

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

Flyber

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

- **What are taxis used for?**

Taxis are used for transporting anyone from a point to another point. Usually taxis are utilized for someone who want to travel a relatively long distance

- **What are the characteristics of the users that leverage them?**

Usually there are used by corporate customers, who usually do business trip and have their taxis fare reimbursed by their company, or usually for family trip but they have no car. Taxis are usually used for airport transfer as well since many people would prefer not bringing their vehicle to the airport

- **What are existing pain points with taxis?**

It's hard for us call the taxi, since we have to go to taxi shelter first. Also the argometer which determine the transportation fee are tied to the traffic jam, so if the road are congested, the transport fee would be higher

Answer Slide

- What are the existing pain points with digital ride-sharing services?

Currently the pain point would be it would be a little uncomfortable if the the passengers didn't fit with the other passenger attitude, it would make the journey uncomfortable. Also ride-sharing services transport time tends to be longer since we have to pick up and drop off another passengers

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

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

Having a flying taxi would decrease the traffic jam trouble since air would be less congested. More fleet can be deployed and more service can be enjoyed by our market

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

Since the capability is not on the land transportation, but it also involves air transportation, relatively short distance inter-island transportation would be benefited by flying taxi.

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

- How many records are in the dataset : 145,864 records
- What does each record represent? Each records represents one journey of a taxi ride
- What is the primary key? Primary key is ID, I would assume this is the ride ID
- What date range is your dataset bound to? The data set is bound to for **pickup time** (from 2016-01-01T00:04:00.000Z to 2016-06-30T23:58:00.000Z) while for **dropoff time**, its (from 2016-01-01T00:12:00.000Z to 2016-07-01T23:02:00.000Z)
- 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? Most data points are **centralized** at New York. ,Brooklyn, and Queens, **It does not include** Staten Island, New jersey, and The bronx

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

Price are derived from:

- Basic Fee 3\$
- 1.56\$ per KM

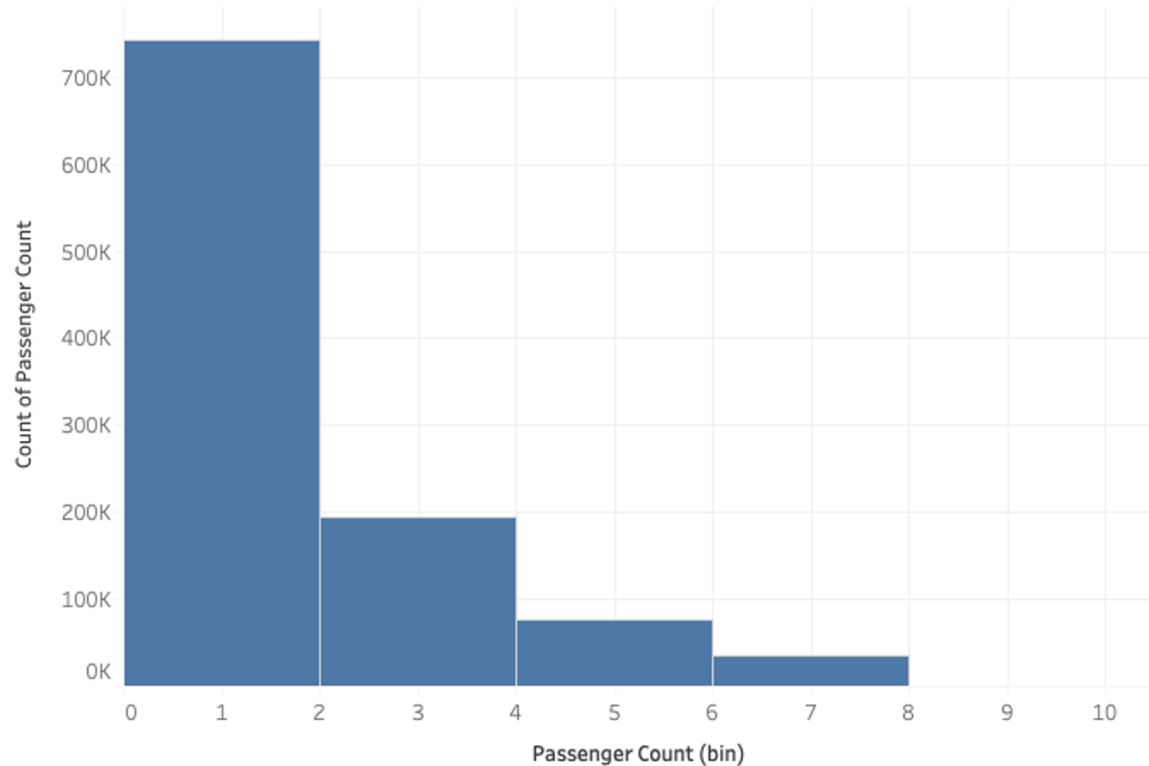
Field	Average	Median	First Std Dev	Second Std Dev
Duration	945.58	662.00	3106.09	6212.18
Distance	3.43	2.09	4.38	8.76
Passenger Counts	1.66	1.00	1.31	2.62
Distance_to_duration Ratio	0.40%	0.36%	0.56%	1.12%
Price	8.35	6.27	6.82	13.63

