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

- What are taxis used for?

Taxis are private vehicles rented for a short periods of time to provide transportation to desired destination for a fee.

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

Users have disposable income to pay a premium for the reduced transit time and personal comfort in a private ride as compared to communal modes of transportation like buses and subways.

- What are existing pain points with taxis?

With many taxis offering their services for premium fare, roads inevitably becomes congested, more damage to roads requiring repairs and hence increase transit time, road closures and costs for repairs.

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

Availability of hireable car depends on time of day (way less during early hours of the day compared to typical business hours) and costs fluctuate depending on demand (surge pricing). Safety is also not issue as these drivers are not licensed.

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

Users of flying taxi service may benefit from the better transit time as traffic jams and road closures might be avoided and distance travelled reduced as a more direct path can be journeyed.

Existing taxi service industry would expand to the skies which will require rigorous regulations and licensing from the FAA, limiting the entry of freelance ride-sharing competitors who had disrupted the traditional taxi service, thereby giving them more control and dominance in this particular mode of transport.

With fewer taxis and even ride-sharing private cars on the road as more travelers choose the flying taxis for speed and convenience, roads should become less congested and be of better condition, requiring fewer repairs and road closures. If road usage dropped to an significantly low level, sections of roads can be converted to pedestrian only, giving room for more shops and parks, encouraging more foot traffic and commerce.

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

1,048,468

- What does each record represent?

Each record represent a taxi ride with taxi vendor id, pickup and dropoff times and locations, number of passengers, trip duration and distance.

- What is the primary key?

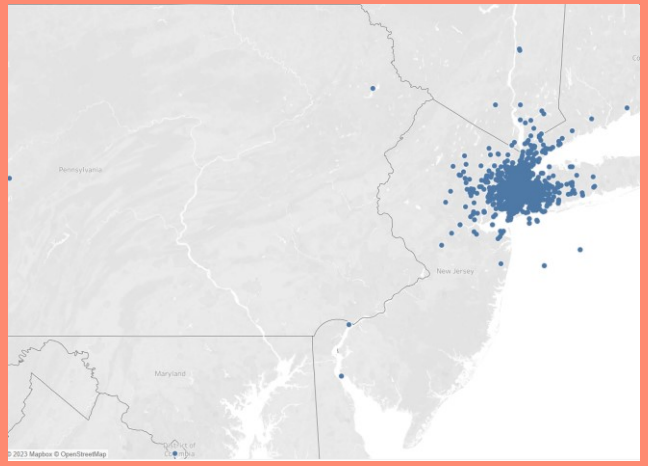
Primary key is the Id for each trip.

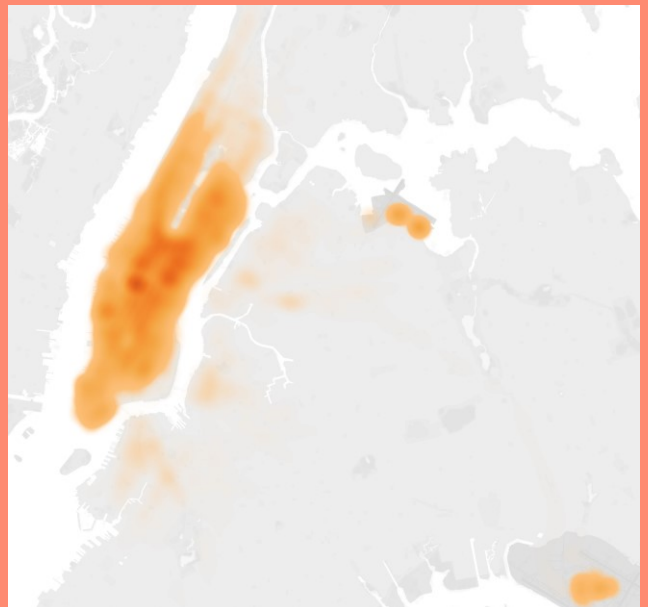
- What date range is your dataset bound to?

Data ranged from January 1 2016 to June 30 2016

- 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?

Centralized to Manhattan and fewer in Brooklyn, Bronx and Queens with the exception of JFK and LGA airports. There's also outliers in PA, NJ, DC and even 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

Duration in minutes

STDEV([Duration])*2	11,706.59
Avg. Duration in minutes	16.04
Median Duration in minutes	11.03
Std. dev. of Duration in minutes	97.55
STDEV([Duration in minutes])*2	195.11

Distance in miles

Avg. Distance	3.44
Median Distance	2.09
Std. dev. of Distance	4.38
STDEV([Distance])*2	8.76

Passenger Count

Avg. Passenger Count	1.66
Median Passenger Count	1.00
Std. dev. of Passenger Count	1.31
STDEV([Passenger Count])*2	2.63

Duration-to-Distance in minutes to miles

Avg. duration-to-distance ratio	78.11
Median duration-to-distance ratio	4.68
Std. dev. of duration-to-distance ratio	15,406.22
STDEV([duration-to-distance ratio])*2	30,812.45

Price in USD

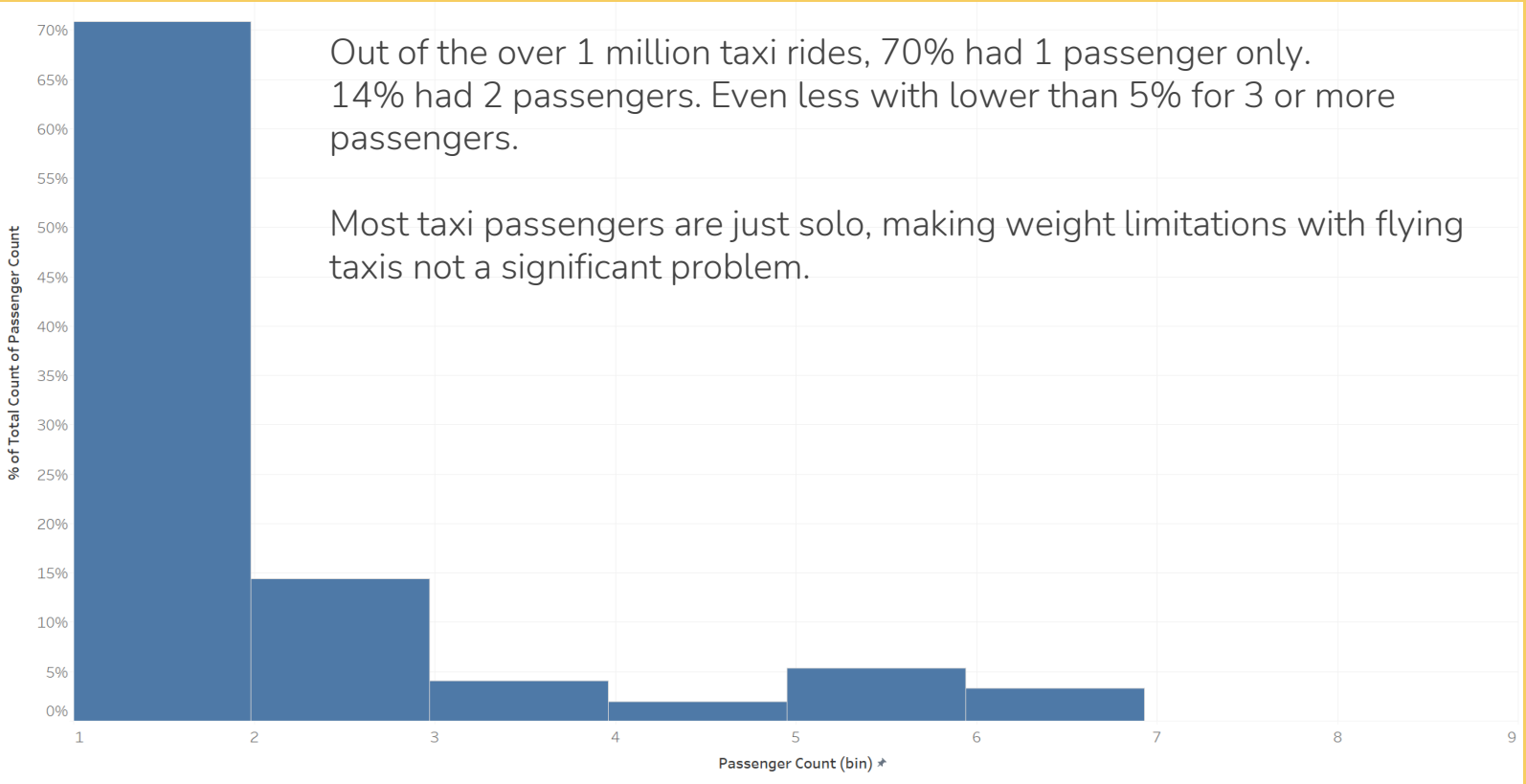
Avg. Price	18.98
Median Price	12.54
Std. dev. of Price	55.50
STDEV([Price])*2	111.00

