SpaceX Rocket Launch Analysis

Andrea Wenman

December 19, 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

Executive Summary

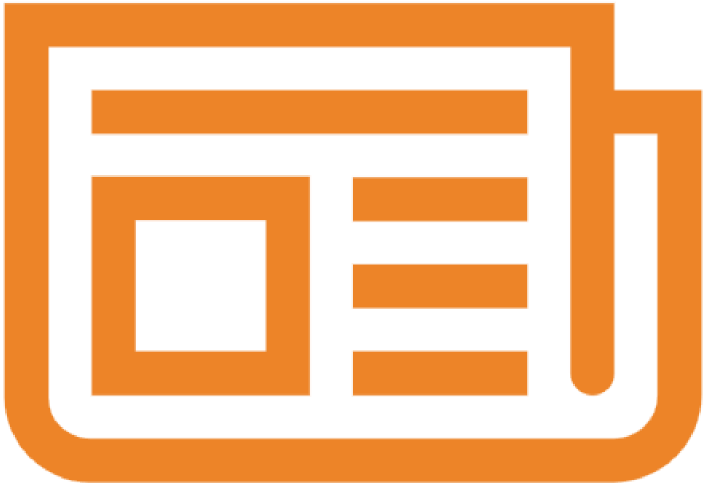
- Summary of Methodologies
  - This project follows the following steps:
    - Data collection
    - Data Wrangling
    - Exploratory Data Analysis
    - Interactive Visual Analysis
    - Predictive Analytics (Classification)
- Summary of Results:
  - Exploratory Data Analysis Results
  - Geospatial Analytics
  - Interactive Dashboard
  - Predictive Analysis of Classification Models

Introduction

- Many companies are working to reduce the cost of space travel, making it more affordable for all. SpaceY is assessing competitor SpaceX's launch data to assess the price of rocket launches.
- SpaceX is a leader in the industry, standing out with their cost-effective approach to rocket launches by reusing the Falcon 9's first stage and working to lower launch costs.
- To determine the launch price for SpaceY, we need to:
  - Analyze the launch data for SpaceX
  - Predict the reusability of Falcon 9's first stage
- We will do this by:
  - Gathering publicly available SpaceX data
  - Utilizing exploratory analysis, data visualization, and ML techniques

Methodology

- Data Collection from the following sources:
  - SpaceX REST API
  - Webscraping records from a HTML table on Wikipedia
- SpaceX Rest API included detailed data about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - https://api.spacexdata.com/v4/
- Wikipedia table included valuable data on Falcon 9 launch records
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Some data cleanup needed:
  - Removing Falcon 1 launch data
  - Clean up NULL values
    - Replace nulls with the mean for PayloadMass data
    - LandingPad will keep NULL values.
  - Convert outcome to a 0 or 1 value
    - 0 is a failure, 1 is a success



Data Collection - API

- Process of collecting API Data
  - Request data from API https://api.spacexdata.com/v4/
  - Decode response text using .json() and turn it into a data frame for easier use and manipulation
  - Request information about launches from the API using custom functions
  - Creating a dictionary from the data, then creating a dataframe from the dictionary
  - Cleaning the data to remove data from launches that were not Falcon 9 launches
  - Replacing missing values of Payload Mass with the mean Payload Mass for the dataset
  - Export into CSV for later use

Github: DataCollectionAndCleaning

Data Collection - Web Scraping

- Data obtained from SpaceX API:
  - Flight Number, Launch Site, Payload, PayloadMass, Orbit, Customer, Launch Outcome, Version Booster, Booster Landing, Date, Time
- Process of collecting API Data
  - Request data from Wikipedia : https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Create a BeautifulSoup object from the HTML response
  - Extract all column names from the HTML table header
  - Collect the data by parsing HTML tables
  - Create a dictionary with the data and create a data frame from the dictionary
  - Export into a CSV
- Github: Data_Collection_WebScraping

Data Manipulation & Data Wrangling

- The SpaceX data contains data for several launch sites, as specified in the LaunchSite column
  - Using value_counts(), we find the number of launches at each site.
- Each launch aims to a dedicated orbit. The orbit type is found in the Orbit column.
  - Using value_counts(), we find the number of launches to each orbit

Data Manipulation & Data Wrangling

Each launch has a mission outcome as described in the Outcome column. The Outcome column has two pieces. The first piece is a BOOLEAN selection, where True indicates a successful landing and False indicates an unsuccessful landing. The second piece indicates where the landing should take place, for instance Ocean for a region in the ocean or RTLS for a ground pad.

To determine whether a booster will successfully land, we converted this column to a binary column – 1 or 0. Here a 1 indicates a successful landing and 0 indicates a failure. This is done by:

