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
```

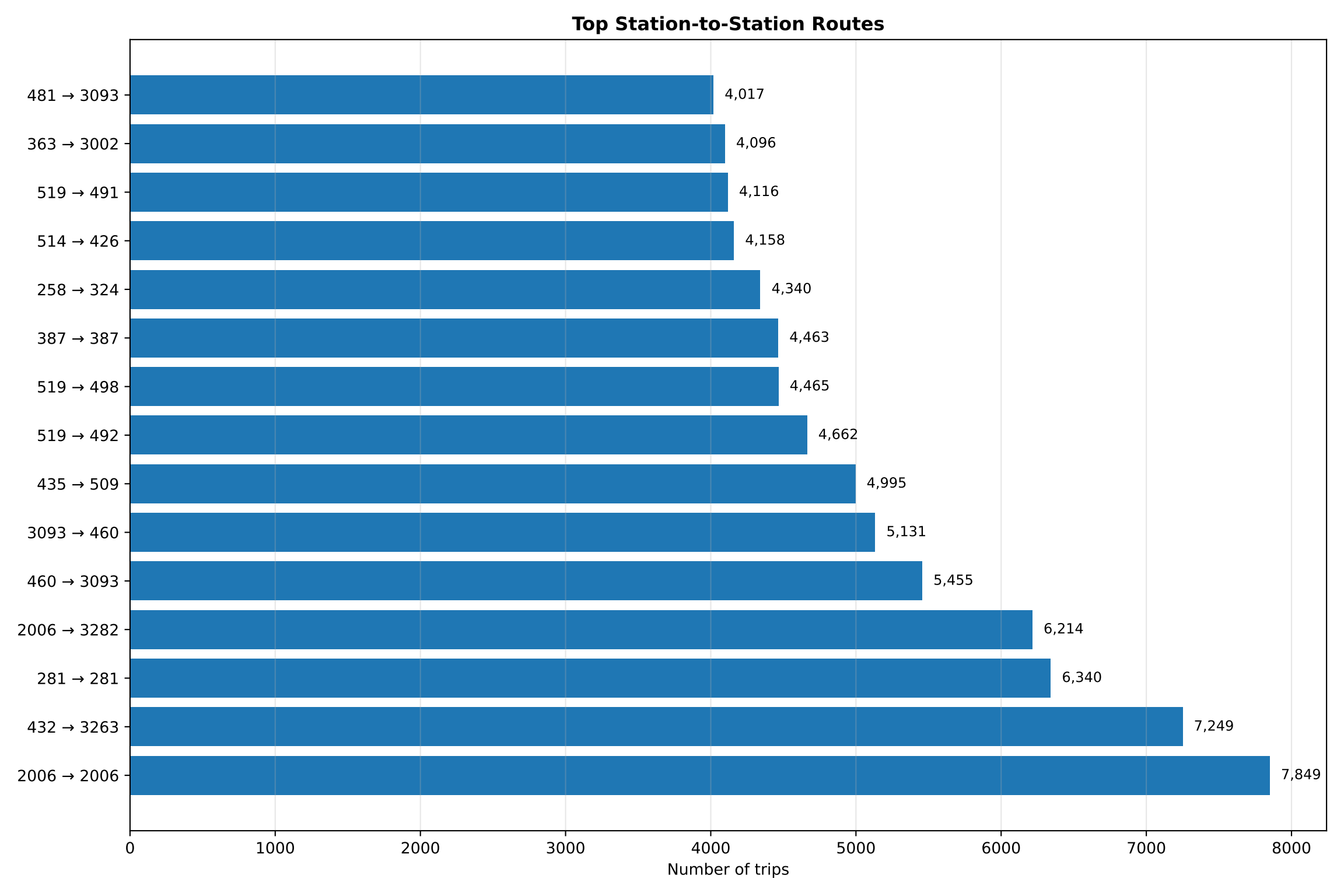
Weather Correlation Analysis

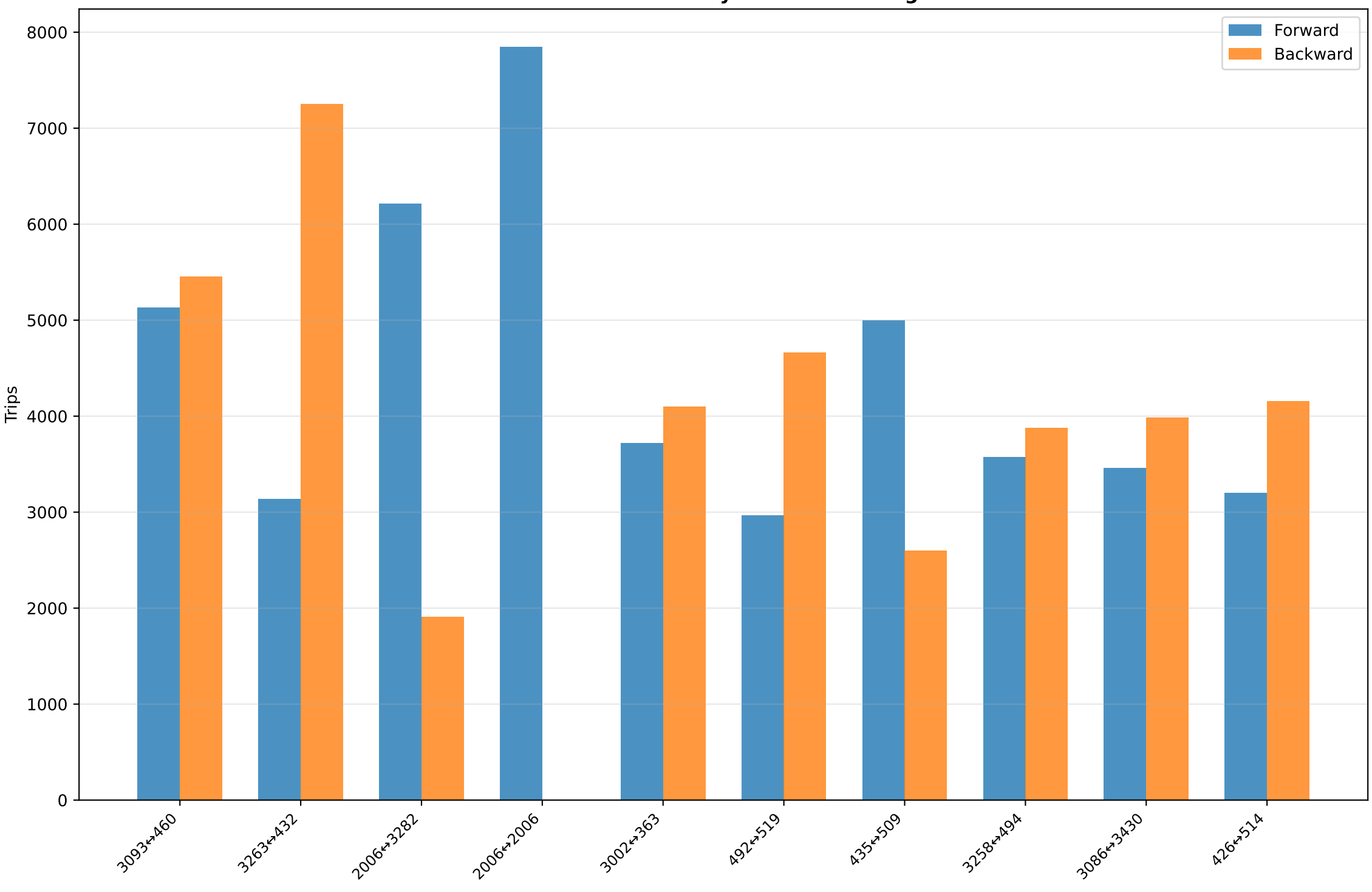
PEARSON CORRELATION: WEATHER vs TRIP DEMAND

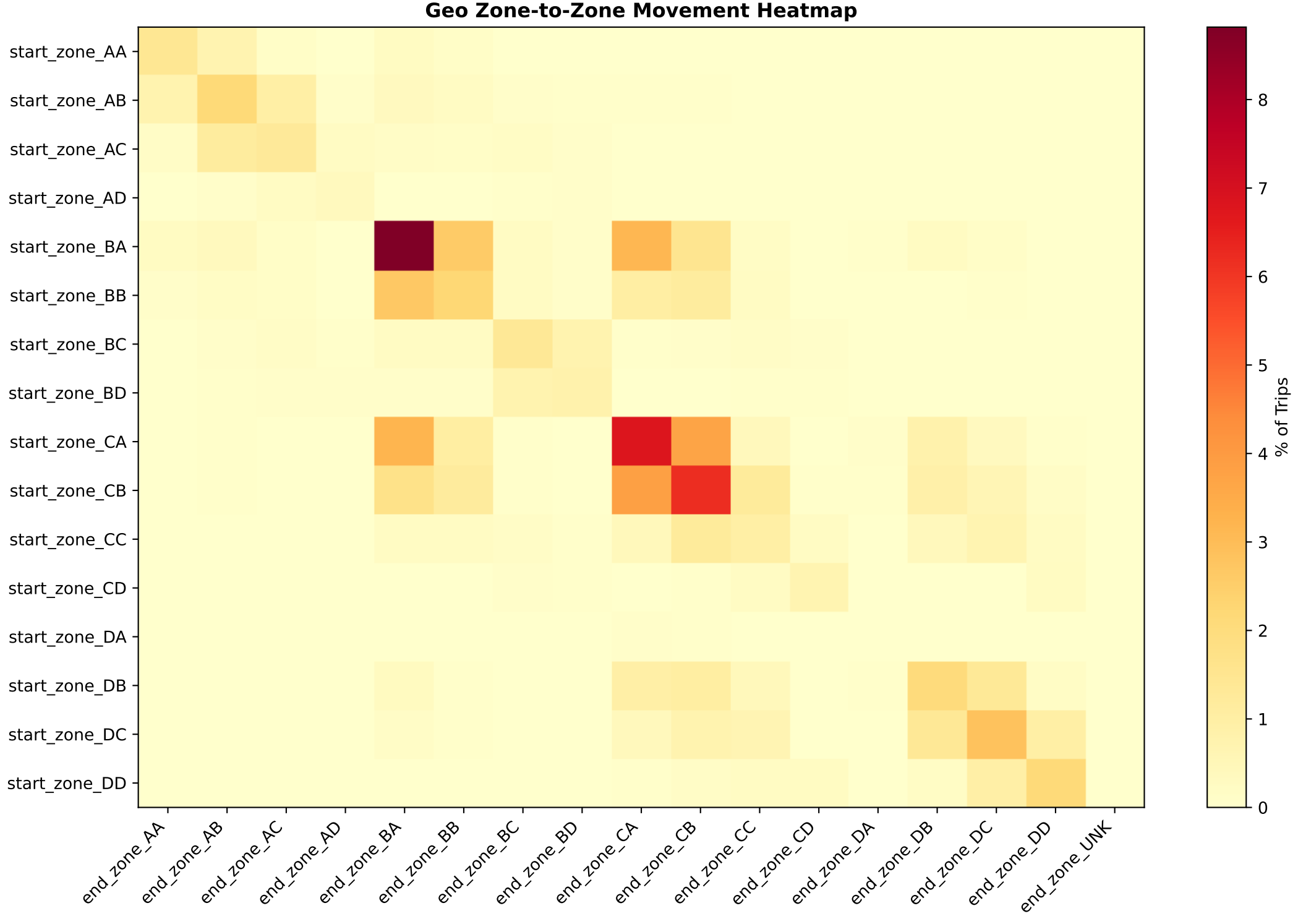
Variable	Correlation	p-value	Significance
temp_celsius	0.4425	0.00e+00	***
apparent_temperature	0.4099	0.00e+00	***
cycling_score	0.3415	9.20e-238	***
is_dry	0.1279	3.36e-33	***
wind_kmh	-0.0141	1.88e-01	n.s.
cloud_cover	-0.0310	3.80e-03	**
snowfall	-0.0850	1.69e-15	***
rain	-0.1280	2.88e-33	***
