Data Product Manager Nanodegree

Applying Data Science to Product Management
Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service

Moinak Pyne

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

Answer 1: Taxis provide a convenient mode of transportation with multiple benefits, chief among them: taxis are a personalized system where ride-sharing or specific pick-up/drop of points are not specified; can book in advance; can potentially travel anywhere 24hrs and 7 days a week.

Answer 2: The user characteristics can be gleamed from the User Research Data: firstly, more women tend to use taxis than males (around 50%); mostly an even age distribution between 18 and 76 years of age; majority of taxi users fall in the income bracket of $40,001- $80,000

Answer 3: The main issues with taxis are: availability, surge pricing, safety of passengers, and traffic conditions

Answer 4: Digital ride sharing has issues such as: almost anyone can be a driver and they don't require any specific qualifications (hence safety); insurance gray area as most ride sharing taxis are personal cars.; and, surge pricing.

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

User Improvements- Two main user improvements which can happen due to a flying taxi service are: reduced travel time for a marginal price increase (taxis in New York costs are significant). Apart from daily commute, flying taxis provide a different view from a tourism point hence generating a new revenue source as well.

Market Improvements- In terms of market, the the market will evolve into adopting more technologies such as alternative fuels, AI and other tech. Also, infrastructure will see more investments leading to better development/

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

Answer 1: total records are 1048468

Answer 2: Each record represents an attribute regarding taxi rides in New York. The fields are: ID, Vendor_ID, Pickup_Datetime, Dropoff_Datetime, passanger_count, pickup-longitude, pickup_latitude, dropoff_longitude, dropoff_latitude, store_fwd_flag, duration, distance

Answer 3: Primary key is ID + Pickup_Datetime

Answer 4: The date range is bound between 1-1-2016 and 6-30-2016

Answer 5: The dataset mainly focusses on Manhattan as well as Bronx, but areas such as Brooklyn, Queens, Staten Island and New Jersey are not included. Majority of the data points are focused on Manhattan. Outliers can be found near John F Kennedy airport and East Elmhurst and Newark airport.

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

1st SD = Mean + SD
2nd SD = Mean +2*SD

Duration
Mean: 962.2 sec, Median: 662 sec, 1st SD: 6815.49 sec, 2nd SD: 12668.78 sec

Distance
Mean: 3.442 miles, Median: 2.095 miles, 1st SD: 7.824 miles, 2nd SD: 12.206 miles

passenger counts
Mean: 1.664 people, Median: 1 person, 1st SD: 2.978 people, 2nd SD: 4.292 people

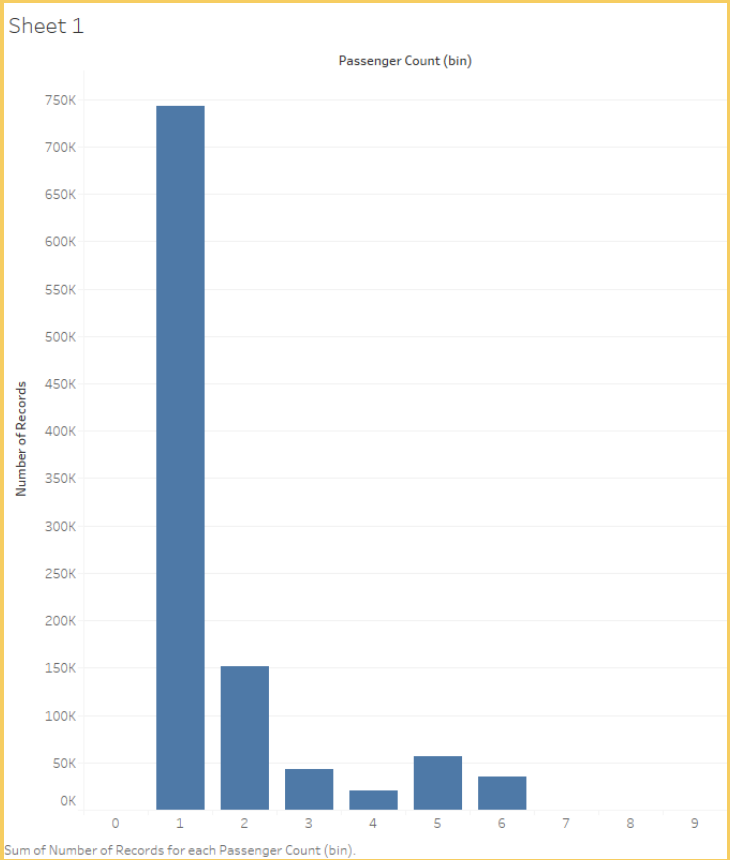
duration-to-distance ratio
Mean: 4,687 sec/mile, Median: 280.8 sec/mile , 1st SD: 929060 sec/mile, 2nd SD: 1853433 sec/mile

price [2.50 + (0.50 * ([Duration]/60)) + (2.5 * [Distance]) + (3.30 *[Passenger Count]) + 2.5]
Mean: $27.12, Median: $21.59, 1st SD: $78.22, 2nd SD: $129.32

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



Based on the histogram, passenger counts of 1 is the most popular one, hence representing a large market.

