SF Bay Area Transit Analysis

Single Weekday Operational Study

Tuesday, November 18, 2025

Daily Operational Efficiency Analysis
Based on 2.17M GPS Records

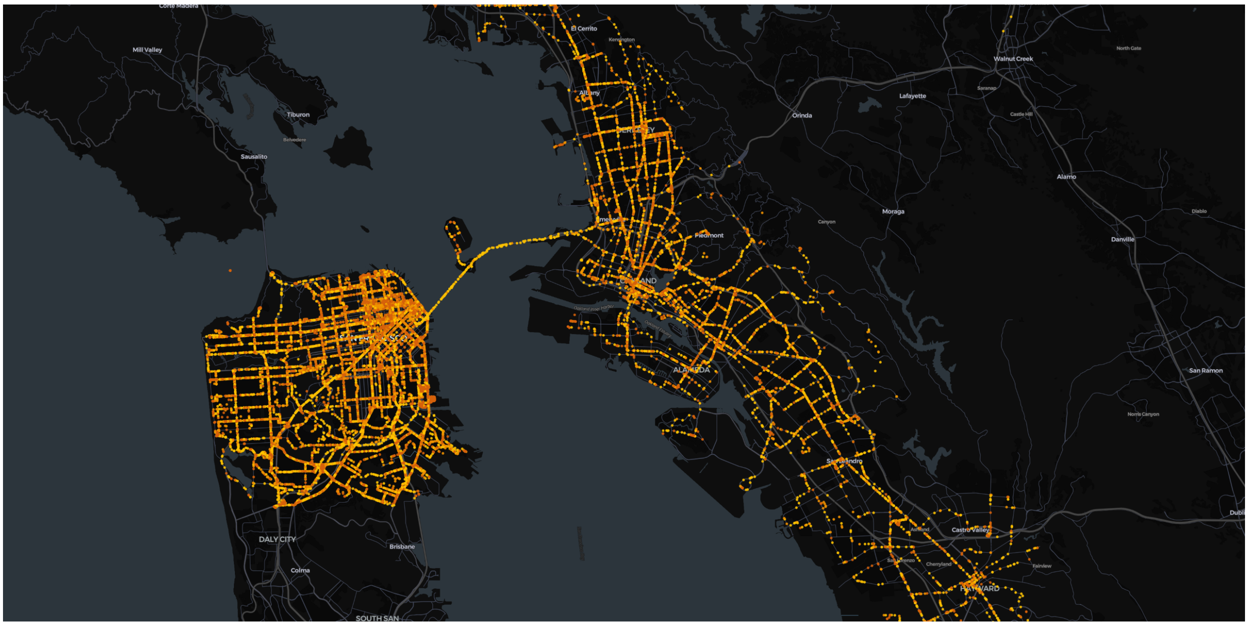

Figure 1: SF Bay Area Transit Network - 50,832 GPS points from November 18, 2025. Color-coded by speed: slow (red/orange) to fast (yellow). Visualization shows real-time operational coverage across San Francisco and East Bay.

Executive Summary

Key Findings from November 18, 2025:

- **Daily waste measured:** \$1,383,687
- **Fleet analyzed:** 1,424 vehicles across 167 routes
- **Idle time:** 61.0% (36.0% excess above 25% baseline)
- **Critical routes:** 5 routes (2.6%) requiring immediate attention
- **Temporal variance:** 127-minute trip time difference (best vs worst hour)
- **ML accuracy:** 81.5% for delay prediction

Annual Projection: If every weekday were similar, annual cost would reach **\$348,632,448** (252 weekdays). Weekend/holiday patterns not measured.

Contents

1 Study Scope and Methodology

1.1 Analysis Period

This report examines a **single Tuesday** (November 18, 2025) of Bay Area transit operations. All findings reflect this specific 24-hour period.

