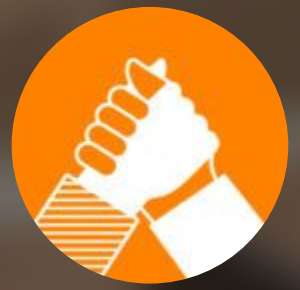
Analyzing the Los Angeles Shared Bike Data

Weilin Ouyang, Mengjun Wang, Tianhao Wu, Haoyu Li

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE



Content

- Motivation
- Objective
- Data
- Methodology
- Results and Conclusion
- Limitations & Future Vision
- Work Distribution

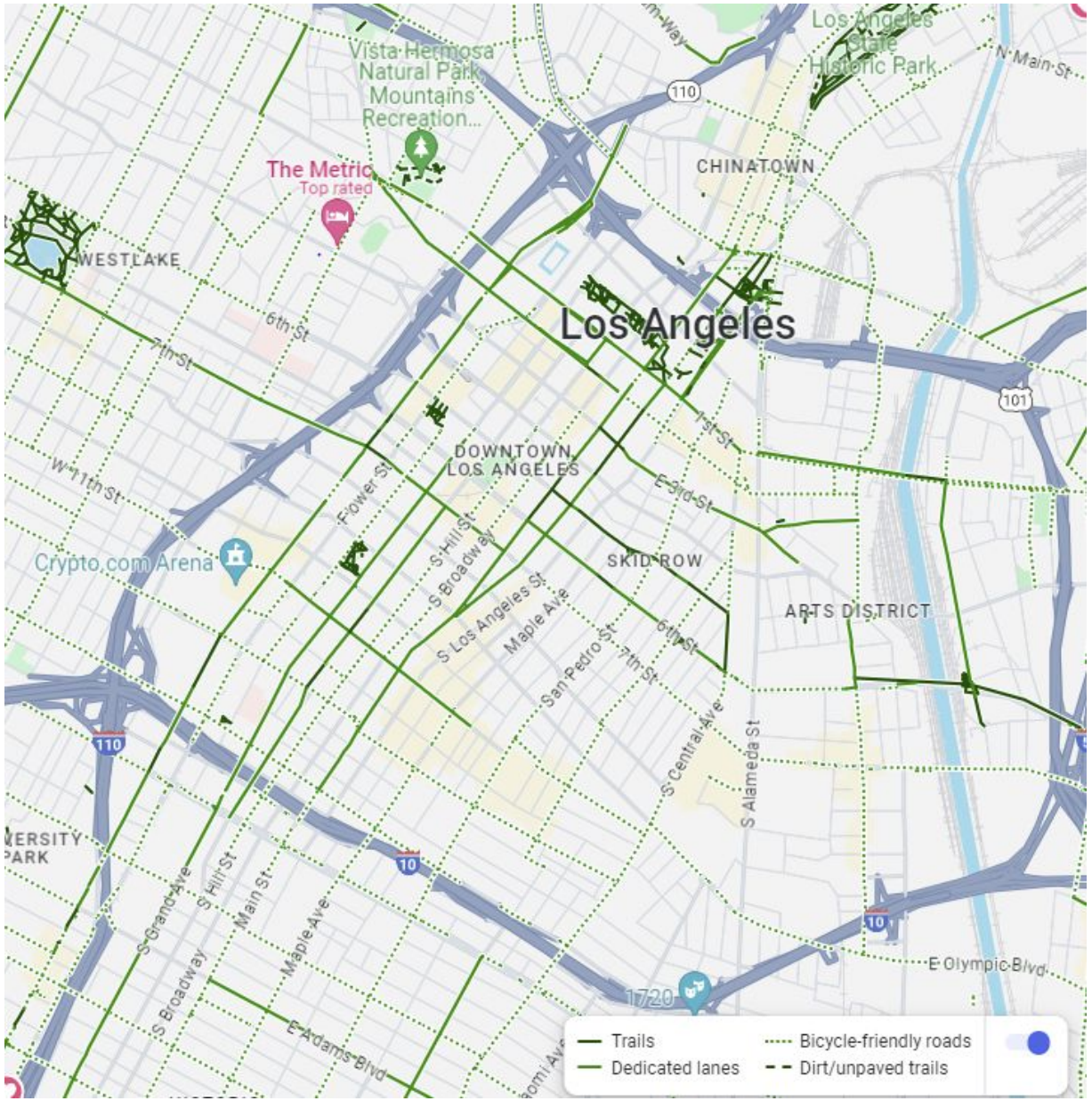
Motivation

Los Angeles has the 6th worst traffic in the US

- The traffic congestion, air pollution issue
- Broad the cover range of shared bike service
- Expand the scale to Knoxville

Objective

- Understanding usage patterns
- Improving bike-sharing services
- Optimizing operations and improving user satisfaction



Data Collection

Data Source: Dataset from Metro Bike Share, Los Angeles area.

