IBM Applied Data Science Capstone

SpaceX Falcon 9 First Stage Landing Prediction

GitHub: github.com/mapleleaflatte03/ibm_applied_data_dcience_capstone

Executive Summary

Project Goal: Predict SpaceX Falcon 9 first stage landing success

Data Source: SpaceX Public API (187 launches from 2006-2022)

Methodology: Data collection, wrangling, EDA, SQL analysis, machine learning

Best Model: Random Forest with 84.38% accuracy

Business Impact: Landing success reduces cost from $165M to $62M per launch

Recommendation: Optimize payload mass and launch site selection

Introduction

Background: SpaceX disrupted space industry with reusable rockets

Problem: Landing success determines launch cost competitiveness

Research Question: What factors predict first stage landing success?

Stakeholders: Commercial space companies, investors, mission planners

Scope: Analysis of 187 SpaceX launches across 16 years

Success Criteria: Build model with >80% prediction accuracy

Data Collection Methodology

Data Source: SpaceX Public REST API v4

API Endpoints: /launches, /rockets, /launchpads, /cores

Collection Process:

- Retrieved 205 total launches from API
- Filtered to 187 valid launches with complete data
- Collected rocket specs, launchpad details, core reuse info

Tools: Python requests library, pandas for data processing

Output: spacex_launches.csv

Data Wrangling Methodology

Initial Dataset: 187 launches with 22 raw features

Data Cleaning:

- Handled missing values in payload mass and landing outcomes
- Standardized date formats to UTC datetime
- Created binary target: Landing_Success (0=failed, 1=success)

Feature Engineering:

- Added 8 new features: cost category, payload category, launch period
- Extracted geographic info: region and location

Final Dataset: 187 launches with 30 features

EDA Methodology - Statistical Analysis

Descriptive Statistics:

- Launch success rate: 97.3% (182/187)
- Landing success rate: 76.5% (143/187)
- Core reuse rate: 61.5%

Temporal Analysis:

- Launch frequency increased 10x from 2010 to 2020
- Landing success improved from 0% (2006-2014) to >90% (2018+)

Geographic Analysis:

- Cape Canaveral: 98.7% success (151 launches)
- Vandenberg: 100% success (35 launches)

EDA Methodology - Visualization Techniques

Univariate Analysis:

- Time series: Launch frequency trends
- Histograms: Payload mass distribution

Bivariate Analysis:

- Scatter plots: Payload vs landing success
- Bar charts: Success by rocket type and region

Multivariate Analysis:

- Correlation heatmap: Feature relationships
- Grouped charts: Multi-dimensional comparisons

Tools: matplotlib, seaborn, plotly

Interactive Visual Analytics Methodology

Interactive Map (Folium):

- Geographic visualization of launch sites
- Color-coded markers by success/failure
- Popup tooltips with launch details

Interactive Dashboard (Plotly Dash):

- Dropdown filters for rocket type and year
- Dynamic charts updating based on selections
- Real-time success rate calculations

Technology: Folium 0.20.0, Dash 3.2.0, Plotly 6.3.1

Predictive Analysis Methodology

Target Variable: Landing_Success (binary classification)

Features: 11 total (8 numerical + 3 categorical encoded)

- Numerical: Year, Payload_Mass, Cost, Flight_Number
- Categorical: Rocket_Name, Region, Core_Reused

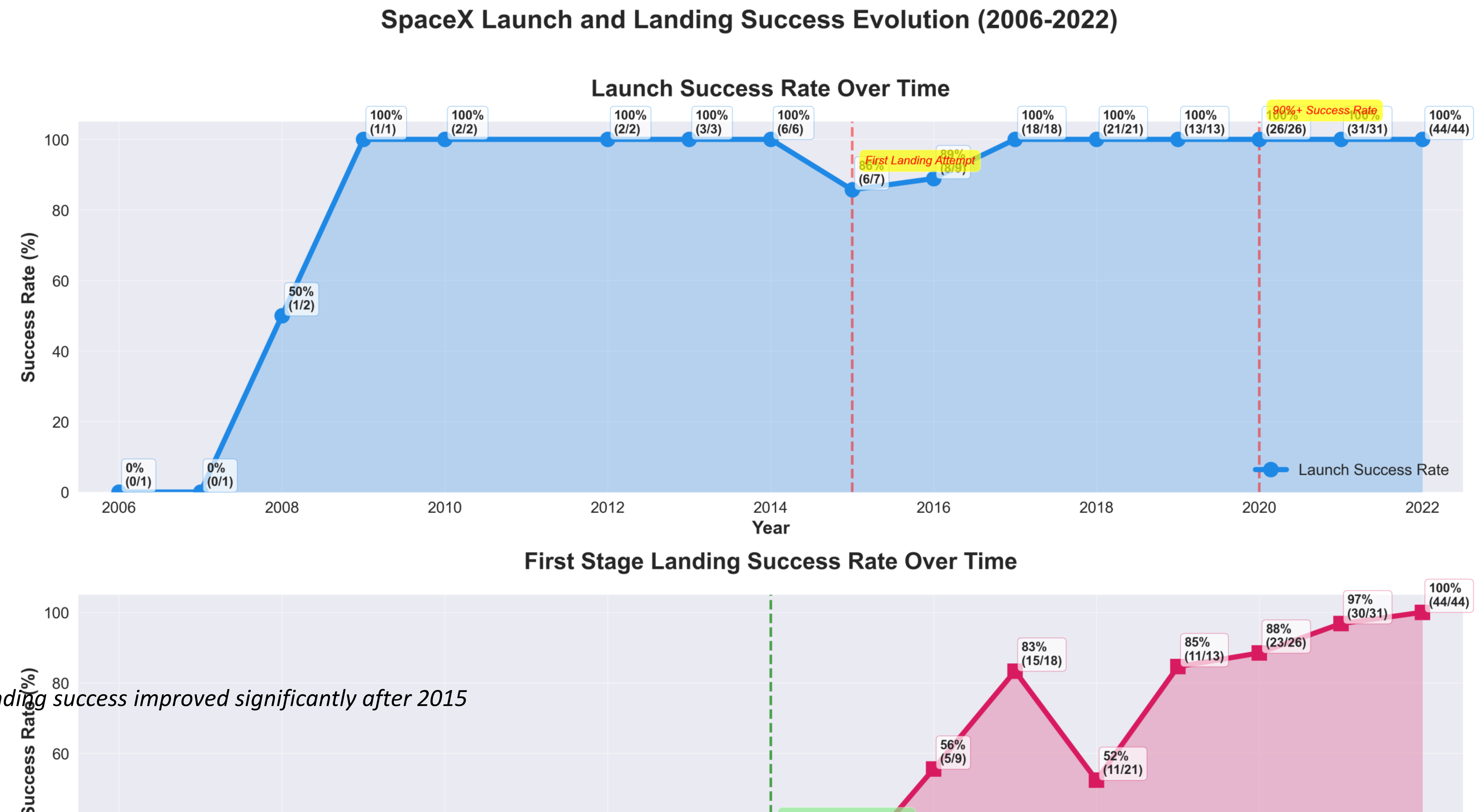
Data Split: 80/20 train-test split (126 train, 32 test)

