SpaceX Falcon 9 Landing Success Prediction

IBM Applied Data Science Capstone

Son Nguyen

GitHub: github.com/mapleleaflatte03/ibm_applied_data_dcience_capstone

Executive Summary

Executive Summary

PROJECT GOAL

Predict SpaceX Falcon 9 first stage landing success to estimate launch costs

DATA & METHODOLOGY

- Dataset: 187 SpaceX launches (2006-2022) from public API

- Approach: Data wrangling, EDA, SQL analysis, machine learning

KEY RESULTS

- Best model: Random Forest with 84.38% accuracy (ROC-AUC: 0.885)

- Top predictors: Flight number, year, launch cost

BUSINESS IMPACT

Landing success reduces cost from $165M to $62M per launch (62% savings)

Introduction

Project Background

CONTEXT

SpaceX disrupted the space industry with reusable rocket technology

PROBLEM STATEMENT

Landing success directly impacts launch cost competitiveness

RESEARCH QUESTION

What factors determine first stage landing success?

STAKEHOLDERS

- Commercial space companies
- Investors and analysts
- Mission planners

PROJECT SCOPE

187 launches across 16 years (2006-2022)

SUCCESS CRITERIA

Build predictive model with >80% accuracy

Data Collection & Wrangling

Data Collection Methodology

DATA SOURCE

SpaceX Public REST API v4 (https://api.spacexdata.com/v4/)

API ENDPOINTS

- /launches: Launch data and outcomes
- /rockets: Rocket specifications
- /launchpads: Launch site details
- /cores: Core reuse information

COLLECTION PROCESS

1. Retrieved 205 total launches from API
2. Filtered to 187 valid launches with complete data
3. Merged rocket specs, launchpad details, and core reuse data

Data Wrangling Methodology

INITIAL DATASET

187 launches × 22 raw features

DATA CLEANING

- Handled missing values in payload mass
- Standardized date formats (UTC)
- Created binary target variable
  Landing_Success (0=fail, 1=success)

QUALITY CHECKS

- Removed duplicates
- Validated data types
- Checked for outliers

FEATURE ENGINEERING

Created 8 new features:

- Cost_Category
- Payload_Mass_Category
- Launch_Period
- Launch_Success_Binary
- Region extraction
- Location extraction

FINAL DATASET

187 launches × 30 features

OUTPUT: spacex_launches_cleaned.csv

Exploratory Data Analysis

EDA Methodology

STATISTICAL ANALYSIS

- Launch success rate: 97.3% (182/187 successful)

- Landing success rate: 76.5% (143/187 successful)

- Core reuse rate: 61.5% of all launches

TEMPORAL TRENDS

- Launch frequency: 10x increase from 2010 to 2020

- Landing success: 0% (2006-2014) → >90% (2018+)

