IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdullah Alshalan
March 9, 2025

Outline

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

Executive Summary

- The data was collected from the SpaceX Wikipedia page
- The results showed a correlation between the predicted landing outcome and set of features
- implying that some tuning we can raise the successful landing rate

Introduction

- The cost of aircrafts varies between manufacturers, and it seems that landing outcome has a strong influence on the cost of aircrafts.
- with that in mind we aim to find the reasons for failed landing and how to overcome them in order to reduce cost.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was web-scraped using Beautiful soup python library
- Perform data wrangling
  - Removing missing values and performing data preprocessing techniques.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - With classification models to find the correlated features with landing outcome

Data Collection

The dataset was collected using web scraping techniques. Below is an overview of the process:

- Data Source: The data was obtained from the SpaceX Wikipedia page, which contains details about launch events, landing outcomes, and relevant flight parameters.
- Web Scraping: The Python BeautifulSoup library was used to extract structured tabular data from the Wikipedia page. Requests were made to the webpage, and the HTML content was parsed.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- [GitHub URL of the completed SpaceX API calls notebook]()

Start

▼

Identify Data Source (SpaceX Wikipedia)

▼

Request Webpage Content (Using Requests Library)

▼

Prepare to parse the HTML content

▼

End

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- [GitHub URL of the completed web scraping notebook]

Start

▼

Parse HTML Content (Using BeautifulSoup)

▼

Extract Tabular Data (Landing Outcomes, Dates, etc.)

▼

Clean & Structure Data (Using Pandas)

▼

Store Data (CSV File)

▼

Use Data for EDA & Model Training

▼

End

Data Wrangling

- Describe how data were processed
- Merged datasets, handled missing values, and formatted data for analysis.
- GitHub URL of completed data wrangling

EDA with Data Visualization

- Created interactive visualizations to explore data distributions and relationships, such as:
  - Scatter point charts for Payload and Launch Site
  - bar chart to find relationship between success rate of each orbit type
  - Line charts to Visualize the launch success yearly trend

- GitHub URL of completed EDA with data visualization notebook

EDA with SQL

Performed EDA using SQL queries to identify key patterns, example for queries used:

- SELECT DISTINCT(Launch_Site) FROM SPACEXTABLE;

- SELECT min(Date) FROM SPACEXTABLE

  where Landing_Outcome = 'Success (ground pad)';

- SELECT avg(PAYLOAD_MASS__KG_) FROM SPACEXTABLE

  where Booster_Version like 'F9 v1.0%';

you can find more queries in [GitHub URL of your completed EDA with SQL notebook]

Build an Interactive Map with Folium

The objects used to enhance the visualization of SpaceX launch sites and landing outcomes:

- Markers added at each launch site to pinpoint exact locations, using custom icons and popups displaying site names, making it easier to identify key locations.
- Circle Markers were placed around the launch sites, color-coded based on success rates with green indicating a high success rates and red representing lower success rates.
- polylines (lines) were drawn to connect launch sites to booster landing locations, whether on a drone ship or a ground pad

These elements collectively provide an intuitive and interactive way to analyze launch locations, and landing patterns, making the data more accessible and visually insightful.

GitHub URL of completed interactive map with Folium map

Build a Dashboard with Plotly Dash

The dash bord was filled with king of graphs:

1. Pie chart for Launch Sites Success Rate, with an option for a specified location or an overview for all locations

2. Scatter chart to show correlation between Payload and Success for all Sites or a chosen site

[GitHub URL of completed Plotly Dash lab]

Predictive Analysis (Classification)

- Conducted using ML models to classify whether a Falcon 9 booster landing would be successful or not.
- The notebook demonstrates how different models, including Logistic Regression, Support Vector Machines, Decision Trees, and KNN, were trained and evaluated.
- Confusion matrix was used to visualize model predictions.
- The final model selection was based on predictive accuracy and generalizability to unseen data.

Results

Through extensive data analysis, several insights were uncovered.

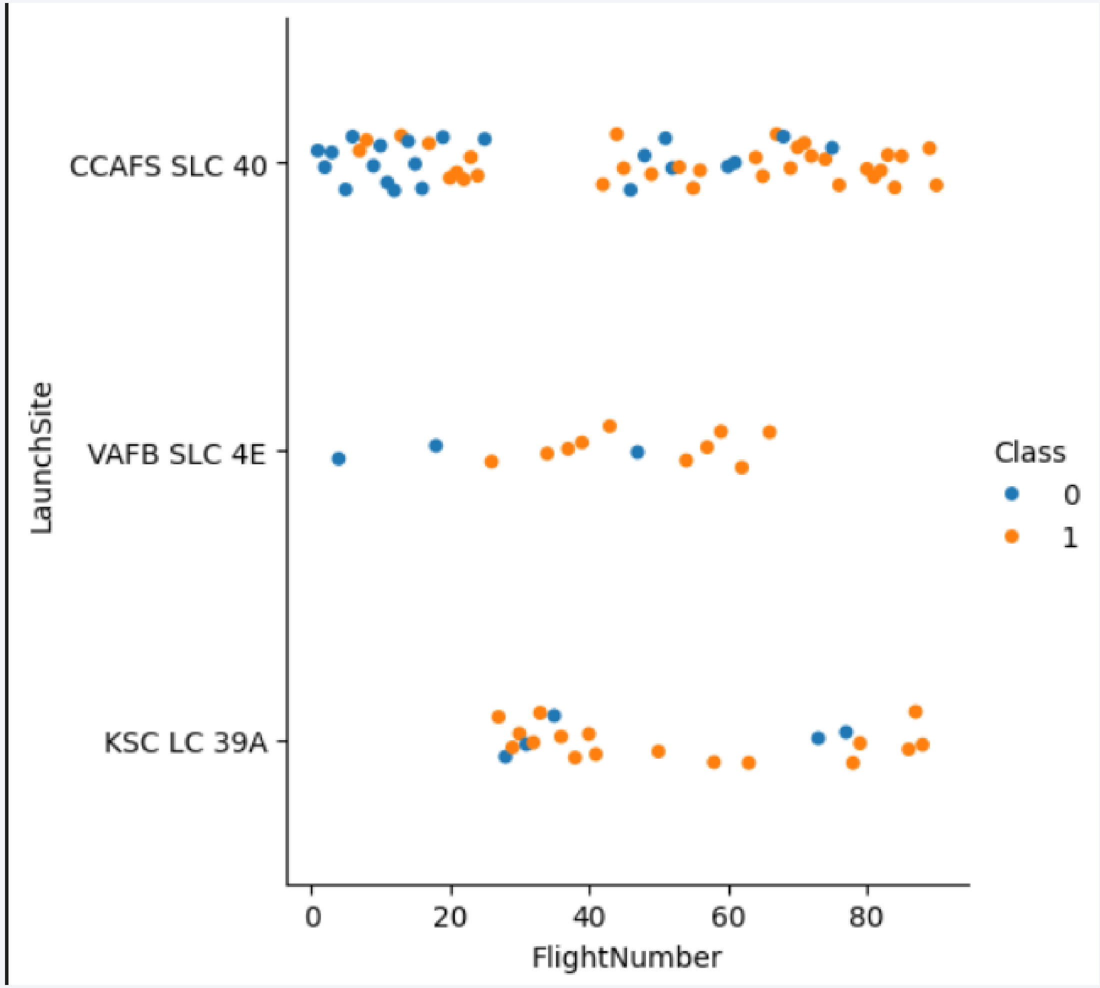
- The landing success rate varied significantly based on launch site selection, suggesting that some sites provide more favorable conditions for recovery.
- Drone-ship landings had a lower success rate than ground-based landings, likely due to oceanic weather conditions and movement dynamics.
- Revealed correlations between payload mass and landing success, where heavier payloads resulted in lower success probabilities.

