IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Amar Pandit>
<6/5/22>

Outline

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

Executive Summary

- Summary of methodologies
  - Webscraping and the SpaceX API are used to obtain and manipulate the data.
  - Machine Learning was used to predict the best possible outcomes.
  - Data visualization theory and the creation of interactive visuals are used to reach conclusions about the data
- Summary of all results
  - Using the above methodologies and available data, it is possible to determine the best features that will predict the success of the launchings.
  - This is supported by predictions made through machine learning to determine important features.

Introduction

- Project background and context
  - Determine if the new company, SpaceY can successfully compete with SpaceX
- Problems you want to find answers
  - Seeking to lower overall costs for launches by determining if landings will be successful.
  - Finding the best locations for launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Used the SpaceX API, https://api.spacexdata.com/v4/rockets/
  - Performed WebScraping on the provided data tables.
- Perform data wrangling
  - Data was evaluated in order to determine confidence levels by which we could predict favorable launch conditions.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Perform predictive analysis using classification models

- A train/test split was created from the collected data and then four classification models were used to verify the results.

Data Collection

- Using Web Scraping and the SpaceX API, we were able to collect and manipulate the data.

Data Collection – SpaceX API

- Obtained data from the public SpaceX API which was used as shown in figure 1

- https://github.com/Zamplifier/class10repo/blob/master/Class%2010%20week%201.ipynb

Figure 1: API Call Flowchart

Data Collection - Scraping

- Used Wikipedia as a source of the data.
- Data were downloaded as shown in figure 2.
- https://github.com/Zamplifier/class10repo/blob/master/Week%201%20Data%20Collection%20with%20WebScraping.ipynb

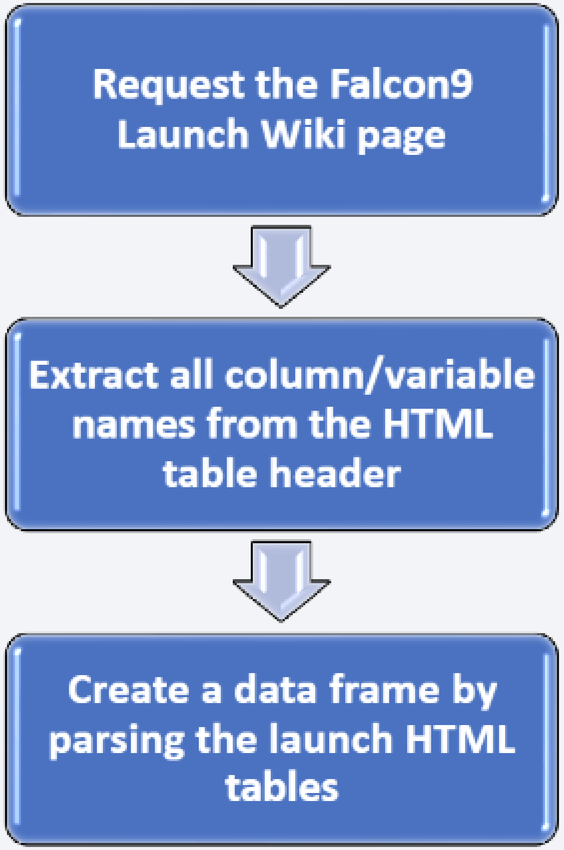

Figure 2: WebscrapingFlowchart

Data Wrangling

- As shown in Figure 3, after completing initial EDA, key data was calculated and summarized including mission outcome and orbit occurrences.
- From the outcome data, a landing outcome label was created.

Figure 3: Data Wrangling Flowchart

https://github.com/Zamplifier/class10repo/blob/master/Class%2010%20Week%201%20-%20Data%20Wrangling.ipynb

EDA with Data Visualization

- The data was visualized using scatterplots (Figure 4) and bar graphs to show the correlation between the following features;
  - Launch Site and Flight Number
  - Launch Site and Payload Mass
  - Orbit and Flight Number
  - Payload Mass and Flight Number
  - Payload and Orbit

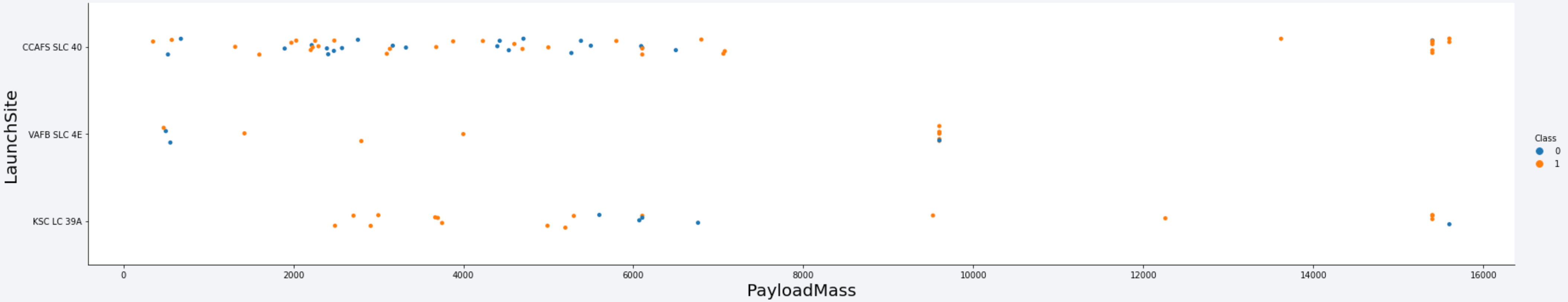

Figure 4: Launch Site vs Payload Mass

https://github.com/Zamplifier/class10repo/blob/master/Coursera%20Week%2010%20Week%202%20-%20EDA%20with%20Visualization.ipynb