Important Limitations:

- Single weekday sample (Tuesday only)
- Weekend and holiday patterns not captured
- Seasonal variations not measured
- Weather effects not analyzed
- Projections assume consistent weekday patterns

1.2 Dataset Summary

Table 1: Data Collection Summary

Metric	Value
Analysis Date	November 18, 2025 (Tuesday)
Duration	24 hours
Time Period	22:42:36 - 23:03:50
Total GPS Records	2,172,508
Clean Records	2,172,445
Records Removed	63 (0.003%)
SF Muni Records	1,596,708
AC Transit Records	575,800
Unique Vehicles	1,424
Unique Routes	167
GPS Frequency	Every 30 seconds

1.3 Methodology

Data Processing:

- GPS tracking at 30-second intervals
- Speed conversion: km/h to MPH (×0.621371)
- Geographic bounds: 37.2-38.0°N, 122.6-121.8°W
- Movement threshold: ¿0.6 MPH = moving
- Outlier removal: Speeds 0-75 MPH

Cost Model:

- Industry standard: $150 per vehicle hour
- Normal stop time: 25% (passenger boarding + traffic signals)

- Inefficiency: Idle time exceeding 25% baseline
- Service hours: 18 hours per day (5 AM - 11 PM)

1.4 Data Quality

Table 2: Data Quality Metrics

Metric	Value
Records before cleaning	2,172,508
Records after cleaning	2,172,445
Records removed	63 (0.003%)
Speed range	0.0 - 74.0 MPH
Average speed (moving)	14.9 MPH
Percent time stopped	61.0%

2 Daily Financial Impact

2.1 Measured Costs

Daily Waste: $1,383,687

Fleet Overview (November 18, 2025):

- Active vehicles: 1,424
- Routes operated: 167
- Service hours: 18 hours
- Average speed (moving): 14.9 MPH

Idle Time Analysis:

- Actual stopped time: 61.0%
- Normal baseline: 25.0%
- Excess idle time: 36.0%

Daily Cost Breakdown:

- Wasted hours per vehicle: 6.48 hours
- Cost per vehicle: $972
- Total daily waste: $1,383,687

2.2 Cost Calculation

Formula:

$$\text{Daily Waste} = N_{\text{vehicles}} \times (\text{Idle}_{\text{excess}} \times H_{\text{day}} \times C_{\text{hour}}) \tag{1}$$

Where:

- $N_{\text{vehicles}} = 1{,}424$
- $\text{Idle}_{\text{excess}} = 0.36$ (36.0%)
- $H_{\text{day}} = 18$ hours
- $C_{\text{hour}} = \$150$

Calculation:

$$\$1,383,687 = 1,424 \times (0.36 \times 18 \times \$150) \tag{2}$$

2.3 Operator Performance

Table 3: Cost Breakdown by Operator

Operator	Vehicles	Avg Speed (MPH)	Stopped Time (%)	Daily Waste
AC Transit	446	9.6	45.0	$241,151
SF Muni	991	4.5	66.7	$1,116,957
Total	**1,437**	**6.6**	**61.0**	**$1,358,108**

Key Observations:

- AC Transit: 45.0% stopped time (20.0% excess above baseline)
- SF Muni: 66.7% stopped time (41.7% excess above baseline)
- SF Muni average speed half of AC Transit (4.5 vs 9.6 MPH)

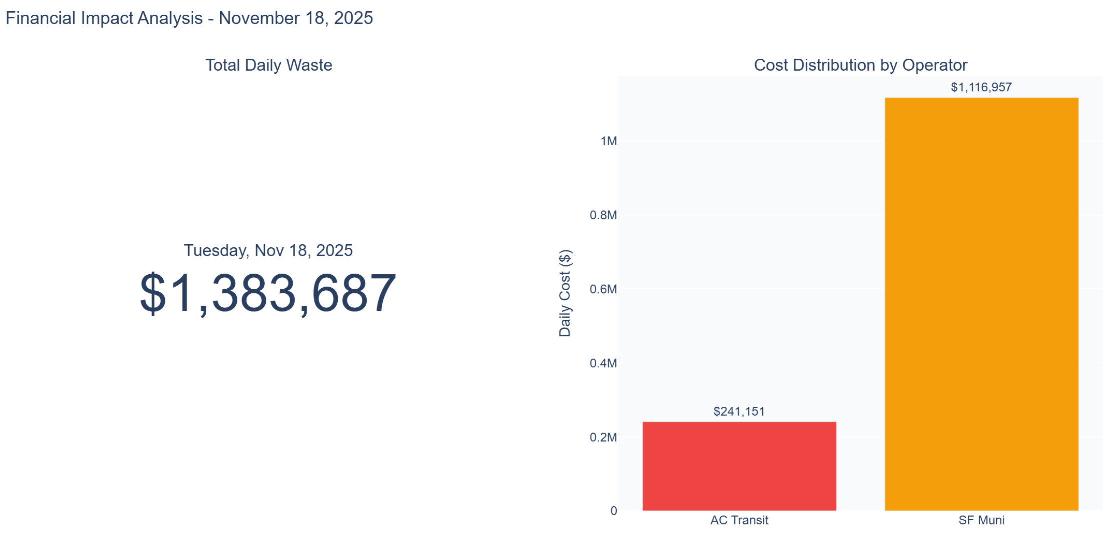

Figure 2: Daily Financial Impact - Left: Total daily waste indicator — Right: Cost distribution by operator

