

Organización de Datos - 2do Cuatrimestre 2020 7506-2020-2

PEDRO PEREZ

October 4, 2020

0.1 Finger 1 Pedro Pérez

0.1.1 Ejercicio:

A partir del set de datos considerado en el enunciado del TP 1 el cual se encuentra publicado en piazza realizar, utilizando pandas realizar 1. La apertura y carga del set de datos en un dataframe. 2. Mostrar, para la región EMEA, los datos: ID, Territory, Delivery_Quarter, Delivery_Year para los 10 proyectos con mayor Total Amount. 3. Calcular el promedio del precio (Columna: Price) para el territorio (Columna: Territory) cuyo valor sea “Germany”.

```
[25]: import pandas as pd
import numpy as np
import seaborn as sns
import os
from pandas_profiling import ProfileReport
import matplotlib.pyplot as plt
%matplotlib inline
```

0.1.2 1. La apertura y carga del set de datos en un dataframe.

```
[2]: directorio = 'dataset'
archivo = 'Entrenamieto_ECI_2020.csv'
fname = os.path.join(directorio,archivo)
#De esta forma puedo filtrar columnas.
df = pd.read_csv(fname, parse_dates= ['Account_Created_Date', 'Delivery_Year',
                                     'Quote_Expiry_Date', 'Month',
                                     'Actual_Delivery_Date',
                                     'Planned_Delivery_End_Date',
                                     'Planned_Delivery_Start_Date',
                                     'Opportunity_Created_Date',
                                     'Last_Modified_Date'] )
```

0.1.3 2. Mostrar, para la región EMEA, los datos: ID, Territory, Delivery_Quarter, Delivery_Year para los 10 proyectos con mayor Total Amount.

```
[5]: # Filtro el dataset con las columnas seleccionadas
columns = ['Region', 'Territory', 'Delivery_Quarter', 'Delivery_Year', 'Total_Amount']
df_selected = df[columns]
df_selected[df_selected['Region'] == 'EMEA'].nlargest(10, 'Total_Amount')
```

	Region	Territory	Delivery_Quarter	Delivery_Year	Total_Amount
2179	EMEA	None	Q2	2016-01-01	2.162400e+08
3667	EMEA	SE America	Q3	2017-01-01	1.539788e+08
2806	EMEA	UAE (Dubai)	Q1	2017-01-01	1.237501e+08
1933	EMEA	Kazakhstan	Q2	2016-01-01	3.960007e+07
2957	EMEA	Romania	Q2	2016-01-01	3.603600e+07
598	EMEA	France	Q4	2017-01-01	3.480420e+07
7898	EMEA	Belgium	Q1	2018-01-01	3.332493e+07
1817	EMEA	Romania	Q3	2017-01-01	2.880198e+07
3684	EMEA	Italy	Q4	2016-01-01	2.879760e+07
5669	EMEA	Germany	Q1	2017-01-01	2.829750e+07

0.1.4 3. Calcular el promedio del precio (Columna: Price) para el territorio (Columna:Territory) cuyo valor sea “Germany”.

```
[18]: df_precios_germany = df.loc[df_selected['Territory'] == 'Germany', ['Price']]
```

```
[19]: df_precios_germany.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1682 entries, 63 to 16938
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  -
0   Price   1682 non-null      object
dtypes: object(1)
memory usage: 26.3+ KB
```

```
[20]: # Convierto los precios a un formato numérico y calculo el promedio
df_precios_germany = pd.to_numeric(df_precios_germany['Price'],
errors='coerce').mean()
```

```
[23]: print('El precio promedio para la región de Germany es:'f'{df_precios_germany:
->11.2f}')
```

El precio promedio para la región de Germany es: 0.26

```
[ ]:
```