IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akash Walivkar
25-Oct-2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- In this project will we try to predict if the Falcon 9 first stage will land successfully using different machine learning classification algorithm

- We used following steps for the project

    1. Data Collection & wrangling
    2. EDA
    3. Data Visualization
    4. Machine learning

- In data Visualization some features have correlation with the outcome of the launches.

- In Machine learning we performed for decision tree, K- Nearest Neighbour, SVM, Logistic regression

Introduction

- We will predict if the falcon 9 first stage will land successfully, The main mission here to reuse the rockets and save partial cost for each launch there fore the first stage launch is important.
- Most of unsuccessful landing are planned
- predicting whether the first stage of a Falcon 9 rocket will land successfully given features such as payload mass, orbit type, launch site, etc.—the task can be reframed as a binary classification problem in machine learning

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- The API use:- https://api.spacexdata.com/v4/rockets/.
- Api provided data for different rockets
- We have 90 instances and 17 features

Data Collection – SpaceX API

First we go for latest launch

then we go for rocket json to fetch rocket id and then payload id

Data Collection - Scraping

- The data is scrapped from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- The link have a data of falcon
- We use beautiful soup 4 for the web scrapping

Data Wrangling

- The Data wrangling includes no missing values and categorical features are encoaded
- Extra column called class added & class contain 0 for failed launches & 1 for all the successful
- In the data we have 90 rows & 83 columns

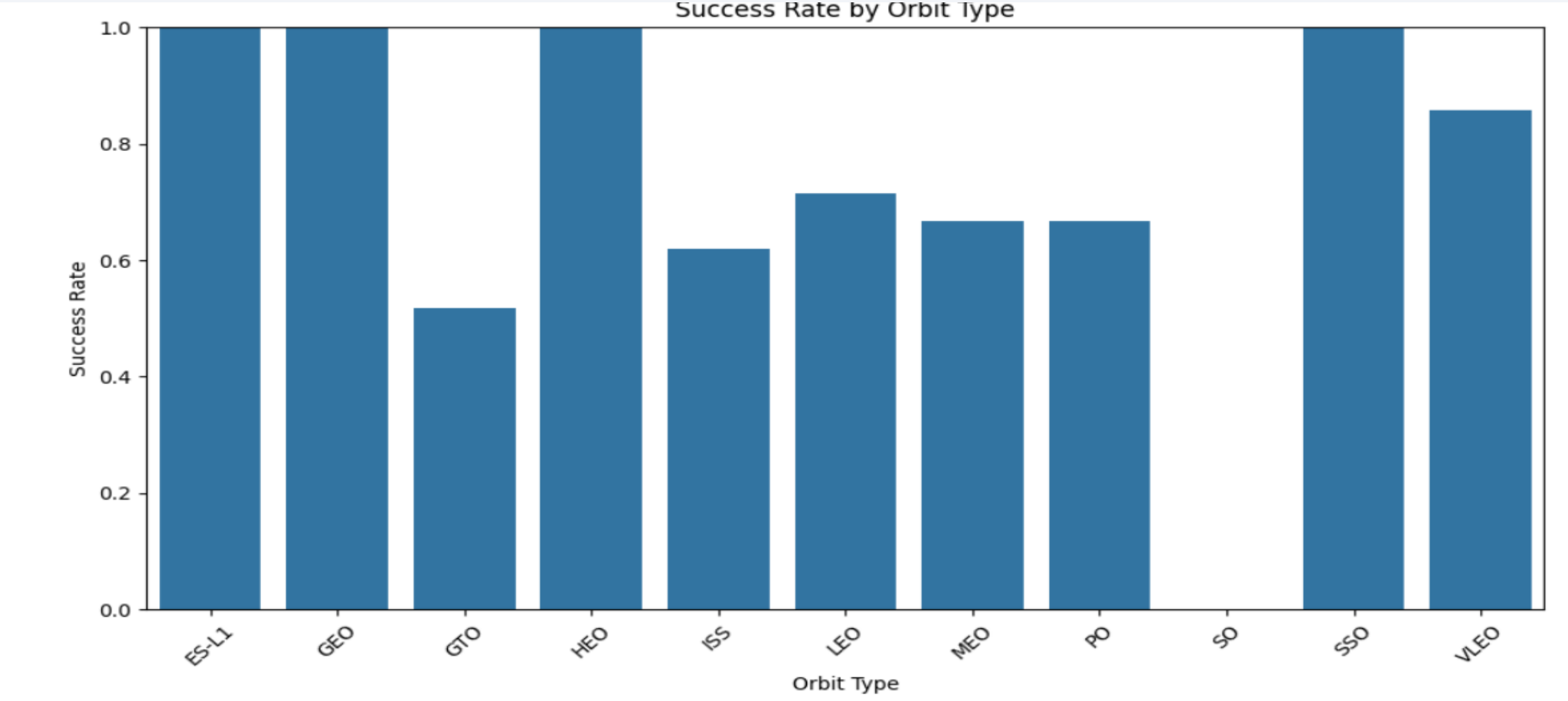
EDA with Data Visualization

- Pandas and NumPy are used to analyze launch data, allowing you to count the number of launches for each launch site.
- You can compute the frequency of each orbit type to understand their distribution.
- They also help summarize the number and frequency of each mission outcome, providing a clear overview of launch results.

EDA with Data Visualization

- Matplotlib and Seaborn were used to create visualizations such as scatterplots, bar charts, and line charts.
- These plots helped explore how **flight number**, **payload mass**, and **launch sites** are related.
- The charts also revealed insights into **success rates** across different **orbit types**.

EDA with Data Visualization



EDA with SQL

SQL was used to list all unique SpaceX launch sites
It calculated total payload mass carried on NASA (CRS) missions.
It computed the average payload mass for Falcon 9 v1.1 booster launches.

Average_Payload_Mass
2928.4

First_Successful_Ground_Landing
2015-12-22

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Build an Interactive Map with Folium

