IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nayeli Rocha
07 /November/ 2023

Outline

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

Executive Summary

Space Y a new rocket company would like to compete with SpaceX , which launches its Falcon 9 rocket for $62 million. Space Y has hired a data scientist to identify the factors for a successful rocket landing, compute the rate of successful landings over time and built a model to predict whether Space X Falcon 9 first stage will land successfully. To achieve the objectives, the data scientist follow the next methodology:

1. ***Data collection***. SpaceX launch data was gathered from an API, requesting the data specifically from SpaceX REST API by the get request () method, and the Falcon 9 Launch data Wiki pages, using web-scrapping methodology
2. ***Wrangling data:*** .once with the raw data from the previous step; throughout wrangling the raw data using an API, sampling data, and dealing with Nulls the raw dataset was transformed into a clean dataset which provides meaningful data such as create a Boolean success/fail landing variable.
3. ***Explore the processed data,*** throughout:
    - EDA. Considering the following features: PayloadMass, LaunchSite, FlightNumber, Orbit Type, and yearly trend, success rate.

Executive Summary

- SQL skills, calculating the total payload range for successful rates, and total number of successful and failed outcomes.

4. ***Basic statistical analysis, data visualization and Dashboards.*** To see how variables might be related to each other and launch sites proximity to geographical markers

5. ***Build, evaluate, and refine predictive models,*** splitting the data into categorical variables , training data and test data to predict landing outcomes finding the best hyperparameters for SVM, classification Tree, Logistic regression and K-nearest neighbor.

Results:

- For launch Sites CCAFS SLC 40 & KSC LC 39A, as the FlightNumber increases, the first stage is more likely to land successfully.

- Launch success has improved over time.

- CCA FSL LC-40 and CCAFS SLC-40 are near the equator, and all sites are located near the coast,

- Launches with a payloadmass range between 2000kg and 5000 kg tended to reach a success landing

- All models performed similar on the test-set, however, classification tree slightly outperformed.

Introduction

Today, companies are trying to make space travel affordable for everyone. One of the most successful is SpaceX, which announces on its website launches of the Falcon 9 rocket for $62 million; other companies cost upwards of 165 million dollars each. Much of the savings is because SpaceX can reuse the first stage. Therefore, by knowing whether the first stage will land, it is possible to determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- A new rocket company, called SPACE Y, would like to compete with SpaceX.

Introduction

As a data scientist working for SPACE Y, it is important to identify and determine the factors for a successful rocket landing price of each launch. Therefore, using the Space X data, the following objectives should be covered in this capstone project:

1. Which factors or features are correlated with a landing success.
2. Rate of successful landings over time
3. Predict whether Space X Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The SpaceX launch data was gathered from:
    - A SpaceX API, throughout a get request to the SpaceX REST API:
    - The Falcon 9 Launch data Wiki pages, throughout the Web Scrapping Method:
- Perform data wrangling
  - To provide meaningful data, the data was cleaned performing: Wrangling Data using an API, Sampling Data, and Dealing with Nulls.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - To predict if the Falcon9 first stage will land successfully.

Methodology

Executive Summary

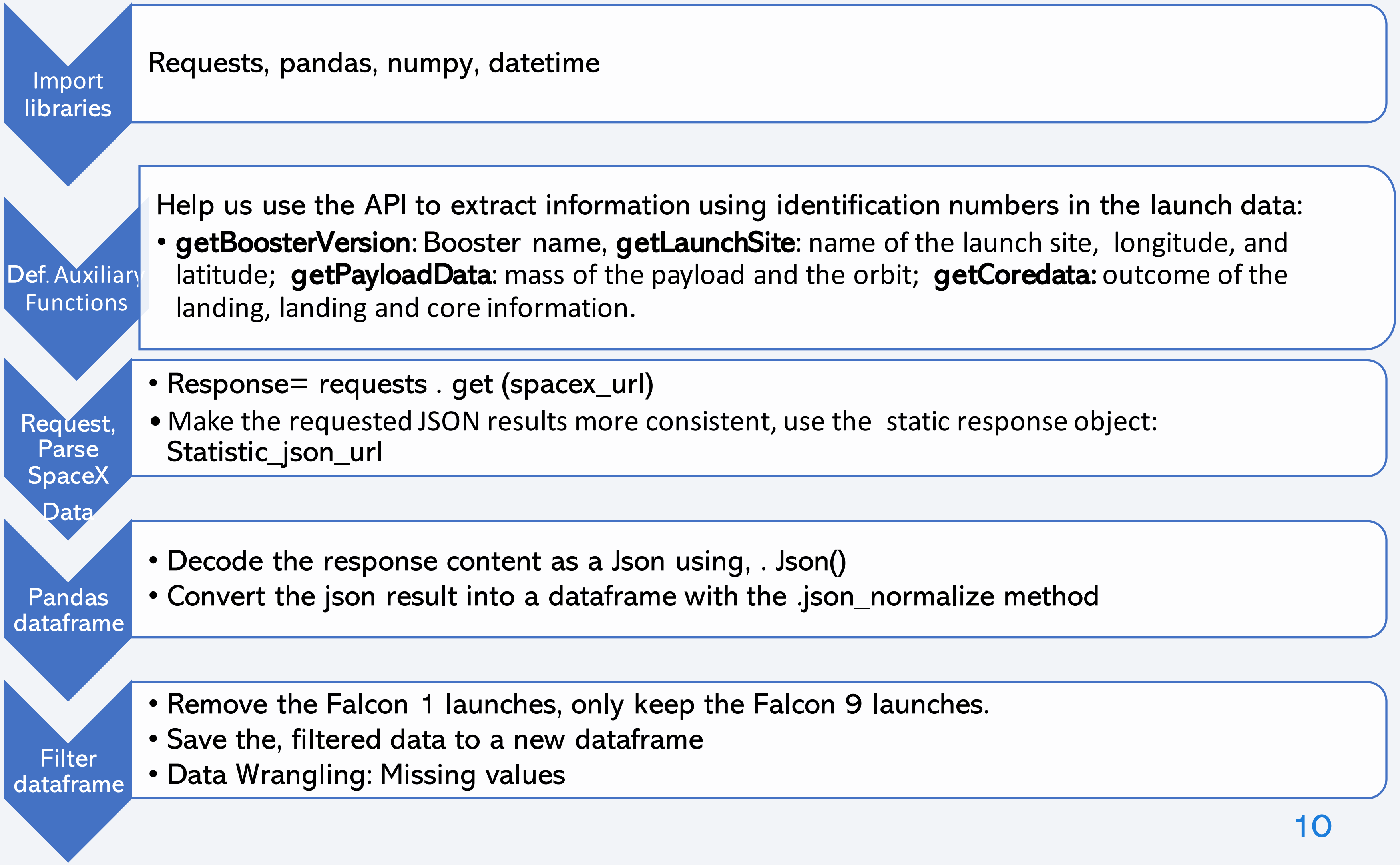
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Classification Models were built to predict landing outcomes, then those models were evaluated to find the best parameters and identify the best models according accuracy and test_accuracy.

Data Collection - SpaceX API

Data was collected from the SpaceX REST API.

It was performed a get request to obtain the launch data.

The data collection process is presented in the Flowchart



Import libraries

Requests, pandas, numpy, datetime

Def. Auxiliary Functions

Help us use the API to extract information using identification numbers in the launch data:

- **getBoosterVersion**: Booster name, **getLaunchSite**: name of the launch site, longitude, and latitude; **getPayloadData**: mass of the payload and the orbit; **getCoredata:** outcome of the landing, landing and core information.

Request, Parse SpaceX Data

- Response= requests . get (spacex_url)
- Make the requested JSON results more consistent, use the static response object: Statistic_json_url

Pandas dataframe

- Decode the response content as a Json using, . Json()
- Convert the json result into a dataframe with the .json_normalize method

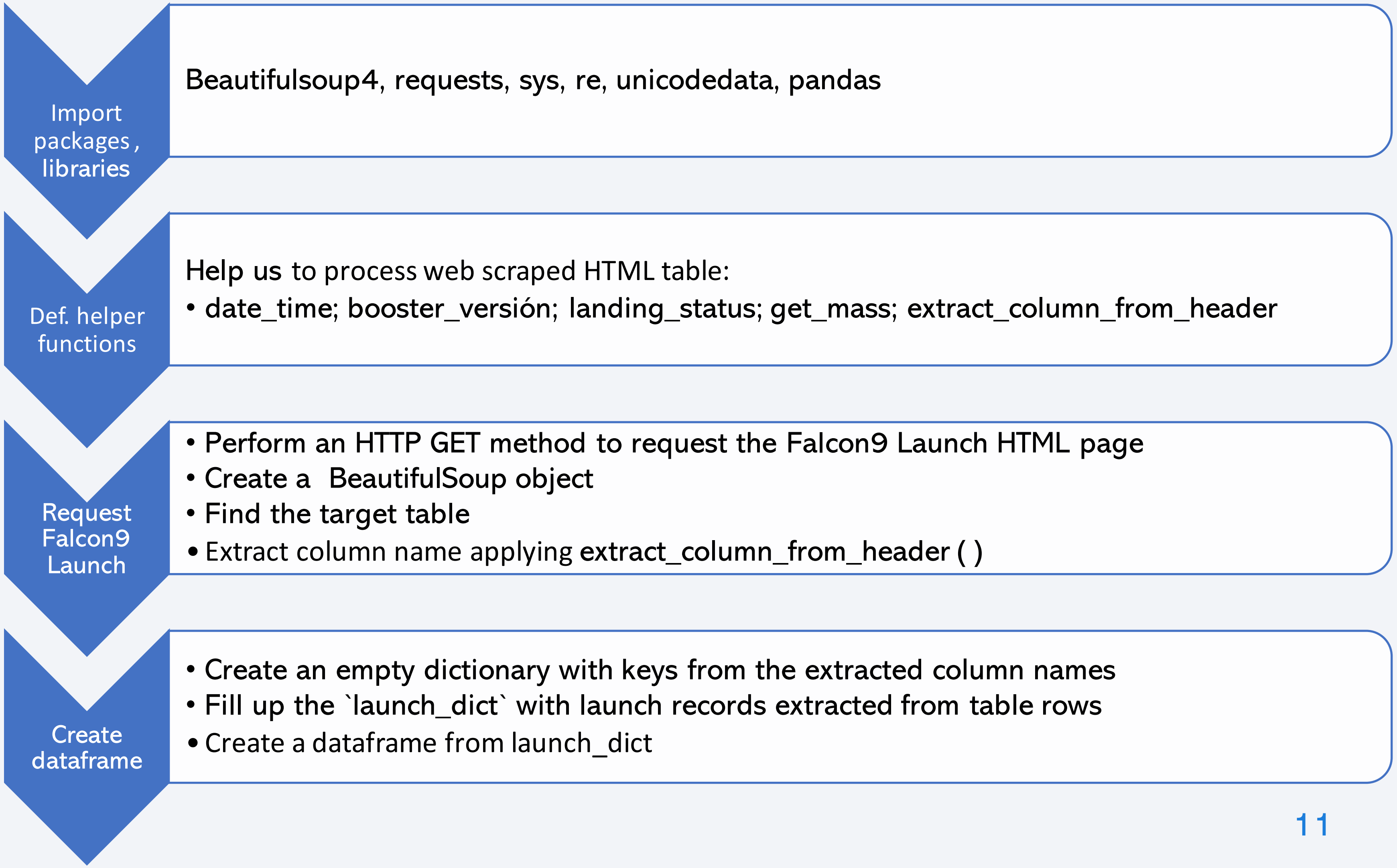
Filter dataframe

- Remove the Falcon 1 launches, only keep the Falcon 9 launches.
- Save the, filtered data to a new dataframe
- Data Wrangling: Missing values

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/data-collection-API.ipynb

Data Collection - Scrapping

- It was used Python BeautifulSoup package to web scrape HTML tables that contain Falcon 9 launch records.
- The data collection process is presented in the Flowchart



Import packages , libraries

Beautifulsoup4, requests, sys, re, unicodedata, pandas

Def. helper functions

Help us to process web scraped HTML table:
- date_time; booster_versión; landing_status; get_mass; extract_column_from_header

Request Falcon9 Launch

- Perform an HTTP GET method to request the Falcon9 Launch HTML page
- Create a BeautifulSoup object
- Find the target table
- Extract column name applying **extract_column_from_header ( )**

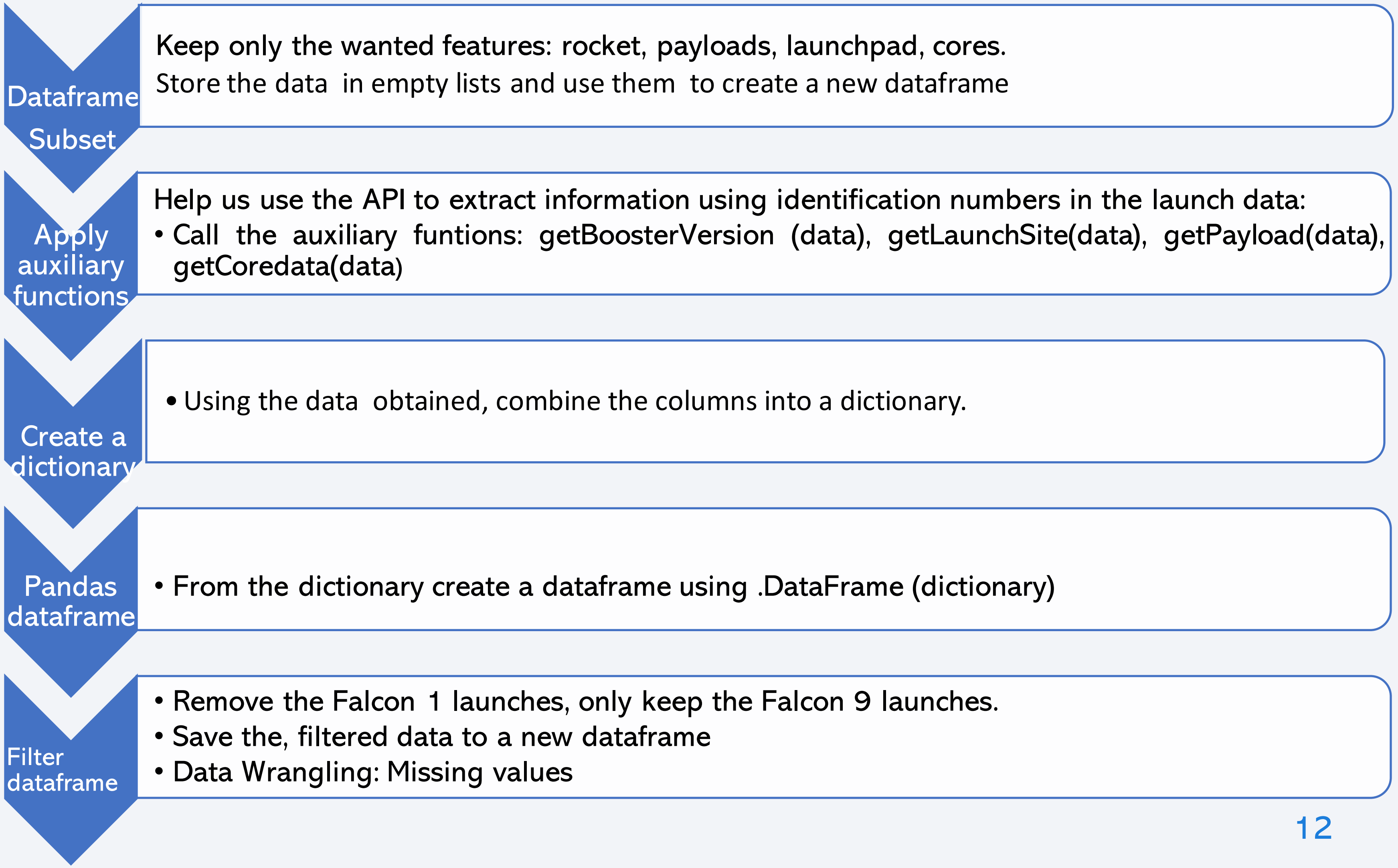
Create dataframe

- Create an empty dictionary with keys from the extracted column names
- Fill up the `launch_dict` with launch records extracted from table rows
- Create a dataframe from launch_dict

GitHub URL : https://github.com/Nayero/Space-X-Capstone-Project/blob/main/Data%20collection%20webscraping.ipynb

Data Wrangling - using an API

- Some columns, have an ID number, not actual data.
- Therefore, the API target another endpoint to gather specific data for each ID number



Dataframe Subset

Keep only the wanted features: rocket, payloads, launchpad, cores.
Store the data in empty lists and use them to create a new dataframe

Apply auxiliary functions

Help us use the API to extract information using identification numbers in the launch data:

- Call the auxiliary funtions: getBoosterVersion (data), getLaunchSite(data), getPayload(data), getCoredata(data)

Create a dictionary

- Using the data obtained, combine the columns into a dictionary.

