

Bike Sharing Analysis - Summary Statistics

DATASET OVERVIEW

=====

Total Trips: 17,527,883
Date Range: 2018-01-01 → 2018-12-31
Total Users: 212 unique user groups

PATTERN MINING RESULTS

=====

Station Pairs: 347,178
Bidirectional Pairs: 200,466
Zone Connections: 268
Association Rules: 258
Frequent Itemsets: 278

TOP 10 ROUTES

=====

1. 2006 → 2006: 7,849 trips (0.04%)
2. 432 → 3263: 7,249 trips (0.04%)
3. 281 → 281: 6,340 trips (0.04%)
4. 2006 → 3282: 6,214 trips (0.04%)
5. 460 → 3093: 5,455 trips (0.03%)
6. 3093 → 460: 5,131 trips (0.03%)
7. 435 → 509: 4,995 trips (0.03%)
8. 519 → 492: 4,662 trips (0.03%)
9. 519 → 498: 4,465 trips (0.03%)
10. 387 → 387: 4,463 trips (0.03%)

COMMUTE PATTERNS

=====

Morning Rush (7-9 AM): 3,330,180 (19.0%)
Evening Rush (5-7 PM): 4,458,497 (25.4%)
Non-Commute Hours: 9,739,206 (55.6%)

WEATHER IMPACT

=====

Good Weather Trips: 6,389,700 (36.5%)
Poor Weather Trips: 11,138,183 (63.5%)

CLUSTERING RESULTS

=====

Optimal Number of Clusters: 4
Clustering Method: average
Davies-Bouldin Score: 0.176

User Clustering - Detailed Summary

CLUSTER CHARACTERISTICS

CLUSTER 1: 209 users (98.6%)

Age: 51.0 years
Trips/month: 25754.3
Trip duration: 15.1 min
Weekend usage: 23.2%
Station diversity: 339.9
Weather score: 56.9/100

CLUSTER 2: 1 users (0.5%)

Age: 49.0 years
Trips/month: 459984.0
Trip duration: 26.1 min
Weekend usage: 41.4%
Station diversity: 764.5
Weather score: 58.3/100

CLUSTER 3: 1 users (0.5%)

Age: 88.0 years
Trips/month: 2.0
Trip duration: 13.2 min
Weekend usage: 83.3%
Station diversity: 3.0
Weather score: 34.8/100

CLUSTER 4: 1 users (0.5%)

Age: 73.0 years
Trips/month: 1.3
Trip duration: 60.5 min
Weekend usage: 100.0%
Station diversity: 3.0
Weather score: 57.5/100

CLUSTER SUMMARY TABLE

	age	trip_count	avg_trip_min	weekend_ratio	station_diversity	avg_cycling_score
cluster						
1	50.95	77262.77	15.13	0.23	339.95	56.910000
2	49.00	1379952.00	26.14	0.41	764.50	58.310001
3	88.00	6.00	13.16	0.83	3.00	34.820000
4	73.00	4.00	60.51	1.00	3.00	57.540001

Weather Correlation Analysis

PEARSON CORRELATION: WEATHER vs TRIP DEMAND

=====

Variable	Correlation	p-value	Significance

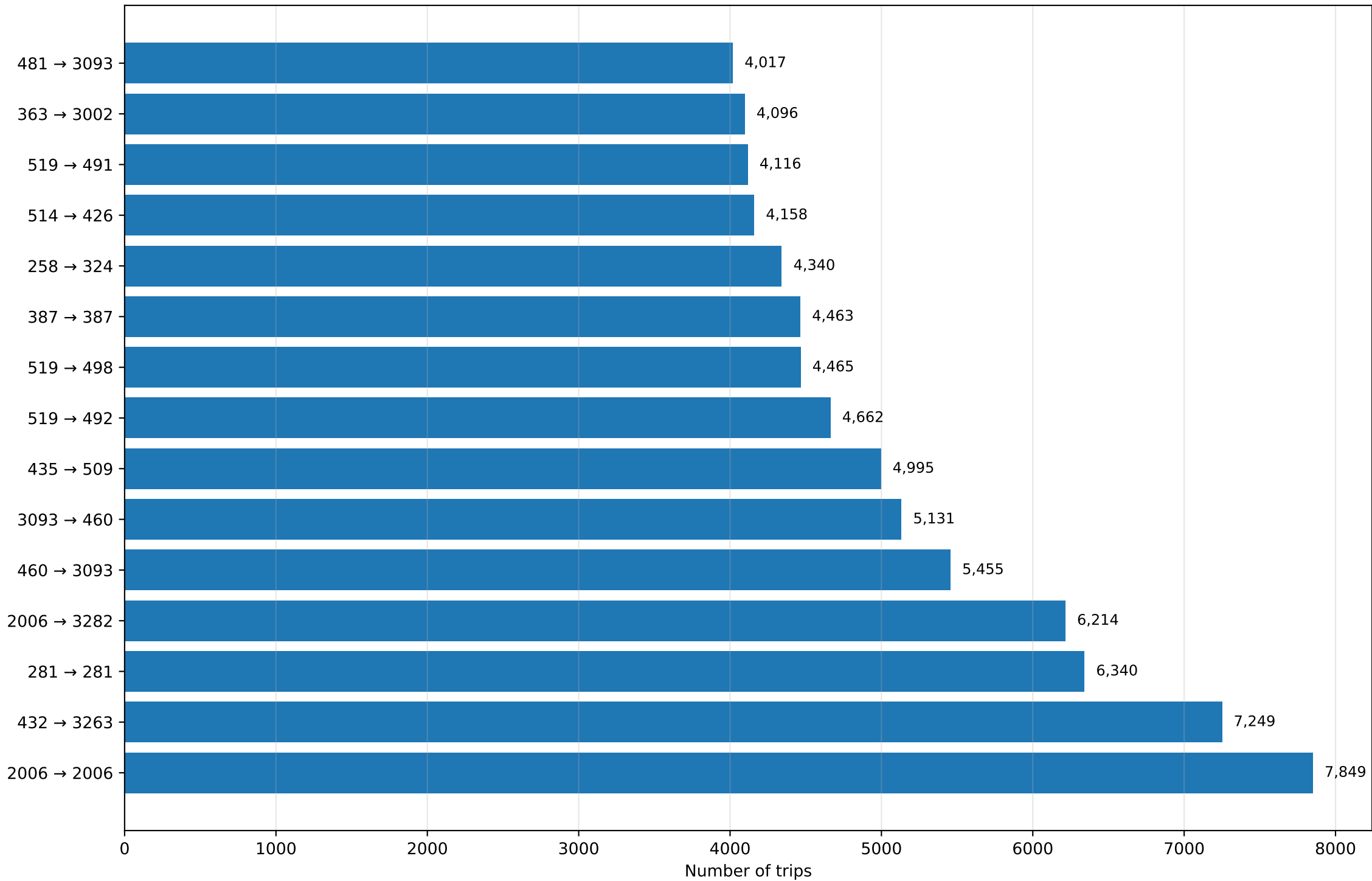
temp_celsius	0.4425	0.00e+00	***
apparent_temperature	0.4099	0.00e+00	***
cycling_score	0.3415	9.20e-238	***
is_dry	0.1279	3.36e-33	***
wind_kmh	-0.0141	1.88e-01	n.s.
cloud_cover	-0.0310	3.80e-03	**
snowfall	-0.0850	1.69e-15	***
rain	-0.1280	2.88e-33	***
relative_humidity_2m	-0.2730	3.20e-149	***

