IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

EL ADNANI ABDELILAH
16/11/2025

Outline

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

Executive Summary

- Summary of methodology
  - Data collection – collected data using SpaceX API and web-scraping on Wikipedia
  - Data wrangling – transformed collected data from JSON object and HTML tables into Pandas dataframe
  - Exploratory data analysis using Matplotlib and SQL
  - Interactive visual analytics using Folium and Plotly Dash
  - Predictive analysis using logistic regression, support vector machine (SVM), decision tree, and k nearest neighbor (knn) classification models
- Summary of all results
  - Exploratory data analysis (EDA) results – determine what attributes are correlated with successful launch outcomes
  - Interactive analytics results – use interactive maps to uncover patterns and identify the optimal launch site
  - Predictive analysis results – employ machine learning models to predict the likelihood of successful launch outcomes

Introduction

The commercial space age is here.

Companies like Rocket Lab, Virgin Galactic, and Blue Origin have made great strides in space travel. Space X has become a household name and the dominant player in modern rocket launches, having achieved significant milestones with a near maniac at its helm.

At Space Y, we know we can do more and do better.

By analyzing a series of Space X Falcon 9 first stage landings and using machine learning to predict the likelihood of future successful landings, we've established an industry standard that we aim to meet and exceed. This careful analysis will enable us to

- Determine the cost of each launch and offer competitive pricing
- Accurately predict the outcomes of our own first stage landings
- Develop a competitive strategy to challenge the market share currently dominated by Space X
- Increase our own operational efficiency by identifying the key factors that contribute to successful landing outcomes

Section 1

Methodology

Methodology

Data Collection	Data Wrangling	Exploratory data analysis (EDA)	Interactive visual analytics	Predictive analysis
Used the SpaceX API and web-scraped relevant Wikipedia pages	Filtered data to include only Falcon 9 landings Handled missing values Converted outcomes to binary class	Used Matplotlib and SQL to gain first insights into SpaceX data	Used Folium and Plotly Dash to visualize data and recognize patterns	Trained and tested various machine learning models and fine-tuned with hyperparameters

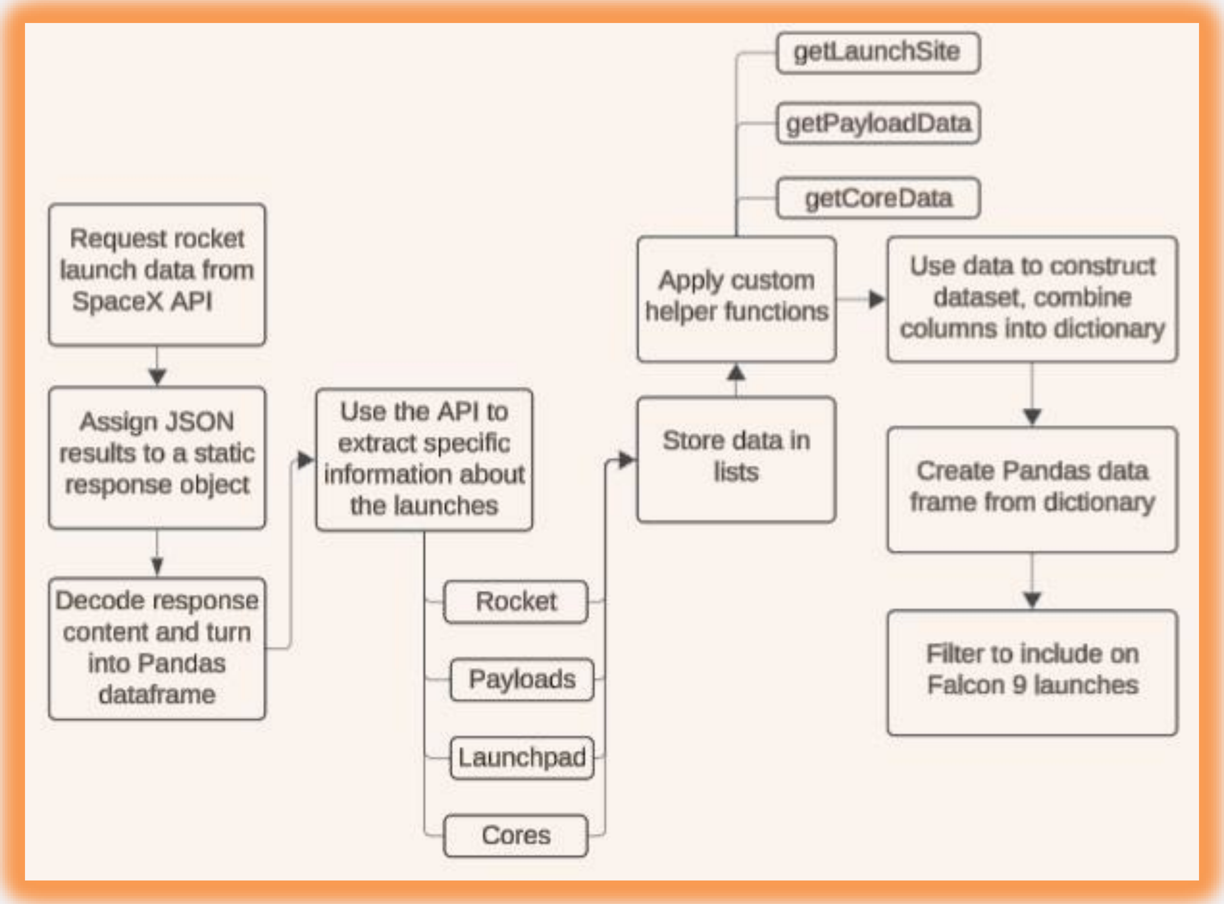
Data Collection

Gathered SpaceX launch data from the SpaceX REST API, including information about

- Rocket used
- Payload delivered
- Launch specifications
- Landing specs
- Landing outcome

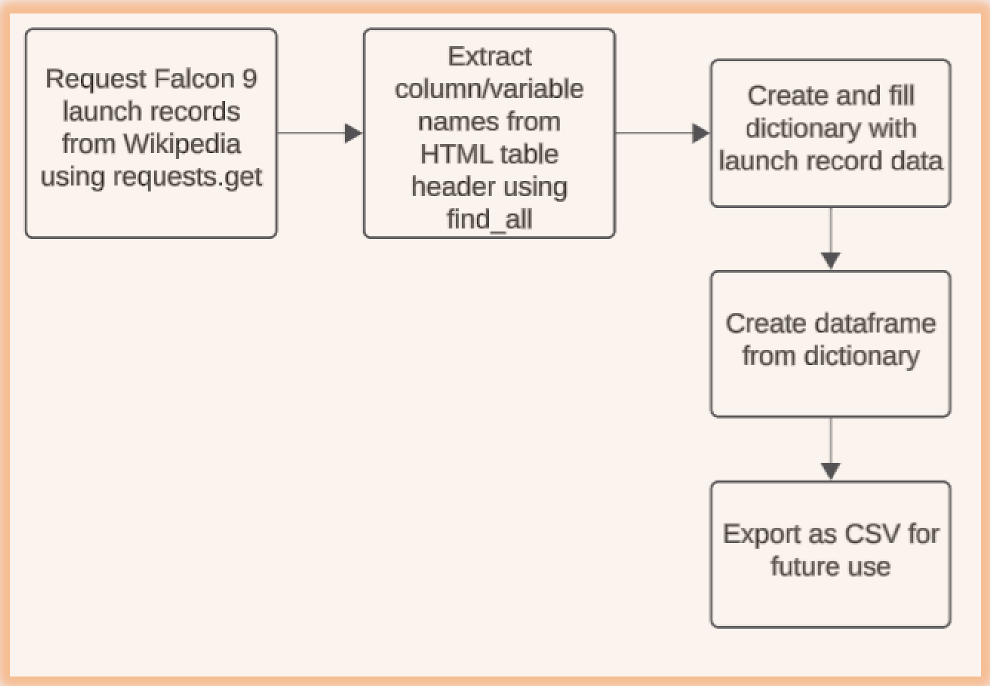
Data Collection – SpaceX API

- Made GET request to SpaceX API
- Defined a series of helper functions to extract relevant information
- Filtered results to only include Falcon 9 launches



Data Collection: Web Scraping

- Web scraped Falcon 9 launch records from Wikipedia using BeautifulSoup
- Extracted and used column names as keys to create dictionary of launch records
- Converted dictionary to dataframe