Pandas dataframe

- From the dictionary create a dataframe using .DataFrame (dictionary)

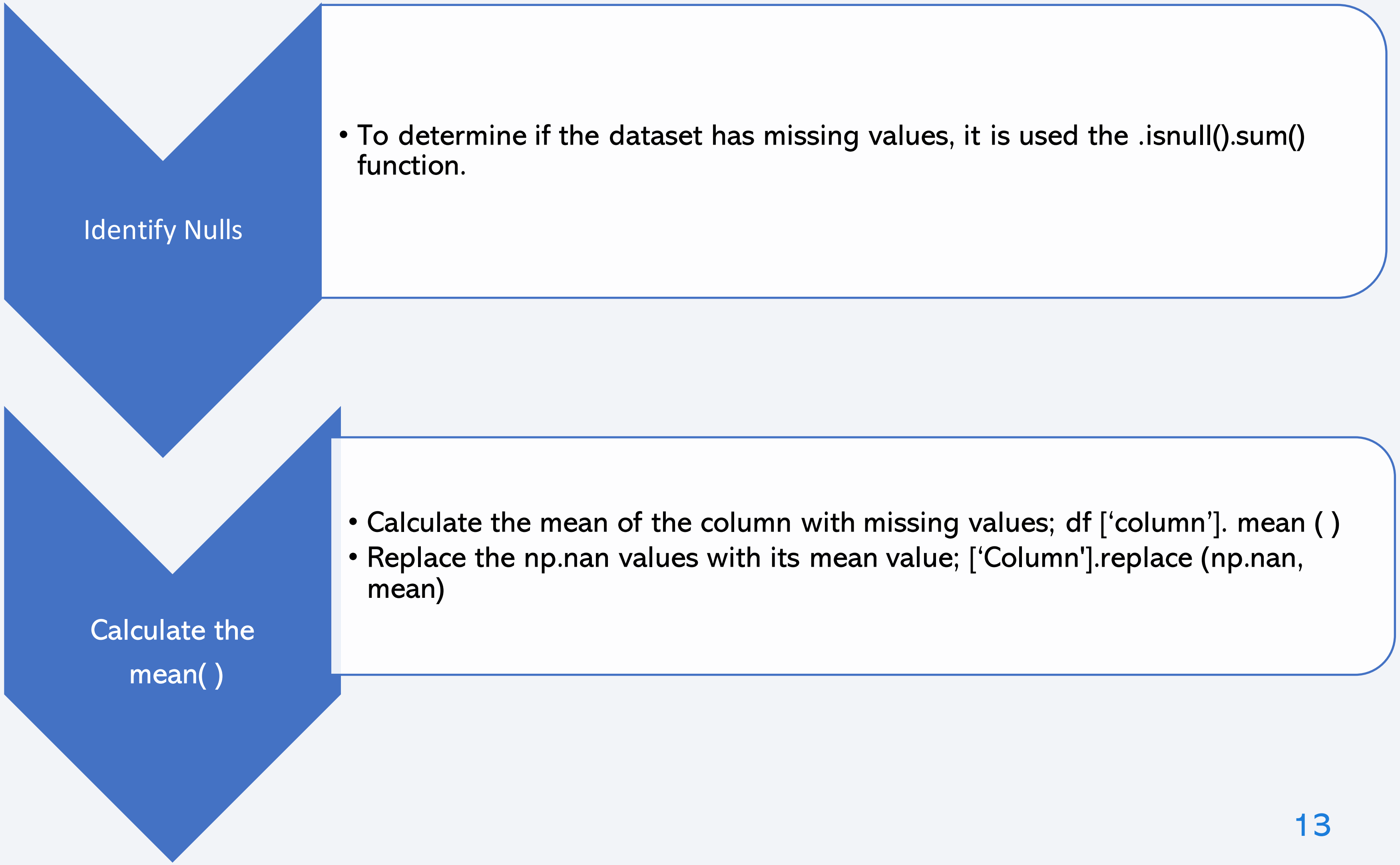
Filter dataframe

- Remove the Falcon 1 launches, only keep the Falcon 9 launches.
- Save the, filtered data to a new dataframe
- Data Wrangling: Missing values

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/data-collection-API.ipynb

Data Wrangling - Dealing with Nulls

- Some columns, have missing values. To deal with missing values follow the flowchart.



Identify Nulls

- To determine if the dataset has missing values, it is used the .isnull().sum() function.

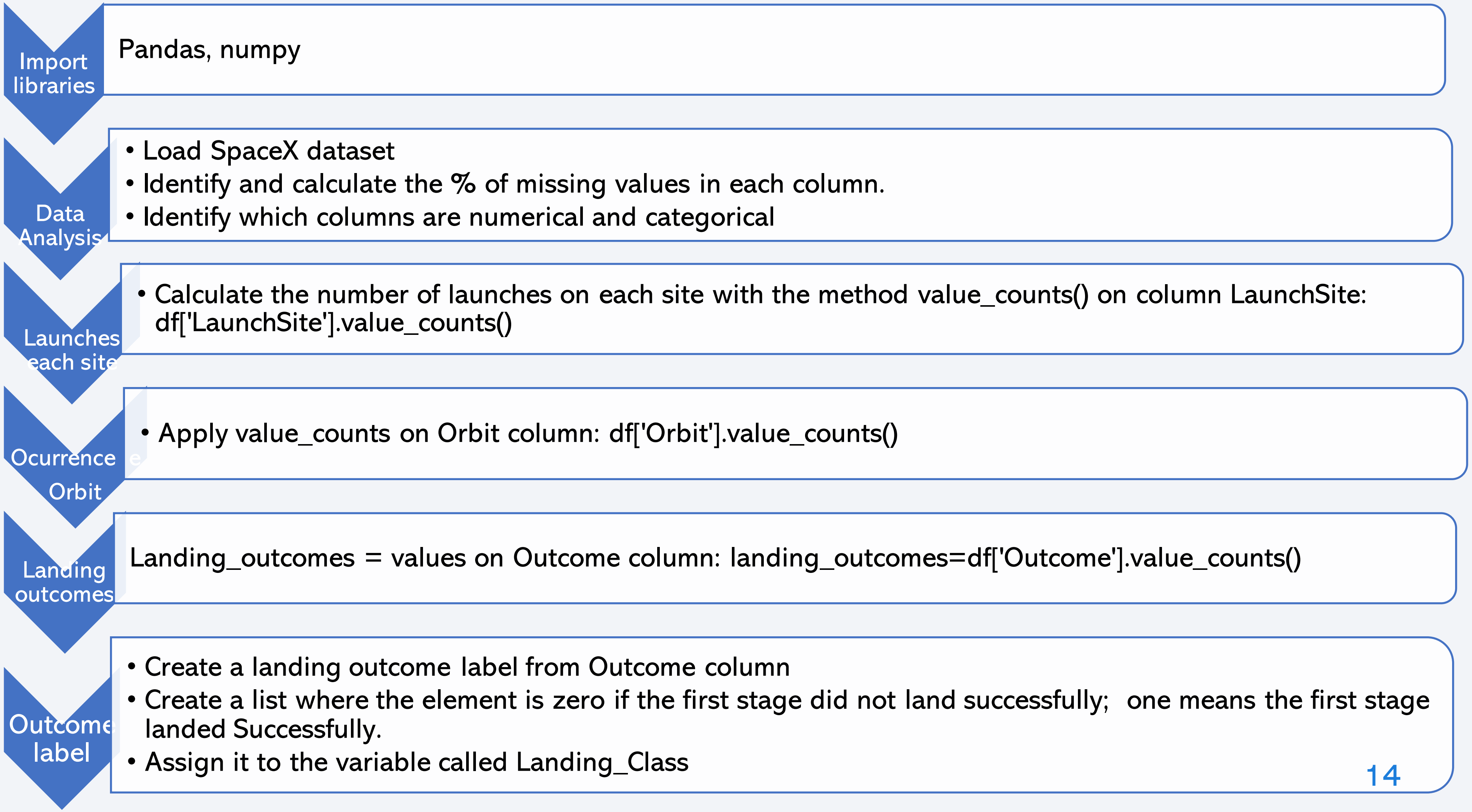
Calculate the mean( )

- Calculate the mean of the column with missing values; df [‘column’]. mean ( )
- Replace the np.nan values with its mean value; [‘Column'].replace (np.nan, mean)

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/data-collection-API.ipynb

Data Wrangling

The collected data, once was analyzed and summarized features, it was improved by creating a landing outcome label



Import libraries

Pandas, numpy

Data Analysis

- Load SpaceX dataset
- Identify and calculate the % of missing values in each column.
- Identify which columns are numerical and categorical

Launches each site

- Calculate the number of launches on each site with the method value_counts() on column LaunchSite: df['LaunchSite'].value_counts()

Ocurrence Orbit

- Apply value_counts on Orbit column: df['Orbit'].value_counts()

Landing outcomes

Landing_outcomes = values on Outcome column: landing_outcomes=df['Outcome'].value_counts()

Outcome label

- Create a landing outcome label from Outcome column
- Create a list where the element is zero if the first stage did not land successfully; one means the first stage landed Successfully.
- Assign it to the variable called Landing_Class

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/Data%20wrangling.ipynb

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

Chart Plotted	Features plotted	Reason
Scatter plot	FlightNumber vs PayloadMass	To visualize how those features would affect the launch outcome
Scatter plot	FlightNumber vs. FlightNumber	Visualize the relationship between features
Scatter plot	PayloadMass vs. FlightNumber	Visualize if there is any relationship between features
Bar chart	Success rate of each orbit type	Visualize if there are any relationship between success rate of each orbit type
Scatter plot	FlightNumber vs. Orbit type	To see if there is any relationship between FlightNumber and Orbit type.
Scatter plot	Payload vs. Orbit type	Visualize the relationship between Payload and Orbit type
Line plot	Year vs. average launch success	Visualize the launch success yearly trend

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/EDA-datavisualization.ipynb

EDA with SQL

1. Load the Spacex dataset and save it as .cvs file.
2. Load the SQL extension and establish a connection with the database.
3. Summarize the SQL queries you performed.

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

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

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

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

   3.5 List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

   3.6 List the total number of successful and failure mission outcomes

   3.7 List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

   3.8 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.

   3.9 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.

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/EDA-sql.ipynb

Build an Interactive Map with Folium

1. The launch Sites were marked with circles on the its specific coordinates (Latitude and Longitude) and a popup label:

  - Blue Circles at Nasa Johnson Space Center
  - Orange Circles at all launch sites.

- 2. Launch Outcomes were marked at each launch site, to indicate which launch site have high success rates:

  - Green colored cluster marker to indicate successful launches
  - Red Colored cluster marker to indicate unsuccessful launches

3. The distances between the CCAFS SLC-40 Launch Site and its proximities were marked with blue Polylines

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/Launch_site_location_Folium.ipynb

Build a Dashboard with Plotly Dash

Plots/Graphs /Charts

- Pie plot ´to visualize the share of successful launches of each launch site, to illustrate the numerical proportion of successful and unsuccessful launches per launch site
- Scatter plot to identify if there is any correlation between PayloadMass vs. Success Rate by Booster version

Interactions

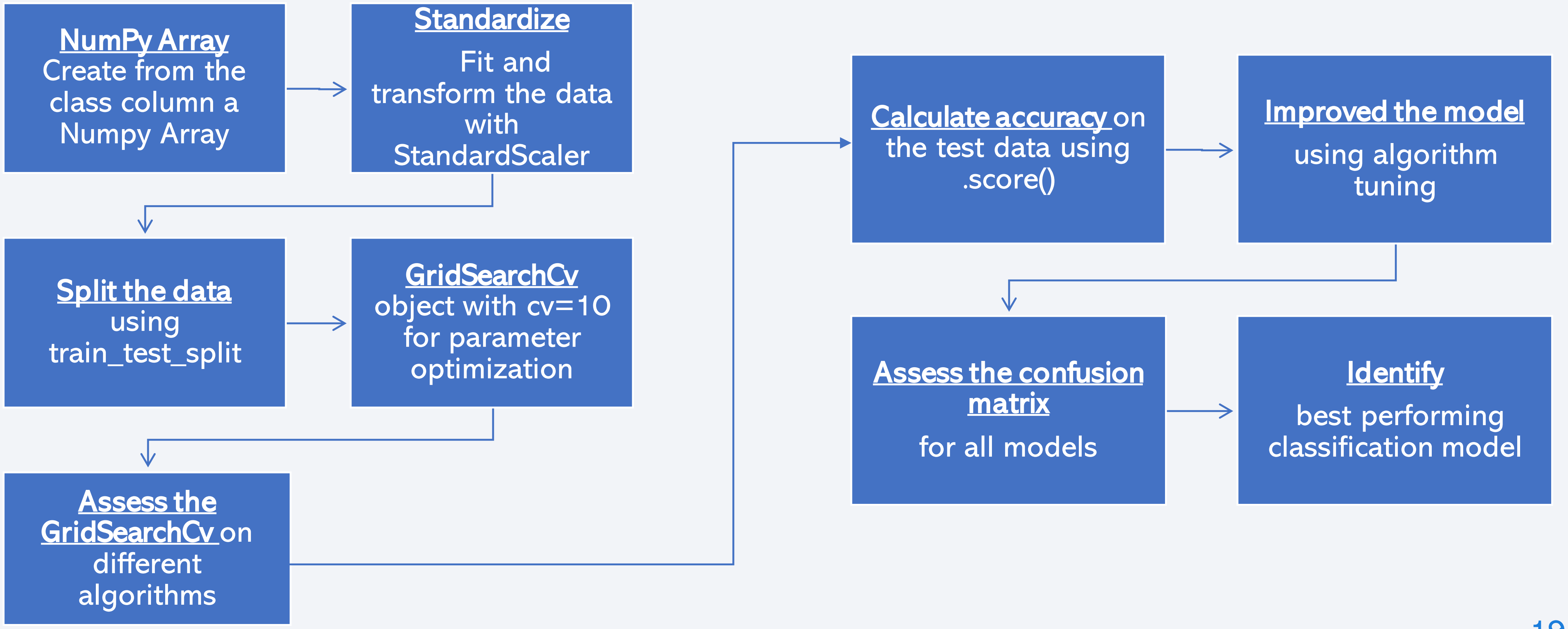
- Dropdown list with Launch sites, allowing users to select either all launches at once or a unique launch site.
- Slider of Payload Mass Range, allowing users to select a certain Payload Mass range to apply on the scatterplot.

