IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mat
20.08.2023

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection via API
  - Data Collection via Web scraping
  - Data Wrangling Exploratory Analysis Using SQL & Visualization
  - Data Visualization with Folium
  - Interactive Dashboard with Plotly Dash
  - Predictive Analysis (Log Regression, SVM, Decision Tree, KNN)
- **Summary of all results**
  - All models gave similar results

Introduction

- In this project, we will predict if the Falcon 9 first stage will land successfully.
- 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.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

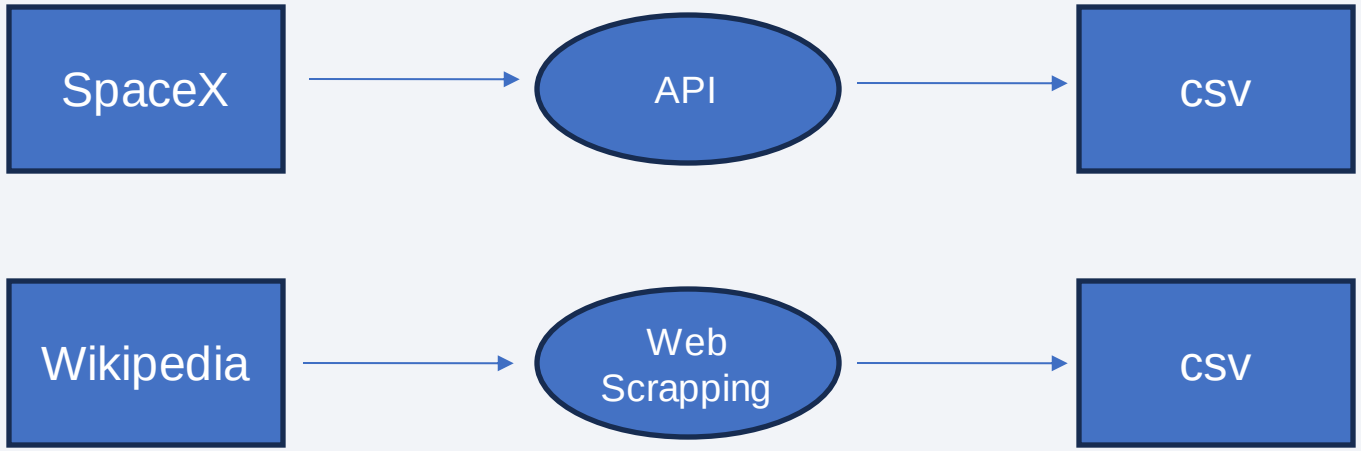
Methodology

Executive Summary

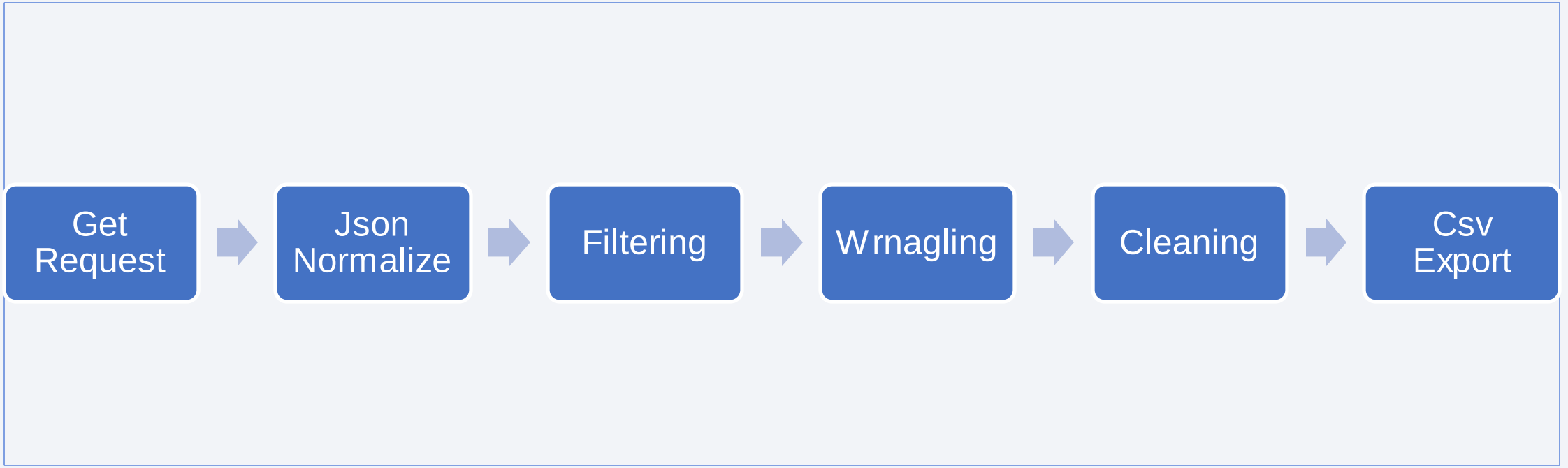
- Data collection methodology:
  - Data has been collected via SpaceX API & Wikipedia Web Scrapping
- Perform data wrangling
  - Data has been reviewed, cleaned, standardized & enriched
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Four classification models: Log Regression, SVM, Decision Tree, KNN
  - Best parameters have been chosen using GridSearch

