



PROCESO GESTIÓN DEL TALENTO HUMANO
FORMATO INFORME MENSUAL EJECUCIÓN CONTRACTUAL

Medellín, 31 de mayo del 2022

Señor (a)
John Jairo Castro Maldonado
SUPERVISOR(A) CONTRATO No. CO1.PCCNTR.3367636
Instructor Grado 16
Centro de Servicios y Gestión Empresarial
Medellín

Asunto: Informe mensual de ejecución contractual de mayo del año 2022.

Referencia: No 3367636 del año 2022

Jesus Arbey Benavides Guevara, identificado con la cédula de ciudadanía No. 1.085.260.713 de Pasto, en mi calidad de Contratista del SENA, en Centro de Servicios y Gestión Empresarial, en cumplimiento del Contrato de Prestación de Servicios de la referencia, a continuación presento el Informe de actividades realizadas en el mes objeto de cobro.

Valor y forma de Pago: Se fija como valor total para el contrato la suma de TREINTA Y CUATRO MILLONES TRESCIENTOS MIL PESOS MCTE (\$ 34.300.000) Incluido IVA. Esta suma será pagada por el SENA al contratista de la siguiente manera: Valor por mes: CUATRO MILLONES NOVECIENTOS MIL PESOS MCTE (\$ 4.900.000), y el equivalente por fracción de mes.

Plazo: Será hasta el treinta (30) de agosto (08) de 2022.

OBJETO: Prestar servicios profesionales para el desarrollo de las actividades de levantamiento de indicadores, bases de datos y el modelo matemático con base en analítica de datos y Big Data en el prototipo de software para modelo de Innovación en el desarrollo del proyecto de I+D+i SGPS 9748

Obligaciones Específicas:

No	Obligaciones	Acciones realizadas	Evidencias
1	Apoyar el desarrollo de aplicaciones o algoritmos usados en la actualidad para solucionar de forma eficiente y eficaz problemas o	Se realizó avances en la etapa de comprensión respecto a las dimensiones del modelo de indicadores para la gestión de la innovación, con enfoque en la dimensión productividad	ANEXOS 4. Reuniones de equipo - discusión de dimensiones en modelo de indicadores ANEXOS 6. Actas de reuniones



	retos identificados en Big Data y analítica de datos con aplicación en gestión de la innovación	<p>con respecto a las variables del modelo relacional.</p> <p>Se implementó modelos de Kmeans y PCA para la determinación de conglomerados en la clasificación de centros de formación en torno a su productividad.</p> <p>Se realizó mapas de visualización de variables estadísticos, valores de frecuencia en cantidad de productos; cantidad de proyectos y clustering por centros de formación.</p>	<p>ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory</p> <p>ANEXOS 5. Presentación de avances técnicos mayo 2022.</p> <p>ANEXOS 6. Actas de reuniones</p> <p>ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory</p> <p>ANEXOS 5. Presentación de avances técnicos mayo 2022.</p> <p>ANEXOS 6. Actas de reuniones</p>
2	Revisar las herramientas tecnológicas o digitales de Big Data y analítica de datos necesarias para dar cumplimiento a los objetivos propuestos dentro del proyecto.	Se implementó etapas de preprocesamiento de los datos, aumentando la cantidad de información para los modelos de machine learning no supervisados en la plataforma Colaboratory google.	ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory
3	Revisar y actualizar indicadores para la gestión de la innovación según objetivos del proyecto	Se determinó dos nuevos <i>dataframes</i> en pro de la determinación de cuantificar indicadores para la gestión de la innovación - proyecto SGPS 9748 prototipo	<p>ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory</p> <p>ANEXOS 4. Reuniones de equipo - discusión de dimensiones en modelo de indicadores</p>



		innovación, desde las dimensiones del modelo teórico a partir de Principal_Productos; Principal_Proyectos ; Principal_Talento_Humano; Principal_Equipos. Se realiza una etapa de preprocesamiento.	
4	Realizar pruebas relacionadas con la compatibilidad de la tecnología y la aplicación de los sistemas de Big Data y analítica de datos para la evaluación y valoración de la Gestión de CTel para los Centros SENA Regional Antioquia	Se implementó etapas de preprocesamiento de los datos, aumentando la cantidad de información para los modelos de machine learning no supervisados en la plataforma Colaboratory google.	ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory
5	Apoyar la elaboración del manual del sistema y usuario y las actividades de transferencia	Gestión por realizar	Gestión por realizar
6	Presentar los informes de seguimiento del proyecto de acuerdo con las solicitudes del líder del área o supervisor	Se elaboró el informe de pago mensual GTH-F-062 V08. Se elaboró diapositivas en presentación de informes técnicos para el supervisor y coordinador del contrato del mes en curso.	Informe en formato pdf. ANEXOS 5. Presentación de avances técnicos mayo 2022
7	Divulgar en eventos académicos/empresariales los	Se participó y se asesoró en la modalidad de coautor en un (1) evento de divulgación de ciencia, tecnología e innovación	ANEXO 1. Medios de divulgación: Apoyo y participación de semilleristas en evento Redcolsi



	resultados del proyecto.	para aprendices de semilleros, evento Redcolsi XXI ENCuentro DEPARTAMENTAL DE SEMILLEROS DE INVESTIGACIÓN. Nodo Antioquia.	
8	Redactar el informe de investigación y de resultados del proyecto.	Se elaboró un (1) informe de pago mensual GTH-F-062 V08. Se elaboró diapositivas en presentación de informes técnicos para el supervisor y coordinador del contrato del mes en curso.	Informe en formato pdf. ANEXOS 5. Presentación de avances técnicos mayo 2022.
9	Redactar y someter a publicación artículo de investigación, libro de investigación o capítulo de libro de investigación relacionado con el proyecto, en medio reconocido por Minciencias.	Se inició un primer borrador de la elaboración de artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación.	ANEXOS 3. Inicio de redacción artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación (brainstorming).
10	Apoyar actividades de Grupo de Investigación, Semilleros de investigación y SENNOVA que sean propias de su área de conocimiento y en relación con el objeto contractual	Se apoyó desde la parte técnica en la visualización de mapas de actividades de CTel para la presentación de la aprendiz Karen Estrada en el evento Redcolsi XXI ENCuentro DEPARTAMENTAL DE SEMILLEROS DE INVESTIGACIÓN. Nodo Antioquia.	ANEXO 1. Medios de divulgación: Apoyo y participación de semilleristas en evento Redcolsi ANEXOS 5. Presentación de avances técnicos mayo 2022.



11	Apoyar la evaluación técnica, administrativa y presupuestal de los proyectos de investigación aplicada, desarrollo tecnológico e innovación que se formulen por el SENA para el desarrollo de actividades misionales, acorde con las indicaciones dadas por la Dirección General.	Gestión por realizar	Gestión por realizar
12	Desarrollar sus actuaciones en el marco de la normatividad vigente y presentar los informes mensuales de acuerdo con el objeto del presente contrato.	Se elaboró un (1) informe de pago mensual GTH-F-062 V08.	Informe en formato pdf.

A continuación relaciono los desplazamientos que realicé previo a la presentación de este informe. Una vez finalizado cada desplazamiento presenté al ordenador del gasto el informe en el Formato Informe Legalización Desplazamiento Contratista GTH-F-087, en el que se describieron las actividades desarrolladas y los resultados de cada desplazamiento. Cada informe cuenta con el visto bueno del Supervisor.

Se lista a continuación el soporte de la legalización de los desplazamientos realizados, los cuales forman parte integral del presente informe de ejecución contractual.

ITEM	No DE LA ORDEN DE VIAJE	LUGAR DE DESPLAZAMIENTO	FECHAS DE DESPLAZAMIENTOS	
			INICIO	FINALIZACIÓN
1.		No he realizado ningún desplazamiento		



Nota 1: Por cada desplazamiento que haya realizado el contratista, adjuntará el respectivo informe que la soporte. En caso de haber realizado el desplazamiento en fecha posterior a la presentación del informe de ejecución contractual, deberá reportarlo en el siguiente informe de ejecución contractual.

Para el trámite de la cuenta me permito adjuntar: Documentos electrónicos enunciados como evidencias del cumplimiento de las obligaciones contractuales y los desplazamientos realizados y el No.58568570 de la planilla, miplanilla correspondiente al mes de abril del año 2022. (Decreto Ley 2106 de 2019 – “Decreto Ley Antitrámites”)

Declaración bajo la gravedad de juramento.

Me permito declarar **bajo la gravedad de juramento** que durante el periodo objeto de cobro SI___NO_X_ he suscrito otro u otros contratos con entidad(es) del Estado el (los) cual(es) se encuentra(n) en ejecución; la anterior declaración de acuerdo con lo establecido en el numeral 5º de la Directiva Presidencial No. 001 de 2022, a fin de identificar potenciales conflictos de interés y hacer la gestión del riesgos para la toma de decisiones dentro del principio de la debida diligencia frente a la ejecución de los contratos de prestación de servicios profesionales o de apoyo a la gestión.

Nota: En el evento de señalar **SI**, el análisis efectuado por el supervisor y ordenador de pago deberá quedar documentado y publicado en el SECOP II en los documentos relacionados en el archivo Gestión Contractual GC_ del periodo objeto de cobro.

Evidencias en 158 folios:

- ANEXOS.
- <https://drive.google.com/drive/folders/1j61SX1uD8L8vP3UX7BeYt0NUtED3ZxwB?usp=sharing>

Cordialmente,

JESUS ARBEY BENAVIDES GUEVARA
Contratista
C.C. No. 1.085.260.713 DE PASTO (NARIÑO)



Recibí a satisfacción:

JOHN JAIRO CASTRO MALDONADO
SUPERVISOR(A) CONTRATO No. CO1.PCCNTR.3367636 DE 2022
INSTRUCTOR GRADO 16
C.C. No. 88.265.552



ANEXOS



ANEXO 1. Medios de divulgación: Apoyo y participación de semilleristas en evento Redcolsi

XXI ENCUENTRO DEPARTAMENTAL DE SEMILLEROS DE INVESTIGACIÓN.

Fecha de inicio: 2022-05-11

Fecha de finalización: 2022-05-13

Medellín, Colombia

Organiza: Fundación Redcolsi

http://fundacionredcolsi.org/eventos/index.php?evento_id=330



Mapas estadística descriptiva y analítica



Jesus Arbey Benavide Guevara <jabenavidesgu@misena.edu.co>
para kestrada97, ERIKA

mié, 4 may, 15:51 (hace 9 días)



Estimada Karen,

sigue a continuación los archivos en formato .html de los mapas. Recordando lo tratado el día de hoy:

Mapas de estadística descriptiva:

1. Productos_Centros
2. Poyectos_Centro
3. Mapas_Gráficas_Costos_Proyectos

Analítica -Machine Learning
Clusters_Centros

No dude en escribir cualquier duda.

Cordialmente,
Arbey



ANEXO 2. Modelos de Machine Learning No Supervisado y visualización en mapas interactivos - Collaboratory

 Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

```
[32] import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

import folium # plotting library
from folium import plugins
from folium.plugins import MarkerCluster

from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
import sklearn.metrics as sm

#Cluster
from yellowbrick.cluster import SilhouetteVisualizer
from yellowbrick.cluster import KElbowVisualizer
from sklearn.cluster import KMeans
from sklearn.datasets import make_blobs

#Graficas
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
```

Completed at 08:41



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Comentar

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

```
+ Código + Texto
[32] from google.colab import drive
[33] drive.mount('/content/drive')
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
[34] from google.colab import files
path = "/content/drive/MyDrive/Proyecto Sennova borrador/Copia de Datos_proyecto_R.xlsx"

tipologia_productos = pd.read_excel(path,sheet_name= 'Tipologia_Productos')
tipo_contratacion = pd.read_excel(path,sheet_name= 'Tipo_Contratacion')
semilleros = pd.read_excel(path,sheet_name= 'Semilleros')
rubro = pd.read_excel(path,sheet_name= 'Rubro')
rol_sennova = pd.read_excel(path,sheet_name= 'Rol_Sennova')
grupo_investigacion = pd.read_excel(path,sheet_name= 'Grupo_investigacion')
area_conocimiento_ocde = pd.read_excel(path,sheet_name= 'Area_conocimiento_OCDE')
estado_proyecto = pd.read_excel(path,sheet_name= 'Estado_Proyecto')
nivel_academico = pd.read_excel(path,sheet_name= 'Nivel_Academico')
municipio = pd.read_excel(path,sheet_name= 'Municipio')
centros_formacion= pd.read_excel(path,sheet_name= 'Centros_Formacion')
departamento = pd.read_excel(path,sheet_name= 'Departamento')
regionales = pd.read_excel(path,sheet_name= 'Regionales')
tipo_equipo = pd.read_excel(path,sheet_name= 'Tipo_equipo')
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

```
+ Código + Texto
[34] centros_formacion= pd.read_excel(path,sheet_name= 'Centros_Formacion')
departamento = pd.read_excel(path,sheet_name= 'Departamento')
regionales = pd.read_excel(path,sheet_name= 'Regionales')
tipo_equipo = pd.read_excel(path,sheet_name= 'Tipo_equipo')
rol_proyectos = pd.read_excel(path,sheet_name= 'Rol_Proyectos')
estado_contratacion = pd.read_excel(path,sheet_name= 'Estado_Contratacion')

#Dataframes
principal_proyectos= pd.read_excel(path,sheet_name= 'Principal_Proyectos')
principal_talento_humano = pd.read_excel(path,sheet_name= 'Principal_Talento_Humano')
principal_productos = pd.read_excel(path,sheet_name= 'Principal_Productos')
principal Equipos = pd.read_excel(path,sheet_name= 'Principal_Equipos')

[35] prod_1 = principal_proyectos.merge(centros_formacion,on='Codigo_Centro',how='inner')
prod_1.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 500 entries, 0 to 499
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Codigo_Proyecto        500 non-null    int64
1   Nombre_Proyecto        500 non-null    object
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto



```
0 s dtypes: datetime64[ns](2), float64(2), int64(15), object(2)
memory usage: 85.9+ KB
```

```
✓ [36] f = prod_1.groupby(['Nombre_Centro']).count()
0 s f1 = f[['Nombre_Proyecto']].reset_index()
f1
```

	Nombre_Centro	Nombre_Proyecto
0	Centro_Tecnológico_del_Mobiliario	28
1	Centro_Textil_y_de_Gestión_Industrial	37
2	Centro_de_Comercio	37
3	Centro_de_Formación_Minero_Ambiental	23
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32
5	Centro_de_Servicios_de_Salud	27
6	Centro_de_Servicios_y_Gestión_Empresarial	30
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37
9	Centro_de_los_Recursos_Naturales_Renovables_La	42





Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

```
+ Código + Texto

[37] prod_2 = principal_productos.merge(centros_formacion,on='Codigo_Centro',how='inner')
prod_2.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 500 entries, 0 to 499
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype  
---  -
0   Codigo_Producto                       500 non-null    int64  
1   Nombre_Producto                       500 non-null    object  
2   Descripcion_producto                  0 non-null      float64 
3   Codigo_Regional_x                     500 non-null    int64  
4   Codigo_Centro                         500 non-null    int64  
5   Codigo_Rubro                           500 non-null    int64  
6   Codigo_Proyecto                       500 non-null    int64  
7   Nombre_Proyecto                       500 non-null    int64  
8   Codigo_Grupo_Investigacion            500 non-null    object  
9   Codigo_Semillero                       500 non-null    int64  
10  Codigo_persona                         500 non-null    int64  
11  Codigo_area_OCDE                       500 non-null    int64  
12  Fecha_registro                         500 non-null    datetime64[ns]
13  Nombre_Centro                         500 non-null    object  
14  Codigo_Regional_y                     500 non-null    int64  
15  Codigo_Municipio                       500 non-null    int64  
16  Codigo_Departamento                   500 non-null    int64  
17  latitud                               500 non-null    float64 
18  longitud                              500 non-null    float64
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

Comentario Compartir

```
+ Código + Texto

[38] 14 Complejo_Tecnológico_Turístico_y_Agroindustria... 34
     15 Complejo_Tecnológico_para_la_Gestión_Agroempre... 26

[39] proy_centro = principal_proyectos.merge(centros_formacion,on='Codigo_Centro',how='inner')
proy_centro

Codigo_Proyecto  Nombre_Proyecto  Valor_Proyecto  Codigo_Rubro  Fecha_inicio_Proyecto  Fecha_Cierre_Proyecto  Codigo_Regional_x  Codigo_Centro  Codigo_Estado
0               10651  Corporación de tecnología Carahsoft  147801416      65      2010-04-30      2017-11-06          5          9501
1               1028      Infosys  45734896       23      2019-11-18      2011-02-26          5          9501
2               11490  DuPontmakers Technology Inc  166624849      23      2010-11-19      2013-08-09          5          9501
3               7349  NetStandard Inc  230668625      23      2016-10-21      2013-05-31          5          9501
4               3393  SoftRock Inc  175797775      65      2021-11-25      2014-04-04          5          9501
...           ...           ...           ...           ...           ...           ...           ...           ...
495           1319  Señales de alta tecnología  48766169      82      2013-12-28      2016-09-22          5          9504
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆



Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)



+ Código + Texto



15 Complejo_Tecnológico_para_la_Gestión_Agroempre... 7.996707 -75.196365

```
[41] df1 = pd.DataFrame()
df1['Nombre_Centro'] = f1['Nombre_Centro']
df1['Cantidad_Proyectos'] = f1['Nombre_Proyecto']
df1['Cantidad_Productos'] = g1['Nombre_Producto']
df1['latitud'] = h2['latitud']
df1['longitud'] = h2['longitud']
df1
```

	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365
2	Centro_de_Comercio	37	29	6.254794	-75.575466
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

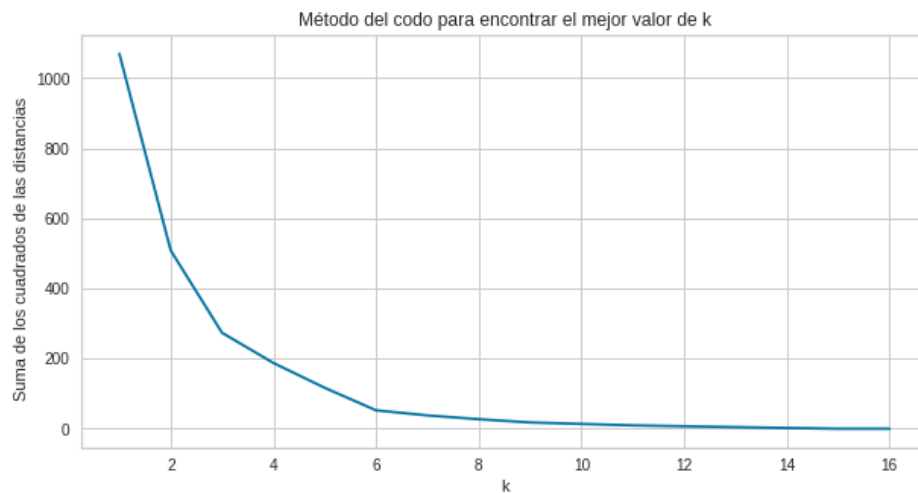
+ Código + Texto

casi horizontal,ese es el número óptimo de conglomerados.

```
codes = df1[['Nombre_Centro']]
types = df1.drop(['Nombre_Centro'], axis= 1)
distortion = []

K = range(1,17)
for k in K:
    kmean = KMeans(n_clusters=k, random_state=0, n_init = 50, max_iter = 500)
    kmean.fit(types)
    distortion.append(kmean.inertia_)
```

```
[43] plt.figure(figsize=(10,5))
plt.plot(K, distortion, 'bx-')
plt.xlabel('k')
plt.ylabel('Suma de los cuadrados de las distancias ')
plt.title('Método del codo para encontrar el mejor valor de k')
plt.show()
```



```
[44] # run k-means clustering
kmeans = KMeans(n_clusters=3, random_state=0).fit(types)

codes['cluster'] = kmeans.labels_
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb



Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)



+ Código + Texto



```
[44] # run k-means clustering
kmeans = KMeans(n_clusters=3, random_state=0).fit(types)

codes['cluster'] = kmeans.labels_

codes
```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
after removing the cwd from sys.path.

Nombre_Centro cluster



0	Centro_Tecnológico_del_Mobiliario	2
1	Centro_Textil_y_de_Gestión_Industrial	1
2	Centro_de_Comercio	1
3	Centro_de_Formación_Minero_Ambiental	0
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	0
5	Centro_de_Servicios_de_Salud	2
6	Centro_de_Servicios_y_Gestión_Empresarial	0
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	1
9	Centro_de_los_Recursos_Naturales_Renovables_La...	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	2
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	0
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	0
13	Complejo_Tecnológico_Minero_Agroempresarial_...	0
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	0
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	1





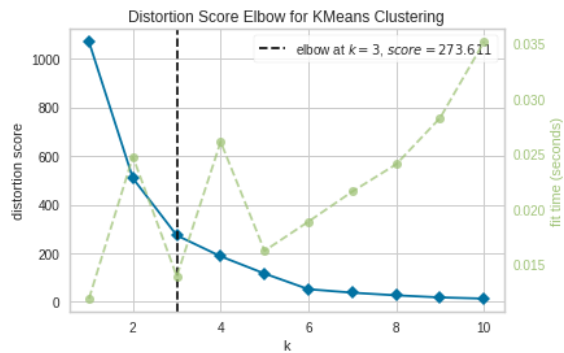
Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

Métrica *distortion*, que calcula la suma de las distancias al cuadrado desde cada punto hasta su centro asignado

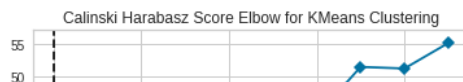
```
[45] #km_transformed = mms.transform(km)
model = KMeans(random_state=42)
visualizer = KElbowVisualizer(model, k=(1,11))
visualizer.fit(types)
visualizer.show();
```



La métrica *calinski_harabasz* calcula la relación de dispersión entre y dentro de los grupos.

```
[46]
model = KMeans()
visualizer = KElbowVisualizer(
    model, k=(2,12), metric='calinski_harabasz', timings=False
)

visualizer.fit(types)      # Fit the data to the visualizer
visualizer.show()          # Finalize and render the figure
```





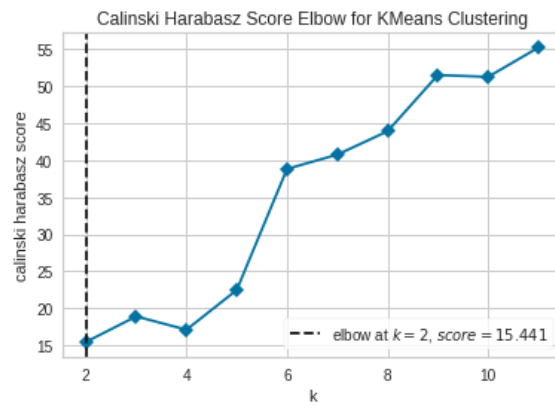
Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

```
[46]
model = KMeans()
visualizer = KElbowVisualizer(
    model, k=(2,12), metric='calinski_harabasz', timings=False
)

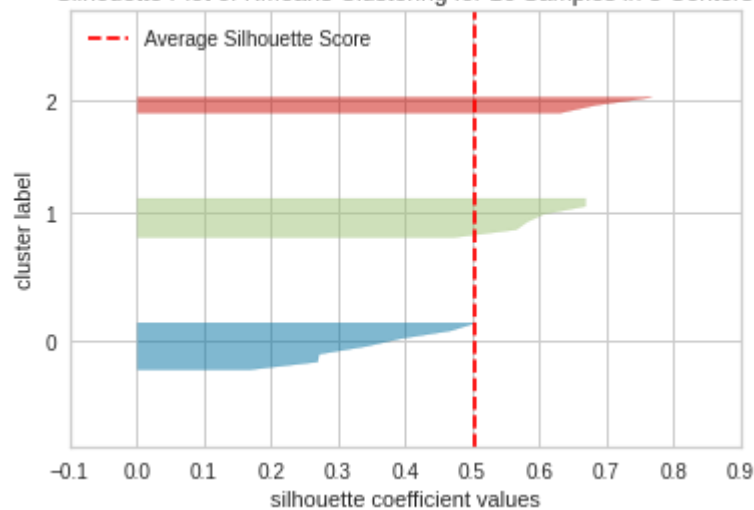
visualizer.fit(types)      # Fit the data to the visualizer
visualizer.show()         # Finalize and render the figure
```



<matplotlib.axes._subplots.AxesSubplot at 0x7f263ae17d50>

<https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>

Silhouette Plot of KMeans Clustering for 16 Samples in 3 Centers





Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

dimensión. Adicionalmente, podemos ver si las características de nuestros datos están linealmente relacionadas entre sí.

```
[52] pca = PCA().fit(types)
      pca_types = pca.transform(types)
      print("Varianza explicada por cada componente (%): ")
      for i in range(len(pca.explained_variance_ratio_)):
          print("\n",i+1,"º:", pca.explained_variance_ratio_[i]*100)
      print("Suma total (%): ",sum(pca.explained_variance_ratio_)*100)
      print("Varianza explicada de los dos primeros componentes (%): ",sum(pca.explained_variance_ratio_[0:1])*100)
```

Varianza explicada por cada componente (%):

1 º: 69.22433522843373

2 º: 29.934838638550627

3 º: 0.5967323711309355

4 º: 0.2440937618847172

Suma total (%): 100.00000000000003

Varianza explicada de los dos primeros componentes (%): 69.22433522843373

Podemos observar que los 5 componentes principales (PC) llevan al 100 % de la información original que permiten explicar y modelar los datos.

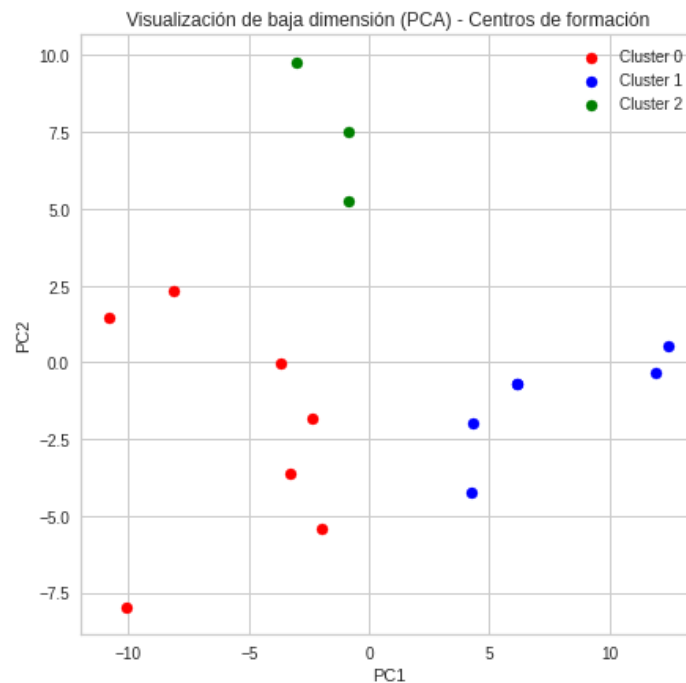
Podemos ver que utilizando los dos primeros componentes somos capaces de conservar el 68.72% de la información original, por lo tanto, reduce la dimensionalidad de nuestros datos. Usamos, por lo tanto, estas PC para visualizar nuestros clústeres.

```
[53] c0 = []
      c1 = []
      c2 = []

      for i in range(len(pca_types)):
          if kmeans.labels_[i] == 0:
              c0.append(pca_types[i])
          if kmeans.labels_[i] == 1:
              c1.append(pca_types[i])
          if kmeans.labels_[i] == 2:
              c2.append(pca_types[i])

      c0 = np.array(c0)
      c1 = np.array(c1)
      c2 = np.array(c2)

      plt.figure(figsize=(7,7))
      plt.scatter(c0[:,0], c0[:,1], c='red', label='Cluster 0')
      plt.scatter(c1[:,0], c1[:,1], c='blue', label='Cluster 1')
      plt.scatter(c2[:,0], c2[:,1], c='green', label='Cluster 2')
      plt.legend()
      plt.xlabel('PC1')
      plt.ylabel('PC2')
      plt.title('Visualización de baja dimensión (PCA) - Centros de formación');
```



Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

```
[54] types['cluster'] = kmeans.labels_  
types.head()
```

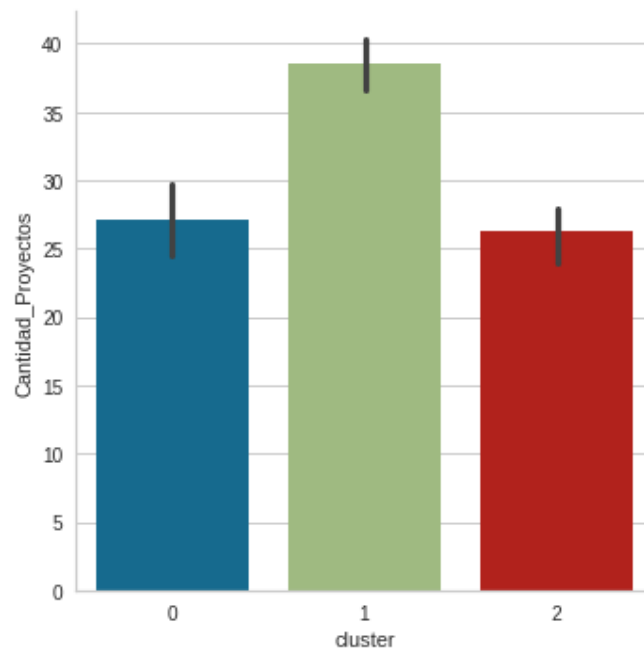
	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud	cluster
0	28	27	6.180750	-75.606430	2
1	37	29	6.302387	-75.567365	1
2	37	29	6.254794	-75.575466	1
3	23	33	7.584361	-74.800660	0
4	32	37	6.181273	-75.605935	0

```
[56] types.groupby('cluster').mean()
```

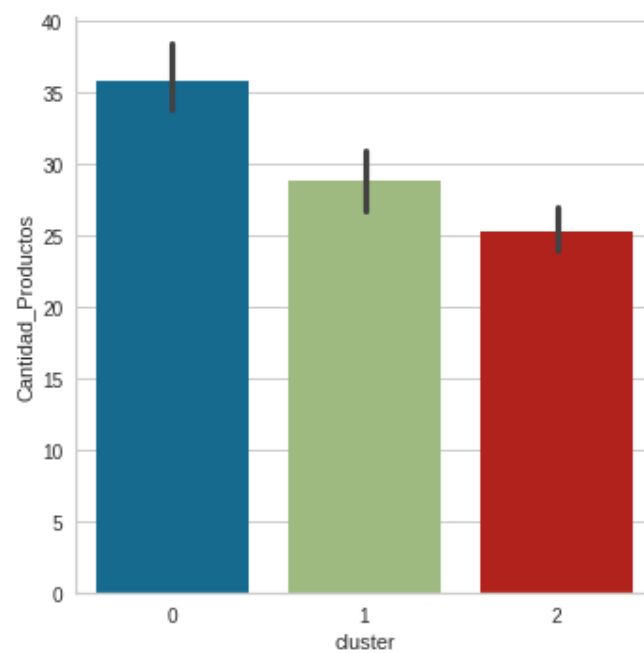
	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud
cluster				
0	27.142857	35.857143	6.836398	-75.529837
1	38.500000	28.833333	6.510331	-75.484777
2	26.333333	25.333333	6.204311	-75.594628



✓ [57] `sns.catplot(x='cluster', y='Cantidad_Proyectos', data=types, kind='bar');`
0 s



✓ [58] `sns.catplot(x='cluster', y='Cantidad_Productos', data=types, kind='bar');`
0 s





[59]

	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud	cluster
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430	2
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365	1
2	Centro_de_Comercio	37	29	6.254794	-75.575466	1
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660	0
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935	0
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571	2
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671	0
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	31	6.300949	-75.568664	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37	33	6.153989	-75.372317	1
9	Centro_de_los_Recurso_Naturales_Renovables_La...	42	25	6.053157	-75.628486	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	24	6.180794	-75.605884	2
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	26	43	6.303487	-75.567591	0
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	33	7.890489	-76.631035	0
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	35	7.080799	-74.699840	0
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	34	6.559073	-75.829128	0
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	26	7.996707	-75.196365	1

```
[62] ocurr_product = df1.drop(['Cantidad_Proyectos'], axis =1)
      ocurr_product
```

	Nombre_Centro	Cantidad_Productos	latitud	longitud	cluster
0	Centro_Tecnológico_del_Mobiliario	27	6.180750	-75.606430	2
1	Centro_Textil_y_de_Gestión_Industrial	29	6.302387	-75.567365	1
2	Centro_de_Comercio	29	6.254794	-75.575466	1
3	Centro_de_Formación_Minero_Ambiental	33	7.584361	-74.800660	0
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	37	6.181273	-75.605935	0
5	Centro_de_Servicios_de_Salud	25	6.251389	-75.571571	2
6	Centro_de_Servicios_y_Gestión_Empresarial	36	6.255308	-75.574671	0
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	31	6.300949	-75.568664	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	33	6.153989	-75.372317	1
9	Centro_de_los_Recurso_Naturales_Renovables_La...	25	6.053157	-75.628486	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	6.180794	-75.605884	2
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	43	6.303487	-75.567591	0
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	33	7.890489	-76.631035	0
13	Complejo_Tecnológico_Minero_Agroempresarial_	35	7.080799	-74.699840	0
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	34	6.559073	-75.829128	0
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	26	7.996707	-75.196365	1



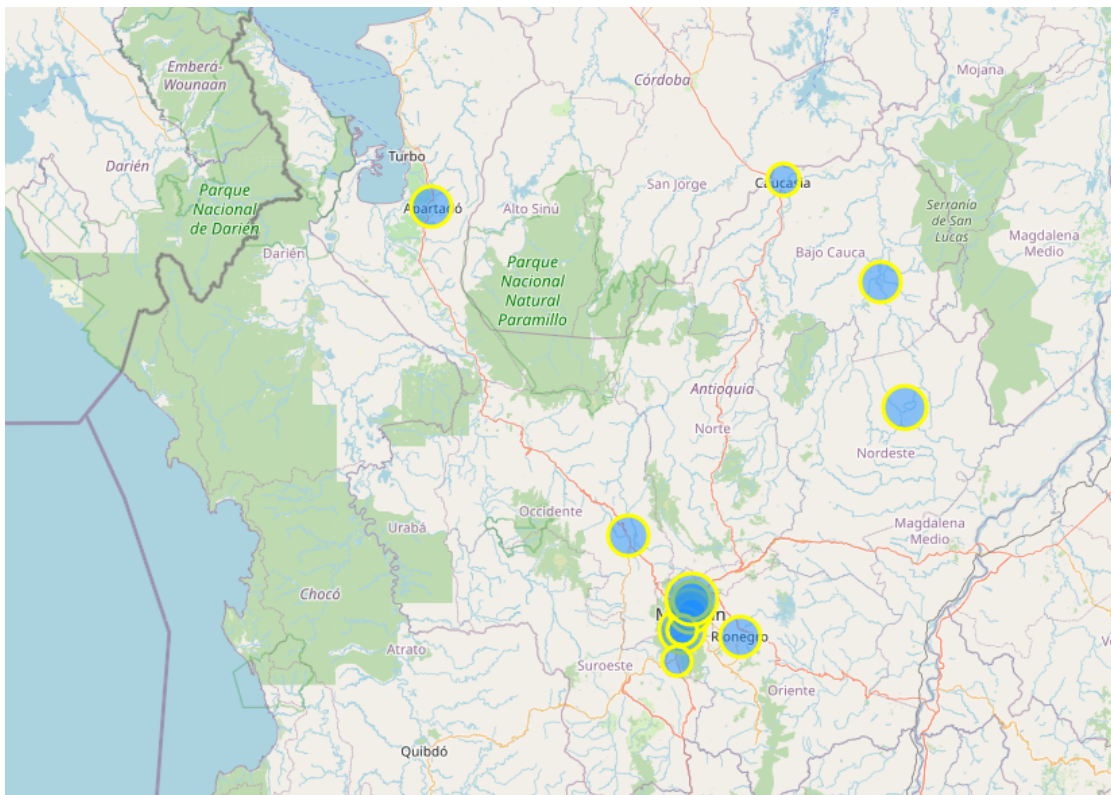
```
[63] #latitude = 7.199941990952388
#longitude = -75.75991310406549
latitude = 6.51033052
longitude = -75.48477721

map_Antioquia = folium.Map(location=[latitude, longitude], zoom_start=8)

productos = folium.map.FeatureGroup()
n_mean = occur_product['Cantidad_Productos'].mean()

for lat, lng, number, cent in zip(occur_product['latitud'],
                                occur_product['longitud'],
                                occur_product['Cantidad_Productos'],
                                occur_product['Nombre_Centro']):
    productos.add_child(
        folium.vector_layers.CircleMarker(
            [lat, lng],
            radius=number/n_mean*14, # define qué tan grande quieres que sean los marcadores circulares
            color='Yellow',
            fill=True,
            fill_color='DodgerBlue',
            fill_opacity=0.5,
            tooltip = str(cent)+'', productos: '+str(number)
        )
    )

