IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ismet Okan Celik
07/14/2023

Outline

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

Executive Summary

- Summary of methodologies:
  - Data Collection with SpaceX API
  - Data Collection with Web Scraping
  - Data Wrangling using Pandas
  - Exploratory Data Analysis (EDA) using SQL
  - EDA with Data Visualization
  - Launch Sites Location Analysis with Folium and Creating Interactive Dashboards with Plotly Dash
  - SpaceX Machine Learning Landing Prediction
- Summary of all results
  - Exploratory Data Analysis Results
  - Interactive Dashboard Results
  - SpaceX Flacon-9 Landing Prediction with Machine Learning

Introduction

- Project background and context

  SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars. However other providers cost upward of 165 million dollars each, much of the savings is because SpaceX manufactures reusable first-stage rockets. If we would able to determine whether the first stage will land then we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- Problems you want to find answers

  In this project, our main goal is to predict whether Falcon-9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is collected from two sources, these are:
    - SpaceX API
    - Wikipedia by using web scraping approaches
- Perform data wrangling
  - Collected data was organized by using pandas and missing values replaced with columns mean values. Landing outcome feature is created. Categorical variables turn into new features.
- 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 is normalized and split into training and testing data. After that different machine learning models developed. Each model evaluated with different combinations of hyperparamaters

Data Collection

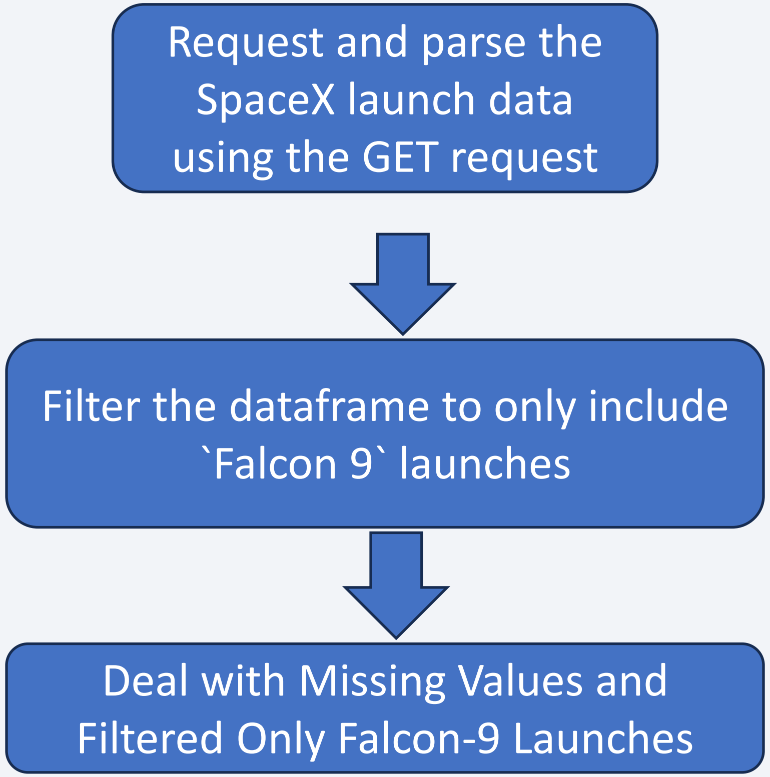
- Description of data collection process for SpaceX Falcon-9 Rockets
    - Data was collected with request.get() method by using a REST API for SpaceX. By using JSON, data content is turned into a Pandas data frame with .json_normalize() method. In this first data frame, most of the data was defined with ID numbers. To be able to solve this problem, API is used again to get information about launches using the IDs given for each launch. Specifically, *rocket, payloads, launchpad, and cores* columns are used with helper functions to pull the numeric data by using ID numbers for launches.
    - Another, method was collecting data was web scraping. Historical data was collected for Falcon 9 from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches). With the help of BeautifulSoup and Request libraries. Falcon-9 launched HTML table from the website page, parsed the table and converted it into a Pandas data frame

Data Collection – SpaceX API

- Presenting data collection with SpaceX REST calls with key phrases and flowcharts:

  Data Collected using SpaceX API, with get request and extracted content decoded to JSON file. JSON filed turned into Pandas data frame. After that, data is filtered and only Falcon-9 launches are kept. Missing values are replaced with mean values of the column.

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/1-jupyter-labs-spacex-data-collection-api.ipynb

