Presentation

INTELLIGENT TRAFFIC ANALYTICS IN BISHKEK: DATA, MODELS, AND PRACTICAL SOLUTIONS FOR URBAN MOBILITY IN KYRGYZSTAN

COMPREHENSIVE ML STRATEGY:
FROM DATA CLEANING TO STRATEGIC INFRASTRUCTURE PLANNING.

Manasova Gulum MATDAIS23
Totoeva Aidai MATDAIS23

PROBLEM STATEMENT AND PROJECT GOAL

The Problem: Growing traffic congestion and a lack of proactive, data-driven tools for traffic management in Bishkek.

The Goal: To build a reproducible, Machine Learning-based framework to understand, predict, and optimize the city's road network.

DATA PREPARATION AND INITIAL INSIGHTS

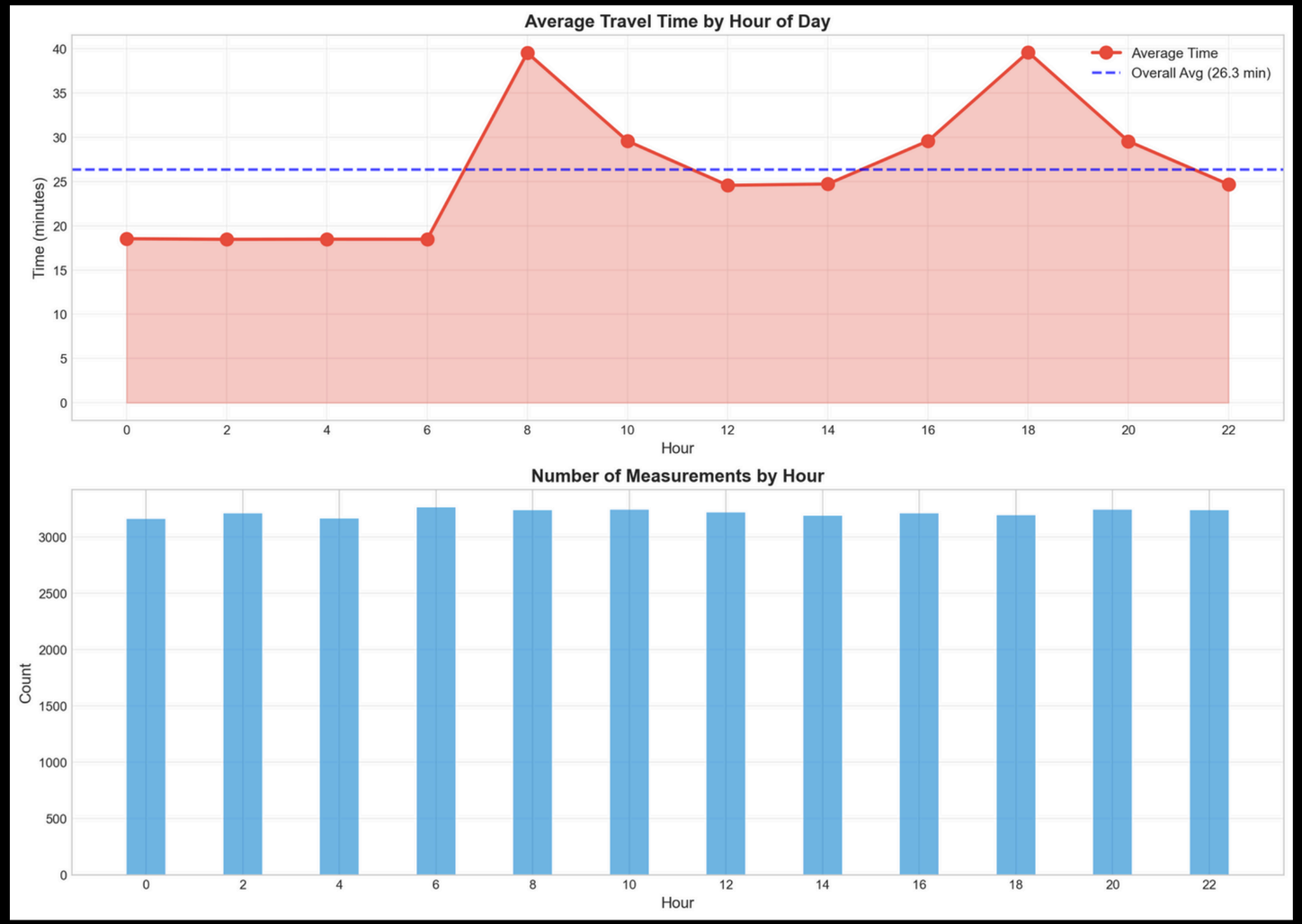
Data Cleaning: Merging, type conversion, and feature creation (hour, speed_cat).

