IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Fajar Tawfik
April 27, 2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection API – SpaceX launch data
  - Data Wrangling using API
  - Exploratory Analysis Using SQL
  - Exploratory Data Analysis for Data Visualization
  - Interactive Visual Analytics and Dashboard with Folium and plotly dash
  - Predictive Analysis with machine learning

- Summary of all results
  - Exploratory Data Analysis
  - Interactive Analytics
  - Predictive Analytics Results

Introduction

- Project background and context

  We aim to predict whether the Falcon 9 first stage will land successfully. SpaceX advertises its Falcon 9 rocket launches at a cost of $62 million. In contrast, other providers charge upwards of $165 million per launch. A significant portion of SpaceX's savings comes from the ability to reuse the first stage of the rocket. By accurately predicting the success of the first stage landing, we can estimate the cost of a launch. This prediction model could be valuable for other companies aiming to bid against SpaceX for rocket launches, providing insights into potential cost savings and competitive pricing.

- Problems you want to find answers
  - Prediction of Successful First Stage Landing?
  - What factors influence a successful landing?
  - The interactions between features that influence success rate?

Section 1

Methodology

Methodology

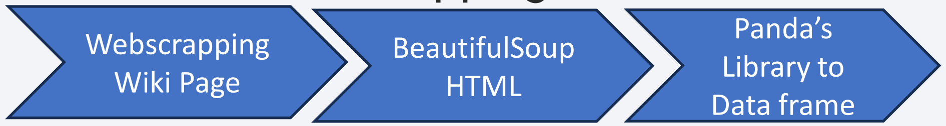
Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX rest API and web-scrapping from SpaceX launch Wiki
- Perform data wrangling
  - One-hot encoding (0 or 1) was applied on features of the dataset
- 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

- Describe how data sets were collected.
  - Data Was collected using SpaceX API and web scrapping techniques
  - Requested rocked launch data using Get request.
  - Filtered data to only include falcon 9 launches
  - Replaced missing values in payload mass with the mean
  - Used beautiful soup library to parse Wiki page containing SpaceX launch data
  - Extracted HTML Launch data tables and converted it to data frames using Panda’s library
- Flowcharts
  - API Call

