

# SpaceX Launch Success Rates

IBM Understanding Data Science  
Capstone Project



- Executive Summary
- Introduction
- Methodologies
- Data Exploration
- Model Results
- Conclusion
- Appendix

## AGENDA

# Executive Summary - Turning Historical Launch Data into Predictive Insight

---



- **Objective:** Predict Launch Success
- **Data Science Methodology**
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis
  - Machine Learning Modeling
  - Model Evaluation
- **Data Scope:**
  - 90 SpaceX Launches over 10 Yrs (2010-20)
  - 3 Launch Sites in California, Texas and Florida
- **Conclusion:**
  - Best Model: 83.33% Test accuracy
  - Predicted: 88.93% Cross Validation Accuracy
  - Assume: Similar Rocket Design and Flight Parameters

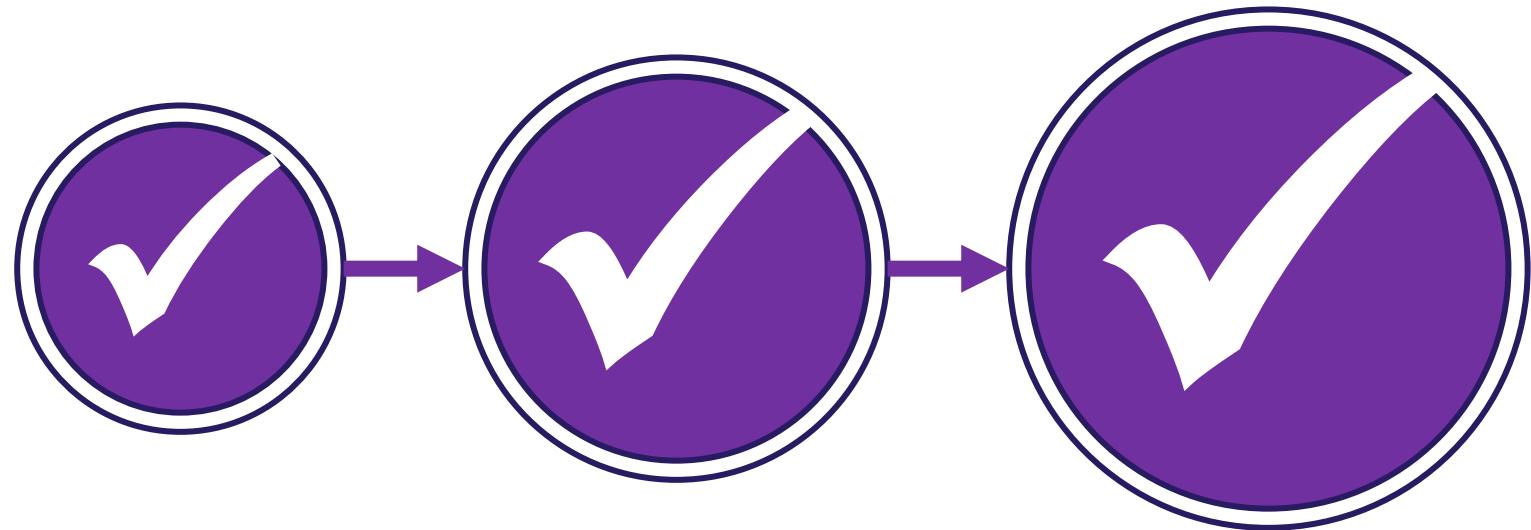


# Introduction – About this Project

---



- Launch **reliability** drives mission success.
- SpaceX data reveals **performance patterns**.
- Historical trends inform predictive **insight**.
- Objective: **Forecast** outcomes before liftoff.
- Data **reduces** operational risk.
- Models were **rigorously** evaluated.
- Built for robustness and **real-world use**.
- Key **success drivers** were identified.
- **Predictive power** was demonstrated.
- Analytics strengthens **aerospace strategy**.



# METHODOLGIES

# Overall Project Methodology

---

## 1 Business Problem Definition

Translate stakeholder goals into a measurable ML problem with defined success metrics

## 2 Data Acquisition & Inventory

Identify and catalog relevant data sources across internal systems, third-party feeds, and public datasets

## 3 Exploratory Data Analysis (EDA)

Profile distributions, detect anomalies, and uncover relationships through statistical summaries and visualizations

## 4 Data Preprocessing & Feature Engineering

Handle missing values, encode categoricals, and engineer domain-informed variables to strengthen model signal

## 5 Model Selection & Baseline Establishment

Evaluate candidate algorithms against a simple baseline to guide complexity tradeoffs

## 6 Model Training & Cross-Validation

Train models using k-fold cross-validation to ensure robust, generalizable performance

## 7 Hyperparameter Tuning & Optimization

Apply grid search, random search, or Bayesian optimization to improve model performance

## 8 Model Evaluation & Interpretability

Assess final performance on a held-out test set; use SHAP or LIME for explainability



# Data Collection & Wrangling Methodology

## Step 1: Data Understanding

- Identify and catalog data sources
- Explore available fields and formats
- Assess quality and identify challenges
- Document limitations

## Step 2: Data Collection

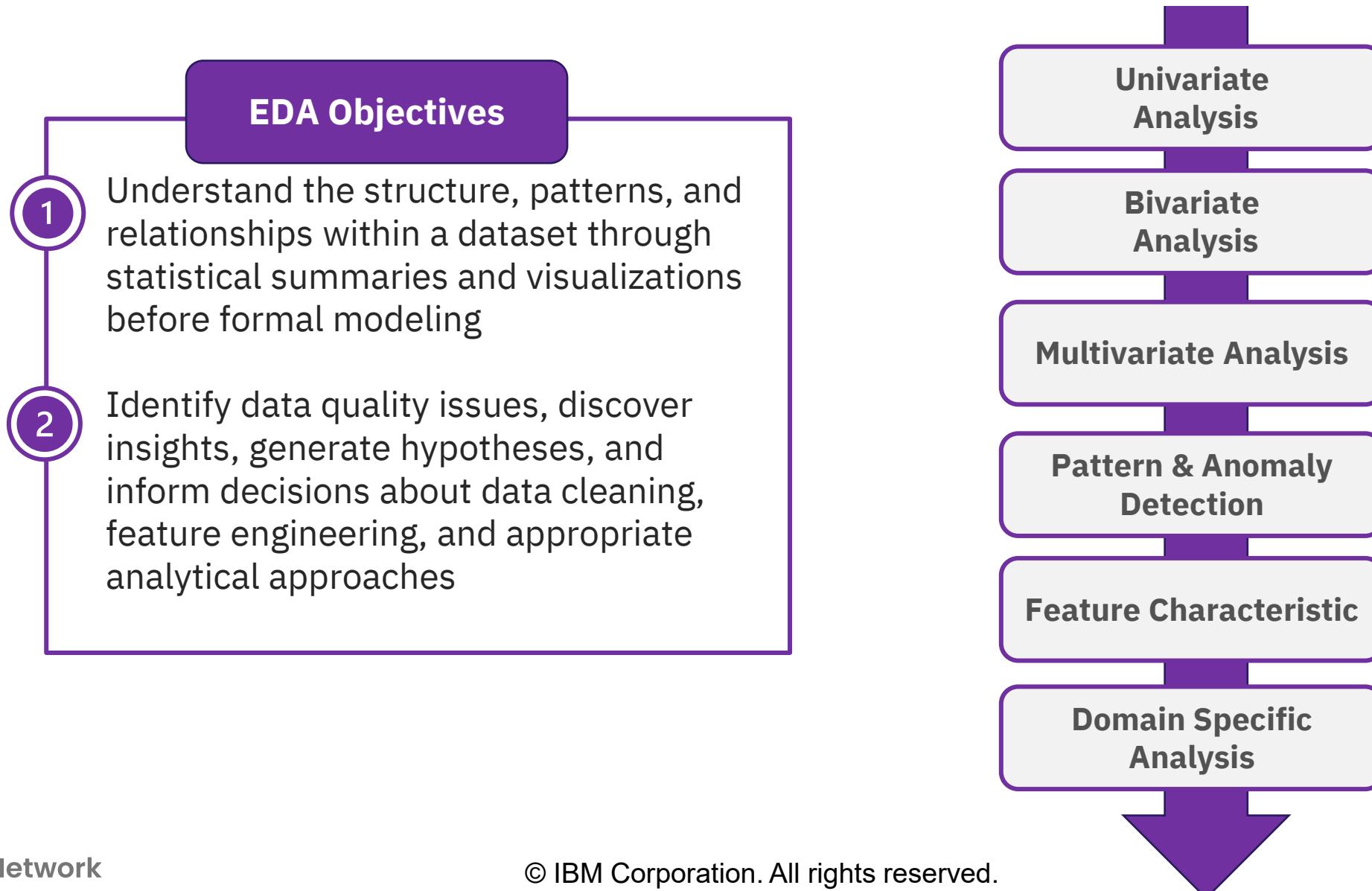
- Design collection strategy
- Implement extraction process
- Validate during collection
- Handle errors and log activities

## Step 3: Data Wrangling

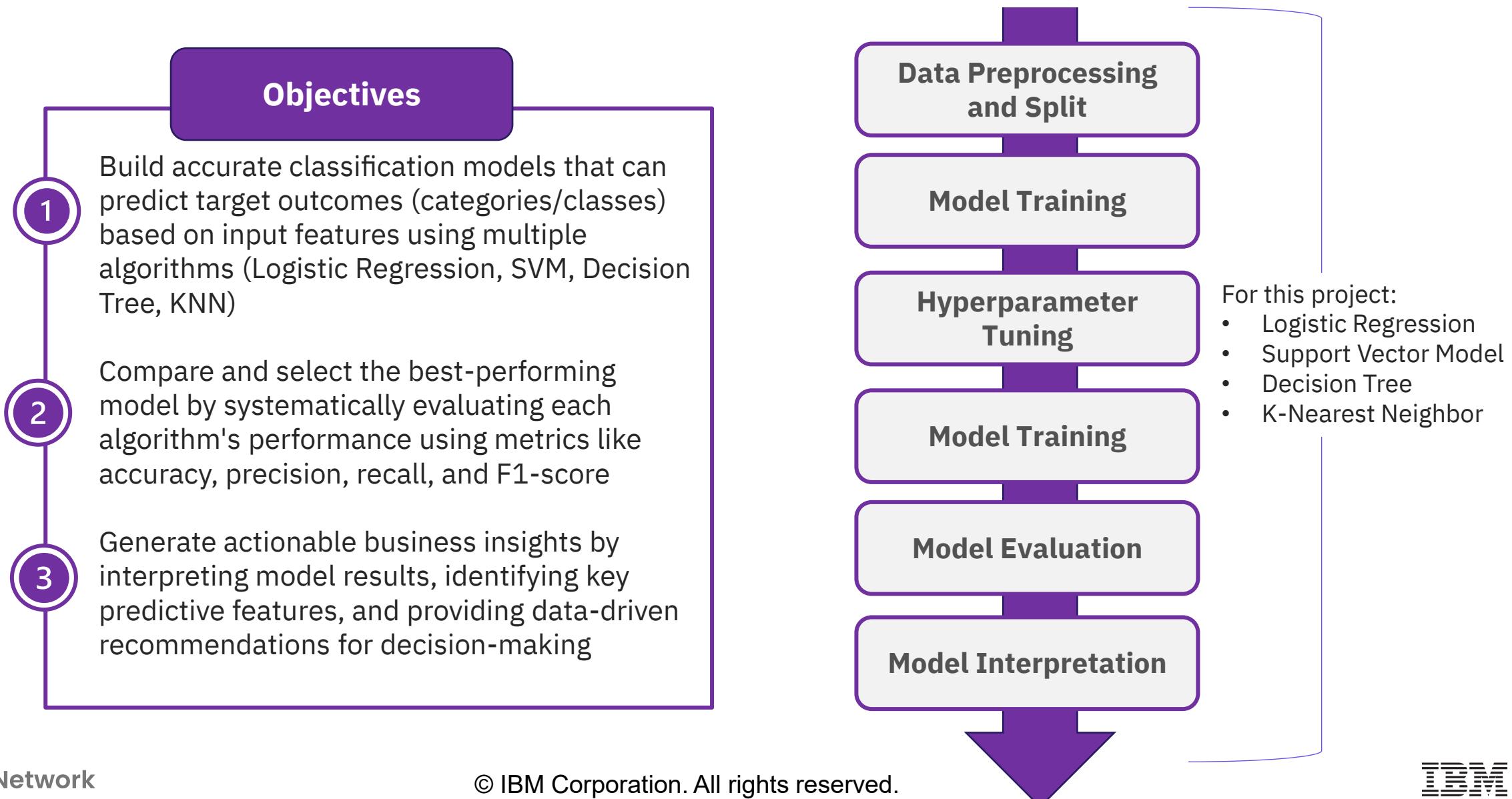
- Clean data (missing values, errors, duplicates)
- Transform data (engineering, encoding, scaling)
- Integrate multiple sources
- Validate quality comprehensively
- Prepare final dataset



