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

1. Taxi is used for public transport. People use them to get to where they want to go.
2. Users insensitive to the price, and sensitive to the time, who do not want to drive by her/his own.
3. Taxi service is expensive, sometimes encounters traffic problem, and demand exceeds supply.
4. Uncertainty in the ride-sharing service. The time prediction may not be accurate. Rejection, users may get rejection when they need the service, slow-moving traffic also existed in ride-sharing service.

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

1. User- improvement: More economic and time-saving transportation.
2. Market-improvement: Reduce traffic, require less maintenance on road infrastructure. Taxi-service: encourage existing taxi service to improve their service and decrease price.

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

1. 1048468 records in the dataset.
2. Each entry recorded each taxi trip information, including location and time of dropoff and pickup, duration, distance and number of passengers, and vendor ID and ride id.
3. Primary key - Id.
4. Date range : 2016-01-01 to 2016-07-01.
5. Geographical bounds: Most of the data points centralized in Manhattan, few located in New Jersey and other states. Outliers existed in PA, VA, MA, etc.

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

Avg. Distance	3.4423
Avg. Duration	962.1956
Avg. Passenger Count	1.6644
Avg. Price	22.9488
Avg. duration_to_distance	4,686.8335

Median Distance	2.0946
Median Duration	662.0000
Median Passenger Count	1.0000
Median Price	16.2303
Median duration_to_dist..	280.8214

	1st	2nd
Std. dev. of Distance	4.3821	8.7642
Std. dev. of Duration	5,853.30	11706.5936
Std. dev. of Passenger Count	1.3142	2.6284
Std. dev. of Price	61.0583	122.1166
Std. dev. of duration_to_dista	924,373.47	1848746.941

Units

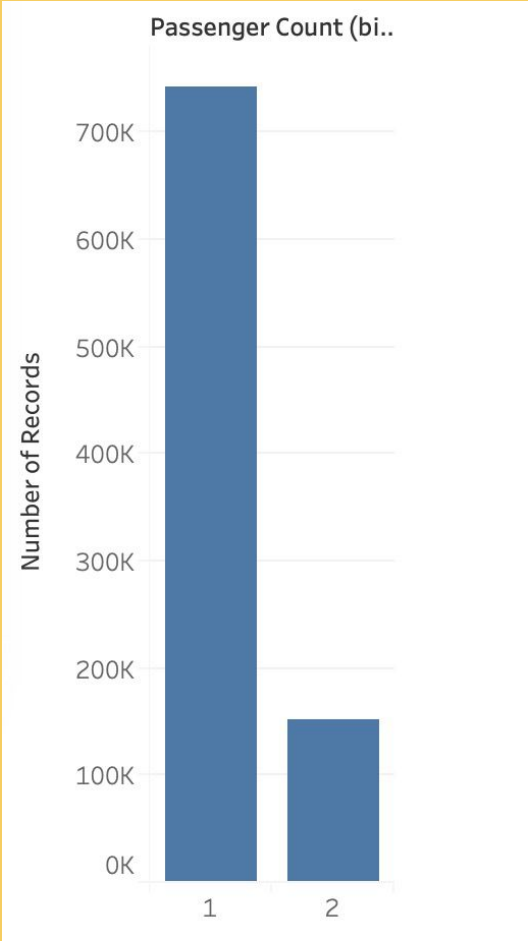
- Duration - seconds
- Distance - miles
- passenger counts - per person
- duration-to-distance ratio - Seconds/Miles
- Price - dollars

