IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alejandro Copado
01/30/25

Outline

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

Executive Summary

- Employed Methodologies:
  - Data Collection – Completed via SpaceX API & Beautiful Soup Web Scrapping(Wikipedia)
  - Data Wrangling – Via Pandas.DataFrame()
  - EDA – SQL w/ Python utilizing SQLite3
  - EDA – Visualization w/ Matplotlib & Seaborn
  - Interactive Visualization – Folium, Dashboard w/ Plotly Dash
  - Predictive Analysis - Classification
- Summary of all results:
  - Highest degree of success is based on payload, location & orbit.

Introduction

- What is our Purpose?
  - Given the success of SpaceX's innovative reusable launch vehicles, at a significantly reduced cost compared to it competitors, a market exists for a lower cost high success reuse of first stage boosters.
- What do we need to know?
  - We need to identify the variables that alter the rate of success for first stage boosters safely landing in their designated area, i.e.; payload mass, launch site & orbit type.
  - What are the best Classification models for predicting success.

Section 1

Methodology

Methodology

Executive Summary

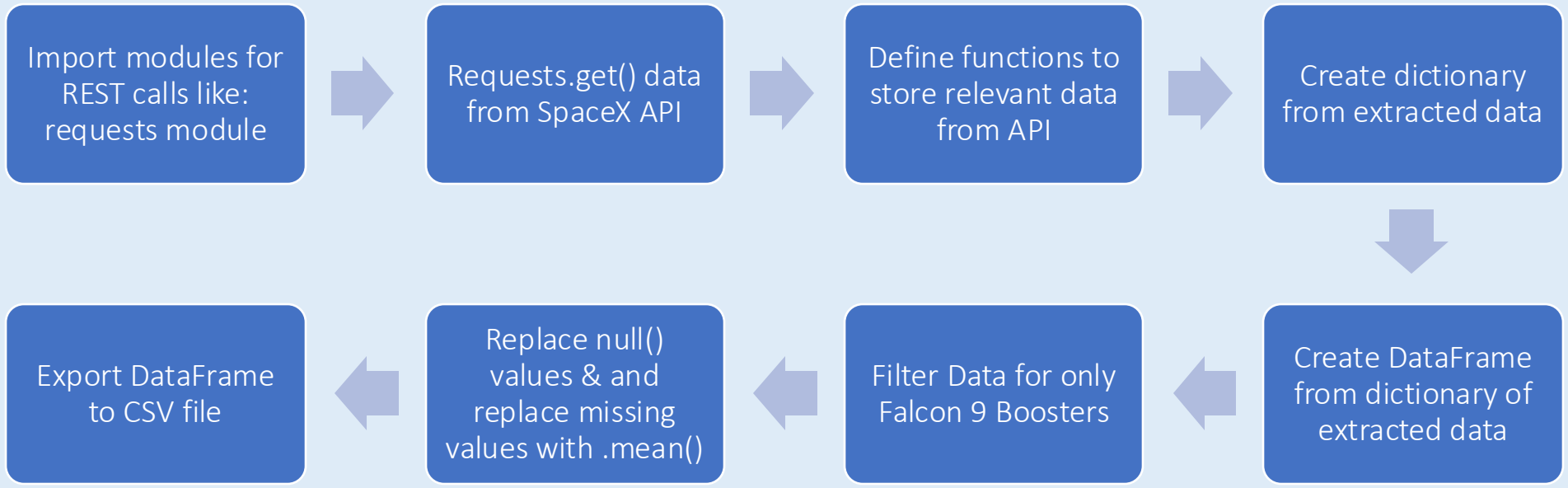
- Data collection methodology:
  - Completed via SpaceX API & Beautiful Soup Web Scrapping(Wikipedia)
- Perform data wrangling
  - Data was stored in tabular form via Pandas.DataFrame().
  - Data was cleaned of null & missing values as well as ensuring categorical data was transformed into binary values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Utilized Python Scikit-learn to determine classification models: Logistic Regression, Decision Tree, KNN, SVM

Data Collection

- Collection Process
  - Utilized SpaceX API via a requests.get() from the requests module of Python.
  - Data was stored in tabular form via Pandas.DataFrame().

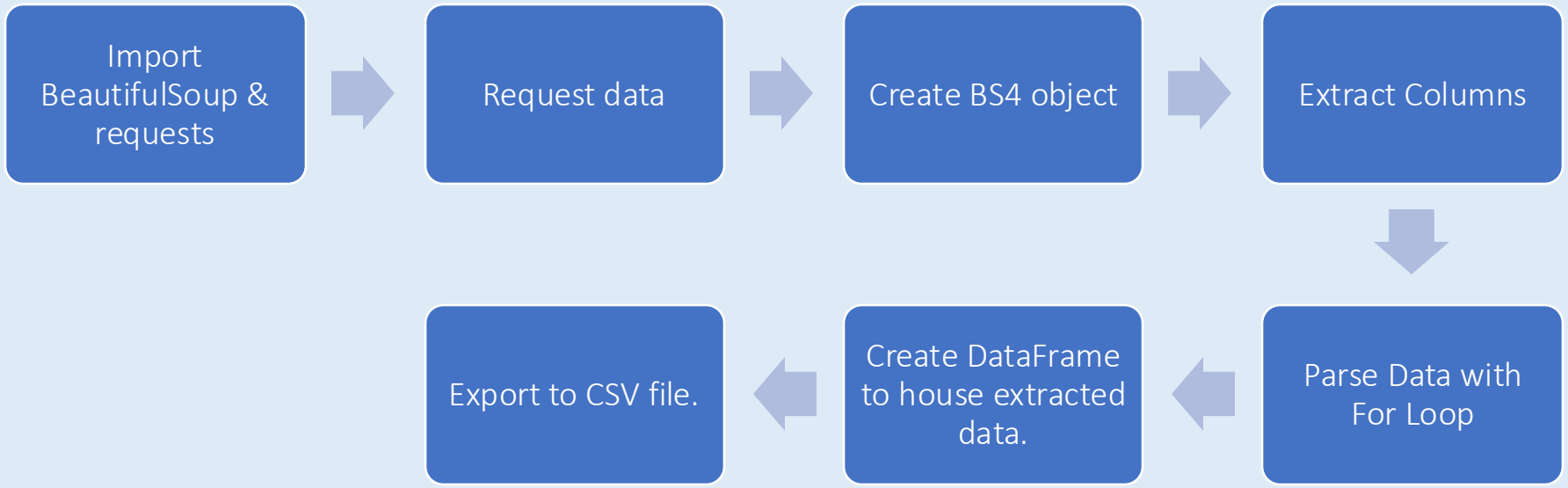
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API



GitHub URL: Data Collection API

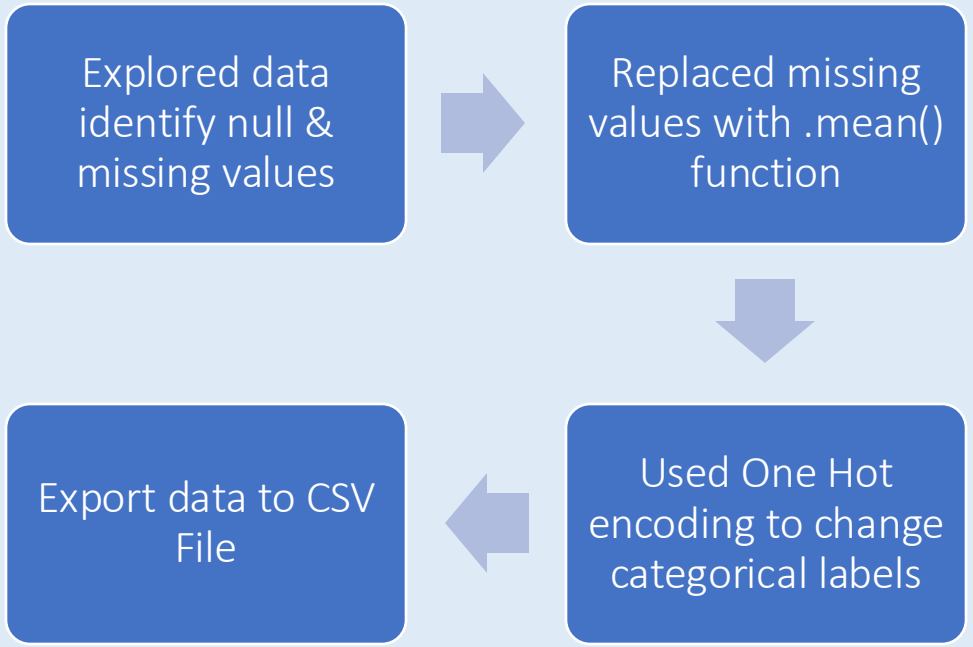
Data Collection - Scraping



GitHub URL: Webscrapping

Data Wrangling

- Data was extracted then cleaned of null values, missing values were replaced using Python .mean() function. Categorical values were transformed to binary values using One Hot Encoding were "1" indicated the booster successfully landed and "0" meant the booster was unsuccessful.