P_Centro = map_Antioquia.add_child(productos)
P_Centro
```





```
✓ #Descarga um arquivo .html  
0s P_Centro.save("Productos_Centro.html")
```

```
✓ [65] ocurr_project = df1.drop(['Cantidad_Productos'], axis =1)  
0s ocurr_project
```

	Nombre_Centro	Cantidad_Proyectos	latitud	longitud	cluster
0	Centro_Tecnológico_del_Mobiliario	28	6.180750	-75.606430	2
1	Centro_Textil_y_de_Gestión_Industrial	37	6.302387	-75.567365	1
2	Centro_de_Comercio	37	6.254794	-75.575466	1
3	Centro_de_Formación_Minero_Ambiental	23	7.584361	-74.800660	0
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	6.181273	-75.605935	0
5	Centro_de_Servicios_de_Salud	27	6.251389	-75.571571	2
6	Centro_de_Servicios_y_Gestión_Empresarial	30	6.255308	-75.574671	0
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	6.300949	-75.568664	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la...	37	6.153989	-75.372317	1
9	Centro_de_los_Recurso_Naturales_Renovables_La...	42	6.053157	-75.628486	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	6.180794	-75.605884	2
11	Centro_para_el_Desarrollo_del_Habitat_y_la_Con...	26	6.303487	-75.567591	0
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	7.890489	-76.631035	0
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	7.080799	-74.699840	0
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	6.559073	-75.829128	0
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	7.996707	-75.196365	1

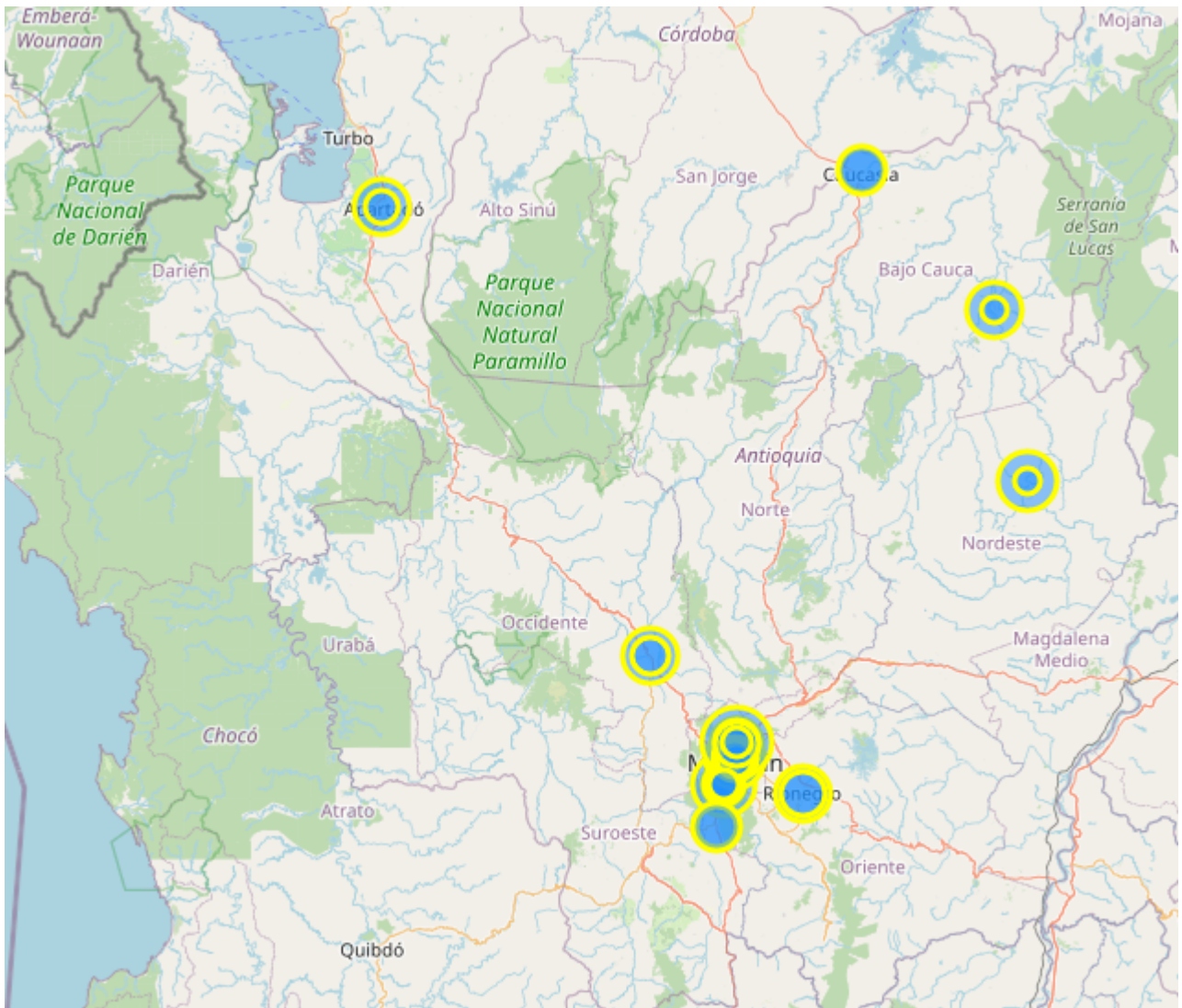


Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

```
✓ [66] proyectos = folium.map.FeatureGroup()  
0s n_mean = ocurr_project['Cantidad_Proyectos'].mean()  
  
for lat, lng, number, cent in zip(ocurr_project['latitud'],  
                                ocurr_project['longitud'],  
                                ocurr_project['Cantidad_Proyectos'],  
                                ocurr_project['Nombre_Centro']):  
  
    proyectos.add_child(  
        folium.vector_layers.CircleMarker(  
            [lat, lng],  
            radius=number/n_mean*10, # define qué tan grande quieres que sean los marcadores circulares  
            color= "yellow",  
            fill=True,  
            fill_color= "DodgerBlue",  
            fill_opacity=0.5,  
            tooltip = str(cent)+'', proyectos: '+str(number)  
        )  
    )  
  
Pry_Centro = map_Antioquia.add_child(proyectos)  
Pry_Centro
```





```
[67] Pry_Centro.save("Proyectos_Centro.html")
```

```
[68] #Mapa con formas para modificar
      #draw = plugins.Draw(export=True)
      #draw.add_to(map_Antioquia)
      #map_Antioquia
```

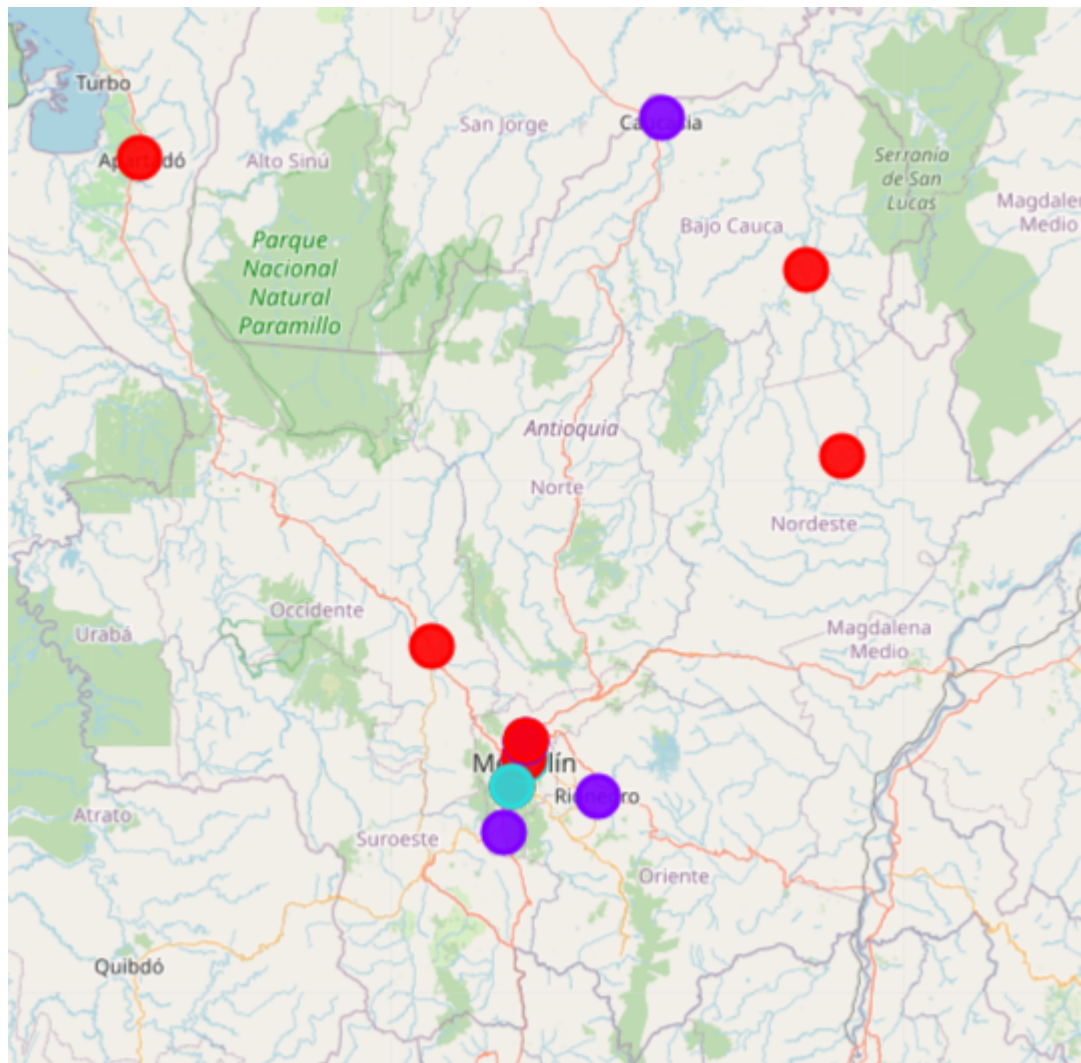
Mapas de Clusters centros de formación

```
[69] map_clusters = folium.Map(location=[latitude, longitude], zoom_start=4)

      # set color scheme for the clusters
      x = np.arange(4)
      ys = [i + x + (i*x)**2 for i in range(4)]
      colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
      rainbow = [colors.rgb2hex(i) for i in colors_array]

      # add markers to the map
      markers_colors = []
      for lat, lng, cluster, cent in zip(ocurr_product['latitud'],
                                         occur_product['longitud'],
                                         occur_product['cluster'],
                                         occur_product['Nombre_Centro']):

          label = folium.Popup(str(cent) + '- Cluster ' + str(cluster), parse_html=True)
          folium.vector_layers.CircleMarker(
              [lat, lng],
              radius=10,
              #popup=label,
              tooltip = str(cent) + '- Cluster ' + str(cluster),
              color=rainbow[cluster-1],
              fill=True,
              fill_color=rainbow[cluster-1],
              fill_opacity=0.9).add_to(map_clusters)
```





Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

+ Código + Texto

0 s map_clusters.save("Clusters_Centros.html")

Exploración de base de datos con dummies



0 s [71] principal_productos.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Codigo_Producto                      500 non-null    int64
1   Nombre_Producto                      500 non-null    object
2   Descripcion_producto                 0 non-null      float64
3   Codigo_Regional                      500 non-null    int64
4   Codigo_Centro                       500 non-null    int64
5   Codigo_Rubro                        500 non-null    int64
6   Codigo_Proyecto                     500 non-null    int64
7   Nombre_Proyecto                     500 non-null    int64
8   Codigo_Grupo_Investigacion          500 non-null    object
9   Codigo_Semillero                     500 non-null    int64
10  Codigo_persona                       500 non-null    int64
11  Codigo_area_OCDE                     500 non-null    int64
12  Fecha_registro                       500 non-null    datetime64[ns]
dtypes: datetime64[ns](1), float64(1), int64(9), object(2)
memory usage: 50.9+ KB
```

0 s [72] df_pro = principal_productos[['Nombre_Producto', 'Codigo_Centro']]
df_pro = centros_formacion.merge(df_pro, on='Codigo_Centro', how='inner')
df_pro = df_pro[['Nombre_Centro', 'Nombre_Producto', 'latitud', 'longitud']]
df_pro

```
prod_dummies1 = pd.get_dummies(df_pro['Nombre_Producto'])
df_pro1 = pd.concat([df_pro, prod_dummies1 ], axis=1)
df_pro2 = df_pro1.drop(['Nombre_Producto', 'latitud', 'longitud'], axis = 1)
df_pro2
```

<>



CO

Sena_Antioquia_Cluster_KNN_Pruebas.ipynb

☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

Comentario Compartir Configuración

+ Código + Texto

[72]

	Nombre_Centro	Apropiación_Social	Artículo_Sometido_en_revista_Indexada	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital	Informe_de_in
0	Centro_de_los_Recursos_Naturales_Renovables_La...	0	0	0	0
1	Centro_de_los_Recursos_Naturales_Renovables_La...	0	0	1	1
2	Centro_de_los_Recursos_Naturales_Renovables_La...	0	0	0	0
3	Centro_de_los_Recursos_Naturales_Renovables_La...	0	0	1	1
4	Centro_de_los_Recursos_Naturales_Renovables_La...	0	0	0	0
...	...	...	...	...	...
485	Complejo_Tecnológico_Turístico_y_Agroindustria...	1	0	0	0
486	Complejo_Tecnológico_Turístico_y_Agroindustria...	0	0	0	0
487	Complejo_Tecnológico_Turístico_y_Agroindustria...	0	0	0	0
488	Complejo_Tecnológico_Turístico_y_Agroindustria...	1	0	0	0
489	Complejo_Tecnológico_Turístico_y_Agroindustria...	0	0	0	0

500 rows x 7 columns

[73]

```
df_pro3 = df_pro1.drop(['Nombre_Producto'],axis = 1)
df_pro3.head()
#df_pro3 = df_pro2.drop(['Nombre_Centro'],axis = 1)
y1 = df_pro3['Nombre_Centro']
X1 = df_pro3.drop(['Nombre_Centro'],axis=1)
```

[74]

```
Sum_of_squared_distances = []
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(X1)
    Sum_of_squared_distances.append(km.inertia_)
```

[75]

```
plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Suma de los cuadrados de las distancias ')
plt.title('Método del codo para encontrar el mejor valor de k')
plt.show()
```

0 s completado a las 21:41

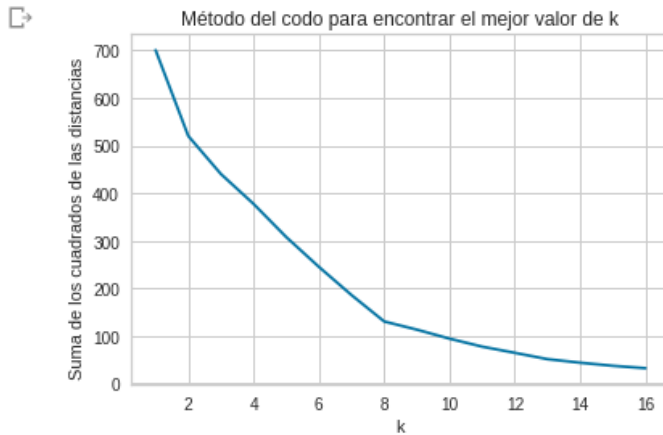


Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)

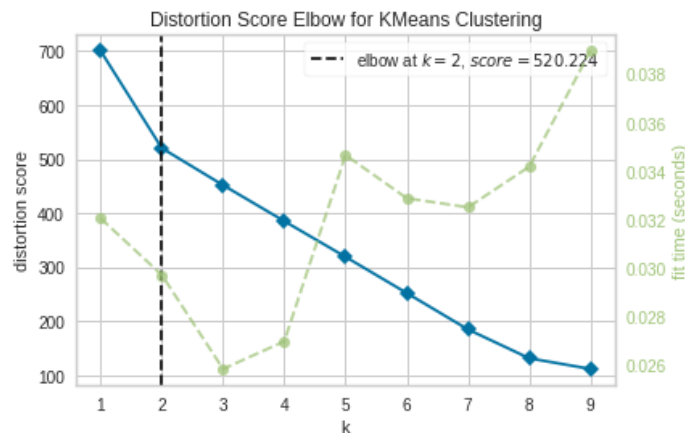
+ Código + Texto

```
[75] plt.title('Método del codo para encontrar el mejor valor de k')  
plt.show()
```



```
[ ] %pip install -U -q dtreeviz seaborn yellowbrick
```

```
[76] from yellowbrick.cluster import KElbowVisualizer  
#km_transformed = mms.transform(km)  
model = KMeans(random_state=42)  
visualizer = KElbowVisualizer(model, k=(1,10))  
visualizer.fit(X1)  
visualizer.show();
```





Sena_Antioquia_Cluster_KNN_Pruebas.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se han guardado todos los cambios](#)



+ Código + Texto



0 s 1 2 3 4 5 6 7 8 9
k

```
[77] # agrupando con KMeans  
km = KMeans(n_clusters = 2)  
km.fit(X1)
```

KMeans(n_clusters=2)

```
[78] #Calcular índices de validação externos do agrupamento com KMeans
```

```
labels_true = y1  
rotulos_km = km.labels_  
h = sm.homogeneity_score(labels_true, rotulos_km)  
print("Homogeneidade: %0.3f" % h)  
c = sm.completeness_score(labels_true, rotulos_km)  
print("Completeza: %0.3f" % c)  
v = sm.v_measure_score(labels_true, rotulos_km)  
print("Medida V: %0.3f" % v )  
ari = sm.adjusted_rand_score(labels_true, rotulos_km)  
print("Índice Rand corrigido: %0.3f" % ari)  
ami = sm.adjusted_mutual_info_score(labels_true, rotulos_km, average_method='arithmetic')  
print("Adjusted Mutual Information: %0.3f" % ami )  
fm = sm.fowlkes_mallows_score(labels_true, rotulos_km)  
print("Índice Fowlkes-Mallows: %0.3f" % fm )
```

Homogeneidade: 0.205
Completeza: 1.000
Medida V: 0.341
Índice Rand corrigido: 0.078
Adjusted Mutual Information: 0.335
Índice Fowlkes-Mallows: 0.317



ANEXOS 3. Inicio de redacción artículo científico orientado en la visualización de mapas para actividades de ciencia tecnología e innovación (*brainstorming*).

Títulos tentativos:

Estrategias de visualización de información en mapas para la presentación de indicadores de Ciencia, Tecnología e Innovación

Estrategias de visualización en mapas para la presentación de información con técnicas de estadísticas descriptiva y analítica de datos

Estrategias de visualización en mapas para la presentación de información con técnicas de estadísticas descriptiva y analítica de datos

Generar un camino para dar pautas para la visualización de datos de información

Introducción

facilidad de comprender grandes cantidades de datos a través de imágenes para la toma de decisiones

Wang, Wang y Alejandro, (2015) publicó que el análisis y la visualización de datos deben integrarse a la perfección para que puede funcionar mejor en cualquier proyecto

Importancia y formas de medir el desempeño en procesos de investigación o en la Ciencia Tecnología e Innovación

Importancia de presentar la información de forma visual en instituciones con presencia a nivel nacional y realizar comparaciones entre sus sedes para quienes es importante (subdirectores ,funcionarios, administrativos y público en general) que desconocen el uso de herramientas para visualizar de manera geográfica grandes volúmenes de información

Tipos de visualización geográfica:

1. Desde la estadística descriptiva: Códigos de visualización por centro de formación con estadística descriptiva
2. Desde la analítica de datos

Herramientas de visualización utilizadas



Materiales y métodos

Diseño del estudio

Etapas del algoritmo

Colección de datos

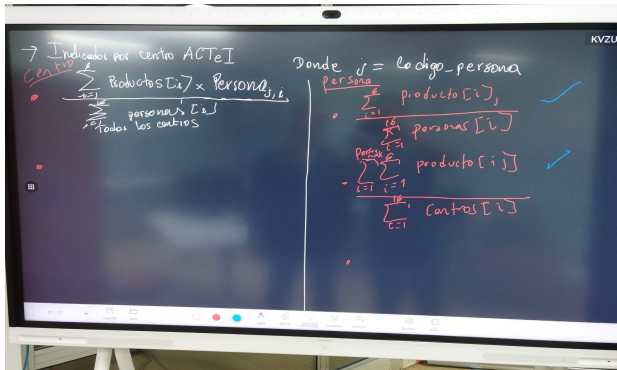
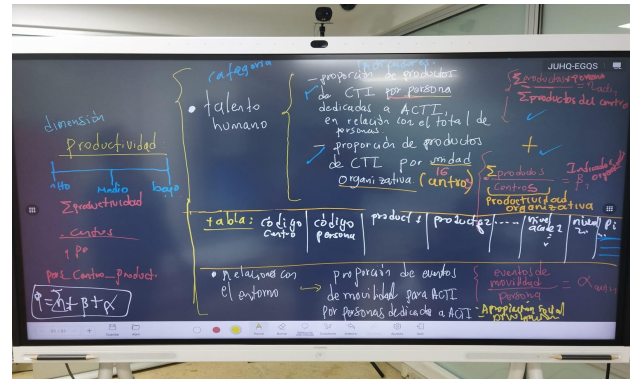
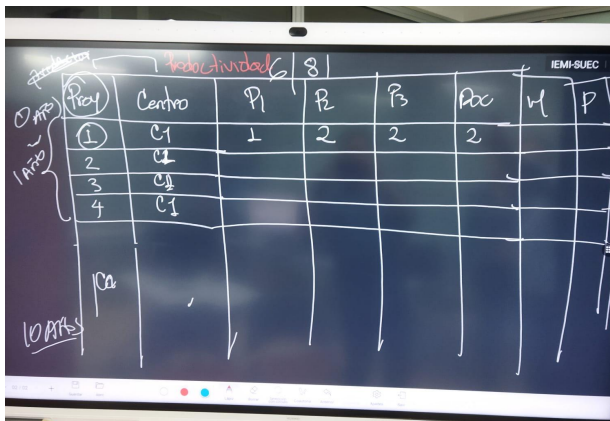
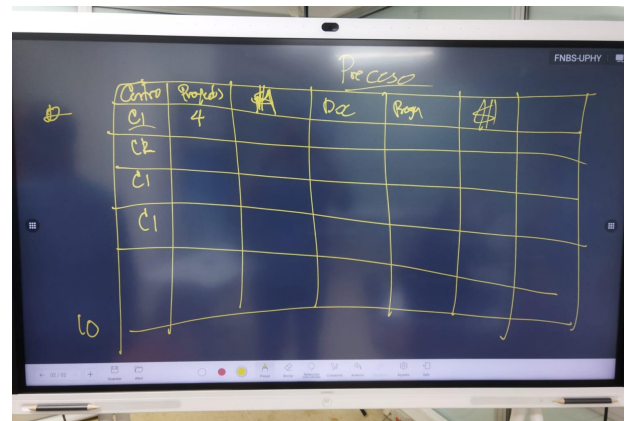
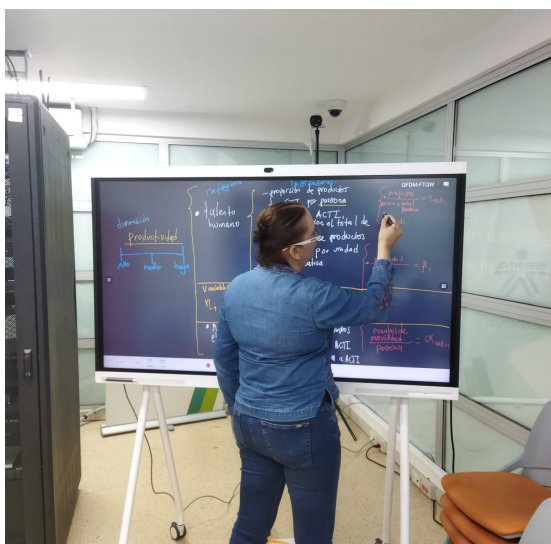
Los datos utilizados fueron en parte de diferentes áreas del SENA como base para la generación de datos sintéticos por medio de un algoritmo aleatorizado

Configuración del estudio

Se utilizó Google Collaboratory (procesamiento de datos)



ANEXOS 4. Reuniones de equipo - discusión de dimensiones en modelo de indicadores





ANEXOS 5. Presentación de avances técnicos mayo 2022.



El empleo
es de todos

Mintrabajo

Proyecto SGPS 9748

Prototipo Innovación

Presentación de avances 07

Modelo de evaluación y valoración de la Gestión de CTel para los Centros SENA

Regional Antioquia aplicando analítica de datos

3 de mayo del 2022



@SENAcomunica

www.sena.edu.co



Contenido

1. Etapas Modelo CRISP - DM - Proyecto SGPS-9748
2. Cronograma 2022 - Proyecto SGPS-9748
3. Actualización y exploración de base de datos
4. Visualización de datos en mapas -KNN Cluster-PCA
5. Modelos de analítica de datos
 - 5.1. Machine learning no supervisado
 - 5.2. Machine learning supervisado
6. Actividades de divulgación



@SENAcomunica

www.sena.edu.co

1. Etapas modelo CRISP - DM - Proyecto SGPS-9748

1. ~~Comprensión del negocio~~
2. ~~Comprensión de los datos~~
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Implementación
7. Modelo de analítica de datos
8. Validación del modelo



2. Cronograma 2022 - Proyecto SGPS-9748



	Fechas programadas	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Fase 1 : Avance de la metodología del Big Data	3 marzo	100%						
Fase 2 : Identificación y selección del Dataset análogo al trabajo en particular	15 abril		100 %					
Fase 3: Caracterizar y seleccionar modelos analítica	19 mayo			45%				
Fase 4: Aplicación de API's para vincular con el desarrollo web	20 junio							
Fase 5 : Puesta en marcha del desarrollo web con los modelos de segmentación y clusterización	21 julio							
Entrega final	15 agosto							



3. Actualización y exploración de base de datos



@SENAcomunica

www.sena.edu.co

Dataframes - Actualización



idigo_Prodect	Nombre_Producto	Descripcio_producto	Codigo_Regional	Codigo_Centro	Codigo_Resbo	Codigo_Proyecto	Nombre_Proyecto	Nombre_Grupo_Investigacion
80033	Articulo_Sometido_en_revista_Indexada		5	32001	62	34602	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80023	Productos_tecnológicos_certificados_o_validados		5	32001	65	44465	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80019	Articulo_Sometido_en_revista_Indexada		5	35003	62	112005	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80016	Libros_resultados_de_investigación		5	32003	65	104041	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80020	Apropiación_Social		5	34001	65	11700	3939	Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA
80013	Apropiación_Social		5	35002	66	43702	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80031	Productos_tecnológicos_certificados_o_validados		5	35004	23	120075	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80016	Libros_resultados_de_investigación		5	35433	66	10102	3939	GIDOCA
80012	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	31001	66	36447	3939	Grupo_de_Investigación_La_Saluda_CILS
80003	Informe_de_investigación		5	32006	65	53448	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_textil_y_a_la_química_GIAITE
80023	Libros_resultados_de_investigación		5	35002	62	607	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80016	Libros_resultados_de_investigación		5	31001	23	25344	3939	Grupo_de_Investigación_La_Saluda_CILS
80035	Informe_de_investigación		5	31017	62	33356	3939	GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bogre
80017	Apropiación_Social		5	34002	23	122551	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_tec
80003	Apropiación_Social		5	32006	23	38222	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_textil_y_a_la_química_GIAITE
80001	Productos_tecnológicos_certificados_o_validados		5	31001	62	55444	3939	Grupo_de_Investigación_La_Saluda_CILS
80001	Libros_resultados_de_investigación		5	32001	23	30338	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80021	Libros_resultados_de_investigación		5	35004	65	103921	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80025	Articulo_Sometido_en_revista_Indexada		5	35003	65	35441	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80020	Productos_tecnológicos_certificados_o_validados		5	32006	62	74744	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_textil_y_a_la_química_GIAITE
80015	Articulo_Sometido_en_revista_Indexada		5	32005	23	27603	3939	Materiales_para_el_Mobiliario_MATERMOB
80019	Informe_de_investigación		5	35003	62	110708	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80040	Productos_tecnológicos_certificados_o_validados		5	35004	62	16511	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80039	Informe_de_investigación		5	32006	62	36556	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_textil_y_a_la_química_GIAITE
80013	Informe_de_investigación		5	31001	65	736	3939	Grupo_de_Investigación_La_Saluda_CILS
80026	Productos_tecnológicos_certificados_o_validados		5	35002	65	1636	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80010	Apropiación_Social		5	32003	62	110010	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80025	Productos_tecnológicos_certificados_o_validados		5	32003	62	118005	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80020	Informe_de_investigación		5	34002	65	30002	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_tec
80019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	32003	23	38335	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80025	Libros_resultados_de_investigación		5	31017	65	30057	3939	GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bogre
80019	Libros_resultados_de_investigación		5	32003	23	12330	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80031	Apropiación_Social		5	35004	23	13966	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80036	Libros_resultados_de_investigación		5	31001	23	84666	3939	Grupo_de_Investigación_La_Saluda_CILS
80028	Apropiación_Social		5	31001	23	1215	3939	Grupo_de_Investigación_La_Saluda_CILS
80007	Libros_resultados_de_investigación		5	32003	66	6117	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80038	Informe_de_investigación		5	34001	65	10736	3939	Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA
80005	Apropiación_Social		5	34002	65	8496	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_tec
80007	Libros_resultados_de_investigación		5	32003	23	62204	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80026	Informe_de_investigación		5	32002	62	88332	3939	INAMOD_Grupo_de_Investigación_Aplicada_en_Moda_y_Diseño
80019	Libros_resultados_de_investigación		5	32005	65	52444	3939	Materiales_para_el_Mobiliario_MATERMOB
80039	Articulo_Sometido_en_revista_Indexada		5	32006	23	536	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_textil_y_a_la_química_GIAITE

Actualización Dataframe Principal_Productos

Referencia: <https://drive.google.com/drive/folders/1YiB1WYLYRgvG5PVmidIX9ZwrsP517ysQ>

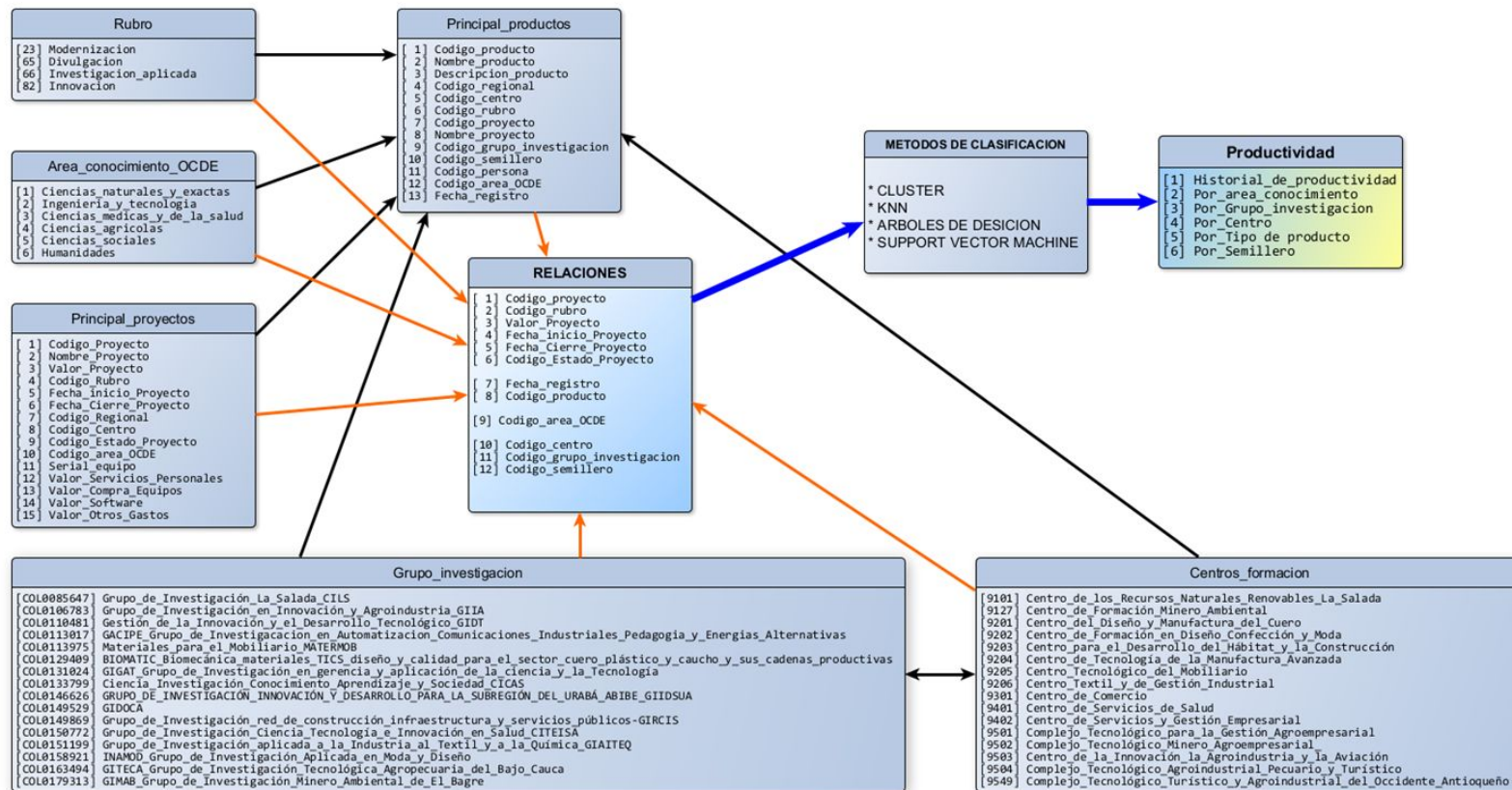
Dataframes identificados en base datos.

Principal_Productos	Principal_Proyectos	Principal_Talento_Humano	Principal_Equipos
Codigo_Producto Nombre_Producto Descripcion_producto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipologia Codigo_Proyecto Nombre_Proyecto Codigo_Grupo_Investigacion Codigo_Semillero Codigo_persona Codigo_area_OCDE Fecha_registro	Codigo_Proyecto Nombre_Proyecto Valor_Proyecto Codigo_Rubro Fecha_inicio_Proyecto Fecha_Cierre_Proyecto Codigo_Regional Codigo_Centro Codigo_Estado_Proyecto Codigo_area_OCDE Serial_equipo Valor_Servicios_Personales Valor_Compra_Equipos Valor_Software Valor_Otros_Gastos	Codigo_Persona Codigo_Nivel_Academico Codigo_Tipo_Contrato Objeto_Contrato Codigo_Rubro Codigo_Area_OCDE Codigo_Estado_Contratacion Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Producto	Serial_equipo Nombre_del_equipo Valor_equipo Fecha_de_compra Codigo_Proyecto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipo_Equipo Codigo_area_OCDE Codigo_persona

Dataframes identificados en base datos.

