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?

- Taxis are used for public transportation. People can use them to get to the areas they want to go, without any stops in between and less limitations on where they want to get on and get off compared to other public transportations.

- Taxi usage rates are higher among to customer who have 40.000 dollars and more income annually. Also females are more likely to use them according to our user research data.

- Using taxi services increases the traffic density. Both drivers and customers can and up spending hours on the road, especially during rush hours. Also taxi availability can be an issue especially in crowded cities.

- Same issues apply with ride-sharing sharing services. In addition to that, most of the drivers are not professionals, which means this service may not be as safe.

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

- With flying taxi service, customers will not spend as much time on the road because they won't have to deal with traffic issues. Since there are more options with this service, they will also be less likely to not find an available vehicle.

- Since there will be less individual cars on the roads, it will improve traffic conditions for other services and drivers.

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

- There are 1.048.468 records on dataset.
- Each record present an individual taxi ride in New York city.
- Id is the primary key of the dataset.
- Dataset contains rides from 01.01.2016 to 01.07.2016.

- *(I wasn't able to use tableau map future so I had to minimize the data set and get the records with pickup date between 6/15/2016, 12:00:00 AM and 6/30/2016, 11:59:00 PM. Also removed the records with distance 0.)*

- While the pick up data centralized at Manhattan, especially in Midtown Manhattan, there are still records from other parts of New York, such as Long Island and Queens. We also see a density on airport regions.

- There are also some outliers in the dataset. I have eliminated the records with more than 24 hours and less than 60 seconds duration and with 0 distance.

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`.

- With the reference of this website, I have created this formula and I checked it with some of the records. $2.5 is the base price for all taxi rides.
  Formula:
  Price = (2.5 +1.2*(distance)+0.4*(duration/60))*Passenger Count

- I will calculate the duration to distance ratio as below:
  distance to duration = distance / (duration/60)

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
- distance-to-duration ratio
- price

Duration

- Average: 958.4
- Median: 666.0
- First standard deviation of the mean: 3185
- Second standard deviation of the mean: 6370

Distance

- Average: 3454
- Median: 2102
- First standard deviation of the mean: 4385
- Second standard deviation of the mean: 8770

Passenger Count

- Average: 1665
- Median: 1000
- First standard deviation of the mean: 1315
- Second standard deviation of the mean: 2630

Price

- Average: 22.20
- Median: 12.35
- First standard deviation of the mean: 61.14
- Second standard deviation of the mean: 122.28