Data Collection

Data has been collected using SpaceX API and Scrapping Wikipedia

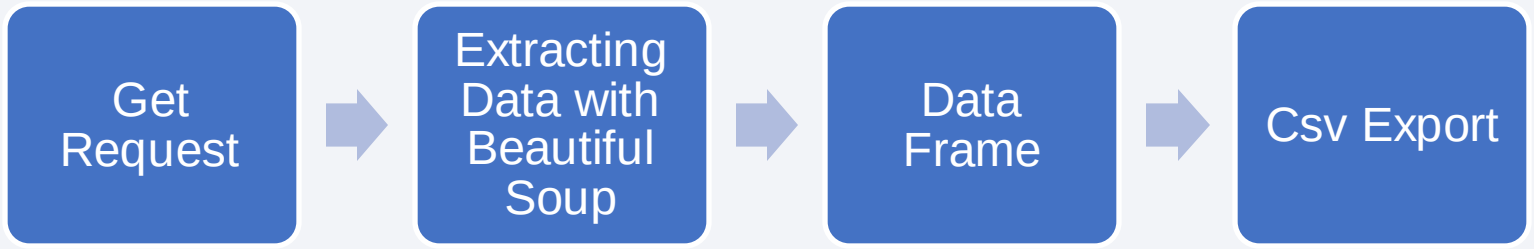


Data Collection – SpaceX API



- https://github.com/Mat-wsw/Coursera---Capstone

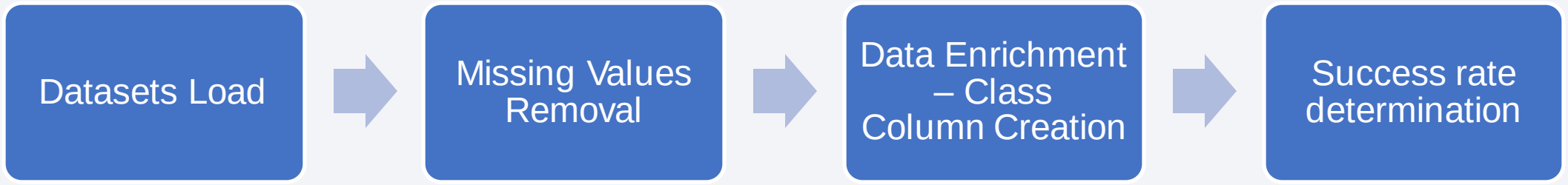
Data Collection - Scraping



- https://github.com/Mat-wsw/Coursera---Capstone

Data Wrangling

Exploratory Data Analysis (EDA) has been conducted to find patterns in the data and determine what would be the label for training supervised models



- https://github.com/Mat-wsw/Coursera---Capstone

EDA with Data Visualization

1) **Flight Number vs. Payload** - to see how the Flight Number (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome.
2) **Flight Number vs Launch Site** – to asses outcome for each site
3) **Payload Vs. Launch Site -** to observe if there is any relationship between launch sites and their payload mass
4) **Bar chart for the success rate of each orbit** – to check if there are any relationship between success rate and orbit type.
5) **Flight Number and Orbit type -** to observe if there is any relationship between Flight Number and Orbit type
6) **Payload and Orbit type -** to observe if there is any relationship between Payload and Orbit type
7) **Launch Success Yearly Trend**

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship booster versions, launch_site for the months in year 2015.
- 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.

Build an Interactive Map with Folium

- The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories.
- Finding an optimal location for building a launch site certainly involves many factors and we could discover some of the factors by analyzing the existing launch site locations.
- Tasks completed:
  - Marked all launch sites on a map with a circle and marker
  - Marked the success/failed launches for each site on the map
  - Calculated the distances between a launch site to its proximities

Build a Dashboard with Plotly Dash

Plots added:

- A Launch Site Drop-down Input Component
- A success-pie-chart based on selected site dropdown
- A Range Slider to Select Payload
- A success-payload-scatter-chart scatter plot

The plots have been added to assess the following:

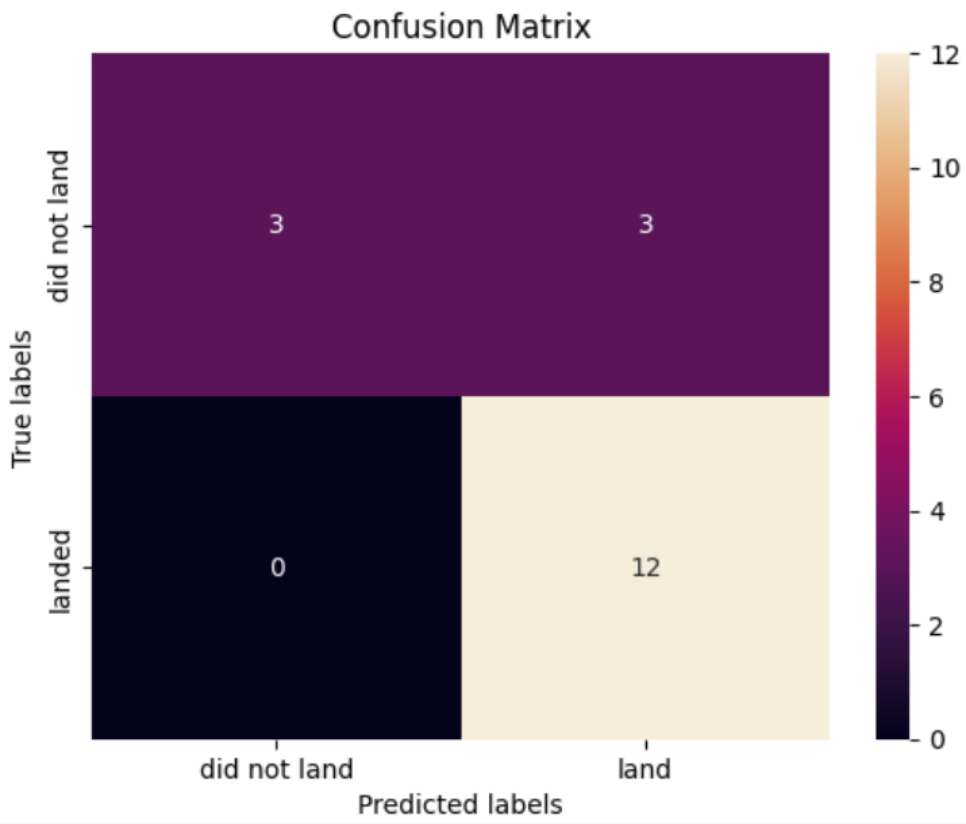
- A site with the largest successful launches
- A site with the highest launch success rate
- A payload range(s) with the highest launch success rate
- A payload range(s) with the lowest launch success rate
- Booster version (v1.0, v1.1, FT, B4, B5, etc.) with the highest launch success rate

Predictive Analysis (Classification)

1) Performed exploratory Data Analysis and determined Training Labels
2) Created a column for the class
3) Standardized the data
4) Split into training data and test data
5) Found best Hyperparameter for SVM, Classification Trees and Logistic Regression
6) Found the method performing best using test data

