IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jacob Hurley

May 25, 2025

Outline

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

Executive Summary

- In This project we perform Data collection through the use of REST API's and webscraping, data wrangling, Exploratory Data Analysis using matplotlib, seaborn for static visualizations, folium and plotly for dynamic visualizations, and scikit-learn machine learning models for predictions.
- The Decision Tree Classifier performed the best on the test data with an accuracy of about 94%

Introduction

- The Goal of this project is to determine, given some predictors, whether a rocket will have a successful landing when returning from space.
- Some of the ideas that we would like to answer are which predictors are most important for determining the results of the landing and how accurately can we predict the landing results.

Section 1

Methodology

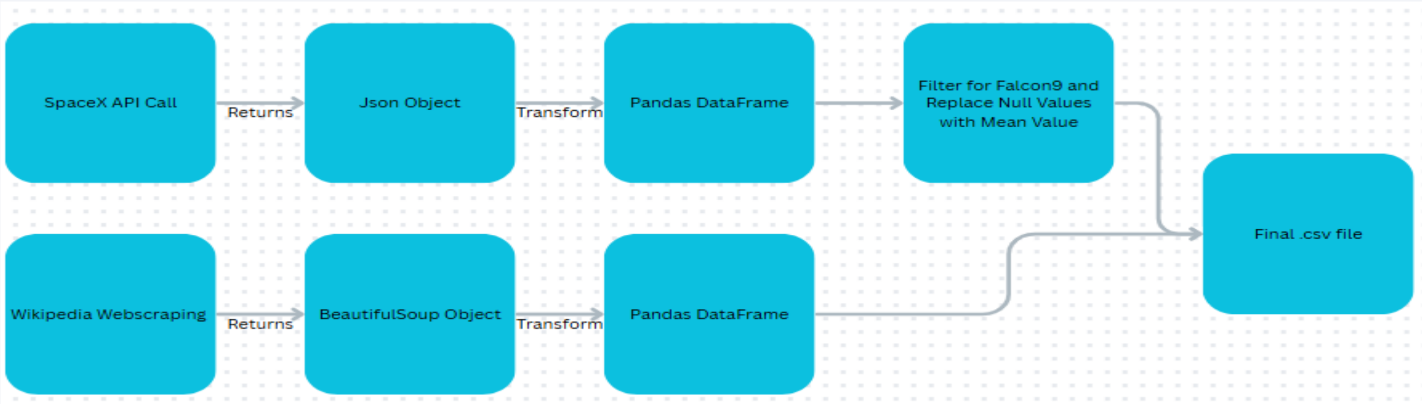
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - WebScraping
- Perform data wrangling
  - When processing the data, we replace null values from payloadmass with the mean value.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build models using scikitLearn, tune using gridSearch, and evaluate using test prediction accuracy.

Data Collection

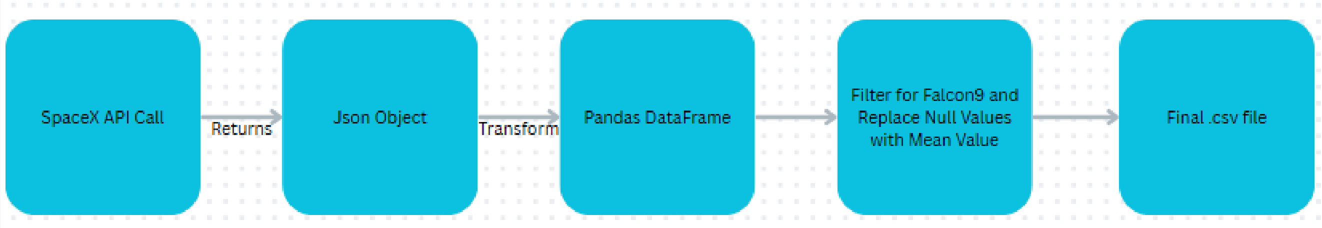
- The Data was collected in two ways
  - SpaceX API Calls - Using https://api.spacexdata.com we can get much data about previous launches in the form of a json object and convert it into a pandas DataFrame. We can perform more API requests to obtain different information about the individual launches and filter out from our DataFrame any that are not Falcon9. Finally, we replace missing values from payload Mass with the mean value of the column.
  - Wikipedia Webscraping – An alternative way of data collection is to use the public Wikipedia page that contains information about launches and parse the html of the page. We can then combine our results into a final .csv file to use for our Exploratory Data Analysis and predictive models



Data Collection – SpaceX API

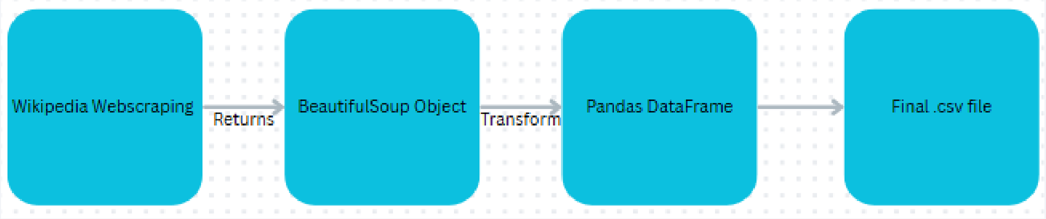
- Using SpaceX's Rest API we were able to obtain information about their F9 rocket launches and convert the data into a .csv file with different predictors as columns

- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%201%20Collecting%20the%20Data/jupyter-labs-spacex-data-collection-api.ipynb



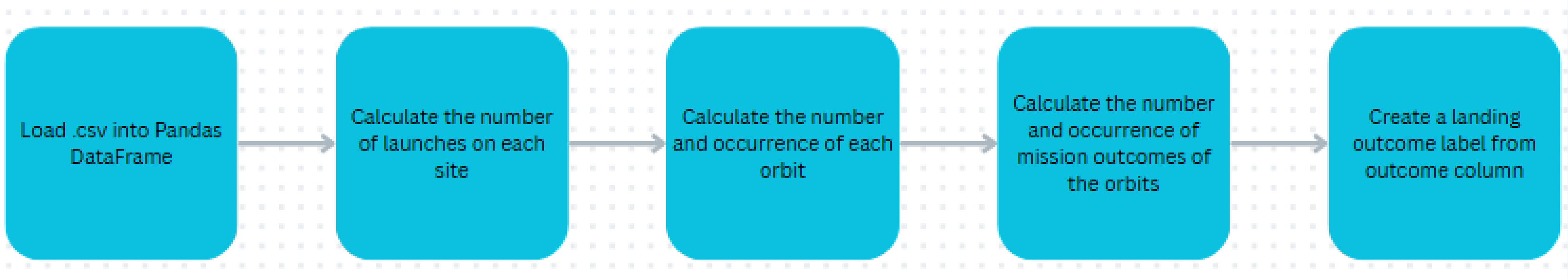
Data Collection - Scraping

- Using Wikipedia, we can scrape the HTML and convert it into a string/BeautifulSoup object. We can then split and parse this object to create a pandas DataFrame and finally convert it into a .csv file
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%201%20Collecting%20the%20Data/jupyter-labs-webscraping.ipynb



