IBM Developer
SKILLS NETWORK

# SpaceX Falcon 9
# First Stage Landing Prediction:

## A Data Science Approach

Sumanth Bhaskar Hegde
20 March 2025

# Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Interactive Visualisation
- Predictive Analysis
- Conclusion
- Appendix

# Executive Summary

**Project Goal:**
Predict successful Falcon 9 first stage landings to optimize launch costs.

**Key Methodologies:**
- Data Collection: SpaceX API, Web Scraping
- EDA: Visualization (Seaborn, Matplotlib), SQL
- Interactive Analysis: Folium Maps, Plotly Dash
- Predictive Modeling: Logistic Regression, SVM, Decision Tree, KNN

**Key Findings:**
- Launch success improves over time.
- Launch site & orbit significantly affect landing outcomes.
- Decision Tree slightly outperformed other models.

**Value:**
Insights can inform launch strategy and cost reduction.

# Introduction

**The Space Race 2.0:**
Driven by commercial interests, reusability is key.

**SpaceX's Advantage:**
Falcon 9's reusable first stage lowers launch costs significantly (~$62M vs. $165M).

**Project Goal:**
Develop a model to predict successful first stage landings.
Insights enable cost optimization and competitive bidding.

**Key Questions:**
- How do payload, launch site, and orbit affect landing success?
- Can we accurately predict landing outcomes?
- What are the most influential factors?

Section 1

# Methodology

# Methodology

| Step | Description | Tools/Techniques |
|---|---|---|
| Data Collection | Gathered launch data from two sources. | SpaceX API, Web Scraping |
| Data Wrangling | Cleaned, transformed, and prepared the data for analysis. Created the 'Class' label. | Pandas |
| EDA | Explored data patterns and relationships. | Seaborn, Matplotlib, SQL |
| Interactive Analysis | Created interactive visualizations for exploration. | Folium Maps, Plotly Dash |
| Predictive Analysis | Built and evaluated classification models to predict landing success. Tuned hyperparameters with GridSearchCV. | Logistic Regression, SVM, Decision Tree, KNN, GridSearchCV |

# Data Collection

**Objective:**
Gather comprehensive SpaceX launch data.

**Sources:**
- ***SpaceX API:***
  - Direct access to structured launch data.
  - Retrieved details on rockets, payloads, launchpads, and cores.
- ***Web Scraping (Wikipedia):***
  - Collected historical launch records.
  - Required parsing HTML tables.

**Key Data Points:**
- Booster Version, Payload Mass, Orbit, Launch Site
- Outcome, Landing Type, Core Details.

**Challenges:**
- Data inconsistencies in web scraping source.
- Data Structure is nested.

# Data Wrangling

**Objective:** Prepare raw data for analysis & modeling.

**Key Actions:**
- Handle Missing Values
- Convert Data Types
- Feature Engineering
- Data Transformation

**Details:**
- Missing Values: Addressed in Payload Mass & Landing Pad columns.
- "Class" Label: Created binary target (1=Success, 0=Failure).
- One-Hot Encoding: Applied to Orbit, Launch Site, etc.
- Standardization: Scaled numerical features.

**Outcome:** Clean & preprocessed dataset, ready for next steps.

# EDA with Data Visualization

Key Visuals:

Flight Number vs. Launch Site:
> Launch frequency per site.

Payload Mass vs. Launch Site:
> Typical payload ranges.

Success Rate vs. Orbit Type:
> Orbits with higher success.

Launch Success Yearly Trend:
> Improving success over time.

Tools: Seaborn, Matplotlib

# EDA with SQL

We used SQL to interrogate the SpaceX dataset, focusing on key areas to understand mission characteristics and outcomes. Some key aspects are:

**Launch Sites:** We began by identifying all unique launch locations used by SpaceX.

**Launch Sites (CCA):** To further understand launch locations, we listed five records where the launch site names began with the characters "CCA".

**NASA Payload:** SQL allowed us to determine the total payload mass carried by boosters launched on NASA (CRS) missions. This helps quantify SpaceX's support for critical resupply efforts.

**Average Payload (F9 v1.1)**: We also determined the average payload mass carried by the F9 v1.1 booster version, to understand the booster's payload capabilities.

**First Ground Landing:** We pinpointed the date of the first successful landing outcome on a ground pad, a key milestone in SpaceX's reusability program.

# Build an Interactive Map with Folium

- **Global Launch Site Overview:** All launch sites were marked on a world map to show proximity to the equator and coastlines. Proximity to water may be important to landing success for safety, and proximity to the equator provides a natural boost to the launch.

- **Success/Failure by Launch Site:** Markers were color-coded to indicate launch success/failure rates, helping visualize high-performing locations. Marker clusters were used to manage overlapping points and declutter the display.

- **Site Proximity Analysis:** We calculated and displayed distances from launch sites to nearby features (coastlines, railways, highways), using MousePosition to help determine coordinates. We marked on the selected closest coastline point using the polyline between launch site and coastline.

# Build a Dashboard with Plotly Dash

The Plotly Dash dashboard allows users to:

- **Select a Launch Site:**
  > Examine launch records for a specific site.

- **Filter by Payload Range:**
  > Analyze the correlation between payload mass and success.

- **Visualize Success Rates:**
  > See successful and unsuccessful launches as percentages.

Key Visuals:

1. Pie chart (success rate by site)
2. Scatter plot (payload vs. outcome, colored by booster version)

# Predictive Analysis (Classification)

**Models Evaluated:**
- Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN).

**Key Steps:**
- Engineered the "Class" (landing success)
- Split data to test and train.
- Standardized all the data.
- Used Grid Search for optimization.

**Results:**
- All models achieved similar test accuracy (~83%).
- Decision Tree had a slight edge on validation.
- False positives remain a challenge.

**Insights:** Model performance is limited by dataset size, and further feature engineering might improve accuracy and help address false positives.

# Results

Our analysis reveals several factors influencing SpaceX Falcon 9 first-stage landing

**Outcomes:**

- Launch success has improved over time, but the landing outcome remains complex.
- Launch site and orbit type are key factors, and more success comes with sites closer to the equator, and coasts.
- Interactive visualization (Folium, Plotly Dash) enables dynamic exploration of launch data.