Results

- All models gave similar results

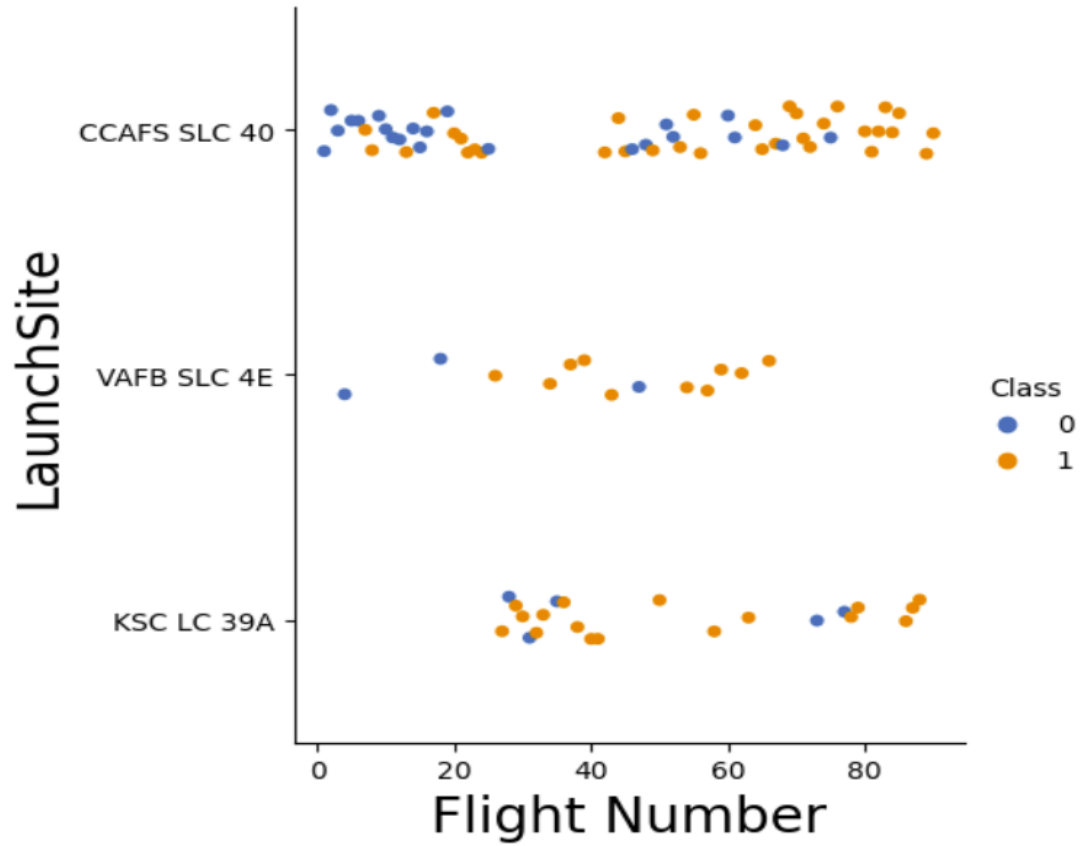


- Accuracy Score = 0.83

Section 2

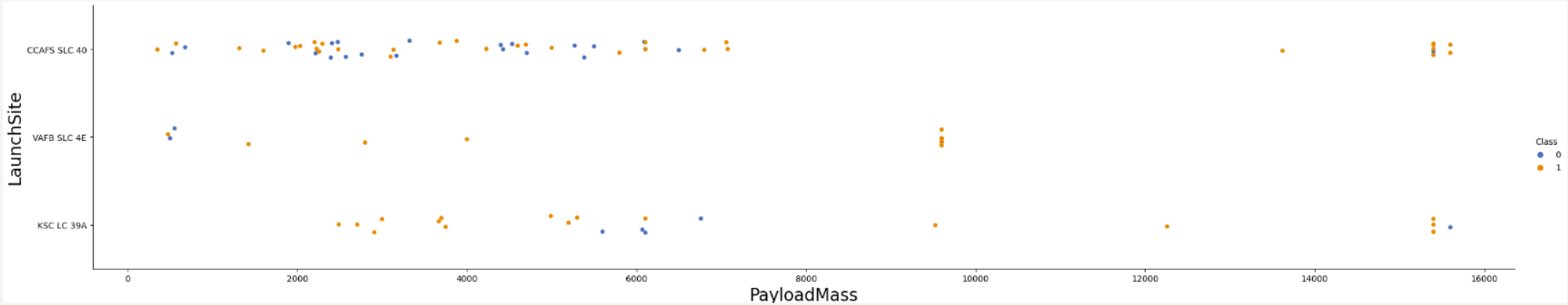
Insights drawn from EDA

Flight Number vs. Launch Site



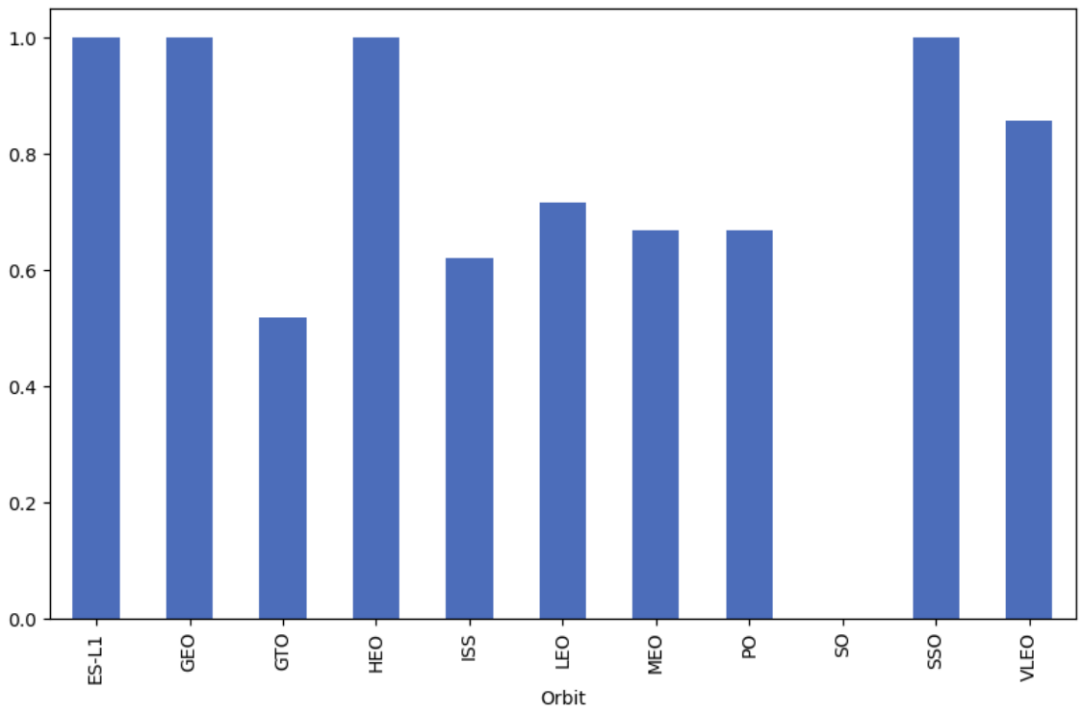
- CCAFS SLC 40 had the most launches
- Succes rate increased with time and flights

