

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

Ryber

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

Back to the basics of product management, identify your customer and their pain points:
What are taxis used for? – Taxis are used for commuting from one place to other for the purpose of visiting a friend/relative, shopping, going to office/college or enjoying a trip to tourist destinations.

What are the characteristics of the users that leverage them? – People aged between 15-75 (more females than males), midlevel income group (\$40000-\$200000), primarily from neighbourhoods in midtown(>50%) and lower Manhattan regions.

What are existing pain points with taxis? – City traffic is the primary pain point which costs the users time and money. They feel exhausted and frustrated when stuck in traffic.

What are the existing pain points with digital ride-sharing services? – Existing digital ride-sharing service optimize the path based on the travel in similar direction. It aligns the new request to the existing taxis that are currently on similar routes. Multiple pick-up and drop points can cause the route to get extended unnecessarily which sometime can cause more travel time for fellow travellers. It result in customer dissatisfaction and wastage of time & money. Too many vehicle stuck in roads is not good for environment either.

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? – Flying taxi service would be point to point pick-up and drop facility without have to worry about the traffic congestion or route diversion. It would not have uncertainty of travel time unlike existing state of taxis has today which would save a lot of time & money and keep the users delighted. Users can plan their travel and work better as the travel time can be accurately anticipated.

What market improvements do you hypothesize a flying taxi service would have the existing taxi service industry & physical road infrastructure today? – Flying taxi service would help taking cars out of the currently congested routes which in turn will make the traffic relatively smoother in those routes. The cars can be deployed to other places to pick/drop the people from flying pick-up/drop points for onwards journey which will help equally distribute the traffic amongst city roads. Less congestion in road will save energy and money which can be utilized in other areas. Furthermore, the state taxi service will invest in road infrastructure and better route planning in order to compete with flying taxi services. To make price competitive they may explore low cost clean energy solutions like electric taxis.

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

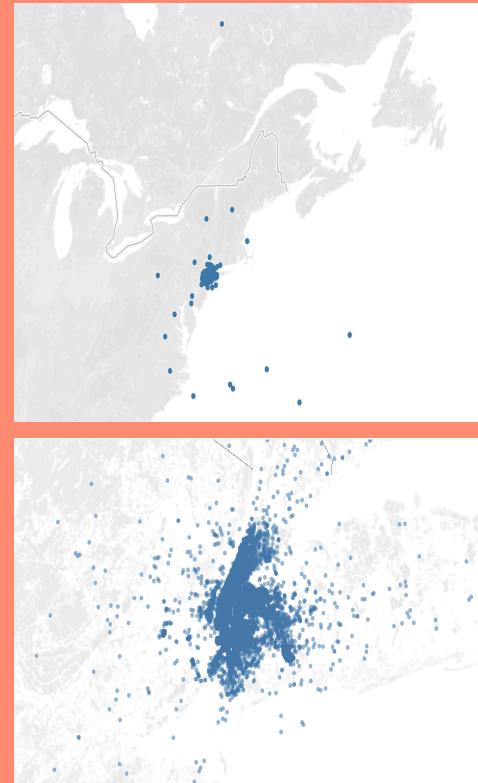
How many records are in the dataset? - 1048468

What does each record represent? – Each record represent a travel event of taxi ride with related details e.g. pick-up/drop time/location, travel duration, distance travelled and vendor identification.

What is the primary key? - id

What date range is your dataset bound to? – 01/01/2016 to 30/06/2016

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? – No, it is not limited to Manhattan. There are many data points from Brooklyn, Queens and Bronx, few from Staten Island and New Jersey as well. Most of the data points are centralized at Manhattan. Yes, there are few outliers too.



