



TECNOLÓGICO
NACIONAL DE MÉXICO



SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA

INSTITUTO TECNOLÓGICO DE HERMOSILLO

MATERIA:
MINERÍA DE DATOS

PROFESOR:
EDUARDO ANTONIO HINOJOSA PALAFOX

ALUMNO:
KEVIN RICARDO VILLAGRÁN MORALES

PROYECTO FINAL:
AUMENTO DE TEMPERATURA EN MEXICO DESDE 1835

GRUPO:
S7C

UNIDAD 5

CARRERA:
ING. EN SISTEMAS COMPUTACIONALES

HERMOSILLO, SONORA.

14 DE DICIEMBRE 2022

Introducción

En este documento implementaremos y desarrollaremos algunos de los métodos vistos durante el curso de la materia, con la ayuda de google colab y Python.

Python es un lenguaje de programación interpretado de tipado dinámico cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multiparadigma y disponible en varias plataformas. Detecta los errores en tiempo de ejecución.

Google Colaboratory es una plataforma en línea de Google que permite la creación, ejecución, almacenamiento y compartición de programas en Python en espacios de trabajo amigables y sencillos de utilizar. Con estas mencionadas herramientas elaboraremos el siguiente problema.

Tema: Aumento de la temperatura en México desde 1835

Conjunto de Datos: <https://www.kaggle.com/code/luisferfranco/exploring-mexico-temperature-changes-from-1835/notebook>

Predicciones: Verificar si hay un aumento en la temperatura de manera gradual o enorme

Descripción del problema a desarrollar

Durante años y determinada por varios datos verídicos sobre el aumento de la temperatura del planeta, que esto ha ocasionado cambios en algunas zonas geográficas del mundo, para este problema nos enfocaremos solo en el aumento de temperatura de México, en la plataforma Kaggle se encontró un documento con varios archivos que denotan el aumento de la temperatura de los países desde 1835, por lo que con la ayuda de la minería de datos, se dará una observación de esos datos de manera más gráfica para una mayor comprensión y apreciación.

Con esto se busca dar a conocer que el aumento de la temperatura ha incrementado de una manera exponencial y los promedios de dichas temperaturas.

Descripción del conjunto de datos

Como se mencionó anteriormente en la página web Kaggle, se buscó un archivo de un conjunto de datos relacionados sobre el tema, tras haber indagado en varios archivos que son referentes al tema, se considera que este conjunto de datos es el más apropiado para aplicar estas observaciones ya que abarca un gran rango y además es de los más precisos que se han encontrado.

Descripción de la solución propuesta.

Para este problema se busca dar a conocer los cambios de temperatura que han ido incrementando durante varios años en México, con apoyo de Python se mostrarán

una graficas que darán un mejor entendimiento de estos, aplicando métodos comparativos, promedios y agrupación.

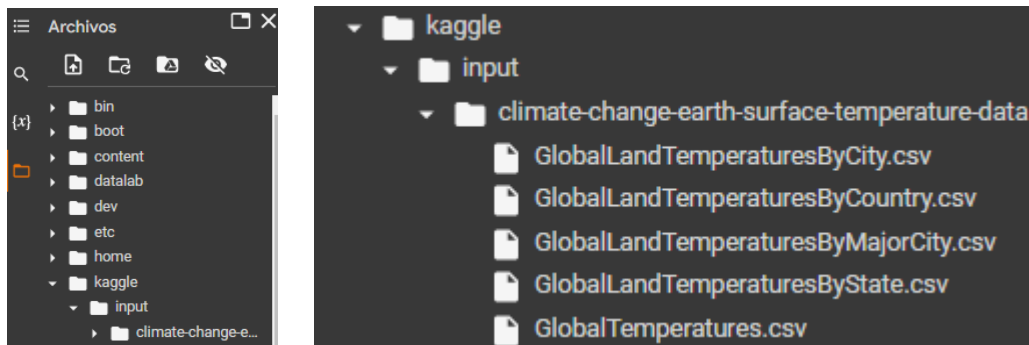
Descripción del código utilizado

Paso 1: Abrimos un nuevo cuaderno en google colab e importamos la librerías correspondientes

```
import numpy as np
import pandas as pd
import datetime
import math
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

Paso 2: En el apartado de explorador de archivos salimos de la carpeta content y crearemos una carpeta que se llame Kaggle y dentro de esa carpeta crearemos una que se llame input y dentro de esa importaremos la carpeta climate-change-earth-surface-data aquí es donde importaremos los 5 csv.



Paso 3: Una vez importados los conjuntos de datos, como en este problema contamos con varios archivos creamos este método para llamar a todos los archivos importados y a su vez guardarlos.

