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

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. Taxis are used for moving from one point to another with less effort.

2. The users have urgent work to do and are mostly middle-aged and above. They want to reach in less time and with less effort.

3. Taxis' painpoints:
   a) More time for less distance
   b) High Fares.
   c) Unsafe
   d) Commute to Taxis parked area.

4) Ride-sharing painpoints:
   a) More time for less distance
   b) Some times more passengers.
   c) Surge Pricing.

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. Flyber will usher in
   a) Great time savings for the user.
   b) Higher Safety standards.
   c) VFM (Value for money)

2. Flyber will bring in
   a) Lots of breathing space for Physical infrastructure to be improved
      drastically.
   b) Taxis industry will become Lean and efficient to compete with
      Flyber, once it becomes a mass value proposition. It's do or die for
      taxis to be competitive.

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

1048468 rows viz. One million rows are there.

Each record represents a ride.

Id is the primary key, wherein it denotes every unique Ride.

The Date Range is from Jan 1 2016 – June 30 2016.

No, it's not limited to Manhattan but includes Brooklyn, Queens, Staten Island, the Bronx, and New Jersey too. The data points are more from Bronx till Financial District. Also Outliers aren't there.

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`.

Clarification

I have updated the calculated field "Price" based on "duration" and "distance" viz. Price=((Distance)*1.75) + ((Duration)*(0.35/60)) from Investopedia url **https://www.investopedia.com/articles/personal-finance/021015/uber-versus-yellow-cabs-new-york-city.asp** and the Passenger count doesn't affect price.

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: avg – 953 s, median – 662 s, 1st std. dev - 3177 s, 2nd std. dev – 6354 s.

Distance: avg – 3.437 miles, median – 2.095 miles, 1st std. dev – 3.956 miles, 2nd std. dev – 7.912 miles.

Passenger count : avg – 1.664, median – 1, 1st std. dev – 1.314, 2nd std. dev – 2.628.

Duration-to-distance ratio: avg – 4685, median – 280.8, 1st std. dev – 924378, 2nd std. dev – 1848756.

Price: avg – $11.57, median – $7.72 , 1st std. dev – $20.82, 2nd std. dev – $41.64.

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

The Potential For low passenger count (1-2) is the highest when histogram was plotted as seen below. 7,43,095 passengers travelled in this category and the nearest competitior in the bin (2-4) had 1,94,093 passengers.

