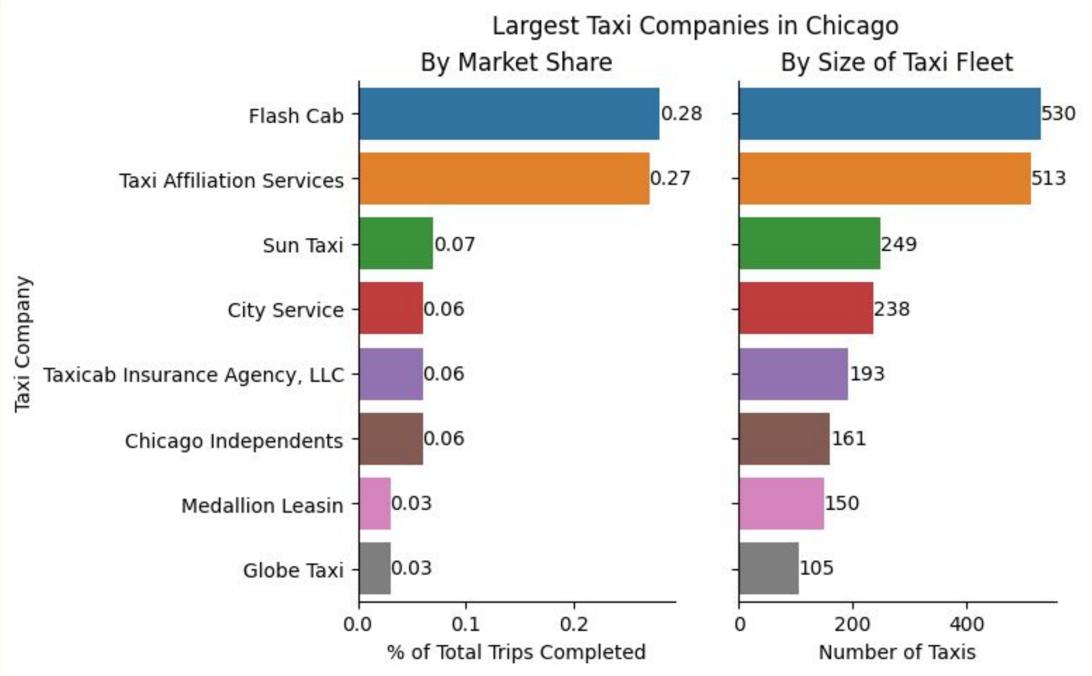
Question 1:

Chicago Mobility Service Launch Evaluation Report

Part 1: Competitor Landscape



<u>Recommendation</u>

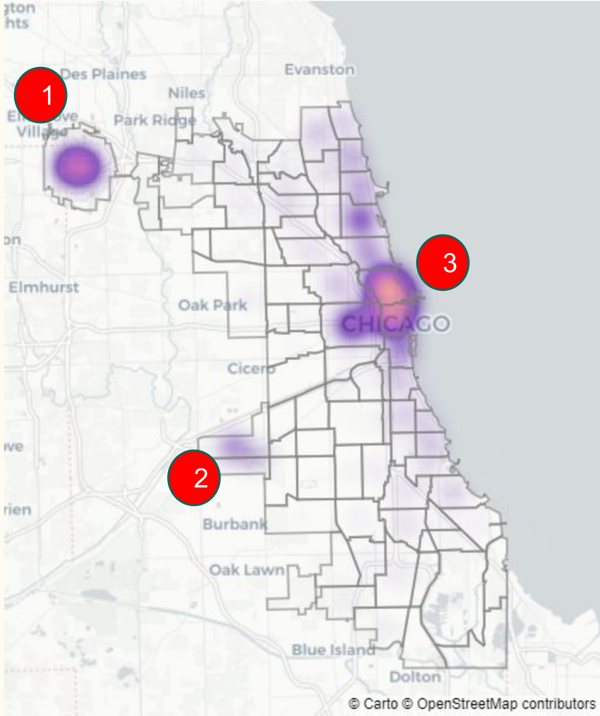
Conduct in-depth business analysis on these top 2 companies to better understand:

- Any product/feature gaps?
- Any underserved population?