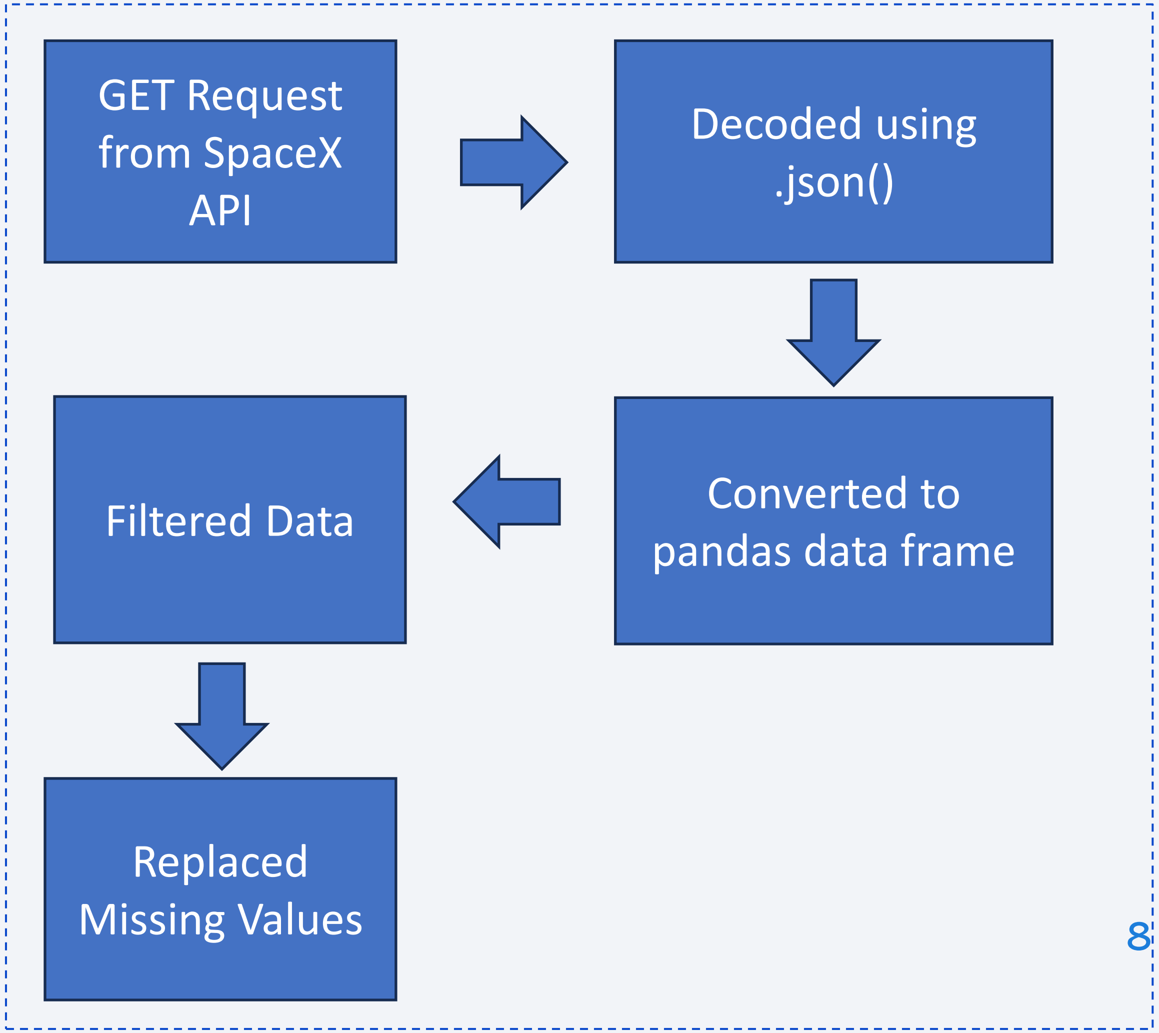
Webscrapping



Data Collection – SpaceX API

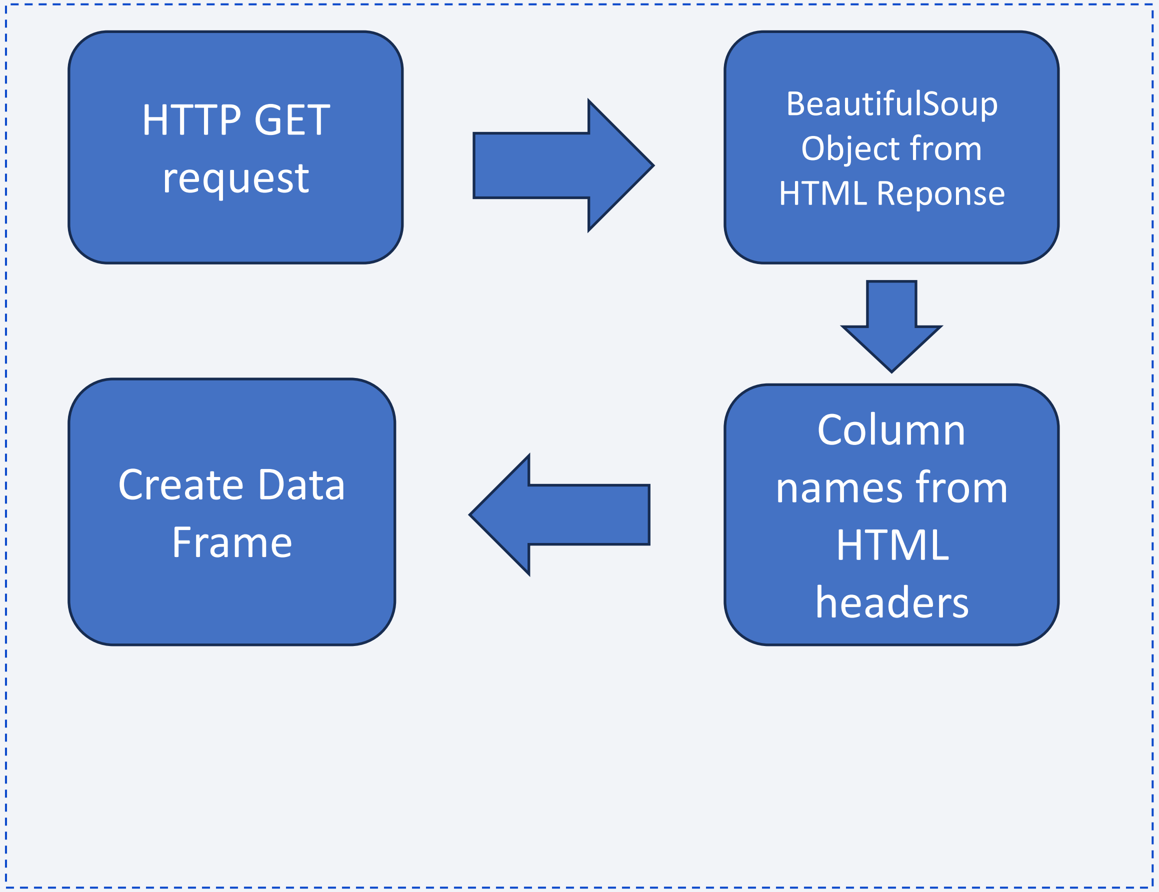
- Made a GET request to SpaceX Rest API, decoded using.
- Json() then converted to Pandas data frame using .json_normalize().
- Filtered Data to use only Falcon 9 Launches.
- Replaced missing values for payloadmass using it's mean

- GitHub URL of the completed SpaceX API calls notebook: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/spacex-data-collection-api.ipynb



