IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sean R.
05.11.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection via API
  - Data Wrangling
  - Exploratory Analysis using SQL, Pandas, and Matplotlib
  - Dashboarding with Folium and Dash
  - Predictive Analysis
- Summary of all results
  - Select most accurate model

Introduction

- Project background and context
  - SpaceX offers launches at a fraction of the cost of its competitors. The firm is able to do this because of the Falcon 9's reusable first stage. In this case study, a competing firm, SpaceY, is interested in predicting and modeling SpaceX launches to make their own market decisions.
- Problems you want to find answers
  - How to predict the outcome of SpaceX launches?
  - What factors influence launch success?

Section 1

Methodology

Methodology

Executive Summary

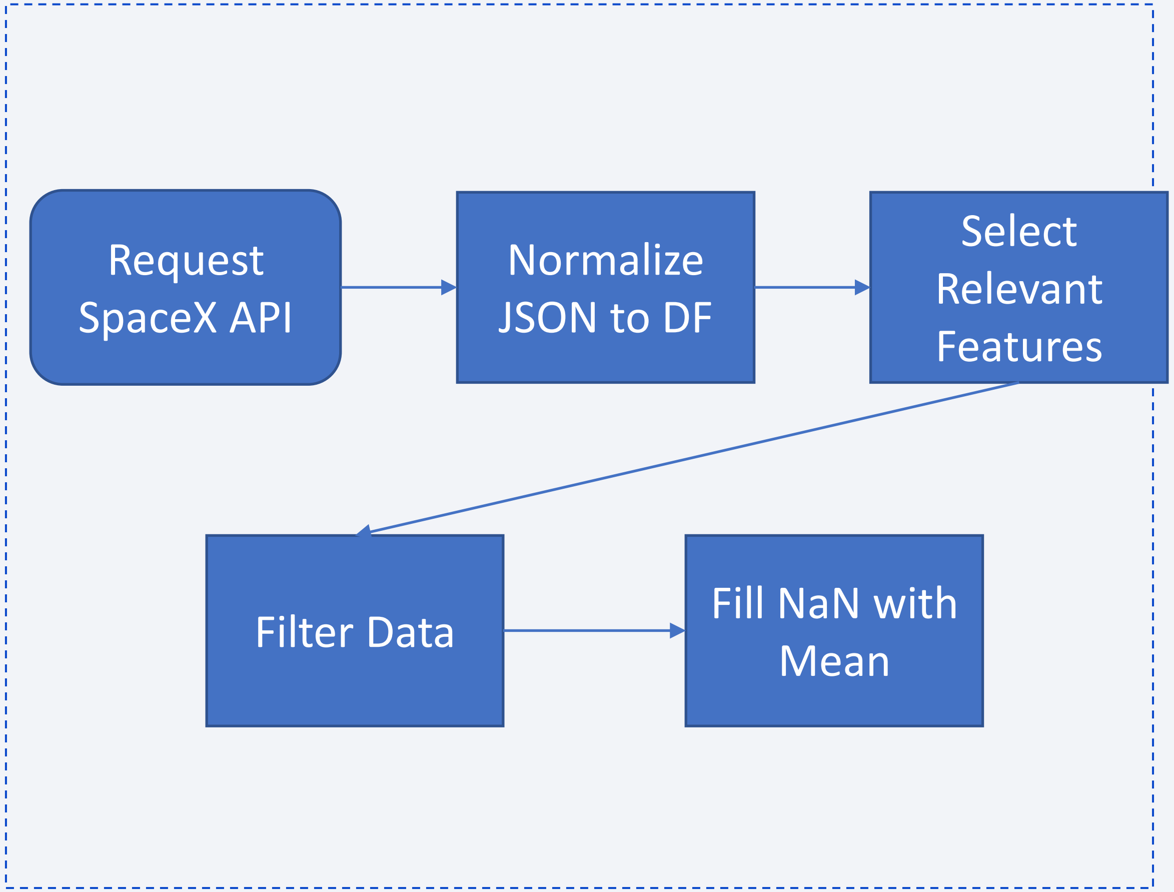
- Data collection methodology:
    - API call and Web Scraping
- Perform data wrangling
    - Cleaned and Calculated
- Perform exploratory data analysis (EDA) using visualization and SQL
    - Understand the data
- Perform interactive visual analytics using Folium and Plotly Dash
    - Make analysis accessible
- Perform predictive analysis using classification models
    - Tune and select models

Data Collection

- SpaceX API Call
  - Requests to call API
  - JSON Normalization to Data Frame
  - Data Processing
- Web Scrape of Wikipedia
  - Requests and Beautiful Soup
  - Parse HTML
  - Dictionary of table elements to Data Frame

