IBM Developer
SKILLS NETWORK

Space Race with Data Science

Essa Mujammami
12/01/2023

Outline

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

Executive Summary

SpaceX made a revolution in space industry by introducing the concept of reusable launcher/booster: Falcon9/Falcon 9 Heavy. The main advantage of this concept is the significant reduction in cost per kg. However, Reliability issues remain compared to classic launchers. In this regards, the success of Falcon9 mission needs to be well-defined in both quantitative and qualitative aspects. The major criteria of success definition is the successful recovery or landing of the booster so that the “low cost” competitive advantage is maintained.

Executive Summary

This requires a definition of the overall system features that includes (Orbit, payload mass , booster versions Launching sites. For this purpose, the system was modeled using four different Machine Learning supervised classification models to predict booster recovery outcome. Based on the above problem statement, the methodology was the following:

- Data collection from SpaceX-API and Webscraping of SpaceX Wikipedia page
- Data Wrangling: Missing Values replaced by mean values
- Exploratory Data Analysis: outcome by (orbit, payload mass and booster versions)

Executive Summary

- Visual Analysis: different types of plots and charts, as well as, map by site
- Interactive Dashboard: Analysis by Site, Payload and booster version
- Predictive Analysis Using Classification: Logistic Regression, SVM, Decision Tree, KNN

Hence, the findings were promising indicating high success rate for higher orbits and payload mass. In addition, versions FT, B4, B5 have higher success rate compared to booster versions v1.0, v1.1. The supervised classification models developed in this project predict the booster recovery outcome with an accuracy ranges between 83.3% to 94%.

Introduction

- Background and context
  - Based on the reusability of the first stage (ability to recover part of rocket (Stage 1)), SpaceX provides low-cost Falcon 9 rocket launchers (average of $62m vs. $165m of competitor)

- Question to be answered
  - Is it possible to accurately predict the success of the first stage will land? The, the cost of a launch can be determined accordingly.

Methodology

Methodology

- Data collection
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Methodology: Data Collection

- SpaceX REST API

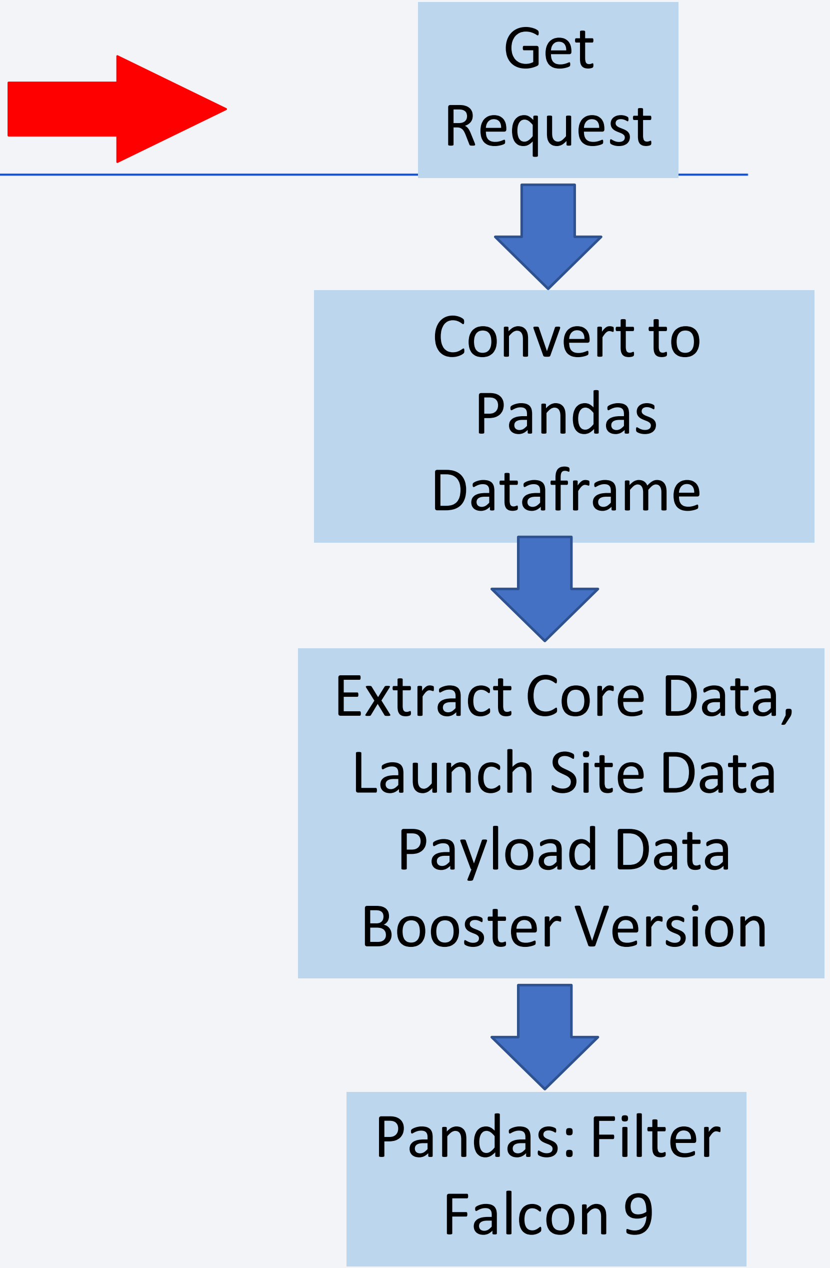
- Webscraping of SpaceX Wikipedia Page

Data Collection SpaceX API

(RESTful API)

SpaceX REST API:

- RESTful Interface
- Get Core Data
- Get Booster Version
- Get Launch Site Data
- Get Payload Data



Data Collection Scraping

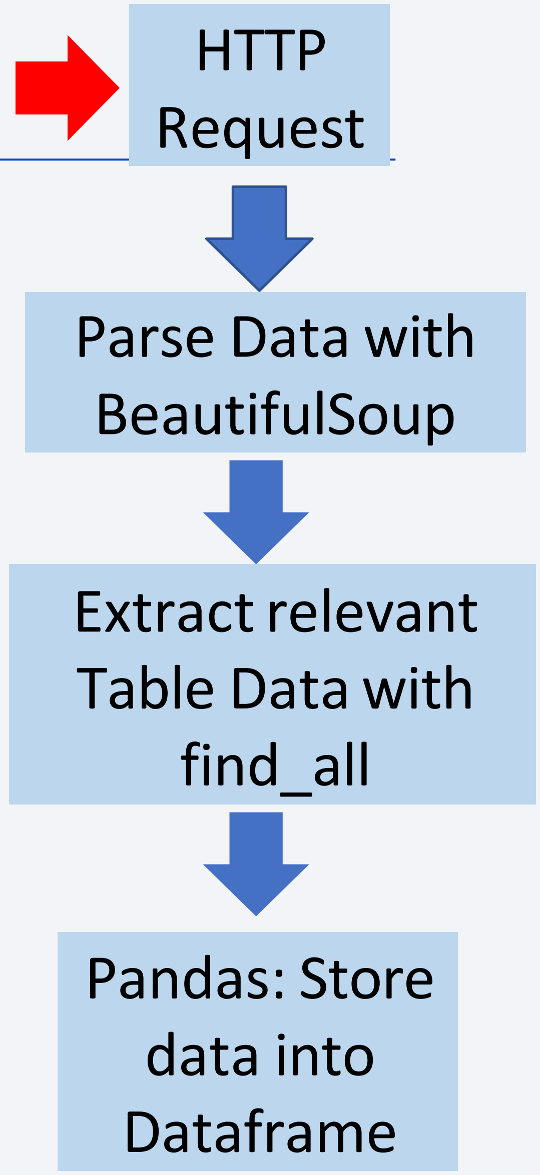
(Webscraping)

Webscraping (SpaceX Wikipedia Page):

- HTML Requests (HTTP-Get)
- Package for Webscraping:

  - Python

  - BeautifulSoup
- Extract Column Names from HTML table header



Methodology : Data Wrangling

- Dealing with missing values:

  - Retake them if it is possible,

  - Replace them with the mean value of the column

or - discard them.