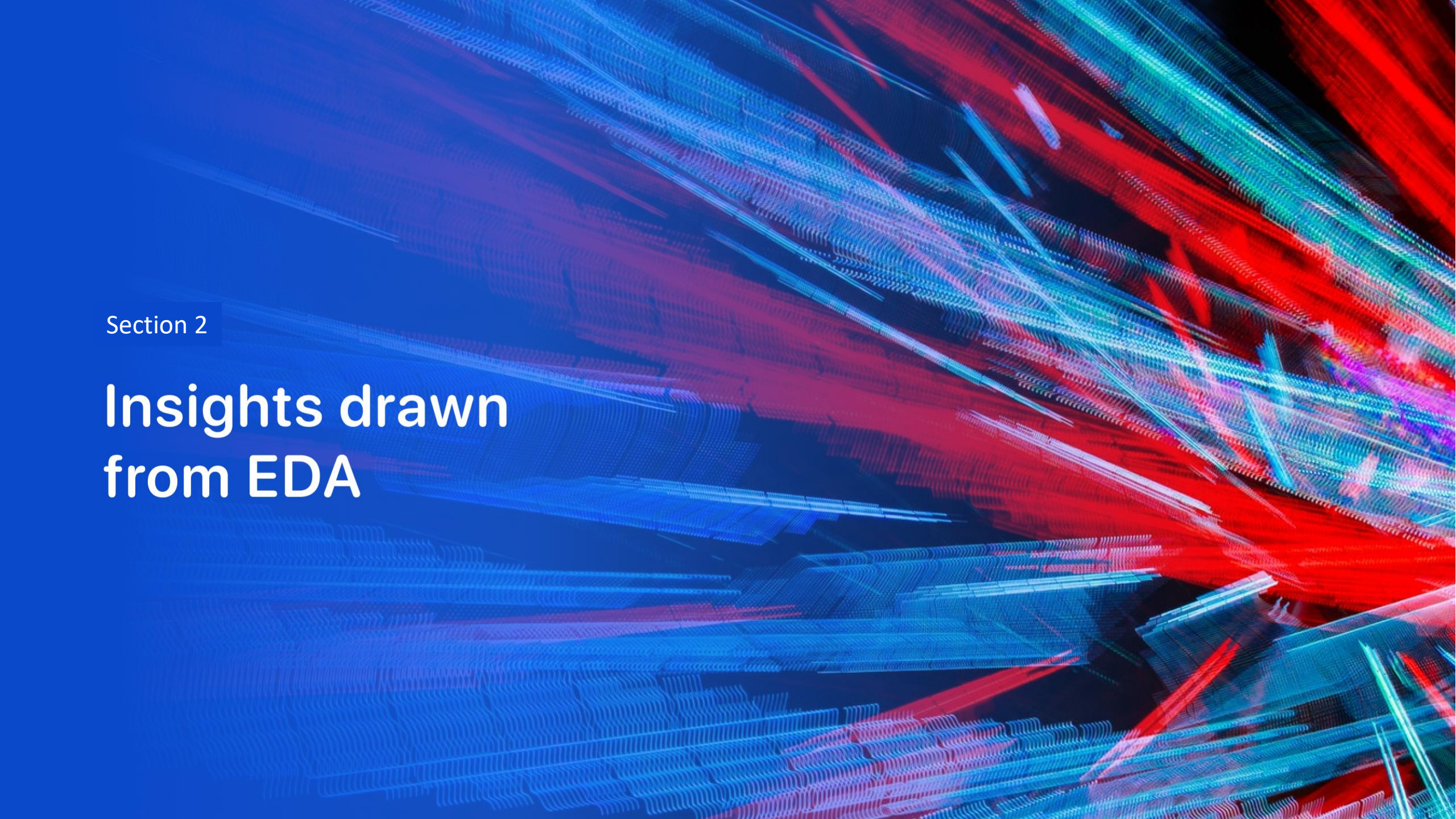
Folium created **interactive maps** to visualize the geographic locations of SpaceX launch pads.
Markers were placed at each **launch site** to display site names and mission outcomes.
Additional features like **circle markers** and **distance overlays** helped analyze proximity to nearby infrastructure, which supports launch safety and logistics.

Build a Dashboard with Plotly Dash

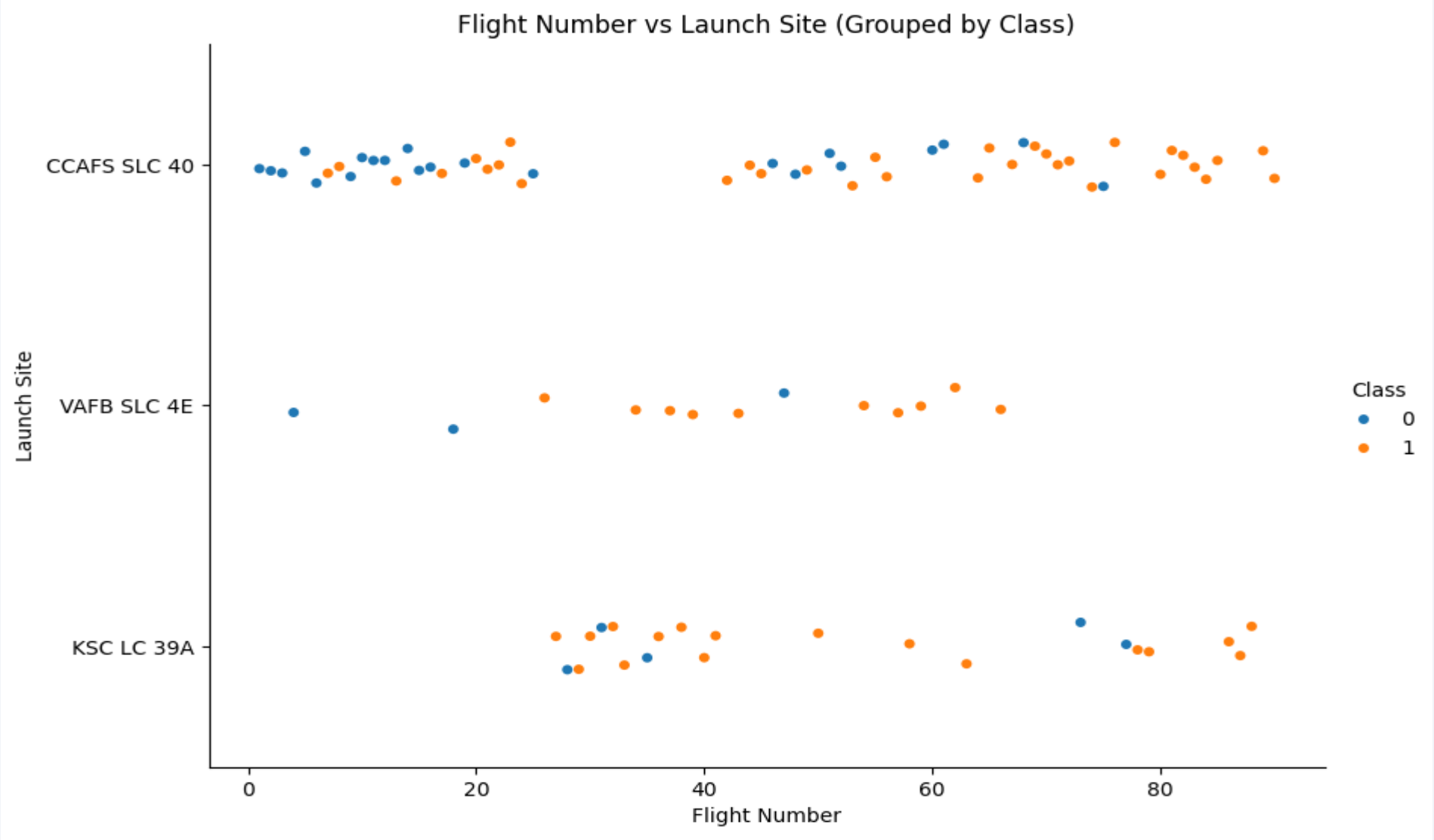
- Dash enables building an **interactive dashboard** to explore launch statistics in real-time.
- Users can **filter by launch site and payload range** to dynamically update charts.
- It provides **visual insights** into launch success rates and payload performance across different SpaceX sites.