Section 2

Insights drawn from EDA

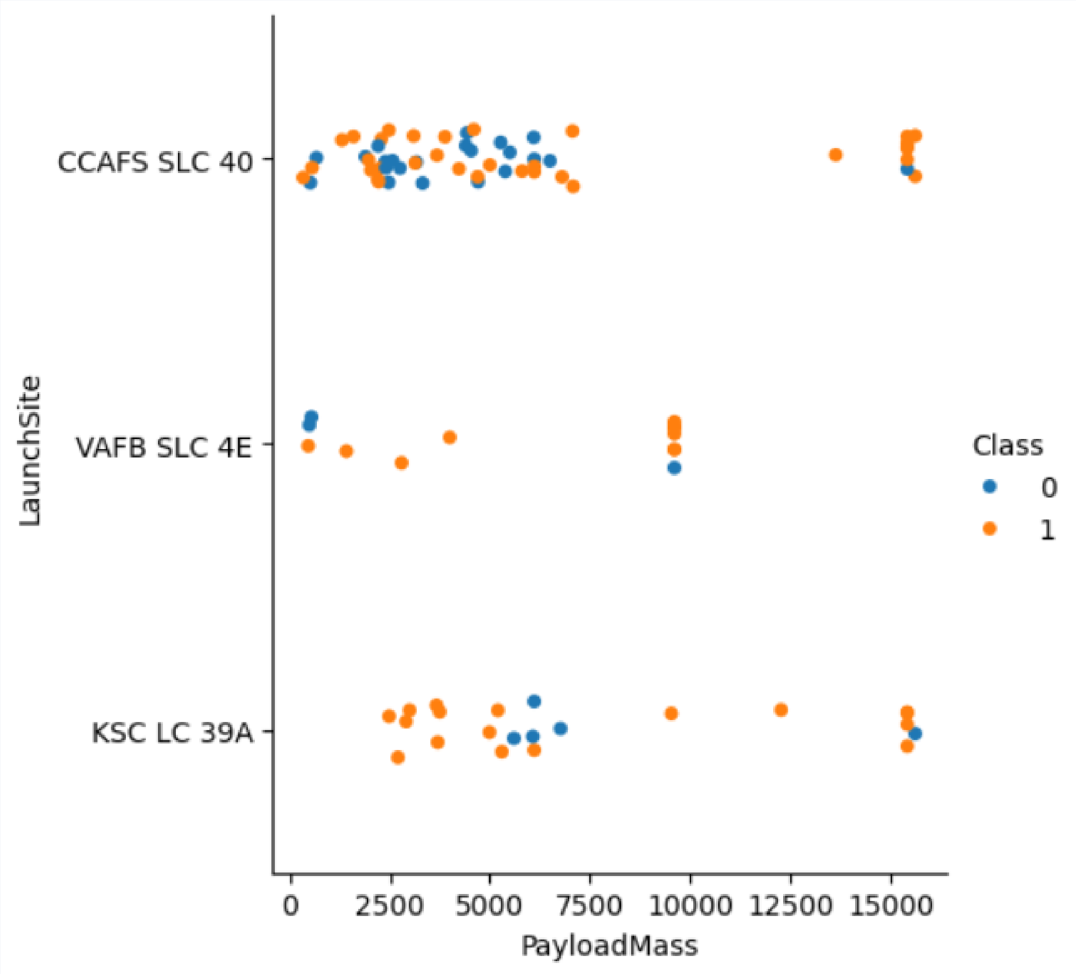
Flight Number vs. Launch Site

Flight number has no correlation with launch site, but interestingly in VAFB it seems that higher Flight number suggests successful landing



