IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hassan Riaz
1st June, 2023

Outline

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

Executive Summary

Summary of methodologies:

- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification) Summary of all results
- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Summary of all results:

- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars which is much cheaper than other companies this savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.
- What launch sites have a high success rate?
- Which payload range has the lowest success rate?
- What is the overall success rate for all launches combined?

Section 1

Methodology

Methodology

Executive Summary

- Data collection:
  - SpaceX API and web scraping Wikipedia using beautiful soup was used to collect data.
- Data wrangling
  - Main highlight of data wrangling was creating a new column with binary values based on outcome.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardize data, split into training and testing sets then fitting into models.

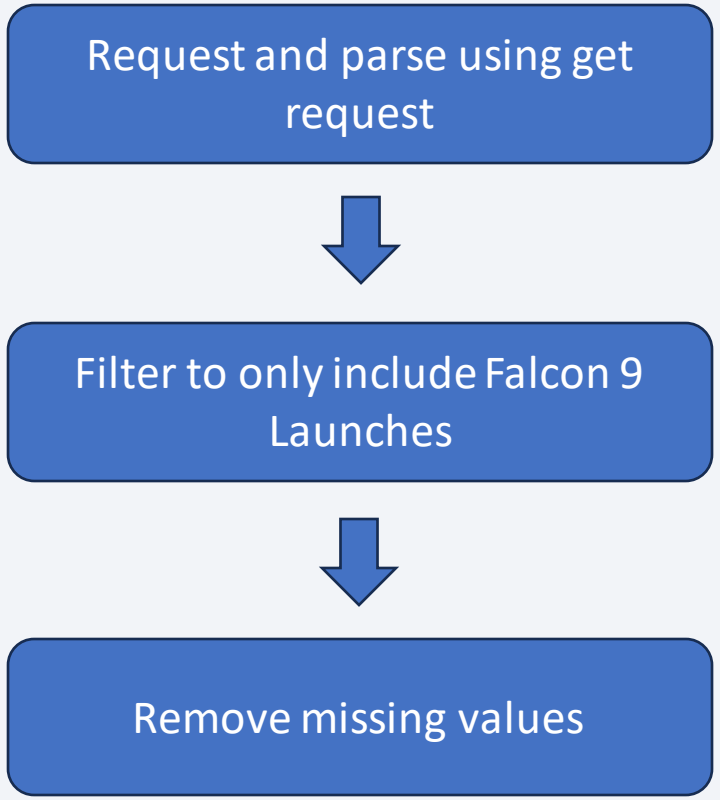
Data Collection

- Main source of data collection was kept SpaceX API, however we also used beautiful soup which is a python library used for web-scraping and data from Wikipedia page was scraped.
- Using the SpaceX API following data was collected:
  - Booster name of rocket.
  - Payload mass and orbit that rocket is going to.
  - Launch site being used by rocket.
  - Outcome (if the rocket landed or not).

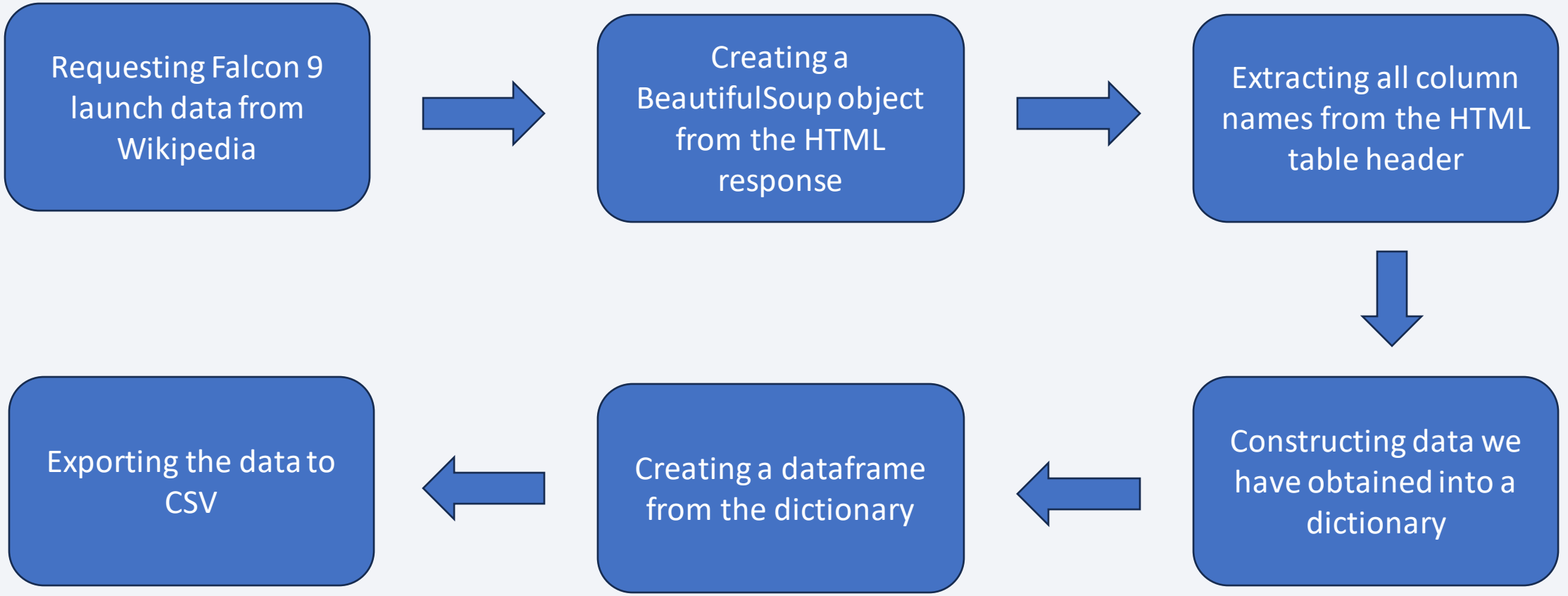
Data Collection – SpaceX API

- To collect data using SpaceX's REST API, I made HTTP requests to their endpoints and receive responses in JSON format.

