IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akash Kumar L
20/02/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive Analysis (Classification)

Summary of all results

- EDA results
- Interactive analytics
- Predictive analysis

Introduction

- Project background and context

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars. Other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers

The task is to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

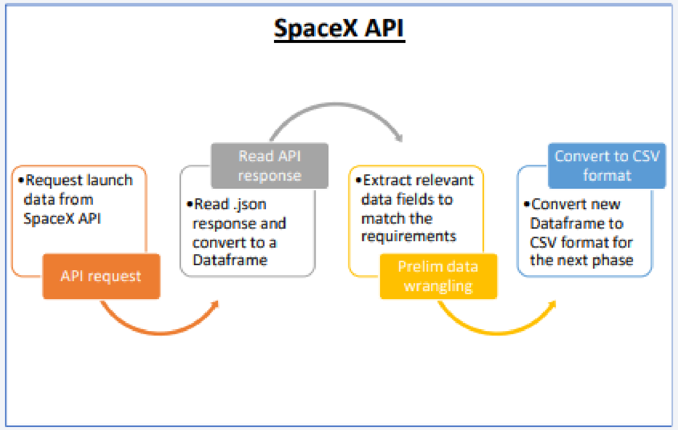
Methodology

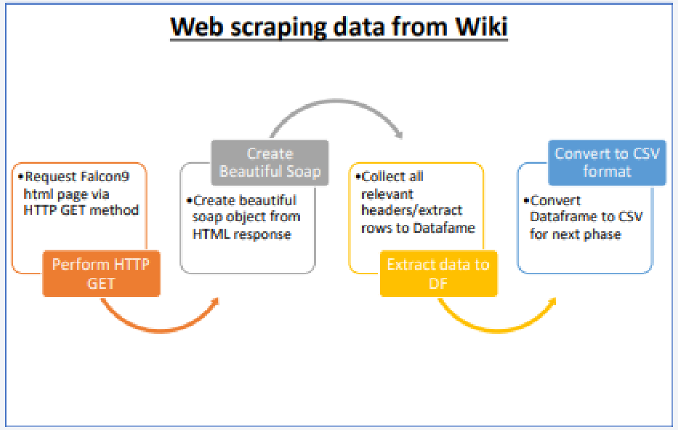
Executive Summary

- Data collection methodology:
    - Data was collected using SpaceX REST API, Web scraping from Wikipedia
- Perform data wrangling
    - Filled missing values with Mean, one-hot encoding data columns for ML
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Built, tuned and evaluated classification models

Data Collection

- SpaceX launch data is collected from SpaceX REST API
- The API includes data about launches, payload mass, landing outcomes etc
- The Falcon 9 launch data is collected from Wikipedia using BeautifulSoup.





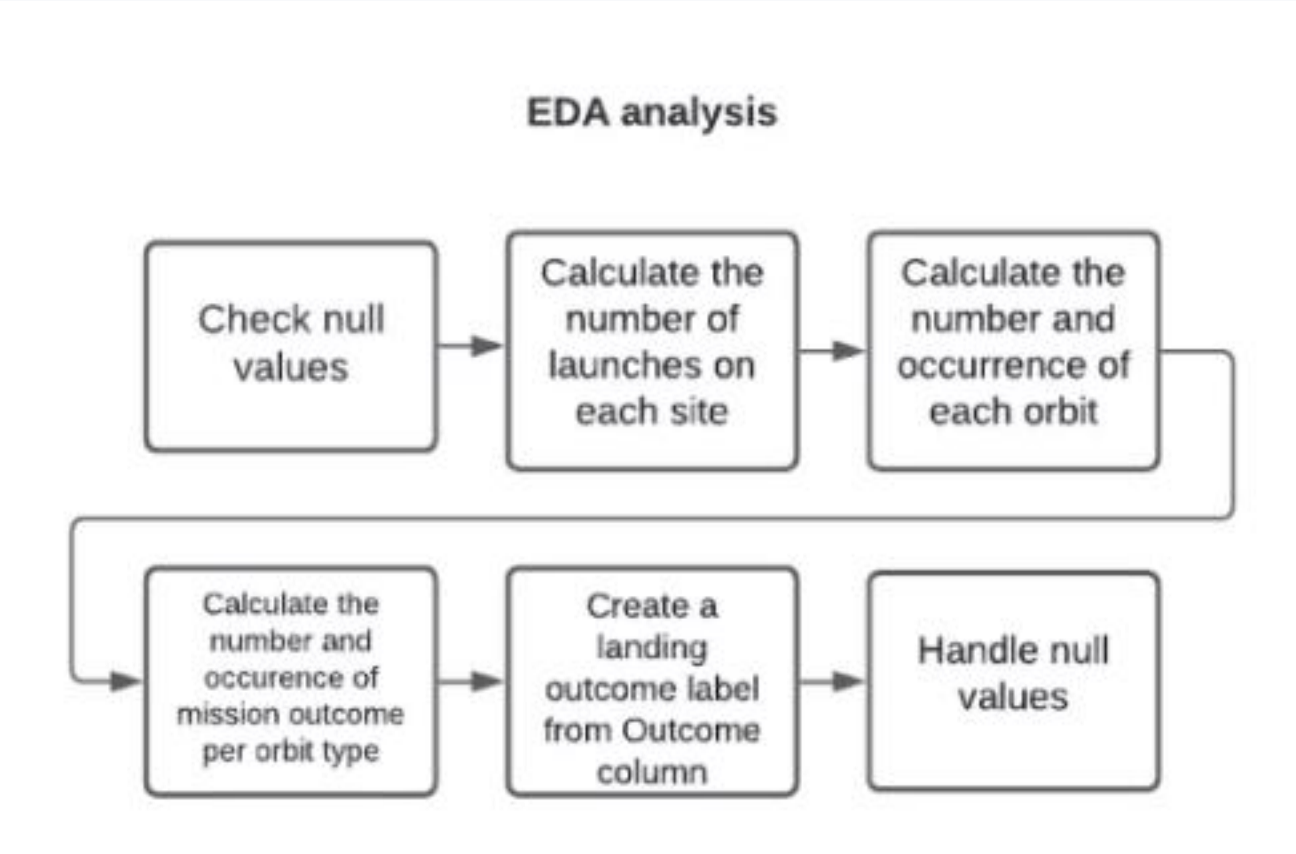
Data Collection – SpaceX API

- Create a API GET request, normalize and read it into a Dataframe
- Declare global variable lists to store data returned by functions
- Call helper functions to get relevant data
- Construct dataset and combine columns into a dictionary and convert it to a Dataframe
- Save the dataframe as a csv file
- Github Link

Data Collection - Scraping

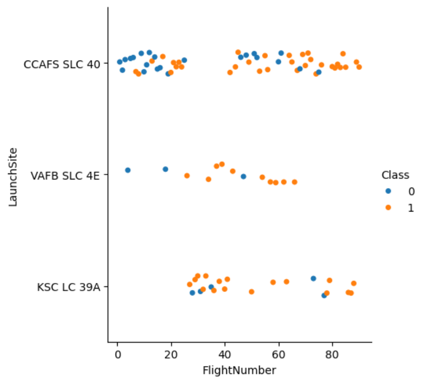
- Create an API GET method to request Falcon 9 Launch HTML page
- Using Beautiful Soup object, parse the html data and retrieve the table columns
- Extract the table rows and store it in a Dataframe
- Convert Dataframe to csv file
- Github Link