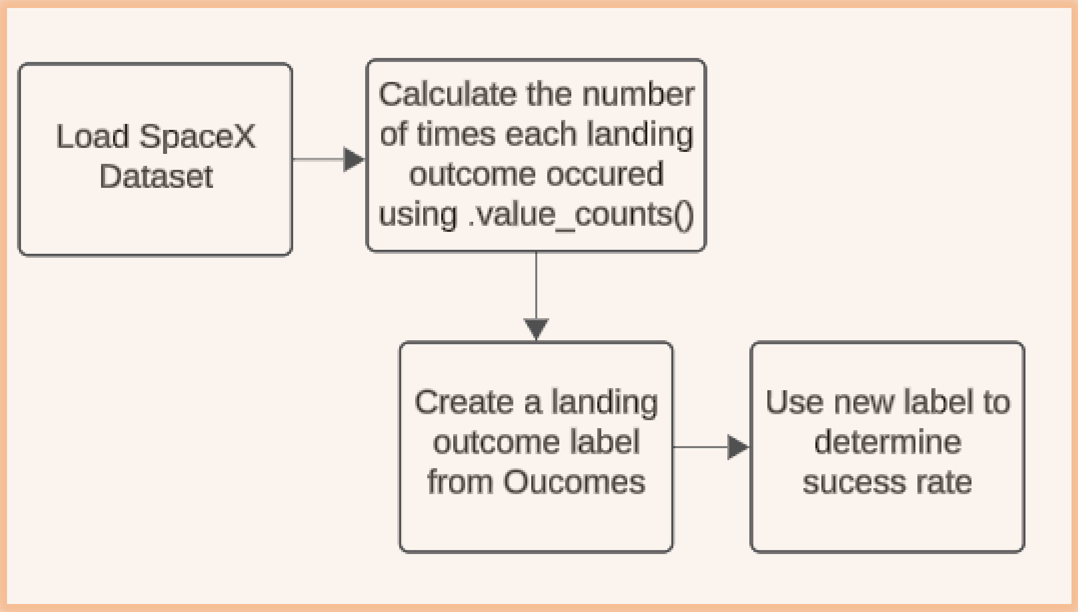


Data Wrangling

- Performed exploratory data analysis (EDA) to identify patterns in data and determine labels for training supervised models
  - Converted landing outcomes to Class *y*, wherein

    0 = unsuccessful landing and

    1 = successful landing



Used Pandas and Matplotlib to produce visualizations

- Scatter plot to visualize relationship between Flight Number and Launch Site
  - There were more unsuccessful landings in earlier flights among all launch sites and more successful landings in later flights at each launch site
- Scatter plot to visualize relationship between Payload Mass and Launch Site
  - No rockets launched for payloads >10000 kg at the VAFB SLC launch site
- Bar chart to visualize relationship between orbit type and landing outcome
  - ES-L1, GEO, HEO, and SSO demonstrate the highest success rates among all orbit types
- Scatter plot to visualize relationship between Flight Number and Orbit
  - In low earth orbits (LEO), success seems to be relational to the number of flights
- Scatter plot to visualize relationship between Payload Mass and Orbit
  - Heavy payloads saw higher landing success in Polar, LEO, and ISS orbits.
- Line Chart to visualize the yearly trend of launch success
  - Success rates improved from 2013-2020

Used SQL queries to learn more about rocket performance and outcomes

- %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE
- %sql SELECT * FROM SPACEXTABLE WHERE "LAUNCH_SITE" LIKE "CCA%" LIMIT 5
- %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE CUSTOMER="NASA (CRS)“
- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE BOOSTER_VERSION="F9 v1.1“
- %sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE LANDING_OUTCOME="Success (ground pad)“
- %sql SELECT PAYLOAD, BOOSTER_VERSION FROM SPACEXTABLE WHERE LANDING_OUTCOME=("Success (drone ship)") AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
- %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number FROM SPACEXTABLE GROUP BY MISSION_OUTCOME
- %sql SELECT BOOSTER_VERSION FROM SPACEXTABLE WHERE PAYLOAD_MASS KG_ = (SELECT MAX(PAYLOAD_MASS KG_) FROM SPACEXTABLE)
- %sql SELECT substr(Date,6,2) as month, substr(Date,0,5) as year, DATE, BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME FROM SPACEXTABLE WHERE LANDING_OUTCOME="Failure (drone ship)" and substr(Date,0,5)="2015“
- %sql SELECT LANDING_OUTCOME, COUNT(*) AS OUTCOME_COUNT FROM SPACEXTBL WHERE DATE between '2010-06-04' and '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY OUTCOME_COUNT DESC

Interactive Visualizations with Folium Map

- Created a marker and circle for to represent each launch site on the site map
- Enhanced our understanding of the map by adding launch outcomes for each site
    - Used marker clusters to organize multiple markers at one coordinate and color-labeled markers within the clusters to visualize launch outcome success rates
        - Successful launch outcomes were represented by green markers
        - Unsuccessful launch outcomes were represented by red markers
- Calculated proximity of launch site to highway, railway, and city and used lines to visualize these distances

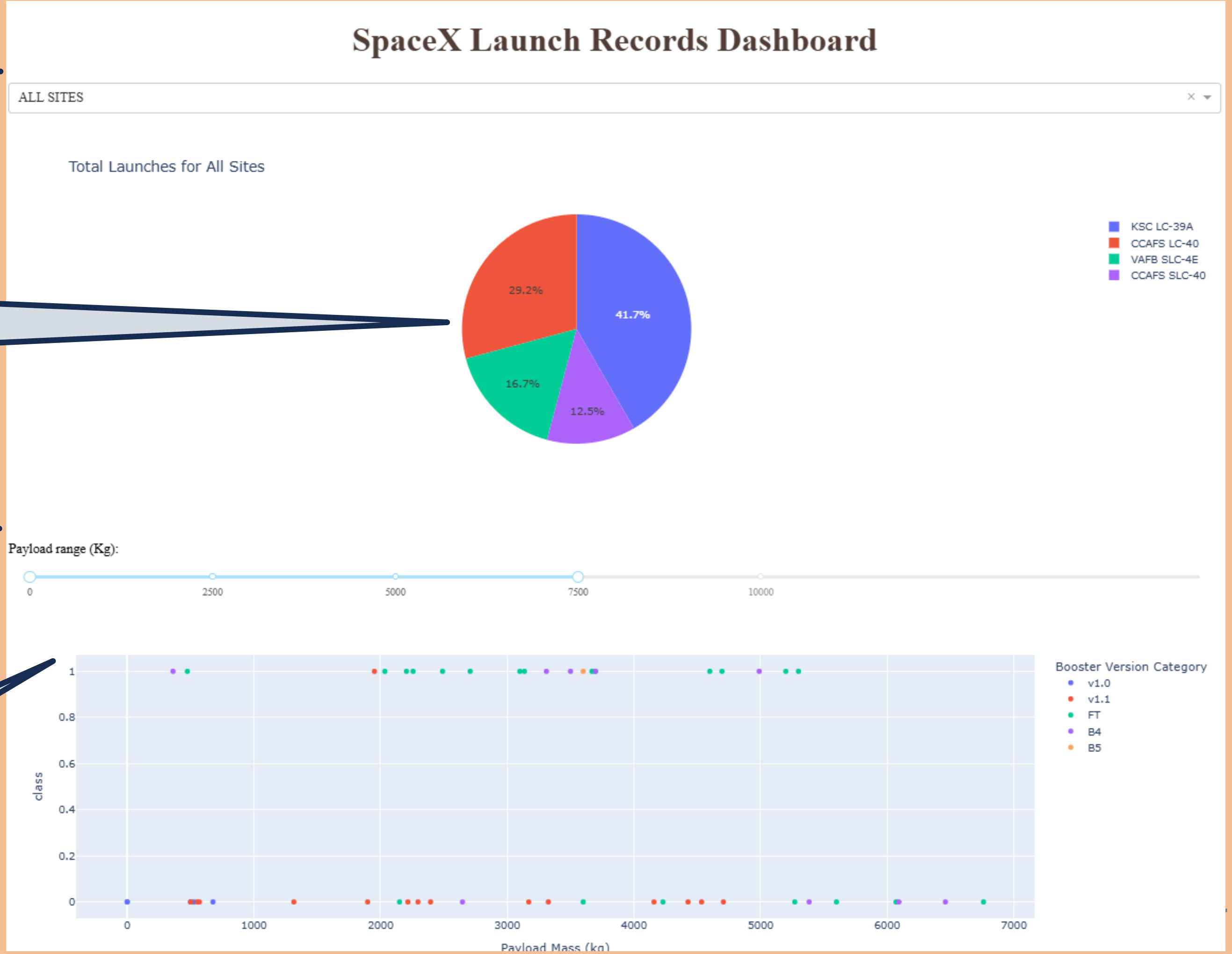
SpaceX Launch Records Dashboard with Plotly Dash

Drop down menu – allows user to select and view data for all launch sites or one specific launch site

Pie chart – represents the percentage of successful launch outcomes (1 - purple) and unsuccessful launch outcomes (0 – red)

