IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Daniel Eduardo López
July 28, 2022

Outline

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

Executive Summary

- By using Python 3 and its libraries, historical data of **Falcon 9 launches from 2010-04-06 to 2020-06-12** was retrieved through calls to the SpaceX's API and web scraping to the Wikipedia's article on List of Falcon 9 and Falcon Heavy launches.

- Data was wrangled and explored through SQL queries, visualization and dashboarding, allowing to identify that SpaceX has a **success rate** of about **80%**; **KSC LC-39A** is the launch site with the highest number of successful missions, whereas **VAFB SLC-4E** is the one with the lowest; **ES-L1**, **GEO**, **HEO** and **SSO** were the orbit types with the highest success rates (100%) but with a small number of missions; **VLEO** was the most common orbit type in recent years; the **payload range** with the highest launch success rate was from 1952 kg to 5300 kg; and **FT** and **B5** are the **booster versions** with the highest launch success rate.

Executive Summary

- Finally, several classification models using Logistic Regression, Support Vector Machines, Decision Trees and K-Nearest Neighbors were set up and tuned, allowing to conclude that most of the launches from SpaceX **will land successfully**, thus suggesting that the **cost of a Falcon 9 rocket launch** should be set in **62 million dollars**.

Introduction

- SpaceX offers Falcon 9 rocket launches with a cost of 62 million dollars while other providers cost upward of 165 million dollars each. A significant amount of the savings is due to SpaceX's ability to reuse the first stage.
- In this context, the main goal of the present study is to **predict the first stage landing of the SpaceX Falcon 9 rocket launch** by using a classification model in order to determine the cost of a launch.
- Furthermore, it is also desirable to obtain other insights from the SpaceX Falcon 9 rocket launches.

Section 1

Methodology

Methodology

Executive Summary

- **Python 3** and its libraries were used in the entire project.
- First, **data was collected** through REST calls to the SpaceX API using Requests and by performing web scraping to the Wikipedia's article on List of Falcon 9 and Falcon Heavy launches using BeatifulSoup. Data covers a time period **from 2010-04-06 to 2020-06-12**.
- Then, a process of **data wrangling** was performed using Pandas and Numpy in order to filter the data to the Falcon 9 launches only, imputing missing values with the mean, creating the landing outcome labels; as well as selecting the features for modeling and obtaining dummy variables for the categorical variables.
- Next, an **exploratory data analysis (EDA)** was performed by means of visualization using Matplotlib and SQL.

Methodology

Executive Summary

- After that, an **interactive visual analysis** using Folium and Plotly Dash was carried out.
- Finally, a **predictive analysis** was performed by means of a Classification Model using Scipy and Scikit-learn. To do so:
  - Data was split in a training and test sets.
  - The model was built by using several machine learning techniques: Logistic Regression, Support Vector Machines (SVM), Decision Trees and K-Nearest Neighbors (KNN).
  - The models were tuned by using GridSearchCV, in which several parameters were tested, and the data was cross-validated in a 10-fold scheme.
  - The best model was selected based on the criteria of outcome of the confusion matrix, precision, recall, f1-score and accuracy.

Data Collection

- Data sets were collected through SpaceX API and Web Scraping in Python 3 by using the libraries Requests, Pandas, Numpy, BeautifulSoup, among others.

Figure 1. Data sources. Own elaboration.

Data Collection – SpaceX API

- The purpose of this step was to make a get request to the SpaceX API and clean the requested data.
- **Link of the data source:**
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- **GitHub URL:**
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/1-Spacex-data-collection-api.ipynb