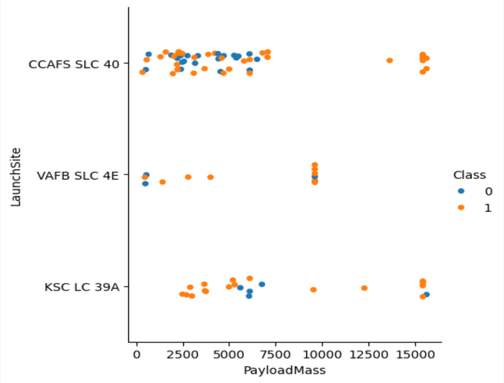
Data Wrangling

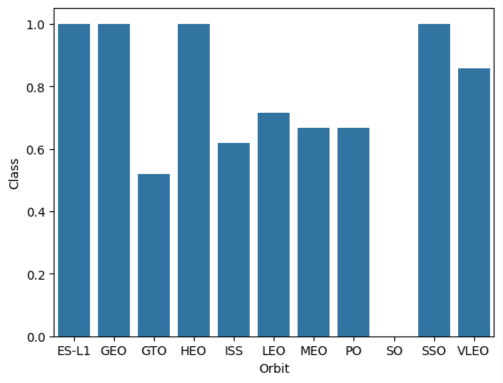


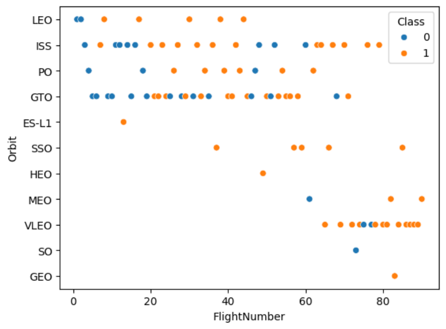
- [Github Link]

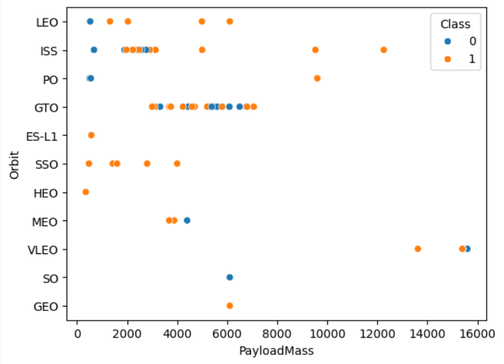
EDA with Data Visualization

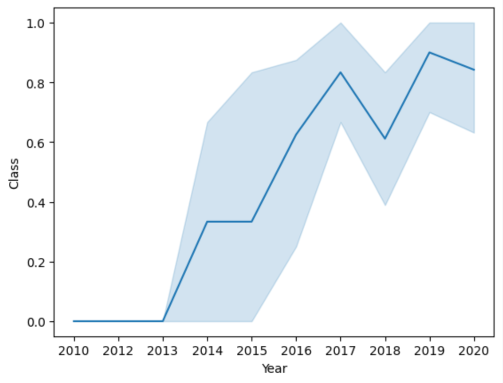












- Scatter plots, bar charts, line chart are used to understand relationship between different factors
- Github Link

EDA with SQL

- SQL Queries:

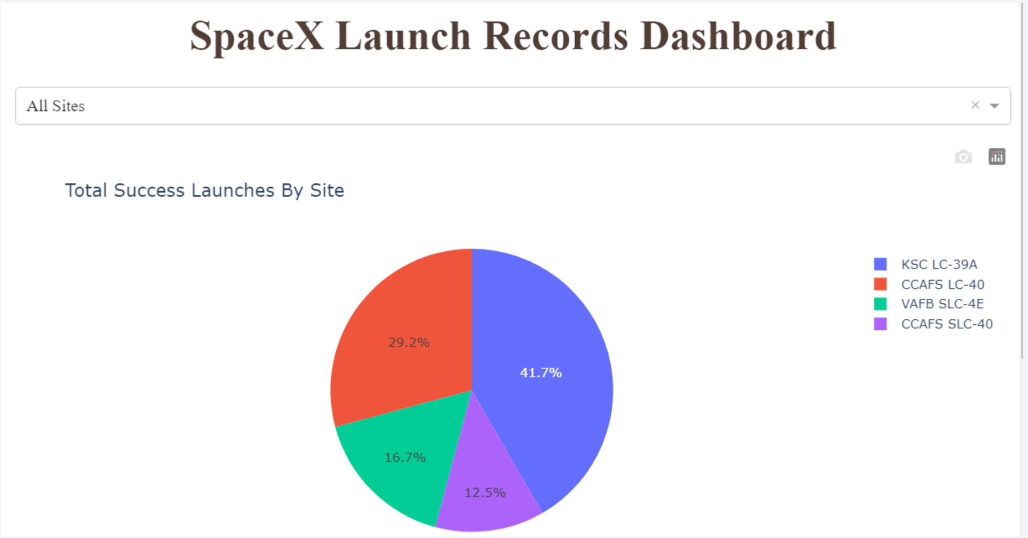
1. Display the name of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string ‘CCA’
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in ground pad was achieved
6. List the names of the boosters which have success in drone ship and payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster versions which have carried the maximum payload mass
9. List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the month in year 2015
10. 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

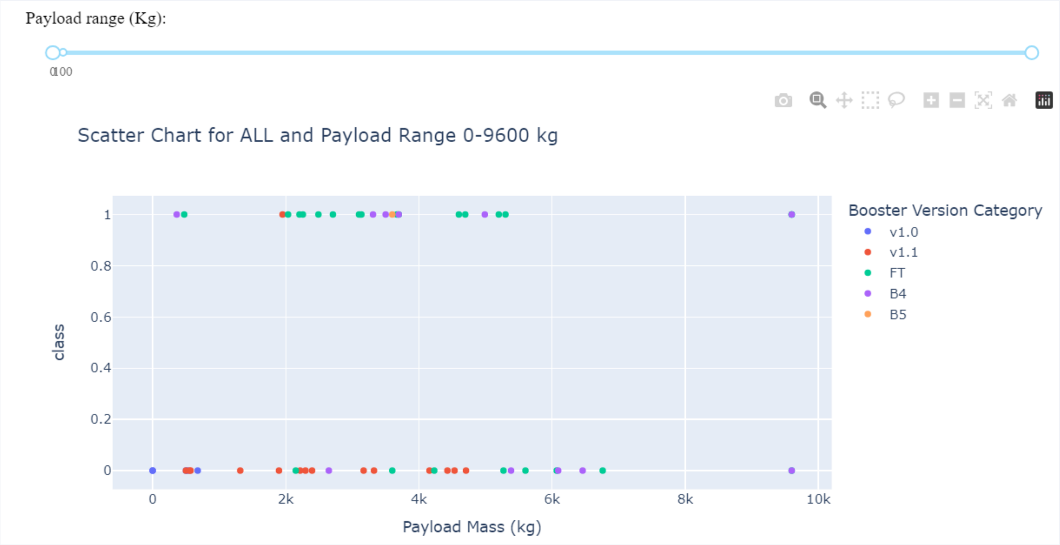
Github Link

Build an Interactive Map with Folium

- Marker clusters, poly lines are added to the map, in order to find the optimal location for building a launch site.
- Github Link

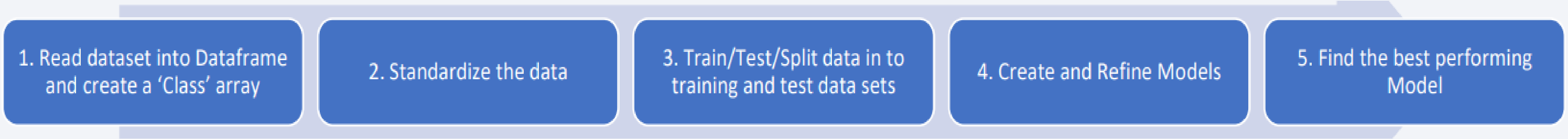
Build a Dashboard with Plotly Dash





- Used Pie chart and Scatter plot using Plotly Dash for interactive visualization
- From the pie chart, we could see that KSC LC-39A has the most successful launches. In scatter plot, it is evident that the success rate is higher for low payload mass range
- Github Link