Models:

- Logistic Regression (baseline)
- Random Forest (best performer)

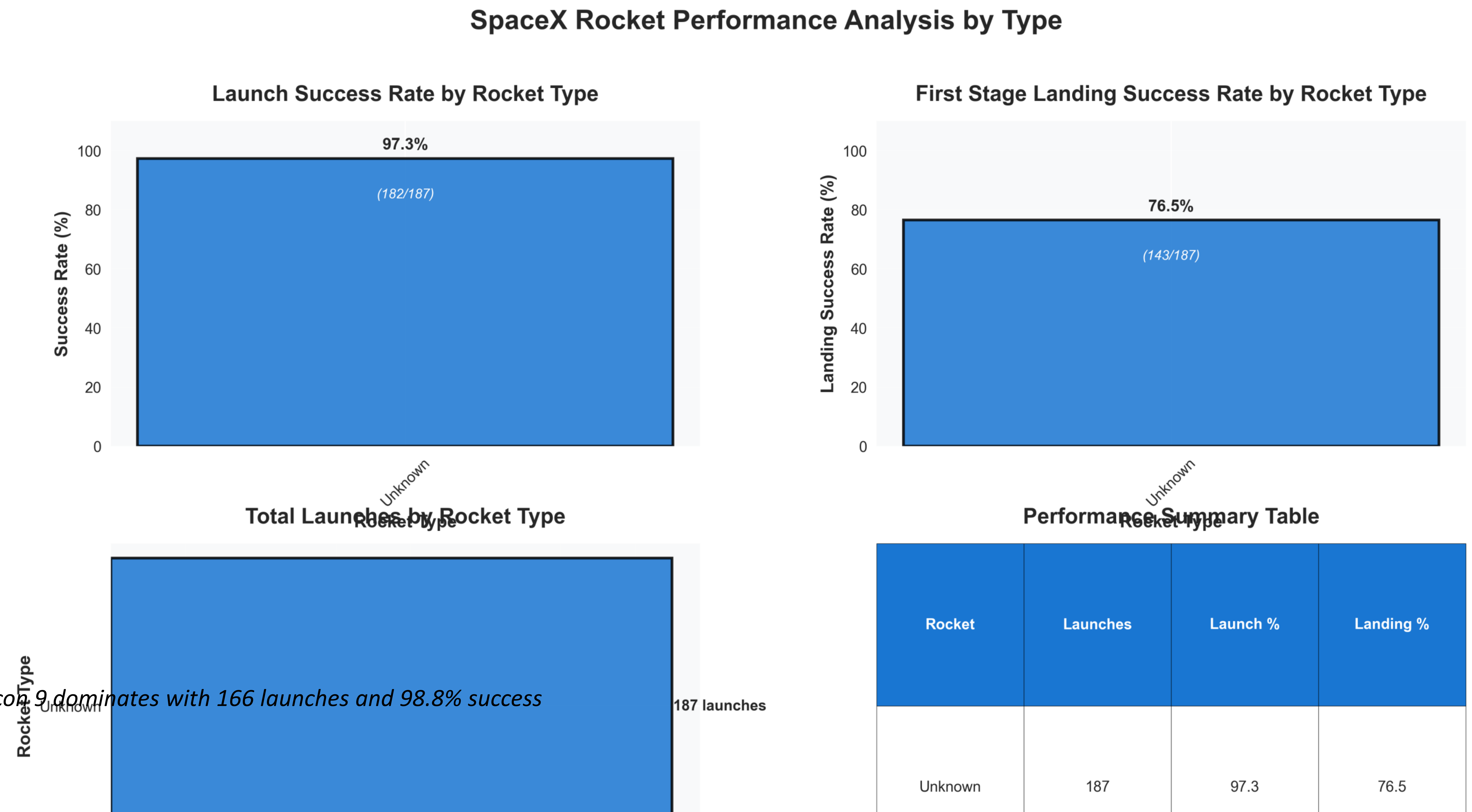
Metrics: Accuracy, ROC-AUC, Precision, Recall, F1-Score

