

# **Data Analysis for Marketing to Convert Casuals into Members with Python, SQL and R**



## 1. Data Cleaning Process for Creating Valuable Data

## 2. Insights through Data Analysis



# **1.Data cleaning process**

# 1.Data cleaning process

Checking for missing values, the results are as follows

```
print(data.isnull().sum())
```

- start\_station\_name 29283
- start\_station\_id 29283
- end\_station\_name 31158
- end\_station\_id 31158



# complete missing data

use the unique latitude and longitude of starting and ending points to extract unique station information

Find rows where starting point and ending point names and IDs are missing, and fill them with unique station information corresponding to the latitude/longitude.

```
분하여 유일한 역 정보를 추출  
    .groupby('start_station_name').  
    .head(1).reset_index()  
    .groupby('end_station_name').  
    .head(1).reset_index()
```

```
:  
lon_name']) or pd.isnull(row['start_station'])  
stations[(unique_start_stations['start_lat']
```

```
_station_name'] = match.iloc[0]['start_'
'_station_id'] = match.iloc[0]['start_st
```

```
:  
_name']) or pd.isnull(row['end_station_  
ons[(unique_end_stations['end_lat'] == r
```

```
station_name'] = match.iloc[0]['end_station_name']
station_id'] = match.iloc[0]['end_station_id']
```

## Data completion results

- start\_station\_name 18896
- start\_station\_id 18896
- end\_station\_name 20230
- end\_station\_id 20230

**Recovery of 10,387 datas**



## **Convert start and end time data into date/time format for trip\_duration**

- `data['started_at'] = pd.to_datetime(data['started_at'])`

## **Add Some Rows**

- `['trip_duration'] = data['ended_at'] - data['started_at']`  
Calculate trip duration based on data transformation.  
Remove trip durations that are negative.
- `['distance'] , ['Day of week'], ['hour_of_day'], ['Frequency']`, and so on

## **2. Insights through Data Analysis**

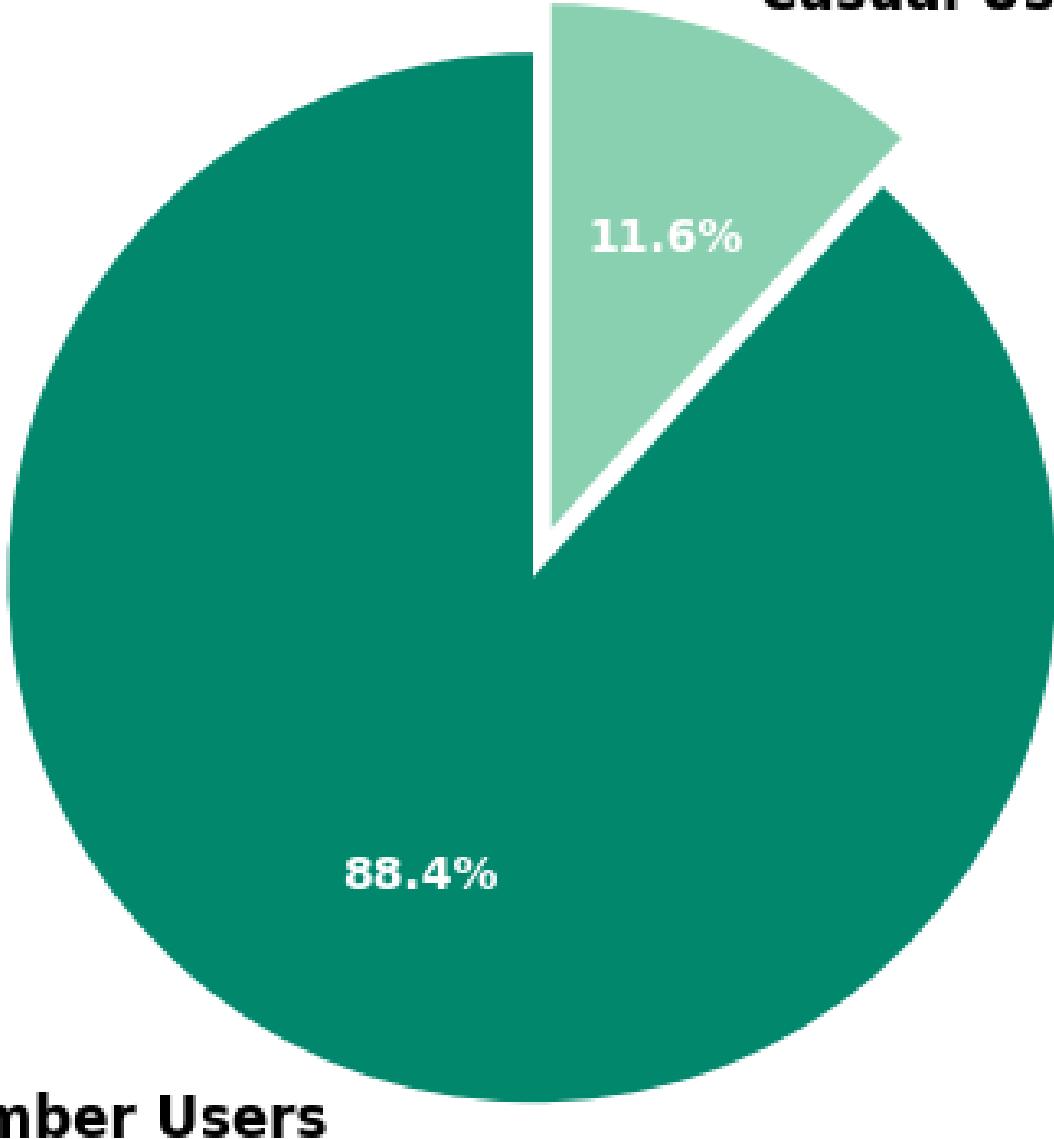
- Differences in Patterns Between Members and Casual Users**
- Marketing Strategy for Casual Riders**

