



PROCESO GESTIÓN DEL TALENTO HUMANO
FORMATO INFORME MENSUAL EJECUCIÓN CONTRACTUAL

Medellín, 31 de marzo 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 marzo 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	Se realizaron avances en la etapa de comprensión del problema así como en el entendimiento de los datos y preparación de	ANEXO 3. Preprocesamiento de datos en python - Collaboratory ANEXOS 6. Actas de reuniones técnicas



	para solucionar de forma eficiente y eficaz problemas o retos identificados en Big Data y analítica de datos con aplicación en gestión de la innovación	<p>estos siguiendo la metodología CRISP-DM.</p> <p>Organización de datasets, variables y modelo relacional, para determinación de variables críticas.</p> <p>Inicio de etapa de preprocesamiento de datos desde python -colab.</p>	<p>ANEXO 3. Preprocesamiento de datos en python - Collaboratory</p> <p>ANEXOS 6. Actas de reuniones técnicas</p> <p>ANEXO 3. Preprocesamiento de datos en python - Collaboratory</p> <p>ANEXOS 6. Actas de reuniones técnicas</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 instaló la siguiente lista de software para el desarrollo del proyecto por medio de solicitud de mesa de servicios. SQL-Server; SQL;Management Studio; Node js; Java Runtime ; Git ;Anaconda;Rstudio;R; Python.	Correo institucional
3	Revisar y actualizar indicadores para la gestión de la innovación según objetivos del proyecto	Se hizo revisión de documentación ofrecida por el supervisor/coordinador de indicadores para la gestión de la innovación, proyecto SGPS 9748 desarrollado durante el 2021 .	Correo institucional



		Se revisa los modelos de indicadores y de evaluación	
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	Reunión para evaluación de modelo de indicadores para 2021 y acople con modelo relacional.	ANEXO 2. Acta de reunión de planeación técnica del desarrollo del prototipo de innovación proyecto SGPS 9748 ANEXO 4. Reunión Seguimiento Proyecto SGPS 9748 Prototipo Innovación - Consultas Modelo ANEXOS 6. Actas de reuniones técnicas
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 elaboraron 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 marzo 2022
7	Divulgar en eventos académicos/empresariales los resultados del proyecto.	Se identificó un (1) evento de divulgación de ciencia, tecnología e innovación para aprendices de semilleros.	ANEXO 1. Medios de divulgación tentativos en evaluación



8	Redactar el informe de investigación y de resultados del proyecto.	<p>Se elaboró un (1) informe de pago mensual GTH-F-062 V08.</p> <p>Se elaboraron diapositivas en presentación de informes técnicos para el supervisor y coordinador del contrato del mes en curso.</p>	<p>Informe en formato pdf.</p> <p>ANEXOS 5. Presentación de avances técnicos marzo 2022</p>
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 revisó la Metodología de la etapa del proyecto CRISP-DM proyecto SGPS 9748 para el 2021.	ANEXO 4. Reunión Seguimiento Proyecto SGPS 9748 Prototipo Innovación - Consultas Modelo
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	Participación en evento virtual "Ciclo de webinars sobre inteligencia artificial" ofrecido entre Sena y Huawei.	Evidencia en formato de participación: https://isurvey.huawei.com/survey/#/question-page?surveyId=33161&customerId=0&placementId=0&channel=0
11	Apoyar la evaluación técnica, administrativa y presupuestal de los proyectos de investigación aplicada, desarrollo tecnológico e innovación que se formulen por el	Gestión por realizar	Gestión por realizar



	SENA para el desarrollo de actividades misionales, acorde con las indicaciones dadas por la Dirección General.		
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.57071092 de la planilla, miplanilla correspondiente al mes de febrero 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 5° 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 100 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 tentativos en evaluación

Expociencia 2022. Inscripción hasta el 15 de marzo de 2022.

Ibagué 20, 21 y 22 de Abril de 2022.

Organiza: AvanCiencia, Universidad de Ibagué, Minciencias, Sena y la Gobernación del Tolima.

<https://www.expociencia.co>

The screenshot displays the Expociencia 2022 website. The top navigation bar includes links for Home, Nosotros, Inscripción de Proyectos, Red de Tutores, and Exposiciones. The main banner features a large orange and green background with a woman sitting on a couch using a laptop. The text on the banner reads: "¿Interesado en inscribir tus proyectos para participar en Expociencia Tolima 2022? Insíbete hasta el 15 de marzo del 2022. ¡Apresúrate!". Below this, there are logos for the organizing institutions: AvanCiencia, Universidad de Ibagué, Minciencias, Sena, and the Gobernación del Tolima. A table shows the event dates as April 20, 21, and 22, 2022, in Ibagué. Another table indicates 2 months, 9 days, and 4531 visitors. The section "¿QUÉ VAS A ENCONTRAR?" lists four scenarios: ExpoInfantil y ExpoJuvenil, ExpoInstitucional y ExpoEmpresarial, Divulgación y Cultura, and Encuentro de Experiencias. Each scenario is accompanied by a DNA helix icon and a brief description of the activities. At the bottom, there are four buttons labeled "Ver más" for each scenario.

Home Nosotros Inscripción de Proyectos Red de Tutores Exposiciones

¿Interesado en inscribir tus proyectos para participar en Expociencia Tolima 2022?

Inscríbete hasta el 15 de marzo del 2022.

¡Apresúrate!

Organizan: AvanCiencia, Universidad de Ibagué, Minciencias, Sena, Gobernación del Tolima

Apoyan: [Logos of supporting organizations]

Ibagué
20, 21 y 22 de Abril de 2022

Meses	Días	Visitantes
2	9	4531

¿QUÉ VAS A ENCONTRAR?

Cuatro escenarios donde te acercaras a la ciencia y la tecnología, construirás redes y obtendrás grandes aprendizajes.

EXPOINFANTIL Y EXPOJUVENIL

Escenario dirigido a niños, jóvenes y semilleros de investigación escolar, con múltiples espacios de exhibición y participación.

[Ver más](#)

EXPOINSTITUCIONAL Y EXPOEMPRESARIAL

Escenario que materializa el compromiso con la ciencia, la tecnología y la innovación, por parte de universidades, centros de investigación, instituciones gubernamentales y sector productivo tanto público como privado.

[Ver más](#)

DIVULGACIÓN Y CULTURA

Escenario de apropiación social del conocimiento, donde conocerás de la mano de científicos regionales, nacionales e internacionales, sus experiencias y resultados en investigación, desarrollo e innovación.

[Ver más](#)

ENCUENTRO DE EXPERIENCIAS

Escenario donde encontrarás talleres y actividades lúdicas realizadas por universidades y organizaciones relacionadas a la ciencia, tecnología e innovación.

[Ver más](#)



ANEXO 2. Acta de reunión de planeación técnica del desarrollo del prototipo de innovación proyecto SGPS 9748

ACTA No. 025		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 04 de marzo de 2022	HORA INICIO: 9:00 a.m.	HORA FIN: 10:00 a.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. Análisis del estado actual del prototipo 2. Estrategias de visualización alternas a Power Bi 3. Roles de trabajo en el prototipo		
OBJETIVO(S) DE LA REUNIÓN: Realizar reunión de planeación técnica del desarrollo del prototipo de innovación proyecto SGPS 9748		
DESARROLLO DE LA REUNIÓN		



1. Análisis del estado actual del prototipo

Se observa el estado actual del prototipo desarrollado en la vigencia 2021. Se evalúan las diferentes etapas relevantes para la continuidad en la vigencia actual. Como primer paso, se evalúa la base de datos generada a partir del modelo relacional. Con respecto a esto, el ingeniero John Ferla menciona la importancia de determinar las variables críticas que determinarán las características del diseño del mockup y dashboard del prototipo trabajado en 2021.

2. Estrategias de visualización alternas a Power Bi

Se plantea la posibilidad de nuevas estrategias de visualización por medio de librerías de python:

Plotly Express: Es un módulo que contiene funciones que pueden crear figuras dinámicas que permiten una óptima visualización para la representación de modelos desde una perspectiva de estadística descriptiva e inferencial en pro de los resultados obtenidos desde el machine learning.

<https://plotly.com/python/plotly-express/>

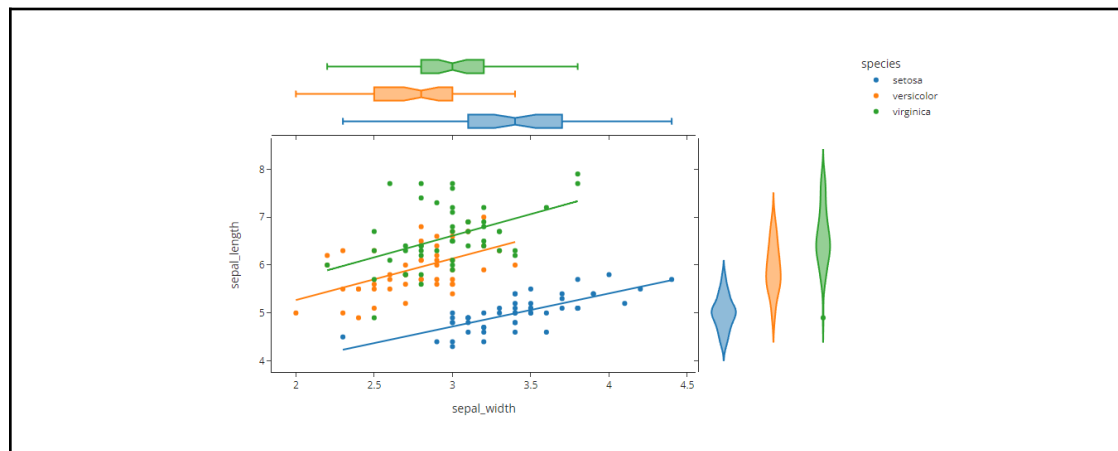


Figura 1. Gráfica ejemplo de visualización de datos.

Folium: Es una librería de python ampliamente utilizada para la visualización de datos en zonas geográficas específicas. En el futuro permitirá dar una forma visual de métodos de clustering para centros respectivos de investigación.

<https://www.cursosgis.com/visualizando-datos-en-un-mapa-con-el-paquete-de-python-folium/>

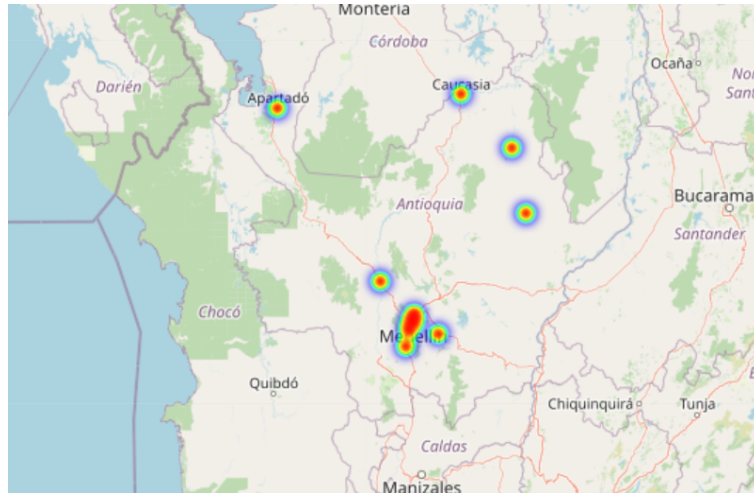


Figura 2. Mapa de representación de centros de investigación en el departamento de Antioquia

3. Roles de trabajo en el prototipo

Se determinan los roles de trabajo respectivos a la parte de analítica de datos así como del backend y frontend del prototipo de innovación. Para ello el ingeniero Ferla se compromete en la revisión de los prototipos anteriormente desarrollados y en una nueva propuesta adecuada a las necesidades actuales del proyecto. Por otro lado, desde el equipo de analítica conformado por Paola Buitrago y Jesus Benavides se plantea una idea posibles librerías como soporte en paralelo con PowerBi.

CONCLUSIONES

- Se observa la necesidad la importancia de la evaluación de la base datos registrada en pasados periodos, así como la generación de datos sintéticos que simulan de manera adecuada modelos analíticos en pro de cuantificar los indicadores definidos desde una perspectiva teórica, así como las variables críticas que definirán el desarrollo de una propuesta en modelos de machine learning.
- Se discutió posibles alternativas a PowerBi que permitan aumentar el rendimiento entre la etapa de software con la parte analítica, por medio de librerías que permiten generar visualizaciones dinámicas tales como Plotly Express y Folium de python.
- Se definen roles, entre los que se acopla el trabajo de el ingeniero John Ferla en lo respectivo al desarrollo de plataforma y dashboard en paralelo con el trabajo por parte de los investigadores Paola Buitrago y Jesus Benavides.



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
	Erika Restrepo	04 03 2022
	Paola Buitrago	04 03 2022



	John Ferla	04 03 2022
	Jesus Arbey Benavides Guevara	04 03 2022
ASISTENTES: (Incorporar registro de asistencia)		



ANEXO 3. Preprocesamiento de datos en python - Collaboratory

Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

▾ Proyecto SGPS 9748 - Prototipo Innovación

Modelo de evaluación y valoración de la Gestión de CTel para los Centros SENA Regional Antioquia aplicando analítica de datos 23 de febrero del 2022

▾ Versión Python

```
[ ] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import branca
import plotly.express as px
#Nube de Palabras
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
#mapas
import folium
from folium.plugins import HeatMap

[ ] # upload de archivo desde Drive
from google.colab import files
path = "/content/drive/MyDrive/Proyecto SGPS 9748/v2_Dataset_FINAL_Python.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')
principal_talento_humano = pd.read_excel(path,sheet_name= 'Principal_Talento_Humano')
estado_proyecto = pd.read_excel(path,sheet_name= 'Estado_Proyecto')
nivel_academico = pd.read_excel(path,sheet_name= 'Nivel_Academico')
```



Python_SGPS 9748 _Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo



+ Código + Texto



Descripción

tipologia_productos

Describe la tipología de los productos clasificados por Sena

```
[ ] tipologia_productos.head()
```

	_Codigo_Tipologia	_Nombre_Tipologia
0	7500.0	2.2.1.1 Artículos de investigación A1, A2, B y...
1	7501.0	2.2.1.1 Artículos de investigación D.
2	7502.0	2.2.1.2 Notas científicas.
3	7503.0	2.2.1.3 Libros resultados de investigación.
4	7504.0	2.2.1.4 Capítulos en libro res...

```
[ ] tipologia_productos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   _Codigo_Tipologia     30 non-null    float64
1   _Nombre_Tipologia     30 non-null    object
```



Python_SGPS 9748 _Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
[ ] <class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   _Codigo_Tipologia      30 non-null    float64
1   _Nombre_Tipologia      30 non-null    object
dtypes: float64(1), object(1)
memory usage: 608.0+ bytes
```

```
[ ] tipologia_productos.describe()
```

_Codigo_Tipologia	
count	30.000000
mean	7514.500000
std	8.803408
min	7500.000000
25%	7507.250000
50%	7514.500000
75%	7521.750000
max	7529.000000

```
[ ] # Verificacion de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(tipologia_productos.notna())
```





```
[ ] # Verificacion de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(tipologia_productos.notna())
```





Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
[ ] # Remoción de duplicados
tipologia_productos.drop_duplicates()
```

	_Codigo_Tipologia	_Nombre_Tipologia
0	7500.0	2.2.1.1 Artículos de investigación A1, A2, B y...
1	7501.0	2.2.1.1 Artículos de investigación D.
2	7502.0	2.2.1.2 Notas científicas.
3	7503.0	2.2.1.3 Libros resultados de investigación.
4	7504.0	2.2.1.4 Capítulos en libro res...
5	7505.0	2.2.3.6.1 Libros de Formación Q1.
6	7506.0	2.2.1.5 Productos tecnológicos patentados ...
7	7507.0	2.2.1.6 Variedades vegetales, nuevas ...
8	7508.0	2.2.1.7 Productos resultados de la creaci...
9	7509.0	2.2.2.1 Productos tecnológicos certifica...
10	7510.0	2.2.2.2 Productos empresariales.
11	7511.0	2.2.2.3 Regulaciones, normas, reglamento...
12	7512.0	Los requerimientos son mencionados en la Tabla...
13	7513.0	2.2.2.4 Conceptos técnicos.
14	7514.0	2.2.2.5 Registros de Acuerdos de licencia...
15	7515.0	2.2.3.3 Procesos de Apropiación Social...
16	7516.0	2.2.3.4 Circulación de conocimiento especializ...
17	7517.0	2.2.3.5 Divulgación Pública de la CTel. Public...
18	7518.0	2.2.3.6 Producción Bibliográfica.
19	7519.0	2.2.4.1 Direcciones de Tesis de doctorado Dire...
20	7520.0	Los requerimientos son mencionados en la Tabla...
21	7521.0	2.2.4.2 Direcciones de Trabajo de grado de mae...
22	7522.0	2.2.4.3 Direcciones de Trabajo de pregrado Dir...
23	7523.0	2.2.4.4 Proyectos de Investigación y Desarroll...
24	7524.0	2.2.4.4 Proyectos de Investigación-Creación Pr...
25	7525.0	2.2.4.4 Proyectos de Investigación, Desarr...
26	7526.0	2.2.4.5 Proyectos de extensión ...