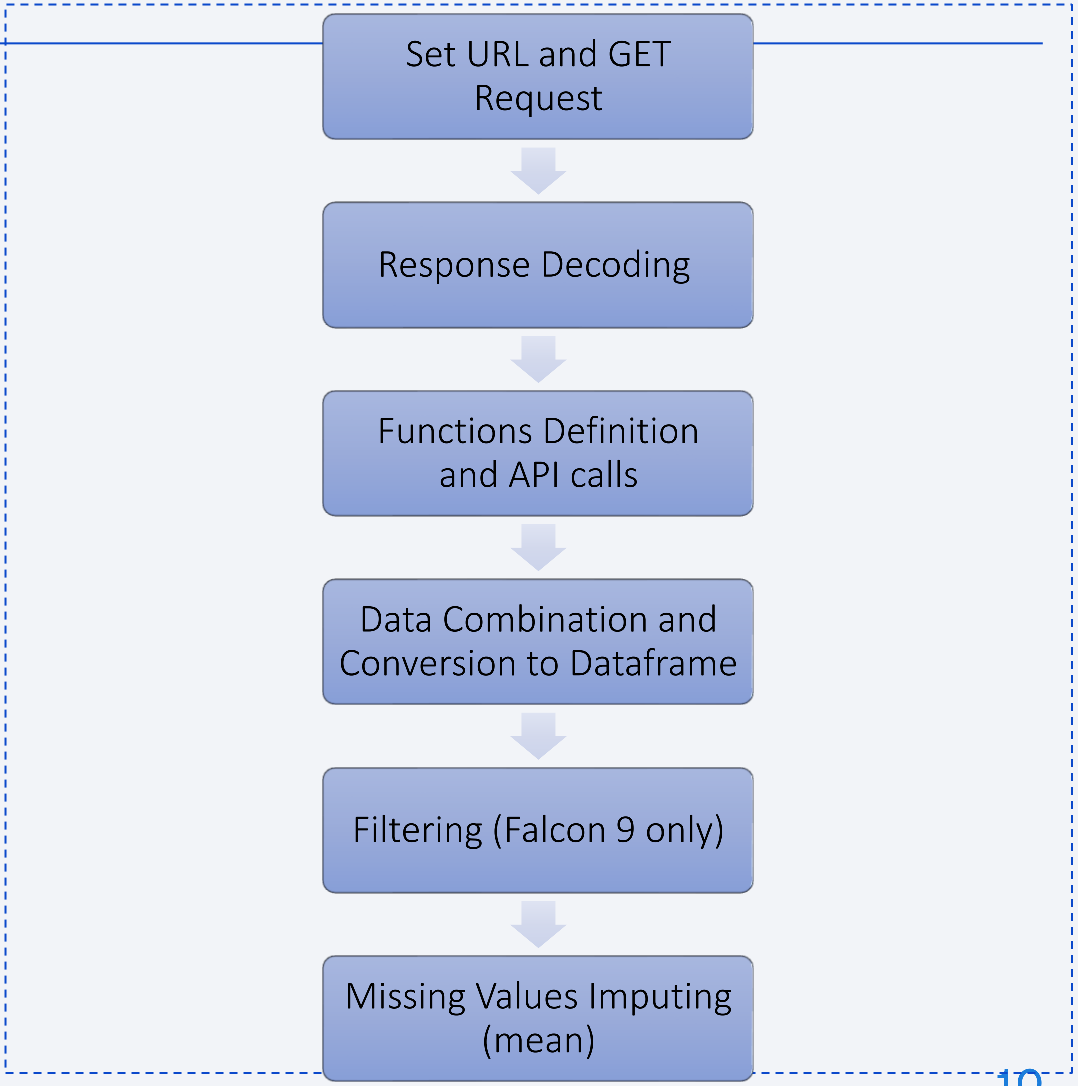


Figure 2. Data collection process for SpaceX API. Own elaboration.

Data Collection - Scraping

- The purpose of this step was to extract the Falcon 9 launch records from Wikipedia.
- **Link of the data source:**
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- **GitHub URL:**
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/2-Webscraping.ipynb

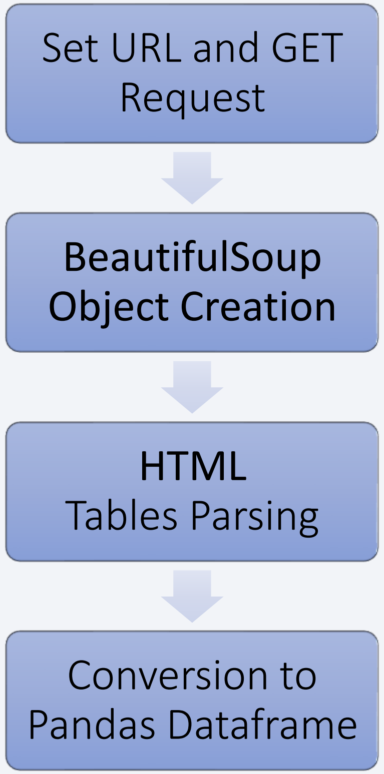


Figure 3. Data collection process for web scraping. Own elaboration.