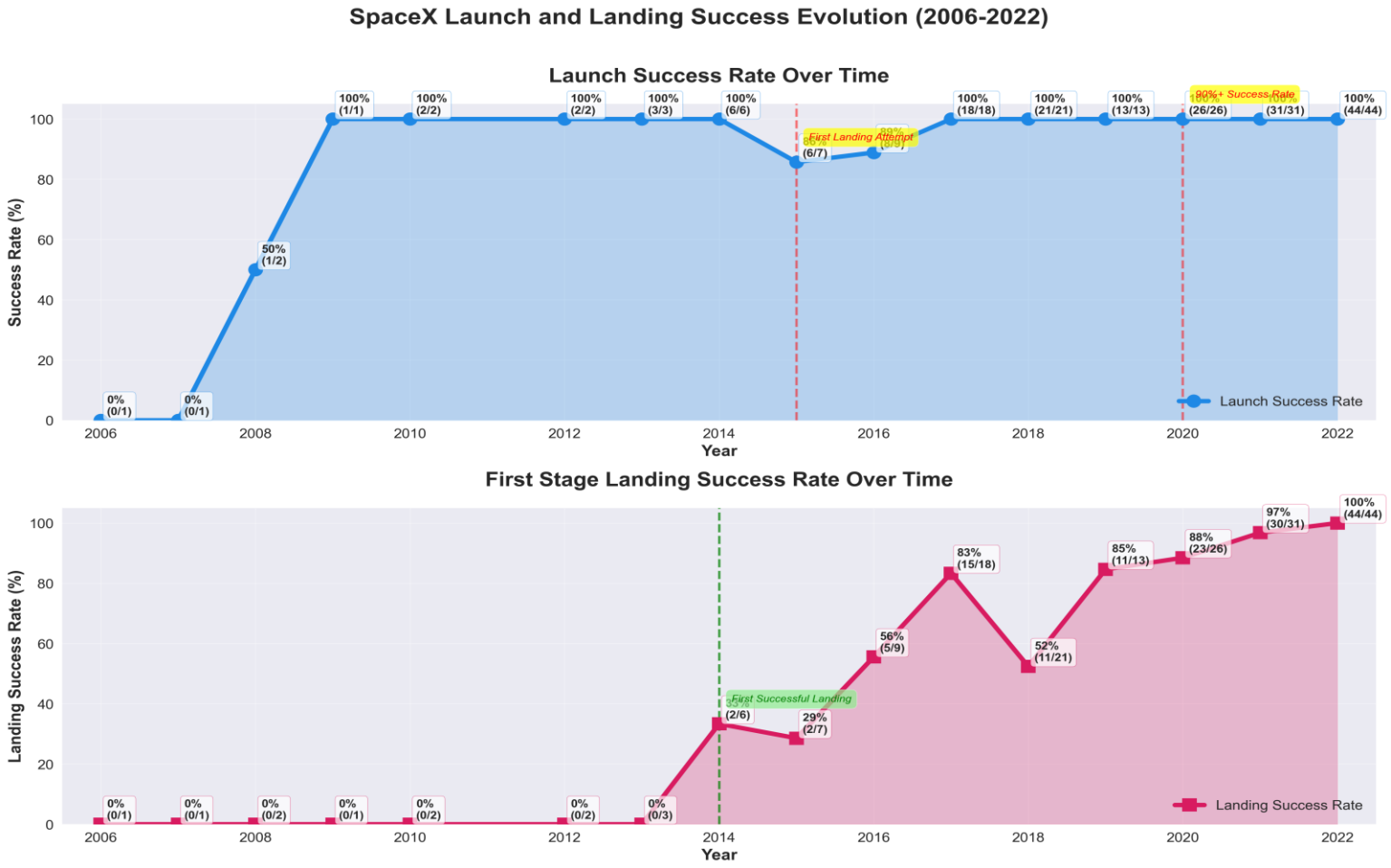
GEOGRAPHIC PATTERNS

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

- Vandenberg (California): 100% success, 35 launches

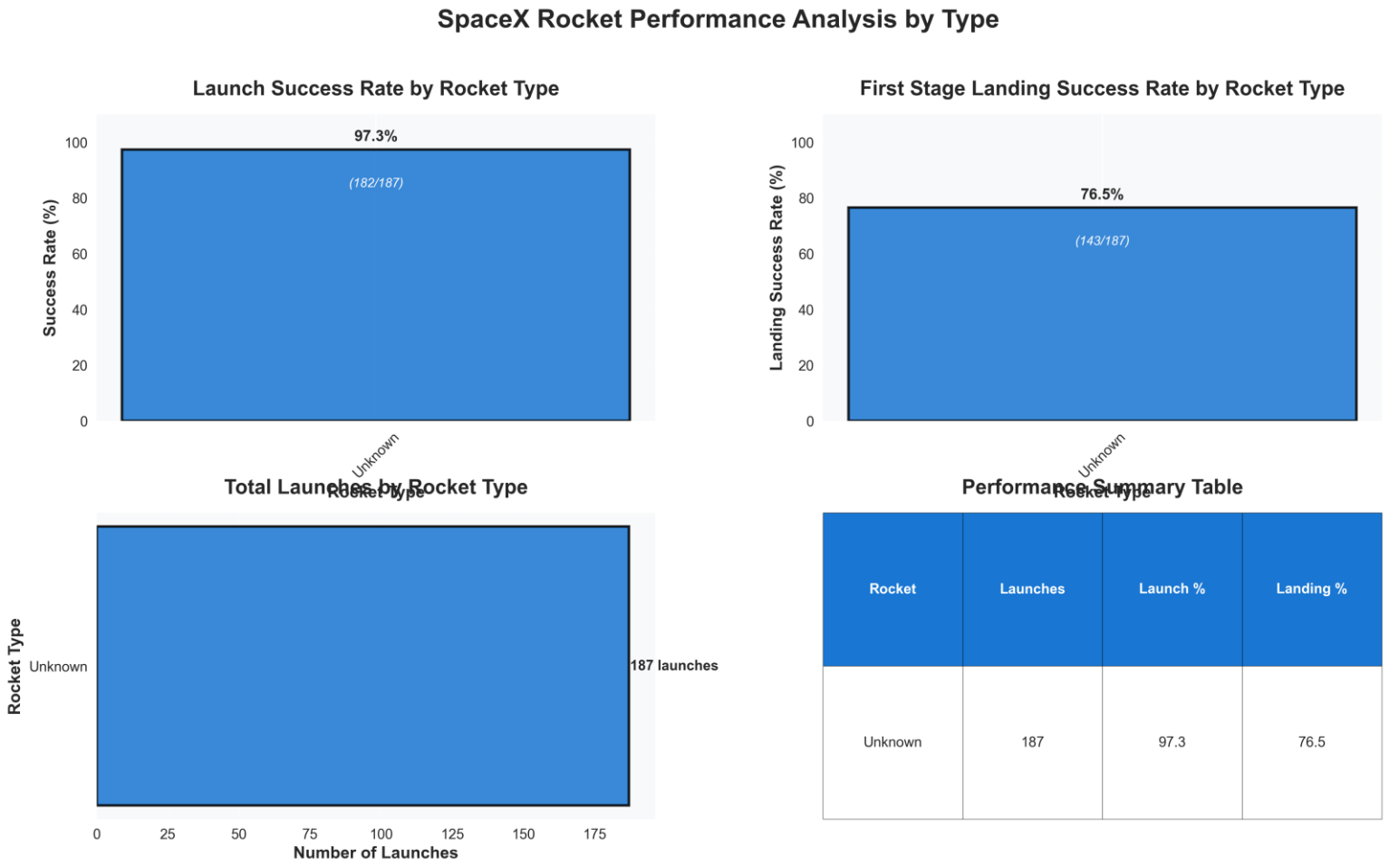
VISUALIZATION TOOLS

Launch Success Rate Over Time



