Data Product Manager Nanodegree

Applying Data Science to Product Management

Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service

Welcome to your first week at Flyber


Aybar

In this project, you will apply the skills acquired in this course to create the MVP launch strategy for the first flying car taxi service, Flyber, in one of the most congested cities in America -- New York City.

You are responsible for bringing the first flying car taxi service to market by analyzing data and building a product proposal.

You will need to use the SQL workspace provided in the Classroom, and Tableau Public, in order to successfully complete the project.

You'll present your answers, findings, and insights in the Answer Slides found in this deck. Feel free to include any additional slides, if needed.

Section 1: Data Exploration

Back to the basics of product management, identify your customer and their pain points:

- What are taxis used for?
- What are the characteristics of the users that leverage them?
- What are existing pain points with taxis?
- What are the existing pain points with digital ride-sharing services?

Answer Slide

The users who utilize taxi services are those who may not be having a personal vehicle available to ride in or may choose not to use it even if it's available, because of the issues like finding a place to park by themselves, maintenance and lack of need.
Given that Taxis are being used mostly for evening rides, and for relatively short rides, users doesn't seem to need a car aside of leisure activities. And since they belong to a middle to higher class, they choose to use taxis instead by choice and not because they don't have other options.
The main pain points with taxis are the waiting time, the elevated cost and the traffic jams, especially in Manhattan and during peak hours.
On the other hand, the existing pain points with digital ride-sharing services are also the waiting time and the traffic jams, but on the contrary, the cost can be an advantage since it can be considerably lower but the waiting times and detours for dropping off another person are a disadvantage.

What user improvements do you hypothesize a flying taxi service would have over the existing state of taxis today?

What market improvements do you hypothesize a flying taxi service would have the existing taxi service industry & physical road infrastructure today?

Answer Slide

A service like the one that Flyber proposes, would be a refreshing way of travelling for most people in Manhattan, since there won't be no traffic jams in the air. And for people moving outside of Manhattan, will make the distances shorter with their increase on speed.

It will not only improve quality of service for the passengers that chose Flyber. It will also improve the existing taxi service industry & physical road infrastructure by freeing the traffic jams that are so common in the center of Manhattan.

Upload this dataset into Tableau Online.

Ensure the fields are parsed correctly; field headers are included in the first row of the CSV.

Let’s begin exploration!

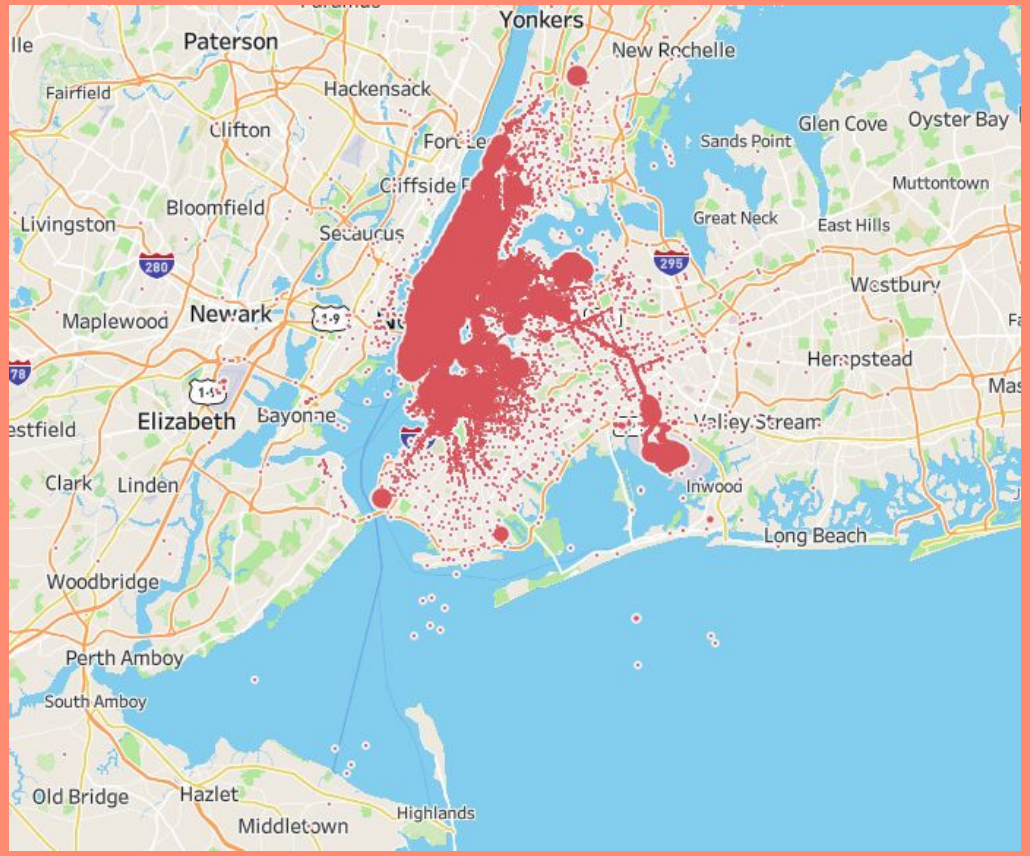
Acquire a high-level understanding of the granularity and scope of the dataset, to inform the basis for your analyses:

- How many records are in the dataset
- What does each record represent?
- What is the primary key?
- What date range is your dataset bound to?
- What are the geographical bounds of this dataset? Is it limited to Manhattan, or is Brooklyn, Queens, Staten Island, the Bronx, and New Jersey included? Where are most of the data points centralized at? Are there outliers?

Answer Slide

The dataset has **10.48.468** records, where each record (granularity) represents a taxi trip.
Each record is uniquely identified by **id** (primary key).
The dataset includes data within the date range from **01/01/2016 12:00 AM to 30/06/2016 11:59 PM.**

Answer Slide



The geographical bounds of this dataset are mainly Manhattan, but also included Brooklyn, Queens, Staten Island, the Bronx, and New Jersey. Most of the data points are centralized in Manhattan and Brooklyn.
There are some outliers, since many points are located in the ocean.

You notice that the dataset does not contain explicit data points out-of the-box, we'll need to enrich the dataset with relevant fields:

- You notice that ride price is not included, but figure it could be derived. Based on information about New York taxi prices gleaned from the internet, create a calculated field called `price` using the `duration`, `distance`, and `passenger count` fields.
- You hypothesize your target users will be those who take a relatively longer time getting to a destination that is relatively close, due to heavy traffic conditions and/or limitations to physical road infrastructure. To be able to analyze where this is happening, you will need to create a calculated field called `distance-to-duration ratio`.

Let’s understand the scope and distribution various dimensions within the dataset. Calculate the **average**, **median**, and the **first & second standard deviation of the mean** for the following measures:

- duration
- distance
- passenger counts
- duration-to-distance ratio
- price

Answer Slide

Median measures

Median Duration(Min)	11.10
Median Distance	2.11
Median Passenger Count	1.00
Median duration-to-distance ratio	4.68
Median Taxi_fare	13.61

1º Std. dev. measures

Std. dev. of Duration(Min)	53.07
Std. dev. of Distance	4.39
Std. dev. of Passenger Count	1.32
Std. dev. of duration-to-distance r..	76.65
Std. dev. of Taxi_fare	30.19

Average measures

Avg. Duration(Min)	15.97
Avg. Distance	3.46
Avg. Passenger Count	1.67
Avg. duration-to-distance ratio	7.66
Avg. Taxi_fare	19.18

2º Std. dev. measures

Percentile (95) of Duration(Min)	106.1
Percentile (95) of Distance	8.8
Percentile (95) of Passenger Count	2.63
Percentile (95) of duration-to-distance ratio	153.3
Percentile (95) of Taxi_fare	60.4

