IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alireza Choobindari
August 2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies:
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis (Classification)

- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- 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 because SpaceX can reuse the first stage.

- In this project task we try to predict the first stage of the SpaceX Falcon9 rocket will land successfully or not.

Section 1

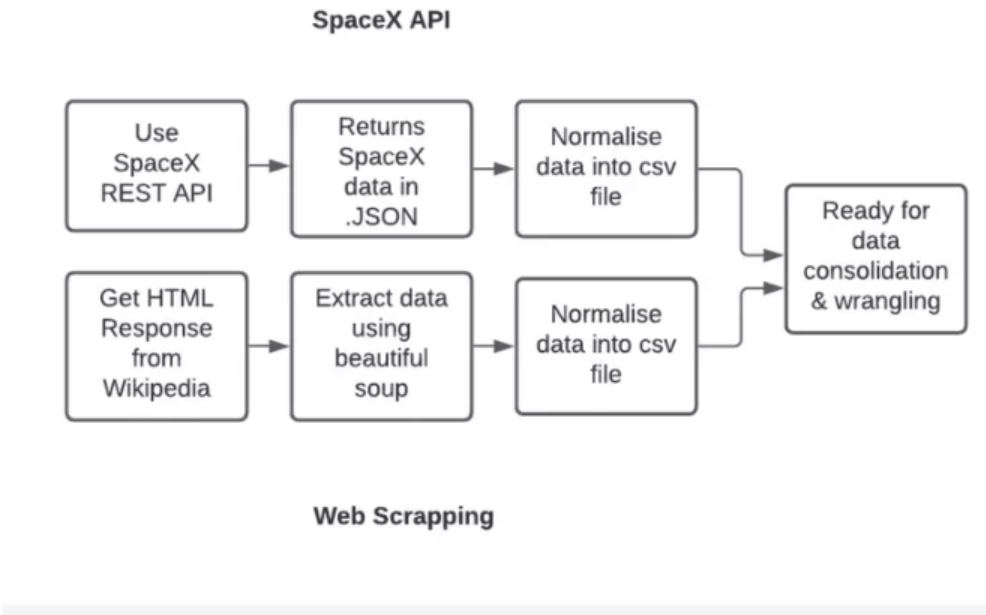
Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - One Hot Encoding data field for Machine Learning and data cleaning of null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM, and DT models have been built and evaluated for the best classifier

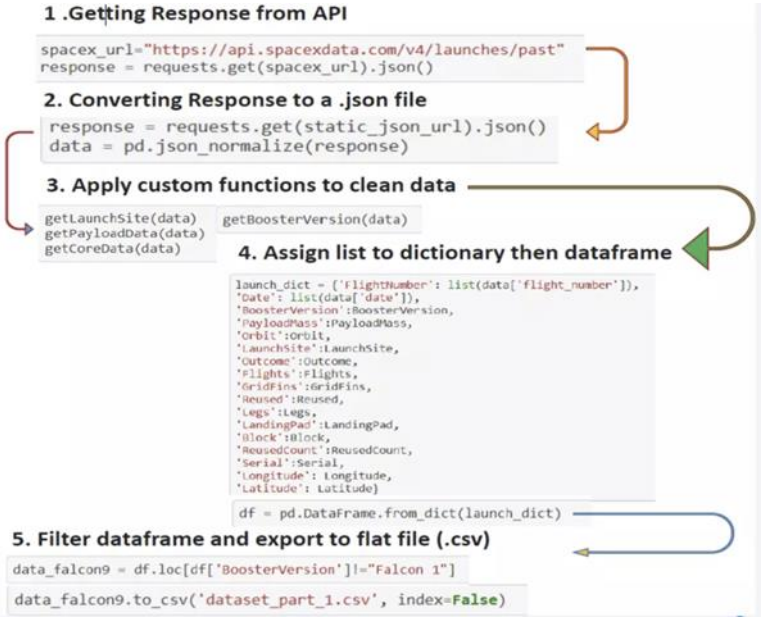
Data Collection

- The following datasets were collected:
  - SpaceX launch data that is gathered from the SpaceX REST API.
  - This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - The SpaceX REST API endpoints, or URL, starts with api.spacex.com/v4/.
  - Another popular source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

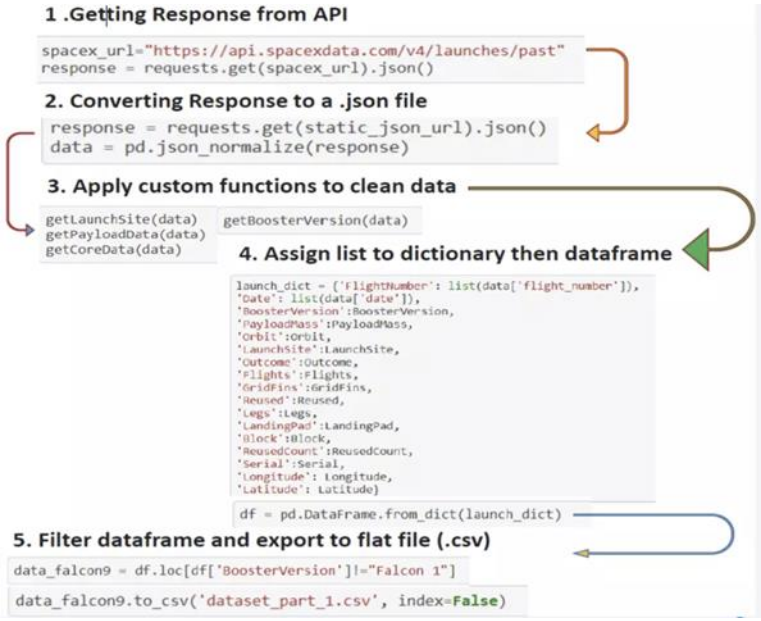
- Data collection with SpaceX REST calls



- https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/Data-Collection-API.ipynb

Data Collection - Scraping

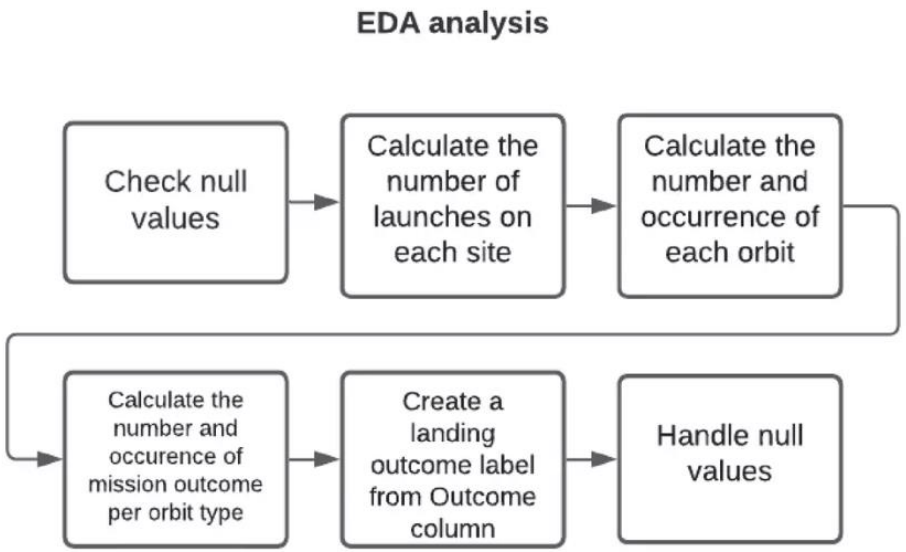
- Web Scrapping from Wikipedia



https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/Data-Collection-With-Web-Scraping.ipynb

Data Wrangling

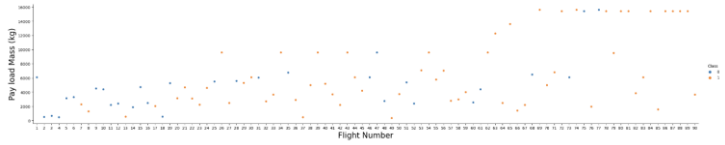
- cleaning and unifying messy and complex data sets for easy access and Exploratory Data Analysis (EDA).

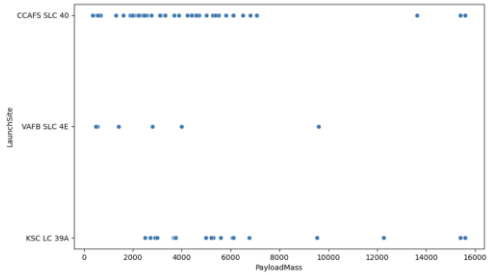


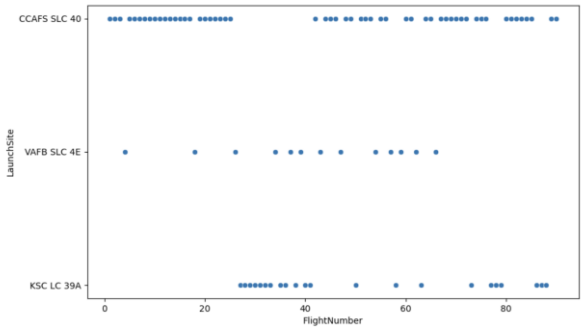
- https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/Data-Wrangling.ipynb

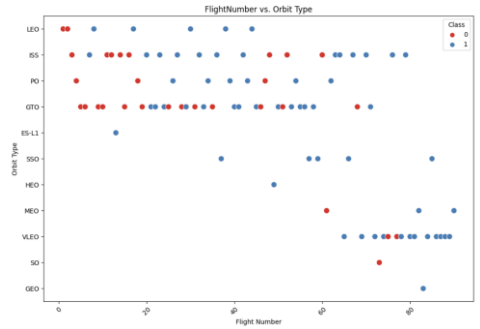
EDA with Data Visualization

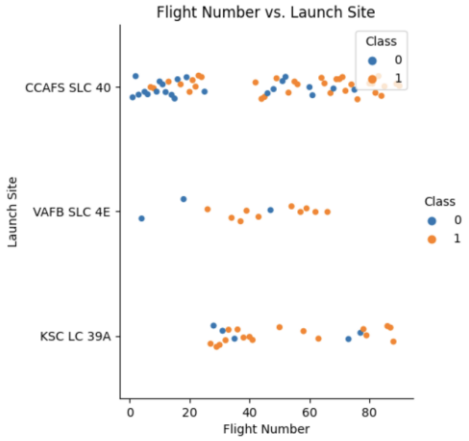
- We first started by using scatter graph to find the relationship between the attributes such as between:
  - Payload and Flight Number.
  - Flight Number and Launch Site.
  - Payload and Launch Site.
  - Flight Number and Orbit Type.
  - Payload and Orbit Type.