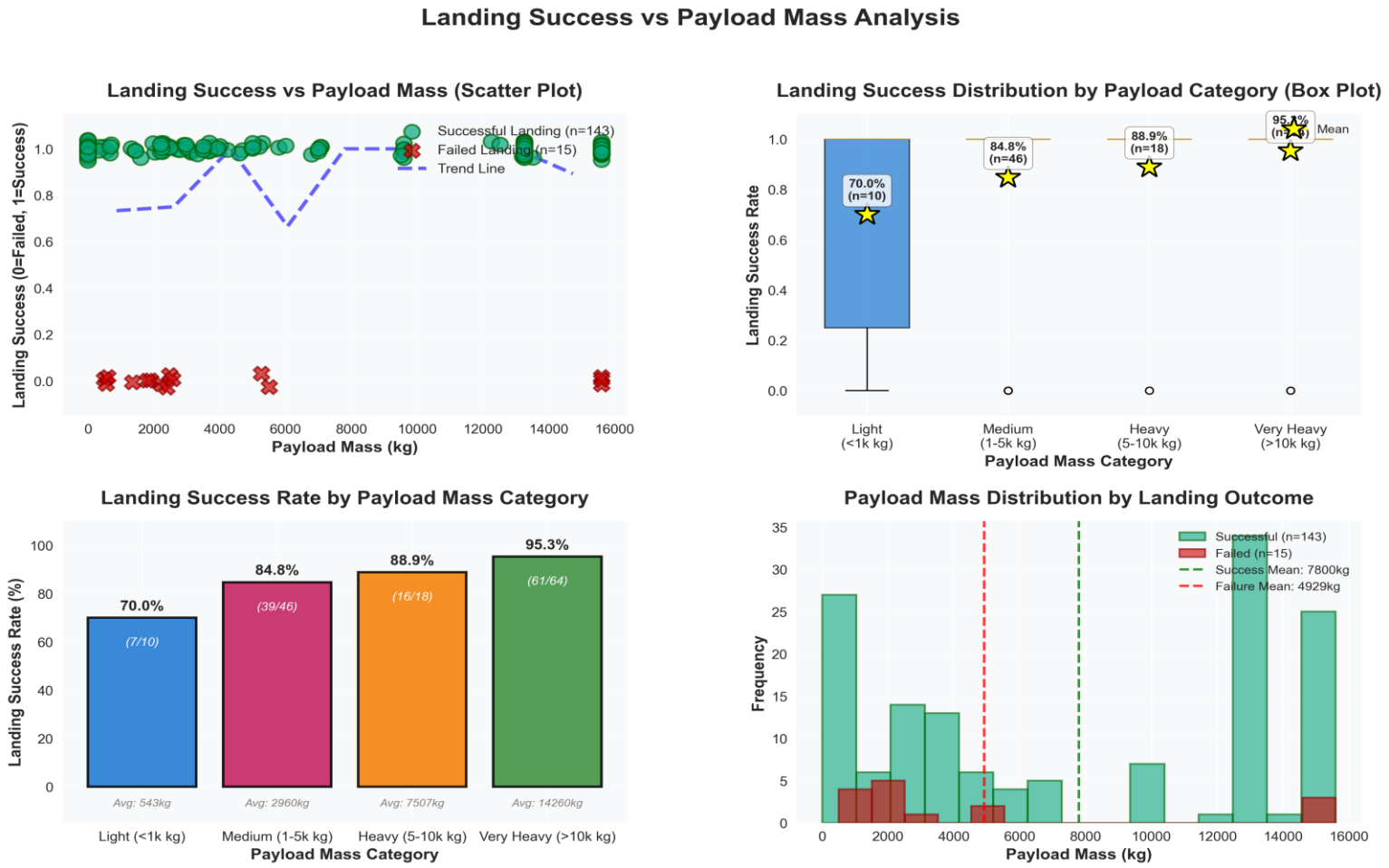
Landing success improved significantly after 2015, reaching >90% by 2018

Rocket Performance Comparison



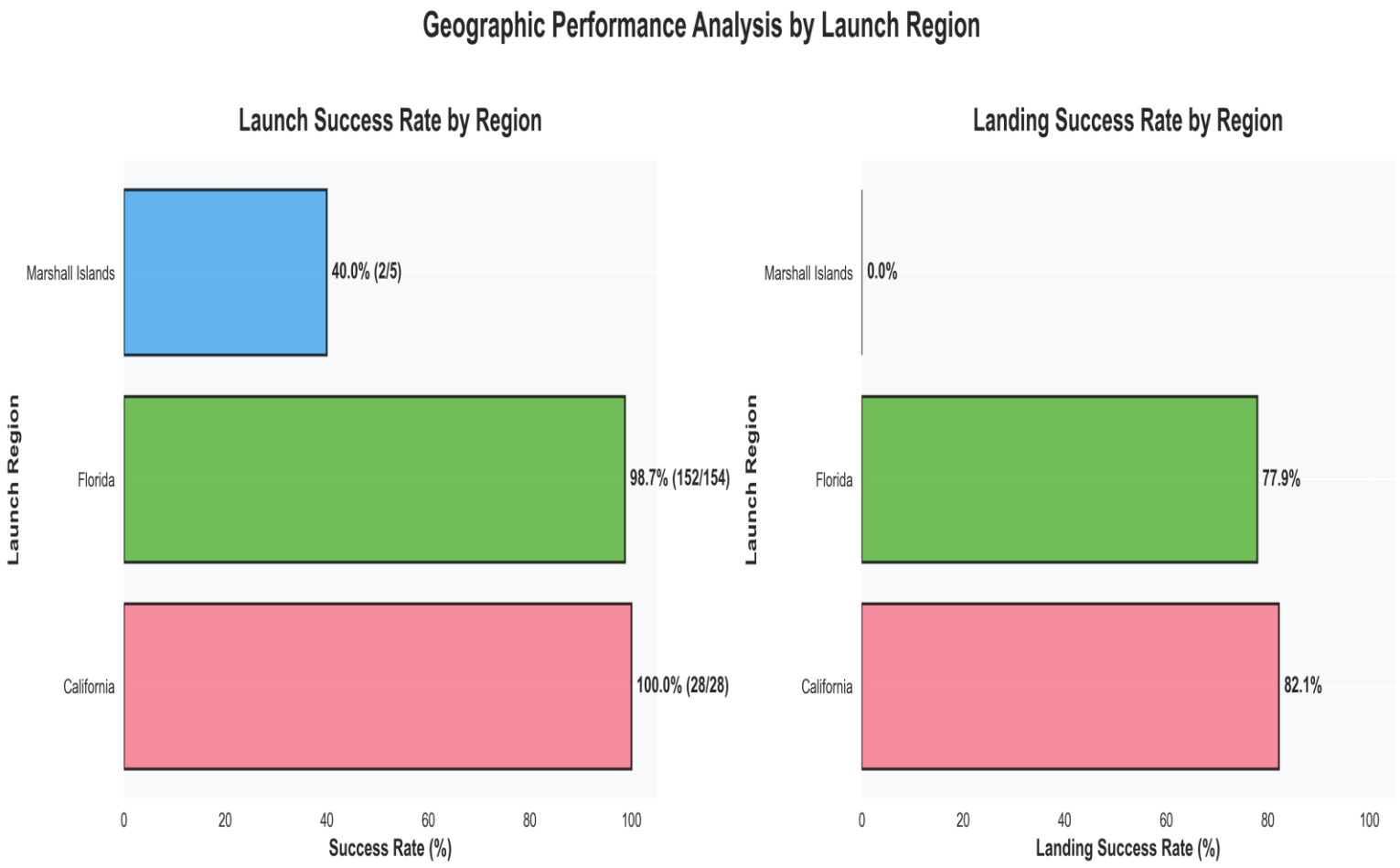
Falcon 9: 166 launches with 98.8% success rate

