IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Amy Parrish
11/30/2023

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection via API and Web Scraping
  - Exploratory Data Analysis with Data Visualization
  - EDA with SQL
  - Interactive Map with Folium
  - Dashboards with Plotly Dash
  - Predictive Analysis
- Summary of all results
  - Explorary Data Analysis results
  - Interactive maps and dashboard
  - Predictive results

# Introduction

- Project background and context:
  - The primary goal of this endeavor is to forecast the successful landing of the Falcon 9's initial stage. According to SpaceX's website, the launch cost for the Falcon 9 rocket stands at 62 million dollars, whereas other providers charge over 165 million dollars per launch. This substantial price gap is attributed to SpaceX's capability to recycle the first stage. Evaluating the likelihood of the stage's successful landing enables us to ascertain the launch cost. This data holds significance for any company aspiring to rival SpaceX in the rocket launch domain.
- Problems we will find answers to:
  - What are the main characteristics of a failed or successful launch/landing?
  - What are the conditions that allow for the best possible outcome regarding landing success rate?
  - What are the effects of other variables (rocket etc) on the success or failure of a landing?

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scraping from Wikipedia
- Perform data wrangling
  - Drop unnecessary columns
  - Remove null values
  - Perform One Hot Encoding for normalization of data for use in classification models
- 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

- Describe how data sets were collected.
  - Data sets are collected from the SpaceX REST API and Web Scraping Wikipedia page for Falcon launches
    - Information includes such data points as: payload information, launch site, and orbit

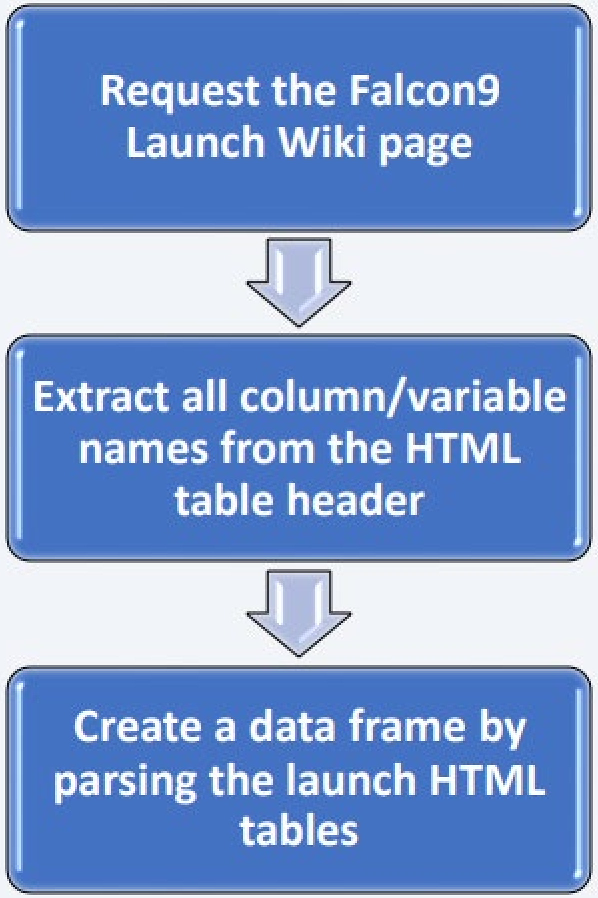
# Data Collection – SpaceX API

- SpaceX provides a publicly accessible API where users can retrieve data for their utilization.


[Link]

# Data Collection - Scraping

- Data regarding SpaceX launches can be obtained from the Falcon 9 Wikipedia page




Link

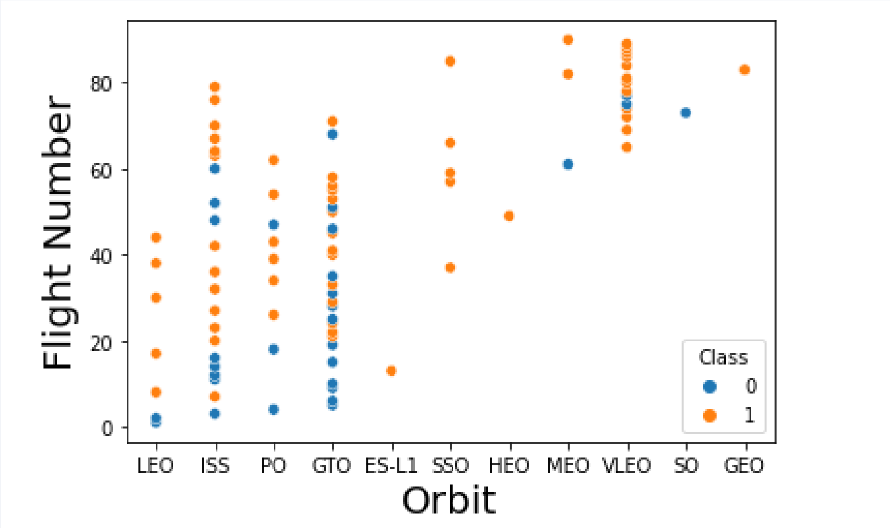
# Data Wrangling

- Exploratory Data Analysis (EDA) was performed on the dataset.
  - The totals of launches at each site, frequency of orbits, and occurrences of mission outcomes for each orbit type were computed.
- Finally, the Landing Outcome Label was created based on the information in the Outcome column.


