IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Uzoma Ewurum

31-12-2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

SpaceX has revolutionized the aerospace industry with reusable rockets and cost-efficient launches. This report aims to analyze historical SpaceX launch data to:

1. Identify trends and patterns.
2. Predict launch outcomes using machine learning models.

Section 1

Methodology

Methodology

Executive Summary

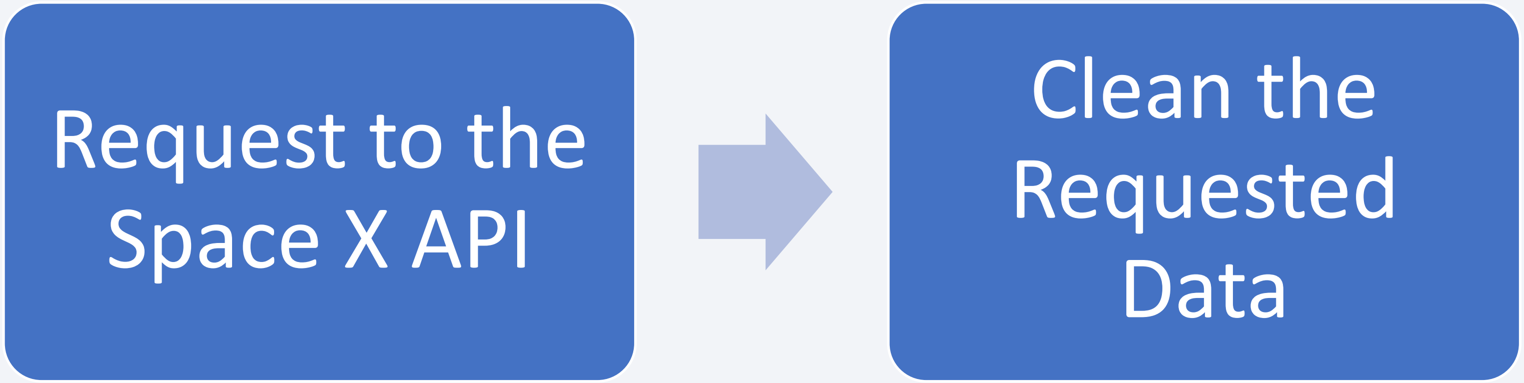
- Data collection methodology:
  - Falcon 9 historical launch records was collected by webscraping the Wikipedia page titled List of Falcon 9 and Falcon Heavy Launches.
  - Data Extraction using identification number in the Launch data by sending a get request to SpaceX API
- Perform data wrangling
  - Dataframe was filtered to include only Falcon 9 Launches and missing values (5) in the payloadmass column was replaced by the mean

Methodology

Executive Summary

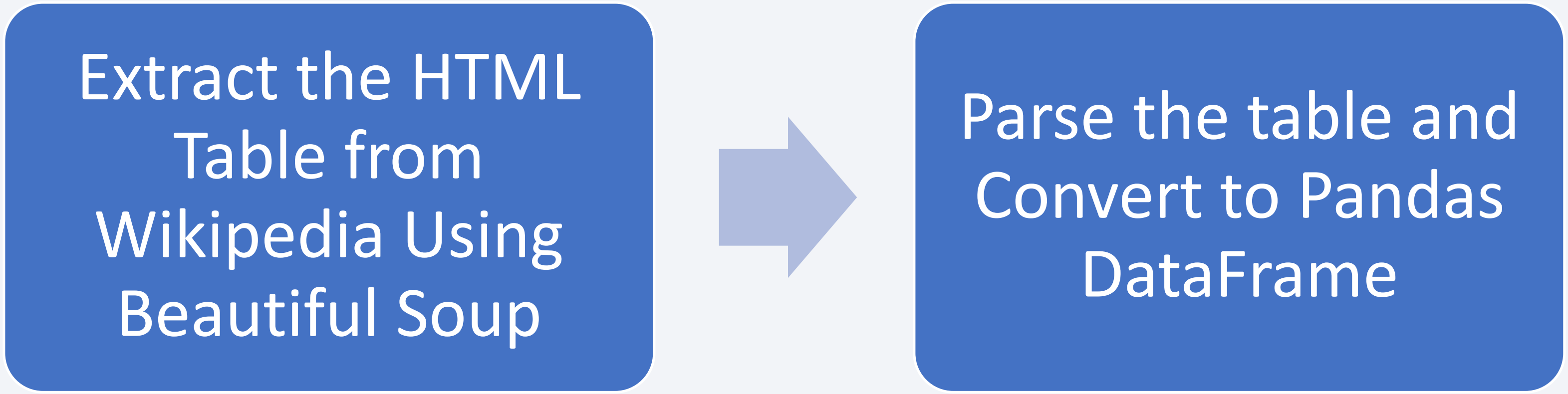
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API



https://github.com/uzilon/Data-science-Coursera/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb

Data Collection - Scraping



https://github.com/uzilon/Data-science-Coursera/blob/main/jupyter-labs-webscraping%20(1).ipynb

Data Wrangling

Fitter Data Frame to only include Falcon9 Launches

Replace Missing Values in the PayLoadMass column by its mean

https://github.com/uzilon/Data-science-Coursera/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb

EDA with Data Visualization

- Scatter Plot to understand the relationship between;
  - Flight number and Launch site
  - Payload mass and Launch Site
  - Flight number and Orbit type
  - Payload mass and Orbit type
- Bar Plot to compare success rate of each orbit type
- Line Plot to visualize launch success yearly trend

https://github.com/uzilon/Data-science-Coursera/blob/main/edadataviz.ipynb

EDA with SQL

https://github.com/uzilon/Data-science-Coursera/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb

- Name display of unique launch sites in space mission
- 5 records display where launch sites begin with the string 'CCA'
- Display of total payload mass carried by boosters launched by NASA (CRS)
- Display of average payload mass carried by booster version F9 v1.1
- Display showing date when the first successful landing outcome in ground pad was achieved.
- Name display of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Display total number of successful and failure mission outcomes
- Display list of names of the booster_versions which have carried the maximum payload mass.
- Display list of records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Display 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.

