IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Sam Cubberly>
<7/17/2024>

Outline

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

Executive Summary

- Summary of methodologies
  - Used Pandas Library to load data
  - Scraped the web with pandas, requests, and BeautifulSoup
  - Used SpaceX API to get more data
  - Used Seaborn and Folium to represent data
  - Used Dash to create a results dashboard
  - Sklearn to determine whether the first stage will be reused
- Summary of all results
  - Determined it is possible to predict first stage reusibility

Introduction

- SpaceY is a rocket company
- They are competing for space travel with SpaceX
- The goal of the case study is to estimate the price of rocket launches, depending on multiple factors, such as:
  - Previous launch outcomes
  - Booster types
  - Launch location
  - Much more

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

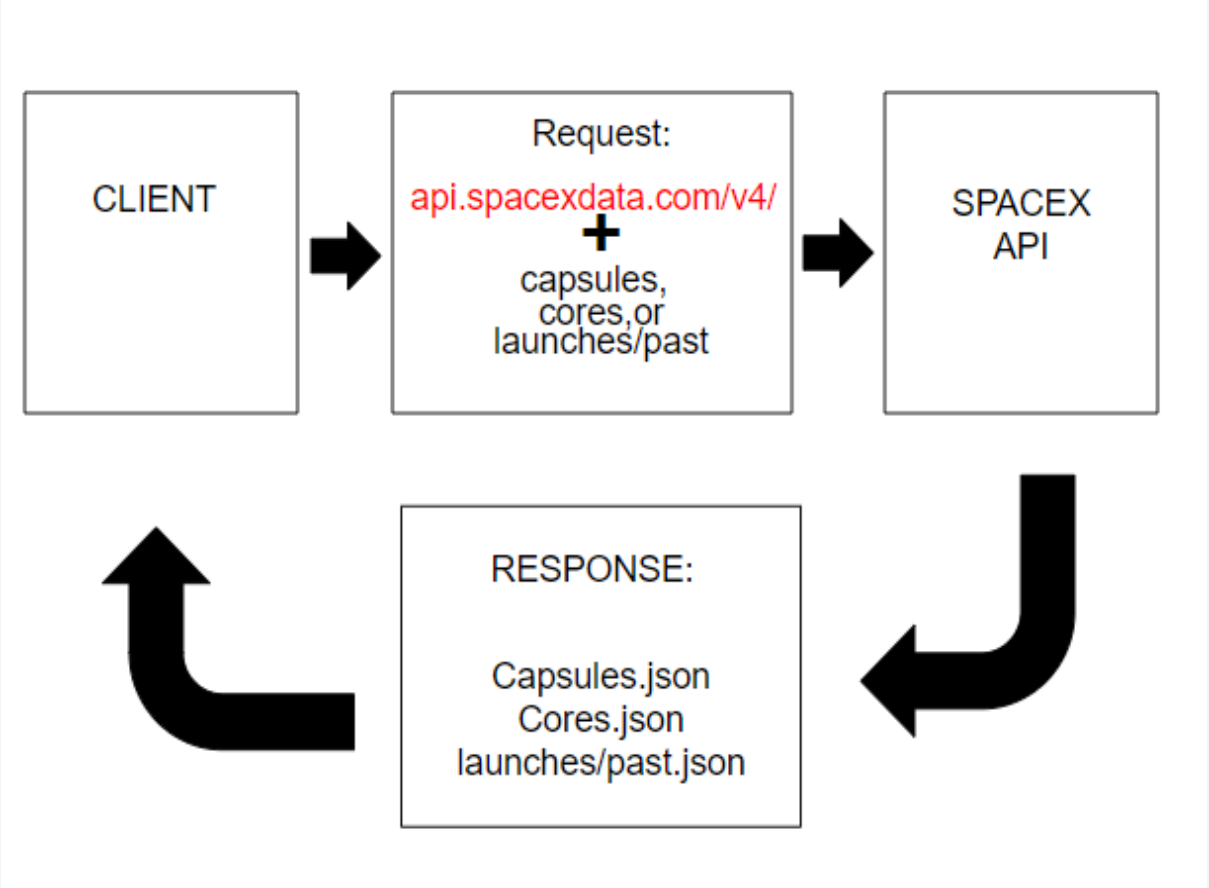
Data Collection

- Describes how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Used the Requests library to request one of three JSON files from the API
- Used Pandas libraries to convert JSON to dataframes
- Transformed the data to a more useable format

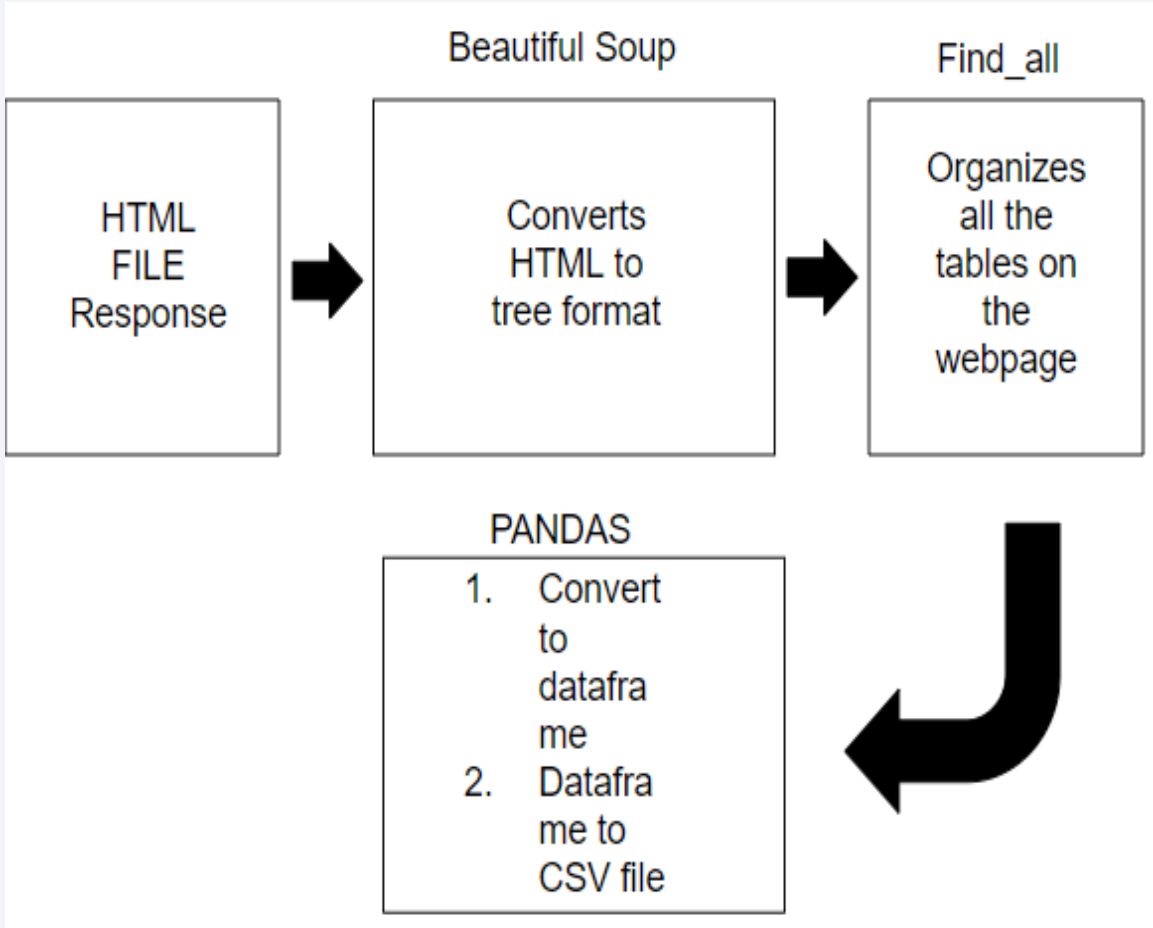
- GitHub URL: https://github.com/SamCubz/CaseStudy/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Scraped the HTML response using BeautifulSoup
- Converted reults into a CSV file

