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

- Taxis are used for commuting between different geographical locations within the same city.

- The characteristics of the users that make use of this service are usually people in a mid high level socioeconomic level that use the service on a regular bases for commuting for work purposes (meeting clients, base work location, etc). The other type of users are more for entertainment purposes which means be for tourists or locals.

- The existing pain points with taxis is mainly that the limitation to moving with this service is the dependency there is on road infrastructure and the congestion that can come along with this when travelling at peak time sand according to how well the traffic management is in the city. The other paint point for taxis is that for longer distances travel, the time of travel is limited not only by the road infrastructure, highways, etc, but also by the speed limits and the maximum velocity a car can travel.

- The pain points with digital sharing service is that it is really hard to match offer and demand from drivers and riders at the same time. Usually people require this service in peak times of the day and there might be a saturation of the offer.

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

- The main user improvement a flying taxi service would be to reduce the total time travel from origin to destination.

- With regards to market improvement, the flying taxi service would not require such a high infrastructure building and maintenance costs, as the road infrastructural convenential taxi requires.

  The non-recurring costs of building the infrastructure is huge as well as limited by the space there is in land for dense cities. However, when thinking of a flying taxi service, the fact that we can use a 3D routes means not only that the infrastructure costs is reduced to the take off and landing areas, but also that the amount of trajectories to go from point to point are much larger and therefore reducing congestion constraints and consequently reducing time travel.

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!

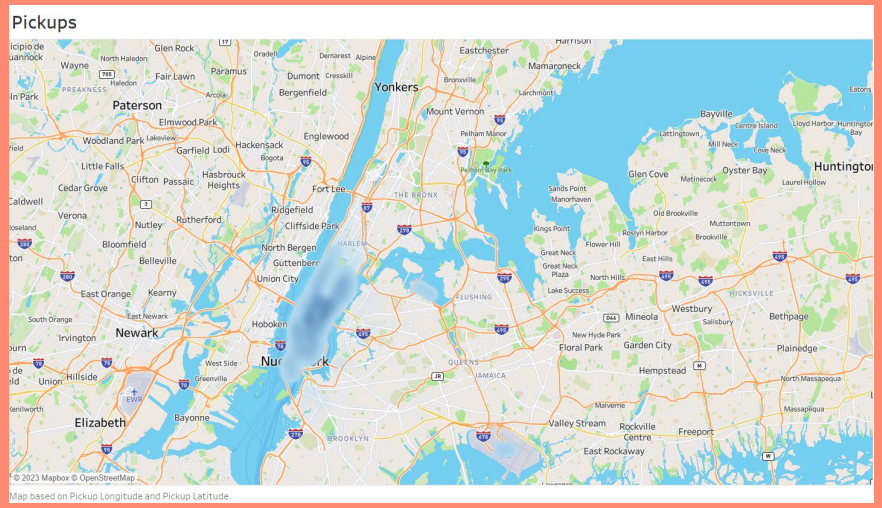
Acquire a high-level understanding of the granularity and scope of the dataset, to inform the basis for your analyses:

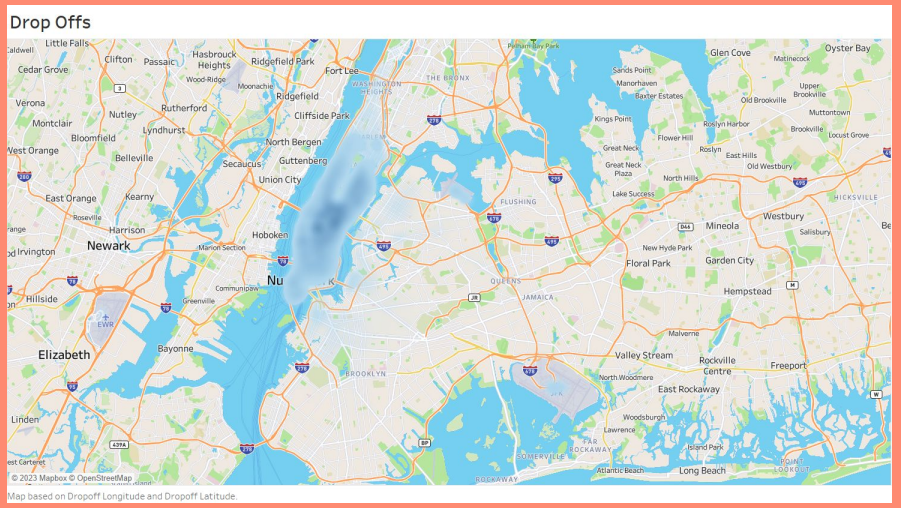
- How many records are in the dataset
- What does each record represent?
- What is the primary key?
- What date range is your dataset bound to?
- What are the geographical bounds of this dataset? Is it limited to Manhattan, or is Brooklyn, Queens, Staten Island, the Bronx, and New Jersey included? Where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

Answer Slide

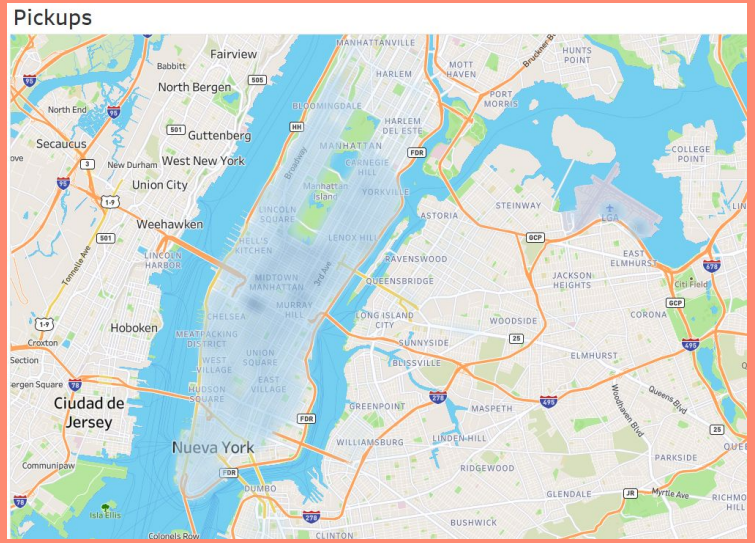
- There are 1 048 468 records in the table.
- Each record represents a taxi ride with the coordinates of the pick up point and drop off points, their corresponding times and the number of passengers in the taxi.
- The primary key for the table is the id column.
  This is done by making sure
  SELECT count(distinct id) FROM taxi_rides = SELECT * FROM taxi_rides
- The date range for the dataset is (1st January 2016, 1st July 2016)
- The geographical bounds for this dataset are mainly for Manhattan JFK, La Guardia and Newark Liberty airports when we look at the density map plots in the next slide. Of course there are drop offs and picks outside of these areas (e.g. Queens) but mainly the majority of the movement happen in this area. There are also some points that are clearly outliers in the data that we will remove

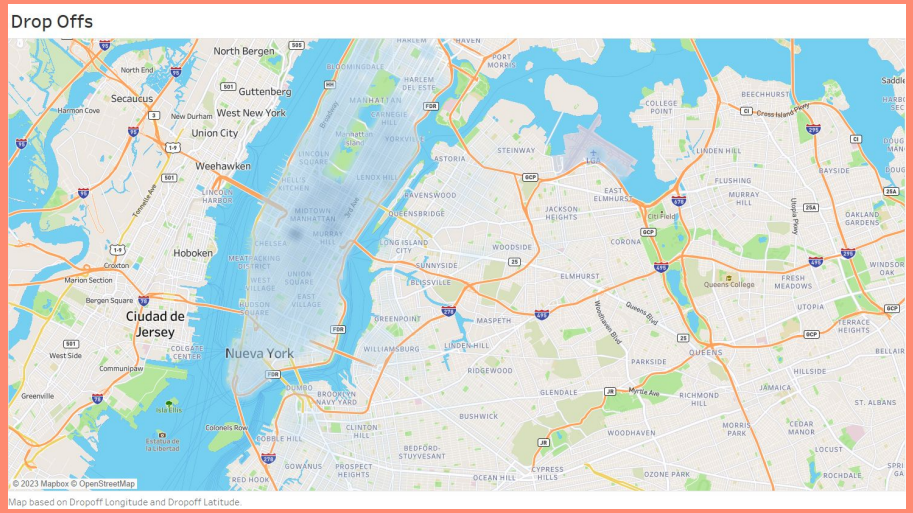
Answer Slide





Answer Slide





Here we can see what are the neighbourhoods with the highest density pick ups and drop offs