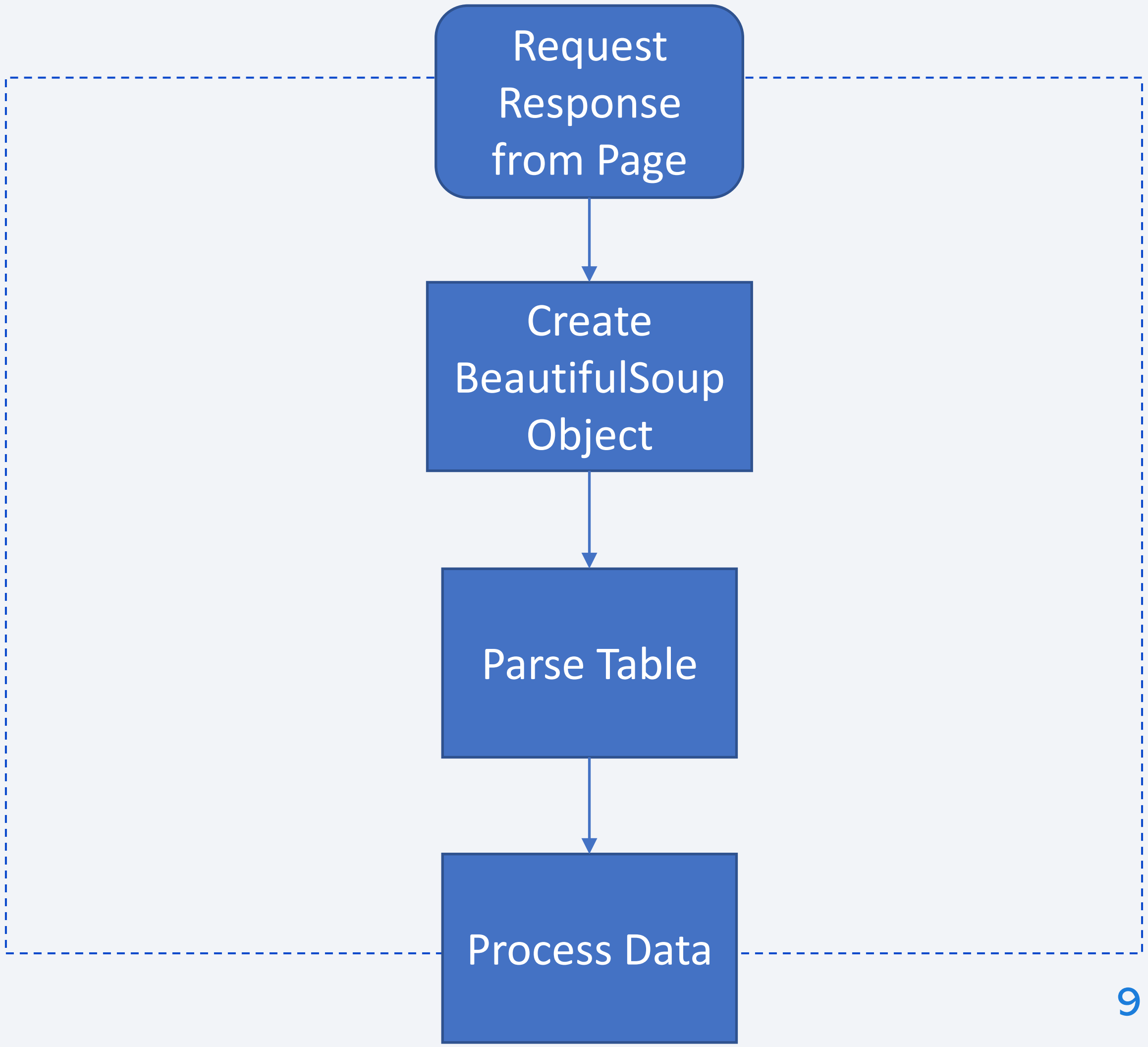
Data Collection – SpaceX API

- The API was called with the Requests package
- The raw JSON data was normalized and written to a data frame
- Processing steps were taken to limit rows to only relevant data
- Missing Payload Mass values were populated with the mean
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/00DataCollection-API.ipynb



Data Collection - Scraping

- Web page was scraped with Requests
- Response was used to create BeautifulSoup object
- Table of interest was parsedto Data Frame from BeautifulSoup Object
- Data processing
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/01DataCollection-WebScraping.ipynb



Data Wrangling

- EDA to find:
  - Missing values
  - Number of launches at each site
  - Number and occurrence of each orbit
  - Mission outcomes for each orbit
- Additional “Class” column derived from “Outcome”
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/03Data Wrangling.ipynb

EDA with Data Visualization

- Plots:
  - Flight Number and Launch Site – Scatter
  - Payload and Launch Site – Scatter
  - Success Rate by Orbit Type – Bar
  - Flight Number and Orbit Type – Scatter
  - Payload and Orbit Type – Scatter
  - Launch Success Over Time – Line
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/05Matplotlib.ipynb

EDA with SQL

- Queries Performed:
  - Unique Launch Sites
  - 5 Launch Sites that begin with “CCA”
  - Sum of Payload Mass for boosters launched by NASA (CRS)
  - Average Payload Mass carried by F9 v1.1 Boosters
  - First date of successful ground pad landing
  - Boosters with successful drone ship landings whose payload mas is > 4000 and < 6000
  - Totals of mission outcome Successes and Failures
  - Boosters that have carried the maximum Payload Mass
  - Failed landings in 2015
  - Descending ranked count of successful landing outcomes between June 4, 2010 and March 20, 2017
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/04SQL.ipynb

Build an Interactive Map with Folium

- Objects Used:
  - Markers – to identify launch sites
  - Marker Clusters – to identify multiple launches at a single site
  - Lines – to show distance between two Markers
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/06Folium.ipynb

Build a Dashboard with Plotly Dash

- Plots:
  - Total Launches by Launch Site – Pie Chart
  - Success by Payload Mass for Booster Versions – Scatter
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/07Dash.py

Predictive Analysis (Classification)

- Process:
  - Standardize “Class” column
  - Split data into training and testing sets
  - Create a model with training data:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree
    - KNN
  - Calculate model accuracy using test data
  - Evaluate highest scoring model
- https://github.com/smryd/IBMDataScienceCapstone/blob/main/08ML.ipynb

Results

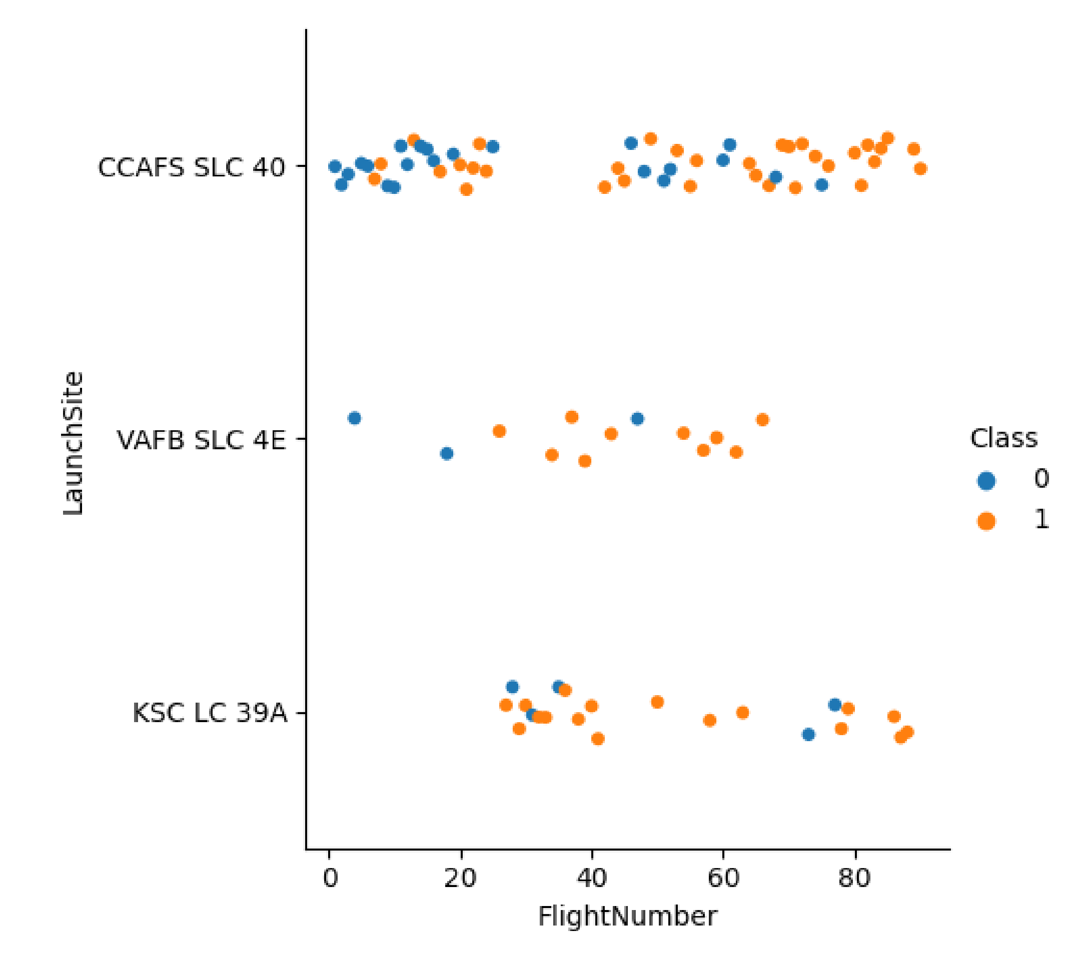
- EDA Results:
  - CCAFS SLC 40 had the most launches (55)
  - The most common orbit was GTO (27)
  - The Success Rate of launches was roughly 67%
  - The Average Payload Mass carried by F9 v1.1 boosters was 2928.4 kg
- Predictive analysis results
  - The Decision Tree model had the highest in-sample accuracy of 0.875 but all models performed the same on the test data, with 0.833