Predictive Analysis (Classification)

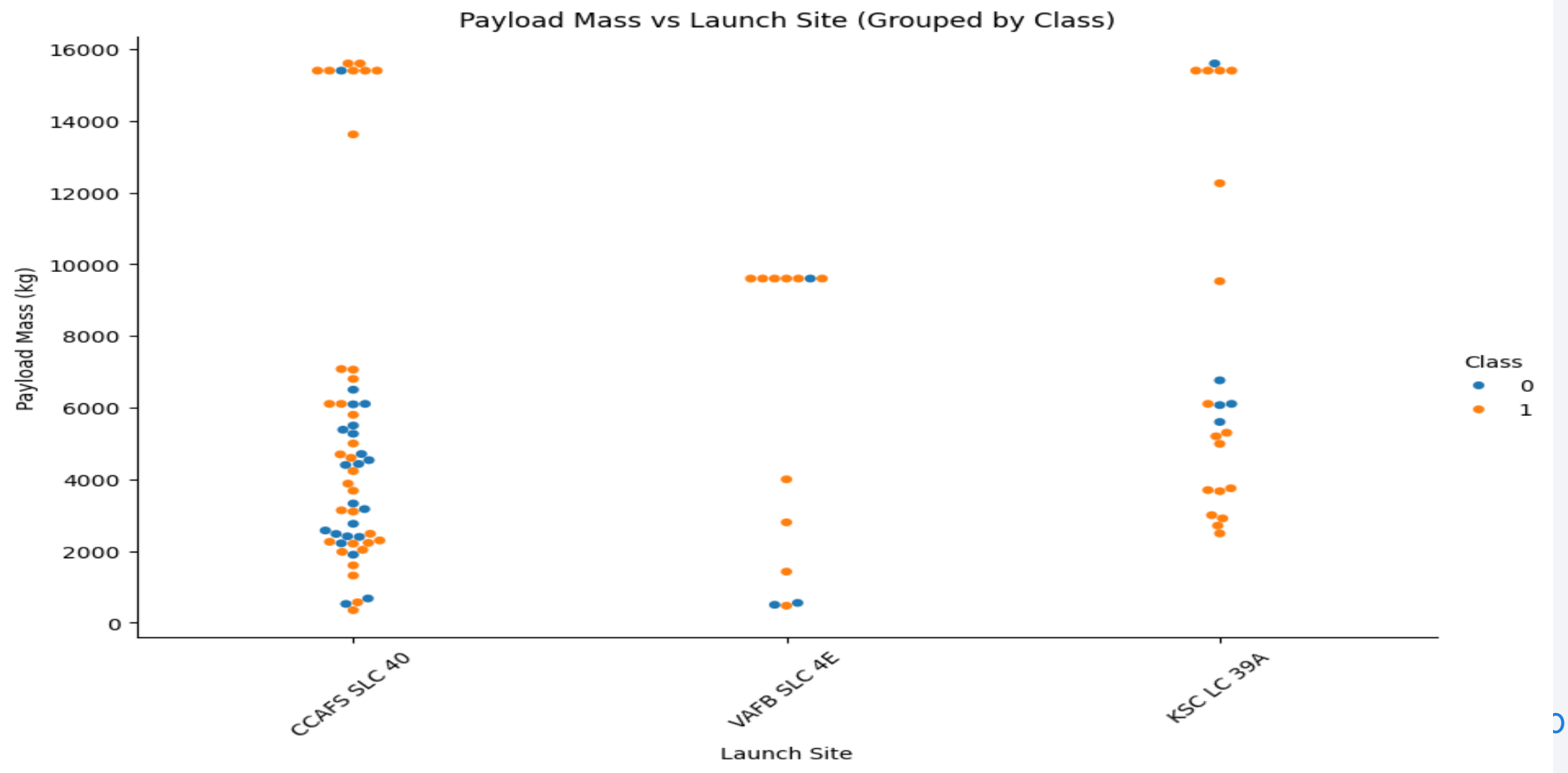
- We used four model linear regression, Decision Trees, KNN neighbor and SVM
- From the analysis we found that logistic regression gives best result
- I couldn't run SVM


