IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tiago Amador Coelho
01/01/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Collecting the Data on the Falcon 9 first-stage landings
  - Data Wrangling
  - Exploratory Analysis Using SQL
  - Visual analytics and dashboard to analyze launch records interactively
  - Create Models to determine if the first stage of Falcon 9 will land successfully
- Summary of all results
  - Launch success has improved over time
  - Space X Falcon 9 First Stage Landing predict

Introduction

Background and context

SpaceX is an aerospace company founded in 2002 that stands out worldwide in space transportation services. Its success is due to the Falcon 9 rockets, in which Space X can reuse its first stage. Other companies have a cost of 165 million dollars for each rocket launched, but because Space X can reunite the first stage of the Falcon 9, the cost of launching each rocket is 62 million dollars.

Problem to answer

It is possible to predict Space X Falcon 9 First Stage Landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

SpaceX REST API

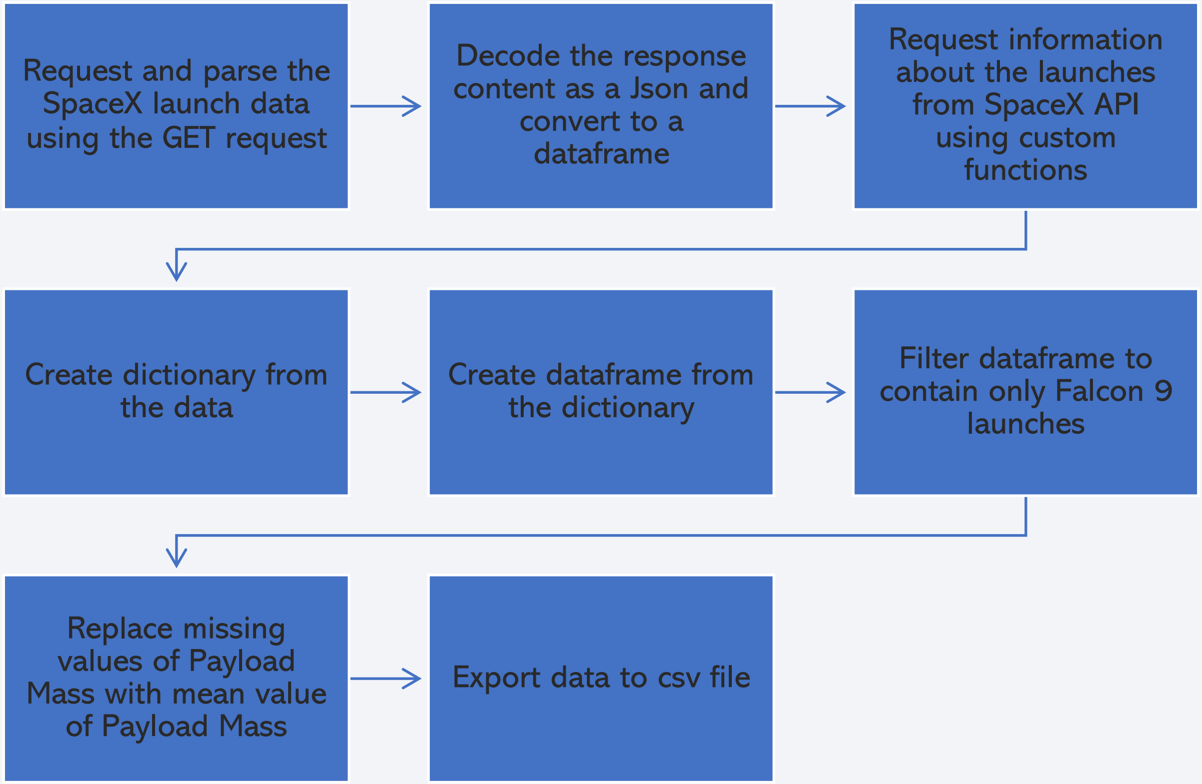
Request and parse the SpaceX launch data → Filter the dataframe to only include `Falcon 9` launches → Dealing with Missing Values → Export it to a CSV

Web scraping Wikipedia

Request the Falcon9 Launch Wiki page from its URL → Extract all column/variable names from the HTML table header → Create a data frame by parsing the launch HTML tables → Export it to a CSV

Data Collection – SpaceX API

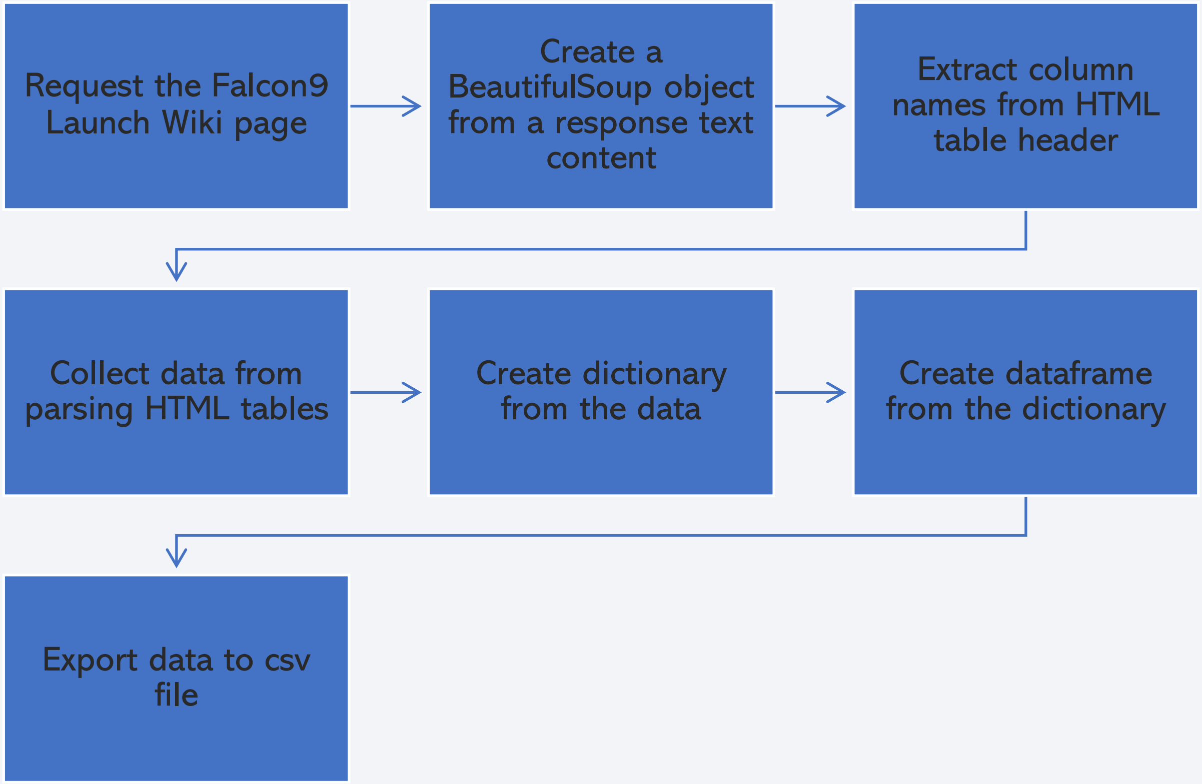
- Steps
  - Request and parse the SpaceX launch data using the GET request
  - Decode the response content as a Json and convert to a dataframe
  - Request information about the launches from SpaceX API using custom functions
  - Create dictionary from the data
  - Create dataframe from the dictionary
  - Filter dataframe to contain only Falcon 9 launches
  - Replace missing values of Payload Mass with mean value of Payload Mass
  - Export data to csv file



