IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Javier Contreras
22-11-2022

Outline

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

Executive Summary

- Summary of methodologies
  - ➢ Data Collection with SpaceX API and web scraping
  - ➢ Data analysis with SQL, data visualization and interactive dashboard
  - ➢ Machine Learning predictions
- Summary of all results
  - ➢ Exploratory data analysis
  - ➢ Predictive analysis

Introduction

The objective of this project is to predict if the SpaceX Falcon 9 rocket first stage will land successfully in order to determine the cost of launch. So, if another space company wants to bid against them for a rocket launch, they can use these predictions.

Principal problems:

- Identify the relation of all the variables that affects the outcome
- What is the best classification model to predict the outcome?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - SpaceX Rest API
    - Web Scraping
- Perform data wrangling
    - Filtering Falcon 9 cases
    - Missing values
    - One-hot encoding to categorical variables

Methodology

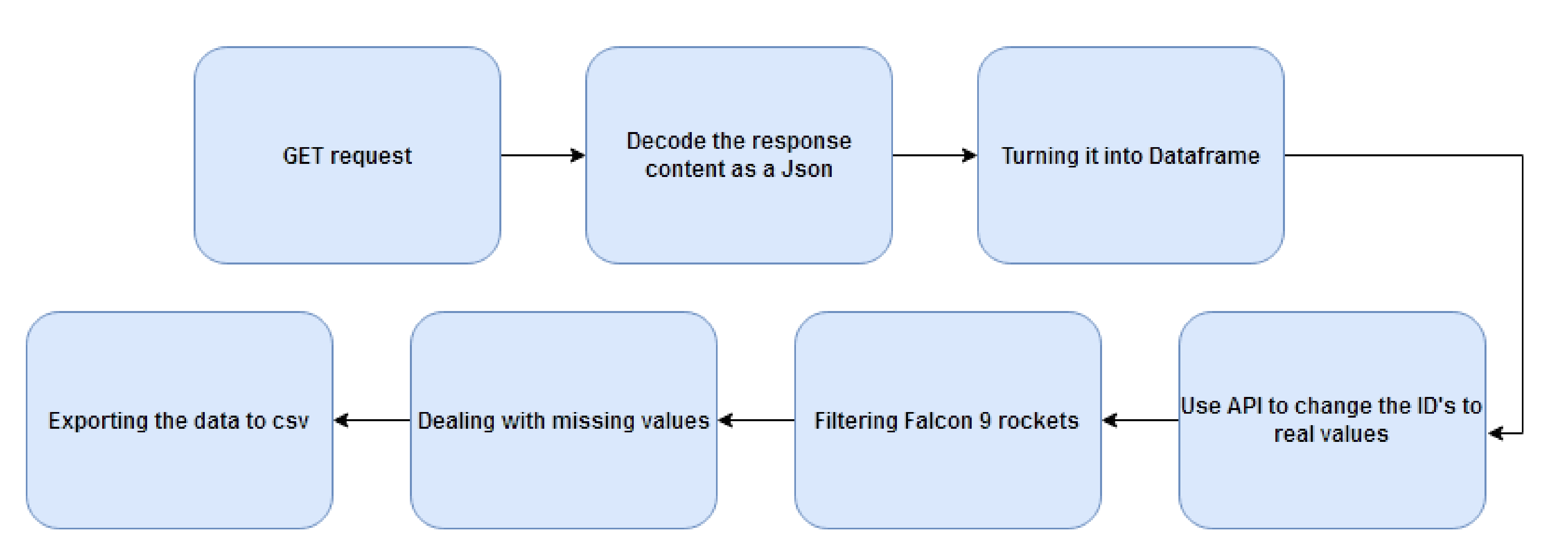
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data was normalized, splitted in train and test data sets and using GridSearch to find the best hyperparameters the data was trained with Classification Models such as Logistic Regression, Support Vector Machine, Decision Tree and K-Nearest Neighbors.

Data Collection

The data was collected from:

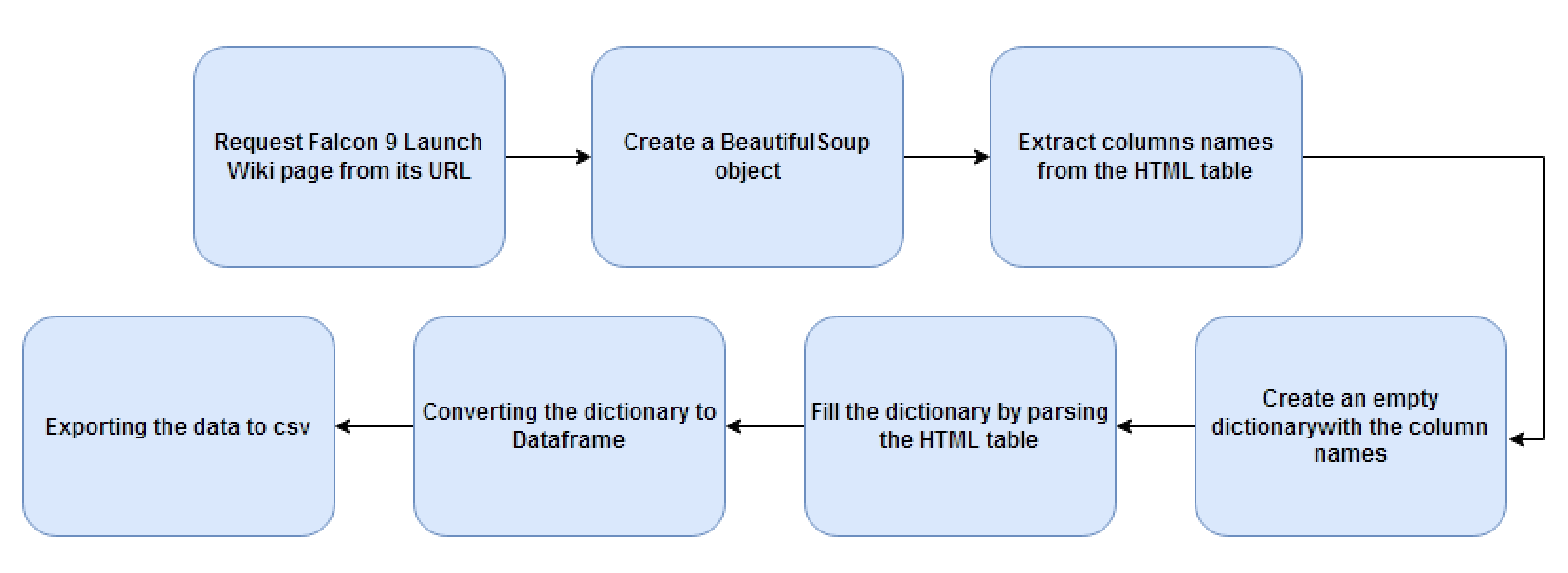
- SpaceX REST API using the get request and turning the content to a Pandas Dataframe.
- Wikipedia using BeautifulSoup extracting the records as an HTML table, parsing it and converting it to a Pandas Dataframe

Data Collection – SpaceX API



GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Exploratory Data Analysis was performed to find patterns and to determine the label of the Supervised models.

The number of launches on each site, number and occurrence of each orbit and mission outcome type were calculated.