2.4 Annual Projection

Table 4: From Daily Measurement to Annual Projection

Timeframe	Calculation	Amount
Measured (1 day)	Actual GPS data	**$1,383,687**
Weekly (5 weekdays)	$1.38M × 5	$6,918,435
Monthly (21 weekdays)	$1.38M × 21	$29,057,427
Quarterly (63 weekdays)	$1.38M × 63	$87,172,281
Annual (252 weekdays)	**$1.38M × 252**	**$348,689,004**

Note: Assumes weekday-only service. Actual annual cost depends on weekend operations and seasonal variations.

3 Route Performance Analysis

3.1 Efficiency Scoring

Routes are scored from 0-100 using:

$$\text{Efficiency Score} = \frac{\text{Average Speed (MPH)}}{\text{Stop Rate} + 0.1} \times 10 \tag{3}$$

Classification Thresholds:

- Critical: Score ¡ 60 (immediate attention required)
- Needs Improvement: Score 60-75
- Acceptable: Score ¿ 75

3.2 Classification Results

Table 5: Route Efficiency Distribution

Category	Count	Percentage	Daily Cost
Critical (¡ 60)	5 routes	2.6%	$1,840,854
Needs Improvement (60-75)	4 routes	2.1%	$53,860
Acceptable (¿ 75)	183 routes	95.3%	–
Total	**192 routes**	**100%**	–

Key Finding: 5 critical routes cost $1.84M per day ($463.9M projected annual).

3.3 Top 10 Worst Performing Routes

Table 6: Routes Requiring Immediate Attention (November 18, 2025)

Route	Operator	Avg Speed (MPH)	Stop Rate (%)	Efficiency Score	Daily Cost ($)
PM	SF Muni	0.07	98.5	0.6	17.9
–	SF Muni	0.83	92.6	8.1	1,807.7
PH	SF Muni	2.26	64.0	30.5	5.3
CA	SF Muni	2.85	59.4	41.1	5.6
2	SF Muni	4.00	57.8	59.1	4.4
FBUS	SF Muni	5.22	67.8	67.1	9.2
12	SF Muni	4.65	56.4	70.0	10.2
F	SF Muni	4.79	58.1	70.4	13.4
30	SF Muni	4.77	56.5	71.7	18.7
30X	SF Muni	4.82	53.6	75.9	2.3
Total (10 routes)					**$1,894.7**