- Define the set of bad outcomes: bad_outcomes
- Create a list landing_class, where the element is 0 if the corresponding row in Outcome is the set bad_outcome, otherwise it is 1.
- Create a Class column that contains the values from the landing_class

```
df['Class']=landing_class
df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

Exploratory Data Analysis - Visualization

Scatter Charts

This data visualization was used to show the relationships between two different variables, such as:

- Flight Number and Launch Site
- Orbit Type and Flight Number
- Payload and Orbit Type

Bar Chart

Bar charts are great for comparing the values of different categories or groups to view trends and variations. We used this for showing the relationship between:

- Success Rate and Orbit Type

Line Chart

These visualizations are excellent for displaying trends, patterns, or changes in data over a continuous interval. We used these to see patterns in:

- Success Rate and Year

Exploratory Data Analysis - SQL

- We utilized SQL to further analyze the dataset. We explore the following:
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v 1.1
  - Find the date for the first successful landing in ground pad
  - List the names of boosters which have success in drone ship and have a payload mass greater than 4000
  - Find the total number of successful and failure mission outcomes
  - Find the names of booster versions which have carried the maximum payload mass

Appendix B

Interactive Dashboard - Plotly

- An interactive dashboard was built with Plotly to provide a hands-on visualization of the data. It includes:
- Pie Chart showing the total successful launches per site
  - Allows for easy identification of the most successful sites
  - Chart can be filtered to see the success/failure ratio for an individual site
- Scatter Graph showing the correlation between outcome and payload mass (kg)
  - Can be filtered using the slider to view ranges of payload masses
  - Can be filtered by booster version

Results

- Exploratory Data Analysis
- Interactive Analytics Dashboard
- Predictive Analytics Results

Predictive Analytics - Classification

Model Development

- To prepare the dataset for model development
  - Load dataset
  - Perform data transformation
  - Split data into training and testing sets
  - Decide which type of ML algorithms were appropriate
- For each algorithm:
  - Create a GridSearchCV object and a dictionary of parameters
  - Fit the object to the parameters
  - Use the training data set to train the mode.

Model Evaluation

- For each algorithm:
  - Using the output GridSearchCV object to check the tuned hyper paramters and the accuracy.
  - Plot and examine the confusion matrix.

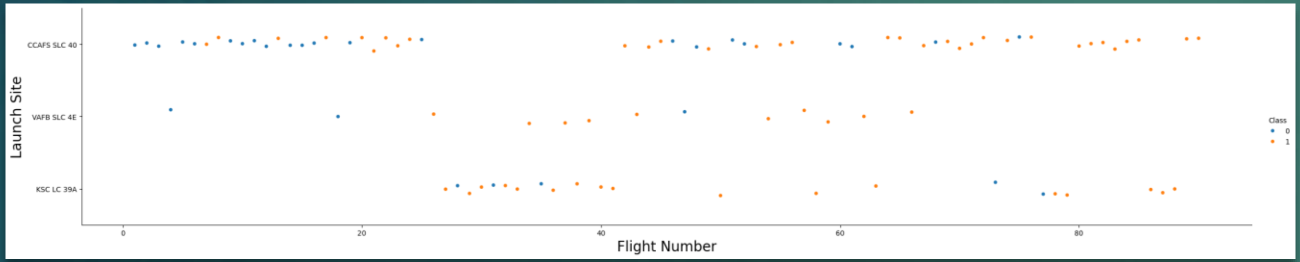
Finding the Best Model

- Review the accuracy scores for all algorithms
- The model with the highest accuracy score is determined as the best performing model.

Exploratory Data Analysis with Data Visualization

GITHUB: DASHBOARD_BUILDING

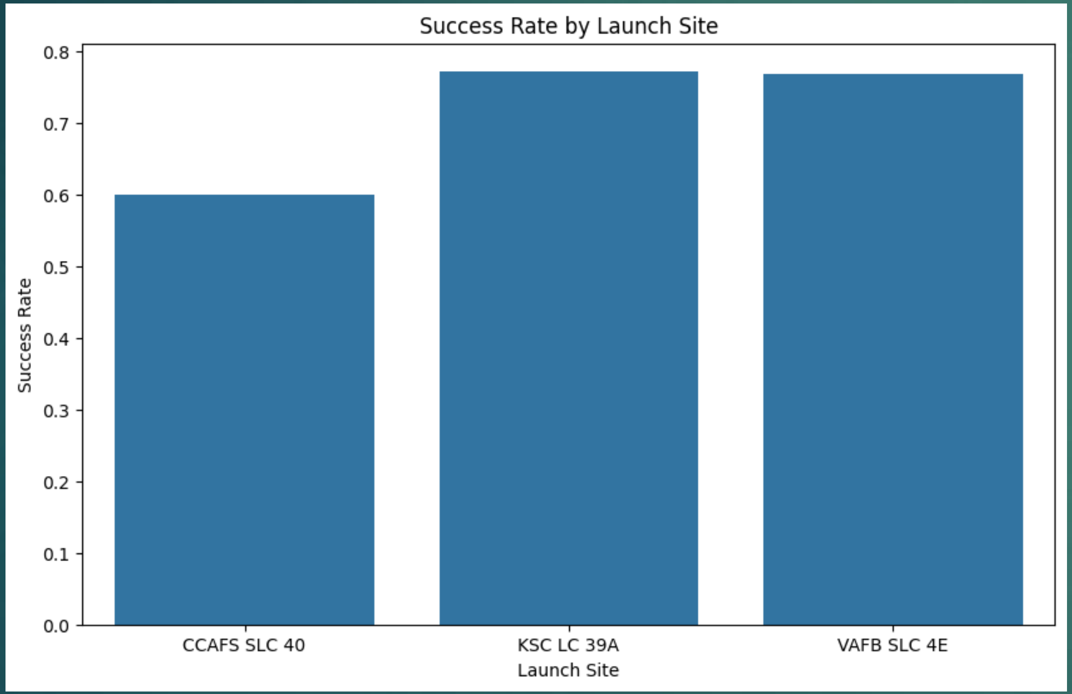
Exploratory Data Analysis – Launch Site by Flight Number



Explanation and Analysis

- As number of flights increase, the success rate of that launch site increases
