**PROBLEM STATEMENT**

The predicting of fare amount (inclusive of tolls) for a taxi ride, given the pickup and dropoff locations,the pickup date time and many other attributes given below. The description of all the attributes is given below. The target variable is **“fare\_amount “**

Understand the data and do necessary data exploration and try creating new features and build a machine learning model to predict the fare amount.

| Variable | Description |
| --- | --- |
| TID | Unique ID |
| Vendor\_ID | Technology service vendor associated with cab company |
| New\_User | If a new user is taking the ride |
| toll\_price | toll tax amount |
| tip\_amount | tip given to driver (if any) |
| tax | applicable tax |
| pickup\_timestamp | time at which the ride started |
| dropoff\_timestamp | time at which ride ended |
| passenger\_count | number of passenger during the ride |
| pickup\_longitude | pickup location longitude data |
| pickup\_latitude | pickup location latitude data |
| rate\_category | category assigned to different rates at which a customer is charged |
| store\_and\_fwd | if driver stored the data offline and later forwarded |
| dropoff\_longitude | drop off longitude data |
| dropoff\_latitude | drop off latitude data |
| payment\_type | payment mode used by the customer (CRD = Credit Card, CSH - Cash, DIS - dispute, NOC - No Charge, UNK - Unknown) |
| surcharge | surchage applicable on the trip |
| fare\_amount | trip fare (to be predicted) |