## **- Differences in Patterns Between Members and Casual Users**

- Revenue share by User Type
- Day of the Week Usage Frequency
- Hourly Usage Distribution : Member vs Casual

# Revenue Share by User Type

Casual User:

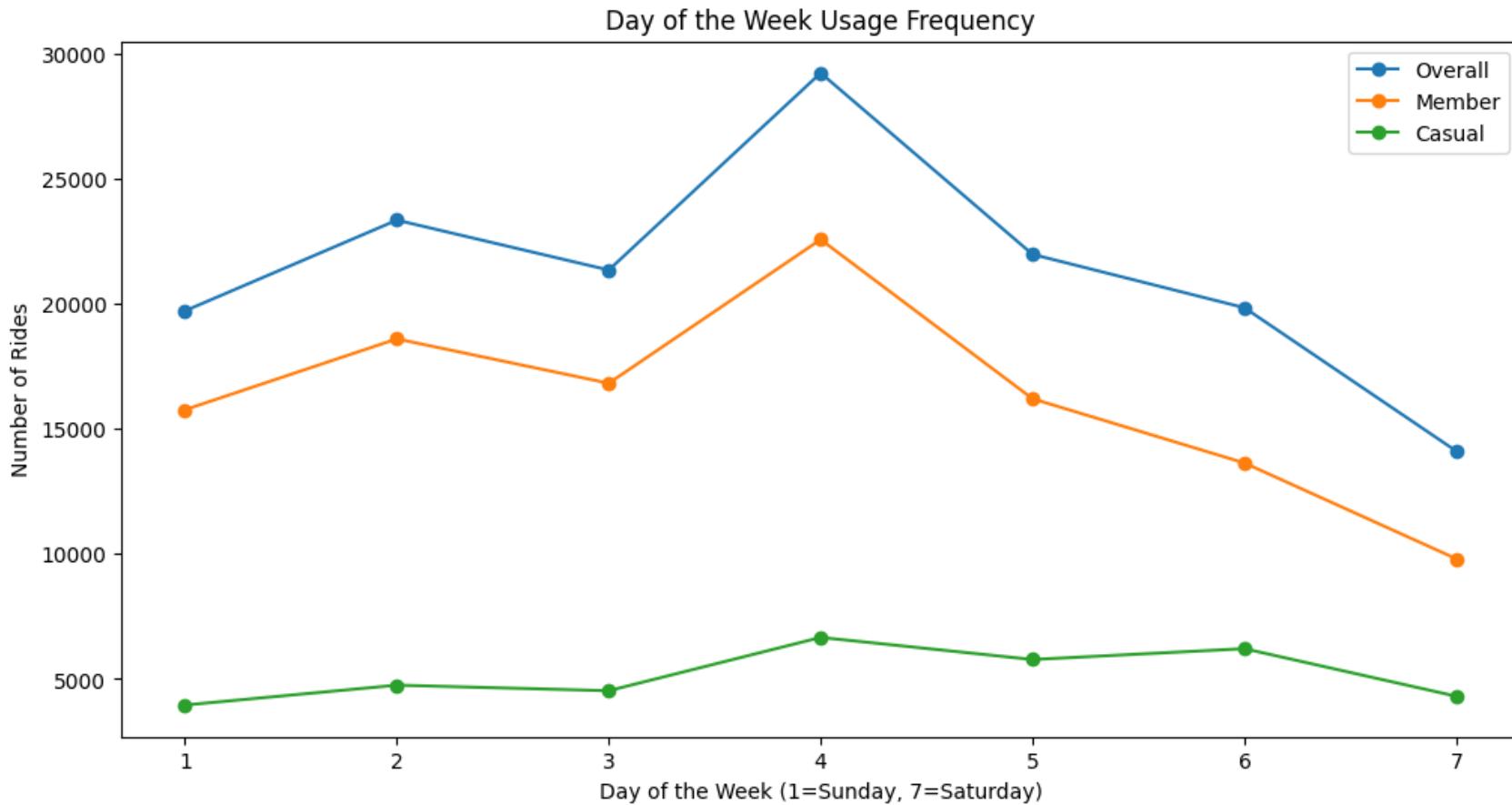


## Revenue share by User type.

Through a pie chart, we confirmed that annual members generate more revenue compared to casual users.

Based on this, we conclude that establishing a marketing strategy to convert casual users into members will contribute to an overall increase in revenue.

