

Bike Sharing Analysis - Summary Statistics

DATASET OVERVIEW

Total Trips: 17,548,339

Date Range: 2018-01-01 → 2018-12-31

Total Users: 263 unique user groups

PATTERN MINING RESULTS

Station Pairs: 347,869

Bidirectional Pairs: 201,016

Zone Connections: 268

Association Rules: 258

Frequent Itemsets: 278

TOP 10 ROUTES

1. 2006 → 2006: 7,852 trips (0.04%)
2. 432 → 3263: 7,249 trips (0.04%)
3. 281 → 281: 6,344 trips (0.04%)
4. 2006 → 3282: 6,215 trips (0.04%)
5. 460 → 3093: 5,456 trips (0.03%)
6. 3093 → 460: 5,137 trips (0.03%)
7. 435 → 509: 4,996 trips (0.03%)
8. 519 → 492: 4,663 trips (0.03%)
9. 519 → 498: 4,465 trips (0.03%)
10. 387 → 387: 4,464 trips (0.03%)

COMMUTE PATTERNS

Morning Rush (7-9 AM): 3,332,116 (19.0%)

Evening Rush (5-7 PM): 4,463,897 (25.4%)

Non-Commute Hours: 9,749,829 (55.6%)

WEATHER IMPACT

Good Weather Trips: 6,396,349 (36.5%)

Poor Weather Trips: 11,149,493 (63.5%)

CLUSTERING RESULTS

Optimal Number of Clusters: 3

Clustering Method: average

Davies-Bouldin Score: 0.161

User Clustering - Detailed Summary

CLUSTER CHARACTERISTICS

CLUSTER 1: 261 users (99.2%)

Age: 63.0 years
Trips/month: 20647.3
Trip duration: 24.7 min
Weekend usage: 25.2%
Station diversity: 275.0
Weather score: 57.5/100

CLUSTER 2: 1 users (0.4%)

Age: 49.0 years
Trips/month: 460481.0
Trip duration: 35.3 min
Weekend usage: 41.4%
Station diversity: 765.5
Weather score: 58.3/100

CLUSTER 3: 1 users (0.4%)

Age: 131.0 years
Trips/month: 5.7
Trip duration: 1591.9 min
Weekend usage: 42.2%
Station diversity: 3.5
Weather score: 81.0/100

CLUSTER SUMMARY TABLE

cluster	age	trip_count	avg_trip_min	weekend_ratio	station_diversity	avg_cycling_score
1	62.98	61942.02	24.67	0.25	274.96	57.500000
2	49.00	1381443.00	35.27	0.41	765.50	58.320000
3	131.00	17.00	1591.89	0.42	3.50	80.989998

