IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Danish Nurhira Putra
19/ September/ 2024

Outline

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

Executive Summary

Summary of methodologies

- Data was collected using a combination of a Wikipedia website and a SpaceX API. Numpy, pandas, sci-kit learn, seaborn, SQL, Matplotlib, Folium, and Dash were all used to wrangle, clean, explore, visualize and create predictive models to determine landing outcomes of the Falcon 9 rockets.

Summary of Results

- The FT rocket booster has the highest success rate for landing outcomes compared to all other boosters, especially with a payload mass of 2000 to 4000 kg. KSC LC-39A has shown to have the highest chance of a successful landing compared to the other sites. Combining these two observations could yield a reliable outcome for landing success.

Introduction

- For most of history, rockets have been single use rockets. Designing a rocket that can land the first stage of the rocket for reuse can greatly reduce the waste and cost for future space flights. This project analyzes Falcon 9 launch data to predict if the first stage of the rocket will land successfully. The ability to predict the success of a landing outcome can narrow the variables that have the highest impact on successful landings.
- We would like to answer in this project, what variables have the greatest impact on landing success? Is there a specific launch site that has a higher landing success rate? What rocket booster versions have the highest success? Does the mass of the payload being carried impact landing success? How does the flight orbit affect landing outcomes?

Section 1

Methodology

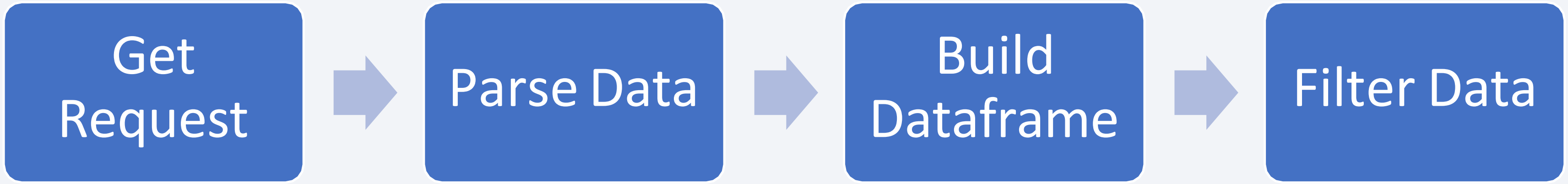
Methodology

Executive Summary

- Data collection methodology:
  - Web scraping from Wikipedia - List of Falcon 9 and Falcon Heavy Launches
  - Web scraping from SpaceX API
- Perform data wrangling
  - Determined number and occurrence of launches, orbits, and outcomes
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

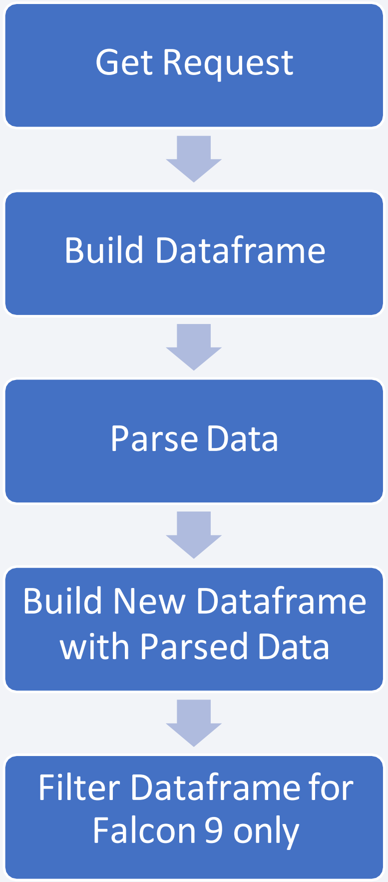
Data Collection

- Data was collected by using a get requests on a SpaceX API and a Wikipedia page. Functions were used to parse the data and build that data into dataframes. SpaceX API data needed to be filtered to only include Falcon 9 data as this was the focus of this project.