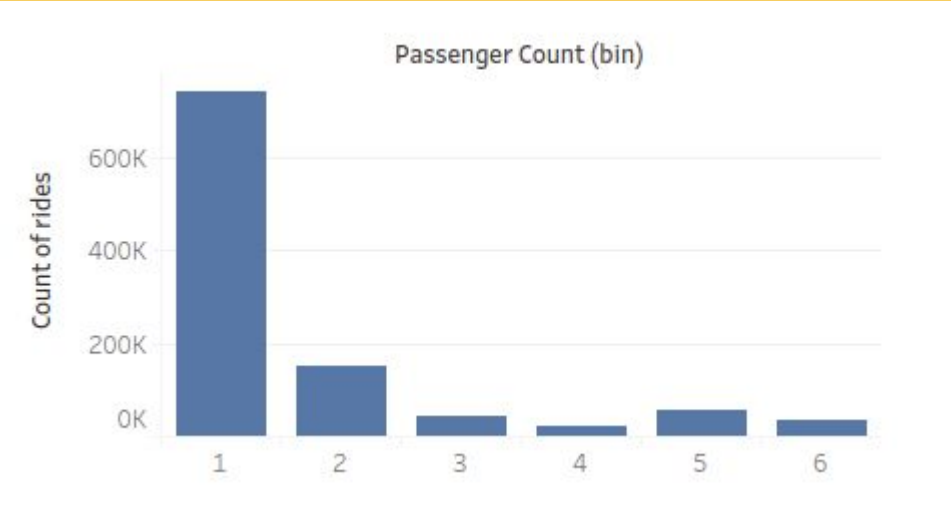
Taxi_fare = (2.5 +(1.56[Distance]*1.61)+([Duration (sec)]/3600)*30)

* duration-to-distance ratio = Duration (Min) / Distance

Flying cars may have to have to be a lower weight for efficiency & take-off. Or you may just decide to leverage mini-copters for your initial MVP.

Create a histogram that visualizes the number of total rides grouped by passenger counts to analyze the potential market volume of low passenger pickups (1-2 passengers).

Answer Slide



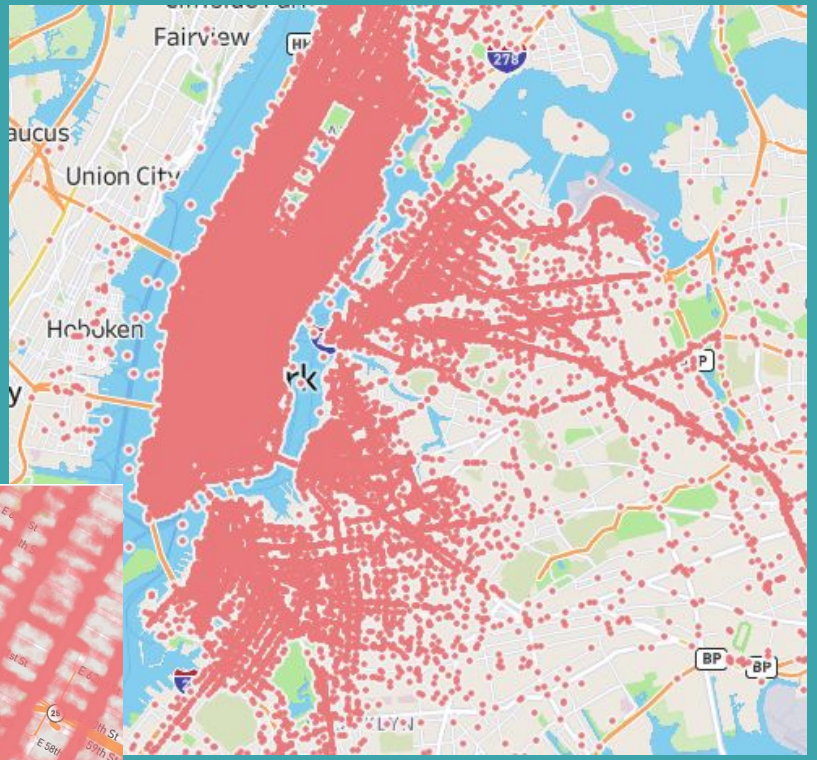
We can see in the histogram that the mayor amount of rides is done by 1 passenger only, followed by 2. Taking this in consideration, we can state that the potential market volume of low passenger pickups (1-2 passengers) represents 84% of the total market.

For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?
- Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off?
- For any of the neighborhoods identified, are there any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off? What makes them suitable?

