Winning Space Race with Data Science

Adrian Simms
2023-07-21

Outline

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

Section 1: Executive Summary

Executive Summary

- Summary of methodologies
  - We sourced the Spacex launch data using an API, populated a Dataframe, filtered on Falcon 9, and cleaned up missing data.We also web scraped Falcon 9 data from the Falcon 9 launch history wikipedia page.
  - We performed exploratory data analysis utilizing SQL
  - We performed exploratory data analysis using visualizations and charts
  - We Utilized Folium to plot the Launch Sites on a map and find their proximity to points of interest such as railways, coastlines and cities.
  - We created an interactive Plotly Dash to evaluate and identify Most successful Launch sites and any relationship of payload mass and or booster on success.
  - In an attempt to determine if we could predict successful launch we applied various machine learning models and utilized hyperparametization to optimize the models and decide on the best model to use in our prediction.

- Summary of all results
  - After filtering on Falcon9 launches we were left with 90 rows of data. There were 5 missing Payload mass values that were replaced with the Mean of PayLoad Mass. Landing Pad field had 26 missing values remaining. After web scraping we were left with a 121 row data frame as well.
  - Thee are 4 distinct Spacex Falcon 9 Launch Sites. NASA is a customer of SpaceX Falcon 9. Space x has launched 45,596 kg of payload mass for NASA. The average payload mass for Falcon 9 booster version 1.1 is 2,928 kg. The first successful ground pad landing was Dec 12th 2015. Space x has had 100 successful missions. 12 different Falcon 9 booster B5 variations have carried the maximum payload of 15,600 kg. In 2015 Launch site CCAAF LC-40 had 2 failed drone ship launches. Drone ship and ground ship produced the most successful launch outcomes at 5 launches each.
  - Launch Number and Payload mass both are key features in affecting Launch Success. Launch Site VAFB has no Launches for heavy payloads over 10,000 kg. Orbits SSO, HEO, GEO and ES-L1 have a 100% success rate. Success rate trend increased from 2013 to 2020.
  - The Launch sites were all near Ocean coasts with 3 in Florida and 1 in California. They were also near Highways and railways and further away from major cities.
  - Launch Site KSC Lc-39 had both the most successful launches and the highest successful launch rate at 10 and 76.9% respectively. The most successful booster versions was FT and Payload range with the most success was between 3,000 and 4,000 kg.
  - The Decision Tree Classification model performed the best with an accuracy score of 0.833 on our test data and 0.875 on out training data.

Section 2: Introduction

Introduction

- Project background and context
  - SpaceX is a private rocket company that has been able to distinguish itself in the competitive market by charging $65 Million per Rocket launch whereas its competitors charge upwards of $165 Million per Launch. SpaceX is able to keep its cost low by being the first private company to establish a process by which it can re-use the first stage of its Falcon 9 rockets. As a competitor analyzing the market and preparing for bidding we would like to utilize Machine learning to predict the outcome of a launch. This predictive capably could assist a competitor in figuring out the true cost of a launch.
- Problems you want to find answers
  - Can we identify a source for Falcon 9 launch Data?
  - What is the success rate for Falcon 9 Launches?
  - With what accuracy can we predict the success of Falcon 9 launches?

Section 3: Methodology

Methodology

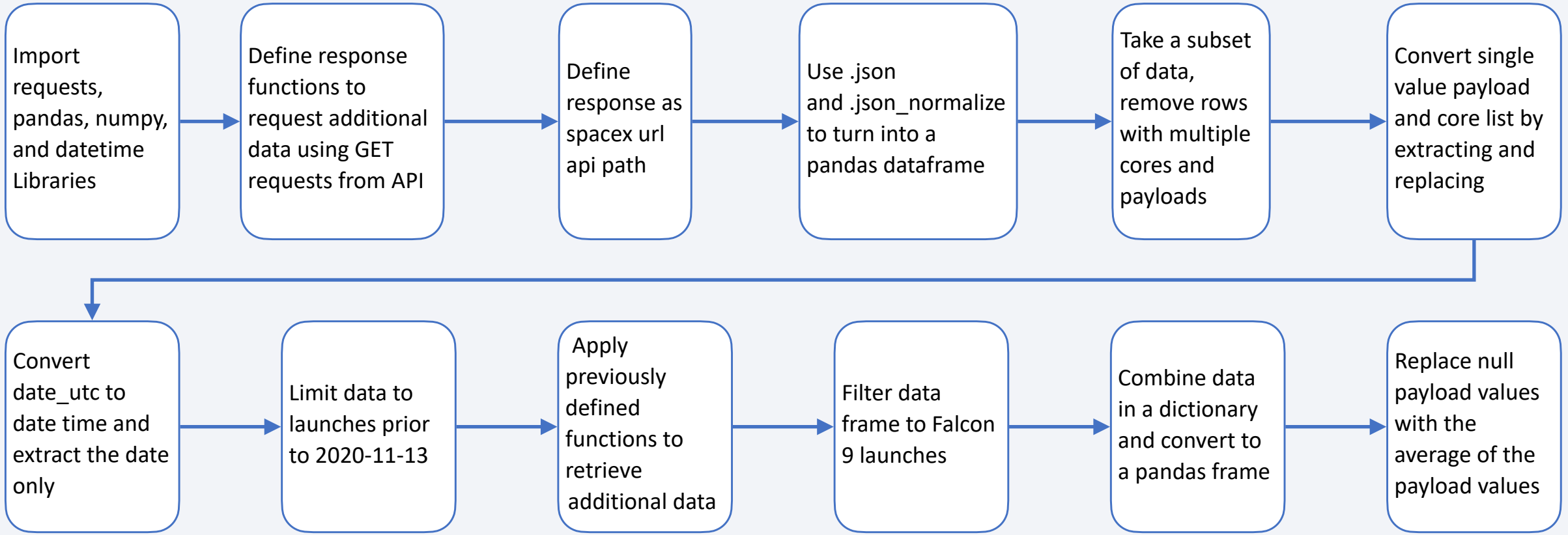
Executive Summary

- Data collection methodology:
  - The SpaceX launch data was initially collected from the SpaceX launch data API.
  - Then we used a get request to request and parse the data.
  - We then filtered the data frame to only include Falcon 9 launches.
  - We updated missing payload mass values with the average Payload Mass value.
  - We used web scraping from the Falcon 9 Launch history Wikipedia page to also pull data.
  - We parsed and placed the web-scraped data in a Dataframe.
- Perform data wrangling:
  - Calculate the Number of Launches at each site
  - Determine the number of Landing Outcomes
  - Create a landing outcome label
- 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
  - Tune
  - Evaluate classification models

Data Collection

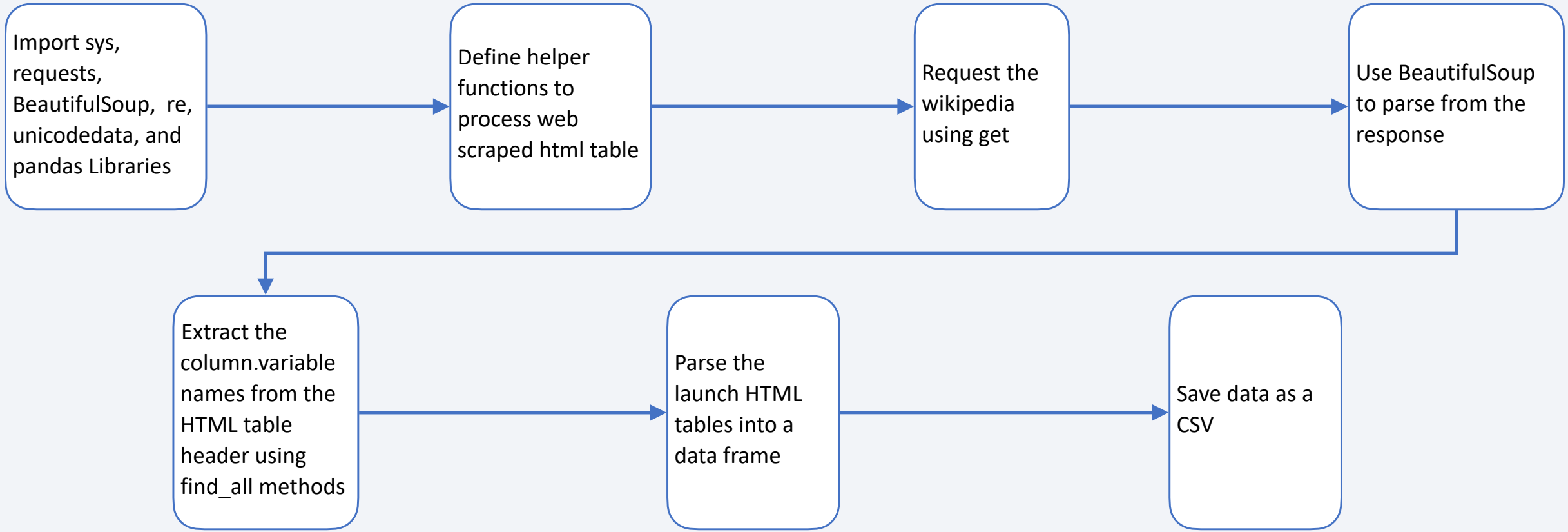
- Datasets were collected by utilizing two methods (Collection FlowChart on next two slide):
    1. API Request (Dataset Part 1)
        - Initial Data Pull: https://api.spacexdata.com/v4/launches/past
        - Additional Data:
            - Rocket: https://api.spacexdata.com/v4/rockets/
            - Launchpad: https://api.spacexdata.com/v4/launchpads/
            - Payloads: https://api.spacexdata.com/v4/payloads/
            - Cores: https://api.spacexdata.com/v4/cores/
    2. Web scraping (Dataset Part 2)
        - List of Falcon 9 and Falcon Heavy launches Wikipedia Web Page:
            - 2020 Launch Data: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
            - 2010 to 2019 Launch Data: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches_(2010%E2%80%932019)