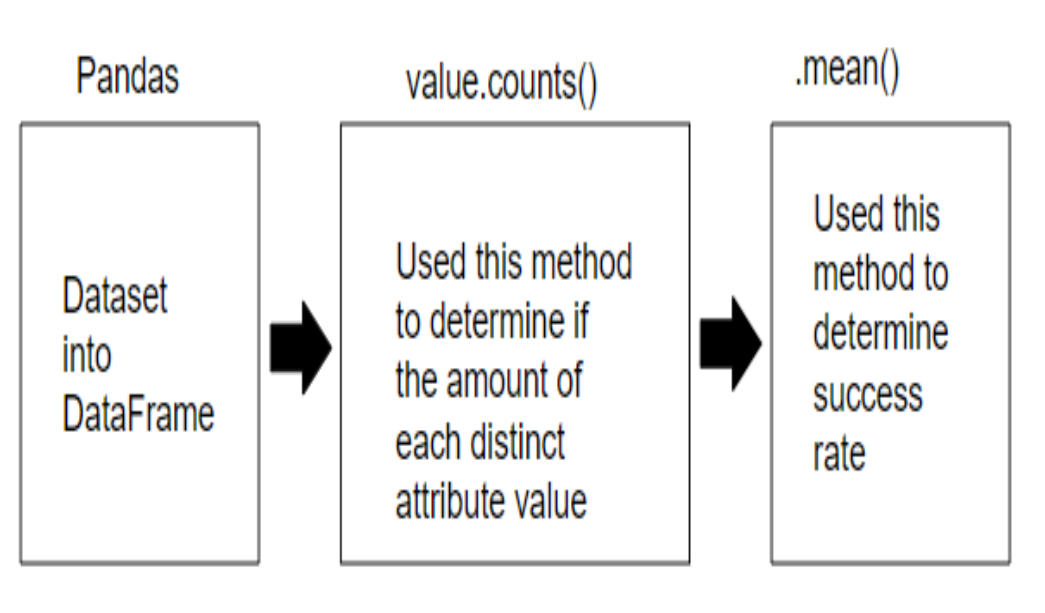
- GitHub URL: https://github.com/SamCubz/CaseStudy/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- The Dataset was processed using pandas library.
- Pandas Library used to determine the values and counts for each attribute in the dataset
- Determining the rates of success using .mean() function

- GITHUB URL: https://github.com/SamCubz/CaseStudy/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Seaborn scatterplots were used whenever Flight Number and Payload Mass were the independent variable.
    - This is because it maps change over a set of values.
- Seaborn bar plots were used whenever the independent variable was categorical, such as orbit types.
- GITHUB
  URL: https://github.com/SamCubz/CaseStudy/blob/main/edadataviz%20(1).ipynb

EDA with SQL

- Queries
  - Selected distinct launch sites
  - Found launch sites with similar prefixes
  - Selected total payload mass
  - Selected average payload mass
  - Determined first successful landing date
  - Filtered booster versions by success and mass
  - Compared successes vs failures
  - Ranked the successes
- GITHUB URL:
- https://github.com/SamCubz/CaseStudy/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb

Build an Interactive Map with Folium

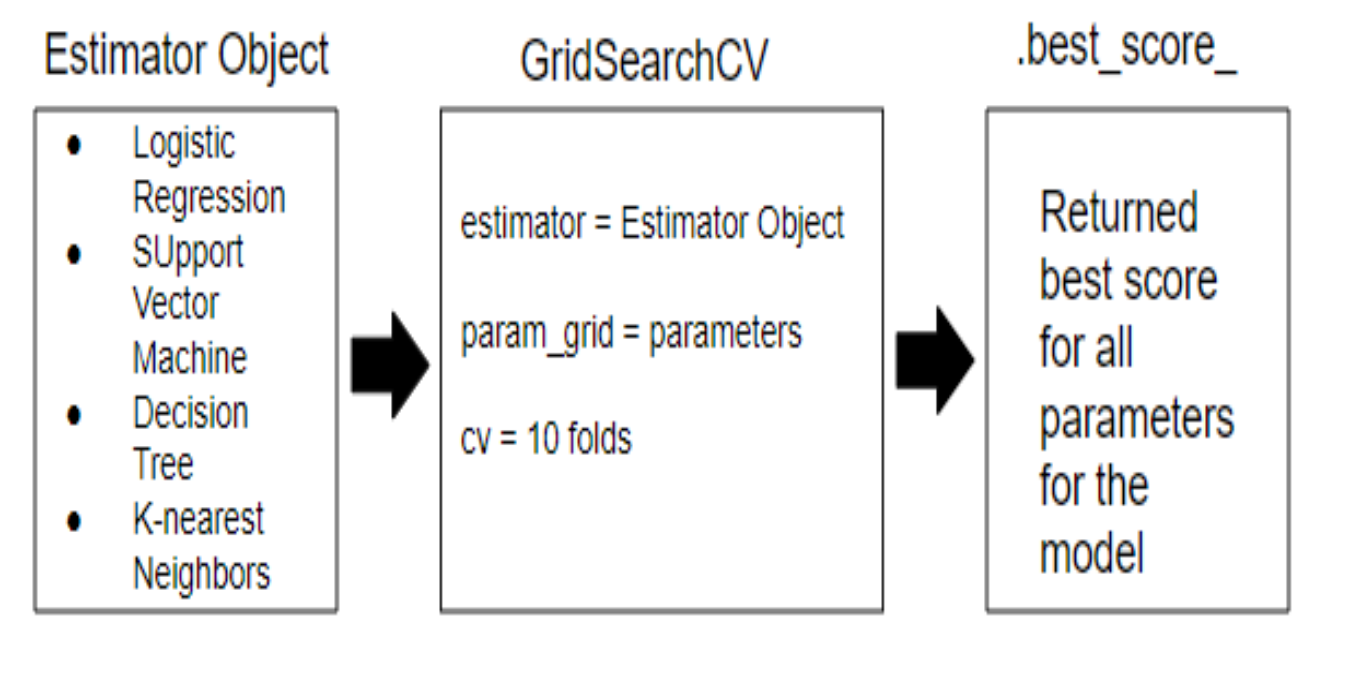
- Created Markers / Labels
  - Marked the names of the launch sites
- Created Circles
  - To help visualize the area that the launch site is contained in
- Lines
  - Helped to determine the distances between the shore and a launch site
  - Helped to determine the distance between the launch site and a city
- GITHUB
  URL: https://github.com/SamCubz/CaseStudy/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Pie Chat Showcasing Launch Sites and Their Respective Success Proportion
  - This helped determine where a rocket is the most likely to have successfully launched from
  - The user can select either ALL sites, or a specific site to determine successes
    - Helped aid in the discovery of the dataset
    - A dropdown was used to help with this
- A slider and a scatter chart were added for Payload Mass vs the Class
  - The slider helped set the range of the payload masses for viewing
  - The scatter chart was used to show how the payload mass affected the class at different masses
- GITHUB
  URL: https://github.com/SamCubz/CaseStudy/blob/main/spacex_dash_app.py

