IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Deric Green
July 6th, 2022

Outline

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

Executive Summary

- This analysis uses a variety of analytic and data science techniques to predict the likelihood a SpaceX Falcon9 first stage booster will successfully land
- Data was collected from public sources
- Four data science models supported by Grid Search Cross-Validation were all statistically significant in predicting Falcon9 booster results
- Logistic Regression, Support Vector Machines, K-Nearest Neighbor, and Decision Tree model each had favorable results

Introduction

- SpaceX has become a major player in the space industry
- Some of the best prices to deliver a payload orbit is offered by SpaceX
- Falcon9 launches are advertised at cost of $62 million dollars, significantly less than comparable competitors
- Reusable first stage rocket boosters is a major contributor to cost savings
- In this analysis, we will explore the data of reusable first state boosters, and predict the likelihood that any given Falcon9 launch will successfully land its first stage

Section 1

Methodology

Methodology

Executive Summary

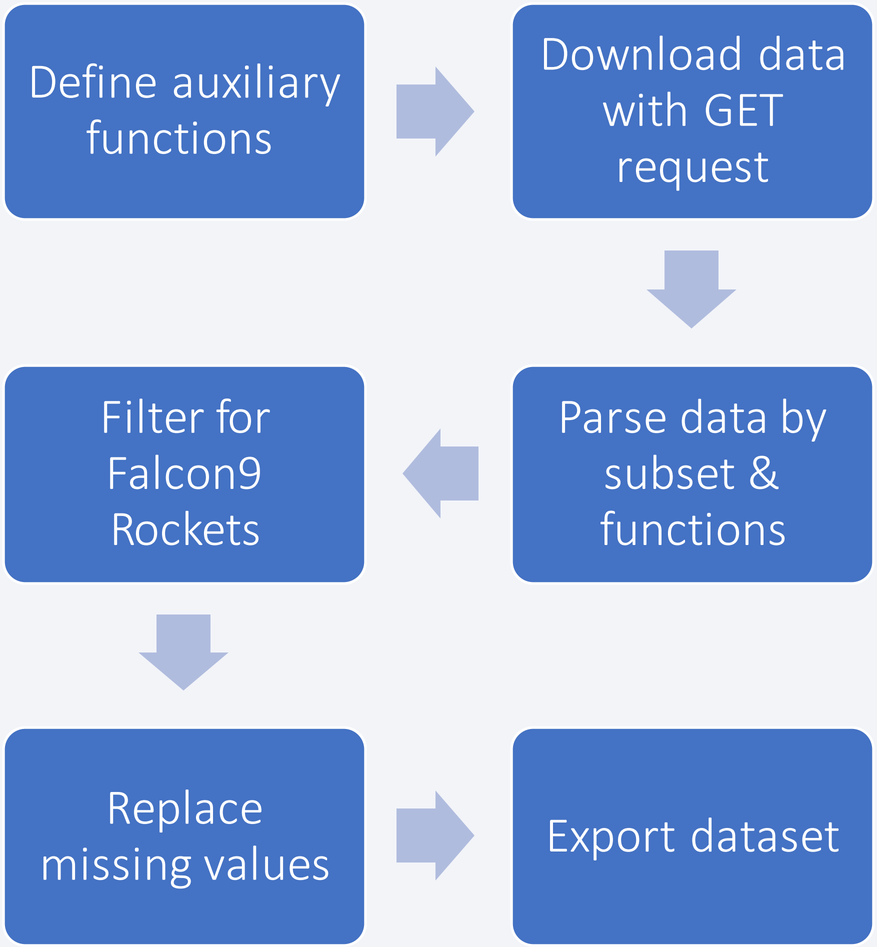
- Data collection methodology
  - Data was collected from public sources: SpaceX API and Wikipedia
- Perform data wrangling
  - Data was cleaned via using EDA to inform labelling and transformations of data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- SpaceX data launch is publicly available via the SpaceX API
  - To collect data, Python code was written using the built-in 'requests' library
  - Data includes information on the date, time, launch payload, capsule type, rocket type, etc.
- SpaceX data is also available on Wikipedia via web-scrapping
  - Data on Falcon9 launches was web-scarped from Wikipedia using the BeautifulSoup, a Python library with powerful web-scraping capabilities
  - Useable data was extracted by Interacting with HTML data pulled by BeautifulSoup

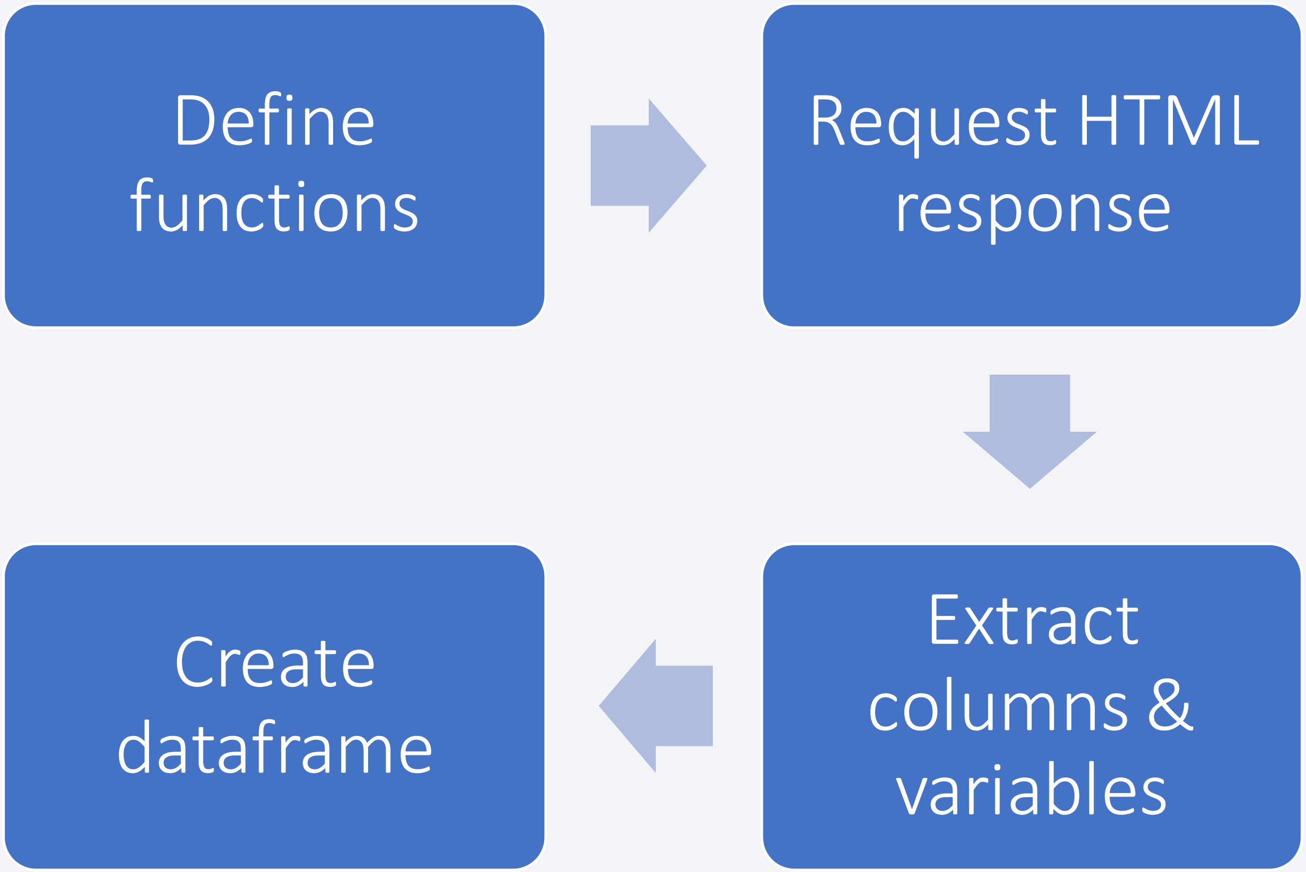
Data Collection – SpaceX API

- Using a GET request, we return data from the SpaceX API
- Using functions and Pandas library, we reduce the dataset
- Replace missing values with an average
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/Data%20Collection%20API.ipynb