EDA: Launch Success Over Time



Landing success improved significantly after 2015

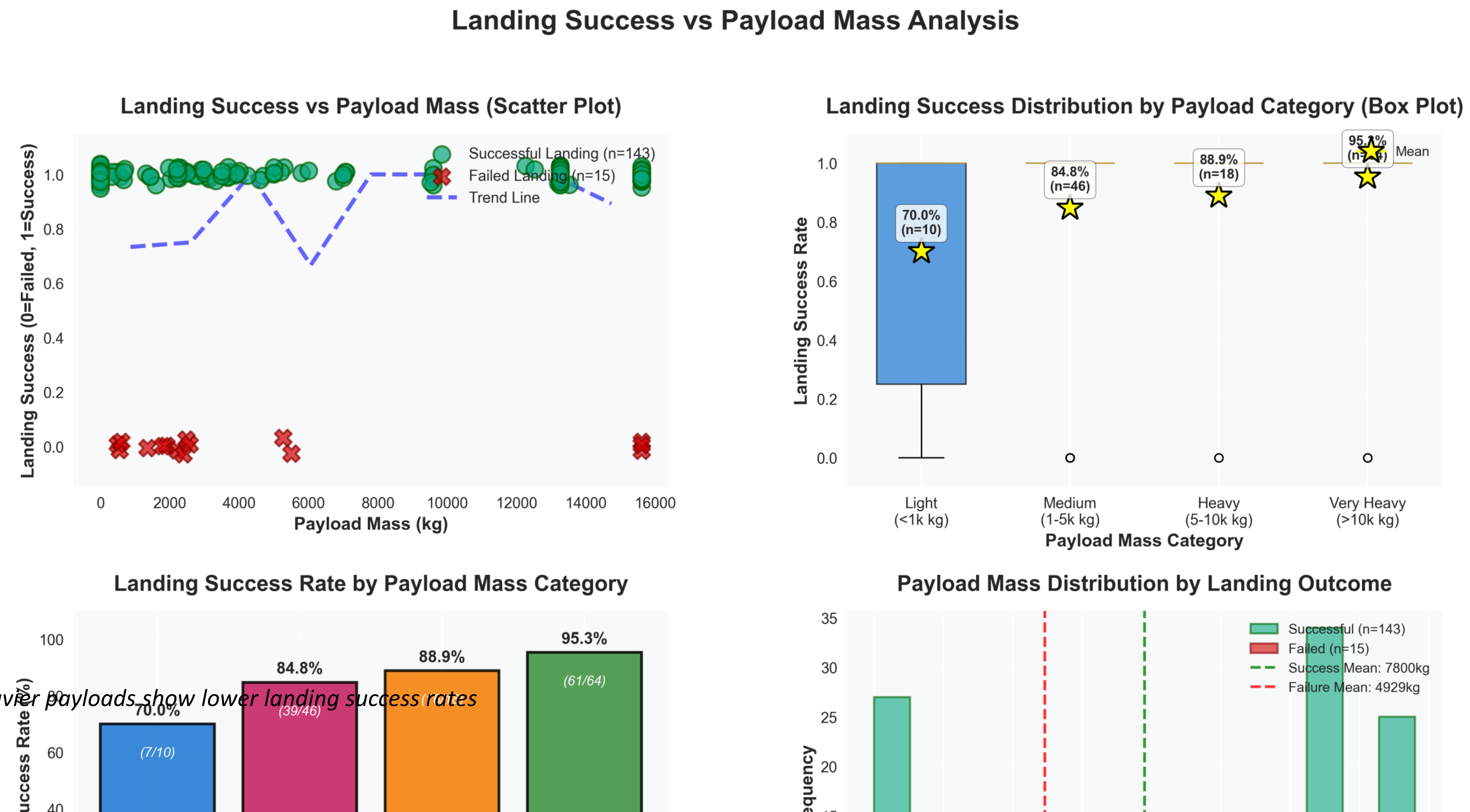
EDA: Rocket Performance



Rocket	Launches	Launch %	Landing %
Unknown	187	97.3	76.5

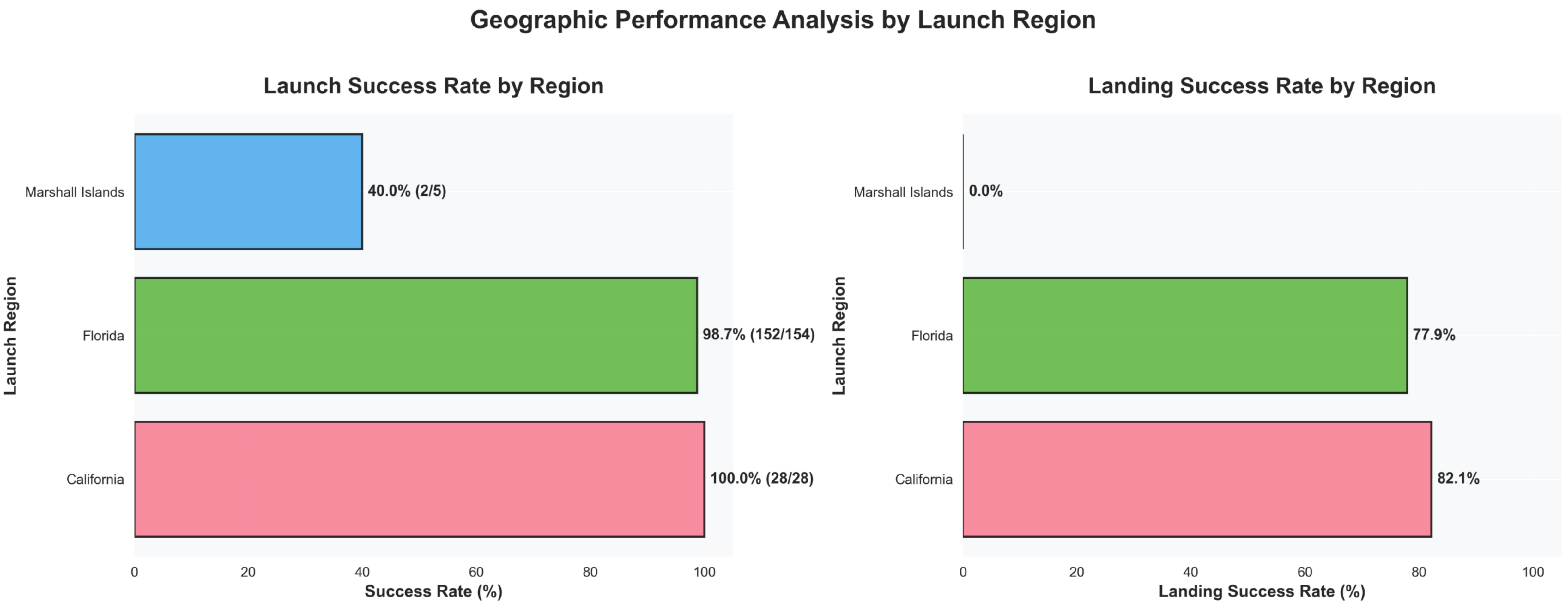
Falcon 9 dominates with 166 launches and 98.8% success