Passenger count bins

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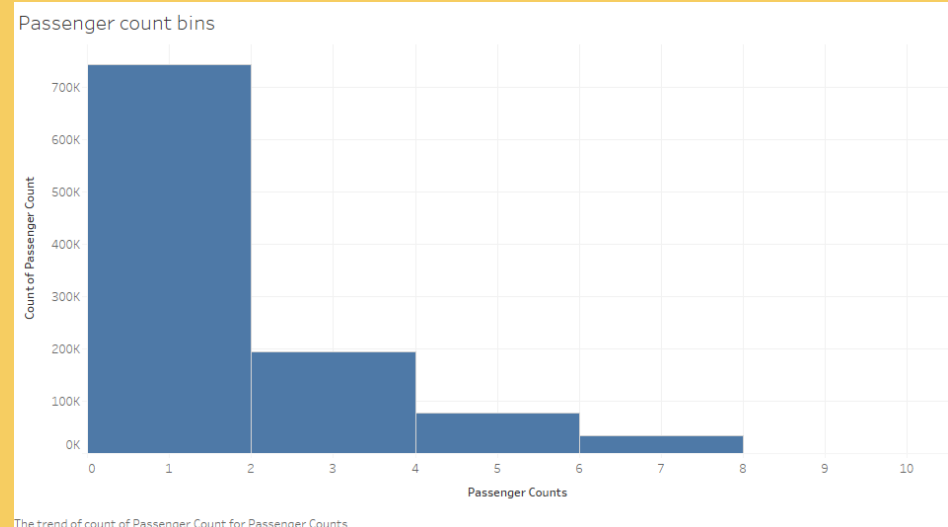
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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. The neighbourhood of Midtown Manhattan, Theater District, Midtown
   South, Koreatown, Flatiron District, Nomad, Rosehill, Midtown East
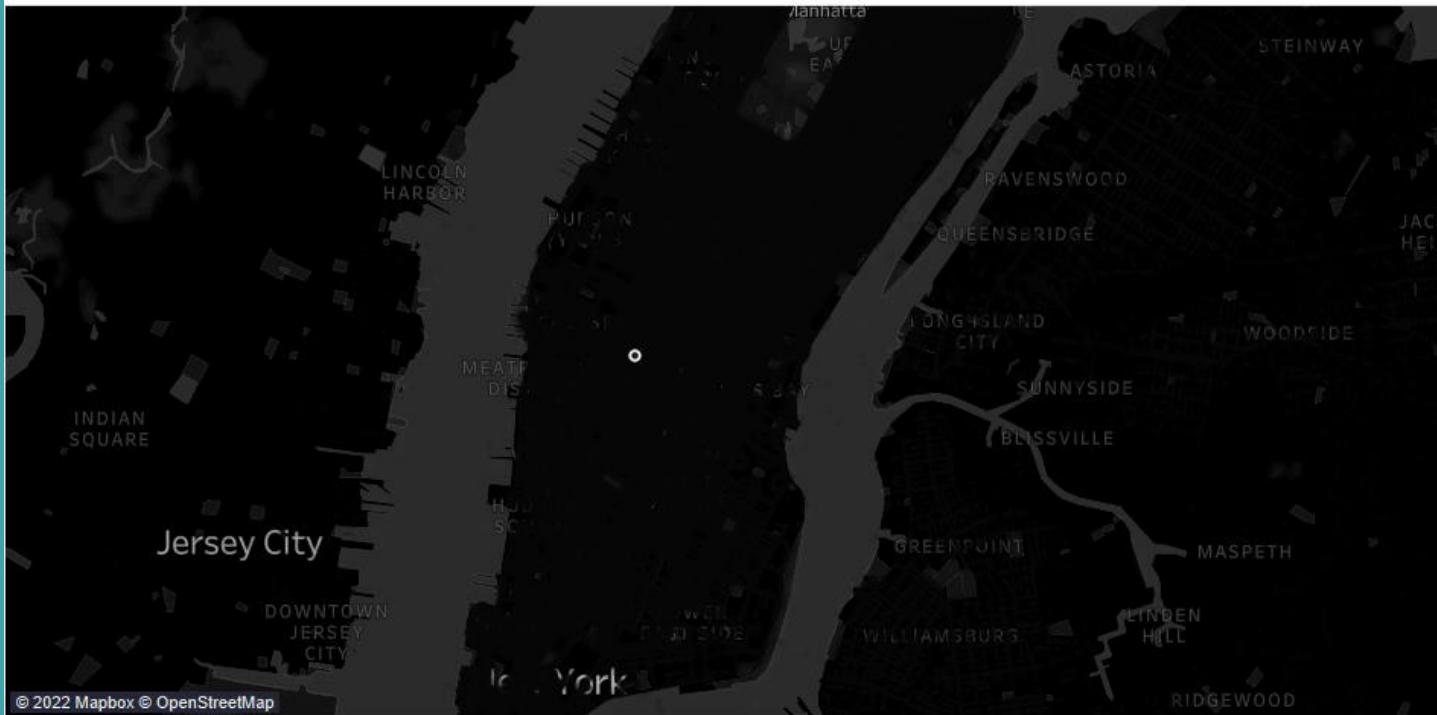   and Columbus Circle have relatively higher density of pickups.

Map based on Pickup Longitude and Pickup Latitude.

Answer Slide

2. The neighbourhood of West Village, Hudson Square, East Village, Union Square, Midtown Manhattan, Murray Hill, Turtle Bay, Chelsea and Lincoln Square have relatively higher density of drop-offs.

Higher Density Dropoffs

© 2022 Mapbox © OpenStreetMap

Map based on Dropoff Longitude and Dropoff Latitude.

Answer Slide

3. The neighbourhood of Midtown Manhattan, Nomad, Murray Hill, Midtown East and Columbus Circle have relatively higher duration-to-distance ratios for pickups.

Map based on Pickup Longitude and Pickup Latitude. Size shows average of Duration-to-Distance.

Answer Slide

4. The neighbourhood of Union Square, Midtown East, Midtown Manhattan, Murray Hill, Flatiron District and Nomad have relatively higher duration-to-distance ratios for dropoffs.

