Data Product Manager Nanodegree

Applying Data Science to Product Management
Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service

Mark Shannon

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

According to our data sets:

- 79.15% of females interviewed currently use taxis.
- 84.90% of males interviewed currently use taxis.

- Of those who currently use taxis AND ride-shares the highest % earn between $40,000 to $200,000 (74.30%)
  - 31.78% earn $40,001 - $80,000
  - 21.96% earn $80,001 - $120,00
  - 20.56% earn $120,001 - $200,000

Taxi Pain Points:

- Taxi drivers often struggle with the high distance driven without passengers vs distance driven with passengers. (Example) In 2018 Seattle's average daily km driven per taxi was 283km, of which only 30% of the distance was with passengers.
- Taxi's bring uncertainty for customers who try to hail them. Unlike TNC or digital-based services you are not able to pre-book, or be aware of the taxis that are in your area. This can bring pain to those who are in a rush.

Ride-sharing pain points:

- Surge pricing and limited availability are two pain points that customers face. In dense cities at popular times, prices can rise 2x as demand exceeds supply. With a more efficient service like flying taxis, we could aim to provide quicker trips, allowing pilots to have more trips per hour.

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

I believe that a flying taxi service could allow drivers to have more trips per hour, reducing the amount of surge pricing for customers as pilots can pick-up and drop-off more customers than with traditional taxi services.

I believe that the distance to duration will be lower, as there will be less traffic for flying taxis. This will again lead to more trips per hour and a better service for customers.

Overall trip duration will also be less as flying taxis will be able to avoid the traffic by using the skies however this might be impacted by less areas that flying taxis can pickup and dropoff.

Customers should end up paying less for their trips due to less distance to duration and overall shorter trips unless there is a premium added due to a more expensive method of transport.

If flying taxis were a success, roads would become less busy with vehicles due to people using the skies. This could lead to improved road conditions or more pedestrian areas, resulting in a better urban experience for people who live there.

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

- There are 1,048,468 records in the taxi_rides.csv.
- Each record represents a taxi ride and includes properties about that taxi ride including pick up, drop off, date times etc
- Primary Key = ID
- January 2016 to July 2016
- Queens seems to be where the data coverage is less and this could be due to less taxis been taken there or due to a lack of data in that area. Brooklyn is another area where the data starts to reduce. Staten Island has very few data points. The Bronx also contains a lot less data points. The majority of the data is focused on Manhattan.
    - There are some outliers close to North New Jersey State.
    - There are also some outliers in the sea, which we can assume is bad data.

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
- distance-to-duration ratio
- price

Answer Slide

According to our data sets:

Average duration = 962 seconds
Median duration = 662 seconds
1st deviation = 5853 seconds
2nd deviation = 2104 seconds

Average distance = 3.442 miles
Median distance = 2.095 miles
Standard deviation = 4.382 miles
2nd deviation = 11 miles

Average passenger count = 1.664 passengers
Median passenger count = 1 passenger
1st standard deviation = 1.314 passengers
2nd standard deviation = 5 passengers

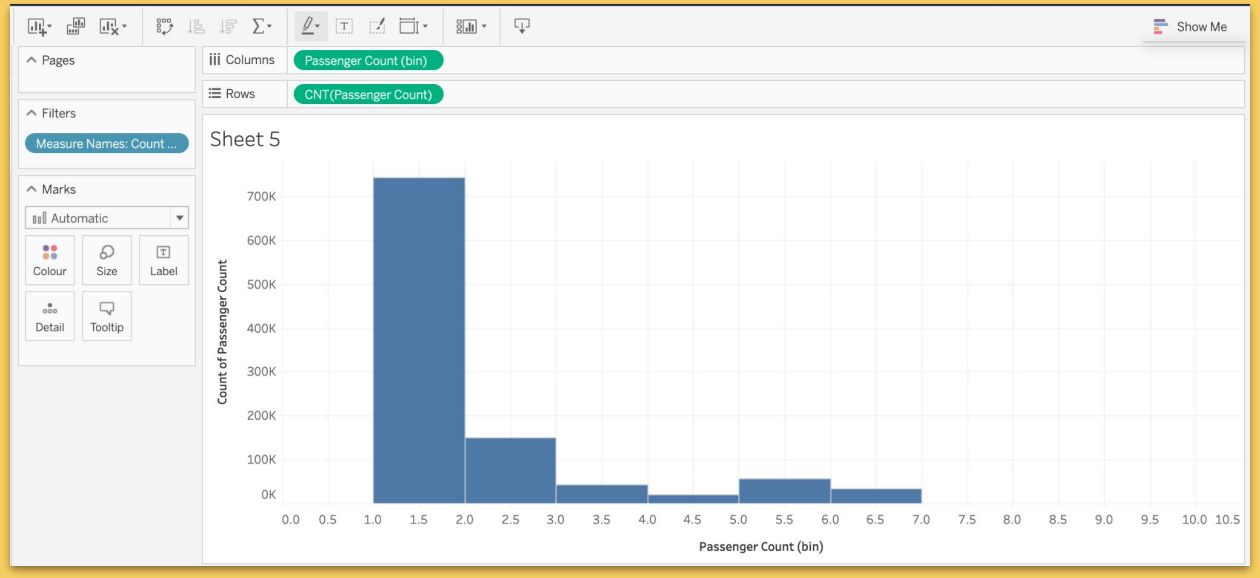
Average duration to distance = 6.129 minutes per mile
Median duration to distance = 4.636 minutes per mile
Standard deviation = 35.19
2nd deviation 11.21