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

1 Passenger : 743,056
2 Passengers: 151,060



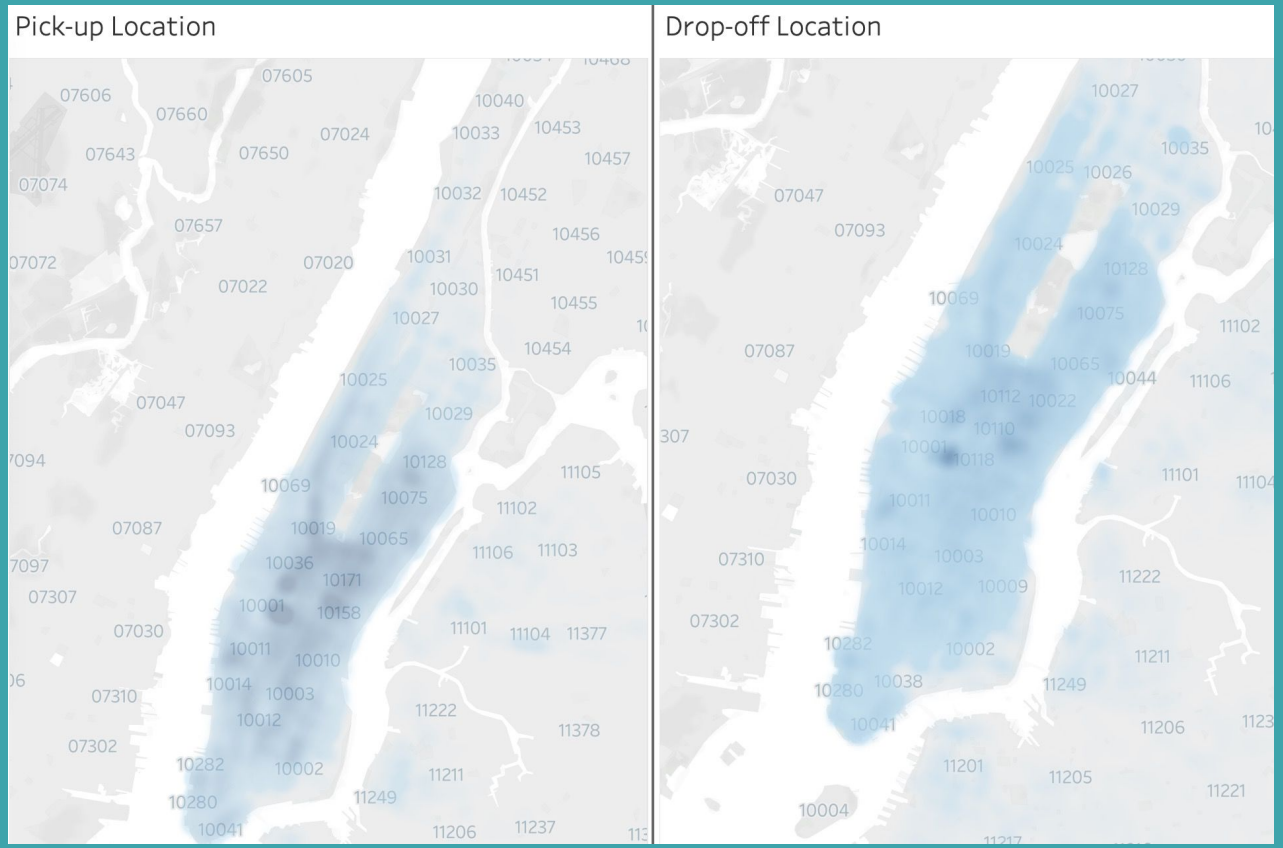
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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?