Higher Density duration_to_distance ratio for dropoffs

Avg. Duration-to-Distance
- 3
- 100,000,000
- 200,000,000
- 308,292,940

© 2022 Mapbox © OpenStreetMap

Map based on Dropoff Longitude and Dropoff Latitude. Size shows average of Duration-to-Distance.

Answer Slide

5. Meatpacking District, Upper Eastside, Lennox Hill, Upper Westside and Kips bay have relatively higher duration-to-distance ratios and density for pickups and dropoffs and hence selected.

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

From 6pm – 10 pm the ride pickup volumes are very high, 6pm and 7pm it being the highest.

From Wednesday till Sunday it's comparatively high, Saturday being the highest in terms of ride pickups.

There is no seasonality in trend when the ride pickup volume is observed throughout the year.

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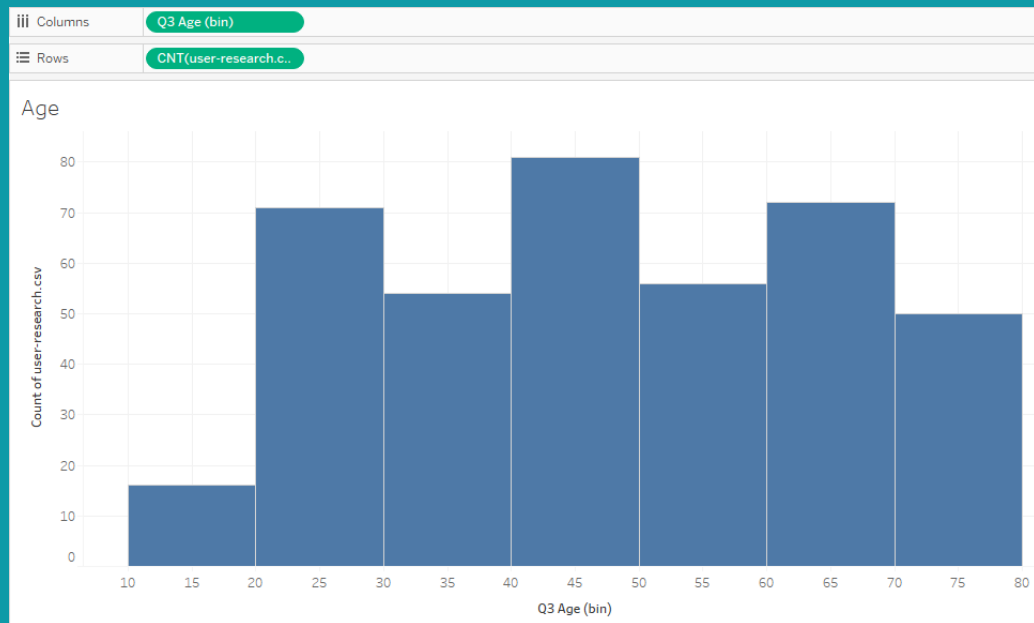Q10 - If no to Q8, what is the reason?

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

Yes, there is an inclination. 243 females compared to 157 males favoured it. The Age groups 40-50, 60-70 and 20-30 favoured with 40-50 being the most favourable. The income groups from $40,001 till $200,000 favoured with $40,001 - $80,000 favouring the most and the neighbourhood of Midtown, Battery Park City and Financial District favoured with Midtown on top.

Answer Slide

Pages

Columns | Q4 Income

Rows | Q8 Flying Taxi

Filters

Q8 Flying Taxi: Y

Income

	Q4 Income					
Q8 Flying Taxi	$40,001 - $80,0..	$80,000 - $120,..	$120,000 - $2..	$20,001 - $40,0..	> $200,000	$0 - $20,000
Y	106	94	85	53	42	20