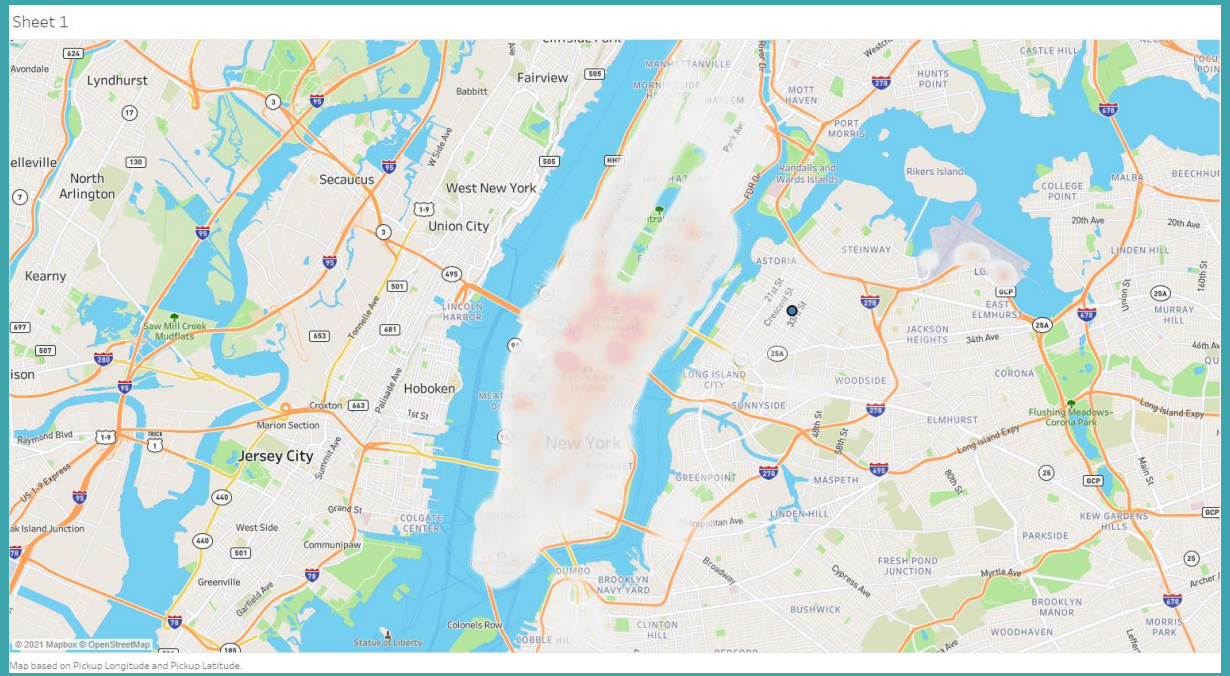
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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

Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?