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
- distance-to-duration ratio
- price

Answer Slide

	Mean	Median	1 std	2 stds
Duration (seconds)	962.2	662	5853.30	11706.60
Distance (miles)	3.44	2.09	4.38	8.76
Passenger Counts (#)	1.66	1	1.31	2.62
Duration to Distance Ratio (mins/mile)	78.11	4.68	15406.22	30812.44
Price ($)	19.16	13.55	50.90	101.80

Price = 2.5 + (1.56*Distance*1.61) + (Duration/3600)*30

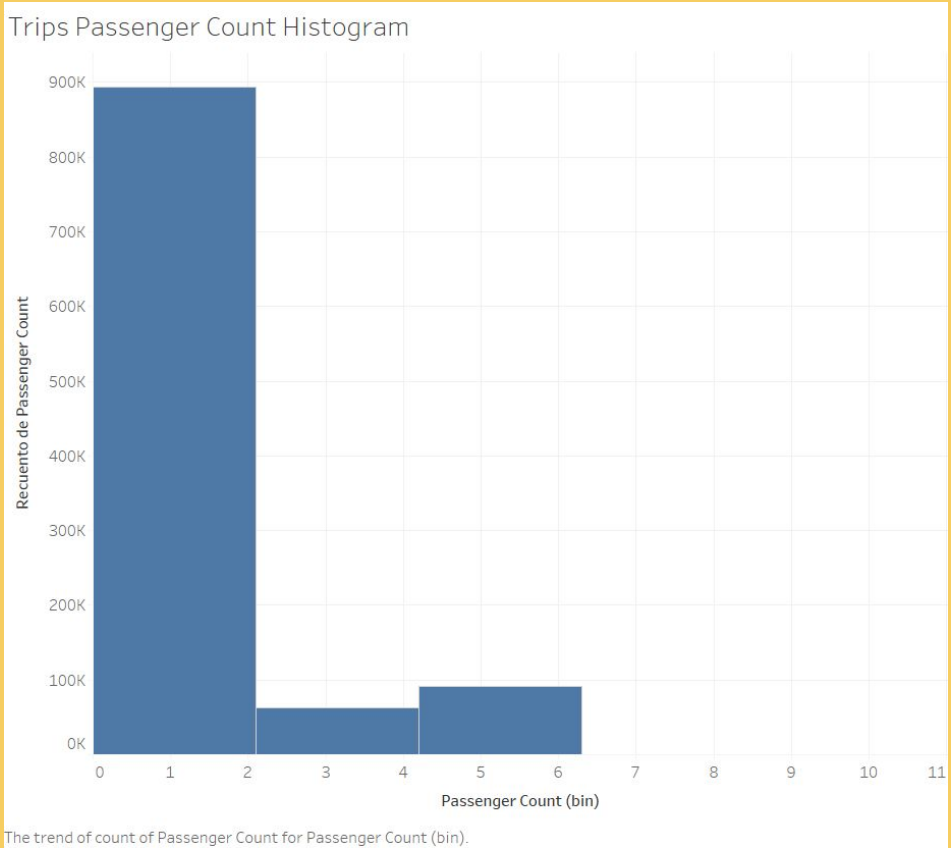
https://knowledge.udacity.com/questions/645559

https://www.estimate.taxi/rates/united-states/new-york

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



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:

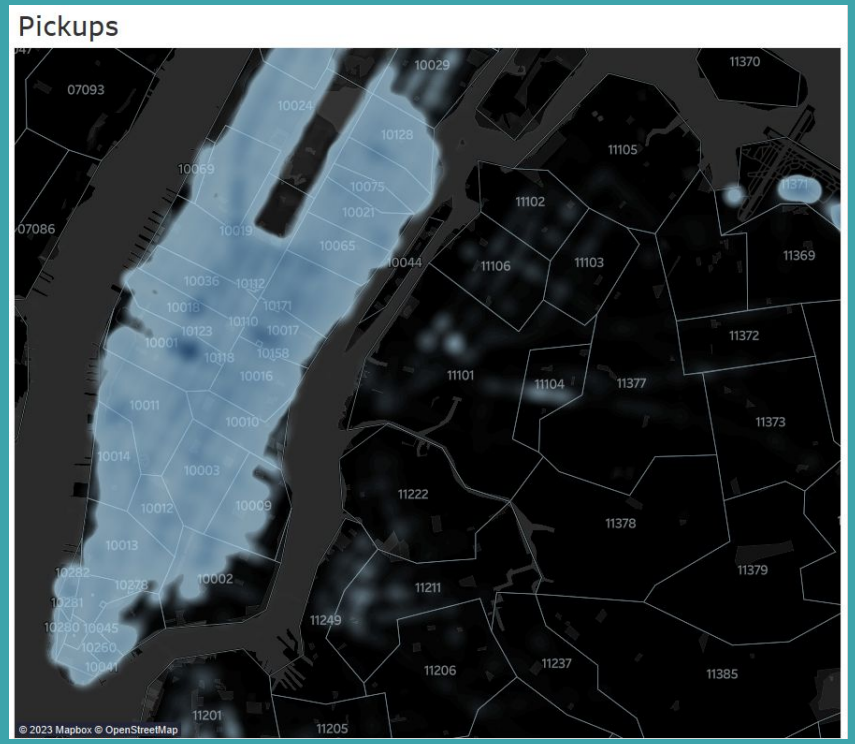
- Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?
- Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?
- Which neighborhoods/zip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

Highest density of Pick ups Neighbourhoods are in Manhattan: Midtown Manhattan, SOHO