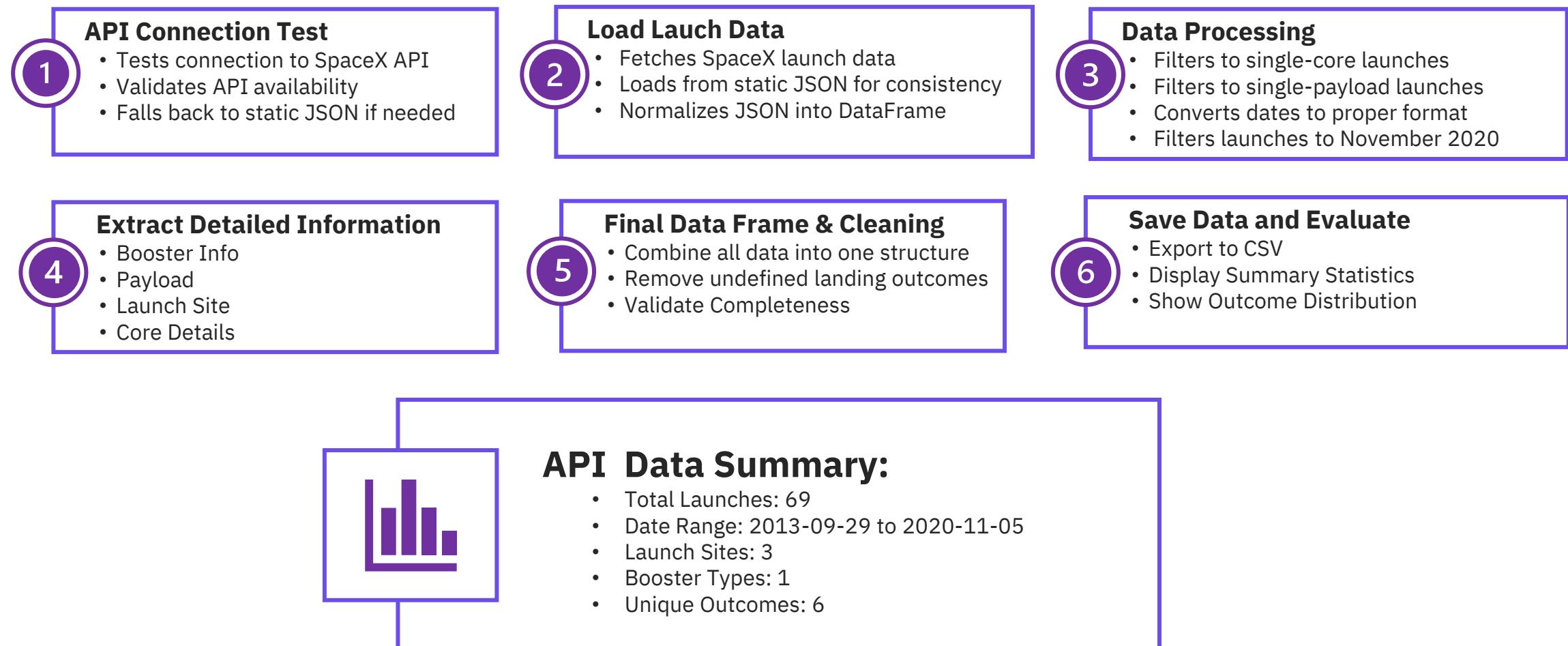
# Exploratory Data Analysis Methodology

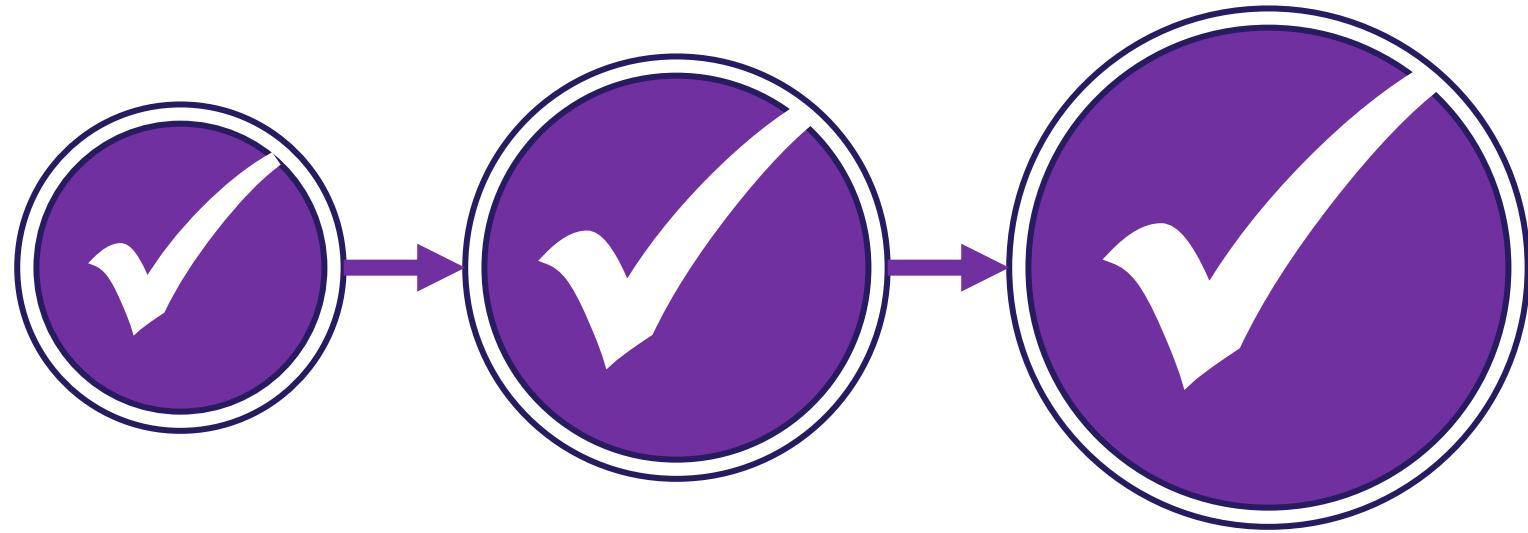


# Predictive Modeling Methodology



# API Data Collection Steps

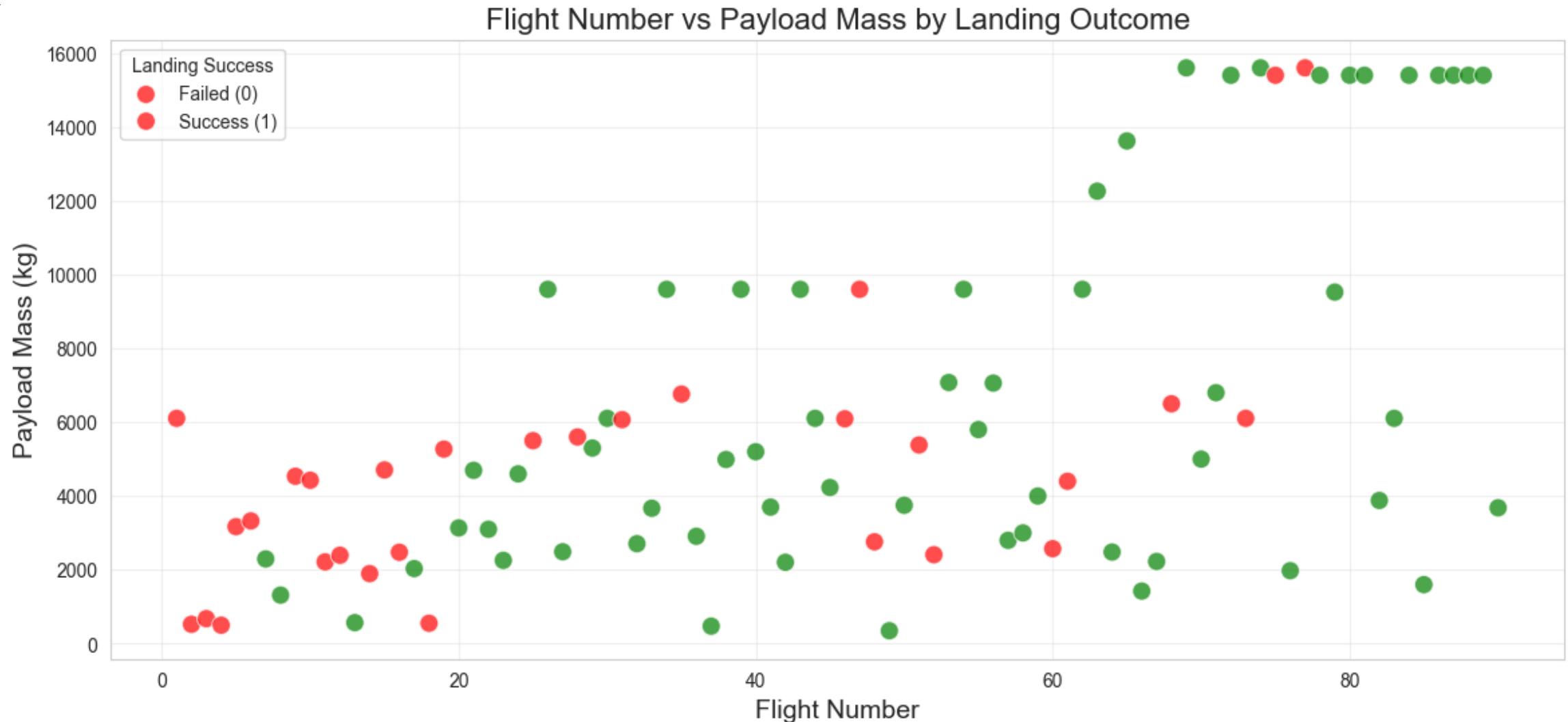




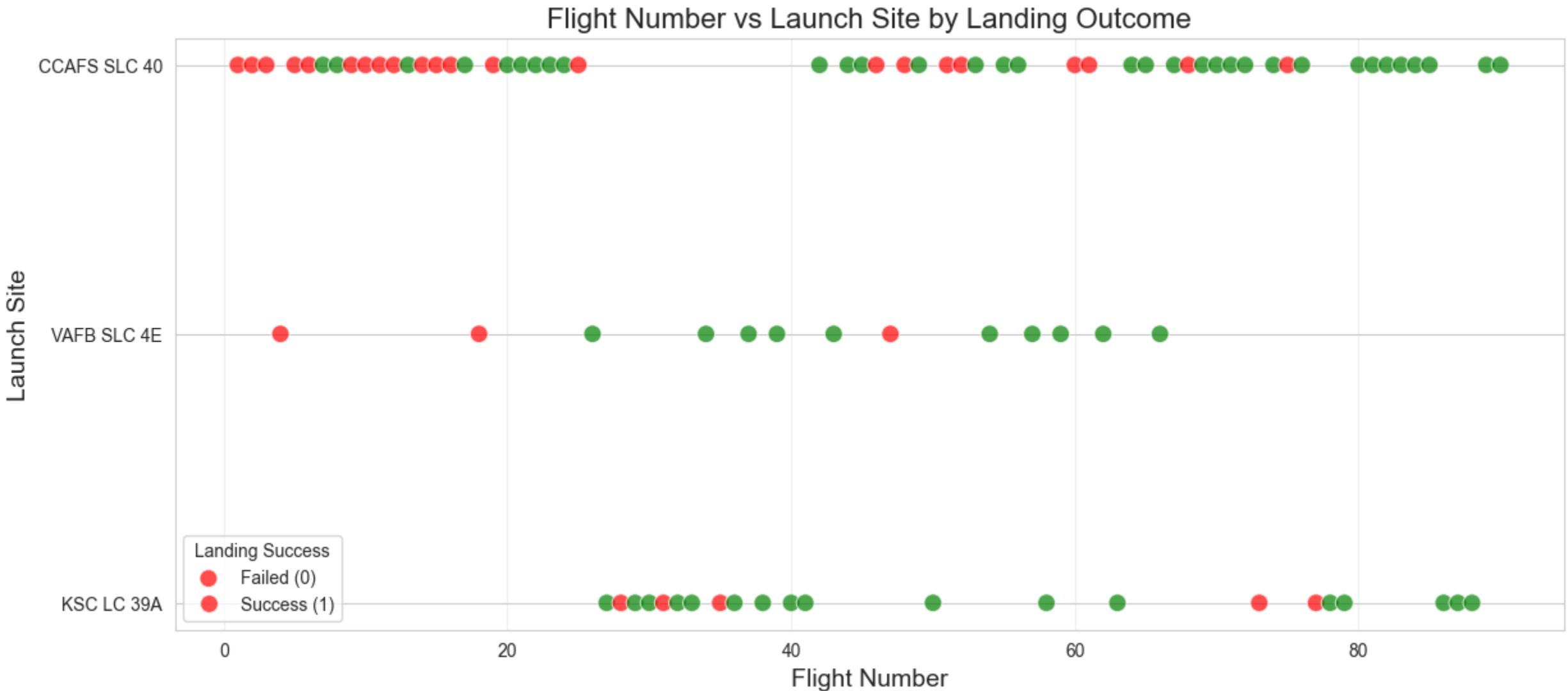
# EDA Visualizations

Module 2 – Lab 2

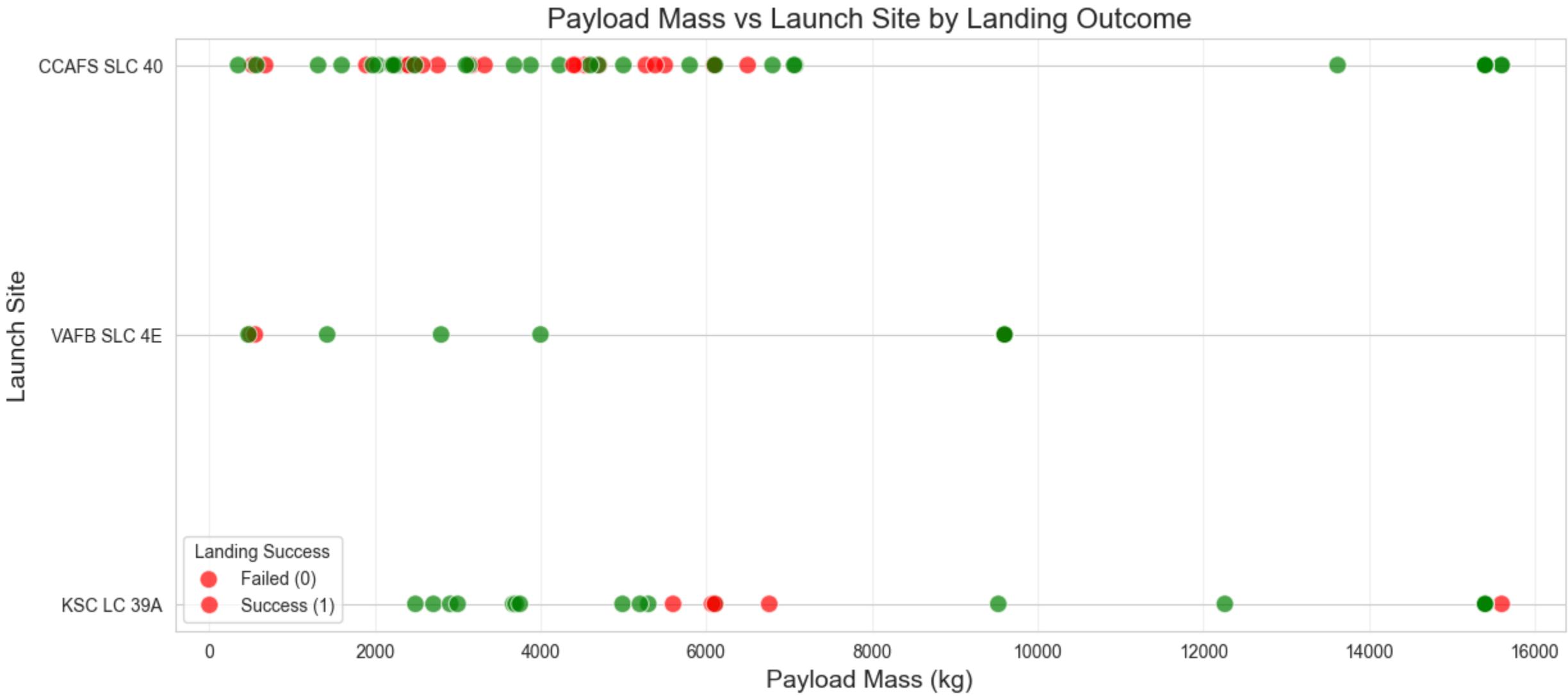
# EDA With Visualization Results



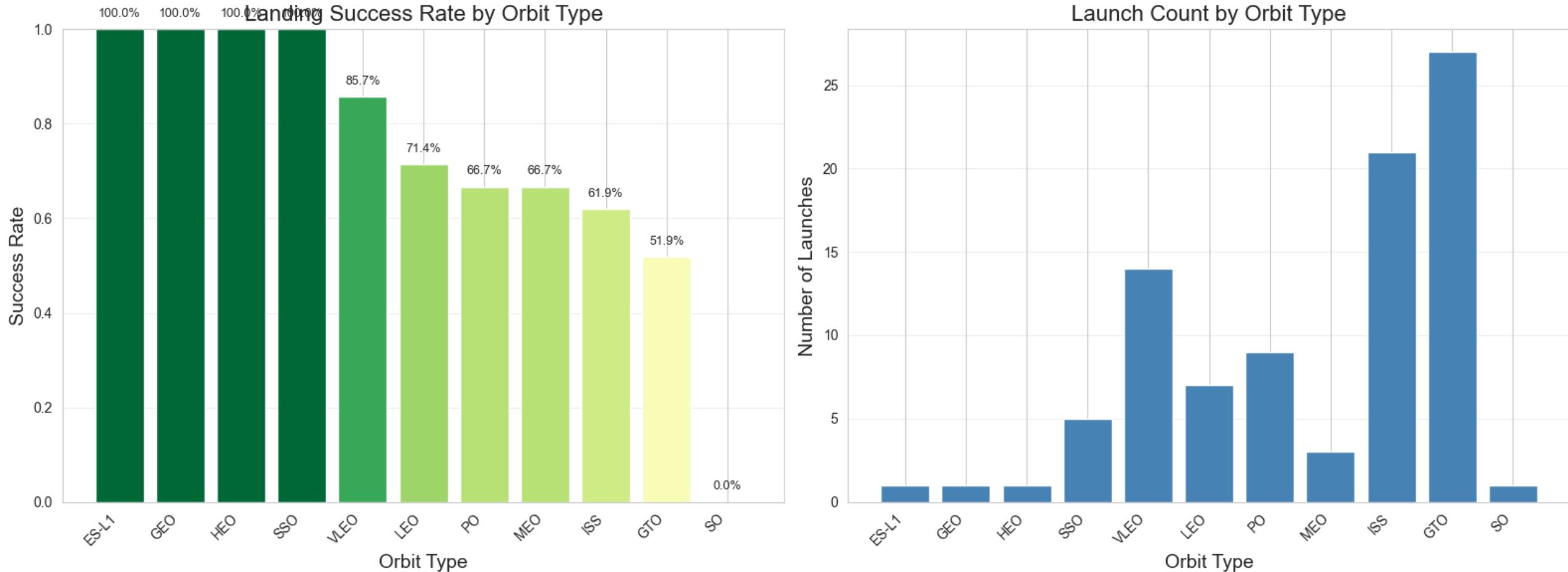
# EDA With Visualization Results



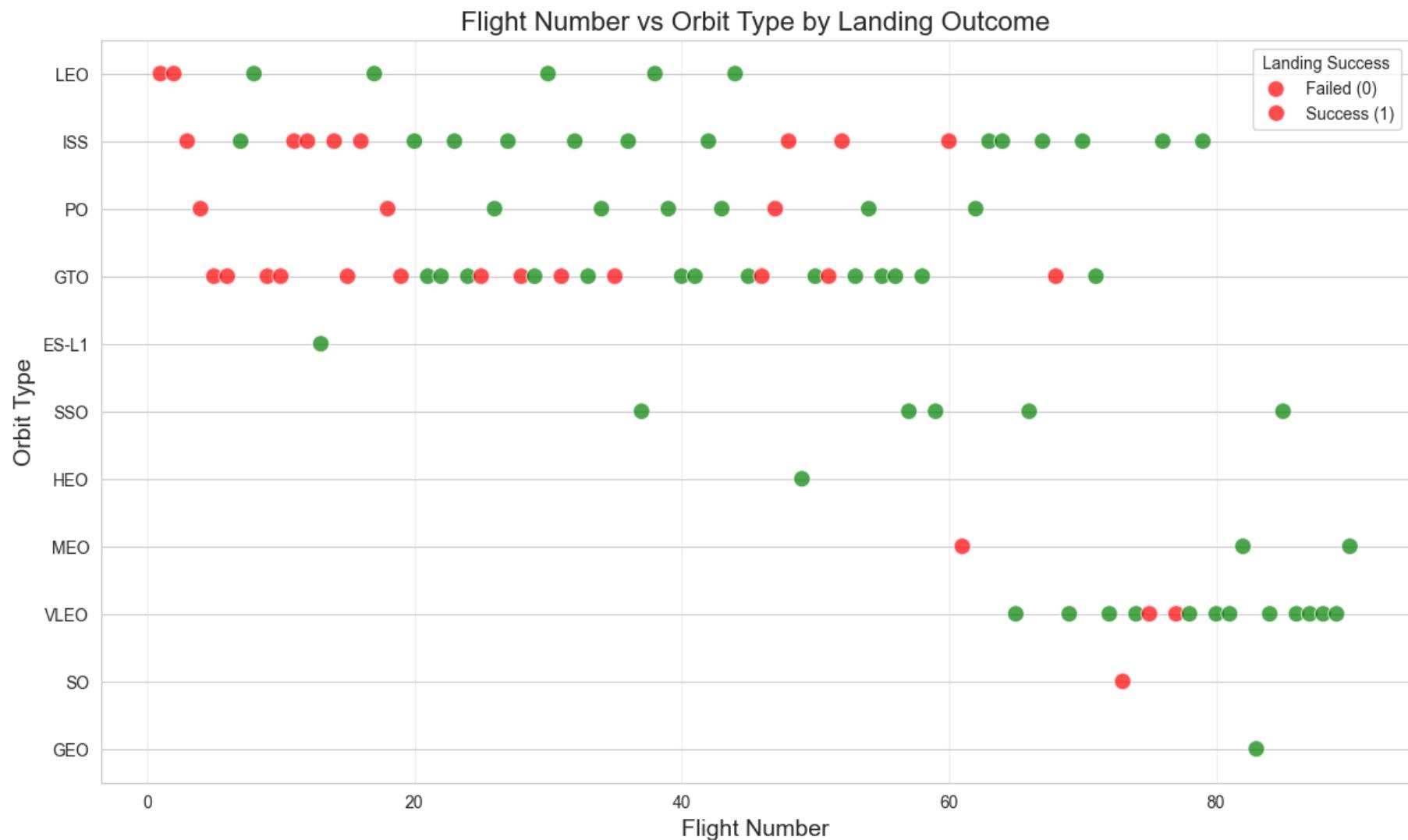
# EDA With Visualization Results



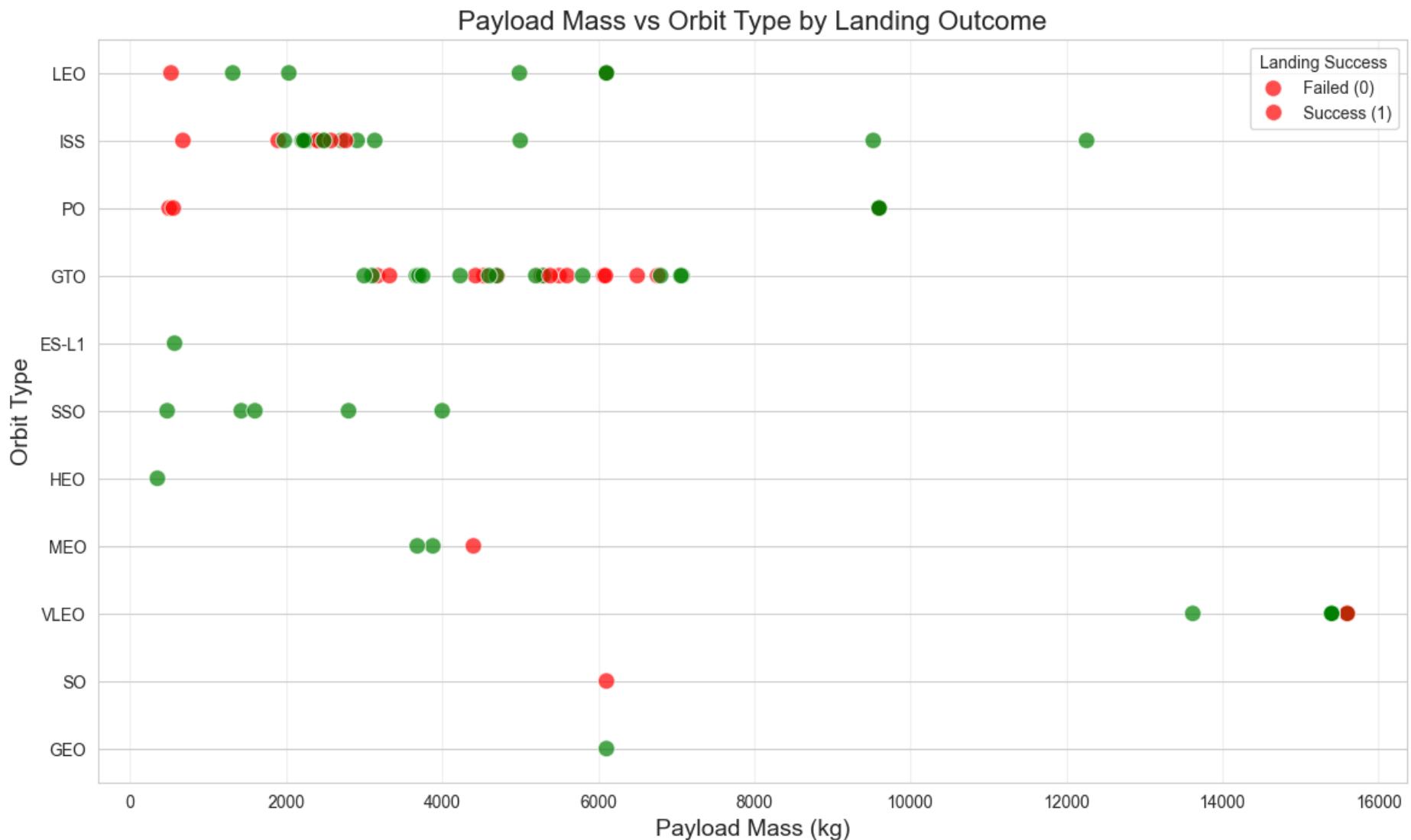