Build an Interactive Map with Folium

- Use of marker object to mark coordinates on map
- Use of circle object to add a highlighted circle area with a text label on a specific coordinates.
- Use of marker cluster object to simplify map containing many markers with the same coordinate.
- Use of mouseposition object to get coordinate for a mouse over a point on the map
- Use of lines to connect markers with highways, railways, coastline and cities and help calculate distance.

https://github.com/uzilon/Data-science-Coursera/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Dropdown showing all launch sites to make dashboard interactive.
- Payload range slider to interactively visualize choice payload ranges.
- Pie chart showing total success launches by site for all launch sites and success and failure for specific launch site
- Correlation between payload mass and success for specific launch site

https://github.com/uzilon/Data-science-Coursera/blob/main/Spacex_dash_app.py

Predictive Analysis (Classification)

- Create Dummy Variables for categorical data
- Extract the class column and convert it to a numpy array as Y
- Standardize the data and assign to X
- Train/Test Split
- Create Grid Search Object for each Model Object
- Model Fit
- Model Evaluation

Models

- Logistic Regression
- Support Vector Machine
- Decision Tree Classifier
- K Nearest Neighbour

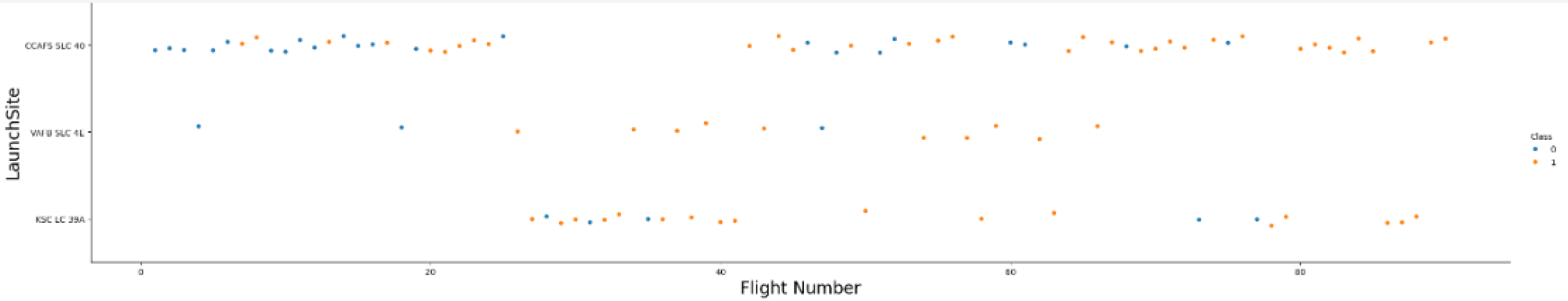
Results

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

Section 2

Insights drawn from EDA

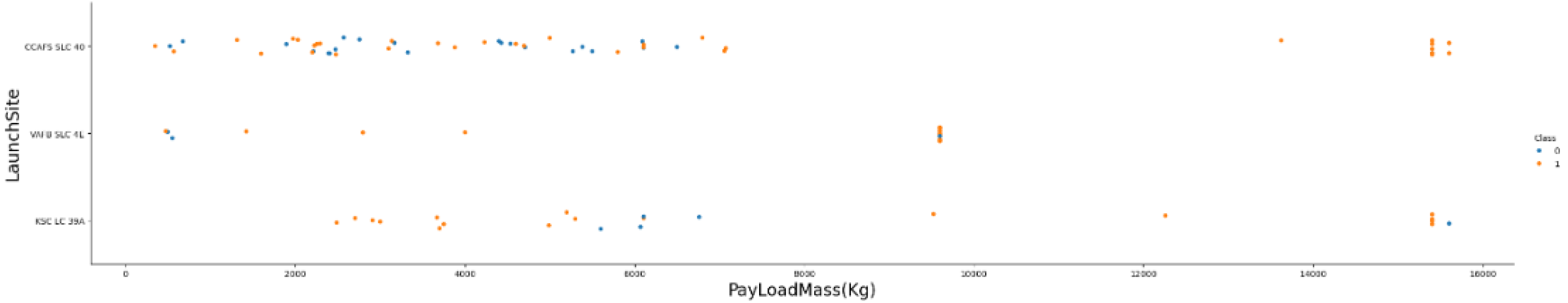
Flight Number vs. Launch Site



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

As flight number increases, the first stage is likely to land successfully

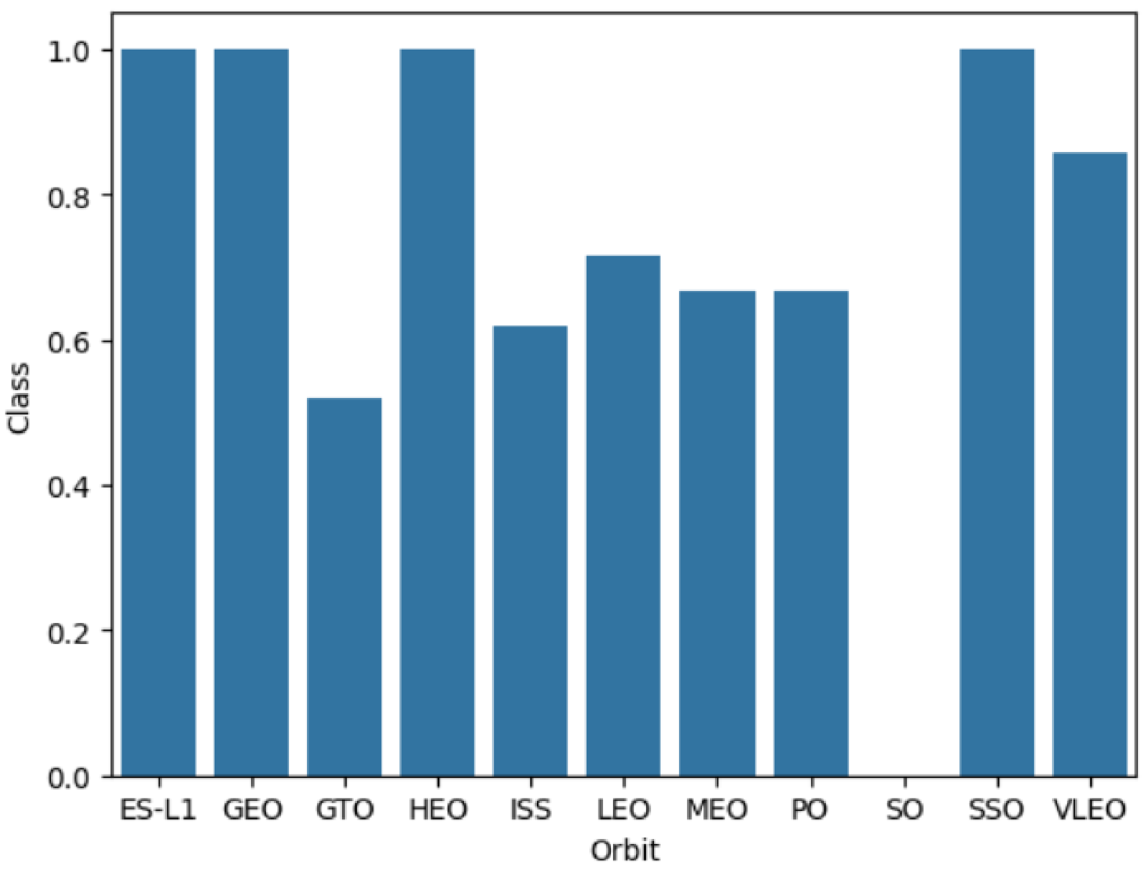
Payload vs. Launch Site



Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

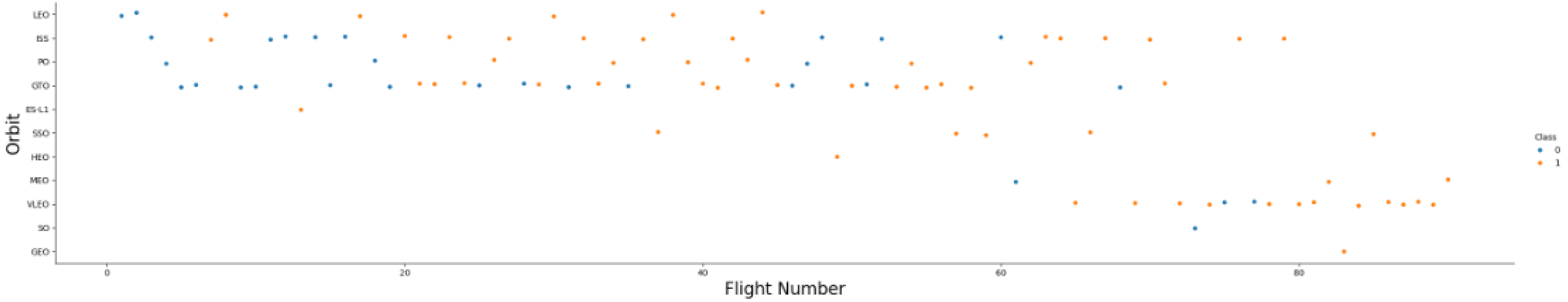
<AxesSubplot:xlabel='Orbit', ylabel='Class'>



Analyze the plotted bar chart to identify which orbits have the highest success rates.

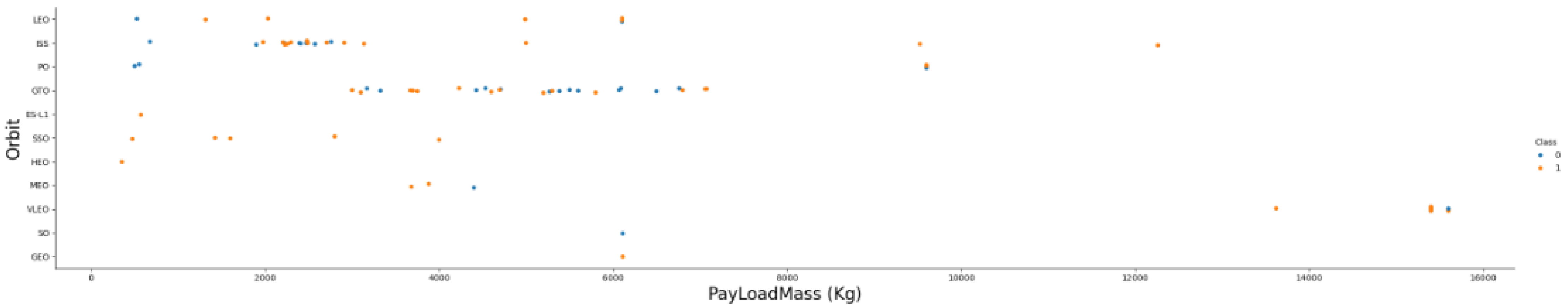
ES_LI, GEO, HEO, SSO had the highest success rates

Flight Number vs. Orbit Type



You can observe that in 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.