Data Wrangling

- The purpose of this step was to find some patterns in the data and determine the label for training supervised models.

- **GitHub URL:**
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/3-Spacex_Data_wrangling.ipynb

Missing Values Identification

↓

Column Types Identification

↓

Number of Launches per Site

↓

Number and occurrence of each orbit

↓

Number and occurence of mission outcome per orbit type

↓

Landing Outcome Label Creation

Figure 4. Data wrangling process. Own elaboration.

EDA with Data Visualization

- The purpose of this step was to perform Exploratory Data Analysis with Pandas and Matplotlib.
- The following charts were plotted:
  - **Scatter Plot of "Flight Number" vs. "Payload Mass":** To assess the likelihood of the first stage return in function of the Payload Mass and the Flight Number.
  - **Scatter Plot of "Flight Number" vs. "Launch Site":** To assess the relationship among Launch Site, Flight Number, and the first stage return.
  - **Scatter Plot of "Payload Mass" vs. "Launch Site":** To assess the relationship among Launch Site, Payload Mass and the first stage return.

EDA with Data Visualization

- **Bar chart of the "Orbit" "Success Rate":** To assess the sucess rate of each orbit.
- **Scatter Plot of "Flight Number" vs. "Orbit":** To assess the relationship among Flight Number, Orbit, and the first stage return.
- **Scatter Plot of "Payload Mass" vs. "Orbit":** To assess the relationship among Payload Mass, Orbit, and the first stage return.
- **Line Chart of "Payload Mass" vs. "Orbit":** To visualize the launch success yearly trend.

GitHub URL: https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/5-EDA_dataviz.ipynb

Feature Engineering

- In addition, Feature Engineering was carried out to select the features that were used in the classification model. The following ones were selected:
- `'FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial'`
- Then, dummy variables were created for the categorical variables `Orbits`, `LaunchSite`, `LandingPad`, and `Serial` using the **get_dummies()** function.
- Finally, the whole dataset was casted to float64 datatype.
- **GitHub URL:** https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/5-EDA_dataviz.ipynb

EDA with SQL

- Furthermore, the following SQL queries were performed after loading the dataset into a corresponding table in a database using the libraries Sqlite3 and Sqlalchemy:
  - `SELECT Launch_Site, COUNT(Launch_Site) AS Count FROM SPACEXTBL GROUP BY Launch_Site;`
  - `SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;`
  - `SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';`
  - `SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';`
  - `SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (ground pad)';`

EDA with SQL

- `SELECT Booster_Version FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;`
- `SELECT Mission_Outcome, COUNT(Payload) FROM SPACEXTBL GROUP BY Mission_Outcome;`
- `SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);`
-
```
SELECT SUBSTR(Date,1,4) AS Year, COUNT("Landing
_Outcome") AS Successful_landing_outcomes
FROM SPACEXTBL
WHERE DATE > '2010-06-04' AND DATE < '2017-03-20' AND "Landing
_Outcome" LIKE 'Success%'
GROUP BY SUBSTR(Date,1,4)
ORDER BY SUBSTR(Date,1,4) DESC;
```

EDA with SQL

```
SELECT DISTINCT (CASE when SUBSTR(Date, 6, 2) = '01' then 'January'
                      when SUBSTR(Date, 6, 2) = '02' then 'February'
                      when SUBSTR(Date, 6, 2) = '03' then 'March'
                      when SUBSTR(Date, 6, 2) = '04' then 'April'
                      when SUBSTR(Date, 6, 2) = '05' then 'May'
                      when SUBSTR(Date, 6, 2) = '06' then 'June'
                      when SUBSTR(Date, 6, 2) = '07' then 'July'
                      when SUBSTR(Date, 6, 2) = '08' then 'August'
                      when SUBSTR(Date, 6, 2) = '09' then 'September'
                      when SUBSTR(Date, 6, 2) = '10' then 'October'
                      when SUBSTR(Date, 6, 2) = '11' then 'November'
                      when SUBSTR(Date, 6, 2) = '12' then 'December'
                 END) AS Month, "Landing _Outcome", Booster_Version,
Launch_Site FROM SPACEXTBL WHERE SUBSTR(Date,1,4)='2015' AND "Landing
_Outcome" = 'Failure (drone ship)’;
```

