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


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

* These flying taxis are being utilized in order to eliminiate wasted time & energy on road traffic. Flying taxis also avoids the possibility of ‘road rage’ which is quite common in NYC.

* The target market for Flyber is quite diverse. During the initial phase, it is highly likely that majority would be those with extreme curiousity. After the ‘curiosity phase’ have faded, it would be a high % of busy professionals who are ‘on the go’ for work/business. After all, its NYC.

* The paint points for regular taxis are:
a. expensive
b. bad customer service
c. disorganized dispatching service
d. lack of technology-based communication from driver to passenger and vice versa
e. lack of technology-based communication between driver and dispatcher

* The existing pain points for flying taxis plenty are :
a.keeping the regulations and licenses. Since flying cars and taxis are not that common in most parts of the world yet, this may be an ongoing issue during the initial phase.
b. Keeping the ‘spark of interest’ ongoing among target audience. Our goal is to eventually make sure that this project goes beyong ‘fad’. Eventually it would be the ‘uber, zoom, AirBnb and cloud computing’ of transporation’s ‘new normal’.

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

A hypotethical user improvement is the ease of travel and almost guarantee of arriving on time. Both driver (assuming it is not a self-driving flying taxi) and the passenger need not worry about traffic lights. Of course, on-going air traffic rules do apply.

A very good hypothesis that I would say as far as market improvements is the lack of road traffic. Although flyign cars have just recently started in certain US states, literally there would be almost zero traffic for flying taxis. This is based on a good assumption that there is still zero to minimal ‘air traffic’ that exists within commercial travel in NYC.

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 – Understanding a Deeper Granularity of the DataSet

* The total # of records is 1,048,468

* Each record represents the basic and vital details of the dataset. This means that each record represents detaisl of each taxi ride that occurs in NYC. Each record represents an attribute regarding taxi rides within NYC. These are:
-> ID, Vendor_ID, Pickup_Datetime, Dropoff_Datetime, passenger_count, pickup_longitude, pickup_lattitude, dropoff_longitude, dropoff_latittude, store_fwd_flag, duration, distance

* The Primary Key for this table is a composite key, the combination of ID and PickUp_Datetime. This is determined because 'ID' could only do 1 specific Pick Up within that exact PickUp_Datetime.
The SQL query below determines the total, distinct # of IDs:

```
select COUNT(DISTINCT ID)
from taxi_rides;
OR
SELECT count(*)
FROM taxi_rides;
```

****The above 2 queries yields to the number of distinct values : 145864**

I used another query using the 'id' and 'vendor_id' as a form of composite key (primary key).

```
SELECT count(distinct(concat(id,vendor_id)))
FROM taxi_rides
```

and yields : 145864

I used another query to make sure that I do get the correct primary key:

Answer Slide - continuation

SELECT count(distinct(concat(id, pickup_datetime)))
FROM taxi_rides
yields the result:

* The Date Range is simply bound within: PickUp_Datetime and DropOff_Datetime.
This simply means that a taxi ride occurs once the passenger starts & boards the taxi (PickUp_Datetime) & ends once he/she exits (Dropoff_Datetime).

Query # 1 shows both the MAX() and MIN() pick up datetimes:
select max(pickup_datetime),min(pickup_datetime)
from taxi_rides;

RESULT:

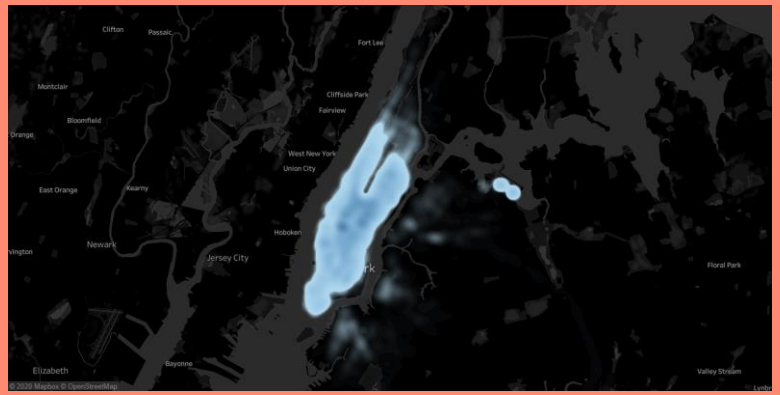
2016-06-30T23:58:00.000Z (max)	2016-01-01T00:04:00.000Z (min)

Query # 2 shows both the MAX() and MIN() drop-off datetimes:
select max(dropoff_datetime),min(dropoff_datetime)
from taxi_rides;
RESLUT:

2016-07-01T23:02:00.000Z (max)	2016-01-01T00:12:00.000Z (min)

The georgraphical bounds of this specific data set is mostly focused within New York, specifically within NYC. There are ofcours,e some outliers. The screen shots below taken from Tableau Public would prove this point:

A 'Bird's Eye View' of traffic concentration, It is very clear that there are certain areas that are considered outliers due to their locations.



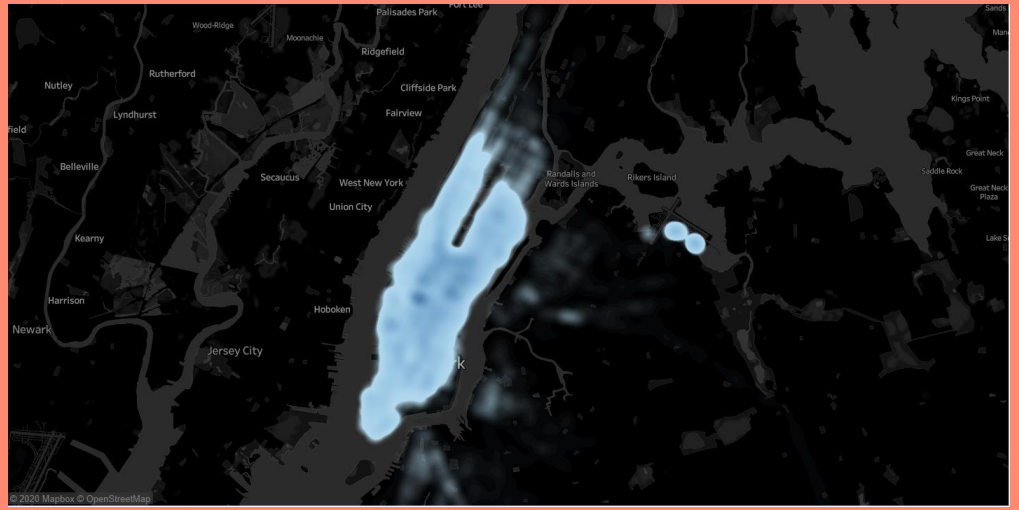
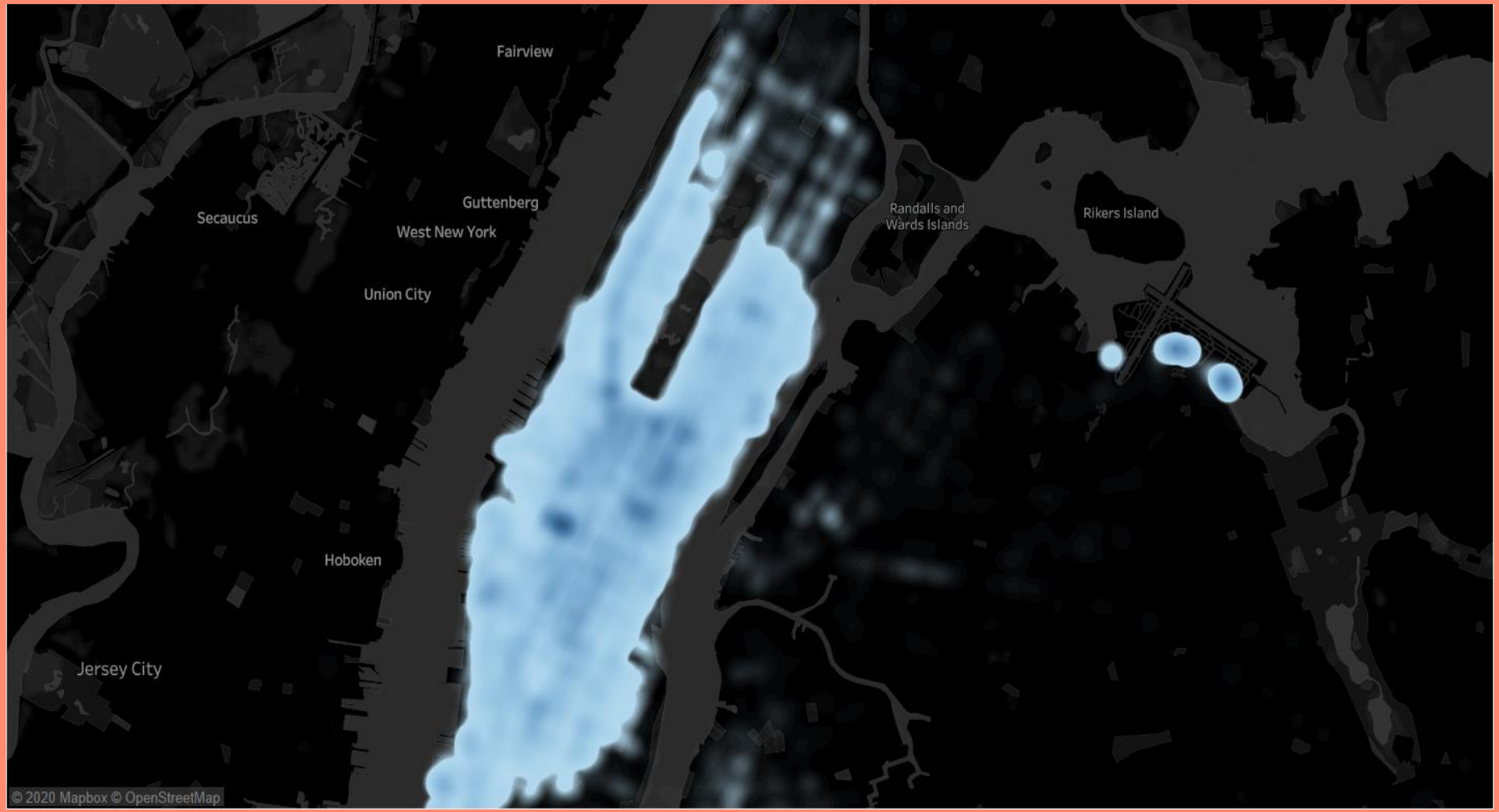
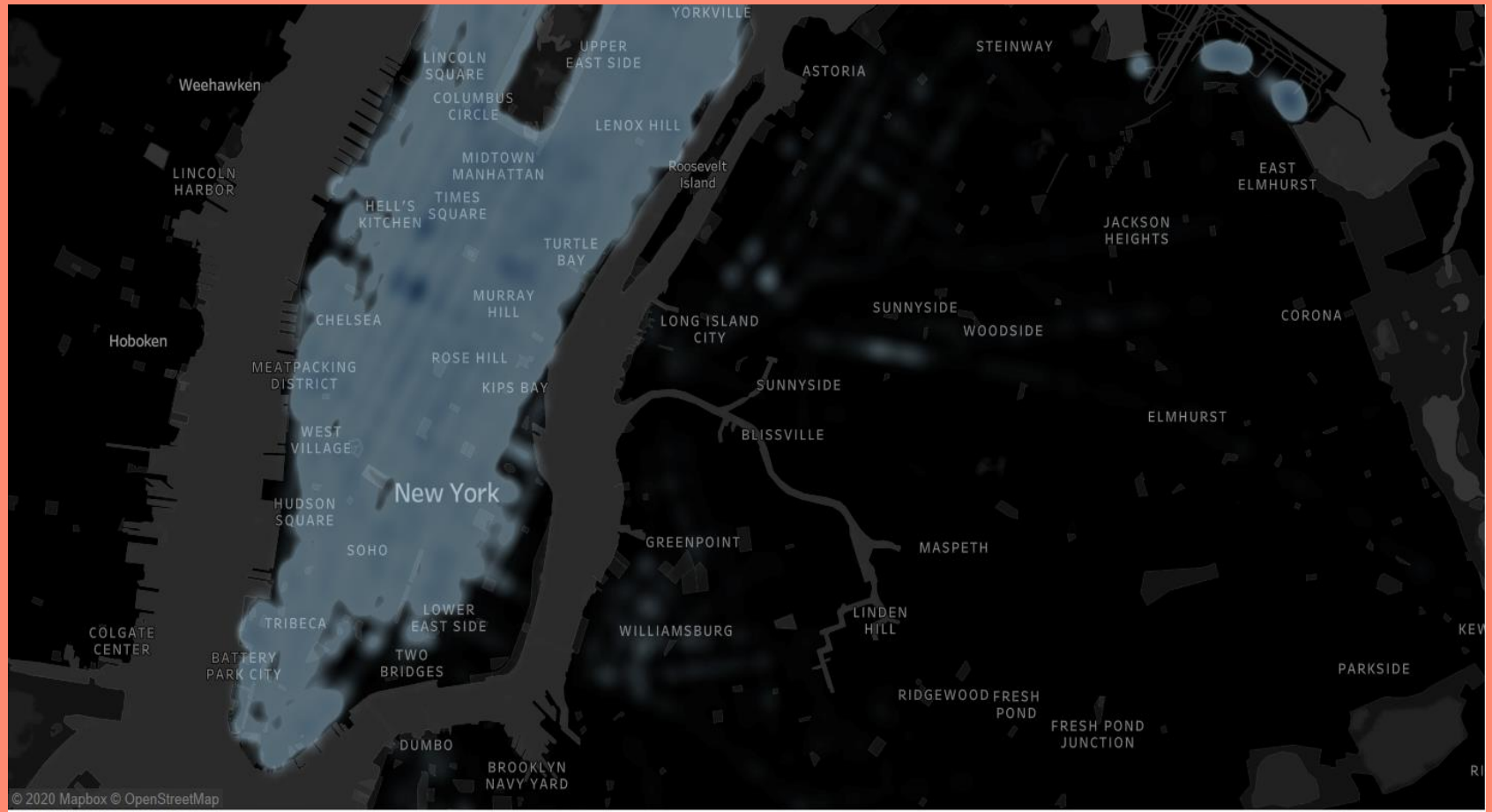