System Capacity: Approximately 1,400 bicycles. 93 stations across Downtown L.A., Venice, and the Port of L.A.

Data Timeline: Data spans from July 2017 to September 2023.

Data Organization: Quarterly data, each in a separate .csv file.

Data Details: Includes trip ID, duration, start/end times and stations, geolocation of stations, details on bike ID, passholder type, and trip route category.

https://bikeshare.metro.net/about/data/

Data Processing

1. Reclassify the pass types, ’Flex Pass’ —-> ’Annual Pass’
2. Exclude the outlier trips — too long, short, special user ID
3. Remove stations with special ID
4. Reclassify the bike types
5. Standardize the time units
6. Quarterly datasets combination

	Starting Station ID	Starting Station Latitude	Starting Station Longitude	Ending Station ID	Ending Station Latitude	Ending Station Longitude
0	3014.0	34.056610	-118.23721	3014.0	34.056610	-118.23721
1	3014.0	34.056610	-118.23721	3014.0	34.056610	-118.23721
2	3016.0	34.052898	-118.24156	3016.0	34.052898	-118.24156
3	3016.0	34.052898	-118.24156	3016.0	34.052898	-118.24156
4	3032.0	34.049889	-118.25588	3032.0	34.049889	-118.25588

Methodology

Distribution of Hot Stations Analysis

- Analyzed LA Metro Bike Share data for station usage patterns.
- Used a scatter plot to visualize geographical distribution of stations.
- Focused on key areas in LA, identifying popular stations.

Rush Hour Analysis

- Cleaned data and extracted time information from bike trip start times.
- Utilized line charts to display rush hour trends.
- Visualized peak hours for bike usage.

Methodology

Round Trip/One Way Comparison Analysis

- Segregated data into 'Round Trip' and 'One Way' categories.
- Compared departure times and passholder types using line and pie charts.
- Analyzed trip durations for both categories.

Relevancy Analysis between Duration & Passholder Type

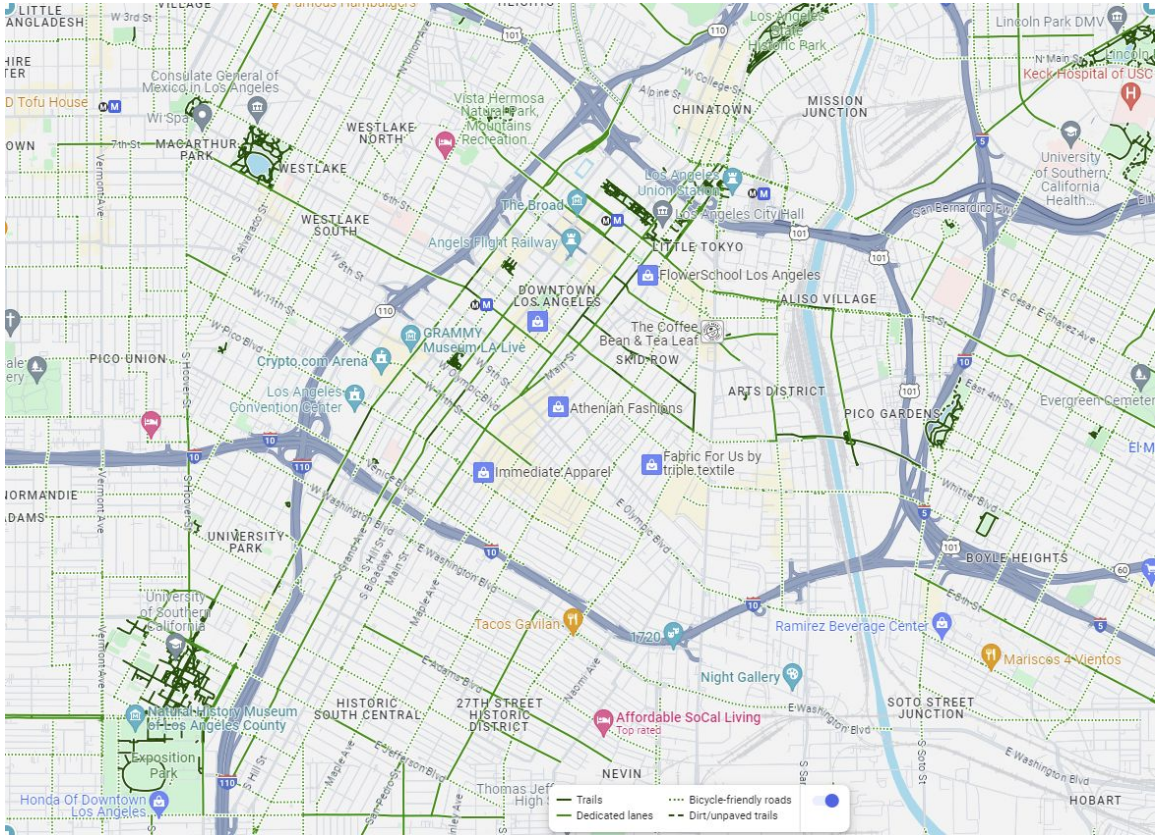
- Employed K-Means clustering to categorize trip durations.
- Analyzed the distribution of passholder types across duration categories.
- Visualized findings with a radar chart.

```
plt.axis([-118.30, -118.20, 34.01, 34.07
])
```