Weather Correlation Analysis

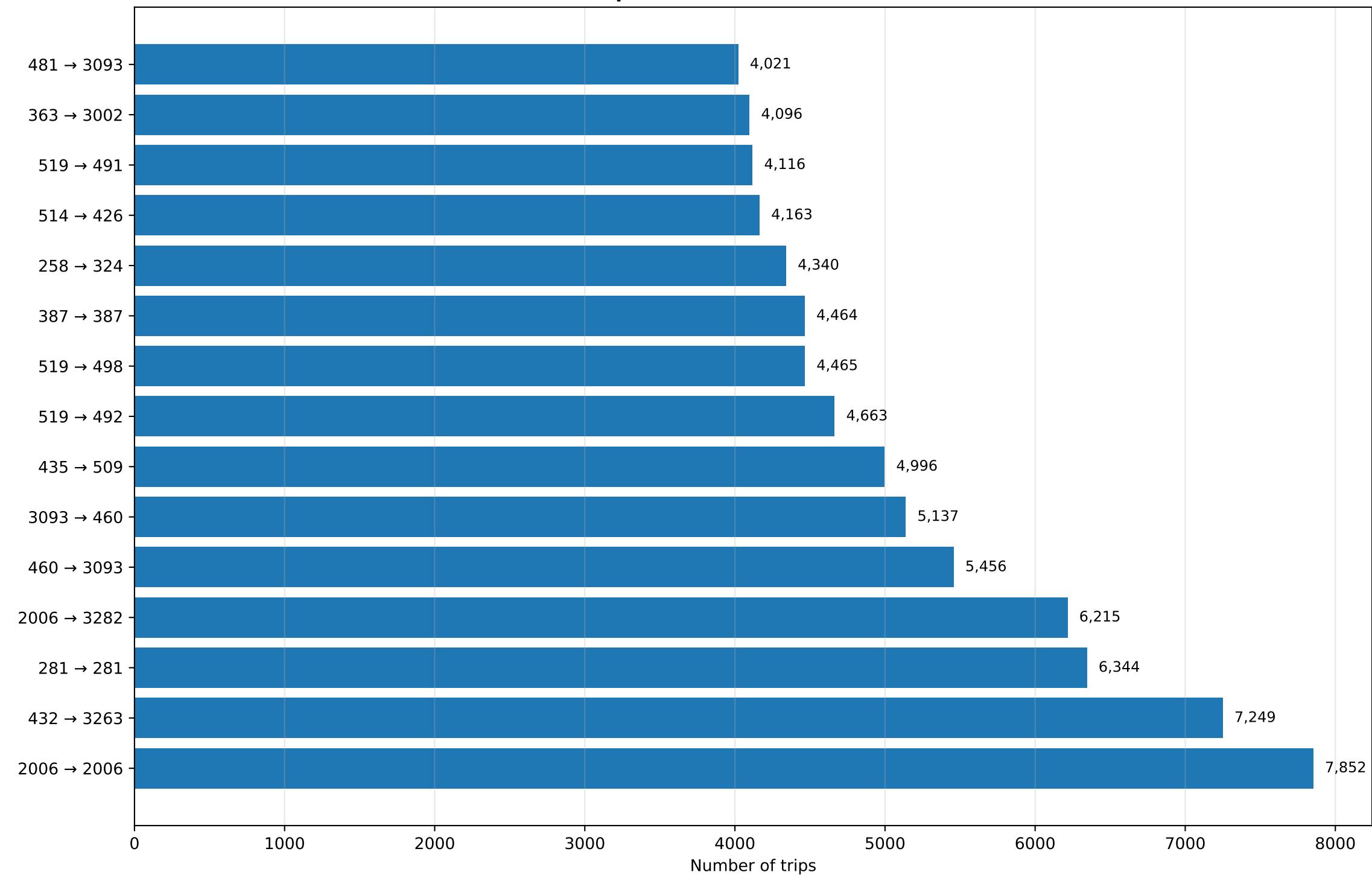
PEARSON CORRELATION: WEATHER vs TRIP DEMAND

Variable	Correlation	p-value	Significance
temp_celsius	0.4425	0.00e+00	***
apparent_temperature	0.4099	0.00e+00	***
cycling_score	0.3415	9.38e-238	***
is_dry	0.1279	3.18e-33	***
wind_kmh	-0.0140	1.90e-01	n.s.
cloud_cover	-0.0310	3.78e-03	**
snowfall	-0.0850	1.69e-15	***
rain	-0.1280	2.80e-33	***
relative_humidity_2m	-0.2731	2.51e-149	***

