IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christopher Crawford
1/16/2025

https://github.com/ChrisACr/CapstoneRepo.git

Outline

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

Executive Summary

Summary of methodologies

- Data Collection API
- Web Scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Visualization (i.e. Pandas, Matplotlib)
- Visual Analytics with Folium (interactive maps)
- Interactive Dashboard
- Machine Learning Predictive Analysis

Summary of all results

- Nearly 100% Mission Success with estimated 66% Successful Landings
- Most flights had a Moderate to Low Payload Mass
- Payload Mass heavily Influenced Orbit type
- Launch Sites close to Coast Lines and far from Cities
- Decision Tree model slightly out preformed other models

Introduction

The SpaceX company has revolutionized rocket launches due to their ability to reuse the first stage of the rocket by having it land back on earth safely. By doing this, they have significantly lowered the price of rocket launches with their successful recovery of the first stage a launch would cost approximately $62 million dollars. Opposed to the $165 million dollar price tag that others must pay from not recovering the first stage.

Problems you want to find answers

- How features such as orbit and payload affect the success rate of successful landings
- Find a way to predict if a landing will be successful based on the features

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

Request rocket launch data from SpaceX API: https://api.spacexdata.com

Web Scrape Falcon 9 historical launches from Wikipedia: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Perform data wrangling

Find patterns in the data and determine the label for training supervised models

Perform exploratory data analysis (EDA) using visualization and SQL

Understand the dataset and load it into a DB2 database

Perform interactive visual analytics using Folium and Plotly Dash

Explore launch locations on a map

Build a dashboard analyzing feature comparison

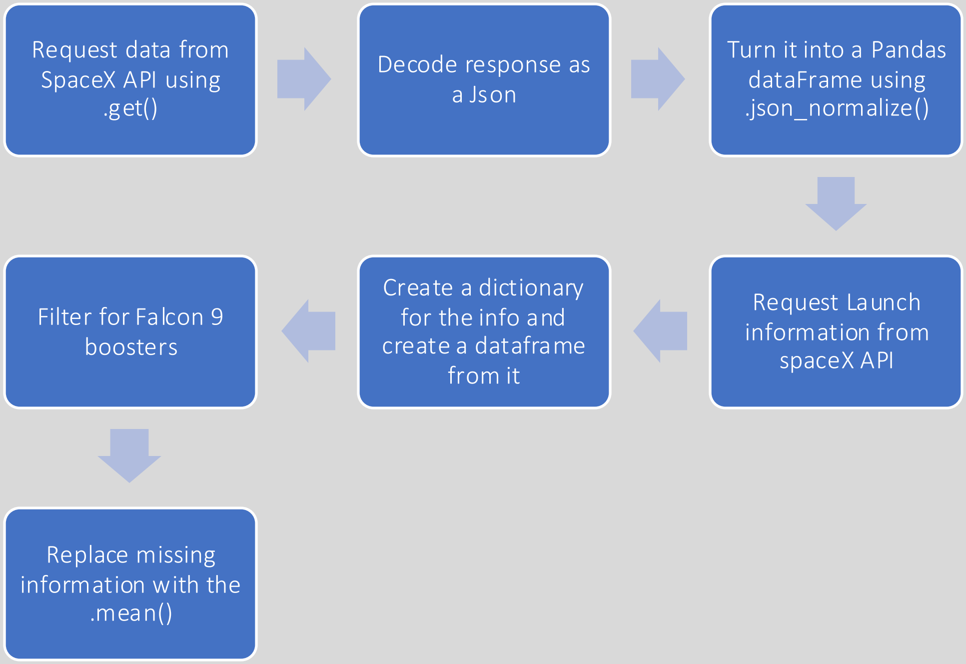
Perform predictive analysis using classification models

Use the data to train and test logistic regression, support vector machine, k-nearest neighbor, and decision trees to predict the successful landings

Data Collection – SpaceX API

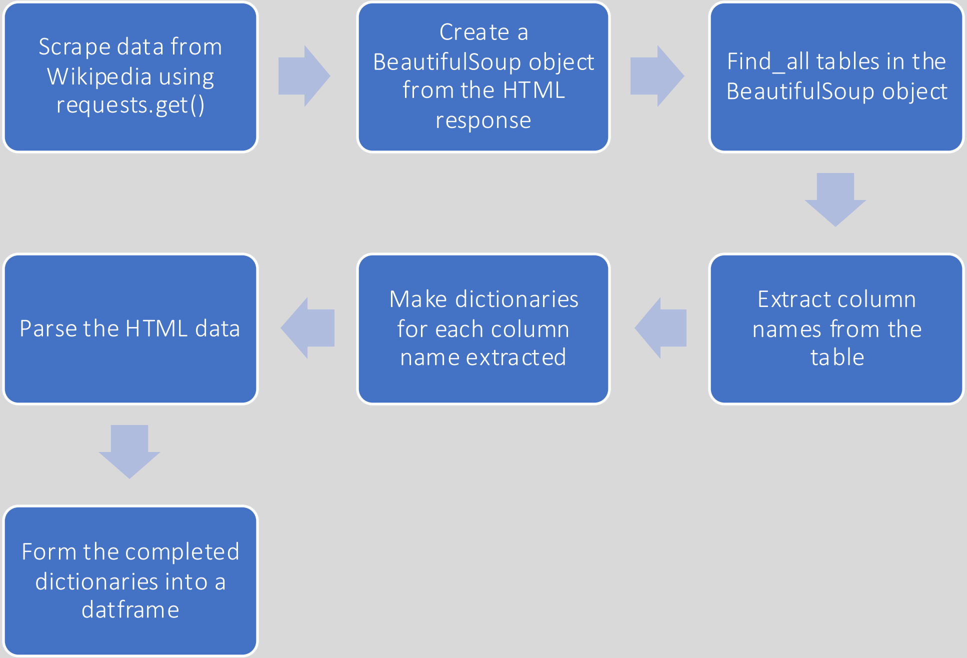
- Data collection came from the SpaceX API: https://api.spacexdata.com. Using the first request to get launch IDs and the second to get launch information about the IDs. Once it was properly normalized and placed into a dataframe, filtering for Falcon 9 booster and replacing missing values can begin.

- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/jupyter-labs-spacex-data-collection-api.ipynb



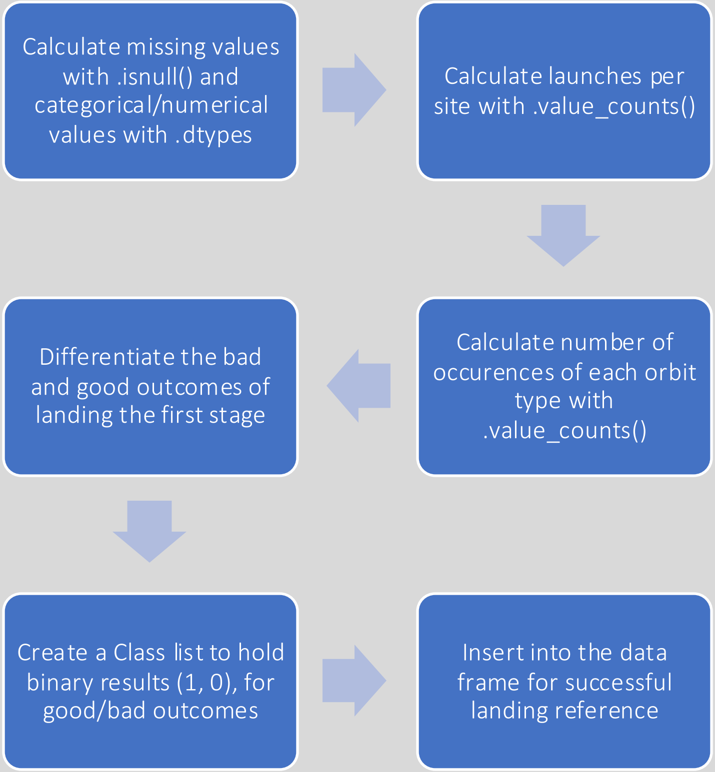
Data Collection - Scraping

- Scrape data with requests.get()
- Create BeautifulSoup object
- Use find_all() to locate all tables
- Iterate through <th> elemtns to gather colunm names
- Make dictionaries for each column name and fill with parsed HTM<L data
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/jupyter-labs-webscraping.ipynb



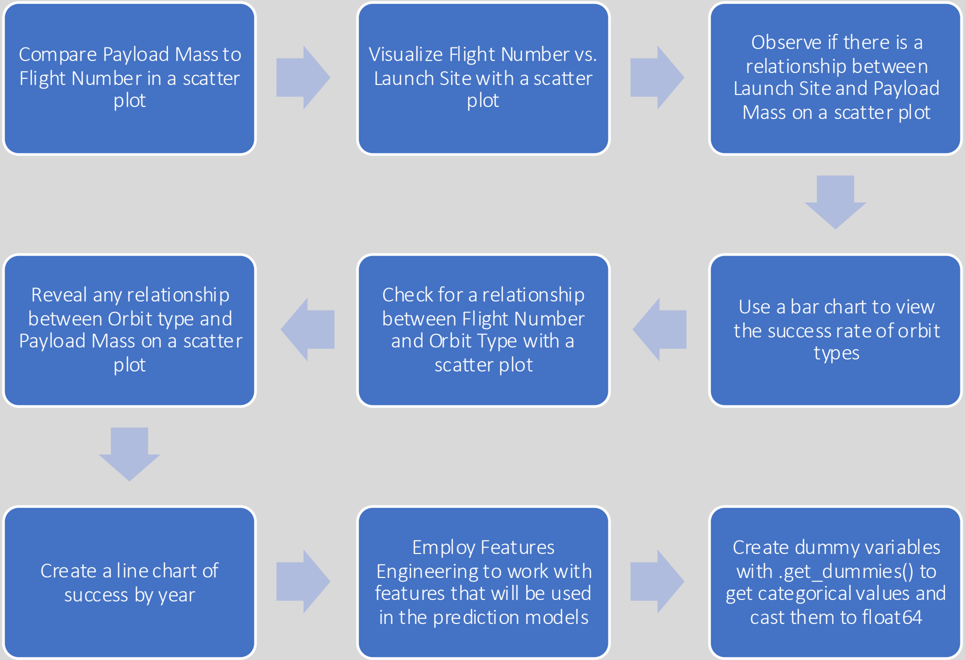
Data Wrangling

- After identifying missing values and noting the data types, we calculate the launches per site and occurrences of each orbit type. Since there are 7 types of landing results, with only 3 being successful, we filter out the unsuccessful ones by using their key number. Then in a new list, apply successful landings "1" and unsuccessful "0". Which gets added to the dataframe as a new column 'Class'
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- In order to analyze relationships between Payload/Flight Number, Flight Number/ Launch Site, Launch Site/Payload, Orbit/Payload, and Flight Number/Orbit, scatter plots were used as they are a great method for identifying possible relationships between variables. A bar chart for the Orbit/Success as it shows good distribution. A line chart for success by year to see the upward trend.
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/edadataviz.ipynb



EDA with SQL

- Determine the names of the unique launch sites for SpaceX
- View records of launches at CCA launch site
- View total payload mass carried for customer NASA (CRS)
- View average payload for booster version F9 v1.1
- Find the date of the first successful landing in ground pad
- List booster names that have success in drone ship with payload between 4000 and 6000
- List total number of successful and failure mission outcomes
- List boosters that have carried maximum payload capacity
- Find the month, booster version, and launch site of failure drone ship in 2015
- Rank the count of landing outcomes
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

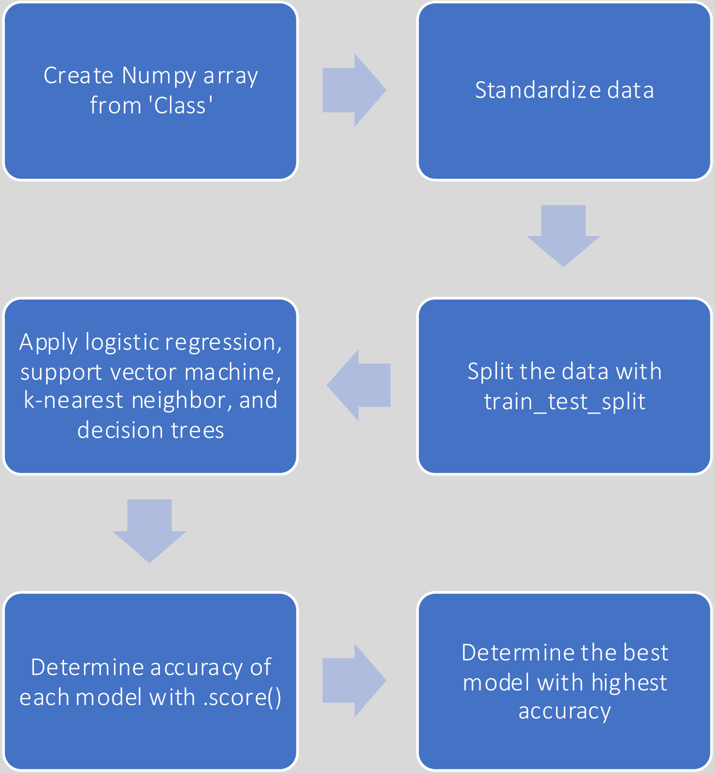
- With folium I added a circle and marker for locations of four different SpaceX launch sites. Then a cluster on each site representing the successes and failures. Next, I calculated the distances between the launch sites and respective coastlines, railroads and highways by placing a line from launch site, to the closest city, coastline and highway.
- The circles around the launch sites were to boldly demonstrate the area, as well as encapsulate the cluster of successes and failures at each respective site. Next the lines helped to calculate the distances that each site was away from coastlines, cities, and highways.
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- On the plotly Dashboard, the first thing was implementing a dropdown menu. Next there are pie charts in order to display successes at different launch sites. Lastly there is a combination of a range slider for payload mass and a scatter plot revealing success compared to payload mass.
- The dropdown menu allows the user to choose between viewing statistics from all site combined, or any individual site selected. The pie chart vividly displays and compares the quantity of success per launch site. Lastly the range slider and scatter plot make it easy to view specific payload ranges as well as compare successes that may depend on payload.
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/spacex_dash_app.py