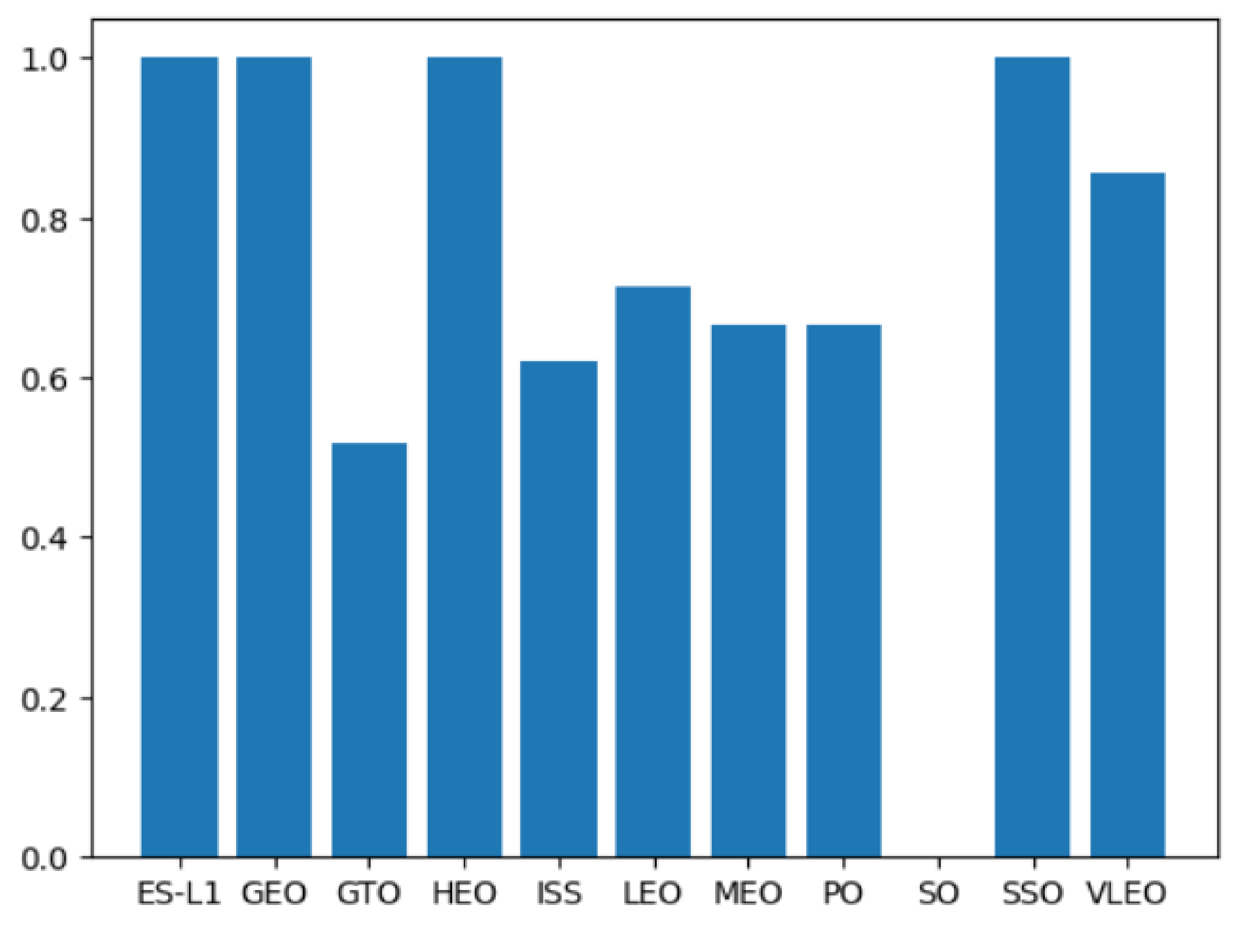
Payload vs. Launch Site

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



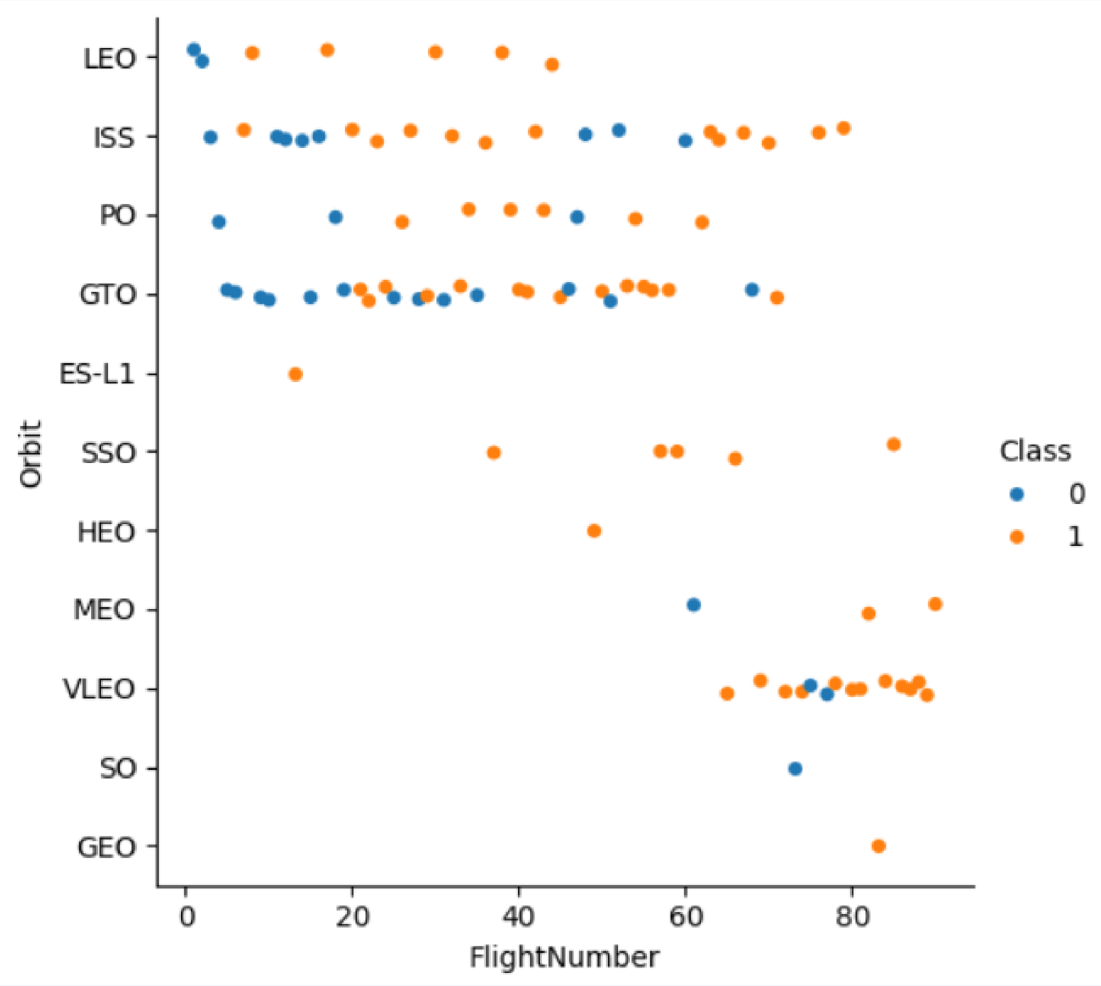
Success Rate vs. Orbit Type

ES-L1, GEO, HEO, and SSO are the higher successful rate orbit types



Flight Number vs. Orbit Type