Methodology: EDA with Data Visualization

Charts:

- Payload mass vs. Flight number vs. Success rate
- Launch site vs. Flight number vs. Success rate
- Launch site vs. Payload mass vs. Success rate
- Orbit type vs. Success rate
- Orbit type vs. Flight number vs. Success rate
- Orbit type vs. Payload mass vs. Success rate
- Success rate vs. Year

Methodology: EDA with SQL

- SQL queries
  - Extract a list of all launch sites
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved

EDA with SQL (Data Analysis with SQL)

- SQL queries
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
  - List the total number of successful and failure mission outcomes.
  - List the names of the booster_versions which carried the maximum payload mass.
  - List the failed landing_outcomes in drone ship and launch site names for in year 2015.
  - Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20

Methodology : Build an Interactive Map with Folium

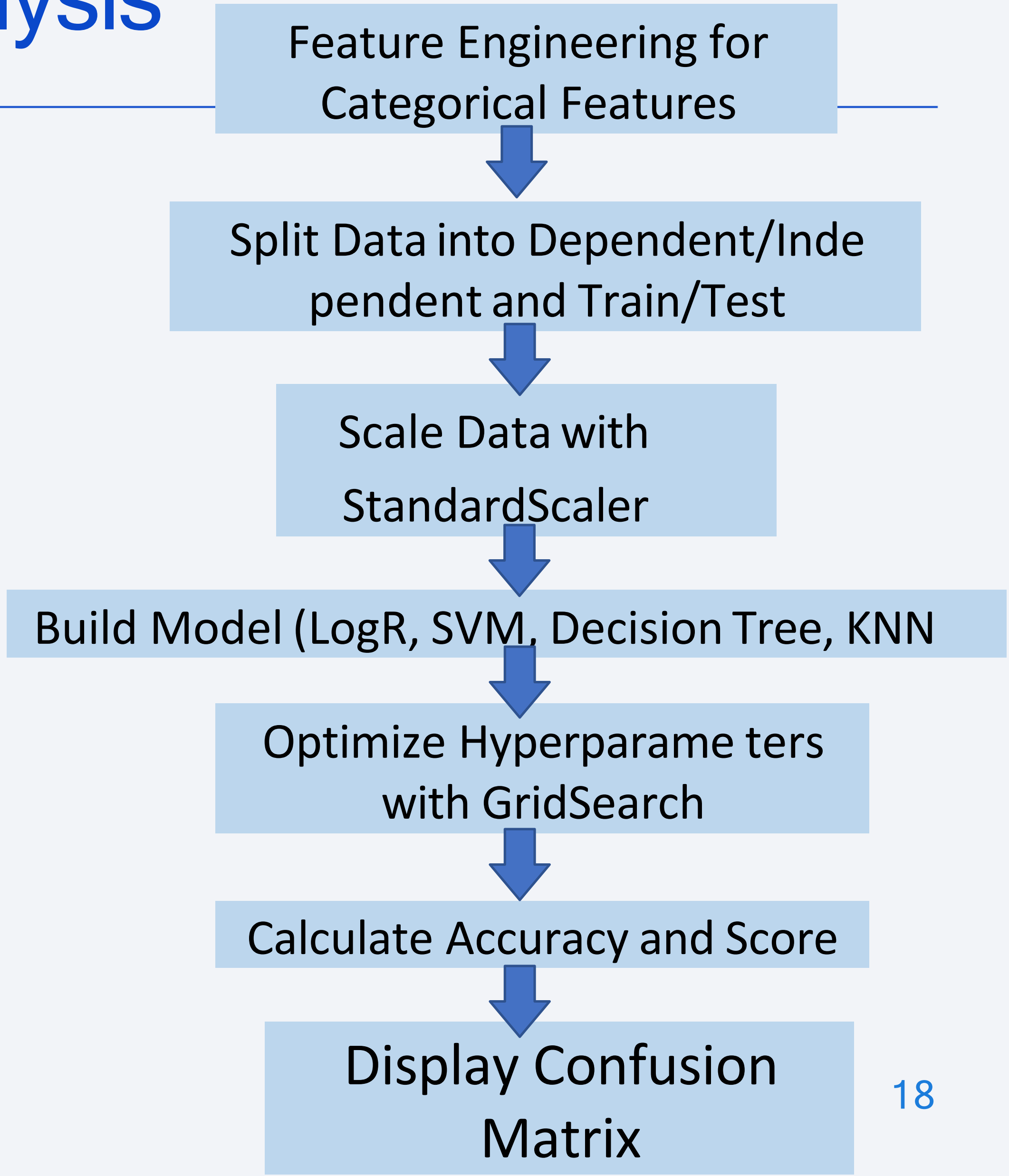
- Map Objects
  - Edged Circles (radius 1000m): Space launch sites
  - Markers: for labeling all objects
  - MarkerCluster: for creating a bunch of markers around space launch sites to indicate success (green) or failure (red) of the landing of the rocket's first stage
  - Lines: Measure the distince between the launch site and the next coast or next city

Build an Interactive Dashboard with Plotly Dash

- Input :
  - Dropdown list for the launch site (with option to select all)
  - RangeSlider for selecting the payload mass
- Output :
  - PieChart: for showing the success rate of (each launch site / all sites are selected) showing the percentage of successful landing outcomess
  - Scatterplot: Show success/failure by payload and booster version

Methodology: Predictive Analysis

- Preprocessing
- Model Building for each Method
- Optimization
- Evaluation

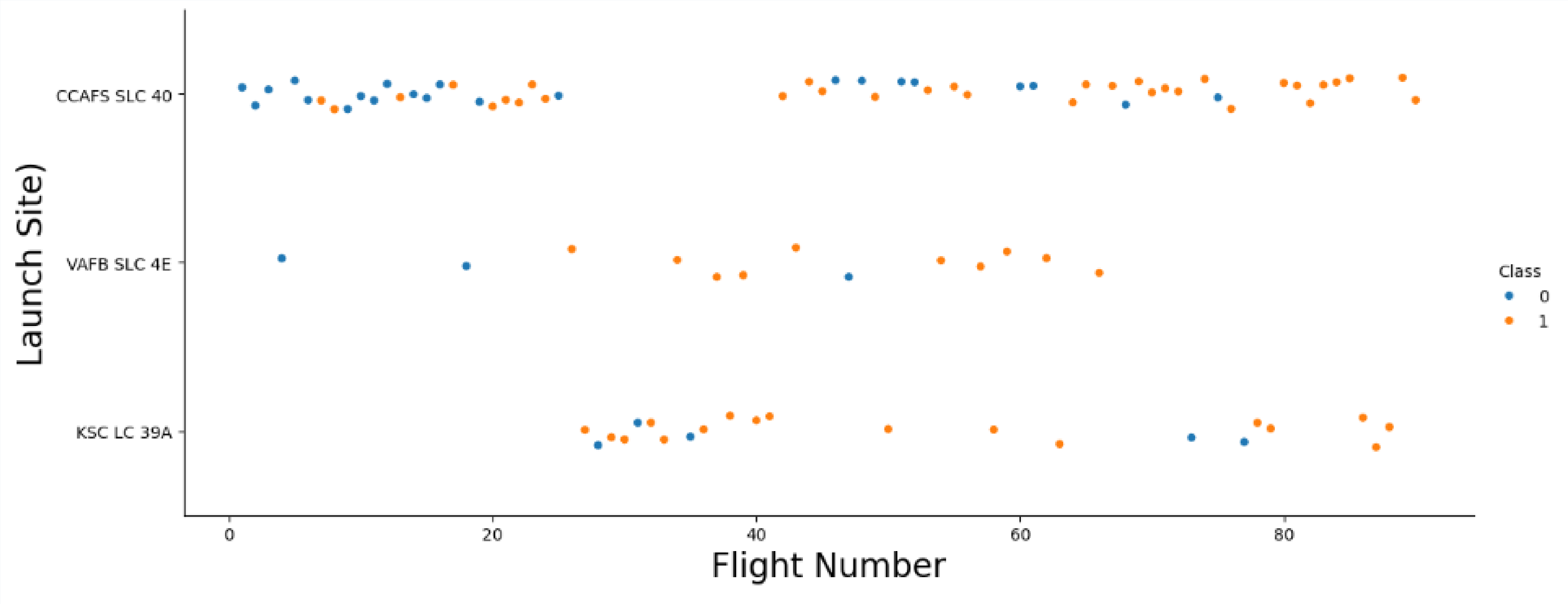


Results

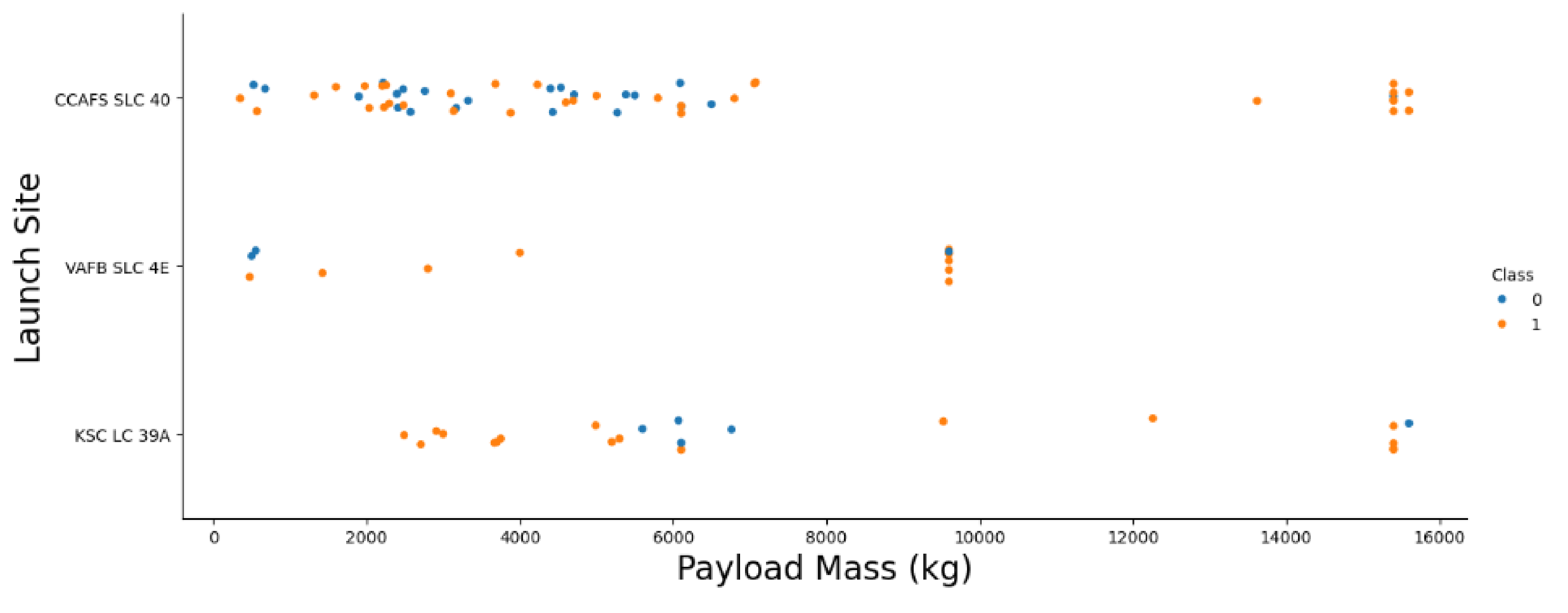
Results

Launch success rate increases over time and higher success rate for higher orbits, as well as, higher payload mass. Low success rate for booster versions v1.0, v1.1, while high success rate for FT, B4, B5. The Best prediction results with k nearest neighbors (KNN) and Support Vector Machine (SVM) predict the booster recovery outcome with an accuracy close to 94%.