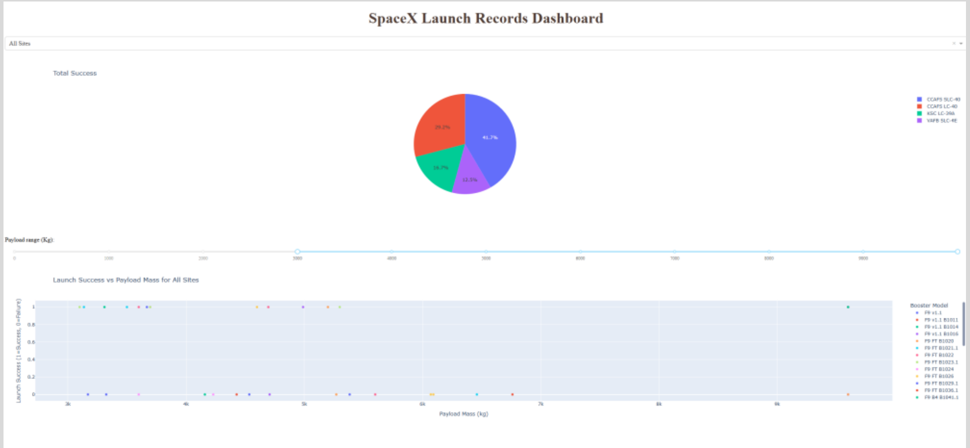
Predictive Analysis (Classification)

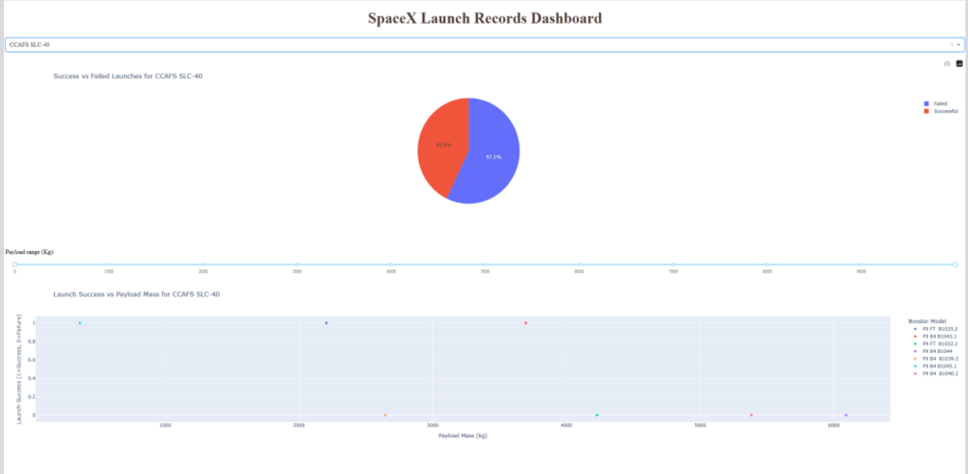
- Create the dependant variable Y from the feature 'Class'. Standardize the X data with StandarScalar(). Split into X_train, Y_train, X_test, Y_test using train_test_split. Formulate models including logistic regression, support vector machine, k-nearest neighbor, and decision trees. Calculate the accuracy of the models in order to determine which one should be used.
- https://github.com/ChrisACr/CapstoneRepo/blob/3e4c603cdf7d7ea9bff33bd884e09d1c72285856/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb



Results

- Exploratory data analysis results
  - It was discovered that SpaceX has a nearly 100% mission success rate and roughly a 66% successful landing of the first stage. Some orbits like GEO, HEO and SSO have 100% success while others like SO have only failure. Orbit type and payload mass significantly influence each other. Lastly, SpaceX has seen a nearly constant increase in success since 2010
- Predictive analysis results
  - Decission tree is the best model for prediction due to its higher accuracy than the others.