Marks

Automatic

Columns

Rows | Q5 Neighborhood

Neighbourhood

Q5 Neighborhood

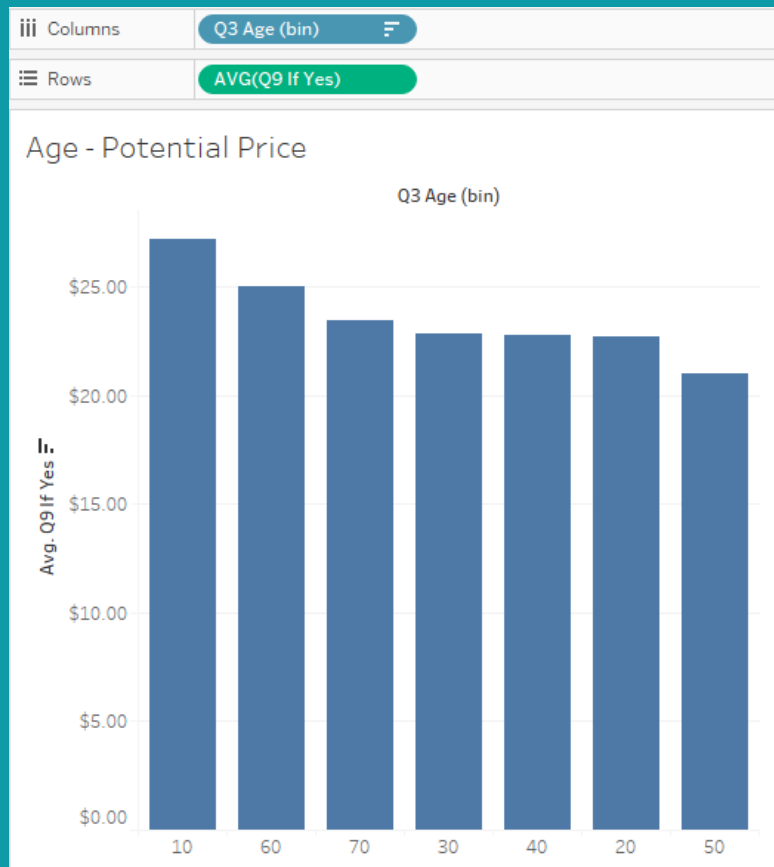| Midtown	15
Battery Park City	14
Financial District	12

Answer Slide

For Potential Price, the distribution was almost same for both females and males viz. $23.29 and $23.01 respectively. While the age groups 10-20, 60-70 and 70-80 had higher prices with 10-20 leading with $27.19. For income groups and price there was strong positive correlation with price increasing with increasing income groups and > $200,000 registering $37.10 price. The neighbourhood of Inwood, Financial district and Bowery were the top three with Inwood registering $34.00.

Columns	
Rows	Q2 Gender

Gender - Potential Price

Q2 Gender	
Null	
F	$23.29
M	$23.01

Columns	
Rows	Q5 Neighborhood

Neighbourhood - Potential Price

Q5 Neighborhood	
Inwood	34.00
Financial District	32.83
Bowery	32.80

Answer Slide

Income - Potential Price

Columns: Q4 Income
Rows: AVG(Q9 If Yes)

Age - Potential Price

Columns: Q3 Age (bin)
Rows: AVG(Q9 If Yes)

Answer Slide

Females, the age groups 20-30, 30-40 and 50-60, the Income groups $40,001-$80,000 and $80,000-$120,000 and users from the Neighbourhoods West Harlem, Tribeca and Sutton Place denote the Personas with Negative sentiments towards the Flying Car Project.

Personas - Gender

Q10 If No	Q2 Gender F	M
I think it is more unsafe than taxis	16	7
Wouldn't trust the person flying it	11	4
It would be too expensive	8	7
Not worth the extra money	6	7
It's unsafe	6	2
My commute is already straightforward	4	2
It's dangerous	5	1
The skies would get just as crowded as the streets	2	3
I think it's more dangerous	4	1
My commute is fairly efficient	2	1

