
Collaboration Strategies

— By Indu Mittal —

Product Development Life Cycle

Stage 1	Stage 2	Stage 3	Stage 4
Definition Phase	Development Phase	Experimentation Phase	Global Launch
• Product/Business Requirement Proposal • User Research • Wireframes/Prototype • New Policy/TOS Proposal (if needed) • Technical Design Document	• Design and Develop the MVP (minimum viable product) • Test and debug • Document User Journey Mapping • Compliance and Security Checks • Threats and Risk Mitigation	• Internal Alpha Testing (Dogfooding) • Beta Customer Advisory Board testing • Release to few customer segments • Instrumentation Plan • Overview of results and compare with success criteria and feature goal	• Launch and Operations Plan • Analytics (Metrics and Reporting) • Security Model • Customer Communication Model • Automation and ML Scope

Customer Feedback Model



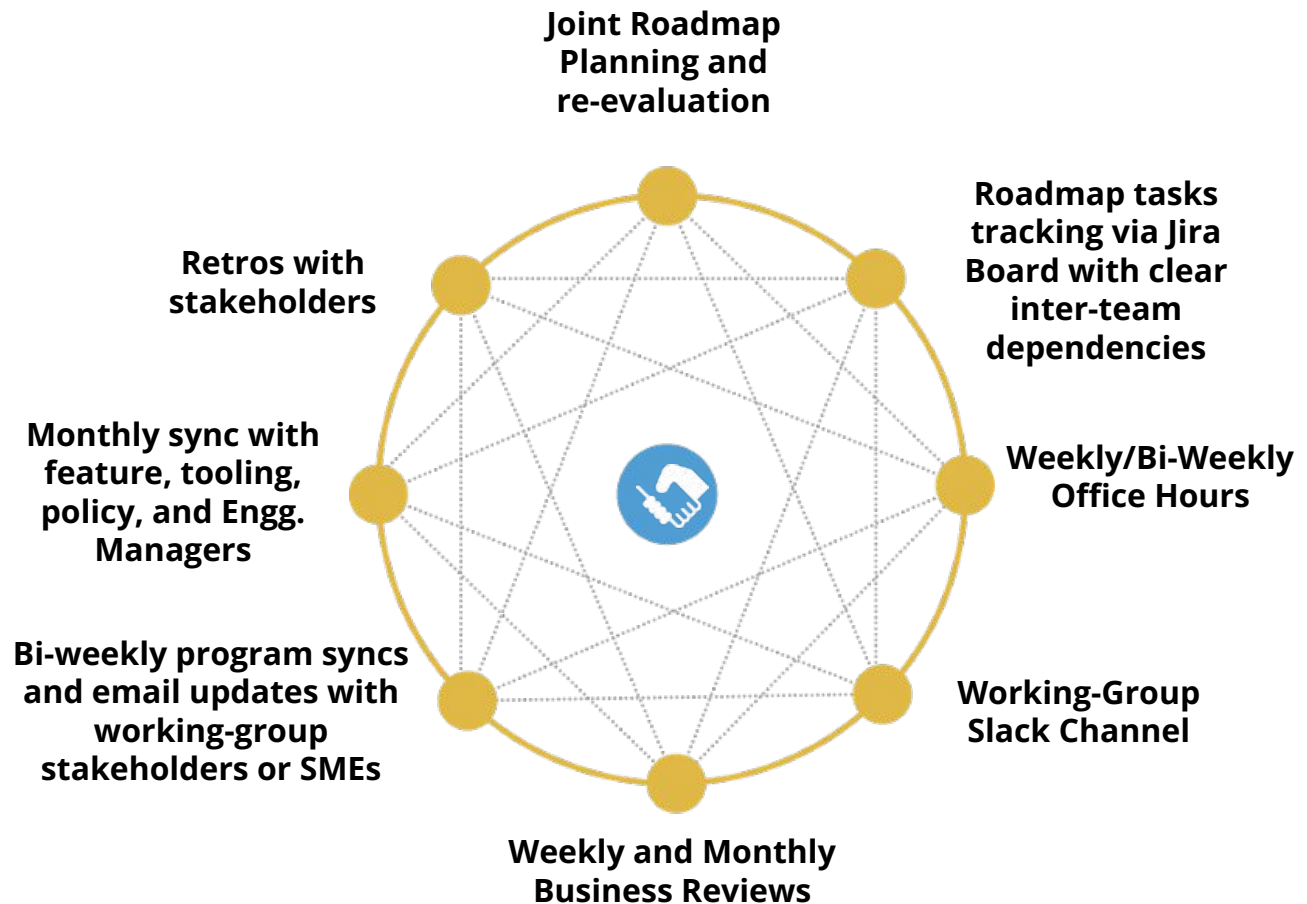
**Customer
Escalations by
different channels**

**Customer Reports
volume by
workflow/issues
and SSAT/CSAT
feedback reviews**

**VOC Reports of insights
build via customers
advisory board or various
customer-segment
interviews**

**@Flying-Uber
Twitter-Support
or other external
channels listening**

Cross-Team Collaboration Model



Appendix

Product Proposal





Product Discovery - Introduction and Hypothesis

Introduction

In this project, we have created the MVP launch strategy for the first flying car taxi service, Flying Uber, in one of the most congested cities in America -- New York City.

We will start with building a product proposal including initial data analysis and then proposing the data strategy. We will also be applying iterative design principles to a live product and perform funnel & cohort analyses to surface the improvement opportunities along with an experimentation plan.

As a Data Product Manager/Analyst, working with multiple teams and stakeholders is imperative to success. To understand what the needs of the customers are, what scale the company is growing at, and how to build for the future, we need to consider all relevant stakeholders. In this proposal, I am presenting the findings along with the analysis and reasoning behind the choices made in order to help Flying Uber continue its success. I have utilised some datasets from a few sites like Kaggle and Udacity to support my hypothesis and research.

Usage of Taxis

It's a semi-private or shared mode of transport utilized by a passenger or group of passengers to commute between locations of their choice.

It is utilized mostly in metro cities mainly as an alternative to private mode of transport.

User Personas

Characteristics of the users that leverage the taxis	Existing pain points with taxis	Existing pain points with digital ride-sharing services
Tourists (visiting the city for exploration and fun)	Lower service quality due to lack of training or accountability	Sometimes riders don't have local phone connections or app accounts to utilize the service or driver is new and not aware of the city which leads to hassle for riders
Daily Commuters (passengers who travel between distances for work or study)	Expensive for long distances or during rush hours	Sometimes the taxi drivers artificially inflate the prices. Unreliable availability. A recent article by <u>Natalie B. Compton in WP</u> highlights a non-dependable service offered by ride-sharing companies.
Occasional riders or technophobics (riders travel very specific trips few times in a month/year and would prefer to walk or ride a bike whenever permissible or people who not tech savvy)	Taxis are not always available during odd hours or when the riders have to make multiple stops for some reasons	Car service hired using rideshare companies don't follow the same regulations as taxi service so the crime risks are higher

Impact Hypothesis

User Improvement over the Existing State of Taxis

Flying Uber can potentially help in lowering the operating and maintenance costs of taxis as well as drivers' driving conditions by reducing the distance-to-duration ratio.

Market Improvements on Existing Taxi Service Industry & Physical Road Infrastructure

Flying Uber can possibly improve the reliability and availability of the taxi service as well as reduce the congestion, road expansion pressure and pollution problem by easing the traffic burden on the roads.



Data Exploration - Market and User Insights

Data Exploration

High-level Understanding of the Dataset

- Number of records in the dataset
1,048,468
- Record representation
Each record shows each taxi ride information taken by the rider(s)
- Primary key
Id
- Date range of the dataset
Dec-27-2015 to Jun-26-2016 (6 months of data)
- Geographical bounds of the dataset
Most of the data points are concentrated at New York area with few outliers in JFK (John F. Kennedy Airport) and LGA (Laguardia Airport) locations.

Distribution of Major Dimensions

Duration Spread

Avg. Duration	962.20
Median Duration	662.00
1st Std. Dev. of Duration	5,853.30
2nd Std. Dev of Duration	11,706.59

Distance Spread

Avg. Distance	3.44
Median Distance	2.09
1st Std. dev. of Distance	4.38
2nd Std. Dev of Distance	8.76

Passengers Count Spread

Avg. Passenger Count	2
Median Passenger Count	1
1st Std. Dev. of Passenger Count	1
2nd Std. Dev of passenger Count	3

Price Spread

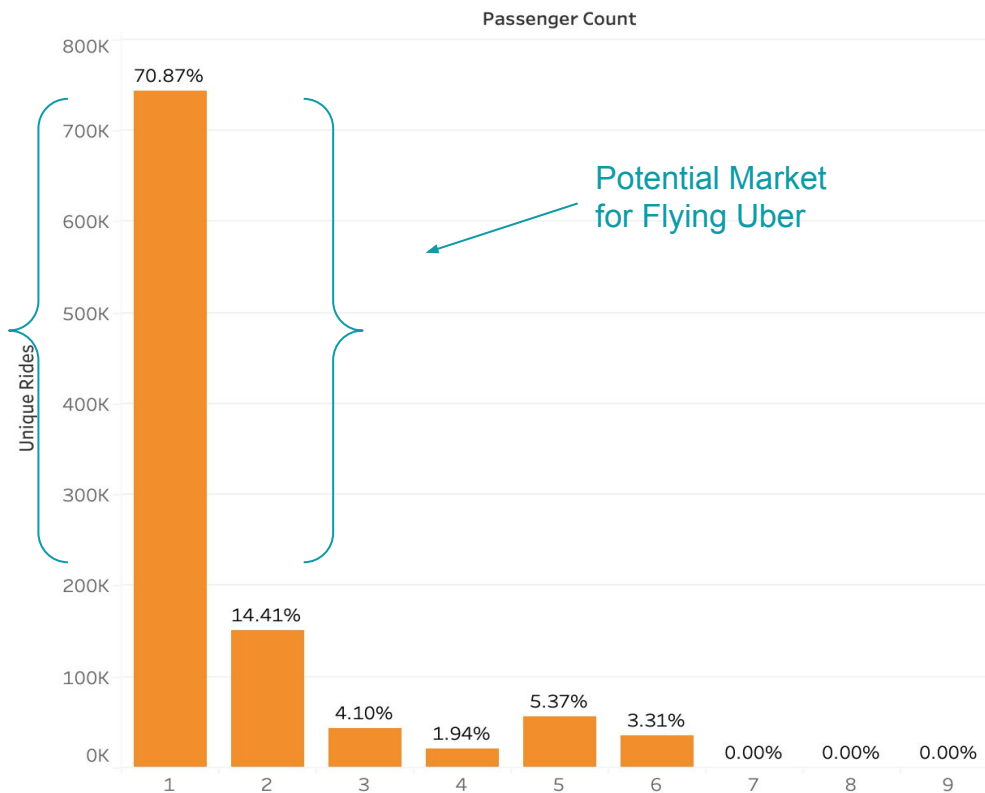
Avg. Price	8.0
Median Price	5.9
1st. Std. dev. of Price	6.7
2nd Std. Dev of Price	13.4