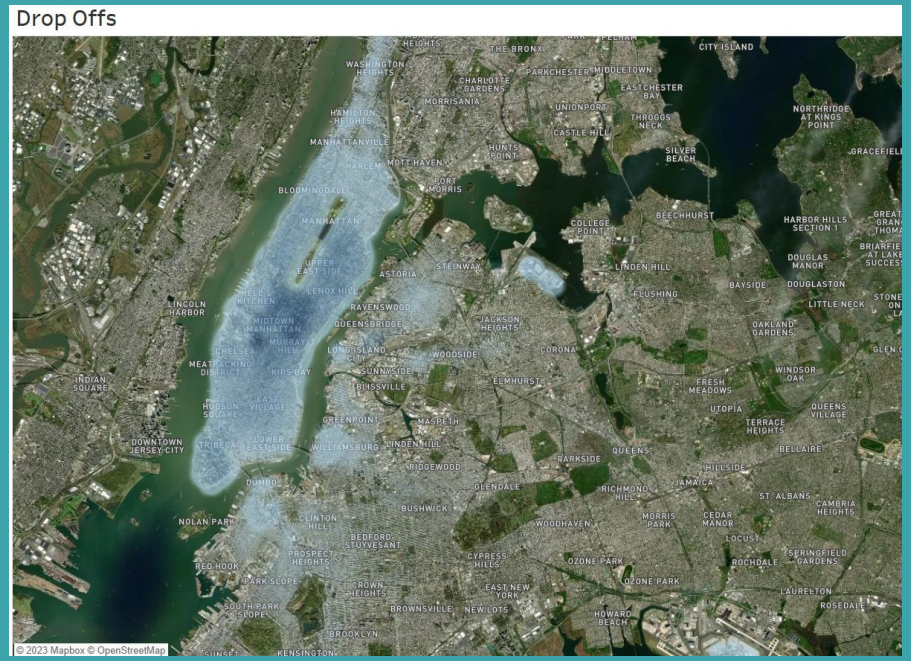
Answer Slide

Highest density of pickups ZIP Codes are: 10001, 10178, 10018



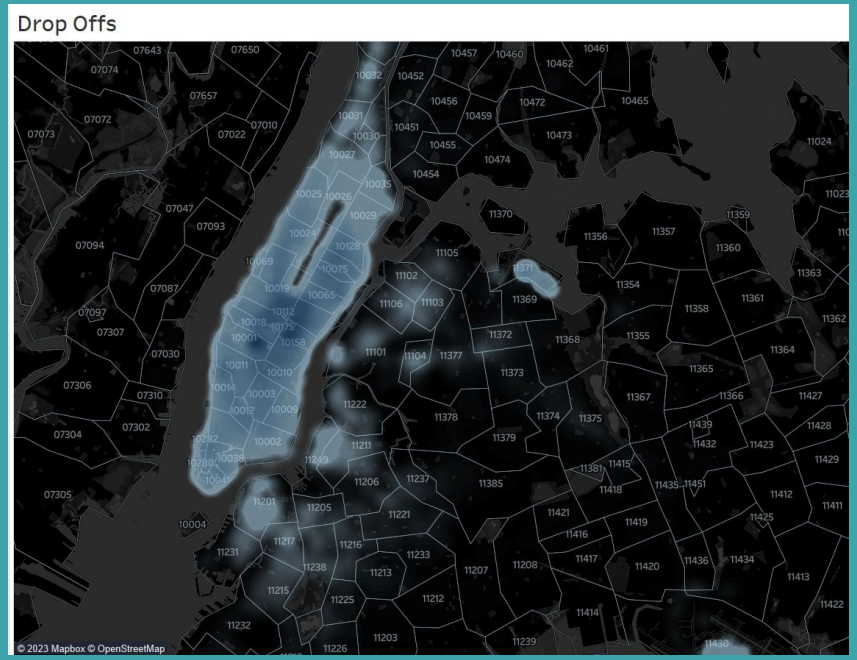
Answer Slide

Highest density of Drop Offs Neighbourhoods are in Manhattan: Midtown Manhattan, SOHO, airports



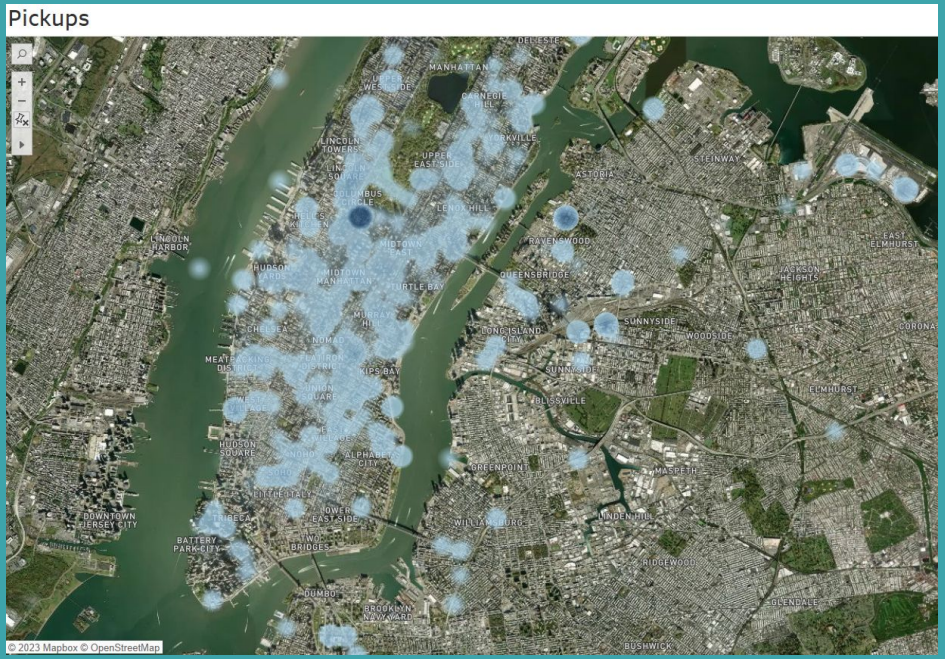
Answer Slide

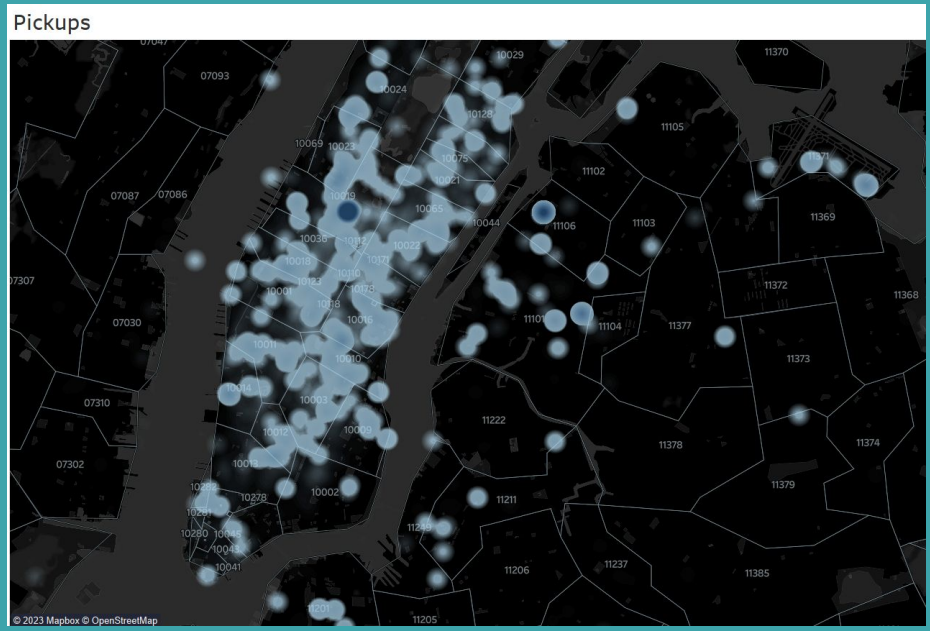
Highest density of drop offs ZIP Codes are: 10175, 10001, 10112



Answer Slide

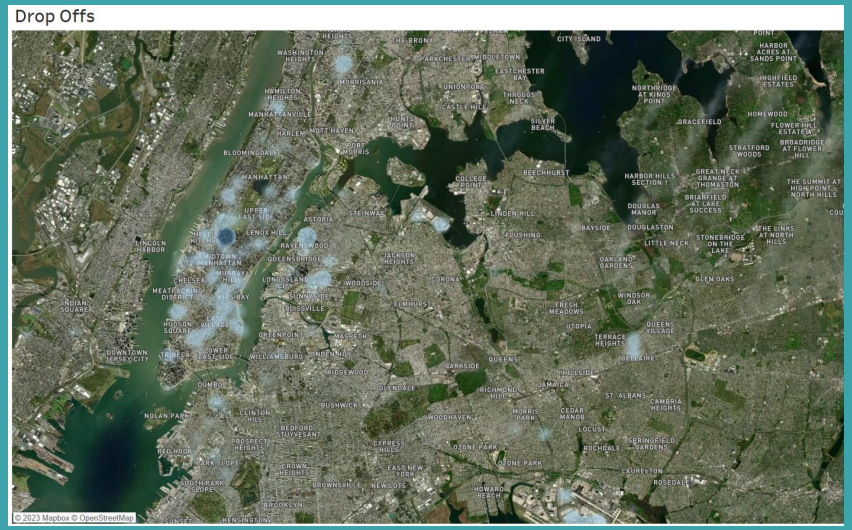
Highest density of Pick ups Neighbourhoods based on Duration to Distance ratio are:
Between Midtown Manhattan & Columbus Circle and Ravenswood; Zip: 10019 & 11106