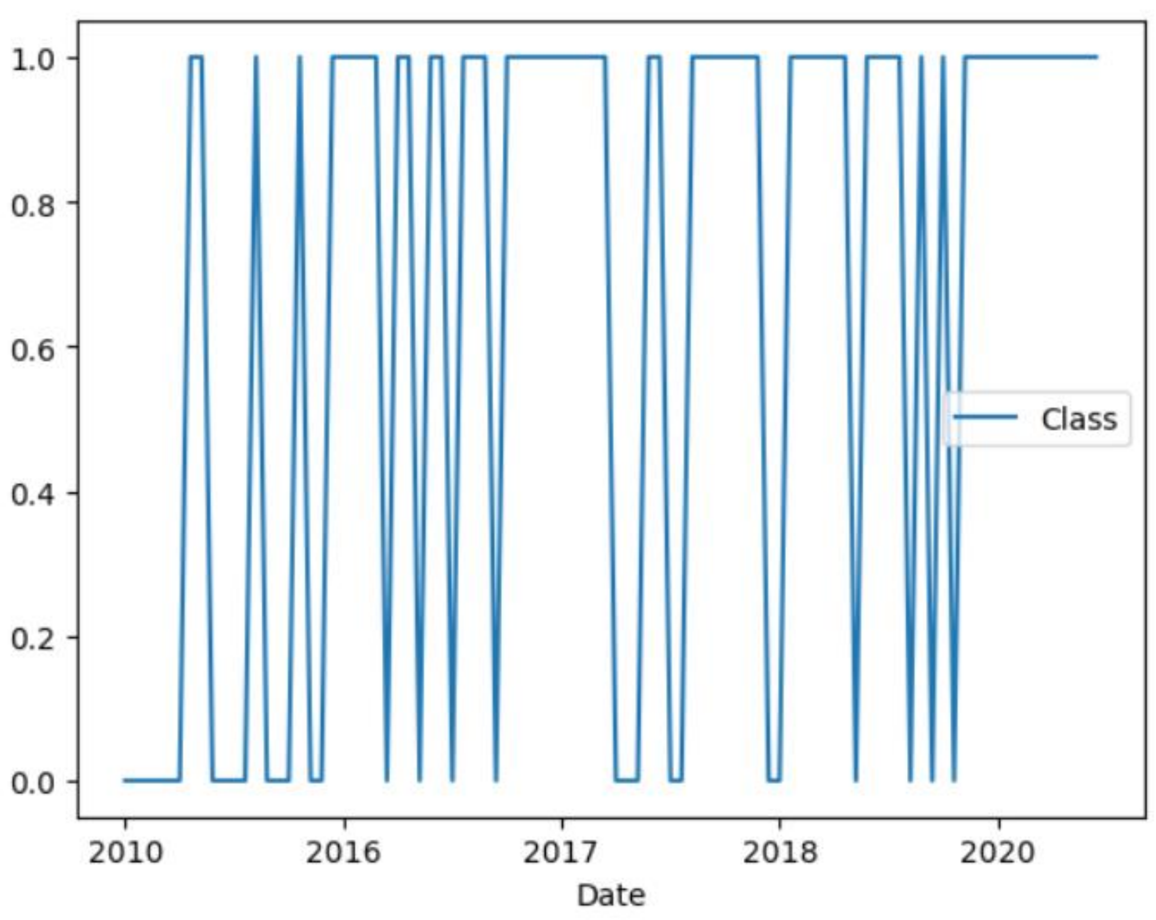
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



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

All Launch Site Names

%sql select distinct(Launch_Site) from SPACEXTABLE

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

Four distinct launch sites

Launch Site Names Begin with 'CCA'

%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

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

5 Records display of launch site CCAFS LC-40

Total Payload Mass

%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer like 'NASA%'

SUM(PAYLOAD_MASS__KG_)
99980

Total Payload Mass of 99,980 for all NASA customers

Average Payload Mass by F9 v1.1

%sql select Avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version == 'F9 v1.1'

Avg(PAYLOAD_MASS__KG_)
2928.4

Average Payload Mass for Booster Version is 2,927.4

First Successful Ground Landing Date

%sql select min(Date) from SPACEXTABLE where Landing_Outcome == 'Success (ground pad)'

min(Date)
2015-12-22

First Successful Ground Landing Date was December 22nd 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

%sql select Booster_Version from SPACEXTABLE where Landing_Outcome == 'Success (drone ship)' and PAYLOAD_MASS__KG_ between '4000' and '6000'

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

