IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nelson MASANGANISE
17/05/2025

Outline

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

Executive Summary

- In this capstone, we will predict if the SpaceX Falcon 9 first stage will land successfully.
- If we can determine if the first stage will land, we can determine the cost of a launch.
- The methodology used in this project involves:
  - ❖ Data collection and data wrangling
  - ❖ Exploratory data analysis
  - ❖ Data visualization
  - ❖ Model development and model evaluation
  - ❖ Reporting your results to stakeholders.
- The results of this project showed us that some rocket launch features are correlated to the outcome of the launch on whether is a success or failure.

Introduction

- Space Y that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
- We will also determine if SpaceX will reuse the first stage.
- We will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX launch data was gathered from an API, specifically the SpaceX REST API.
  - Falcon 9 Launch data was obtained by web scraping related Wiki pages.
- Perform data wrangling:
  - Calculate the number of launches on each site.
  - Calculate the number and occurrence on each orbit and the number and occurrence of mission outcome of the orbits.
  - Convert the landing outcomes to Classes y (either 0 or 1).
  - 0 is a bad outcome, that is, the booster did not land. 1 is a good outcome, that is, the booster did land.
  - The variable Y will represent the classification variable that represents the outcome of each launch.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL:
  - Create scatter plots and bar charts by writing Python code to analyse data in a Pandas data frame
  - Write Python code to conduct exploratory data analysis by manipulating data in a Pandas data frame
  - Create and execute SQL queries to select and sort data
  - Utilize data visualization skills to visualize the data and extract meaningful patterns to guide the modelling process.

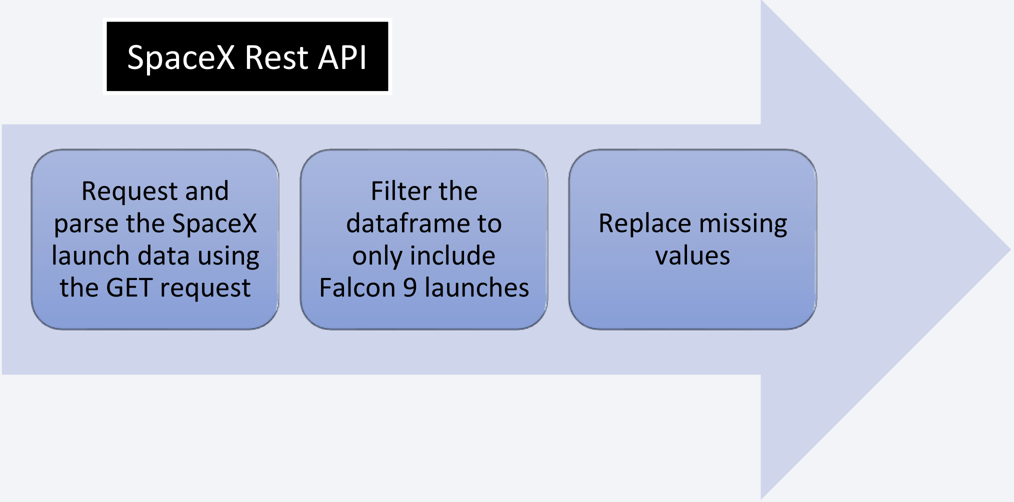
Methodology

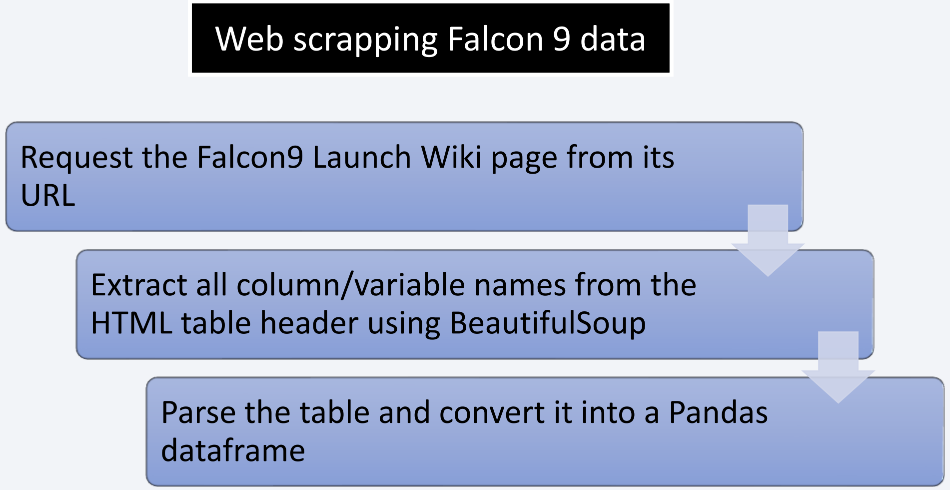
Executive Summary

- Perform predictive analysis using classification models:
  - Build a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.
  - This will include:
    - ❖ Preprocessing, allowing us to standardize our data.
    - ❖ Train_test_split, allowing us to split our data into training and testing data.
    - ❖ Train the model and perform Grid Search, allowing us to find the hyperparameters that allow a given algorithm to perform best
  - Use the best hyperparameter values to determine the model with the best accuracy using the training data.
  - The following models will be tested: Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors
  - Plot the confusion matrix to evaluate the performance of the models.