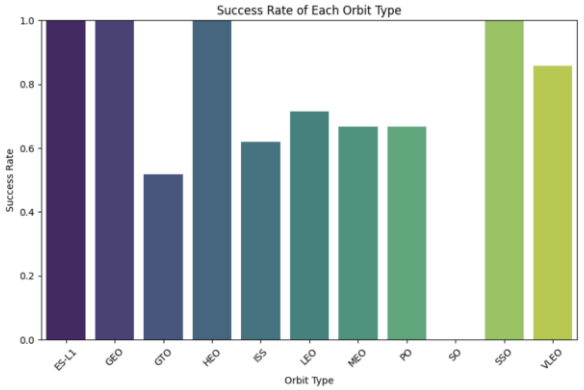












https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/EDA-With-Visualization.ipynb

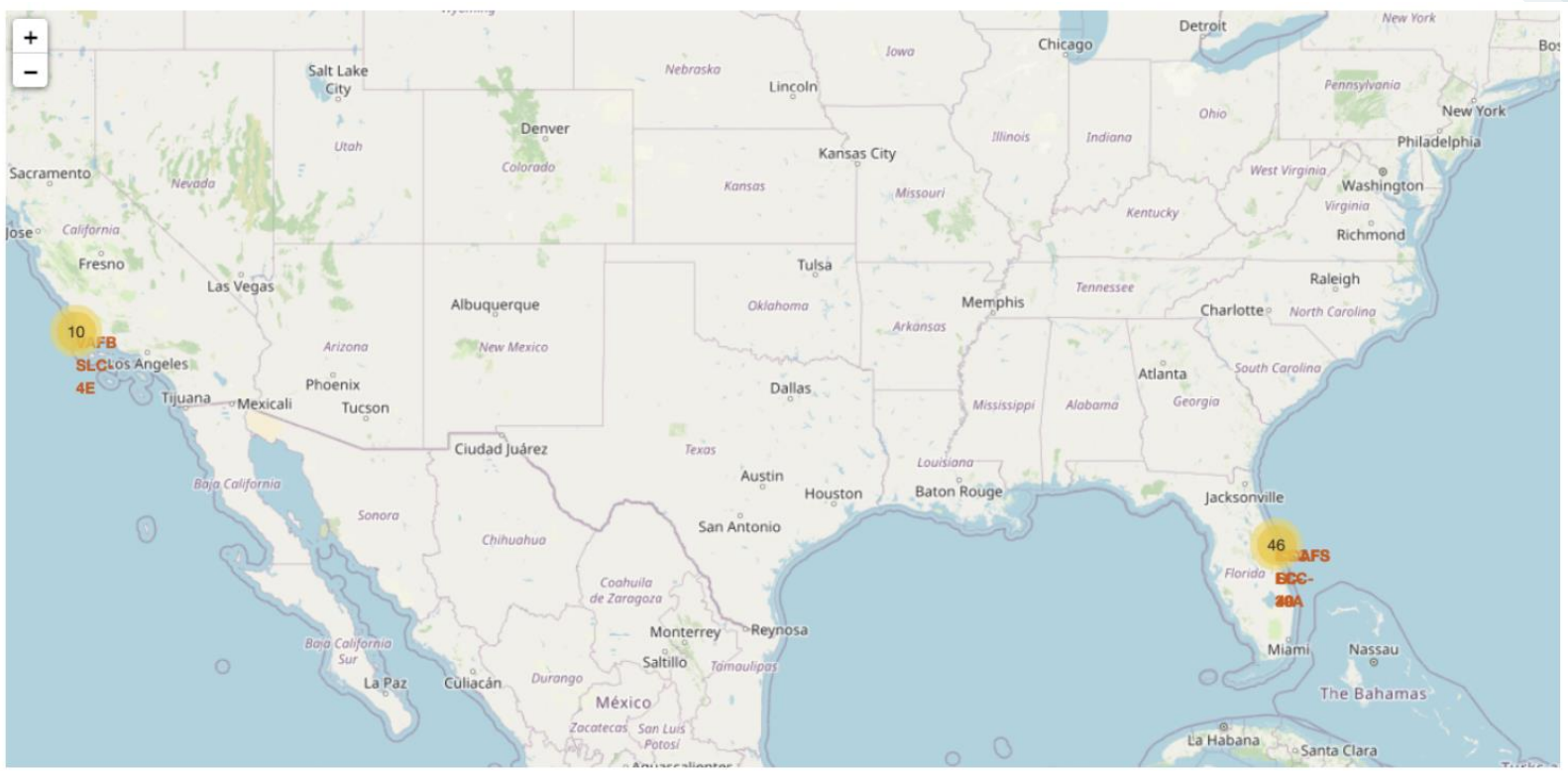
EDA with SQL

- **SQL queries performed include:**
  - - Displaying the names of the launch sites.
  - - Displaying 5 records where launch sites begin with the string ‘CCA’.
  - - Displaying the total payload mass carried by the booster launched by NASA (CRS).
  - - Displaying the average payload mass carried by booster version F9 v1.1.
  - - Listing the date when the first successful landing outcome in the ground pad was achieved.
  - - List the names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000.
  - - Listing the total number of successful and failed 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 year 2015.
  - - Rank the count of landing outcomes or success between the dates 2010-06-04 and 2017-03-20, in descending order.

- https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/EDA-With-SQL.ipynb

Build an Interactive Map with Folium

- Map markers have been added to the map with the aim of finding an optimal location for building a launch site

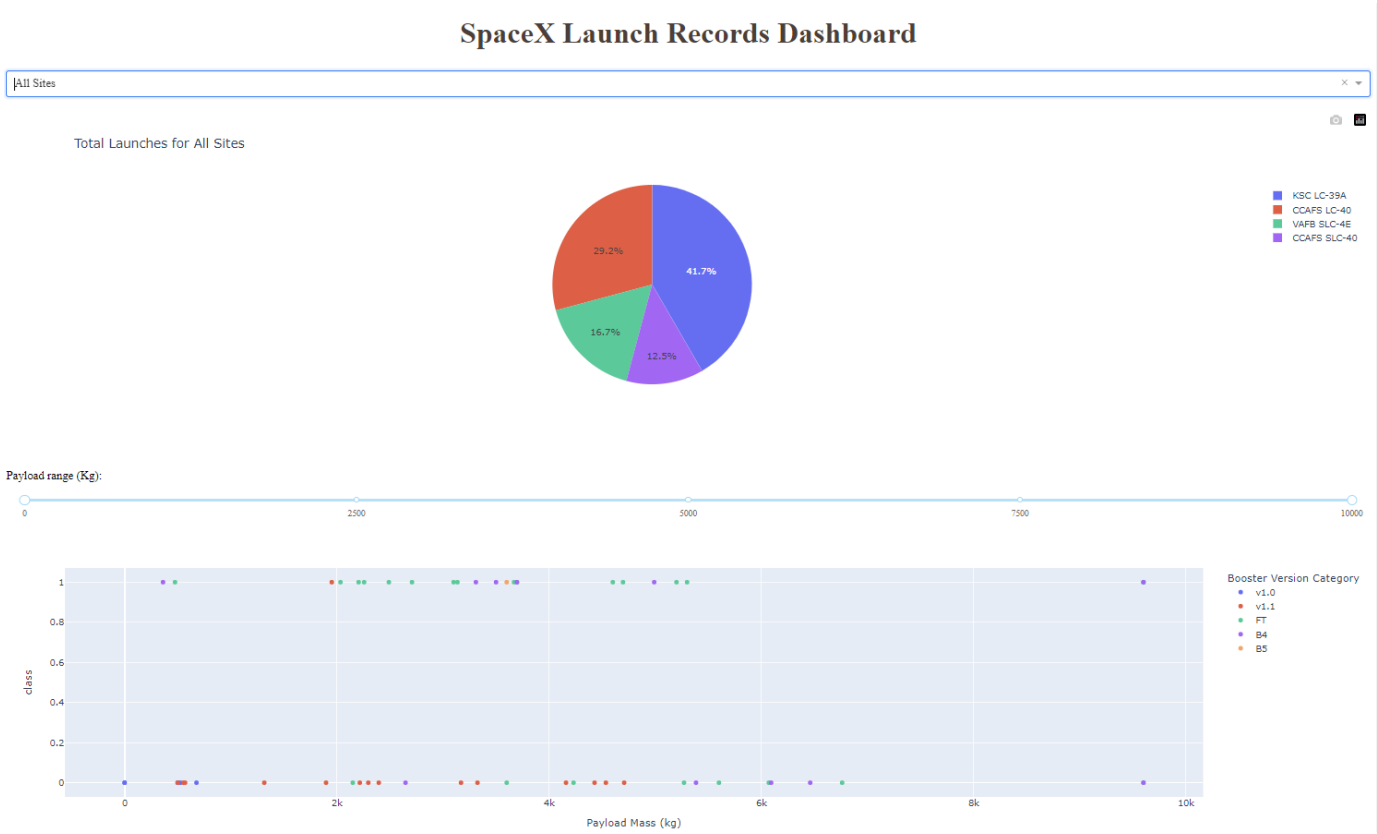


https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/Interactive-Visual-Analytics-with-Folium.ipynb