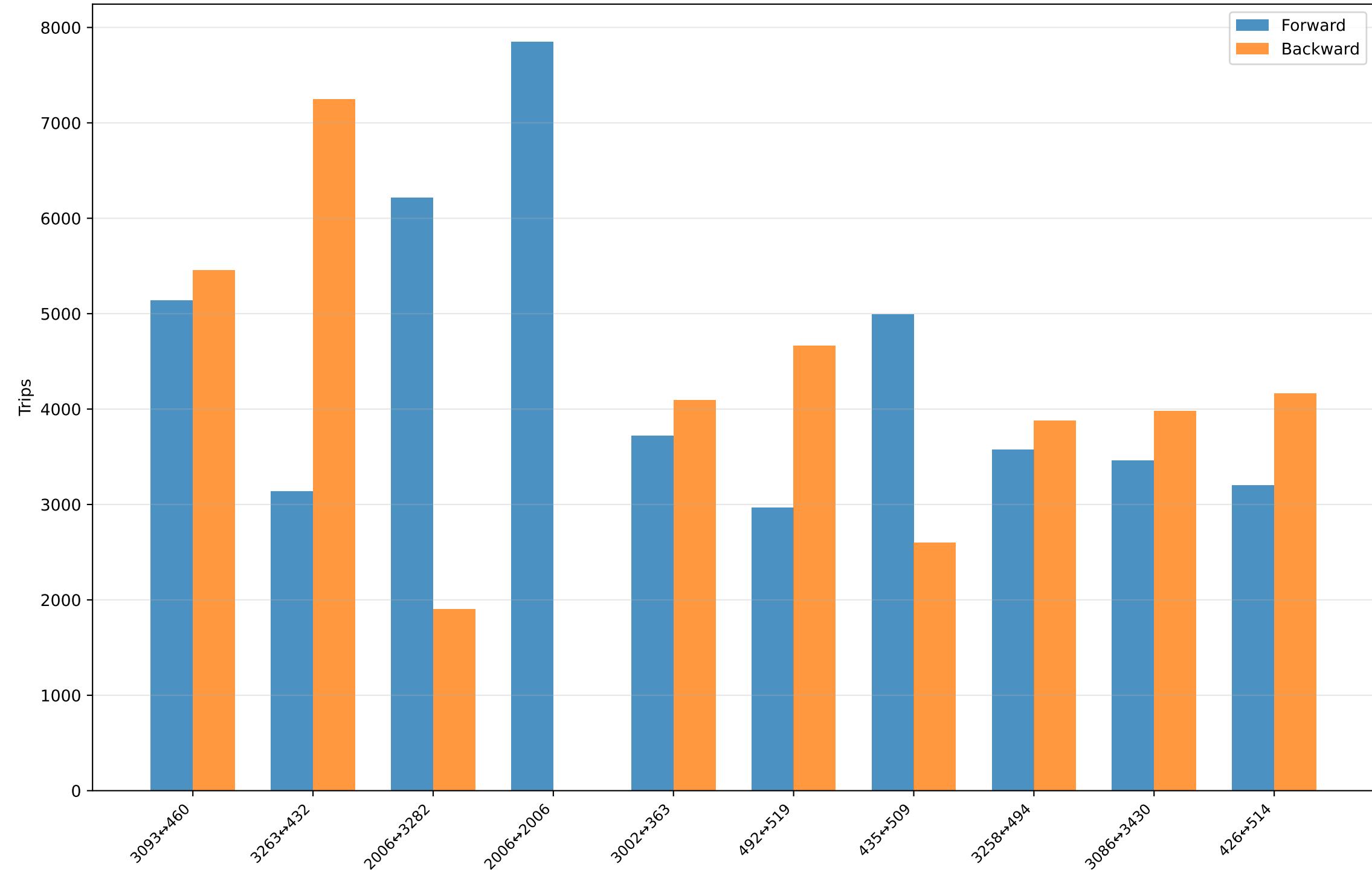
KEY INSIGHTS

- ✓ 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 ($p < 0.001$)