**Predictive Modeling:**

- Several classification models (Logistic Regression, SVM, Decision Tree, KNN) achieved similar test accuracy (~83%). This accuracy may be linked to the nature of the dataset
- Decision Tree Model is better than the other three models.

Section 2

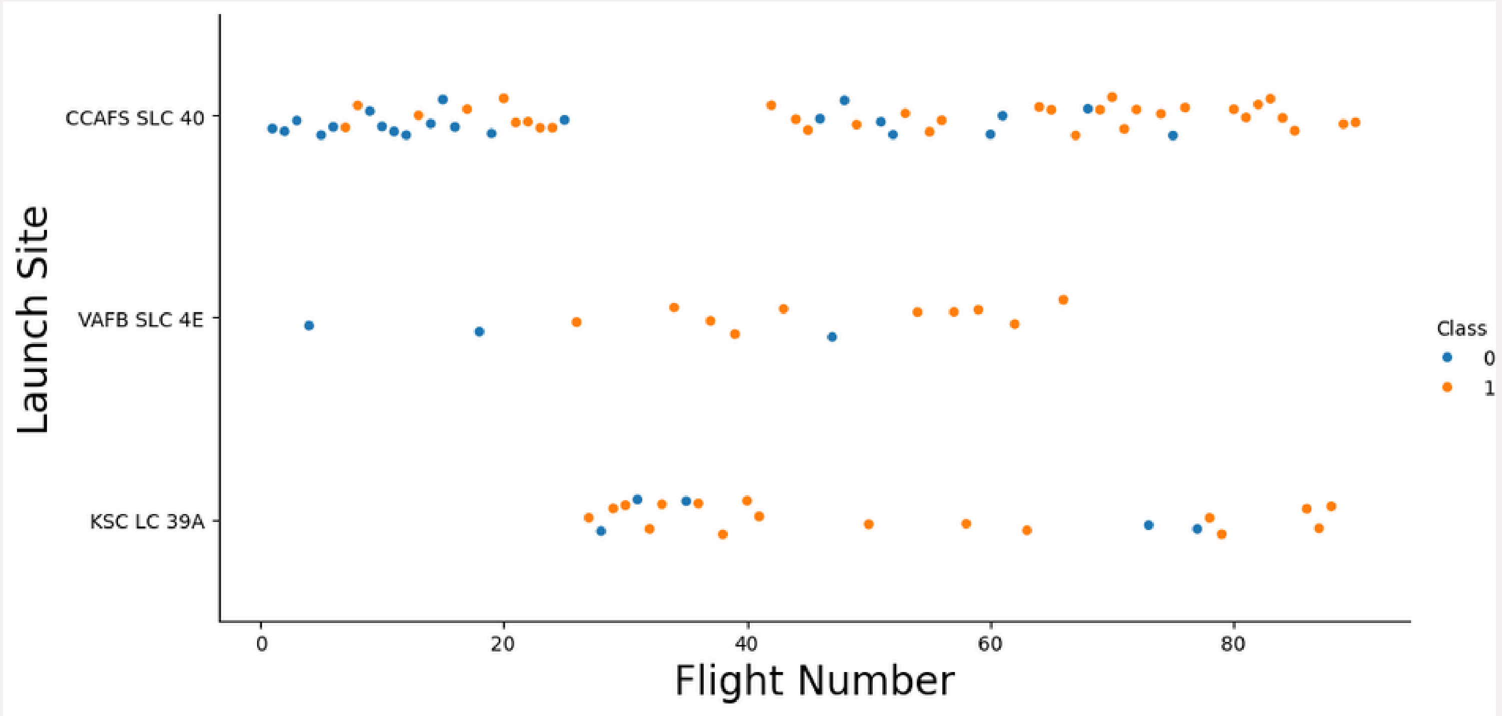
# Insights drawn from EDA

# Flight Number vs. Launch Site

**Key Finding:**
- CCAFS SLC 40 was more commonly used in the initial launches, and is used less in later launches. This data provides information that it had more launches in early days as compared to today.

**Visual:**
- Scatter plot of Flight Number vs. Launch Site

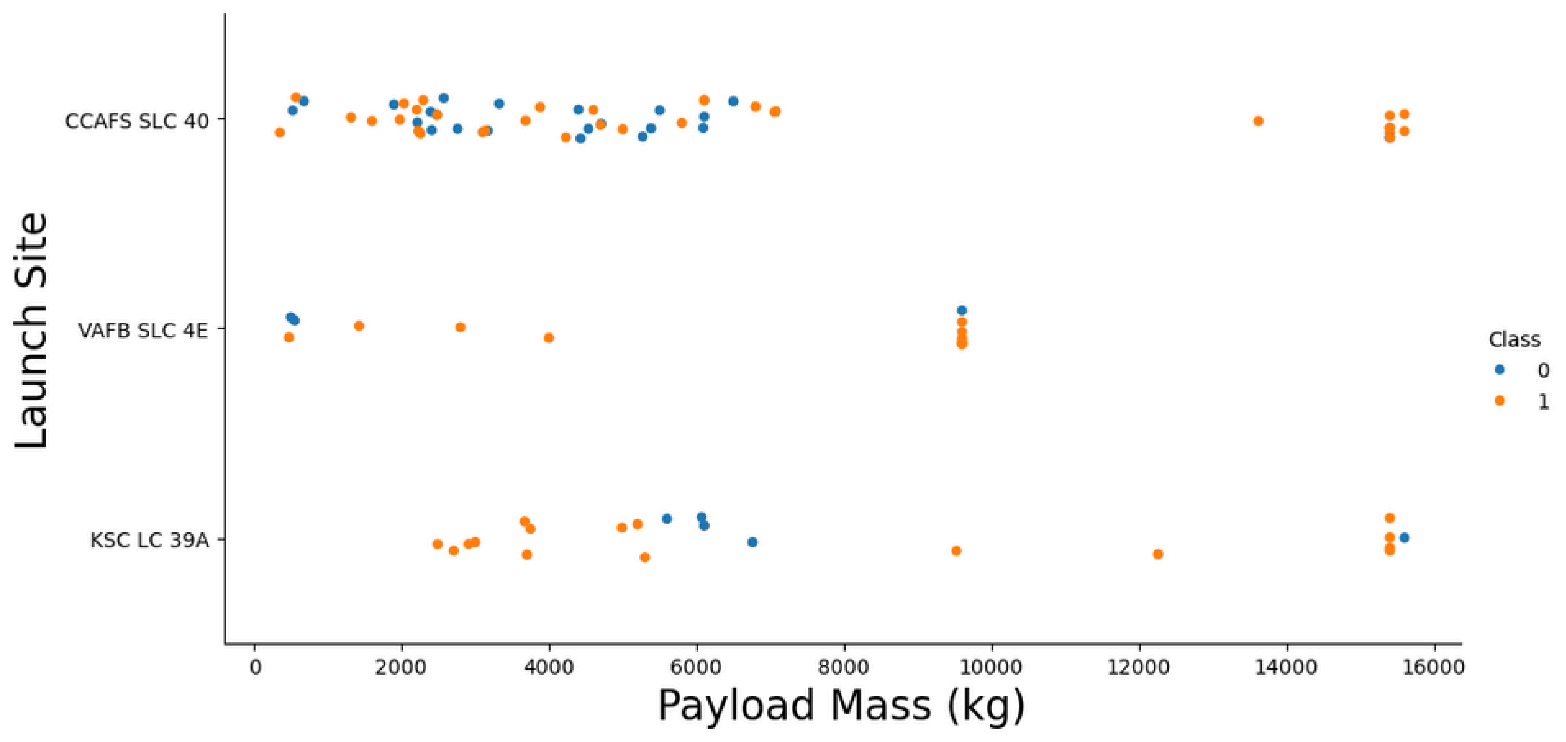
This visual shows the distribution of launches across different launch sites over time.