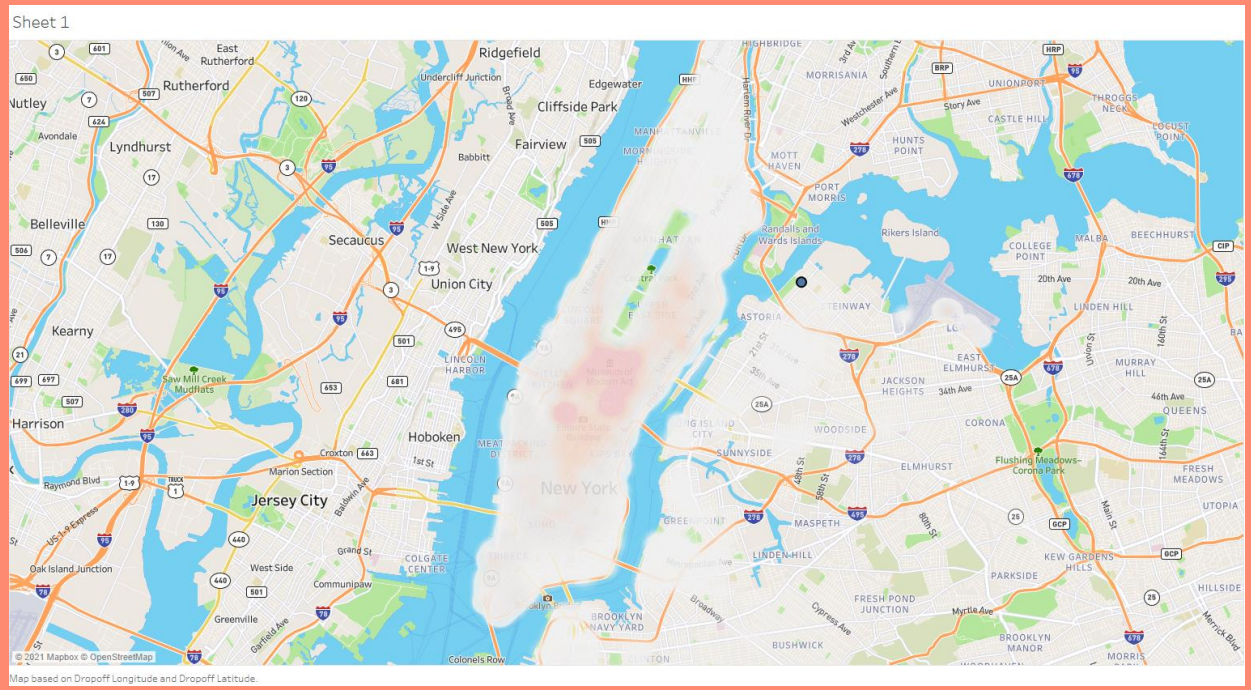
Answer: Herald Square, Murray Hill, Manhattan



Answer Slide

Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?

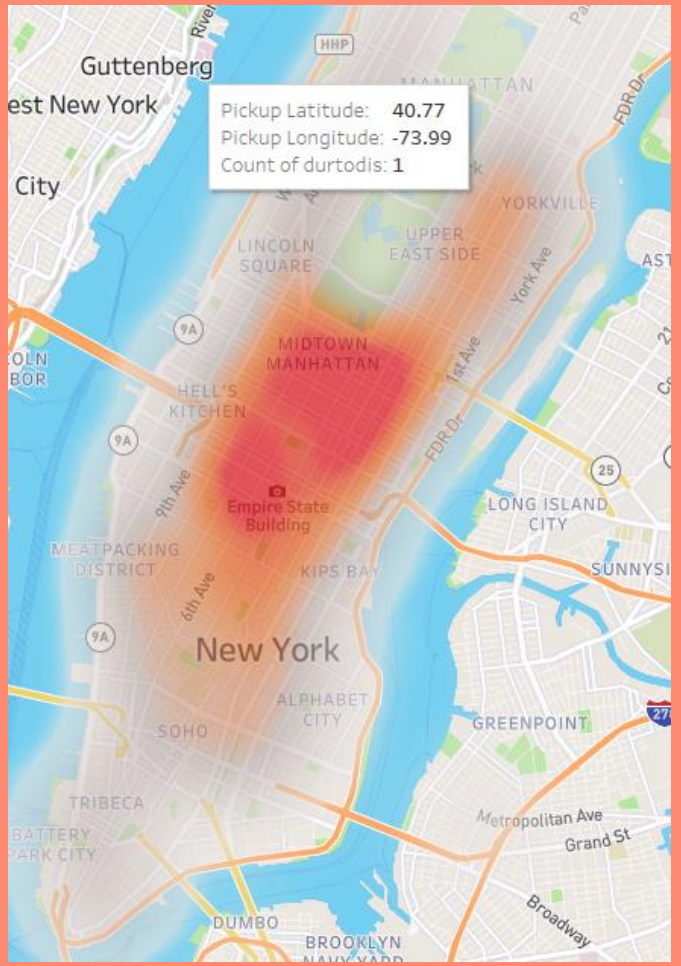
Answer: Herald Square, Manhattan



Answer Slide

Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?