Data Collection

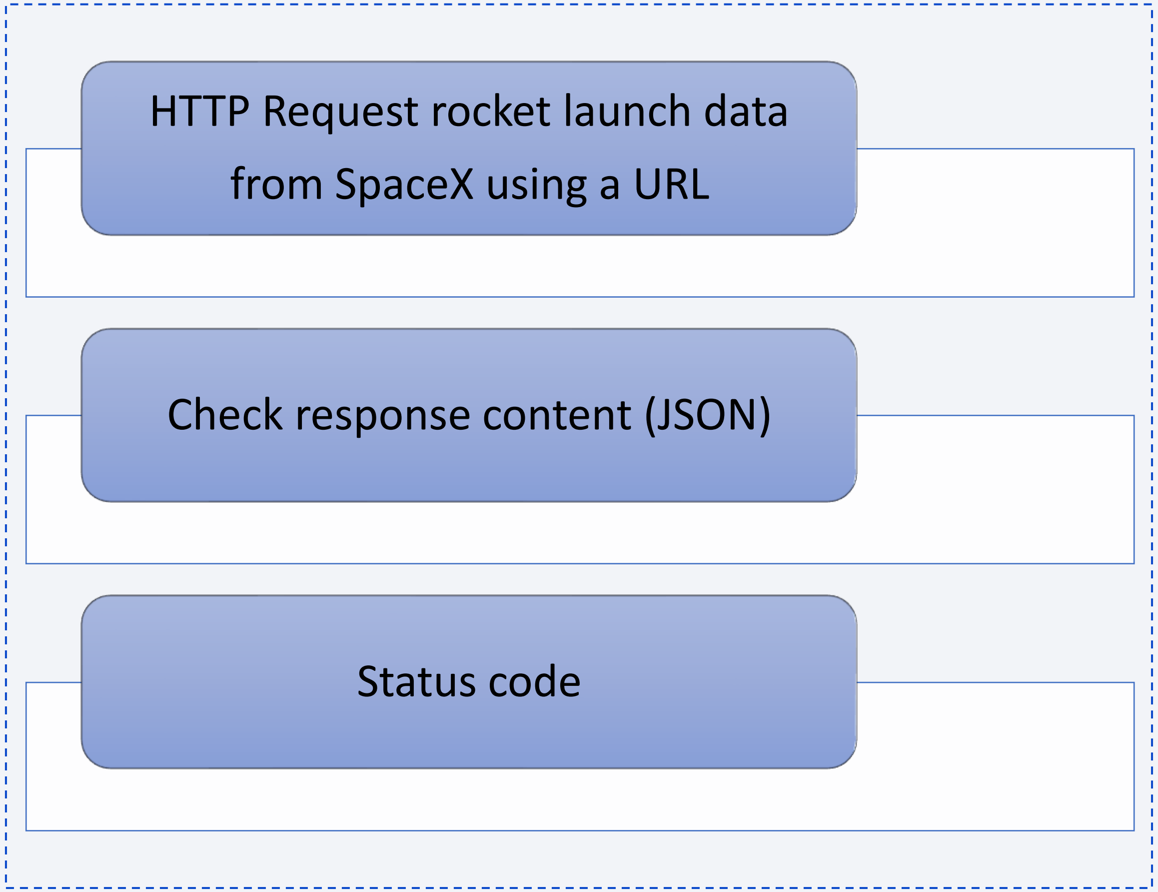
- SpaceX REST API gave us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Web scrapping was used to scrape Falcon 9 Launch data from related Wiki pages.
- The flowcharts are as follows:





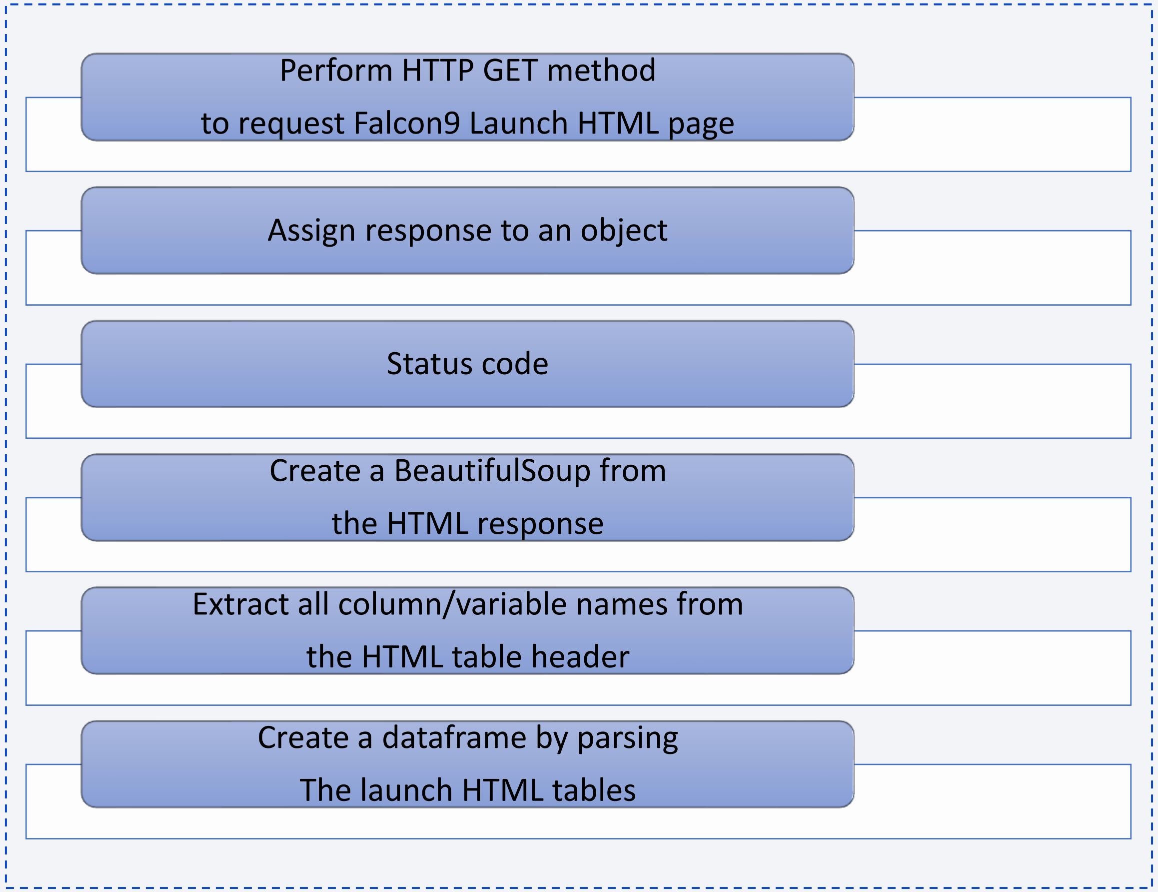
Data Collection – SpaceX API

- The steps for data collection using the SpaceX REST API are shown in the flow diagram.
- The response content is decoded using .json and is turned into a Pandas dataframe using .json_normalize()
- The completed notebook is available on this URL: [Data Collection API]