GitHub URL:

Predictive Analysis (Classification)

Build the Model

Evaluate and Improve the Model



NumPy Array Create from the class column a Numpy Array

Standardize Fit and transform the data with StandardScaler

Split the data using train_test_split

GridSearchCv object with cv=10 for parameter optimization

Assess the GridSearchCv on different algorithms

Calculate accuracy on the test data using .score()

Improved the model using algorithm tuning

Assess the confusion matrix for all models

Identify best performing classification model

GitHub URL: https://github.com/Nayero/Space-X-Capstone-Project/blob/main/SpaceX_Machine_Learning_Prediction.ipynb

Results

Exploratory data analysis results

- Launch success rate has increased overtime.
- As the FlightNumber increases, the first stage is more likely to land successfully
- In overall, KSC LC-39A has the highest success rate among landing sites. Moreover; it recorded 100% success rate for launches with PayloadMass less than aprox 5,500Kg:
- Most launches sites with a PayloadMass 7000Kg recorded successful landings in first stage.
- Orbits ES-L1, GEO, HEO, and SSO (SO) have 100% success rate: The success rate for each orbit tend to be increase as the FlightsNumber increases. However, for GTO orbit is hard to follow a trend.

Interactive analytics demo in screenshots

- KSC LC 39 -A Launch site had the highest launch success ratio (76.9%)
- Launch sites do not represent a risk for humans due to are close to the coast and not close to highways, cities and railways
- Payloads between 2000 Kg and 5000 kg shown the highest success rate.

Predictive analysis results

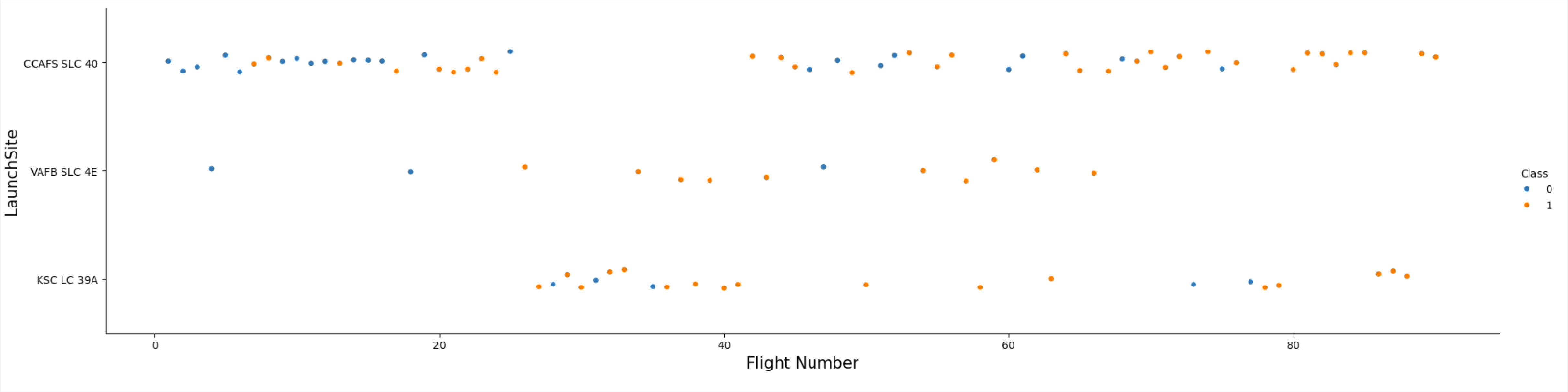
Decision tree model is the best predictive model for the dataset.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

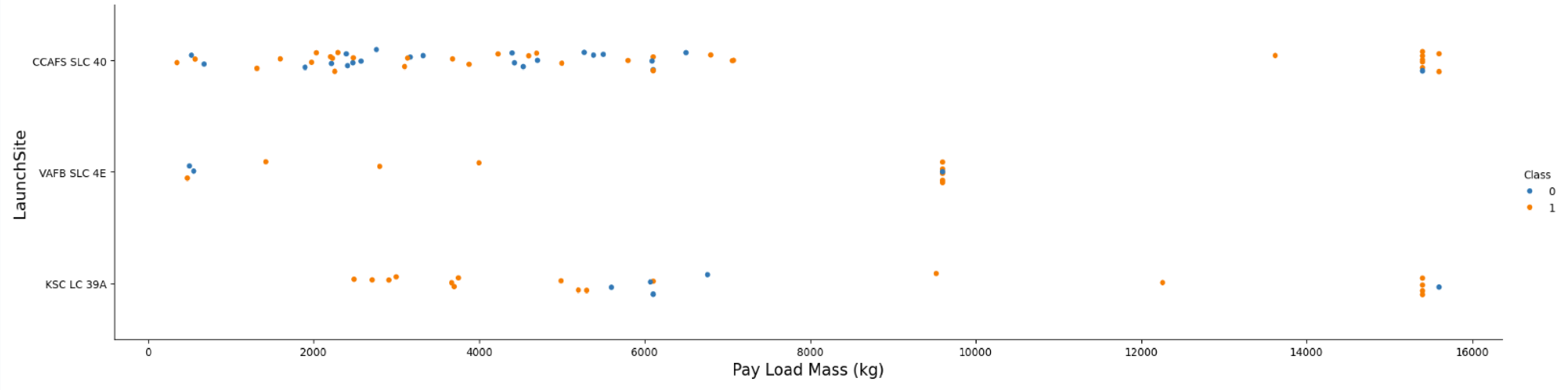
Show a scatter plot of Flight Number vs. Launch Site



- For launchSites CCAFS SLC 40 & KSC LC 39A, as the flight number increases, the first stage is more likely to land successfully.
- On the other hand, when FlightNumbers are lower than 20: CCAFS SLC 40 tended to fail, and there are no launches from KSC LC.
- For the VAFB-SLC launchSite there are no rockets launches for FlightNumbers greater than 70.

Payload vs. Launch Site

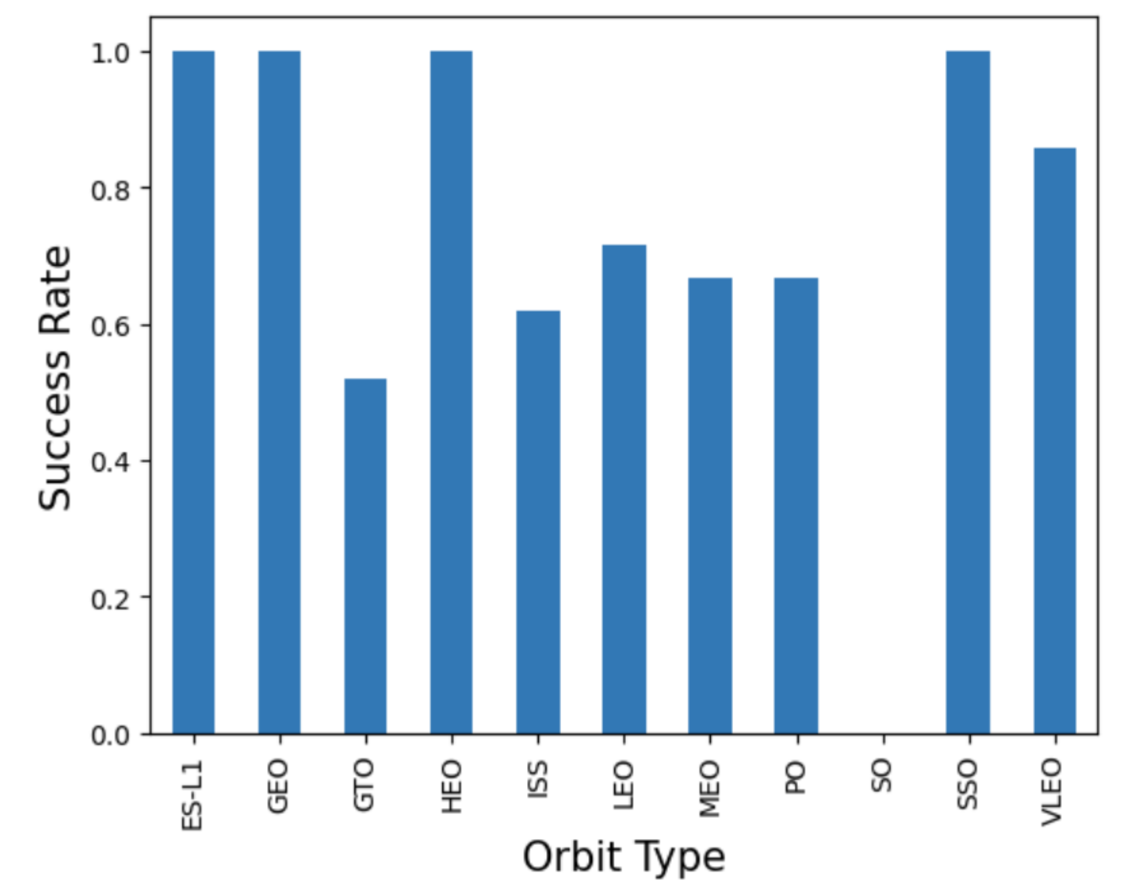
Show a scatter plot of Payload vs. Launch Site



- In the VAFB-SLC launchsite, there are no rockets launched for PayloadMass greater than 10000.
- Payloads that approach MAX(Payload) tended to launch successfully from CCAFS SLC 40 & KSC LC 39A
- Payloads less than 8000 kg tended to fail at a higher rate when launched from CCAFS SLC 40
- KSC LC 39A has a 100% success rate for launches less than aprox 5500Kg

Success Rate vs. Orbit Type

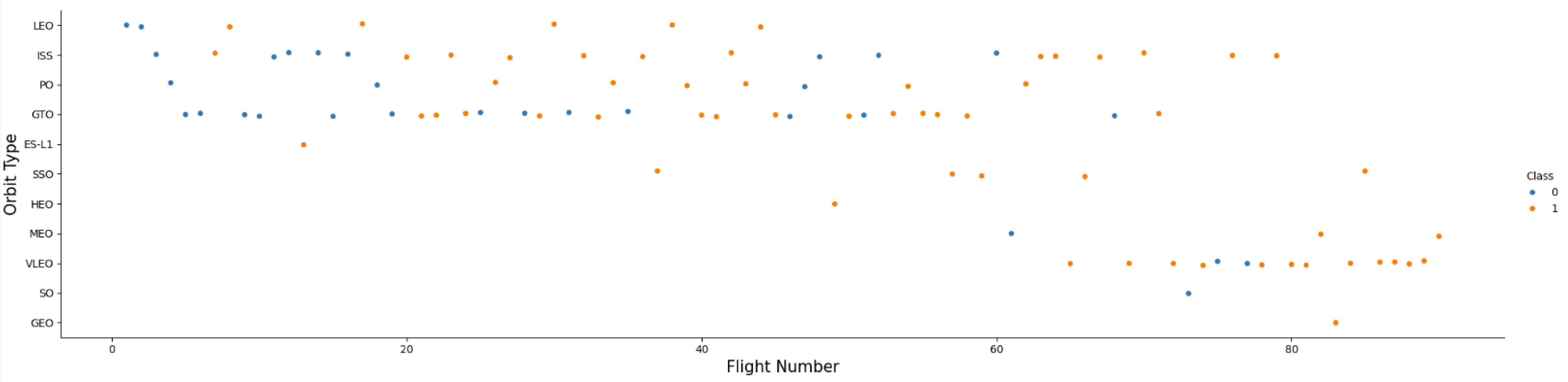
Show a bar chart for the success rate of each orbit type



- The orbits with higher success rate are ES-L1, GEO, HEO, SSO or SO.
- GTO is the orbit with lower success rate.

Flight Number vs. Orbit Type

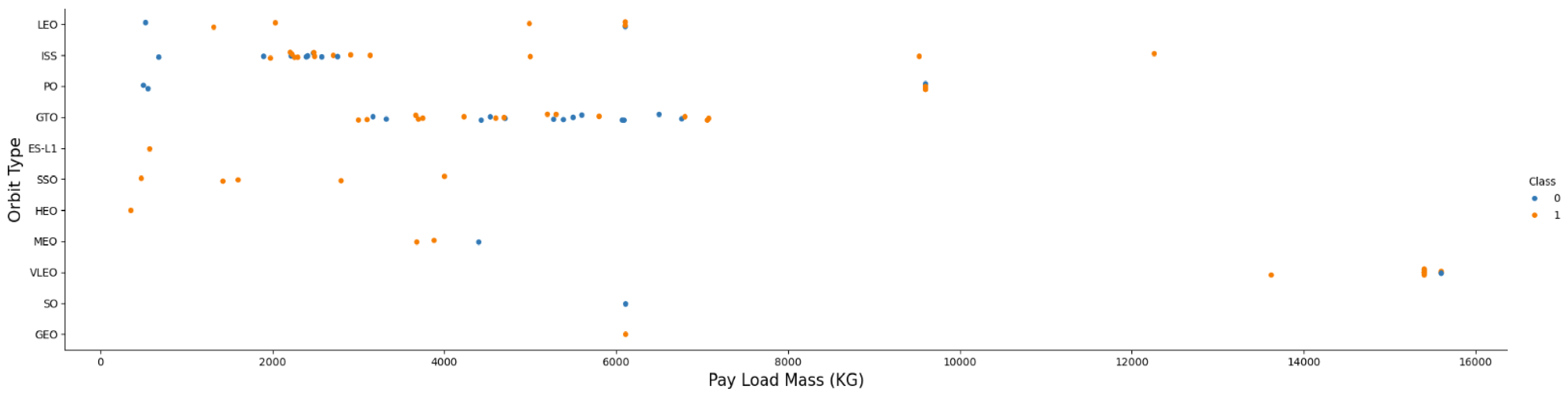
Show a scatter point of Flight number vs. Orbit type



- For most orbits, as the FlightNumber increases, the landing success rate for each orbit also increases. This is evident in the Leo Orbit. Not the case for GTO orbit, which seems to not follow this trend.