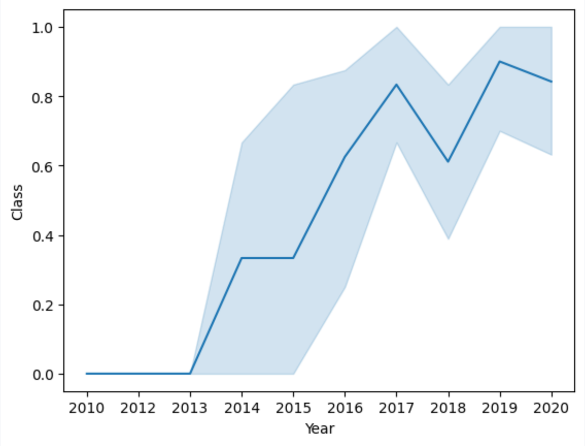
Predictive Analysis (Classification)



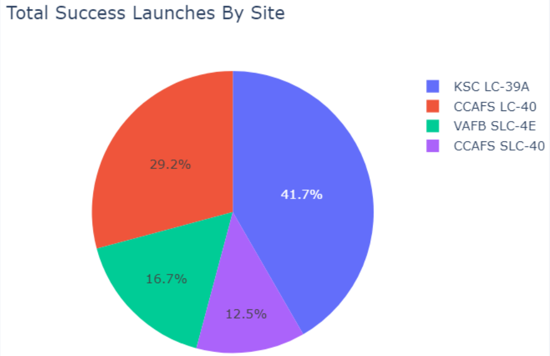
- The Logistic Regression and K-Nearest Neighbors model have achieved the highest classification accuracy compared to other models (Support Vector Machine, Decision Tree)
- In terms of overall metrics, KNN performs better than other models
- Github Link

Results

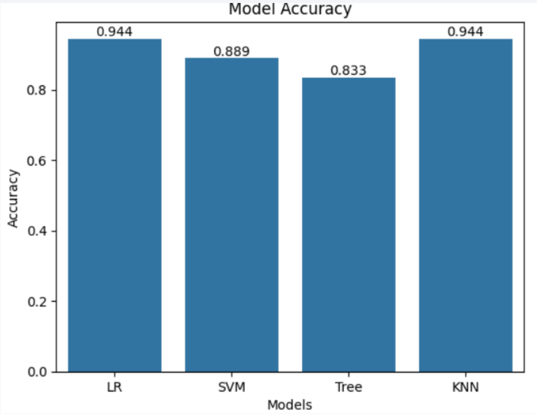
- Exploratory data analysis results:



- Interactive analytics results:



- Predictive analysis results:



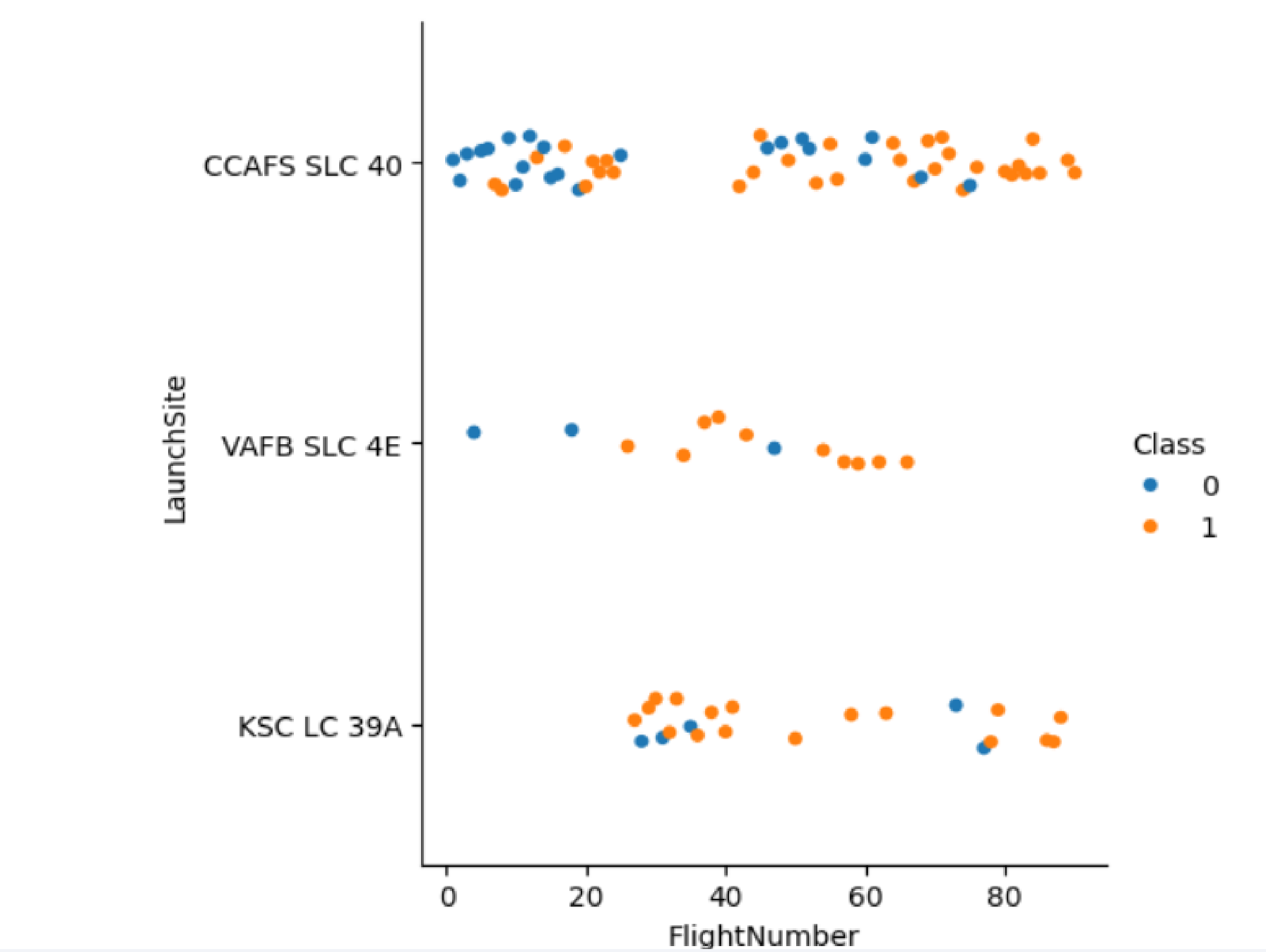
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

As the Flight number increases, the success rate is also increases.