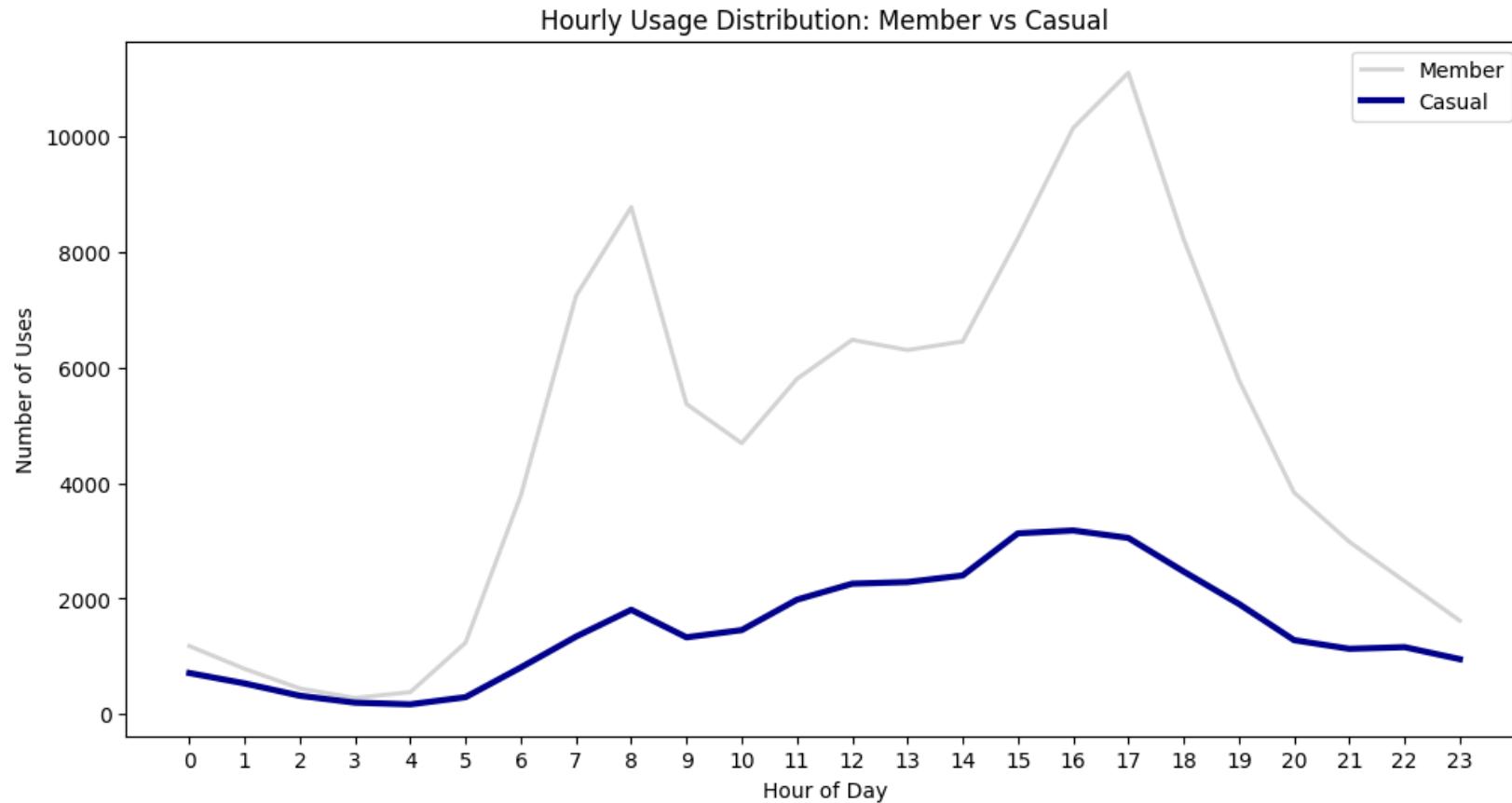




## The usage patterns of members and casual users

We found that for members, usage is highest on Mondays and Wednesdays, with a significant drop in usage on Thursdays, Fridays, and Saturdays.