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

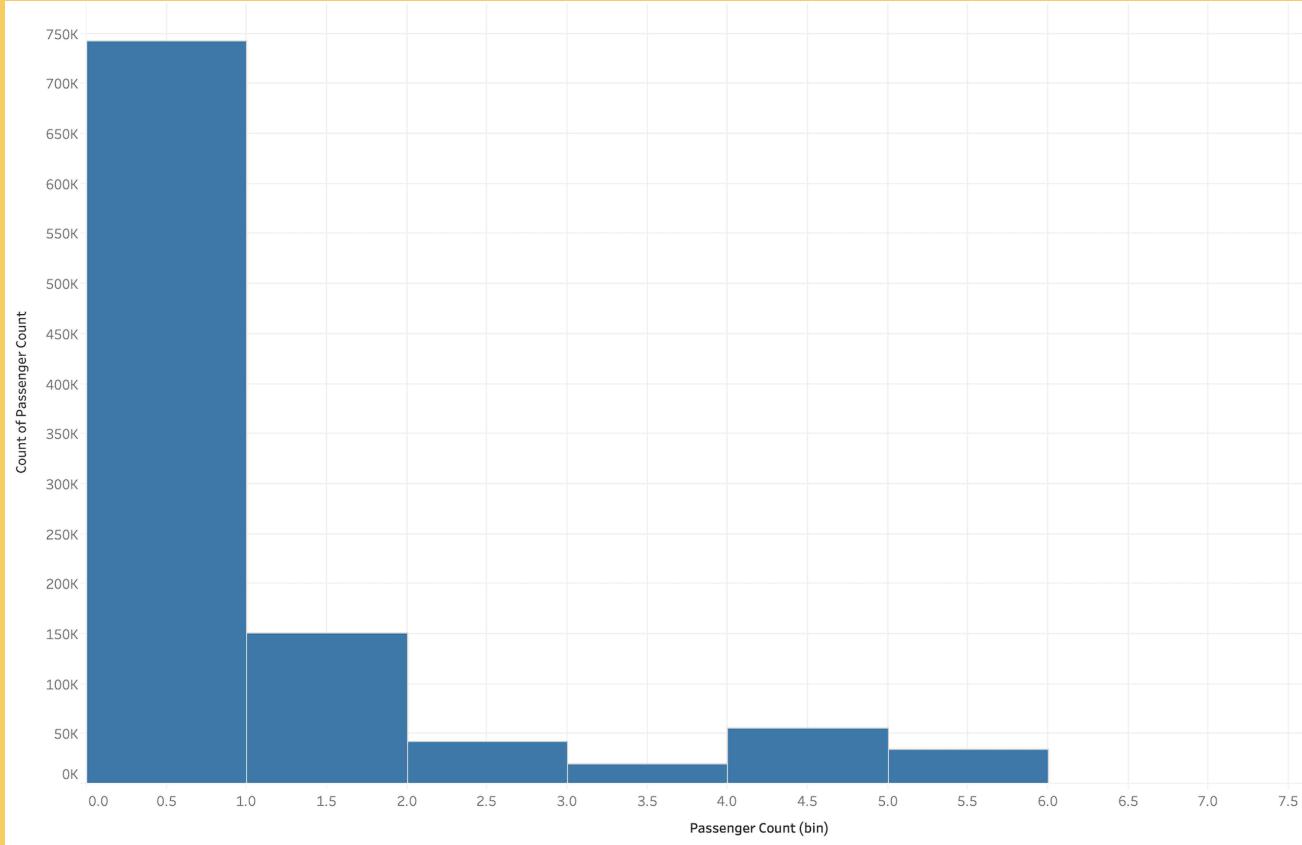
	AVG	MEDIAN	STDEV	2 STDEV
duration (seconds)	906.8	662	2469.2	4938.36
distance (mile)	3.42	2.11	3.85	7.7
passenger count	1.67	1	1.31	2.62
Distance to duration ratio (seconds per mile)	372.4	280.59	2106.7	4213.44
price (\$)	\$19.66	14.61	24.49	48.98

This is the best I could achieve by eliminating multiple outliers, (trust me) I spent entire one day just to identify and eliminate the outliers and learnt a lot more about them – 0 passengers, 0 distance, 0 ‘distance to duration’ ratio, pick-up/drop in water (which could be justified as it could be related to the nearby location) and events from far fetched places and other states. I think as based on the density map, we are going to focus on density map, the outliers would anyways become irrelevant.

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



Most (~90%) of the rides are having 1-2 passengers on-board. So, it would make sense to leverage mini copters for initial MVP.