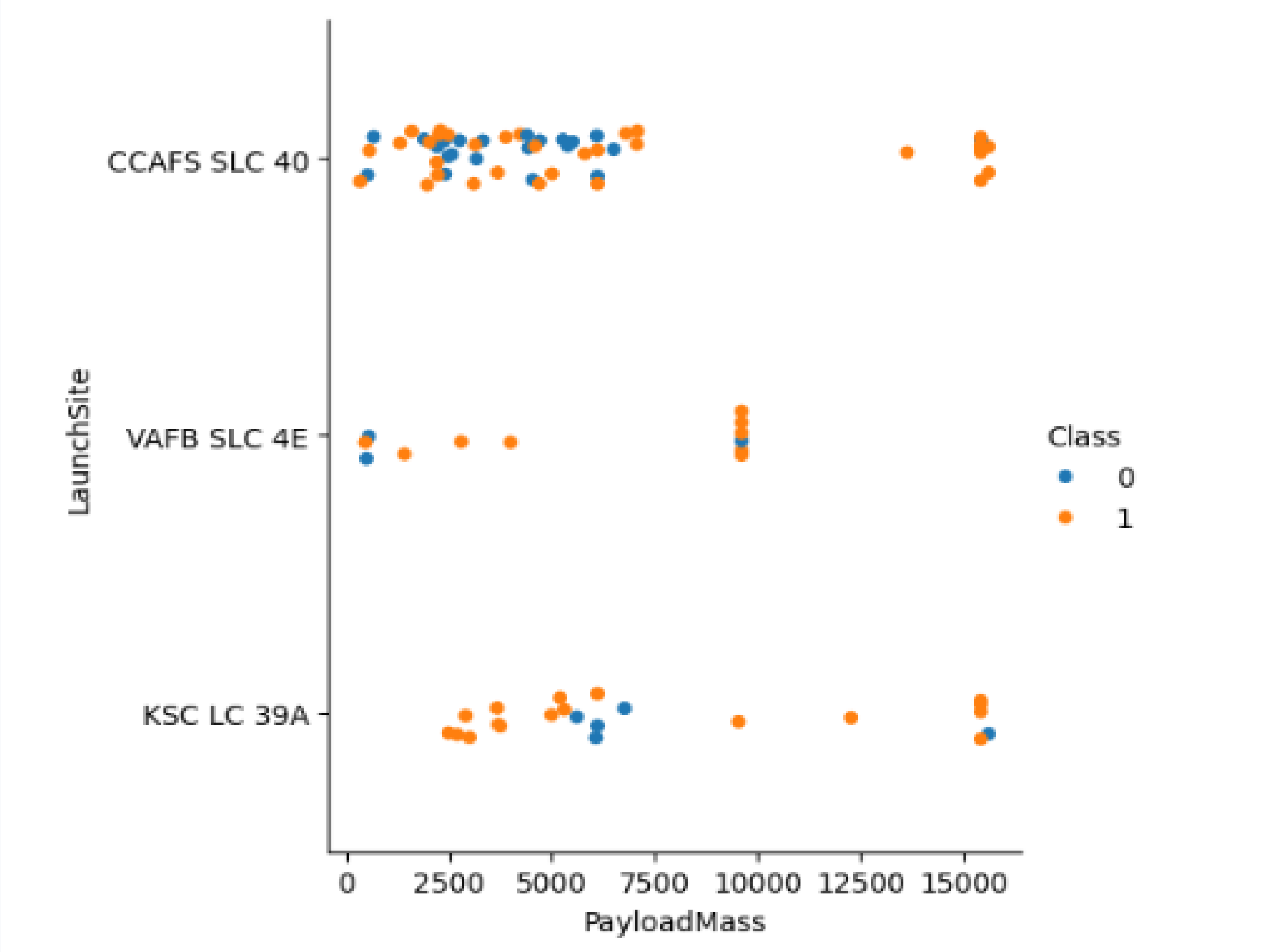
The VAFB SLC 4E & KSC LC 39A has the highest success rate.



Payload vs. Launch Site

For VAFB SLC 4E, there are no rockets launched with payload mass greater than 10,000 KG

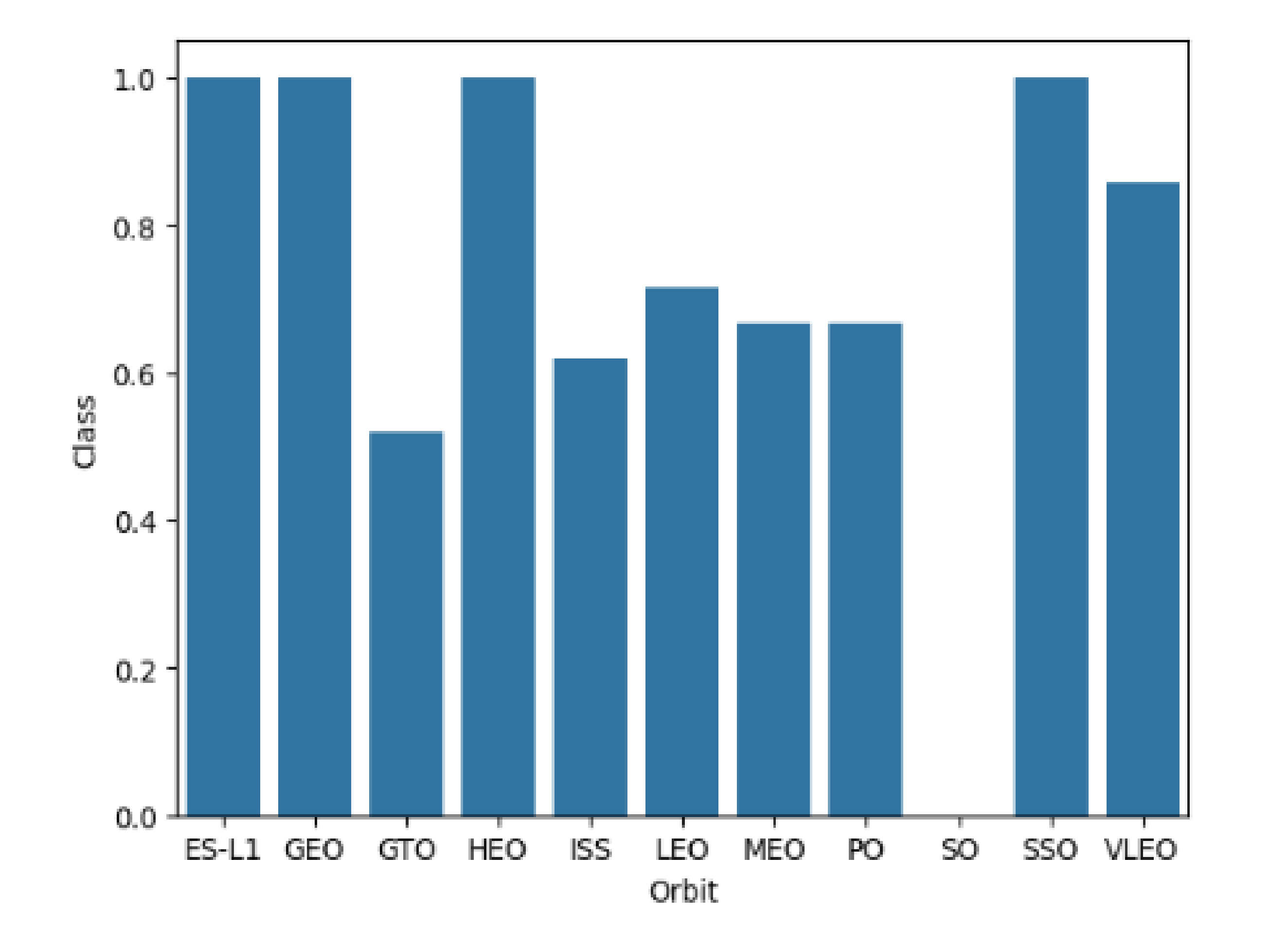
As the Payload Mass increases, there is an increase in success rate