Principal_Productos	Principal_Proyectos	Principal_Talento_Humano	Principal_Equipos
Codigo_Producto Nombre_Producto Descripcion_producto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipologia Codigo_Proyecto Nombre_Proyecto Codigo_Grupo_Investigacion Codigo_Semillero Codigo_persona Codigo_area_OCDE Fecha_registro	Codigo_Proyecto Nombre_Proyecto Valor_Proyecto Codigo_Rubro Fecha_inicio_Proyecto Fecha_Cierre_Proyecto Codigo_Regional Codigo_Centro Codigo_Estado_Proyecto Codigo_area_OCDE Serial_equipo Valor_Servicios_Personales Valor_Compra_Equipos Valor_Software Valor_Otros_Gastos	Codigo_Persona Codigo_Nivel_Academico Codigo_Tipo_Contrato Objeto_Contrato Codigo_Rubro Codigo_Area_OCDE Codigo_Estado_Contratacion Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Producto	Serial_equipo Nombre_del_equipo Valor_equipo Fecha_de_compra Codigo_Proyecto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipo_Equipo Codigo_area_OCDE Codigo_persona

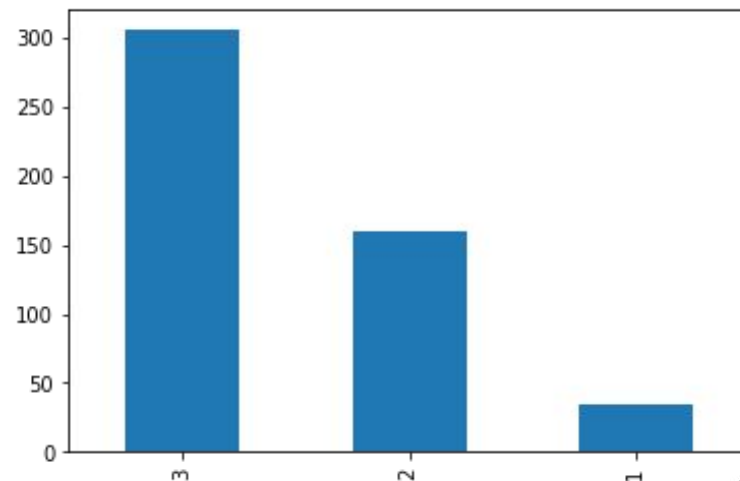
Identificación de base datos modelo relacional



Creación de jerarquías en conceptos de inversión

```
# máximo $309.289.800, Mínimo de proyecto $215.540.900
data4["Categoria_Inversion"] = np.nan
Categoria_Inversion = []
for rows in data4['Valor_Proyecto']:
    if rows > 200000000:
        Categoria_Inversion.append("Alta Inversión")
    elif 50000000 < rows <= 200000000 :
        Categoria_Inversion.append("Media Inversión")
    elif rows <= 50000000:
        Categoria_Inversion.append("Baja Inversión")
    else:
        Categoria_Inversion.append("Sin Información")
print(len(Categoria_Inversion))
data4['Categoria_Inversion'] = Categoria_Inversion
```

500



_Centro	ValorMedio_Proyecto	Cantidad_Proyectos	Valor_Proyecto	latitud	longitud	Año_inicio	Mes_inicio	Año_fin	Mes_fin	Categoria_Inversion
Mobiliario	2.925324e+08	28	417559959	6.18075	-75.60643	2015	4	2019	9	Alta Inversión
Mobiliario	2.925324e+08	28	474990272	6.18075	-75.60643	2014	11	2013	6	Alta Inversión
Mobiliario	2.925324e+08	28	422069442	6.18075	-75.60643	2010	11	2019	7	Alta Inversión
Mobiliario	2.925324e+08	28	389518376	6.18075	-75.60643	2014	7	2013	6	Alta Inversión
Mobiliario	2.925324e+08	28	35859865	6.18075	-75.60643	2015	3	2017	10	Baja Inversión

Medición de indicadores de CTel mediante métodos supervisados por categorización de productos de Minciencias





5. Visualización de datos en mapas - KNN Cluster-PCA

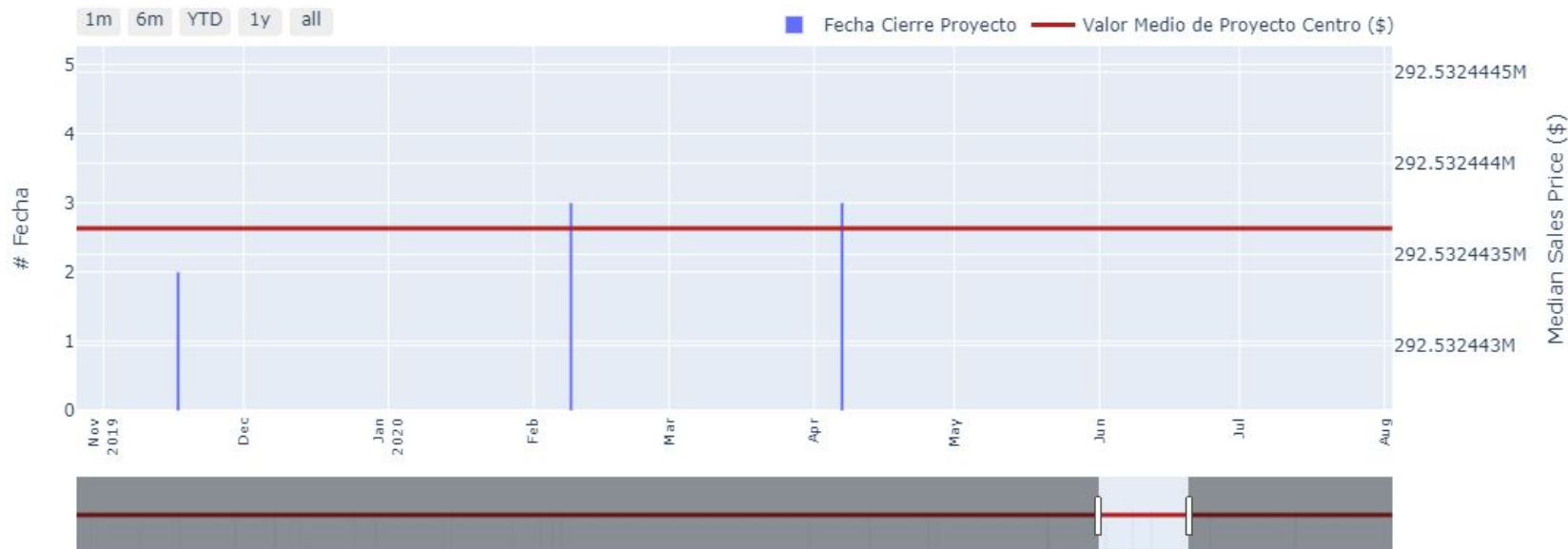


@SENAcomunica

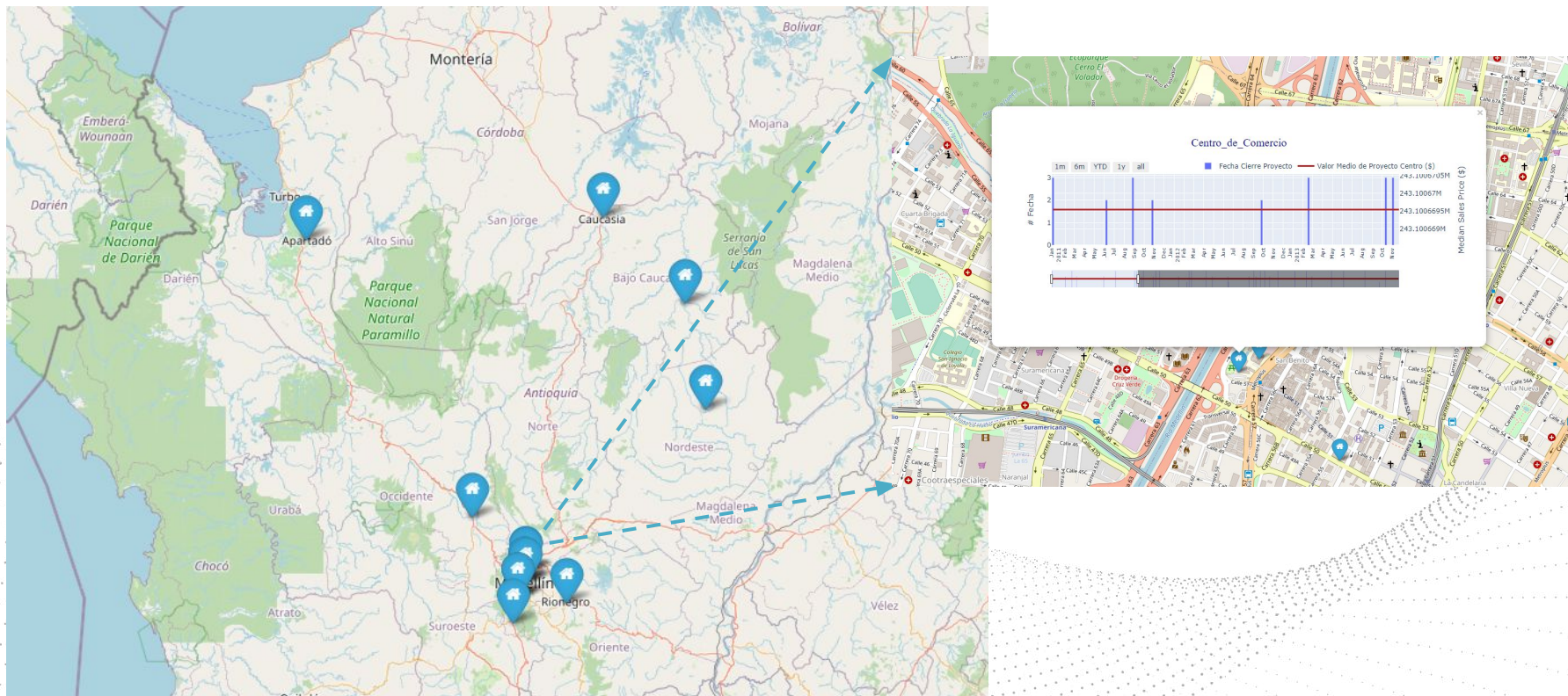
www.sena.edu.co

Serie temporal Costos de inversión - Valor medio

Complejo_Tecnológico_para_la_Gestión_Agroempresarial



Visualización estadística en mapas



Visualización de datos en mapas - KNN Cluster-PCA

1. En primer lugar, echaremos un vistazo rápido a nuestros datos para entender con qué estamos trabajando. Y, luego, ¡lo limpiaremos/filtraremos!
2. Luego, aplicaremos la agrupación de k-Means para identificar centro de formación similares en función de la cantidad de producción de productos y proyectos a lo largo del tiempo para centros de formación en particular.
3. Además, vamos a utilizar el análisis de componentes principales para visualizar datos de alta dimensión, de modo que podamos ver cómo se relacionan nuestros grupos en el espacio original.
4. Finalmente, el resultado final de los centros de formación agrupados se ilustra utilizando la biblioteca Seaborn y Folium Maps.

Limpieza y filtrado de datos

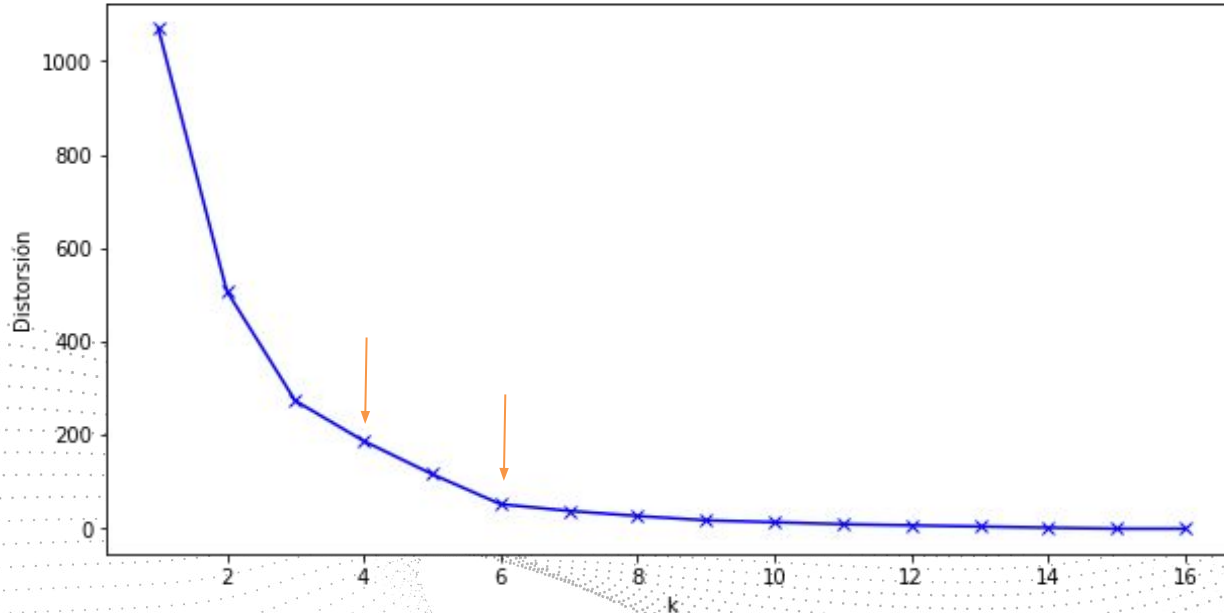
	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365
2	Centro_de_Comercio	37	29	6.254794	-75.575466
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	31	6.300949	-75.568664
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37	33	6.153989	-75.372317
9	Centro_de_los_Recursos_Naturales_Renovables_La...	42	25	6.053157	-75.628486
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	24	6.180794	-75.605884
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	26	43	6.303487	-75.567591
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	33	7.890489	-76.631035
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	35	7.080799	-74.699840
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	34	6.559073	-75.829128
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	26	7.996707	-75.196365

Tabla 1. Limpieza y filtrado de datos de acuerdo a cantidad de proyectos y cantidad de productos totales por centro de formación

Método del codo - óptimo de conglomerados

Para identificar el número óptimo de grupos, necesitamos usar el **método del codo**. Cuando la pendiente de la línea tangente comienza a ser casi horizontal, ese es el número óptimo de conglomerados.

El método del codo



El método del codo
sugiere de **4 o 6 conglomerados**.

	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud	cluster
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430	0
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365	1
2	Centro_de_Comercio	37	29	6.254794	-75.575466	1
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660	3
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935	2
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571	0
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671	2
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	31	6.300949	-75.568664	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37	33	6.153989	-75.372317	1
9	Centro_de_los_Recursos_Naturales_Renovables_La...	42	25	6.053157	-75.628486	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	24	6.180794	-75.605884	0
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	26	43	6.303487	-75.567591	2
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	33	7.890489	-76.631035	2
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	35	7.080799	-74.699840	3
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	34	6.559073	-75.829128	2
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	26	7.996707	-75.196365	1

Tabla 2. Cantidad de proyectos y cantidad de productos totales por centro de formación con cluster.

Aplicamos algunas técnicas de reducción de dimensionalidad para visualizar cómo se relacionan nuestros grupos en el espacio original de alta dimensión. Adicionalmente, podemos ver si las características de nuestros datos están linealmente relacionadas entre sí.

Varianza explicada por cada componente (%):

1 º: 68.72437475021606

2 º: 29.890759114263073

3 º: 0.9507412392611251

4 º: 0.29915093446266766

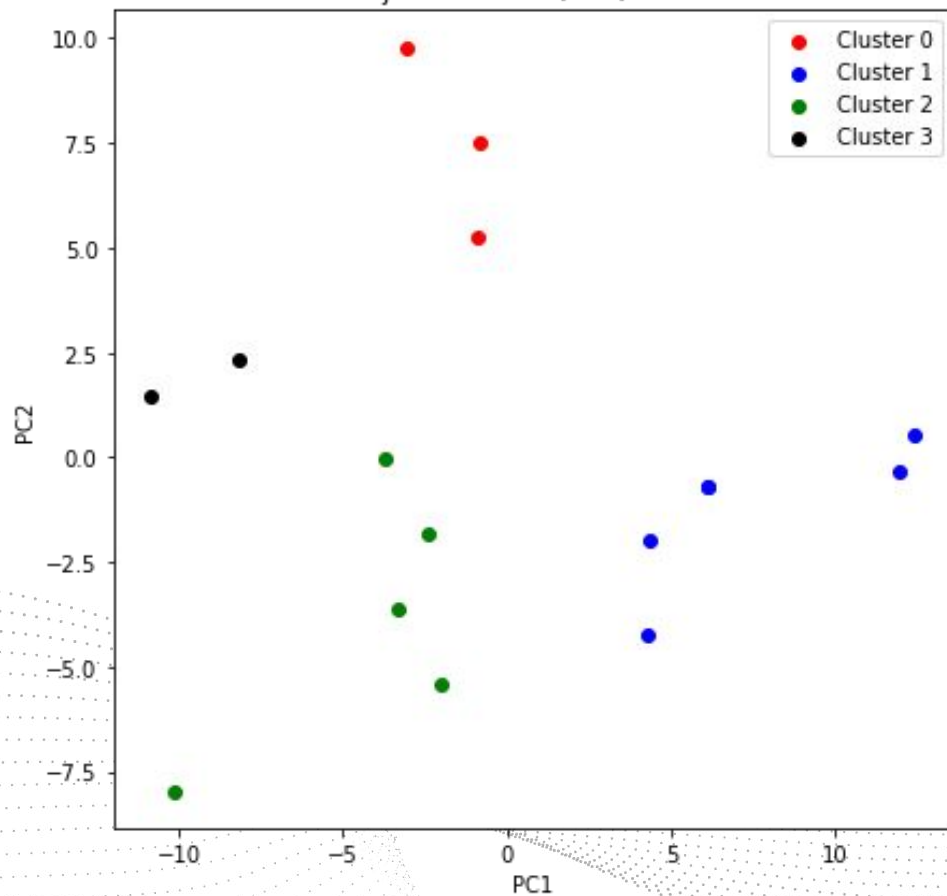
5 º: 0.13497396179707852

Suma total (%): 100.0

Varianza explicada de los dos primeros componentes (%): 68.72437475021606

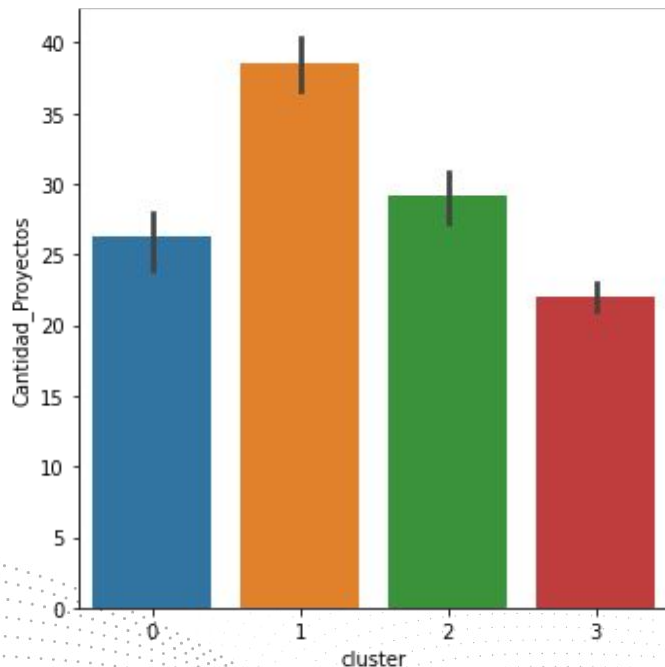
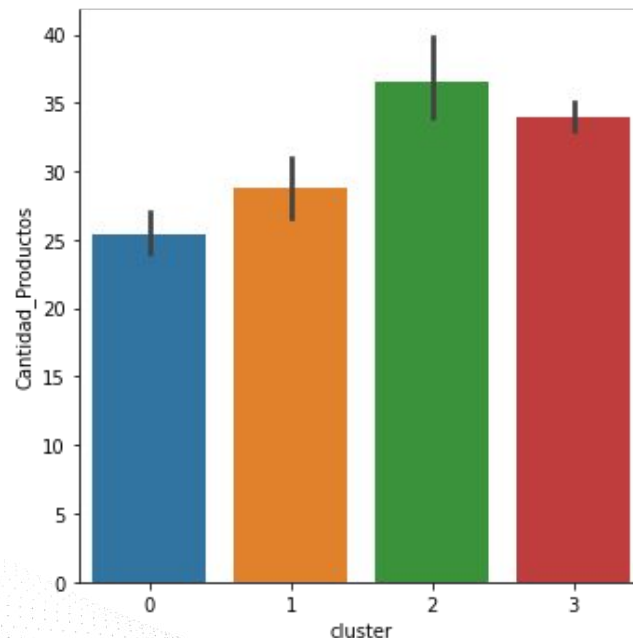
- Podemos observar que los **5 componentes principales (PC)** llevan al **100 %** de la información original que permiten explicar y modelar los datos.
- Podemos ver que utilizando los dos primeros componentes somos capaces de conservar el 68.72% de la información original, por lo tanto, reduce la dimensionalidad de nuestros datos.
- Usamos, por lo tanto, estas PC para visualizar nuestros clústeres.

Visualización de baja dimensión (PCA) - Centros de formación



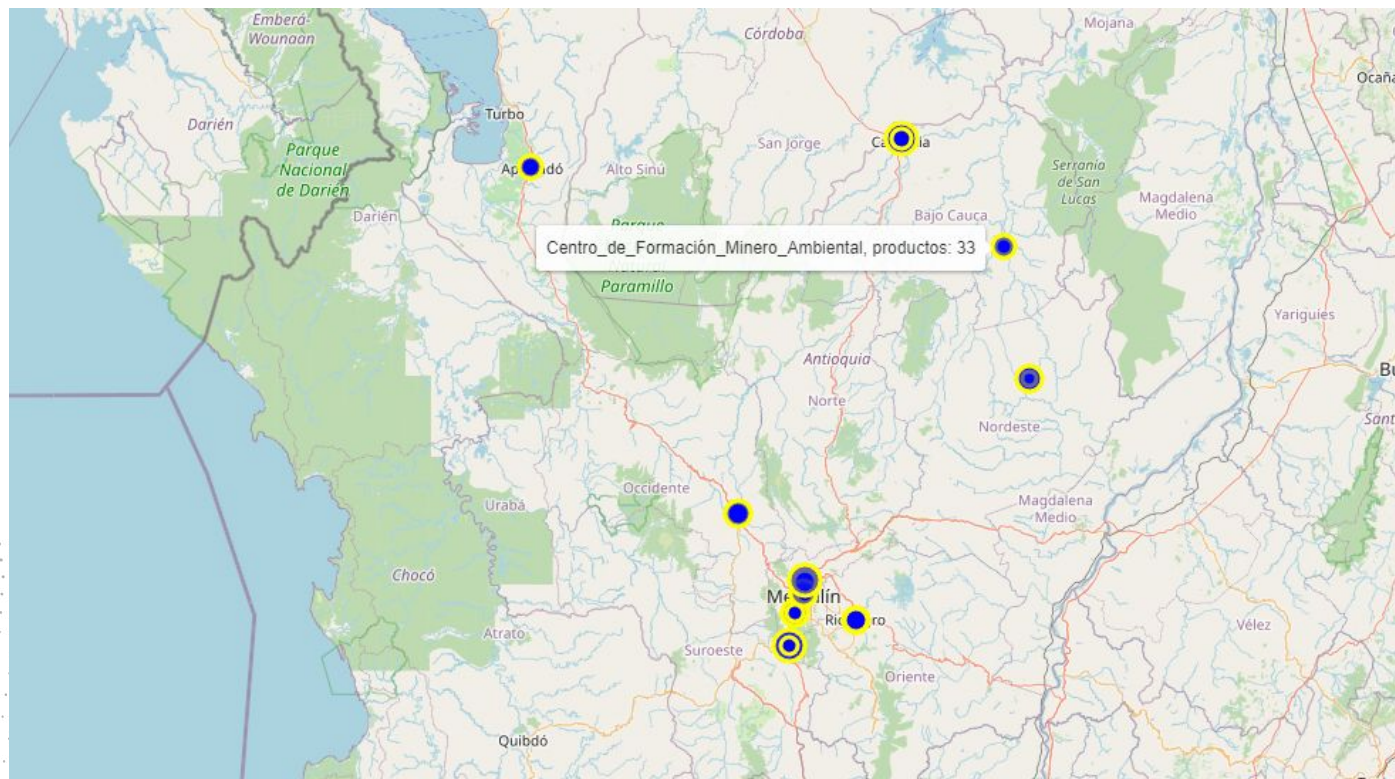
4 grupos parecen ser razonables para identificar muestras similares dentro de nuestros datos.

Clusters por proyectos y productos

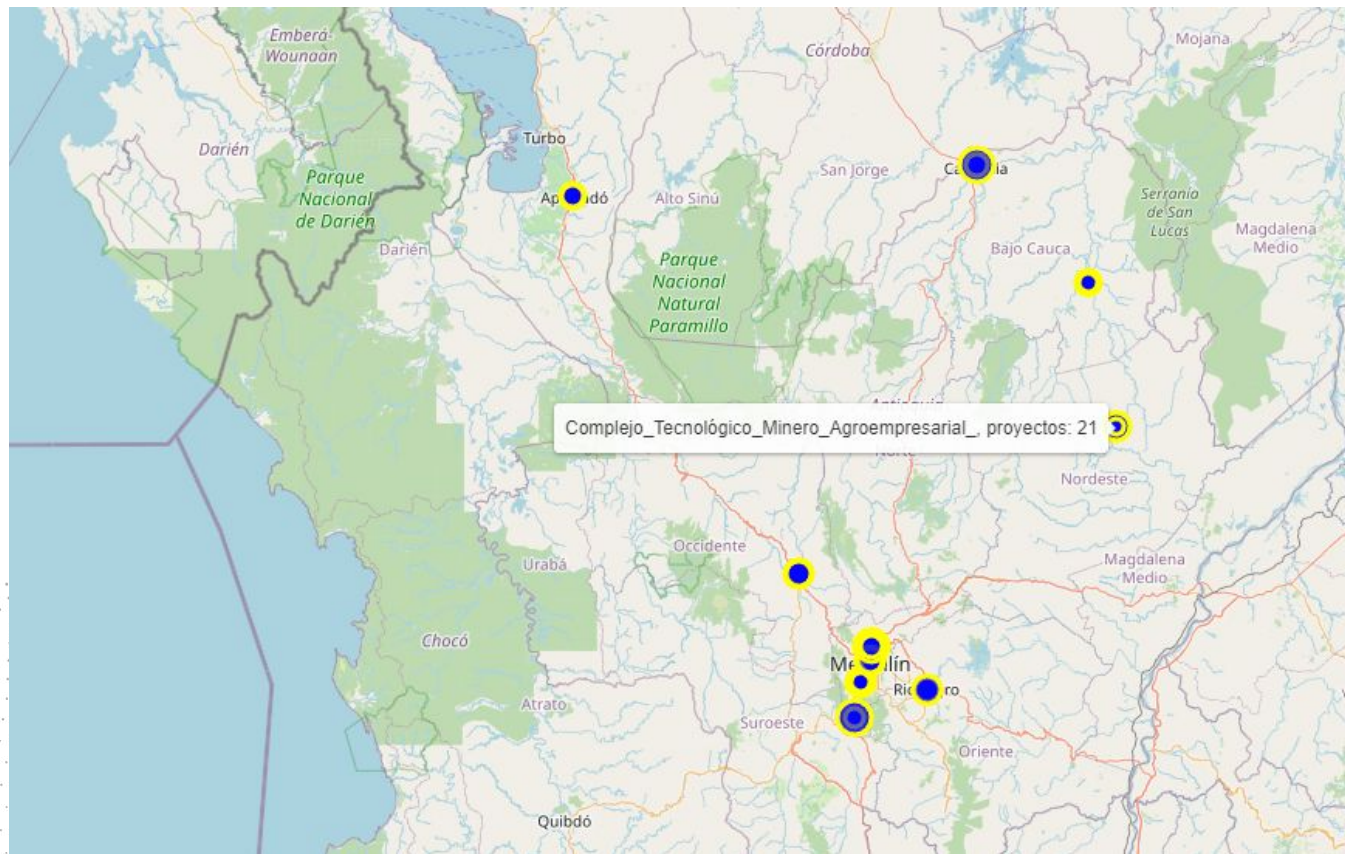


Podemos ver que el **grupo 1** que **mayor cantidad de proyectos** proyectos. Por otro lado, en términos de productos el **grupo 2** es el que **mayor cantidad de productos** presenta.

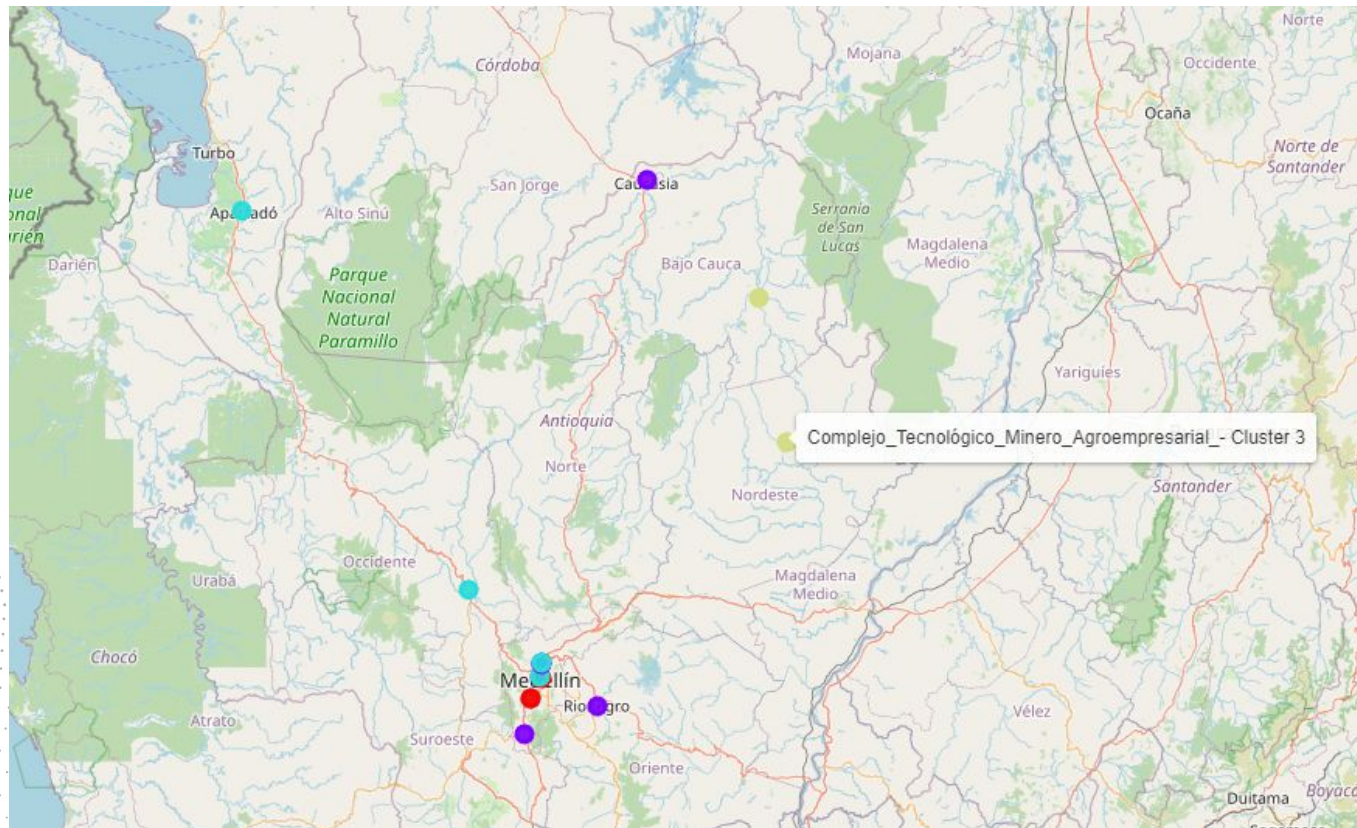
Mapa de cantidad de productos por centro



Mapa de cantidad de proyectos por centro



Mapa de clusters por centro de formación



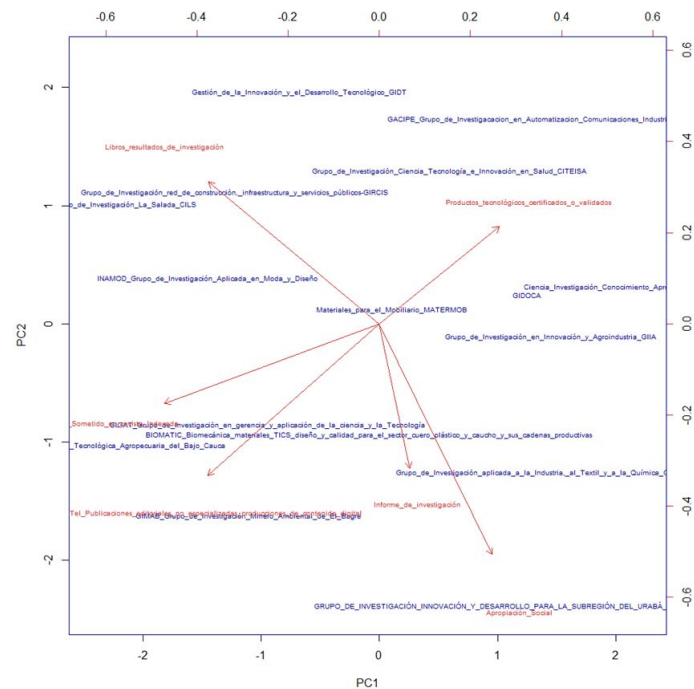


4. Modelos de analítica de datos



@SENAcomunica

www.sena.edu.co



Referencia: <https://drive.google.com/drive/folders/1YiB1WyLYRqvG5PVmidIX9Zwrsp5I7ysQ>

Análisis de componentes principales y Circulo de correlación actualizado

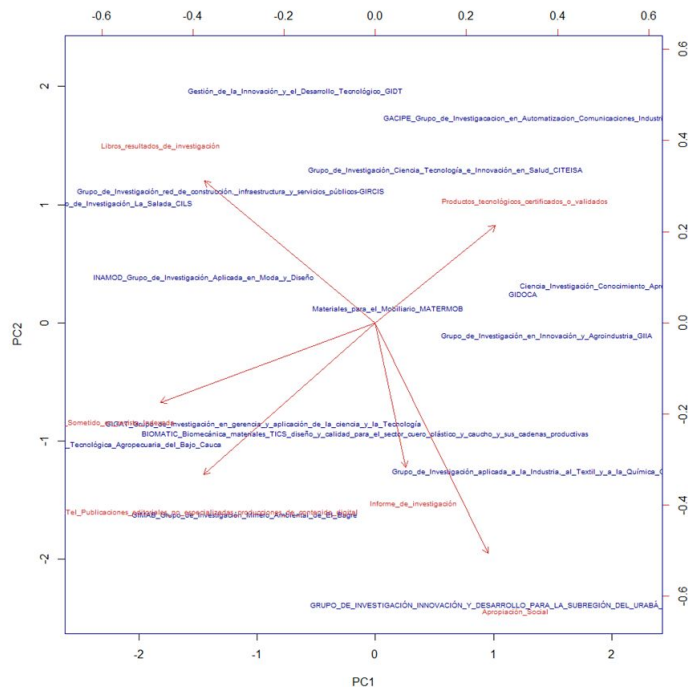
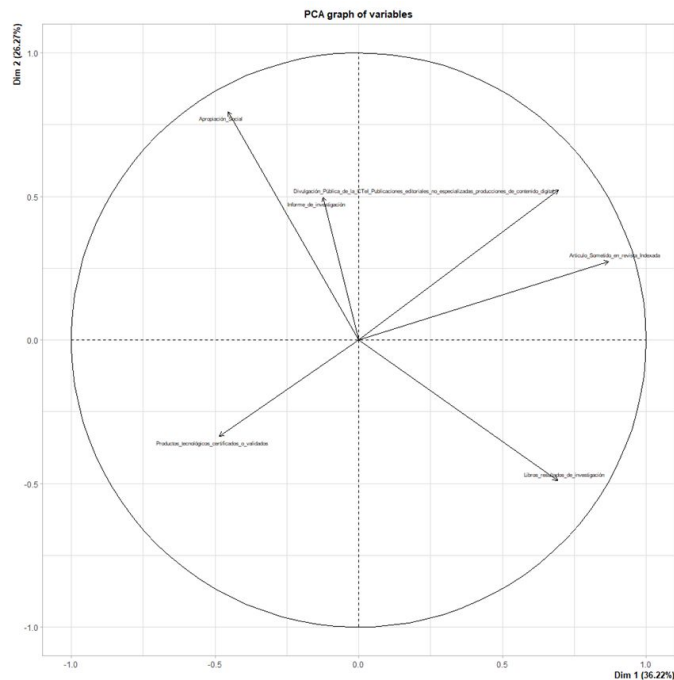


Figura 1: Gráfica de componentes principales y círculo de correlación de variables por productos de CTel

Referencia: <https://drive.google.com/drive/folders/1YiB1WvLYRgvG5PvMidIX9ZwrsP5I7ysQ>

Análisis de componentes principales y Círculo de correlación



	Apropiación Social	Artículo_Sometido_en_revista_Indexada	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital	Informe_de_investigación	Libros_resultados_de_investigación_o_validados	Productos_tecnológicos_certificados
Grupo_de_Investigación_La_Salud_CILS	3	3	4	12	5	7
Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA	2	7	5	8	3	3
Gestión_de_la_Innovación_y_el_Developo_Tecnológico_GIDT	3	1	1	6	5	7
GACIPE_Grupo_de_Investigación_en_Automatización_Comunicaciones_Industriales_Pedagogía_y_Energías_Alternativas	2	3	4	3	9	5
Materiales_para_el_Mobiliario_MATERMOB	7	7	7	4	3	2
BIOMATIC_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero_plástico_y_caucho_y_sus_cadenas	2	7	6	2	8	4
GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tecnología	6	3	6	8	5	3
Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS	8	4	3	4	3	7
GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUBREGIÓN_DEL_URABÁ_ABIBE_GIDS	5	3	3	2	6	6
GIDOCA	8	3	5	3	7	5
Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos-GIRCIS	4	5	8	4	8	7
Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA	5	2	3	3	2	3
Grupo_de_Investigación_aplicada_a_la_Industria_al_Textil_y_a_la_Química_GIAITEQ	13	4	6	7	3	6
INAMOD_Grupo_de_Investigación_Aplicada_en_Moda_y_Diseño	5	8	7	8	8	2
GITECA_Grupo_de_Investigación_Tecnológica_Agropecuaria_del_Bajo_Cauca	6	8	5	6	3	10
GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bagre	5	6	7	7	5	5

Tabla 1: Caracterización de grupos de investigación por productos de CTeI

	Apropiación Social	Artículo_Sometido_en_revista_Indexada	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital	Informe_de_investigación	Libros_resultados_de_investigación	Productos_tecnológicos_certificados_o_validados
Grupo_de_Investigación_La_Salud_CILS	2	7	6	2	8	4
Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA	8	4	3	4	3	7
Gestión_de_la_Innovación_y_el_Developo_Tecnológico_GIDT	2	3	4	3	9	5
GACIPE_Grupo_de_Investigación_en_Automatización_Comunicaciones_Industriales_Pedagogía_y_Energías_Alternativas	3	1	1	6	5	7
Materiales_para_el_Mobiliario_MATERMOB	8	3	5	3	7	5
BIOMATIC_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero_plástico_y_caucho_y_sus_cadenas	6	3	6	8	5	3
GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tecnología	5	6	7	7	5	5
Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS	6	0	5	6	3	10
GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUBREGIÓN_DEL_URABÁ_ABIBE_GIDSUA	13	4	6	7	3	6
GIDOCA	5	2	3	3	2	3
Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos-GIRCIS	2	7	5	8	9	9
Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA	5	3	3	2	6	6
Grupo_de_Investigación_aplicada_a_la_Industria_al_Textil_y_a_la_Química_GIAITEQ	9	3	4	12	5	7
INAMOD_Grupo_de_Investigación_Aplicada_en_Moda_y_Diseño	3	5	8	4	8	7
GITECA_Grupo_de_Investigación_Tecnológica_Agropecuaria_del_Bajo_Cauca	5	8	7	8	8	2
GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bagre	7	7	7	4	3	2

Referencia: <https://drive.google.com/drive/folders/1YiB1WvLYRqG5PVmidIX9Zwrs517ysQ>

Análisis de clustering

Los tres clusters se agrupan según los grupos establecidos en los cuatro cuadrantes de la PCA. Y se distribuyen por productos según los patrones de productos de CTeI descritos en los ítems 1,2,3 y 4.