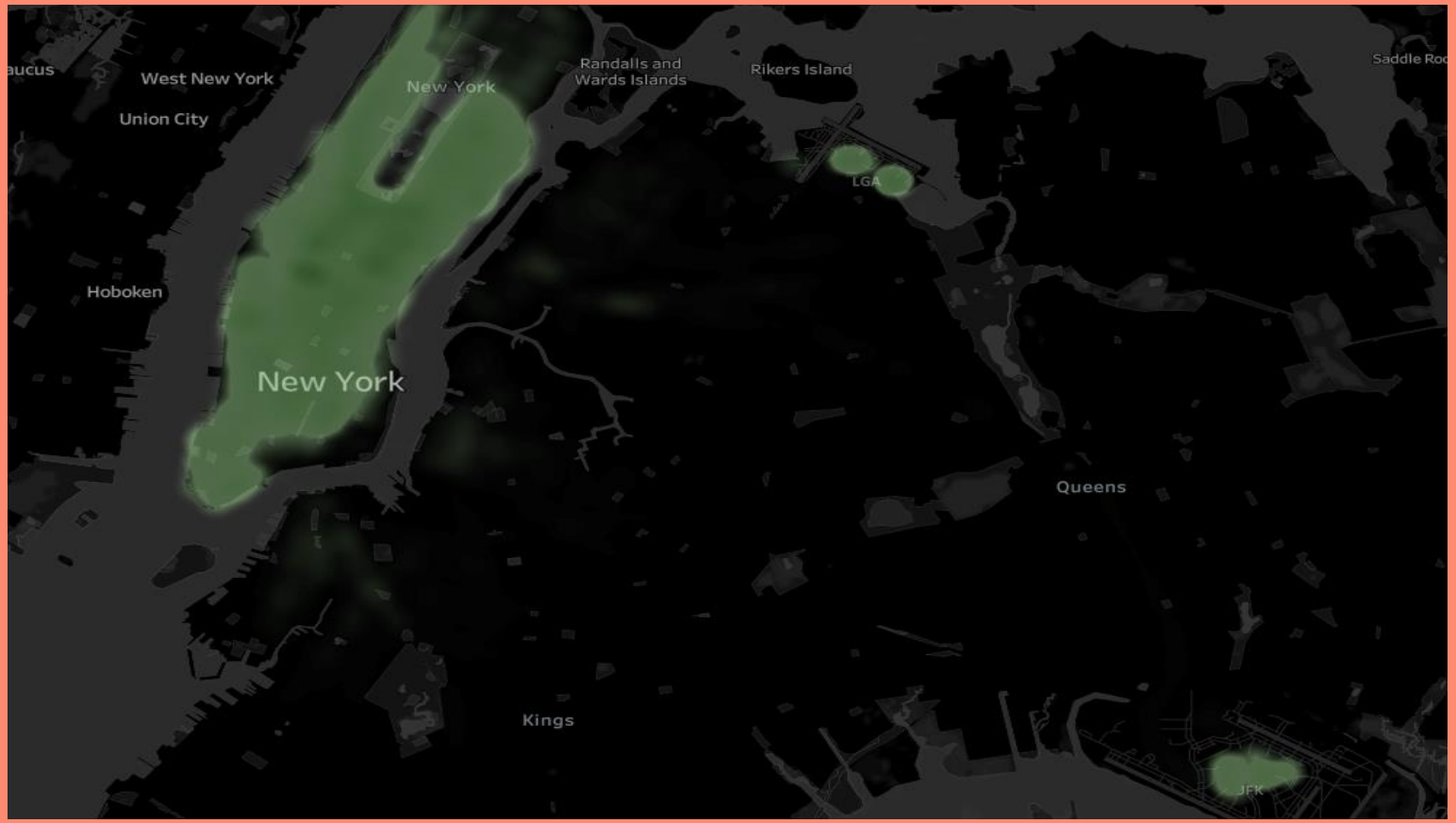
Image below shows the taxi usage from 6pm (1800 hrs) to 8 pm (2000 hrs)



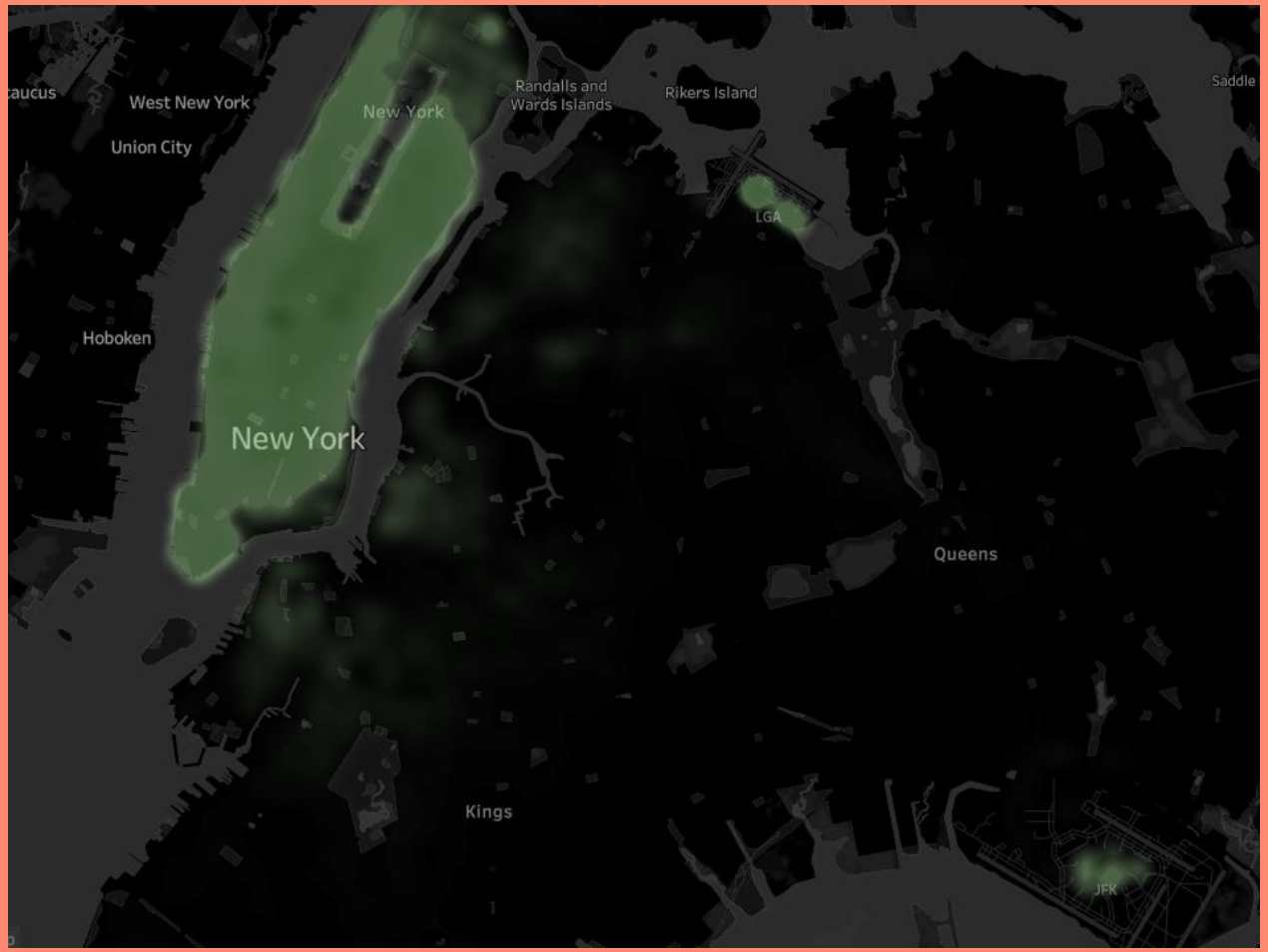
The tableau map image below shows the neighbourhoods that are of main focus within the heatmap