Data Collection - Scraping

- Using BeautifulSoup we return HTML data from Wikipedia
- Parsing the HTML response to extract relevant details
- Transform HTML into a useable dataframe
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Using Exploratory Data Analysis (EDA) we can get a better understanding of the data
  - Find the number of launches at each site from the total of 90
  - Find the number of each orbit type
  - Find the occurrences of outcome per orbit type
- Landing Outcome (of the first stage) is our target variable
  - Create a label for each launch called 'landing_class', identifying if it was a successful or not
  - Reduce the many outcomes provided by the data to a binary
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/Data%20Wrangling.ipynb

EDA with Data Visualization

- By studying the relationships between variables in our dataset with visualizations, we can identify a number of features about our data. The visualizations show us:
  - The success of landings at each site
  - Which sites launched flights with high or low payloads mass
  - Which orbits have higher landing success rates
  - Payload mass and launch sites relationship to the flight number
  - If the size of the payload has a positive or negative relationship with successful landings, including a breakdown by each orbit type
  - The yearly landing success trend
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

Using SQL allows us to return additional insights into the data. Queries include:

- The unique launch sites
- Five records where the launch site begins with "CCA%"
- Total payload mass of boosters launched by NASA
- Average payload mass launched by the F9 v1.1 Booster
- The date when the first successful landing took place
- The names of boosters which have successfully landed on drone ships with a payload mass between 4000 and 6000 kg
- The total number of successful and failed mission outcomes
- The names of the boosters which carried the maximum payload mass
- Failed outcomes on drone ships, including the booster version and launch site since 2005

- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

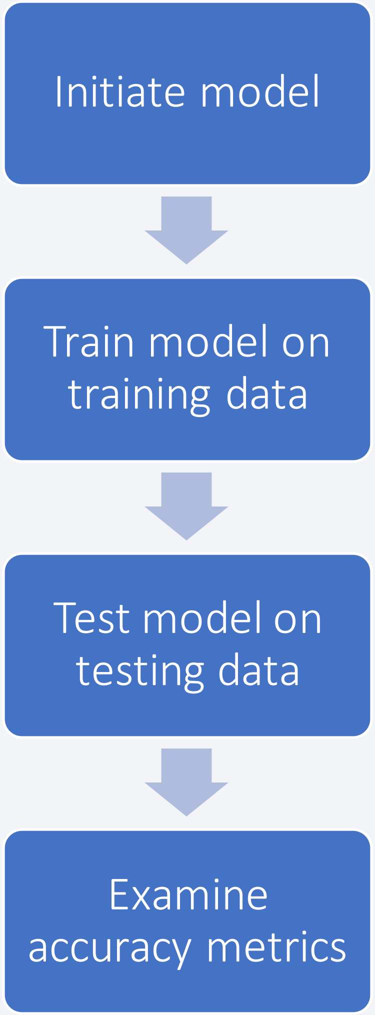
- Folium mapping allows us to visualize launch sites and data with Jupyter Notebooks
  - Mark all launch sites on a map and the successful or failed launches at each site
  - Calculate the distance between the California launch site to a nearby coastline, railroad, city, and highway. Draw a line and display the distance to each of these nearby locations
  - Visualizing this allows us to interactively explore launch data, and well as get an understanding of where launch sites are relative to other locations
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/Launch%20Sites%20Location%20Analysis%20with%20Folium.ipynb
  - *A note to reviewers: Folium maps do not display correctly in the GitHub viewer. To see the maps, please download the .ipynb file or referance the screenshots included later in this presentation.*
    Source: https://github.com/python-visualization/folium/issues/1072

Build a Dashboard with Plotly Dash

- Using Ploty Dash two interactive plots were created
- The first plot show the Successful Launches by Site in a pie chart. Any individual site can be selected to see the percentage of successes
- The second plot shows the relationship between Payload and Successful Launches, this plot can be filtered for individual sites and for the Payload kg
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/spacex_dashboard_app.py

Predictive Analysis (Classification)

- Four Machine Learning models were built: Logistic Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors
- The best model parameters were found using Grid Search Cross Validation
- Accuracy metrics were used to identify the performance of each model
- https://github.com/DericGreen/SpaceX-Data-Science-Project/blob/9ec5cb8474de76365864f9d415b65ddd750f8f75/Machine%20Learning%20Prediction.ipynb



Results

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

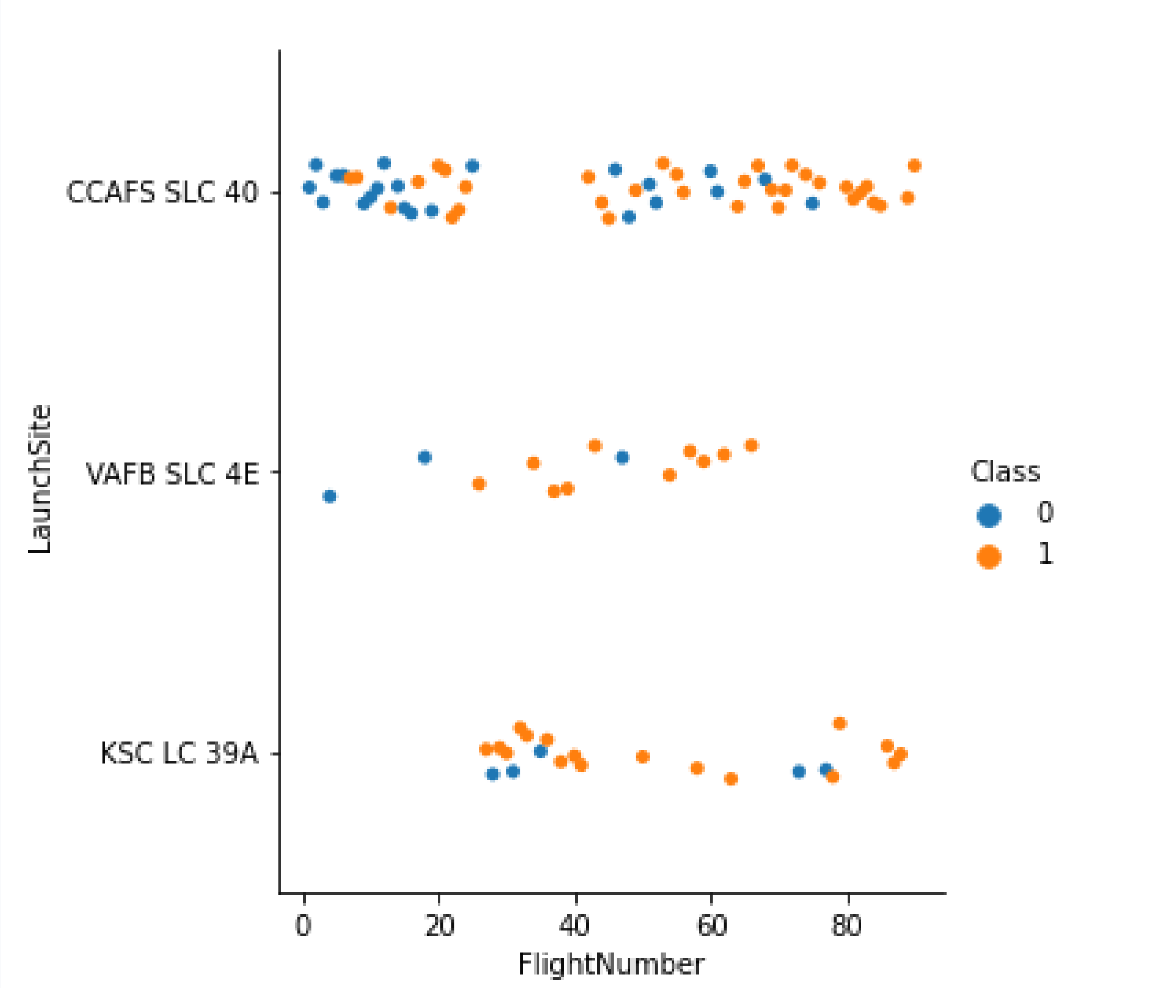
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Site VAFB SLC 4E has had fewer Falcon9 Flights than other sites