- Most of the early flights launched out of CCAFS SLC 40 and were generally not successful.
- No early flights were launched from KSC LC 39A, this site appears to have a higher success rate but this is partially because there were no early flights from that site.

Exploratory Data Analysis: Success Rate by Launch Site



Explanation and Analysis:

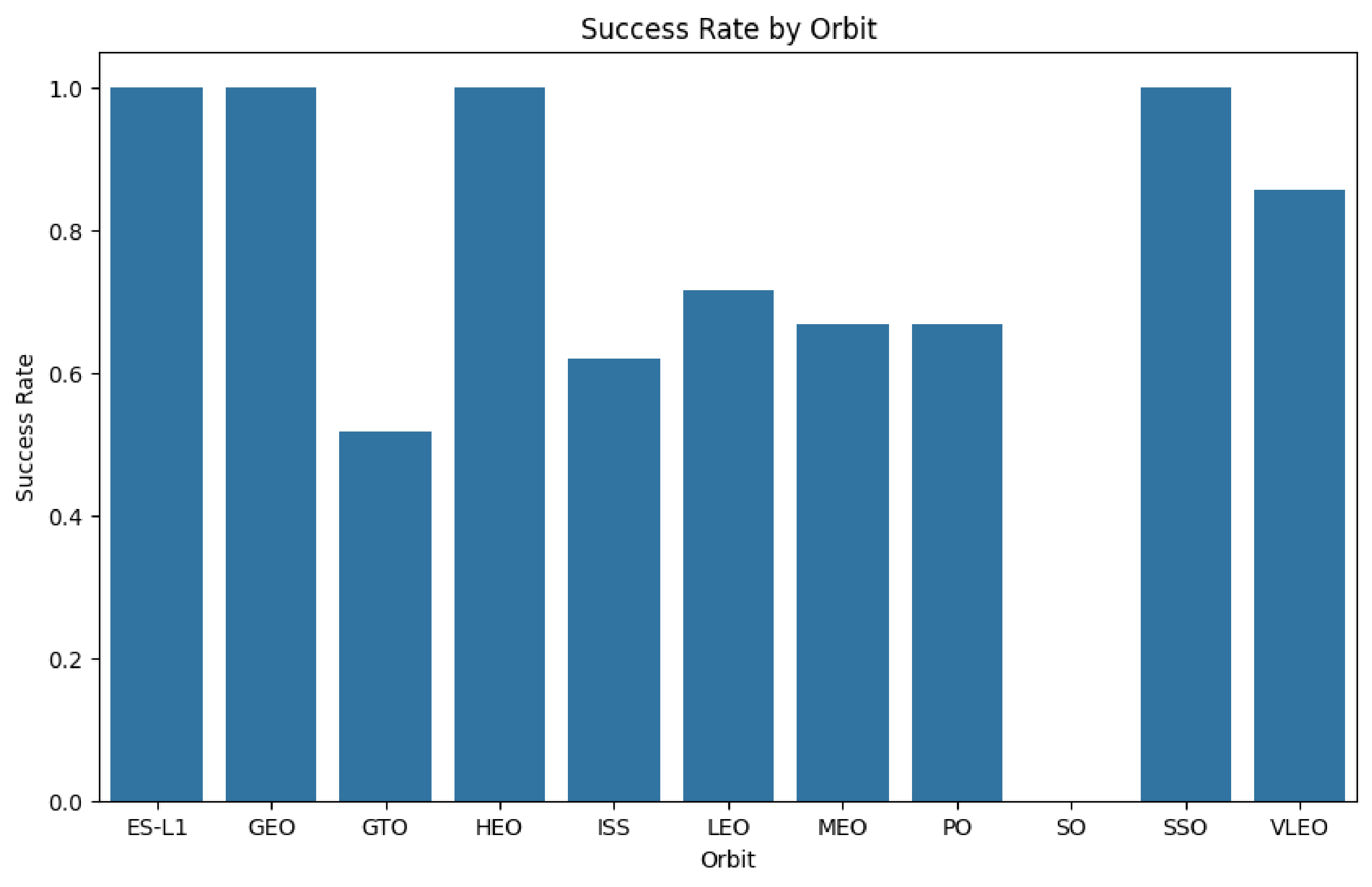
- Site with the lowest success rate:
  - CCAFS SLC 40 – 60%
- Sites with the best success rate:
  - KSC LC 39A - 77%
  - VAFB SLC 4E – 77%
- Implication: Further analysis into what differentiates the CCAFS SLC 40 launch site from the other two could provide some insight into what makes a launch site more likely to have a successful launch.

Exploratory Data Analysis: Success Rate by Orbit

Explanation and Analysis:

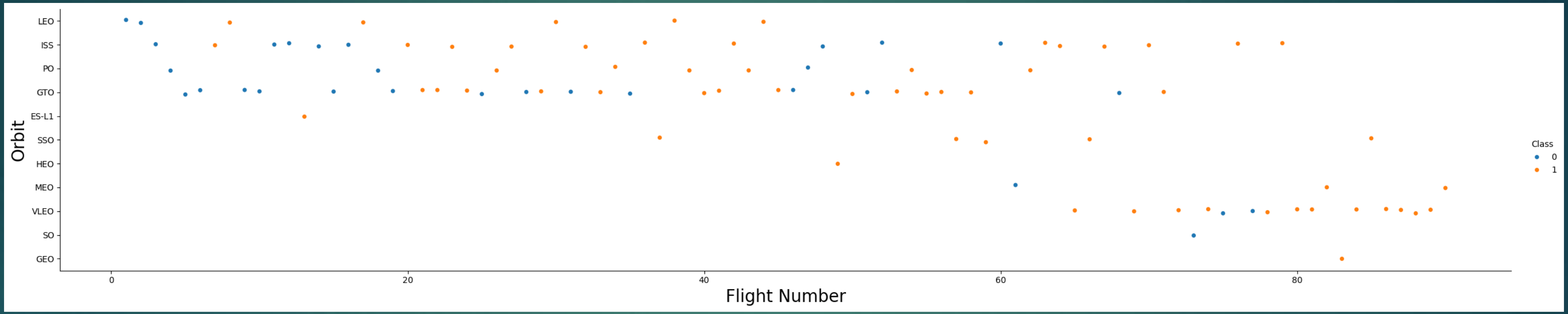
- Orbits with 100% Success:
  - ES – L1
  - GEO
  - HEO
  - SSO

- Orbits with 0% Success Rate:
  - SO



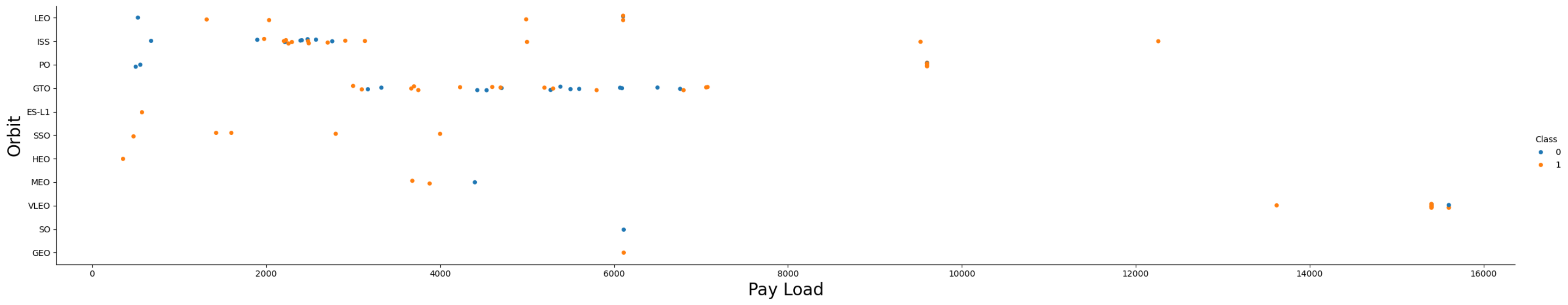
Appendix A

Exploratory Data Analysis: Flight Number Success Rate by Orbit



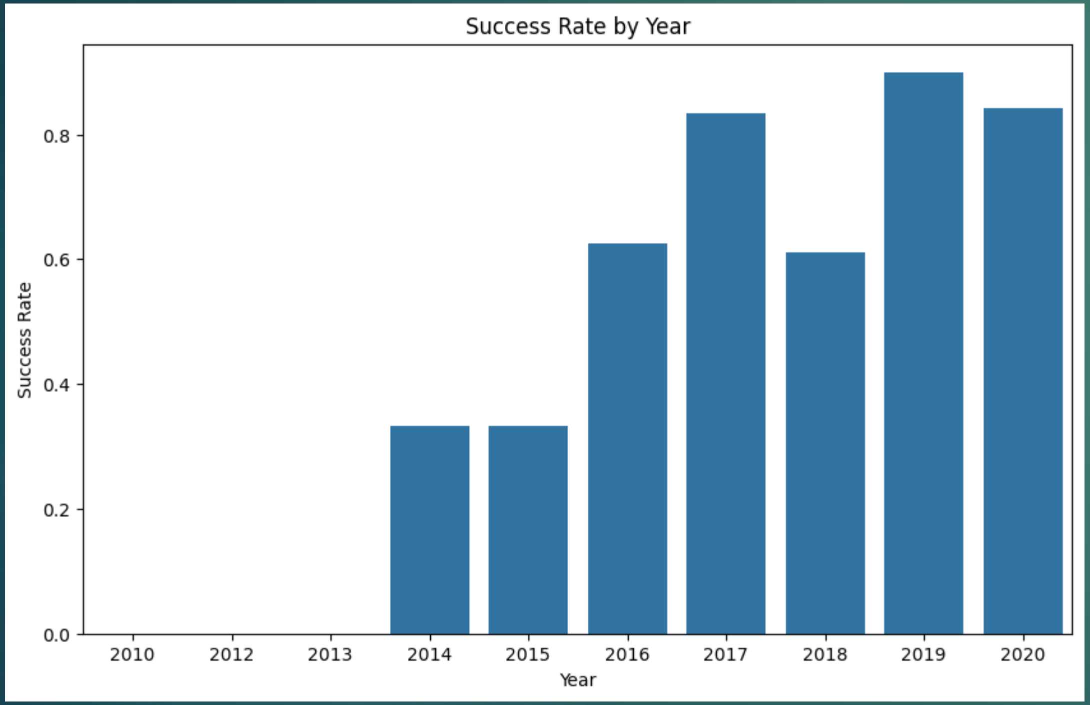
- Orbits such as GTO and IS did not appear to have a relationship between flight number and success rate, having failures and successes spread throughout. Meanwhile, orbits such as LEO had better success as the flight number increased.
- SO had a 0% success rate, but only had one flight.
- GEO, ES-L1, and HEO had a 100% success rate, but didn't have a large quantity of flights to base the success rate off of.