# EDA With Visualization Results



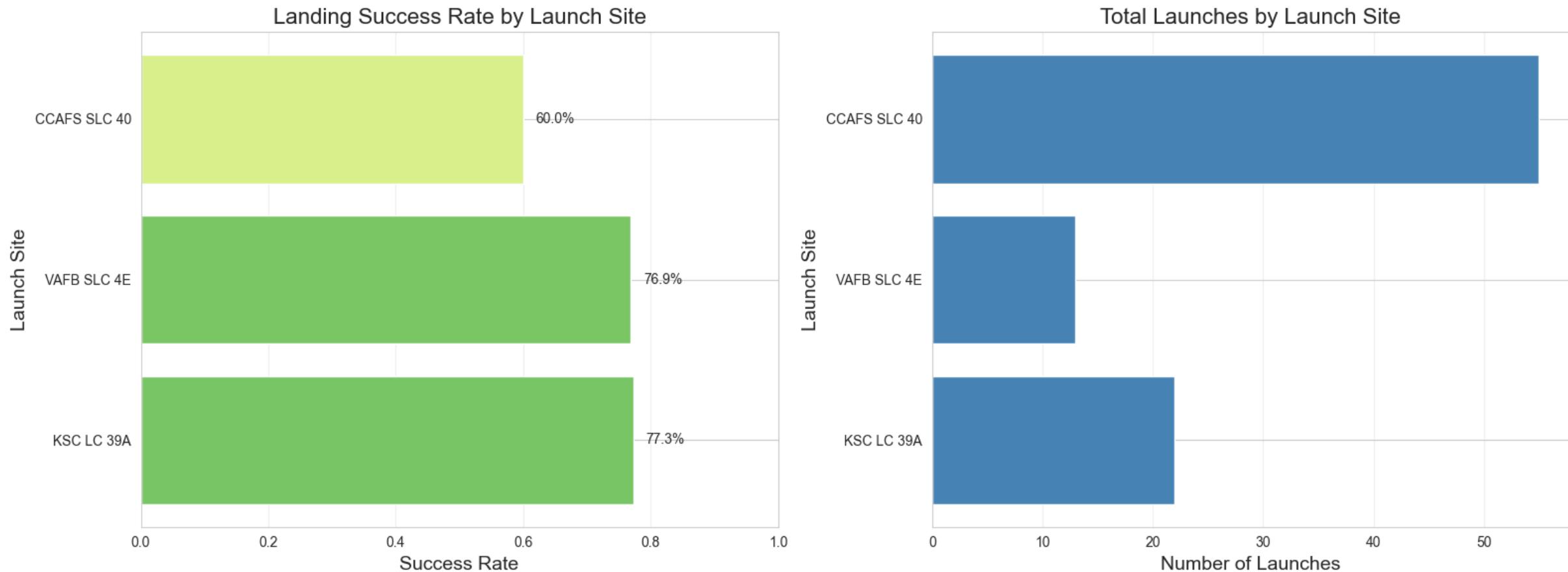
# EDA With Visualization Results



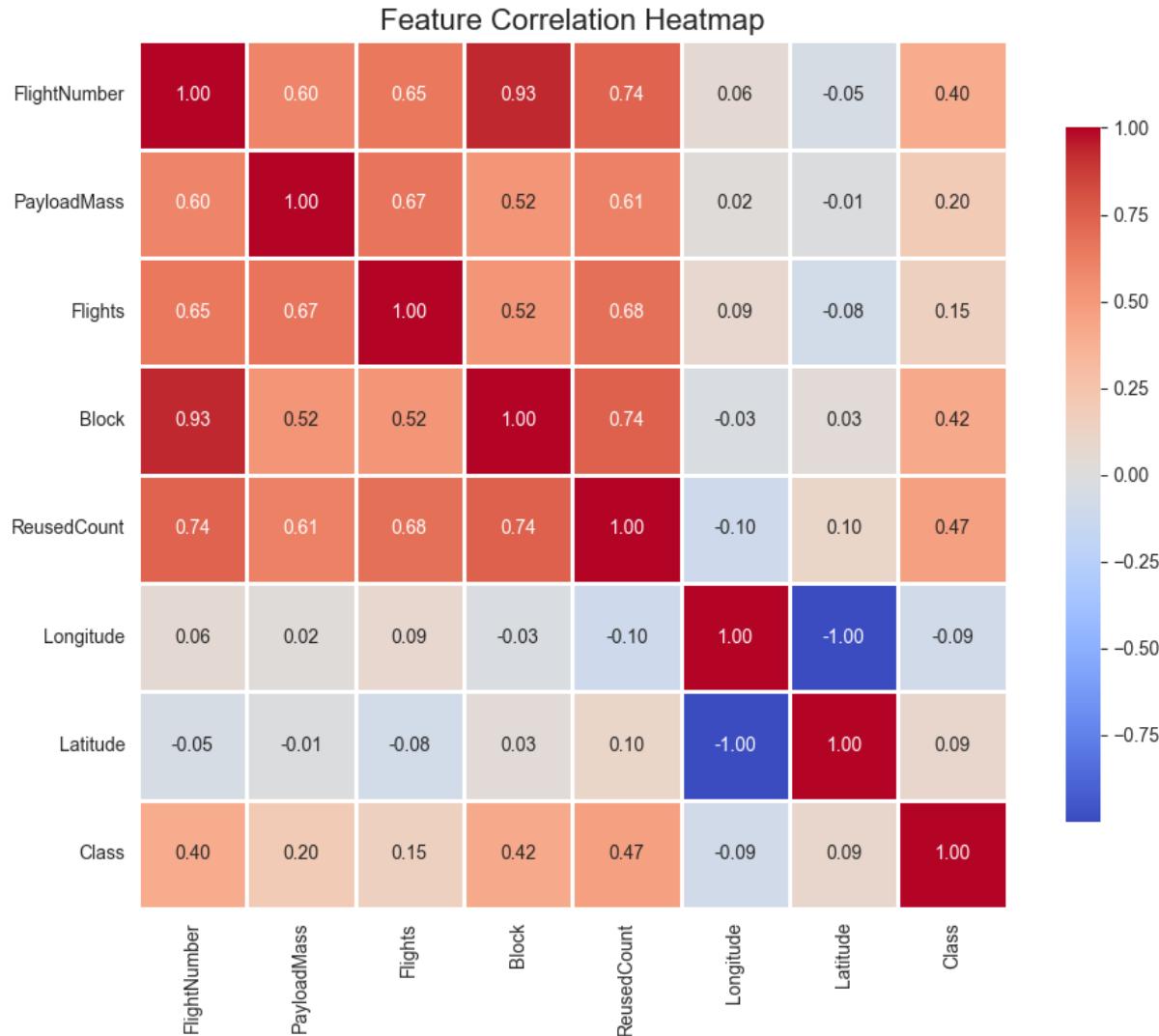
# EDA With Visualization Results



# EDA With Visualization Results



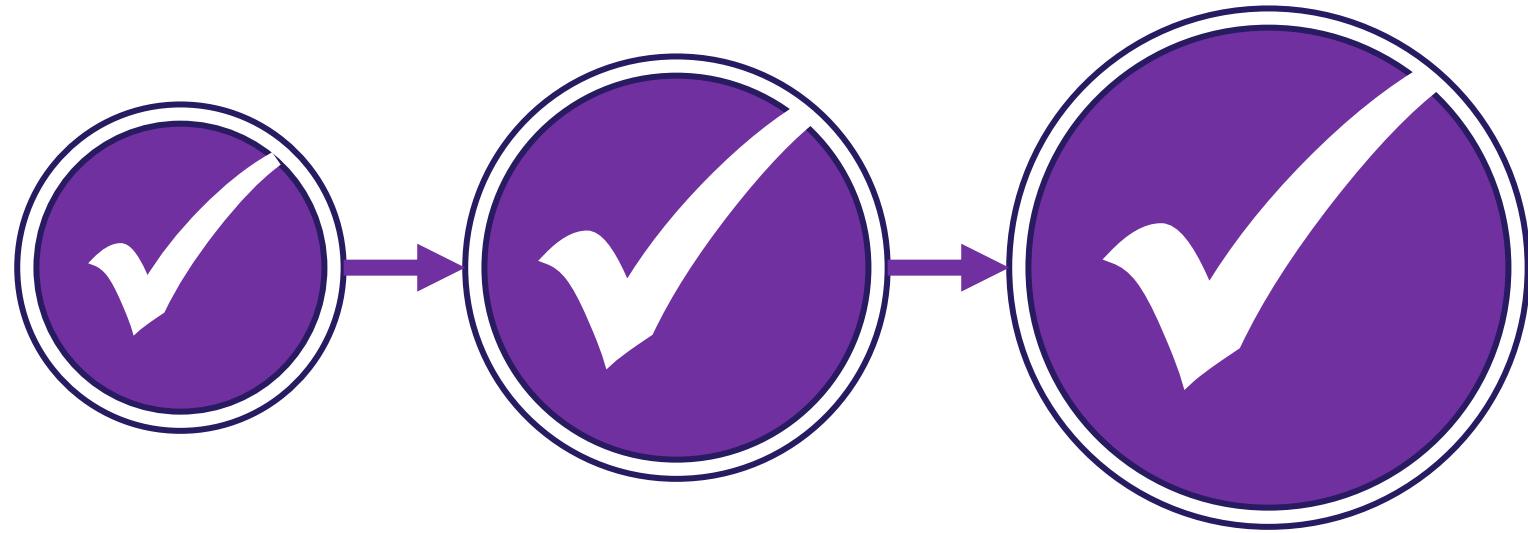
# EDA With Visualization Results



# EDA With Visualization – Key Findings

---

Metric	Value
Total Launches Analyzed	90
Overall Success Rate	66.67%
Number of Launch Sites	3
Number of Orbit Types	11
Payload Mass Range	350 - 15,600 kg
Flight Number Range	1 - 90