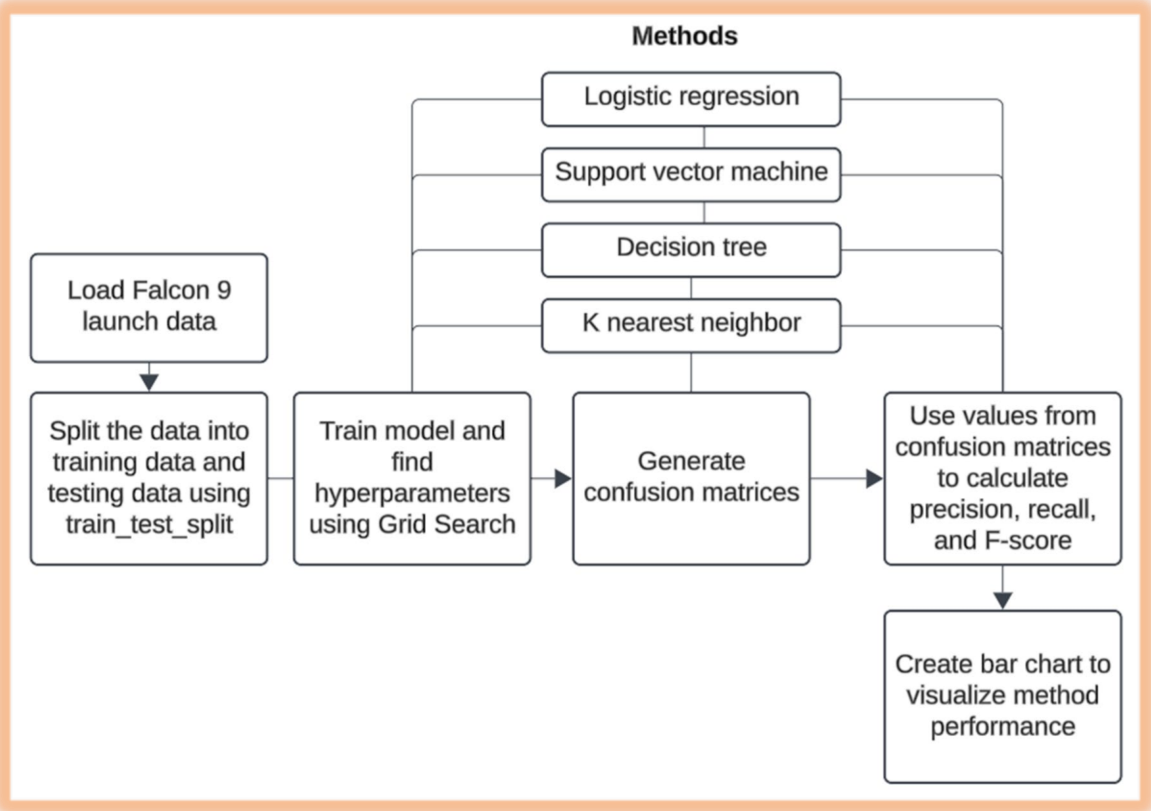
Range slider – allows user to select various payload ranges to better visualize the correlation between payload and outcome

Scatter plot – used in conjunction with the slider, used to visualize correlations between payload and outcome



Predictive Analysis with Classification

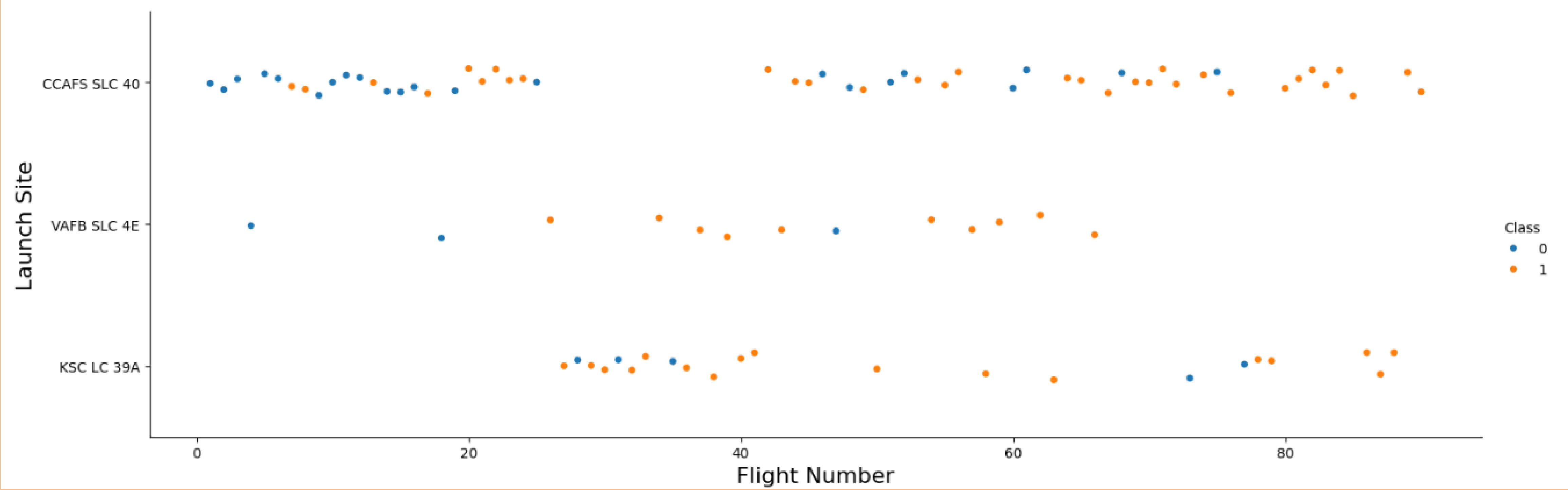
- Considered multiple methods to for predictive analysis
  - Logistic regression
  - Support vector machine (SVM)
  - Decision tree
  - K nearest neighbor (knn)
- Split data into (1) training set and (2) testing set
- Found hyperparameters using Grid Search to optimize performance of each method
- Used confusion matrices to count false positives, false negatives, and true positives
  - Used values to calculate precision and recall
- Identified best performing method (based on F-Score)



Section 2

Insights drawn from EDA

Flight Number and Lunch Site



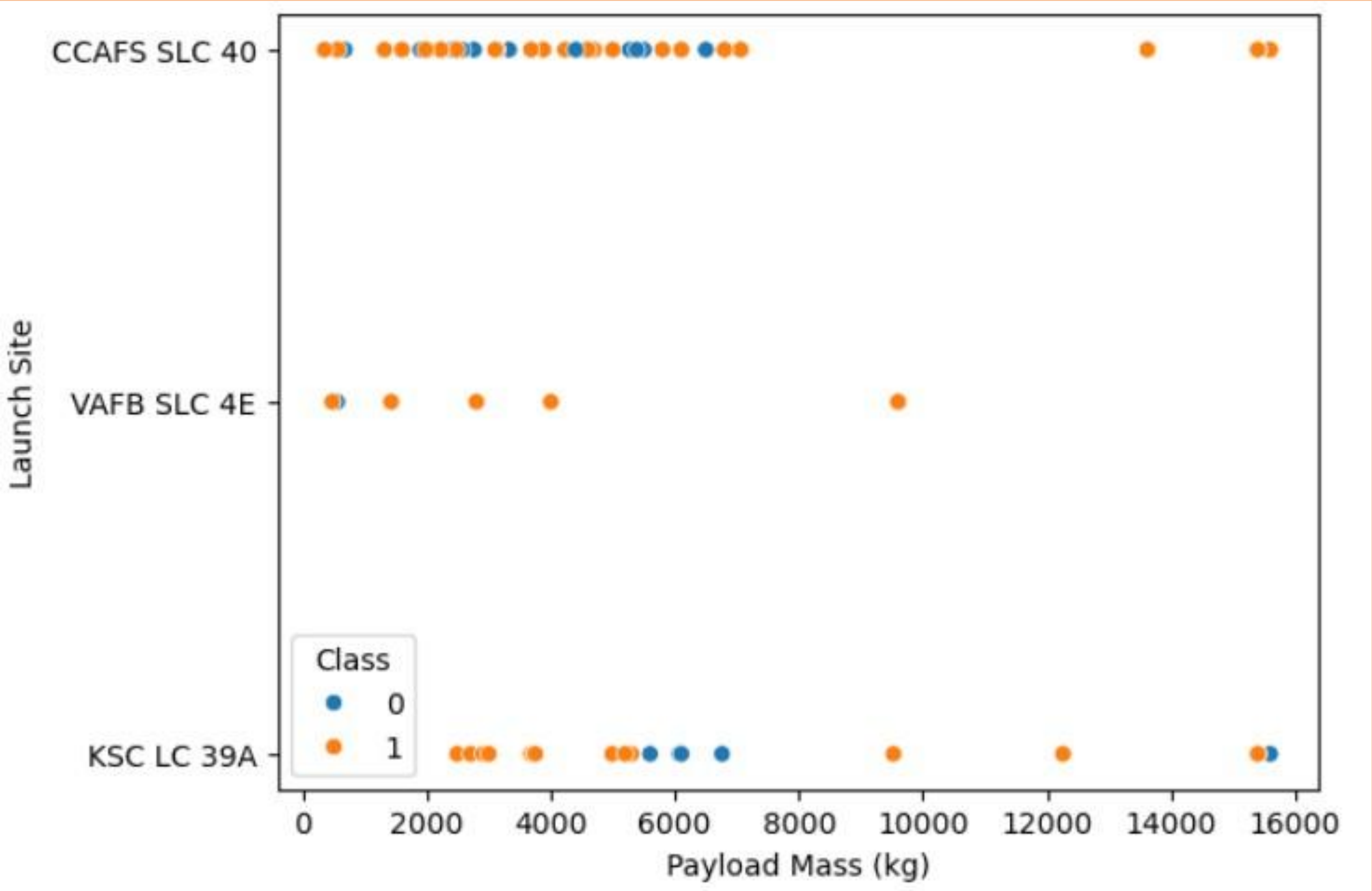
Scatter plot to visualize relationship between Flight Number and Launch Site