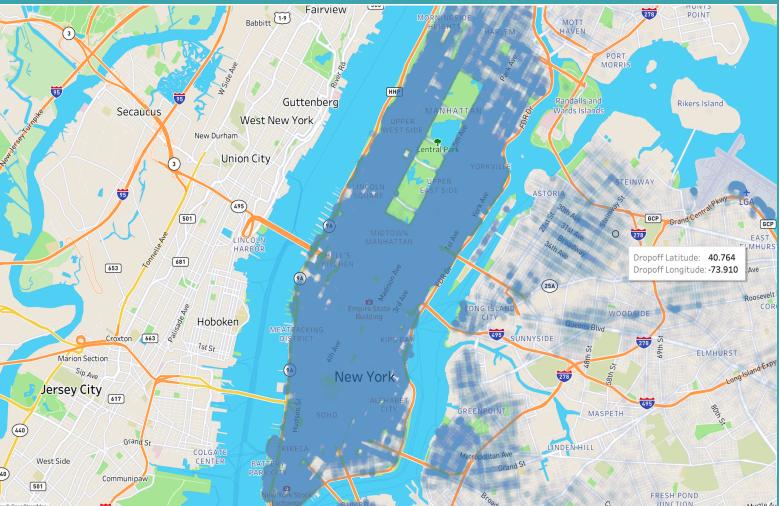
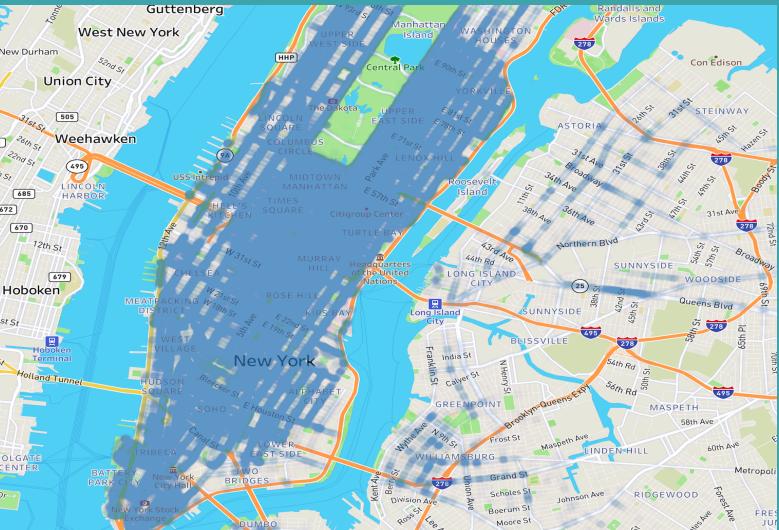
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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? – Neighbourhoods from Midtown and lower Manhattan e.g. Columbus circle, Times square, Korea town, Rose hill and SoHO, Financial district etc.

Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs? – Again neighbourhoods from Midtown and lower Manhattan e.g. Upper east side, Lenox hill, Turtle bay, Times square, Korea town, Murray hill, SoHO and Financial district etc.



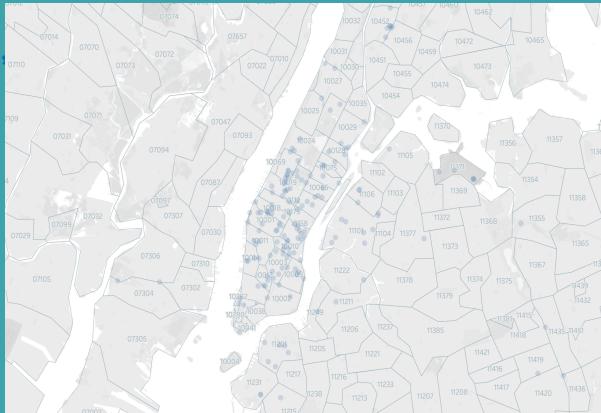
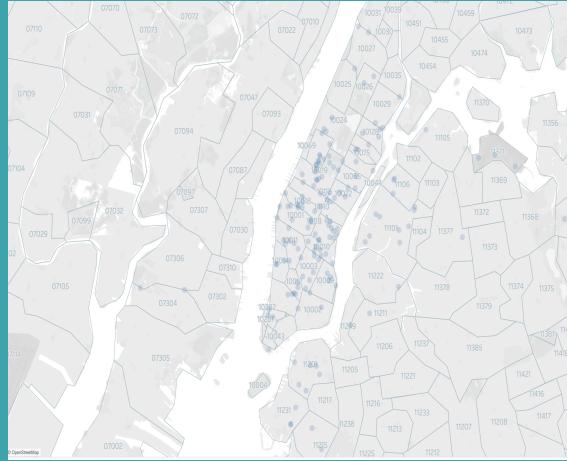
Answer Slide

Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up? – Columbus circle, SoHO, Empire state building

Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off? – Lincoln square, Hell's kitchen, Murray hill