Price estimated according to https://www.gobytaxi.com/north-america/united-states/new-york/new-york-city
Base fee = $4.30, Price per miles = $2.99, Price per hour = $32.50, no surcharge with more passengers.

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



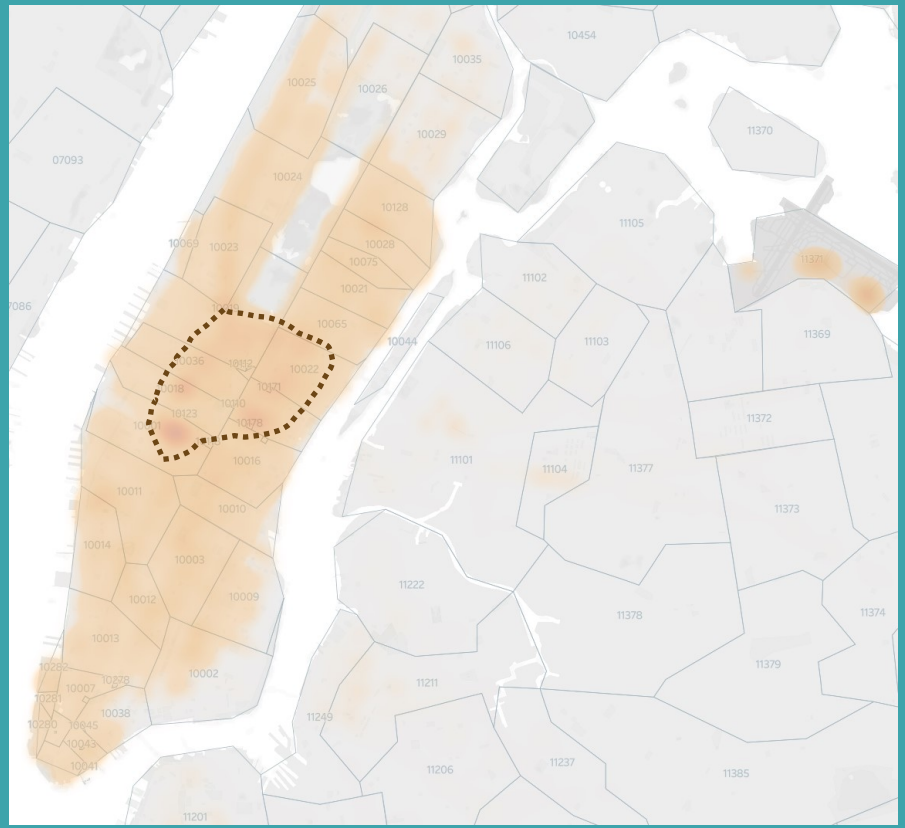
Out of the over 1 million taxi rides, 70% had 1 passenger only. 14% had 2 passengers. Even less with lower than 5% for 3 or more passengers.

Most taxi passengers are just solo, making weight limitations with flying taxis not a significant problem.

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?

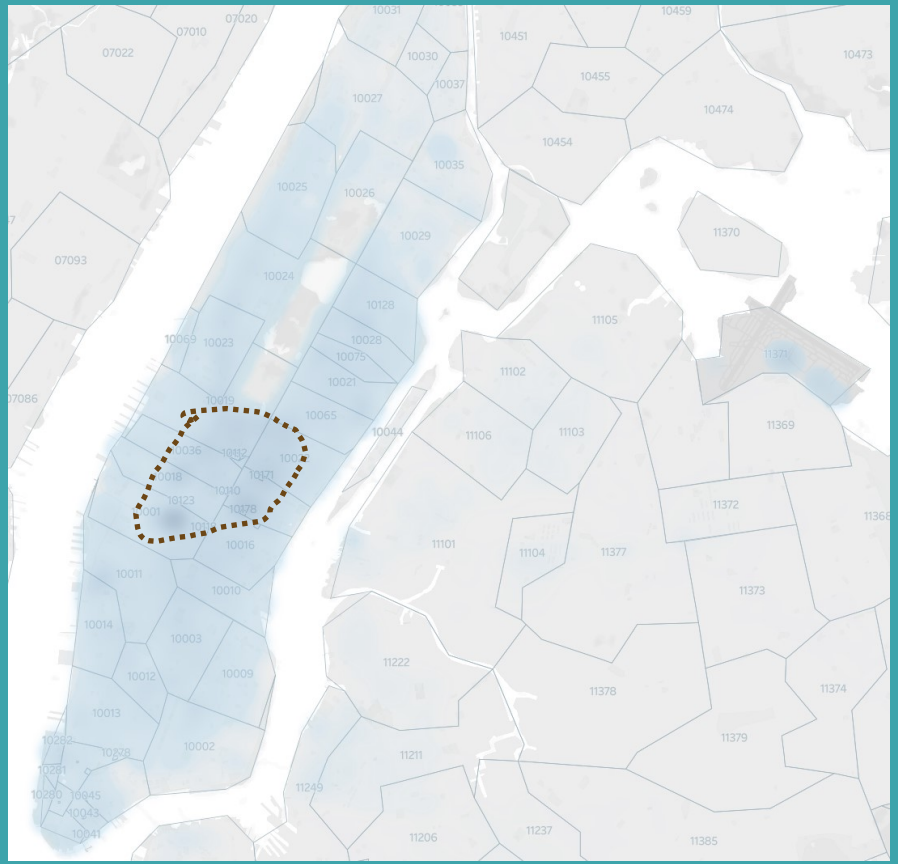
Answer Slide



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

Zip codes 10001, 10018, 10178, 10036, 10022, 10110.
Along with LGA and JFK airports.