- **GitHub URL:** https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/4-EDA_sql_sqllite.ipynb

Build an Interactive Map with Folium

- An Interactive Map with the Folium library was created, and the following maps objects were created and added to the map:
  - **Markers and circles for all launch sites**: To visualize the location of the launch sites and assess their characteristics.
  - **Marker Clusters for success/failed launches for each site**: To assess if there was a relationship among the launch sites and their success/fail missions.
  - **Markers and lines between a launch site to its proximities**: To assess the characteristics of the proximities to the launch sites.

- **GitHub URL:** https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/6-Folium_Launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- An Interactive Dashboard with the Plotly and Dash libraries was built, and the following plots/graphs and interactions were added to the dashboard:
  - **Dropdown**: In order to enable the Launch Site selection for the user.
  - **Pie chart:** To show the Total Success Launches by Site.
  - **Range Slider:** To enable the Payload Mass selection.
  - **Scatter Plot:** To show the relationship among Payload Mass and mission outcomes for the selected Sites.

- **GitHub URL:**
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/7-spacex_dash_app.py
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/8-SpaceX_Dashboard.ipynb

Predictive Analysis (Classification)

- The purpose of this step was to build a Classification model for predicting the landing of the first phase; as well as finding the best hyperparameters and the best model.

- **GitHub URL**: https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/9-SpaceX_Machine%20Learning%20Prediction.ipynb

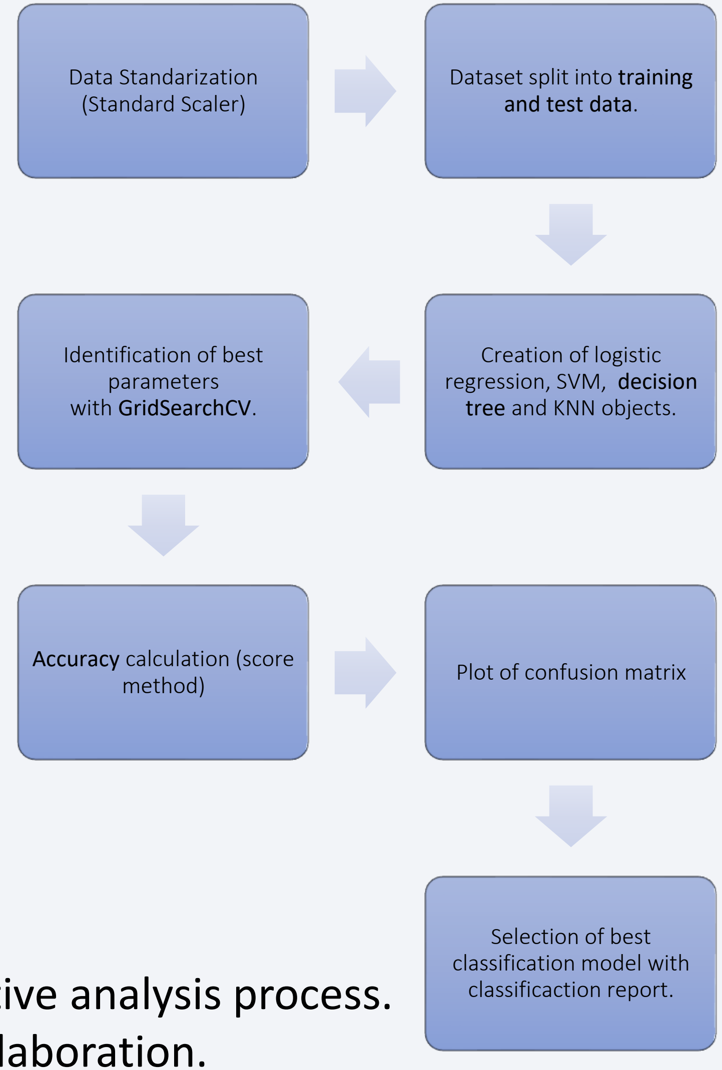


Figure 5. Predictive analysis process. Own elaboration.