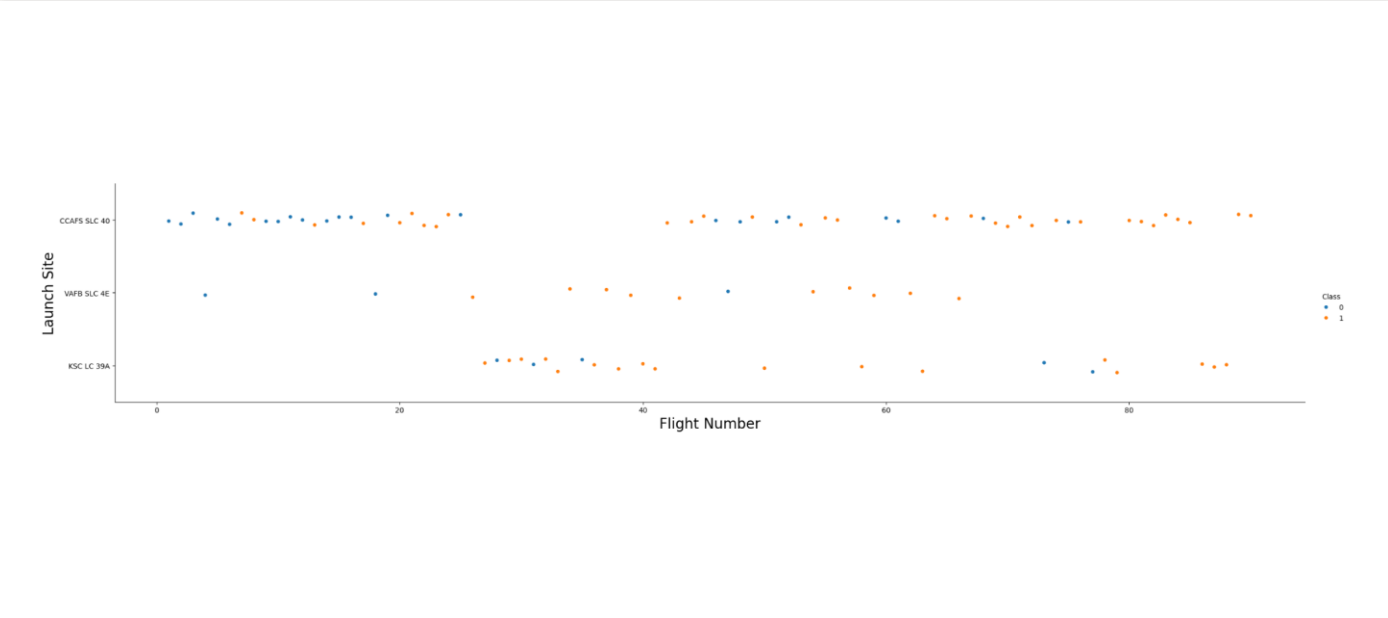




Section 2

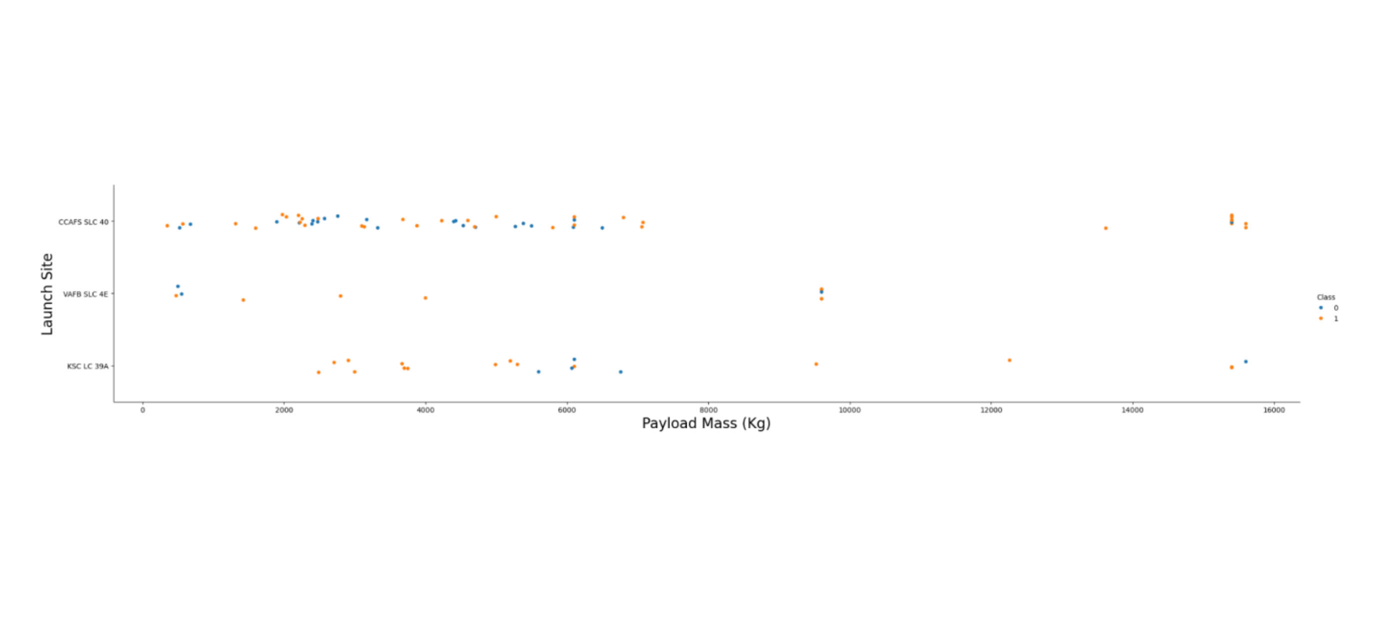
Insights drawn from EDA

Flight Number vs. Launch Site



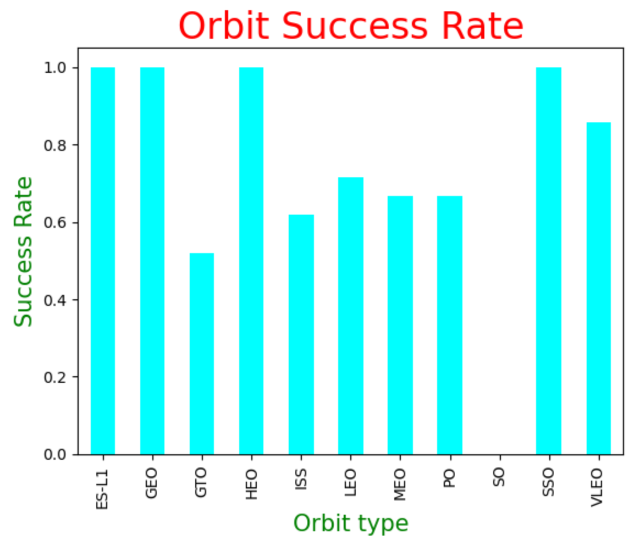
- CCAFS SLC 40 seems to be their most commonly used site
- All launch sites seem to increase success rate over time
- VAFB SLC 4E received the least launches, despite the high success

Payload vs. Launch Site



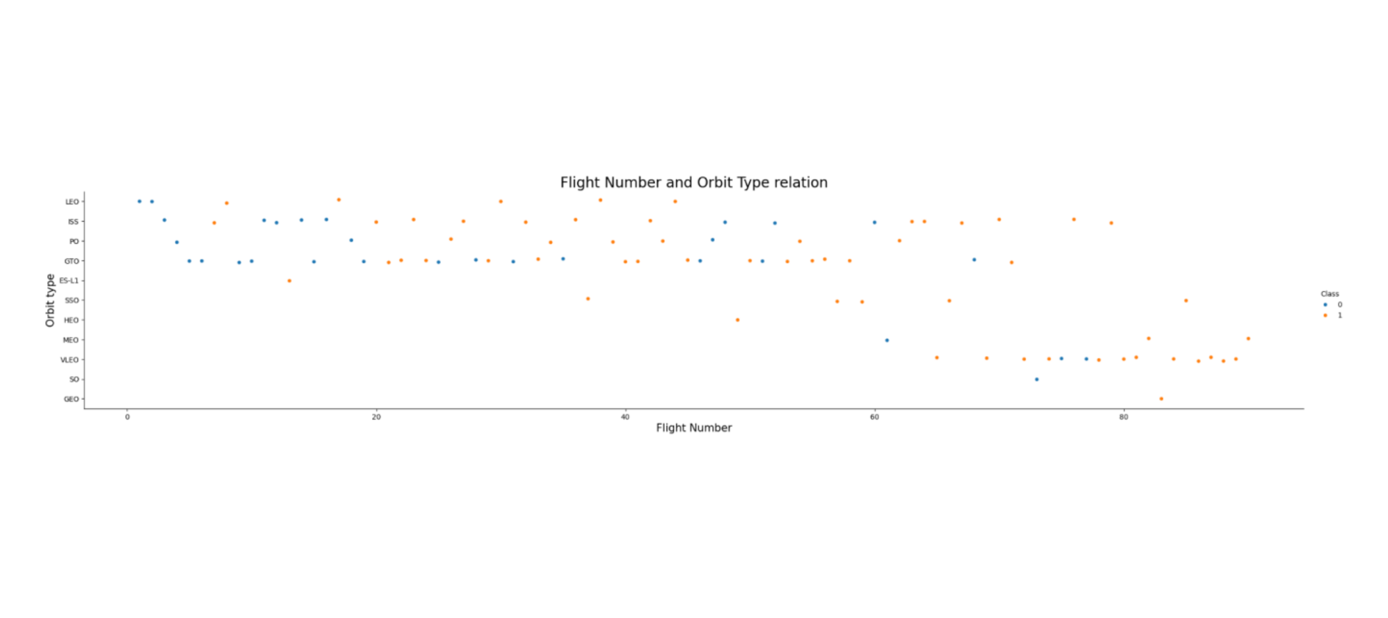
- Majority of payload mass was below 8000 Kg
- 2000 – 6000 Kg was the most common payload mass
- Minimal testing toward 10000 Kg still received some success