Based on the 2 heat maps, the main outliers for both Pickups and Dropoffs are : La-Guardia Airport (LGA), John F. Kennedy Airport (JFK), and some surrounding neighbourhoods within NYC.
PickUps and Outliers



An image of Drop Offs and Outliers: As is very evident, both La-Guardia and JFK are outliers either pickups or drop offs. This is quite obvious due the nature of NYC's activities and business operations.



The image shows a closer look (zoomed-in) at the HIGHEST DENSITIES of both pickups and drop-offs.

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 – Calculation of Values (Using Tableau)

Duration
a. Average – > 962.2 Seconds
b. Median -> 662.0 Seconds
c. 1st Std Dev - > 5,853 Seconds
d. 2nd Std Dev - > 11,706 Seconds

Distance
a. Average -> 3.442 Miles
b Median -> 2.095 Miles
c. 1st Std -> 4.382 Miles
d. 2nd Std Dev - > 8.764 Miles

Passenger Counts
a. Average - > 1.664 passengers OR persons
b Median -> 1.000 passengers OR persons
c. 1st Std -> 1.314 passengers OR persons
d. 2nd Std Dev - > 2.628 passengers OR persons

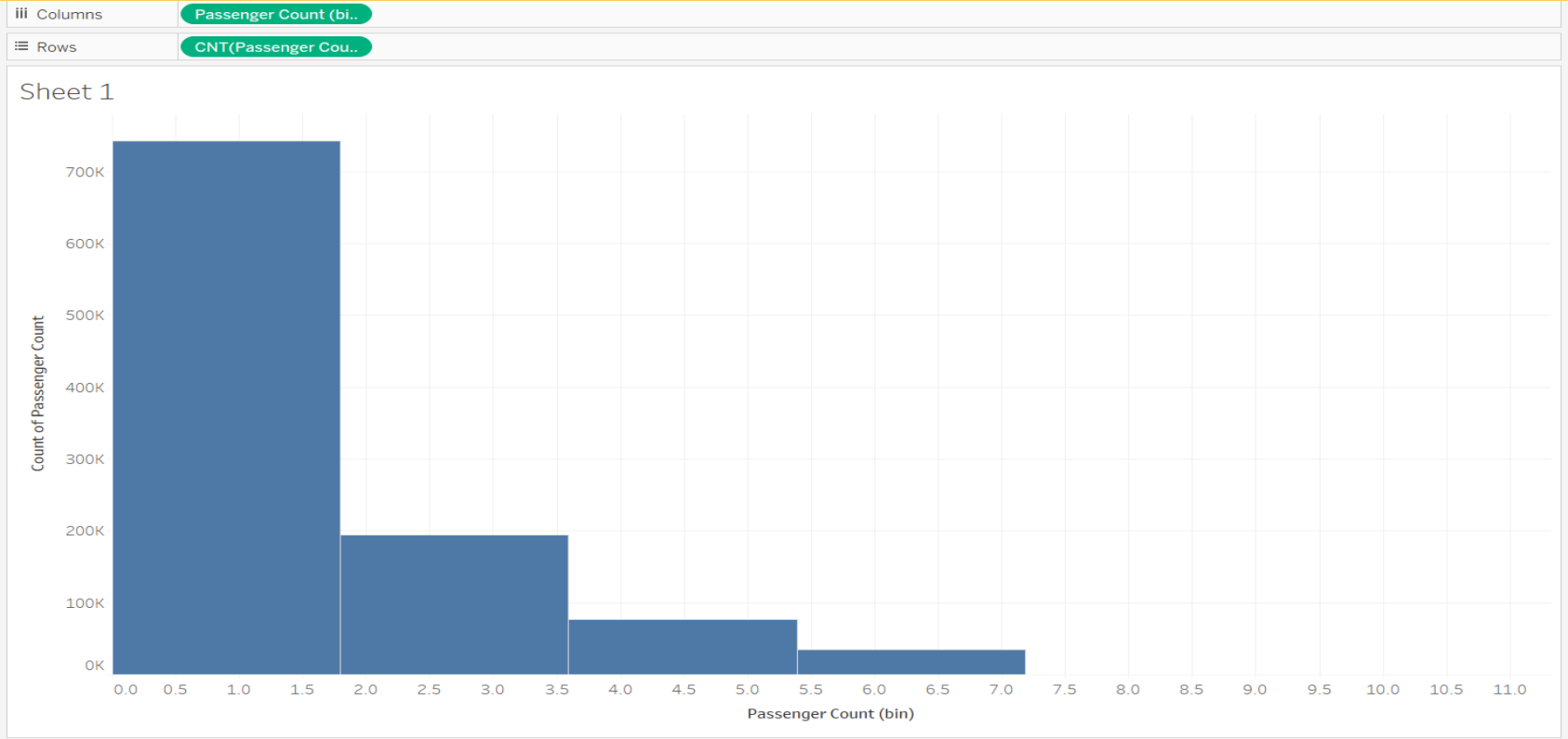
Duration to Distance Ratio
a. Average - > 4,687 miles/min
b Median -> 280.8 miles/min
c. 1st Std -> 924,373 miles/min
d. 2nd Std Dev - > 1,848,746 miles/min

PRICE = (2.5 + (1.56*[DISTANCE]*1.61)+([DURATION]/3600)*30)
a.Average = $ 6.968
b. Median = $ 5.408
c.1st Std Dev = $16.72
d. 2nd Std Dev = $33.44
** Explanation : Basic Taxi fee between $2.5 to 3.00 ; km price is $.156 , standing & waiting times is $30.00 per hour, there are 3600 secs in hour (calculation was derived from 1 of the answers in our forums)

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.

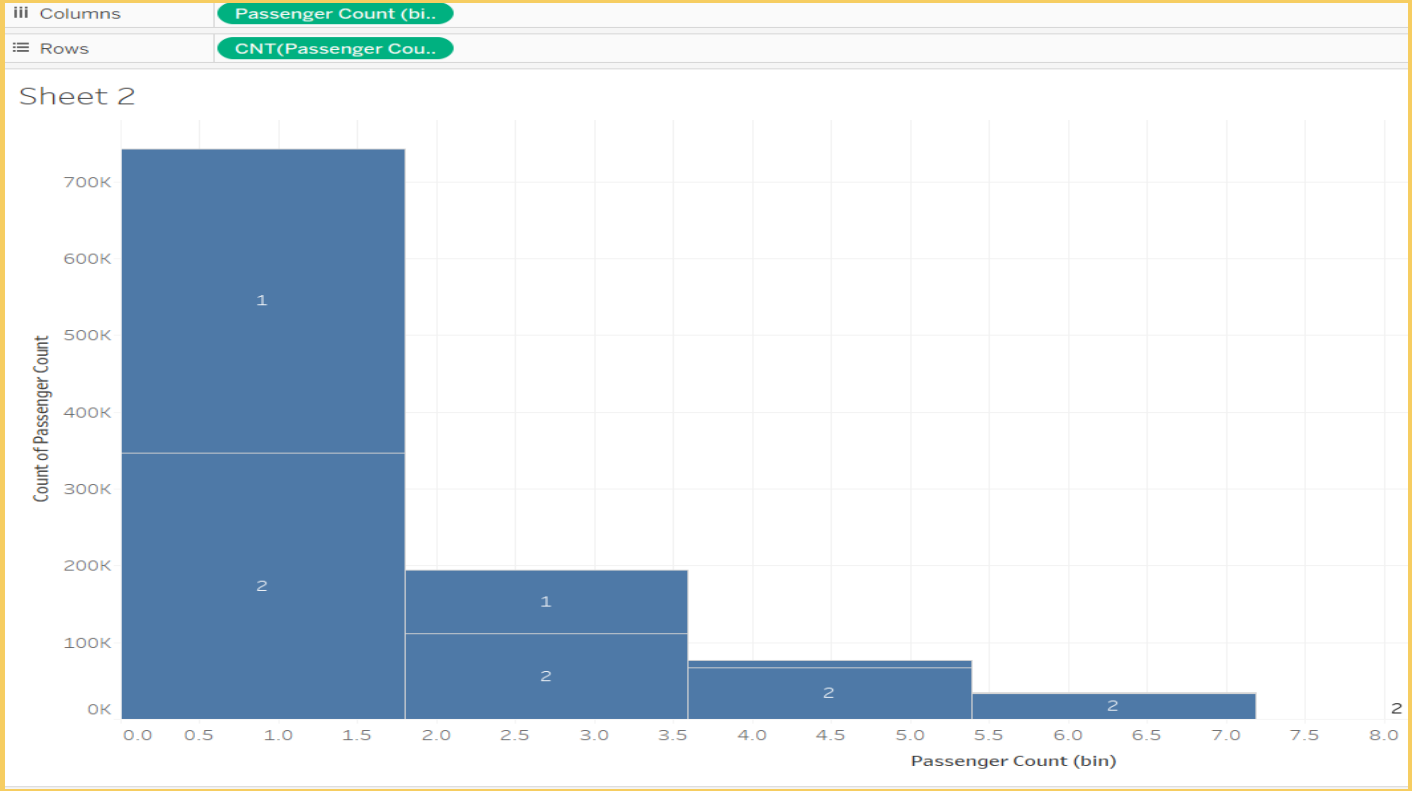
Create a histogram that visualizes the number of total rides grouped by passenger counts to analyze the potential market volume of low passenger pickups (1-2 passengers).