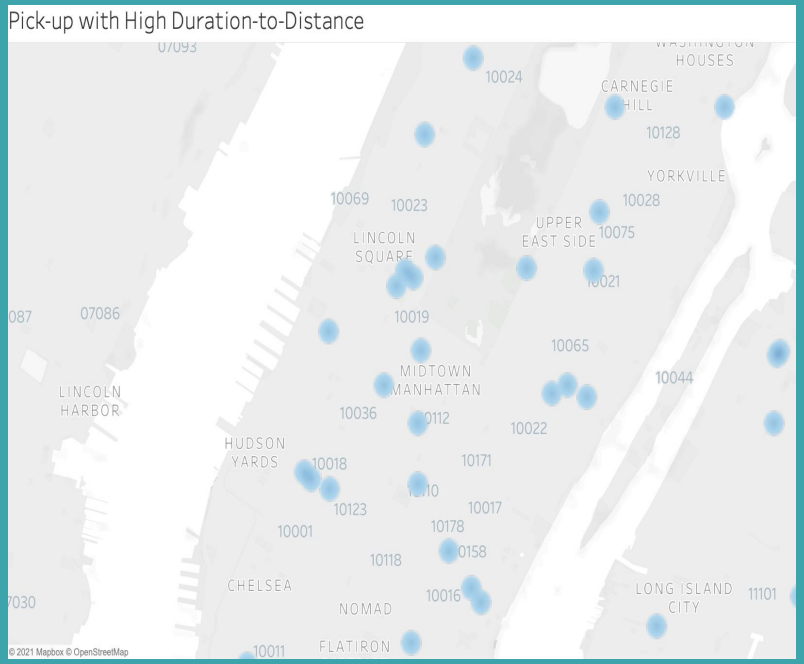
Answer Slide

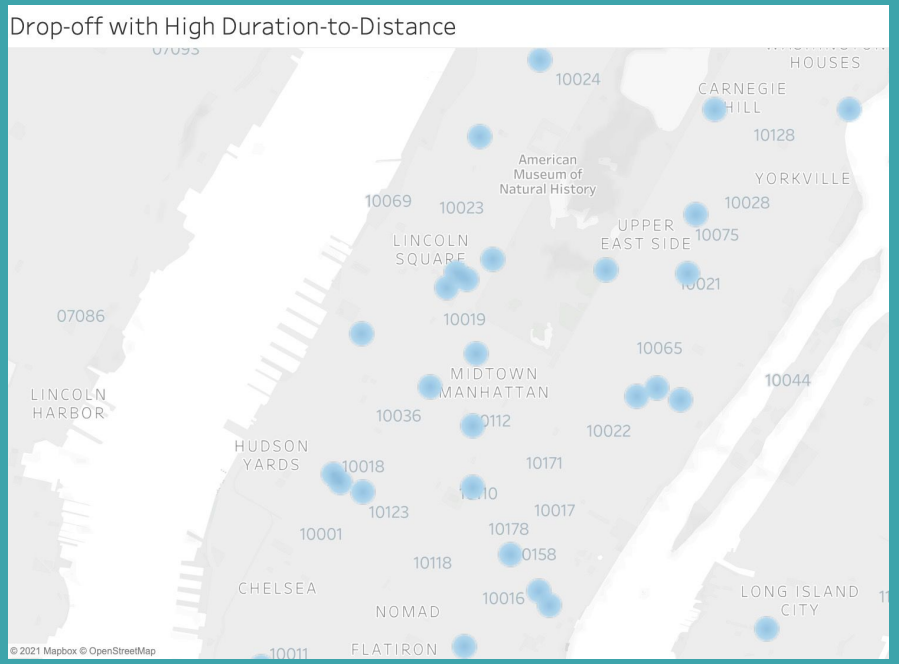
1. Pick-ups: 10001, 10118, around Chelsea neighborhood.
2. Drop-offs: 10001, 10123, around Chelsea, Nomad and Hudson neighborhood.
3. Pick-up Highest ratio: 10019, 10036, around Columbus Circle and Midtown Manhattan.
4. Drop-off Highest ratio: 10019, 10036, around Columbus Circle and Midtown Manhattan.
5. Midtown Manhattan and Columbus circle are potential place for Flying taxi service. Because these areas have lots of rides, and also took long time for a ride, (the duration to distance is high) which shows traffic issue. (See next slide of graphs)

Answer Slide



Answer Slide - Filter out high ratios





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

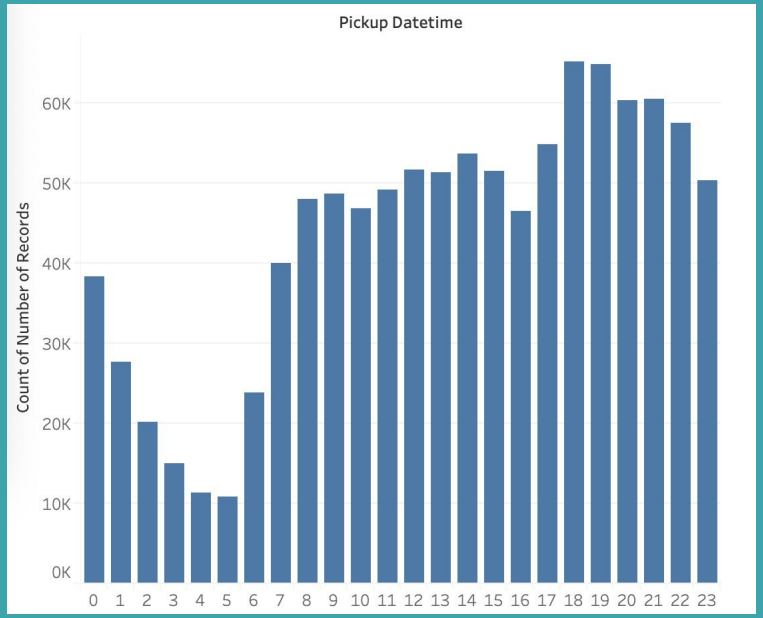
- What times throughout the day experience relatively higher volumes of ride pick-ups?
- What days throughout the week experience relatively higher volumes of ride pick-ups?
- Pinpoint any periods throughout the year that experience trend fluctuation or seasonality around ride pick-up volumes. This will help us in our post-launch analyses to determine if any spikes or dips were influenced by seasonality or through actual feature adoption/regression.