<u>Key Insights</u>

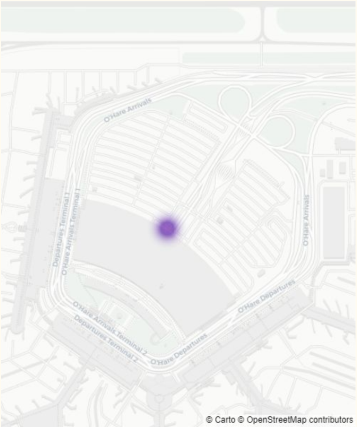
Top 2 companies (*i.e: Flash Cab, Taxi Afflication Services*) command >50% of total market share

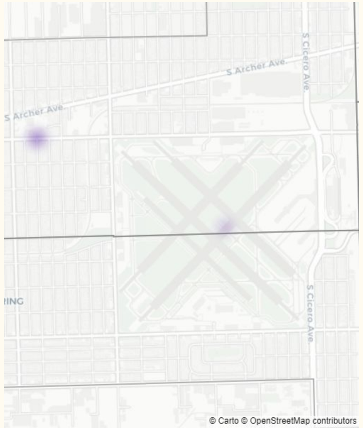
Part 2: Mobility Market (Demand)

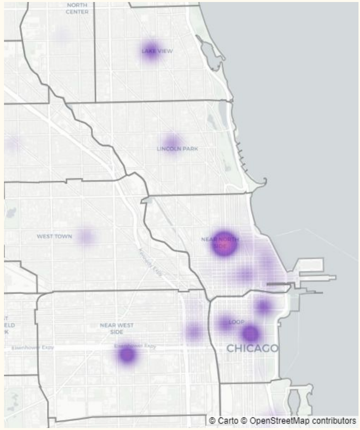


<u>Key Insights</u>: Top pickup hotspots

<u>Recommendation</u>: Prioritise these locations for initial launch







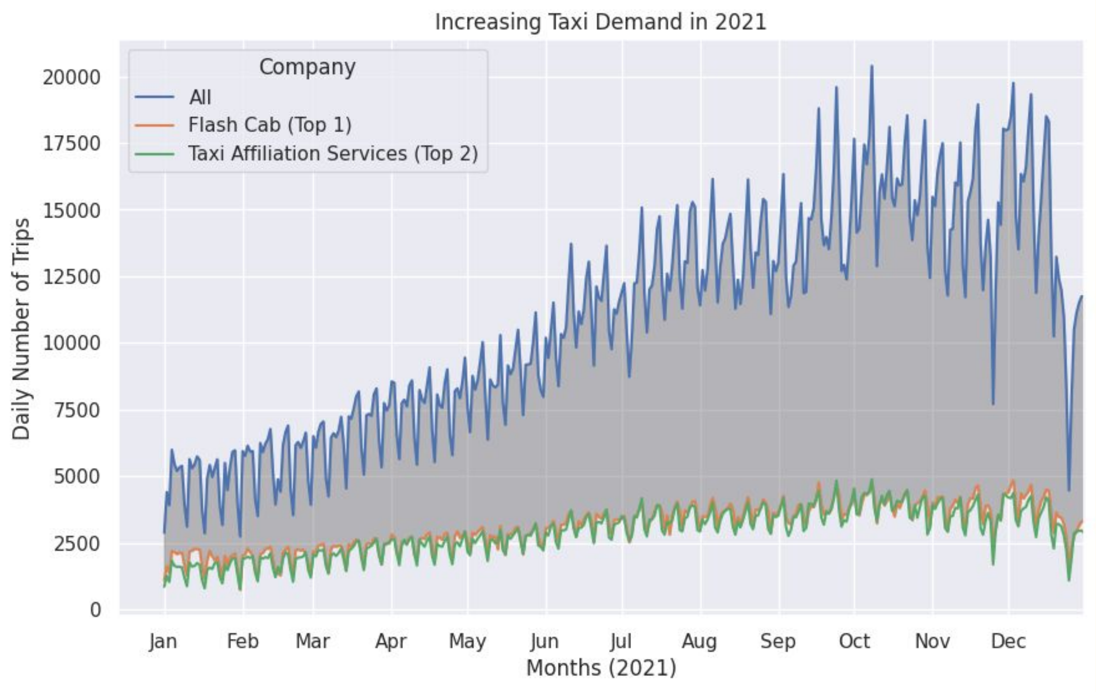
1 O'Hare Airport

2 Midway Airport

3 Popular Community Areas

- *Near North Side*
- *Near West Side*
- *Loop*
- *Lake View*

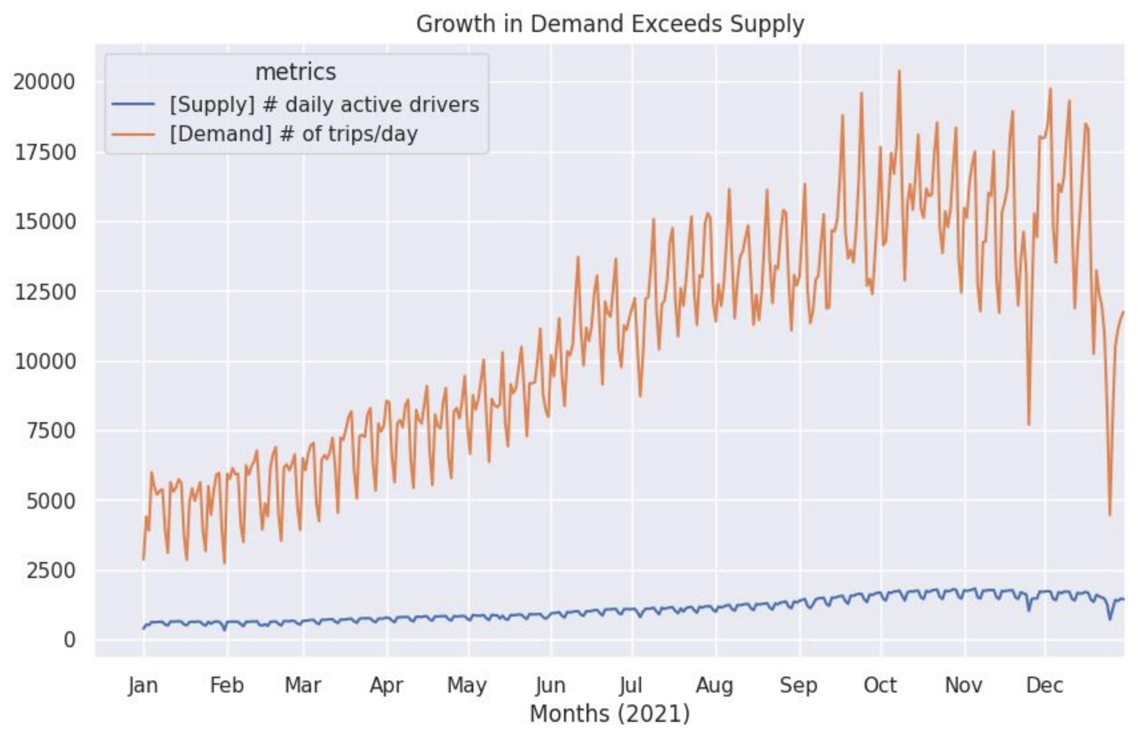
Part 2: Mobility Market (Demand)



<u>Key Insights</u>

- **Clear increasing trend** in demand throughout the years (likely due to economy recovery post-covid)
- The top two companies are unable to pick up the **pace of demand growth**.
    - This demand gap can be translated into **business opportunities**

Part 2: Mobility Market (Supply)



Key Insights

- The growth in **demand** is much **steeper** than **supply**

Recommendation

- Laser focus in **expanding supply** to to utilise this opportunity to **win the market**

Part 3: Summary

<u>Summary</u>

1. **Top Competitors** (which holds >50% of current market shares)
   - *Flash Cab*
   - *Taxi Afflication Services*
2. Strong signals of **supply crunch** in current market
   - *Good timing to enter the market if we can tackle the supply issues*
3. Initial Launch Areas Recommendations
   - Populous community area
     - *Near North Side*
     - *Near West Side*
     - *Loop*
     - *Lake View*
   - Airports
     - *O'Hare Airport*
     - *Midway Airport*

Question 2:

Chicago Ride Pricing Model

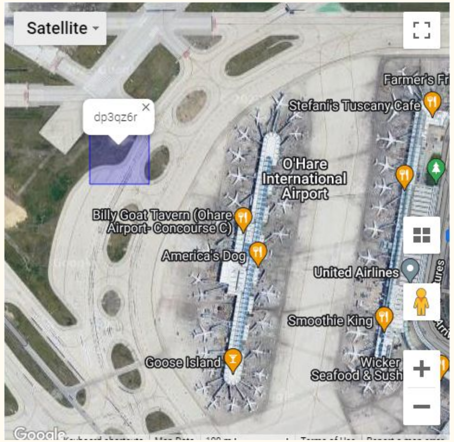
Methodology: Feature Engineering

<u>Demand Score</u>

- Idea: Create a table that maps `*day_of_week*`, `*hour*`, `*geohash*` to a <u>**demand score**</u>
- Method:
  - Step 1: Calculate the **number of trips** aggregated by \`day_of_week\`, \`hour\` and \`***pickup_geohash***\` using data from **training period**
  - Step 2: **Normalise** number of trips as an indicator of demand

Example: Top 5 highest demand score

day_of_week	hour	geohash	num_trips	demand_score
0	19	dp3qz6r	276	10.487691
0	20	dp3qz6r	271	10.288883
0	18	dp3qz6r	261	9.891267
5	16	dp3wmge	258	9.771982
5	17	dp3wmge	254	9.612936



Top highest demand area

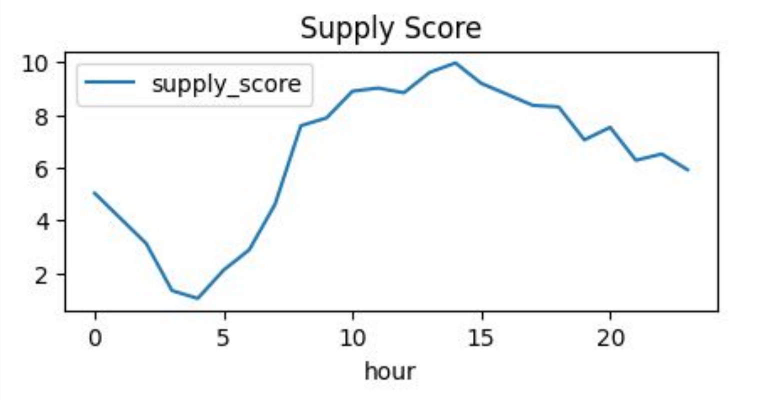
- **O'hare airport** (on Sunday from 1800-2000)

Methodology: Feature Engineering

<u>Supply Score</u>

- Idea: Create a table that maps `*day_of_week*`, `*hour*`, `*geohash*` to a **<u>supply score</u>**
- Method:
    - Step 1: Calculate the **number of drivers** aggregated by `*day_of_week*`, `*hour*` and `***dropoff_geohash***` from data during **training period**
    - Step 2: **Normalise** number of drivers as an indicator of demand

Sanity Check:



Observation: The trend of supply score seems reasonable

- Peaked during working hour
- Dropped during midnight

Methodology: Feature Engineering

Traffic Score

- Idea: Create a table that maps `*day_of_week*`, `*hour*` to a **traffic score** (of Chicago city in general)
- Method:
  - Step 1: Calculate the **mean driving speed of trip** (miles per hours = `*Trip Miles*` / `*Trip Seconds*`)
  - Step 2: Calculate **average driving speed** aggregated by `*day_of_week*`, `*hour*` from data during the **training period** as an indicator of **traffic situation** of Chicago city as a whole

Example: Top 5 highest traffic score
(higher means less congested)

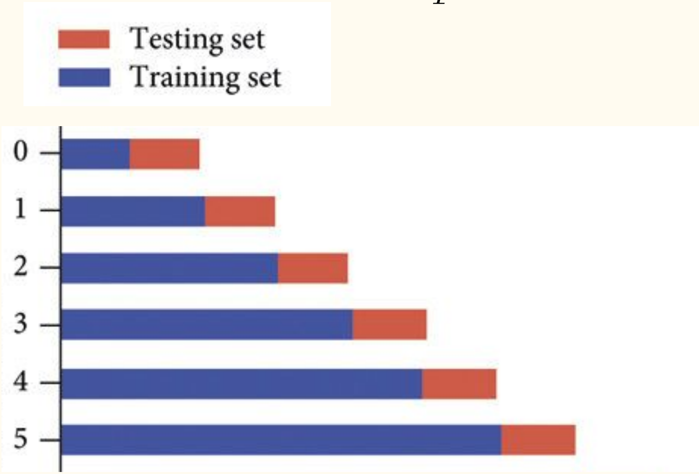
day_of_week	hour	traffic_score
6	5	0.036008
1	22	0.020970
3	23	0.015931
5	1	0.014406
3	12	0.012402