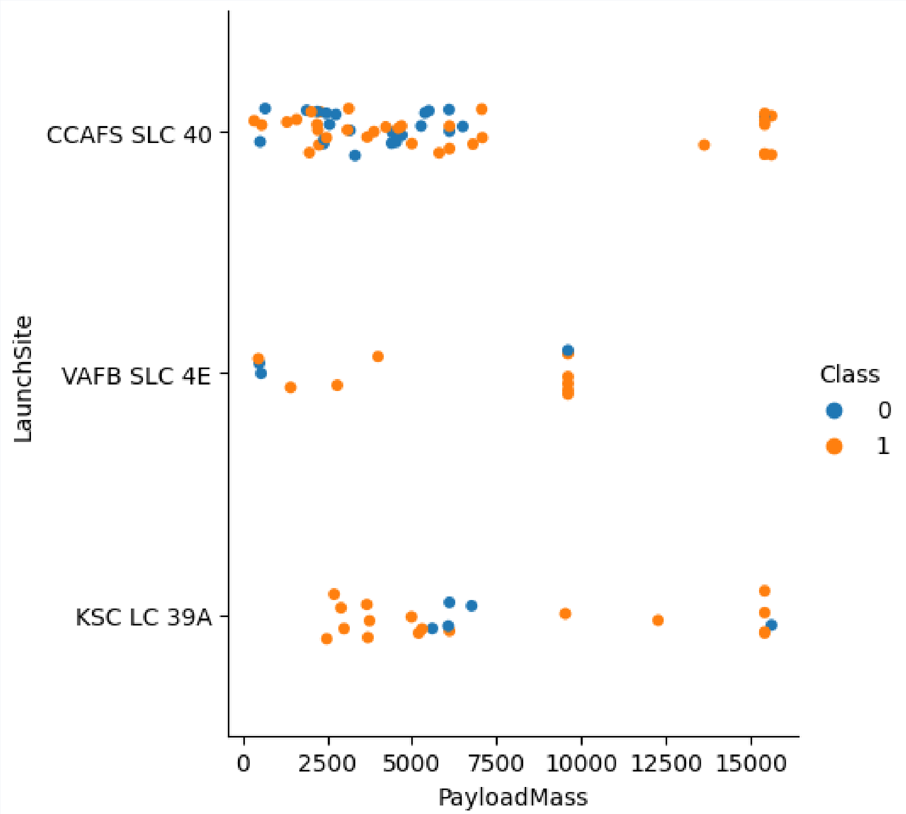
The values of the "Outcome" variable were used to create a new variable "Class" equal to 1 if it was a successful landing or 0 unsuccessful landing.

GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

The charts were plotted to identify patterns between features, those are:

FlightNumber vs PayloadMass, FlightNumber vs LaunchSite, PayloadMass vs LaunchSite, success rate of each orbit type, FlightNumber vs Orbit, PayloadMass vs Orbit, launch success yearly trend.



GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries

- Display the names of the unique launch sites in the space mission.
- Display 5 records where launch sites begin with the string.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1.
- List the date when the first successful landing outcome in ground pad was achieved.
- 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

The following map objects were added to the folium map:

- Markers: indicates launch sites.
- Circles: highlighted area with a text label.
- Lines: indicate distances between coordinates.
- Marker clusters: success/failed launches on each launch site.
    - Green mark: success
    - Red mark: failed

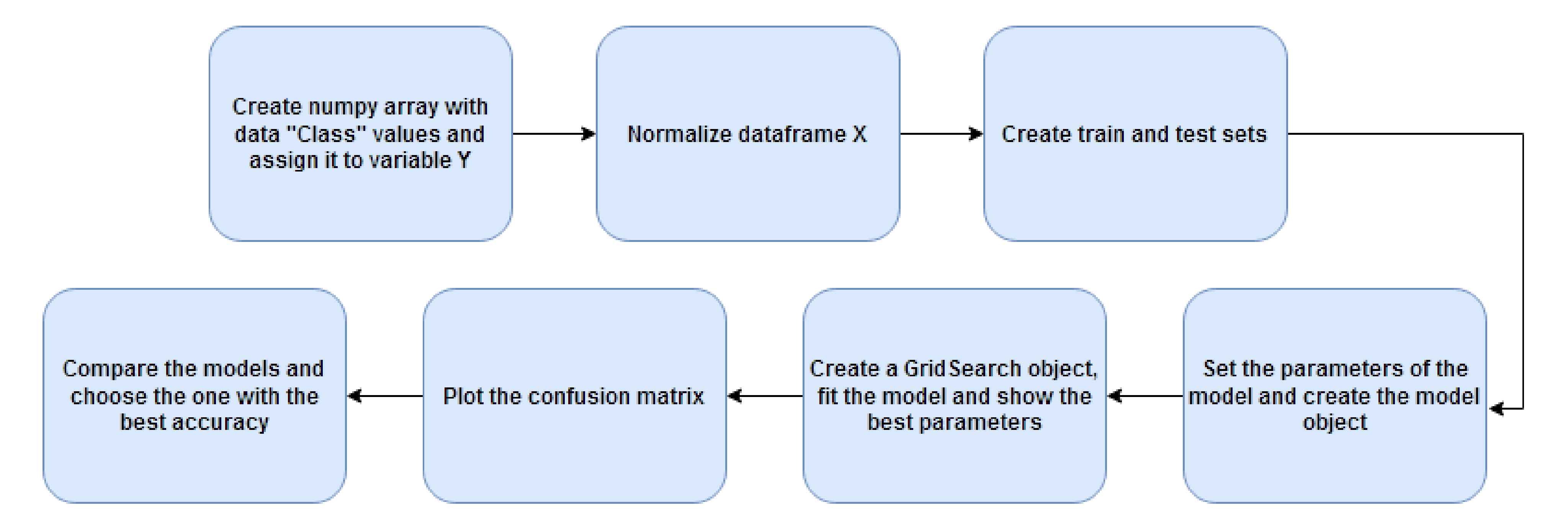
GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Searchable dropdown list to filter by a specific launch site.
- Pie chart showing the total success launches per-site.
- Slider to select a payload range.
- Scatter plot Payload Mass vs class by Booster Version category.

GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/plotly%20dash.py

Predictive Analysis (Classification)



GitHub code: https://github.com/javier-icvds/Applied-Data-Science-Captsone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

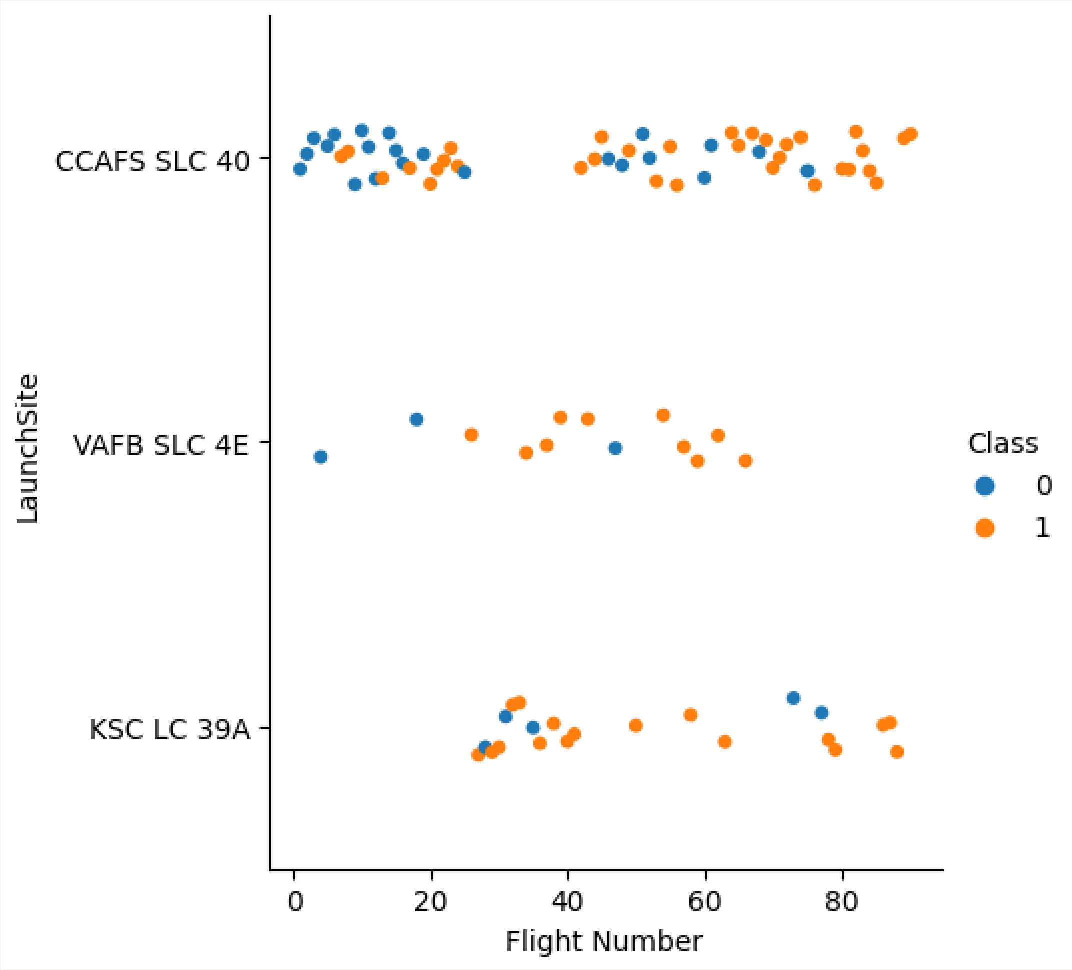
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