Section 2

Insights drawn from EDA

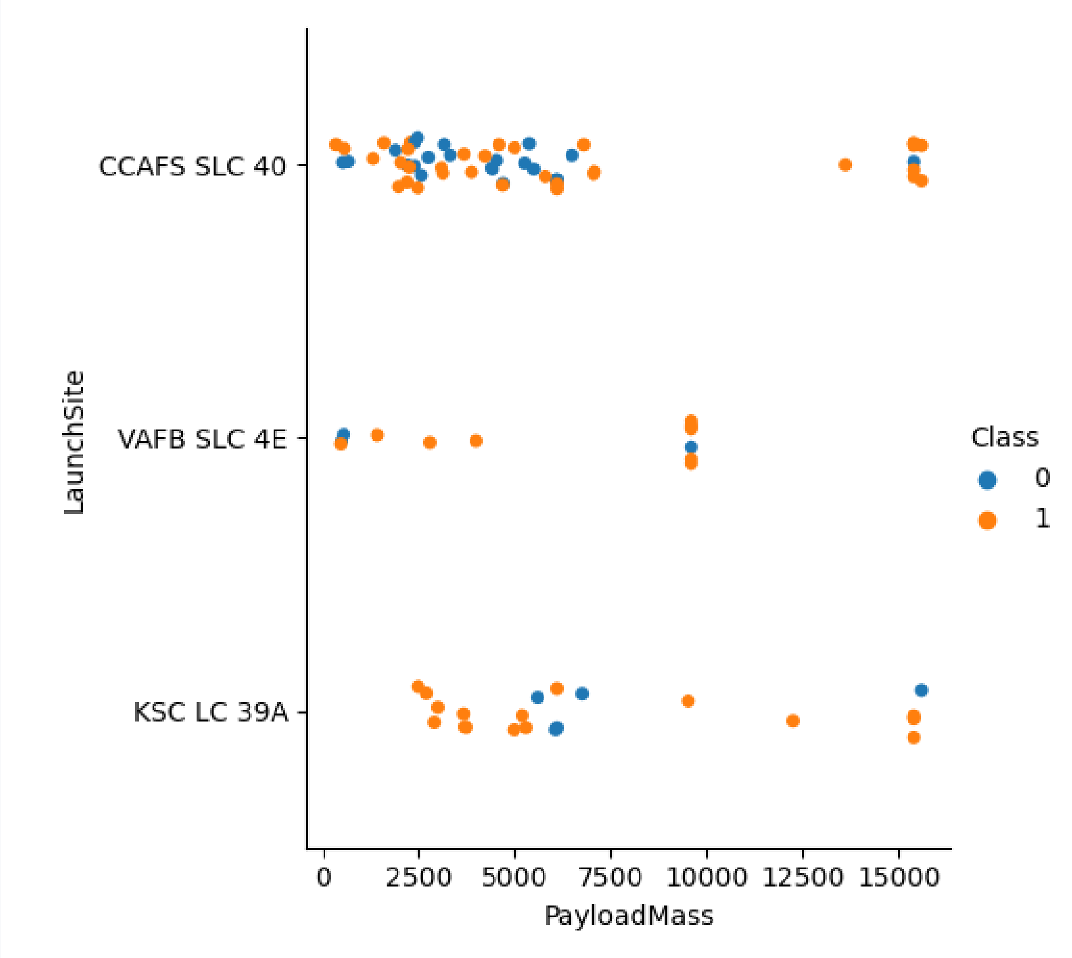
Flight Number vs. Launch Site

- This scatter plot shows the Successes (1) and Failures (0) of Flight Numbers at Launch Sites



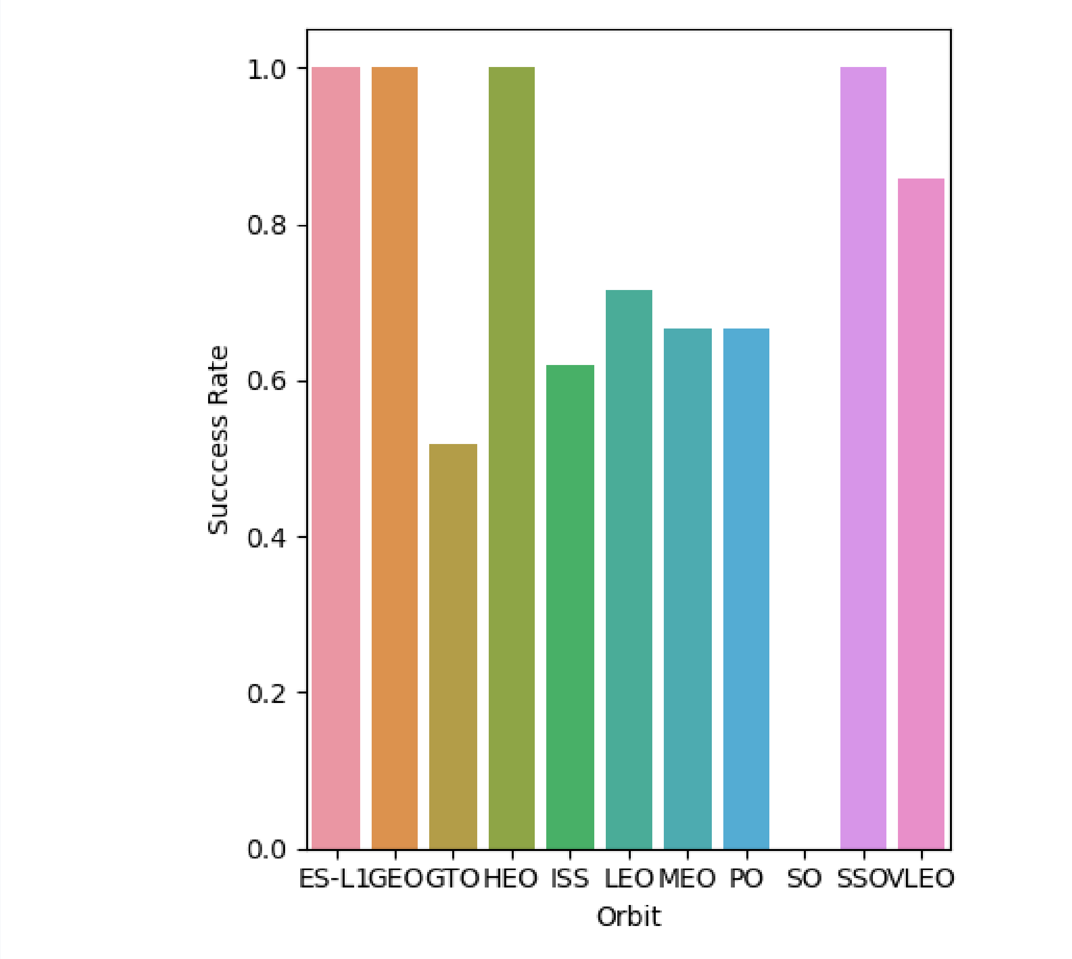
Payload vs. Launch Site

- This scatter plot shows the Successes (1) and Failures (0) of Payload Masses at Launch Sites