# Payload vs. Launch Site

**Key Finding:**
- No launch sites has used heavy mass with greater than 10000 Payload Mass. The visual suggests there is a limited launch of high mass.

**Visual:**
- Scatter plot of Payload Mass vs. Launch Site



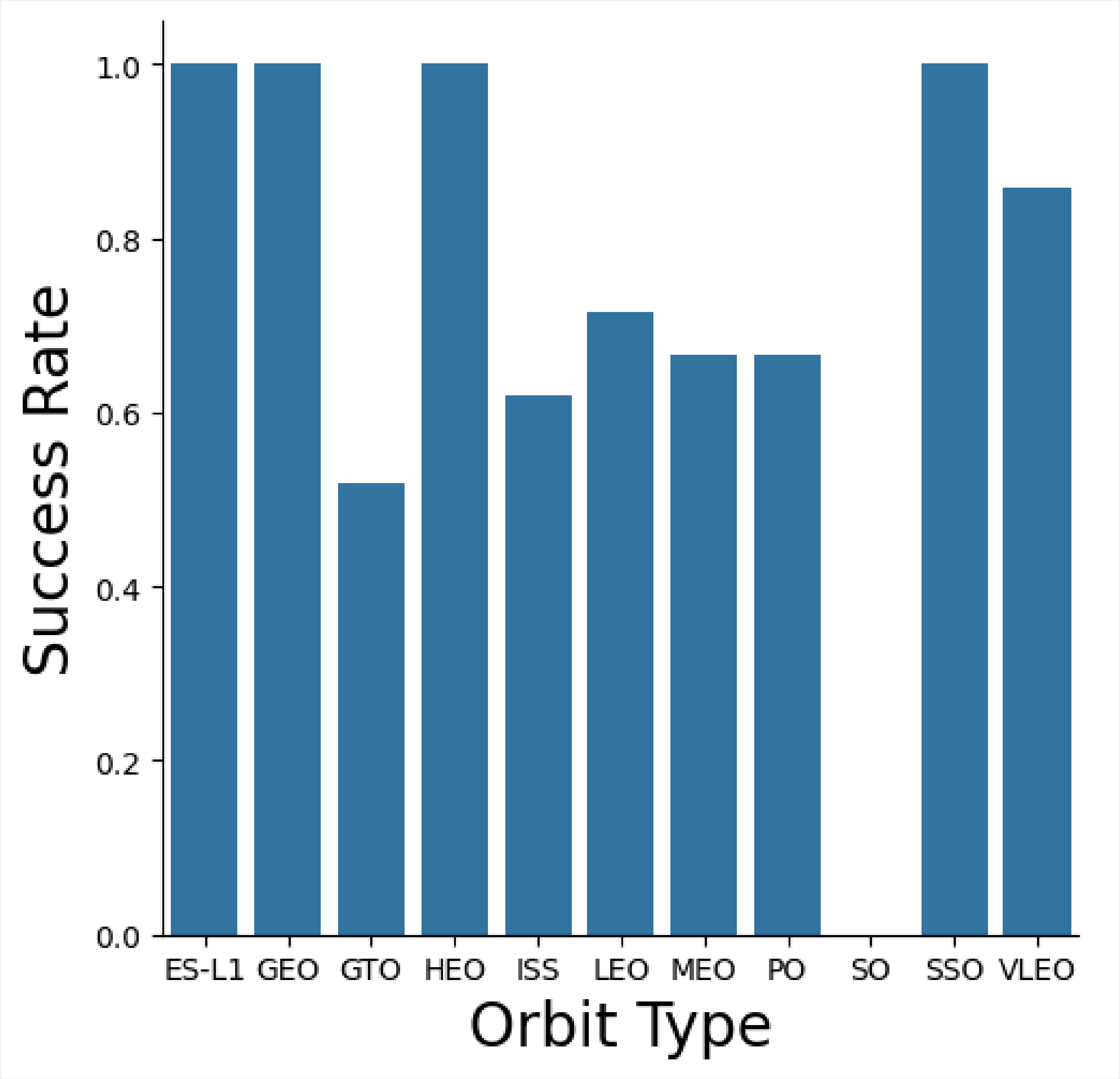

This visual maps payload mass to each launch site.