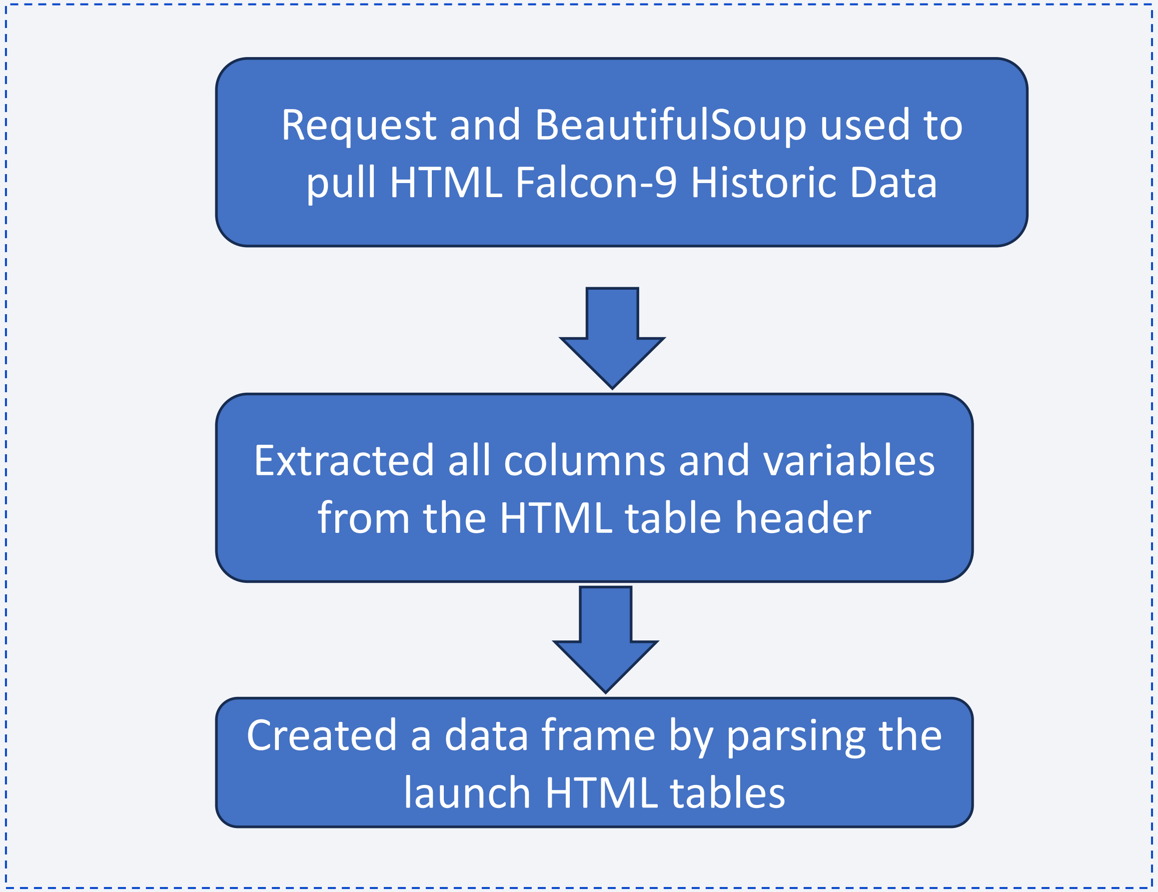


Data Collection - Scraping

- Presenting web scraping process with key phrases and flowcharts:

  Web scraping performed for Falcon-9 launches with BeautifulSoup and Request library from Wikipedia. Historical launch data pulled from the website and HTML table turned into Pandas data frame

- GitHub URL of the completed web scraping notebook:https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/2-jupyter-labs-webscraping.ipynb



Data Wrangling

- Exploratory Data Analysis (EDA) was performed to see some of the patterns in the dataset.
- Identified the percentage of missing values and datatypes.
- Calculated Launches for per site, occurrences of each orbit and occurrences of mission outcome per orbit type.
- Lastly, created a landing outcome label from outcome column
- Present the data wrangling process with key phrases and flowcharts:

EDA → Summarizations and Calculations → Creating Landing Outcome Column

- GitHub URL of your completed data wrangling related notebooks: https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/3-module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Summarizing what charts were plotted and why those charts used:
  - <u>**FlightNumber vs. PayloadMass with Outcome (ScatterPlot):**</u> Observing relationship between FlightNumber and PayloadMass and their effect on the success rate of the launches.
  - <u>**FlightNumber vs. LaunchSite with Outcome (ScatterPlot):**</u> Observing relationship between FlightNumber and LaunchSites and their effect on the success rate of the launches.
  - <u>**PayloadMass vs. LaunchSite with Outcome (ScatterPlot):**</u> Observing relationship between PayloadMass and LaunchSites and their effect on success rate of the launches.
  - <u>**SuccessRate vs. Orbit Type (BarPlot):**</u> Observing what is the success rate of different orbits.
  - <u>**FlightNumber vs. OrbitType (ScatterPlot):**</u> Observing relationship between FlightNumber and OribitType
  - <u>**PayloadMass vs. OrbitType (ScatterPlot):**</u> Observing relationship between PayloadMass and OribitType
  - <u>**SuccessRate vs. Year (LinePlot):**</u> Observing success rate by yearly trend.
- GitHub URL of completed EDA with data visualization notebook: https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/5-jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Summarizing the SQL queries which performed during EDA:
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string 'CCA'
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - Listing the date when the first succesful landing outcome in ground pad was acheived
  - Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Listing the total number of successful and failure mission outcomes
  - Listing the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - Listing the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.¶
  - Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- GitHub URL of completed EDA with SQL notebook:https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/4-jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb

Build an Interactive Map with Folium

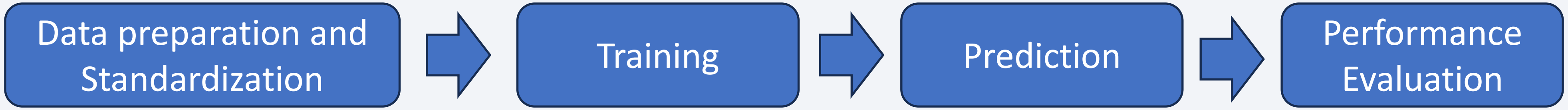
- Markers, circles, lines, and marker clusters created and added to a folium map:
  - Markers represent Launch Sites
  - Circles represenT areas around the Launch Sites
  - Marker clusters represents a group of events that occurred in specific launch site
  - Lines represent the distances between launch sites and specific location such as coast, train rail, highway etc.
- GitHub URL of completed interactive map with Folium map:https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/6-module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summarizing what plots/graphs and interactions added to a dashboard:
  - Dropdown menu for Launch Sites added, which helps user to choose specific launch site.
  - Interactive pie chart for Launch Sites added which shows number of launches for ALL and Specific Launch Sites.
  - Payload Mass (kg) vs. Landing Outcome scatter plot added with Booster Versions in different color. User can see what is the successful landing rate for different boosters and Payloads.
  - Payload Range Slidder added. User can change the Payload Range to see success rate for different ranges and different booster types.
- GitHub URL of completed Plotly Dash lab: https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/7-spacex_dash_app.py.ipynb

Predictive Analysis (Classification)

- Summarizing how the model built, evaluated, improved, and found the best performing classification model:
    - Data is turned into data frame, and a numpy array is created for the Class column.
    - Data is standardized by using StandardScaler() and saved as Training data as X without class labels. Class labels saved as Y variable.
    - Data is splitted 80% as training data and 20% as test data.
    - Four different machine learning approaches were applied (logistic regression, support vector machine, decision tree, and k-nearest neighbors.
    - To be able to improve the performance of the models, hyperparameters picked by GridSearch.
    - Models are evaluated by using score method for both training and prediction. Also confusion matrix plotted to see outcome of the predictions.
- Model Development Flow Chart:



- GitHub URL of your completed predictive analysis: https://github.com/okisna93/Applied_Data_Science_Capstone_Project/blob/main/8-module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

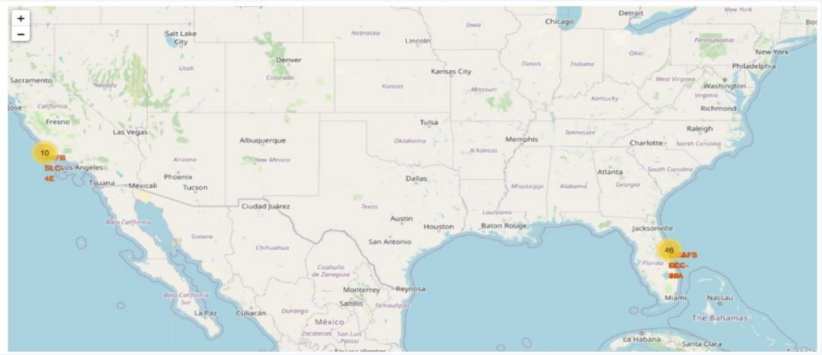
Results

- Exploratory data analysis results
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA
  - The average payload of F9 v1.1 booster is 2,928 kg
  - The first success landing outcome happened in 2015 fiver year after the first launch
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  - Almost 100% of mission outcomes were successful
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
  - The number of landing outcomes became as better as years passed.