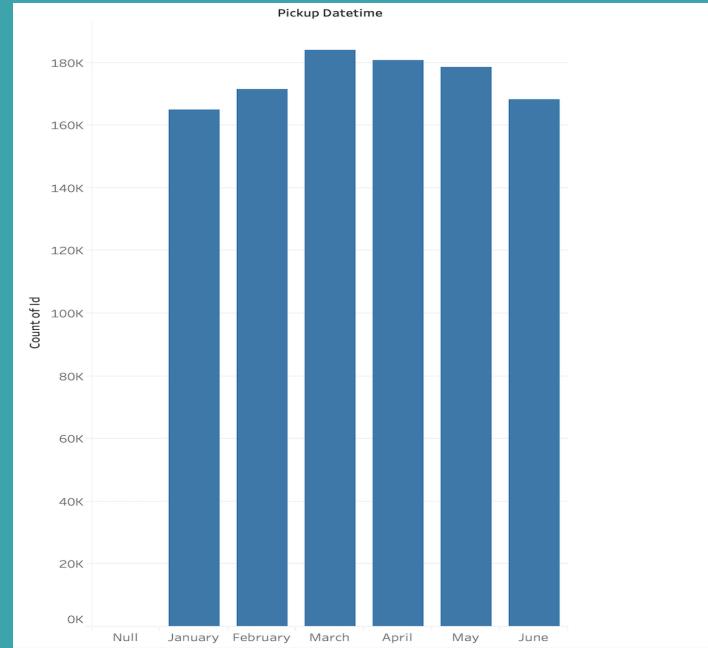
For any of the neighborhoods identified, are there any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off? What makes them suitable? – Columbus circle, Murray hill, Hell's kitchen, LIC tower, SoHO etc. Below mentioned points makes them suitable for flying pick-up/drop-off:

- High density of pick-ups and drops
- High density of distance to duration ratio for pick-ups as well as drops
- Famous shopping destinations, administrative buildings, and tourist places nearby



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.

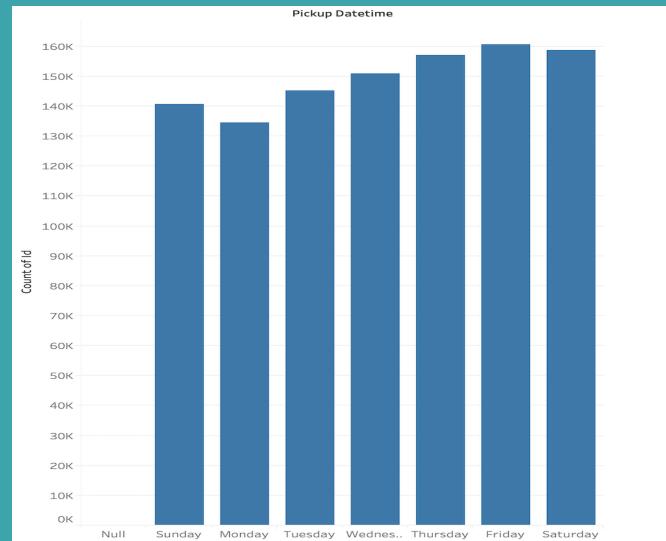
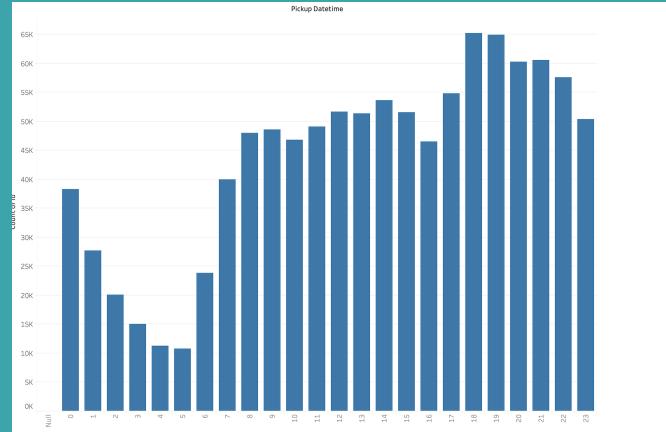


Answer Slide

What times throughout the day experience relatively higher volumes of ride pick-ups? – 6-7PM experience relatively higher volume of ride pick-ups during the day.

What days throughout the week experience relatively higher volumes of ride pick-ups? – Friday experience relatively higher volumes of ride pick-ups during the week.

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. – Based on the six month data (plot in previous slide), the volume starts picking up from the month of January and peaks in the month of March. The volume then start declining slowly from March to June.



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?