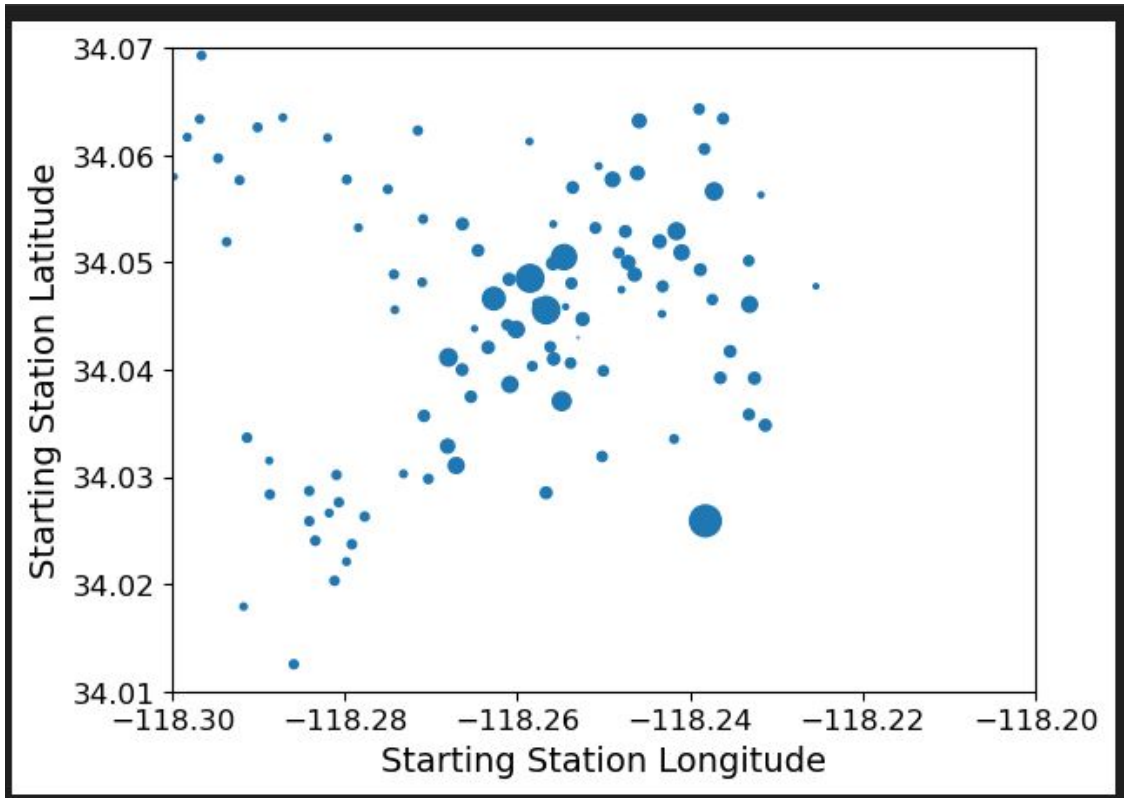
Results & Conclusion

1. From: (34°0'36"N -118°18'0"E) To: (34°4'12"N -118°18'0"E)
   Distance is: 4.15 miles / 6.67 kilometers / 3.60 nautical miles
2. From: (34°0'36"N -118°18'0"E) To: (34°0'36"N -118°12'0"E)
   Distance is: 5.73 miles / 9.22 kilometers / 4.97 nautical miles