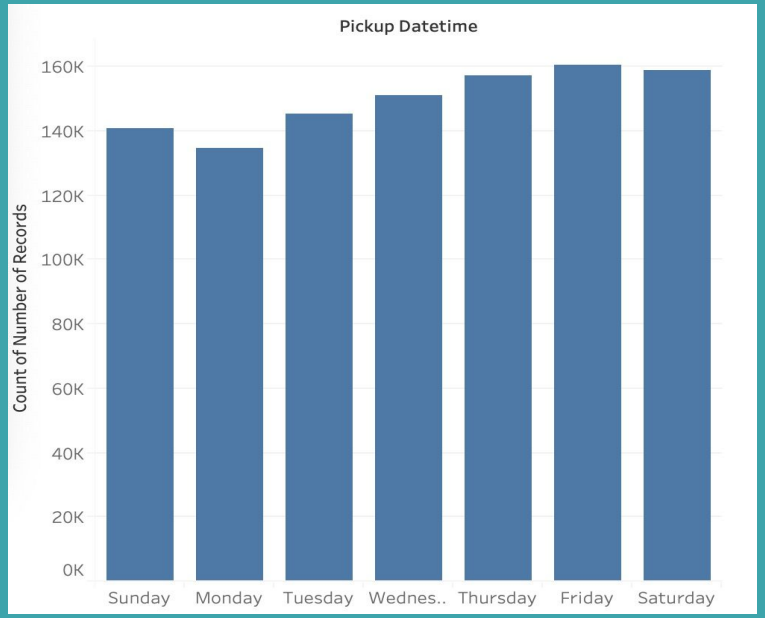
Answer Slide

1, From Histogram, we can see time at 18-21 experiences relatively higher volumes of pick-up.
2. Friday experiences relatively higher volumes of pick-up.
3. We can see that March experiences relatively higher volumes of pick-up. Jan and June experiences relatively lower volumes.

See next slides with Tableau chart

Answer Slide





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

1. Through analysis of Q8, we can see that differences of percentage of answered ‘Yes’ in gender is not obvious. But Age under 30 showed inclined to Flyber service than other ages. Income larger than 200,000 showed less interested in Flyber service than other groups (reasonable, because the group is insensitive to the price.) Neighborhood in West Harlem and Sutton place showed less interested in Flyber service.
2. The distribution of price per mile in Male group are very similar to Female group. For age under 36, people are willing pay more price per mile in average. The income and price per mile has positive relationships, people with higher income are willing to pay more price per mile. The data for each neighborhood is small size, we cannot see any significant relationship.
3. Please see next slides

Answer Slide

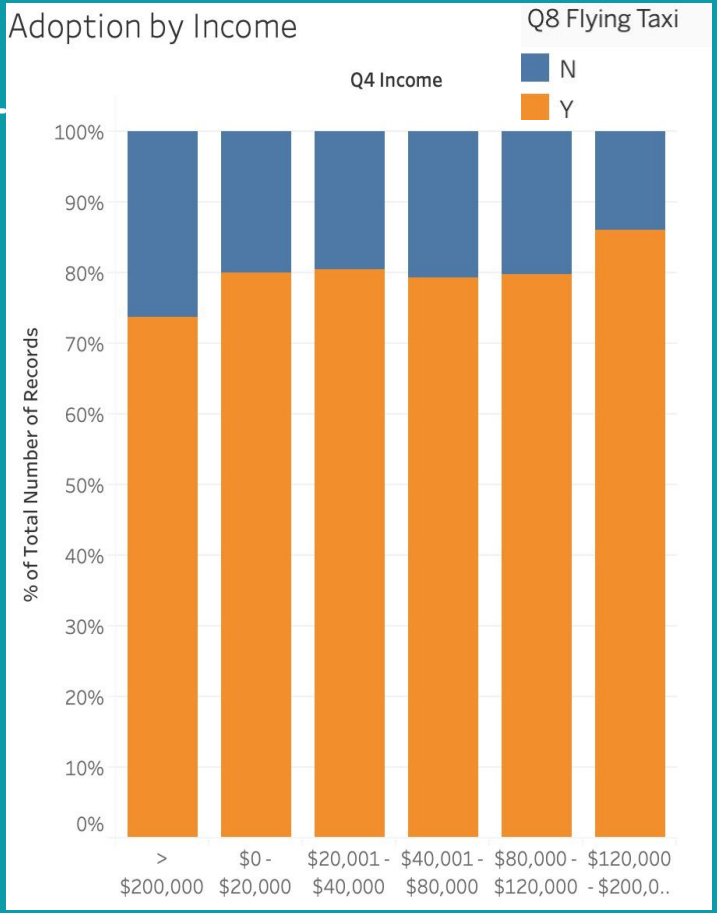
3. We can see people with high income (>200,000) are less interested in Flyber service. From the answer with Q10, we noticed that 2 main issues with Flyber service for people are safety and price. Some people cared much about safety and people are very sensitive to the price tend to not using Flyber service.

