**Come si muovono i taxi a New York?**

In questo progetto svolgiamo un'analisi dei taxi a New York. In particolare, siamo curiosi di rispondere ad alcune specifiche Research Questions (RQ) che possono aiutare i tassisti a pianificare i loro spostamenti in città e ai clienti ad avere suggerimenti sulla convenienza dell’utilizzo di questo servizio.

Per questo progetto utilizziamo i dati pubblici delle rotte dei Taxi a NYC disponibili su <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>.   
Per rispondere alle RQ teniamo conto dei dati relativi ai Yellow Taxi per l'anno 2020.

**Prima di iniziare**

1. Fare il download di Yellow Taxi Trip Records (CSV) 2020. Iniziate sul dataset di January.

2. A causa delle dimensioni dei file, vedrete che inserire tutti dati in memoria è difficile. Se siete in grado di caricare i dati in memoria, sarà difficile eseguire qualche analisi. Per questo motivo, prima di iniziare vi invitiamo a dare un'occhiata alle RQ e a pensare ad una strategia per trattare i dati. Come liberare la memoria: <https://teamtreehouse.com/community/how-to-delete-a-variable-in-python>

3. Per poter effettuare l'analisi sulle borough (distretto) è necessario combinare i dati delle rotte dei taxi con il set di dati che si trova in taxi\_zone\_lookup.csv. Date un'occhiata qui: <https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html>

4. Studiare il dataset:

* Per capire in che cosa consistono i dati leggete la leggenda: <https://www1.nyc.gov/assets/tlc/downloads/pdf/data_dictionary_trip_records_yellow.pdf>
* Leggete qualche informazione su come funzionano i taxi a New York per poter fare osservazioni significative: <http://www.nyc.gov/assets/tlc/downloads/pdf/taxi_information.pdf>

5. Controllate per bene i dati. Bisogna cambiare qualcosa? Per esempio: ci sono rotte con distanza pari a 0, hanno senso? Controllate se sono presenti valori NaN e decidete come trattarli. Ci sono informazioni inutili?

**!!!Molto Importante!!!**

1. Leggete tutte le RQ prima di inziare a scrivere codice.

2. Come ogni attività di analisi dei dati, non esiste un modo unico e corretto per rispondere alle RQ. Poiché i risultati che otterrete per ogni RQ possono dipendere dalle scelte che farete durante l'analisi, è molto importante (e necessario) che descriviate ogni singola decisione che prenderete e tutti i passi che farete.

3. All'inizio della vostra analisi, scegliete e indicate chiaramente quale borough considerate per la conduzione dell'analisi: il borough di partenza o quello di arrivo.

**Research questions**

**RQ5**

La tariffa al miglio cambia attraverso i borough di New York? Vogliamo scoprire se le spese di un utente che usufruisce dei taxi in una zona sono diverse da quelle di chi li usa in un'altra. Considerando il fare\_amount:

* Calcolare il Prezzo al miglio (PM) per ogni corsa.
* Calcolare la media e la deviazione standard della nuova variabile per ogni borough. Poi tracciate la distribuzione con un boxplot.

Il prezzo per miglio potrebbe dipendere dal traffico che il Taxi trova lunga la strada. Quindi, cerchiamo di mitigare questo effetto. È probabile che la durata del viaggio dica qualcosa sulla congestione della città, soprattutto se combinata con le distanze percorse. Potrebbe essere una buona idea ponderare il prezzo per miglio utilizzando il tempo necessario per completare il viaggio (T). Così, invece di PM, si può usare PMT=. Svolgere le stesse analisi fatte su PM ma su PMT e confrontare i risultati.

**Suggerimenti**

1. Concentratevi su un solo mese, eseguendo tutte le vostre analisi su un unico set di dati.

2. Ripetere la stessa analisi negli altri mesi.

3. Combinate i risultati.