Payload vs. Launch Site



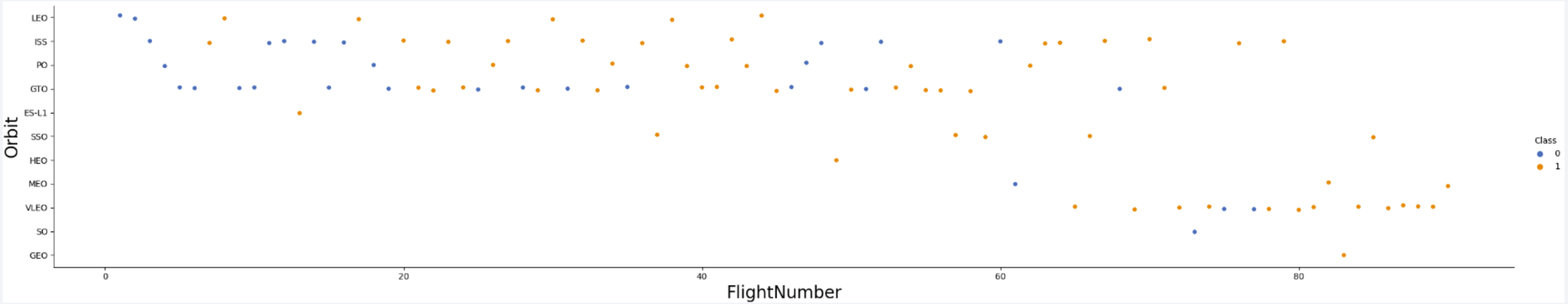
- For AFB-SLC launchsite there are no rockets launched for heavypayload mass
- CCAFS SLC 40 has the biggest payload range

Success Rate vs. Orbit Type



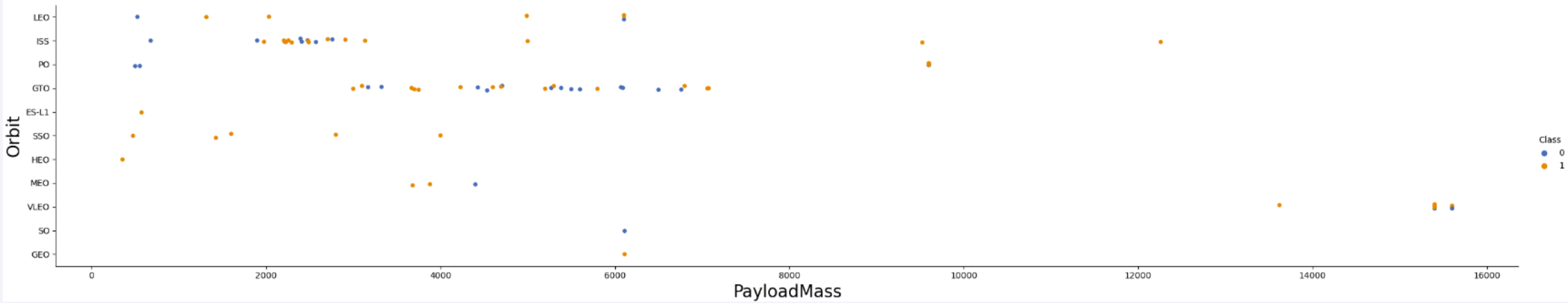
- ES-L1, GEO, HEO, SSO have highest success rates
- SO had no success
- Avg above 0.5

Flight Number vs. Orbit Type



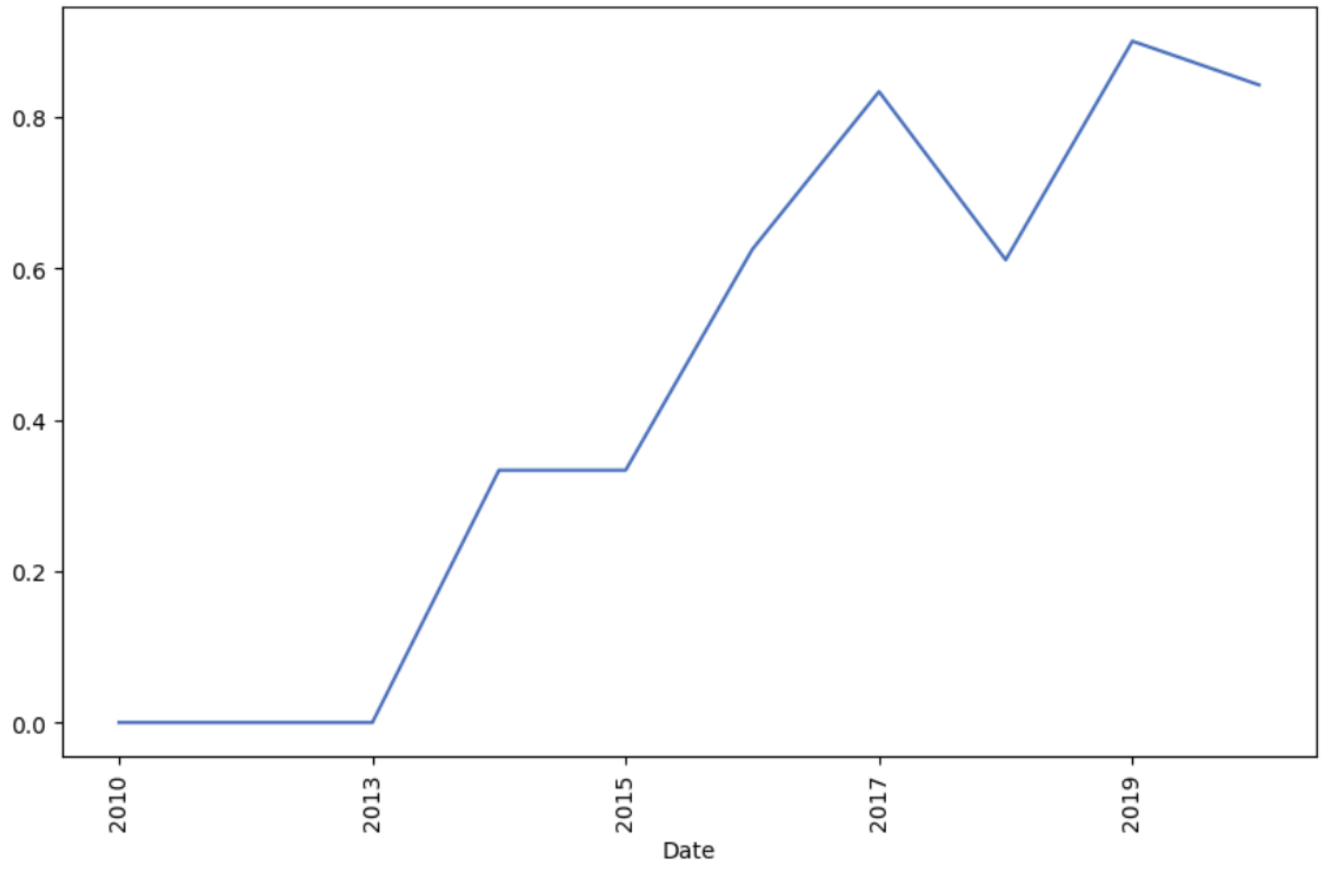
- For LEO orbit the Success appears related to the number of flights
- No relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- For GTO we cannot distinguish this well as both positive landing rate and negative landing are both there.