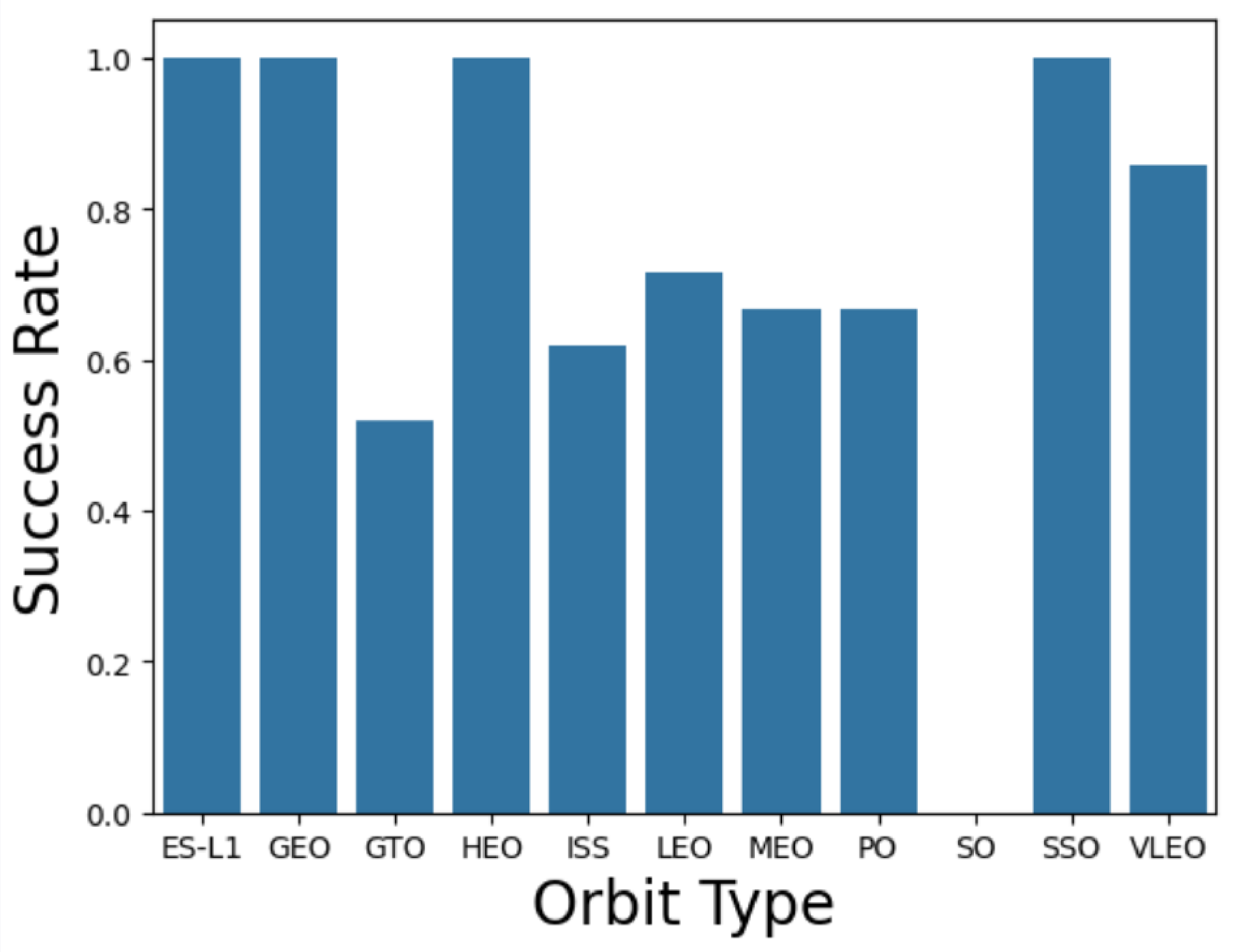
Flight Number vs. Launch Site



Payload vs. Launch Site

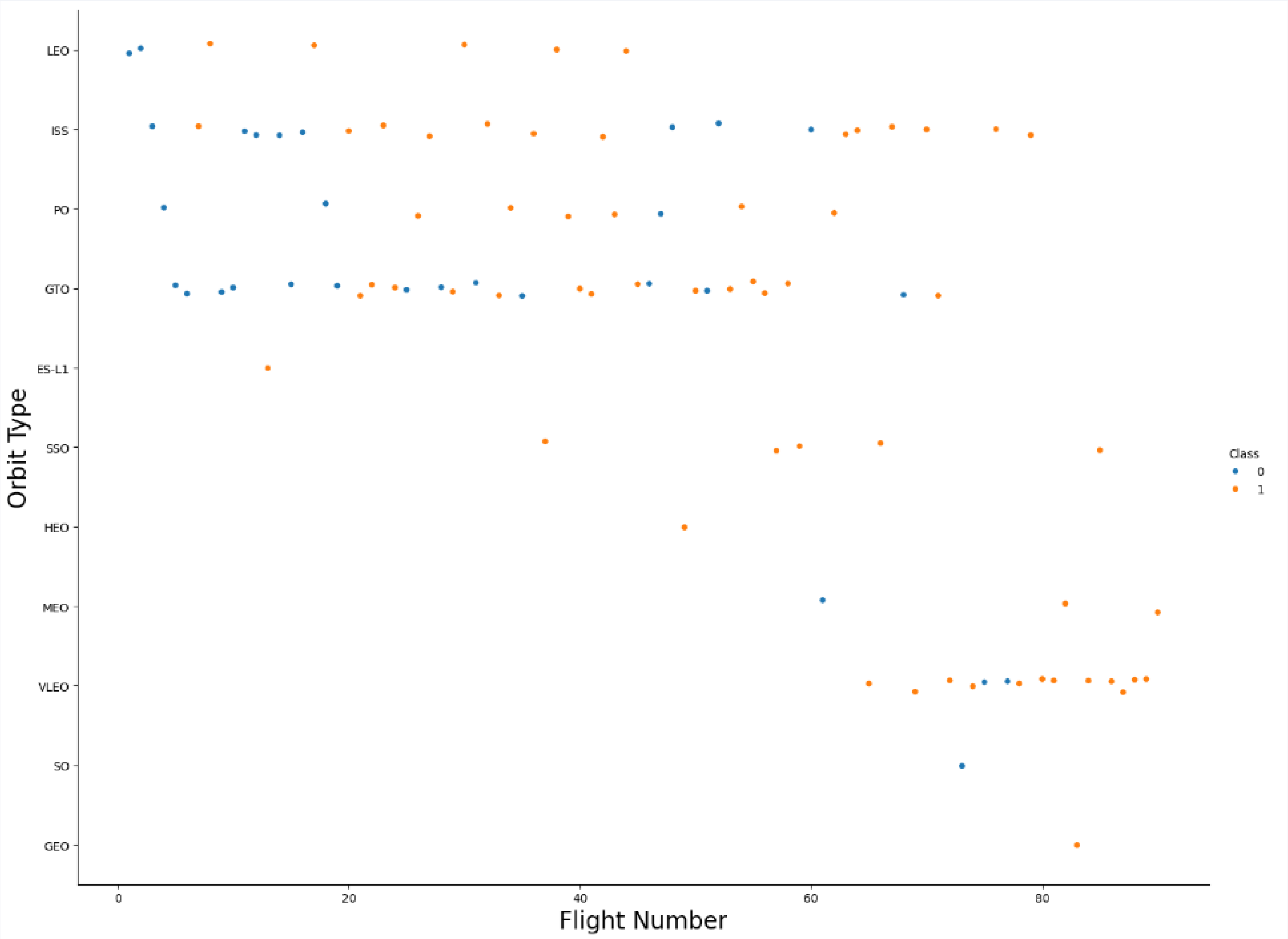


Success Rate vs. Orbit Type

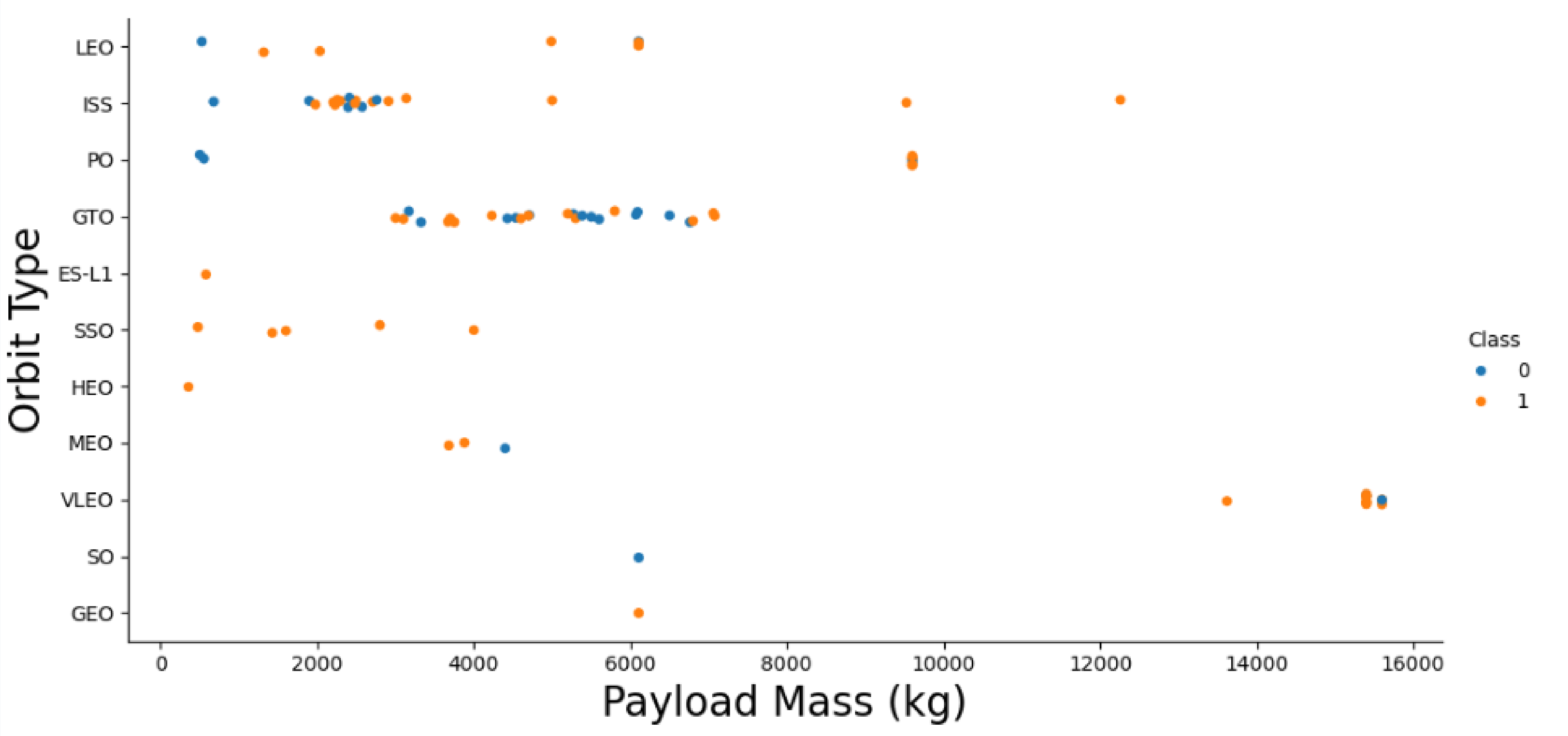


Flight Number vs. Orbit Type

- Success rate has increased over time for all orbit types.



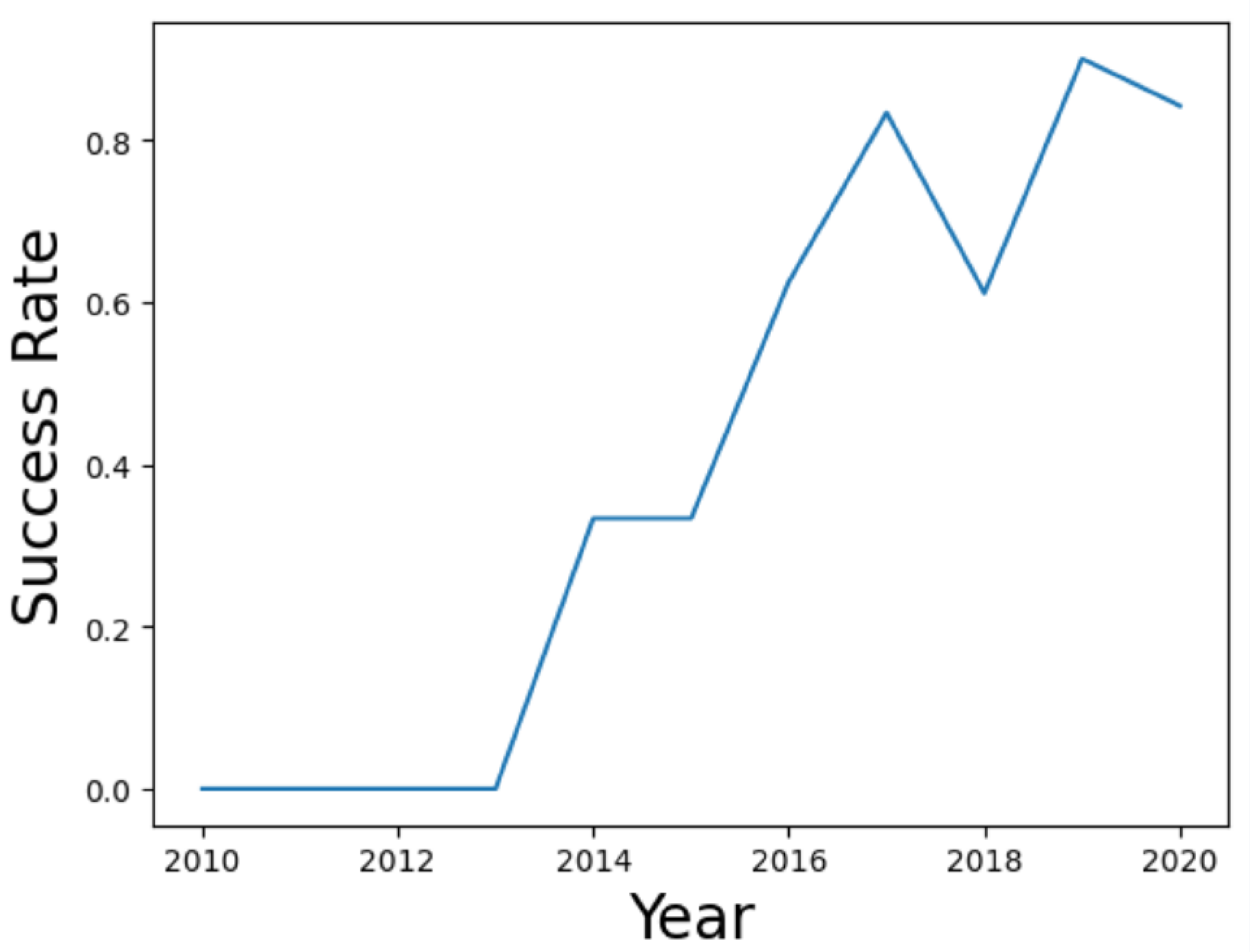
Payload vs. Orbit Type



Launch Success Yearly Trend

Success generally increases over time since 2013 with a slight dip in 2018

Recently, success rate is around 80%



All Launch Site Names

	Launch Site
0	CCAFS LC-40
1	CCAFS SLC-40
2	KSC LC-39A
3	VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Some sample records for starts 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 carried by boosters from NASA