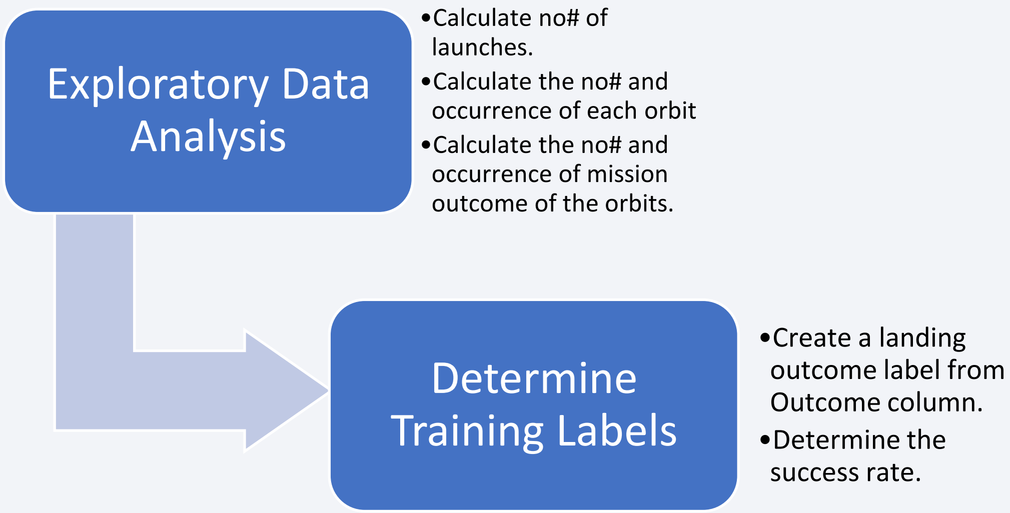
Data Collection - Scraping

- The steps for data collection by scrapping the ‘List of Falcon 9 and Falcon Heavy launches’ Wikipage are shown in the flow diagram.
- The completed notebook is available on this URL: [Data Collection with Web Scrapping]



Data Wrangling

- Once the data was collected, we performed data wrangling as per the steps shown in the process flow.
- The completed notebook is available on this URL: [Data Wrangling]



EDA with Data Visualization

- Pandas and Matplotlib were used perform exploratory Data Analysis and Feature Engineering
- The charts listed below were used in this step and the completed notebook is available on the following URL: EDA

Type of Chart	Reason
Scatter plot	• To see how the FlightNumber and Payload variables would affect the launch outcome. • To visualize the relationship between Flight Number and Launch Site. • To visualize the relationship between Payload Mass and Launch Site. • To visualize the relationship between FlightNumber and Orbit type. • To visualize the relationship between Payload Mass and Orbit type.
Bar chart	• To visualize the relationship between success rate of each orbit type.
Line chart	• To visualize the launch success yearly trend.

EDA with SQL

- Several SQL queries were performed in this step to achieve the following:
- Display the names of the unique launch sites in the space mission using DISTINCT.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the total number of successful and failure mission outcomes.
- List all the booster_versions that have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank 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.
- The complete notebook is available here: [EDA SQL]

Build an Interactive Map with Folium

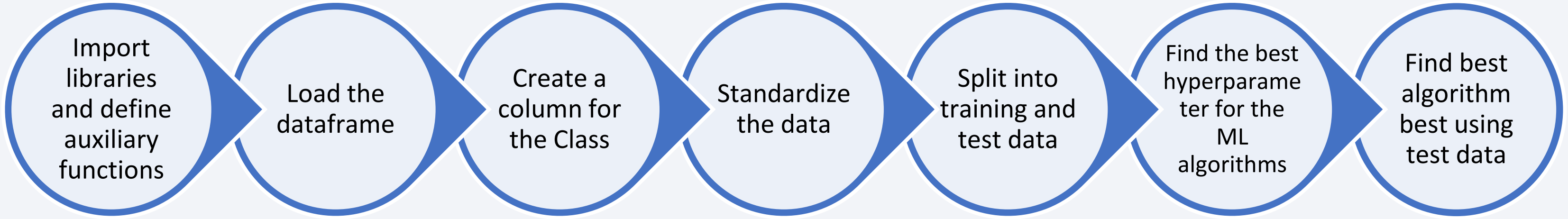
- To mark all launch sites on a map, we used the following objects: Map, Circle, and Marker.
- To mark the success/failed launches for each site on the map, we used the following objects: MarkerCluster.
- To calculate the distances between a launch site to its proximities, we used th following objects: Marker and PolyLine.
- The complete notebook can be found here: Map with Folium

Build a Dashboard with Plotly Dash

- For the dashboard, we added a Launch Site Drop-down Input Component. This is required to select one specific site and check its detailed success rate (class=0 vs. class=1).
- We also added a callback function to render success-pie-chart based on selected site dropdown.
- To select the Payload, we added a Range Slider. This enables us to find whether the variable payload is correlated to mission outcome.
- We added a callback function to render the success-payload-scatter-chart scatter plot. This enables use to visually observe how payload may be correlated with mission outcomes for selected site(s).
- The completed notebook can be found here: [Dashboard with Plotly Dash]

Predictive Analysis (Classification)

- The model development process adopted for this step is summarized below.



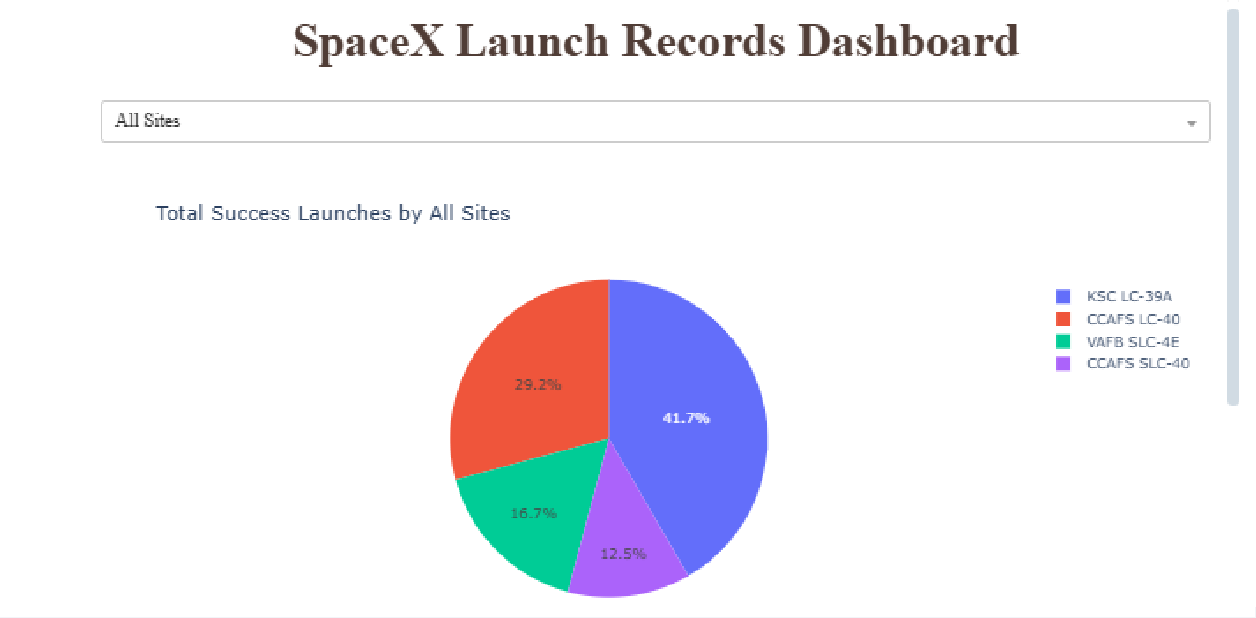
- The models tested in this step include SVM, Classification Trees, and Logistic Regression.
- The complete notebook is available here: [Machine Learning]