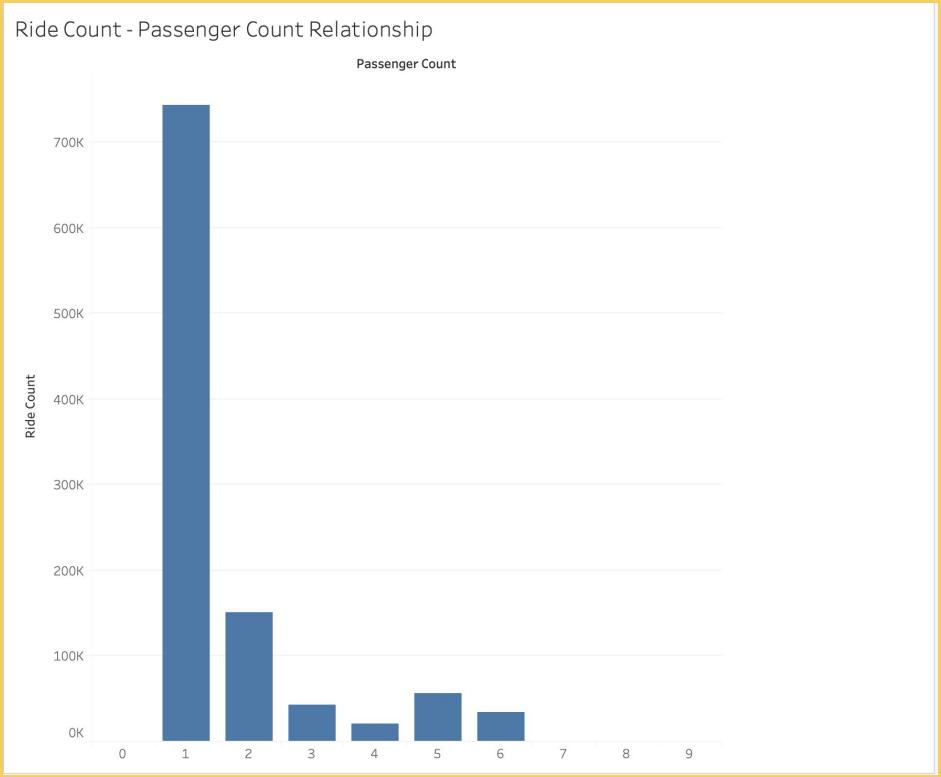
Distance to Duration Ratio

- Average: 6181
- Median: 4659
- First standard deviation of the mean: 35.34
- Second standard deviation of the mean: 70.68

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).

Ride and Passenger Count Relationship



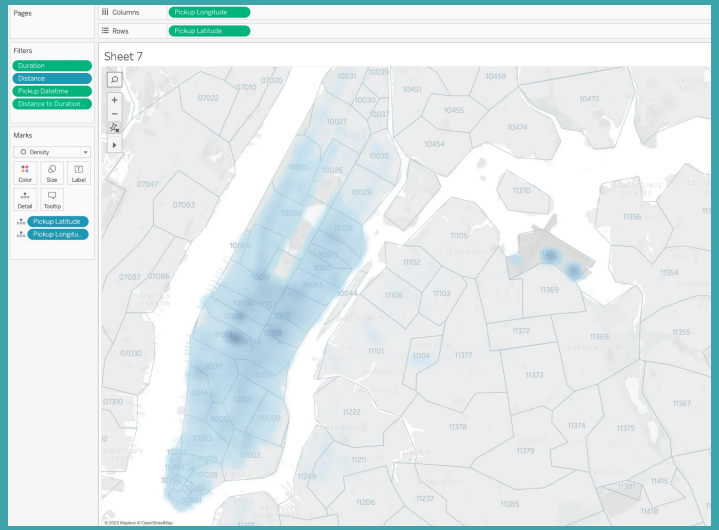
On traditional taxi rides majority of the time, there are 1 or 2 passengers per ride. So lower passenger pick-up shouldn't be a big problem for the MVP phase.

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:

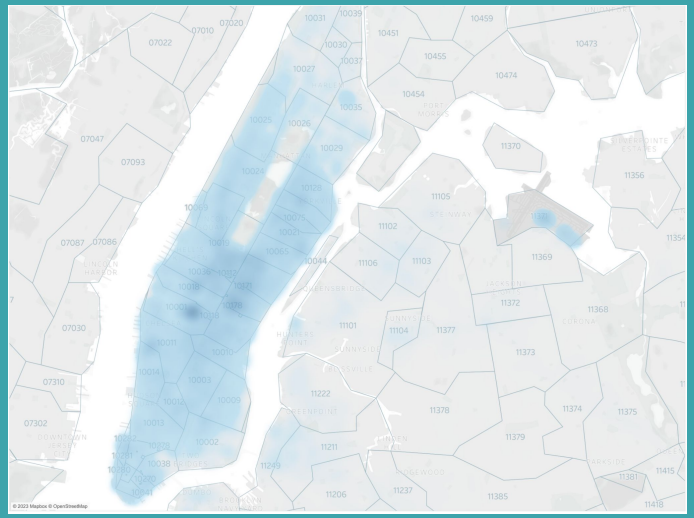
- Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?
- Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?
- Which neighborhoods/zip 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?

(I wasn't able to use tableau map future so I had to minimize the data set and get the records with pickup date between 6/15/2016, 12:00:00 AM and 6/30/2016, 11:59:00 PM. Also removed the records with distance 0.)