Answer Slide



Answer Slide

A more detailed histogram with the respective Vendor IDs

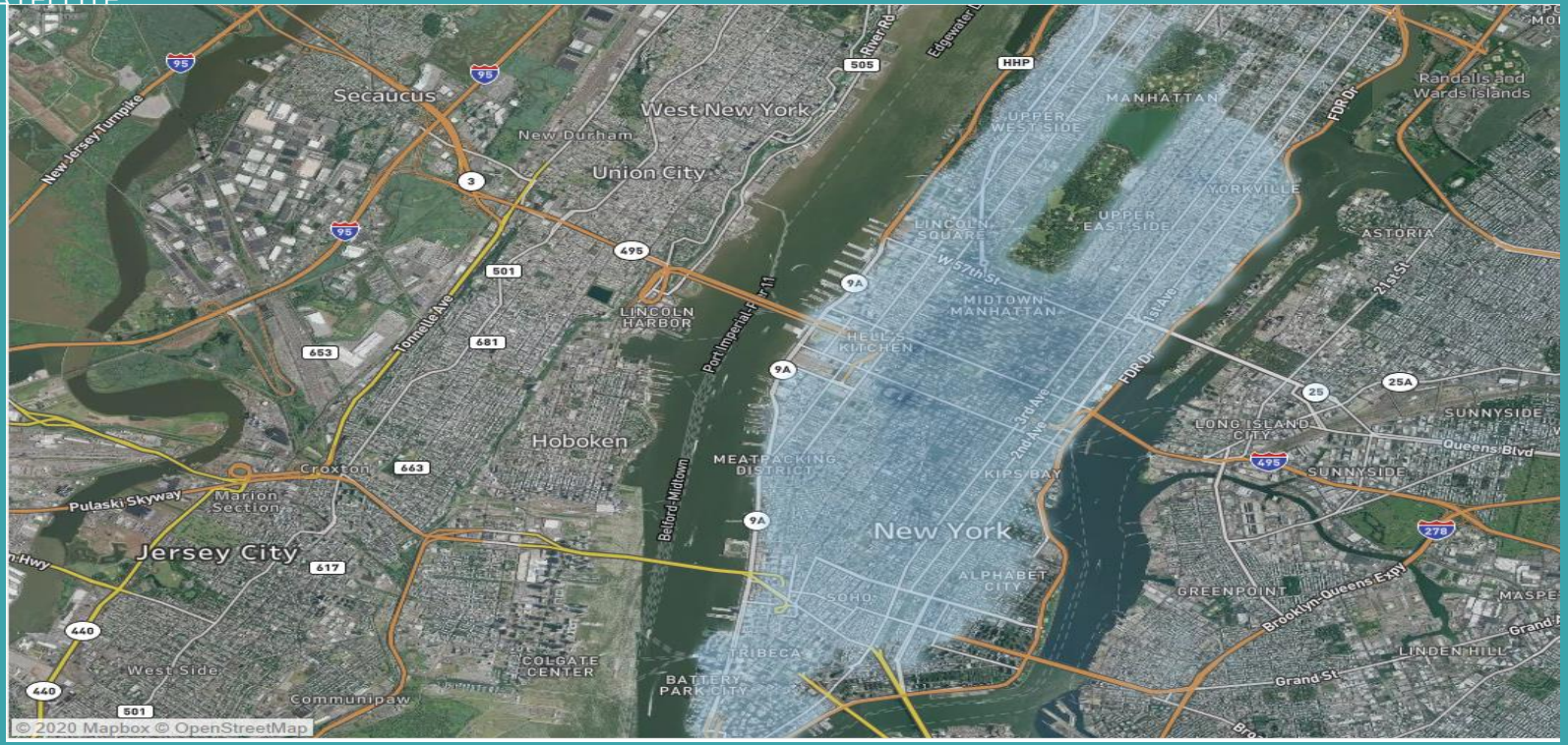


For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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 Densities of Pickups -> Satellite

Areas/localities with the highest densities of pickups: Alphabet City, SOHO, Tribeca, Kips Bay, Upper Westside, Yorkville , Hell's Kitchen , Yorkville, Upper Eastside, Meatpacking District and some parts of Battery Park City -> SATELLITE

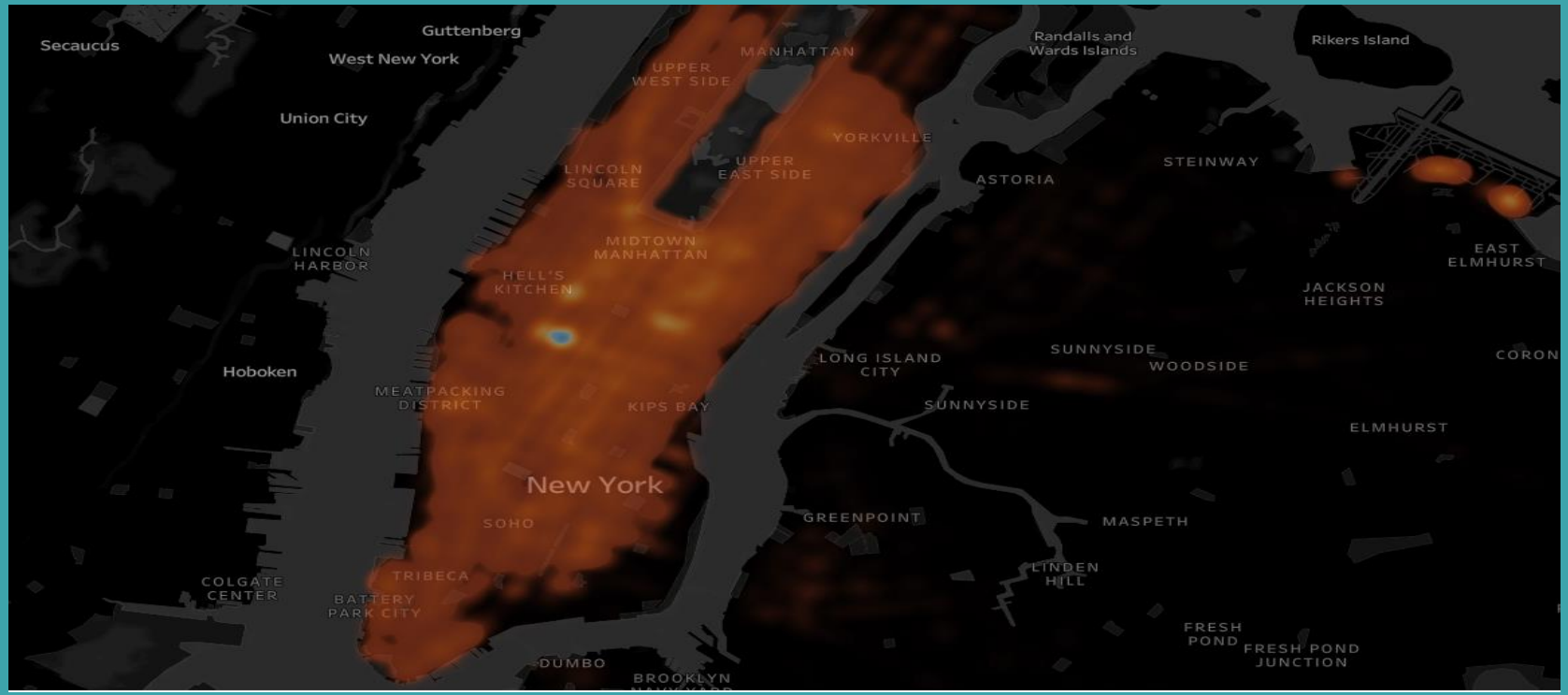


Answer Slide- High Densities of Pickups -> DarkMap(Zip)

Areas/localities with the highest densities of pickups: Alphabet City, SOHO, Tribeca, Kips Bay, Upper Westside, Yorkville , Hell's Kitchen , Yorkville, Upper Eastside, Meatpacking District and some parts of Battery Park City -> SATELLITE

Answer Slide- High Densities of Pickups -> DarkMap(Cities)

Areas/localities with the highest densities of pickups: Alphabet City, SOHO, Tribeca, Kips Bay, Upper Westside, Yorkville , Hell's Kitchen , Yorkville, Upper Eastside, Meatpacking District and some parts of Battery Park City -> SATELLITE