There were more unsuccessful landings in earlier flights among all launch sites and more successful landings in later flights at each launch site

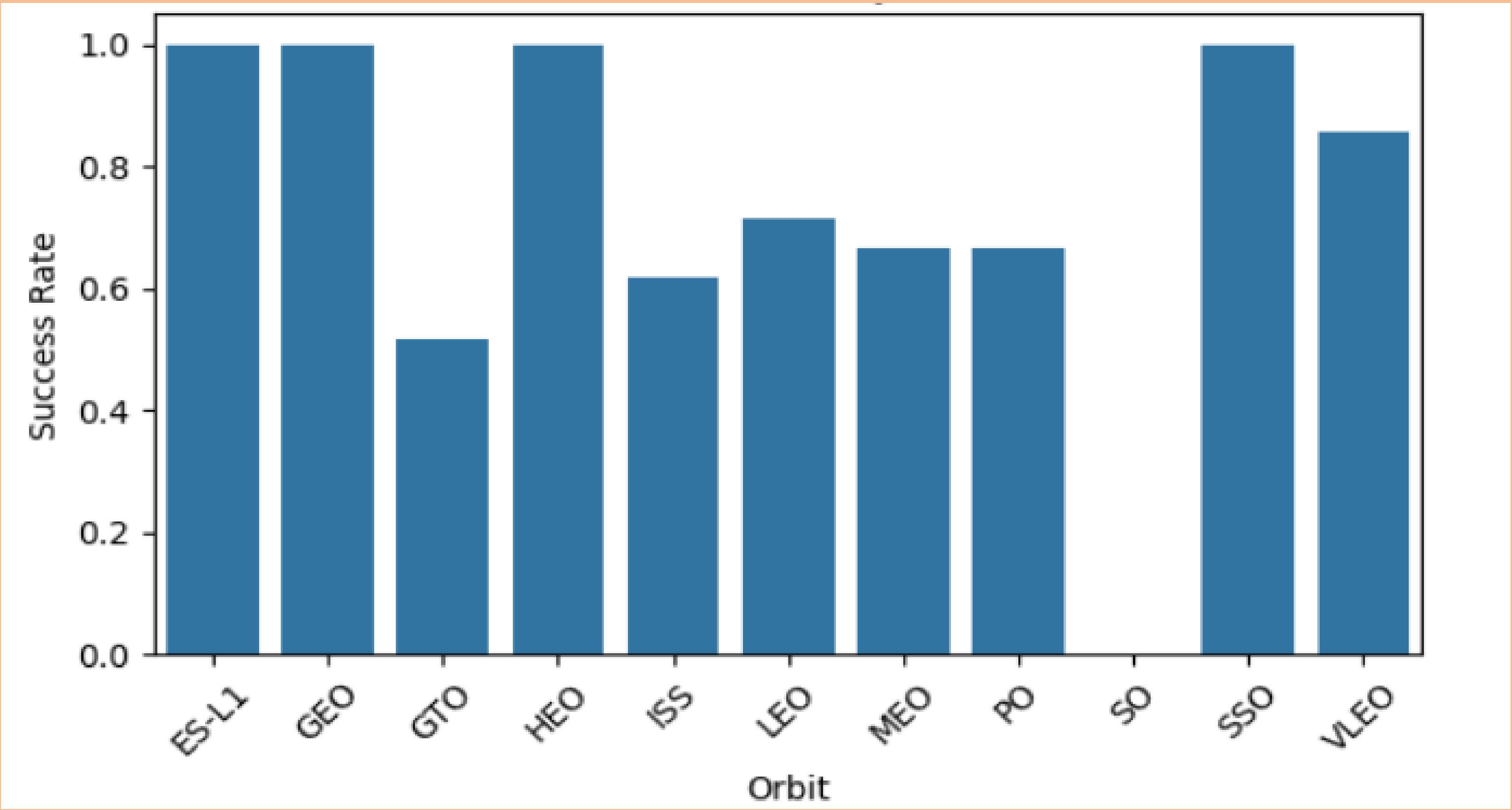
Payload Mass and Launch site

Scatter plot to visualize relationship between payload and launch site

- There were no rockets launched for heavy payload mass (>10000 kg) at the VAFB-SLC launch site
- CCAFS SLC 40 had general success in launching heavy payloads, though the instances were few

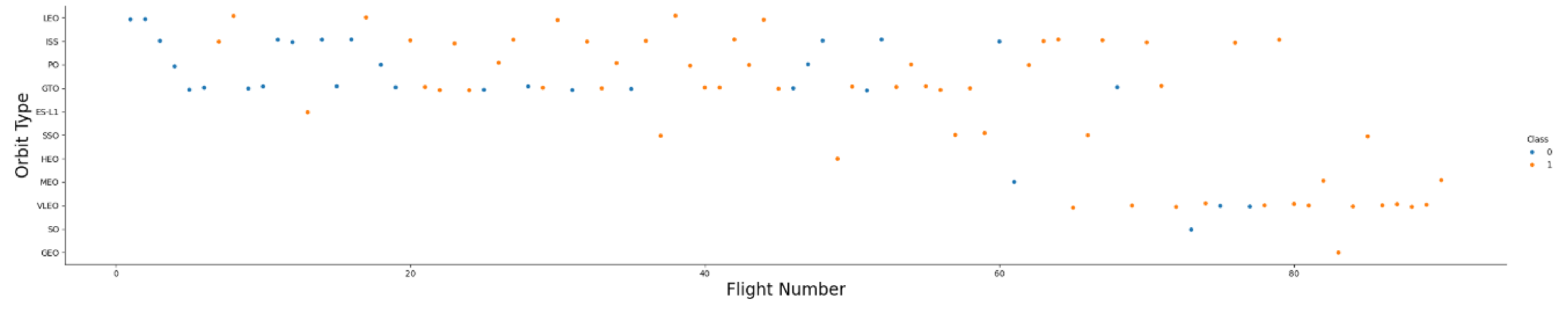


Success Rate by Orbit Type



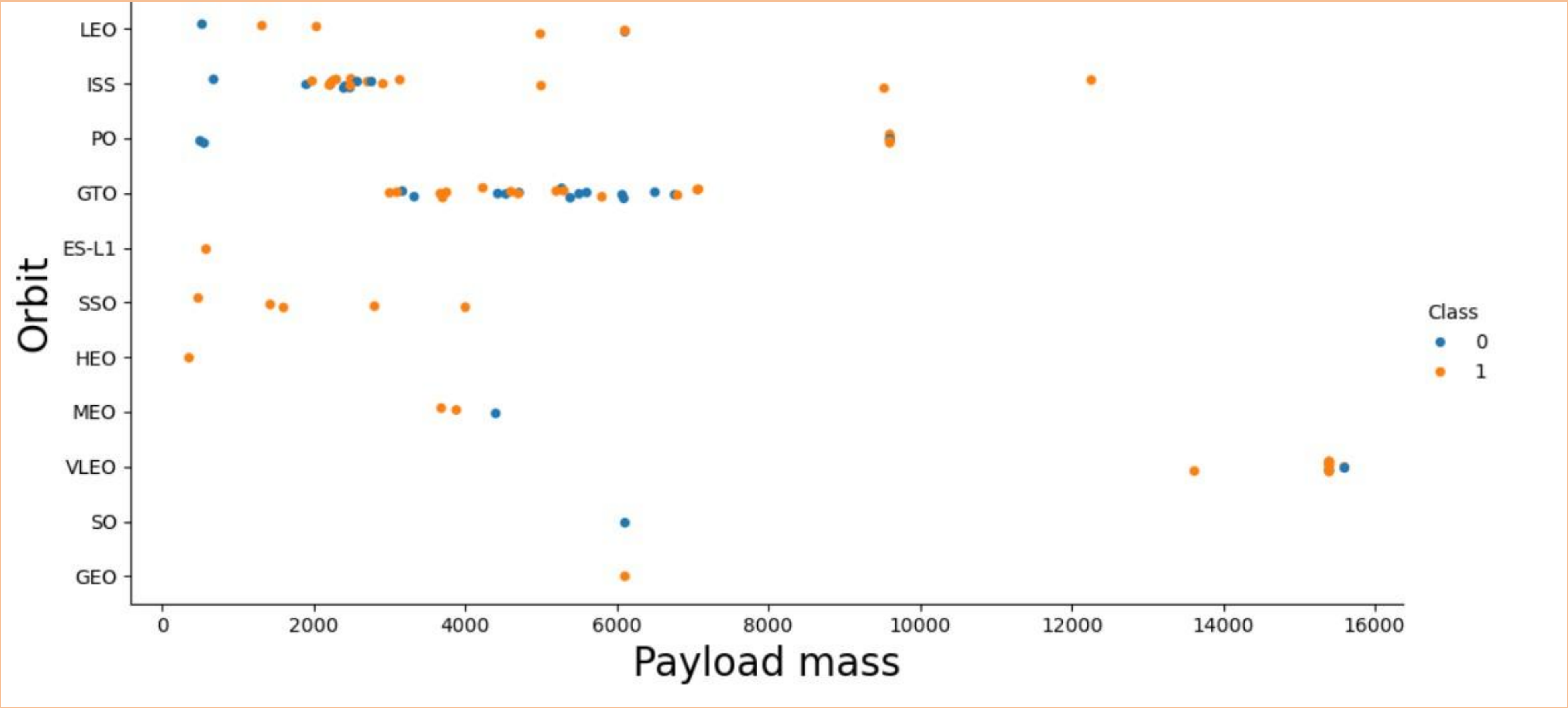
ES-L1, GEO, HEO, and SSO demonstrate the highest success rates among all orbit types