https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

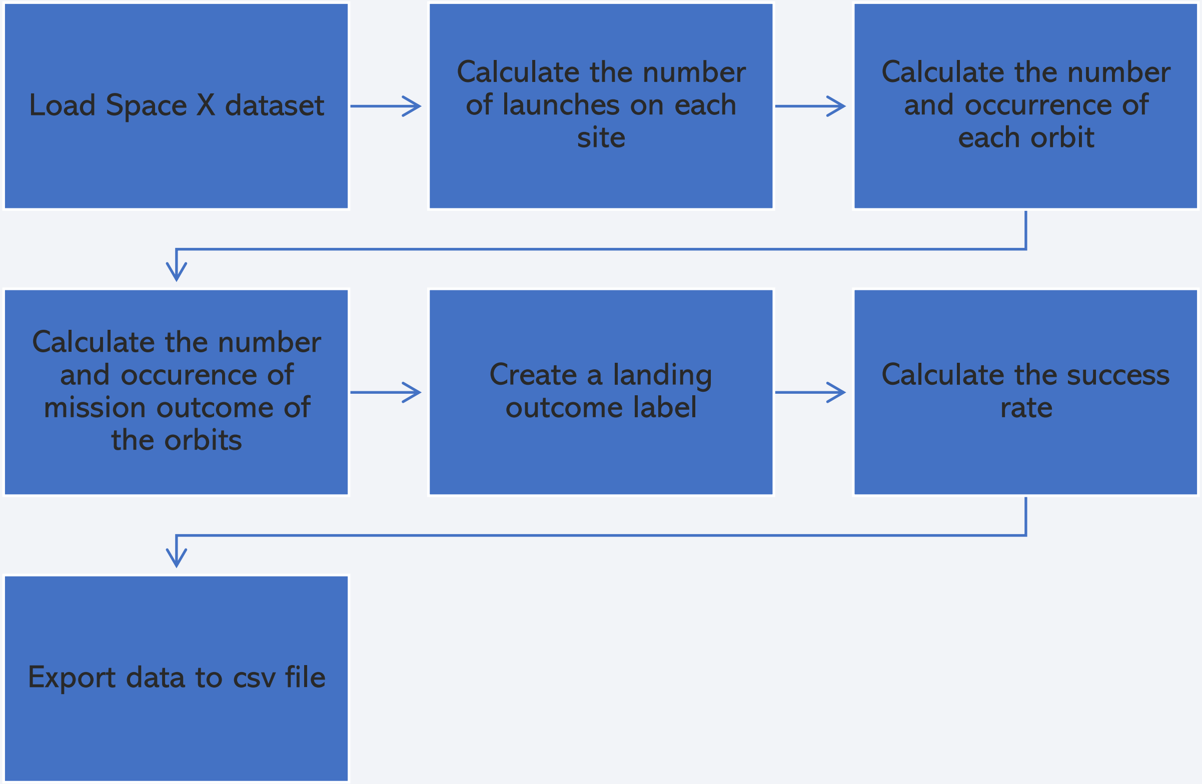
- Steps
  - Request the Falcon9 Launch Wiki page
  - Create a BeautifulSoup object from a response text content
  - Extract column names from HTML table header
  - Collect data from parsing HTML tables
  - Create dictionary from the data
  - Create dataframe from the dictionary
  - Export data to csv file



https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Steps
  - Load Space X dataset
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurence of mission outcome of the orbits
  - Create a landing outcome label
  - Calculate the success rate
  - Export data to csv file



https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Some charts were chosen to analyze data to summarize main characteristics, gain a better understanding of the data set, uncover relationships between different variables, and extract important variables for the problem
  - Flight Number vs Pay Load Mass
  - Flight Number vs Launch Site
  - Pay Load Mass vs Launch Site
  - Orbit Type vs Success Rate
  - Flight Number vs Orbit
  - Pay Load Mass vs Orbit
  - Date vs Class
  - Year vs Success Rate

https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

Queries

Display

- 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:

- 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
- 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.

https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Create a blue circle at NASA Johnson Space Center's coordinate
- For each launch site, add a Circle object based on its coordinate (Lat, Long) values
- Mark the success(green)/failed(red) launches for each site on the map
- Calculate the distances between a launch site to its proximities (city, railway, highway)
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

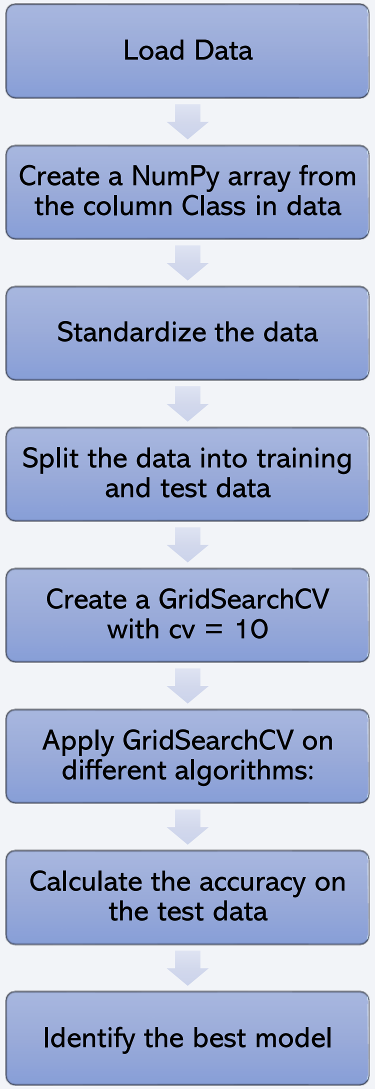
- A dropdown list to enable Launch Site selection
- A pie chart to show the total successful launches
- A slider to select payload range
- A scatter chart to show the correlation between payload and launch success

https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis

- Load Space X Data from Data Base.
- Create a NumPy array from the column Class in data.
- Preprocess the data with StandardScaler(), fit and transform it.
- Split the data into training and test data.
- Create a GridSearchCV with cv = 10.
- Apply GridSearchCV on different algorithms: logistic regression, SVC, decision tree, KNN.
- Calculate the accuracy from all models on the test data.
- Identify the best model using accuracy.

https://github.com/tiagoamador/IBM_Data_Science_Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