- Github Notebook (Data collection)



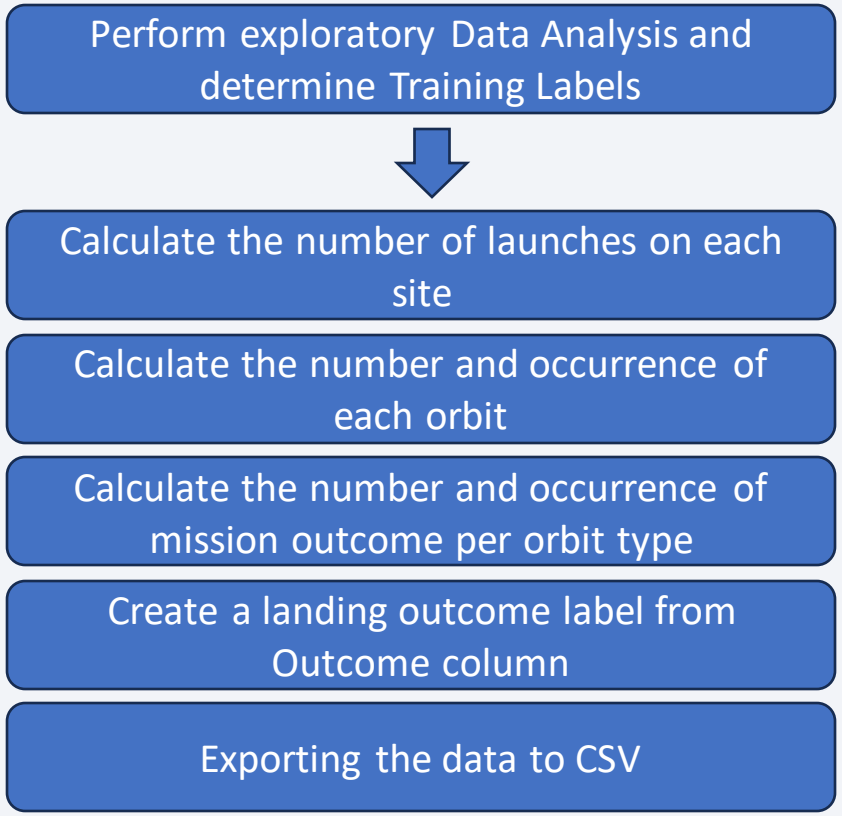
Data Collection - Scraping



Github - Webscraping

Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accidents.
- We mainly converted those outcomes into Training Labels with “1” means the booster successfully landed, “0” means it was unsuccessful.



Github - Wrangling

EDA with Data Visualization

Charts were plotted:

- Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend
- Scatter plots show the relationship between variables. If a relationship exists, they could be used in machine learning model.
- Bar charts show comparisons among discrete categories. The goal is to show the relationship between the specific categories being compared and a measured value.
- Line charts show trends in data over time (time series).

Github - EDA Visualization

EDA with SQL

Performed SQL queries:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes

Github - EDA SQL

Build an Interactive Map with Folium

Markers of all Launch Sites:

- Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location. - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

Colored Markers of the launch outcomes for each Launch Site:

- Added colored Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates

Github - Interactive Map with Folium

Build a Dashboard with Plotly Dash

Launch Sites Dropdown List:

- Added a dropdown list to enable Launch Site selection. Pie Chart showing Success

Launches (All Sites/Certain Site):

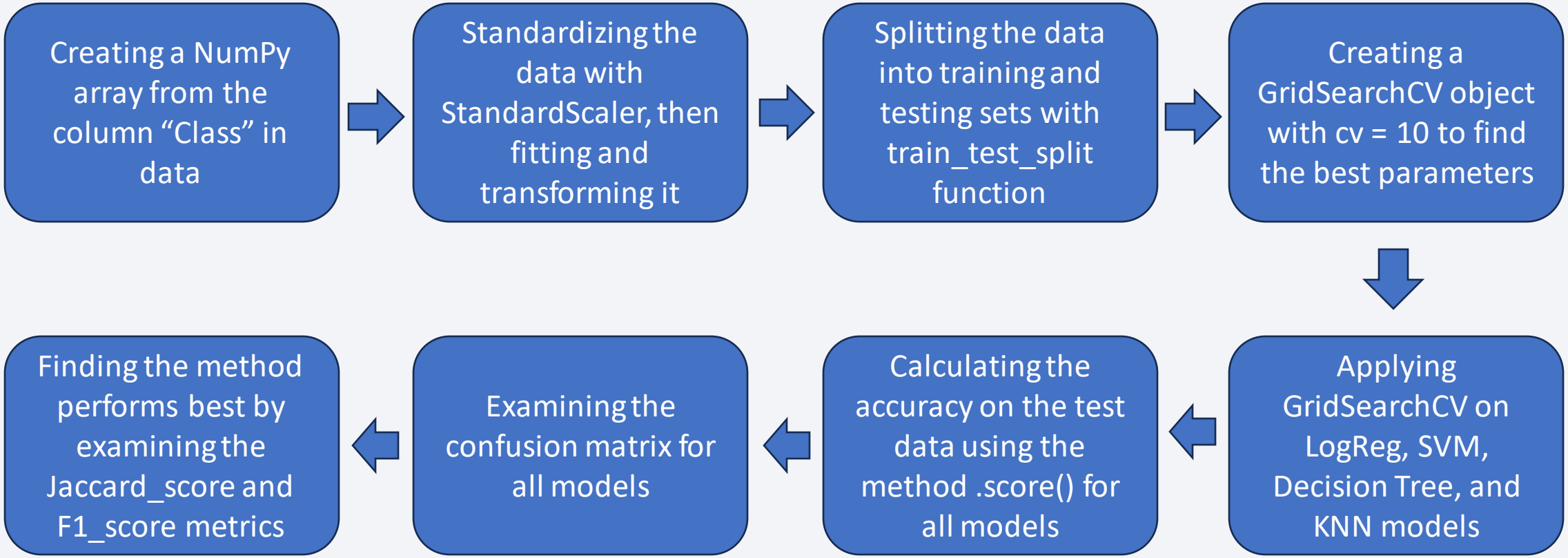
- Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

Slider of Payload Mass Range:

- Added a slider to select Payload range.