[Link]

# EDA with Data Visualization

- To analyze the data, scatterplots and barplots were employed to visually represent the correlation between pairs of features, including:
  - Payload and Orbit
  - Launch Site and Flight Number
  - Launch Site and Payload Mass
  - Payload Mass and Flight Number
  - Flight number and Orbit




[Link](#)

# EDA with SQL

- SQL queries were used to gather and understand the dataset's information.
  - Selected the names of the boosters which had success and payload mass greater than 4000 but less than 6000
  - Selected the total number of successful and failure mission outcomes
  - Selected the names of the booster versions that have carried the maximum payload mass
  - Listed the failed landing outcomes, their booster versions and launch site names for 2015
  - Ranked the count of landing outcomes between 6/4/2010 and 3/20/2017 in descending order

Link

# Build an Interactive Map with Folium

- Folium is used to build a leaflet map to show the following:
  - Red circles at coordinates such as NASA Johnson Space Center in Houston, TX
  - Markers to show successful and unsuccessful landings (Green is successful and Red is unsuccessful)
  - Markers to show distance between launch sites and key locations such as highways and cities, plot lines were added between the points of interest.


[Link]

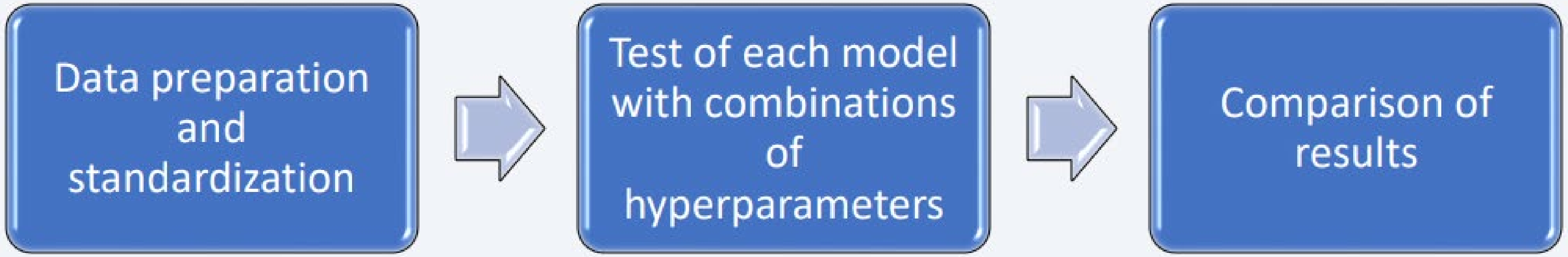
# Build a Dashboard with Plotly Dash

- The following graphs were built
    - Pie Chart showing total successful launches by site
    - Range Slider chart showing payload mass and range
    - Pie Chart showing total launches at site CCAFS LC-40

Link

# Predictive Analysis (Classification)

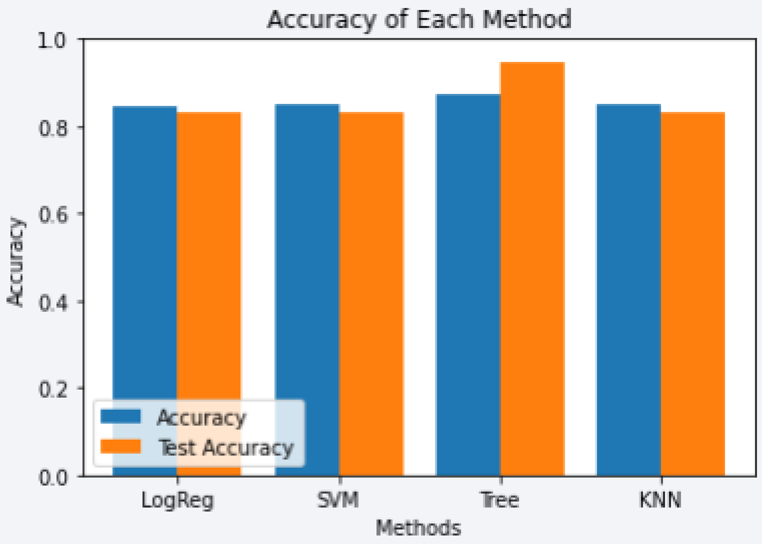
- The following classification models were compared: Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbors




Link

# Results

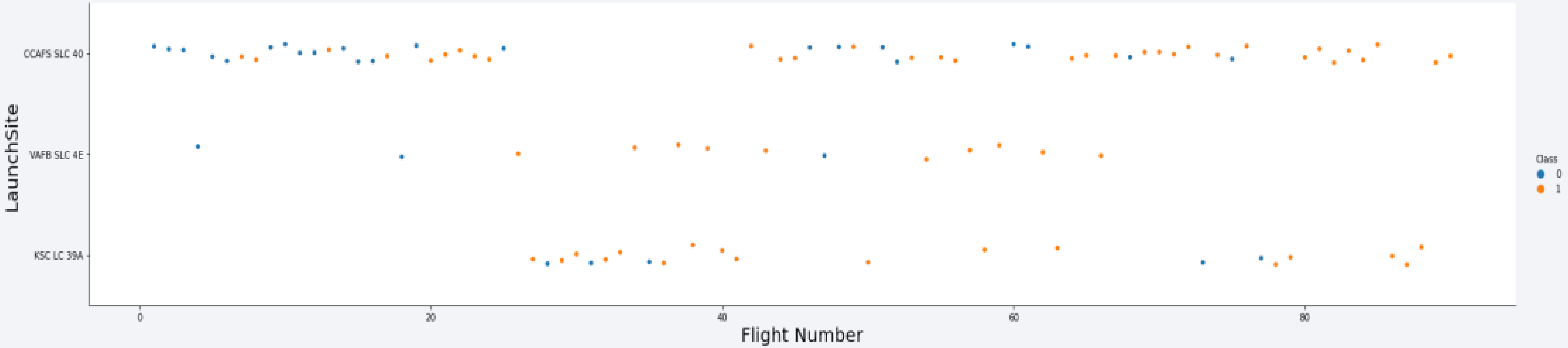
- Space X uses 4 different launch sites.
- Most launches happen on the east coast.
- The first success landing outcome happened in 2013.
- Many Falcon 9 booster versions were successful at landing on drone ships having payload above the average.
- Heavier payloads seem to have the highest success rate.
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.
- The success rate of Landing Outcomes has improved over the years.
- The most accurate model is the Decision Tree Model having accuracy over 87%.