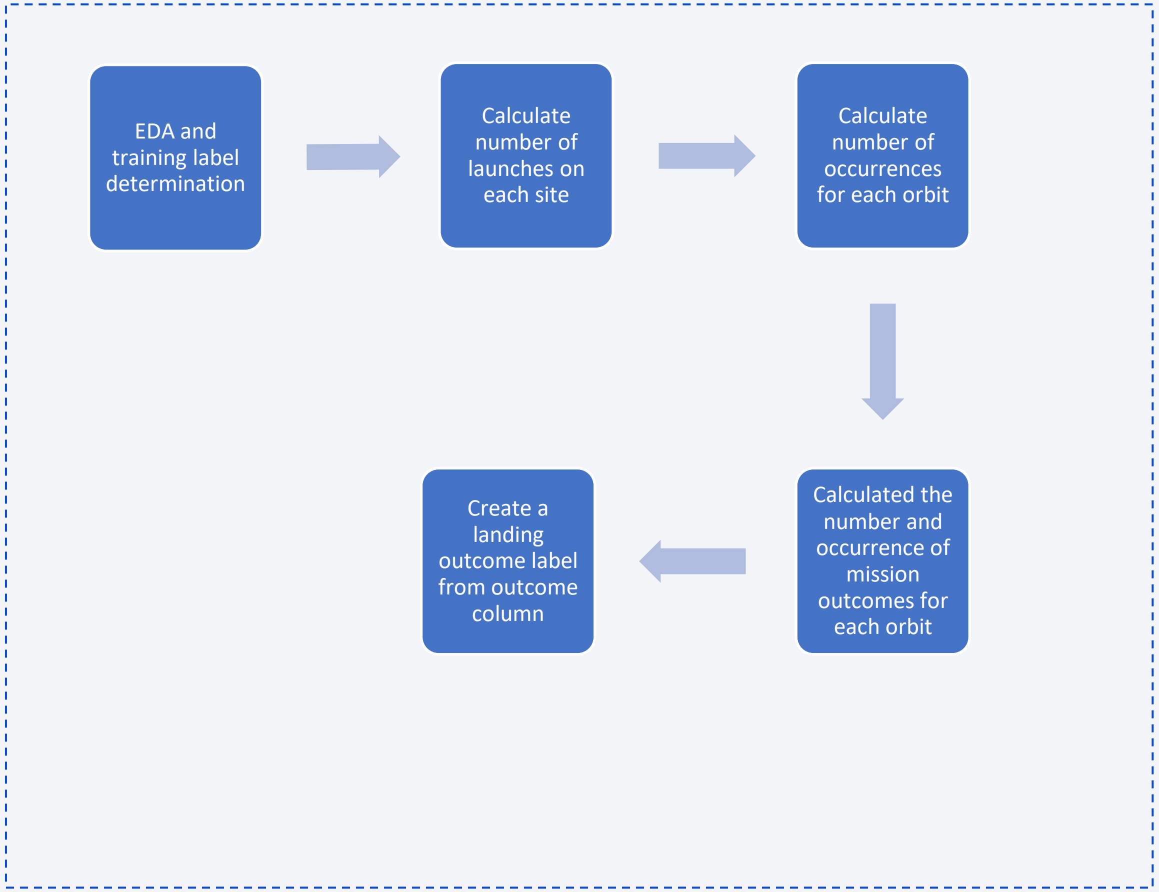
Data Collection - Scraping

- Used HTTP GET method to request Falcon9 Launch
- Created BeautifulSoup object from HTML Response
- Extracted column/variable names from HTML table header
- Created a data frame by parsing the launch HTML tables
- GitHub URL of the completed web scraping notebook: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/webscraping.ipynb



Data Wrangling

- Exploratory Data Analysis (EDA) to find patterns in the data and label for training supervised models
- Calculated the number of launches on each site, occurrences of each orbit, occurrence of mission outcome of the orbits, and landing outcome label from outcome column
- GitHub URL of your completed data wrangling related notebooks: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Relationships between Flight Number & Launch Site, Payload & Launch Site, Success rate of orbit type, Payload & Orbit type, and Launch Success rate Yearly were Visualized to find patterns and correlations
- Add the GitHub URL of your completed EDA with data visualization notebook:https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/edadataviz.ipynb

EDA with SQL

- The following SQL queries were preformed:
  - Names of unique launch sites
  - Displayed 5 records where launch sites began with the string ‘CCA’
  - Displayed the total payload mass carried by boosters launch by NASA (CRS)
  - Displayed average payload mass carried by boosters version F9 v1.1
  - Displayed average payload mass by booster version F9 v1.1
  - Listed date of first successful landing outcome in ground pad
  - Listed names of boosters which have success in drone and with a payload mass greater than 4000 but less than 6000
  - Listed the total number of successful and failure mission outcomes
  - Listed booster version that have carried the maximum payload mass
  - Listetd records that displayed month booster version, landing outcome, and launch site
  - Ranked the landing oucome between two dates

- Add the GitHub URL of your completed EDA with SQL notebook: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We marked all launch sites, and added map objects such as markers, circles, lines to mark the success or failure of launches for each site on the folium map.
- We assigned the feature launch outcomes (failure or success) to class 0 and 1.i.e., 0 for failure, and 1 for success.
- Using the color-labeled marker clusters, we identified which launch sites have relatively high success rate.
- We calculated the distances between a launch site to its proximities. We answered some question for instance:
  - Are launch sites near railways, highways and coastlines.
  - Do launch sites keep certain distance away from cities.
- Add the GitHub URL of your completed interactive map with Folium map: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash
- We plotted pie charts showing the total launches by a certain sites
- We plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.
- GitHub URL of completed Plotly Dash lab: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

- Loaded the data using numpy and pandas, transformed the data, split our data into training and testing.
- Built different machine learning models and tune different hyperparameters using GridSearchCV.
- Used accuracy as the metric for our model, improved the model using feature engineering and algorithm tuning.
- Found the best performing classification model.
- GitHub URL of completed predictive analysis lab: https://github.com/FajarTawfik/Capstone-SpaceX-Falcon-9-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

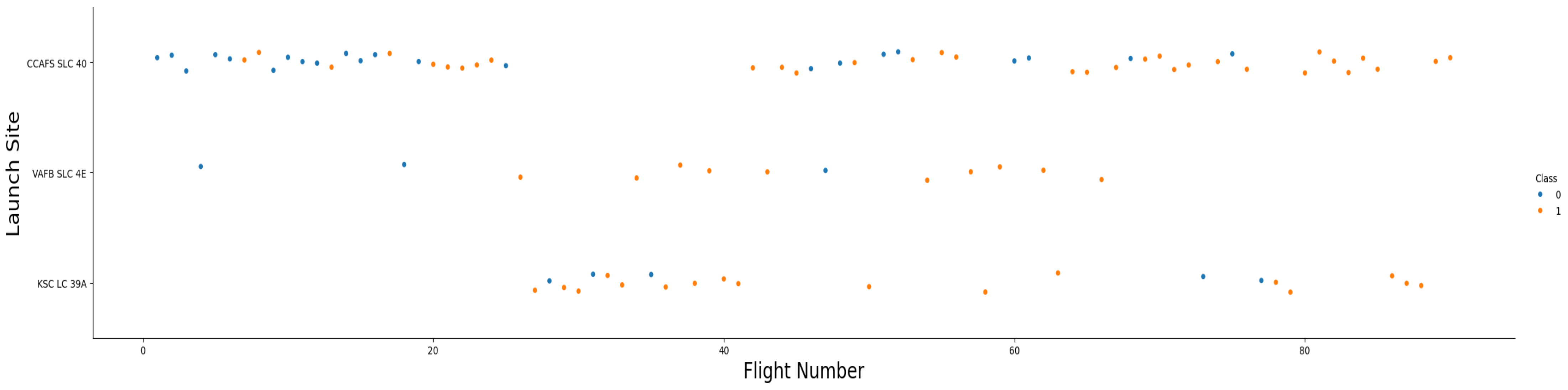
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