Success Rate vs. Orbit Type



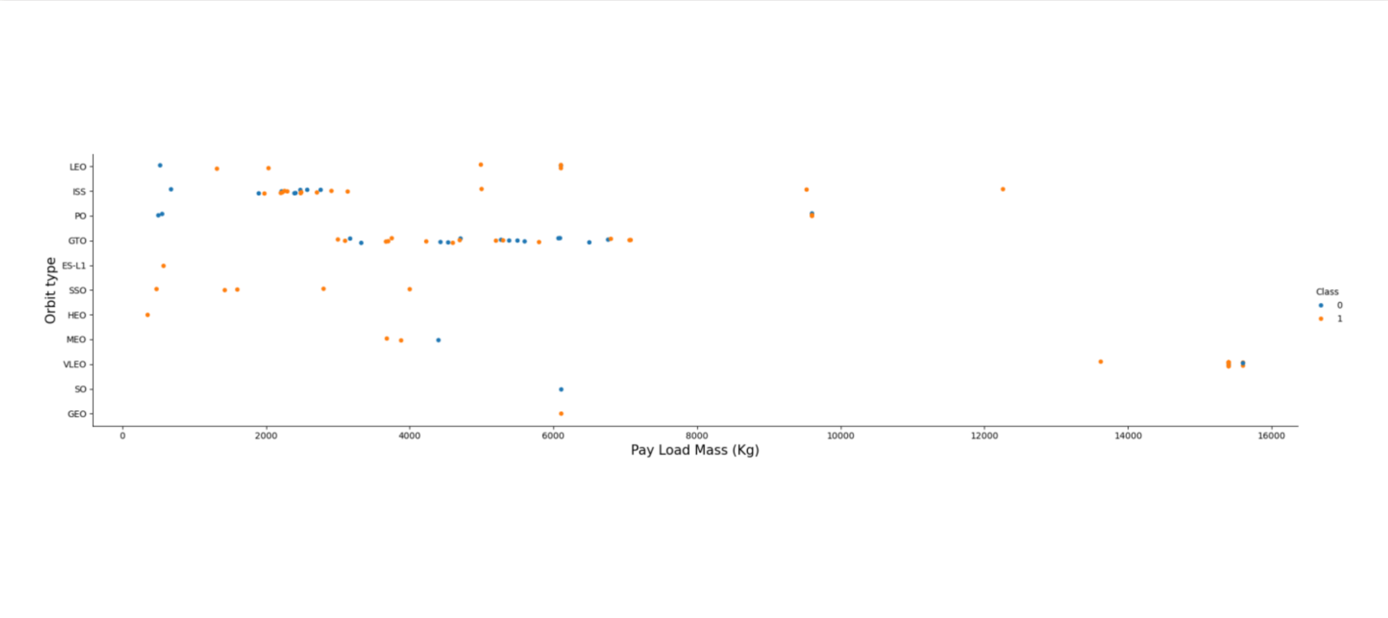
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate
- SO orbit has had no success
- Orbits GTO, ISS, LEO, MEO and PO range from 50% - 75% success with VLEO above 80%

Flight Number vs. Orbit Type



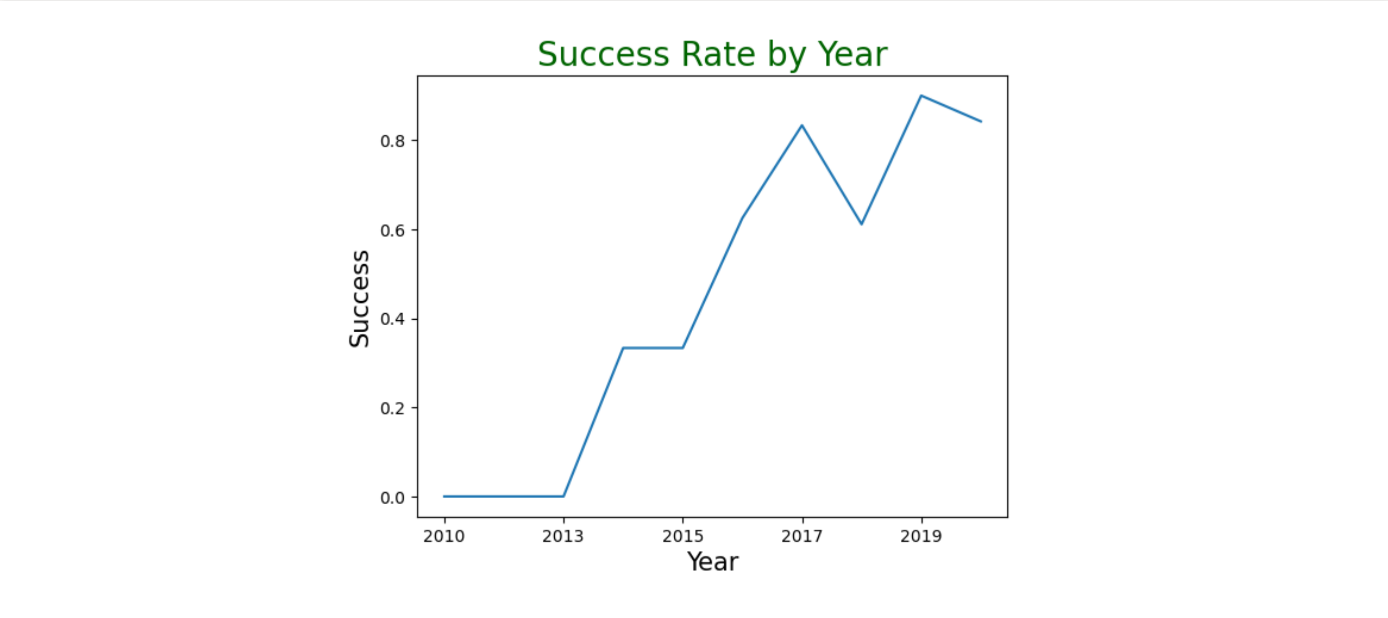
- For more than half of the first launches, LEO, ISS, PO and GTO were primarily used
- VLEO was not used for the first 60 launches; once used, yeilded success
- Aside from the orbits named above, the others were not used much