Section 2

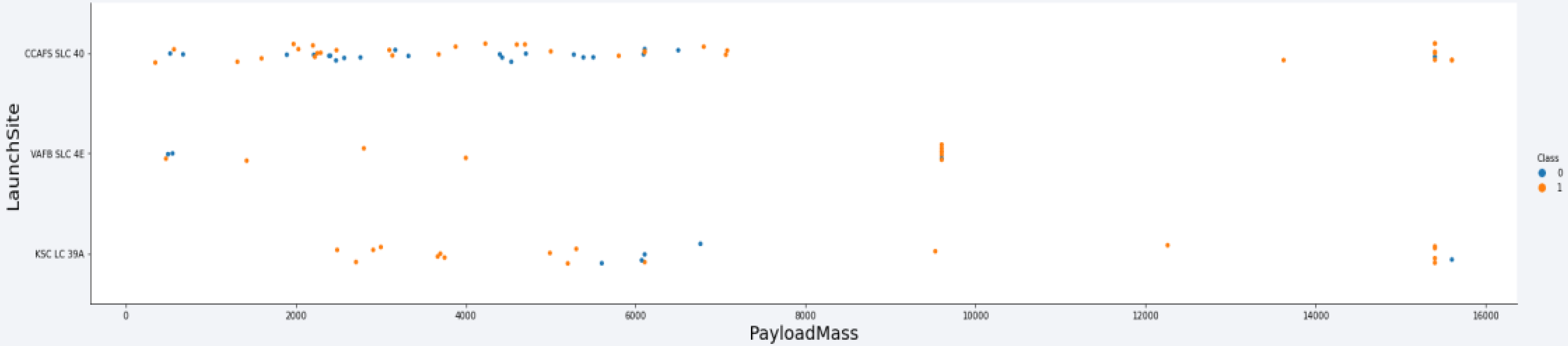
# Insights drawn from EDA

# Flight Number vs. Launch Site



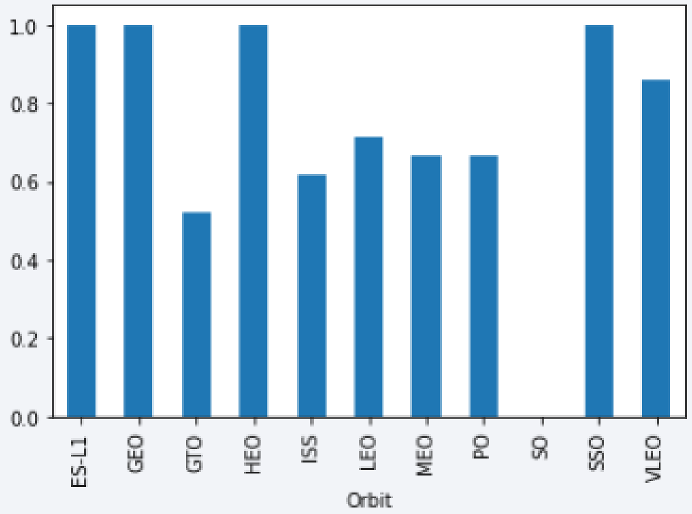

- The most successful launch site is CCAFS SLC 40.
- The plot above also shows that the success rate improved over time.

# Payload vs. Launch Site



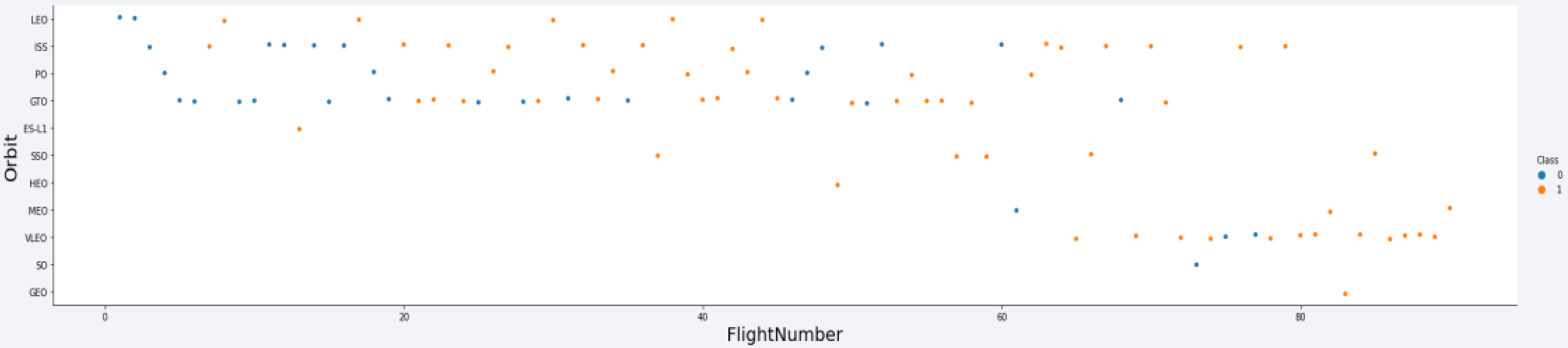

- Payloads over 9000kg have the best success rate, and the success rate goes up as the payloads get heavier.