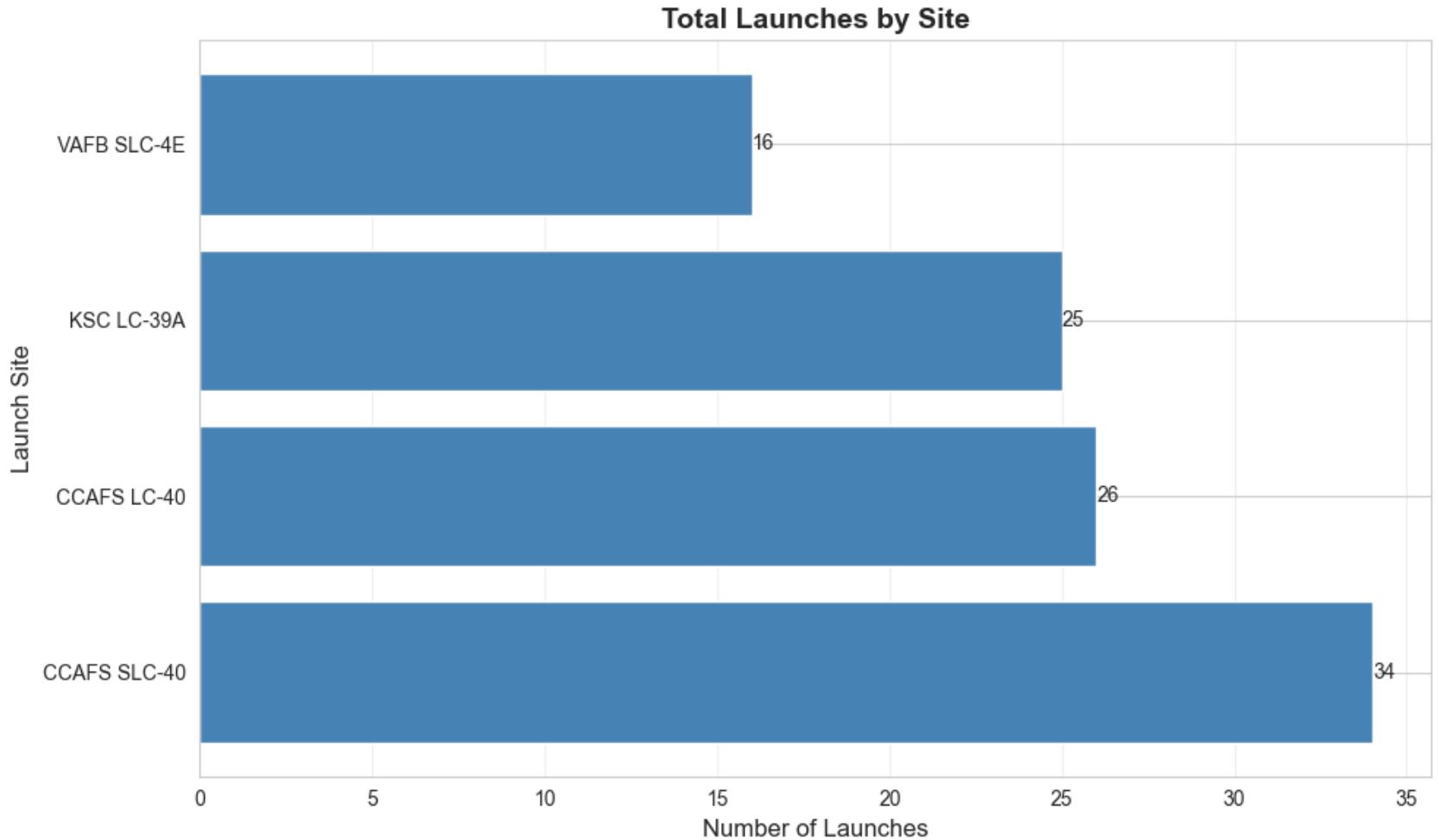
# EDA with SQL Visualizations

Module 2 – Lab 1

# EDA With SQL Results – Task 1

## Launch Sites

- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40



# EDA With SQL Results – Task 2 Results

## First 5 records from Cape Canaveral (CCA)

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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	(Payload not provided)	(Not provided)	(Unknown)	(Not provided)	(Not provided)	(Not provided)

# EDA With SQL Results – Task 3

---

Total payload mass for NASA (CRS)

**45596**

# EDA With SQL Results – Task 4

---

Average payload mass for F9 v1.1

**2928.4**

# EDA With SQL Results – Task 5

---

**FIRST\_SUCCESSFUL\_GROUND\_PAD\_LANDING\_DATE:**  
**2015-12-22**

# EDA With SQL Results – Task 6

---

Boosters with successful drone ship landings (4000-6000 kg)

## **Booster\_Version**

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Create a visualization showing average for each booster



# EDA With SQL Results – Task 7

Count of successful vs failed missions

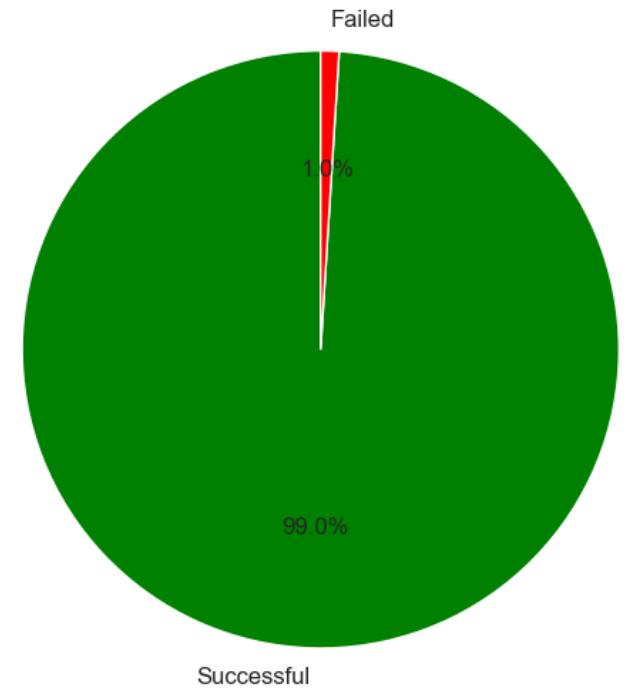
SUCCESSFUL: 100

FAILED: 1

Success Rate:  
99.01% (100/101)



Mission Success Rate



# EDA With SQL Results – Task 8

---

Boosters with successful drone ship landings (4000-6000 kg)

Results (4 rows):

**Booster\_Version**

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

# EDA With SQL Results – Task 9

---

Failed drone ship landings in 2015

<b>MONTH_NAME</b>	<b>Landing_Outcome</b>	<b>Booster_Version</b>	<b>Launch_Site</b>
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

# EDA With SQL Results – Task 10 Results

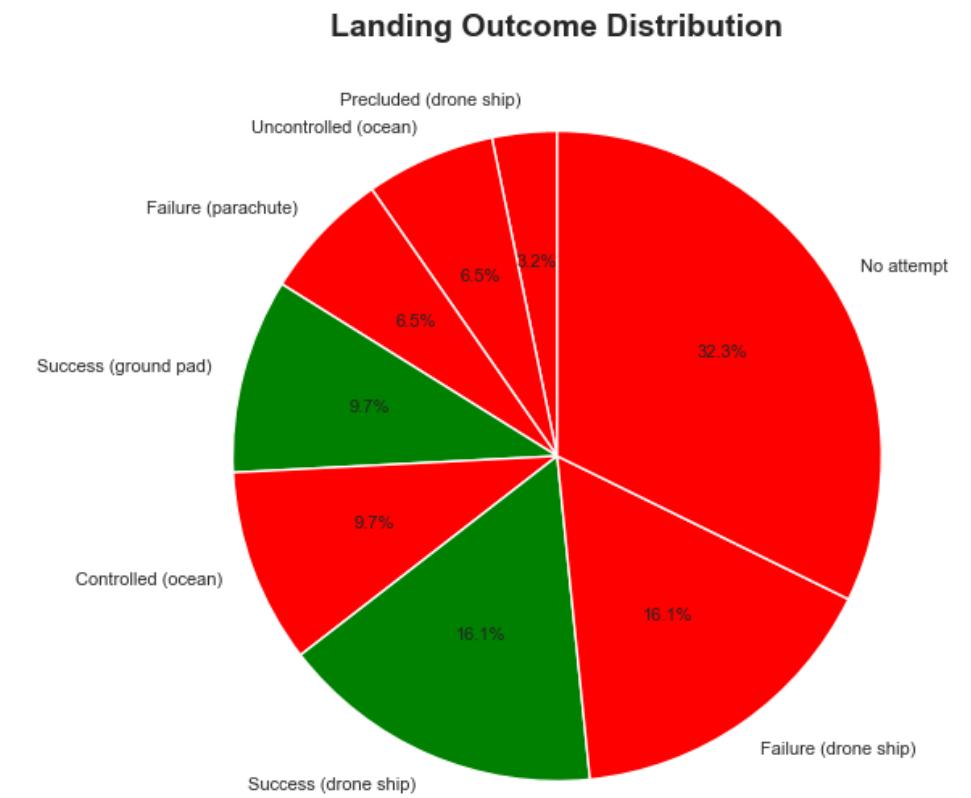
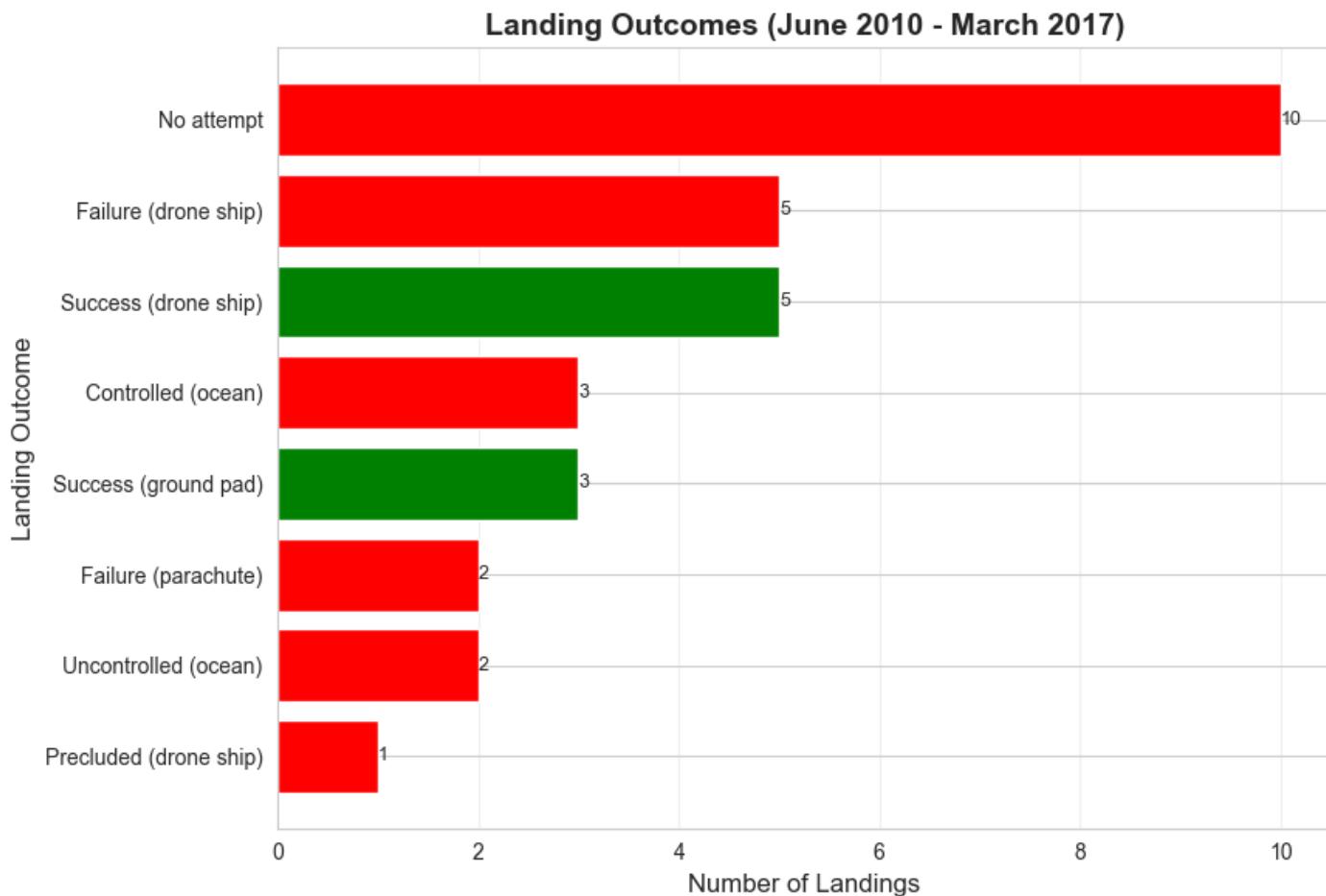
---

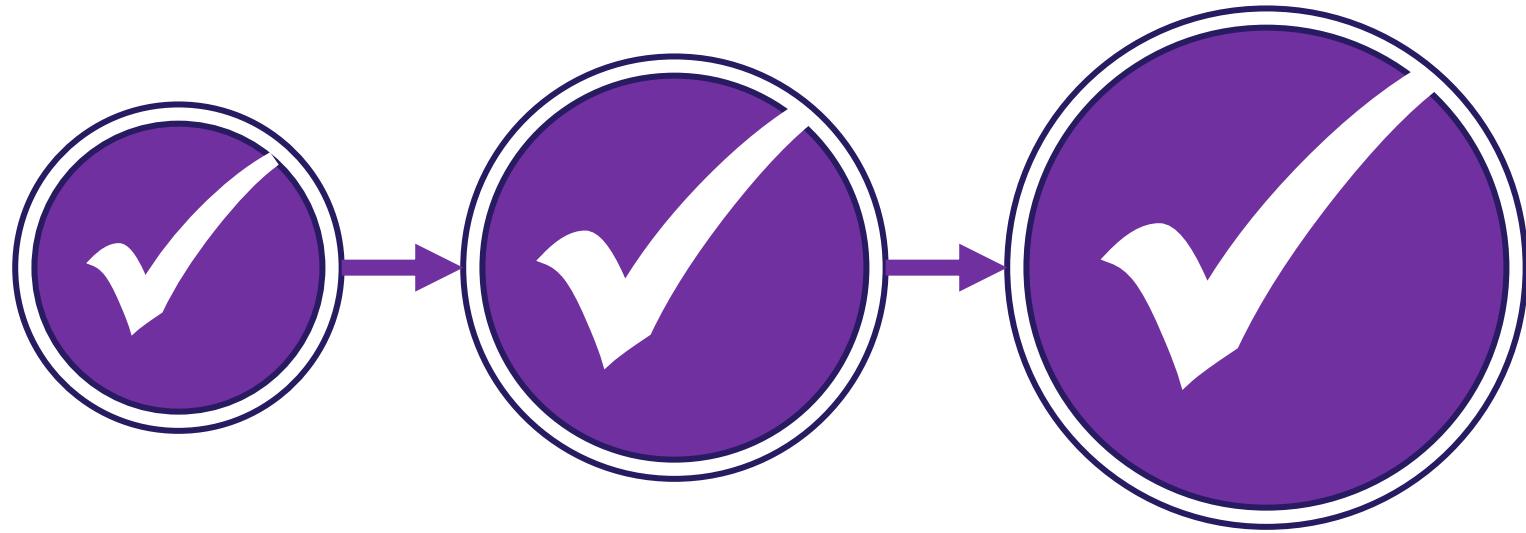
Landing outcomes between June 2010 - March 2017