Average Price = $19

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 visual for the number of total rides grouped by passenger counts to analyze the potential market volume of low passenger pickups (1-2 passengers).

Passenger count per rides



- 743,000 of the total rides had just 1 passenger.
- 151,060 of the total rides had 2 passengers
- 43,033 of the total rides had 3 passengers
- 20,334 of the total rides had 4 passengers
- 56,249 of the total rides had 5 passengers
- 34,692 of the total rides had 6 passengers

- The total addressable market for 1-2 passengers is **894,060**

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

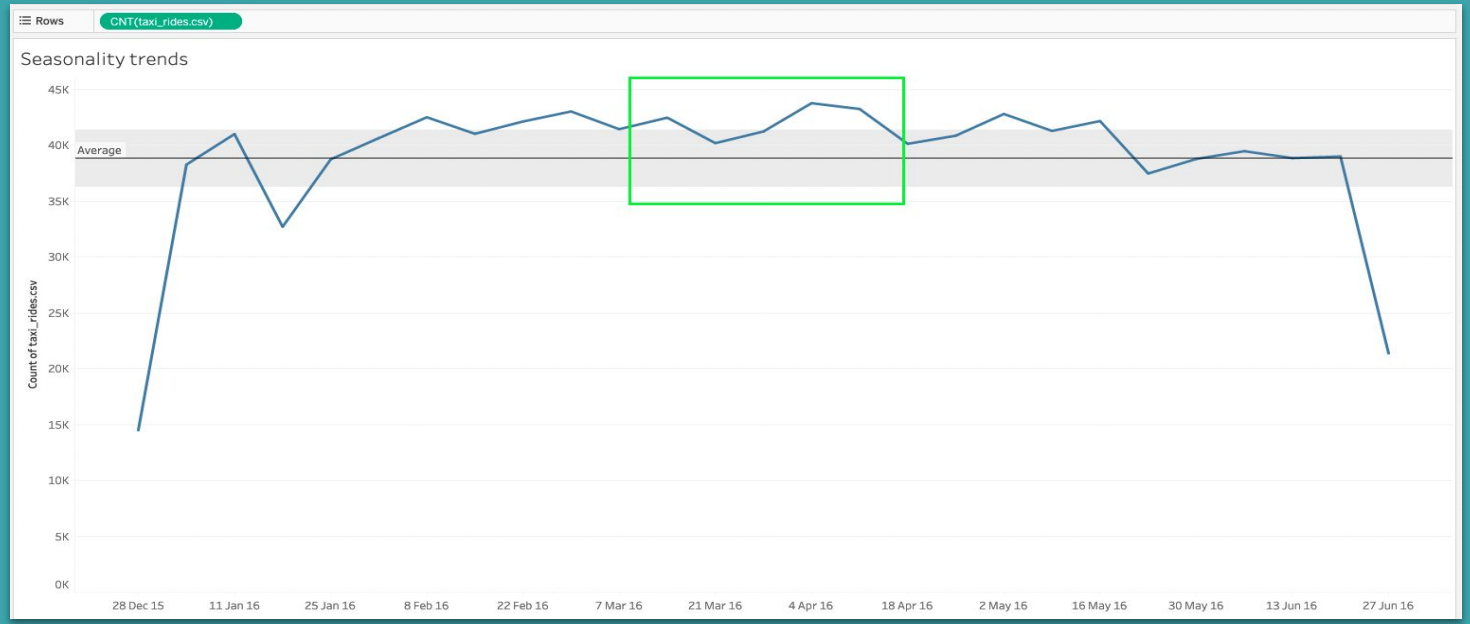
- High density pickups per neighbourhoods -> Upper West Side, Upper EastSide, Hells Kitchen, Chelsea, Flatiron District, Gamecry. 10281, 10280, 10282, 10069, 10023
- High density drop-off per neighbourhoods. -> Upper West Side, Upper EastSide, Hells Kitchen, Chelsea, Flatiron District, Gamecry. 10281, 10280, 10282, 10069, 10023
- Garment district (Madison Sq Garden) has one of the highest duration to distance, based on pick up. Hells Kitchen. La Guardia Airport and JFK.
- Hell's Kitchen, Garment District and La Gaurdia have the highest duration to distance, based on drop-off.
- I don't believe the airport locations mentioned previous are suitable for flying taxis due to airport and flying regulations. One potential good spot could be at the West Midtown Ferry terminal within Hell's Kitchen. This location would have space due to the already high passenger throughput and be able to hand pickup and drop offs of customers. It would also be a good place for new customers, due to the footfall. According to data, Since launching on May 1 2017, NYC Ferry has transported over 7.7 million passengers (as of the end of 2018).