# Success Rate vs. Orbit Type



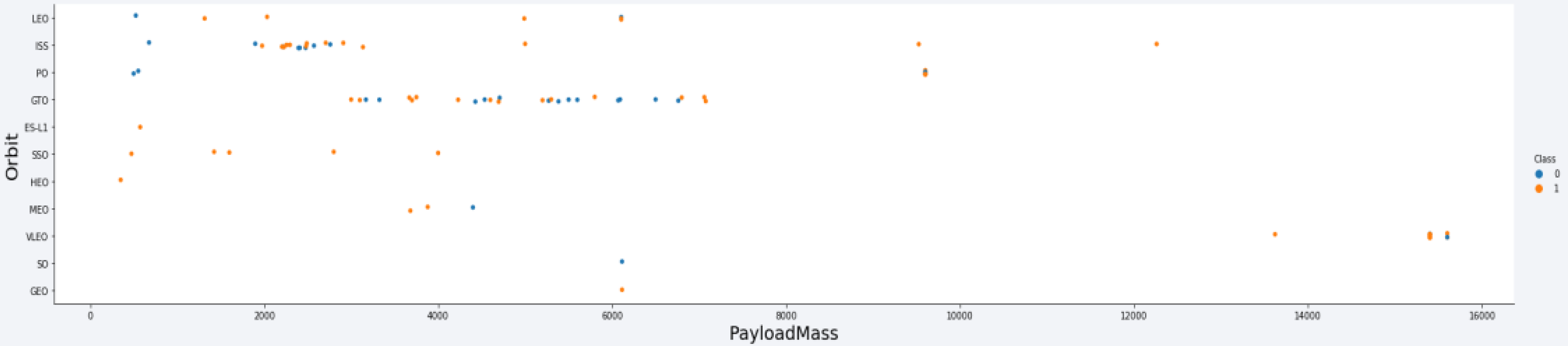

- Orbit ES-L1 (Lagrange point 1) has the best success rate
- GEO, HEO and SSO are also highly successful.

# Flight Number vs. Orbit Type



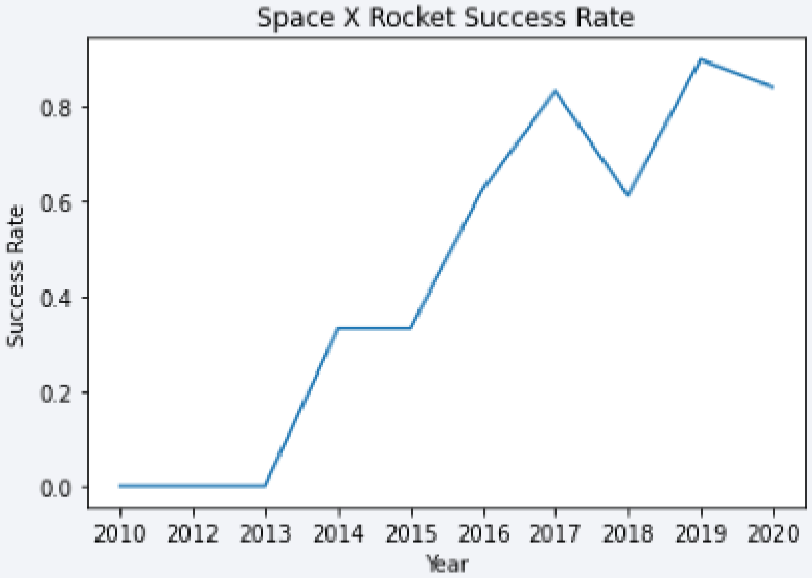

- As the number of flights for the LEO orbit rises, the success rate also increases.
- Increased experience and knowledge has led to an overall increase in success rate for all orbits.

# Payload vs. Orbit Type




- The payload weight may significantly impact the success rate of launches in specific orbits.
- Some orbits have so few launches that there aren't enough data points to compare.

# Launch Success Yearly Trend




- Between 2013 and 2020 there has been a steady increase in success rate.