Results

- It was noted from the EDA that the launch success rate since 2013 kept increasing till 2020.
- This was based on the EDA using Pandas.
- From the EDA using SQL, it was observed that 5 boosters have success in drone ship and have payload mass greater than 4000 kg but less than 6000 kg.

Results

- Interactive analytics demo in screenshots



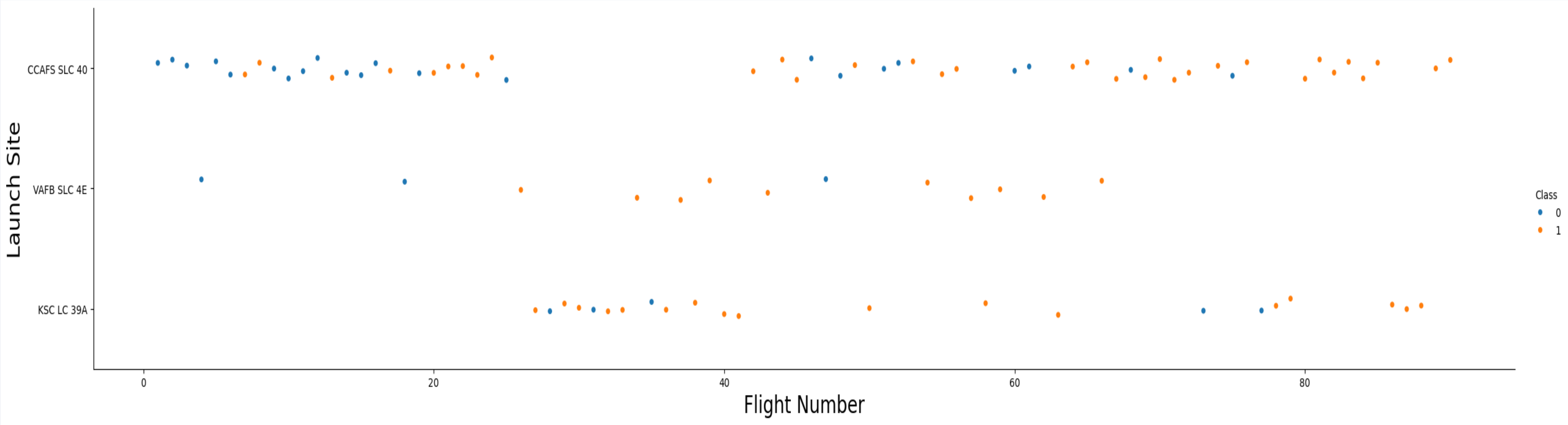
Results

- The following results were obtained from the predictive analysis results:
- 18 records were in the test sample.
- The sigmoid kernel had the best result on the validation dataset using the SVM.
- The data accuracy achieved for the decision tree classifier using the validation data was 83.33%

Section 2

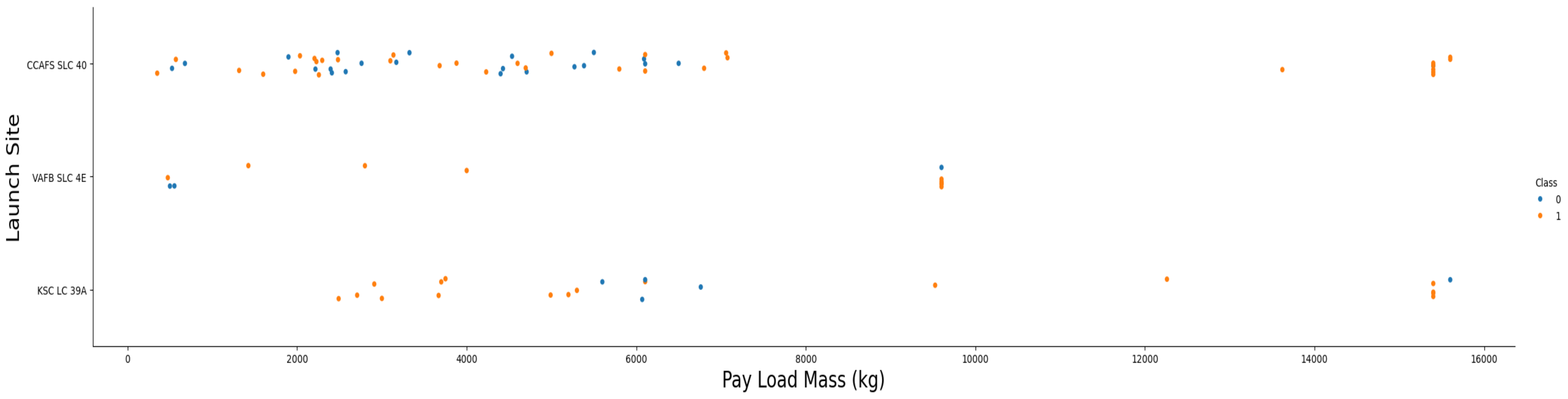
Insights drawn from EDA

Flight Number vs. Launch Site



- It can be seen that the higher the number of flights at a launch site, the higher the successful launches.
- Furthermore, the CCAFS SLC 40, had a balanced number of flights ranging from 1 up to > 80.