Final Conclusions-part 1

The graph shows the proportion of Adoption of Flying taxi by income.

All the income group shows the majority of accepting concept of flying taxi.

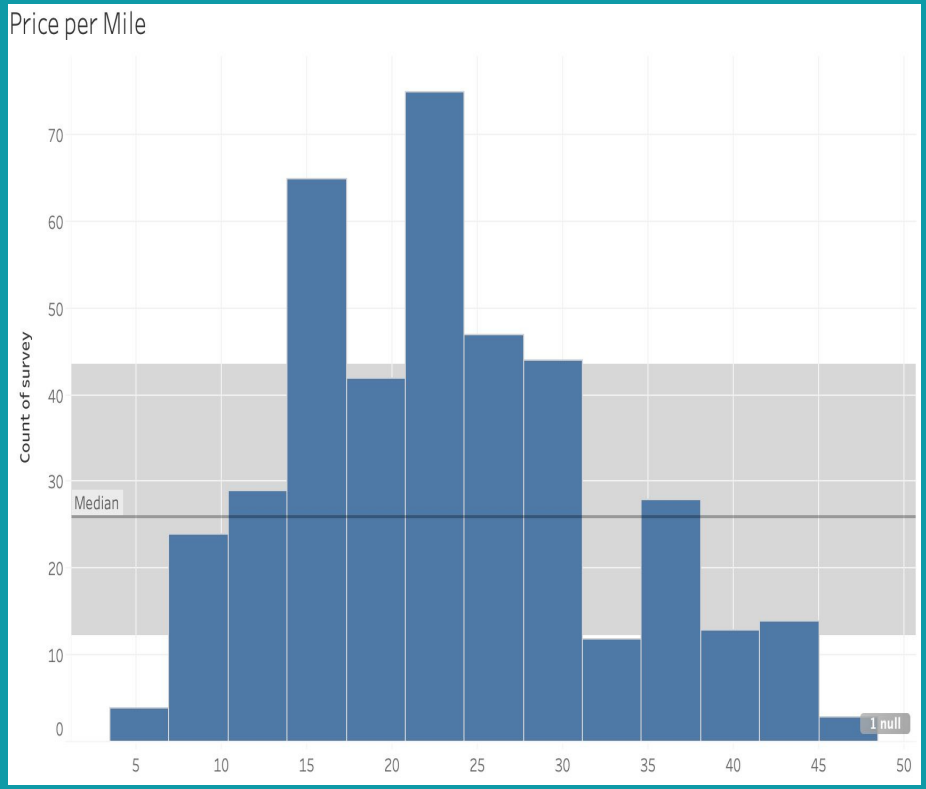
People with income $120,000-200,000 shows the highest proportion of accepting flying taxi. And people with income >$200,000 shows the lowest proportion of accepting flying taxi.



Final Conclusions-part 2

Potential Price:
The graph shows the median price per mile users willing to pay is $26, and most of users can accept $28 per mile.

And according to the distribution, we can conduct experiments on potential price of $10-$28, which are acceptable for majority of users.



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

Objectives: User Acquisitions
Attract enough customers to use Flyber service.

Reasons: User Acquisitions is important because we need find new users of Flyber at the initial stage, in this way, we can get enough data and feedback to analyze customer needs. Other areas like Engagement and Retention, profitability are also important, but these happens after customer acquisitions and we can revisit these fields in the later stages.

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. Number of rides per week.
2. Number of duration per week.
3. Total distance traveled in the City per week.