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

Passenger Count 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?

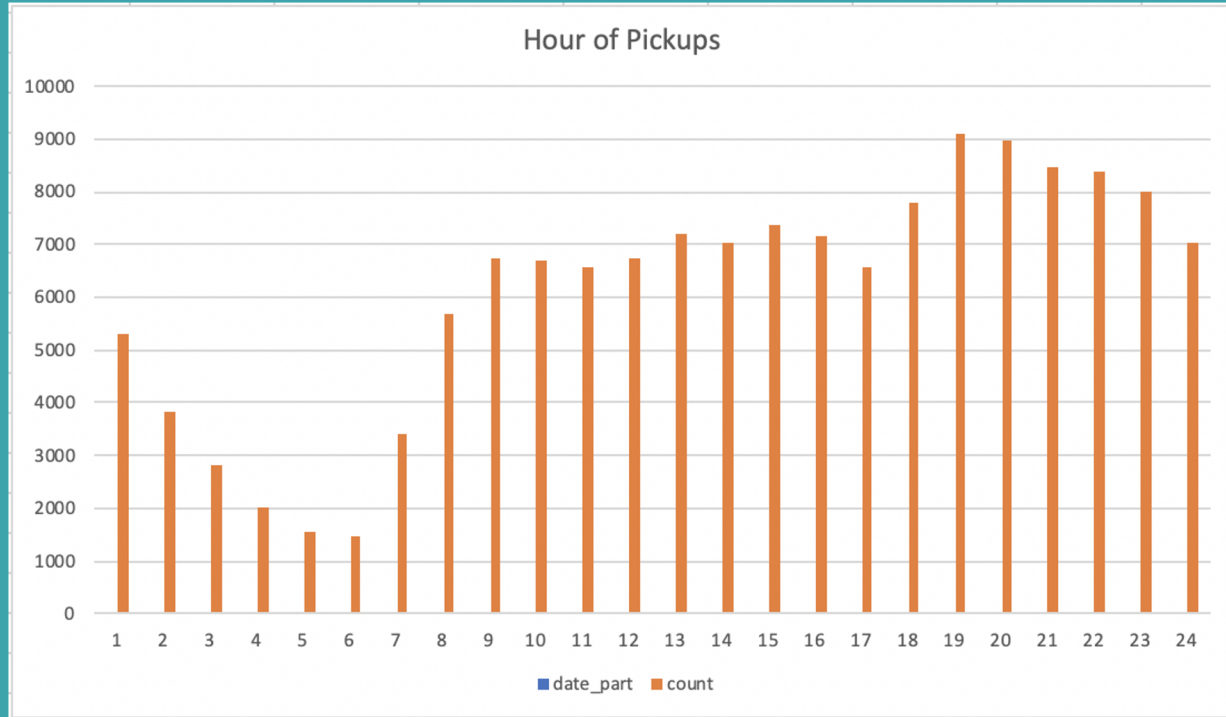
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups? Zip codes 11371, 11430, and whole New York
- Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs? Zip codes 11371, 11430, and whole New York
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up? Zip codes 11371, 11430, and whole New York
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off? Zip codes 11371, 11430, and whole New York
- 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? It's on Zip Codes 11371 and 11430