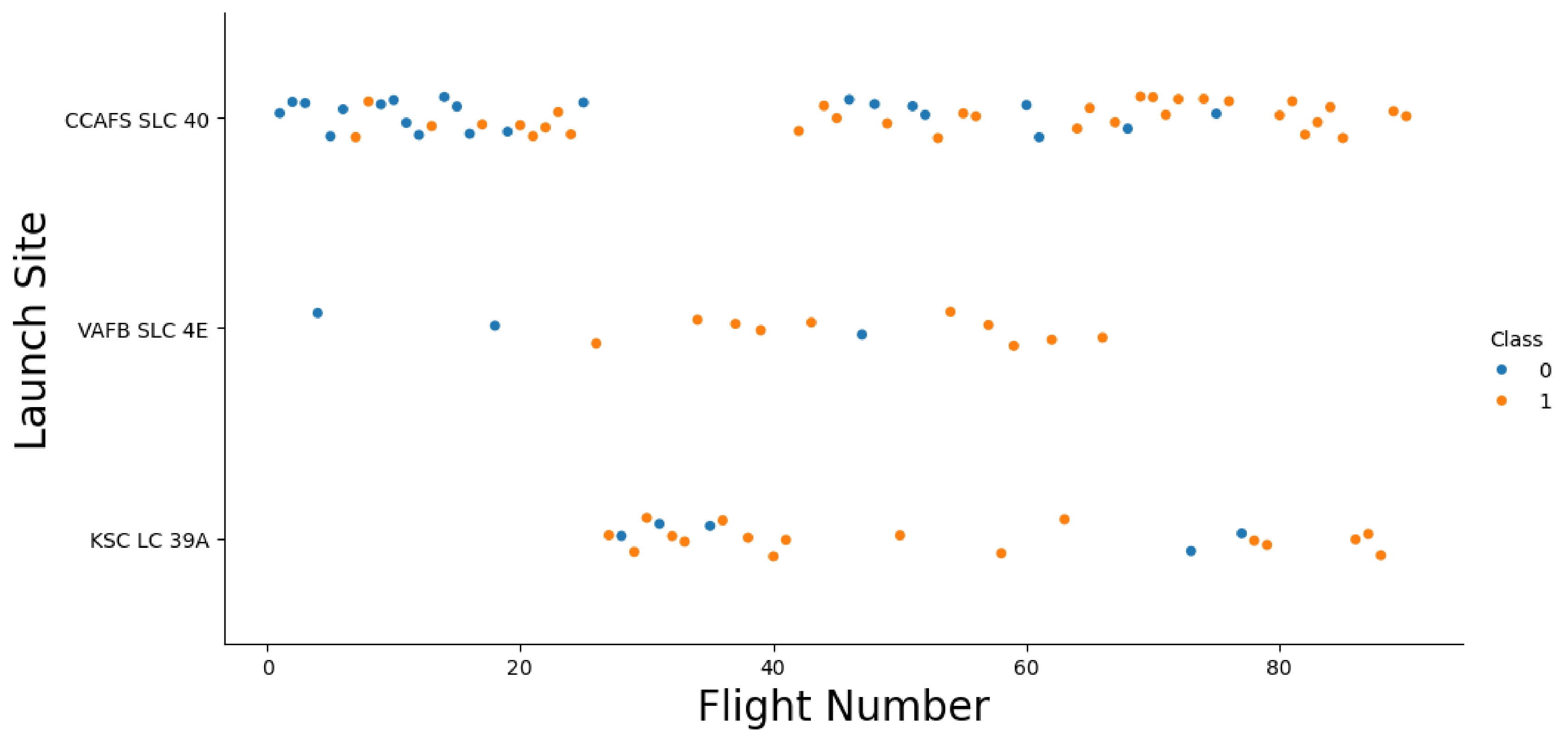
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

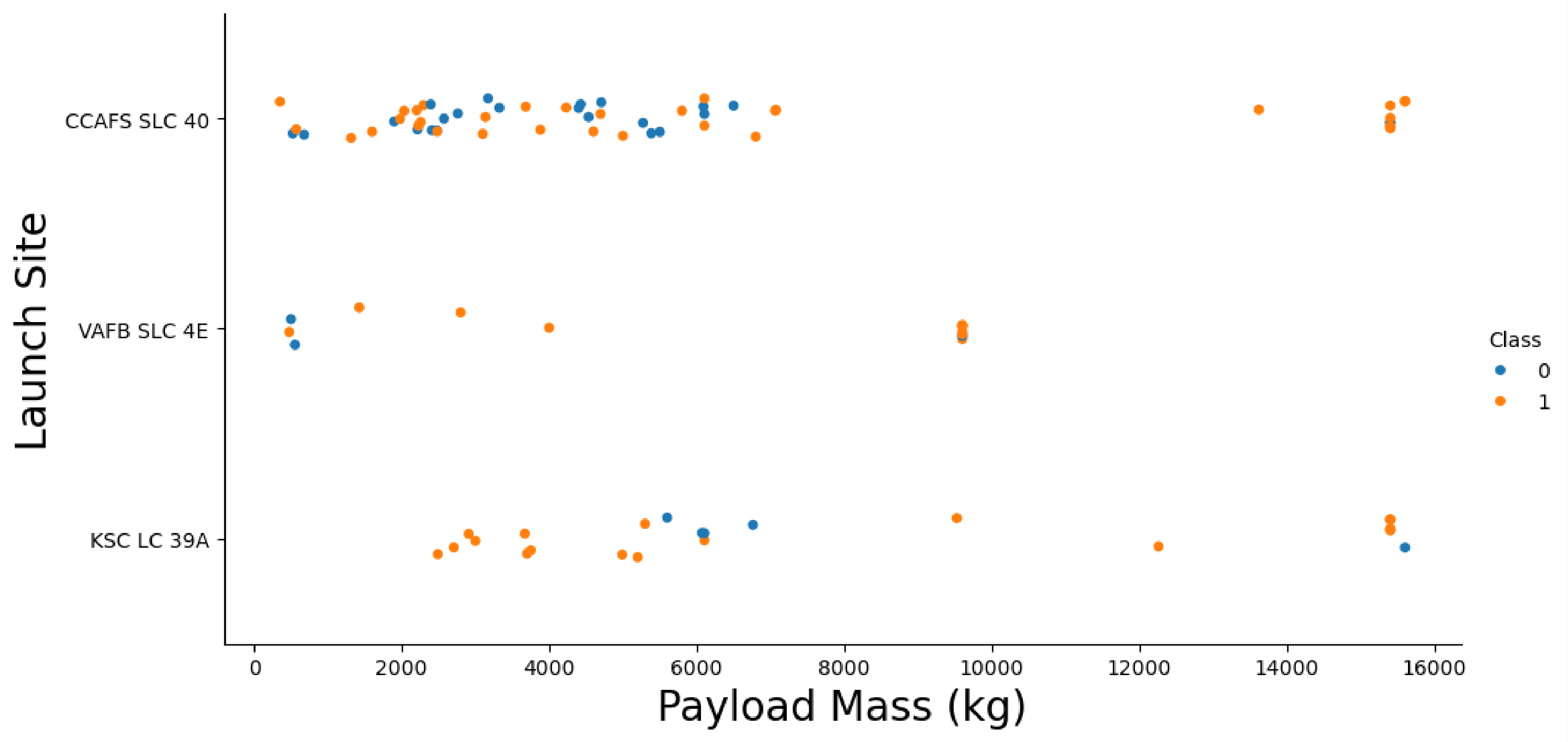
Insights drawn from EDA

Flight Number vs. Launch Site



- Class 0 (blue) = fail
- Class 1 (orange) = success
- New launches have a higher success rate

