

Data Product Manager Nanodegree

Applying Data Science to Product Management

Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service

Welcome to your first week at Flyber

Flyber

In this project, you will apply the skills acquired in this course to create the MVP launch strategy for the first flying car taxi service, Flyber, in one of the most congested cities in America -- New York City.

You are responsible for bringing the first flying car taxi service to market by analyzing data and building a product proposal.

You will need to use the SQL workspace provided in the Classroom, and [Tableau Public](#), in order to successfully complete the project.

You'll present your answers, findings, and insights in the Answer Slides found in this deck. Feel free to include any additional slides, if needed.

Section 1: Data Exploration

Back to the basics of product management, identify your customer and their pain points:

- What are taxis used for?
- What are the characteristics of the users that leverage them?
- What are existing pain points with taxis?
- What are the existing pain points with digital ride-sharing services?

Answer Slide

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. $\text{Price} = ((\text{Distance}) * 1.75) + ((\text{Duration}) * (0.35/60))$ 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.

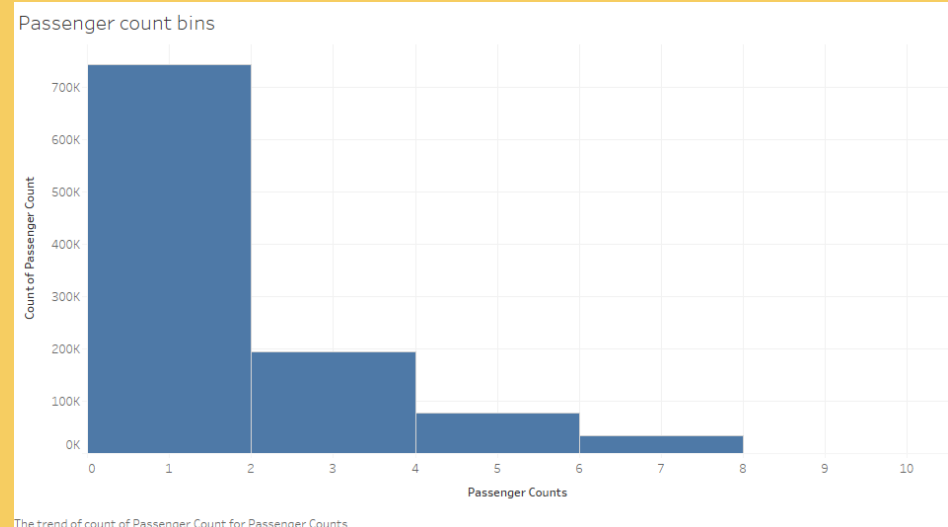
Price: avg – \$11.57, median – \$7.72 , 1st std. dev – \$20.82, 2nd std. dev – \$41.64.

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 competitor in the bin (2-4) had 1,94,093 passengers.



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.