- In general, the later flights are more successful. We see this is true for this visualization, as well.

Exploratory Data Analysis: Payload Success Rate by Orbit



- We see only VLEO orbit had a payload over 14000 and had both successful and unsuccessful launches.
- Most of the orbits had a pay load of 8000 or less
- The Polar, LEO and ISS orbit types have more success with heavy payloads.
- For GTO, the relationship between payloadmass and success rate is not clear.

Exploratory Data Analysis: Success Rate by Year



- Here we see launch data starts at 2014 and goes until 2020
- Success rate trends upwards, generally, with a few dips in 2018 and 2020.
- It could be worth investigating 2018 further to determine why that year had such a dramatic dip in success.

Exploratory Data Analysis with SQL

GITHUB: DASHBOARD_BUILDING

Exploratory Data Analysis: SQL:

Average Payload Mass by F9 v1.1

- Here we calculated the average payload mass carried by booster version F9 v1.1

Result: 2,928.40

Date of First Successful Landing

- Here we determined the date when the first successful landing outcome in ground pad was achieved.

Result: 12/22/2015

Exploratory Data Analysis: SQL:

Drone Ship Successful Boosters with Payload Mass Between 4,000 and 6,000

- We listed the names of the boosters that have has success in drone ship and have a payload mass between 4,000 and 6,000

Result:

- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

Boosters Which Have Carried Maximum Payload Mass

- Here we gathered a list of the booster versions that have carried the maximum payload mass

Result in Appendix B II

