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

- **What are taxis used for?**
  - Private transportation between 2 locations for people
- **What are the characteristics of the users that leverage them?**
  - People who are in a rush
  - People who do not have cars
  - People who do not wish to take other forms of transportation
  - People who wish for privacy and economical way of transportation
- **What are existing pain points with taxis?**
  - You cannot find one when you need one
  - Pricing uncertainty
- **What are the existing pain points with digital ride-sharing services?**
  - Inconsistent quality of rides
  - Safety
  - Low margins for drivers

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

What user improvements do you hypothesize a flying taxi service would have over the existing state of taxis today?

- Much faster commute times
- More fancy joy-ride

What market improvements do you hypothesize a flying taxi service would have the existing taxi service industry & physical road infrastructure today?

- Existing taxi services will be optimize for shorter trips/last mile
- More infrastructure spending will be directed to improve more “local” commutes

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

- **How many records are in the dataset**
  - 1,048,468
- **What does each record represent?**
  - One ride
- **What is the primary key?**
  - Id
- **What date range is your dataset bound to?**
  - 1/1/2016 to 30/6/2016
- **What are the geographical bounds of this dataset?**
  - New York City
- It is not limited to Manhattan. Brooklyn, Queens, Staten Island, the Bronx, and New Jersey are included
- **Where are most of the data points centralized at?**
  - New York, New York
- **Are there outliers?**
  - Yes