Success Rate vs. Orbit Type

The ES-L1, GEO, HEO, SSO orbit types have 100% success rate

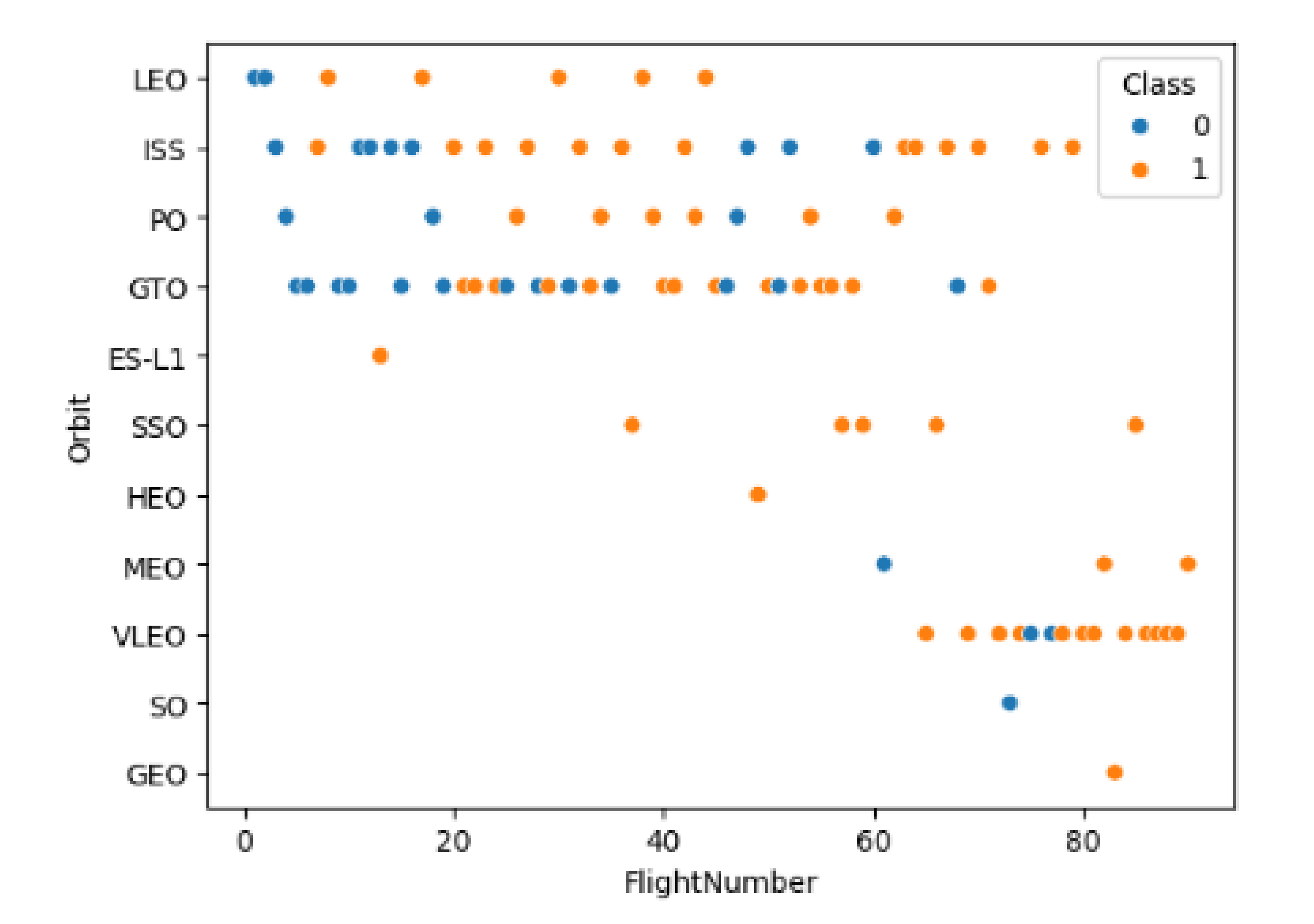
The SO orbit type has the lowest success rate



Flight Number vs. Orbit Type

The LEO orbit type's Success rate seems to be correlated with the Flight number

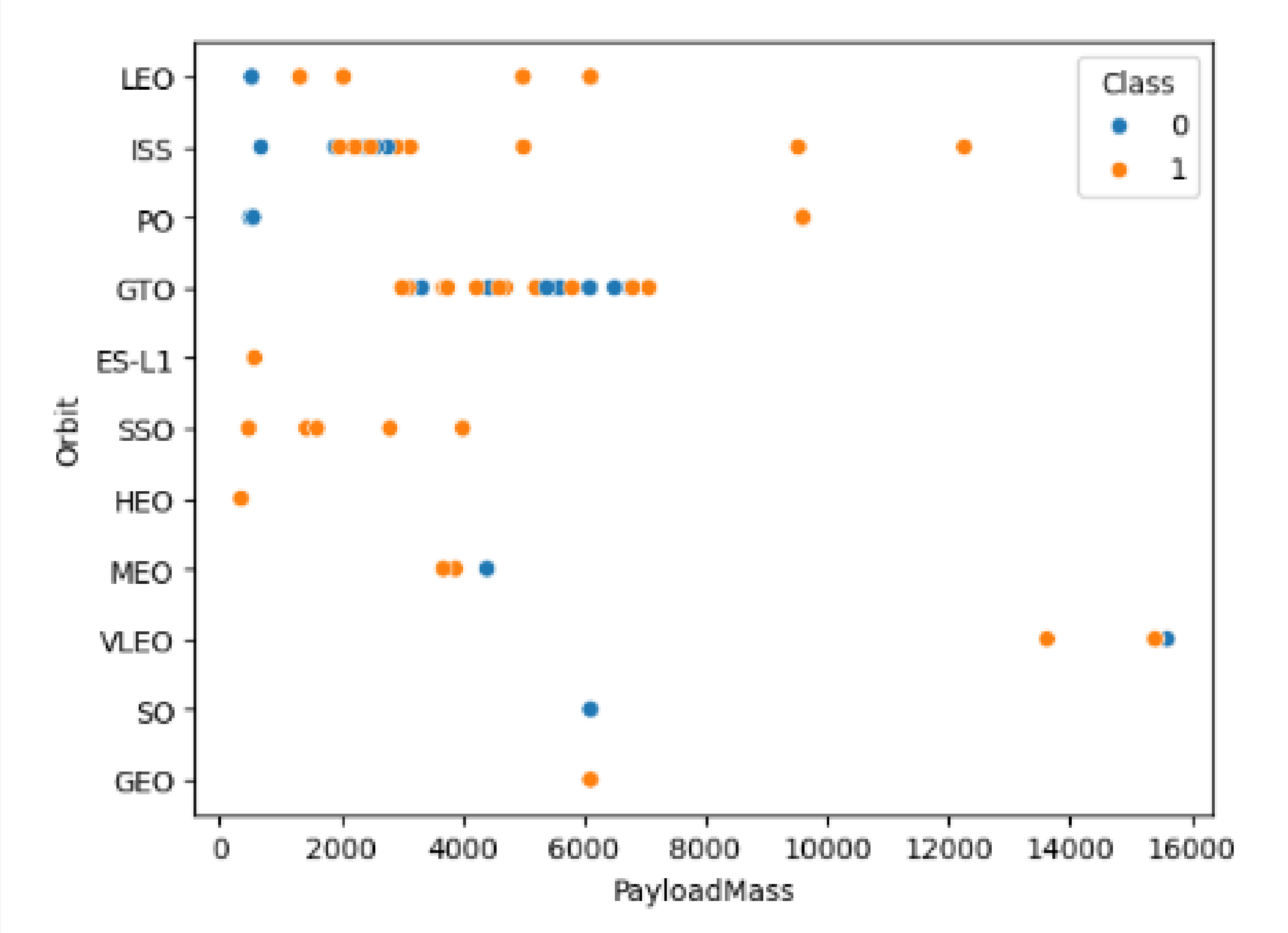
There is no correlation in GTO orbit type