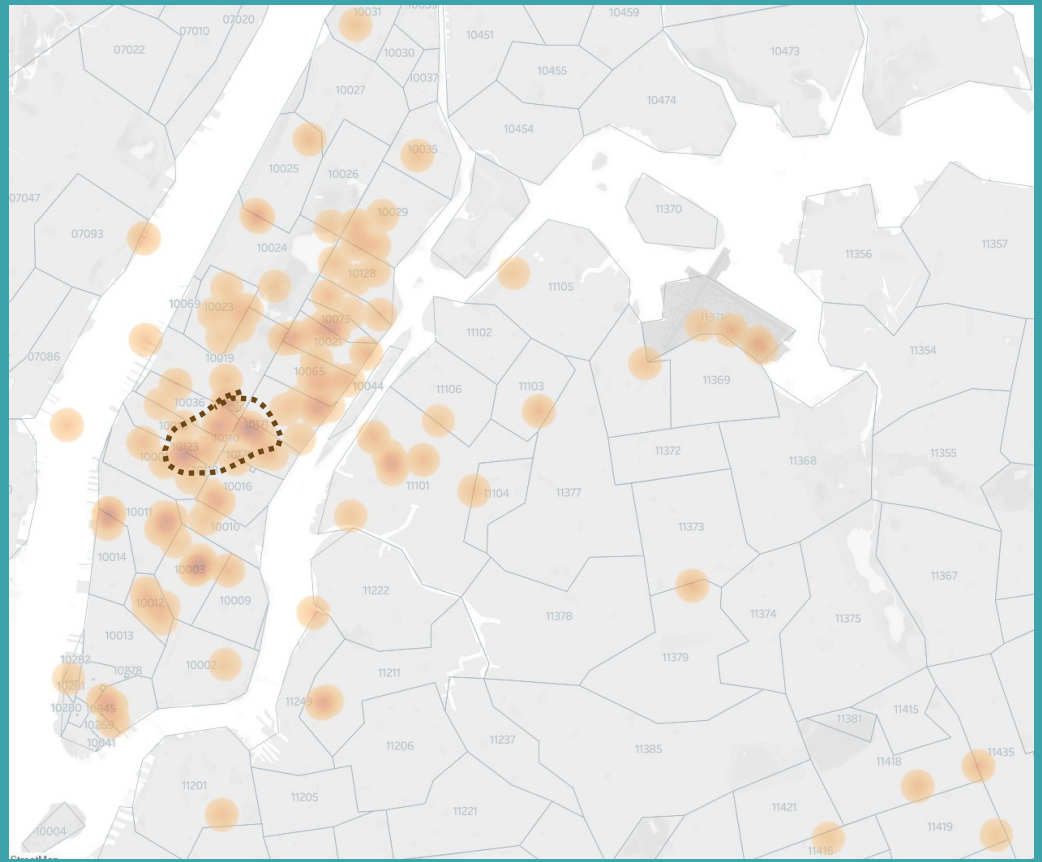
Answer Slide



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

Similar to pickup, zip codes 10001, 10018, 10178, 10036, 10022, 10110.
Along with LGA and JFK airports.

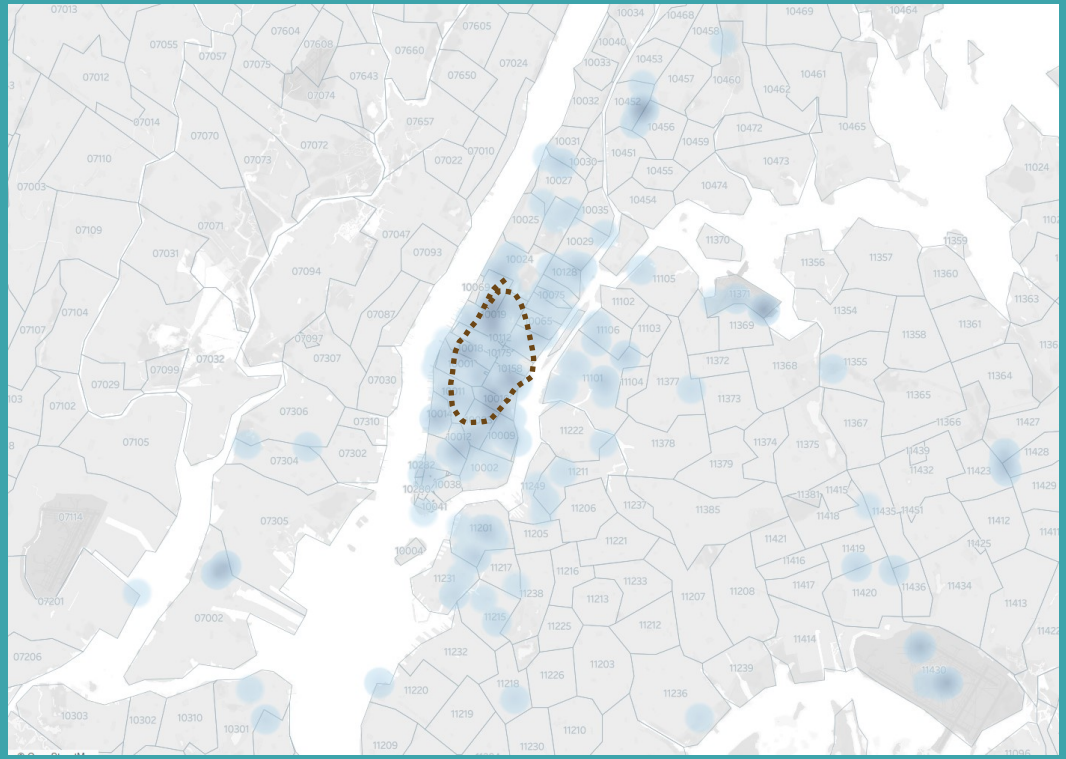
Answer Slide



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

Outliers in PA, NJ, DC and in ocean are removed.
Limiting the duration-to-distance ratios to between 50 and 100k, majority are concentrated in zipcodes 10171, 10123, 10003, 10110.
Along with LGA and JFK airports.

Answer Slide



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

Outliers in PA, NJ, DC and in ocean are removed.
Limiting the duration-to-distance ratios to between 100k and 500k, majority are concentrated in zipcodes 10019, 10158, 10010, 10018.
Also LGA and JFK airports too.

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?

Common areas for both pickup and dropoff are 10001, 10018, 10178, 10036, 10110, as well as the JFK and LGA airports.

