IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jan Pabiszczak
14 December 2024 r.

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- ISSUE:

This project seeks to predict the success of the Falcon 9 first stage landing, addressing the key question:

What factors influence the successful landing of Falcon 9's first stage, and how can this guide competitive bidding decisions for rocket launches against SpaceX?

- APPROACH:

We propose leveraging machine learning techniques to analyze historical launch data, including factors such as payload, weather conditions, and mission type, to identify patterns and predict the likelihood of a successful landing.

- IMPACT:

This model will empower data-driven decisions in competitive bidding by providing insights into the key factors influencing the success of Falcon 9's first stage landing. Additionally, it will serve as a robust tool to evaluate the risks and opportunities when competing with SpaceX for rocket launch contracts.

Introduction

- SpaceX, a pioneer in the space industry, aims to make space travel accessible to all. Its notable achievements include delivering spacecraft to the International Space Station, deploying a satellite constellation to provide global internet access, and launching manned missions into space.
- In this project, we will predict if the Falcon 9 first stage will land successfully.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- If we can determine whether the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

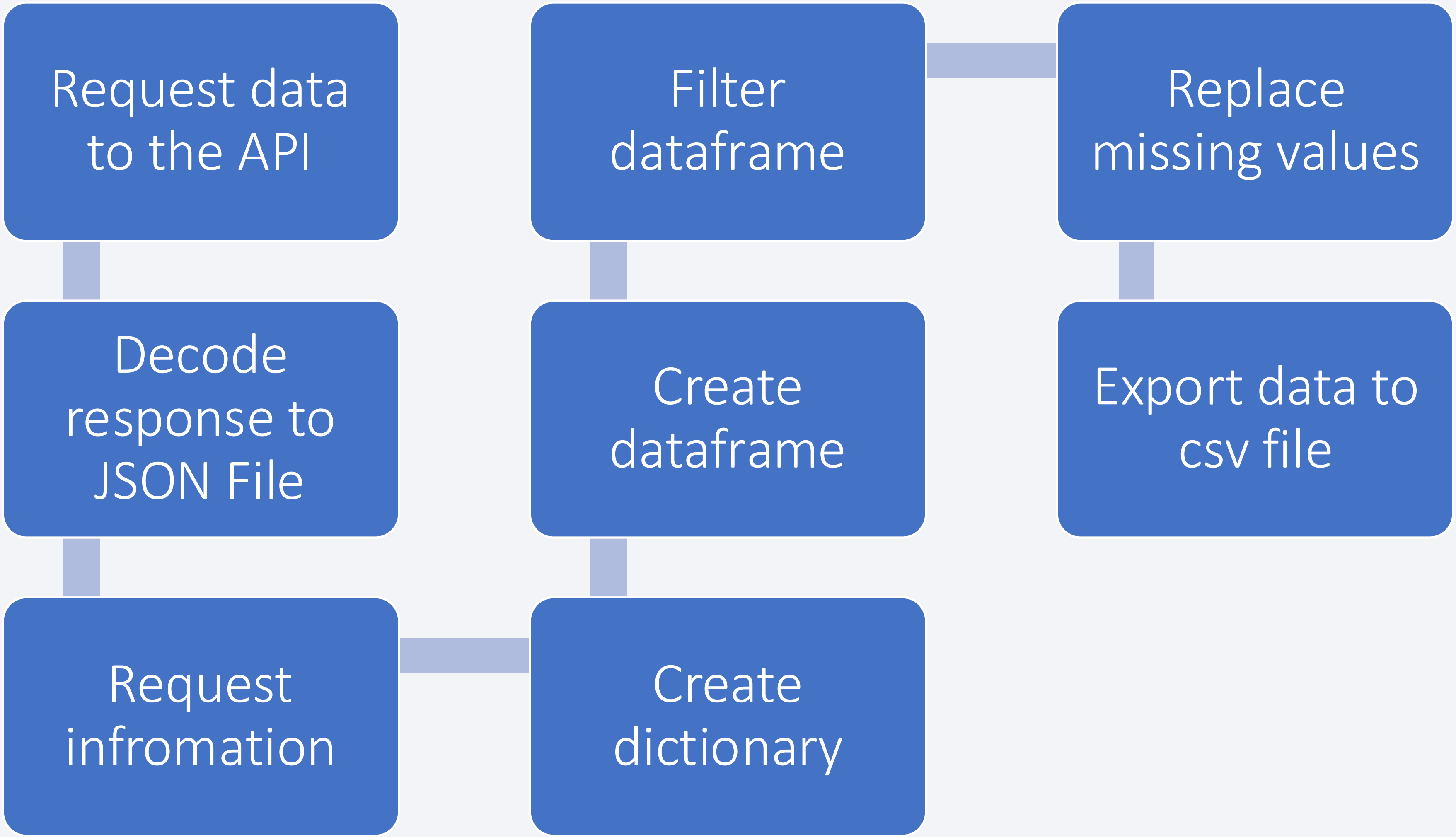
Methodology

Methodology

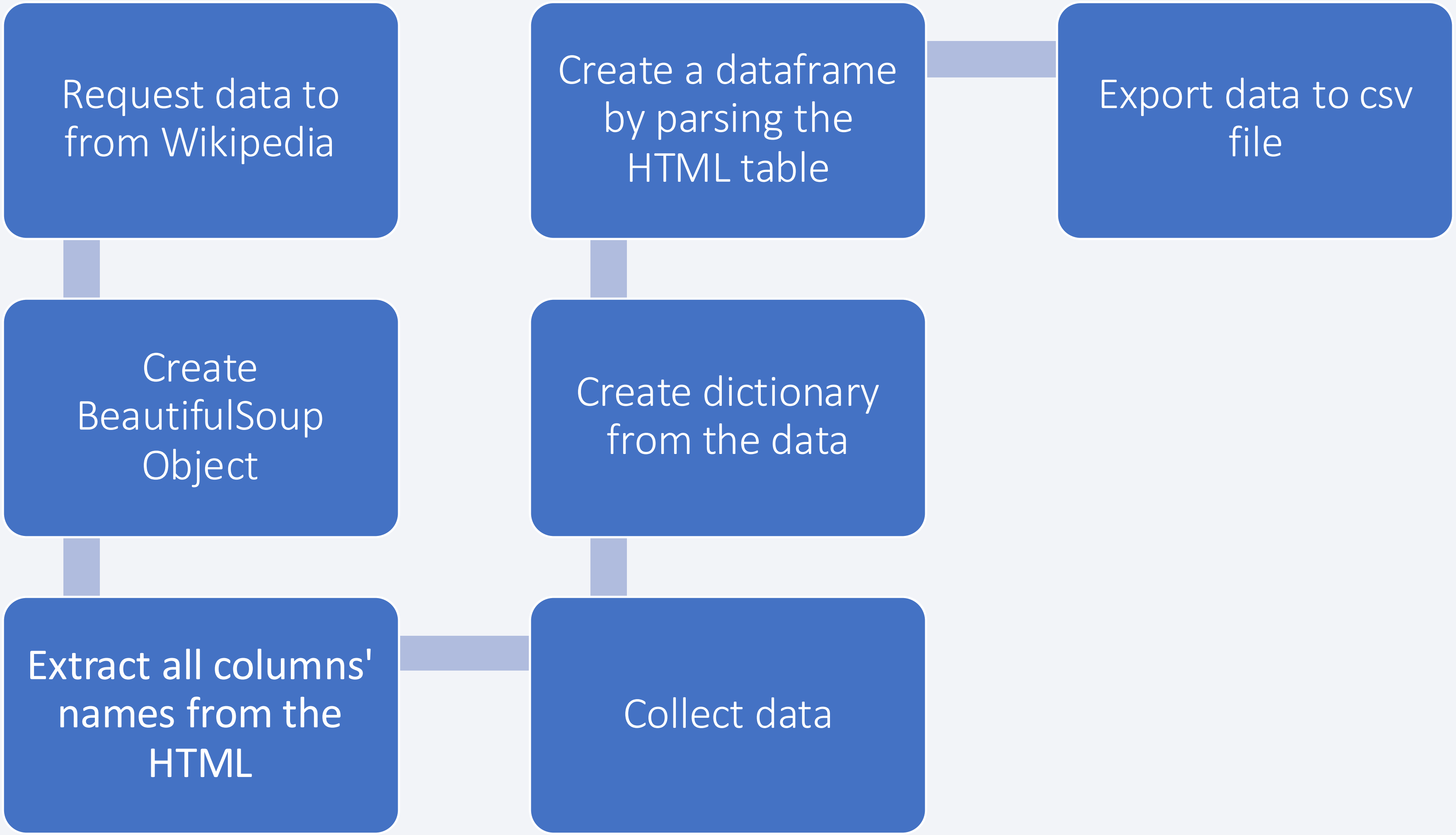
- Data collection, wrangling and formatting:
  - SpaceX API
  - Web scraping
- Exploratory Data Analysis
  - Pandas and NumPy
  - SQL
- Visualization:
  - Matplotlib and Seaborn
  - Folium
  - Dash
- Predictive Analysis (Classification):
  - Logistic regression
  - Support vector machine (SVM)
  - Decision tree
  - K-nearest neighbors (KNN)