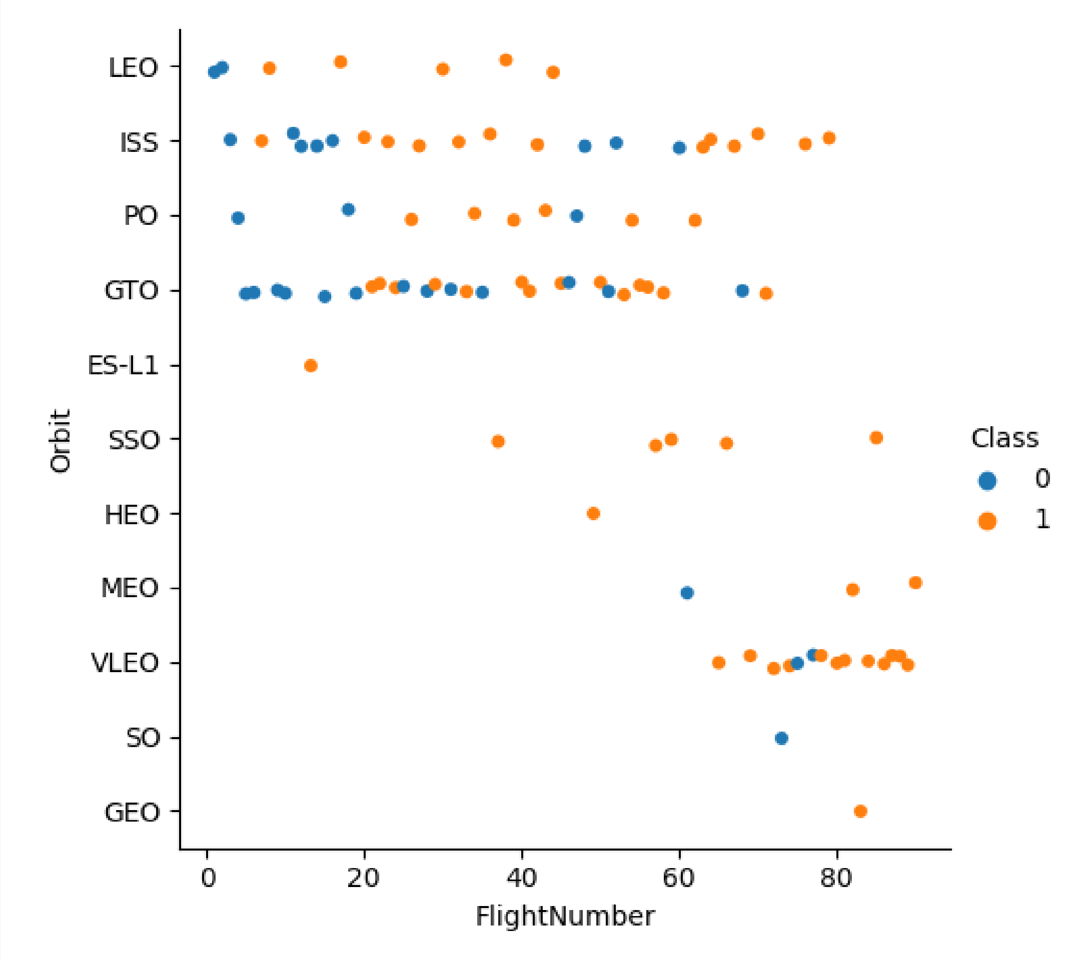
Success Rate vs. Orbit Type

- This bar chart shows the Success Rates for each Orbit Type



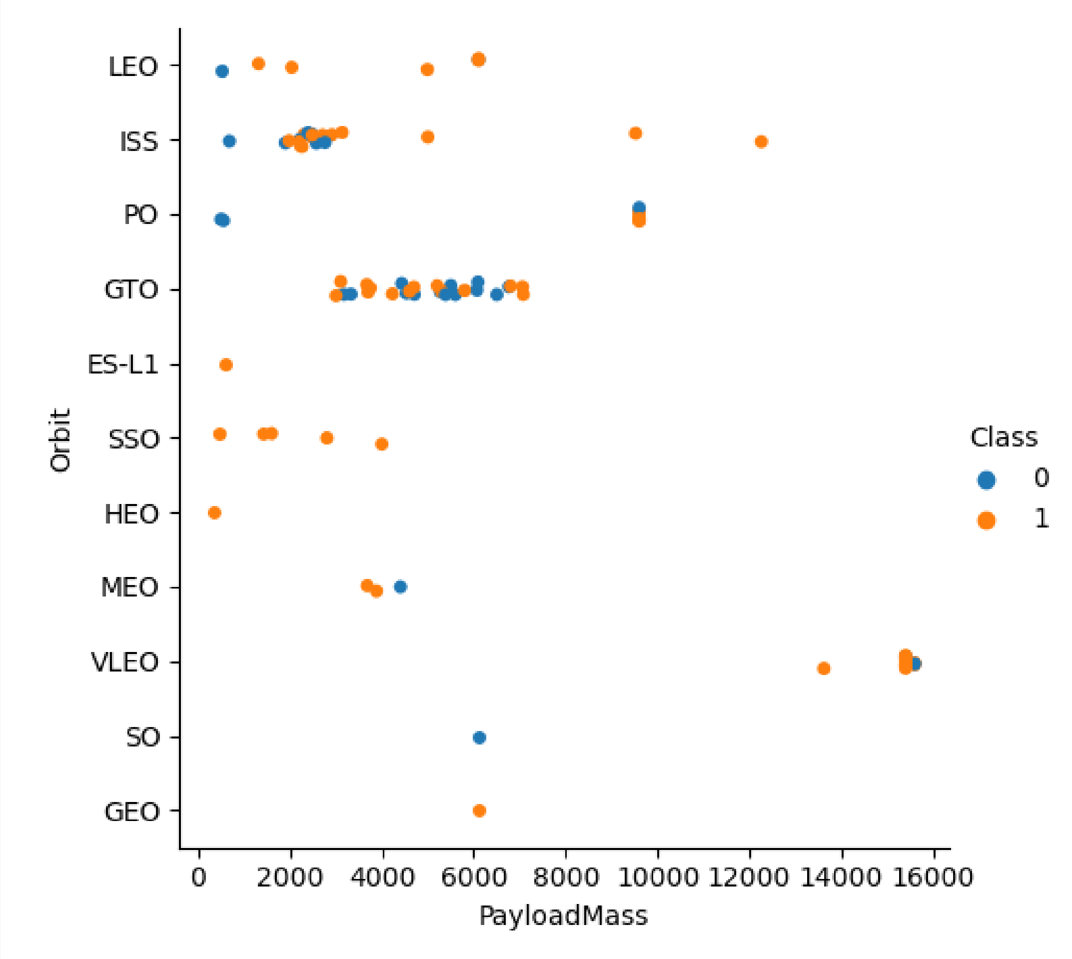
Flight Number vs. Orbit Type

- This scatter plot shows the Successes (1) and Failures (0) of Flight Numbers by Orbit Type