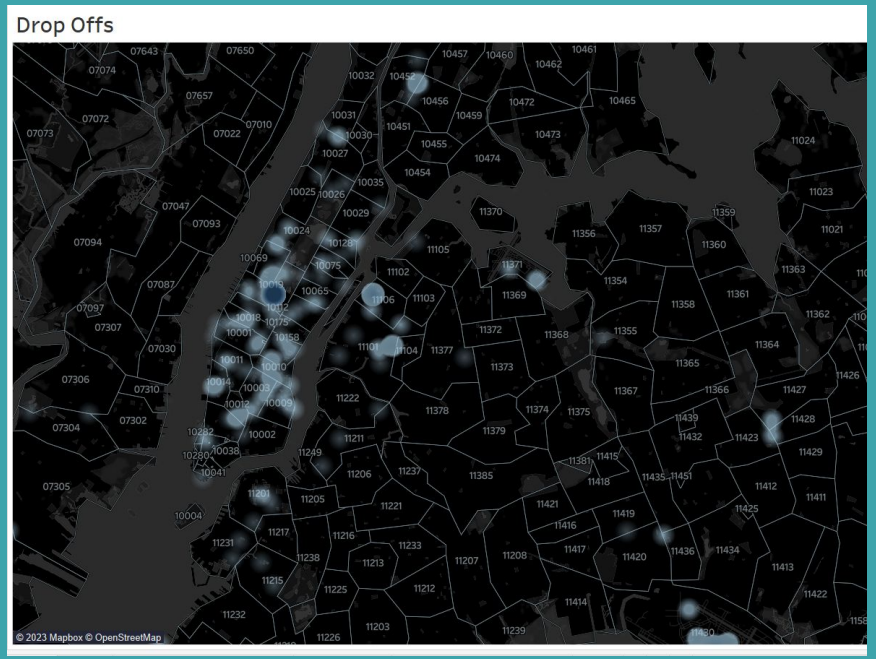




Answer Slide

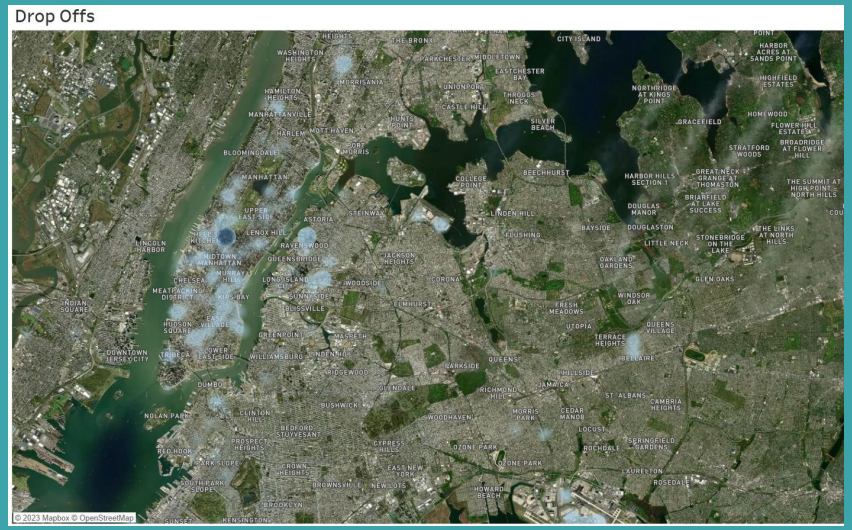
Highest density of Drop Offs Neighbourhoods based on Duration to Distance ratio are: Between Midtown Manhattan & Columbus Circle ; Zip codes:10019

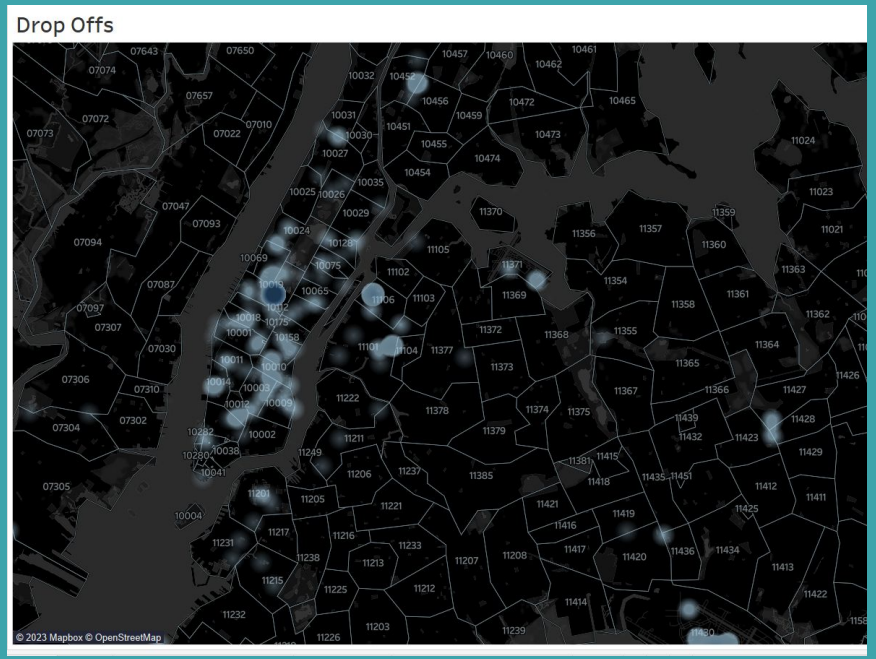




Answer Slide

Highest density of Drop Offs Neighbourhoods based on Duration to Distance ratio are: Between Midtown Manhattan & Columbus Circle ; Zip codes:10019





Answer Slide

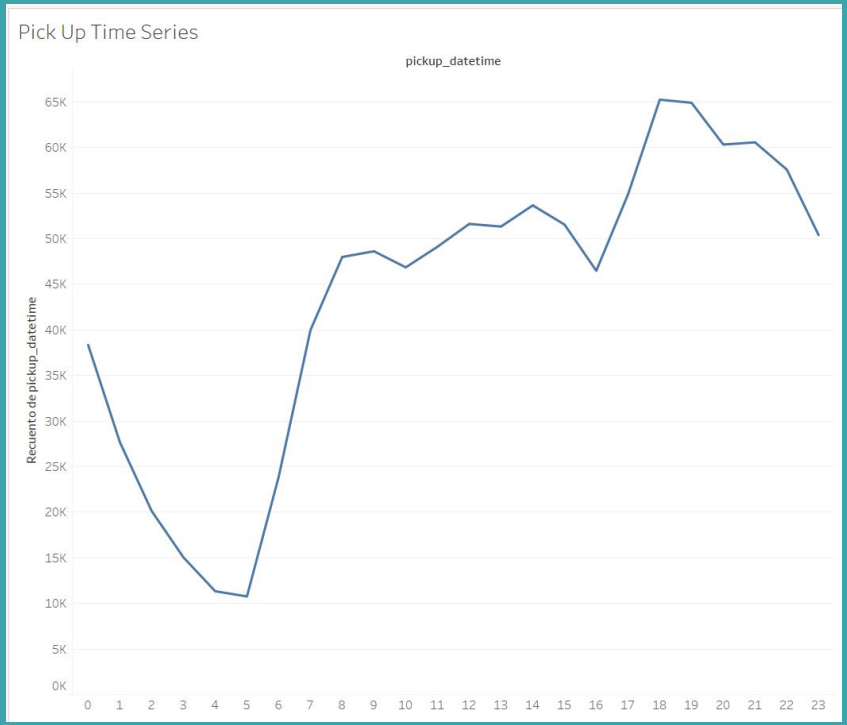
The potential neighbourhoods for both pick ups and drop off should definitely include the middle part of Manhattan since that is where we can see a higher duration to distance ratio which makes sense because there will be higher traffic in these areas.

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.