TOTAL_PAYLOAD_MASS__KG
45596

Average Payload Mass Carried by booster version F9 v1.1

AVERAGE_PAYLOAD_MASS__KG
2928.4

Date of First Successful Ground Landing Date

min(DATE)
2015-12-22

Successful Drone Ship Landing with Payload Between 4000 and 6000 KG

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	count(*)
Failure	1
Success	100

Boosters Maximum Carried 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

Landing_Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
Failure (drone ship)	F9 FT B1020	CCAFS LC-40
Failure (drone ship)	F9 FT B1024	CCAFS LC-40

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

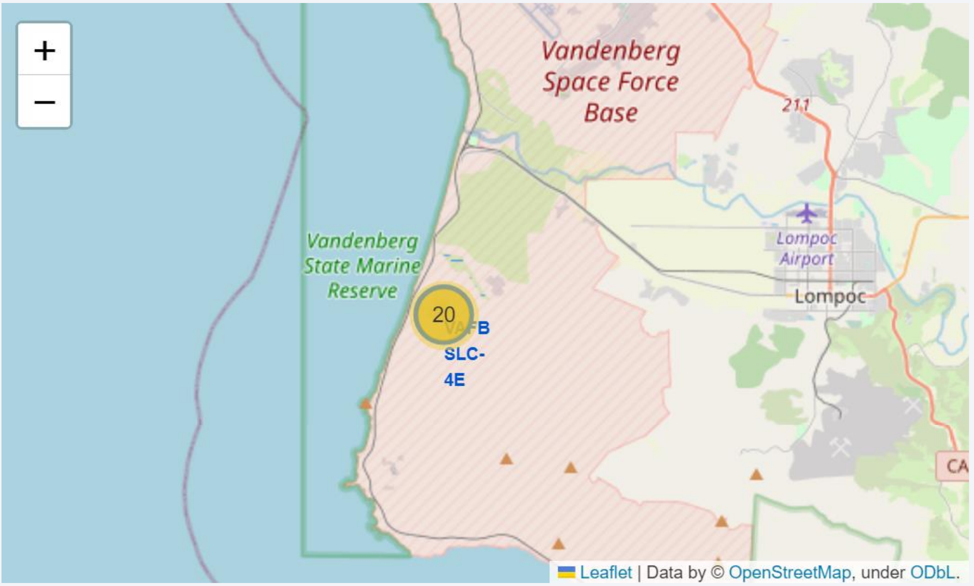
Landing_Outcome	count(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1