GitHub URL: Data Wrangling

EDA with Data Visualization

Visualizations:

1. Flight # vs. Launch Site
2. Payload Vs Launch Site
3. Success Rate of Orbit
4. Flight # & Orbit Type
5. Payload & Orbit Type
6. Yearly success rate

Data Visualization was plotted using scatter plots to help identify relationships between multiple variables. Which can help with identifying variables to test with machine learning.

GitHUB URL - Data Visualizations

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- • Displaying 5 records where launch sites begin with the string ‘CCA'
- • Displaying the total payload mass carried by boosters launched by NASA (CRS)
- • Displaying average payload mass carried by booster version F9 v1.1
- • Listing the date when the first successful landing outcome in ground pad was achieved
- • Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but
- less than 6000
- • Listing the total number of successful and failure mission outcomes
- • Listing the names of the booster versions which have carried the maximum payload mass
- • Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in
- year 2015
- • Ranking 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 URL: EDA w/ SQL

Build an Interactive Map with Folium

- All launch sites were created as objects with markers, circles, Popup Labels & text labels.
- Additionally colored markers were added shoing success and failure of launchs.
- Added colored lines to Launch site to show proximity to infastructure.

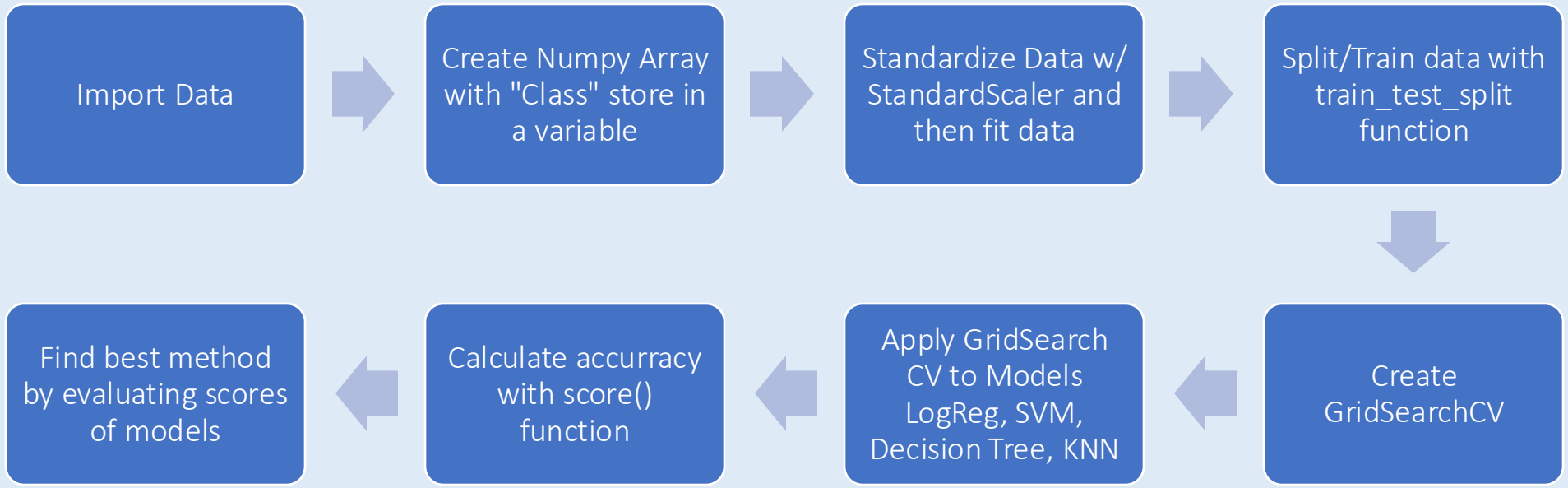
GitHub URL: Interactive Map

Build a Dashboard with Plotly Dash

- Dropdown list created to select Launch Site.
- Pie Chart showing Rate of Success of all Launch Sites.
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

GitHub URL: Dashboard

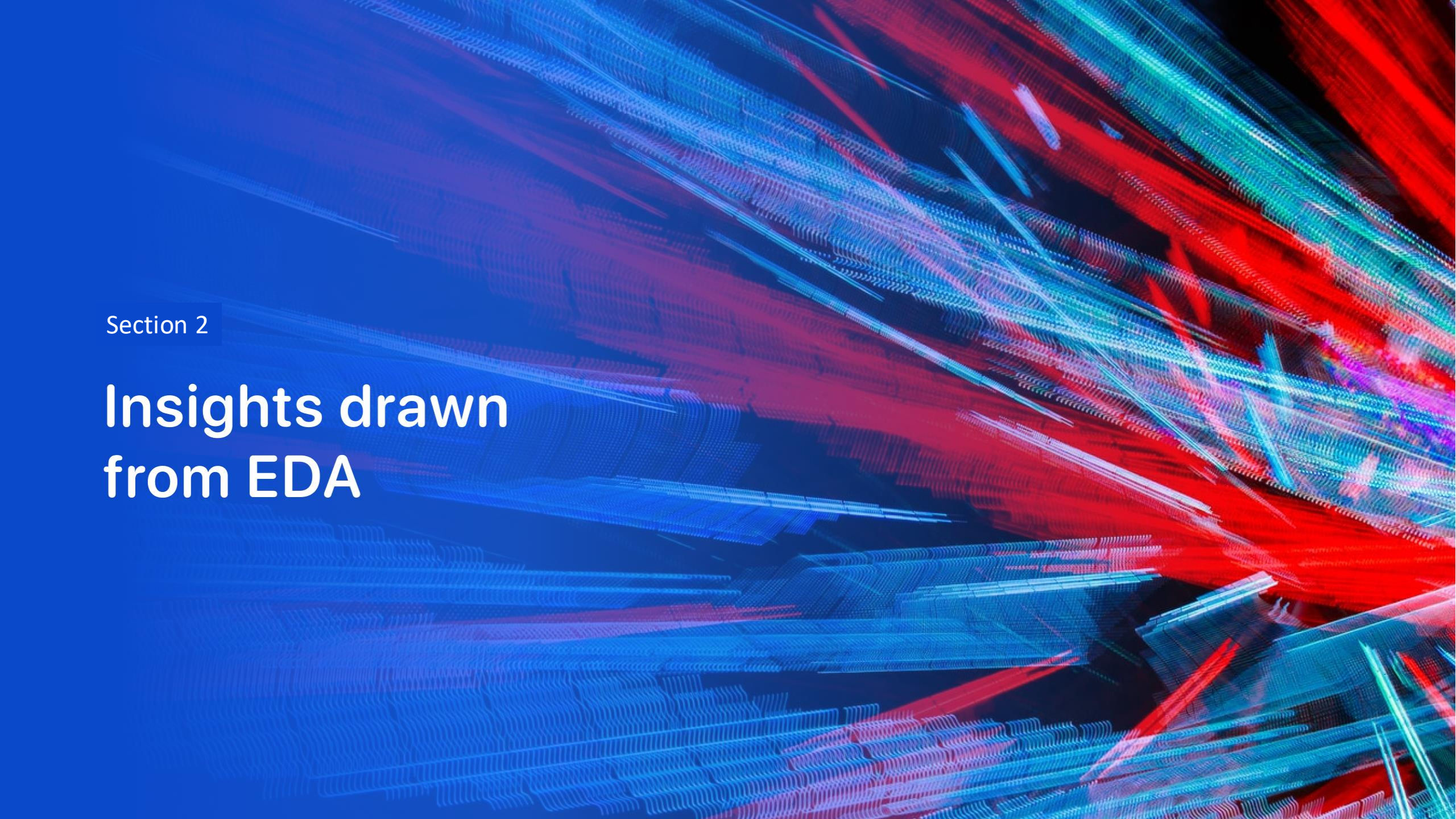
Predictive Analysis (Classification)