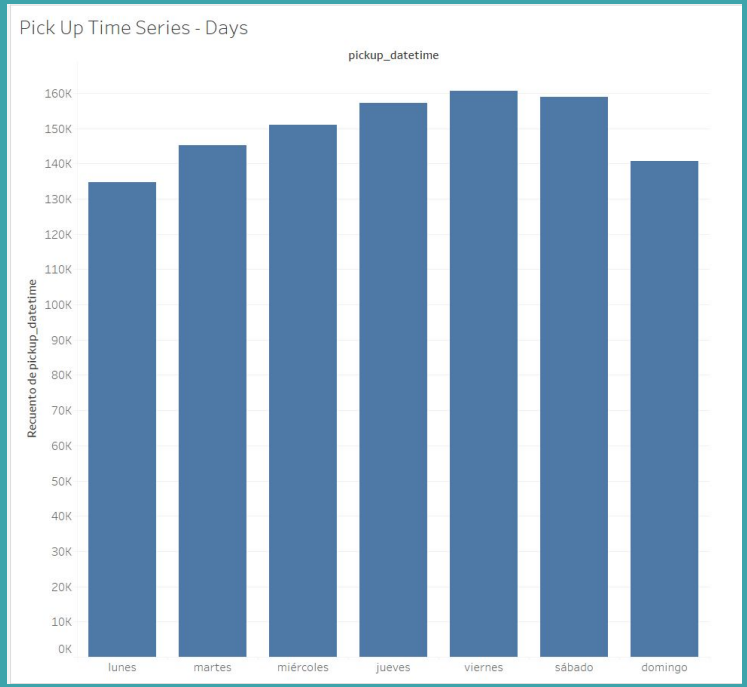
Answer Slide

The highest peak of pick up times occurs during 18:00 and 19:00 and more broadly between 18:00 and 22:00



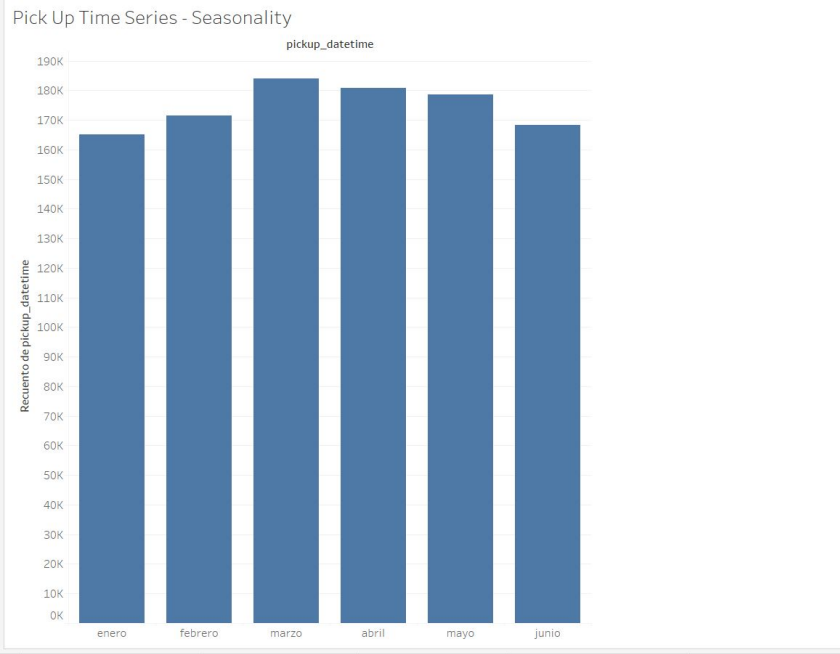
Answer Slide

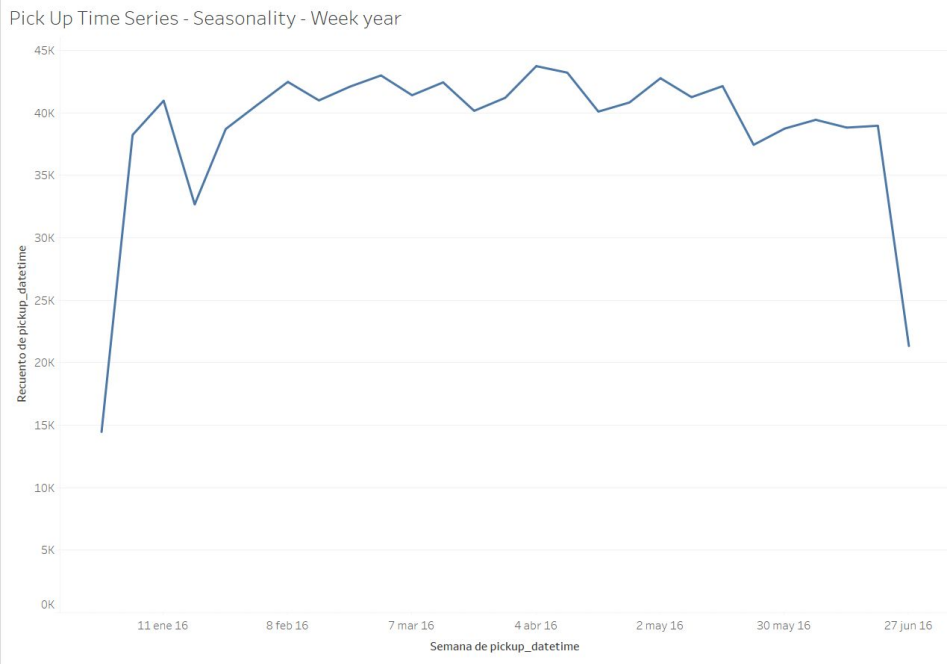
The highest peak of pick up times occurs on Friday first then Saturday followed by Thursday



Answer Slide

The highest peak of pick up times occurs in March followed by April then May. We also see a bigger decrement in pick ups in the 3rd week of January





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:

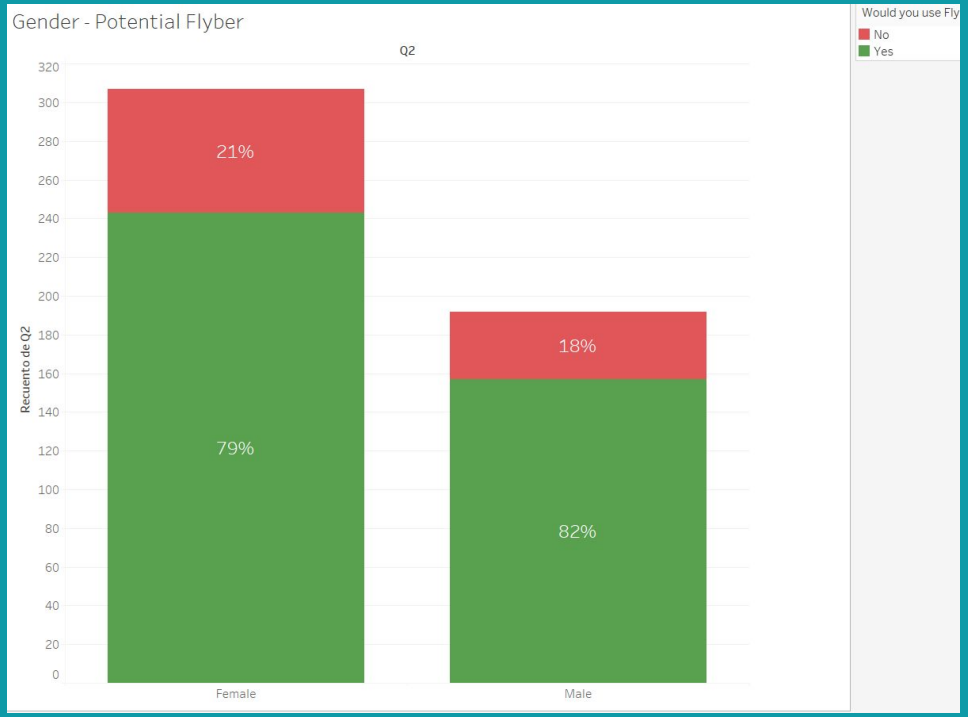
Q1 - What is your email?
Q2 - What gender do you identify as?
Q3 - What is your age?
Q4 - What is your annual income? (income bands)
Q5 - What neighborhood do you reside in?
Q6 - Do you currently use taxis? (Y/N)
Q7 - Do you currently use ridesharing services? (Y/N)
Q8 - Would you use a flying taxi service, 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?