Distance to Duration Ratio Spread

Avg. Distance to Duration	6.12
Median Distance to Duration	4.64
1st Std. dev. of Distance to Duration Ratio	34.85
2nd Std. Dev of Distance to Duration Ratio	69.71

Data Exploration

Potential Market Volume of Passenger (1-2 each ride)
Pickups is ~85%



Data Exploration

Possible Flying Uber pick-up / drop-off nodes

- Zip Codes 11371 and 10018 experience highest number of pickups and dropoffs
- Zip code 11430 (highlighted in red below) has the highest duration-to-distance for pickup location and zip code 11371 has the highest for drop-off location.
- Setting up nodes in these neighbourhoods prove most optimum as it can reduce the taxi congestion and wait times for riders which are the some of the biggest pain points of taxi riders.

Summary by Top pickup points

Zip Code	Total Trips	Distance	Duration	Distance to Duration ratio
11371	8,753	84,010	16,428,786	0.00511
10018	6,818	18,234	7,601,368	0.00240
11430	6,135	105,562	15,258,768	0.00692
11369	4,340	43,736	8,883,763	0.00492
100036	713	1,994	809,539	0.00246
10001	654	1,723	585,817	0.00294
10028	185	498	181,955	0.00274
10014	161	435	135,110	0.00322

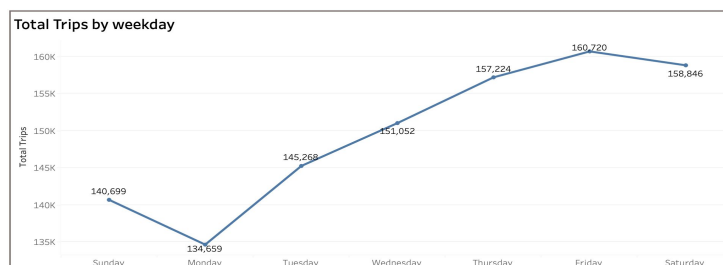
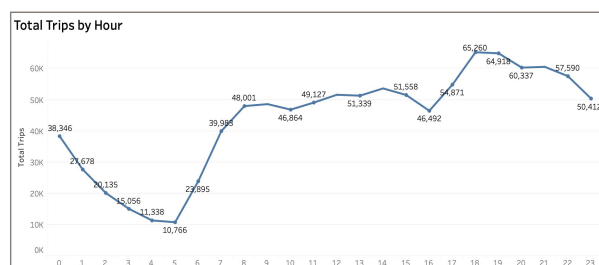
Summary by Top dropoff points

Zip Code	Total Trips	Distance	Duration	Distance to Duration ratio
10018	4,067	9,026	3,761,633	0.00240
11371	2,774	26,537	4,673,237	0.00568
10001	2,470	5,511	2,152,195	0.00256
11369	2,424	24,285	4,702,974	0.00516
100036	643	1,521	620,793	0.00245
100014	506	1,162	496,119	0.00234
10014	404	1,214	407,836	0.00298
11430	208	3,901	551,447	0.00707
10017	180	427	127,729	0.00334

Data Exploration

Week and Hour Taxi Usage Trends

- 1800 hrs (6pm) to 2100 hrs (9pm) are the rush hours of the day with the maximum demand on Fridays and Saturdays. Launching the Flying Uber service during these time periods can potentially maximize the ROI of the project. The heat-map on the bottom right shows the pickup details by hour and day.



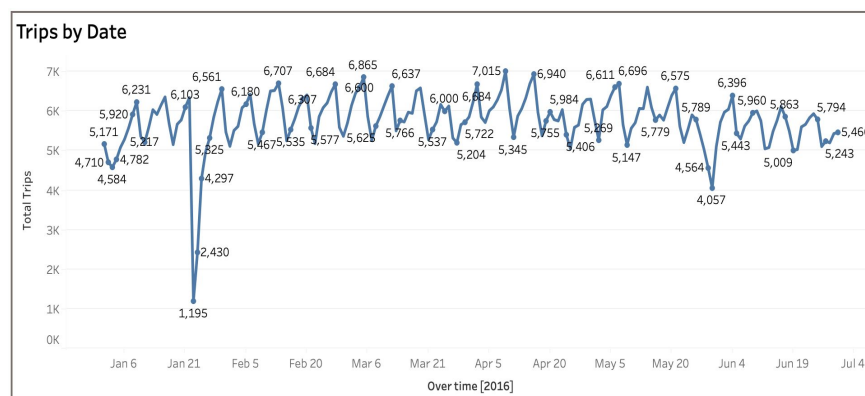
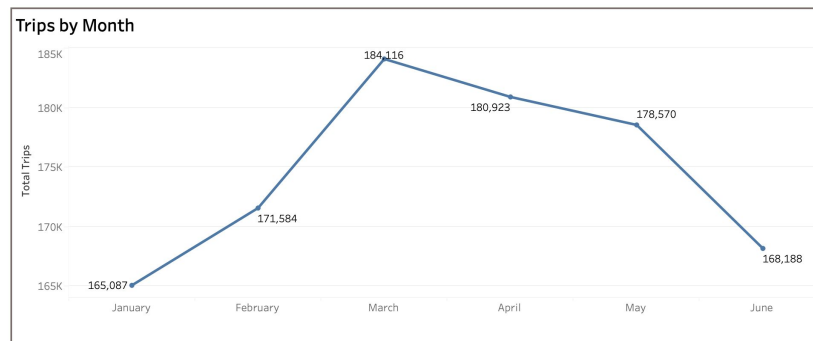
Heatmap - Trips %

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	% Total
0	0.85%	0.30%	0.30%	0.36%	0.44%	0.60%	0.81%	3.66%
1	0.74%	0.17%	0.17%	0.21%	0.25%	0.40%	0.69%	2.64%
2	0.59%	0.11%	0.10%	0.13%	0.15%	0.27%	0.56%	1.92%
3	0.47%	0.08%	0.07%	0.09%	0.11%	0.20%	0.41%	1.44%
4	0.31%	0.09%	0.08%	0.09%	0.10%	0.16%	0.26%	1.08%
5	0.13%	0.15%	0.15%	0.15%	0.15%	0.17%	0.14%	1.03%
6	0.15%	0.37%	0.39%	0.40%	0.42%	0.38%	0.17%	2.28%
7	0.20%	0.59%	0.69%	0.71%	0.71%	0.66%	0.25%	3.81%
8	0.29%	0.71%	0.79%	0.82%	0.81%	0.78%	0.37%	4.58%
9	0.44%	0.69%	0.75%	0.75%	0.75%	0.72%	0.54%	4.64%
10	0.59%	0.59%	0.64%	0.67%	0.68%	0.66%	0.65%	4.47%
11	0.70%	0.60%	0.66%	0.67%	0.69%	0.66%	0.70%	4.69%
12	0.73%	0.62%	0.67%	0.71%	0.72%	0.70%	0.77%	4.92%
13	0.73%	0.62%	0.67%	0.70%	0.71%	0.70%	0.78%	4.90%
14	0.73%	0.68%	0.72%	0.73%	0.75%	0.76%	0.75%	5.12%
15	0.69%	0.68%	0.69%	0.69%	0.71%	0.70%	0.75%	4.92%
16	0.68%	0.63%	0.62%	0.61%	0.60%	0.60%	0.69%	4.43%
17	0.74%	0.74%	0.76%	0.74%	0.73%	0.75%	0.78%	5.23%
18	0.76%	0.89%	0.92%	0.91%	0.93%	0.92%	0.89%	6.22%
19	0.69%	0.85%	0.90%	0.93%	0.95%	0.97%	0.90%	6.19%
20	0.62%	0.78%	0.87%	0.91%	0.93%	0.88%	0.76%	5.75%
21	0.60%	0.77%	0.87%	0.92%	0.98%	0.86%	0.78%	5.78%
22	0.54%	0.66%	0.79%	0.85%	0.91%	0.91%	0.84%	5.49%
23	0.44%	0.48%	0.60%	0.66%	0.81%	0.91%	0.91%	4.81%
Grand Total	13.42%	12.84%	13.86%	14.41%	15.00%	15.33%	15.15%	100.00%

Data Exploration

Seasonal Trends

- Based on the monthly and weekly pickup trends, March has the highest demand followed by April and May. The weekly needs are pretty consistent with major outlier in Jan 3rd week as shown in the weekly trend below. Variations beyond below shown trends may explain adoption gaps post launch of Flying Uber.