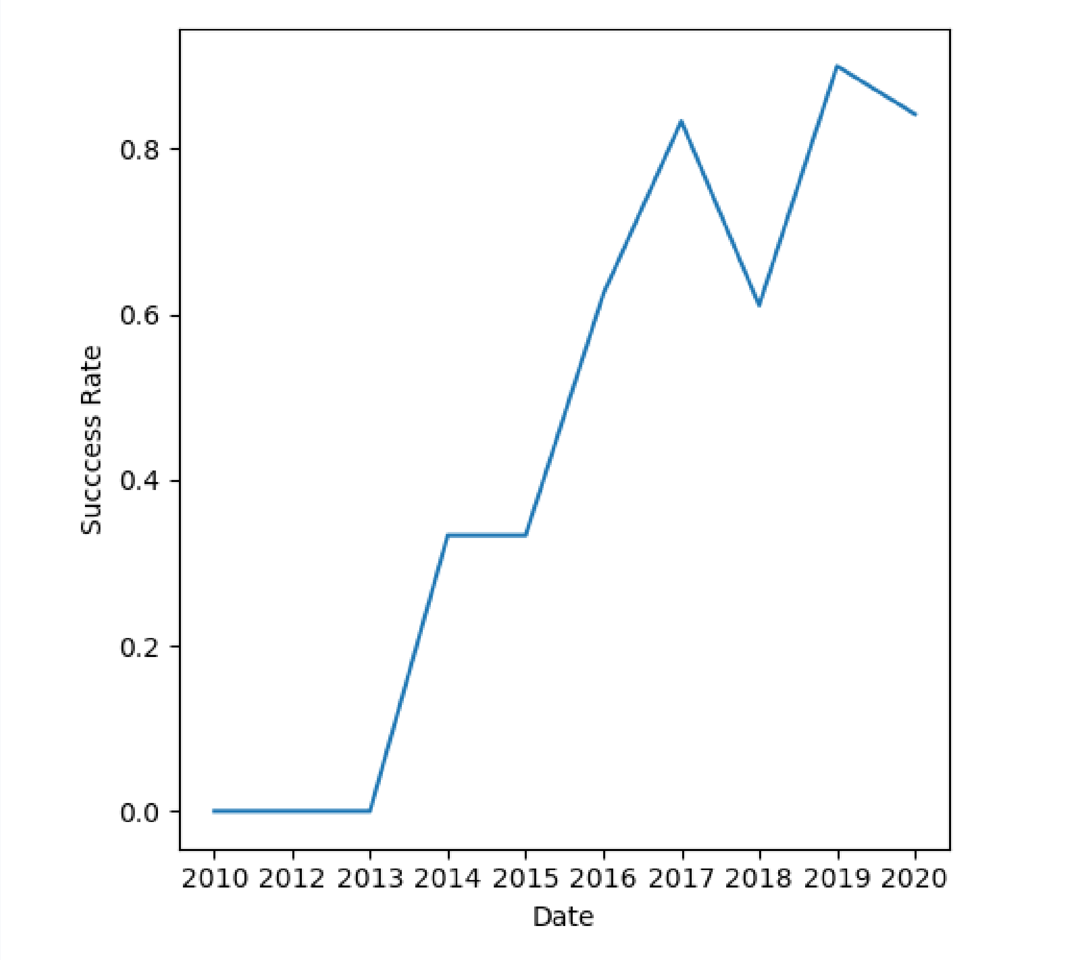
Payload vs. Orbit Type

- This scatter plot shows the Successes (1) and Failures (0) of Payload Masses by Orbit Type



Launch Success Yearly Trend

- This line chart shows the yearly trend of launch Success Rates



All Launch Site Names

- The table below contains a Distinct selection of Launch Sites:

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

Launch Site Names Begin with 'CCA'

- 5 Records whose Launch Site begins with “CCA”

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

MIN(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

COUNT(*)
1
98
1
1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

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

2015 Launch Records

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

substr(Date, 4, 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

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

- Count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, ranked in descending order (First 5)

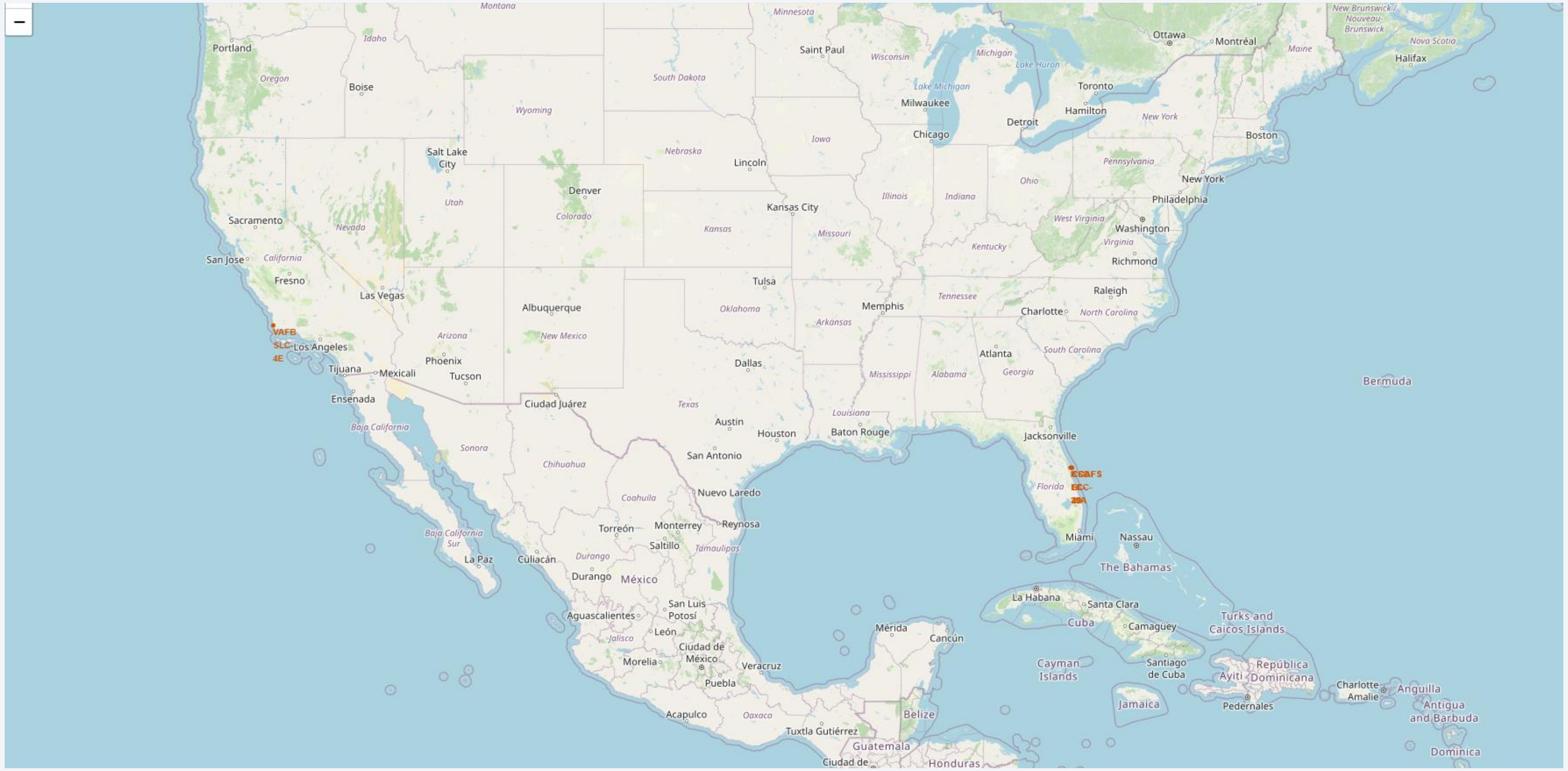
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)