Data Wrangling

- We can first find the number of launches that occur from each site using the value counts.
- We can then find the number of occurrences of each of the different orbits to see what level we are launching the satellites to.
- We can determine the number of occurrences of each mission outcome and then convert each label into a class (0/bad or 1/good)
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%201%20Data%20Wrangling/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Flight Number vs Launch site
  - There does not seem to be a change in launch sites as SpaceX launches more flights.
- Payload Mass vs Launch site
  - VAFB launch site has a lower payload mass on average, but high success rate.
- Success Rate vs Orbit Type
  - Most orbits have a success rate over 50%, some with 100%, and SO as 0%.
- Flight Number vs Orbit Type
  - Does the orbit that SpaceX launches to change through the years
- Payload Mass vs Orbit Type
  - VLEO has a higher payload mass but also a higher success rate
- Launch success yearly Trend
  - We see more successful launches as time advances
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%202%20Exploratory%20Analysis%20Using%20Pandas%20and%20Matplotlib/edadataviz.ipynb

EDA with SQL

- Display Unique Launch Sites
- 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.
- List the total number of successful and failure mission outcomes
- List all the booster_versions that have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%202%20Exploratory%20Analysis%20Using%20SQL/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

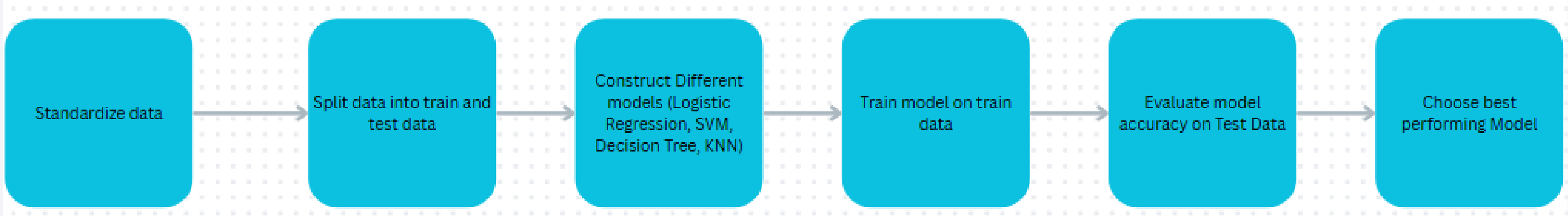
- I added circles to each of the launch sites of the rockets.
- I then added markers for the number of success/failed launches at each site
- Lastly, I added lines between the closest landmarks to each Launch Site
- These additions give us a simple visualization to where each launch site is, the success rate of each, and how close different infrastructures are to each.
- https://nbviewer.org/github/JacobHurley/IBMCapstone/blob/main/Module%203%20Interactive%20Visual%20Analytics%20and%20Dashboard/lab-jupyter-launch-site-location-v2.ipynb
  - Note that since github can not render the interactive elements, I have linked a notebook that does have them rendered, the original github link is
    - https://github.com/JacobHurley/IBMCapstone/blob/main/Module%203%20Interactive%20Visual%20Analytics%20and%20Dashboard/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- I added a dropdown menu to select different launch sites
- I then added a pie chart to display the success rate of launch sites
- I added a slider to filter by payload mass
- Finally, I added a scatter plot to view the success by payload mass and booster version
- These interactive visuals will allow us to further identify relationships between our predictors.
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%203%20Interactive%20Visual%20Analytics%20and%20Dashboard/spacex-dash-app.py

Predictive Analysis (Classification)

- I chose different machine learning models built into the Scikit-Learn library to determine which model had the highest validation accuracy for binary classification.
- The way I determined within each model which performed best (improved) was by using the GridCV method to test different hyper parameters for each model
- https://github.com/JacobHurley/IBMCapstone/blob/main/Module%204%20Predictive%20Analysis%20(Classification)/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

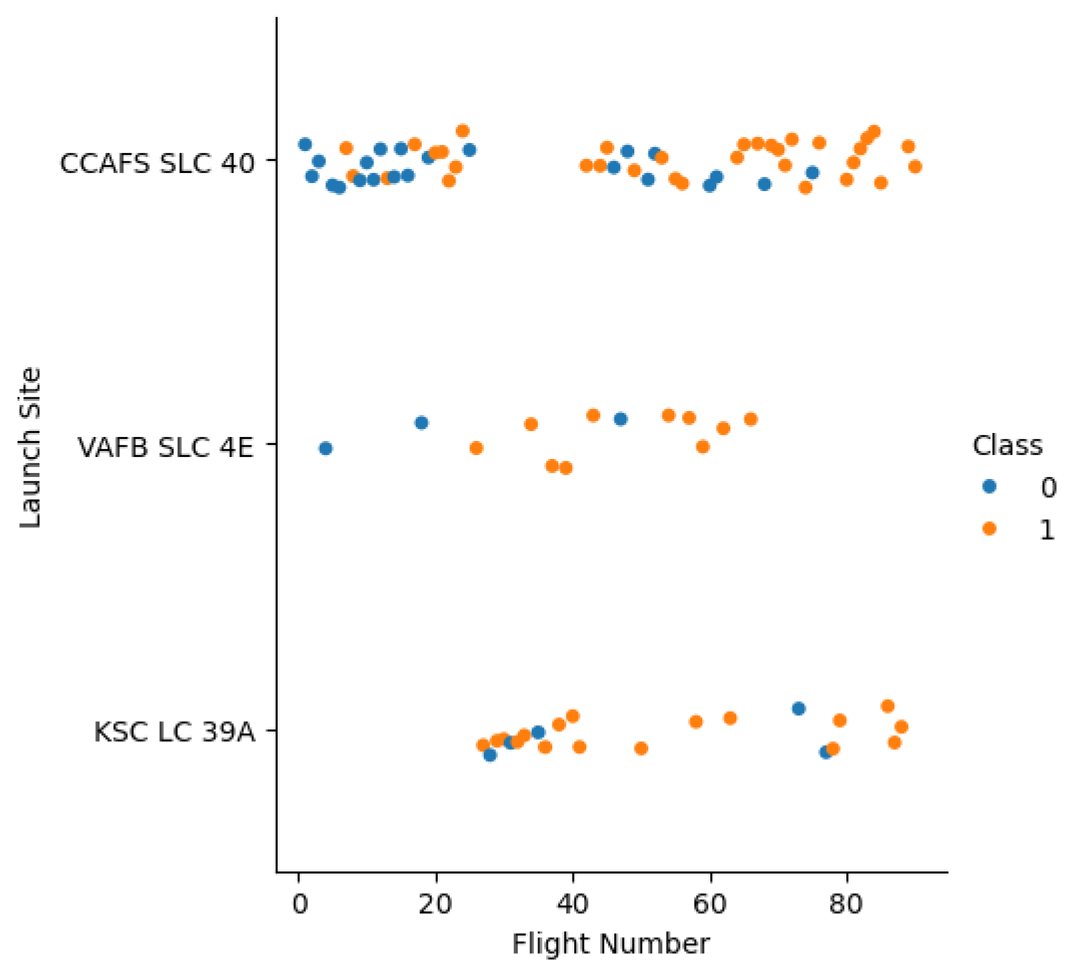
- Success Rate of launches and landing improve as time and number of launches increases.
- The Decision Tree Model had the best classification rate of about 94% compared to other models 83%.

