Índice

[Introducción: 2](#_Toc154187161)

[Importando librerías 2](#_Toc154187162)

[Importación de bibliotecas: 2](#_Toc154187163)

[Creación de las funciones 3](#_Toc154187164)

[Estimados de localización y variabilidad 3](#_Toc154187165)

[Introducción a la visualización de datos: Distribuciones 5](#_Toc154187166)

[Exploración de variables categóricas y análisis multivariable 5](#_Toc154187167)

[Correlaciones y Regresión Linear Simple 5](#_Toc154187168)

[Distribuciones Muestrales y Técnicas de Evaluación de Modelos 5](#_Toc154187169)

[Visualización de datos avanzada 5](#_Toc154187170)

[Pruebas A/B y Procesamiento de Lenguaje Natural 5](#_Toc154187171)

[Introducción a Machine Learning: Clasificación No Supervisada y Supervisada 5](#_Toc154187172)

[Conclusión 5](#_Toc154187173)

[Bibliografía 5](#_Toc154187174)

# Introducción:

Realice este proyecto sola, lo hice de temática de la astronomía porque desde el 2016 pertenezco a un club de astronomía que realiza todos los años a nivel nacional observaciones de los astros soy joven voluntaria se llama: “Noche de las estrellas” y cada año nos hacen preguntas de los planetas y otros astros mezcle mi hobbie con este proyecto.

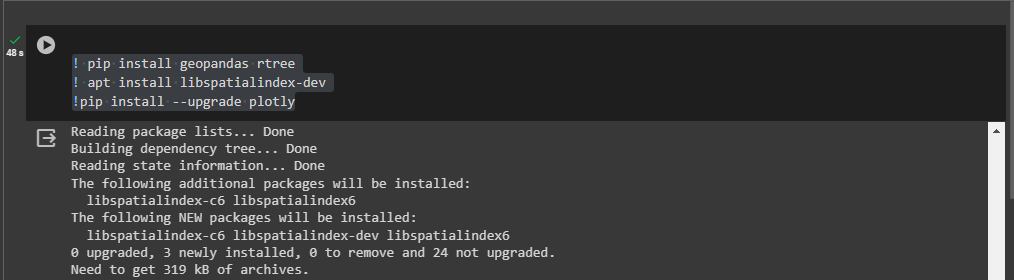
# Importando librerías

!pip install --upgrade plotly

! pip install geopandas rtree

! apt install libspatialindex-dev

!pip install --upgrade plotly



# Importación de bibliotecas:

import pandas as pd

import geopandas as gpd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy import stats

import plotly.express as px

from sklearn.linear\_model import LinearRegression

from sklearn.linear\_model import LogisticRegression

from sklearn.model\_selection import train\_test\_split

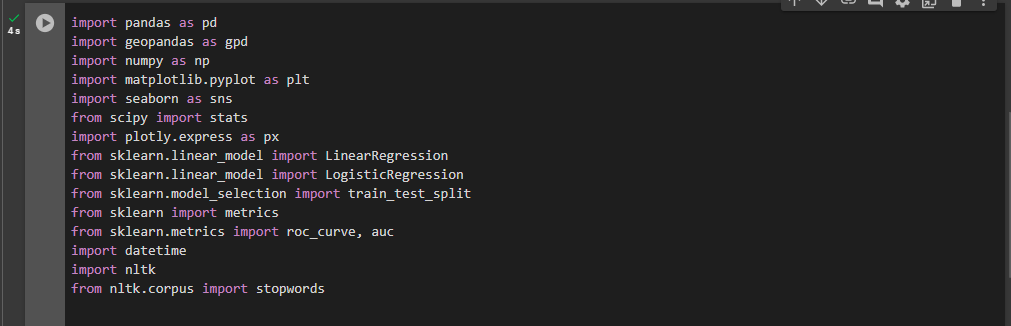
from sklearn import metrics

from sklearn.metrics import roc\_curve, auc

import datetime

import nltk

from nltk.corpus import stopwords



# Creación de las funciones

# Mensaje de bienvenida

print("Programa con información de astronomía\n")

# (Numero de entrada) Planeta a analizar

saturno="Total de lunas"