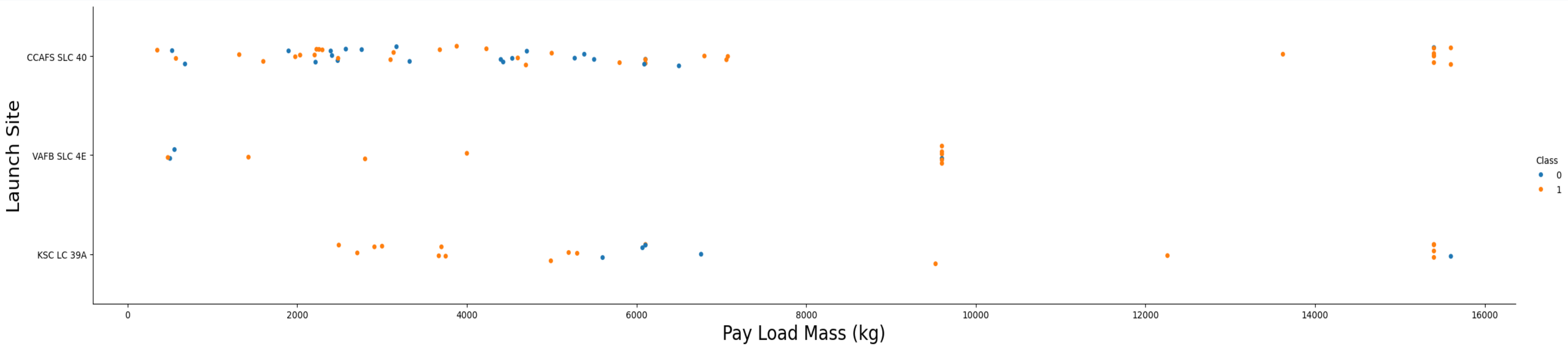
Flight Number vs. Launch Site

- A scatter plot of Flight Number vs. Launch Site shows a direct correlation between flight number and launch sites
  - Higher Flight number correlates to higher success rate



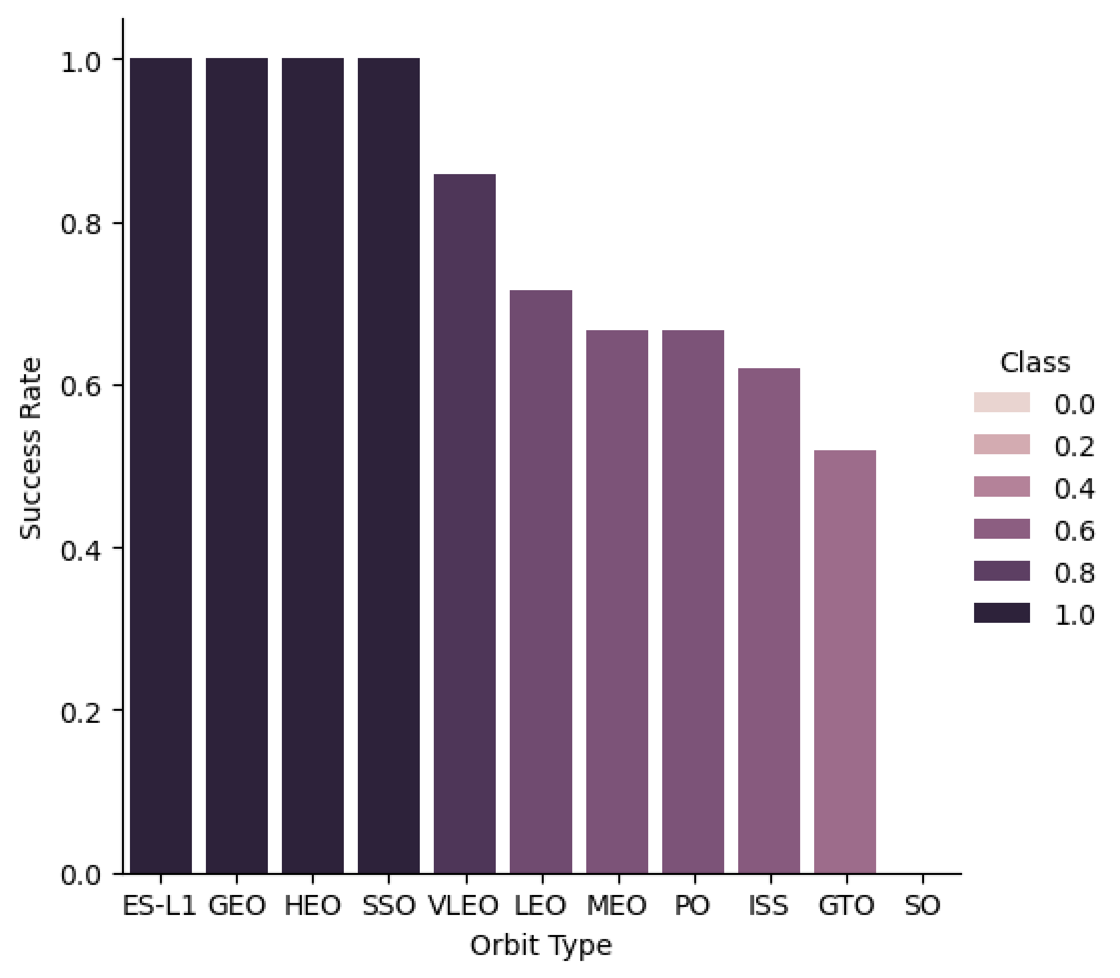
Payload vs. Launch Site

- Scatter plot of Payload vs Launch Site shows the success rate is higher when the mass is below 5000 and above 7000 for KSC LC 39A



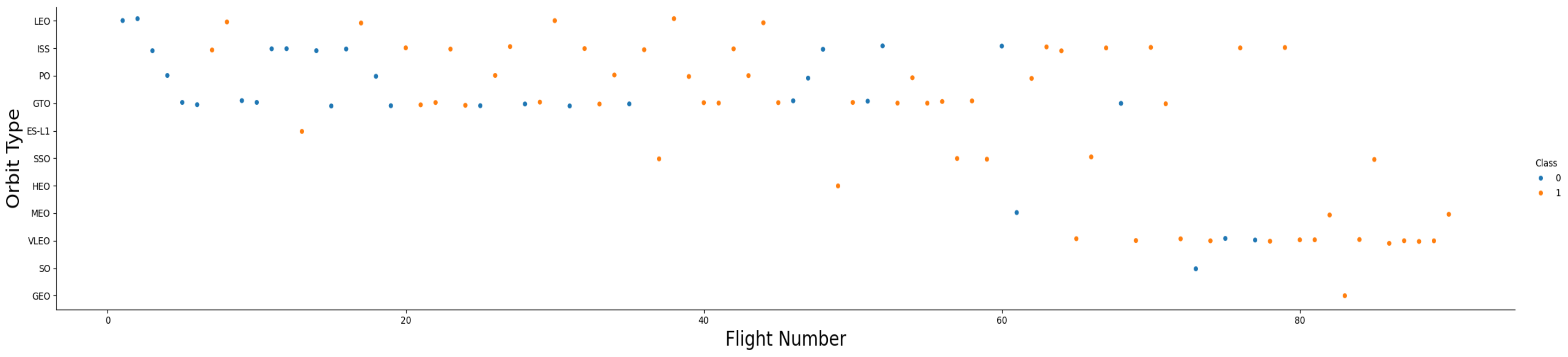
Success Rate vs. Orbit Type

- The bar chart shows the ES-L1 has the highest success rate and GTO has the lowest of approximately 0.6