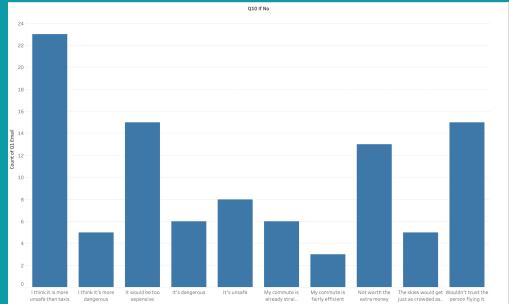
Answer Slide

Is there an inclination of better Flyber adoption based on gender, age, income level, or neighbourhood of residence? – Inclination of better Flyber adoption is observed based on age and neighbourhood of residence. Inclination based on gender and income level is in proportion to the data points available for the same so it doesn't give clear indication(plot in next slide).

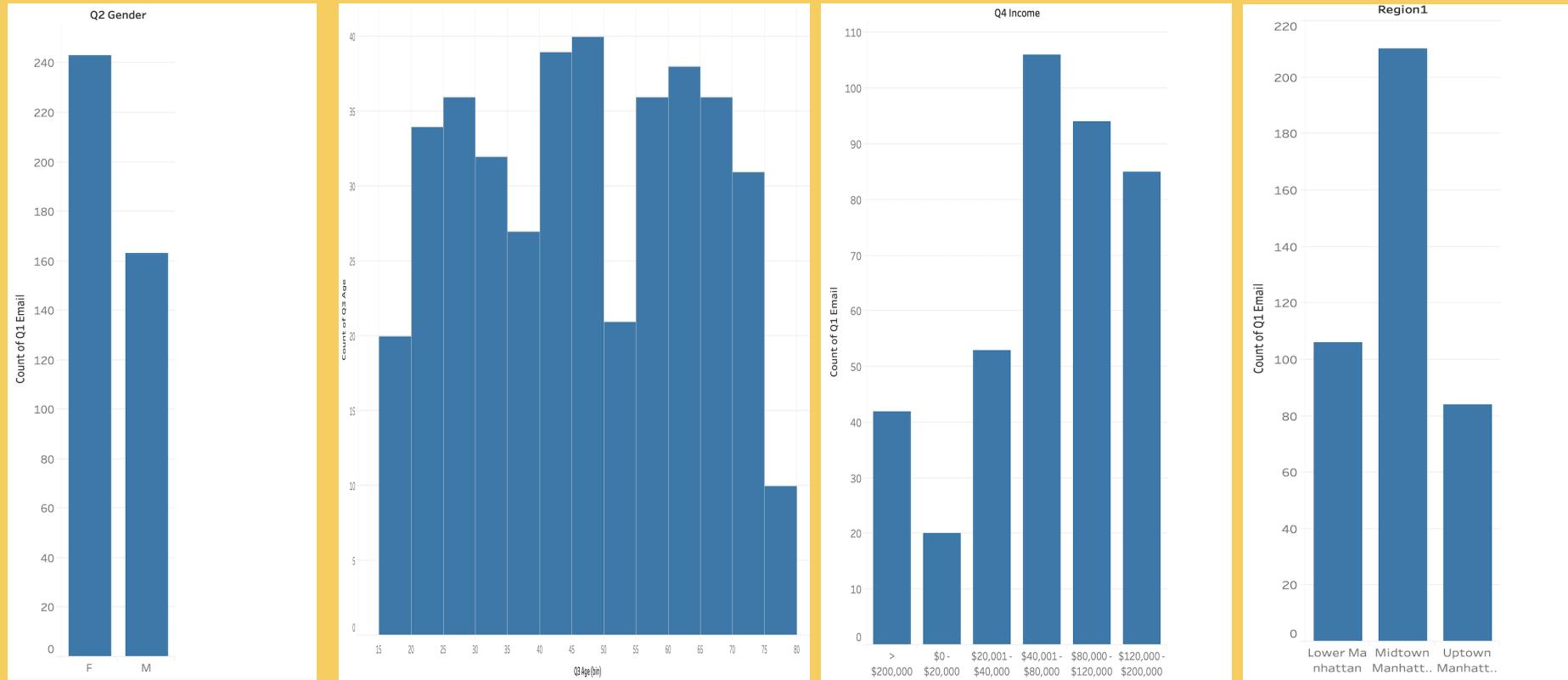
What is the distribution of potential price per mile based on gender, age, income level, and neighbourhood of residence? – Please refer attached table. The price people are willing pay for Flyber is higher than the average for taxis for almost all the dimensions and the spread is relatively less which means people across the market segments are willing to use the service and pay extra cost.