Outcome	Count
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

# EDA With SQL Results – Task 10 Visualized

Landing outcomes between June 2010 - March 2017

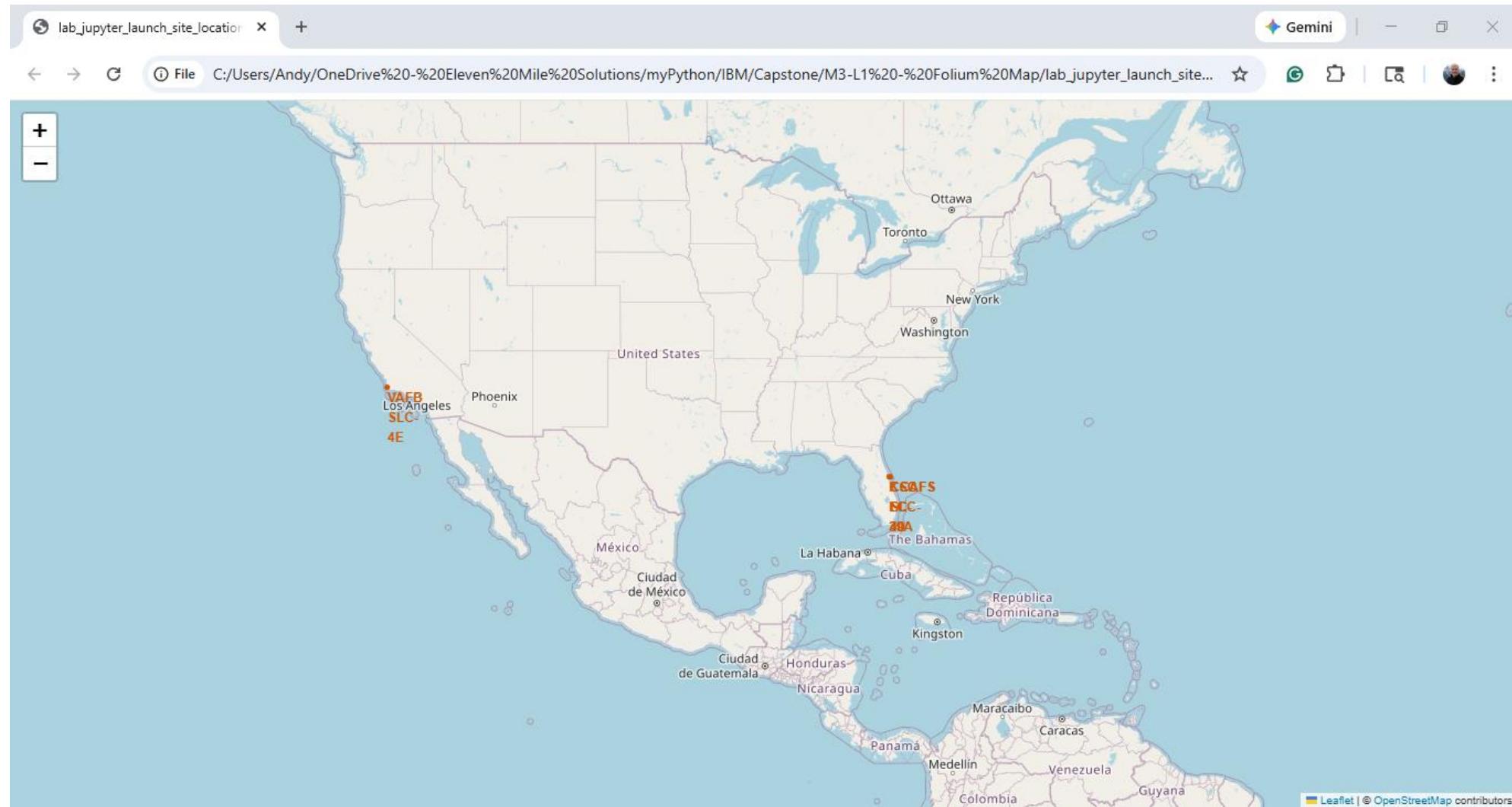




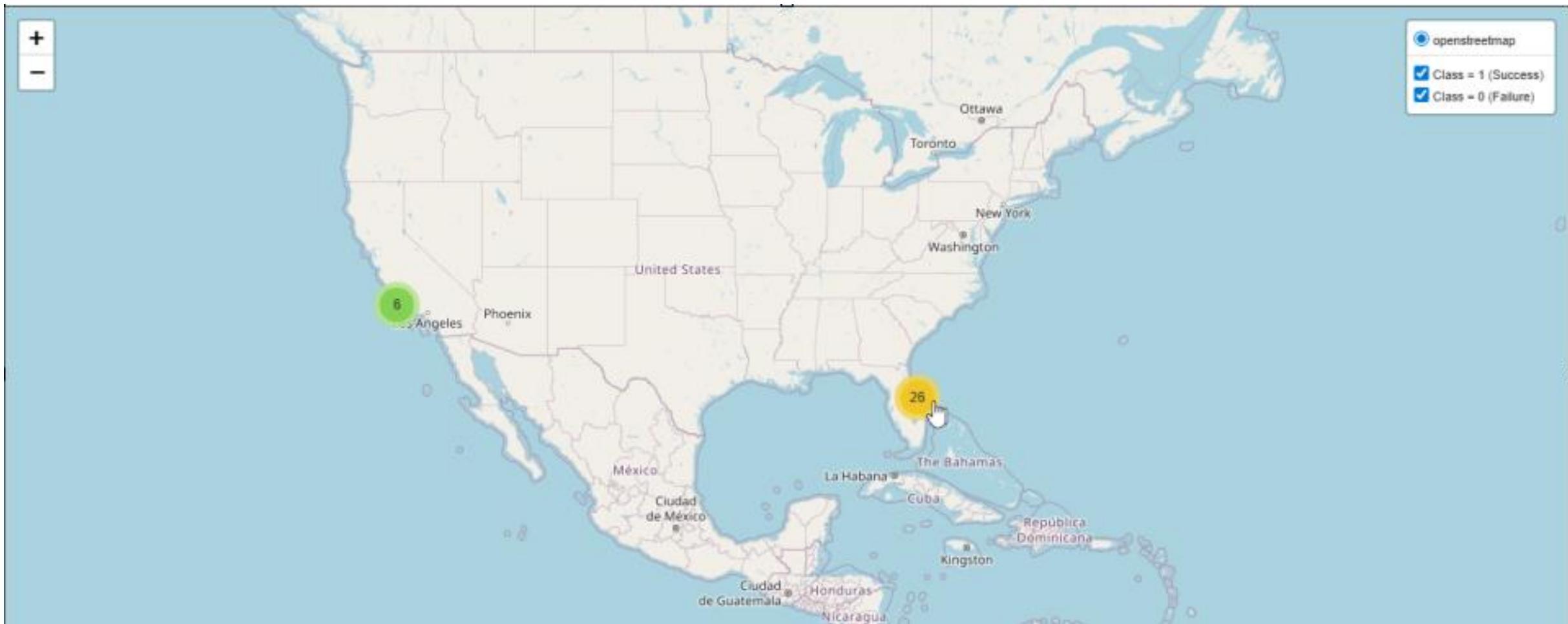
# EDA with Folium Maps

Module 3 – Lab 1

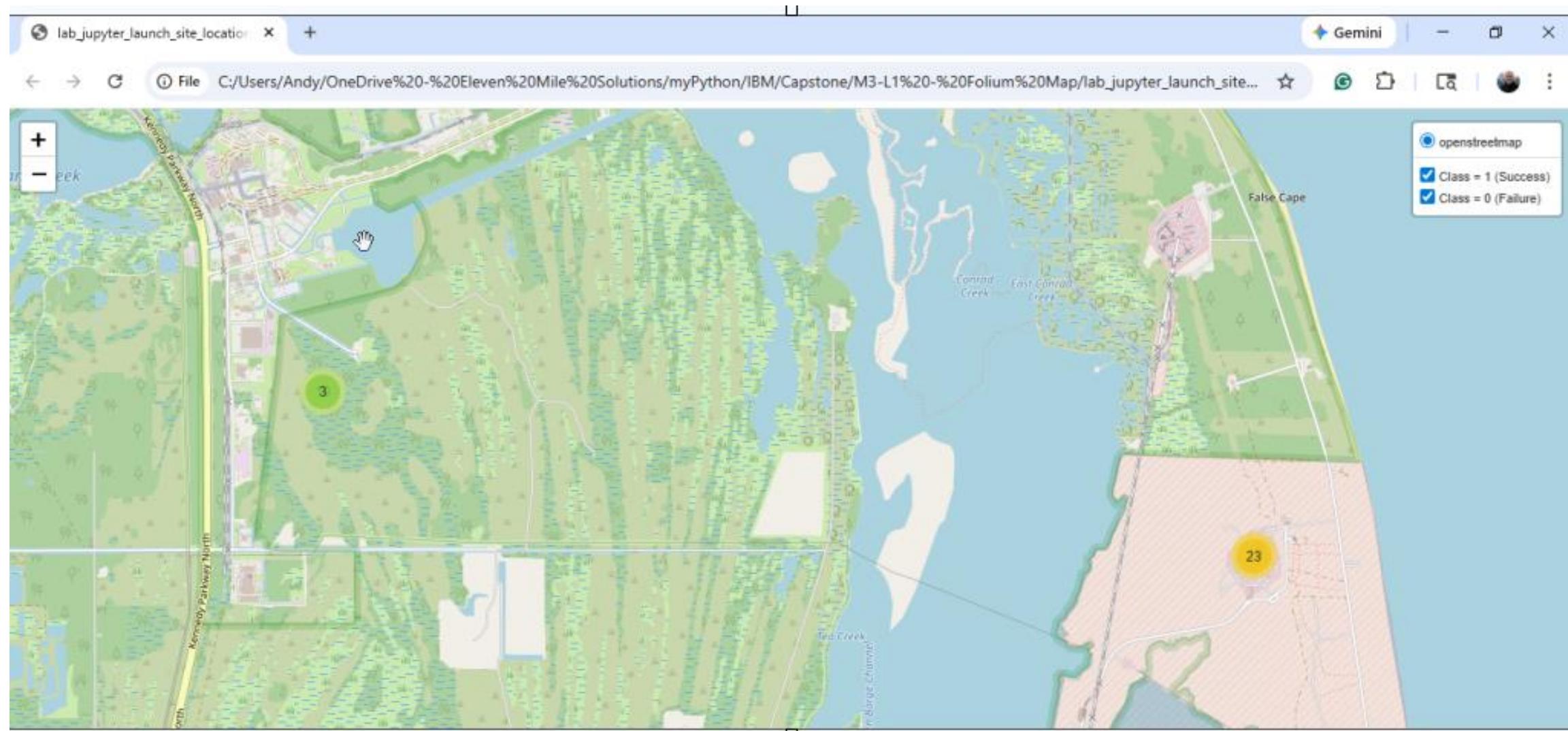
# EDA With Folium Map – Task 1 - Marking Launch Sites



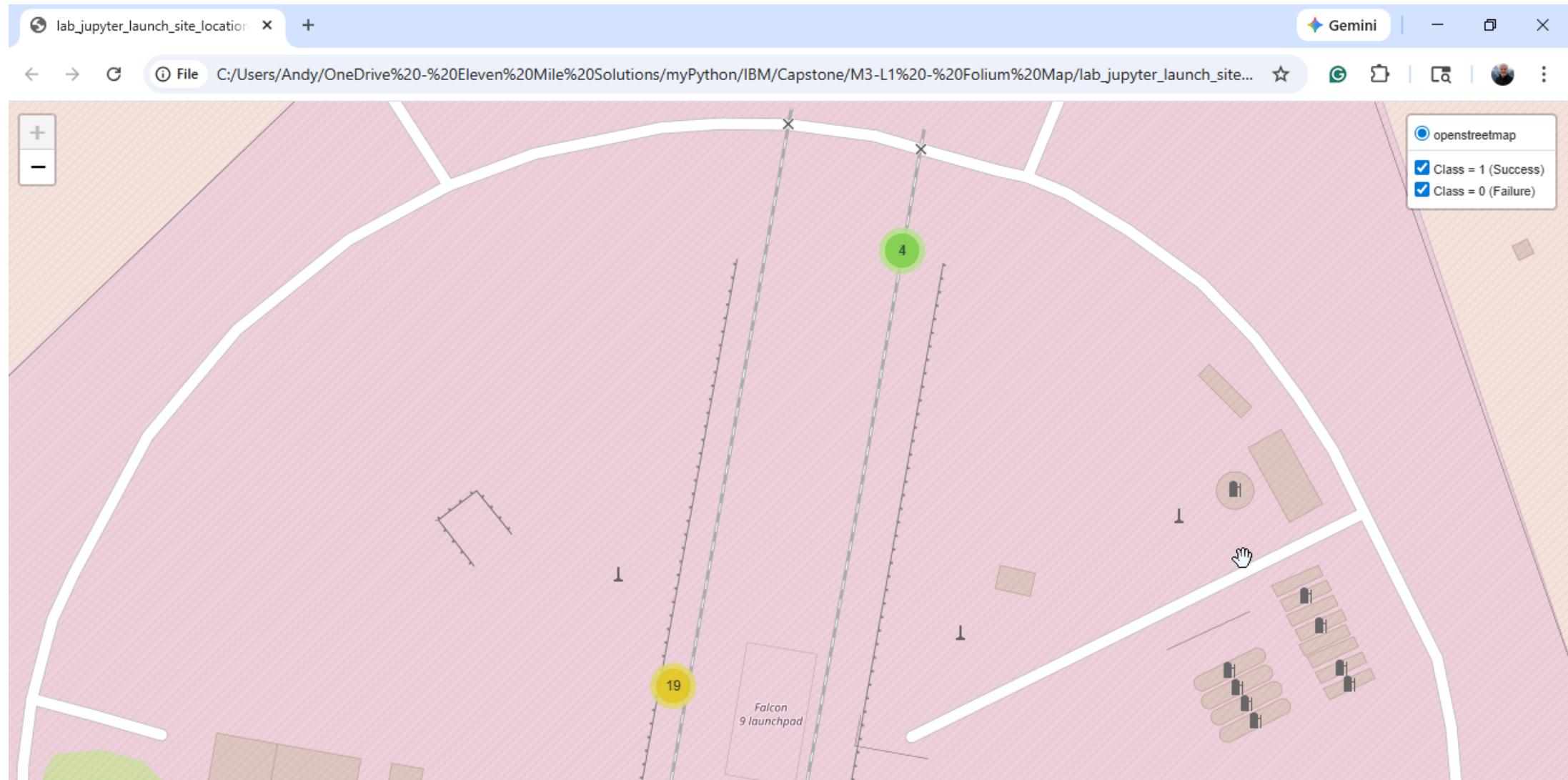
# EDA With Folium Map – Task 2 - Marking Launch Outcomes 1



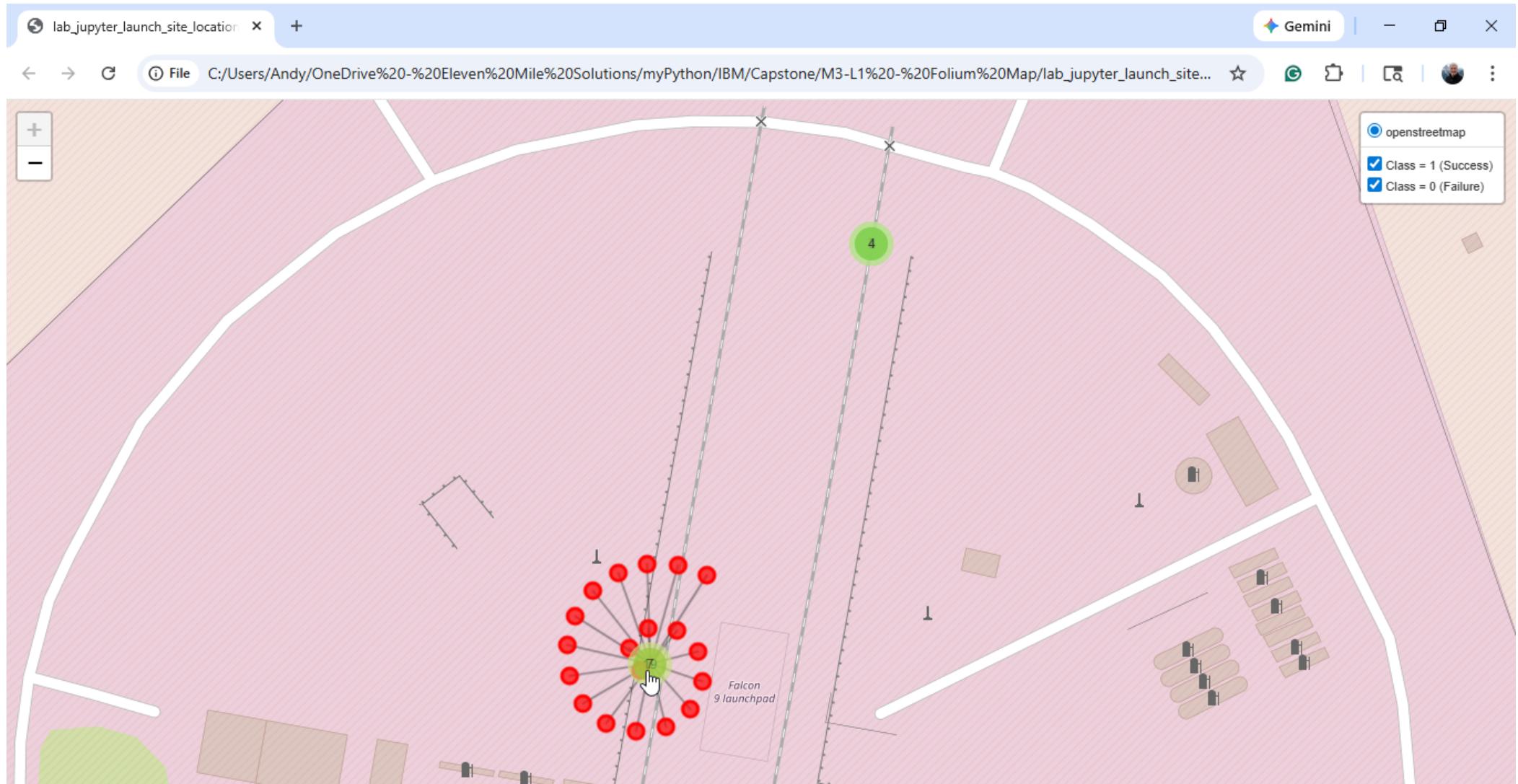
# EDA With Folium Map – Task 2 Marking Launch Outcomes 2