Payload vs. Orbit Type



- VLEO was the only orbit tested above 13000 Kg, with mostly success
- GTO and ISS have many launches
- At 6000 Kg, SO had 1/1 failure while GEO had 1/1 success

Launch Success Yearly Trend



- From 2013, SpaceX has shown a consistant increase in success
- Despite the 2018 decrease, they recovered and yeilded even higher success

All Launch Site Names

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

- The data shows us that there are 4 launch site locations used by SpaceX
- All of the lauches covered will be from these four locations

Launch Site Names 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	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

- These are the first 5 launch sites that come up beginning with 'CCA'

Total
48213

Total Payload Mass

- The total calculated payload mass in Kg for the customer NASA was 48,213 Kg.

Average
2534.666666666665

Average Payload Mass by F9 v1.1

- On average, the F9 v1.1 would have a payload mass of about 2535 Kg

MIN("Date")
2015-12-22

First Successful Ground Landing Date

- December 22nd of 2015 was SpaceX's first successful ground pad landing

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030

Successful Drone Ship Landing with Payload between 4000 and 6000

- These are the rocket booster versions that had successful drone ship landing with a payload mass between 4000 Kg and 6000 Kg

total_success	total_failure
100	1

Total Number of Successful and Failure Mission Outcomes

- Aside from landing success, SpaceX has a nearly 100% mission success rate with only one occurrence of a failure to date

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

Boosters Carried Maximum Payload

- Compared to the 4000-6000Kg payload boosters, these boosters seem to be newer versions; meaning that SpaceX improved their equiptment overtime in order to do more rigorous testing.

2015 Launch Records

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- In 2015, there were two occurences of drone ship failures. One in January and one in April

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

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

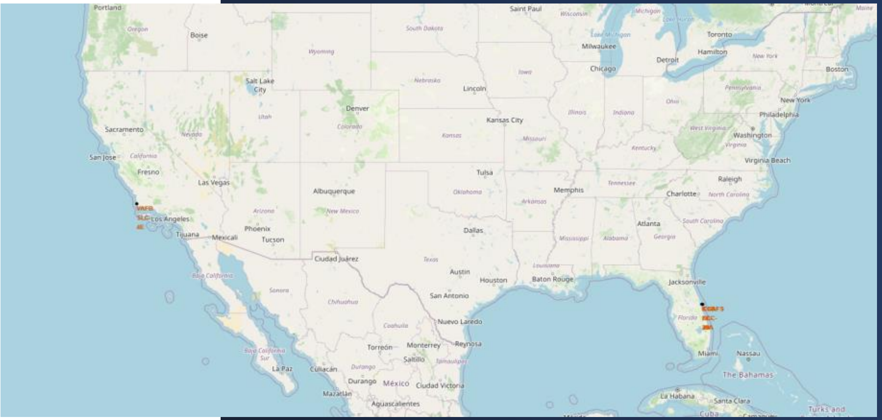
- Roughly a third of landings were not even attempted, more than likely teaching the SpaceX team in return
- Drone ship has the most success with 5
- Ground pad and ocean both yield 3 success each

Section 3

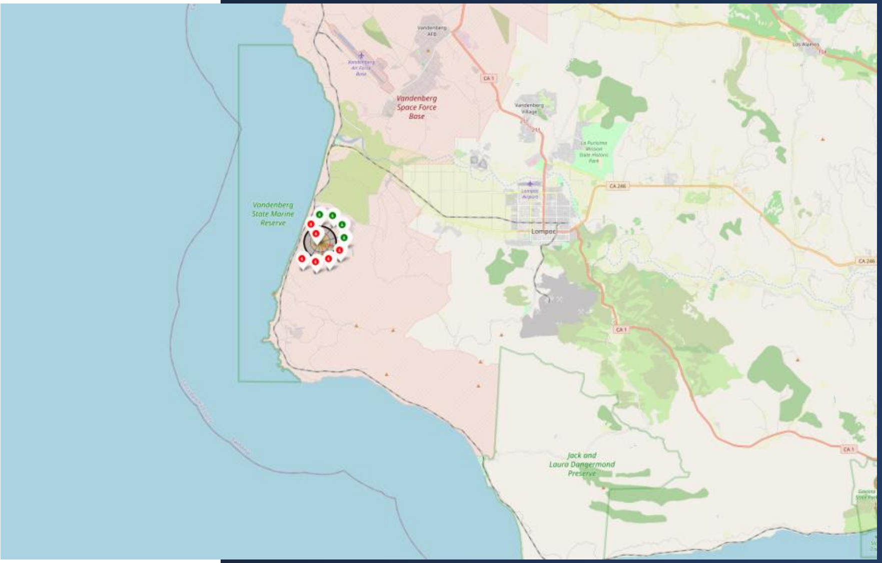
Launch Sites Proximities Analysis

Launch locations on Map

- This shows the locations of SpaceX launch sites on a map of The United States.
- From here we can draw insight into their location regarding launch patterens, and viscinity to other objects/locations



Success and Failure on Map



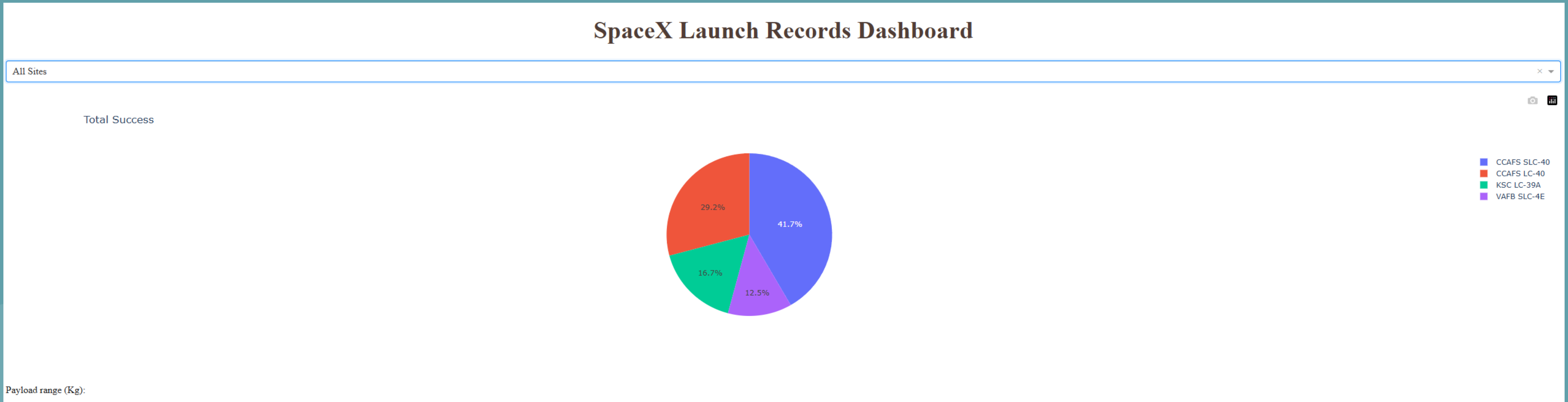
- The first picture shows a cluster of successes and failures at the launch location on the western coast.
- The second picture details one of the launch locations on the eastern coast