relative_humidity_2m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

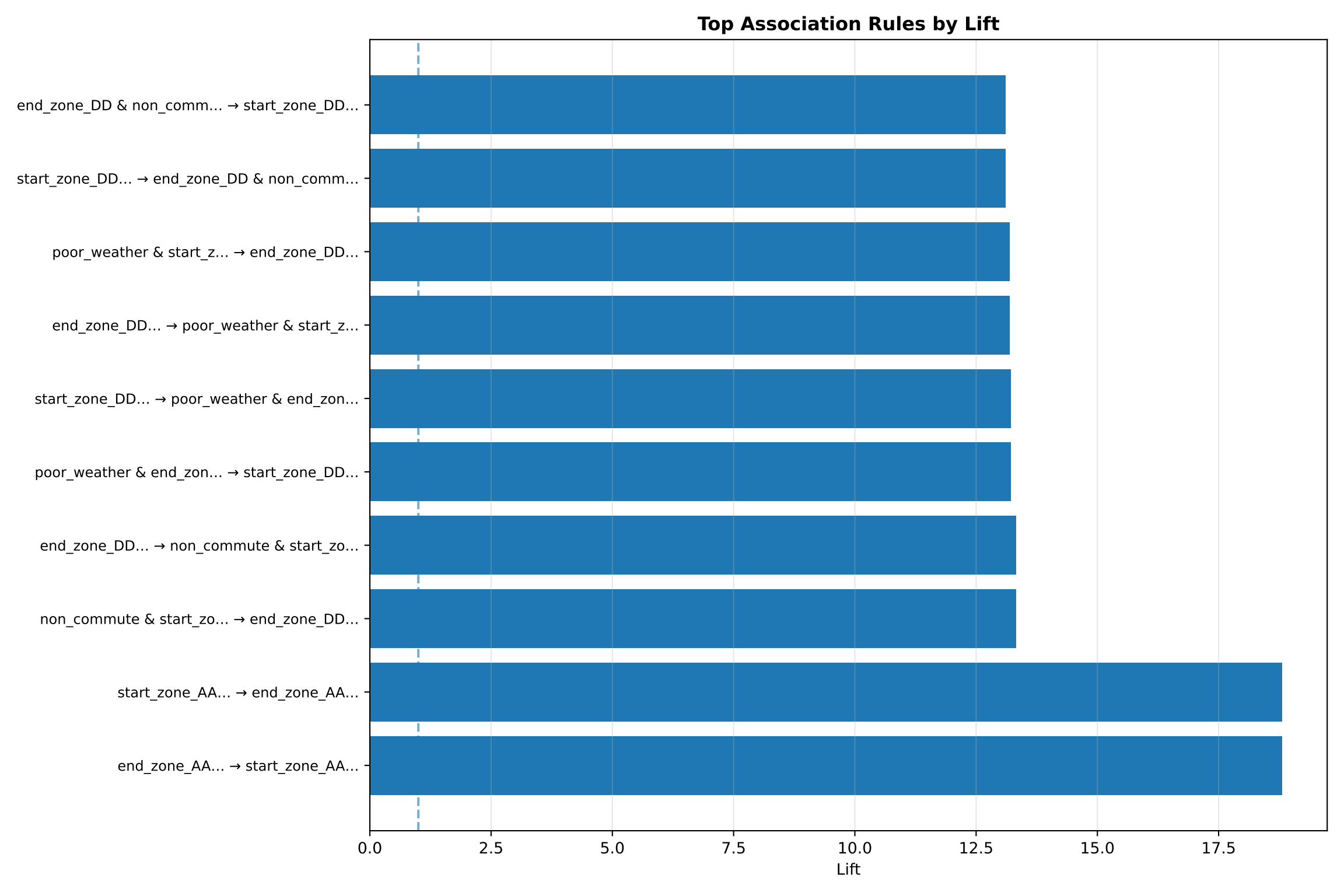
KEY INSIGHTS

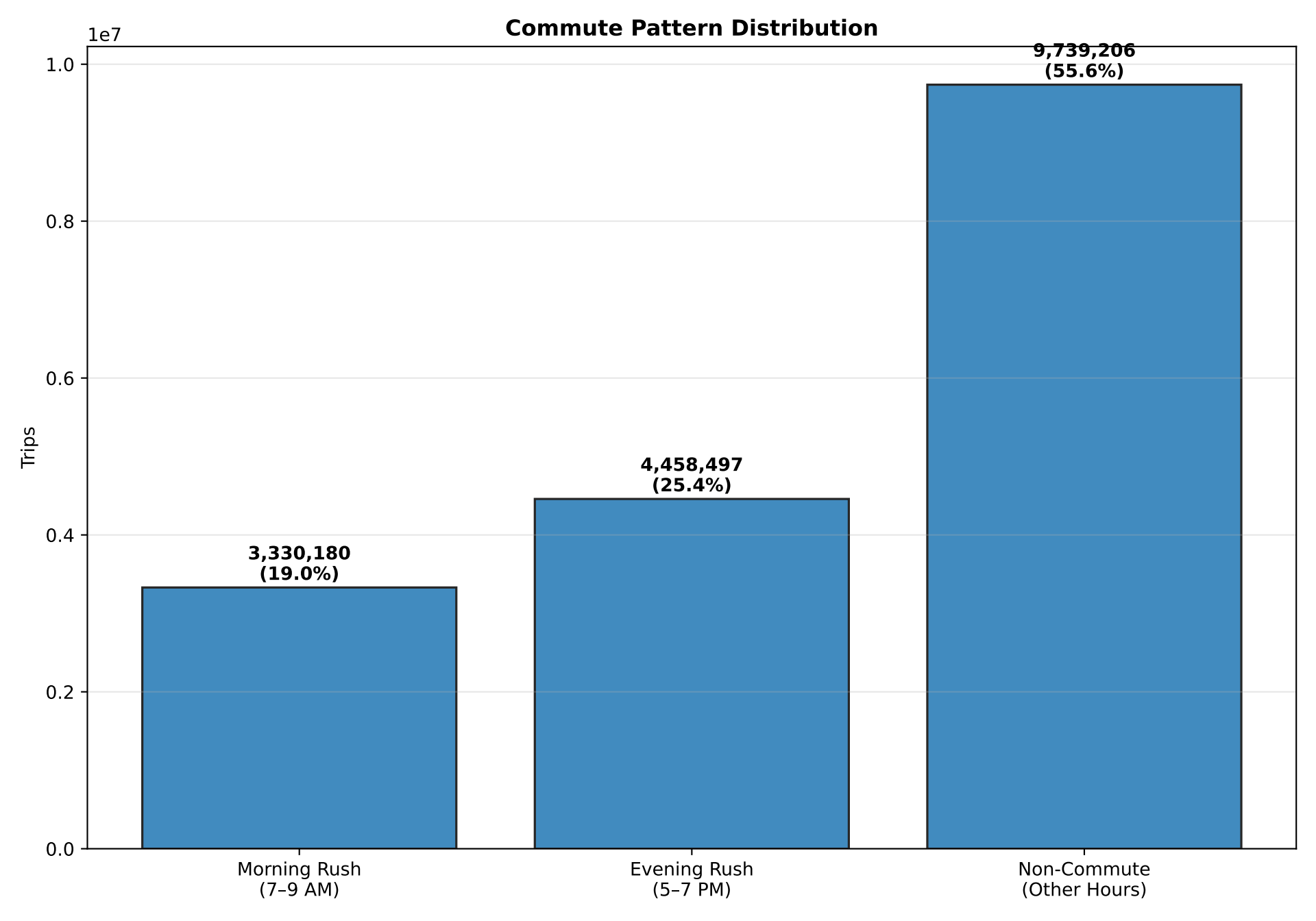
✓ Strongest positive: temp_celsius (r=0.4425)
✓ 8 variables statistically significant (p < 0.05)
✓ 3 variables show moderate-to-strong correlation (|r| > 0.3)
✓ Cycling score explains 11.7% of variance
✓ 7 variables highly significant (p < 0.001)

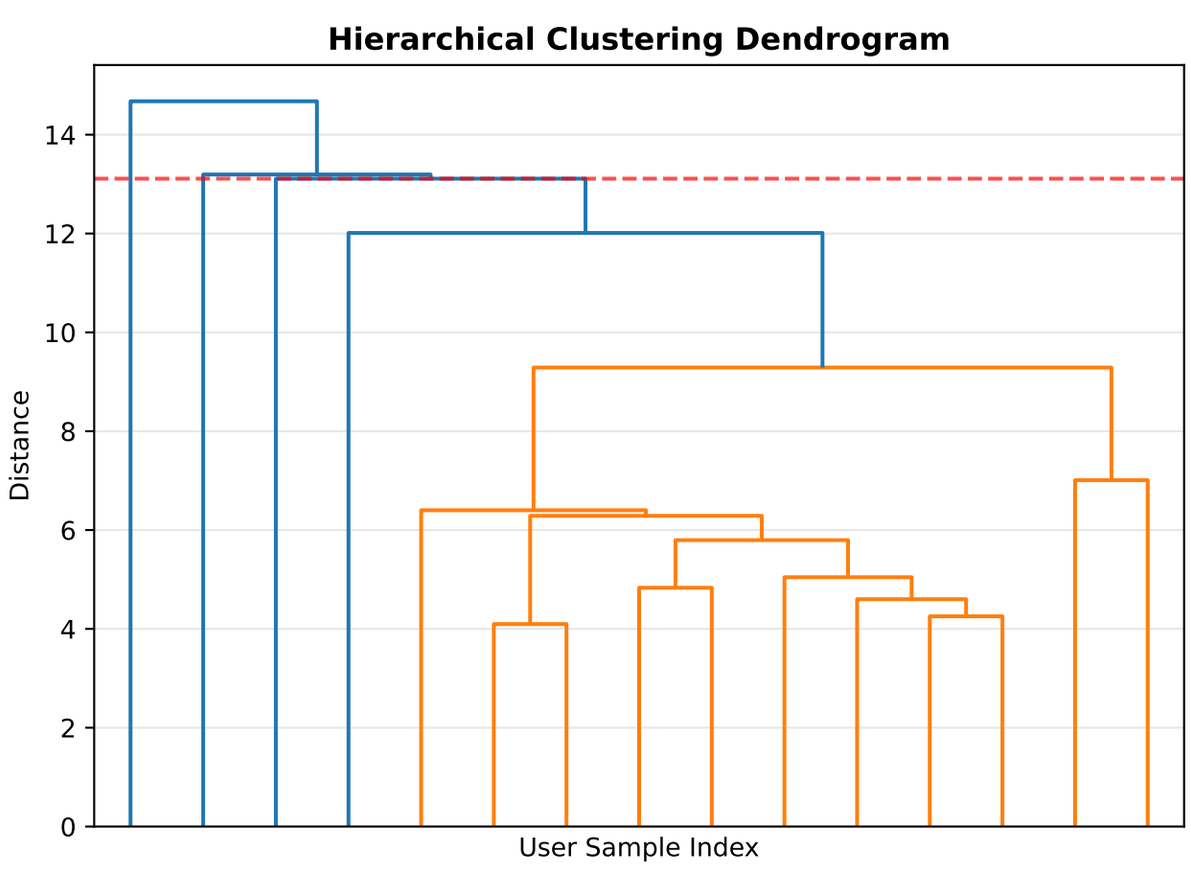

Top Station-to-Station Routes
481 → 3093
4,017
363 → 3002
4,096
519 → 491
4,116
514 → 426
4,158
258 → 324
4,340
387 → 387
4,463
519 → 498
4,465
519 → 492
4,662
435 → 509
