IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jeff Barlow-Spady
September 2024

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Accurate landing predictions can lead to cost savings by enabling the reuse of rocket stages, impacting competitive bidding for launches.

Logistic Regression and SVM models show promising accuracy, enhancing prediction reliability.

Introduction

Project background and context

- Space X 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 Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

Problems you want to find answers

- What factors determine if the rocket will land successfully?
- The interaction amongst various features that determine the success rate of a successful landing.
- What operating conditions needs to be in place to ensure a successful landing program.

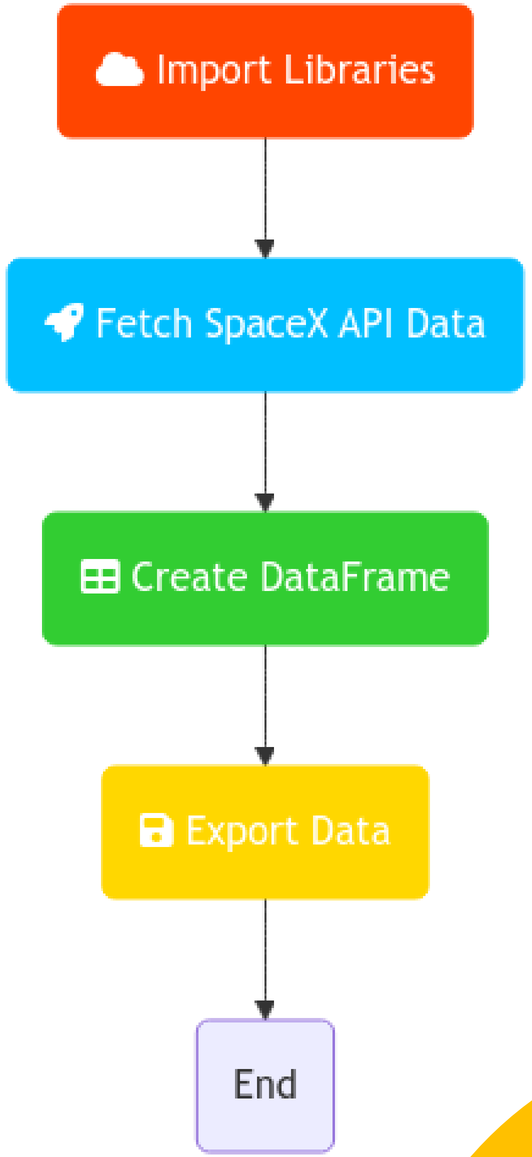
Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

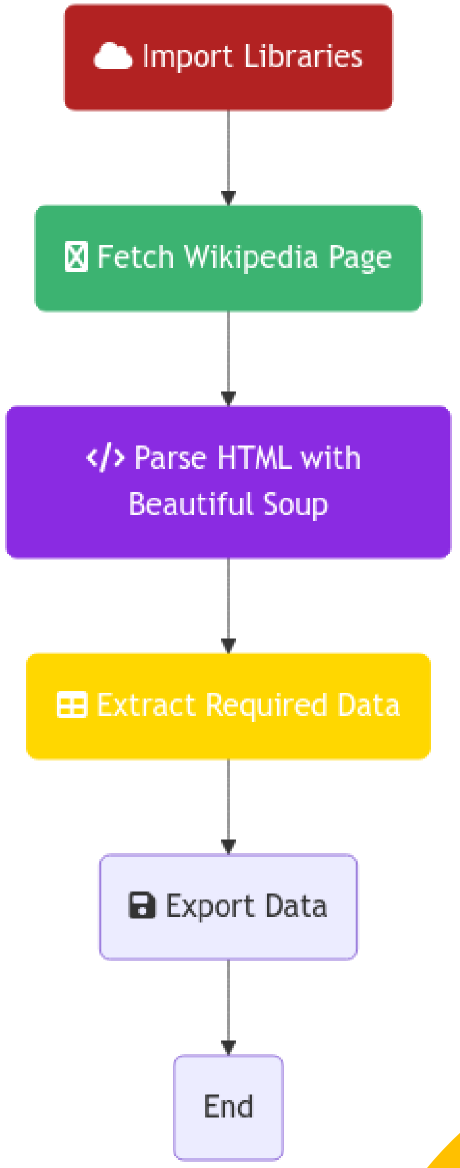
Data Collection – SpaceX API



Data was sourced from SpaceX's API, focusing on rocket launch metrics such as payload mass and landing outcomes

https://github.com/Jeff-Barlow-Spady/notebooks/blob/master/final/files/ipynb/jupyter-labs-spacex-data-collection-api%20(1).ipynb

Data Collection - Scraping



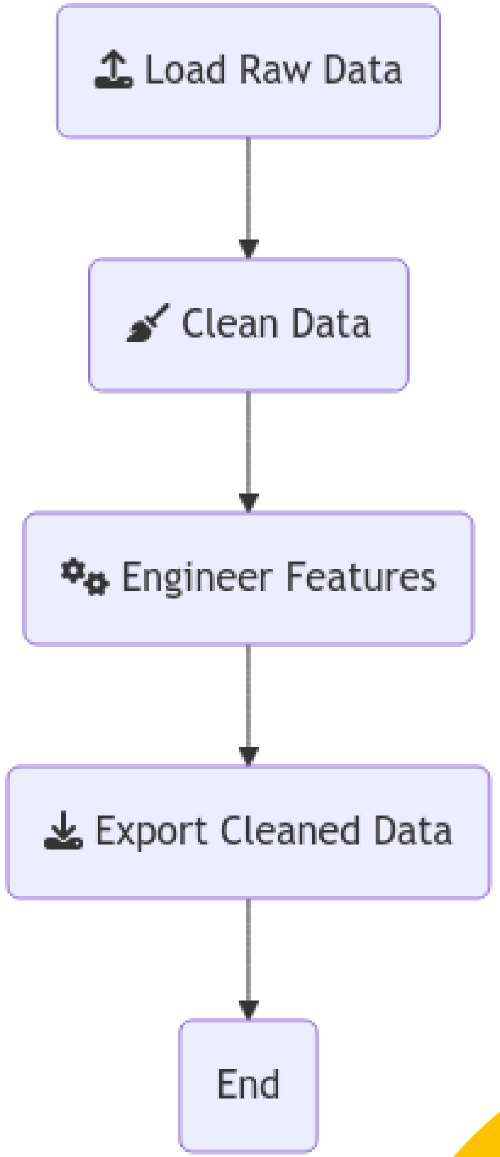
Data was sourced from SpaceX's wikipedia, focusing on rocket launch metrics such as payload mass and landing outcomes

https://github.com/Jeff-Barlow-Spady/notebooks/blob/master/final/files/ipynb/jupyter-labs-spacex-data-collection-api%20(1).ipynb