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

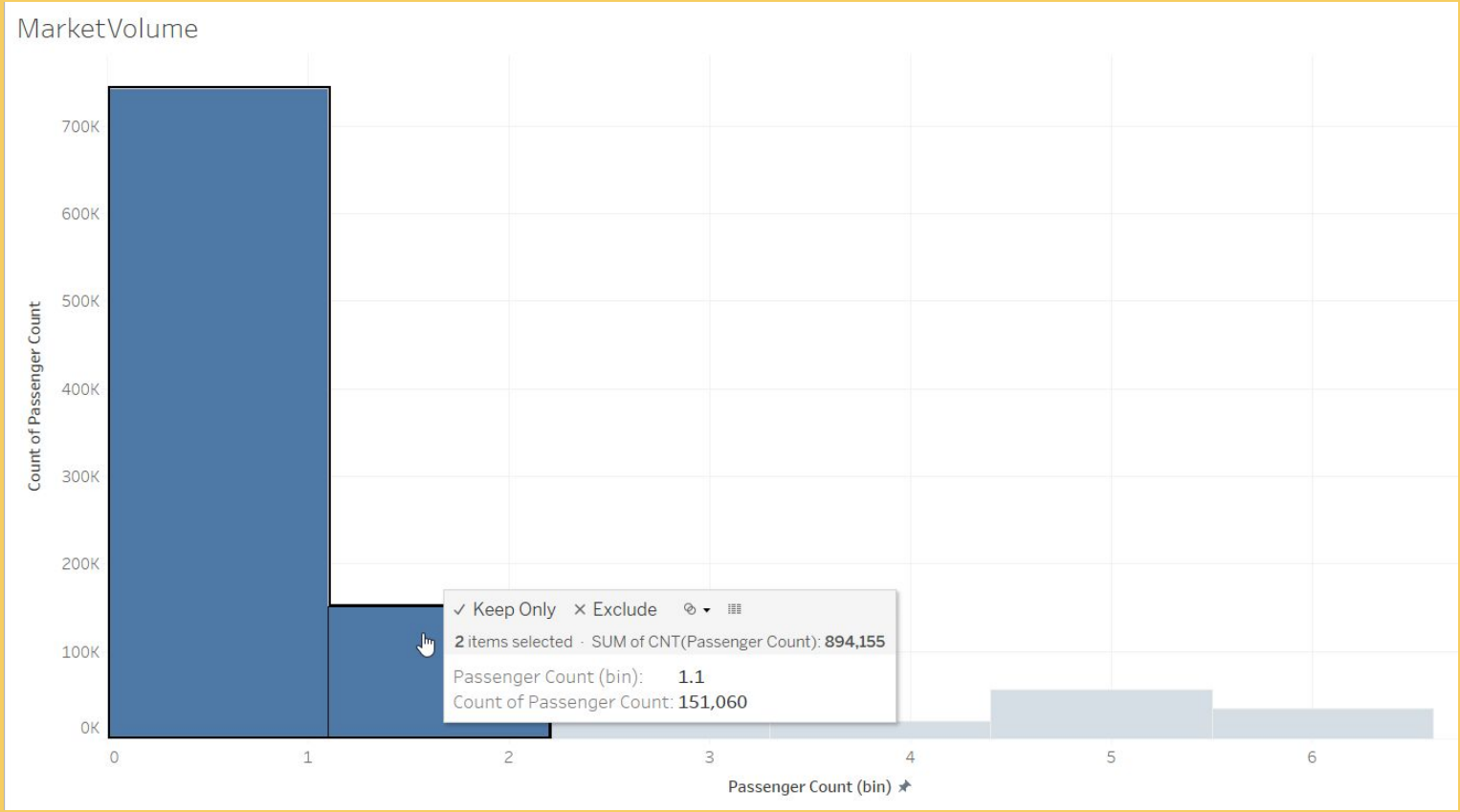
Answer Slide

	duration	distance	passenger counts	duration-to-distance ratio	price
Mean	911	3.46	1.664	416	22.85
Median	665	2.11	1	279.87	17.35
Std Dev	2482	4.38	1.314	4250.73	18.33
2 Std Dev	5564	8.76	2.628	9500	122.12

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

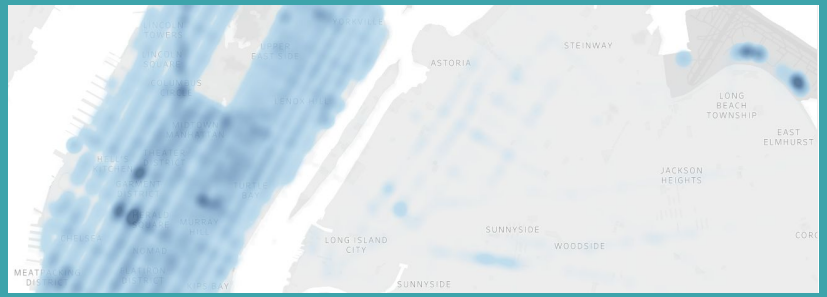


For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- 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

- **Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?**
  - Midtown, theater district, garment district, Herald Square



- **Based on Drop-offs**
  - Herald Square, Theater District, Murray Hill, Turtle Bay

Answer Slide

- **Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?**
  - JFK Airport, Garment District, Herald Square, Turtle Bay

- **Based on drop-off?**
  - Herald Square, Turtle Bay

Answer Slide

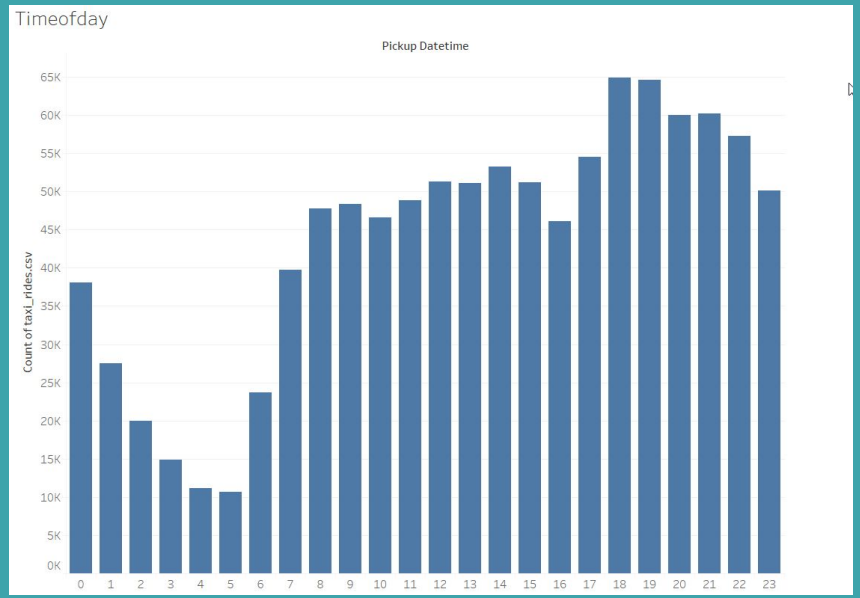
- For any of the neighborhoods identified, are there any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off? What makes them suitable?
    - JFK Airport to Manhattan NY, Takeoff and Landing Infrastructure

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.

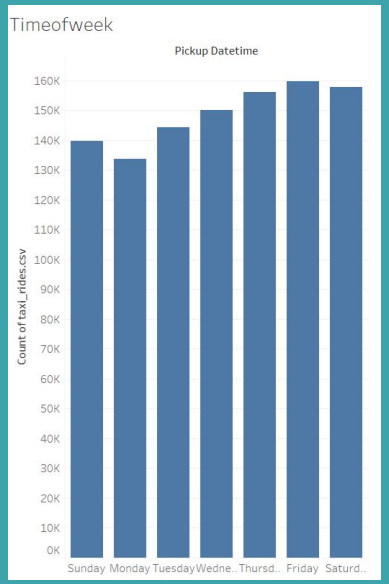
Answer Slide

- **What times throughout the day experience relatively higher volumes of ride pick-ups?**
  - 1700 - 2300