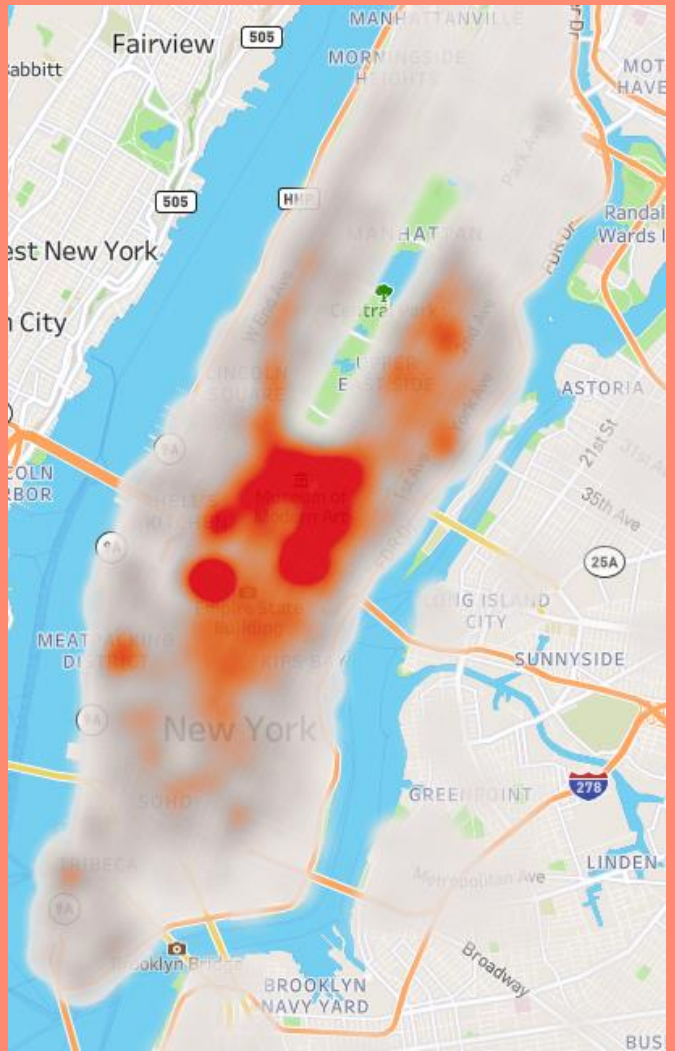
Answer: Manhattan, Midtown South, Hell's Kitchen, Diamond District, Grand Central Terminal, Midtown East



Answer Slide

Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off?

Answer: Manhattan, Grand Central Terminal, Murray Hill, Midtown East



Answer Slide

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: As flying taxies will potentially be operating at an altitude of 300 meters, potentially take-off and landing above skyscrapers are suitable for this kind of service**. Hence Manhattan with its tall buildings seem to be highly suitable.

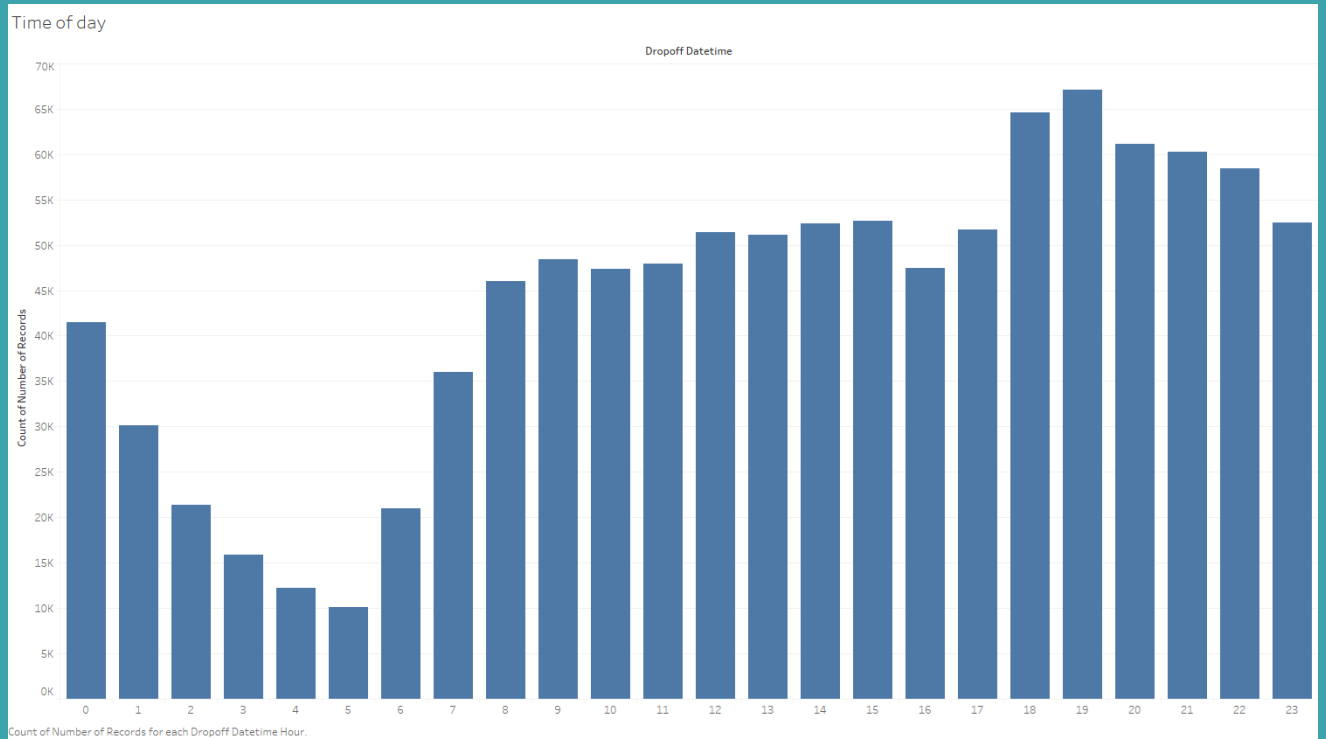
**“Summary+Mission+and+Requirements_Uber.pdf”