However, for casual users, usage is lower on Mondays and Tuesdays but higher on Wednesdays, Thursdays, and Fridays.



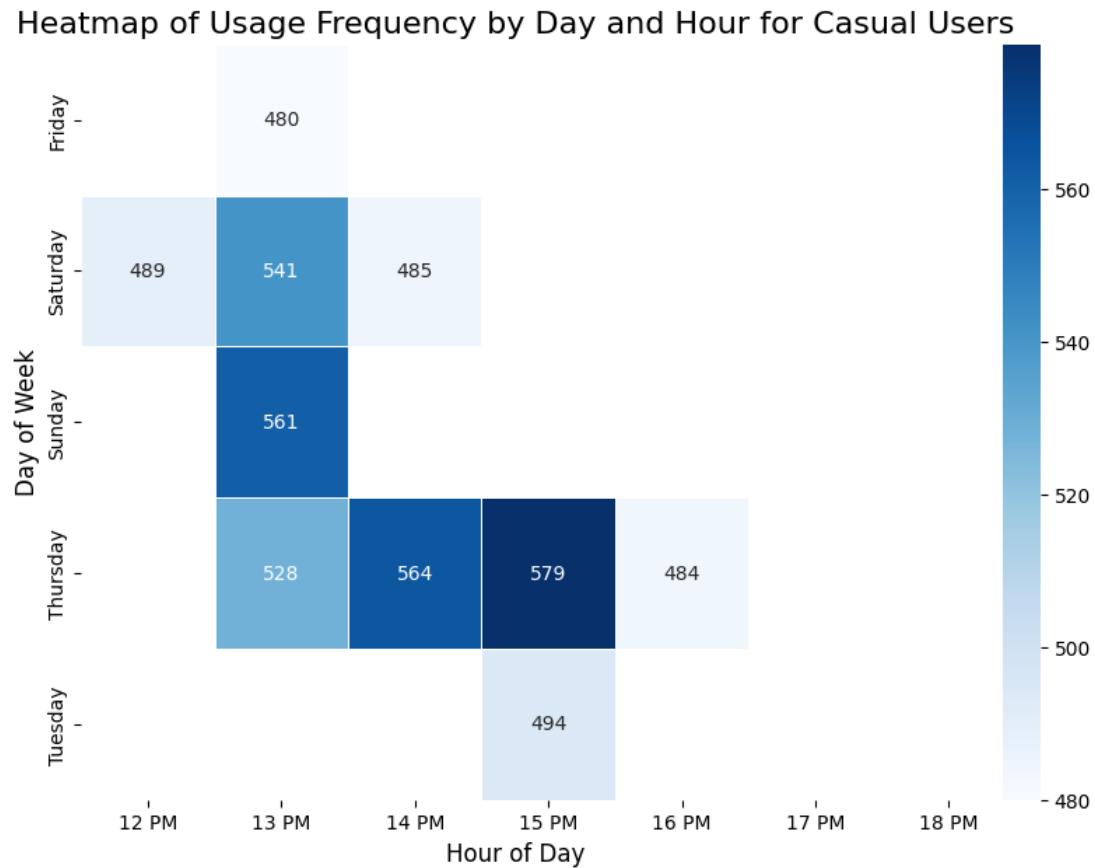
## Hourly Usage Distribution : Member vs Casual

