Applying Data Science to Product Management
Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service
Responsible for bringing the first flying car taxi service to market by analyzing data and building a product proposal.

By Preyash Parikh

Section 1: Data Exploration

Identifying your customer and their pain points

→ Taxis are used for facilitating public transport.
→ Few characteristics of user who leverages them are : Who do not have transportation facilities nearby the start of their journey, one who does not own some means of transport such as car, users who do not want to self ride because of congestion in New York City.

→ Pain points with the taxis are : It takes time to reach the destination due to congestion, Sometimes the cab drivers are not professionals.

→ With the digital ride sharing app, the cab is not allotted instantly and it takes time to arrive for the pickup. Also, the cab drivers do not know the destination before they pickup the passengers . The price of the fare fluctuates with change in congestion or demand or bad weather.

Improvements hypothesized a flying taxi service would have the existing taxi service industry & physical road infrastructure today?

→ In my opinion, flying taxis would save lot of travelling time of passengers And the fares would not depend on congestion,demand and bad weather. Also users will be able to travel longer journey with the help of flying taxi.

→ Employment would increase in transport and engineering sector. Less accidents and less traffic as number of cars on the road will be low. As taxis will be obsolete, car manufacturing companies would lower the car prices and this will benefit customers wishing to own a private car.

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!

Informed the basis for analyses by understanding granularity and scope of the dataset - Tableau, SQL

→ There are 1048468 records.
→ Each record represents taxi rides.
→ id -
Query ⇒ SELECT count(distinct(id)) FROM taxi_rides;

→ Max date range ⇒ 6/30/2016 11:59:00 PM
Min date range ⇒ 1/1/2016 12:00:00 AM

→ Geographical bounds are from Toronto to North Carolina to Concord and sea. New Jersey is not included. Most of data points are centralized near New York and yes, there are outliers which can be seen in the portion of water.

Enriched the Dataset

- Ride price was not included in the dataset, but I derived from New York taxi prices which is gleaned from the internet.

  Calculated fields : Price

  (3+(1.56*[Distance]*1.61)+([Duration]/3600)*30)

  Source : https://www1.nyc.gov/site/tlc/passengers/taxi-fare.page

- Hypothesized that 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, I created a calculated field called `distance-to-duration ratio`.

  Calculated fields : Duration to Distance Ratio

  IIF([Distance]>0.007 AND [Distance]<100 AND [Duration]>0.1 AND [Duration]<=101060, [Duration]/([Distance]),0)

  This field is adjusted after cleaning data

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

Understanding the scope and distribution of various dimensions within the dataset

Duration :
Average = 962.2 Seconds
Median = 662 Seconds
1 std Deviation = 5853 seconds
2 std deviation = 12668.2 seconds

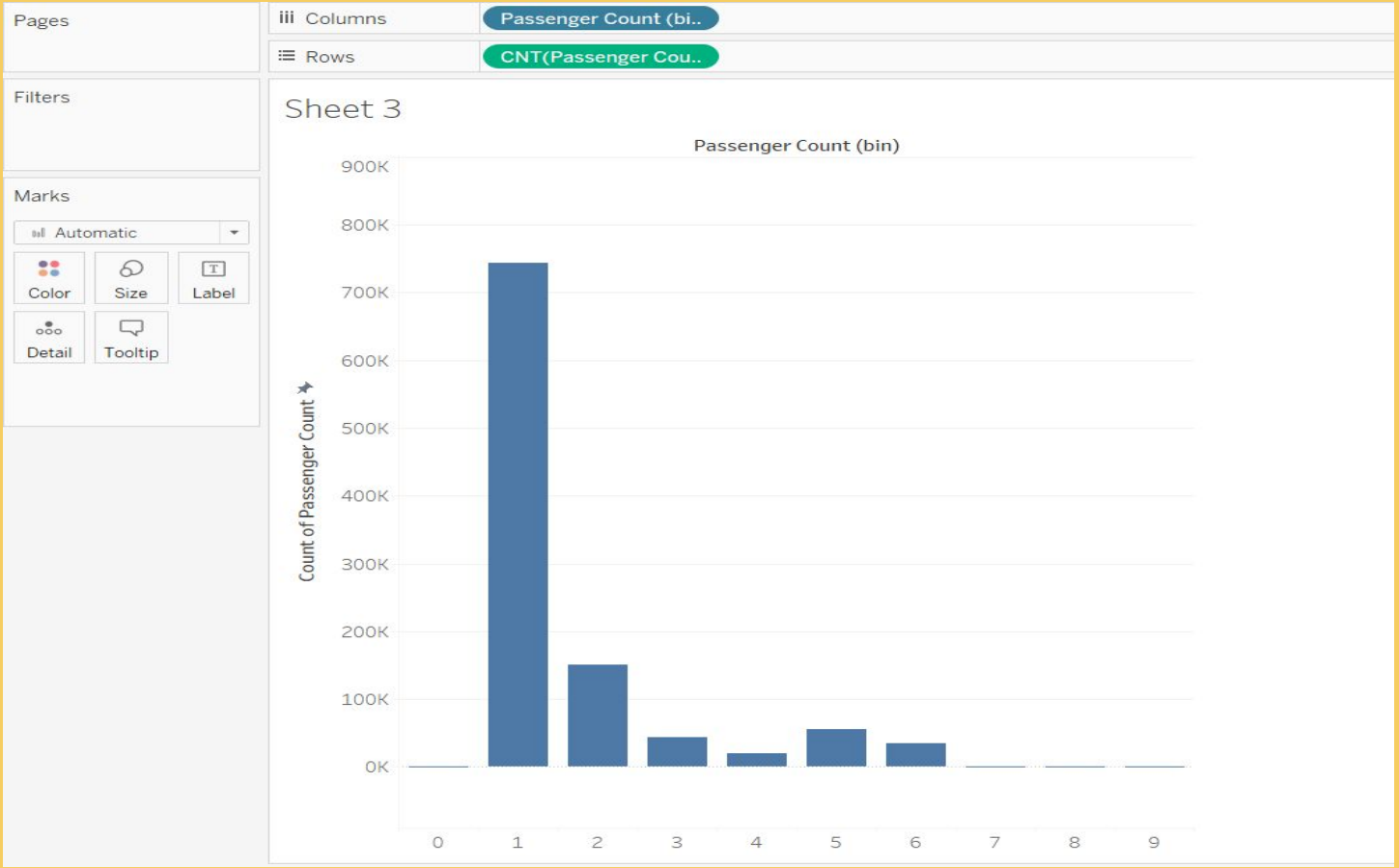
Distance :
Average = 3.4 miles
Median = 2.1 miles
1 std Deviation = 4.382 miles
2 std Deviation = (1 std deviation)* 2 + mean
= 12.164 miles

Passenger Count :
Average = 1.664
Median = 1
1 std Deviation = 1.314
2 std deviation = 4.292

Duration to Distance ratio :
Average = 469 seconds per mile
Median = 279.5 seconds per mile
1 std deviation = 13553 seconds per mile
2 std deviation = 27575 seconds per mile

Price :
Average = 19.66
Median = 14.05
1 std Deviation = 50.90
2 std deviation = 121.46 // (2 std deviation = (1 std deviation * 2)+mean