# All Launch Site Names

- There are 4 launch sites.

| Launch Sites |
|---|
| CCAFS LC-40 |
| CCAFS SLC-40 |
| VAFB SLC-4E |
| KSC LC-39A |

- The launch site names are obtained by using the following query:
  - SELECT DISTINCT “LAUNCH_SITE” FROM SPACEXTBL

# Launch Site Names Begin with 'CCA'

| Date | Time (UTC) | Booster Version | Launch Site | Payload | Payload Mass (Kg) | Orbit | Customer |
|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45 | F9 V1 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX |
| 08-12-2010 | 15:43 | F9 V1 B0004 | CCAFS LC-40 | Dragon Demo flight C1, 2 Cubesats, Barrell of Brouere cheese | 0 | LEO (ISS) | NASA(COTS)NRO |
| 22-05-2012 | 07:44 | F9 V1 B0005 | CCAFS LC-40 | Dragon Demo Flight C2 | 525 | LEO (ISS) | NASA(COTS) |
| 08-10-2012 | 00:35 | F9 V1 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA(CRS) |
| 01-03-2013 | 15:10 | F9 V1 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA(CRS) |

- The results above are obtained with the following SQL query:
  - SELECT * FROM SPACEXTBL WHERE “LAUNCH_SITE” LIKE %CCA% LIMIT 5

# Total Payload Mass

- The total payload carried by boosters from NASA:

| Total Payload |
| --- |
| 45596Kg |

- The information above was obtained using the following SQL query:
  - SELECT SUM(“PAYLOAD_MASS_KG”) FROM SPACEXTBL WHERE “CUSTOMER” = “NASA (CRS)”

# Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1:

| Avg Payload Mass |
| --- |
| 2534.6666666666665Kg |

- The information above was obtained using the following SQL query:
  - SELECT AVG("PAYLOAD_MASS_KG") FROM SPACEXTBL WHERE "BOOSTER_VERSION" LIKE "%F9 V1.1%"

# First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad is:

| Date |
| --- |
| 01/05/2017 |

- The information above was obtained using the following SQL query:
  - SELECT MIN("DATE") FROM SPACEXTBL WHERE "LANDING_OUTCOME" LIKE "%SUCCESS%"

# 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:

| Booster Version |
|---|
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

- The information above was obtained using the following SQL query:
  - SELECT “BOOSTER_VERSION” FROM SPACEXTBL WHERE “LANDING_OUTCOME” = “SUCCESS (DRONE SHIP)” AND “PAYLOAD_MASS_KG” > 4000 AND “PAYLOAD_MASS_KG” < 6000

# Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes:

| Outcome | Count |
|---|---|
| Success | 100 |
| Failure | 1 |

- The information above was obtained using the following SQL query:
  - SELECT (SELECT COUNT(“MISSION_OUTCOME”) FROM SPACEXTBL WHERE “MISSION_OUTCOME” LIKE “%SUCCESS%”) AS SUCCESS (SELECT COUNT(“MISSION_OUTCOME”) FROM SPACEXTBL WHERE “MISSION_OUTCOME” LIKE “%FAILURE%”) AS FAILURE

# Boosters Carried Maximum Payload

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

| 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 |

| Booster Version |
|---|
| F9 B5 B1049.5 |
| F9 B5 B1049.7 |
| F9 B5 B1051.6 |
| F9 B5 B1058.3 |
| F9 B5 B1060.2 |
| F9 B5 B1060.3 |

- The information above was obtained using the following SQL query:
  - SELECT DISTINCT “BOOSTER_VERSION” FROM SPACEXTBL WHERE “PAYLOAD_MASS_KG” = (SELECT MAX(“PAYLOAD_MASS_KG”) FROM SPACEXTBL)

# 2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

| MONTH | BOOSTER VERSION | LAUNCH SITE |
|---|---|---|
| 01 | F9 V1.1 B1012 | CCAFS LC-40 |
| 04 | F9 V1.4 B1015 | CCAFS LC-40 |

- The information above was obtained using the following SQL query:
  - SELECT SUBSTR("DATE", 4, 2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL WHERE "LANDING_OUTCOME" = "FAILURE (DRONE SHIP)" AND SUBSTR("DATE", 7 , 4) = "2015"

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

- 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:

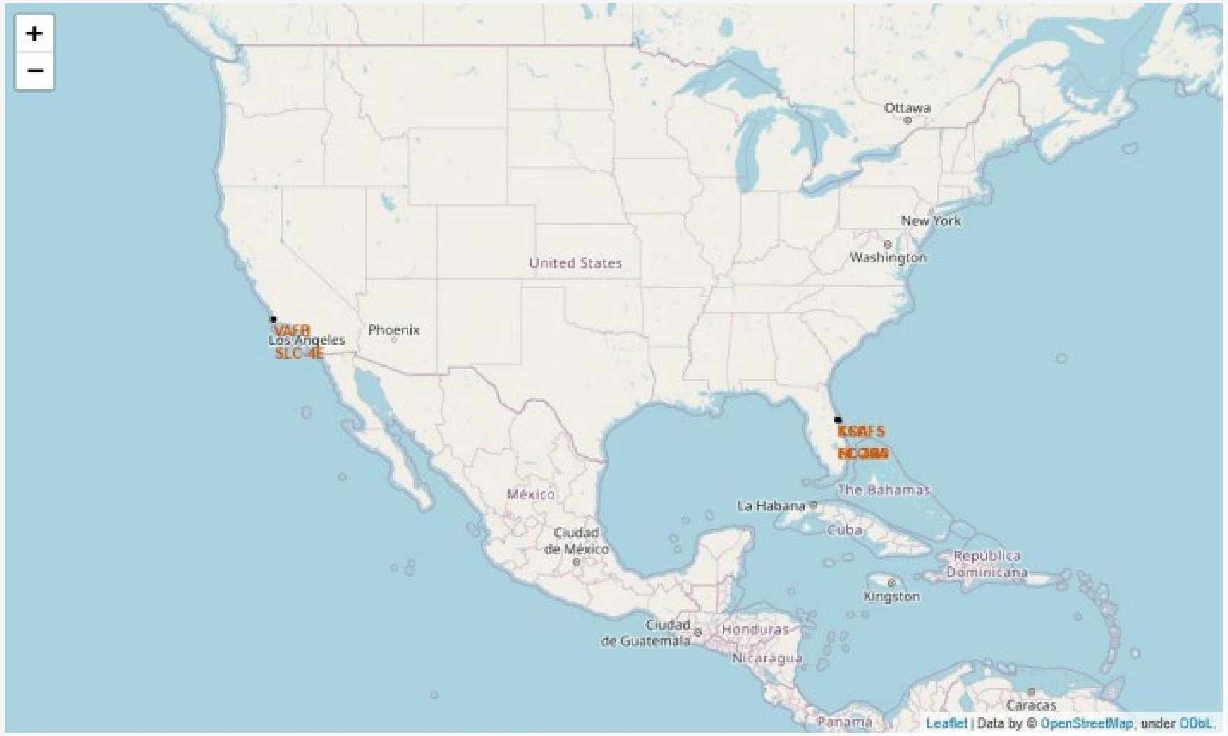
| LANDING OUTCOME | COUNT |
|---|---|
| NO ATTEMPT | 10 |
| SUCCESS (DRONE SHIP) | 5 |
| SUCCESS (GROUND PAD) | 3 |
| CONTROLLED (OCEAN) | 3 |
| UNCONTROLLED (OCEAN) | 2 |
| PRECLUDED (DRONE SHIP) | 1 |
| FAILURE (PARACHUTE) | 2 |
| FAILURE (DRONE SHIP) | 5 |