Payload vs. Orbit Type

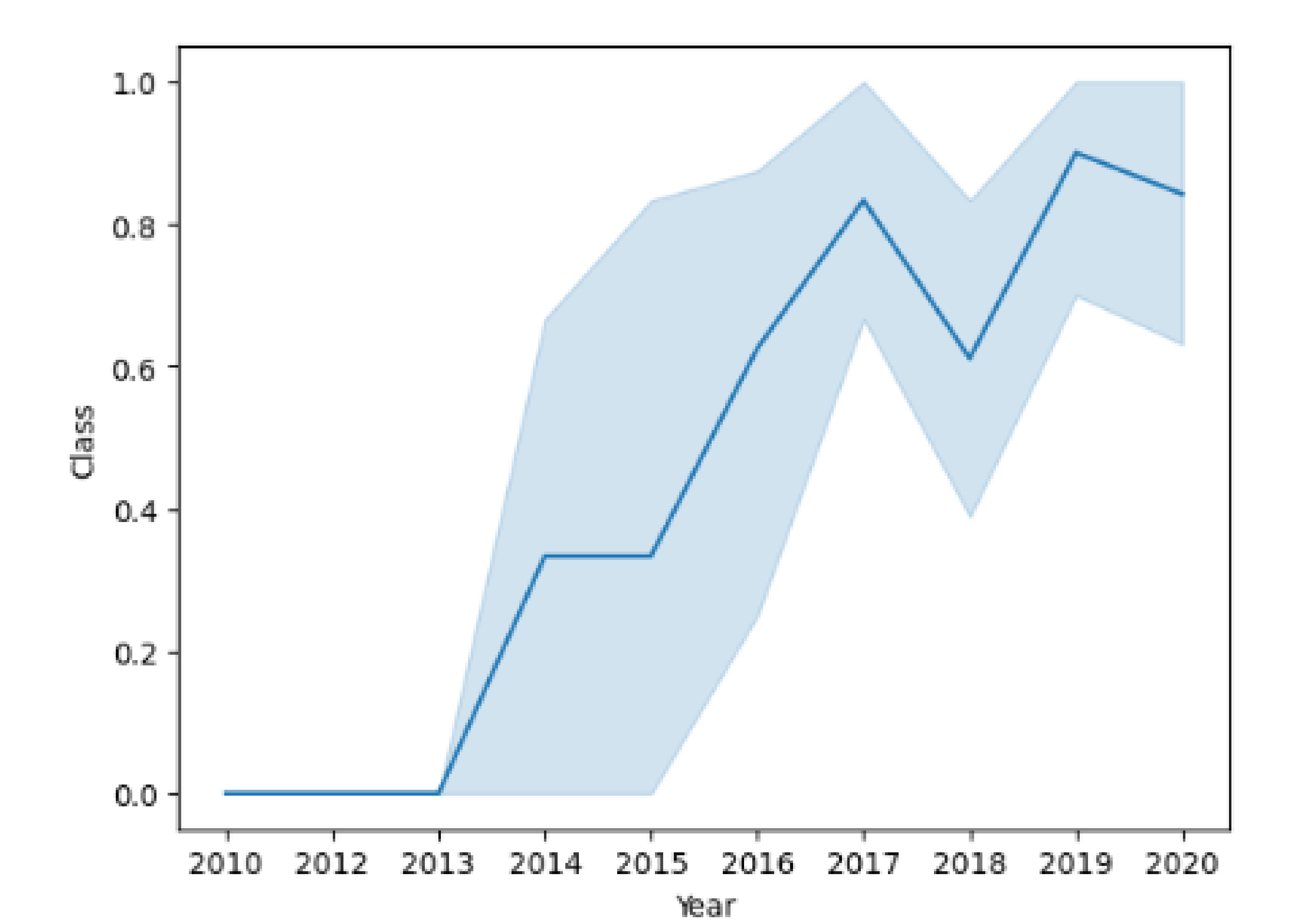
With heavy payloads, the successful landing rate are more for LEO, PO and ISS

However, There is no relationship for GTO orbit type



Launch Success Yearly Trend

- There is a steady increase in Success rate from 2013 to 2017 (stable in 2014)
- There is a decrease in 20% success rate from 2017 to 2018 and regained in 2019



All Launch Site Names

- Unique launch sites:

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

- There are total four launch sites for SpaceX across United States

Launch Site Names 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

It contains information about Booster version, Payload mass, Orbit type, Customer, Mission and Landing outcome

Total Payload Mass

The total payload mass carried by boosters launched by NASA (CRS) is 45,596 Kg

Customer	Total_Payload
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

The table shows Average payload mass carried by booster version F9 v1.1

The highest carrying capacity boosters are B1016, B1011, B1014

Booster_Version	Average_Payload
F9 v1.1	2928.4
F9 v1.1 B1003	500.0
F9 v1.1 B1010	2216.0
F9 v1.1 B1011	4428.0
F9 v1.1 B1012	2395.0
F9 v1.1 B1013	570.0
F9 v1.1 B1014	4159.0
F9 v1.1 B1015	1898.0
F9 v1.1 B1016	4707.0
F9 v1.1 B1017	553.0
F9 v1.1 B1018	1952.0

First Successful Ground Landing Date

The First successful landing outcome in ground pad was achieved on January 22, 2015

min(date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are shown in the figure

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes are 101

Total
101

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass are shown in the table

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

The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

month	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

The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

Landing_Outcome	Count
Failure (drone ship)	5
Success (ground pad)	3

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

There are 4 launch sites in the United States of America

All the locations are situated near the coastal regions

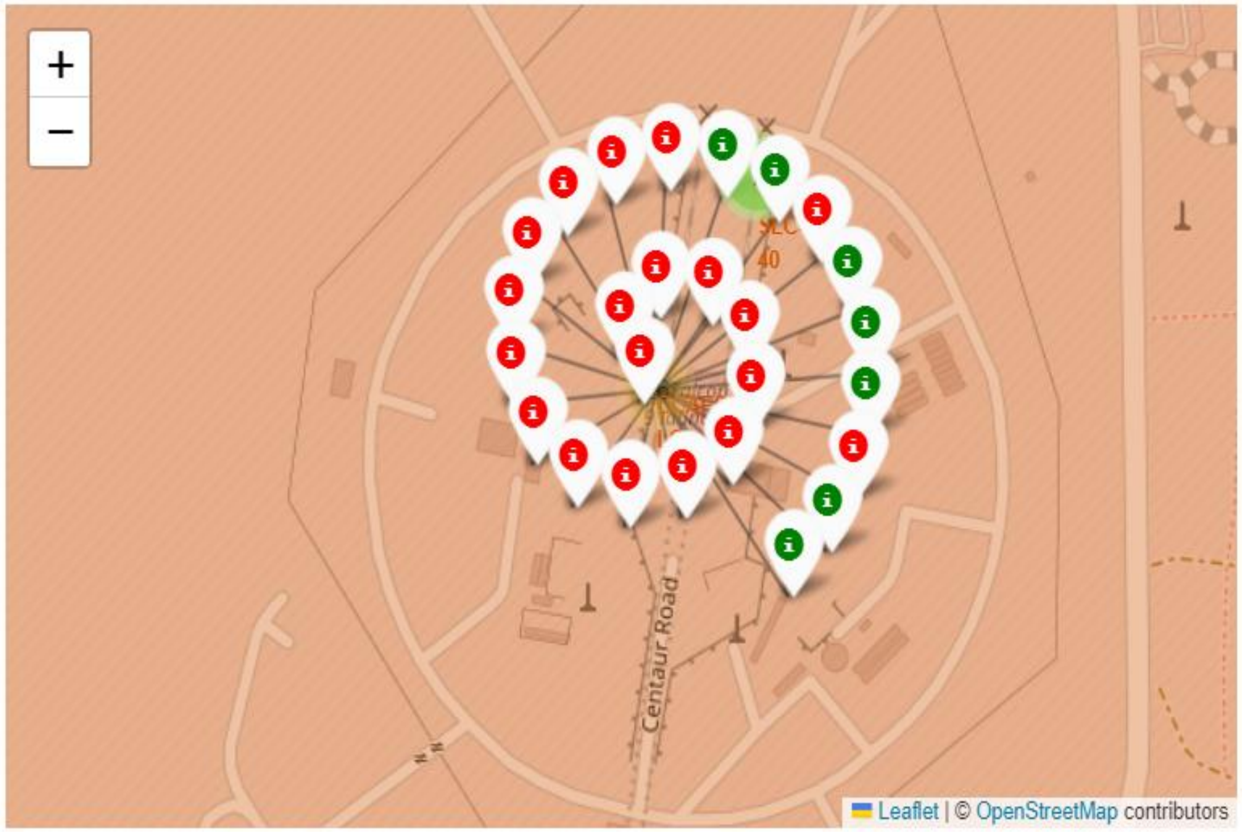
Three of them are situated near to each other and in Florida state.