- **Member Usage:** Peaks at 8 AM and 5-6 PM, indicating commuting patterns.
- **Casual Usage:** Increases to peak at 3 PM, steady until after 6 PM, likely leisure use.
- **Data Analysis Insight:** Different usage patterns suggest targeted marketing strategies.
- **Marketing Strategy:** Offer commuter incentives to members; leisure-time promotions to casual users to boost off-peak usage and membership conversion.

## **-Marketing Strategy for Casual Riders**

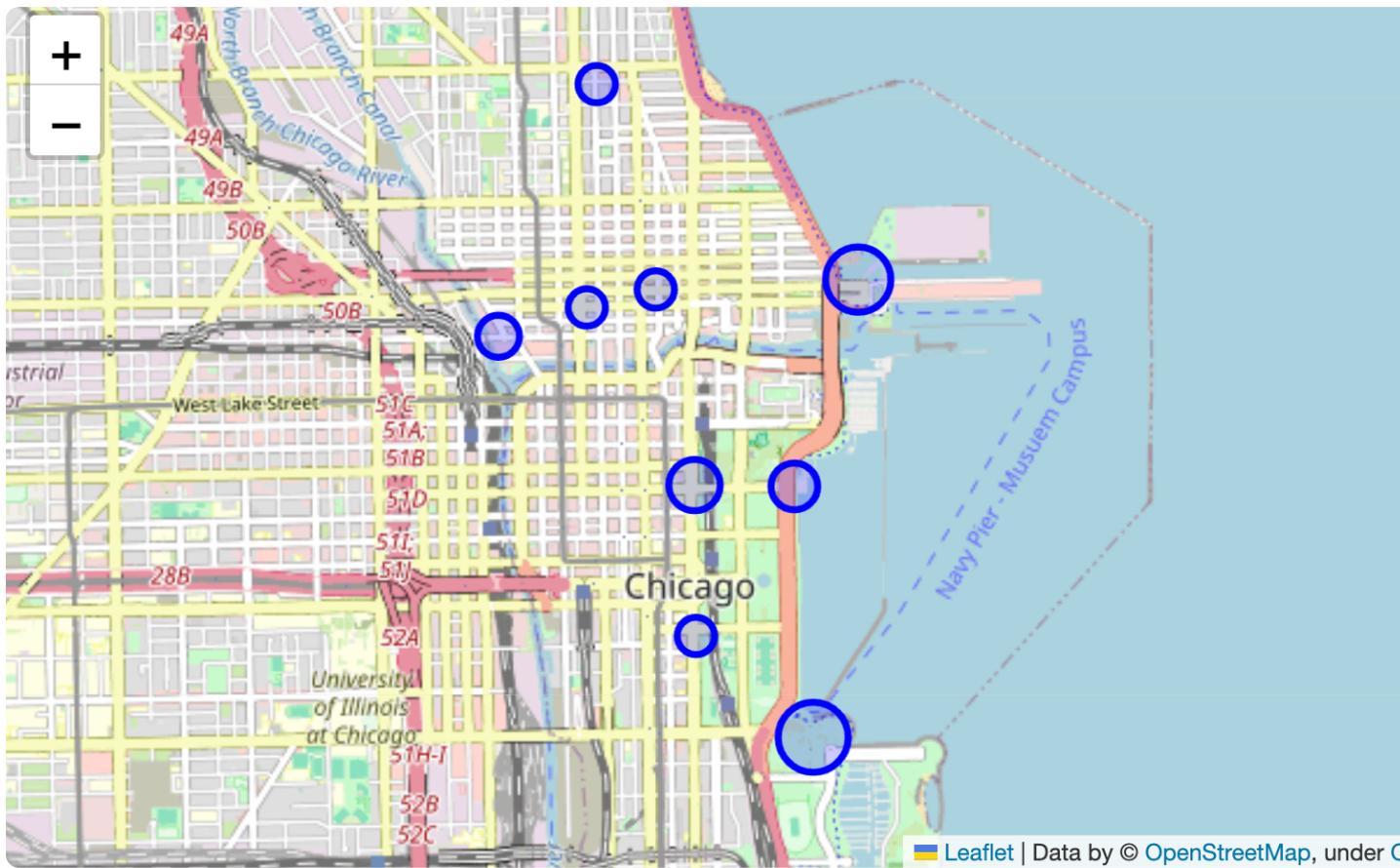
- Identify Key Usage Stations
- Determine Usage Time Slots