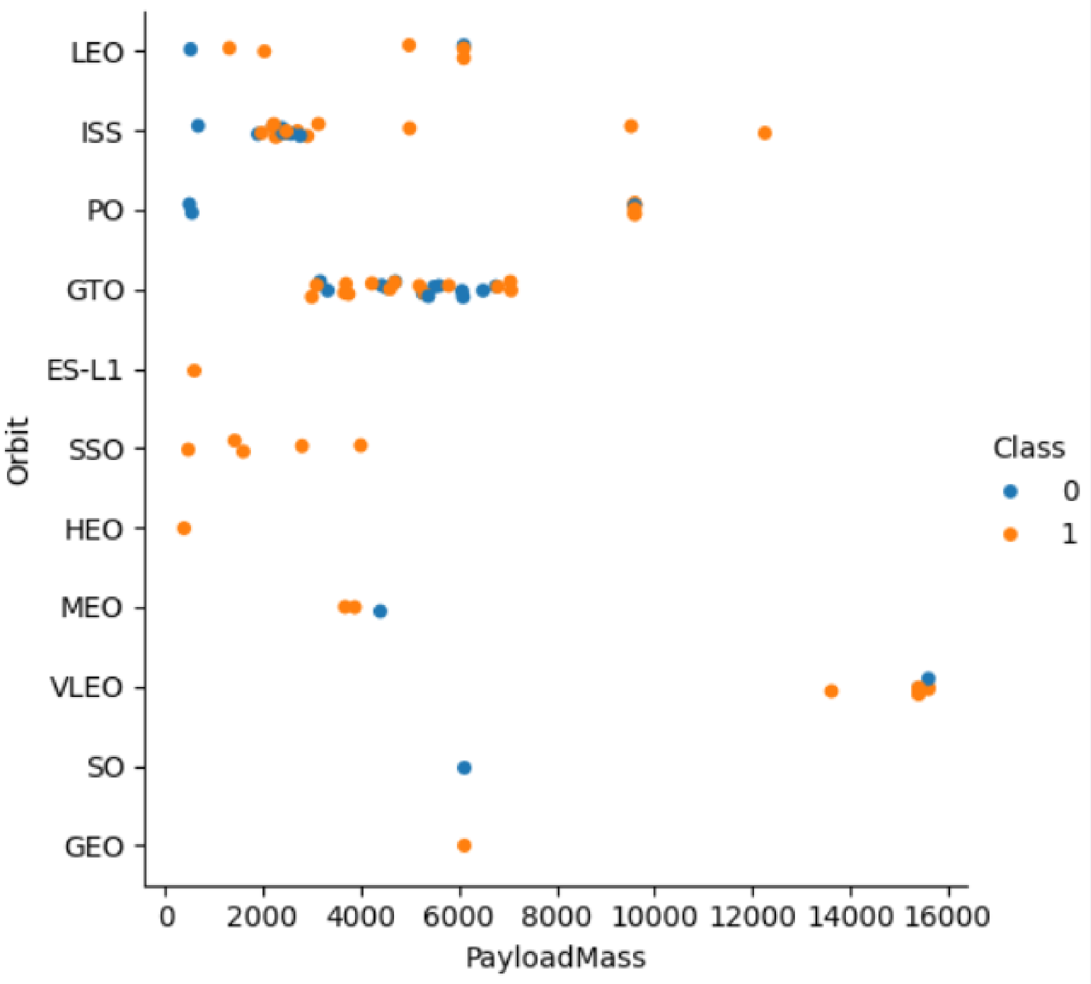
You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

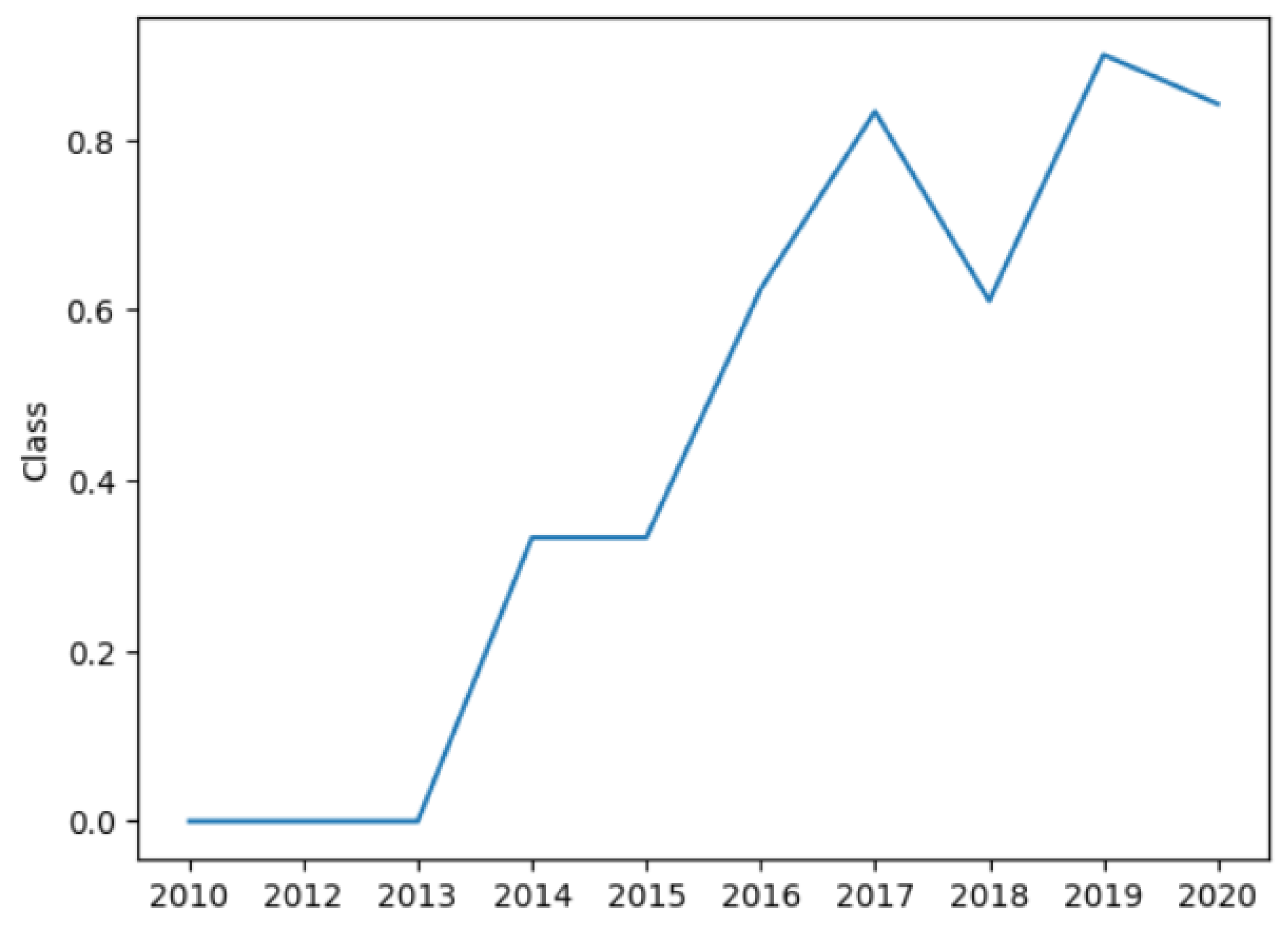
With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

However, for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.