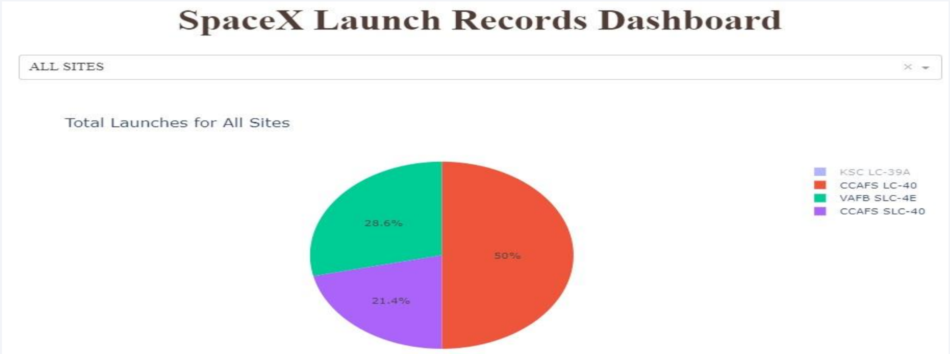
Results

- Interactive analytics demo in screenshots

  By using interactive analytics we see that launch sites have similar characteristic in terms of proximity to the coast, highway, train rails, and cities in terms of logistic infrastructure, safety issues and transportation.



- Interactive dashboards make easier to compare the effect of different features on successful landing



Results

- Predictive analysis results:

  As we can see the accuracy of the Decision Tree is higher than other Machine Learning algorithms. However, if we compare the test accuracy of the all the models that we can see that all models performs the same.

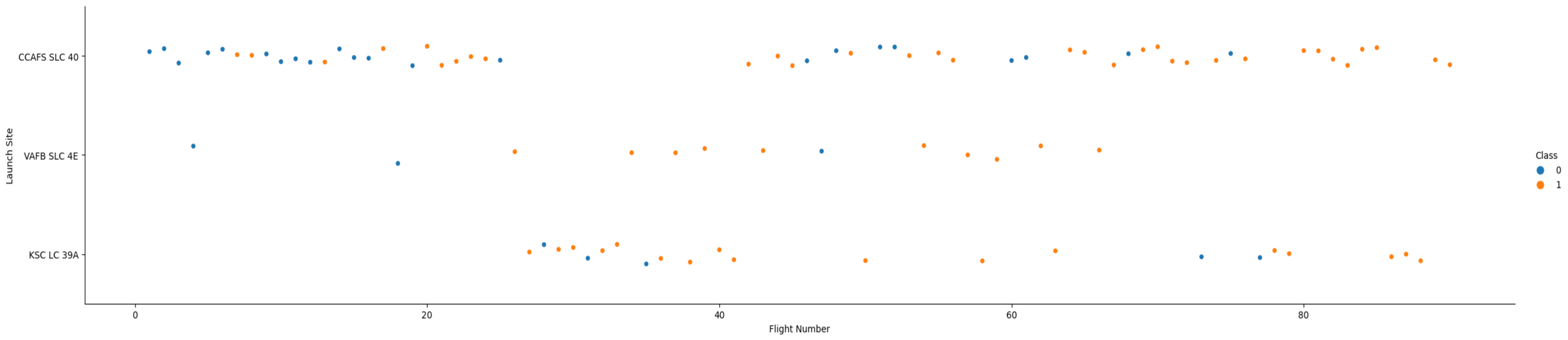
	Logistic Regression	SVM	Decision Tree	KNN
Training Acccuracy	0.846429	0.848214	0.875000	0.848214
Testing Accuracy	0.833333	0.833333	0.833333	0.833333

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site

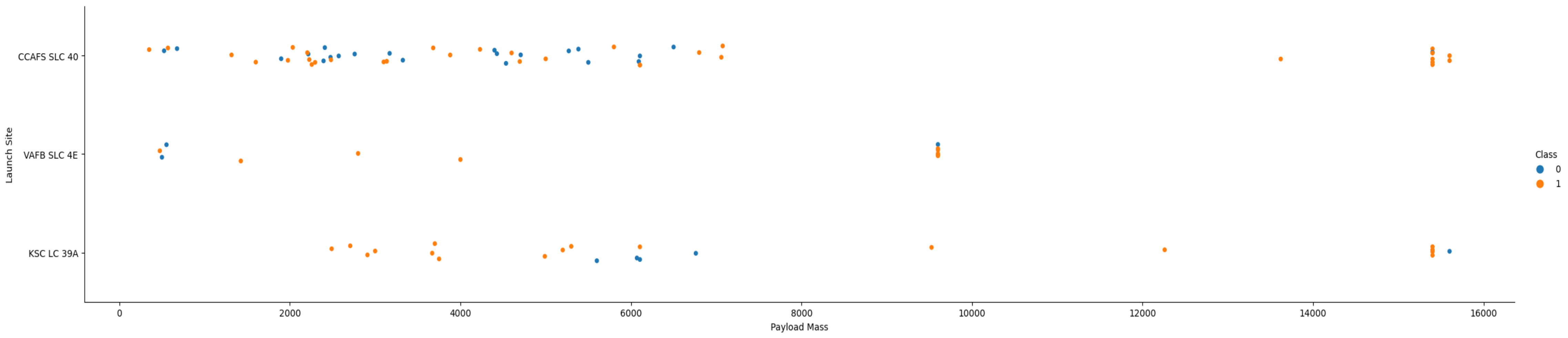


- Explanations:

As we can see from the graph above, different launch site has different success rate, but the general outcome is if the number of flight attempt increases, the success rate also increases linearly. However, if we compare three different launch sites. VAFB SLCE 4E, KSC LC 39A hosted a more successful launch than CCAAFS SLC 40, but CCAAFS SLC 40 launch site was used more than other sites.

Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site

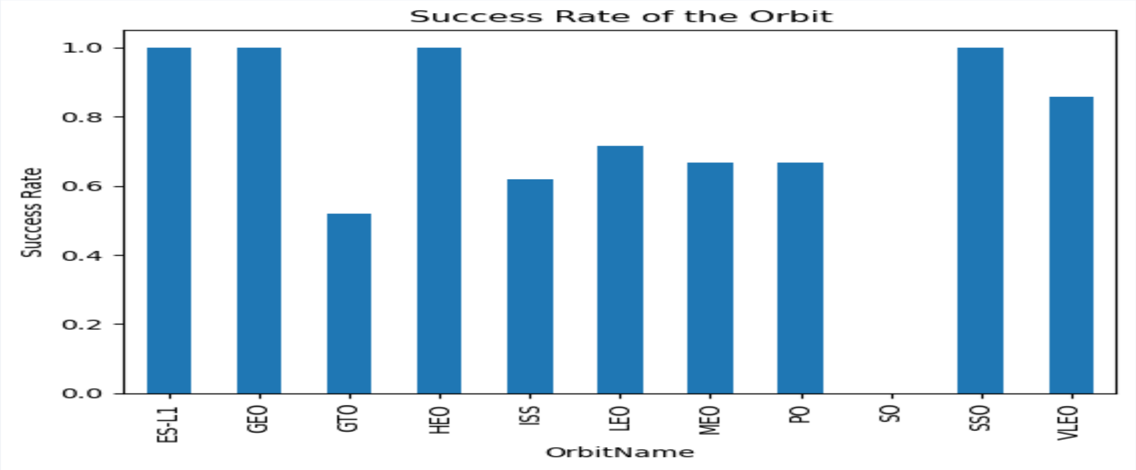


- Explanations:

If you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type

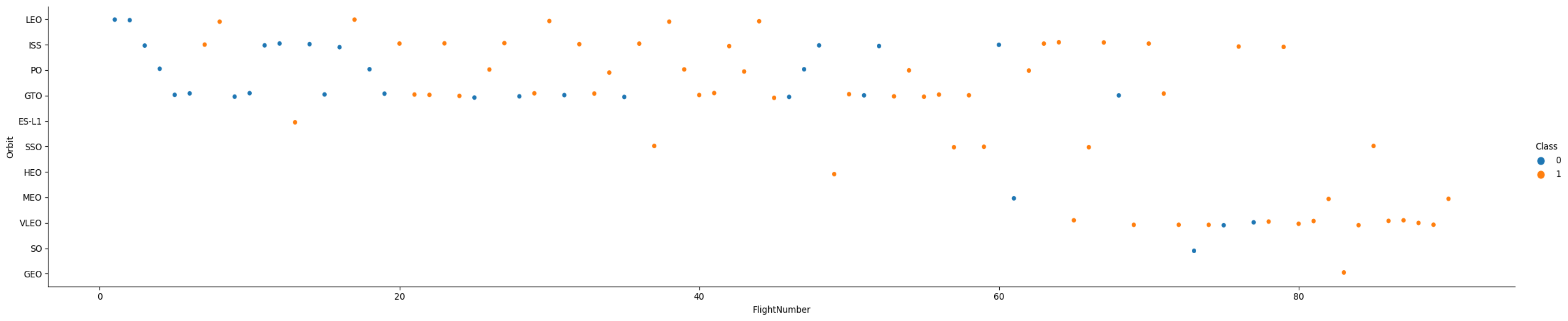


- Explanations:

As we can see ES-L1, GEO, HEO, SSO orbits are the most successful orbits

Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type

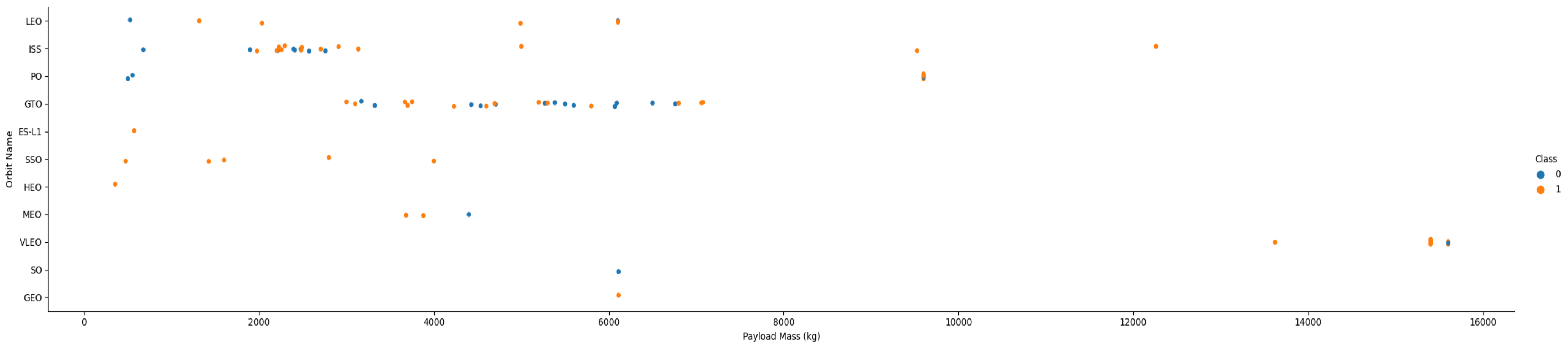


- Explanations:

LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

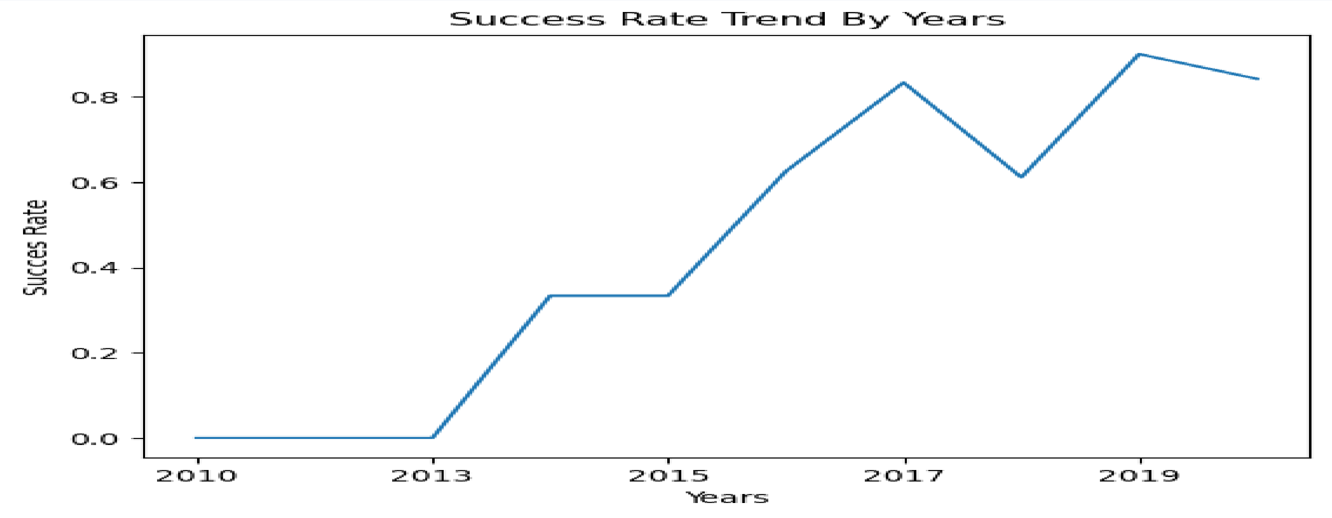
- Show a scatter point of payload vs. orbit type