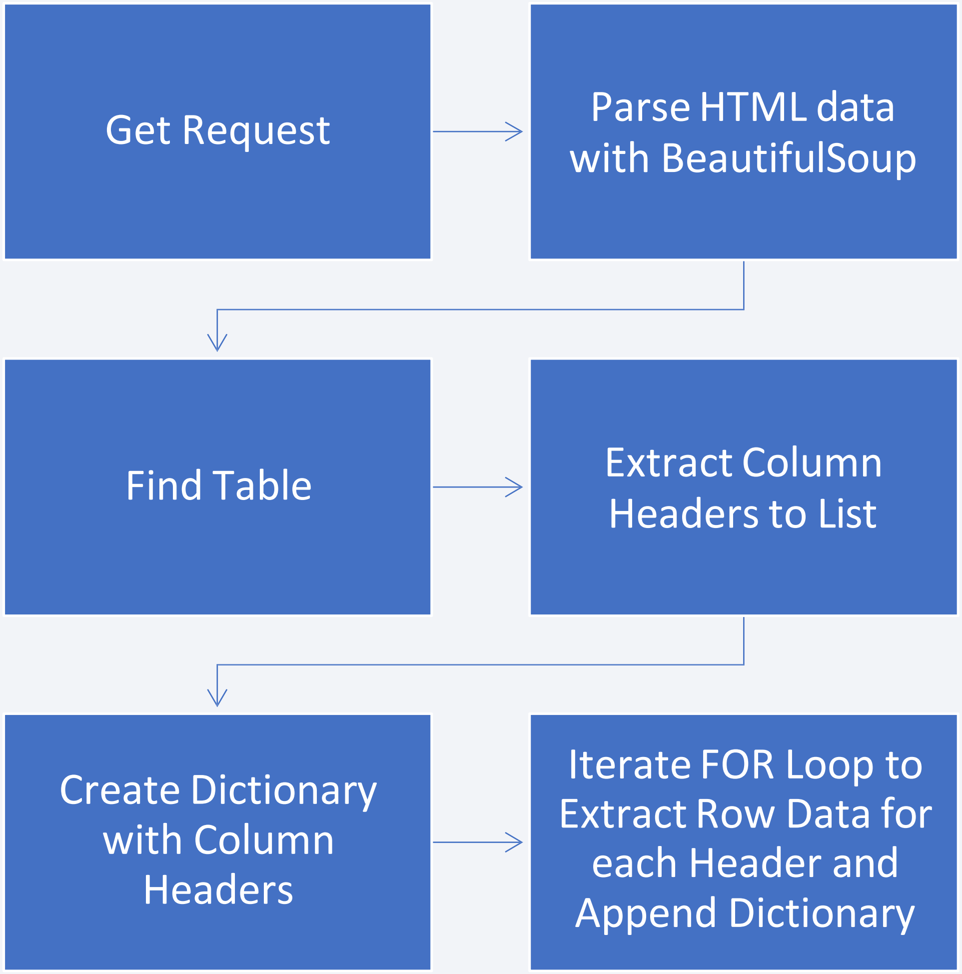
Data Collection - SpaceX API

- [Capstone-Project-Coursera/SpaceX Data Collection.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)]



Data Collection - Scraping

- [Capstone-Project-Coursera/SpaceX Webscraping.ipynb at main ·]



Data Wrangling

- First, I checked the number of occurrences for each launch site and orbit. Next, I found the number of mission outcomes by string name. Creating a list of the failed outcomes, I assigned each outcome occurrence a 0 (failure) or 1 (successful), and assigned that to a landing class list. I appended the dataframe to include the landing class list.
- [Capstone-Project-Coursera/SpaceX Data Wrangling.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)]

Value Count Launchsite Column → Value Count Orbit Column → Value Count Outcome Column → Assign Failed Outcome Strings to a Set → Create landing_class list → Loop Set and Append List with 0 and 1 → Add New landing_class List to Dataframe

EDA with Data Visualization

- Multiple scatter plots were made with the landing class as the color hue for each occurance. The scatter plots created were Flight Number vs (Payload Mass, Launch Site, and Orbit) and Payload Mass vs (Launch Site and Orbit). These plots were used to visually identify if there were any relationships between the columns of data.
- A bar chart was created to determine which orbits had the highest success rates with 4 having 100% success rates.
- Finally, a line plot was created to determine the success rate for flights between 2010 and 2020.
- Capstone-Project-Coursera/Exploring and Preparing Data Visualizations.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)

EDA with SQL

- Listed each launch site
- Pulled 5 rows of data for launch sites beginning with ‘CCA’
- Calculated the total payload mass launched by NASA (CRS)
- Found the average payload mass carried by F9 v1.1 booster
- Found the first date when a successful ground pad landing occurred
- Determined the booster versions with payloads between 4000-6000 kg and successful drone ship landing outcomes
- Found the total success and failure mission outcomes
- Listed the booster that carried the heaviest payload mass
- Listed records of the failed drone ship outcomes for the year of 2015
- Ranked the landing outcomes by count between 6/4/2010 and 3/20/2017
- Capstone-Project-Coursera/Data Exploration with SQLlite.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)

Build an Interactive Map with Folium

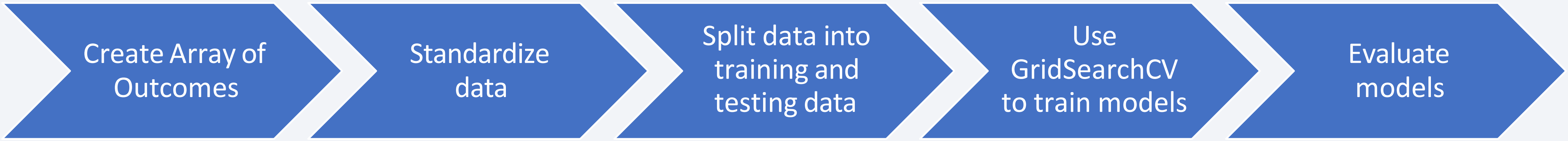
- Launch sites were added to a folium map of the United States that detailed out how many successful and failed landing outcomes occurred.
- Each launch site also shows the distance to the nearest coast, railroad, highway, and city.
- Each of these objects displays a lot of information in a relatively clean and understandable fashion allowing the user to understand each site better.
- Capstone-Project-Coursera/Folium Site Locations.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)

Build a Dashboard with Plotly Dash

- Two plots were created using Plotly Dash: pie chart and scatter plot
- The pie chart showed either the successful flights of all sites or each sites success rate
- The scatter plot showed the landing outcomes for either all sites or one by booster version and payload mass.
- The payload mass, launch site, and booster versions showed varying degrees of influence on the landing outcome success. Combining all of these variables into interactive plots made it easier to see how each variable affected the outcome.
- Capstone-Project-Coursera/spacex_dash_app.py at main · DanishNurhira/Capstone-Project-Coursera (github.com)

Predictive Analysis (Classification)

- Each prediction model was built by comparing the landing outcome to the other standardized variables. The data were split into training and testing sets where 80% of the data were used to train each model.
- The models used were linear regression, SVC, decision tree, and K-nearest neighbor. The models were created using GridSearchCV and the training data. The models were evaluated by their best parameters and their accuracy using best_params_ and best_score_. Confusion matrices were plotted to visualize the accuracy of the models.
- Capstone-Project-Coursera/SpaceX_Machine Learning Prediction.ipynb at main · DanishNurhira/Capstone-Project-Coursera (github.com)