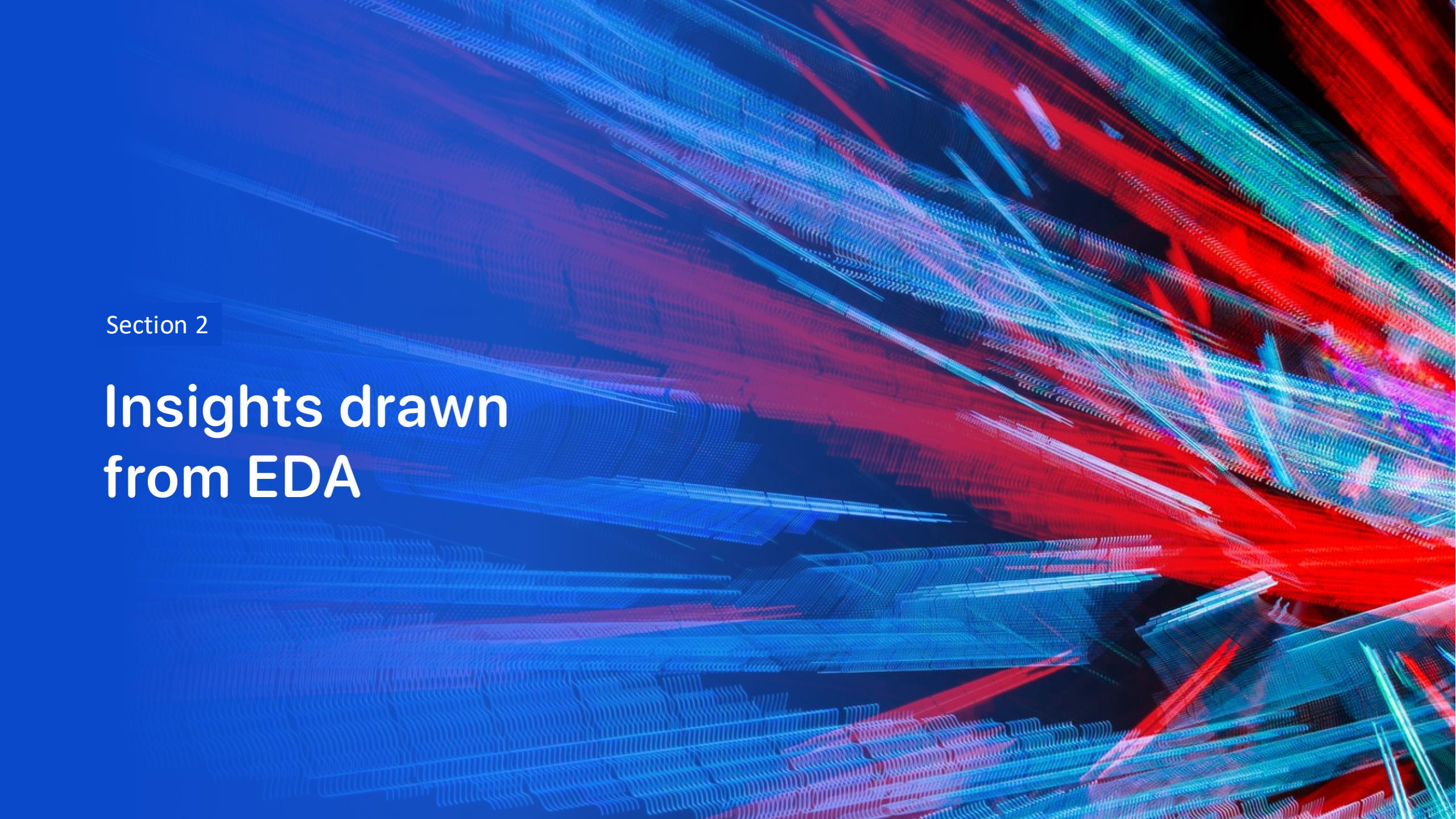
Predictive Analysis (Classification)

- Used Grid Search to determine best parameters for many different classification models
- Determined the highest score through cross-validation
- GitHub URL: https://github.com/SamCubz/CaseStudy/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(2).ipynb



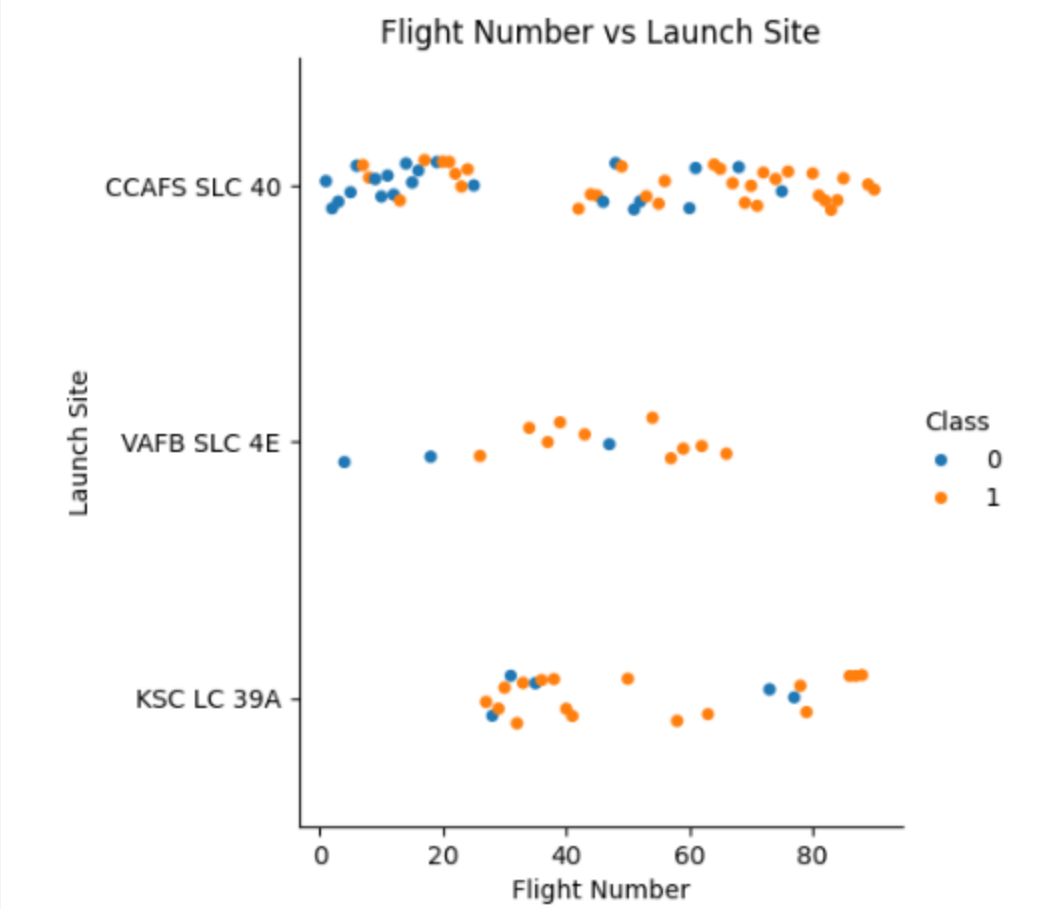
Results

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


Section 2
Insights drawn from EDA

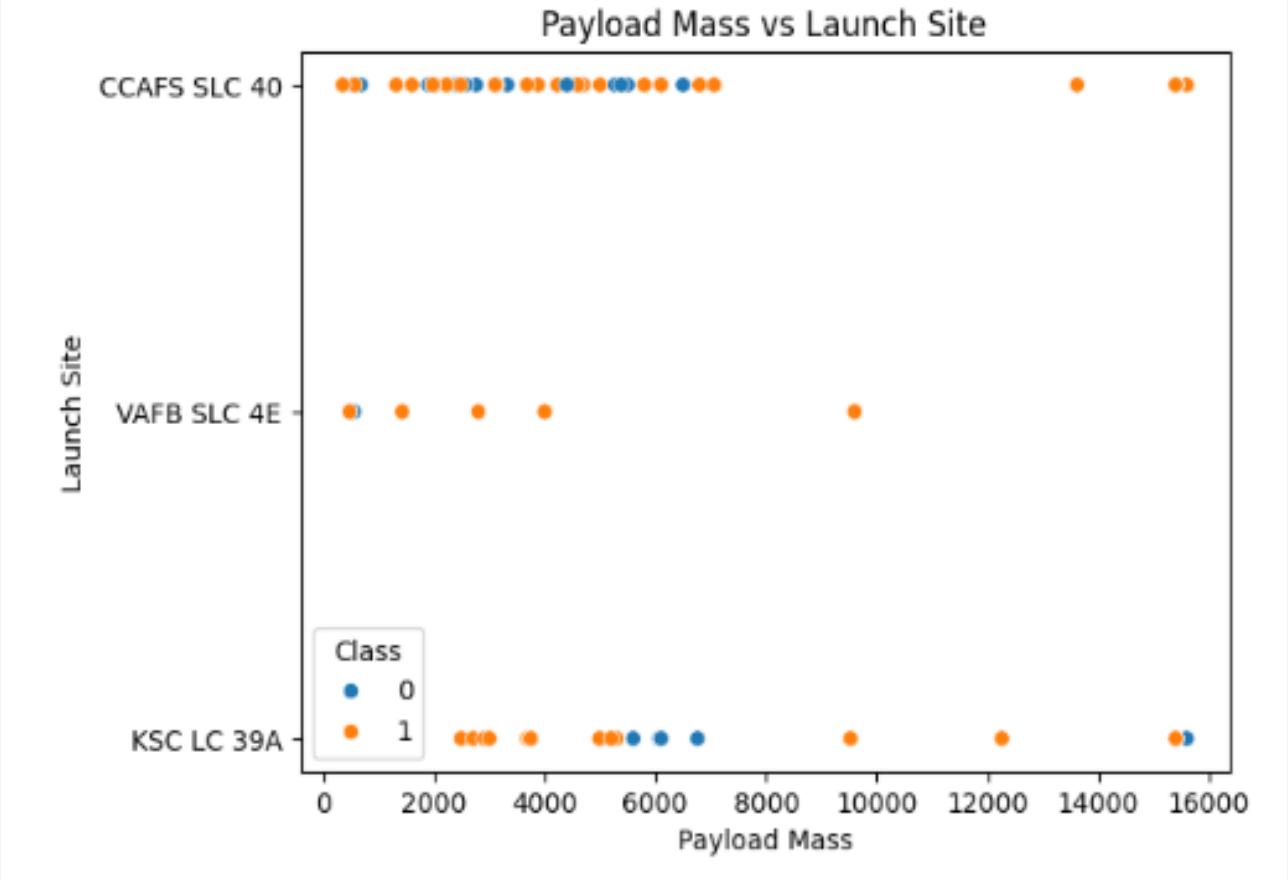
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site
  - This shows the flight number on the x axis, and the launch site on the y axis
  - The class is shown by the color
  - For example, as flight number increases, the class of VAFBSLC4E appears to head toward 1, which means that more flights makes the flights more likely to succeed