# Success Rate vs. Orbit Type

**Key Finding:**
- Orbits ES-L1, GEO, HEO, and SSO had a 100% success rate. Other Orbits were low. The orbits with high rate suggest, those orbits are easily maintained.  success orbit.

**Visual:**
Bar chart of Success Rate vs. Orbit Type

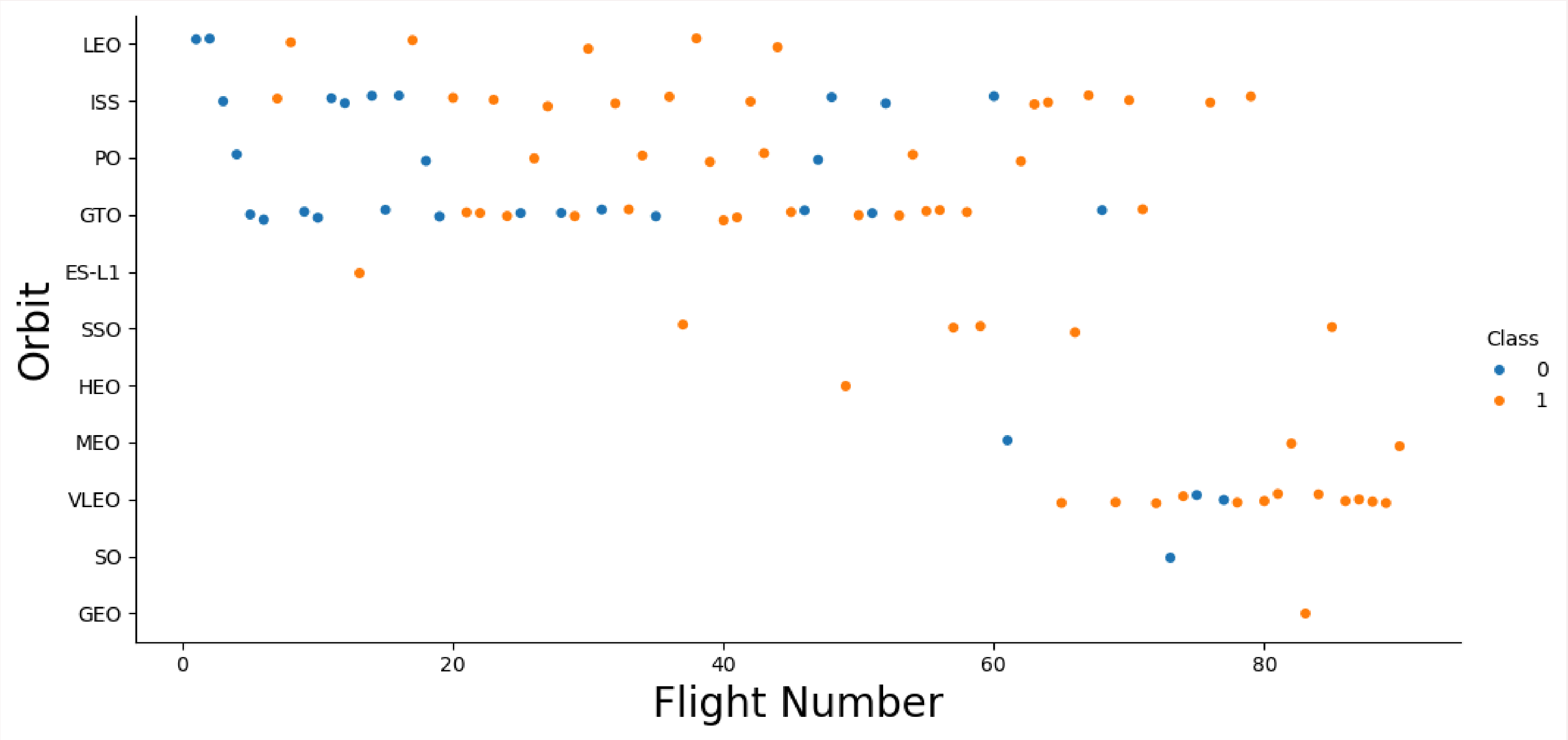
This visual identifies high success orbit.



# Flight Number vs. Orbit Type

**Key Finding:**
- The success rate typically increases with the number of flights for each orbit. There is a trend in orbit.