Section 2

Insights drawn from EDA

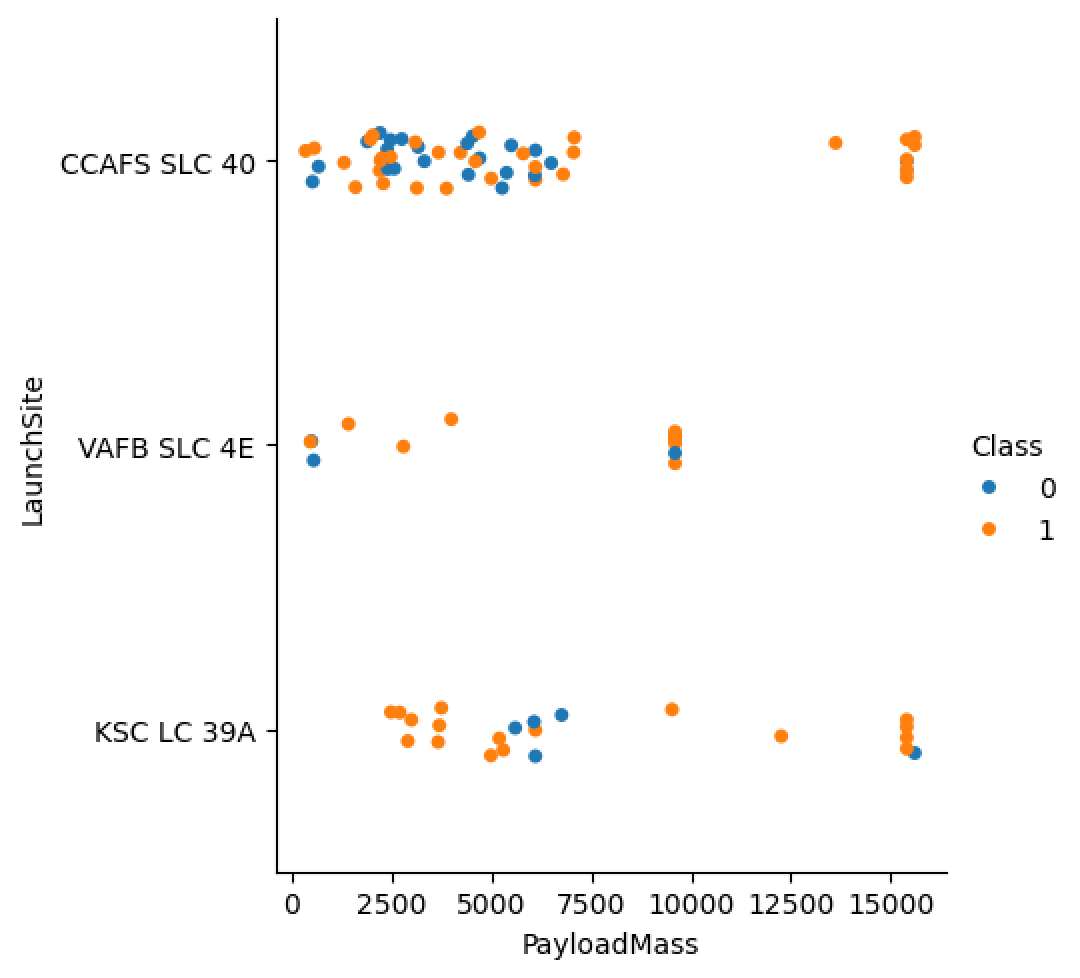
Flight Number vs. Launch Site

- VAFB launch site seems to be fazed out as flight number increase.
- There also appear to be a higher success rate as more rockets are launched.



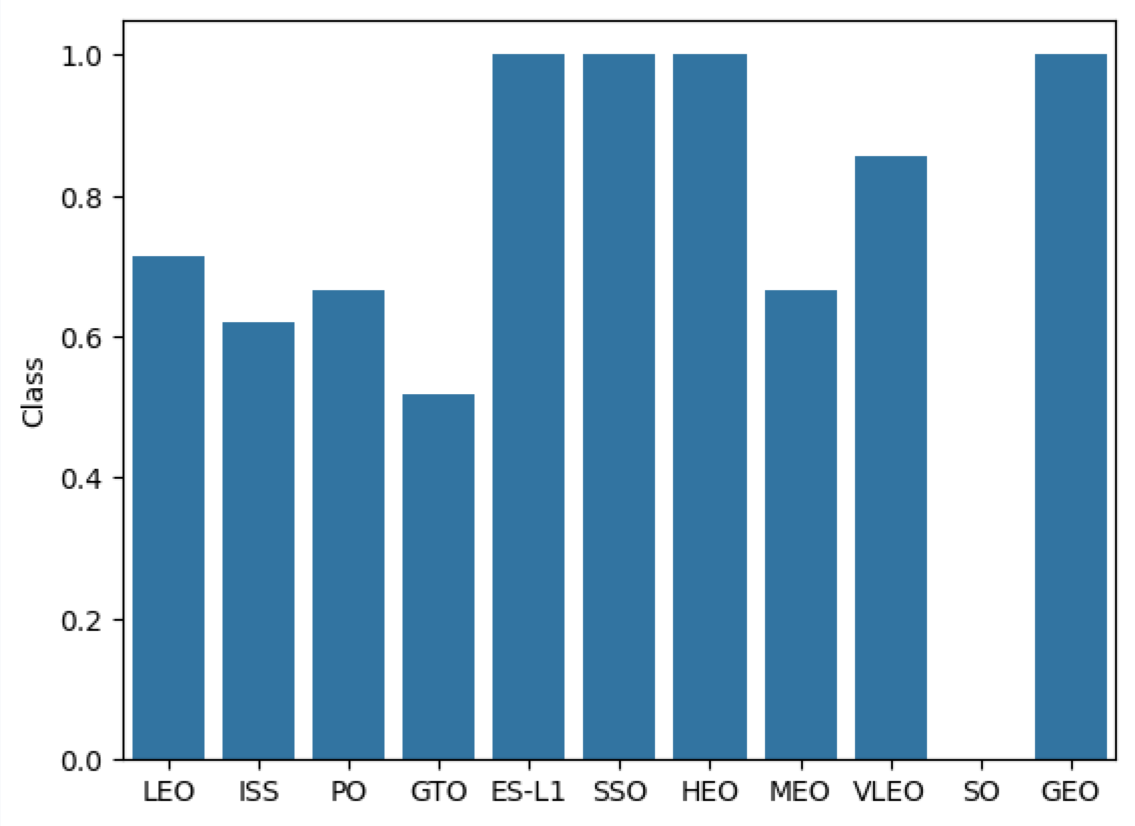
Payload vs. Launch Site

- VAFB appears to receive the lightest payloads.
- There appears to be a high success rate with heavy payloads.