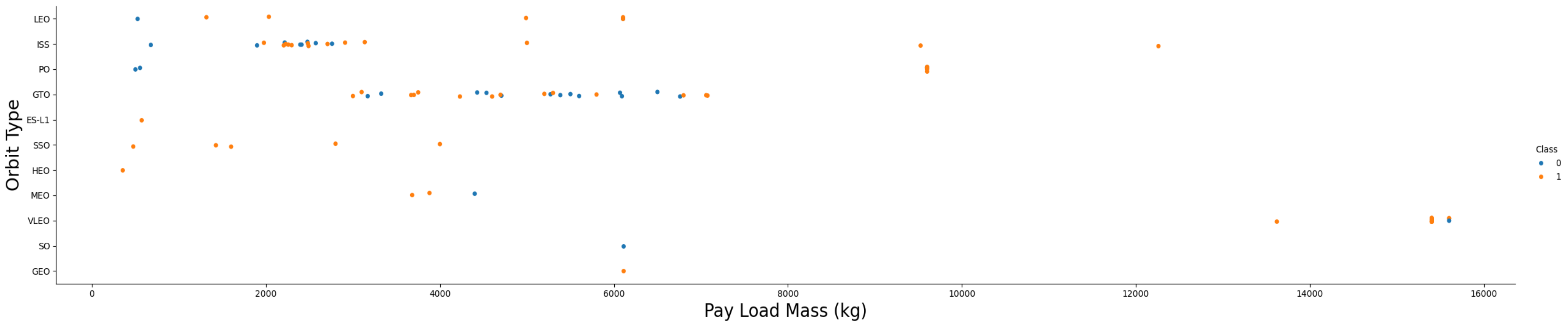
Flight Number vs. Orbit Type

- The scatter plot shows that in the LEO orbit success seems to be related to the number of flights however in GTO orbit there appears to be no relationship between flight number and success



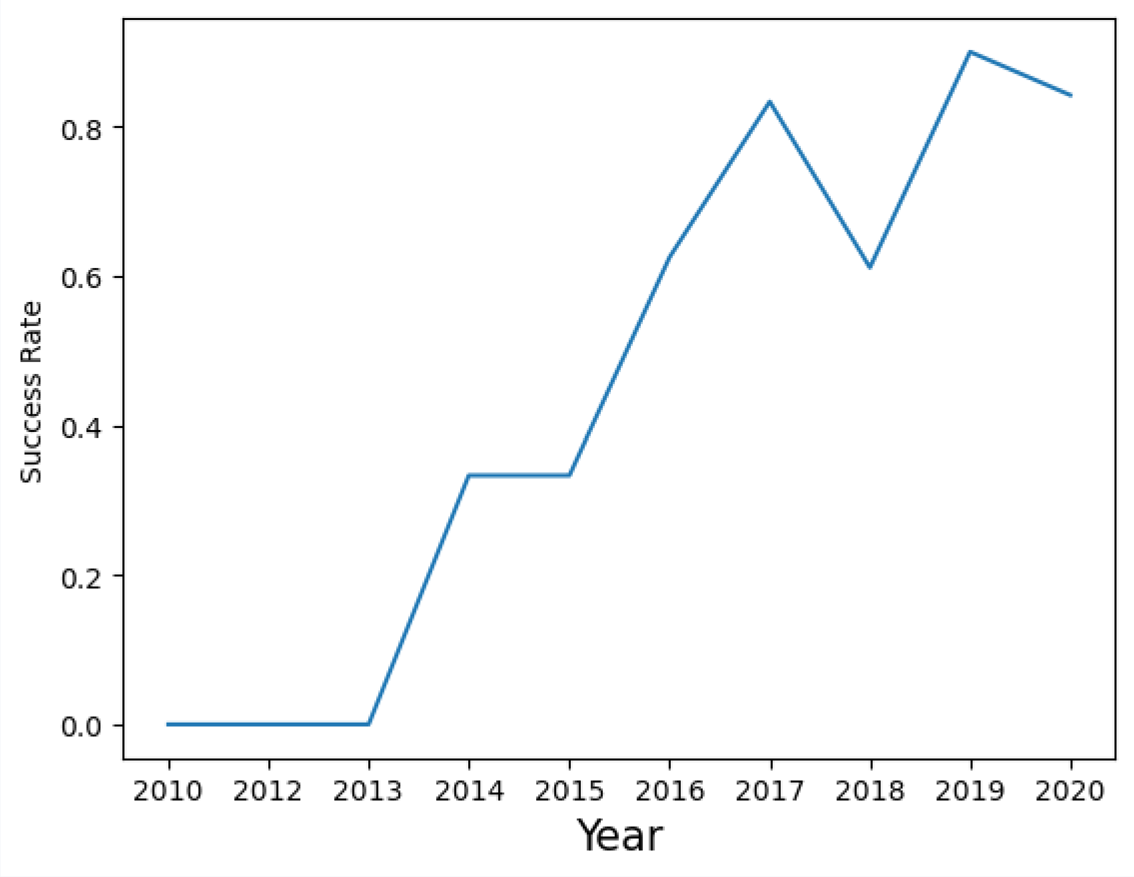
Payload vs. Orbit Type

- The figure shows with heavy payload 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

- The following line chart shows that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- The unique launch sites in the data set are as follows:

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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

- Calculate the total payload carried by boosters from NASA
- The total payload mass was 45596

Total payload Mass By Nasa (CRS)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- The average payload mass was 2534.67

Average paylod mass by F9 v1.1
2534.6666666666665

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- The first Successful landing outcome on ground pad was June 4th, 2010 this was achieve by using the min fucntion