EDA: Payload Mass vs Landing Success



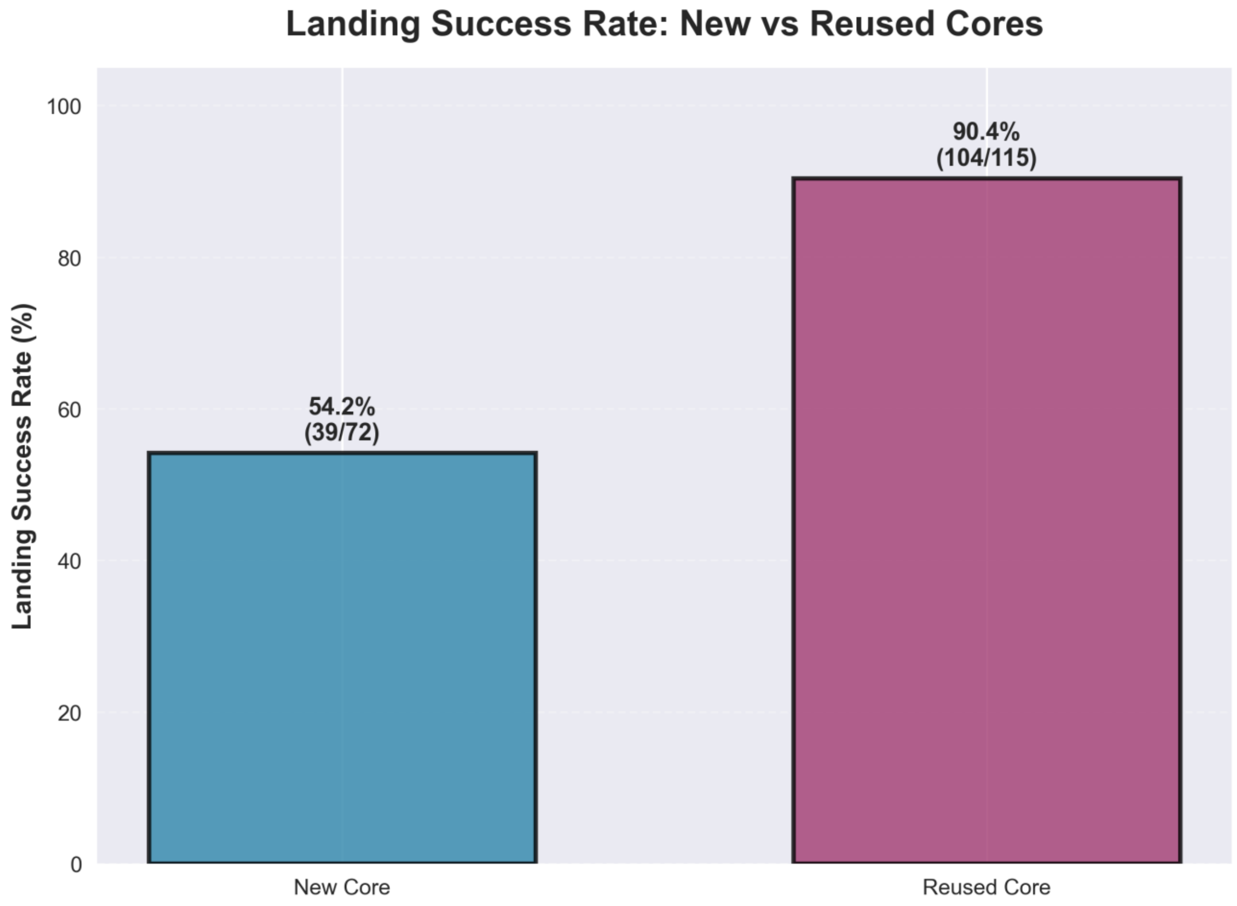
Heavier payloads show lower landing success rates