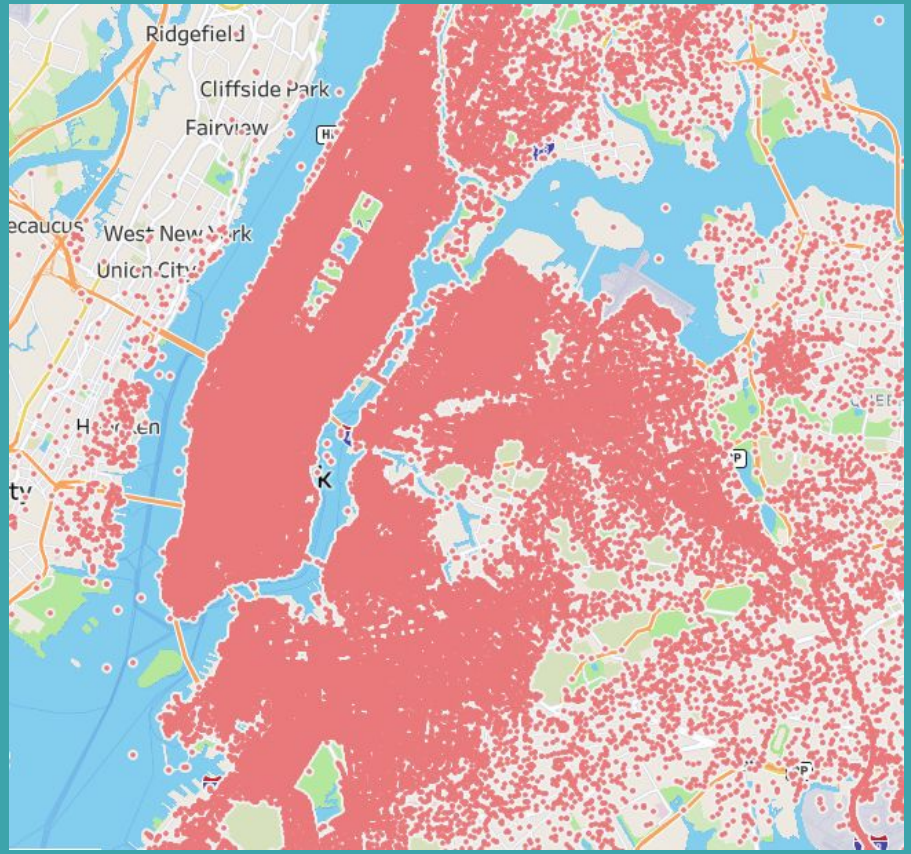
Answer Slide

The neighborhoods with the higher density of rides based on pickup location are all around Manhattan. In particular, in the area around Midtown.



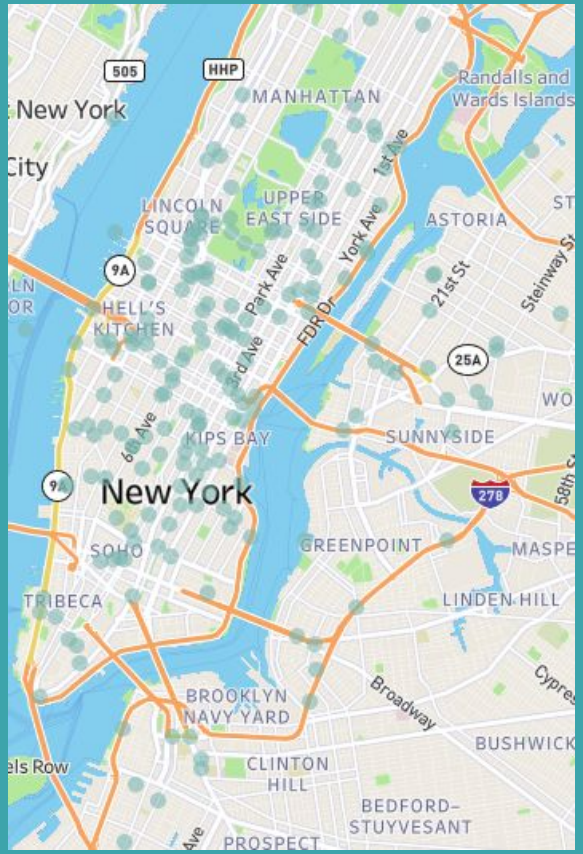
Answer Slide

On the other hand, if we evaluate the drop off instead, the neighborhoods with the higher density of rides are still all around Manhattan, but include as well Brooklyn.