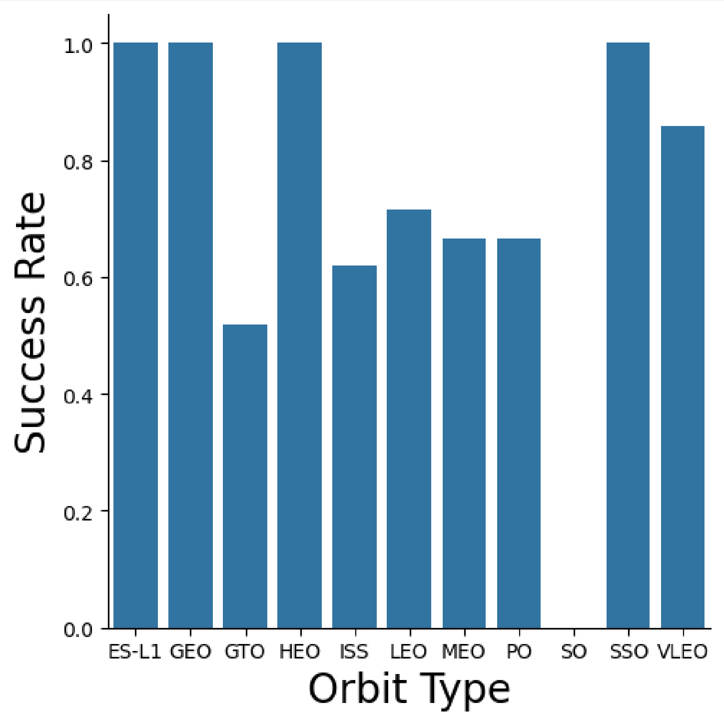
Payload vs. Launch Site



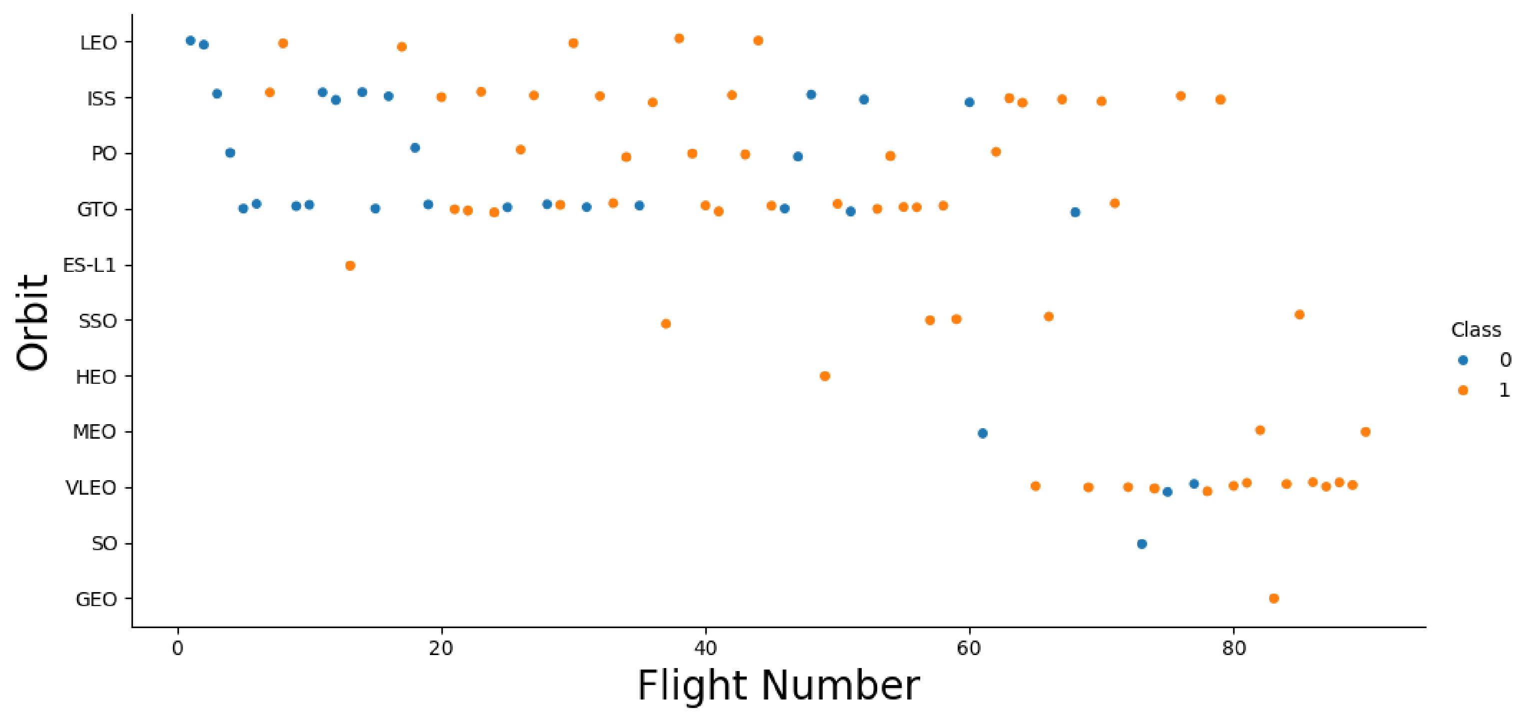
- Rockets with a payload greater than 7000 have a better chance of success

Success Rate vs. Orbit Type

- There are 4 orbits in which all launches were successful
- There is 1 orbit that was not successfully launched

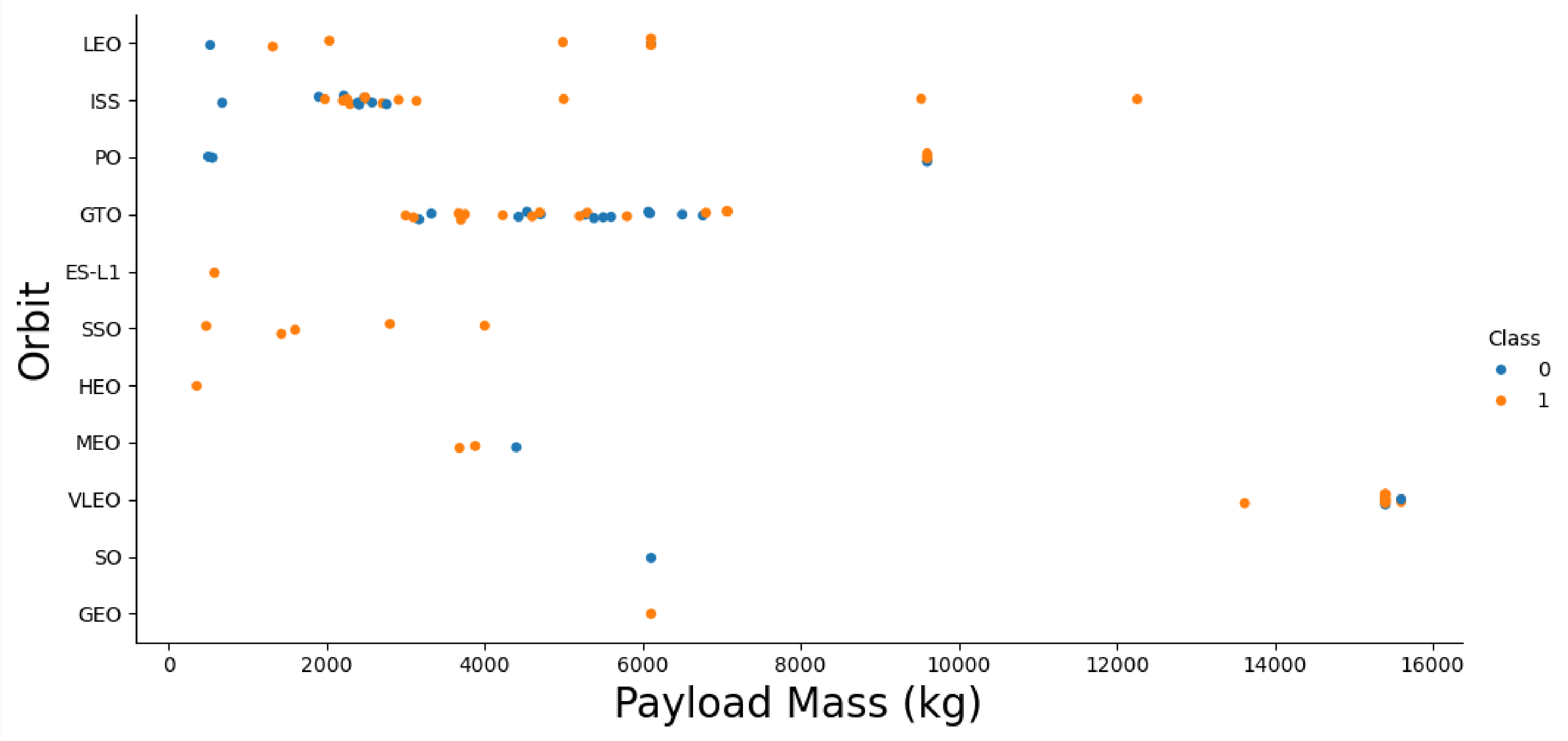


Flight Number vs. Orbit Type



- In the first 6 launches there was fail
- The fail rate decrease with the number of flights