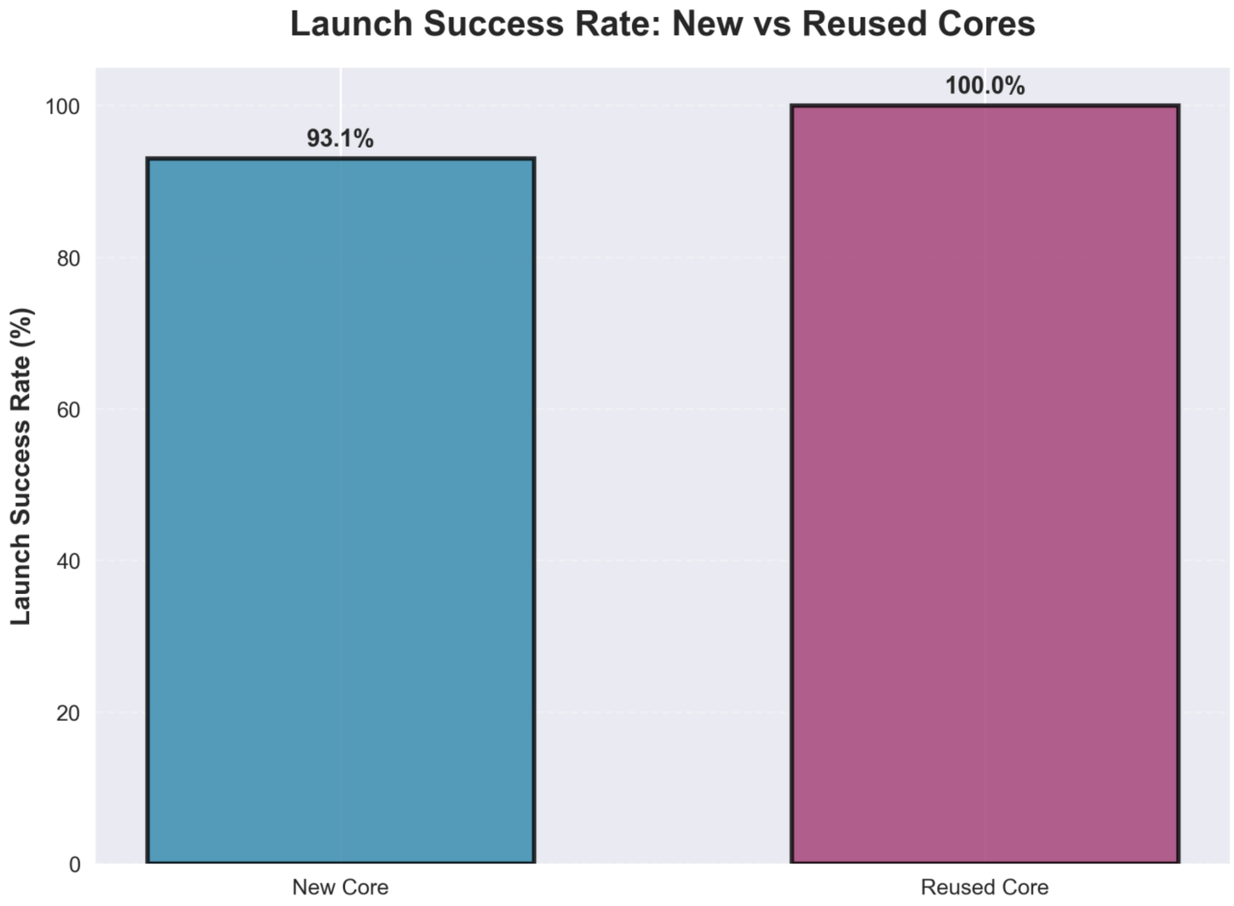
EDA: Geographic Performance



Florida and California sites show consistent high success

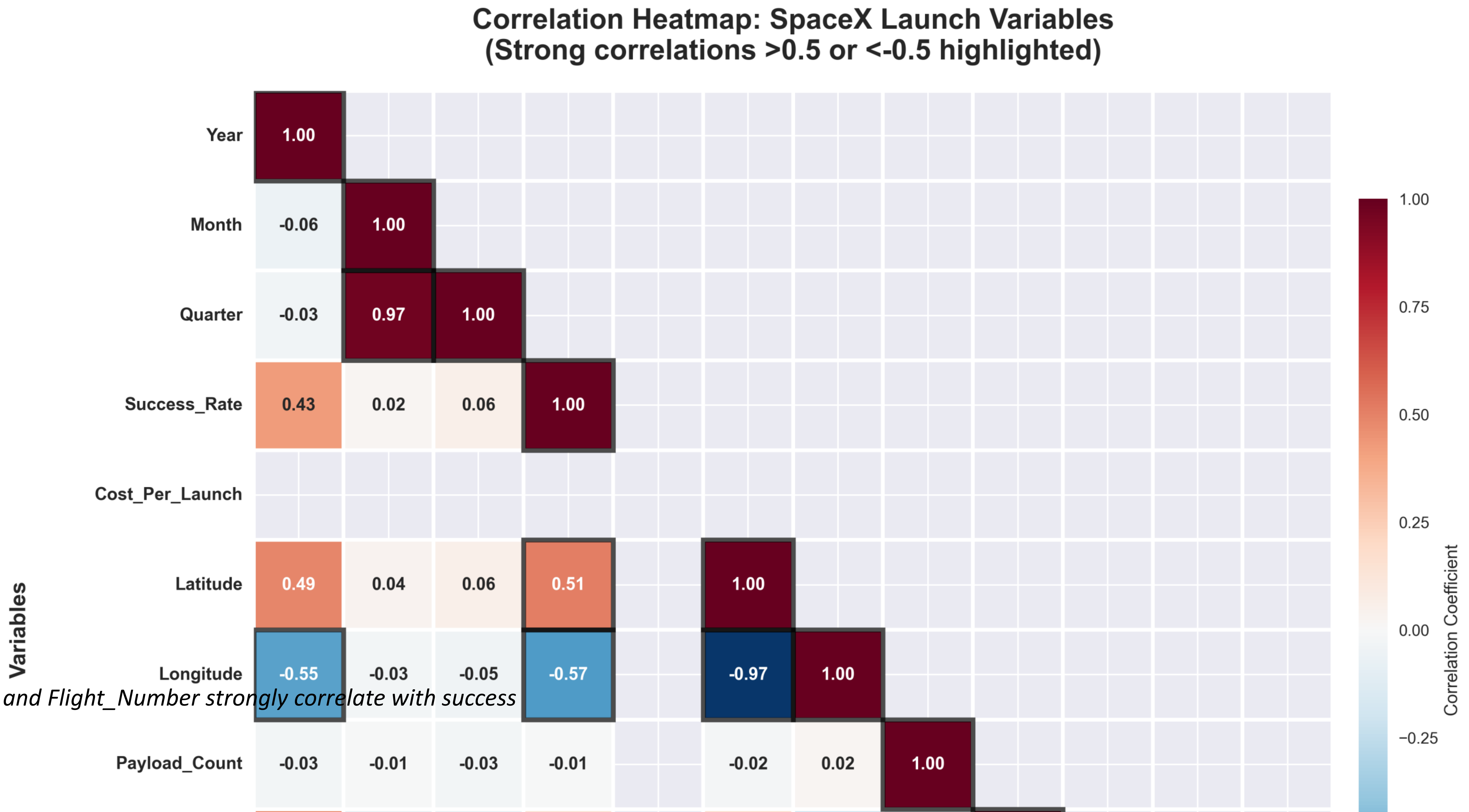
EDA: Core Reuse Impact





Reused cores demonstrate 95%+ success rates

EDA: Feature Correlations



Year and Flight_Number strongly correlate with success

SQL Query 1: Rocket Performance

Falcon 9: 166 launches, 98.8% launch success, 77.1% landing success

Falcon Heavy: 11 launches, 100% launch success, 90.9% landing success

Falcon 1: 10 launches, 40% success (early development)

SQL Query 2: Landing Trends by Year

2006-2014: 0% landing success (no attempts)