Data Collection – SpaceX API



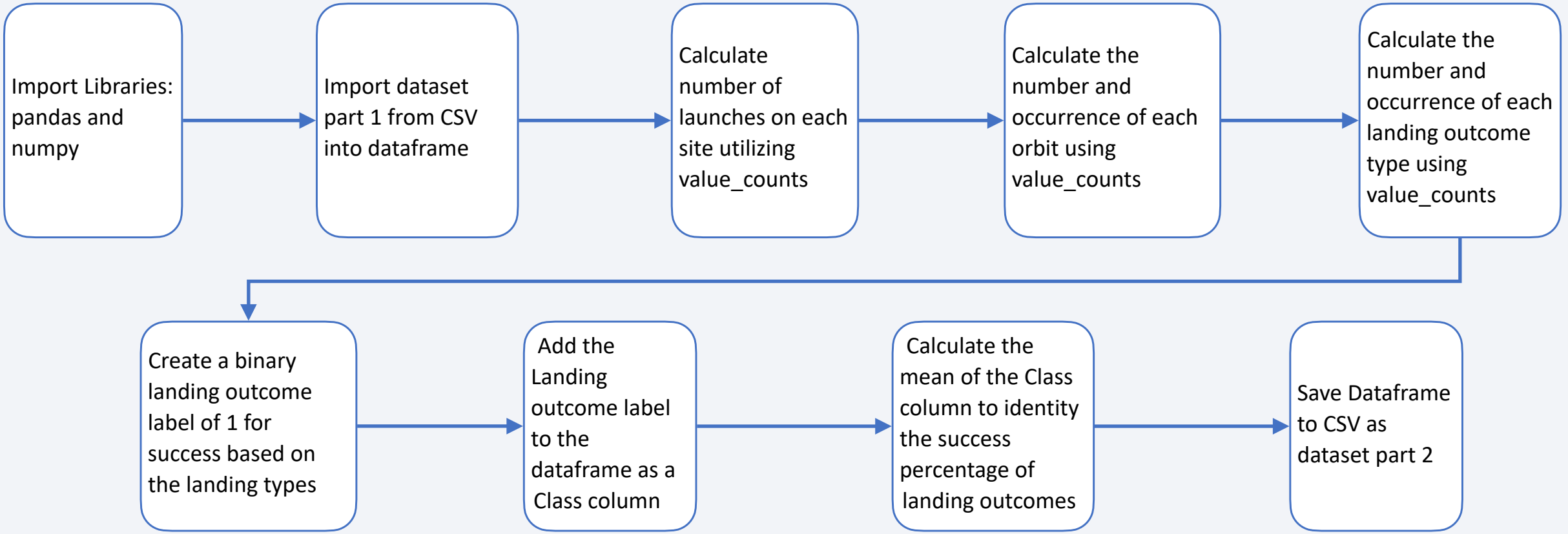
- The GitHub URL of the completed SpaceX API calls notebook:
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



- The GitHub URL of the completed web scraping notebook:
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



- The GitHub URL of the completed data wrangling related notebook:
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.ipynb

EDA with Data Visualization

- Charts plotted and purpose:
  - Plotted a Scatter Chart to see how the FlightNumber and Payload Mass affect the Launch Outcome
  - Plotted a Scatter Chart to see how the FlightNumber and Launch Site affect the Launch Outcome
  - Plotted a Scatter Chart to visualize the relationship between Launch Site and Payload Mass and the affect on Launch Outcomes
  - Plotted a Horizontal Bar Chart to compare the Success Rates for different Orbit Types
  - Plotted a Scatter chart to see if there is any relationship between Orbit Type and Flight Number and the affect on Launch Outcomes
  - Plotted a Scatter Chart to visualize the relationship between Orbit Type and Payload Mass and their affect on Launch Outcomes
  - Plotted a Line Chart visualize the trend of the Success Rate over time (by Year)
- GitHub URL of completed EDA with data visualization notebook
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- SQL queries performed:
  - Displayed the names of the unique launch sites in the space mission
  - Displayed 5 records where the launch site began with ‘CCA’
  - Displayed the total payload mass carried by boosters launched by NASA (CRS)
  - Displayed the average payload mass carried by booster version F9 v1.1
  - Listed the dated when the first successful landing outcome in ground pad was achieved
  - Listed the names of the boosters which have success in drone ship and had payload mass greater than 4000 and less than 6000 kg
  - Listed the total number of successful and failure mission outcomes
  - Listed the names of the booster versions which carried the maximum payload mass
  - Listed the records occurring in 2015 with Failure landing outcomes in drone ship. I displayed the Month, Year, Landing Outcome, Booster Version and Launch Site
  - Ranked the count of landing outcomes by their count of launches between the dates of 2010-06-04 and 2017-03-20 in descending order
- GitHub URL of completed EDA with SQL notebook
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Launch Outcome success may also be dependent on geographic location of the launch site as well as launch site proximities. In order to observe these geographic considerations we utilize folium to create an interactive map.
    - First we marked all launch sites on the map utilizing their Latitude and longitude data points with Circle markers
    - We then utilized MarkerClusters to mark successful/failed launches for each launch site
    - Lastly we calculated the distance between the launch sites and the following proximities and added them to the map using PolyLine:
        - Railways
        - Oceanic coasts
        - Highways
        - Major cities
- The GitHub URL of the completed interactive map with Folium map:
    - https://github.com/BlessedMath/DataScienceCapStone/blob/main/lab_jupyter_launch_site_location.ipynb