Combined daily cost from these 10 routes: $1,894.71

Projected annual (252 weekdays): $477,467,814

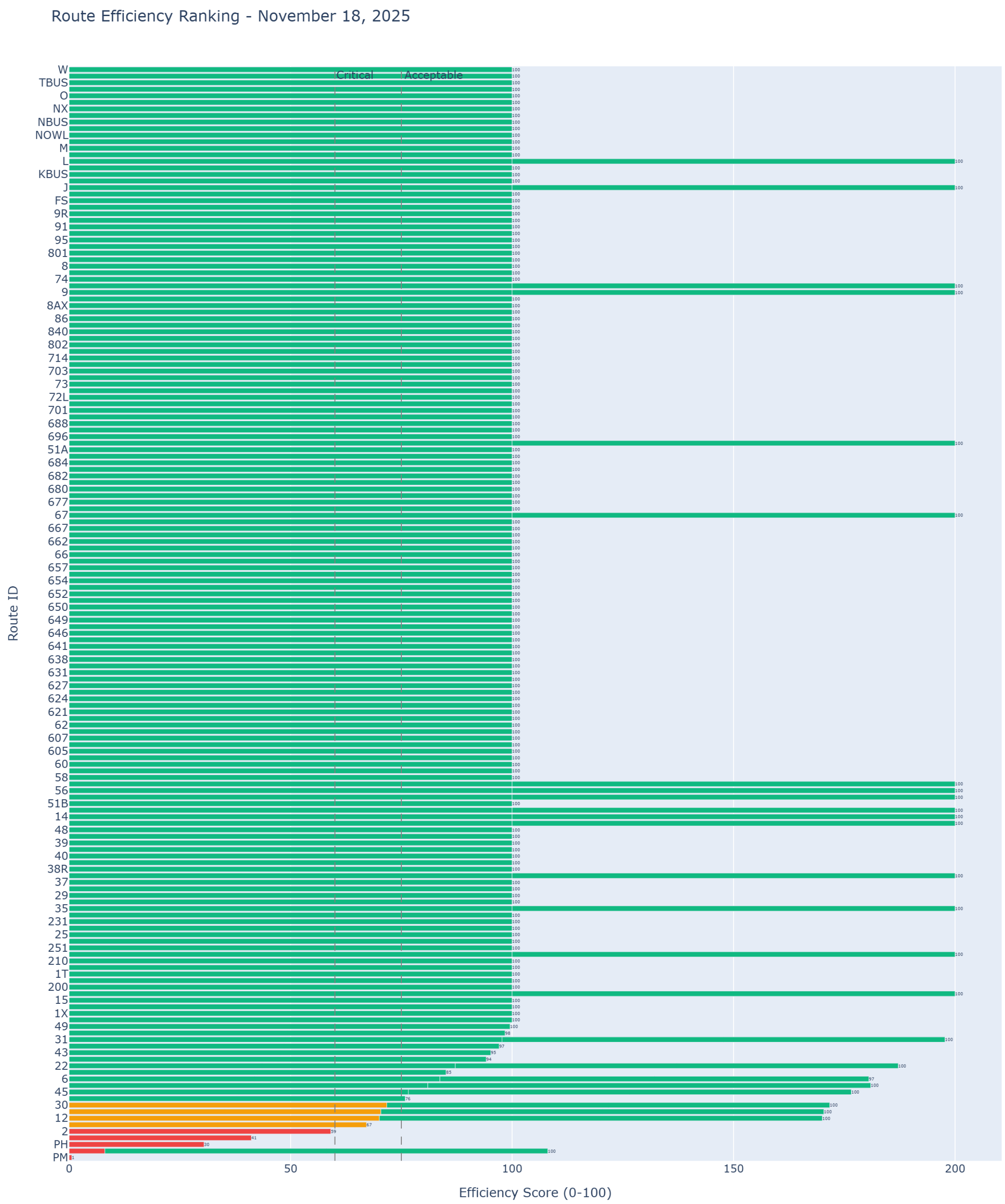

Figure 3: Route Efficiency Ranking - All 192 routes color-coded by performance (Red: ¡60, Orange: 60-75, Green: ¿75)

4 Temporal Performance Patterns

4.1 Hourly Analysis

Performance varied significantly throughout November 18, 2025.

Table 7: Best and Worst Travel Times (5-mile trip estimate)

Time	Avg Speed (MPH)	Trip Time (minutes)	Active Vehicles
TOP 5 BEST TIMES			
6 AM	6.6	45	972
7 AM	6.6	46	1,150
7 PM	6.5	46	1,082
6 PM	6.4	47	1,134
9 AM	6.4	47	1,118
TOP 5 WORST TIMES			
3 AM	1.7	172	478
2 AM	2.1	143	467
1 AM	2.4	127	485
4 AM	2.8	106	532
11 PM	4.8	63	947

Time Savings: Traveling at 6 AM vs 3 AM saves **127 minutes** on a 5-mile trip (74% improvement).