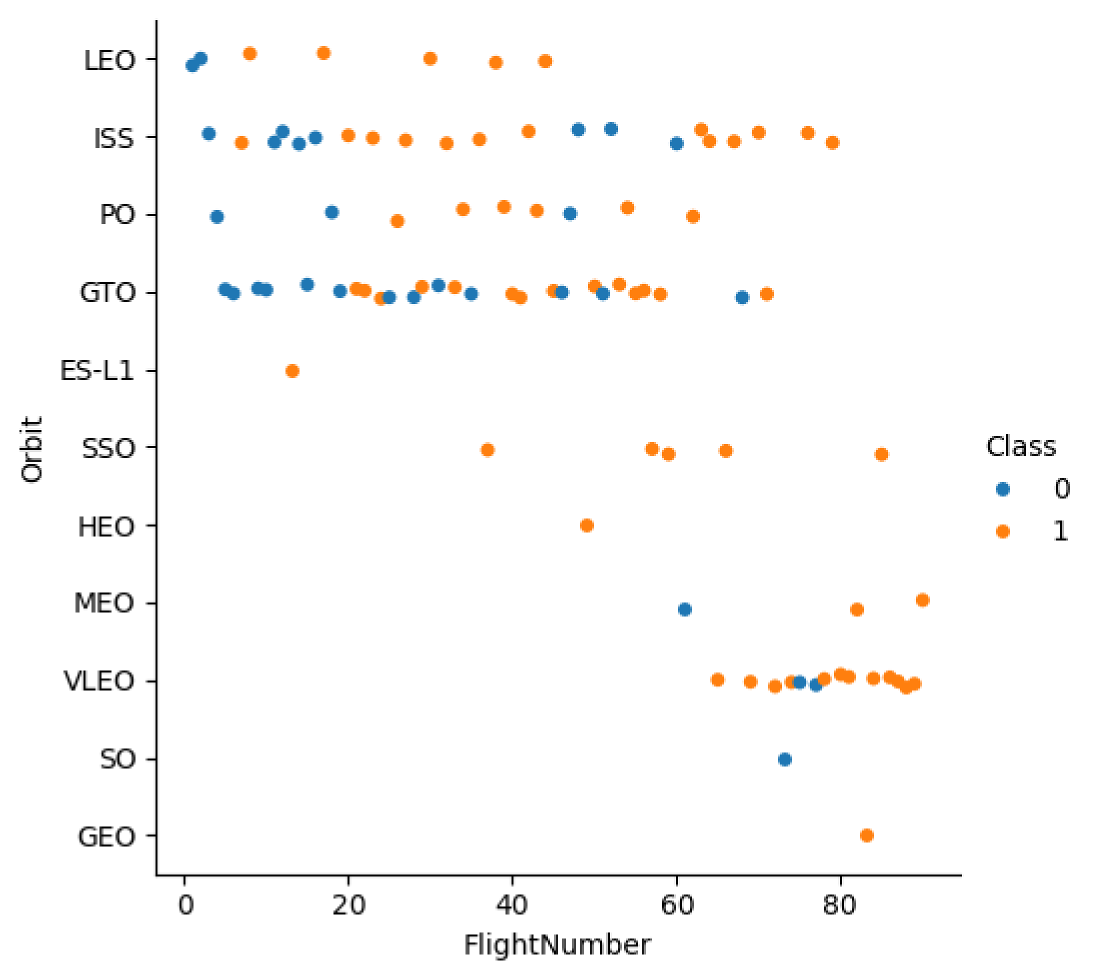
Success Rate vs. Orbit Type

- Most orbits have a success rate over 50%.
- SO has a success rate of 0%
- ES-L1, SSO, HEO, GEO all have a success rate of 100%