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

- 6pm to 10pm experience the highest pick-ups according to the data.
- Wednesday, Thursday, Friday experience the highest number of pickups
- Late March and Early April are the best months for launching our MVP service.



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

- 25% of females said **no** to flying taxi service, whilst 23.53% of male said **no**. 75% of females said **Yes**, whilst 76.47% of males said **Yes**. Very minimal differences.
- 75%-80% of the those who live in ChinaTown, Garment District, Koreatown and Little Italy have said Yes to using a flying taxi service. 50% of those living in Lower East Side have said Yes. There is no real indication of a strong pull towards any neighbourhood.
- Those people aged 20-25 and 40-45 plus those aged 60-70 are the highest adopters.
- Those making $40,001 to $200,000 were more inclined to use a flying taxi service.
- 15% of all Yes answers were people living in 10282, close to Tribeca. Some of the other high results were 10021, 10023, 10024, 10025

- The negative sentiment seems to focus mostly on the safety and trust of the flying taxi service. Some people also already have a straightforward commute and answered no, so that is fine.

Tableau Link

https://public.tableau.com/views/FlyberLaunchStrategy/Sheet1?:language=en-GB&:display_count=n&:origin=viz_share_link

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

User acquisition + retention: Flyber want to become the number 1 taxi service used in high density areas of NYC, through the repeat booking of riders who want to get to their destination faster.

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

Answer Slide

3-5 KPIs for Flyber:

Duration to Distance ratio - to help us understand the time taken to get from point A to B.
Weekly active riders per neighbourhood - to help us measure our impact in each neighbourhood and especially the high density areas.
Retention rate per neighborhood - to help us measure our churn in the neighbourhoods that are important to us.
Gross bookings

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

Answer Slide

Duration to Distance ratio -> We believe that reducing the duration to distance ratio for customers in high density areas will improve customer and driver experience, therefore increasing the retention rate of those customers.

Duration to Distance ratio -> We believe that reducing the duration to distance ratio for customers in high density areas will reduce surge charging at peak hours, allowing customers to travel for less and providing more trip opportunities for pilots.