Results

- Exploratory data analysis results

Table 1. Launch Site Counting and Success Rate.

Launch Site	Count	Success Rate
CCAFS SLC 40	55	60.00%
KSC LC 39A	22	77.27%
VAFB SLC 4E	13	76.92%

Table 2. Successful Landing Counting from 2015-2017.

Year	Successful Landing Outcomes
2017	4
2016	5
2015	1

Table 3. Orbit Type Counting and Success Rate.

Orbit	Count	Success Rate
GTO	27	51.85%
ISS	21	61.90%
VLEO	14	85.71%
PO	9	66.67%
LEO	7	71.43%
SSO	5	100.00%
MEO	3	66.67%
ES-L1	1	100.00%
HEO	1	100.00%
SO	1	0.00%
GEO	1	100.00%

Results

- Exploratory data analysis results

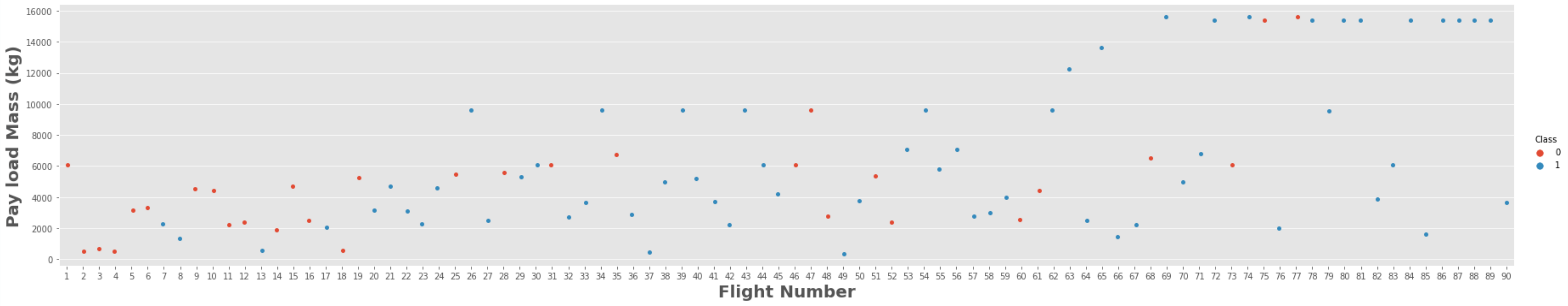


Figure 6. Flight Number vs. Payload Mass. Own elaboration.

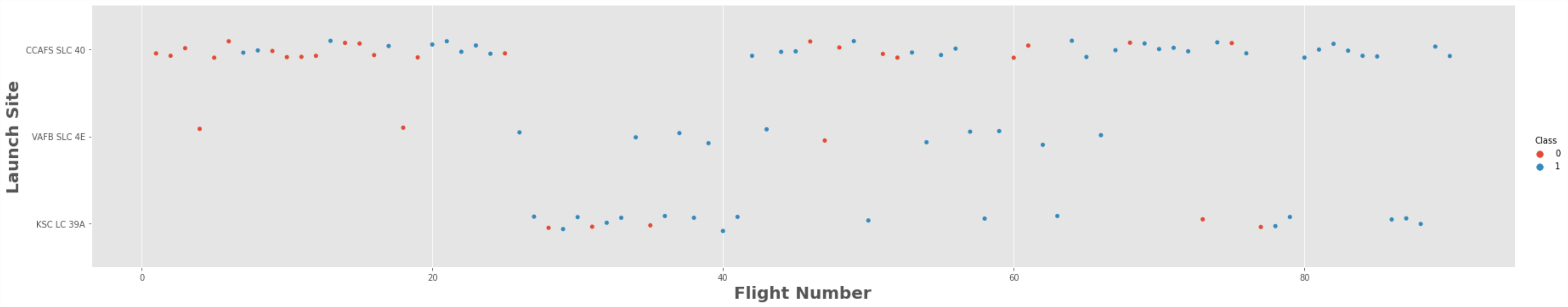


Figure 7. Flight Number vs. Launch Site. Own elaboration.