What is the different personas/segments of negative sentiment towards not using a flying taxi car service? – Please refer attached picture. More than 50% of negative responses are related to safety (more unsafe than taxis, dangerous, unsafe, don't trust the person flying it). Extra cost is not an issue for most part of population so this is an ideal region to launch and run a pilot.

Potential price per mile distribution:			
Pay per mile	AVG	MEDIAN	STDEV
Gender			
Female	23.29	23.00	9.38
Male	23.01	22.00	9.15
Age range			
15-20	26.75	25.5	9.7
20-25	25.24	25.5	9.46
25-30	21.47	18.5	10.12
30-35	22.81	21.5	10.5
35-40	19.63	19	7.46
40-45	23.21	23	9.17
45-50	23.08	21	9.67
50-55	23.1	23	8.9
55-60	19.22	18.5	7.38
60-65	25.08	23.5	8.35
65-70	26.92	26.5	9.25
70-75	21.94	21	8.97
75-80	23.9	24.5	8.96
Income group			
\$0 - \$20,000	11.7	11	4.66
\$20,001 - \$40,000	15.43	15	5.38
\$40,001 - \$80,000	18.33	18	5.46
\$80,000 - \$120,000	22.13	22	5.13
\$120,000 - \$200,000	31.05	30	6.46
> \$200,000	37.1	38	6.13
Region (Neighbourhood)			
Lower Manhattan	25.91	25.00	10.18
Midtown Manhattan	22.46	22.00	9.10
Uptown Manhattan	21.54	21.00	7.77



Potention Flyber adoption inclination based on gender, age, income and region:



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

Objective – To enable users for a safe transportation in New York City to quickly travel from one place to other place across the city without having to worry about the uncertainties of travel time due to heavy traffic conditions.

Reasoning – As discovered in the user insight analysis in part I, it is evident that more than 50% of the user population survey is willing to go for an alternative and willing to pay for the same. Now, since the product is in its infancy stage so it would make sense to inform the users about the offerings and engage them for how they can use it and contribute in improving new age transportation alternative for the city. We saw from the data that people in midtown and low Manhattan areas (relatively rich population as well) are eager to use Flyber taxi so better user acquisition and engagement would be key to be successful in the region and beyond. User retention and profitability can be focused later when we have a relatively matured product and good user base.

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

Answer Slide

Daily active users (>500)– There should be at least 500 user login on daily basis to start with. Considering the trend for enthusiasm and high duration travel pick-up/drop density in Manhattan area this should be a good indicator to measure success of the product.

Trips completed daily (>12)– There should be at least 12 trips a day across the city to start with depending upon the customer requests. Higher is the better.

Passengers served daily (>24)– There should be two passengers on average served in each trip to be cost effective.

NPS score (>90%)– Consumer satisfaction is extremely important in initial days for more traction and brand building so it should be no less than 90%.

Daily conversion rate (>5%) – Daily conversion rate is at least 5% due to limited number of routes and information available to the users initially. It is likely to go up on daily basis.

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

Answer Slide

From the data insights in part I, the average duration per trip is more than 16 mins and average distance travelled is 3.4 mile i. e. ~12 mile per hour which is truly an average speed for a city like New York City. There are more than 44% of travel events that are slower than the average speed.

Reducing the number of travel events to daily basis and increase in taxi usage since 2016, Daily active user count (500) is an indication of the interest the service is able to generate among potential users.

Other KPIs are related as if we manage to maintain at least 5% of conversion rate on daily basis then the passenger server would be atleast 25 per day which in turn is 12 trips a day (consider 2 persons per trip on an average)

As the product is in its fancy stage, we can't compromise with the customer satisfaction and a delighted customer will refer more and more customers so it has to be more than 90%.