GitHub URL: Classification

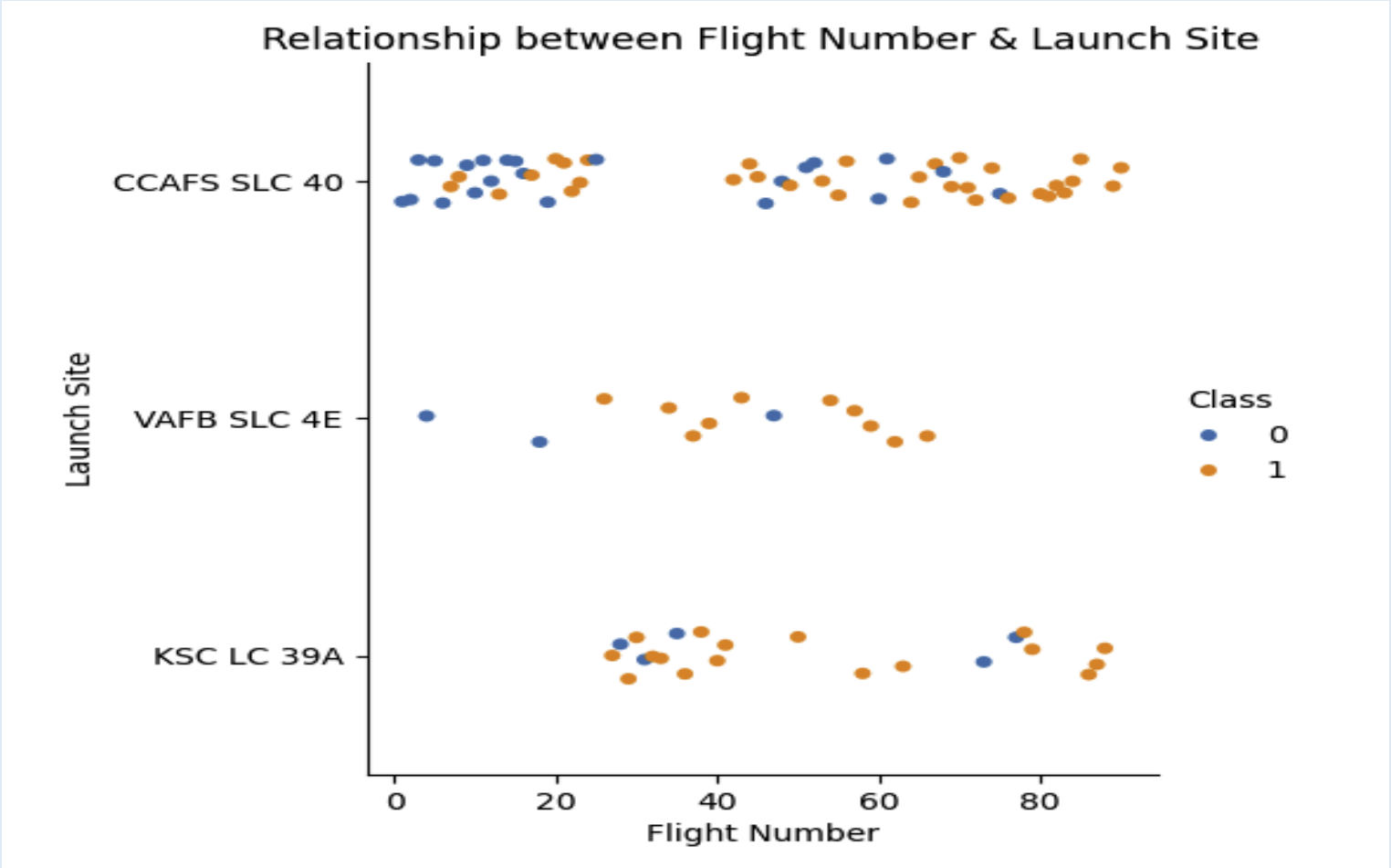
Results

- Highest degree of success is based on payload, location & orbit.
- Decision Tree Model appears to be the most appropriate model to test with.


Section 2
Insights drawn from EDA

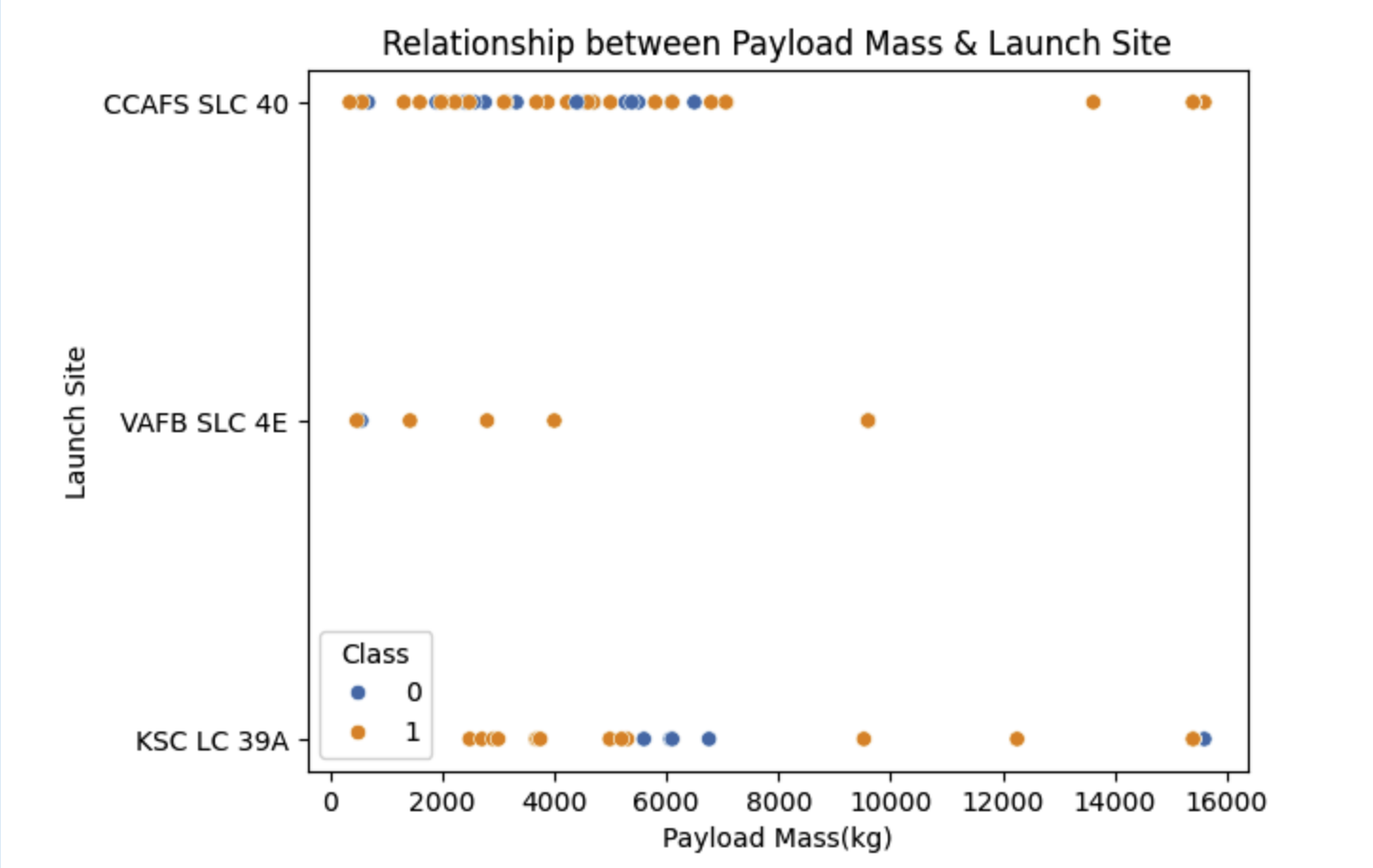
Flight Number vs. Launch Site

- VAFB SLC 4E have high success rate
- Earlier flights often failed
- KSC LC 39A also has high success rate