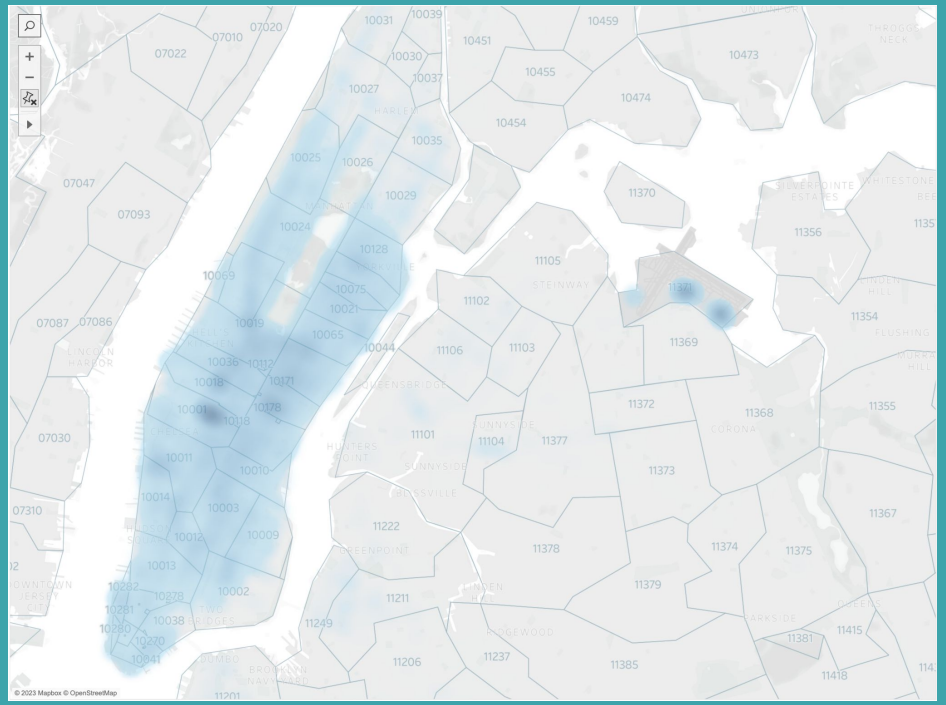
- Zip codes 10001, 10123, 10018, 11371 has more density on pickup rates. Midtown Manhattan area and La Guardia Airport region.



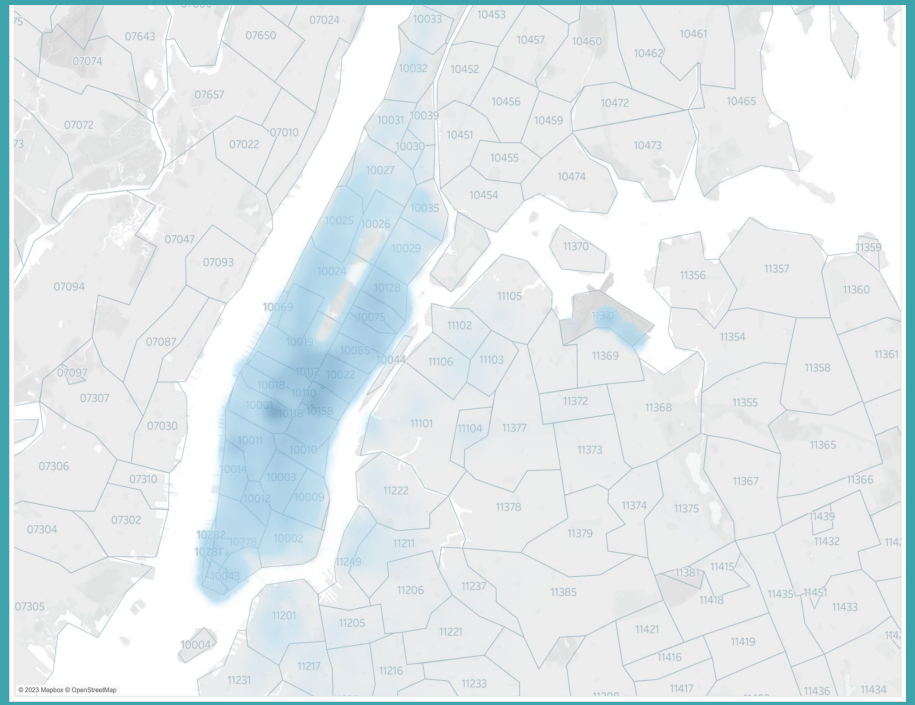
- Zip codes 10001, 10112, 10010 has more density on drop off rates. Midtown Manhattan area and La Guardia Airport region.