**Visual:**
- Scatter plot of Flight Number vs. Orbit Type

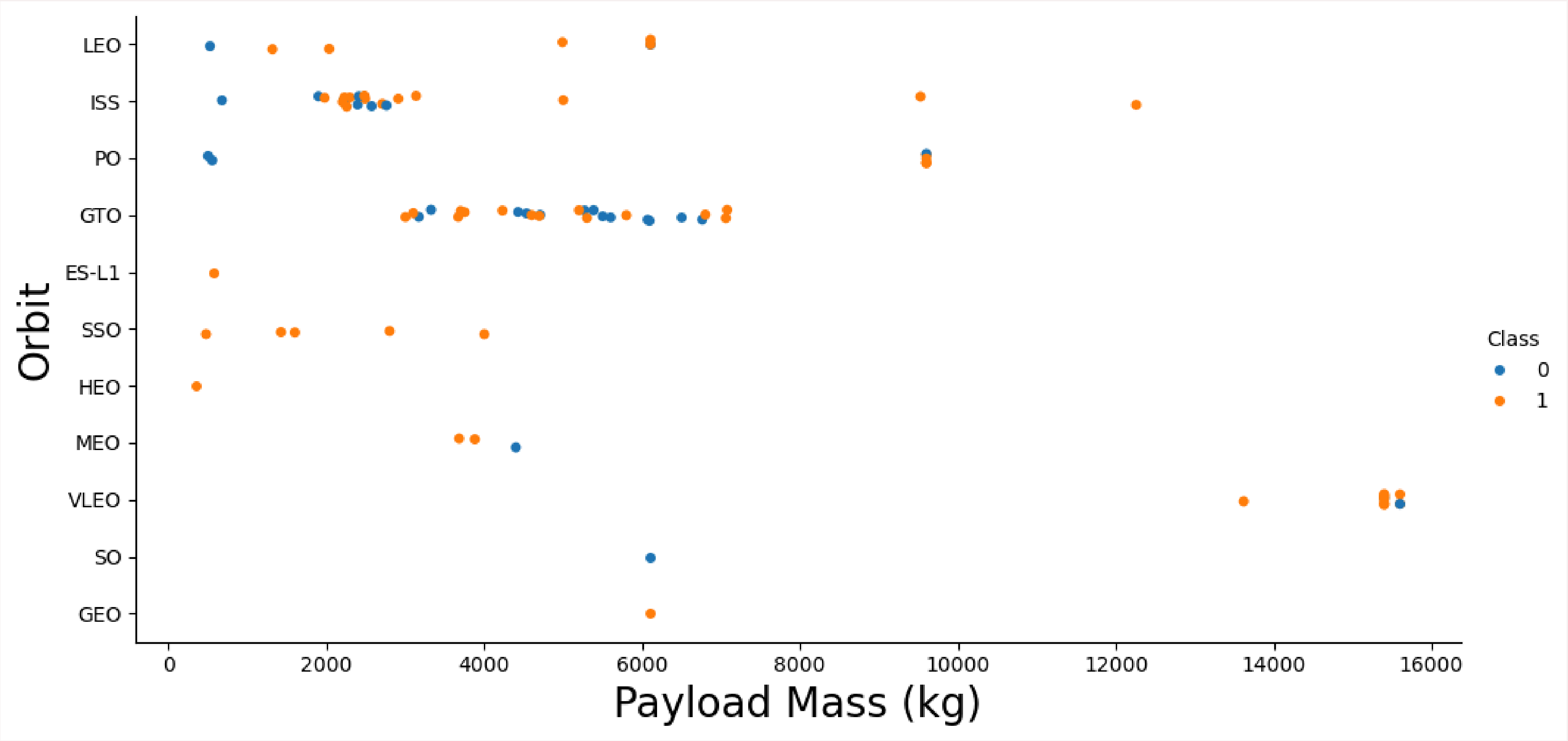



This plot demonstrates if orbits become more reliable over time.

# Payload vs. Orbit Type

**Key Finding:**
- Orbits LEO, ISS and PO were the only orbits to have heavy payloads. This shows what type of orbits can carry more mass.

**Visual:**
- Scatter plot of Payload Mass vs. Orbit Type




This chart maps payload capacity to orbit type.

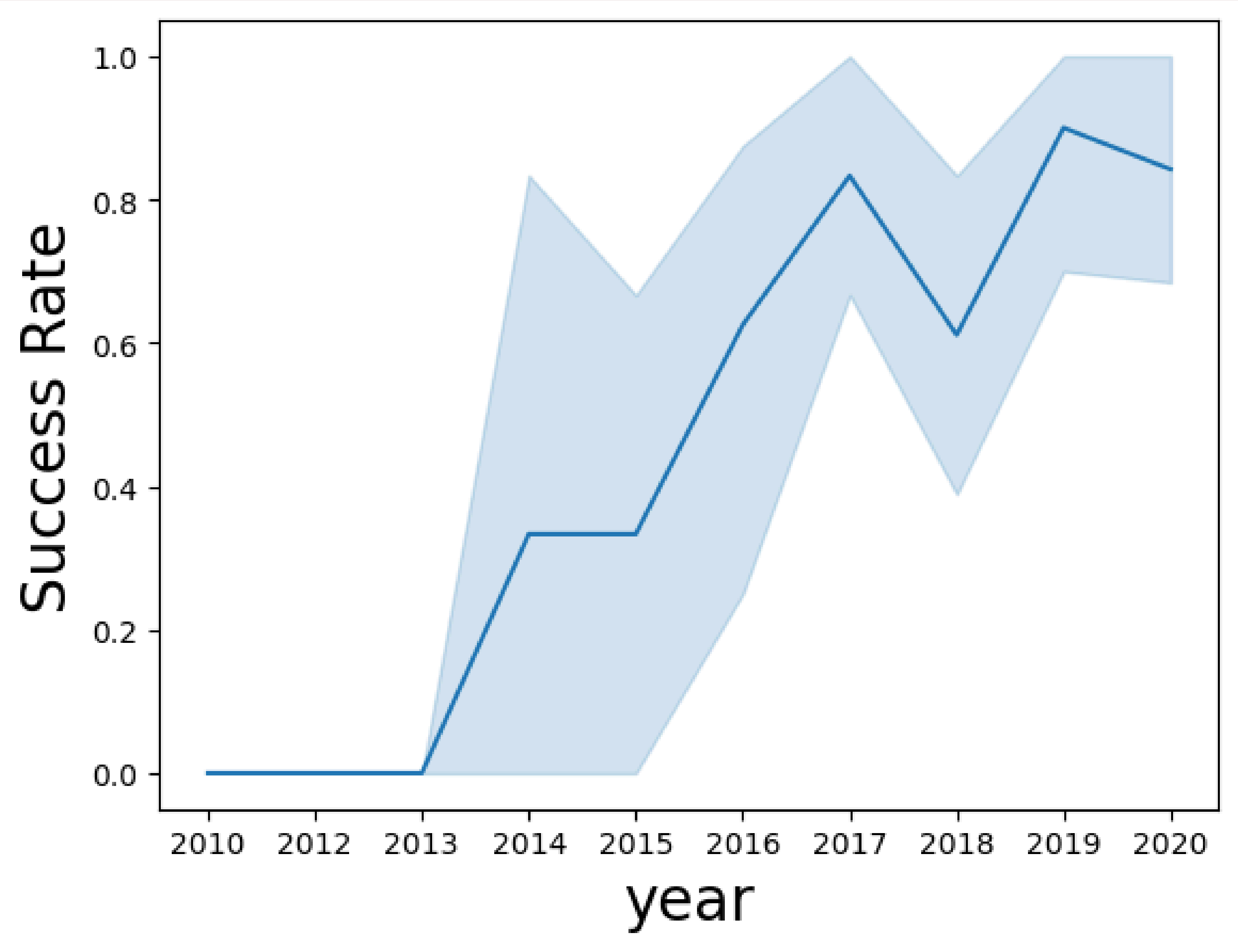
# Launch Success Yearly Trend

**Key Finding:**
-Overall, the success rate has improved since 2013. There is an obvious trend that it is improving.

**Visual:**

- Line chart of Yearly Success Rate

This graph tracks the success rate trajectory.