4,995
3093 → 460
5,131
460 → 3093
5,455
2006 → 3282
6,214
281 → 281
6,340
432 → 3263
7,249
2006 → 2006
7,849
0
1000
2000
3000
4000
5000
6000
7000
8000
Number of trips


Bidirectional Route Analysis - Rebalancing Needs
Forward
Backward
Trips
8000
7000
6000
5000
4000
3000
2000
1000
0
3093↔460
3263↔432
2006↔3282
2006↔2006
3002↔363
492↔519
435↔509
3258↔494
3086↔3430
426↔514

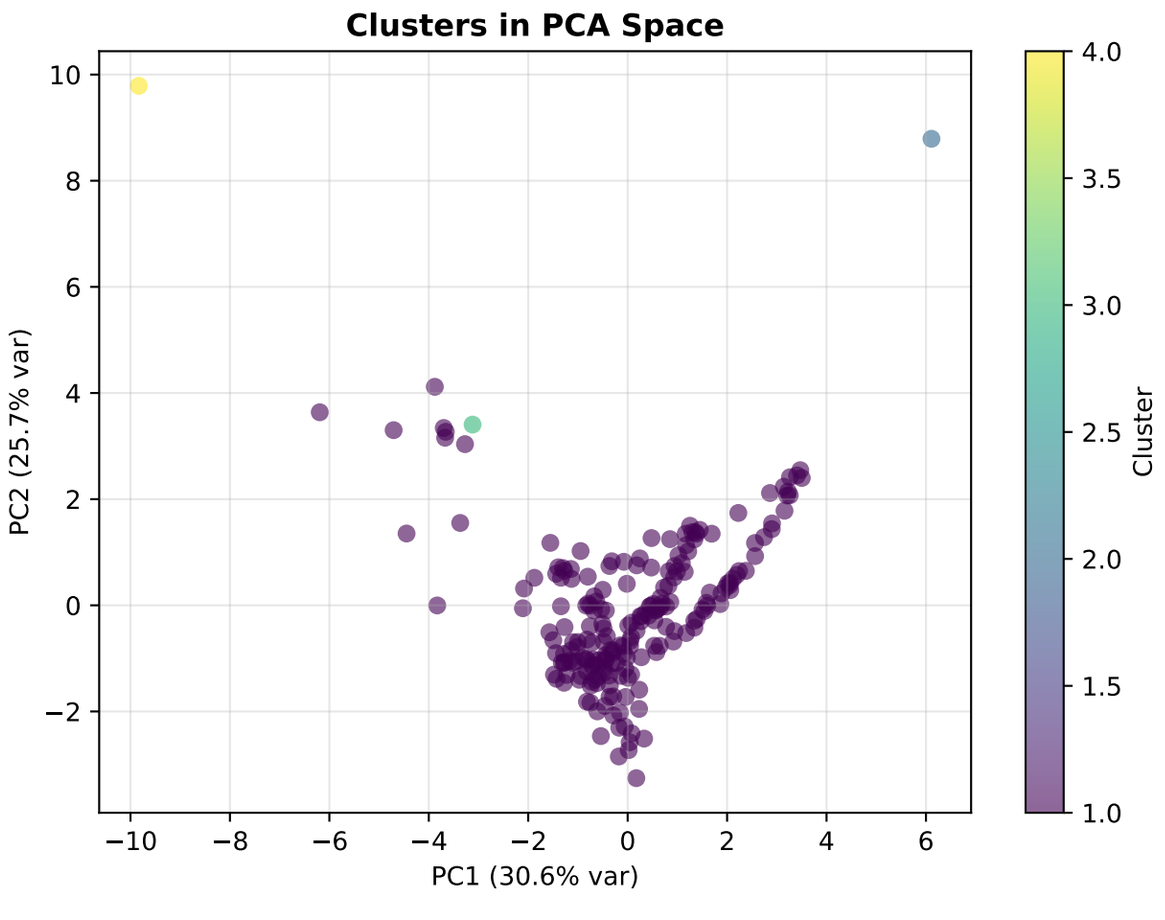

Geo Zone-to-Zone Movement Heatmap
start_zone_AA
start_zone_AB
start_zone_AC
start_zone_AD
start_zone_BA
start_zone_BB
start_zone_BC
start_zone_BD
start_zone_CA
start_zone_CB