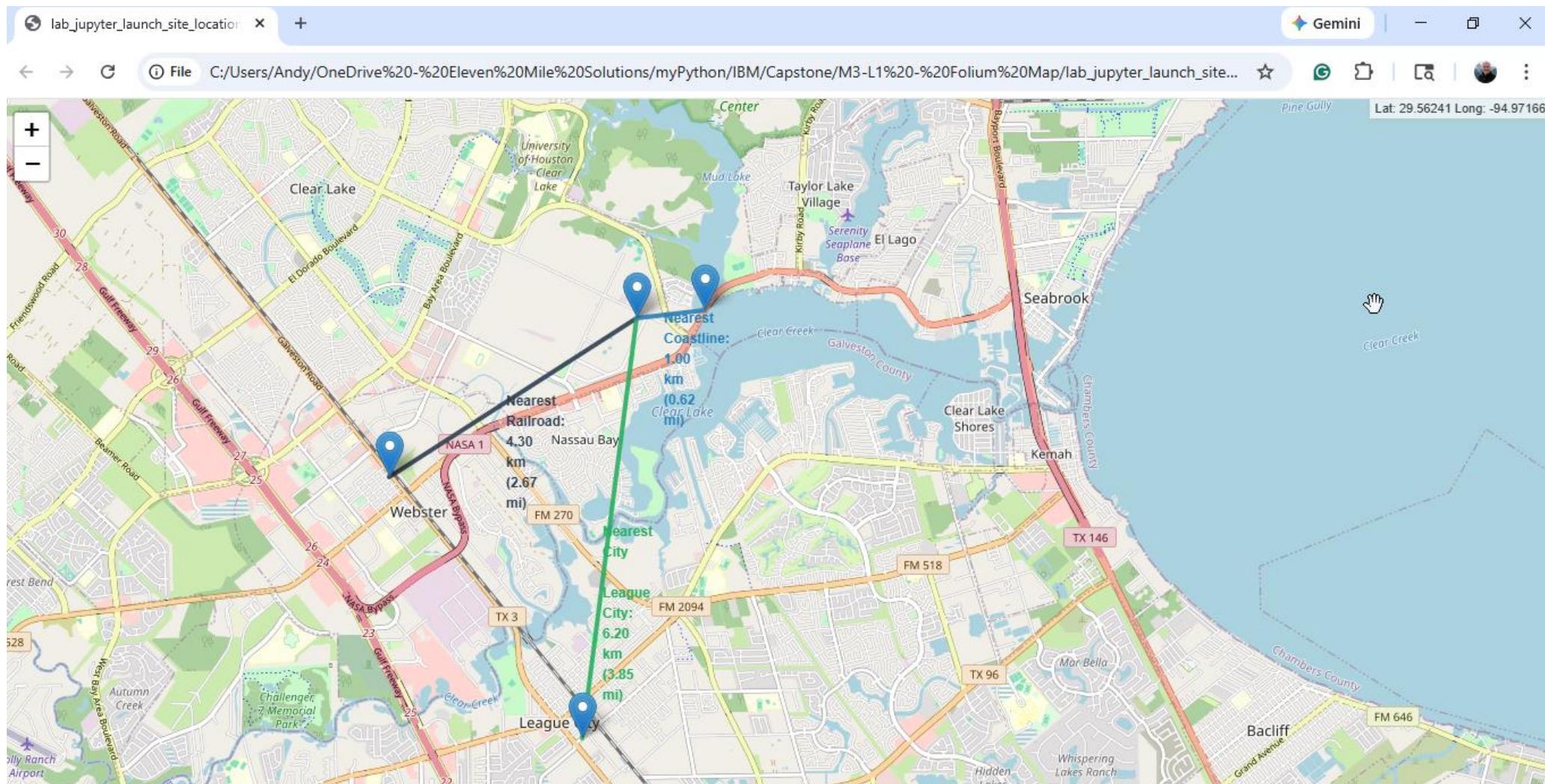
# EDA With Folium Map – Task 2 - Marking Launch Outcomes 3

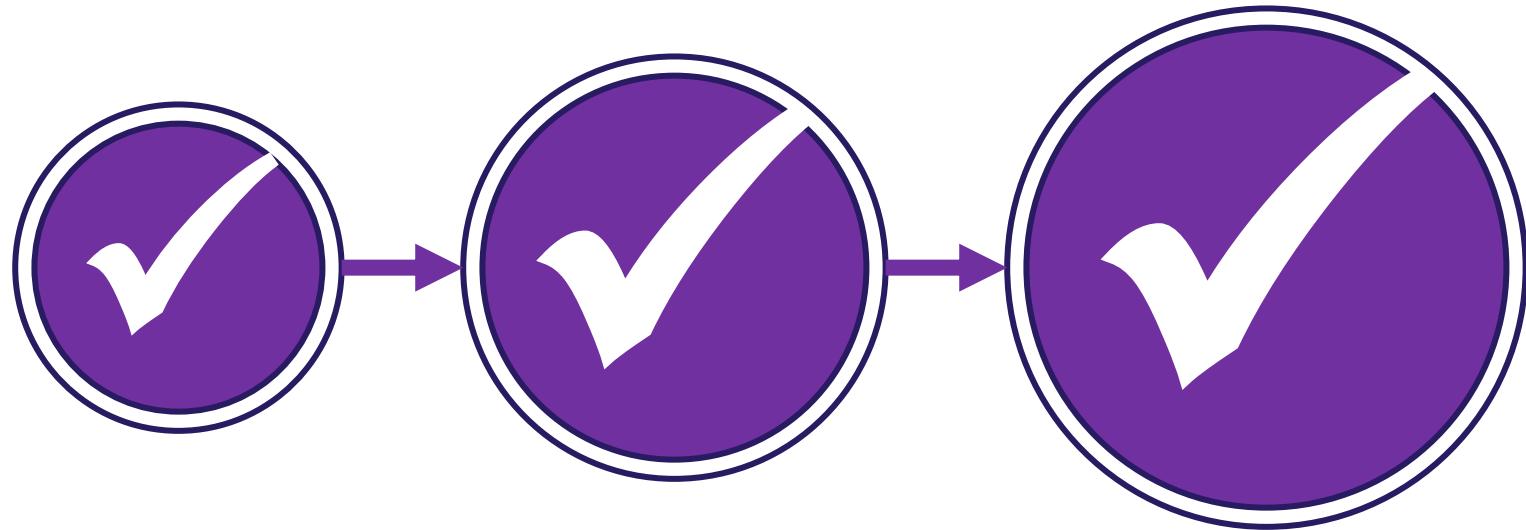


# EDA With Folium Map– Task 2 - Marking Launch Outcomes 4



# EDA With Folium Map– Task 3 – Polyline Proximity

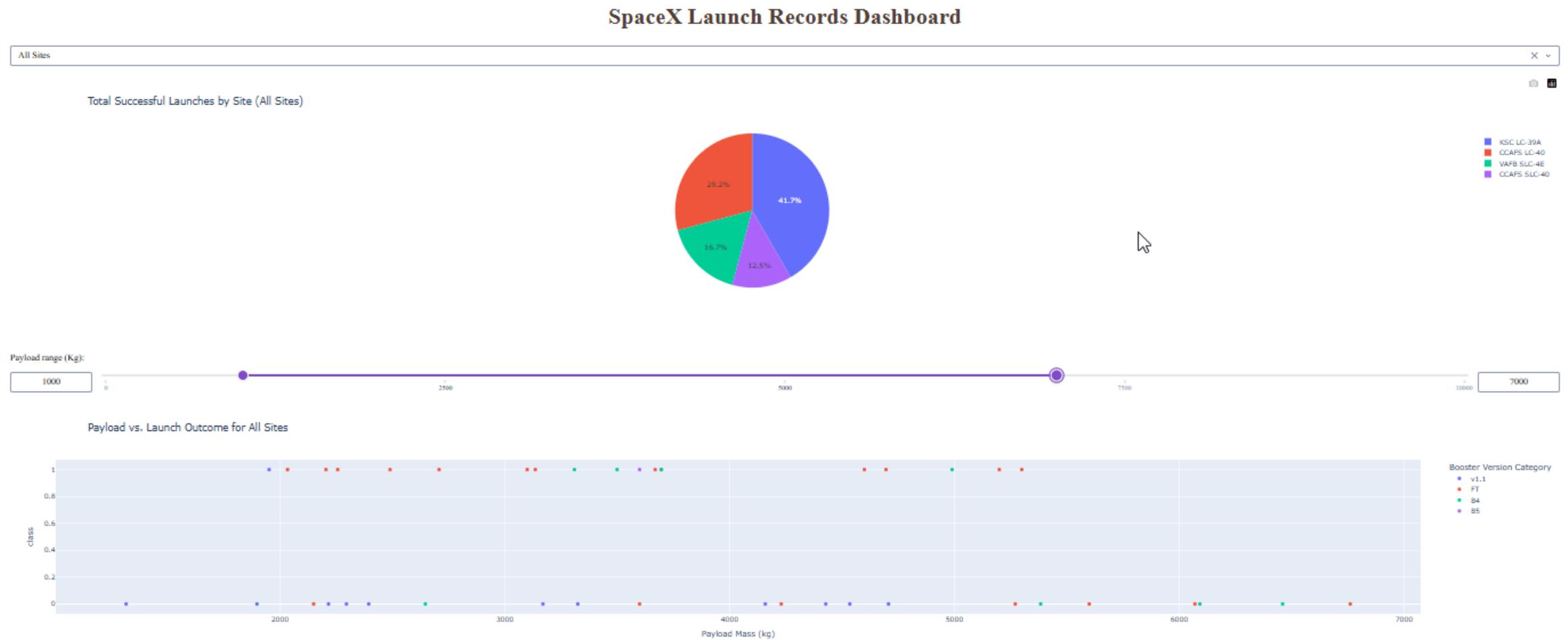




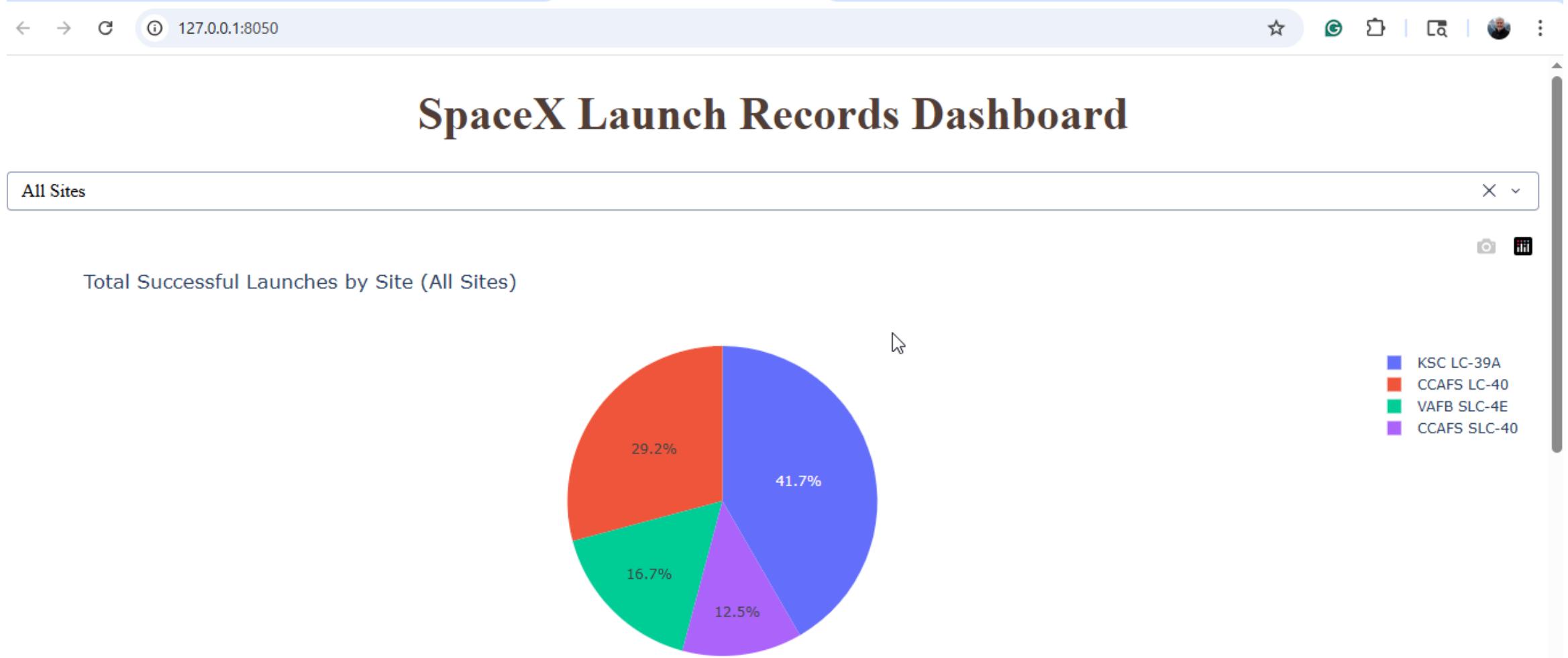
# EDA with Plotly Interactive Dashboard

Module 3 – Lab 2

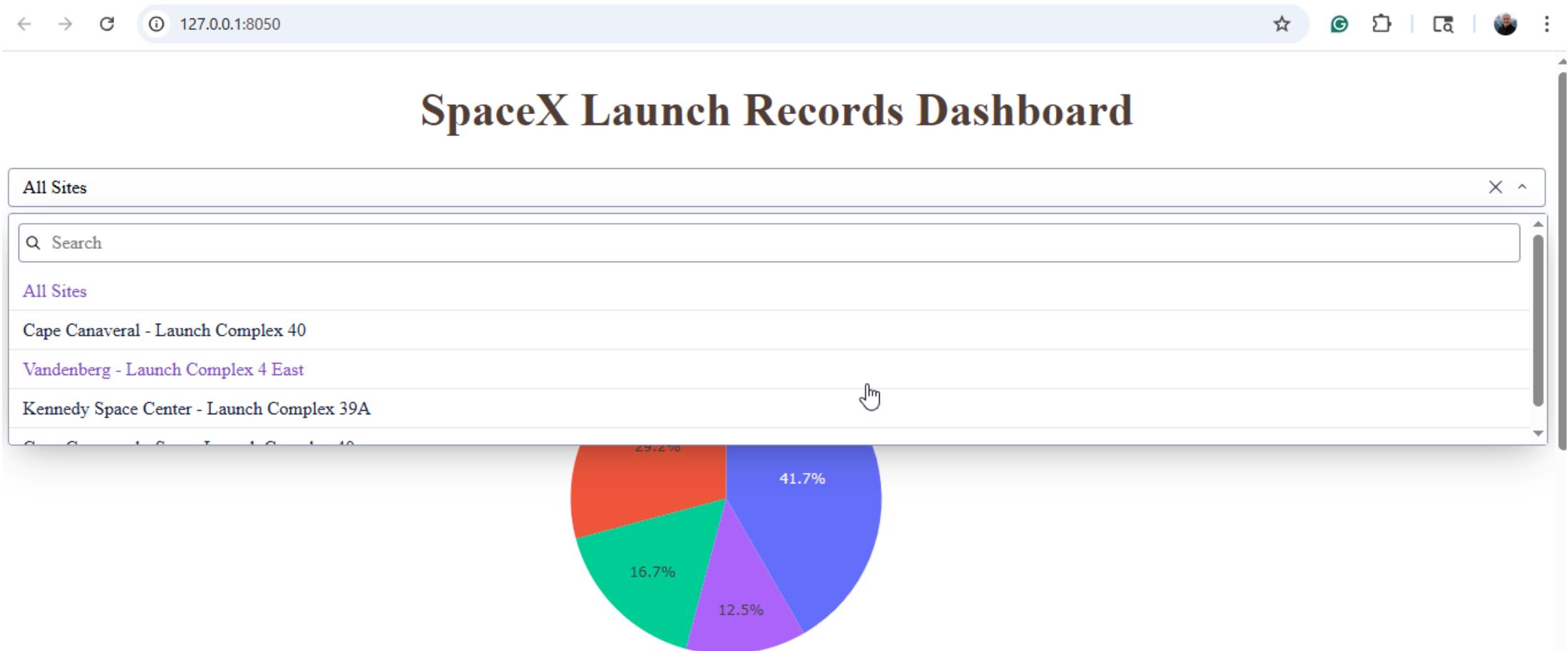
# Plotly Interactive Dashboard – Visualization 1



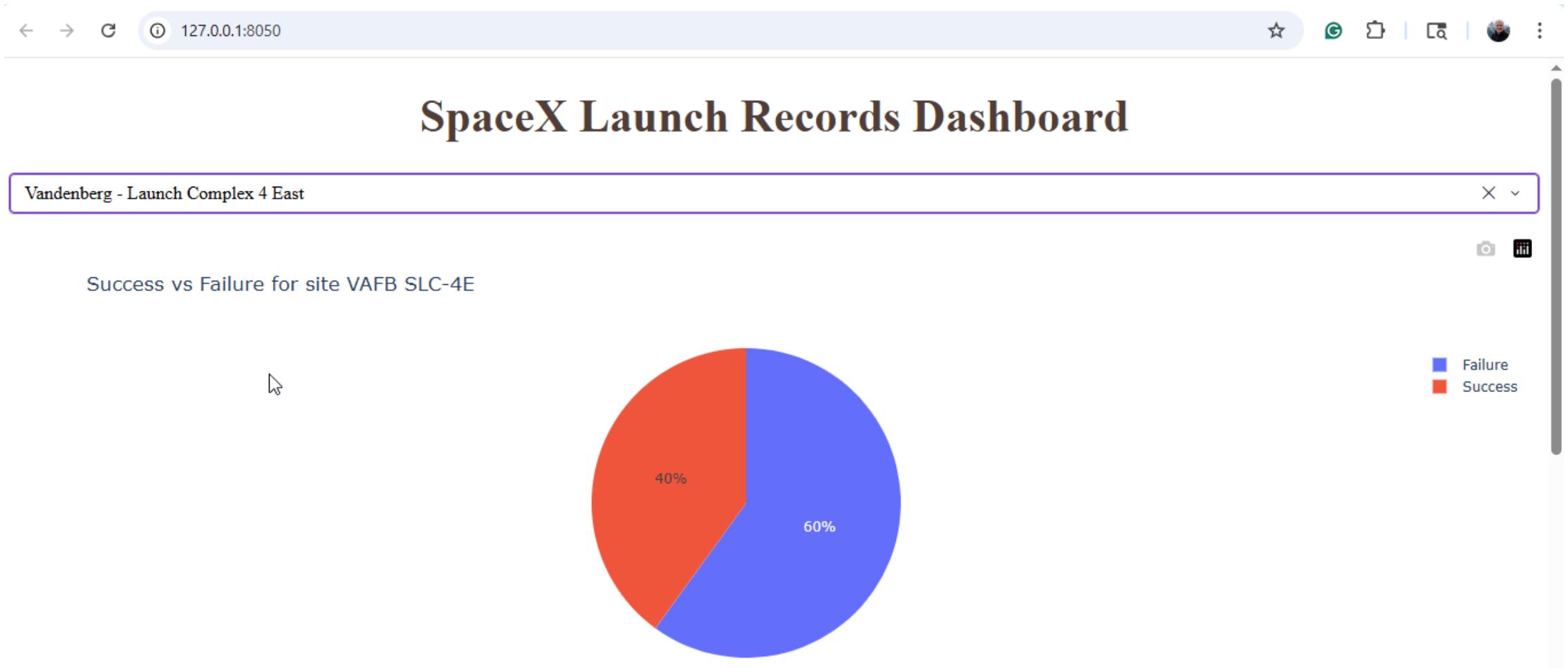
# Plotly Interactive Dashboard – Visualization 2