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

22	7522.0	2.2.4.3 Direcciones de Trabajo de pregrado Dir...
23	7523.0	2.2.4.4 Proyectos de Investigación y Desarroll...
24	7524.0	2.2.4.4 Proyectos de Investigación-Creación Pr...
25	7525.0	2.2.4.6 Proyectos de Investigación, Desarr...
26	7526.0	2.2.4.5 Proyectos de extensión ...
27	7527.0	2.2.4.6 Apoyos a la creación de programas y cu...
28	7528.0	2.2.4.7 Acompañamientos y asesorías de li...
29	7529.0	Informe de investigación

```
[ ] #Palabras más mencionadas en tipología de productos
#Stopword list:
stopwords = set(STOPWORDS)
stopwords.update(["de", "en", "la", "son", "Los", "del", "los", "el", "una", "un", "uno", "las", "con", "Tabla",
                  'trabajo', 'productos', 'anexo', 'resultado', 'requerimientos', 'mencionados'])

# Word cloud image

text = ' '.join(tipologia_productos['Nombre_Tipologia'].astype('str'))
wordcloud = WordCloud(width = 1500, height = 1000, stopwords=stopwords, max_font_size=400, max_words=100, background_color="white").generate(text)
plt.figure()
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```



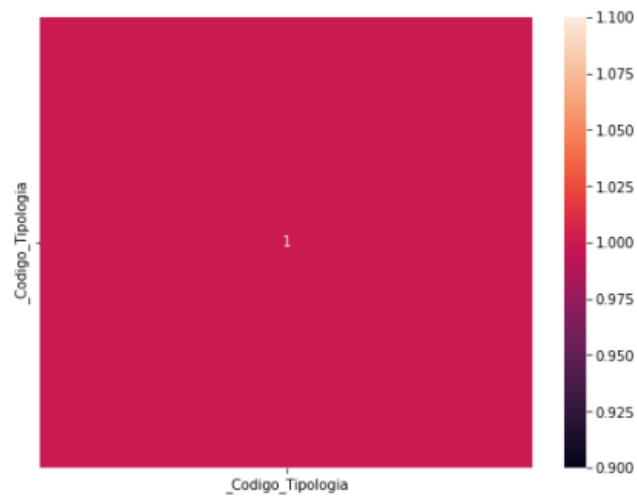


Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Última modificación: 11 de marzo](#)

+ Código + Texto

```
[ ] # Correlación de Pearson
corr_df = tipologia_productos.corr(method='pearson')
plt.figure(figsize=(8, 6))
sns.heatmap(corr_df, annot=True)
plt.show()
```



tipo_contratacion

```
[ ] tipo_contratacion.head()
```

	_Codigo_Tipo_Contrato	_Nombre_Tipo_Contrato
0	2001.0	Aprendiz
1	2002.0	Contratista
2	2003.0	Planta
3	2004.0	Temporal
4	2005.0	Provisional



Python_SGPS 9748 _Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo



+ Código + Texto



tipo_contratacion



```
[ ] tipo_contratacion.head()
```



	_Codigo_Tipo_Contrato	_Nombre_Tipo_Contrato
0	2001.0	Aprendiz
1	2002.0	Contratista
2	2003.0	Planta
3	2004.0	Temporal
4	2005.0	Provisional

```
[ ] tipo_contratacion.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   _Codigo_Tipo_Contrato  5 non-null     float64
1   _Nombre_Tipo_Contrato  5 non-null     object
dtypes: float64(1), object(1)
memory usage: 208.0+ bytes
```

```
[ ] tipo_contratacion.describe()
```

	_Codigo_Tipo_Contrato
count	5.000000
mean	2003.000000
std	1.581139
min	2001.000000
25%	2002.000000
50%	2003.000000
75%	2004.000000
max	2005.000000

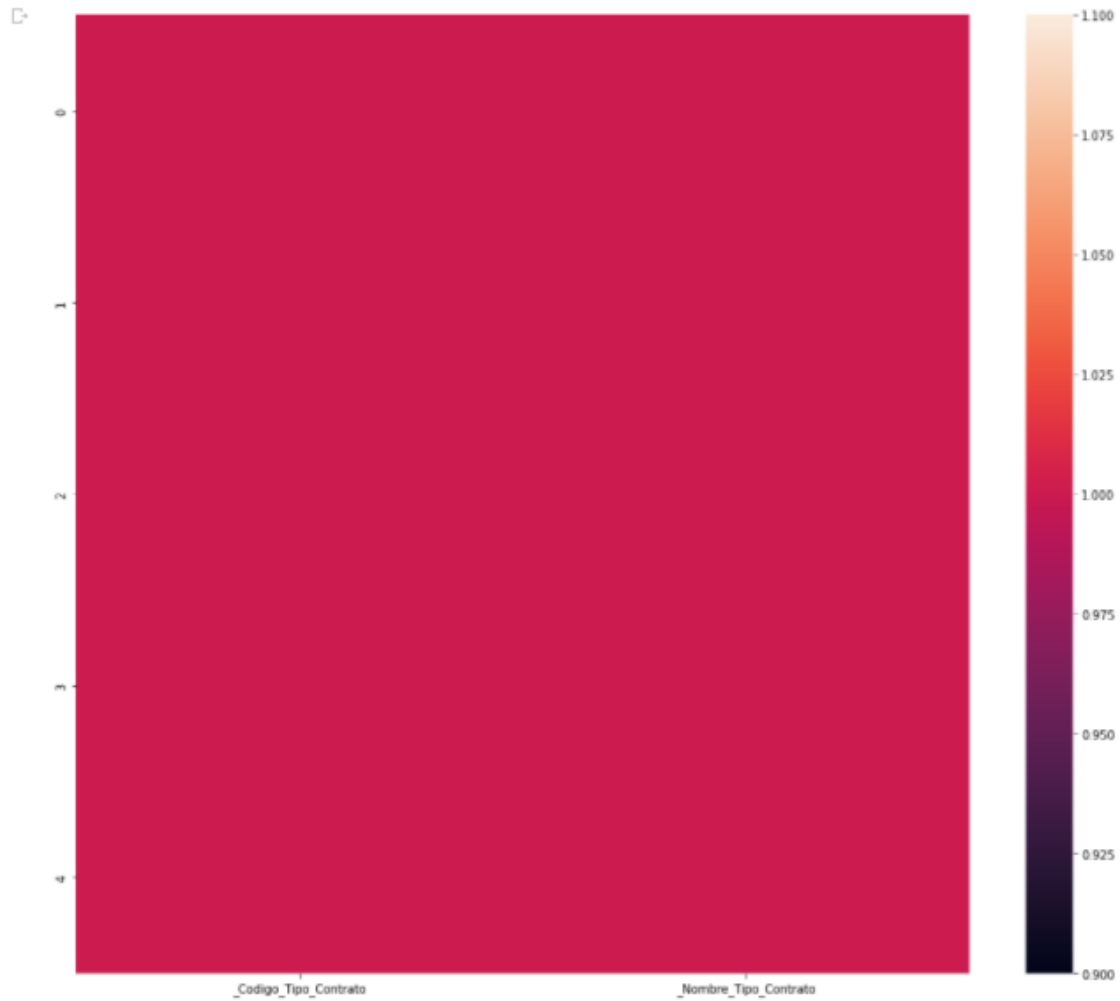


Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
# Verificacion de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(tipo_contratacion.notna())
```



```
[ ] semilleros.head()
```

	_Codigo_Semillero	_Nombre_Semillero	_Codigo_Grupo_Investigacion	_Codigo_Centro	_Codigo_Regional
0	8888.0	REDTIS (Semillero de Investigación en Redes, t...	9999.0	9402.0	5.0
1	8889.0	SIALTA (Semillero de Investigación en Logístic...	9999.0	9402.0	5.0
2	8890.0	SIDIN (Semillero de Investigación en Desarroll...	9999.0	9402.0	5.0
3	8891.0	SIG-SENA (Semillero de investigación en sistem...	9999.0	9402.0	5.0
4	8892.0	SIPAEM (Semillero de Investigación del Program...	9999.0	9402.0	5.0



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

semilleros.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 5 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   _Codigo_Semillero    11 non-null    float64
 1   _Nombre_Semillero    11 non-null    object
 2   _Codigo_Grupo_Investigacion  11 non-null    float64
 3   _Codigo_Centro       11 non-null    float64
 4   _Codigo_Regional     11 non-null    float64
dtypes: float64(4), object(1)
memory usage: 568.0+ bytes
```

#Palabras más mencionadas en tipología de productos

#Stopword list:

stopwords = set(STOPWORDS)

stopwords.update(["de", "en", "la", "son", "Los", "del", "los", "el", "una", "un", "uno", "las", "con", "Semillero"])

Word cloud image

text = ' '.join(semilleros['_Nombre_Semillero'].astype('str'))

wordcloud = WordCloud(width = 1500, height = 1000, stopwords=stopwords, max_font_size=600, max_words=200, background_color="white").generate(text)

plt.figure()

plt.imshow(wordcloud, interpolation="bilinear")

plt.axis("off")

plt.show()



rubro.head()

	_Codigo_Rubro	_Nombre_Rubro
0	23.0	23 - Modernización
1	85.0	85 - Divulgación
2	86.0	86 - Investigación aplicada
3	82.0	82 - Innovación

rubro.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   _Codigo_Rubro  4 non-null    float64
 1   _Nombre_Rubro  4 non-null    object
dtypes: float64(1), object(1)
memory usage: 192.0+ bytes
```



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

rubro.head()

	_Codigo_Rubro	_Nombre_Rubro
0	23.0	23 - Modernización
1	65.0	65 - Divulgación
2	66.0	66 - Investigación aplicada
3	82.0	82 - Innovación

[] rubro.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   _Codigo_Rubro    4 non-null      float64
1   _Nombre_Rubro    4 non-null      object
dtypes: float64(1), object(1)
memory usage: 192.0+ bytes
```

[] rol_sennova.head()

	_Codigo_Rol	_Nombre_Rol
0	1.0	Aprendiz
1	2.0	Semillero
2	3.0	Aprendiz Sennova con Contrato de aprendizaje
3	4.0	Monitor
4	5.0	Dinamizador del conocimiento

[] rol_sennova.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13 entries, 0 to 12
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   _Codigo_Rol     13 non-null     float64
1   _Nombre_Rol     13 non-null     object
dtypes: float64(1), object(1)
memory usage: 336.0+ bytes
```

[] tipo_contratacion.head()

	_Codigo_Tipo_Contrato	_Nombre_Tipo_Contrato
0	2001.0	Aprendiz
1	2002.0	Contratista
2	2003.0	Planta
3	2004.0	Temporal
4	2005.0	Provisional



+ Código + Texto

```
[ ] 3      2004.0      Temporal
    4      2005.0      Provisional
```

```
[ ] tipo_contratacion.info()
```

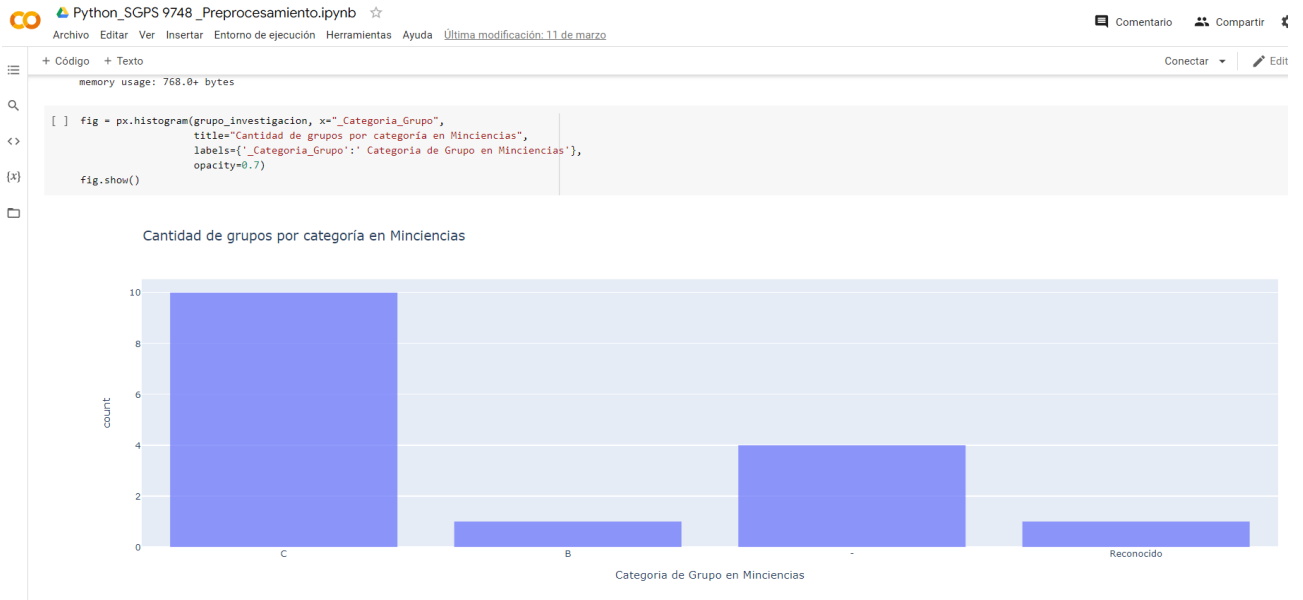
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Codigo_Tipo_Contrato  5 non-null     float64
1   _Nombre_Tipo_Contrato  5 non-null     object
dtypes: float64(1), object(1)
memory usage: 208.0+ bytes
```

```
[ ] grupo_investigacion.head(12)
```

	_Codigo_Grupo_Investigacion	_Nombre_Grupo_Investigacion	_Categoria_Grupo	_Codigo_Centro	_Codigo_Regional
0	COL0085647	Grupo de Investigación La Salada CILS	C	9101.0	5.0
1	COL0106783	Grupo de Investigación en Innovación y Agroind...	B	9503.0	5.0
2	COL0110481	Gestión de la Innovación y el Desarrollo Tecno...	C	9301.0	5.0
3	COL0113017	GACIPE (Grupo de Investigación en Automatiza...	C	9204.0	5.0
4	COL0113975	Materiales para el Mobiliario - MATERMOB	C	9205.0	5.0
5	COL0129409	BIOMATIC (Biomecánica, materiales, TICS, diseñ...	C	9201.0	5.0
6	COL0131024	GIGAT-Grupo de Investigación en gerencia y apl...	C	9402.0	5.0
7	COL0133799	Ciencia, Investigación, Conocimiento, Aprendiz...	-	9502.0	5.0
8	COL0148628	GRUPO DE INVESTIGACIÓN, INNOVACIÓN Y DESARROLL...	C	9504.0	5.0
9	COL0149529	GIDOCA	C	9549.0	5.0
10	COL0149889	Grupo de Investigación red de construcción, in...	-	9203.0	5.0
11	COL0150772	Grupo de Investigación Ciencia, Tecnología e I...	C	9401.0	5.0

```
[ ] grupo_investigacion.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Codigo_Grupo_Investigacion  16 non-null     object
1   _Nombre_Grupo_Investigacion  16 non-null     object
2   _Categoria_Grupo            16 non-null     object
3   _Codigo_Centro              16 non-null     float64
4   _Codigo_Regional            16 non-null     float64
dtypes: float64(2), object(3)
memory usage: 768.0+ bytes
```



Python_SGPS 9748_Preprocesamiento.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

Categoría de 'grupo' en Minciencias

```
[ ] area_conocimiento_ocde.head()
```

	_Codigo_area_OCDE	_Nombre_area_OCDE
0	1.0	CIENCIAS NATURALES Y EXACTAS
1	2.0	INGENIERÍAS Y TECNOLOGÍAS
2	3.0	CIENCIAS MÉDICAS Y DE LA SALUD
3	4.0	CIENCIAS AGRÍCOLAS
4	5.0	CIENCIAS SOCIALES

```
[ ] area_conocimiento_ocde.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  --
0   _Codigo_area_OCDE      6 non-null      float64
1   _Nombre_area_OCDE      6 non-null      object
dtypes: float64(1), object(1)
memory usage: 224.0+ bytes
```

```
[ ] principal_talento_humano.head()
```

	_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
0	3000.0	5009.0	2004.0	NaN	66.0	NaN	1.0	5.0	9101.0	3412.0	8000.0
1	3001.0	5005.0	2002.0	NaN	66.0	NaN	1.0	5.0	9127.0	3412.0	8000.0
2	3002.0	5008.0	2002.0	NaN	66.0	NaN	1.0	5.0	9201.0	3412.0	8000.0
3	3008.0	5005.0	2002.0	NaN	65.0	NaN	1.0	5.0	9202.0	3515.0	8012.0
4	3010.0	5009.0	2003.0	NaN	65.0	NaN	1.0	5.0	9203.0	3515.0	8012.0



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Última modificación: 11 de marzo](#)



+ Código + Texto



```
[ ] principal_talento_humano.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 534 entries, 0 to 533
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   _Codigo_Persona                       534 non-null   float64
1   _Codigo_Nivel_Academico               534 non-null   float64
2   _Codigo_Tipo_Contrato                 534 non-null   float64
3   _Objeto_Contrato                      0 non-null     float64
4   _Codigo_Rubro                         534 non-null   float64
5   _Codigo_Area_OCDE                     0 non-null     float64
6   _Codigo_Estado_Contratacion           534 non-null   float64
7   _Codigo_Regional                      534 non-null   float64
8   _Codigo_Centro                       534 non-null   float64
9   _Codigo_Proyecto                      534 non-null   float64
10  _Codigo_Producto                      534 non-null   float64
dtypes: float64(11)
memory usage: 46.0 KB
```

```
[ ] estado_proyecto.head()
```

	_Codigo_Estado_Proyecto	_Nombre_Estado_Proyecto
0	1.0	Sin iniciar
1	2.0	Ejecución
2	3.0	Terminado
3	4.0	Suspedido

```
[ ] estado_proyecto.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 2 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   _Codigo_Estado_Proyecto               4 non-null     float64
1   _Nombre_Estado_Proyecto               4 non-null     object
dtypes: float64(1), object(1)
memory usage: 192.0+ bytes
```



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo



+ Código + Texto



```
[ ] 2          3.0          Terminado
    3          4.0          Suspedido
```

```
[ ] estado_proyecto.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 2 columns):
#   Column                      Non-Null Count  Dtype
---  ---                      ---
0   _Codigo_Estado_Proyecto      4 non-null     float64
1   _Nombre_Estado_Proyecto      4 non-null     object
dtypes: float64(1), object(1)
memory usage: 192.0+ bytes
```

```
[ ] nivel_academico.head(10)
```

	_Codigo_Nivel_Academico	_Nombre_Nivel_Academico
0	5001.0	Bachiller
1	5002.0	Técnico certificado
2	5003.0	Técnico profesional
3	5004.0	Técnico profesional con especialización
4	5005.0	Tecnólogo
5	5006.0	Tecnólogo con especialización
6	5007.0	Profesional
7	5008.0	Profesional con especialización
8	5009.0	Profesional con maestría
9	5010.0	Profesional con doctorado

```
[ ] nivel_academico.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 2 columns):
#   Column                      Non-Null Count  Dtype
---  ---                      ---
0   _Codigo_Nivel_Academico      10 non-null     float64
1   _Nombre_Nivel_Academico      10 non-null     object
dtypes: float64(1), object(1)
memory usage: 288.0+ bytes
```





Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
[ ] dtypes: float64(1), object(1)
memory usage: 288.0+ bytes
```

```
[ ] municipio.head()
```

	_Codigo_Municipio	_Nombre_Municipio	_Codigo_Departamento
0	1.0	MEDELLIN	232.0
1	2.0	ABEJORRAL	232.0
2	4.0	ABRIAQUI	232.0
3	21.0	ALEJANDRIA	232.0
4	30.0	AMAGA	232.0

```
[ ] municipio.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 125 entries, 0 to 124
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  -
0   _Codigo_Municipio      125 non-null   float64
1   _Nombre_Municipio      125 non-null   object
2   _Codigo_Departamento   125 non-null   float64
dtypes: float64(2), object(1)
memory usage: 3.1+ KB
```

```
[ ] centros_formacion.head(12)
```

	_Codigo_Centro	_Nombre_Centro	_Codigo_Regional	_Codigo_Municipio	_Codigo_Departamento	latitud	longitud
0	9101.0	Centro de los Recursos Naturales Renovables -L...	1.0	129.0	5.0	6.053157	-75.628486
1	9127.0	Centro de Formación Minero Ambiental	1.0	250.0	5.0	7.584361	-74.800660
2	9201.0	Centro del Diseño y Manufactura del Cuero	1.0	360.0	5.0	6.180794	-75.605884
3	9202.0	Centro de Formación en Diseño, Confección y Moda	1.0	360.0	5.0	6.181273	-75.605935
4	9203.0	Centro para el Desarrollo del Hábitat y la Con...	1.0	1.0	5.0	6.303487	-75.567591
5	9204.0	Centro de Tecnología de la Manufactura Avanzada	1.0	1.0	5.0	6.300949	-75.568664
6	9205.0	Centro Tecnológico del Mobiliario	1.0	360.0	5.0	6.180750	-75.606430
7	9206.0	Centro Textil y de Gestión Industrial	1.0	1.0	5.0	6.302387	-75.567365
8	9301.0	Centro de Comercio	1.0	1.0	5.0	6.254794	-75.575466
9	9401.0	Centro de Servicios de Salud	1.0	1.0	5.0	6.251389	-75.571571



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

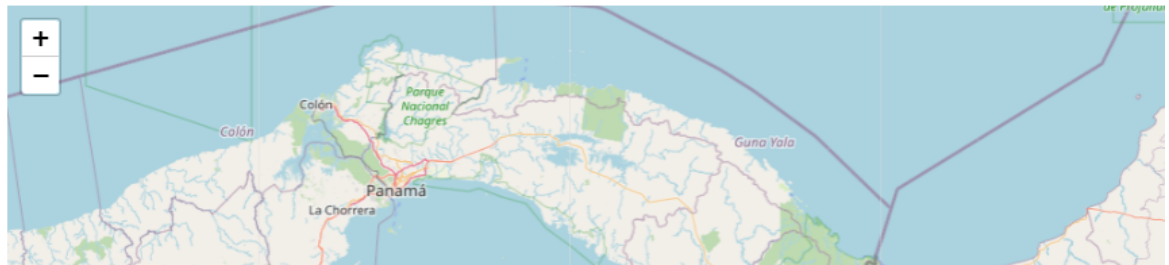
```
[ ] centros_formacion.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Codigo_Centro        16 non-null    float64
1   _Nombre_Centro        16 non-null    object
2   _Codigo_Regional      16 non-null    float64
3   _Codigo_Municipio     16 non-null    float64
4   _Codigo_Departamento  16 non-null    float64
5   latitud               16 non-null    float64
6   longitud              16 non-null    float64
dtypes: float64(6), object(1)
memory usage: 1.0+ KB
```

```
[ ] #Datos de coordenadas geoespaciais
my_zip = zip(centros_formacion['latitud'], centros_formacion['longitud'])
list_of_my_zip = list(my_zip)
```

```
[ ] hm_layer = HeatMap(list_of_my_zip,
                        # These are parameters that we tweak manually to adjust color
                        # See folium docs for more information
                        min_opacity=3,
                        radius= 20,
                        blur= 30,
                        )
```

```
[ ] folium_map = folium.Map()
ny_coords = [6.55, -75.817 ] # lat, long
folium_map = folium.Map(location=ny_coords,zoom_start=8) #tiles="Stamen Terrain")
folium_map.add_child(hm_layer)
```