Answer Slide

Personas - Age
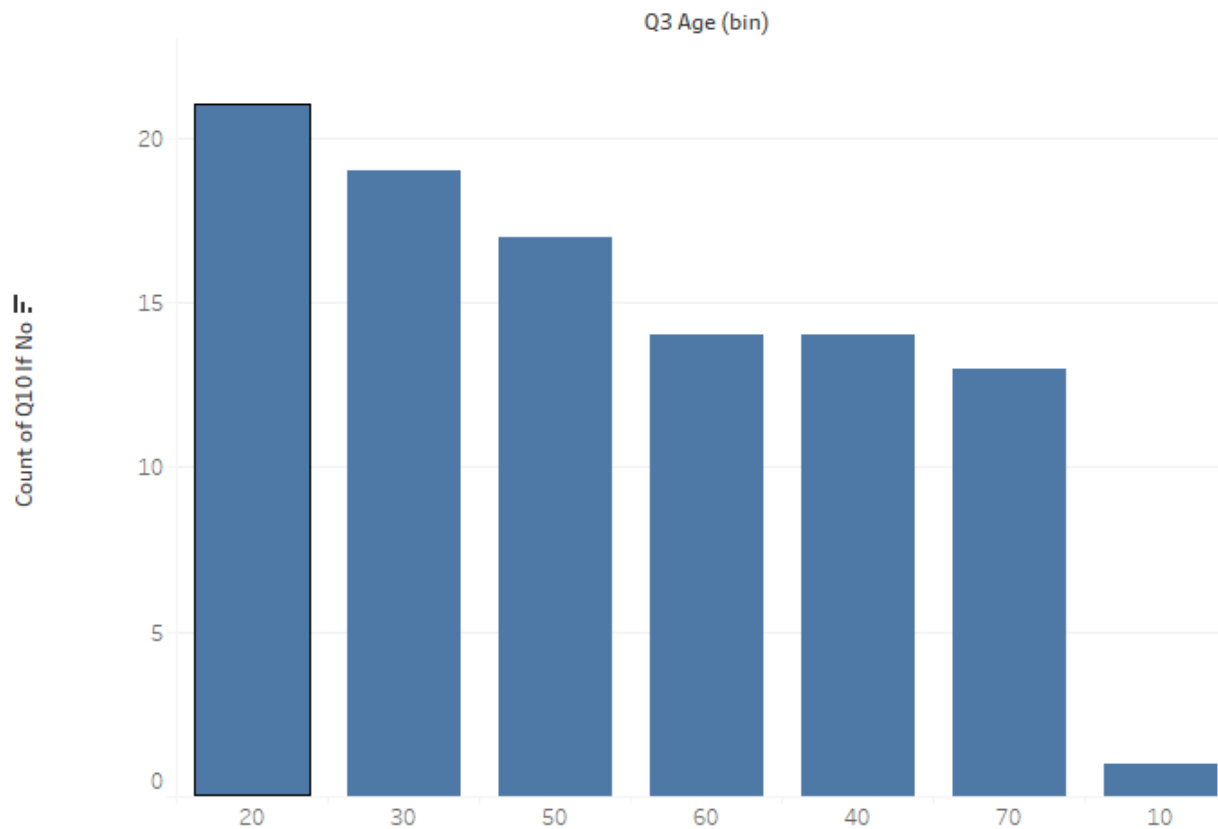
Answer Slide

Answer Slide
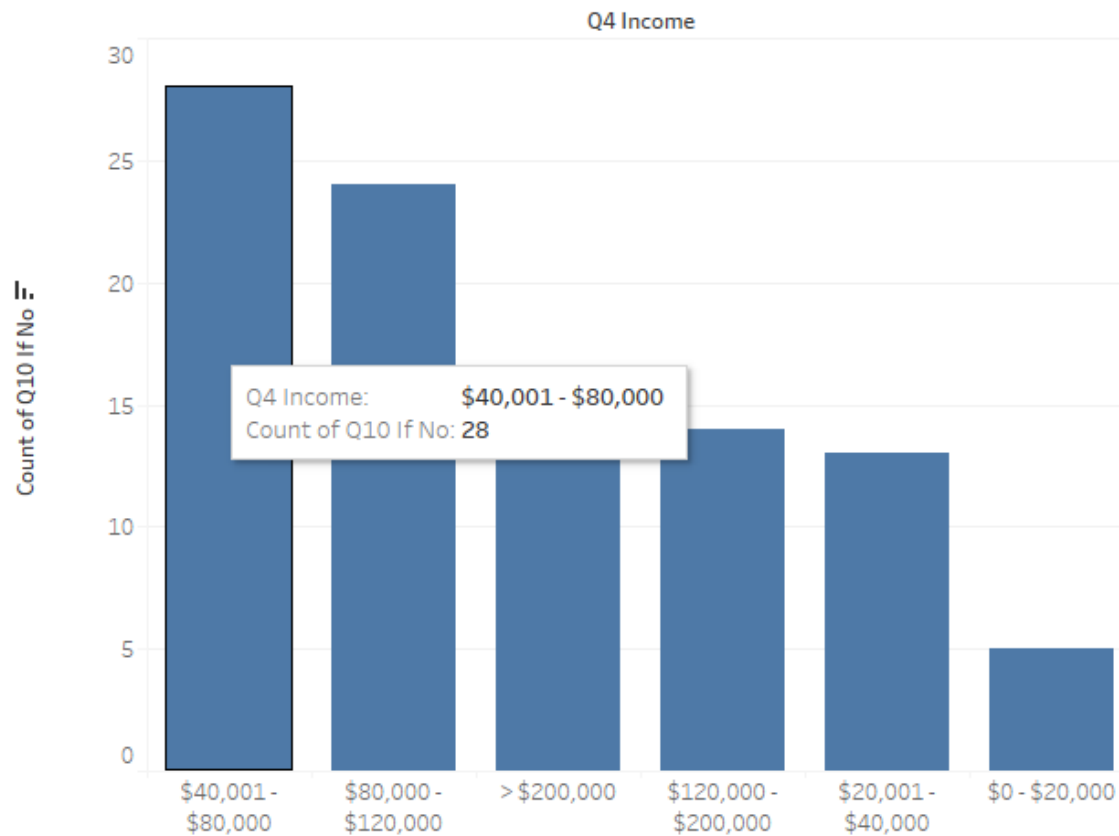
Personas - Neighbourhood

Q5 Neighborhood	
West Harlem	5
Tribeca	5
Sutton Place	5
Midtown	5
Times Square	4
Tenderloin	4
Inwood	3
Herald Square	3

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

Flyber should focus on User Acquisition as it's in the first stage of Product Development Cycle and once it has the expected user base then only it can focus on the remaining focus areas.

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

Answer Slide
KPIs

1. We need to monitor DAU (Daily Active Users) and MAU (Monthly Active Users).

2. Net Promoter Score (NPS)

3. No. of Safe Drives (SDn)

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

Answer Slide

1. Our estimated daily users as per the flight duration average of 953 secs is 210 and flight days per month is 20 hence DAU <= 210 & MAU <= 4200

2. NPS <=10 (Learnt from Marty Cagan's Book "Inspired" that states an NPS score of 5 is more than enough)

3. Estimated Safe Drive nos. SDn <= 105, wherein 105 is the estimated total rides per day if we take the flight duration average of 953 secs for 4 hours daily with 7 pickup/dropoff nodes.

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

1. As per the insights derived in Slide no. 27 the Flyber services should operate during 6pm – 10pm from Mid-week till weekend (Wed.–Sun.)

2. We should start with 7 pick-up/Drop-Off nodes. Refer Slide nos. 23 & 24.

3. The nodes should be at Midtown Manhattan, Nomad, Murray Hill, Midtown East, Columbus Circle, Union Square and Flatiron Dist. Refer Slide nos. 23 & 24.