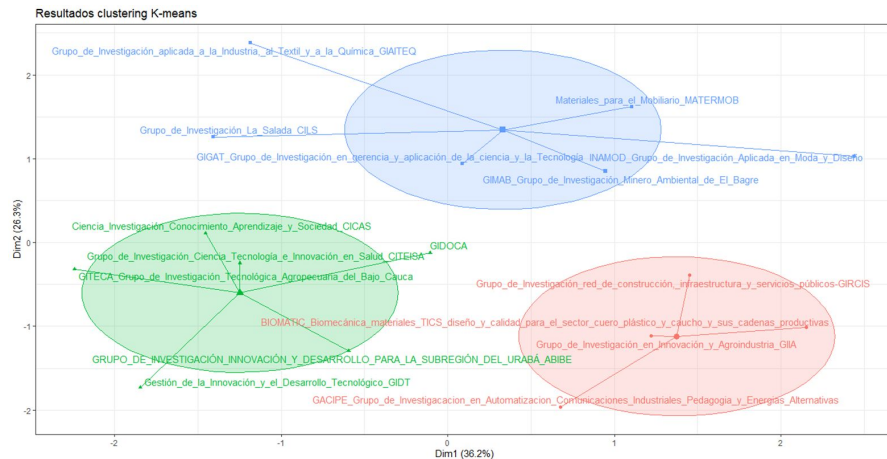


Figura 2: Clustering por grupos de investigación

Referencia: <https://drive.google.com/drive/folders/1YiB1WYLYRqGvG5PVmidIX9ZwrsP5I7ysQ>

```
> km_clusters$size  
[1] 7 3 6
```

Dendograma

Herarchical clustering

Distancia euclídea, Lincage complete, K=3

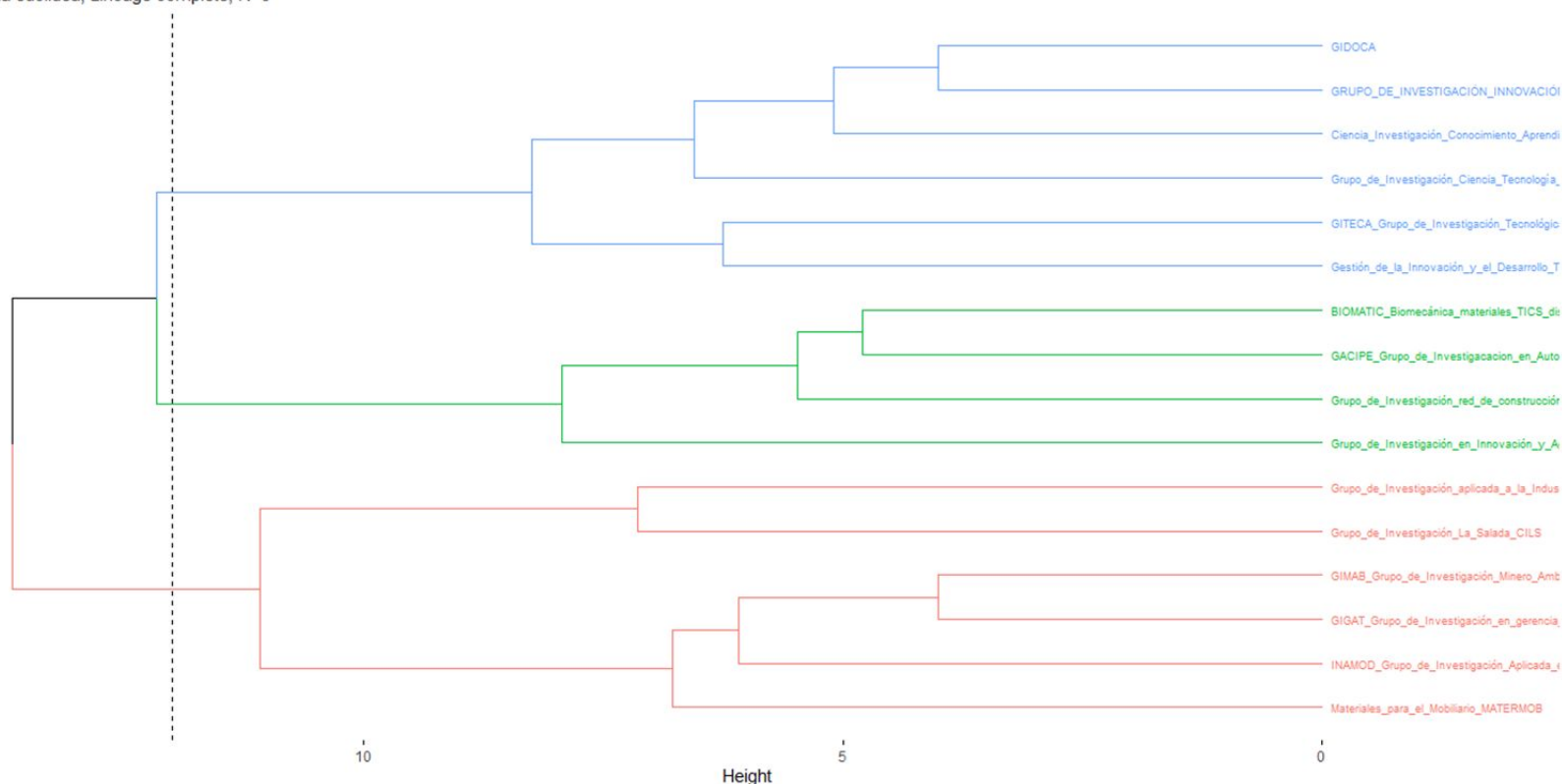


Figura 3: Clustering Jerárquico por grupos de investigación

Referencia: <https://drive.google.com/drive/folders/1YiB1WyLYRgvG5PVmidlX9Zwrsp5l7ysQ>

Mapa de calor de productos de CTel por cluster (Datos sin escalar)



```
> center_df
  cluster Apropiación_Social Artículo_Sometido_en_revista_Indexada
1        1          2.500000                      5.500000
2        2          5.833333                      2.166667
3        3          7.500000                      5.166667
  Divulgación_Pública_de_la_CTEI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital
1                                                                5.750000
2                                                                3.333333
3                                                                6.166667
  Informe_de_investigación Libros_resultados_de_investigación Productos_tecnológicos_certificados_o_validados
1          4.250000                      8.500000                      6.250000
2          4.000000                      4.333333                      6.333333
3          7.666667                      4.833333                      4.166667
```

Tabla 3: Productos de CTel por clustering (Datos sin escalar)

```
> km_clusters.df$size
[1] 4 6 6
>
```

Tabla 4: Tamaño de observaciones por clustering

Mapa de calor de productos de CTel por cluster (Datos sin escalar)

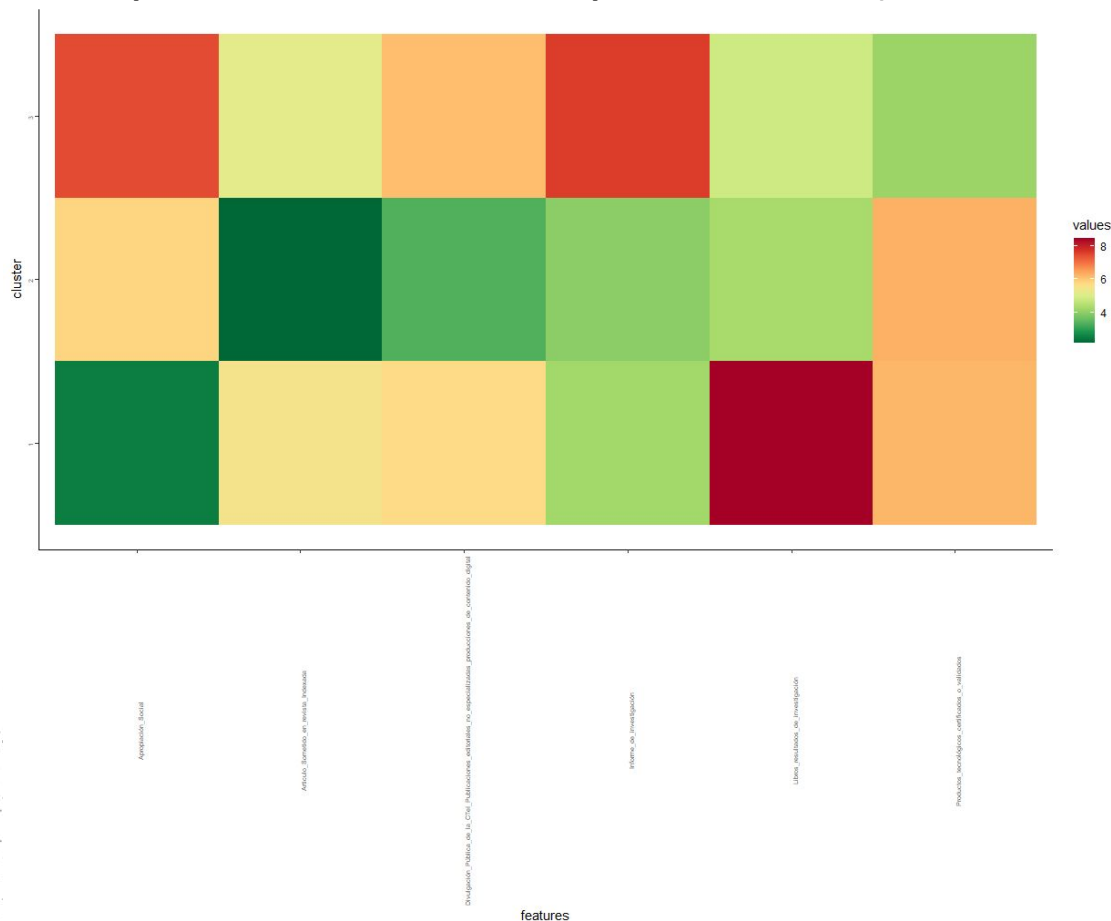


Figura 5: Mapa de calor de los Clustering por productos de CTel, datos sin escalar

Referencia: <https://drive.google.com/drive/folders/1YiB1WylYRqvG5PVmidIX9Zwrsp5l7ysQ>

Mapa de calor de productos de CTel por cluster (Datos escalados)



cluster		features	values
1	1	Apropiación_Social_scal	1.26087116
2	2	Apropiación_Social_scal	-0.40679792
3	3	Apropiación_Social_scal	-0.15583801
4	1	Articulo_Sometido_en_revista_Indexada_scal	-0.76746568
5	2	Articulo_Sometido_en_revista_Indexada_scal	0.86435503
6	3	Articulo_Sometido_en_revista_Indexada_scal	-0.62468137
7	1	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital_scal	0.00000000
8	2	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital_scal	0.82821558
9	3	Divulgación_Pública_de_la_CTeI_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital_scal	-0.96625151
10	1	Productos_tecnológicos_certificados_o_validados_scal	0.92668217
11	2	Productos_tecnológicos_certificados_o_validados_scal	-0.39714950
12	3	Productos_tecnológicos_certificados_o_validados_scal	0.00000000
13	1	Informe_de_investigación_scal	1.04161500
14	2	Informe_de_investigación_scal	0.15094318
15	3	Informe_de_investigación_scal	-0.69690788
16	1	Libros_resultados_de_investigación_scal	-0.80143401
17	2	Libros_resultados_de_investigación_scal	0.42650884
18	3	Libros_resultados_de_investigación_scal	-0.09687664

Tabla 5: Productos de CTel por clustering (Datos escalados)

```
> km_clusters$size
[1] 7 3 6
>
```

Tabla 6: Tamaño de observaciones por clustering

Mapa de calor de productos de CTel por cluster (Datos escalados)

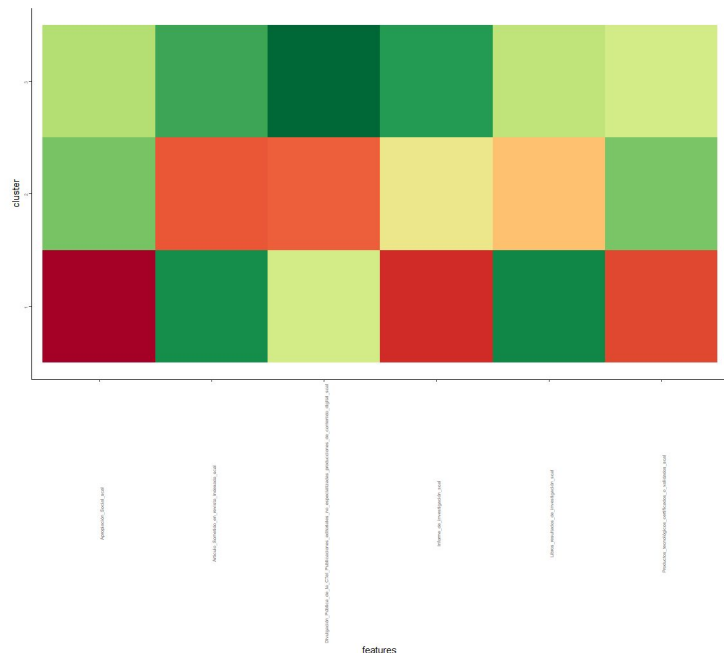
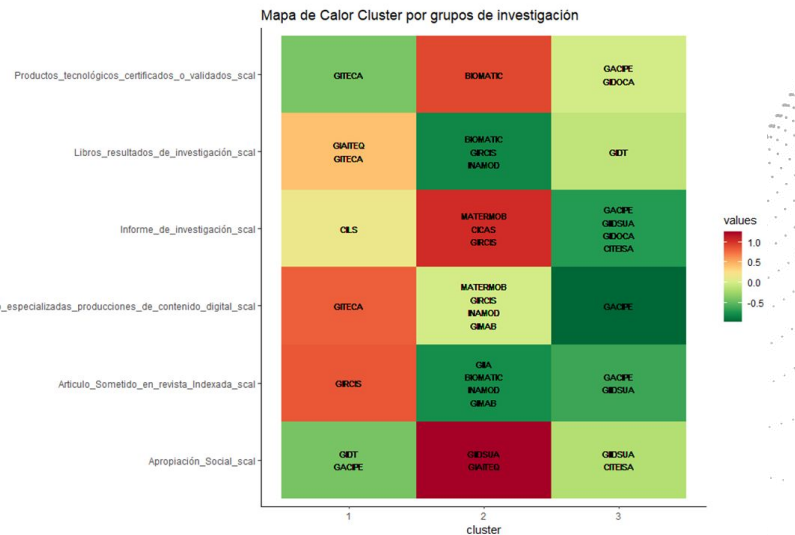
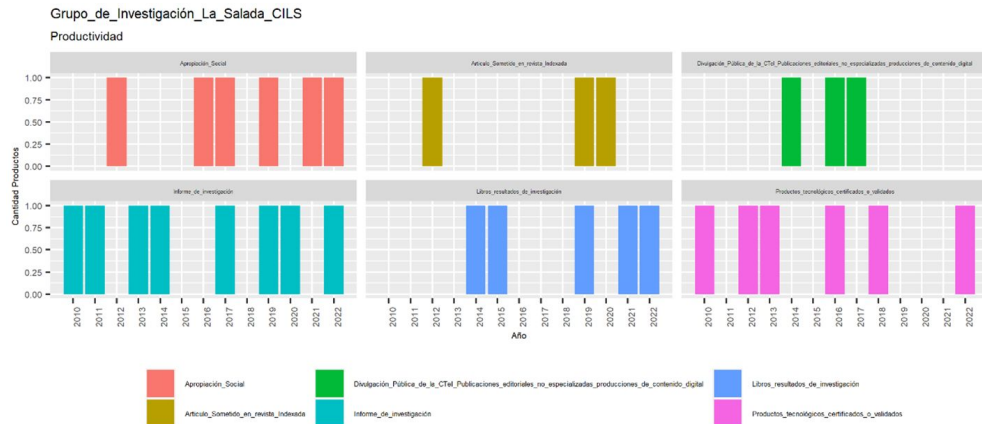
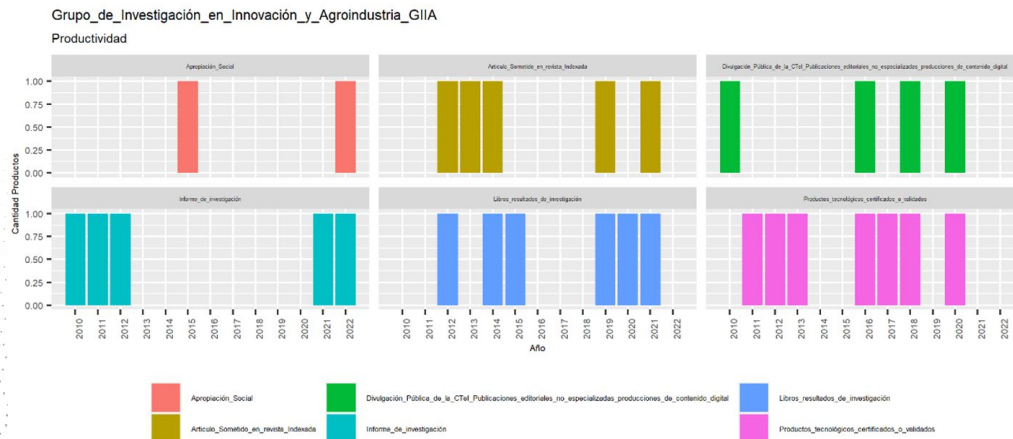


Figura 4: Mapa de calor de los Clustering por productos de CTel, datos escalados.

Referencia: <https://drive.google.com/drive/folders/1YiB1WylYRqvG5PVmidIX9Zwrsp5l7ysQ>

Análisis estadístico descriptivo por grupos de investigación



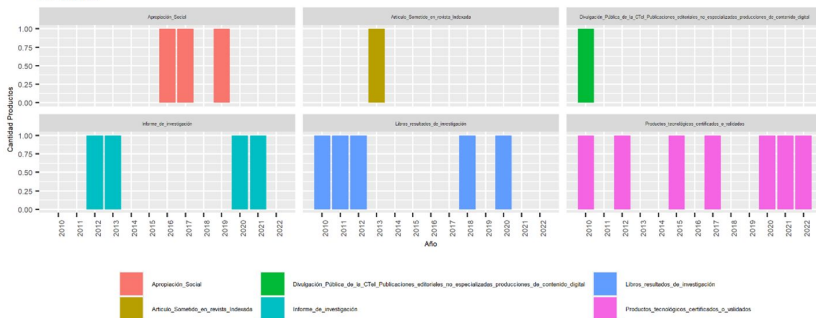
Referencia: <https://drive.google.com/drive/folders/1YiB1WylYRqvG5PVmidlX9Zwrsp5l7ysQ>

Análisis estadístico descriptivo por grupos de investigación



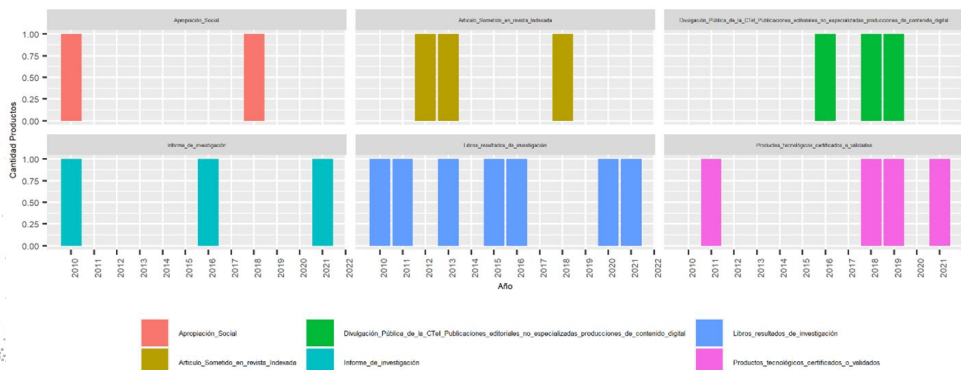
Gestión_de_la_Innovación_y_el_Desarrollo_Tecnológico_GIDT

Productividad



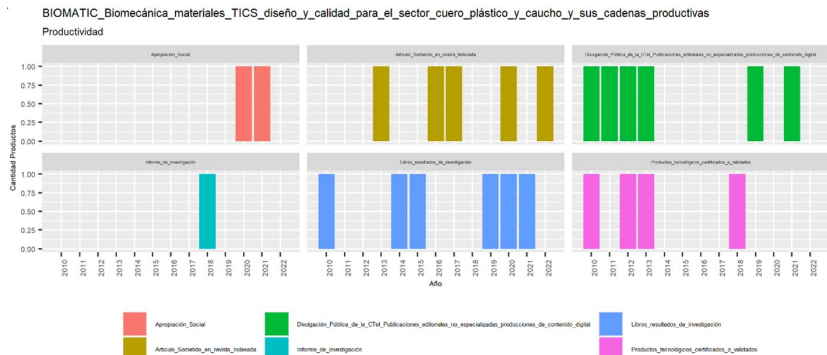
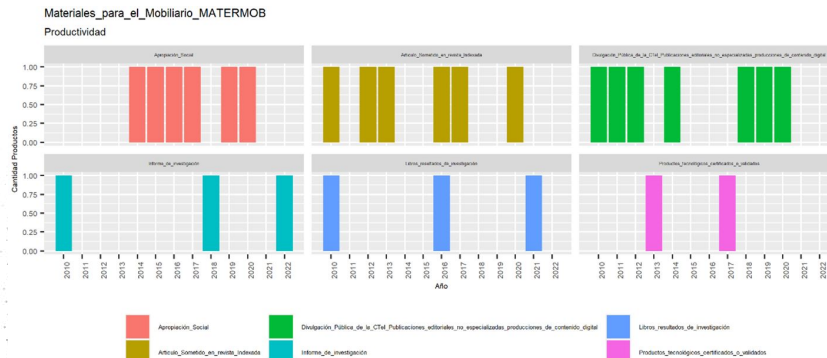
GACIPE_Grupo_de_Investigacion_en_Automatizacion_Comunicaciones_Industriales_Pedagogia_y_Energias_Alternativas

Productividad



Referencia: <https://drive.google.com/drive/folders/1YIB1WylYRqvG5PVmidlX9Zwrsp5l7ysQ>

Análisis estadístico descriptivo por grupos de investigación



Referencia: <https://drive.google.com/drive/folders/1YIB1WylYRqvG5PVmidlX9Zwrsp5l7ysQ>

Implementación modelo Knn



	Libros_resultados_de_investigación	Productos_tecnológicos_certificados_o_validados	Peso relativo	Productividad	Nombre_Grupo_Investigacion
12	5	7	0.8813559	1	Grupo_de_Investigación_La_Salada_CILS
8	9	9	0.6186441	2	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIIA
6	5	7	0.1949153	3	Gestión_de_la_Innovación_y_el_Desarrollo_Tecnológico_GIDT
3	9	5	0.3728814	2	GACIPE_Grupo_de_Investigación_en_Automatización_Com...
4	3	2	0.3305085	2	Materiales_para_el_Mobiliario_MATERMOB
2	8	4	0.2966102	3	BIOMATIC_Biomecánica_materiales_TICS_diseño_y_calidad_p...
8	5	3	0.4745763	2	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_d...
4	3	7	0.4630508	2	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Socieda...
2	6	6	0.3898305	2	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO...
3	7	5	0.7542373	1	GIDOCA

Showing 1 to 10 of 16 entries, 9 total columns

Referencia: <https://drive.google.com/drive/folders/1YiB1WylYRqvG5PVmidIX9ZwrsP5l7ysQ>

Implementación modelo Knn



```
#####  
#install.packages("kknn")  
library(kknn)  
library(class)  
#####  
  
#####  
#Mediante la librería kknn encontramos el k óptimo  
#####  
  
#####  
modelo1 <- train.kknn(Productividad ~ ., data = entrenamiento, kmax = 9)  
modelo1  
names(modelo1)  
view(modelo1$data)  
#####
```

Implementación modelo Knn



Numero óptimo de k-vecinos cercanos: k=7

```
> #-----  
> modelo1 <- train.kknn(Productividad ~ ., data = entrenamiento, kmax = 9)  
> modelo1
```

```
Call:  
train.kknn(formula = Productividad ~ ., data = entrenamiento,      kmax = 9)
```

```
Type of response variable: nominal
```

```
Minimal misclassification: 0.4
```

```
Best kernel: optimal
```

```
Best k: 7
```

```
as.factor(mach12.predictions(Productividad))
```

```
#-----
```

Matriz de predicción Knn



	Productividad	Nombre_Grupo_Investigacion
12	Baja	Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA
16	Media	GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_El_Bagre
7	Media	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tecnología
6	Baja	BIOMATIC_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero_plástico_y_caucho_y_sus_cadenas_productivas
8	Media	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
1	Alta	Grupo_de_Investigación_La_Salada_CILS
11	Media	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos-GIRCIS
4	Media	GACIPE_Grupo_de_Investigacion_en_Automatizacion_Comunicaciones_Industriales_Pedagogia_y_Energias_Alternativas
2	Media	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIIA
13	Alta	Grupo_de_Investigación_aplicada_a_la_Industria_al_Textil_y_a_la_Química_GIAITEQ

Referencia: <https://drive.google.com/drive/folders/1YiB1WylYRqvG5PVmidIX9Zwrsp5I7ysQ>

Clasificación Knn



Clasificación realizada mediante el k-nn

Apropiación Social	Artículo Sometido en revista indexada	Divulgación Pública de la CTeL Publicaciones editoriales no especializadas producciones de contenido digital	Informe de investigación	Libros resultados de investigación	Productos tecnológicos certificados o validados	Peso relativo	Productividad	Nombre Grupo Investigación
5	2	3	3	2	3	0	Bajo	Grupo de Investigación Ciencia Tecnología e Innovación en Salud CITEISA
5	6	7	7	5	5	0,466102	Medio	GIMAB Grupo de Investigación Minero Ambiental de El Bague
6	3	6	8	5	3	0,474576	Medio	GIGAT Grupo de Investigación en gerencia y aplicación de la ciencia y la Tecnología
2	7	6	2	8	4	0,29661	Bajo	BIOMATIC Biomecánica materiales TICS diseño y calidad para el sector cuero plástico y sus cadenas productivas
6	4	3	4	3	7	0,483051	Medio	Ciencia Investigación Conocimiento Aprendizaje y Sociedad CICAS
9	3	4	12	5	7	0,881356	Alto	Grupo de Investigación La Salud CILS
4	5	8	4	8	7	0,652542	Medio	Grupo de Investigación red de construcción infraestructura y servicios públicos-GIRCIS
2	3	4	3	9	5	0,372881	Medio	GACIPE Grupo de Investigación en Automatización Comunicaciones Industriales Pedagogía y Energías Alternativas
2	7	5	8	9	9	0,618644	Medio	Grupo de Investigación en Innovación y Agroindustria GIA
13	4	6	7	3	6	1	Alto	Grupo de Investigación aplicación a la Industria al Textil y a la Química GIAITEQ

Referencia: <https://drive.google.com/drive/folders/1YIB1WylYRqvG5PVmidIX9Zwrs5I7ysQ>

Análisis Bibliográfico de técnicas a implementar



Using the data mining method to assess the innovation.pdf - Adobe Acrobat Reader DC (64-bit)

Archivo Edición Ver Firmar Ventana Ayuda

Inicio Herramientas Using the data mini...

80 (1 de 18) 75,5%

Marcadores

- Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country
 - 1. Introduction
 - 2. Methodology
 - 2.1. Support vector machines-based machine learning
 - 2.2. Assessing attributes for the emerging industry
 - 2.2.1. Ownership of platform technologies
 - 2.2.2. Globalization intention
 - 2.2.3. International knowledge

Technological Forecasting & Social Change 119 (2017) 80–97

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country

Dejing Kong ^{a,b}, Yuan Zhou ^{b,*}, Yufei Liu ^d, Lan Xue ^b

^a School of Modern Post, Beijing University of Posts and Telecommunications, Beijing, China
^b School of Public Policy and Management, Tsinghua University, Beijing, China
^c The CIC Center for Strategic Studies, Chinese Academy of Engineering, Beijing, China
^d College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 20 September 2016
Accepted 20 February 2017
Available online 16 March 2017

Keywords:
Innovation gap assessment
Support vector machines-based classifier
High-quality patents
Industry robot
China

ABSTRACT

It is critical for “catching-up” countries to narrow innovation gaps with developed countries by developing emerging industries. This research introduces a data mining-based method to systematically assess the national innovation gap that is specifically for emerging industries. The method examines the five key attributes of emerging industries, including the ownership of platform technologies, globalization intention, international knowledge position, university-industry linkage, and cross-disciplinary technology development. In particular, this method combines data mining with experts’ knowledge to build patent training examples, and then uses a support vector machine-based classifier to single out all high-quality patents for each innovation attribute. Based on the selected high-quality patents, the authors utilize a factorial design analysis to systematically evaluate the innovation gap between countries. This method can significantly reduce measurement bias of traditional single patent indicators. In addition, it also can robustly adjust measuring weights in response to the specific of each innovation attribute, while traditional multi-attribute evaluation methods cannot. As a result, this research empirically shows that China’s industrial robot sector has apparent innovation gaps compared to developed economies, specifically in university-industry linkage, cross-disciplinary competence, and globalization intention, and this calls for the attention of policy makers and industrial experts.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Innovation is necessary for “catching-up” countries (Fan, 2005).

1976). Further, an effective diffusion of innovation is vital for the economic development of many countries’ operating periods under different social and economic systems. (Guan et al., 2005). Every country is a

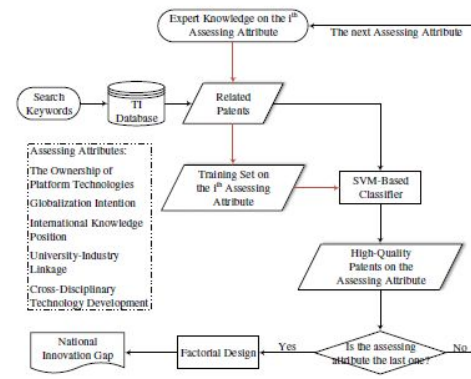


Fig. 1. The framework for innovation gap assessment.

Referencia: <https://drive.google.com/drive/folders/1YIB1WvLYRQvG5PVmidIX9ZwrspsI7ysQ>

Conclusiones

La implementación de la gestión de los recursos humanos en las empresas de servicios de salud en Colombia, requiere de un enfoque integral que considere tanto la dimensión técnica como la humana. Es fundamental que las organizaciones cuenten con líderes comprometidos y capacitados, así como con personal de salud altamente calificado y motivado. La implementación exitosa de estas estrategias dependerá de la capacidad de las organizaciones para adaptarse a los cambios constantes en el sector de la salud y para fomentar una cultura de mejora continua y trabajo en equipo.

https://docs.google.com/presentation/d/1neieCwpeMABexVmtblgNuEMuE_9a_uNCk/edit#slide=id.p1

Curso complementario Formulación de Proyectos Hub de Innovación





GRACIAS

Línea de atención al ciudadano: 018000 910270
Línea de atención al empresario: 018000 910682



@SENAcomunica

www.sena.edu.co



El empleo
es de todos

Mintrabajo

Proyecto SGPS 9748

Prototipo Innovación

Presentación de avances 08

Modelo de evaluación y valoración de la Gestión de CTel para los Centros SENA

Regional Antioquia aplicando analítica de datos

10 de mayo del 2022



@SENAcomunica

www.sena.edu.co



Contenido

1. Etapas Modelo CRISP - DM - Proyecto SGPS-9748
2. Cronograma 2022 - Proyecto SGPS-9748
3. Actualización y exploración de base de datos
4. Visualización de datos en mapas -KNN Cluster-PCA
5. Modelos de analítica de datos
 1. Machine learning no supervisado
 2. Machine learning supervisado
6. Actividades de divulgación



@SENAcomunica

www.sena.edu.co

1. Etapas modelo CRISP - DM - Proyecto SGPS-9748

1. ~~Comprensión del negocio~~
2. ~~Comprensión de los datos~~
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Implementación
7. Modelo de analítica de datos
8. Validación del modelo



2. Cronograma 2022 - Proyecto SGPS-9748



	Fechas programadas	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Fase 1 : Avance de la metodología del Big Data	3 marzo	100%						
Fase 2 : Identificación y selección del Dataset análogo al trabajo en particular	15 abril		100 %					
Fase 3: Caracterizar y seleccionar modelos analítica	19 mayo			48%				
Fase 4: Aplicación de API's para vincular con el desarrollo web	20 junio							
Fase 5 : Puesta en marcha del desarrollo web con los modelos de segmentación y clusterización	21 julio							
Entrega final	15 agosto							



3. Actualización y exploración de base de datos



@SENAcomunica

www.sena.edu.co

Dataframes - Actualización



idigo_Prodect	Nombre_Producto	Descripcio_producto	Codigo_Regional	Codigo_Centro	Codigo_Resbo	Codigo_Proyecto	Nombre_Proyecto	Nombre_Grupo_Investigacion
80033	Articulo_Sometido_en_revista_Indexada		5	32001	62	34602	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80023	Productos_tecnológicos_certificados_o_validados		5	32001	65	44465	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80019	Articulo_Sometido_en_revista_Indexada		5	35003	62	112005	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80016	Libros_resultados_de_investigación		5	32003	65	104041	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80020	Apropiación_Social		5	34001	65	11700	3939	Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA
80013	Apropiación_Social		5	35002	66	43702	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80031	Productos_tecnológicos_certificados_o_validados		5	35004	23	120075	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80016	Libros_resultados_de_investigación		5	35433	66	10102	3939	GIDOCA
80012	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	31001	66	36447	3939	Grupo_de_Investigación_La_Saluda_CILS
80003	Informe_de_investigación		5	32006	65	53448	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_Textil_y_a_la_Química_GIAITE
80023	Libros_resultados_de_investigación		5	35002	62	607	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80016	Libros_resultados_de_investigación		5	31001	23	25344	3939	Grupo_de_Investigación_La_Saluda_CILS
80035	Informe_de_investigación		5	31017	62	33356	3939	GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bogre
80017	Apropiación_Social		5	34002	23	122551	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tec
80003	Apropiación_Social		5	32006	23	38222	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_Textil_y_a_la_Química_GIAITE
80001	Productos_tecnológicos_certificados_o_validados		5	31001	62	55444	3939	Grupo_de_Investigación_La_Saluda_CILS
80001	Libros_resultados_de_investigación		5	32001	23	30338	3939	BIOMATICO_Biomecánica_materiales_TICS_diseño_y_calidad_para_el_sector_cuero
80021	Libros_resultados_de_investigación		5	35004	65	103921	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80025	Articulo_Sometido_en_revista_Indexada		5	35003	65	35441	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80020	Productos_tecnológicos_certificados_o_validados		5	32006	62	74744	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_Textil_y_a_la_Química_GIAITE
80015	Articulo_Sometido_en_revista_Indexada		5	32005	23	27663	3939	Materiales_para_el_Mobiliario_MATERMOB
80019	Informe_de_investigación		5	35003	62	110708	3939	Grupo_de_Investigación_en_Innovación_y_Agroindustria_GIA
80040	Productos_tecnológicos_certificados_o_validados		5	35004	62	16511	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80039	Informe_de_investigación		5	32006	62	36556	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_Textil_y_a_la_Química_GIAITE
80013	Informe_de_investigación		5	31001	65	736	3939	Grupo_de_Investigación_La_Saluda_CILS
80026	Productos_tecnológicos_certificados_o_validados		5	35002	65	1636	3939	Ciencia_Investigación_Conocimiento_Aprendizaje_y_Sociedad_CICAS
80010	Apropiación_Social		5	32003	62	110010	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80025	Productos_tecnológicos_certificados_o_validados		5	32003	62	118005	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80020	Informe_de_investigación		5	34002	65	30002	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tec
80019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	32003	23	38335	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80025	Libros_resultados_de_investigación		5	31017	65	30057	3939	GIMAB_Grupo_de_Investigación_Minero_Ambiental_de_EL_Bogre
80019	Libros_resultados_de_investigación		5	32003	23	12330	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80031	Apropiación_Social		5	35004	23	13966	3939	GRUPO_DE_INVESTIGACIÓN_INNOVACIÓN_Y_DESARROLLO_PARA_LA_SUEB
80036	Libros_resultados_de_investigación		5	31001	23	84466	3939	Grupo_de_Investigación_La_Saluda_CILS
80028	Apropiación_Social		5	31001	23	1215	3939	Grupo_de_Investigación_La_Saluda_CILS
80007	Libros_resultados_de_investigación		5	32003	66	6117	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80038	Informe_de_investigación		5	34001	65	10736	3939	Grupo_de_Investigación_Ciencia_Tecnología_e_Innovación_en_Salud_CITEISA
80005	Apropiación_Social		5	34002	65	84466	3939	GIGAT_Grupo_de_Investigación_en_gerencia_y_aplicación_de_la_ciencia_y_la_Tec
80007	Libros_resultados_de_investigación		5	32003	23	62204	3939	Grupo_de_Investigación_red_de_construcción_infraestructura_y_servicios_públicos
80026	Informe_de_investigación		5	32002	62	88332	3939	INAMOD_Grupo_de_Investigación_Aplicada_en_Moda_y_Diseño
80019	Libros_resultados_de_investigación		5	32005	65	52444	3939	Materiales_para_el_Mobiliario_MATERMOB
80039	Articulo_Sometido_en_revista_Indexada		5	32006	23	536	3939	Grupo_de_Investigación_aplicada_a_la_industria_del_Textil_y_a_la_Química_GIAITE

Actualización Dataframe Principal_Productos

Referencia: <https://drive.google.com/drive/folders/1YiB1WYLYRgvG5PVmidIX9ZwrsP517ysQ>

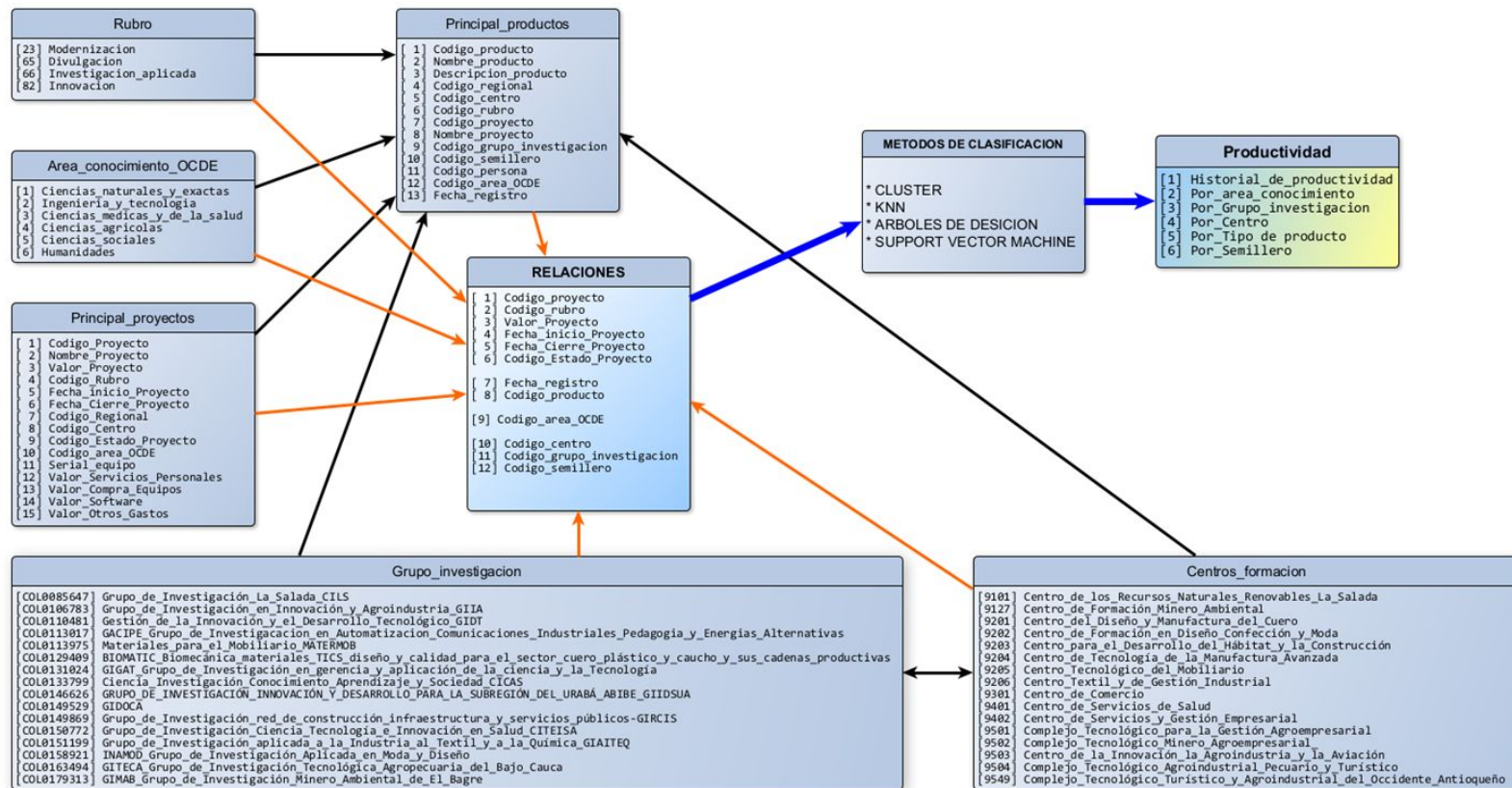
Dataframes identificados en base datos.