- The information above was obtained using the following SQL query:
  - SELECT “LANDING_OUTCOME”, COUNT(“LANDING_OUTCOME”) FROM SPACEXTBL WHERE “DATE” >= “04-06-2010” AND DATE <= “20-03-2017” GROUP BY “LANDING_OUTCOME” ORDER BY COUNT(“LANDING_OUTCOME”) DESC;

Section 3

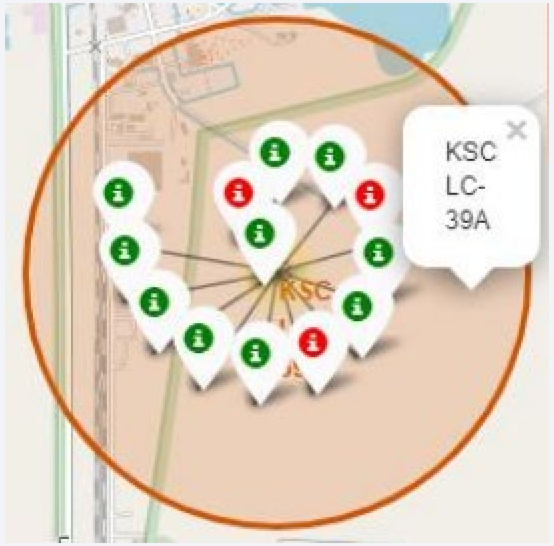
# Launch Sites Proximities Analysis

# All Launch Sites



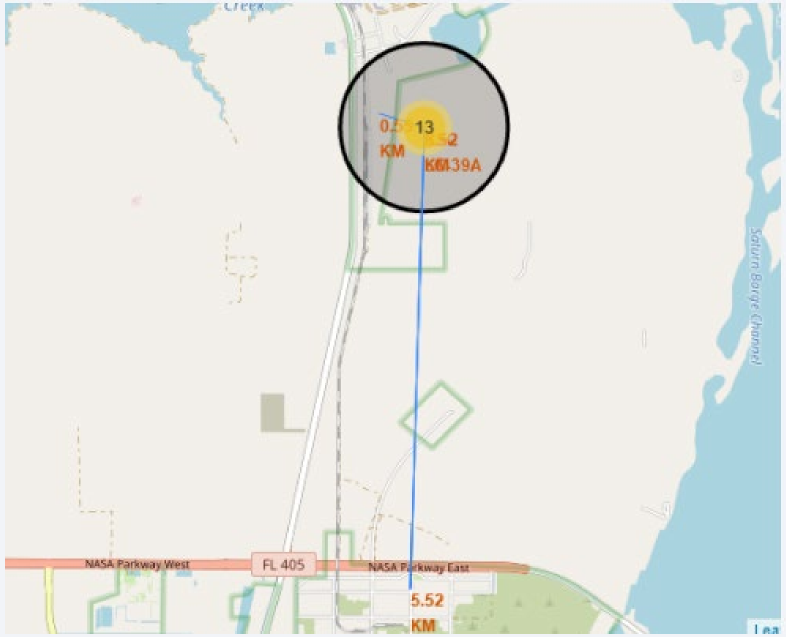

- Launch sites are indicated on the map above.
- Most sites are located along the coasts

# Launch Outcomes




- Green markers indicate successful launches, while Red markers indicate failures.
- The site in the picture above has a high launch success rate.

