Codebook for Taxi Trips Dataset

# Data Overview

## Credentials

This data set can be downloaded from Kaggle data sets.

NYC taxi timings + NYC weather report data link: https://www.kaggle.com/purna007/intermidiate-data

## Business goal

This data was collected to answer the question:

## How does the average travel distance affected by the amount of precipitation?

## Data description

This data set is a data frame of 20 variables (21 columns, first column is an index) X 945 rows.   
Each row represents an taxi ride with all its parameters.

There are missing values in the data represented by “T”.

# Variables description

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Type | Possible values |
| year | The year of the ride | Numeric integer | 2016 |
| month | The month of the ride | Numeric integer | 1-12, months |
| date | The full date of the ride,  format: m/dd/yyy | Date | 2/20/2016 |
| hour | The hour of the ride, 24 hr | Numeric integer | 0-23 |
| id | The given ride\_id, format: idXXXXXXX | String | id1205949 |
| vendor\_id | The vendor who supplied the ride | Numeric integer | 1, 2 |
| pickup\_datetime | Pick up time and date, format: m/dd/yyyy, hh:mm | Date | 2/7/2016 18:15 |
| dropoff\_datetime | Drop off up time and date, format: m/dd/yyyy, hh:mm | Date | 5/22/2016 18:06 |
| passenger\_count | Number of passengers in the taxi,  1<= passengers <=6 | Numeric integer | 1, 2, 3, 4, 5, 6 |
| pickup\_longitude | Pickup longitude position | Numeric Float | -73.99120331 |
| pickup\_latitude | Pickup latitude position | Numeric Float | 40.74955368 |
| dropoff\_longitude | Drop off longitude position | Numeric Float | -74.00937653 |
| dropoff\_latitude | Drop off latitude position | Numeric Float | 40.74667358 |
| store\_and\_fwd\_flag | Irrelevant column- all the rows have the same value | string | N |
| trip\_duration | Trip duration in seconds | Numeric Float | 589 |
| avg\_temp | The average temperature in Fahrenheit | numeric | 50.5 |
| snow\_depth | Snow depth during the ride | string | 0, 2, T-missing |
| snow\_fall | The quantity of snow falling within a trip in % | string |  |
| precipitate | Precipitate amount in %/inches | string | 0.91 |
| distance | Trip distance in miles | Numeric Float | 40.12910039 |

Data set: Taxi trips https://www.kaggle.com/purna007/intermidiate-data

Owner: PurnaRaoMallepaddi | Kaggle.com

1) How does the average travel duration affect by amount of precipitation?

2) How does the average travel distance affect by the amount of precipitation?

3) How is the trip duration affected by average temperatures?

4) What is the ideal distance of travel to maximize the number of passengers per hour in rush hour?

Who needs to review this data?

Taxi company managers - predicting when demand for services will go up, thus adjusting taxi number on shift and prices.

Taxi passengers - getting a clearer picture on price, taxis near their location, and adjustment to commute time.

Article:

Tsapakis I., Cheng T. & Bolbol A. (April 2013). Impact of weather conditions on macroscopic urban travel times. Journal of Transport Geography, Volume28, pp. 204-211.

https://www.sciencedirect.com/science/article/pii/S0966692312002694

Main conclusions from the article:

The researchers found that the impact of rain and snow is a function of their intensity. the range of travel time increase due to light, moderate and heavy rain/snow.

The researchers also found that temperature had negligible effects on travel time.