In [65]:
```
# Load clean data
df = pd.read_csv('clean_traffic_data.csv')
df['datetime'] = pd.to_datetime(df['datetime'])
df['date'] = pd.to_datetime(df['date'])
print(df)
```

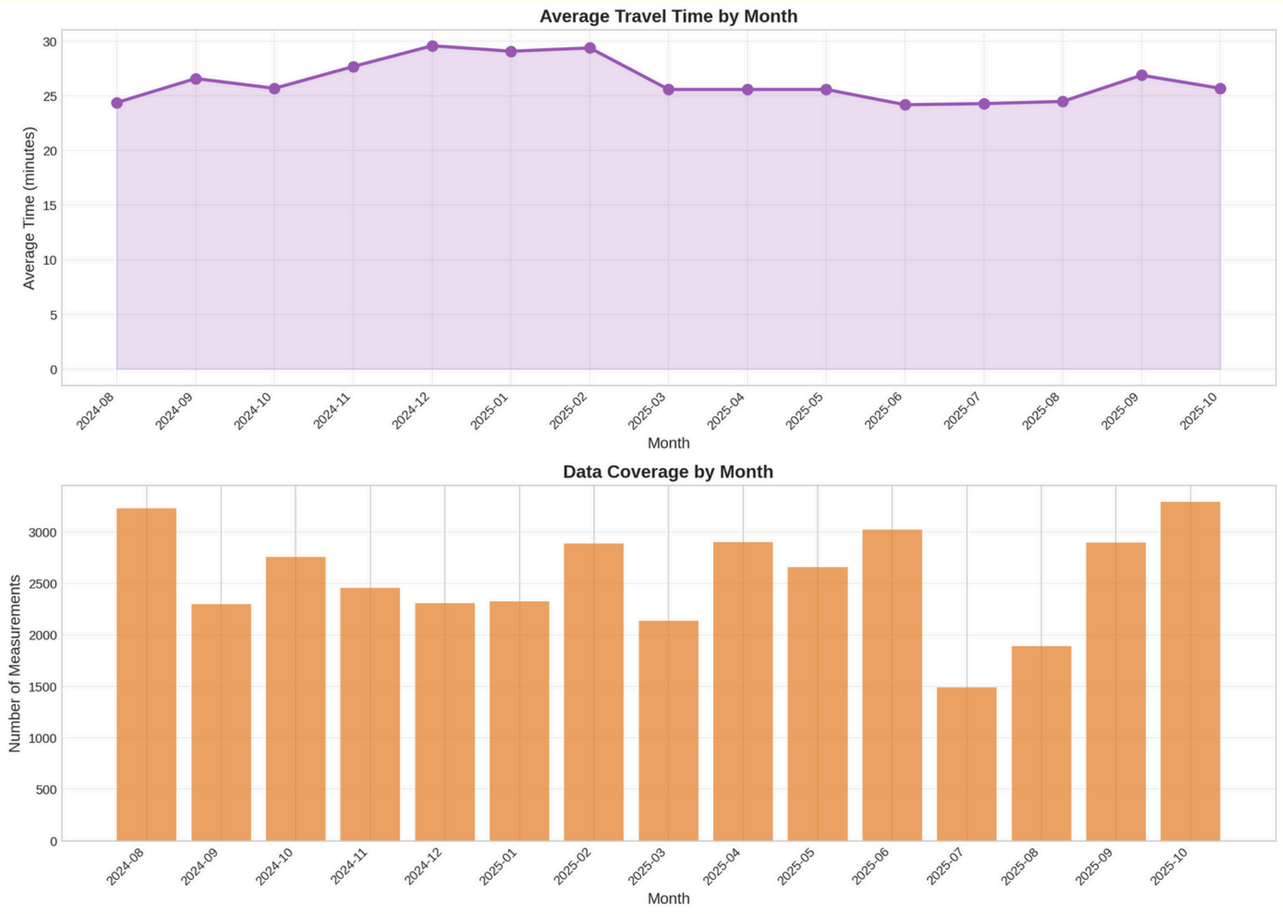
```
            date            datetime             route_name  distance_m  \
0     2024-08-01 2024-08-01 00:00:00           12_mkr-TSUM        8450
1     2024-08-01 2024-08-01 00:00:00     Ala_Too_Sq-Asanbai        7956
2     2024-08-01 2024-08-01 00:00:00  Ak_Orgo-Hyatt_Regency       13119
3     2024-08-01 2024-08-01 00:00:00             PVT-Globus        9234
4     2024-08-01 2024-08-01 00:00:00  Yuzhnye_Vorota-Dordoi       12761
...          ...                 ...                    ...         ...
38537 2025-10-31 2025-10-31 22:00:00  Yuzhnye_Vorota-Dordoi       12761
38538 2025-10-31 2025-10-31 22:00:00        Dzhal-Osh_Bazar        6848
38539 2025-10-31 2025-10-31 22:00:00   Dordoi-Togolok_Moldo       11543
38540 2025-10-31 2025-10-31 22:00:00   Tunguch-Philharmonia       10660
38541 2025-10-31 2025-10-31 22:00:00           Vefa-Ortosay        8912

       duration_sec  duration_min  avg_speed_kmh  hour  day_of_week  month  \
0            936.63         15.61          32.48     0            3      8
1            787.66         13.13          36.36     0            3      8
2           1853.10         30.88          25.49     0            3      8
3            901.86         15.03          36.86     0            3      8
4           1516.77         25.28          30.29     0            3      8
...             ...           ...            ...   ...          ...    ...
38537       2189.70         36.50          20.98    22            4     10
38538        859.85         14.33          28.67    22            4     10
38539       2192.44         36.54          18.95    22            4     10
38540       1465.95         24.43          26.18    22            4     10
38541       1418.17         23.64          22.62    22            4     10

       year  is_weekend time_period  rush_hour  season month_name speed_cat
0      2024       False       Night      False  Summer     August    Normal
1      2024       False       Night      False  Summer     August      Fast
2      2024       False       Night      False  Summer     August    Normal
3      2024       False       Night      False  Summer     August      Fast
4      2024       False       Night      False  Summer     August    Normal
...     ...         ...         ...        ...     ...        ...       ...
38537  2025       False     Evening      False  Autumn    October      Slow
38538  2025       False     Evening      False  Autumn    October    Normal
38539  2025       False     Evening      False  Autumn    October      Slow
38540  2025       False     Evening      False  Autumn    October    Normal
38541  2025       False     Evening      False  Autumn    October      Slow

[38542 rows x 17 columns]
```