- 10001 and 10018 has the most distance to duration ratio based on pickups. 11371 is also visible since it is the airport region.



- 10001 and 10178 has the most distance to duration ratio based on drop offs. 11371 is also visible since it is the airport region.



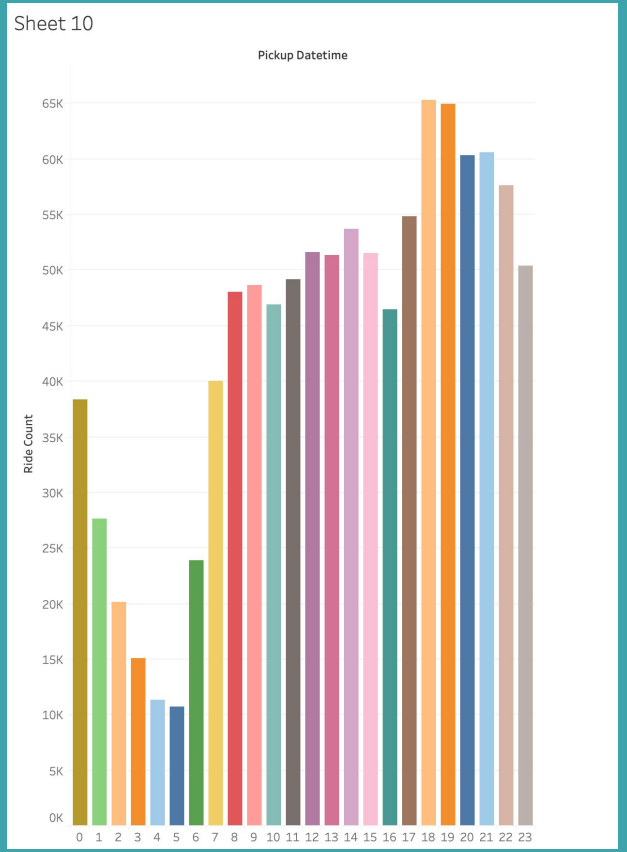
- Neighborhoods 10001 seems to be the most popular location for both drop offs and pickups. Most of the traffic density is located on Midtown Manhattan area. So it would make sense to choose this area as pilot region, since both taxi demand and duration-distance ratio is the highest.

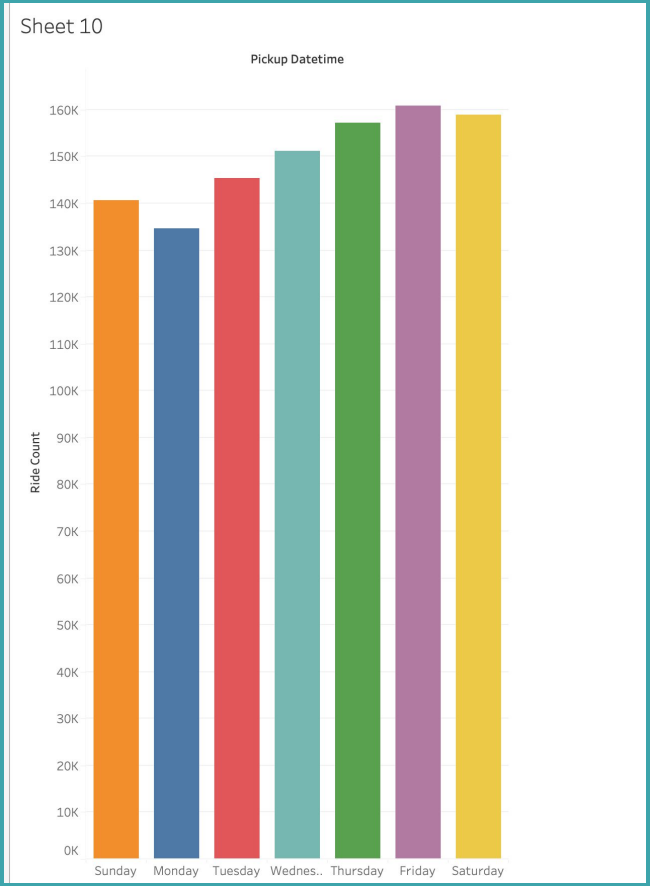
It may not make operational sense to have the service running 24/7, for now.