- Explanations:
  - With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
  - However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- Explanations:

You can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Unique launch sites:

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

- Explanation:

They are obtained by selecting unique occurrences of “launch_site” values from the dataset

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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

Total Payload Mass

- Total payload carried by boosters from NASA
- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA

Customer	TOTAL_MASS_KG
NASA (CRS)	45596.0

Average Payload Mass by F9 v1.1

- Calculated the average payload mass carried by booster version F9 v1.1
- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg

Booster_Version	AVERAGE_MASS_KG
F9 v1.1	2928.4

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad
- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence

Date_First_Succesful_Landing
01/07/2020

Successful Drone Ship Landing with Payload between 4000 and 6000

- Listing the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Selecting distinct booster versions according to the filters above, these 4 are the result

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

- Calculate the total number of successful and failure mission outcomes
- Grouping mission outcomes and counting records for each group led us to the summary above

Mission_Outcome	COUNT(Mission_Outcome)
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Listing the names of the booster which have carried the maximum payload mass
- These are the boosters which have carried the maximum payload mass

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

- Listing the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- There are two occasions that landing in drone ship was failed in 2015

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

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

- Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending orde
- As we can see that there are almost 29 successful lading has been made.

Date	Landing_Outcome	Count
08/07/2018	Success	19
10/08/2012	No attempt	9
18/07/2016	Success (ground pad)	5
14/08/2016	Success (drone ship)	5
14/04/2015	Failure (drone ship)	3
12/05/2018	Failure	3
06/04/2010	Failure (parachute)	2
18/04/2014	Controlled (ocean)	2
08/06/2019	No attempt	1

Section 3

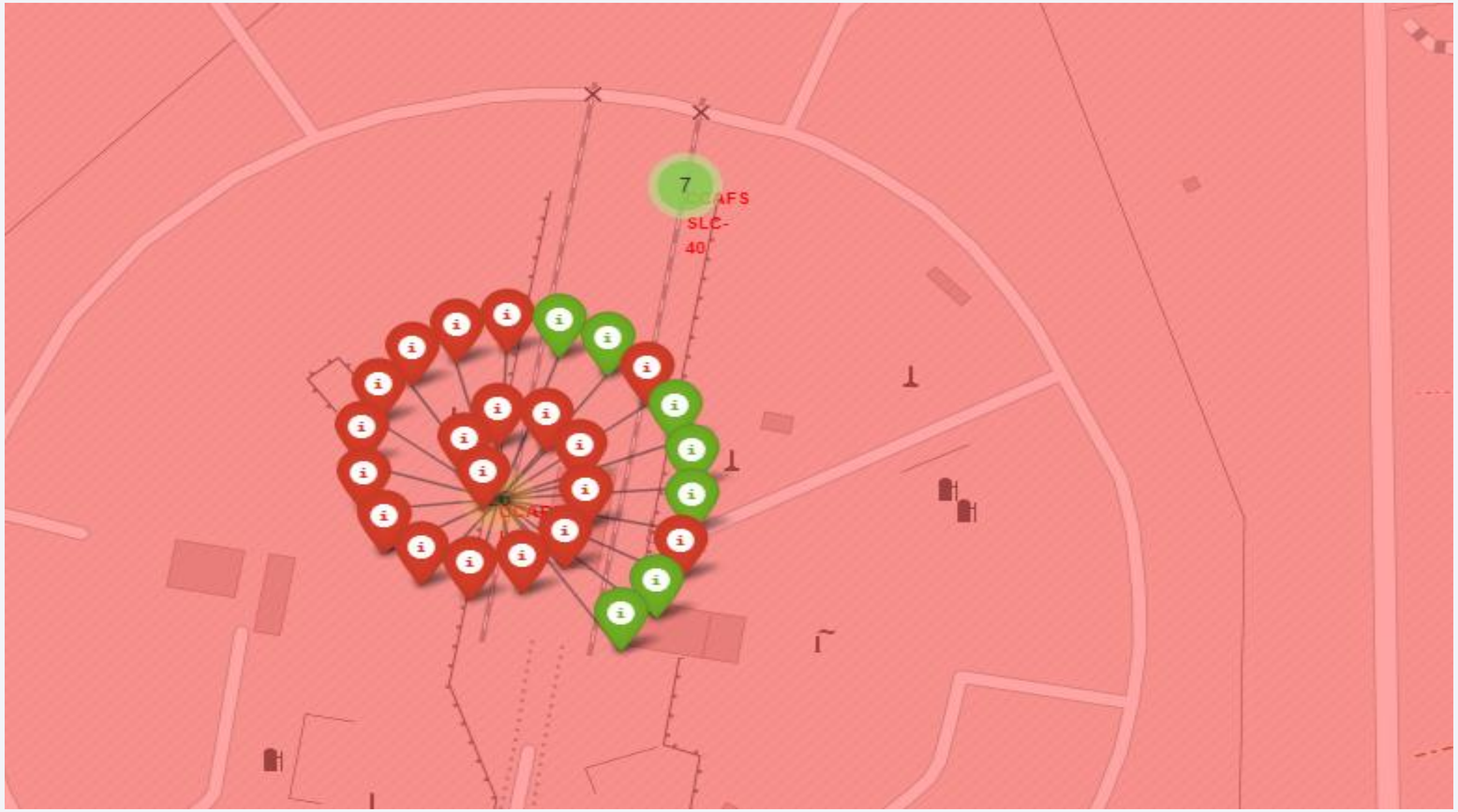
Launch Sites Proximities Analysis

All Launch Sites

- Locations of launch sites follow the same pattern. They are near the coast and far from the cities for safety reasons. Close to the railroad and highway to make it easier to carry cargo to rockets and human transportation to launch sites.