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.

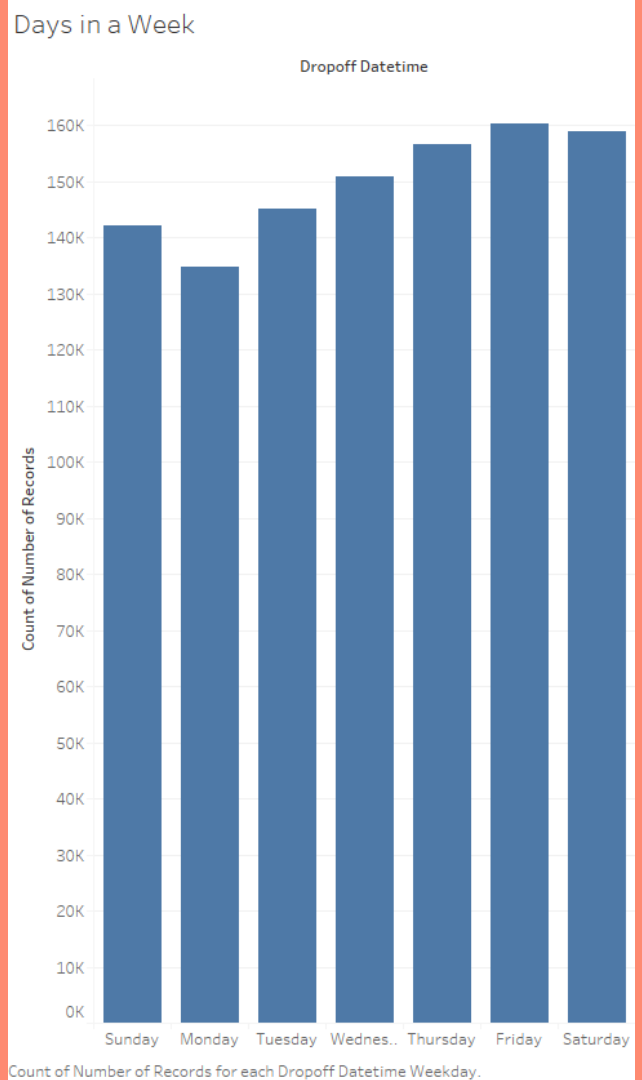
Answer Slide

Answer 1: During 1800 Hrs. and 1900 Hrs. of the day, higher volumes of ride pick-ups are observed.