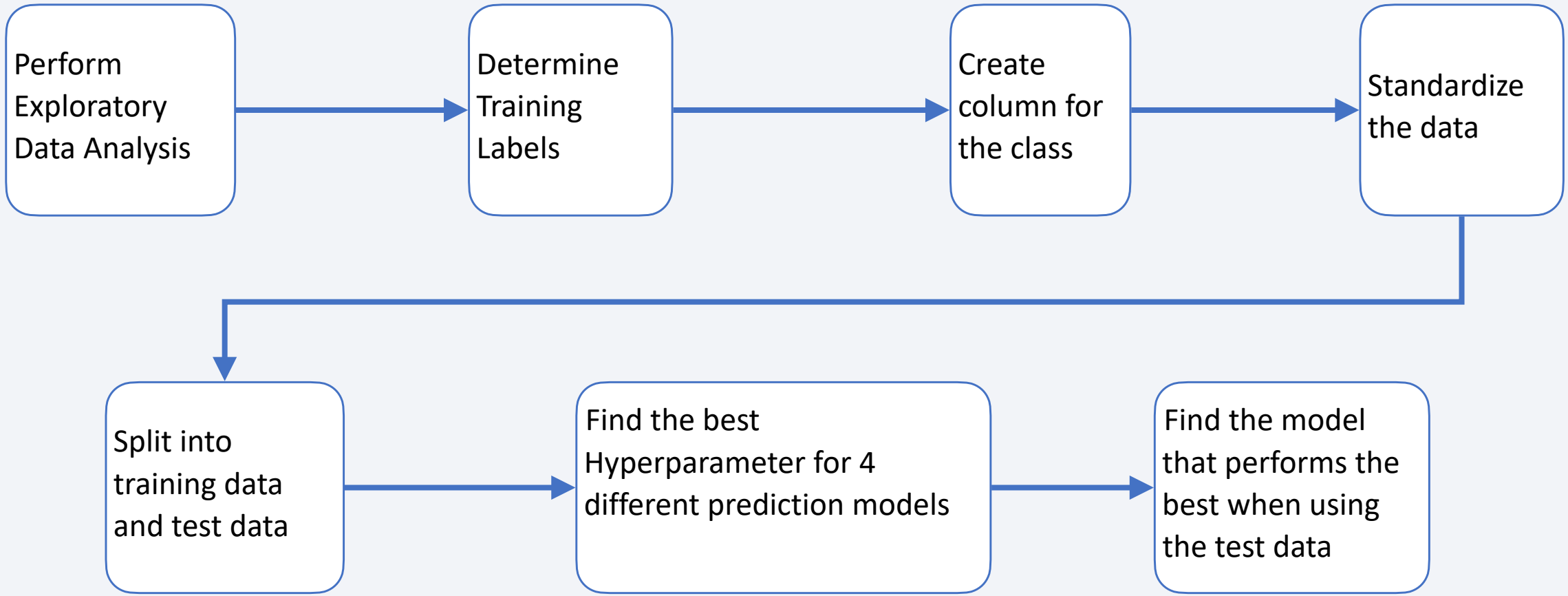
Build a Dashboard with Plotly Dash

- In order to perform interactive analytics we build a dashboard with Plotly Dash. The following elements will be added:
  - Interactions:
    - A Drop down to select Launch Site
    - A Range Slider to select Payload Ranges
  - Charts:
    - A Pie chart to display counts of success and success rate
    - A Scatter chart comparing Success with Payload Mass.
- Callback functions will be used to render and update the charts based on the Interaction inputs.
- The GitHub URL of the completed Plotly Dash lab:
  - https://github.com/BlessedMath/DataScienceCapStone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- After bringing in the SpaceX Launch Data, we identified our target variable (Class). Once assigning it to the value Y, we standardized the independent features hereafter called X. Using the train test split method we split our data into train and test data with 20% of the data assigned to test for both X and Y. Starting with Logististic Regression we ran Grid Search across 4 different Machine Learning Models to identify the best parameters to use for each model.We then tested the accuracy on both the training and test data sets to identify the model which performed best. With the help of both the accuracy score on the test data and the confusion matrix of each model we were able to choose the best performing model with the best parameters.
- Simplified Model Development Process flow chart on next slide.
- The GitHub URL of the completed predictive analysis lab
    - https://github.com/BlessedMath/DataScienceCapStone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.ipynb

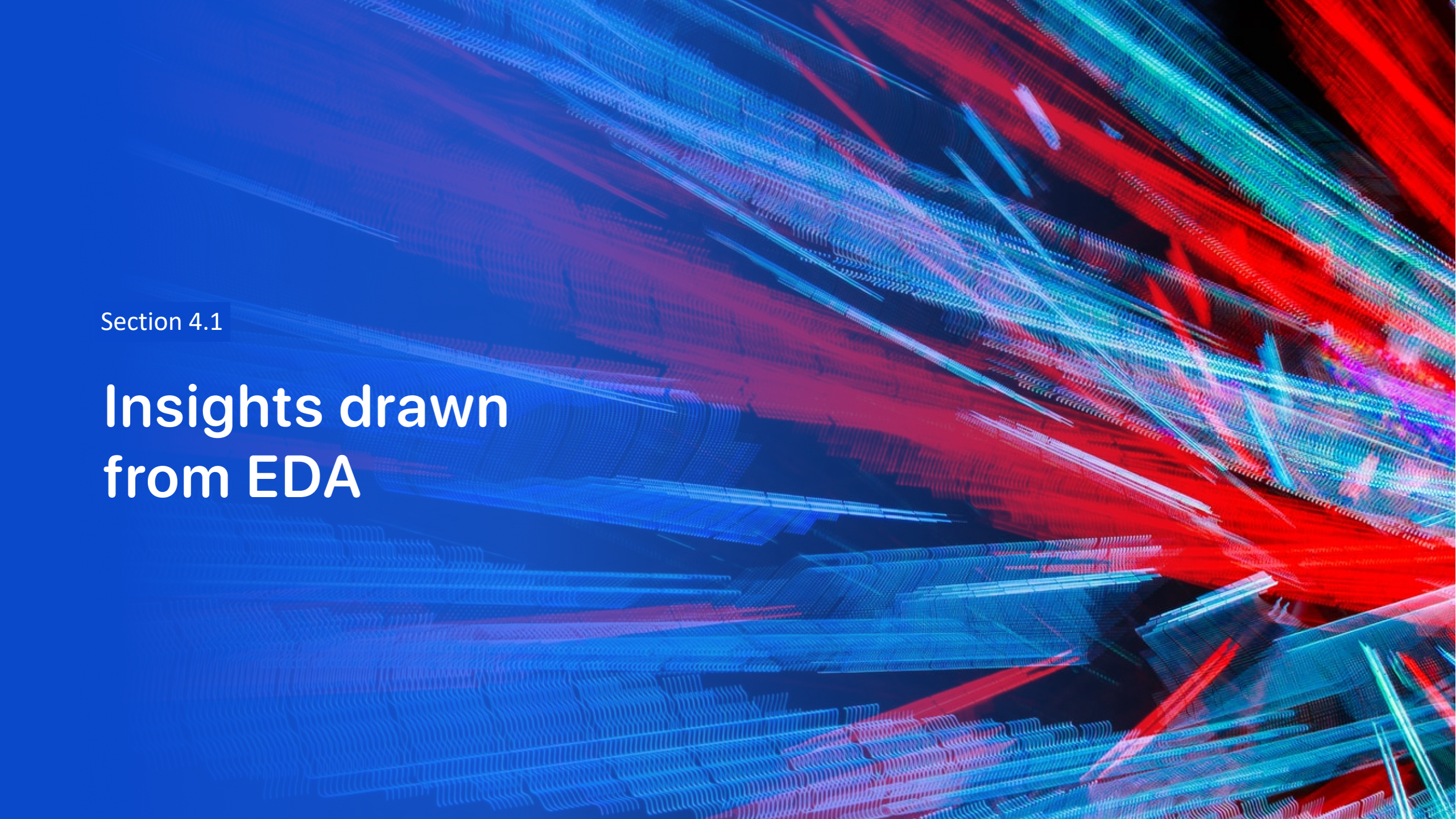
Predictive Analysis (Model Development Process: Flow Chart)