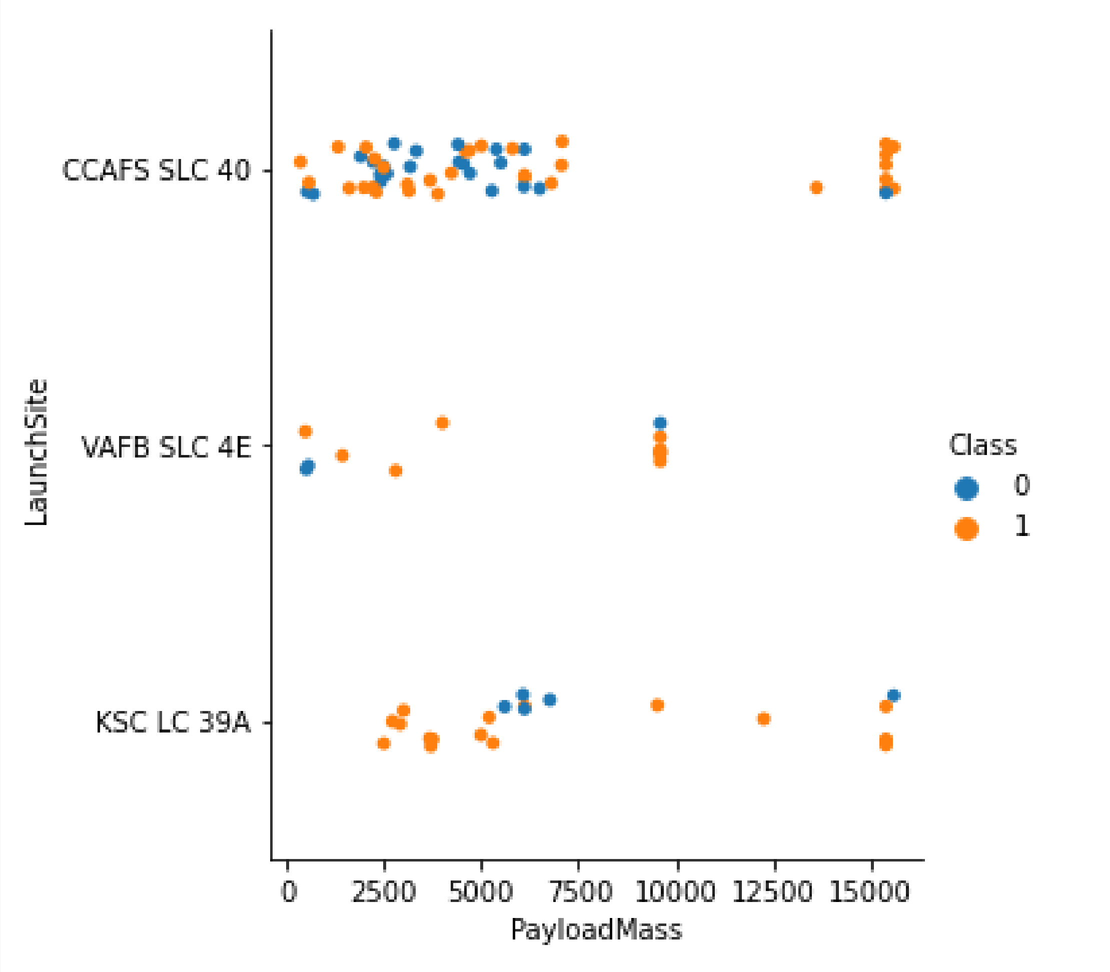
- Successful launches (class=1) is positively correlated with the Flight Number



Payload vs. Launch Site

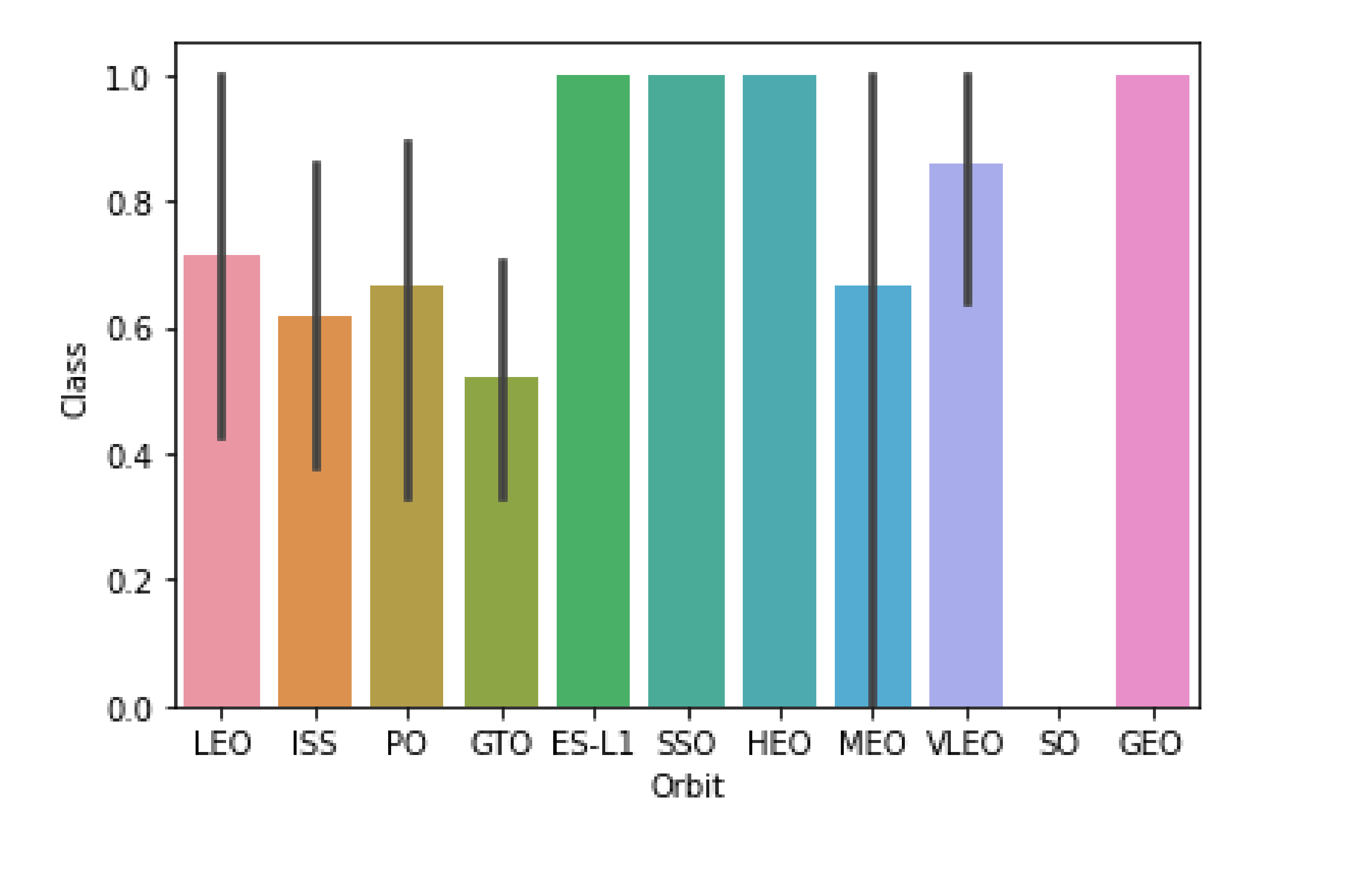
- Site VAFB SLC 4E does not accommodate heavy payload mass launches