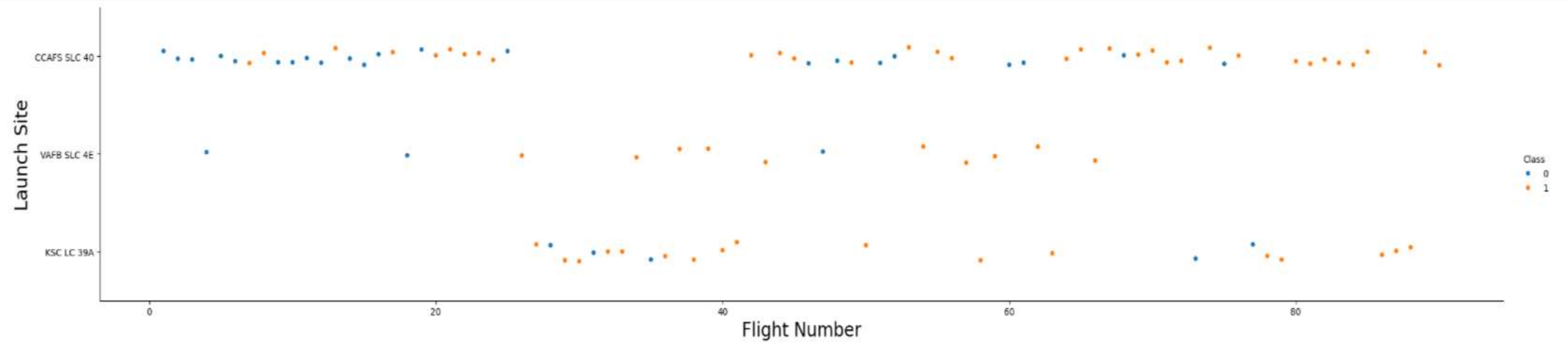
Results

- Exploratory data analysis showed that there were relationships between orbit, payload mass, and launch site that affected the landing outcome.
- The interactive analytics demonstrations further portrayed these relationships with greater insight and uncovered the affect of the booster version contributed to the success of landing.
- The results of both the exploratory data analysis using Matplotlib and SQL and the interactive analytics using Folium and Plotly Dash showed that the highest probability of a successful landing was using an FT booster launched from KSC LC-39A with a payload between 2000 to 4000 kg reaching orbits: ES-L1, GEO, HEO, or SSO.
- All the predictive analytic models performed well with all predicting most of the landing outcomes correctly but all with 3 false positive landing outcomes.

Section 2

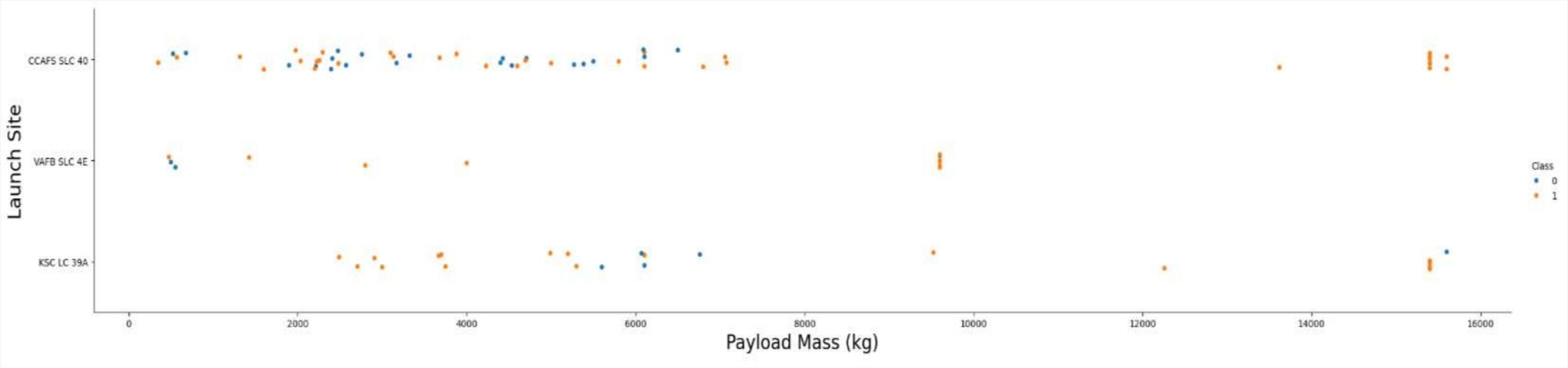
Insights drawn from EDA

Flight Number vs. Launch Site



- The plot above shows the successful (Orange) and unsuccessful (Blue) landing outcomes for 3 different launch sites (CCAFS SLC-40, VAFB SLC 4E, and KSC LC 39A) vs the corresponding flight number.