The distribution of hot stations analysis



Actual map



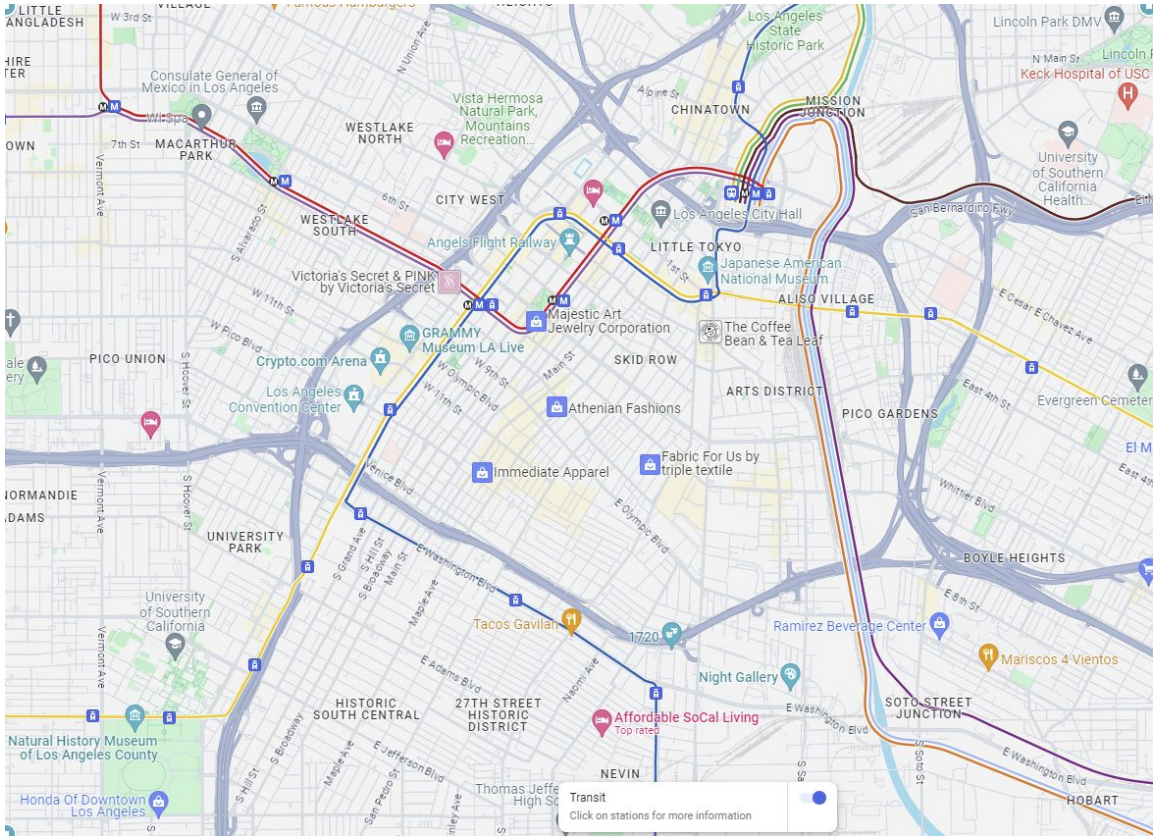
Starting Station map

```
plt.axis([-118.30, -118.20, 34.01, 34.07
])
```

