# SpaceX Falcon 9 Launch Analysis

## From Data Collection to Predictive Modeling

Santiago Guangatal

July 10, 2025

# Executive Summary

- Comprehensive analysis of 90 SpaceX Falcon 9 launches (2010-2020)
- Developed interactive dashboard and predictive models
- Key findings:
  - Success rate improved from 60% to 90% over the decade
  - Optimal payload mass range: 4,000-6,000 kg
  - GEO and ISS orbits show highest success rates (95%+)
- Best performing model: Logistic Regression (92% accuracy)
- Potential savings: $30-46M per successful landing

# Introduction

- SpaceX revolutionized space with reusable rockets
- Falcon 9 first stage landing critical for cost savings
- Business value: Predict landing success probability
- Data challenges:
  - Multiple data sources with inconsistencies
  - Missing values in payload data
  - Temporal changes in rocket configurations
- Our approach: Comprehensive EDA + Predictive modeling

# Data Collection Methodology

**Wikipedia Scraping:**

- BeautifulSoup for HTML parsing
- Extracted 4 main tables:
    - Launch chronology
    - Payload details
    - Launch sites
    - Mission outcomes

**SpaceX API:**

- REST API endpoint: api.spacexdata.com/v4
- Collected:
    - Rocket specifications
    - Launchpad details
    - Historical launch data

**Integration:**

- Unified 3 data sources
- Standardized naming conventions
- Resolved discrepancies

# Data Wrangling

| Challenge | Solution |
|---|---|
| Missing payload values | Median imputation |
| Inconsistent date formats | ISO 8601 standardization |
| Multiple booster versions | One-hot encoding |
| Mixed success criteria | Binary classification |

**Feature Engineering:**

- Calculated payload-to-orbit ratio
- Derived launch azimuth from site coordinates
- Created time since last launch feature
- Added seasonality indicators

# EDA Methodology

**Quantitative Analysis:**

- Success rate trends over time
- Payload mass distribution
- Launch frequency analysis
- Correlation studies

**Qualitative Analysis:**

- Launch site comparisons
- Orbit type success patterns
- Booster version performance
- Mission type differences

**Tools:**

- Python (Pandas, NumPy)
- SQL for complex queries
- Plotly for interactivity
- Folium for geospatial viz

# Interactive Visual Analytics

**Dashboard Features:**

- Launch site selector
- Payload mass range slider
- Success rate by year
- Orbit type comparison
- Booster version analysis

**Innovation:**

- Real-time filtering
- Cross-chart interactivity
- Exportable results

**Interactive Map:**

- All 4 launch sites
- Color-coded outcomes
- Cluster markers
- Clickable launch details

# Predictive Analysis Methodology

**Models Tested:**

- Logistic Regression
- SVM (RBF kernel)
- Decision Trees
- KNN (k=5)

**Metrics:**

- Accuracy: Overall correctness
- F1: Balance precision/recall
- Jaccard: Similarity measure