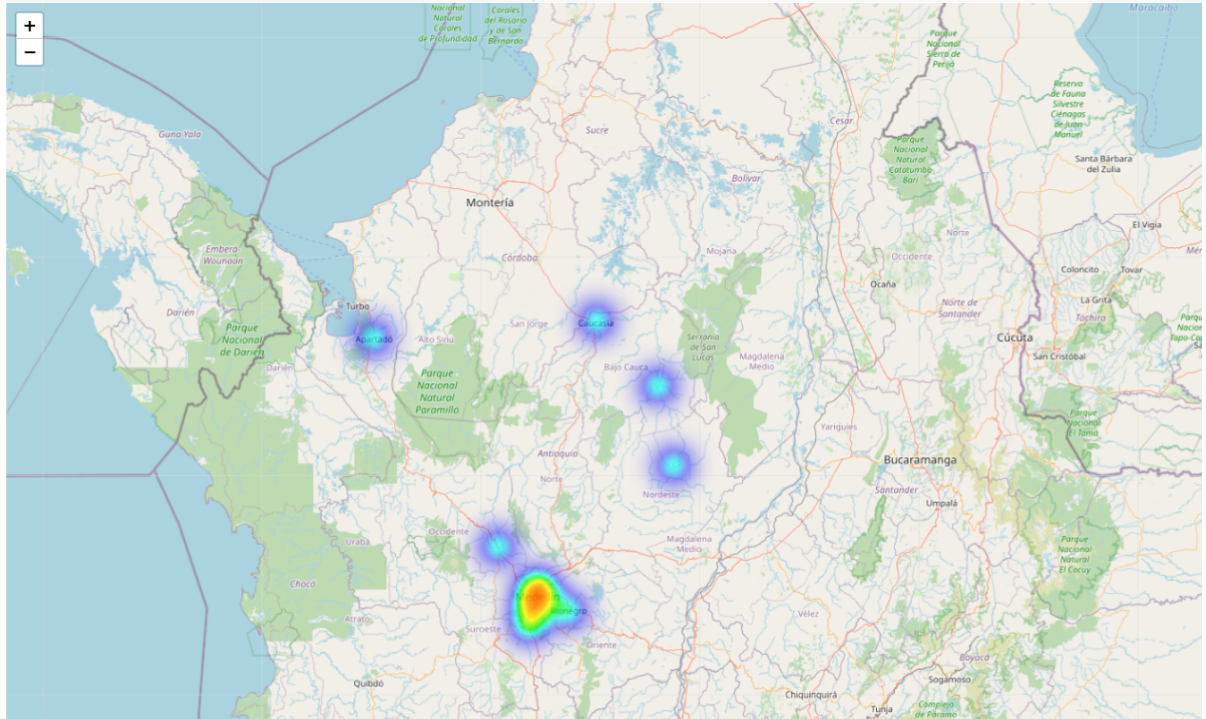


Python_SGPS 9748_Preprocesamiento.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
[ ] folium_map = folium.Map()
ny_coords = [6.55, -75.817 ] # lat, long
folium_map = folium.Map(location=ny_coords,zoom_start=8) #tiles="Stamen Terrain"
folium_map.add_child(hm_layer)
```



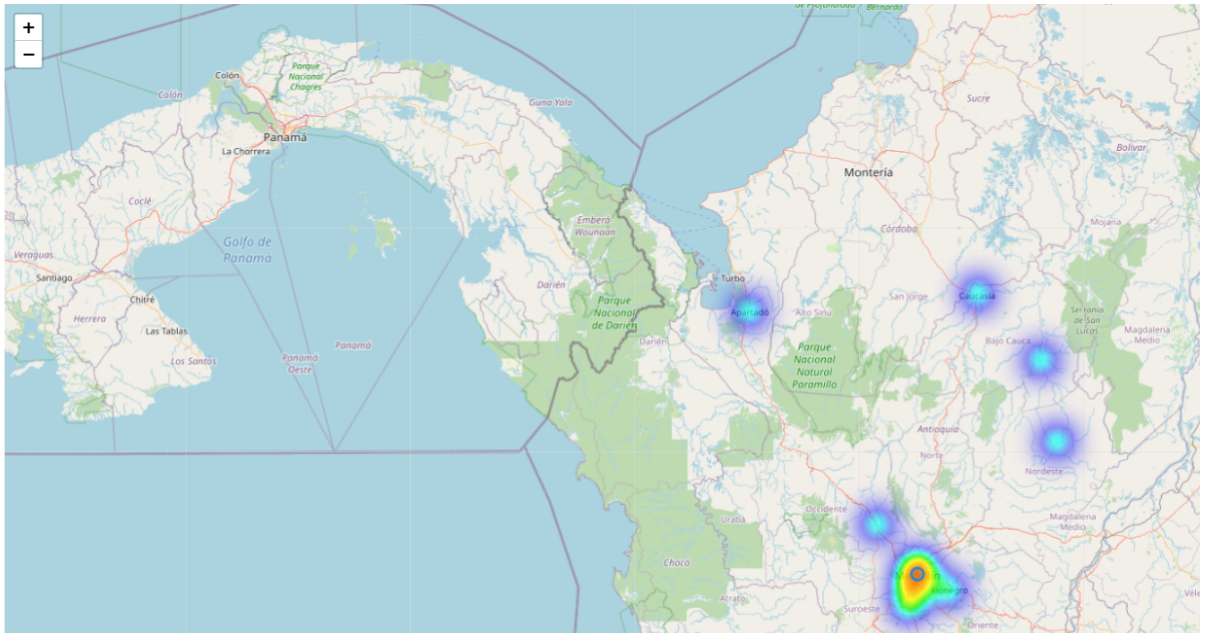


Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
[ ] #Ubicación centro de servicios y gestión empresarial
folium.CircleMarker(
    location=[6.255546889481773, -75.57468130270018],
    radius=7,
    popup="Laurelhurst Park",
    color="#3186cc",
    fill=True,
    fill_color="#3186cc",
).add_to(folium_map)
folium_map
```



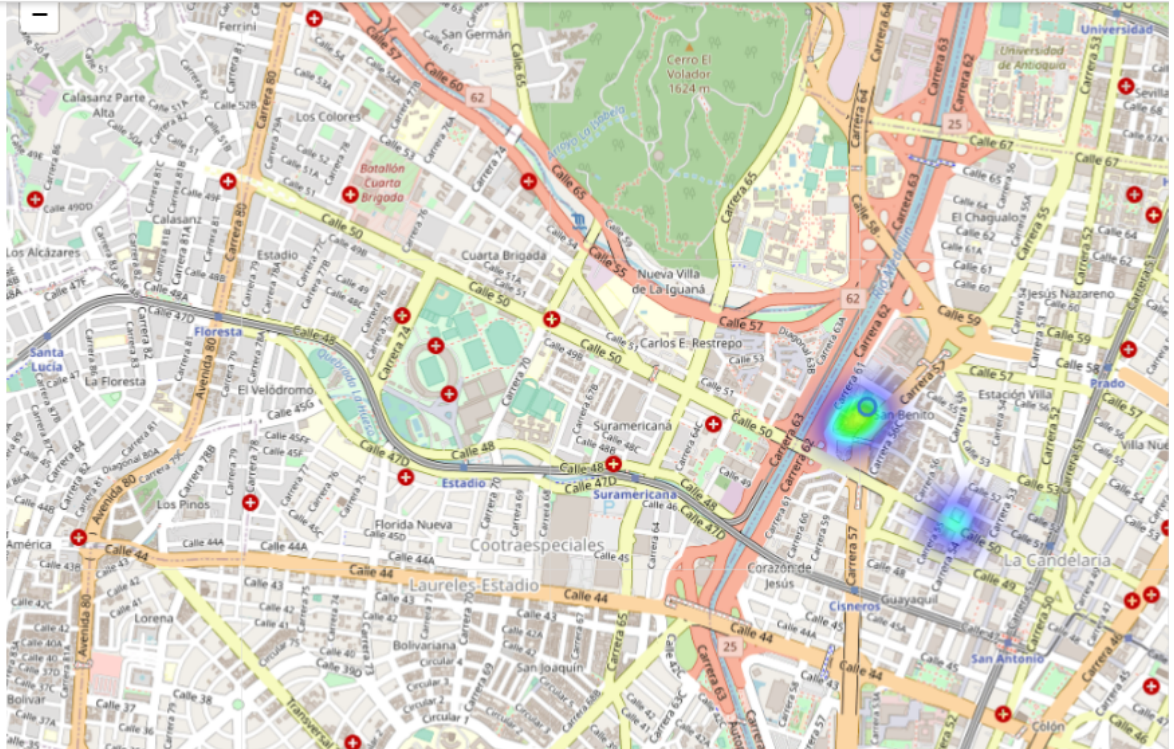


Python_SGPS 9748_Preprocesamiento.ipynb



Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto





Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto



```
[ ] #Descarga en un archivo .html
folium_map.save("Centros de Formación Antioquia.html")
```

```
[ ] departamento.head()
```

	_Codigo_Departamento	_Nombre_Departamento
0	8.0	ATLANTICO
1	11.0	BOGOTA
2	13.0	BOLIVAR
3	15.0	BOYACA
4	17.0	CALDAS

```
[ ] departamento.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33 entries, 0 to 32
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Codigo_Departamento  33 non-null    float64
1   _Nombre_Departamento  33 non-null    object
dtypes: float64(1), object(1)
memory usage: 656.0+ bytes
```

```
[ ] regionales.head()
```

	_Codigo_Regional	_Nombre_Regional	_Codigo_Departamento	_Codigo_Municipio
0	5.0	Regional Antioquía	232.0	NaN
1	8.0	Regional Atlántico	8.0	NaN
2	11.0	Regional Distrito Capital - Bogotá	11.0	NaN
3	13.0	Regional Bolívar	13.0	NaN
4	15.0	Regional Boyacá	15.0	NaN



Python_SGPS 9748_Preprocesamiento.ipynb

ComentarioCompartir

ArchivoEditarVerInsertarEntorno de ejecuciónHerramientasAyudaÚltima modificación: 11 de marzo

+ Código + Texto

[] 1 Nombre_Departamento 32 non-null object
dtypes: float64(1), object(1)
memory usage: 656.0+ bytes

[] regionales.head()

	_Codigo_Regional	_Nombre_Regional	_Codigo_Departamento	_Codigo_Municipio
0	5.0	Regional Antioquia	232.0	NaN
1	8.0	Regional Atlántico	8.0	NaN
2	11.0	Regional Distrito Capital - Bogotá	11.0	NaN
3	13.0	Regional Bolívar	13.0	NaN
4	15.0	Regional Boyacá	15.0	NaN

[] regionales.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 32 entries, 0 to 31  
Data columns (total 4 columns):  
#   Column              Non-Null Count  Dtype  
---  ---  
0   _Codigo_Regional    32 non-null     float64  
1   _Nombre_Regional    32 non-null     object  
2   _Codigo_Departamento 32 non-null     float64  
3   _Codigo_Municipio   0 non-null      float64  
dtypes: float64(3), object(1)  
memory usage: 1.1+ KB
```

[] principal_productos.head(20)

	_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	
0	8000.0	Artículo Sometido en revista Indexada		NaN	5.0	9101.0	23.0	7500.0	244.0	9999	8893.0	NaN	3412.0	NaN	2018-04-11
1	8001.0	Artículo Sometido en revista Indexada		NaN	5.0	9127.0	66.0	7500.0	3412.0	9999	8892.0	NaN	3432.0	NaN	2019-12-11
2	8002.0	Artículo Sometido en revista Indexada		NaN	5.0	9201.0	66.0	7500.0	3432.0	9999	8898.0	NaN	3497.0	NaN	2019-12-11
3	8003.0	Artículo Sometido en revista Indexada		NaN	5.0	9202.0	66.0	7500.0	3439.0	9999	8898.0	NaN	3497.0	NaN	2019-12-11
4	8004.0	Artículo Sometido en revista Indexada		NaN	5.0	9203.0	82.0	7500.0	3497.0	9999	8897.0	NaN	3510.0	NaN	2019-12-11
5	8005.0	Artículo Sometido en revista Indexada		NaN	5.0	9204.0	82.0	7500.0	3510.0	9999	8897.0	NaN	3510.0	NaN	2019-12-11



Python_SGPS 9748 _Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

[] en revista indexada

[] principal_productos.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   _Codigo_Producto                     234 non-null   float64
1   _Nombre_Producto                     234 non-null   object
2   _Descripcion_producto                0 non-null     float64
3   _Codigo_Regional                     234 non-null   float64
4   _Codigo_Centro                      234 non-null   float64
5   _Codigo_Rubro                        228 non-null   float64
6   _Codigo_Tipologia                    234 non-null   float64
7   _Codigo_Proyecto                     234 non-null   float64
8   _Nombre_Proyecto                     234 non-null   int64
9   _Codigo_Grupo_Investigacion         234 non-null   float64
10  _Codigo_Semillero                    0 non-null     float64
11  _Codigo_persona                      234 non-null   float64
12  _Codigo_area_OCDE                    0 non-null     float64
13  _Fecha_registro                      234 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(11), int64(1), object(1)
memory usage: 25.7+ KB
```

[] principal_productos.isna().sum()

```
_Codigo_Producto          0
_Nombre_Producto          0
_Descripcion_producto     234
_Codigo_Regional          0
_Codigo_Centro            0
_Codigo_Rubro             6
_Codigo_Tipologia         0
_Codigo_Proyecto          0
_Nombre_Proyecto          0
_Codigo_Grupo_Investigacion 0
_Codigo_Semillero         234
_Codigo_persona           0
_Codigo_area_OCDE         234
_Fecha_registro           0
dtype: int64
```

[] # Nombre_Producto
principal_productos["_Nombre_Producto"].value_counts()

```
Articulo Sometido en revista Indexada          96
Informe de investigación                        66
Apropiación Social                             30
Productos tecnológicos certificados o validados 30
```



Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Última modificación: 11 de marzo](#)

+ Código + Texto

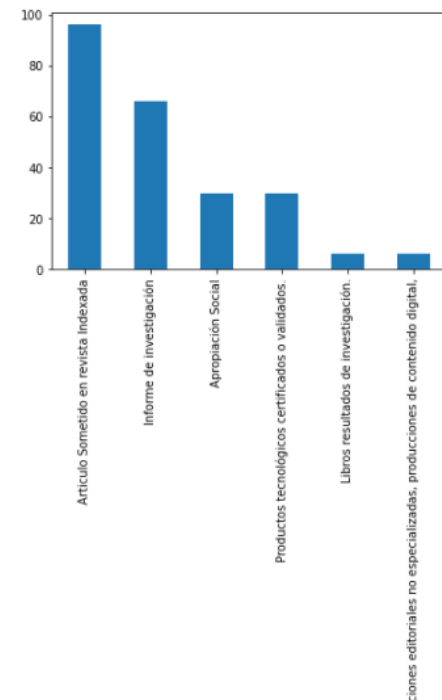
dtype: int64

```
[ ] # Nombre_Producto
principal_productos["_Nombre_Producto"].value_counts()
```

```
Articulo Sometido en revista Indexada      96
Informe de investigación                  66
Apropiación Social                        30
Productos tecnológicos certificados o validados. 30
Libros resultados de investigación.         6
Divulgación Pública de la CTeI. Publicaciones editoriales no especializadas, producciones de contenido digital, 6
Name: _Nombre_Producto, dtype: int64
```

```
[ ] # Nombre_Producto
plot_principal_productos = principal_productos["_Nombre_Producto"].value_counts()
plot_principal_productos.plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f89d98bd310>





Python_SGPS 9748_Preprocesamiento.ipynb ☆

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo



+ Código + Texto

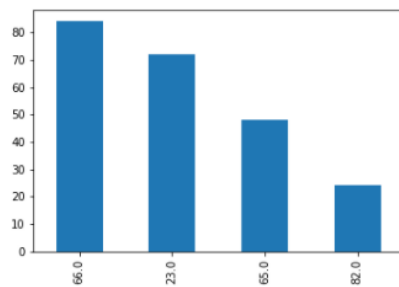


```
[ ] # Rubro
principal_productos["_Codigo_Rubro"].value_counts()

66.0    84
23.0    72
65.0    48
82.0    24
Name: _Codigo_Rubro, dtype: int64
```

```
[ ] plot_Codigo_Rubro = principal_productos["_Codigo_Rubro"].value_counts()
plot_Codigo_Rubro.plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f89d92d5dd0>



```
[ ] #Cantidad de Grupos de investigación
principal_productos["_Codigo_Grupo_Investigacion"].value_counts()

8897.0    54
8893.0    44
8892.0    41
8898.0    41
8894.0    22
8888.0    18
8889.0    14
Name: _Codigo_Grupo_Investigacion, dtype: int64
```



Python_SGPS 9748 _Preprocesamiento.ipynb ☆

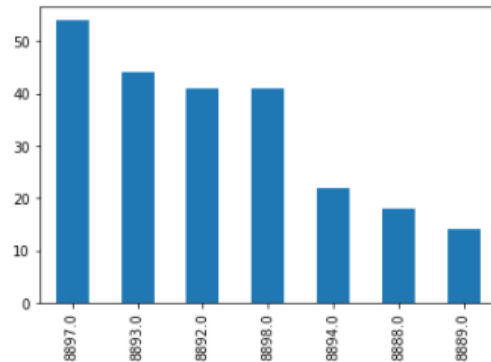
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

```
8897.0 14
Name: _Codigo_Grupo_Investigacion, dtype: int64
```

```
[ ] plot_Codigo_Grupo_Investigacion = principal_productos["_Codigo_Grupo_Investigacion"].value_counts()
plot_Codigo_Grupo_Investigacion.plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f89d995e390>



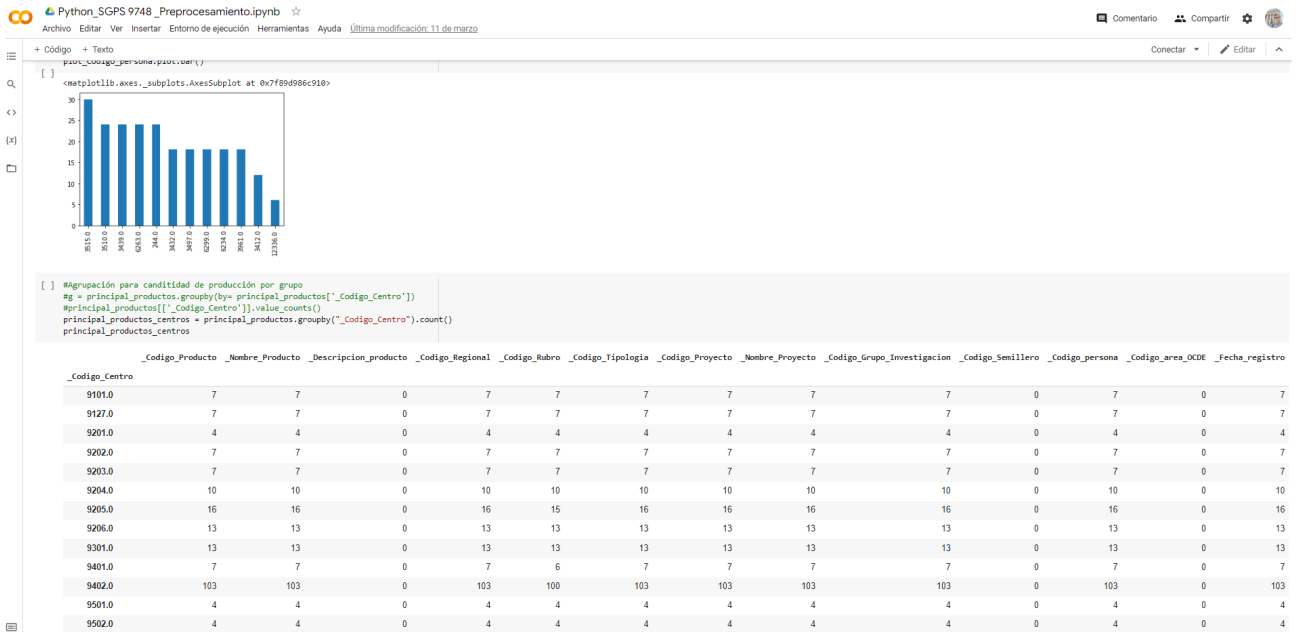
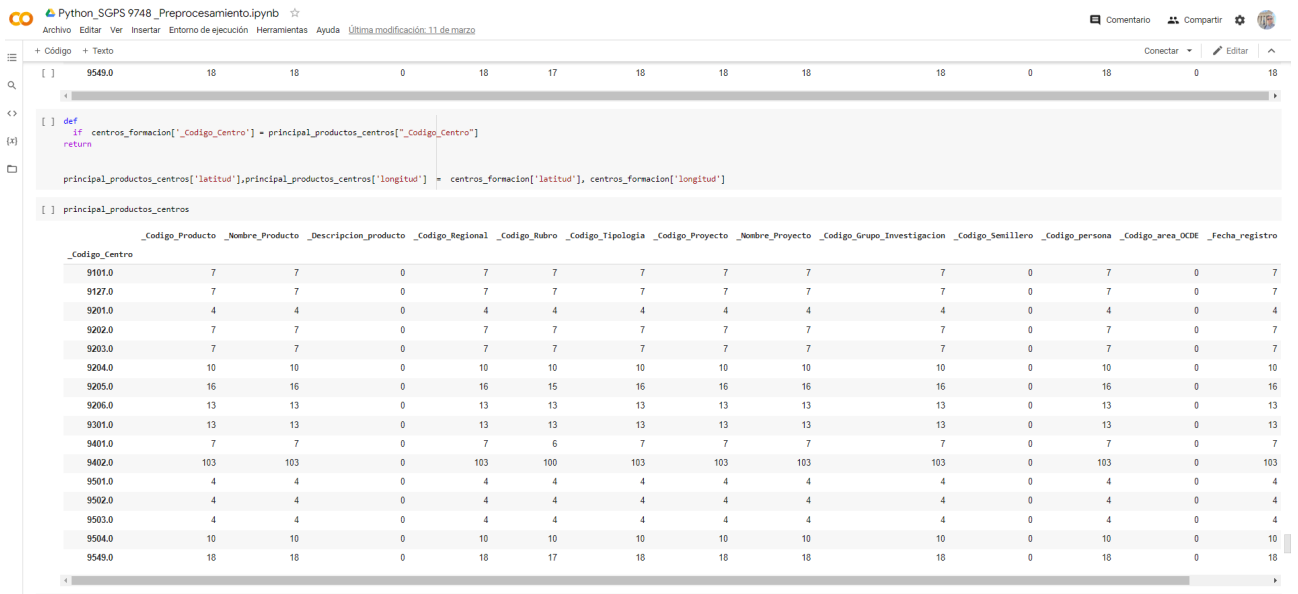
```
[ ] #Cantidad de productos por persona
principal_productos["_Codigo_persona"].value_counts()
```