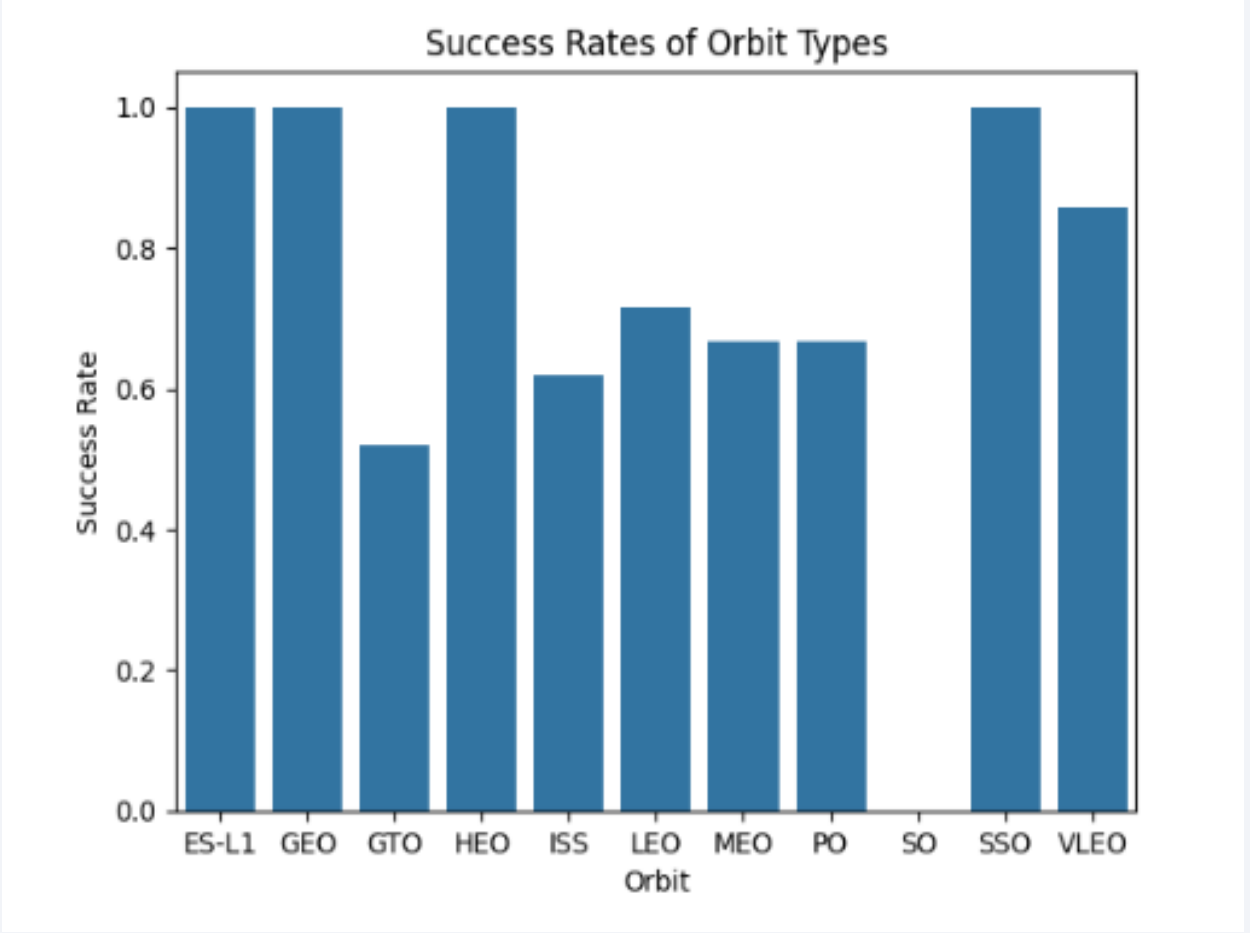
Payload vs. Launch Site

- Scatter Point chart of Payload Mass vs Launch Site
  - Payload Mass is independent variable
  - Launch site is the dependent variable
  - If appears that as the mass increases, the amount of VAFB SLC 4E launches decreases



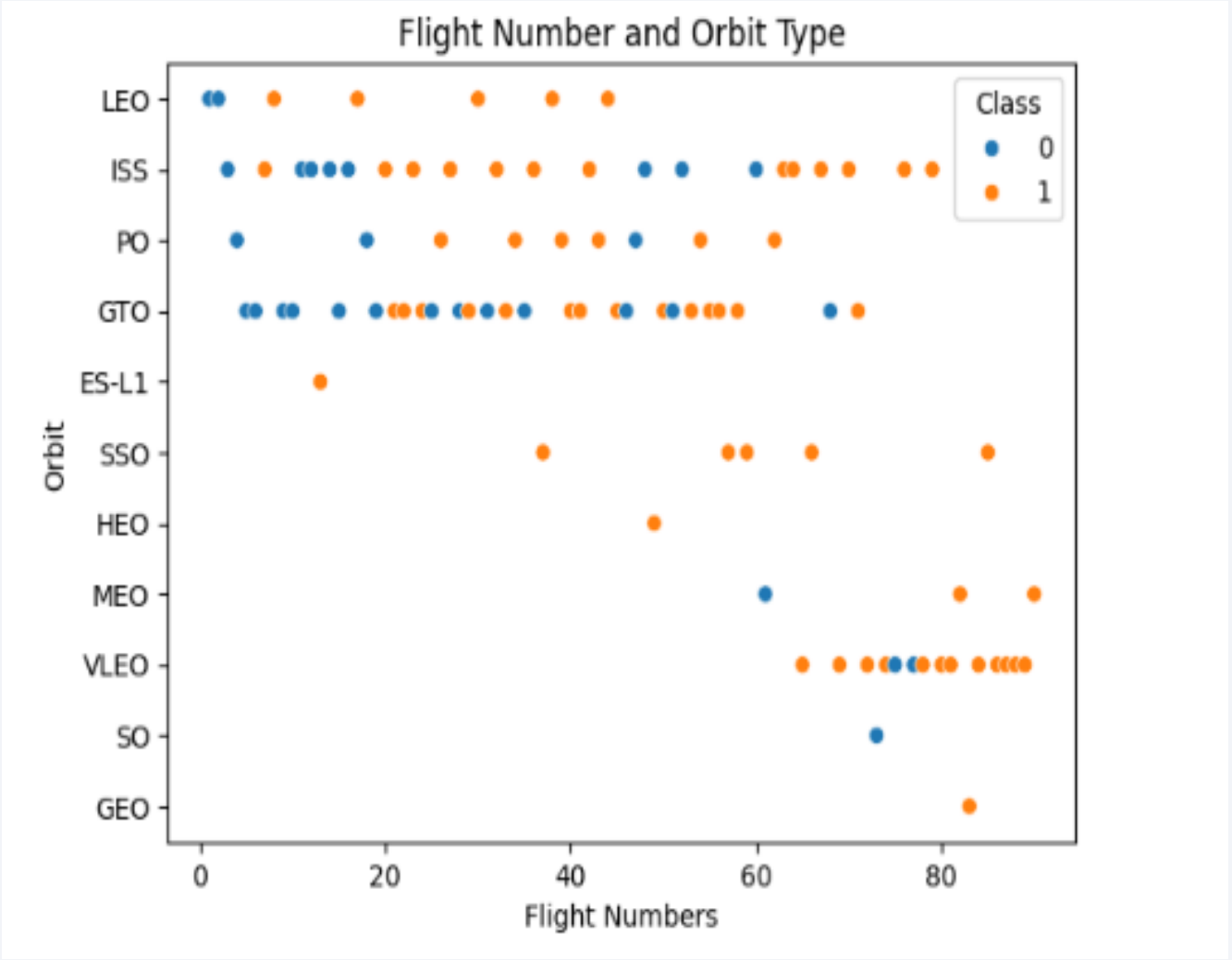
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
  - Orbit type on the x axis
  - Success Rate on the Y axis
  - SO has the worst success rates, while ES-L1, GEO, HEO, and SSO all have the same, 100% success rate