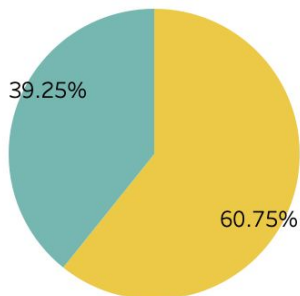


User Research Insights

Potential Flying Uber Adoption Rate is 80%

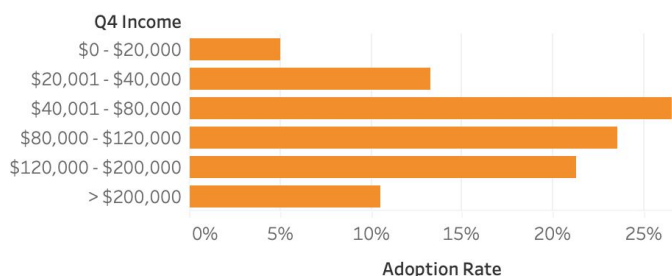
- Female riders have ~61% adoption rate as compared to males with ~39% adoption
- ~27% of the interested riders have income level range of \$40K - \$80K (both male and female) and 24% are having range of \$80K - \$120K
- ~92% of the adopters are between the age of 18-58
- Majority of the Adopters reside in Midtown, Battery Park City and Financial District

Adoption by Gender

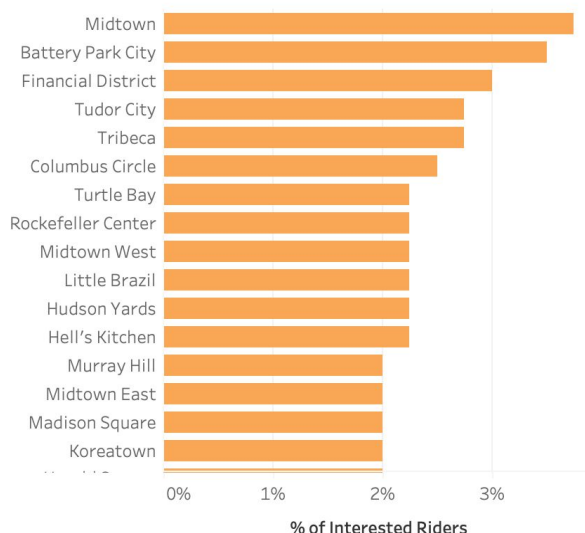


Gender
F
M

Adoption by Income Range



Adoption by Neighborhood



Adoption by Age

Age Range	
18-22	30.50%
36-40	30.50%
54-58	31.00%
72-76	8.00%

User Research Insights

On an Average riders are willing to pay USD \$23 per mile

Avg. Price/Mile by Gender		Avg. Price/Mile by Age		Avg. Price/Mile by Income	
Q2 Gender		Age Range		Q4 Income	
Female	23	18-22	24	\$0 - \$20,000	12
Male	23	36-40	22	\$20,001 - \$40,000	15
		54-58	24	\$40,001 - \$80,000	18
		72-76	23	\$80,000 - \$120,000	22
				\$120,000 - \$200,000	31
				> \$200,000	37

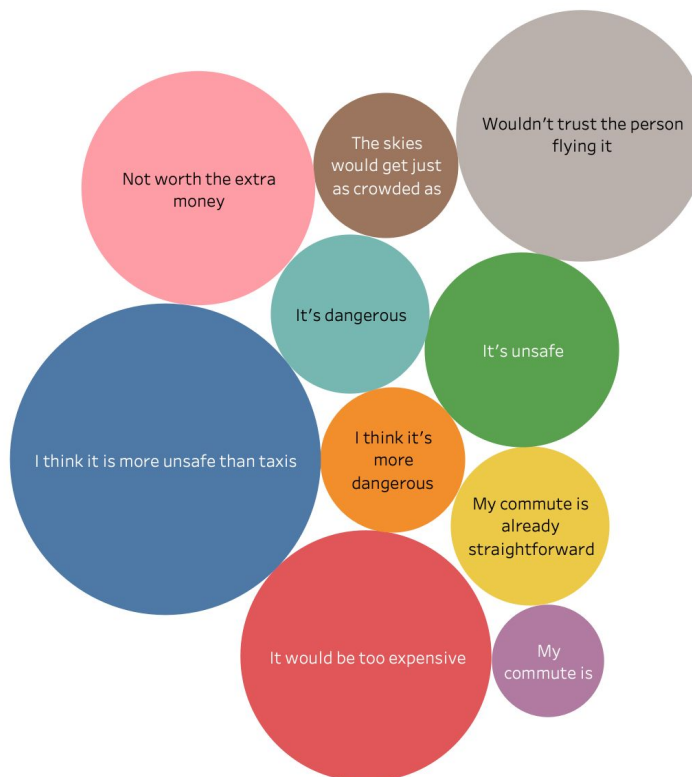
Avg. Price/Mile by Neighborhood

Neighborhood	
Inwood	34
Financial District	33
Bowery	33
Battery Park City	33
Columbus Circle	30
Tribeca	29
Lower East Side	28
West Village	27
Downtown Manhattan	27
Diamond District	26
Midtown East	25
Little Italy	25
Harlem	25
Fort George	25
Midtown West	25
Five Points	25
Lenox Hill	25
Madison Square	25
Little Australia	25
Nolita	25
St. Nicholas Historic Distr..	25
Upper East Side	25
SoHo	24
Garment District	24
Le Petit Senegal (Little Se..	24
Tudor City	24
Greenwich Village	24

User Research Insights

Safety and Cost of the service are the dominating themes in negative sentiment

Segments of Negative Sentiments





Proposal Synthesis

Proposal Synthesis

Vision:

Flying Uber aims to improve the city taxi service for its riders by providing access to safe and affordable air-taxi service that will ameliorate their travel conditions by reducing travel time for taxi riders, and reduce the road traffic congestion.

Objectives:

Our objectives are primarily focusing on user acquisition and engagement as adoption and usage of the new service will help build trust in the company as generate interest in stakeholders for future investments.

Objective-1

Though effective marketing and promotions, the company should facilitate adoption of air-taxi-service that ensures safety, reliability and availability to its riders.

Objective-2

Provide an air-taxi-service that reduces the wait-time and travel-time for its riders and hence improve customer satisfaction index

Key Performance Indicators

KPI 1 - Increase in number of trips taken by riders monthly.

KPI 2 - Increase in average CSAT (Customer Satisfaction) score

KPI 3 - Reduction in average wait time for taxi-riders

KPI 4 - Reduction in average travel-time per trip compared to road-taxi service

KPI 5 - Reduction in average Distance to Duration Ratio

Objective-1 Key Results

- With effective branding and marketing efforts, Flying Uber should be able to complete 10% of the current road-taxi service trips per month by the end of 3 months of launch date.
- Attain monthly average CSAT (Customer Satisfaction) Rating of 3 stars (*).

Objective-2 Key Results

- Reduce average wait-time per trip to 3 min
- Reduce average travel-time per trip to 10 min
- Reduce average Distance to Duration Ratio by 10%

Proposal Synthesis

MVP Plan

	Plan	Justification
Days of Operation	- Thursdays, Fridays and Saturdays 18 to 22 hrs i.e. from 6 PM to 11 PM	Road-taxi service has peak demand during these times
Pick-up & Drop-off Nodes Locations	2 nodes at zip codes 11371 and 11369 in Flushing and East Elmhurst, NY neighborhood 1 node at zip code 11430 in Jamaica, NY neighborhood	These locations has highest number of pickups and dropoffs as well as highest duration-to-distance ratio
Initial Equipment	Copters	To save the initial launch cost and time, Flying Uber should consider launching service using Copters
Pricing	Fixed Price plan	Fixed pricing plan will invoke interest among maximum number of passengers

Proposal Synthesis

Instrumentation Plan

Event	Event Metadata	Event Triggered	KPIs	Formula
TripBooked	- rider_id/login_userid - rider_location - passenger_count - booking_timestamp	Triggered when the rider login to the app and hit book a ride button	Number of unique customers	Count of unique rider id passenger count each ride.
TripStarted	- rider_id/login_userid - Trip booking number - node_pickup_location - start_timestamp	Triggered when rider has scanned the QR code and boarded the taxi	Avg. wait time	Wait time = duration between booking time and start time
TripCompleted	- rider_id/login_userid - node_dropoff_location - End_timestamp - Distance travelled	Triggered when rider has scanned the QR code after offboarding from the taxi	Average travel-time per trip Average Distance to Duration Ratio	Travel time = duration between start time and trip end time Distance to Duration Ratio = duration / distance
Feedback Rating	- rider_id/login_userid - rating for Q1-1,2,3 - option for Q2 - opinion for Q3 - comments text - overall rating	Triggered when rider responds to the feedback survey in the app. Incomplete surveys will also be captured.	Average CSAT (Customer Satisfaction) score	Avg CSAT score = sum of scores / total count of unique ratings

Proposal Synthesis

Feedback Survey Questions

Did you find the ride service experience?

- 1) Rate the booking * * * * *
- 2) Rate the ride * * * * *
- 3) Rate the driver * * * * *

Why do you most often use a taxi?

- Airport, Rail or Port transfer
- To and from entertainment events
- Commute to work
- Following a road accident or breakdown
- Other:

In your opinion, are taxi prices affordable?

- Rightly priced
- Would pay more
- Expensive

Overall rating: How would you rate overall experience?

* * * * *

Comments: Please let us know your thoughts about the ride experience



Proposal Synthesis

Flying Uber Proposal

Pain Points

- Long travel time for riders
- High operating costs and unhealthy driving conditions for taxi drivers due to traffic congestion especially during peak hours.

Target Customers

- Primary: Male and female taxi-riders between the age group 18-58 and have an income range above \$40K
- Secondary: Road-taxi drivers who face unhealthy driving conditions due to traffic congestion

Solution

Safe and reliable air-taxi service that reduces the travel time for its riders, and improve driving conditions for taxi drivers by reducing traffic congestion

Customer Impact

- Reduce travel time for taxi-riders
- Improve customer satisfaction among taxi-riders
- Increase wellness among taxi-drivers.

Market Impact

- Supplement traditional road-taxi service
- Improve living conditions for city residents and ensure safe road conditions for pedestrians and cyclists
- Reduce city noise and air pollution, as well as need for road infra expansion by easing the traffic congestion on the roads.

Business Impact

- Align with company's vision of improving the current road-taxi service conditions
- Reduce churn rate for taxi-service industry by addressing the emotional worries around the current road taxi system. Improves customer satisfaction by offering safe, reliable, and availability without long wait times.