Section 4: Results

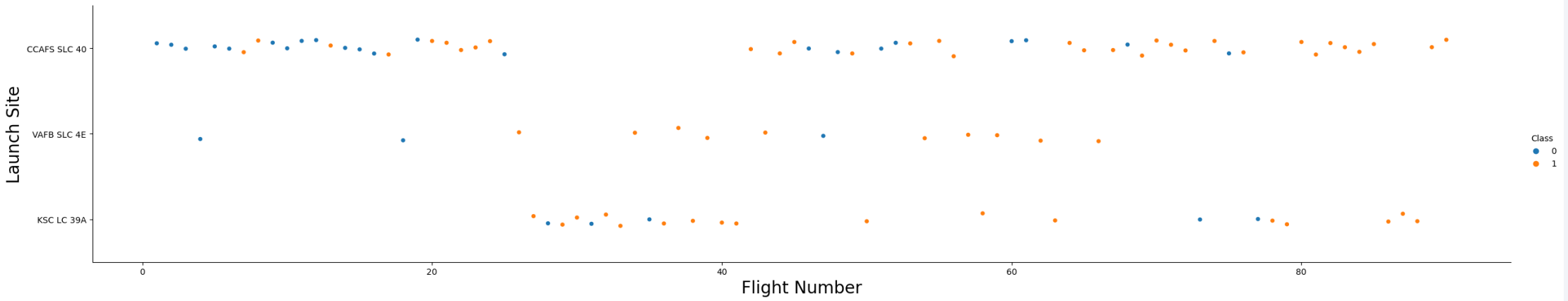
Results

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


Section 4.1
Insights drawn from EDA

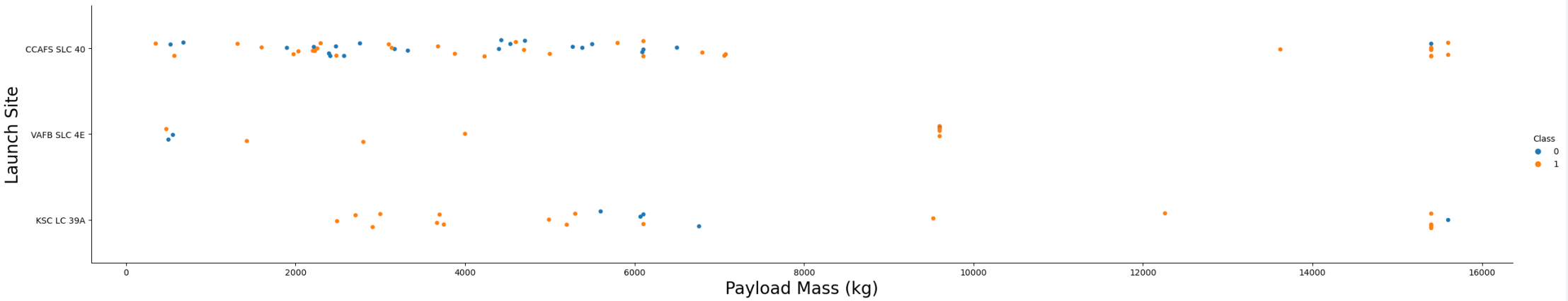
Flight Number vs. Launch Site

- The scatter plot has Flight Number on the x axis and Launch Site on the y axis. The orange dots represent successful launches and the blue unsuccessful launches.
- Success is correlated with flight number. After 20 flights the successful launches began to increase at all Launch sites.
- Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%



Payload vs. Launch Site

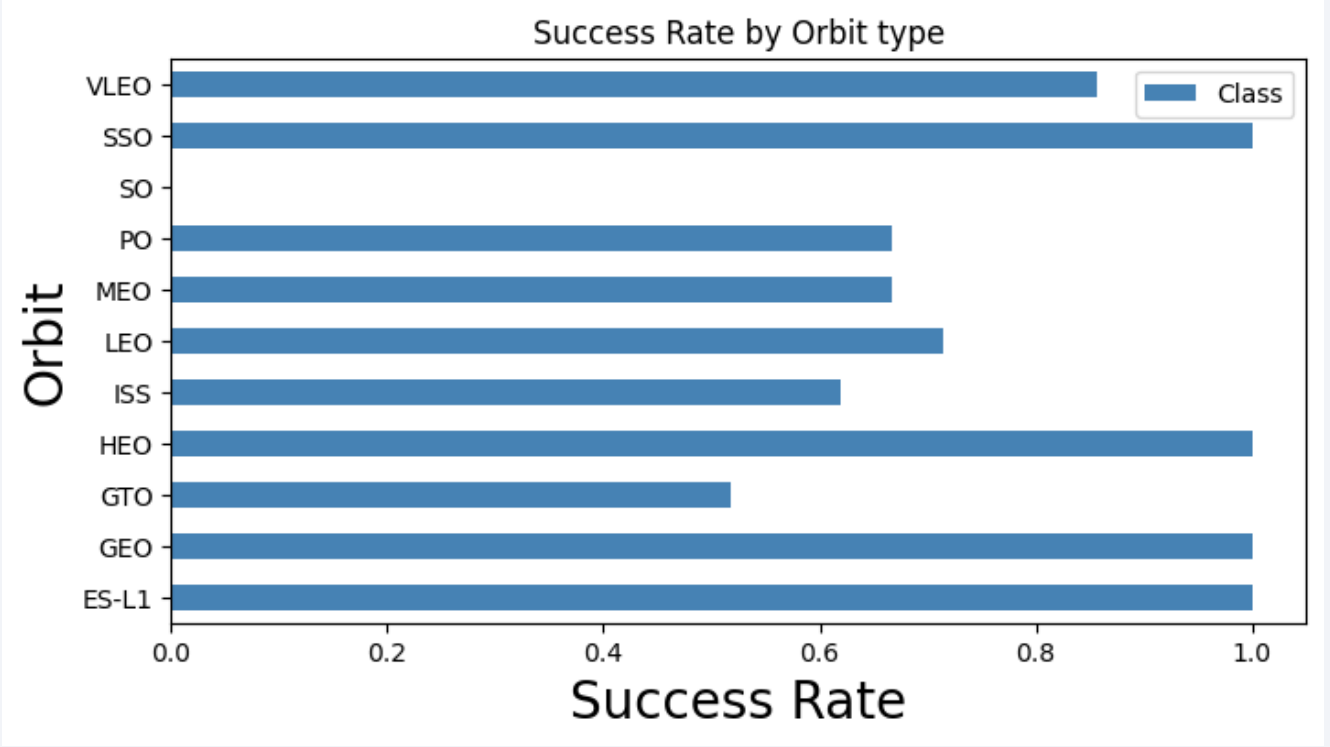
- The scatter plot shows Payload Mass on the x axis and Launch Site on the y axis. The orange dots represent successful launches and the blue unsuccessful launches.
- For Launch site KSC LC-39A success is correlated with Payload Mass. The Launches with mass under 6000 kg and over 7000 kg are successful.
- For launch site VAFB SLC-4E the Payload Mass over 10000 kg are successful
- For the CCAFS SLC-40 launch site the success seems to spike after Payload Mass of 6500 kg.



Success Rate vs. Orbit Type

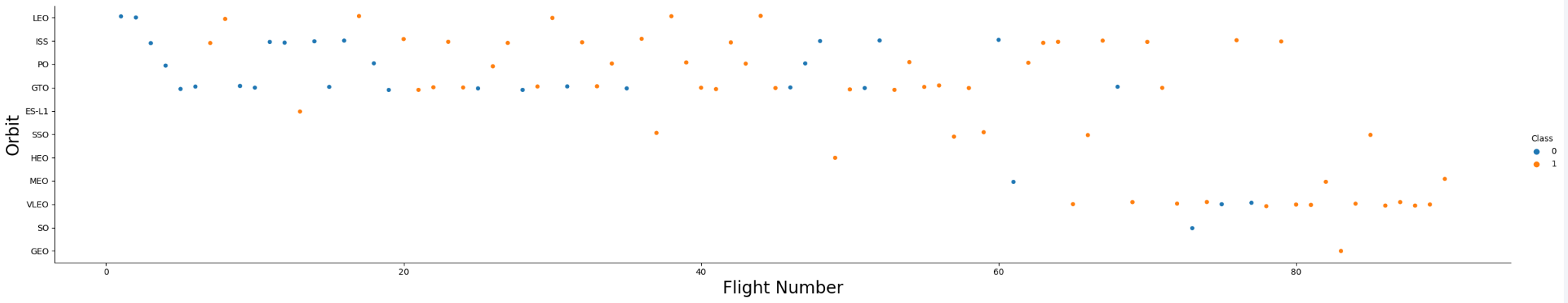
- The bar chart displays the success rate of each orbit type

- The following Orbits have a 100% success rate:
  - SSO (Sun Synchronous Orbit)
  - HEO (Highly Elliptical Orbit)
  - GEO (Geostationary Orbit)
  - ES-L1 (European Space Agency Lagrange Point 1)



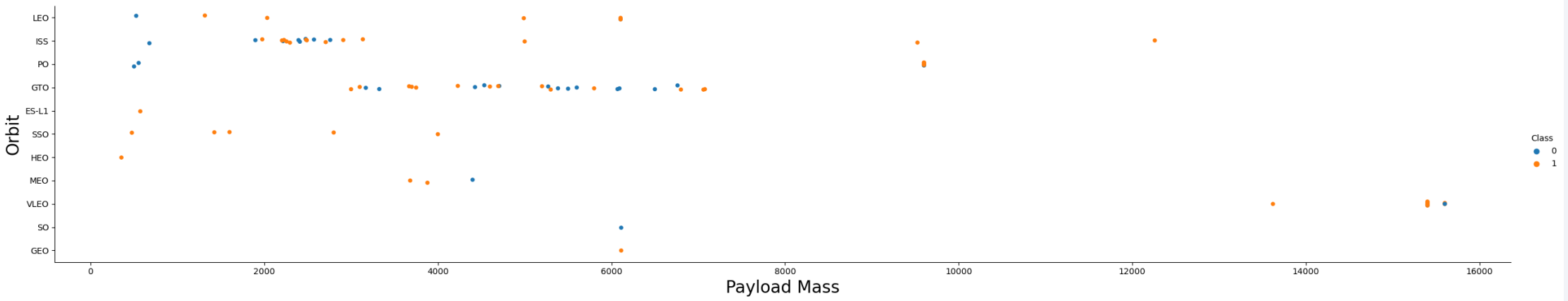
Flight Number vs. Orbit Type

- The scatter plot has Flight Number on the x axis and Orbit on the y axis. The orange dots represent successful launches and the blue unsuccessful launches.
- The Orbits is not as correlated with flight number. LEO has been successful since flight number 5. VLEO flight are more prevalent in modern flights.