Analyzed the potential market volume for MVP



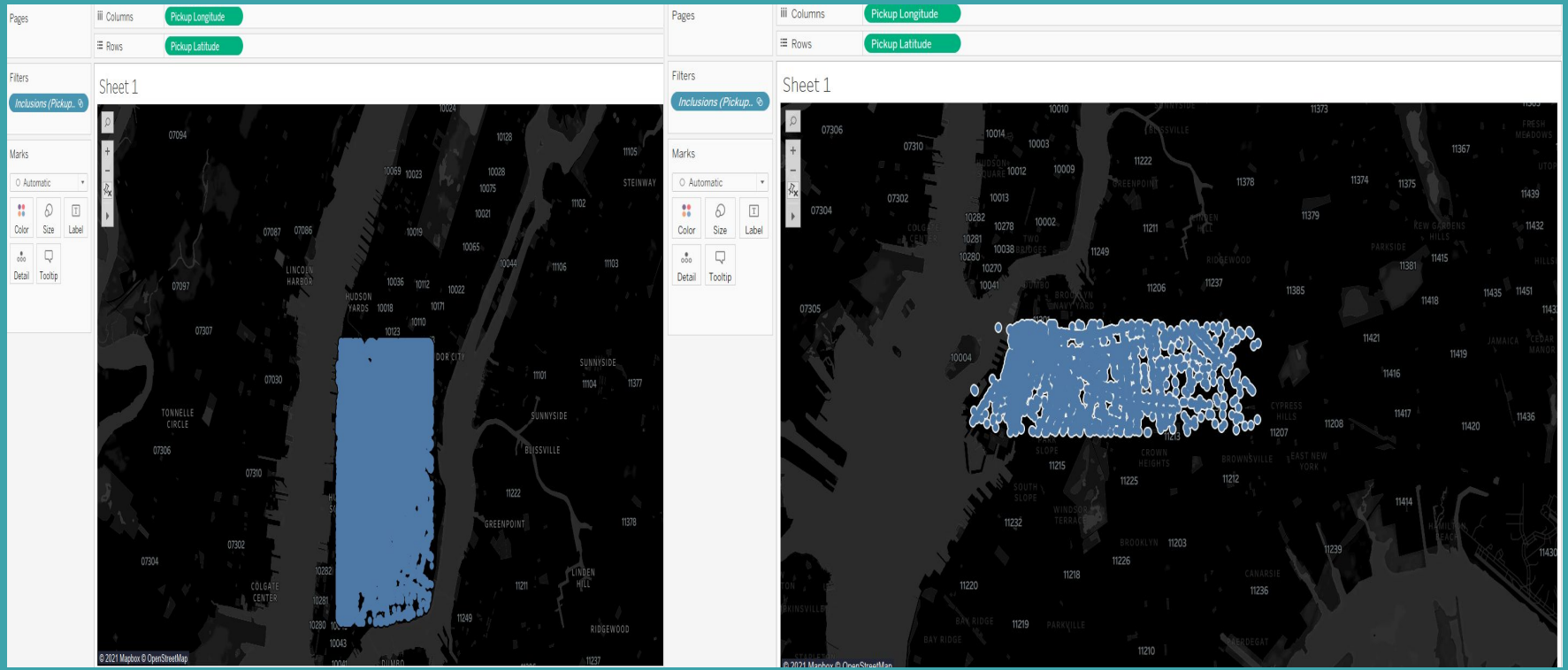
Analyzed the potential market volume of low passenger pickups by visualizing the number of total rides grouped by passenger counts From the graph we can infer that the 1 passenger has highest count. The 4 has the lowest.

For **the initial MVP launch**, since at this moment we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. I **strategize the locations** where we can place nodes based on :

- 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?

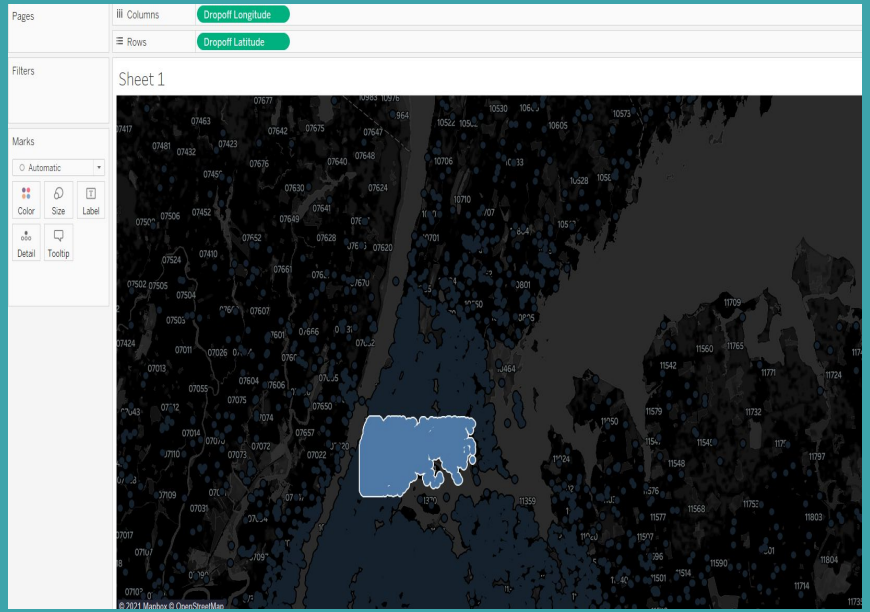
Pick Off

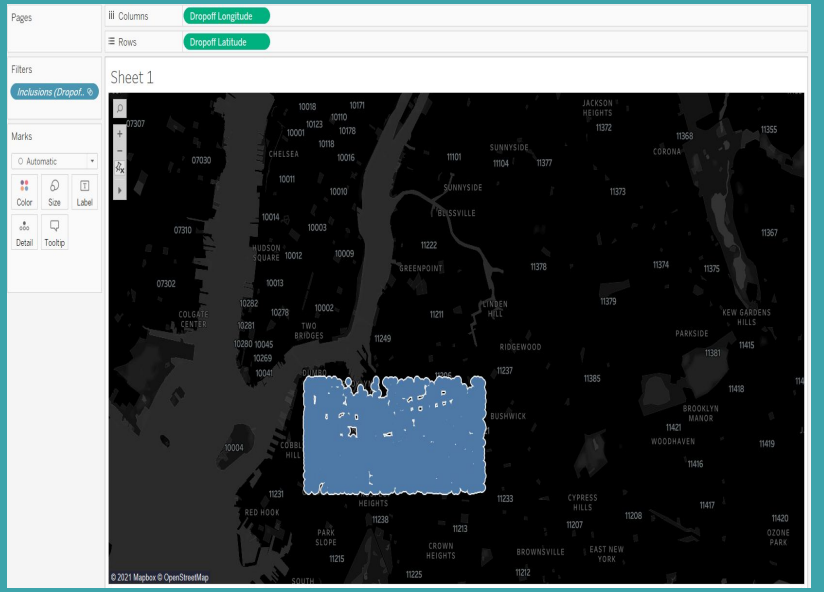
→ Washington Square village, 10010,1123, bedford stuyvesant