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

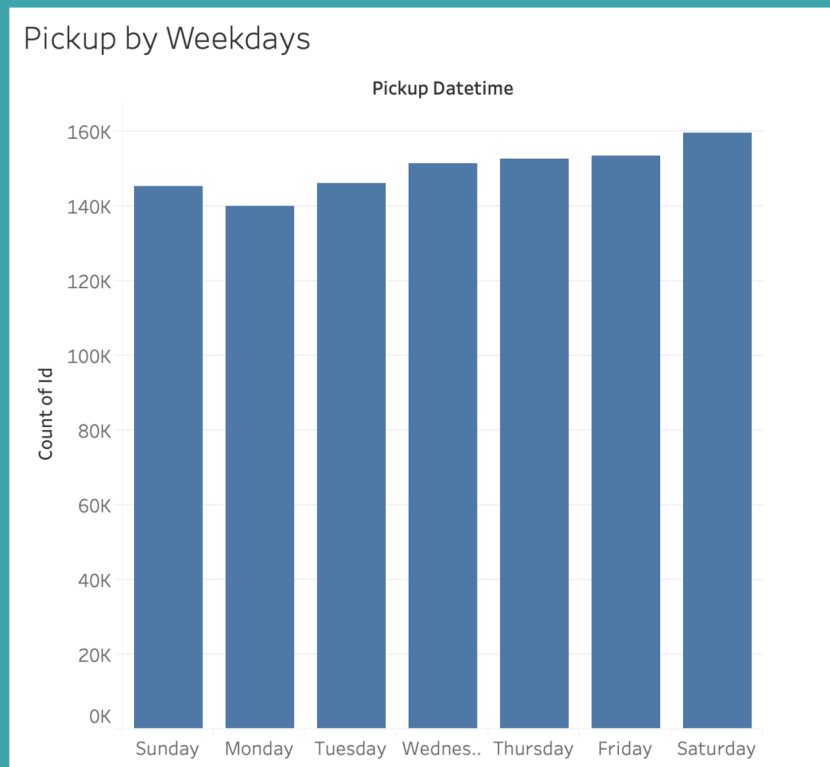
- What times throughout the day experience relatively higher volumes of ride pick-ups? It's generally on evening (18:00-22:00) but for the biggest It's on 18:00 (9086 pickups)
- What days throughout the week experience relatively higher volumes of ride pick-ups? It's saturday (with number of pickup 159,743)
- 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. Actually the rides have more demand in the first half of the year (January - June), started decreasing rapidly on July and being steady until December