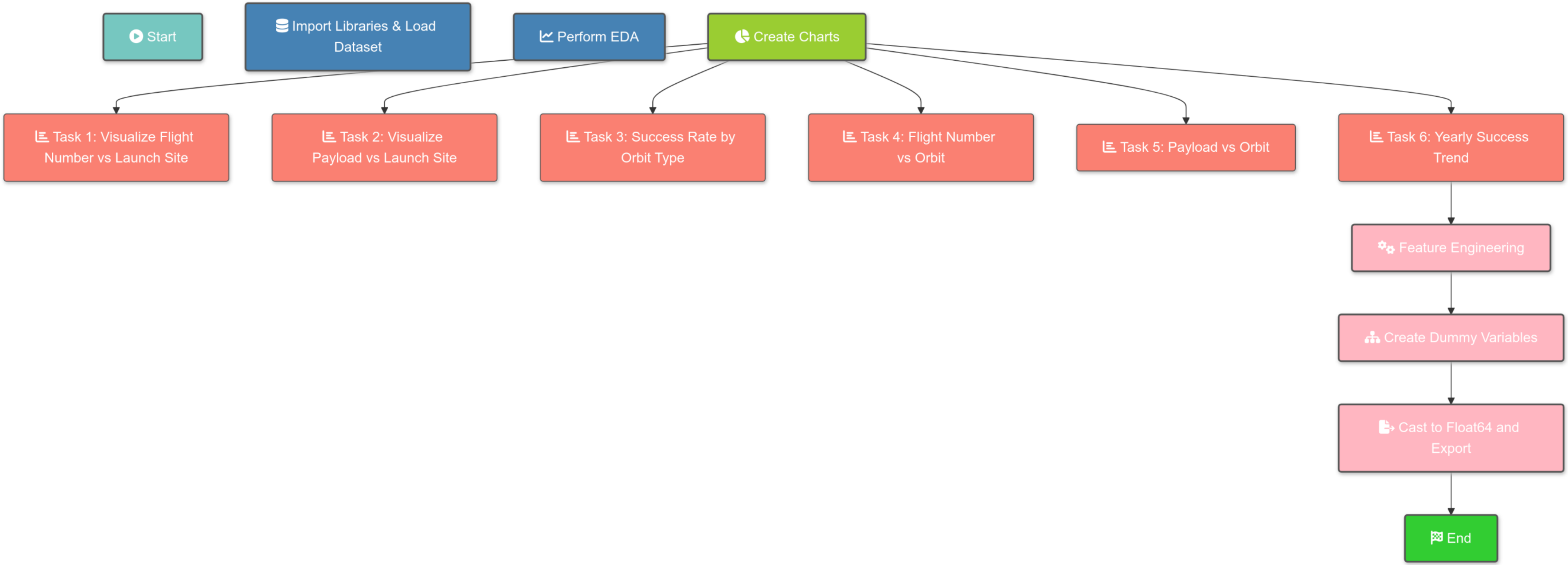
Data Wrangling

Utilized pandas for cleaning and preparing the data by handling missing values and transforming features

https://github.com/Jeff-Barlow-Spady/notebooks/blob/master/final/files/ipynb/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization



https://github.com/Jeff-Barlow-Spady/notebooks/blob/master/final/files/ipynb/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

[Task 1: Retrieve Unique Launch Sites];

[Task 2: Filter Launch Sites by Prefix];

[Task 3: Calculate Payload Mass for NASA];

[Task 4: Average Payload Mass for Booster v1.1];

[Task 5: First Successful Ground Landing Date];

[Task 6: Select Boosters with Criteria];

[Task 7: Successful vs Failed Missions];

[Task 8: Max Payload Mass Boosters];

[Task 9: Filter 2015 Failures by Month];

[Task 10: Rank Landing Outcomes];

https://github.com/Jeff-Barlow-Spady/notebooks/blob/master/final/files/ipynb/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Functions from the Folium libraries are used to visualize the data through interactive maps.
- The Folium library is used to:
  - Mark all launch sites on a map
  - Mark the succeeded launches and failed launches for each site on the map
  - Mark the distances between a launch site to its proximities such as the nearest city, railway, or highway

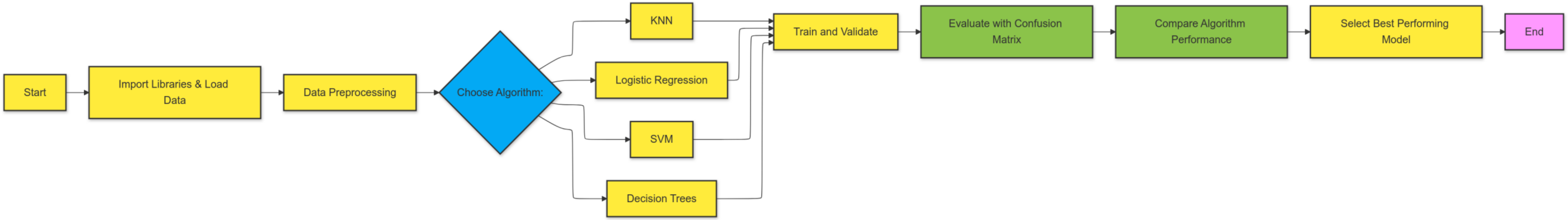
Build a Dashboard with Plotly Dash

Created an interactive, styled dashboard to communicate insights

Pie Chart With Success Rate For Launch Sites

Scatter chart showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.

Predictive Analysis (Classification)



https://github.com/Jeff-Barlow-Spady/DS-NB/blob/master/notebooks/final/dash/spacex_dash_app%20(1).py