Weekly active riders per neighbourhood -> We believe that if we increase our weekly active riders, month over month, we will move towards becoming the number one taxi service in high density areas

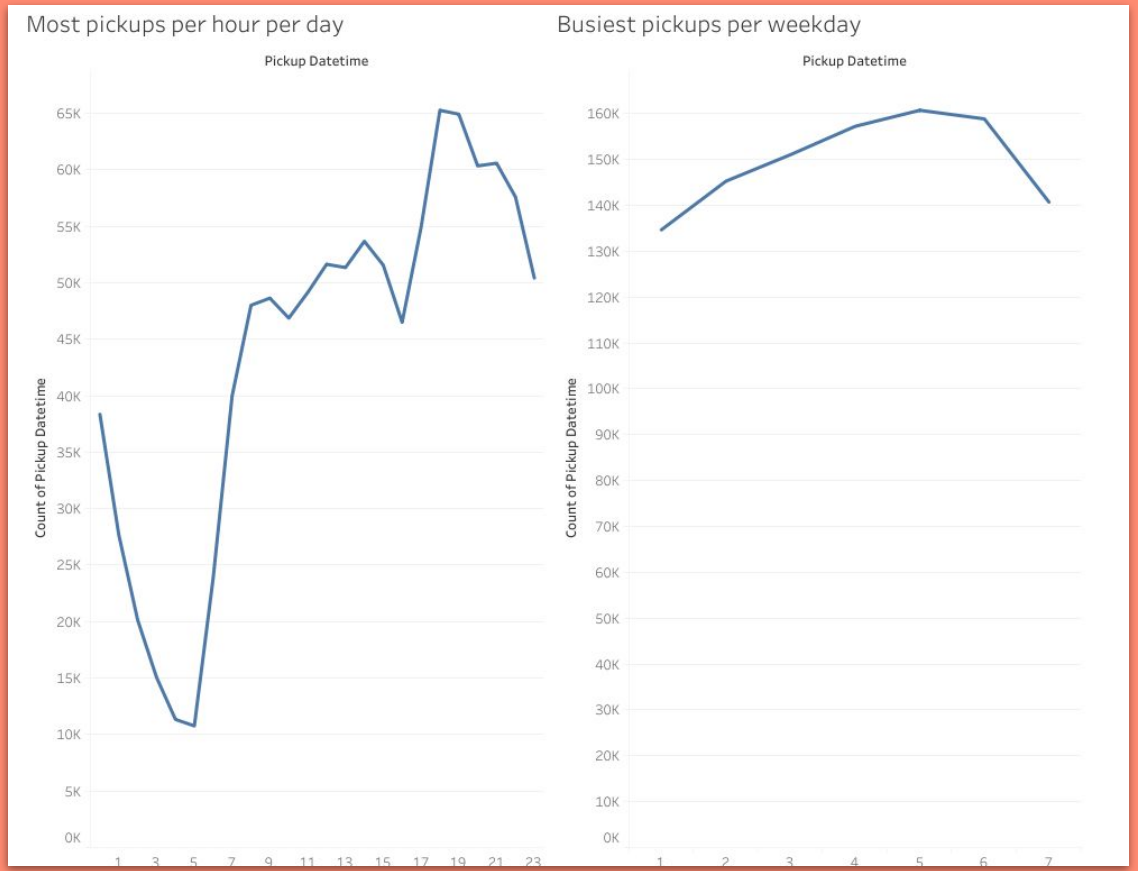
Retention rate per neighborhood -> We believe that increasing our retention rates in high density neighborhoods will help to increase our market share in those areas.

Gross bookings -> We believe that if we increase our week on week gross bookings, flyber will move towards becoming the number one taxi service in high density areas

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?

Best times to operate our service.



According with our data the best days to operate would be Wednesday through Saturday.

The best times to operate could be:

10am through 2pm

4pm through 7pm

How many pick up / drop off nodes?

- High density pickups per neighbourhoods -> Upper West Side, Upper EastSide, Hells Kitchen, Chelsea, Flatiron District, Gamecry. 10281, 10280, 10282, 10069, 10023
- High density drop-off per neighbourhoods. -> Upper West Side, Upper EastSide, Hells Kitchen, Chelsea, Flatiron District, Gamecry. 10281, 10280, 10282, 10069, 10023

Based on our goal to be the number 1 taxi service in high density areas of NYC we should aim to have at least 1 pick up and drop off node in each of these high density areas. That would be a total of 6, which would mean we are not heavily investing yet however we are validating our assumptions in the areas we need to be.

Should we initially use copters or homegrown hardware?

Depending on our circumstances and time to market with homegrown hardware, we might want to start with copters. This would provide us with the opportunity to appeal and change the mind of those who are not inclined to take a flying taxi service. Using more common and industry standard tech like a copter, might entice the 25% of females who said **no** to flying taxi service, and the 23.53% of male said **no** to take a ride before we move to homegrown hardware. This provides us with an opportunity to show them the benefits of not using traditional taxis and helping to build customer acquisition and retention.