Payload vs. Orbit Type

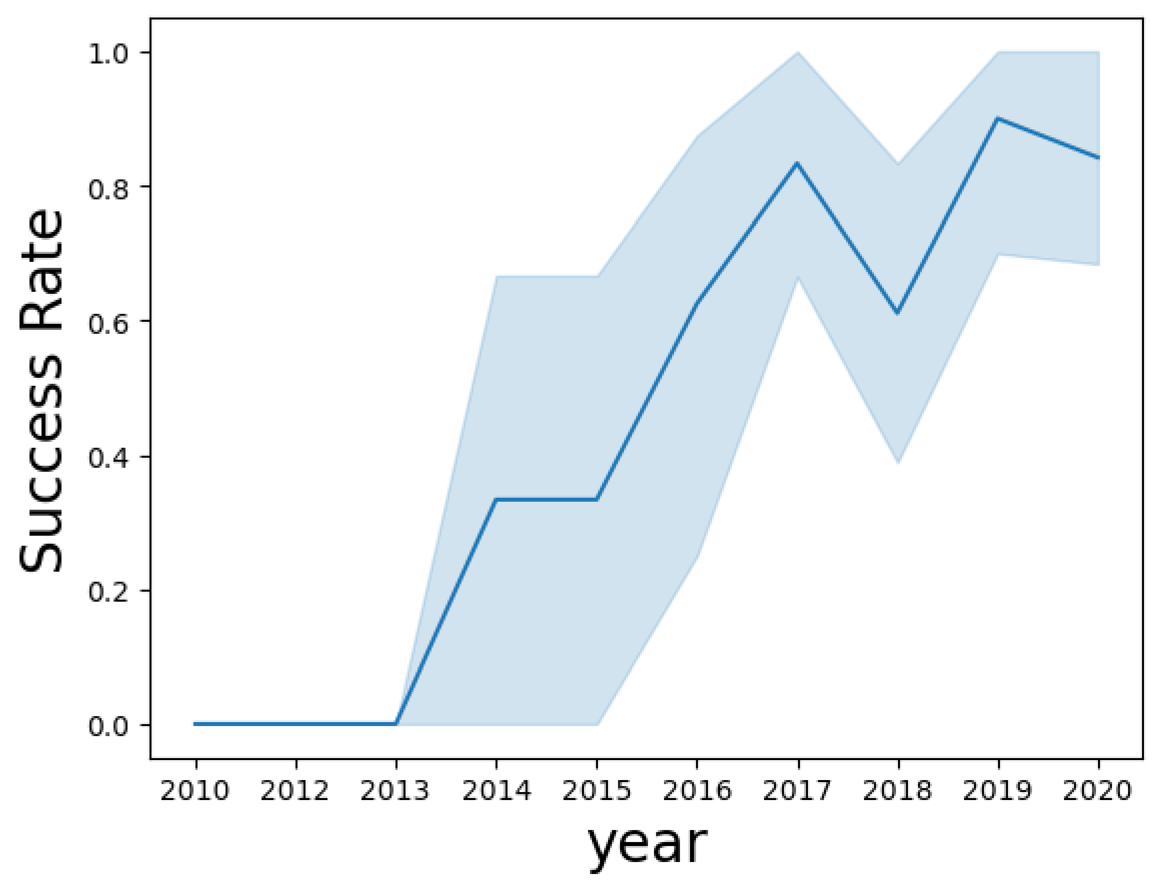


- Orbits that were not used with a payload greater than 8000:

    LEO, ES-L1, GTO, SSO, HEO, MEO, SO, GEO

Launch Success Yearly Trend

- The success rate has been on a growth trend since 2013, with small declines in 2017 and 2020



All Launch Site Names

- Launch sites names
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

```
In [8]:  %sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTABLE;

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

Out[8]:

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

Launch Site Names Begin with 'CCA'

- Display 5 site names with CCA:

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

In [9]:
```
%sql SELECT * FROM SPACEXTABLE WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
```

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

Out[9]:

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)

```
In [10]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE CUSTOMER = 'NASA (CRS)';

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

Out[10]:

SUM(PAYLOAD_MASS__KG_)
45596

45596,0 kg was carried by boosters launched by NASA

Average Payload Mass by F9 v1.1

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

In [11]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

Out[11]:

AVG(PAYLOAD_MASS__KG_)
2928.4

2928,8 kg was carried by boosters launched by F9 v1.1

First Successful Ground Landing Date

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

Hint:Use min function

In [12]:
```
%sql SELECT MIN(DATE) FROM SPACEXTABLE WHERE LANDING_OUTCOME = 'Success (ground pad)'
```

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

Out[12]:

MIN(DATE)
2015-12-22

- Firs Successful Ground Landing Date:
  - 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have success in drone ship and have payload mass greater than 4000 but less than 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 PAYLOAD FROM SPACEXTABLE WHERE LANDING_OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- List the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number FROM SPACEXTABLE GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	total_number
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

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_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * 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

- 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.

```
%sql SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] FROM SPACEXTABLE where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	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.

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_outcomes FROM SPACEXTABLE WHERE DATE between '2010-06-04' and '2017-03-20' group by [Landing_Outcome] order by count_outcomes DESC;
```

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

Landing_Outcome	count_outcomes
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

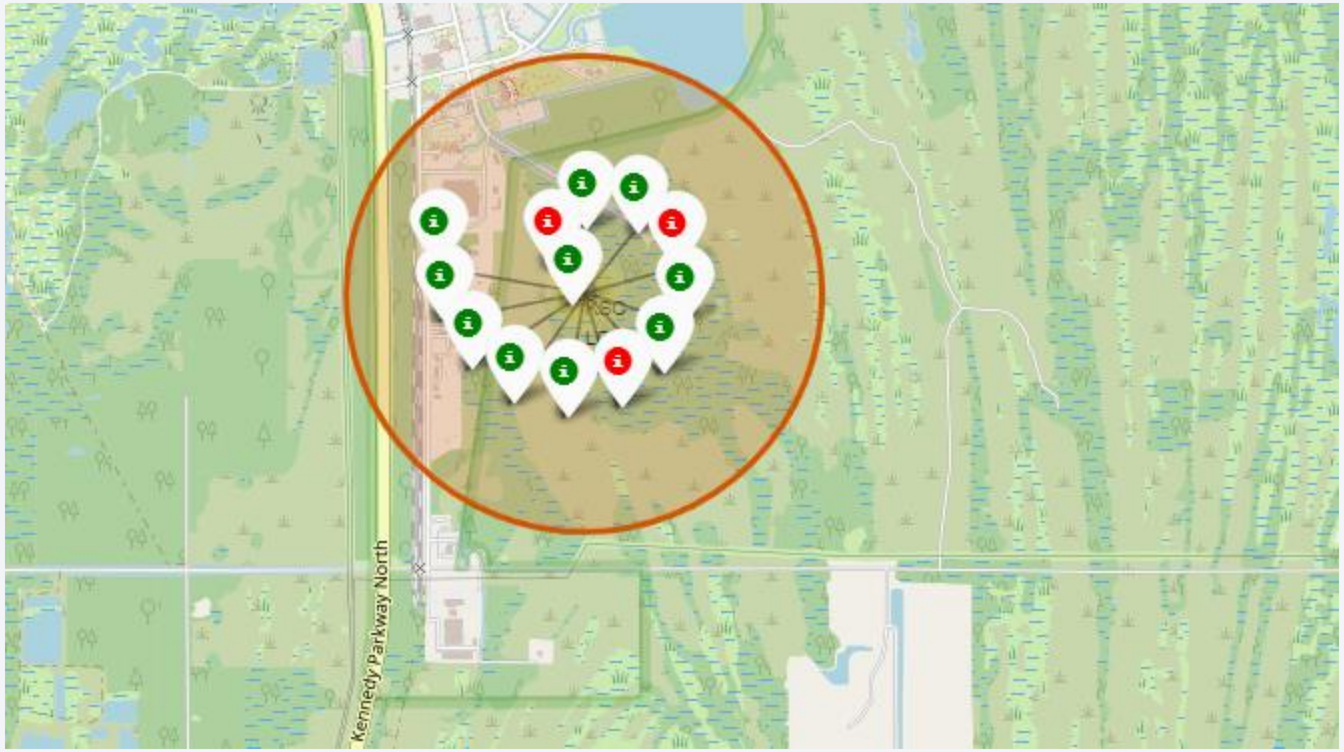
Launch Sites Proximities Analysis

Launch Sites in US

- Launch sites are located near bodies of water to provide a safe area for rocket stages to fall back to Earth
- Launch sites are often located near the Equator for several strategic reasons that enhance the efficiency and effectiveness of space launches.