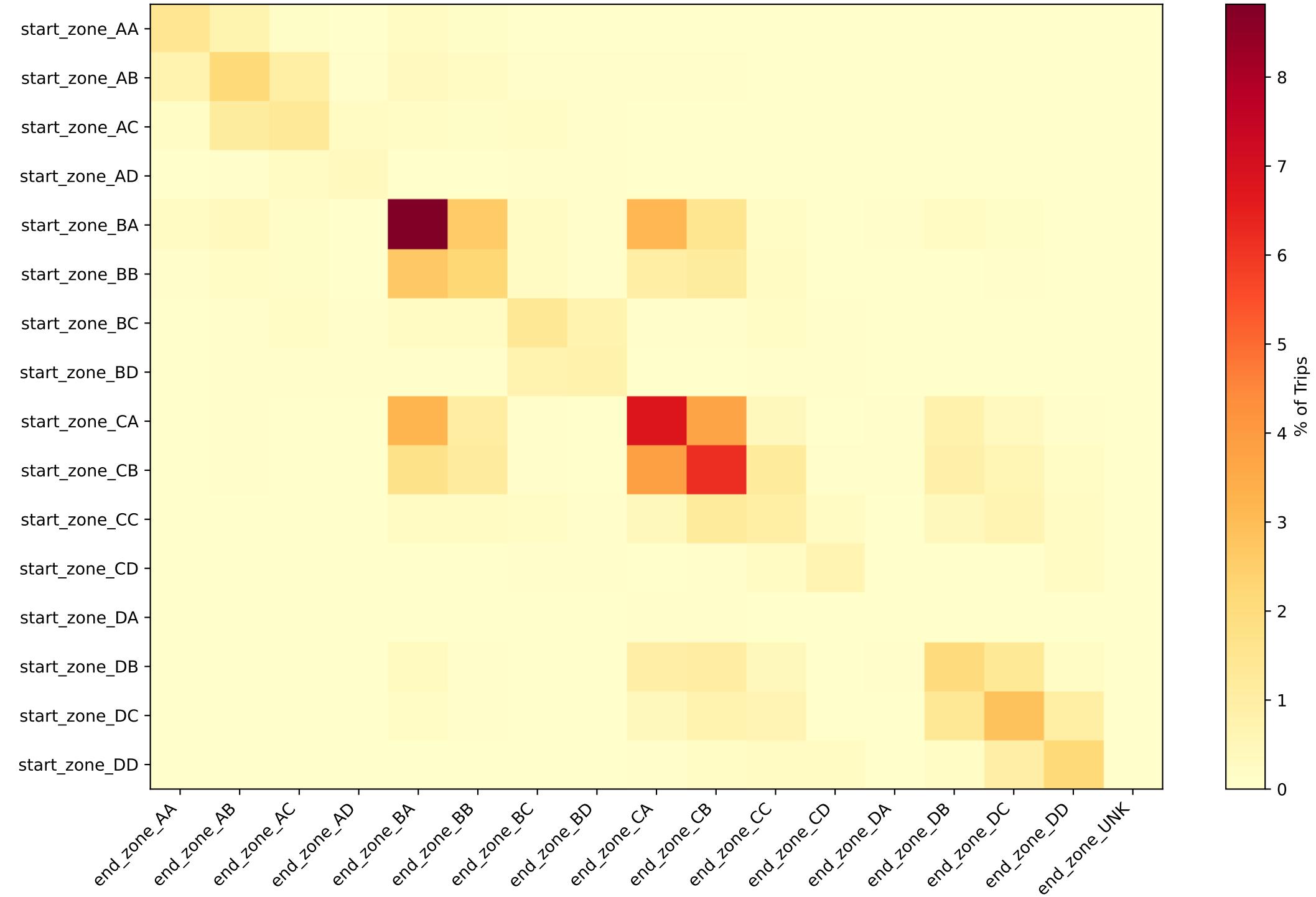
Top Station-to-Station Routes