start_zone_CC
start_zone_CD
start_zone_DA
start_zone_DB
start_zone_DC
start_zone_DD
end_zone_AA
end_zone_AB
end_zone_AC
end_zone_AD
end_zone_BA
end_zone_BB
end_zone_BC
end_zone_BD
end_zone_CA
end_zone_CB
end_zone_CC
end_zone_CD
end_zone_DA
end_zone_DB
end_zone_DC
end_zone_DD
end_zone_UNK
% of Trips
0
1
2
3
4
5
6
7
8

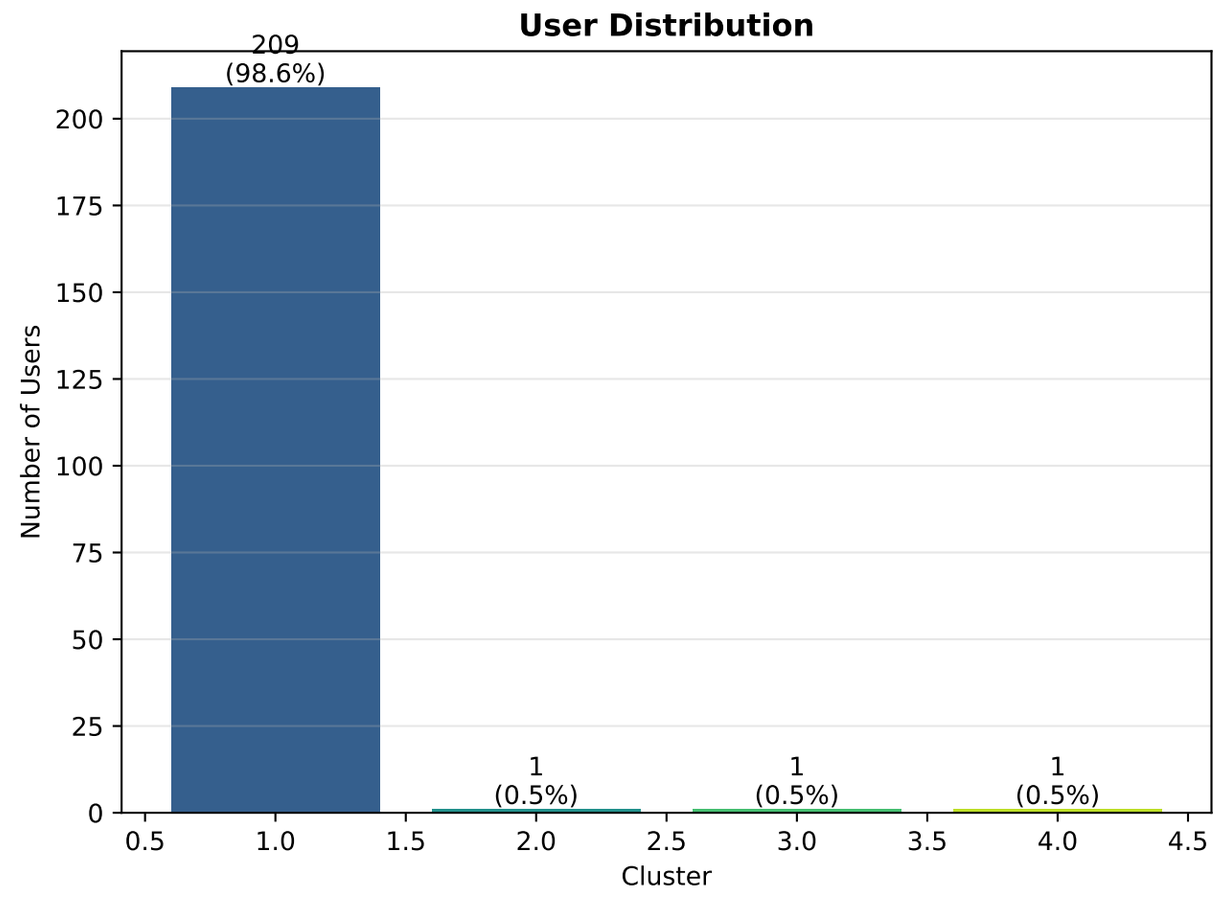

Top Association Rules by Lift
end_zone_DD & non_comm… → start_zone_DD…
start_zone_DD… → end_zone_DD & non_comm…
poor_weather & start_z… → end_zone_DD…
end_zone_DD… → poor_weather & start_z…
start_zone_DD… → poor_weather & end_zon…