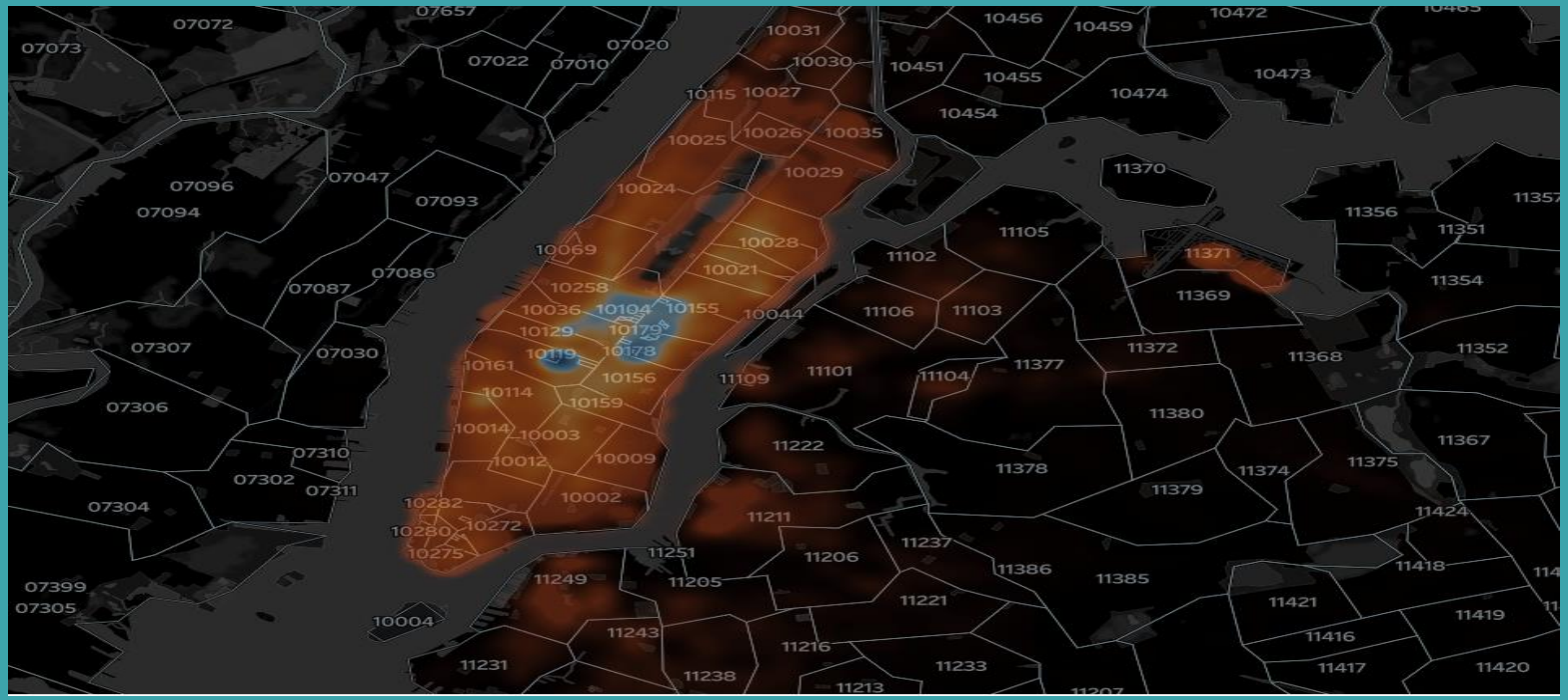


Answer Slide – High Densities of Dropoffs -> Satellite

Areas/ localities with high % of drop-offs: (majority are the same as the pickups) SOHO, Hell's Kitchen, Tribeca, Manhattan , half of Meatpacking district, Empire State area, La-Guardia Airport, KIPS Bay, Alphabet City, DUMBO, Lincoln Square, Morningside Heights

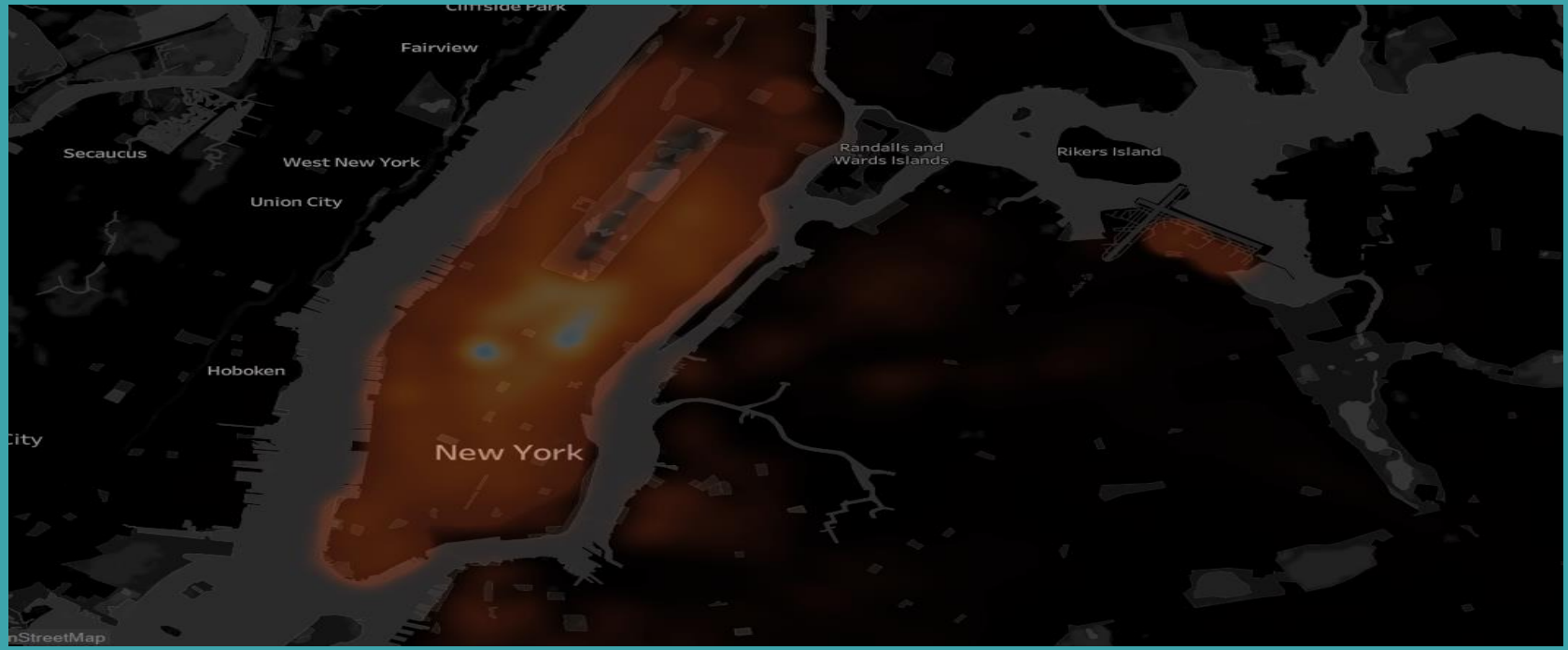
Answer Slide – High Densities of Dropoffs -> DarkMap(Zip)

Areas/ localities with high % of drop-offs: (majority are the same as the pickups) SOHO, Hell's Kitchen, Tribeca, Manhattan , half of Meatpacking district, Empire State area, La-Guardia Airport, KIPS Bay, Alphabet City, DUMBO, Lincoln Square, Morningside Heights



Answer Slide – High Densities of Dropoffs -> DarkMap(Cities)

Areas/ localities with high % of drop-offs: (majority are the same as the pickups) SOHO, Hell's Kitchen, Tribeca, Manhattan , half of Meatpacking district, Empire State area, La-Guardia Airport, KIPS Bay, Alphabet City, DUMBO, Lincoln Square, Morningside Heights



Answer Slide – highest distance-to-duration ratios based on pickups

From this image, we can see that the highest densities of distance to duration pickups are: Hell's kitchen, Upper East Side, Yorkville, Midtown Manhattan, Lincoln Square, Meatpacking District, Kip's Bay, SOHO, Tribeca and Battery Park City . Also included are LGA and JFK Airports

Answer Slide – Highest Distance-to-Duration Ratio based on dropoffs

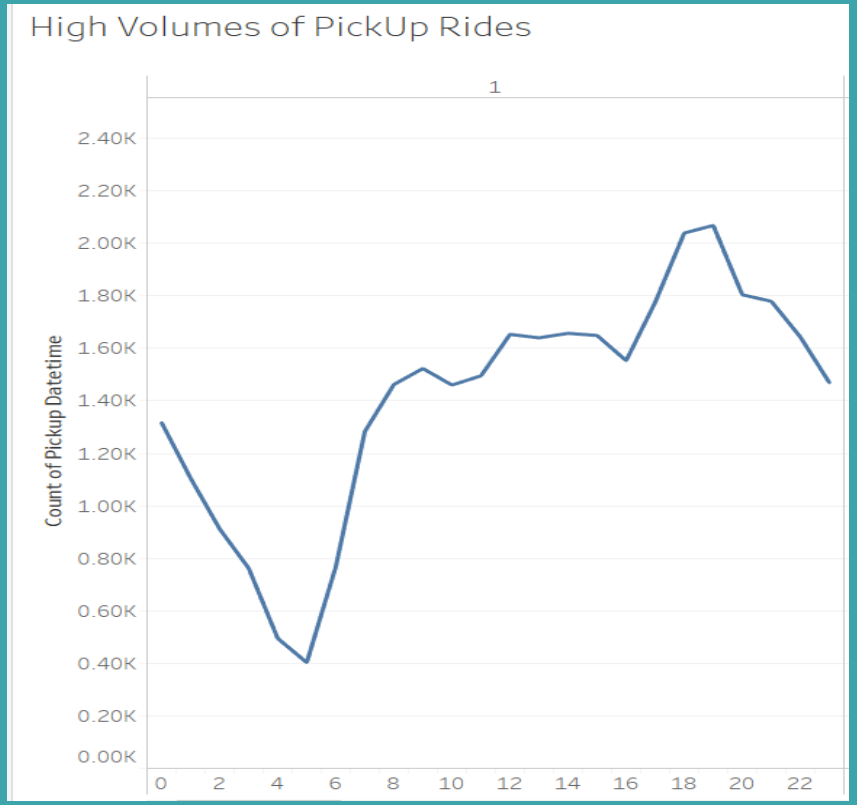
From this image, we can see that the highest densities of distance to duration dropoffs are: Hell's kitchen, Upper East Side, Yorkville, Midtown Manhattan, Lincoln Square, Meatpacking District, Kip's Bay, SOHO, Tribeca and Battery Park City . Also included are LGA and JFK Airports

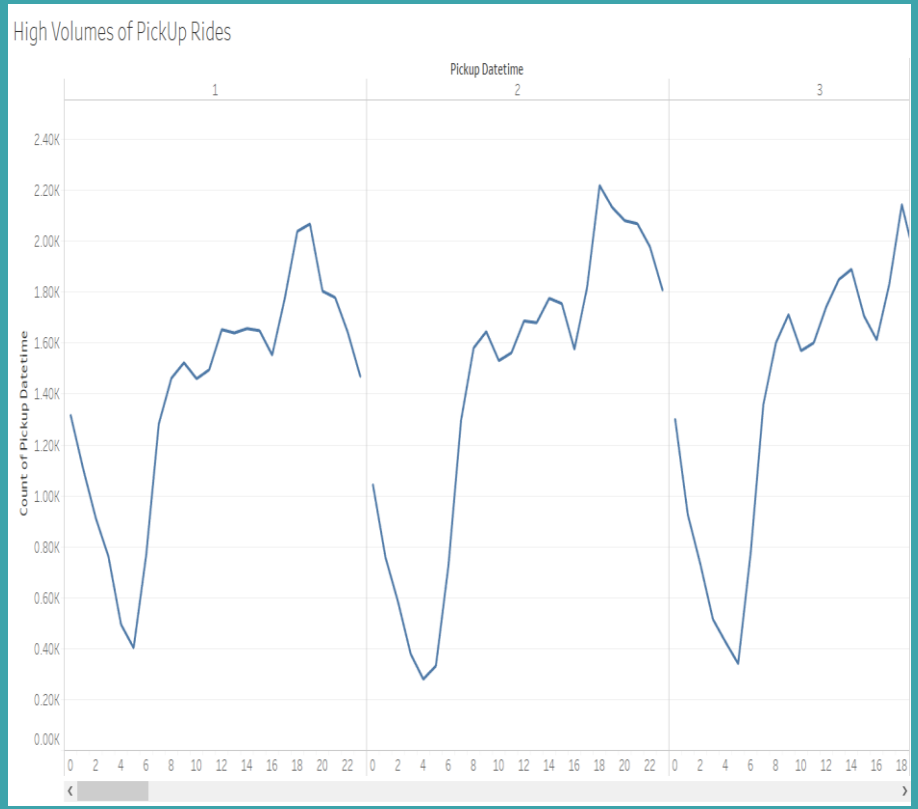
It may not make operational sense to have the service running 24/7, for now.

