

Author: Jaime Andres Diaz Crosthwaite

Data: 07 di Giugno 2023

### Dataset 1: Intrattenimento

CCO: Public Domain

<https://www.kaggle.com/datasets/mauryansshivam/netflix-ott-revenue-and-subscribers>

Informazioni sul set di dati: "Netflix OTT (Over the Tops) Revenue and Number of Paid Subscribers dataset contains the revenue of Netflix in dollars and number of paid subscribers region wise."

Si tratta di un dataset strutturato poiché è organizzato in una tabella che contiene dati qualitativi e quantitativi.

Qualitativi: Data

Quantitativi: Il profitto globale

Nei diversi attributi del database, si osserva il profitto globale di netflix anche per regione come segue:

UACN : United States and Canada

EMEA : Europe, Middle East and Africa

APAC : Asia-Pacific

LATM : Latin America

Analisi: Utilizzerei queste informazioni per analizzare la crescita dell'intrattenimento in streaming e stimare i guadagni netti in diverse regioni del mondo.

### Dataset 2: Economia

Economia dei paesi PIL(GDP)

<https://www.kaggle.com/datasets/darshanprabhu09/countries-economy-gdp-and-everything>

"Un dataset in formato CSV che descrive la diversa economia di vari paesi del mondo tenendo in conto del PIL"

Dataset strutturato poiché è organizzato in una tabella che contiene dati qualitativi e quantitativi.

Attributi:

20 Colonne

- Decimali: 15
- Interi: 3
- Stringa: 2

Qualitativo	Quantitativo
Nome dei Paesi	Popolazione, Area, Densità di popolazione, Migrazione, Mortalità infantile, PIL

Analisi: Sarebbe interessante analizzare questo database per misurare il potere d'acquisto della popolazione, in questo caso utilizzando il PIL pro capite. Se qualche azienda vuole esportare i propri

servizi o prodotti possiamo analizzare quali sono quelli paesi che hanno in media, un potere d'acquisto maggiore.

Utilizzerei l'indice dei telefoni in un determinato paese per segmentare il mercato e lanciare pubblicità attraverso il marketing digitale.

### **Dataset 3: Vendite di biciclette**

<https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe>

“Dataset di informazioni sulle vendite di biciclette in ogni paese.”

Dataset strutturato

Attributi:

- Colonne: 18
- Intero: 9
- Testo: 7
- Data: 1
- Altri: 1

Analisi:

Determinare l'uso della bicicletta nelle diverse età delle persone al fine di segmentare il mercato. Questo dataset inizia dall'anno 2011 all'anno 2016 quindi Sommerei il totale delle vendite accumulate di ogni anno e misurerei l'impatto dell'uso delle biciclette sulla qualità dell'aria.

Stimolare l'uso della bicicletta nella popolazione giovane per migliorare la qualità dell'aria.

### **Dataset 4: Recensioni di Hotel**

<https://www.kaggle.com/datasets/datafiniti/hotel-reviews>

“Questo è un dataset di 1.000 hotel e le loro recensioni fornite dal Business Database di Datafiniti. Il set di dati include la posizione dell'hotel, il nome, la valutazione, i dati delle recensioni, il titolo, il nome utente e altro.

Licenze: CCO

Semi strutturato

#### **Analisi:**

Si rilevano:

- Dati strutturati come le coordinate della posizione di ogni Hotel.
- Dati non strutturati come i diversi commenti dei clienti.

Si Analizzano:

I commenti degli hotel con una valutazione negativa, estraendo parole chiave che possano rilevare i punti deboli e le criticità col fine di migliorare il servizio delle strutture.

## **Dataset 5: Italian Signs**

“Un dataset contenente immagini del contesto stradale fatte dalla fotocamera di un telefono”

<https://www.kaggle.com/datasets/officialprojecto/italiansigns>

Dataset Semi Strutturato (misto)

Attributi

- Colonne: 6
- Intero: 5
- Stringa: 1

### **Analisi:**

Si rilevano:

- Dati non strutturati: immagini
- Dati strutturati: una tabella strutturata con l'ora in cui ciascuna immagine è stata scattata.