Launch Success Yearly Trend



- Sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
%sql select distinct "Launch_Site" from SPACEXTABLE
```

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

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

- 4 unique launch sites

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE where "Launch_Site" like "CCA%" LIMIT 5
```

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

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

- All relate to the same launch site

Total Payload Mass

```
%sql select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
```

sum("PAYLOAD_MASS__KG_")
619967

- Seems heavy

Average Payload Mass by F9 v1.1

```
%sql select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Booster_Version" = "F9 v1.1"
```

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

avg("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTABLE where "Landing_Outcome" like "%Success%"
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select Booster_Version from SPACEXTABLE where "Landing_Outcome" = "Success (drone ship)" and "PAYLOAD_MASS__KG_" between 4000 and 6000
```

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

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

- 4 unique boosters

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

- Close to 100% were successful

Boosters Carried Maximum Payload

```
%sql select distinct "Booster_Version", "PAYLOAD_MASS__KG_" from SPACEXTABLE where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTABLE)
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

```
%sql select substr(Date, 6, 2) as month, substr(Date,0,5) as year, "Landing_Outcome", "Booster_Version", "Launch_Site" from SPACEXTABLE where year = "2015" and "Landing_Outcome" = "Failure (drone ship)"
```

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

month	year	Landing_Outcome	Booster_Version	Launch_Site
10	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

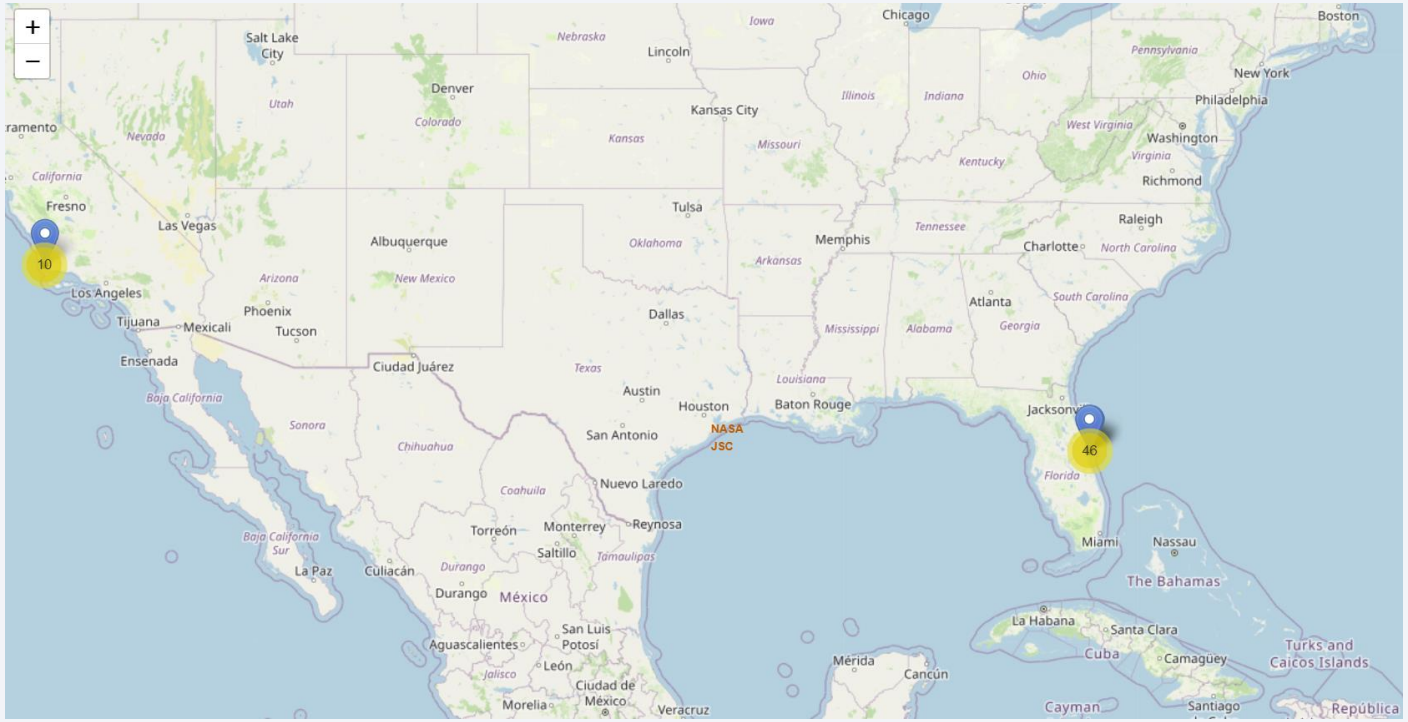
```
%sql select date, "Landing_Outcome", count(*) as "count" from SPACEXTABLE where date between "2010-06-04" and "2017-03-20" group by "Landing_Outcome" order by "count" desc
```