- 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.

Answer Slide

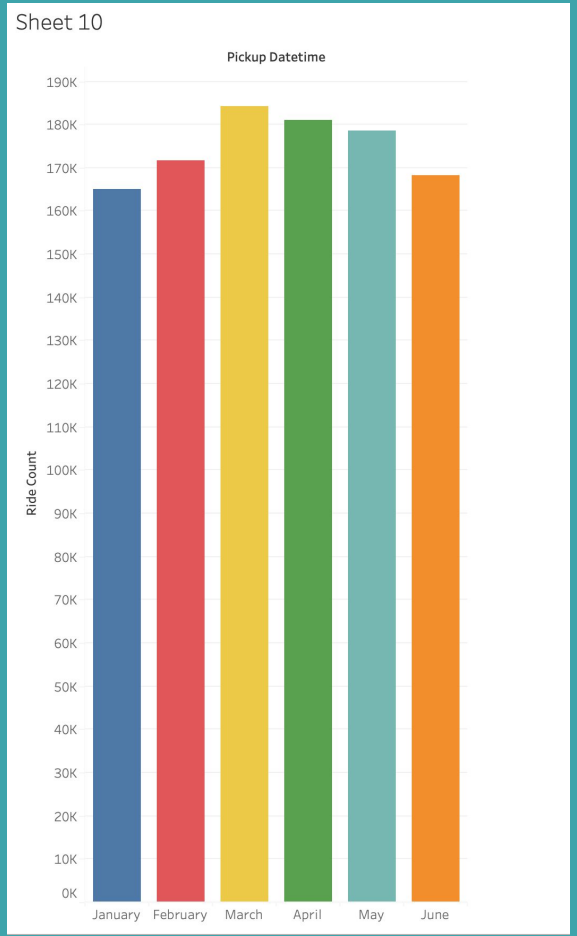
- 18.00 and 19.00 has the most volumes of pickups. Period between 18.00 and 00.00 has the most taxi ride count compared other time of the day.





- Fridays and Saturdays has the most amount of pickups compared to rest of the week

- Provided data restricted to 6 month period, so it is not very possible to make an all year analyses. Spring months seems to be the months where we can see an increase on total rides. Precipitation rates might be an affect since they are higher in winter months.



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:

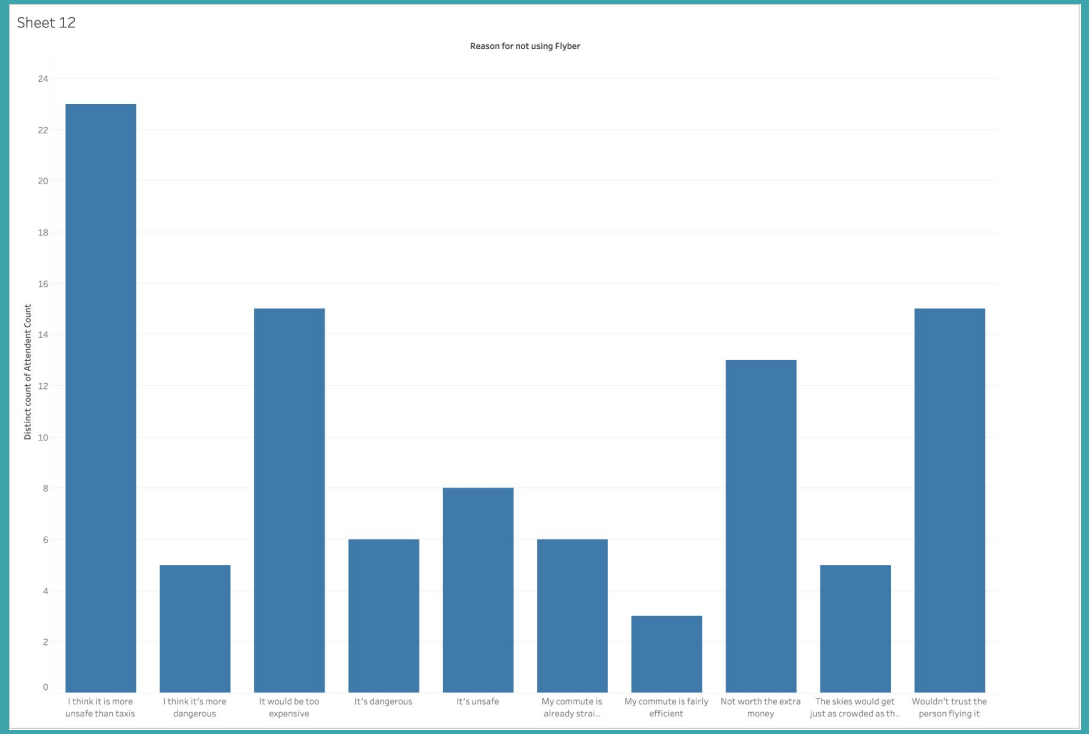
- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?
- What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?
- What is the different personas/segments of negative sentiment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