Launch Outcomes

- KSC LC-39A with 10 success (green) and 3 fails (red) launches



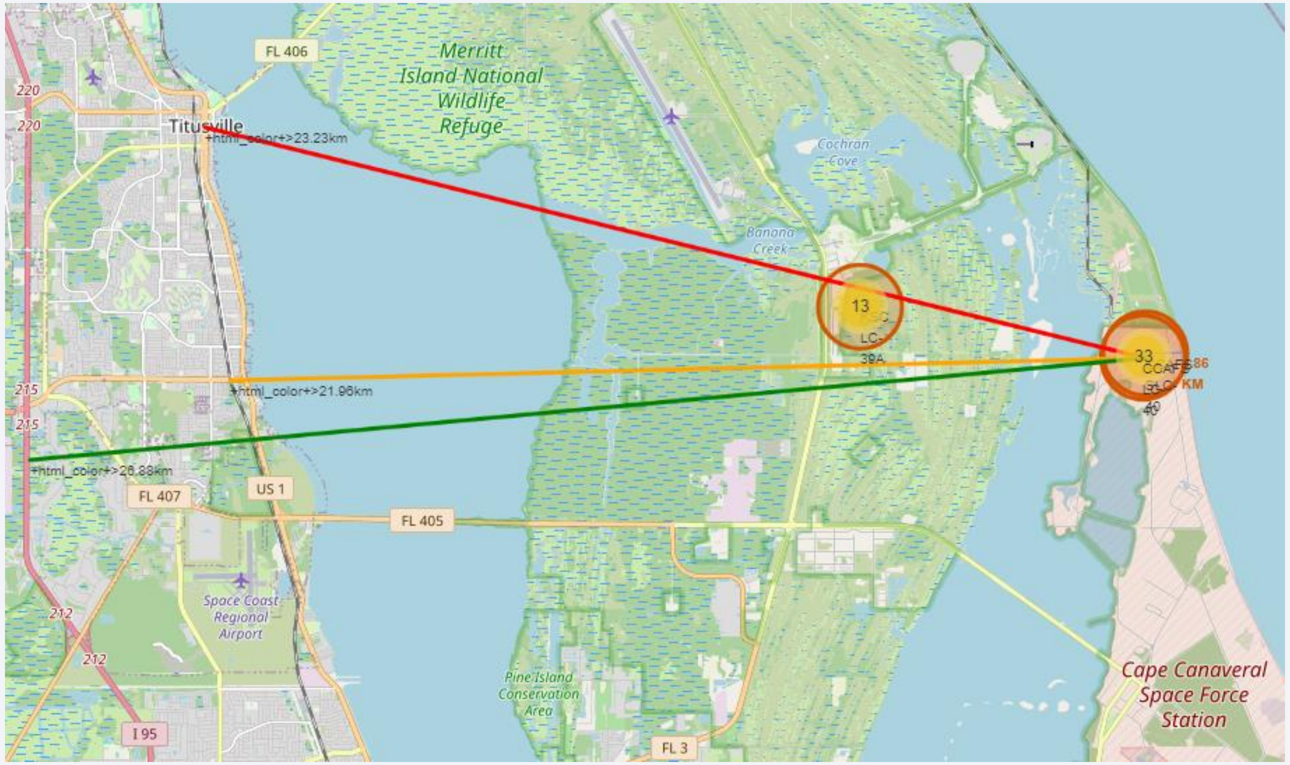
Distance to a closest city, railway, highway,

Launch sites are situated near railways, highways and coastlines to facilitate transport of rocket components, propellants and payloads.

The proximity to the coast allows for the safe discharge of spent rocket stages into the ocean, minimizing risks to inhabited areas.

Distancing launch sites from urban areas reduces the risks associated with launch failures, noise and intense vibrations.

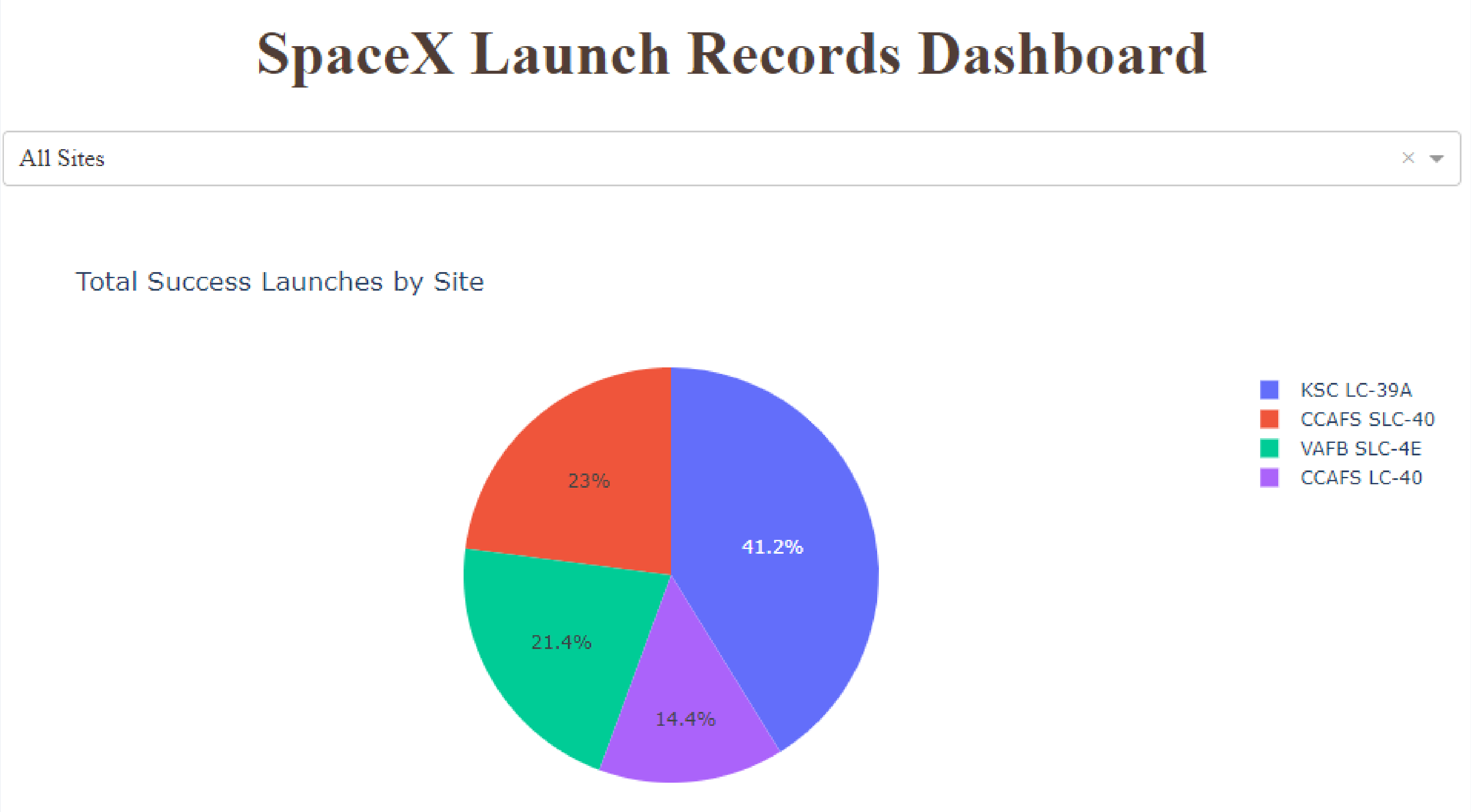
The remote location also contributes to security, access control and compliance with environmental regulations and international agreements, ensuring the success of space missions and minimizing negative impacts.