## This is top 10 of usage frequency by day and hour for casual users.



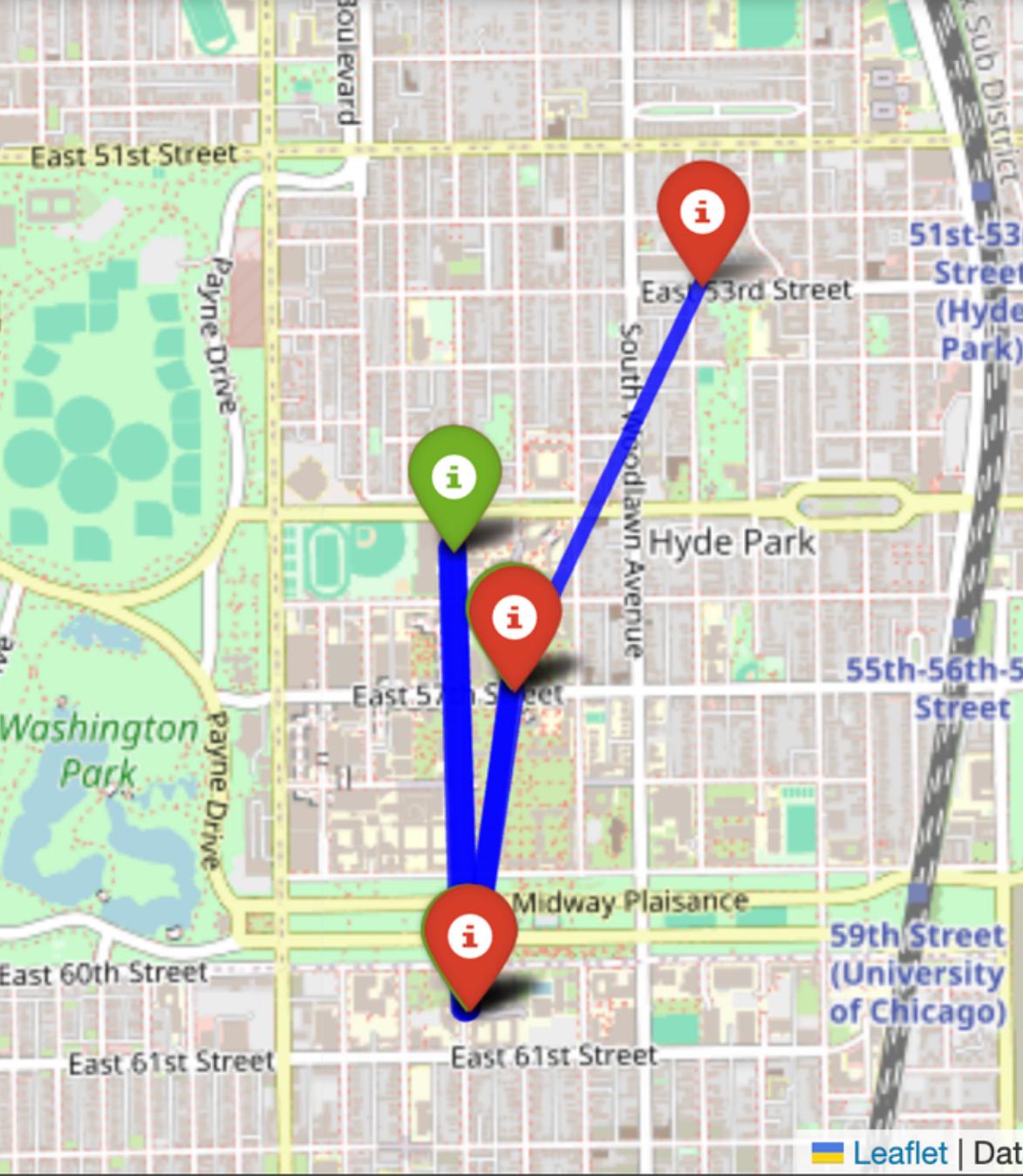
- The highest usage frequency is on Thursday at 15 PM (3 PM), with a count of 579.
- Weekdays from 14 PM to 16 PM (2 PM to 4 PM) seem to have consistently higher usage compared to other times.