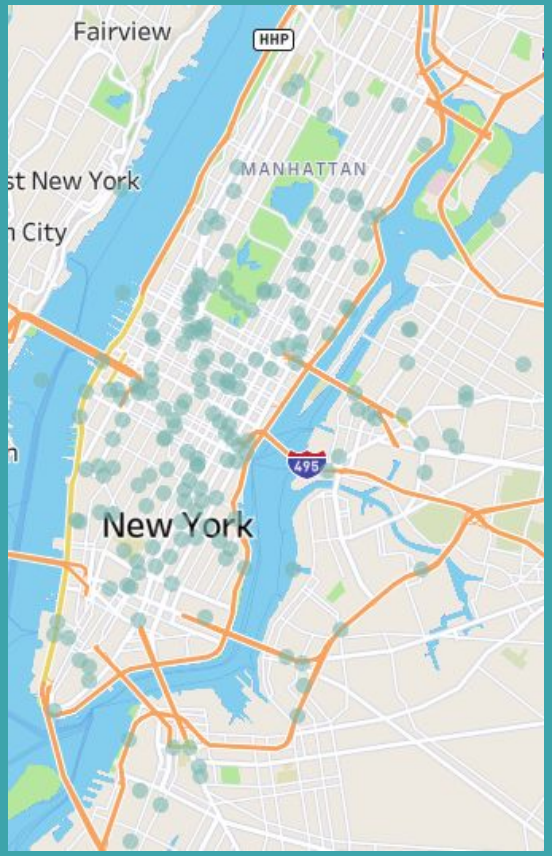
Answer Slide

The neighborhoods with the higher duration to distance ratio based on pickup location are around Rose Hill, Hell's Kitchen, Colombus CIrcle and Midtown.



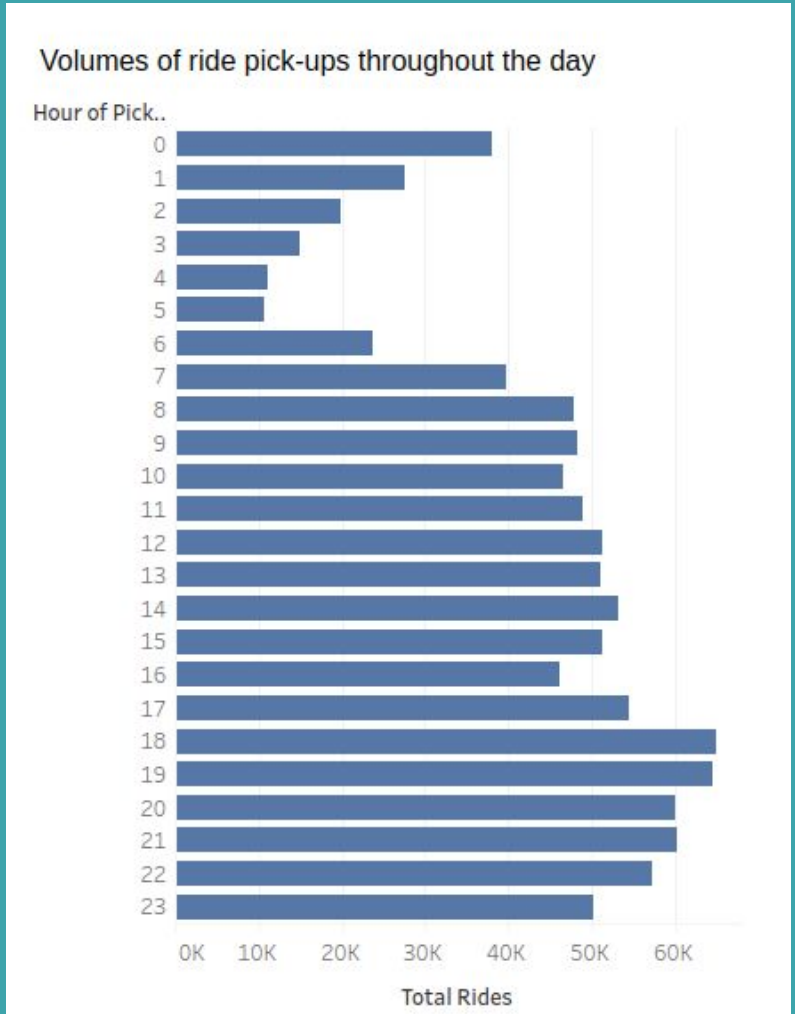
Answer Slide

The neighborhoods with the higher duration to distance ratio based on drop off location are around Colombus CIrcle, Lincoln Square, Hell's Kitchen and Midtown.
Based on all of these findings, any area in between Lincoln Square and Rosie Hill seem to be a good target, because it's where most rides are being taken, but also where most of the traffic is.
Also, it would be a good idea to focus on longer trips, from these same areas to far away points as Long Island.