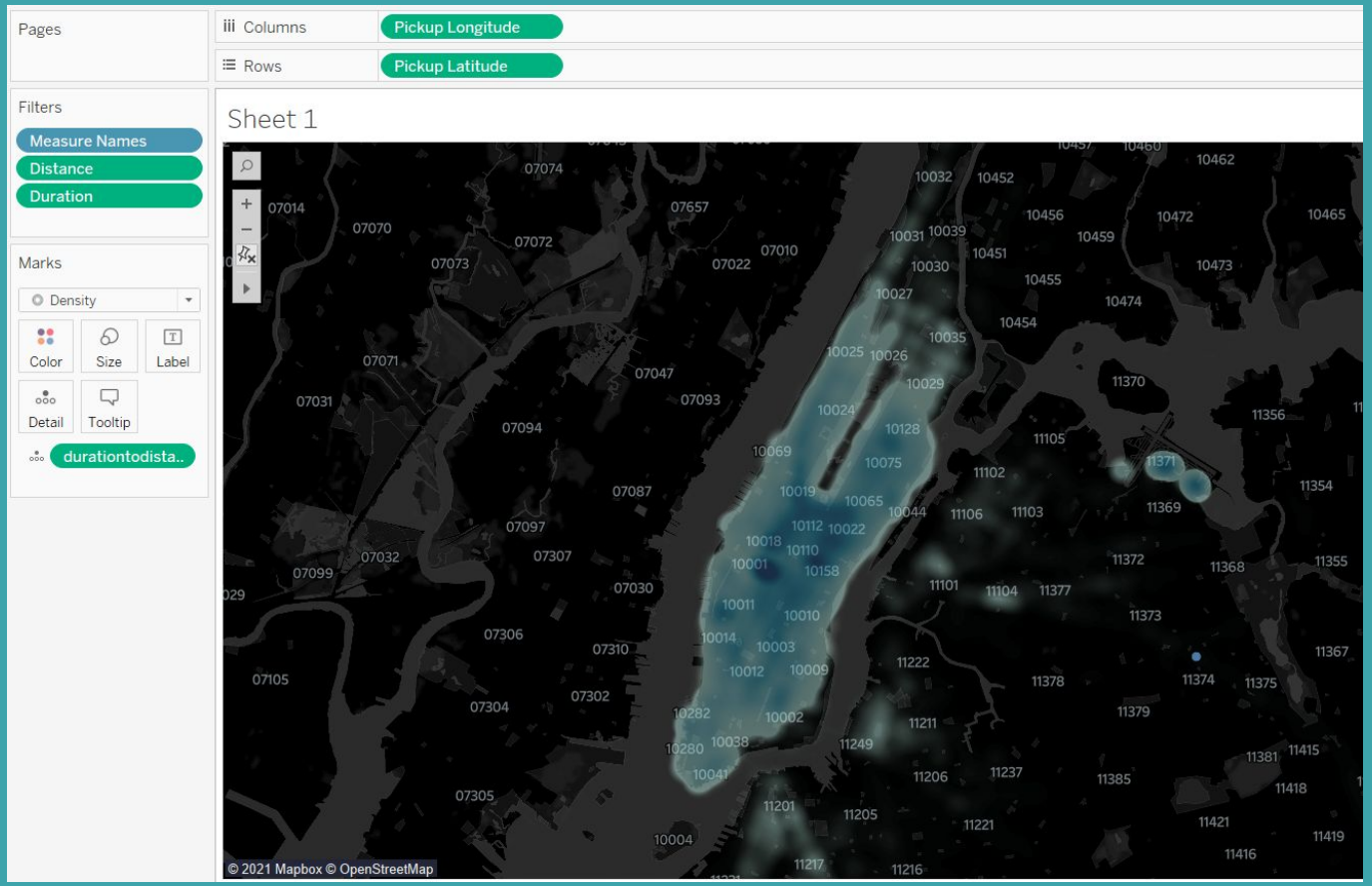
Drop Off

→ 10451, 10456, Fort Greene, Clinton Hill,11221

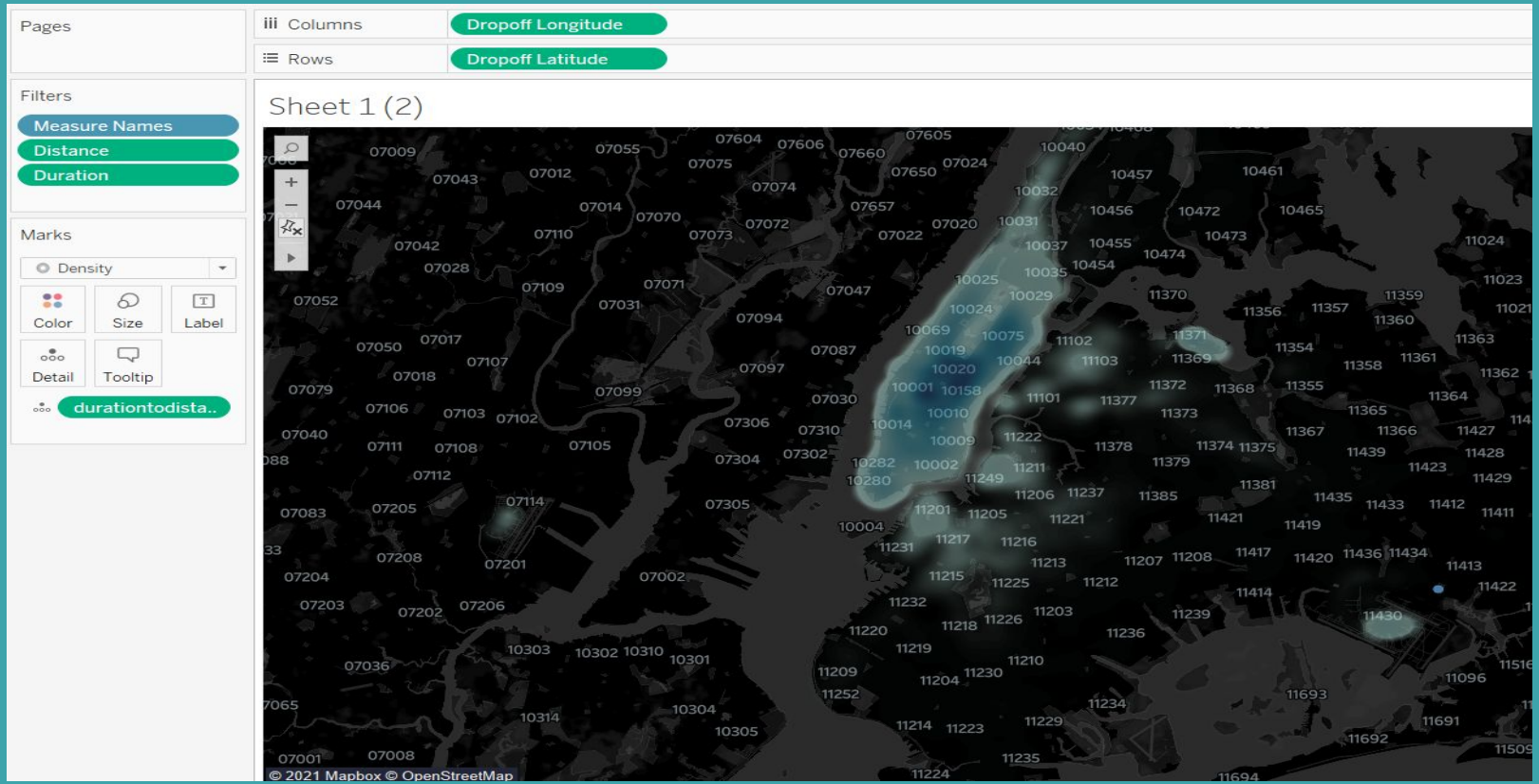




Highest duration-to-distance ratios, based on pick-up



Neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off



Neighborhoods identified, any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off

Drop off -10020, 10158, 10001

Pick up - 10001, 10158, 10112

These are the zip codes where the MVP should be launched for pick up and drop off since these are the most optimal area as per duration to distance ratio.

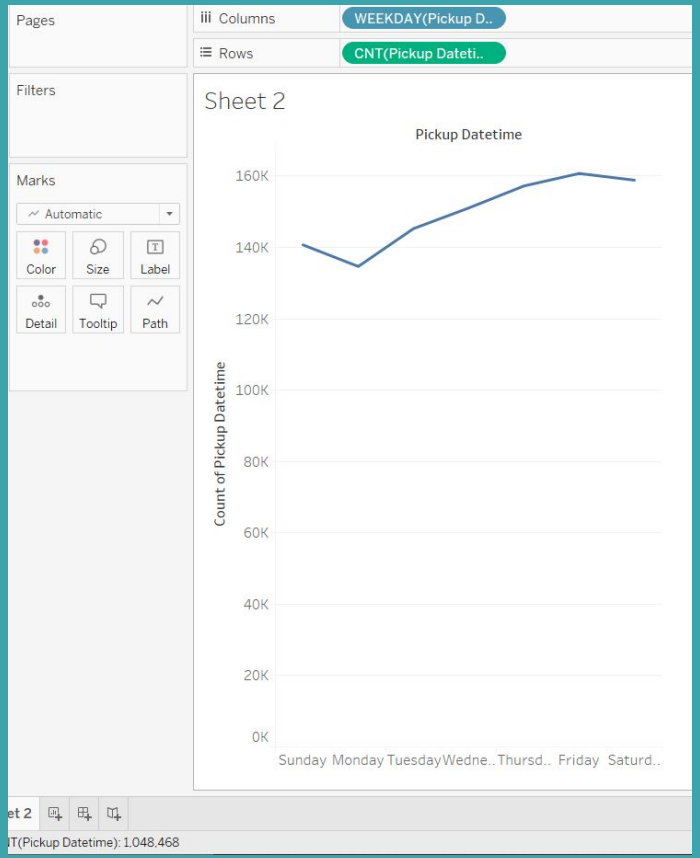
It does **not make operational sense** to have the service running 24/7, for now. I analyzed below things for the same :

- 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.

Daily

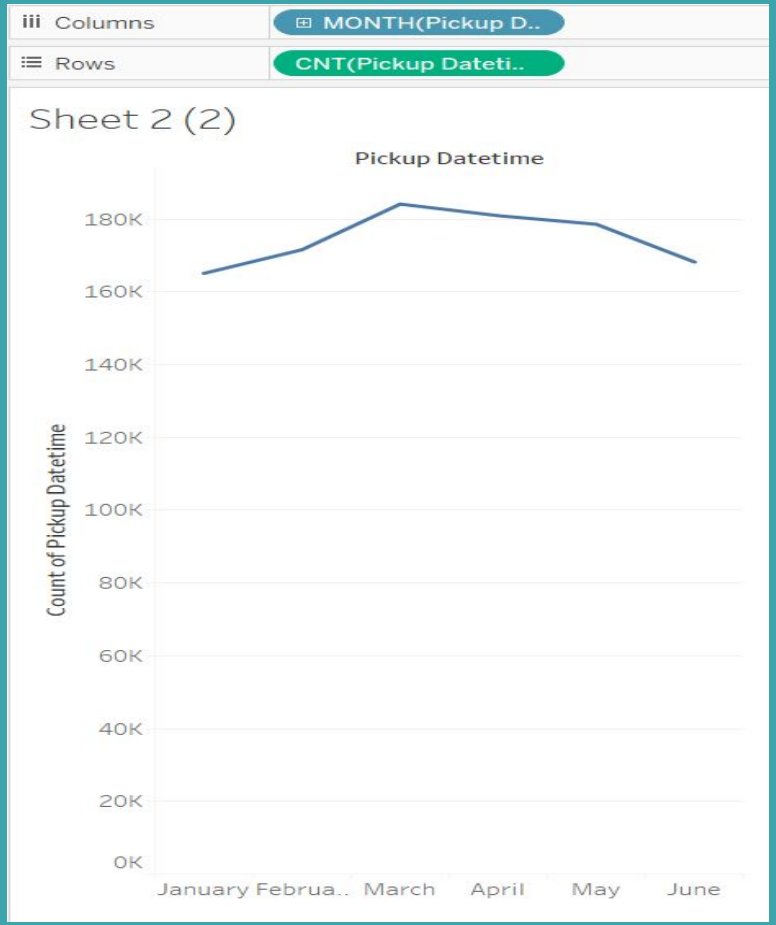
The maximum volume can be observed in 18-19th hour that is around 6-7 PM in NY with 65k.