MIN(DATE)
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- The majority resulted in a successful mission outcome

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the 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

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

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

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	OUTCOME_COUNT	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

Section 3

Launch Sites Proximities Analysis

All Launch Site Global map Markers

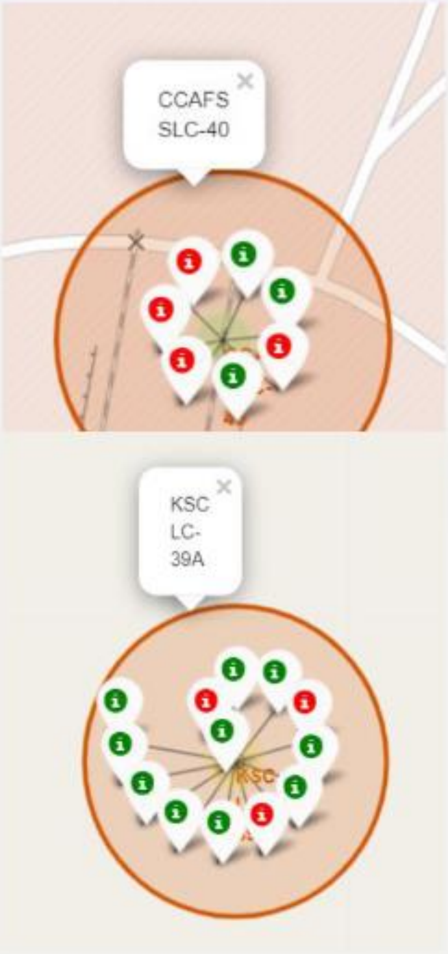
Launch sites markers

- Replace <Folium map screenshot 2> title with an appropriate title

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

- Explain the important elements and findings on the screenshot

Markers showing launch sites with color labels



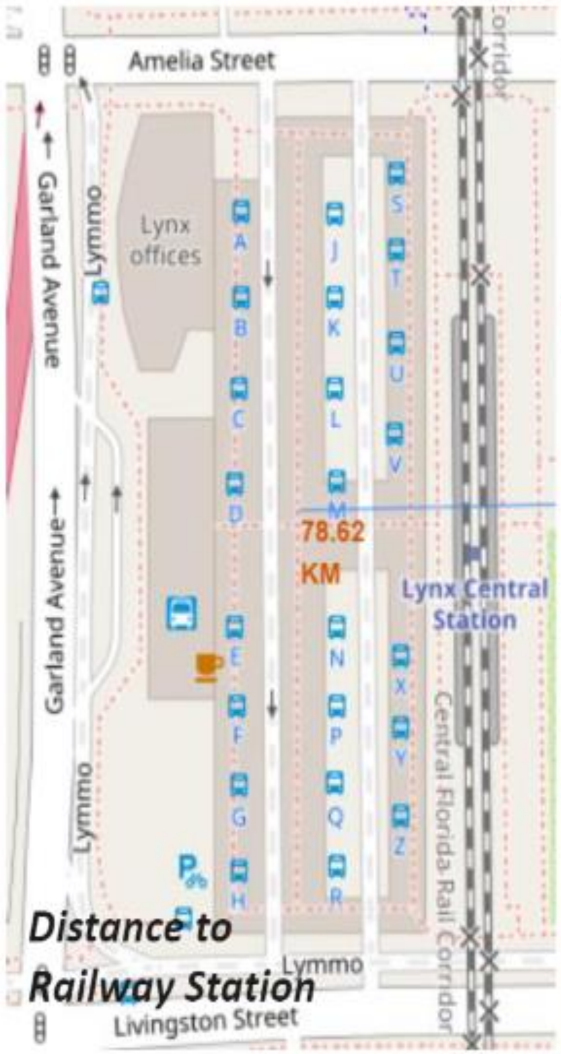
Florida Launch Sites

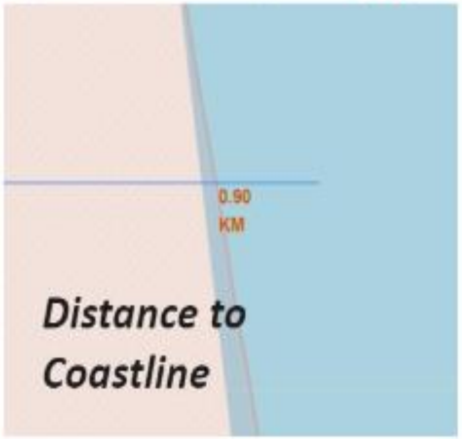
Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