Data Collection

Data Collection – SpaceX API



Data Collection - Scraping



Data Wrangling

1. Calculate the number of launches on each site
2. Calculate the number and occurence of mission outcome of the orbits
3. Create binary landing outcome column (dependent variable)
4. Export data to csv file

EDA with SQL

Display:

- the names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'¶
- the total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1

List:

- the date when the first succesful landing outcome in ground pad was acheived
- the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- the total number of successful and failure mission outcomes
- the names of the booster_versions which have carried the maximum payload mass.
- count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

EDA with Data Visualization

Scatter Graphs:

- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- FlightNumber vs Orbit type
- Payload Mass vs Orbit type

Scatter plots show the relationship between variables (correlation).

Bar graphs:

- success rate of each orbit type

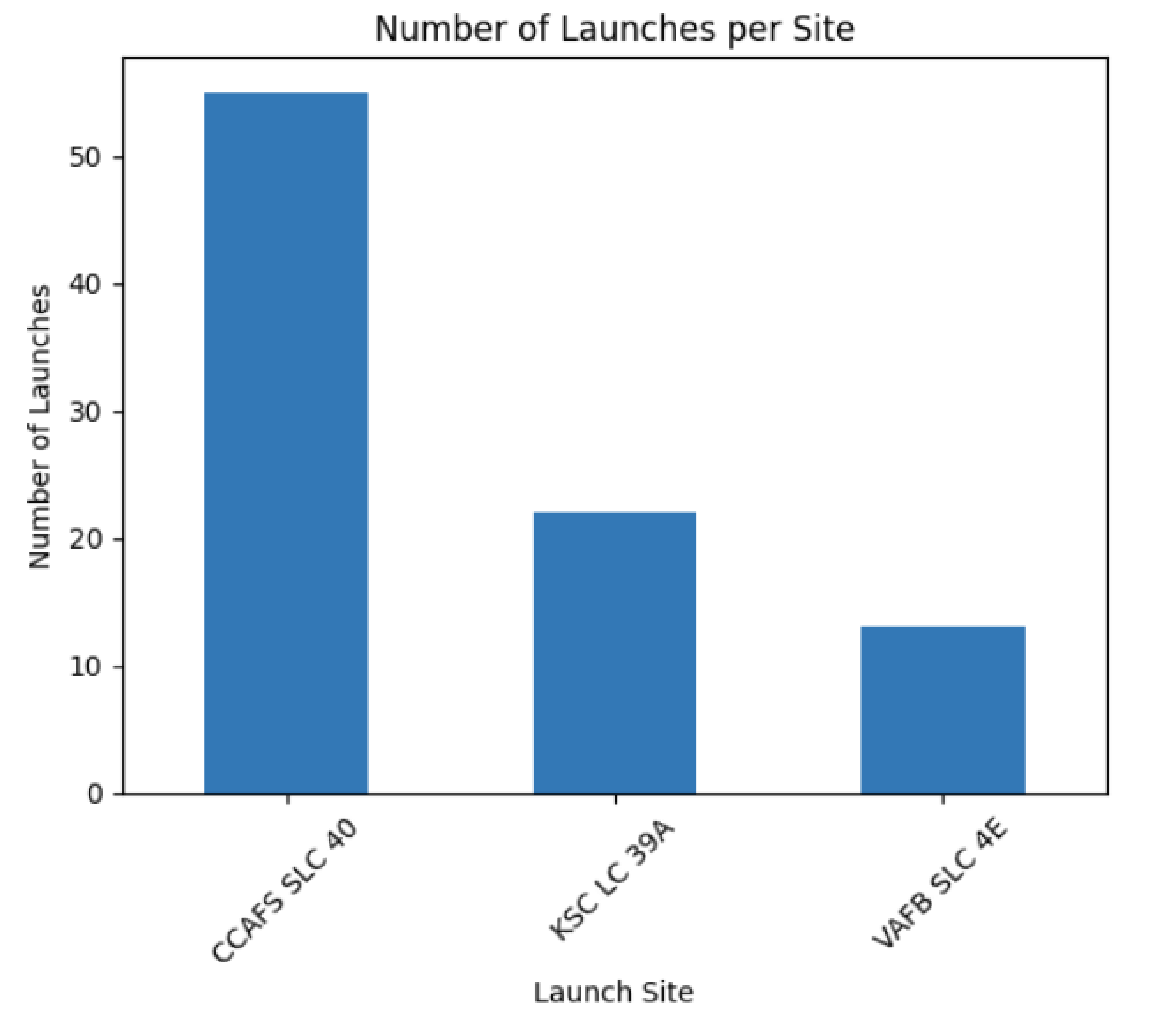
Bar graphs show the relationship between numeric and categoric variables

Line chart:

- the launch success yearly trend

Line graphs are the tool to indicate data variables and their trends.