These areas have the most demand in middle of Manhattan and airports and spend longer time to commute relative to distance.

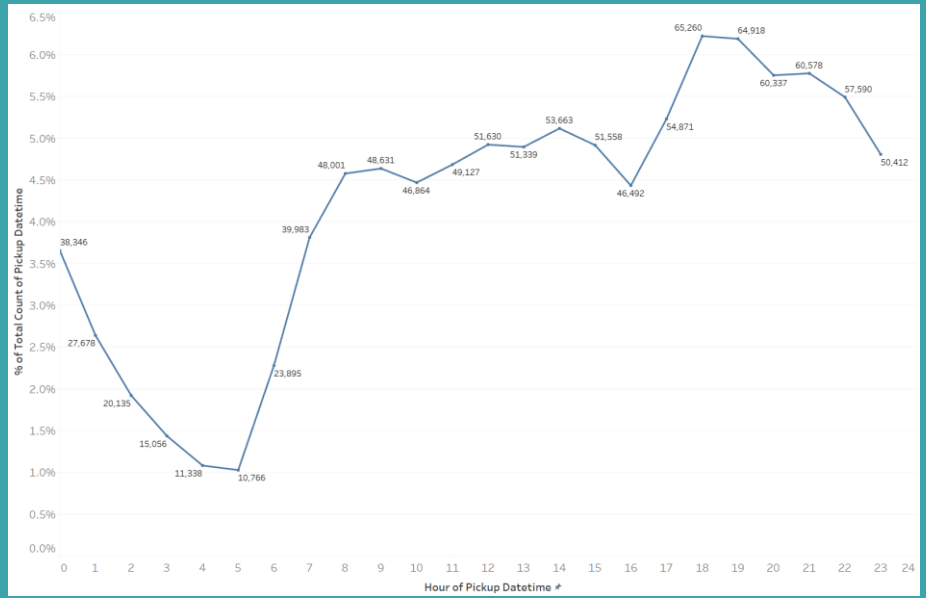
It may not make operational sense to have the service running 24/7, for now.

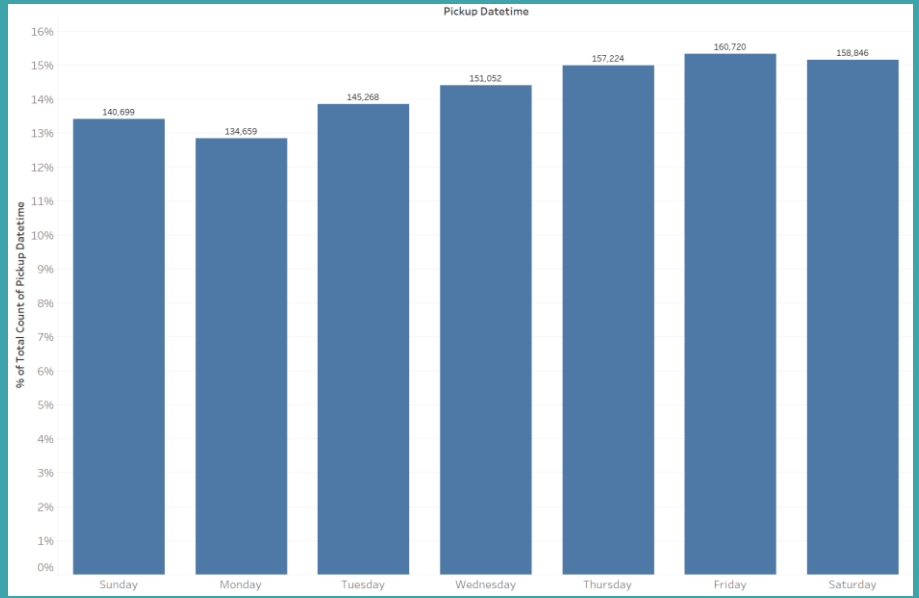
- What times throughout the day experience relatively higher volumes of ride pick-ups?
- What days throughout the week experience relatively higher volumes of ride pick-ups?
- Pinpoint any periods throughout the year that experience trend fluctuation or seasonality around ride pick-up volumes. This will 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

Higher volumes after 5pm to 10pm. Drops significantly from 11pm to5am after which it start picking up as people start going to places for the new day.

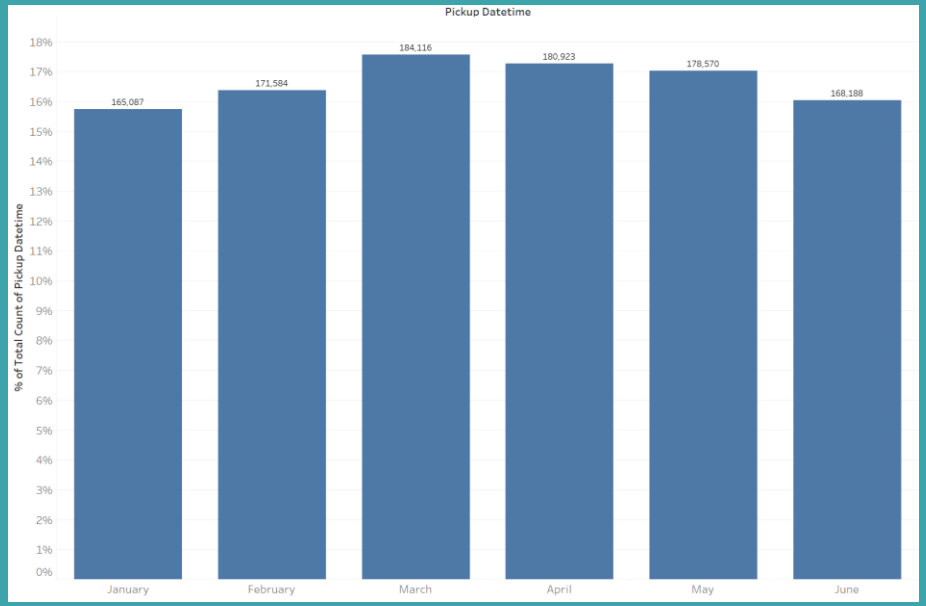
Highest on Friday and Saturday as people are out and about trying to get to their destinations. Lowest on Monday and Sunday as people tend to be home, although the difference in ride volume isn't large.