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

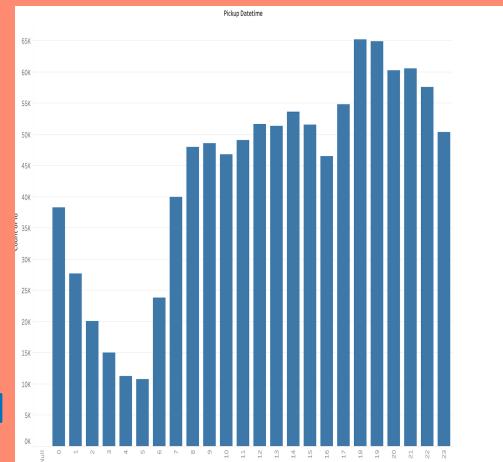
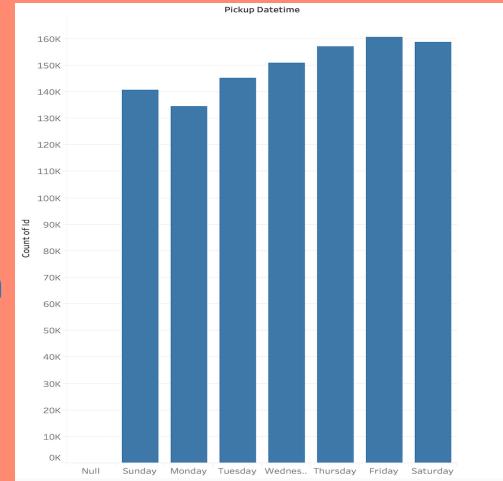
What times/days of operation should the service run for? – 7AM – 7PM considering safety and potential adjustment of couple of hours by the people falling out of this range, all days with readiness for extra trips on Thurs, Fri and Sat.

How many pick-up / drop-off nodes should we have? – initially 2 in the places where there is high density of travel events having greater ‘distance to duration’ ratio populated.

Where should the nodes be located? SoHO, Columbus circle

Should we initially use copters or homegrown hardware? – copters (MVP), as homegrown hardware will be time consuming and require heavy investment.

Should the pricing be fixed or dynamic? At what rates? – Fixed at \$22 (median price from the distribution table in part I for the target population, midtown & lower Manhattan which is relatively rich population, should be comfortably willing to pay) initially to promote the service and keep it simple for customer acquisition, dynamic based on the demand later.



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

Answer Slide

Based on the data explained in earlier slides, sample size per variation is 6900 and duration to conclude the hypothesis is 61 days considering we'll only use 50% of the visitors for the test and using the calculator prescribed.

Baseline Conversion Rate
 %
Your control group's expected conversion rate. [?]

Minimum Detectable Effect
 %
The minimum relative change in conversion rate you would like to be able to detect. [?]

Statistical Significance

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [?]

[EDIT](#)

Sample Size per Variation
6 , 900

How long should you run your A/B Test?

Estimated existing conversion rate (%)
 %

Minimum improvement in conversion rate you want to detect (%)
 %

Number of variations/combinations (including control)

Average number of daily visitors

Percent visitors included in test?
 %

Total number of days to run the test: **61 days**

CALCULATE TEST DURATION

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

Answer Slide

Events and properties listed below:

Event : LogIn	Event - checkedIn
Definition - First touch point when user launches the app and logs in	Definition - Triggered when entry is confirmed at the check-in counter at the pick-up location.
Properties – userID, device, browser, date/time	Properties – userID, pick-up location, checkin date/time
Event - scheduledFlight	Event - boardedFlight
Definition– User schedule flight and check for availability in the app.	Definition – Triggered when boarding confirmation is done in the app and rider is ready to take off.
Propeties - source, destination, schedule date/time, event date/time	Properties - pilotID, tripID, riderID boarding date/time, destination, pick-up location
Event - bookedFlight	Event - landedFlight
Definition – User makes the payment and book the flight in the app.	Definition - Event triggered when landing is confirmed using app at the destination.
Properties – bookingId, payment method, date/time, payment status	Properties – pilotID, riderID, tripID, landing date/time, source, drop location
Event - cancelledFlight	Event - checkedOut
Definition – User cancels the flight and provider reason in the app.	Definition - Event is triggered when user's exit from drop-off location is confirmed in the app.
Properties - bookingId, cancellation reason, date/time	Properties - – userID, tripID, drop location, checkout date/time

Answer Slide

KPI calculation and other insights using the events:

Daily active user – can be calculated by the number of login events on daily basis.

Trips completed daily – can be calculated by distinct tripID generated.

Passengers served daily – can be calculated by checkOut event generated.

Conversion rate – number of bookedFlight events/number of login events

NPS score – it can be gathered by using the user survey or capturing feedback (positive, neutral, negative) at the end of each stage in the flow.

Other events will help us determine where (which step) the users are dropping in the flow and help us analyze the reason for the same. Then, we can analyze if there is something that needs to be changed which can be gathered by user feedback at the later stage. This is helpful to acquire, engage and retain the users in order to increase profitability.