Answer Slide

- Taxi usage rates are higher among to customer who have 40.000 dollars and more income annually.
  Females are more likely to use them.
  Midtown East area has the most taxi usage rate.

- People reside in Harlem area are the ones that are willing to pay the most by average of 33.5$ per mile.
  Males and females price range are similar. Males are slightly willing to pay more.
  People with 200.000$ and more annual income are willing to pay the highest with 37.35$ per mile. Average price decreases when the income drops.

In general, there are worries about it’s safety, price. Also being pleased with the existing system is one of the reasons.



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

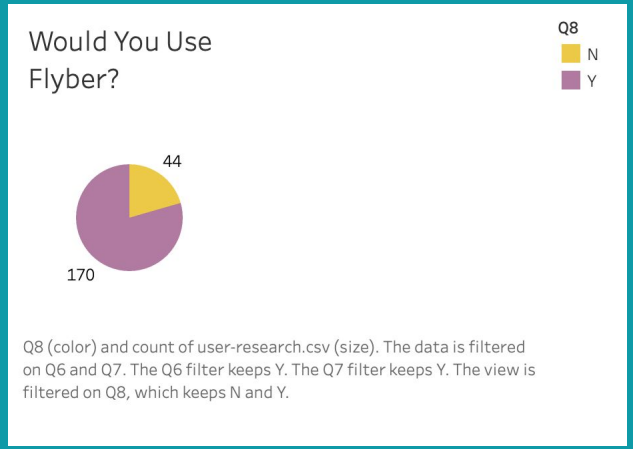
Identify a product objective for Flyber's launch. Your product objective will guide your KPIs, so identify what Flyber should optimize for. Your objective should be centered around one the following focus areas:

- 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

User acquisition will be our main objective on this MVP.
Nearly 25% of the users who uses taxi and ride sharing services are not willing to use Flyber.
Since this is a unique and new system, we want users to accept Flyber as a practical and usable transportation way. So this objective should be our main priority among others.



When we check the detailed answers on why people wouldn't use Flyber, we have two main issues so these can be our viable objectives.

1) People don't think it is as safe as the existing taxi and ride sharing systems. They don't trust the person flying it.

2) They think it will be much more expensive and wouldn't worth extra money.

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 new user count of Flyber. (Weekly New Active Users - WNAU)

2) Weekly flight count of Flyber. (Weekly Flight Count - WFC)

3) How much of the users would recommend Flyber to their friends and family? (Average Recommendation Rate - ARR)

4) What percent of the users converted from taxi or ride sharing services to Flyber? (Conversion Rate - CR)

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

Answer Slide

1) We should have around 50 new subscribers each week.

2)There are around 40.000 weekly taxi rides on the taxi_rides dataset. Considering that, there should be about 400 flyber rides weekly, since this is an MVP.

3)50% of the weekly users should recommend it to others so the customer community can grow.