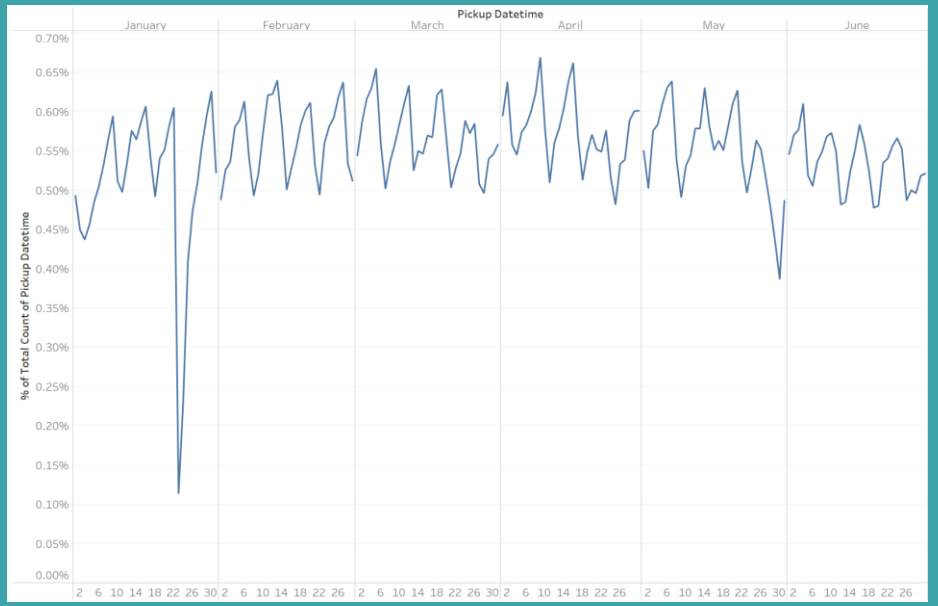


Answer Slide

No large fluctuations in ride volumes from Jan to Jun 2016. Highest was in March and lowest in Jan.



Daily fluctuations per month cycles with lowest for Monday, raising each day to peak on Friday before dropping again. Note the sharp drop on Jan 23 2016 due to travel ban in NYC due to severe weather. May 30 2016 is a public holiday Memorial day, businesses are mostly closed.



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?

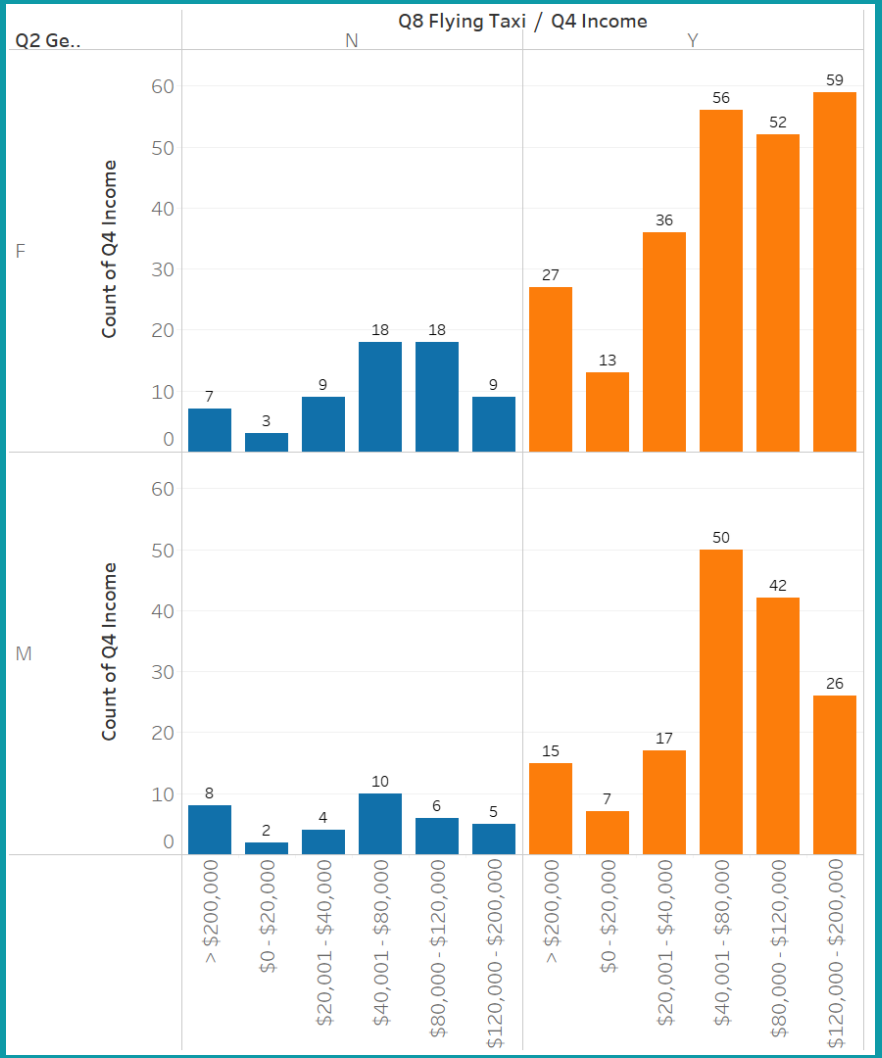
To inform our future product marketing efforts, we'll want to extract the following:

- 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?

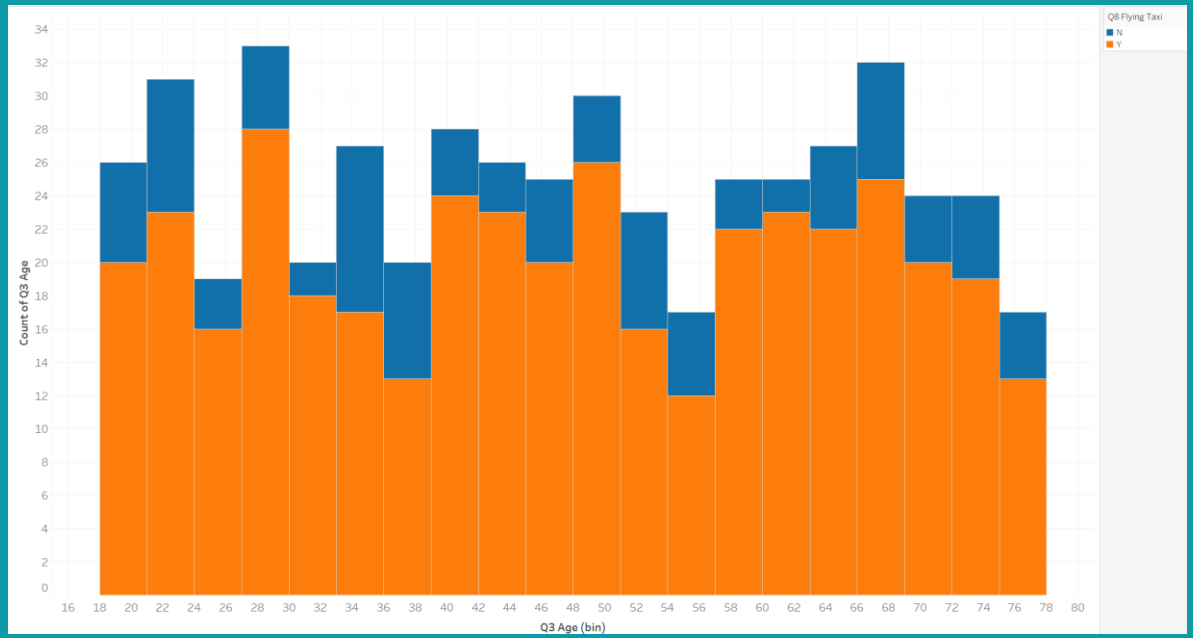
Answer Slide

- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?