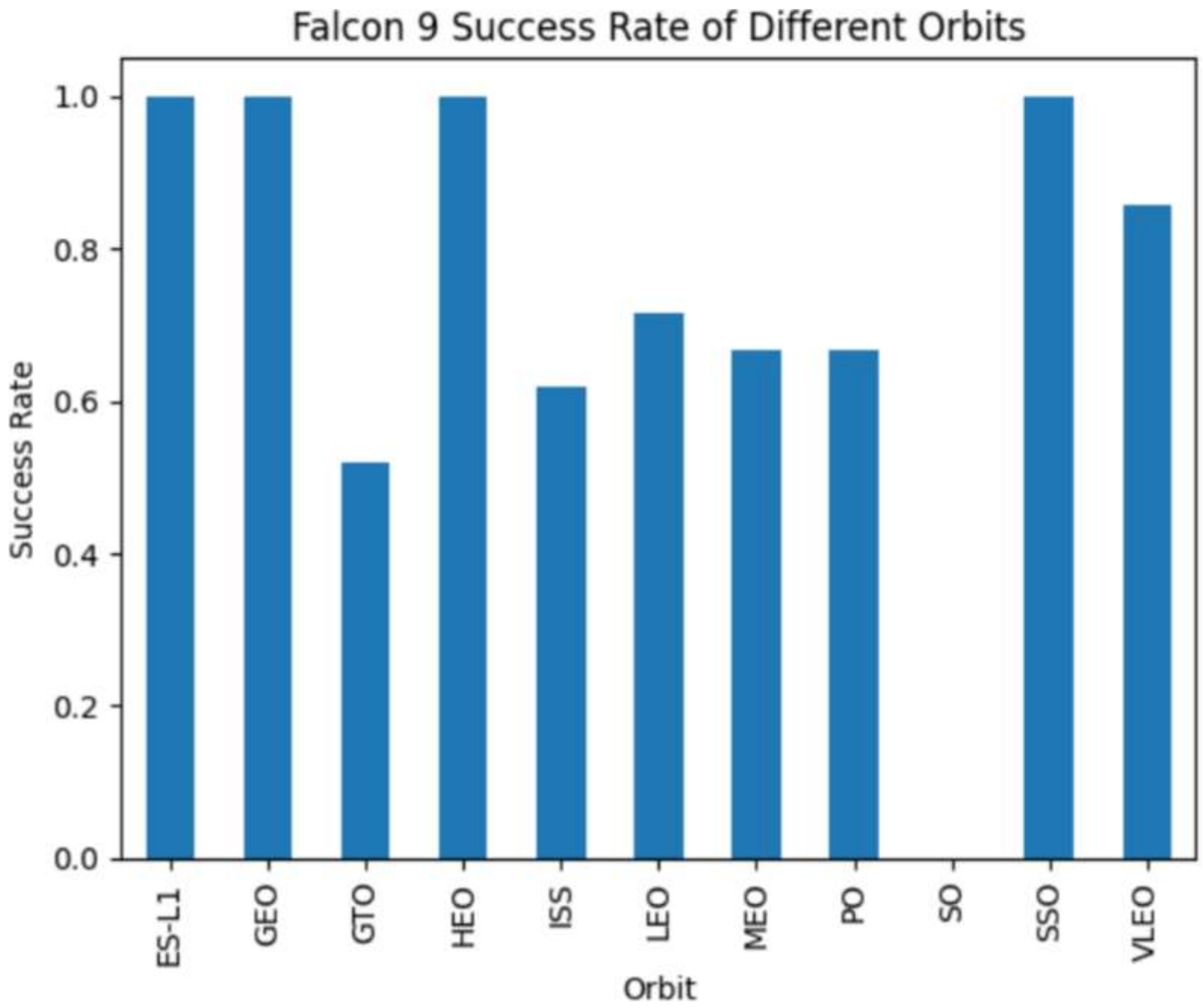
Payload vs. Launch Site



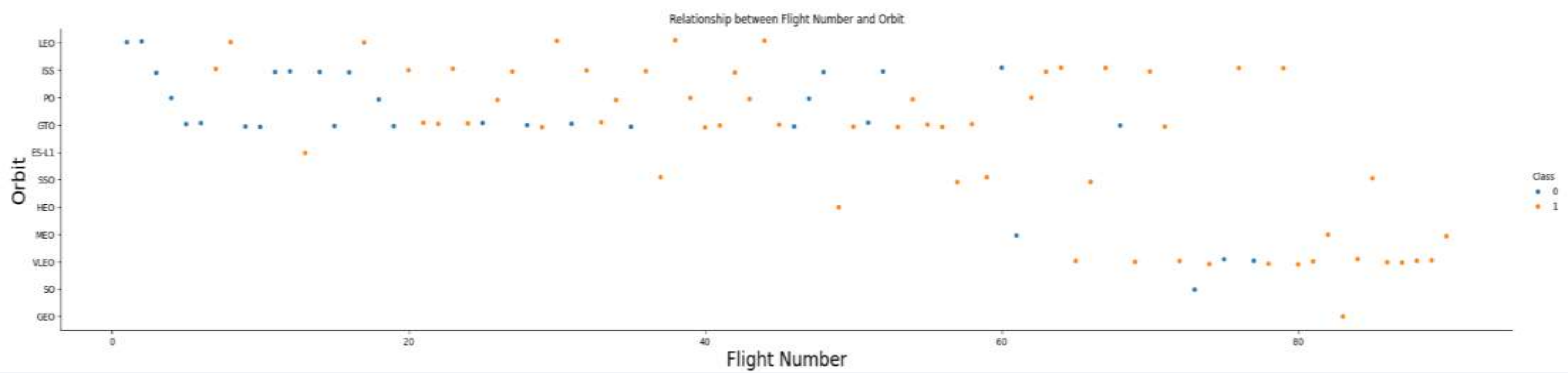
- This plot shows the successful and unsuccessful landing outcomes for each launch site depending on the mass in kg of the payload the rocket was carrying.

Success Rate vs. Orbit Type

- The bar chart on the right shows the successful landing rate for each orbit that the Falcon 9 rockets reached.

