

# Relazione Big Data

Il progetto svolto è incentrato sull'analisi di un dataset proveniente da Kaggle.com, contenente informazioni dettagliate riguardanti il mondo automobilistico degli anni '70 e '80. <https://www.kaggle.com/datasets/tawfikelmetwally/automobile-dataset>. L'obiettivo principale è condurre operazioni statistiche su regioni specifiche e successivamente estendere l'analisi a livello mondiale. Nel corso del progetto, sono state utilizzate librerie Python come Pandas, Matplotlib e Seaborn per manipolare i dati e creare rappresentazioni grafiche.

## 1. Caricamento e Analisi dei Dati

In questa sezione, vengono importate le librerie necessarie e i dati vengono caricati da un file CSV chiamato 'Dataset.csv' utilizzando Pandas. Successivamente, viene visualizzato un riepilogo delle informazioni sul dataset tramite `info()`, e vengono calcolate e stampate le statistiche descrittive con `describe()`.

## 2. Grafico a barre dei modelli più prodotti per Origine

Si filtrano i dati dagli anni '70 agli '82. Vengono poi identificati i modelli più prodotti nelle regioni (USA, Europa, Giappone) durante questo periodo e visualizzati su un grafico a barre. Il grafico è colorato in base alla regione, e sono aggiunte etichette e un titolo per una migliore comprensione.

## 3. Grafico a barre della Potenza Media dei modelli per Origine

In questa sezione, viene calcolata la potenza media per ciascuna origine (USA, Europa, Giappone) e visualizzata tramite un grafico a barre utilizzando **Seaborn**. Le etichette numeriche sono aggiunte sopra ciascuna barra per indicare il valore numerico corrispondente.

## 4. Grafici dei Trend nel Tempo

Vengono creati tre grafici distinti per rappresentare i trend nel tempo relativi alla potenza media, al consumo di carburante medio e alla cilindrata media. Ciascun grafico ha le etichette e i titoli appropriati per una chiara interpretazione.

## 5. Grafico Lineare per la Progressione della Potenza tra i modelli più prodotti al Mondo

Si identificano i modelli più prodotti a livello mondiale, il dataset viene filtrato di conseguenza, e viene creato un grafico lineare utilizzando **Seaborn** per mostrare la progressione della potenza nel tempo per questi modelli. Il grafico include legende, titoli, etichette e linee guida per una migliore interpretazione.

### Documentazione sulle Funzioni utilizzate

Documentazione Sezione 1: Utilizzo della funzione **read\_csv** di Pandas per caricare i dati da un file CSV chiamato 'Dataset.csv'. Questo metodo è ampiamente utilizzato per importare dati tabulari in formato CSV in un DataFrame Pandas.

Documentazione Sezione 2: Utilizzo della libreria Matplotlib per creare un grafico a barre con tre serie di dati sovrapposte. La funzione **subplots()** consente di creare un layout di subplot, e l'asse **ax** viene utilizzato per posizionare le barre delle diverse regioni affiancate.

La funzione **plt.xlim()** è utilizzata per impostare i limiti dell'asse x in un grafico creato con la libreria Matplotlib. In questo caso specifico, l'uso della funzione **plt.xlim(left=-0.5, right=4.5)** è finalizzato ad aggiungere spazio sia a sinistra che a destra dell'asse x per migliorare la visualizzazione delle barre in un grafico a barre.

Documentazione Sezione 3: Utilizzo della libreria Seaborn per creare un grafico a barre della potenza media per origine. La funzione **barplot** prende i dati da un DataFrame e utilizza il parametro **palette** per personalizzare i colori delle barre. La funzione **text** viene utilizzata per aggiungere etichette numeriche sopra ciascuna barra.

Documentazione Sezione 4: Utilizzo della funzione **tight\_layout** di Matplotlib per migliorare la disposizione dei subplot in modo che non si sovrappongano. Questo è particolarmente utile quando si hanno più grafici all'interno di una singola figura.

Documentazione Sezione 5: Utilizzo della funzione `axhline` di Matplotlib per aggiungere una linea orizzontale alla posizione  $y=0$  nel grafico lineare. In questo caso, la linea è utilizzata per separare il titolo dal resto del grafico, migliorando la chiarezza visiva.