# Suma de las lunas

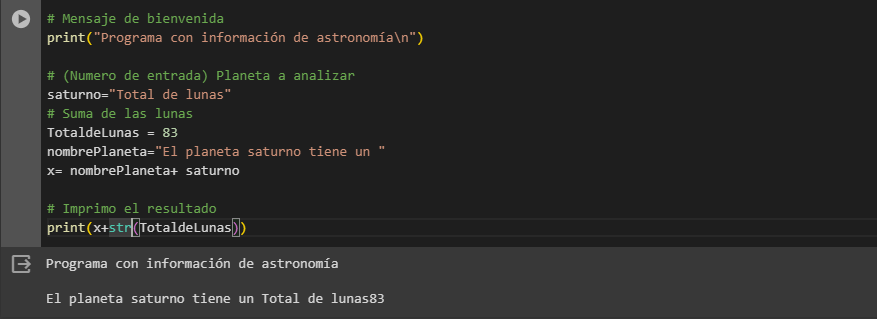
TotaldeLunas = 83

nombrePlaneta="El planeta saturno tiene un "

x= nombrePlaneta+ saturno

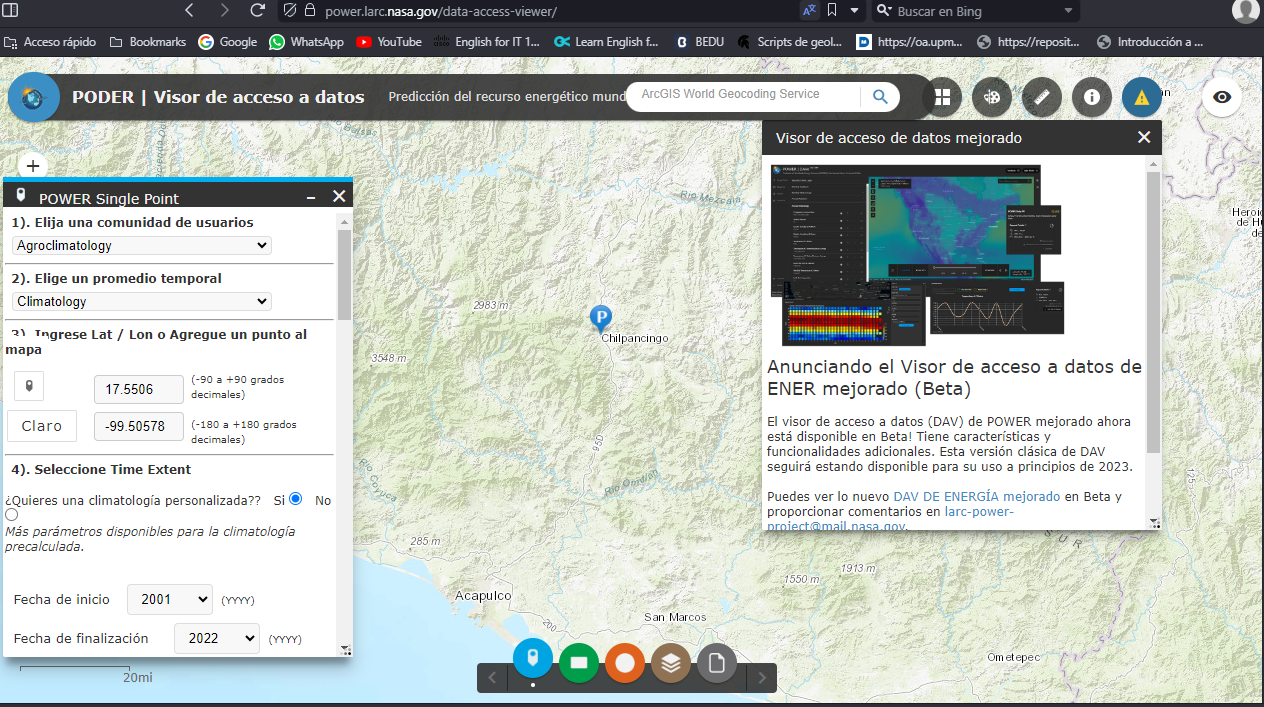
# Imprimo el resultado

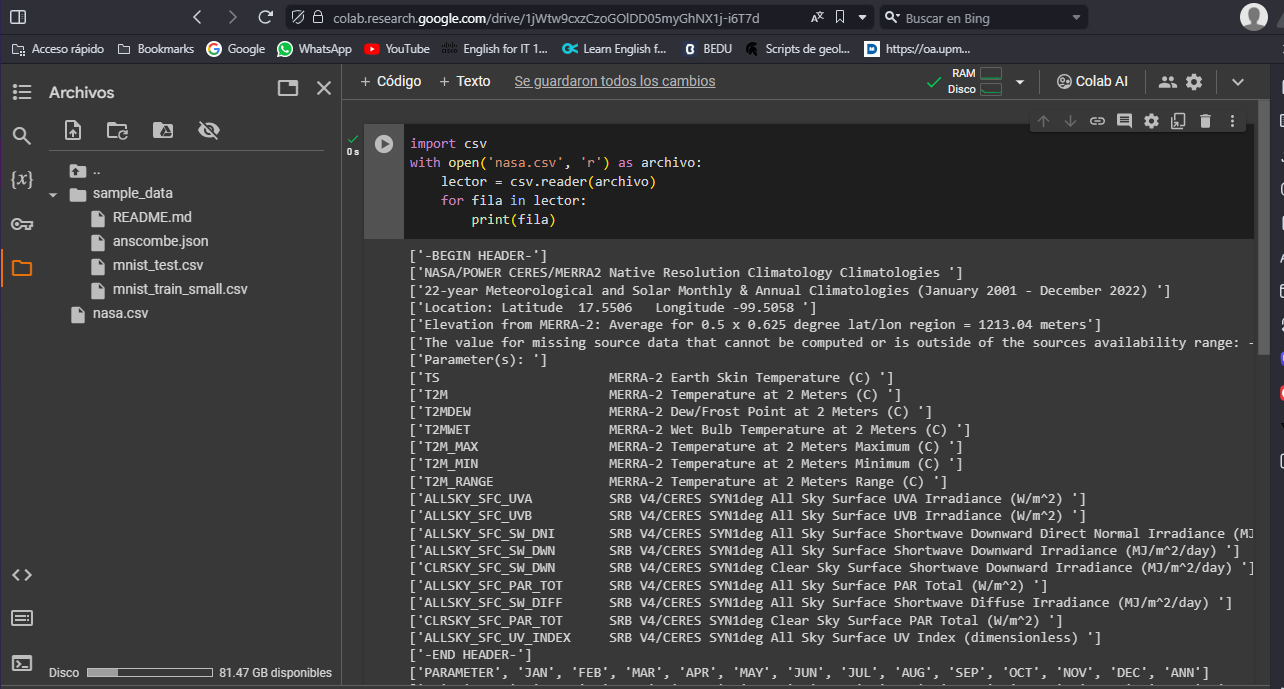
print(x+str(TotaldeLunas))



# Estimados de localización y variabilidad

Se tomaron los datos de mi cuidad de origen en la bibliografía anexo la página de donde se me genero el archivo .csv





import csv

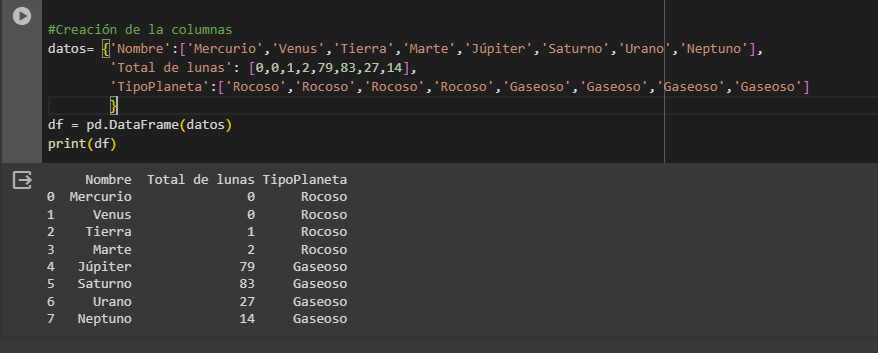
with open('nasa.csv', 'r') as archivo:

    lector = csv.reader(archivo)

    for fila in lector:

        print(fila)

# Introducción a la visualización de datos: Distribuciones



import pandas as pd

#Creación de la columnas

datos= {'Nombre':['Mercurio','Venus','Tierra','Marte','Júpiter','Saturno','Urano','Neptuno'],

        'Total de lunas': [0,0,1,2,79,83,27,14],

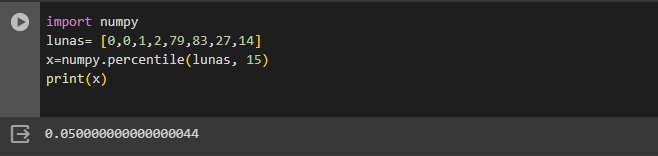
        'TipoPlaneta':['Rocoso','Rocoso','Rocoso','Rocoso','Gaseoso','Gaseoso','Gaseoso','Gaseoso']

        }

df = pd.DataFrame(datos)

print(df)