```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

/kaggle/input/climate-change-earth-surface-temperature-data/GlobalLandTemperaturesByCity.csv
/kaggle/input/climate-change-earth-surface-temperature-data/GlobalLandTemperaturesByMajorCity.csv
/kaggle/input/climate-change-earth-surface-temperature-data/GlobalLandTemperaturesByState.csv
/kaggle/input/climate-change-earth-surface-temperature-data/GlobalLandTemperaturesByCountry.csv
/kaggle/input/climate-change-earth-surface-temperature-data/GlobalTemperatures.csv
```

Paso 4: Ahora llamamos el archivo de GlobalLandTemperaturesByCountry.csv, junto a ello implementamos el método de dataParser para establecer los valores de las fechas basado en cadenas de las fechas del documento.

```
dateParser = lambda x: pd.datetime.strptime(x, '%Y-%m-%d')

clima = pd.read_csv('/kaggle/input/climate-change-earth-surface-temperature-data/GlobalLandTemperaturesByCountry.csv',
                    parse_dates=['dt'],
                    date_parser=dateParser,
                    low_memory=False
                    )
```

Paso 5: Verificamos que se el conjunto de datos este correctamente.

```
clima.head()
```

	dt	AverageTemperature	AverageTemperatureUncertainty	Country
0	1743-11-01	4.384	2.294	Åland
1	1743-12-01	NaN	NaN	Åland
2	1744-01-01	NaN	NaN	Åland
3	1744-02-01	NaN	NaN	Åland
4	1744-03-01	NaN	NaN	Åland

Paso 6: De manera estética cambiamos el nombre de las columnas para una mayor comprensión.

```
clima.columns = ['fecha', 'temp', 'incertidumbre', 'pais']
```

Paso 7: Hacemos un conteo del total de datos que hay en el conjunto de datos.

```
clima.fecha.describe()
```

```
<ipython-input-63-3a129c708011>
clima.fecha.describe()
count          577462
unique          3239
top      2013-09-01 00:00:00
freq           243
first    1743-11-01 00:00:00
last     2013-09-01 00:00:00
Name: fecha, dtype: object
```

Paso 8: Como nuestro objetivo es Mexico vamos a mostrar todos los datos que correspondan con él.

```
mexico = pd.DataFrame(clima[clima.pais == 'Mexico'])
mexico.head()
```

	fecha	temp	incertidumbre	pais
338438	1835-01-01	15.810	1.484	Mexico
338439	1835-02-01	14.911	1.579	Mexico
338440	1835-03-01	16.826	1.858	Mexico
338441	1835-04-01	19.925	1.966	Mexico
338442	1835-05-01	22.037	1.609	Mexico

Paso 9: Ahora para obtener unos resultados más precisos, separaremos las fecha en año, mes y década, con ello será más fácil el muestreo de los datos de manera gráfica.

```
mexico['year'] = pd.DatetimeIndex(mexico['fecha']).year
mexico['mes'] = pd.DatetimeIndex(mexico['fecha']).month

mexico['decada'] = (pd.DatetimeIndex(mexico.fecha).year-1)/10
mexico.decada = mexico.decada.apply(lambda x: math.floor(x)*10)

mexico.head()
```

	fecha	temp	incertidumbre	pais	year	mes	decada
338438	1835-01-01	15.810	1.484	Mexico	1835	1	1830
338439	1835-02-01	14.911	1.579	Mexico	1835	2	1830
338440	1835-03-01	16.826	1.858	Mexico	1835	3	1830
338441	1835-04-01	19.925	1.966	Mexico	1835	4	1830
338442	1835-05-01	22.037	1.609	Mexico	1835	5	1830

Paso 10: Ahora para representar los datos de manera gráfica implementamos el plt.figure y agregamos los parámetros que queremos mostrar, como queremos aplicarlo solo a Mexico agregamos un inicio y final, para finalizar agregamos las etiquetas.