# Plotly Interactive Dashboard – Visualization 3



# Plotly Interactive Dashboard – Visualization 4

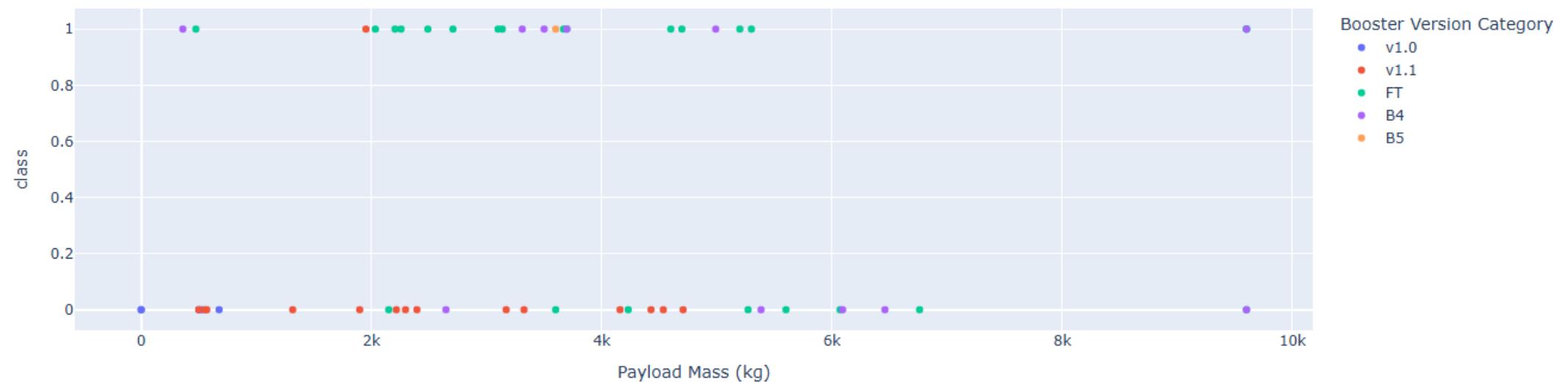


# Plotly Interactive Dashboard – Visualization 5

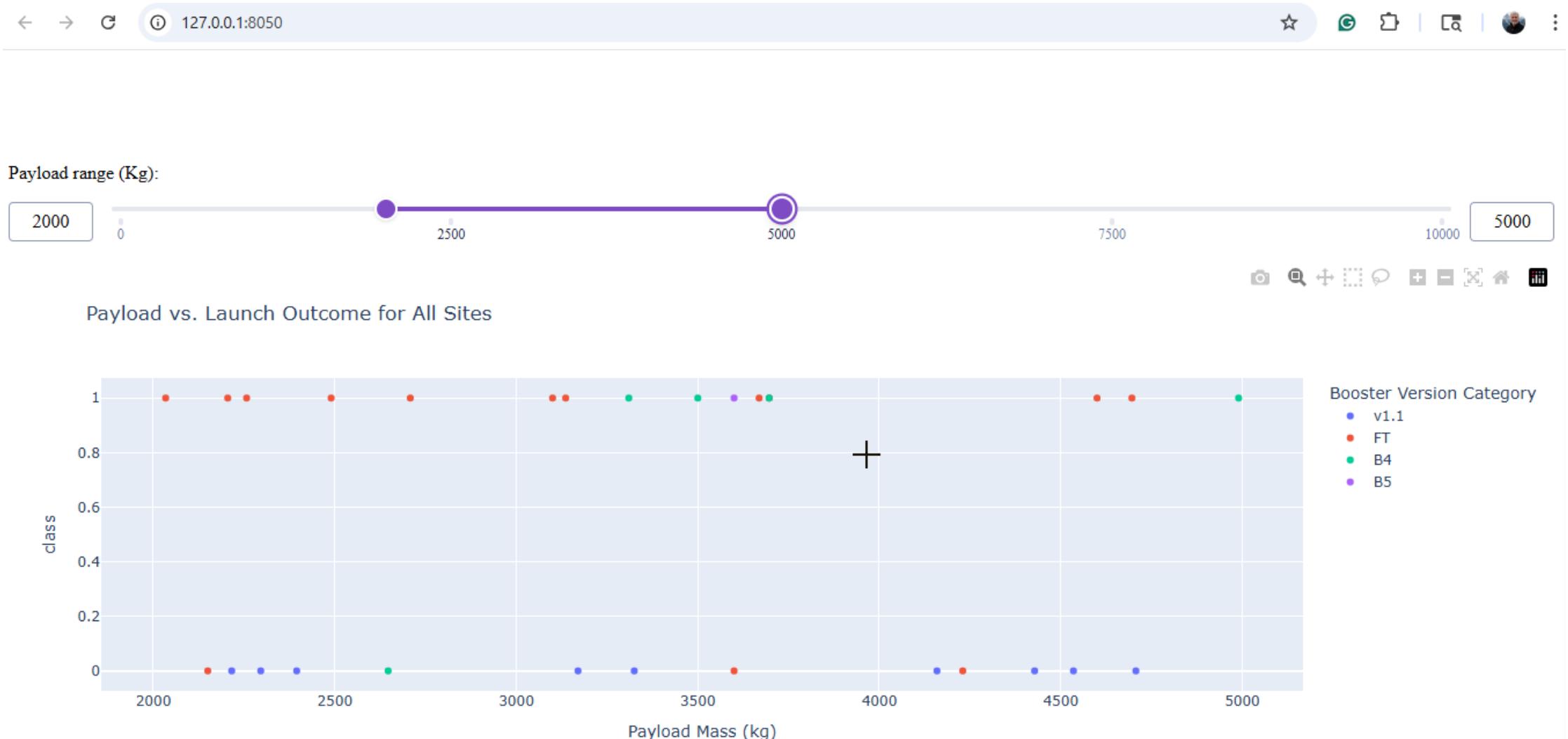
Payload range (Kg):

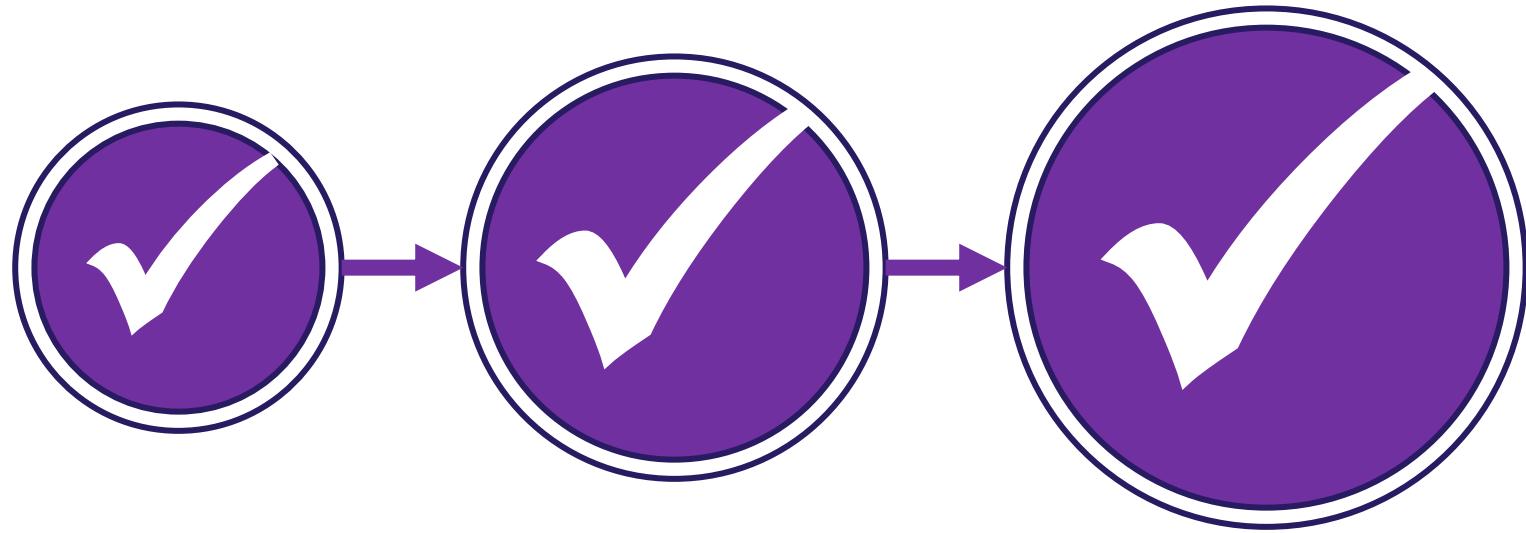


Payload vs. Launch Outcome for All Sites



# Plotly Interactive Dashboard – Visualization 6





# Predictive Analytics with Machine Learning

## Module 4 – Lab 1

# Logistic Regression – Key Metrics



## 📊 BASIC METRICS:

Accuracy: 0.7778 (77.78%)  
Precision: 0.8333  
Recall: 0.7778  
F1-Score: 0.7381

## 📈 PROBABILITY METRICS:

ROC AUC Score: 0.8611  
Log Loss: 0.4993

## 🔍 CROSS-VALIDATION:

CV Accuracy: 0.8625

## ⚙️ BEST HYPERPARAMETERS:

C: 0.01  
max\_iter: 1000  
solver: lbfgs



# Logistic Regression - Confusion Matrix



## CONFUSION MATRIX METRICS:

True Positives: 12

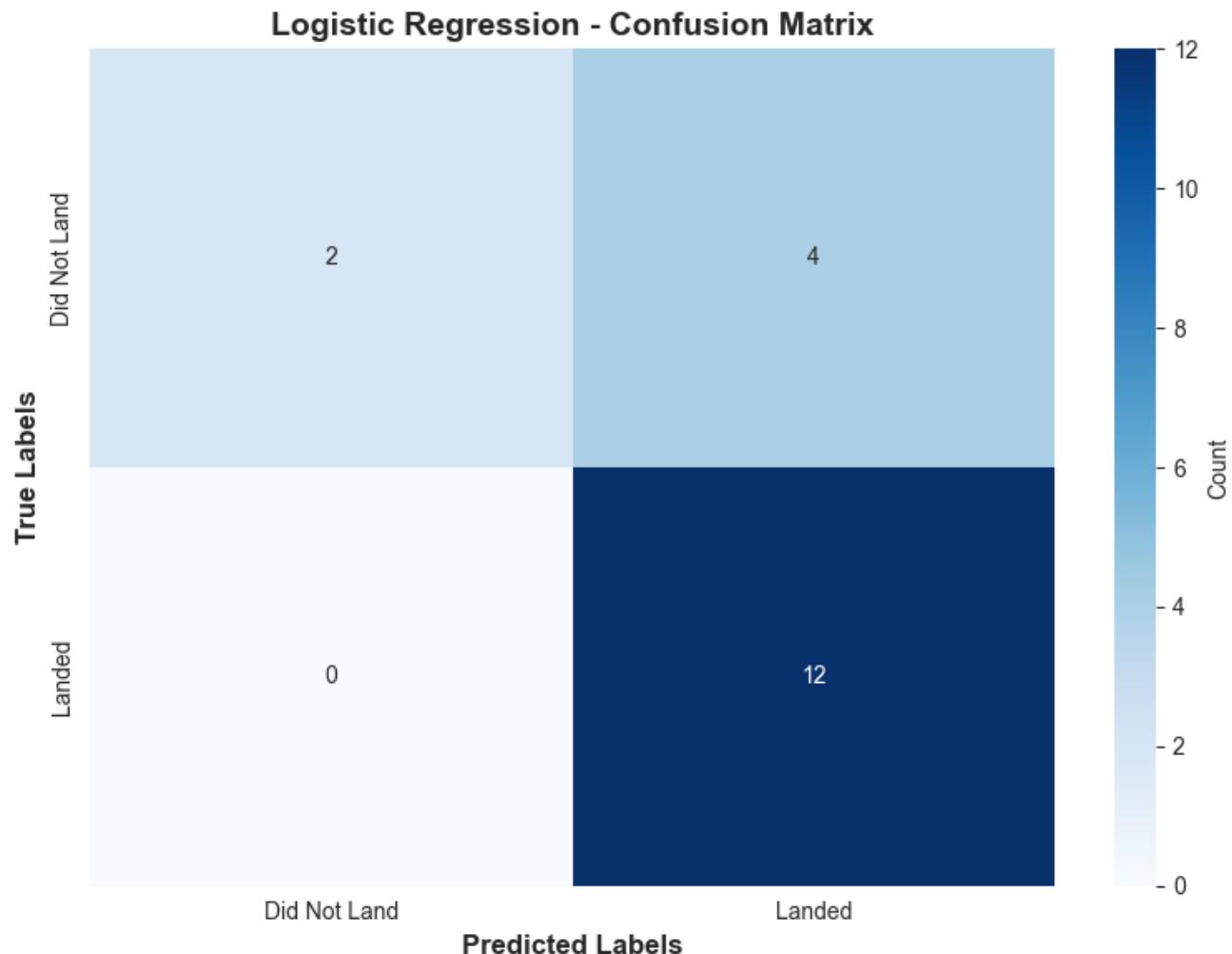
True Negatives: 2

False Positives: 4

False Negatives: 0

Sensitivity: 1.0000

Specificity: 0.3333



# Support Vector Machine (SVM) – Key Metrics



## BASIC METRICS:

Accuracy: 0.7778 (77.78%)  
Precision: 0.8333  
Recall: 0.7778  
F1-Score: 0.7381

## PROBABILITY METRICS:

ROC AUC Score: 0.8472  
Log Loss: 0.4270

## CROSS-VALIDATION:

CV Accuracy: 0.8500

## BEST HYPERPARAMETERS:

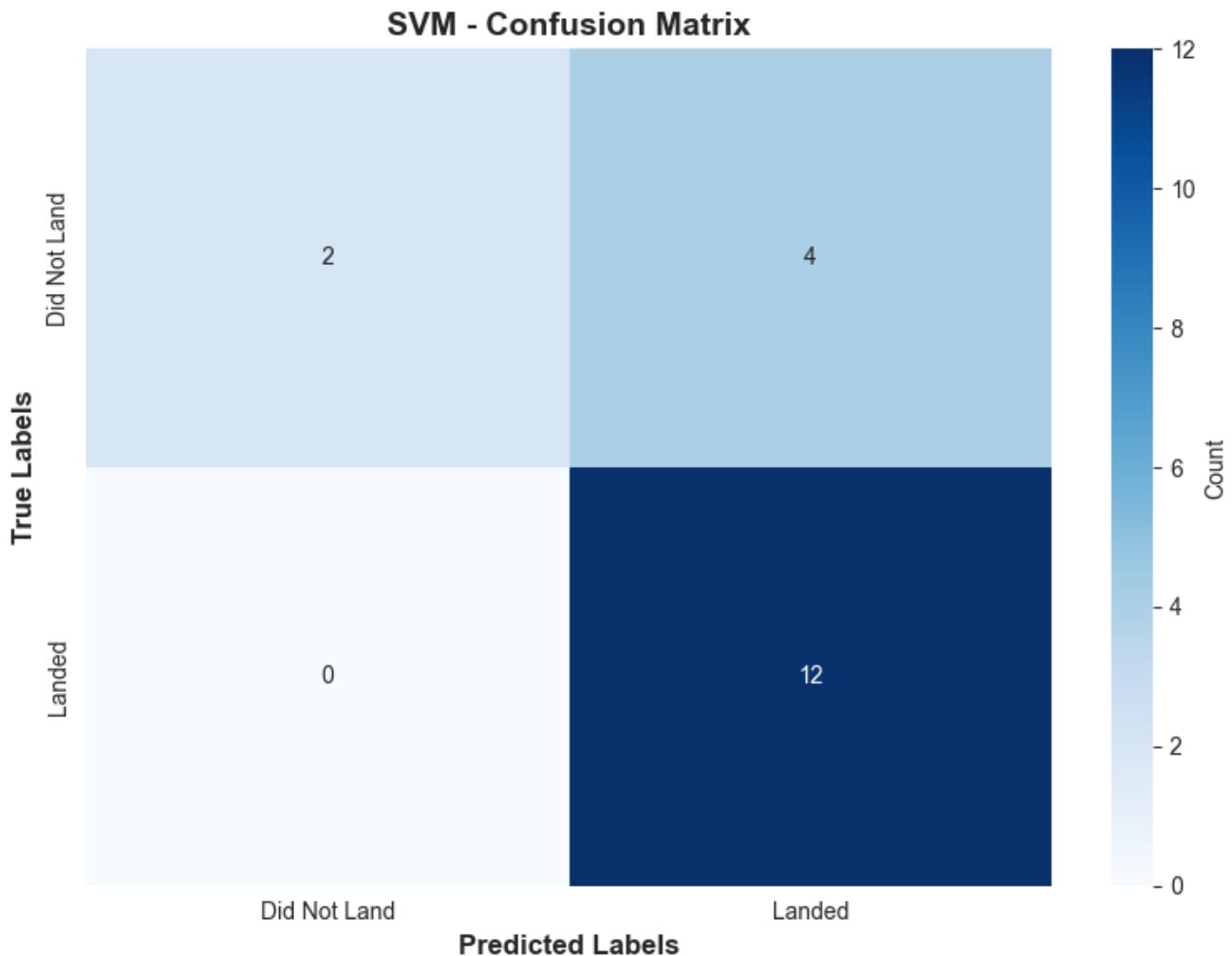
C: 1  
gamma: scale  
kernel: rbf

# Support Vector Machine (SVM) – Confusion Matrix



## CONFUSION MATRIX METRICS:

True Positives: 12  
True Negatives: 2  
False Positives: 4  
False Negatives: 0  
Sensitivity: 1.0000  
Specificity: 0.3333