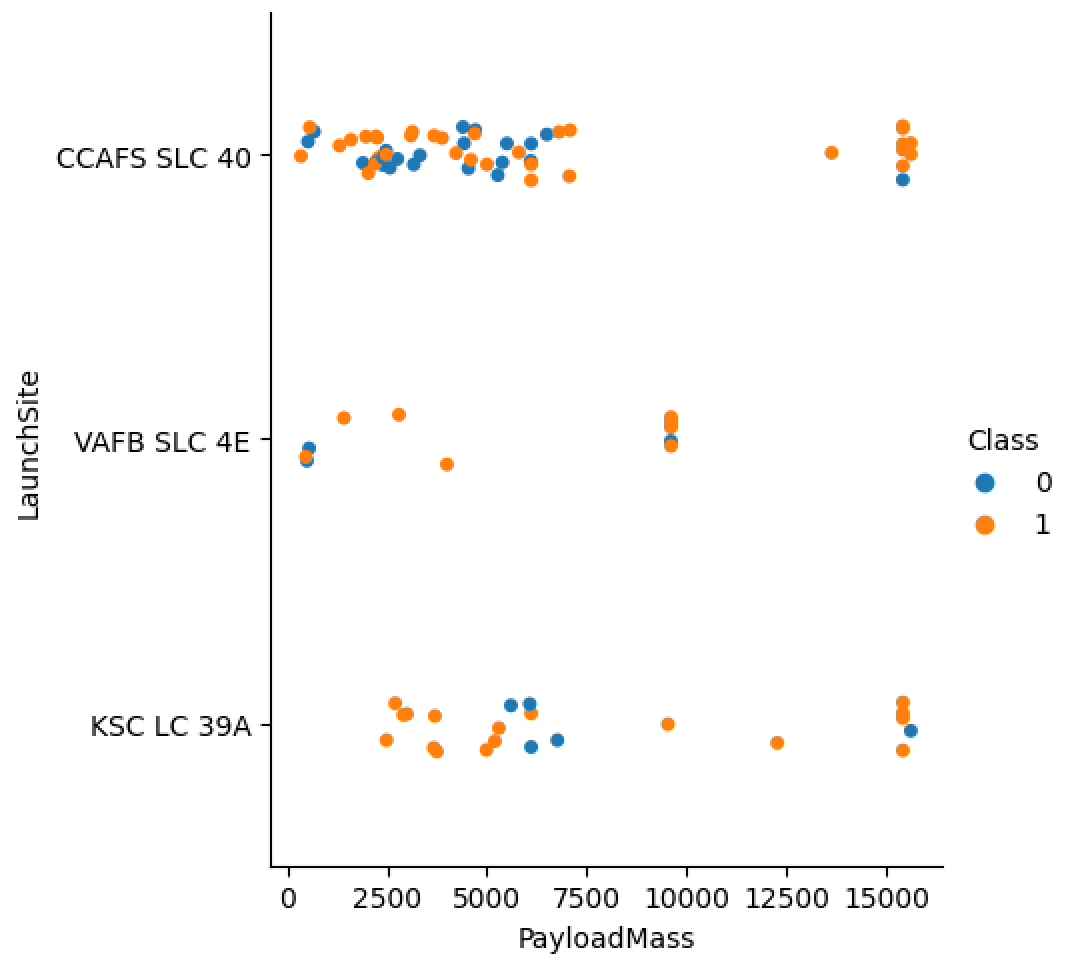
Insights drawn from EDA

Flight Number vs. Launch Site

- Early launches were unsuccessful while the latest were successful.
- As the time goes by, the launches are more likely to succeed.