Observation: Scores seem intuitive

- Road is less congested during off-peak hours like:
  - Saturday midnight (5am)
  - Monday night (10pm)
  - etc

Approach: Modelling

<u>Data Sampling:</u>
Time-series Split

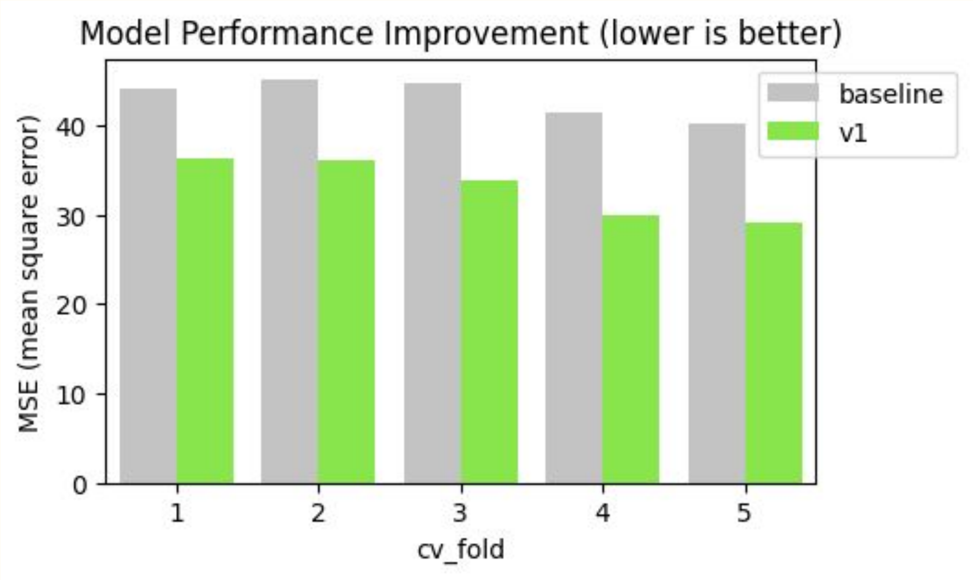


Reason: To ensure no data leakage in train-test via the score tables

<u>Regression Model</u>:

	Baseline	Version 1
Regression Model	XGBoost	
Loss Function	MSE (mean square error)	
Features	- *Trip Miles* - *day_of_week* - *hour*	- *Trip Miles* - *day_of_week* - *hour* - ***demand_score*** **- *supply_score*** **- *traffic_score***

Results



Observation

- Version 1 model **consistently performed better** than baseline across all cross-validation folds.
- This proved that the ***supply, demand, traffic signals*** are important for the pricing.

Future Work

1. **Signals: Demand, Supply, Traffic**
    - Current limitation:
        i. Score tables are essentially **dictionary** that map **discrete states** to a score
            - This can quickly turn out to be impractical as the number of states grow (e.g: 7-precision geohashes regionally)

        ii. Limited source of data
            - e.g: getting the route from pickup to dropoff can certain give better traffic score estimates)
    - Future direction:
        i. **Approximate score functions** - so that we can allow **continuous states**
        ii. Expand **data sources** to get more relevant data regarding **demand, supply and traffic conditions**
2. **Loss Function**
    - Current limitation:
        i. Framing this as a **regression problem** limits us to be only able to **replicate competitor pricing at best.**
    - Future direction:
        i. Ideally, pricing should also take **internal business metrics** into account (e.g: to maximise GMV, profits, etc)
3. **Other Areas:**
    - *Model Interpretability* - to explain pricing to users (e.g: SHAP, LIME)
    - *Personalised Pricing* - based on users price elasticity to maximise profit