Results

EXPLORATORY DATA ANALYSIS RESULTS

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

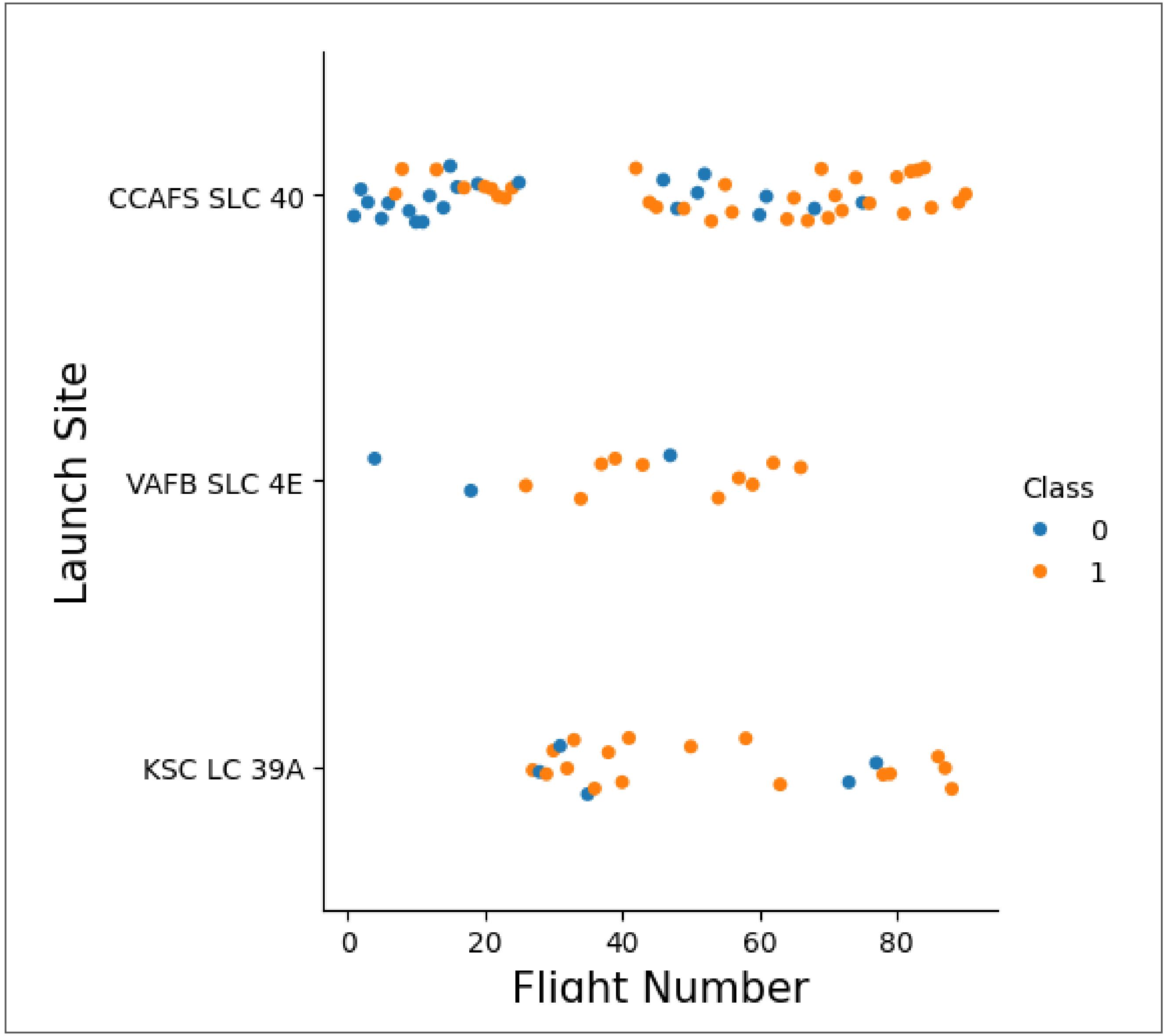
PREDICTIVE ANALYSIS RESULTS

Section 2

Insights drawn from EDA

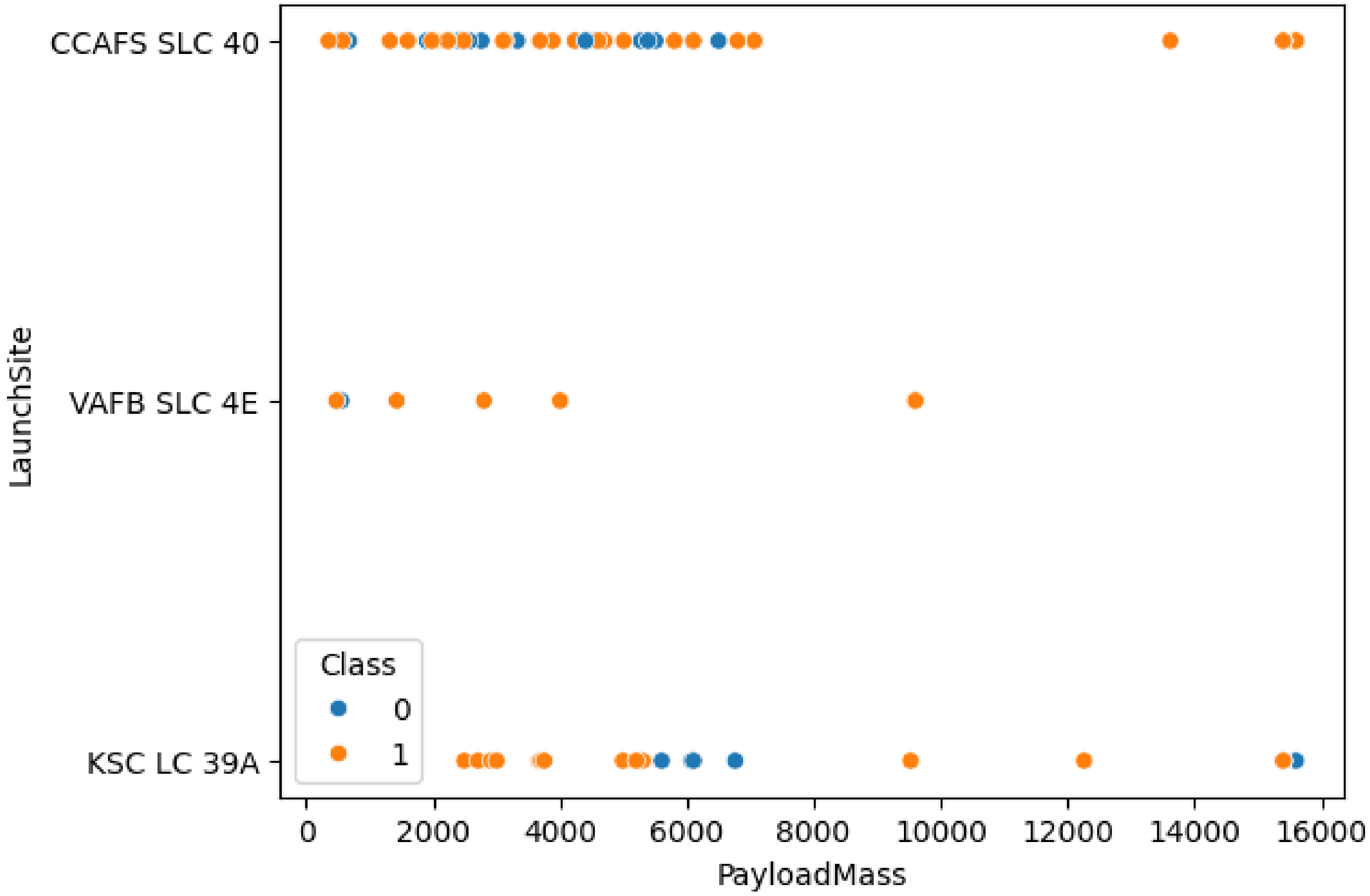
Flight Number vs. Launch Site

CHART DEMONSTRATING CORRELATION BETWEEN FLIGHT NUMBER AND LAUNCHSITE



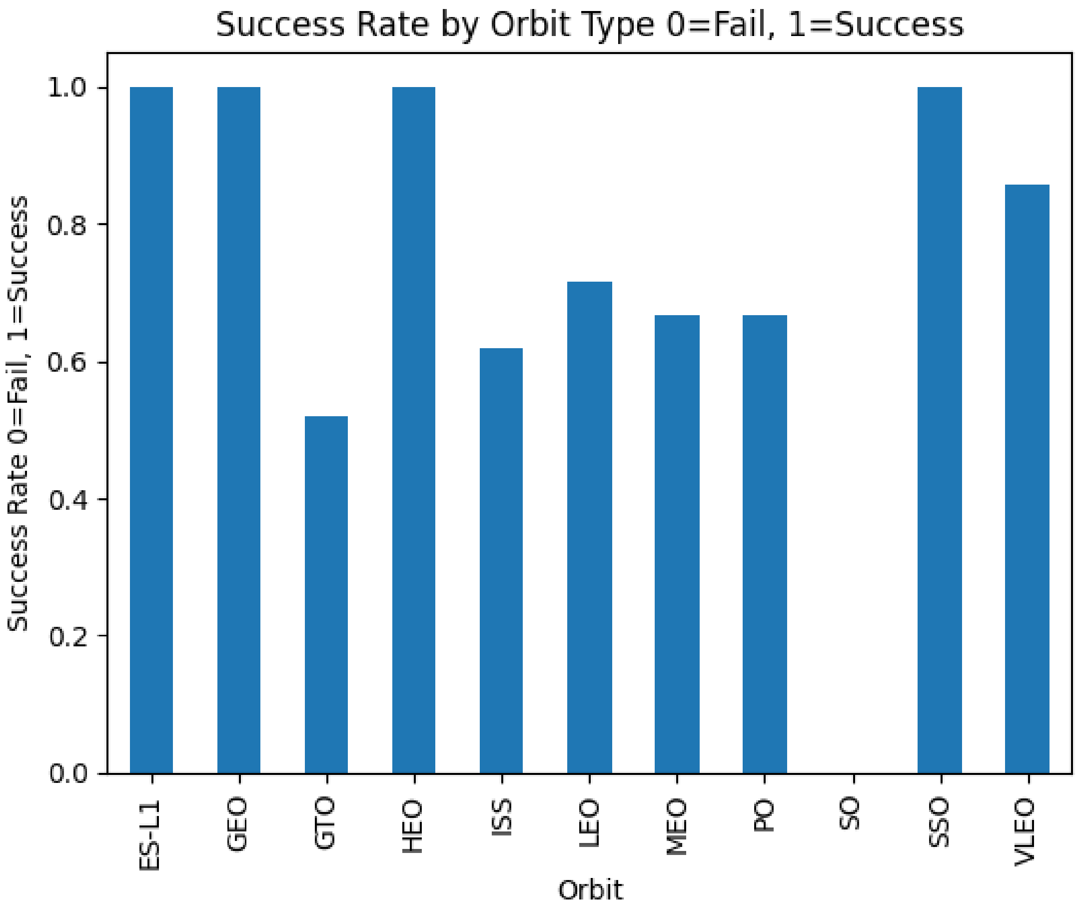
Payload vs. Launch Site

Relationship between launch site and payload size



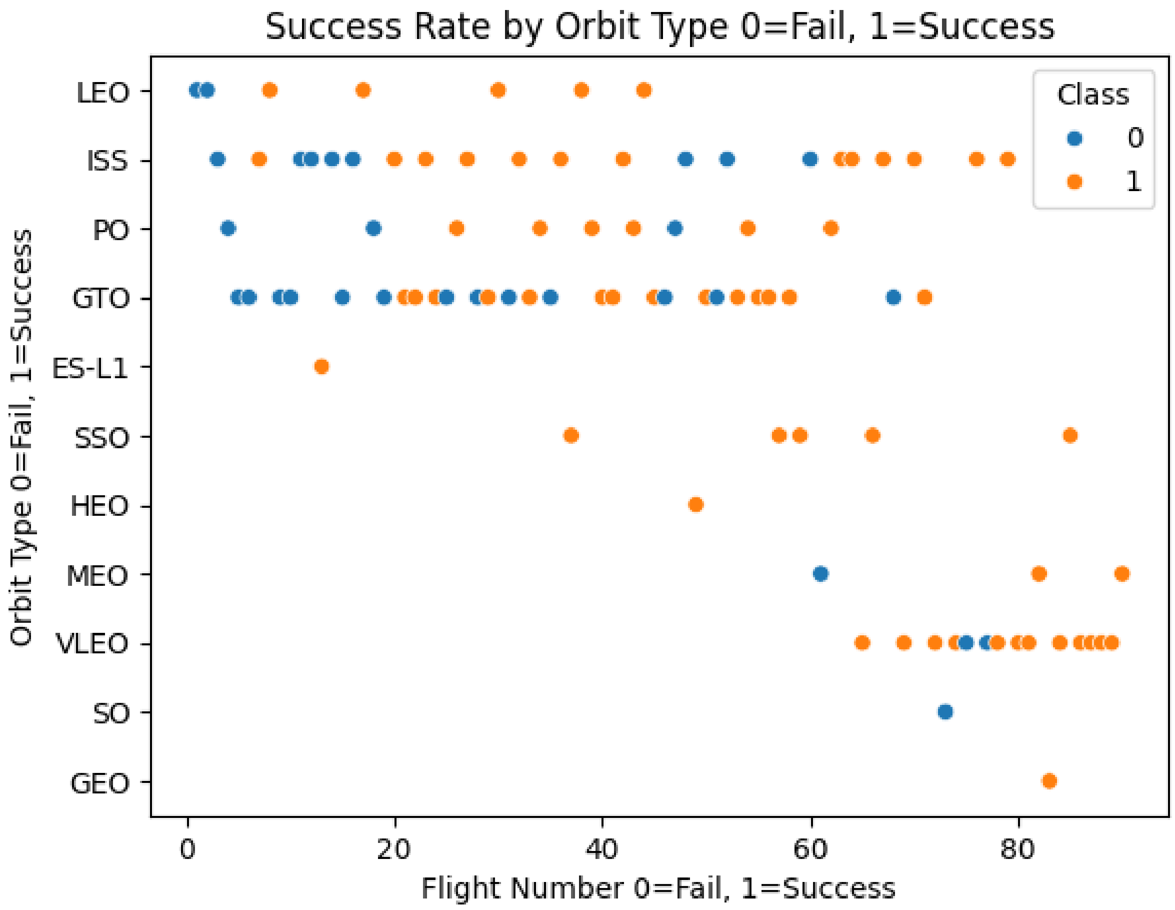
Success Rate vs. Orbit Type

ES-L1, GEO, ISS and SSO had the heightest success rate



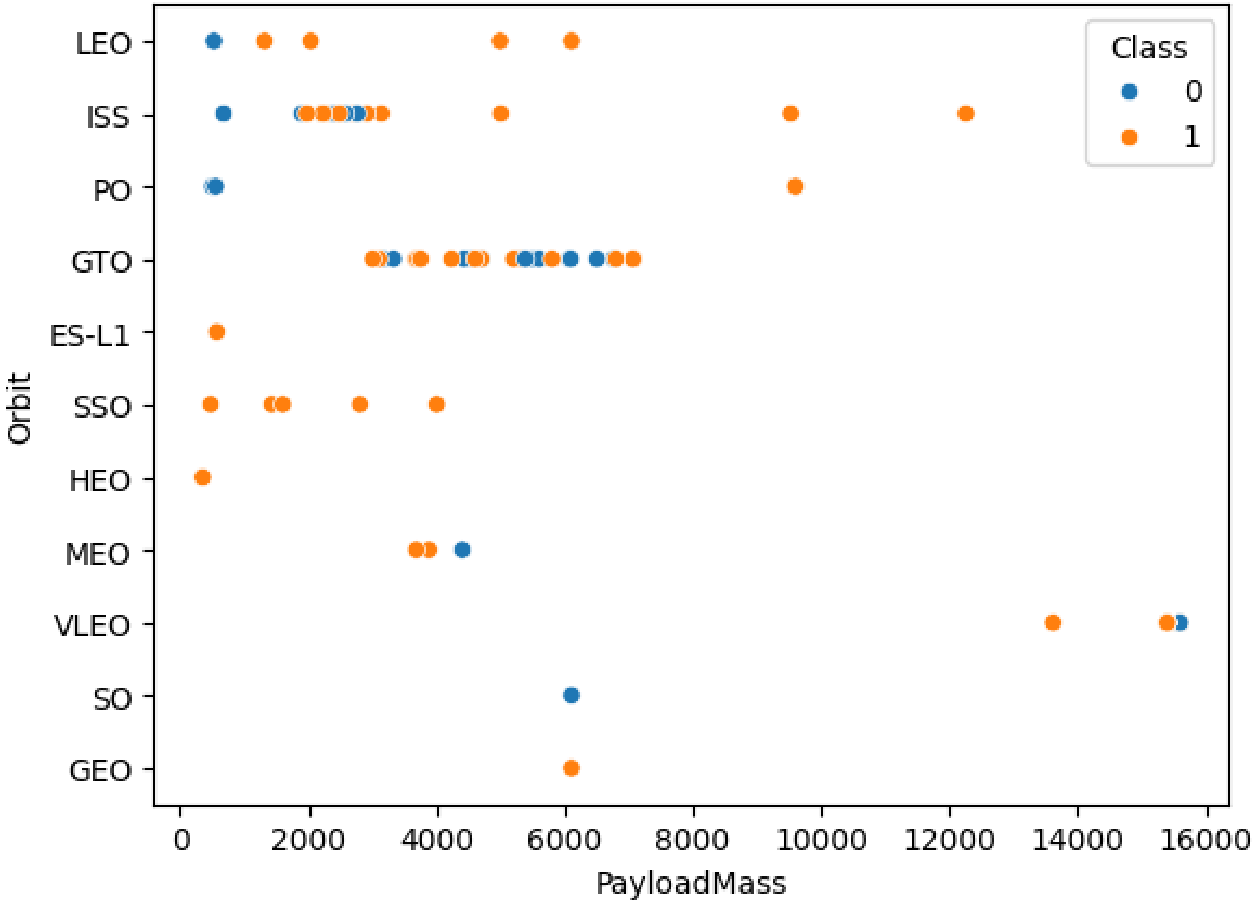
Flight Number vs. Orbit Type

Success rate by orbit type

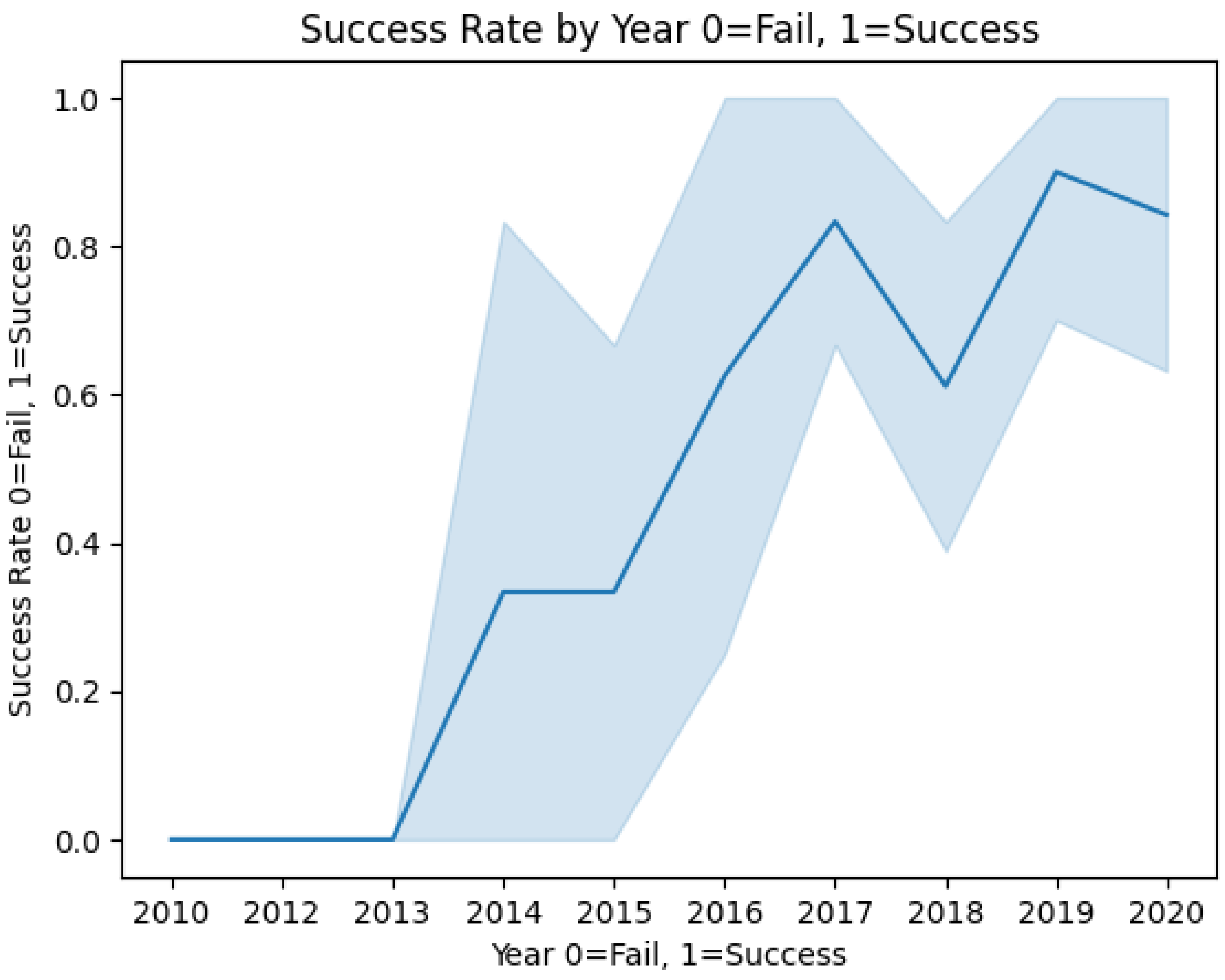


Payload vs. Orbit Type

Payload mass vs orbit type



Launch Success Yearly Trend



All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%%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

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%%sql
