

Task 1

Task 1a — provate ad individuare 1 dataset online e a definire 5 domande (da analista dati), sempre con in mente lo schema: Chi, Cosa, Dove, Quando e Perché

Per questo task abbiamo individuato un dataset che traccia l'evoluzione del valore delle azioni di Tesla in dollari americani in un periodo compreso tra giugno 2010 e marzo 2022

Fonte del dataset: <https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021>

Domande del Data Analyst sul dataset di Tesla:

1. Qual è il valore medio delle azioni di Tesla per ogni anno?
2. C'è una ciclicità nell'evoluzione delle quotazioni di Tesla nel corso del periodo osservato? Se sì, quale correlazione esiste?
3. C'è una ciclicità nell'evoluzione del volume di stocks vendute?
4. C'è una correlazione tra la media mensile delle stocks vendute e la media mensile del prezzo?
5. Nel periodo osservato (2010-2022) com'è evoluto il dato di high price?

Task 2a — Provate ora ad individuare 1 big dataset online e a definire le medesime 5 domande

Abbiamo individuato un dataset che traccia i trip record data dei taxi della città di New York

Fonte del dataset: <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

NB. La struttura dei vari componenti del dataset si trova in fondo alla pagina visitabile dal link sopra stante, sottoforma di pdf consultabili contenenti gli "attributi" e la descrizione ad essi relativa

Questo costituisce un big dataset poiché:

1. Il volume è tale da avere informazioni su milioni di corse avvenute nella città di New York

2. Si tratta di dati che vengono inseriti ogni giorno
3. Ci sono dati di svariata natura (durata delle corse, distanza, luoghi di partenza e destinazione delle corse, metodi di pagamento...) e di diversa appartenenza (abbiamo i dati sui for hire vehicles, sui large for hire vehicles, sui taxi verdi e sui taxi gialli. Questo ci permette di effettuare anche degli studi incrociati tra i vari tipi di veicolo)

Domande:

1. Esiste una relazione tra quantità mancia riscossa e tempi di percorrenza di una tratta, sia per i taxi gialli e verdi?
2. Esiste una correlazione tra il metodo di pagamento e il total amount? Inoltre, com'è cambiata la distribuzione dei metodi di pagamento dal 2019 ad oggi?
3. Tramite analisi degli orari di pick up, è possibile individuare dei momenti di picco della giornata durante i quali si assiste ad un aumento delle richieste dei vari servizi taxi e hiring?
4. Qual'è la distribuzione geografica delle richieste di pick up nell'area di NYC?
5. Che evoluzione ha subito, negli anni, la media giornaliera del tempo impiegato in una corsa sia dai taxi verdi che dai taxi gialli?