# All Launch Site Names

| Launch Site | Description |
|---|---|
| CCAFS LC-40 | Cape Canaveral Air Force Station Launch Complex 40 |
| CCAFS SLC-40 | Cape Canaveral Air Force Station Space Launch Complex 40 |
| KSC LC-39A | Kennedy Space Center Launch Complex 39A |
| VAFB SLC-4E | Vandenberg Air Force Base Space Launch Complex 4E |

# Launch Site Names Begin with 'CCA'

Objective: Display the launch sites beginning with "CCA."

Key Launch Sites:
- CCAFS LC-40
- CCAFS SLC-40

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 2010-12-08 | 15:43:00 | F9 v1.0 B0004 | CCAFS LC-40 | Dragon demo flight C1, two CubeSats, barrel of Brouere cheese | 0 | LEO (ISS) | NASA (COTS) NRO | Success | Failure (parachute) |
| 2012-05-22 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 0:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

Highlights:
- The query returns 5 records, but the sites are not unique. These sites have the code name CCA.

# Total Payload Mass

Objective: Calculate the total payload mass carried on NASA (CRS) missions.

Total Payload: 45596 kg

| SUM(PAYLOAD_MASS__KG_) |
| --- |
| 45596 |

Significance: This number reflects SpaceX's contribution to NASA's cargo resupply efforts, and what it carries to support space exploration.

# Average Payload Mass by F9 v1.1

Objective: Determine typical payload for F9 v1.1 launches.

Average Payload: 2928.4 kg

Significance: This average helps define the typical payload capacity of the F9 v1.1 booster, helping understand the specifications.

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

# First Successful Ground Landing Date

Objective: Identify the date of SpaceX's first ground landing.

Landing Date: 2015-12-22

Significance: This marked a key milestone in reusable rocket technology, which can reduce the total rocket cost.

| FIRST DATE |
|---|
| 2015-12-22 |

## Successful Drone Ship Landing with Payload between 4000 and 6000

Objective: Identify boosters successfully landing on drone ships with specific payload.

Boosters: JCSAT-14, JCSAT-16, SES-10, SES-11 / EchoStar 105

Significance: These missions demonstrate SpaceX's ability to land boosters on drone ships, for specific payload weights.

| Payload |
|---|
| JCSAT-14 |
| JCSAT-16 |
| SES-10 |
| SES-11 / EchoStar 105 |

# Total Number of Successful and Failure Mission Outcomes

Objective: Show distribution of mission outcomes (success vs. failure).

Outcomes:
- Success: 99
- Failure (in flight): 1
- Success (payload status unclear): 1

Significance: This highlights high mission success, and failures are very low, demonstrating reliability.

| Mission Outcome | total_number |
|---|---|
| failure (in flight) | 1 |
| success | 99 |
| success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

Objective: List boosters that carried the maximum payload mass.

Boosters: F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3, F9 B5 B1056.4, F9 B5 B1048.5, F9 B5 B1051.4, F9 B5 B1049.5, F9 B5 B1060.2, F9 B5 B1058.3, F9 B5 B1051.6, F9 B5 B1060.3, F9 B5 B1049.7

Significance: Highlighting all the boosters with a large payload.

| Booster_Version |
|---|
| F9 B5 B1048.4 |
| F9 B5 B1049.4 |
| F9 B5 B1051.3 |
| F9 B5 B1056.4 |
| F9 B5 B1048.5 |
| F9 B5 B1051.4 |
| F9 B5 B1049.5 |
| F9 B5 B1060.2 |
| F9 B5 B1058.3 |
| F9 B5 B1051.6 |
| F9 B5 B1060.3 |
| F9 B5 B1049.7 |

# 2015 Launch Records

Objective: Show drone ship landing failures in 2015.

Details:
- Month: 01, Date: 2015-01-10, Booster: F9 v1.1 B1012, Site: CCAFS LC-40, Outcome: Failure (drone ship)
- Month: 04, Date: 2015-04-14, Booster: F9 v1.1 B1015, Site: CCAFS LC-40, Outcome: Failure (drone ship)

Significance: This gives important context to when and where the incidents occurred.

| Month | Date | Booster_Version | Launch_Site | Landing_Outcome |
|---|---|---|---|---|
| 01 | 2015-01-10 | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| 04 | 2015-04-14 | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

# Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Objective: Show a ranked list of landing outcomes during the period.

Top Outcomes:
1. Success: 20
2. No attempt: 10
3. Success (drone ship): 8

Significance: During that time the success are high, but attempts are low, so this information is important.

| Landing_Outcome | count_outcomes |
|---|---|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 2 |
| Precluded (drone ship) | 1 |

Section 3

# Launch Sites Proximities Analysis

# Launch Site Location Overview

**The map shows...**

- Proximity: All sites are near the coast. Access to water is important.
- Latitude: There's a clustering towards the equator. Boost comes from rotation.