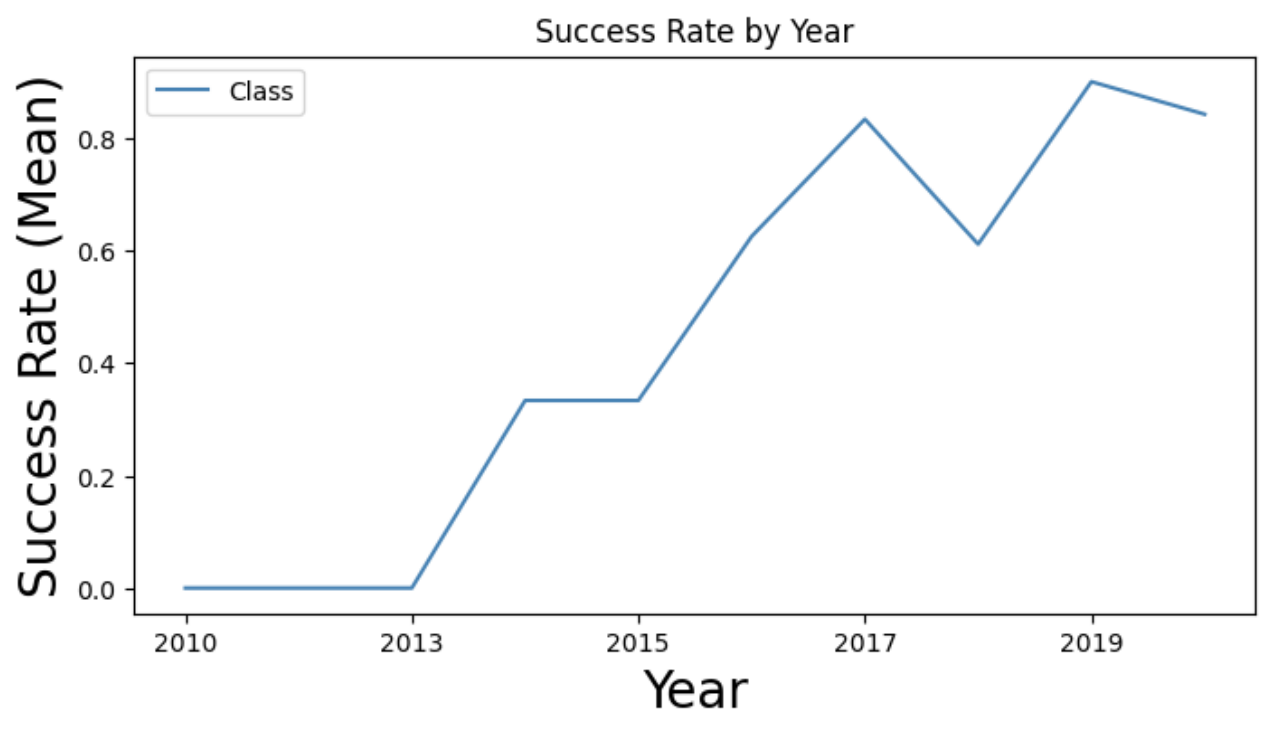
Payload vs. Orbit Type

- The scatter plot has Payload Mass on the x axis and Orbit on the y axis. The orange dots represent successful launches and the blue unsuccessful launches.
- There is not an observed correlation of Payload Mass with Orbit.



Launch Success Yearly Trend

- The Line chart shows the Success Rate of Launches over time.
- Launch rate success began to rapidly improve starting in 2013 and appears to have normalized around 80% since 2017.



All Launch Site Names

- The table displays the unique names of Launch Sites
- Getting the list of launch sites is important as we will later try to identify if launch success is dependent upon the location of launch.
- There are 4 distinct sites and a large portion of records with None listed as the site.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

- The table shows 5 records where launch sites begin with `CCA`
- This query allows us to get an understanding of the important fields and shows that NASA is a common customer from the CCAFS LC-40 Launch Site.

Total Payload Mass

- This table presents the total payload carried by boosters across all launches for NASA as a customer.
- 45,596 kg worth of payload has been launched for NASA.

Customer	Total_Payload_Mass
NASA (CRS)	45596.0

Average Payload Mass by F9 v1.1

- The table displays the Average Payload Mass carried by booster version F9 v1.1
- The average Mass for booster version v1.1 is 2,928 kgs.

Booster_Version	AVG_PAYLOAD_MASS
F9 v1.1	2928.4

First Successful Ground Landing Date

- The table displays the first successful ground pad landing
- The first successful ground pad landing was December 12th 2015.

Min_Date	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- The table shows the names of all boosters which successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.
- All of successful landings were booster version FT.

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696.0	Success (drone ship)
F9 FT B1026	4600.0	Success (drone ship)
F9 FT B1021.2	5300.0	Success (drone ship)
F9 FT B1031.2	5200.0	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Summarized_Mission_Outcome	Mission_Count
Failure	899
Success	100

- The table shows the total number of successful and failure mission outcomes
- Approximately 10% of the missions were a success.
- SpaceX is responsible for 100 successful missions
- Out of the 899 Failure Missions 898 records were listed with a Launch site of None and the other 1 failure was from the CCAFS LC-40 site.

Boosters Carried Maximum Payload

- The table lists the names of the boosters which carried the maximum payload mass of 15,600 kg.
- The B5 booster versions are responsible for all 12 maximum capacity Pay Load launches.

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600.0
F9 B5 B1049.4	15600.0
F9 B5 B1051.3	15600.0
F9 B5 B1056.4	15600.0
F9 B5 B1048.5	15600.0
F9 B5 B1051.4	15600.0
F9 B5 B1049.5	15600.0
F9 B5 B1060.2	15600.0
F9 B5 B1058.3	15600.0
F9 B5 B1051.6	15600.0
F9 B5 B1060.3	15600.0
F9 B5 B1049.7	15600.0

2015 Launch Records

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

- The table shows the 2 failed landing outcomes in drone ship, their booster versions, and launch site names in the year 2015.

- Both of the failed drone ship landings in 2015 were out of the CCAFS LC-40 Florida site with the v1.1 booster version.

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

- The table shows Landing Outcomes ranked by count between the dates 2010-06-04 and 2017-03-20.

- The number one with 10 outcomes are No Attempt.
- There is a 3-way tied for second rank between drone ship Failure, drone ship Success and ground pad Success landings with 5 counts each.

Landing_Outcome	Landing_Outcome_Count	Min_Date	Max_Date	Rank
No attempt	10	2012-05-22	2017-03-16	1
Failure (drone ship)	5	2015-04-14	2016-06-15	2
Success (drone ship)	5	2016-05-27	2017-01-14	2
Success (ground pad)	5	2015-12-22	2017-03-06	2
Controlled (ocean)	3	2014-04-18	2015-11-02	5
Uncontrolled (ocean)	2	2013-09-29	2014-09-21	6
Failure (parachute)	1	2010-08-12	2010-08-12	7
Precluded (drone ship)	1	2015-06-28	2015-06-28	7