DATA PREPARATION AND INITIAL INSIGHTS

Core Finding (EDA): Peak Hours are clearly defined: 8:00–9:00 AM and 6:00–7:00 PM

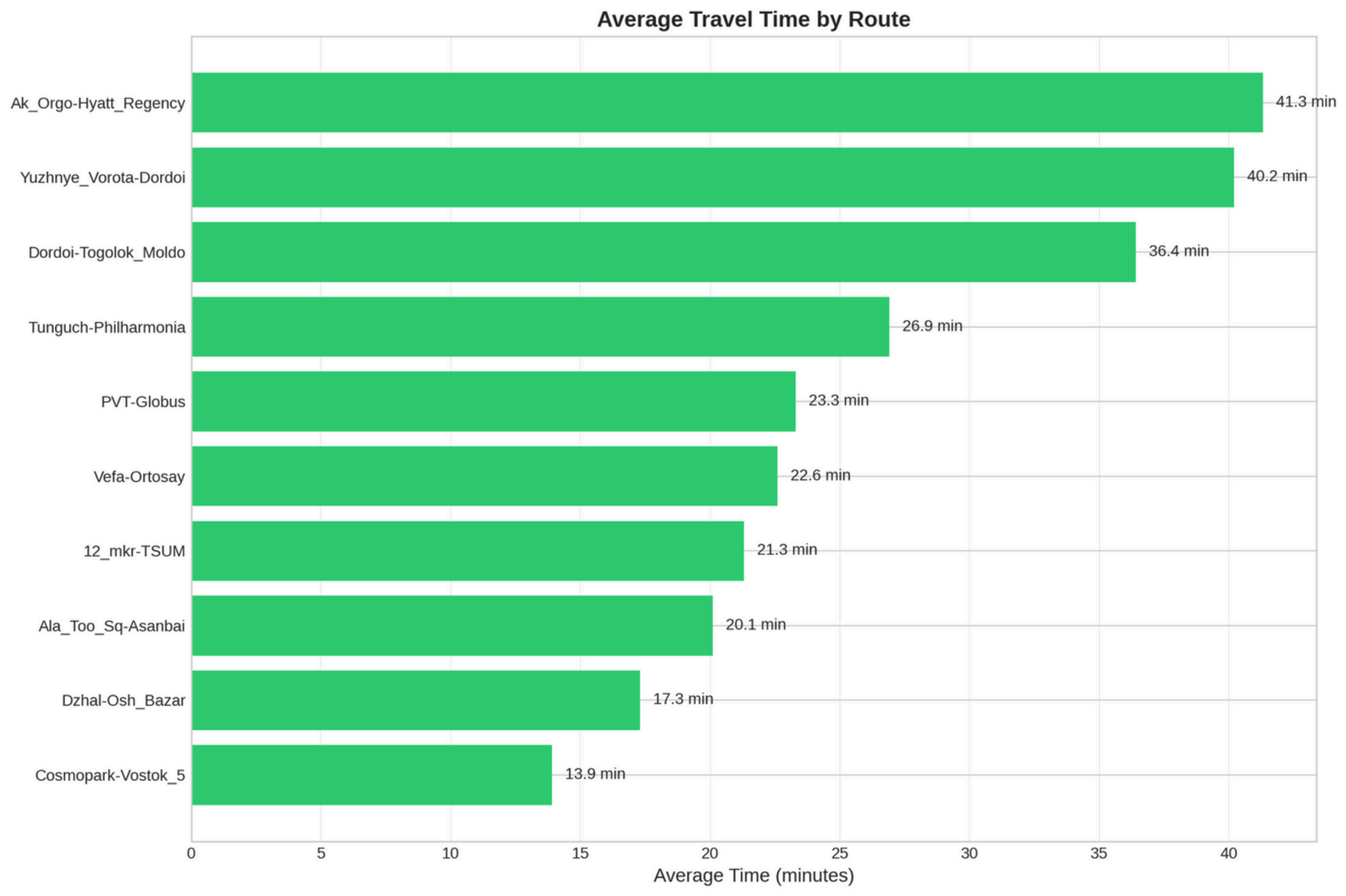


DATA PREPARATION AND INITIAL INSIGHTS



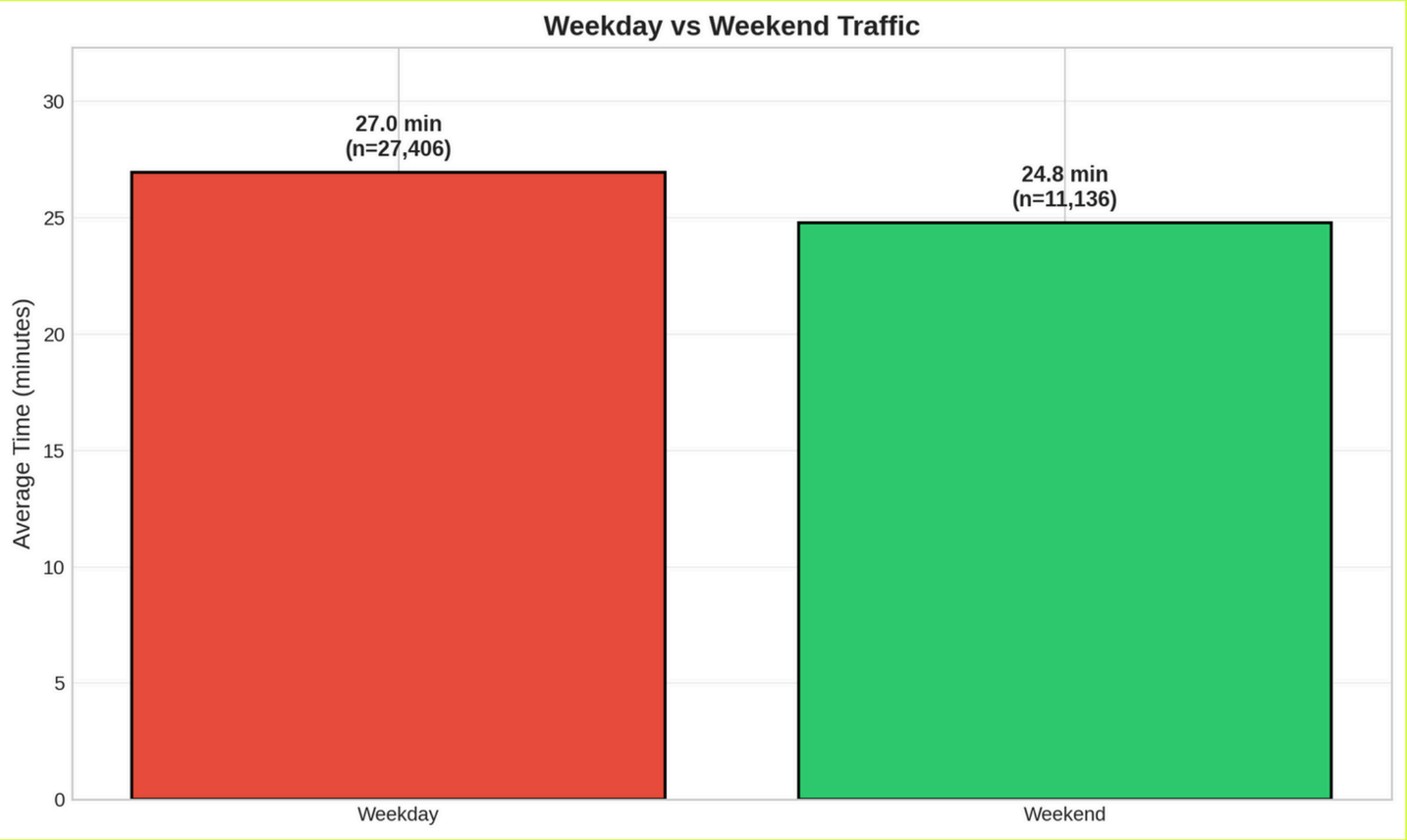
15 months of data. Winter months higher - slower because snow. Summer lower - faster traffic.

DATA PREPARATION AND INITIAL INSIGHTS



Slowest: Ak_Orgo-Hyatt_Regency, 41 minutes.
Fastest: Cosmopark-Vostok, 14 minutes. City should fix slow routes first.