Names of successful drone ship landing boosters with payload between 4000 & 6000

Total Number of Successful and Failure Mission Outcomes

%sql select Landing_Outcome, count(Landing_Outcome) as Count from SPACEXTABLE where Landing_Outcome in ('Success', 'Failure') Group by 1

Landing_Outcome	Count
Failure	3
Success	38

Total number of successful and failure missions

Boosters Carried Maximum Payload

%sql select Booster_Version, max(PAYLOAD_MASS__KG_) from (select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE) as Table1

Booster_Version	max(PAYLOAD_MASS__KG_)
F9 B5 B1048.4	15600

The above booster carried maximum payload

2015 Launch Records

%sql select substr(Date, 6,2), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where Landing_Outcome == 'Failure (drone ship)' AND substr(Date,0,5)='2015'

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Two Launches in 2015, One in January and the other in April

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

%sql select Landing_Outcome, count(Landing_Outcome) as Count from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' Group by 1 order by 2 desc

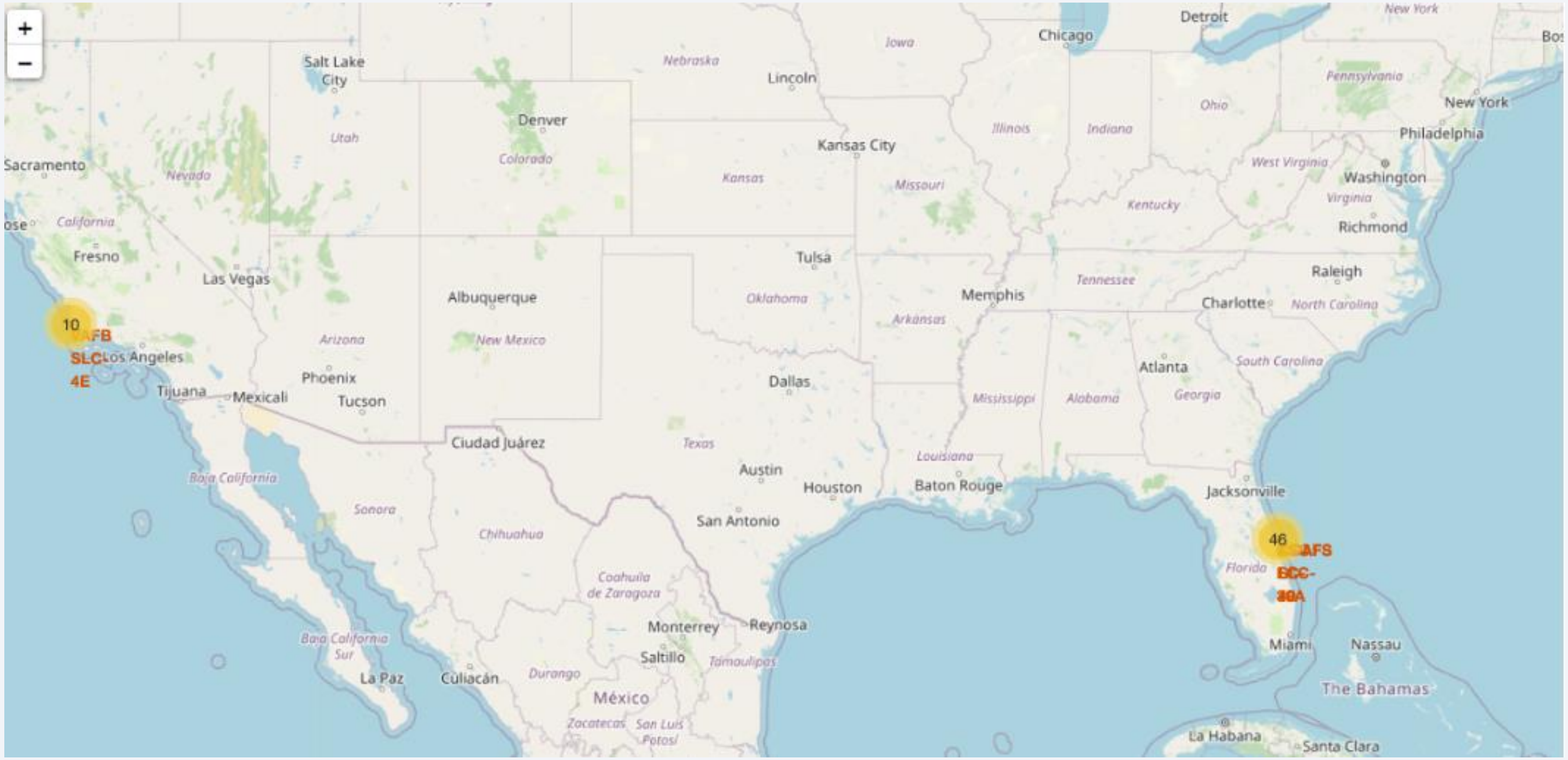
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Landing outcomes ranking between 2010-06-04 and 2017-03-20