Results

- Exploratory data analysis results

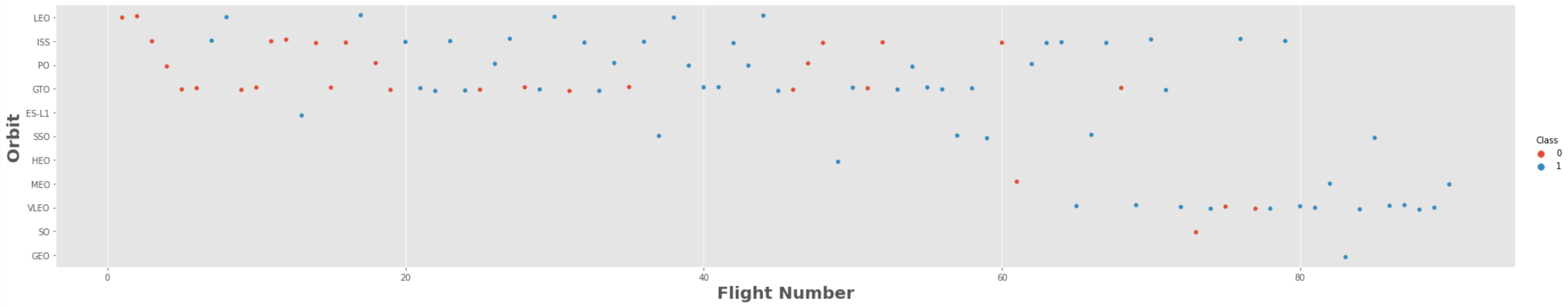


Figure 8. Flight Number vs. Orbit. Own elaboration.

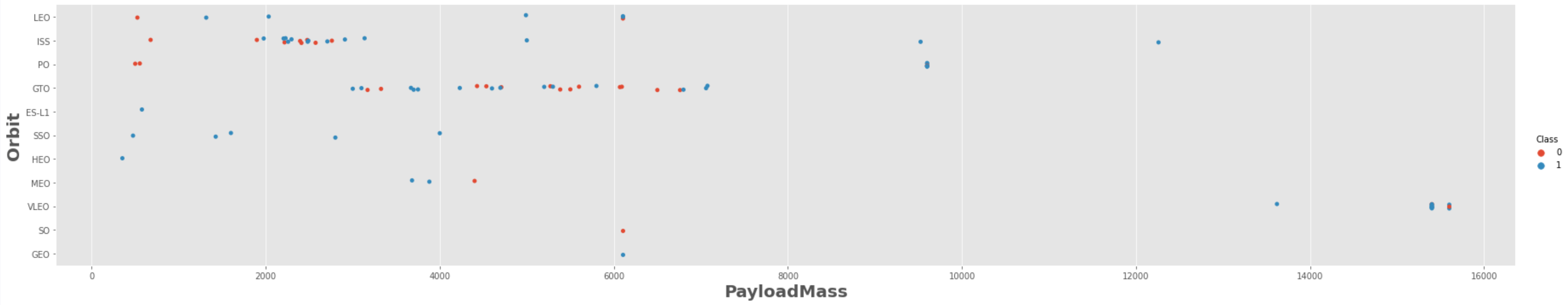


Figure 9. Payload Mass vs. Orbit. Own elaboration.