Data Wrangling



Build a Dashboard with Plotly Dash

Objective: Provide an interactive tool to analyze SpaceX launch data for insights into success rates and payload relationships.

Features:

- **Dropdown for Launch Sites**:
  - Select "All Sites" or a specific site to view filtered data.
- **Pie Chart for Launch Outcomes**:
  - Visualizes the percentage of successful and unsuccessful launches for the selected site.
- **Range Slider for Payload Mass**:
  - Adjust payload mass range to filter data dynamically.
- **Scatter Plot for Payload vs. Success**:
  - Displays the correlation between payload mass and launch success, grouped by booster version.

Predictive Analysis (Classification)

- **Data Preparation**:
  - Load and normalize the dataset to ensure consistency.
  - Split the data into training and test sets for robust evaluation.
- **Model Training**:
  - Select appropriate machine learning algorithms (e.g., Logistic Regression, SVM, etc.).
  - Define hyperparameter ranges for each algorithm using **GridSearchCV**.
  - Train models using the training dataset to optimize performance.
- **Model Evaluation**:
  - Extract the best hyperparameters for each model.
  - Compute test accuracy to assess performance.
  - Visualize results using Confusion Matrices for deeper insights
- **Model Comparison**:
  - Compare models based on accuracy metrics and consistency.
  - Select the **best-performing model** for deployment or further analysis (see Notebook for detailed results).

Results

Exploratory Data Analysis:

- KSC LC-39A has the highest success rate among landing sites (considering total launches)
- the sucess rate since 2013 kept increasing till 2020
- orbits ES-L1, GEO, HEO and SSO have a 100% success rate

Launch_Site	Total_Launches	Successful_Launches	Success_Rate
CCAFS SLC-40	34	32	94.11764705882354
KSC LC-39A	25	25	100.0
CCAFS LC-40	26	25	96.15384615384616
VAFB SLC-4E	16	16	100.0

Predictive Analysis:

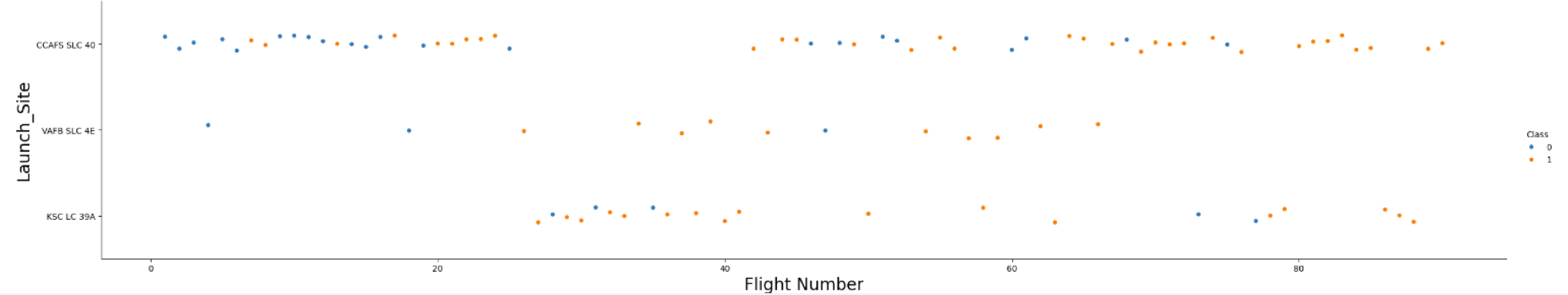
- From all the used models, Decision Tree model has the most accurate prediction power

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

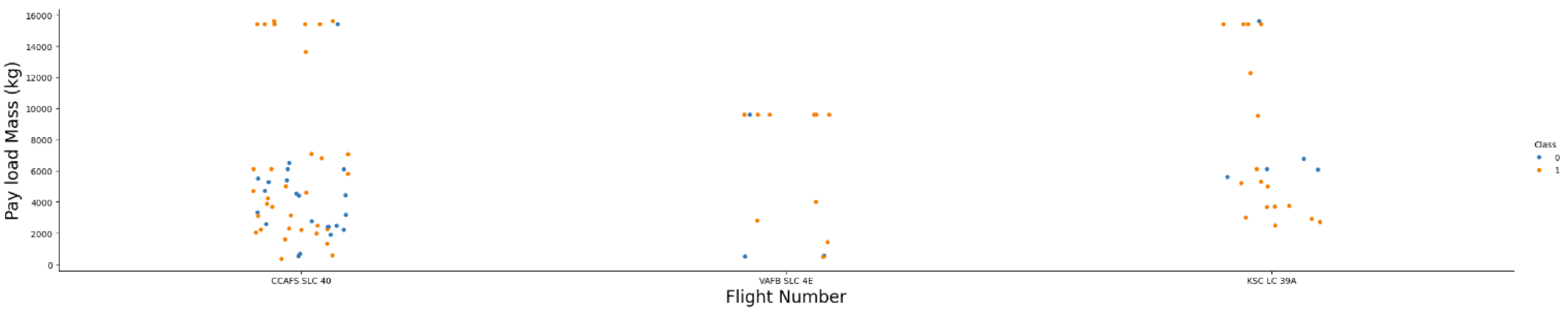
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