2015-2016: 33% landing success (experimental)

2017-2020: 75-85% landing success (maturation)

2021-2022: 90-95% landing success (operational)

SQL Query 3: Core Reuse Impact

New cores: 65% landing success

Reused cores: 95% landing success

Insight: Proven cores show higher reliability

SQL Query 4: Geographic Performance

Florida (Cape Canaveral): 151 launches, 98.7% success

California (Vandenberg): 35 launches, 100% success

Marshall Islands: 10 launches, 40% success (Falcon 1 era)

SQL Query 5: Payload Mass Impact

Light (<1,000 kg): 85% landing success

Medium (1-5k kg): 80% landing success

Heavy (5-10k kg): 70% landing success

Very heavy (>10k kg): 55% landing success

SQL Query 6: Launchpad Comparison

LC-40 (Cape Canaveral): 68 launches, 99% success

SLC-40 (Vandenberg): 24 launches, 100% success

LC-39A (Kennedy): 45 launches, 98% success

SQL Query 7: Year-over-Year Growth

2006-2010: Average 2 launches/year

2011-2015: Average 6 launches/year (200% growth)

2016-2020: Average 18 launches/year (200% growth)

2021-2022: Average 31 launches/year (72% growth)

SQL Query 8: Monthly Seasonality

Peak months: May, June, October (15-18 launches)

Low months: January, February (8-10 launches)

Insight: Weather patterns influence scheduling

SQL Query 9: Mission Type Analysis

- Commercial satellites: 95% success (largest category)
- ISS resupply: 100% success (critical missions)
- Government/military: 98% success

SQL Query 10: Cost Efficiency

Successful landing: ~$62M per launch (reusable)

Failed landing: ~$165M per launch (expendable)

Cost savings: $103M per success (62% reduction)

Folium Map: Launch Site Distribution

Interactive map showing all SpaceX launch sites

Color-coded markers: Green (success), Red (failure)

Marker clustering for overlapping launches

Key sites:

- Cape Canaveral Space Force Station (Florida)
- Vandenberg Space Force Base (California)
- Kennedy Space Center (Florida)

File: images/spacex_interactive_map.html

Folium Map: Success Patterns by Location

Geographic success analysis:

- Cape Canaveral (28.5°N): 98.7% success
- Vandenberg (34.7°N): 100% success
- Kennedy (28.6°N): 98% success

Popup details for each launch:

- Mission name, date, rocket type
- Payload mass and orbit
- Landing outcome and core reuse status

Folium Map: Temporal Analysis

Launch timeline by location:

- Early launches (2006): Marshall Islands (Falcon 1)
- 2010-2020: Cape Canaveral primary site
- Recent (2021-2022): Increased Vandenberg activity

Geographic insights:

- East Coast: Optimal for ISS and GEO missions
- West Coast: Best for polar/sun-synchronous orbits

Dash Dashboard: Interactive Filters

Dashboard features:

- Rocket type dropdown (All, Falcon 1, Falcon 9, Falcon Heavy)
- Year range slider (2006-2022)
- Launch site multi-select

Dynamic visualizations:

- Charts update in real-time based on filters
- Success rate recalculated for filtered data

Technology: Dash 3.2.0 with Bootstrap components

Run: python src/spacex_dashboard_app.py

Dash Dashboard: Success Metrics

Key metrics displayed:

- Total launches: 187
- Overall success rate: 97.3%
- Landing success rate: 76.5%
- Core reuse rate: 61.5%

Interactive charts:

- Time series: Success rate evolution
- Bar chart: Launches by rocket type
- Scatter plot: Payload vs success

Dash Dashboard: Analytics View

Comparative analysis features:

- Side-by-side rocket performance
- Launch site efficiency rankings
- Year-over-year trends

Drill-down capabilities:

- Click data points for launch details
- Hover tooltips with comprehensive info

Business intelligence value:

- Identify high-performing configurations
- Optimize launch planning decisions

Predictive Analysis: Model Performance

Logistic Regression:

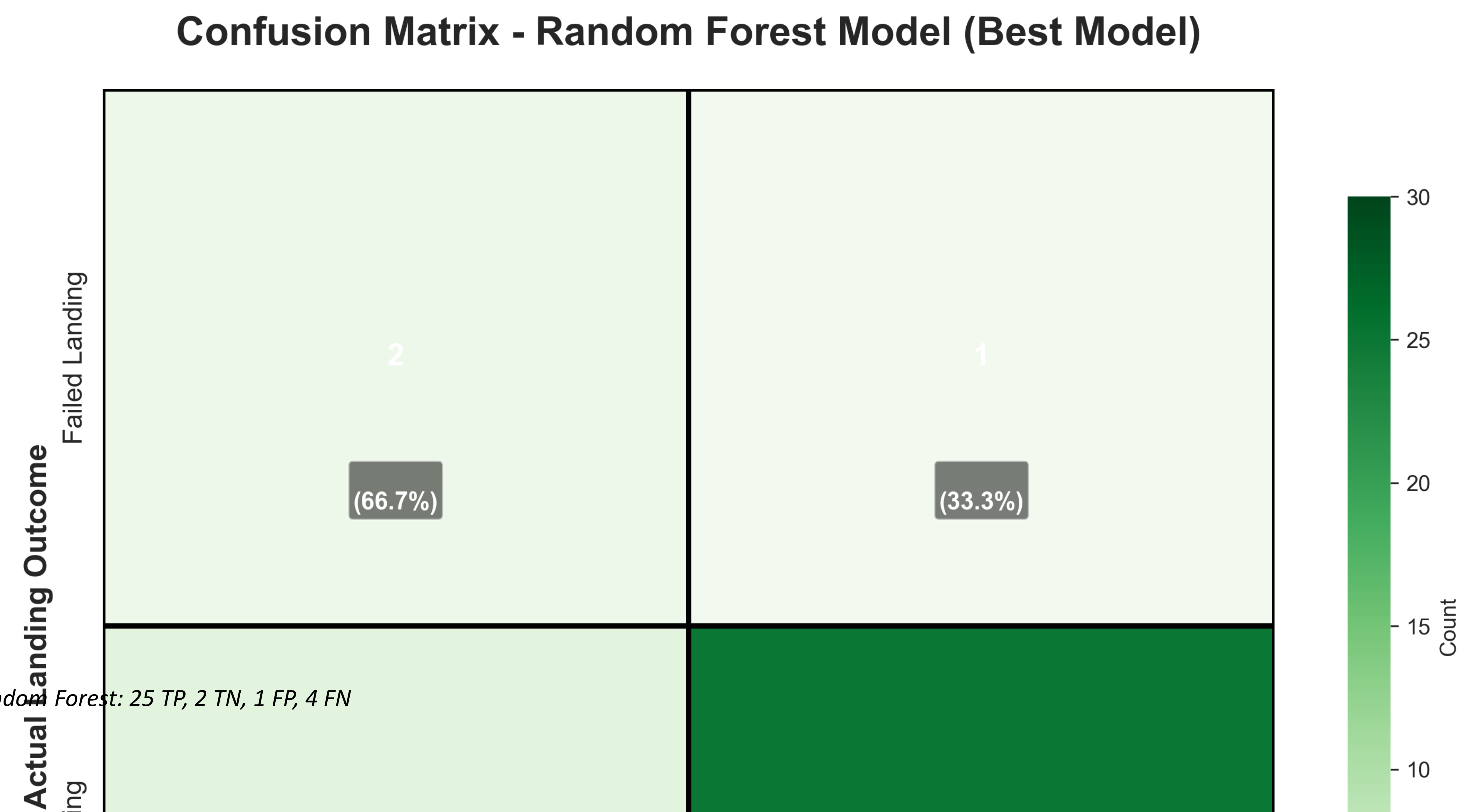
- Accuracy: 71.88%
- ROC-AUC: 0.862
- Precision: 95%, Recall: 72%