EDA with SQL

- Performed the below SQL queries and found;
  - Names of the unique launch sites.
  - 5 launch sites with names beginning with “CCA”.
  - The total payload mass carried by NASA boosters.
  - The Average payload mass carried by booster version F9 v1.1.
  - Date of first successful ground pad landing
  - Booster names that were successful in drone ship operations.
  - Payload mass between 4000 and 6000kg
  - Numbers of successful and failed missions
  - Booster versions which have carried the maximum payload mass
  - Drone ship failed landing outcomes
  - Failed landing outcome booster versions and launch site names in 2015.
  - Successful and failed landing outcomes between 2010 and 2017.

https://github.com/Zamplifier/class10repo/blob/master/Coursera%20Class%2010%20Week%202%20-%20EDA%20with%20SQL%20.ipynb

Build an Interactive Map with Folium

- The created folium maps were fit with the following features;
  - Launch sites were indicated by markers
  - The Space Centers were indicated by circles and highlights
  - Marker clusters show event groups such as launches and launch sites
  - Lines show the distance between locations.

https://github.com/Zamplifier/class10repo/blob/master/Class10Week3%20%20InteractiveVisualAnalyticswithFolium.ipynb

Build a Dashboard with Plotly Dash

- The below plots and graphs were created;
  - Successful launches by site
  - Payload mass between 0kg and 9600kg
- With this information, we can see the relationship between the launch sites and payloads. This will allow us to find the best launch location.

- https://github.com/Zamplifier/class10repo/blob/master/Interactive%20Dashboard%20with%20Plotly%20Dash

Predictive Analysis (Classification)

- The following classification models were used (Process shown in Figure 5);
  - Logistic Regression
  - Decision Tree
  - K Nearest Neighbors

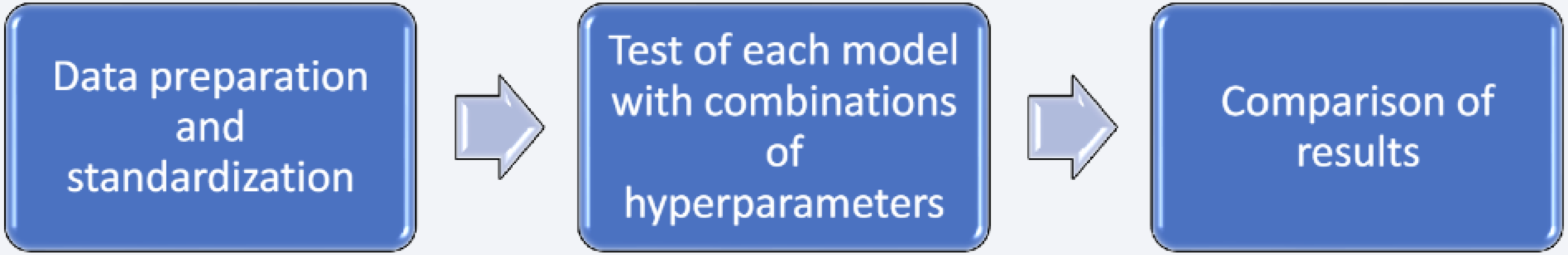

Figure 5: Predictive Analysis Flow Chart

- https://github.com/Zamplifier/class10repo/blob/master/Class%2010%20Week%204%20%20Machine%20Learning%20Predictions.ipynb

Results

- Exploratory data analysis results:
  - The first successful landing happened in 2015
  - There is a near 100% mission success rate
  - The later launches show better landing outcomes
  - There are four locations used by SpaceX for launches
  - A significant number of Falcon 9 booster versions landed successfully.
  - Booster versions F9 v1.1 B1012 and F9 v1.1 B1015 failed at landing in drone ships

Results

- Using folium maps, we can see that most of the launches are planned for the Florida launch site, where the infrastructure is favorable.