Hour of Pickups



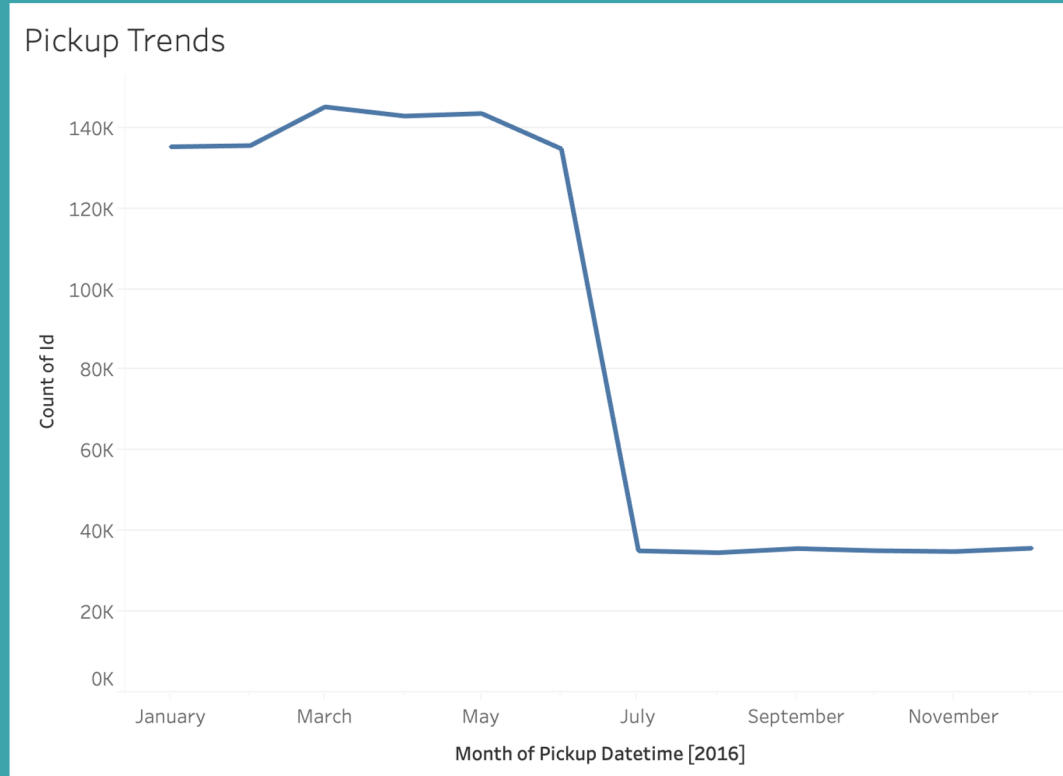
Hour with the most pickup It's generally on evening (18:00-22:00) but for the biggest It's on 19:00 (9086 pickups)

Pickup by Weekdays



On saturday we have relatively higher volume of pickup (with number of pickup 159,743)

Trend of pickup on month level



- Actually the rides have more demand in the first half of the year (January - June), started decreasing rapidly on July and being steady until December

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)

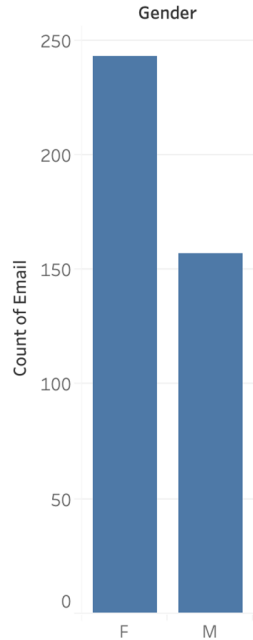
Q10 - If no to Q8, what is the reason?

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?

