ARGOMENTO 1:

Ho ipotizzato che una ditta di caffè voglia ottimizzare la propria produzione e distribuzione nei vari Paesi. Mi sono servita quindi di un dataset sul consumo domestico di caffè in diversi Paesi nel corso degli anni (dal 1990 al 2020).

<https://www.kaggle.com/datasets/michals22/coffee-dataset>

**L’identificatore** è il Paese, gli **attributi** sono:

-Tipologia di caffè (arabica, robusta..)

-Gli anni (consumo per quell’anno specifico, in quel Paese)

Si possono così conoscere i Paesi con consumo maggiore, la tipologia preferita per ogni Paese, il trend nel corso degli anni. Tutto ciò potrebbe permettere una migliore distribuzione del caffè nei differenti Paesi e una produzione più specifica diminuendo gli scarti o i tempi di permanenza in magazzino.

Come data set non strutturato ho invece ipotizzato ci siano dei questionari di gradimento online a risposta aperta. Non avendo modo di dividerli in maniera specifica ho pensato di fare una ricerca per parole chiave (es. buono, ottimo, tipo di caffè…) per aggiungere ulteriori dati.

ARGOMENTO 2:

Ho analizzato un dataset su kaggle relativo alla serie tv “the last of us” :

<https://www.kaggle.com/datasets/ulisesmontoyacanales/the-last-of-us-tweets>

Il dataset contiene i twitter relativi alla serie e alcune informazioni legate a likes o retweets. E’ inoltre presente una colonna che credo si possa definire DATO NON STRUTTTURATO legato al testo dei tweet (da questa parte si potrebbero cercare parole specifiche o hashtag per capire quali siano i più usati)

**L’identificatore** è ID , gli **attributi** sono:

-id dell’autore

-testo (non strutturato)

-data creazione

-numero di likes

-numero di retweets

-numero di risposte

-impressions

-numero di volte in cui è stato citato il tweet

-lingua

ARGOMENTO 3:

Analisi dei punteggi e delle recensioni dei libri su Goodreads.

<https://www.kaggle.com/datasets/jealousleopard/goodreadsbooks>

**L’identificatore** è ID del libro ma anche ISBN e ISBN13 potrebbero essere usati come codici univoci per identificare il libro , gli **attributi** sono:

-titolo

-autore

-voto medio

-lingua

-numero pagine

-numero di voti ricevuti

-numero recensioni scritte

-data pubblicazione

-nome editore

Per i dati non strutturati ancora una volta cercherei delle parole o hashtag specifici nelle recensioni scritte per capire i più usati.