Most users surveyed favor use of flying taxis.
Overall users who would use flying taxis had higher incomes for both genders.
Users with incomes between $40,001 and $200,000 favor the use of flying taxis.



Answer Slide

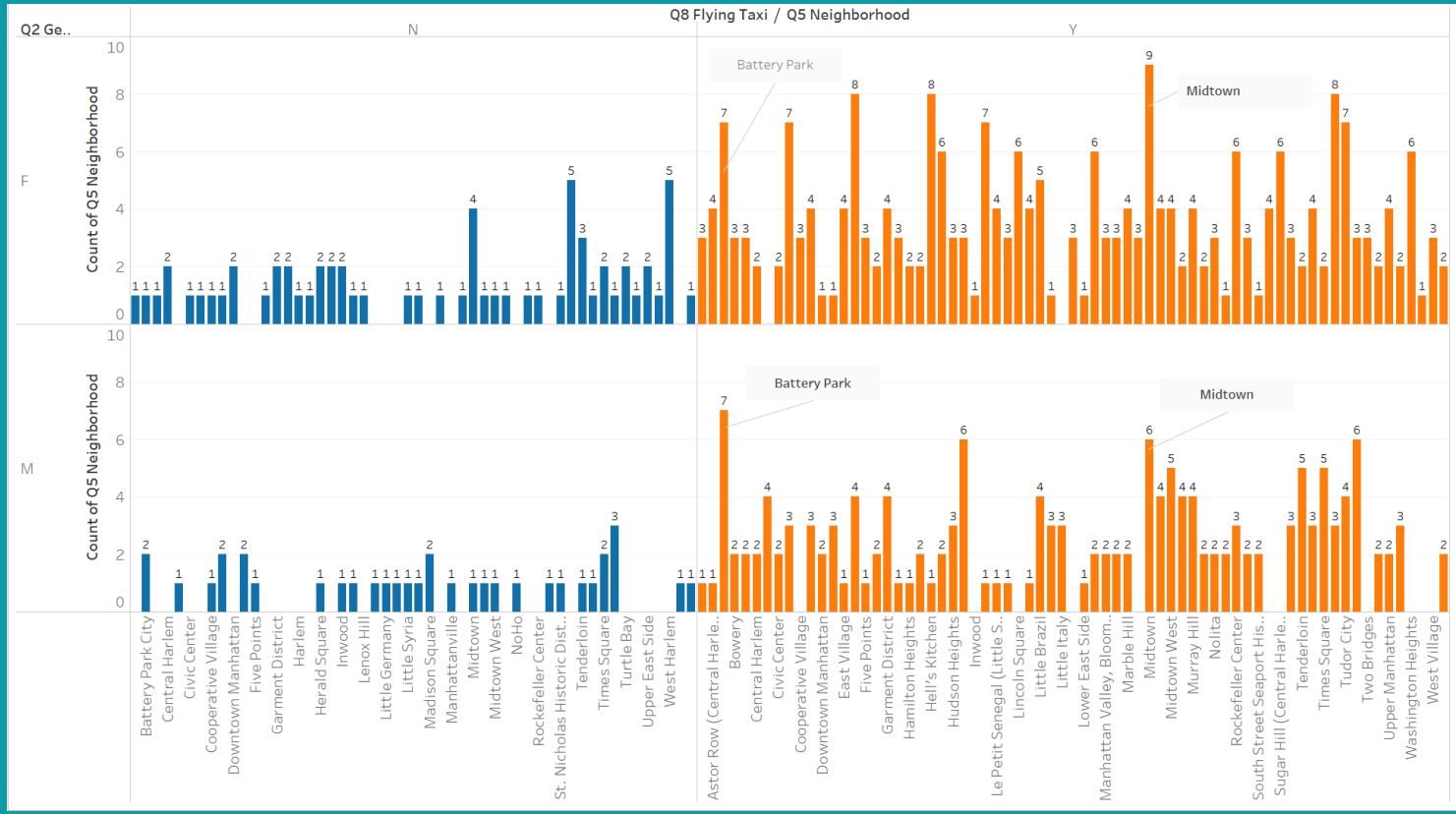


- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?

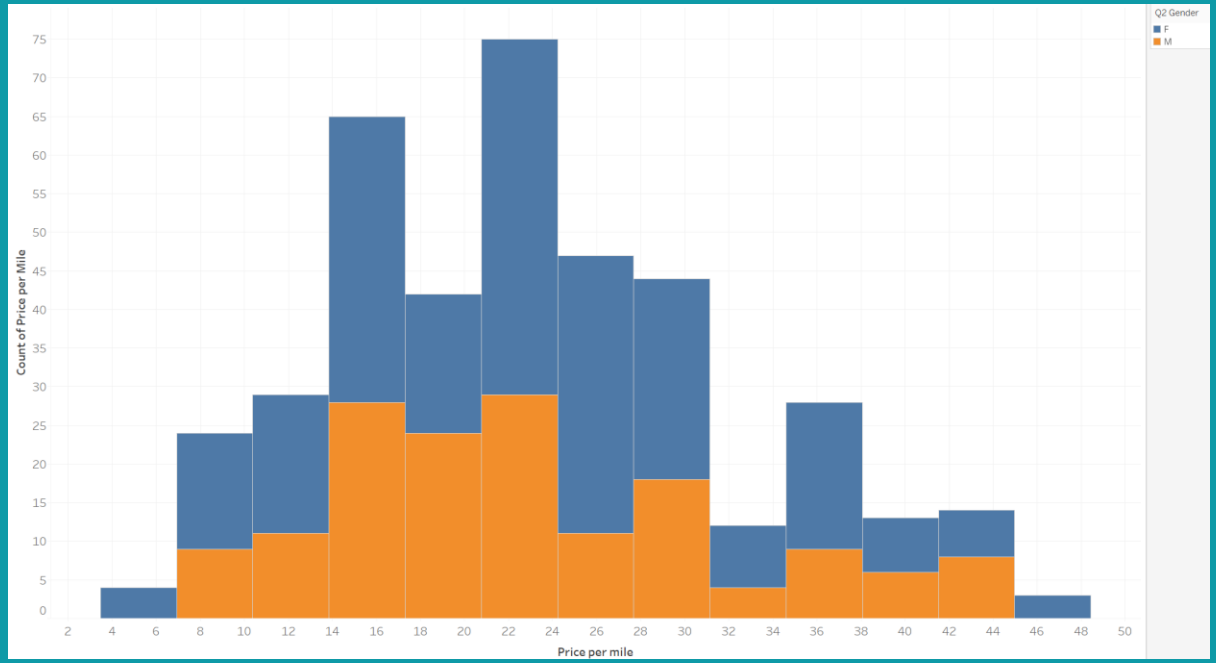
Age distributions of users with favor and disfavor flying taxis are about the same. No age disparities.

Answer Slide

Both genders in Battery Park and Midtown highly favor flying taxis.