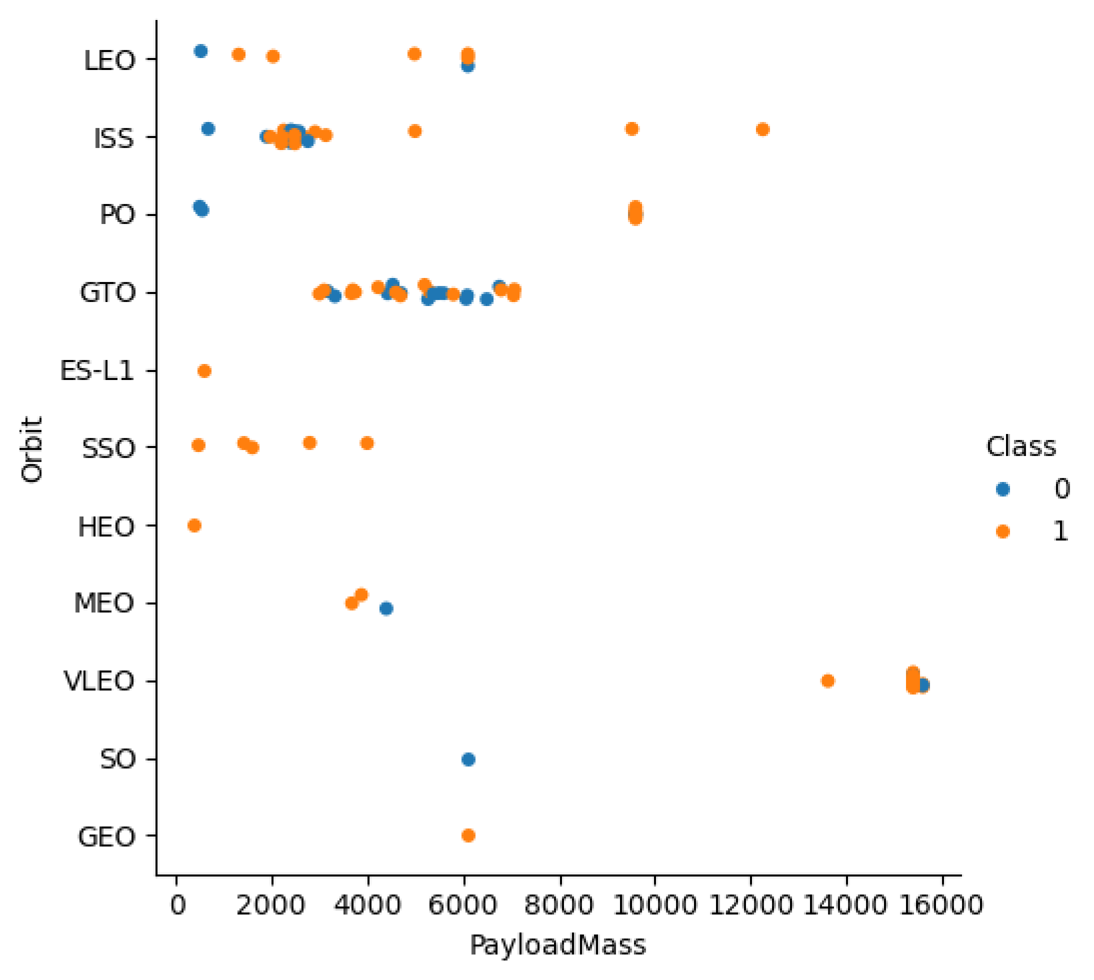
Flight Number vs. Orbit Type

- The orbit chosen appears to change to VLEA as SpaceX launches more flights.
- LEO also has not been launched to since early flights.



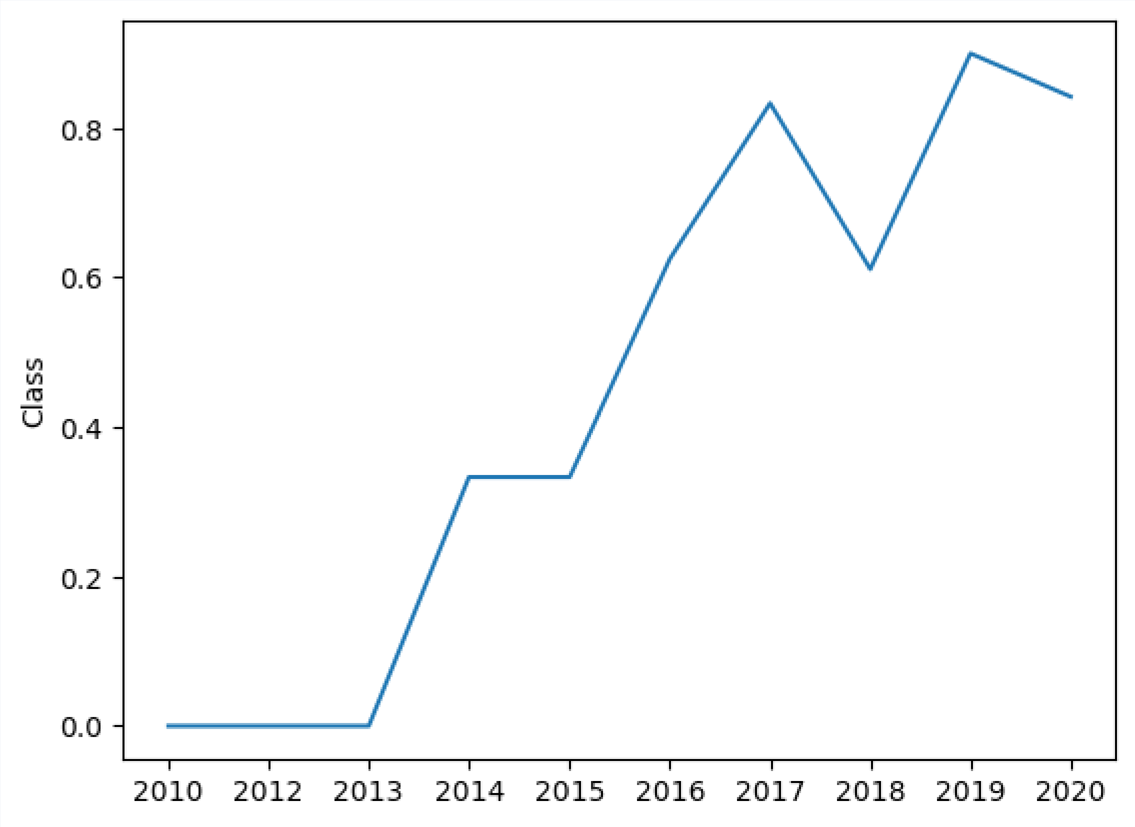
Payload vs. Orbit Type

- The heavier payloads are sent to VLEO orbit
- GTO has many mid-lighter size payloads



Launch Success Yearly Trend

- The success rate of each launch improves on average as the years increase.



All Launch Site Names

- The Unique Launch Sites include
  - 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

Launch Site Names Begin with 'CCA'

- The first 5 launches from a site with CCA were from 2010-2013 and all had successful mission outcomes and failed or not attempted landing outcomes.

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 is 45,596 KG. This is calculated by summing the Payload Mass in kilograms of rockets whose customer was nasa (CRS)

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average Payload Mass in kilograms for F9 1.1 Boosters is 2534.66.

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The first successful landing of the rockets occurred on December 22nd, 2015.

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The booster versions that have had a successful landing with a payload between 4000 and 6000 are
  - F9 FT B1022
  - F9 FT b1026
  - F9 FT B1021.2
  - F9 FT B1031.2

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

Total Number of Successful and Failure Mission Outcomes

- There were 100 successful missions and only 1 mission that failed.

Count(Mission_Outcome)	Count(Mission_Outcome)
100	1

Boosters Carried Maximum Payload

- There have been many boosters that have carried a maximum payload
- All of these boosters are F9 B5

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

- All of the failed landing outcomes in 2015 came from the same Launch Site
- The majority contained a v1.1 booster version and one contained a FT booster version.

substr(Date, 6, 2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

- No attempt is the largest landing outcome between 2010 and 2017
- There were 6 successes and 5 failures with 5 more landing in the ocean.