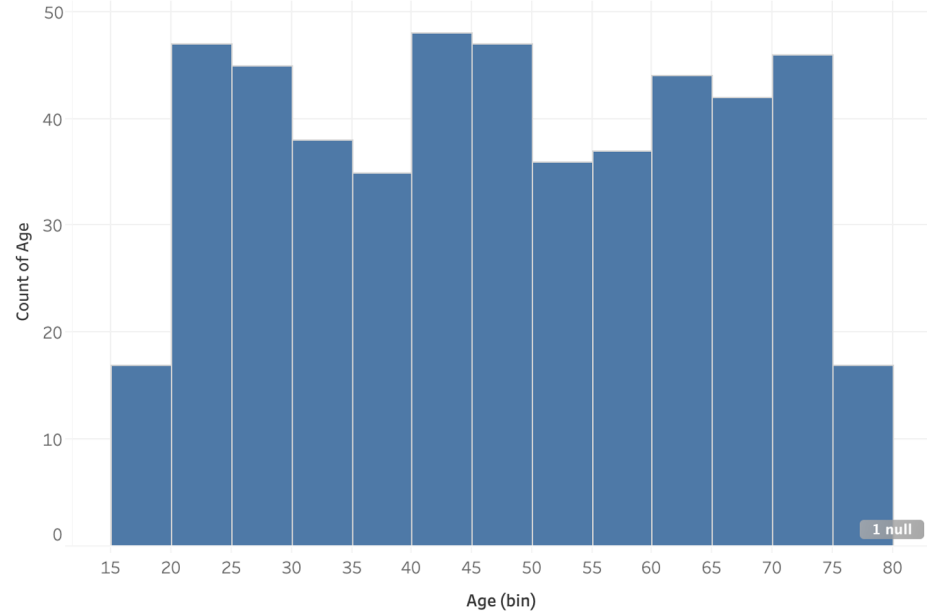
Answer Slide

Interest in Using Flying Taxi Based on Gender



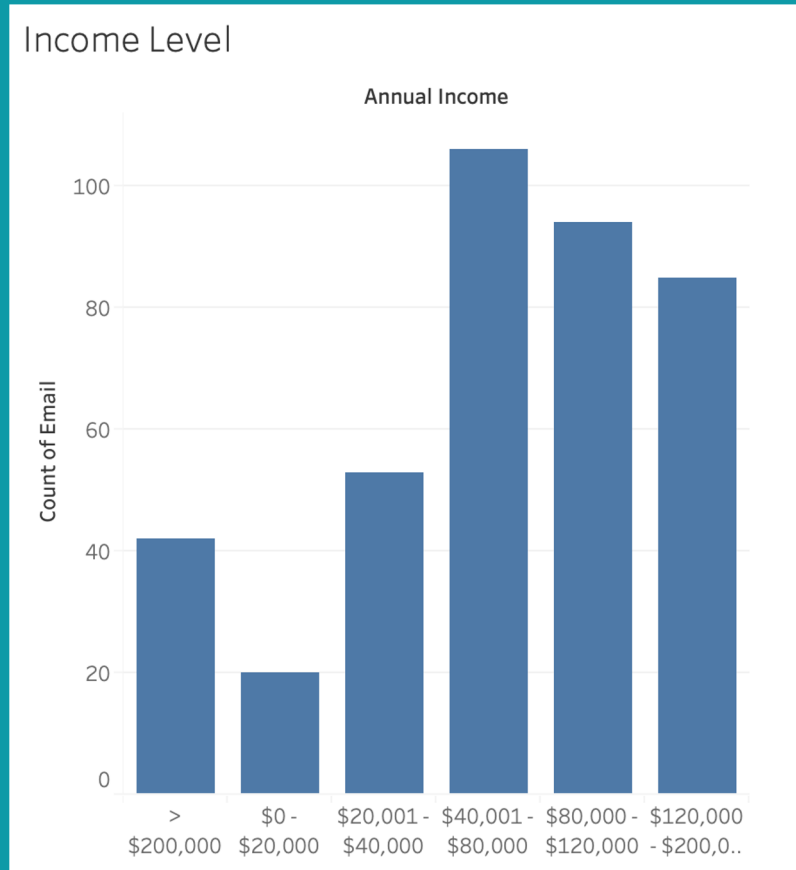
We check the one who said yes when asked interested in flying taxi or not, turns out that Female is more interested than Male

Age Range Histogram



This is histogram with 5 age range, Taken from the one who are interested on flying taxis, Age range 20-25, 40-50, and also 70-75 are interested in that