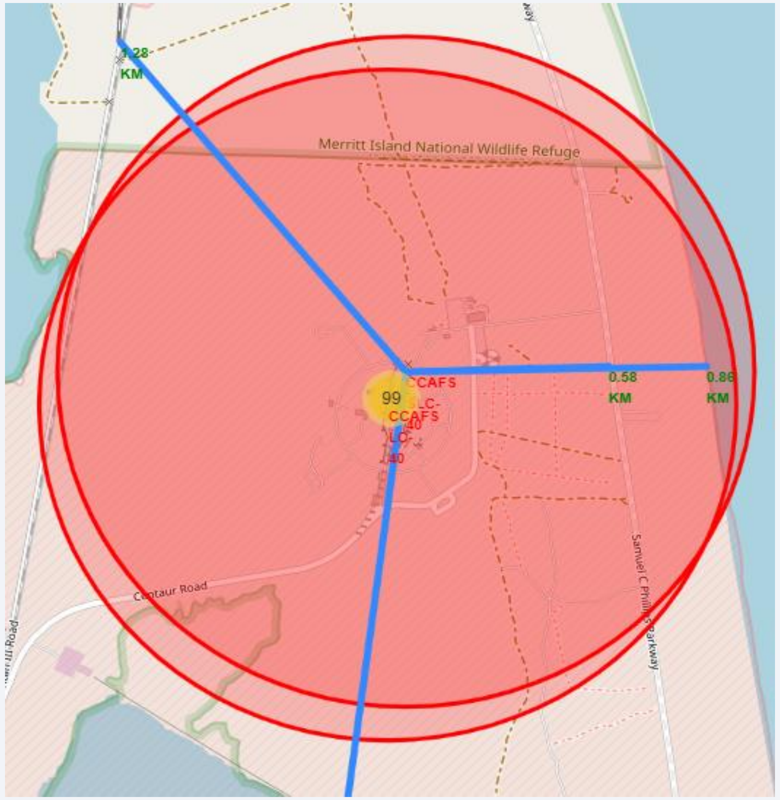
Launch Outcomes for Sites

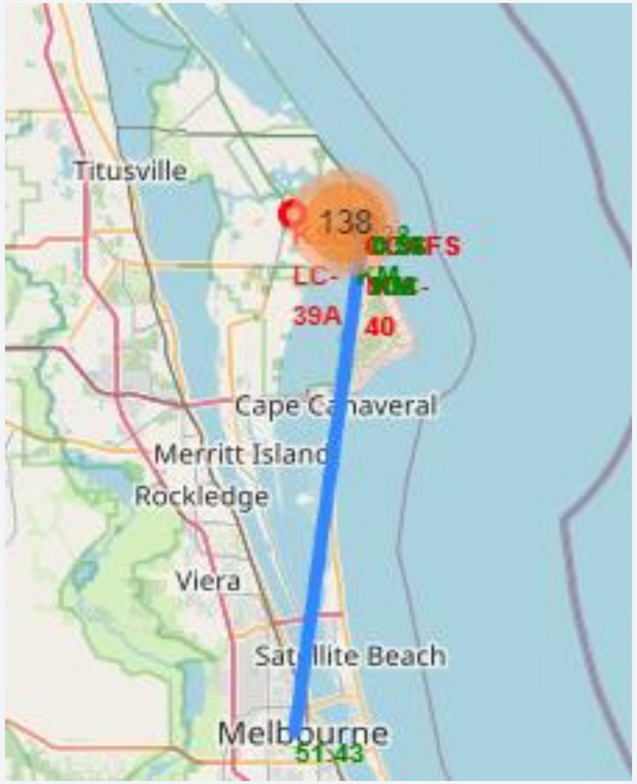
- As we can see in the picture, number of launches made for each facility shows with a number. The total launches for CCAFS-SLC-40 is 7. If we click on this number we can see how many launches from that site are successful and how many of them failed. The green icon shows success, and the red icon shows failure.



Distance to Highway, City, and Railroad

- As we can see from the picture that CCAFS-SLC-40 is almost the middle of the coast, highway, and railroad. However, it's far from the city. The second picture shows the distance between Launch Site and Melbourne.