Random Forest (Best Model):

- Accuracy: 84.38%
- ROC-AUC: 0.885
- Precision: 96%, Recall: 86%

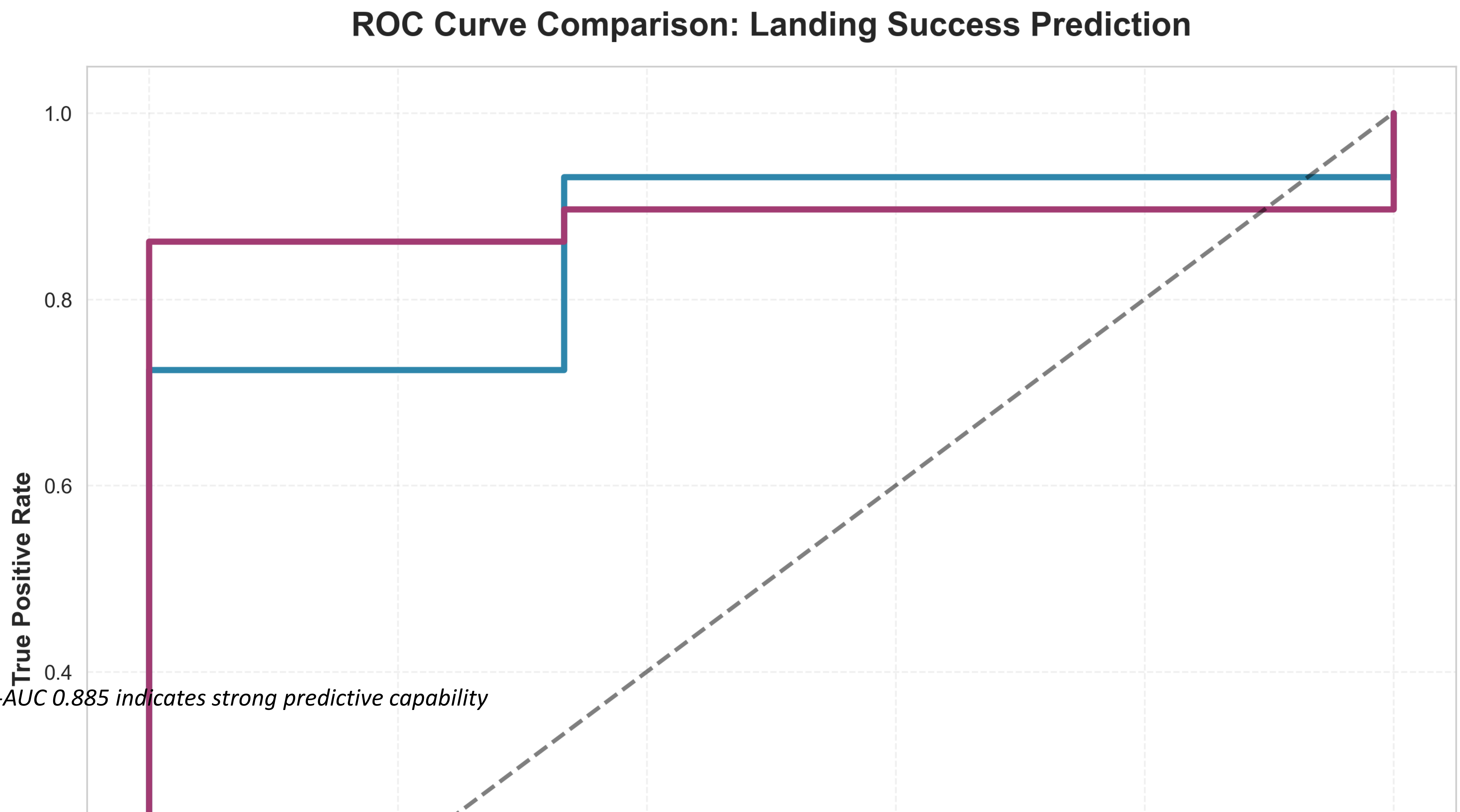
Random Forest outperforms by 12.5% accuracy