Principal_Productos	Principal_Proyectos	Principal_Talento_Humano	Principal_Equipos
Codigo_Producto Nombre_Producto Descripcion_producto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipologia Codigo_Proyecto Nombre_Proyecto Codigo_Grupo_Investigacion Codigo_Semillero Codigo_persona Codigo_area_OCDE Fecha_registro	Codigo_Proyecto Nombre_Proyecto Valor_Proyecto Codigo_Rubro Fecha_inicio_Proyecto Fecha_Cierre_Proyecto Codigo_Regional Codigo_Centro Codigo_Estado_Proyecto Codigo_area_OCDE Serial_equipo Valor_Servicios_Personales Valor_Compra_Equipos Valor_Software Valor_Otros_Gastos	Codigo_Persona Codigo_Nivel_Academico Codigo_Tipo_Contrato Objeto_Contrato Codigo_Rubro Codigo_Area_OCDE Codigo_Estado_Contratacion Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Producto	Serial_equipo Nombre_del_equipo Valor_equipo Fecha_de_compra Codigo_Proyecto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipo_Equipo Codigo_area_OCDE Codigo_persona

Dataframes identificados en base datos.

Principal_Productos	Principal_Proyectos	Principal_Talento_Humano	Principal_Equipos
Codigo_Producto Nombre_Producto Descripcion_producto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipologia Codigo_Proyecto Nombre_Proyecto Codigo_Grupo_Investigacion Codigo_Semillero Codigo_persona Codigo_area_OCDE Fecha_registro	Codigo_Proyecto Nombre_Proyecto Valor_Proyecto Codigo_Rubro Fecha_inicio_Proyecto Fecha_Cierre_Proyecto Codigo_Regional Codigo_Centro Codigo_Estado_Proyecto Codigo_area_OCDE Serial_equipo Valor_Servicios_Personales Valor_Compra_Equipos Valor_Software Valor_Otros_Gastos	Codigo_Persona Codigo_Nivel_Academico Codigo_Tipo_Contrato Objeto_Contrato Codigo_Rubro Codigo_Area_OCDE Codigo_Estado_Contratacion Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Producto	Serial_equipo Nombre_del_equipo Valor_equipo Fecha_de_compra Codigo_Proyecto Codigo_Regional Codigo_Centro Codigo_Rubro Codigo_Tipo_Equipo Codigo_area_OCDE Codigo_persona

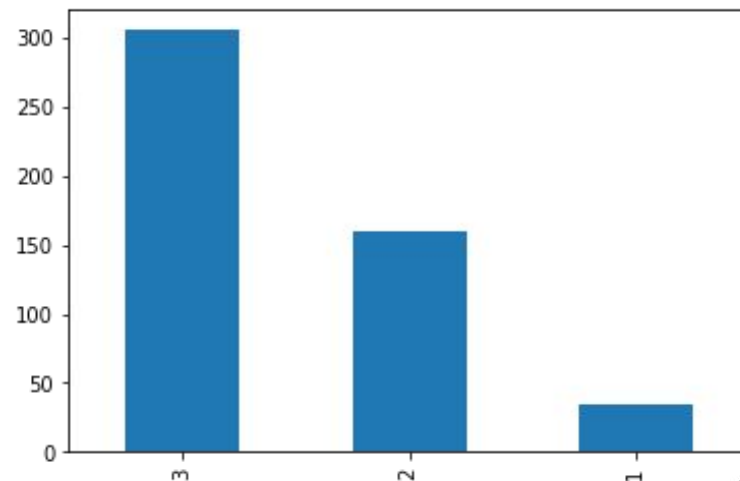
Identificación de base datos modelo relacional



Creación de jerarquías en conceptos de inversión

```
# máximo $309.289.800, Mínimo de proyecto $215.540.900
data4["Categoria_Inversion"] = np.nan
Categoria_Inversion = []
for rows in data4['Valor_Proyecto']:
    if rows > 200000000:
        Categoria_Inversion.append("Alta Inversión")
    elif 50000000 < rows <= 200000000 :
        Categoria_Inversion.append("Media Inversión")
    elif rows <= 50000000:
        Categoria_Inversion.append("Baja Inversión")
    else:
        Categoria_Inversion.append("Sin Información")
print(len(Categoria_Inversion))
data4['Categoria_Inversion'] = Categoria_Inversion
```

500



_Centro	ValorMedio_Proyecto	Cantidad_Proyectos	Valor_Proyecto	latitud	longitud	Año_inicio	Mes_inicio	Año_fin	Mes_fin	Categoria_Inversion
Mobiliario	2.925324e+08	28	417559959	6.18075	-75.60643	2015	4	2019	9	Alta Inversión
Mobiliario	2.925324e+08	28	474990272	6.18075	-75.60643	2014	11	2013	6	Alta Inversión
Mobiliario	2.925324e+08	28	422069442	6.18075	-75.60643	2010	11	2019	7	Alta Inversión
Mobiliario	2.925324e+08	28	389518376	6.18075	-75.60643	2014	7	2013	6	Alta Inversión
Mobiliario	2.925324e+08	28	35859865	6.18075	-75.60643	2015	3	2017	10	Baja Inversión

Medición de indicadores de CTel mediante métodos supervisados por categorización de productos de Minciencias





5. Visualización de datos en mapas - KNN Cluster-PCA



@SENAcomunica

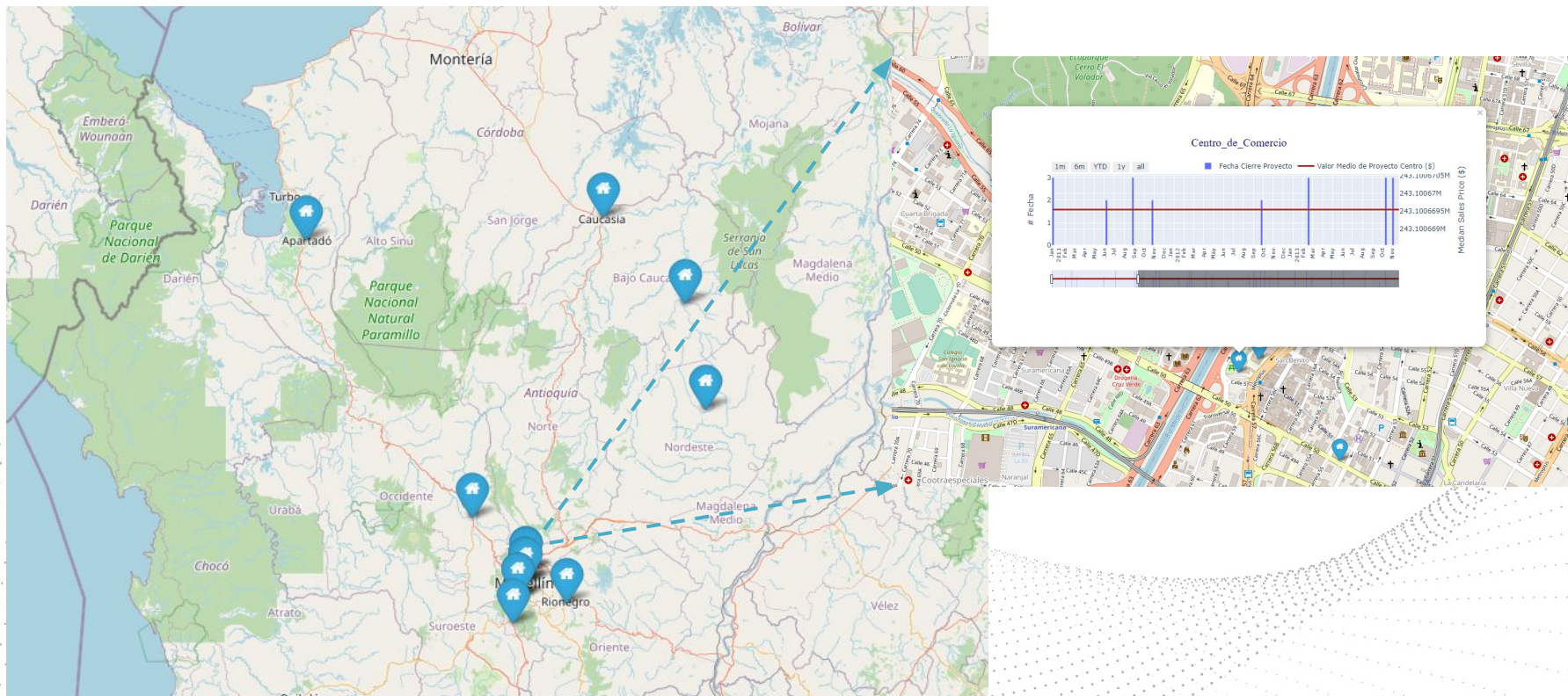
www.sena.edu.co

Serie temporal Costos de inversión - Valor medio

Complejo_Tecnológico_para_la_Gestión_Agroempresarial



Visualización estadística en mapas



Visualización de datos en mapas - KNN Cluster-PCA

1. En primer lugar, echaremos un vistazo rápido a nuestros datos para entender con qué estamos trabajando. Y, luego, ¡lo limpiaremos/filtraremos!
2. Luego, aplicaremos la agrupación de k-Means para identificar centro de formación similares en función de la cantidad de producción de productos y proyectos a lo largo del tiempo para centros de formación en particular.
3. Además, vamos a utilizar el análisis de componentes principales para visualizar datos de alta dimensión, de modo que podamos ver cómo se relacionan nuestros grupos en el espacio original.
4. Finalmente, el resultado final de los centros de formación agrupados se ilustra utilizando la biblioteca Seaborn y Folium Maps.

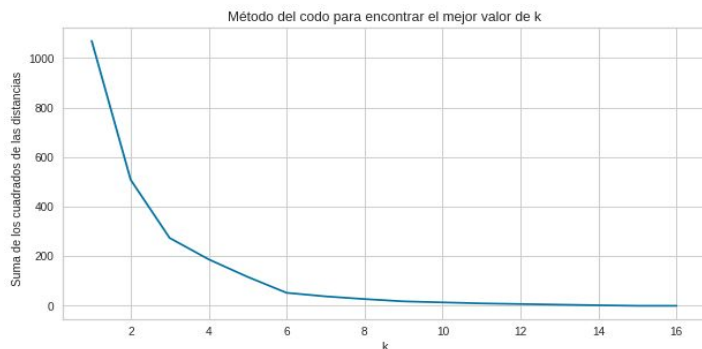
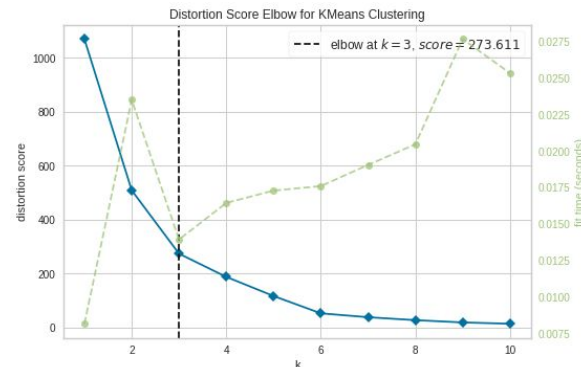
Limpieza y filtrado de datos

	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365
2	Centro_de_Comercio	37	29	6.254794	-75.575466
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	31	6.300949	-75.568664
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37	33	6.153989	-75.372317
9	Centro_de_los_Recursos_Naturales_Renovables_La...	42	25	6.053157	-75.628486
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	24	6.180794	-75.605884
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	26	43	6.303487	-75.567591
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	33	7.890489	-76.631035
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	35	7.080799	-74.699840
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	34	6.559073	-75.829128
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	26	7.996707	-75.196365

Tabla 1. Limpieza y filtrado de datos de acuerdo a cantidad de proyectos y cantidad de productos totales por centro de formación

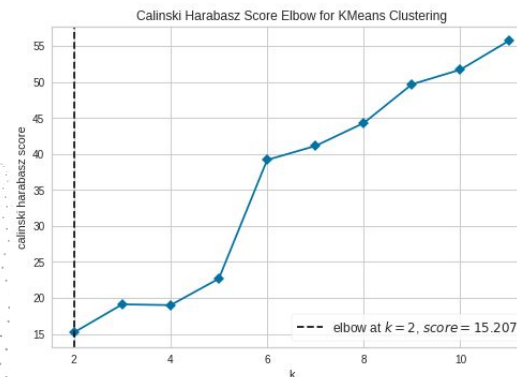
Método del codo - óptimo de conglomerados

Para identificar el número óptimo de grupos, necesitamos usar el **método del codo**. Cuando la pendiente de la línea tangente comienza a ser casi horizontal, ese es el número óptimo de conglomerados.



Calculando el punto de inflexión por medio de la librería KElbowVisualizer observamos que:

- k es óptimo con **3 conglomerados** desde la métrica Métrica *distortion*, que calcula la suma de las distancias al cuadrado desde cada punto hasta su centro asignado.
- k es óptimo con **2 conglomerados** desde la métrica *calinski_harabasz* calcula la relación de dispersión entre y dentro de los grupos.



Métrica *distortion* $k = 3$



	Nombre_Centro	Cantidad_Proyectos	Cantidad_Productos	latitud	longitud	cluster
0	Centro_Tecnológico_del_Mobiliario	28	27	6.180750	-75.606430	2
1	Centro_Textil_y_de_Gestión_Industrial	37	29	6.302387	-75.567365	1
2	Centro_de_Comercio	37	29	6.254794	-75.575466	1
3	Centro_de_Formación_Minero_Ambiental	23	33	7.584361	-74.800660	0
4	Centro_de_Formación_en_Diseño_Confección_y_Moda	32	37	6.181273	-75.605935	0
5	Centro_de_Servicios_de_Salud	27	25	6.251389	-75.571571	2
6	Centro_de_Servicios_y_Gestión_Empresarial	30	36	6.255308	-75.574671	0
7	Centro_de_Tecnología_de_la_Manufactura_Avanzada	36	31	6.300949	-75.568664	1
8	Centro_de_la_Innovación_la_Agroindustria_y_la_...	37	33	6.153989	-75.372317	1
9	Centro_de_los_Recursos_Naturales_Renovables_La...	42	25	6.053157	-75.628486	1
10	Centro_del_Diseño_y_Manufactura_del_Cuero	24	24	6.180794	-75.605884	2
11	Centro_para_el_Desarrollo_del_Hábitat_y_la_Con...	26	43	6.303487	-75.567591	0
12	Complejo_Tecnológico_Agroindustrial_Pecuario_y...	28	33	7.890489	-76.631035	0
13	Complejo_Tecnológico_Minero_Agroempresarial_	21	35	7.080799	-74.699840	0
14	Complejo_Tecnológico_Turístico_y_Agroindustria...	30	34	6.559073	-75.829128	0
15	Complejo_Tecnológico_para_la_Gestión_Agroempre...	42	26	7.996707	-75.196365	1

Tabla 2. Cantidad de proyectos y cantidad de productos totales por centro de formación con cluster.

Índices de validación externos del agrupamiento con KMeans

Homogeneidad: Este procedimiento intenta identificar grupos relativamente homogéneos en términos de la distribución de puntos

Complejidad: se calcula la distancia entre todos los posibles pares formados por una observación del cluster A y una del cluster B

Índices de validación externos del agrupamiento con KMeans

Medida V: El V de Cramer es una medida de la fuerza de asociación entre dos variables nominales
Va de 0 a 1 donde:

0 indica que no hay asociación entre las dos variables.
1 indica una fuerte asociación entre las dos variables.

Índice Fowlkes-Mallows: El índice de Fowlkes-Mallows es un método de evaluación externo que se utiliza para determinar la similitud entre dos agrupaciones, y también una métrica para medir matrices de confusión

Índices de validación externos del agrupamiento con KMeans

Información mutua ajustada: Información Mutua Ajustada (AMI) es un ajuste de la puntuación de Información Mutua (MI) para tener en cuenta el azar. Explica el hecho de que el IM es generalmente más alto para dos agrupaciones con un mayor número de grupos, independientemente de si hay realmente más información compartida.

Calcular índices de validación externos del agrupamiento con KMeans

Homogeneidad: 0.205

Completitud: 1.000

Medida V: 0.341

Índice Rand corregido: 0.078

Información mutua ajustada: 0.335

Índice Fowlkes-Mallows: 0.317

Aplicamos algunas técnicas de reducción de dimensionalidad para visualizar cómo se relacionan nuestros grupos en el espacio original de alta dimensión. Adicionalmente, podemos ver si las características de nuestros datos están linealmente relacionadas entre sí.

Varianza explicada por cada componente (%):

1 º: 68.72437475021606

2 º: 29.890759114263073

3 º: 0.9507412392611251

4 º: 0.29915093446266766

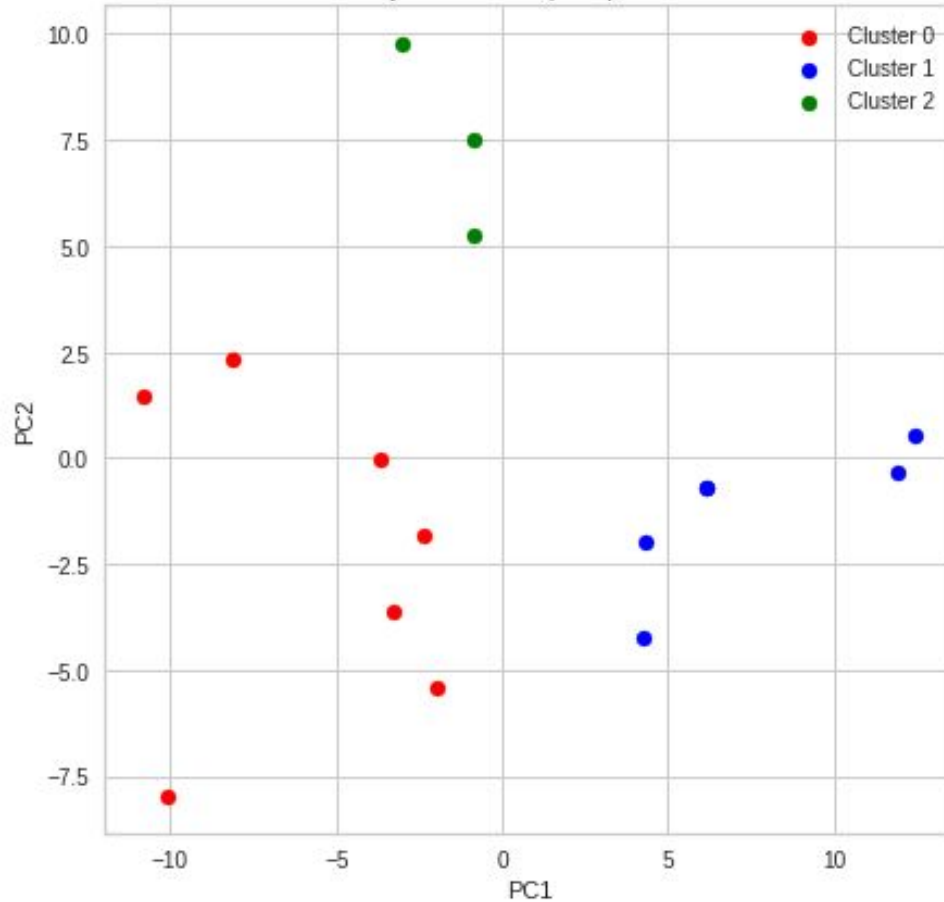
5 º: 0.13497396179707852

Suma total (%): 100.0

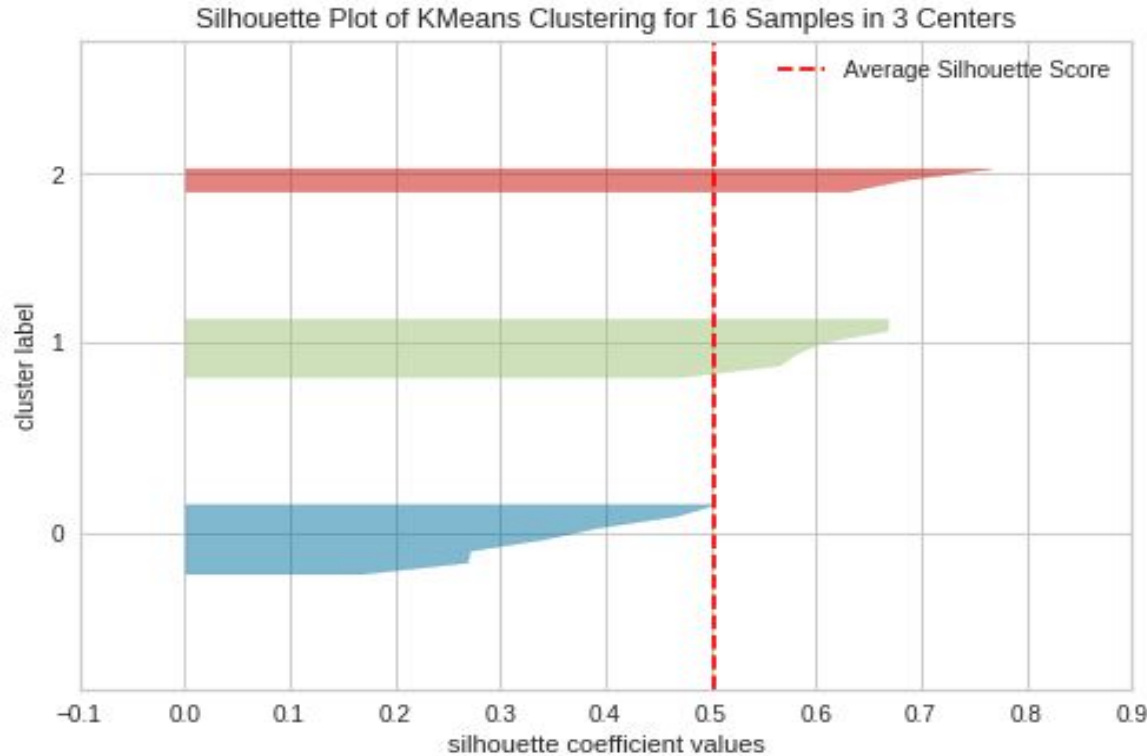
Varianza explicada de los dos primeros componentes (%): 68.72437475021606

- Podemos observar que los **5 componentes principales (PC)** llevan al **100 %** de la información original que permiten explicar y modelar los datos.
- Podemos ver que utilizando los dos primeros componentes somos capaces de conservar el 68.72% de la información original, por lo tanto, reduce la dimensionalidad de nuestros datos.
- Usamos, por lo tanto, estas PC para visualizar nuestros clústeres.

Visualización de baja dimensión (PCA) - Centros de formación



3 grupos parecen ser razonables para identificar muestras similares dentro de nuestros datos.



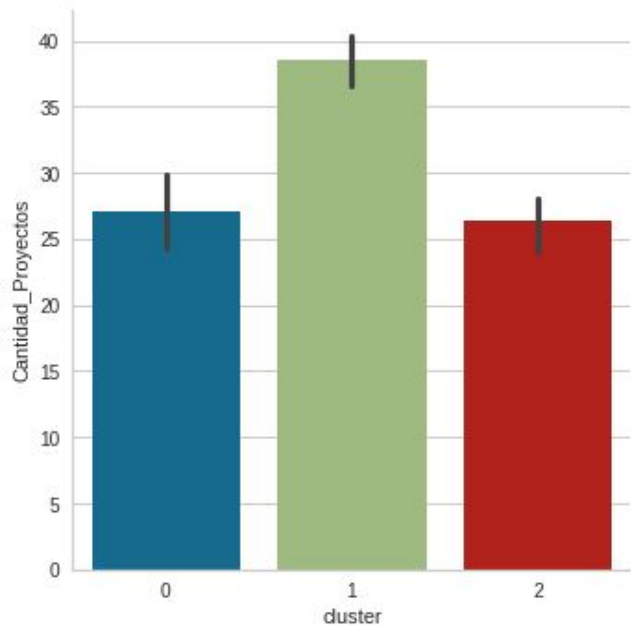
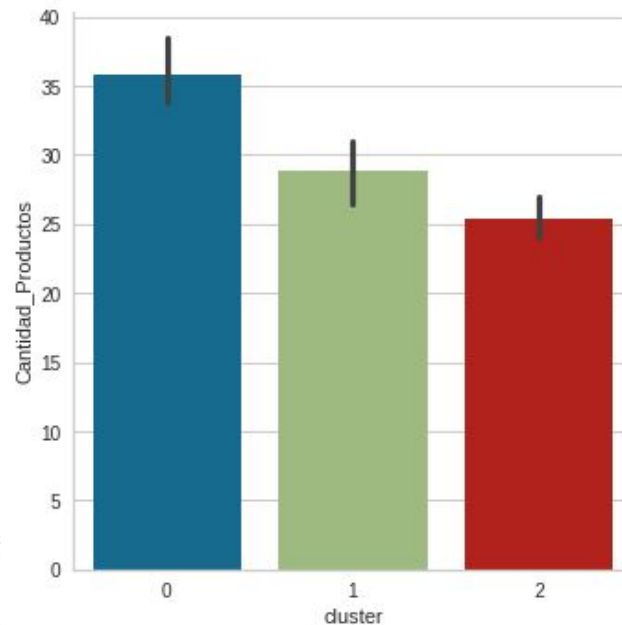
El coeficiente de silueta se utiliza cuando se desconoce la verdad sobre el terreno del conjunto de datos y **calcula la densidad de los clústeres** calculados por el modelo.

La puntuación se calcula promediando el coeficiente de silueta para cada muestra, calculada como la diferencia entre la distancia promedio dentro del grupo y la distancia media del grupo más cercano para cada muestra, normalizada por el valor máximo. Esto produce una puntuación entre 1 y -1, donde 1 son clústeres muy densos y -1 es un agrupamiento completamente incorrecto.

Calinski_harabasz: 146.905
Davies Bouldin: 1.237

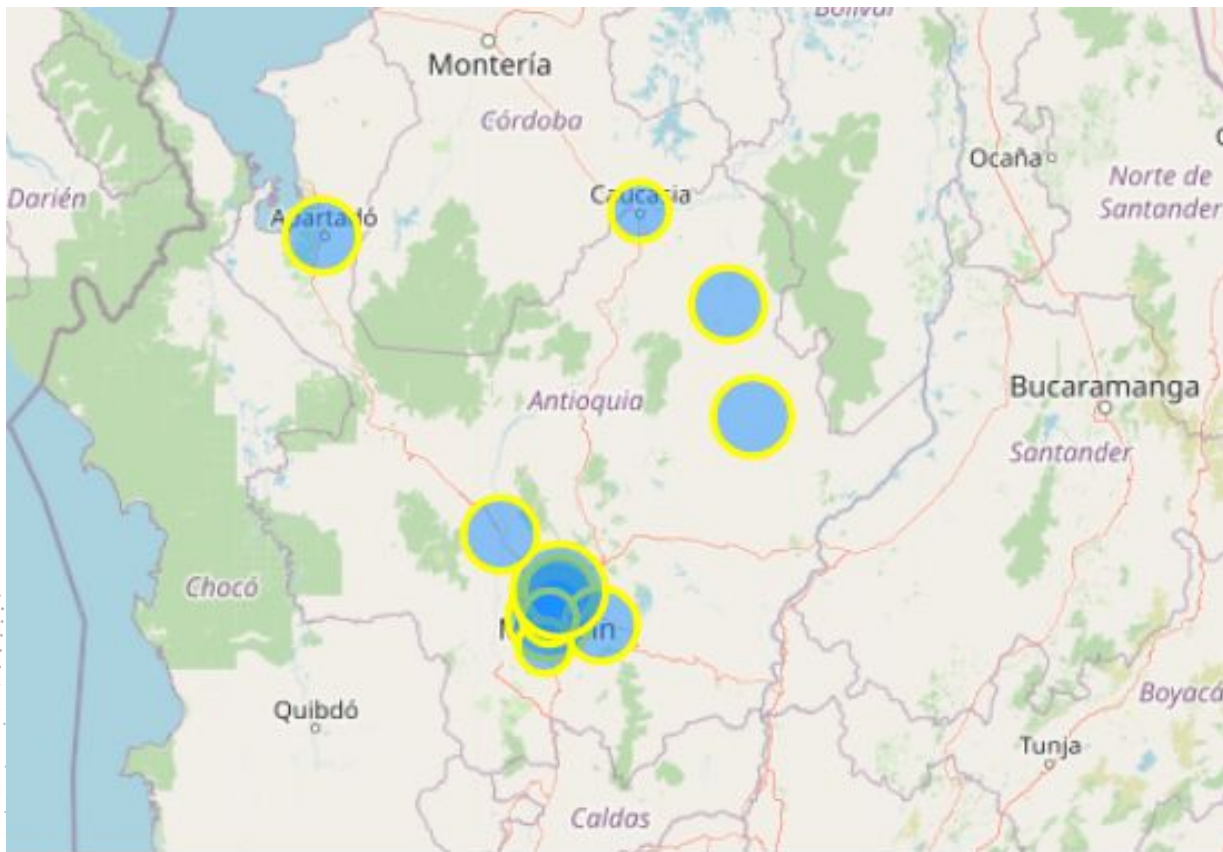
Davies Bouldin: Este es un esquema de evaluación interno, donde la validación de qué tan bien se ha realizado el agrupamiento se realiza utilizando cantidades y características inherentes al conjunto de datos.

Clusters por proyectos y productos

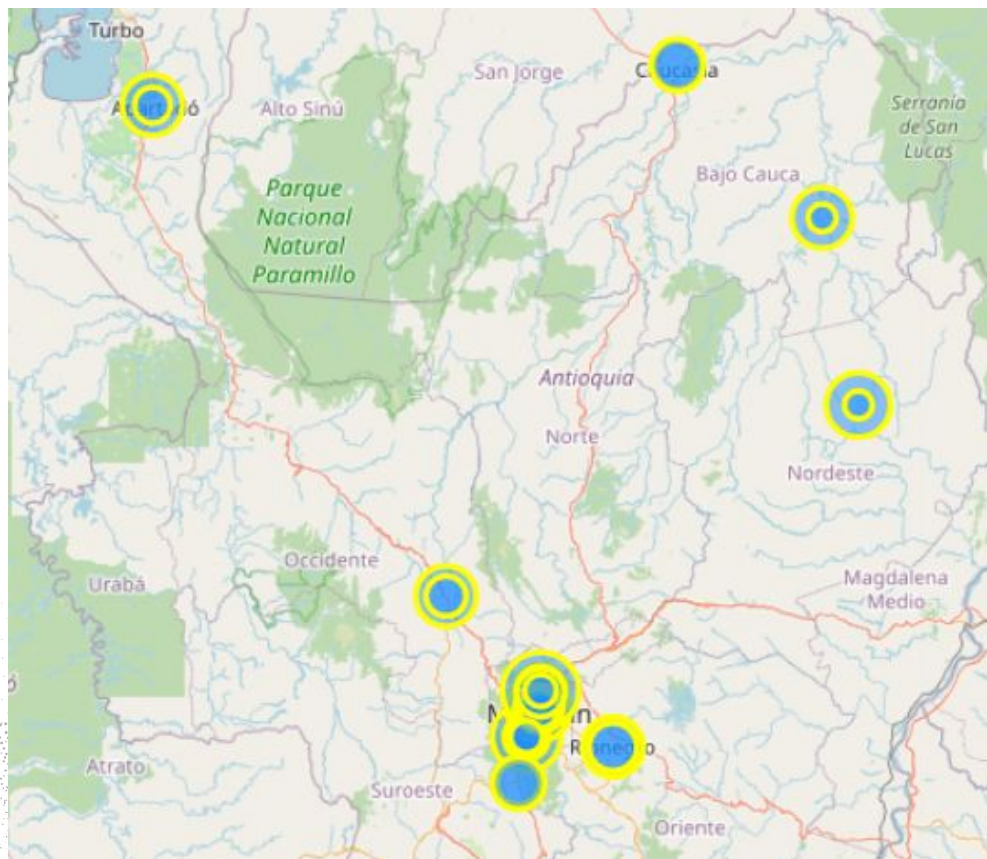


Podemos ver que el **cluster 1** tiene **mayor cantidad de proyectos**. Por otro lado, en términos de productos el **cluster 0** es el que **mayor cantidad de productos** presenta.