- 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 – Operational Value for 24/7 Services - > Hours Via Days

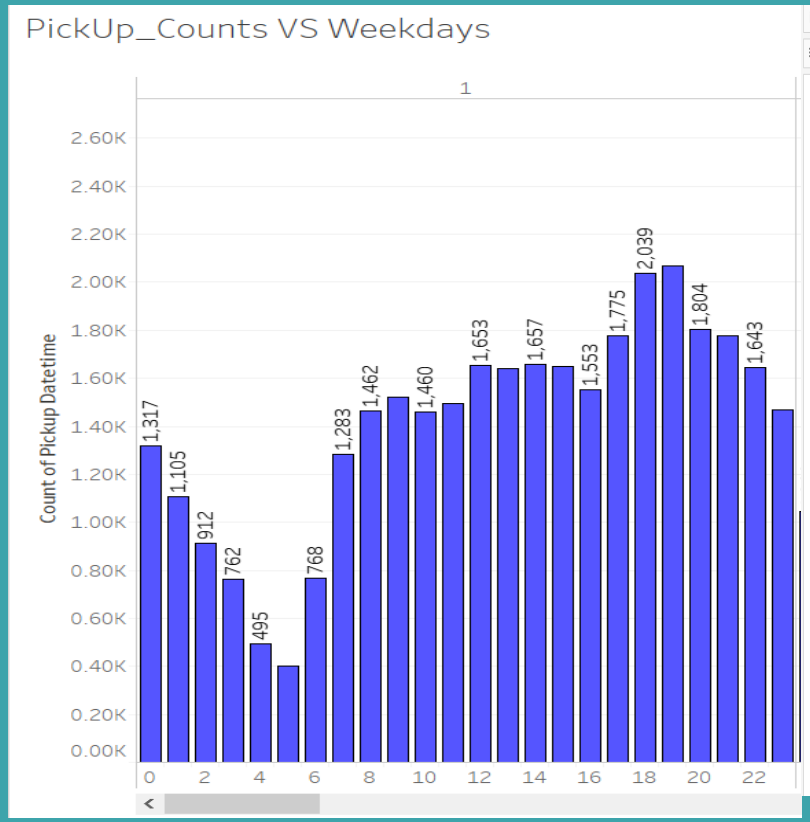
What times throughout the day experience relatively higher volumes of ride pick-ups? The line graph below shows the times throughout the day where there are high pickup volumes: 1800h to 2000h (6pm to 8pm)

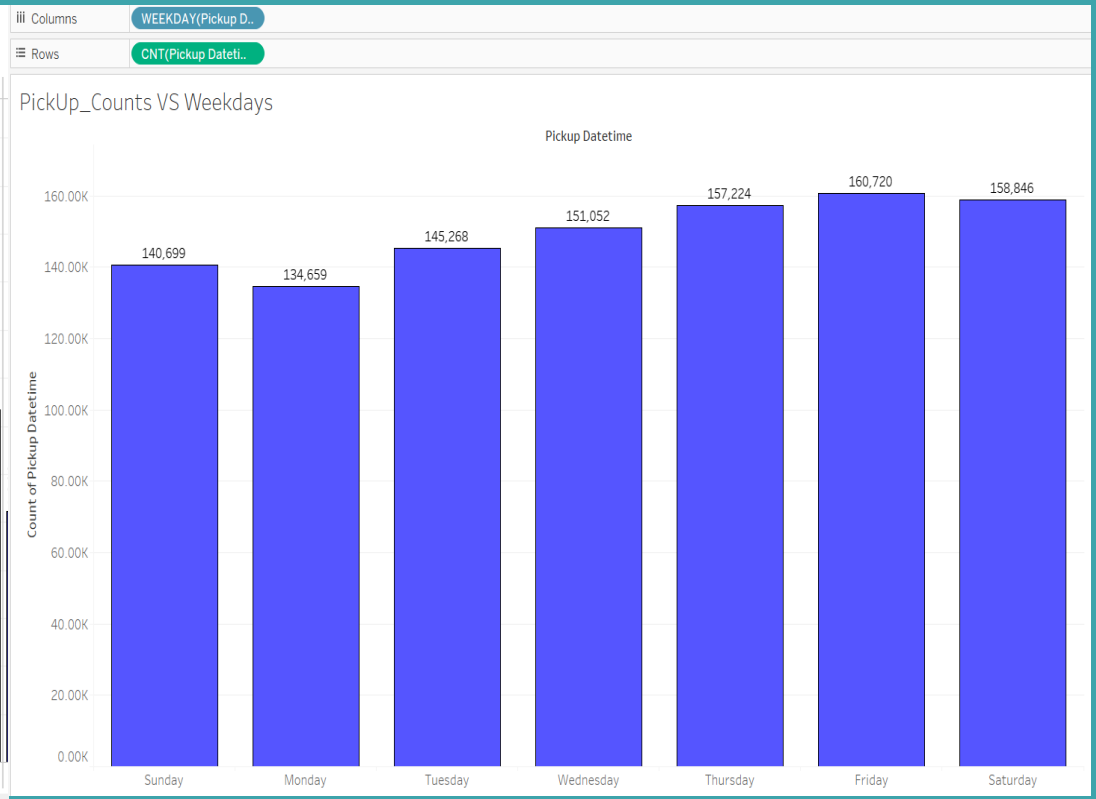




Answer Slide – Operational Value for 24/7 Services -> Days Via Week

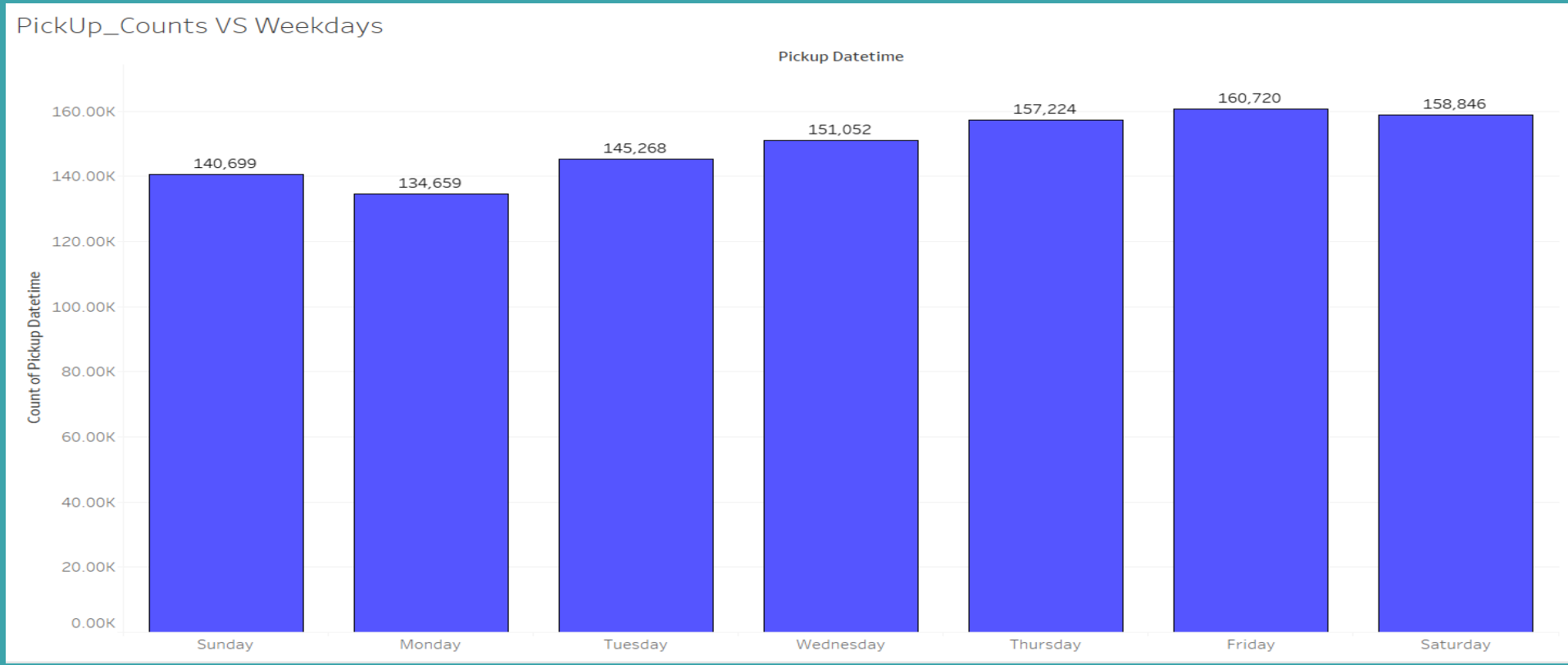
What days throughout the week experience relatively higher volumes of ride pick-ups?The bar graph below shows the days throughout the week where there are high pickup volumes: (starting on a Sunday)