Flight Number and Orbit Type



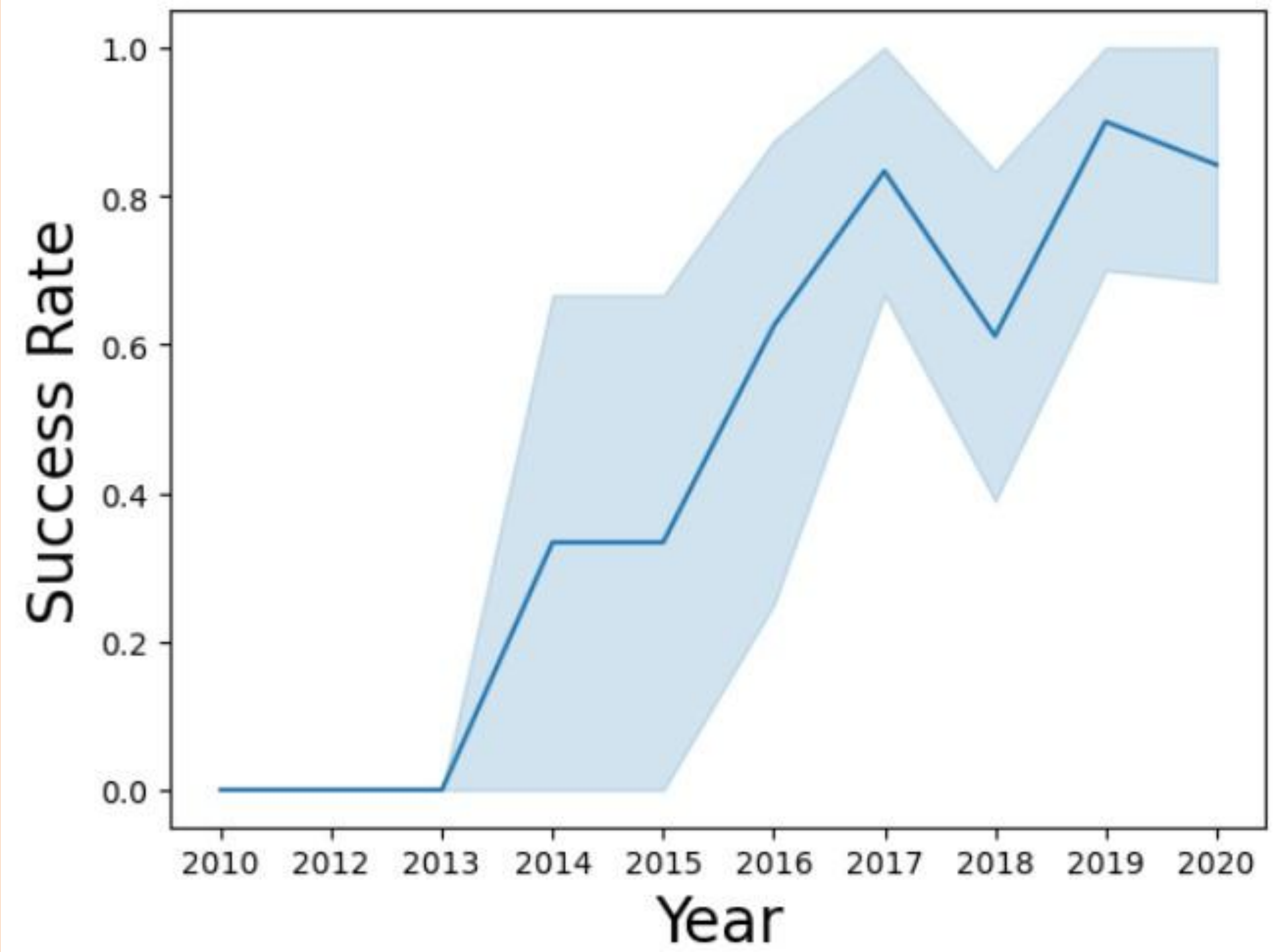
In low earth orbits (LEO), success seems to be relational to the number of flights, whereas in geostationary transfer orbits (GTO), there seems to be no relationship between a flight number and its success

Payload Mass and Orbit Type



Polar, LEO, and ISS orbits demonstrate higher success rates with heavy lades than other orbits.

Launch Success by Year



Success rates generally improved from 2013-2020, despite a minor decrease in success in 2018

All Site Launches

Objective

Display the names of the unique launch sites in the space mission

SQL Query

%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE

Outcome

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

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

Cape Canaveral Launch Sites (CCA)

Objective

Display 5 records where launch sites begin with ‘CCA’

SQL Query

%sql SELECT * FROM SPACEXTABLE WHERE "LAUNCH_SITE" LIKE "CCA%" LIMIT 5

Limited to 5 rows for ease of viewing. To view all records with launch sites beginning with CCA, omit LIMIT 5 from end of query

Outcome

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 on Falcon 9 Launches

Objective

Display the total payload mass carried by NASA-launched boosters (CRS)

SQL Query

%sql SELECT SUM(PAYLOAD_MASS_KG_) FROM SPACEXTABLE WHERE CUSTOMER="NASA (CRS)“

Outcome

Total payload mass = 45596 kg

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by Falcon 9 v1.1

- Objective
- Display the average payload mass carried by booster version F9 v1.1 SQL Query
  - %sql SELECT AVG(PAYLOAD_MASS_KG_) FROM SPACEXTABLE WHERE BOOSTER_VERSION="F9 v1.1“
- Outcome
  - Average payload mass = 2928.4

First Successful Ground Landing

- Objective
- Display the date of the first successful landing on a ground pad SQL Query
  - %sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE LANDING_OUTCOME="Success (ground pad)“
- Outcome
  - December 22, 2015

Successful Drone Ship Landing with Payload Mass 4000 kg – 6000 kg

Objective

List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

SQL Query

%sql SELECT PAYLOAD, BOOSTER_VERSION FROM SPACEXTABLE WHERE LANDING_OUTCOME=("Success (drone ship)") AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000

Outcome

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Payload	Booster_Version	PAYLOAD_MASS__KG_
JCSAT-14	F9 FT B1022	4696
JCSAT-16	F9 FT B1026	4600
SES-10	F9 FT B1021.2	5300
SES-11 / EchoStar 105	F9 FT B1031.2	5200

Total Mission Outcomes by Success and Failure

Objective

Calculate the total number of successful and failure mission outcomes

SQL Query

%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number FROM SPACEXTABLE GROUP BY MISSION_OUTCOME