Should the pricing be fixed or dynamic? Why?

We might want to look at a dynamic pricing model based on demand, that has a fixed base price. Our service will be more expensive to run per minute than a normal taxi service, but by reducing the duration to distance ratio, the end price for customer should still be less.

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

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

Using the quantitative research provided, in the zips that we propose for our drop off and pick up locations, we only have 16 and 5 people, who answered Yes to using a flying taxi service. That wouldn't provide us with a good sample size for MVP, however I recognise that the user research was not conducted with a large sample.

Based on our earlier research we know that from the data the following is true:

- 743,000 of the total rides had just 1 passenger.
- 151,060 of the total rides had 2 passengers

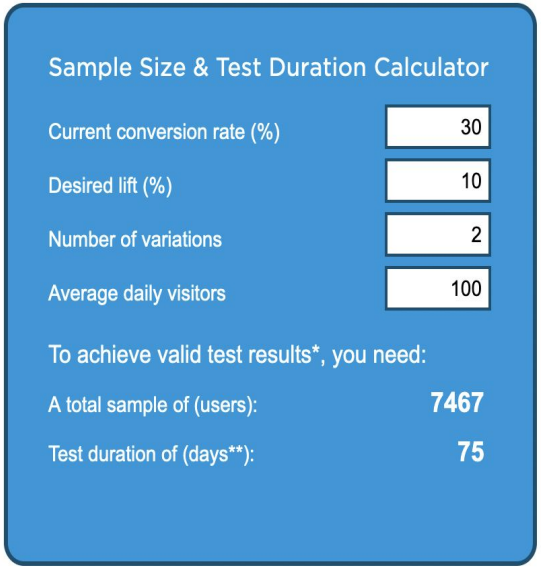
Our copters for the MVP might hold between 1-3 passengers so I believe choosing the sample size from these rides will be realistic. Using data from here I was able to determine that the total population for our high density areas is approx 80,000.

10023 - Total population; 60,762. Based on our earlier research we know that 20-25, 40-45 and 60-70 are a good target demographic. Based on this, there is about 52,364 of the population in our demographic. Using our 75% of Yes answers as an assumption we would have a target population of 39,273
10280 - Total Population 8,817. Based on our earlier research we know that 20-25, 40-45 and 60-70 are a good target demographic. Based on this, there is about 4,800 of the population in our demographic. Using our 75% of Yes answers as an assumption we would have a target population of 3,600
10282 - Total population 5903. Based on our earlier research we know that 20-25, 40-45 and 60-70 are a good target demographic. Based on this, there is about 2,400 of the population in our demographic. Using our 75% of Yes answers as an assumption we would have a target population of 1,800
10069 - Total Population 5,523. Based on our earlier research we know that 20-25, 40-45 and 60-70 are a good target demographic. Based on this, there is about 2,800 of the population in our demographic. Using our 75% of Yes answers as an assumption we would have a target population of 2100

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

Fill out the form and find out how long you need to run your test in order to get valid test results. This will help you avoid false positives and increase the quality of your A/B testing.

Pro tip! Play around with the calculator and get a better feel for how baseline conversion rate and desired lift affect sample size and test duration. E.g. a conversion rate of e.g. 2% vs. 10%. Or a lift of 5% vs 30%.



Based on a bench-mark 30% conversion rate and a desired lift of 10% e.g. moving from 30% to 33% - our sample size should be around 8000 people over the course of 75 days.

Based on that and using the population numbers from the previous slide we could look to use the following neighborhoods:

10280
10282
10069

They would give us close to the number of people required for our MVP validation.

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

Events orchestration for KPIs

Event: **userSignIn**
Definition: Triggers when a user signs into their account
Properties: passenger_name, passenger_id, device_type

Event: **tripRequested**
Definition: Triggers when a user requests a trip
Properties: passenger_name, passenger_id, device_type, pick_up_lat, pick_up_long, pick_up_time, trip_ID

Used to help understand how often a user books a trip. Using the trip_ID and passenger_ID allows us to understand if a rider repeats their bookings over a period of time.