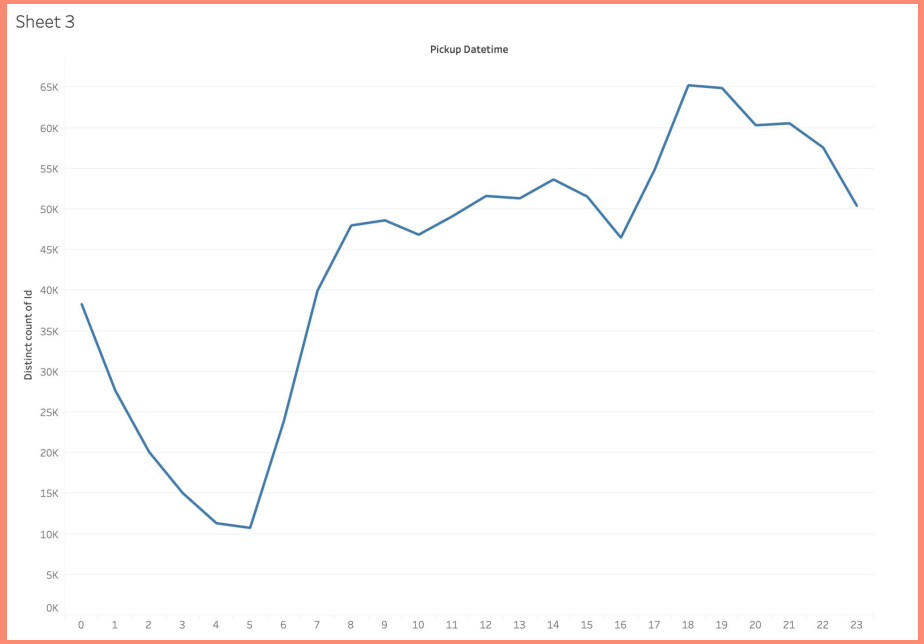
4) 10% of the users should convert from regular taxi service.

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:

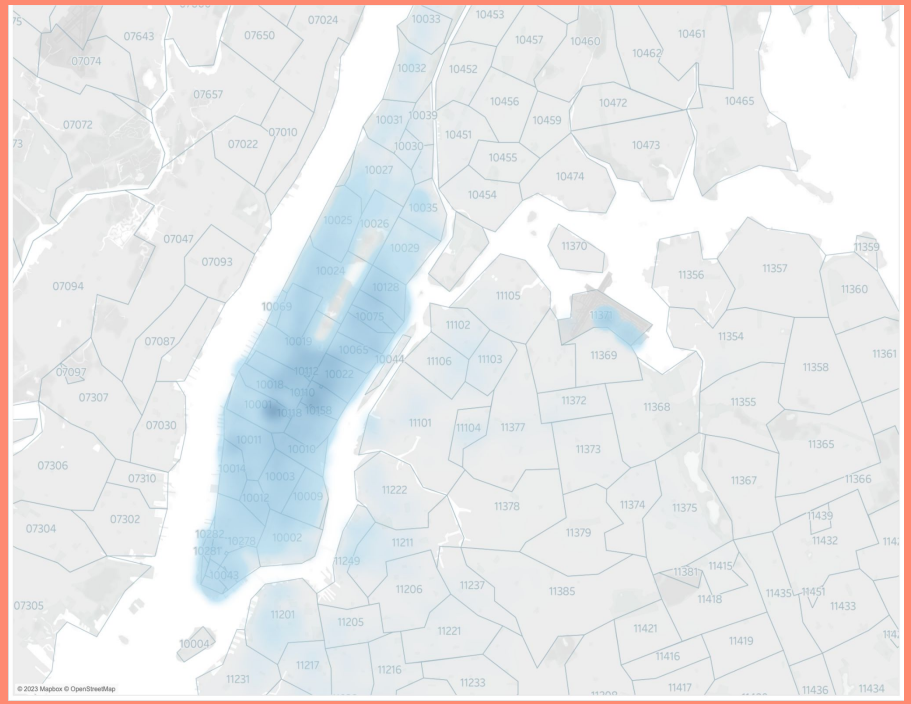
- What times/days of operation should the service run for?
- How many pick-up / drop-off nodes should we have?
- Where should the nodes be located?
- Should we initially use copters or homegrown hardware?
- Should the pricing be fixed or dynamic? At what rates?