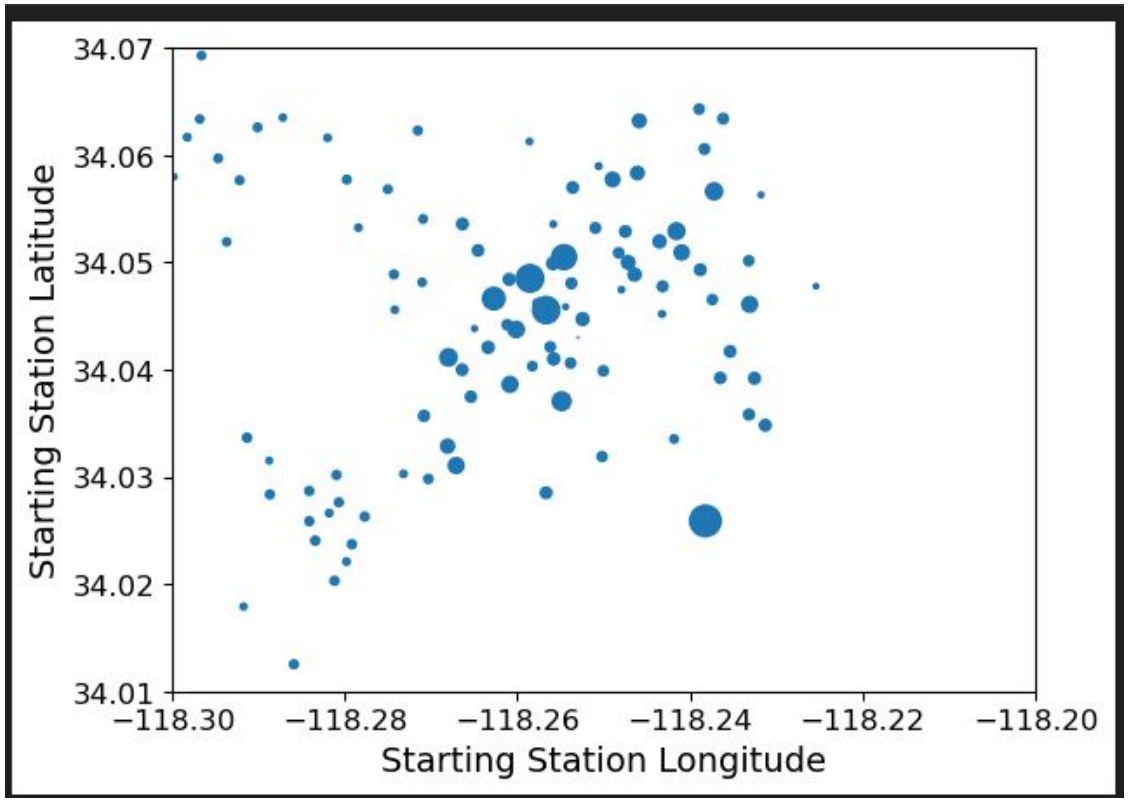
Results & Conclusion

1. From: (34°0'36"N -118°18'0"E) To: (34°4'12"N -118°18'0"E)
   Distance is: 4.15 miles / 6.67 kilometers / 3.60 nautical miles
2. From: (34°0'36"N -118°18'0"E) To: (34°0'36"N -118°12'0"E)
   Distance is: 5.73 miles / 9.22 kilometers / 4.97 nautical miles

