

# About Dataset

This data will try to address the challenges faced by taxi operators in quoting the right fare to customers before starting the trip. However, the details are shared with taxi drivers or operators related to the trip, they find it difficult to quote the right fare because of uncertainties and calculation complexities. The same issue is faced by passengers as well because of inaccurate or irrelevant fares quoted. To find a solution for this, a historical dataset to analyse that includes records of taxi trip details and fares of those trips. Using this dataset, you need to build machine learning models for predicting the trip fare based on the given other useful features of the trip.

Overall, it involves using a dataset, finding the best set of features from the dataset, building a machine learning model to predict trip fare based on other trip features and evaluating the predictions using mean squared error and finally submitting the predictions in the given template.

## Data description:

Trip\_distance: The elapsed trip distance in miles reported by the taximeter.

Rate\_code: The final rate code is in effect at the end of the trip. 1= Standard rate,2=JFK,3=Newark, 4=Nassau or Westchester, 5=Negotiated fare,6=Group ride

Storeandfwd\_flag: This flag indicates whether the trip record was held in vehicle memory before sending it to the vendor and determines if the trip was stored in the server and forwarded to the vendor. Y= store and forward trip N= not a store and forward trip

Payment\_type: A numeric code signifying how the passenger paid for the trip. 1= Credit card,2= Cash, 3= No charge, 4= Dispute, 5= Unknown, 6= Voided trip

Fare\_amount: The time-and-distance fare calculated by the meter

Extra: Miscellaneous extras and surcharges.

Mta\_tax: \$0.50 MTA tax that is automatically triggered based on the metered rate in use.

Tip\_amount: Tip amount credited to the driver for credit card transactions.

Tolls\_amount: Total amount of all tolls paid in the trip.

Imp\_surcharge: \$0.30 extra charges added automatically to all rides

Total\_amount: The total amount charged to passengers. Does not include cash tips

Pickuplocationid: TLC Taxi Zone in which the taximeter was engaged

Dropofflocationid: TLC Taxi Zone in which the taximeter was disengaged

Year: The year in which the taxi trip was taken.

Month: The month on which the taxi trip was taken.

Day: The day on which the taxi trip was taken.

Day\_of\_week: The day of the week on which the taxi trip was taken

Hour\_of\_day: Used to determine the hour of the day in 24 hours format

Trip\_duration: The total duration of the trip in seconds

calculated\_total\_amount: The total amount the customer has to pay for the taxi.