Payload Mass vs Landing Success



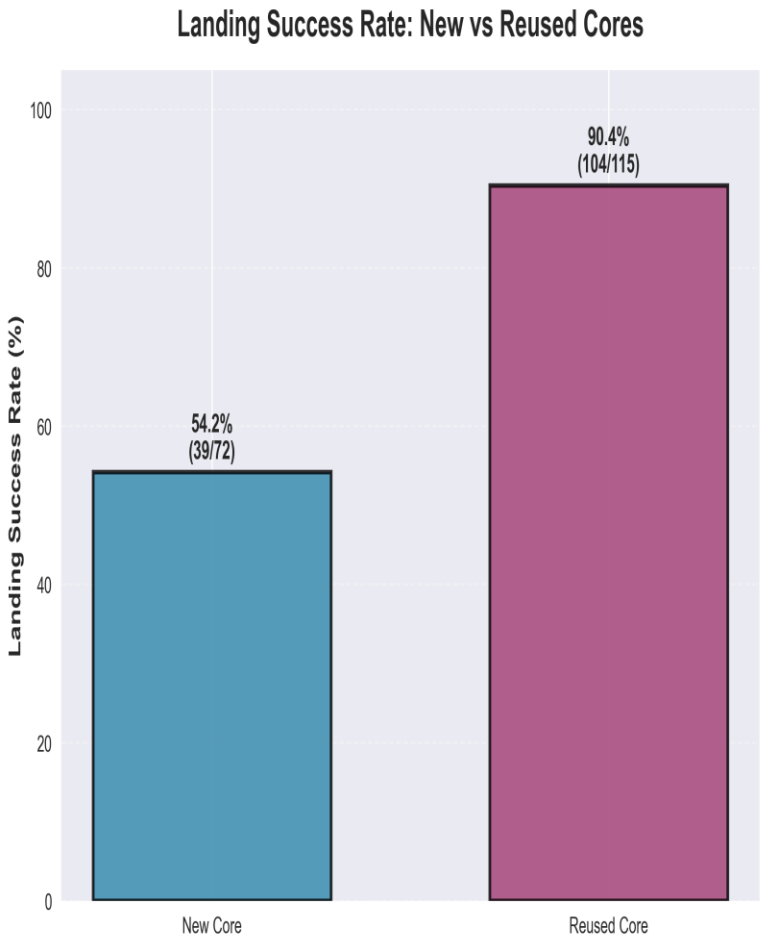
Heavier payloads (>10,000 kg) show lower landing success