Answer Slide



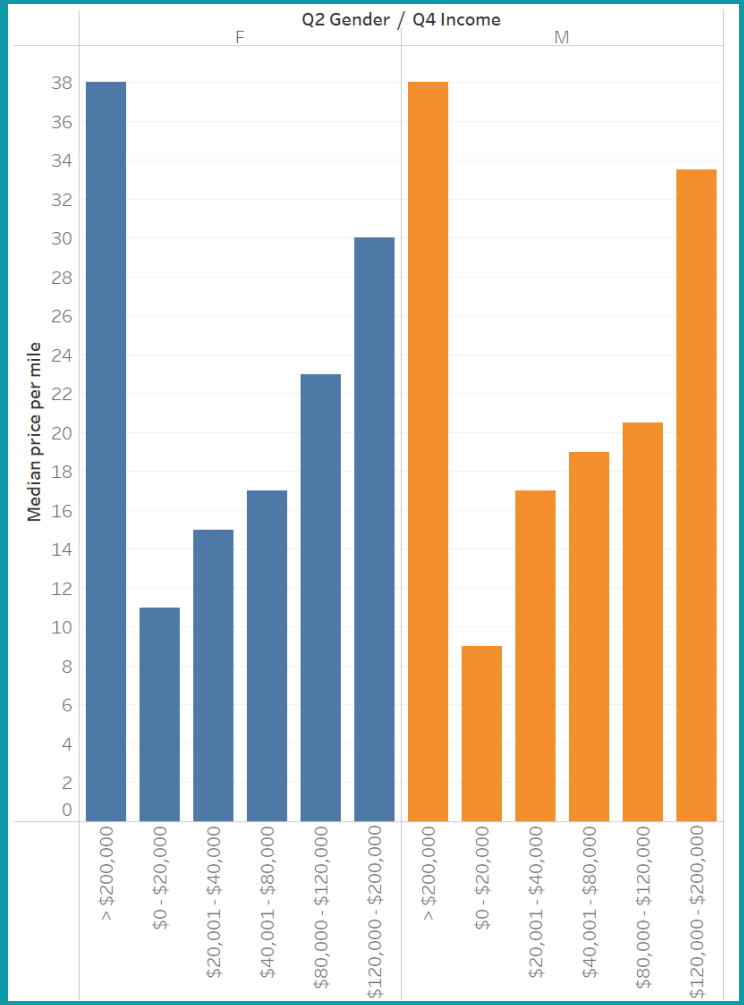
- What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?

Potential price per mile distributions of users gender are about the same. No gender disparities.

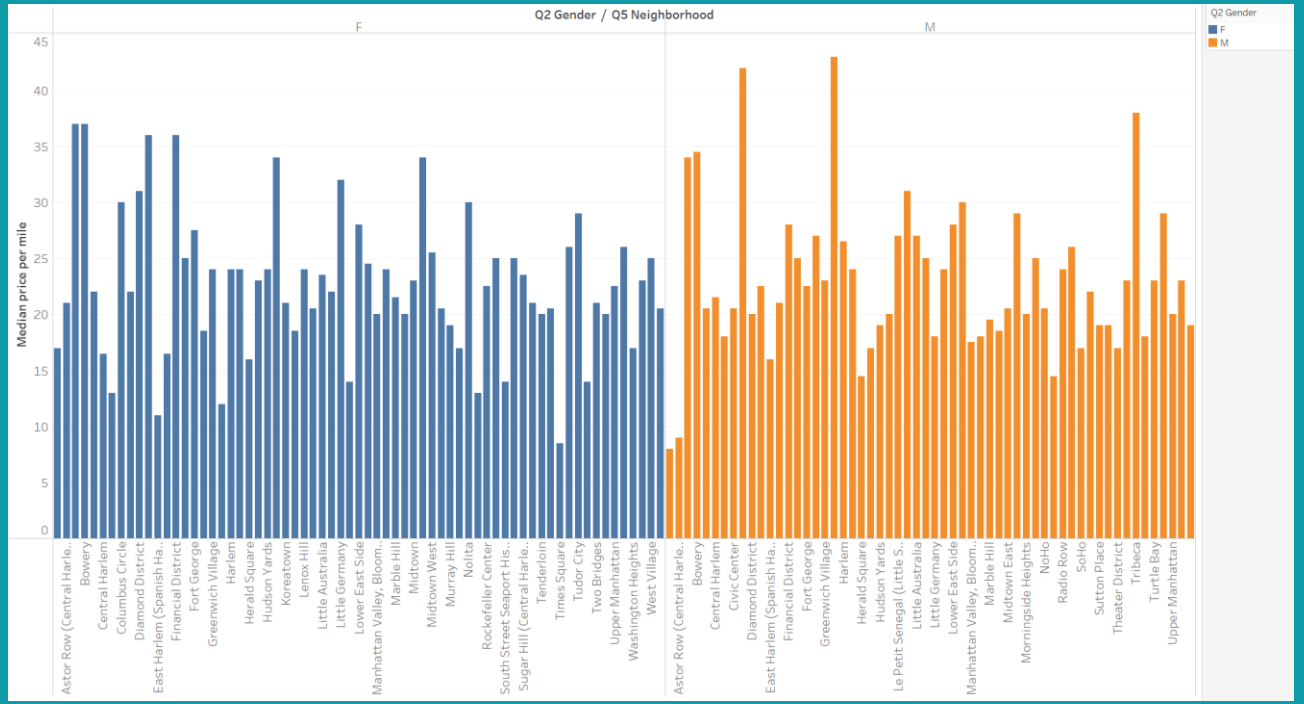
Answer Slide

Median potential price per mile increases with raising income for both genders.

Folks with higher income are willing to pay more for flying taxi fares.