SELECT * FROM SPACEXTABLE WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

Python

* 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 Payload Mass

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

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTABLE WHERE CUSTOMER = 'NASA (CRS)';
```

* sqlite:///my_data1.db
Done.

TOTAL_PAYLOAD_MASS
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_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTABLE WHERE BOOSTER_VERSION = 'F9 v1.1';
```

* sqlite:///my_data1.db
Done.

AVERAGE_PAYLOAD_MASS
2928.4

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) AS FIRST_SUCCESSFUL_LANDING_DATE FROM SPACEXTABLE WHERE
Landing_Outcome = 'Success (ground pad)';
```

Python

* sqlite:///my_data1.db

Done.

FIRST_SUCCESSFUL_LANDING_DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Click to add a breakpoint

```
SELECT substr(DATE, 6, 2) AS MONTH, LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTABLE WHERE LANDING_OUTCOME =
'Failure (drone ship)' AND substr(DATE, 0, 5) = '2015';
```

* 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

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT MISSION_OUTCOME, COUNT(*) AS TOTAL_COUNT FROM SPACEXTABLE GROUP BY MISSION_OUTCOME;
```

* sqlite:///my_data1.db
Done.

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

Boosters Carried Maximum Payload

Click to add a breakpoint

```sql
SELECT BOOSTER_VERSION FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (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

2015 Launch Records

Click to add a breakpoint

```sql
SELECT substr(DATE, 6, 2) AS MONTH, LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTABLE WHERE LANDING_OUTCOME =
'Failure (drone ship)' AND substr(DATE, 0, 5) = '2015';
```

* 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

Click to add a breakpoint

```
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.

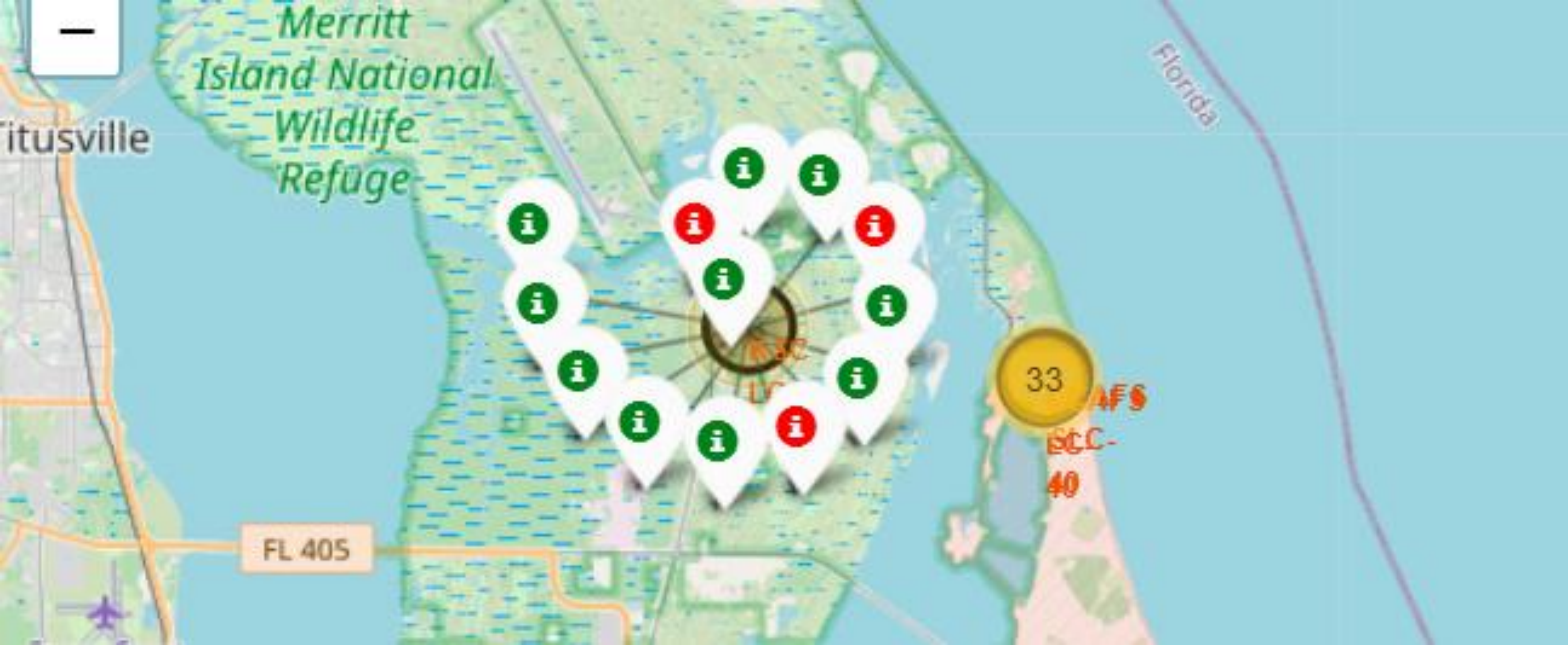
Landing_Outcome	COUNT
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