```
3515.0 30
3510.0 24
3439.0 24
6263.0 24
244.0 24
3432.0 18
3497.0 18
6299.0 18
6234.0 18
3961.0 18
3412.0 12
12336.0 6
Name: _Codigo_persona, dtype: int64
```

```
[ ] plot_Codigo_persona = principal_productos["_Codigo_persona"].value_counts()
plot_Codigo_persona.plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f89d986c910>







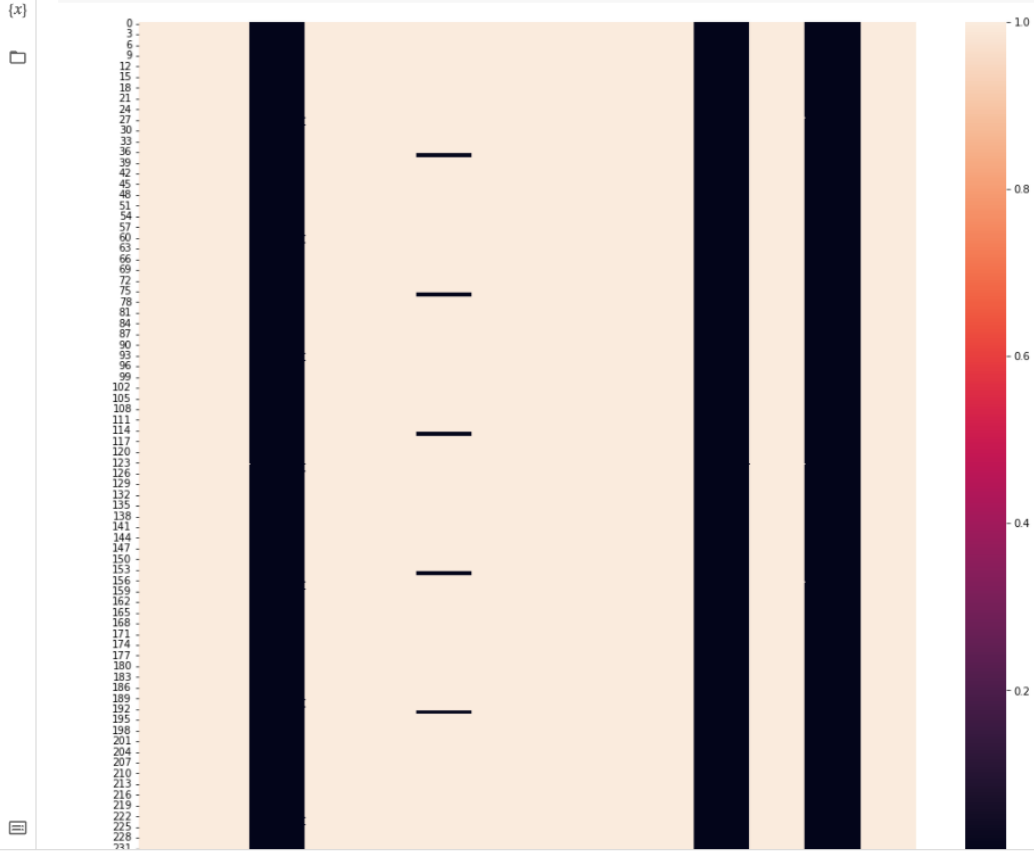
Python_SGPS 9748_Preprocesamiento.ipynb ☆

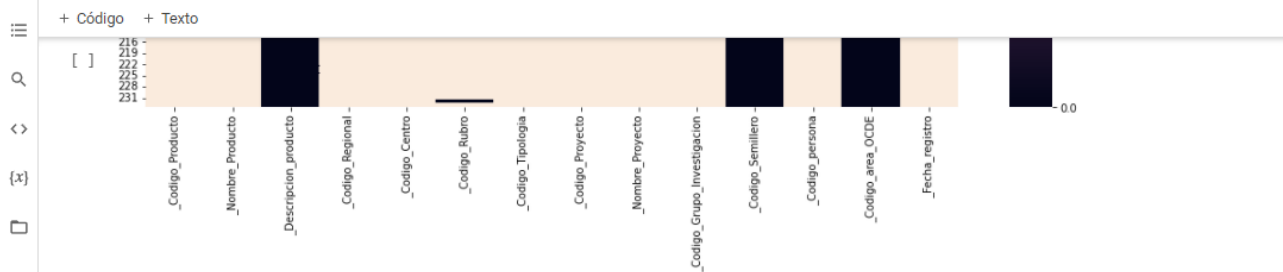
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

SEARCH STACK OVERFLOW

```
[ ] # Verificacion de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(principal_productos.notna())
```





```
[ ] tipo_equipo.head()
```

```
_Codigo_Tipo_Equipo  _Nombre_Tipo_Equipo
```

```
[ ] tipo_equipo.info()
```

```
xclass 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   _Codigo_Tipo_Equipo    0 non-null     object
1   _Nombre_Tipo_Equipo    0 non-null     object
dtypes: object(2)
memory usage: 0.0+ bytes
```

```
[ ] rol_proyectos.head()
```

_Codigo_Rol_Proyecto	_Nombre_Rol
1	Administrador
2	Usuario
3	Invitado

```
[ ] rol_proyectos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
0   _Codigo_Rol_Proyecto  0 non-null     object
1   Nombre_Rol            0 non-null     object
dtypes: object(2)
memory usage: 0.0+ bytes
```

```
[ ] estado_contratacion.head()
```



```
Python_SGPS 9748_Preprocesamiento.ipynb ☆
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Última modificación: 11 de marzo

+ Código + Texto

[ ] estado_contratacion.head()

   _Codigo_Estado_Contratacion  _Nombre_Estado_Contratacion

[ ] estado_contratacion.info()

<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Codigo_Estado_Contratacion  0 non-null    object
1   _Nombre_Estado_Contratacion  0 non-null    object
dtypes: object(2)
memory usage: 0.0+ bytes

[ ] principal_equipos.head()

   _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

[ ] principal_equipos.info()

<class 'pandas.core.frame.DataFrame'>
Index: 0 entries
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   _Serial_equipo        0 non-null    object
1   _Nombre_del_equipo    0 non-null    object
2   _Valor_equipo         0 non-null    object
3   _Fecha_de_compra      0 non-null    object
4   _Codigo_Proyecto      0 non-null    object
5   _Codigo_Regional      0 non-null    object
6   _Codigo_Centro        0 non-null    object
7   _Codigo_Rubro         0 non-null    object
8   _Codigo_Tipo_Equipo   0 non-null    object
9   _Codigo_area_OCDE     0 non-null    object
10  _Codigo_persona       0 non-null    object
dtypes: object(11)
memory usage: 0.0+ bytes
```



ANEXO 4. Reunión Seguimiento Proyecto SGPS 9748 Prototipo Innovación - Consultas Modelo

DIMENSION	CATEGORIA	INDICADOR	INDICADOR FINAL	OBSERVACIÓN	EVENTOS
RECURSOS ECONÓMICOS	GASTO/INVERSIÓN EN ACTI	Suma de los diferentes rubros ejecutados.			46 - Investigación aplicada 47 - Innovación 48 - Divulgación 49 - Modernización
	NÚMERO DE LABORATORIOS PARA ACTI	AMBIENTES DE APRENDIZAJE (Laboratorios dotados con recursos de SENNOVA)			
RECURSOS FÍSICOS Y TECNOLÓGICOS	INVERSIÓN DE EQUIPOS DE PARA ACTI	Suma del rubro de equipos y maquinaria.		Rubro de equipos, /Equipos de Computo/maquinaria	Acta de cierr
	NÚMERO DE LICENCIAS (campo de conocimiento según OCDE) (Por programas de educación)	Software de aplicaciones especializadas Software de programación	Cantidad de procesos adquiridos por los procesos SENNOVA (Se han usado para investigación)		Informe Pre Modernizaci Software
TALENTO HUMANO					

Entradas

Proceso

Salidas

Modelo (tasa rendimiento)

NÚMERO DE LABORATORIOS PARA ACTI

INVERSIÓN DE EQUIPOS DE PARA ACTI

NÚMERO DE LICENCIAS (campo de conocimiento según OCDE) (Por programas de educación)



The screenshot shows a Zoom meeting interface. On the left, a diagram titled 'Modelo (tasa rendimiento)' is displayed. It consists of a box labeled 'Proceso' connected to 'Entradas' (Inputs) on the left and 'Salidas' (Outputs) on the right. Below 'Proceso' is another box labeled 'Modelo (tasa rendimiento)'. The Zoom toolbar at the bottom shows the time as 20:37. On the right, a chat window titled 'Chat de la reunión' is open, showing a list of participants and a link to a Microsoft Forms survey: <https://forms.office.com/r/zAWHpypc5>. The bottom status bar shows the time as 3:13 p.m. on 2/03/2022.

The screenshot shows a Microsoft Teams meeting interface. The main content area displays a presentation slide with a table of indicators. The table has four columns: 'RECURSOS ECONÓMICOS', 'GASTO/INVERSIÓN en ACTI', 'Suma de los diferentes rubros ejecutados.', and '82 - Innovación 65 - Divulgación 23 - Modernización'. The table contains several rows of data, including 'NÚMERO DE LABORATORIOS PARA ACTI', 'INVERSIÓN DE EQUIPOS DE PARA ACTI', and 'NÚMERO DE LICENCIAS (Campo de conocimiento según la OCDE)'. The Microsoft Teams toolbar at the bottom shows the time as 41:52. The bottom status bar shows the time as 3:35 p.m. on 2/03/2022.

ANEXOS 5. Presentación de avances técnicos marzo 2022.



**El empleo
es de todos**

Mintrabajo

Presentación de avances 03

Fase 2: Identificación y selección del Dataset

Proyecto SGPS 9748 - Prototipo Innovación

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

Regional Antioquia aplicando analítica de datos

09 de marzo del 2022



@SENAcomunica

www.sena.edu.co

Contenido



- 1. Fase 2 del modelo CRISP - DM - Proyecto SGPS-9748**
- 2. Diagrama de flujo del modelo CRISP - DM**
- 3. Modelo conceptual**
- 4. Instalación de software**
- 5. Base de datos**
 - 5.1. Organización de base de datos**
 - 5.2. Extracción de base de datos SQL**
 - 5.3. Dataset**
 - 5.4. Identificación de variables**
 - 5.5. Base de datos preprocesamiento**
 - 5.6. Propuesta de librerías de visualización para el front end**
- 6. Sugerencias**



1. Fase 2 del 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. ENTENDIMIENTO DE LOS DATOS	2.1 Recolección inicial de datos	2.1.1 Informe inicial de recopilación de datos
	2.2 Descripción de los datos	2.2.1 Informe de descripción de datos
	2.3 Exploración de los datos	2.3.1 Informe de exploración de datos
	2.4 Verificación de calidad de los datos	2.4.1 Informe de calidad de datos
3. PREPARACION DE DATOS	3.0 Tareas preparatorias	3.0.1 Conjunto de datos 3.0.2 Descripción del conjunto de datos
	3.1 Selección de datos	3.1.1 Justificación de la inclusión / exclusión
	3.2 Limpieza de datos	3.2.1 Informe de limpieza de datos
	3.3 Construcción de datos	3.3.1 Atributos derivados 3.3.2 Registros generados
	3.4 Integración de los datos	3.4.1 Datos combinados
	3.5 Formato de datos	3.5.1 Datos reformateadas

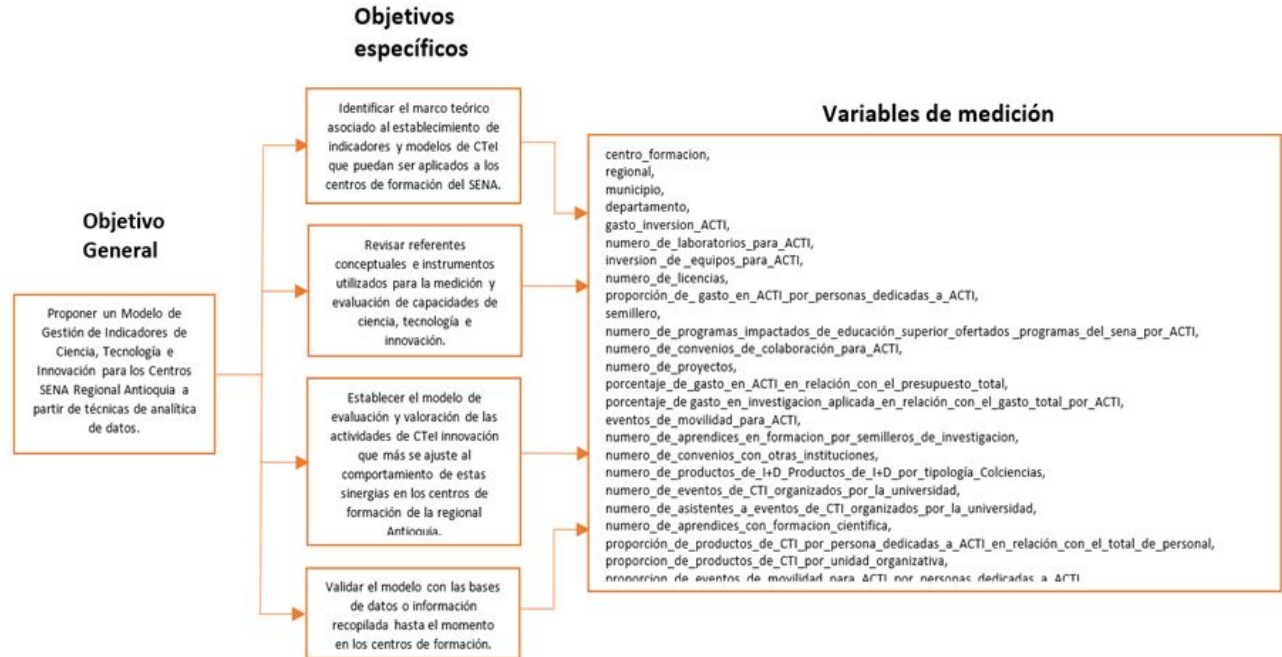


1. Fase 2 del modelo CRISP - DM - Proyecto SGPS-9748

Metodología CRISP-DM

Fase 1: comprensión del negocio, en esta etapa se establecen las variables que permiten medir los indicadores de CTel, y se procede a validar y revisar la correspondencia de las mismas con los objetivos específicos del proyecto que permite la implementación del modelo de gestión de CTel, para dar solución al problema.

https://docs.google.com/spreadsheets/d/1vL_HVo1sFDdggC_gYSwjY8axuVdo8fJenDftpicF_WN0/edit?usp=sharing



1. Indicadores y correspondencia de variables de entrada y salida del modelo



Nro	DIMENSIÓN (Proceso de entradas y salidas)	CATEGORIA	INDICADOR (Variables)	INDICADOR FINAL				
1	Insumos para actividades de Ciencia, Tecnología e Innovación (ACTI) (Entradas en el modelo)	RECURSOS ECONÓMICOS	GASTO/ INVERSIÓN en ACTI	Suma de los diferentes rubros ejecutados.	9	Proceso (Planta del modelo)	PROYECTOS	NÚMERO DE PROYECTOS
2		RECURSOS FÍSICOS Y TECNOLÓGICOS	NÚMERO DE LABORATORIOS PARA ACTI	AMBIENTES DE APRENDIZAJE (Laboratorios dotados con recursos de SENNOVA)	10		TALENTO HUMANO	PROPORCIÓN DE GASTO EN ACTI POR PERSONAS DEDICADAS A ACTI
3			INVERSIÓN DE EQUIPOS DE PARA ACTI	Suma del rubro de equipos y maquinaria.	11		RECURSOS ECONÓMICOS	PORCENTAJE DE GASTO EN ACTI en relación con el PRESUPUESTO TOTAL
4			NÚMERO DE LICENCIAS (Campo de conocimiento según la OCDE) (Por programas de educación)	-Software de aplicaciones especializadas -Software de programación	12		RELACIONES CON EL ENTORNO	Porcentaje de gasto en INVESTIGACIÓN APLICADA en relación con el gasto total por ACTI
5		TALENTO HUMANO	NÚMERO DE PERSONAS DEDICADAS A ACTI		13	Resultado (Salidas del modelo)	FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	EVENTOS DE MOVILIDAD PARA ACTI
6					14		RELACIONES CON EL ENTORNO	Formación de aprendices en semilleros de investigación
7		RECURSOS ESTRUCTURALES Y RELACIONALES	NÚMERO DE UNIDADES ORGANIZATIVAS (GRUPOS DE INVESTIGACIÓN - SEMILLEROS)		15		PRODUCTOS RESULTADOS DE LAS ACTIVIDADES DE CTI	Convenios con otras instituciones
8			NÚMERO DE PROGRAMAS IMPACTADOS DE EDUCACIÓN SUPERIOR OFERTADOS (PROGRAMAS DEL SENA POR ACTI)		16		TRANSFERENCIA DE CONOCIMIENTO	NÚMERO DE PRODUCTOS DE I+D (Productos de I+D por tipología Colciencias)
			NÚMERO DE CONVENIOS DE COLABORACIÓN PARA ACTI		17	Productividad (Rendimiento del modelo)	FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	NÚMERO DE EVENTOS DE CTI ORGANIZADOS POR LA UNIVERSIDAD (Número de eventos de CTI por clasificación ACTI)
					19		TALENTO HUMANO	Formación de aprendices con formación científica
					20		RELACIONES CON EL ENTORNO	PROPORCIÓN DE PRODUCTOS DE CTI POR PERSONA DEDICADAS A ACTI, EN RELACIÓN CON EL TOTAL DE PERSONAL
					21		TALENTO HUMANO	PROPORCIÓN DE PRODUCTOS DE CTI POR UNIDAD ORGANIZATIVA
					22		RELACIONES CON EL ENTORNO	PROPORCIÓN DE EVENTOS DE MOVILIDAD PARA ACTI POR PERSONAS DEDICADAS A ACTI

Identificación de indicadores: Indicadores y correspondencia de variables de entrada y salida del modelo

https://docs.google.com/spreadsheets/d/1vL_HVo1sFDqqCgYSwjY8axuVdo8fJenDftpicF_WN0/edit?usp=sharing

2. Diagrama de flujo del modelo CRISP - DM



Metodología CRISP-DM aplicada al modelo de medición de indicadores de CTel



Primer paso: Comprensión de datos

Tipo de característica: nominal, numérica, etc.
Escalamiento de datos
Estadísticas básicas para detectar anomalías
Correlaciones entre características

Segundo paso: Preparación del conjunto de datos

Reconstrucción de datos faltantes
Limpieza de datos
Definición de filtros de umbral

Tercer paso: Modelado de los datos

Manipulación de datos
Creación de características, mediante filtros
Dividir el conjunto de datos en uno de prueba y otro de entrenamiento
Entrenar el modelo de regresión
Validar el modelo de regresión

Cuarto paso: Evaluación de modelos e interpretación de resultados

Interpretar las métricas generadas por el modelo
Reevaluar modelo si el resultado no es el esperado

Quinto paso: Análisis general del modelo

En este paso es importante volver a mirar el modelo y si algo más se puede agregar es el momento de tener el panorama general del modelo.