KEY INSIGHTS

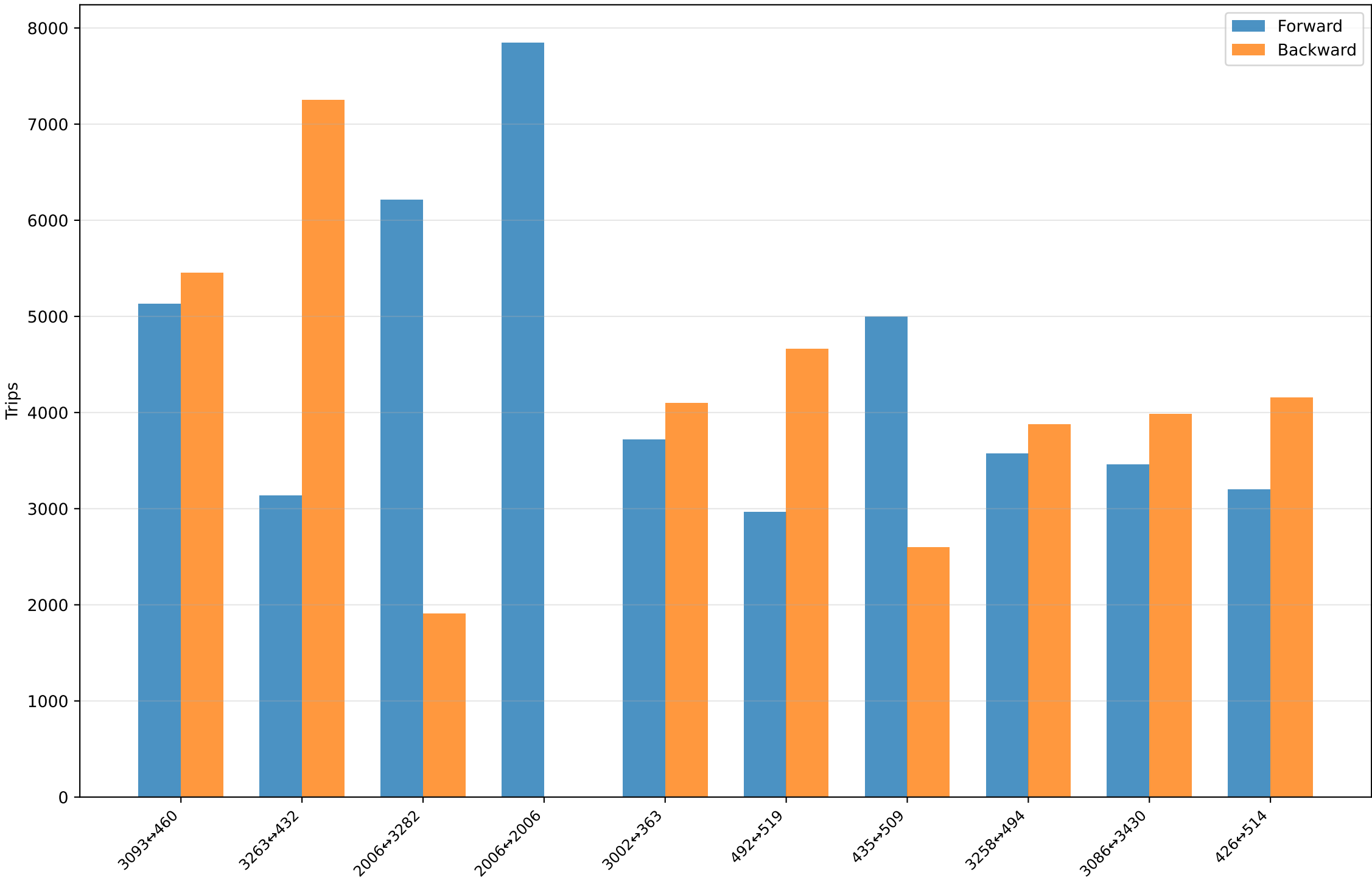
=====

- ✓ Strongest positive: temp_celsius (r=0.4425)
- ✓ 8 variables statistically significant (p < 0.05)
- ✓ 3 variables show moderate-to-strong correlation ($|r| > 0.3$)
- ✓ Cycling score explains 11.7% of variance
- ✓ 7 variables highly significant (p < 0.001)