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

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

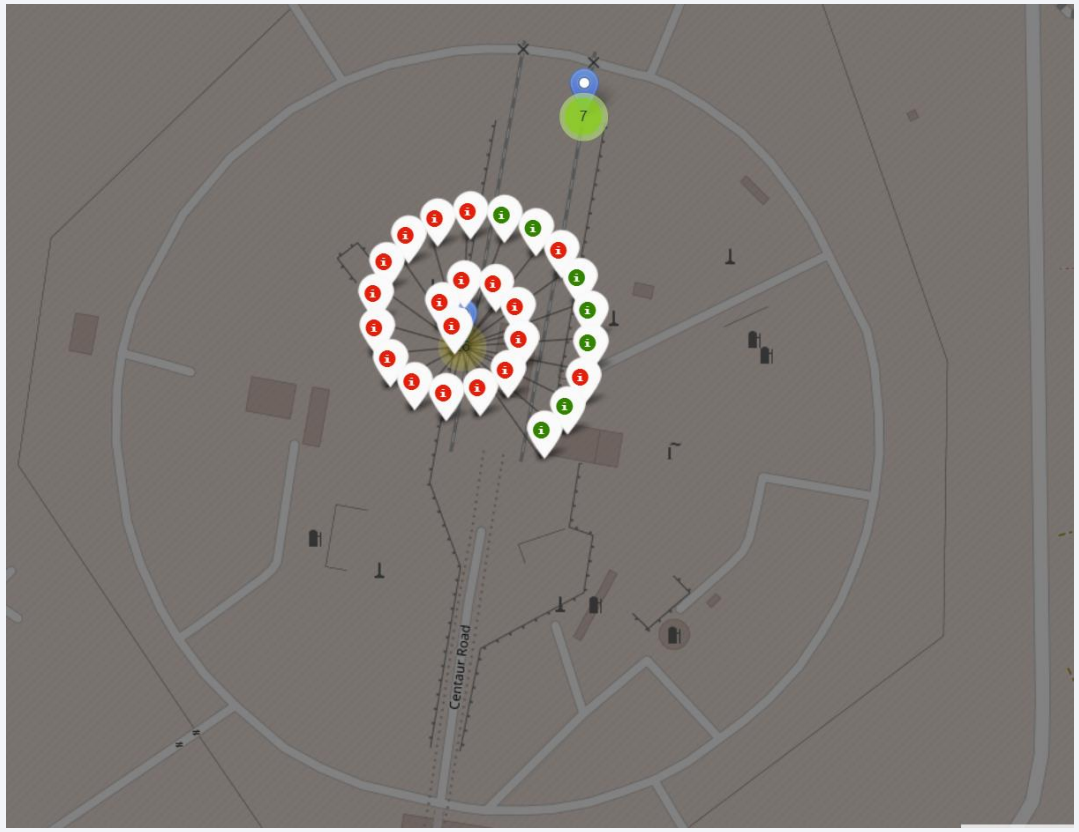

Section 3
Launch Sites
Proximities Analysis

Launch Site Locations



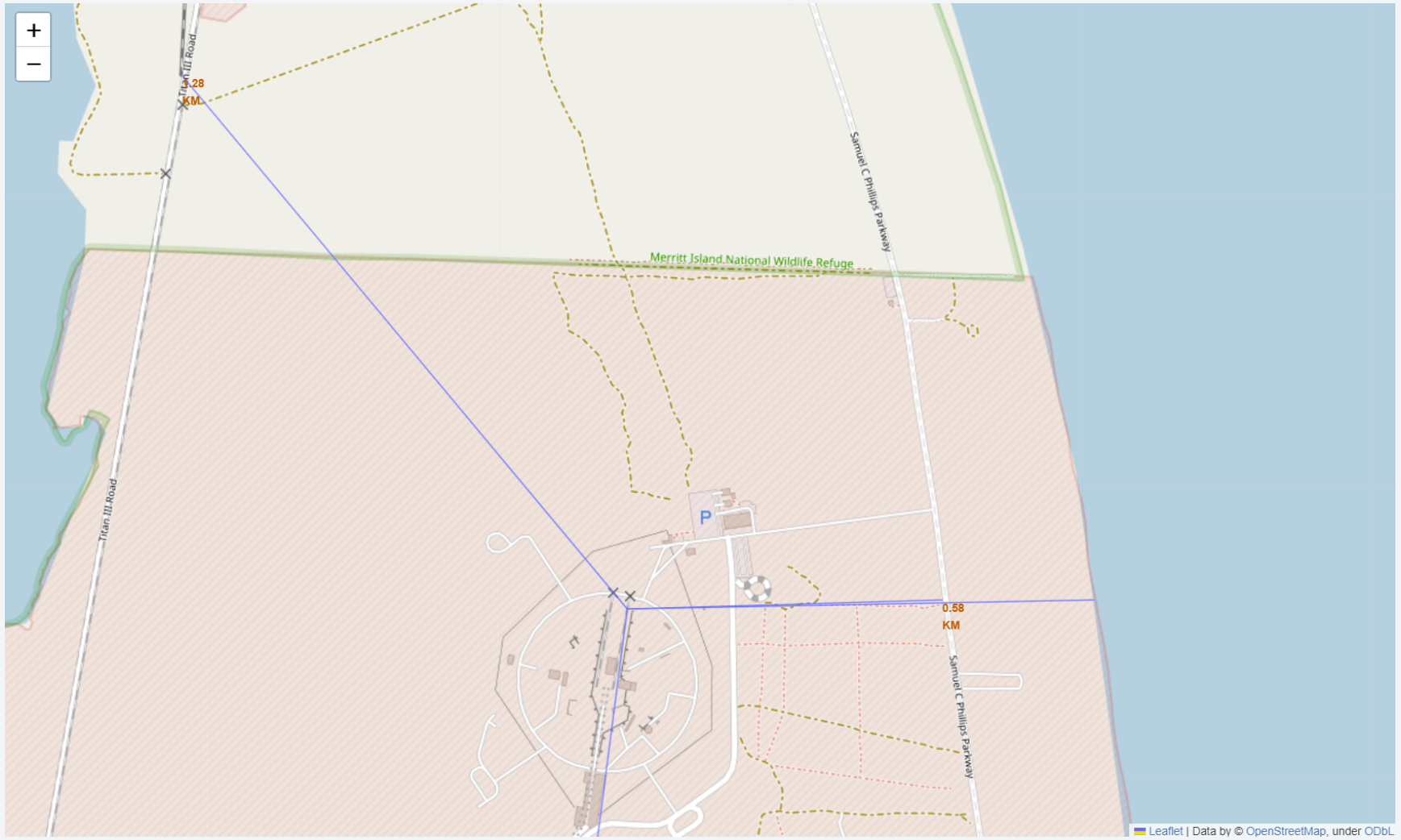
- All launches concentrated in two areas
- Majority were launched on east coast
- 56 launches in total