Weekdays



Highest volume in the weekdays can be seen at Friday. Stands around 160k

Trend fluctuation/ Seasonality



This graph shows how the trend fluctuates over the period of 6 months from Jan to June. The highest volume can be seen in March.

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.
Ensure the fields are parsed correctly, field headers are included in the first row of the CSV.

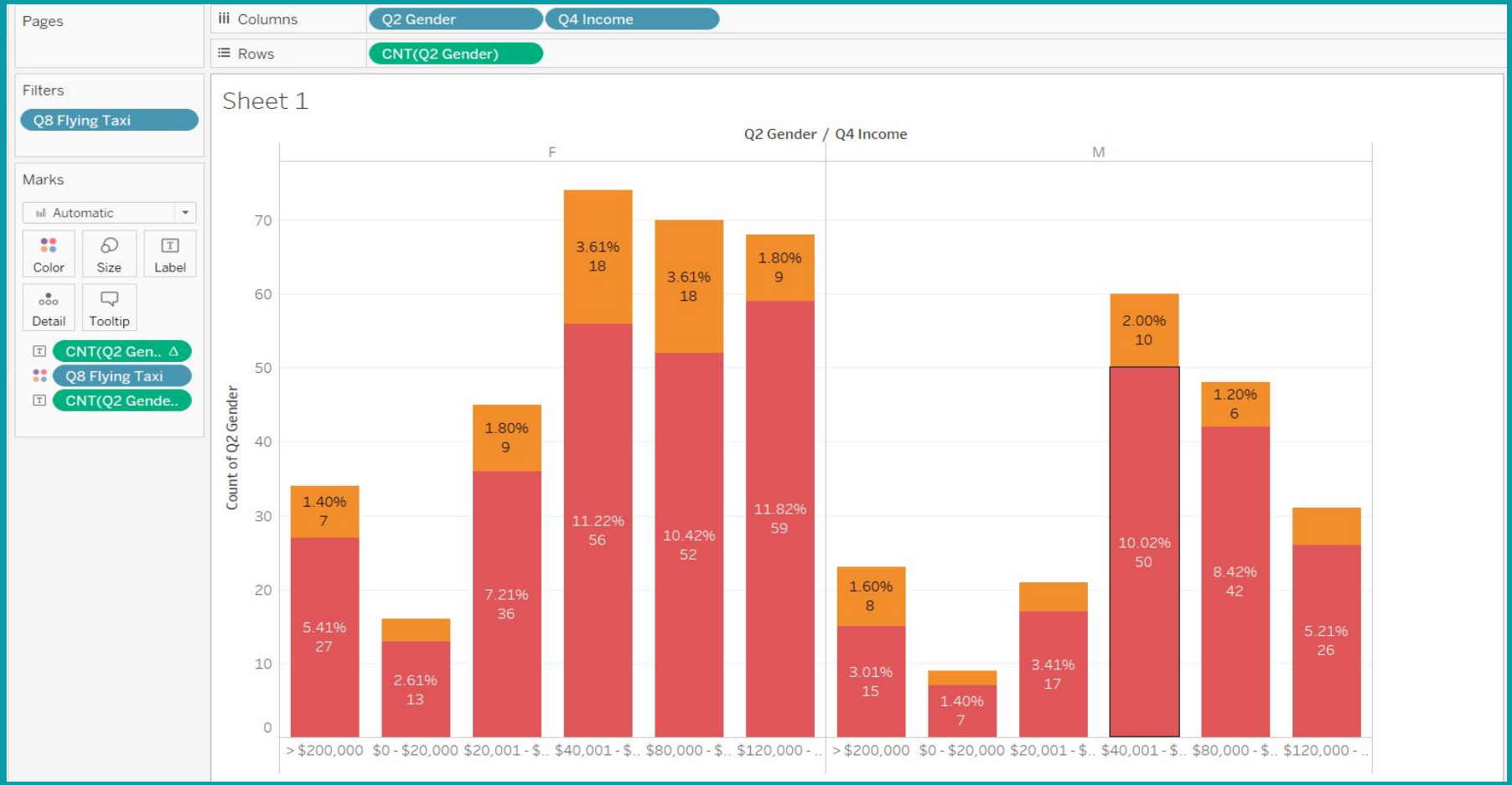
Question schema:

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?

In order to Inform future product marketing efforts, I extracted the following information from the dataset:

- Any inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence for target marketing.
- The distribution of potential price per mile based on gender, age, income level, and neighborhood of residence.
- The different personas/segments of negative sentiment towards not using a flying taxi car service.

Flyber Adoption - Gender and Income



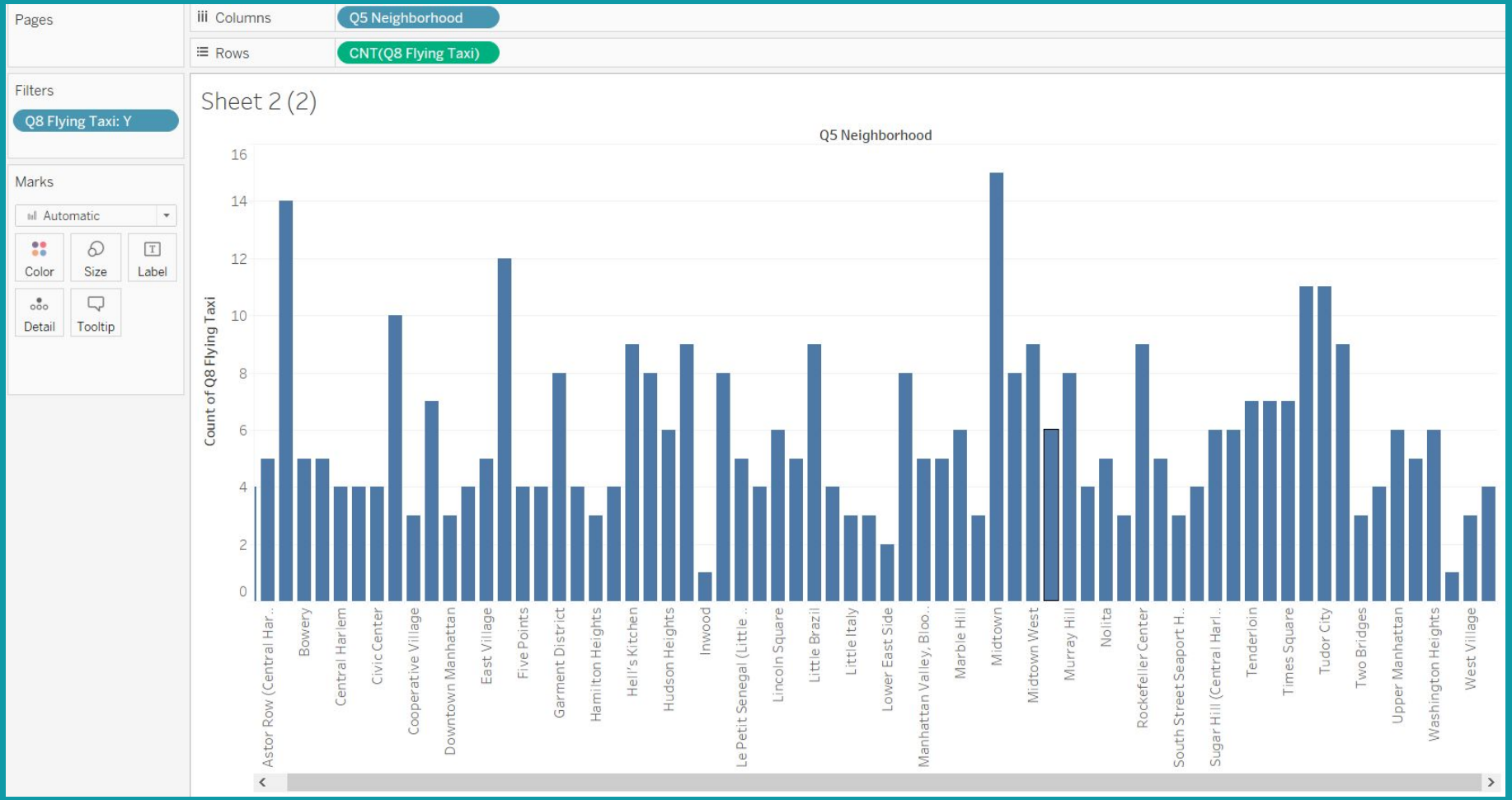
Flyber Adoption - Gender and Income

→ Left half of the graph shows Female users while right half shows male users. The red color depicts the gender group who said “yes” to flyber and the orange depicts the passengers who said “no” to flyber.