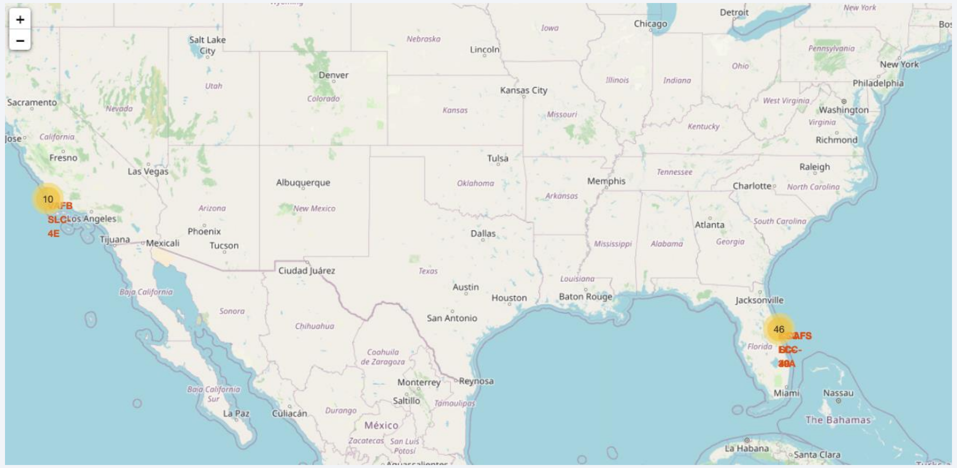

Figure 6: SpaceX Launch Sites

Figure 7: Most Favorable Launch Locations

Results

- Our analysis determined that the Decision Tree Classification is the best prediction model for successful landings.
- The accuracy of the Decision Tree Classifier was 90%

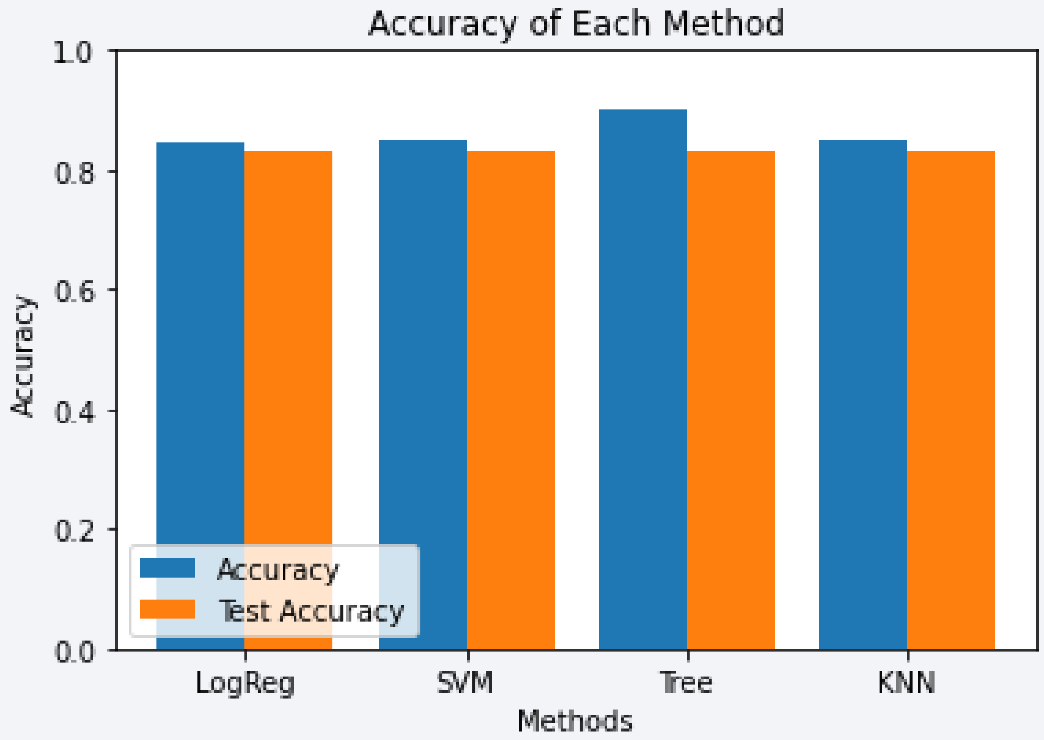

Figure 8: Evaluation of Accuracies of Testing Methods

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- As shown in Figure 9, CCAF5 SLC 40 seems to be the location of the most successful launches.
- Success rate has generally improved over time

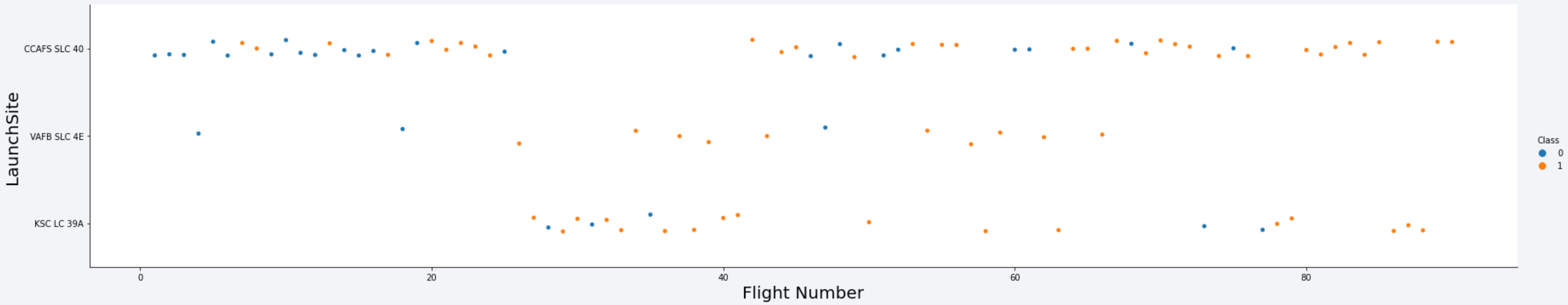

Figure 9: Flight Number vs. Launch Site

Payload vs. Launch Site

- Figure 10 shows that the high payloads, those >8000kg, seem to have the most success.
- At >12,000kg, only CCAFS SLC 40 and KSC LC 39 seem to be the only possible launch sites.