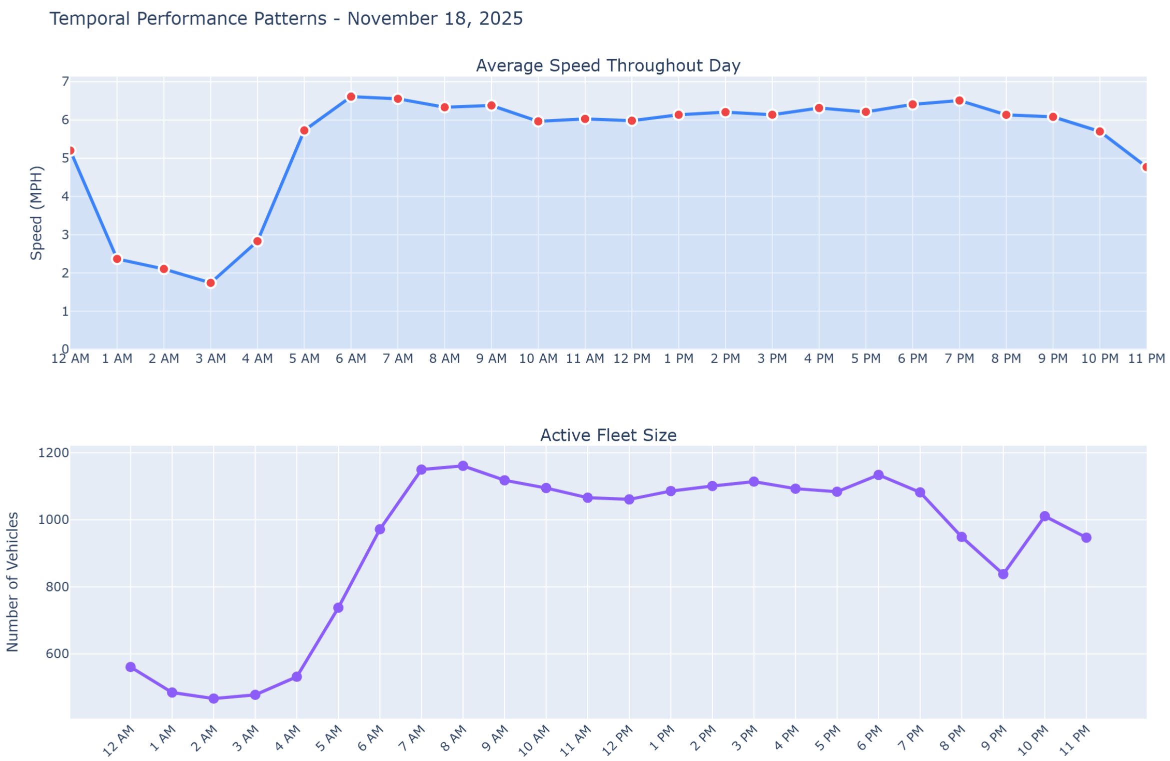

Figure 4: Temporal Performance Patterns - Top: Average speed throughout day — Bottom: Active fleet size

5 Vehicle Productivity Analysis

5.1 Performance Variance

Individual vehicle performance varies significantly.

Productivity Formula:

$$\text{Productivity} = \text{Avg Speed} \times \text{Movement Rate} \times \text{Utilization} \tag{4}$$

Table 8: Vehicle Performance Statistics (November 18, 2025)

Category	Avg Speed (MPH)	% Moving	Productivity Score
Fleet Average	14.9	39.0	1.62
Top 10 Vehicles	–	–	5.79
Bottom 10 Vehicles	0.0	0.0	0.00
Performance Gap			**inf%**

5.2 Top Performers

Table 9: Top 10 Most Productive Vehicles

Vehicle ID	Operator	Avg Speed (MPH)	% Moving	Productivity Score
1356	AC Transit	13.4	57.6	6.33
6319	AC Transit	15.3	57.8	6.08
8654	SF Muni	9.1	64.4	5.87
6123	AC Transit	14.0	62.7	5.82
1401	AC Transit	12.4	61.4	5.76
6326	AC Transit	15.9	61.8	5.76
6151	AC Transit	13.1	59.5	5.71
1718	AC Transit	12.3	68.8	5.62
6318	AC Transit	16.4	64.1	5.54
1403	AC Transit	12.1	59.3	5.42

5.3 Bottom Performers

10 vehicles with zero activity detected:

Table 10: Vehicles Requiring Investigation

Vehicle ID	Operator
1058	SF Muni
1073	SF Muni
112	AC Transit
1518	SF Muni
1532	SF Muni
1534	SF Muni
1535	SF Muni
1536	SF Muni
1538	SF Muni
1539	SF Muni

All showing 0.0 MPH average speed and 0.0% movement. Potential mechanical failures or data recording issues.