Payload vs. Orbit Type

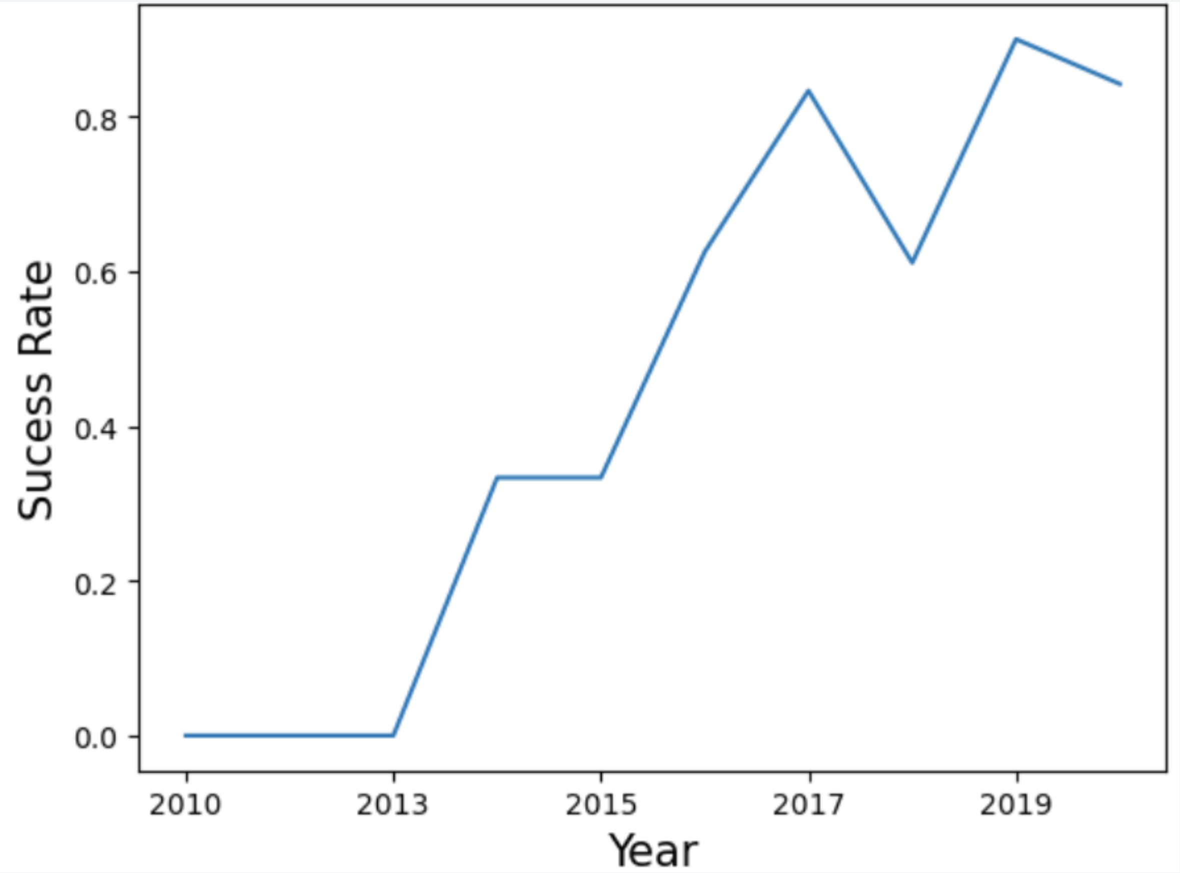
Show a scatter point of payload vs. orbit type



- With heavy payloads the successful landing or positive landing rate is higher for Polar, LEO and ISS orbits
- However, for GTO cannot distinguish this well, as both positive landing rate and negative landing (unsuccessful mission) are both there.

Launch Success Yearly Trend

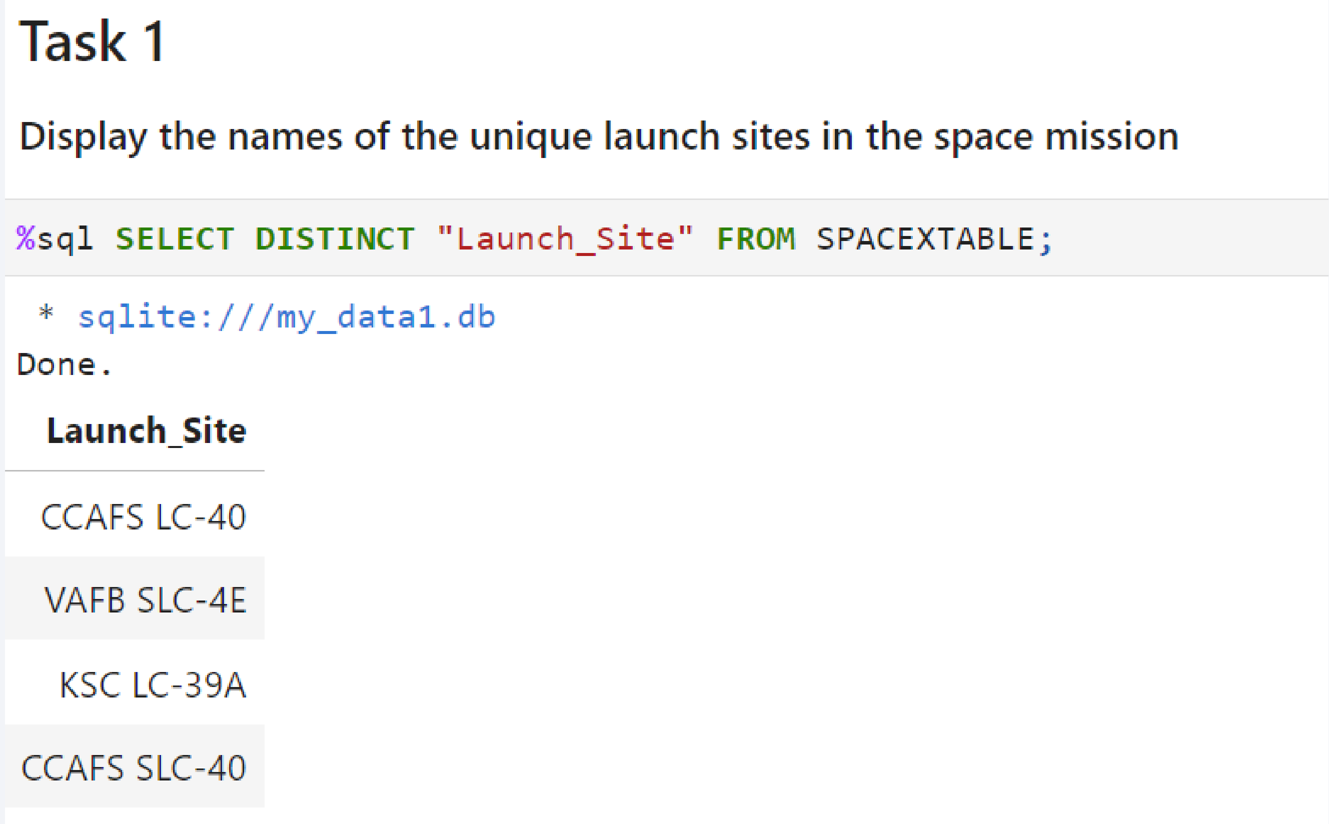
Show a line chart of yearly average success rate



Although the success rate from 2017 to 2018 decreased. Since 2013, the landing success rate kept increasing till 2020.

All Launch Site Names

Find the names of the unique launch sites



Task 1

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

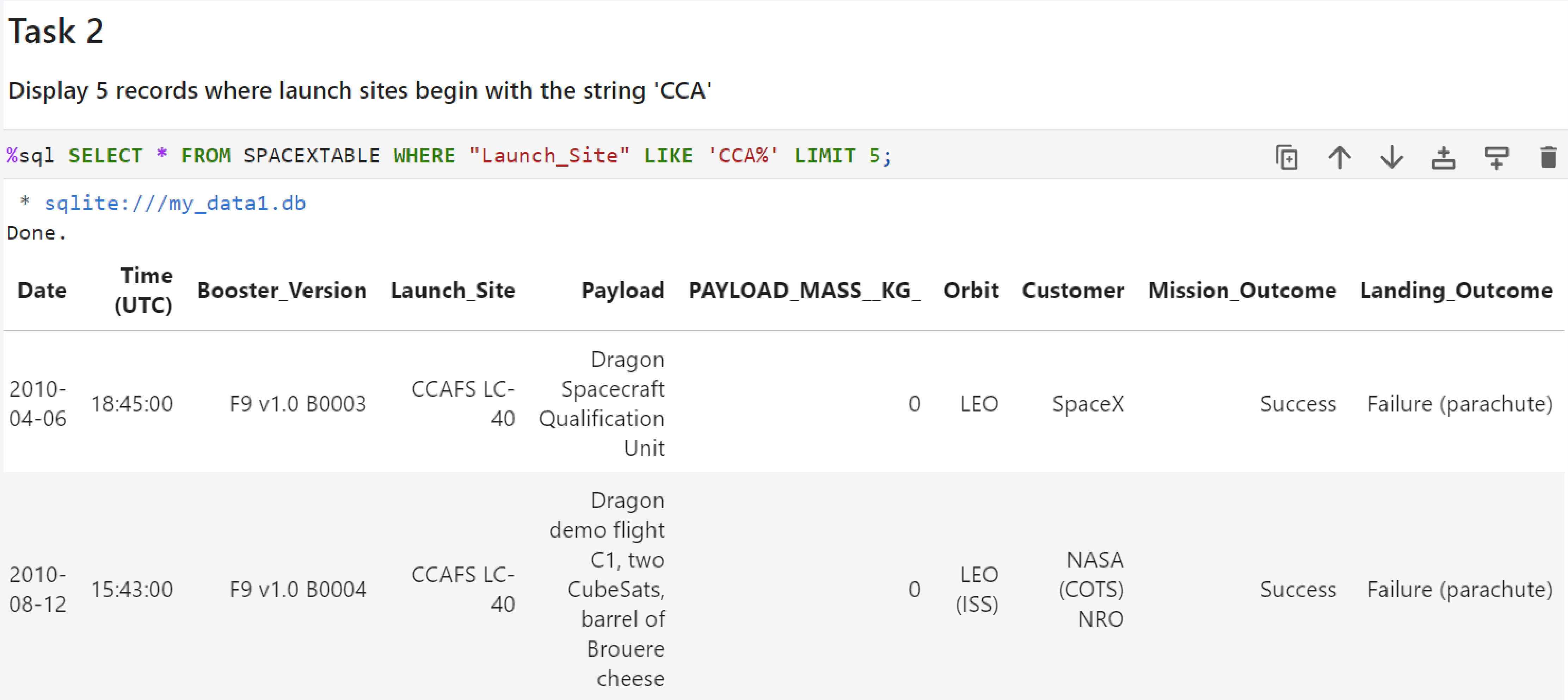
```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

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

The query was made using the Distinct function to obtain the unique launch sites name

Launch Site Names Begin with 'CCA'



Task 2

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

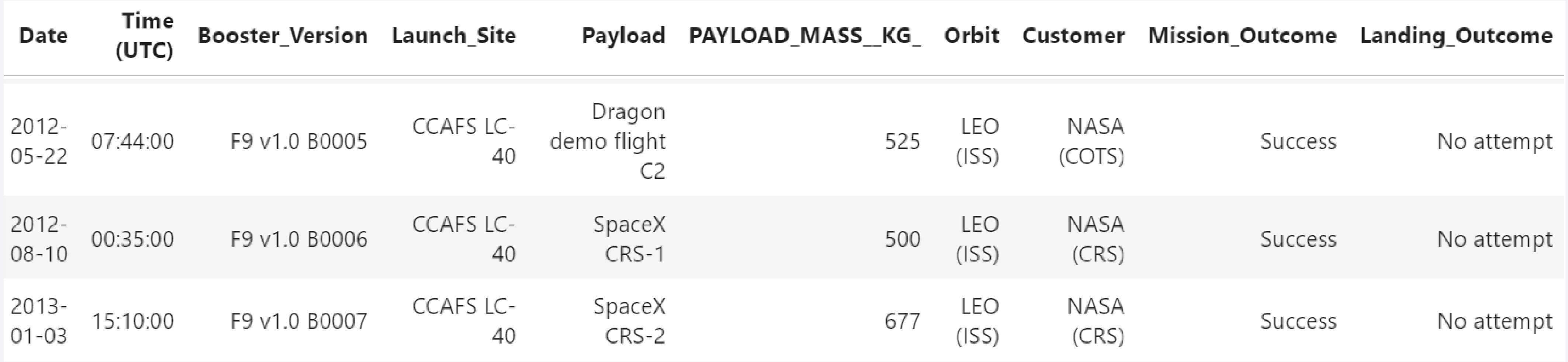
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	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)

Launch Site Names Begin with 'CCA'

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

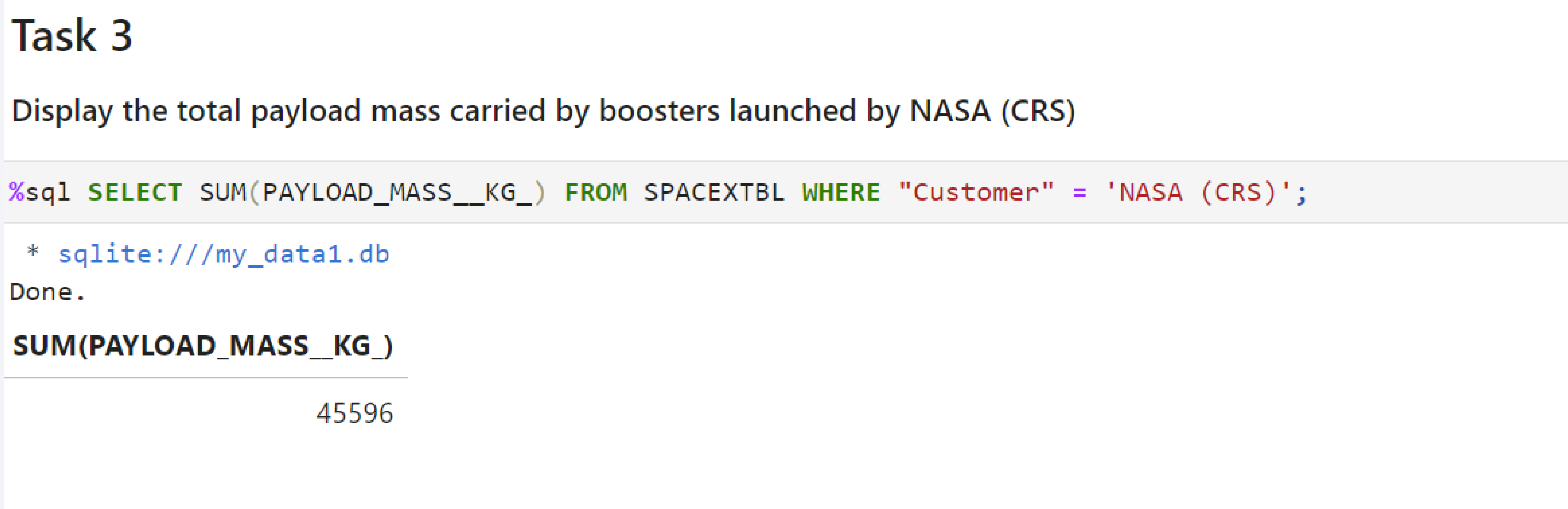


Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	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

In the query we used the like function to filter the launch sites names that start with CCA and limit the results to the first 5 rows.

Total Payload Mass

Calculate the total payload carried by boosters from NASA



Task 3

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)';
```