- Generally, heavy payload mass launches have are successful (class=1)



Success Rate vs. Orbit Type

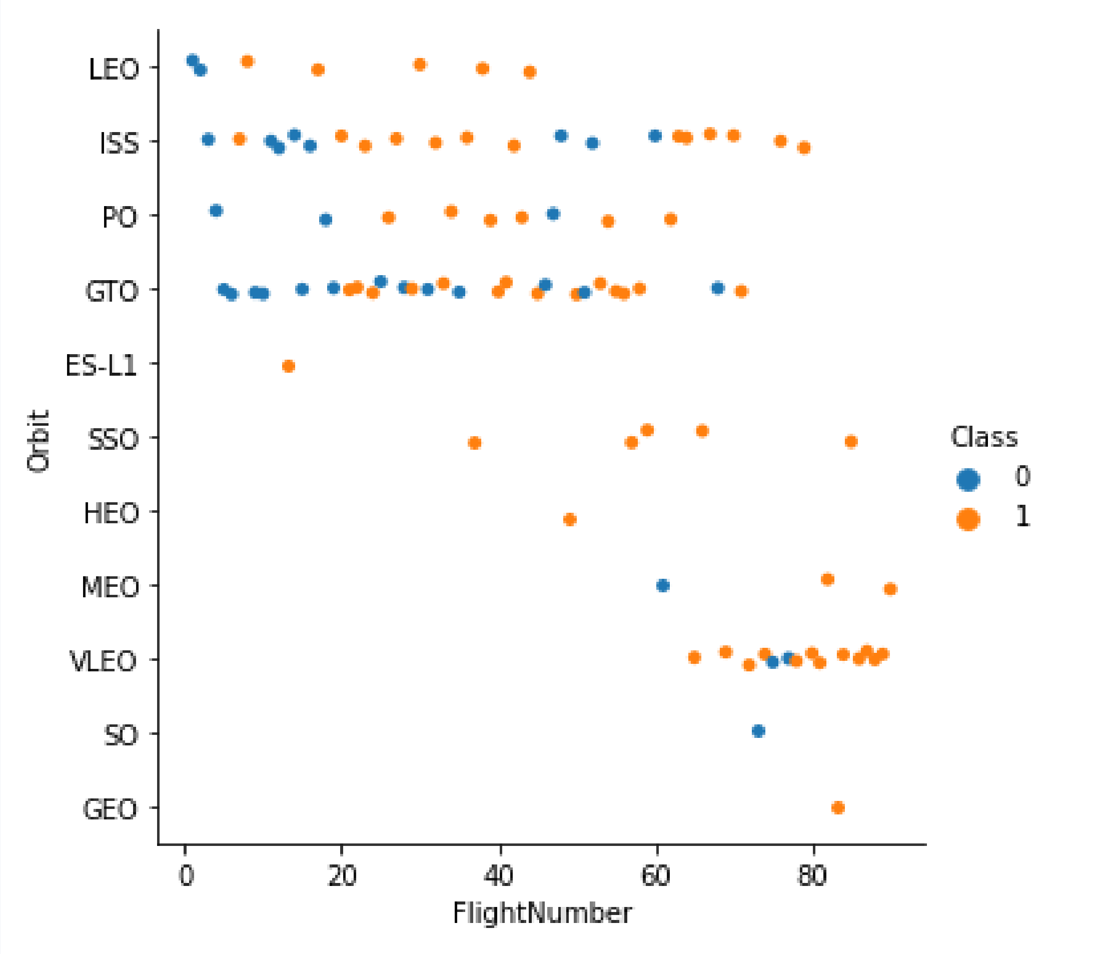
- Particular orbit types, like ES-L1, SSO, HEO, GEO, and VLEO have a high success rate



Flight Number vs. Orbit Type

- LEO successes appear to be positively related to the flight number

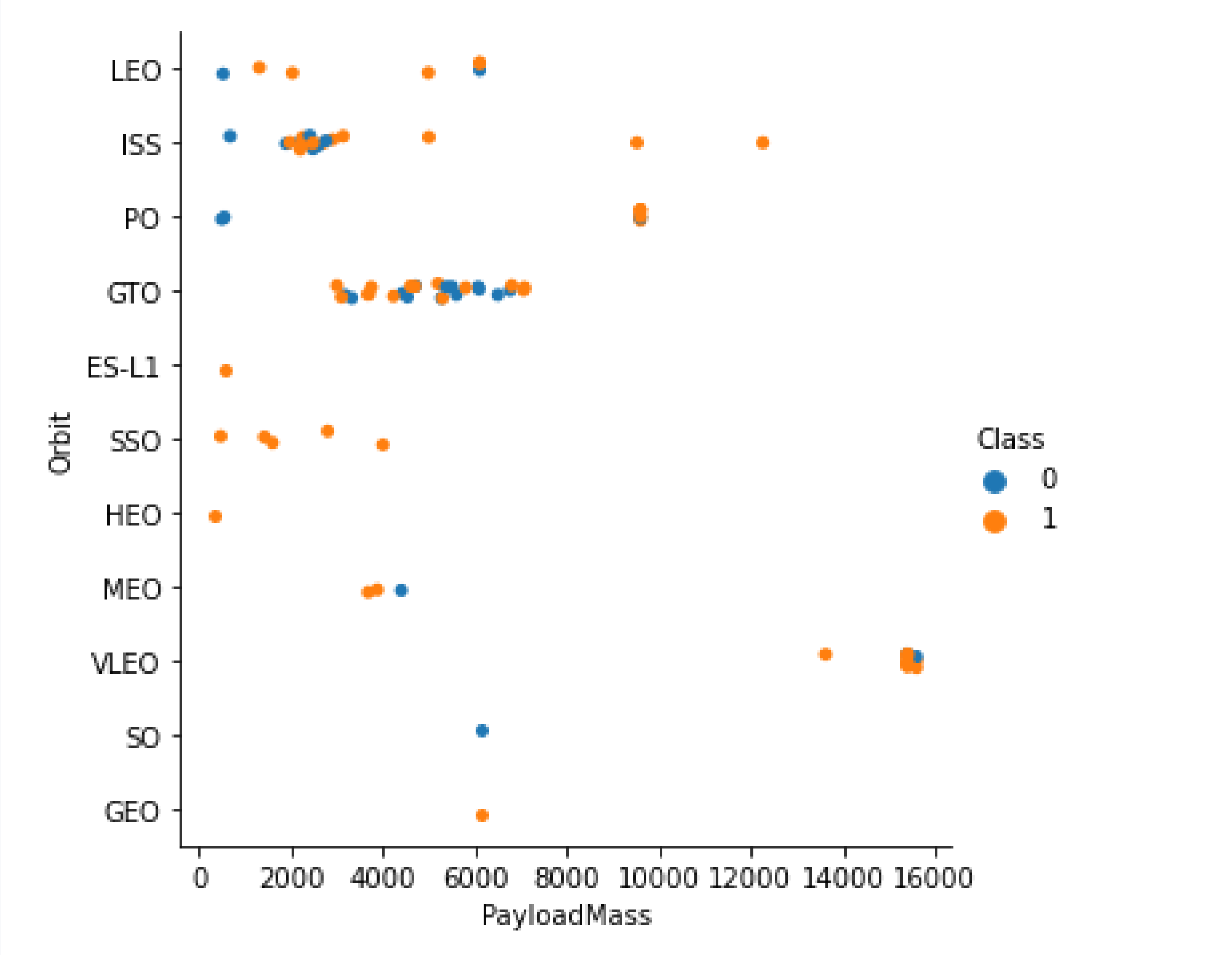
- VLEO launches are plentiful but only occur later in the flight numbers