poor_weather & end_zon… → start_zone_DD…
end_zone_DD… → non_commute & start_zo…
non_commute & start_zo… → end_zone_DD…
start_zone_AA… → end_zone_AA…
end_zone_AA… → start_zone_AA…
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Lift

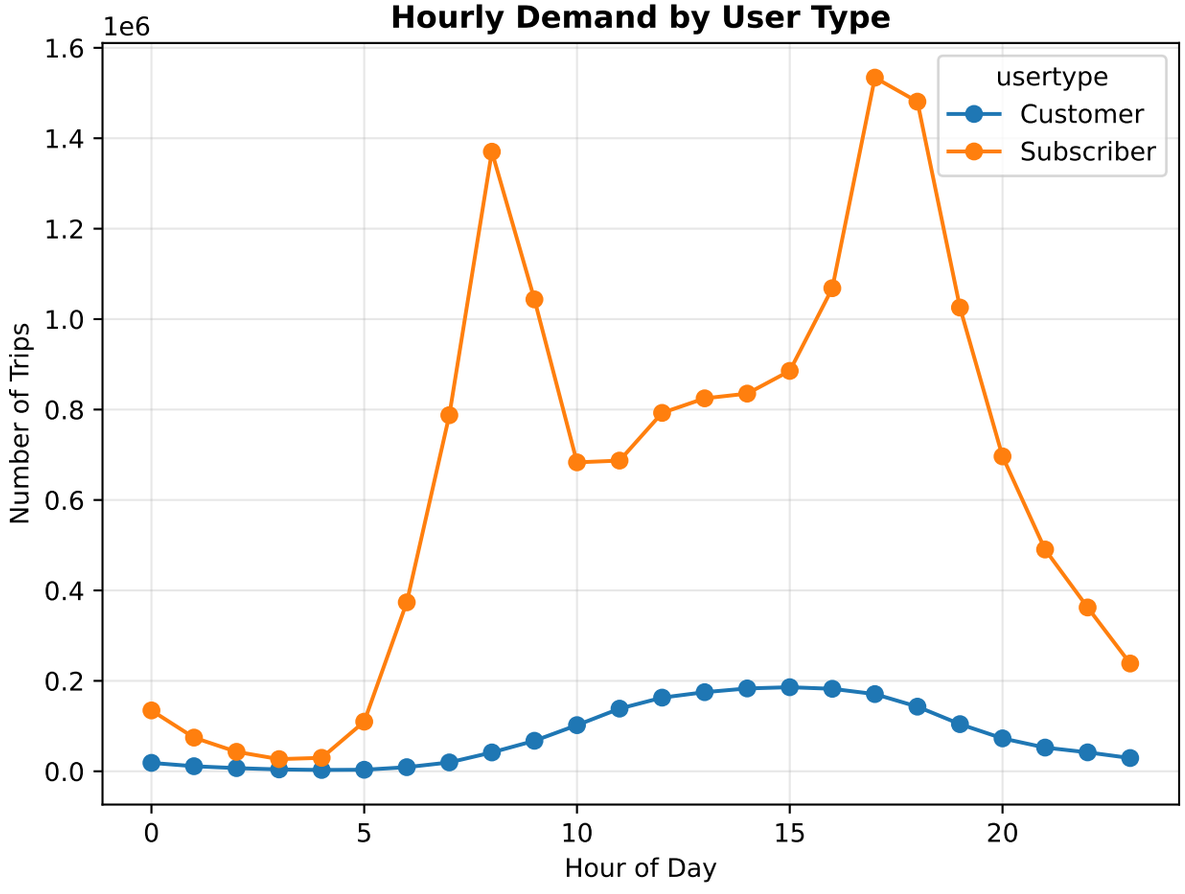

Commute Pattern Distribution
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Trips
3,330,180
(19.0%)
4,458,497
(25.4%)
9,739,206
(55.6%)
Morning Rush
(7–9 AM)
Evening Rush
(5–7 PM)
Non-Commute
(Other Hours)

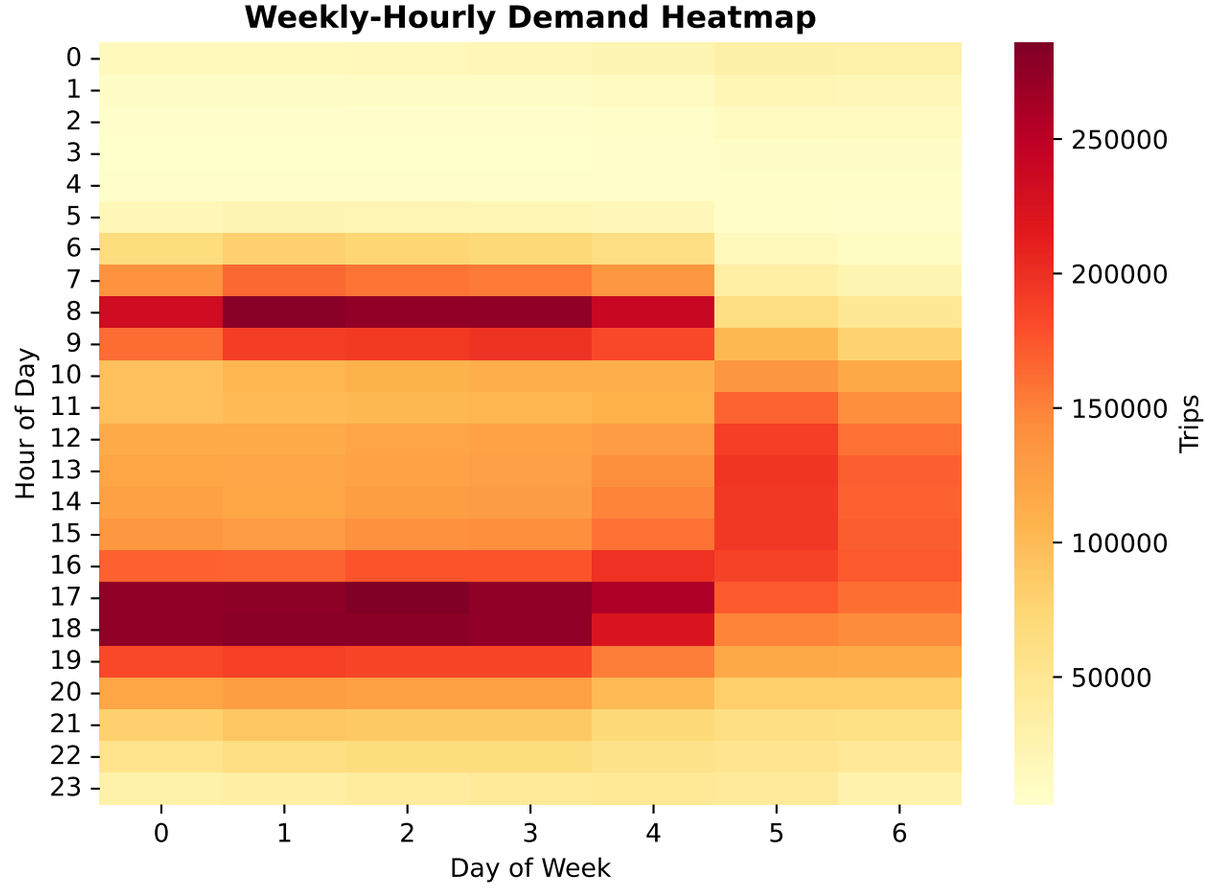

Hierarchical Clustering Dendrogram
Distance
User Sample Index