Section 3

Launch Sites Proximities Analysis

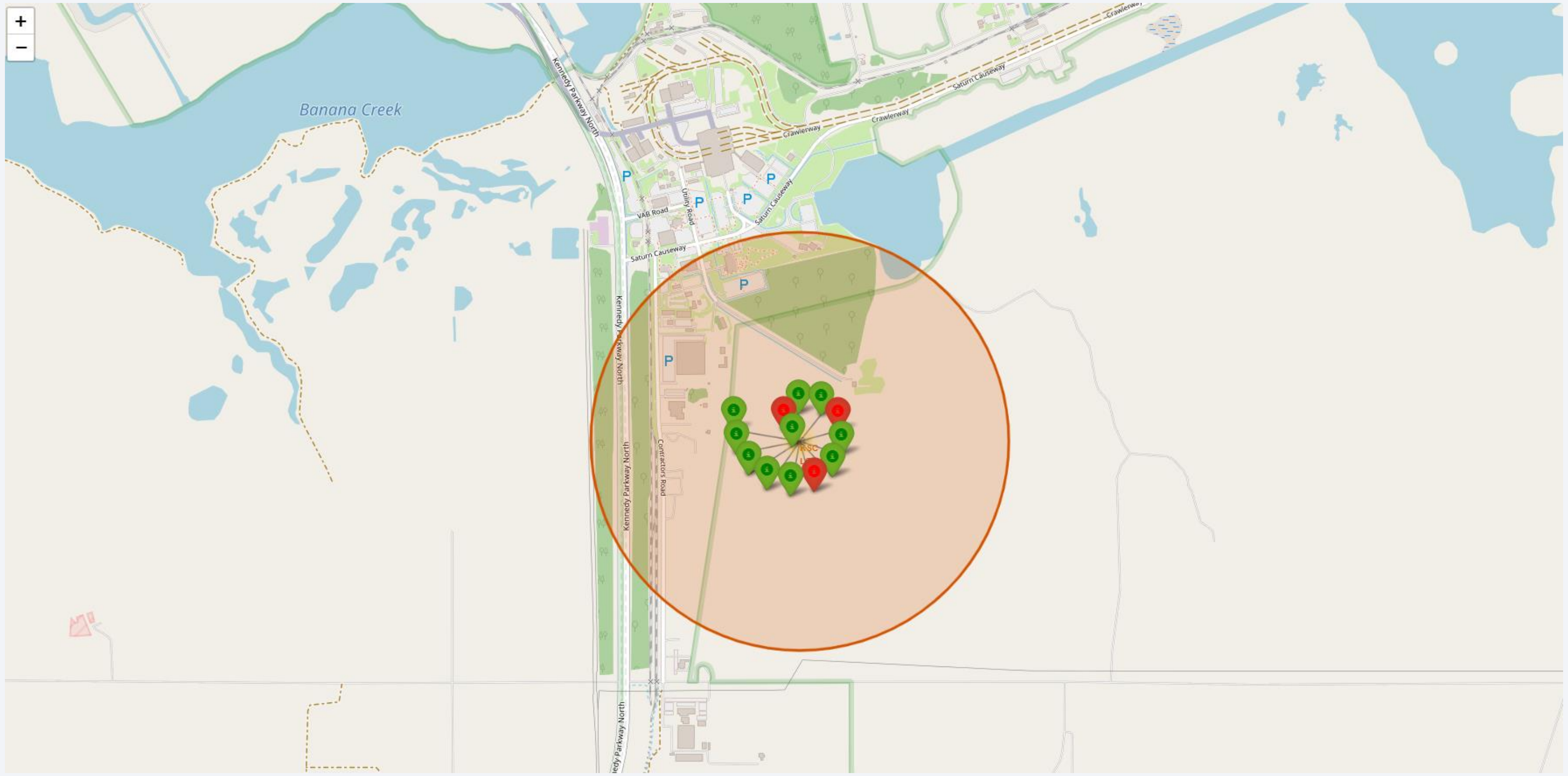
Launch Sites

- All launch sites' location markers on a global map



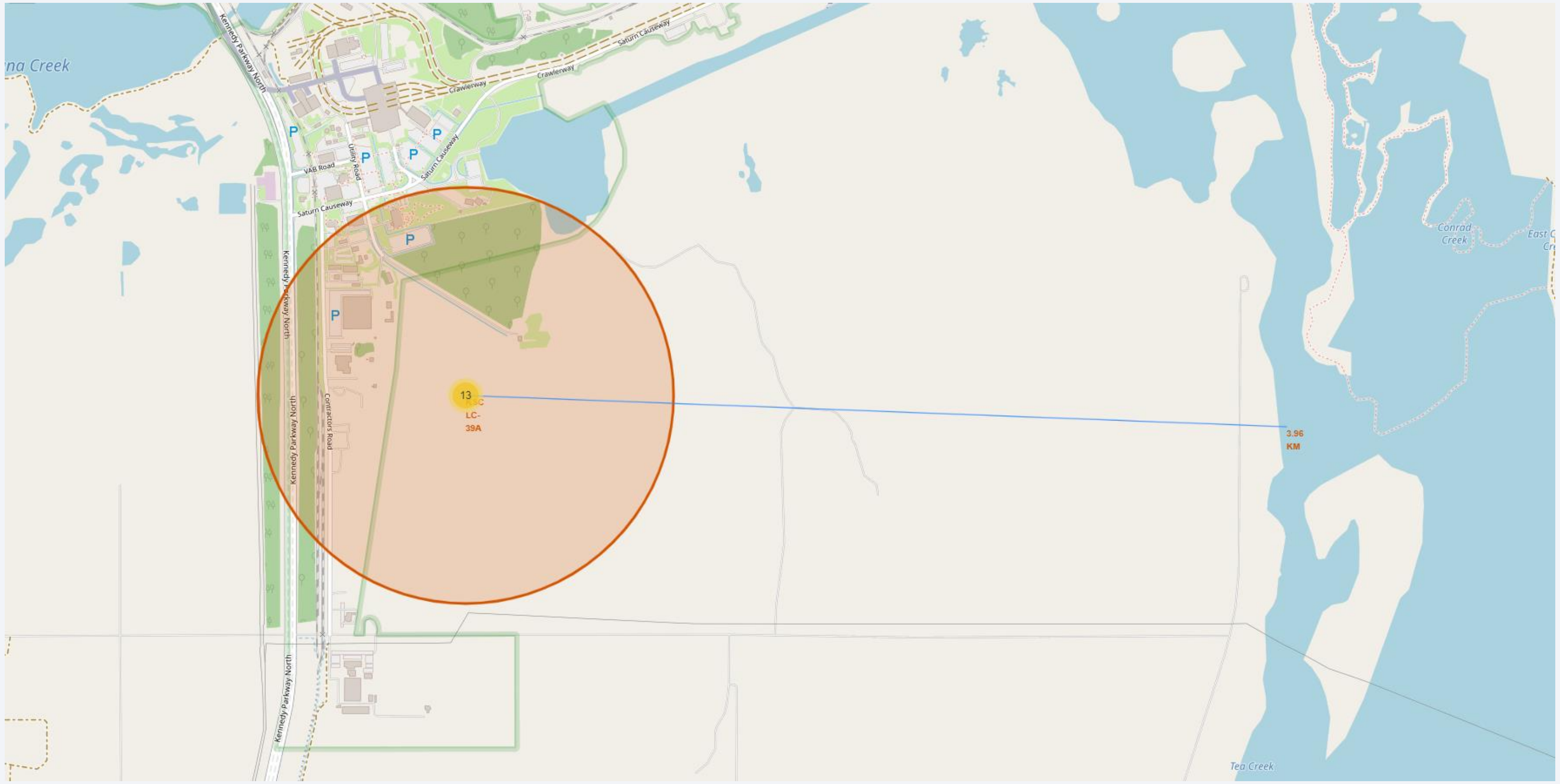
KSC LC-39A Launch Outcomes

- Color-labeled launch outcomes for KSC LC-39A



KSC LC-39A Distance to Coastline

- KSC LC-39A proximity to coastline, with distance calculated and displayed