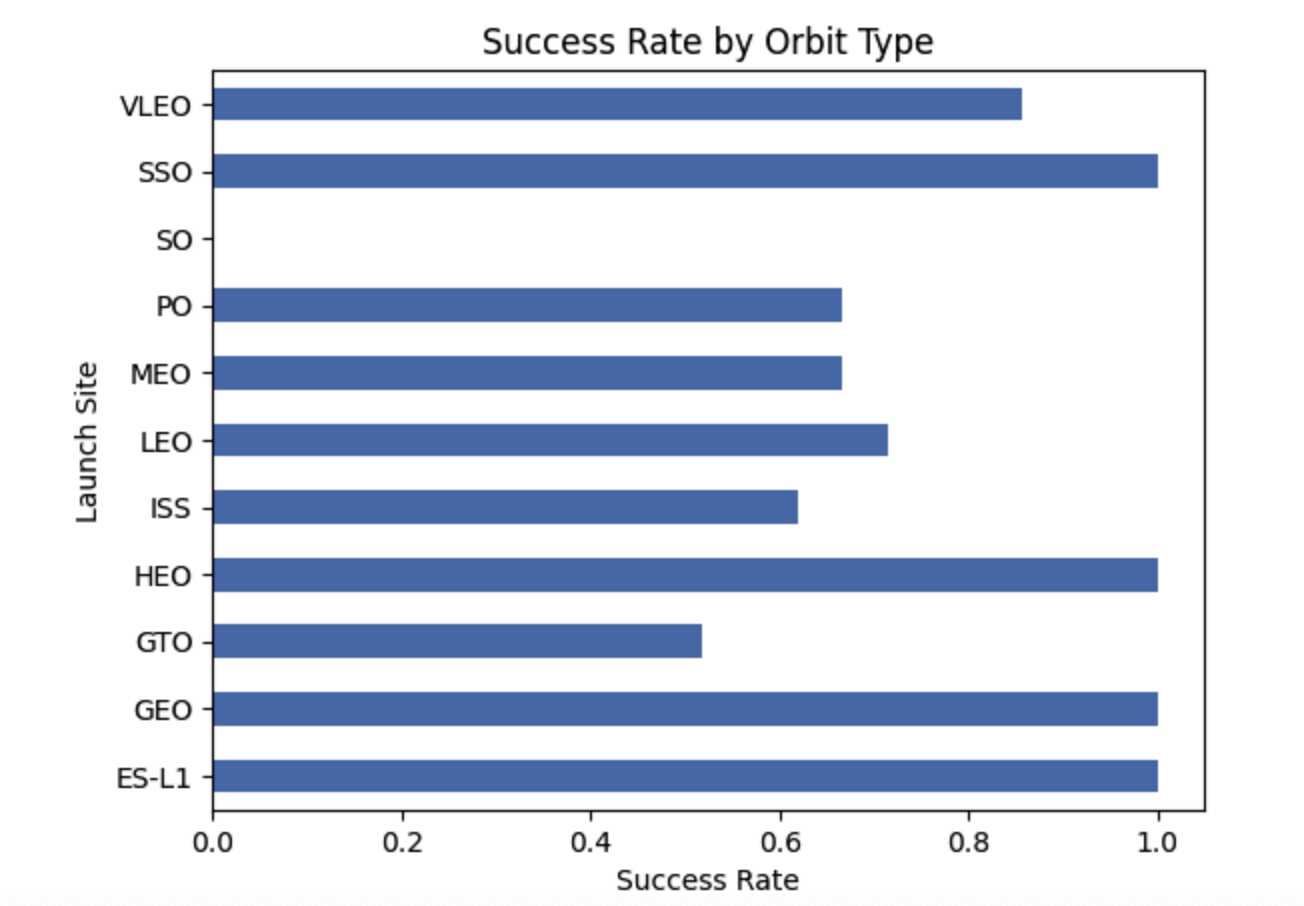
Payload vs. Launch Site

- CCAFS SLC 40 had the highest success with heavy payloads.
- Payloads of less than 10,000kg were the most successful



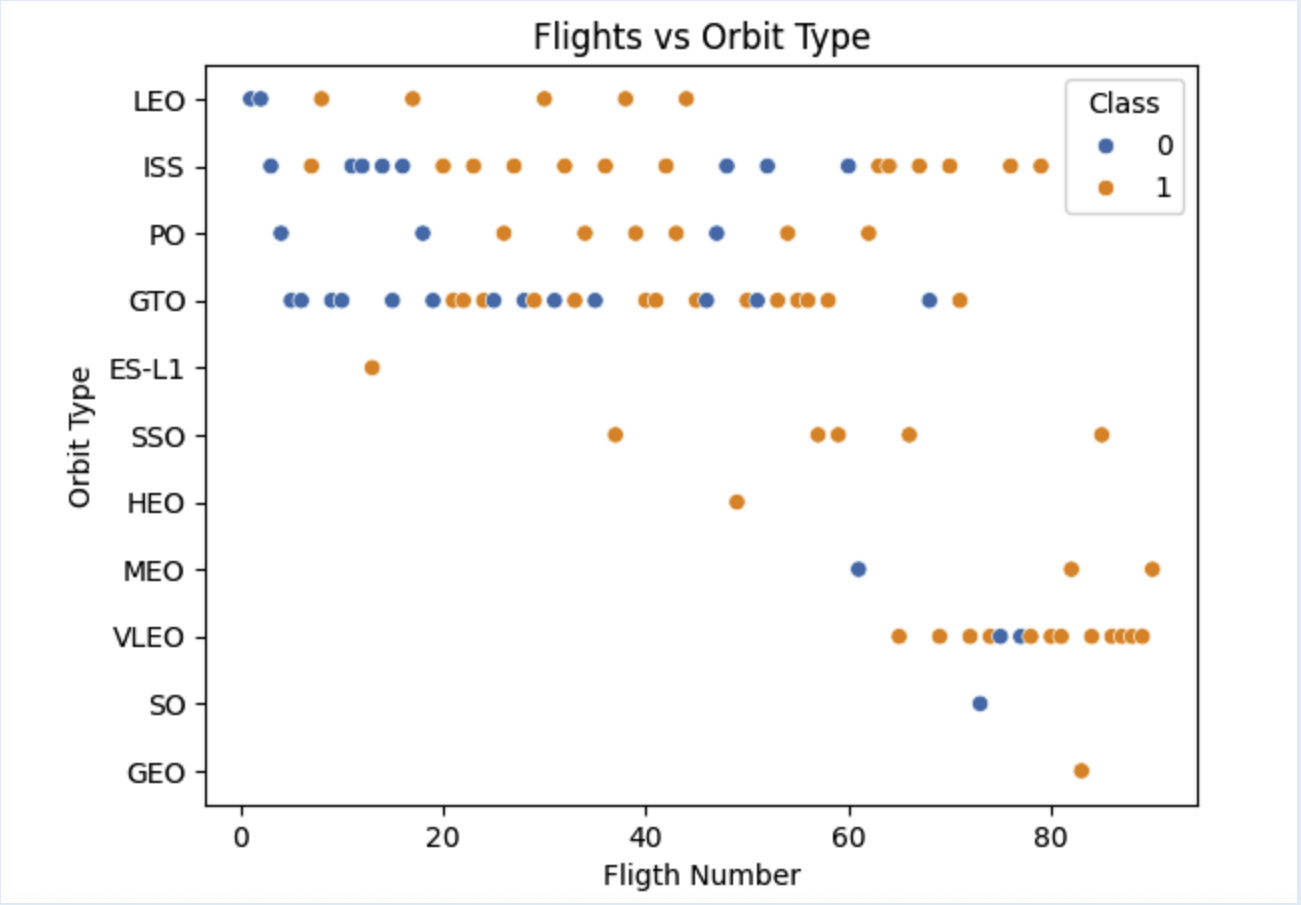
Success Rate vs. Orbit Type

- SSO, ES-L1, GEO, HEO had 100% success rate.
- GTO had the lowest success rate.