Event: **tripAccepted**
Definition: Triggers when a driver accepts a user's trip request
Properties: passenger_name, passenger_id, device_type, pick_up_lat, pick_up_long, pick_up_time, drop_off_lat, drop_off_long, drop_off_time, trip_ID

Event: **tripStarted**
Definition: Triggers when a user's trip starts
Properties: passenger_name, passenger_id, device_type, pick_up_lat, pick_up_long, pick_up_time, drop_off_lat, drop_off_long, drop_off_time, trip_ID

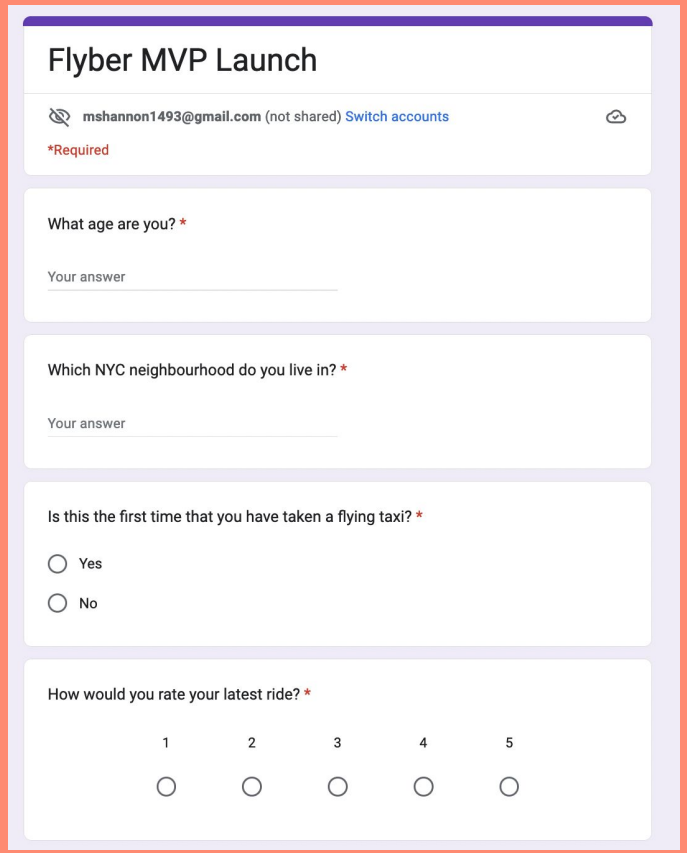
Event: **tripEnded**
Definition: Triggers when a user's trip ends
Properties: passenger_name, passenger_id, device_type, pick_up_lat, pick_up_long, pick_up_time, drop_off_lat, drop_off_long, drop_off_time,trip_total_distance, trip_total_duration, trip_total_price, trip_ID

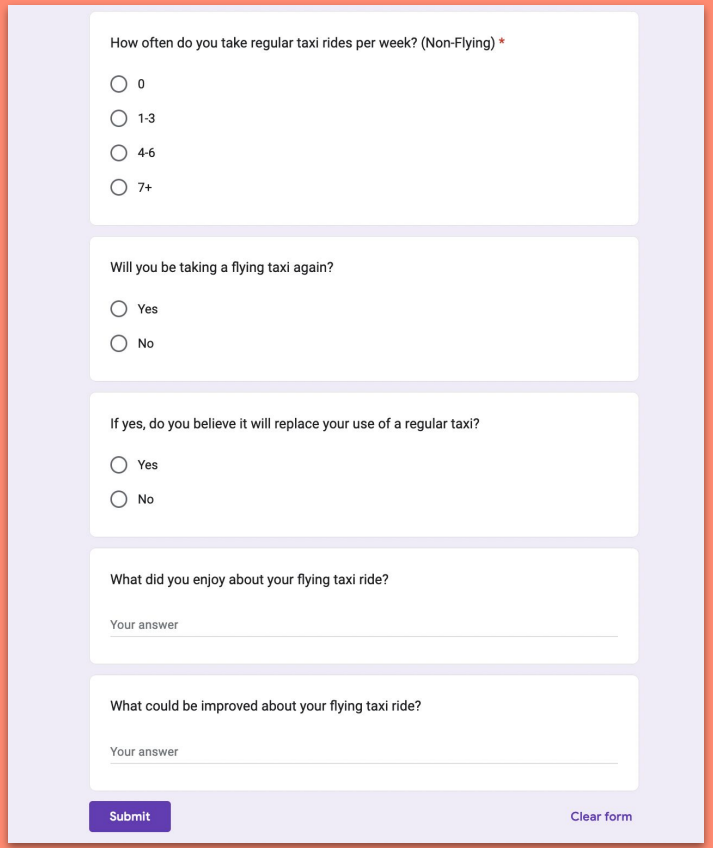
Used to help calculate the duration to distance ratio as well as the price, to help validate our KPI on reducing duration to distance ratio.