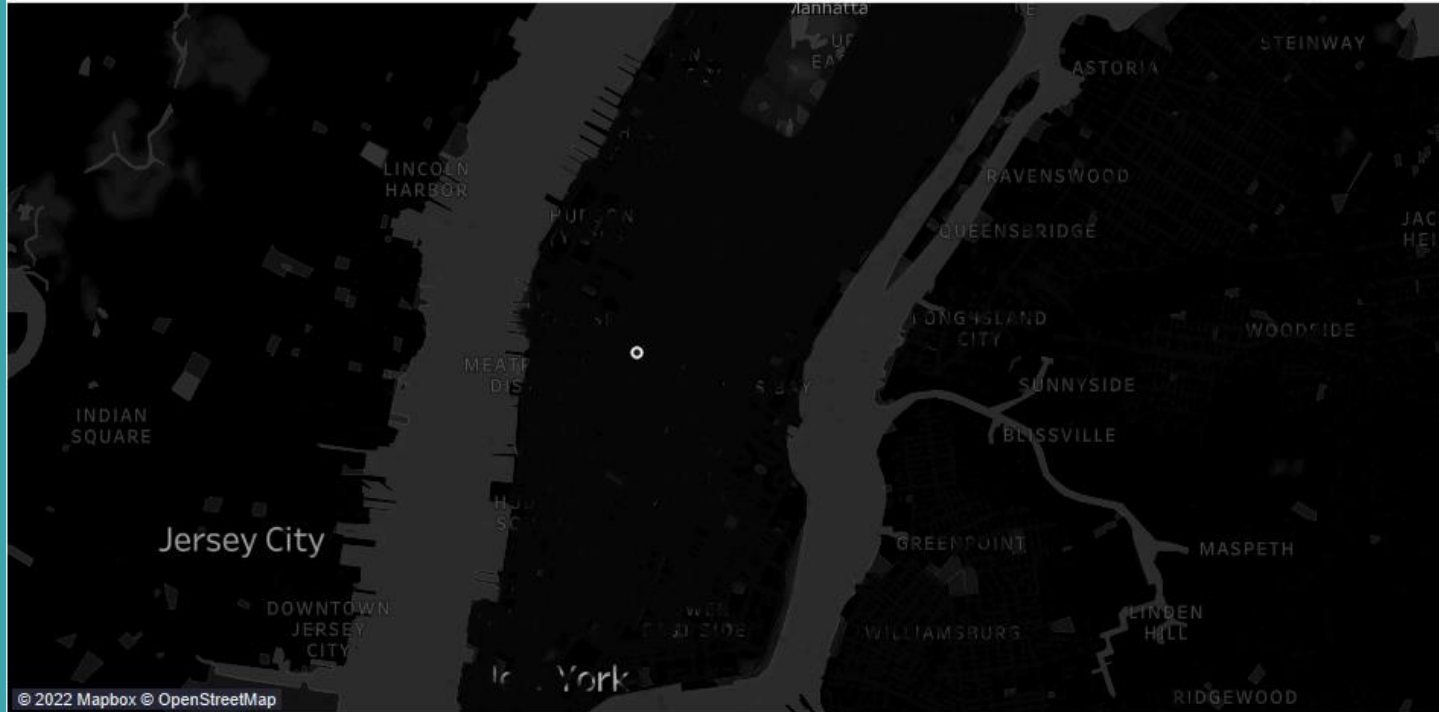
Higher Density Pickups



Map based on Pickup Longitude and Pickup Latitude.

Answer Slide

Higher Density Dropoffs



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.

Higher Density duration_to_distance ratio 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



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?

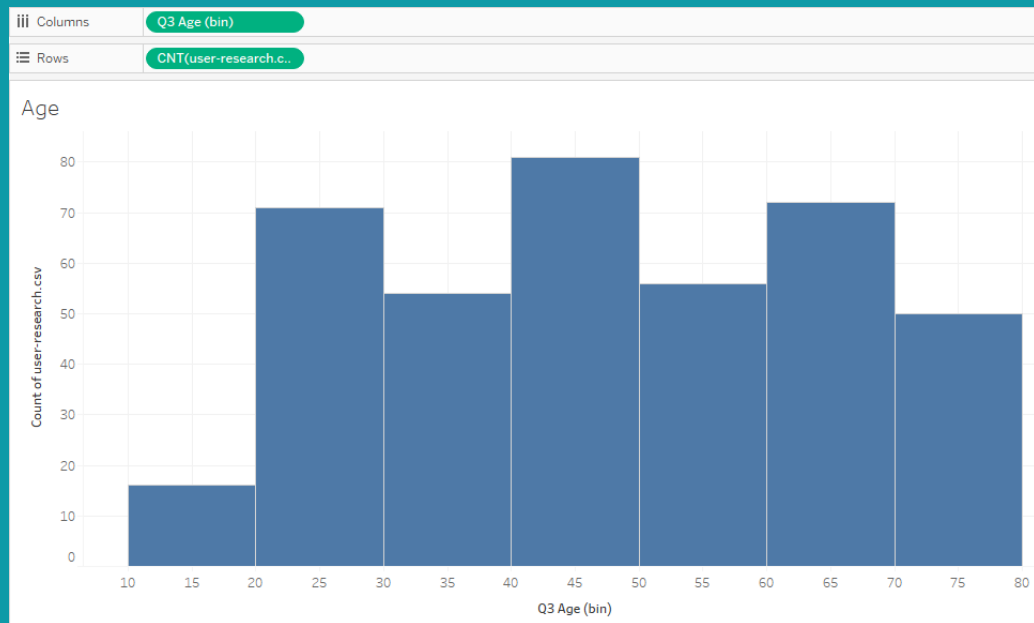
To inform our future product marketing efforts, we'll want to extract the following:

- 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.