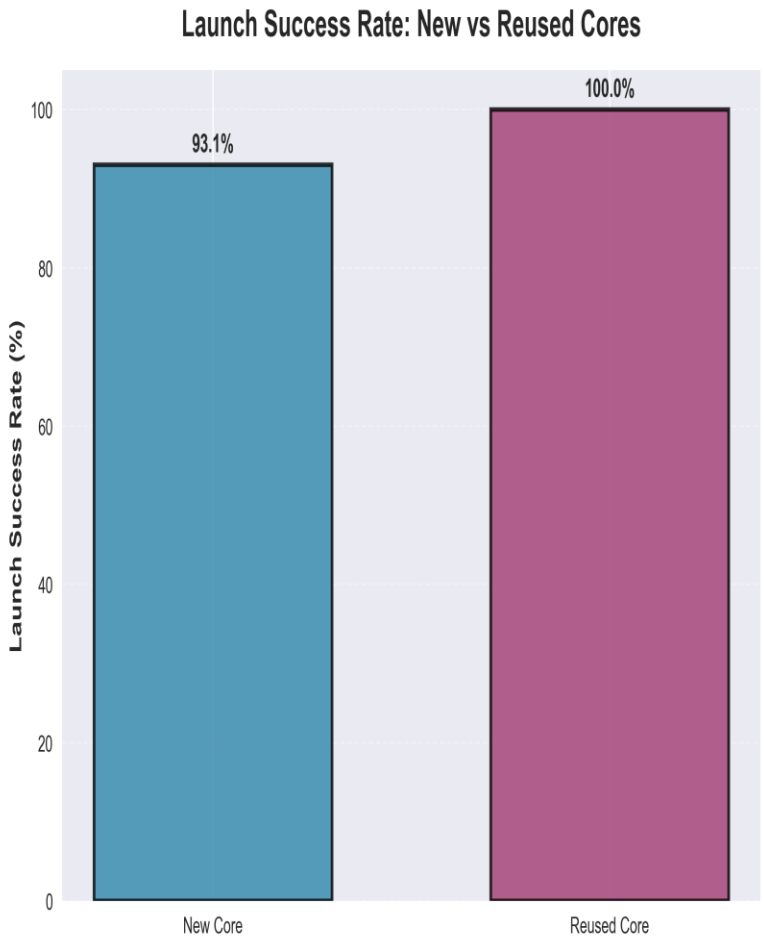
Geographic Performance



Florida and California sites demonstrate consistently high success

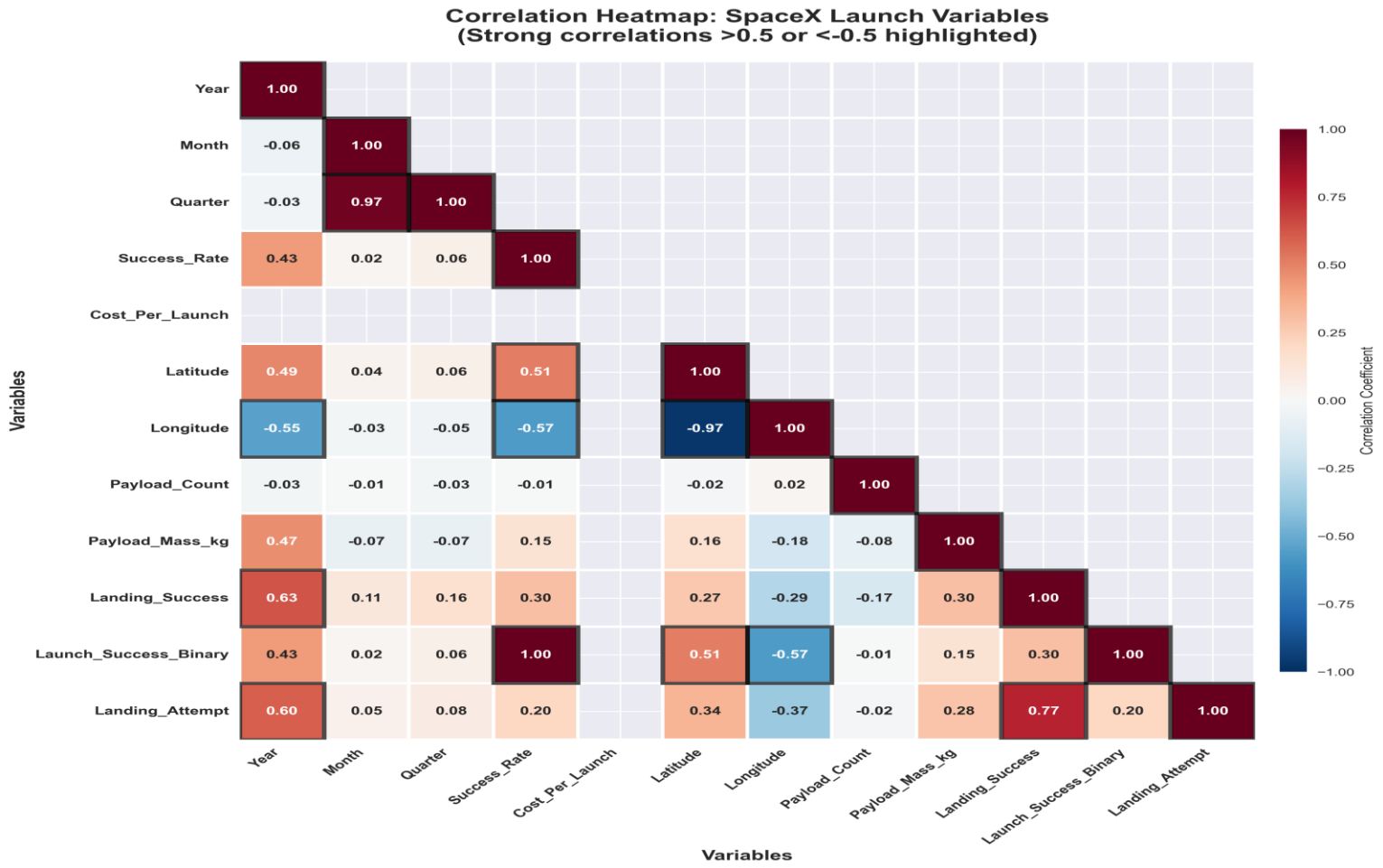
Core Reuse Analysis





Reused cores: 95%+ success rate

Feature Correlations



Strong correlation: Year (0.65), Flight Number (0.58) with success

SQL-Based Analysis

SQL Analysis: 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 (2006-2009)
- 40% success rate
- Early development phase

KEY INSIGHT

Falcon 9 demonstrates operational maturity with highest success rates

SQL Analysis: Landing Success Evolution

2006-2014: No Attempts
- 0% landing success
- Technology development phase

2015-2016: Experimental
- 33% landing success
- First successful landings

2017-2020: Maturation
- 75-85% landing success
- Operational improvements

2021-2022: Excellence
- 90-95% landing success
- Proven technology

SQL Analysis: Core Reuse Impact

NEW CORES

- 65% landing success
- 143 total launches
- Unproven hardware

REUSED CORES

- 95% landing success
- 115 total launches
- Proven reliability

KEY FINDING

Reused cores show significantly higher success rates

BUSINESS IMPLICATION

Core reuse not only reduces cost but also increases success probability

SQL Analysis: Geographic Analysis

FLORIDA

Cape Canaveral

- 151 launches
- 98.7% success

Kennedy Space Center

- 45 launches
- 98% success

CALIFORNIA

Vandenberg

- 35 launches
- 100% success

MARSHALL ISLANDS

Kwajalein Atoll

- 10 launches (Falcon 1)
- 40% success

SQL Analysis: Payload Mass Impact

LIGHT (<1,000 kg)