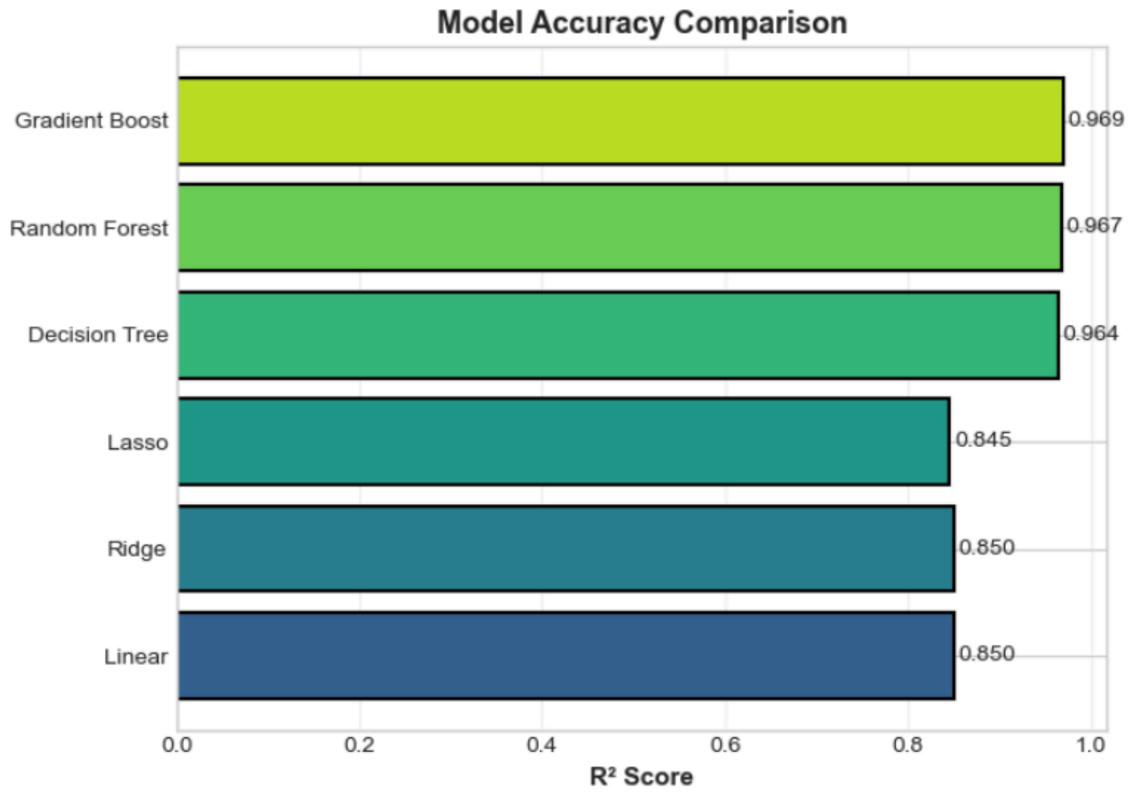
DATA PREPARATION AND INITIAL INSIGHTS

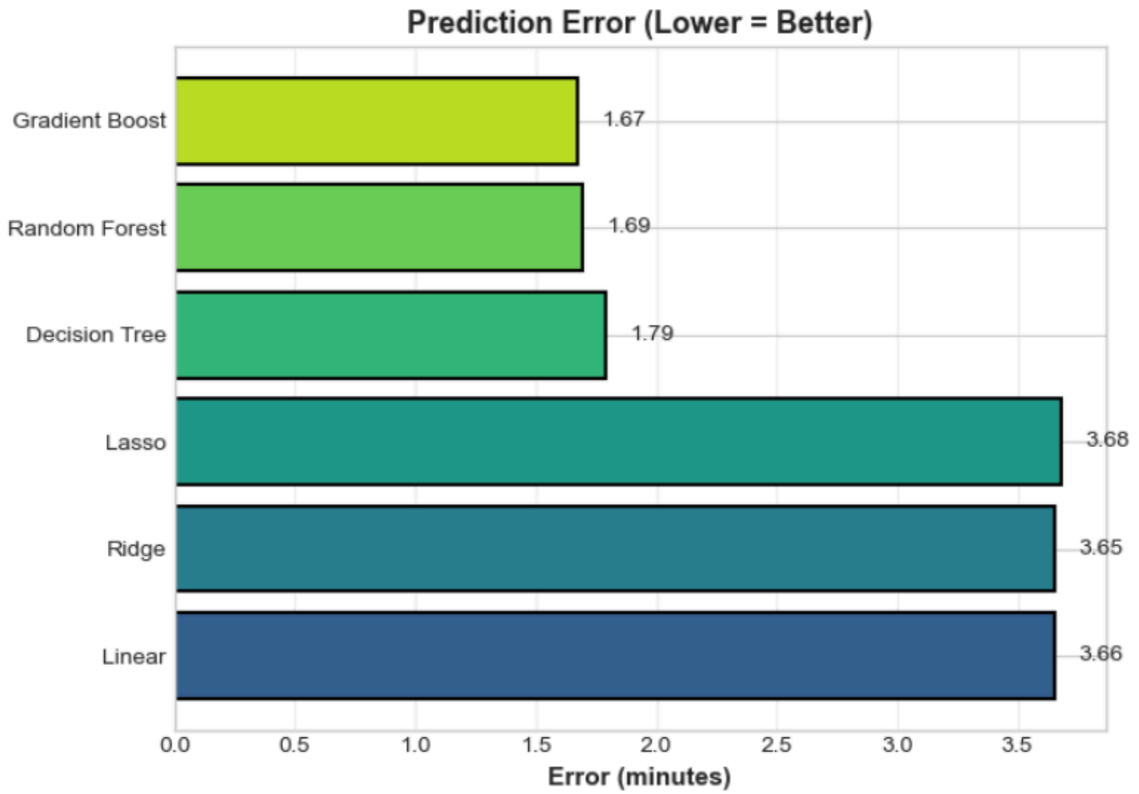


Weekdays 27 min, weekends 24.8 min. Weekends 9% faster - less traffic.