Section 4

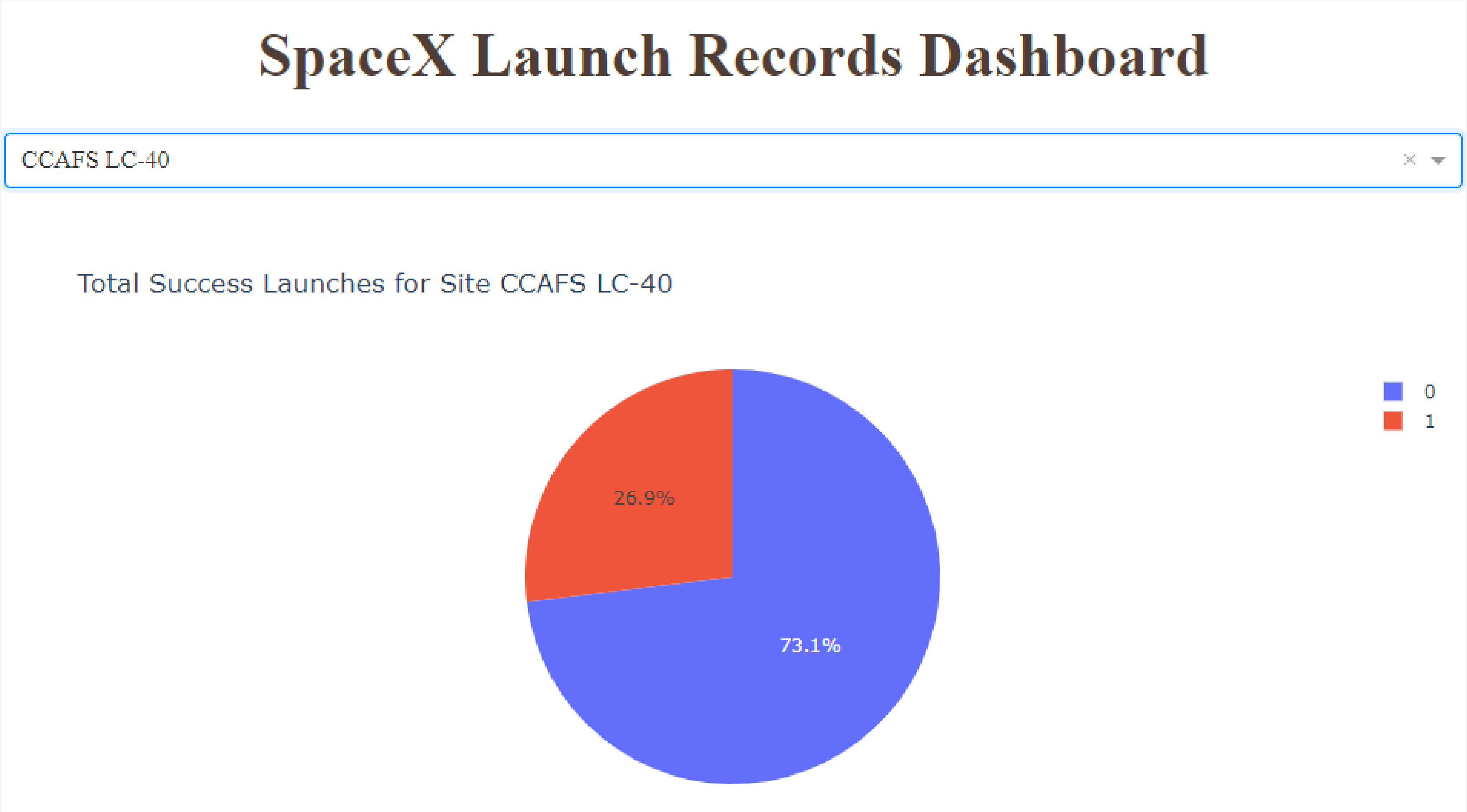
Build a Dashboard with Plotly Dash

Total Success Launches by Site



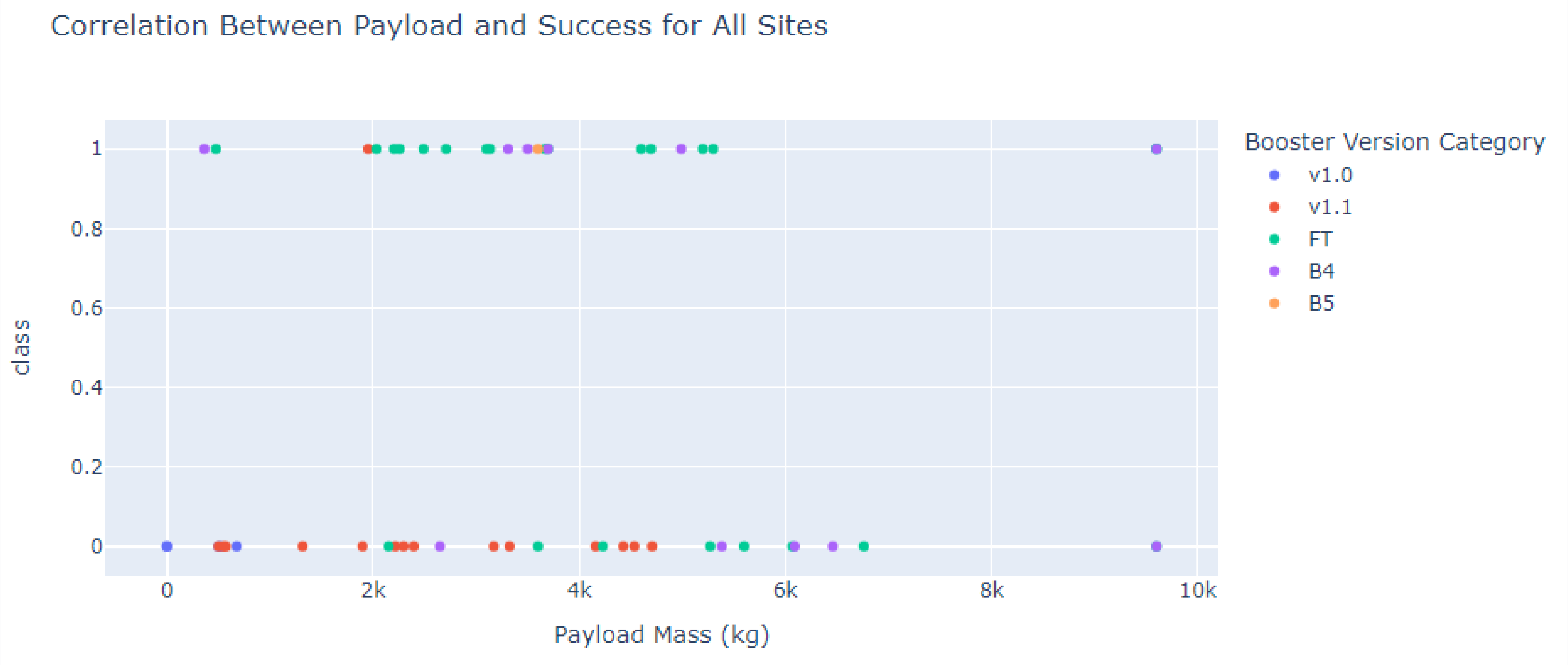
KSC LC-39A é o site com a maior taxa de sucesso de lançamento