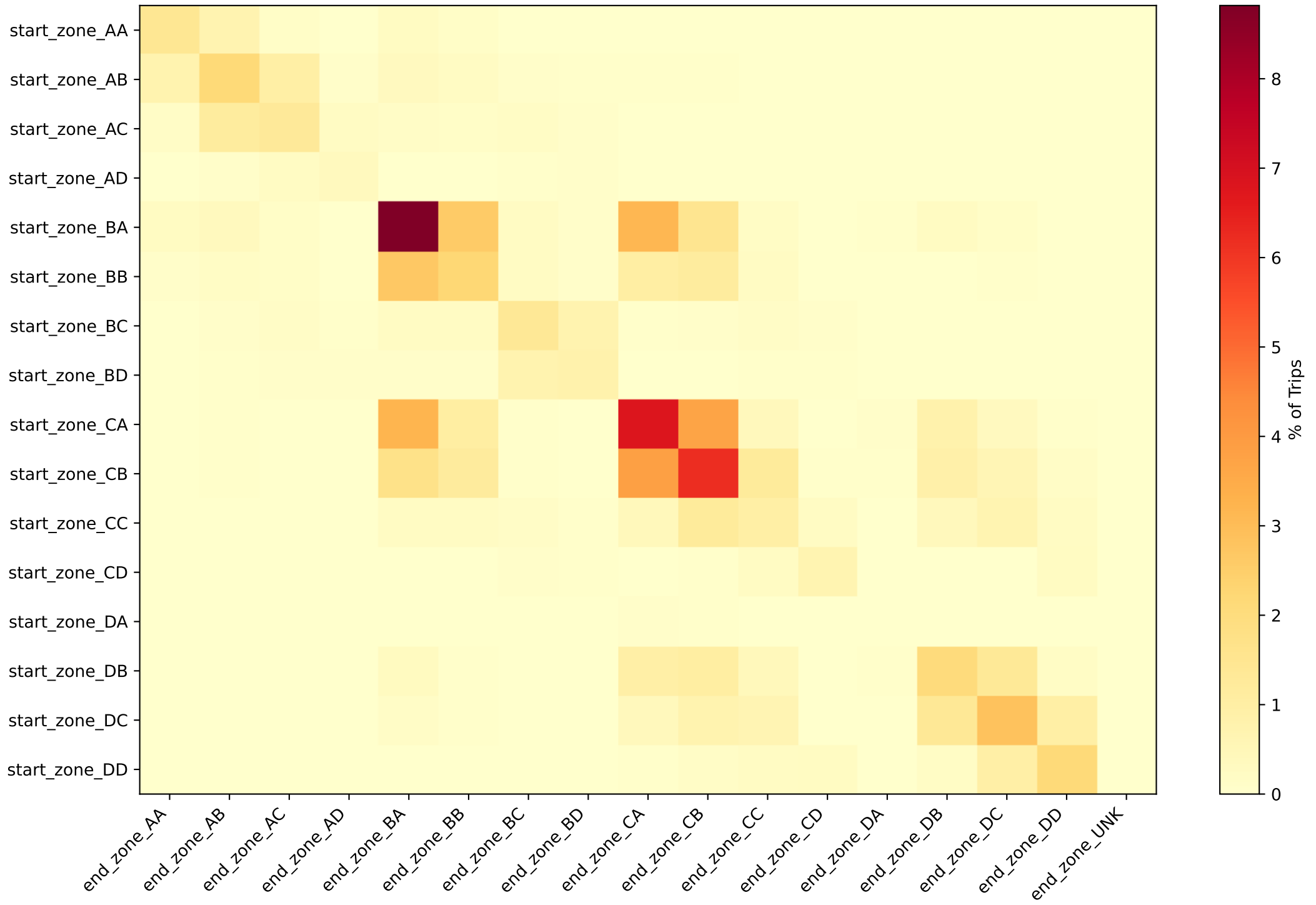
Top Station-to-Station Routes



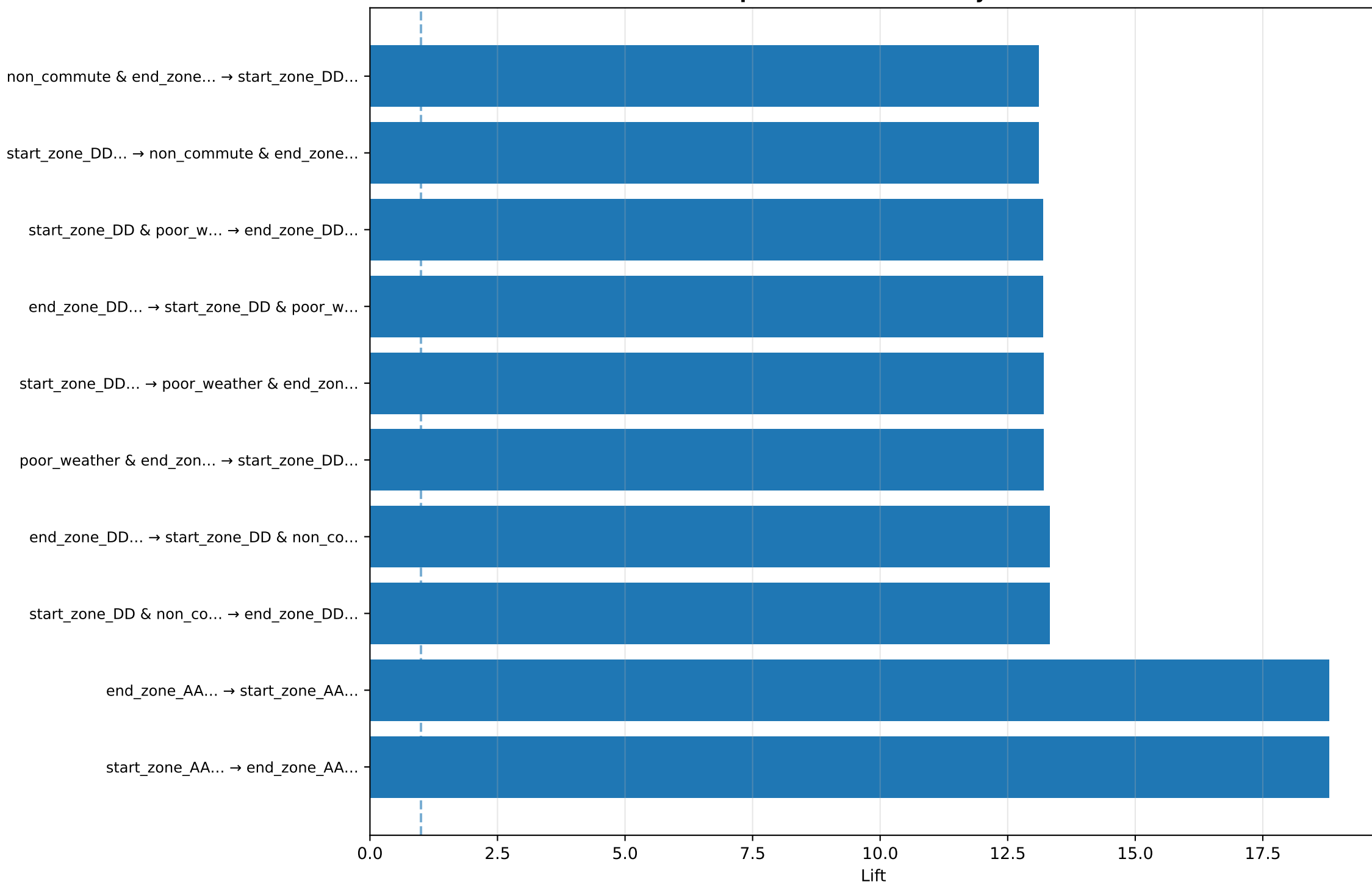
Bidirectional Route Analysis - Rebalancing Needs