The distribution of hot stations analysis



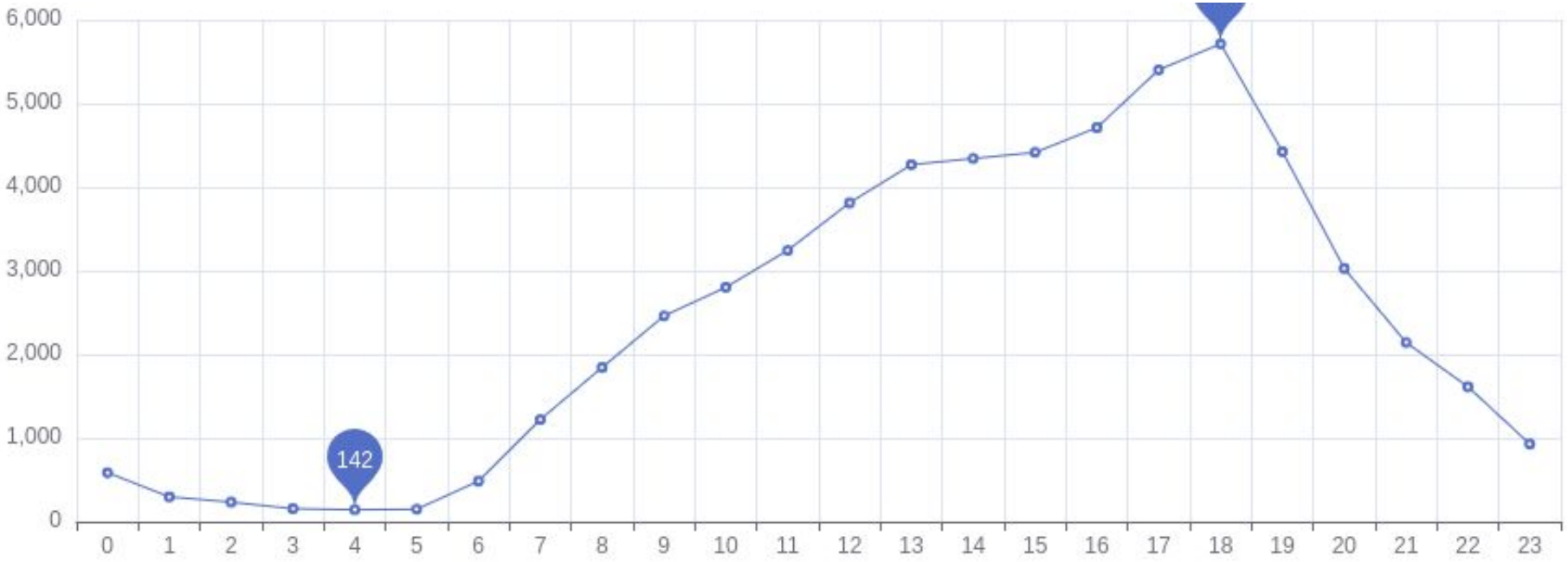
Actual map



Starting Station map

Results & Conclusion

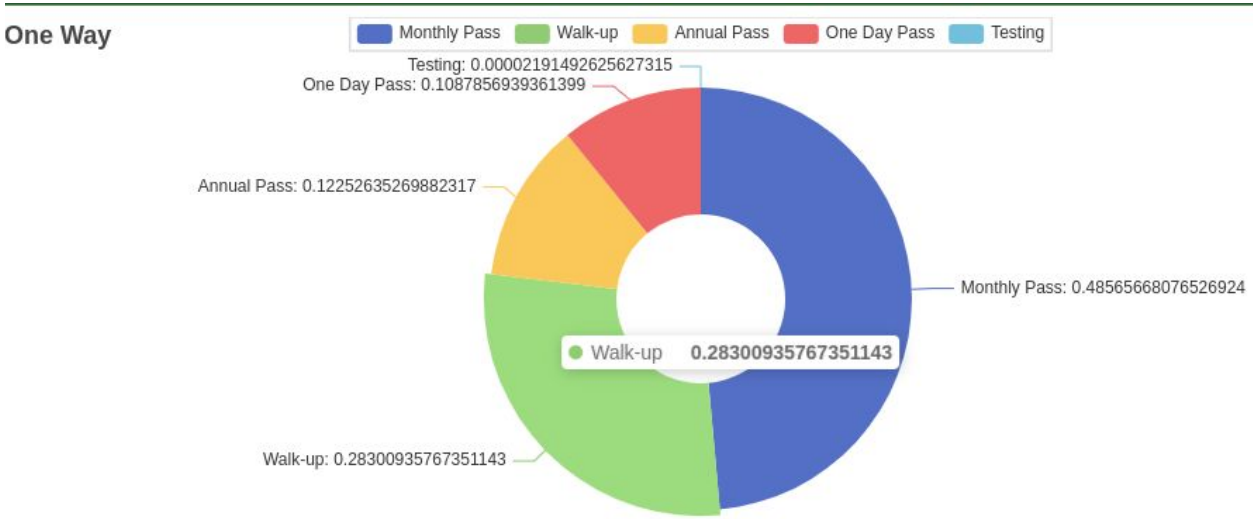
The rush hour analysis

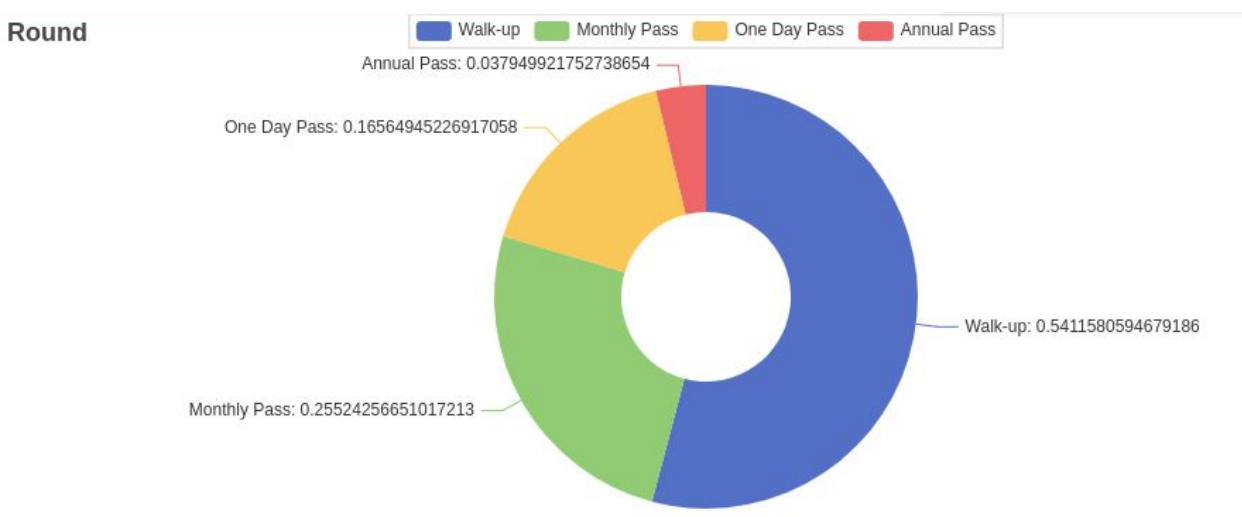


24 hours shared-bike use

Results & Conclusion

The round trip/one way comparison analysis

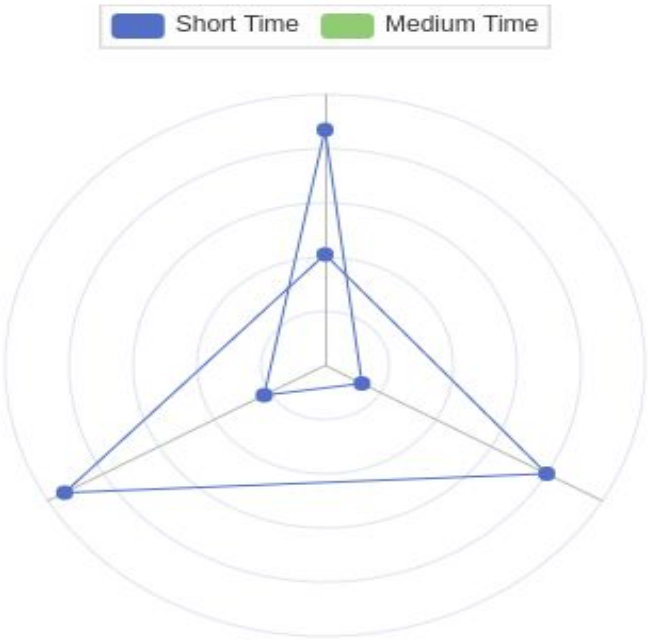




The proportion of users with membership cards

Results & Conclusion

The relevancy analysis between duration & passholder type



Results & Conclusion

Analysis of the relationship between the usage duration of One Way Bicycle and Round Trip Bicycle

	One Way	Round
count	45631.000000	12780.000000
mean	39.475839	68.067136
std	136.662651	129.701820
min	1.000000	1.000000
25%	8.000000	18.000000
50%	15.000000	35.000000
75%	27.000000	75.000000
max	1440.000000	1440.000000

Limitation

1. Managing and filtering the **vast dataset** to eliminate irrelevant or misleading information proved more challenging than anticipated.
2. As traffic analysis can be influenced by environmental factors, this study focused solely on analyzing the existing data and its characteristics. Consequently, it may **lack robustness in adapting to changes.**

Future Work

1. This study concentrates on analyzing data to identify key characteristics that could inform recommendations for **bike traffic distribution in Los Angeles**. However, t**he current approach has limitations** in adapting to environmental changes and does not account for unforeseen factors. In future research, we aim to develop a more resilient analytical method. This may involve **incorporating machine learning prediction models** to yield more reliable and insightful results.

Workflow Chart

Org Chart

01

Mengjun Wang

Gathering, preparing, do the data preprocessing and written reports.

02

Weilin Ouyang

Statistical tests, modeling, meaningful conclusion from data and written reports.

03

Tianhao Wu

Document project progress, methodologies and written reports.

04

Haoyu Li

Data analysis, result visualizations and written reports..