4. We should use Copters eventhough it's expensive than homegrown hardware as Safety and Reliability are paramount. Later we can think of reducing cost in other ways.

5. We should go for Dynamic pricing as evident from insights from User research. Refer Slide no. 33.

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

Answer Slide

As per one article, 5-20% new users will go for a company's new product. Hence min. value I have taken is 5% and the baseline 20% for new users for our new project Flyber. The calculated sample size is 25000 and estimated duration = 25000/Est. daily users = 25000/210 = 119 days. 119 days equivalent to 6 months as we are operating 20 days a month. Refer Slide No. 47 for estimated daily users.

Baseline Conversion Rate

20 %

Your control group's expected conversion rate. [?]

Minimum Detectable Effect

5 %

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

Statistical Significance

95%

Edit

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

Sample Size per Variation

25,000

Sample size calculator

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

Answer Slide

Regarding Instrumentation plan, we should collect data for the Events:

1) userBooked - The event is triggered when the user books a ride on the Flyber app.

2) userPickedUp – The event is triggered when the Driver selects the option "User is Picked up" on the Flyber app.

3) userDropped - The event is triggered when the Driver selects the same on the Flyber app.

4) userCancelled – The event is triggered when the user cancels the ride on the app.

Answer Slide

5) userNpsScore – The event is triggered when the user gives a rating on the question of "How likely will you refer Flyber to a friend or acquaintance?" at the end of the Trip on the Flyber app.