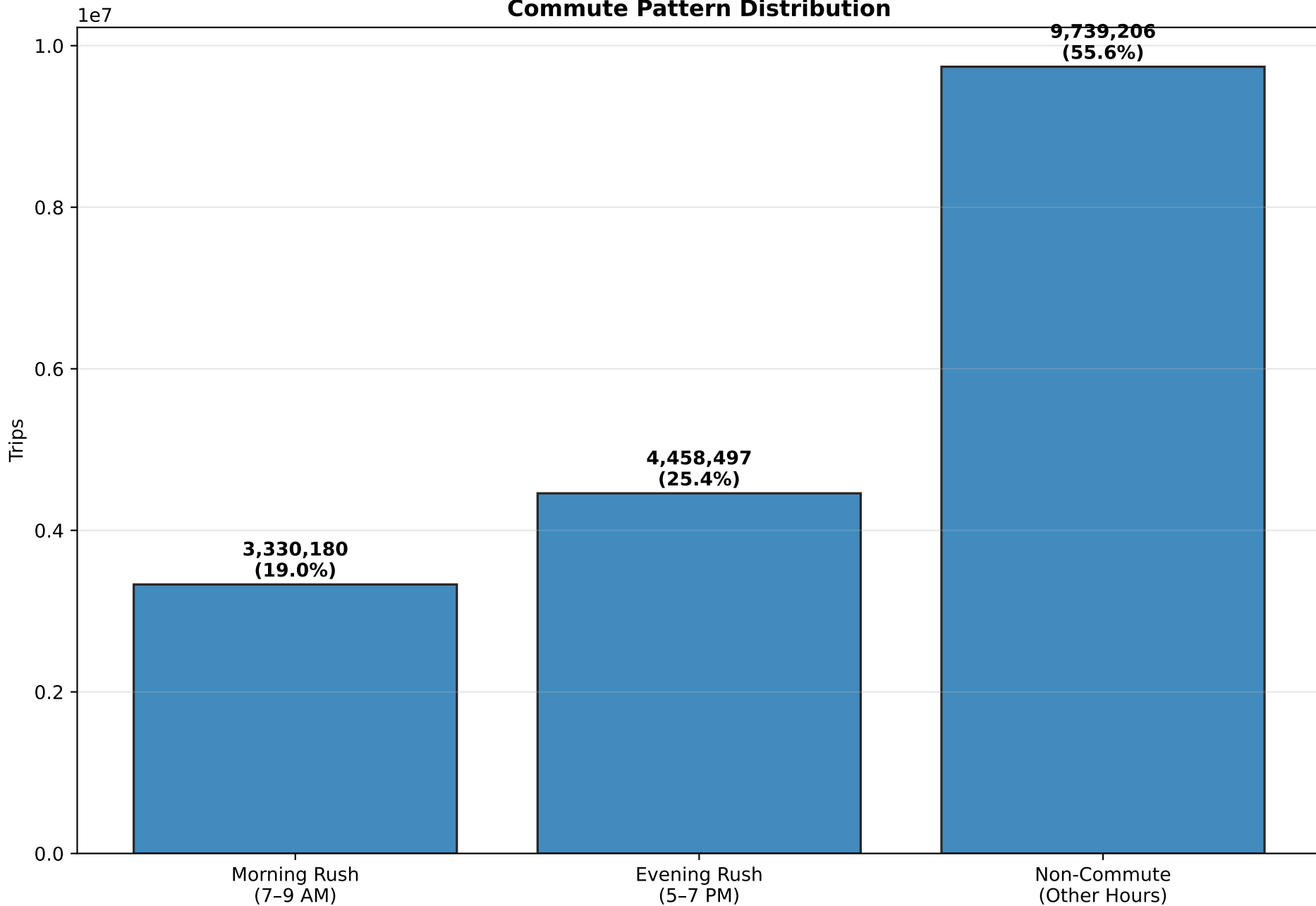
Geo Zone-to-Zone Movement Heatmap



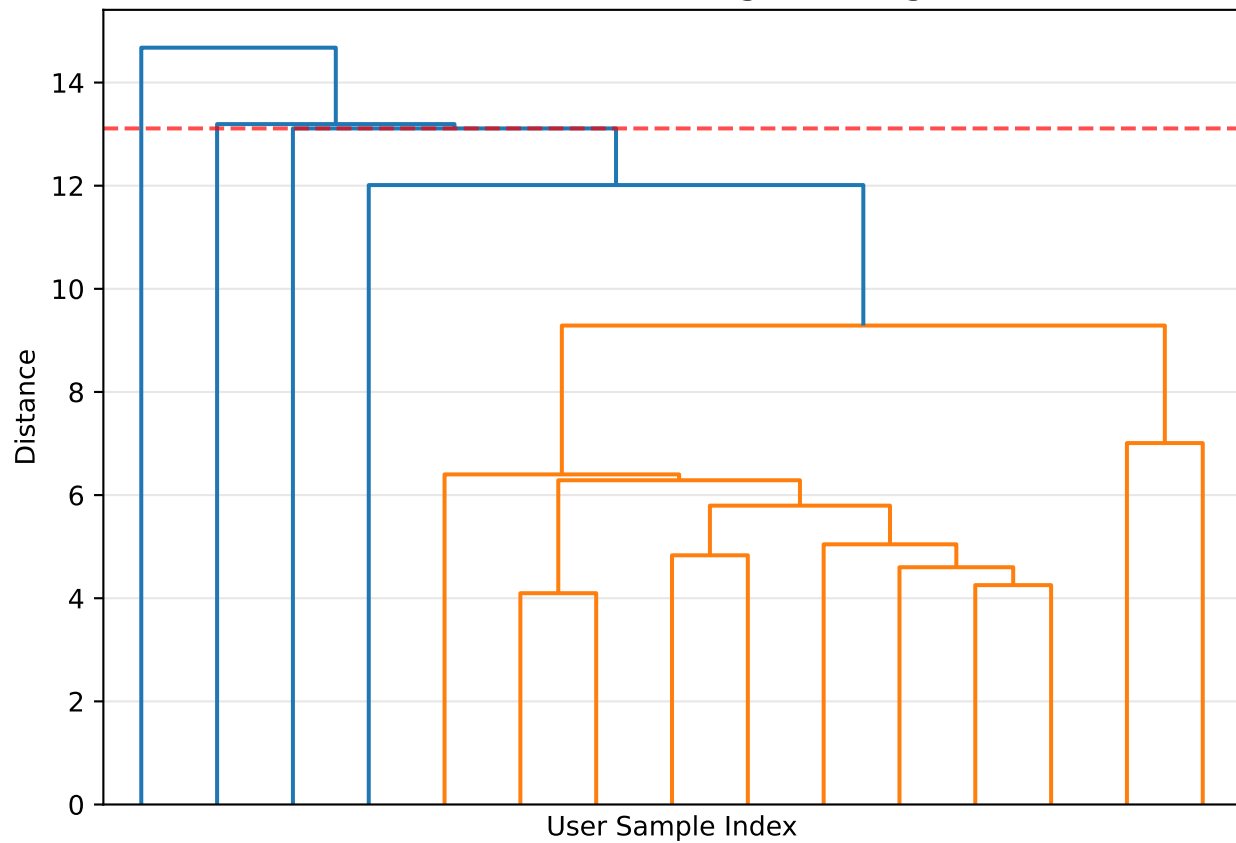