Income Level

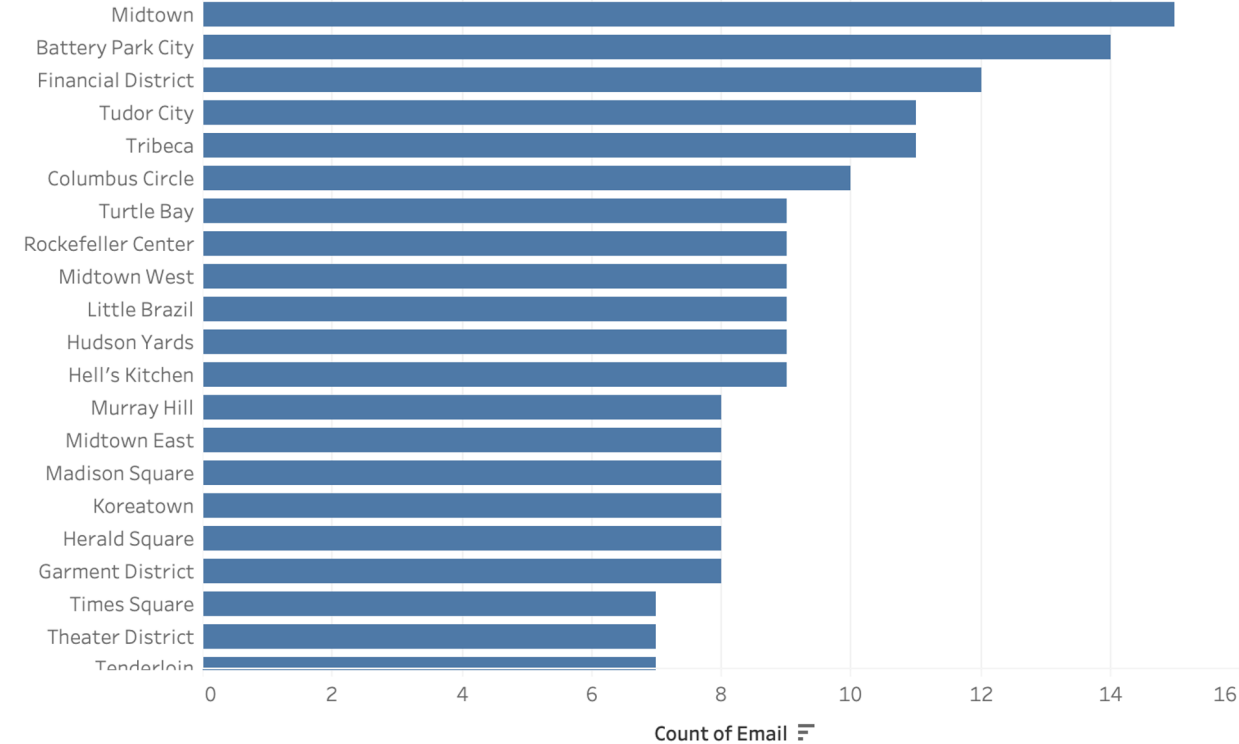


The most people who are interested in flying taxis are on the income level \$ 40,001-\$80,000