Columns			Q2 Gender	
Rows			Q8 Flying Taxi	
Gender			Q2 Gender	
Q8 Flying T..			F	M
Y			243	157



Answer Slide

Pages

Columns

Q4 Income

Rows

Q8 Flying Taxi

Filters

Q8 Flying Taxi: Y

Marks

Automatic

Income

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

Columns

Rows

Q5 Neighborhood

Neighbourhood

Q5 Neighborhood	
Midtown	15
Battery Park City	14
Financial District	12

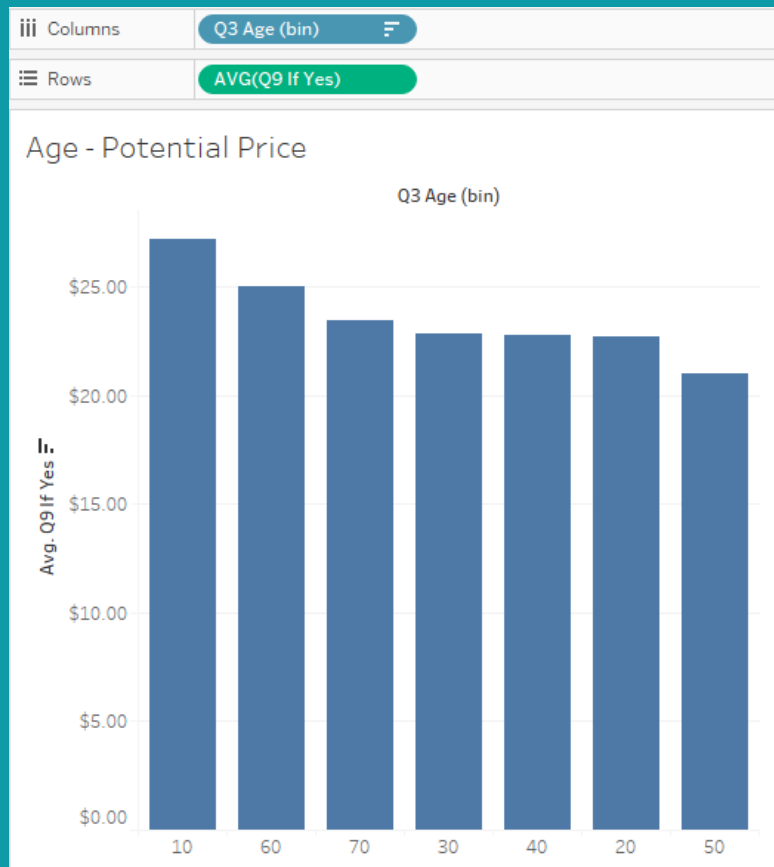
Answer Slide

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



Answer Slide

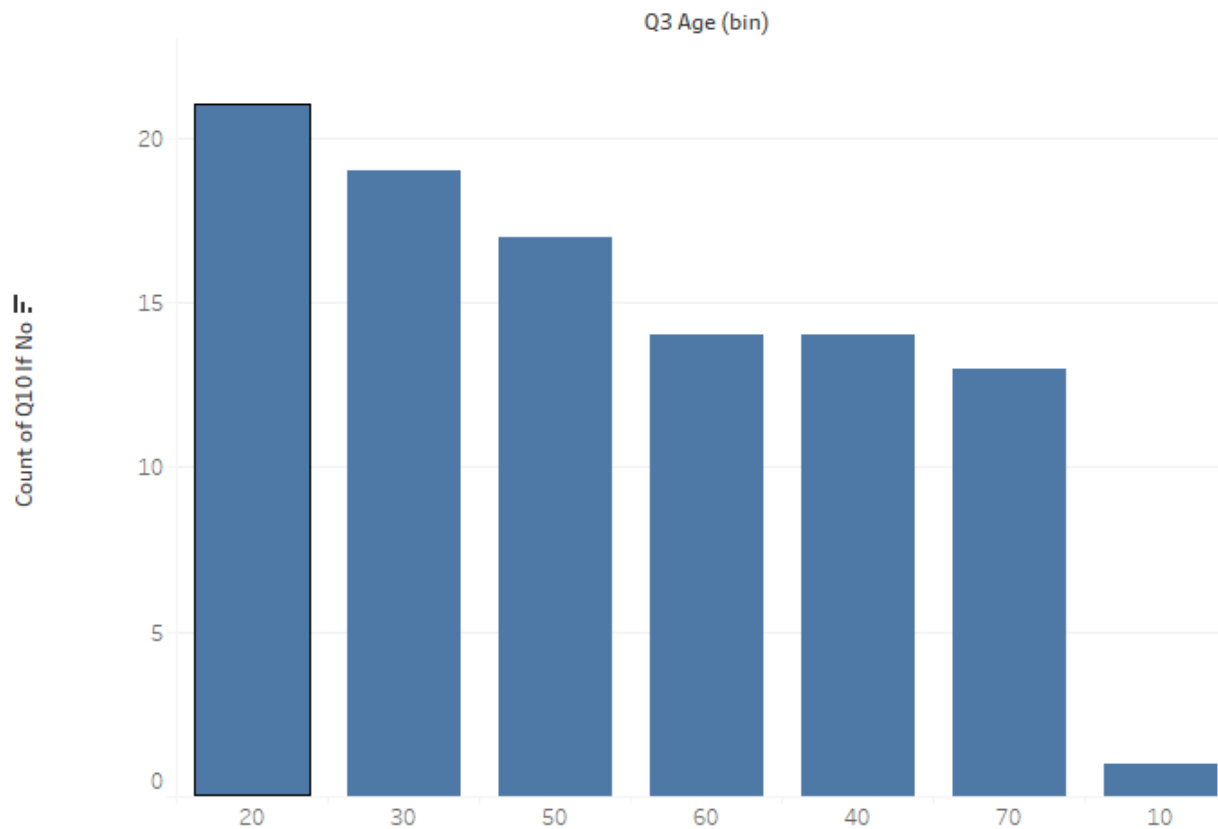
Females, the age groups 20-30, 30-40 and 50-60, the Income groups \$40,001-\$80,000 and \$80,000-\$120,000 and users from the Neighbourhoods West Harlem, Tribeca and Sutton Place denote the Personas with Negative sentiments towards the Flying Car Project.