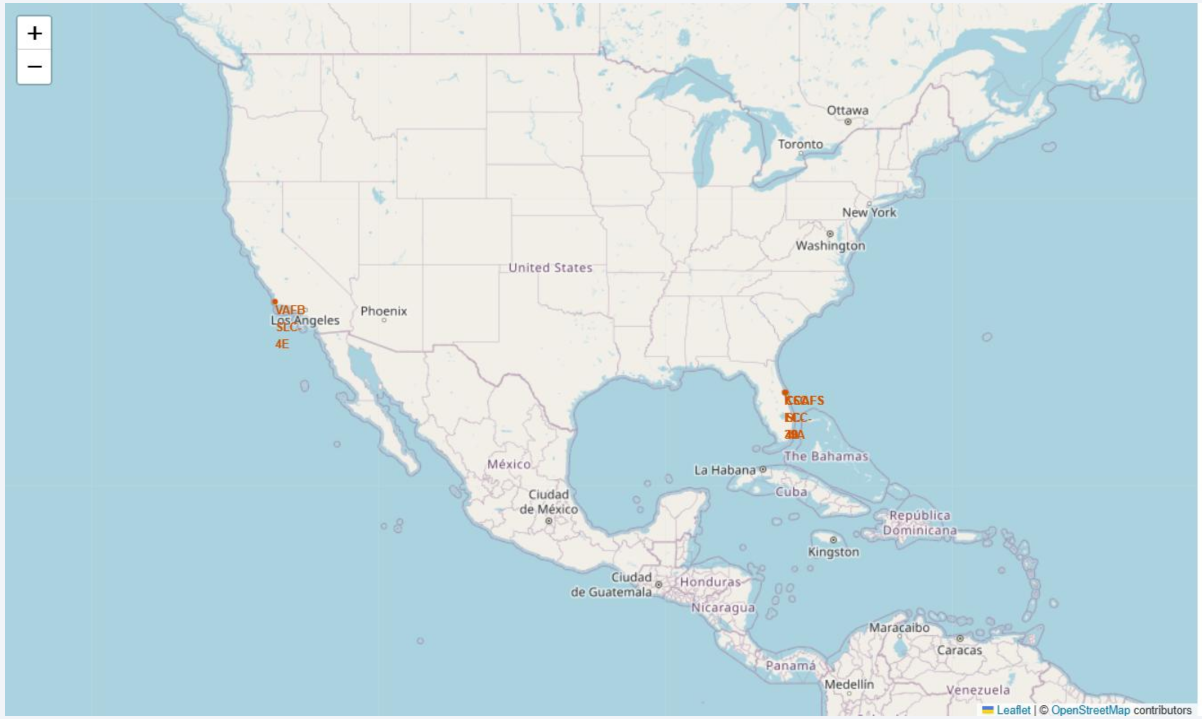
Landing_Outcome	Count(Landing_Outcome)
Controlled (ocean)	3
Failure (drone ship)	5
No attempt	9
Precluded (drone ship)	1
Success (drone ship)	4
Success (ground pad)	2
Uncontrolled (ocean)	2

Section 3

Launch Sites Proximities Analysis

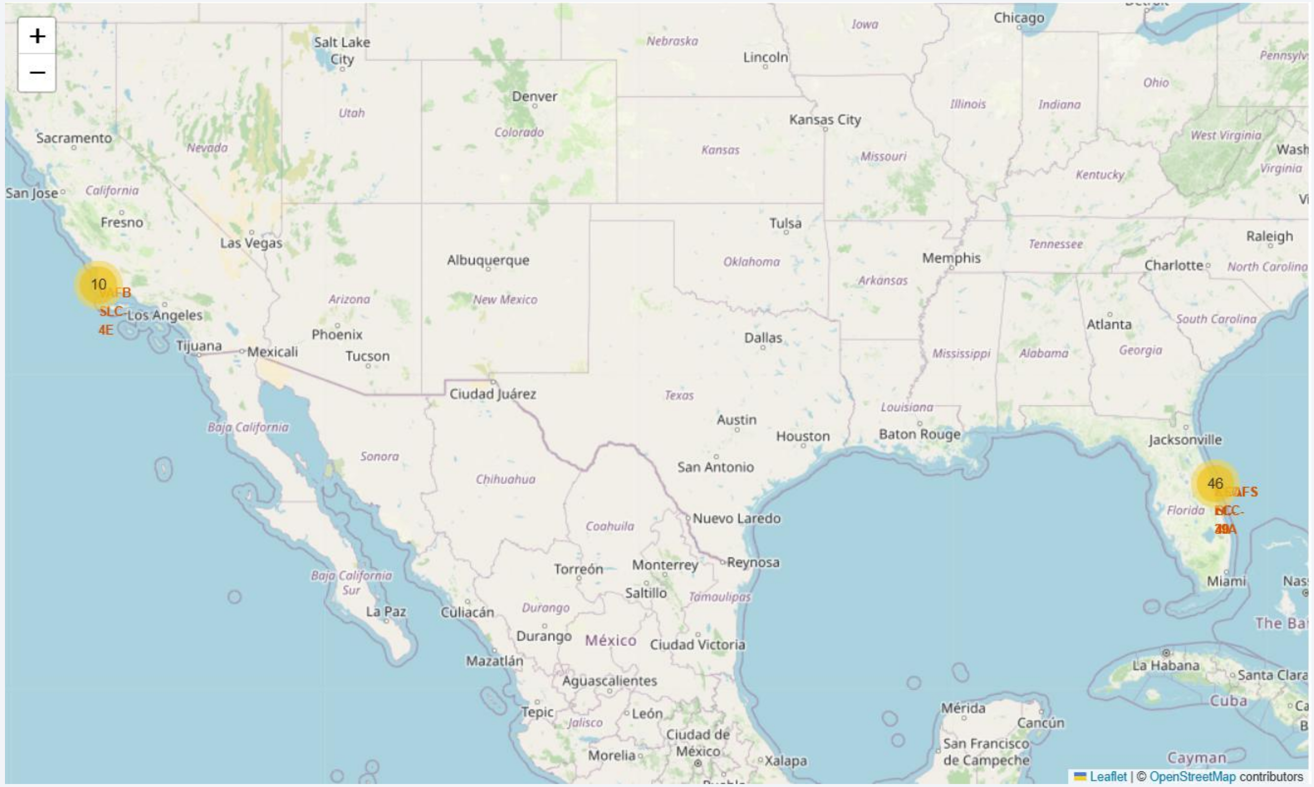
Launch Site Locations

- We Can see that the launch sites are focused between southern California and Florida.