Proposal Synthesis

Flying Uber Proposal

Risks and Assumptions

- Establishing trust and safety among initial adopters will be preeminent for the success or growth of the concept
- Compliance with current air traffic regulations
- The long term development costs are high and needs a lot of testing and continuous improvement
- Special training and certification program cost for the drivers or hire trained drivers who may charge higher wages.

Cross-functional Stakeholder teams

- CEO and Investors
- Product Development team
- User Experience team
- Operations Management team
- Customer Success team

Launch Strategy

Soft launch with promotional pricing to gain early adopters among different personas during peak hours who are willing to share product critique and success stories

Data Strategy



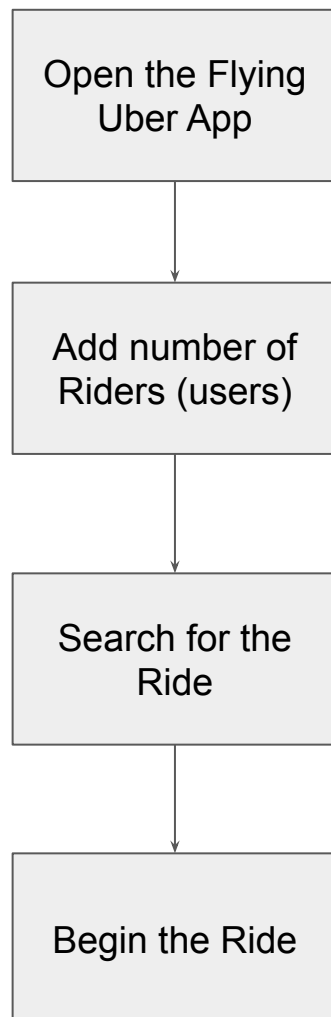


Data Customers, their Needs, and Data Model

Customer Funnel

Flying Uber is a two-sided platform. Riders are the customers and drivers/pilots are partners. For the Minimum Viable Product, we will be focusing on the Riders side of the business. To build an end to end data pipeline the very first step is to understand who needs data and why they need that data. Within Flying Uber, we need to identify who will be the primary data customers/stakeholders, why they are your primary data stakeholders and how they want to use the data (primary use-cases).

Overview of Riders Process Flow (Customer funnel)



Internal Stakeholders and Their Needs

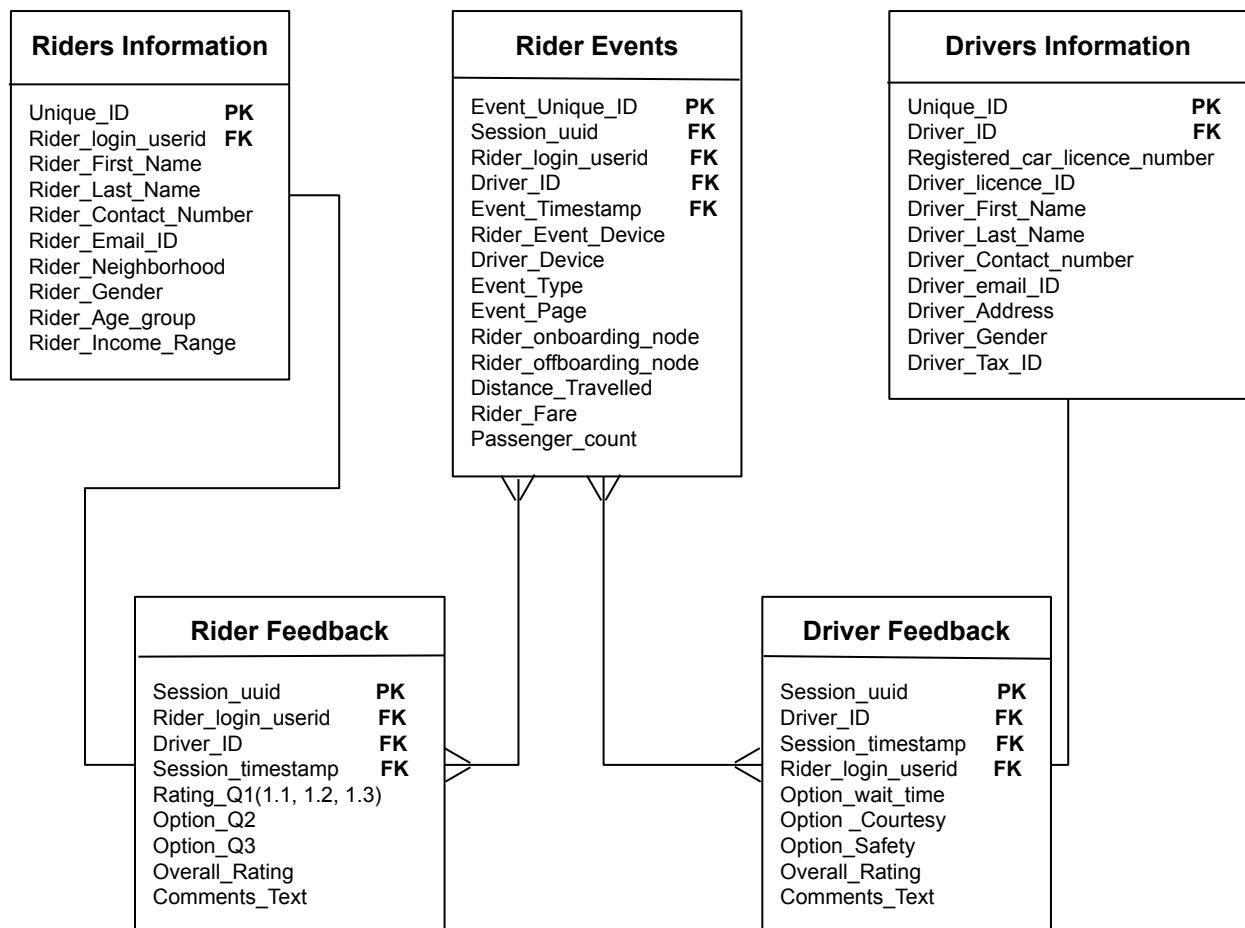
Stakeholder	Why are they primary stakeholders?	Use-Case
Engineering	They are responsible for building and maintaining a working App/product	Monitoring app performance. Tracking app or feature usage.
Product Management	They need to understand the usage of the product to drive customer acquisition and retention	Identifying riders pain points or product expansion opportunities
Marketing	They are responsible for new customer acquisition and increase product reach.	Developing target marketing campaigns to raise product awareness and facilitate adoption
Customer Success/Customer Operations Management	They need to provide personalized customer resolutions and resolve their grievances. Also they will act as a feedback loop for other stakeholders or teams	Resolving or addressing current riders issues. Build inferences from customer feedback system
Logistics team	They are managing the availability of the rides	Monitoring rides availability during peak hours and track riders/customers wait times
Finance/Accounting	They are Managing billing system and reporting on company's financial health	Building profit and loss statements and resolving billing issues when they arise.

Data Collection

Stakeholder	Use-Case	Data	Why is this the primary use-case?
Engineering	Monitoring app performance	Event, Entity	Having a working app is core to the business as it enables the riders to utilize the offered air-taxi service.
Product Management	Identifying riders pain points or product expansion opportunities	Entity, event	Addressing pain points will ensure consistent business growth and build a successful business model overtime.
Marketing	Developing campaigns to raise product awareness	Entity, event	Accurate targeting and building awareness among right personas and help with customer acquisition and product expansion.
Customer Success/Customer Operations Management	Resolving or addressing current riders issues. Build inferences from customer feedback system	Event, Entity	Addressing customer issues where they are will strengthen the retention and hence help with product growth.
Logistics team	Make sure rides are available for booking	Event	The team ensures the effective management of supply chain ie. availability of air-taxis
Finance/Accounting	Building profit and loss statements and resolving billing issues when they arise	Entity, Event	Help effectively sustain the business by measuring the financial health of the product including revenue and profit margins.

Data Modelling

ER Diagram of Tables



Data Modelling

Table 1:

Rider Information

Unique_ID	Primary Key
Rider_login_userid	Foreign Key
Rider_First_Name	
Rider_Last_Name	
Rider_Contact_Number	
Rider_Email_ID	
Rider_Neighborhood	
Rider_Gender	
Rider_Age_group	

Rationale for Choosing Primary and Foreign Keys for the Table 1:

To facilitate frictionless onboarding, the team may decide to onboard customers with phone only, or email only, or login_userid only. In that case the Unique_ID field will serve in identifying unique customers.

A primary key uniquely identifies every record in the table and cannot be null, while foreign key creates a relationship with other tables.

Table 2:

Drivers Information

Unique_ID	Primary Key
Driver_ID	Foreign Key
Registered_Car_licence Number	
Driver_Licence_ID	
Driver_First_Name	
Driver_Last_Name	
Driver_Contact_number	
Driver_email_ID	
Driver_Address	
Driver_Gender	
Driver_Tax_ID	

To facilitate frictionless onboarding, the team may decide to onboard drivers with phone only, or email only, or login_userid only. In that case the Unique_ID field will serve in identifying unique drivers.

Data Modelling

Table 3:

Rider Events

<i>Event_Unique_ID</i>	Primary Key
<i>session_uuid</i>	Foreign Key
<i>Rider_login_userid</i>	Foreign Key
<i>Driver_ID</i>	Foreign Key
<i>Event_Timestamp</i>	
<i>Rider_Event_Device</i>	
<i>Driver_Device</i>	
<i>Event_Type</i>	
<i>Event_Page</i>	
<i>Rider_onboarding_node</i>	
<i>Rider_offboarding_node</i>	
<i>Distance_Travelled</i>	
<i>Rider_Fare</i>	
<i>Passenger_count</i>	

Table 4:

Rider Feedback

<i>session_uuid</i>	Primary Key
<i>Rider_login_userid</i>	Foreign Key
<i>Driver_ID</i>	Foreign Key
<i>Session_timestamp</i>	
<i>Rating_Q1(1.1, 1.2, 1.3)</i>	
<i>Option_Q2</i>	
<i>Option_Q3</i>	
<i>Overall_Rating</i>	
<i>Comments_Text</i>	

Table 5:

Driver Feedback

<i>session_uuid</i>	Primary Key
<i>Driver_ID</i>	Foreign Key
<i>Rider_login_userid</i>	Foreign Key
<i>Session_timestamp</i>	
<i>Option_wait_time</i>	
<i>Option_Courtesy</i>	
<i>Option_Safety</i>	
<i>Overall_Rating</i>	
<i>Comments_Text</i>	

Making Dataset Extraction Decisions

For making business decisions, a data consumer would like to have all the data they want. However, for any ecosystem, it is impossible to collect or provide everything that the customers need. All the resources are not always available to get what you need. Most of the time we need to get creative and get the most insights with a minimal data set.