Other in Los Angeles

Success / Failed launch

From the color-labeled markers in marker cluster, it is easy to identify which launch sites have high success rates

CCAFS LC-40 Launch site have very low success rate as shown on the figure



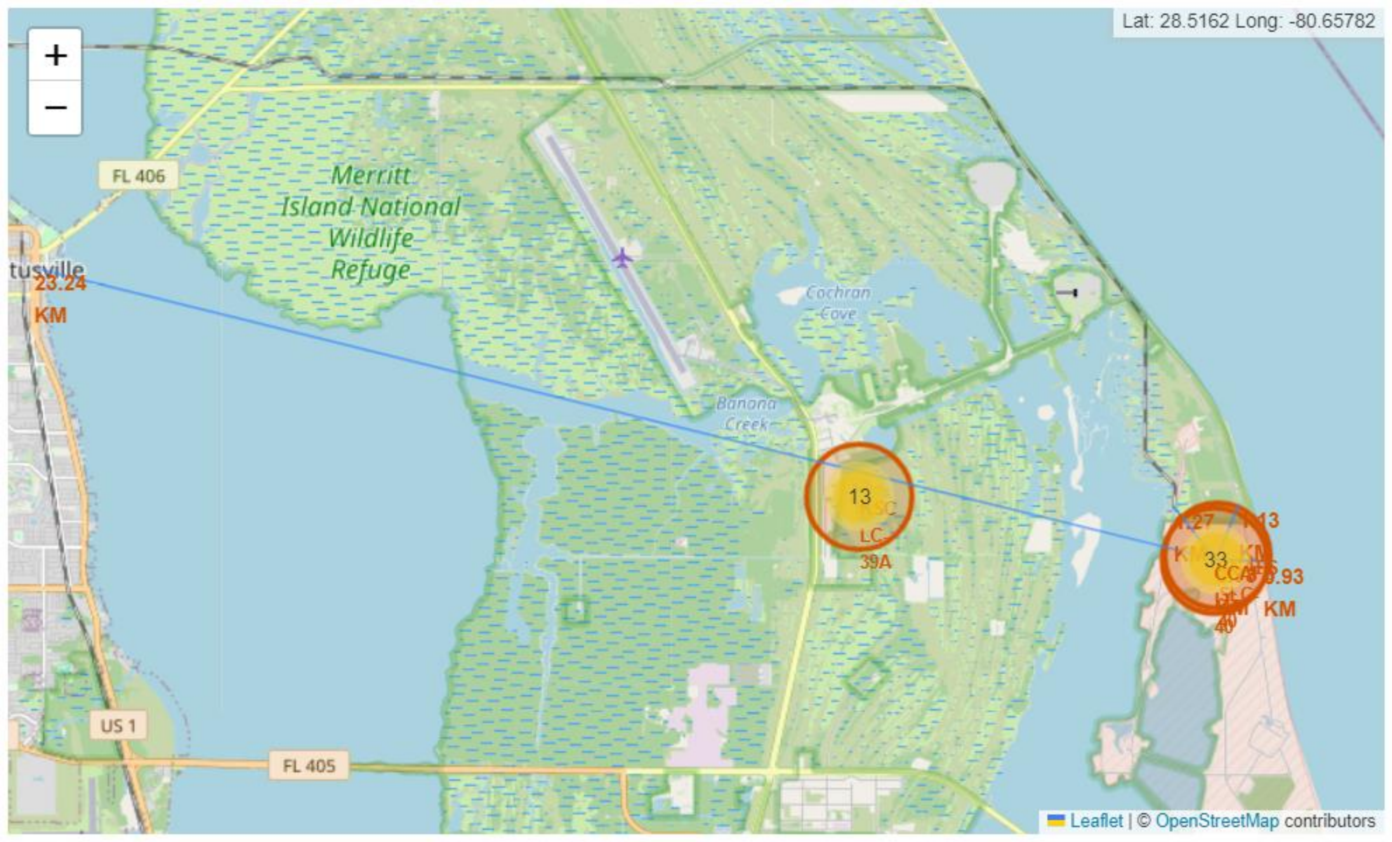
Distance from Launch Site to nearby areas

The nearest coastline from the launch site is 0.93 KM

The nearest highway from the launch site is 1.13 KM

The nearest railway from the launch site is 1.27 KM

The nearest city from the launch site is 23.24 KM



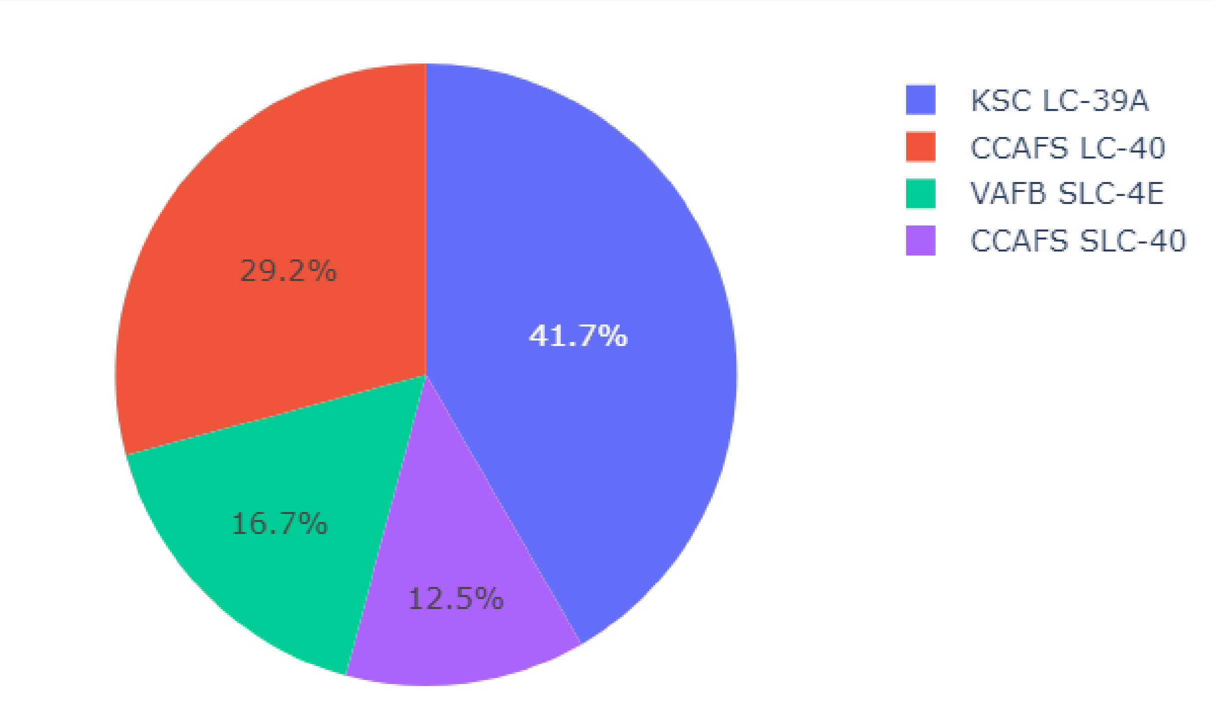
Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site

The KSC LC-39A has the highest successful launches with 41.7% of the total proportion

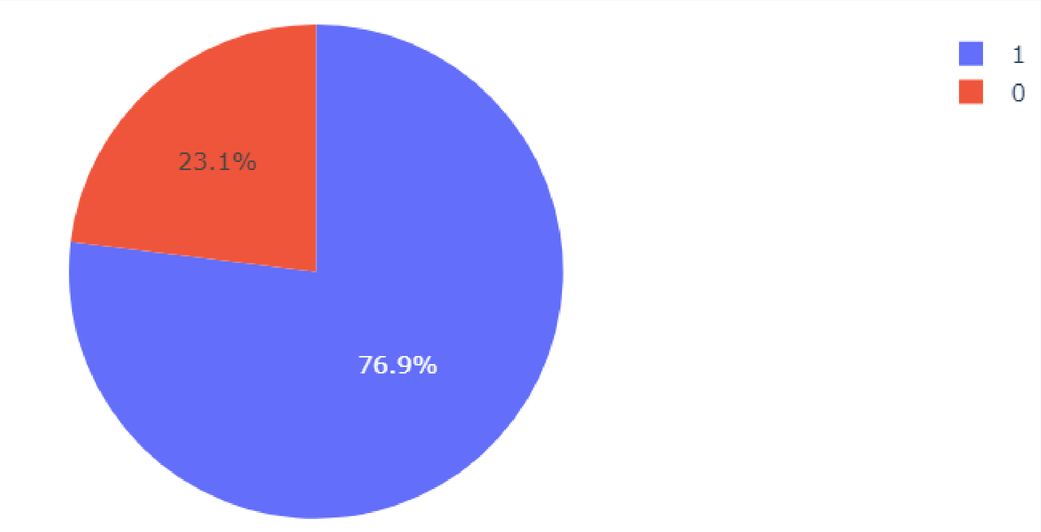
The CCAFS LC-40 ranks second with 29.2 % of the total proportion