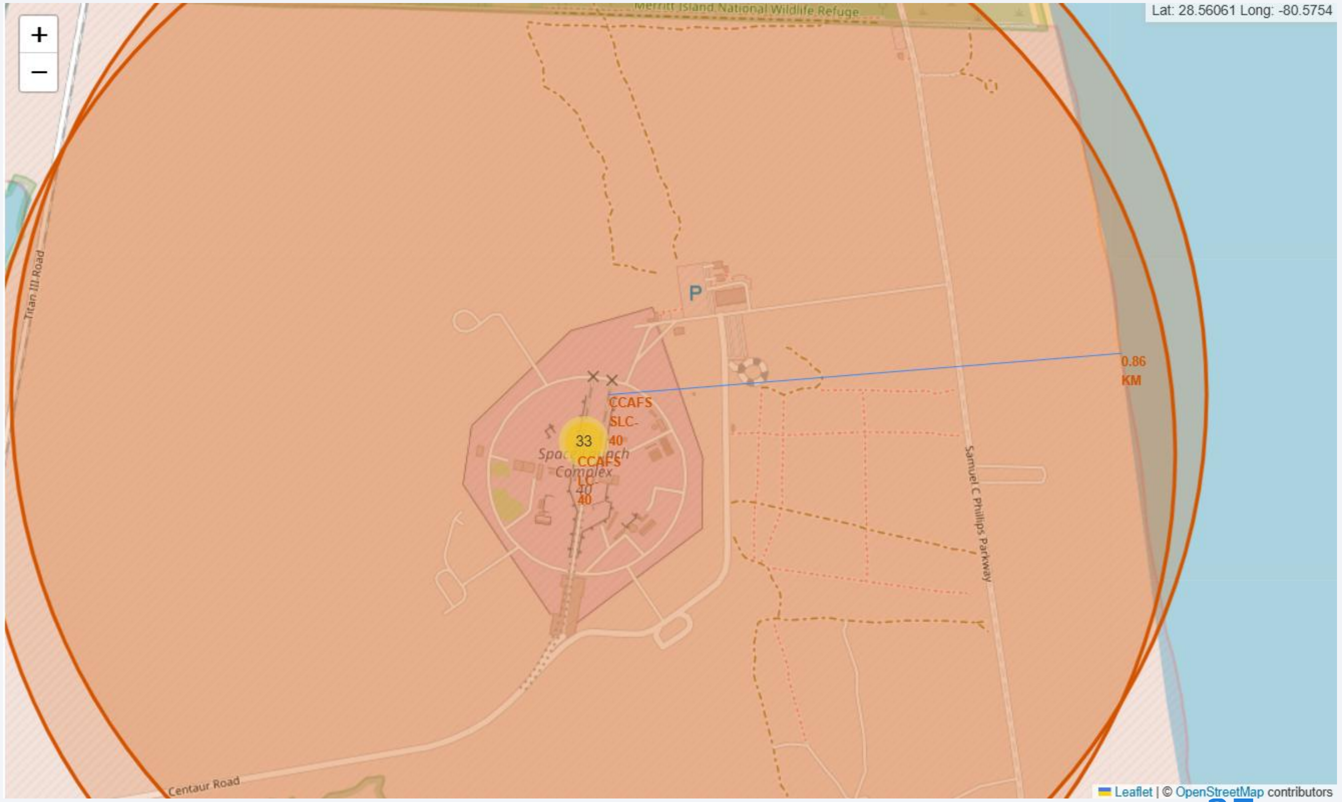
Launch Success by Launch Location

- Florida has more successful Launches than California.
- We see that Florida has 46 successful launches



CCAFS SLC-40 Distance from coastline

- We can see that the CCAFS SLC-40 Launch site is less than a kilometer from the coastline.



Section 4

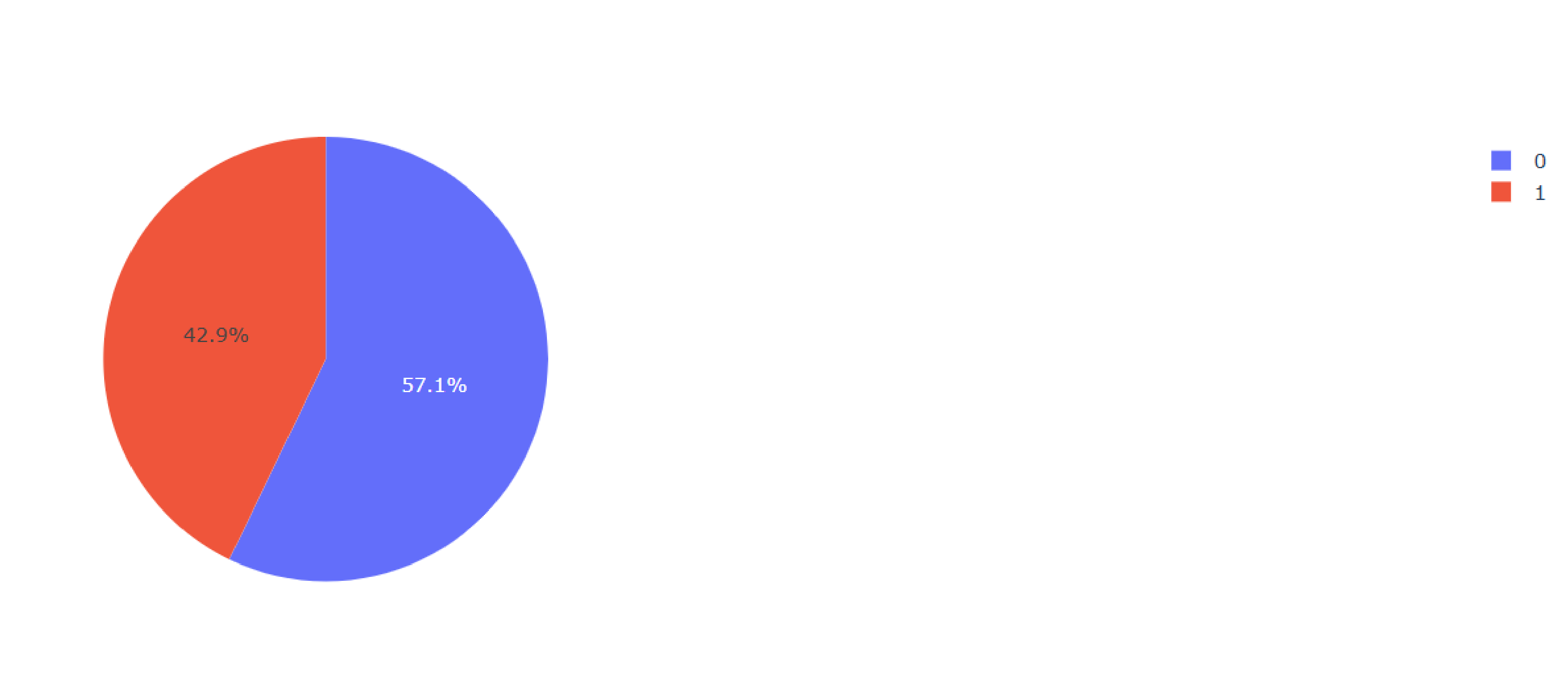
Build a Dashboard with Plotly Dash

Launch Site Launch Success percentage rate

- We see the KSC LC-39A accounts for 41.7% of the total successful launches of F9 Rockets.
- CCAFS SLC-40 accounts for the least number of successful launches at 12.5%

CCAFS SLC-40 Launch Success Rate

- The highest success rate of launch sites comes from CCAFS SLC-40.
- This has a ratio of 42.9% success and 57.1% failure