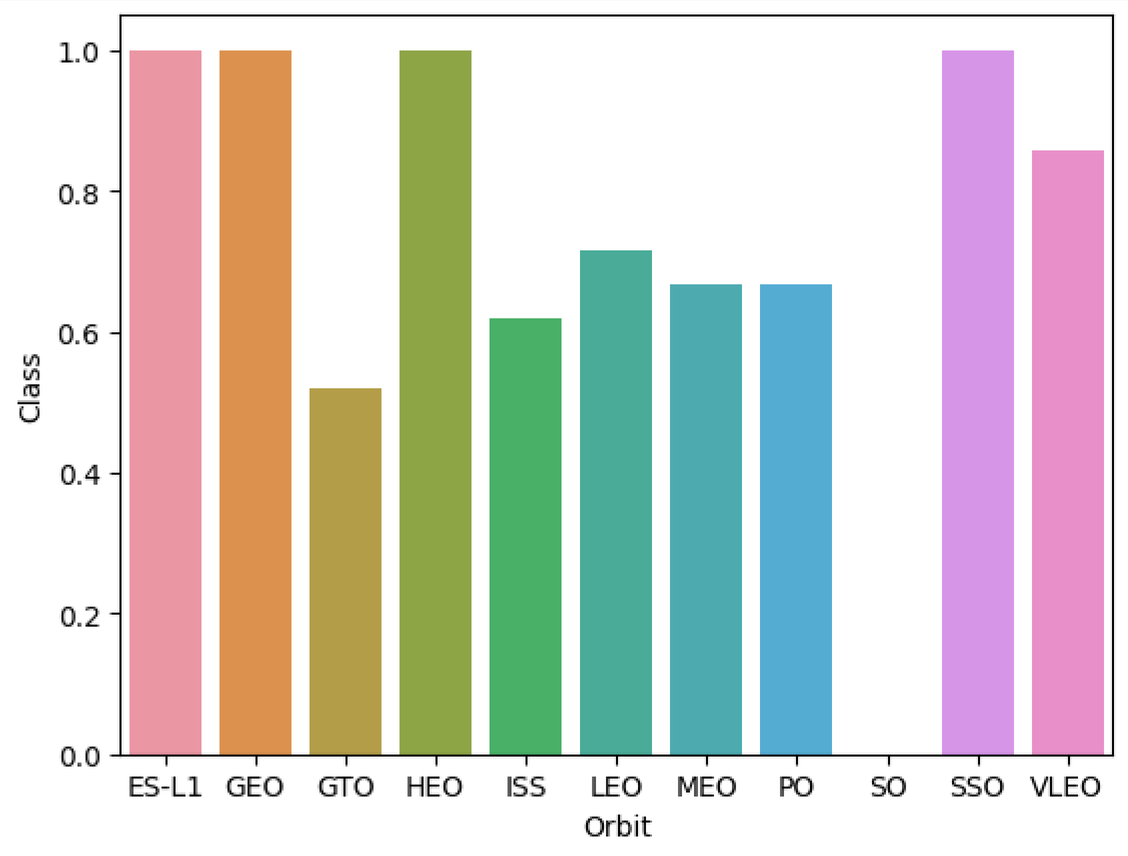
Payload vs. Launch Site



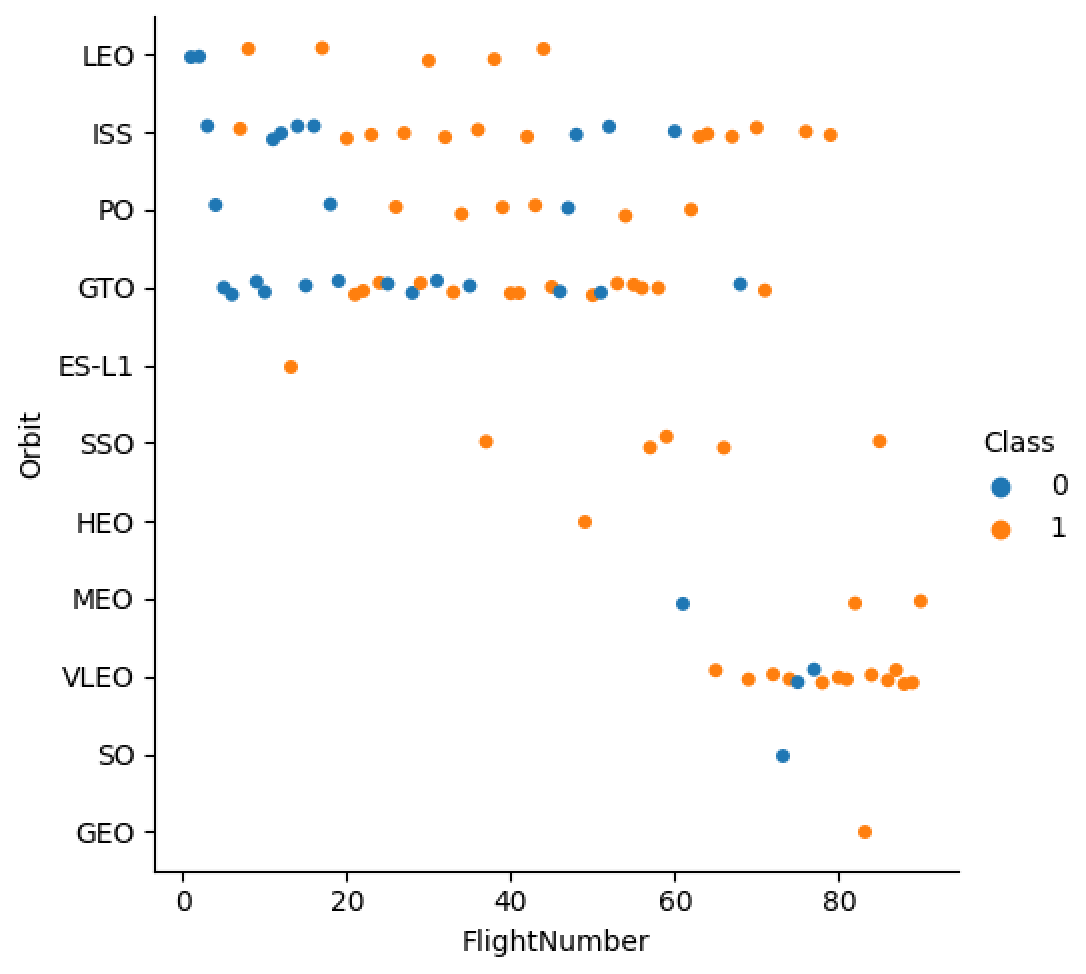
- Launches with a higher payload mass are in majority succeed.
- KSC LC 39A have a high probability to succeed in payloads mass below 6000.

Success Rate vs. Orbit Type

- Orbits with 100% success rate
  - ES-L1, GEO, HEO, SSO
- Orbit SO with 0% success rate
- Followed by VLEO(~88%), LEO (~70%), MEO and PO (~66%), ISS (~61%), HEO (~55%).