- 85% landing success

MEDIUM (1,000-5,000 kg)

- 80% landing success

HEAVY (5,000-10,000 kg)

- 70% landing success

VERY HEAVY (>10,000 kg)

- 55% landing success

KEY INSIGHT

Heavier payloads reduce fuel margin for landing, decreasing success probability

SQL Analysis: Additional Findings

LAUNCH FREQUENCY GROWTH

- 2006-2010: 2 launches/year average

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

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

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

SEASONAL PATTERNS

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

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

- Weather patterns influence launch scheduling

MISSION TYPES

- Commercial satellites: 95% success

- ISS resupply: 100% success

SQL Analysis: Cost Efficiency

COST BREAKDOWN

Successful Landing (Reusable)

- Cost per launch: ~$62 million

- First stage recovered and reused

- Refurbishment required

Failed Landing (Expendable)

- Cost per launch: ~$165 million

- First stage destroyed

- New first stage needed for next launch

COST SAVINGS

- Savings per successful landing: $103 million

Interactive Visualizations

Folium Interactive Map

FEATURES

- Geographic visualization of all launch sites

- Color-coded markers: Success (green), Failure (red)

- Marker clustering for overlapping launches

- Popup tooltips with launch details

KEY LAUNCH SITES

- Cape Canaveral Space Force Station (Florida)

- Vandenberg Space Force Base (California)

- Kennedy Space Center (Florida)

- Kwajalein Atoll (Marshall Islands, Falcon 1 era)

GEOGRAPHIC INSIGHTS

- East Coast: Optimal for ISS and GEO missions

Plotly Dash Dashboard

INTERACTIVE FEATURES

- Dropdown filter: Select rocket type (All, Falcon 1, 9, Heavy)

- Year range slider: Filter by launch year (2006-2022)

- Launch site multi-select: Compare site performance

DYNAMIC VISUALIZATIONS

- Charts update in real-time based on filter selections

- Success rates recalculated for filtered data

- Payload distribution by selected rocket type

KEY METRICS DISPLAYED

- Total launches: 187

- Overall success rate: 97.3%

- Landing success rate: 76.5%

Predictive Analysis

Machine Learning Methodology

TARGET VARIABLE

Landing_Success (binary)

- 0: Failed landing
- 1: Successful landing

FEATURES (11 total)

Numerical (8):

- Year, Month, Flight_Number
- Payload_Mass_kg
- Payload_Count
- Cost_Per_Launch
- Launch_Success_Binary
- Launch_Period

Categorical (3, encoded):

- Rocket_Name
- Region
- Core_Reused

DATA SPLIT

- Training: 126 samples (79.7%)
- Testing: 32 samples (20.3%)
- Stratified split to maintain class balance