3. Modelo conceptual

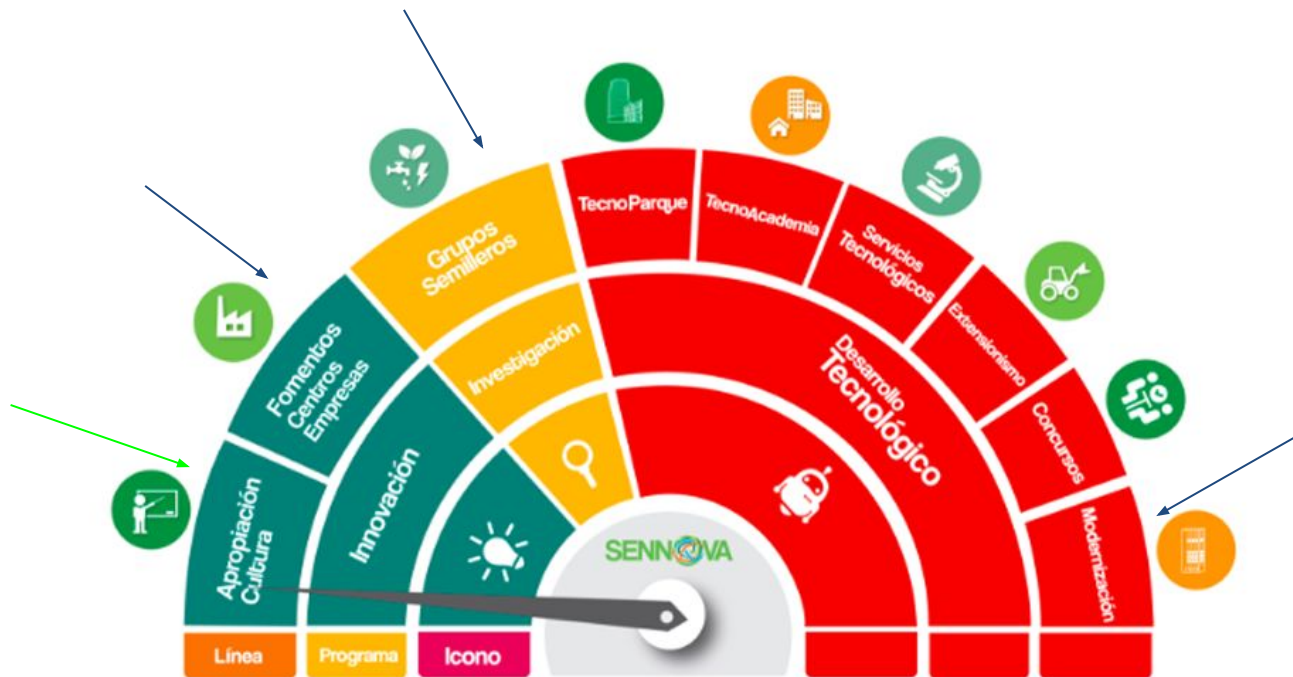
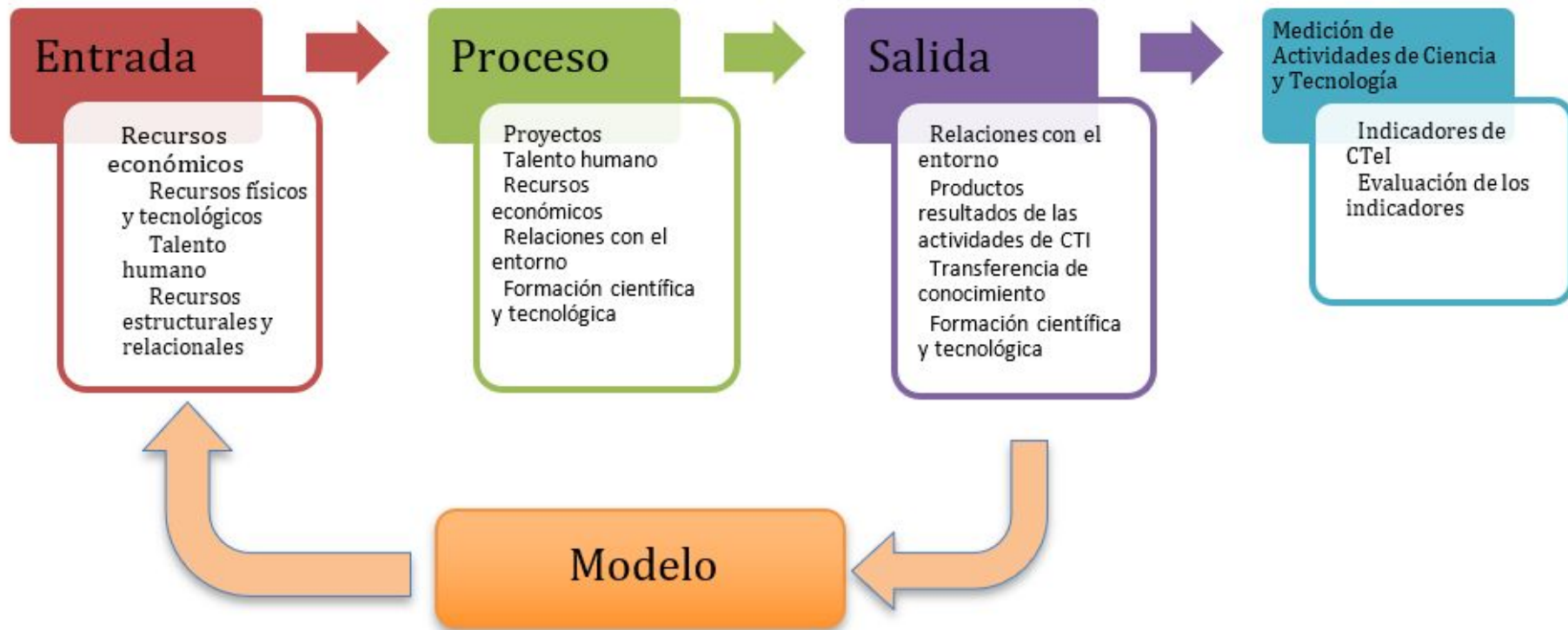


Figura 1. Programas y líneas programáticas de la estrategia SENNOVA.

Fuente: <http://Sennova.senaedu.edu.co/>

3. Modelo conceptual



4. Instalación de software

Instalados

SQL-Server

SQL Management Studio

Node js

Java Runtime

Git

Anaconda

Rstudio

R

Python



5. Base de datos

- 5.1. Organización de base de datos
- 5.2. Extracción de base de datos SQL
- 5.3. Dataset
- 5.4. Identificación de variables críticas
- 5.5. Base de datos preprocesamiento
- 5.6. Propuesta de librerías de visualización para el front end

5.1. Organización de base de datos

Organización de base datos: Variables organizadas por hojas

Se organizaron las variables en hojas dentro de un documento de archivo de acuerdo al modelo relacional.

https://docs.google.com/spreadsheets/d/1nQzejQe2ZgNTMHWBuq_-9Wgs4xHMBWve/edit?usp=sharing&oid=110720597759867780864&rtpof=true&sd=true

	A	B	C	D	E	F	G	H	I	J	K	L	M
		talento_hu_mano_codigo_persona											
	talento_hu_mano_codigo_persona	talento_hu_mano_persona	talento_hu_mano_direccion	talento_hu_mano_duracion_contrato	talento_hu_mano_tipo_pago_dedicacion_semestral	talento_hu_mano_valor_mensual_contrato	talento_hu_mano_valor_total_contrato	talento_hu_mano_seneca_emico	talento_hu_mano_nivel_academico	talento_hu_mano_tipo_documento	talento_hu_mano_estado_contrato	talento_hu_mano_estado_documento	talento_hu_mano_estado_documento
3000									5009	2004			
3001									5005	2002			
3002									5008	2002			
3008									5005	2002			
3010									5009	2003			
3011									5005	2002			
3001									5005	2002			
3013									5008	2002			
3014									5008	2002			
3015									5009	2002			
3016									5007	2002			
3017									5009	2002			
3018									5007	2002			
3019									5009	2002			
3000									5009	2004			
3027									5007	2002			
3028									5007				
3029									5007				
3030									5007				
3020									5005				
3021									5008				
3022									5007				
3023									5005	2003			
3024									5007	2002			
3025									5009	2003			
3000									5009	2004			
3020									5005	2002			
3021									5007	2002			
3022									5001	2001			
3023									5001	2001			
3022									5008	2003			

regional ▼ centro_formacion ▼ registro ▼ estado_proyecto ▼ grupo_i ◀ ▶

5.1. Organización de base de datos

SENA OneDrive

Juliett Restrepo Zapata

Mis archivos

Nuestros archivos

Buscar

+ Nuevo Cargar Compartir Copiar vínculo Sincronizar Descargar Añadir acceso directo a Mis archivos

Mis archivos > SENNOVA 2022 > PROTOTIPO INNOVACIÓN SENA 2022 > Documentos 2021 > INDICADORES CTel > RECOLECCIÓN -DATOS

Nombre	Modificado	Modificado por	Tamaño de archi...	Compartir
Actas de Cierre - Centro Servicios	16 de febrero	Erika Juliatt Restrepo Zapa	1 elemento	Compartido
INFORMACIÓN SENNOVA - DATOS - STEPH...	16 de febrero	Erika Juliatt Restrepo Zapa	9 elementos	Compartido
Prototipo V1 - ALFA	16 de febrero	Erika Juliatt Restrepo Zapa	6 elementos	Compartido
Centros de la Regional Antioquia.xlsx	16 de febrero	Erika Juliatt Restrepo Zapa	18,0 KB	Compartido
Faltantes.xlsx	16 de febrero	Erika Juliatt Restrepo Zapa	17,2 KB	Compartido
Informe Presupuestal SENNOVA.xlsx	1 de marzo	Paola Andrea Buitrago Car	39,5 KB	Compartido

Referencia: [RECOLECCIÓN -DATOS](#)

5.2. Extracción de base de datos SQL

5.3. Identificación inicial de variables

Regionales _Codigo_Regional _Nombre_Regional _Codigo_Departamento _Codigo_Municipio	Grupo_investigacion _Codigo_Grupo_Investigacion _Nombre_Grupo_Investigacion _Categoria_Grupo _Codigo_Centro _Codigo_Regional	Semilleros _Codigo_Semillero _Nombre_Semillero _Codigo_Grupo_Investigacion _Codigo_Centro _Codigo_Regional
Rubro _Codigo_Rubro _Nombre_Rubro	Rol_Sennova _Codigo_Rol _Nombre_Rol	Tipo_Contratacion _Codigo_Tipo_Contrato _Nombre_Tipo_Contrato
Principal_Productos _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	Principal_Proyectos _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	Principal_Talento_Humano _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
Area_conocimiento_OCDE _Codigo_area_OCDE _Nombre_area_OCDE	Estado_Proyecto _Codigo_Estado_Proyecto _Nombre_Estado_Proyecto	Nivel_Academico _Codigo_Nivel_Academico _Nombre_Nivel_Academico

5.3. Identificación inicial de variables

Municipio _Codigo_Municipio _Nombre_Municipio _Codigo_Departamento	Centros_Formacion _Codigo_Centro _Nombre_Centro _Codigo_Regional _Codigo_Municipio _Codigo_Departamento _latitud _longitud	Principal_Equipos _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
Tipologia_Productos _Codigo_Tipologia _Nombre_Tipologia	Tipo_Contratación _Codigo_Tipo_Contrato _Nombre_Tipo_Contrato	Tipo_equipo _Codigo_Tipo_Equipo _Nombre_Tipo_Equipo
Rol_Proyectos _Codigo_Rol_Proyecto _Nombre_Rol	Estado_Contratación _Codigo_Estado_Contratacion _Nombre_Estado_Contratacion	Departamento _Codigo_Departamento _Nombre_Departamento

5.3. Dataset



Buscar en el menú (Alt+/)

100% € % .0 .00 123 ▾ Predetermi... 11 ▾ B I G A

A	B
Código_Tipología	
Código_Tipología	Nombre_Tipología
7500	2.2.1.1 Artículos de investigación A1, A2, B y C. Artículos en revistas indexadas en los índices bibliográficos de citas y índices bibliográficos. Los requerimientos son mencionados en la Tabla I del ANEXO 1.
7501	2.2.1.1 Artículos de investigación D.
7502	2.2.1.2 Notas científicas.
7503	2.2.1.3 Libros resultados de investigación.
7504	2.2.1.4 Capítulos en libro resultado de investigación.
7505	2.2.3.6.1 Libros de Formación Q1.
7506	2.2.1.5 Productos tecnológicos patentados o en proceso de solicitud de patente.
7507	2.2.1.6 Variedades vegetales, nuevas razas animales y poblaciones mejoradas de razas pecuarias.
7508	2.2.1.7 Productos resultados de la creación o investigación-creación.
7509	2.2.2.1 Productos tecnológicos certificados o validados.
7510	2.2.2.2 Productos empresariales.
7511	2.2.2.3 Regulaciones, normas, reglamentos o legislaciones.
7512	Los requerimientos son mencionados en la Tabla XII del ANEXO 1.
7513	2.2.2.4 Conceptos técnicos.
7514	2.2.2.5 Registros de Acuerdos de licencia para explotación de obras de Investigación + Creación en Artes, Arquitectura y Diseño protegidas por derechos de autor.
7515	2.2.3.3 Procesos de Apropiación Social del Conocimiento.
7516	2.2.3.4 Circulación de conocimiento especializado. Eventos científicos con componentes de apropiación, participación en redes de conocimiento especializado, talleres de creación, eventos culturales y artísticos, documentos de trabajo (working papers), nueva
7517	2.2.3.5 Divulgación Pública de la CTel. Publicaciones editoriales no especializadas, producciones de contenido digital, producción de estrategias y contenido transmedia, y desarrollos web. Los requerimientos son mencionados en la Tabla XVII del ANEXO 1.
7518	2.2.3.6 Producción Bibliográfica.
7519	2.2.4.1 Direcciones de Tesis de doctorado Dirección/Tutoría y Codirección/Cotutoría de Tesis de Doctorado, se diferencian las tesis con reconocimiento de las aprobadas.
7520	Los requerimientos son mencionados en la Tabla XIX del ANEXO 1.
7521	2.2.4.2 Direcciones de Trabajo de grado de maestría Dirección/Tutoría y Codirección/Cotutoría de Trabajo de grado de maestría, se diferencian los trabajos con reconocimiento de los aprobados.
7522	2.2.4.3 Direcciones de Trabajo de pregrado Dirección/Tutoría y Codirección/Cotutoría de Trabajo de grado de pregrado, se diferencian los trabajos con reconocimiento de los aprobados.
7523	2.2.4.4 Proyectos de Investigación y Desarrollo Proyectos ejecutados (culminados) clasificados de acuerdo con las fuentes de financiación.
7524	2.2.4.4 Proyectos de Investigación-Creación Proyectos ejecutados (culminados) clasificados de acuerdo con las fuentes de financiación.
7525	2.2.4.4 Proyectos de Investigación, Desarrollo e Innovación (IDI)
7526	2.2.4.5 Proyectos de extensión y de responsabilidad social en CTel
7527	2.2.4.6 Apoyos a la creación de programas y cursos de formación de investigadores
7528	2.2.4.7 Acompañamientos y asesorías de línea temática del Programa Ondas
7529	Informe de investigación

Tipologia_Productos OK Tipo_Contratación OK Semilleros OK Rubro OK Rol_Sennova OK Tipo_Contratacion OK Grupo_investigacion OK

Referencia: <https://docs.google.com/spreadsheets/d/1gyhg5gr7NzDALxnKiTjvOZGcphKKtVhs/edit?usp=sharing&ouid=110720597759867780864&rtpof=true&sd=true>

5.4. Base de datos preprocesamiento



Upload de documentos

```
Python_SGPS 9748_Preprocesamiento.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

+ Código + Texto

Versión Python

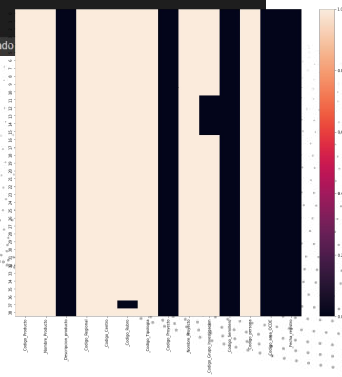
[153] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import branca
#Nube de Palabras
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
#mapas
import folium
from folium.plugins import HeatMap

[154] # upload de archivo desde Drive
from google.colab import files
path = "/content/drive/MyDrive/Proyecto SGPS 9748/Dataset FINAL.xlsx"

tipologia_productos = pd.read_excel(path,sheet_name= 'Tipologia_Productos OK')
tipo_contratacion = pd.read_excel(path,sheet_name= 'Tipo_Contratación OK')
semilleros = pd.read_excel(path,sheet_name= 'Semilleros OK')
rubro = pd.read_excel(path,sheet_name= 'Rubro OK')
rol_sennova = pd.read_excel(path,sheet_name= 'Rol_Sennova OK')
grupo_investigacion = pd.read_excel(path,sheet_name= 'Grupo_Investigación OK')
area_conocimiento_ocde = pd.read_excel(path,sheet_name= 'Area_conocimiento_OCDE OK')
principal_talento_humano = pd.read_excel(path,sheet_name= 'Principal_Talento_Humano OK')
estado_proyecto = pd.read_excel(path,sheet_name= 'Estado_Proyecto OK')
nivel_academico = pd.read_excel(path,sheet_name= 'Nivel_Academico OK')
municipio = pd.read_excel(path,sheet_name= 'Municipio OK')
centros_formacion = pd.read_excel(path,sheet_name= 'Centros_Formación OK')
departamento = pd.read_excel(path,sheet_name= 'Departamento OK')
confesion = pd.read_excel(path,sheet_name= 'Confesión OK')
```

Revisión de valores nulos

```
# Verificación de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(principal_productos.notna())
```