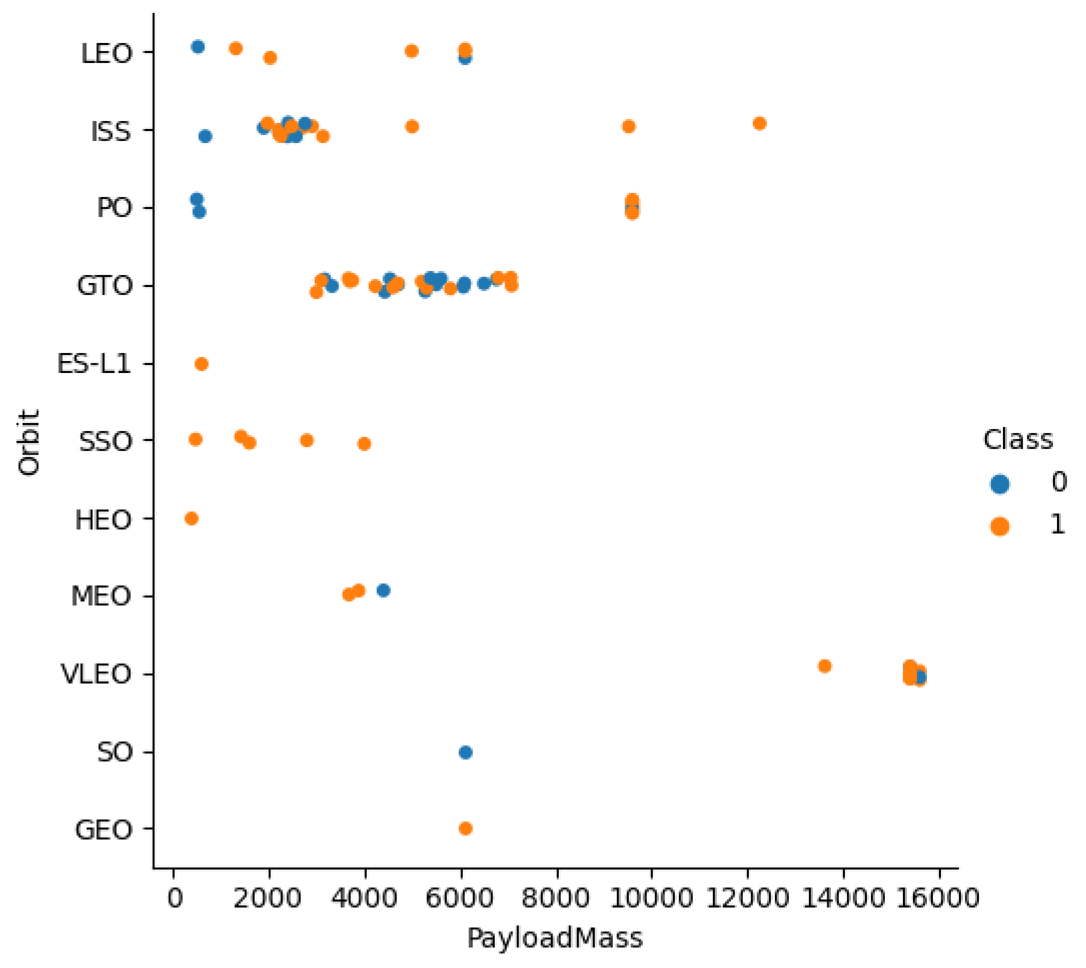
Flight Number vs. Orbit Type



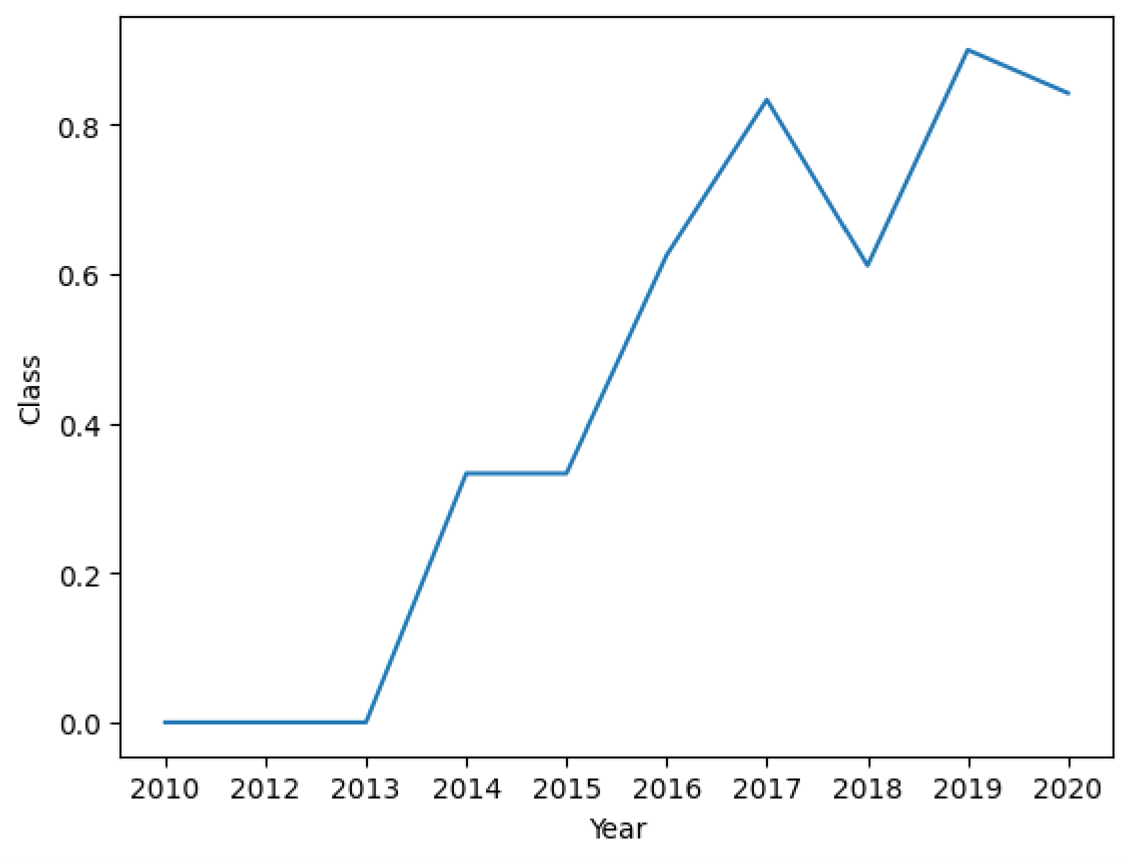
- Success rate improves as time goes by.
- VLEO is the most attractive orbit to ensure a successful launch due to its frequency launches and succeed rate.
- ES-L1, HEO, MEO, SO, GEO should not be considered because they have very few records.

Payload vs. Orbit Type

- The success rate increases significantly when the payload mass is greater than aprox. 9000 for ISS, PO anda VLEO.
- For orbit SSO the success rate is 100% for payload mass lower than 4000.
- For orbit VLEO the success rate is higher when the payload mass is greater than 13000.



Launch Success Yearly Trend



- The success rate has increased since 2013 from 0% to 80% in 2020.

All Launch Site Names

- The names of all site names are the following:

```
%sql SELECT DISTINCT(Launch_Site) from SPACEXTBL
```

```
 * 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'

- 5 records where launch sites begin with `CCA`:

```
%sql SELECT * from SPACEXTBL 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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA (CRS):

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

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Booster_Version" LIKE '%F9 v1.1%'
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

```
%sql SELECT MIN(DATE) as First_successful_landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)'
✓ 0.8s

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

First_successful_landing
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql SELECT "BOOSTER_VERSION" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)'
AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000
✓ 0.7s

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

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

```
%%sql SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS,
(SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE
```

✓ 0.8s

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

SUCCESS	FAILURE
100	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:

```
%%sql SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL
WHERE "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_") FROM SPACEXTBL)
```

✓ 0.1s

```
 * 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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015:

```
%%sql SELECT substr("DATE", 4, 2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Failure (drone ship)' and substr("DATE",7,4) = '2015'
```

✓ 0.7s

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

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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:

```
%%sql SELECT "Landing_Outcome", COUNT("Landing_Outcome") FROM SPACEXTBL
WHERE "DATE" >= '04-06-2010' and "DATE" <= '20-03-2017' and "Landing_Outcome" LIKE '%Success%'
GROUP BY "Landing_Outcome"
ORDER BY COUNT("Landing_Outcome") DESC
```

✓ 0.5s

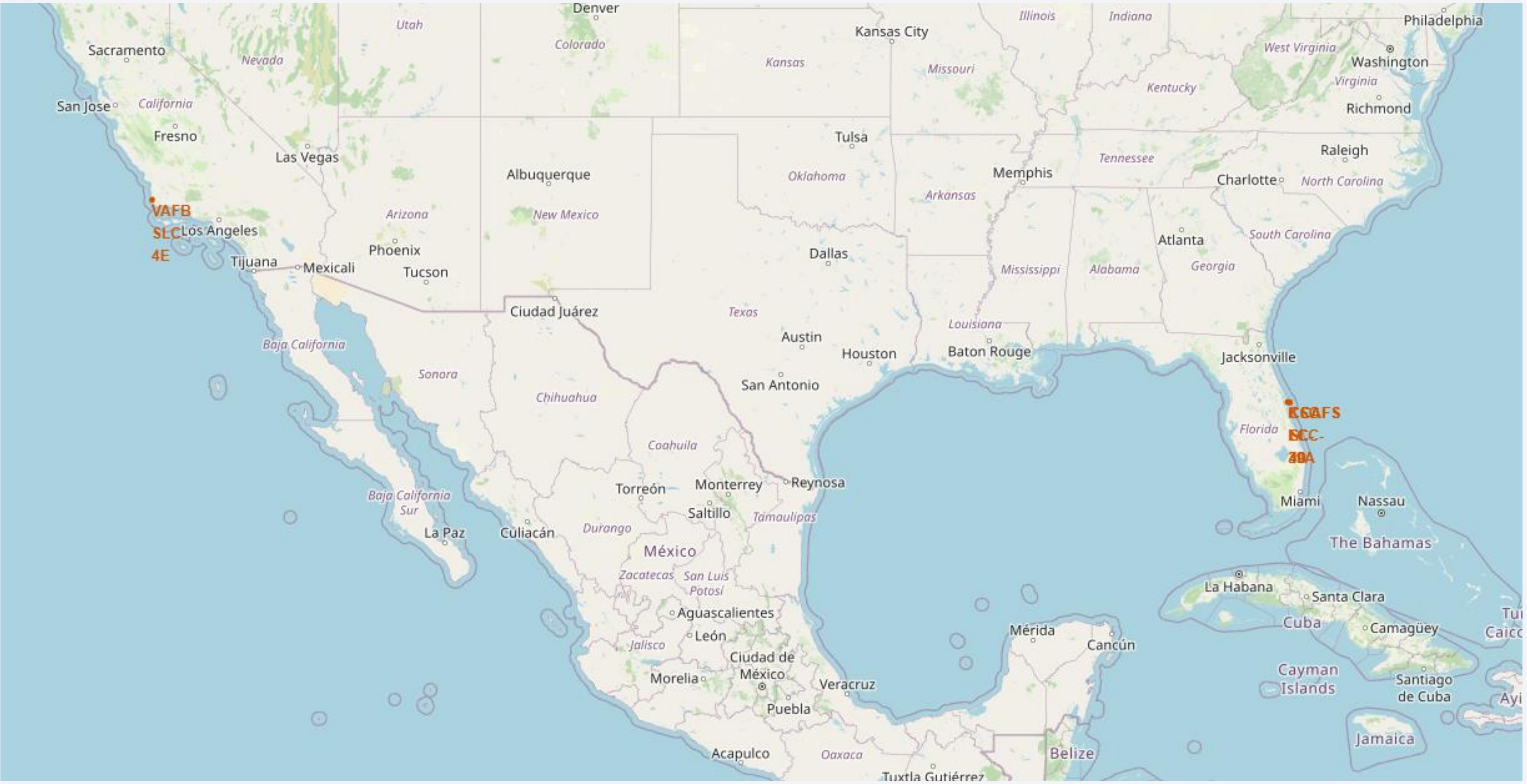
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	COUNT(Landing_Outcome)
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