Exploratory Data Analysis: SQL:

Ranking the Count of Landing Outcomes Between June 2010 and March 2017

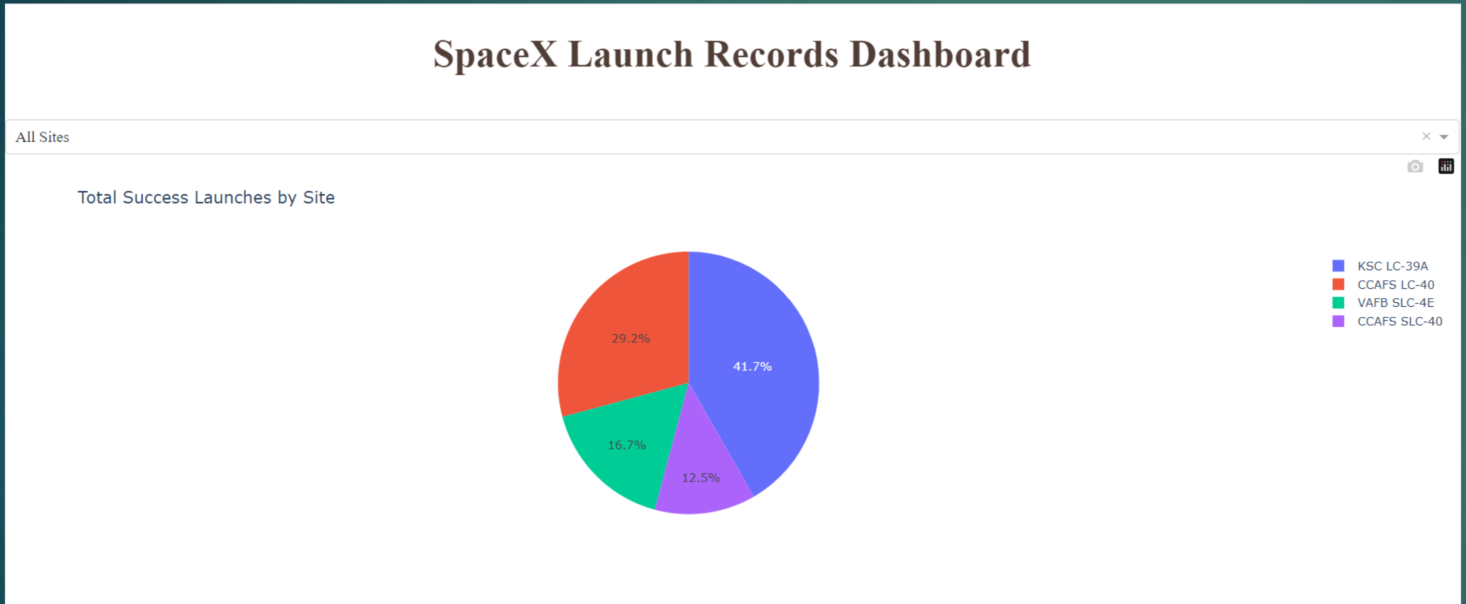
- Next we ranked the count of the landing outcomes between the dates June 4, 2010 and March 20, 2017 in descending order.

Landing_Outcome	outcome_count
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Dashboard

GITHUB: DASHBOARD_BUILDING

Dashboard Tab 1



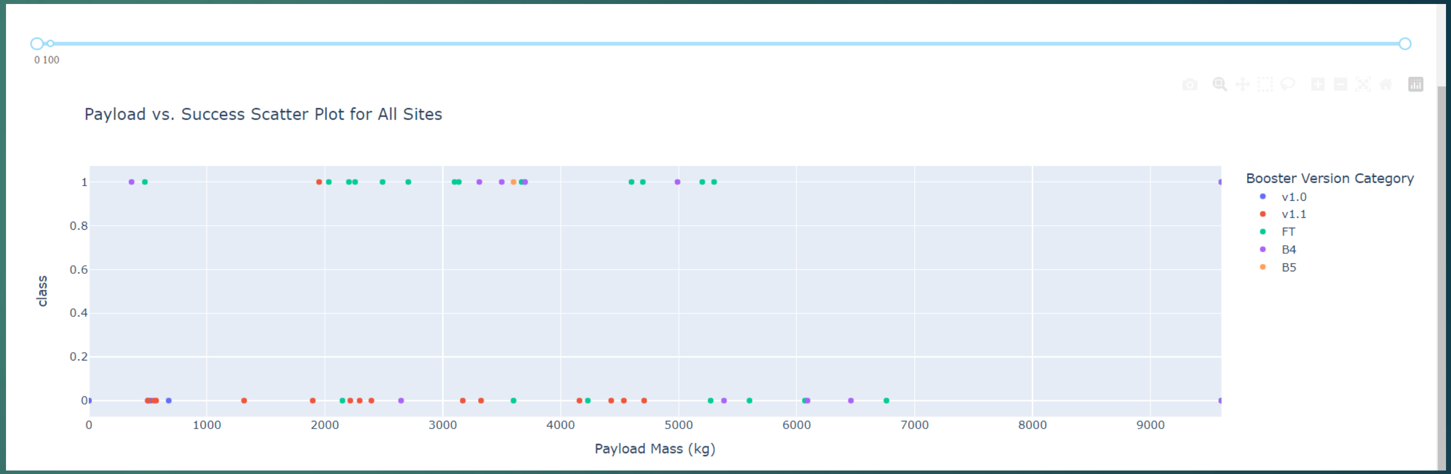
This Page of the interactive dashboard allows you to view a breakdown of the launch success count by launch site. You can also filter by Launch Site.

- You can see the launch site KSC LC-39A has the most successful launches, with 41.7% of the successful launches.

Dashboard Tab 2

Launch Outcome vs. Payload Scatter Plot

- The Next portion of the dashboard displays the launch outcome vs. Payload for all sites. The slider scale filter adjusts for payload range (Kg)
- We can see with this visualization that success for massive payloads is lower than that for low payloads.
- Was can also determine which booster types have been launched with massive payloads.



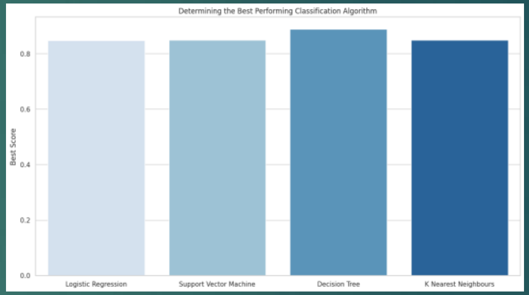
Predictive Analysis

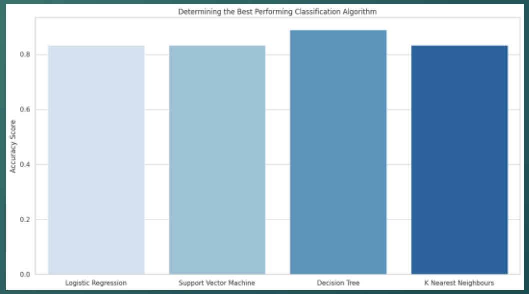
GITHUB: DASHBOARD_BUILDING

Predictive Analytics

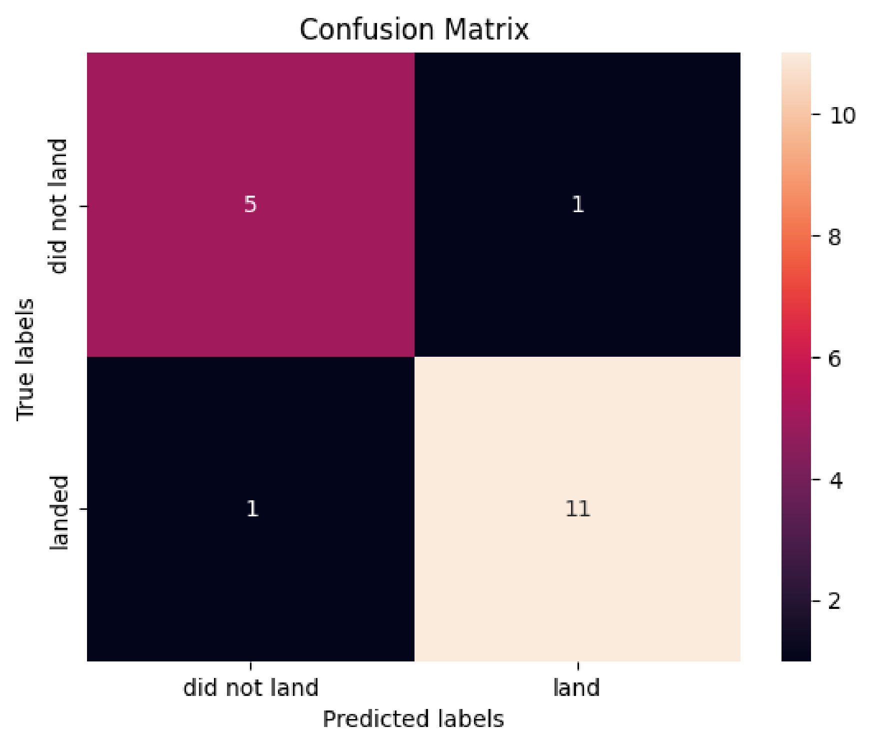
Plotting the accuracy score between the 4 models helps us determine which model was most accurate.

- We see the Decision Tree model was the most accurate with the highest score for both Accuracy Score (88.89%) and the Best Score (84.8%)





Predictive Analytics