Section 4

Build a Dashboard with Plotly Dash

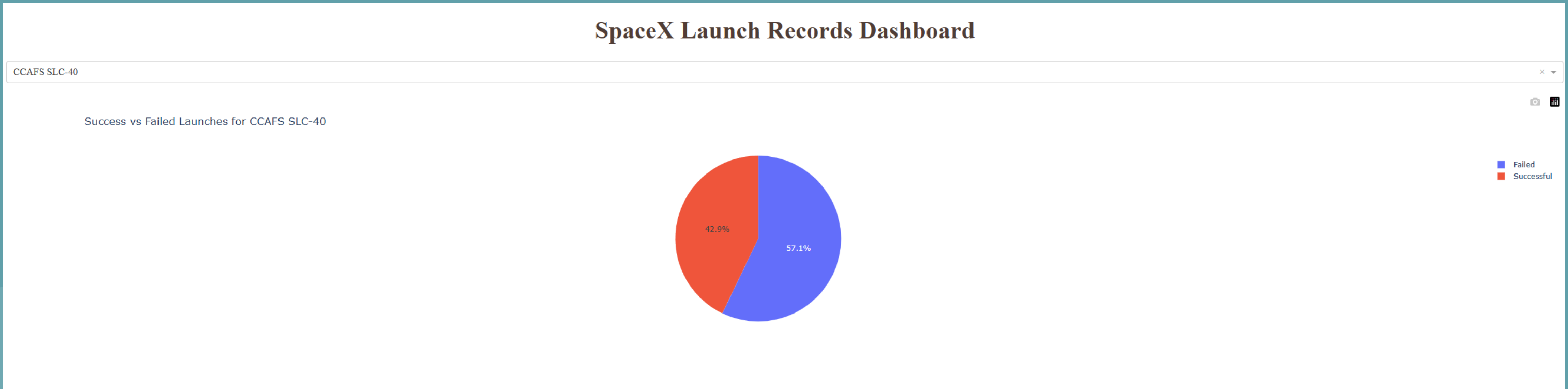
Launch Success for All Sites

- On the dashboard application, we can compare the different ratios of success per different launch site.
- Here we see with all sites displayed, that CCAFS SLC-40 has the highest success among them



Highest Success Rate Launch Site

- Here is a closer look at the CCAFS SLC-40 launch site success ratio.
- We can see that nearly half of the launches were a success



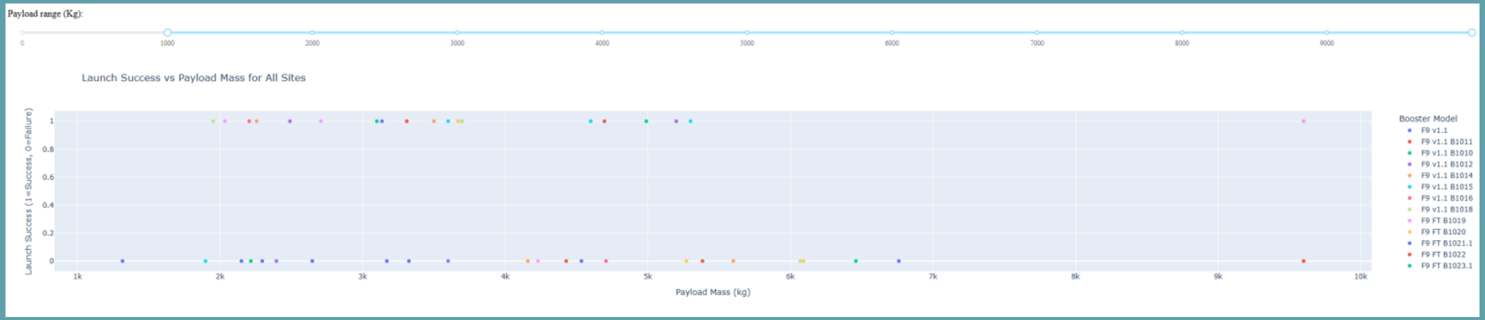
Payload vs. Launch Outcome (All Sites)

- On this portion of the dashboard, we can change the range of payload mass to display their individual success, colored by booster model.