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

Answer Slide

1. 1,000 rides per week. (We can see from the existing data, taxi services are roughly around 40,000 per week.)
2. 150 minutes per week. (The existing data showed average duration for each week is around 6000 minutes.)
3. 3,000 miles per week. (The existing data showed average distance for each week is around 140,000 miles.)

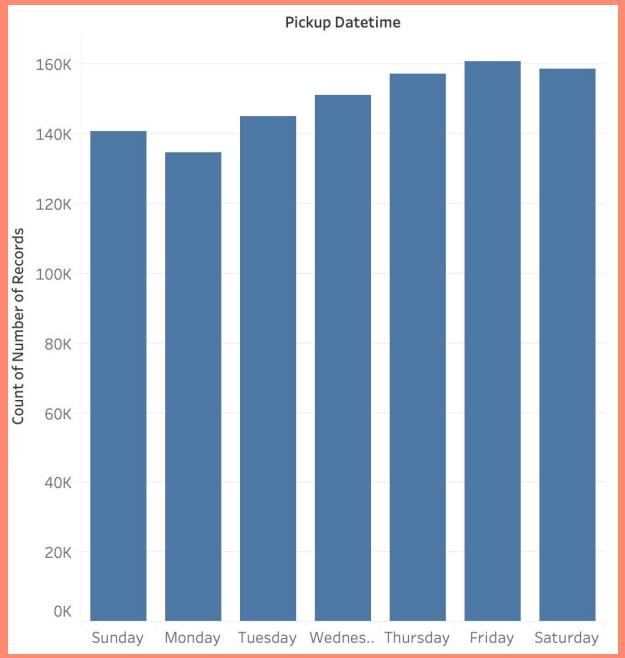
As the product manager, you make decisions based on the insights you extract, we'll need to know the feature set we'll include in the MVP to measure viability, while keeping operational expenditure under control:

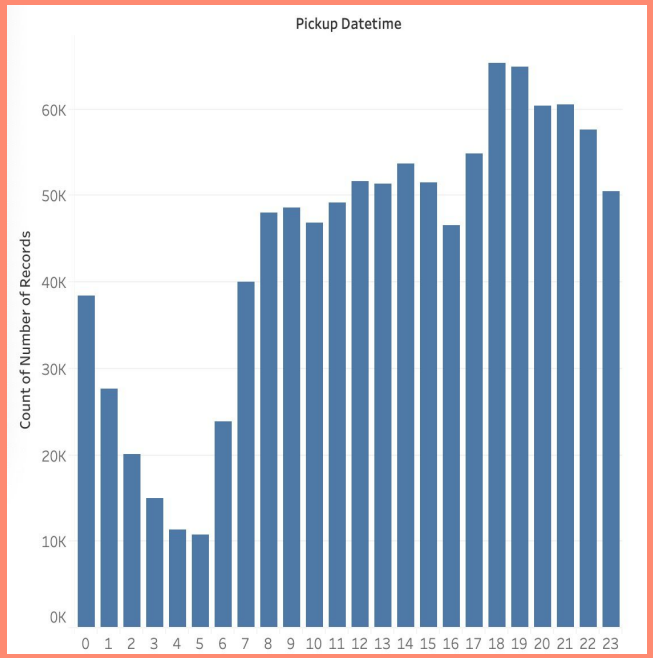
- What times/days of operation should the service run for?
- How many pick-up / drop-off nodes should we have?
- Where should the nodes be located?
- Should we initially use copters or homegrown hardware?
- Should the pricing be fixed or dynamic? At what rates?

Answer Slide

- What times/days of operation should the service run for?

The analysis shows the relatively high proportion of rides through 18:00-22:00. And more rides during Tuesday to Saturday, so we should run for 18:00-22:00 during Tuesday to Saturday.