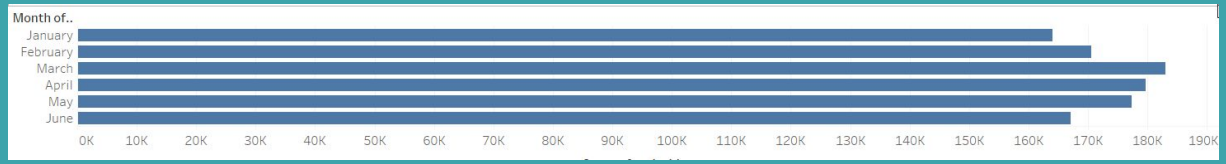
Answer Slide

- **What days throughout the week experience relatively higher volumes of ride pick-ups?**
  - Thurs - Saturday



Answer Slide

- **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.**
  - March, April, May (Data only available for 1st half of 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

- Is there an inclination of better Flyber adoption based on:
  - Gender - Both Genders are equally likely to adopt
  - Age - 54-64 (86.42% likely to adopt)
  - income level - 120k - 200k (85.86% likely)
  - neighborhood of residence

Adoption(Age)

	Q2	
Q8	F	M
N	20.85%	18.23%
Y	79.15%	81.77%

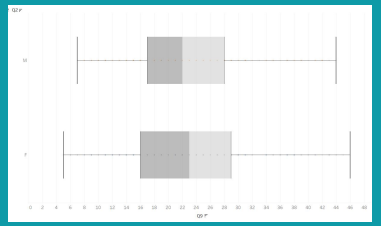
Washington Heights
Upper Manhattan
Tudor City
Sugar Hill (Central Harlem)
SoHo
Nolita
Morningside Heights
Marble Hill
Little Australia
Lincoln Square
Le Petit Senegal (Little Senegal)
Hudson Heights
Greenwich Village
Financial District
East Village
Bowery
Astor Row (Central Harlem)

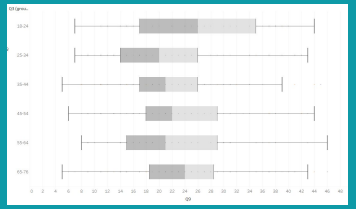
Adoption(Age)

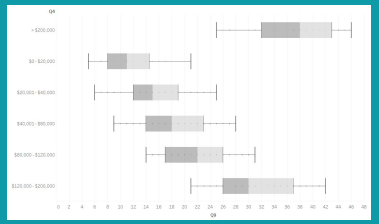
	Q3 (group)					
Q8	18-24	25-34	35-44	45-54	55-64	65-76
N	23.44%	20.48%	19.28%	21.69%	13.58%	20.95%
Y	76.56%	79.52%	80.72%	78.31%	86.42%	79.05%

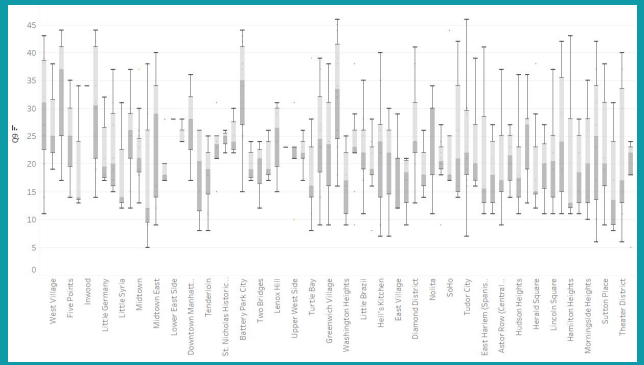
Answer Slide

- 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?
  - Female, 35-44, 40k-120k, Female, 65-76, 40k-80k
  - Male, 45-54, 40k - 80k
  - Female, 65 -76, West Harlem

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.

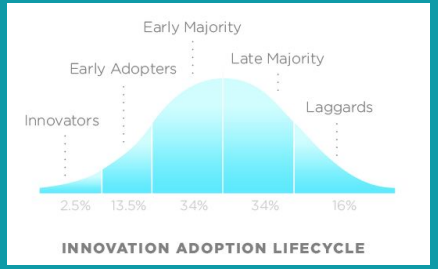
Section 2: Proposal Synthesis

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 Retention

I believe that User Retention should be the key focus area for Flyber at this time of the product development cycle. As flying taxi services are a new product, the adoption would likely follow the innovation adoption cycle. Concurrently, from the survey results, it shows that people earning between 120k to above 200k have the highest willingness to pay (WTP) while, age group of 54-64 has the highest willingness to subscribe to the service.