Top Association Rules by Lift



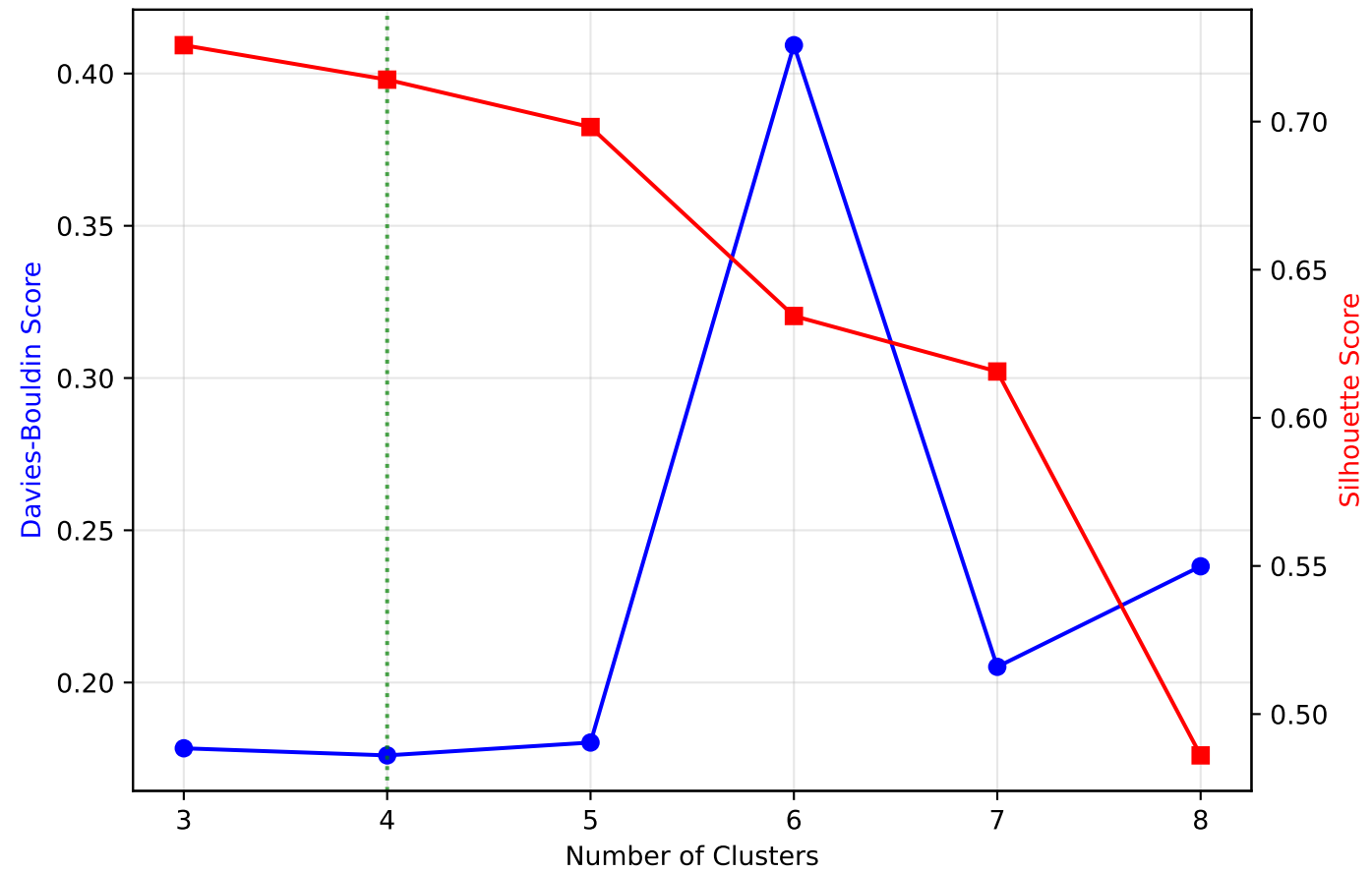
Commute Pattern Distribution



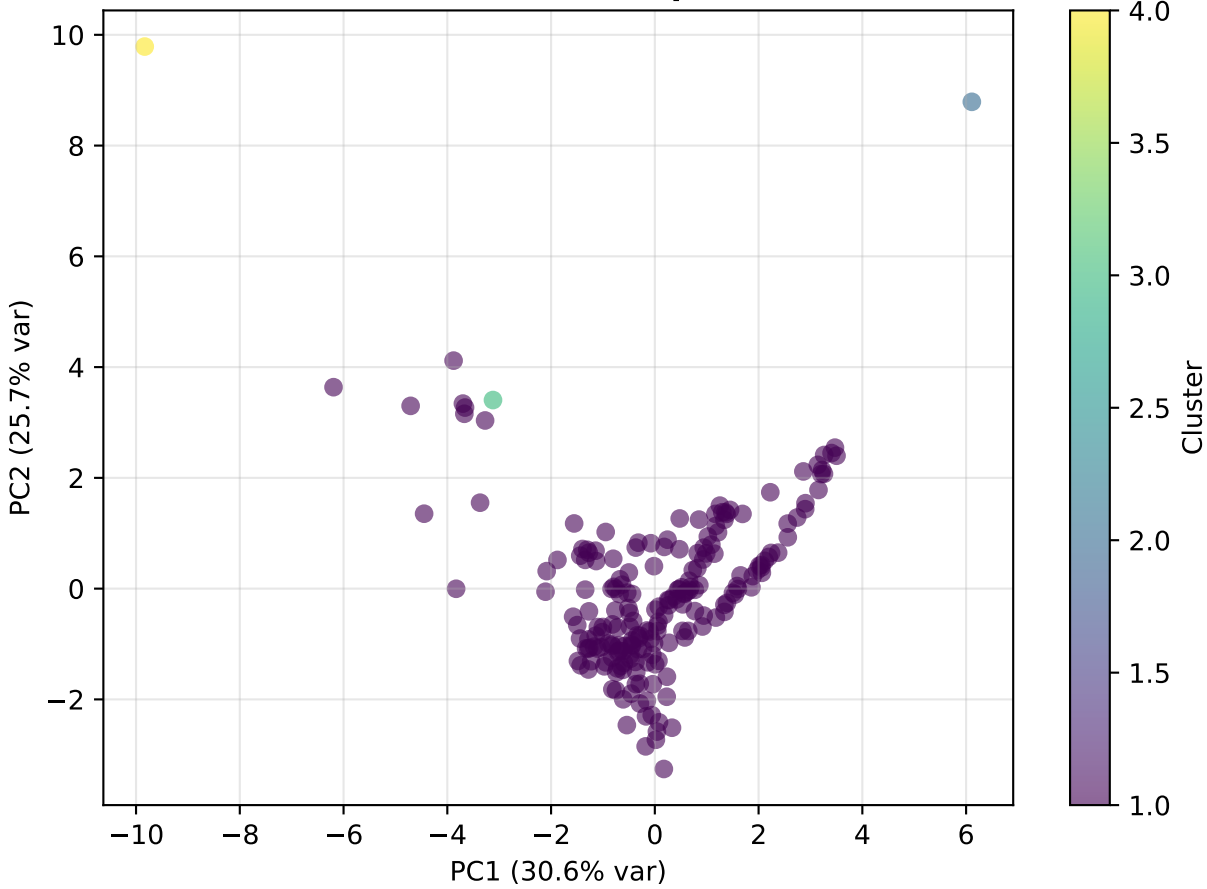