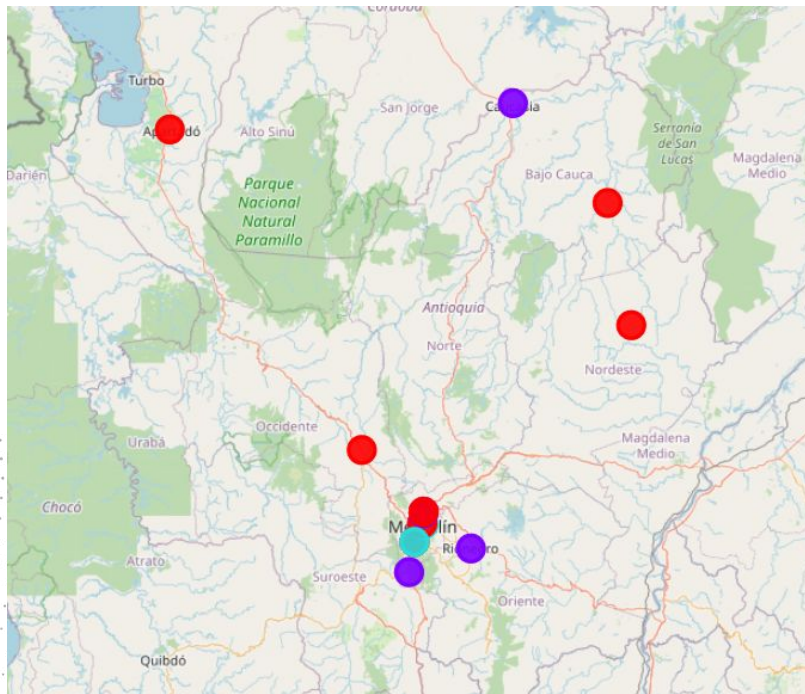
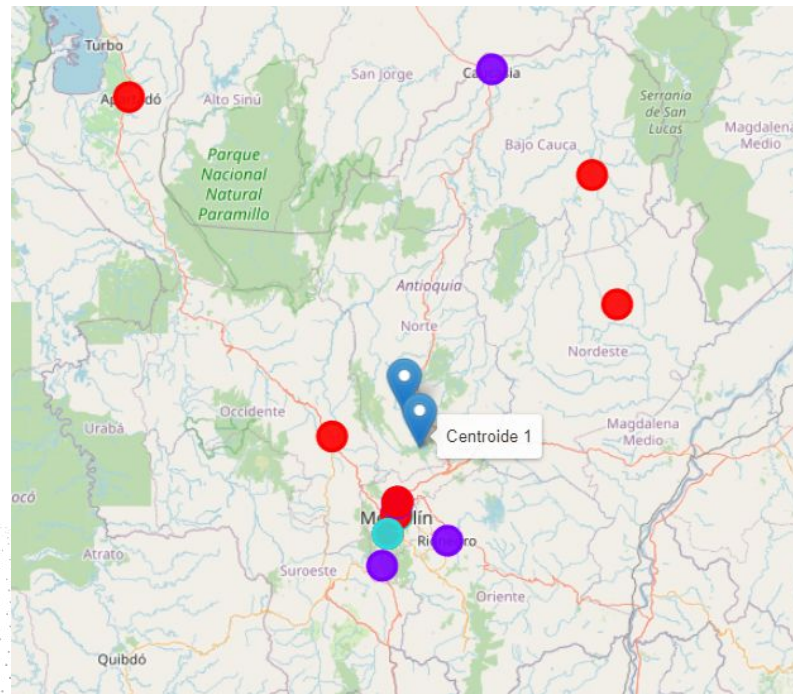
Mapa de cantidad de productos por centro



Mapa de cantidad de proyectos por centro



Mapa de clusters por centro de formación



¿Posibles relación entre centros de formación y centroides de clusters? Se necesita extender a datos reales



6. Modelos de analítica de datos



@SENAcomunica

www.sena.edu.co



Temáticas

1. Preprocesado de bases de datos
2. Bases de datos generadas en R Statistics
3. Estadística descriptiva de datos
 1. Principal_Proyectos
 2. Principal Productos
4. Visualización estadística de datos



@SENAcomunica

www.sena.edu.co



Temáticas

5. Modelo de analítica de datos

1. Machine learning no supervisado: Clustering



@SENAcomunica

www.sena.edu.co

1. Procesado de Bases de Datos

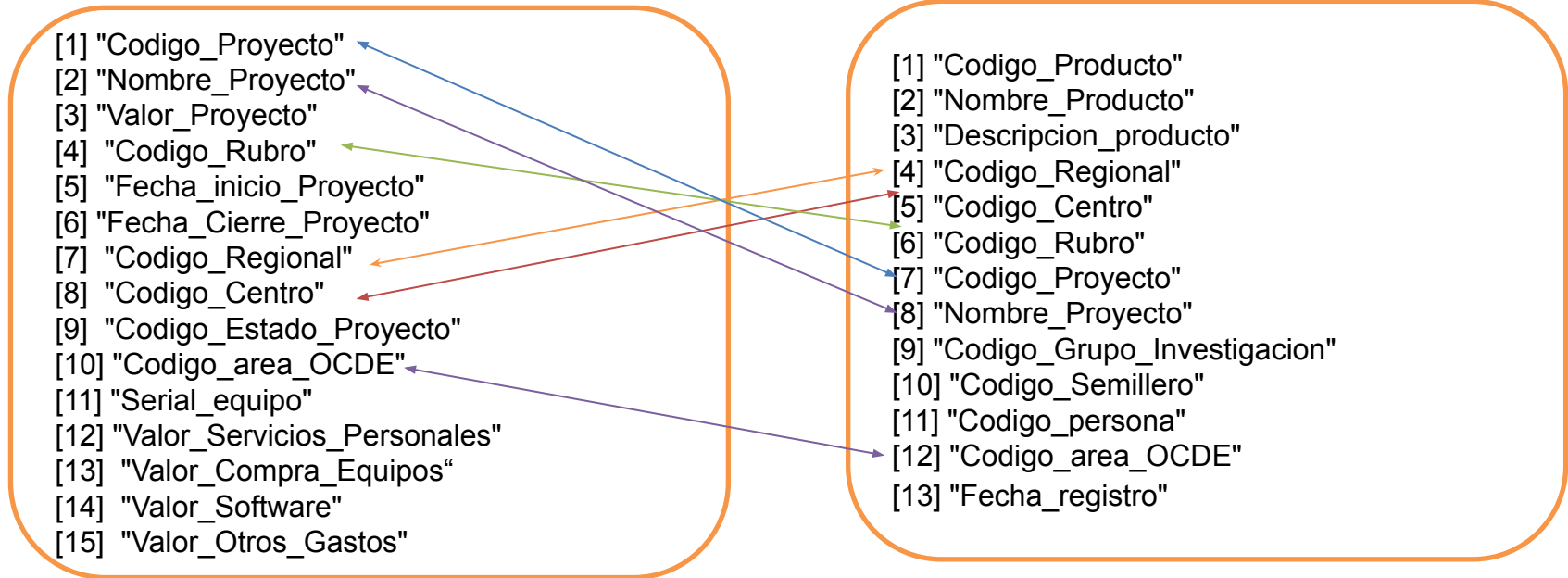
Las tablas a reproducir son las siguientes:

1. Principal_Proyectos: los campos que esta deben tener son:

- [1] "Codigo_Proyecto"
- [2] "Nombre_Proyecto"
- [3] "Valor_Proyecto"
- [4] "Codigo_Rubro"
- [5] "Fecha_inicio_Proyecto"
- [6] "Fecha_Cierre_Proyecto"
- [7] "Codigo_Regional"
- [8] "Codigo_Centro"
- [9] "Codigo_Estado_Proyecto"
- [10] "Codigo_area_OCDE"
- [11] "Serial_equipo"
- [12] "Valor_Servicios_Personales"
- [13] "Valor_Compra_Equipos"
- [14] "Valor_Software"
- [15] "Valor_Otros_Gastos"

2. Principal_Productos: los campos que esta deben tener son:

- [1] "Codigo_Producto"
- [2] "Nombre_Producto"
- [3] "Descripcion_producto"
- [4] "Codigo_Regional"
- [5] "Codigo_Centro"
- [6] "Codigo_Rubro"
- [7] "Codigo_Proyecto"
- [8] "Nombre_Proyecto"
- [9] "Codigo_Grupo_Investigacion"
- [10] "Codigo_Semillero"
- [11] "Codigo_persona"
- [12] "Codigo_area_OCDE"
- [13] "Fecha_registro"



1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Proyectos

Formato de datos:

[1] **"Codigo_Proyecto"** y [2] **"Nombre_Proyecto"**: Estos son campos únicos por proyectos.

[1] **"Codigo_Proyecto"**: Número de cuatro dígitos: 1984 - 2483

[3] **"Valor_Proyecto"**: El valor total en dinero que usa cada centro en un rango de \$400,000,000 a \$600,000,000, por lo tanto, no puede aparecer proyectos que superen todo este monto, además de que la suma total del monto otorgado para todos los proyectos de CTel, deben sumar esa cifra por centro.

Ojo téngase en cuenta que solo hay 16 centros y están en la hoja de Excel **Centros_Formacion**

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Proyectos

Formato de datos:

[4] **"Codigo_Rubro"** debe ser único por proyecto, sacarlo de la tabla de excel: **Rubro**. Cabe aclarar que proyectos distintos si pueden tener el mismo número de Codigo_Rubro.

[5] **"Fecha_inicio_Proyecto"** y [6] **"Fecha_Cierre_Proyecto"**: la duración total debe ser entre 6 y 18 meses

[7] **"Codigo_Regional"** y [8] **"Codigo_Centro"**: Cada proyecto tiene un único código para cada uno de estos campos, pero es posible tener varios proyectos que compartan los mismos códigos Regional y de Centro.

[9] **"Codigo_Estado_Proyecto"**, se encuentra en la hoja de excel de **Estado_Proyecto**

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: **Principal_Proyectos**

Formato de datos:

[10] "**Codigo_area_OCDE**", es distinto para cada proyecto, sin embargo, proyectos distintos podrían tener el mismo código de área OCDE.

[11] "**Serial_equipo**", es un campo que depende de la tabla **Principal_Equipos**, sin embargo, para el caso se genero inicialmente aleatorio en un rango de 6456 - 8527

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Proyectos

Formato de datos:

Los siguientes campos, al ser valores monetarios, del Valor_Proyecto, no pueden superar el valor total destinado a ese proyecto, es decir, que hay que tener como restricción que debe haber un valor en dinero destinado a los productos de CTel.

[12] "Valor_Servicios_Personales"

[13] "Valor_Compra_Equipos"

[14] "Valor_Software"

[15] "Valor_Otros_Gastos"

Son valores cuya suma total debe dar igual al presupuesto asignado a un cada proyecto, pero cuyos valores individuales son distintos, para ello es posible aplicar el método de las particiones o hacerlo numéricamente, mediante algoritmos. Para el caso se realizó un algoritmo que generara los valores enteros aleatoriamente.

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Productos

El tamaño de esta tabla es aleatorio ya que la productividad en el tiempo de los proyectos de investigación por centro no es fija, debido a que son 16 centros y seis tipologías de productos en un rango de 13 años.

[1] **"Codigo_Producto"** y [2] **"Nombre_Producto"**: Ambos campos son únicos para cada Nombre_Producto

Nombre_Producto

- [1,] "Apropiación_Social"
- [2,] "Articulo_Sometido_en_revista_Indexada"
- [3,]
"Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital"
- [4,] "Informe_de_investigación"
- [5,] "Libros_resultados_de_investigación"
- [6,] "Productos_tecnológicos_certificados_o_validados"

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Productos

[3] "**Descripcion_producto**": Este campo esta vacío y no se usará

[4] "**Codigo_Regional**": Campo aleatorio de filas llenas con el número 5, que representan la regional Antioquia.

[5] "**Codigo_Centro**": Son 16 códigos en total para Antioquia, y cada Codigo_Proyecto, con su respectivo Nombre_Proyecto, debe tener un único código de centro asignado, no puede existir un proyecto para varios centros, pero si varios proyectos, distintos en el mismo centro.

[6] "**Codigo_Rubro**": Cada proyecto debe tener un único Codigo_Rubro, pero distintos proyectos pueden coincidir en el mismo Codigo_Rubro.

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Productos

[7] **"Codigo_Proyecto"**: Este campo es único para un proyecto en particular.

[8] **"Nombre_Proyecto"**: Para un proyecto en particular, este campo tiene asociado un Grupo de investigación único, y está asociado a un **Codigo_Centro** único.

[9] **"Codigo_Grupo_Investigacion"**: este campo es único por centro y proyecto.

[10] **"Codigo_Semillero"**: solo se tienen los semilleros del código de centro: 9402, Centro_de_Servicios_y_Gestión_Empresaria, de SENNOVA, donde actualmente se está ejecutando el proyecto: Proyecto SGPS 9748

Para el proyecto indicadores de CTel con Analítica de datos, en los demás centros no se cuenta con información, por lo tanto, inicialmente se deja esta fila: Codigo_Semillero vacía

1. Procesado de Bases de Datos

Restricciones y análisis para la Base de Datos: Principal_Productos

[11] **"Codigo_persona"**: Debe coincidir con el Codigo_Persona de la base de datos Principal_Talento_Humano, sugiero dejar esta base de datos lista luego.

[12] **"Codigo_area_OCDE"**: este campo es único para un proyecto en específico, pero distintos proyectos si pueden tener el mismo código de área OCDE.

[13] **"Fecha_registro"**: es aquella donde se registran la productividad generada por un proyecto.



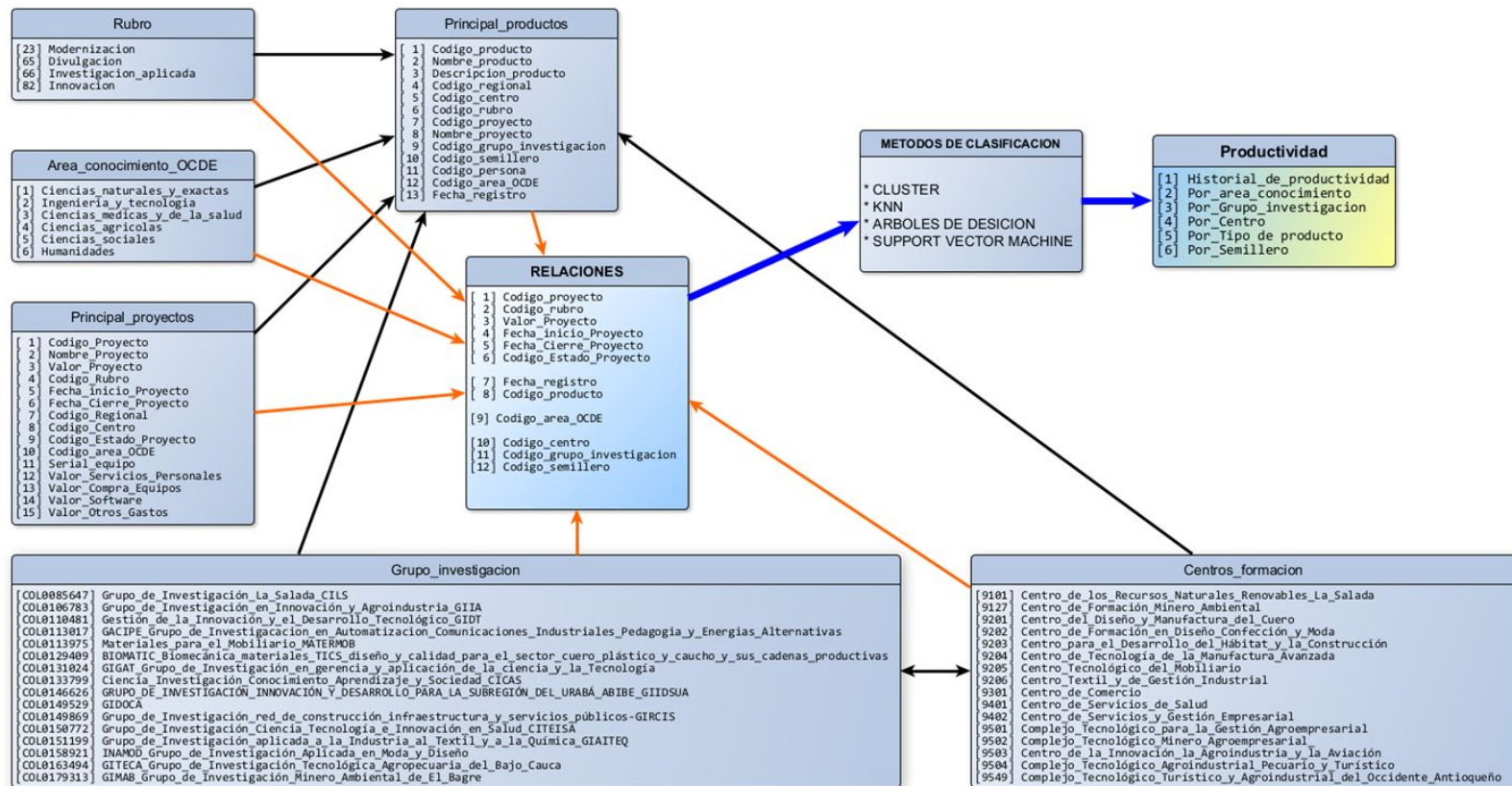
2. Bases de datos generadas en R statistics



@SENAcomunica

www.sena.edu.co

Identificación de base datos modelo relacional



Identificación de campos en base datos

Principal_Proyectos	Principal_Productos
<ul style="list-style-type: none">[1] "Codigo_Proyecto"[2] "Nombre_Proyecto"[3] "Valor_Proyecto"[4] "Codigo_Rubro"[5] "Fecha_inicio_Proyecto"[6] "Fecha_Cierre_Proyecto"[7] "Codigo_Regional"[8] "Codigo_Centro"[9] "Codigo_Estado_Proyecto"[10] "Codigo_area_OCDE"[11] "Serial_equipo"[12] "Valor_Servicios_Personales"[13] "Valor_Compra_Equipos"[14] "Valor_Software"[15] "Valor_Otros_Gastos"	<ul style="list-style-type: none">[1] "Codigo_Producto"[2] "Nombre_Producto"[3] "Descripcion_producto"[4] "Codigo_Regional"[5] "Codigo_Centro"[6] "Codigo_Rubro"[7] "Codigo_Proyecto"[8] "Nombre_Proyecto"[9] "Codigo_Grupo_Investigacion"[10] "Codigo_Semillero"[11] "Codigo_persona"[12] "Codigo_area_OCDE"[13] "Fecha_registro"

Base de datos principal proyectos



	Codigo_Proyecto	Nombre_Proyecto	Id_Proyecto	Id_Rubrica	Inicio_Proyecto	Cierre_Proyecto	Id_Región	Id_Centro_Estado	Id_Provincia	Id_Area_OC	Id_Equipos	Servicios_Personal	Compra_Equipos	Software	Valor_Otros_Gastos	
2	2005	EMXQ23HF	503601821	65	2010-02-20	2011-04-16	5	9101	3	4	7632	235335197,3	57902755,06	83351582	57012266,04	
3	1931	KDWV85RX	189342427	82	2010-02-10	2010-11-07	5	9127	3	1	7888	110389312,4	31163544,33	27360001,8	20422266,04	
4	2013	PFXN30EU	242702291	66	2010-02-26	2011-05-22	5	9127	3	3	8363	141052660,4	33827436,59	35726386,2	26039588,14	
5	1938	MGRE44SJ	442236533	82	2010-02-17	2010-11-14	5	9201	3	1	8193	218520305,6	78519755,06	82775458,6	62481019,75	
6	1984	WSCU80DH	104240335	66	2010-02-01	2011-07-26	5	9202	3	1	7644	51863855,28	17558578,08	14401880,9	20410020,44	
7	2007	KKIM95VB	142131146	65	2010-02-21	2011-03-18	5	9202	3	1	8429	70724273,83	23941028,52	19636888,6	27828955,12	
8	2014	IGFR31EC	211503962	23	2010-02-27	2010-03-25	5	9202	3	6	7540	105244106,7	35626412,61	29221463,7	4141938,52	
9	1935	AXNP43SJ	163527436	23	2010-02-14	2011-03-11	5	9203	3	3	6456	75343790,42	27753013,1	33133883	33290743,72	
10	2002	TSXV81MZ	182531201	65	2010-02-18	2010-12-15	5	9203	5	6360	81132235,38	29882816,31	35676622	35845526,34	23116888,66	
11	2003	MIPN51DA	148343200	65	2010-02-19	2011-03-16	5	9204	3	5	7463	78521909,18	24128064,39	16582336,9	23116888,66	
12	2004	UZPC16OI	135617258	82	2010-02-19	2010-03-17	5	9204	3	3	7391	71782834,4	22057294,73	1519172,2	26617956,34	
13	2015	DLTT23YG	342360419	23	2010-02-28	2010-11-25	5	9204	3	2	7646	181212333,9	55682770,48	38268732,2	67195382,32	
14	1936	ABGJ8THG	407305180	65	2010-02-15	2010-03-13	5	9205	3	6	7709	232106253,7	44705065,62	78156767,4	52337032,88	
15	2011	XZS7VITZQ	448405947	65	2010-02-24	2011-07-19	5	9206	3	4	7601	253233764	51074707,61	54336162,3	83701313,46	
16	1985	IUYX78AX	158621119	65	2010-02-02	2011-05-28	5	9301	3	2	7767	67539066,04	23103334,04	30930285,7	30387893,61	
17	1939	KZBB40JO	132407062	65	2010-02-17	2010-11-14	5	9301	3	1	6481	56377337,54	24234147,34	25868756,6	25868753,82	
18	2006	OBUC93BM	283756032	82	2010-02-20	2011-01-16	5	9301	3	2	6572	120820066,6	52063782,56	55438261,2	55433981,97	
19	1932	JSXQ25FI	145264566	66	2010-02-11	2010-03-09	5	9401	4	7504	67328522,23	24035009,13	27530788,3	26310246,57	26310246,57	
20	1933	WZBC50TN	111526310	82	2010-02-13	2011-06-08	5	9401	3	3	8417	51631212,84	18438850,18	21136656,4	20193590,16	
21	2009	LEIG33YD	301361769	23	2010-02-22	2011-04-18	5	9401	3	2	6747	133677851,7	43986825,82	57114596,3	54582434,86	
22	1936	RSAG49VB	210287835	82	2010-02-04	2011-07-29	5	9402	2	8166	103570421,7	371727384,62	23308114,8	33681974,36	33681974,36	
23	2010	CUCI60T2	143443667	23	2010-02-23	2011-07-18	5	9402	3	2	6510	74744383,32	25736054,52	15839827,7	27063334,44	
24	1988	NWYD51XG	208582829	65	2010-02-04	2010-12-31	5	9501	2	7536	112482718,9	38280385,83	33070378,6	24749345,65	24749345,65	
25	2012	GELY55UH	177538275	65	2010-02-24	2011-06-19	5	9501	3	4	8014	35714284,33	32582301,07	28145328,5	21065761,58	
26	1934	SRYG03PM	132211844	65	2010-02-13	2011-07-08	5	9502	2	7886	63283679	24480666,59	22166267,3	16231030,81	16231030,81	
27	2000	VNCF63HD	129543594	65	2010-02-17	2011-07-12	5	9502	3	2	7177	61885421,6	23986802,86	21716916	15952453,12	
28	2001	XACB40VY	317828760	23	2010-02-17	2010-12-14	5	9502	4	7439	166553503,8	58850427,19	53286279,5	33138549,89	33138549,89	
29	1961	UBJT90KA	173937835	65	2010-02-04	2011-06-29	5	9503	4	8035	88801562,72	31051234,03	23466508,5	24618530,19	24618530,19	
30	1930	ASSD67JH	140406144	66	2010-02-08	2011-01-04	5	9503	2	8361	93387174,22	32864686,66	31187410,1	26056300,38	26056300,38	
31	1983	ZKEO34WM	479817844	23	2010-02-06	2011-05-02	5	9504	2	7161	223070592,6	92440886,66	92911533,4	71034381,32	71034381,32	
32	1931	PXJD36LP	262535947	23	2010-02-15	2011-03-12	5	9549	3	1	6735	159620733,9	45117538,38	31308126,6	26549447,85	26549447,85
33	2008	DNKK2IMF	215835976	82	2010-02-21	2010-12-18	5	9549	4	7378	13119741,6	37083556,62	2573356,7	21821945,51	21821945,51	
34	2038	BEZM53AU	101055326	66	2011-02-17	2012-07-11	5	9101	5	6357	50601225,07	14645432,43	1855561,1	17302573,69	17302573,69	
35	2044	HDXO74KG	106071238	65	2011-02-21	2010-11-19	5	9101	5	6372	53086565,33	15364823,65	19467034,5	18528913,25	18528913,25	
36	2053	PJN141YD	249393825	82	2011-02-28	2012-01-24	5	9101	5	8223	125119390,1	36213428,92	45881950	42784456,12	42784456,12	
37	2034	XUDR03GE	15516314	66	2011-02-15	2011-11-14	5	9127	3	6	8311	73241437,15	14863053,39	14190471,1	13222000,33	13222000,33
38	2036	CWF030EL	73283548,6	65	2011-02-16	2012-06-10	5	9127	3	6	6464	44644125,77	3429075,778	9002353,75	8387933,33	8387933,33
39	2043	FUV161YK	107563287	23	2011-02-21	2012-03-17	5	9127	3	2	7882	68211262,66	13842274,1	13215828,5	1210321,91	1210321,91
40	2046	UKHD52NS	28216088	23	2011-02-22	2011-12-19	5	9127	3	6	7544	178870663,2	36239651,2	34655320,3	32230847,09	32230847,09
41	2031	DQRY42ZS	150297267	65	2011-02-13	2012-07-07	5	9201	5	6652	86663295,74	13510771,46	19428006,2	24635193,12	24635193,12	
42	2040	BJEY10RI	163107576	82	2011-02-20	2012-07-14	5	9201	5	7650	34334739,64	21251388,46	21161233,5	26838328,6	26838328,6	
43	2043	GEED98VT	363854470	82	2011-02-26	2012-02-21	5	9202	3	4	6499	203483891,3	70337804,6	38540324,5	51292445,78	51292445,78
	Principal Proyectos		Estado Proyecto		Nivel Academico		Municipio		Centros Formacion		De ...					

Base de datos obtenida en R y exportada a Excel **Principal_Proyectos**

Base de datos principal proyectos en R



stering.R x Tablas de Datos CTeI New.R x Principal_Proyectos x PPCY x NPCY x PCY x datos.scale x TTPC_2010 x TPyPdPCY x df x PyCY x Indicador_centros x

Filter

	Codigo_Proyecto	Nombre_Proyecto	Valor_Proyecto	Codigo_Rubro	Fecha_inicio_Proyecto	Fecha_Cierre_Proyecto	Codigo_Regional	Codigo_Centro	Codigo_Estado_Proyecto	Codigo_area_OCDE
1	2005	EMXQ29HF	503601821	65	2010-02-20	2011-04-16	5	9101	3	
18	1991	KDWW85RX	189942427	82	2010-02-10	2010-11-07	5	9127	3	
2	2013	PFXN90EU	242702291	66	2010-02-26	2011-05-22	5	9127	3	
3	1998	MGRE44SJ	442296539	82	2010-02-17	2010-11-14	5	9201	3	
453	1984	WSCU80DH	104240335	66	2010-02-01	2011-07-26	5	9202	3	
4	2007	KKIM95VB	142131146	65	2010-02-21	2011-03-18	5	9202	3	
24	2014	IGFR31EC	211503982	23	2010-02-27	2010-09-25	5	9202	3	
5	1995	AXNP43SJ	169527436	23	2010-02-14	2011-03-11	5	9203	3	
463	2002	TSXW81MZ	182537201	65	2010-02-18	2010-12-15	5	9203	3	
6	2003	MIPN51DA	148349200	65	2010-02-19	2011-03-16	5	9204	3	
448	2004	U7PC16OI	135617258	82	2010-02-19	2010-09-17	5	9204	3	

Showing 1 to 11 of 500 entries, 15 total columns

Dataframe obtenido en R **Principal_Proyectos**

Base de datos principal productos exportados a excel desde R



A	B	C	D	E	F	G	H	I	J	K	L	M
Idigo_Produc	Nombre_Producto	scripcion_produccion	Idigo_Regional	Idigo_Ceato	Idigo_Rebo	Idigo_Projec	Idigo_Grupo	Idigo_Investig	Idigo_Semille	Idigo_persona	Idigo_area_OC	Fecha_registro
8020	Apropiación_Social		5	3101	65	2005	EMX022HF	COL0085647	8863	3114	4	2010-12-14
8019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	3101	65	2005	EMX022HF	COL0085647	8863	3750	4	2011-05-04
8016	Libros_resultados_de_investigación		5	3101	65	2005	EMX022HF	COL0085647	8866	3503	4	2011-03-08
8031	Productos_tecnológicos_certificados_o_validados		5	3101	65	2005	EMX022HF	COL0085647	8870	3784	4	2011-07-30
8021	Artículo_Sometido_en_revista_Indexada		5	3101	65	2005	EMX022HF	COL0085647	8863	3204	4	2010-11-06
8025	Artículo_Sometido_en_revista_Indexada		5	3101	65	2005	EMX022HF	COL0085647	8861	3061	4	2010-10-06
8016	Libros_resultados_de_investigación		5	3127	82	1931	KDWW85RX	COL0173313	8858	3836	1	2010-09-15
8020	Apropiación_Social		5	3127	82	1931	KDWW85RX	COL0173313	8873	3503	1	2011-03-19
8019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	3127	82	1931	KDWW85RX	COL0173313	8882	3874	1	2011-07-03
8016	Libros_resultados_de_investigación		5	3127	66	2013	PFYN90EU	COL0173313	8853	3623	3	2011-06-18
8033	Informe_de_investigación		5	3127	66	2013	PFYN90EU	COL0173313	8862	3041	3	2011-01-07
8025	Artículo_Sometido_en_revista_Indexada		5	3127	66	2013	PFYN90EU	COL0173313	8873	3863	3	2011-05-19
8016	Libros_resultados_de_investigación		5	3127	66	2013	PFYN90EU	COL0173313	8873	3273	3	2011-04-12
8031	Productos_tecnológicos_certificados_o_validados		5	3127	66	2013	PFYN90EU	COL0173313	8870	3185	3	2011-05-26
8031	Productos_tecnológicos_certificados_o_validados		5	3127	66	2013	PFYN90EU	COL0173313	8874	3383	3	2011-01-26
8016	Libros_resultados_de_investigación		5	3201	82	1936	MGRE44SJ	COL0123409	8880	3650	1	2011-06-18
8020	Apropiación_Social		5	3201	82	1936	MGRE44SJ	COL0123409	8886	3636	1	2011-07-08
8019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	3201	82	1936	MGRE44SJ	COL0123409	8888	3433	1	2011-07-05
8016	Libros_resultados_de_investigación		5	3201	82	1936	MGRE44SJ	COL0123409	8863	3009	1	2010-10-21
8016	Libros_resultados_de_investigación		5	3201	82	1936	MGRE44SJ	COL0123409	8888	3137	1	2010-08-23
8019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	3202	66	1984	WSCU80DH	COL0158921	8862	3386	1	2011-07-02
8025	Artículo_Sometido_en_revista_Indexada		5	3202	66	1984	WSCU80DH	COL0158921	8877	3514	1	2010-10-25
8020	Apropiación_Social		5	3202	66	1984	WSCU80DH	COL0158921	8880	3305	1	2010-08-12
8016	Libros_resultados_de_investigación		5	3202	65	2007	KKIM95VB	COL0158921	8882	3836	1	2010-08-25
8020	Apropiación_Social		5	3202	65	2007	KKIM95VB	COL0158921	8881	3746	1	2010-10-18
8033	Informe_de_investigación		5	3202	65	2007	KKIM95VB	COL0158921	8873	3107	1	2011-03-30
8033	Informe_de_investigación		5	3202	65	2007	KKIM95VB	COL0158921	8853	3430	1	2011-01-15
8020	Apropiación_Social		5	3202	23	2014	IGFR31EC	COL0158921	8875	3477	6	2011-03-31
8016	Libros_resultados_de_investigación		5	3202	23	2014	IGFR31EC	COL0158921	8868	3148	6	2010-03-11
8033	Informe_de_investigación		5	3202	23	2014	IGFR31EC	COL0158921	8880	3545	6	2011-04-20
8020	Apropiación_Social		5	3202	23	2014	IGFR31EC	COL0158921	8878	3837	6	2011-03-17
8019	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales_no_especializadas_producciones_de_contenido_digital		5	3203	23	1935	AXNP43SJ	COL0143663	8875	3162	3	2011-03-09
8033	Informe_de_investigación		5	3203	23	1935	AXNP43SJ	COL0143663	8883	3314	3	2011-01-21
8033	Informe_de_investigación		5	3203	23	1935	AXNP43SJ	COL0143663	8860	3421	3	2011-04-27
8025	Artículo_Sometido_en_revista_Indexada		5	3203	23	1935	AXNP43SJ	COL0143663	8868	3583	3	2011-07-14
8025	Artículo_Sometido_en_revista_Indexada		5	3203	23	1935	AXNP43SJ	COL0143663	8854	3090	3	2011-01-12
8031	Productos_tecnológicos_certificados_o_validados		5	3203	23	1935	AXNP43SJ	COL0143663	8854	3708	3	2010-10-04
8020	Apropiación_Social		5	3203	65	2002	TSXV8IMZ	COL0143663	8858	3419	5	2010-12-02
8033	Informe_de_investigación		5	3203	65	2002	TSXV8IMZ	COL0143663	8884	3857	5	2010-11-11
8025	Artículo_Sometido_en_revista_Indexada		5	3203	65	2002	TSXV8IMZ	COL0143663	8884	3241	5	2011-01-14
8025	Artículo_Sometido_en_revista_Indexada		5	3203	65	2002	TSXV8IMZ	COL0143663	8860	3123	5	2010-08-17
8016	Libros_resultados_de_investigación		5	3203	65	2002	TSXV8IMZ	COL0143663	8888	3435	5	2010-12-05

Base de datos de **Principal_Productos** realizada en R y exportada a excel

Base de datos principal productos



stering.R x Tablas de Datos CTeI New.R x Principal_Productos x Principal_Proyectos x PPCY x NPCY x PCY x datos.scale x TPC2_2010 x TPYPdPCY x df x PyCY x »

Filter

	Codigo_Producto	Nombre_Producto	Descripcion_producto	Codigo_Regional	Codigo_Centro	Codigo_Rubro	Codigo_Proyecto	Nombre_Proyecto	Codigo_Grupo_I
1	8020	Apropiación Social		5	9101	65	2005	EMXQ29HF	COL0085647
2	8019	Divulgación Pública de la CTeI Publicaciones editoriales n...		5	9101	65	2005	EMXQ29HF	COL0085647
3	8016	Libros resultados de investigación		5	9101	65	2005	EMXQ29HF	COL0085647
4	8031	Productos tecnológicos certificados o validados		5	9101	65	2005	EMXQ29HF	COL0085647
5	8031	Productos tecnológicos certificados o validados		5	9101	65	2005	EMXQ29HF	COL0085647
6	8025	Artículo Sometido en revista Indexada		5	9101	65	2005	EMXQ29HF	COL0085647
7	8016	Libros resultados de investigación		5	9127	82	1991	KDWW85RX	COL0179313
8	8020	Apropiación Social		5	9127	82	1991	KDWW85RX	COL0179313
9	8019	Divulgación Pública de la CTeI Publicaciones editoriales n...		5	9127	82	1991	KDWW85RX	COL0179313
10	8016	Libros resultados de investigación		5	9127	66	2013	PFXN90EU	COL0179313
11	8030	Informe de investigación		5	9127	66	2013	PFXN90EU	COL0179313

Showing 1 to 11 of 2,268 entries, 13 total columns

Base de datos obtenido en R de **Principal_Productos**



3. Estaística Descriptiva de datos



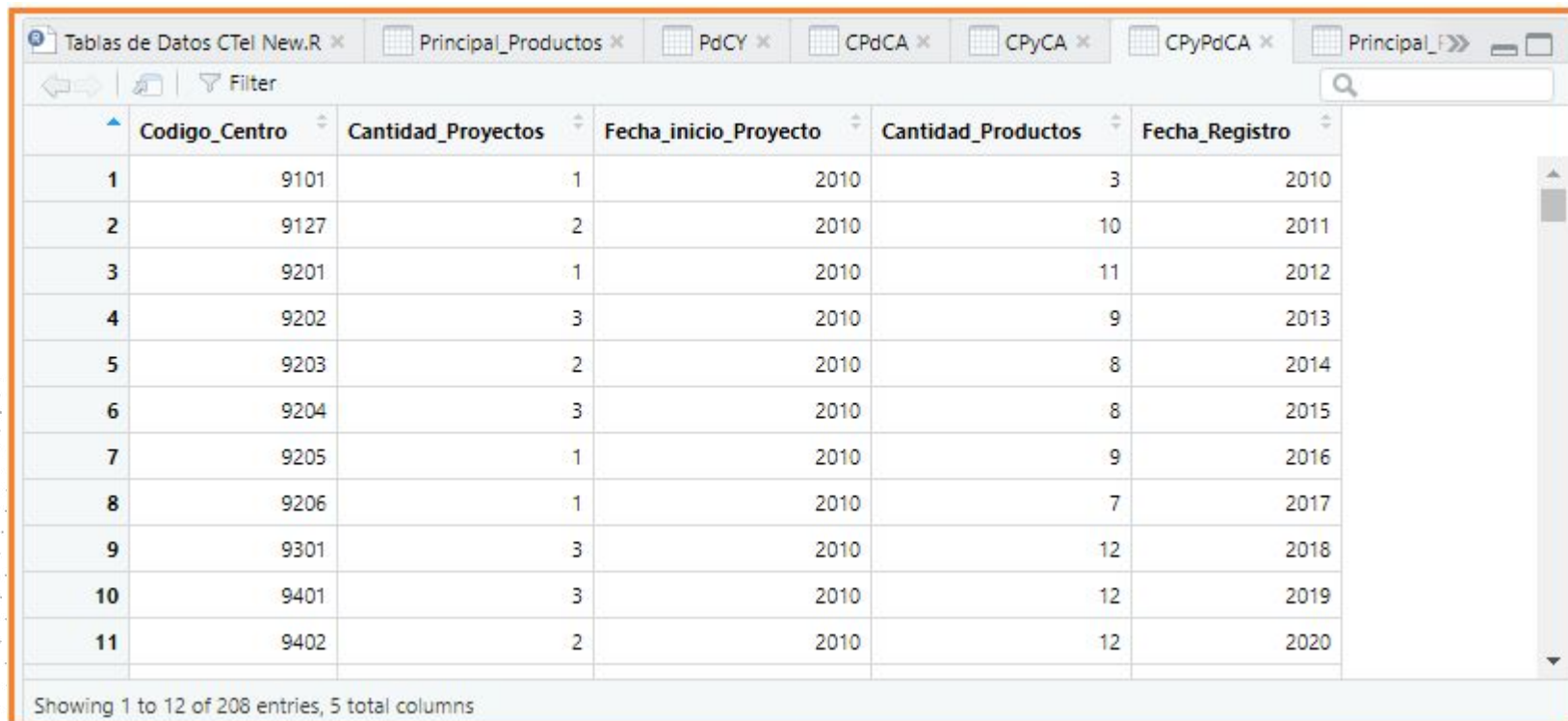
@SENAcomunica

www.sena.edu.co

Cantidad de proyectos por centros, por productos y por años



Se genera una tabla en R statistics, donde se evidencia cada `Codigo_Centro`, con la cantidad de proyectos y productos en un rango de 13 años, desde el 2010 hasta el 2022.