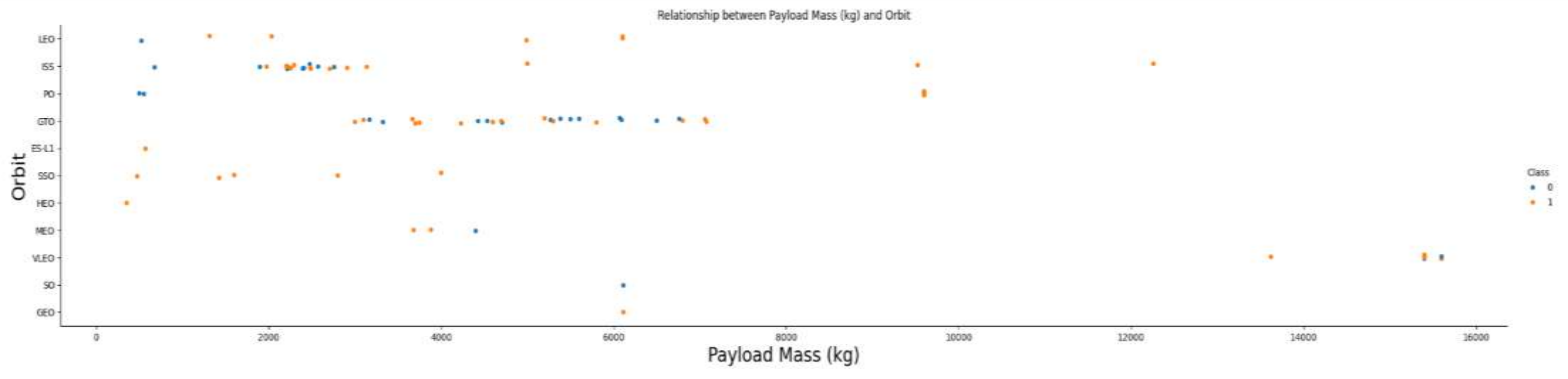


Flight Number vs. Orbit Type



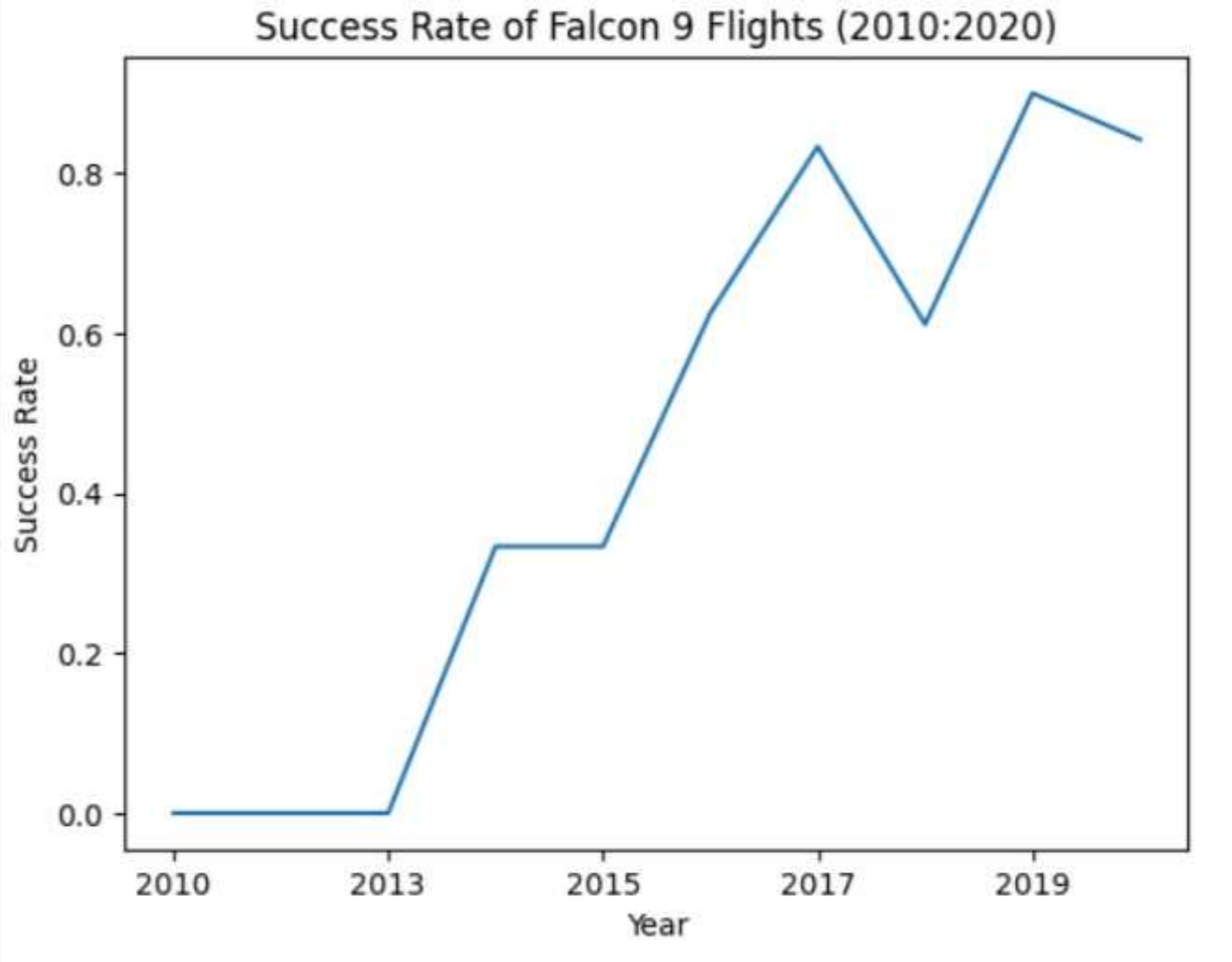
- Above shows the landing outcome based off relationship of the flight number and the orbit the Falcon 9 reached.

Payload vs. Orbit Type



- This plot shows the landing outcome based on the relationship between the payload mass the Falcon 9 was carrying and the orbit the rocket reached throughout the flight.

Launch Success Yearly Trend



- The graph on the right shows the Falcon 9 landing outcome success rate for each year from 2010 to 2020

All Launch Site Names

- SQL was used to find a list of all the launch sites the Falcon 9 flights

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

Launch Site Names Begin with 'CCA'

- Below is a list of 5 records where the Launch Site begins with ‘CCA’

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

- Using SQL, the total payload mass (kg) launched by NASA (CRS) is shown below:

Total Payload Mass Launched by NASA (CRS)
48213

Average Payload Mass by F9 v1.1

- SQL was able to calculate the average payload mass carried by the F9 v1.1 booster version.