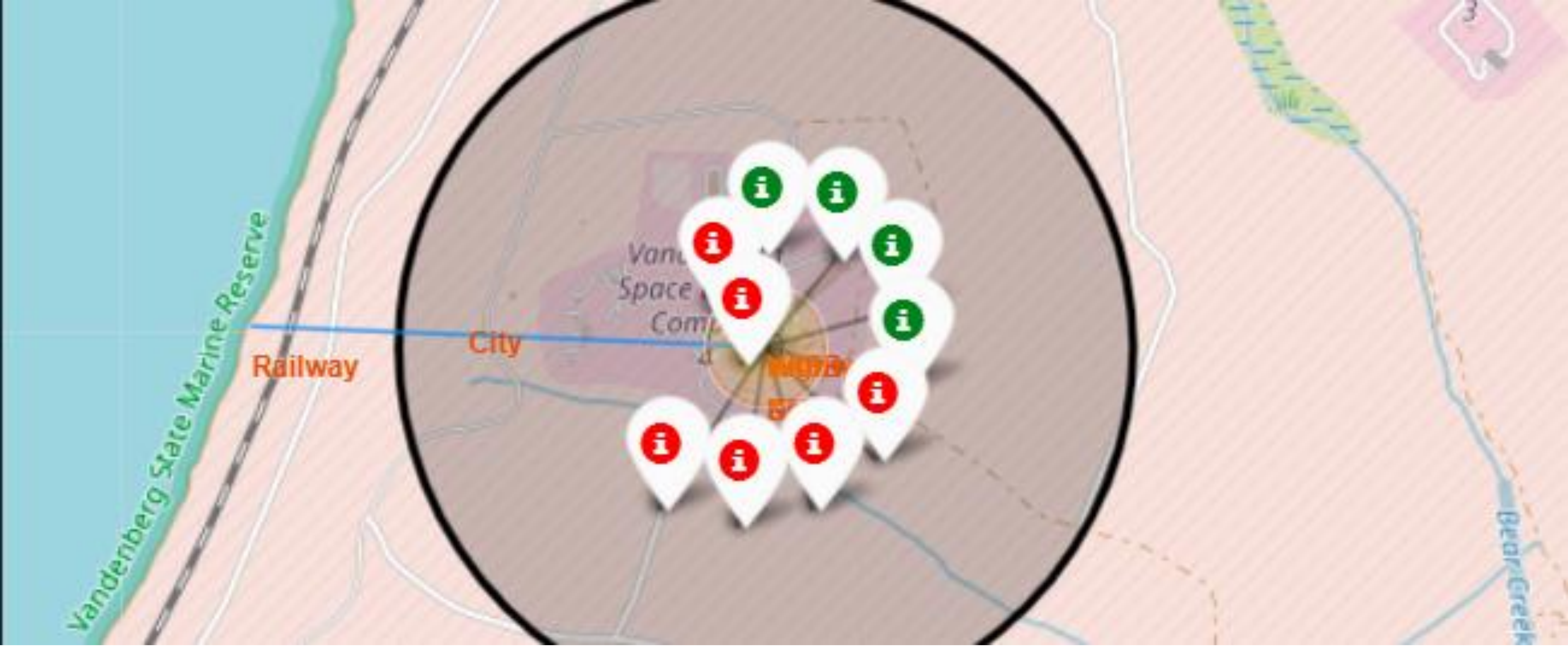
Wide View Showing marker clusters

link to file



Outcome Class as Coloured Marker

[link to file]



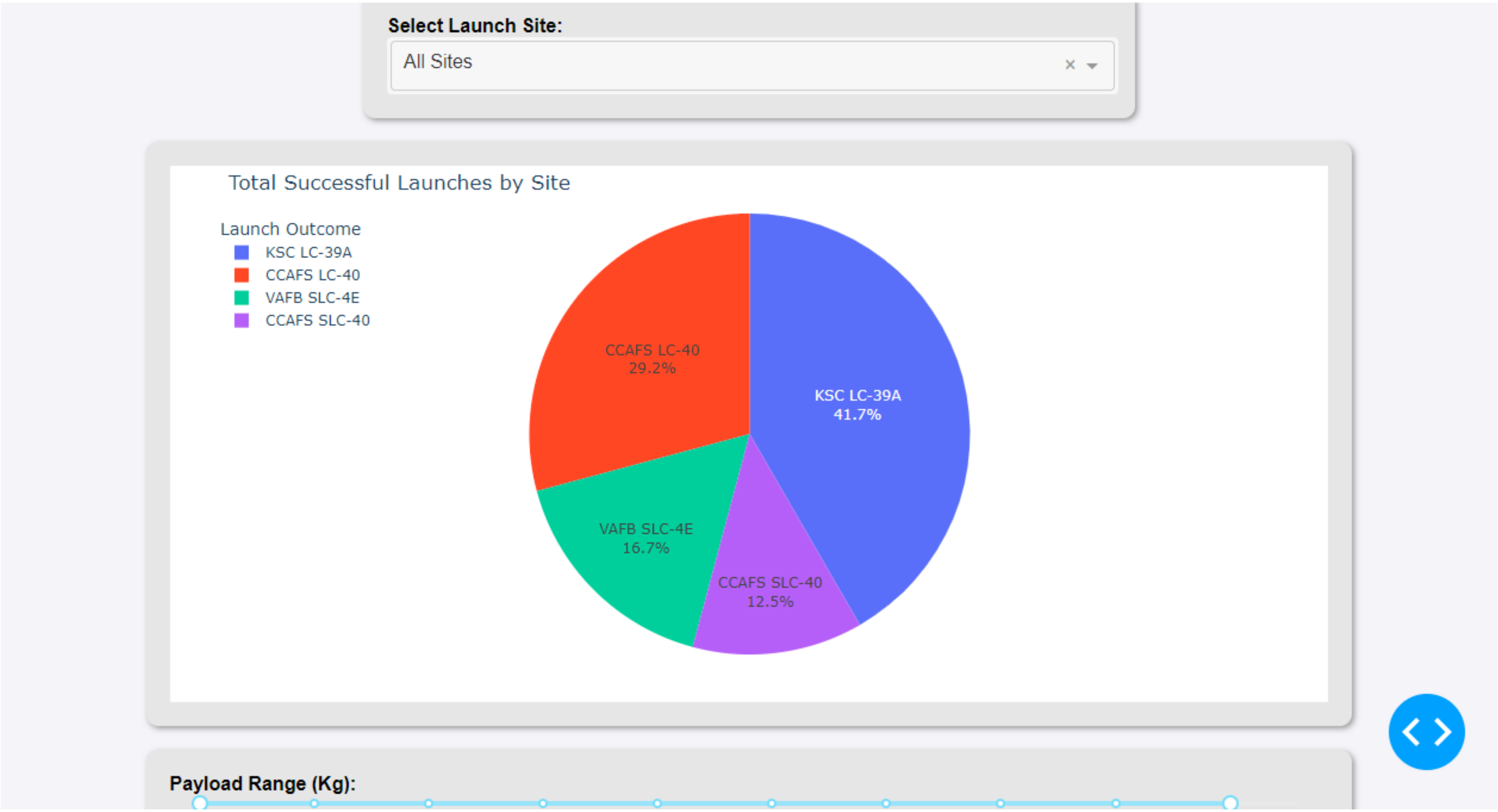
Example of folium annotations

[link to file]