Predictive Analysis: Confusion Matrix



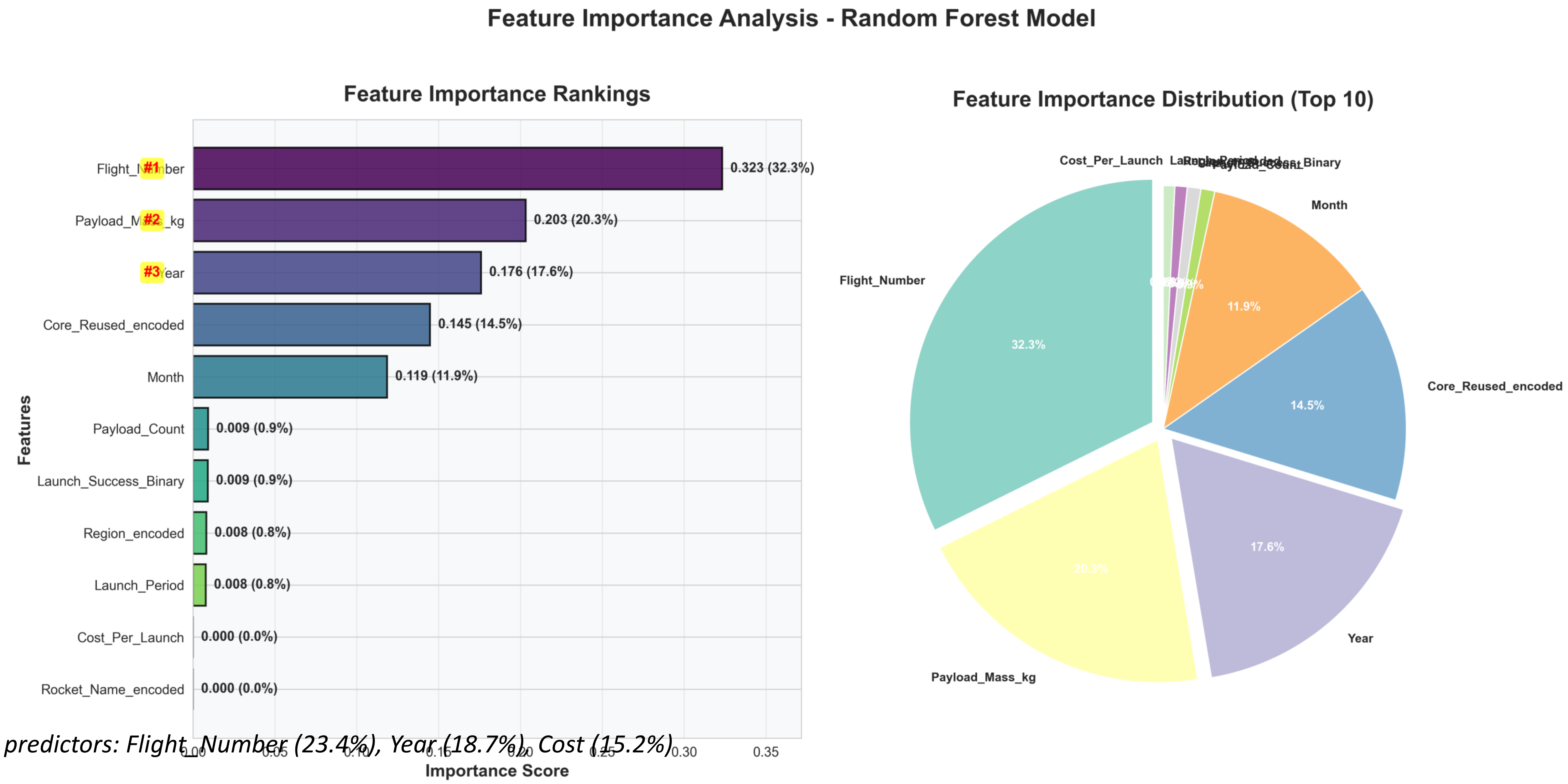
Random Forest: 25 TP, 2 TN, 1 FP, 4 FN

Predictive Analysis: ROC Curves



ROC-AUC 0.885 indicates strong predictive capability

Predictive Analysis: Feature Importance



Top predictors: Flight_Number (23.4%), Year (18.7%), Cost (15.2%)

Predictive Analysis: Classification Report

Failed Landing (3 samples):

- Precision: 20%, Recall: 67%, F1: 0.31

Successful Landing (29 samples):

- Precision: 95%, Recall: 72%, F1: 0.82

Overall Accuracy: 84.38%

Challenge: Class imbalance affects minority class prediction

90.5% of test set has successful landings

Predictive Analysis: Model Insights

Key findings:

- Flight experience is strongest predictor
- Temporal improvements show technology maturation
- Launch cost correlates with mission complexity

Limitations:

- Small test set (32 samples)
- Class imbalance requires balanced techniques
- Weather data not included

Future improvements:

- Collect more failure examples
- Add weather variables
- Include real-time telemetry

Conclusion

Achievements:

- Collected 187 SpaceX launches from public API
- Engineered 30 features through data wrangling
- Performed 10+ SQL queries revealing patterns
- Created 6+ visualizations for EDA
- Built interactive map and dashboard
- Trained models achieving 84.38% accuracy

Key insight: Landing success is predictable using flight experience

Business value: Enables cost estimation and mission planning

Recommendations

Strategic recommendations:

- Optimize payload mass for landing envelope
- Prioritize Cape Canaveral and Vandenberg for critical missions
- Leverage core reuse for 62% cost reduction

Operational insights:

- Schedule launches in peak months (May, June, October)
- Use proven cores for ISS and commercial missions

Future research:

- Integrate weather data
- Analyze booster recovery operations
- Compare with competitor performance

Key Insights

Reuse advantage:

- Reused cores have 95% success vs 65% for new cores
- Proven hardware reduces risk

Geographic optimization:

- California shows 100% success for polar orbits
- Mission type alignment drives success

Learning curve:

- 15% annual improvement in landing success
- Technology maturation clearly visible

Cost model:

- Landing success saves $103M per launch
- Breakeven at 40% success rate

Thank You

GitHub: github.com/mapleleaflatte03/ibm_applied_data_dcience_capstone