Choose some of the following prompts may help in choosing the most relevant information to proceed further:

- How many events are being recorded per day?
- How many events of each event type per day?
- How many events per device type per day?
- How many events per page type per day?
- How many events for each location per day?
- How much is the customer data increasing?
- How much is the transactional data increasing?
- How much is the event log data increasing?

Commonly, the system generated following types of data-logs:

- Event Log Data
- Transactional Data
- Customer Data

Making Dataset Extraction Decisions

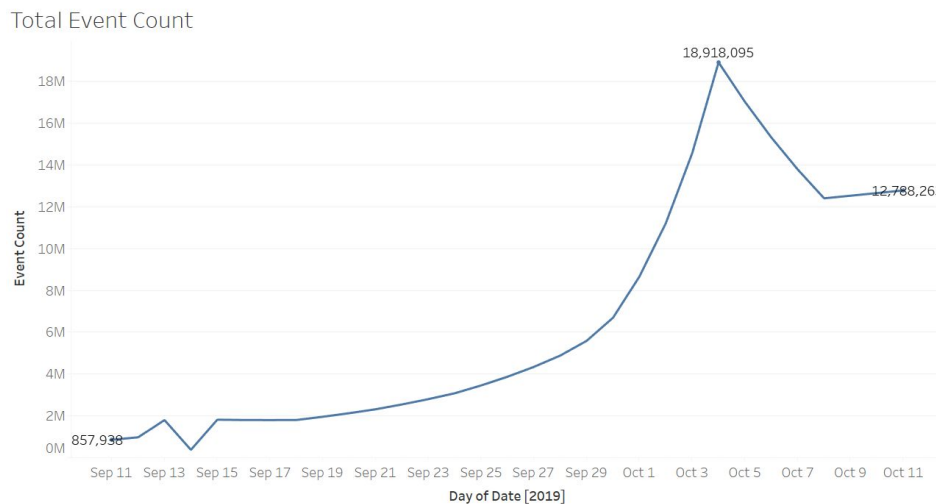
- Event log data: Tracks the user activities in the app like searches, bookings, or rides completed by the customers.
- Transactional data: Transactions track a complete activity of customers e.g. Aggregated transaction data showing users count who have created new accounts, or rides completed per day. Transactional data will help build insights like:
 - When were the rides taken?
 - By whom?
 - How much fare is generated by each location/neighbourhood?
- Customer data: Tracks the lifetime customer base or specific attributes of customers like names, addresses, contact details. Customer data will help solve specific use cases for stakeholders like marketing, customer care or product while maintaining better security and control over PII data.

Prioritizing ETL automation for pipelines where the data growth is highest and which provide holistic view of day-to-day activity of the customers makes most sense as it will reduce the most frequently needed critical manual efforts. Most of the time, events data provides the holistic view of customer activities in the app.

In the next section, we will be using the provided sample events data to visualize the trends and build initial understanding of the business growth.

[Aggregated Events Data](#)

Business Insights from the Data Load



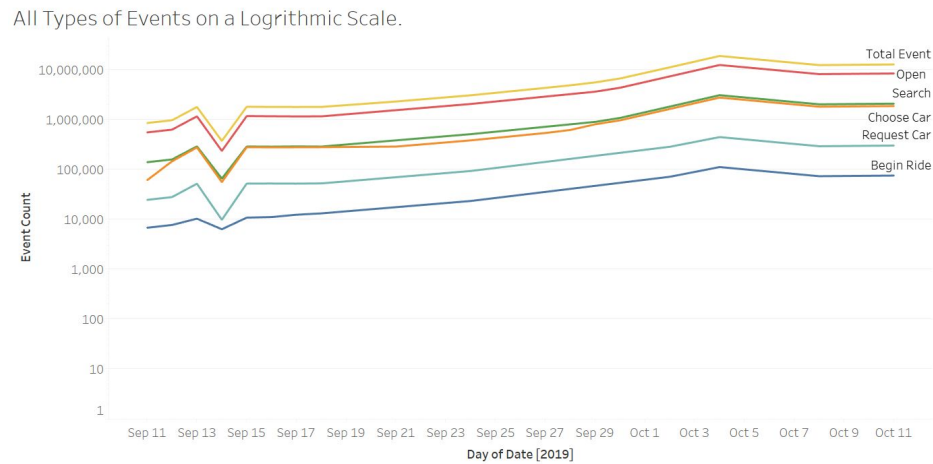
***Compound Growth Rate of Total Events: 9.7%**

Observations:

- All Events are growing showing compound growth rate at 9.7% from the launch day. It infers a healthy customer acquisition trend in its first month.
- Looking at the Event count graph, Sep 14 shows a drastic dip which might be a result of some technical app issue. Deep diving into the issue can help with identifying the reasons and design a remediation plan to avoid its recurrence.
- There is a remarkable spike on Oct 4 which might be a result of some marketing campaign. A close collaboration between different stakeholders (product, marketing, engineering) during the campaign launch period can ensure positive customer experience in the app and success of the product.

*Compound Growth Rate = $(\text{total events on 10-11-2019} / \text{total events 09-11-2019})^{(1 / \text{total days})} - 1$

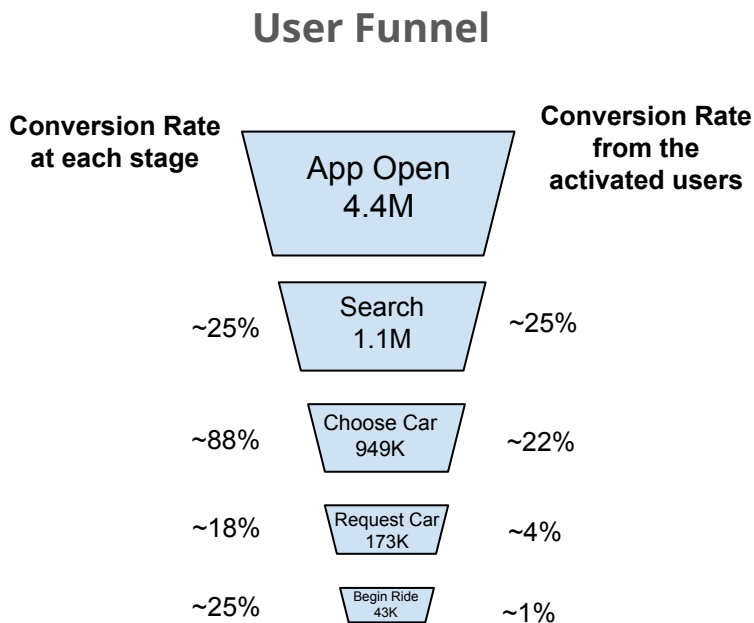
Business Insights from the Data Load



Observations:

- Overall the graph shows an upward trend and depicting a healthy activation rate.
- A sudden dip on Sep 14 might have occurred due to some unexpected technical glitch or uncontrollable external event. Understanding the reason can help with avoiding future recurrences and hence saving Flying Uber market reputation.
- There is a significant growth on Oct 4 that might be a result of a marketing campaign. It's a good practice to facilitate the campaign launch with close collaboration with various company departments/teams to ensure customers have a positive experience in the app and the company collects all events data for future growth strategies.
- Collecting data through planned marketing campaign may help Flying Uber with:
 - Understand customer behavior and identify the inflection point and prioritize data-informed development of the product.
 - Build a cost effective scalable data infrastructure
 - Effectively understanding customer needs and behavior to build predictive BI solutions that show peak hours trends, customer habits and preferences, trends, and many more insights.

Business Insights from the Data Load



Observations:

- User Funnel - where to invest:
Flying Uber will generate revenue only when customers actually utilize the service i.e. take rides. Looking at the user funnel, only 1% of the total customers are actually using the service. Deep diving further and identifying and remediating friction points at each stage is necessary for the success of the business.
- Prioritizing optimization at the ride search and request car stage could help with doubling the conversion rate to the bottom of the funnel quickly.



Data Infrastructure Strategy

Data Infrastructure Strategy

Choosing data warehouse solution

So far we have:

- Identified data stakeholders and their data needs.
- Identified what data is currently being collected and what data needs to be collected.
- Identified data insights and growth trends.

Now, it's time to tie all the loose threads together and bring this process to its logical conclusion by suggesting which Data Warehouse (DWH) Flying Uber should invest in and why.

Data helps with removing guesswork during decision making however, there is a difference between being data-aware and insights-driven. In the big data world where there are multiple data producers and consumers, selecting and running an effective data warehouse system comes with carefully selecting tradeoffs.