Build a Dashboard with Plotly Dash

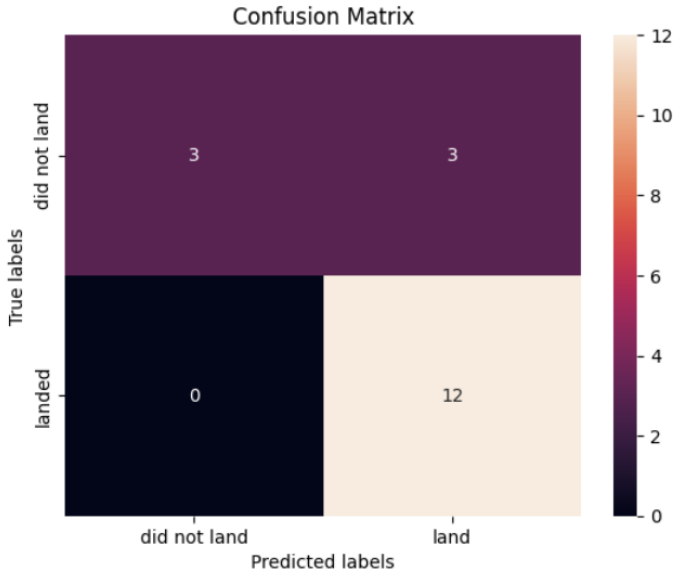
- We plotted pie charts showing the total launches by a certain site.
- We then plotted a scatter graph showing the relationship between Outcome and Payload Mass (Kg) for the different booster versions.

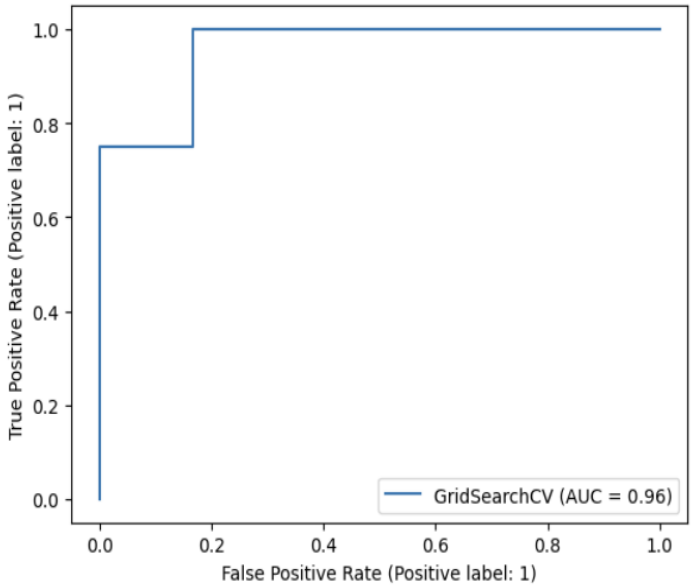
https://github.com/kaveh86/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py



Predictive Analysis (Classification)

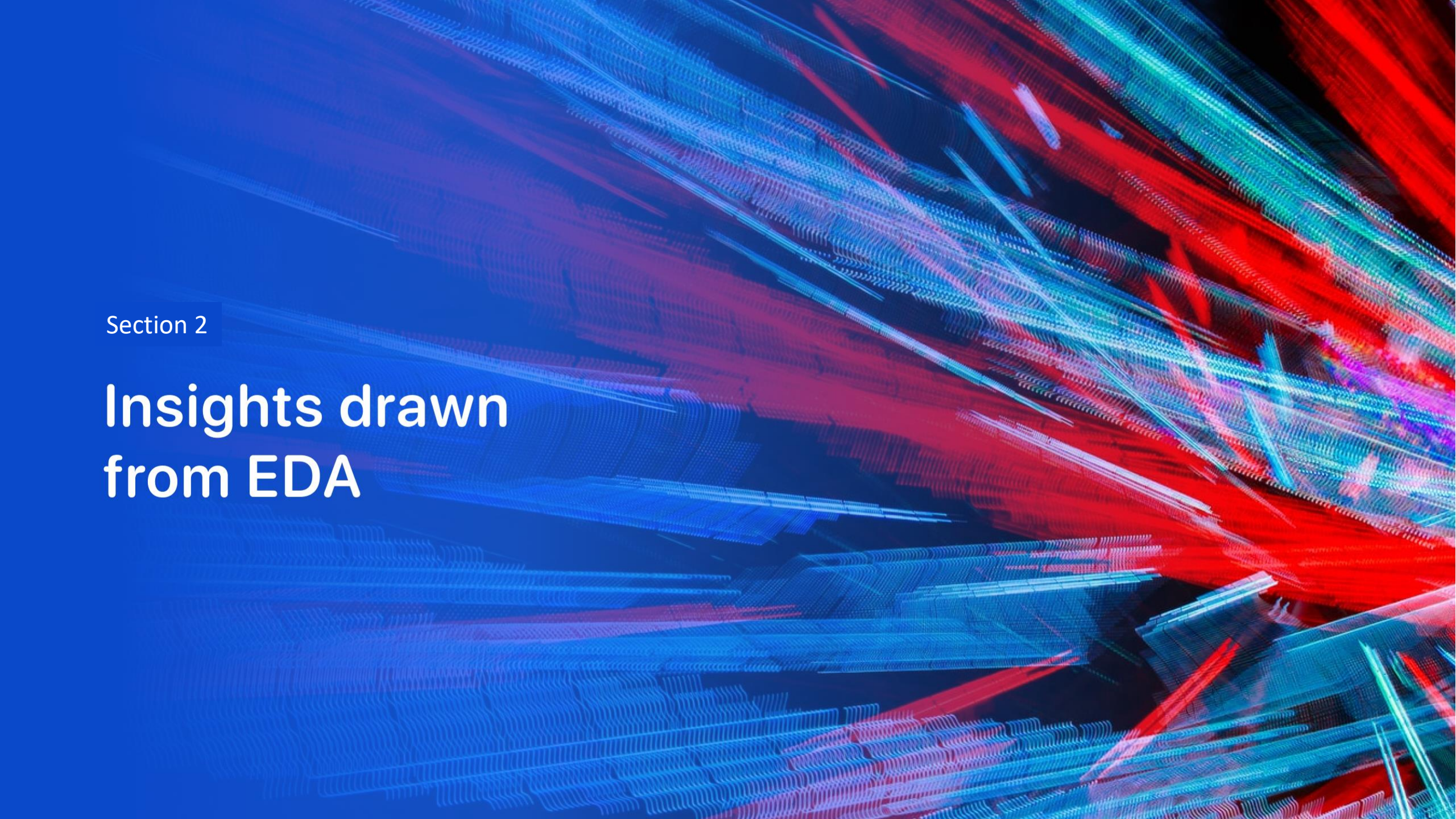
- The SVM, KNN, Logistic regression, and Decision Tree models achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under Curve at 0.96