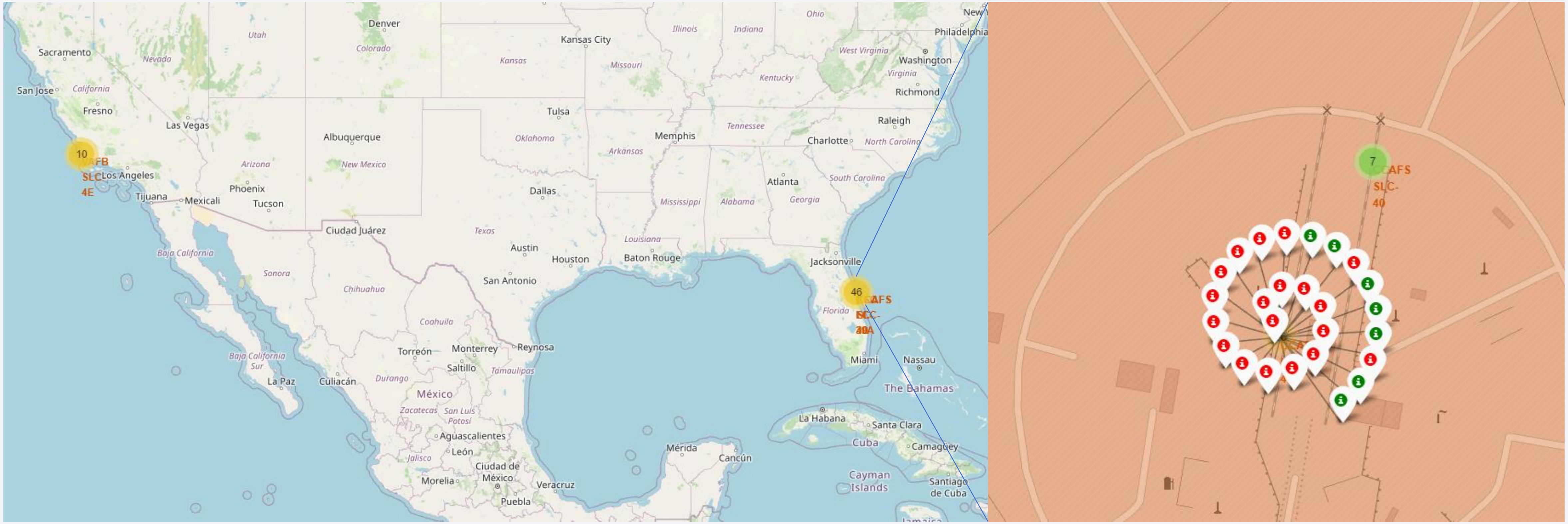
Launch Sites Proximities Analysis

Launch Site Locations



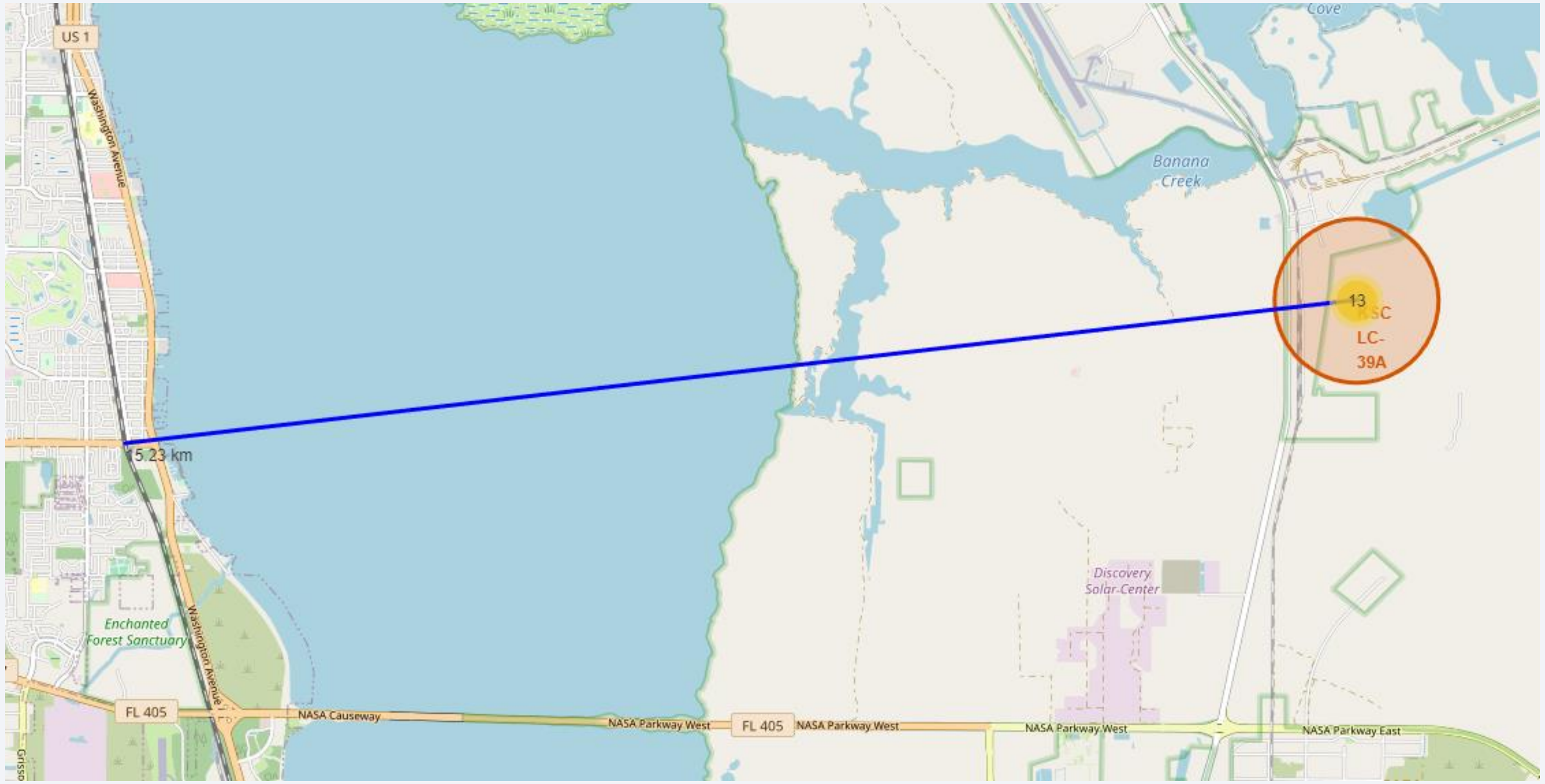
- All the locations are nearly the sea.