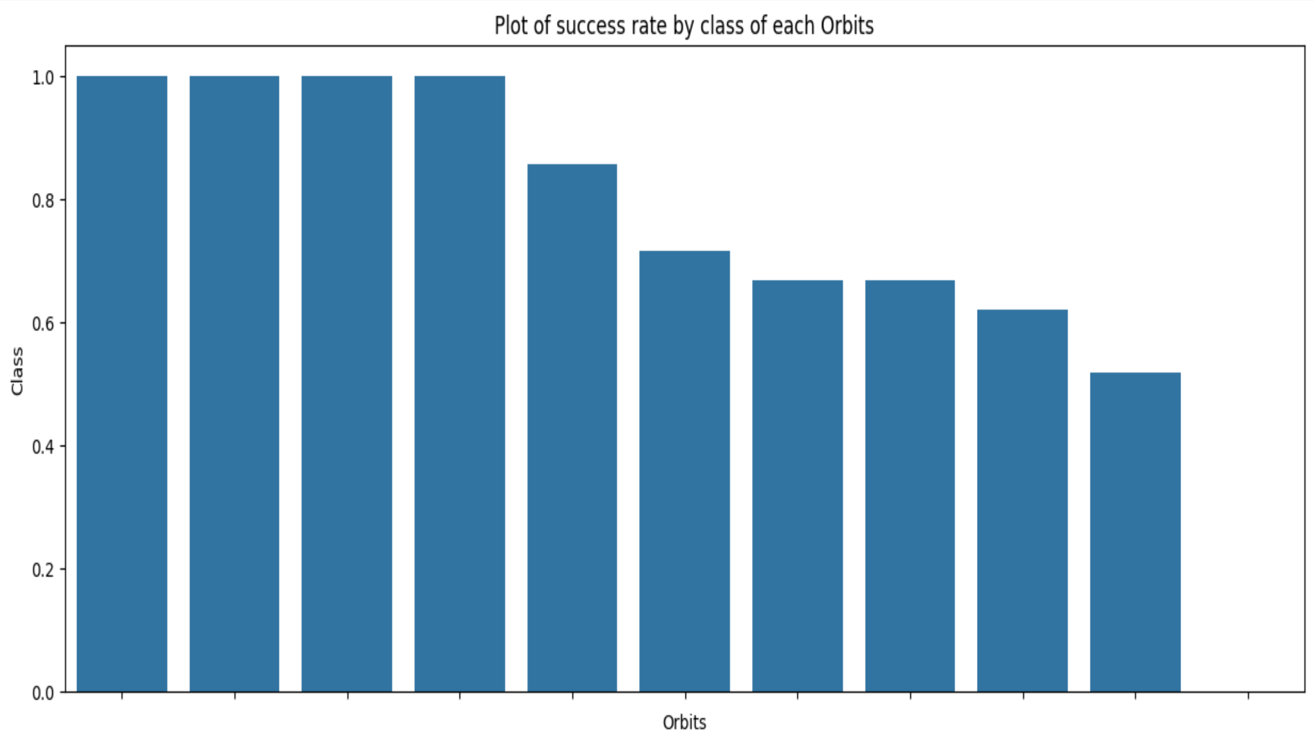
Payload vs. Launch Site



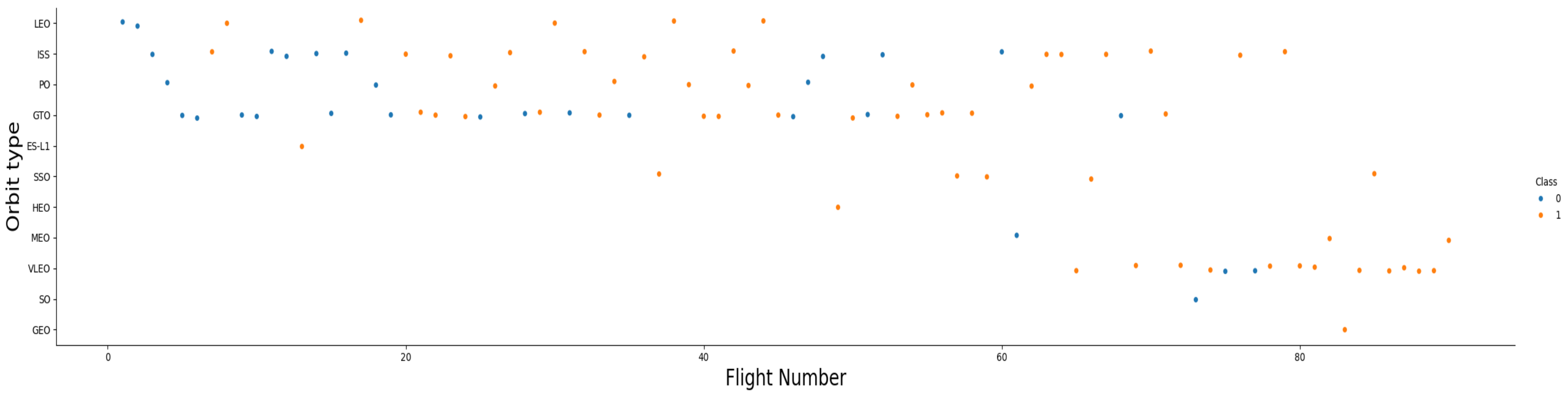
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000 kg).
- The CCAFS-SLC launch site exhibited more successful launches at higher payloads > 12300 kg compared to other sites.

Success Rate vs. Orbit Type

- Orbits ES-L1, SSO, HEO and GEO exhibited a 100% successful launch rate.