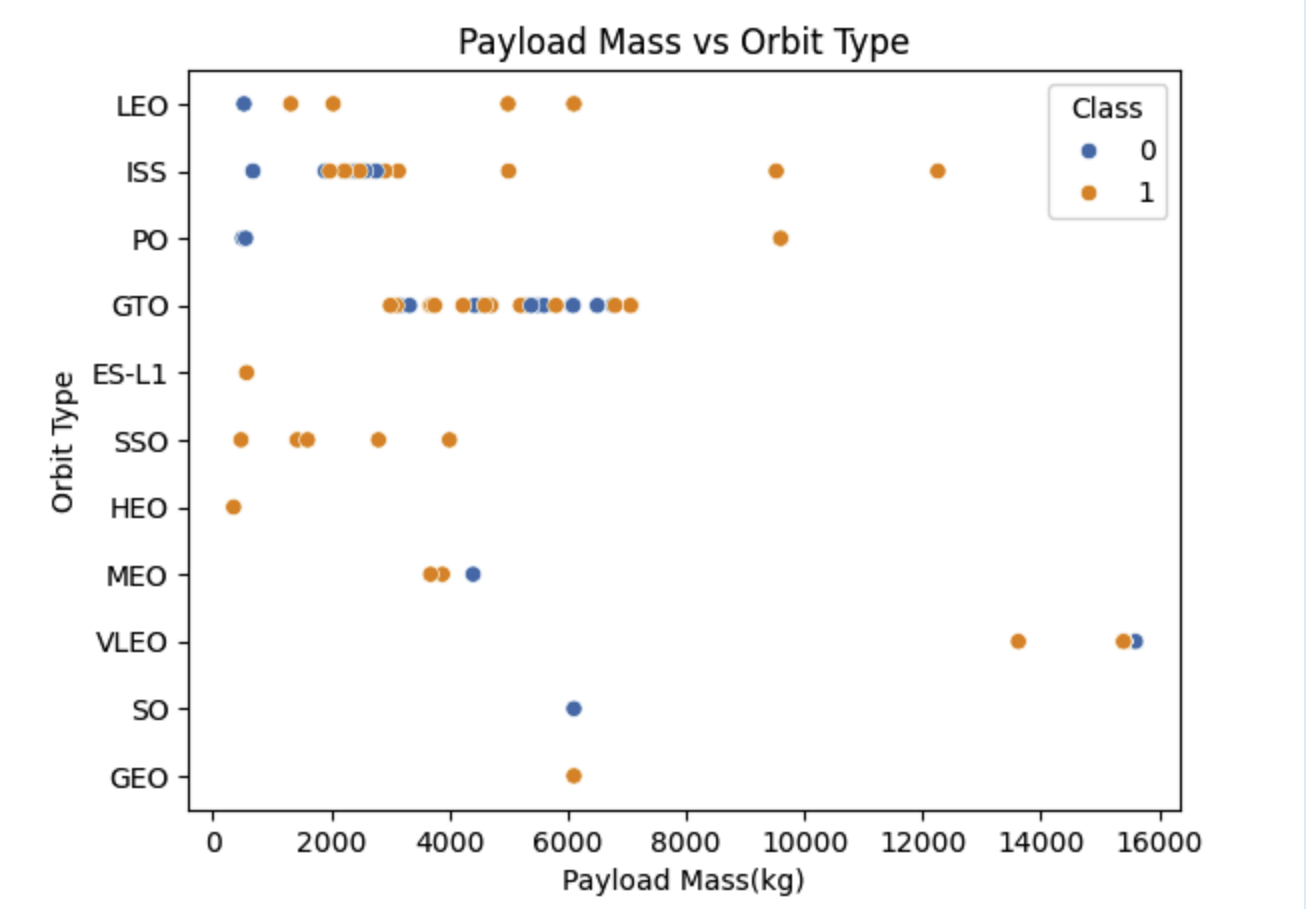
Flight Number vs. Orbit Type

- VLEO & SSO have the highest success rate the more flights completed.
- No relationship exists with GTO & number of flights.



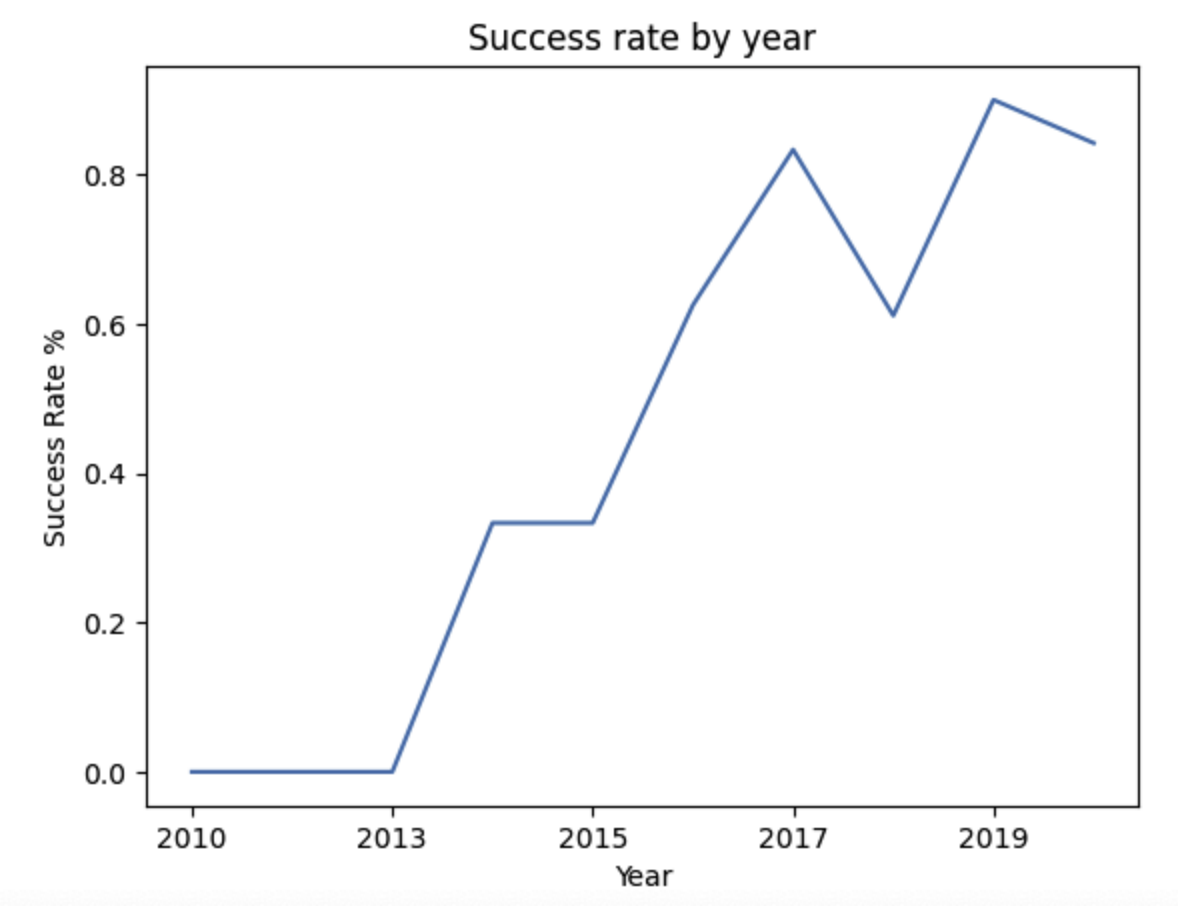
Payload vs. Orbit Type

- Lower weight payloads are the most successful.
- SSO payloads less than or equal to 4,000kg have 100% success rate