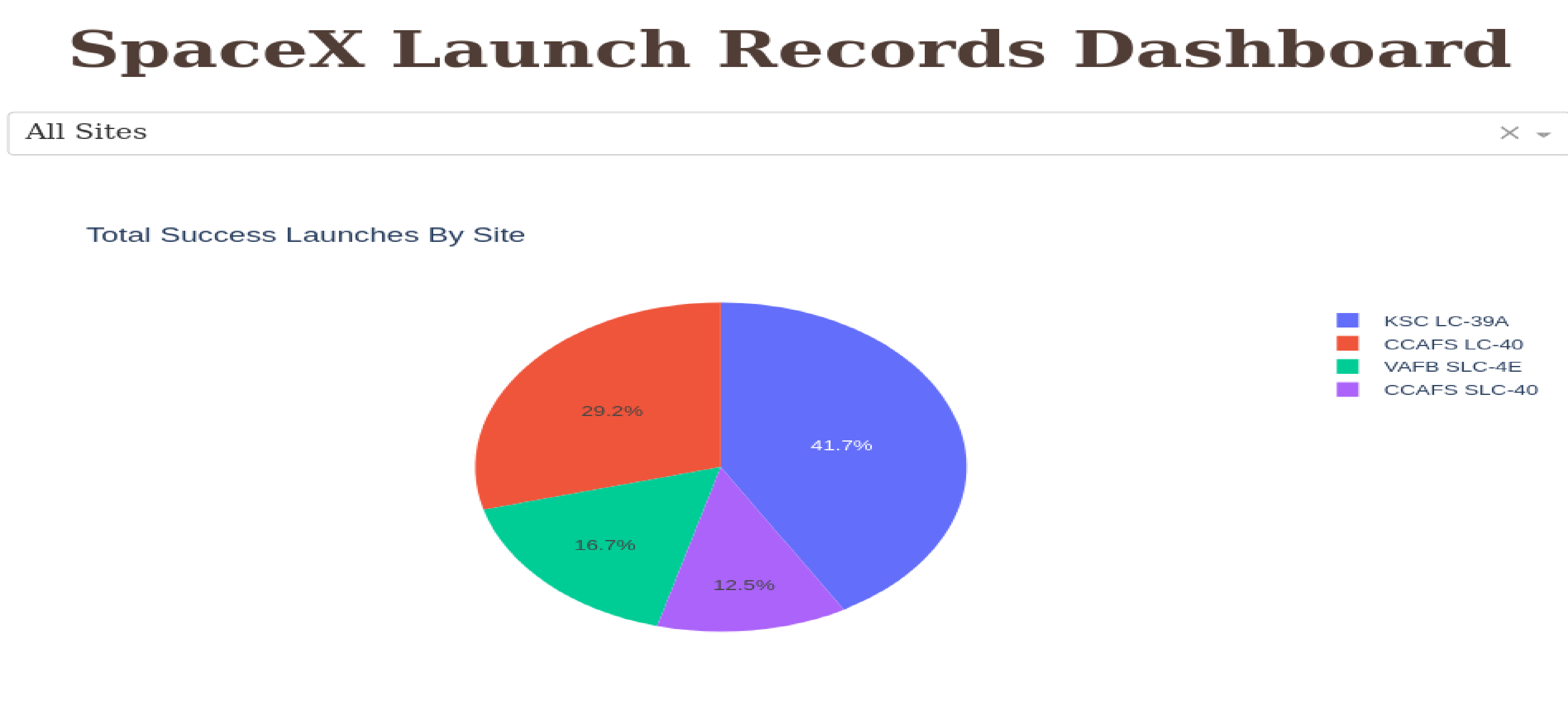


Section 4

Build a Dashboard with Plotly Dash

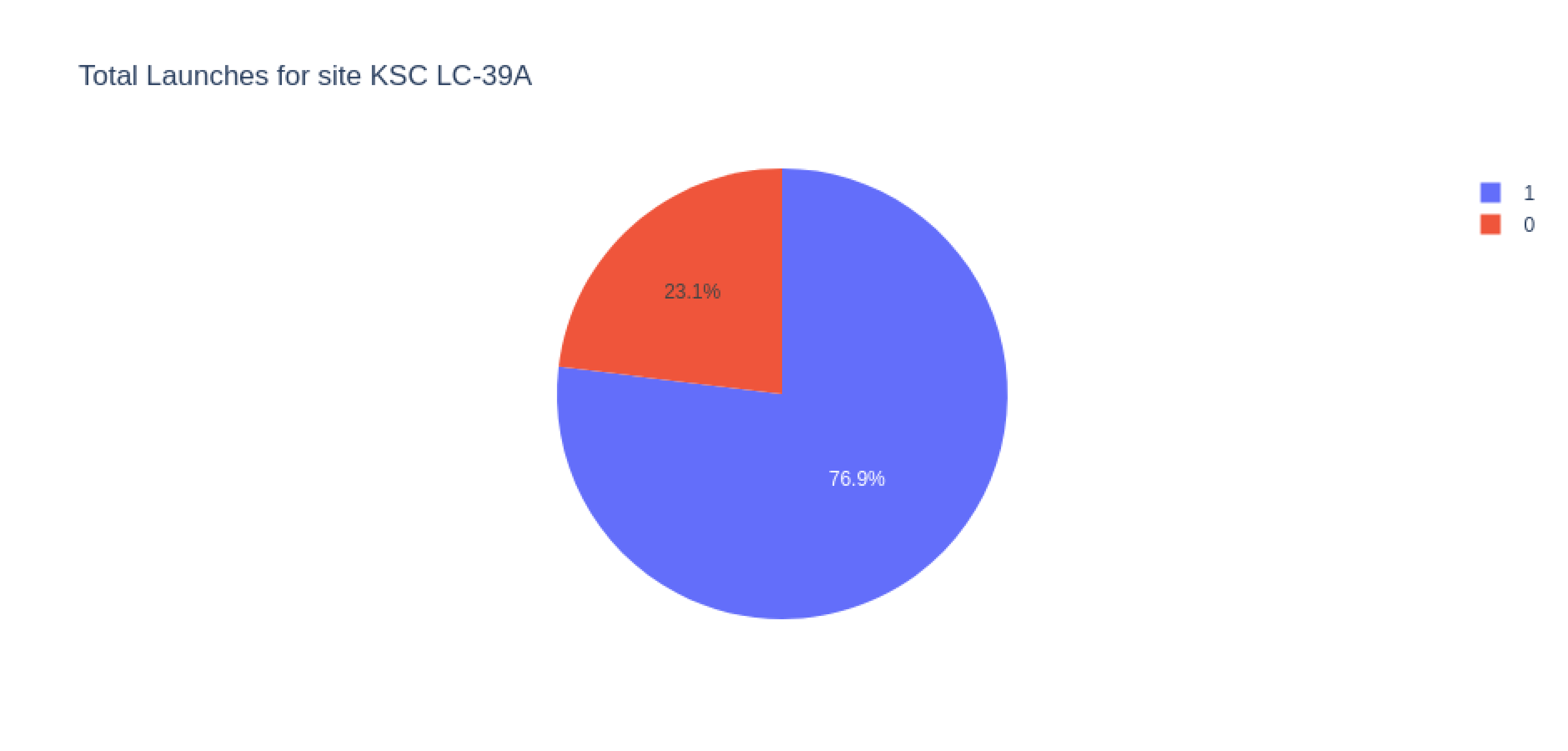
Successful Launches for All Sites

- Each site has a different success rate. The most successful site seems to be KSC LC-39