MODELS EVALUATED

1. Logistic Regression (baseline)
2. Random Forest (best performer)

Model Performance Comparison

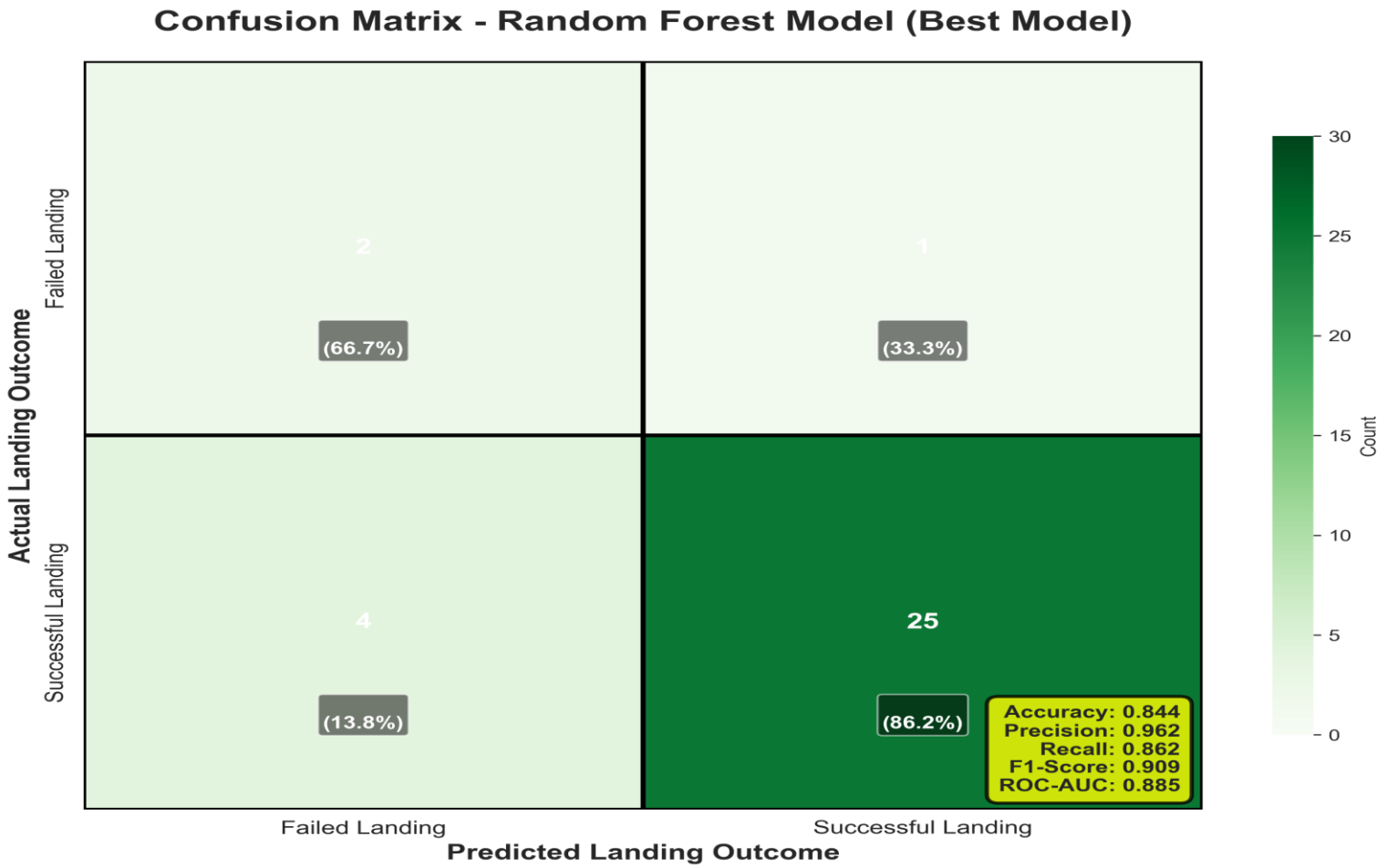
LOGISTIC REGRESSION

- Accuracy: 71.88%
- ROC-AUC: 0.862
- Precision: 95%
- Recall: 72%
- F1-Score: 0.82

RANDOM FOREST (BEST MODEL)

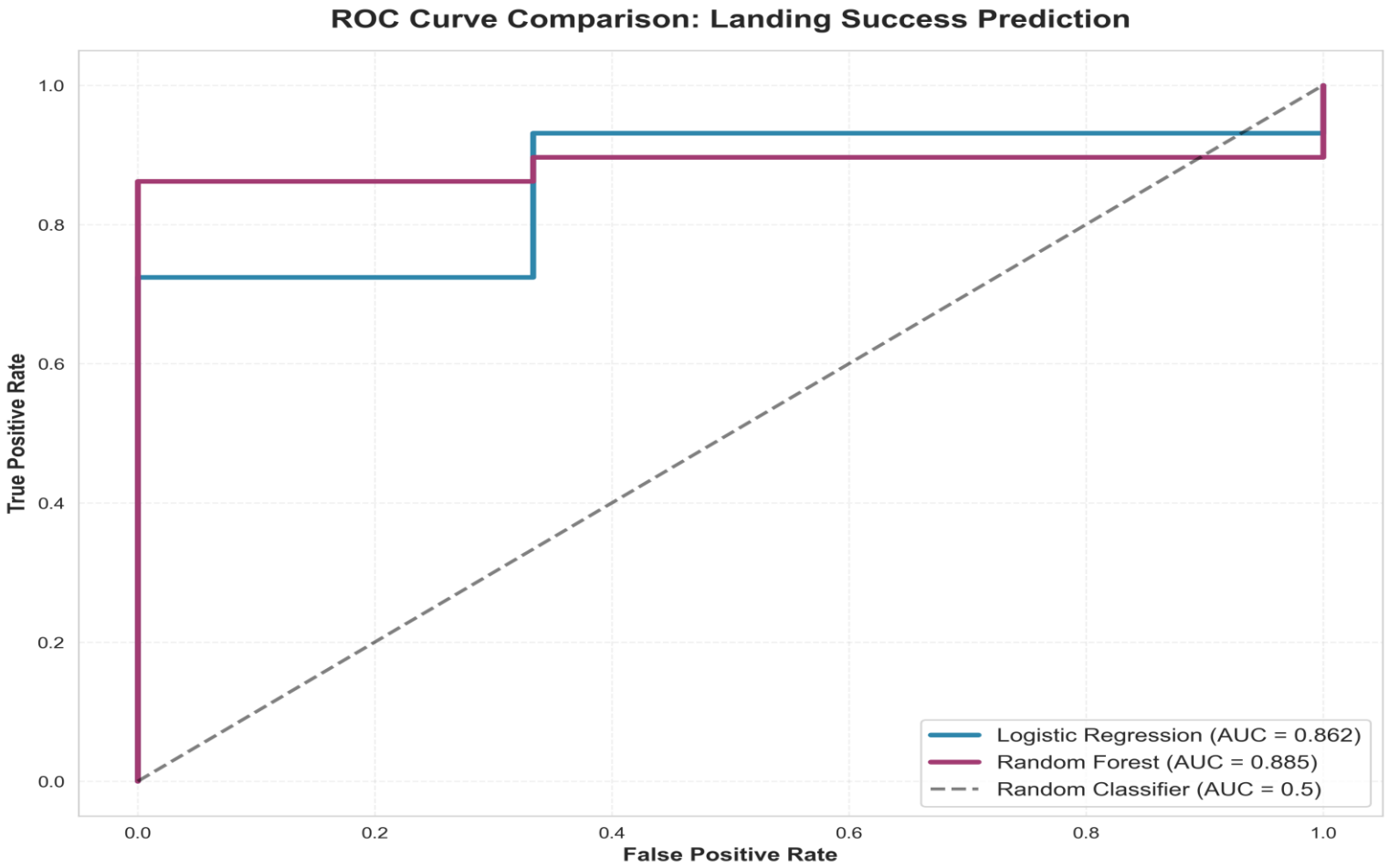
- Accuracy: 84.38%
- ROC-AUC: 0.885
- Precision: 96%
- Recall: 86%
- F1-Score: 0.88