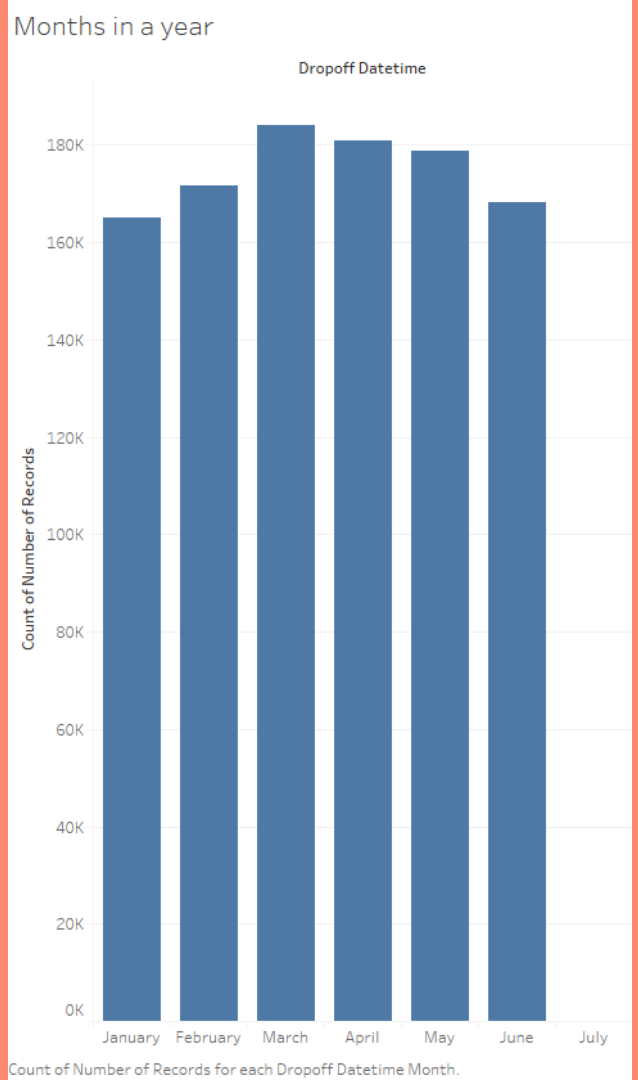
Answer Slide

Answer 2: Saturdays and Sundays seem to have higher pick-up volumes in all weeks



Answer Slide

Answer 3: March seems to have the highest pick-ups with counts decreasing in April in May. Hence, during summer the maximum volume of pick-ups are observed.



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:

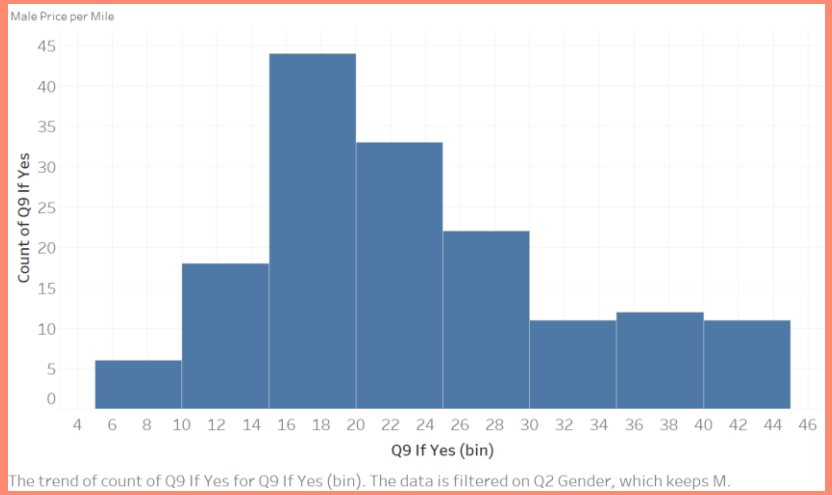
- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?
- What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?
- What is the different personas/segments of negative sentiment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

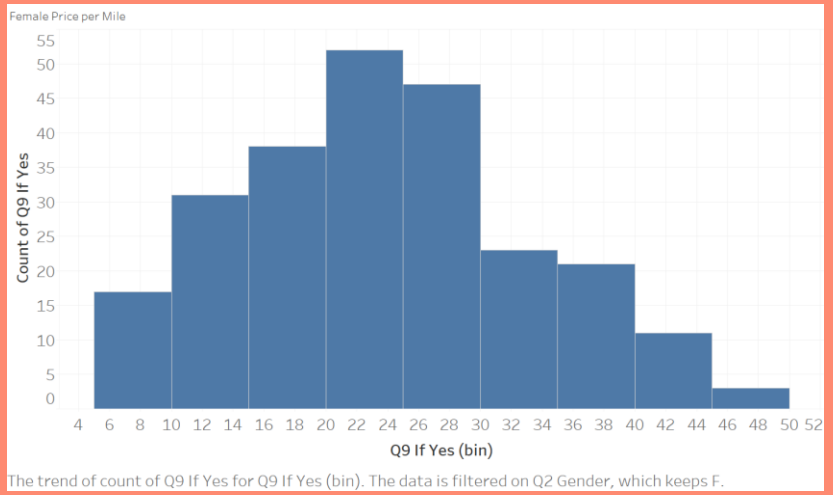
Answer Slide

Answer 1:
81.77% of Men and 79.15% women said yes to Flyber adoption, hence both genders have similar inclination
Age groups with the lowest inclination are 20, 33, 37, 52 and 55
Out of all age groups, the highest percentage of Flyber adoption of 85% came from the income group of 120,000-200,000
Majority neighborhood have high adoption rates except Inwood (25%) and West Harlem (16.66%)