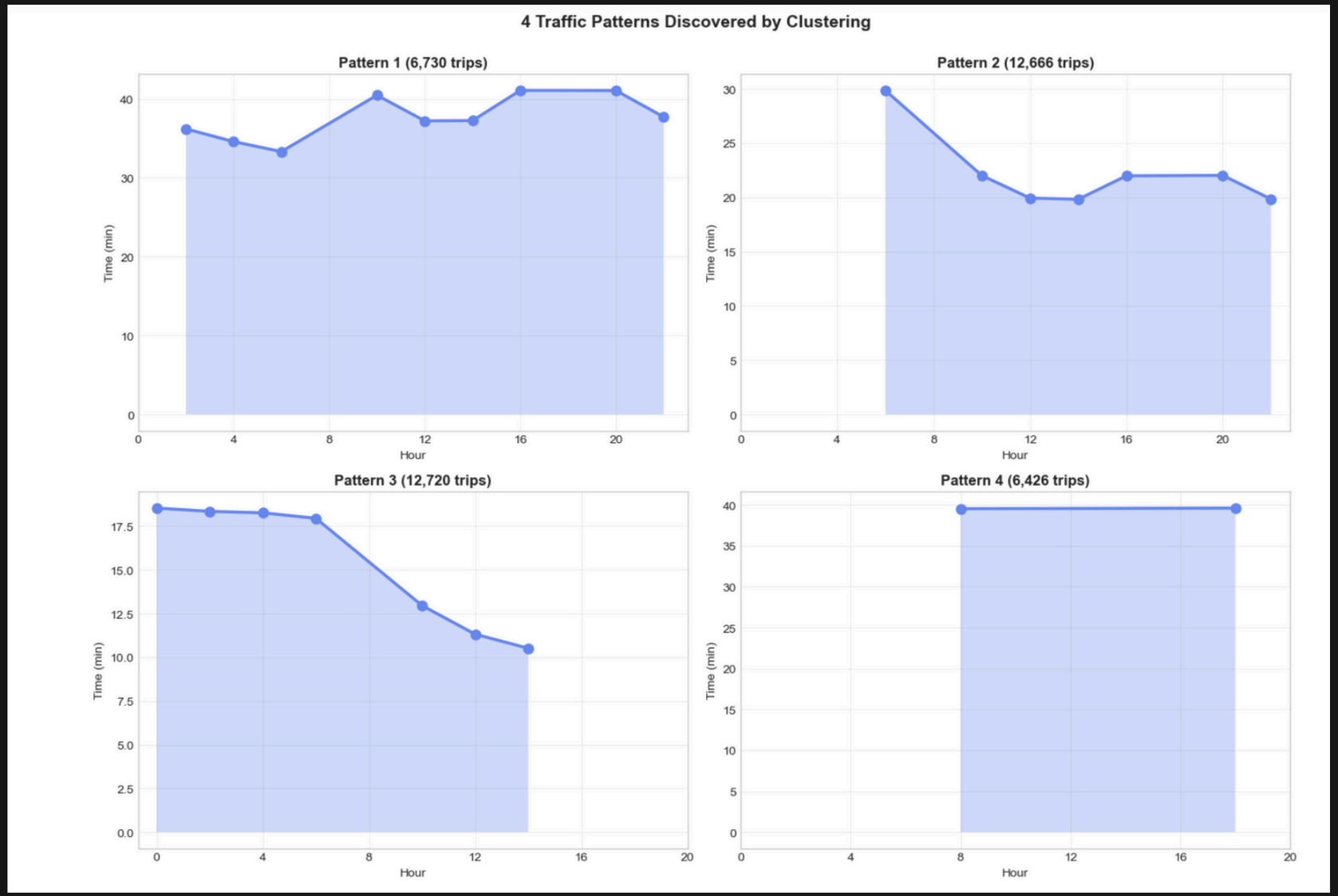
MODEL TRAINING AND EVALUATION

Tested 6 models. Simple models: 85% accuracy.
Advanced models: 96%.
Winner: Gradient Boost - 96.9% with only 1.67 min error!