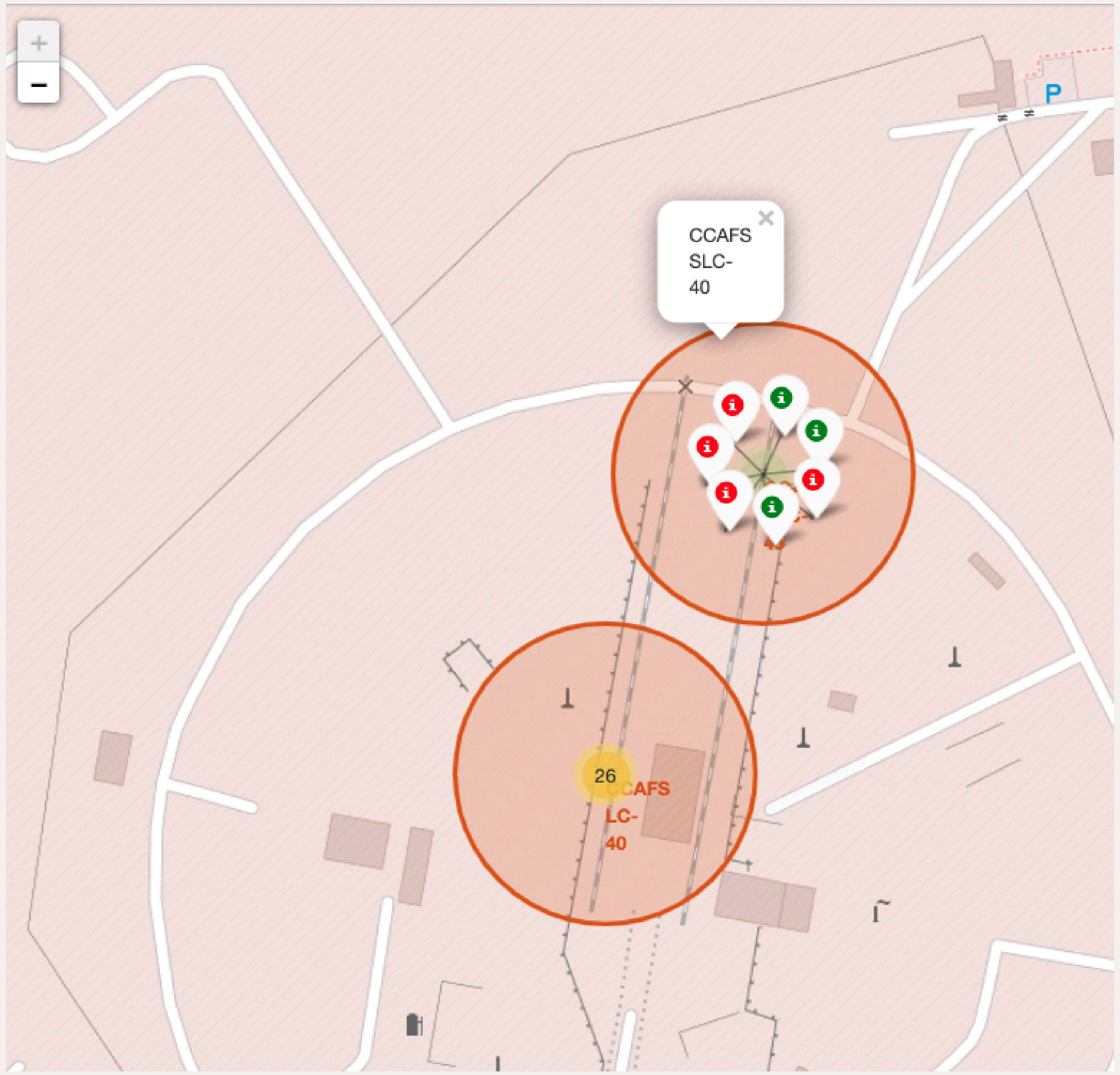

Location matters! We want to keep to the aspects needed to consider is location and why it matters.

# Visualizing Launch Success by Site

**The map shows...**

- CCAFS SLC-40 Success: The site has more red markers than the rest, so it indicates a mix of success.

Analysis of launch outcomes makes each site's value visible.