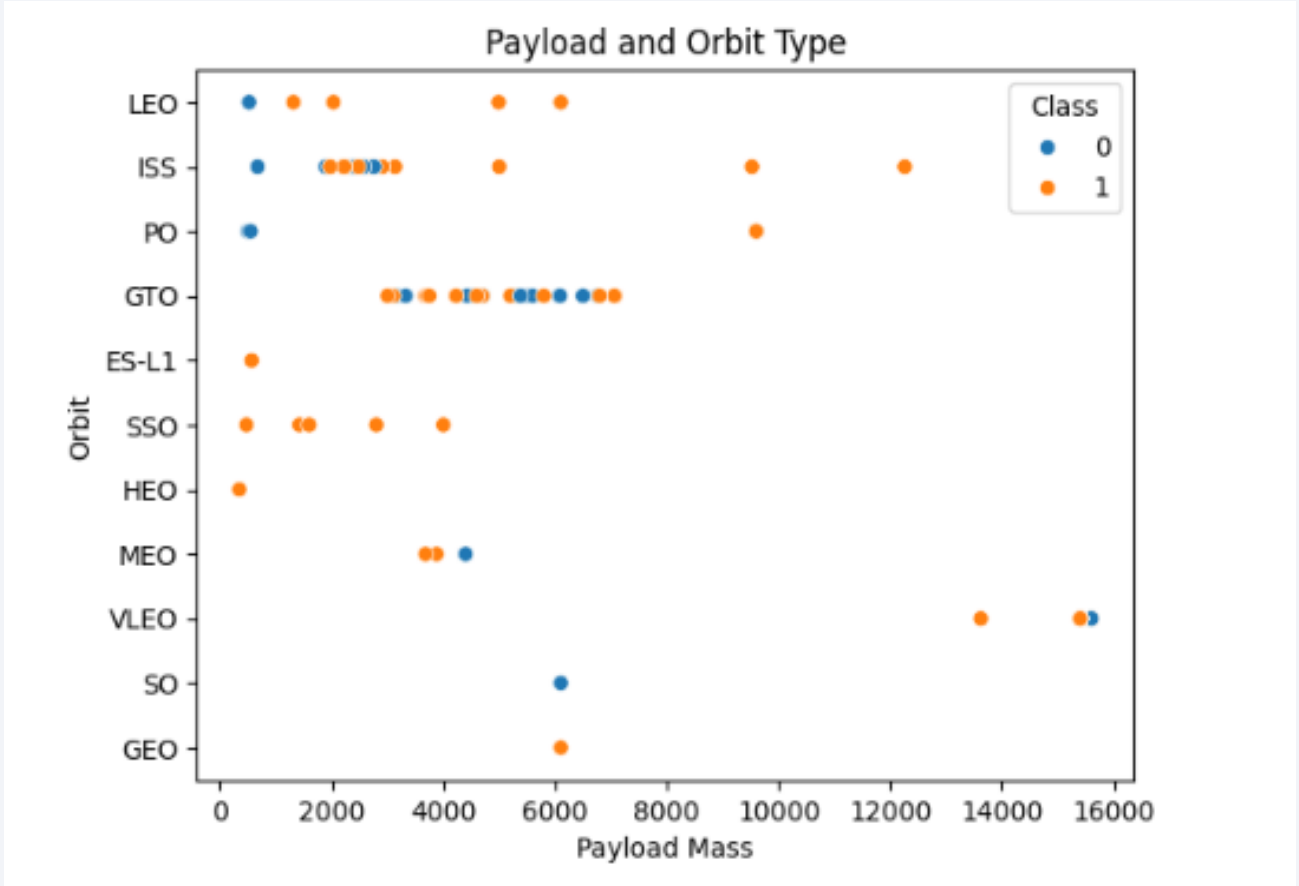
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type
  - Flight Numbers on the X axis
  - Orbit on the Y axis,
  - The color determines if the flight was a success or not
  - LEO appears to be more successful as the flight number increases, while GTO appears to have random failure throughout all flight numbers.



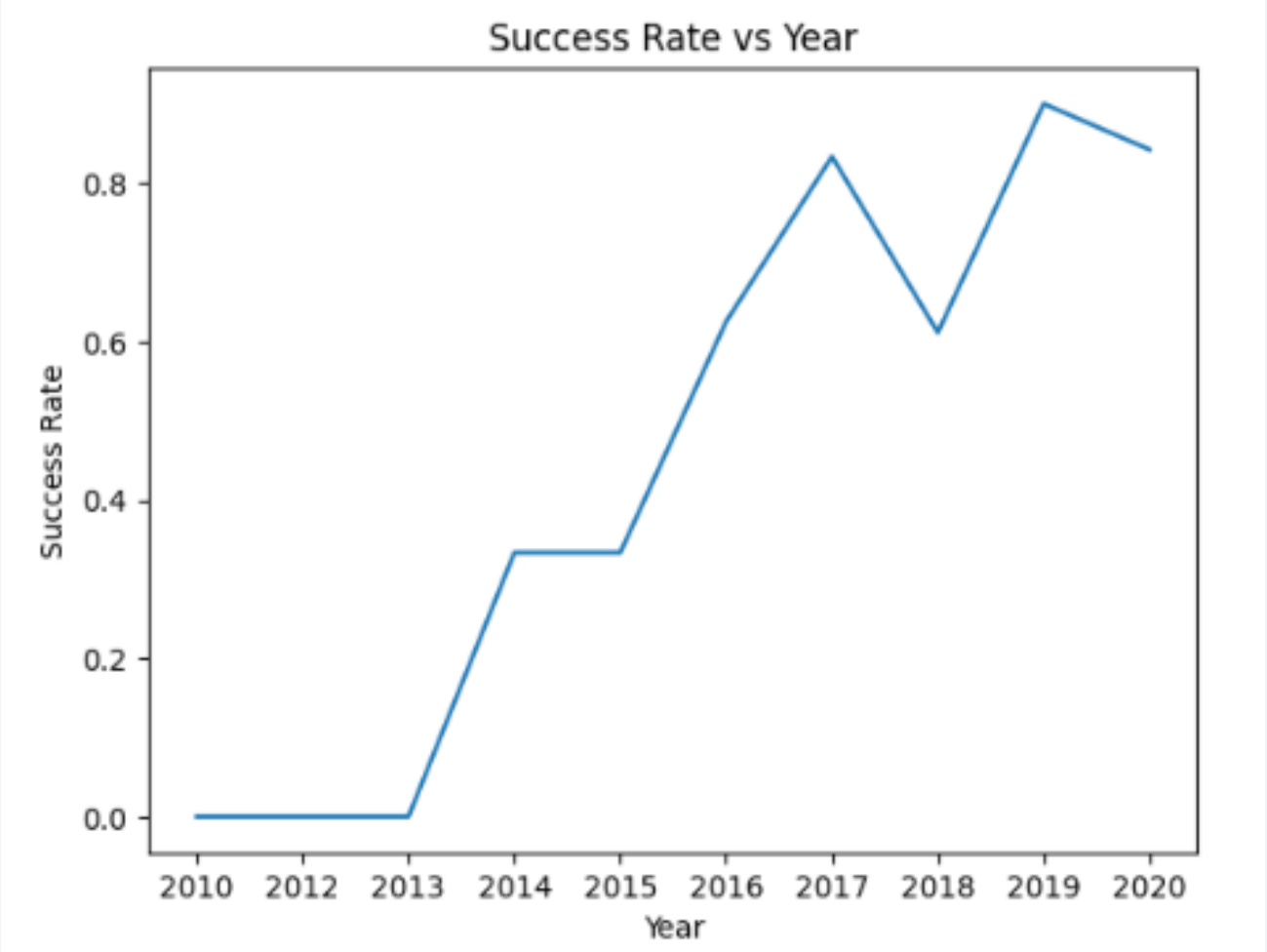
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type
  - Payload mass as independent variable
  - Orbit as the dependent variable
  - Color represents if it was a success or not
  - As payload mass increases going into ISS orbit, the number of successes increases. In GTO, it appears to be random regardless of mass