0
2
4
6
8
10
12
14

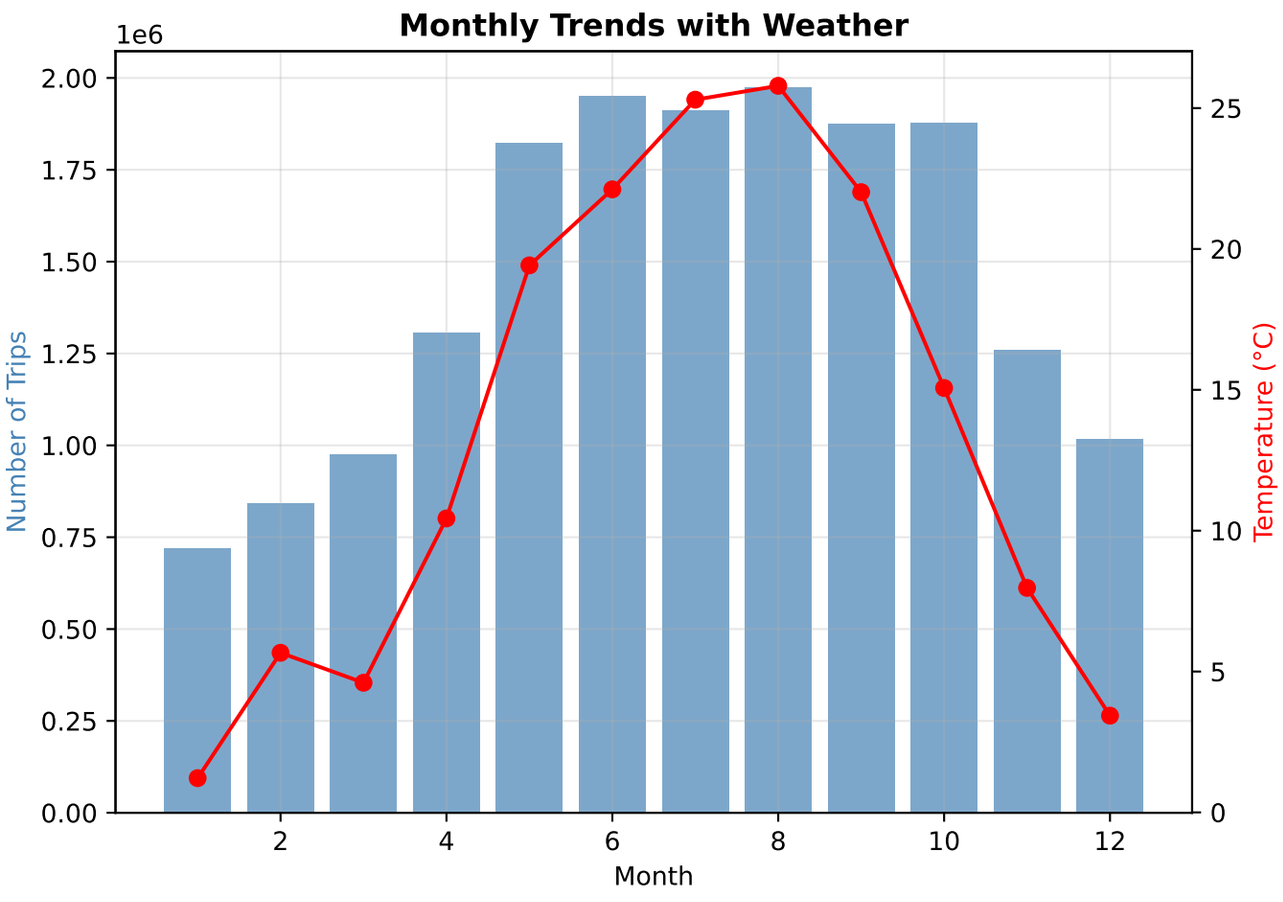

Cluster Evaluation Metrics
Davies-Bouldin Score
Silhouette Score
Number of Clusters
0.20
0.25
0.30
0.35
0.40
0.50
0.55
0.60
0.65
0.70
3
4
5
6
7
8

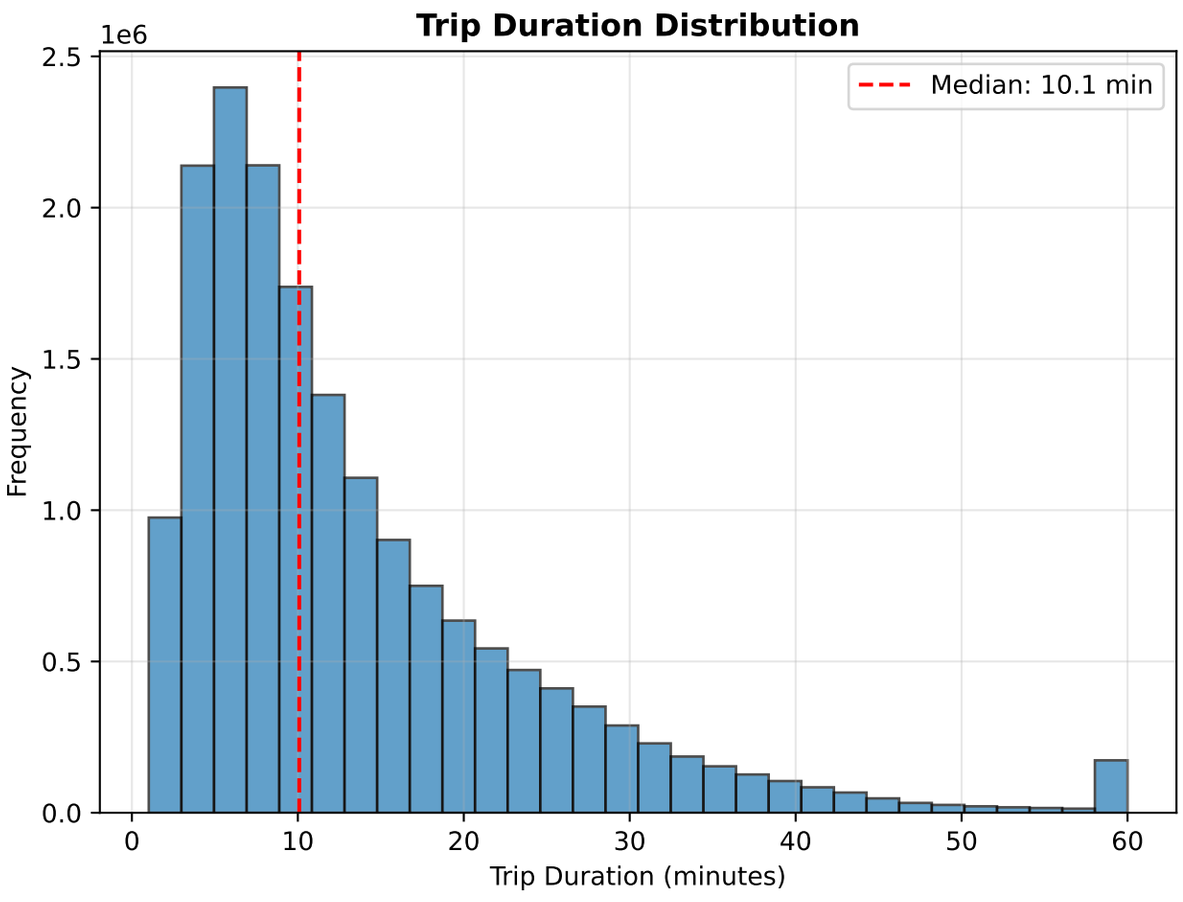

Clusters in PCA Space
PC1 (30.6% var)
PC2 (25.7% var)
Cluster
−10
−8
−6
−4
−2
0
2
4
6
8
10
1.0
1.5
2.0
2.5
3.0
3.5
4.0

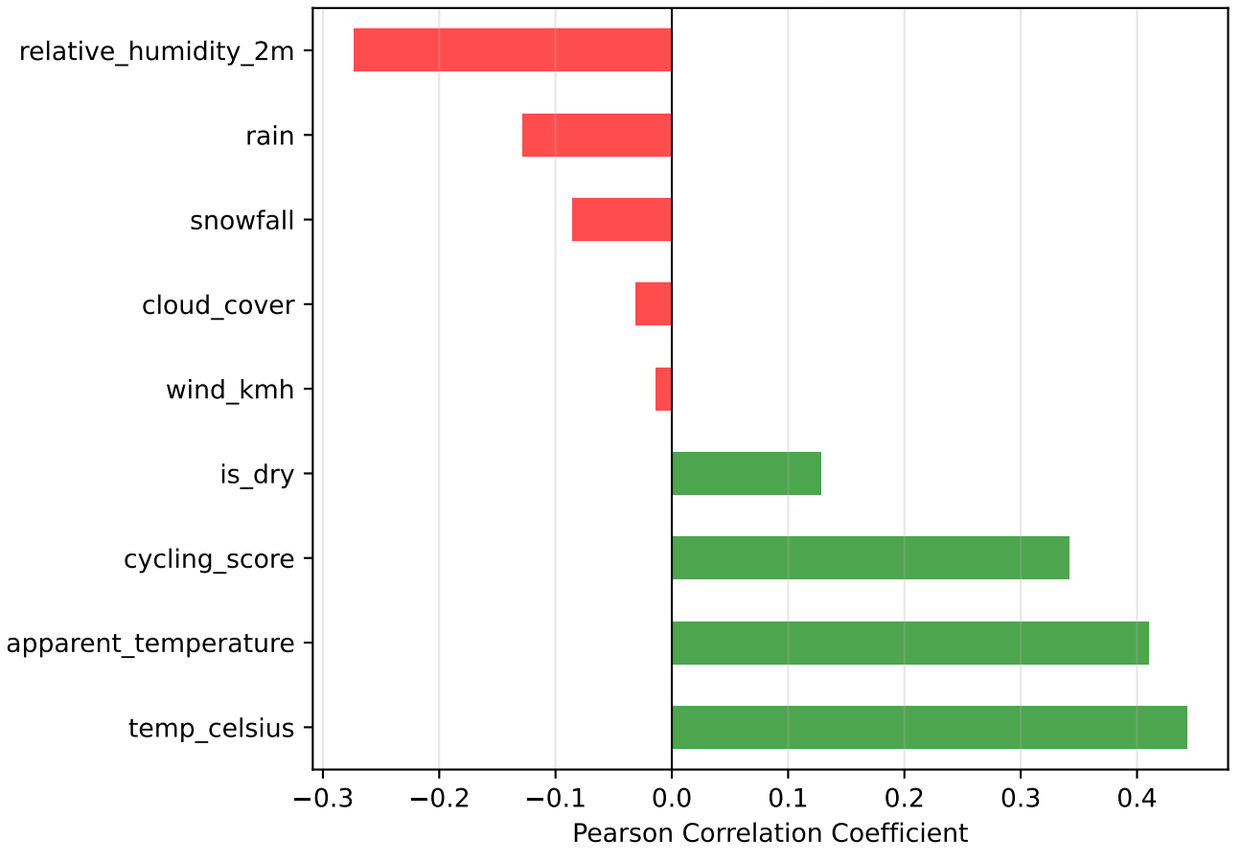

User Distribution
Number of Users
Cluster
209
(98.6%)
1
(0.5%)
1
(0.5%)
1
(0.5%)
0
25
50
75
100
125
150
175
200
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5