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

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

https://forms.gle/cHv3pZS2inM2mLcq9





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

Target Population

- As Flyber wants to become the number 1 taxi service in high density areas of NYC, we decided to target people aged between; 20-25, 40-45 and 60-70 in high dense neighborhoods of NYC. Our reasoning for the age population was due to earlier research pointing to those demographics being high adopters. We chose high dense areas as we believe that if we can reduce the distance to duration ratio in these areas we will deliver an excellent customer experience that will increase customer acquisition and retention.

Pain Points

- Taxi drivers often struggle with the high distance driven without passengers vs distance driven with passengers. An example from Seattle in 2018 showed that drivers drove an average of 283km per day, with 30% of that distance including passengers.
- Surge pricing and limited availability are two pain points that customers face. In dense cities at populate times, prices can rise 2x as demand exceeds supply. With a more efficient service like flying taxis, we could aim to provide quicker trips, allowing pilots to have more trips per hour.
- More efficient trips also provides our pilots with a better work life balance and revenue opportunity. They can drive for less total time and do more total trips.
- 23-25% of males + females surveyed said no to a flying taxi service. We aim to start with copters to prove our MVP, allowing us to build trust and change the mindset of those that are hesitant to use our service. This increases the accessible market available to us.

Average trip distance

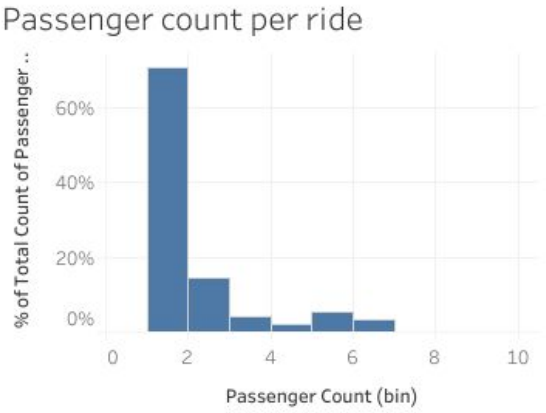
- The average distance per trip is 3.4miles. The average duration to distance is 6.1 mins per mile.

3.4mi

6.1mins /mile

The opportunity for pooling

- 84% of all taxi rides in NYC had between 1-2 passengers per ride. There is a huge opportunity for Flyber to deliver a pooling service that will increase ride efficiency.



Uber and Lyft prices rising

- Since 2018, Uber and Lyft prices **have risen 92%.** This can be attributed to less drivers + the distance to duration average at peak times reducing supply.

Addressable market size

- We believe that there is an addressable target market of around 45,000 people in just 4 NYC alone.

45,000

Our proposal

- We want to become the number 1 taxi service in high density neighborhoods of NYC by reducing the duration to distance ratio through the use of flying taxis. To reduce the risk, we will start our MVP service using copters and validate our assumptions with a target addressable market of 45,000 people over 2-3 months.

Key Risks

- Capital expenditure for pick up and drop off sites. Starting with copters means that we could use existing sites to validate our MVP however if we move to a more proprietary flying taxi service, we will need to likely build our own pick and drop off sites.
- Legislation. There are risks of legislation being brought forward that would restrict the areas that we could operate in due to safety. Similar to the legislation brought in against drones, we could be subject to new rules and laws.
- Ability to source pilots to meet demand. Our service will reduce distance to duration and therefore provide the ability to service more passengers however we will still need to meet demand of pilots. Training and license barriers could slow the recruitment and impact our ability to meet demand.

Who is involved?

- Product Manager - will oversee the product development required to deliver our MVP service.
- Engineers - will build and ship our MVP features including testing.
- Business sponsors - will sign-off on expenditures required to launch our MVP service.
- Marketing
- Training staff
- Data Protection Officer
- Legal