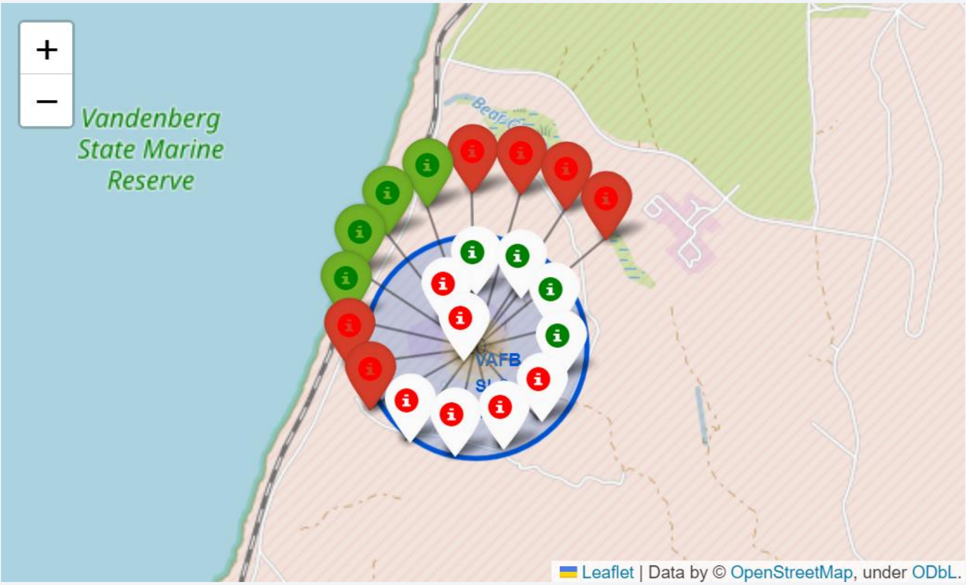
Launch Sites Proximities Analysis

Folium Map: Launch Sites

The map shows all launch sites relative US map

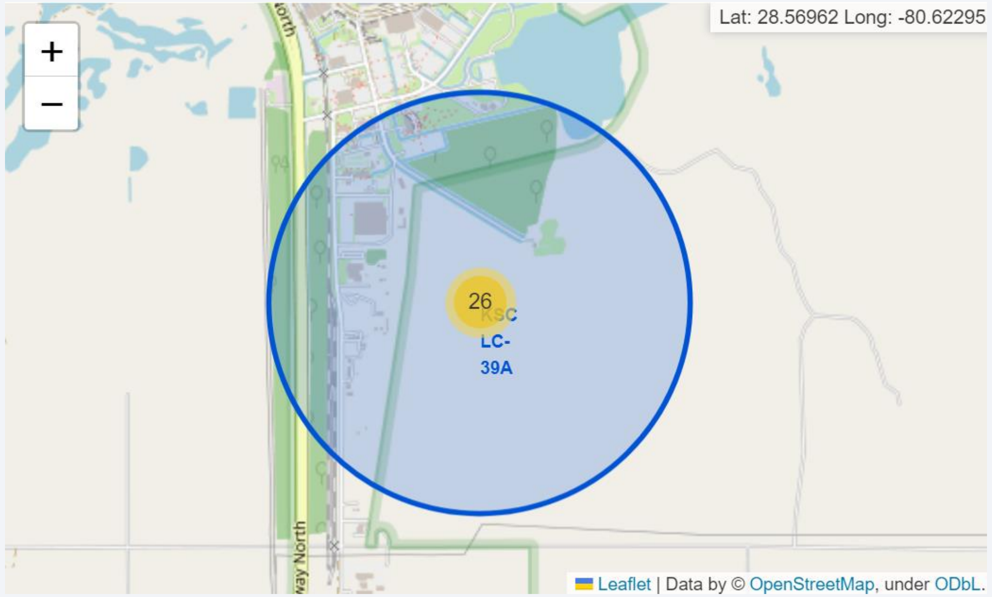
Folium Map: Proximity Vandenburg AFB

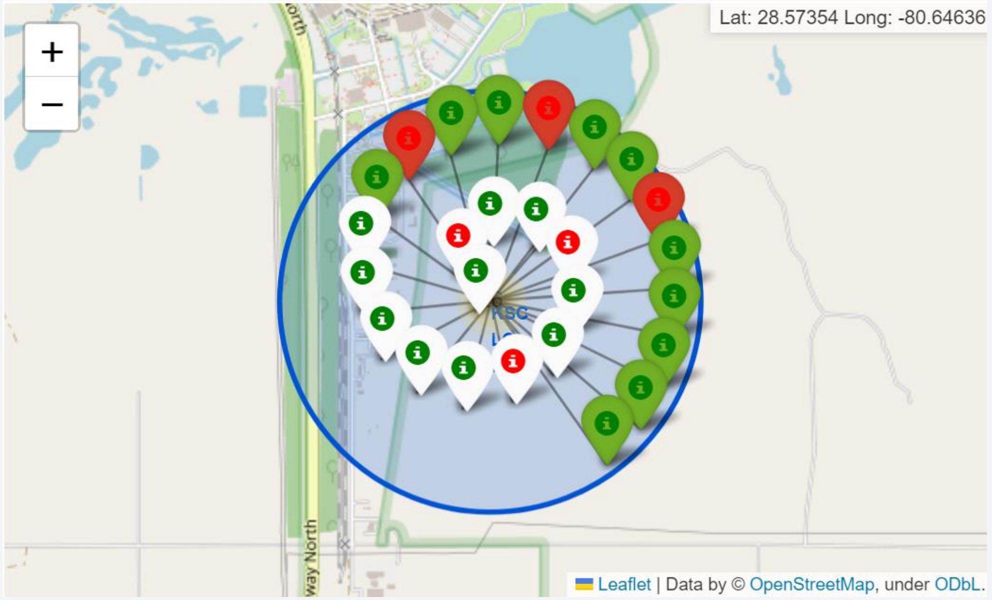




Clusters (map in the right) on Folium map can be clicked on to display each successful landing (green icon) and failed landing (red icon) (map in the left). In this example VAFB SLC-4E shows 4 successful landings and 6 failed landings

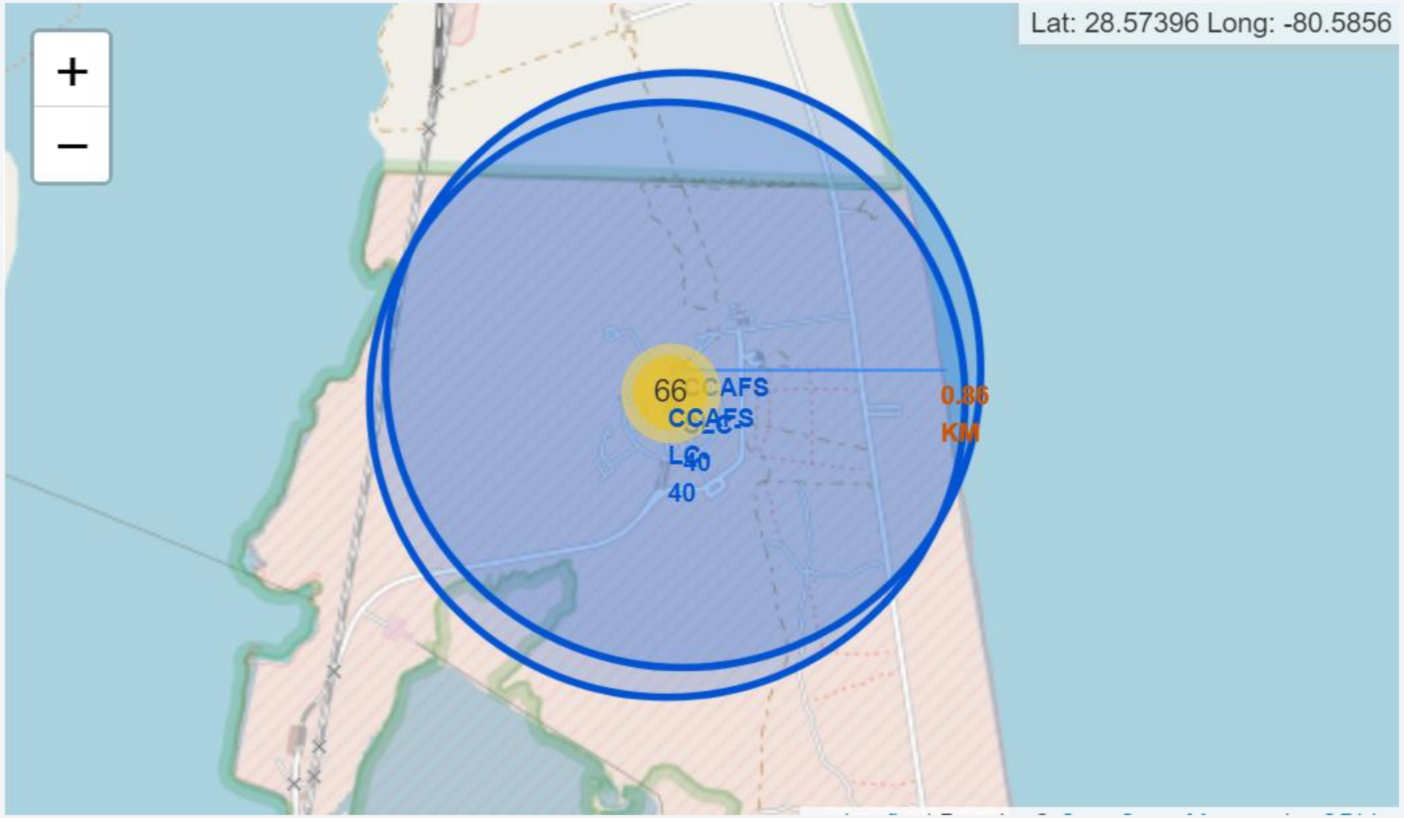
Folium Map: Proximity Kennedy SC / Cape Canaveral





Cluster in the right map denotes KSC LC-39A launch site while each successful landing (green icon) and failed landing (red icon) is shown in the left map.