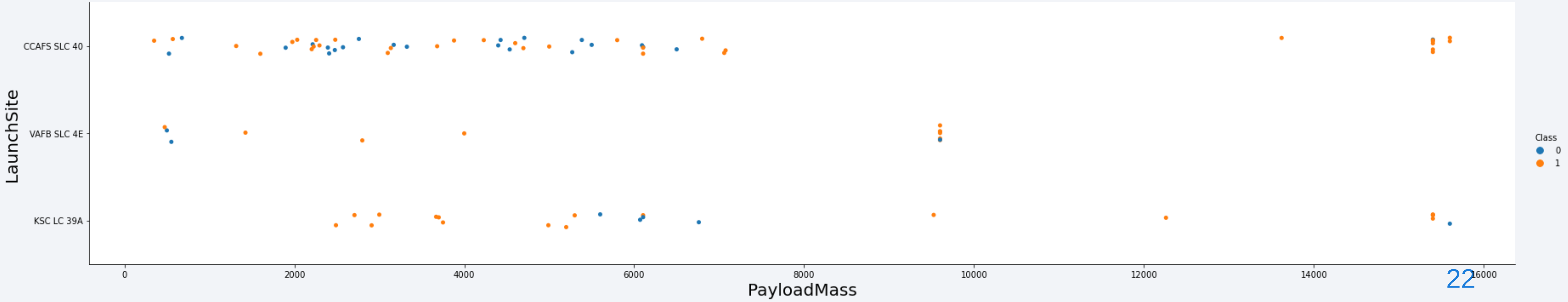

Figure 10: Payload vs. Launch Site

Success Rate vs. Orbit Type

- Figure 11 shows that the following orbits had the highest success rates;
  - ES-L1
  - GEO
  - HEO
  - SSO

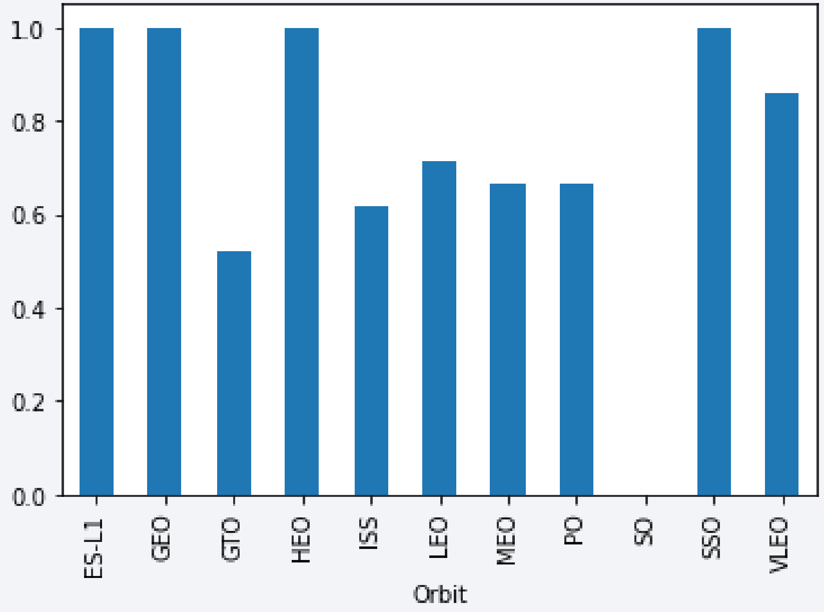

Figure 11: Success Rate vs. Orbit Type

Flight Number vs. Orbit Type

- In Figure 12. VLEO shows increased occurrence, which could be utilized in future endeavors
- The trend supports that success rate improves over time.

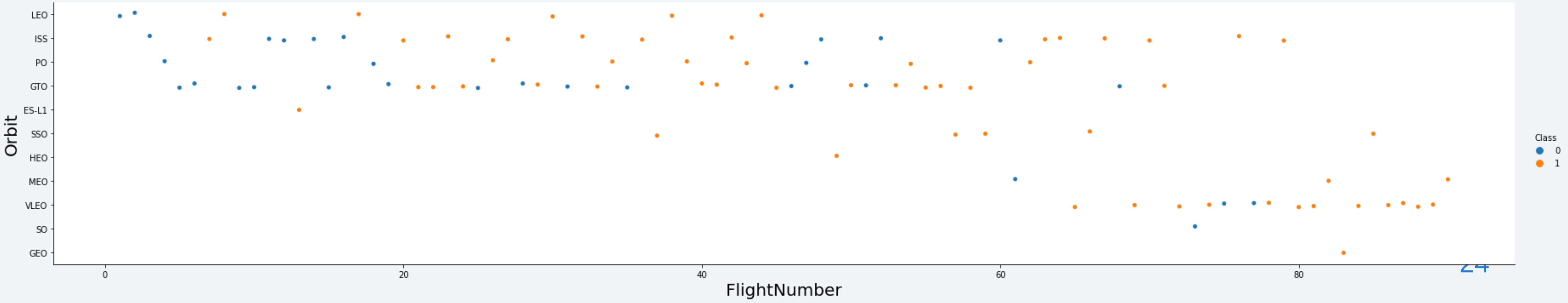

Figure 12: Flight Number vs. Orbit Type

Payload vs. Orbit Type

- In Figure 13, we can clearly see that the ISS orbit can accommodate a large payload range with a reasonable success rate.
- While orbit GTO has success across a wide payload mass, it is very inconsistent with successes.
- More data is needed to draw any conclusions on orbits SO and GEO.

