

Analisi delle spese del 2023 usando Python e Tableau

Raccolta dei dati

Dati provenienti dal mio conto corrente online relativi alle uscite tracciate nel corso del 2023, i dati non sono completi ed esaustivi ma comunque rappresentativi, alcuni numeri potrebbero essere stati modificati per ragioni di privacy. Obiettivo di fondo è fare esperienza con l'utilizzo di **python** per la manipolazione e l'analisi dei dati.

Strumenti utilizzati: *Pycharm community edition e Pandas*

1) Importazione del dataset CSV in Pycharm con il nome Expenses2023

```
import pandas as pd

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)

df = pd.read_csv("C:\\Users\\Lenovo\\Downloads\\Expenses2023.csv")

print(df)
```

Le categorie delle spese sono state raggruppate nei seguenti gruppi come da schermata qui sotto:

input:

```
import pandas as pd

pd.set_option('display.max_rows', 10)
pd.set_option('display.max_columns', 10)

df = pd.read_csv("C:\\Users\\Lenovo\\Downloads\\Expenses2023.csv")

types = df.groupby("Categoria")

for name, group in types:
    print(name)
```

Output del comando **groupby** sulla colonna Categoria:

Banking

Bills
Car
Clothes
Education
Entertainment
Grocery
Health
Travel
Various

2) Data Cleaning and Pre-processing

Il dataset è suddiviso nelle seguenti colonne:

Input:

```
df =  
pd.read_csv("C:\\Users\\Lenovo\\Downloads\\Expenses2023.csv")  
  
nomi_colonne = df.columns  
  
print(nomi_colonne)
```

Ouput:

Index(['DATA CONT.', 'DESCRIZIONE', 'Categoria', 'IMPORTO(€)'], dtype='object')

Rinomino la colonna "DATA CONT" in "DATA" - Trasformo il titolo della
colonna "Categoria" in upper case:

input:

```
df= df.rename(columns={"DATA CONT.": 'DATA'})  
print(df)  
  
df.columns = df.columns.str.replace('Categoria',  
'CATEGORIA').str.upper()
```

output:

	DATA	DESCRIZIONE	CATEGORIA	\
0	20/6/2023	Addebito Tassazione Su Transazioni Finanziarie...	Banking	
1	3/7/2023	Addebito Tassazione Su Transazioni Finanziarie...	Banking	
2	10/7/2023	Addebito Tassazione Su Transazioni Finanziarie...	Banking	
3	12/9/2023	Addebito Tassazione Su Transazioni Finanziarie...	Banking	
4	26/9/2023	Addebito Tassazione Su Transazioni Finanziarie...	Banking	

```
# Controllo il tipo di variabile:
```

input

```
tipi_colonne = df.dtypes
print(tipi_colonne)
```

output

```
DATA      object
DESCRIZIONE  object
CATEGORIA   object
IMPORTO(€)  object
dtype: object
```

```
# rimuovo gli spazi bianchi dal nome delle colonne
```

input

```
df.columns = df.columns.str.strip()
```

```
# Trasformo la colonna DATA in formato datetime:
```

input

```
df['DATA'] = pd.to_datetime(df['DATA'], format='%d/%m/%Y',
errors='coerce')
```

output

```
DATA
0 2023-06-20
1 2023-07-03
```

```
2 2023-07-10
3 2023-09-12
4 2023-09-2
```

```
# Aggiungo una colonna da 1 a 12 per il mese indicato in datetime:
```

input:

```
df['MESE'] = df['DATA'].dt.month
```

output:

	IMPORTO(€)	MESE
0	1.13	6.0
1	1.17	7.0
2	1.32	7.0
3	1.14	9.0
4	1.21	9.0
..
279	1.00	12.0
280	1.00	12.0
281	1.24	10.0
282	1.11	10.0
283	240,00	NaN

```
# Aggiungo una nuova colonna con il trimestre di appartenenza
per avere una rappresentazione anche basata sul trimestre:
```

input:

```
df['TRIMESTRE'] = df['DATA'].dt.quarter

print(df)
```

output:

	IMPORTO(€)	MESE	TRIMESTRE
0	1.13	6.0	2.0
1	1.17	7.0	3.0
2	1.32	7.0	3.0
3	1.14	9.0	3.0
4	1.21	9.0	3.0
..

```
279    1.00 12.0    4.0
280    1.00 12.0    4.0
281    1.24 10.0    4.0
282    1.11 10.0    4.0
```

```
# Controllo il tipo di variabili
```

input

```
tipi_colonne = df.dtypes

print(tipi_colonne)
```

output

```
DATA          datetime64[ns]
DESCRIZIONE    object
CATEGORIA      object
IMPORTO(€)     object
MESE           float64
TRIMESTRE      float64
dtype: object
```

```
# converto colonna IMPORTO(€) in float
```

```
df['IMPORTO(€)'] = df['IMPORTO(€)'].astype(float)
```

3) Analisi dei dati

Vorrei capire dall'analisi dei dati quali sono su base mensile e trimestrale:

- le principali categorie di spesa
- come si sono evolute sulla base del mese e del trimestre, quali sono i trend emergenti
- come posso rendere piu' efficienti le mie uscite il prossimo anno

```
# Calcolo la somma dei valori nella colonna IMPORTO
```

input

```
somma_importo = df['IMPORTO (€)'].sum()
```

output

La somma degli importi convertiti in float è: 5730.76

```
# Raggruppamento in base alla categoria e calcolo la somma, la media, max e min per ogni mese dell'anno da 1 a 12
```

input

```
raggruppato = df.groupby(['CATEGORIA',  
df['DATA'].dt.month]).agg({'IMPORTO (€)': ['sum', 'mean', 'min',  
'max']}).reset_index()  
  
raggruppato.columns = ['CATEGORIA', 'MESE', 'SOMMA_IMPORTO',  
'MEDIA_IMPORTO', 'MIN_IMPORTO', 'MAX_IMPORTO']  
  
print(raggruppato)
```

output

	CATEGORIA	MESE	SOMMA_IMPORTO	MEDIA_IMPORTO	MIN_IMPORTO	MAX_IMPORTO
0	Banking	1	16.01	5.336667	0.55	14.11
1	Banking	2	6.00	1.000000	1.00	1.00
2	Banking	4	18.81	6.270000	1.00	16.51
3	Banking	5	7.68	1.280000	1.03	1.39
4	Banking	6	6.34	0.792500	0.11	1.35
..
74	Various	8	60.00	20.000000	20.00	20.00
75	Various	9	115.02	38.340000	15.02	60.00
76	Various	10	57.30	19.100000	5.00	40.00
77	Various	11	254.54	50.908000	4.54	150.00
78	Various	12	33.00	16.500000	13.00	20.00

```
# Creo una Tabella pivot con i dati suddivisi per categoria e ordinati per mese e la somma per mese suddivisa per categorie
```

input

```
df_piv = df.pivot_table(index='CATEGORIA', columns='MESE',
values=['IMPORTO (€)'], aggfunc='sum')
```

```
df_piv['TOTALE_MESE'] = df_piv.sum(axis=1)
```

output

MESE	1.0	2.0	3.0	4.0	5.0	6.0	7.0
CATEGORIA							
Banking	16.01	6.00	NaN	18.81	7.68	6.34	20.60
Bills	10.00	NaN	10.00	10.00	10.00	10.00	10.00
Car	178.10	262.38	287.98	80.20	223.05	37.40	5.02
Clothes	153.43	134.52	NaN	10.19	NaN	NaN	NaN
Education	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Entertainment	15.39	NaN	NaN	NaN	36.00	18.00	NaN
Groceries	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Grocery	48.10	43.80	72.90	97.58	55.36	38.52	56.76
Healt	NaN	NaN	NaN	NaN	NaN	NaN	35.00
Health	13.00	NaN	NaN	NaN	NaN	NaN	NaN
Travel	7.00	252.59	4.14	NaN	NaN	184.38	82.42
Various	40.00	230.00	60.00	130.00	47.65	470.00	120.00

MESE	8.0	9.0	10.0	11.0	12.0	
CATEGORIA						
Banking	NaN	4.35	15.28	4.18	4.00	103.25
Bills	10.00	NaN	15.00	10.00	NaN	95.00
Car	27.79	35.89	376.82	52.11	20.46	1587.20
Clothes	NaN	NaN	64.88	42.97	39.28	445.27
Education	NaN	44.00	NaN	NaN	NaN	44.00
Entertainment	NaN	NaN	72.00	36.00	18.00	195.39
Groceries	NaN	NaN	2.35	NaN	NaN	2.35
Grocery	62.72	53.61	115.74	51.04	32.70	728.83
Healt	NaN	NaN	NaN	NaN	NaN	35.00
Health	NaN	NaN	NaN	NaN	NaN	13.00
Travel	NaN	NaN	134.28	111.40	87.75	863.96
Various	60.00	115.02	57.30	254.54	33.00	1617.51

```
# Calcolo la somma delle uscite complessive per ogni mese del
calendario
```

input

```
df_pivTotal= df.pivot_table(index='MESE',
values=[ 'IMPORTO(€) '], aggfunc='sum')
print(df_pivTotal)
```

output

MESE	IMPORTO(€)
1.0	481.03
2.0	929.29
3.0	435.02
4.0	346.78
5.0	379.74
...	...
8.0	160.51
9.0	252.87
10.0	853.65
11.0	562.24
12.0	235.19

4) Visualisation

...

5) Insights

- **Qual è la principale categoria di spesa?**
La principale voce di spesa resta la categoria “car” a livello dei 12 mesi insieme alla categoria “various” che non è per sua natura riconducibile ad una precisa categoria di spesa, in terza posizione la categoria “travel” che comprende le spese pre viaggi vari e vacanze.
- **Come si sono evolute le categorie di spesa sulla base del mese e del trimestre, quali sono i trend emergenti?**
Sicuramente i primi mesi dell’anno hanno visto una maggiore spesa per “auto” dovuta agli spostamenti per lavoro, anche i mesi di maggio e giugno hanno una spesa maggiore per la voce “car” dovuta al pagamento del bollo e dell’assicurazione.

Anche la categoria "**various**" e "**travel**" si sono mosse in maniera simile quindi le spese della categoria various sono probabilmente collegate anche ai viaggi. La categoria clothes ha valori maggiori nei mesi invernali mentre la categoria grocery resta stabile durante tutto l'anno.

- ***Come posso rendere piu' efficienti le mie uscite il prossimo anno?***

Posso ridurre le spese per l'auto utilizzando i mezzi pubblici laddove possibile per andare a lavoro e soprattutto fare pagamenti con la carta elettronica in modo da avere dati maggiormente attendibili per il tracciamento delle spese