Answer Slide

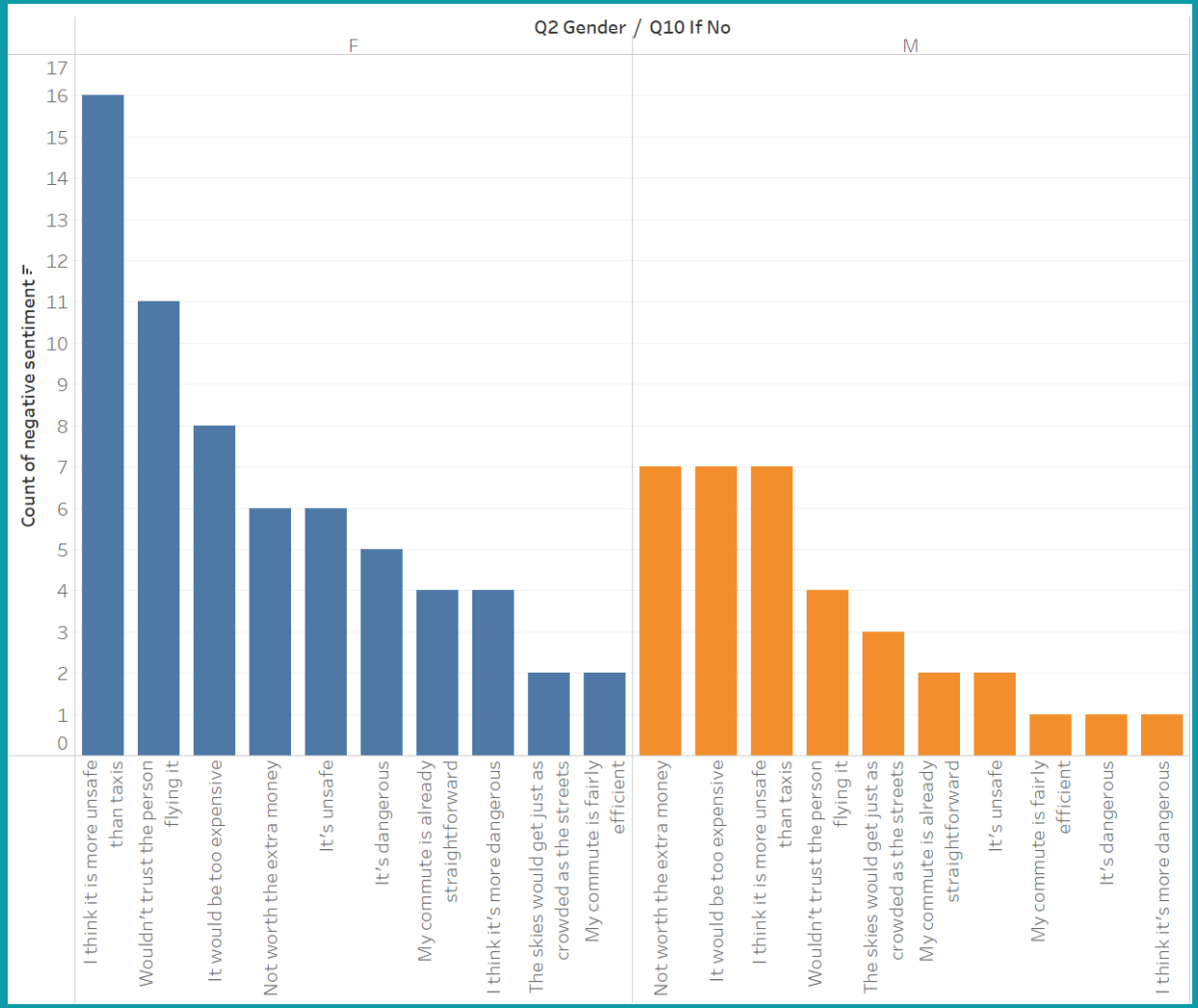


Median price per mile for female users consistently peaked at $35 to $40 for users in Battery Park, Bowery, Downtown and Financial District. For male users, the top median prices are $40 to $45 for users in Columbus Circle and Hamlton Heights.

Answer Slide

- What is the different personas/segments of negative sentiment towards not using a flying taxi car service?

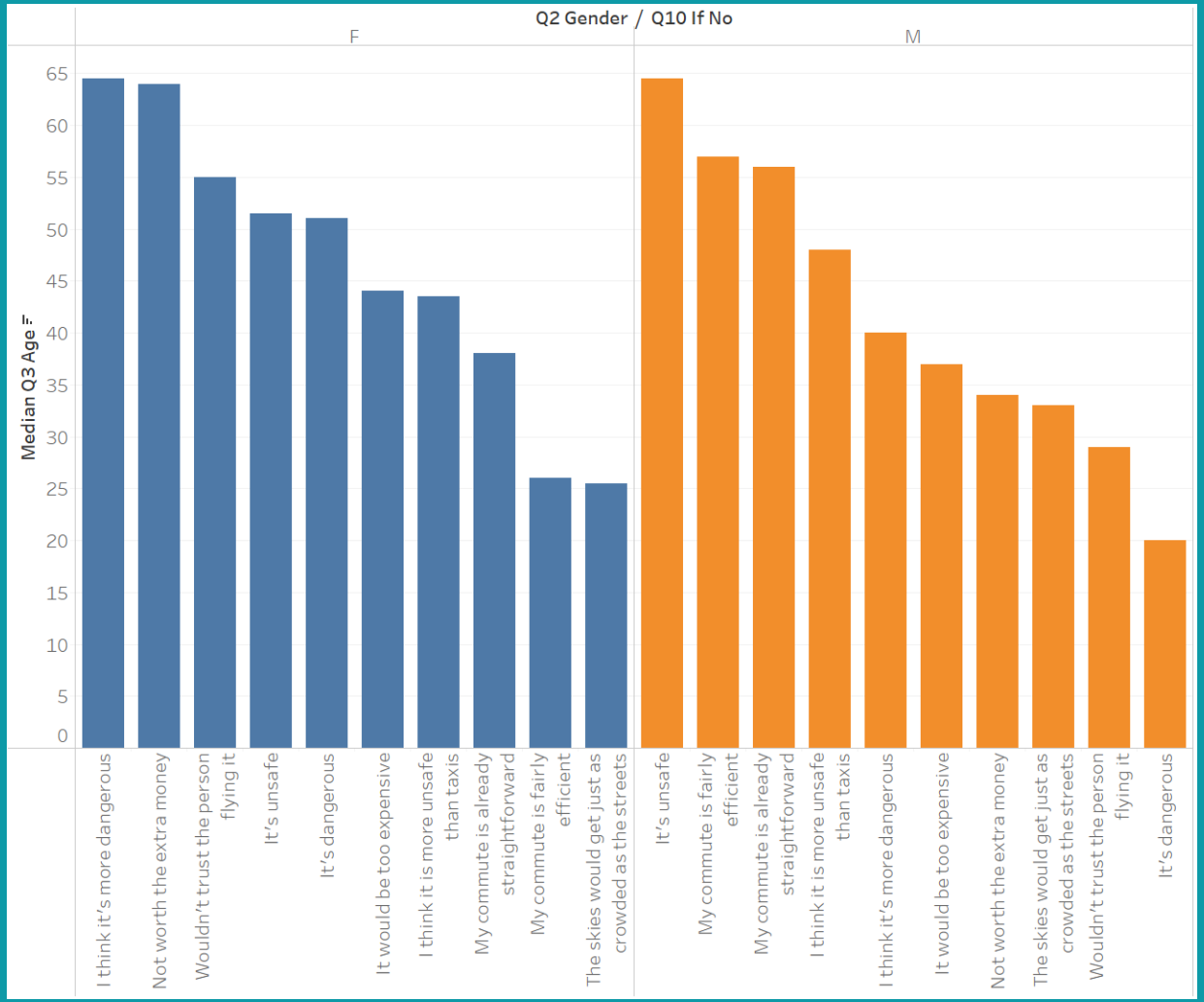
Female users are more concerned about safety than potential expense, while male users are the opposite.



Answer Slide

Older and female users are more concerned about safety and expense, while younger female users care more about commute efficiency and concerns of congestion of the skies.

Older male users worry about safety and questions need for flying taxis while younger males have safety concerns.



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.