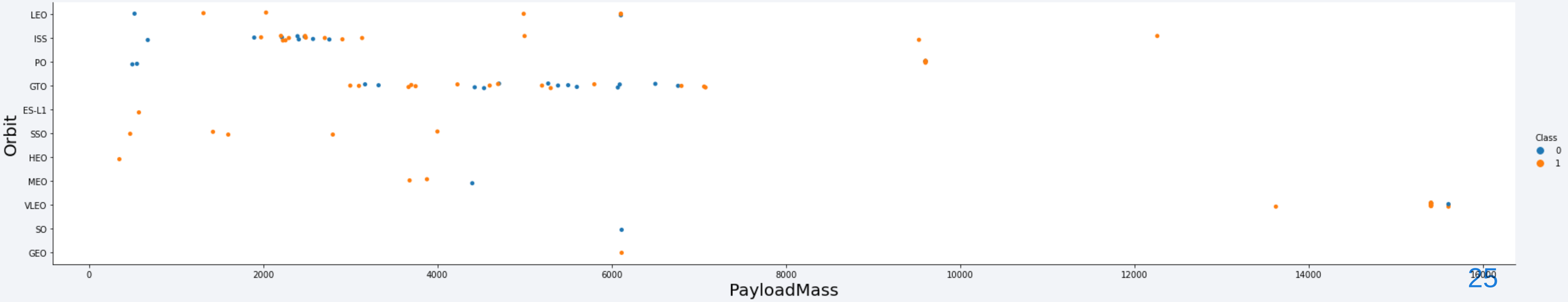

Figure 13: Payload vs. Orbit Type

Launch Success Yearly Trend

- Figure 14 shows that from 2013 onward, there is a sharp increase in success rate.
- This seems to taper off near 2020, but more data is needed to confirm whether or not this is just a local minimum.
- The rapid increase in success rate from 2013 to 2019 could be due to the process being highly iterative in the early years until most features were optimized.

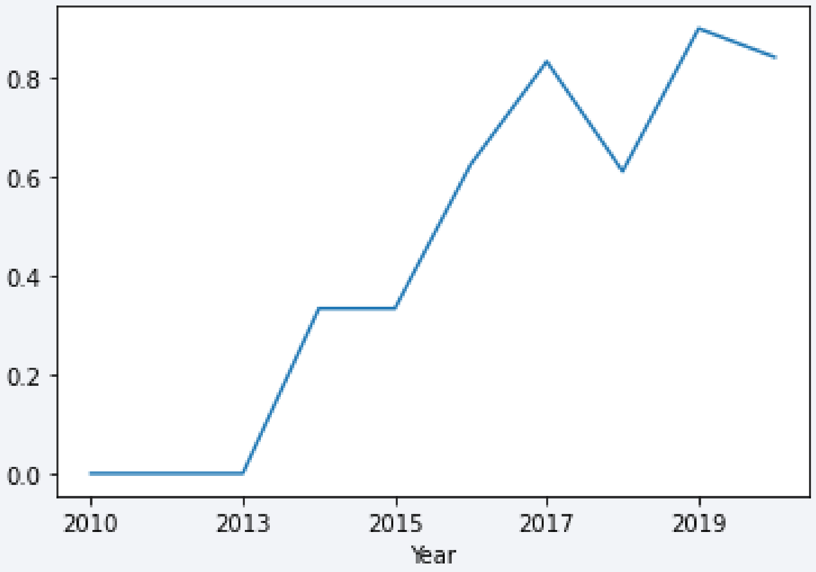

Figure 14: Yearly Trend of Launch Successes

All Launch Site Names

- Four unique launch sites are available as shown in Figure 15.
- From the data, filtering code was used to select only unique launch sites.

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

Figure 15: Unique Launch Sites

Launch Site Names Begin with 'CCA'

- Figure 16 show five launches from Cape Canaveral, Florida

]:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Figure 16: Five Launches from Cape Canaveral

Total Payload Mass

- Figure 17 shows the total payload mass in Kg for payloads carried by NASA boosters.
- This was calculated by filtering the payloads by site codes containing “CRS” which refer to NASA.

total_payload
111268

Figure 17: Total Payload carried by NASA boosters in Kg

Average Payload Mass by F9 v1.1

- Figure 18 shows that the average payload mass carried by booster F9 v1.1 is 2,928kg
- This was calculated by selecting the booster version specifically in the data and averaging the payloads.

avg_payload
2928

Figure 18: Average Payload Mass by F9 v1.1

First Successful Ground Landing Date

- Figure 19 shows the first successful ground pad landing.
- This was obtained by using the MIN function on the DATE column and selecting for rows containing ‘Success (ground_pad).

first_success_gp
2015-12-22

Figure 19: Date of First Successful Ground Pad Landing

Successful Drone Ship Landing with Payload between 4000 and 6000

- Figure 20 show the boosters which have successfully landed on a drone ship with a payload between 4000kg and 6000kg.
- This was obtained by selecting distinct booster versions within the payload mass constraints.