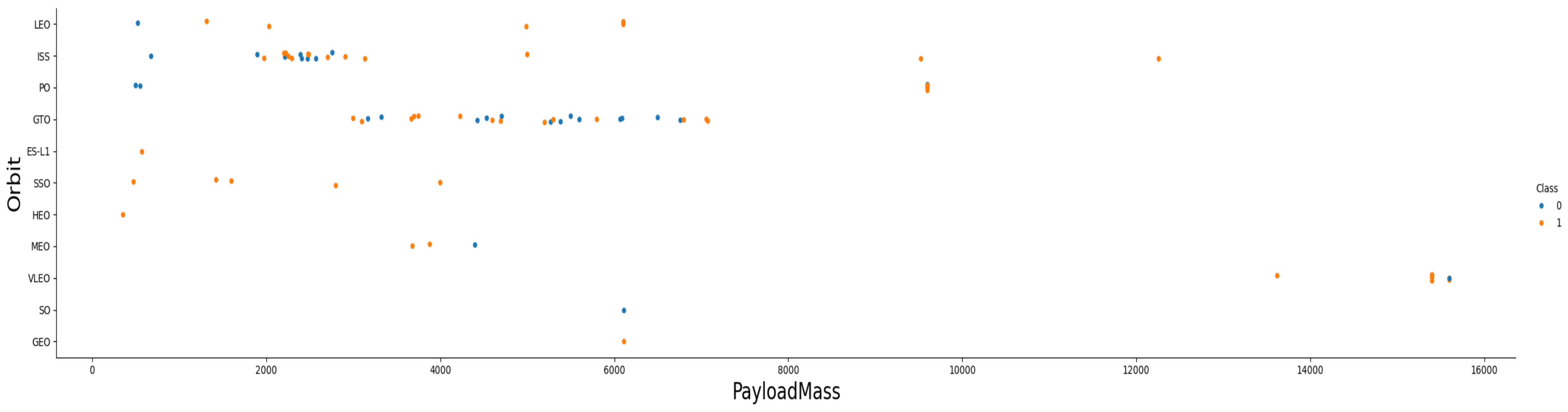


Flight Number vs. Orbit Type



- The LEO orbit, success seems to be related to the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

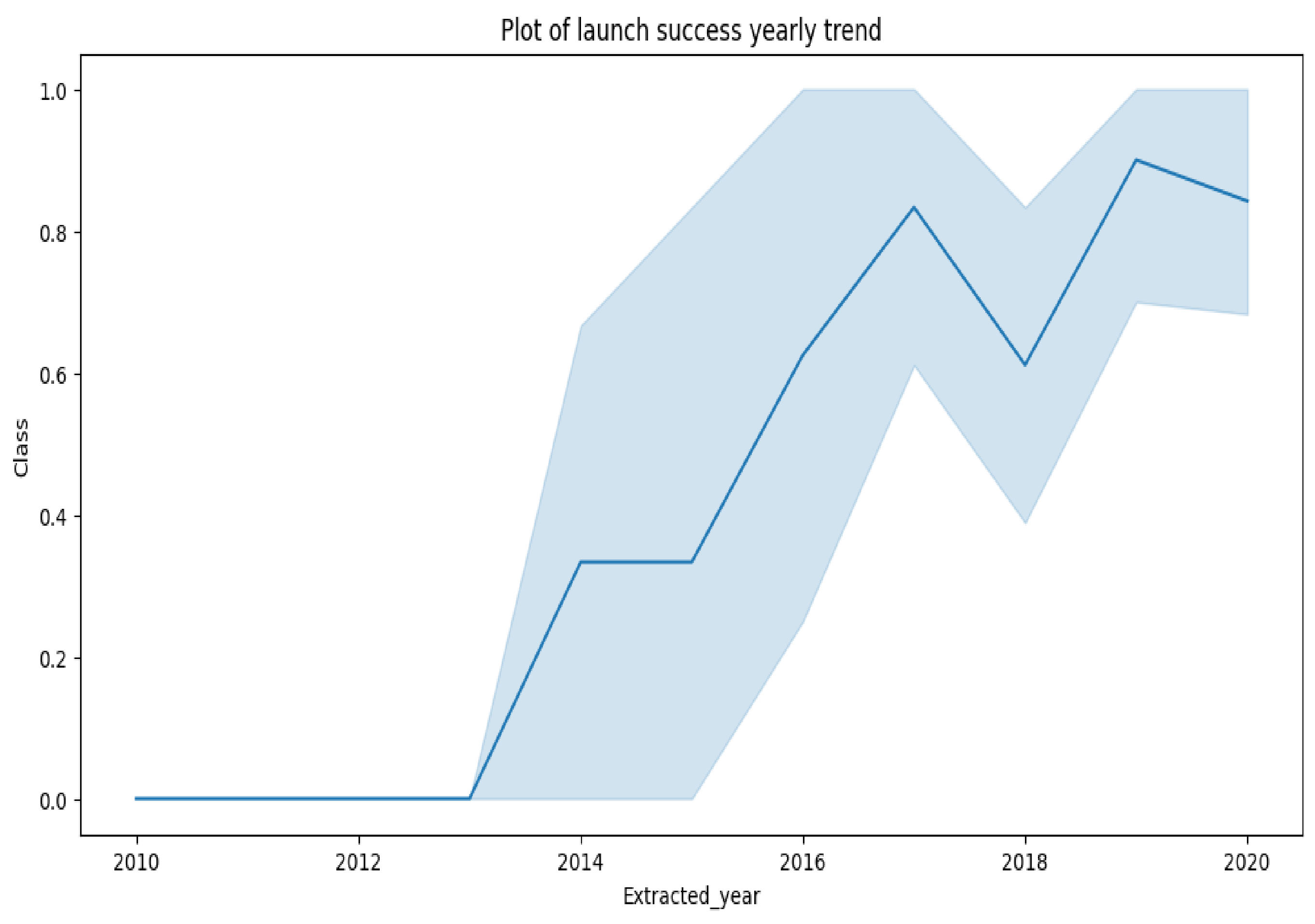
Payload vs. Orbit Type



- The successful landing or positive landing rate are more for Polar, LEO, and ISS considering heavy payloads.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- We can observe that the sucess rate since 2013 kept increasing till 2020.
- In addition, a steep increase is noted between 2013 and 2014.



All Launch Site Names

- The names of the unique launch sites are listed in the table below.
- Only 4 launch sites were considered in this project.

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

Launch Site Names Begin with 'CCA'

- The first 5 launch site names beginning with ‘CCA’ are listed below.

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

- The total payload carried by boosters from NASA was 45 596 kg.

Total_Payload
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2928.4 kg.

Average_Payload
2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was 22 December 2015.

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are listed below.

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are shown below.

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are as follows:

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

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are only 2 as shown below.

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

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

- The table below ranks 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.

Landing_Outcome	Total Count
Success (ground pad)	3
Failure (drone ship)	5

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

- The launch sites were located to the west and east of the USA.

Success/Failed Launches for Each Site

- Successful launches are indicated in green, whereas failed launches are indicated in red.

