EDA – PROJECT (INT 353)

Submitted by : Vikash Kumar Shrivastava.

Section : K20RU.

Roll No. : RK20RUA07.

Registration no. : 12018607.

Submitted to : Abhijeet Dutta Sir.

UBER DATASET.

Insights of Dataset :

]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47
...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

6745 rows × 6 columns

There are six attributes associated with each request made by a customer :

1. Request id : A unique identifier of the request.
2. Time of request : The date and time at which the customer made the trip request.
3. Drop – off time : The drop – off date and time, in case the trip was completed.
4. Pick – up point : The point from which the request was made.
5. Driver id : The unique identification number of the driver.
6. Status of the request : The final status of the trip, that cab be either completed, cancelled by the driver or no car available.

WHY ANALYSIS ON THIS DATASET IS IMPORTANT ?

One may have some experience of travelling to and from the airport (or any location) , there is a higher possibility that the passenger will surely book a cab for him to reach to the final destination. So, there is huge possibility that they face the problem of cancellation of the booking by the driver or there is non – availability of cars.

Well, if these are the problem faced by the customers, these very issue also impact the business of the Uber. If drivers cancel the request of the riders or if cars are unavailable , Uber loses out on its revenue.

So it's a prime area of concern for the company like Uber , which make it a necessary for them to analyse the data captured by them to review it and find the cause and solution for the same.

MY APPROACH :

1. Understanding the data.
2. Analysing all the columns.

3. Adding some columns if required.
4. Cleaning the dataset.
5. Perform Univariate, Bivariate and Multivariate analysis if required.
6. Finding insights on basis of different analysis.
7. Conclusion.

OBJECTIVE :

The aim of analysis is to identify the root cause of the problem (i.e. cancellation and non – availability of cars) and recommend ways to improve the situation. As a result of the analysis one should present the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

HOW CAN THIS ANALYSIS HELP THE ORGANIZATION.

By this analysis the organization with similar business model can come to know priorly the problems faced by already established companies.

So that they can execute there plan bit smoothly as compared to one who are facing the issues.

Such analysis can help the company to limit there loses and help them to work on the causes which are leading to the loss of companies revenue.

THANK YOU.