Section 3

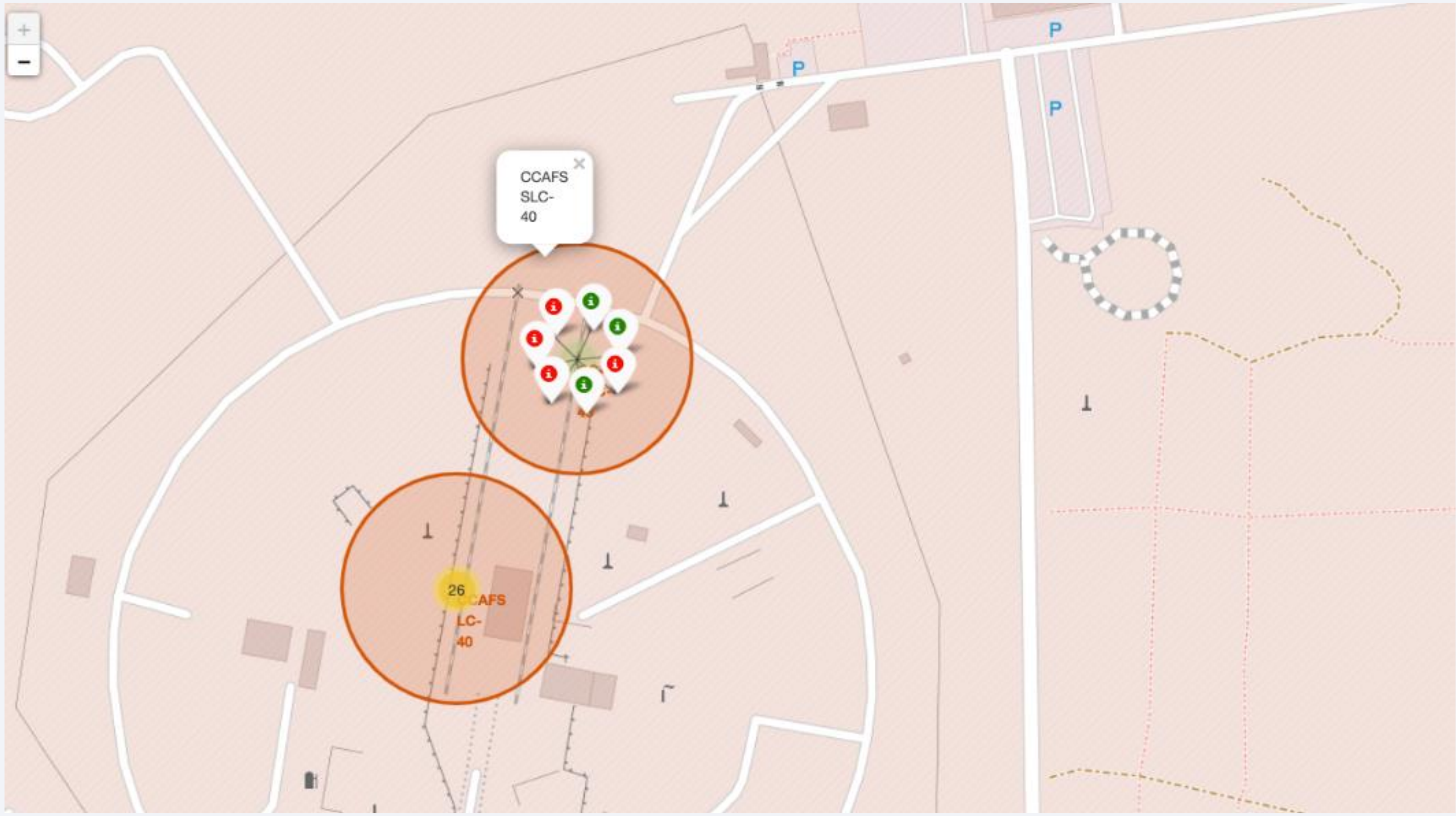
Launch Sites Proximities Analysis

Global Map Launch Sites Location Marker



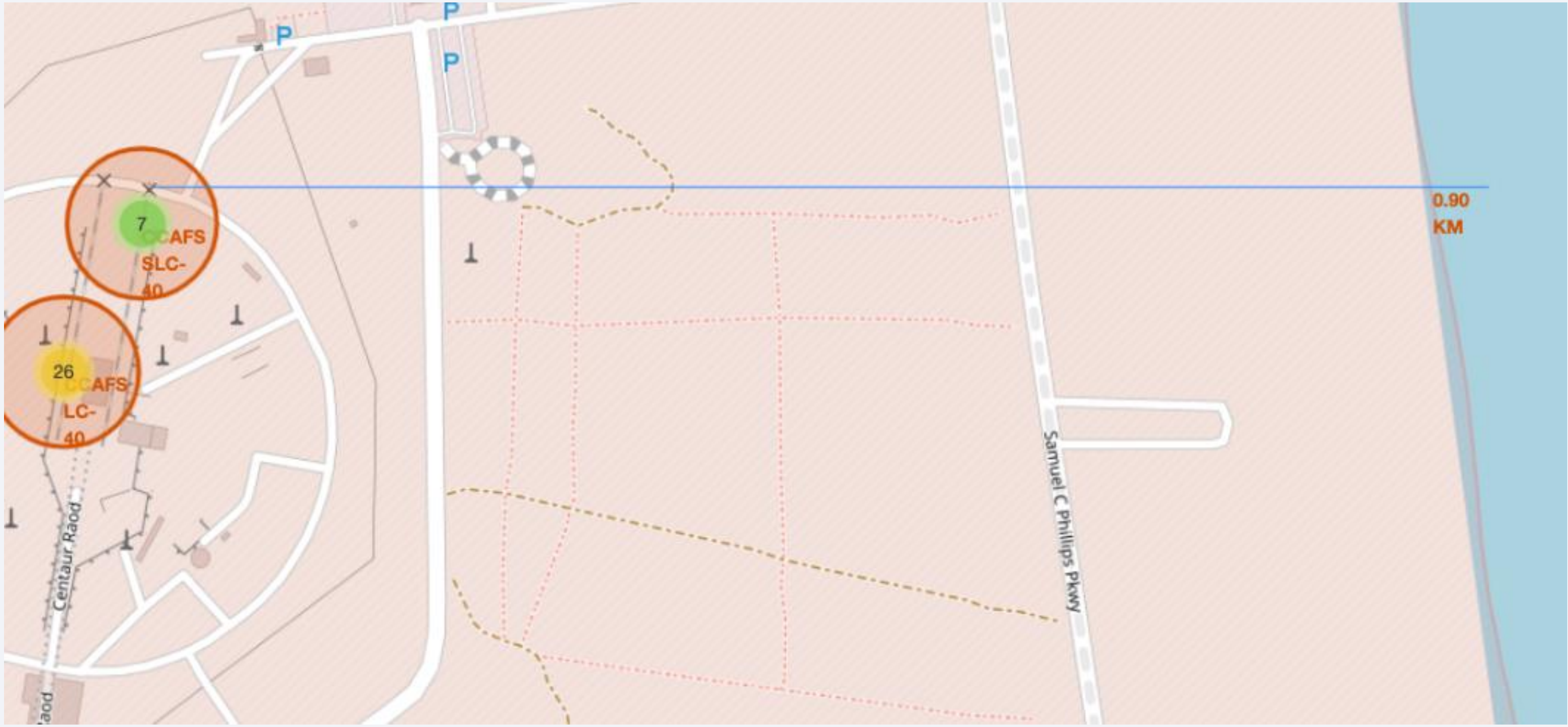
Location marker to help explore location launch sites showing a total of 56 Launch sites with 46 near Florida and 10 near Los Angeles

Color Label Launch Outcomes



Color Label Launch Outcomes to help easily identify which Launch Sites have relatively high success rate.