**Percentiles en base a las lunas**



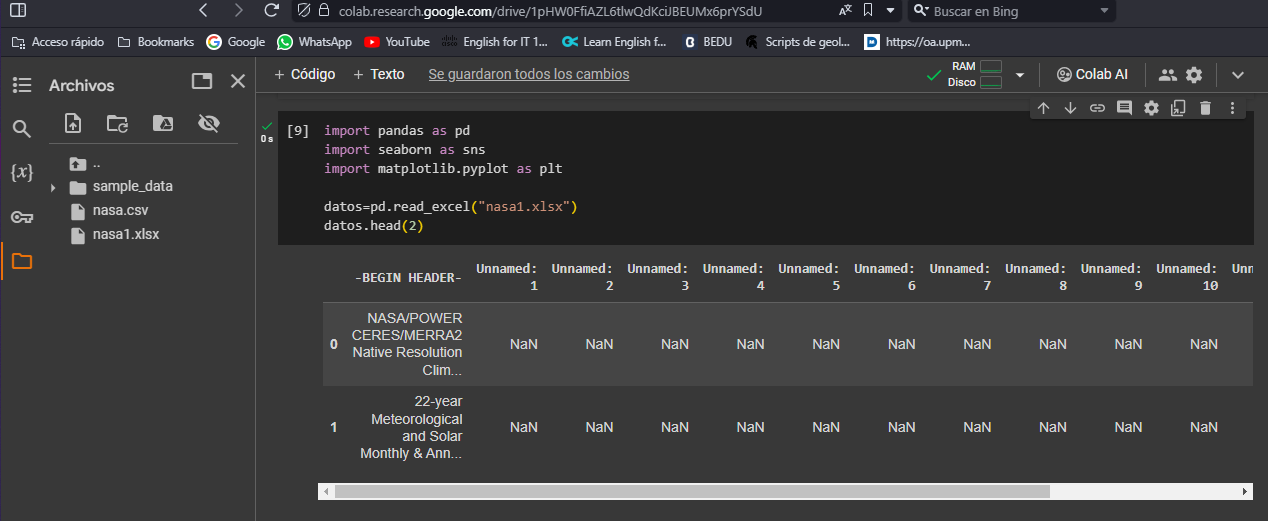
import numpy

lunas= [0,0,1,2,79,83,27,14]

x=numpy.percentile(lunas, 15)

print(x)

Tomando los 2 primeros datos del Excel



import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

datos=pd.read\_excel("nasa1.xlsx")

datos.head(2)

Uso de boxplots

import matplotlib.pyplot as plt

plt.figure(figsize=[6,3])

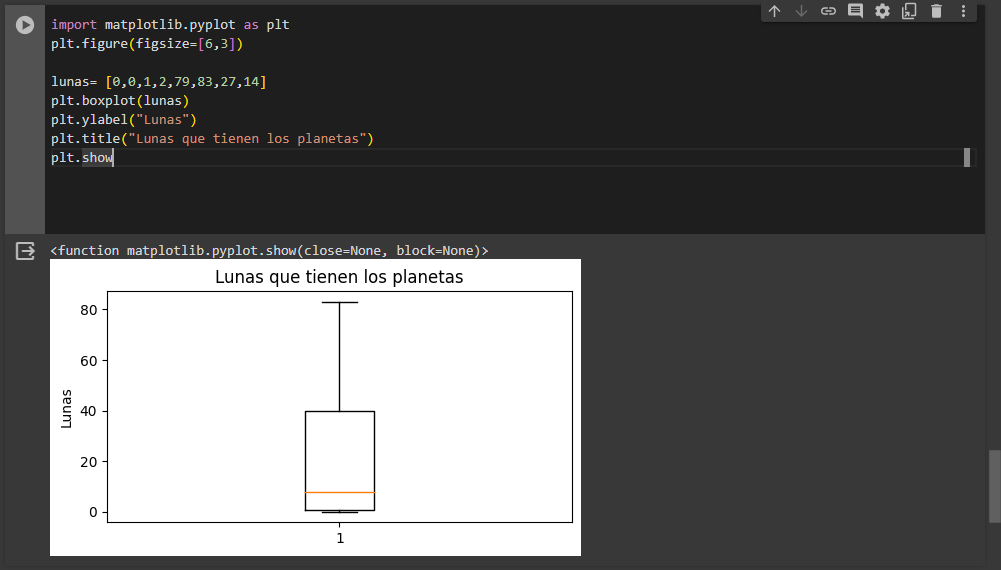
lunas= [0,0,1,2,79,83,27,14]

plt.boxplot(lunas)

plt.ylabel("Lunas")

plt.title("Lunas que tienen los planetas")

plt.show



# Exploración de variables categóricas y análisis multivariable

Bueno en está parte me confundi un poco y no pude hacer una gráfica con mis datos asi que me base e la documentación de Python para hacer el ejemplo