# Site locations

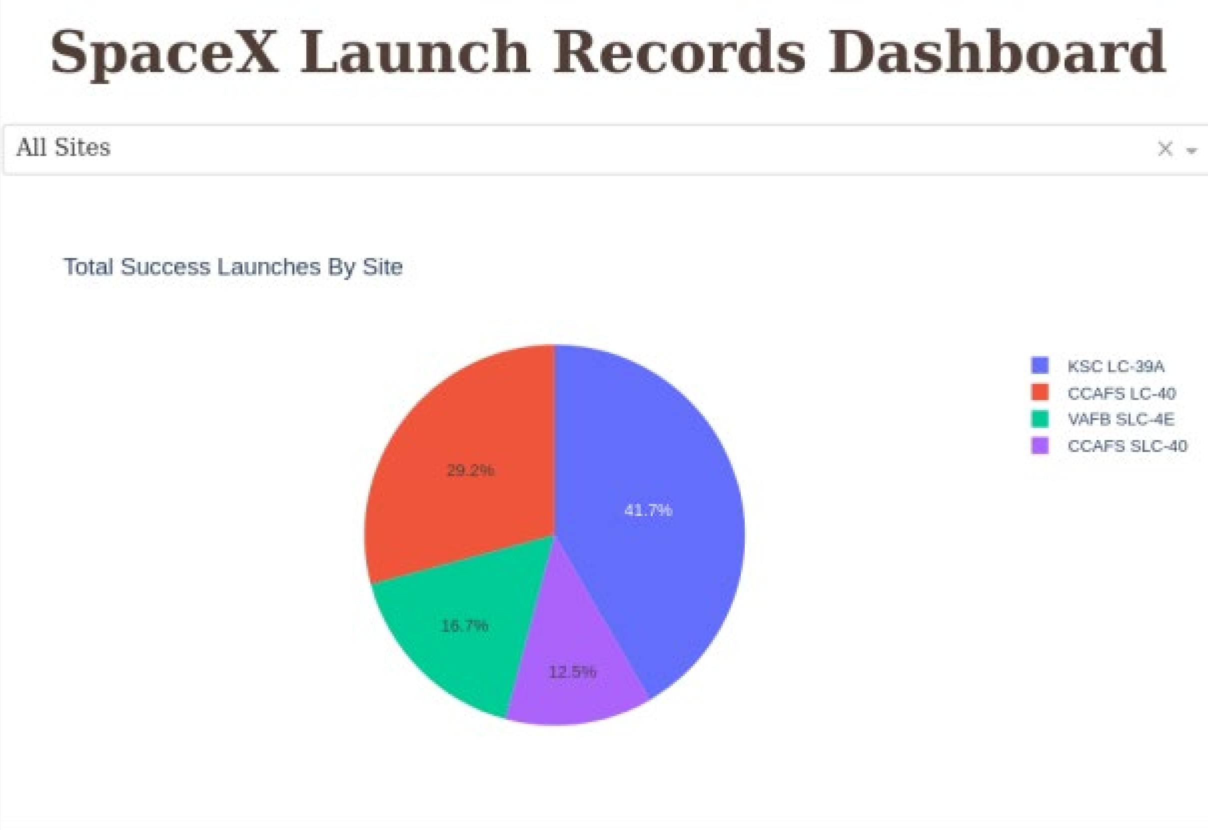



- It is important to choose sites that are close enough to be accessible, but not so close that the general public is endangered.  Here you can see that this site is far away from an inhabited area.