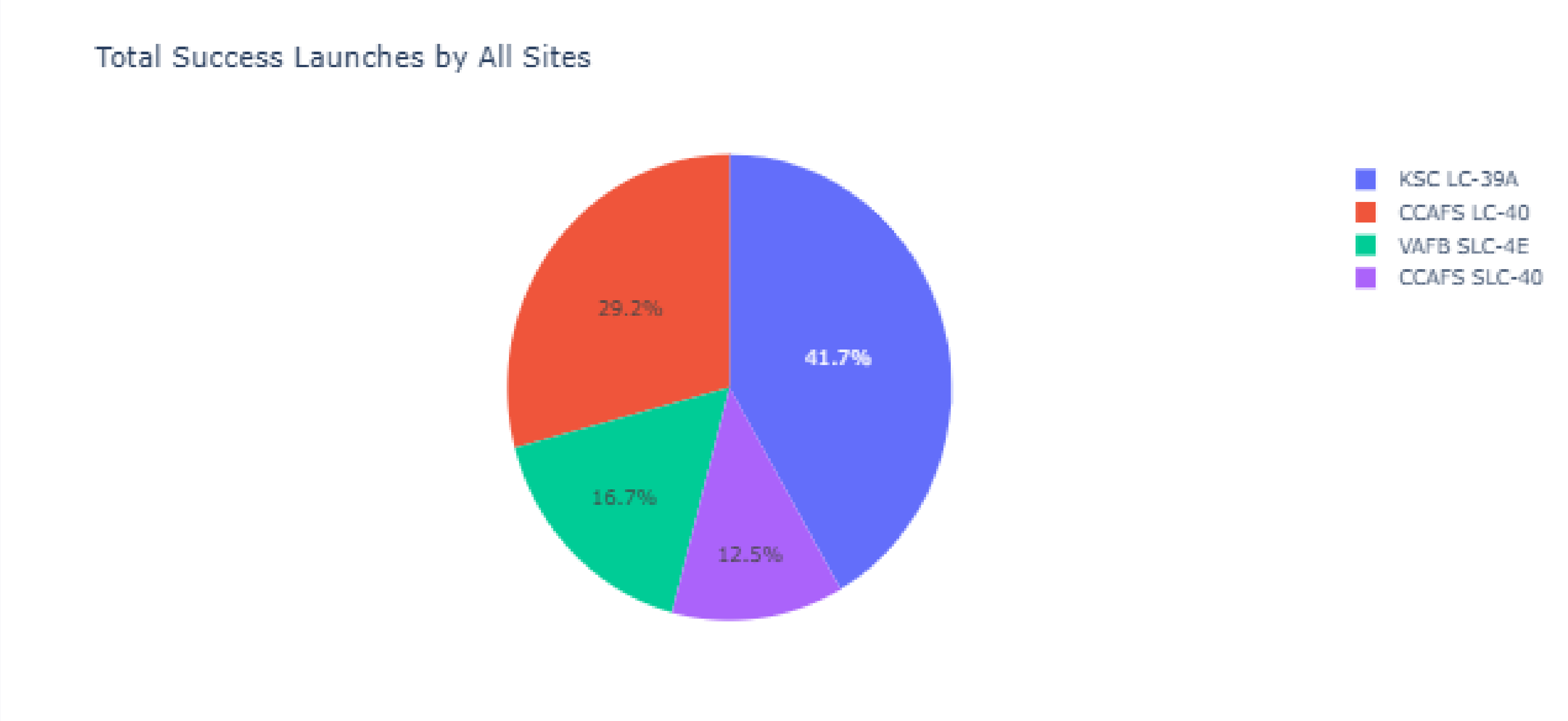
Map with Proximities

- Landmarks shown on the map include railway and airforce base.

Section 4

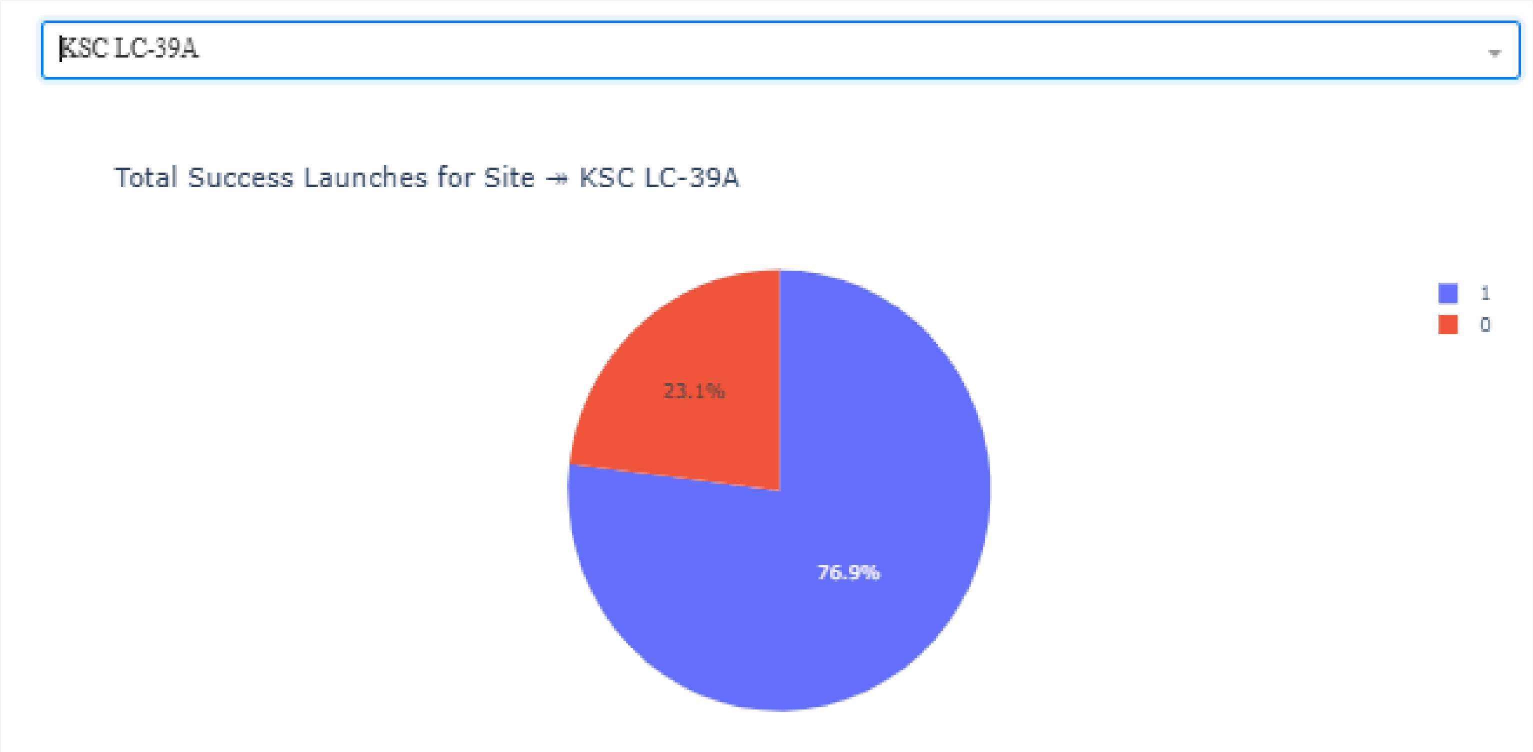
Build a Dashboard with Plotly Dash

Total Success Launches by All Sites



- KSC LC-39A had the highest percentage of successful launches at 41.7% followed by CCAFS LC-40 at 29.2%.