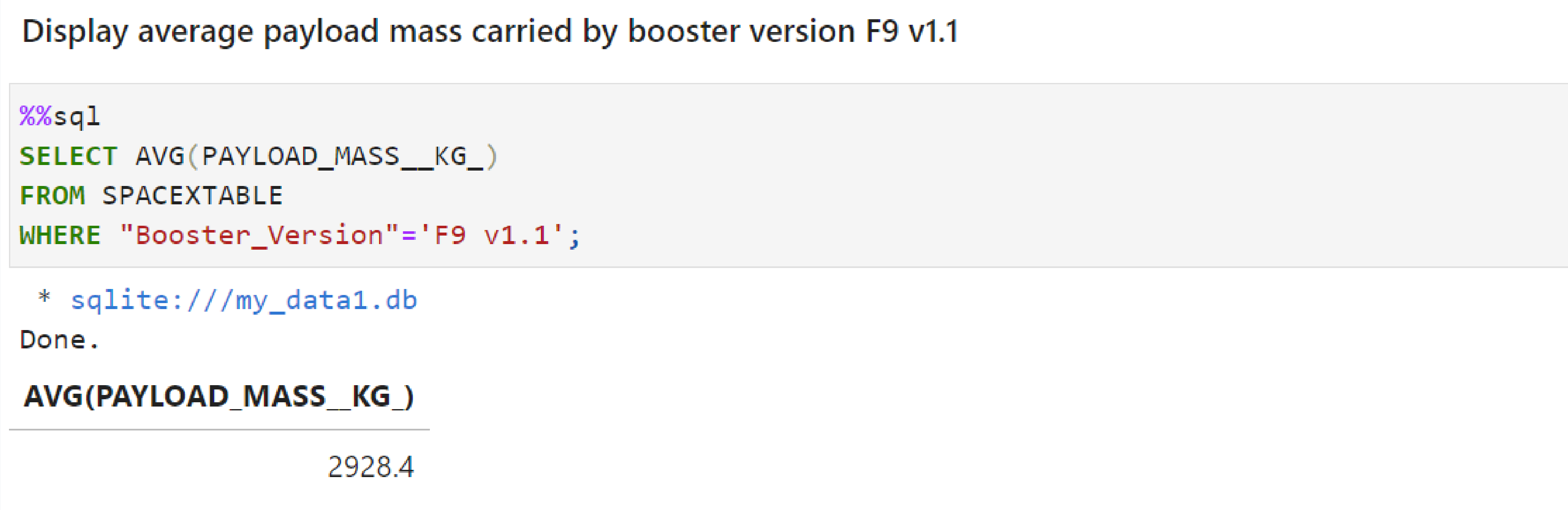
```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

The total PayloadMass was conditioned to be carried by boosters from (CRS), therefore we used the where function with this parameter.

Average Payload Mass by F9 v1.1

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



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

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE
WHERE "Booster_Version"='F9 v1.1';
```

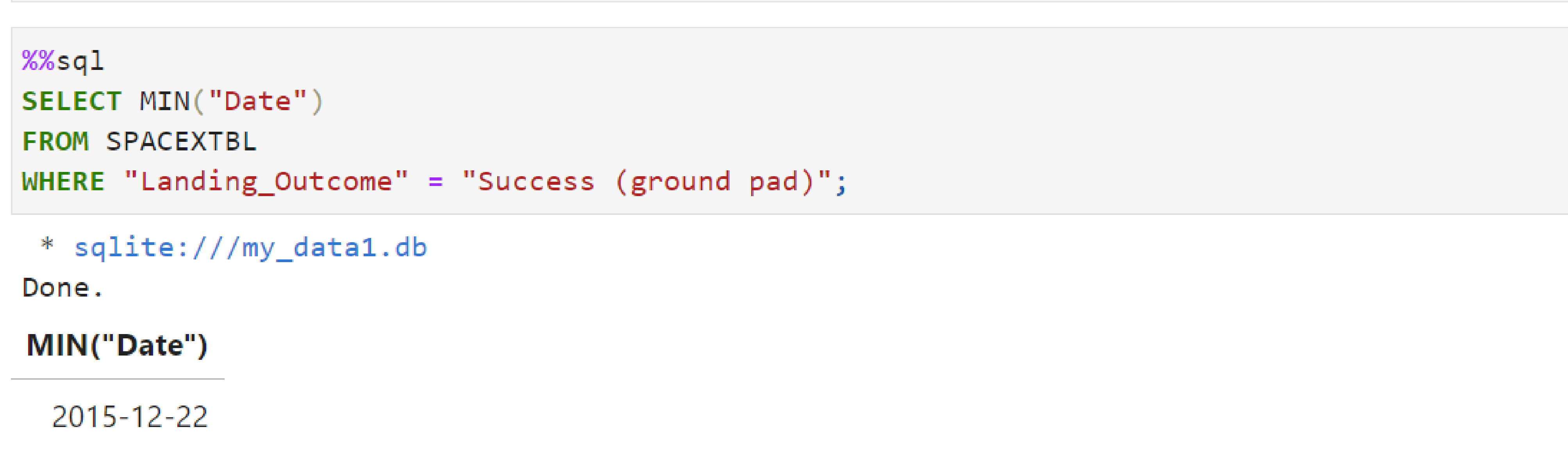
```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

The average PayloadMass was conditioned to be carried by booster version F9 v1.1 therefore we used the where function with this parameter.

First Successful Ground Landing Date

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



```
%%sql
SELECT MIN("Date")
FROM SPACEXTBL
WHERE "Landing_Outcome" = "Success (ground pad)";
```

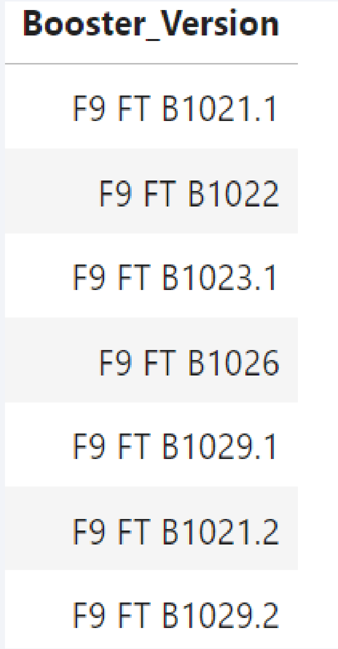
```
 * sqlite:///my_data1.db
Done.
```

MIN("Date")
2015-12-22

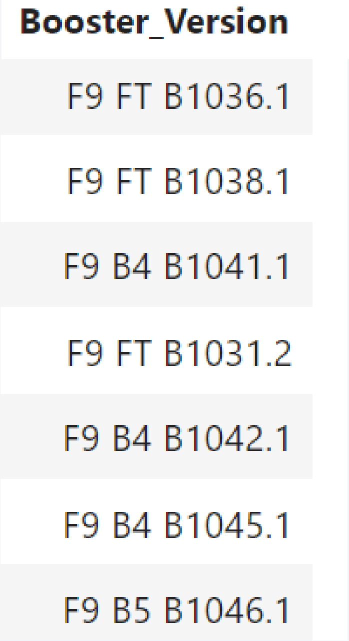
The query was made using success landing_outcome on ground pad as parameter in the where function to filter the first successful Ground landing date

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

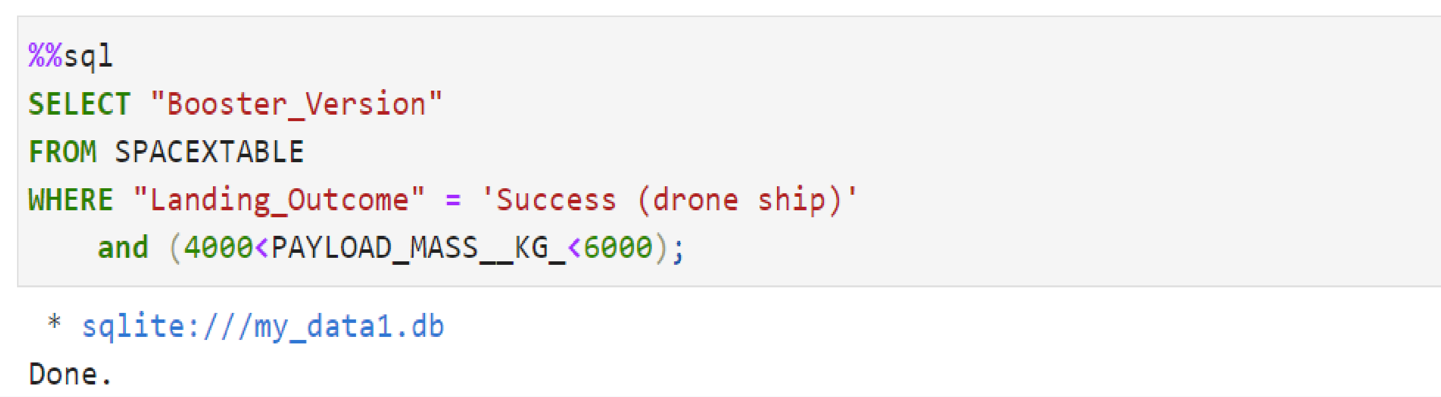


Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2



Booster_Version
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

- Present your query result with a short explanation here



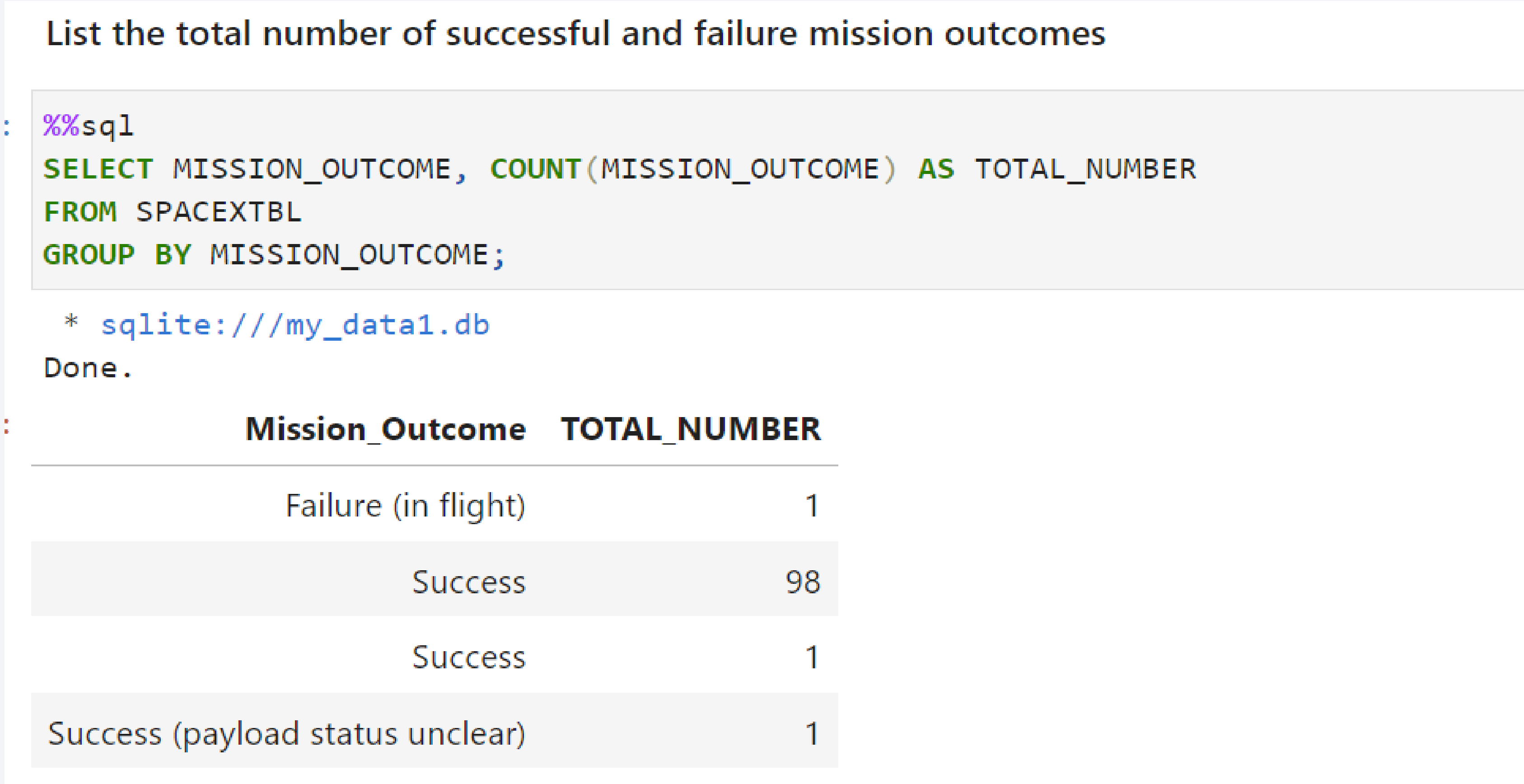
```
%%sql
SELECT "Booster_Version"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (drone ship)'
    and (4000<PAYLOAD_MASS__KG_<6000);
```

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

The query was made using two conditions as parameter success landing_outcome on drone ship and a specific payloadmass range as parameters

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes



List the total number of successful and failure mission outcomes

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

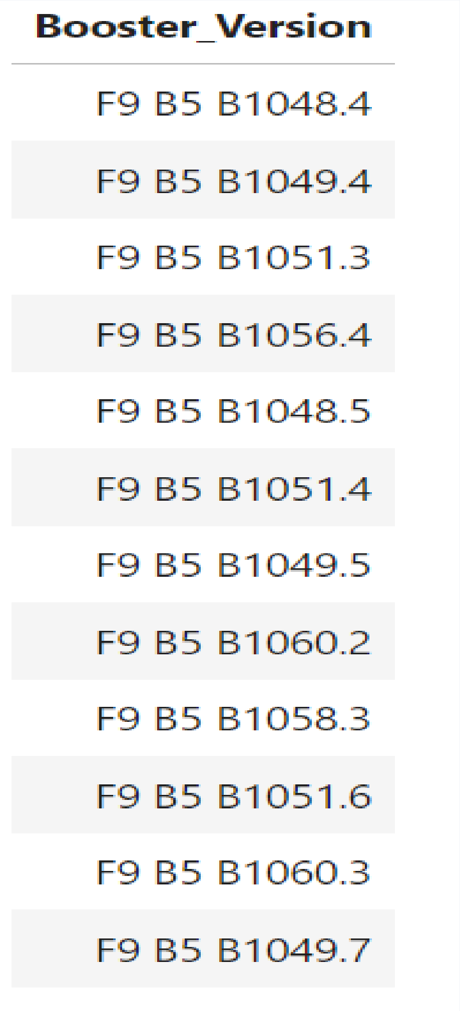
```
 * sqlite:///my_data1.db
Done.
```

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

The query used groupby function to get the total number of Mission outcomes per successful and failure classification

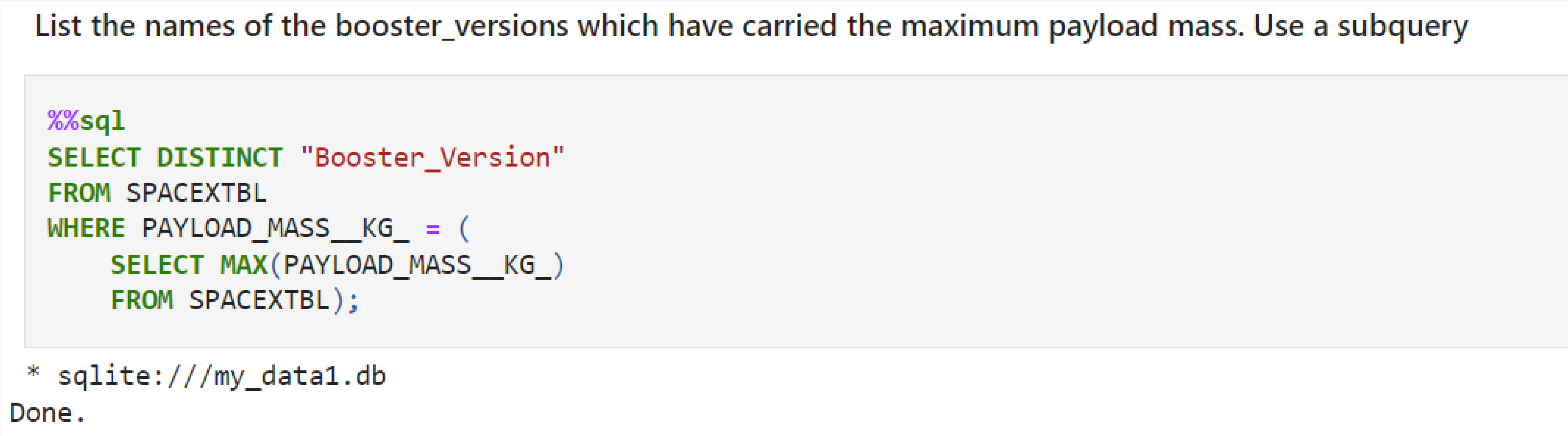
Boosters Carried Maximum Payload

- 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

- Present your query result with a short explanation here



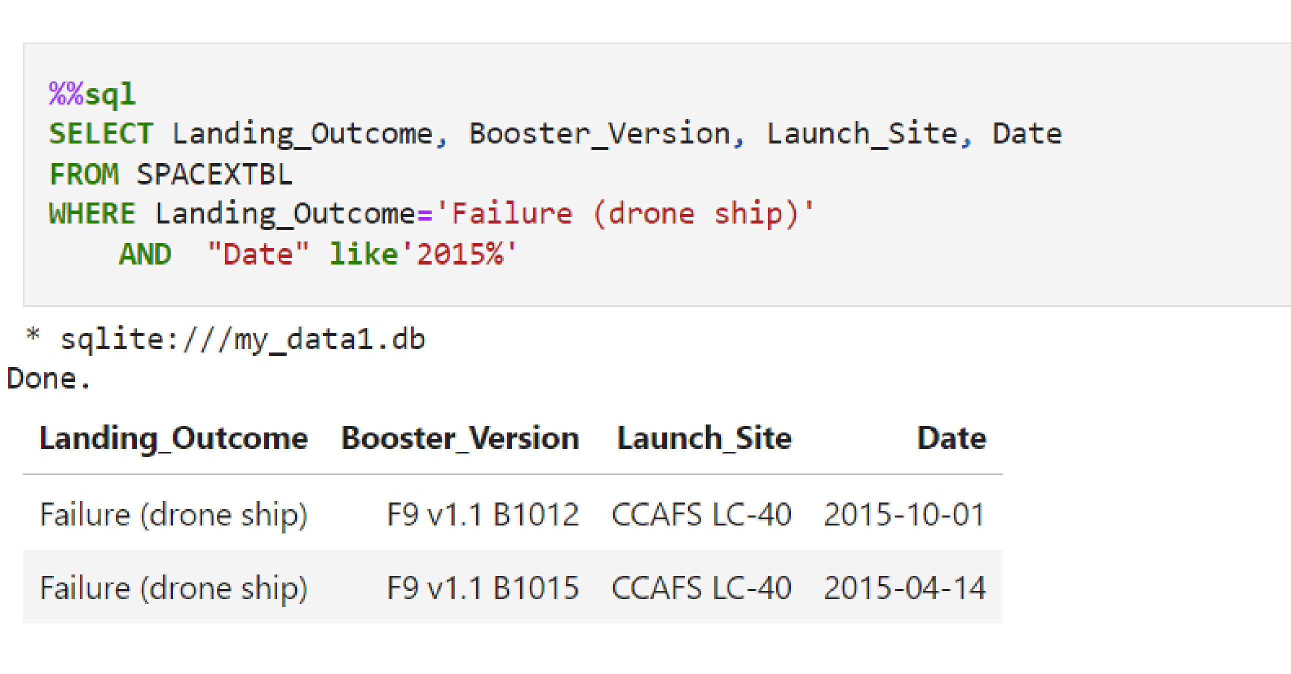
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL);
```