Section 4.2

Launch Sites Proximities Analysis

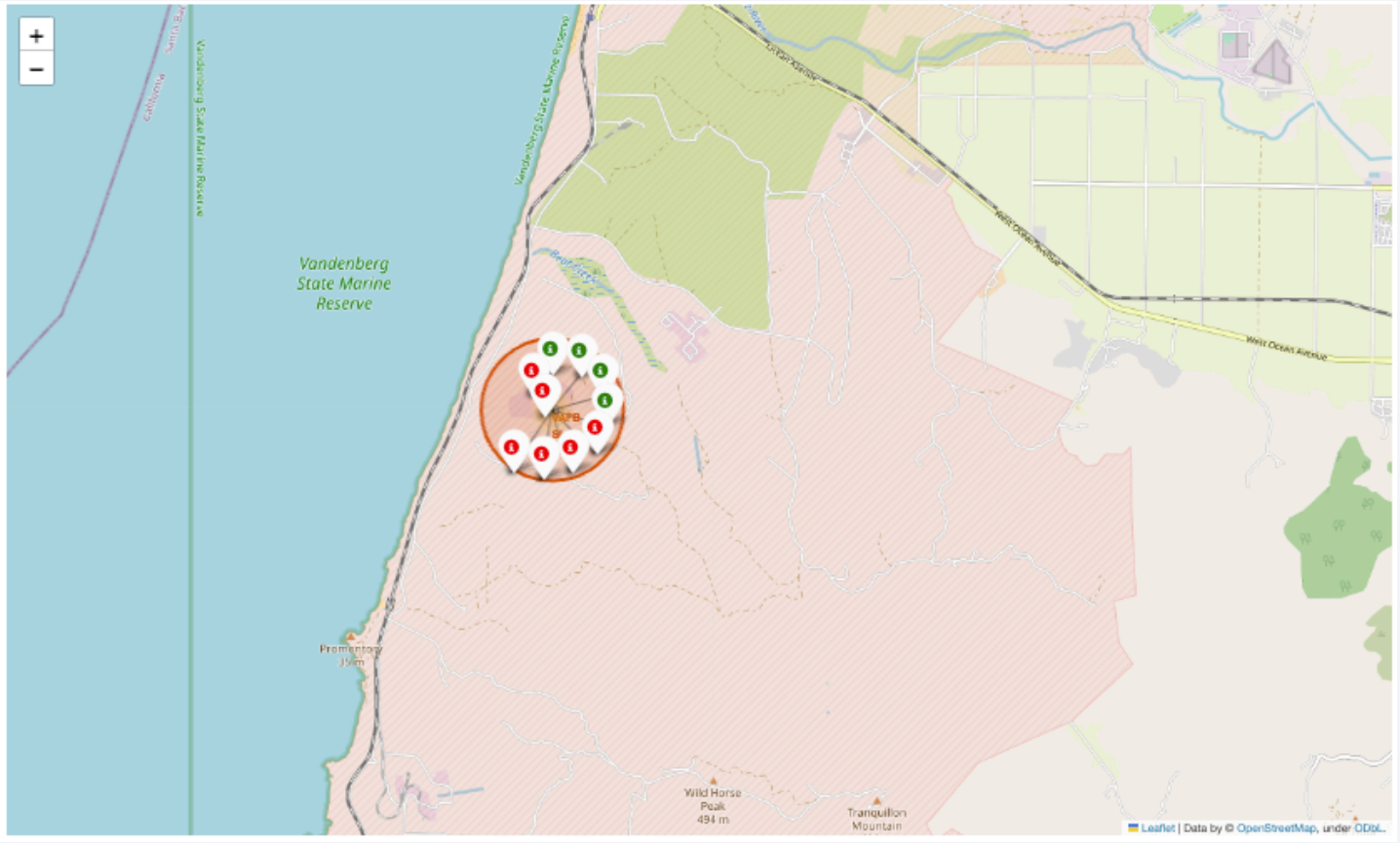
Launch Site Locations

- The map provided displays the locations of the Launch Sites
- Three of the sites are located on the coast of Florida, while the VAFB SLC-4E site is on the opposite coast of California

Launch Site with color Labeled Outcomes

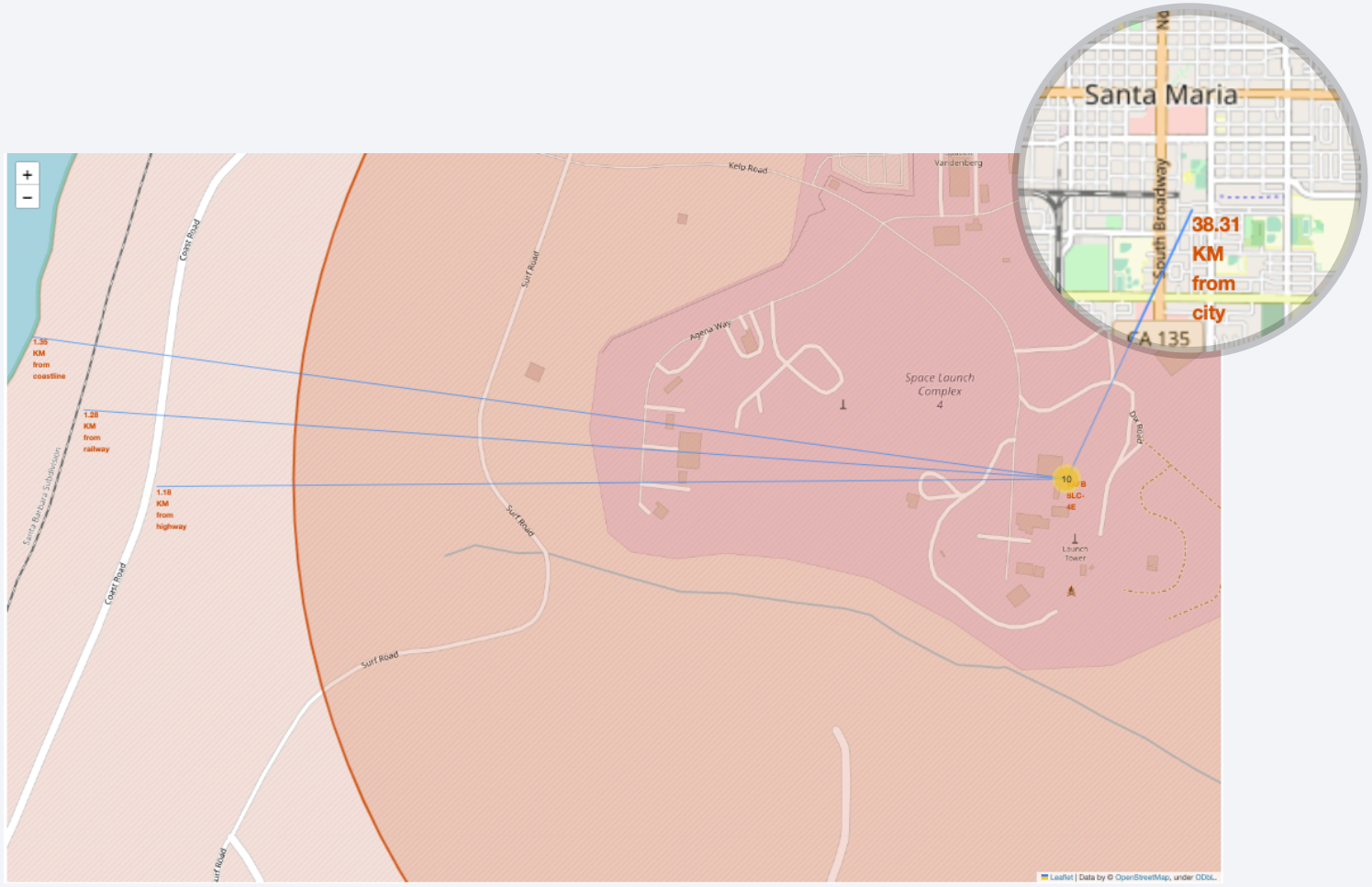
- Zooming into the California VAFB SLC-4E site we can see the outcomes of the Launches represented as a cluster of icons.