Event Properties: userBooked – 1) user_id
                     userPickedUp – 1) passenger_count 2) user_id
                     3) driver_id 4) pickup_timestamp 5) pickup_lat
                     6) pickup_long
                     userDropped – 1) passenger_count 2) user_id
                     3) driver_id 4) dropoff_timestamp 5) dropoff_lat
                     6) dropoff_long
                     userCancelled – 1) user_id 2) user_cancelled_soon
                     3) user_cancelled_later (For future analysis)
                     userNpsScore – 1) user_id 2) user_nps_score

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

Answer Slide

1) Will you take Flyber again? If no, then take another survey path.

2) Did you find Value and Enjoyment in the flight?

3) Did you face any issues during flight? If no then go to question 5

4) What issues did you face and how it can be resolved?

5) Will you refer your friend for Flyber service?

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

The Target users are from the neighbourhood of Midtown Manhattan, Nomad, Murray Hill, Midtown East, Columbus Circle, Union Square and Flatiron Dist. where the duration-to-distance ratios for Pick-up/Drop-offs are the highest. The pain-points being time and money lost as not able to reach on time and the associated physical and mental stress.

Answer Slide

<u>Flyber Product Proposal</u>

Problems:

Target Users – Users from the neighbourhood of Midtown Manhattan, Nomad, Murray Hill, Midtown East, Columbus Circle, Union Square and Flatiron Dist.

Pain Points – Lot of Time and Money also wasted, accompanied with Physical and mental stress.

Justifications:

User Impact – Will save the users time, get VFM and gets rid of the commute related stress. They can focus on the important things in life.

Answer Slide

Market Impact – It opens up a new market for other players also. It forces the taxi and ride-sharing industry to up their game to provide quality services to the users to remain competitive. This creates multiple options for the users and an increase in user traffic too.

Business Impact – Flyber generates good revenue as it solves the customer pain points. It's current features as well as upcoming features based upon user research will pave for customer retention and reduce any future user churn.

Solution:

Risks – The Flyber Service has an air-safety risk which can be mitigated by the usage of reliable, not–so-old copters. Also experienced and retired pilots should be hired. Thorough screening of onboarding passengers to be done.

Answer Slide

Assumptions: Flyber will usher in a viable, usable and feasible solution providing value to the customers as POA in the next slide shows the market potential for 0-2 passenger segment being the highest. The target users are corroborated in the adjoining slide too.

The likely adoption of Flyber service is demonstrated in Slides 31 & 32.

The Flyber service will need to work with the Air Traffic Services of FAA (Federal Aviation Administration) closely for a safe user experience.

Finally Flyber service will increase the User Experience as well as improve the Transportation Industry.
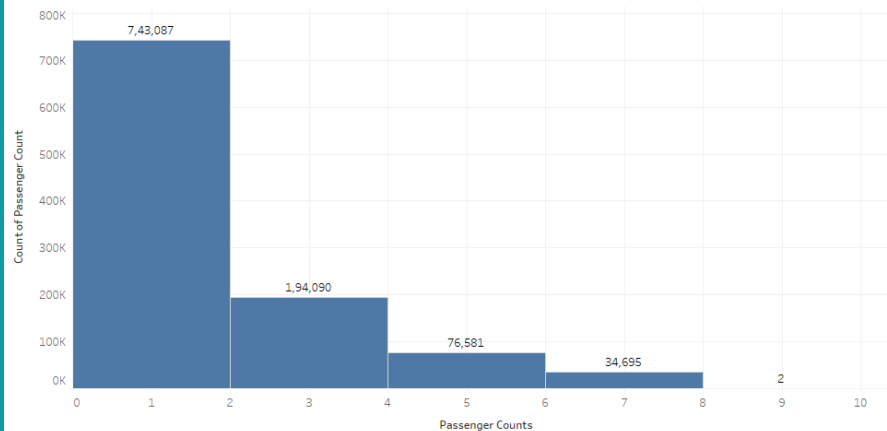
Answer Slide

Higher Density duration_to_distance ratio for dropoffs

Avg. Duration-to-Distance
- 3
- 100,000,000
- 200,000,000
- 308,292,940

based on Dropoff Longitude and Dropoff Latitude. Size shows average of Duration-to-Distance.

Passenger count bins

7,43,087
1,94,090
76,581
34,695
2

The trend of count of Passenger Count for Passenger Counts. The marks are labeled by count of Passenger Count.