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

2015 Launch Records

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



```
%%sql
SELECT Landing_Outcome, Booster_Version, Launch_Site, Date
FROM SPACEXTBL
WHERE Landing_Outcome='Failure (drone ship)'
    AND  "Date" like'2015%'
```

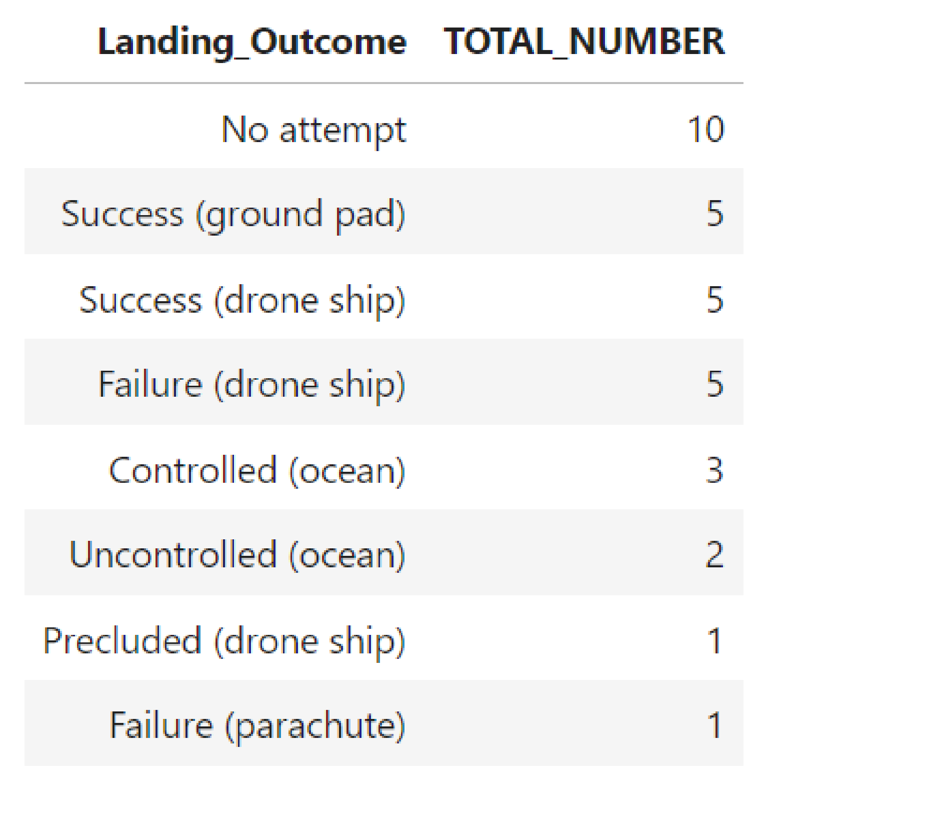
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Booster_Version	Launch_Site	Date
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-10-01
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

The query was made considering the conditions failure drone ship lainding_outcome in the year 2015

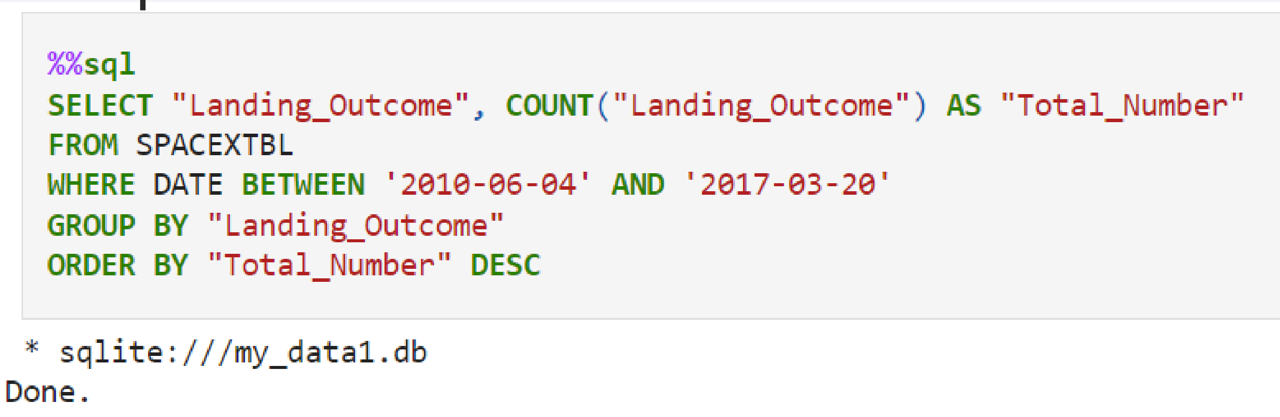
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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	TOTAL_NUMBER
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

- Present your query result with a short explanation here



```
%%sql
SELECT "Landing_Outcome", COUNT("Landing_Outcome") AS "Total_Number"
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY "Total_Number" DESC
```

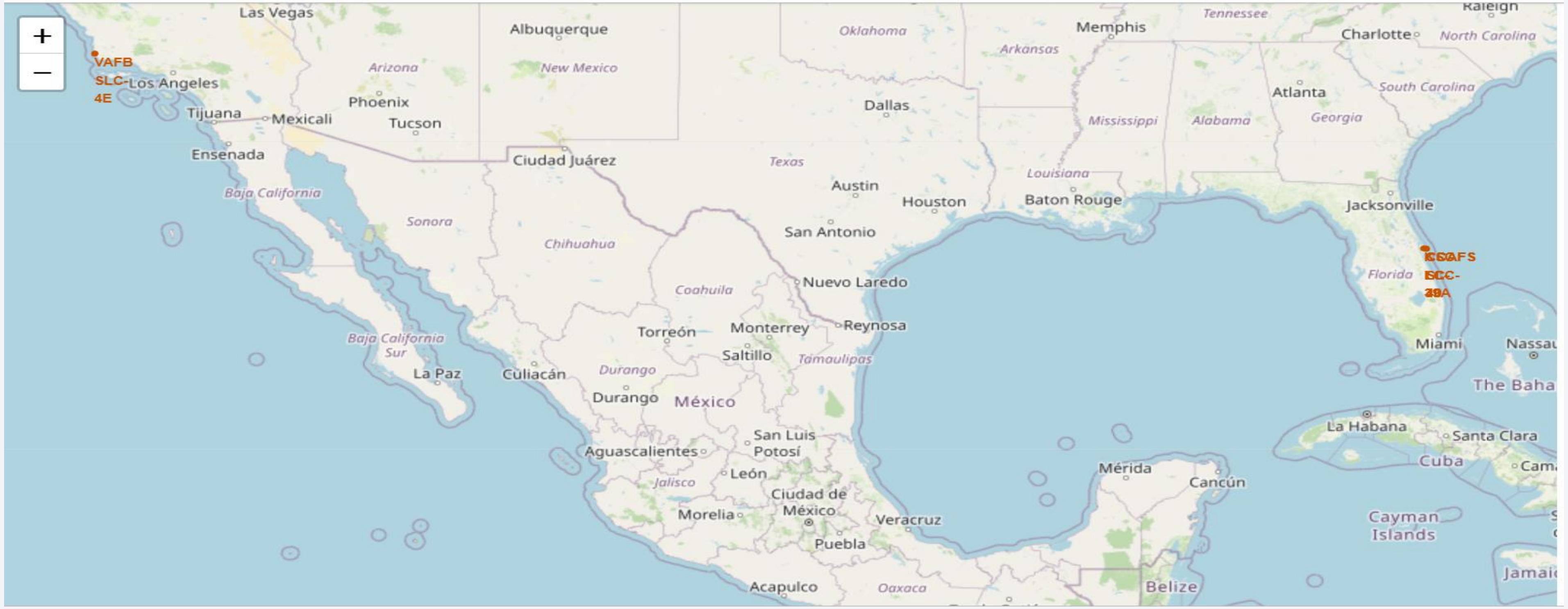
```
 * sqlite:///my_data1.db
Done.
```

The query used the groupby function to get the landing_Outcome records classified by by success and failure, then we used the DESC function to make the rank