Launch Success Yearly Trend

- Success has gone up significantly every year since 2013



All Launch Site Names

- names of the unique launch sites

```
%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'

- Find 5 records where launch sites 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;
```

```
 * 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

- Calculate the total payload carried by boosters from NASA

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

```
%%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
    WHERE "Customer"
    LIKE "NASA (CRS)";
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

average_payload
2534

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

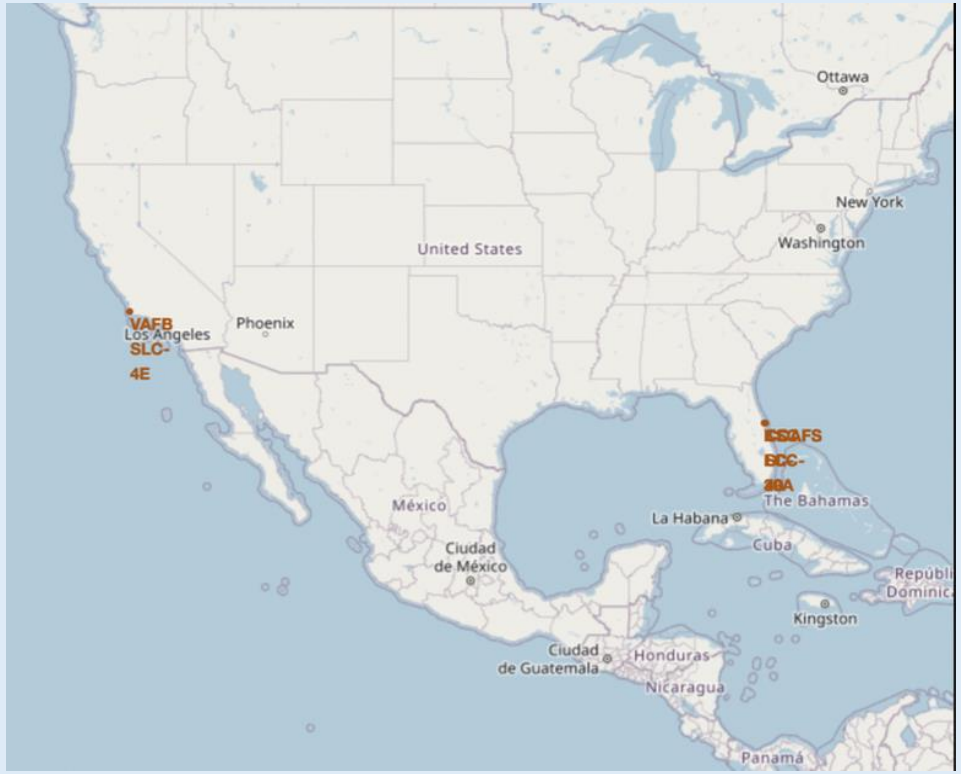
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	count_outcomes
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

Section 3

Launch Sites Proximities Analysis

Launch Sites

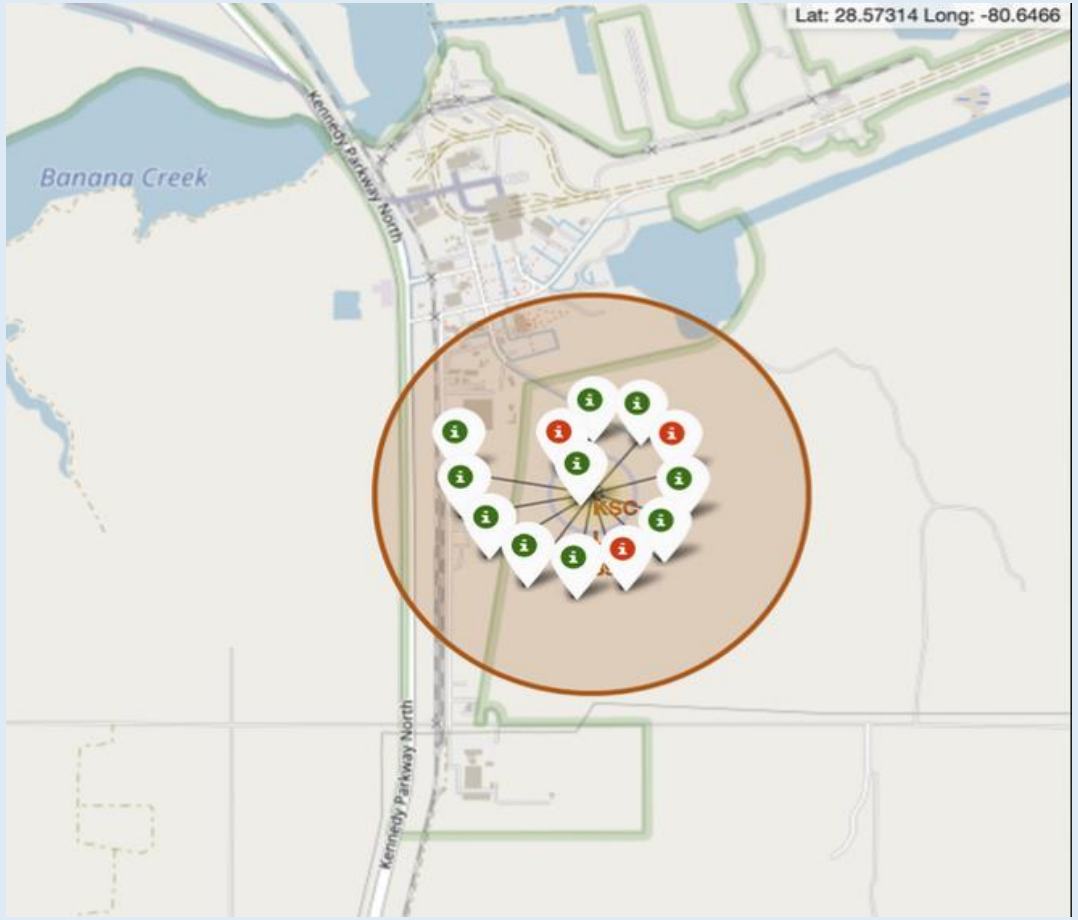


Sites are primarily located near coastal areas and close to the Equator Line.