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

Answer Slide

- Rate the experience – positive, neutral, negative
- What did you like the most? Booking, Onboarding, Travel, Other - please specify
- What did you like the least? Booking, Onboarding, Travel, Other - plese specify
- Would you be willing to use the service again? – Yes, No
- If Yes, why? – 1. Saves time 2. Secure 3. Cost effective 4. Other - please specify. If No, why
1. Costly 2. Not secure 3. Dangerous 4. Other - please specify
- Would you like to have similar service to the other part of city?
- If No, Why - 1. Service is not good 2. Not cost effective 3. Other? If Yes, Choose from next top 3 places (Hell's Kitchen, Murray Hill, Empire State Building) or other – please specify.

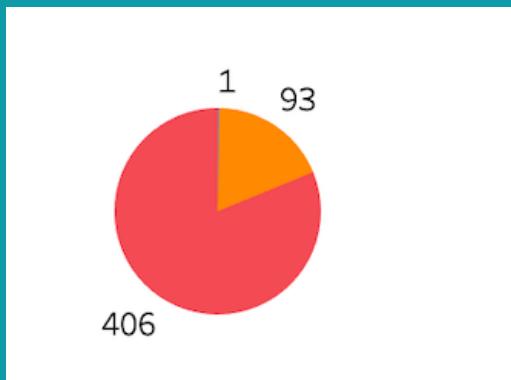
Summarize everything you have learned into your final proposal

- 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

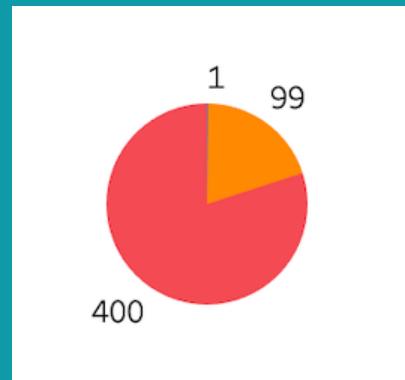
Answer Slide

From the taxi rides data in section I, there are more than 314000 travel events (30%) for which the time taken to travel 1 mile is more than 6 mins. which means the speed of 10 mile an hour. The existing city taxi travel is frustrating, time consuming and costs unnecessary higher amount of taxi bill to the users.

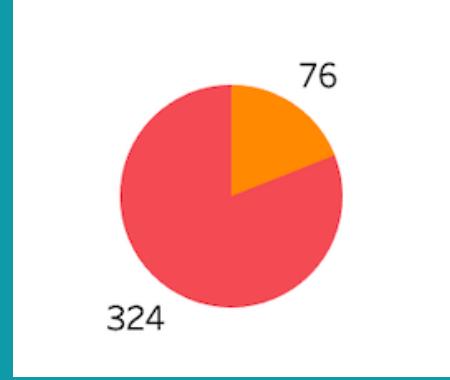
User insight data clearly shows that more than 80% of the population is using existing mode of transportation and looking to replace it with an alternate mode of transportation which is convenient and secure.



Current taxi users



Population willing to use Flyber



Current taxi users willing to use Flyber

Red – Y, Orange – N, Blue - Null

Answer Slide

User Impact: Users will have delighted experience of quickly travelling from one place to other. People can come to Flyber to have a new experience and visit nearby tourist destinations. The hassle free travels saves user time, energy and keep the user happy.

Market Impact: As the flying taxi capture more and more users, the number of taxies on road will reduce. This will make the traffic condition better and users (car driver/ traveler) will have a better experience travelling to city roads. To compete with Flyber, the taxi vendor will explore the low cost electric fuel and launch competitive prices.

Business Impact: With the maturity of MVP, as the product expands to more and more destinations, more and more users will be adopting the flying taxi services resulting in more and more revenue for the company. Delighted customer will refer more customers and help increasing company's brand value.

Answer Slide

Solution: Flyber will be providing a flying taxi service for two of the most densely populated pick-up/drop locations having highest 'distance to duration' ratio.

Risk & assumptions: **Risk** – There can be a risk of getting environmental clearance as the product matures and number of flying taxies increased in the sky.

Assumption – Flying taxi not only will help the daily travelers travel to their destination but will also act as a tourist destination where people come to gain a new experience and use services.

Roll out and launch strategy: The roll out will be for the selected population having point to point route between the two travel points initially and based on the feedback it can be extended for the general public. Also, the travel points will be increased across the city based on the demand and feedback received.

State cross-functional stakeholders: Police department, Environmental clearance department, Disaster management department, Local authority/government to approve location/structure and plan.