```
plt.style.use('fivethirtyeight')

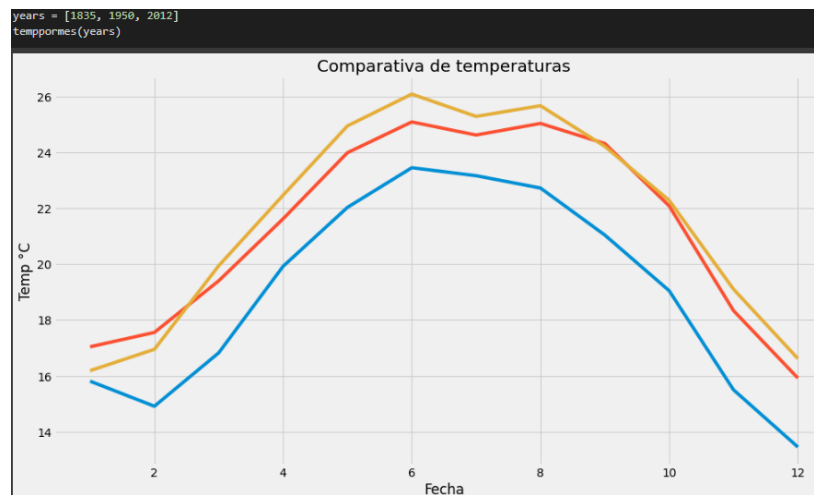
def tempormes(rango):
    plt.figure(figsize=(15, 8))
    for year in rango:
        ini = datetime.datetime.strptime('{}-01-01'.format(year), '%Y-%m-%d')
        fin = datetime.datetime.strptime('{}-12-31'.format(year), '%Y-%m-%d')

        rango = mexico[(mexico.fecha >= ini) & (mexico.fecha <= fin)]

        plt.plot(rango.mes, rango.temp, label=year)

    plt.title('Comparativa de temperaturas')
    plt.xlabel('Fecha')
    plt.ylabel('Temp °C')
    print()
```

Paso 11: Mostramos los años que queremos comparar, en este caso 1835 que es el primer año que contiene datos y así mismo con los demás años y con esto podemos observar la gráfica.



La línea azul representa los años 1835, naranja 1950 y amarillo 2012

Paso 12: Ahora una vez mostrado esta gráfica, mostraremos el aumento de la temperatura mediante el agrupamiento de su promedio anual y década.

```

meant = mexico.groupby(['decada'])['temp'].mean()

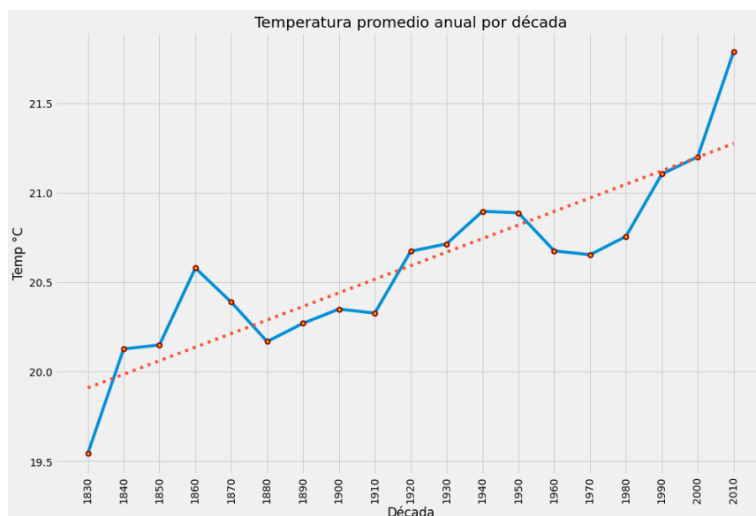
plt.figure(figsize=(15, 10))
plt.title('Temperatura promedio anual por década')
x = np.array(range(1, int(np.floor((2010-1820)/10)+1)))