Section 4

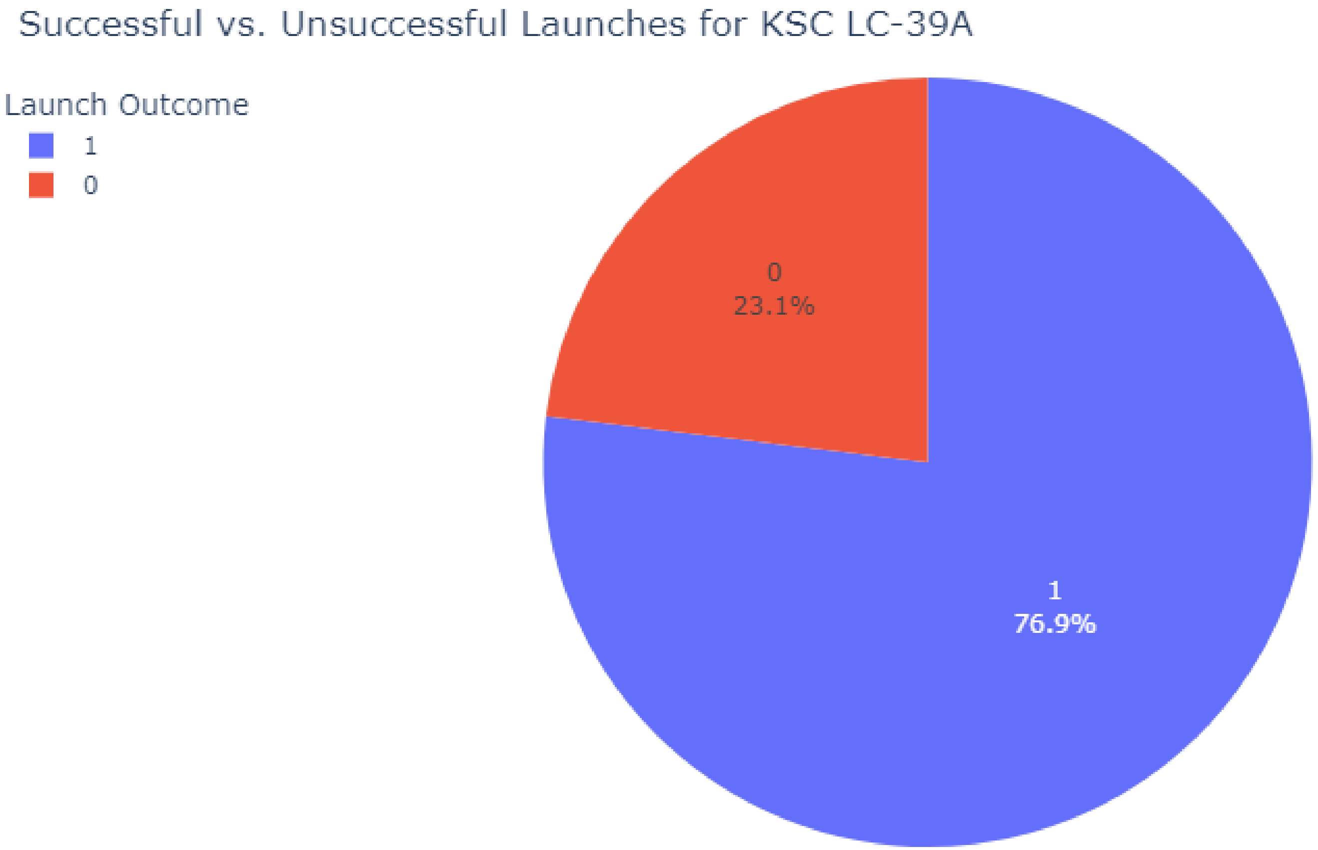
Build a Dashboard with Plotly Dash

Launch Sites



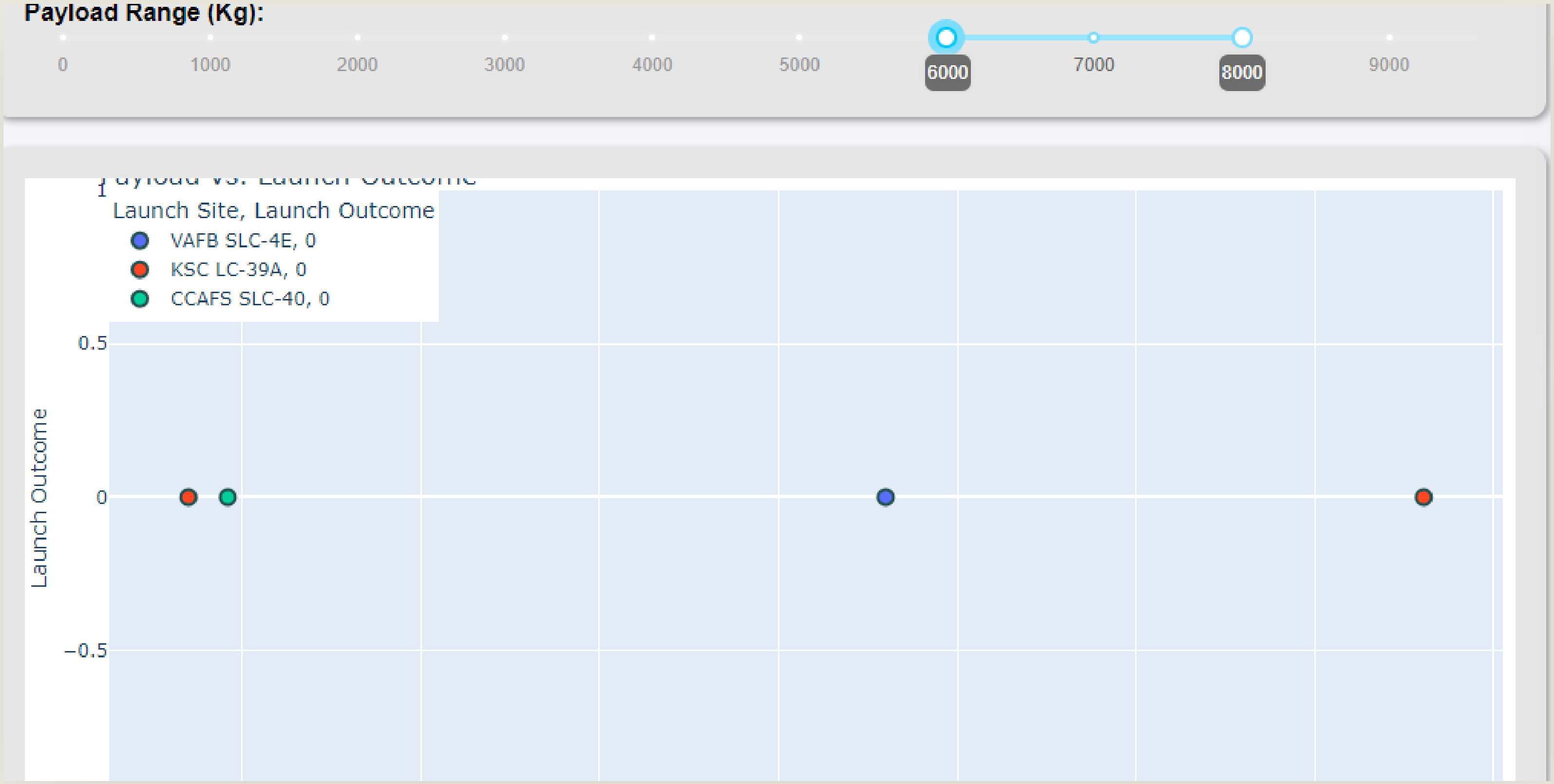
dash script

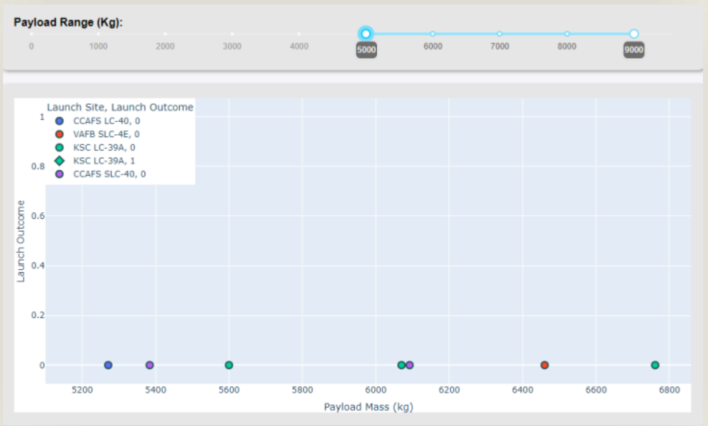
Highest Success Ratio



[dash script]