Github - Plotly Dash App

Predictive Analysis (Classification)



Github – Predictive Analysis

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

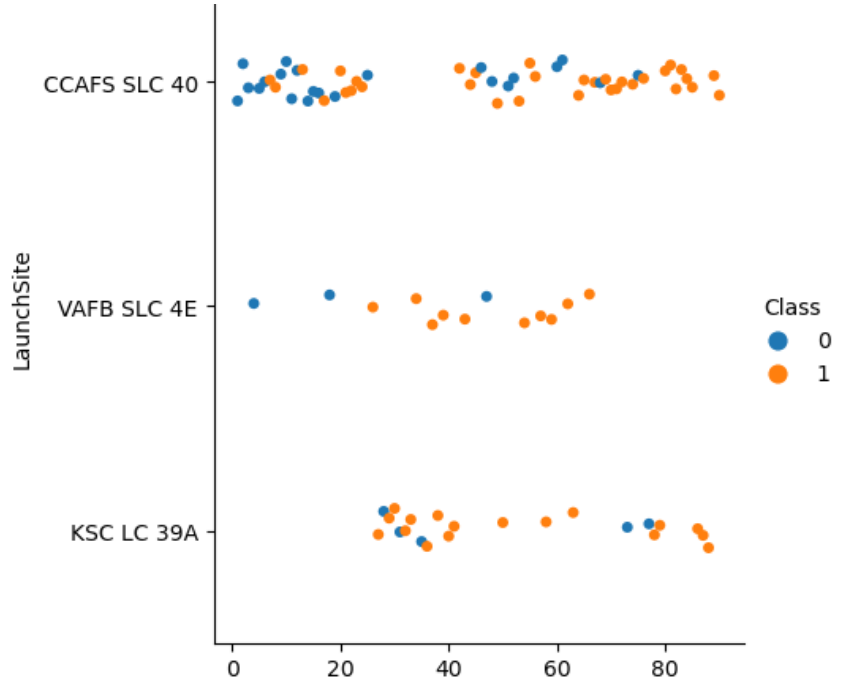
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Explanation:

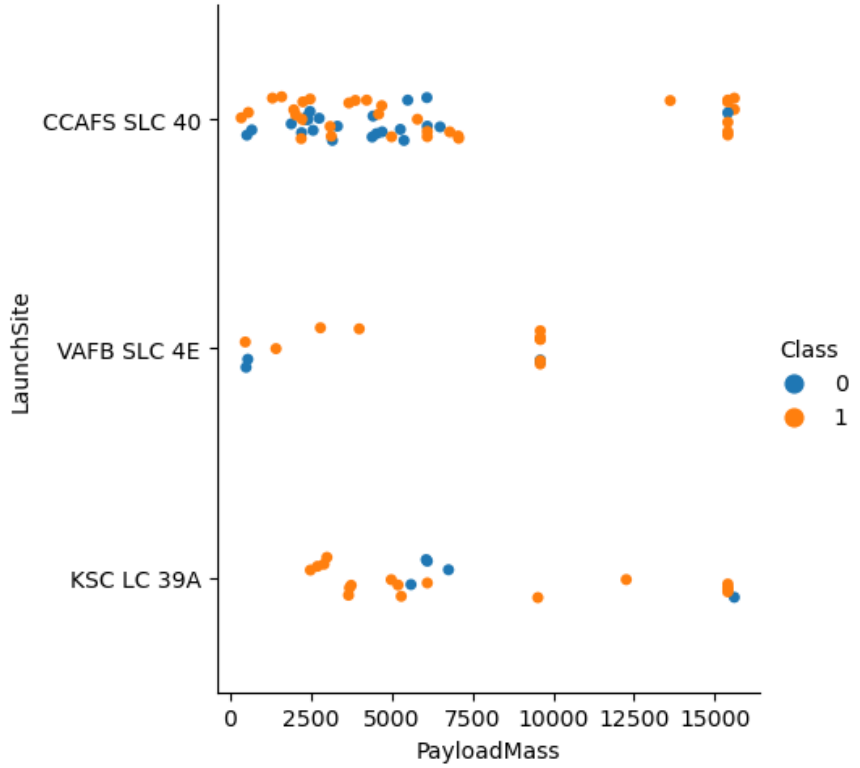
- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch has a higher rate of success.



Payload Mass vs. Launch Site

Explanation:

- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000 kg were successful.
- KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.



Success Rate vs. Orbit Type

Explanation:

- Orbits with 100% success rate:

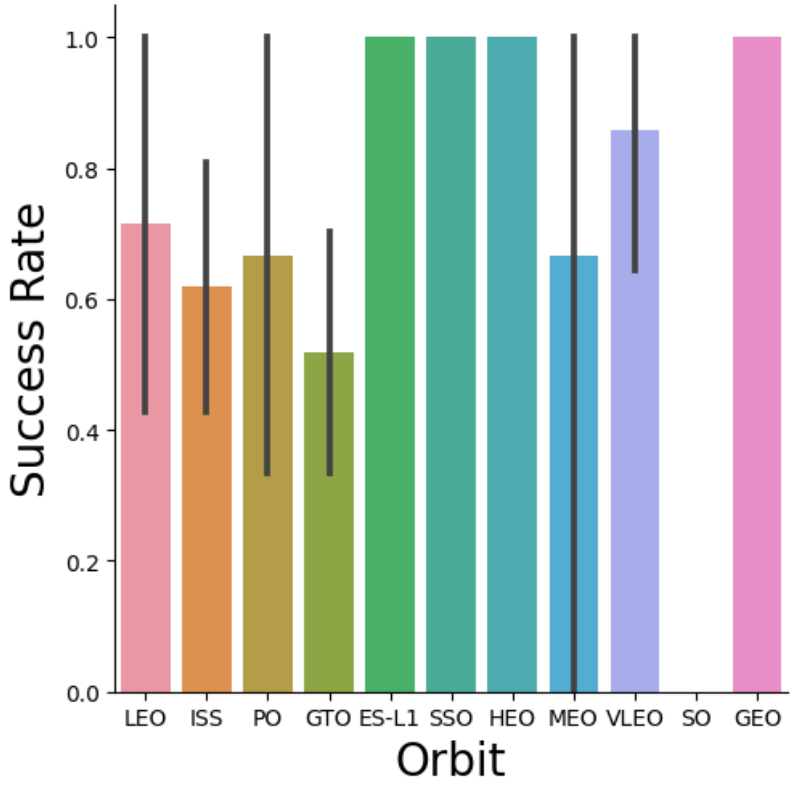
  ES-L1, GEO, HEO, SSO

- Orbits with 0% success rate:

  SO

- Orbits with success rate between 50% and 85%:

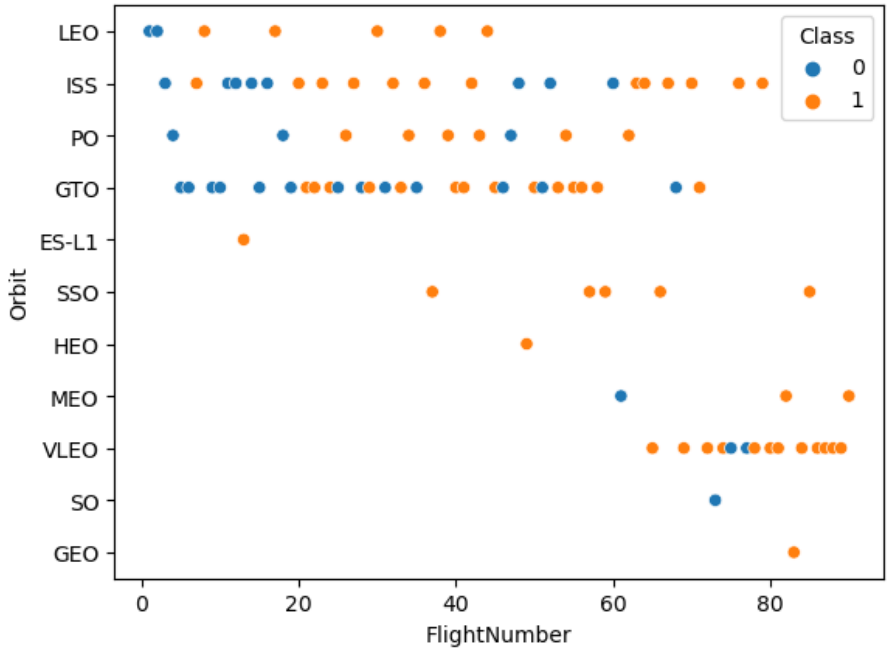
  GTO, ISS, LEO, MEO, PO



Flight Number vs. Orbit Type

Explanation:

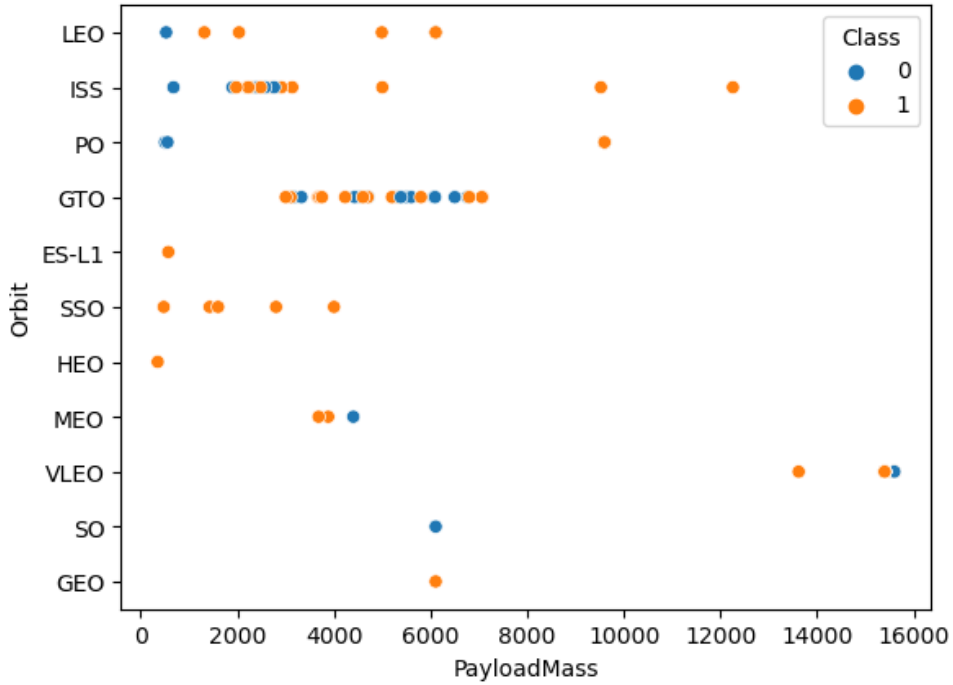
In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

Explanation:

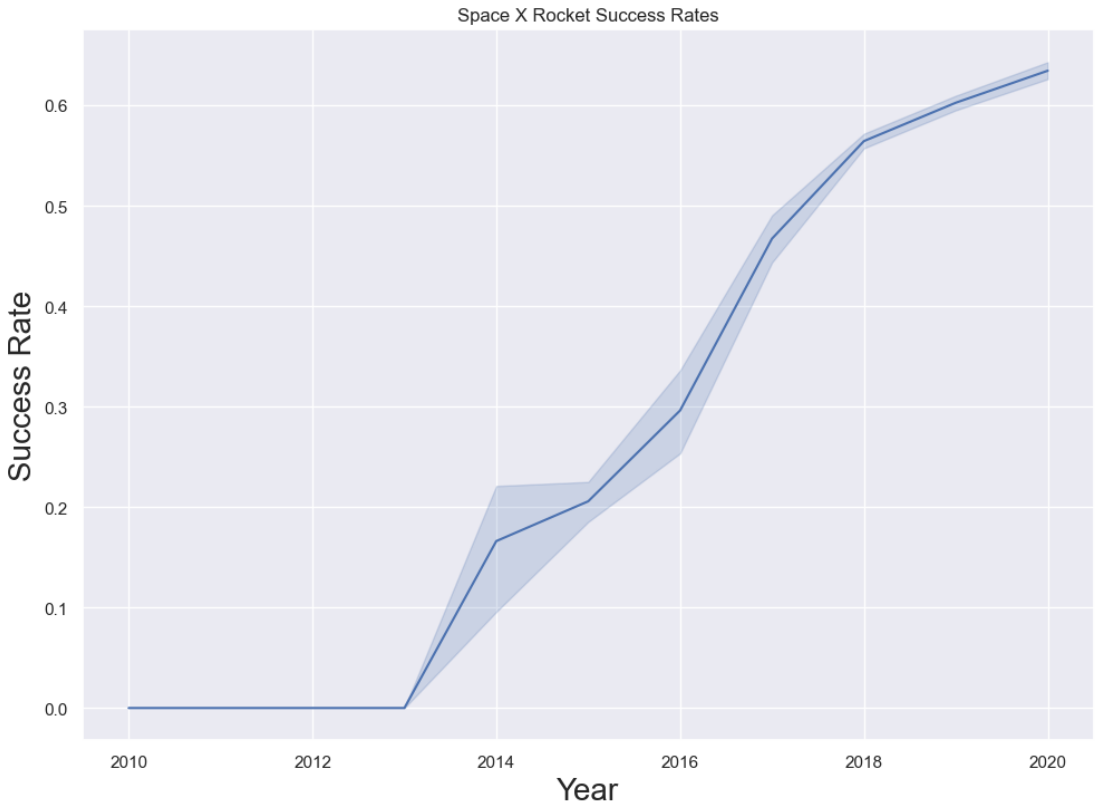
Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch Success Yearly Trend

Explanation:

The success rate since 2013 kept increasing till 2020

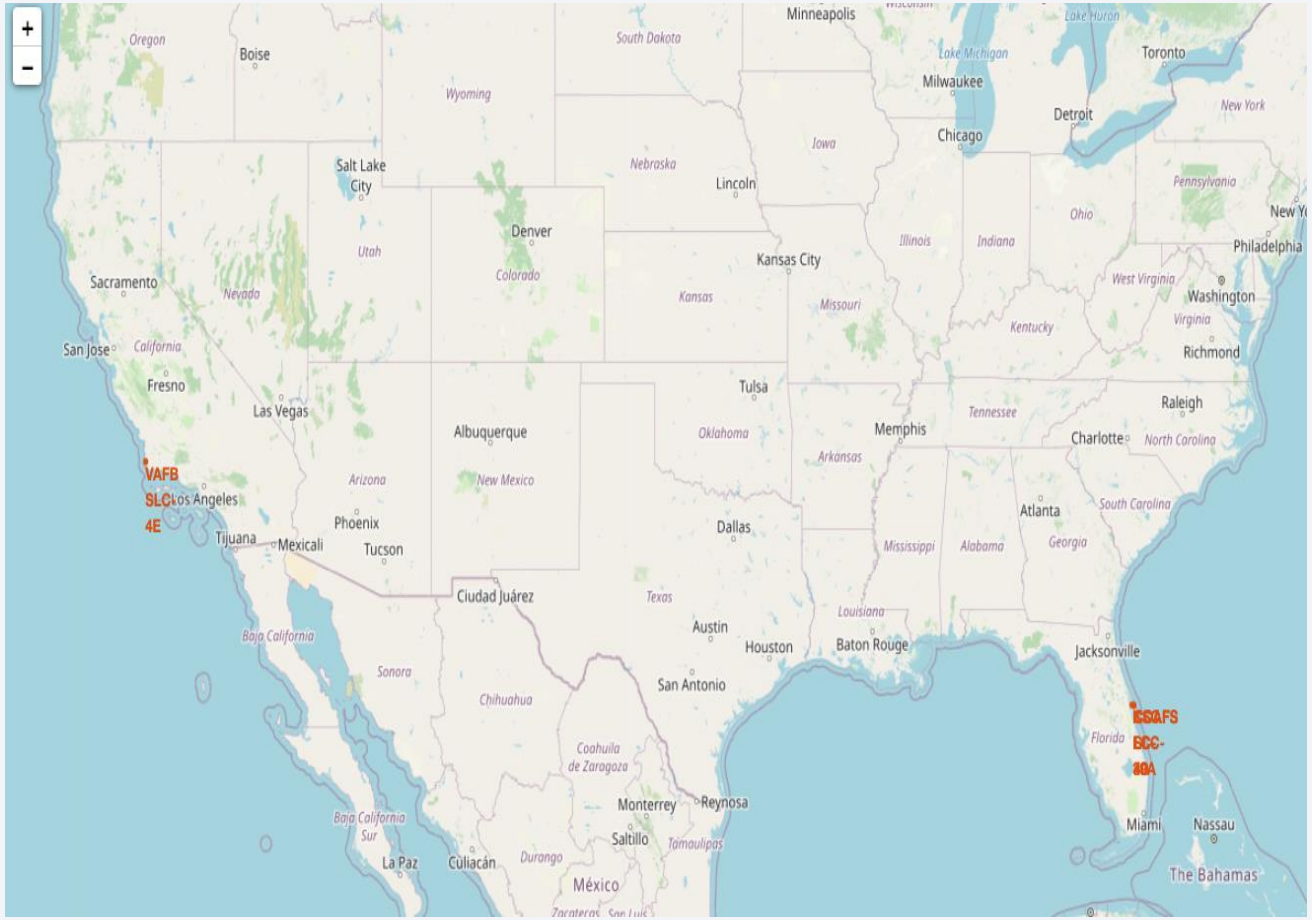



Section 3
Launch Sites
Proximities Analysis

All launch sites' location markers on a global map

- Explanation:

Most of Launch sites are in proximity to the Equator line. The land is moving faster at the equator than any other place on the surface of the Earth. Anything on the surface of the Earth at the equator is already moving at 1670 km/hour. If a ship is launched from the equator it goes up into space, and it is also moving around the Earth at the same speed it was moving before launching.