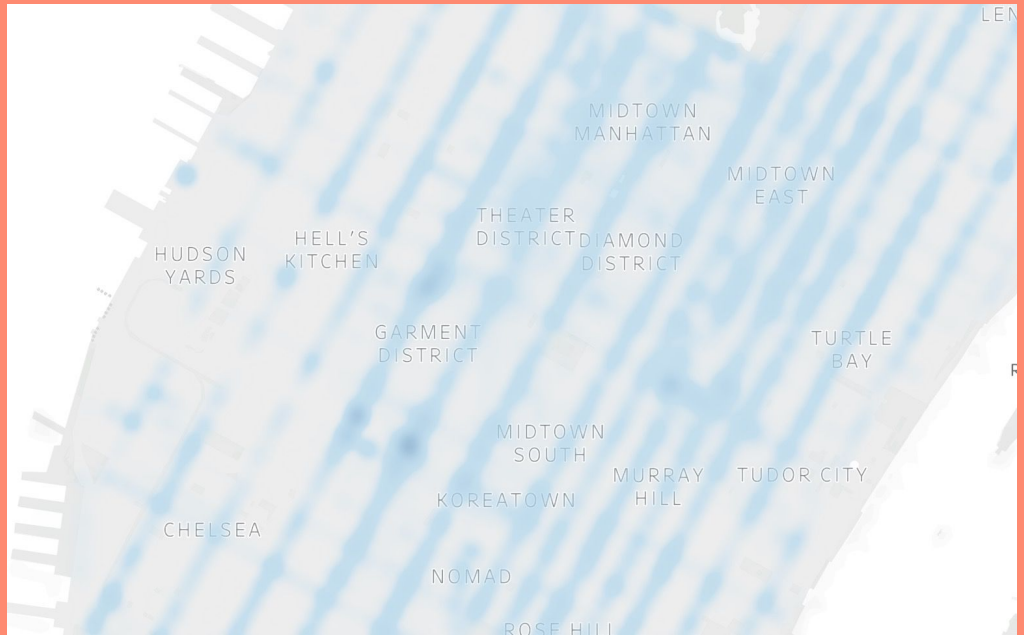


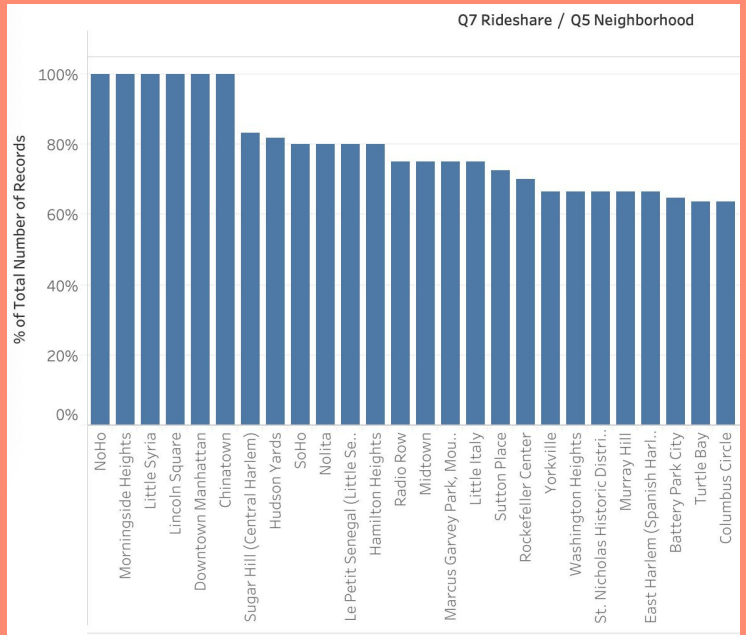


Answer Slide

- How many pick-up / drop-off nodes should we have?
- Where should the nodes be located?

Combine users' acceptance of Flyber with total rides in the city area, we can set 4 nodes first around: Hudson Yards, Midtown South, Garment District and Soho.





Answer Slide

- Should we initially use copters or homegrown hardware?
  Copters, because in general, homegrown hardware would be more expensive than copters for the start of operations.

- Should the pricing be fixed or dynamic? At what rates?
  Dynamic. The price should be calculated on Distance. According to the user research, I would suggest the price per mile is $15 and $5 initial fee. So the total price would be $5+$15*Distance

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

Answer Slide

Sample Size, using conversion rate of 20%, Minimum Detectable Effect of 20%, significance of 95%, the sample size is 1,100.

Estimated experiment time, since the hypothesis is 1,000 rides per week, The Estimated experiment time can be set as 2 weeks.

