

# PROGETTO DI DATA MANIPULATION AND VISUALIZATION

DI NICHOLAS COPPO

FONTE: [HTTPS://WWW.KAGGLE.COM/YUGAGRAWAL95/SAMPLE-MEDIA-SPENDS-DATA](https://www.kaggle.com/yugagrawal95/sample-media-spends-data)

IL PROGETTO PONE LE PROPRIE BASI IN UNA  
POSSIBILE COLLABORAZIONE CON UNA  
ORGANIZZAZIONE NO PROFIT IMPEGNATA  
NELLA DISTRIBUZIONE DI REPORTAGE  
INDIPENDENTI.

**OBIETTIVO:** ANALIZZARE IL DATASET  
FORNITO ATTRAVERSO STRUMENTI DI  
MANIPOLAZIONE E VISUALIZZAZIONE DATI.

**STRUMENTI:** PYTHON E LIBRERIE (PANDAS,  
MATPLOTLIB ,SEABORN , PLOTLY EXPRESS).

LA PRIMA FASE DEL LAVORO PREVEDEVA L'IMPORTAZIONE DEL DATASET E LA RELATIVA ANALISI ESPLORATIVA, PER LA QUALE SONO STATI USATI DEI CODICI AD HOC. ALCUNI DI QUESTI PIU SEMPLICI, ALTRI PIU APPROFONDITI.

QUI DI SEGUITO VIENE PROPOSTO SOLO QUALCHE ESEMPIO

```
df.head()
```

```
df.info()
```

```
print(df.describe())
```

LA SECONDA FASE DEL LAVORO PREVEDEVA LA RICERCA DI EVENTUALI PATTERN, TREND OPPURE OUTLIER IN GRADO DI FORNIRE QUALCHE INFORMAZIONE PIU' PRECISA SULL'ANDAMENTO DELLE VENDITE.

QUI DI SEGUITO QUALCHE ESEMPIO DI CODICE CREATO PER RISPONDERE AI QUESITI CITATI E VISUALIZZARE UN GRAFICO ADATTO:

1) ANDAMENTO DELLE VENDITE (LINEPLOT)

2) DISTRIBUZIONE DELLE VENDITE PER MESE (BOXPLOT)

3) ANDAMENTO DELLE VENDITE TRIMESTRALI (BARPLOT)

```
plt.figure(figsize=(12,6))
sns.lineplot(x='Calendar_Week', y='Sales', data=df, marker='o')
plt.title('Andamento Vendite nel Tempo')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
plt.figure(figsize=(10,5))
sns.boxplot(x='Month', y='Sales', data=df)
plt.title("Distribuzione delle Vendite per Mese")
plt.show()
```

```
plt.figure(figsize=(8,5))
sns.barplot(x='Quarter', y='Sales', data=df, estimator='mean')
plt.title("Vendite Medie per Trimestre")
plt.show()
```

IN QUESTA SLIDE SI PROPONE UN ESEMPIO DI CODICI CREATI PER UNA PARTE PIÙ APPROFONDITA DEL PROGETTO, OVVERO L'ANALISI QUANTITATIVA.

L'INTERESSE ERA QUELLO DI PROVARE A IDENTIFICARE I CANALI PIÙ EFFICACI TRA QUELLI USATI DALLA ORGANIZZAZIONE NO PROFIT.

```
corr_google = df['Sales'].corr(df['Google_Impressions'])
corr_facebook = df['Sales'].corr(df['Facebook_Impressions'])
corr_email = df['Sales'].corr(df['Email_Impressions'])
corr_affiliate = df['Sales'].corr(df['Affiliate_Impressions'])

print(f"Correlazione Google: {corr_google:.2f}")
print(f"Correlazione Facebook: {corr_facebook:.2f}")
print(f"Correlazione Email: {corr_email:.2f}")
print(f"Correlazione Affiliate: {corr_affiliate:.2f}")

sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Matrice di Correlazione")
plt.show()
```

```
df['Google_Eff'] = df['Sales'] / df['Google_Impressions']
df['Facebook_Eff'] = df['Sales'] / df['Facebook_Impressions']
df['Email_Eff'] = df['Sales'] / df['Email_Impressions']
df['Affiliate_Eff'] = df['Sales'] / df['Affiliate_Impressions']

eff_media = df[['Google_Eff', 'Facebook_Eff', 'Email_Eff', 'Affiliate_Eff']].mean()
eff_media.plot(kind='bar', title='Efficienza Media (Vendite per Impression)', figsize=(8,5))
plt.ylabel('Efficienza media')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```