Outcome

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

Boosters Carrying Max Payload

Objectives

List the names of the booster which have carried the maximum payload

SQL Query

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTABLE WHERE PAYLOAD_MASS_KG_ = (SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEXTABLE)
```

Outcome

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

Display launches from 2015 (include month names, failure landing outcomes in drone ships, booster versions, and launch sites)

SQL Query

%sql SELECT substr(Date,6,2) as month, substr(Date,0,5) as year, DATE, BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME FROM SPACEXTABLE WHERE LANDING_OUTCOME="Failure (drone ship)" and substr(Date,0,5)="2015“

Outcome

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

Landing Outcomes Ranked 06/04/2010 – 03/20/2017

Objective

Display the total number of each landing outcomes between June 4, 2010 and March 20, 2017 and sort in descending order

SQL Query

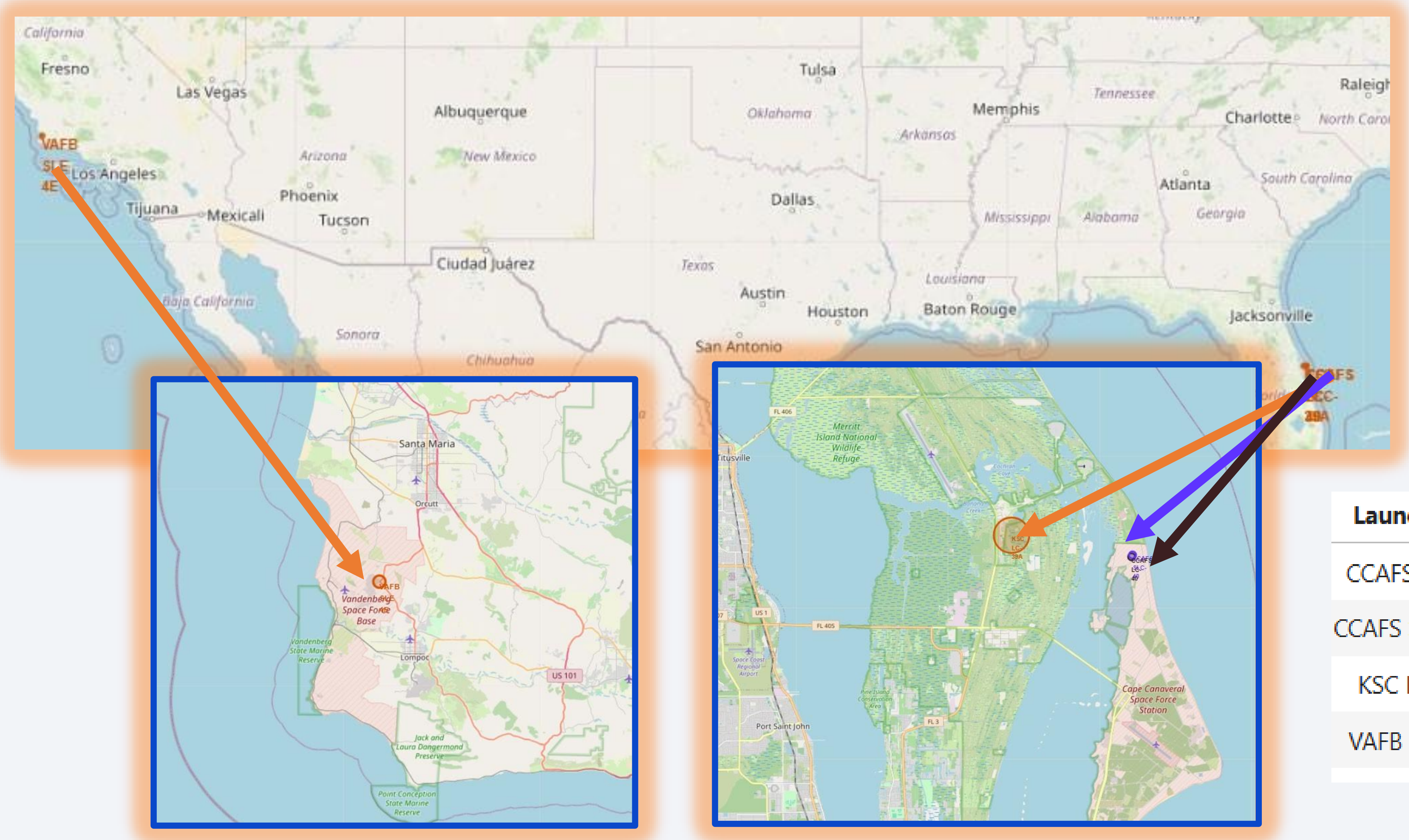
%sql SELECT LANDING_OUTCOME, COUNT(*) AS OUTCOME_COUNT FROM SPACEXTBL WHERE DATE between '2010-06-04' and '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY OUTCOME_COUNT DESC

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

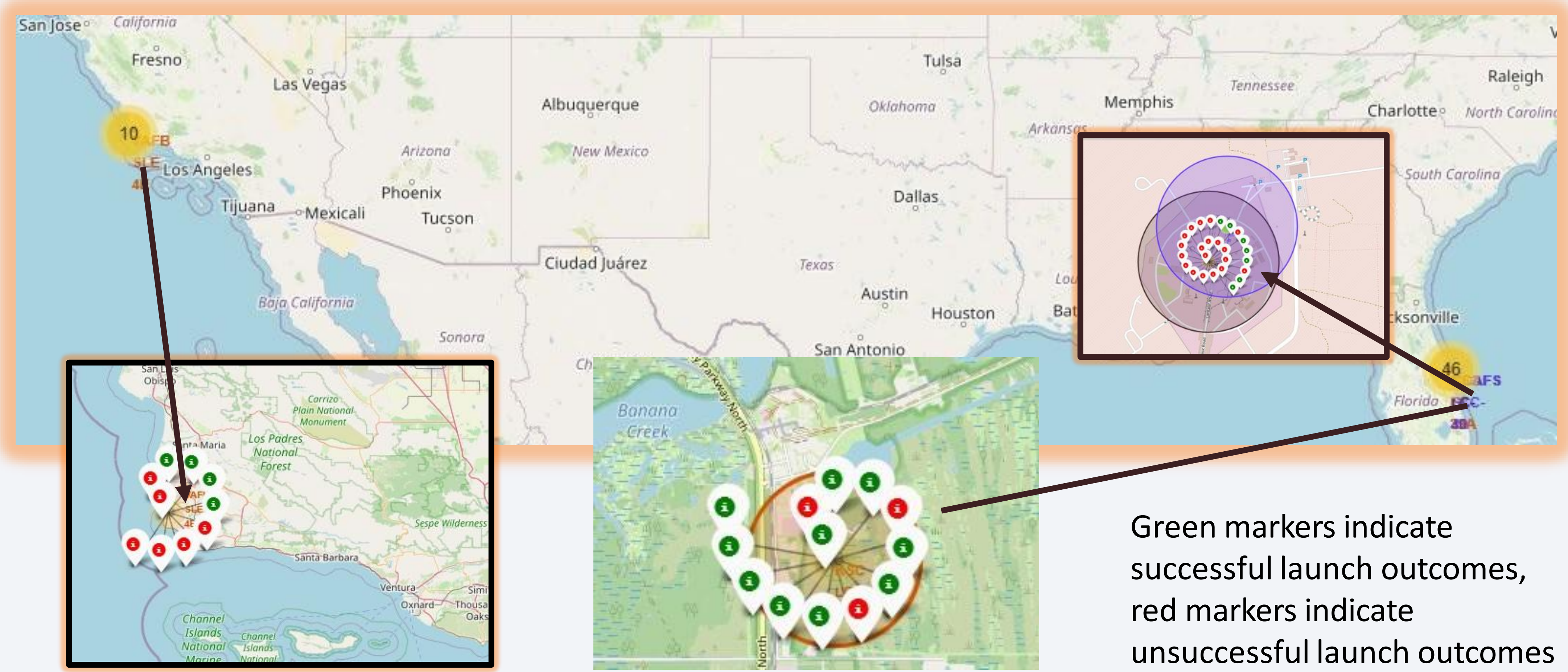
Map of All Launch Sites



Used longitude and latitude to add map markers of launch site locations in California and Florida

Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

Launch Outcome Markers



Green markers indicate successful launch outcomes, red markers indicate unsuccessful launch outcomes

Launch Sites and Proximities

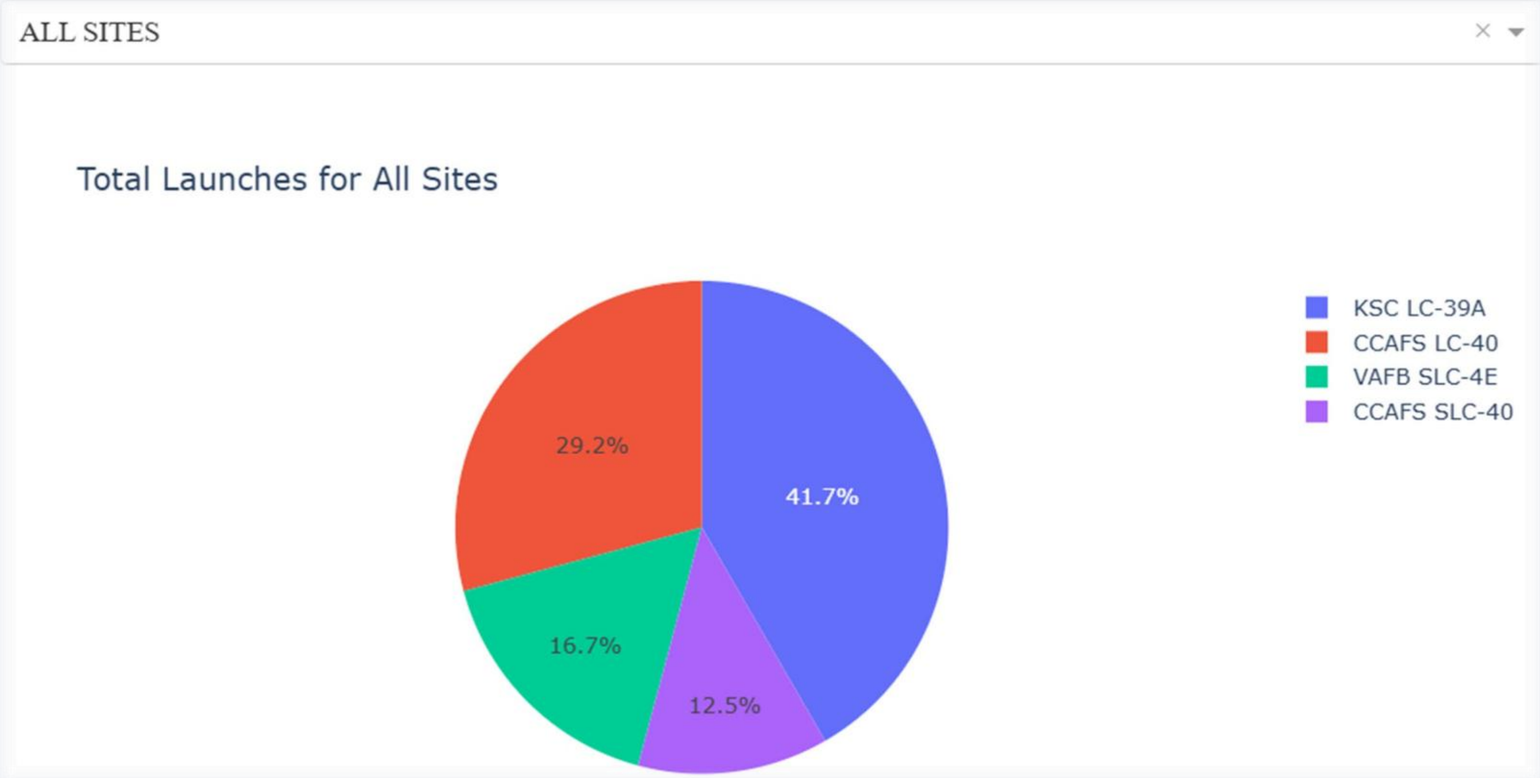
Determined the proximity of Kennedy Space Center to:

- Titusville, FL (~16 km)
- Florida East Coast Railway (~15 km)
- US 1 + FL 405 (~14 km)

Section 4

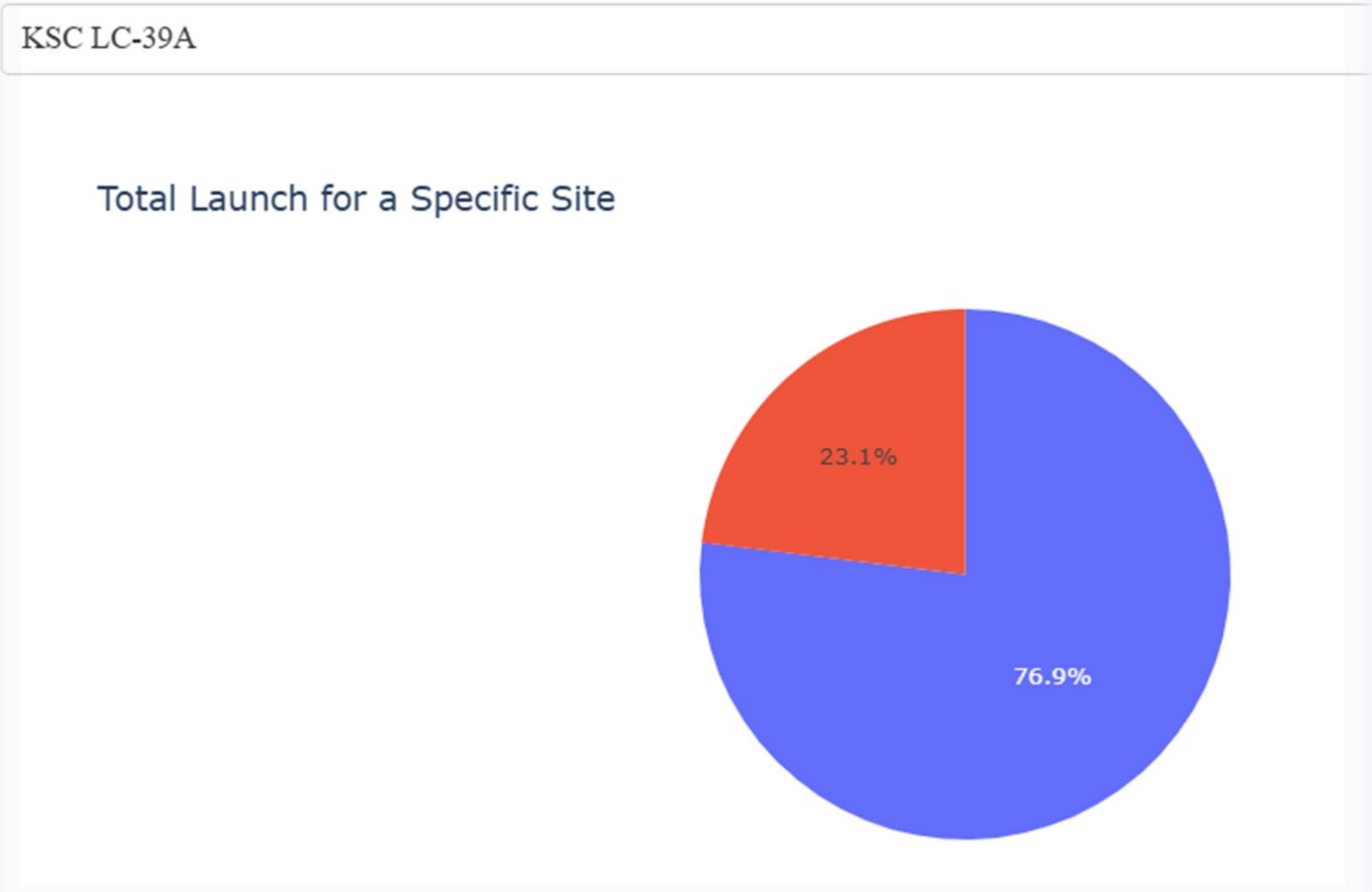
Build a Dashboard with Plotly Dash

Total Launches from All Sites



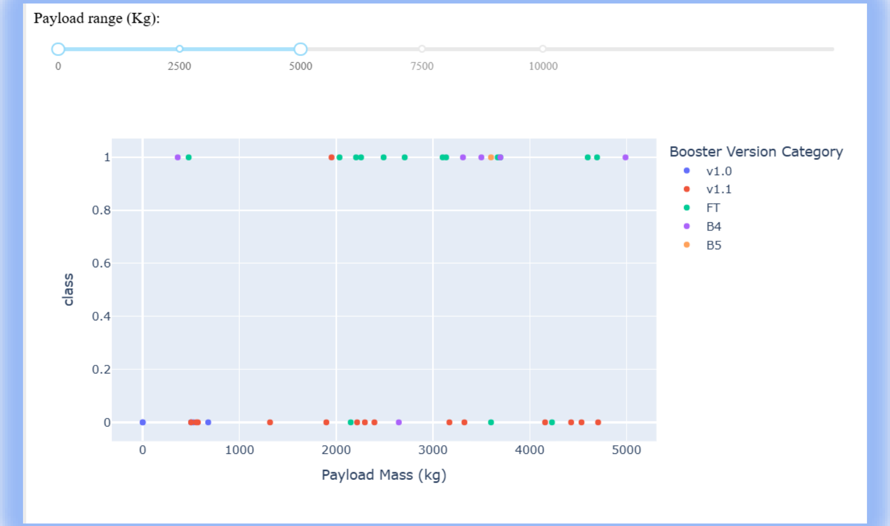
Kennedy Space Center Launch Complex 39 (KSC LC-39A) accounts for nearly 42% of all launches, with the other 58% split between Cape Canaveral Launch Control 40 (CCAFS LC-40), Vandenberg Space Launch Complex 4 (VAFB SLC-4E), and Cape Canaveral Space Launch Complex 40 (CCAFS SCL-40)

Highest Launch Success Rate

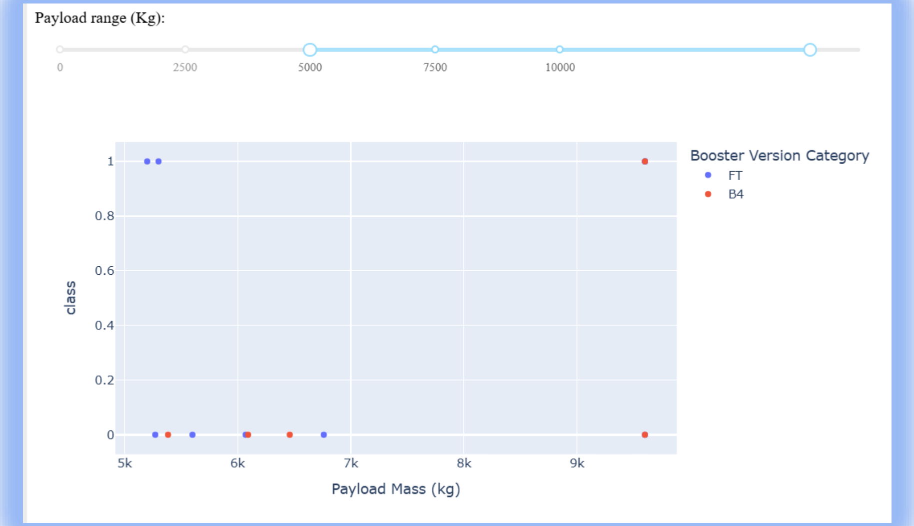


Kennedy Space Center Launch Complex 39 (KSC LC-39A) achieved a nearly 77% success rate, outperforming the next nearest launch site by 3.8%

Payload Mass and Launch Outcomes



Payload Mass 0 kg – 5000 kg



Payload Mass 5000 kg – 10000 kg

The success rate of very low payloads (0 kg – 2000 kg) is low, but begins to increase in slightly higher (but still low-weight) payloads (2000 kg – 4000 kg). The success rate of mid-weight payloads decreases again, though the total overall number of those launches are fewer.