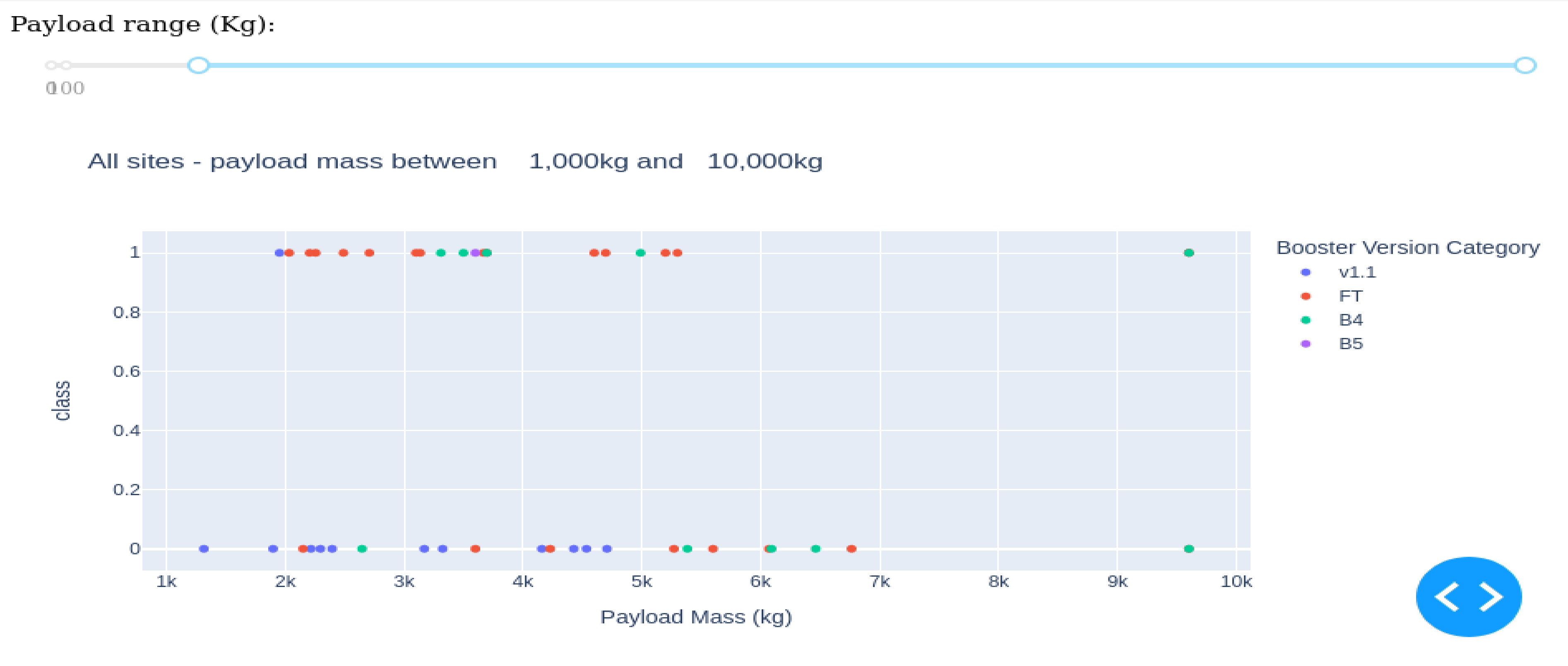
Most Successful Launch Site

- KSC LC-39 is most successful site with 76.9% success rate



Payload vs. Launch Outcome with Booster Version

- Launches have payloads less the 6000 kg seem to be more successful. Most of these launches were made with the FT booster model.



Section 5

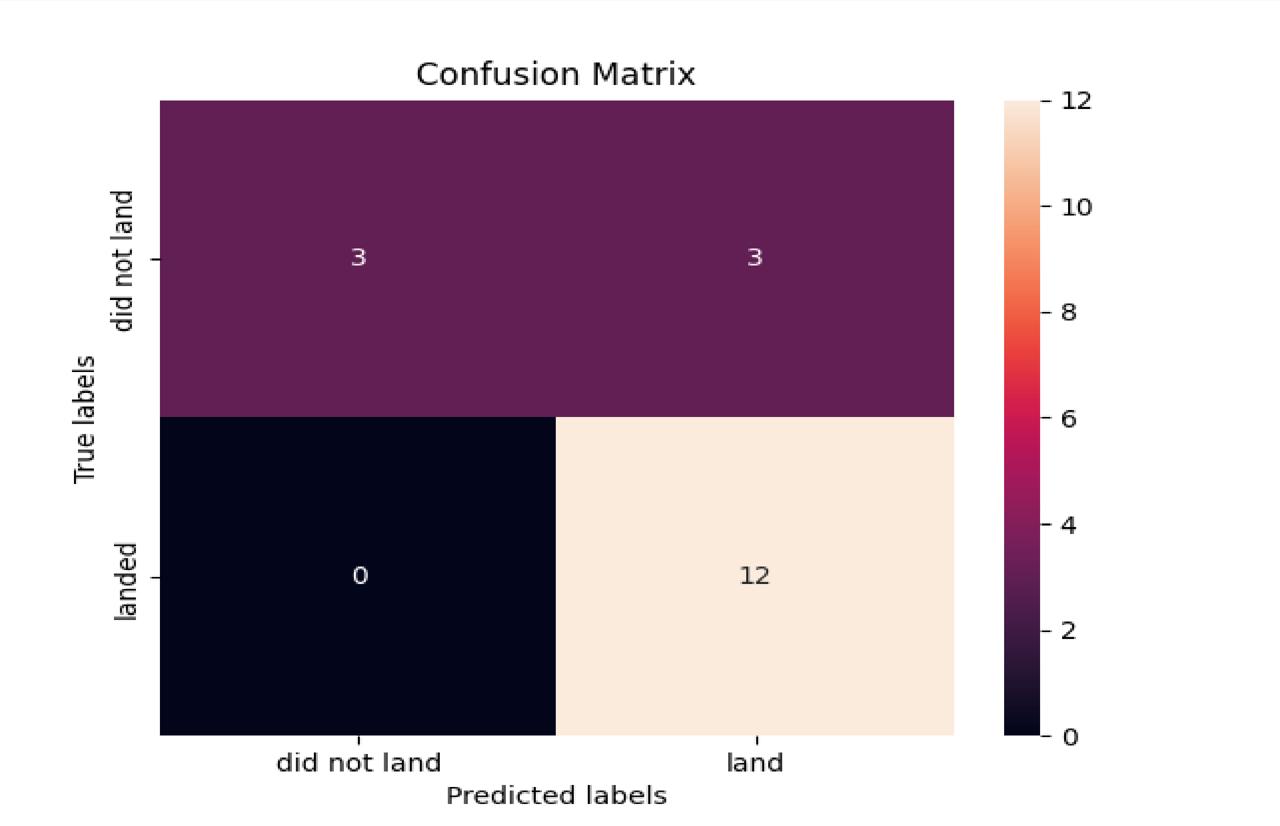
Predictive Analysis (Classification)

Classification Accuracy

- Four different machine learning model developed. Training and testing accuracies plotted.
- Almost every model performs similarly. However, the decision tree has the potential to outperform other models.

Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positives and true negatives compared to the false ones



Conclusions

- Different data sources were analyzed, refining conclusions along the process.
- Most successful launch site is KSC LC-39A;
- All launch site locations follow the same pattern.
- Launches above 7,000kg are less risky.
- Success rate of the launches increased over time, but first big jump was made in 2013
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- Not any outside resources were used during this project.

Thank you!