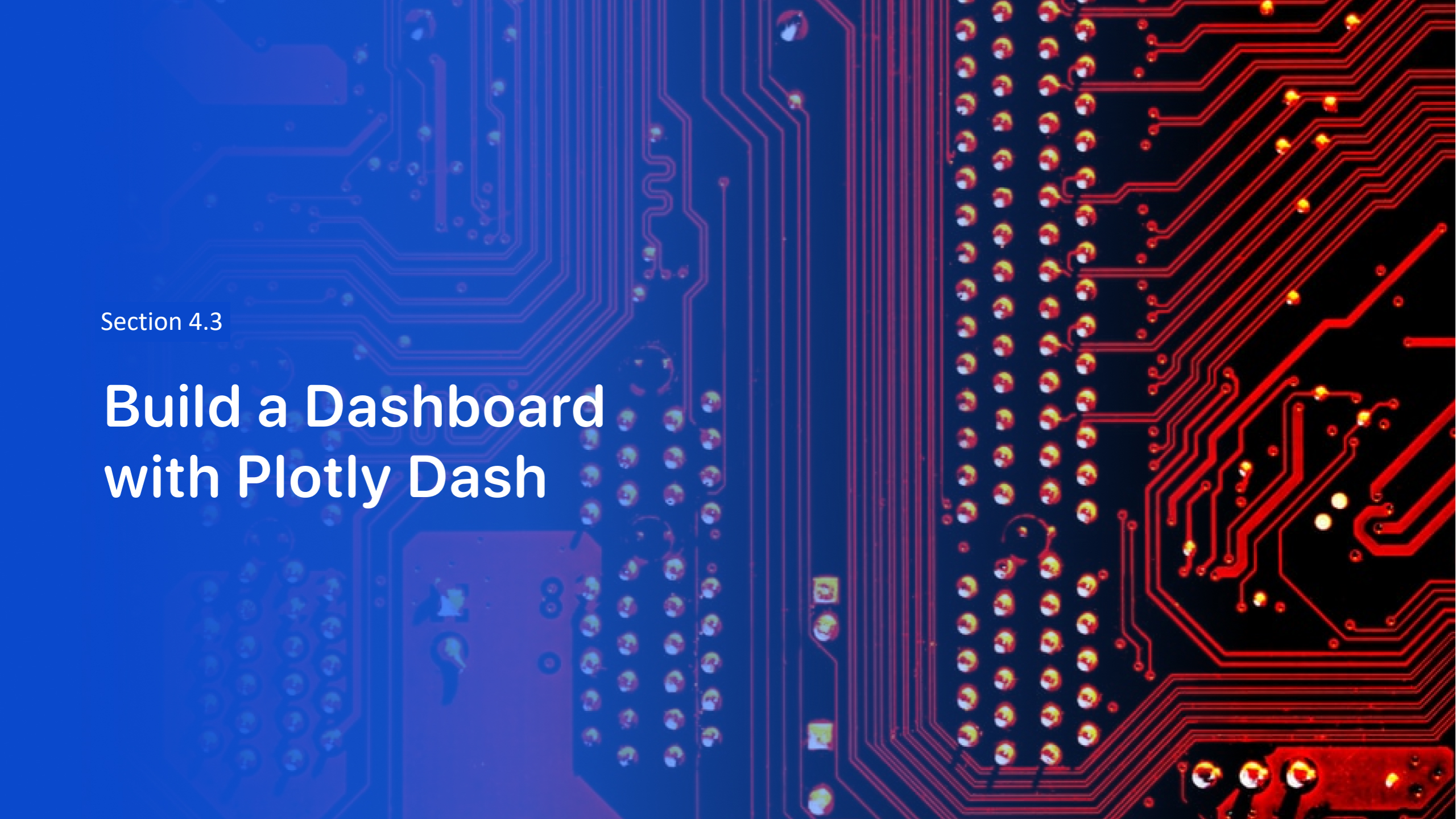
- Green represents successful Launches and Red unsuccessful



Launch site with proximities

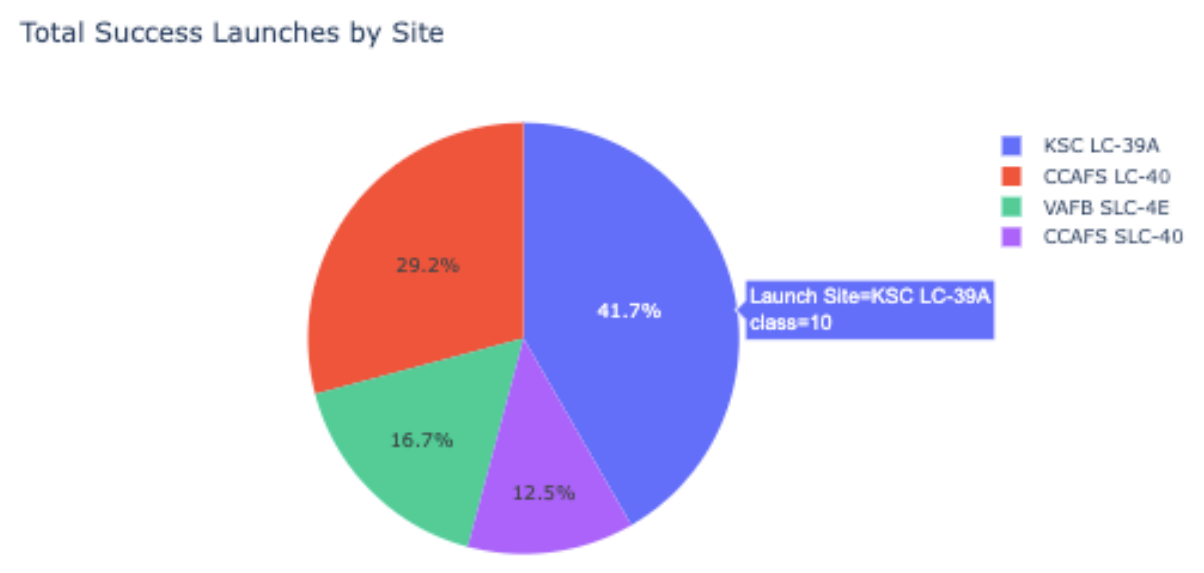
- The map also includes proximities for the VAFB SLC-4E Site
- The sites in general are close to coastlines, railroads and highways and further away from major cities.
- The railroad, highway and coastline were within 2 km of the sites , while the city of Santa Maria is 38 km away.




Section 4.3
Build a Dashboard with Plotly Dash

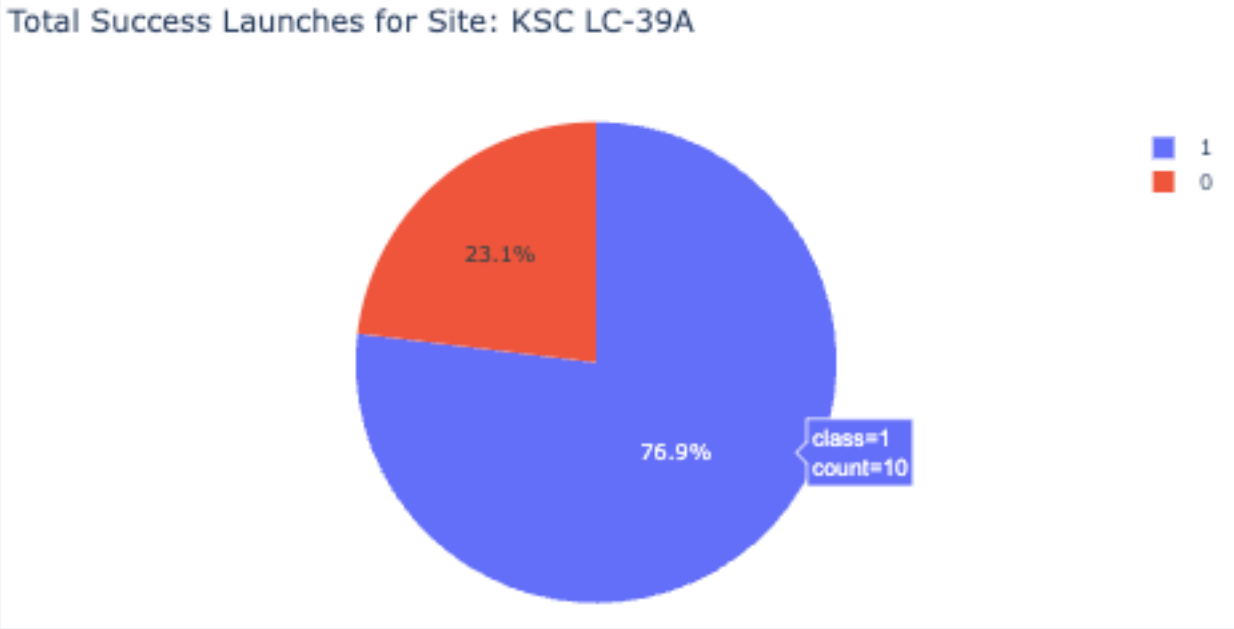
Total Success Launches by Site

- This pie chart uses the count of successful launches to show the portion of successful launches represented by each landing site.
  - KSC LC-39A is the site that has produced the most successful Launches with a count of 10 representing 41.7% of the total successful launches.