It may not make operational sense to have the service running 24/7, for now.

- What times throughout the day experience relatively higher volumes of ride pick-ups?
- What days throughout the week experience relatively higher volumes of ride pick-ups?
- Pinpoint any periods throughout the year that experience trend fluctuation or seasonality around ride pick-up volumes. This will help us in our post-launch analyses to determine if any spikes or dips were influenced by seasonality or through actual feature adoption/regression.



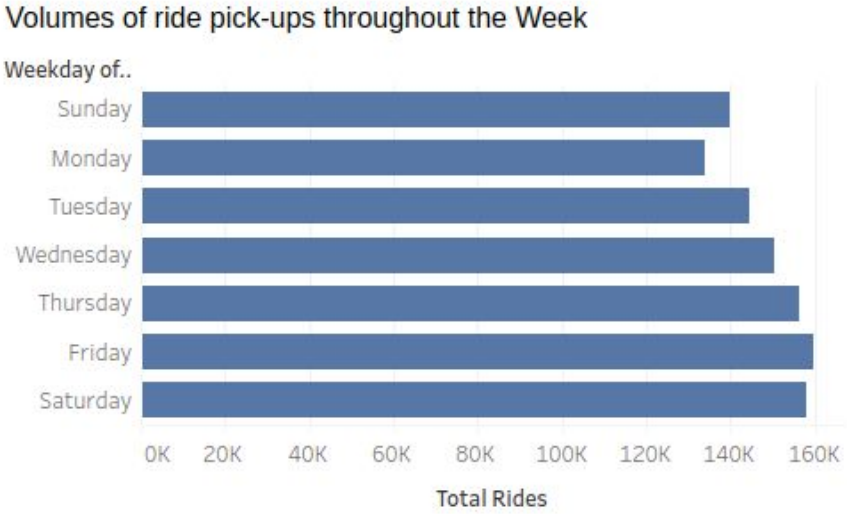
We can see that most pick ups take place between 18 and 22hs, being the peak at 18hs.

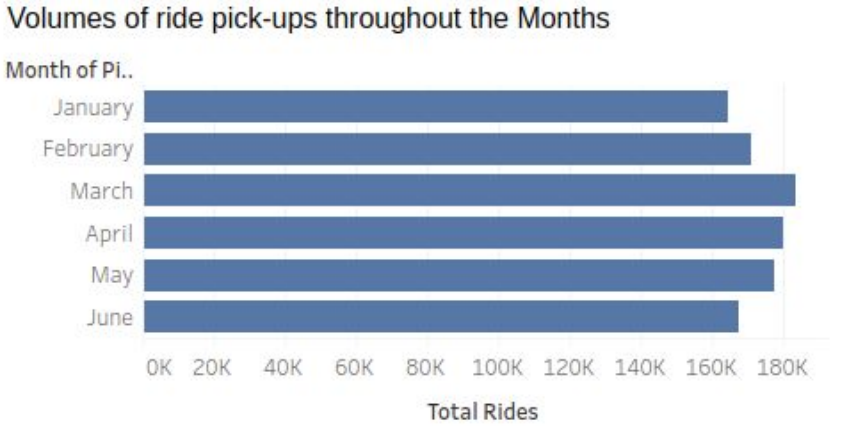
On the other side, the lowest volume can be seen during the early morning between 3 and 5 AM.

Answer Slide

During the week we can see that there's a tendency throughout the week to increase during Thursdays, Fridays and Saturdays. Being Friday the peak of the week and Monday the lowest point.

Regarding seasonality, there seems to be a tendency to lower rides during warmer months, having June the lowest volume of rides. But it's not possible to prove without a full year of data.





You and the user research team ran a quantitative survey on existing taxi and/or rideshare users in New York City to determine sentiment around potentially using a flying taxi service.