Payload vs. Orbit Type

- GTO orbit types plentifully fall into a small range of payload mass, however the success is undetermined

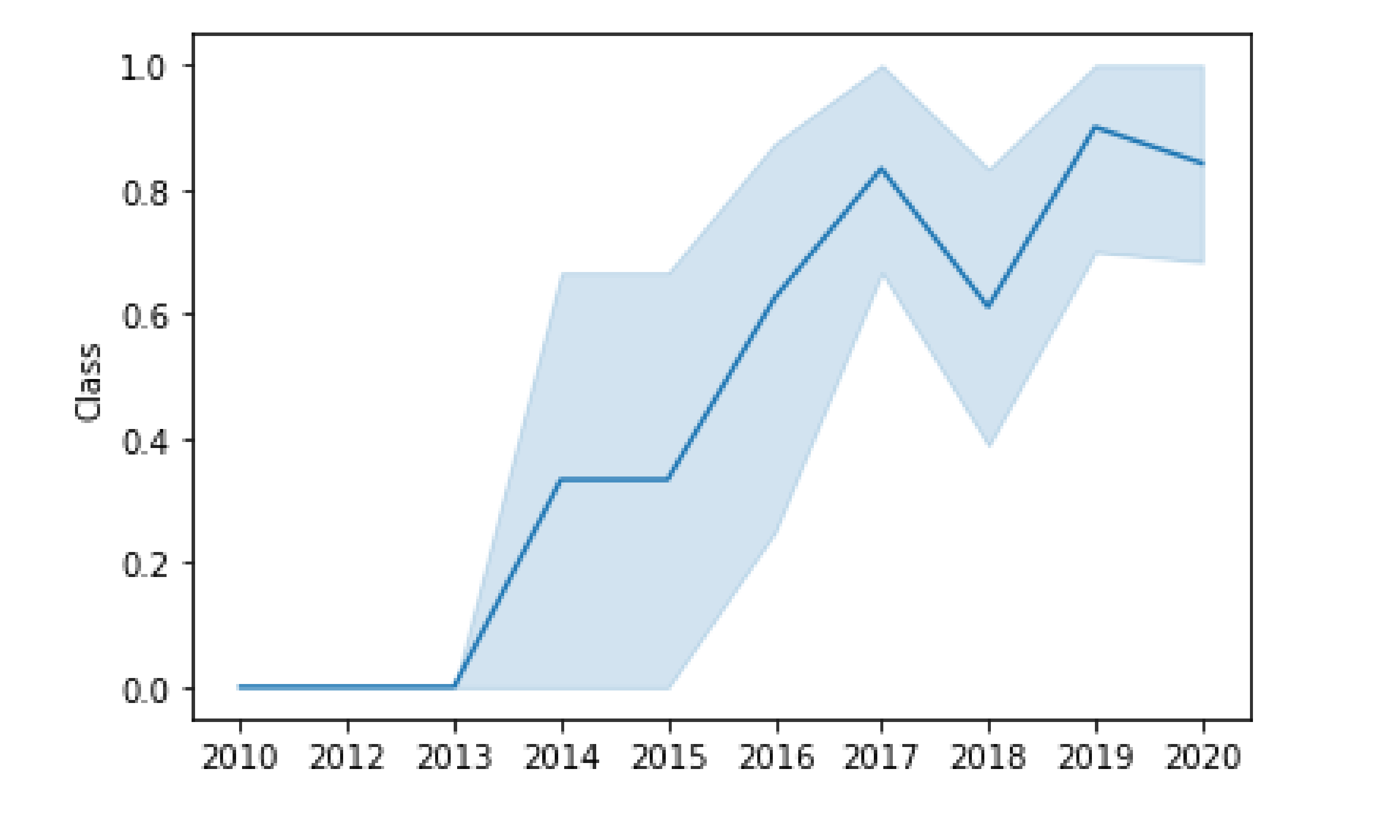
- LEO, ISS, and VLEO launches have more successes at higher payloads



Launch Success Yearly Trend

- The success rate of launches has been trending up since 2013

- Success rates dipped in 2018 and 2020 compared to the trend



All Launch Site Names

- An SQL query There are 4 unique launch sites in the analysis

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

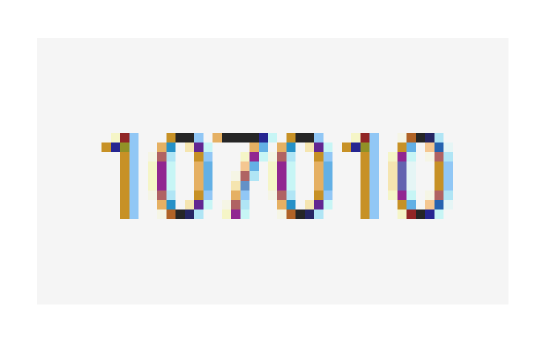
Launch Site Names Begin with 'CCA'

- The below table shows 5 records where the launch site begin 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- In total NASA has launched 107,010 kg of mass on Falcon9 Rockets in our dataset's timeframe.



Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2,928 kg

First Successful Ground Landing Date

- The first successful ground landing occurred om 2018-07-22

DATE
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four booster versions have successfully landed on a drone ship with a 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

- Of the missions explored in the dataset, only 1 has failed
- Note that mission outcome refers to the result of the mission, not the first stage ground landing analyzed elsewhere in this presentation

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- A number of F9 B5 booster versions have carried the maximum mass payload of 15600 kg