Section 4

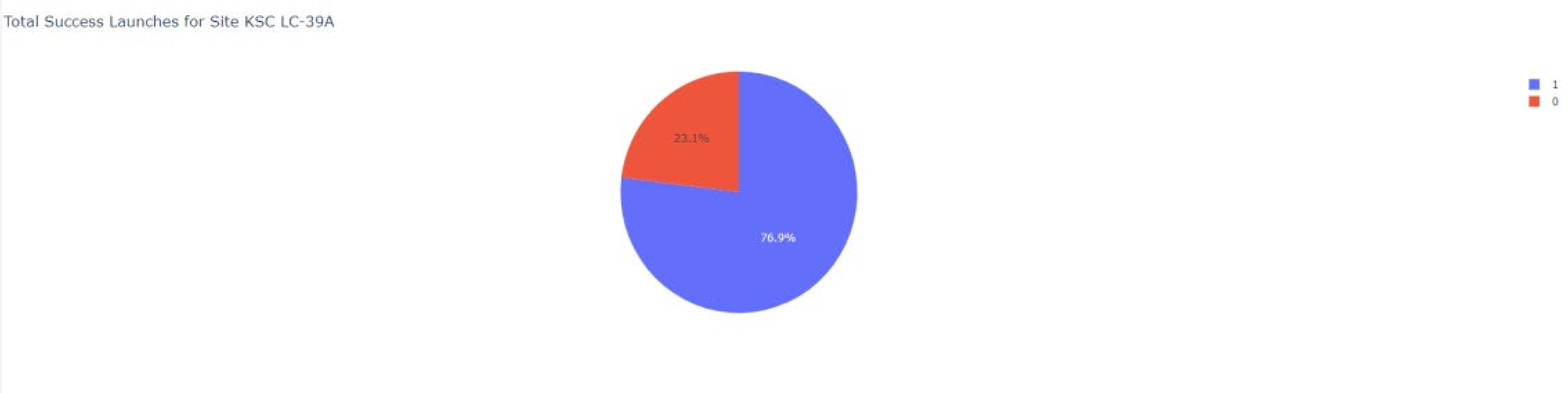
# Build a Dashboard with Plotly Dash

# Plotly Dash – Successful launches by site



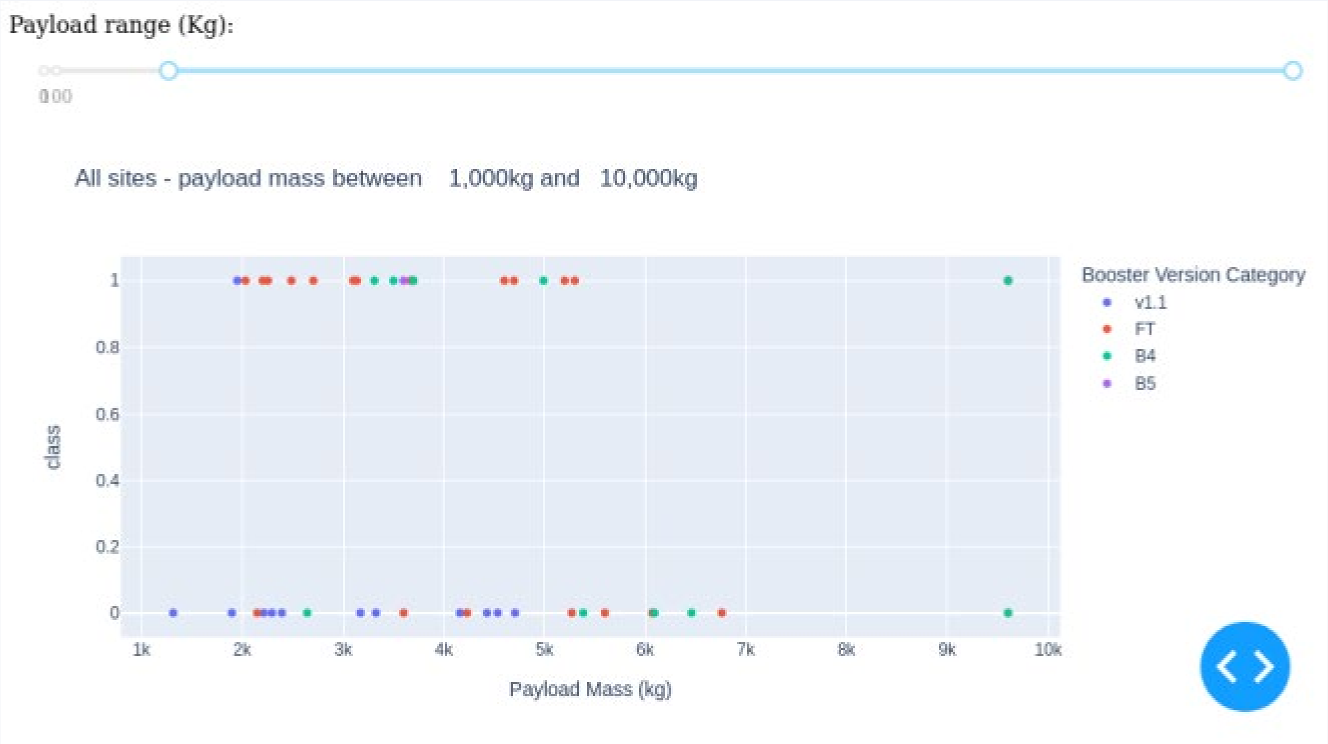

- As show in the chart above, KSC LC-39A has the highest success rate of all the sites.

# Plotly Dash – Successful launches KSC LC-39A




- KSC LC-39A has a 76.9% success rate

# Plotly Dash – Payload Mass vs Launch Outcome




- Booster version FT has the highest success rate

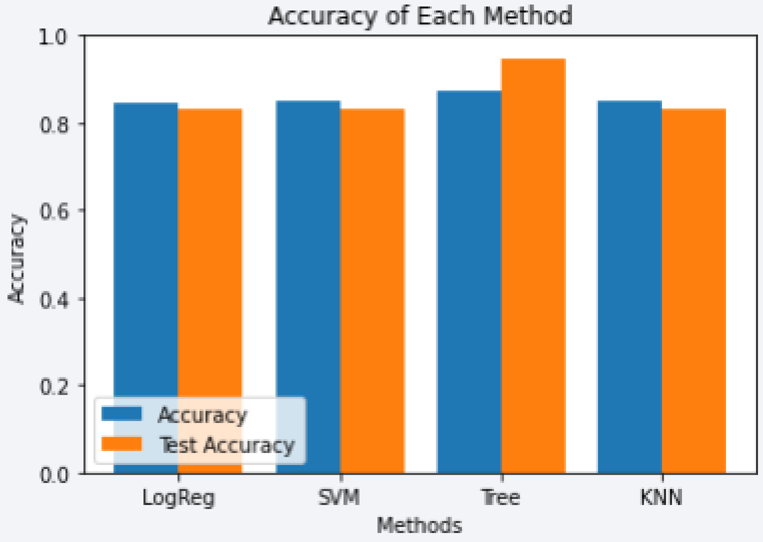
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

- The Decision Tree model has the highest accuracy.
- Using training data, the accuracy for this model was over 87%.
- Using testing data, the accuracy for this model was 83%.

| | Accuracy Train | Accuracy Test |
|---|---|---|
| Tree | 0.876786 | 0.833333 |
| Knn | 0.848214 | 0.833333 |
| Svm | 0.848214 | 0.833333 |
| Logreg | 0.846429 | 0.833333 |



# Confusion Matrix


- There is one instance of a False Positive in the confusion matrix

# Conclusions

- The most successful launch site is KSC LC-39A
- Landing outcomes have improved over time and with better technology and testing.
- Landing outcomes in some cases are orbit and payload dependent.
- ES-L1 is the most successful orbit.

# Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

# Thank you!