Launch Success Yearly Trend

- Line chart of yearly average success rate
  - Independent Variable is the year
  - Dependent variable is the success rate
  - It appears that as the years go on, the rates of success skyrocket



All Launch Site Names

- The query shows 4 distinct launch sites
- These are listed in the Launch Site table

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

Launch Site Names Begin with 'CCA'

- These are 5 of the records with a launch site beginning in "CCA"
- They are all F9 Boosters
- They all go into LEO orbit

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 carried by boosters from NASA is displayed below

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is displayed below

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- This represents the date of the first successful landing outcome on ground pad

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Below is a list of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

done.

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

Total Number of Successful and Failure Mission Outcomes

- Below is the total number of successes and failures of the missions

Successes	Failures
100	1

Boosters Carried Maximum Payload

- Below is a list of 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

- Below is a list of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

- Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	COUNT(*)
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Section 3

Launch Sites Proximities Analysis

Global Launch Sites

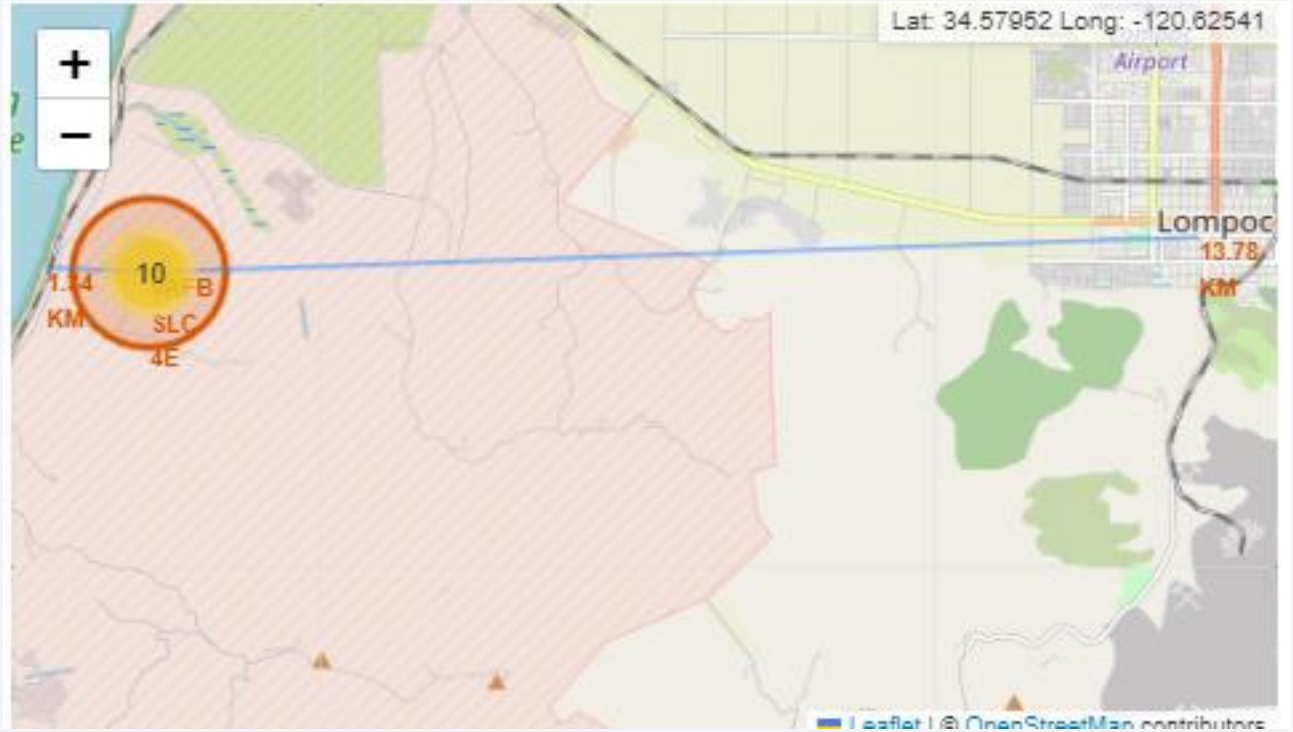
- The SpaceX launch sites are in the United States
  - They are in the southern part, nearest the equator
  - They are near the coastline