Section 2
Insights drawn from EDA

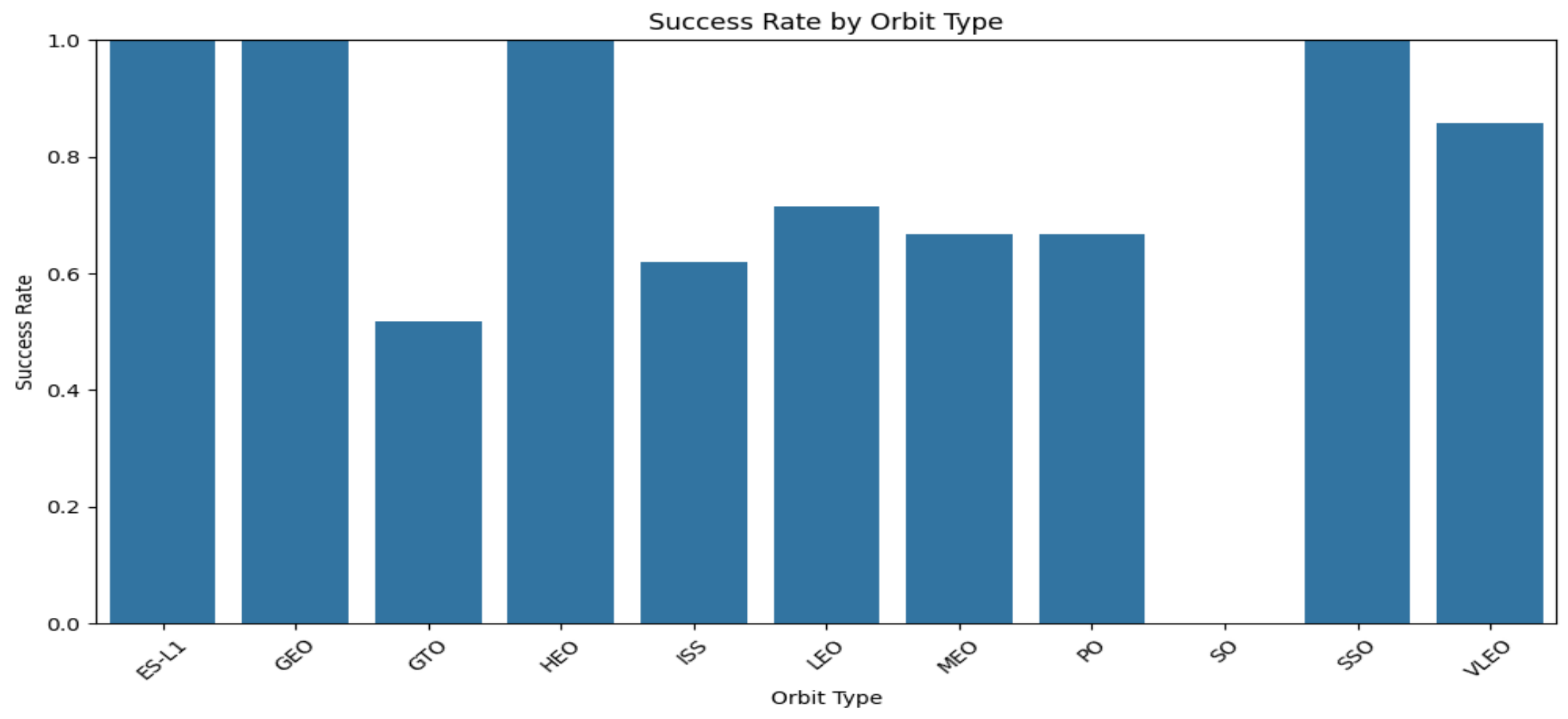
Flight Number vs. Launch Site



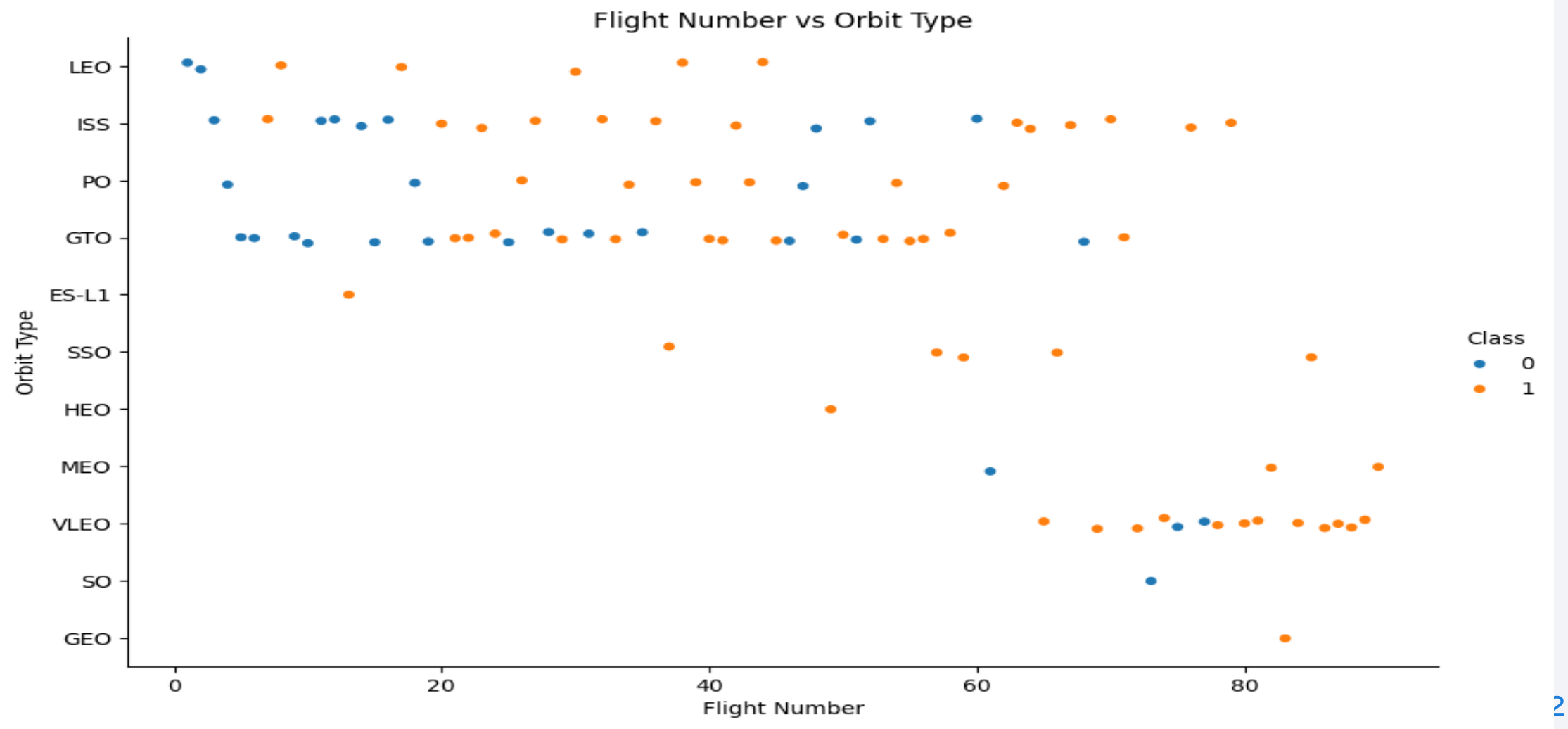
Payload vs. Launch Site



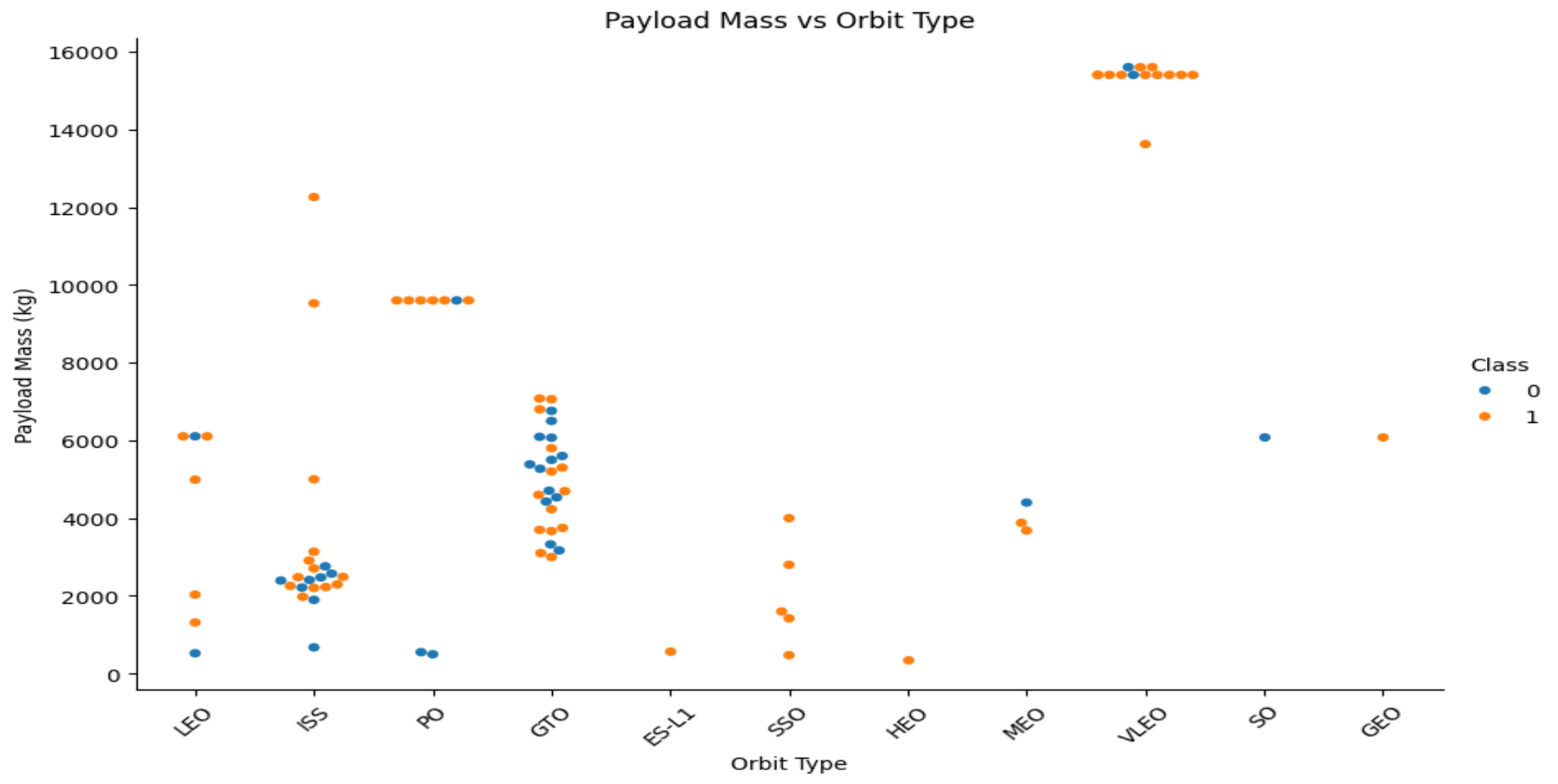
Success Rate vs. Orbit Type



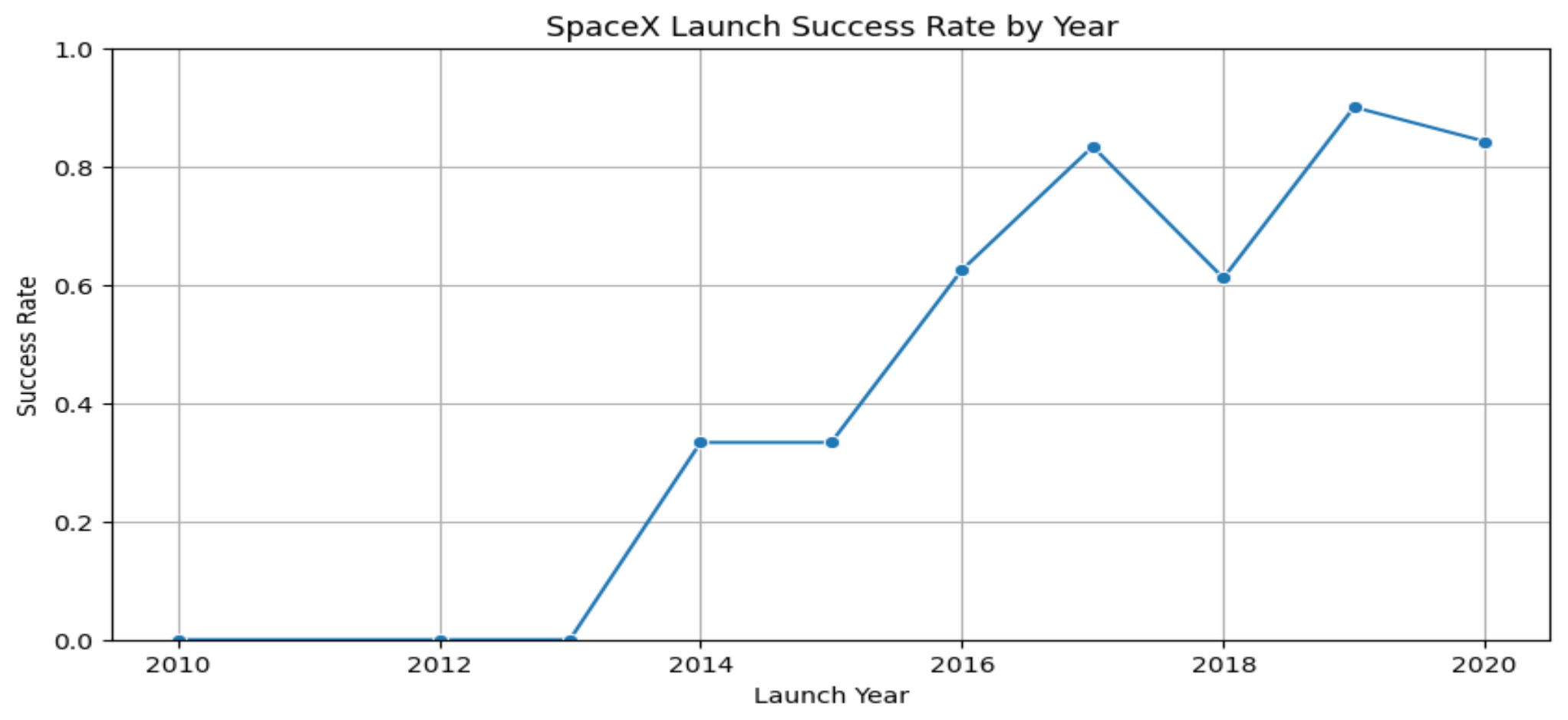
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with '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	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