All Launch Site Names

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

- The query is: SELECT DISTINCT(Launch_Site) FROM SPACEXTABLE ;

- Used the distinct operation to remove duplicate names

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

- The query :

```
SELECT * FROM SPACEXTABLE where Launch_Site like 'CCA%' limit 5;
```

Note that like enables the search with regular expression

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%%sql
SELECT sum(PAYLOAD_MASS__KG_) FROM SPACEXTABLE group by Customer having Customer = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT avg(PAYLOAD_MASS__KG_) FROM SPACEXTABLE where Booster_Version like 'F9 v1.0%';
```

```
 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
340.4

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%%sql
SELECT min(Date) FROM SPACEXTABLE where Landing_Outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%%sql
SELECT * FROM SPACEXTABLE limit 5 ;
```

```
 * sqlite:///my_data1.db
Done.
```

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 Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%%sql
SELECT Mission_Outcome, count(*) FROM SPACEXTABLE group by Mission_Outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

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_versions which have carried the maximum payload mass. Use a subquery

```
%%sql
SELECT Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ in (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE);
```

```
 * sqlite:///my_data1.db
Done.
```

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

```
%%sql
SELECT substr(Date, 6, 2) as month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE where substr(Date, 0, 5) = '2015' and Landing_Outcome = "Failure (drone ship)";
```

```
 * sqlite:///my_data1.db
Done.
```

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

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.