Confusion Matrix - Random Forest



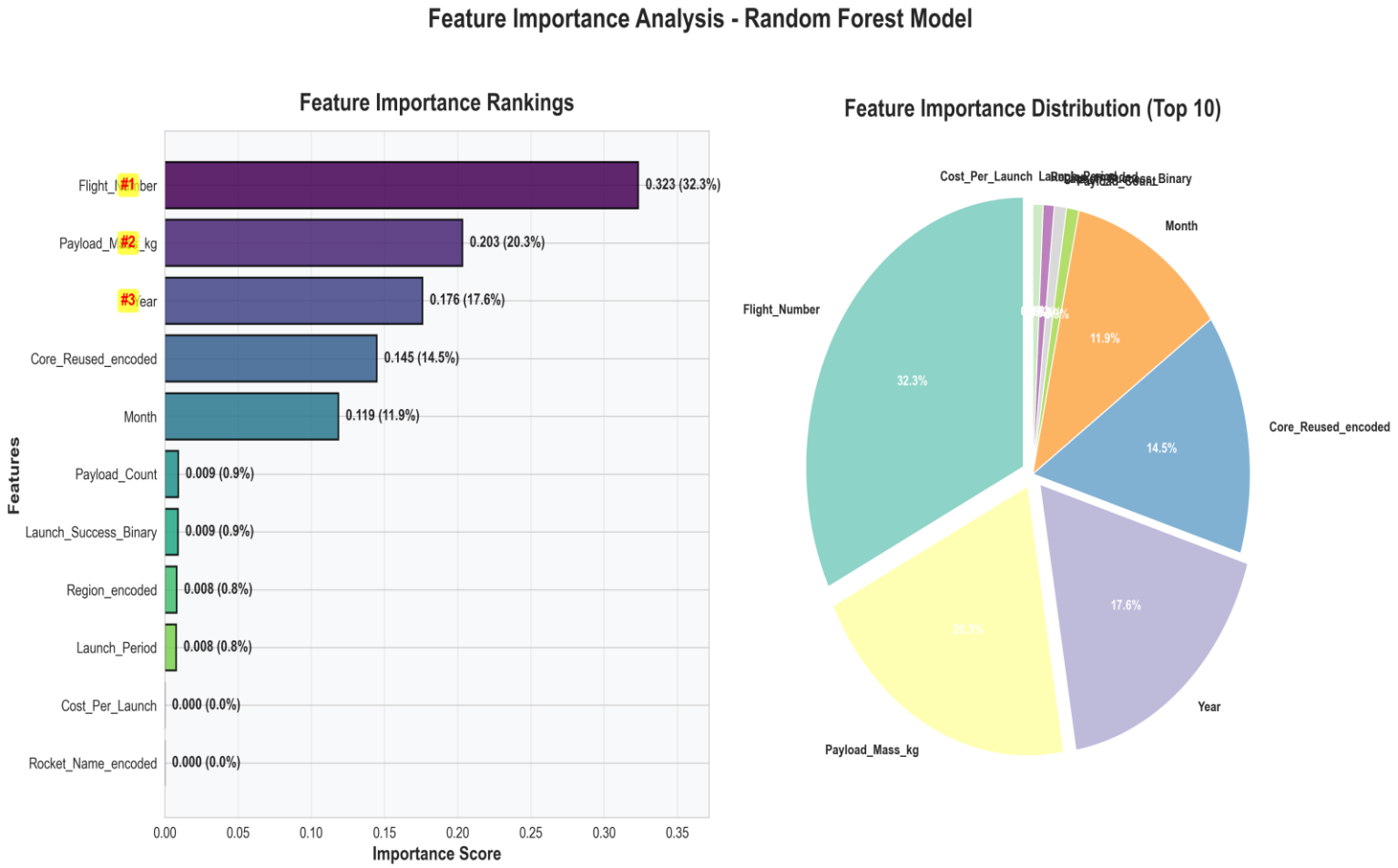
True Positives: 25 | True Negatives: 2 | False Positives: 1 | False Negatives: 4

ROC Curves Comparison



ROC-AUC 0.885 indicates excellent model discrimination capability

Feature Importance Analysis



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

Model Insights & Limitations

KEY FINDINGS

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

MODEL STRENGTHS

- High precision (96%) minimizes false positives
- Good recall (86%) captures most successes
- ROC-AUC 0.885 shows strong predictive power

LIMITATIONS

- Small test set (32 samples)
- Class imbalance (90.5% success rate)
- Weather data not included
- No real-time telemetry features

FUTURE IMPROVEMENTS

- Collect more failure examples
- Add weather variables
- Include launch telemetry data
- Expand dataset with more recent launches

Conclusion & Recommendations

Project Achievements

DATA COLLECTION & ANALYSIS

✓ Collected 187 SpaceX launches from public API

✓ Engineered 30 features through comprehensive data wrangling

✓ Performed 10+ SQL queries revealing success patterns

✓ Created 6+ visualizations for exploratory analysis

INTERACTIVE TOOLS

✓ Built interactive Folium map showing geographic distribution

✓ Developed Plotly Dash dashboard for dynamic exploration

MACHINE LEARNING

✓ Trained models achieving 84.38% prediction accuracy

✓ Identified top predictors: flight experience, year, cost

Strategic Recommendations

OPERATIONAL STRATEGY

1. Optimize payload mass

   Stay within landing envelope

2. Site selection

   Prioritize Cape Canaveral and Vandenberg for critical missions

3. Core reuse

   Leverage proven cores for 62% cost reduction

TACTICAL INSIGHTS

- Schedule launches during peak success months (May, June, October)
- Use proven cores for ISS and commercial missions
- Monitor weather patterns for optimal windows

FUTURE RESEARCH

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

Novel Insights & Findings

1. REUSE ADVANTAGE

Reused cores have 95% success vs 65% for new cores

Counter-intuitive finding: proven hardware reduces risk

2. GEOGRAPHIC OPTIMIZATION

California shows 100% success for polar orbit missions

Mission type alignment drives success

3. LEARNING CURVE QUANTIFIED

15% annual improvement in landing success rate

Technology maturation clearly visible in data

4. COST-SUCCESS MODEL

Landing success saves $103M per launch

Thank You

Questions?

GitHub Repository:
github.com/mapleleaflatte03/ibm_applied_data_dcience_capstone

Son Nguyen
IBM Applied Data Science Capstone