Launch Site with Outcome Markers

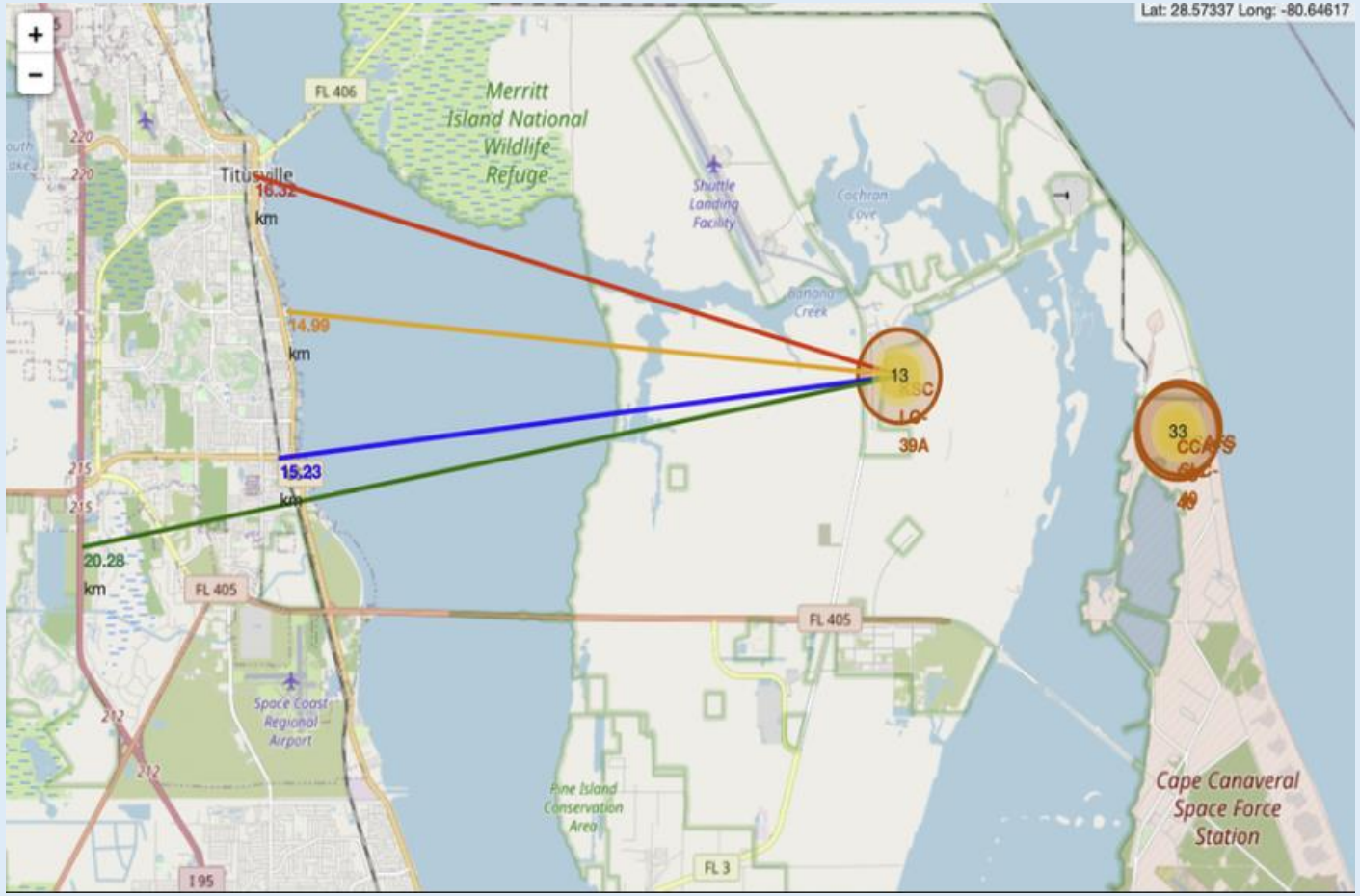
Markers indicating the Launch Sites:

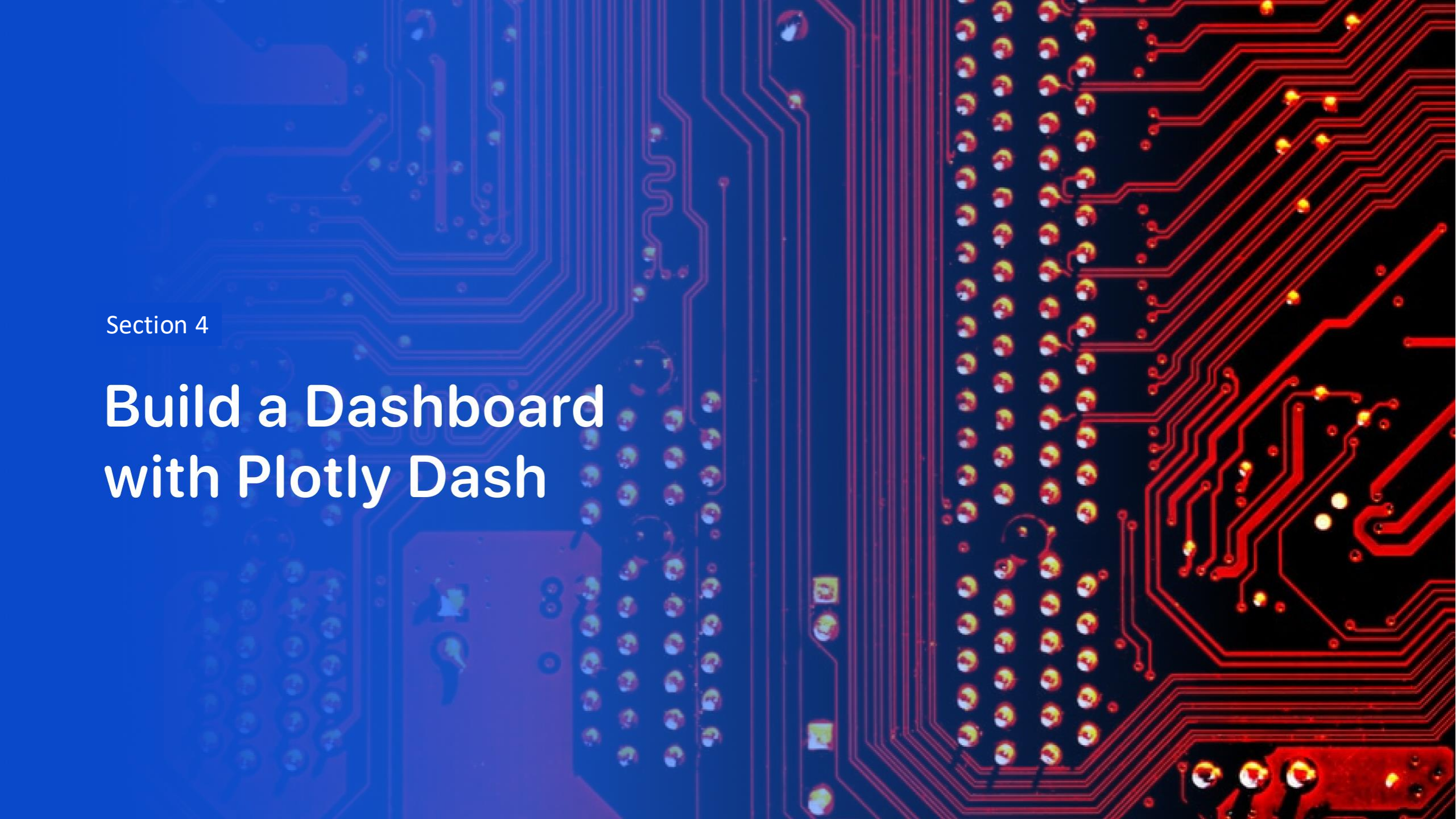
- Red = Failed Launch
- Green = Successful Launch