Average Payload Carried by F9 v1.1 Booster
2534.6666666666665

First Successful Ground Landing Date

- The first ever successful ground landing occurred:

MIN("Date")
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- On the right is a list the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

Booster_Version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- Listed below are the successful and failure missions.

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

Boosters Carried Maximum Payload

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- The list on the left shows the different booster versions that have carried the highest payload by a Falcon 9 rocket.

2015 Launch Records

- There were only two times during 2015 where the Falcon 9 failed to land on a drone ship.

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

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

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

Landing_Outcome	Date	Rank
No attempt	2012-05-22	10
Success (drone ship)	2016-04-08	5
Failure (drone ship)	2015-01-10	5
Success (ground pad)	2015-12-22	3
Controlled (ocean)	2014-04-18	3
Uncontrolled (ocean)	2013-09-29	2
Failure (parachute)	2010-06-04	2
Precluded (drone ship)	2015-06-28	1

Section 3

Launch Sites Proximities Analysis

All Falcon 9 Launch Site

- Shown below is a map of all the Falcon 9 launch site. All sites are located neat the coast in the southern half of the United States. KSC LC-39A, CCAFS LC-40, and CCAFS SLC-40 are all located in Florida while VAFB SLC-4E is in California.

Launch Outcomes of CCAFS LC-40

- Below shows the number of launches, color coded for successful (green) and failed (red) launches

VAFB SLC-4E Relations to Landmarks

- VAFV SLC-4E is located roughly 1.32 km from the pacific coastline, 1.24km from the US Santa Barbara Railroad, 5.91 km from Ocean Avenue highway, and 38.14 km from the city of Santa Maria.

Section 4

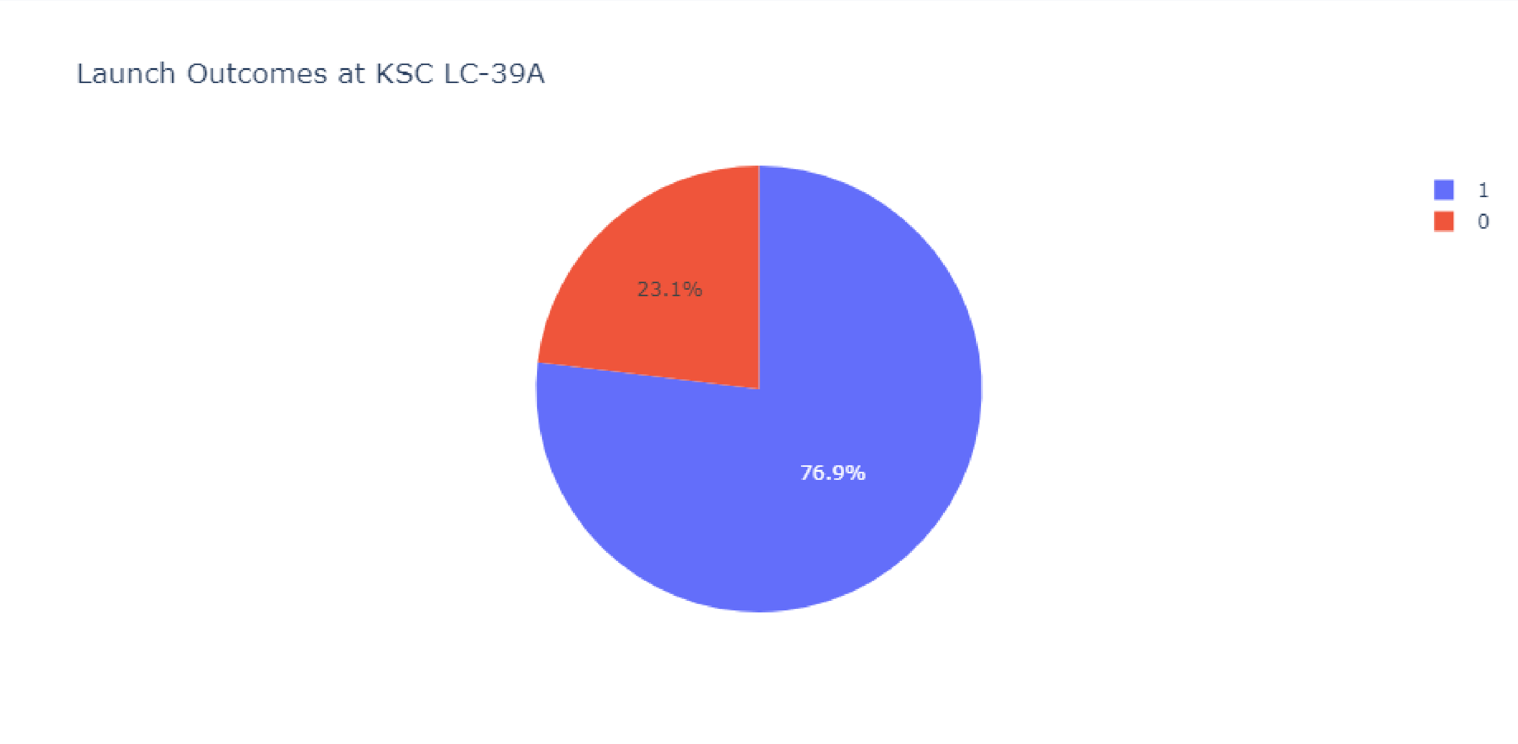
Build a Dashboard with Plotly Dash

Successful Launches by Launch Site

- For all the successful launches, KSC LC-39A has the most successful launches compared to its peers