Results

- Exploratory data analysis results

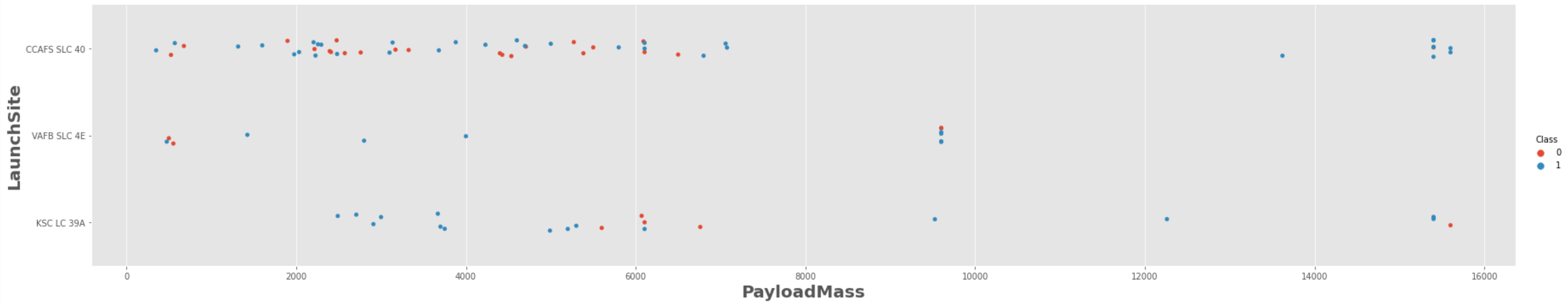
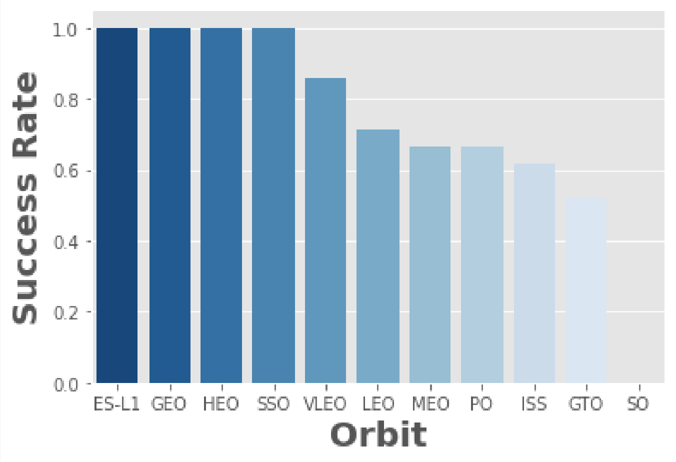


Figure 10. Payload Mass vs. Launch Site. Own elaboration.



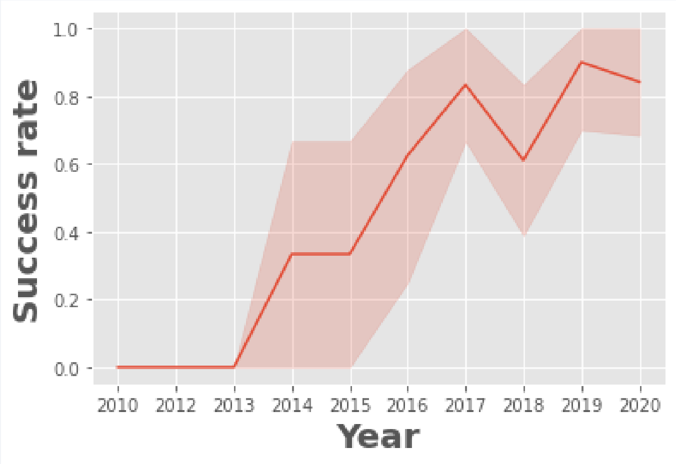


Figure 11. Orbit Success Rate. Own elaboration.

Figure 12. Success Rate Over Time. Own elaboration.

Results

- Interactive analytics demo.