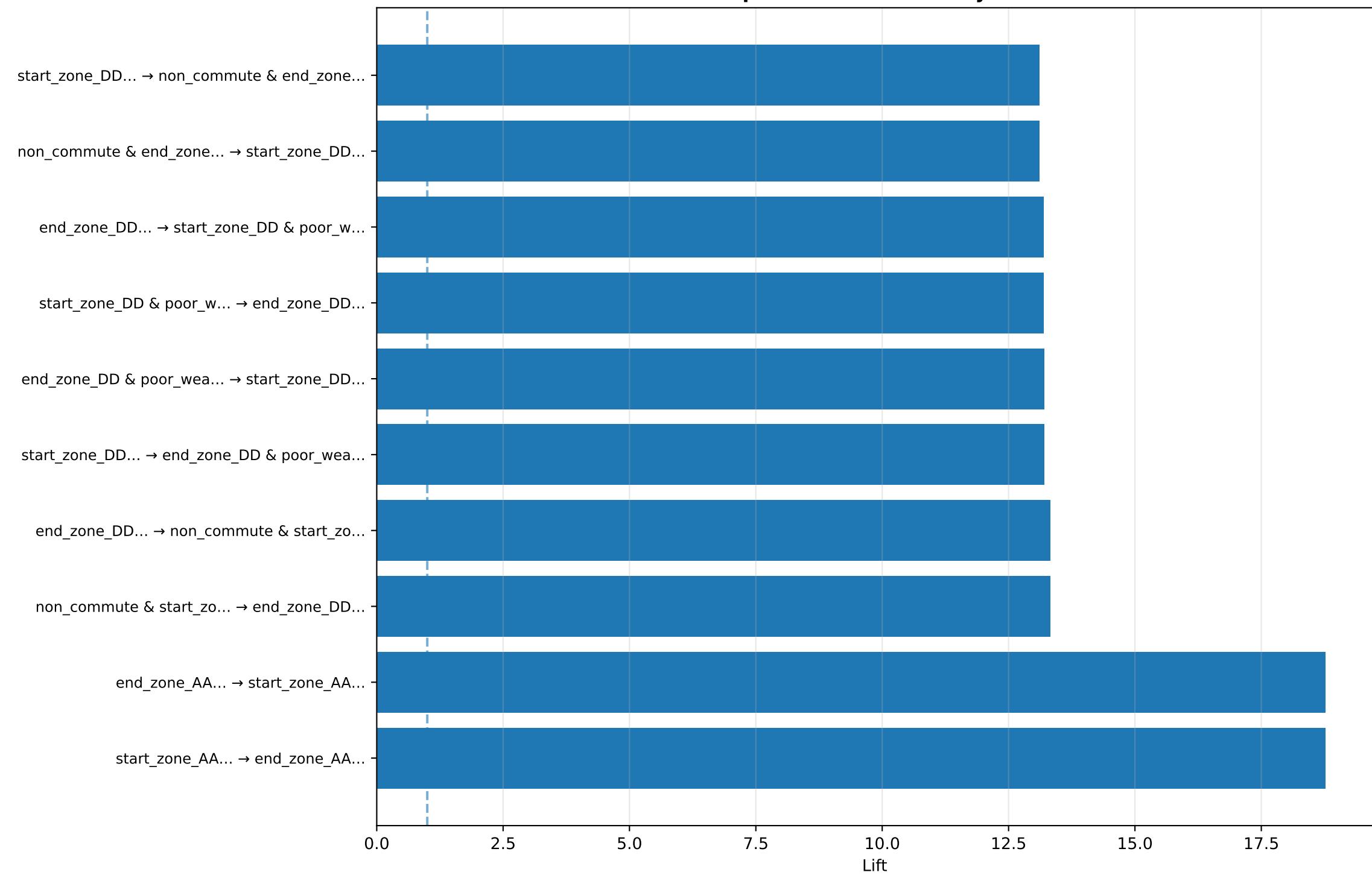
Bidirectional Route Analysis - Rebalancing Needs