Data Infrastructure Strategy

Below table shows a comparison between the cloud and on-prem system on few main factors:

Factors	Cloud	On-Premise
Cost	Comparatively requires low upfront costs. The initial infrastructure setup is cheaper. Cloud software is usually priced under a monthly or annual subscription package, covering training, support, updates and other additional functionalities like BI solutions, etc. However, in the long run, subscription costs may end up costing more in total or if the company decides to move their data to another infrastructure in future.	Requires a large upfront infrastructure investment. The company will be responsible for the ongoing costs of the server hardware, power consumption, space and maintenance.
Scalability	Cloud offers more flexibility in terms of scaling up. As a third party company is maintaining the system, the data storage can be scaled comparatively faster by upgrading the subscription.	Since all resources are deployed in-house, the expansion costs can highly vary over the long term. Scalability may not happen faster for big data especially if the data warehouse is handling more unstructured data.
In-house Expertise	Need for expert IT staff is less as the cloud storage system is managed by another company. Many cloud companies cover initial setup as well maintenance or error handling as well.	Need in-house technical expertise to deploy and maintain complex big data systems. Initial implementation may take longer if the business needs multiple customizations.
Latency / Connectivity	May face higher latency or connectivity issues if the internet setup is non-reliant. The company needs an advanced and efficient internet access system to ensure availability.	Low latency or connectivity issues compared to cloud as the complete infrastructure is maintained on-site or on the intranet and may not be completely dependent on internet connectivity.
Reliability	When data is hosted on a cloud, the data security becomes a shared responsibility of the company and the third party vendor. Strong data encryption and access procedures and policies are needed to stay compliant with industry regulations.	If the company is dealing with more PII data then having an on-prem system ensures more control over the data and hence maintaining high compliance regulations.

Suggested Data Warehouse

Flying Uber has been successful in introducing a new service in the market due to close collaboration of various teams' efforts. For further successful expansion, we need to define a data instrumentation plan that helps the company to strengthen its differentiation strategy i.e. focus on the core value that the company brings to its customers. As defined in the initial proposal, we have two main objectives:

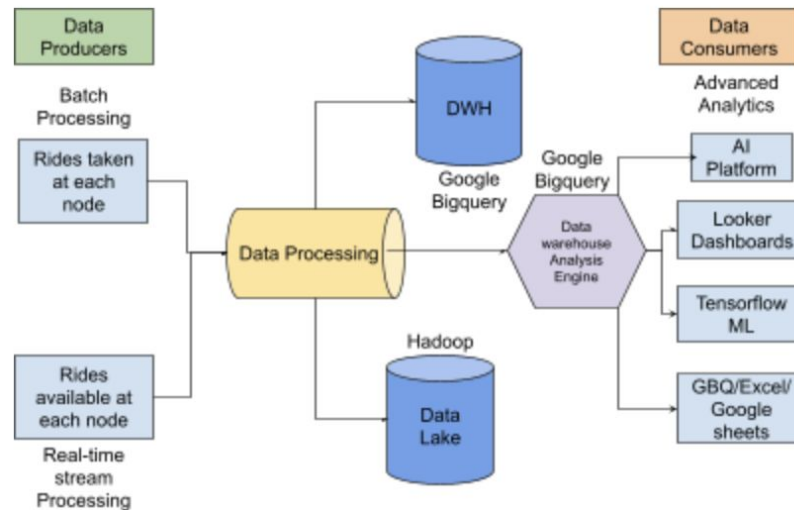
- Objective 1:
Through effective marketing and promotions, the company should facilitate adoption of air-taxi-service that ensures safety, reliability and availability to its riders.
- Objective 2:
Provide an air-taxi-service that reduces the wait-time and travel-time for its riders and hence improve customer satisfaction index

To expand and to offer outstanding experience to our customers, we need to focus on faster expansion in below mentioned areas:

- Running successful marketing campaigns and product management to improve retention and usage,
- Data science backbone to build strong understanding of the customer behaviour,
- Manage an effective supply chain to make sure enough rides are available for bookings.

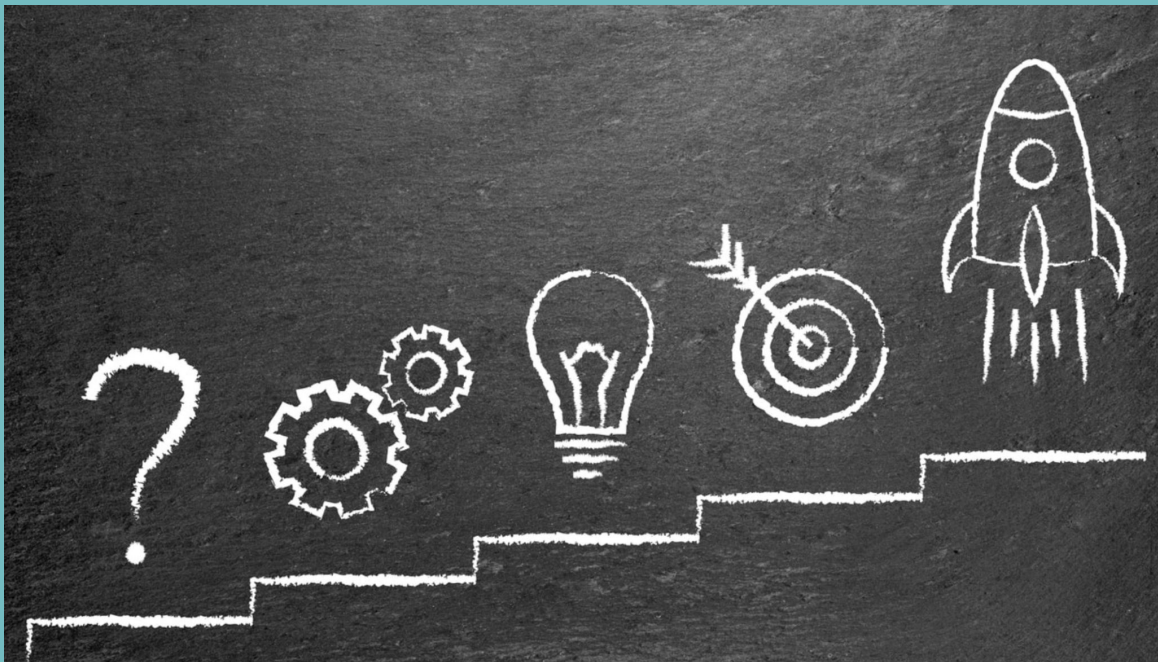
Suggested Data Warehouse

Flying Uber should take a hybrid data-warehouse system approach where OLTP (online transaction processing) system setup can be done on-premise on Hadoop and OLAP (Online Analytical Processing) system can run on a cloud-based Google Bigquery.



Factors	Hadoop	Google Bigquery
Cost	Hadoop is an open source framework that offers parallel processing at a lower cost.	Customers have the choice between on-demand or flat rate pricing. BigQuery uses a query-based pricing model where users are charged for the amount of data queried.
Scalability	Offers vast scalability on any type of data structured, semi-structured or unstructured data by processing from a single server to thousands of machines for large data processing needs. It will help Flying Uber scale up faster in future.	Automatically scale compute resources without manual provisioning. . Easily connects with other Google Cloud products like Google analytics, or looker etc. Hence there is less dependency and burden on in-house administrator expertise.
In-house Expertise	Big data teams may have to invest in more labor for on-premises implementations to provision new nodes and add them to a cluster	The platform manages all resources and automates scalability and availability including resolving scalability issues.
Latency / Connectivity	Stores and processes structured, semi-structured and unstructured data using files-system caching. The platform can be utilized with a spark engine to process time-sensitive jobs.	Google BigQuery supports partitioning of storage, and compute as separate operations thereby resulting in improved query performance.
Reliability	Offers data protection through an embedded fault tolerance system.	The platform encrypts data at rest and in transit by default.

Applying Iterative Design Principles to a Live Product





Step 1
Select KPIs
&
Evaluate Previous
Multivariate
Experiment Results

KPIs for Flying Uber Analysis

Before defining the KPIs, we need to understand the business model of Flying Uber:

There are three common business models:

- eCommerce: It's a commerce model where company generates revenue by selling items, goods or service online like amazon marketplace.
- Freemium: Under the model, the company allows its customers to use basic service or few features of the service free and pay for premium service or upgrades like Spotify.
- Software-as-a-service: Under this model, the customers paid a fixed amount to get access to the service or product for limited period of time like Adobe photoshop.

Since Flying Uber offers flying taxi service to its customers where a customer books a ride using Flying Uber app and pays a fixed price for the ride, its an ecommerce marketplace. The company's primary focus should be to increase the revenue and customer satisfaction.

To define and improve the KPIs, we should consider using Pirate Metrics framework, where we break customer's journey into following categories:

Acquisition	How users come to know about the service?
Activation	Do users have good experience?
Retention	Do users come back?
Revenue	How company makes money?
Referral	Do users spread the word about the service?

KPIs for Flying Uber Analysis

Now let's identify the KPIs based on the data available:

Category	KPI	How to measure
Acquisition	- Daily/Weekly customers trend - Unique Customers count by neighborhood - Unique customer who have opened the service or used the service	It can be derived from the event logs by filtering for customers who have opened the app
Activation	- Unique customers who have searched for the ride	It can be derived from the event logs by filtering for event type "search"
Retention	- Returned customers count	It can be derived from the event logs
Revenue	- Total Fare earned in a day/week - Daily Profit - Acquisition Cost per customer - Customer lifetime value	Can't be derived using current event logs but data instrumentation can be defined
Referral	- Customers accessing the service through referral code	Can't be derived using current event logs but data instrumentation can be defined

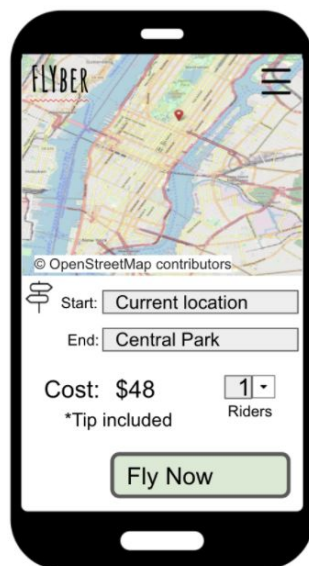
First Multivariate Experiment

In the experiment there are 2 testing elements “Tip Included” and “Fly Now” and 2 options for each element including original version. Different set of customers are exposed to 4 different experiments.

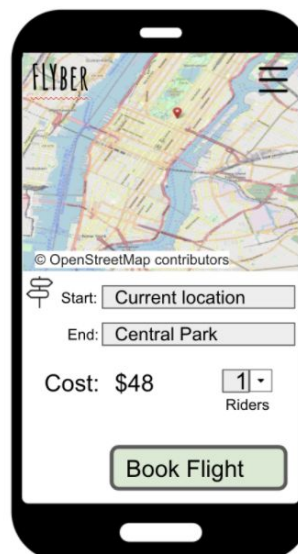
Control



Experiment 1



Experiment 2



Experiment 3



When running experiments, we must make sure that users in the control and the experiment group share the same characteristics to ensure the differences among experiment groups is caused by the change of the product, not because of different user characteristics.