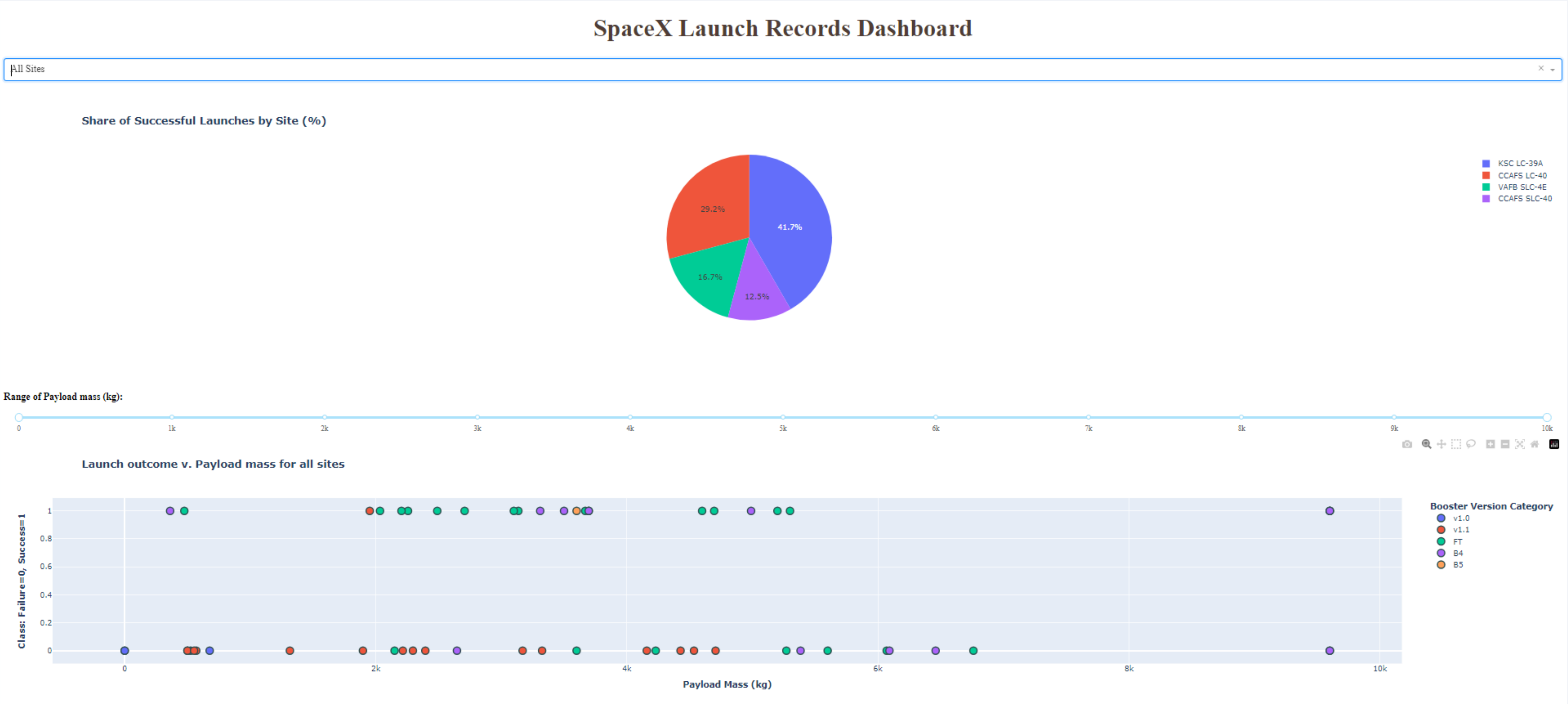
Folium Map: Distance Between The Launch Site The Coastline



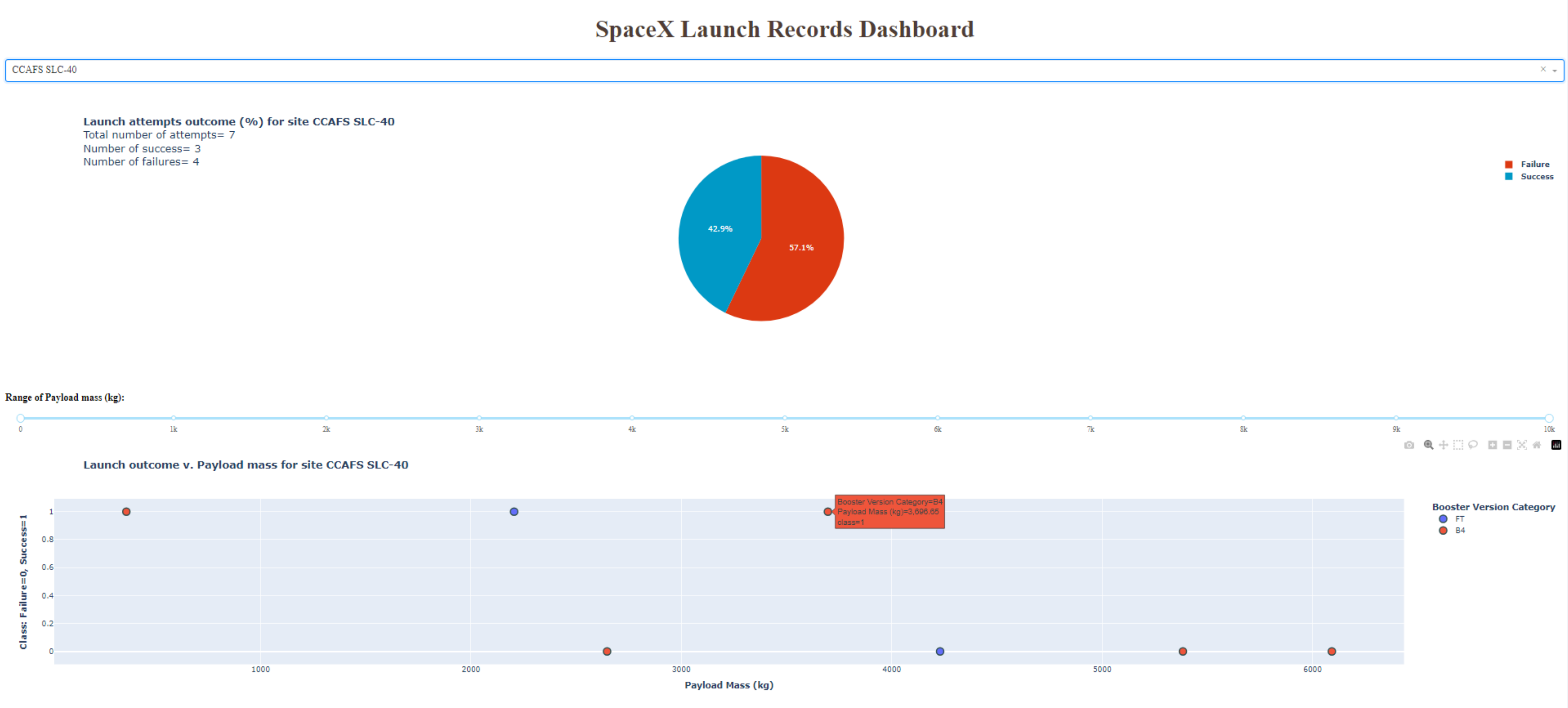
The distance between the coastline point and the launch site a point on the closest coastline.