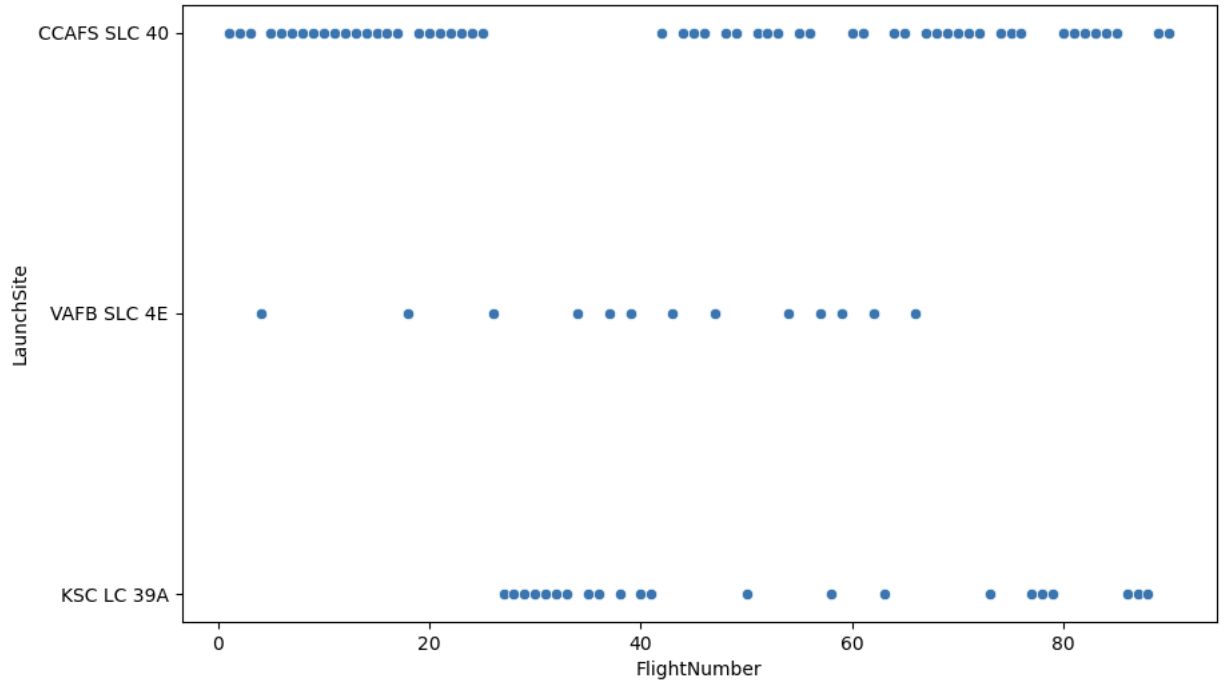


Results

- While all four models exhibit identical accuracy scores, the superiority of the SVM model becomes evident when considering the Area Under Curve (AUC-ROC). This comprehensive evaluation metric establishes SVM as the optimal choice among the models assessed
- Low-weighted payloads perform better than heavier payloads.
- The success rates for SpaceX launches id directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbits GEO, HEO, SSO, and ES L1 have the best Success Rate.

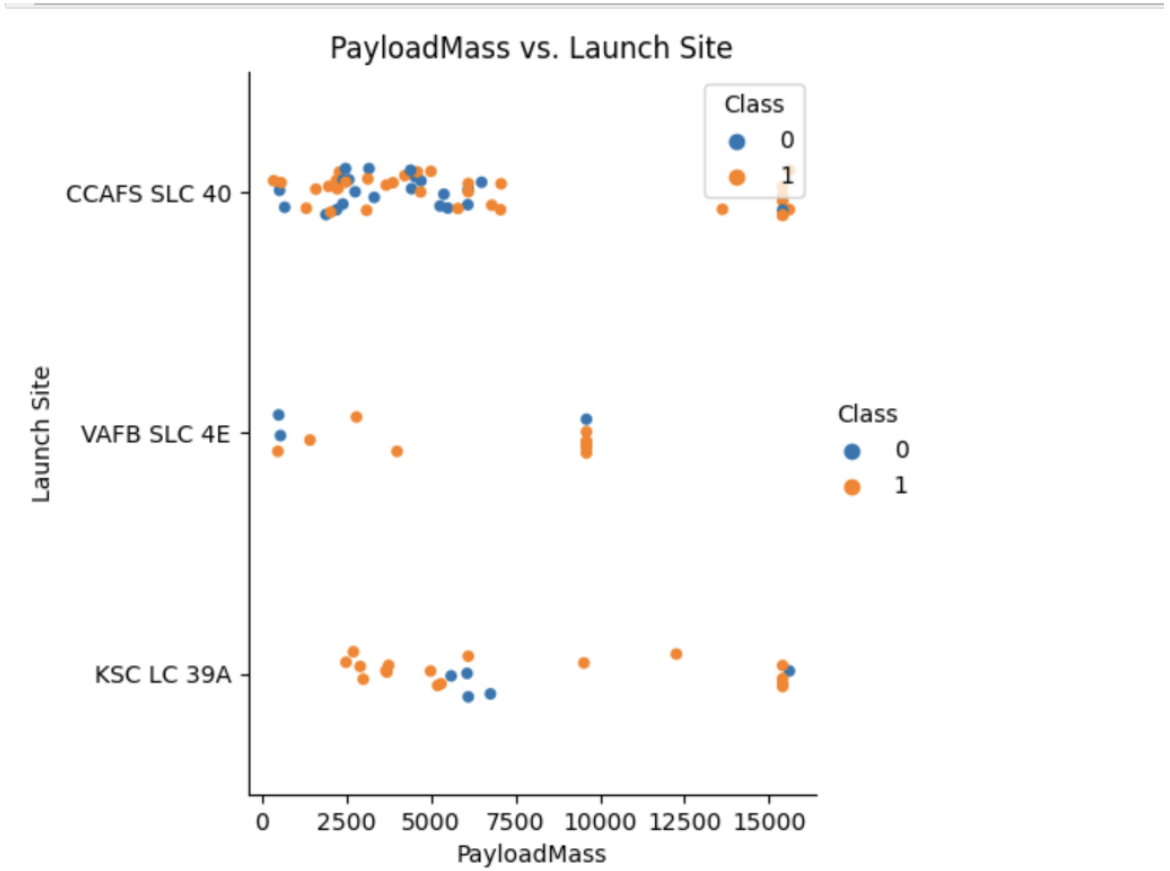

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



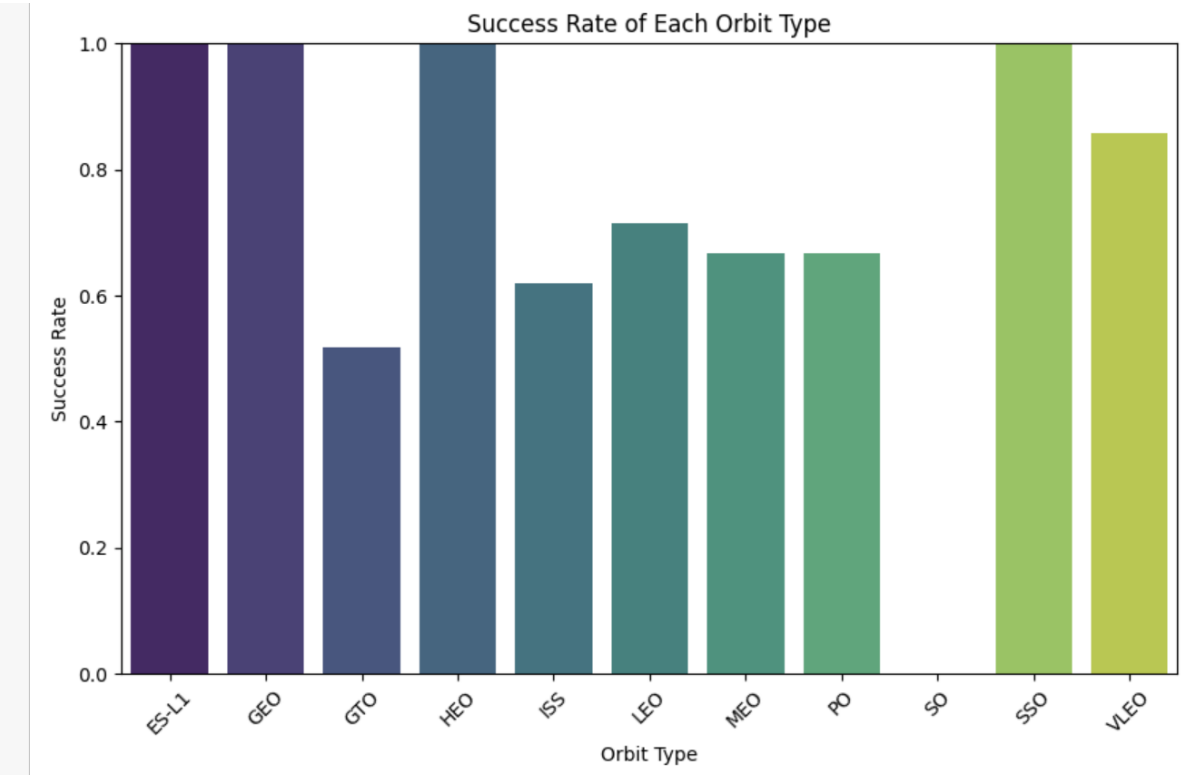
- Launches from the site of CCAFS SLC 40 ate significantly higher than launches from other sites.

Payload vs. Launch Site



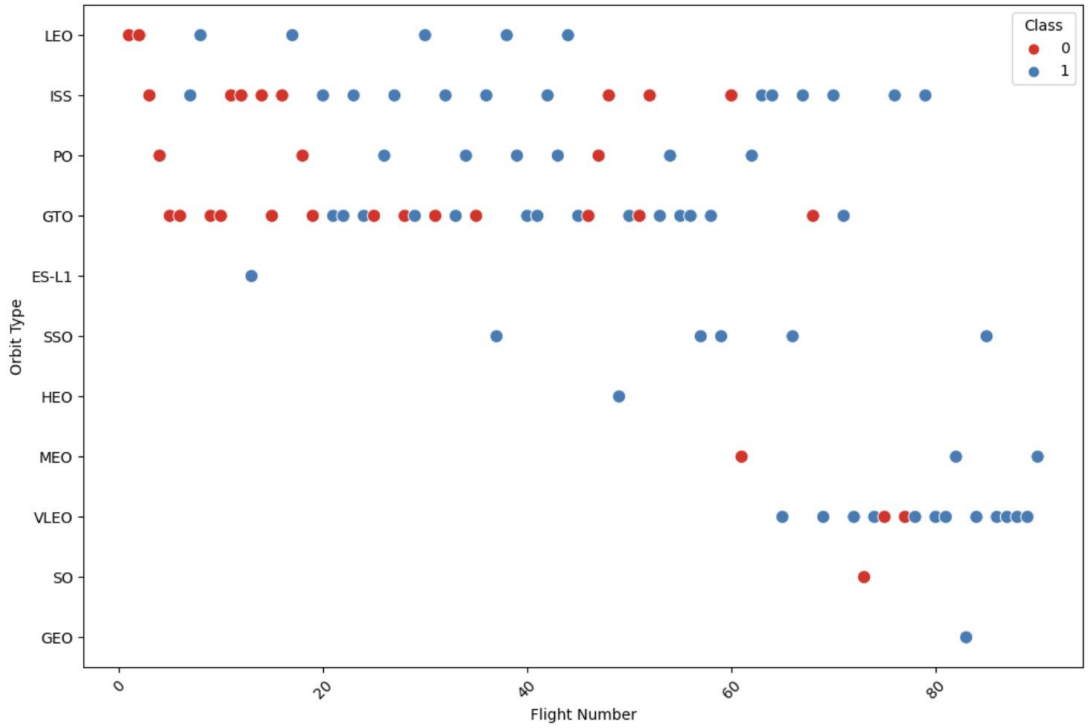
- The Majority of IPay Loads with lower Mass have been launched from CCAFS SLC 40