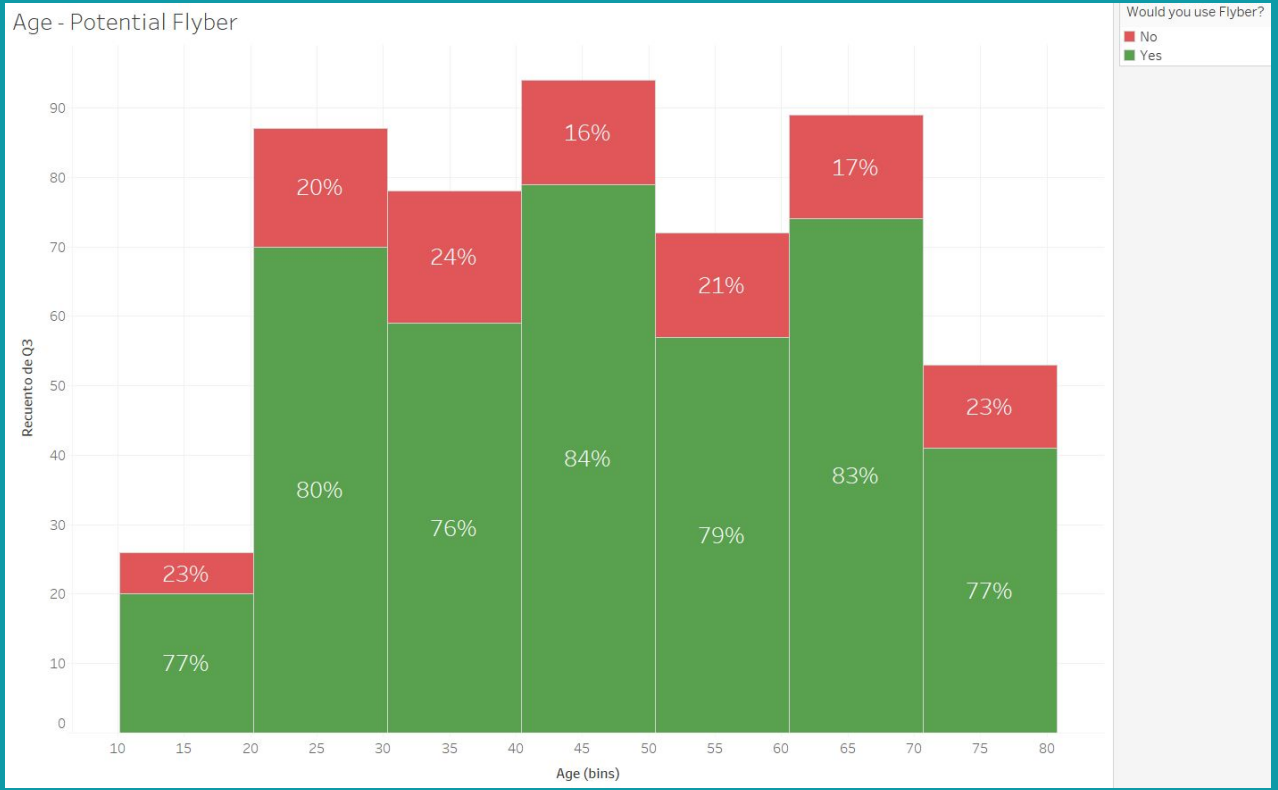
Answer Slide

Flyber potential usage- Gender: Slightly higher percentage of adoption for Male



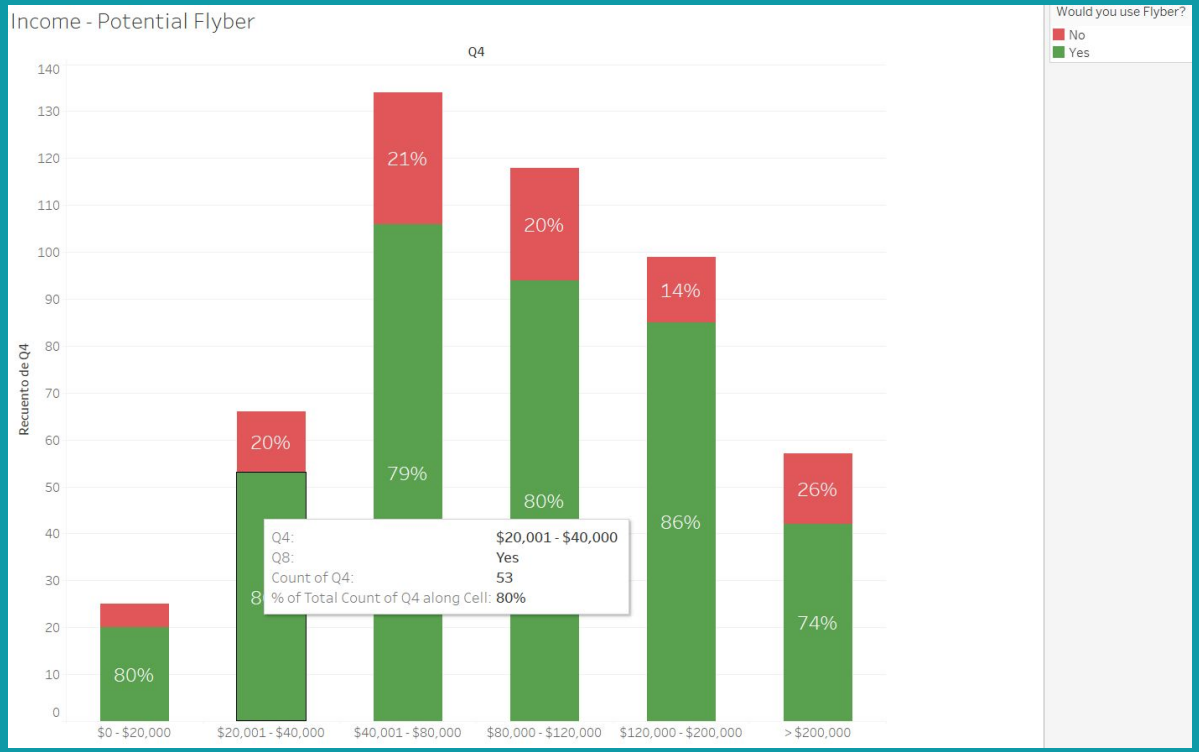
Answer Slide

Flyber potential usage- Age: Highest potential 40 -50 followed by 60 - 70



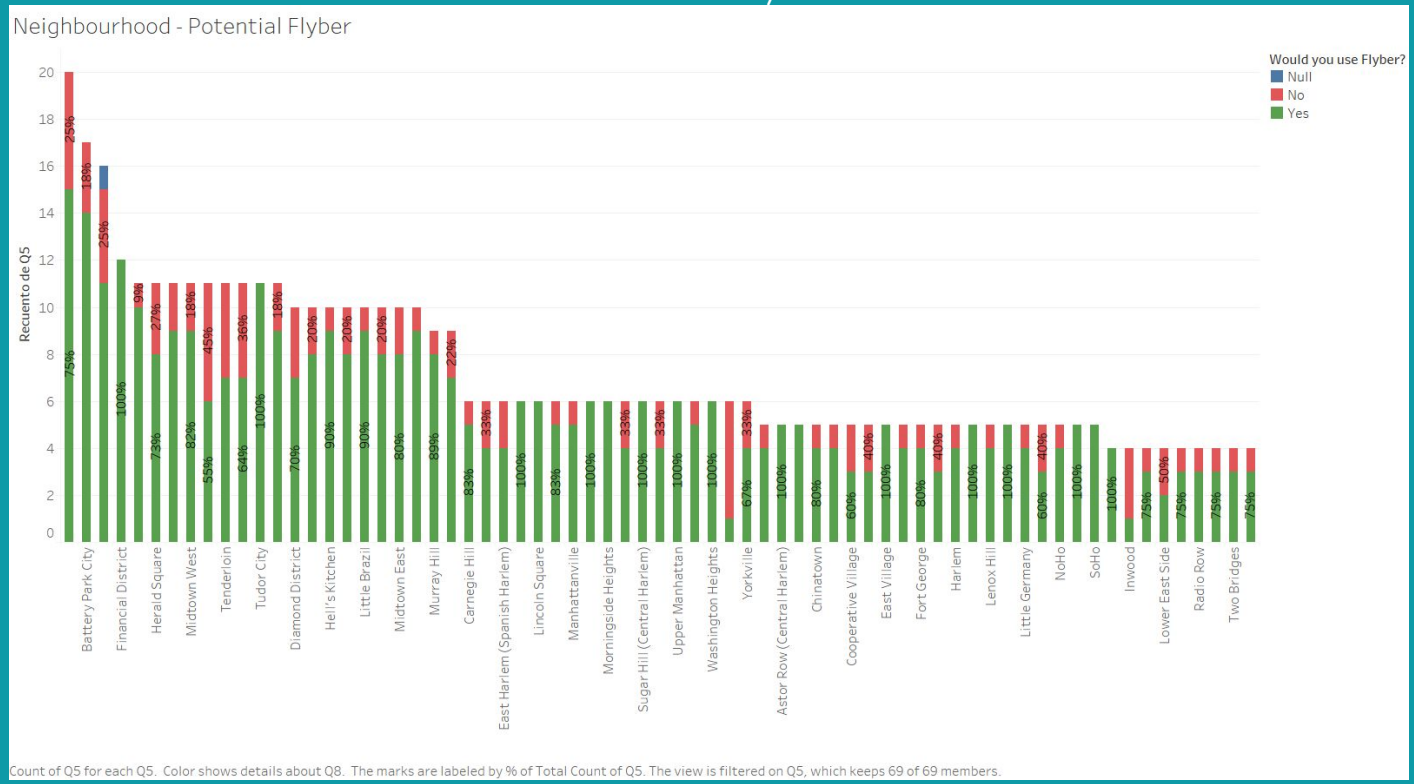