Build a Dashboard with Plotly Dash

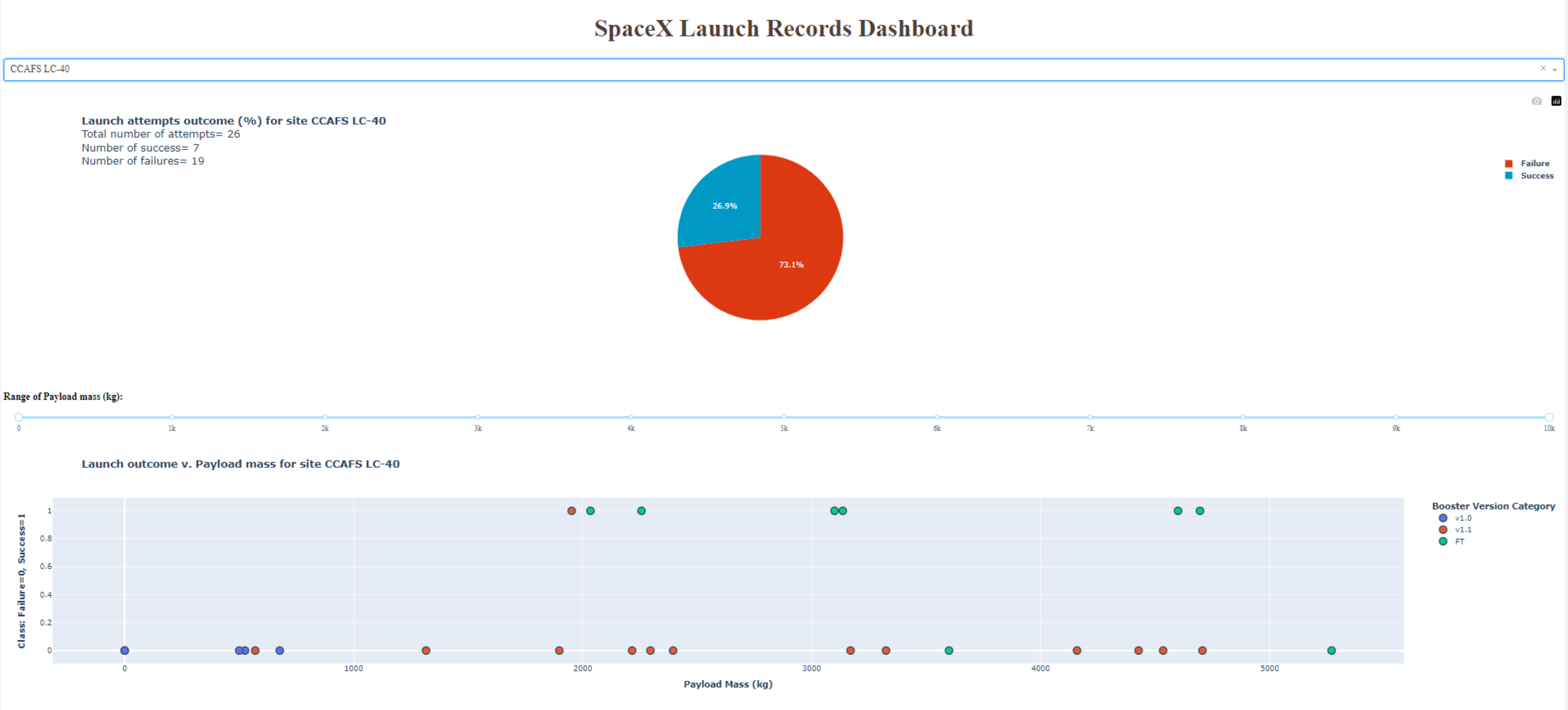
Dashboard: Launch Success Count For All Sites



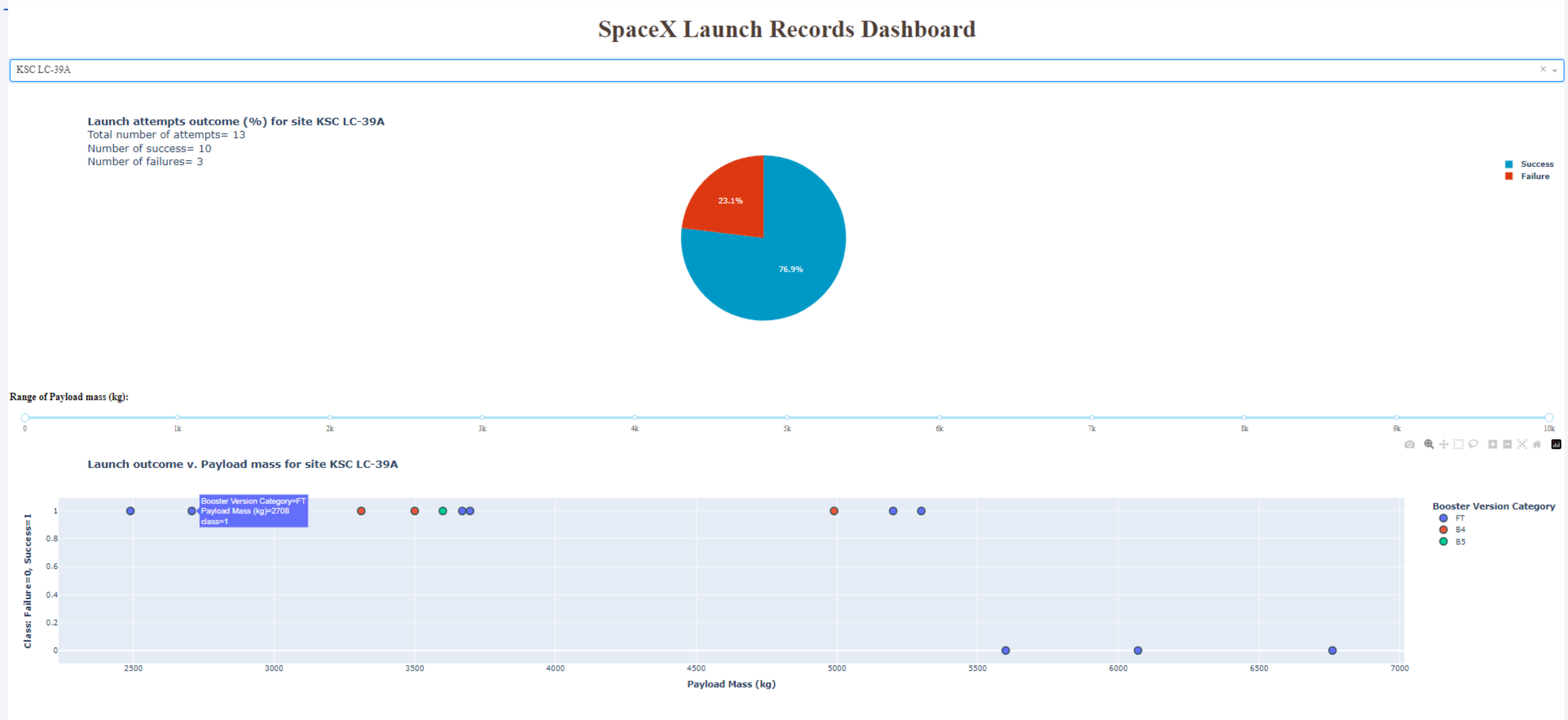
Dashboard: Success Rate Cape Canaveral Launch Center



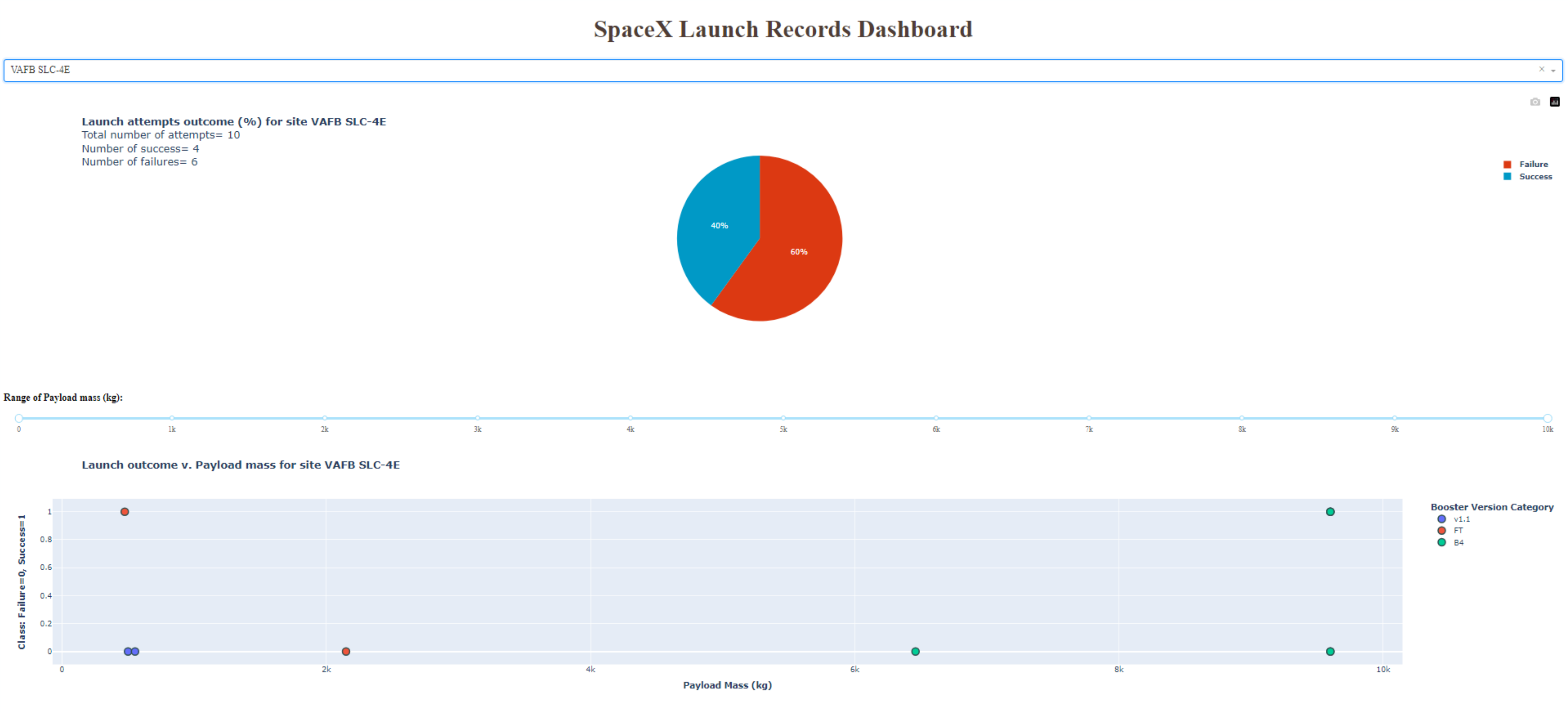
Dashboard: Success Rate Cape Canaveral Launch Center