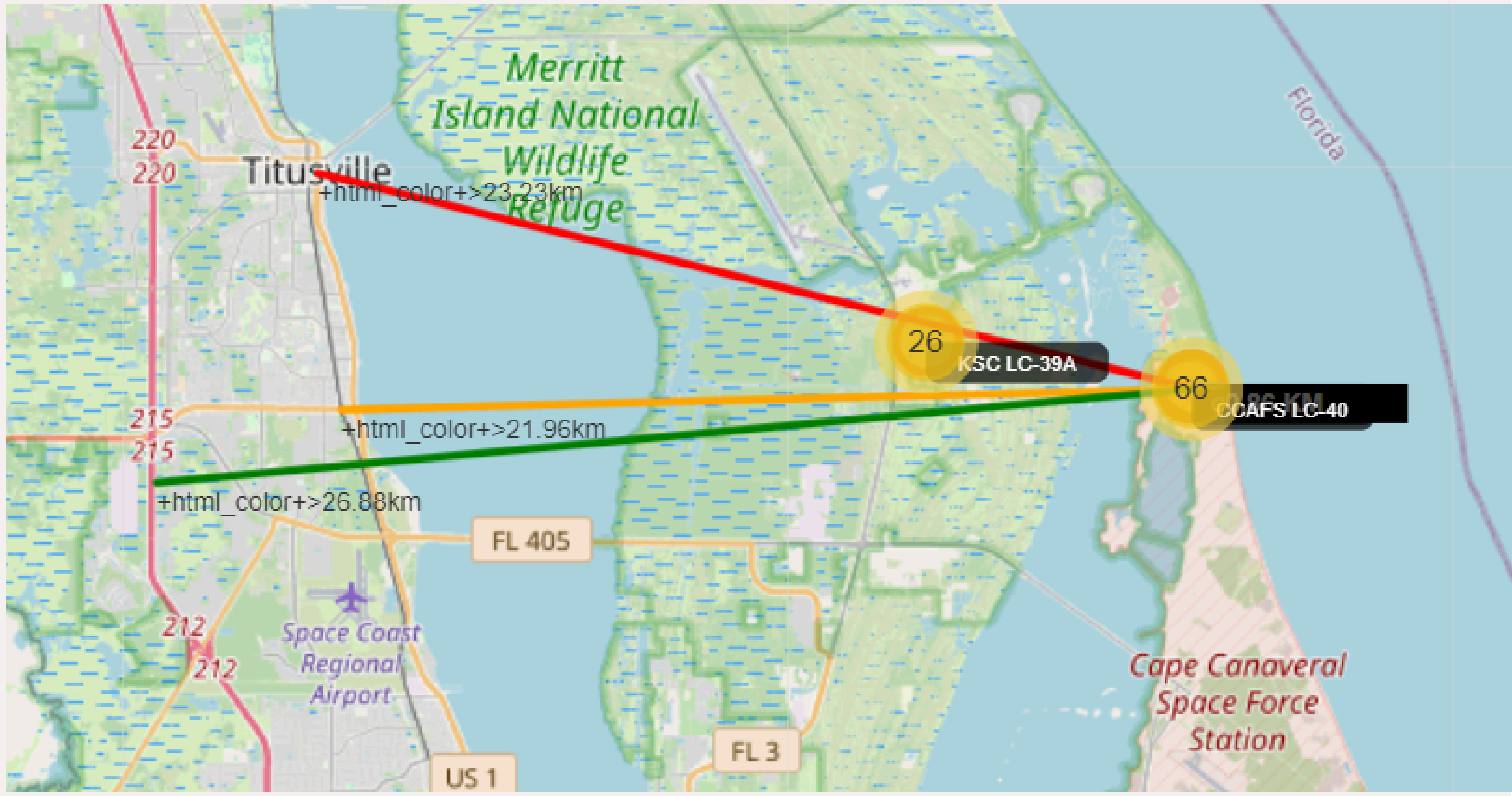


# Proximity Analysis: Location and Key Features

**The map shows...**

- CCAFS SLC-40 Proximity: At CCAFS SLC-40, the nearest coast is 0.86km

Knowing precise distances will help to identify the locations that are far from damages.



Section 4

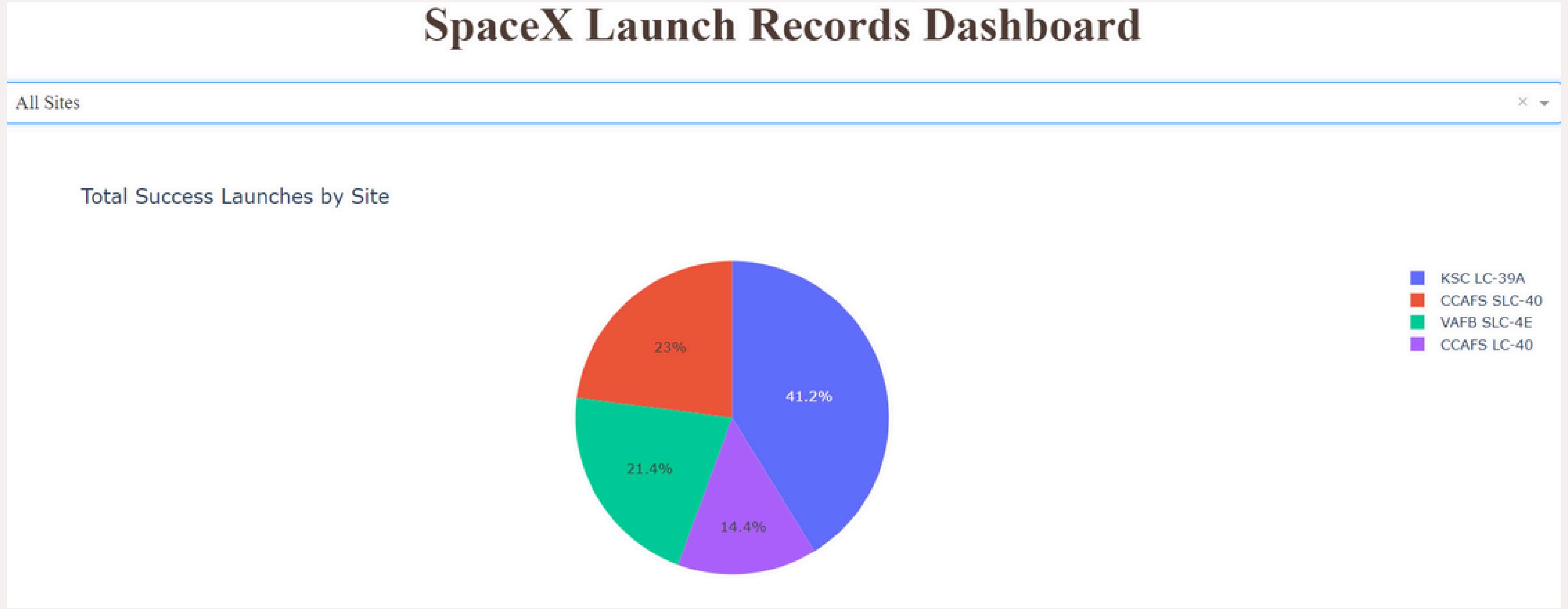
# Build a Dashboard with Plotly Dash

# Launch Success by Site: An Overview

**The dashboard reveals...**

- Overall Success for KSC LC-39A: KSC LC-39A has the most successful launches with 41.2% compared to others.
- There is a disparity between site success.

Having this can help to determine if a site is having more reliability.



# High-Success Launch Site

**The dashboard shows...**

- KSC LC-39A Dominance: KSC LC-39A has the highest success rate for the launches and has most sucessful rate with 76.9%

With KSC-LC-39A the best, a company could invest more in its construction.

# Payload and Outcome Correlation

**The dashboard section shows...**

- No pattern: In this plot, a single area does not show any specific or consistent value.

There should be a need for a more accurate visual.

Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

**Models Evaluated:**
- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbors (KNN)

**Performance Highlights:**
- All models achieved similar test accuracy (~83%).
- Decision Tree had a slight edge on validation data (best score of 0.8875)
- However all models are more or less similar so, model performance is limited by dataset size

**Trade-offs:** While accurate, models have false positives: can improve prediction.

# Confusion Matrix

Key Takeaways (Decision Tree):
- High accuracy can correctly show what happens.
- Model Limitations: False positives represent the primary issue.

Analyzing confusion matrix is essential for refining models.

# Conclusions

**Key Takeaways:**

- Launch success has improved over time.
- KSC LC-39A & certain orbits are more reliable.
- Predictive models show promise, with a small data range, and improvements can come to other models and approaches.

**Next Steps:**

- Data: Bring more sources.
- Features: Add more details.
- Models: Try XGBoost and new aproaches.

*By taking the right steps, this research has important strategic value, and models are important.*

# Appendix

Access the full analysis, code, and data on GitHub:

- Repository Link: Github

This repository contains:

- Jupyter Notebooks: All analysis steps, from data collection to modeling.
- Data Files: The datasets used in this project.
- Supporting Documentation: README file with project details.


Thank You