Section 3

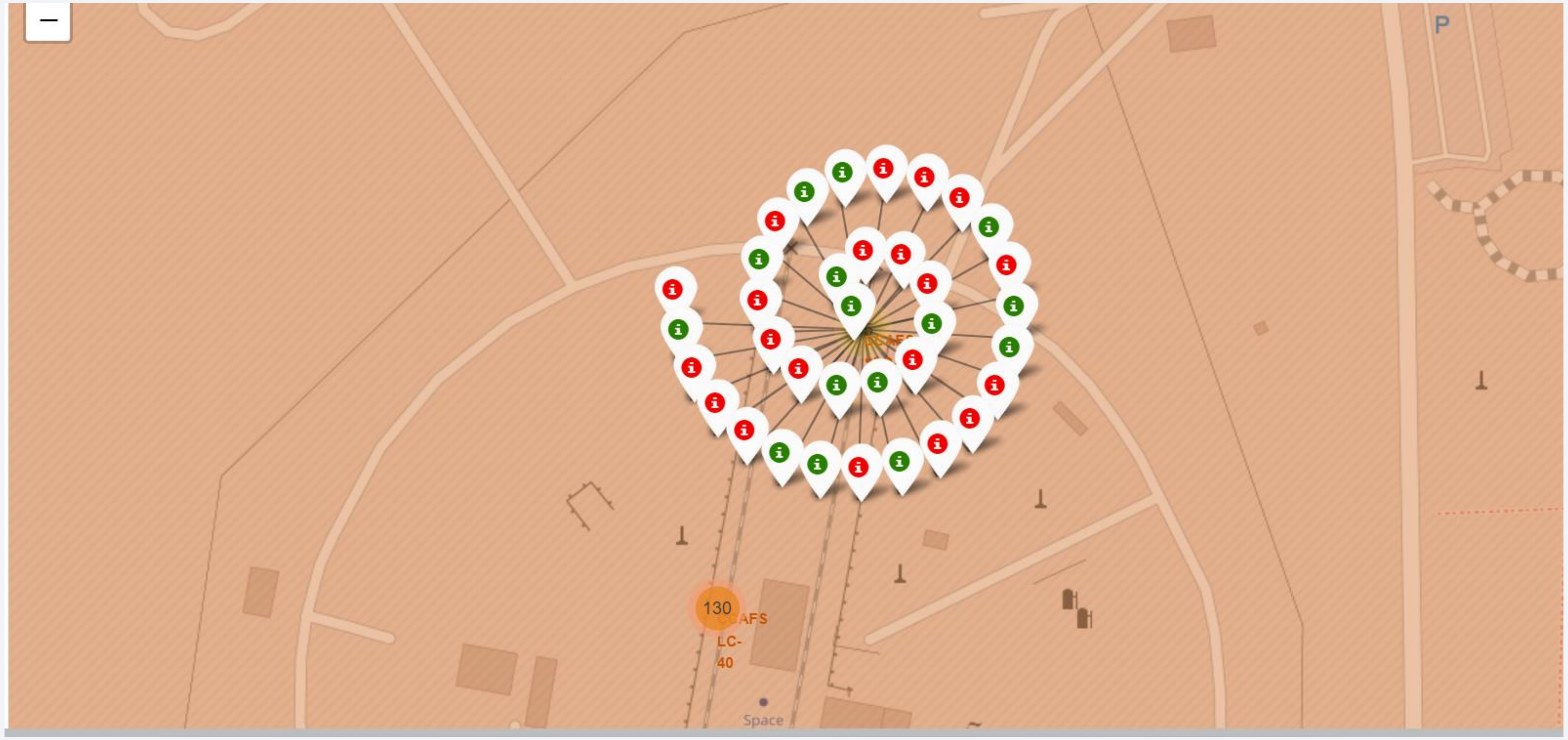
Launch Sites Proximities Analysis

Folium map - Launch Sites locations



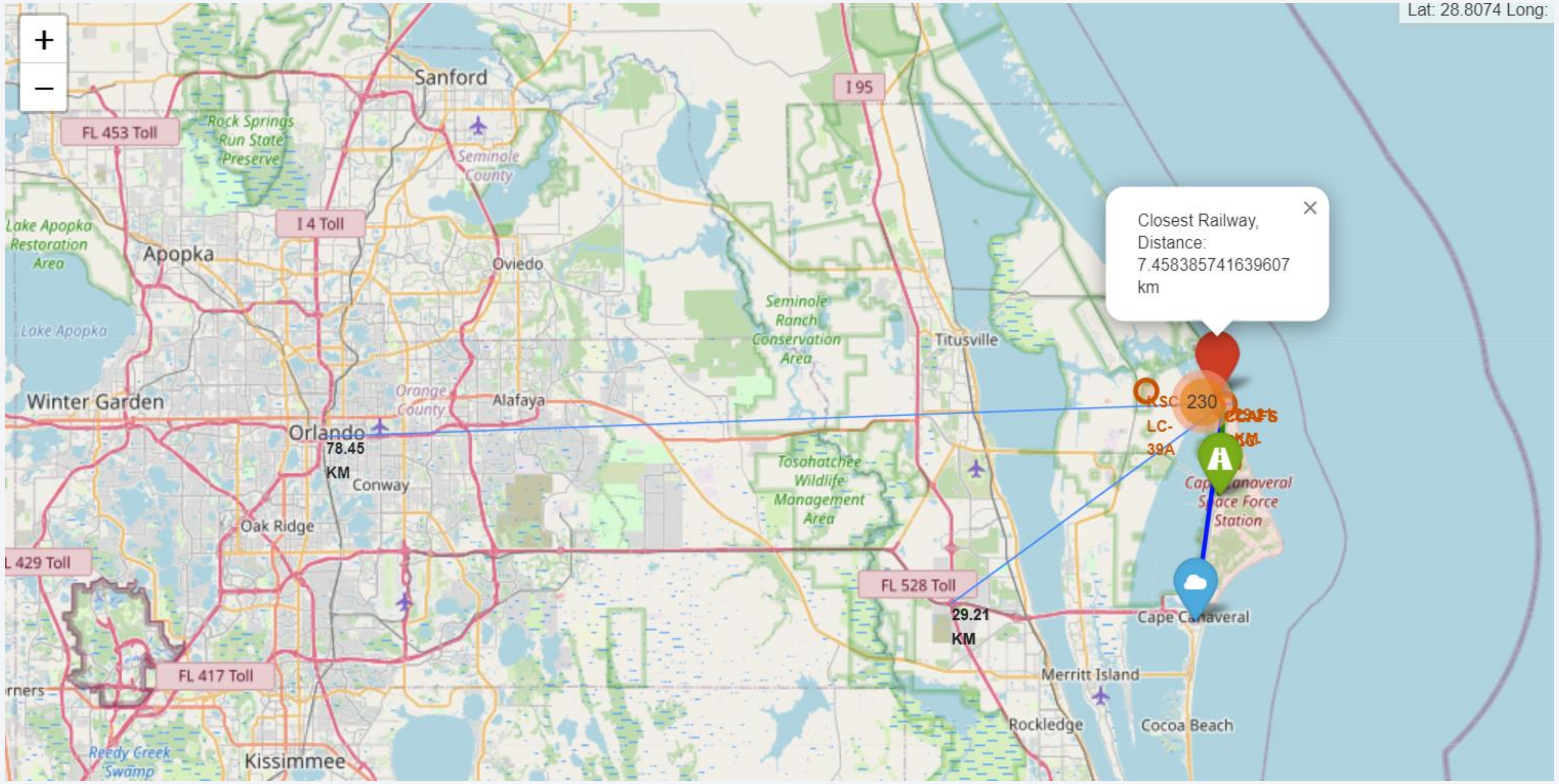
- Overall, all the launch sites are close to the coast, but VAFBSLC – 40 launch site has a very close proximity to the coast.
- The Launch sites are close to the Equator line, so their locations make easier to launch to the equatorial orbit

Folium Map - Launch Outcomes



- CCAFS SLC-40 launch Site has 42.8% success rate (15/35)
- CCAFS LC-40 has a 26.9.% success rate (35/130)
- The green markets indicates a successful launch (class=1)
- The red markers indicates an unsuccessful launch (class=0)

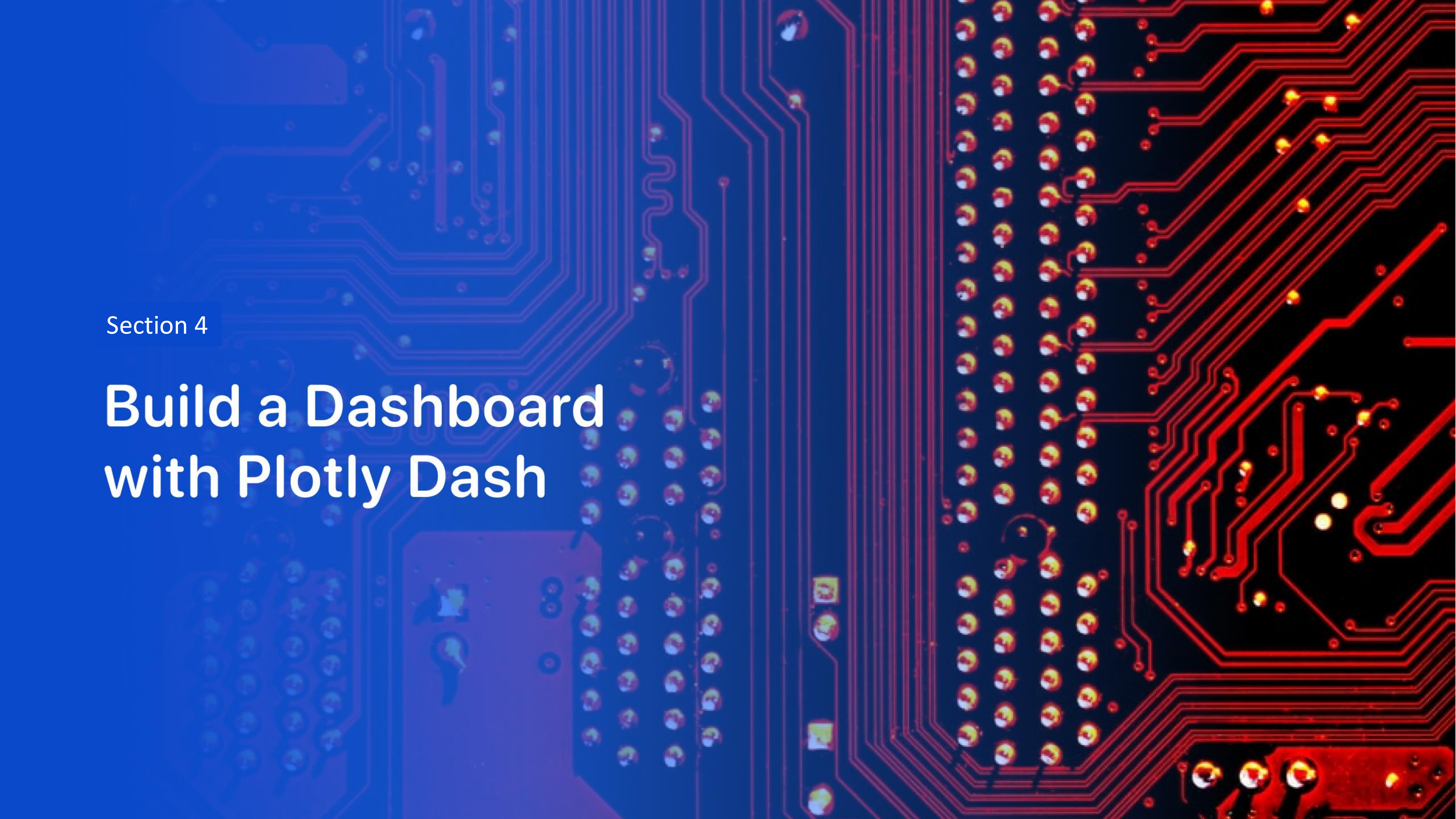
Folium Map - Distance to Proximities



Closest Railway, Distance: 7.458385741639607 km

CCAFS LC- 40 distances to:

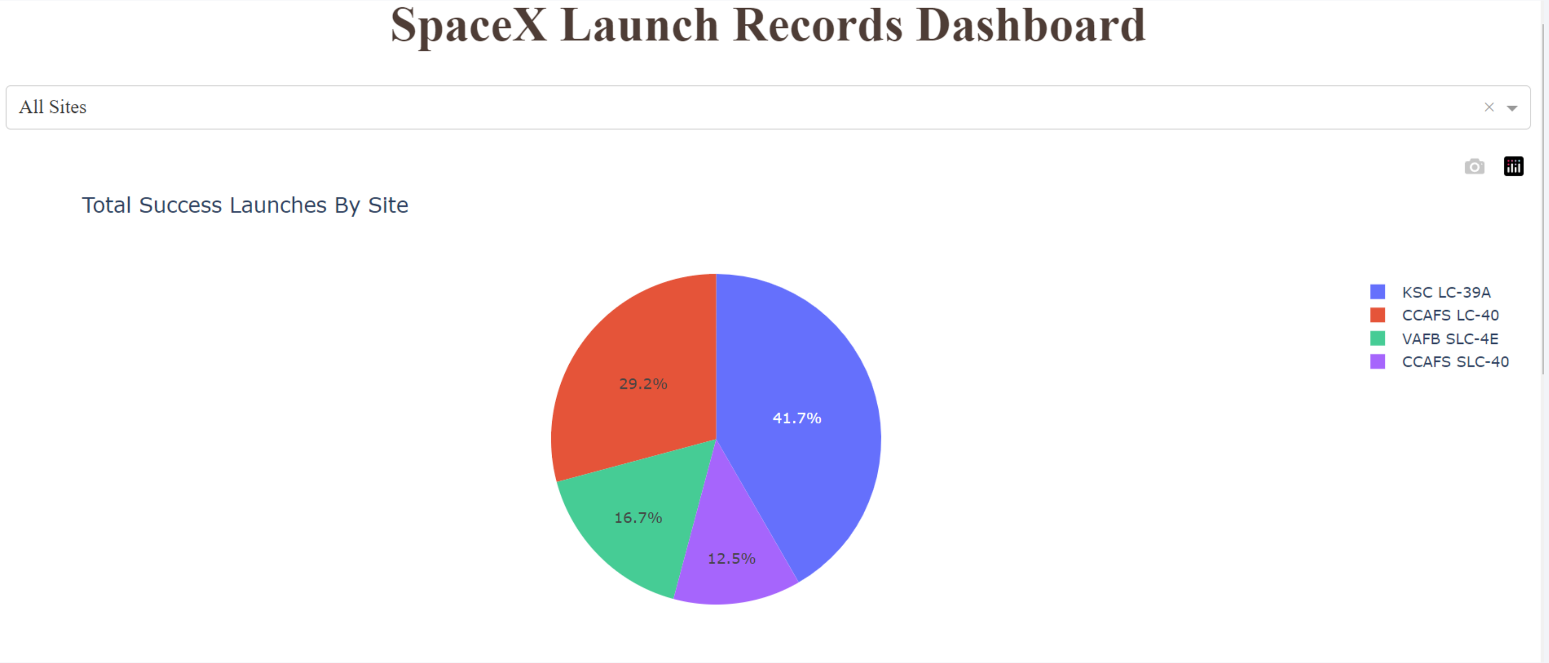
- Closest railway is 7.45 Km
- Closest city is 18.18 km
- Closest Highway: 1.35 km
- Closest coastline is 29.21 km