Hierarchical Clustering Dendrogram



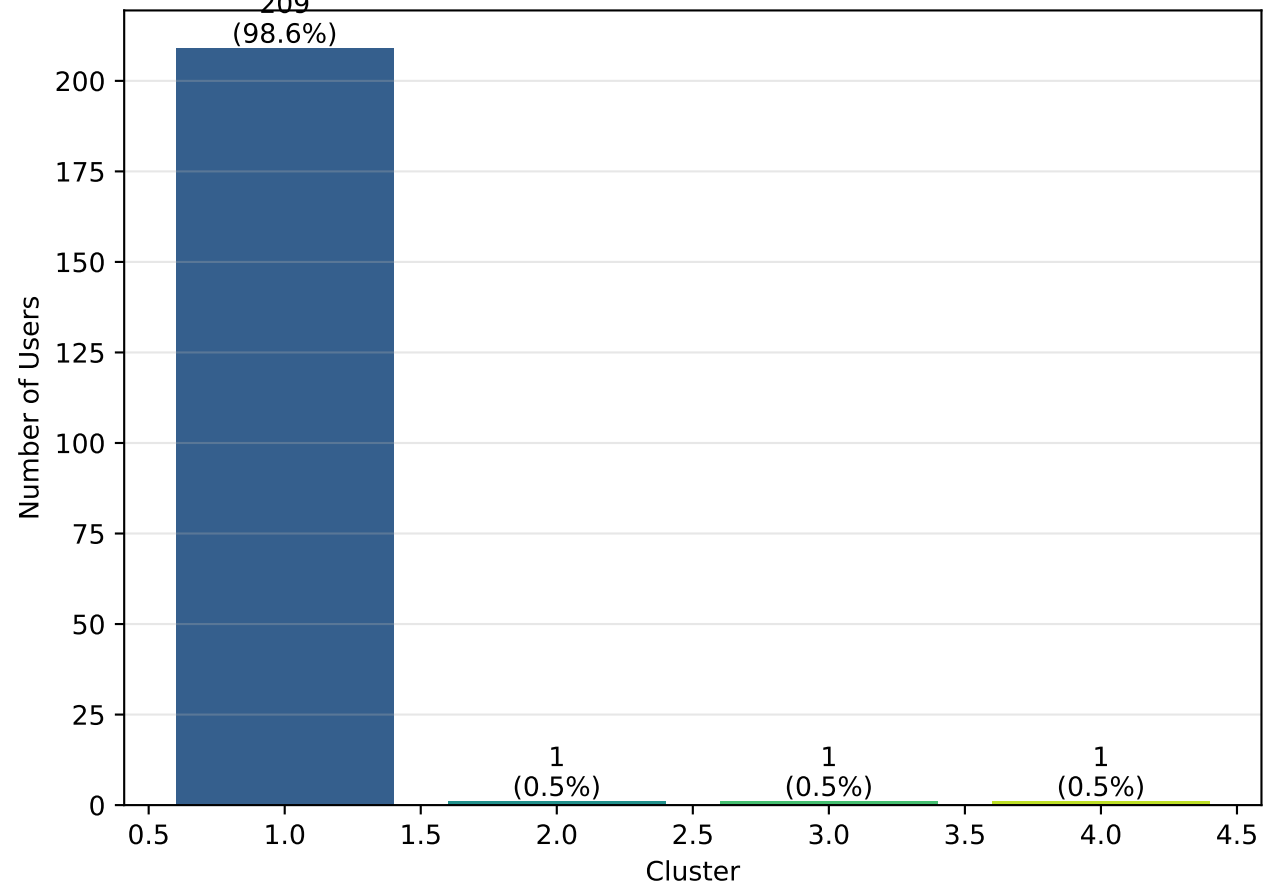
Cluster Evaluation Metrics

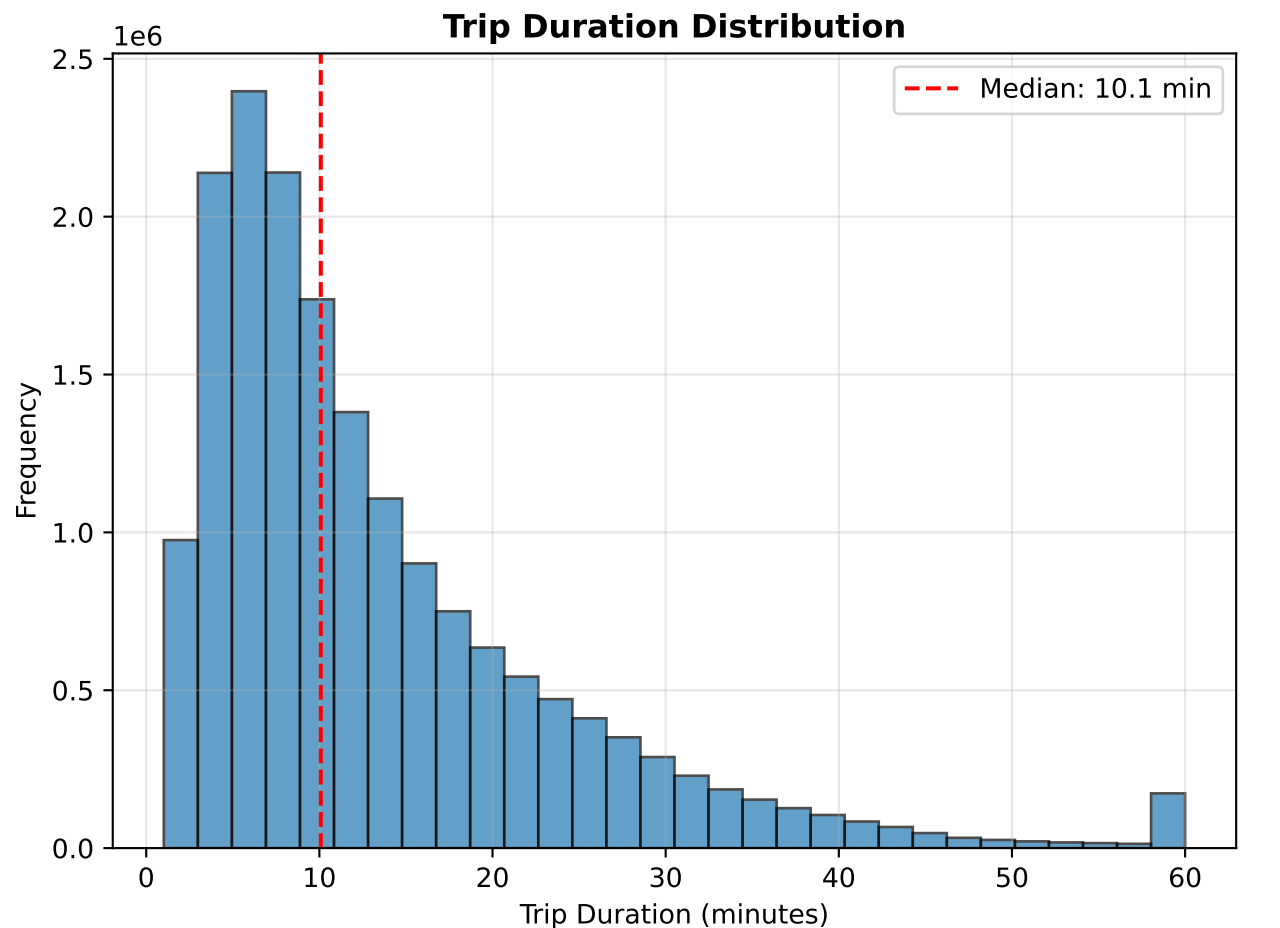