Success Rate vs. Orbit Type



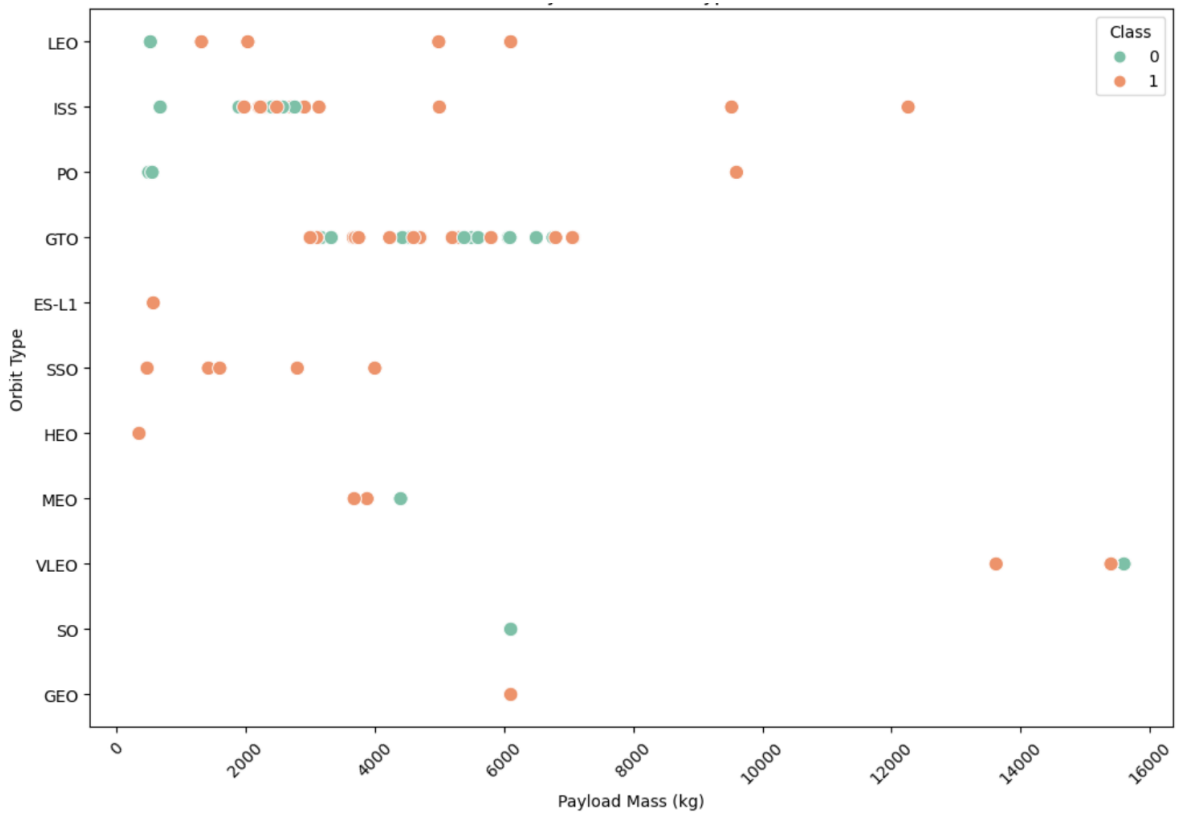
- The orbit types of ES-L1, GEO, HEO, and SSO are the highest success rate.

Flight Number vs. Orbit Type



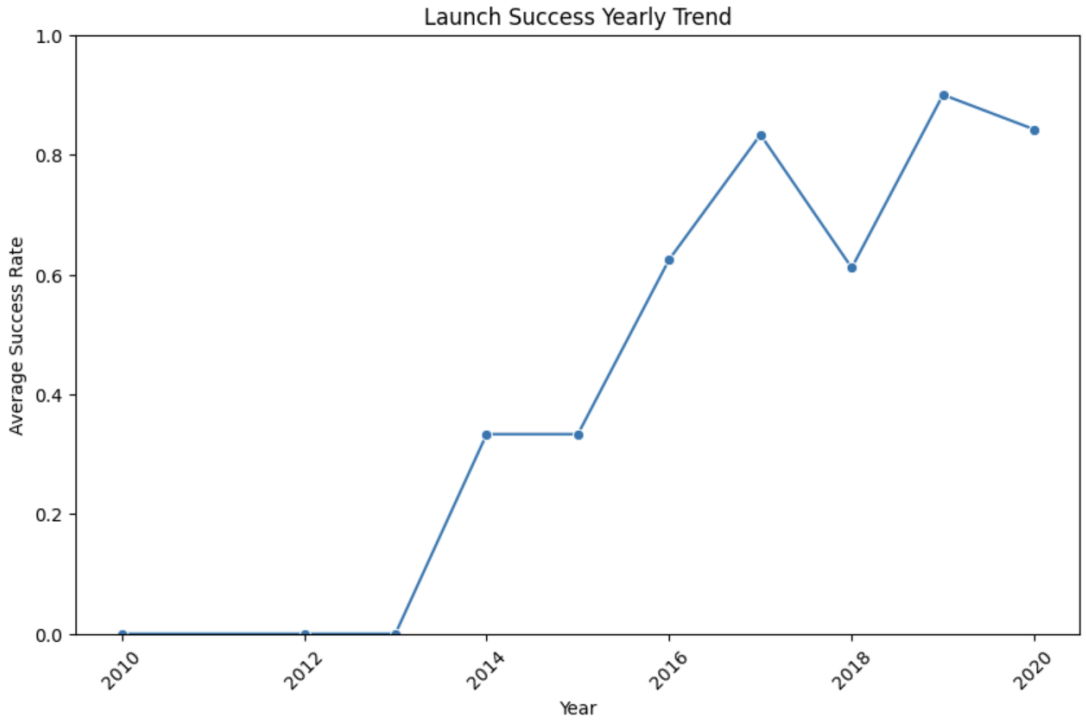
- A trend can be observed of shifting to VLEO launches in recent years.

Payload vs. Orbit Type



- There is a strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stabilized since 2019, potentially due to advances in technology and lessons learned.

All Launch Site Names

- %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

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

Launch Site Names Begin with 'CCA'

- %sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

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

Total Payload Mass

- %sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

- %sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- %sql SELECT min(Date) AS First_Successful_Landing_On_Ground_Pad FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';

First_Successful_Landing_On_Ground_Pad
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_ > 4000 AND
PAYLOAD_MASS__KG_ < 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

- %%sql SELECT Mission_Outcome, COUNT(*) AS Total_Count

FROM SPACEXTABLE

GROUP BY Mission_Outcome;