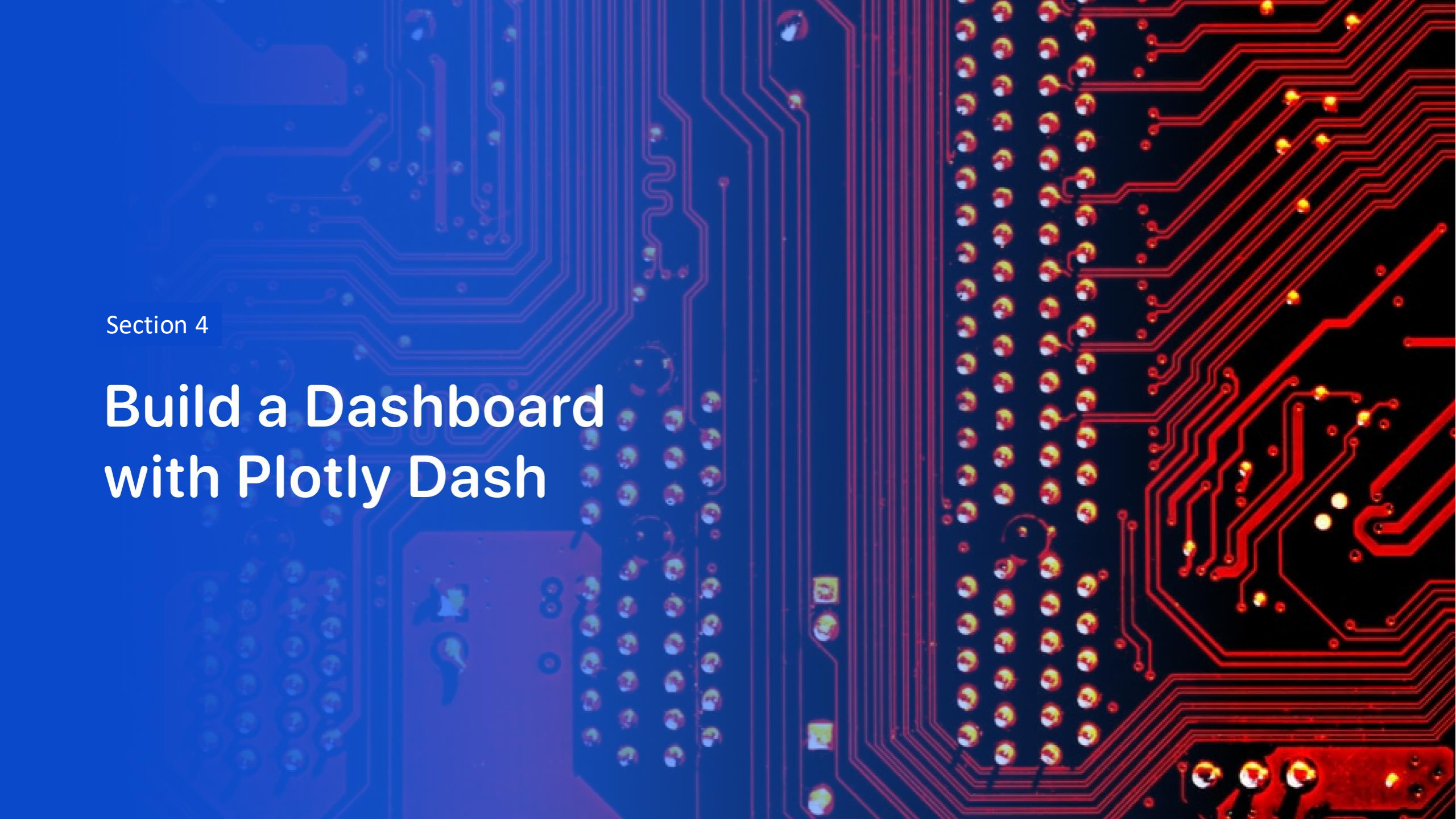
Color-Labeled Outcome of Launches

- You can see that there are far more successful landings on the east coast
  - There are 46 on the launch sites in Florida
  - There are 10 in California

California Launches Proximities

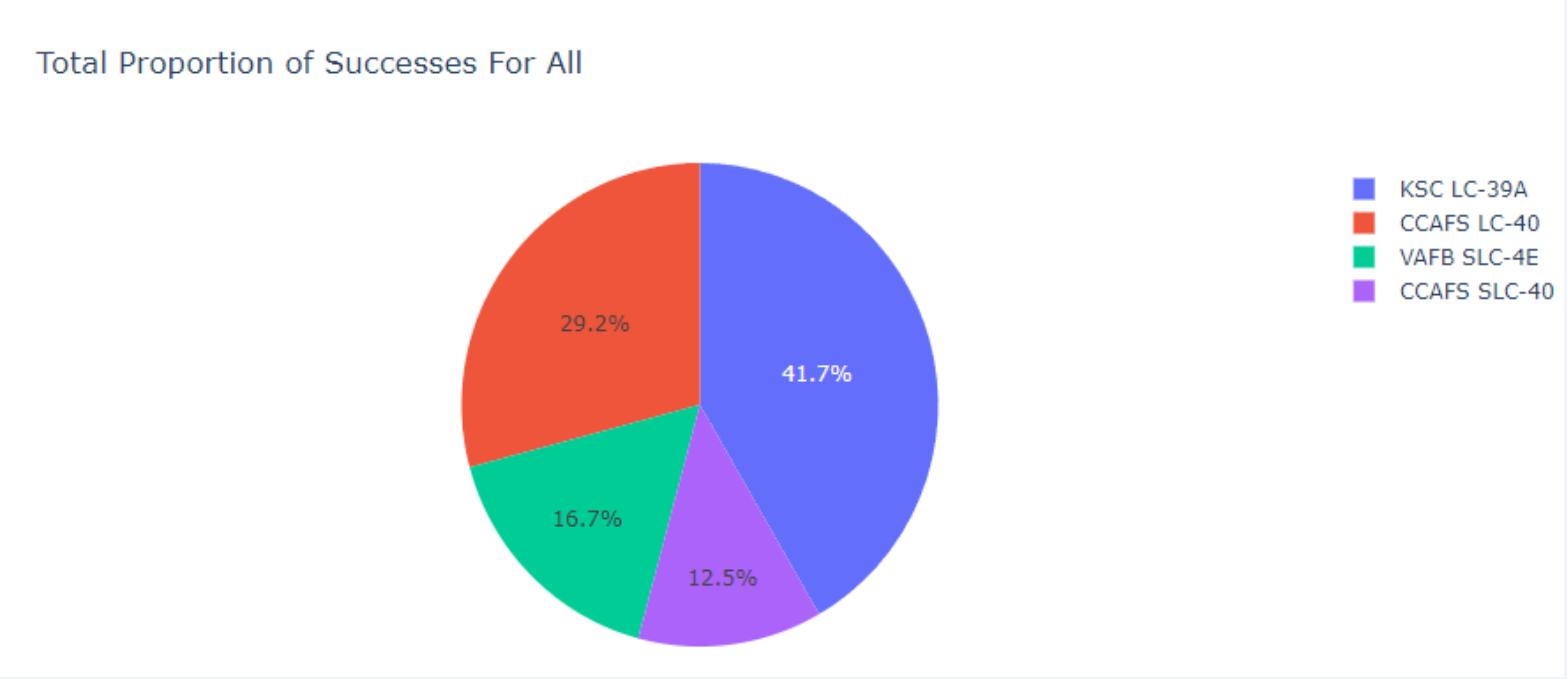
- The launch site is considerable farther away from the nearest city, as it is from the nearest shore




Section 4
Build a Dashboard with Plotly Dash

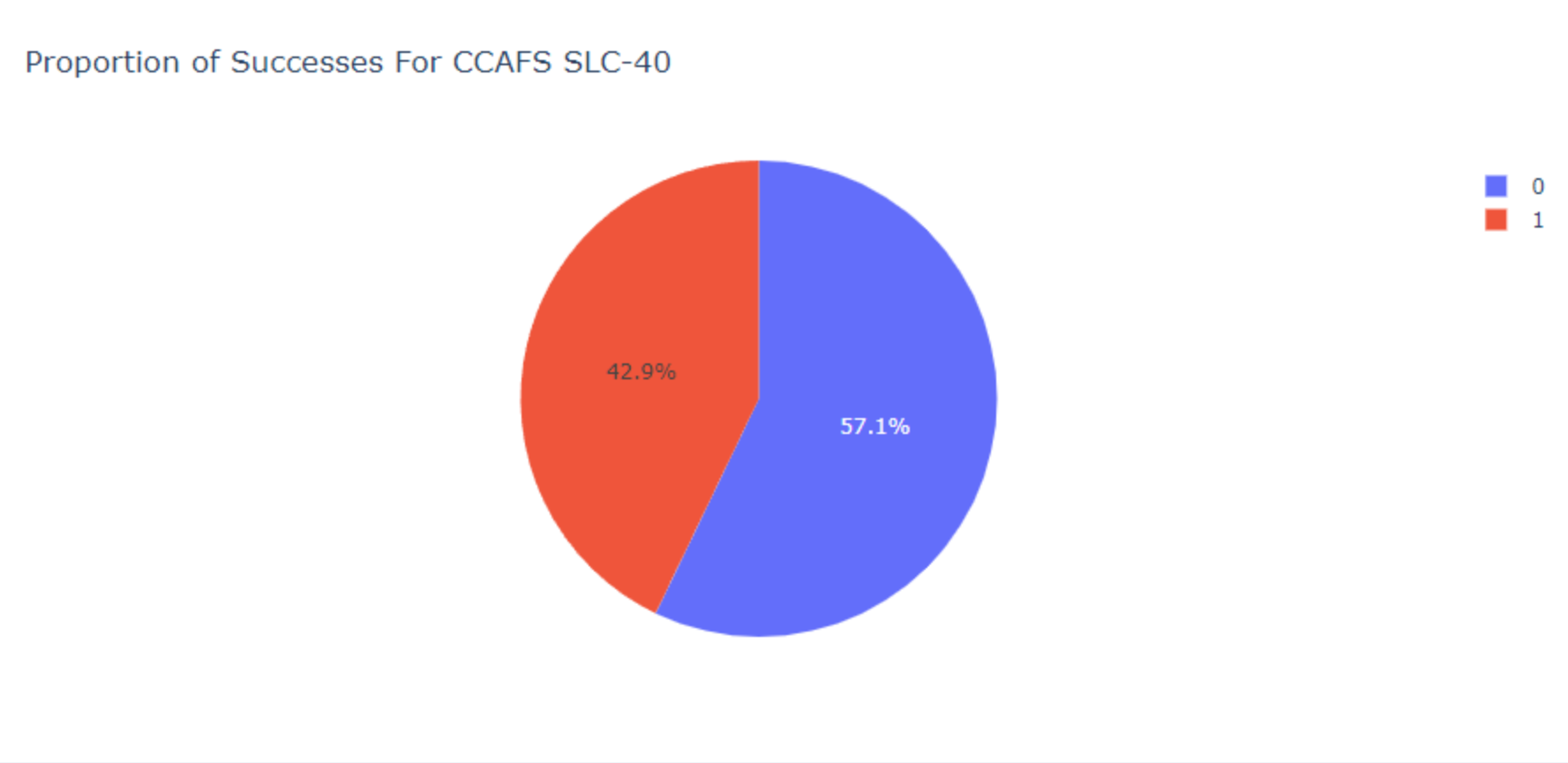
Launch Success Counts For All Sites Pie Chart