Total Payload Mass

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

First_Successful_Ground_Landing

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum 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

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

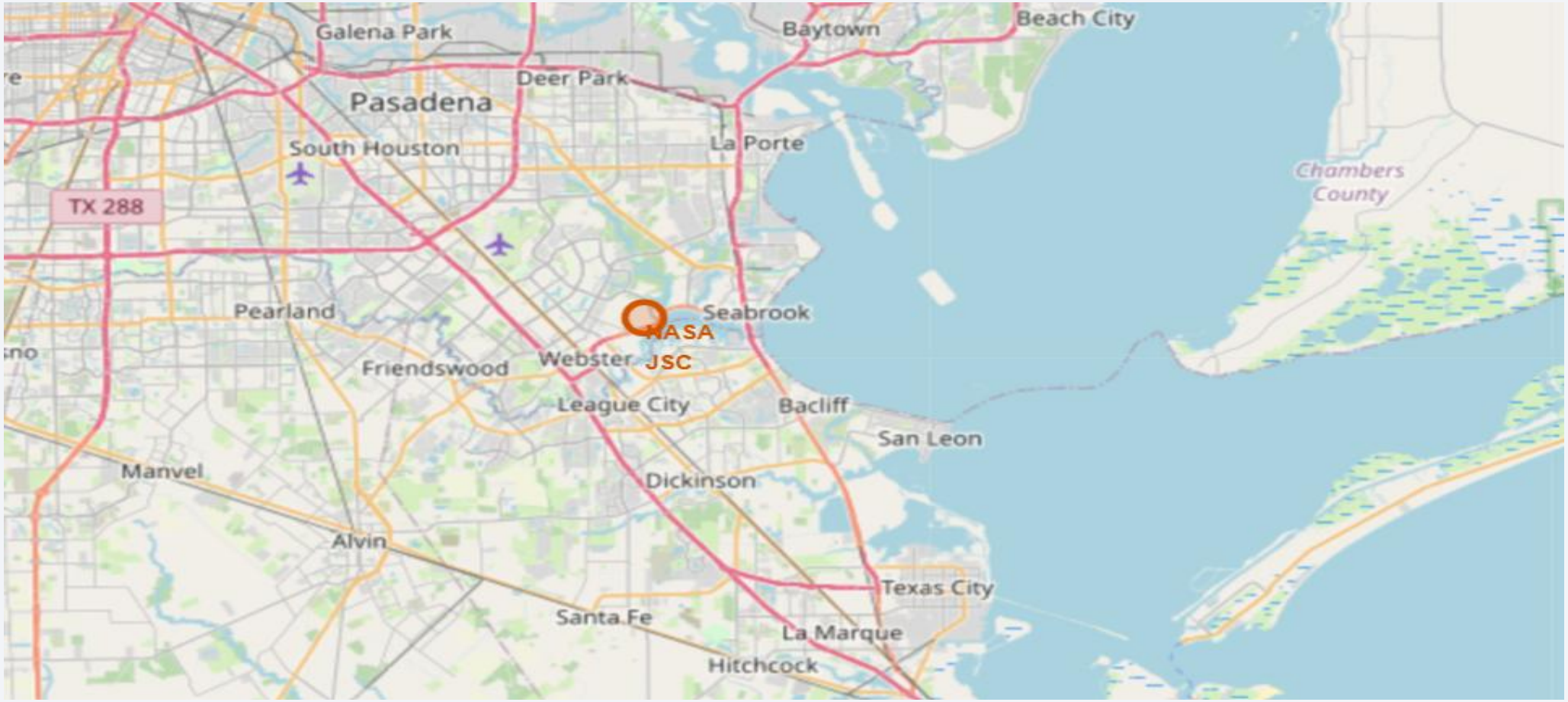
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	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