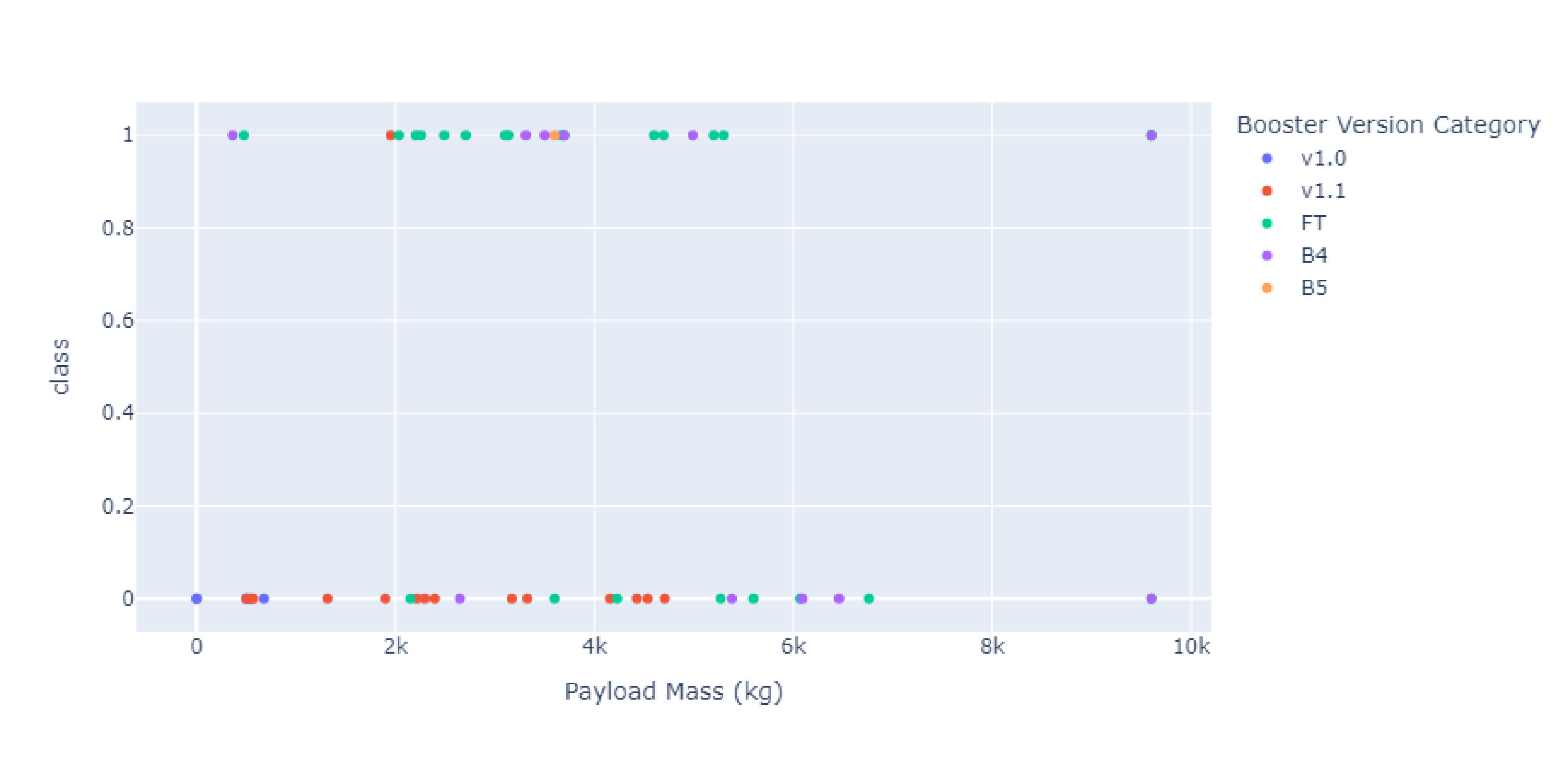
KSC LC-39A

- Not only has KSC LC-39A had the most successful launches compared to the other launch sites, but it also demonstrates the highest success rate of all the sites.



Success Outcome of Booster with Different Payloads

- The FT booster has shown the most success for landing outcomes, especially for the payload mass range of 2000 to 4000 kg.



Section 5

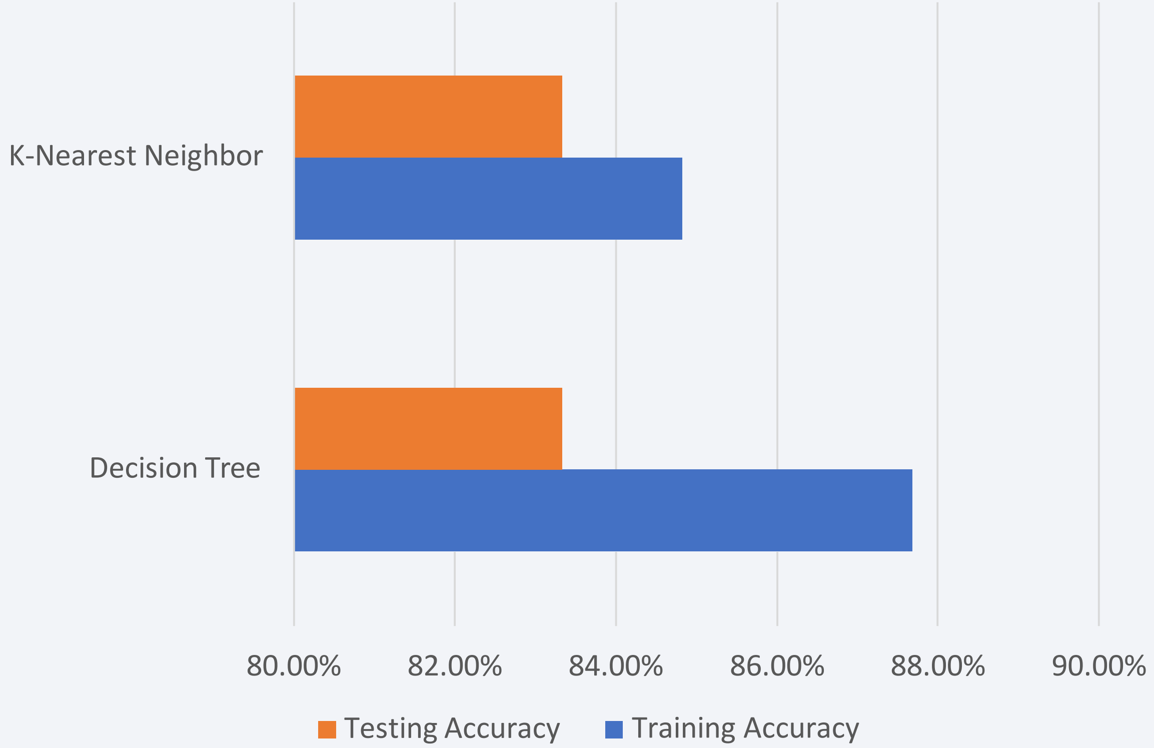
Predictive Analysis (Classification)