- KSC LC-39A had the highest proportion of success
- CCAFS LC-40 had the second highest proportion of success
- VAFB SLC-4E had the third highest proportion of success
- CCAFS SLC-40 was the lowest



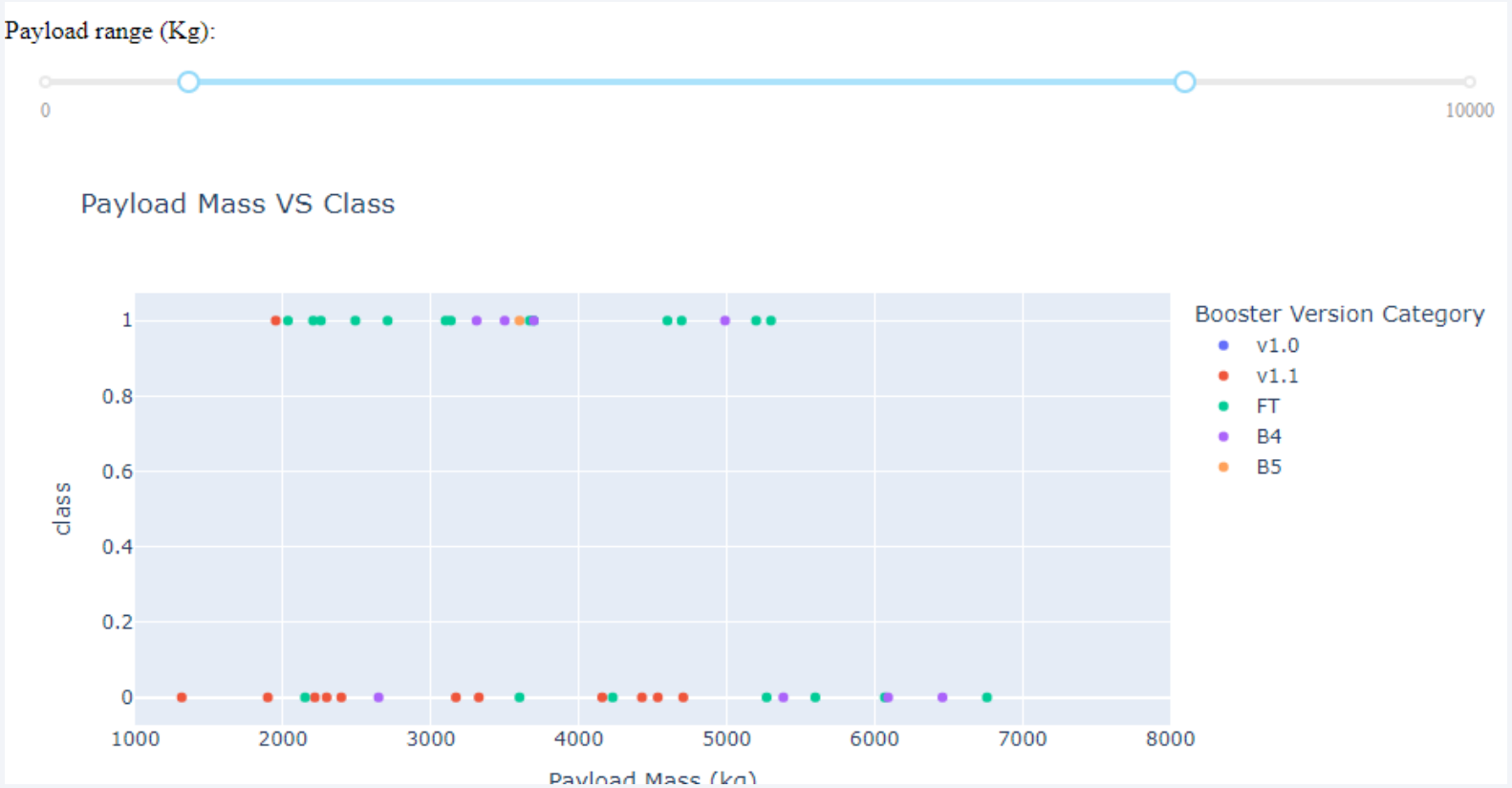
Highest Launch Success Ratio

- For CCAFS SLC-40, 42.9% of landings were a success, while 57.1% were failures



Payload vs. Launch Outcome For All Sites

- As the payload increased, the amount of successes goes down



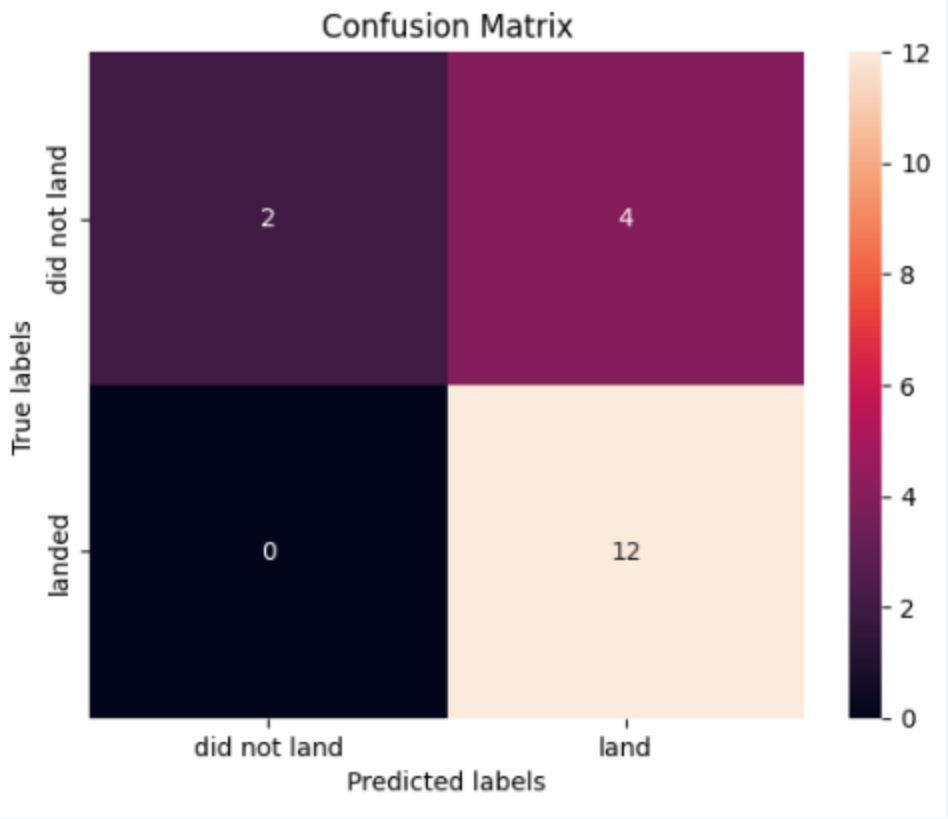
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Ranking from highest best score accuracy to lowest
  - Decision Tree - 91.6%
  - K – Nearest Neighbors - 84.8%
  - Support Vector Machine - 84.8%
  - Logistic Regression - 84.6%

Confusion Matrix – Decision Tree Classifier



Conclusions

- K-nearest neighbor, Support Vector Machine, and Logistic Regression all seem to give nearly identical results.
- The decision Tree classifier was the most effective, and should be used for the model.
- The coefficient of determination was high enough to determine that it is possible to determine the likelihood of the first stage landing.
- By determining if the first stage will land, it is now very likely that the price can be determined.

Thank you!