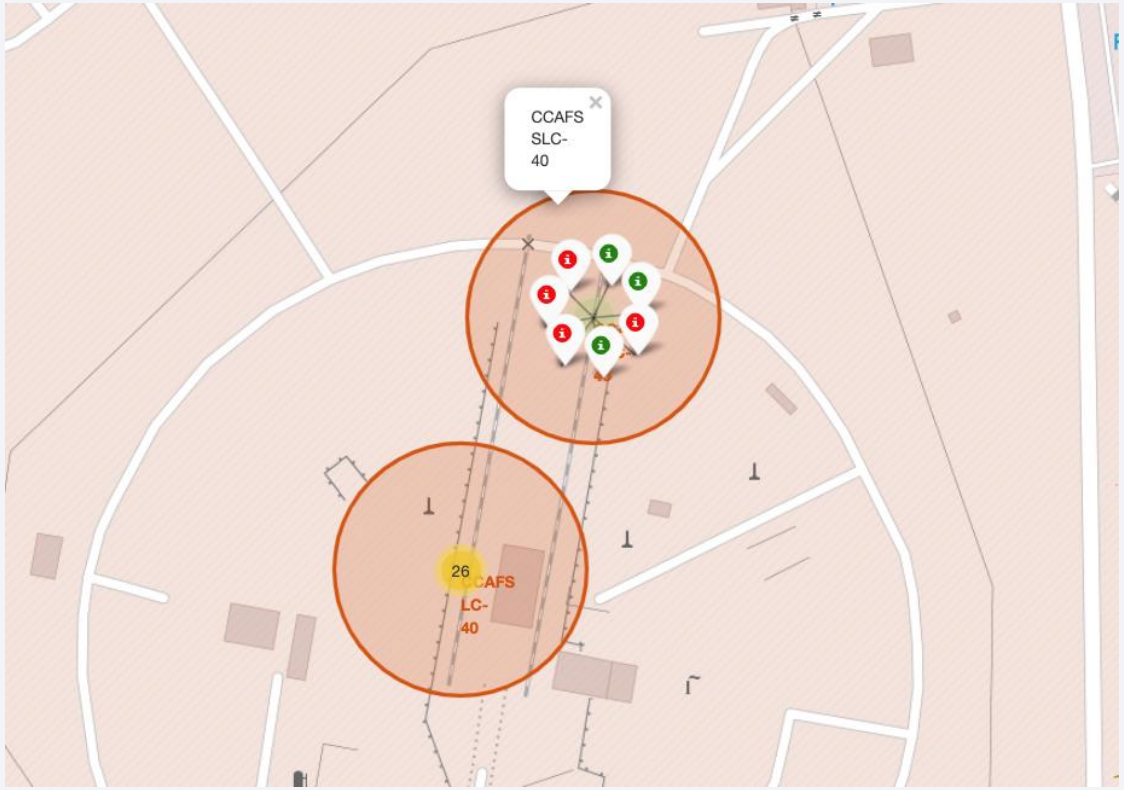
Color-labeled launch records on the map

Explanation:

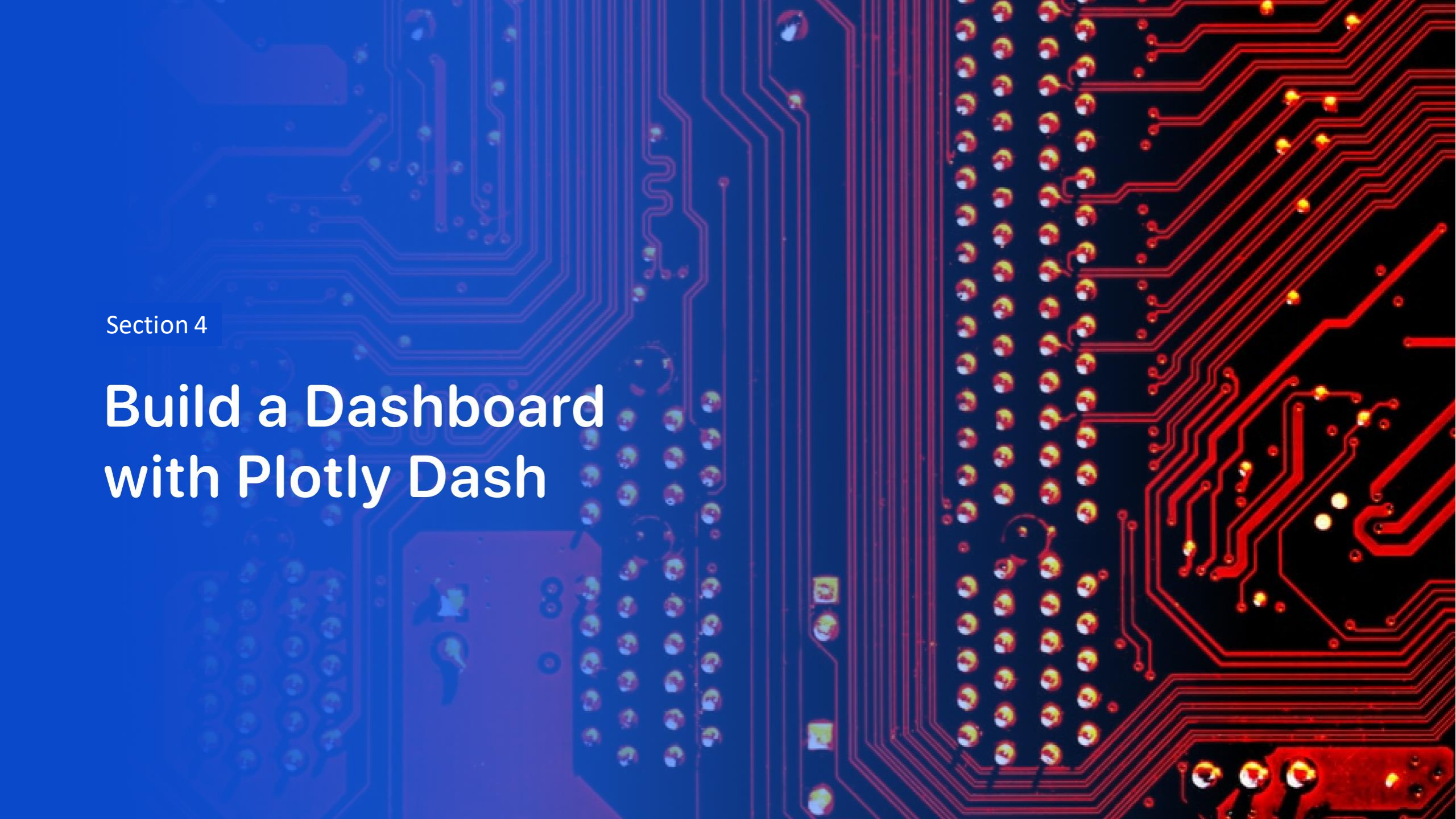
From the color-labeled markers we should be able to easily identify which launch sites have relatively high success rates.