# Casual Key Usage Stations



After calculating the total traffic at each station, I implemented the top 10 stations on a map with latitude and longitude information

"Utilizing these insights, we can implement GPS-based marketing at key stations and enhance our effectiveness by targeting peak hours."



# Casual top 10 routes

*Streeter Dr & Grand Ave to Streeter Dr & Grand Ave*

*Ellis Ave & 60th St to Ellis Ave & 55th St*

*DuSable Lake Shore Dr & Monroe St to DuSable Lake Shore Dr & Monroe St*

*Ellis Ave & 55th St to Ellis Ave & 60th St*

*DuSable Lake Shore Dr & Monroe St to Streeter Dr & Grand Ave*

*Ellis Ave & 60th St to University Ave & 57th St*

*University Ave & 57th St to Ellis Ave & 60th St*

*University Ave & 57th St to Kimbark Ave & 53rd St*

*Streeter Dr & Grand Ave to Millennium Park*

*Sheffield Ave & Fullerton Ave to Greenview Ave & Fullerton Ave*

## Promotions based on casual riders' key routes

1. **Destination-Based Promotions:** Collaborate with frequently visited spots like tourist attractions, shopping centers, and restaurants to offer discounts or promotions
2. **Event-Linked Marketing:** Partner with cultural events or festivals happening near the routes to provide special invites or participation opportunities.
3. **Local Business Partnerships:** Establish partnerships with nearby businesses (cafes, restaurants, shops) to offer special discounts or services to riders upon reaching their destinations, like discount coupons or promotional codes.

These strategies aim to leverage popular routes to enhance rider experiences and encourage usage of the service by adding value through promotions and partnerships.

**To increase the conversion rate of casual riders to members**, focus marketing efforts on offering benefits based on usage times, popular stations, and routes

I have presented data analysis for marketing based on data up to December 2022.

The analysis process and code can be found on my GitHub page.

For the first half of 2023, I have uploaded code and materials for data extraction using SQL and visualization in R.

GitHub URL: <https://github.com/Jun6481/portfolio>