Highest Launch Success Ratio

The KSC LC-39A has the highest launch success ratio out of all other launch sites

It has 76.9 % successful landings and 23.1 % unsuccessful landings



Payload vs Launch Outcome

The figure shows the Scatter chart for all launch sites and Payload range of 1000 – 8000 Kg

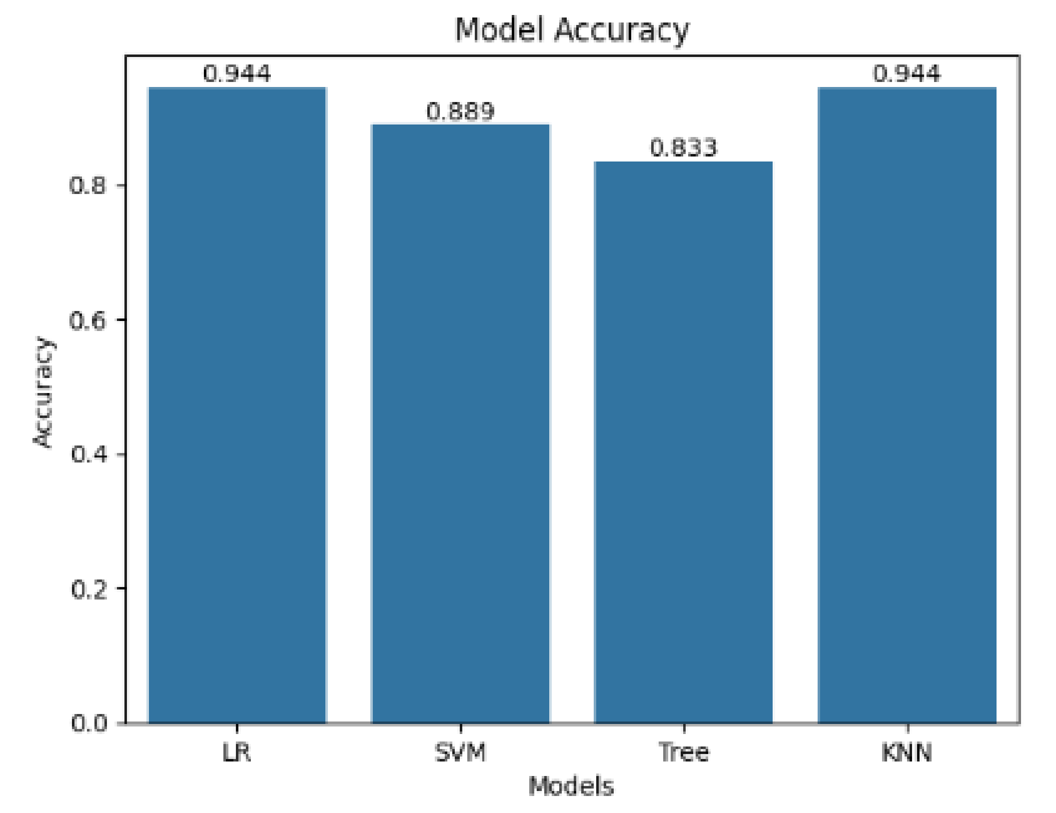
The Booster versions such as FT, B4, B5 have the highest success ratio in the given payload range

Section 5

Predictive Analysis (Classification)

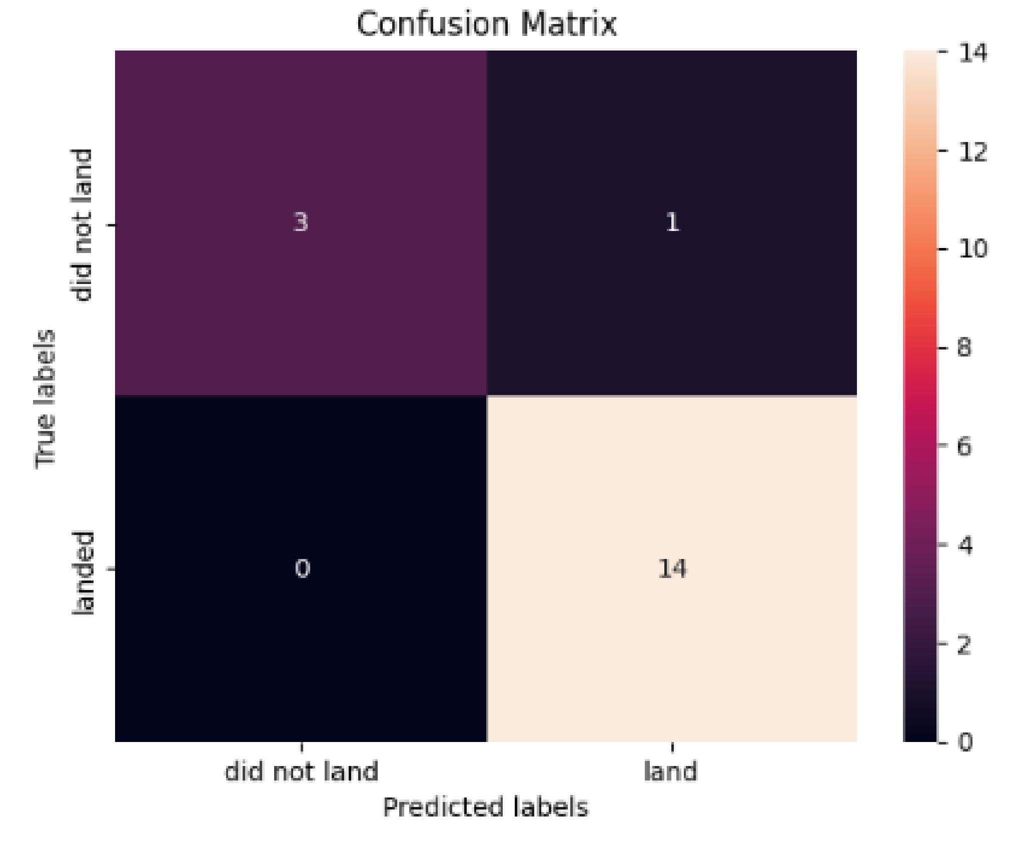
Classification Accuracy

- The accuracy for all 4 classification models is displayed in a bar chart
- LR and KNN model have the highest classification accuracy
- In terms of all overall metrics, KNN performs better than other models



Confusion Matrix

- This is the confusion matrix for KNN model
- The model was able to classify most predictions as correct
- It misclassified only one prediction as False Positive



Conclusions

- As the number of flights increases, the first stage is more likely to land successfully
- Success rate is higher for FT Booster version with Payload mass range between 2000 to 6000 Kg
- There is a steady increase in success rate from 2013 to 2017
- Orbit types such as ES-L1, GEO, HEO, SSO have the highest launch success rates and Orbit GTO recorded the lowest
- Launch Site 'KSC LC-39A' has the highest launch success rate and 'CCAFS SLC-40' has the lowest launch success rate
- Launch sites are located away from the cities and closer to the coastline, railways and highways
- The best performing model in terms of overall metrics is the K-Nearest Neighbors with an 94% accuracy score

Thank you!