From the graph we can depict that Female users are more adopted.

Also, both Male and Female users who belong to 120k -140k $ has adopted the most from all income levels.

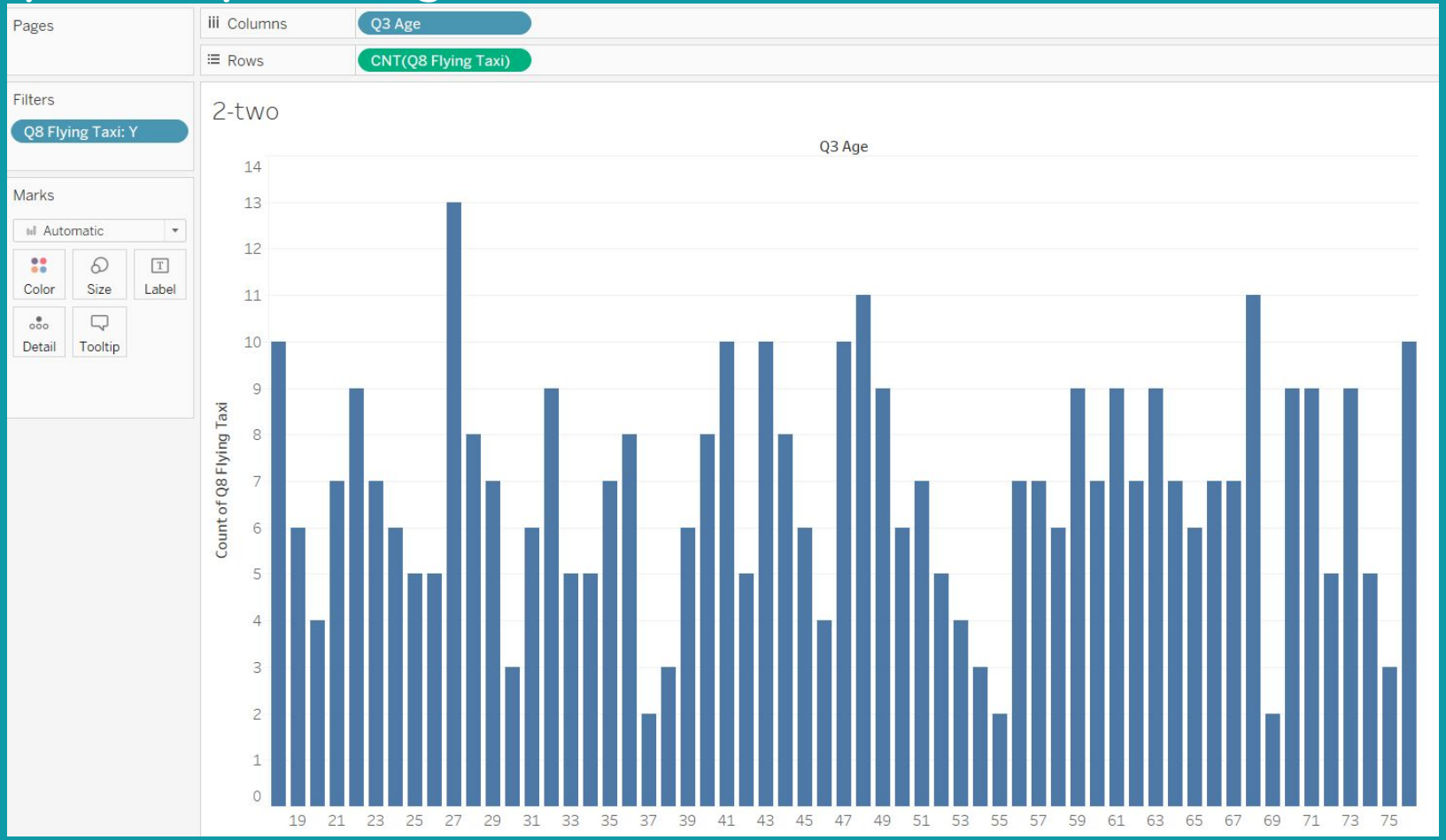
Flyber Adoption - Neighborhood



Flyber Adoption - Neighborhood

→ Midtown adopted the most and followed by neighbourhoods such as Astor Row.