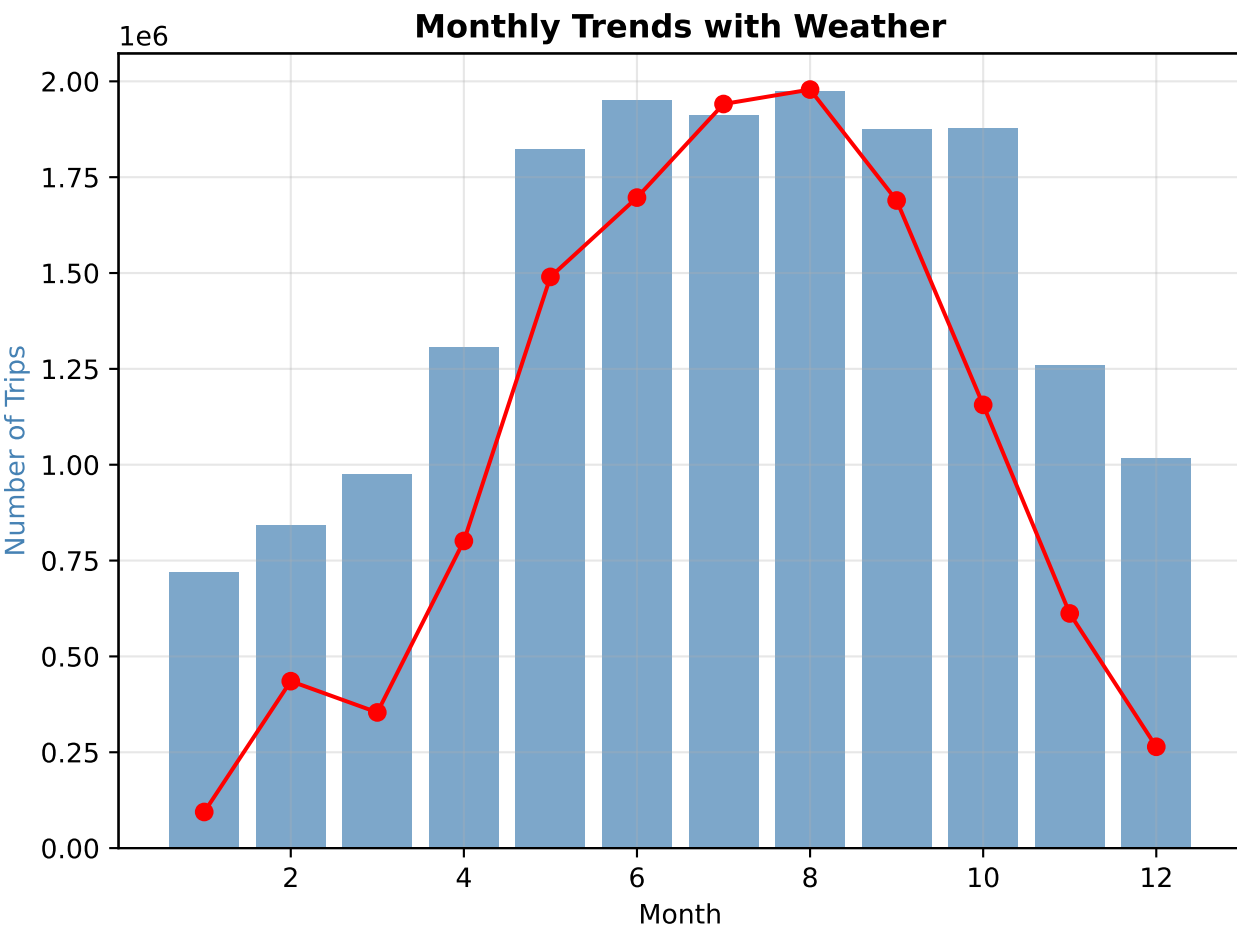
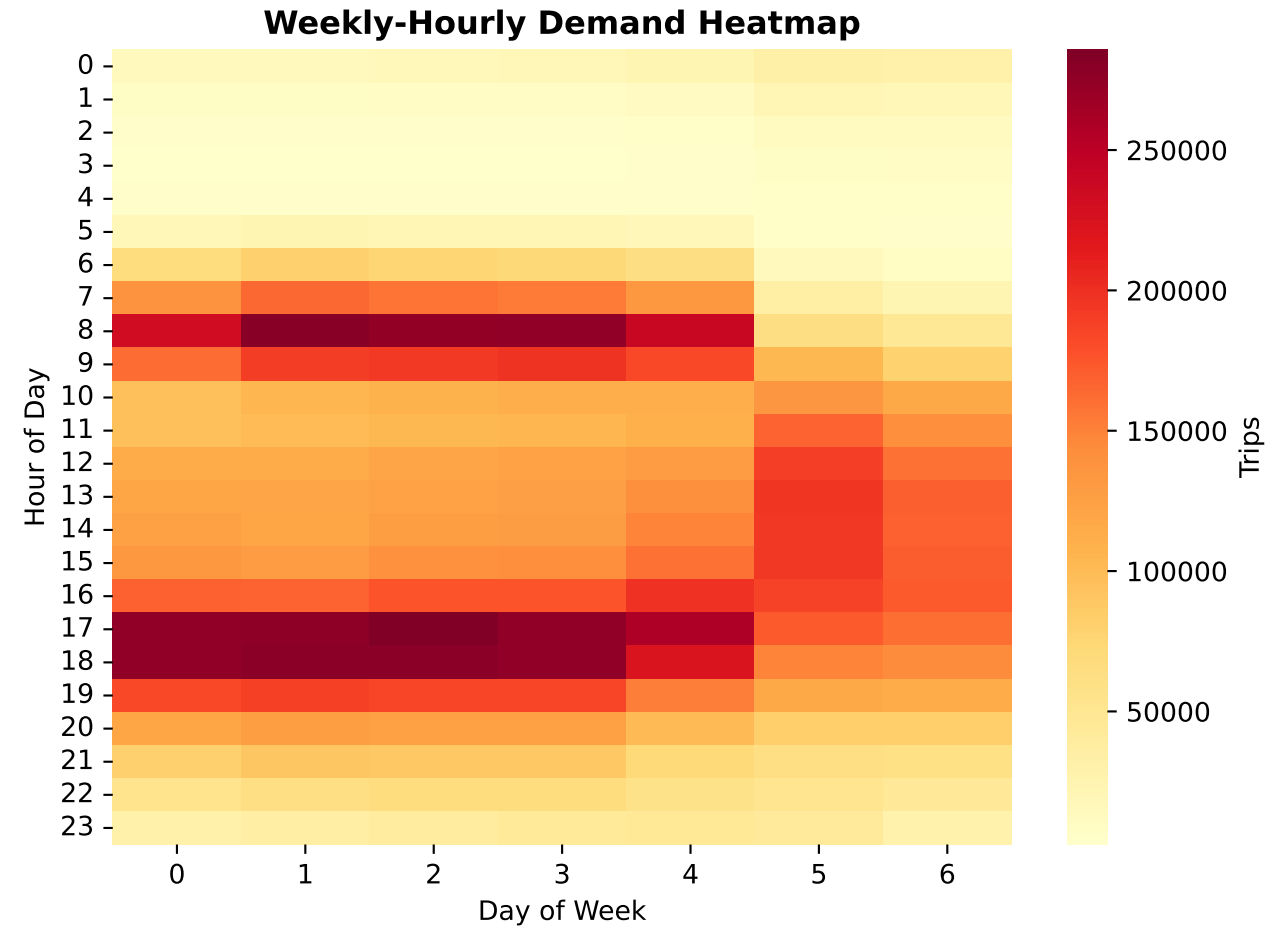
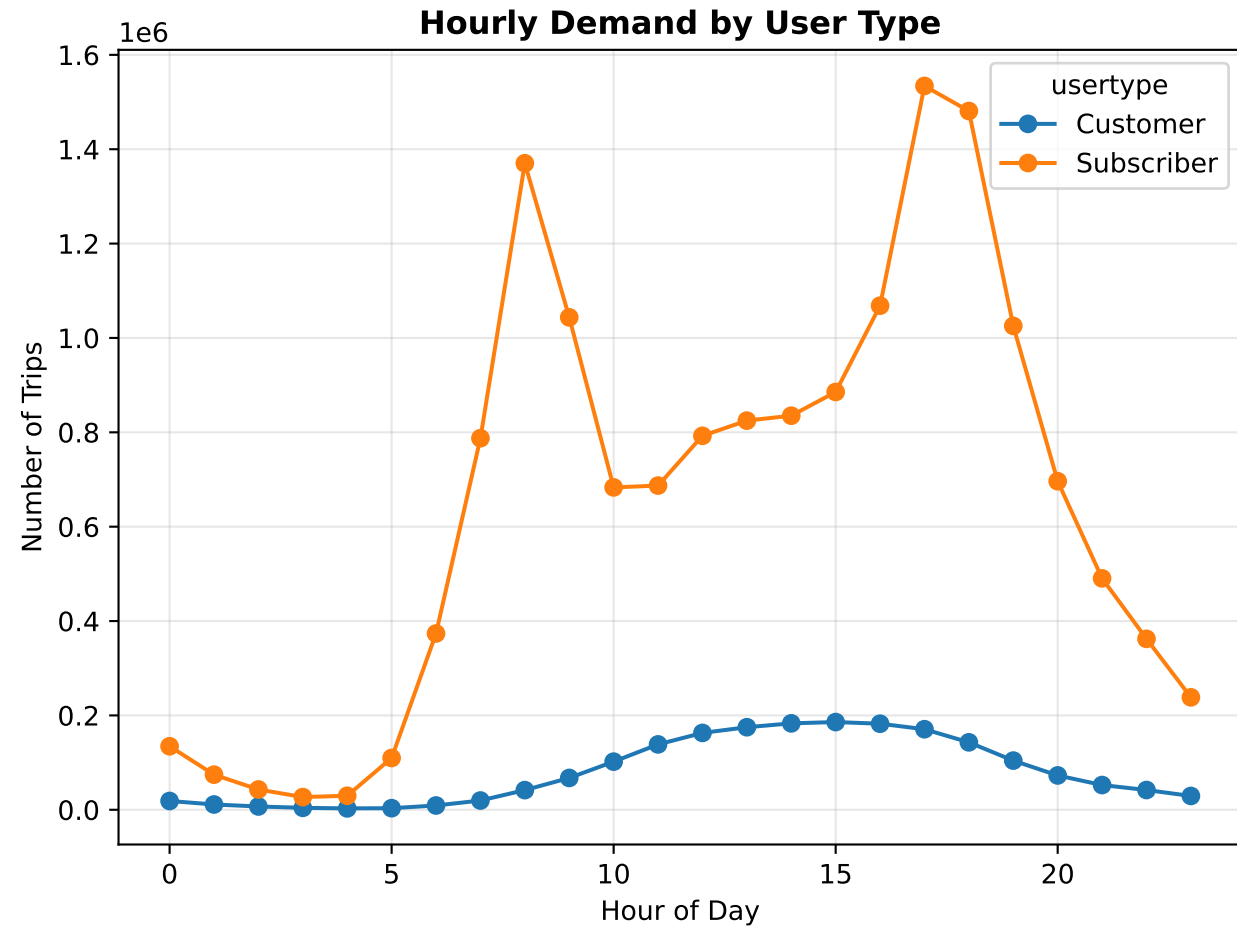