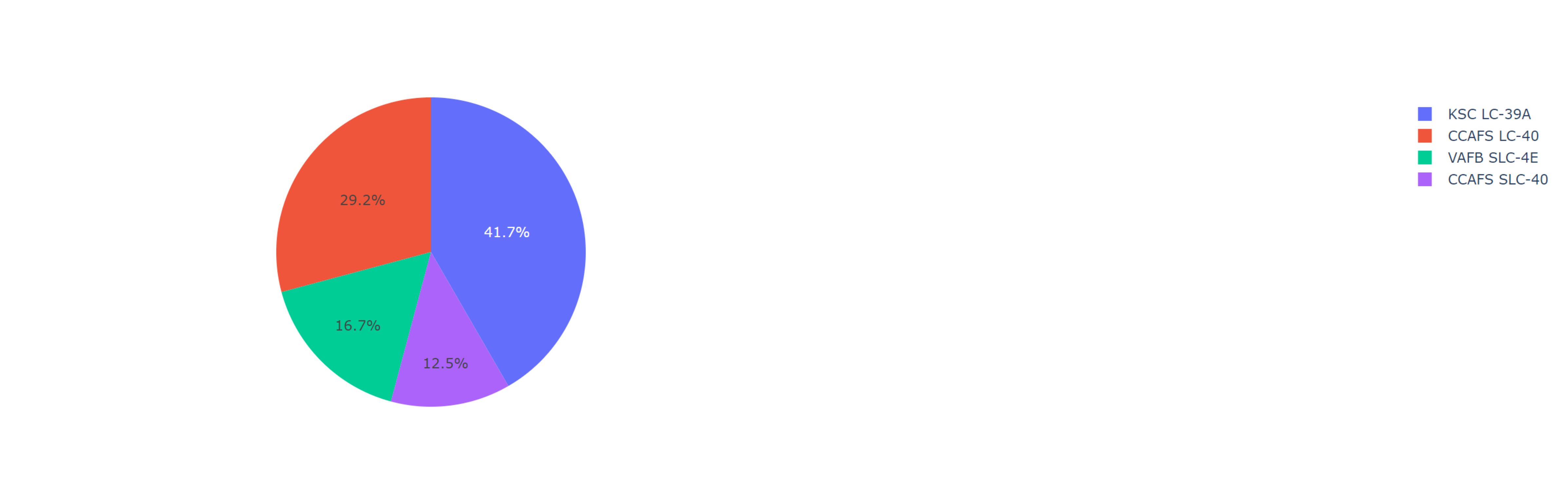


Section 4

Build a Dashboard with Plotly Dash

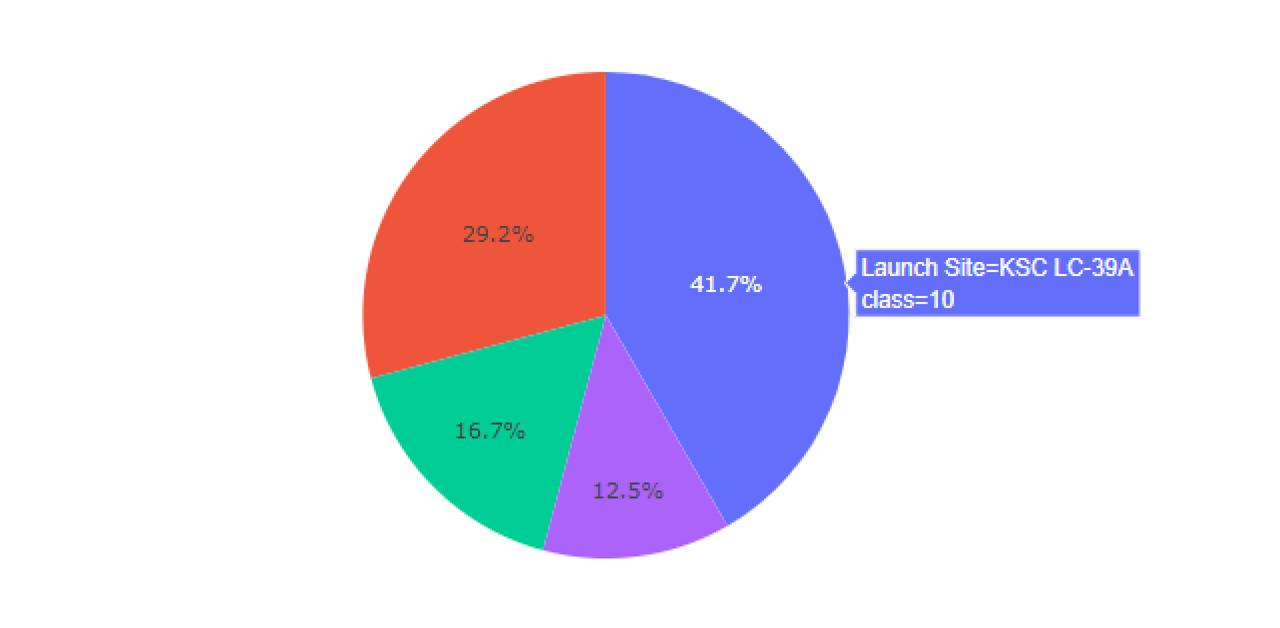
Launch Success by Site

- Launch success count for all sites



Launch Success by Site – KSC LC-39A

- Pie Chart for the launch site with highest launch success ratio



Payload vs. Launch Outcome

- Payload vs. Launch Outcome scatter plot for all sites, with payload > 2000 and < 8000 selected