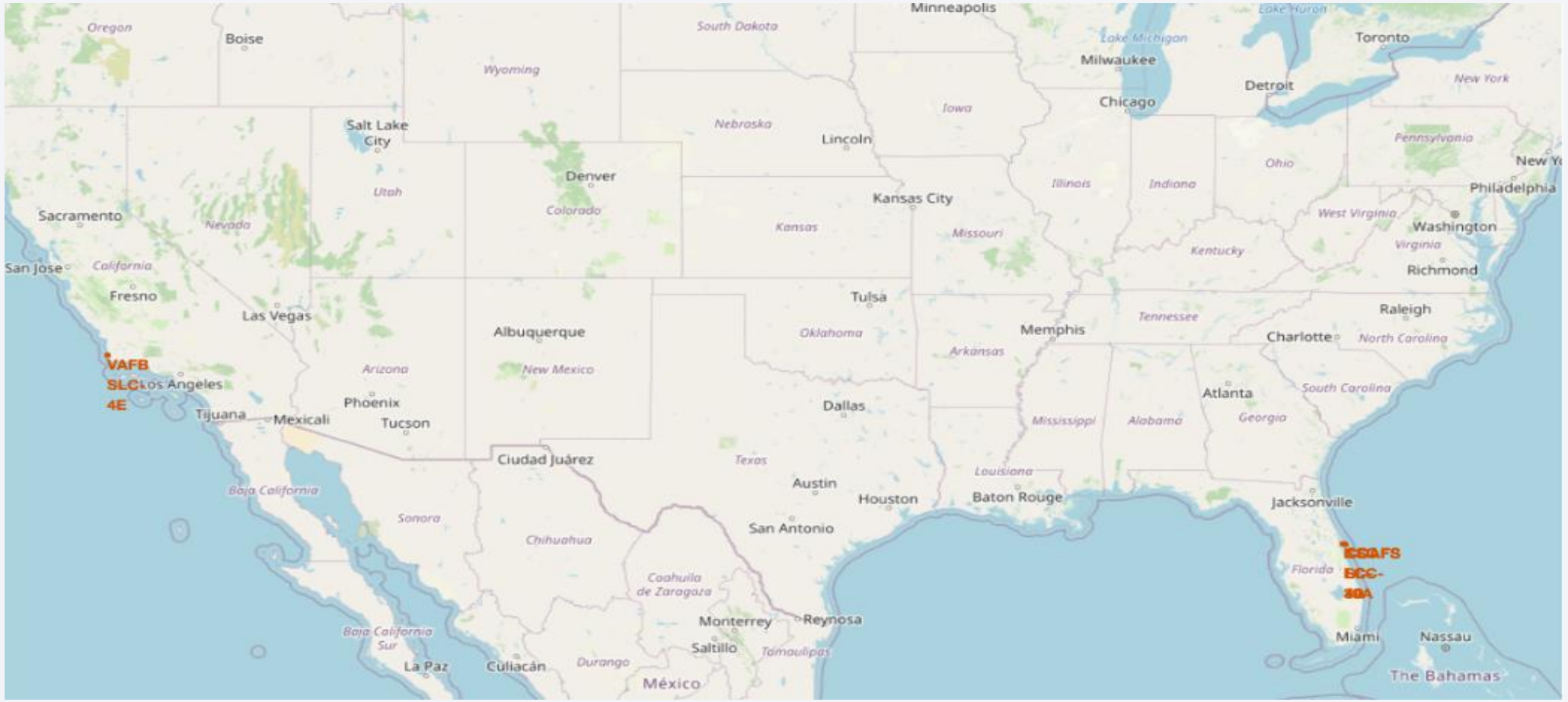
Section 3

Launch Sites Proximities Analysis

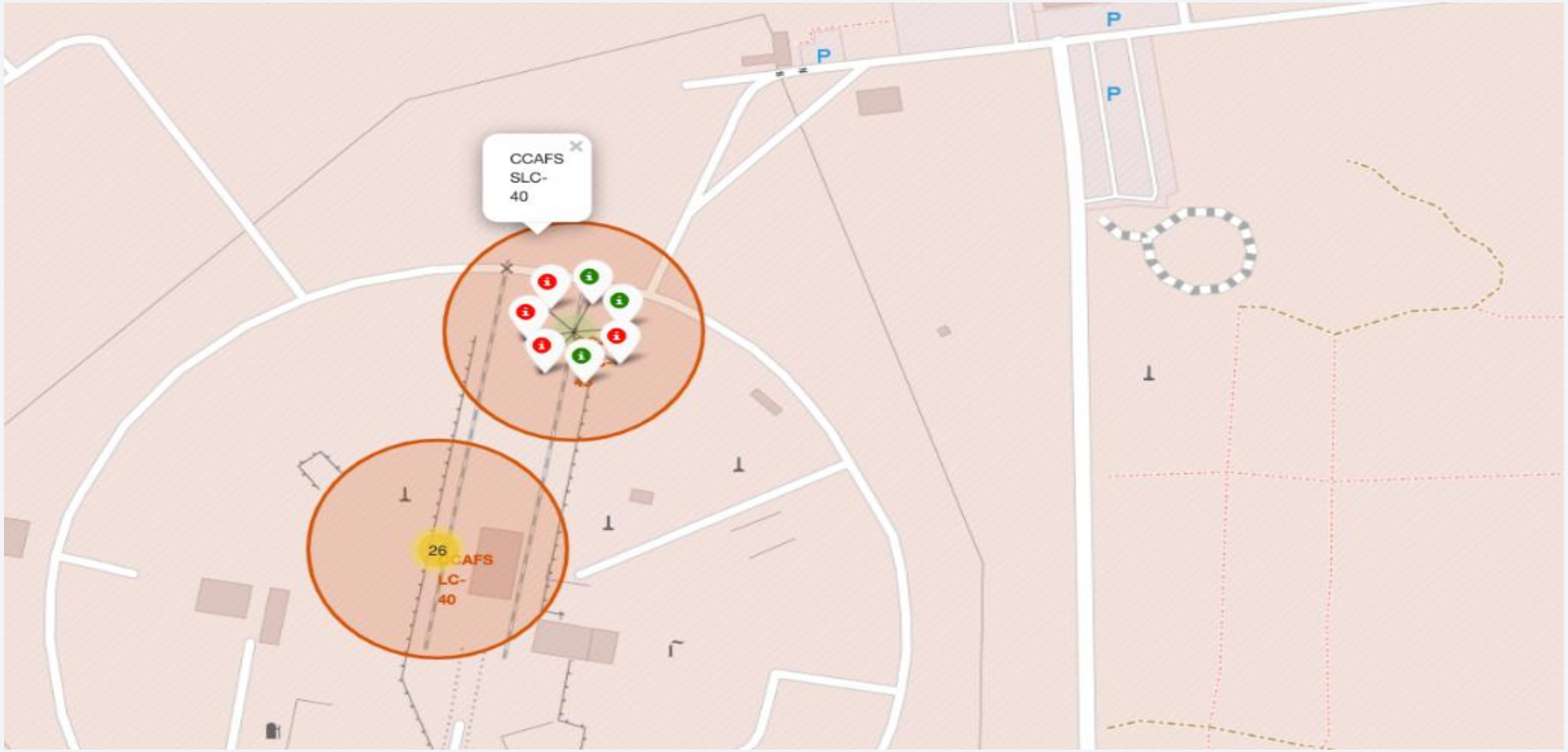
<Folium Map Screenshot 1>

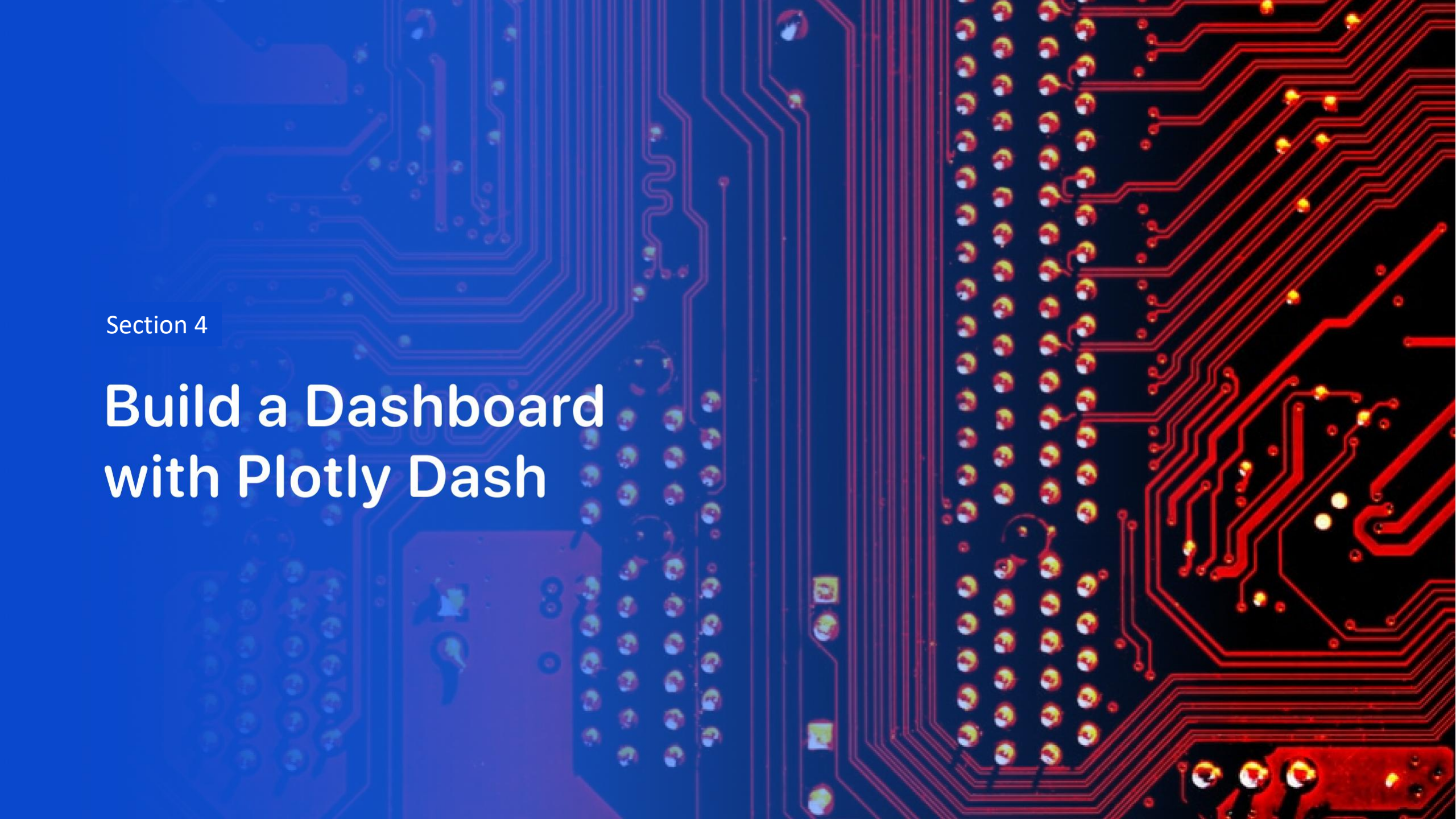


<Folium Map Screenshot 2>

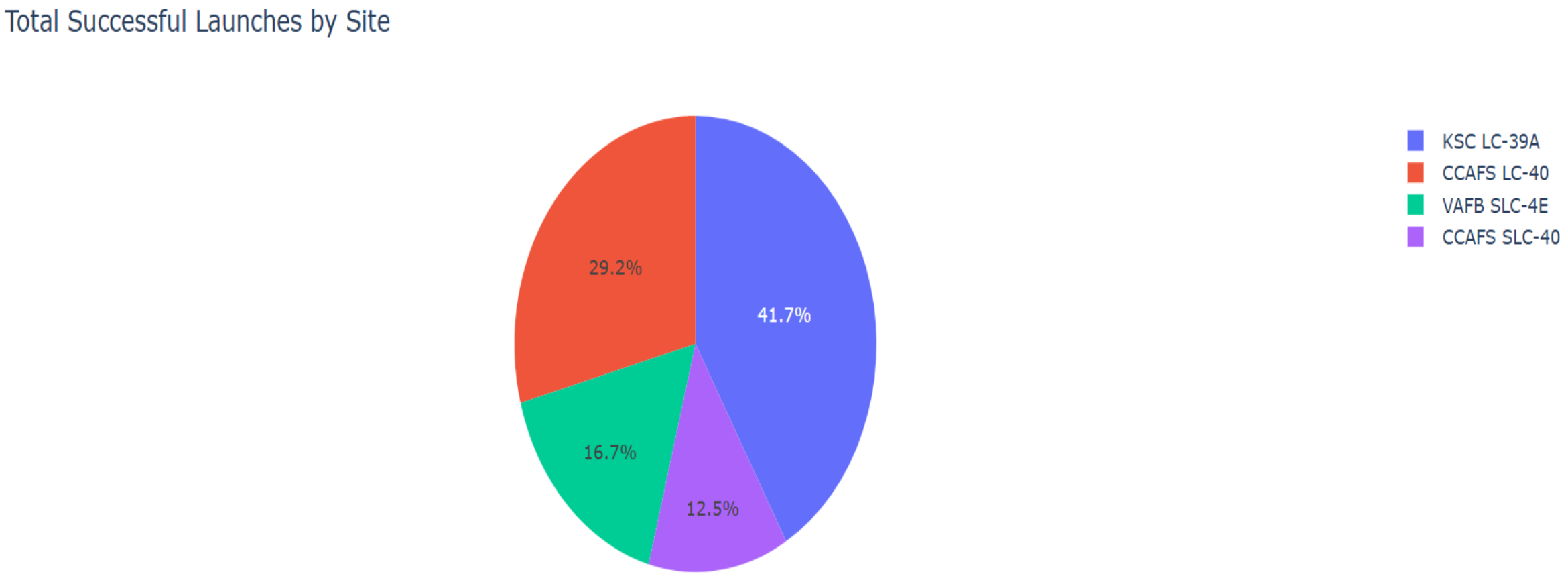


<Folium Map Screenshot 3>

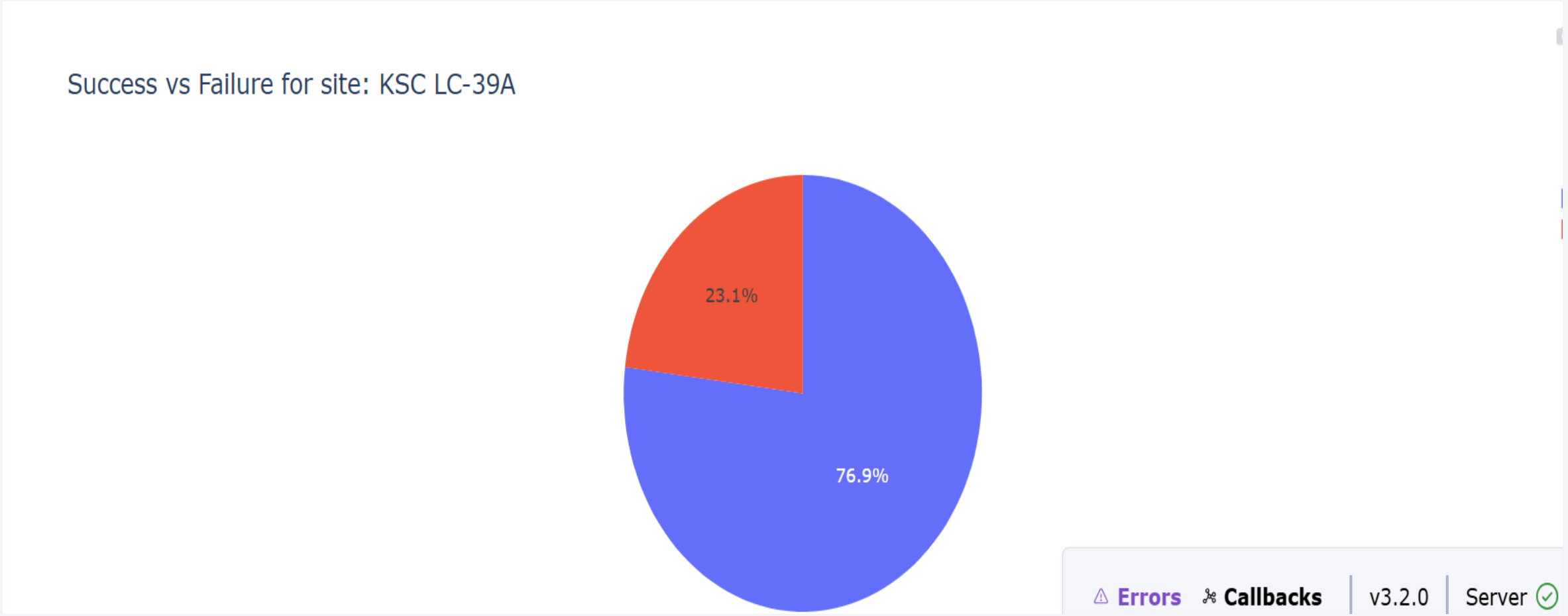



Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



<Dashboard Screenshot 3>