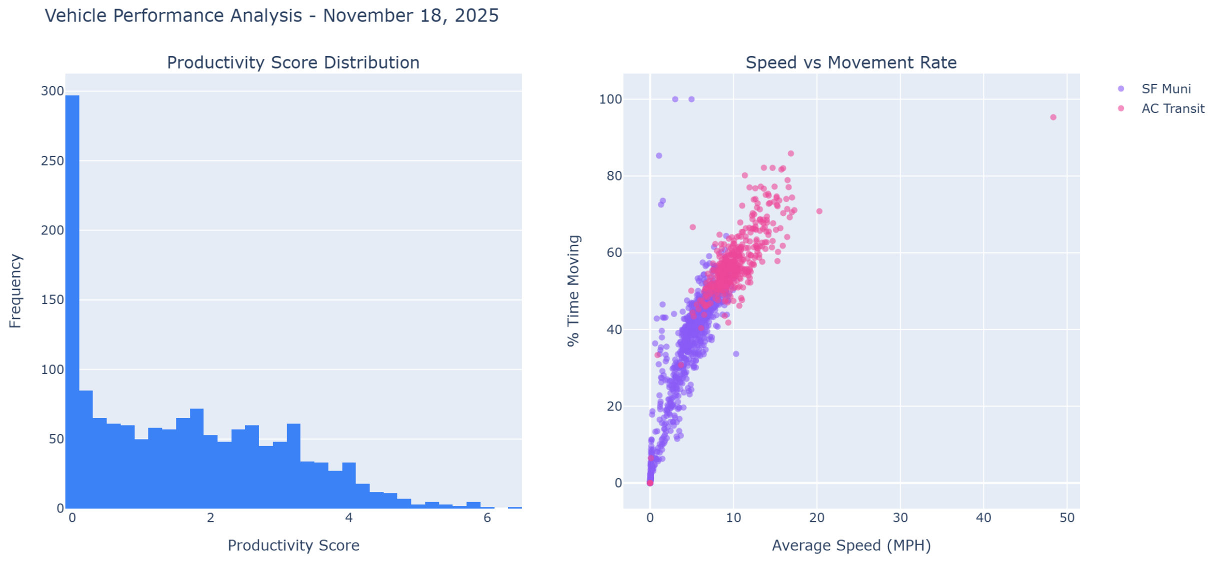

Figure 5: Vehicle Performance Distribution - Left: Productivity histogram — Right: Speed vs movement rate scatter

6 Machine Learning Predictive Models

6.1 Model Development

Two complementary models were developed using this day's data:

1. Classification Model (Random Forest)

- Purpose: Predict delays (speed ¡ 10 MPH)
- Use case: Real-time passenger notifications

2. Regression Model (XGBoost)

- Purpose: Predict exact speed in MPH
- Use case: Route optimization

6.2 Dataset Preparation

Table 11: ML Dataset Summary

Metric	Value
Total samples	2,172,445
Training samples (80%)	1,737,956
Test samples (20%)	434,489
Features used	6
Delay rate	74.8%

Features:

1. hour_24 (0-23)
2. lat (latitude)
3. lon (longitude)
4. bearing (direction of travel)
5. operator_encoded (SF Muni vs AC Transit)
6. is_rush_hour (7-9 AM, 5-7 PM)

6.3 Model Performance

Table 12: Classification Model Results (Delay Prediction)

Metric	Precision	Recall	F1-Score
On Time (10 MPH)	0.74	0.41	0.53
Delayed (¡10 MPH)	0.83	0.95	0.88
Accuracy			**0.815 (81.5%)**

Table 13: Regression Model Results (Speed Prediction)

Metric	Value
RMSE	7.10 MPH
R-squared	0.435
Mean Absolute Error	4.87 MPH

6.4 Feature Importance

Table 14: Top Predictive Features

Feature	Classification Model	Regression Model
operator_encoded	0.0479	0.6961
lon (longitude)	0.3810	0.1006
bearing	0.2594	0.0905
lat (latitude)	0.2761	0.0735
is_rush_hour	0.0017	0.0260
hour_24	0.0339	0.0133

Key Insight: Geographic features (lat, lon) account for 17.4% of prediction power in the regression model.