Clusters in PCA Space

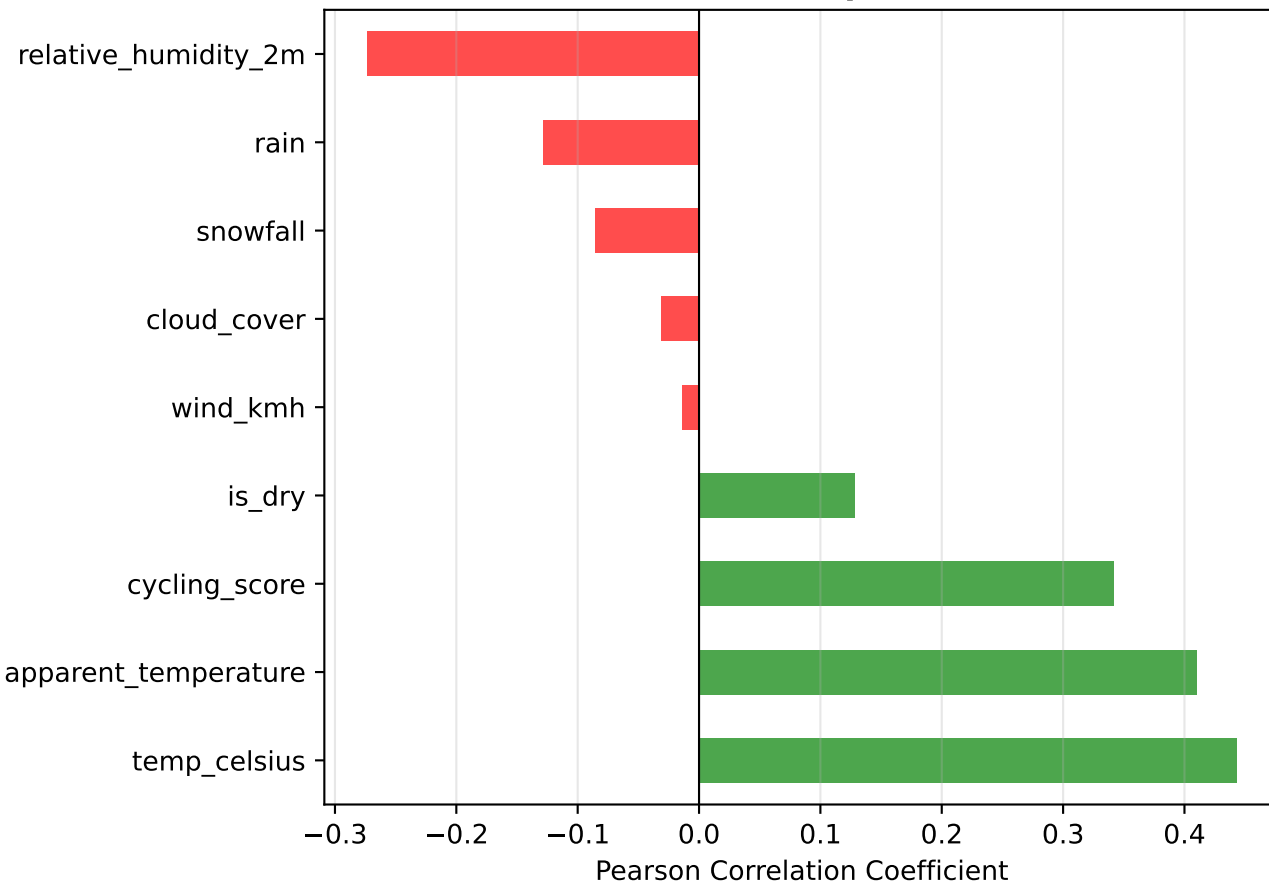


User Distribution

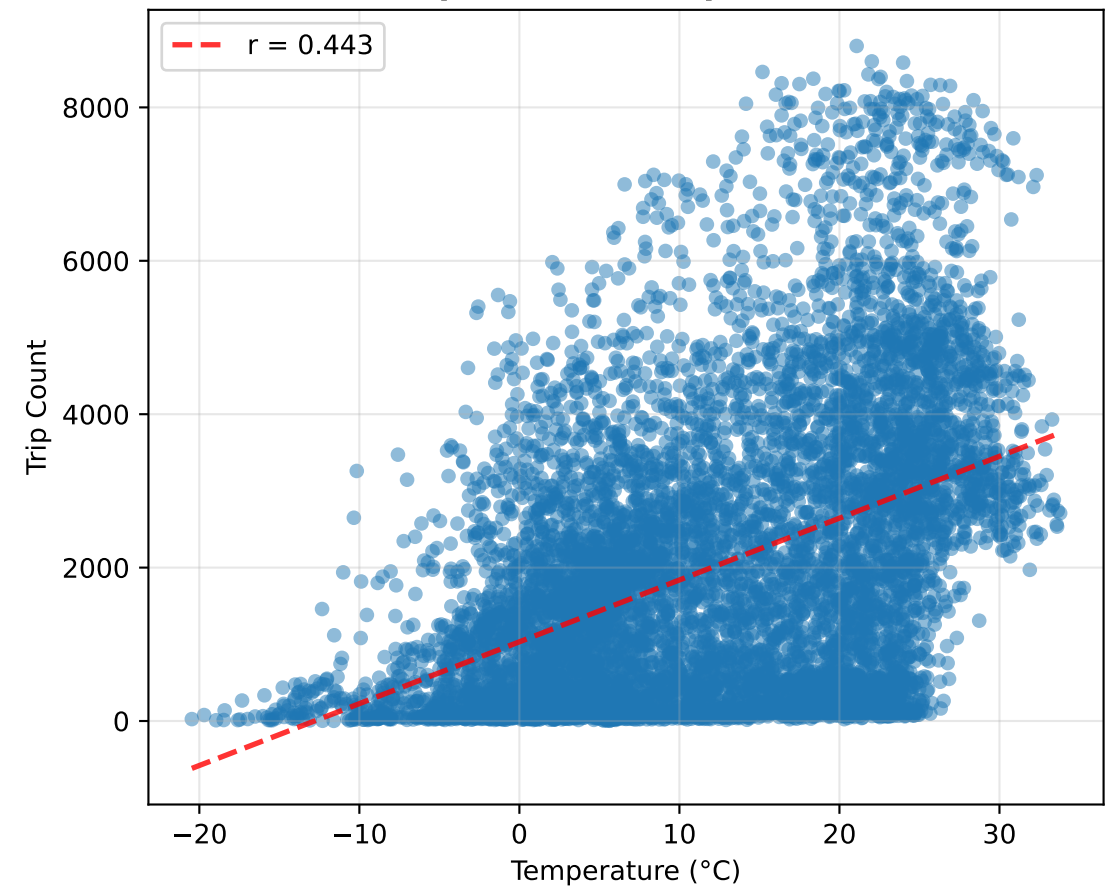




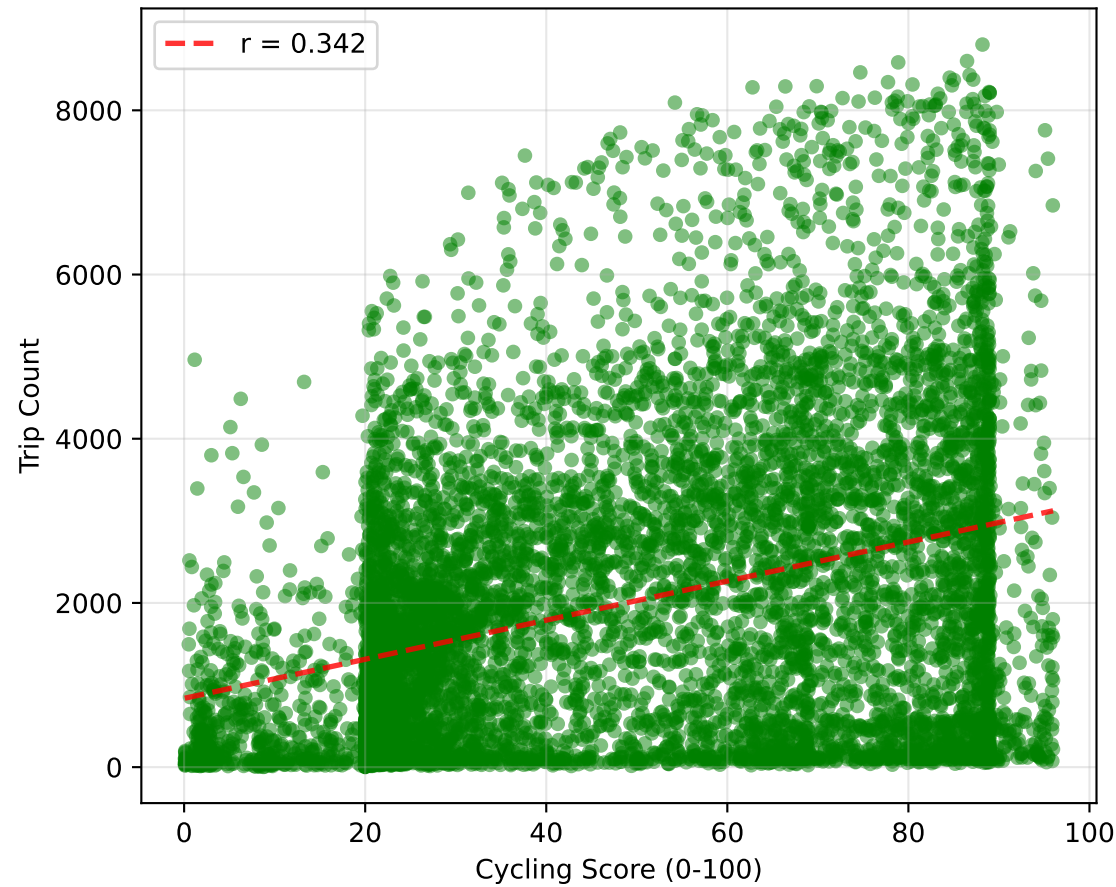
Weather Variables vs Trip Count Correlation



Temperature vs Trip Demand



Cycling Score vs Trip Demand



Weather Correlation Heatmap