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

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
    CASE
        WHEN Date like '%-01-%' THEN 'January'
        WHEN Date like '%-02-%' THEN 'February'
        WHEN Date like '%-03-%' THEN 'March'
        WHEN Date like '%-04-%' THEN 'April'
        WHEN Date like '%-05-%' THEN 'May'
        WHEN Date like '%-06-%' THEN 'June'
        WHEN Date like '%-07-%' THEN 'July'
        WHEN Date like '%-08-%' THEN 'August'
        WHEN Date like '%-09-%' THEN 'September'
        WHEN Date like '%-10-%' THEN 'October'
        WHEN Date like '%-11-%' THEN 'November'
        WHEN Date like '%-12-%' THEN 'December'
    END AS Month,
    Landing_Outcome AS Failure_Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTABLE
WHERE Date like'%2015%' and Landing_Outcome LIKE 'Failure (drone ship)';
```

Month	Failure_Landing_Outcome	Booster_Version	Launch_Site
October	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- %%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count

FROM SPACEXTABLE

WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'

GROUP BY Landing_Outcome

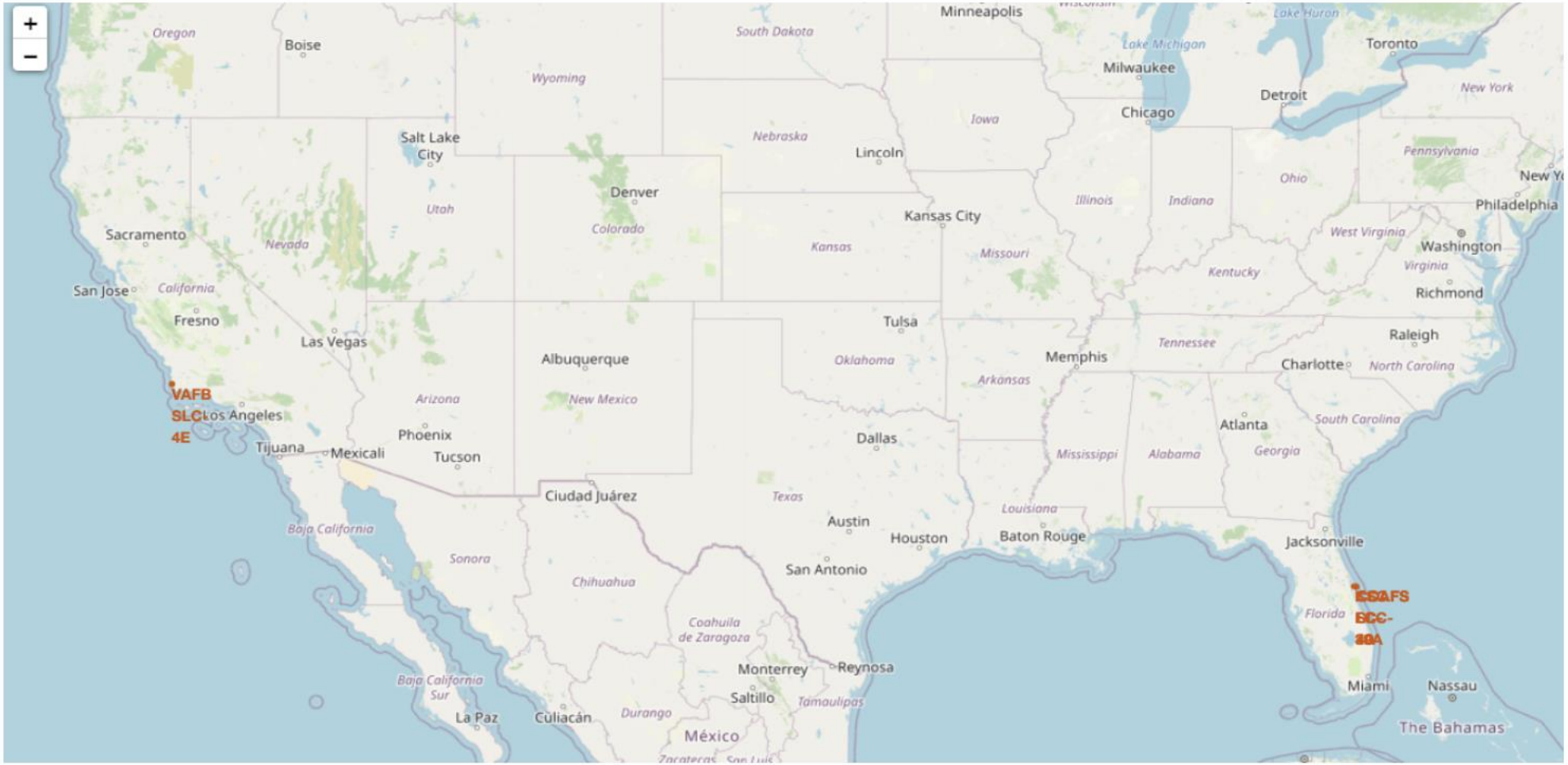
ORDER BY Outcome_Count DESC;

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

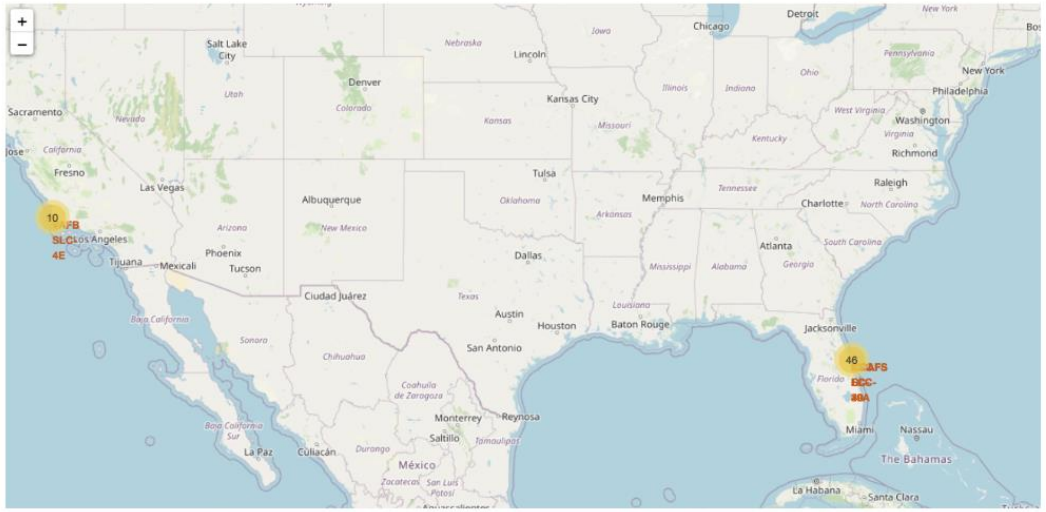
Section 3

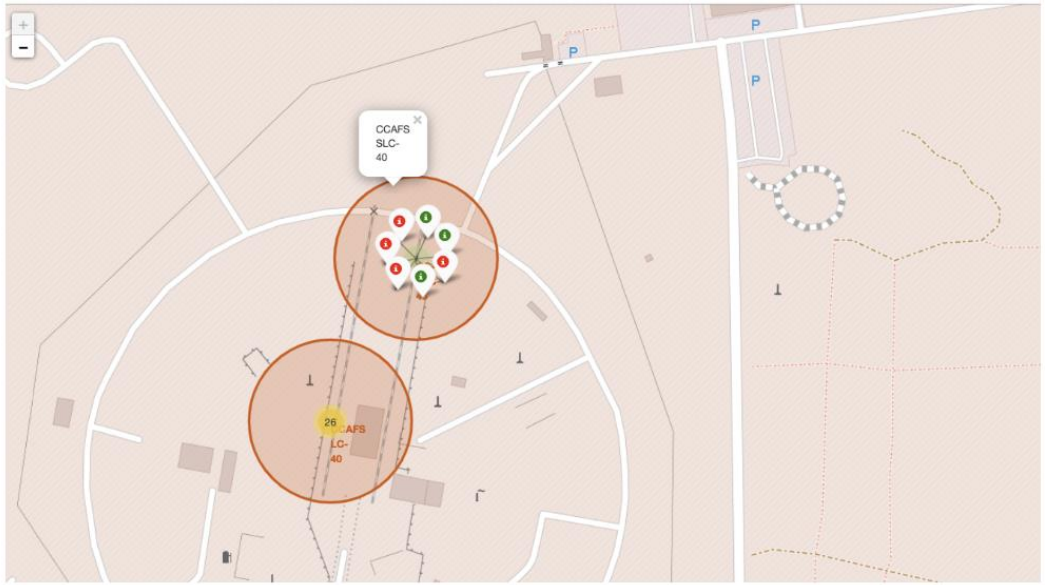
Launch Sites Proximities Analysis

All launch sites marked on map

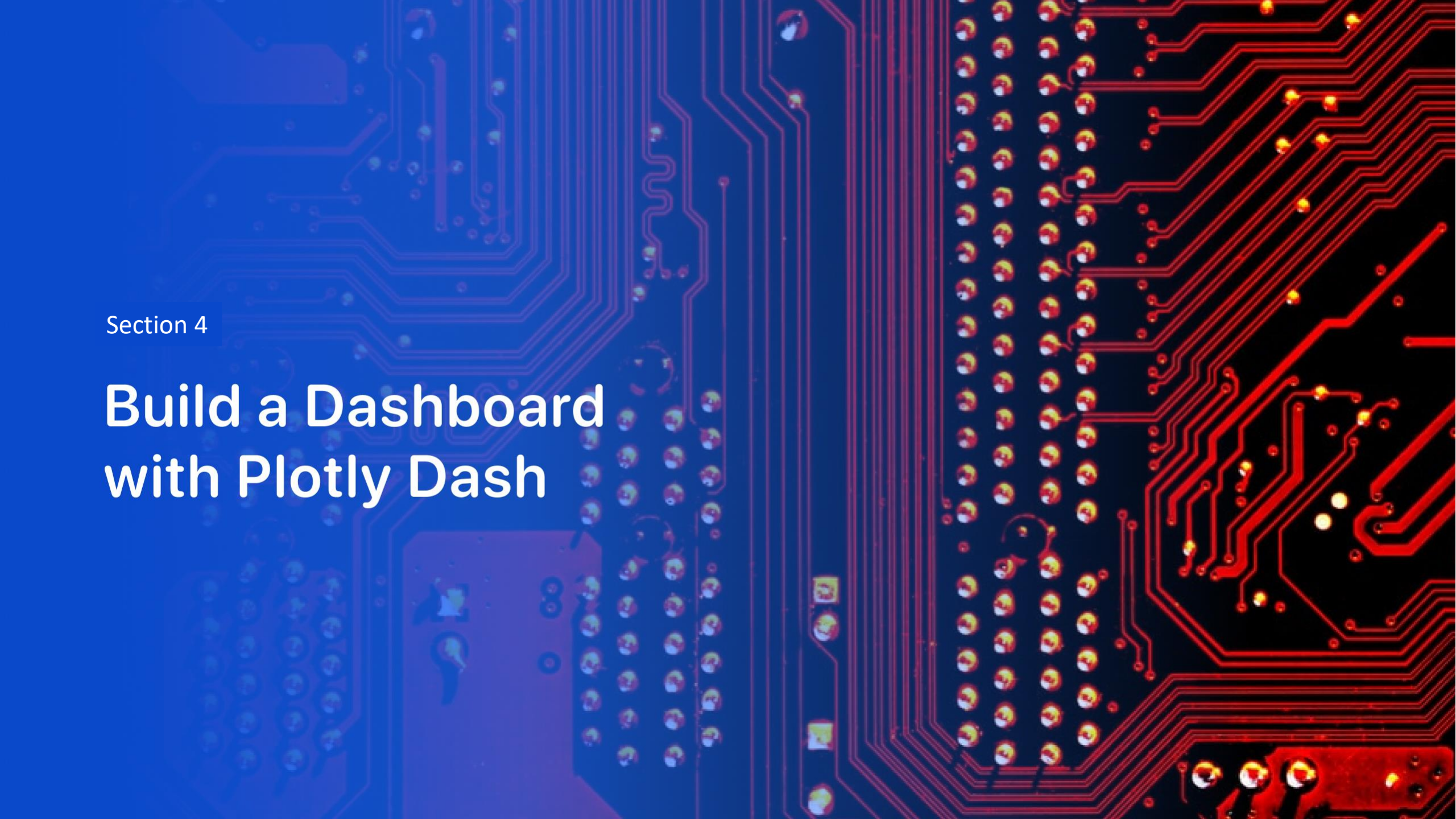


Success/Failed launches marked on the map

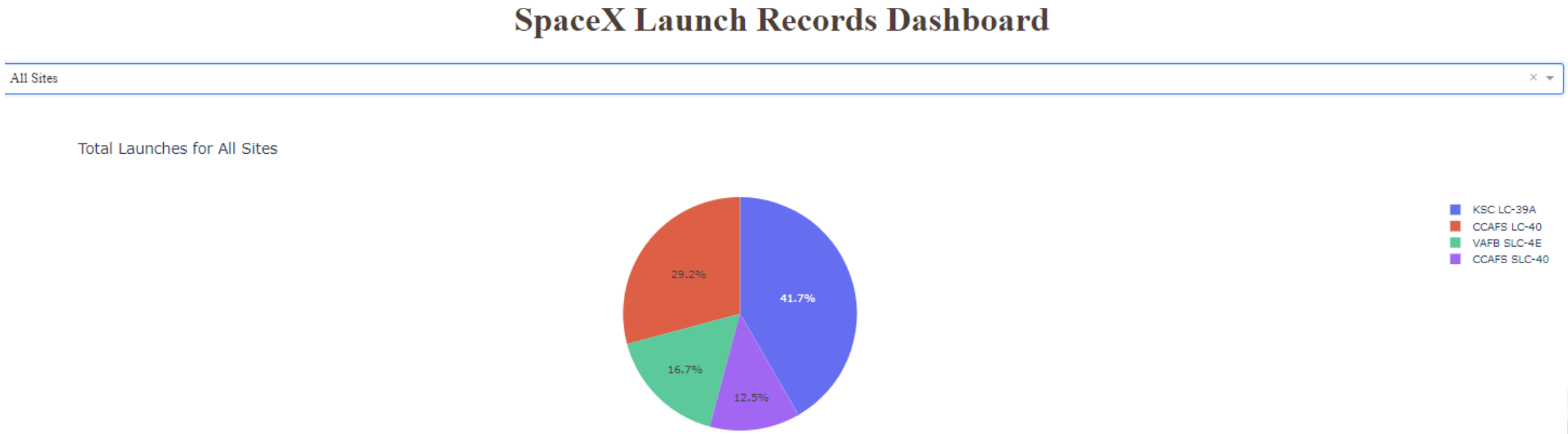




Distances between a launch site to its proximities

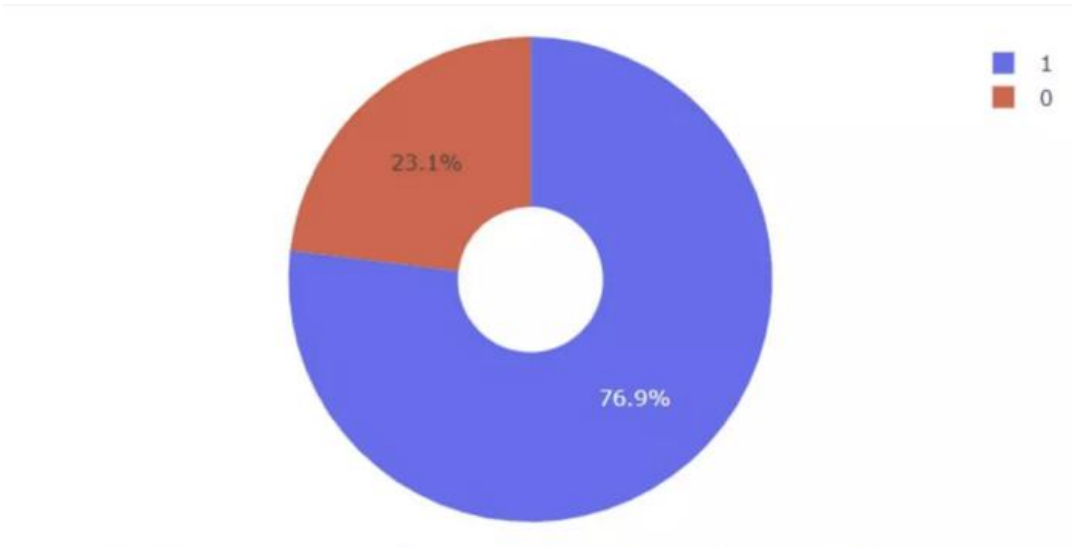

Section 4
Build a Dashboard with Plotly Dash

Total success launches by all sites



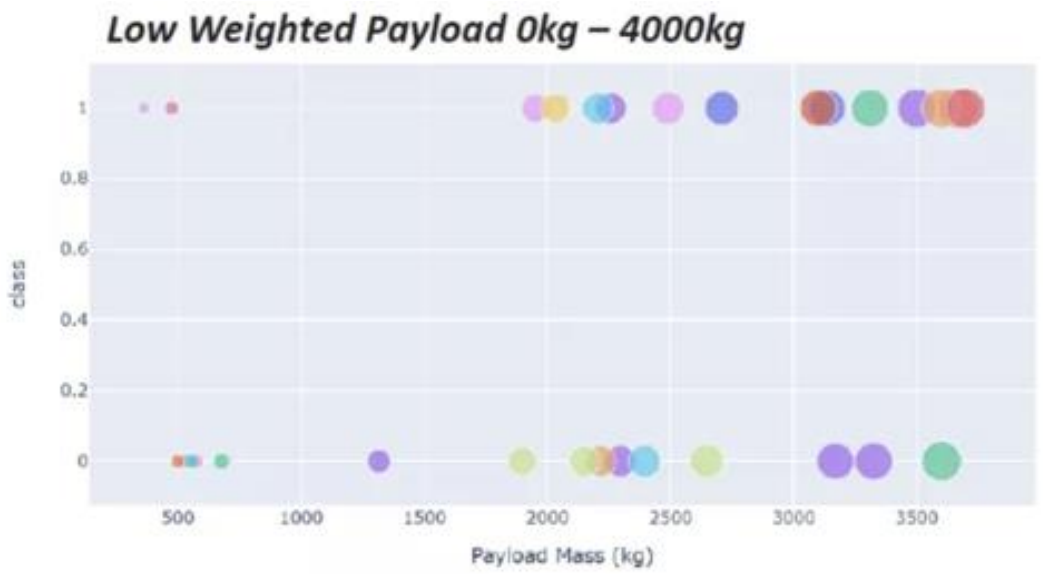
- As we can see KSC LC 39A has the most successful launches

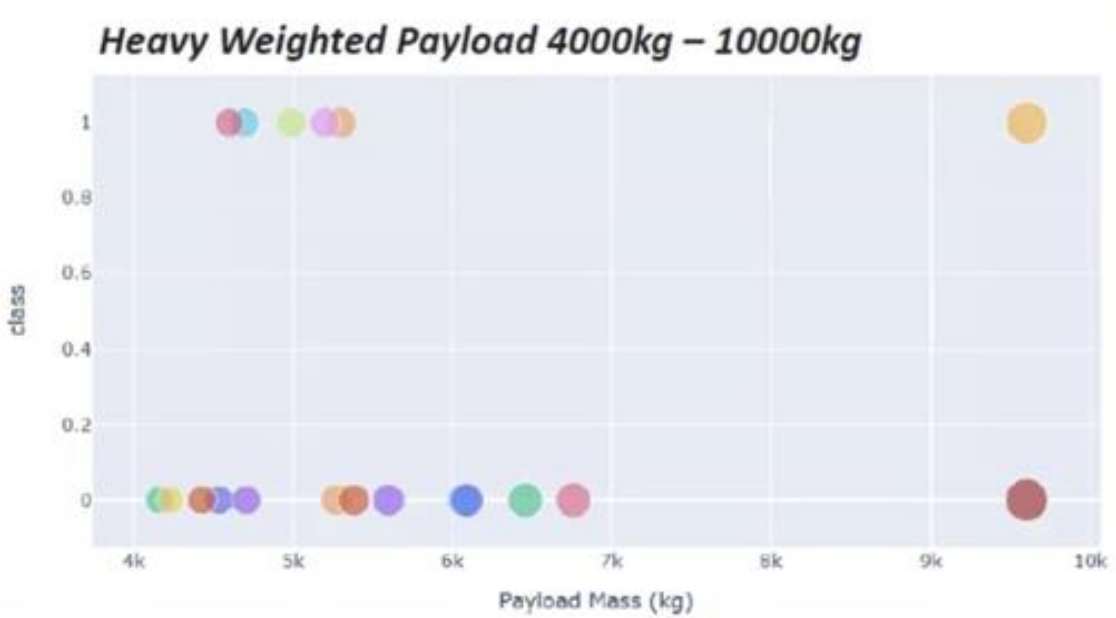
Success rate by site



- KSC LC 39A achieved a 76.9% success rate getting a 23.1% failure rate

Payload vs Launch outcome