Dive into the survey results dataset in order to extract insights from explicit feedback.

Upload this dataset into Tableau Online or a SQL database (the classroom contains a workspace with the data for you as well).

Ensure the fields are parsed correctly, field headers are included in the first row of the CSV.

Question schema:

Q1 - What is your email?
Q2 - What gender do you identify as?
Q3 - What is your age?
Q4 - What is your annual income? (income bands)
Q5 - What neighborhood do you reside in?
Q6 - Do you currently use taxis? (Y/N)
Q7 - Do you currently use ridesharing services? (Y/N)
Q8 - Would you use a flying taxi service, if such a concept existed? (Y/N)
Q9 - If yes to Q8, how much would you be willing to pay per mile for such a service? (USD)
Q10 - If no to Q8, what is the reason?

To inform our future product marketing efforts, we'll want to extract the following:

- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?
- What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?
- What is the different personas/segments of negative sentiment towards not using a flying taxi car service?

Answer Slide

There’s an inclination on the female users that have between 27 and 68 years old.
Users have medium to high salaries, being the annual income band most common between $40,001 and $80,00 and live mostly in Midtown, Battery Park City or the Financial District.
In general, the price that the users are willing to pay lays around 17 and 25 USD per mile. Women are willing to pay more than men, whose most common response was 23 USD per mile, followed by 17.

Among the users that are not interested in Flyber’s services we can find a wide range of ages, being 20 and 33 the most common ones. It can be reduced to one specific age range.

Hooray! End of Section 1.

You will complete Section 2 at the end of this course.

Please submit this file for review for Section 1.

Section 2: Proposal Synthesis

Identify a product objective for Flyber's launch. Your product objective will guide your KPIs, so identify what Flyber should optimize for. Your objective should be centered around one the following focus areas:

- User Acquisition
- User Engagement
- User Retention
- Profitability

Explain your reasoning. Include both why you feel your focus area is more relevant than the others for Flyber at this time of the product development cycle.

Answer Slide

I'll focus my project on User Acquisition, because as a new service with a groundbreaking product it's the area that will enable us to get on the hyper growth path. We need first to acquire users before focus on engagement and retention. And since the potential market is very big in a city like New york, we could focus just in acquisition during our initial phase.

Of course profitability is an important focus, but given that the product is new and has never been marketed before, we need first to invest in convincing people to use our service before we can achieve profitability.

Formulate 3-5 Key Performance Indicators (KPIs), to measure if the product is heading towards the right direction based on your objective

Answer Slide

1. Weekly customer acquisition (`(New customers added in a week / Total customers in a week) x 100`)
2. Daily rides
3. Weekly Active Users

Create hypotheses around what thresholds your KPIs would need to hit in order to determine success

Answer Slide

1. 40% Weekly customer acquisition (`(New customers added in a week / Total customers in a week) x 100`)
2. 1.000 daily rides
3. 5.000 Weekly Active Users in the first 6 months

As the product manager, you make decisions based on the insights you extract, we'll need to know the feature set we'll include in the MVP to measure viability, while keeping operational expenditure under control:

- What times/days of operation should the service run for?
- How many pick-up / drop-off nodes should we have?
- Where should the nodes be located?
- Should we initially use copters or homegrown hardware?
- Should the pricing be fixed or dynamic? At what rates?