Neighbourhood

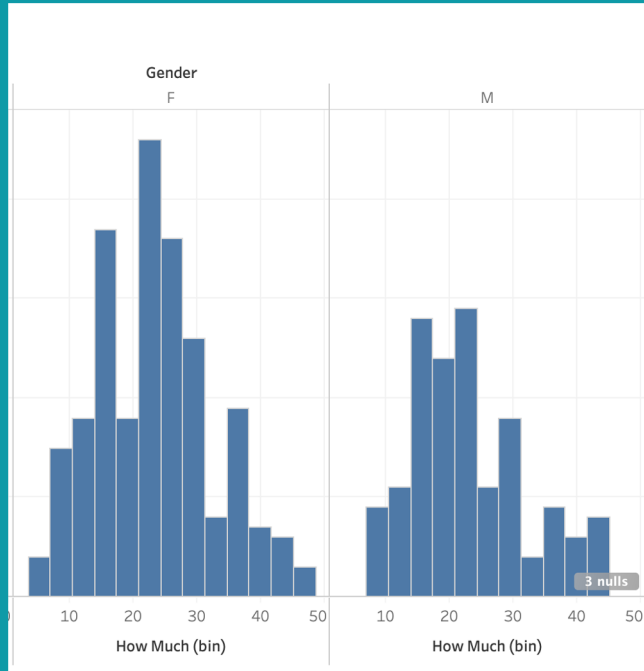
Neighbourhood

Neighborhood



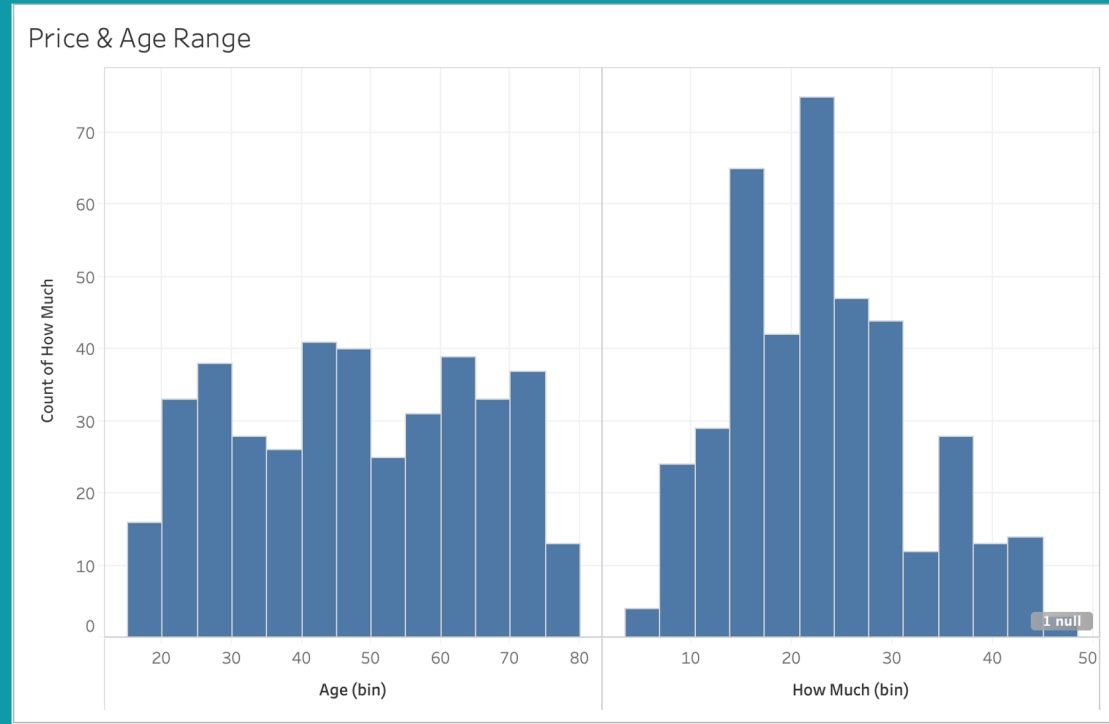
We can see here that top 5 neighbourhood which are interested in flying taxis are Midtown, Battery Park City, Financial District, Tudor City, and Tribeca