Exploración de la información

```
Python_SGPS 9748_Preprocesamiento.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

+ Código + Texto

[160] Nombre_tipologia
Nombre_Tipologia 0
dtype: int64

[161] #Palabras más mencionadas en tipología de productos
#Stopword list:
stopwords = set(STOPWORDS)
stopwords.update(['de', 'en', 'la', 'son', 'los', 'del', 'los', 'el', 'una', 'un', 'uno', 'las', 'con', 'Tabla',
'trabajo', 'productos', 'anexo', 'resultado', 'requerimientos', 'Mencionados'])

# Word cloud image

text = ' '.join(tipologia_productos['Nombre_Tipologia'].astype('str'))
wordcloud = WordCloud(width = 1500, height = 1000, stopwords=stopwords, max_font_size=400, max_words=100, background_color='white').generate(text)
plt.figure()
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



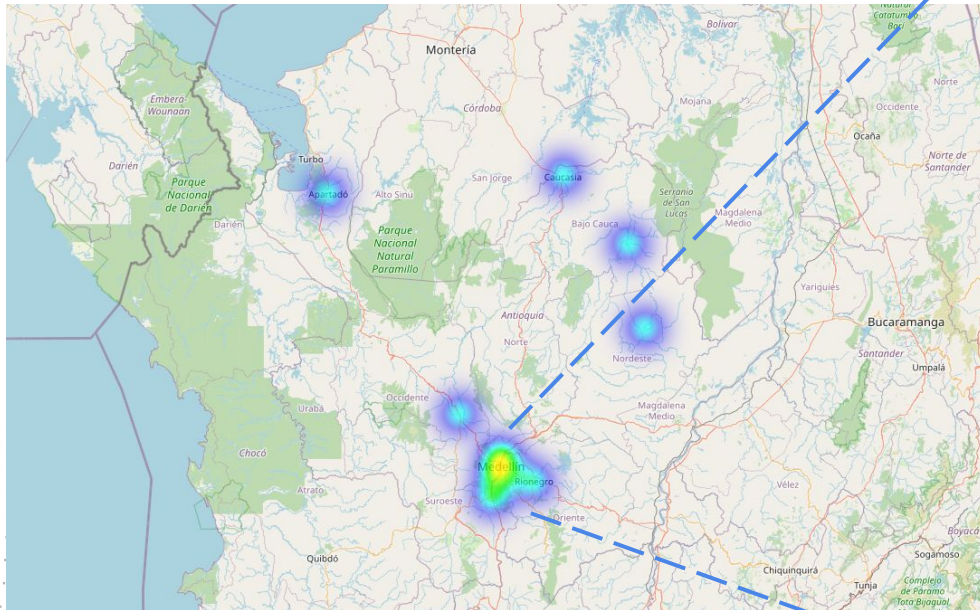
Upload de archivo desde Drive. Python_ SGPS9748

https://colab.research.google.com/drive/1ew11GfbgV9h2j-tsPUJ_VUkXRnCfR6F?usp=sharing

5.4. Base de datos preprocesamiento: Mapas Regional Antioquia



Antioquia



Centro de Servicios y Gestión Empresarial



Upload de archivo desde Drive. Python_ SGPS9748

https://colab.research.google.com/drive/1ew11IGfbqV9h2j-tsPUJ_VUkXRnCFr6F?usp=sharing

5.4. Base de datos preprocesamiento: Upload dataset Python



```
[ ] path="/content/Identificación de indicadores.xlsx"

[ ] excel_sheets("Identificación de indicadores.xlsx")

'Dataset' -> 'Datos' -> 'Indicadores' -> 'centro_formacion' -> 'regional' -> 'municipio' -> 'departamento' -> 'usuario' -> 'rol' -> 'estado' -> 'tipo_de_equipo' -> 'registro_equipos' -> 'estado_proyecto' -> 'grupo_investigacion' -> 'area_conocimiento_ocde' -> 'proyecto_principal' -> 'semillero' -> 'talento_humano' -> 'proyecto_talento_humano' -> 'nivel_academico' -> 'producto_talento_humano' -> 'productos_principal' -> 'rubro' -> 'estado_contrato' -> 'rol_de_proyectos' -> 'tipo_contrato' -> 'rol_sennova' -> 'tipologia_productos'

[ ] excel_Dataset <- read_excel("Identificación de indicadores.xlsx")

New names:
* "" -> ...17
* "" -> ...18
* "" -> ...19
* "" -> ...20
* "" -> ...21
* ...

excel_Dataset %>% str()

tibble [14 x 30] (53: tbl_df/tbl/data.frame)
 $ centro_formacion      : chr [1:14] "centro_formacion_codigo_centro" "centro_formacion_nombre_centro" "centro_formacion_municipio_codigo_municipio" NA ...
 $ regional              : chr [1:14] "regional_codigo_regional" "regional_nombre_regional" NA NA ...
 $ municipio             : chr [1:14] "municipio_codigo_municipio" "municipio_nombre_municipio" "municipio_regional_codigo_regional" "municipio_departamento_codigo_regional" NA NA ...
 $ departamento         : chr [1:14] "departamento_codigo_departamento" "departamento_nombre_departamento" NA NA ...
 $ rubro_codigo_rubro    : chr [1:14] "rubro_codigo_rubro" "rubro_nombre_rubro" NA NA ...
 $ logi                  : logi [1:14] NA NA NA NA NA NA ...
 $ registro_equipos_valor_equipo : chr [1:14] "registro_equipos_valor_equipo" "registro_equipos_tipo_de_equipo_codigo_tipo_equipo" "registro_equipos_proyecto_principal_codigo_proyecto_principal" NA NA ...
 $ talento_humano_nombre_persona : chr [1:14] "talento_humano_nombre_persona" "talento_humano_objeto_contrato" "talento_humano_duracion_contrato" "talento_humano_codigo_persona" NA NA ...
 $ semillero             : chr [1:14] "semillero_codigo_semillero" "semillero_nombre_semillero" "semillero_grupo_investigacion_codigo_grupo_investigacion" NA NA ...
 $ area_conocimiento_ocde_codigo_area_ocde : chr [1:14] "area_conocimiento_ocde_codigo_area_ocde" "area_conocimiento_ocde_nombre_area_ocde" NA NA ...
```

Prueba de *upload* de archivo y creación de *dataframes* para cada hoja de la base de datos en R desde Colab.

Upload de archivo desde Drive. R_ SGPS9748

<https://colab.research.google.com/drive/1A5MquraGf0nCTqDLXEBTzCE7P4ie1oIN?usp=sharing>

5.5. Propuesta de librerías de visualización para el front end

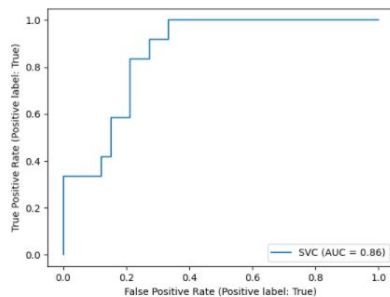
Librería Plotly Express

<https://plotly.com/python/plotly-express/>



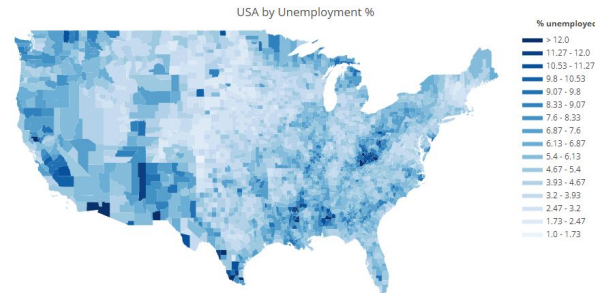
Librería Folium

<https://plotly.com/python/plotly-express/>



Librería scikit-learn

<https://scikit-learn.org/stable/>



6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo

- A. Apoyo con la identificación de la información 2021
- B. Reunión con Juan camilo Zapata
- C. Reunión con John Ferla
- D. Reunión de divulgadores del proyecto 9748 para definir estrategias de apoyo
- E. Participación en reuniones de seguimiento proceso SENNOVA - Actas
- F. Recepción de aprendices en Monitoría
- G. Definir aprendices representan el prototipo en Redcolsi - Pendiente

6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo



Universidad Politécnica
Salesiana de Ecuador

Envío de artículos
hasta: 31/07/2022
Fecha de publicación:
01/10/2022

BASES DE DATOS INTERNACIONALES SELECTIVAS



Emerging Source Citation Index



SciELO Ecuador



Google Scholar



ERIH+



Base Search



Copernicus Index



C.I.R.C.



Trueserials



EBSCO. Fuente Académica Plus



EBSCO. Business Source Ultimate



Latam+



Scientific Indexing Services



EuroPub



Journal Scholar Metrics

6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo



Universidad Politécnica
Salesiana de Ecuador

**Envío de artículos
hasta: 31/07/2022
Fecha de publicación:
01/10/2022**

Retos

© 2019, Universidad Salesiana de Ecuador

Sección (Marcar)
Dossier Monográfico _____
Miscelánea _____

Modelo conceptual para medir la Gestión de CTel aplicando analítica de datos

Conceptual model to measure CTel Management applying data analytics

Resumen

Mínimo 220 y máximo 230 palabras. Debe incluir 1) Justificación del tema; 2) Objetivos; 3) Metodología y muestra; 4) Principales resultados; 5) Principales conclusiones. Ha de estar escrito de manera impersonal. La presente investigación analiza...

La utilización de herramientas de análisis de datos en diferentes ámbitos cada día cobra más importancia, dado que permite utilizar la información generada como base para la toma de decisiones, una de las áreas en la que pueden ser utilizadas dichas herramientas es en la Gestión Ciencia, Tecnología e innovación. El presente trabajo surgió de la necesidad de proponer un modelo como apoyo para la medición del desempeño de las Actividades de Ciencia, Tecnología e Innovación (ACTI) que facilite el diseño de estrategias teniendo en cuenta la información generada y los indicadores definidos. Para ello, luego de haber realizado revisión de la literatura y realizar entrevistas a actores interesados de diferentes niveles, se construyó una modelo conceptual que incluyen diferentes indicadores y el planteamiento de la aplicación de la metodología CRISP-DM, la cual es una de las más utilizadas en laminaria de datos a nivel mundial, esta servirá de complemento al aplicativo desarrollado en el marco del proyecto "Fortalecimiento de capacidades de industria 4.0 para las vocaciones productivas de Salud, Comercio, Turismo y Ambiente del departamento de Antioquia a partir de un modelo de gobernanza TI para desarrollos de software en los Centros SENA".



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

Presentación de avances 03

Fase 2: Identificación y selección del Dataset

Proyecto SGPS 9748 - Prototipo Innovación

**Modelo de evaluación y valoración de la Gestión de CTel para los Centros SENA
Regional Antioquia aplicando analítica de datos**

15 de marzo del 2022



@SENAcomunica

www.sena.edu.co

Contenido



1. Fase 2 del modelo CRISP - DM - Proyecto SGPS-9748
2. Indicadores y correspondencia de variables de entrada y salida del modelo
3. Instalación de software
4. Organización de base de datos
5. Identificación dataframes
6. Propuesta inicial de generación de datos aleatorios
7. Dataframes
8. Base de datos preprocesamiento
9. Sugerencias (Verbal)



1. Fase 2 del 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. ENTENDIMIENTO DE LOS DATOS	2.1 Recolección inicial de datos	2.1.1 Informe inicial de recopilación de datos
	2.2 Descripción de los datos	2.2.1 Informe de descripción de datos
	2.3 Exploración de los datos	2.3.1 Informe de exploración de datos
	2.4 Verificación de calidad de los datos	2.4.1 Informe de calidad de datos
3. PREPARACION DE DATOS	3.0 Tareas preparatorias	3.0.1 Conjunto de datos 3.0.2 Descripción del conjunto de datos
	3.1 Selección de datos	3.1.1 Justificación de la inclusión / exclusión
	3.2 Limpieza de datos	3.2.1 Informe de limpieza de datos
	3.3 Construcción de datos	3.3.1 Atributos derivados 3.3.2 Registros generados
	3.4 Integración de los datos	3.4.1 Datos combinados
	3.5 Formato de datos	3.5.1 Datos reformateadas



2. Indicadores y correspondencia de variables de entrada y salida del modelo



Nro	DIMENSIÓN (Proceso de entradas y salidas)	CATEGORIA	INDICADOR (Variables)	INDICADOR FINAL				
1	Insumos para actividades de Ciencia, Tecnología e Innovación (ACTI) (Entradas en el modelo)	RECURSOS ECONÓMICOS	GASTO/ INVERSIÓN en ACTI	Suma de los diferentes rubros ejecutados.	9	Proceso (Planta del modelo)	PROYECTOS	NÚMERO DE PROYECTOS
2		RECURSOS FÍSICOS Y TECNOLÓGICOS	NÚMERO DE LABORATORIOS PARA ACTI	AMBIENTES DE APRENDIZAJE (Laboratorios dotados con recursos de SENNOVA)	10		TALENTO HUMANO	PROPORCIÓN DE GASTO EN ACTI POR PERSONAS DEDICADAS A ACTI
3			INVERSIÓN DE EQUIPOS DE PARA ACTI	Suma del rubro de equipos y maquinaria.	11		RECURSOS ECONÓMICOS	PORCENTAJE DE GASTO EN ACTI en relación con el PRESUPUESTO TOTAL
4			NÚMERO DE LICENCIAS (Campo de conocimiento según la OCDE) (Por programas de educación)	-Software de aplicaciones especializadas -Software de programación	12		RELACIONES CON EL ENTORNO	Eventos de movilidad para ACTI
5		TALENTO HUMANO	NÚMERO DE PERSONAS DEDICADAS A ACTI		13	Resultado (Salidas del modelo)	FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	Formación de aprendices en semilleros de investigación
6					14		RELACIONES CON EL ENTORNO	Convenios con otras instituciones
7		RECURSOS ESTRUCTURALES Y RELACIONALES	NÚMERO DE UNIDADES ORGANIZATIVAS (GRUPOS DE INVESTIGACIÓN - SEMILLEROS)		15		PRODUCTOS RESULTADOS DE LAS ACTIVIDADES DE CTI	NÚMERO DE PRODUCTOS DE I+D (Productos de I+D por tipología Colciencias)
8			NÚMERO DE PROGRAMAS IMPACTADOS DE EDUCACIÓN SUPERIOR OFERTADOS (PROGRAMAS DEL SENA POR ACTI)		16		TRANSFERENCIA DE CONOCIMIENTO	NÚMERO DE EVENTOS DE CTI ORGANIZADOS POR LA UNIVERSIDAD (Número de eventos de CTI por clasificación ACTI)
			NÚMERO DE CONVENIOS DE COLABORACIÓN PARA ACTI		17	Productividad (Rendimiento del modelo)	FORMACIÓN CIENTÍFICA Y TECNOLÓGICA	Número de aprendices con formación científica
					19		TALENTO HUMANO	PROPORCIÓN DE PRODUCTOS DE CTI POR PERSONA DEDICADAS A ACTI, EN RELACIÓN CON EL TOTAL DE PERSONAL
					20		RELACIONES CON EL ENTORNO	PROPORCIÓN DE PRODUCTOS DE CTI POR UNIDAD ORGANIZATIVA
					21		RELACIONES CON EL ENTORNO	PROPORCIÓN DE EVENTOS DE MOVILIDAD PARA ACTI POR PERSONAS DEDICADAS A ACTI
					22			

Identificación de indicadores: Indicadores y correspondencia de variables de entrada y salida del modelo

https://docs.google.com/spreadsheets/d/1vL_HVo1sFDqqCgYSwjY8axuVdo8fJenDftpicF_WN0/edit?usp=sharing

3. Instalación de software

Instalados

SQL-Server

SQL Management Studio

Node js

Java Runtime

Git

Anaconda

Rstudio

R

Python

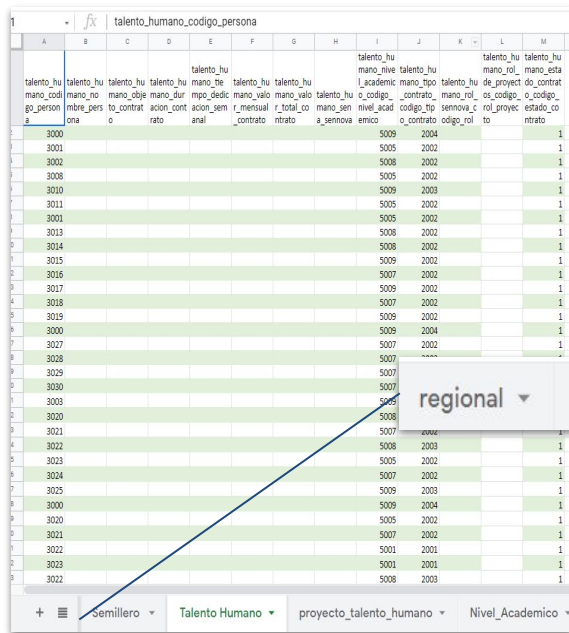


4. Organización de base de datos

Organización de base datos: Variables organizadas por hojas

Se organizaron las variables en hojas dentro de un documento de archivo de acuerdo al modelo relacional.

https://docs.google.com/spreadsheets/d/1nQzejQe2ZgNTMHWBug_-9Wgs4xHMBWve/edit?usp=sharing&oid=110720597759867780864&rtpof=true&sd=true



	A	B	C	D	E	F	G	H	I	J	K	L	M
					talento_hu				talento_hu	talento_hu			
	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu	talento_hu
	mano_codi	mano_no	mano_obje	mano_dur	mano_dedic	mano_valo	mano_valo	talento_hu	o_codigo	contrato	mano_rol	os_codigo	o_codigo
	go_persona	mbre_pers	to_contrat	acion_dur	acion_sem	r_mensual	r_total_co	mano_sen	nivel_acad	o_contrato	senovicia	rol_proyec	estado_co
	a	ona	o	rato	anal	_contrato	ntrato	a_sennova	emico	o_codigo	rol	to	ntrato
1	3000								5009	2004			1
2	3001								5005	2002			1
3	3002								5008	2002			1
4	3008								5005	2002			1
5	3010								5009	2003			1
6	3011								5005	2002			1
7	3001								5005	2002			1
8	3013								5008	2002			1
9	3014								5008	2002			1
10	3015								5009	2002			1
11	3016								5007	2002			1
12	3017								5009	2002			1
13	3018								5007	2002			1
14	3019								5009	2002			1
15	3000								5009	2004			1
16	3027								5007	2002			1
17	3028								5007				
18	3029								5007				
19	3030								5007				
20	3003								5009				
21	3020								5008				
22	3021								5007	2004			1
23	3022								5008	2003			1
24	3023								5005	2002			1
25	3024								5007	2002			1
26	3025								5009	2003			1
27	3000								5009	2004			1
28	3020								5005	2002			1
29	3021								5007	2002			1
30	3022								5001	2001			1
31	3023								5001	2001			1
32	3022								5008	2003			1

regional

centro_formacion

registro

estado_proyecto

grupo_i

4. Organización de base de datos

SENA OneDrive

Juliett Restrepo Zapata

Mis archivos

Nuestros archivos

Buscar

+ Nuevo Cargar Compartir Copiar vínculo Sincronizar Descargar Añadir acceso directo a Mis archivos

Mis archivos > SENNOVA 2022 > PROTOTIPO INNOVACIÓN SENA 2022 > Documentos 2021 > INDICADORES CTel > RECOLECCIÓN -DATOS

Nombre	Modificado	Modificado por	Tamaño de archi...	Compartir
Actas de Cierre - Centro Servicios	16 de febrero	Erika Julieta Restrepo Zapata	1 elemento	Compartido
INFORMACIÓN SENNOVA - DATOS - STEPH...	16 de febrero	Erika Julieta Restrepo Zapata	9 elementos	Compartido
Prototipo V1 - ALFA	16 de febrero	Erika Julieta Restrepo Zapata	6 elementos	Compartido
Centros de la Regional Antioquia.xlsx	16 de febrero	Erika Julieta Restrepo Zapata	18,0 KB	Compartido
Faltantes.xlsx	16 de febrero	Erika Julieta Restrepo Zapata	17,2 KB	Compartido
Informe Presupuestal SENNOVA.xlsx	1 de marzo	Paola Andrea Buitrago Carr	39,5 KB	Compartido

Referencia: [RECOLECCIÓN -DATOS](#)

5. Identificación de *dataframes* en el modelo relacional

Regionales _Codigo_Regional _Nombre_Regional _Codigo_Departamento _Codigo_Municipio	Grupo_investigacion _Codigo_Grupo_Investigacion _Nombre_Grupo_Investigacion _Categoria_Grupo _Codigo_Centro _Codigo_Regional	Semilleros _Codigo_Semillero _Nombre_Semillero _Codigo_Grupo_Investigacion _Codigo_Centro _Codigo_Regional
Rubro _Codigo_Rubro _Nombre_Rubro	Rol_Sennova _Codigo_Rol _Nombre_Rol	Tipo_Contratacion _Codigo_Tipo_Contrato _Nombre_Tipo_Contrato
Principal_Productos _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	Principal_Proyectos _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	Principal_Talento_Humano _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
Area_conocimiento_OCDE _Codigo_area_OCDE _Nombre_area_OCDE	Estado_Proyecto _Codigo_Estado_Proyecto _Nombre_Estado_Proyecto	Nivel_Academico _Codigo_Nivel_Academico _Nombre_Nivel_Academico

5. Identificación de *dataframes* en el modelo relacional

Municipio _Codigo_Municipio _Nombre_Municipio _Codigo_Departamento	Centros_Formacion _Codigo_Centro _Nombre_Centro _Codigo_Regional _Codigo_Municipio _Codigo_Departamento _latitud _longitud	Principal_Equipos _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
Tipologia_Productos _Codigo_Tipologia _Nombre_Tipologia	Tipo_Contratación _Codigo_Tipo_Contrato _Nombre_Tipo_Contrato	Tipo_equipo _Codigo_Tipo_Equipo _Nombre_Tipo_Equipo
Rol_Proyectos _Codigo_Rol_Proyecto _Nombre_Rol	Estado_Contratación _Codigo_Estado_Contratacion _Nombre_Estado_Contratacion	Departamento _Codigo_Departamento _Nombre_Departamento

6. Propuesta inicial de generación de datos aleatorios

- Generadores de datos azar. Costos de proyectos: Intervalo 30000000 - 500000000
 - <http://www.alazar.info/>
 - <http://www.alazar.info/generador-de-numeros-aleatorios>
- Generación de datos aleatorios para centros de formación dentro del intervalo de códigos ya definidos institucionalmente.
- Fechas entre el 1/01/2010 al 31/12/2022
<http://www.alazar.info/generador-de-fechas-aleatorias>

7. Dataframes



Referencia: <https://docs.google.com/spreadsheets/d/1PPSfangM7gA2CJcNDy8YNtzFcZaaXYso/edit?usp=sharing&ouid=110720597759867780864&rtpof=true&sd=true>

8. Base de datos preprocesamiento



Upload de documentos