Launch Sites Proximities

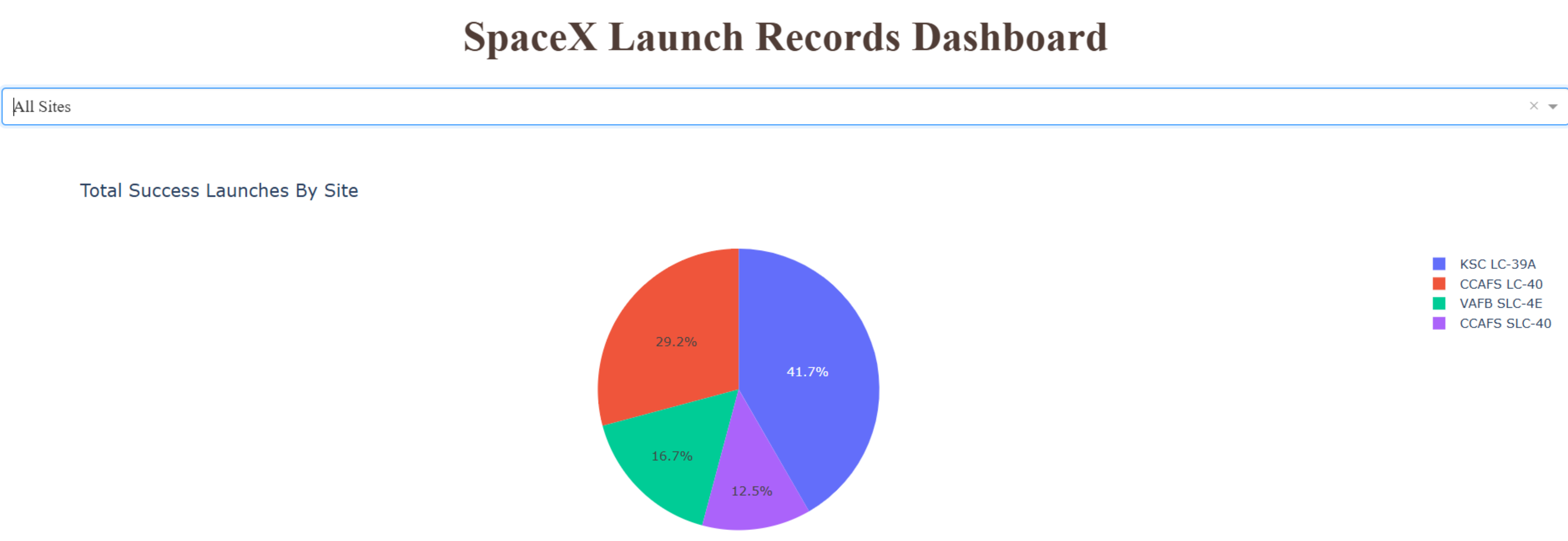


Selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

Section 4

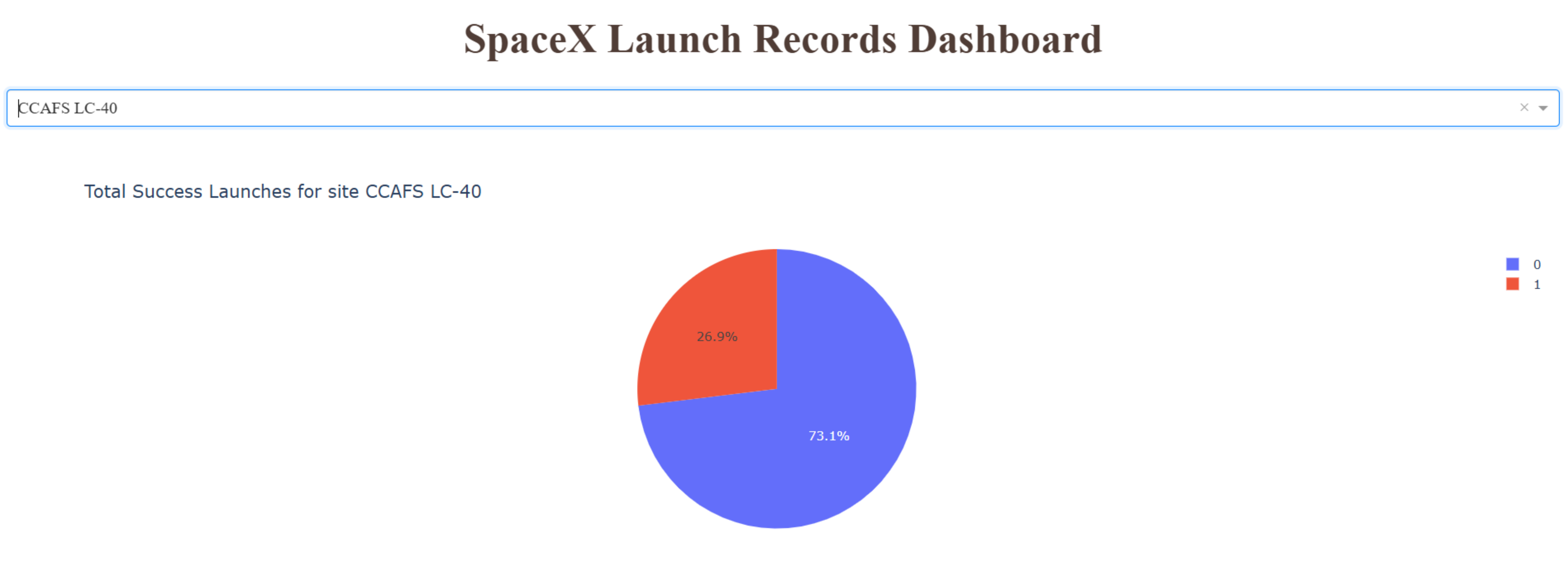
Build a Dashboard with Plotly Dash

All Sites Launch Success Ratio



Launch Success Ratio for all sites with CCAFS SLC-40 as the lowest and KSC LC-39A as the highest.