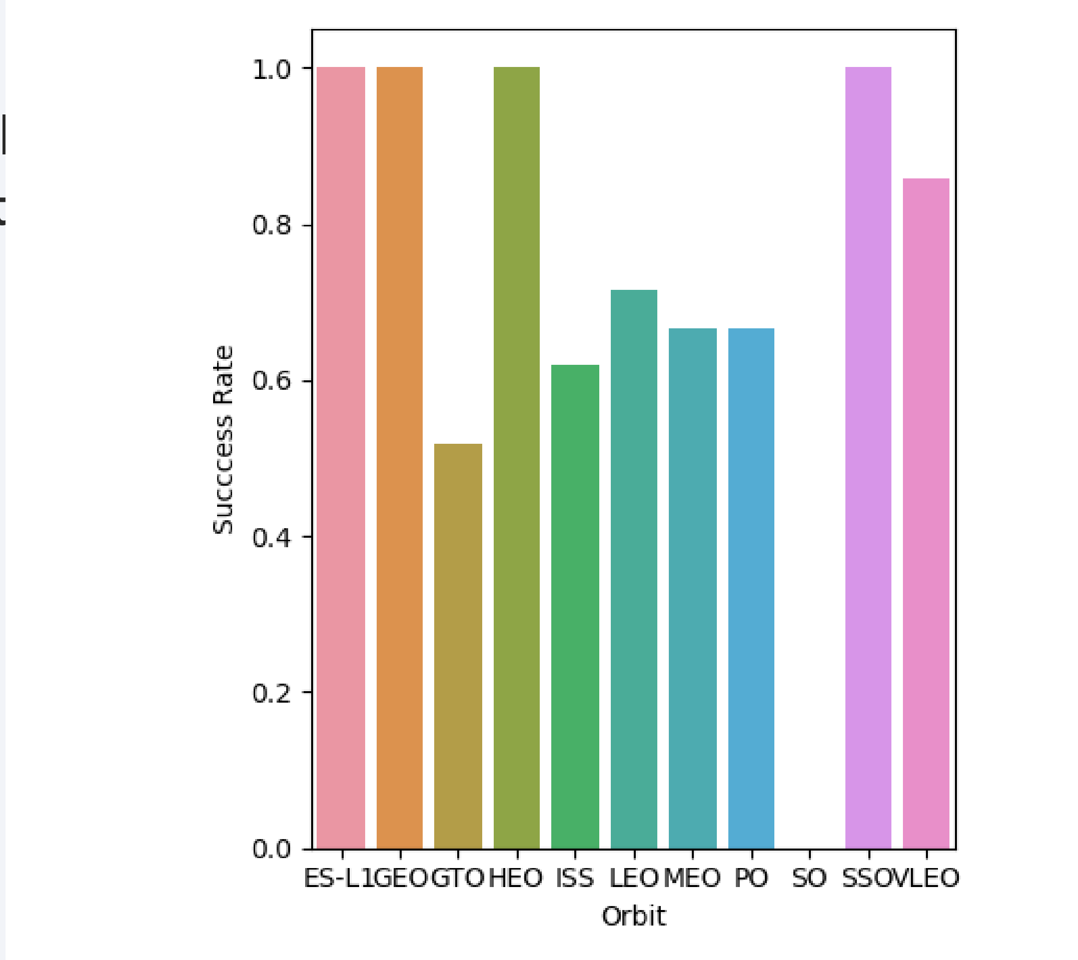
Section 5

Predictive Analysis (Classification)

Classification Accuracy

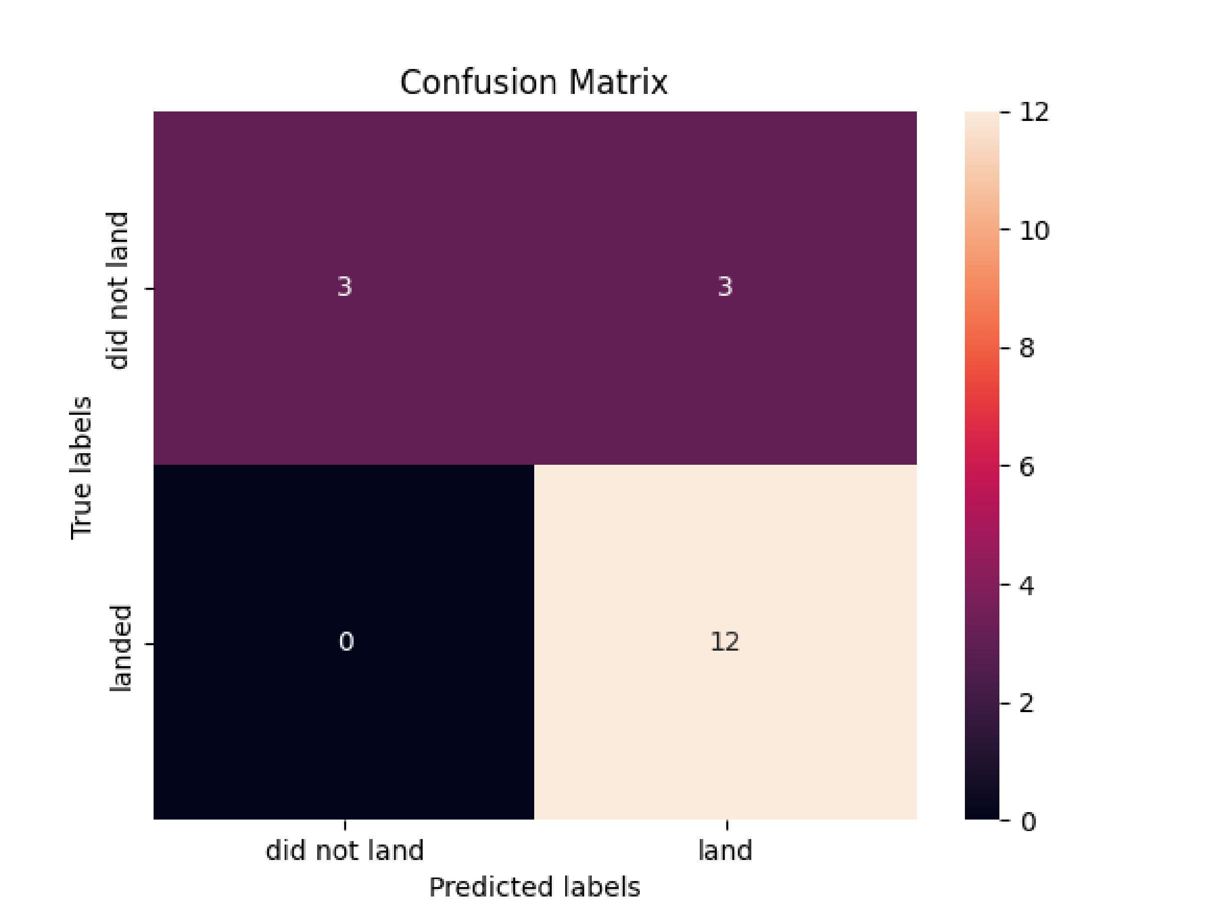
- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy



Confusion Matrix

- Confusion Matrix of the Decision Tree Model



Conclusions

- All models performed the same on the test data
- The model with the highest accuracy on training data was the decision tree model

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!