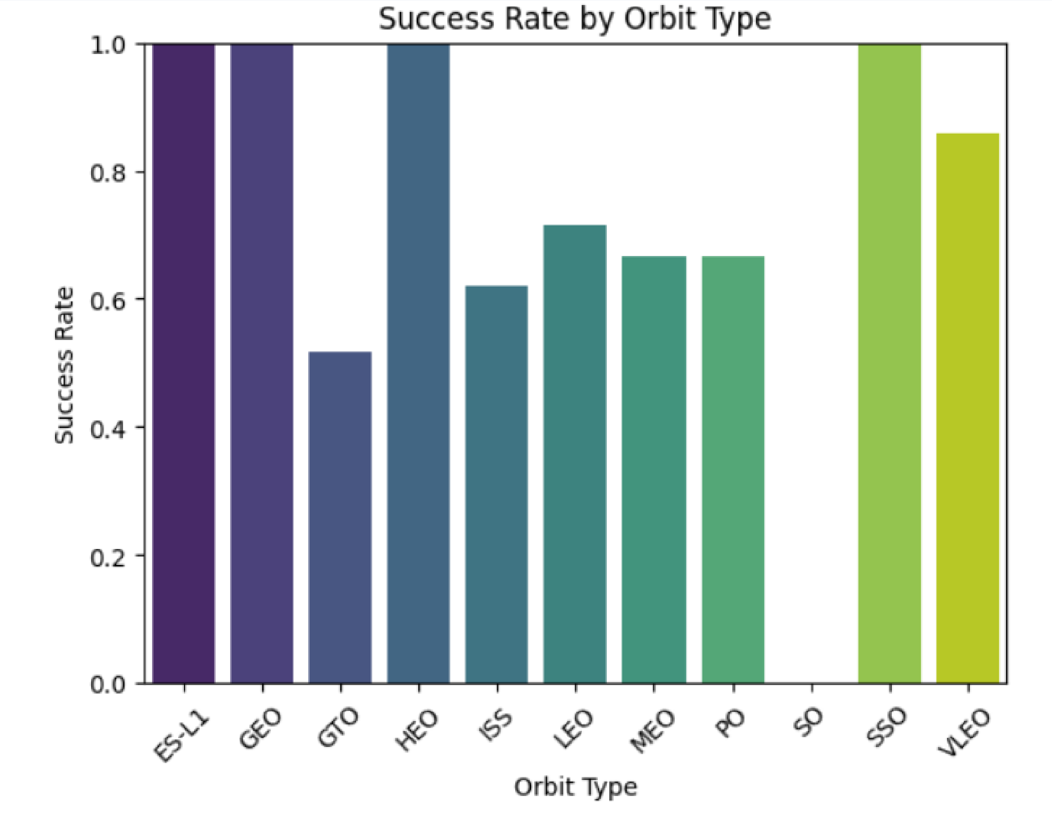


Payload vs. Launch Site



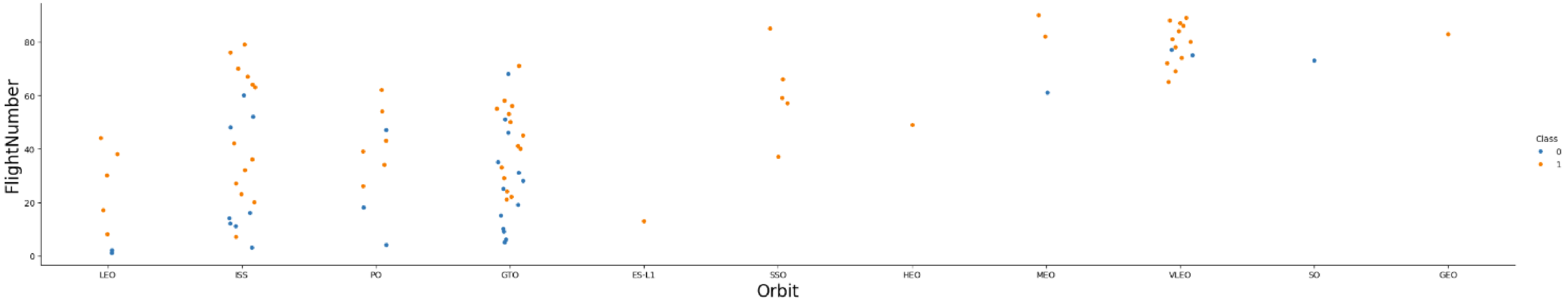
If we observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