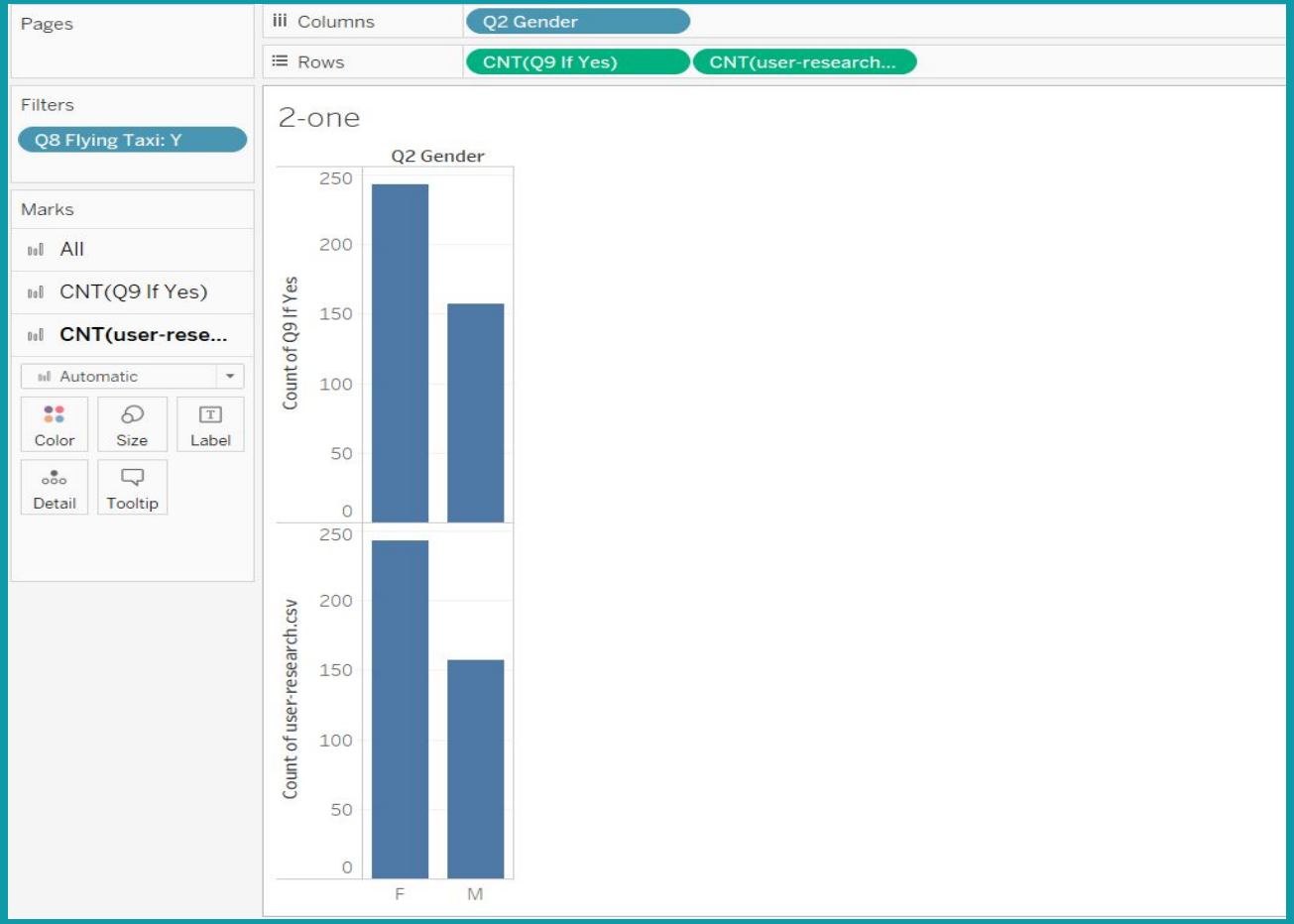
Flyber Adoption - Age



Flyber Adoption - Age

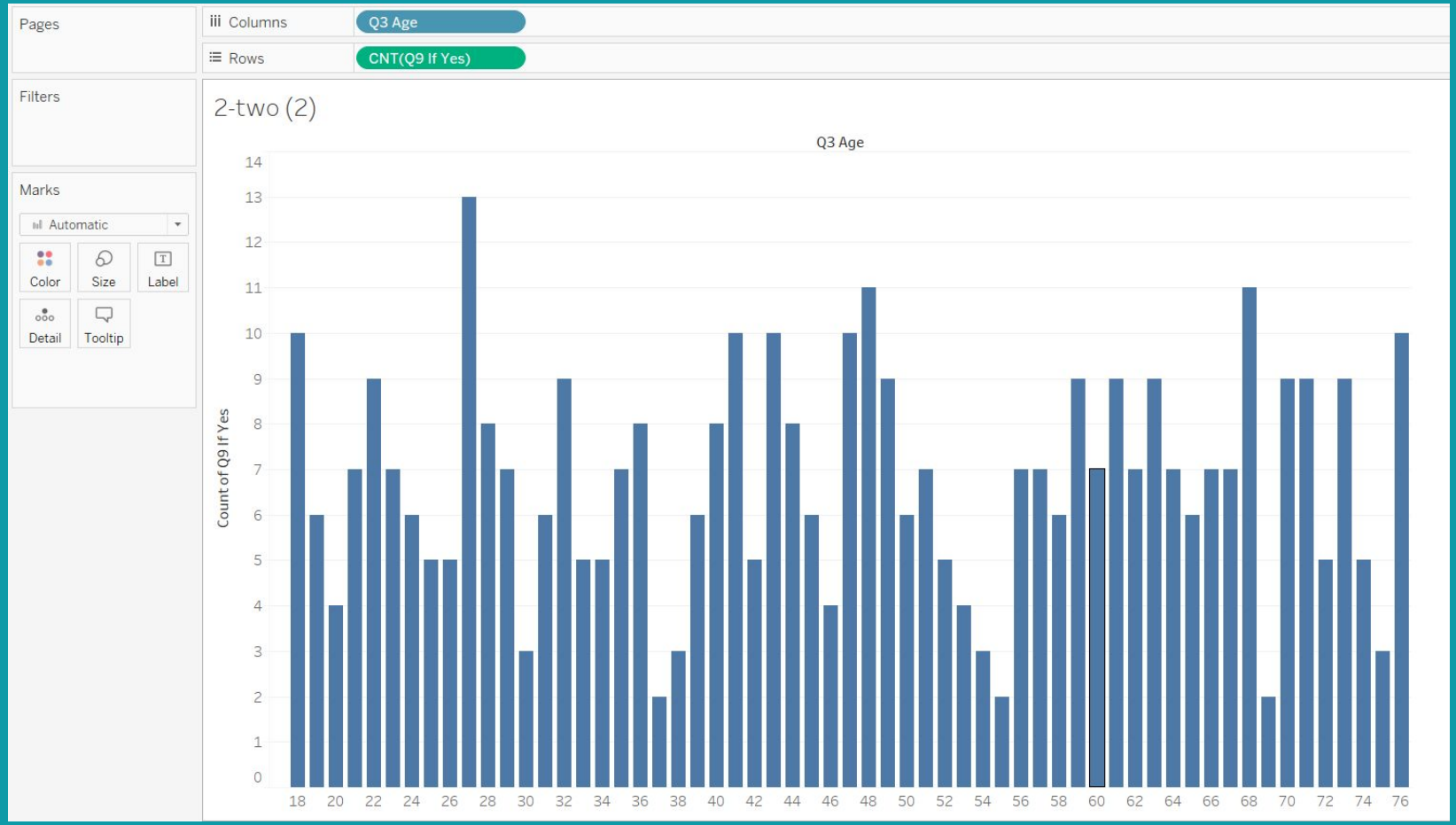
→ 27 years of Age adopted the most followed by 48 and 68 years of Age.

Potential Price Per mile - Gender

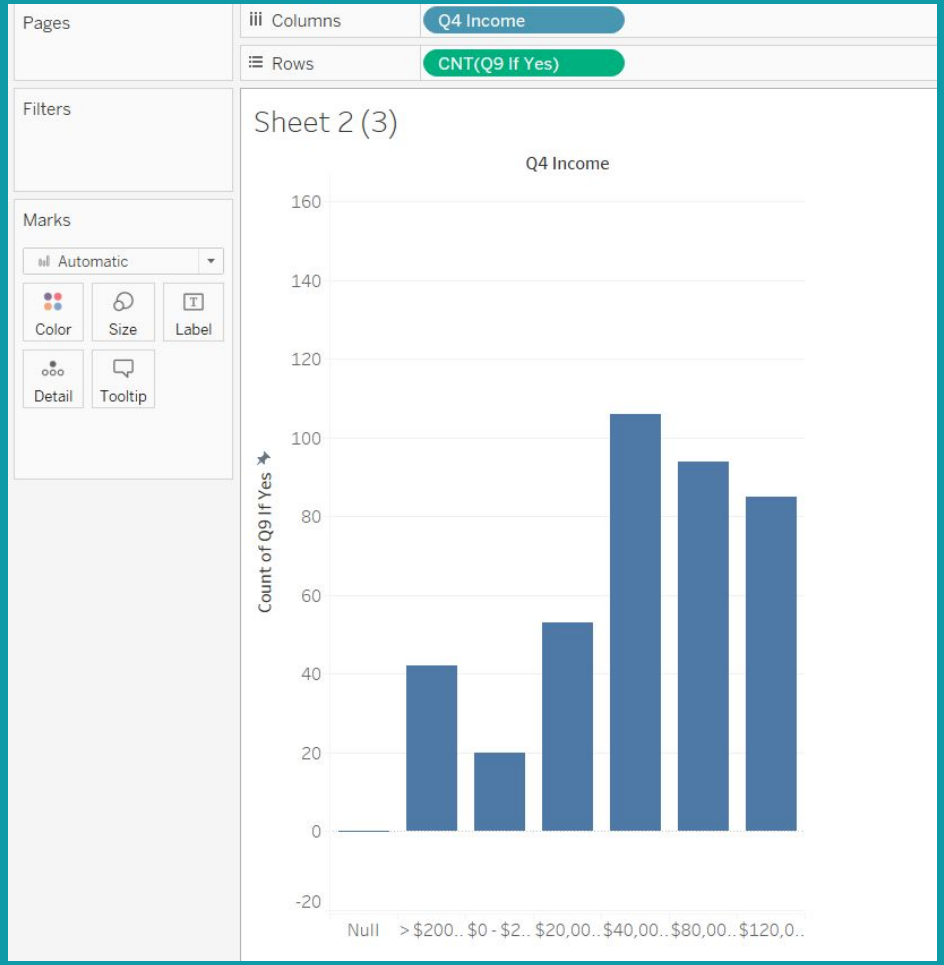


Female were higher for “yes” as compared to male.