Answer Slide

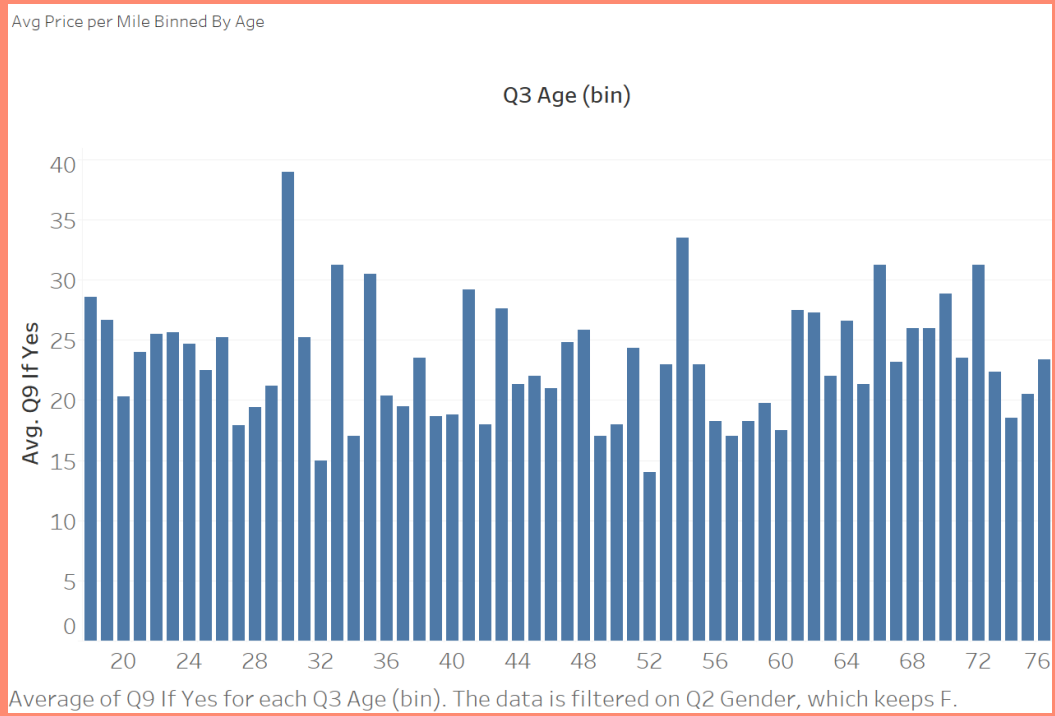
Answer 2:
price per mile based on gender:
Highest price bin in case of Males is 15-20, and for Females is 20-25





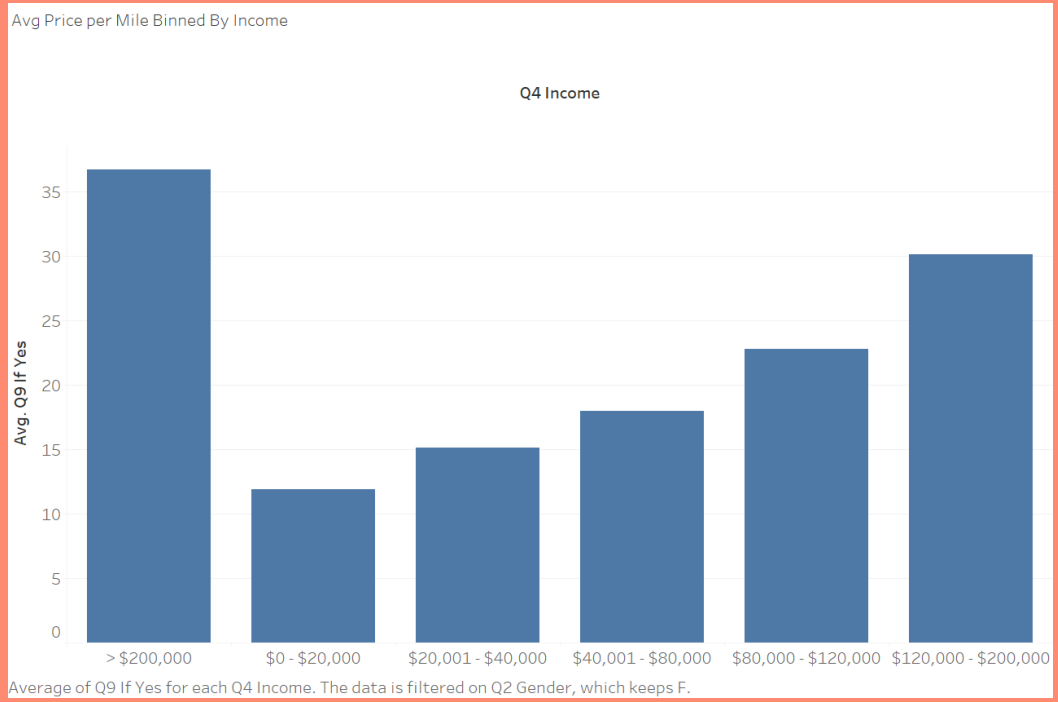
Answer Slide

Answer 2:
potential price per mile based on age shows that the highest average price per mile belongs to the age of 30