Classification Accuracy

The testing outcomes for all algorithms used resulted in an 83.33% accuracy.

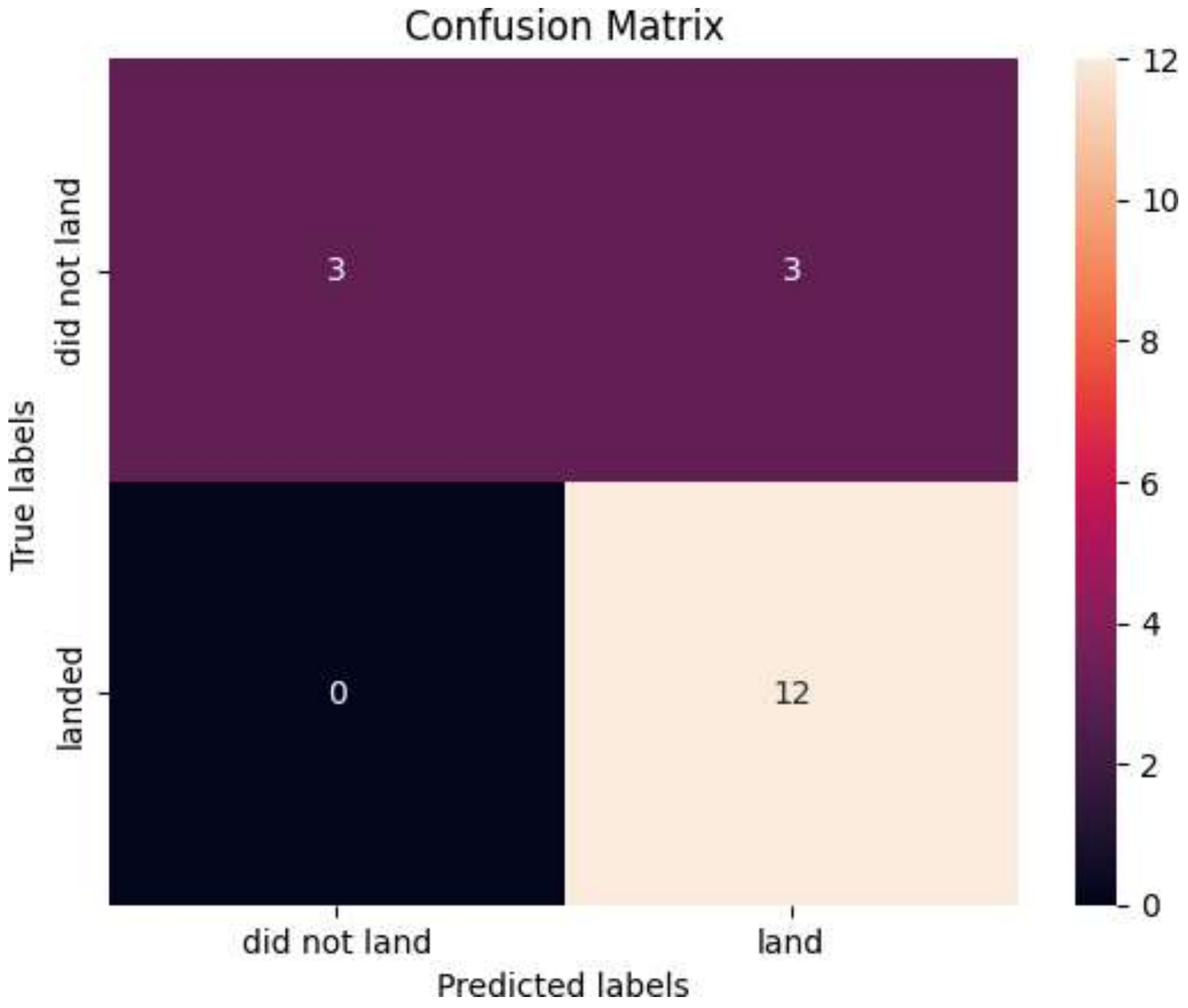
The training accuracy was the determining factor that showed any difference in accuracy.

The best was the Decision Tree model.



Confusion Matrix

- The Decision Tree confusion showed good accuracy but did result in 3 false positive landing outcomes.
- Tweaking some of the training variable might improve the overall accuracy of testing outcomes.



Conclusions

- The testing outcomes for all models resulted in an 83.33% accuracy predicting successful/unsuccessful landing outcomes.
- Adjusting training/testing data splits and model variables could potentially improve testing outcome accuracy
- Overall, the models produce good predictions of landing outcome. However, false positives were observed.
- As more launches occur the models should improve.

Appendix

- For a complete review of all the notebooks used in this Capstone Project please refer to the following website:

[DanishNurhira/Capstone-Project-Coursera (github.com)]

Thank you!