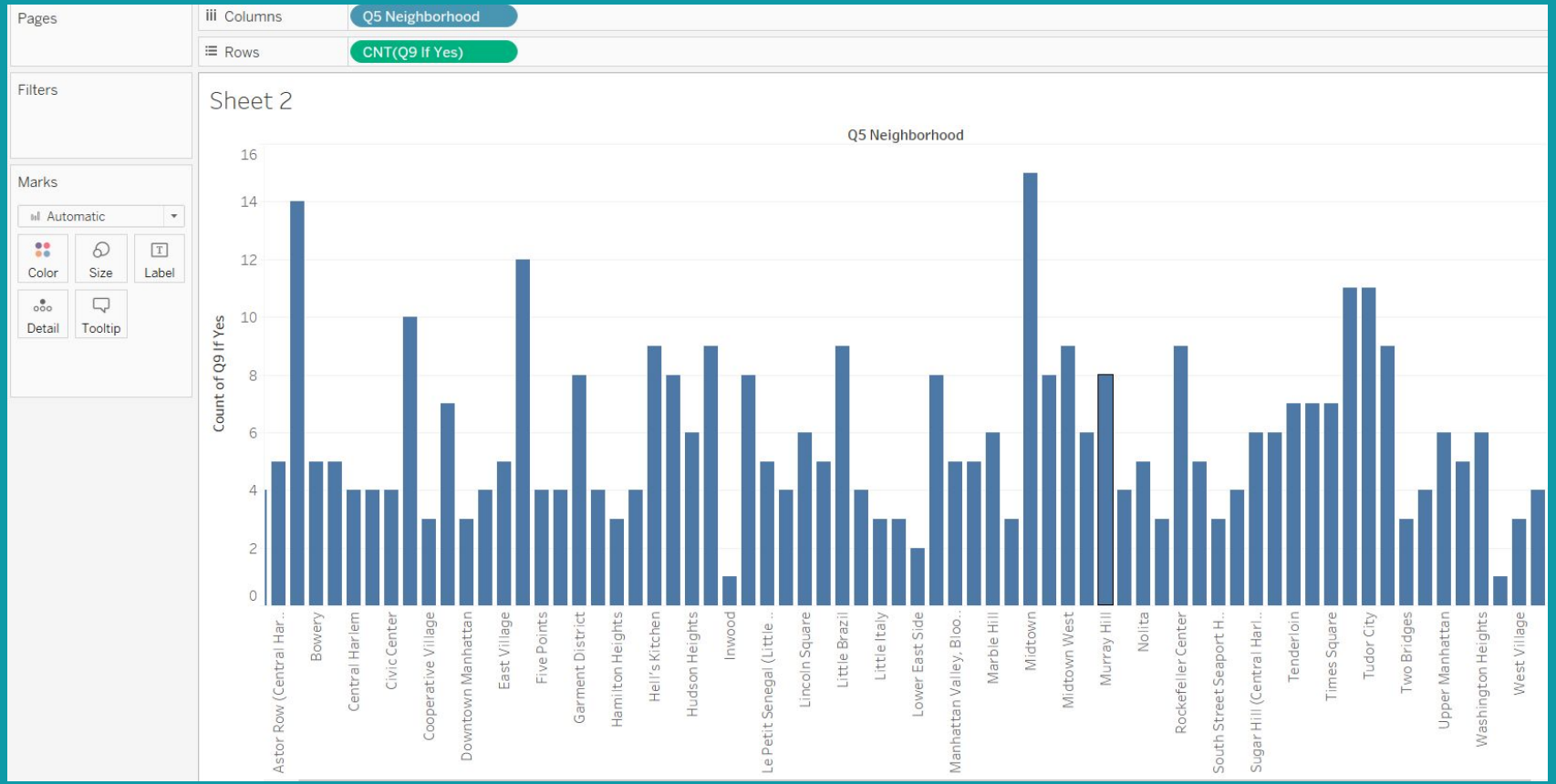
Potential Price Per mile - Age



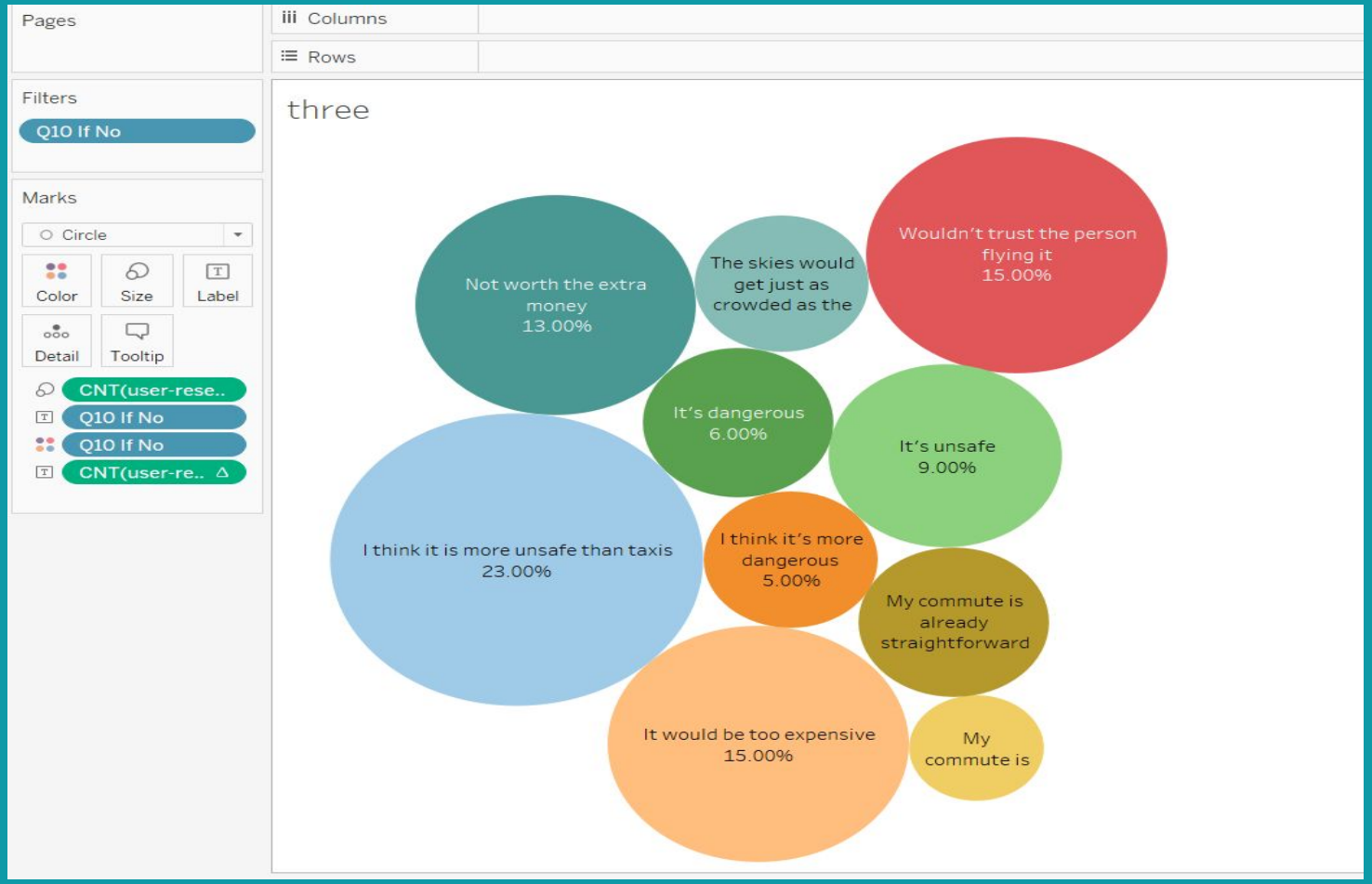
Potential Price Per mile - income



Potential Price Per mile - Neighborhood



Negative Sentiment



This depicts the reason why people were hesitant to use flyber. Unsafe than taxi recorded the most count.

Section 2: Proposal Synthesis

Identified a product objective for Flyber's launch to guide my KPIs. From below objectives, I focused on User Acquisition:

- User Acquisition
- User Engagement
- User Retention
- Profitability

Product Objective - User Acquisition

→ Flyber, objective is to create brand awareness and have diligent marketing in order to acquire users. The Go - To - market strategy has to be exemplary in order to have 10 % market share in NY city at the end of 5 months.

KR 1 - Increasing drop off and pick up location by 20 %
KR 2 - Collaborating with companies and giving discount to their employees for daily commute would bring in 30 % market share of regular taxis.
KR 3 - Referral program with Flyber App interface will bring in more users. Average NPS to be above 80.

→ At this stage, User Acquisition is the most important objective for the Flyber as the company needs to have users. The company only has a data based on research survey. Once the company has acquired enough MAU or DAU then the company can focus on factors such as customer retention, profitability and how to engage users with the product. I think, if the company have an access to actual user data in large amount then there are various possibilities and strategies company can utilise to enhance the product and the business.

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

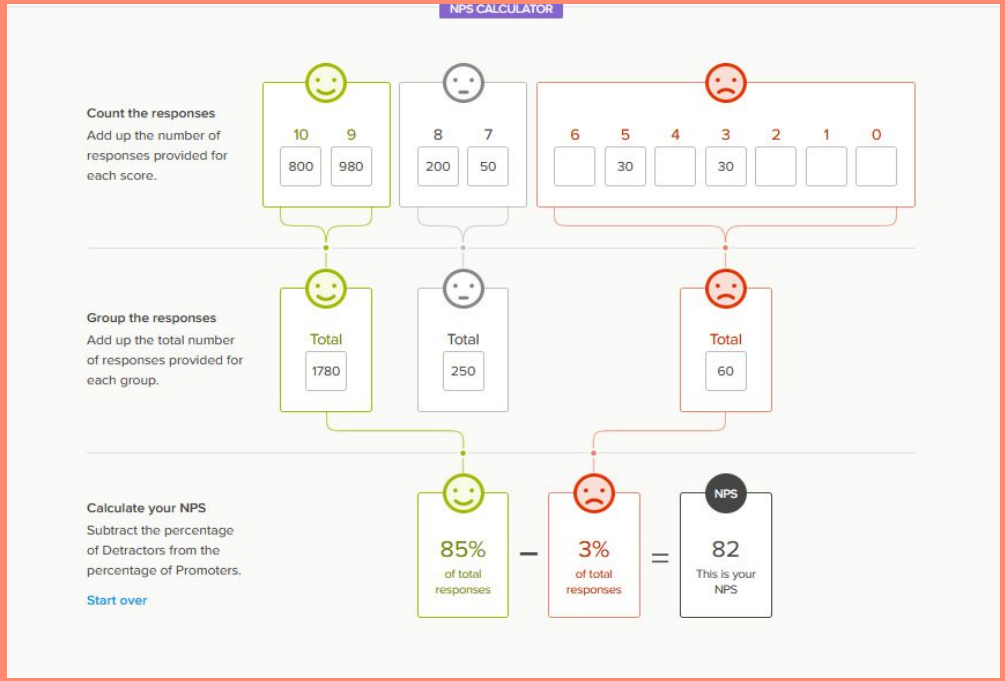
KPI 1:
DAU / MAU a ratio effective at measuring the rate of stickiness,
for products that are meant to be used daily
DAU target : 2000 users in MVP phase (6 month)
MAU: 5500 users in MVP phase
ratio: 36%
*** DAU/MAU ratio of 50% basically means that the average user of our service uses it 15 days from 30. DAU/MAU ratio of < 1% means that all new coming users try our service just once and do not return.**