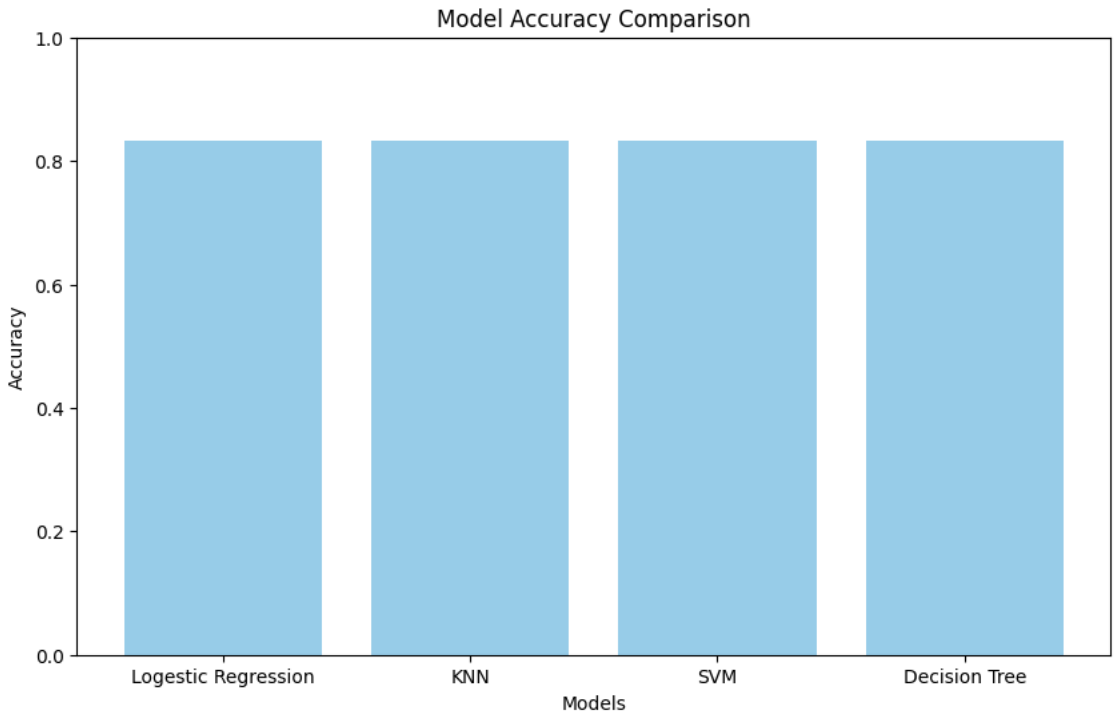
We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

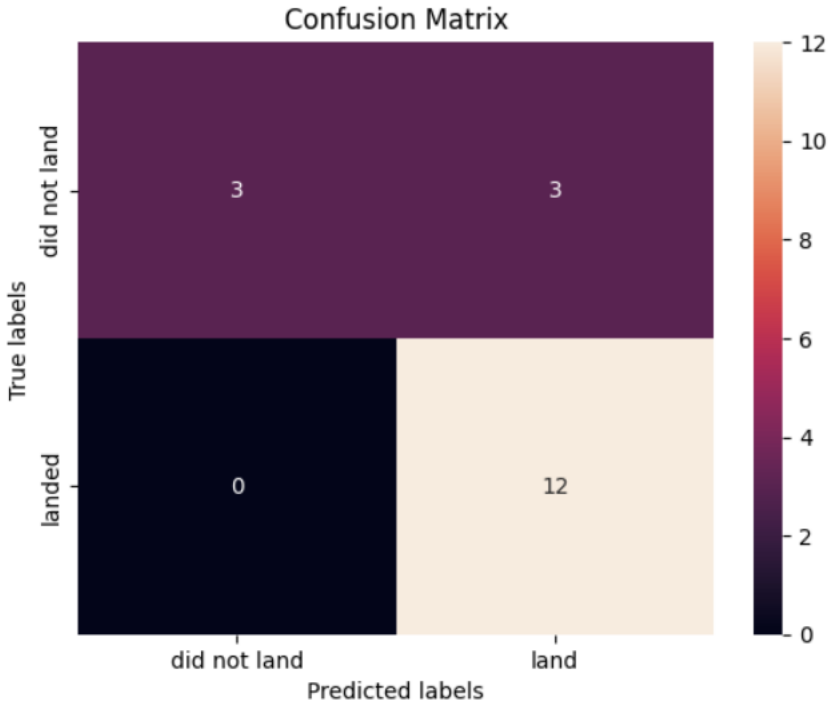
Classification Accuracy

- The accuracy scores for all four models are the same:



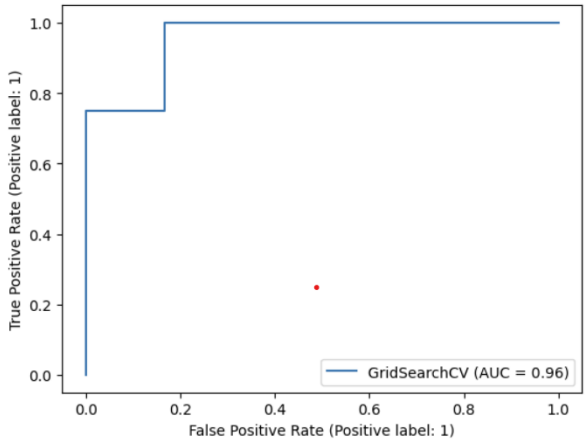
Confusion Matrix

- And Confusion Matrix results are also the same for all four models

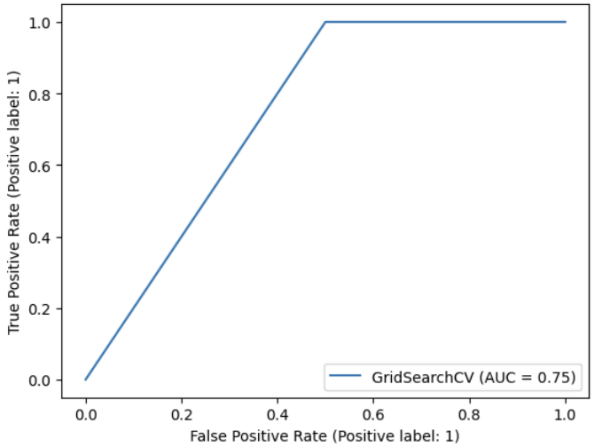


ROC Curve of the models:

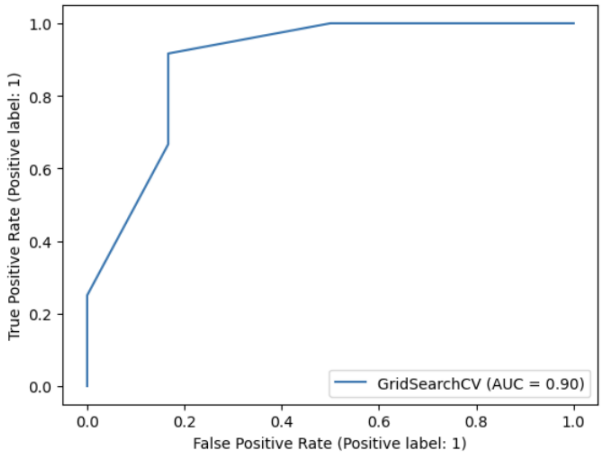
SVM



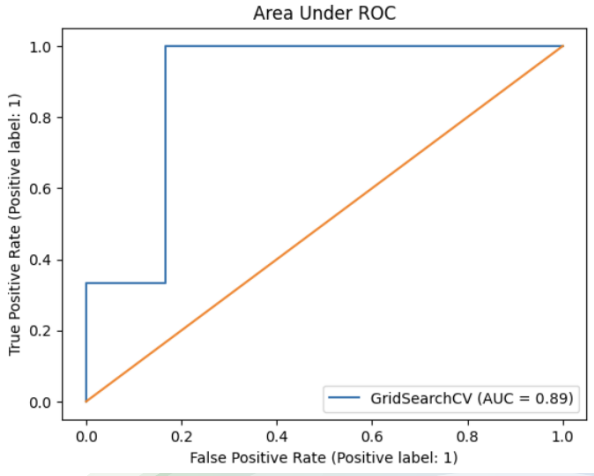
Decision Tree



KNN



Logistic Regression



- The best ROC Curve is for SVM Model

Conclusions:

- Although all Models have the same accuracy and confusion matrix, SVM is the best model for prediction for this dataset because of the better area under the ROC curve.
- Low-weighted payloads perform better than heavier payloads.
- The success rates for SpaceX launches are directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbits GEO, HEO, SSO, and ES L1 have the success rate

Thank you!