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

2015 Launch Records

- In 2015 there were 3 failed landing outcomes on drone ships

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40	2015-06-28

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

- This list shows the count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

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

Section 3

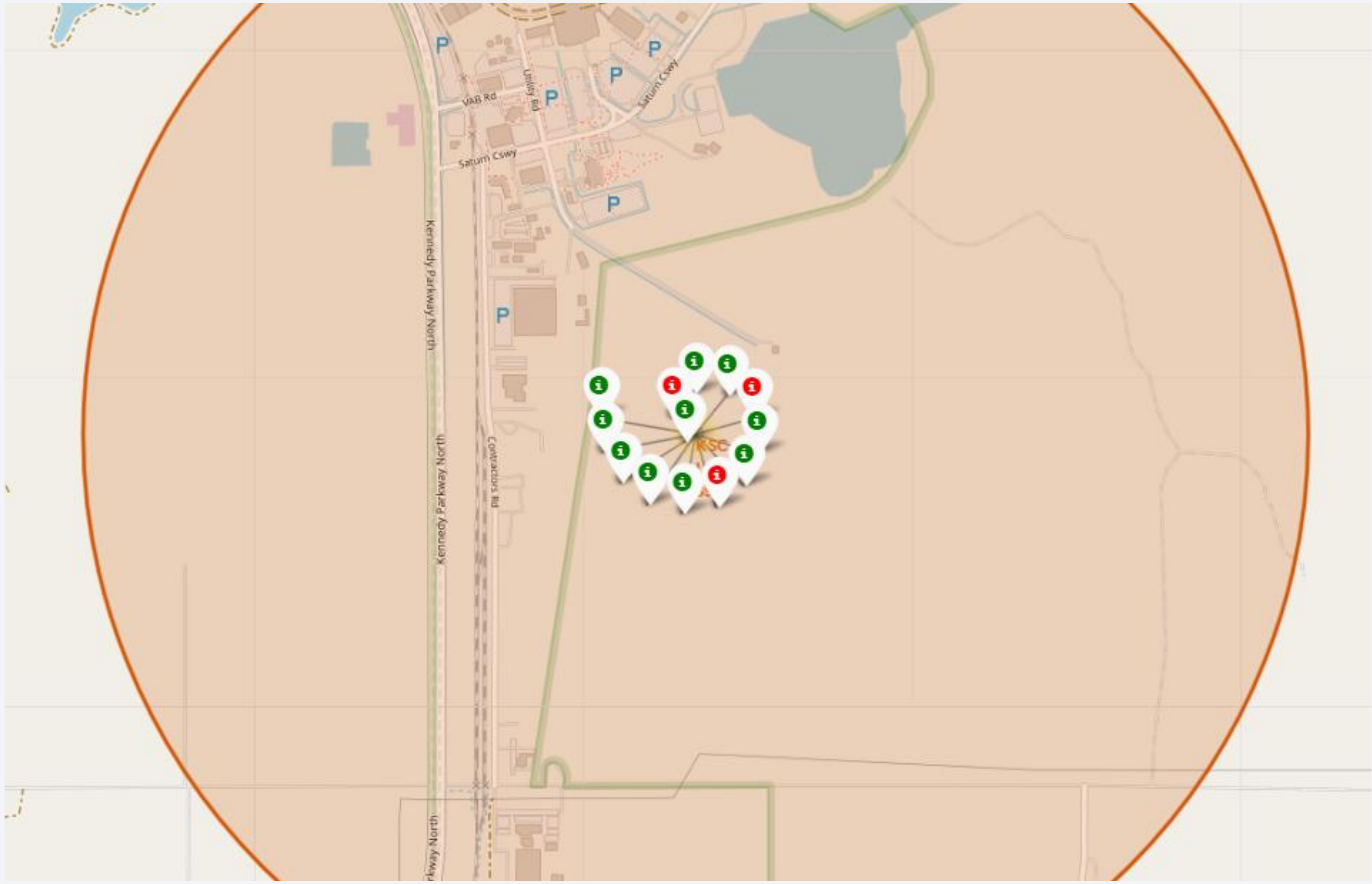
Launch Sites Proximities Analysis

Launch Site Locations in Folium

- We can identify the locations of launch sites in a Folium map. The sites are found on coasts of the United States

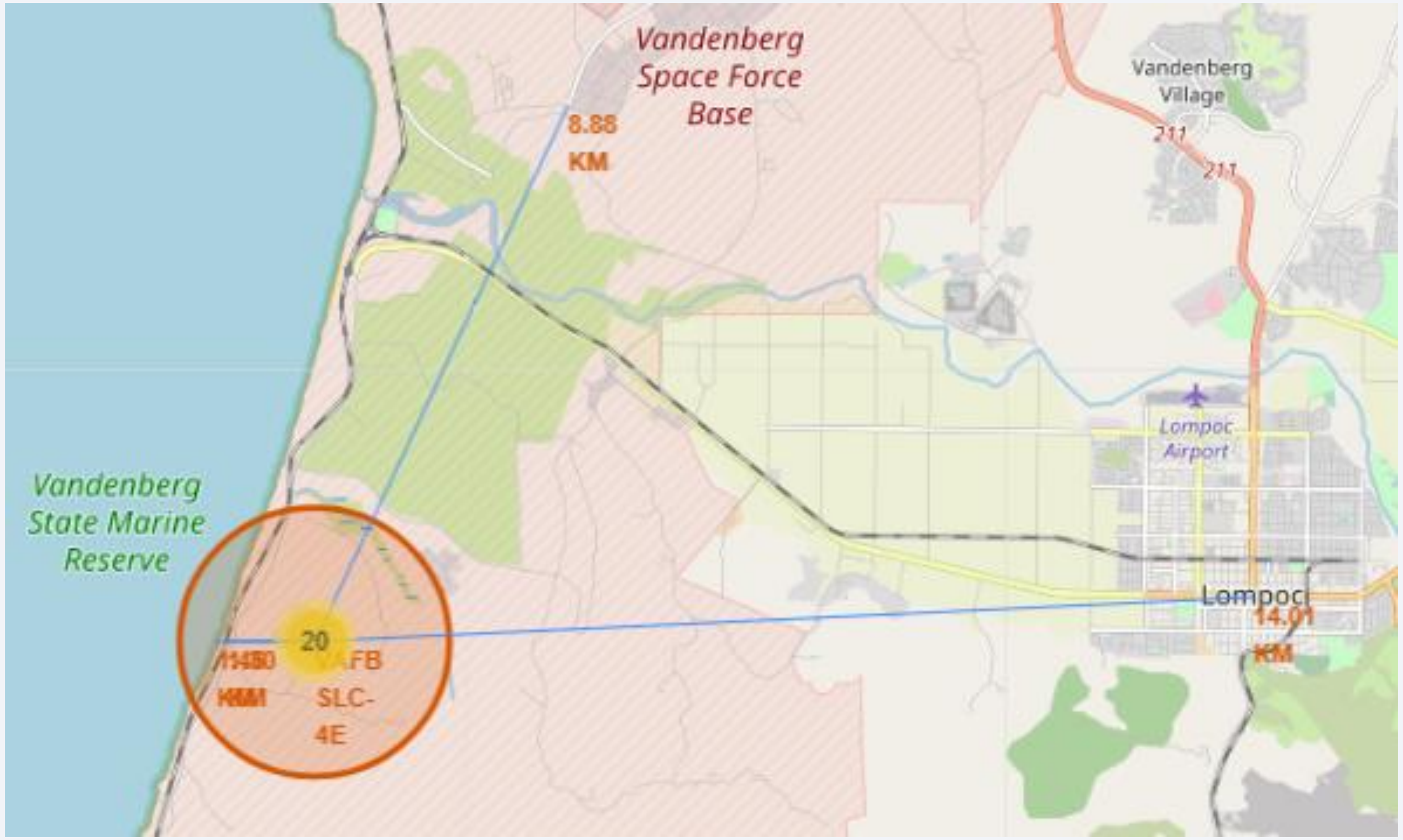
Launch Outcomes in Folium

- We can visualize the launch outcomes for any site in the dataset.
- This screenshot shows us the result of launch outcomes at the KSC LC-39A launch site in Florida



Launch Site Proximity to points of interest

- We can visualize the distance from any given launch site to points of interest.
- This screenshot shows the distance from the California launch site VAFB SLC-4E to nearby railways, coastline, cities, and highways. We can display the distance to these points.