KPI 2:

Monthly recurring revenue (MRR) total amount of estimated revenue that a company will receive on a monthly basis.
Calculation : MRR(t) = Σ Recurring Revenue(t)
Taking into consideration : DAU: 2000, average distance : 4miles, target early users price budget : 30$ see below
Target MRR in MVP phase : 3.6M$

https://www.cobloom.com/blog/saas-metrics

KP 3 - Average NPS of all flying.
Target is 82



Created hypotheses around the thresholds the KPIs would need to hit in order to determine success

Hypothesis template : [Feature name description] solves the user need through qualititive reason], we expect it to move [KPI] by how much you think the KPI will move by

Hypotheses 1
Add flying safety features like safety engine solves the user need through reducing fear from flying services we expect it to move DAU / MAU KPI by 5%.
Hypotheses 2 Add early booking features through mobile app solves the user need through giving control we expect it to move up MRR KPI by 12%.
Hypotheses 3 Add luxury features like 5G connection , laptop workstation solves the user need through enjoy her time we expect it to move up NPS KPI by 5%

Now, I included feature set based on the insights extracted in the MVP to measure viability while keeping operational expenditure under control

- Times/Days of operation should the service run for
- Number of pick-up / drop-off nodes we should have
- Location of the Nodes
- Should we initially use copters or homegrown hardware?
- Should the pricing be fixed or dynamic? At what rates?

→ The operation should run in 18-19 or 21th hour usually on friday / Saturday.

→ Since the company is just forming, it is going to have financial constraints and thus I would prefer to have around 5-6 nodes in total. I will place nodes very demand is high to have revenue boost.
Nodes - Pick up - 10123, diamond district, 10111
Nodes - Drop off - 13068, 13859

→ The company should use copters since at this stage it is very important to know what user thinks about this idea of flying taxi and what is the potential of this business. Homegrown hardware would cost a lot for the company at this stage and in the iteration phase this will prove costly. User validation and initial traction would be easily obtained with the help of copters. Which is the end purpose of MVP.
→ **Dynamic Pricing =**
Total amount of flight time + 0.35(Total amount of flight time) + 1.75(Total miles) + booking fee

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

Baseline conversion rate is the current conversion rate for the page you're testing : Start with my DA/MAU KPI: 36%

Minimum detectable effect (MDE): You need to decide how much change from the baseline (how big or small a lift) you want to detect. If you enter the baseline conversion rate and MDE into the Sample Size Calculator, the calculator will tell you what sample size you need for your original and each variation

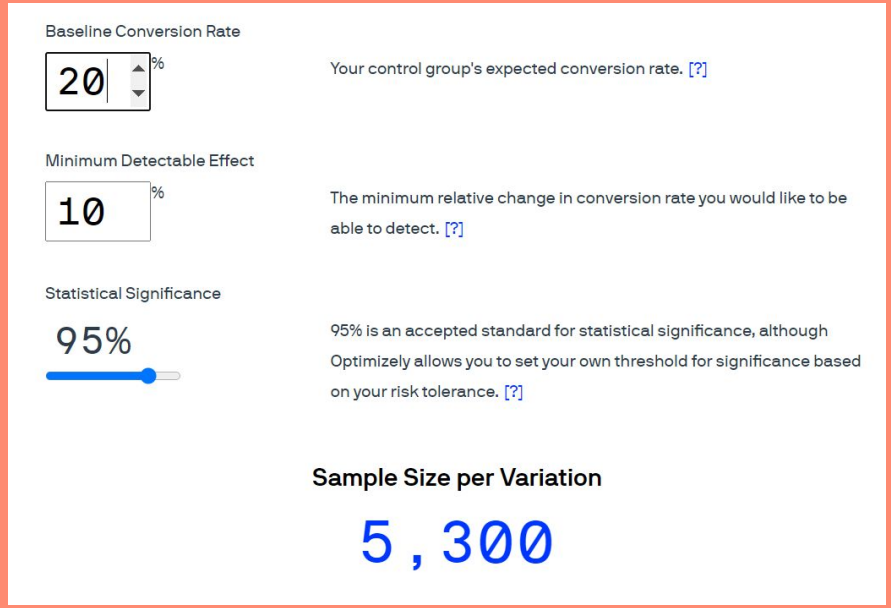
Statistical significance : Statistical significance answers the question, "How likely is it that my experiment results will say I have a winner when I actually don't?" We usually consider 90% statistical significance. Another way to say the same thing is that we will accept a 10% false positive rate, where the result is not real (100% 10% = 90%).

→ 10 % market size in 5 months of NY city ride share - Objective

→ Assuming = 5 million users
Therefore, SOM = 10% * 5 million = 500,000 users
Conversion(booked) = 20% of 500000

DURATION

Sample size = 5300
DAU = 80
Therefore, duration = 5300/80
= 66 Days.



After this, I created an instrumentation plan for the events needed to collect and logged, in order to be able to physically measure the KPIs.

Instrumentation

KPI 1 - KPI 1: DAU / MAU a ratio effective at measuring the rate of stickiness, for products that are meant to be used daily

Instrumentation Strategy 1 -

Event I. userOnboard
Definition - triggers whenever the driver hits the “ User is picked up” button on the Flyber driver app.
Properties - user_count , rider_id , driver_id , pickup_timestamp ,pickup_latitude , pickup_longitude

Event II. userExperience
Definition - triggers whenever the user activte flyber’s experience on the flight like wifi, cold drinks , music
Properties - user_experience_count , rider_id , driver_id , wifi_use ,music_on , colddrinks_on

`

Instrumentation

KPI 3 - Average NPS and the number of downloads per week.
Instrumentation Strategy 1 -

Event = downloadsApp
Properties = user_signup, user_name, user_booking
Definition = Triggered when the App is downloaded from the App store.

Event = appReferral
Definition: Triggered when the user sends a referral link.
Properties = appReferral_Id, appReferral_link, appReferral_sent

The both events will show how many new users we have acquired through Downloading the App and sending the referral.

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

Answer - Feedback

→ Rating - How much on scale of 1 to 5, you would recommend flying taxi ?
→ Rating - How much comfort and safe is to use Flying taxi on scale of 1 to 5.
→ Short Answer - Please type few words where we can improve your ease of use and ease of accessing it.
→ Short Answer - Where do you think the station for pick and drop off should be constructed ?
→ Rating - Rate App interface and usage on scale of 1 to 5

Summarizing everything into the final proposal

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

Answer Slide

Target Population

The target for flyber launch as per analysis would be the person having the **income level** in the range of 40000$ to 120000$. Also, target would include **employee** since we can see surge in using the service around 18:00-19:00.

The pain points faced by these two targets are :

1) The cost of individual taxi is comparatively higher for the people having income between 40000 to 120000$ in NYC. Thus, they are looking for means of shared transportation.
2) This income range are usually of employees, therefore they are looking for a solution which will save their daily commute time from NYC traffic congestion.

Answer - Product Proposals

→ The NYC city is facing issue such as traffic congestion, deaths due to accidents, unpredictability in public transport.

→ Due to the traffic congestion, a person losses more than 5 days due to rush in an year as per TomTom research. As per health official of NY, 292 deaths are recorded each year.

→ The market impact, number of taxis on the road will reduce and thus traffic congestion will be reduced. However, due to flying taxi incorporation, the employment in sectors such as Technology, Infrastructure will see a significant rise. Cars prices would be affordable upto certain point. Assuming the main fuel is Kerosene fuel, the global warming impact would be much higher than all the cars combined in NYC.

Answer - Risks

→ Engineering failure risks
→ Crime risks
→ Maintenance of Air traffic i.e control of Air traffic
→ Human error Risks
→ Regulatory risks because of residential community present

Answer - Cross Functional Stakeholders

→ Data science
→ Date engineering team
→ Marketing Team
→ Business and Sales Team
→ Mechanical and Electrical Engineering team
→ Aerospace Engineering team
→ Structure and Construction Team
→ Software Engineering Team
→ Legal Compliance Team
→ Operators
→ Various Vendors

Section 3: Data Strategy for MVP

Section 4: Applying Iterative Design Principles to a Live Product