Answer Slide

Flyber potential usage- Income: Most potential highest income in the bracket 120 -200 K$



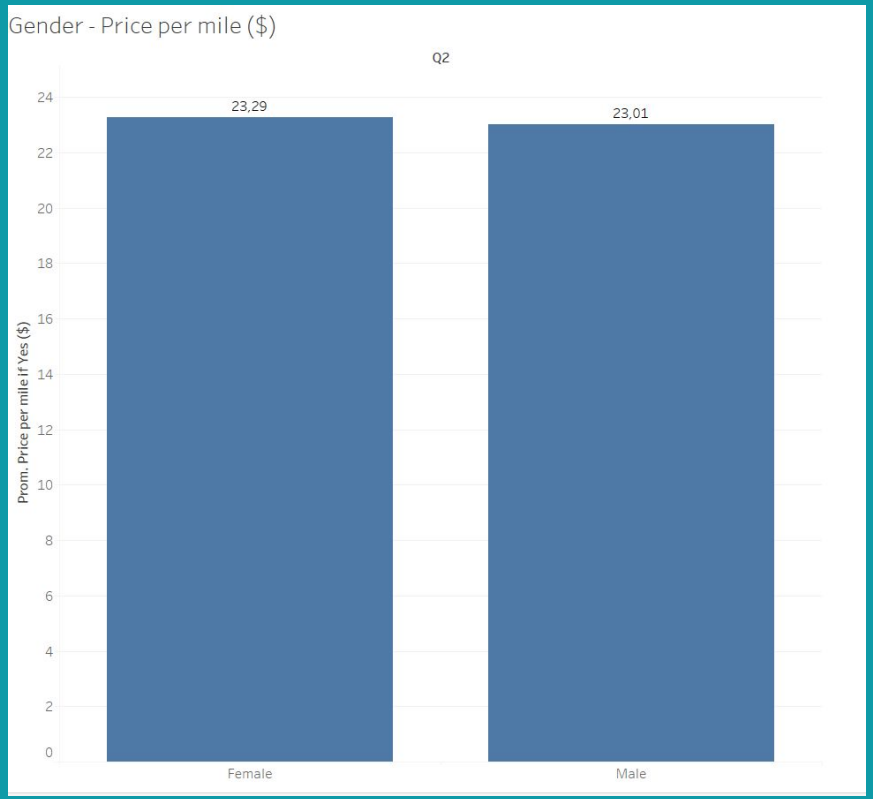
Answer Slide

Flyber potential usage- Neighbourhood: Very heterogeneous neighbourhoods. Most common is Midtown with 20 with 75% that would take Flyber