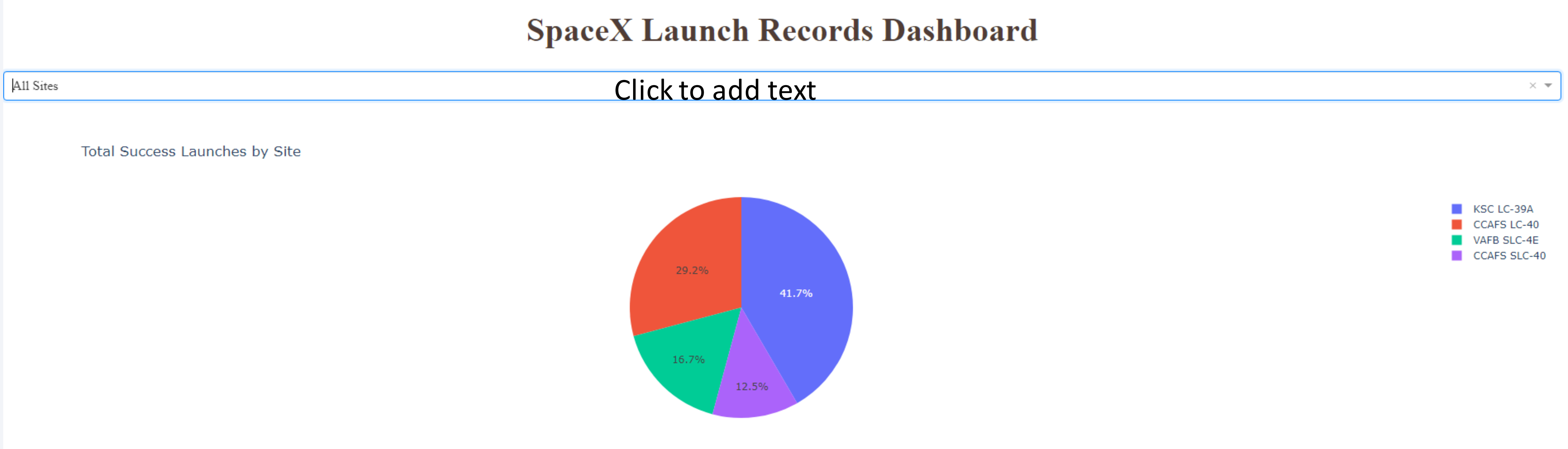


Section 4

Build a Dashboard with Plotly Dash

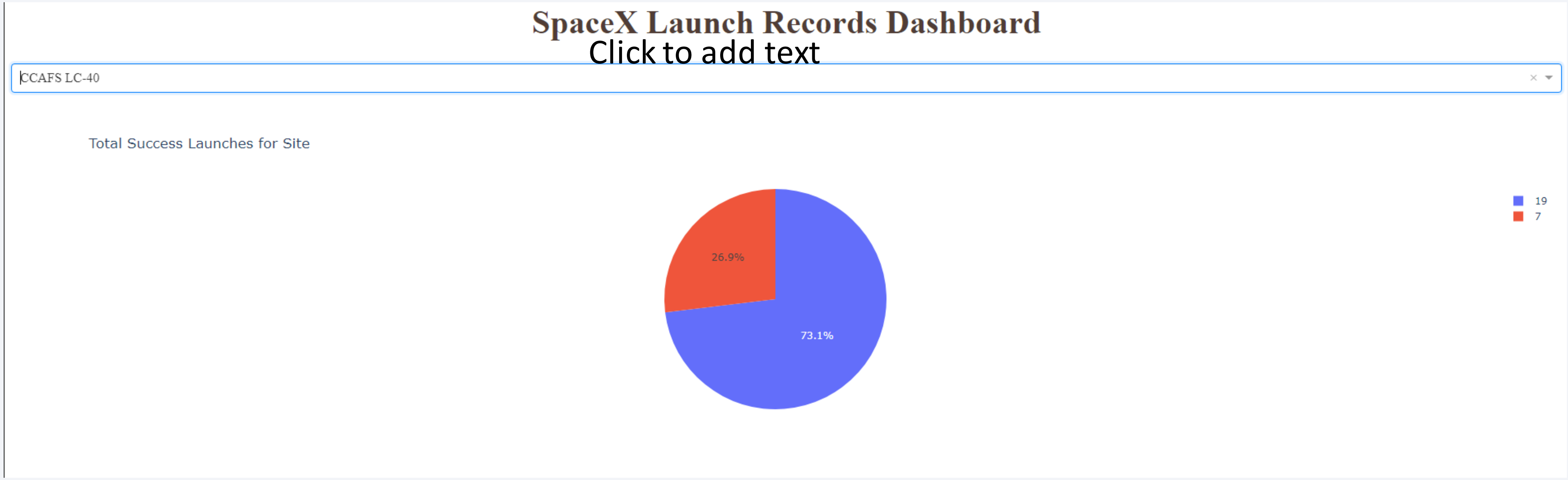
Launch Success by Launch Site

- Using Plotly Dash, we can create an interactive dashboard
- This screenshot of the dashboard shows us launch success rate at each launch site



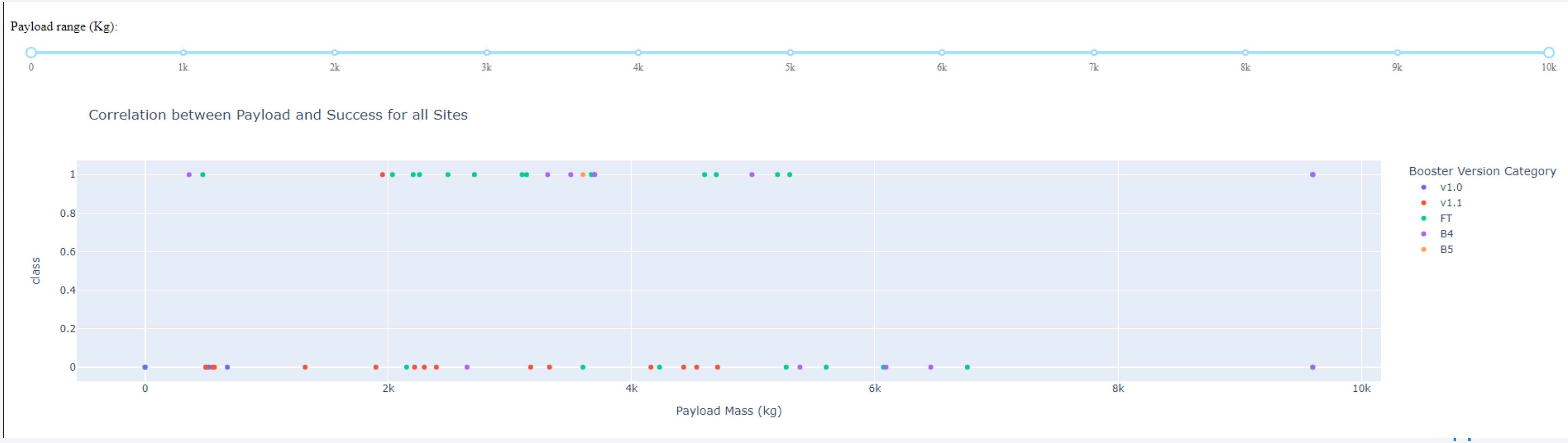
Launch Success at individual sites

- By interacting with our dashboard, we can see the individual rate of launch success at each launch site.
- This screenshots shows the launch success percentage at CCAFS LC-40



Interactive visualization of Payload v. Launch Outcome

- We can visualize the payload (kg) against the outcome class. We can also see the different booster versions through coloring.
- By adding a payload slider and site selection, the user can interactively drill down for the precise information they are looking for.