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

Answer Slide

1. Number of rides per week.
   Events: Rider pickup
   Definitions: Triggers whenever the driver hits the “Passenger is picked up” button on the Flyber driver app
   Properties: Ride_id, Pickup_Datetime, Dropoff_Datetime, Passenger_count, rider_id
2. Number of duration per week.
   Events: Rider pickup
   Definitions: Triggers whenever the driver hits the “Passenger is picked up” button on the Flyber driver app
   Properties: Ride_id, Pickup_Datetime, Dropoff_Datetimne
3. Total distance traveled in the City per week.
   Events: Rides completed
   Definitions: Triggers whenever the driver hits the “Trip is Completed” button on the Flyber driver app
   Properties: Ride_id, Pickup_Location, Dropoff_Location, Distance

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

Answer Slide

Feedback Survey:

1. Rider’s overall rating on each ride, scale 0-5.
2. User’s overall rating on each ride, scale 0-5.
3. For Users: What is the best part of the ride

- Service
- Duration
- Price
- Convenience to get in

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: From Part I distribution, we can target users at age under 45, with income smaller than 200,000. Because user-research shows people with these characteristics have higher acceptance of Flyber service. And user mainly lives/works in the Midtown part of Manhattan. Because Taxi data shows heavy rides and traffic in the area.

Pain points: Too much time spent on heavy traffic. Hard to find parking place. Cause delay because the unpredictable traffic.

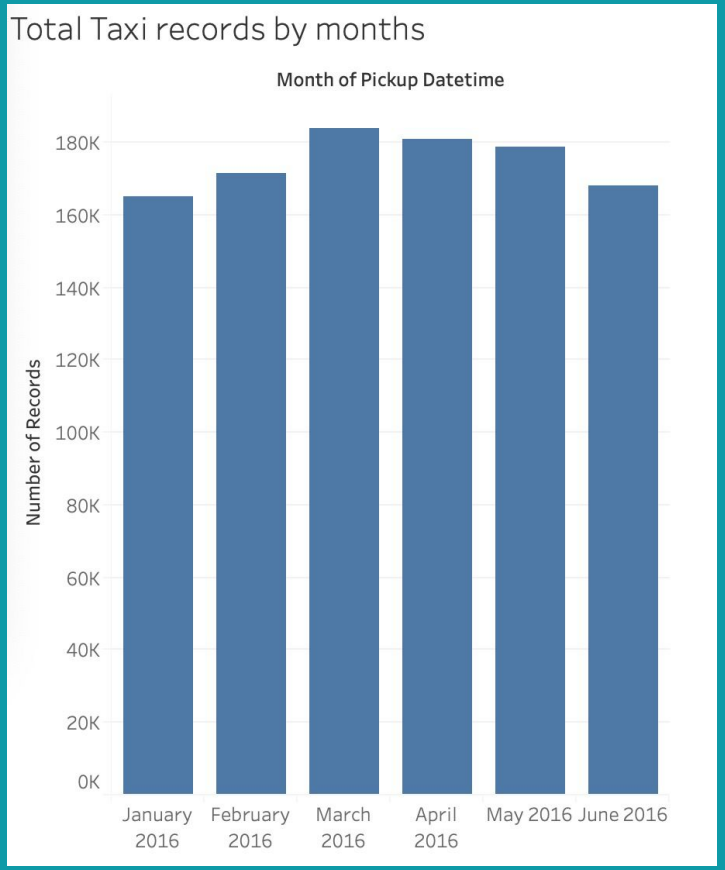
Answer Slide

Claim: Growing transport issues in New York City requires innovative solutions to solve the crisis. Flying taxi service could help solve issues by improving users experience, address their problems on heavy traffic, time delay, etc. Also the service could benefit the public road infrastructure, it relieve the burden of the road maintenance,

Evidence: From the Taxi data, we can see huge demand for taxi service in the NYC. Also, from user-research, we can see majority acceptance of Flying taxi service.

(See charts in the next slides)

Answer Slide



Answer Slide

User impact:

1. Increase customers experience in taxi service, provide more transportation choice.
2. Decrease customers wait time, and delay time in the taxi service.

Market impact:

1. Reduce road traffic, improve the road condition and decrease the cost on infrastructure maintenance,
2. Possible lower price for public transportation, because more competitors.

Answer Slide

Risks and Assumptions:

1. The choice of pickup and drop-off locations for Flying taxi may be difficult, considering crowded parking situation in the NYC.
2. Safety issues: Since the service is new, we need to evaluate safety. And to make sure there are completed rescue and other supplement process.
3. User experience, The Flying taxi service is different from traditional road transportation, so users may encounter unpredictable issues during the trip.

Answer Slide

Cross-functional stakeholder teams:

Data science, Design, Product marketing, User research,
Data engineering, Client engineering,
Legal, privacy, and compliance teams, Finance & business operations,
Sales & customer success, Quality assurance engineering