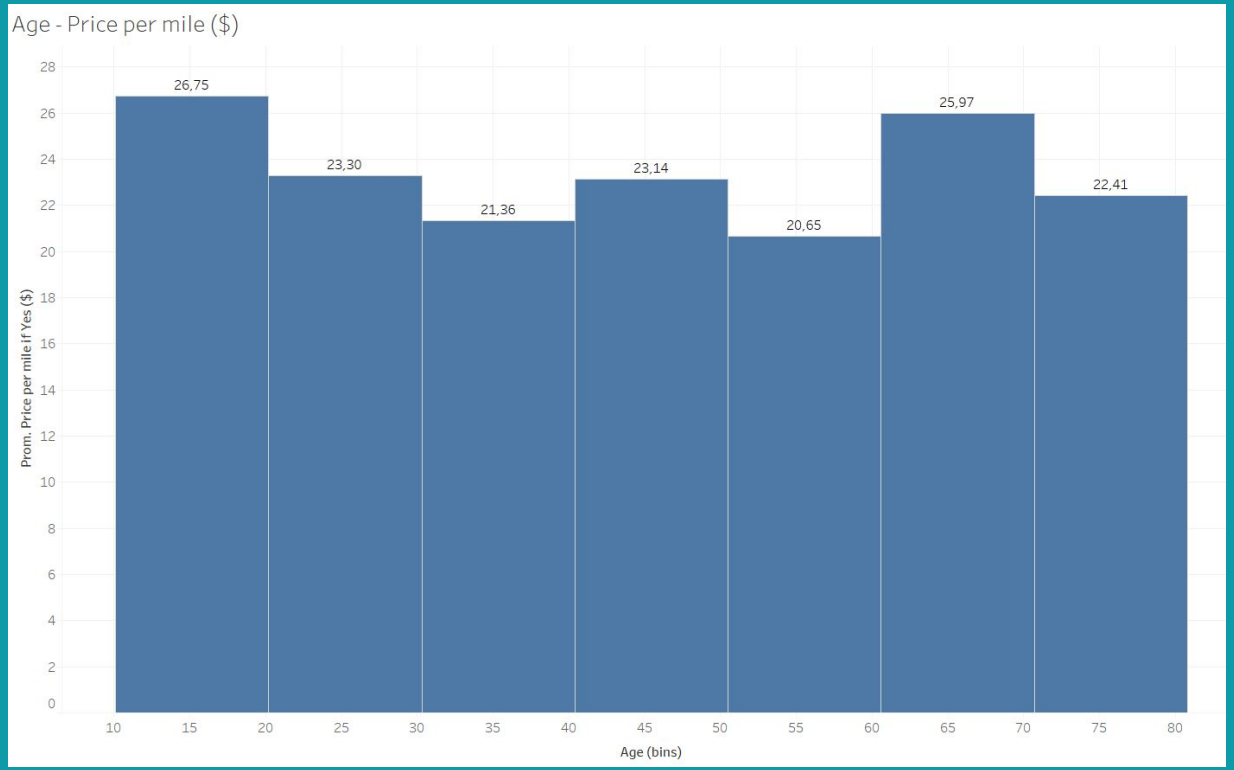
Answer Slide

Flybe Gender - Price per Mile ($): Very similar across genders



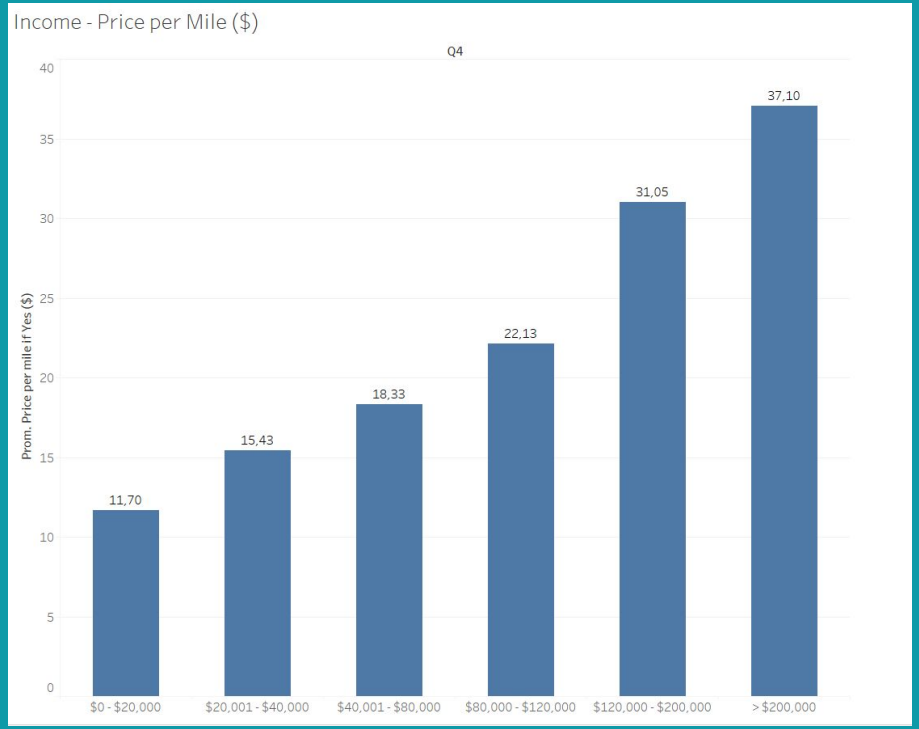
Answer Slide

Flyber Age - Price per Mile ($) : Surprisingly highest is the range from 10-20 years followed by 60 - 70.



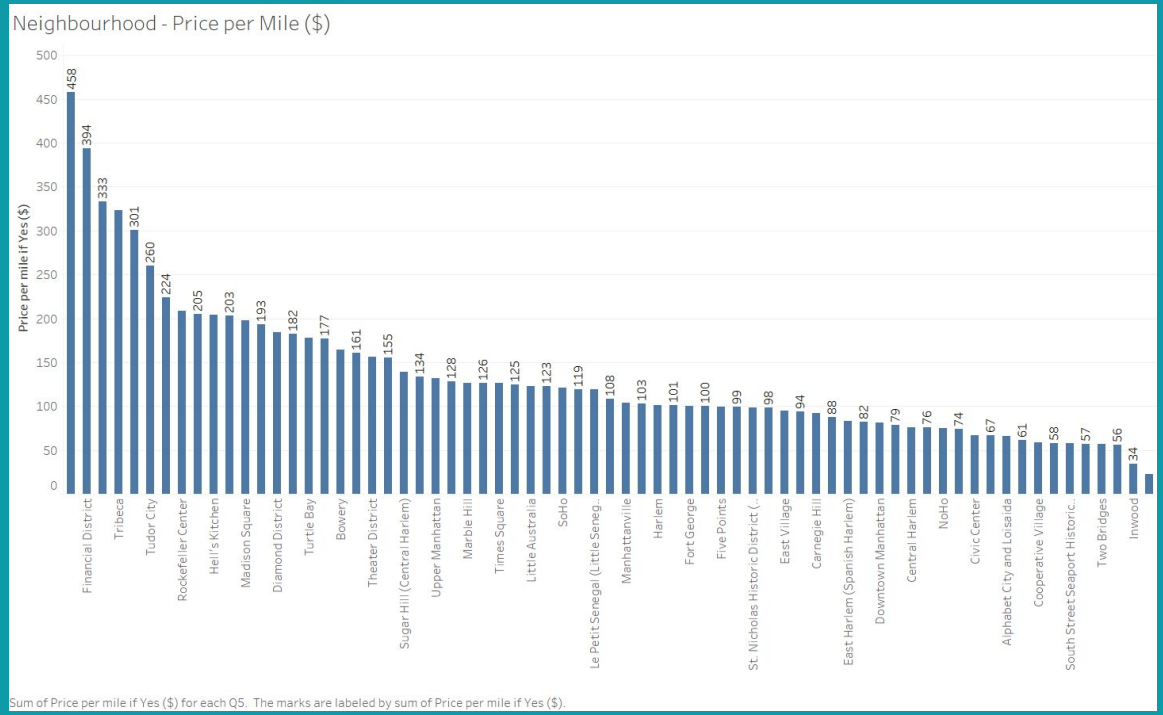
Answer Slide

Flyber Income- Price per Mile ($) : As expected the higher the income the higher the price per mile rate the people are willing to way for Flyber if it existed.



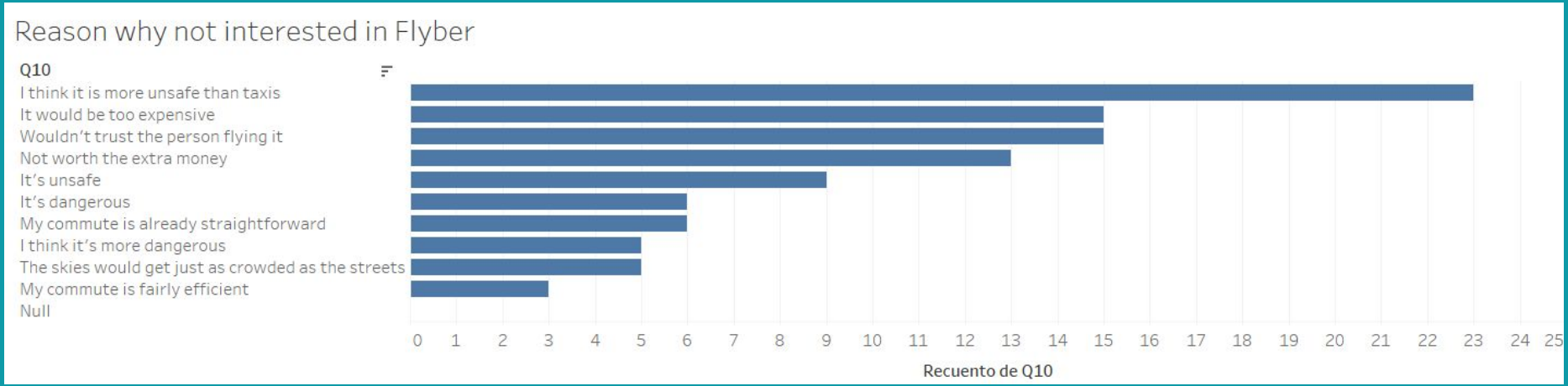
Answer Slide

Flyber Neighbourhood - price per mile ($): This is higher to analyze since the maximum number of people per neighbourhood is 20 for Midtown, but in any case the neighbourhood willing to pay the highest price per mile is Battery Park City



Answer Slide

Why are the potential clients susceptible to Flyber. Again, not a lot data for this, so would need further investigation, but here are the preliminary results from this survey. Based on this the main reason behind the people saying no, is that they believe it would be unsafe and they imagine that it would be too expensive. Other than that there is no clear mapping of a segment that rejects the Flyber be age, gender, income or neighbourhood.



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)