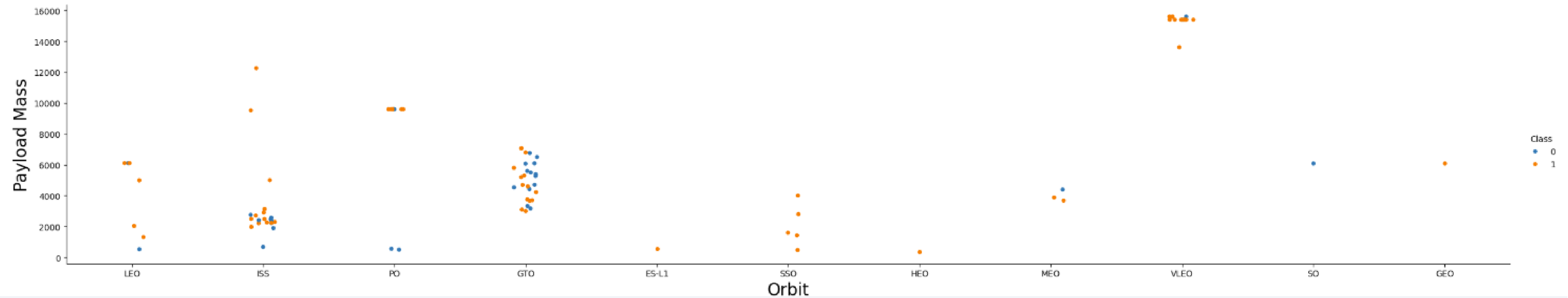
Orbits ES-L1, GEO, HEO and SSO have a 100% success rate

Flight Number vs. Orbit Type



We can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

We can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Displaying 5 records below:

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

The total Payload Mass carried by boosters launched by NASA equals: **45.596 kg.**

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1:

AVGPayloadMass
2928.4

First Successful Ground Landing Date

The dates of the first successful landing outcome on ground pad:

FirstSuccessfulLanding
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

The 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

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

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

The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

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

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

Rank of the landing outcomes count (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order:

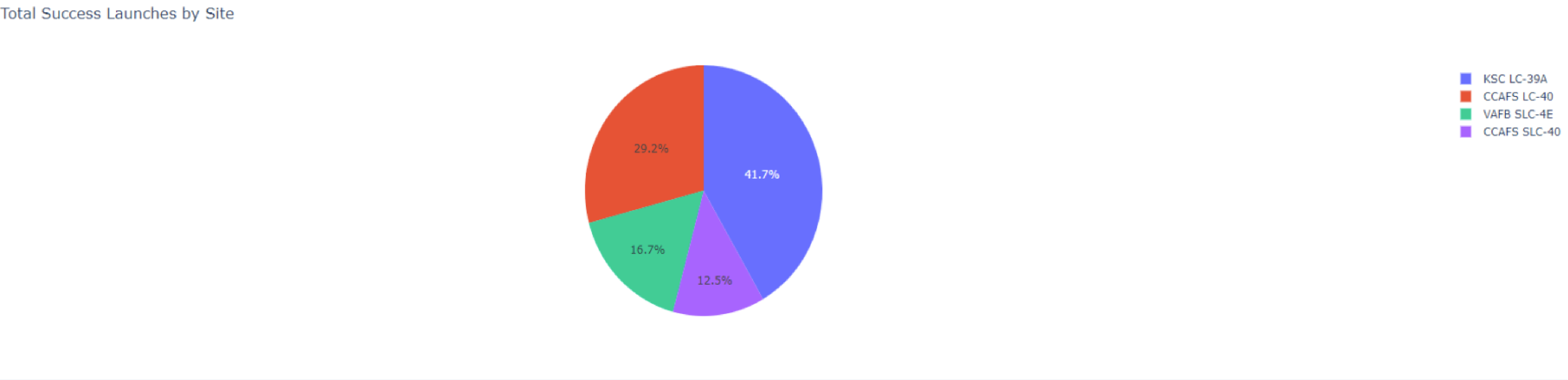
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

