[Fecha]

Ruben Ceballos Ramírez

Proyecto Final

PDT23-VER-Aprendiendo Python 3 - intermedio

# Introduction

Este Proyecto está basado en el aprendizaje que **PDT23-VER-Aprendiendo Python 3 – Intermedio.** La importancia de la ciencia de datos en estos tiempos es de mucha importancia, nos ayuda en la toma de decisiones y al hablar de estos tiempos nos encontramos que la toma de decisiones debe de ser en el momento ya que de esta manera uno puede salvarse de una mala decisión o bien tomar una muy buena decisión, muchos se preguntaran en que puede ser aplicado esto, pues puede ser aplicado en varios sectores, por ejemplo:

1. Manufactura, toma de decisiones para ver si estamos cumpliendo con los objetivos.
2. Salud, hace un tiempo nos enfrentamos en una pandemia.
3. Bienes Raíces, detectar y analizar si estamos logrando las metas.
4. Publico, al momento de darnos cuenta como está el sector.

Hay muchas aplicaciones del análisis de datos y esta herramienta proporciona como puede uno llegar a alcanzar los retos que uno busca. En el menor tiempo posible.

# Proyecto Profeco Quejas

El proyecto que hice tiene que ver con las quejas de cliente, una curiosidad que en ocasiones nos pasa por la mente como cuantas quejas hay y que compañías son las que tienen más quejas encontré la manera de buscar las quejas de las personas. Revisando en la página de la misma.

https://datos.profeco.gob.mx/datos\_abiertos/'

De aquí parte la búsqueda de la información que me llevaría a encontrar las quejas haciendo una búsqueda del archivo mas actual tomo la información más nueva (El archivo al parecer contiene 600,000 registros) con lo cual en milisegundo hace un análisis y encuentra el top 10 de todas las compañías con mayores problemas de servicio al cliente y muestro una gráfica la cual hace ver de una manera muy sencilla los resultados del análisis de datos.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Es de una manera más sencilla de encontrar las empresas con mayores problemas de servicio al cliente por medio de los reportes.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Las tecnologías de Python que se utilizaron fueron las siguientes,

import requests

from bs4 import BeautifulSoup

import pandas as pd

import matplotlib.pyplot as plt

Herramientas que se utilizaron de Microsoft,

Visual Studio Code Icono

Descripción generada automáticamente

Extensiones de Visual Studio Code.

1. Jupyter
2. Python

# Desarrollo del Proyecto

Iniciamos instalando **request** y **BeautifulSoup**

import requests

from bs4 import BeautifulSoup

Posteriormente declaramos una variable con la dirección donde tomaremos la información.

url = 'https://datos.profeco.gob.mx/datos\_abiertos/'

Iniciamos con el GET, con el cual buscaremos donde se encuentra el link o los links,

response = requests.get(url)

#Creamos un objeto beautifulsoup

soup = BeautifulSoup(response.text, 'html.parser')

#Buscamos los elementos que estaremos utilizando para iniciar una descarga posterior

enlaces\_bases\_de\_datos = soup.find\_all('a', href=True)

#file.php es nuestro identificador

enlaces\_filtrados = [enlace['href'] for enlace in enlaces\_bases\_de\_datos if 'file.php?t=' in enlace['href']]

nombre\_archivo = soup.find\_all('a', string=lambda string: string and 'BASE DE DATOS ABIERTOS' in string)

for link in nombre\_archivo:

    nombre = link.get\_text(strip=True) + '.xlsx'

    print(nombre)

    break

#Imprimimos resultados que seran nuestros links

for enlace in enlaces\_filtrados:

    print(enlace)

    break

#preparando la descarga

response = requests.get(url+enlace)

if response.status\_code == 200:

    # Guardar el contenido del archivo en disco

    with open(nombre, 'wb') as file:

        file.write(response.content)

    print(f"Archivo {nombre} descargado exitosamente.")

else:

    print("No se pudo descargar el archivo.")

Estos son o deberán de ser los resultados

BASE DE DATOS ABIERTOS ENERO - JULIO 2023.xlsx

file.php?t=91b447e79caa1aa652a40cf5c7c5e83f

Archivo BASE DE DATOS ABIERTOS ENERO - JULIO 2023.xlsx descargado exitosamente.

Estos datos nos muestran el archivo y el nombre y de paso descargamos el archivo encontrado.

Importamos Pandas

import pandas as pd

En este momento deberé de convertir el archivo xlsx a csv. Ya que siempre estuvimos trabajando con CSV.

xlsx\_file = nombre

nombreCSV = nombre+'.csv'

data = pd.read\_excel(xlsx\_file, engine='openpyxl')

data.to\_csv(nombreCSV, index=False)

print(f'Archivo {xlsx\_file} convertido exitosamente a {nombreCSV}.')

Generare un data table, solo para comprobar que la información es correcta,

dt = pd.read\_csv(nombreCSV)

dt

Iniciamos con las graficas,

Importamos matplotlib

import matplotlib.pyplot as plt

Generamos las graficas

Hare un conteo

conteo = data['NOMBRE\_COMERCIAL'].value\_counts()

haremos la grafica

# Conteo de valores únicos en la columna "NOMBRE\_COMERCIAL"

nombre\_comercial\_counts = dt['NOMBRE\_COMERCIAL'].value\_counts().head(10)

# Crear un gráfico de barras

plt.figure(figsize=(10, 6))

nombre\_comercial\_counts.plot(kind='bar')

plt.xlabel('NOMBRE COMERCIAL')

plt.ylabel('Conteo')

plt.title('Gráfico de Barras: Conteo de NOMBRE COMERCIAL')

plt.xticks(rotation=45)

plt.tight\_layout()

# Mostrar el gráfico

plt.show()

Resultado

Gráfico, Gráfico de barras

Descripción generada automáticamente

# Conclusión

Me pareció muy interesante el curso y muy importante para lograr encontrar información. En mi caso trabajo mucho con BI y esta herramienta se que me puede ayudar a encontrar soluciones a los proyectos que tengo de una manera muy rápida que nos ayudara a ser mejores en el trabajo.

En mi caso no solo hice esta parte, sino que también lo Utilice importando JIRA a lo cual me ayudo en el scrum master y mire que es una herramienta muy poderosa.

De antemano muchas gracias.