Launch outcomes by launch site



Green markers show successful launches and red markers unsuccessful launches.

Proximity to surrounds

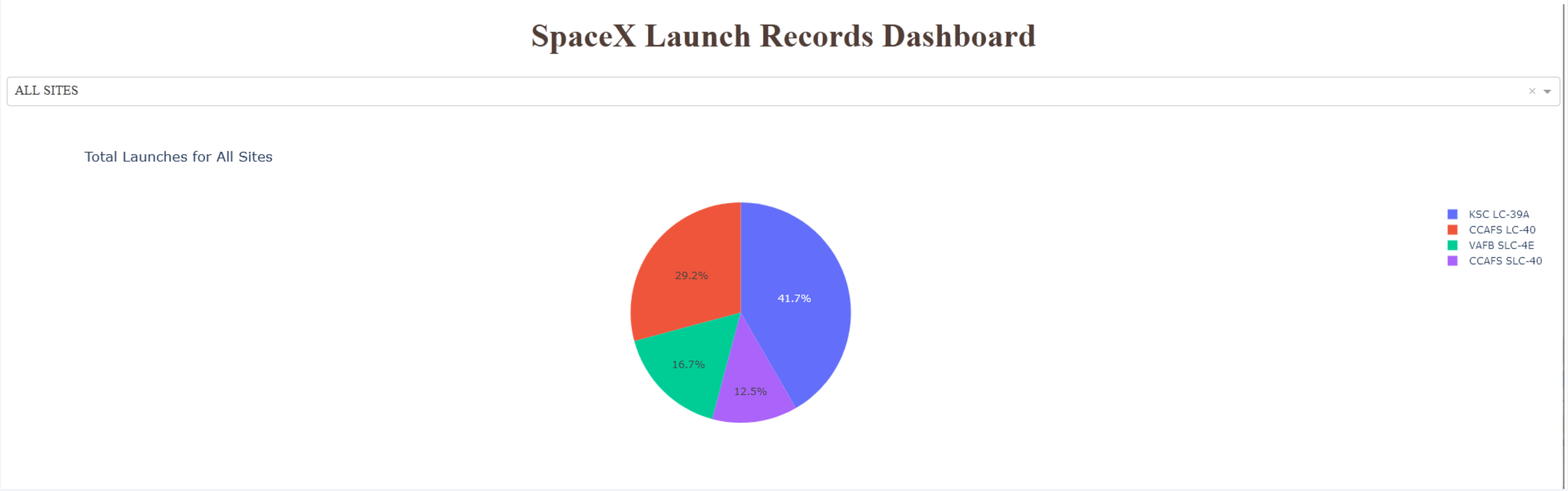


- The launch proximity of KSC LC-39A launch site to the nearest railway is 15.23 km.

Section 4

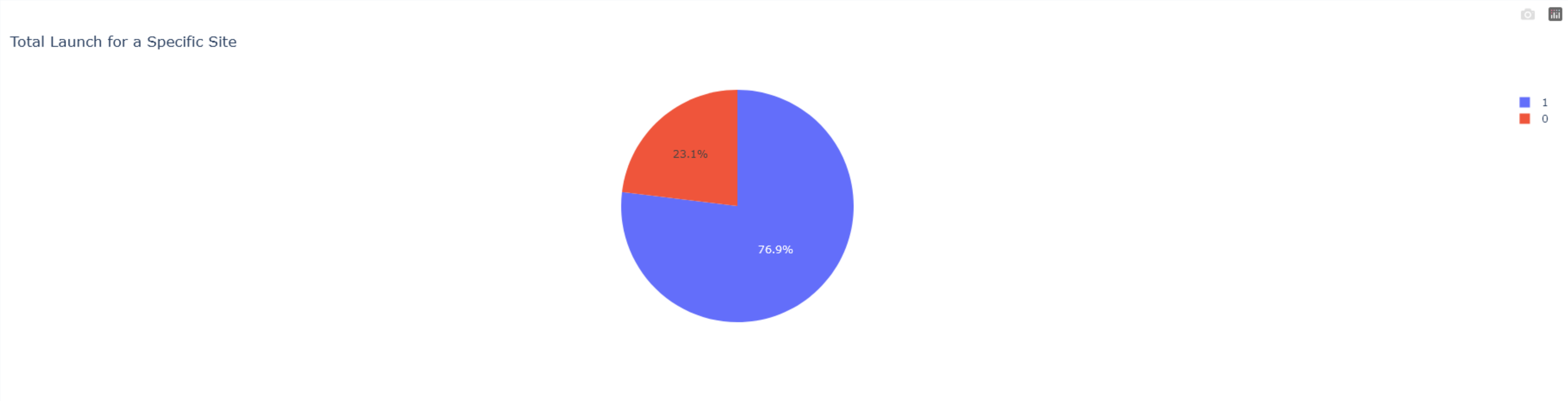
Build a Dashboard with Plotly Dash

Launches ratio per launch site



- As we can observe, KSC LC-39A has the higher launch rate with 41.7% of the total launches.

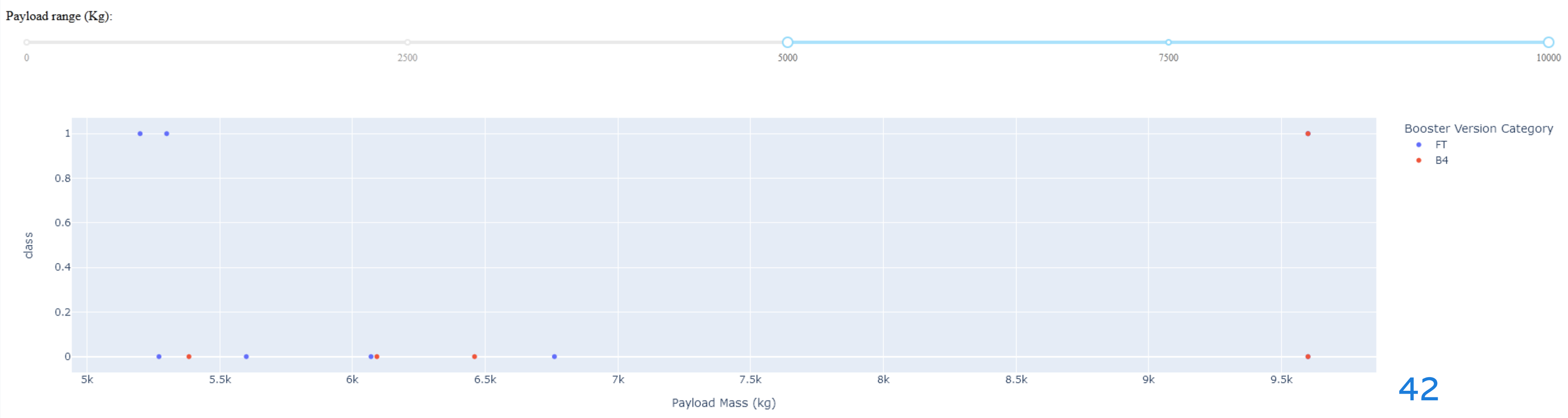
Launches success ratio for KSC LC-39A



- KSC LC-39A has a launch success probability of 76.9%.