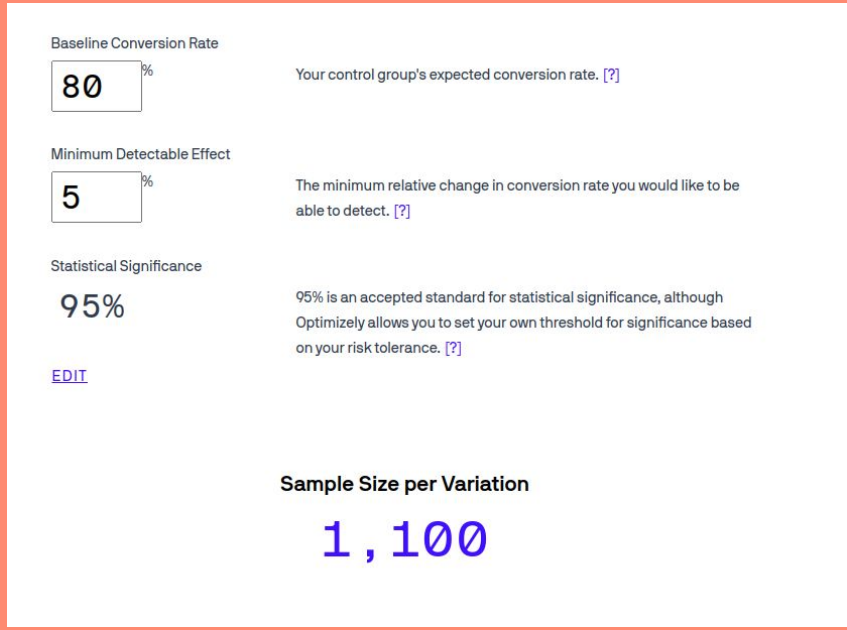
Answer Slide

- The service will run daily from 5pm to 1 am
- We will start with 5 pick-up / drop-off nodes in the areas where taxis are used the most, and one more far away to evaluate the performance of longer rides.
- The nodes will be located in Rose Hill, Hell's Kitchen, Colombus Circle, Midtown and Brooklyn.
- Initially, we will use copters
- We will use fixed pricing, to facilitate understanding and price clarity. Rates will be enough to cover expenses only.

Determine the MVP sample size & time period allotted estimated to come to a conclusion on your hypotheses.

Answer Slide

Since we're targeting the areas and clients that we know already use taxis and will be willing to try our service, we expect a high conversion rate.
Nevertheless, we expect to detect small changes from the baseline.
Hence, the Sample Size per Variation is 1,100
The experiment will run for 1 month



Create an instrumentation plan for the events you need collected and logged, in order to be able to physically measure your KPIs.

Answer Slide

Event - **rideCompleted**
Definition - Triggered whenever the driver hits the “Ride is finished” button on the Flyber driver app.
Properties - ride_id, ride_begin_timestamp, ride_end_timestamp, duration, cost, flyber_id

Event - **signupCompleted**
Definition - Triggered whenever the user completes a signup on the Flyber client app. A signup is considered completed when all of the basic user information has been registered.
Properties - signup_completed_timestamp, flyber_id

Answer Slide

We can measure **weekly active users** with the *ride_begin_timestamp* and the distinct *flyber_id*.
We can measure **newly weekly acquired users** with the *signup_completed_timestamp* and the count of distinct *flyber_id*.
We can measure daily rides with the *ride_begin_timestamp* and the distinct *ride_id*.

Create a qualitative feedback survey questions for users after their ride, to further understand and optimize the product for future iterations.

Answer Slide

- How was your ride? Do you have anything to comment or report?
- How likely is that you would ride again with us from 0 to 5?
- How would you rate your ride from 0 to 5?
- How would you rate your driver from 0 to 5?
- How would you rate the relation quality-rate from 0 to 5?
- What would you like us to improve?

Summarize everything you have learned into your final proposal

- Identify the target population. Why did you select that target population? What are their pain points?
- Create a product proposal containing claim, evidence, estimated impact, and risks
- Claims should be backed by quantitative evidence, impact should assess market needs/benefits
- Risks involve any known unknowns that we'll still need to monitor post-launch
- State cross-functional stakeholder teams that will need to be involved

Answer Slide

Our target population are the current taxi user in the center of Manhattan. They enjoy and can afford using a taxi service, but experience many pain points with the current traffic situation in New york: delays and traffic jams.
This transport concerns in New York City only keep growing and the situation requires innovative solutions to address this problem. Our flying taxi service is not only innovative but also addresses many of the concerns like road accidents, time delay, novelty of experience, less delays and unpredictability in public transport.
Flyber not only represents an immediate solution to these problems, but also, as it expands, the number of taxis on the road will reduce, consequently reducing traffic. Thus the time people usually spent in road traffics would be reduced.

Answer Slide

In order to provide a safe and secure service we will need to include Legal and Compliance team to review the jurisdiction laws and insurance, Operations team to supervise the areas we operate in and the service we provide, Software Engineers for app development, third party copter vendors that will rent the vehicles, Data team to store an analyze the users and rides data, and drivers.