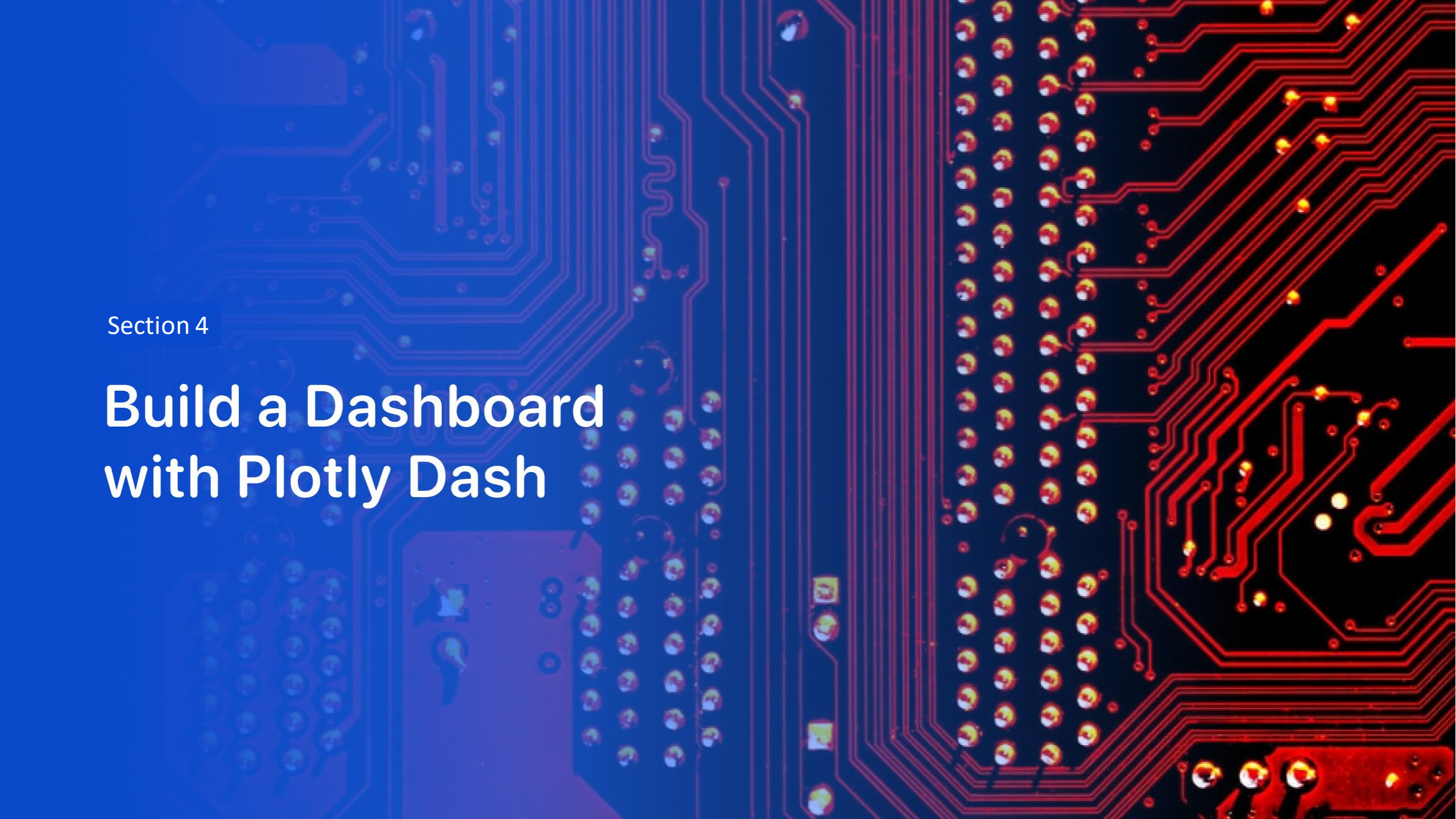
Launch Site Locations – Launch Outcome



- Majortiy on this site was unsuccesful

<Folium Map Screenshot 3>

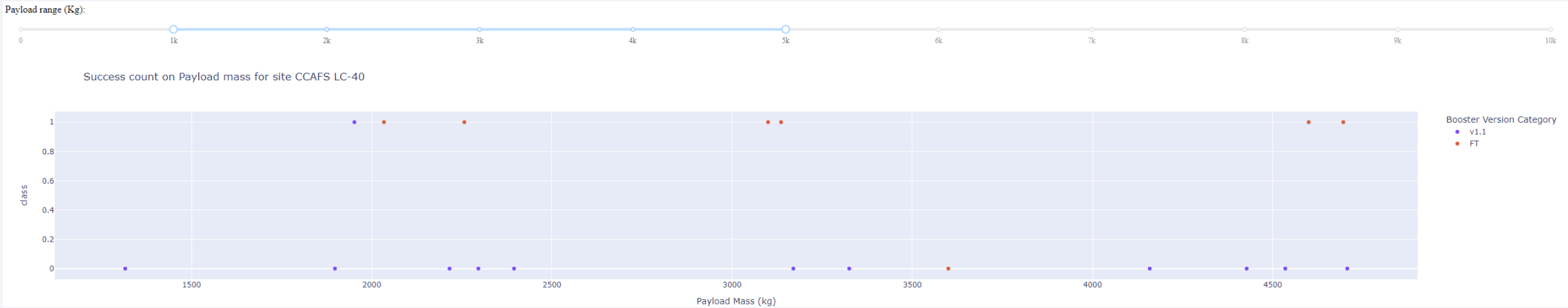



Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

SpaceX Launch Records Dashboard – Highest Success Ratio

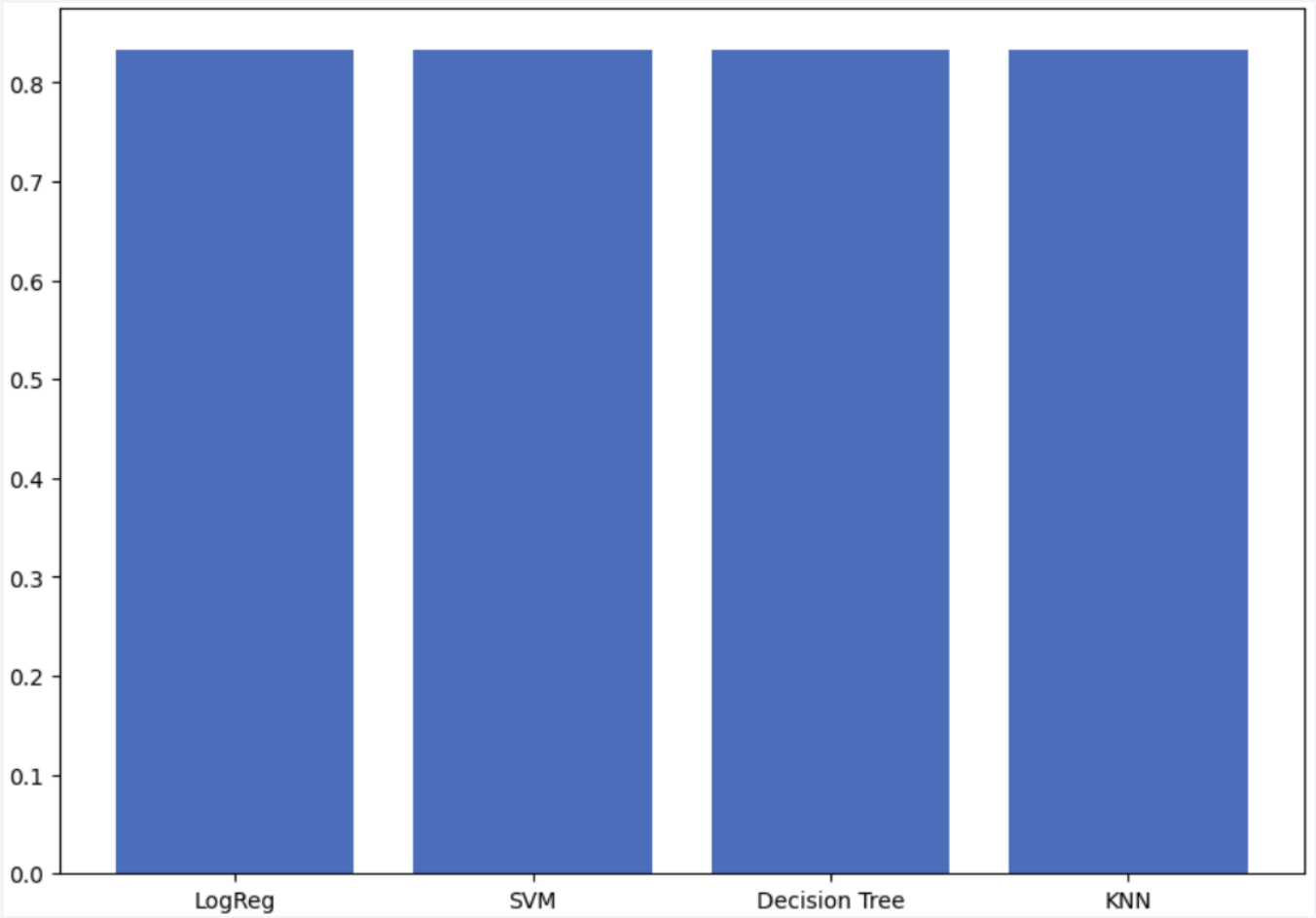
SpaceX Launch Records Dashboard - Payloads



Section 5

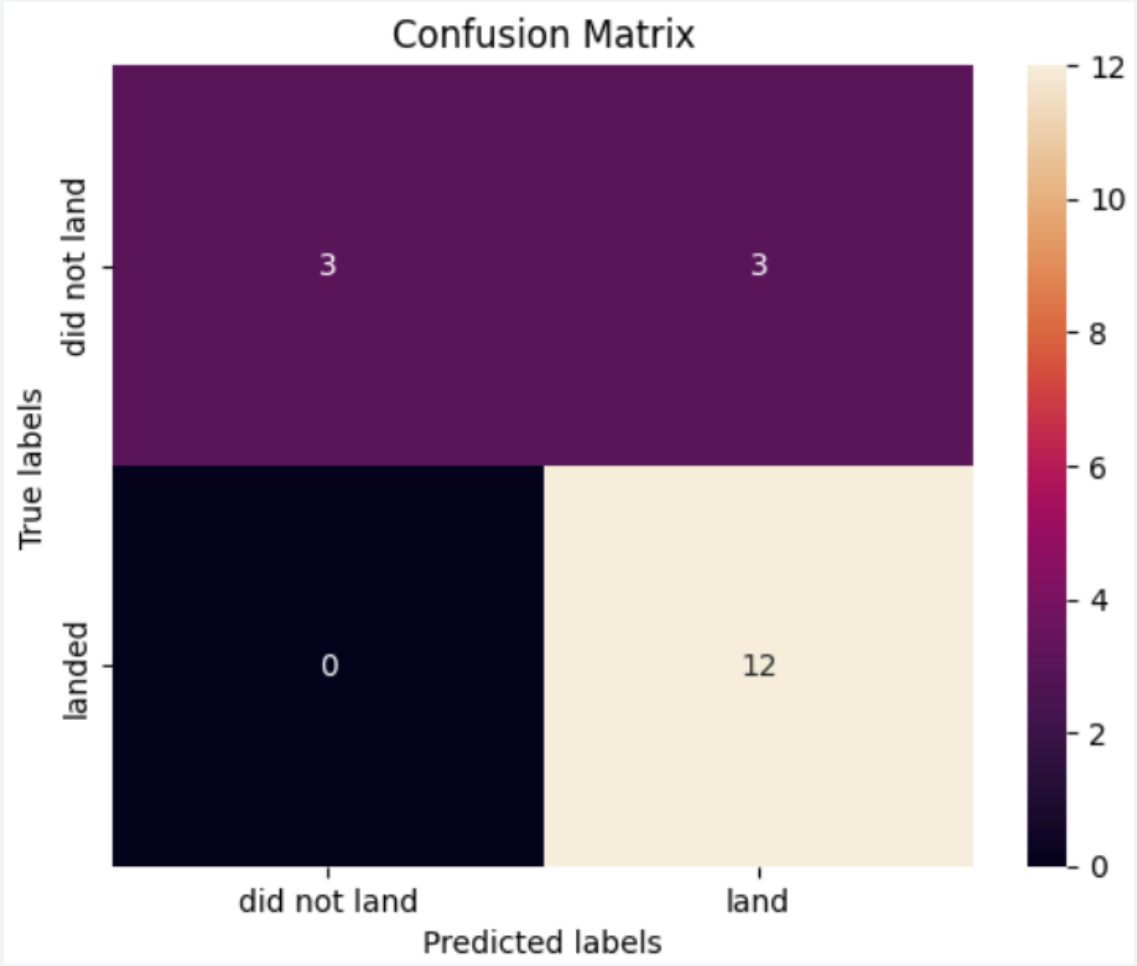
Predictive Analysis (Classification)

Classification Accuracy



- All models have the same accuracy
- Accuracy = 0.83

Confusion Matrix



- All models have the same confusion matrix
- 3 false positives (showing land while in fact it did not)

Conclusions

- All models resulted in the same confusion matrix
- All models have the same accuracy of 0.83
- Proposed models can be used to predict successful landing

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!