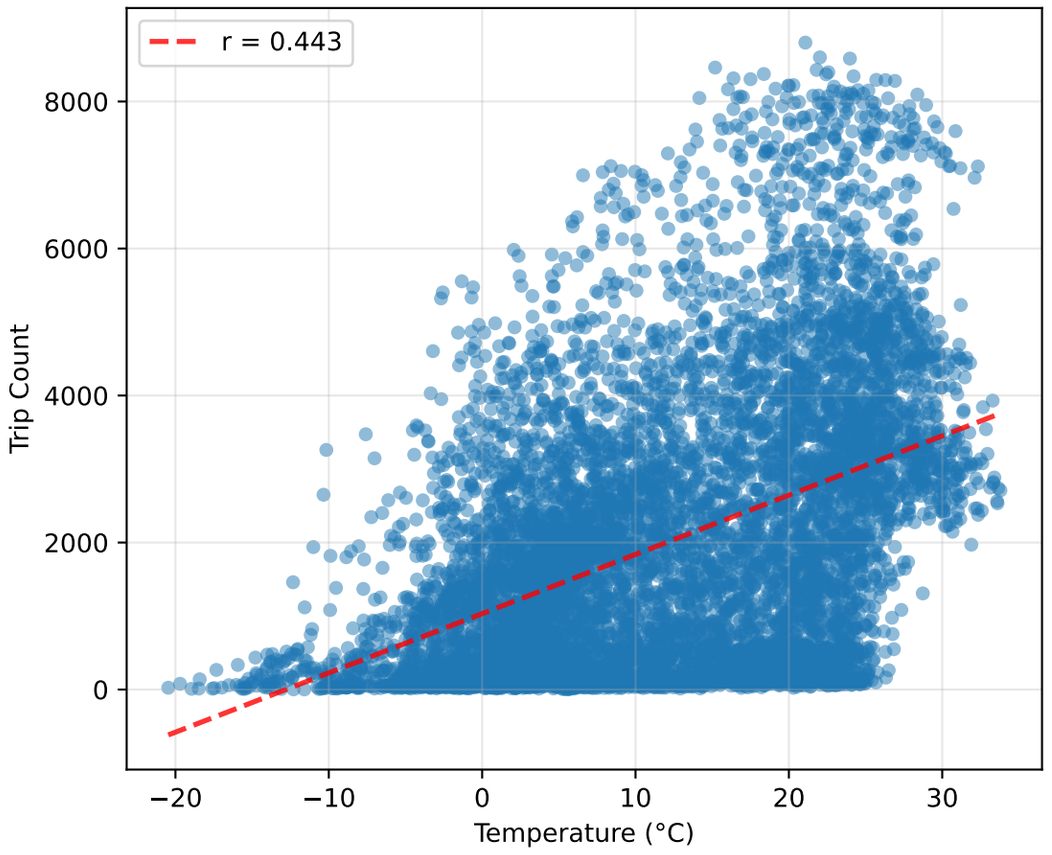

Hourly Demand by User Type
1e6
usertype
Customer
Subscriber
Number of Trips
Hour of Day

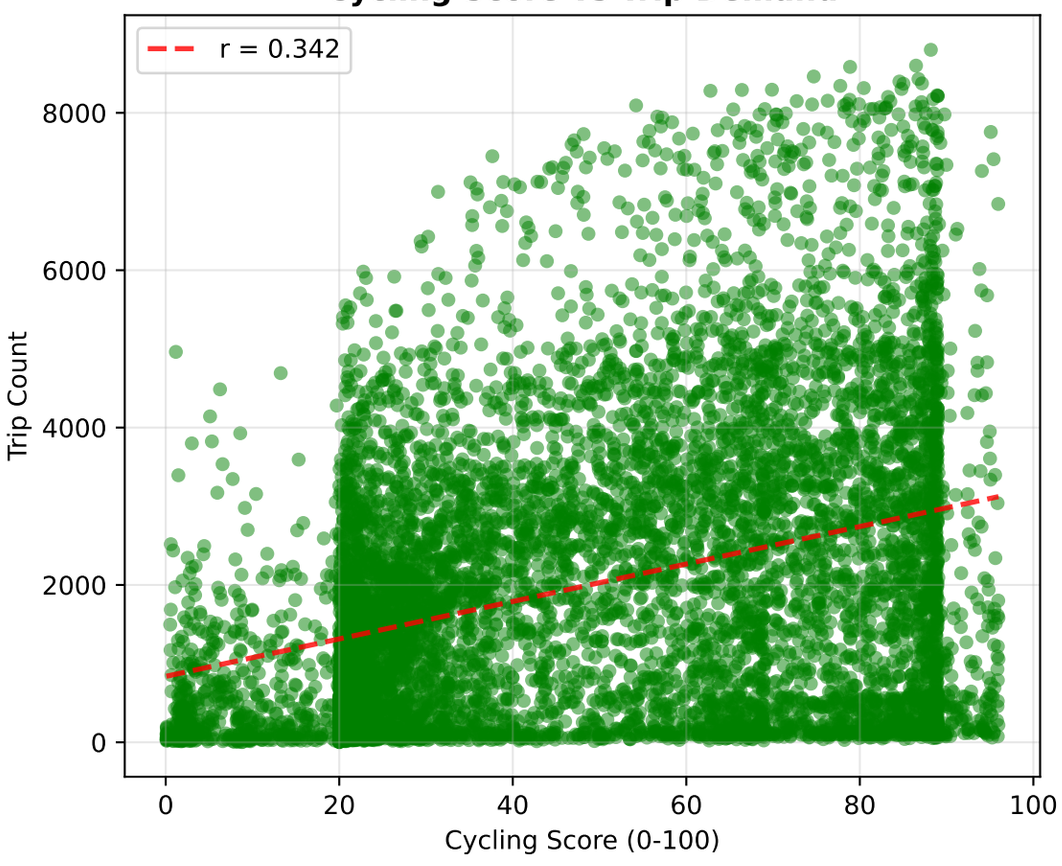

Weekly-Hourly Demand Heatmap
Hour of Day
Day of Week
Trips

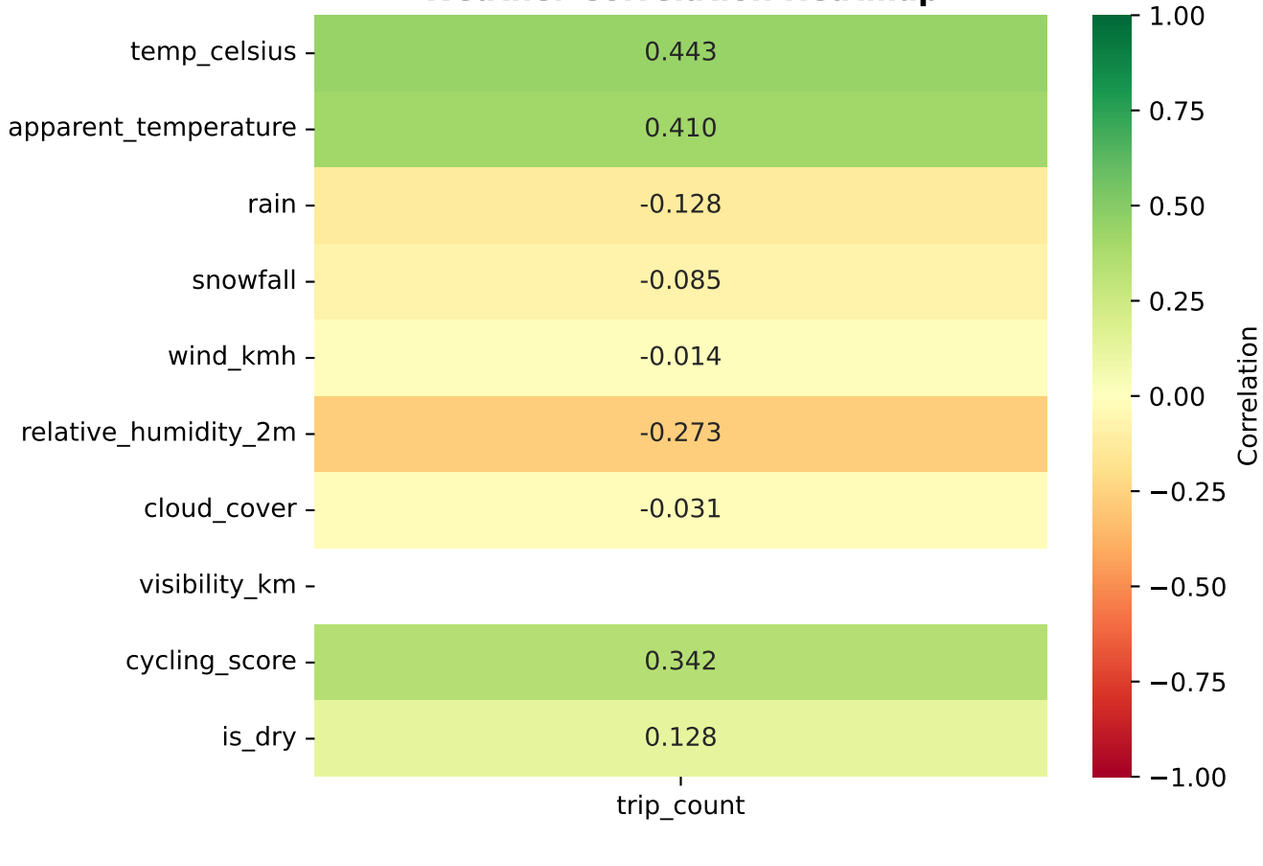

Monthly Trends with Weather
1e6
Number of Trips
Temperature (°C)
Month


Trip Duration Distribution
1e6
Median: 10.1 min
Frequency
Trip Duration (minutes)


Weather Variables vs Trip Count Correlation
relative_humidity_2m
rain
snowfall
cloud_cover
wind_kmh
is_dry
cycling_score
apparent_temperature
temp_celsius
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
Pearson Correlation Coefficient


Temperature vs Trip Demand
r = 0.443
Trip Count
8000
6000
4000
2000
0
−20
−10
0
10
20
30
Temperature (°C)


Cycling Score vs Trip Demand
r = 0.342
Trip Count
8000
6000
4000
2000
0
0
20
40
60
80
100
Cycling Score (0-100)


Weather Correlation Heatmap
temp_celsius
apparent_temperature
rain
snowfall
wind_kmh
relative_humidity_2m
cloud_cover
visibility_km
cycling_score
is_dry
0.443
0.410
-0.128
-0.085
-0.014
-0.273
-0.031
0.342
0.128
trip_count
Correlation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00