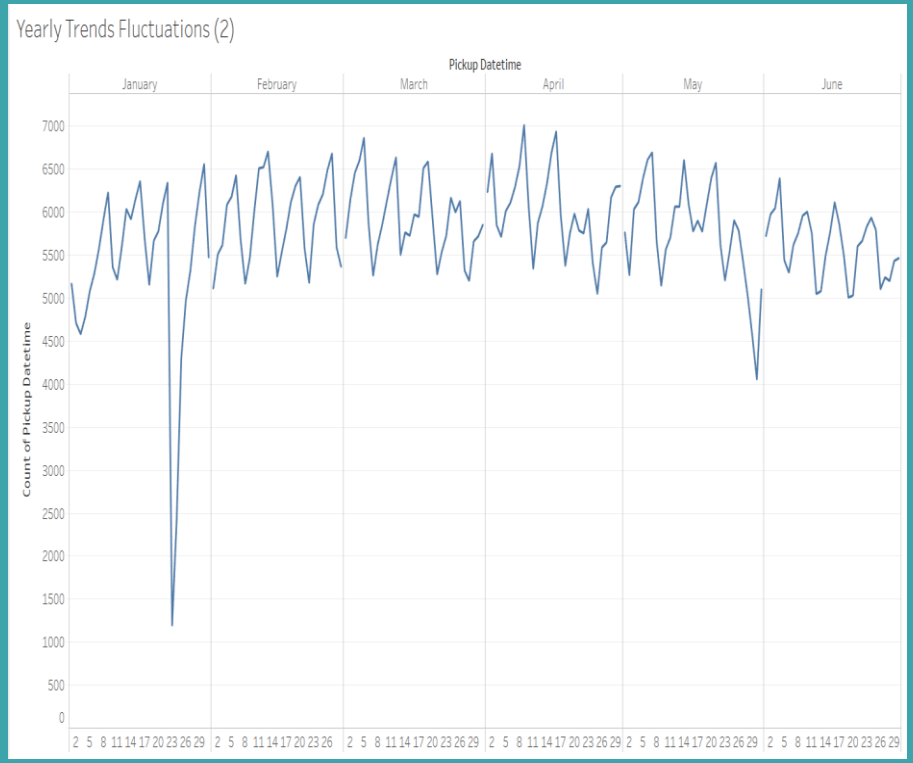
Operational Value for 24/7 Services -> Days in week (MORE DETAILED VERSION)

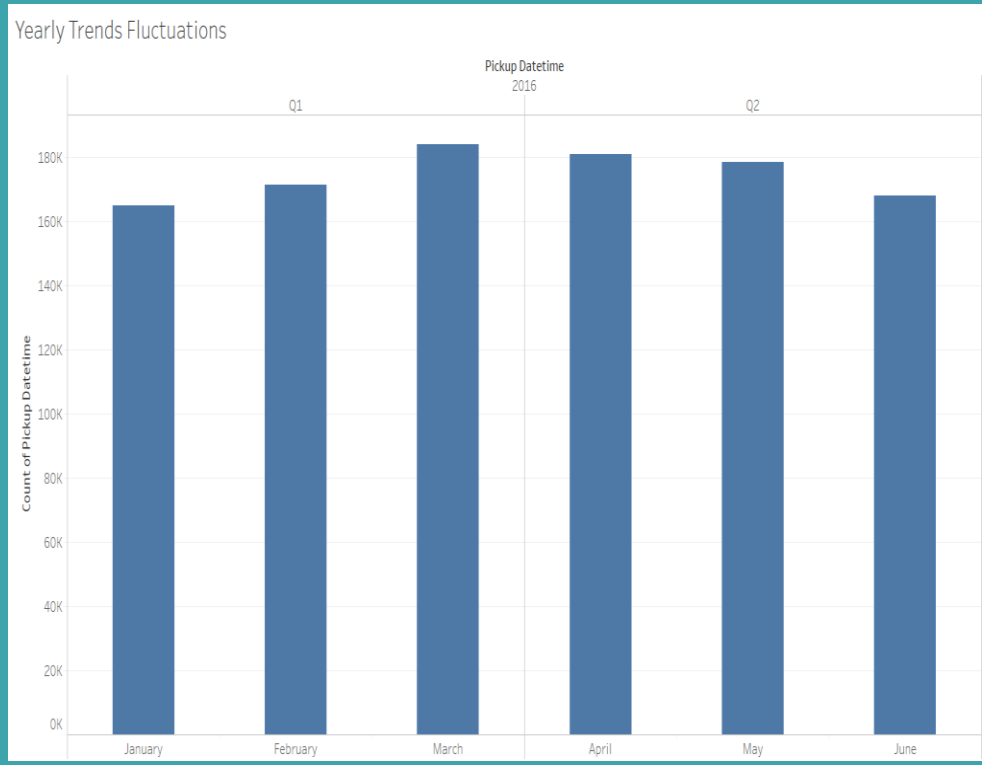
The bar graph below shows the days throughout the week where there are high pickup volumes: (starting on a Sunday). **This shows that Friday had the highest pickups weekly with 160.72k . Immediately followed by Saturday with 158.85k**



Answer Slide – Operational Value for 24/7 Services -> Trend Fluctuation

Analysis of any trend fluctuations throughout the year regarding ride pickup volumes. There appears to be a HUGE DIVE (DECREASE) in rides during 3rd week of January. Also a slight dive/decrease during the last week of May.





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:

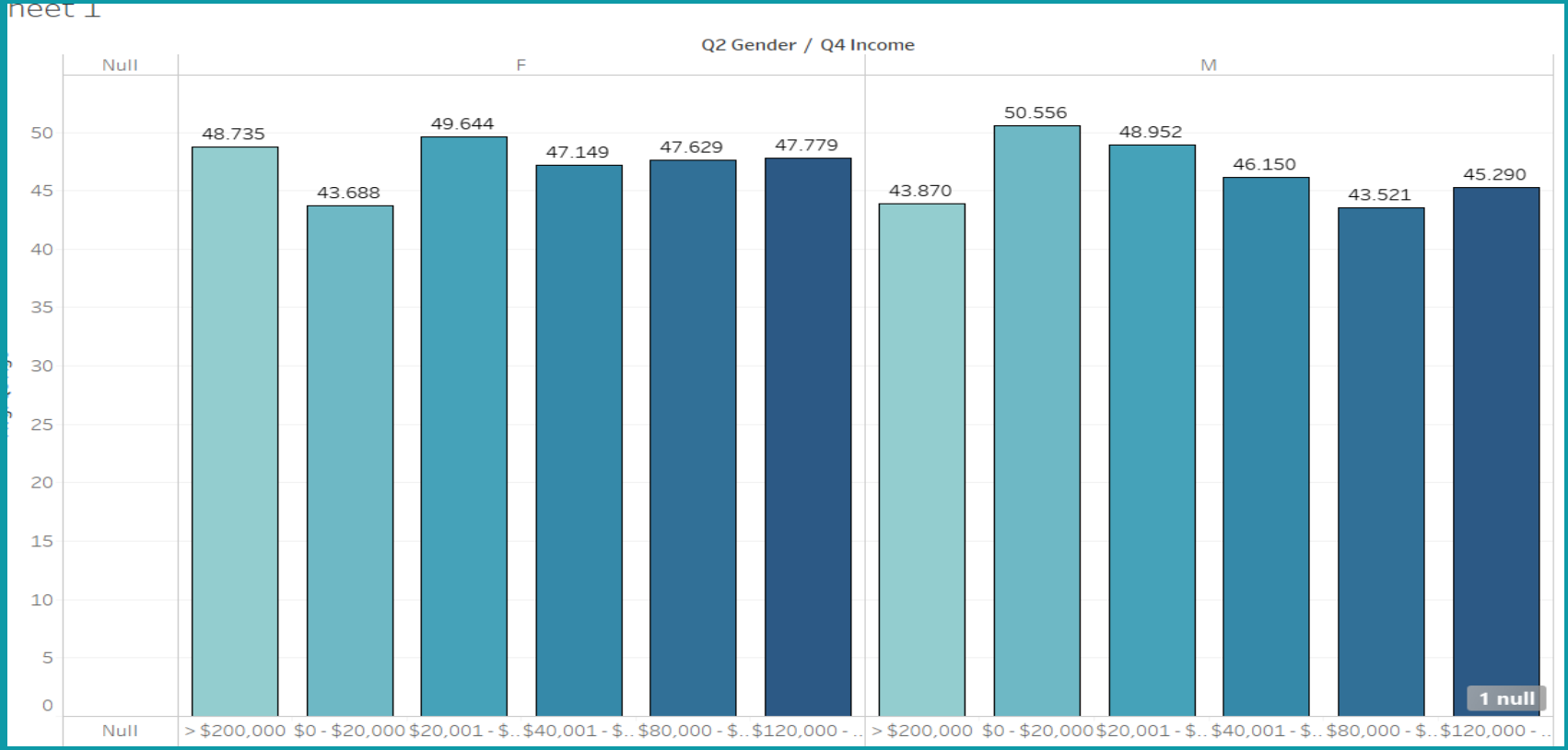
Q1 - What is your email?
Q2 - What gender do you identify as?
Q3 - What is your age?
Q4 - What is your annual income? (income bands)
Q5 - What neighborhood do you reside in?
Q6 - Do you currently use taxis? (Y/N)
Q7 - Do you currently use ridesharing services? (Y/N)
Q8 - Would you use a flying taxi service, 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 –User_Research Data

A simple visualization regarding the breakdown of target audience . Metrics used are: Gender, Age & Income



Answer Slide- User Research Data

A simple query would reveal random selection of data within the User research data set:

```
select *
from user_research
LIMIT 25;
```

Output 25 results

emailofnewyorker+375@gmail.com	F	42	$80,000 - $120,000	Yorkville
emailofnewyorker+94@gmail.com	F	40	$120,000 - $200,000	Midtown
emailofnewyorker+60@gmail.com	M	39	$0 - $20,000	Alphabet City and Loisaida
emailofnewyorker+58@gmail.com	F	61	$0 - $20,000	Tudor City
emailofnewyorker+175@gmail.com	F	28	$20,001 - $40,000	Astor Row (Central Harlem)
emailofnewyorker+86@gmail.com	M	46	$120,000 - $200,000	Theater District
emailofnewyorker+410@gmail.com	M	71	> $200,000	Chinatown

Answer Slide- User Research Data – Demographics Analysis via SQL

Query regarding genders :
select COUNT(*)
from user_research
WHERE q2 = ‘M’;

Result = 192 Male participants

select COUNT(*)
from user_research
WHERE q2 = 'F’;

Result = 307 Female participants

Query determining whether they use ride-sharing services :
select *
from user_research
where q7 = 'Y’;

result = 295 from both genders and all age levels use ride-sharing services, 205 does not use ride-sharing services

Answer Slide- User Research Data – Demographics Analysis via SQL

```
select *
from user_research
where q3 < 30;
```

Result = 109 participants below 30 years old

```
select *
from user_research
where q3 >= 30;
```

Result = 390 participants aged 30 and above

```
select *
from user_research
where q4 > '100000';
```