```
Python_SGPS 9748_Preprocesamiento.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

+ Código + Texto

Versión Python

[153] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import branca
#Nube de Palabras
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
#mapas
import folium
from folium.plugins import HeatMap

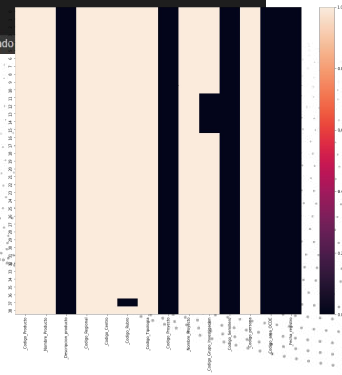
[154] # upload de archivo desde Drive
from google.colab import files
path = "/content/drive/MyDrive/Proyecto SGPS 9748/Dataset FINAL.xlsx"

tipologia_productos = pd.read_excel(path,sheet_name= 'Tipologia_Productos OK')
tipo_contratacion = pd.read_excel(path,sheet_name= 'Tipo_Contratación OK')
semilleros = pd.read_excel(path,sheet_name= 'Semilleros OK')
rubro = pd.read_excel(path,sheet_name= 'Rubro OK')
rol_sennova = pd.read_excel(path,sheet_name= 'Rol_Sennova OK')
grupo_investigacion = pd.read_excel(path,sheet_name= 'Grupo_investigacion OK')
area_conocimiento_ocde = pd.read_excel(path,sheet_name= 'Area_conocimiento_OCDE OK')
principal_talento_humano = pd.read_excel(path,sheet_name= 'Principal_Talento_Humano OK')
estado_proyecto = pd.read_excel(path,sheet_name= 'Estado_Proyecto OK')
nivel_academico = pd.read_excel(path,sheet_name= 'Nivel_Academico OK')
municipio = pd.read_excel(path,sheet_name= 'Municipio OK')
centros_formacion = pd.read_excel(path,sheet_name= 'Centros_Formacion OK')
departamento = pd.read_excel(path,sheet_name= 'Departamento OK')
confesion = pd.read_excel(path,sheet_name= 'Confesion OK')

0 s completado
```

Revisión de valores nulos

```
# Verificación de valores null
f, ax = plt.subplots(figsize=(17, 15))
ax = sns.heatmap(principal_productos.notna())
```



Exploración de la información

```
Python_SGPS 9748_Preprocesamiento.ipynb
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

+ Código + Texto

[160] Adongo tipologia = 0
Nombre_Tipologia = 0
dtype: int64

[161] #Palabras más mencionadas en tipología de productos
#Stopword list:
stopwords = set(STOPWORDS)
stopwords.update(['de', 'en', 'la', 'son', 'los', 'del', 'los', 'el', 'una', 'un', 'uno', 'las', 'con', 'Tabla',
'trabajo', 'productos', 'anexo', 'resultado', 'requerimientos', 'mencionados'])

# Word cloud image

text = ' '.join(tipologia_productos['Nombre_Tipologia'].astype('str'))
wordcloud = WordCloud(width = 1500, height = 1000, stopwords=stopwords, max_font_size=400, max_words=100, background_color='white').generate(text)
plt.figure()
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



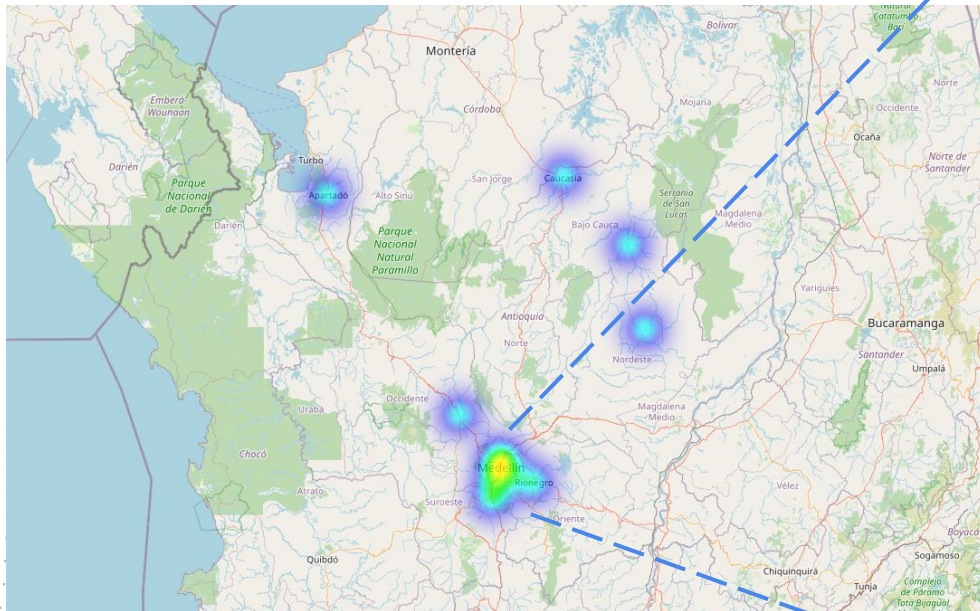
Upload de archivo desde Drive. Python_ SGPS9748

https://colab.research.google.com/drive/1ew11IGfbgV9h2j-tsPUJ_VUkXRnCrR6F?usp=sharing

8. Base de datos preprocesamiento: Mapas Regional Antioquia



Antioquia



Centro de Servicios y Gestión Empresarial



Upload de archivo desde Drive. Python_ SGPS9748

https://colab.research.google.com/drive/1ew11IGfbqV9h2j-tsPUJ_VUkXRnCFr6F?usp=sharing

5.4. Base de datos preprocesamiento: Upload dataset R



```
[ ] path="/content/Identificación de indicadores.xlsx"

[ ] excel_sheets("Identificación de indicadores.xlsx")

'Dataset' -> 'Datos' -> 'Indicadores' -> 'centro_formacion' -> 'regional' -> 'municipio' -> 'departamento' -> 'usuario' -> 'rol' -> 'estado' -> 'tipo_de_equipo' -> 'registro_equipos' -> 'estado_proyecto' -> 'grupo_investigacion' -> 'area_conocimiento_ocde' -> 'proyecto_principal' -> 'semillero' -> 'talento_humano' -> 'proyecto_talento_humano' -> 'nivel_academico' -> 'producto_talento_humano' -> 'productos_principal' -> 'rubro' -> 'estado_contrato' -> 'rol_de_proyectos' -> 'tipo_contrato' -> 'rol_sennova' -> 'tipologia_productos'

[ ] excel_Dataset <- read_excel("Identificación de indicadores.xlsx")

New names:
* "" -> ...17
* "" -> ...18
* "" -> ...19
* "" -> ...20
* "" -> ...21
* ...

excel_Dataset %>% str()

tibble [14 x 30] (53: tbl_df/tbl/data.frame)
 $ centro_formacion      : chr [1:14] "centro_formacion_codigo_centro" "centro_formacion_nombre_centro" "centro_formacion_municipio_codigo_municipio" NA ...
 $ regional              : chr [1:14] "regional_codigo_regional" "regional_nombre_regional" NA NA ...
 $ municipio              : chr [1:14] "municipio_codigo_municipio" "municipio_nombre_municipio" "municipio_regional_codigo_regional" "municipio_departamento_codigo_regional" NA NA ...
 $ departamento          : chr [1:14] "departamento_codigo_departamento" "departamento_nombre_departamento" NA NA ...
 $ gasto_inversion_ACTI   : chr [1:14] "rubro_codigo_rubro" "rubro_nombre_rubro" NA NA ...
 $ numero de laboratorios para ACTI : logi [1:14] NA NA NA NA NA NA ...
 $ inversion de equipos para ACTI    : chr [1:14] "registro_equipos_valor_equipo" "registro_equipos_tipo_de_equipo_codigo_tipo_equipo" "registro_equipos_proyecto_principal_codigo_proyecto_principal" NA NA NA ...
 $ numero de licencias (Campo de conocimiento segun la OCDE) (Por programas de educación) : logi [1:14] NA NA NA NA NA NA ...
 $ proporción de gasto en ACTI por personas dedicadas a ACTI : chr [1:14] "talento_humano_codigo_persona" "talento_humano_nombre_persona" "talento_humano_objeto_contrato" "talento_humano_duracion_contrato" NA NA NA ...
 $ semillero              : chr [1:14] "semillero_codigo_semillero" "semillero_nombre_semillero" "semillero_grupo_investigacion_codigo_grupo_investigacion" NA NA ...
 $ numero de programas impactados de educación superior ofertados (programas del sena por ACTI) : chr [1:14] "area_conocimiento_ocde_codigo_area_ocde" "area_conocimiento_ocde_nombre_area_ocde" NA NA ...
```

Prueba de *upload* de archivo y creación de *dataframes* para cada hoja de la base de datos en R desde Colab.

Upload de archivo desde Drive. R_ SGPS9748

<https://colab.research.google.com/drive/1A5MquraGf0nCTqDLXEBTzCE7P4ie1oIN?usp=sharing>

6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo

- A. Apoyo con la identificación de la información 2021
- B. Reunión con Juan camilo Zapata
- C. Reunión con John Ferla
- D. Reunión de divulgadores del proyecto 9748 para definir estrategias de apoyo
- E. Participación en reuniones de seguimiento proceso SENNOVA - Actas
- F. Recepción de aprendices en Monitoría
- G. Definir aprendices representan el prototipo en Redcolsi - Pendiente

6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo

Retos
Revista de Ciencias de la Educación y la Comunicación

Universidad Politécnica
Salesiana de Ecuador

Envío de artículos
hasta: 31/07/2022
Fecha de publicación:
01/10/2022

BASES DE DATOS INTERNACIONALES SELECTIVAS



Emerging Source Citation Index



SciELO Ecuador



Google Scholar



ERIH+



Base Search



Copernicus Index



C.I.R.C.



Trueserials



EBSCO. Fuente Académica Plus



EBSCO. Business Source Ultimate



Latam+



Scientific Indexing Services



EuroPub



Journal Scholar Metrics

6. Actividades de apoyo a SENNOVA - Divulgación de resultados del prototipo



Universidad Politécnica
Salesiana de Ecuador

**Envío de artículos
hasta: 31/07/2022
Fecha de publicación:
01/10/2022**

Retos

© 2019, Universidad Salesiana de Ecuador

Sección (Marcar)
Dossier Monográfico _____
Miscelánea _____

Modelo conceptual para medir la Gestión de CTel aplicando analítica de datos

Conceptual model to measure CTel Management applying data analytics

Resumen

Mínimo 220 y máximo 230 palabras. Debe incluir 1) Justificación del tema; 2) Objetivos; 3) Metodología y muestra; 4) Principales resultados; 5) Principales conclusiones. Ha de estar escrito de manera impersonal. La presente investigación analiza...

La utilización de herramientas de análisis de datos en diferentes ámbitos cada día cobra más importancia, dado que permite utilizar la información generada como base para la toma de decisiones, una de las áreas en la que pueden ser utilizadas dichas herramientas es en la Gestión Ciencia, Tecnología e innovación. El presente trabajo surgió de la necesidad de proponer un modelo como apoyo para la medición del desempeño de las Actividades de Ciencia, Tecnología e Innovación (ACTI) que facilite el diseño de estrategias teniendo en cuenta la información generada y los indicadores definidos. Para ello, luego de haber realizado revisión de la literatura y realizar entrevistas a actores interesados de diferentes niveles, se construyó una modelo conceptual que incluyen diferentes indicadores y el planteamiento de la aplicación de la metodología CRISP-DM, la cual es una de las más utilizadas en laminaria de datos a nivel mundial, esta servirá de complemento al aplicativo desarrollado en el marco del proyecto "Fortalecimiento de capacidades de industria 4.0 para las vocaciones productivas de Salud, Comercio, Turismo y Ambiente del departamento de Antioquia a partir de un modelo de gobernanza TI para desarrollos de software en los Centros SENA".

4 BUSQUEDA DE INFORMACIÓN

Configura



Buscadores

Son programas dentro de un sitio o página web, los cuales, al ingresar palabras claves, operan dentro de la base de datos del mismo buscador y recopilan todas las páginas posibles, que contengan información relacionada con la que se busca.

Google (www.google.com)

Yahoo (www.yahoo.com)

Bing (www.bing.com)

Ask (www.ask.com)

Gigablast (www.gigablast.com)

Carrot2 (<http://search.carrot2.org>)

Meta-buscadores

Son herramientas de búsqueda que actúan sobre algunos de los buscadores, dirigiendo las consultas a otros buscadores facilitando al máximo y haciendo más eficiente la búsqueda de información.

Izito (www.izito.com)

Ixquick (www.ixquick.com)

Metacrawler (www.metacrawler.com)

PolyMeta (www.polymeta.com)

eTools (www.etools.ch)

iSeek (www.iseek.com)

Exaled (www.exaled.com)

GC-F-004V.01

/ 2:05:52



Monitoreo de Procesos



Monitoreo de Procesos



Monitoreo de Procesos



Monitoreo de Procesos

Tranferencia de conocimientos a aprendices - monitores de excelencia SENNOVA



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 TÉCNICAS



ACTA No. 22		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 02 de marzo de 2022	HORA INICIO: 3:00 p.m.	HORA FIN: 4:30 p.m.
LUGAR Y/O ENLACE: Virtual TEAMS https://bit.ly/3JhXv6B	DIRECCIÓN GENERAL / REGIONAL / CENTRO Centro de Servicios y Gestión Empresarial	
AGENDA O PUNTOS PARA DESARROLLAR: 1. Consulta de la generación de los indicadores 2. Definición del modelo relacional		
OBJETIVO(S) DE LA REUNIÓN: Realizar consultas sobre la generación de los indicadores y del modelo relacional del prototipo innovación – Proyecto SGPS 9748		
DESARROLLO DE LA REUNIÓN		
<p>En la reunión el equipo de trabajo del prototipo tuvo la oportunidad conversar con Juan Camilo Zapata, profesional que trabajó durante la vigencia 2021 en la investigación y desarrollo del prototipo, por esta razón se le realizaron consultas acerca de los siguientes temas:</p> <p>1. Consulta de la generación de los indicadores Se revisó el listado de indicadores y Juan camilo Zapata preciso que los indicadores fueron identificados viendo a SENNOVA como un área que tiene unas entradas o insumos, un proceso, salidas y tasa de rendimiento, se concluyó que en la lista de indicadores (Anexo 46) en la columna de indicadores se consideran variables, esta pregunta se aclaró debido a que se requiere saber de qué manera medir los indicadores. Se aclararon las dudas e inquietudes al respecto de la forma en la que se identificaron y establecieron los indicadores.</p> <p>2. Definición del modelo relacional En cuanto al modelo relacional y la forma en la que se conecta con los indicadores propuestos, se mencionó que en el proceso no se llevó a cabo esta relación, Juan Camilo Zapata sugiere partir del aplicativo que se desarrolló el año pasado e identificar las variables que allí se incluyeron dado que en el transcurso se fueron agregando variable que no fueron actualizados en el archivo de indicadores.</p> <p>Se aclararon las dudas con respecto al archivo que se exportó a Excel de SQL Server, ya que hay unos campos vacíos y se aclaró que están así por la dificultad que se presentó para recolectar la información, además se explicó la finalidad de cada uno de los nombres de los campos y la relación entre ellos.</p> <p>Adicionalmente, Juan Camilo Zapata Giraldo, explicó el esquema para adquisición de la información y sus elementos más relevantes, como proyectos, recursos y formación ACTI. También se compartió una de las principales referencias que se tuvo en la identificación de los indicadores</p> <p>(https://books.google.com.co/books/about/Medici%C3%B3n_de_Capacidades_de_Ciencia_Tecn.html?id=YMGKDwAAQBAJ&printsec=frontcover&source=kp_read_button&hl=es-419&redir_esc=v#v=onepage&q&f=false)</p>		



Se sugiere realizar una reunión con Edgar Fernando Parra Ortega, Stephania Aragón Rojas y Paola Andrea Cárdenas Velásquez, quienes pueden tener acceso a la información que se requiere.

CONCLUSIONES

Se dio respuesta a las inquietudes del equipo de trabajo y quedó abierta la posibilidad de realizar una nueva reunión que permita dar más claridad con respecto a la información generada en el proyecto.

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

ASISTENTES: (Incorporar registro de asistencia)



Seguimiento Proyecto SGPS 9748 Prototipo Innovación - Consultas Modelo

01:33:15

Está grabando Esta grabando esta reunión. Asegúrese de que todos sepan que se les está grabando. [Política de privacidad](#) Descartar

Participantes

Invíte a alguien o marque un número

Compartir invitación

En esta reunión (6)

- Erika Juliet Restrepo Zapata
- Jesus Arbey Benavides Guevara
- John Ferla (Invitado)
- Juan Camilo Zapata Giráldez
- Julian Alberto Patino Murillo
- Paola Andrea Buitrago Cadavid

Otros invitados (2)

- Johnairo Castro Maldonado
- Jhon Alexis Ferla Perez

Otros usuarios del chat (1)

- john.ferla10 (Invitado)

Chat de la reunión

Le grabador de la reunión. Esta reunión se está grabando. Al cerrar, de su consentimiento para que se graba. Política de privacidad

Microsoft Teams

La grabación de la reunión. Esta reunión se está grabando. Al cerrar, de su consentimiento para que se graba. Política de privacidad

Paola Andrea Buitrago Cadavid

Escriba aquí para buscar

Correo: Erika Juli... X Página principal X Listado de Asiste... X Página principal X 01. Reuniones - C X Ver "Herramienta... X Recordings X

https://senad-my.sharepoint.com/personal/erestrepoz_sena_edu_co/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Ferestrepoz_sena_edu_co%2F...