Review Multivariate Test Results: Visualization

Multivariate test results

Experiment Group	Event Type			
	open	#_of_users	search	begin_ride
control		-42.37%	-45.91%	-98.46%
experiment_1		-42.25%	-45.81%	-98.28%
experiment_2		-42.11%	-44.61%	-98.27%
experiment_3		-42.70%	-45.59%	-98.29%

Multivariate Test Results

The results of experiments can be evaluated by a T-Test. T-test helps us understand whether any differences observed in the conversion rate are significant and are likely due to the change we made.

Review Multivariate Test Results: Significance Test

Determine if there was a significant difference between the experiments and control states.

To properly evaluate the results, we need to take following steps:

Step 1: State the null hypothesis. The null hypothesis represents the idea that the control state is not different from a test state, which is what we want to reject. For example, for multivariate test we are running for the Flying Uber, we hypothesize that there is no difference in the ride booking rate between the test state compared to control state.

Step 2: State the alternative hypothesis. The alternative hypothesis is the one we want to accept. Like in case of Flying Uber experiment we define alternative hypothesis that ride booking rate is different between the test and the control group.

Step 3: Set a confidence threshold. Confidence level sets if the experiment results appears statistically valid. For Flying Uber, we will set the confidence level at 95%.

Step 4: Now let compare the multivariate experiments results we ran earlier.

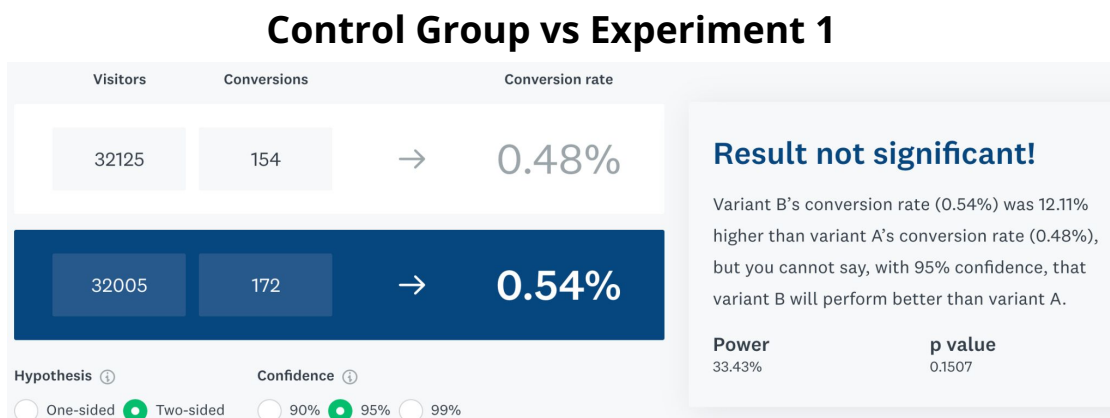
Review Multivariate Test Results: Significance Test

Step 5: Perform a T-test to see if the differences we see in the experiment are significant. We calculate and compare the p-value to find if the experiment is statistically significant. The p-value indicates the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. A smaller p-value means the evidence in favor of the alternative hypothesis is stronger. For Flying Uber multivariate test we ran 3 experiments against the control group.

Multivariate test results

Experiment Group	Event Type			
	open	#_of_users	search	begin_ride
control	32,125	18,514	10,015	154
experiment_1	32,005	18,482	10,015	172
experiment_2	32,216	18,650	10,330	179
experiment_3	32,105	18,397	10,009	171

- For Experiment-1, Using a two-sided T test at a 95% confidence interval, we find that the results of the test are not significant. We know the results are not significant because the p-value generated is 0.151 which is larger than the p-value of 0.025. An image of the p-value calculation using an [online statistical significance calculator](#) is included below:

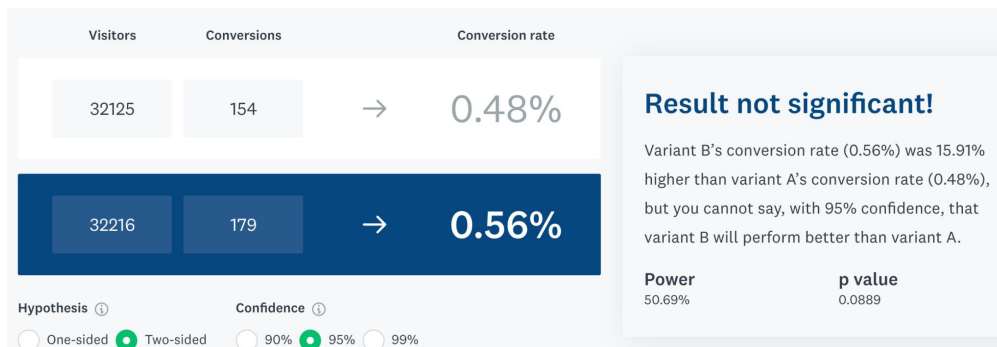


Review Multivariate Test Results: Significance Test

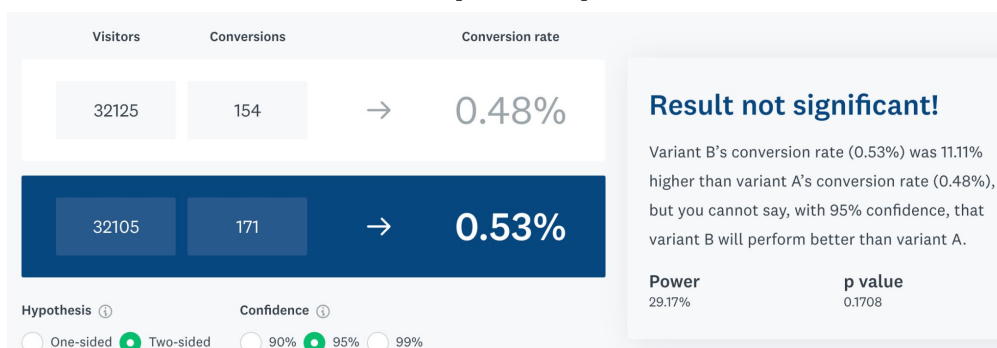
- Now let's evaluate the t-test for other two experiments as well.

Using a two-sided T test at a 95% confidence interval for Experiment 2 and 3, we find that the results of the tests are not significant either. The results are not significant as the p-value for Experiment-2 is 0.089 and Experiment-3 is 0.171 which is larger than the p-value of 0.025. The images of the p-value calculation are included below:

Control Group vs Experiment 2



Control Group vs Experiment 3



Review Multivariate Test Results: Significance Test

Step 6: Based on the T-test results in step-5 the recommendation is not to roll out or expand the changes suggested in experiment pages to the Flying Uber customers as the test conditions did not result in a statistically significant increase in conversions. We can safely assume that the difference between the versions are more likely due to random chance and may not lead to increase in number of rides taken by the customers.