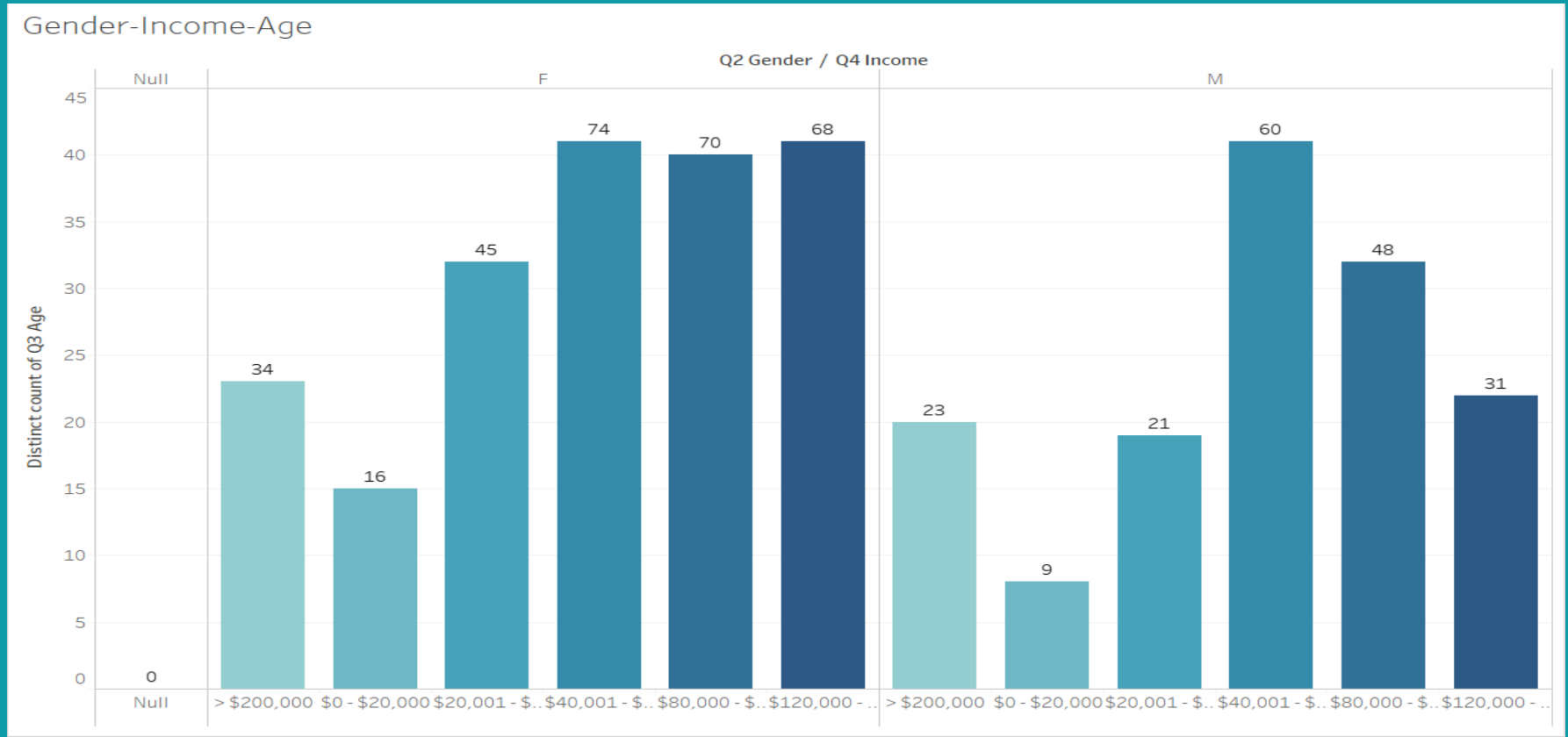
result = 474 participants have yearly incomes higher than 100k (meaning 25 participants earn less than 100k)

```
select *
from user_research
where q6 = 'Y';
```

result = 406 uses taxis, 93 doesn't use it.

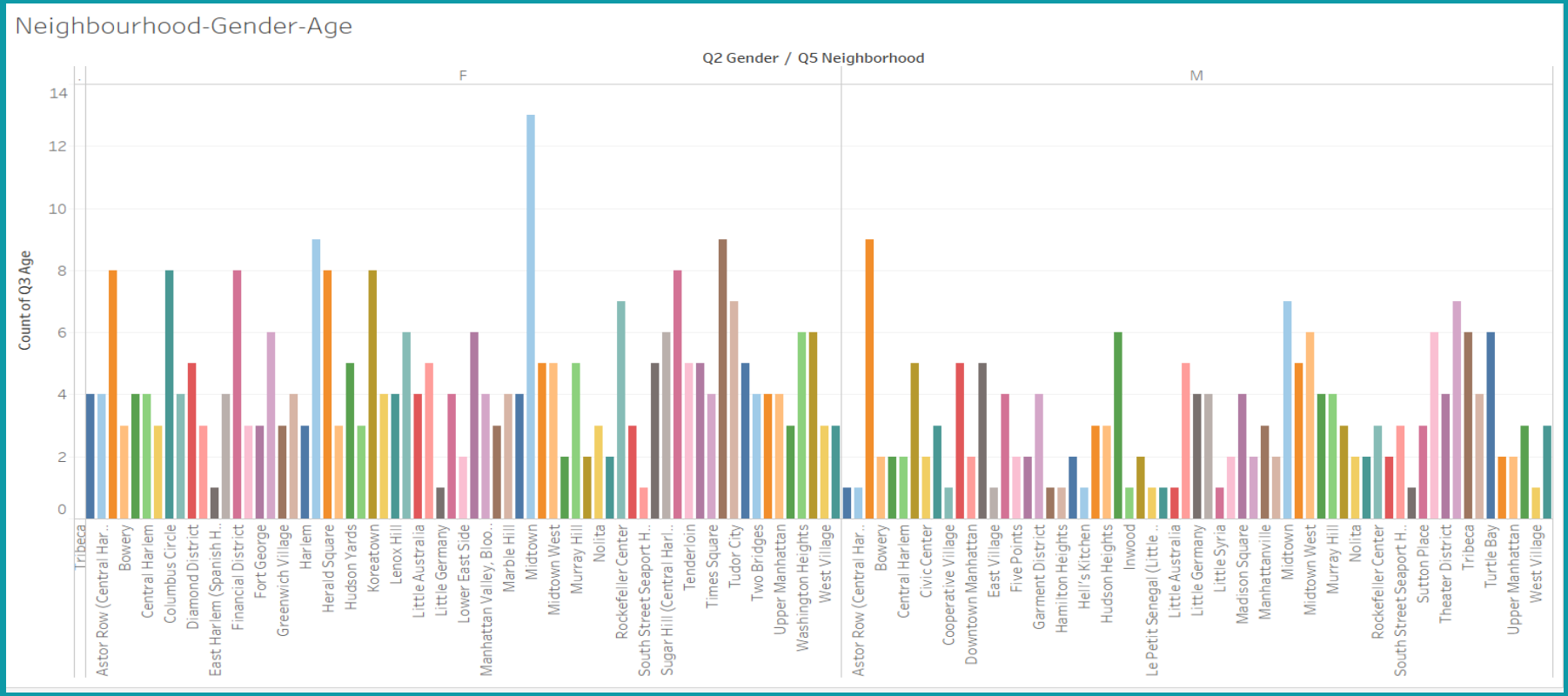
Answer Slide- Graph based on Gender, Income & Age

The graph shows that out of 307 Female passengers, 212 were earning between 40k to 120k+ annually.
Out of 192 Male passengers, 139 were earning between 40k to 120k + annually.



Answer Slide- User Research Data- Neighbourhoods

The dataviz below gives another overview of the diversity in the neighbourhoods where respondents reside in. The highest lives in Midtown Manhattan with 13. Then at second are: Tribeca, Battery Park 7 Hell's Kitchen with 9 respondents.



Answer Slide- Analysis and Findings Regarding Demographics (PART 1)

Based on the over-all findings , that could be proven using the 2 bar graphs and the SQL scripts, here are the results :

1. For both genders, the average yearly income is within the 80k range. This means that the target audience is on the higher bracket.

2. 294 over-all uses ride sharing services, 204 doesn't. This means that more than 50% are willing to spend money on such services.

3. 406 uses taxis and 93 doesn't use. 81% of the respondents use taxis.

4. A total of 204 never used ride-sharing services

5. 99 respondents said they will NEVER use a flying taxi service , while a whopping 400 said YES!

6. Among the respondents who said yes to flying taxi services, the maximum amount willing to be paid was $46 per ride and the minimum was $5

7. Those who said 'NO' they will never use flying taxis, most have mentioned that it was related to 'safety & security' and also a large group said 'it was too expensive'.

Additional Analysis and Findings Regarding Demographics(PART 2)

Based on the several key findings that were observed by using the User_Research DataSet, it could be concluded that:

8. The target audience comes from a very diverse locations in NYC, but a great majority are earning an average of 80k yearly. This means that they are on the upper bracket of employees.

9. More than 50% have used ride sharing services before. Although not by a huge margin, this is still an indication that the target market is somewhat ready for our product.

10. 81% have used taxis before.

11. More than 75% are willing to use a flying taxi service once it is being offered. 157 Males said they were and 243 Females.

12. The 99 respondents who said NO to flying taxis came from a wide variety of neighbourhoods.

13. The same 99 respondents who said NO to flying taxis have a very diverse range of yearly income. Some are eraning 40k yearly while other are on the 120k and a handful were earning 200k yearly.

CONCLUSION: There is NO SPECIFIC target audience where Flying Taxis are veering towards. The genders vary almost equally, the place of residence are also widely distributed, the age group is also very diverse. The same thing applies to their annual income.

The ONLY things that really resonates regarding the 99 respondents who said NO were: they were deeply concerned with the safety & security of flying taxis. A second reason is they have an image/ notion that it would be very expensive.

BONUS SECTION: TABLEAU PUBLIC LINKS:

My Main Tableau Public Link:
https://public.tableau.com/profile/frederick.zoreta.first#!/

Taxi Rides Version 1(Yearly Trends & Fluctuations):
https://public.tableau.com/profile/frederick.zoreta.first#!/vizhome/Udacity_Flyber_Project1/YearlyTrendsFluctuations2

Taxi Rides Version2 (Highest Drop Offs):
https://public.tableau.com/profile/frederick.zoreta.first#!/vizhome/Udacity_FLYBER_Version2/Highest_Dropoffs-DarkMap

User Researc (Neighbourhoods-Gender=Age):
https://public.tableau.com/profile/frederick.zoreta.first#!/vizhome/Udacity_User-Research/Neighbourhood-Gender-Age

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

(Fill out your answer here)

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)