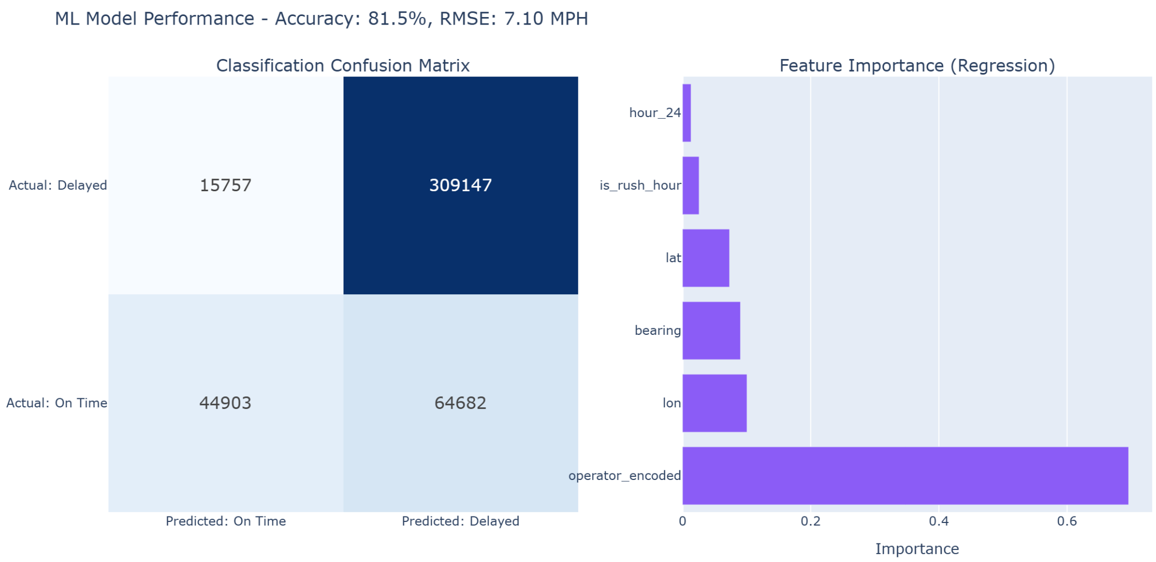

Figure 6: ML Model Performance - Left: Confusion matrix — Right: Feature importance (regression)

7 Summary and Recommendations

7.1 Key Findings

Daily Operational Costs (Measured):

- Total daily waste: $1,383,687
- Cost per vehicle: $972
- Excess idle time: 36.0%
- Total idle time: 61.0%

Route Performance:

- 5 critical routes (2.6%) costing $1.84M per day
- 4 routes needing improvement (2.1%)
- 183 routes performing acceptably (95.3%)

Temporal Patterns:

- Best travel time: 6 AM (6.6 MPH, 45 min for 5 miles)
- Worst travel time: 3 AM (1.7 MPH, 172 min for 5 miles)
- Time savings: 127 minutes (74% improvement)

Vehicle Productivity:

- Fleet average: 1.62

- Top 10 average: 5.79 (258% above average)
- Bottom 10: Zero activity (investigation required)

Predictive Capabilities:

- Classification accuracy: 81.5%
- Regression RMSE: 7.10 MPH
- Geographic features key predictors

7.2 Annual Projection

If weekday patterns remain consistent:

- 252 weekdays × $1,383,687 = **$348,689,004 annual cost**
- Assumes no weekend or holiday service variations
- Requires validation with extended data collection

7.3 Data Limitations

- Single weekday sample (Tuesday only)
- Weekend patterns not captured
- Seasonal variations not measured
- Weather effects not analyzed
- Special events not accounted for

7.4 Recommended Actions

Immediate (Week 1):

1. Investigate 10 zero-activity vehicles ($5K, 2 days)
2. Analyze worst performing route ($0, 3 days)
3. Begin multi-day data collection ($0, continuous)

Short-term (Month 1):

1. Collect 30-day continuous data for validation
2. Compare weekend vs weekday patterns
3. Establish seasonal baseline

Strategic (Quarter 1):

1. If patterns validate: Implement route optimization
2. Deploy real-time ML prediction system
3. Invest in targeted infrastructure improvements

7.5 Conclusion

This single-day analysis of November 18, 2025 provides:

- Baseline measurement of current operations
- Identification of problem routes and vehicles
- Proof-of-concept for ML predictive models
- Framework for ongoing monitoring
- Business case for data-driven optimization

Next Critical Step: Validate findings with 30-day data collection before major investments.

Tuesday, November 18, 2025

Daily Cost: $1,383,687

If representative of typical weekday operations, the opportunity for improvement is substantial.

Validation with multi-day data recommended before action.