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

Figure 20: Boosters for Successful Landings with Payloads between 4000 and 6000kg

Total Number of Successful and Failure Mission Outcomes

- Figure 21 was obtained by using the Count() function and grouping by Mission Outcome.
- There were 99 successes, 1 success with an unclear payload status, and 1 in-flight failure.

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Figure 21: Total Successful and Failed Missions

Boosters Carried Maximum Payload

- Figure 22 shows only the boosters which were able to carry the maximum payload.
- This was obtained by using the MAX() function on the “Payload_Mass_KG” column and selecting for distinct booster versions.

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

Figure 22: Boosters Which Have Carried the Maximum Payload

2015 Launch Records

- Figure 23 shows the failed drone ship landings and launch site names in 2015
- This was obtained by selecting for booster versions whose Landing_Outcome was listed as ‘Failure (drone ship)’

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Figure 23: 2015 Failed Drone ship Landing Launch Site and Booster Version.

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

- Figure 24 shows that ‘No attempt’ landing outcomes are the most common occurrence in this time frame.
- This reinforces the need for favorable launch and landing locations and conditions.

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

Figure 24: Landing outcomes Ranked by Frequency

Section 3

Launch Sites Proximities Analysis

North American Launch Sites

- As shown in Figure 25, the launch sites are located near coasts. This is most likely to ensure that there is plenty of safe landing locations.

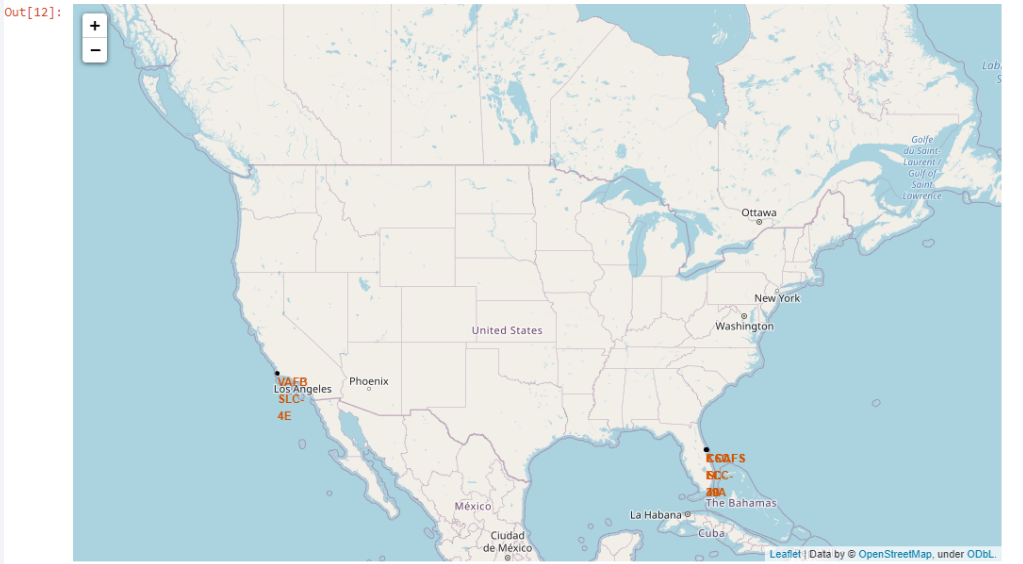

Figure 25: Launch Sites in North America

Launch Outcomes

- The map in Figure 26 show the launch outcomes by site. Green for successes and red for failures.

Figure 26: Launch Outcomes Labeled by Color

<Folium Map Screenshot 3>

- Figure 27 is a good example of the logistical safety needs for launch sites.
- The site itself is distanced from large housing areas and is near roads and railways.

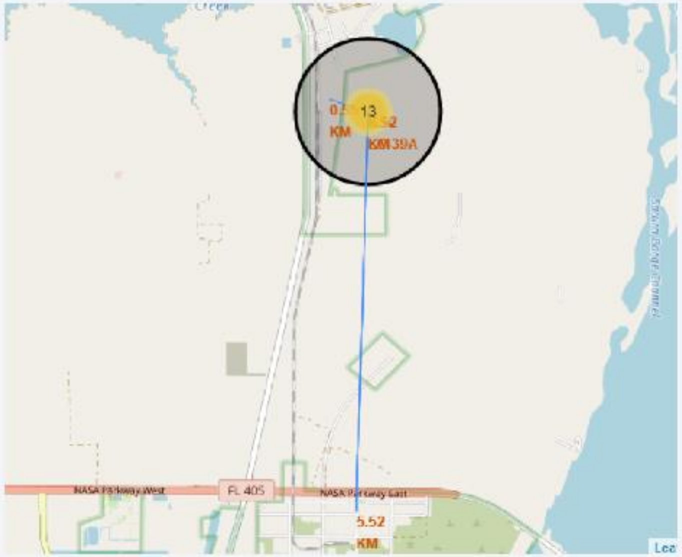

Figure 27: Proximity Map of the KSC LC-39A Launch Site

Section 4

Build a Dashboard with Plotly Dash

Successful Launches for Each Site

- It can be clearly seen in Figure 28 that the KSC LC-39A launch site has a majority of the successful launches
- Second best is the CCAFS LC-40 site, followed by VAFB SLC-4E and CCAFS SLC-40.