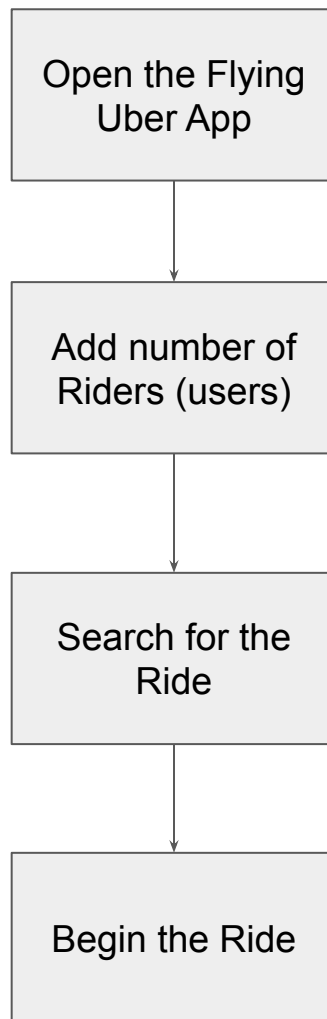
Step 2

Funnel & Cohort Analyses

User Funnel

Identifying the different stages the user funnel

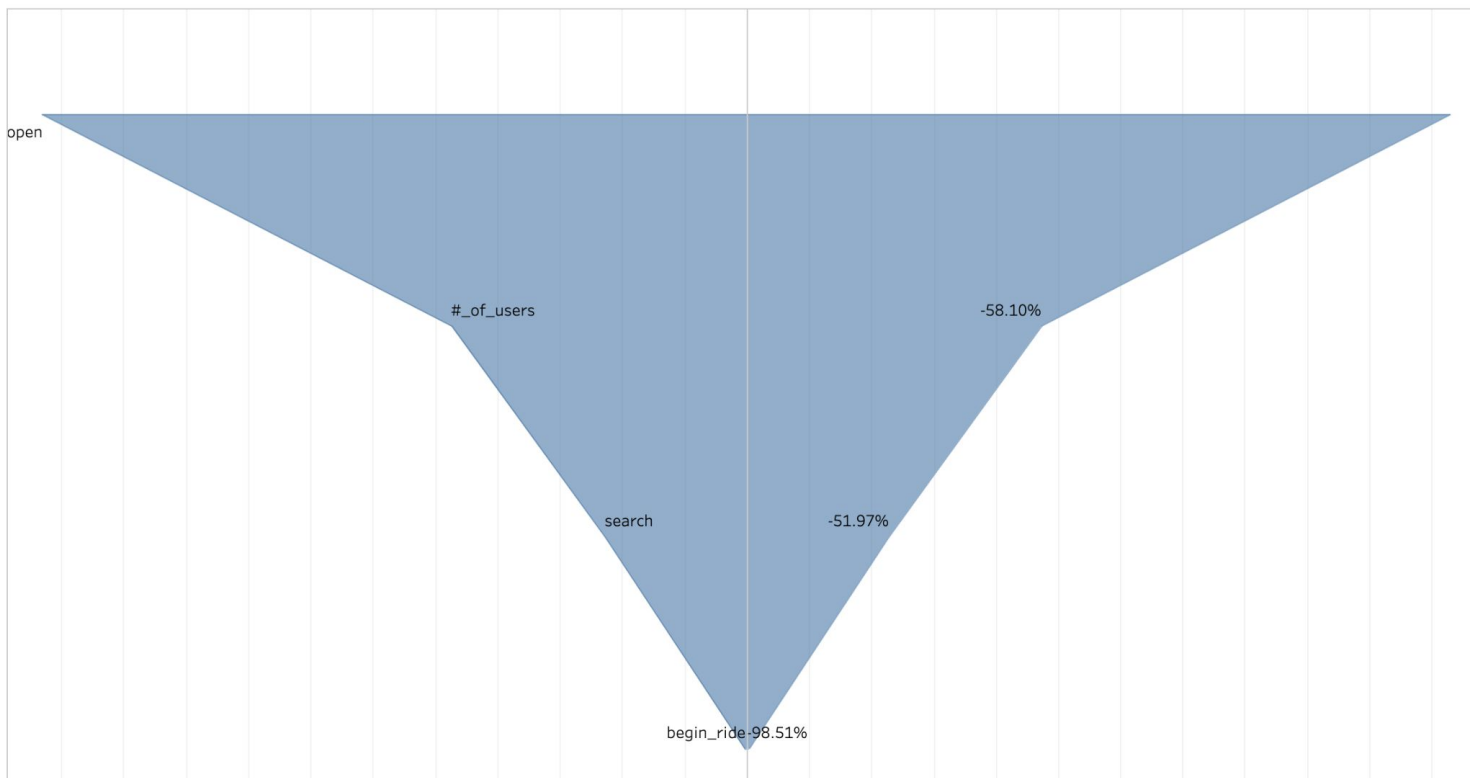
Customer flow in the App



User Funnel

Graph showing the drop-off rate in the user funnel

Funnel Drop Off Rate



Observation:

We notice a significant drop at search and begin_ride stage. Despite of healthy customer acquisition, Flying Uber is unable to retain its customers. We need to deep dive into activation and retention stages to understand the reasons for high drop-offs.

User Segments

User segments are groups of users that share a similar attributes. User segment analysis help identify the patterns in data by different types of customers. A user segment can be considered in three main categories:

- ❖ user acquisition date (cohort analysis),
- ❖ user demographics,
- ❖ user behavioral attributes.

The 2 demographic attributes In the presented data are:

- ❖ User Age range,
- ❖ User Neighborhood

User
Segments by
Age Group

Age	
18-29	28,321
30-39	20,124
40-49	41,774
50+	64,059

User Segments by
Neighborhood

User Neighborhood	
Manhattan	57,110
Brooklyn	16,435
Queens	4,050
Bronx	2,396
Staten Island	1,566

● Observations

- Maximum number of users are in the age group of 50+
- Users living in the Manhattan neighborhood shows maximum interest in the service

Now let's dive further into data to understand the drop-off among different user segments..

Segment Analysis of Funnel

Opportunities for Improvement

Drop Off Rate by Age Group

Event Type	Age			
	18-29	30-39	40-49	50+
open				
#_of_users	-53.59%	-55.27%	-50.40%	-42.15%
search	-33.88%	-35.18%	-32.69%	-62.13%
begin_ride	-98.55%	-98.29%	-98.16%	-98.61%

- The drop-off rate among 50+ age-group is lesser at number of riders and search stage as compared other cohorts. Also please note that the 50+ cohort has the largest number of users who have opened our service. Also, there is ~30% difference in drop-off rate at search stage for the other age-groups compared to 50+ age group.

Drop Off Rate by Neighborhood

Event Type	User Neighborhood				
	Manhattan	Brooklyn	Queens	Bronx	Staten Island
open					
#_of_users	-28.22%	-28.63%	-28.89%	-28.71%	-26.37%
search	-38.89%	-38.61%	-38.30%	-38.35%	-40.50%
begin_ride	-98.11%	-98.14%	-98.26%	-97.82%	-97.96%

- At the search stage, all neighborhoods are consistently underperforming at the same drop-off rate except Staten Island which is at ~41%.
- The drop-off rate at begin_ride stage is consistently high across all the demographics.
- A qualitative research is needed to find out the reasons for drop-off among different demographic groups.



Step 3

Hypothesis &
Suggested
Experimentation Plan

Review Qualitative Data

- Overview of the Qualitative interviews data
 - There are 11 interviews conducted across different user demographics:
 - 7 interviewees are in the age-group of 50+
 - On an average most of the customers use Flying Uber approx. 1+ times a month
 - Almost all customers ride Flying Uber to save travel time
- Reasons for funnel 's underperformance at stage 2 (#of riders) and hypothesis of customers' needs.

One of the theme common among all interviews is related to the non-user-friendly interface of the app. Sharing some of the comments from the customers below:

"Time is money and Flying Uber saves me time! But I let my assistant actually book the Flying Uber because the first few times I tried booking, the instructions were too small."

"Luckily my daughter was around to help me book the ride. I usually just use Uber because it remembers my addresses and has all my favorite places saved, so I guess I always just open that up since it is so convenient and saves me time."

"I call up our local pilot, Bob. He's not always available but I don't need to fiddle around with an app and hitting tiny buttons. He knows where I tend to be and where I want to go."

These comments clearly prove that Flying Uber app interface is not intuitive enough to search for and book a ride. As most of our customers are in the age group of 50+ for whom time and convenience are important needs, we should consider to provide the ability to:

1. Book the ride using voice commands,
2. Option to save the credit card info and frequently traveling locations

Suggested Features & Experimentation Plan

Hypothesis:

We believe, by providing the ability to:

1. Book the ride using voice commands,
2. Option to save the credit card info and frequently traveling locations in the app,

we will be able to improve the usage of ride service among 50+ age-group customers and reduce the drop off rate by ~42% among the Flying Uber customers.

Customer Pain Point / Hypothesis	Possible Solution / Suggested Feature
High drop-off at the number of riders and search stage in the app due to long instructions or small buttons	Add voice command button
unable to save frequently used information	Option to save credit card and frequent travel locations

Multivariate testing framework:

Control Group	Current View
Experiment 1 Group	Voice button on the open page
Experiment 2 Group	Option to save credit card and travel location at the search page
Experiment 3 Group	Voice button on the open page and save CC and travel location option at the search page

Suggested Features & Experimentation Plan

Target Audience:

Random set of existing customers in the age-group of 50+ using android or ios version of the Flying Uber app. Sample size can be selected using [optimizely calculator](#).

Data Instrumentation for tracking the Experiments:

- Customer information: user_id, age-group, name, income group
- Session characteristics: Session Id/event id, date & timestamp, app version, number of riders selected
- Which experiment the customer is exposed to
- Track if the customer clicked on the feature
- Track if customer started the ride.

List of Metrics to track the progress of the features:

- Feature click-through rate: unique count of sessions where feature is clicked
- % customers using the feature: unique count of customers who clicked the feature/total unique customers exposed to the feature
- Conversion rate by the feature: unique users started the ride by using the feature



Step 4

Global Launch Plan

Global Launch Product Requirement

- Flying Uber Launch: PRD
- Policy: High-Level Requirements and Enforcement Guidance
- Compliance New Feature Assessment (NFA)
- JIRA Epic: FU-2738

Status (<i>definitions</i>)	Draft
Creation Date	TBD
Launch Date	TBD
Hard Deadline	TBD

Team <i>(all teams may not be required for launch, but at least one in each area)</i>		Point(s) of Contact <i>(ideally only one per team)</i>	Sign-off Date <i>(work will not begin until PRD is signed off)</i>
Engineering	Dev Tech Lead	TBD	
	3rd Party Tooling	TBD	
	Tools Foundation	TBD	
	Data Science	TBD	
	QA	TBD	
Product Management	PM (Platform)	TBD	
	PM (Health)	TBD	
	PM (Feature)	TBD	
	Legal Counsel (Compliance)	TBD	
Marketing	PMM	TBD	
	Comms	TBD	

Global Launch Product Requirement

Team (all teams may not be required for launch, but at least one in each area)		Point(s) of Contact (ideally only one per team)	Sign-off Date (work will not begin until PRD is signed off)
Customer Success / Operations	Operations Manager	TBD	
	Policy Manager	TBD	
	Vendors Manager	TBD	
	Training & QA Manager	TBD	
	Analytics Manager	TBD	
Logistics	Logistics Manager	TBD	
	Sustainability Manager	TBD	
	Supply Chain SME	TBD	
Finance/Accounting	Managerial Accountant	TBD	
	Financial Accountant	TBD	
	Tax Accountant	TBD	

Global Launch Product Template

- ❑ **Feature Description and Outline**
 - Overview
 - Motivation
 - Success Metrics
- ❑ **Scope and Launch Project Plan with Gantt Charts**
 - Summary of Requirements
 - P0s (blockers for Global Launch)
 - P1s (Fast follows after Global launch)
- ❑ **Detailed Requirements**
 - In-product report option
 - High Level Overall Flow
 - Reporting Flow (In-Product and webforms)
 - Triaging
 - Case Routing for in-product reports
 - Case Management Tool Views and Fields
 - Policy Application
 - Side Effects (Risk Plan)
 - Change Plan
- ❑ **Platform backend**
- ❑ **Security Model**
 - Legal and Compliance
 - Access Regulations
 - Comms Plan
- ❑ **Internal Tooling**
 - Automation Plan
 - AI/ML infra
 - User Communication
 - In-App Push Notifications
 - Help Center
- ❑ **Analytics**
 - Instrumentation Plan
 - Adoption Metrics Dashboard
 - Success Metrics Dashboard
 - Guardrail Metrics Dashboard
- ❑ **Appendix**
 - Open Ended Questions
 - Meeting Notes
 - Other Resources