Confusion Matrix

- Confusion matrices were used to evaluate the accuracy of the models. These help visualize the number of correct and incorrect predictions made by the model.
- This Matrix is for the Decision Tree model. It had only one false positive and one false negative.

Results & Conclusion

Overall Findings and Conclusions

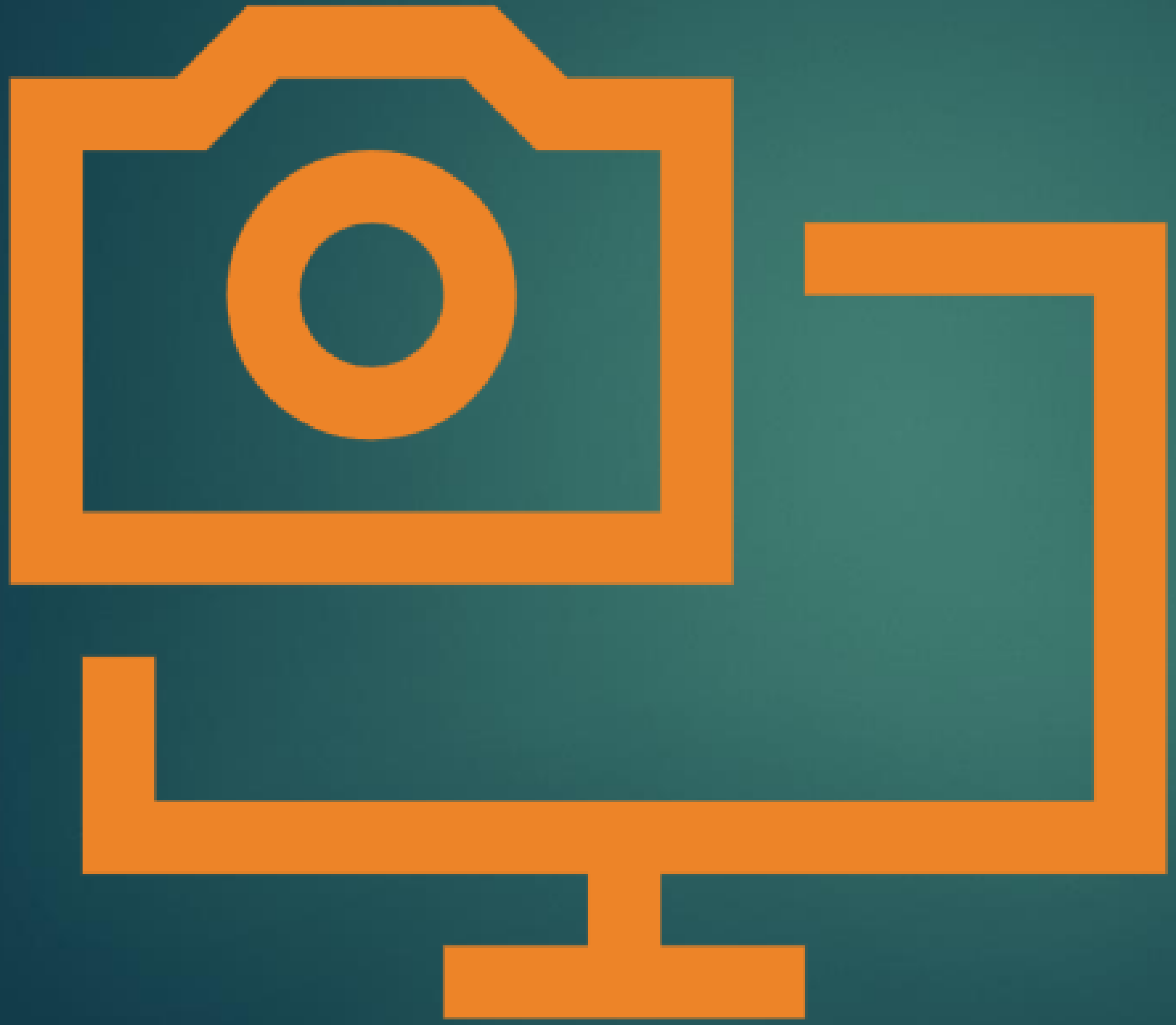
- As flight number increases, success rate increases
  - Earlier flights appear more likely to fail while later flights tend to be more successful
    - Only 3 flights between flights 1 – 19 were successful, only 3 flights between flights 70-90 were unsuccessful
    - After 2015, success rate trended upwards year over year with only a couple of minor dips.

- Certain orbit types tend to be more successful than others
  - Orbits with the highest success rate: ES – L1, GEO, HEO, and SSO
    - However, orbit types ES – LI, GEO, and HEO all had a very small sample size
    - SSO had a higher number of successful flights
  - Orbits PO, ISS, and LEO have more success with heavy payloads
  - Orbits SO and GTO had a success rate of under 60%
    - SO only had 1 flight, so this is based off a small sample size

Overall Findings and Conclusions

- Launch Site CCAFS SLC 40 had the lowest success rate
  - KSC LC 39A and VAFB SLC 4E had success rates of 75%
  - CCAFS SLC 40 had a success rate of 60%

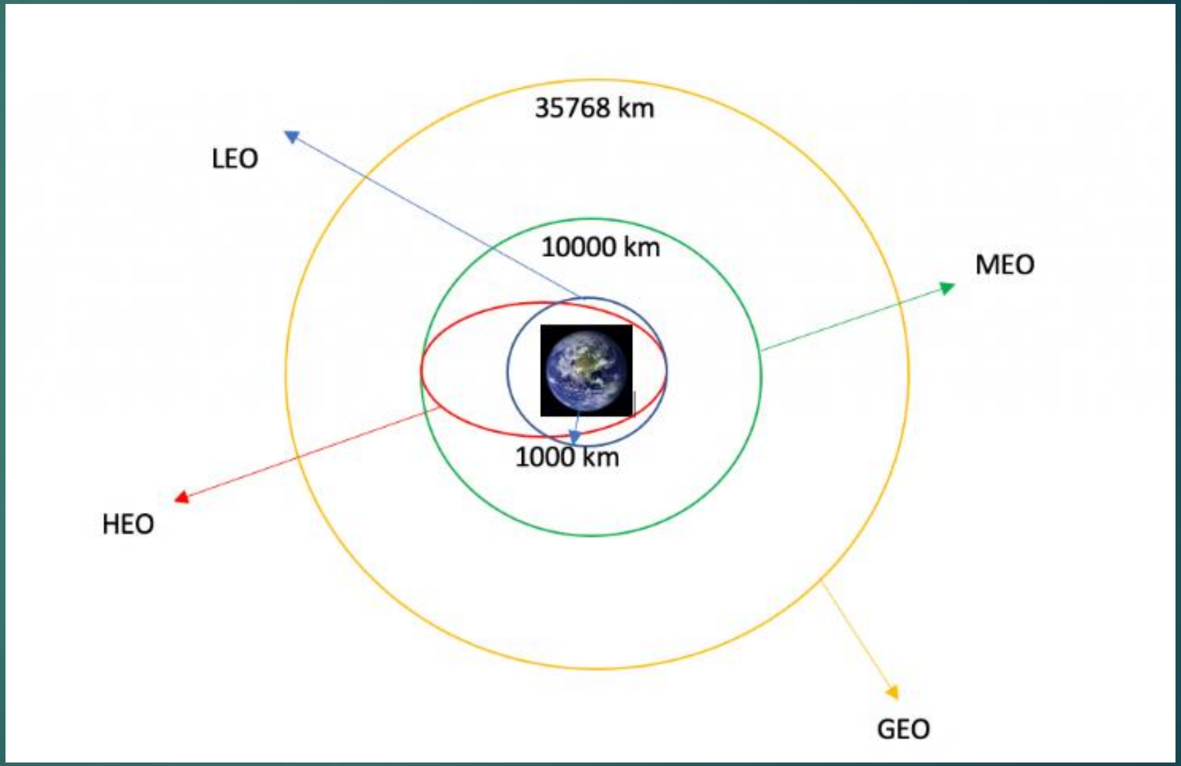
- The decision tree model was the best performing classification model
  - Decision tree accuracy score = 88.89%
  - Decision tree best Score = 84.8%



Appendix

Appendix A