CCAFS LC- 40 Launch Success Ratio



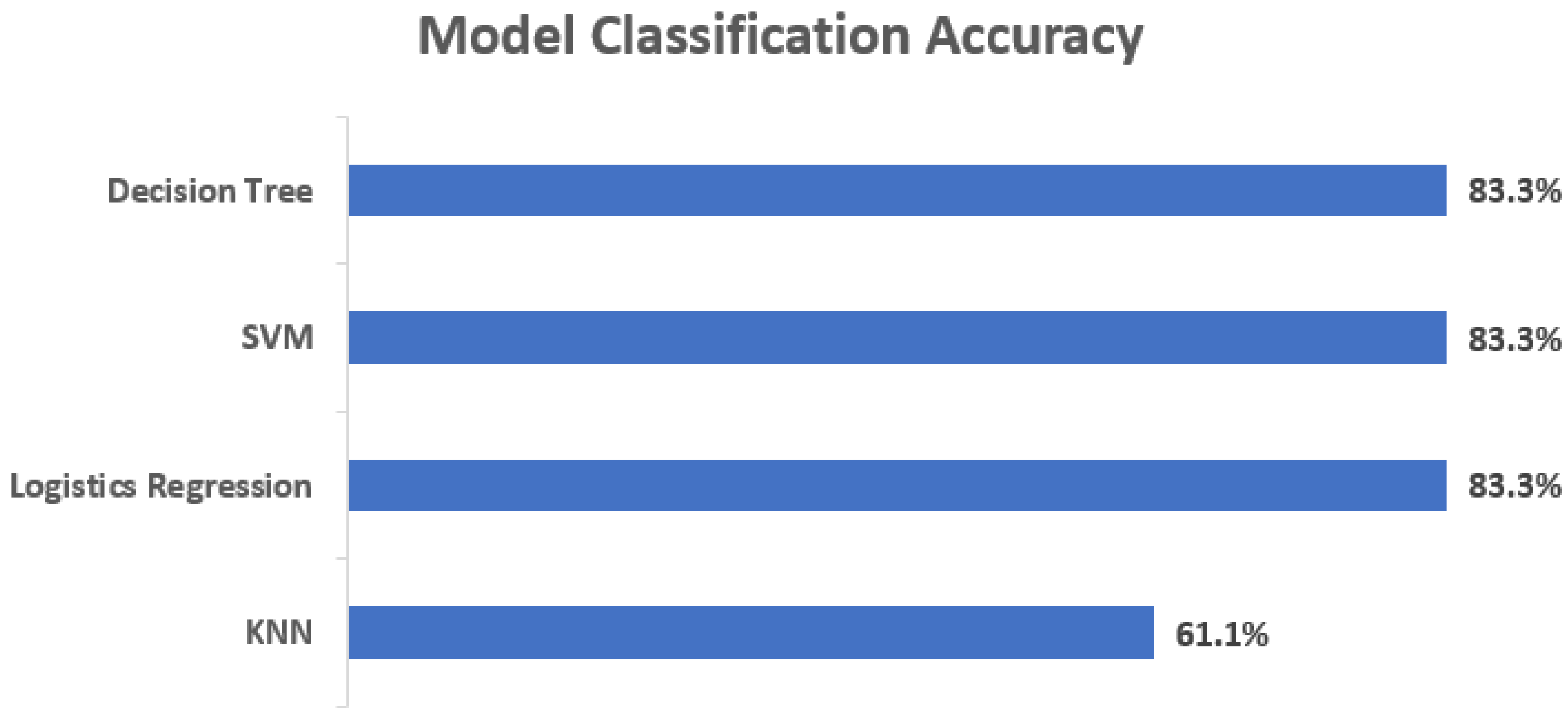
Payload vs. Launch Outcome Scatter Plot

Booster Version F showed a high success outcome at Payload between 2000 -5000

Section 5

Predictive Analysis (Classification)

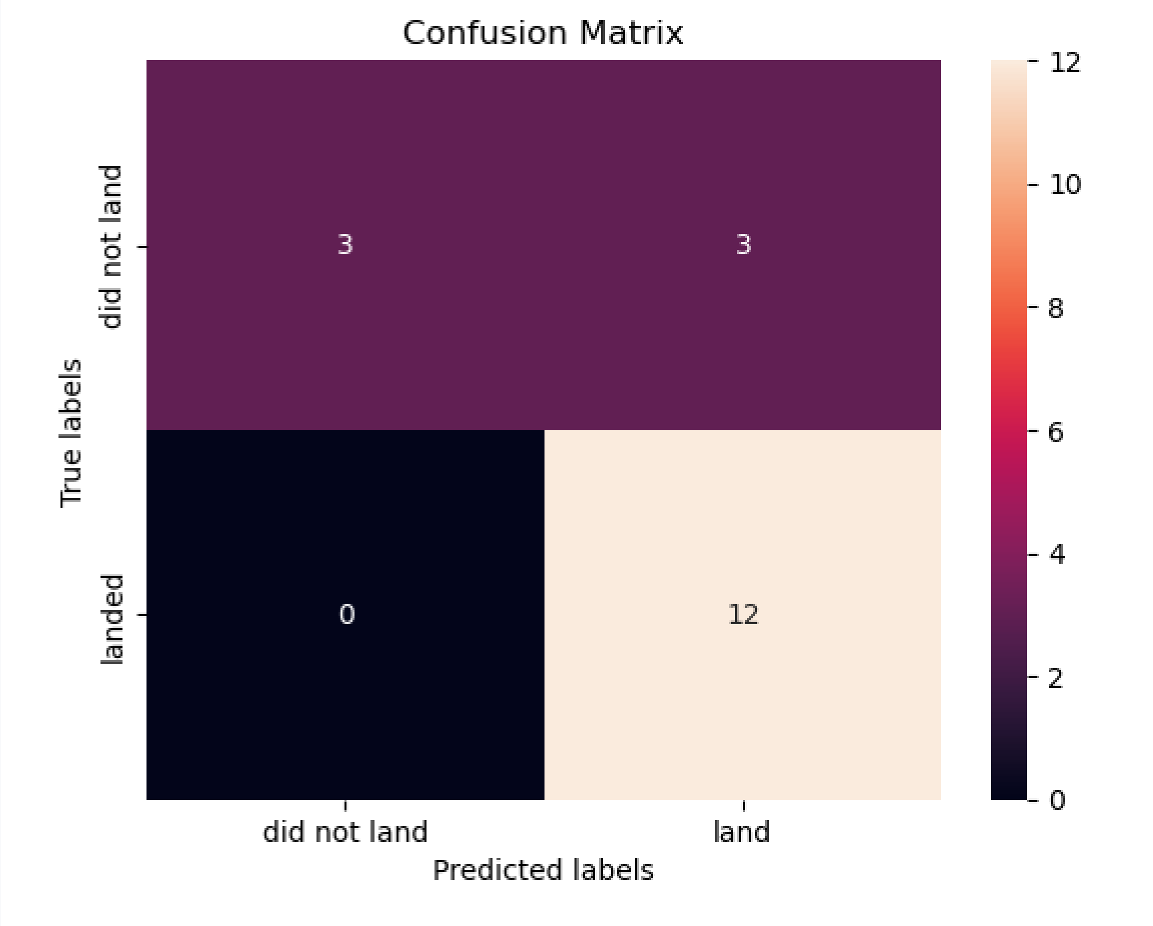
Classification Accuracy



	KNN	Logistics Regression	SVM	Decision Tree
Series1	61.1%	83.3%	83.3%	83.3%

Decision Tree, SVM and Logistic Regression all had same classification accuracy of 83% as the highest

Confusion Matrix



It shows model can distinguish between the different classes, the only problem is the true positive which depicted that 3 landed when it did not land.

Conclusions

Launch Site Success Analysis:
•Four distinct launch sites were identified, with **CCAFS SLC-40** having the lowest success ratio and **KSC LC-39A** achieving the highest success ratio.

Payload Trends:
•Total payload mass carried by NASA customers amounted to 99,980 kg.
•The average payload mass for booster version **F9 v1.1** was 2,927.4 kg.

Significant Milestones:
•The first successful ground landing occurred on **December 22, 2015**.
•Successful drone ship landings were achieved for payloads between **4,000–6,000 kg**.

Performance of Boosters:
•Certain boosters carried maximum payloads, highlighting technological advancements.

Machine Learning Insights:
•Classification models like Decision Tree, SVM, and Logistic Regression achieved the highest accuracy at **83%**.
•Confusion matrix analysis indicated issues with true positives, reflecting challenges in distinguishing specific outcomes.

Appendix

Datasets

dataset_part_1.csv

dataset_part_2.csv

dataset_part_3.csv

ML_data.csv

Spacex.csv

Machine Learning Notebook

ML Notebook

Thank you!