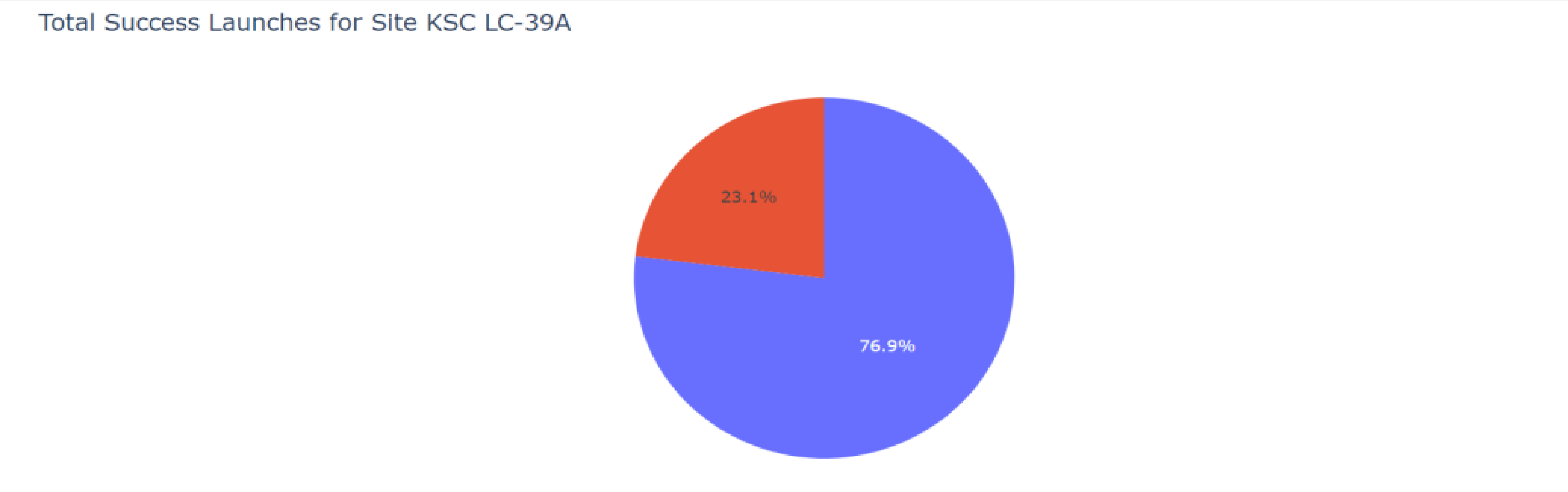
Build a Dashboard with Plotly Dash

Total success by Site

As we can see below KSC LC-39A has **the best success rate** of launches (41.2 %).



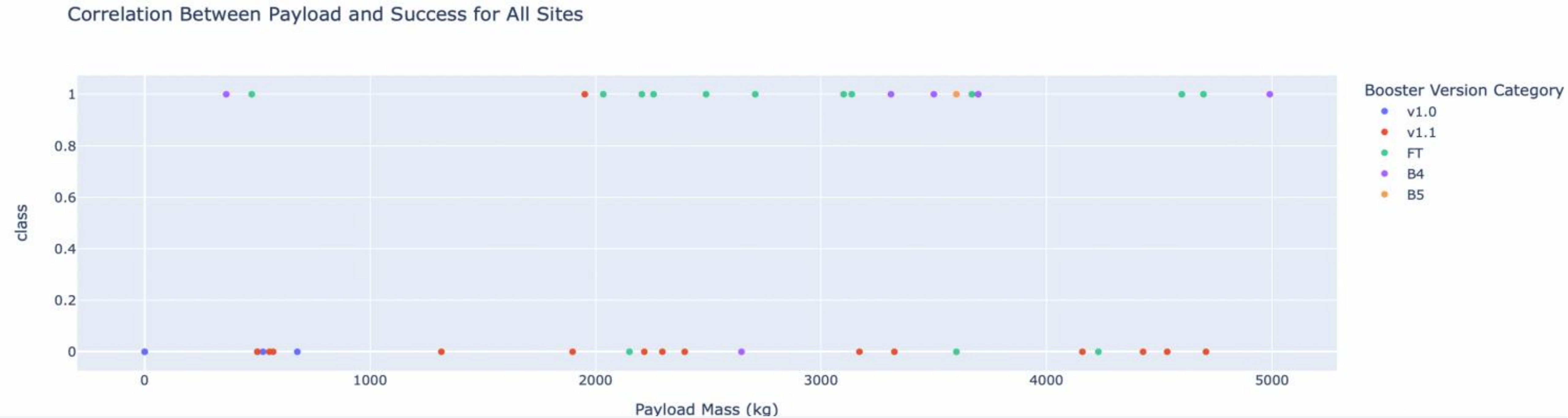
Launch Success for KCS LV-29A



Class 0 = Fail
Class 1 = Success

Payload Mass vs. Launch Outcome for all sites

Low weighted payloads (**between 2,000 kg and 5,000 kg**) have a better success rate than the heavy weighted payloads.