```
%%sql
SELECT Landing_Outcome, count(*) as count FROM SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by count desc;
```

* sqlite:///my_data1.db

Done.

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

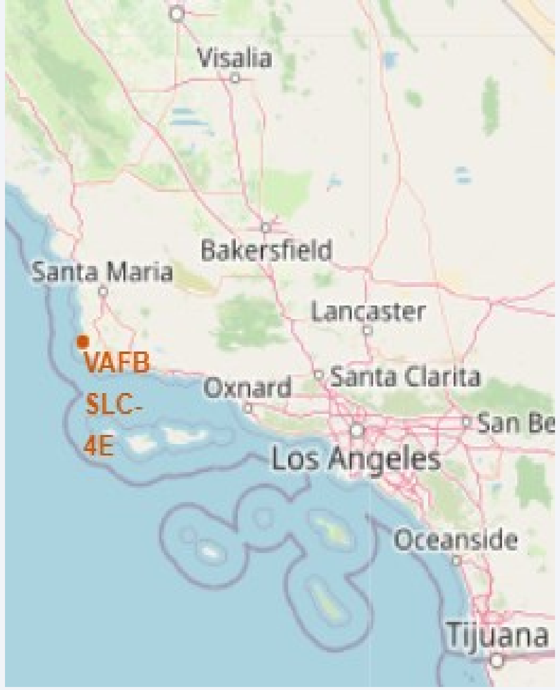
Section 3

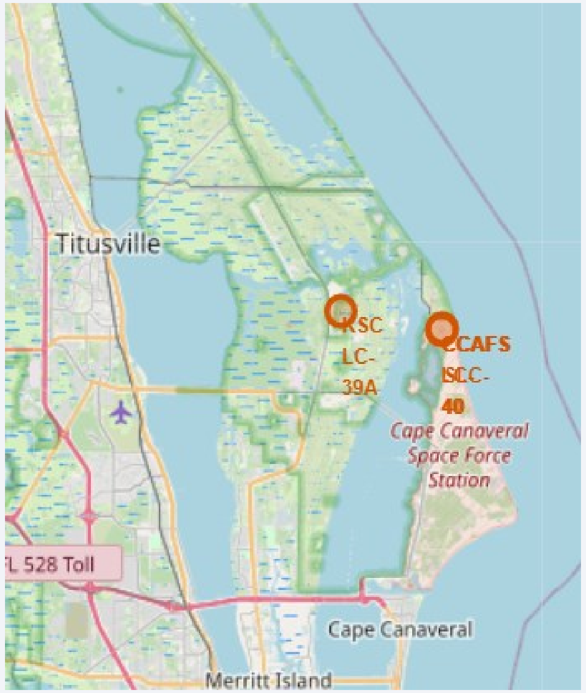
Launch Sites Proximities Analysis

Fist Folium Map

- We can see that all launch sites in very close to the coast

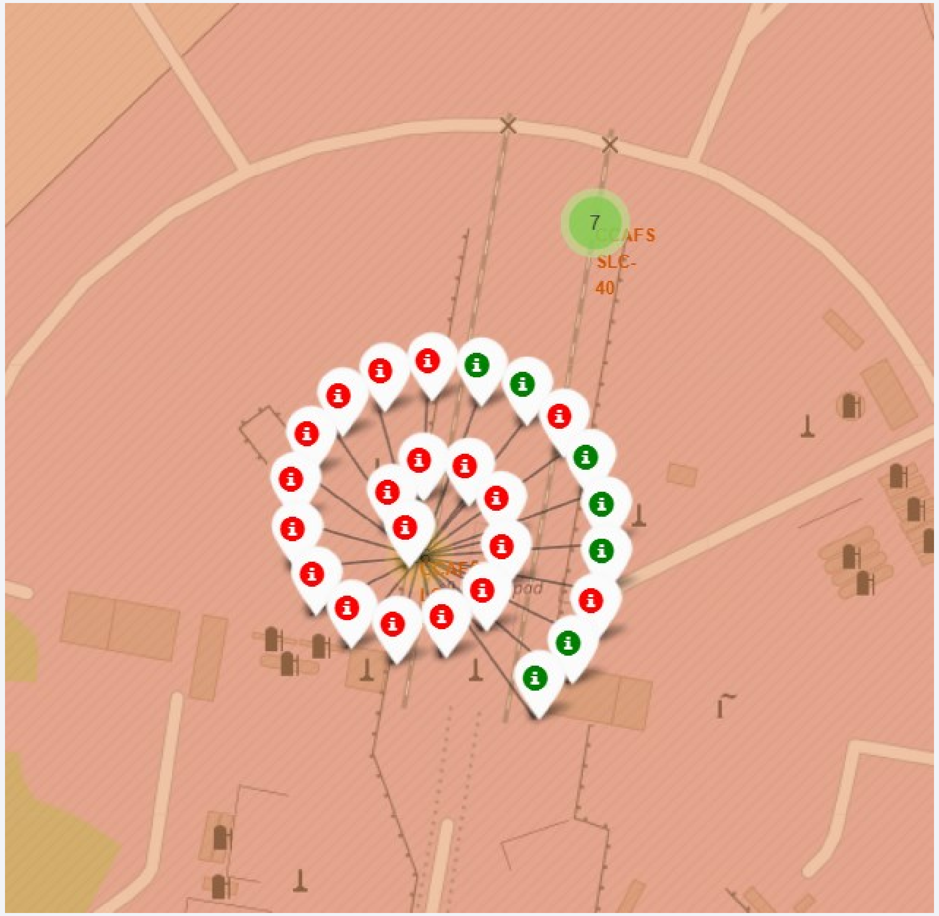
- Also, all launch sites in proximity to the Equator line





Second Folium map

- Enhance the map by Marking the success and failed launches for each site on the map
- assign color to launch outcome green for success, red for failure



Third Folium Map

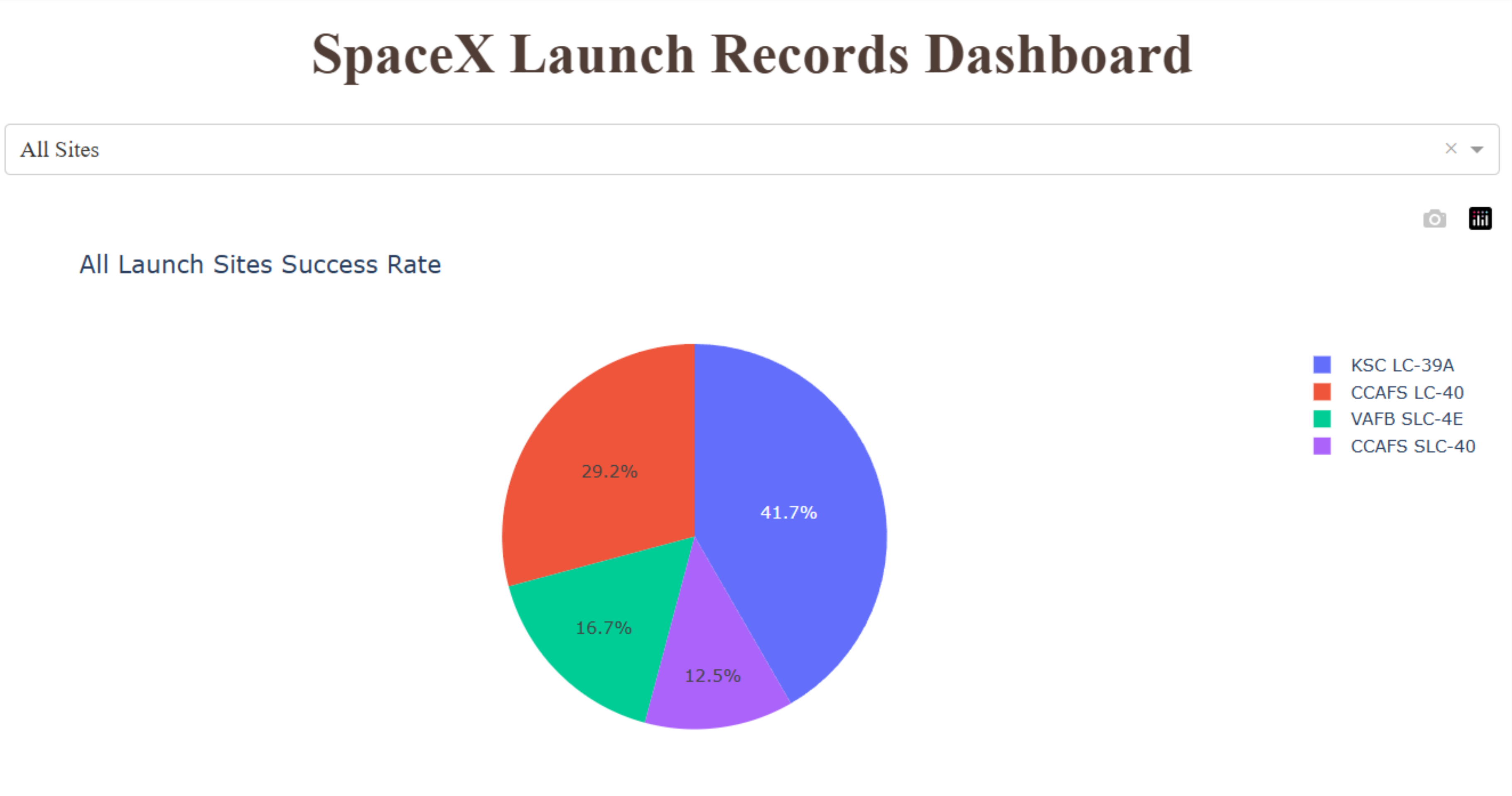
- Enhance the map by adding lines to landmarks
- such as railways, highways, and coastlines
- It’s clear that launch sites keep their distance from cities