Compartir Copiar vínculo Descargar Eliminar Copiar en Historial de versiones 1 de 1

INVERSIÓN DE EQUIPOS DE PAE

Proceso (Pantalla del modelo)

Entradas Proceso Salidas

Modelo (tasa rendimiento)

Resultado (Salidas del modelo)

Productividad (Rendimiento del modelo)

1:19:07 / 1:33:51

Escriba aquí para buscar

25°C

4:17 p. m. 15/03/2022

Enlace grabación:
[Seguimiento Proyecto SGPS 9748 Prototipo Innovación - Consultas Modelo-20220302 150155-Grabación de la reunión.mp4](#)



REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No- 22 DEL DÍA 02 DEL MES DE MARZO DEL AÑO 2022

OBJETIVO (S) Realizar consultas sobre la generación de los indicadores y del modelo relacional del prototipo innovación – Proyecto SGPS 9748											
No.	Autorizo el tratamiento de la Información SI / NO. Si marca No por favor no diligencie el formato.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPACIÓN VIRTUAL
1	Si	Juan Camilo Zapata Giraldo	71526818			Invitado Externo	SURA	jczapatag@sena.edu.co	3006734537	Si	Virtual Teams
2	Si	Julián Alberto Patiño Murillo	8102363		X		SENA - Centro de Servicios y Gestión Empresarial	japatino@sena.edu.co	3013004624	Si	Virtual Teams
3	Si	Paola Andrea Buitrago Cadavid	43641579		X		SENNOVA	pabuitragoc@sena.edu.co	3113671040	Si	Virtual Teams
4	Si	Jesús Arbey Benavides Guevara	1085260713		X		SENA	jabenavides@sena.edu.co	3105595765	Si	Virtual Teams
5	Si	John Alexis Ferla Pérez	1036621874		X		Sena - Cesge	john.ferla10@gmail.com	3146864277	Si	Virtual Teams
6	Si	Erika Juliett Restrepo Zapata	43973982		X		Centro de Servicios y Gestión Empresarial	erestrepoz@sena.edu.co	3005017253	Si	Virtual Teams



ACTA No. 025		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 04 de marzo de 2022	HORA INICIO: 9:00 a.m.	HORA FIN: 10:00 a.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. Análisis del estado actual del prototipo 2. Estrategias de visualización alternas a Power Bi 3. Roles de trabajo en el prototipo		
OBJETIVO(S) DE LA REUNIÓN: Realizar reunión de planeación técnica del desarrollo del prototipo de innovación proyecto SGPS 9748		
DESARROLLO DE LA REUNIÓN		
1. Análisis del estado actual del prototipo Se observa el estado actual del prototipo desarrollado en la vigencia 2021. Se evalúan las diferentes etapas relevantes para la continuidad en la vigencia actual. Como primer paso, se evalúa la base de datos generada a partir del modelo relacional. Con respecto a esto, el ingeniero John Ferla menciona la importancia de determinar las variables críticas que determinarán las características del diseño del mockup y dashboard del prototipo trabajado en 2021.		
2. Estrategias de visualización alternas a Power Bi Se plantea la posibilidad de nuevas estrategias de visualización por medio de librerías de python: Plotly Express: Es un módulo que contiene funciones que pueden crear figuras dinámicas que permiten una óptima visualización para la representación de modelos desde una perspectiva de estadística descriptiva e inferencial en pro de los resultados obtenidos desde el machine learning. https://plotly.com/python/plotly-express/		

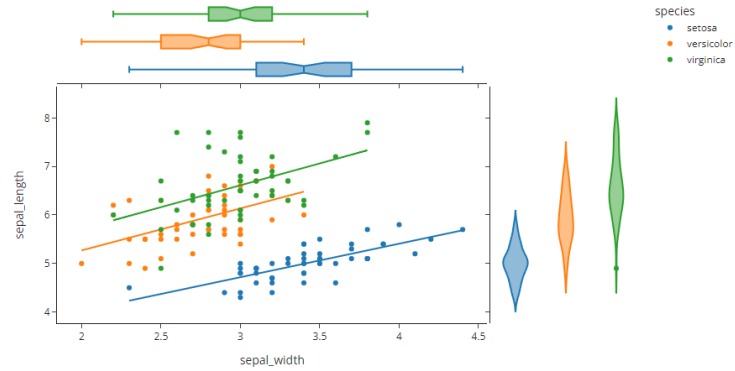


Figura 1. Gráfica ejemplo de visualización de datos.

Folium: Es una librería de python ampliamente utilizada para la visualización de datos en zonas geográficas específicas. En el futuro permitirá dar una forma visual de métodos de clustering para centros respectivos de investigación.

<https://www.cursosgis.com/visualizando-datos-en-un-mapa-con-el-paquete-de-python-folium/>

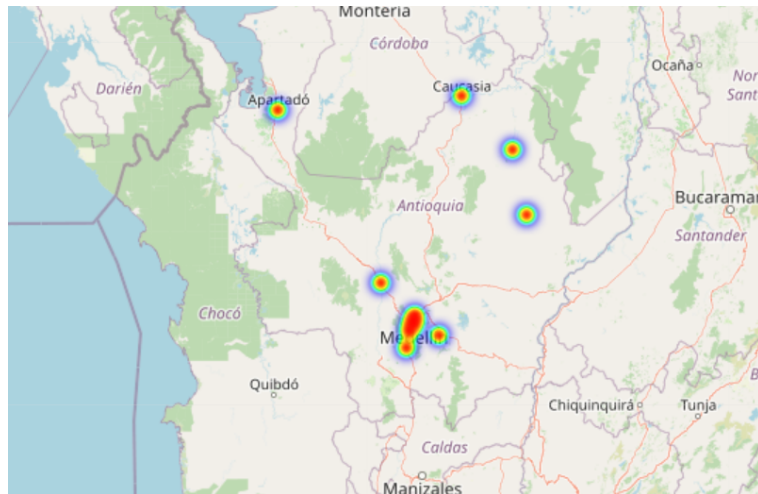


Figura 2. Mapa de representación de centros de investigación en el departamento de Antioquia



3. Roles de trabajo en el prototipo

Se determinan los roles de trabajo respectivos a la parte de analítica de datos así como del backend y frontend del prototipo de innovación. Para ello el ingeniero Ferla se compromete en la revisión de los prototipos anteriormente desarrollados y en una nueva propuesta adecuada a las necesidades actuales del proyecto. Por otro lado, desde el equipo de analítica conformado por Paola Buitrago y Jesus Benavides se plantea una idea posibles librerías como soporte en paralelo con PowerBi.

CONCLUSIONES

- Se observa la necesidad la importancia de la evaluación de la base datos registrada en pasados periodos, así como la generación de datos sintéticos que simulan de manera adecuada modelos analíticos en pro de cuantificar los indicadores definidos desde una perspectiva teórica, así como las variables críticas que definirán el desarrollo de una propuesta en modelos de machine learning.
- Se discutió posibles alternativas a PowerBi que permitan aumentar el rendimiento entre la etapa de software con la parte analítica, por medio de librerías que permiten generar visualizaciones dinámicas tales como Plotly Express y Folium de python.
- Se definen roles, entre los que se acopla el trabajo de el ingeniero John Ferla en lo respectivo al desarrollo de plataforma y dashboard en paralelo con el trabajo por parte de los investigadores Paola Buitrago y Jesus Benavides.

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
	Erika Restrepo	04 03 2022
	Paola Buitrago	04 03 2022



	John Ferla	04 03 2022
	Jesus Arbey Benavides Guevara	04 03 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)											
Reunión técnica de planeación del desarrollo - prototipo de innovación proyecto SGRS 9748											
No.	Autorizo el tratamiento de la Información SI / NO. Si marca No por favor no diligencie el formato.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT. SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPACIÓN VIRTUAL
1	Si	Jesus Arbey Becavides Guerra	1089260713		X		Senca	jahenavides @senca.edu.co	3105545765	Si	Jesus Arbey B
2	Si	John Alexis ferla Perez	1036621874		X		Senca	Jaferla@ senca.edu.co	3146664277	Si	Jaferla
3	Si	Erika Juliette Restrepo Zapata	43973782		X		SENA	erestrepoz @senca.edu.co	3005072253	Si	Erika Zapata



ACTA No. 22		
NOMBRE DEL COMITÉ O DE LA REUNIÓN: Seguimiento Proyecto SGPS 9748 - Prototipo Innovación		
CIUDAD Y FECHA: Medellín, 09 de marzo de 2022	HORA INICIO: 2:00 p.m.	HORA FIN: 4:30 p.m.
LUGAR Y/O ENLACE: Virtual TEAMS https://bit.ly/3lqvFqb	DIRECCIÓN GENERAL / REGIONAL / CENTRO Centro de Servicios y Gestión Empresarial	
AGENDA O PUNTOS PARA DESARROLLAR: 1. Fase 2 del modelo CRISP - DM - Proyecto SGPS-9748 2. Diagrama de flujo del modelo CRISP - DM 3. Modelo conceptual 4. Instalación de software 5. Base de datos 6. Sugerencias		
OBJETIVO(S) DE LA REUNIÓN: Realizar consultas sobre la generación de los indicadores y del modelo relacional del prototipo innovación – Proyecto SGPS 9748		
DESARROLLO DE LA REUNIÓN		
1. Fase 2 del modelo CRISP - DM - Proyecto SGPS-9748 Se presentó el avance del modelo CRISP – DM que luego de realizar la comprensión del negocio, se continua con la comprensión y preparación de los datos, en este aspecto se realizó el análisis de la relación entre los indicadores, la forma de medirlo y la correspondencia con las variables de entrada y salida del modelo.		
2. Diagrama de flujo del modelo CRISP - DM Se presentó el diagrama de flujo con el paso a paso de la metodología CRISP - DM aplicada al modelo de medición de indicadores de Ciencia, Tecnología e Innovación, el cual fue dividido en 5 paso, comprensión de los datos, preparación de los datos, modelado de los datos, evaluación de modelos e interpretación de resultados y por último análisis general del modelo.		

Metodología CRISP-DM aplicada al modelo de medición de indicadores de CTel



Primer paso: Comprensión de datos

Tipo de característica: nominal, numérica, etc.
Escalamiento de datos
Estadísticas básicas para detectar anomalías
Correlaciones entre características

Segundo paso: Preparación del conjunto de datos

Reconstrucción de datos faltantes
Limpieza de datos
Definición de filtros de umbral

Tercer paso: Modelado de los datos

Manipulación de datos
Creación de características, mediante filtros
Dividir el conjunto de datos en uno de prueba y otro de entrenamiento
Entrenar el modelo de regresión
Validar el modelo de regresión

Cuarto paso: Evaluación de modelos e interpretación de resultados

Interpretar las métricas generadas por el modelo
Reevaluar modelo si el resultado no es el esperado

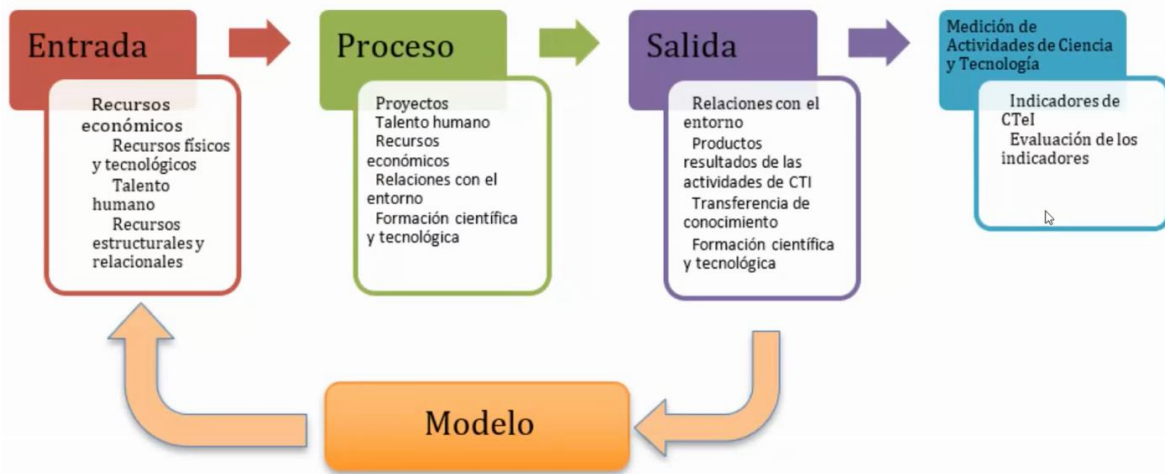
Quinto paso: Análisis general del modelo

En este paso es importante volver a mirar el modelo y si algo más se puede agregar es el momento de tener el panorama general del modelo.

3. Modelo conceptual

En cuanto al modelo conceptual se recordó las cuatro líneas programáticas que se medirán con el modelo que se está trabajando, dichas líneas son Apropiación de la cultura (Divulgación, fomento de la innovación en centros y empresas, investigación y modernización. Además, se socializó el análisis del modelo conceptual donde se presentan entradas, procesos y salidas para llegar a la medición de las actividades de Ciencia y Tecnología.

3. Modelo conceptual



4. Instalación de software

Por otro lado, se socializó que los investigadores Paola Buitrago y Jesús Benavides gestionaron la instalación de los siguientes softwares que son requeridos para el desarrollo del modelo de analítica del prototipo: SQL- Server, SQL Management Studio, Node JS, Java Runtime, GIT, anaconda, Rstudio, R y Python.

5. Base de datos

En cuanto al trabajo con la base de datos se realizó la organización de base de datos, se extrajo la base de datos SQL, se identificó la información y variables a tener en cuenta en el dataset, además de los ajustes que se deben realizar en elementos como los productos, se presentó una base de datos preprocesamiento de datos, propuesta de librerías de visualización para desarrollar el front end.

De las tablas identificadas se sugiere iniciar con principal_productos, principal_proyectos, principal_talento_humano y principal_equipos, con las que se definirían como dataframe para la analítica inicial de datos. Se sugirió a los investigadores para uniformidad de la información se adicione de manera aleatoria las tablas principales definidas como dataframes.

CONCLUSIONES

Se continúa avanzando con la aplicación de la metodología CRISP – DM para diseñar el modelo de analítica de datos, se delimitaron algunas variables para delimitarlas y agilizar el diseño del modelo a utilizar en el prototipo.

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



SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANÁLITICA DE DATOS

25:57

La grabación se ha iniciado. Esta reunión se está grabando. Al unirse, da su consentimiento para que se grabe. [Política de privacidad](#)

Descartar

1. Indicadores y correspondencia de variables de entrada y salida del modelo

Identificación de indicadores: Indicadores y correspondencia de variables de entrada y salida del modelo
https://docs.google.com/spreadsheets/d/1y1_HVo1aFDdgCgY5aY8axuVocofJenDlqpf_WN0/edit?usp=sharing

Jesus Arbey Benavides Guevara

Participantes

Invíte a alguien o marque un número

Compartir invitación

En esta reunión (4)

Silenciar a todos

Erika Juliet Restrepo Zapata

Jesus Arbey Benavides Guevara

John Jairo Castro Maldonado
Organizador

Paola Andrea Buitrago Cadavid (...)
Invitado de la reunión

Otros invitados (1)

Julian Alberto Patino Murillo
Aceptado

SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANÁLITICA DE DATOS

01:08:10

La grabación se ha iniciado. Esta reunión se está grabando. Al unirse, da su consentimiento para que se grabe. [Política de privacidad](#)

Descartar

2. Diagrama de flujo del modelo CRISP - DM

Metodología CRISP-DM aplicada al modelo de medición de indicadores de CTEI

Primer paso: Comprensión de datos

Segundo paso: Preparación del conjunto de datos

Tercer paso: Entendimiento de los datos

Cuarto paso: Evaluación de modelos e interpretación de resultados

Quinto paso: Análisis general del modelo

Participantes

Invíte a alguien o marque un número

Compartir invitación

En esta reunión (4)

Silenciar a todos

Erika Juliet Restrepo Zapata

Jesus Arbey Benavides Guevara

John Jairo Castro Maldonado
Organizador

Paola Andrea Buitrago Cadavid (...)
Invitado de la reunión

Otros invitados (1)

Julian Alberto Patino Murillo
Aceptado



SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS

02:34:26

La grabación se ha iniciado. Esta reunión se está grabando. Al unirse, da su consentimiento para que se grabe. [Política de privacidad](#)

Descartar

Participantes

Invite a alguien o marque un número

Compartir invitación

En esta reunión (4)

Silenciar a todos

Erika Julieta Restrepo Zapata

Jesus Arbey Benavides Guevara

John Jairo Castro Maldonado

Paola Andrea Buitrago Cadavid

Invitado de la reunión

Otros invitados (1)

Julian Alberto Patino Murillo

Aceptado

5.3. Identificación inicial de variables

Regionales Codigo_Regional Nombre_Regional Codigo_Departamento Codigo_Municipio	Grupo_investigacion Codigo_grupo_investigacion Nombre_grupo_investigacion Categoria_grupo Codigo_Centro Codigo_Regional	Semilleros Codigo_Semillero Nombre_Semillero Codigo_grupo_investigacion Codigo_Centro Codigo_Regional
Rubro Codigo_Rubro Nombre_Rubro	Rol_Sennova Codigo_Rol Nombre_Rol	Tipo_Contratacion Codigo_Tipo_Contrato Nombre_Tipo_Contrato
Principal_Productos Codigo_Producto Nombre_Producto Descripcion_producto Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Tipologia Codigo_grupo_investigacion Nombre_grupo_investigacion Codigo_Semillero Codigo_Persona Codigo_area_OCDE Fecha_registro	Principal_Proyectos Codigo_Proyecto Nombre_Proyecto Valor_Proyecto Codigo_Rubro Fecha_inicio_Proyecto Fecha_termino_Proyecto Codigo_Regional Codigo_Centro Codigo_estado_Proyecto Codigo_area_OCDE Serial_equipo Valor_Servicios_Personales Valor_Compra_Equipos Valor_Software Valor_Otros_Gastos	Principal_Talento_Humano Codigo_Persona Codigo_Nivel_Academico Codigo_Tipo_Contrato Codigo_grupo_investigacion Codigo_Rubro Codigo_estado_Contratacion Codigo_Regional Codigo_Centro Codigo_Proyecto Codigo_Producto
Area_conocimiento_OCDE Codigo_area_OCDE Nombre_area_OCDE	Estado_Proyecto Codigo_estado_Proyecto Nombre_estado_Proyecto	Nivel_Academico Codigo_Nivel_Academico Nombre_Nivel_Academico

Jesus Arbey Benavides Guevara

Escribe aquí para buscar

25°C

4:39 p. m.

9/03/2022

Enlace grabación:

[SEGUIMIENTO DE ACTIVIDADES DE PROYECTO DE INDICADORES DE CTEI CON ANALÍTICA DE DATOS-20220309_192131-Meeting Recording.mp4](#)



REGISTRO DE ASISTENCIA Y APROBACIÓN DEL ACTA No- 31 DEL DÍA 09 DEL MES DE MARZO DEL AÑO 2022

OBJETIVO (S) Realizar seguimiento a la ejecución de actividades del proyecto SGPS 9748 – Prototipo Innovación.											
No.	Autorizo el tratamiento de la Información SI / NO. Si marca No por favor no diligencie el formato.	NOMBRES Y APELLIDOS	No. DOCUMENTO	PLANTA	CONTRATISTA	OTRO ¿CUAL?	DEPENDENCIA/ EMPRESA	CORREO ELECTRÓNICO	TELÉFONO/EXT.SENA	AUTORIZA GRABACIÓN	FIRMA O PARTICIPACIÓN VIRTUAL
1	Si	John Jairo Castro Maldonado	88265552	X			SENA - Centro de Servicios y Gestión Empresarial	jcastrom@sena.edu.co	3204185552	Si	Virtual Teams
2	Si	Paola Andrea Buitrago Cadavid	43641579		X		SENNOVA	pabuitragoc@sena.edu.co	3113671040	Si	Virtual Teams
3	Si	Jesús Arbey Benavides Guevara	1085260713		X		SENA	jabenavides@sena.edu.co	3105595765	Si	Virtual Teams
4	Si	Erika Juliett Restrepo Zapata	43973982		X		Centro de Servicios y Gestión Empresarial	erestrepoz@sena.edu.co	3005017253	Si	Virtual Teams