coef = np.polyfit(x, meant, 1)
poly1d_fn = np.poly1d(coef)

plt.plot(x, meant, marker='o', mec='darkred', mfc='orange', mew=2)
plt.plot(x, poly1d_fn(x), ls=":")

plt.xticks(x, np.arange(1830, 2020, 10), rotation=90)
plt.xlabel('Década')
plt.ylabel('Temp °C')
print()

```



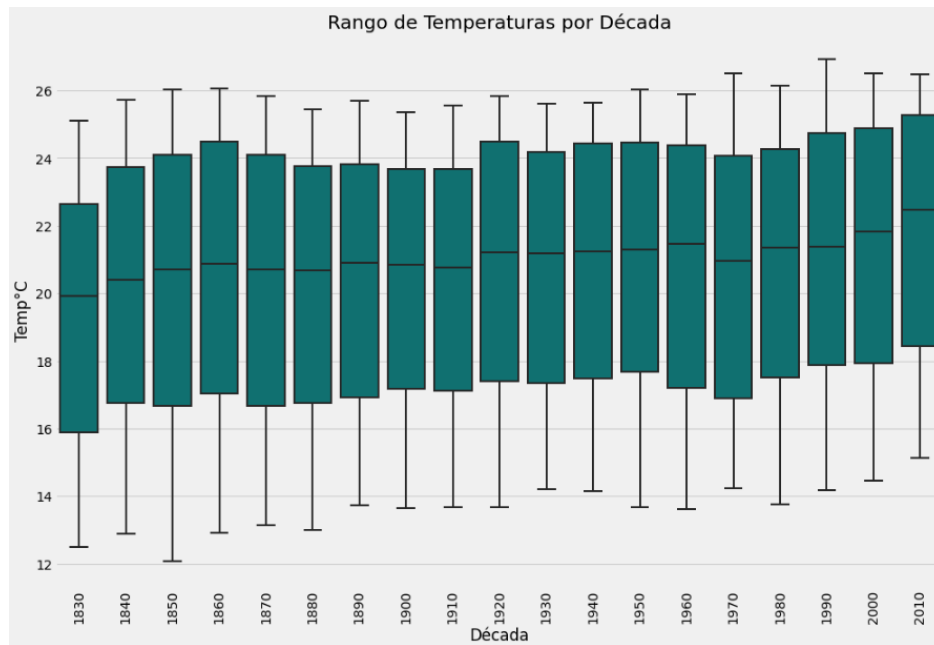
Paso 13: Ahora mostraremos la misma gráfica pero mediante el boxplot

```

plt.figure(figsize=(15, 10))
g = sns.boxplot(x='decada', y='temp', data=mexico, color="teal", linewidth=2)
g.set_xticklabels(g.get_xticklabels(), rotation=90)

g.set_xlabel("Década")
g.set_ylabel("Temp°C")
g.set_title('Rango de Temperaturas por Década')
print()

```



Resultados

Según los métodos aplicados obtenemos los resultados del aumento de temperatura que el mínimo de temperatura ha ido aumentado exponencialmente ya que como se muestra en la última gráfica del paso 13, en 1830 la mínima ha sido de 16 grados y en 2010 ha sido de 18 grados.

Ha estos resultados hay que tomarles un importante impacto según la contaminación que se puede estar generando en el planeta, por lo que el aumento de estos grados es consecuente a que se genere más calor por lo que afecta a la población.

Conclusiones

Para concluir con este proyecto, la ejecución de varios métodos aprendidos durante al curso dejan claramente como la ilustración de conjunto de datos y con los métodos necesarios podemos representar y analizar los datos de mejor manera.

Además la minería de datos nos deja como enseñanza que cada valor puede ser administrable, así mismo puede valer y a su vez no, podemos contemplarlo en algún esquema si es que se necesita contemplar, para ello es la minería de datos que se tenga una buena manipulación de los datos.

Referencias

- luisferfranco. "Exploring Mexico temperature changes from 1835". Kaggle: Your Machine Learning and Data Science Community. <https://www.kaggle.com/code/luisferfranco/exploring-mexico-temperature-changes-from-1835/notebook> (accedido el 14 de diciembre de 2022).