Launch Site with Highest Launch Success Ratio



- KSC LC-39A had a launch success ratio of 76.9% to 23.1%.

Payload vs. Launch Outcome

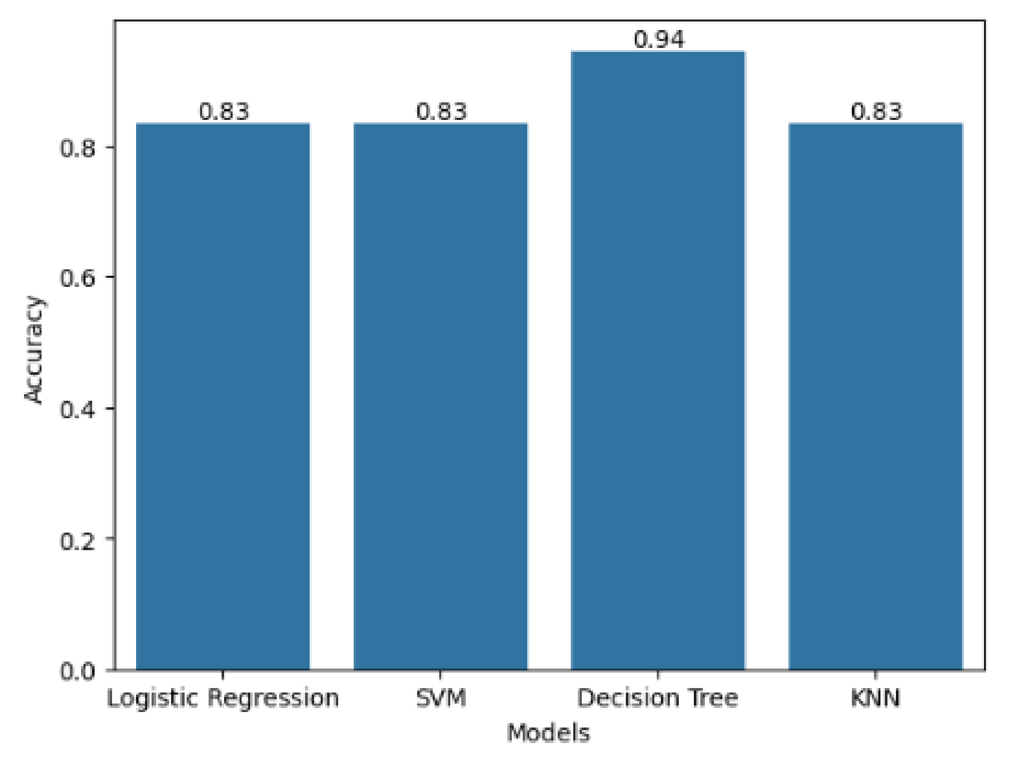
- We can see that as the payload increases, the number of sites with a correlation between the payload and success launch reduces.

Section 5

Predictive Analysis (Classification)

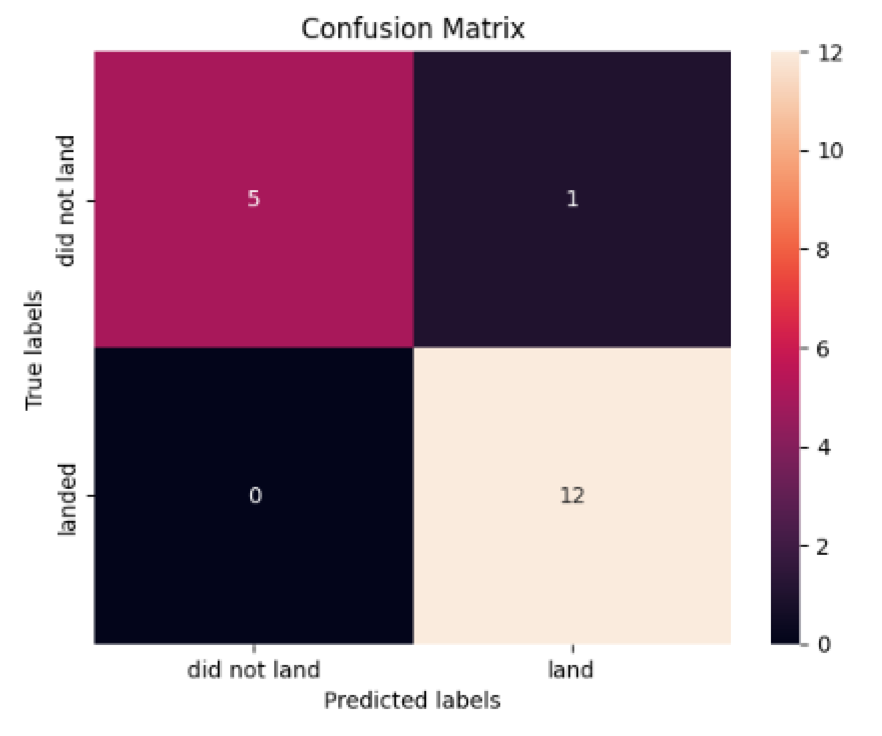
Classification Accuracy

- The Decision Tree model exhibited the highest accuracy at 94%.



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- For orbit types ES-L1, SSO, HEO, and GEO, there is a 100% success rate in landing there.
- The decision tree algorithm is the preferred algorithm to predict successful launches.
- The higher the number of flights at a launch site, the higher the successful launches.
- The CCAFS-SLC launch site exhibited more successful launches at higher payloads > 12300 kg compared to other sites.

Appendix

- The Python code snippets, SQL queries, charts, Notebook outputs, or data sets that were used in this project are available on Github.

Thank you!