Answer Slide

Service should run between 16.00 and 00.00 since these are the times for the most demand on taxi rides, so people can prefer another option



- We should have 5 drop-off and pick-up nodes.
- 2 on Midtown Manhattan area (10001 and 10028 can be locations), 1 on La Guardia Airport, 1 in John F Kennedy Airport and 1 in Long Island area. These are the regions that has the most demands on taxi rides dataset.



- We should use copters for the MVP stage, if we have a stable product, we can later consider having homegrown hardware

- We should have a base price, and then a dynamic adding to that base price depending on distance.

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

Answer Slide

- We should have around 2900 rides, with 10% conversion rate, 20% MDE and 95% statistical significance. (Reference)

- Considering we are aiming for 400 rides per week, 7 week of observation period should be enough to come a conclusion.

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

Answer Slide

- New_user_created → People who created a new account so that we can measure new users.
- Ride_started → When a ride starts, so that we can measure ride count, and active user count.
- Ride_ended → When a ride ends, so that we can measure ride count, and active user count.
- Ride_cancelled → If the user cancels ride, so that we can measure ride count accurately.
- Ride _refunded →If the user has a bad experience, doesn't recommend Flyber, so we refund the ride.

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

Answer Slide

- Rate your ride from 1-5. (If below 4, what could we do better?)
- Rate your driver from 1-5. (If below 4, what could we do better?)
- Would you use Flyber again? Yes - No. (If no, what could we do better?)
- Would you recommend Flyber to friends and family? Yes - No. (If no, what could we do better?)
- Any additional notes would you like to let us know? (Free text)

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 - Target Population

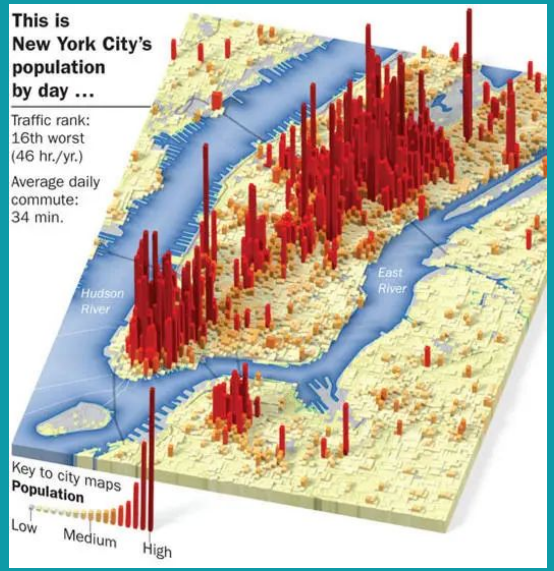
We should have a target audience that already uses taxis and ride sharing services.

Taxi usage rates are higher among to customer who have 40.000 dollars and more income annually so our target audience should have around the same level and more income.

These people more likely to be willing to pay more for a less amount of time spent in the traffic, or not being able to find a ride. Females are more likely to use them so we can give them priority.

Answer Slide - Product Proposal

With the increasing population in New York, traffic became a big problem on public transportation. Average commute time has increased and it creates a huge cost on customer, driver and also on cities end.



Answer Slide - Product Proposal

User Impact:
With Flyber we are expecting to reduce traffic density, so time people spend on traffic will be less.

Market Impact:
With the traffic density reducing, other drivers will be able to spend less time on rides, so they will be able to make more rides.

Business Impact:
Since this is a new method of transportation, it will create new business roles and new employment areas.

Answer Slide - Risks

- Since this is a new way of transportation, there are still concerns about safety. So we need to make sure drivers getting proper education and they are not harming existing air traffic.
- We can have a air traffic congestion in the future, so we need to continue monitoring properly.
- There might be new air traffic rules that contains commercial use.
- We should monitor our hardware and maintenance costs properly, it can get out of control in the first phase.

Answer Slide - State Cross-Functional Stakeholder Teams.

- Legal and Compliance Teams → To get proper authorizations, permits. And to consult them on air traffic rules.
- Learning Platform Teams → To prepare a proper education programme for drivers.
- Data Science & BI teams → To be able to get the proper data and the analyses and reportings.
- Sales & Marketing → We need advertisement and a sales strategy to be able to make profit from our product.
- Customer Experience → We need a department that can take care of customers compliance and also can make user research.