Price Per Mile Based on Gender



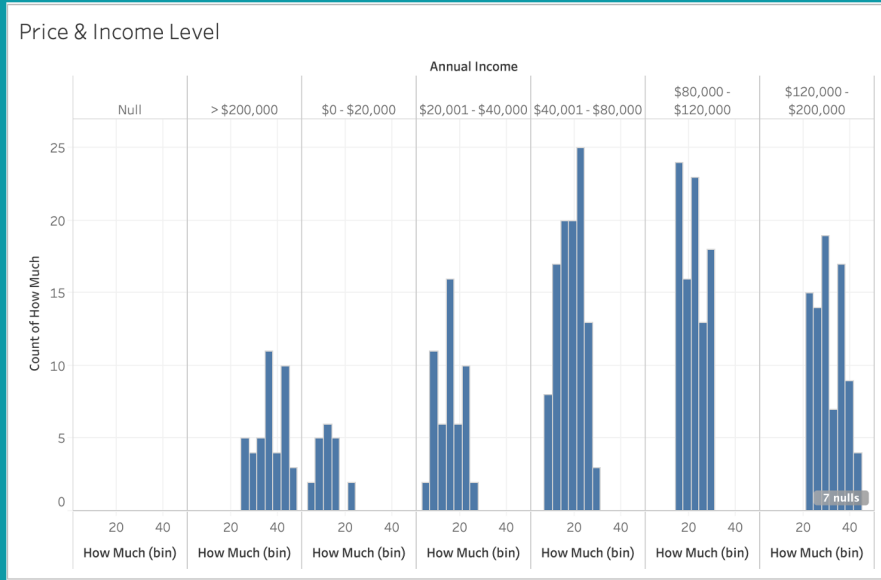
We can see that The distribution is pretty similar between Female and Male. It's just Effected by the total amount of Female and Male only

Price Per Mile Based on Age Range



For the distribution of price per mile based on Age range, for price per mile it's pretty higher up in the price range of 20\$

Price Per Mile Based on Income Range



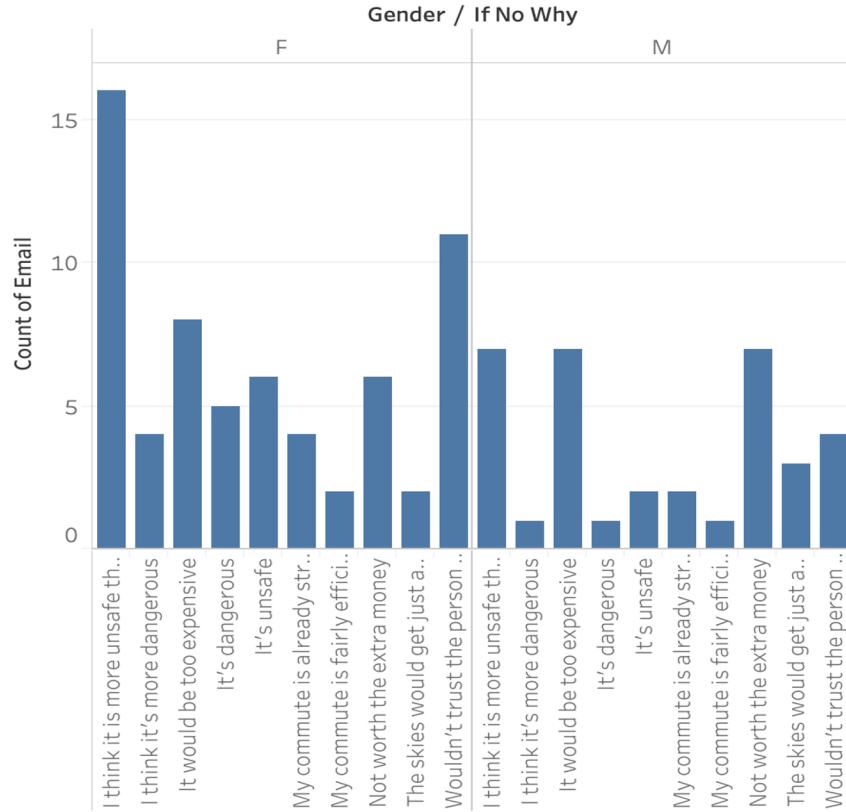
The distribution for price per mile are higher in annual income between \$40,001-\$80,000

Price Per Mile Based on Neighbourhood



The distribution for price per mile per neighbourhood are pretty same, except for upper manhattan, which are pretty high in 20\$ mark

Personas for Negative Segment



For Customer which have negative sentiment, when we're talking about the unsafety of flying taxis, Female are more concern about that then men

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)