1: successful outcome
0: unsuccessful outcome

Section 4

Predictive Analysis (Classification)

Classification Accuracy

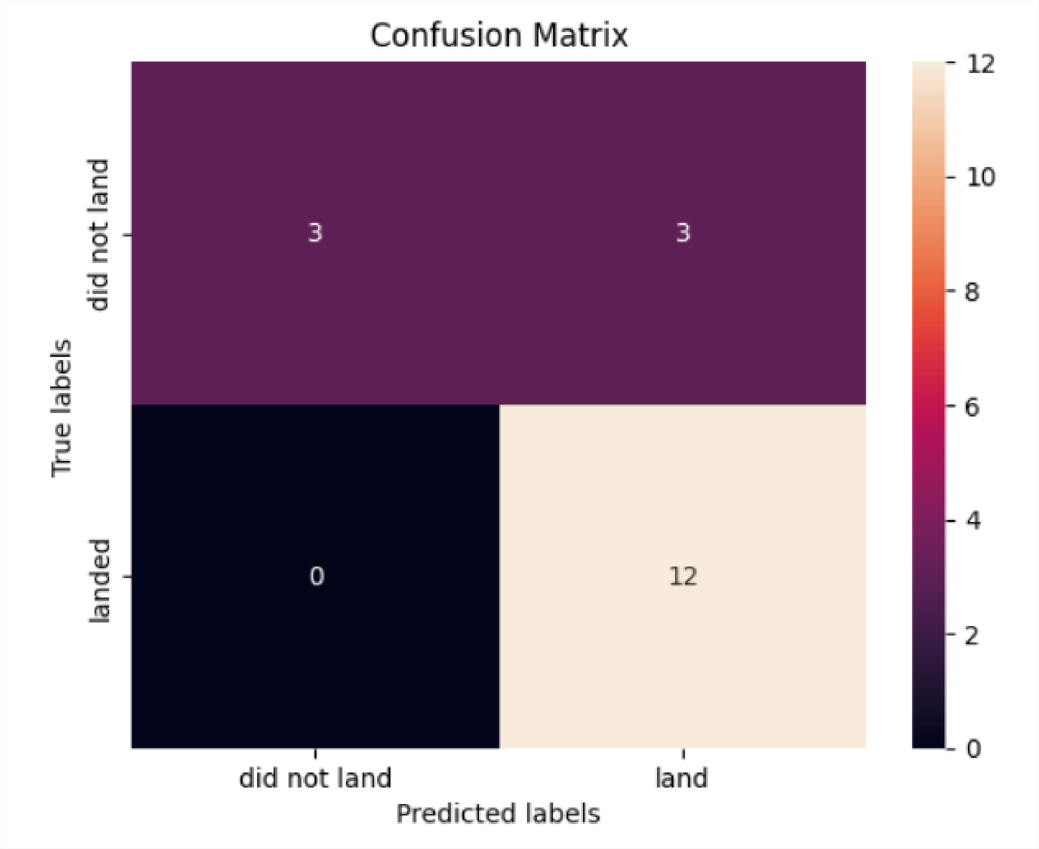
Accuracy Analysis:

- All models demonstrated similar performance, achieving comparable accuracy scores. This consistency is likely influenced by the limited size of the dataset.
- Among the models, the Decision Tree showed a slight edge over the others when evaluating the .best_score_.
- The .best_score_ represents the average cross-validation accuracy across all folds for a specific set of parameters.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.769231	0.800000
F1_Score	0.888889	0.888889	0.869565	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Confusion Matrix

Decision Tree confusion matrix:



Confusion matrix outputs:

- True Positive: 12
- True Negative: 3
- False Positive: 3
- False Negative: 0

Conclusions

- Among the models, the **Decision Tree** showed a slight edge over the others, but all the models demonstrated similar performance, achieving comparable accuracy scores
- **KSC LC-39A** has the highest success rate among landing sites (considering total launches)
- the **sucess rate** of the launches increases over time (since 2013 kept increasing till 2020)
- orbits **ES-L1, GEO, HEO and SSO** have a 100% success rate
- the success rate of a mission **depends on** several factors including among others: the launch site, the orbit and the number of previous launches

Thank you!