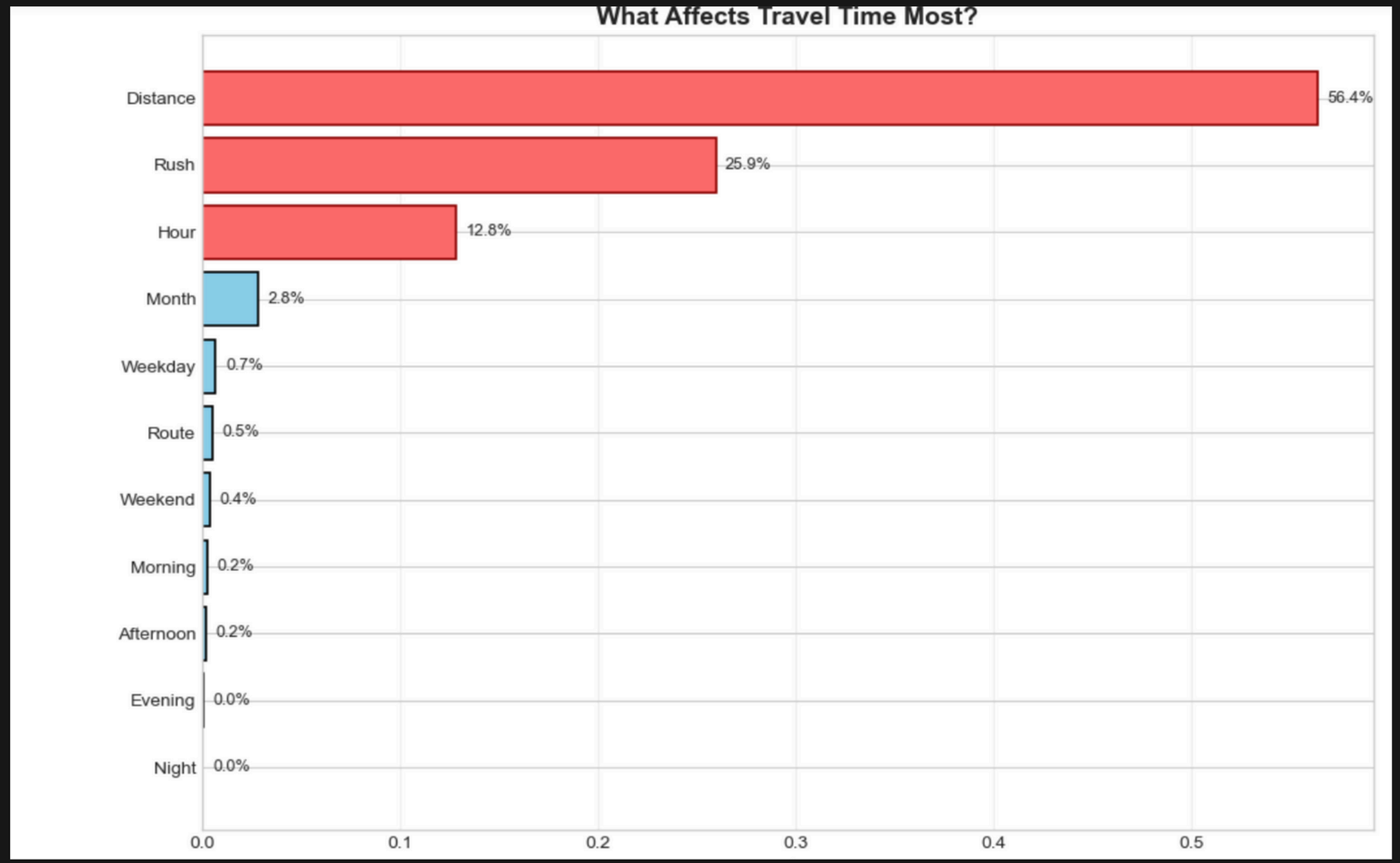


MODEL TRAINING AND EVALUATION



Clustering found 4 patterns. Pattern 3: night traffic, very fast.
Pattern 4: rush hour, always slow.
Time creates different patterns!