Dashboard: Success Rate Kennedy Space Center



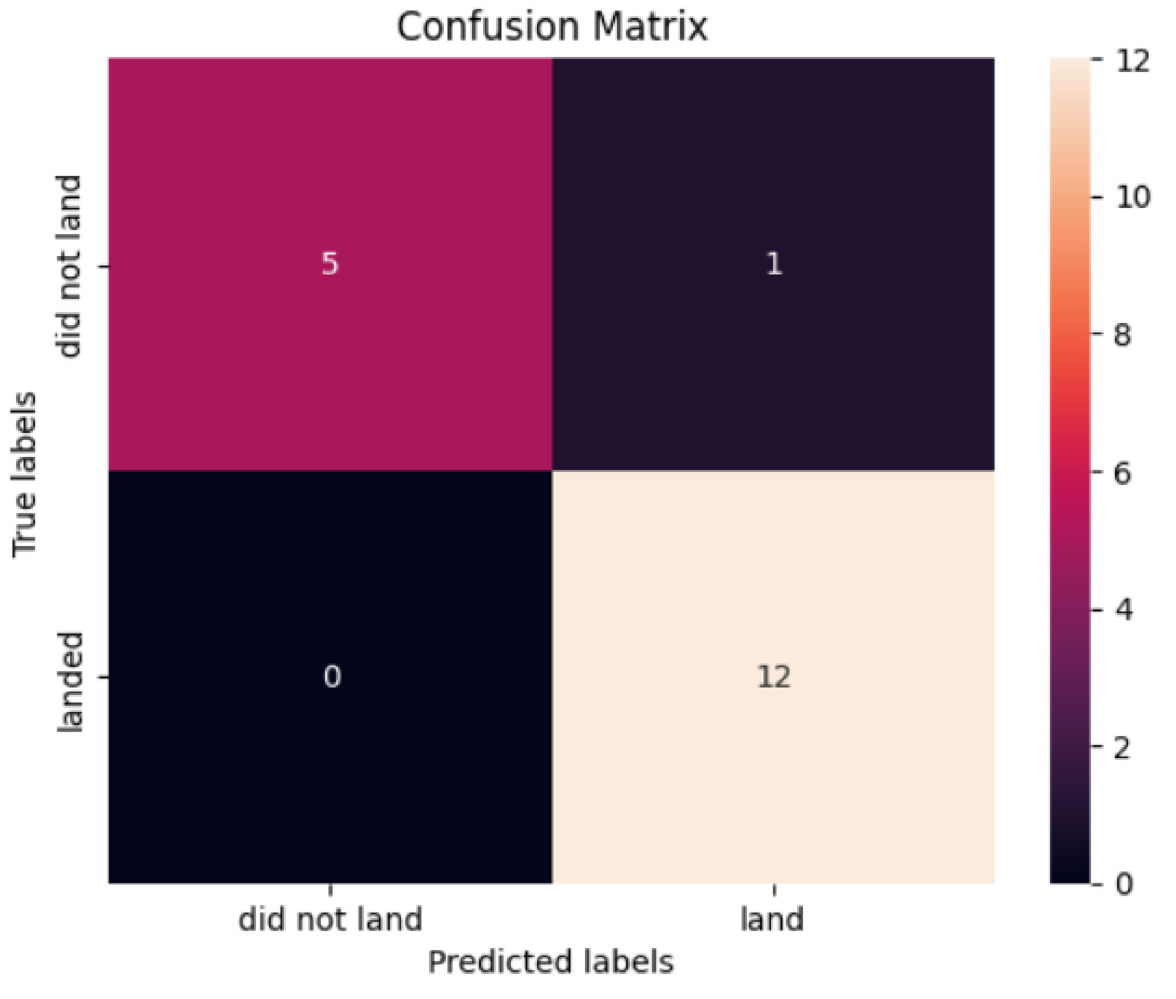
Dashboard: Success Rate Vandenburg Air Force Base



Predictive Analysis

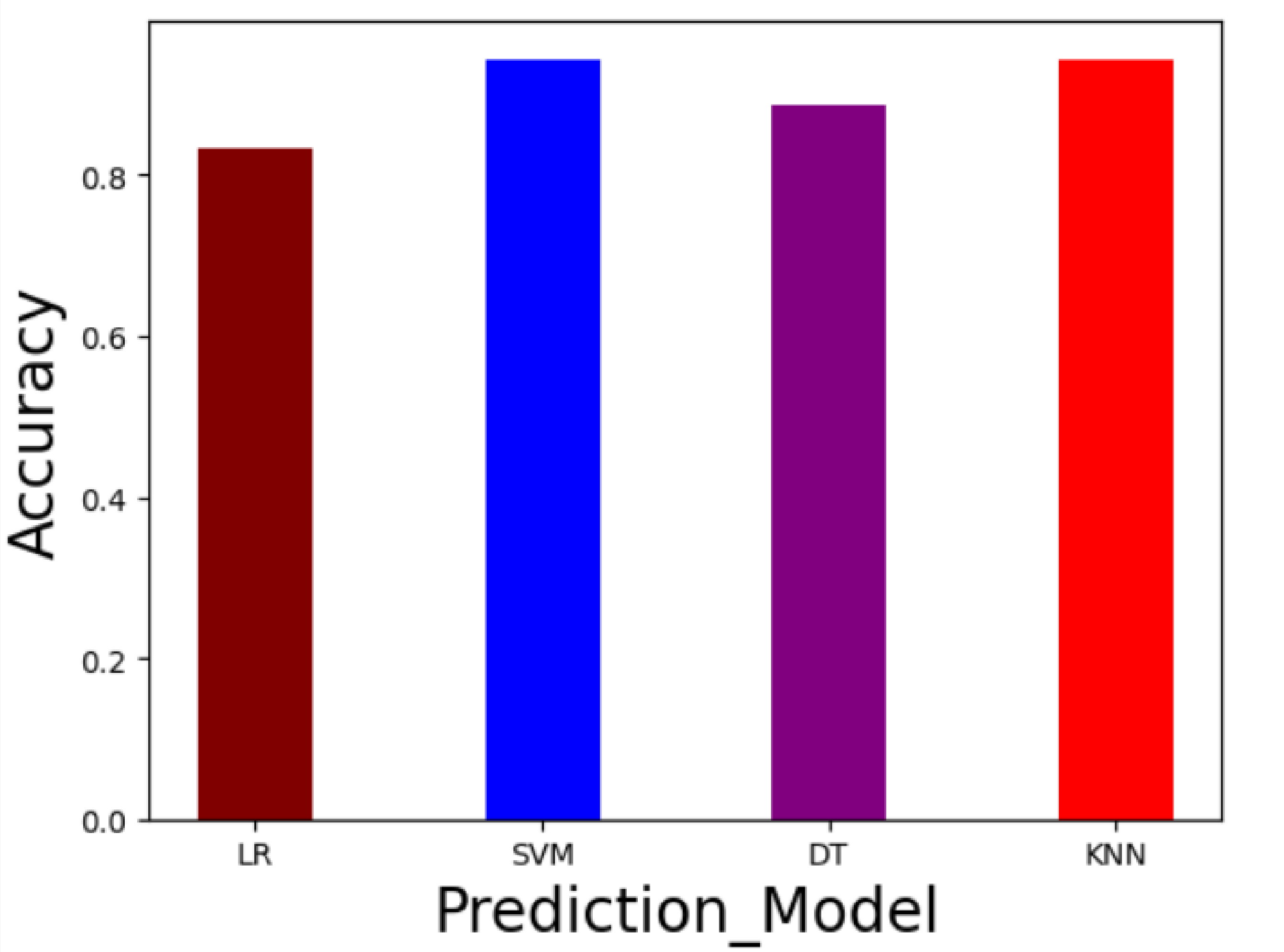
Confusion Matrix For Support Vector Machine

- True Positives: 12
- True Negatives: 5
- False Positives: 1
- False Negatives: 0



Methods Performance Evaluation in Terms of Accuracy

The accuracy based on test data is 94.44% for both of SVM and KNN, while DT and LR come after with 88.88% and 83.33% respectively.



Conclusions

- Support Vector Machine provides a good result for predicting the landing outcome
- None of the models had false negatives
- All models had at least one false positive
- The accuracy based on test data is 94.44% for SVM and KNN. However, SVM has better processing time compared to KNN. For that reason SVM is recommended for system modeling and outcome prediction

Appendix

- All codes, Jupyter notebooks, and the presentation for this project can be inspected and download from the following link:

https://github.com/DayofJudgement/Final-Applied-Data-Science-Capstone-Project

Thank You ...