- The first is at 4000 Kg, with minimal success

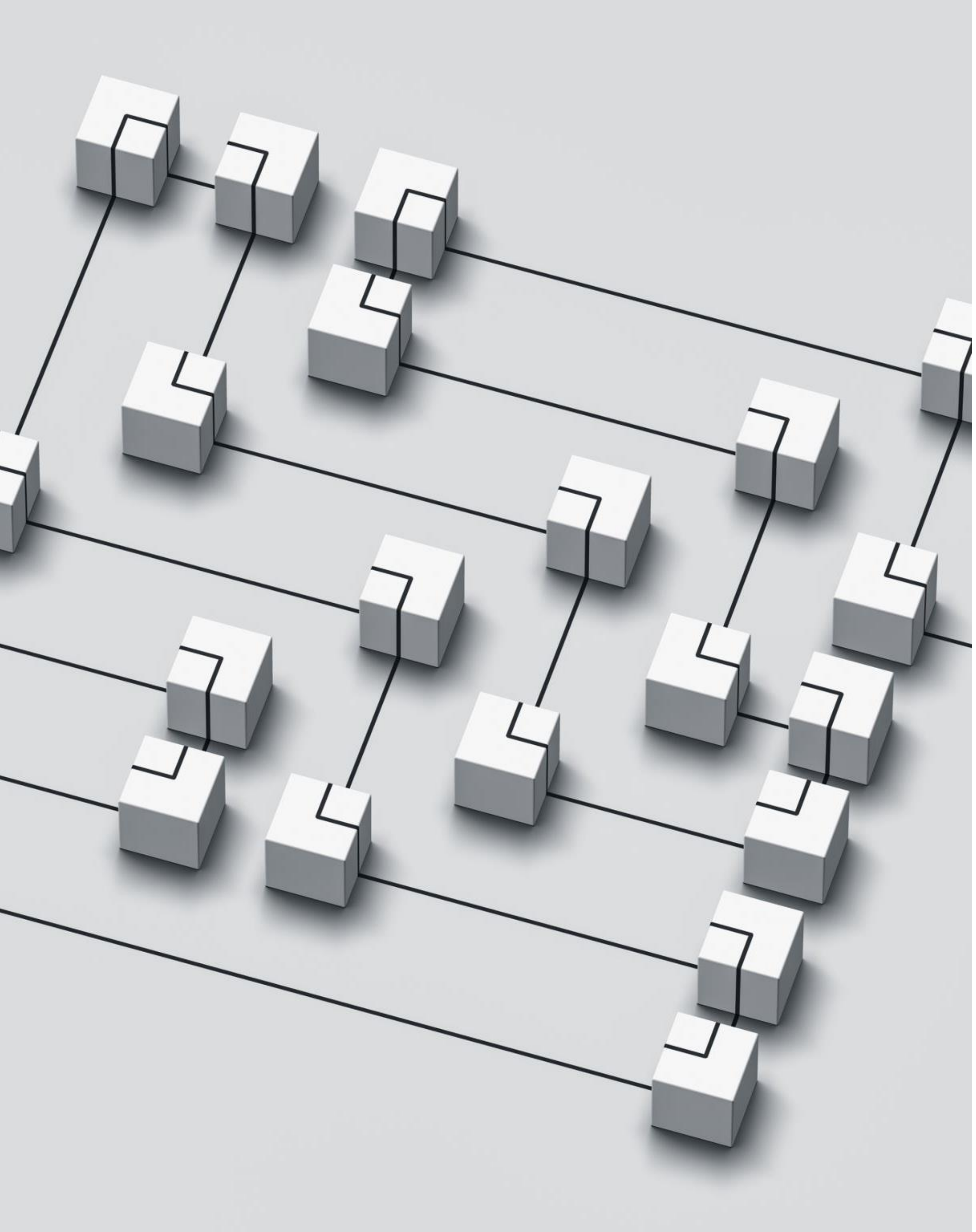
- The second is at a more modest 1000 Kg, noting that there are many more successes in this range

- We can also view which booster model produced these successes e.g. we might see that some models perform better with higher payloads, and some with lower payloads



Section 5

Predictive Analysis (Classification)

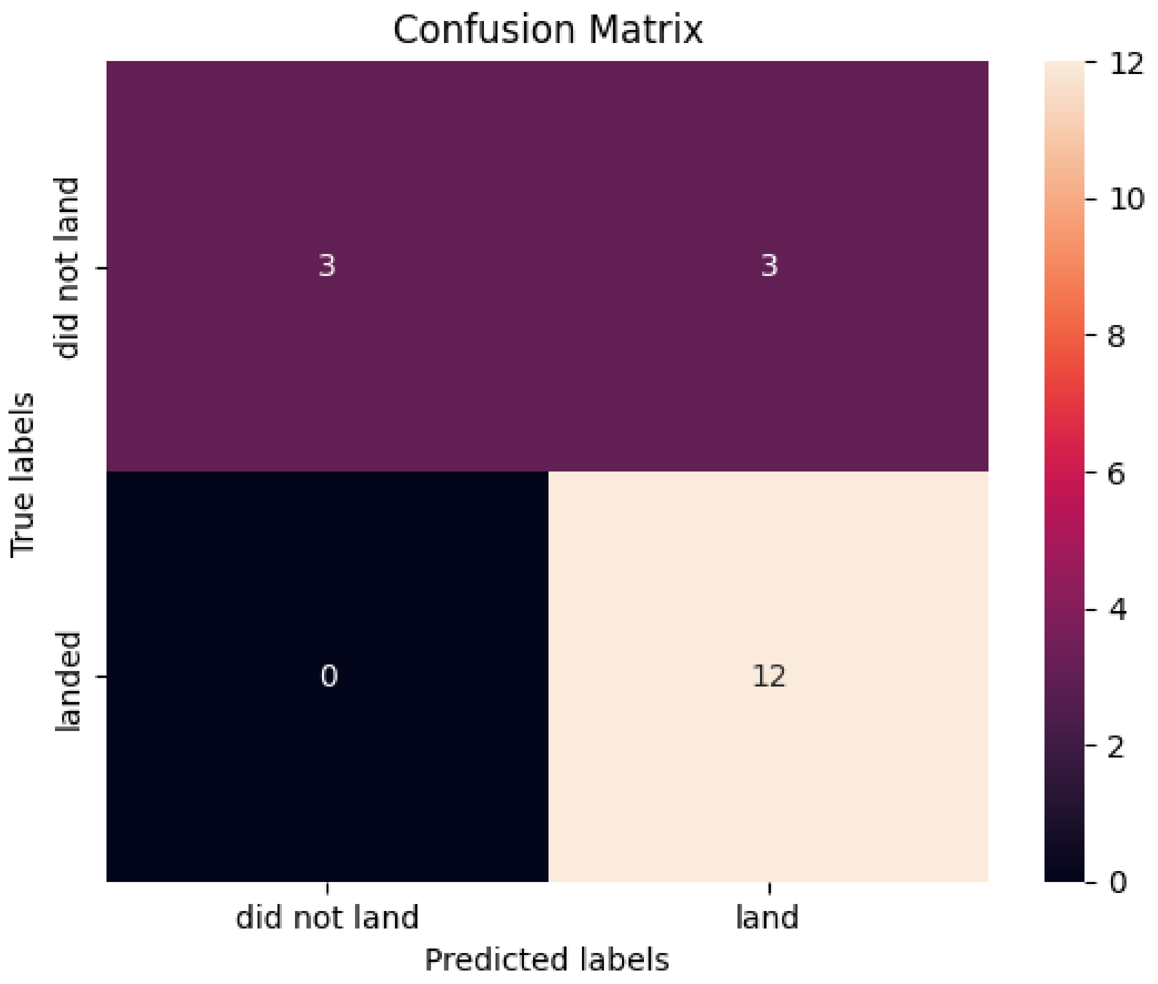


Classification Accuracy

- The models all performed relatively equally with only one slightly outperforming the rest
- In the end the decision tree had slightly better accuracy and therefore is the best model to make predictions on this data set

Confusion Matrix

The best preforming model was the decision tree. The confusion matrix shows us it 3 true negatives and 12 true positives, with only 3 false negatives.



Conclusions

- The decision tree is the best model to make predictions on this data set
- SpaceX has continued to increase success rate and therefore there is more to learn in the future

Appendix

https://github.com/ChrisACr/CapstoneRepo.git

Thank you!

https://github.com/ChrisACr/CapstoneRepo.git