The image is a screenshot of an R Studio window. The top pane shows several open scripts: 'Tablas de Datos CTel New.R', 'Principal_Productos', 'PdCY', 'CPdCA', 'CPyCA', 'CPyPdCA', and 'Principal_'. The bottom pane displays a data frame with 5 columns: 'Codigo_Centro', 'Cantidad_Proyectos', 'Fecha_inicio_Proyecto', 'Cantidad_Productos', and 'Fecha_Registro'. The data is presented in a table with 11 rows, numbered 1 to 11. The status bar at the bottom indicates 'Showing 1 to 12 of 208 entries, 5 total columns'.

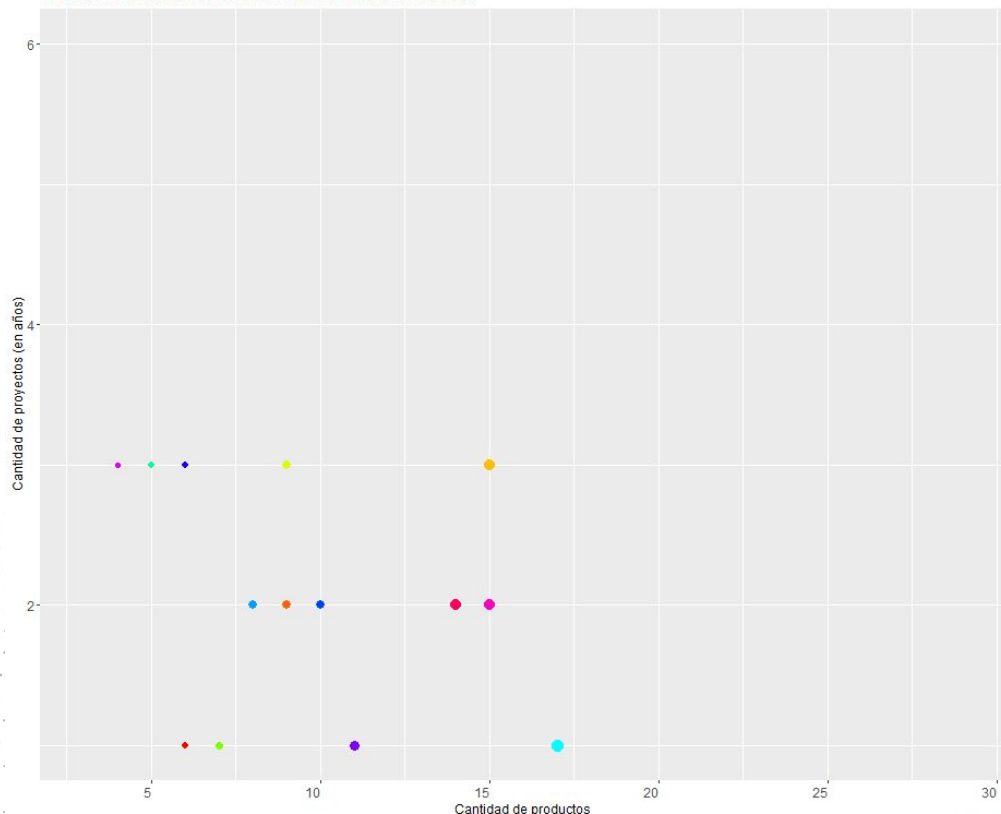
	<code>Codigo_Centro</code>	<code>Cantidad_Proyectos</code>	<code>Fecha_inicio_Proyecto</code>	<code>Cantidad_Productos</code>	<code>Fecha_Registro</code>
1	9101	1	2010	3	2010
2	9127	2	2010	10	2011
3	9201	1	2010	11	2012
4	9202	3	2010	9	2013
5	9203	2	2010	8	2014
6	9204	3	2010	8	2015
7	9205	1	2010	9	2016
8	9206	1	2010	7	2017
9	9301	3	2010	12	2018
10	9401	3	2010	12	2019
11	9402	2	2010	12	2020

Cantidad de proyectos y productos por centros en 13 años



Año: 2009-12-31 19:00:00

Productividad de centros entre los años del 2010 al 2022



Proyecto CTel

Cantidad de productos de CTel proyectos y productos y centros en 13 años



	Codigo_Centro	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	Fecha_Registro	Apropiación_Social	Articulo_Sometido_en_revista_Indexada	Divulgación_Pública_de_la_CTel_Publicaciones_editoriales
1	9101	1	2010	3	2010	1	1	
2	9127	2	2010	10	2011	0	0	
3	9201	1	2010	11	2012	0	0	
4	9202	3	2010	9	2013	2	1	
5	9203	2	2010	8	2014	1	1	
6	9204	3	2010	8	2015	0	1	
7	9205	1	2010	9	2016	0	0	
8	9206	1	2010	7	2017	1	1	
9	9301	3	2010	12	2018	0	0	
10	9401	3	2010	12	2019	2	0	
11	9402	2	2010	12	2020	0	0	

Showing 1 to 11 of 208 entries, 11 total columns

Cantidad de productos de CTel proyectos y productos y centros en 13 años



aring_1.R	Clustering.R	TPPC_2010	TPyPdPCY	TPdCY	df	Principal_Productos	PdCY	CPdCA	CPyCA	CPyPdCA	Principal_Proyectos	Informe.R
Filter												
	Codigo_Centro	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	Fecha_Registro	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031	
1	9101	1	2010	3	2010	1	1	0	0	0	1	
2	9127	2	2010	1	2010	0	0	0	0	1	0	
3	9201	1	2010	2	2010	0	0	0	0	2	0	
4	9202	3	2010	5	2010	2	1	0	0	2	0	
5	9203	2	2010	5	2010	1	1	0	1	1	1	
6	9204	3	2010	4	2010	0	1	0	0	1	2	
7	9205	1	2010	1	2010	0	0	1	0	0	0	
8	9206	1	2010	3	2010	1	1	0	0	1	0	
9	9301	3	2010	3	2010	0	0	0	2	1	0	
10	9401	3	2010	7	2010	2	0	0	0	1	4	
11	9402	2	2010	4	2010	0	0	0	1	3	0	

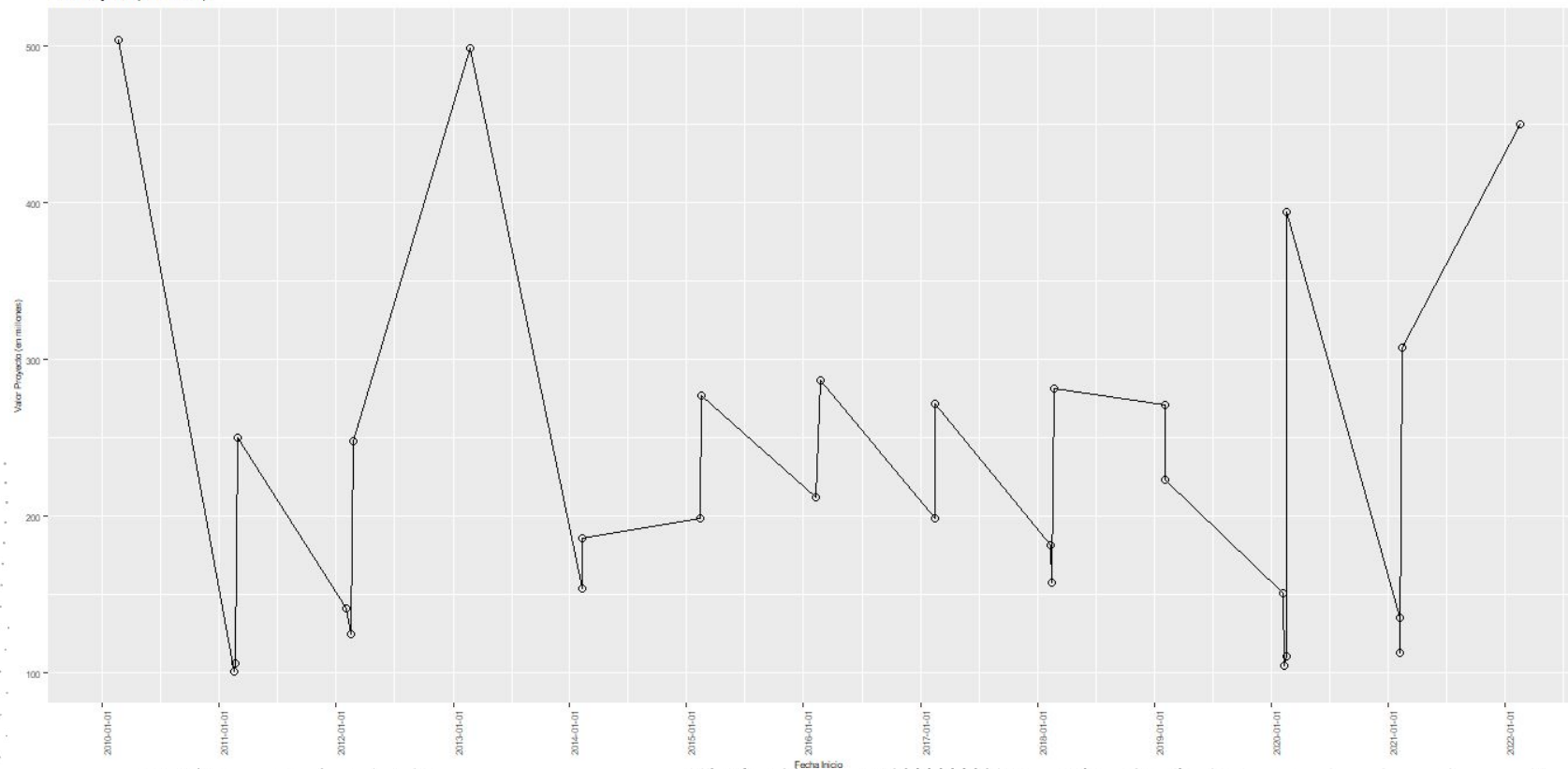
Showing 1 to 12 of 16 entries, 11 total columns

Comportamiento de la inversión por centro en el tiempo



Centro_de_los_Recurso_Naturales_Renovables_La_Salada

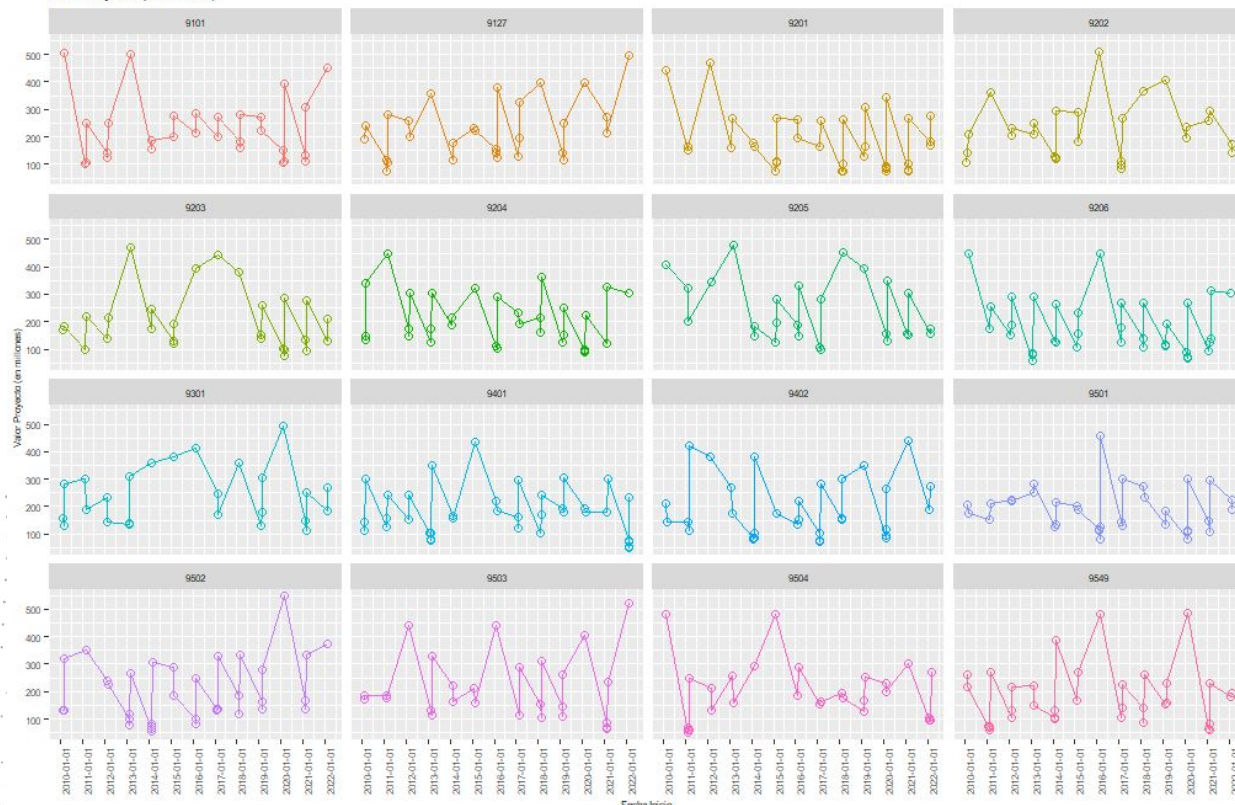
Valor Proyecto (en millones)



Comportamiento de la inversión por centro en el tiempo



Centros de Formación
Valor Proyecto (en millones)



Centro Formación

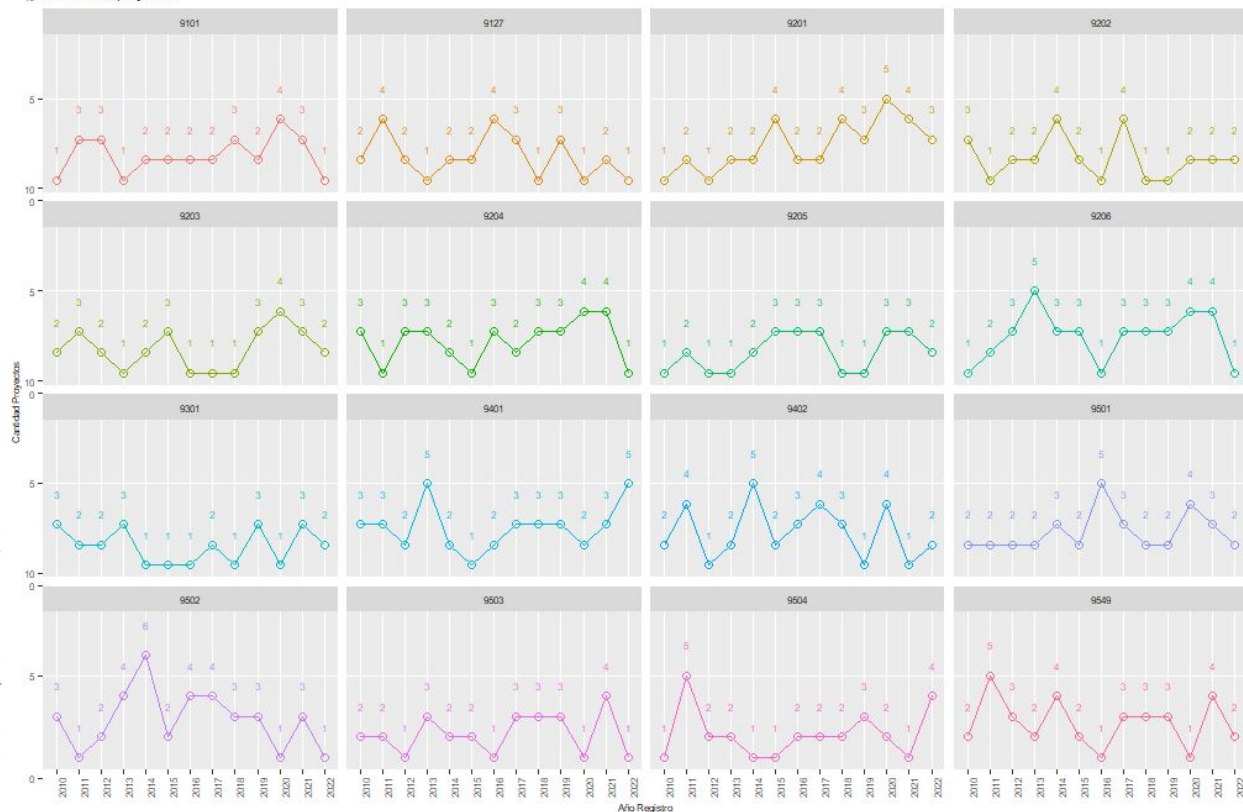
- 9101 - Centro de los Recursos Naturales Renovables La Salada
- 9127 - Centro de Formación Minero Ambiental
- 9201 - Centro del Diseño y Manufactura del Cuero
- 9202 - Centro de Formación en Diseño Confección y Moda
- 9203 - Centro para el Desarrollo del Hábitat y la Construcción
- 9204 - Centro de Tecnología de la Manufactura Avanzada
- 9205 - Centro Tecnológico del Mobiliario
- 9206 - Centro Textil y de Gestión Industrial
- 9301 - Centro de Comercio
- 9401 - Centro de Servicios de Salud
- 9402 - Centro de Servicios y Gestión Empresarial
- 9501 - Complejo Tecnológico para la Gestión Agroempresarial
- 9502 - Complejo Tecnológico Minero Agroempresarial
- 9503 - Centro de la Innovación la Agroindustria y la Aviación
- 9504 - Complejo Tecnológico Agroindustrial Pecuero y Turístico
- 9549 - Complejo Tecnológico Turístico y Agroindustrial del Occidente Antioqueño

Comportamiento de la cantidad de proyectos por centro en el tiempo



Centros de Formación

12-Cantidad de proyectos



Centro Formación

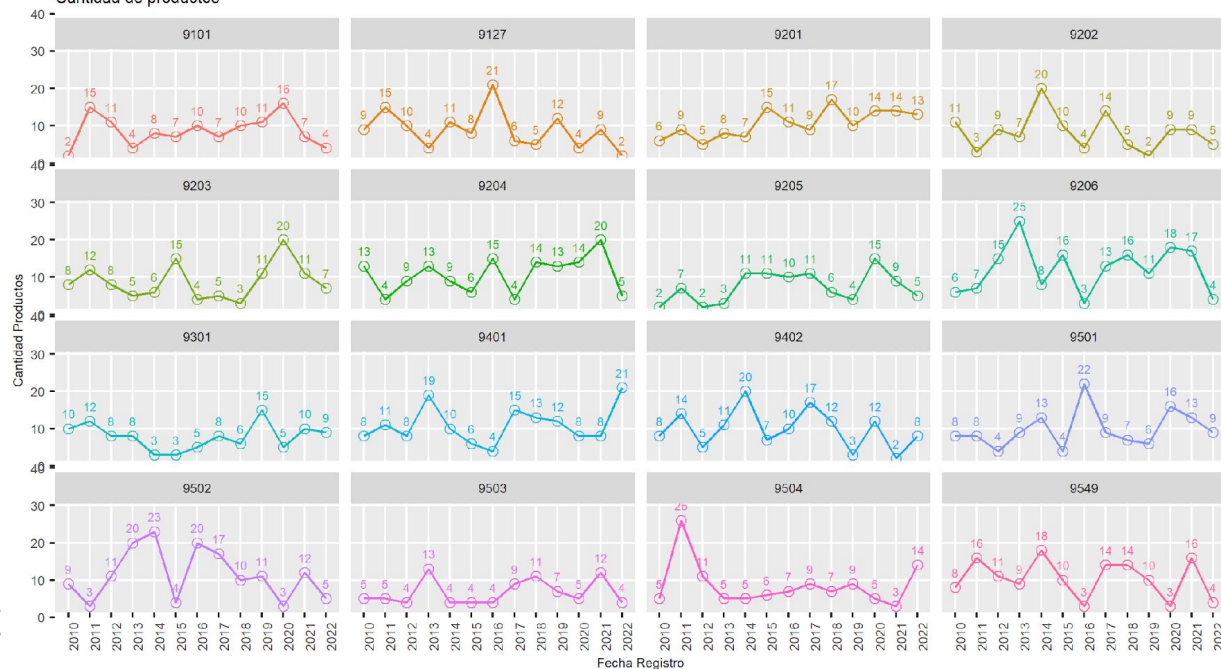
- 9101-Centro de los Recursos Naturales Renovables La Salada
- 9127-Centro de Formación Minero Ambiental
- 9201-Centro del Diseño y Manufactura del Cuero
- 9202-Centro de Formación en Diseño Confección y Moda
- 9203-Centro para el Desarrollo del Hábitat y la Construcción
- 9204-Centro de Tecnología de la Manufactura Avanzada
- 9205-Centro Tecnológico del Mobiliario
- 9206-Centro Textil y de Gestión Industrial
- 9301-Centro de Comercio
- 9401-Centro de Servicios de Salud
- 9402-Centro de Servicios y Gestión Empresarial
- 9501-Complejo Tecnológico para la Gestión Agroempresarial
- 9502-Complejo Tecnológico Minero Agroempresarial
- 9503-Centro de la Innovación la Agroindustria y la Aviación
- 9504-Complejo Tecnológico Agroindustrial Pecuario y Turístico
- 9549-Complejo Tecnológico Turístico y Agroindustrial del Occidente Antioqueño

Comportamiento de la productividad CTel por centro en el tiempo



Centros de Formacion

Cantidad de productos



Centro Formacion

- 9101 - Centro de Recursos Naturales Renovables La Salada
- 9127 - Centro de Formación Minero Ambiental
- 9201 - Centro del Diseño y Manufactura del Cuero
- 9202 - Centro de Formación en Diseño Confección y Moda
- 9203 - Centro para el Desarrollo del Hábitat y la Construcción
- 9204 - Centro de Tecnología de la Manufactura Avanzada
- 9205 - Centro Tecnológico del Mobiliario
- 9206 - Centro Textil y de Gestión Industrial
- 9301 - Centro de Comercio
- 9401 - Centro de Servicios de Salud
- 9402 - Centro de Servicios y Gestión Empresarial
- 9501 - Complejo Tecnológico para la Gestión Agroempresarial
- 9502 - Complejo Tecnológico Minero Agroempresarial
- 9503 - Centro de la Innovación, la Agroindustria y la Aviación
- 9504 - Complejo Tecnológico Agroindustrial Pecuário y Turístico
- 9549 - Complejo Tecnológico Turístico y Agroindustrial del Occidente Antioqueño



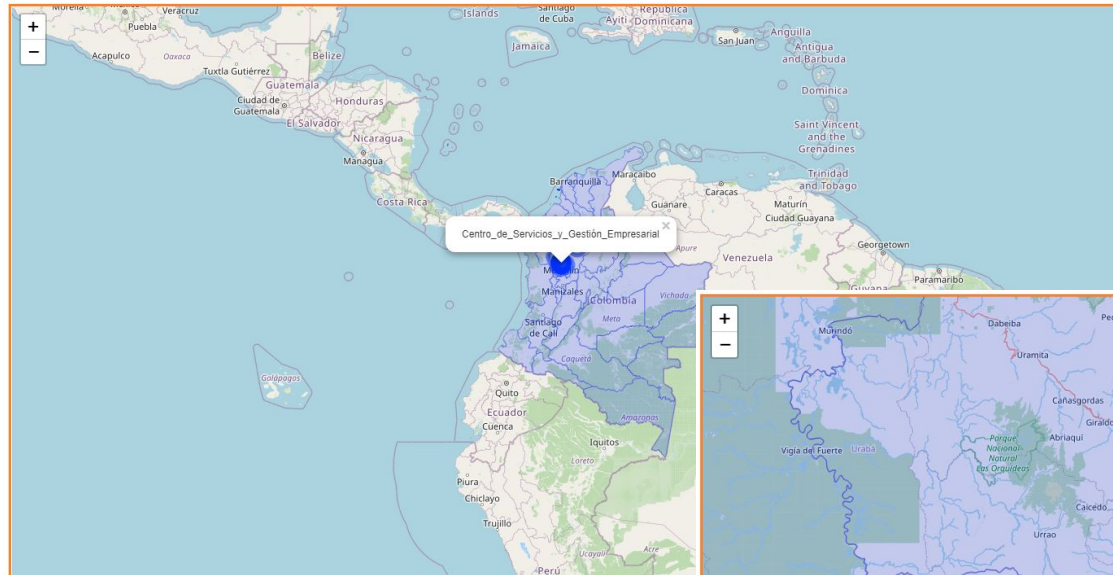
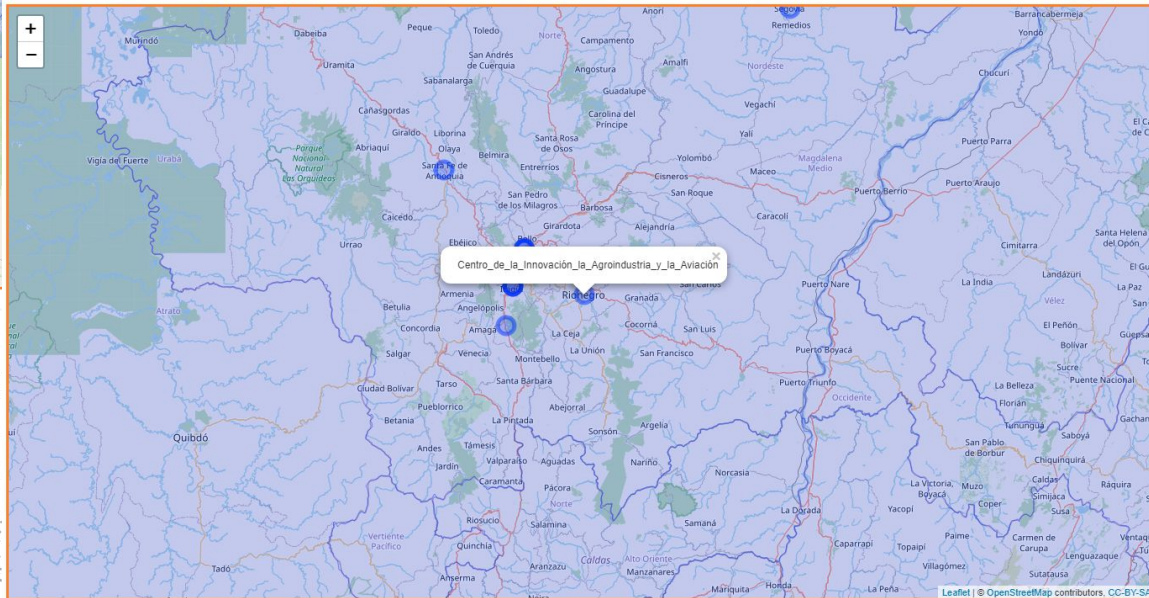
4. Visualización estadística de datos mediante mapas iterativos en R

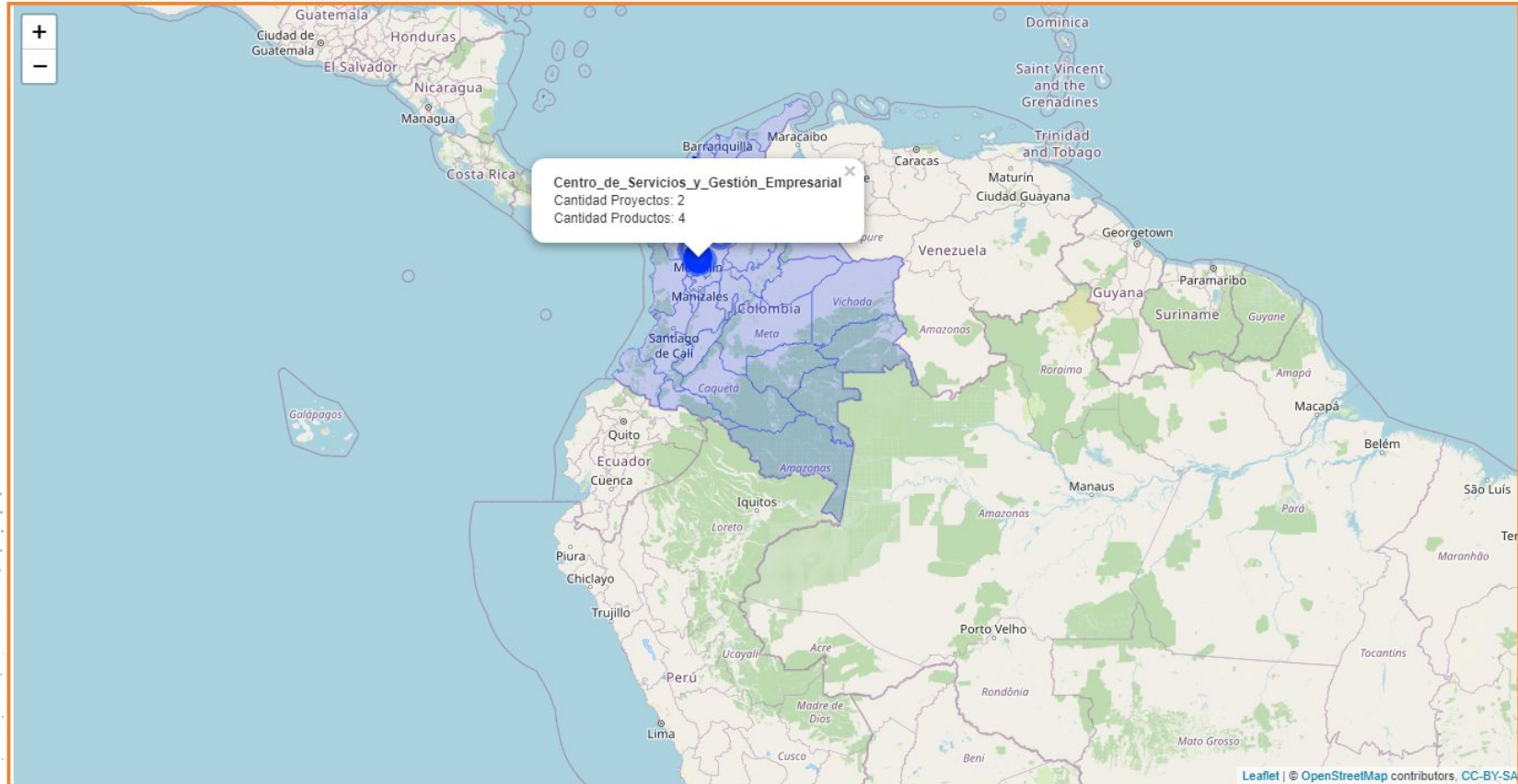


@SENAcomunica

www.sena.edu.co

Visualización de Centros en mapas: Regional Antioquia







5. Modelado de analitica de datos



@SENAcomunica

www.sena.edu.co

Implementación del método de clústering mediante K-means y análisis de componentes principales

1. Se construyen tablas donde se evidencia la productividad en CTel, por centros y por año.
2. Identificación del k óptimo
3. Se aplica a la base de datos analítica de datos, realizando transformación a los datos en los que se identificarán los dendogramas.
4. Para cada dendograma se identifican los respectivos cluster
5. Se aplica algoritmo k-means para identificar aglomerados en función de centros y productos de CTel para el año 2010 inicialmente, de acuerdo a los dendogramas hallados.

Implementación del método de clústering mediante K-means y análisis de componentes principales

6. Se aplica el método de componentes principales se identifica el plano factorial y el círculo de correlación, para identificar clúster formados por centros.
7. Se realiza visualización por mapas de calor para identificar productividad de CTel por centro en el 2010, inicialmente.
8. La validación de lo anterior permitirá evidenciar la diferencia al aplicar el método de clustering mediante series de tiempo.