While at this stage of the process, it is also unlikely that we would want to invest on a large fleet of flying taxi initially. The focus should be on covering routes that are very often travelled to-and-fro by users and getting them to adopt a new behavior/habit of commuting by flying as regularly as possible.

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

Answer Slide

Passenger retention = percentage of passengers who take a ride after the first app event (say, first successful trip).

Trips per active passenger = number of rides/active passengers over time.

Existing Customer Growth Rate = Monthly Revenue Growth Rate = (MRR at the End of Month - MRR at the Start of Month) / MRR at the Start of Month

Repeat Purchase Ratio = Number of Returning Customers / Number of Total Customers

NPS (Net promoter score) = 1-10 ("How likely are you to recommend our company to a friend or colleague?")

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

Create hypotheses around what thresholds your KPIs

Passenger retention > 30%

Trips per active passenger = min. 3x per month

Monthly Revenue Growth Rate = 1.3x

Repeat Purchase Ratio = 0.5x

NPS (Net promoter score) = >7.5

- What times/days of operation should the service run for?
  - Thurs - Saturday

- How many pick-up / drop-off nodes should we have?
  - 5
- Where should the nodes be located?
  - Herald Square, Theater District, JFK Airport, Turtle Bay, Midtown Manhattan
- Should we initially use copters or homegrown hardware?
  - Copters that are white labeled
- Should the pricing be fixed or dynamic? At what rates?
  - Dynamic to better match supply and demand, 17-30 per mile

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

Answer Slide

Daily Pickups NY Taxis: ~140,000
Market Share(Target): 10% (14000)

Baseline Conversion Rate: 10%
Minimum Detectable Effect: 10%
Statistical Significance: 95%
Sample Size per Variation: 14,000

30% user retention across a timeline of 1 month
14,000/0.3 = 47,000 customers (sample size) at the start

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

Answer Slide

Passenger retention = percentage of passengers who take a ride after the first app event (say, first successful trip).

Trips per active passenger = number of rides/active passengers over time.

Monthly Revenue Growth Rate = (MRR at the End of Month - MRR at the Start of Month) / MRR at the Start of Month

Repeat Purchase Ratio = Number of Returning Customers / Number of Total Customers

NPS (Net promoter score) = 1-10 ("How likely are you to recommend our company to a friend or colleague?")

Answer Slide

Event - rideCompleted
Definition - Triggered when the driver hit “Ride Completed” button on driver app
Properties - passenger_count, rider_id, driver_id, dropoff_timestamp, dropoff_latitude, dropoff_longitude, ride_cost

Event - riderPickedUp
Definition - Triggered whenever the driver hits the “Passenger is picked up” button on the Flyber driver app.
Properties - passenger_count, rider_id, driver_id, pickup_timestamp, pickup_latitude, pickup_longitude

Event - initiateApp
Definition - Triggered when the user opens the flyber app
Properties - rider_id, initiate_timestamp, initiate_latitude, initiate_longtitude, duration_counter

Event - searchRide
Definition - Triggered when the user hits search for rides
Properties - rider_id, search_timestamp, search_latitude, search_longitude, search_address, pickup_address, pickup_latitude, pickup_longitude, destination_address, destination_latitude, destination_longitude, estimated_price_of_ride, notes_for_pilot

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

Answer Slide

1. Rate your flight? 1-5 stars
2. What other starting and ending locations you would like us to expand to?
3. How likely will you recommend this service to a friend/colleague? 1-10
4. What other things would you like us to improve?
5. Where do you live?
6. What is your occupation/industry?
7. What is your age?
8. Why do you use flyber?
9. What are the alternatives you considered?

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

Both men and women are likely to use the service (~80%)

54-64 Age Group (~87%)

120k - 200k income level and living in more charming districts of NYC

Pain Points

Lack of Time

Sufficient Disposable Income (Target Group)

Traffic in Congested Roads

Product Proposal

Flyber is the next generation transportation sharing platform that is meant for users who wishes to skip traffic and live the high life. Giving them the ability to buy time back so they can spend them on activities that money cannot buy.

Risks involve any known unknowns (post-launch)

- Too much demand for service
- Too little supply of copters
- Pilot Fatigue (if human operated)
- Hacking (if autonomous)
- Regulation & Compliance to Aerospace safety standards
- Initial Hype around product which may not be sustainable
- Overhead cost for ensuring safety compliance
- High fixed cost for building out infrastructure and copters

Cross Functional Stakeholder Teams

- Data Science
- Design
- Product Marketing
- User Research
- Data engineering
- Client Engineering
- Server Engineering
- Legal, privacy and compliance teams
- Finance & Business Operations
- Quality Assurance Engineering