Section 4

Build a Dashboard with Plotly Dash

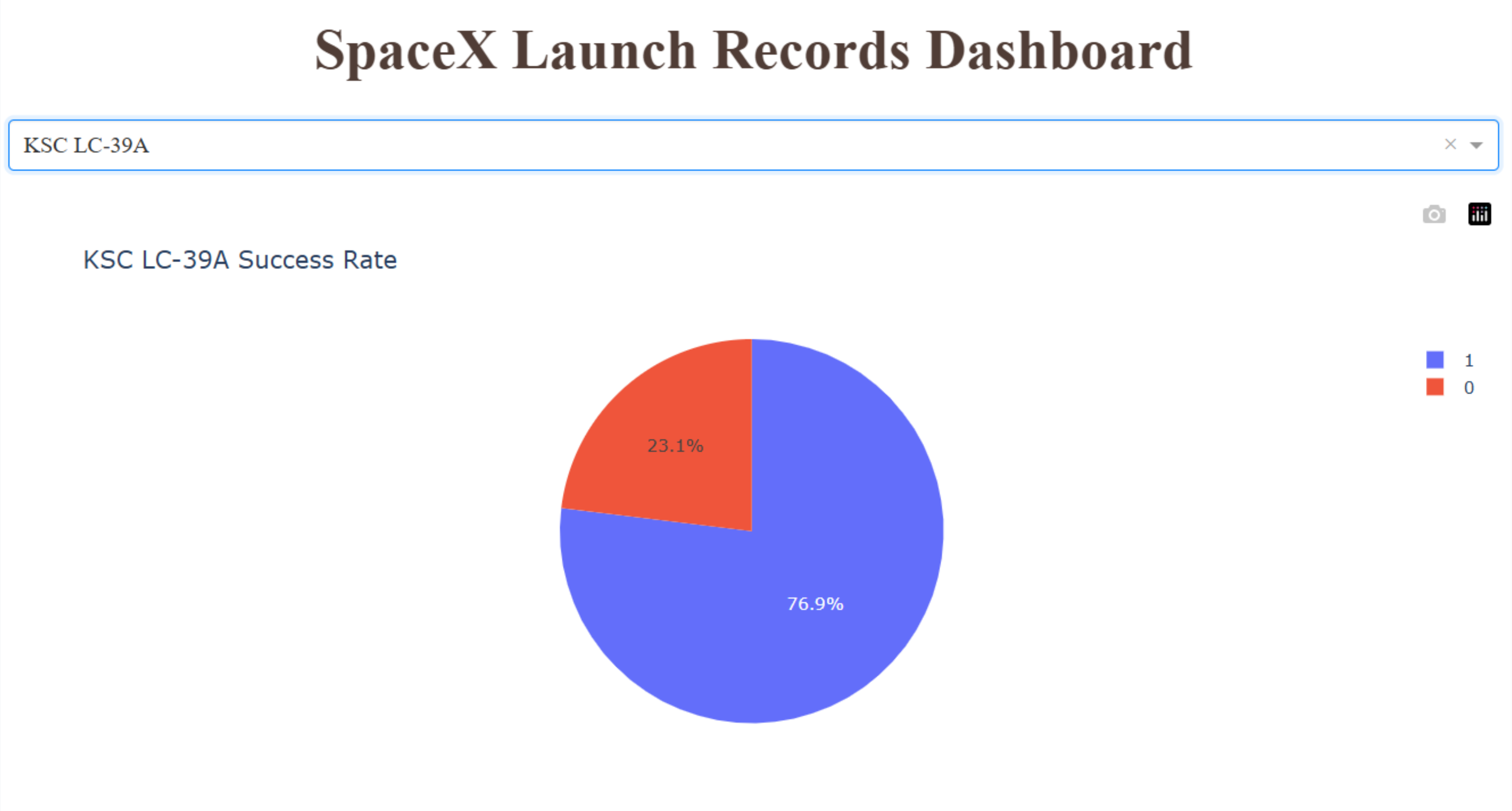
All Launch Sites Success Rates

The KSC LC-39A shows a high success rate followed by CCAS LC-40

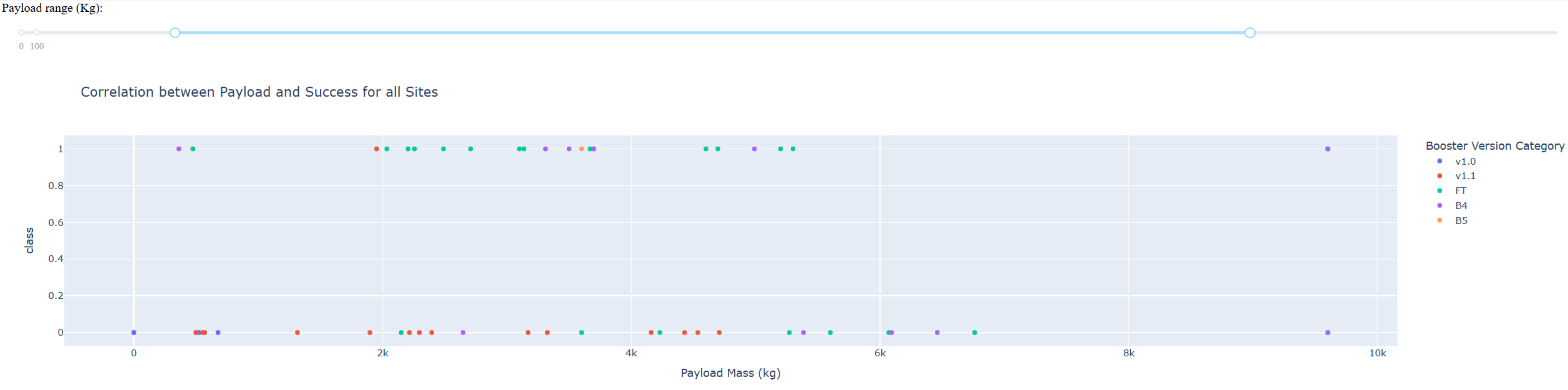


KSC LC-39A Sites Success Rate

in KSC LC-39 77% of the landing attempts were successful



Payload vs. Launch Outcome scatter plot

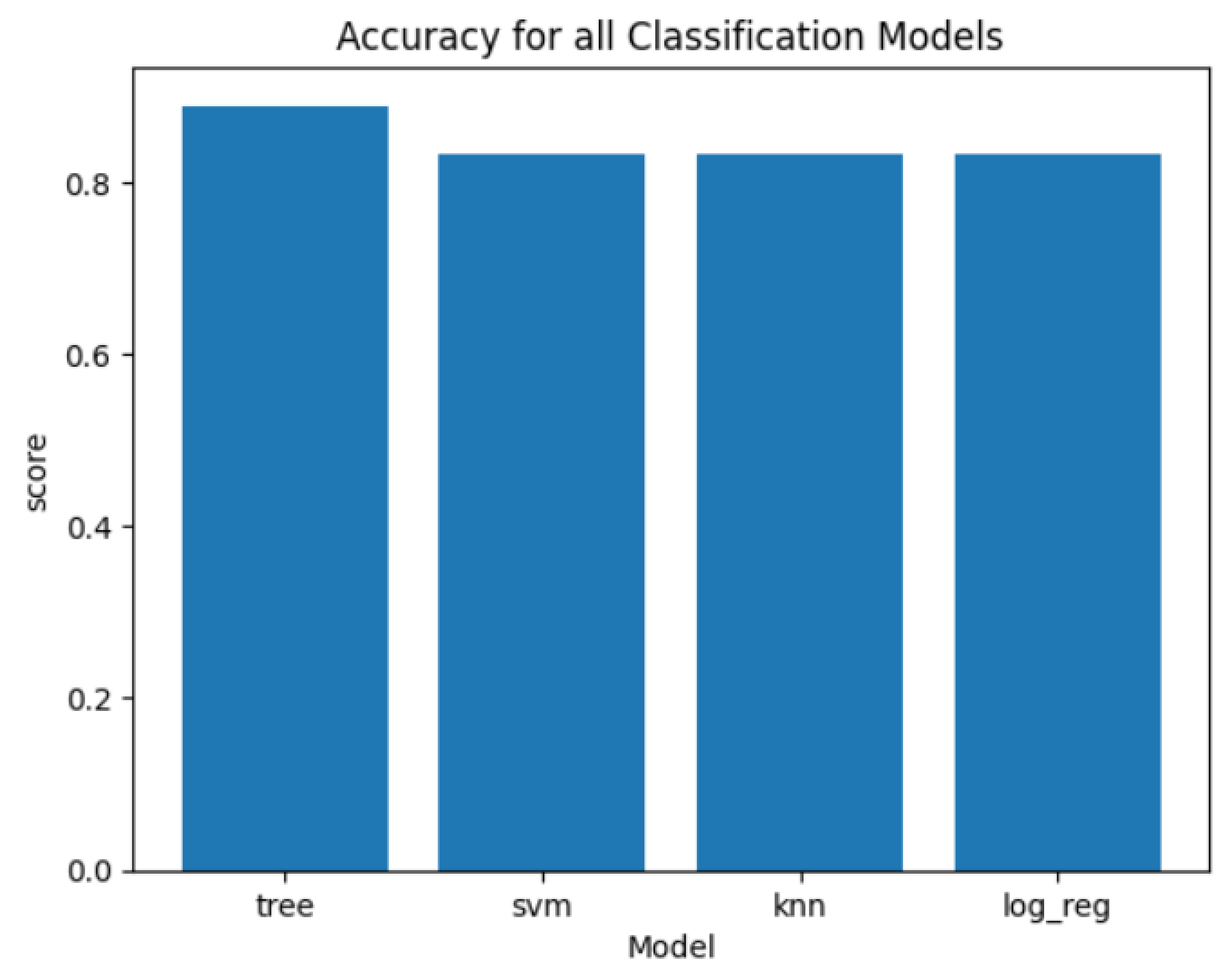


FT Boster with payload mass between 2k – 5k is the best combination for a successful landing

Section 5

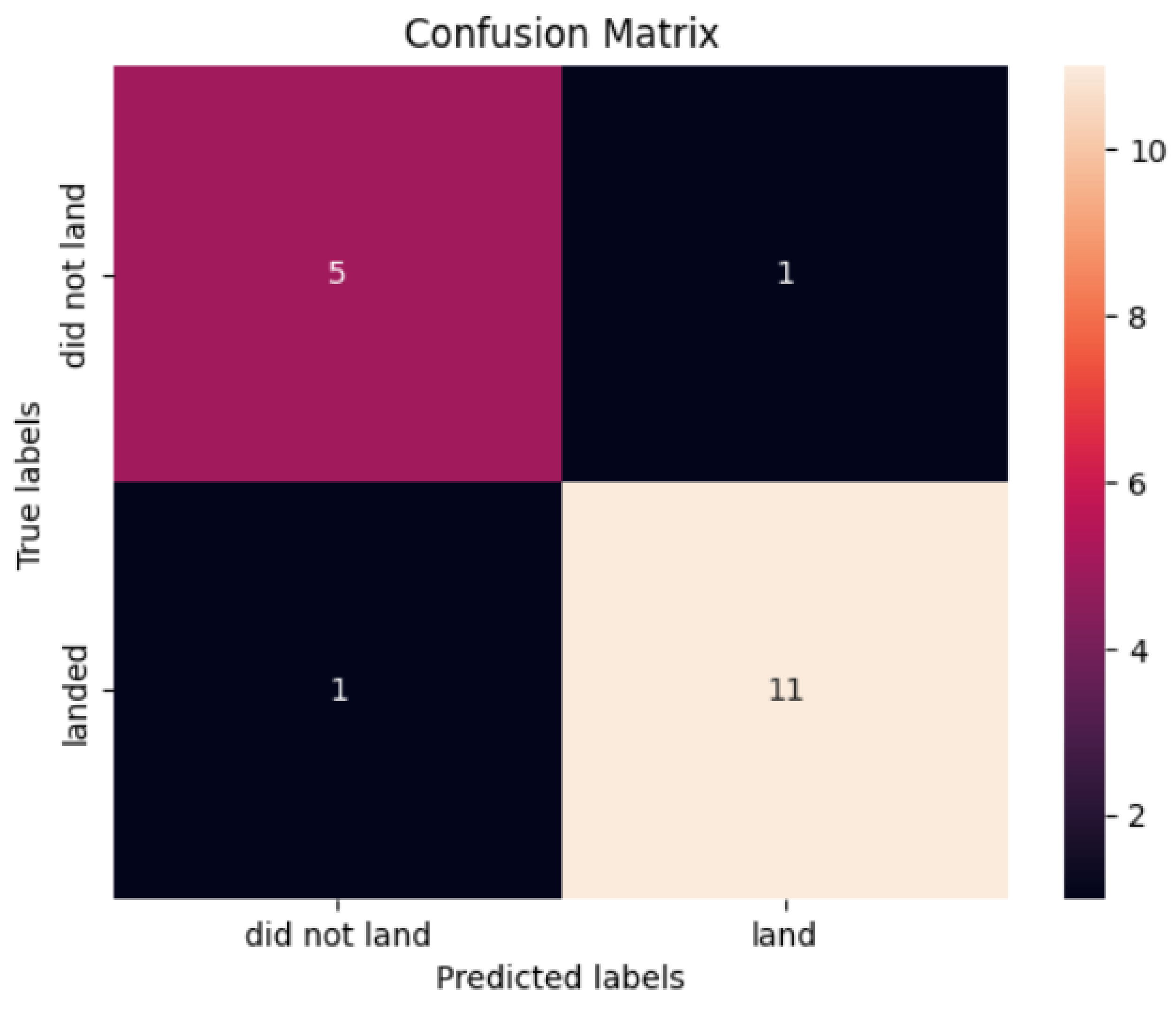
Predictive Analysis (Classification)

Classification Accuracy



From the visual we can infer that Decision tree is the best suited model

Confusion Matrix



- The decision tree shows higher accuracy
- confusion matrix tells the same story with a 2 mislabeled instances
- in contrasted other models mislabeled 3 or more instances

Conclusions

- Machine learning models effectively predict SpaceX Falcon 9 landing outcomes with high accuracy.
- Key factors influencing landing success include booster version, launch site, payload mass, and weather conditions.
- Decision Tree and SVM models performed well, but further optimization can improve predictive accuracy.
- Drone-ship landings had a lower success rate than ground landings due to oceanic conditions and platform stability.
- Heavier payloads were associated with a decreased probability of successful landings.
- Interactive visualizations (Folium maps, Plotly Dash) provided a clear understanding of landing patterns.
- Future improvements could include real-time telemetry data, deep learning models, and additional flight parameters for enhanced predictions.

Appendix

To view the all labs in this course, make sure to visit my GitHub Repository

Thank you!