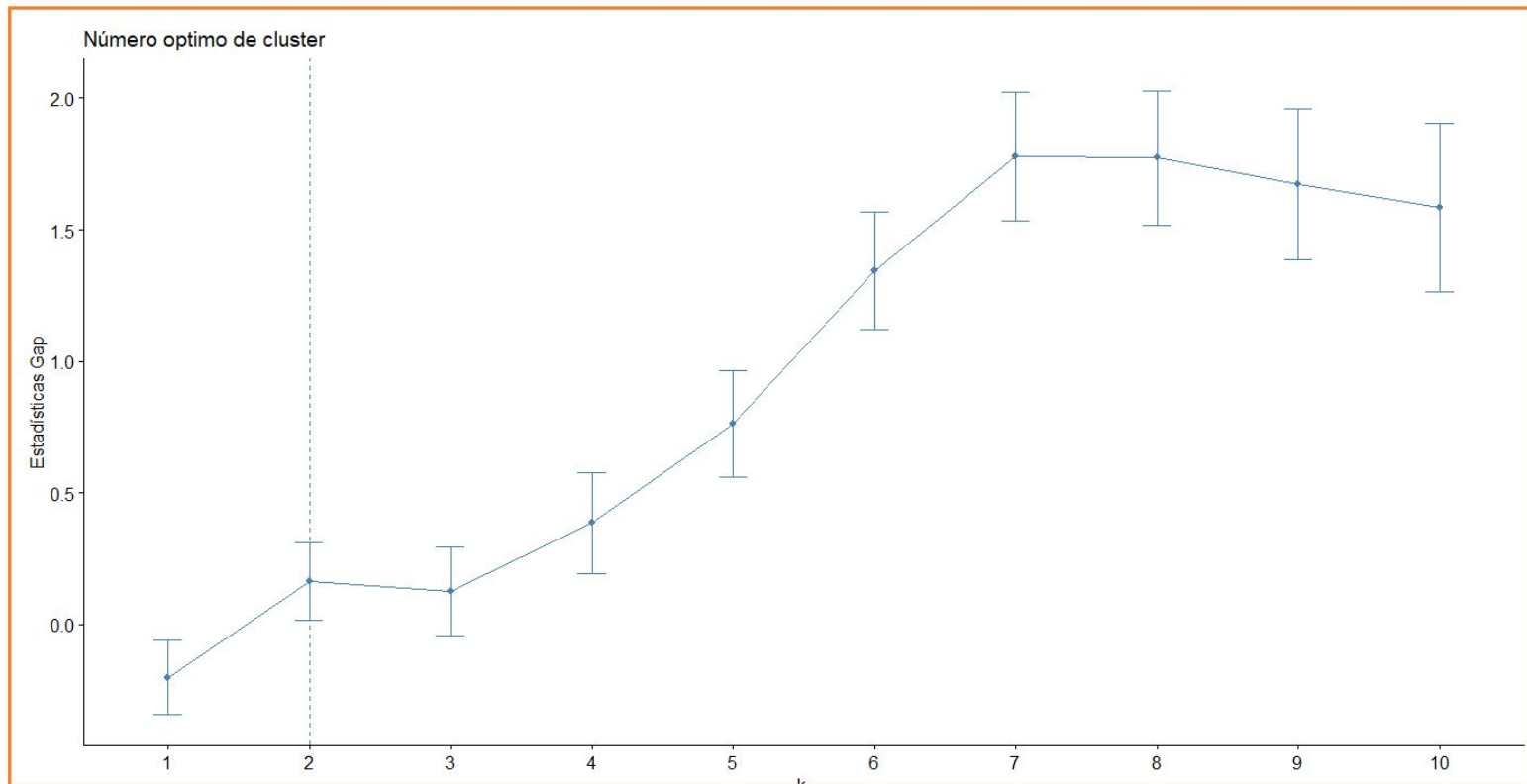
Cantidad de productos de CTel por centros en el 2010



	Codigo_Centro	Cantidad_Proyectos	Fecha_inicio_Proyecto	Cantidad_Productos	Fecha_Registro	P_8020	P_8025	P_8019	P_8039	P_8016	P_8031
1	9101	1	2010	3	2010	1	1	0	0	0	1
2	9127	2	2010	1	2010	0	0	0	0	1	0
3	9201	1	2010	2	2010	0	0	0	0	2	0
4	9202	3	2010	5	2010	2	1	0	0	2	0
5	9203	2	2010	5	2010	1	1	0	1	1	1
6	9204	3	2010	4	2010	0	1	0	0	1	2
7	9205	1	2010	1	2010	0	0	1	0	0	0
8	9206	1	2010	3	2010	1	1	0	0	1	0
9	9301	3	2010	3	2010	0	0	0	2	1	0
10	9401	3	2010	7	2010	2	0	0	0	1	4
11	9402	2	2010	4	2010	0	0	0	1	3	0

Showing 1 to 12 of 16 entries, 11 total columns

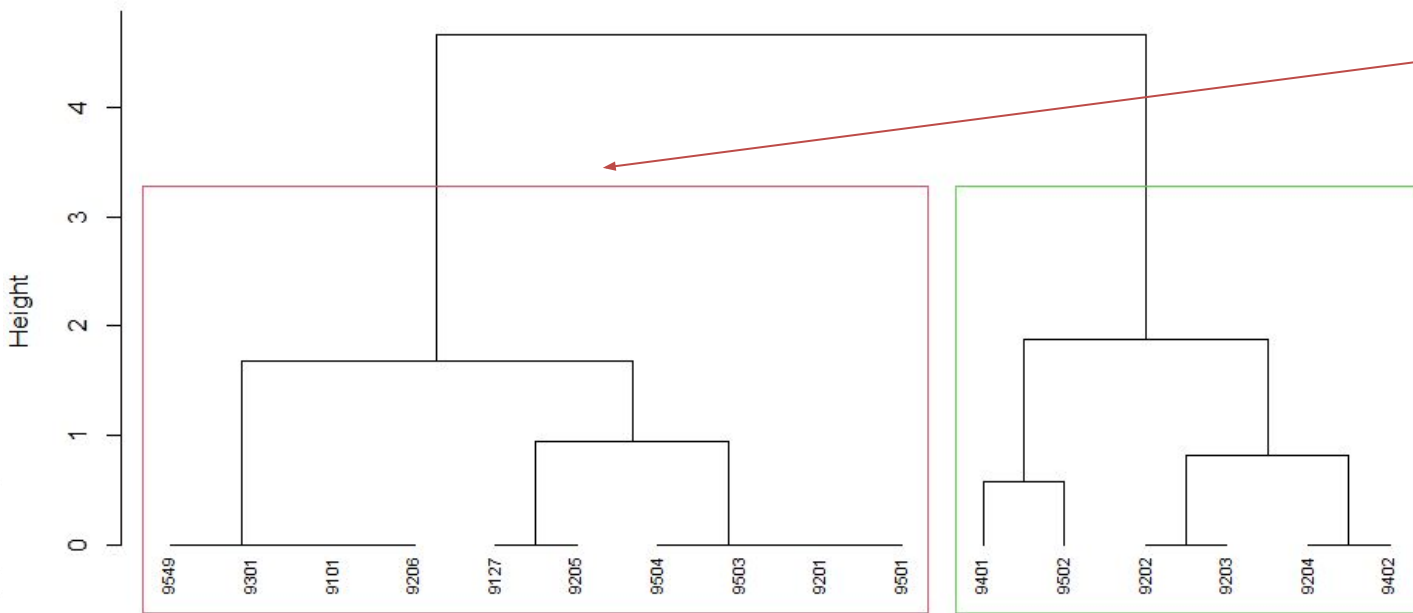
Número óptimo k de clúster mediante Bootstrapping



Dendograma productos por centros en el 2010



Dendograma sobre cantidad de productos por centros en el 2010



Se forman dos
clúster

	Codigo_Centro	Cantidad_Productos
1	9101	-0.1795513
2	9127	-1.3286793
3	9201	-0.7541153
4	9202	0.9695768
5	9203	0.9695768
6	9204	0.3950128
7	9205	-1.3286793
8	9206	-0.1795513
9	9301	-0.1795513
10	9401	2.1187048
11	9402	0.3950128
12	9501	-0.7541153
13	9502	1.5441408
14	9503	-0.7541153
15	9504	-0.7541153
16	9549	-0.1795513

`dist(data$Cantidad_Productos, method = "euclidean")`
`hclust (*, "ward.D2")`

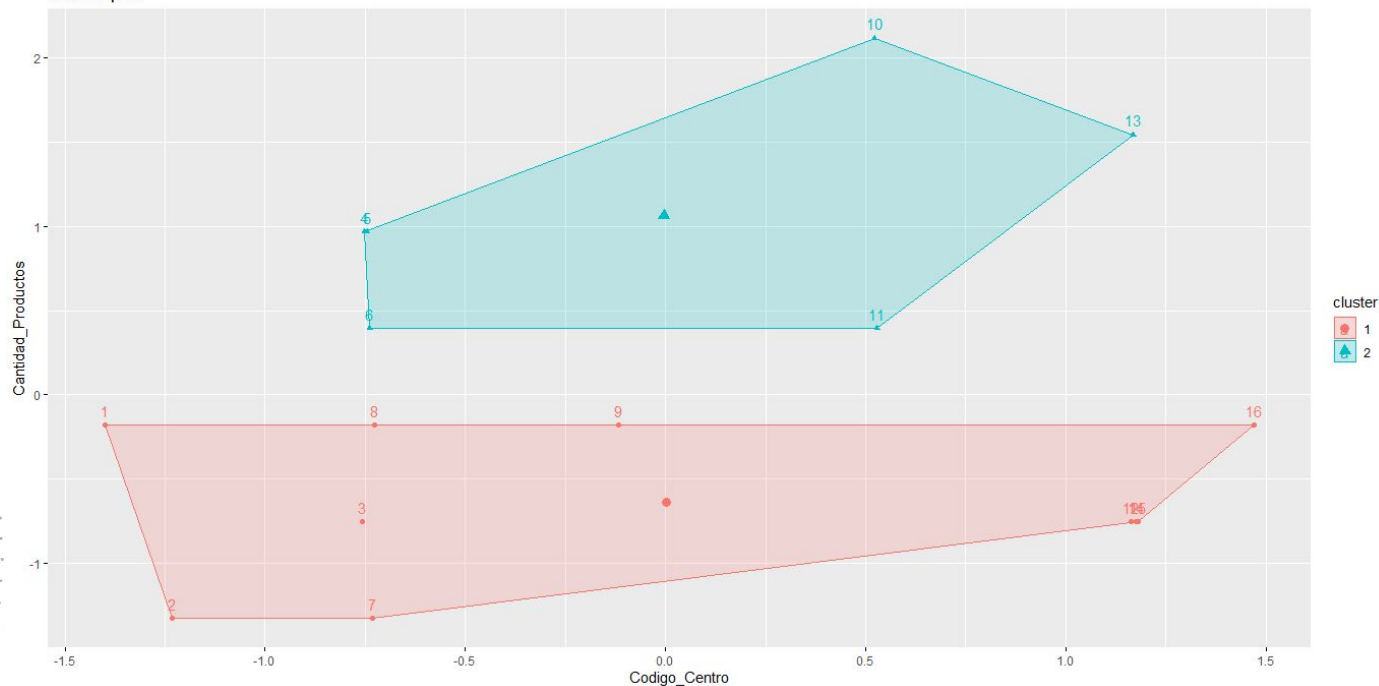
El primero con 10 centros

El segundo con 6 centros

Clustering de centros por productos según dendrograma



Cluster plot

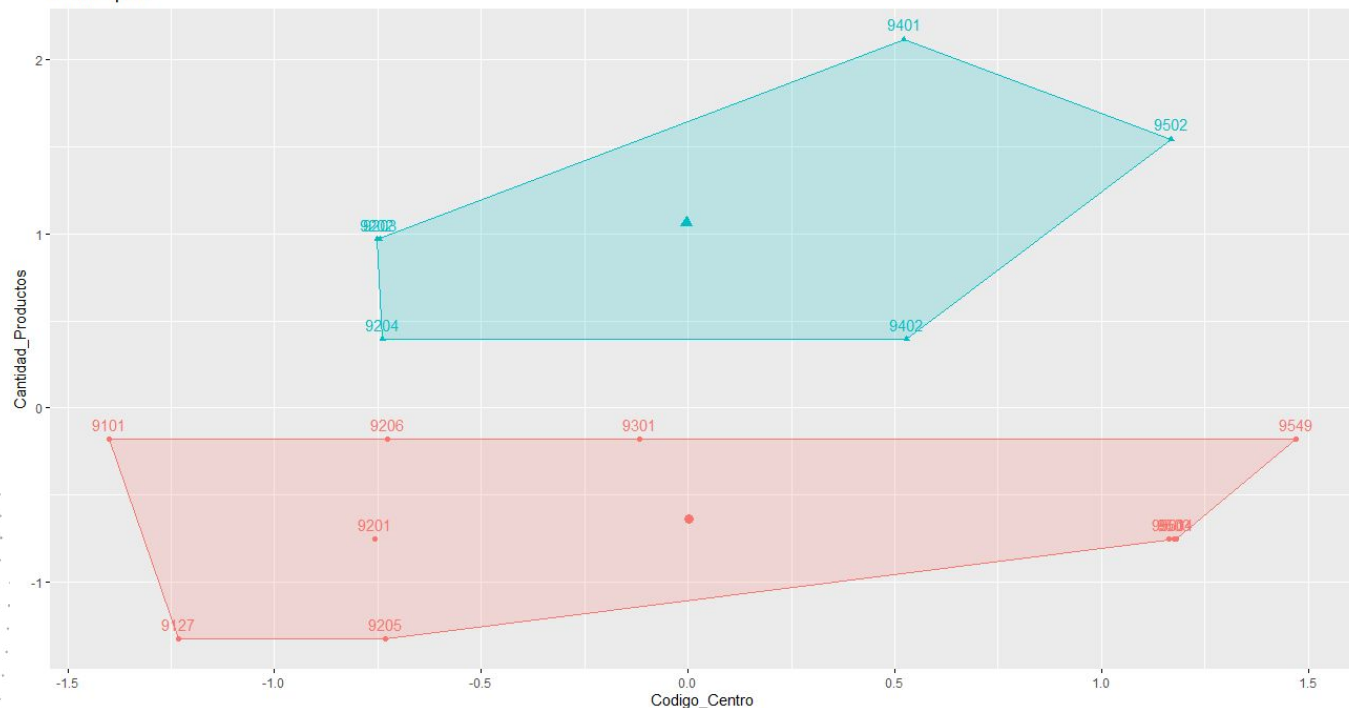


	Codigo_Centro	Cantidad_Productos
1	9101	-0.1795513
2	9127	-1.3286793
3	9201	-0.7541153
4	9202	0.9695768
5	9203	0.9695768
6	9204	0.3950128
7	9205	-1.3286793
8	9206	-0.1795513
9	9301	-0.1795513
10	9401	2.1187048
11	9402	0.3950128
12	9501	-0.7541153
13	9502	1.5441408
14	9503	-0.7541153
15	9504	-0.7541153
16	9549	-0.1795513

Clustering de centros por productos según dendrograma



Cluster plot

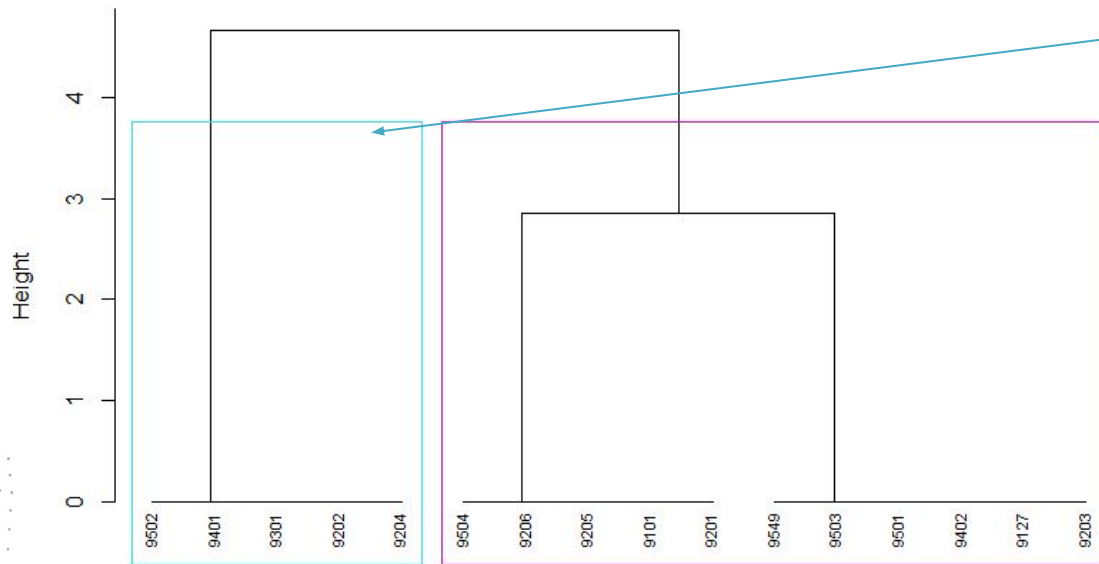


	Codigo_Centro	Cantidad_Productos
1	9101	-0.1795513
2	9127	-1.3286793
3	9201	-0.7541153
4	9202	0.9695768
5	9203	0.9695768
6	9204	0.3950128
7	9205	-1.3286793
8	9206	-0.1795513
9	9301	-0.1795513
10	9401	2.1187048
11	9402	0.3950128
12	9501	-0.7541153
13	9502	1.5441408
14	9503	-0.7541153
15	9504	-0.7541153
16	9549	-0.1795513

Dendrograma proyectos por centros en el 2010



Dendrograma sobre cantidad de proyectos por centros en el 2010



Se forman dos
clúster

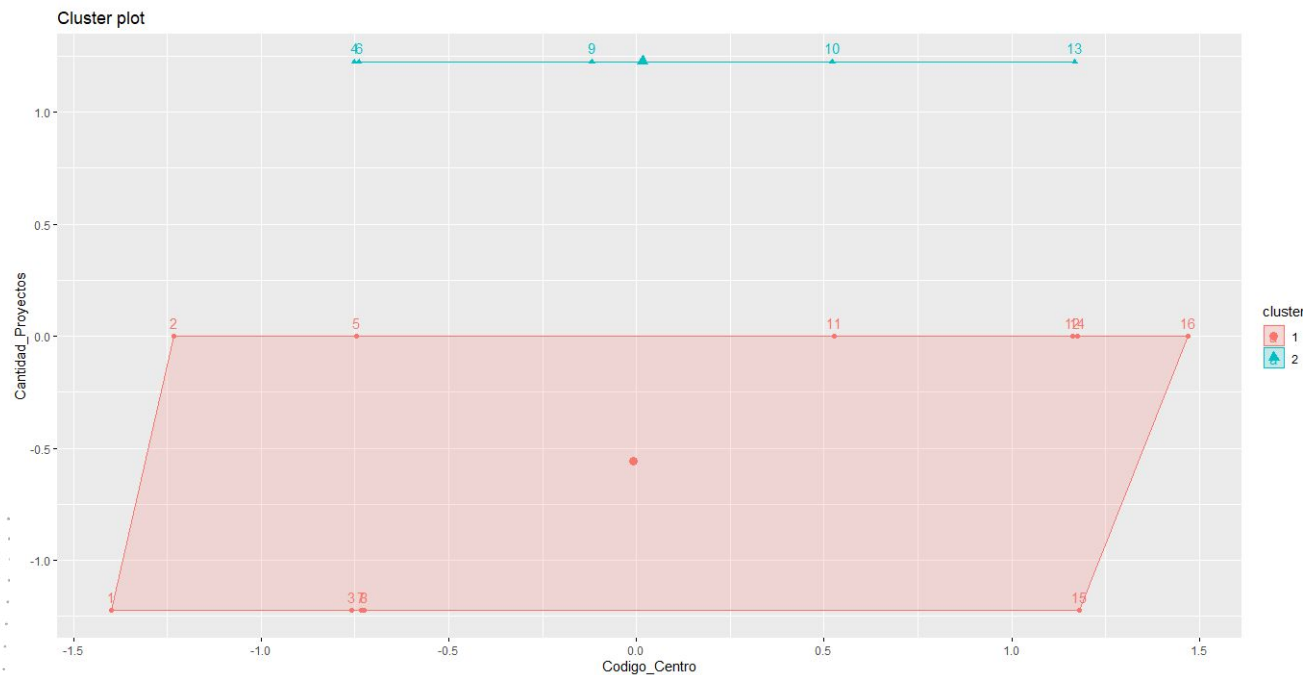
```
dist(data$Cantidad_Proyectos, method = "euclidean")  
hclust (*, "ward.D2")
```

El primero con 5 centros

El segundo con 11 centros

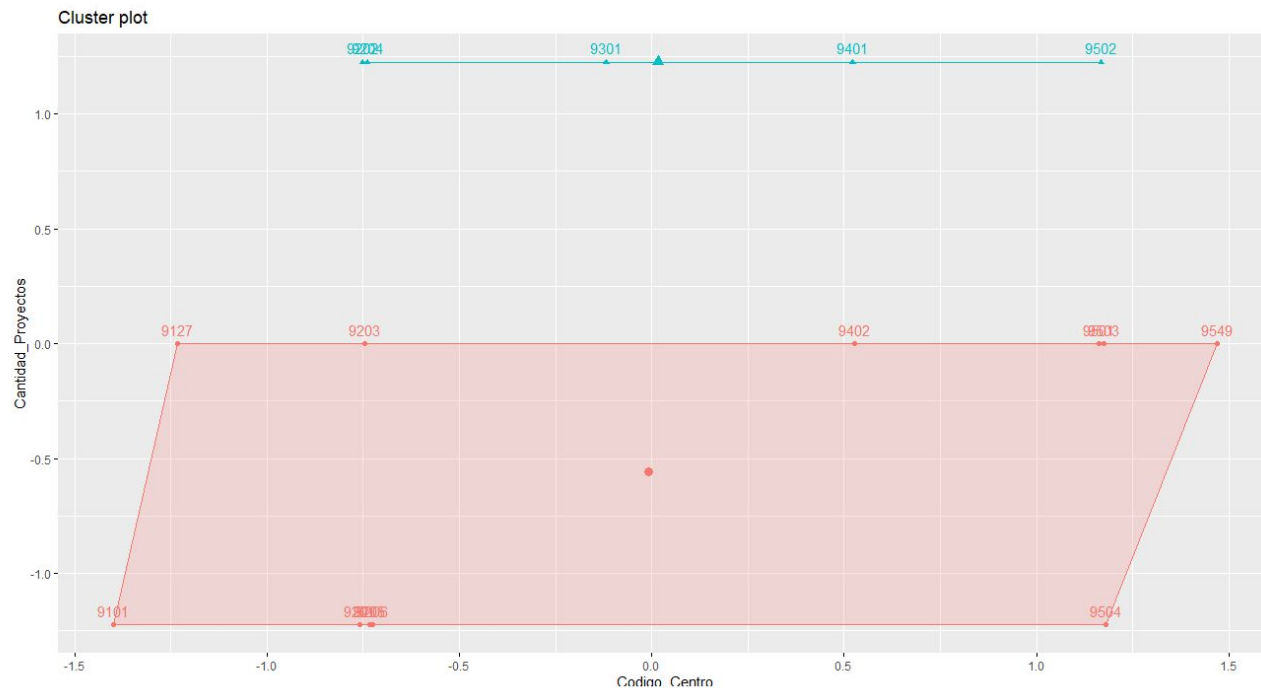
	Codigo_Centro	Cantidad_Proyectos
1	9101	-1.224745
2	9127	0.000000
3	9201	-1.224745
4	9202	1.224745
5	9203	0.000000
6	9204	1.224745
7	9205	-1.224745
8	9206	-1.224745
9	9301	1.224745
10	9401	1.224745
11	9402	0.000000
12	9501	0.000000
13	9502	1.224745
14	9503	0.000000
15	9504	-1.224745
16	9549	0.000000

Clustering de centros por proyectos según dendrograma



	Codigo_Centro	Cantidad_Proyectos
1	9101	-1.224745
2	9127	0.000000
3	9201	-1.224745
4	9202	1.224745
5	9203	0.000000
6	9204	1.224745
7	9205	-1.224745
8	9206	-1.224745
9	9301	1.224745
10	9401	1.224745
11	9402	0.000000
12	9501	0.000000
13	9502	1.224745
14	9503	0.000000
15	9504	-1.224745
16	9549	0.000000

Clustering de centros por proyectos según dendrograma



cluster



	Codigo_Centro	Cantidad_Proyectos
1	9101	-1.224745
2	9127	0.000000
3	9201	-1.224745
4	9202	1.224745
5	9203	0.000000
6	9204	1.224745
7	9205	-1.224745
8	9206	-1.224745
9	9301	1.224745
10	9401	1.224745
11	9402	0.000000
12	9501	0.000000
13	9502	1.224745
14	9503	0.000000
15	9504	-1.224745
16	9549	0.000000



6. Divulgación



@SENAcomunica

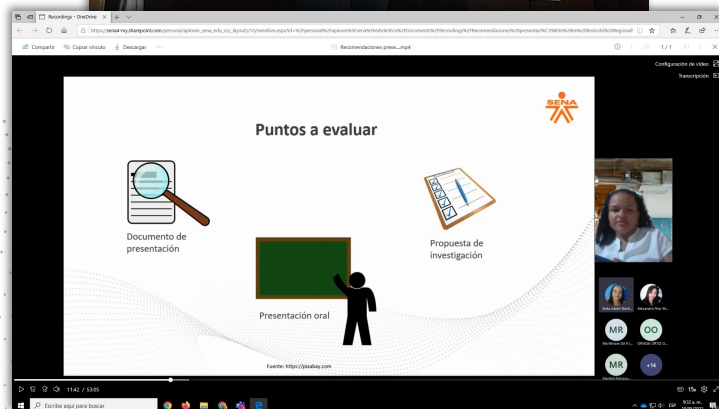
www.sena.edu.co

Divulgación del proyecto

- Gestión de diseño e impresión
- Planeación del artículo de visualización de gráficas en mapas
<https://bit.ly/3w0tDY0>
- Participación en reunión técnica del proyecto y solución de dudas Redcolsi
- Proyecto Promoción de Ecosistemas de Innovación en América Latina - Componentes C3 OTC (Oficina de Transferencia de Conocimiento y C4 (Gestión de Capital Documental)



Recomendaciones Redcolsi Regional 2022





GRACIAS

Línea de atención al ciudadano: 018000 910270
Línea de atención al empresario: 018000 910682



@SENAcomunica

www.sena.edu.co



ANEXOS 6. Actas de reuniones.

ACTA No. 63		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 03 de mayo de 2022	HORA INICIO: 11:00 a.m.	HORA FIN: 1:30 p.m.
LUGAR Y/O ENLACE: Laboratorio de Inteligencia Artificial Piso 5	DIRECCIÓN GENERAL / REGIONAL / CENTRO Centro de Servicios y Gestión Empresarial	
AGENDA O PUNTOS PARA DESARROLLAR: 1. Simulacro de presentación en Redcolsi Regional 2. Avances de los modelos de analítica de datos 3. Avances de actividades de divulgación		
OBJETIVO(S) DE LA REUNIÓN: Realizar seguimiento a la ejecución de las actividades del Proyecto SGPS 9748 Prototipo Innovación		
DESARROLLO DE LA REUNIÓN		
<p>1. Simulacro de presentación en Redcolsi Regional</p> <p>Para iniciar la reunión la aprendiz Karen Estrada realizó la presentación de la exposición como si estuviera en el evento. Luego de esto surgió la recomendación de instalar el aplicativo de recolección de información para mostrarlo en Redcolsi Regional, para los avances de la analítica de datos se mencionó que se pueden utilizar los archivos html para presentar los avances, además se pueden agregar imágenes de los gráficos en póster.</p> <p>2. Avances de los modelos de analítica de datos</p> <p>En cuanto a los modelos de analítica se mencionó que se está trabajando en los pasos de preparación y modelamiento de datos de la Metodología CRISP – DM, en la que se está ajustando la base de datos mediante un algoritmo aleatorio con restricciones que permita tener la información lo más parecido a la realidad posible. Se mencionó, además, que se requiere el apoyo del investigador John Ferla para analizar con su orientación el modelo relacional y algunos otros asuntos relacionados con el desarrollo del prototipo y la integración de las gráficas.</p> <p>Por otro lado, se presentó el cronograma con las grandes tareas acordadas al inicio del año, en él se evidencia las dos primeras actividades completadas y el porcentaje de avance que se lleva hasta el momento.</p>		

2. Cronograma 2022 - Proyecto SGPS-9748



	Fechas programadas	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Fase 1 : Avance de la metodología del Big Data	3 marzo	100%						
Fase 2 : Identificación y selección del Dataset análogo al trabajo en particular	15 abril		100 %					
Fase 3: Caracterizar y seleccionar modelos analítica	19 mayo			45%				
Fase 4: Aplicación de API's para vincular con el desarrollo web	20 junio							
Fase 5 : Puesta en marcha del desarrollo web con los modelos de segmentación y clusterización	21 julio							
Entrega final	15 agosto							

Se está trabajando en la actualización y exploración de las bases de datos utilizando modelos supervisados y no supervisado, lo que se pretende es tener varias opciones para que en el momento que se trabaje con la información real se pueda adaptar a las necesidades que se identifiquen. Se mencionó que para el análisis de los datos se considera pertinente realizarlo por año, ya que normalmente esa es la vigencia de los proyectos y que la comparación se realice por centro para poder obtener como resultado un análisis comparativo de productividad en Ciencia, Tecnología e Innovación de cada centro de formación.

Igualmente, se presentaron los avances en la visualización de la información utilizando mapas, KNN, Cluster y PCA, en los mapas se pueden observar análisis estadísticos

3. Avances de actividades de divulgación

En cuanto a las actividades de divulgación se está gestionando el mejoramiento del diseño del póster para Redcolsi Regional y en la preparación de la aprendiz que representará el proyecto en el evento.

Por otra parte, la investigadora Erika Restrepo orientó la sesión que le correspondió del curso de Formulación de Proyectos organizado por el Hub de Innovación y mencionó que está trabajando en la estructura del informe de investigación para iniciar con su escritura.

También, se acordó que la investigadora Erika Restrepo con el apoyo del equipo de trabajo planteará un artículo relacionado con técnicas de visualización de datos en mapas. EL equipo de trabajo debe identificar una revista que permita publicar de manera rápida el artículo.

Por último, se mencionó que con el fin de solicitar la adición del contrato de los investigadores técnicos se planteará una propuesta que incluya proyectos conexos de Inteligencia Artificial, divulgación técnica del prototipo y asesoramiento en el montaje de datos y recopilación de información real. Dicha propuesta debe incluir justificación, objetivos, y resultados esperados.

CONCLUSIONES

Se continúa avanzando con la aplicación de la metodología CRISP – DM para diseñar el modelo de analítica de datos y la exploración de los modelos supervisados y no supervisados.

Se está apoyando la participación en Redcolsi Regional ya que es un evento que permite la divulgación con integrantes de otras instituciones educativas de la región.

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales se debe garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento. El Servicio Nacional de Aprendizaje SENA solicita la siguiente clasificación de la información:

La información de este documento se debe clasificar como:

PÚBLICA ☐ **PRIVADO** ☐ **SEMIPRIVADO** ☐ **SENSIBLE** ☐

Nota: Antes de contestar esta informacional por favor remitirse al instructivo

COMPROMISOS

ACTIVIDAD	RESPONSABLE	FECHA
Plantear propuestas de artículo de visualización de información	Erika Restrepo	10 de mayo de 2022
Continuar explorando modelos de analítica supervisados y no supervisados	Jesús Arbey Benavides Paola Buitrago	10 de mayo de 2022

ASISTENTES: (Incorporar registro de asistencia)

The screenshot shows a Google Slides presentation titled "SEGUIMIENTO DE ACTIVIDADES". The main slide displays a line chart titled "Serie temporal Costos de inversión - Valor medio". The chart has two data series: "Fecha Cierre Proyecto" (represented by blue vertical bars) and "Valor Medio de Proyecto Centro (S)" (represented by a red line). The x-axis represents time, and the y-axis represents the value. The presentation is being viewed in a browser window with a video player interface at the bottom. On the right side, there is a vertical list of participants with their names and initials: PC, PAOLA BUITRAGO C., and KM.



GD-F-007 V03



REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No- DEL DÍA 03 DEL MES DE Mayo DEL AÑO 2022

OBJETIVO (S)										
No.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PAR VIRT
1	John Jairo Castro M.	88267752	X			SENA CESGE	jcastrom@sena.edu.co	42850	Si	[Firma]
2	Paula Hurtado	43641519		X		SENA	paula.hurtado@betoneur.com		Si	[Firma]
3	Jesus Arbey Jaberayides	1085260713		X		SENA	jaberayides@sen.edu.co		Si	[Firma]
4	Enika Juliette Restrepo	43973982		X		SENA CESGE	erestrepoz@sen.edu.co		Si	[Firma]

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales, el Servicio Nacional de Aprendizaje SENA, se compromete a garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento, y les dará el tratamiento correspondiente en cumplimiento de lo establecido legalmente.

ACTA No. 64		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 04 de mayo de 2022	HORA INICIO: 11:00 a.m.	HORA FIN: 1:30 p.m.
LUGAR Y/O ENLACE: Laboratorio de Inteligencia Artificial Piso 5	DIRECCIÓN GENERAL / REGIONAL / CENTRO Centro de Servicios y Gestión Empresarial	
AGENDA O PUNTOS PARA DESARROLLAR: 1. Estrategias para importación de mapas generados en Python en formato .html en prototipo. 2. Evaluación de modelo relacional con datos aleatorios usados en la analítica de datos. 3. Explicación de los elementos de analítica de datos que se han utilizado hasta el momento.		
OBJETIVO(S) DE LA REUNIÓN: Realizar reunión técnica y de preparación para Redcolsi Regional del Proyecto SGPS 9748 - Prototipo Innovación		
DESARROLLO DE LA REUNIÓN		
1. Estrategias para importación de mapas generados en Python en formato .html en prototipo Se acordó que el investigador Arbey Benavides envíe los siguientes mapas que se han generado hasta el momento en formato .html al desarrollador John Ferla: <ul style="list-style-type: none">• Mapa de costos de proyectos por centros de formación• Mapa de cantidad de productos por centros de formación• Mapa de cantidad de proyectos por centros de formación• Mapa de clustering de centros de investigación en torno a las variables cantidad de productos, cantidad de proyectos, latitud y longitud de centros de formación.		
2. Evaluación de modelo relacional con datos aleatorios usados en la analítica de datos Base de datos utilizada para verificación y actualización de modelo relacional.		
3. Explicación de los elementos de analítica de datos que se han utilizado hasta el momento. Se explica a la aprendiz Karen Estrada los métodos de visualización de avances de proyecto y la interpretación de resultados preliminares de la investigación, desde el enfoque estadístico y el <i>machine learning</i> enfocándose en lo siguiente: <ul style="list-style-type: none">a. Estadística descriptiva: Presentación de mapas como estrategia visual en presentación de posterb. Modelos analíticos: Presentación de mapas, así como de gráficas de calor como estrategia visual en presentación de poster El investigador Arbey Benavides que da con el compromiso de enviar en formato .html a la semillerista Karen: <ul style="list-style-type: none">a. Mapa de costos de proyectos por centros de formaciónb. Mapa de cantidad de productos por centros de formaciónc. Mapa de cantidad de proyectos por centros de formaciónd. Mapa de clustering de centros de investigación en torno a las variables cantidad de productos, cantidad de proyectos, latitud y longitud de centros		

de formación.

La investigadora Entrega de Paola Buitrago enviará las gráficas de clusterización de grupos de investigación en formato .png a Karen Estrada.

CONCLUSIONES

Se continúa avanzando con la aplicación de la metodología CRISP – DM para diseñar el modelo de analítica de datos y la exploración de los modelos supervisados y no supervisados.

Se está apoyando la participación en Redcolsi Regional ya que es un evento que permite la divulgación con integrantes de otras instituciones educativas de la región.

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales se debe garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento. El Servicio Nacional de Aprendizaje SENA solicita la siguiente clasificación de la información:

La información de este documento se debe clasificar como:

PÚBLICA ☐ **PRIVADO** ☐ **SEMIPRIVADO** ☐ **SENSIBLE** ☐

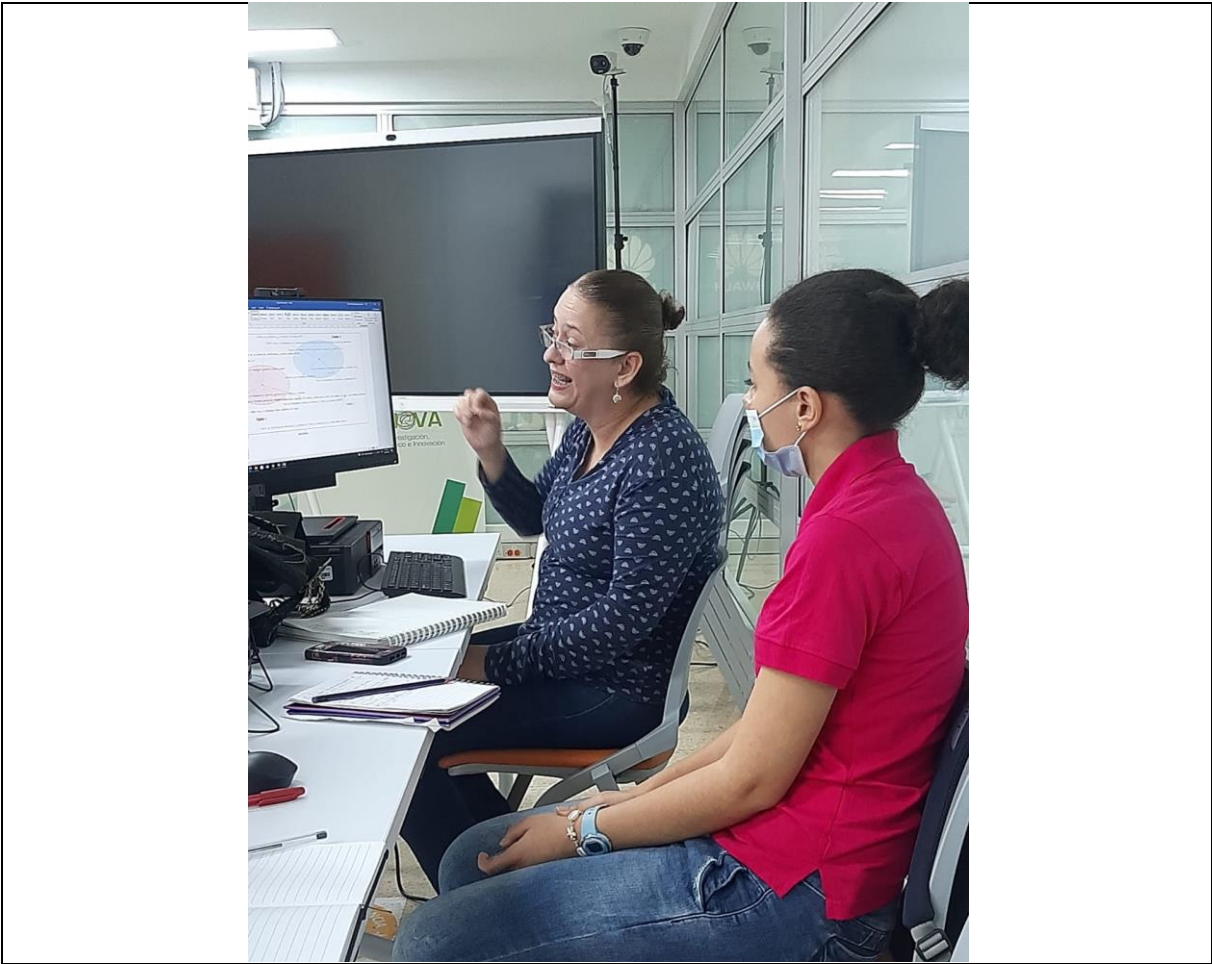
Nota: Antes de contestar esta informacional por favor remitirse al instructivo

COMPROMISOS

ACTIVIDAD	RESPONSABLE	FECHA
Plantear propuestas de artículo de visualización de información	Erika Restrepo	10 de mayo de 2022
Enviar información requerida al desarrollador John Ferla	Jesús Arbey Benavides Paola Buitrago	10 de mayo de 2022

ASISTENTES: (Incorporar registro de asistencia)







REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No- DEL DÍA 04 DEL MES DE mayo DEL AÑO 2022.

OBJETIVO (S)

No.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPA VIRTUAL
1	Jesús Arbecy Benaides	1085260713		X		SENA	jbenavides@sena.edu.co	3105595765	SI	Jesús Arbecy
2	John Ferla	1036621874		X		SENNOVA	Jferla@sena.edu.co	3146864277	SI	John Ferla
3	Karen Y. Estrada	2001760579			Monitor	SENA	kestrada@sena.edu.co	3145117659	SI	Karen Y. Estrada
4	Erika Julieta Restrepo Zapata	43973982		X		SENA	erestropo@sena.edu.co	3005017253	SI	Erika Restrepo
5	Paula Andrea Buitrago	43641579		X		SENA	paula.buitrago.betancur@gmail.com	3113671040	SI	Paula Buitrago

De acuerdo con La Ley 1581 de 2012, Protección de Datos Personales, el Servicio Nacional de Aprendizaje SENA, se compromete a garantizar la seguridad y protección de los datos personales que se encuentran almacenados en este documento, y les dará el tratamiento correspondiente en cumplimiento de lo establecido legalmente.