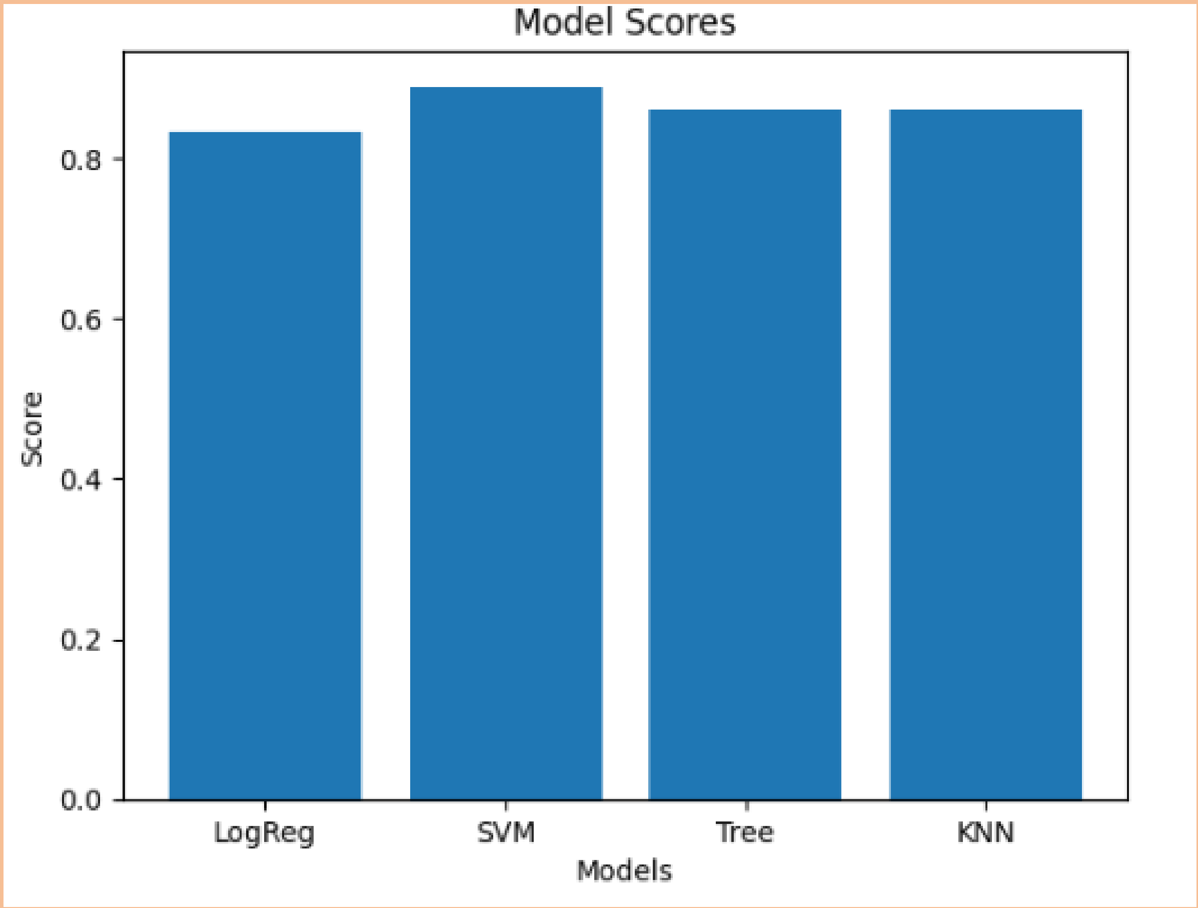
Section 5

Predictive Analysis (Classification)

Classification Accuracy

The Support Vector Machine (SVM) model performed marginally better than all other models

Model	Score
Logistic Regression	0.8333333333333334
Support Vector Machine	0.8888888888888888
Tree	0.8611111111111112
K Next Neighbor	0.8611111111111112



Confusion Matrix

Subject Vector Method (SVM)

True positives = 12

False positives = 3

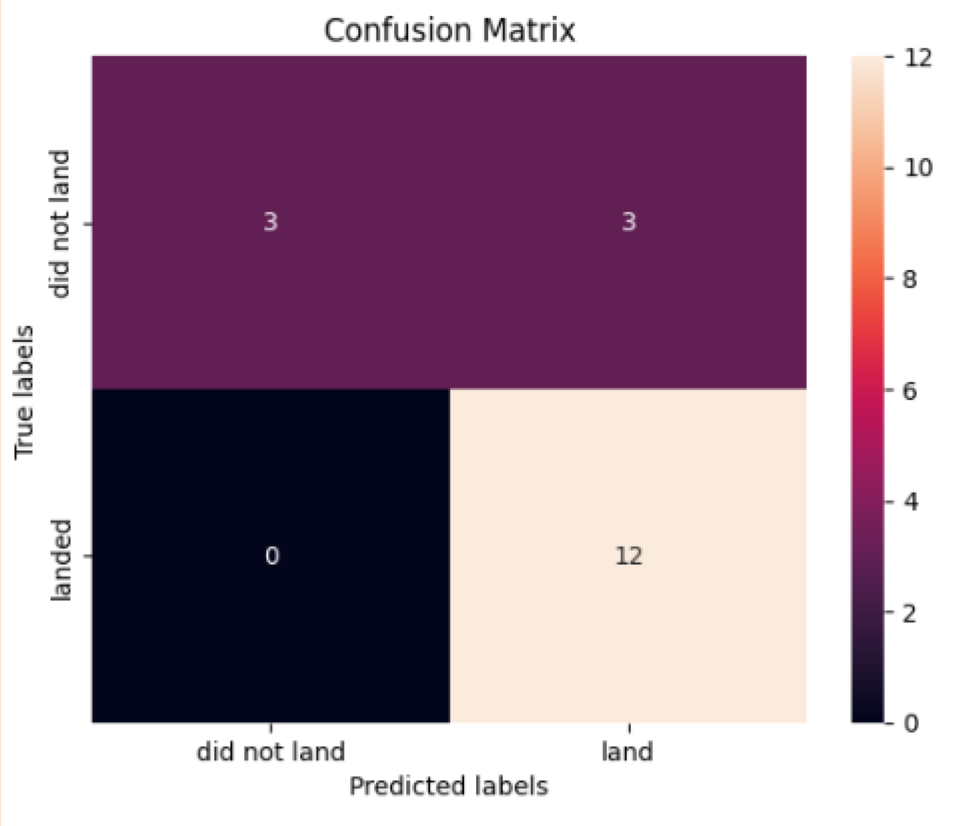
False negatives = 0

True negatives = 3

Precision = 0.8

Recall = 1.0

F-Score = 0.8888



Conclusions

- ES-L1, GEO, HEO, and SSO demonstrate the highest success rates among all orbit types
- Heavy payloads perform well in Polar, LEO, and ISS orbits.
- Success rates improved from 2013-2020, so studying the practices and trends in that time period may provide valuable insight to achieving successful launch outcomes
- Among all launch sites, there were more unsuccessful landings in earlier flights and more successful landings in later flights. Thus, persisting through early, unsuccessful launches is important to stabilizing and reaching successful launch outcomes
- Kennedy Space Center Launch Complex 39 (KSC LC-39A) achieved a nearly 77% success rate, outperforming the next nearest launch site by 3.8%
- The Support Vector Machine (SVM) model performed marginally better than all other models.

Appendix

- See my GitHub link

Thank you!