MODEL TRAINING AND EVALUATION

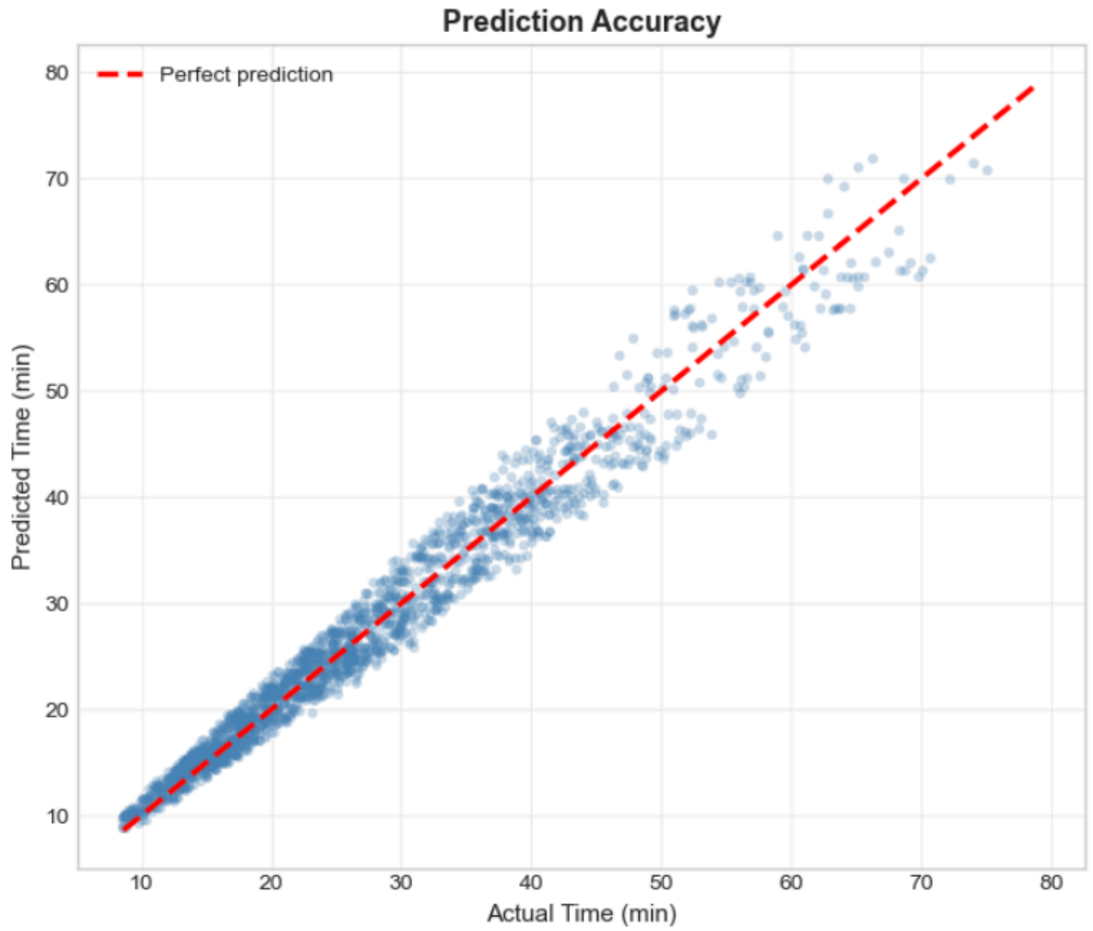


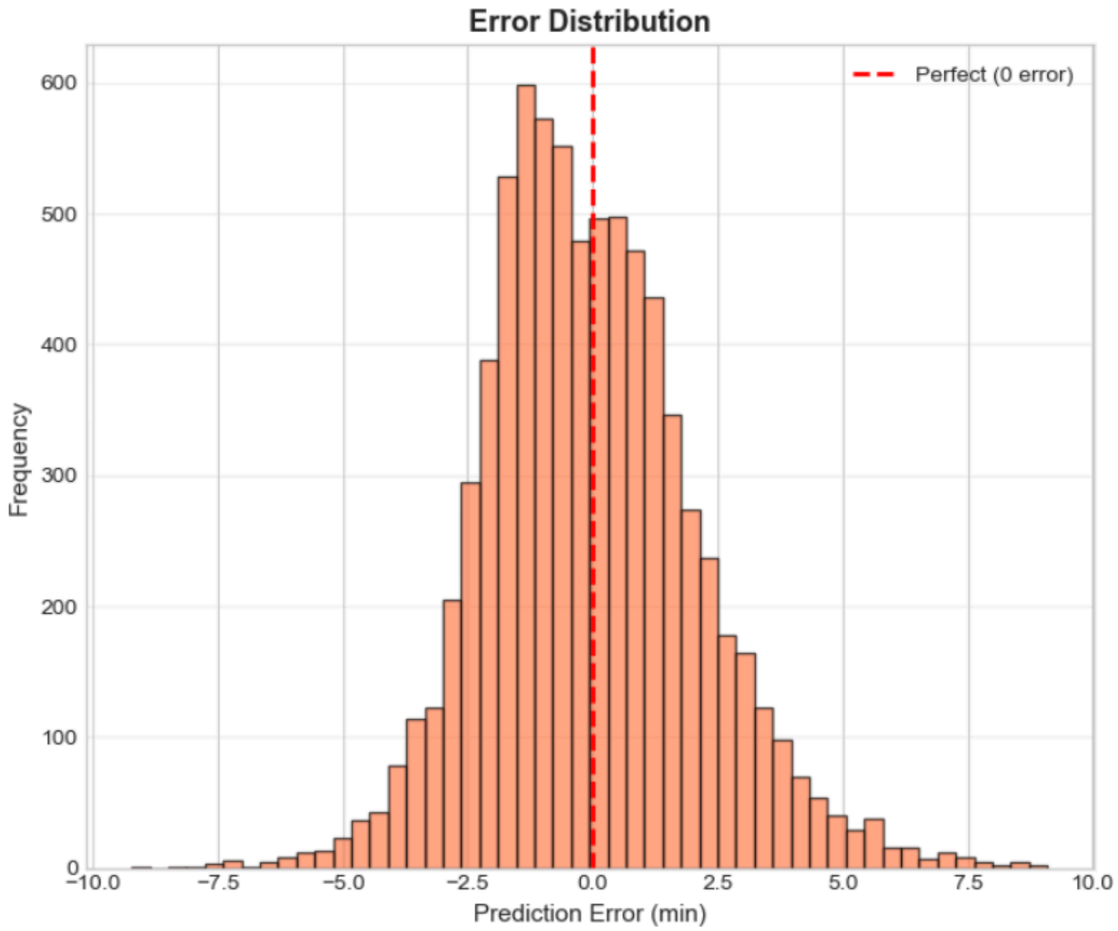
What matters most?
Distance 56% - can't change.
Rush hour 26% - CAN avoid! This is the key finding

MODEL TRAINING AND EVALUATION

Left: dots near line = accurate. Right: most errors near zero.

Result: 96.9% accuracy, 1.7 min error.





SUMMARY

Analyzed 38,542 trips, tested 6 models, 96.9% accuracy.
Key finding: Rush hour is 26% of problem - we can avoid it!
Practical use: Help city and citizens make better decisions.

38,542 - total trips

15 - months

10 - routes

6 - models tested

96.9% - best accuracy

±1.7 min - error

26% - rush hour importance

8 AM, 6 PM - rush hours