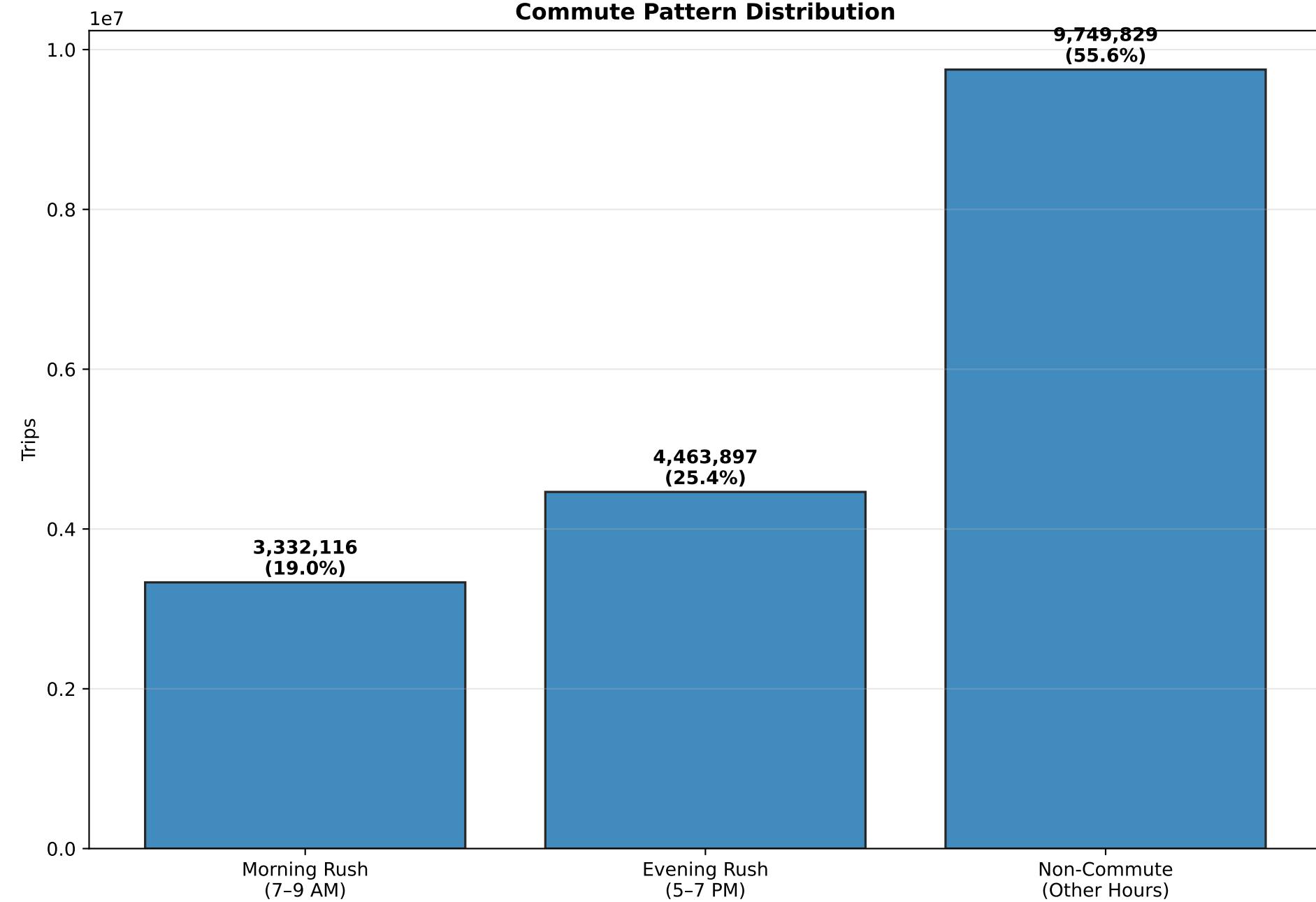
Geo Zone-to-Zone Movement Heatmap



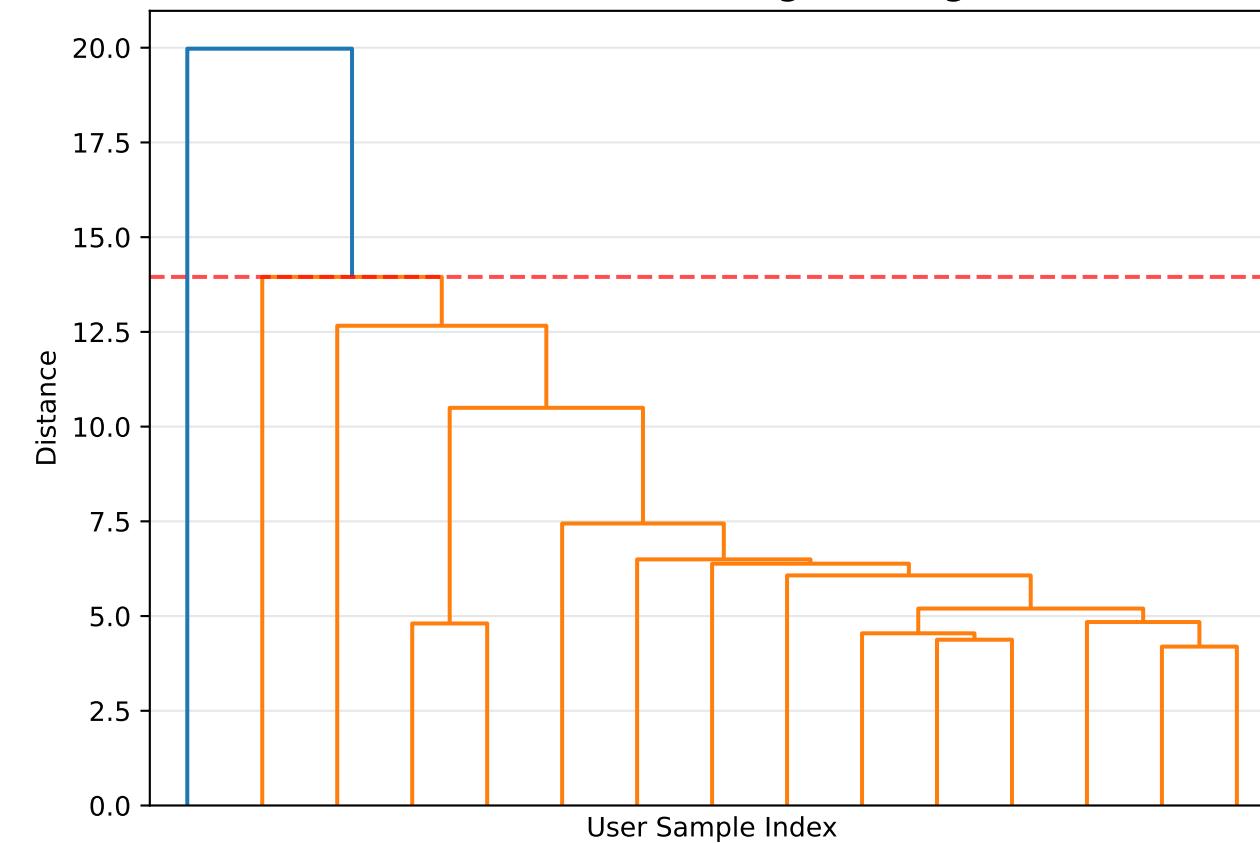