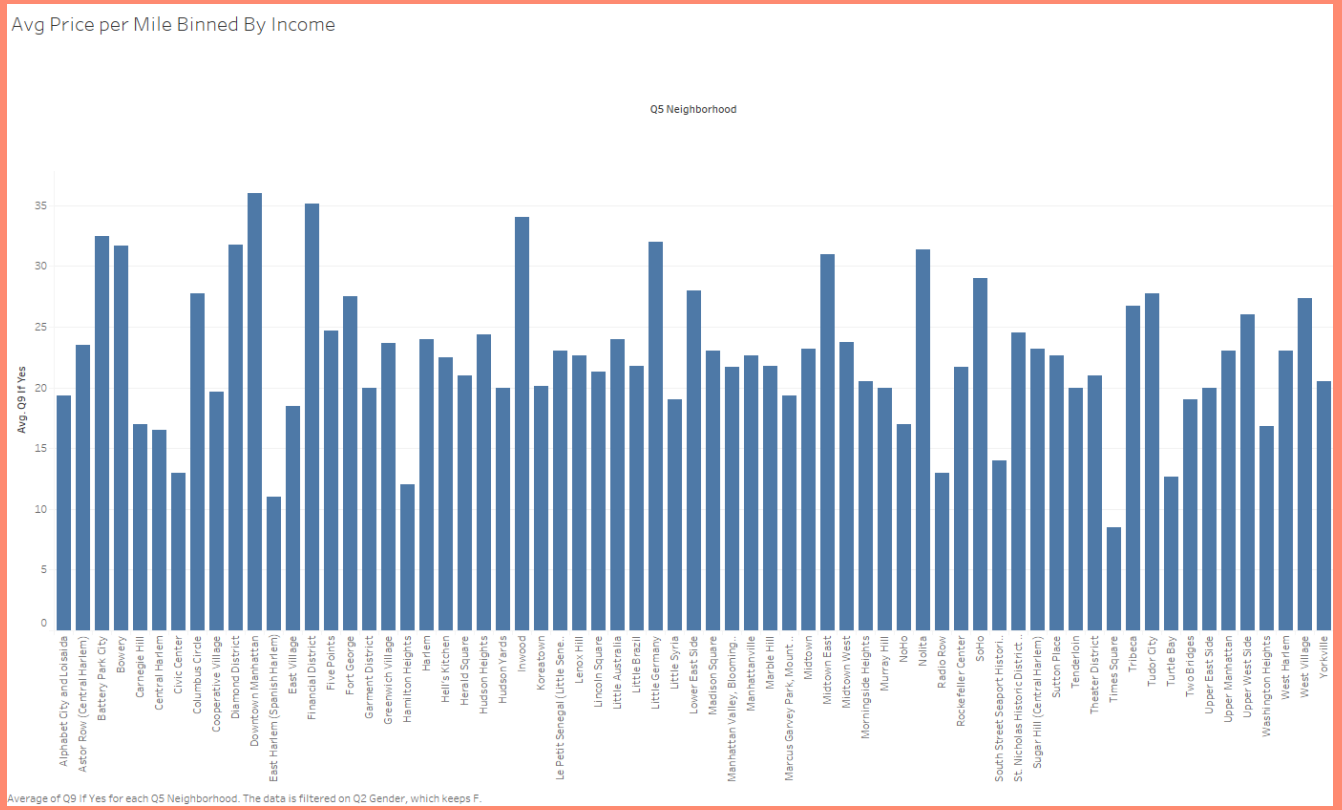
Answer Slide

Answer 2:
potential price per mile based on income shows that the highest average price per mile belongs to the income bracket of >$200,000



Answer Slide

Answer 2:
potential price per mile based on Neighborhood shows that the highest average price per mile belongs to Diamond District.



Answer Slide

Answer 3:
The different personas/segments of negative sentiment are:

Reason	Count
I think it is more unsafe than taxis	23
It would be too expensive	15
Wouldn't trust the person flying it	15
Not worth the extra money	13
It's unsafe	9
It's dangerous	6
My commute is already straightforward	6
I think it's more dangerous	5
The skies would get just as crowded as the streets	5
My commute is fairly efficient	3

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

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

(Fill out your answer here)

Answer Slide

(Fill out your answer here)

Answer Slide

(Fill out your answer here)