Outcome with different payload

- Payloads mass between 2500 and 5000 for FT booster have higher success ratio.
- For payloads grater than 5000, the success ratio decreases.

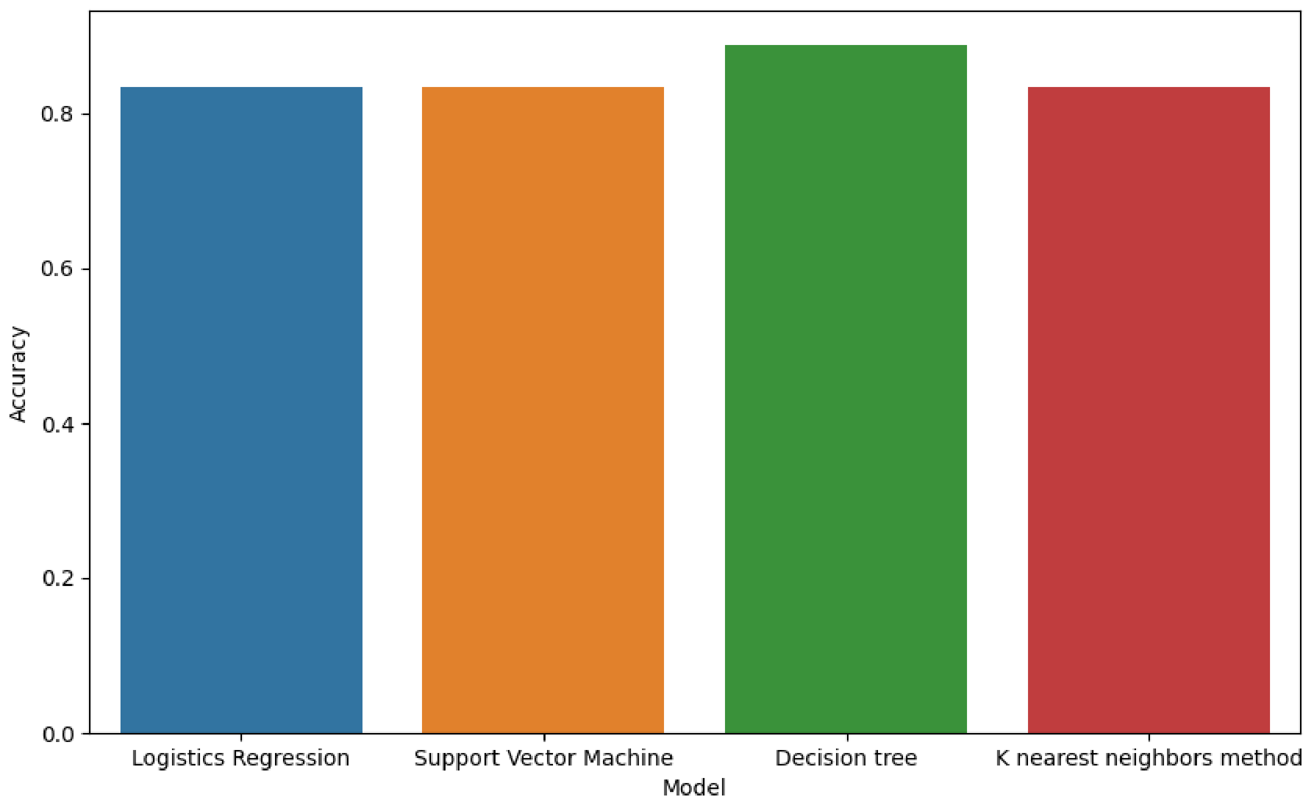


Section 5

Predictive Analysis (Classification)

Classification Accuracy

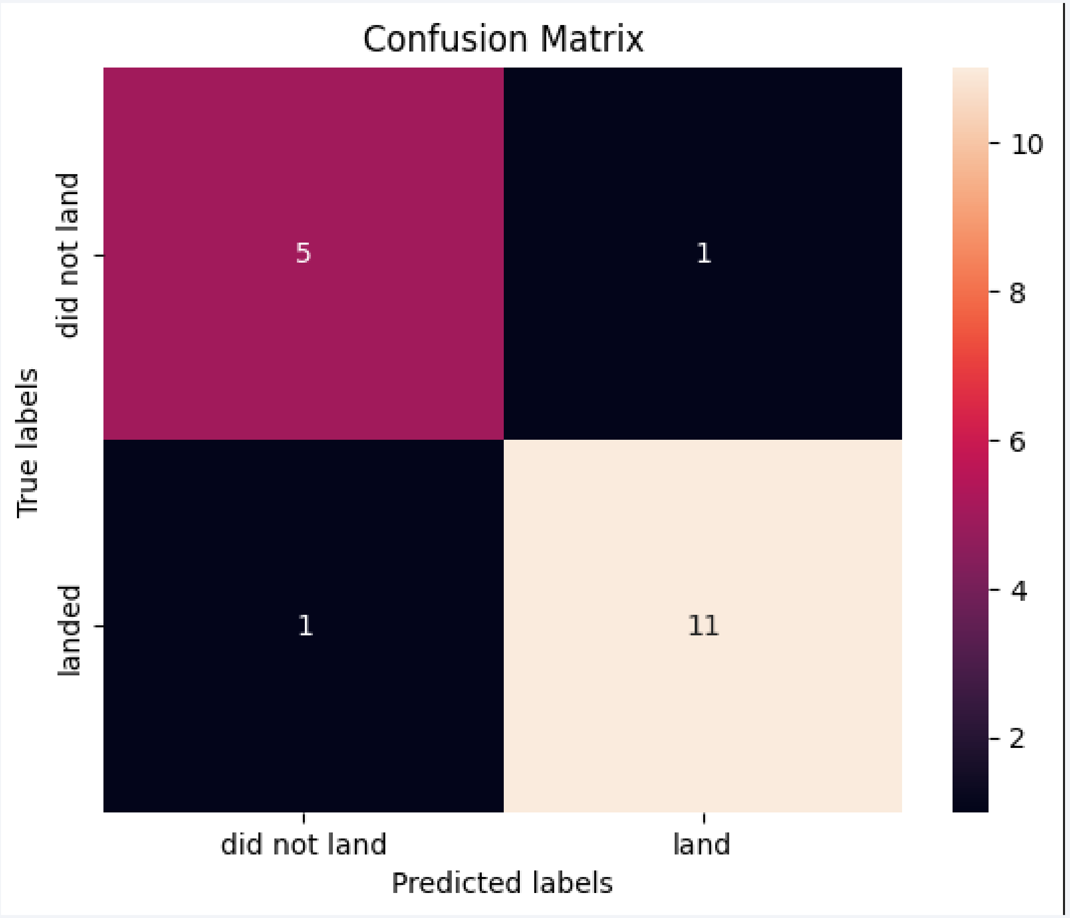
	Model	Accuracy
0	Logistics Regression	0.833333
1	Support Vector Machine	0.833333
2	Decision tree	0.888889
3	K nearest neighbors method	0.833333



- The best model is the Decision Tree with an accuracy of 88.8%.

Confusion Matrix for Decision Tree

- The model correctly predicts that 5 launches did not land and 11 launches land.



Conclusions

- KSC LC 39A is the most successful launch site with 76.9%.
- SSO and VLEO orbit have the best chance of a successful launch.
- Payloads lower than 4000 have a higher probability to success in SSO orbit.
- For orbit VLEO, payloads greater than 13000 have a higher probability to success.
- The best classification model to predict the outcome is Decision Tree.

Thank you!