Green Marker = Successful Launch

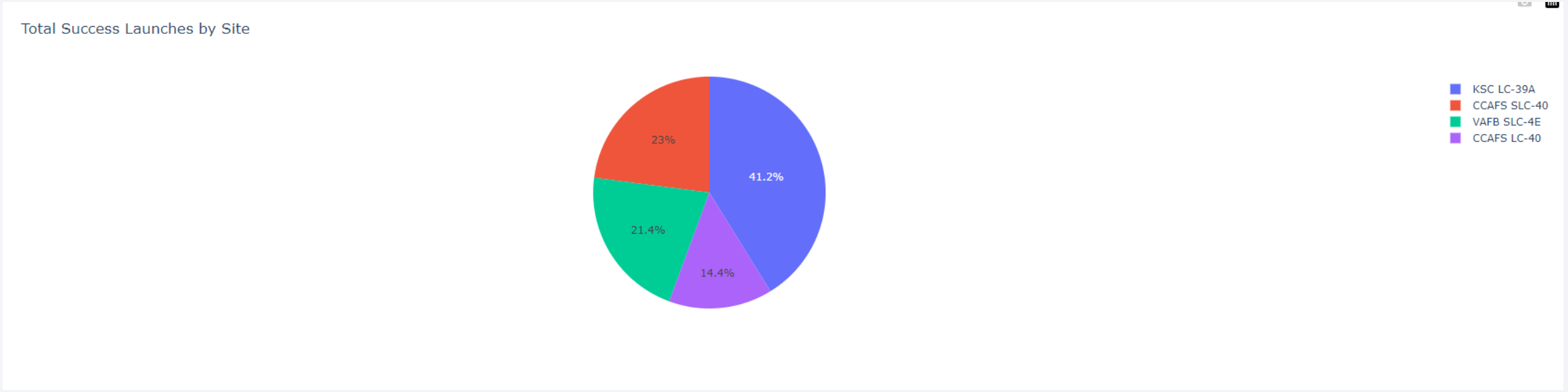
Red Marker = Failed Launch



- Launch Site KSC LC-39A has a very high Success Rate.


Section 4
Build a Dashboard with Plotly Dash

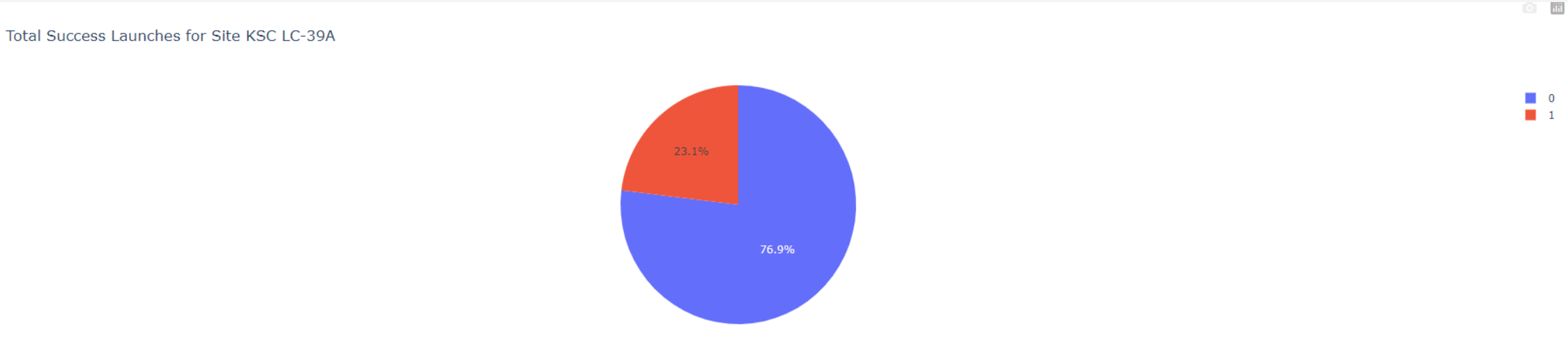
Launch success count for all sites



Explanation:

The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches.

Launch site with highest launch success ratio



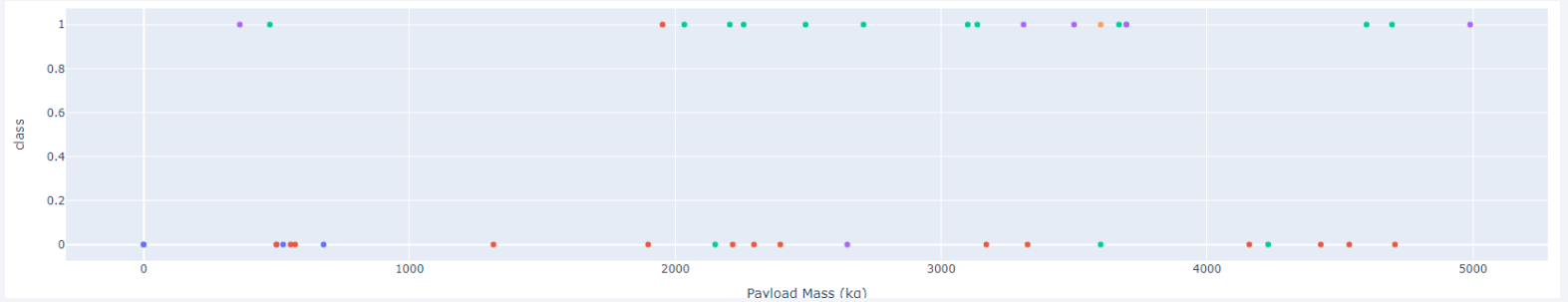
Explanation:

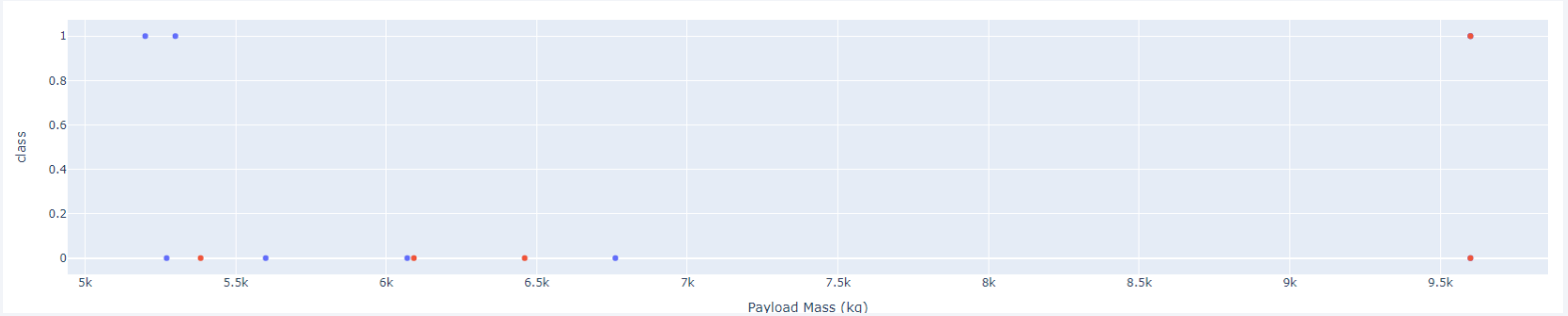
KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings.

Payload Mass vs. Launch Outcome for all sites

Explanation:

- The charts show that payloads between 2000 and 5500 kg have the highest success rate.





Section 5

Predictive Analysis (Classification)

Classification Accuracy

Explanation:

Based on the scores of the Test Set, we can not confirm which method performs best.

Same Test Set scores may be due to the small test sample size (18 samples). Therefore, we tested all methods based on the whole Dataset.

The scores of the whole Dataset confirm that the best model is the Decision Tree Model. This model has not only higher scores, but also the highest accuracy.

Classification Accuracy

Explanation:

Based on the scores of the Test Set, we cannot confirm which method performs best.

Same Test Set scores may be due to the small test sample size (18 samples). Therefore, we tested all methods based on the whole Dataset.

The scores of the whole Dataset confirm that the best model is the Decision Tree Model. This model has not only higher scores, but also the highest accuracy.

Scores and Accuracy of the Test Set

	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

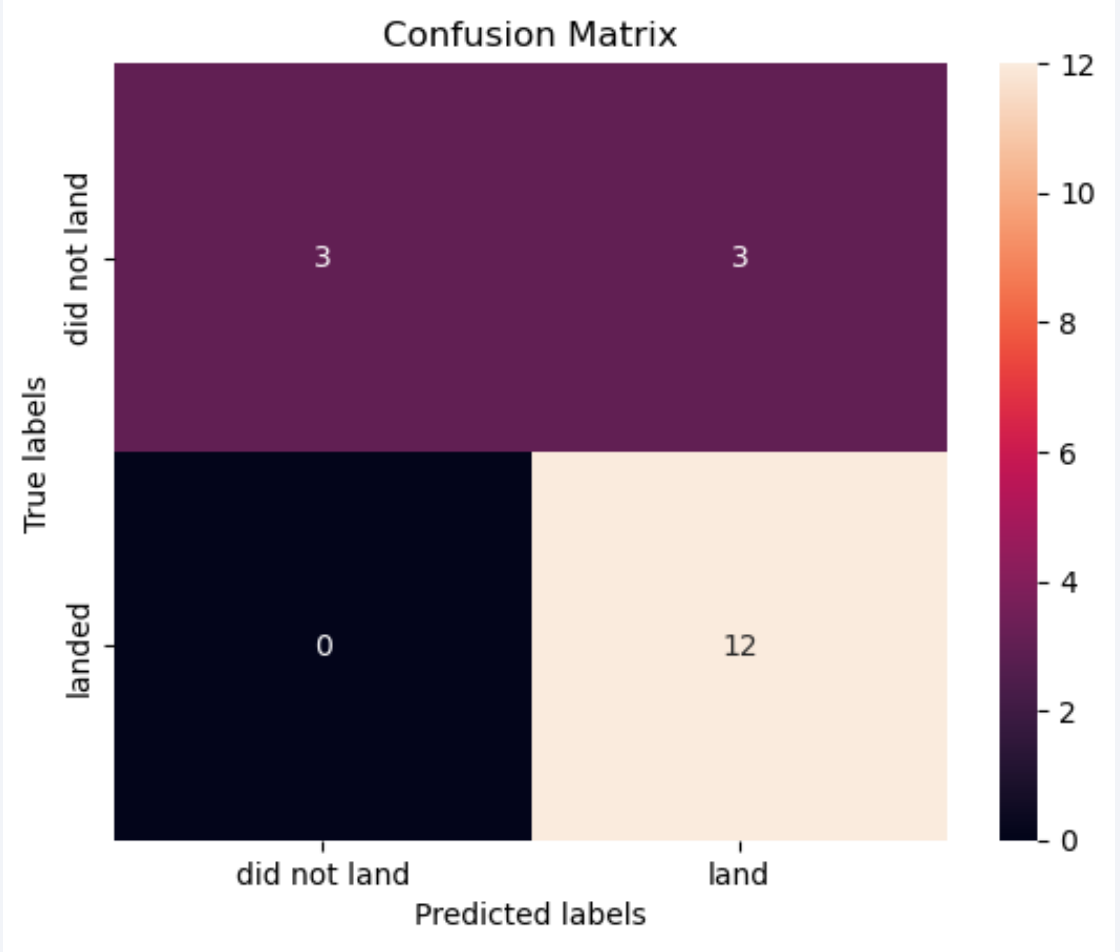
Scores and Accuracy of the Entire Data Set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

Confusion Matrix

Explanation:

- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.



Conclusions

- Decision Tree Model is the best algorithm for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.

Appendix

Special thanks to Instructors, Coursera and IBM

Instructors

Coursera

IBM

Thank you!