PayloadMass and booster success rate

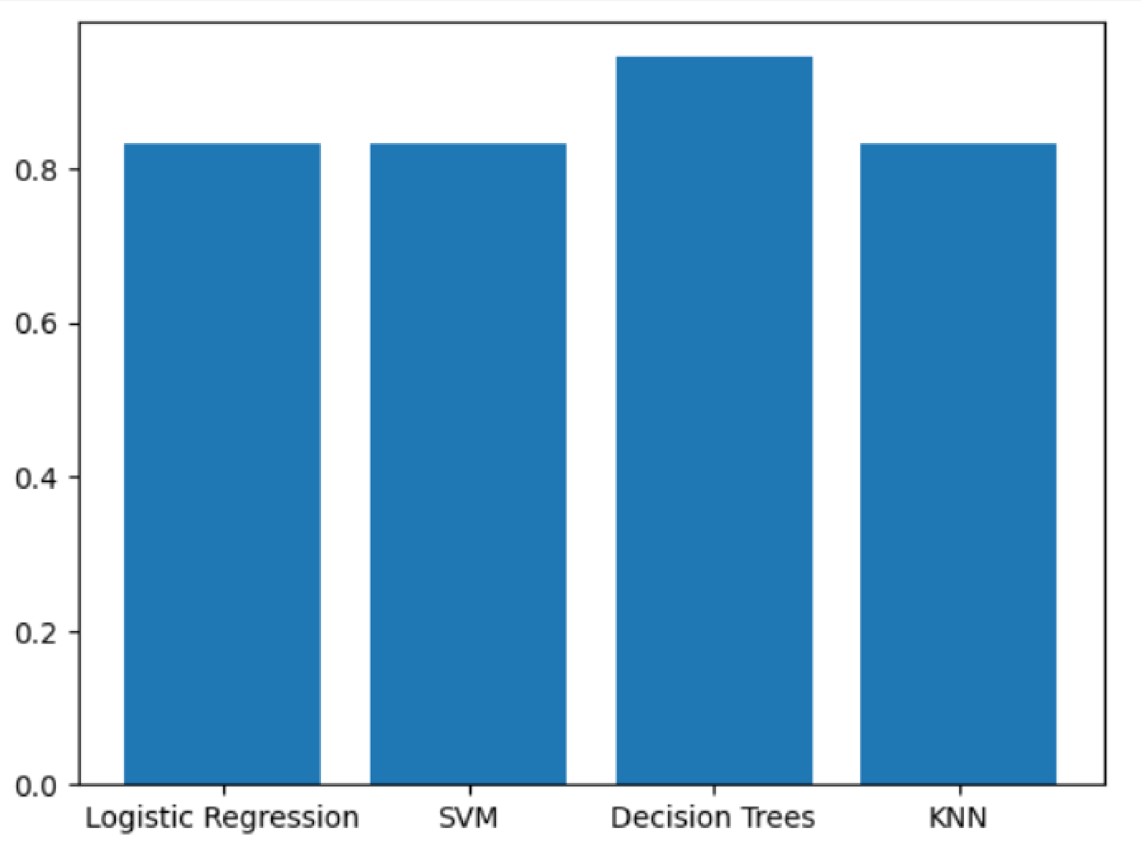
- The highest success rate appears to come for payload masses under 6000 kilograms and FT Boosters.

Section 5

Predictive Analysis (Classification)

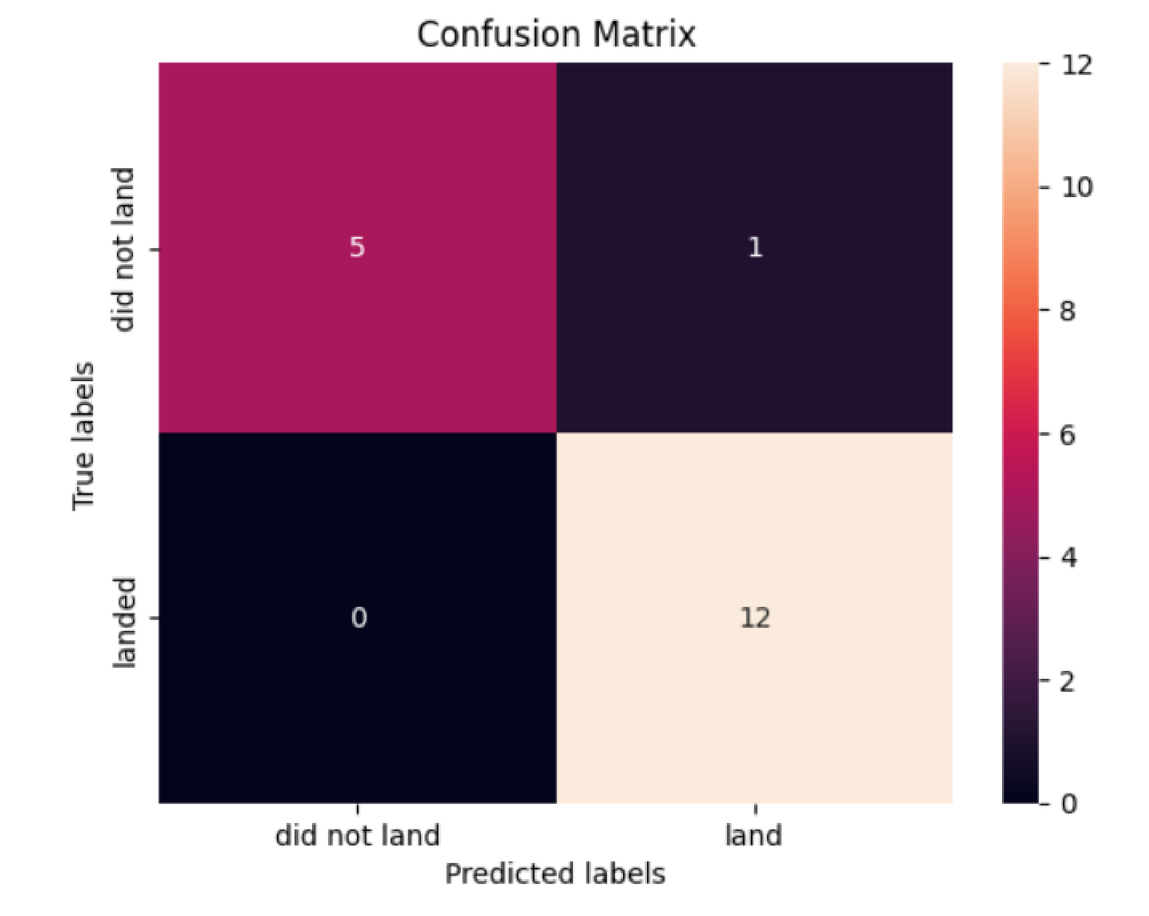
Classification Accuracy

- Decision Trees have the highest classification accuracy at about 94%



Confusion Matrix

- We see that the decision tree only misclassified one point, and it predicted that it landed when it did not actually land.



Conclusions

- Predictors such as launch number (Time), launch site, payload mass, and orbit are important predictors for landing success.
- A Decision tree classifier works best to predict the success of a landing with an accuracy of about 94%. Depending on stakeholder's demands, this may be a suitable model for use in production settings.

Thank you!