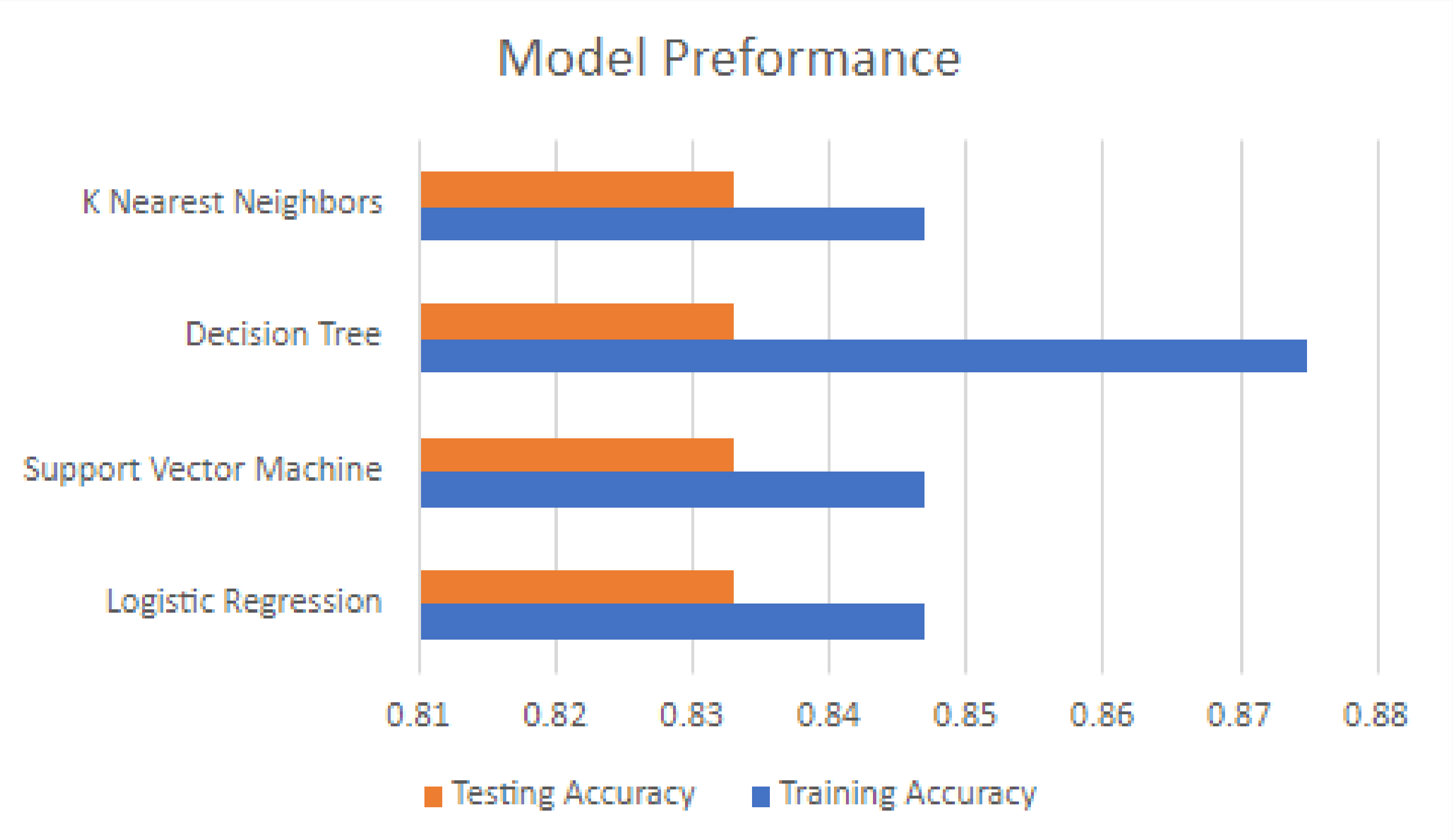


Section 5

Predictive Analysis (Classification)

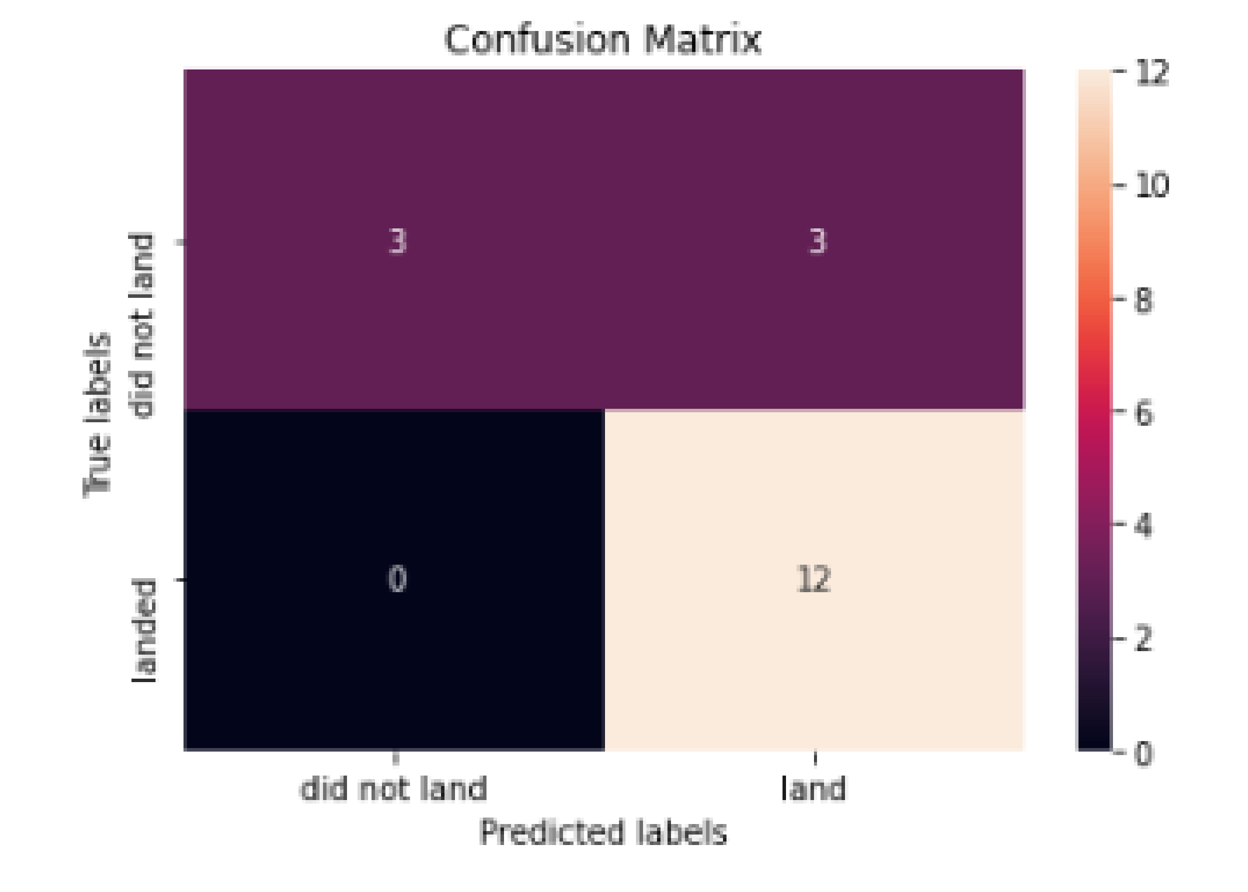
Classification Accuracy

- Of the 4 models built to predict the landing success of SpaceX Falcon9 rockets, all had the same accuracy on training data of 0.833
- The decision tree model did perform better on the testing data



Confusion Matrix – Decision Tree Model

- Each model performed well and would be a suitable choice, however analyzing the decision tree model here is suitable.
- The alongside confusion matrix shows the result of the decision tree model.
- The model does predict a few cases of failure when the launch actually succeeded, but overall is a good model



Conclusions

- The success of SpaceX Falcon9 landings can be modelled with statistical significance
- The models of these results can further inform SpaceX launches, and given the cost savings of landings the employment of these models could be significant for informing business decisions
- A number of different models could be used to inform business decisions.

Appendix

All code, pictures, and notebooks in this analysis are available for peer review at:

https://github.com/DericGreen/SpaceX-Data-Science-Project

This analysis was completed as part of the requirements for the IBM Data Science Professional Certificate by Deric Green.

https://www.coursera.org/professional-certificates/ibm-data-science

Thank you!