Total Success Launches for Site CCAFS LC-40



CCAFS LC-40 has a success rate of over 70%

Correlation Between Payloads and Success for All Sites



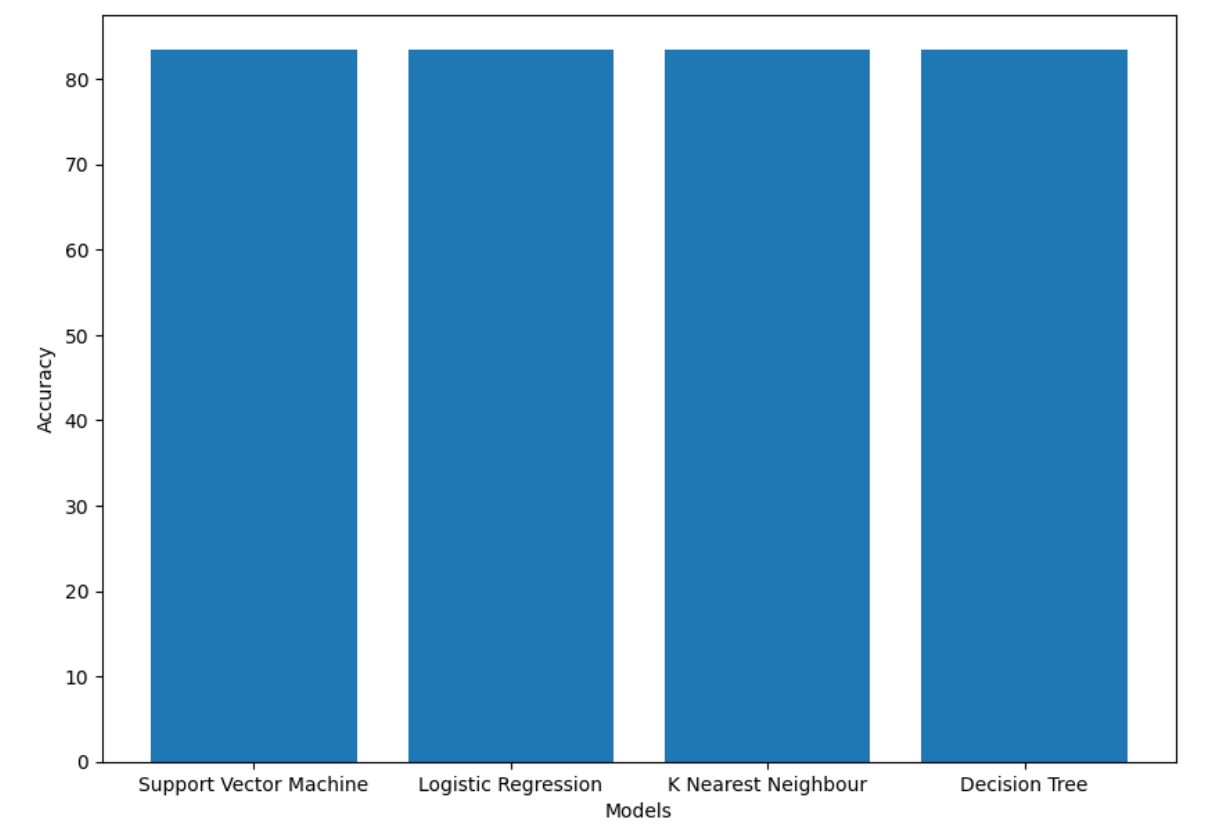
Payloads between 2000 kg and 5000 kg have the highest success rate

Section 5

Predictive Analysis (Classification)

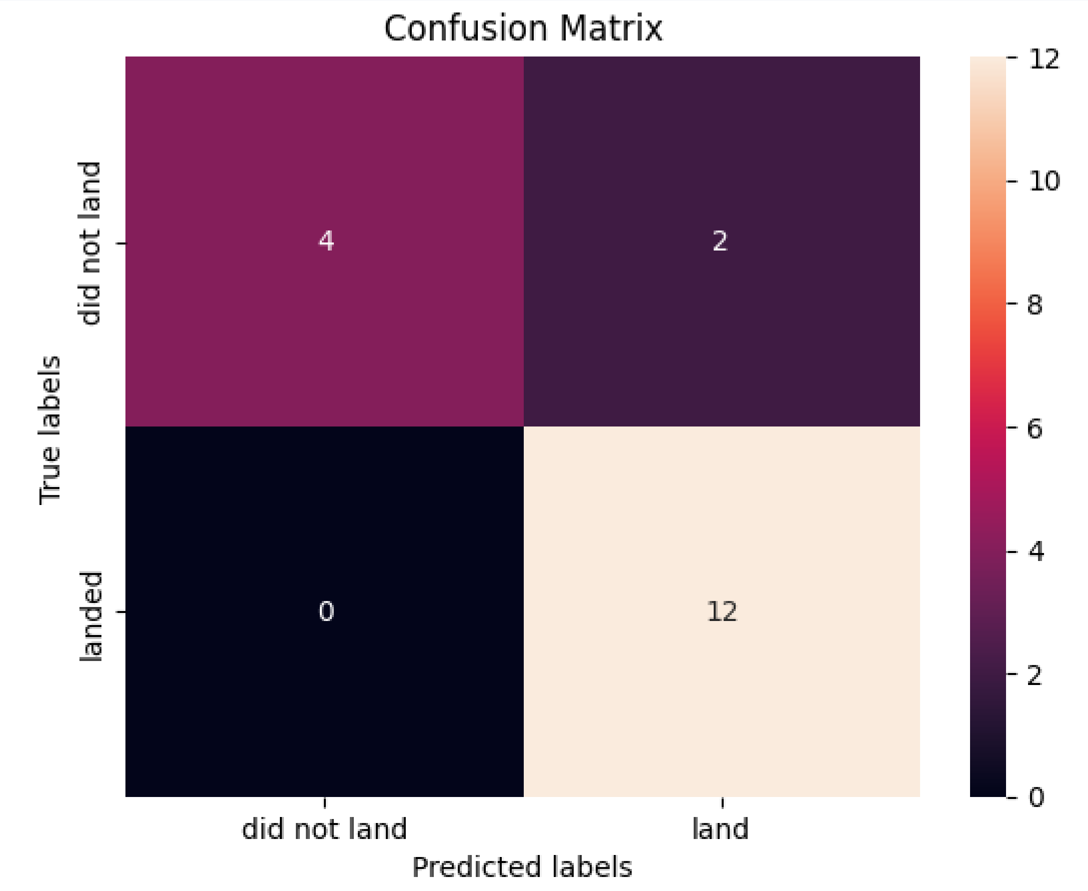
Classification Accuracy

All models performed the same on the test set, but in training the model that had the best accuracy was DecisionTree with 0.86



Confusion Matrix

- With a confusion matrix visualizes and summarizes the performance of a classification algorithm.
  - True Negative: 4
  - False Positive: 2
  - False Negative: 0
  - True Positive: 12



Conclusions

- The success launch rate increased over the time
- Launch sites are located near bodies of water to provide a safe area for rocket stages to fall back to Earth
- KSC LC-39A is the site with the highest launch success rating
- It is possible to predict a successful launch with approximately 84% accuracy

Thank you!