Payload vs Launch Outcome



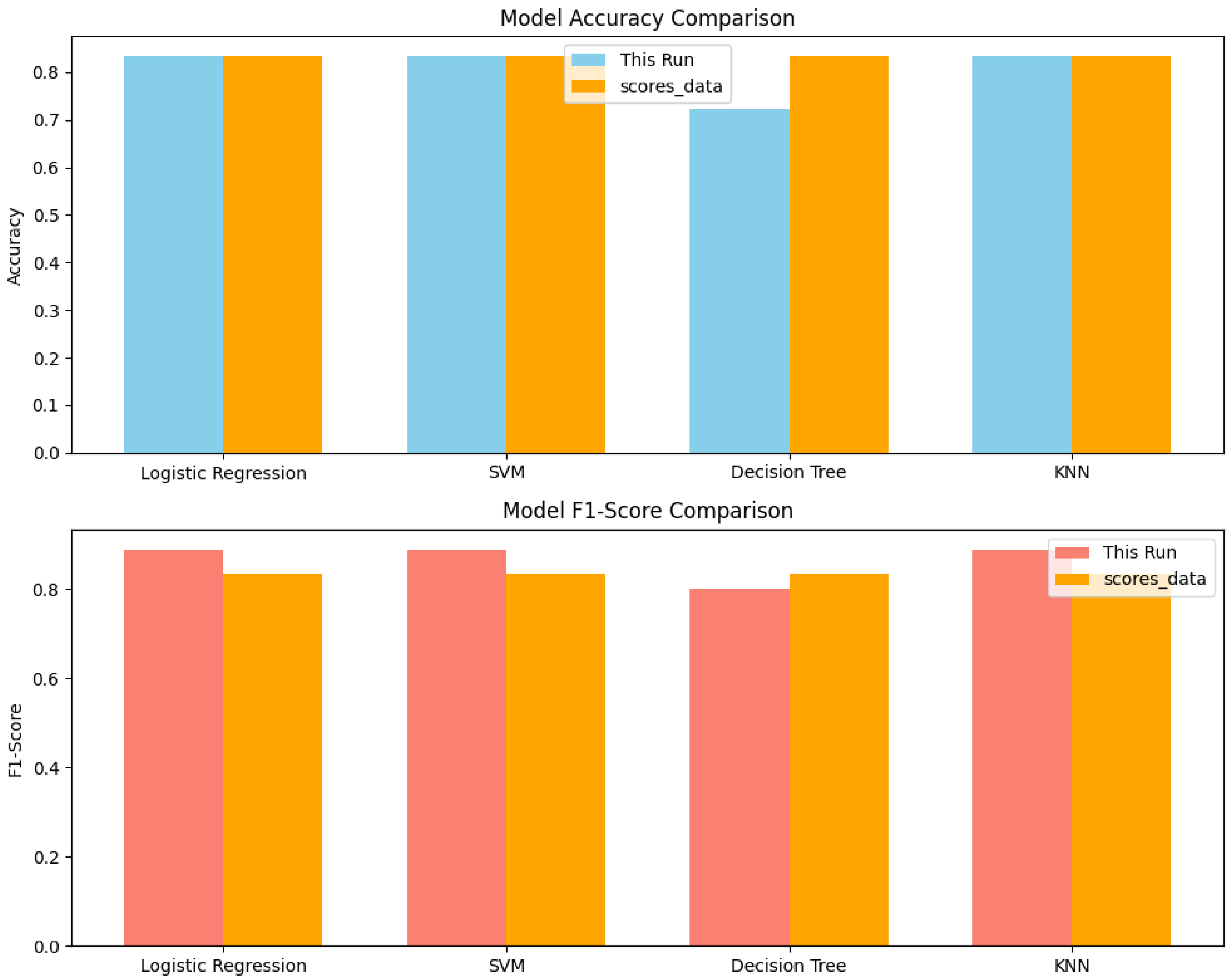


[dash script]

Section 5

Predictive Analysis (Classification)

Classification Accuracy

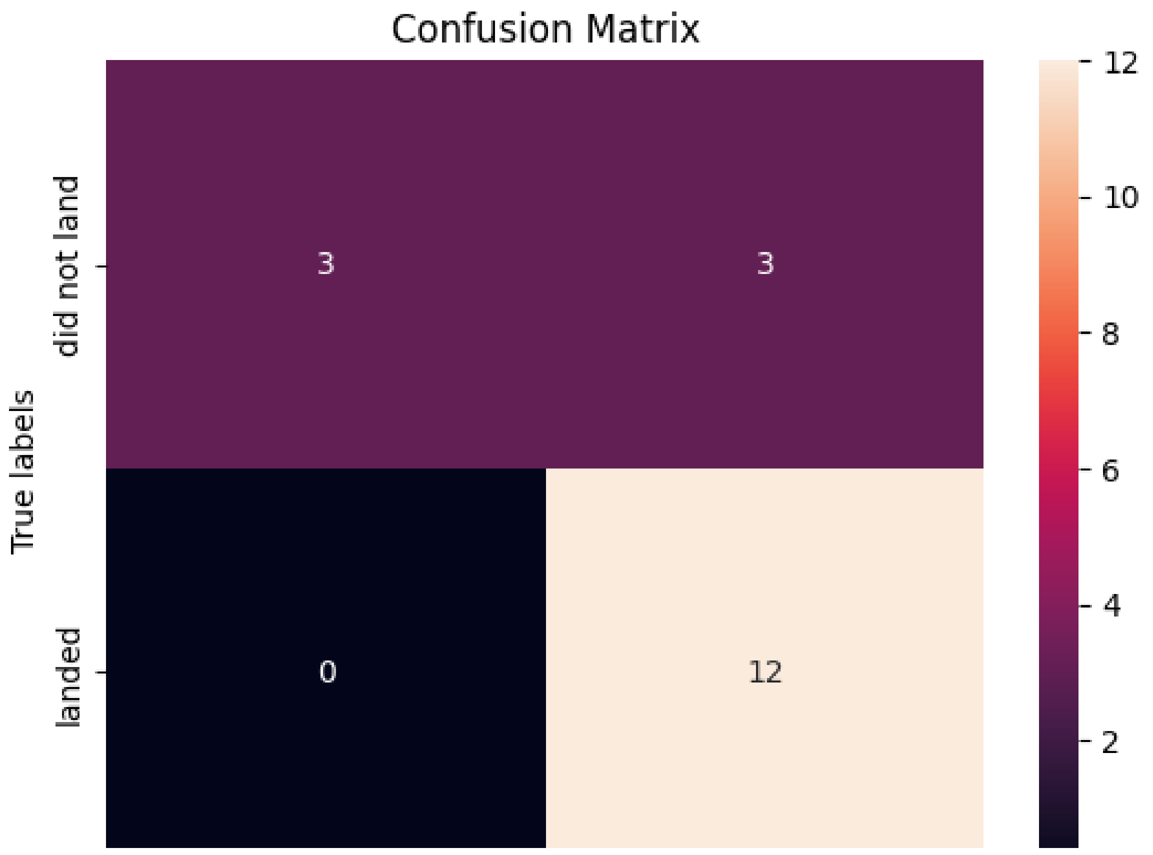


[notebook link]

Confusion Matrix

Logistic Regression, SVM and KNN had the same scores

Model	Score	Accuracy Score	F1 Score
Logistic Regression	0.833333	0.833333	0.888889
SVM	0.833333	0.833333	0.888889
Decision Tree	0.833333	0.722222	0.800000
KNN	0.833333	0.833333	0.888889



Conclusions

The models achieved high accuracy in predicting Falcon 9 landings, showcasing the effectiveness of data-driven approaches in aerospace.

Accurate predictions can significantly reduce launch costs, enhancing SpaceX's competitive edge and enabling informed decision-making.

Hyperparameter tuning optimized model performance, underscoring the value of rigorous model development processes.

Future enhancements could involve advanced techniques and feature enrichment to further improve predictions and practical applications.

notebook link

Appendix

Thank you!