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

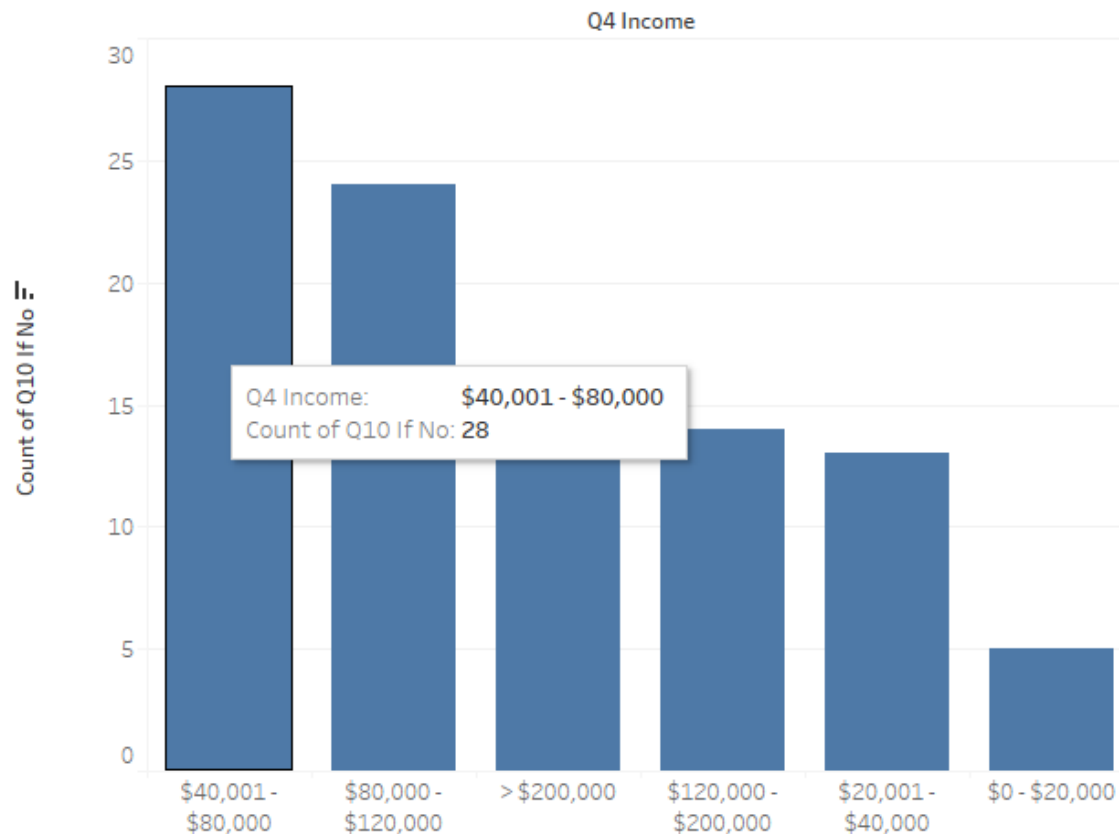
Answer Slide

Personas - Age



Answer Slide

Personas - Income



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

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

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

Answer Slide

(Fill out your answer here)

Summarize everything you have learned into your final proposal

- Identify the target population. Why did you select that target population? What are their pain points?
- Create a product proposal containing claim, evidence, estimated impact, and risks
- Claims should be backed by quantitative evidence, impact should assess market needs/benefits
- Risks involve any known unknowns that we'll still need to monitor post-launch
- State cross-functional stakeholder teams that will need to be involved

Answer Slide

(Fill out your answer here)

Answer Slide

(Fill out your answer here)

Answer Slide

(Fill out your answer here)