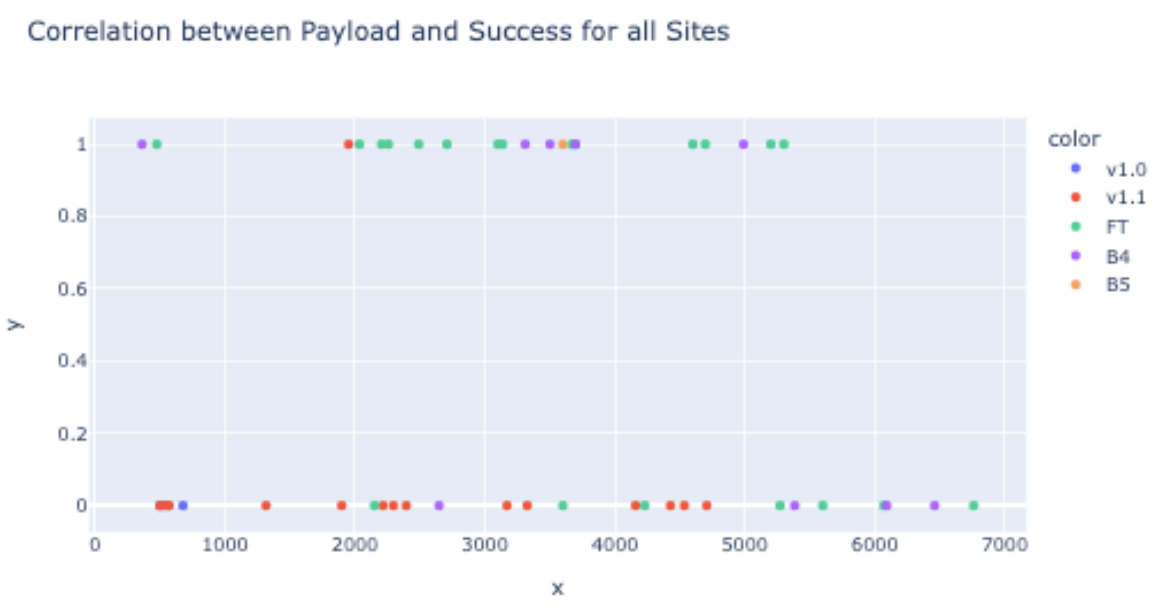


Launch with the highest Success Ratio

- The pie chart compares the success and failure rate for the most successful launch site
  - KSC LC-39, which had the most successful launches also has the highest Launch Success Ratio with 76.9% of its launches being successful



Correlation between Payload and Success for all Sites

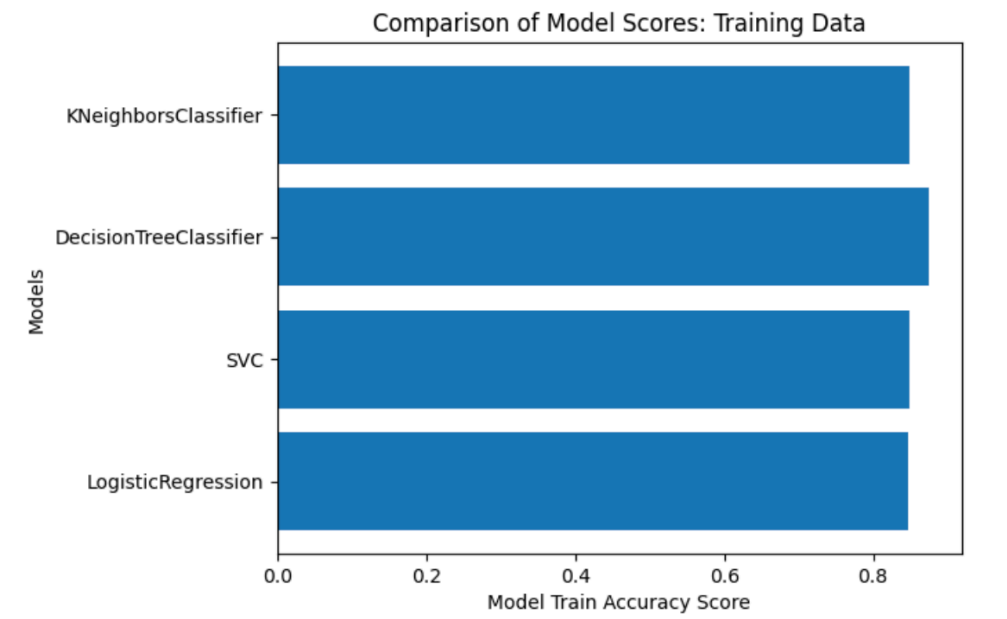


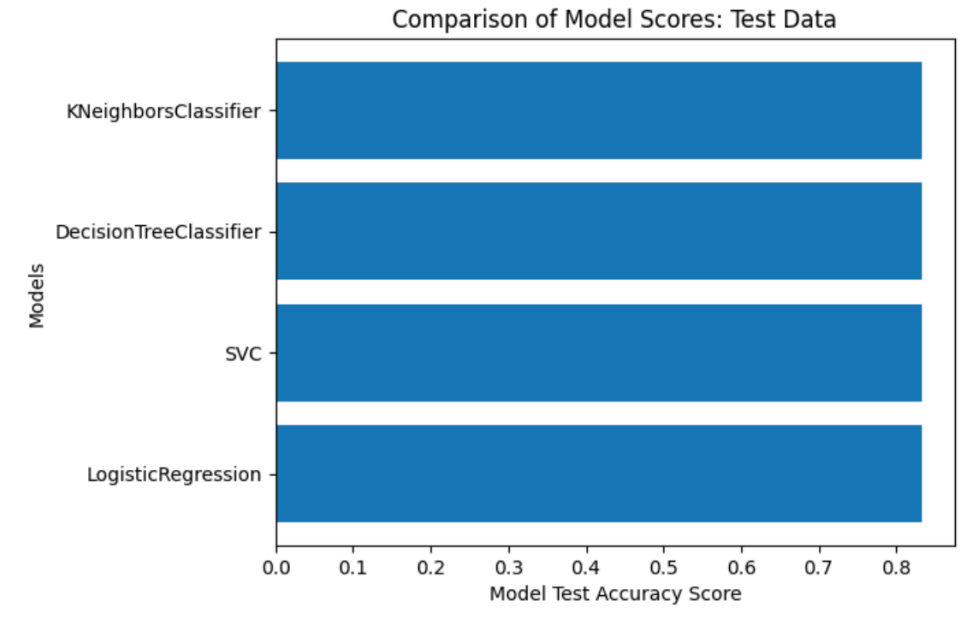
- The scatter charts compare Payload Mass on the x axes to Launch Success on the y axes with 1 being a successful launch and 0 being unsuccessful. The colors representing booster versions. FT booster version is the most successful.
- Filtering the payload Range to 3000 to 4000 kg we observed this payload range had the highest success rate. This success is observed across all booster versions with the exception of booster v1.1.

Section 4.4

Predictive Analysis (Classification)

Classification Accuracy

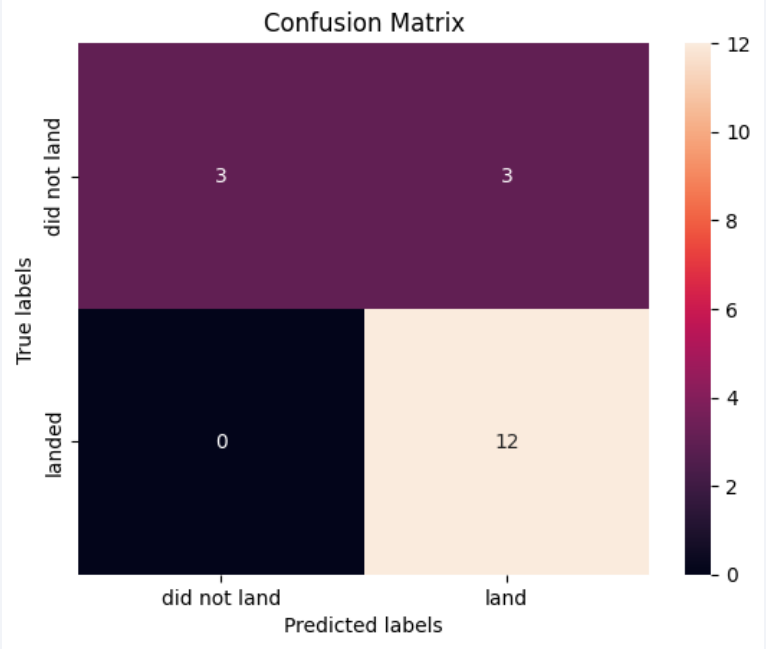




- The bar charts compare the Classification Model Scores on first the Training Data and then the Test Data.
- While all models performed equally on the Test Data, the Decision Tree Classifier performed the best on the Training Data with an accuracy of 0.875.

Confusion Matrix

- The Decision Tree Model had the best accuracy score on the test data of 0.833…
- The model performed well by correctly identifying the 12 launches that landed. However, the model did struggle with False Positives as it categorized 3 of the Launches that did not land as incorrectly landing



Decision Tree Confusion Matrix

Section 5: Conclusion

Conclusions

- Falcon 9 Launch data is available via an api call to the space website as well as via web scraping the wikipedia Falcon 9 Launch tables.
- Falcon 9 Landing outcomes are successful at an overall rate of 66.6% however that success rate has increased to over 80% by 2020.
- It is possible to predict the success of a landing outcome utilizing a machine learning approach. In this model we were able to predict with a reasonable accuracy of 83.333% on our test data via a tuned Decision Tree Classifier.

Section 6: Appendix

Appendix

- Snapshot of Wikipedia Falcon 9 Launches as of 2021-06-09:
  - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!