# Decision Tree - Key Metrics



## BASIC METRICS:

Accuracy: 0.8333 (83.33%)  
Precision: 0.8308  
Recall: 0.8333  
F1-Score: 0.8291

## PROBABILITY METRICS:

ROC AUC Score: 0.7222  
Log Loss: 4.1583

## CROSS-VALIDATION:

CV Accuracy: 0.8893

## BEST HYPERPARAMETERS:

criterion: gini  
max\_depth: 6  
max\_features: sqrt  
min\_samples\_leaf: 1  
min\_samples\_split: 5  
splitter: best



# Decision Tree - Confusion Matrix

## 🎯 CONFUSION MATRIX METRICS:

True Positives: 11

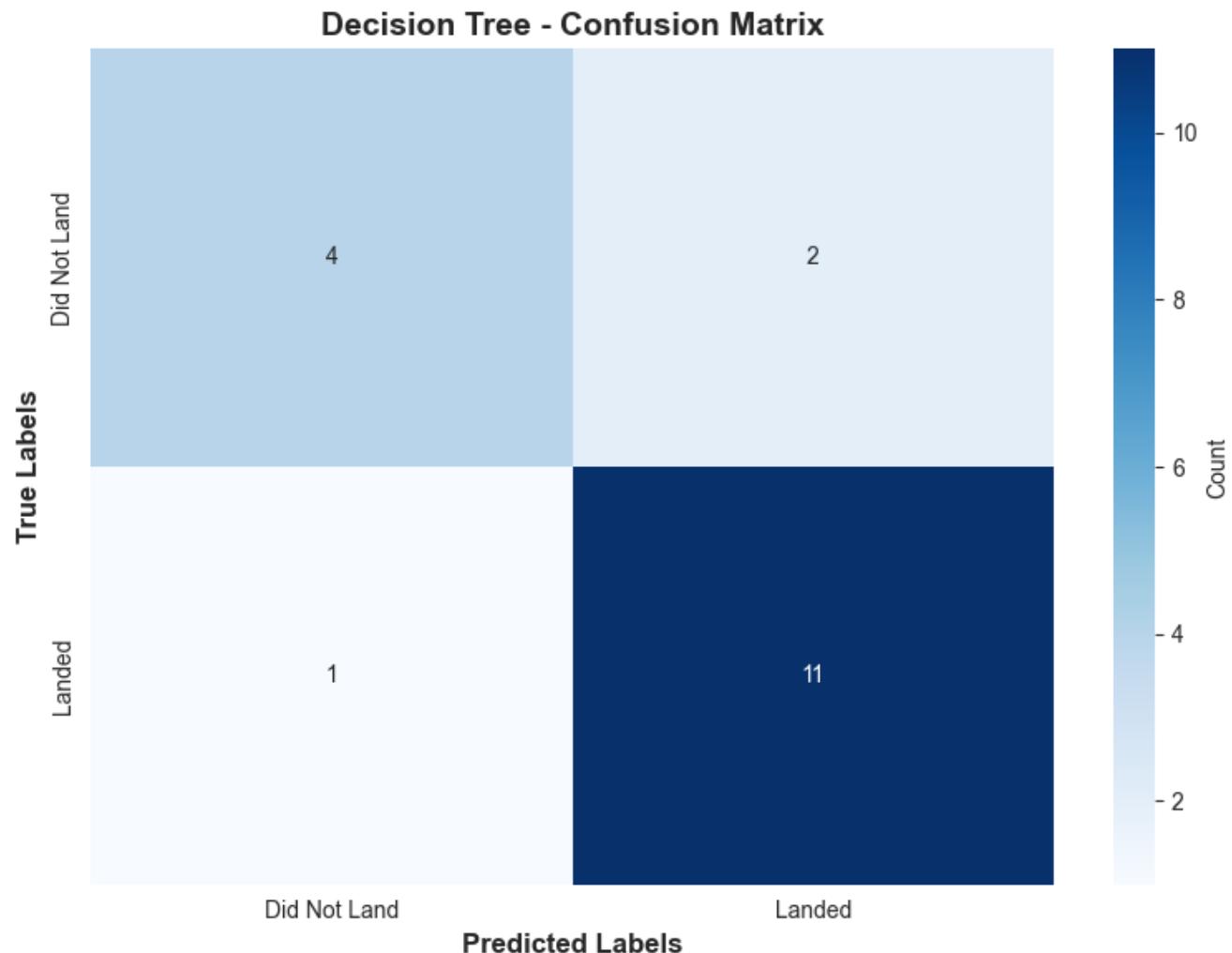
True Negatives: 4

False Positives: 2

False Negatives: 1

Sensitivity: 0.9167

Specificity: 0.6667



# K-Nearest Neighbor – Key Metrics



## BASIC METRICS:

Accuracy: 0.7778 (77.78%)  
Precision: 0.8333  
Recall: 0.7778  
F1-Score: 0.7381

## PROBABILITY METRICS:

ROC AUC Score: 0.8403  
Log Loss: 0.4568

## CROSS-VALIDATION:

CV Accuracy: 0.8786

## BEST HYPERPARAMETERS:

algorithm: auto  
n\_neighbors: 9  
p: 1  
weights: uniform



# K-Nearest Neighbor – Confusion Matrix



## CONFUSION MATRIX METRICS:

True Positives: 12

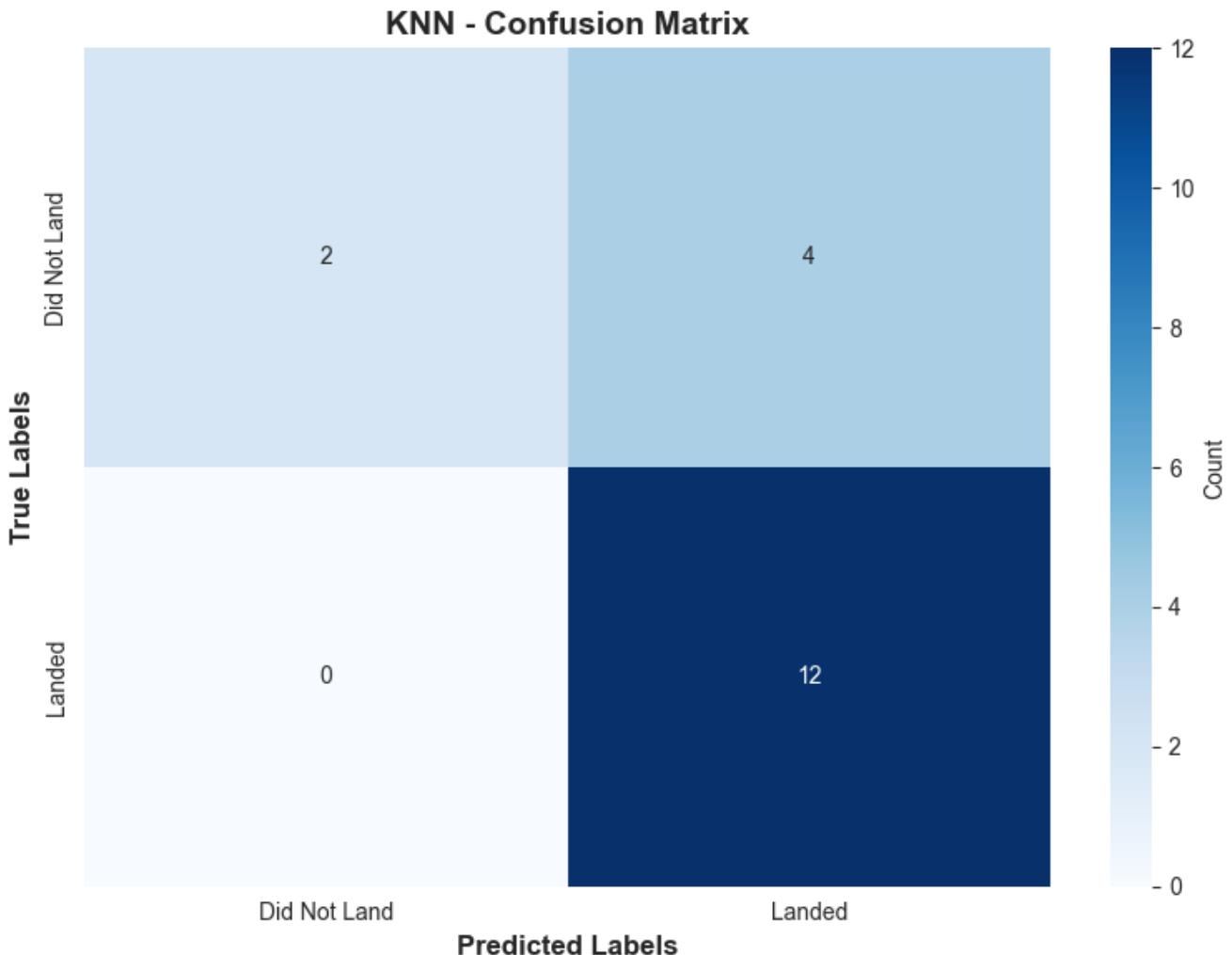
True Negatives: 2

False Positives: 4

False Negatives: 0

Sensitivity: 1.0000

Specificity: 0.3333



# Model Comparison - Conclusion

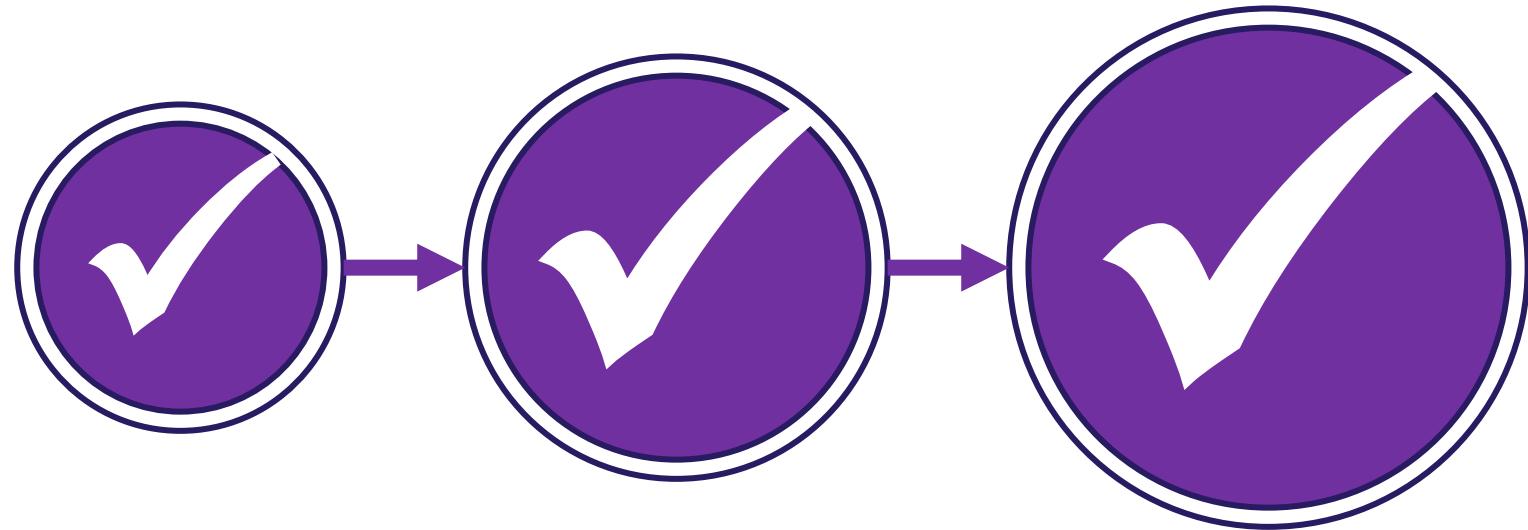
Metric	Logistic Regression	SVM	Decision Tree	KNN
Accuracy	0.7778	0.7778	0.8333	0.7778
Precision	0.8333	0.8333	0.8308	0.8333
Recall	0.7778	0.7778	0.8333	0.7778
F1-Score	0.7381	0.7381	0.8291	0.7381
CV Accuracy	0.8625	0.8500	0.8893	0.8786
Specificity	0.3333	0.3333	0.6667	0.3333
Sensitivity	1.0000	1.0000	0.9167	1.0000
ROC AUC	0.8611	0.8472	0.7222	0.8403
Log Loss	0.4993	N/A	N/A	N/A



**BEST MODEL:** Decision Tree - Test Accuracy: 0.8333 (83.33%)



**BEST CV SCORE:** Decision Tree - CV Accuracy: 0.8893 (88.93%)



# Appendices

# Appendix: GitHub Repository Link

---

**<https://github.com/andyflorence/IBM-Capstone-Project>**

# Appendix: About the SpaceX API Data

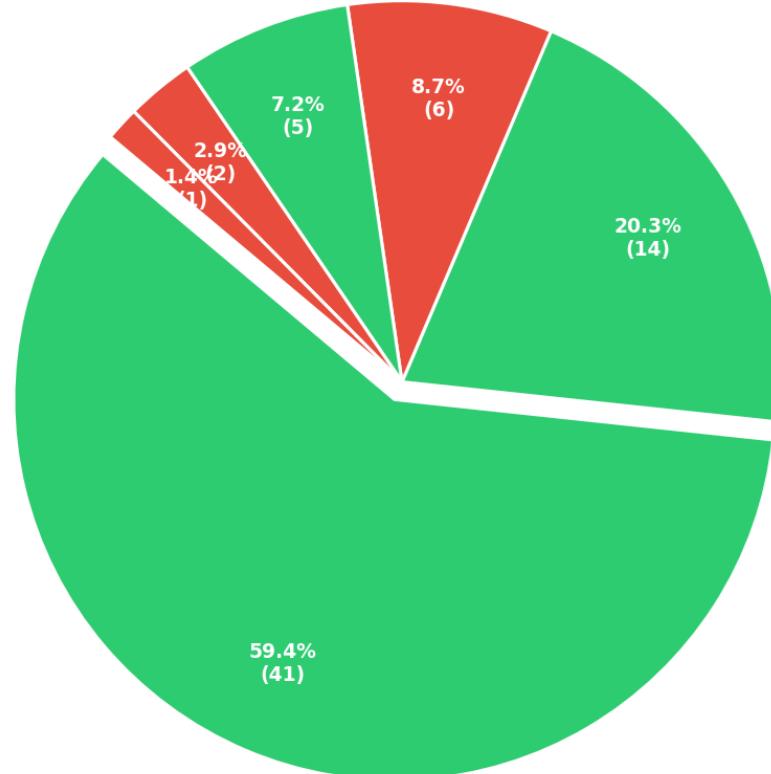
## Data Types:

FlightNumber int64  
Date object  
BoosterVersion object  
PayloadMass float64  
Orbit object  
LaunchSite object  
Outcome object  
Flights int64  
GridFins bool  
Reused bool  
Legs bool  
LandingPad object  
Block float64  
ReusedCount int64  
Serial object  
Longitude float64  
Latitude float64

## Missing Values:

FlightNumber 0  
Date 0  
BoosterVersion 0  
PayloadMass 3  
Orbit 0  
LaunchSite 0  
Outcome 0  
Flights 0  
GridFins 0  
Reused 0  
Legs 0  
LandingPad 7  
Block 0  
ReusedCount 0  
Serial 0  
Longitude 0  
Latitude 0

**SpaceX Falcon 9 — First Stage Landing Outcome Distribution**



Total launches: 69 | Successful landings: 60 | Failed/Other: 9