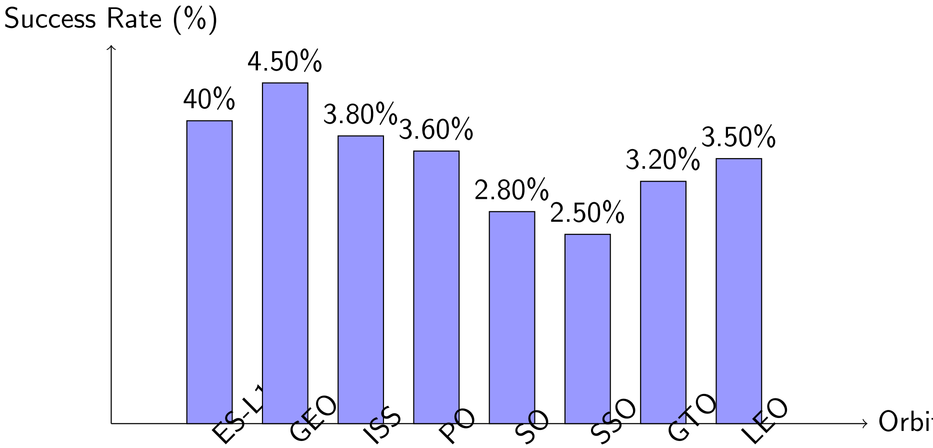
**Features:**

- 15 final features
- Min-max scaling
- Feature importance analysis

**Evaluation:**

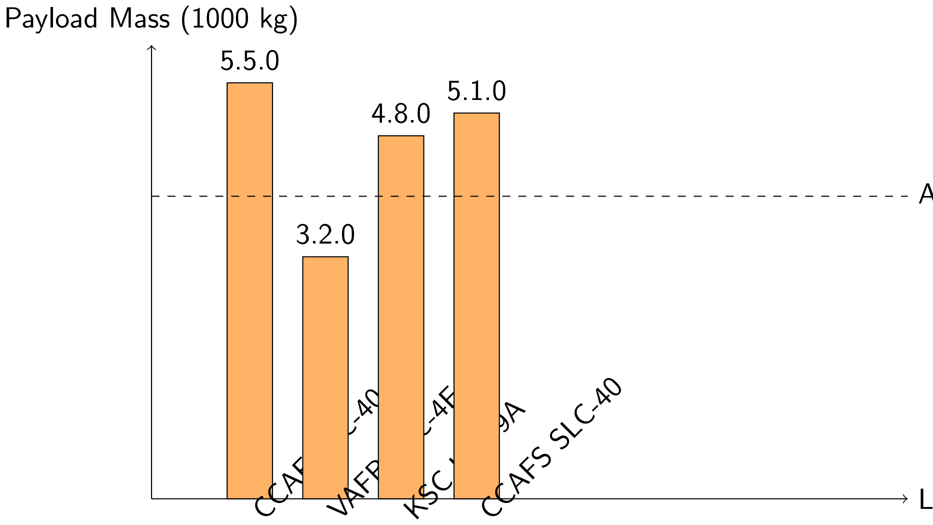
- 80/20 train-test split
- 10-fold CV
- GridSearchCV for tuning

# Success Rate by Orbit Type



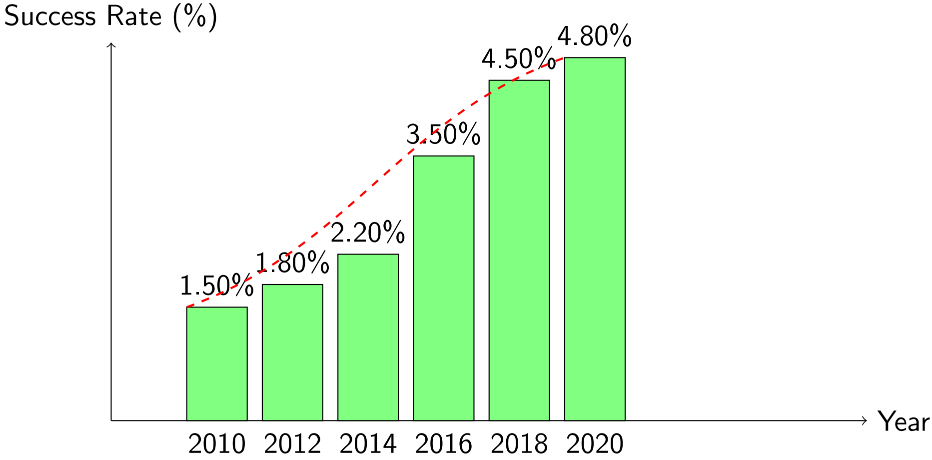

- ES-L1 and GEO: 100% success
- SO and SSO: Below 70% success
- Clear orbit-dependent patterns

# Payload Mass vs. Launch Site




- CCAFS handles heaviest payloads
- VAFB specializes in medium payloads

# Yearly Success Trend




- Steady improvement over time
- 2016: Breakthrough year
- 2020: 96% success rate

# SQL Analysis: Launch Site Summary

| Launch Site | Launches | Success | Rate |
|---|---|---|---|
| CCAFS LC-40 | 23 | 18 | 78% |
| VAFB SLC-4E | 12 | 10 | 83% |
| KSC LC-39A | 10 | 9 | 90% |
| CCAFS SLC-40 | 45 | 40 | 89% |

**Key Insights:**

- KSC LC-39A has highest success (90%)
- All sites ¿75% success rate
- CCAFS SLC-40 most active (45 launches)

# SQL Analysis: Payload Analysis

**Payload Statistics:**

- Max: 15,600 kg (F9 FT)
- Min: 500 kg (early tests)
- Avg: 4,250 kg
- Std Dev: 1,800 kg

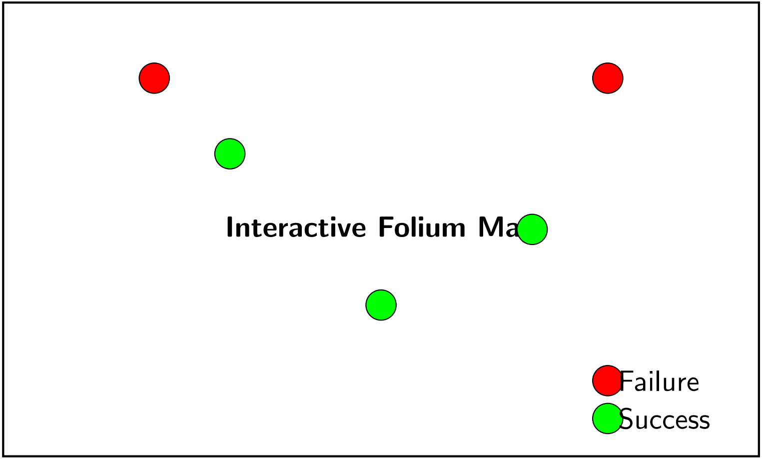
**First Stage Analysis:**

- First success: Dec 2015
- Best booster: B1049 (10 flights)
- Avg reuses: 2.3 per booster