Section 4

Build a Dashboard with Plotly Dash

Dashboard – Launch Success



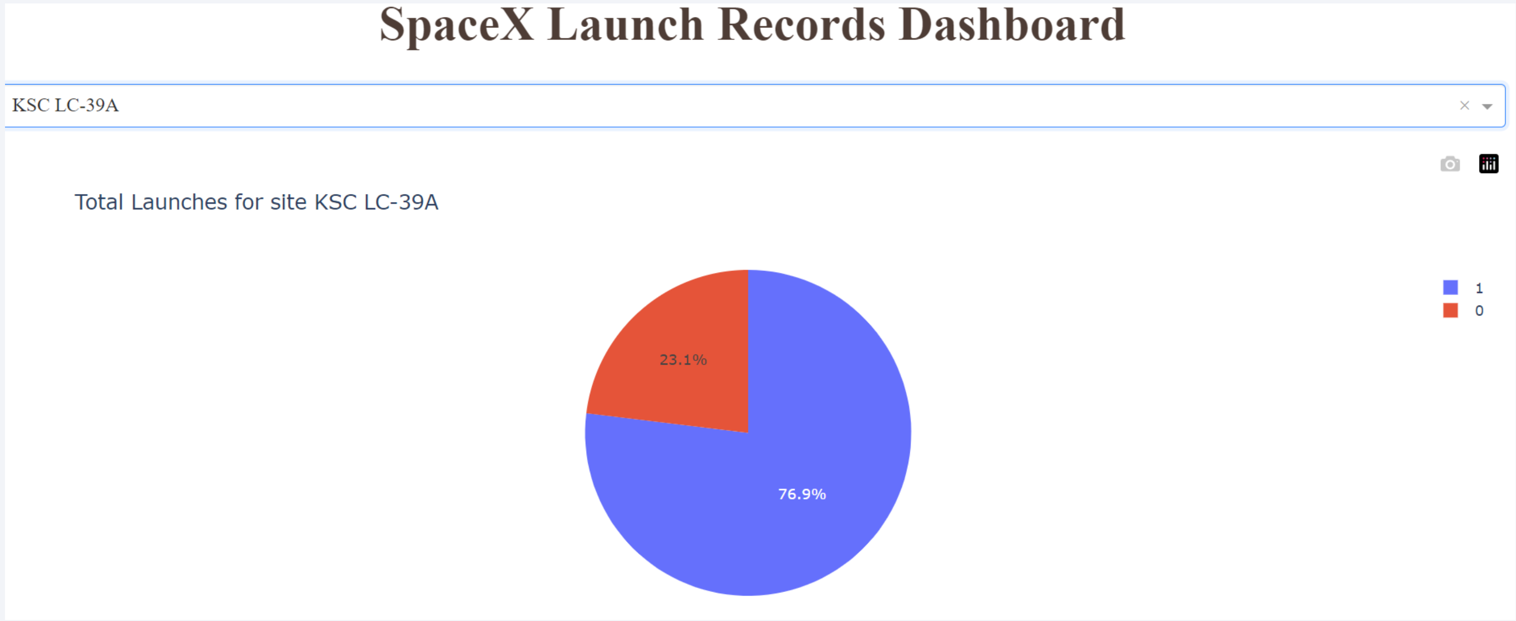
SpaceX Launch Records Dashboard

All Sites

Total Success Launches By Site

- The pie chart describe the success rate count per launch site.
- As is shown in the pie chart, KSC LC 39A has the most successful launches percentage (41.2%), followed by   with …

Dashboard – Launch Success (KSC LC-39A)



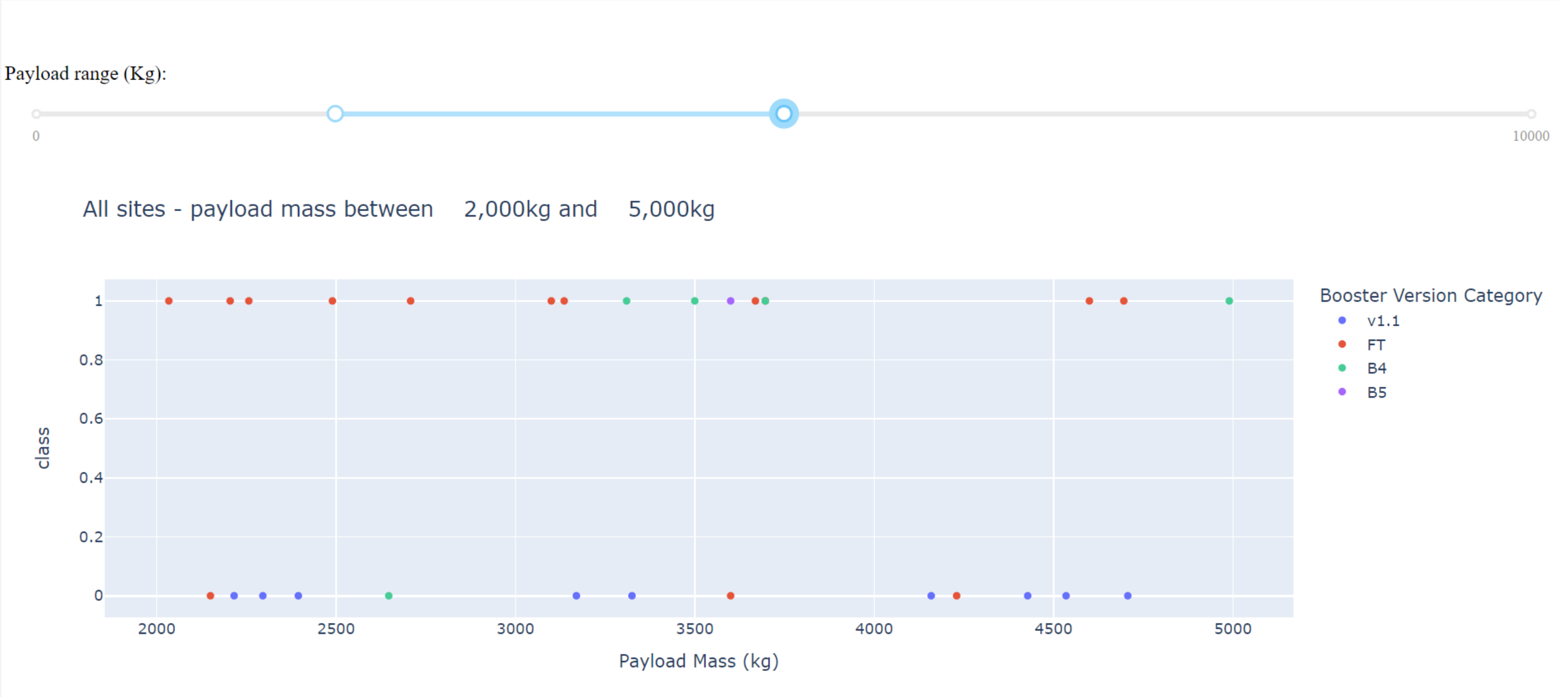
SpaceX Launch Records Dashboard

KSC LC-39A

Total Launches for site KSC LC-39A

The pie chart shown the launch site with the highest success counts, in this case KSC LC-39A. This launch site has the success rate of 76.9%)

Dashboard - Payload vs. Launch Outcome



Payload range (Kg):

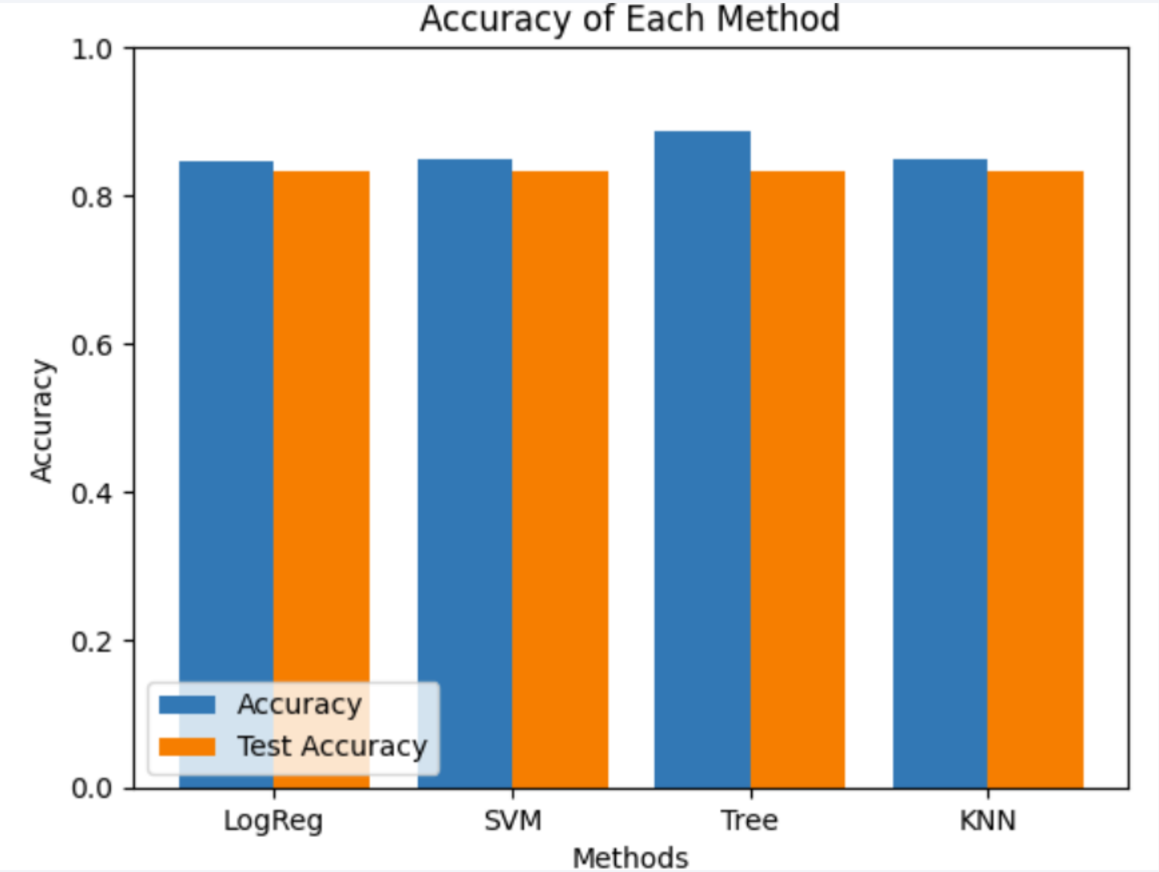
All sites - payload mass between 2,000kg and 5,000kg

- The scatter plot shown how the payload is correlated with the Launch Outcome.
- Payloads between 2000 Kg and 5000 kg shown the highest success rate

Section 5

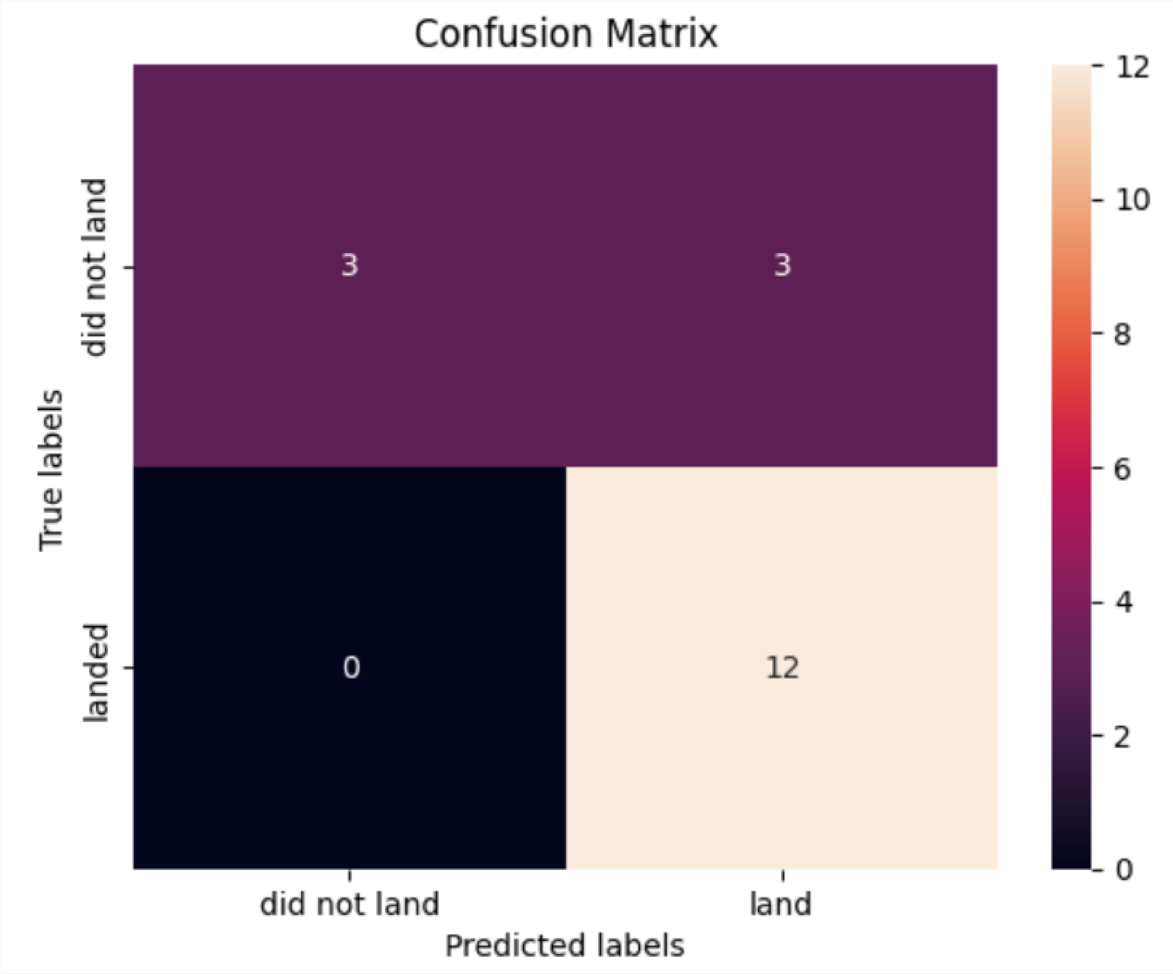
Predictive Analysis (Classification)

Classification Accuracy



The model with the highest classification accuracy is the Decision Tree Model, which slightly outperformed when looking at .best_score_ with 0.8857. However, all the models performed with same test-accuracy results

Confusion Matrix



- All the confusion matrix were identical, all presented Type 1 error:
  - 3 True negative
  - 3 False positive (type 1 error)
  - 0 false negative
  - 12 True Positive

Conclusions

It is concluded:

- FlightNumbers and Payloads are correlated with landing success rate.

- Launch success rate kept improved from 2013 to 2019

- Orbits GEO, LEO, SSO had the most success rate.

- KSC LC- 39A had the most successful launches among all launch sites.

- Launches with a Payload mass range between 2000kg and 5000 kg tended to reach a success landing

- The classification Tree is the best machine learning algorithm to predict landing outcomes.

Thank you!