1. Cab Service Provider Company - Uber

Data Analyst @ Uber:-

Riders Ops Team: - Rider Operations team

**Uber cab:**

1. **Riders**
2. **Drivers**

**2 way marketplace**

Q3- Main Goal/OKR: Objective Key Results:

1. Reduce the waiting time for my riders
2. KR : - Average customer ratings increase by x%

Rider Ops team for Q3:

Look into ride completion times and see if there are a lot of deviations

Completion times are 10mins - 1 hr- 10%

Completion time rest : 30%

Completion time is = 2hrs - 20%

Completion time is = 3hrs - 20%

Completion time is = 4hrs - 20%

Completion time is 20-40 mins 90%

Standard deviation -

1. Look into ride completion times and see if there are a lot of deviations
2. Uber already has a prediction system called as

**Internal Trip Duration Prediction System**

**ITDPS-**

**Rider → Predicted the trip completion time**

1. **Now you want to come up with a ML based system for predicting this**
2. **This ML model should be better than rule based system to be deployed**
3. **Is this model performing good enough to be launched? – feature reliable or not?**
4. **Before investing in ML models, the product managers want to**

**Quickly validate if this idea is worth investing**

**User → completion time–??**

Uber Eats:

1. Delivery folks
2. Restaurants
3. Customers

3 way marketplace

Data analysis: Validation of the idea

To understand if this feature of showing the predicted ride completion time would be useful?

1 hr case study - problem study - ML round

Cost -benefit Analysis:

Cost of building the feature - Benefit of implementing

Analyze something which tells us this idea is worth the investment or not?

1. Check the user experience for some set of users first
2. Historical Reviews: Customer Ratings - bad

If mean customer rating is very bad and the reason has been mentioned as long waiting time,

Took more than what was promised

Customer ratings - good: not add that much

1. Surveys: Driver – > Idea where you’d incentivized? → 92% driver said yes

Surveys: Riders→ If we show you total completion time → would you → 96% customers felt yes

1. Data Analysis:

**1) Check # of tickets where issue is regarding the trip completion time**

#2) Customer waiting time for the cancelled rides – driver availa

3) Analysis of trip completion times →

1)Survey, Customer Ratings, Tickets, Analysis of the trip completion, waiting times

1. **Customer → Opened app→ Selected a ride→10$→ System searches for cab drivers→ customer drops - duration→ greater than 5 mins**

**Took so much time the customer couldn’t wait**

**Not 100% accurate**

What if we had a feature which told customer the accurate time to wait —>

1. Customer → Opened app→ Selected a ride→ System searches for cab drivers→ cab driver allocated-> ride starts-> ride ends→ customer gives low rating

Bad quality of ride, cleanliness of the cab, Trip duration

1. Customer → Opened app→ Selected a ride, 10 mins→ System searches for cab drivers→ cab driver allocated-> customer drops

Cab driver rating, change of minds, cash →

Long waiting time → inaccurate waiting time

User Journey

Reliability of the idea:

Driver ->

If I show a feature which talks about the predicted time →

Predicted times→

Hypothesis: Advantage:

**You re trying to come up with the ideas for the correct predictions**

**But first**

Through the existing data, can you first tell if this idea will work?

Data Analysis:

User app → if people going

Existing data→ if this idea will work or not

Comparison of old vs new system

Looking into the demand for the feature

Driver availability -

Driver -> Uber will say that you complete the ride within 15 mins → incentive to the driver

(5 mins, 10 mins)

Driver will try to complete within that time→ stable trip duration

Time period → Predicted completion time = booking time + driver to reach customers location time + 15 mins

Cancellation Rate:- Will go down!

ML model perfomr

Investment: - lot of problems faced by customer because of long waiting time and inaccurate predictions, and also drivers seem to be very much excited about this feature getting launched because they can earn more incentive, investment cost is under the budget

ML prediction system to predicting the completion times

Factors affecting the trip completion time→ hypothesis

More distance implies more completion time

# of drivers available/area size —> completion time to be less

Current state → future state —> gap

All the data is stored in GCP BigQuery, and there is more advantage of going and building models in GCP BQ ML instead of Python. Now you have no idea of GCP BQ ML and you need to learn and deliver the results within 10 days

Interval:

**Part 2:**

**GCP BigQuery**

**GCP BQ:**

**Steps:**

1)https://cloud.google.com/bigquery/docs/introduction

2) https://cloud.google.com/bigquery/docs/reference/standard-sql/query-syntax#union

<https://hevodata.com/learn/bigquery-jupyter-notebook/>

https://cloud.google.com/bigquery/docs/visualize-jupyter