Top Association Rules by Lift



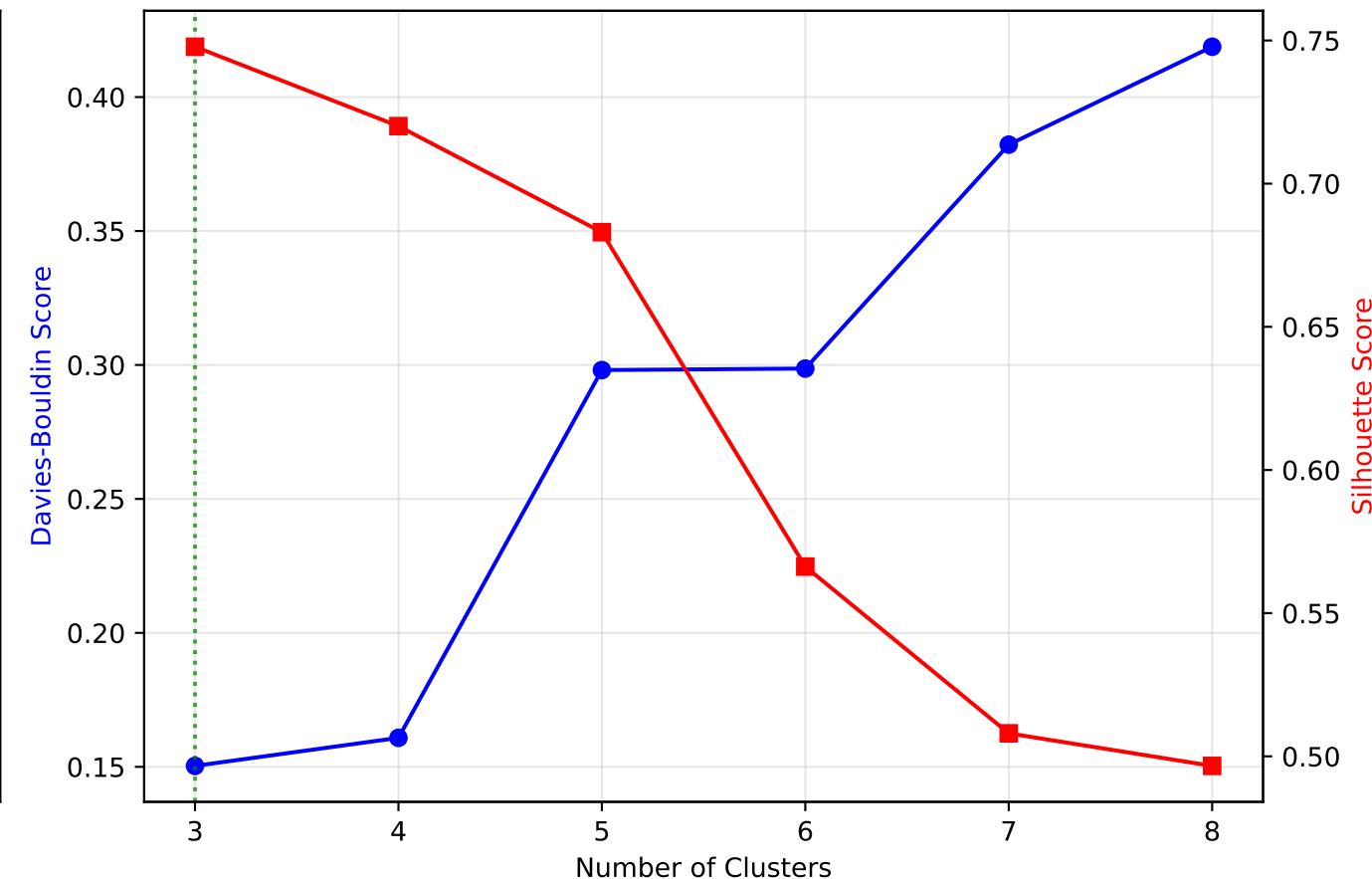
Commute Pattern Distribution



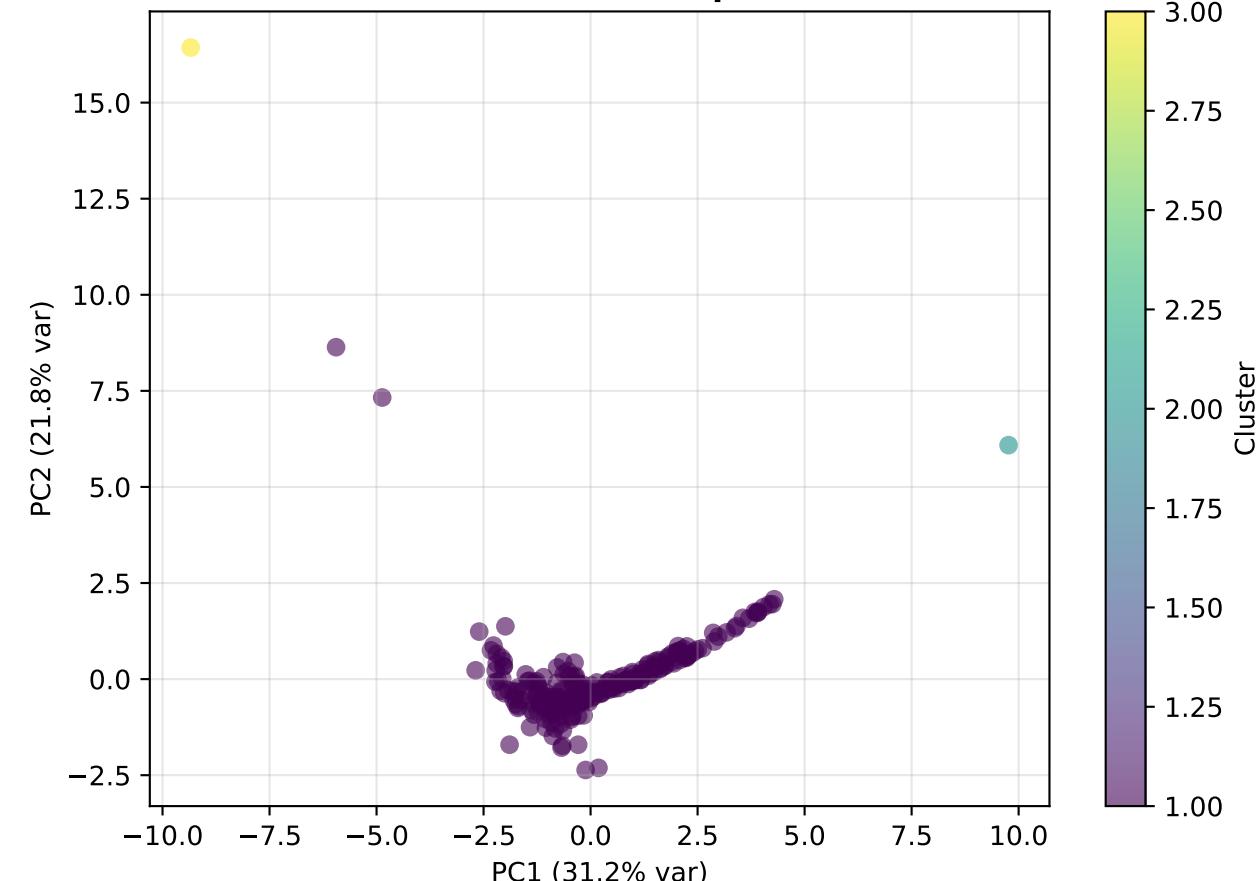
Hierarchical Clustering Dendrogram



Cluster Evaluation Metrics



Clusters in PCA Space



User Distribution