37

Launch Site distance to landmarks



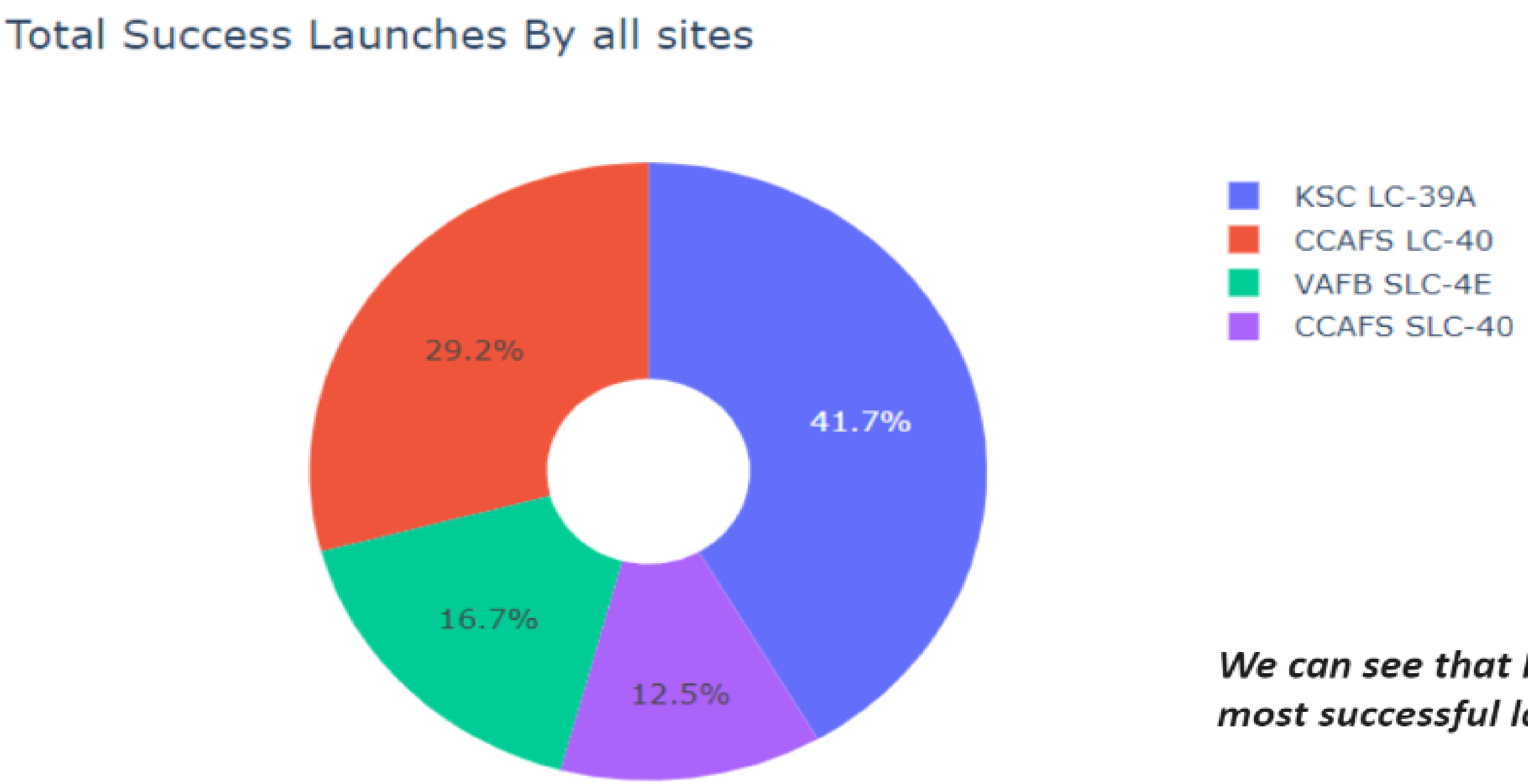


- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

Section 4

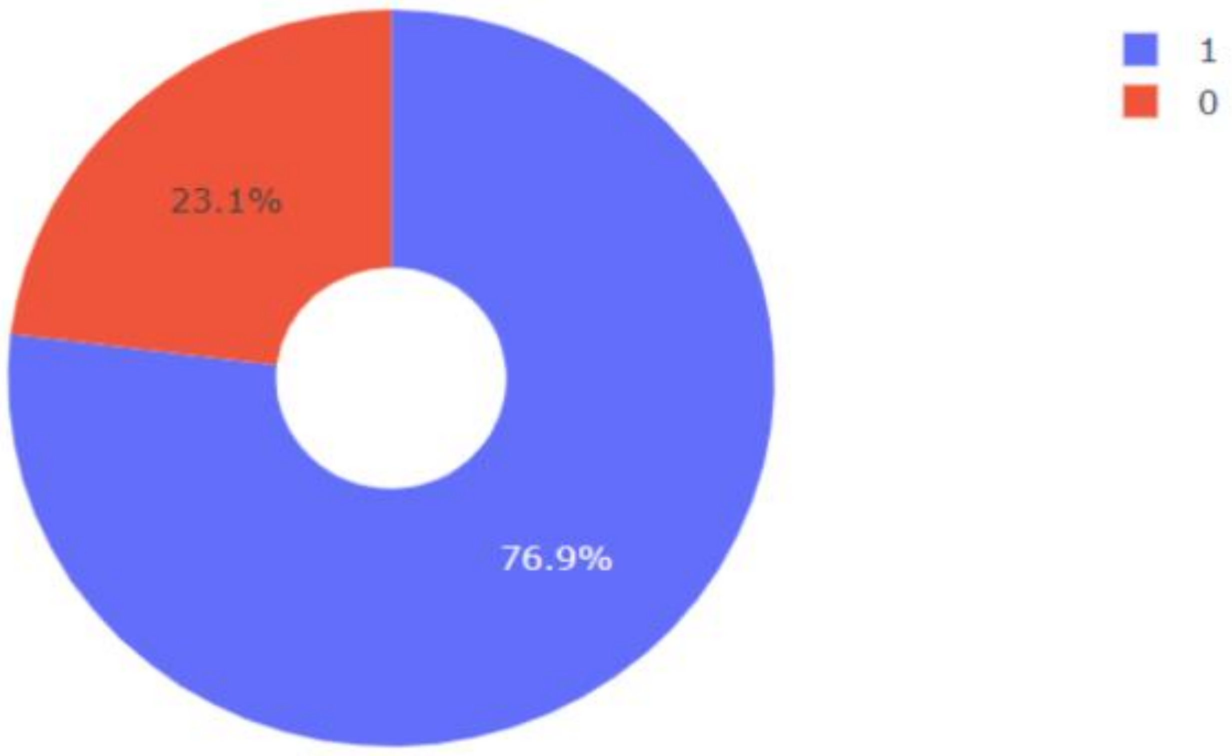
Build a Dashboard with Plotly Dash

Total Success Launch pie chart



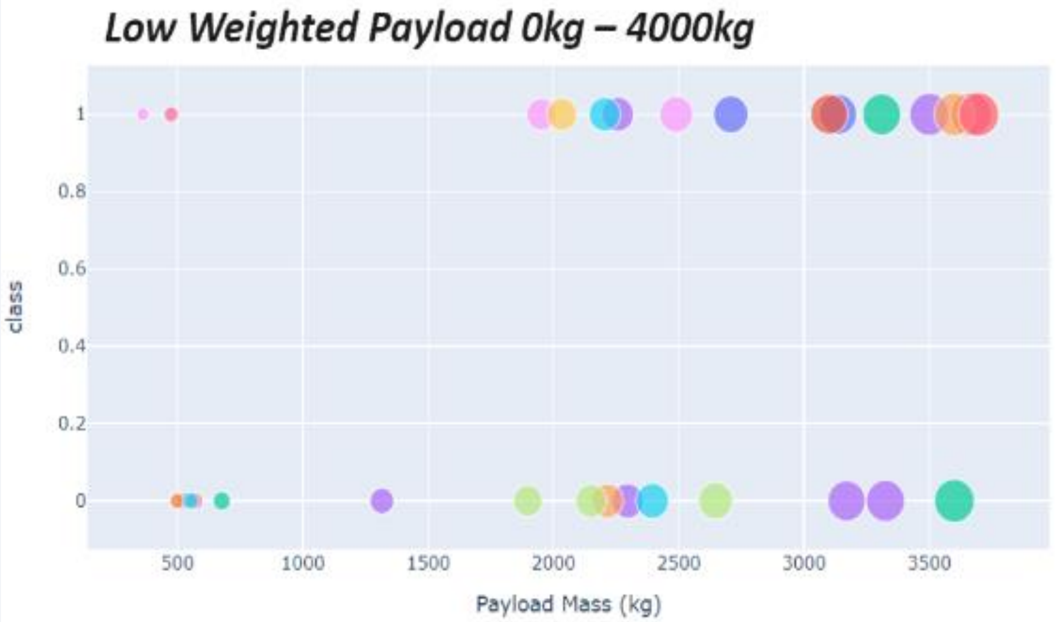
We can see that KSC LC-39A had the most successful launches from all the sites

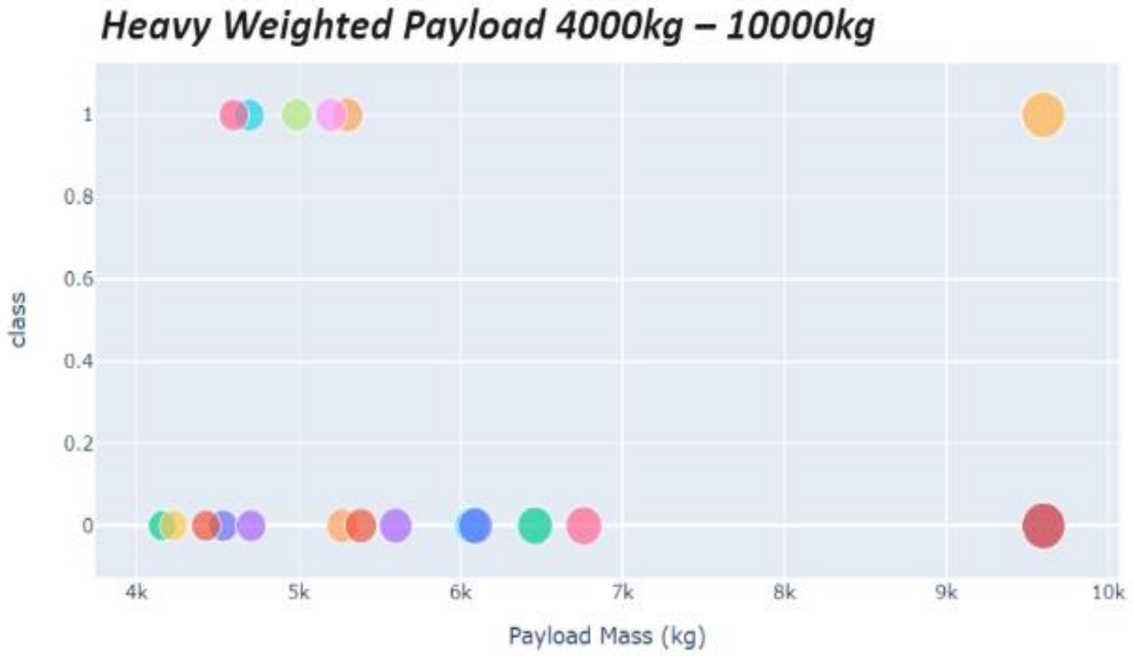
Highest Launch success ratio



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs Launch Outcome for all sites