Distance Markers

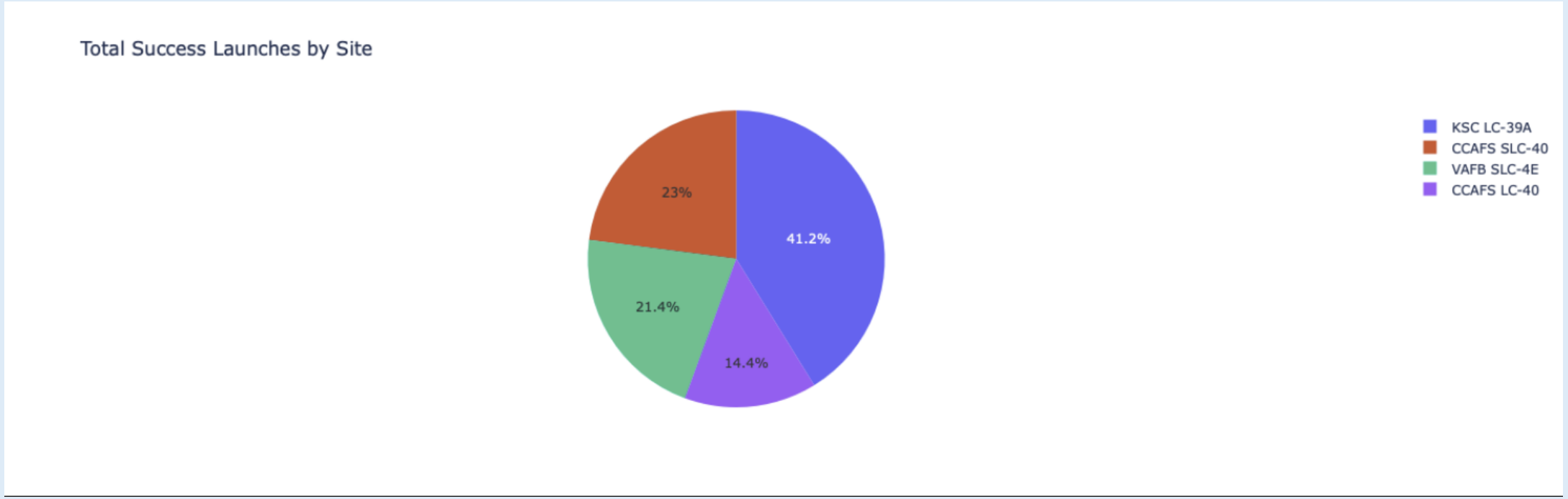
- Distance measuring site KSC LC-39A to railway, coastline & highway.




Section 4
Build a Dashboard with Plotly Dash

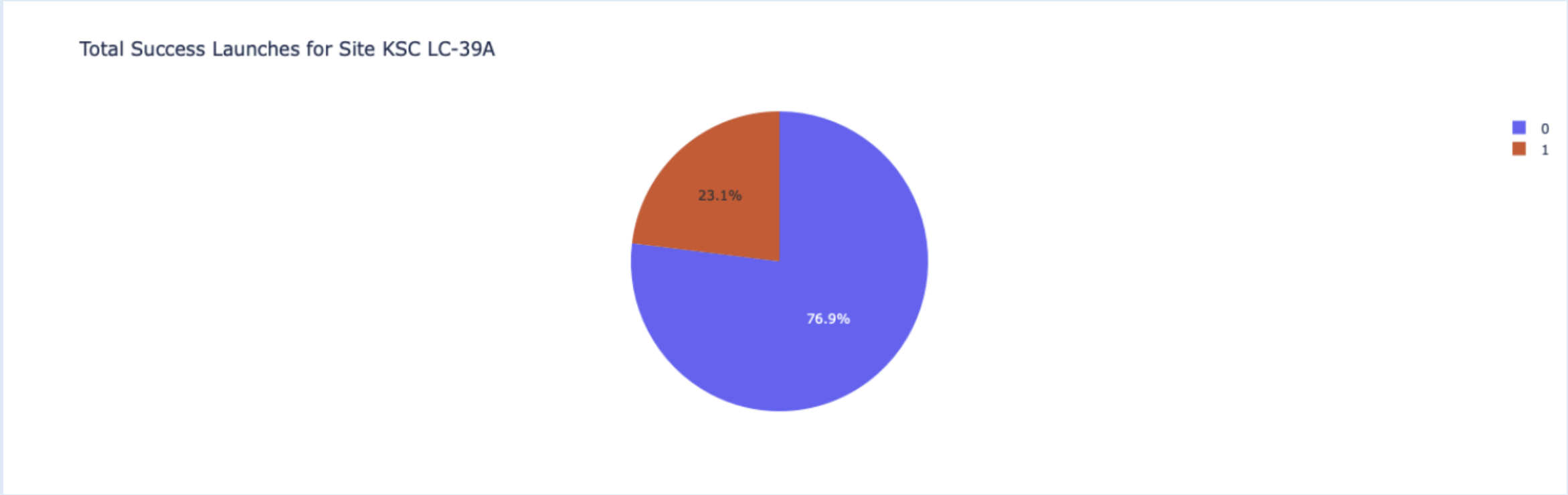
Launch Site success counts

- KSC LC 39A has the highest success rate



Launch site with highest success ratio

- Success ration for KSC LC 39A 76.9%



Payload vs Outcome

- Payloads between 2000 & 5500 have high success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Scores of Classification Accuracy on all models on the test set
- Scores of Classification Accuracy on all models on the entire set

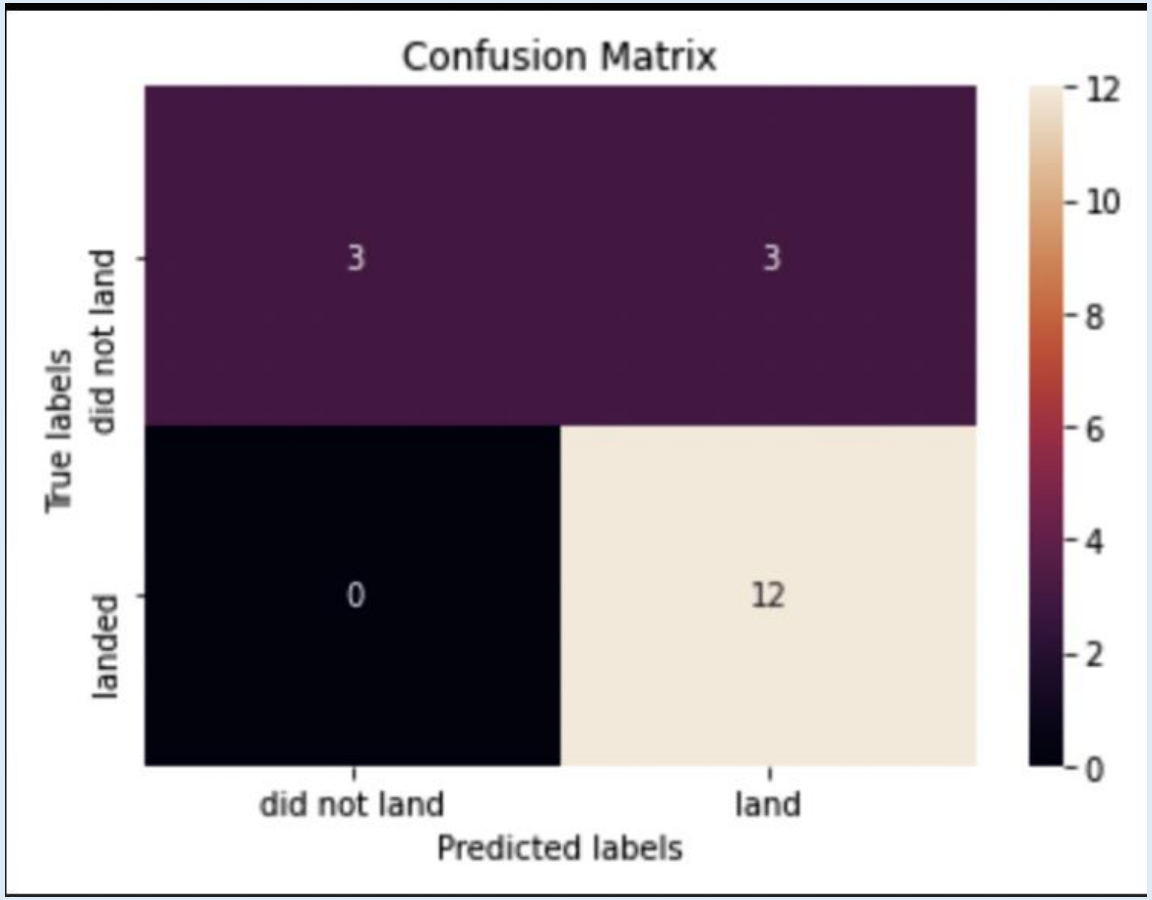
Test Set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Full Data Set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

Confusion Matrix



confusion matrix of highest scoring model.

Conclusions

- Over time the rate of success has increased.
- Ensuring site are near the equator and located in close proximity to a coast ensures higher success.
- Low weight payloads have the best results.
- KSC LC-39A had the best success.

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!