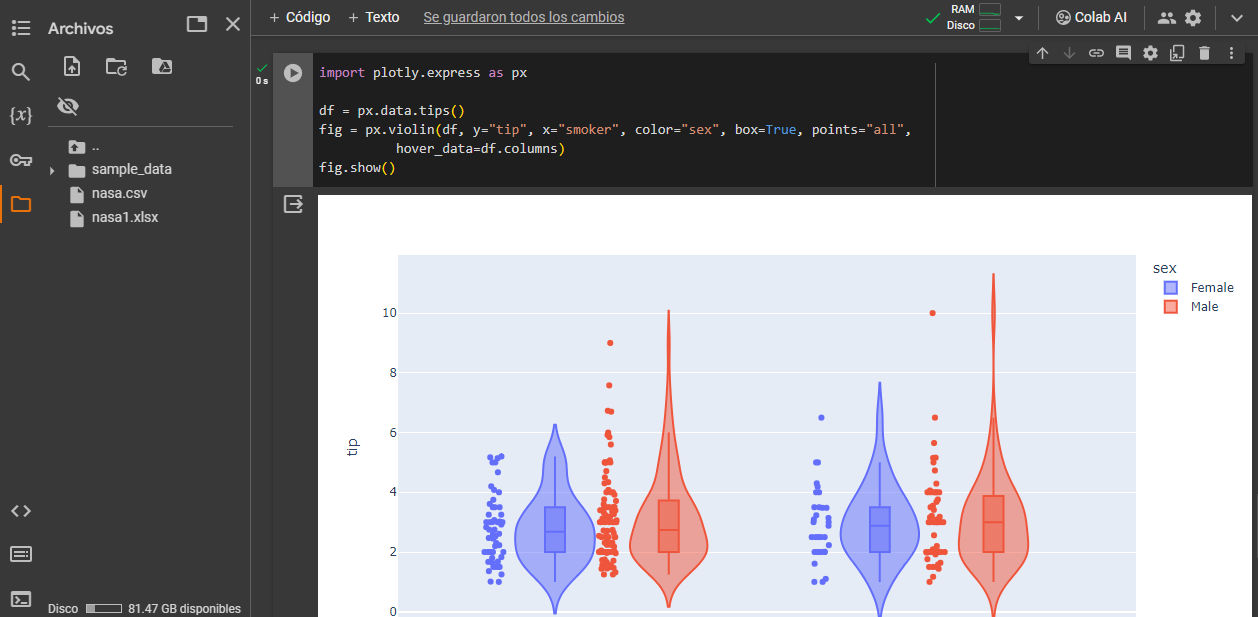
import plotly.express as px

df = px.data.tips()

fig = px.violin(df, y="tip", x="smoker", color="sex", box=True, points="all",

hover\_data=df.columns)

fig.show()



# Correlaciones y Regresión Linear Simple

import matplotlib.pyplot as plt

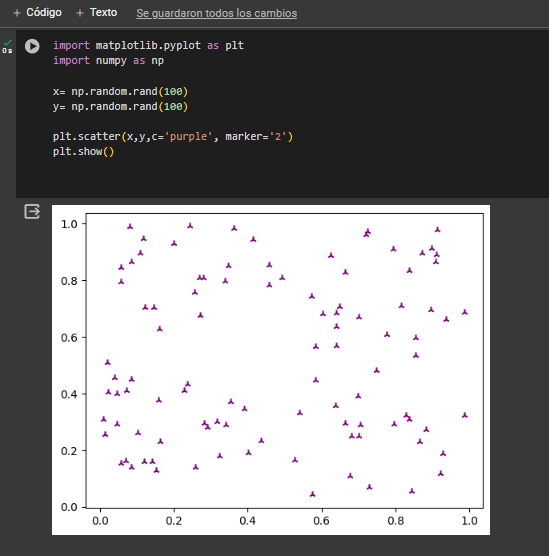
import numpy as np

x= np.random.rand(100)

y= np.random.rand(100)

plt.scatter(x,y,c='purple', marker='2')

plt.show()



# Distribuciones Muestrales y Técnicas de Evaluación de Modelos

# Visualización de datos avanzada

# Pruebas A/B y Procesamiento de Lenguaje Natural

# Introducción a Machine Learning: Clasificación No Supervisada y Supervisada

# Conclusión

Algunos temas se me complicaron me pondré a estudiar para poder ejecutarlos mejor

# Bibliografía

<https://www.nationalgeographic.com.es/ciencia/asi-son-8-planetas-sistema-solar_18432>

<https://www.geodatos.net/coordenadas/mexico/chilpancingo>

<https://power.larc.nasa.gov/data-access-viewer/>