- This visualization displays the different orbits discussed in this analysis



Appendix B I

- Exploratory Data Analysis with SQL

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTABLE WHERE "Customer" = "NASA (CRS)"
```

```
 * sqlite:///my_data1.db
Done.
```

total_payload_mass
45596

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS average_payload_mass FROM SPACEXTABLE WHERE "Booster_Version" = "F9 v1.1"
```

```
 * sqlite:///my_data1.db
Done.
```

average_payload_mass
2928.4

List the date when the first succesful landing outcome in ground pad was acheived.

```
%sql SELECT MIN("Date") AS first_successful_landing FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)"
```

```
 * sqlite:///my_data1.db
Done.
```

first_successful_landing
2015-12-22

Appendix B II

- Exploratory Data Analysis with SQL

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND "PAYLOAD_MASS_
```

```
* sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT("Landing_Outcome") AS Total_Number, "Landing_Outcome" FROM SPACEXTABLE GROUP BY "Landing_Outcome"
```

```
* sqlite:///my_data1.db
Done.
```

Total_Number	Landing_Outcome
5	Controlled (ocean)
3	Failure
5	Failure (drone ship)
2	Failure (parachute)
21	No attempt
1	No attempt
1	Precluded (drone ship)
38	Success
14	Success (drone ship)
9	Success (ground pad)
2	Uncontrolled (ocean)

List the names of the booster_versions which have carried the maximum payload mass.

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLI
```

```
* sqlite:///my_data1.db
Done.
```

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