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Logistic regression

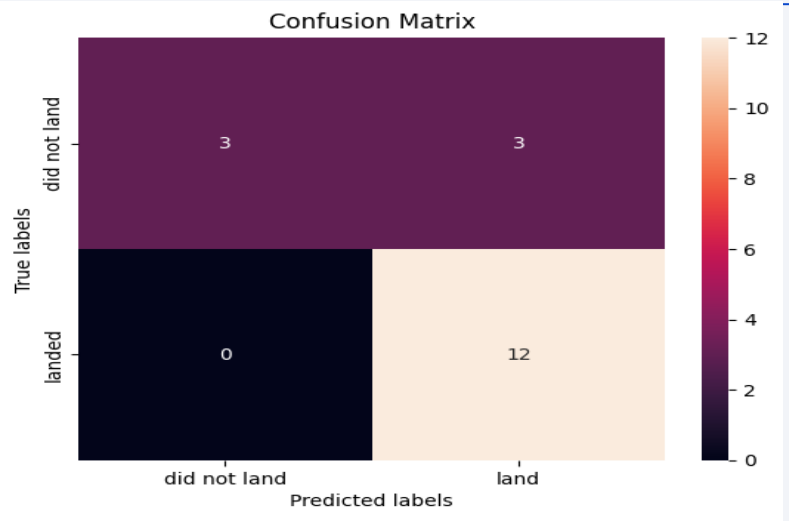
GridSearchCV best score: 0.8190476190476191
Accuracy score on test set: 0.8333333333333334

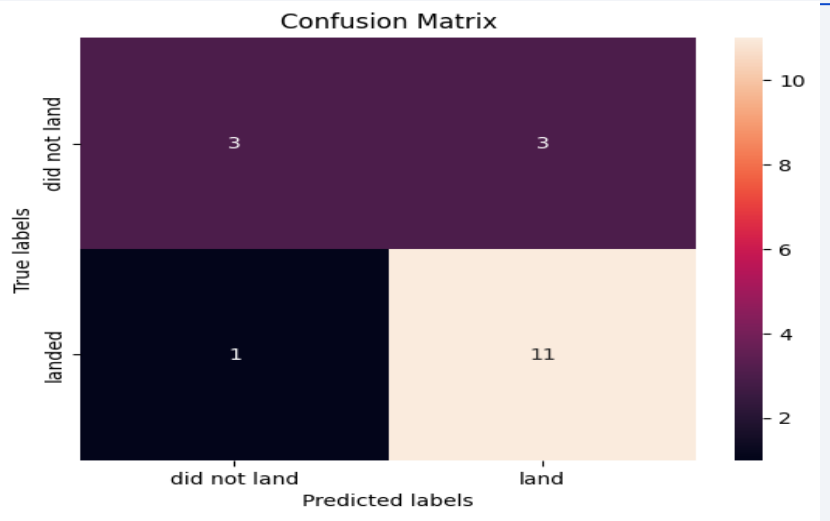
Decision Tree Classifier

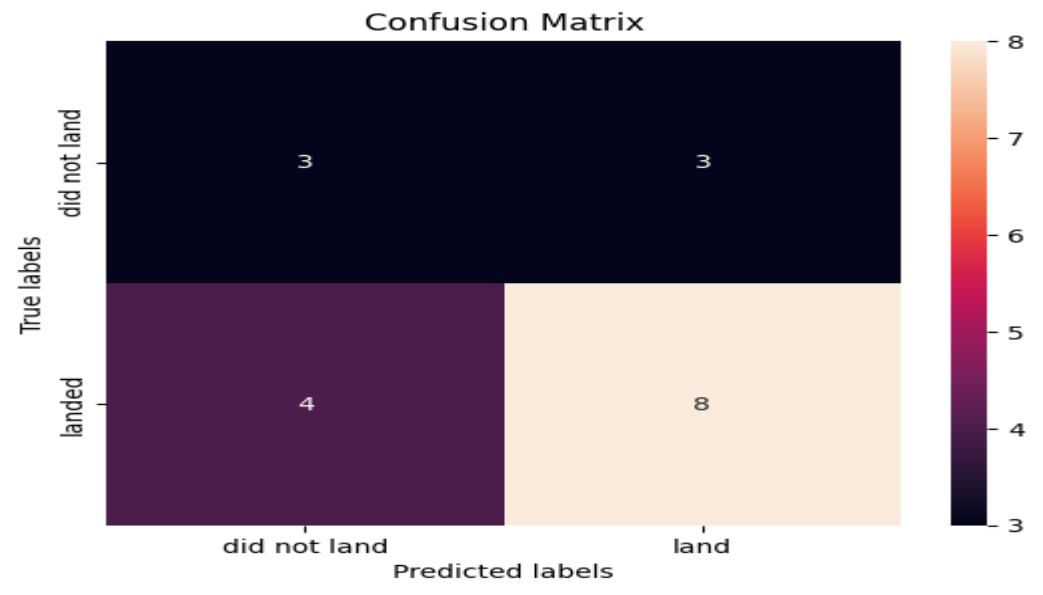
GridSearchCV best score: 0.8732142857142857
Accuracy score on test set: 0.7777777777777778

KNN Classifier
GridSearchCV best score: 0.6642857142857143
Accuracy score on test set: 0.6111111111111112

Confusion Matrix







Conclusions

This project predicts whether the first stage of a Falcon 9 launch will successfully land to estimate launch costs. It analyzes features like payload mass and orbit type that influence mission outcomes. Among four machine learning algorithms tested, the logistic regression model achieved the best performance.

Thank you!