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The decision tree classifier model had the highest classification accuracy of 0.873

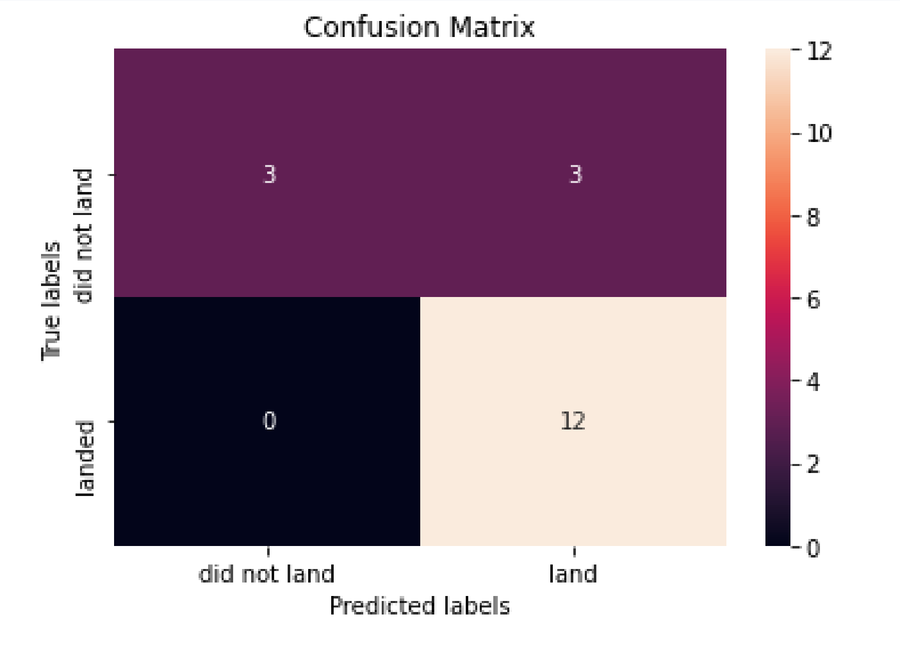
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

Confusion Matrix

- The confusion matrix for the decision tree classifier shows that the classifier can distinguish between the different classes. The major problem is the false positives .i.e., unsuccessful landing marked as successful landing by the classifier.



Conclusions

We can conclude that:

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

Thank you!