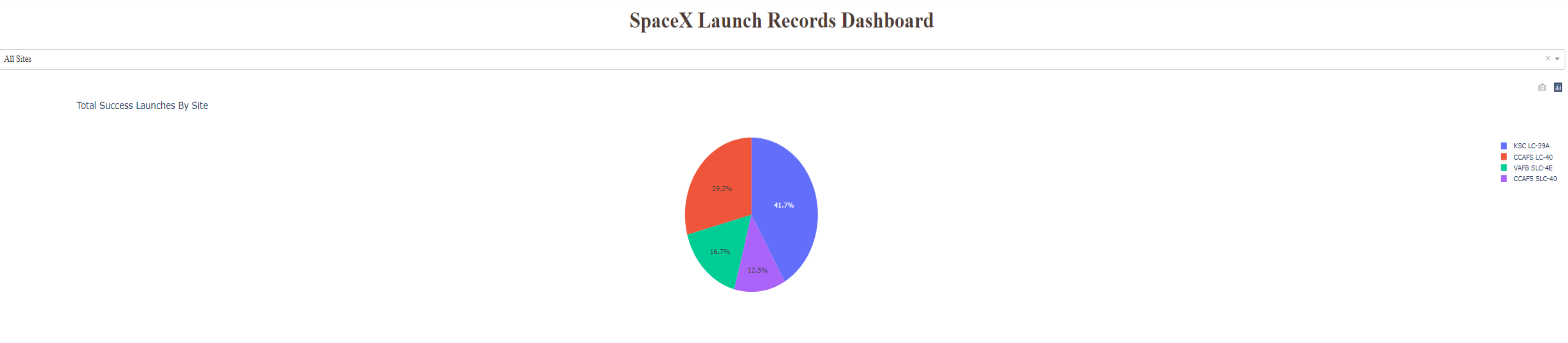

Figure 28: Pie chart Showing Counts of Successful Launches Divided by Launch Site

<Dashboard Screenshot 2>

- Figure 29 shows us a closer look at the most successful launch site, KSC LC-39A.
- 76.9% of launches are successful and 23.1% fail.

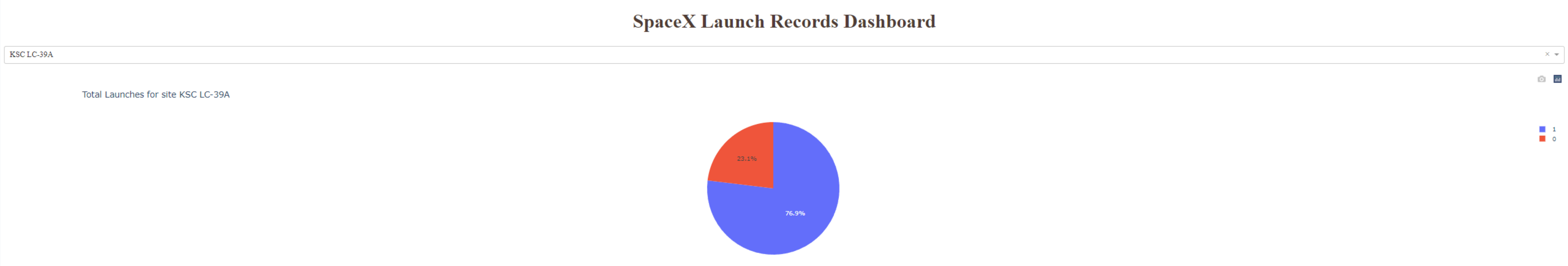

Figure 29: Launch Successes and Failures at the KSC LC-39A Launch Site

Payload vs. Launch Outcome Divided by BV

- In Figure 30, we have plotted the payload mass by launch outcome. We have further categorized the data by showing the booster version by color.
- Based on this, we can see that the FT Booster seems to be the most successful. While the v1.1 seem to be the least successful.
- The most successes with the FT booster are between 2000kg and 4000kg
- There is significantly less data for launches above 6000kg, so further research is needed to determine if there is any usable payload between 6000kg and 10000kg.