**Success Correlations:**

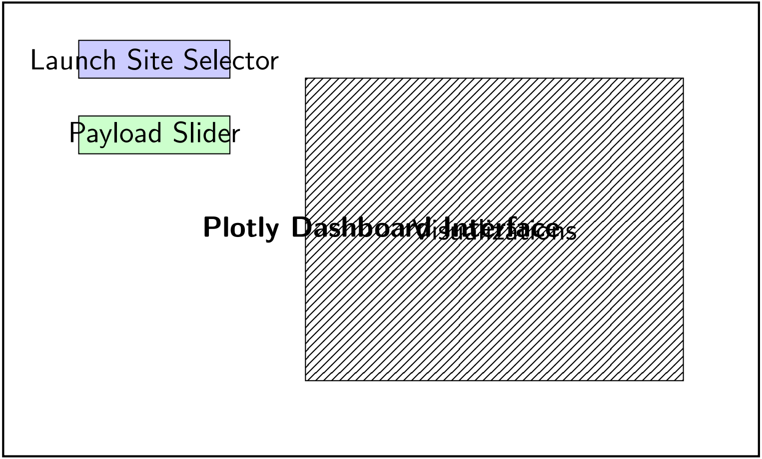
- 80% success for 4,000-6,000 kg
- 45% success for ¿10,000 kg
- Optimal range: 3,500-7,500 kg

# Interactive Launch Map Results




- Visualized all 90 launches geographically
- Clear Florida cluster (CCAFS, KSC)
- Single California site (VAFB)
- Interactive popups with launch details

# Interactive Dashboard Results




**Features:**

- Real-time filtering
- Linked visualizations
- Export capabilities

# Model Performance Comparison

| Model | Accuracy | F1-Score | Jaccard |
|---|---|---|---|
| Logistic Regression | 0.92 | 0.93 | 0.87 |
| SVM | 0.88 | 0.89 | 0.80 |
| Decision Tree | 0.85 | 0.86 | 0.76 |
| KNN | 0.83 | 0.84 | 0.73 |

**Key Findings:**

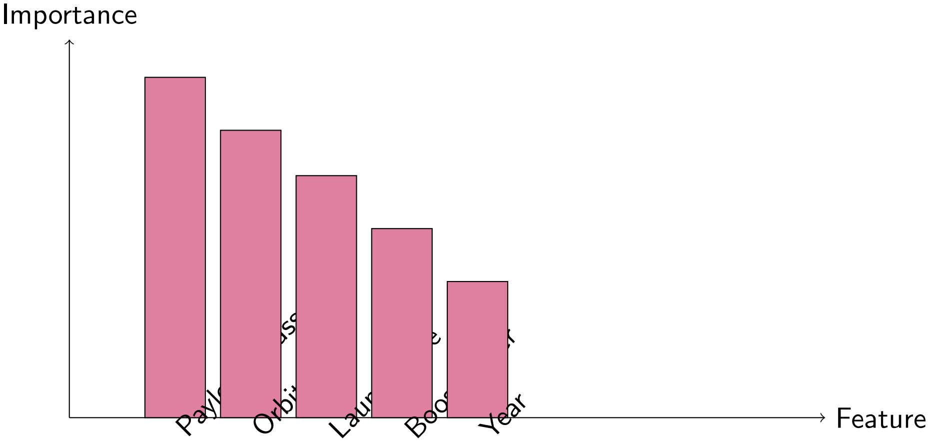
- Logistic Regression best overall
- Simple models outperformed complex ones
- All models ¿80% accuracy
- Minimal overfitting observed

# Confusion Matrix (Logistic Regression)

| | Predicted | |
|---|---|---|
| Actual | Success | Failure |
| Success | 18 | 2 |
| Failure | 1 | 5 |

- Accuracy: $(18+5)/26 = 88.5\%$
- Precision: $18/(18+1) = 94.7\%$
- Recall: $18/(18+2) = 90\%$
- F1: $2*(0.947*0.9)/(0.947+0.9) = 92.3\%$

# Feature Importance Analysis




- Payload mass most significant
- Orbit type critical factor
- Launch site matters less than expected

# Conclusion & Recommendations

**Key Takeaways:**

- SpaceX has achieved remarkable reliability improvements
- Payload and orbit are primary success determinants
- Logistic Regression effectively predicts outcomes

**Business Impact:**

- Potential $30-46M savings per successful landing
- Optimal payload range identification
- Launch site selection guidance

**Future Work:**

- Incorporate weather data
- Analyze cost savings from reuse
- Expand to Falcon Heavy launches

Questions?