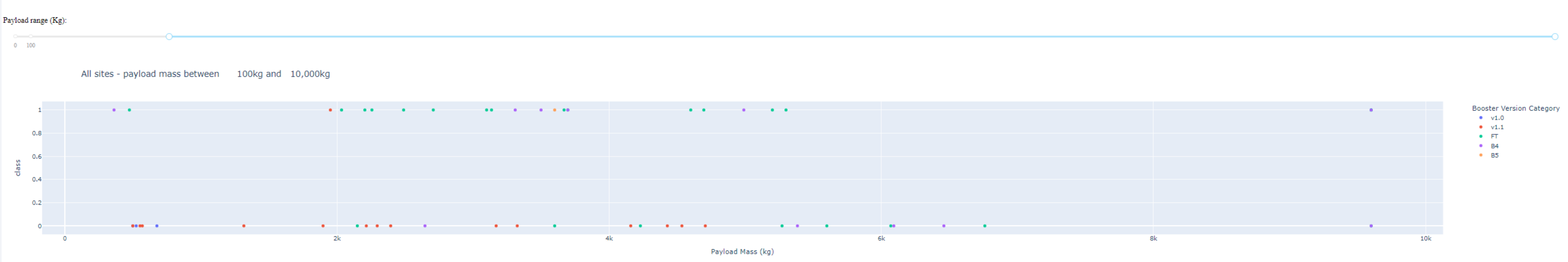

Figure 30: Scatterplot Showing Launch Outcome based on payload and booster version

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- As shown in Figure 8, four classification models were tested and plotted.
- The decision tree classifier is the best overall method, with the highest overall accuracy and comparable test accuracy.
- All tests show accuracies above 80%.

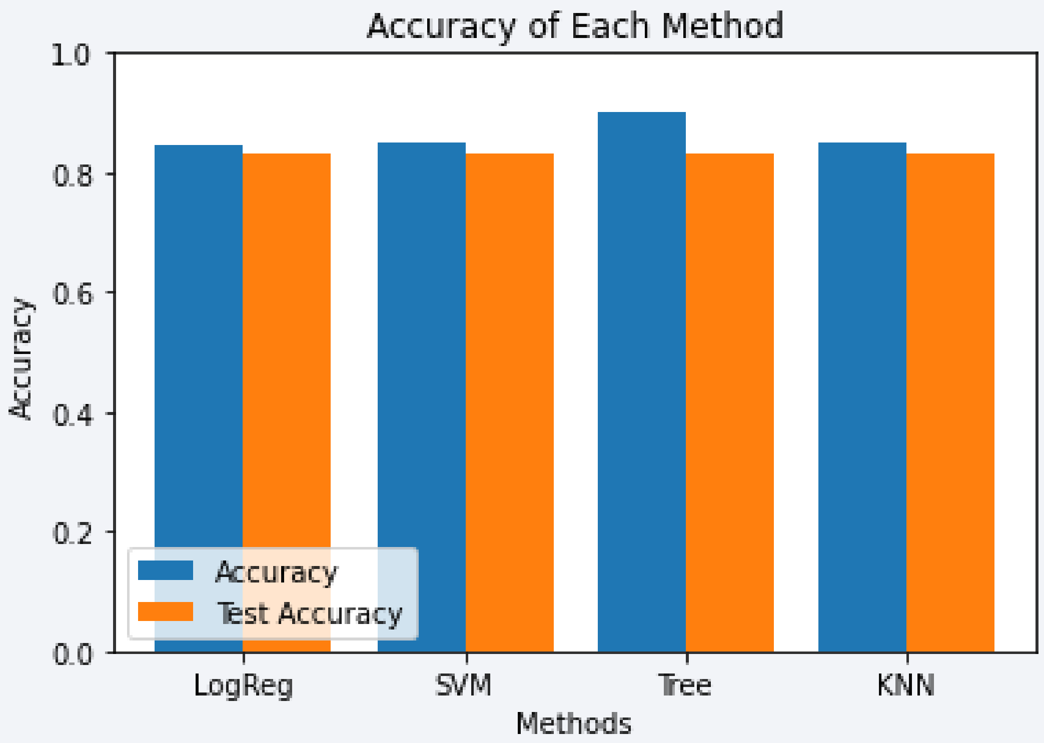

Figure 8: Evaluation of Accuracies of Testing Methods

Confusion Matrix

- In Figure 31 we can see that the data supports Decision Tree Classifier as the best model as there are larger numbers of true positives and true negatives
- The false positives and negatives are very low.

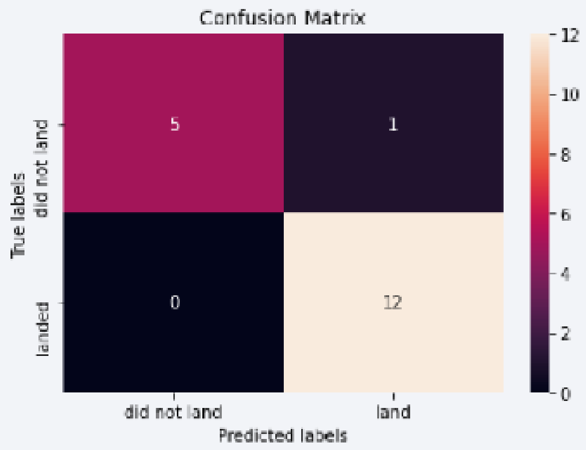

Figure 31: Confusion Matrix of the Decision Tree Model

Conclusions

- From the supporting evidence, the best scenario is as follows;
  - KSC LC-39A as the launch site.
  - Payloads between 2000kg and 4000kg
  - It is more likely to have a successful launch than a successful landing. The outcome of a launch improve with more attempts.
  - Decision Tree Classifier is the best method to determine if a landing will be successful.

Appendix

- Had issues with Folium maps not showing any changes even with correct code.
- Could not fix some issues due to the monthly limited processing quantity allotted by free Jupyter notebooks through IBM.

Thank you!