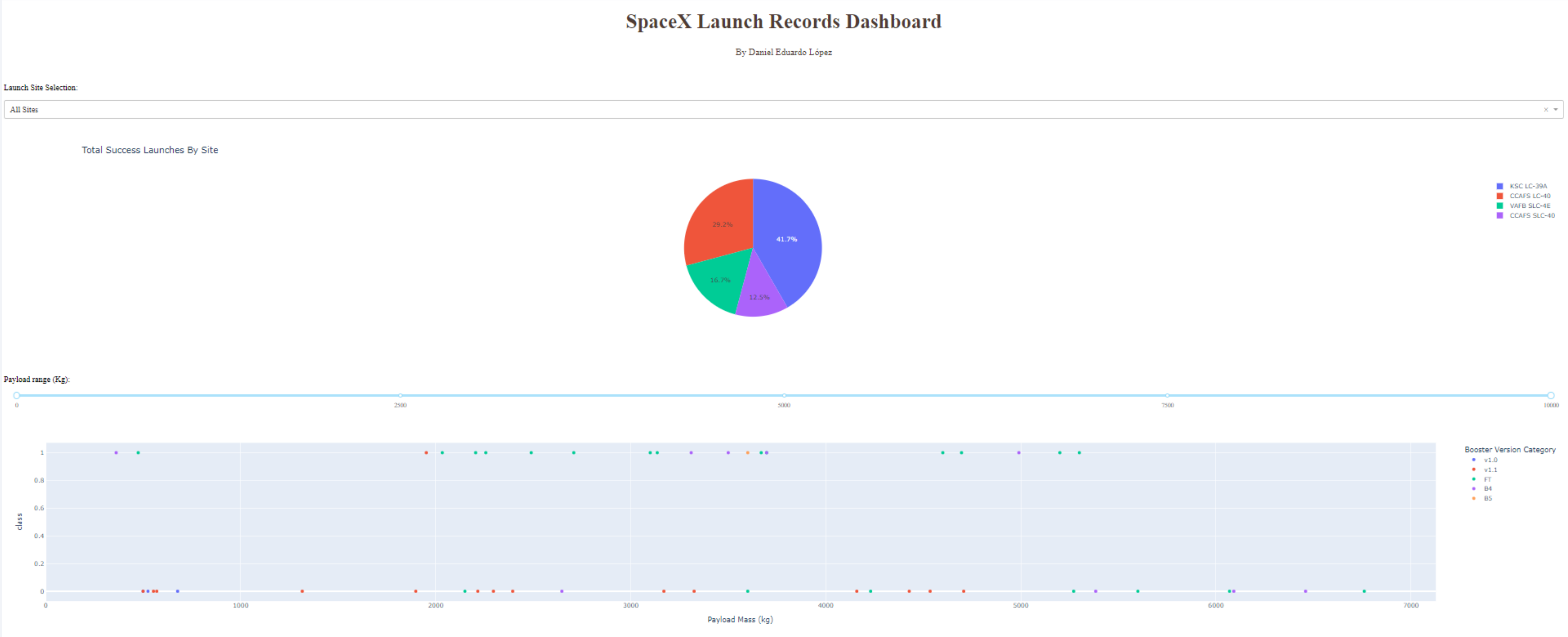


Figure 13. Dashboard overview. Own elaboration.

Results

- Interactive analytics demo.

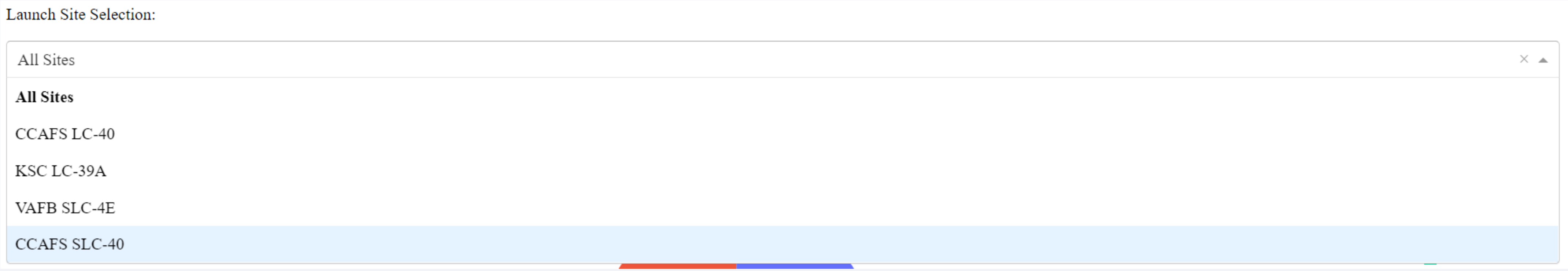


Figure 14. Dashboard's dropdown. Own elaboration.

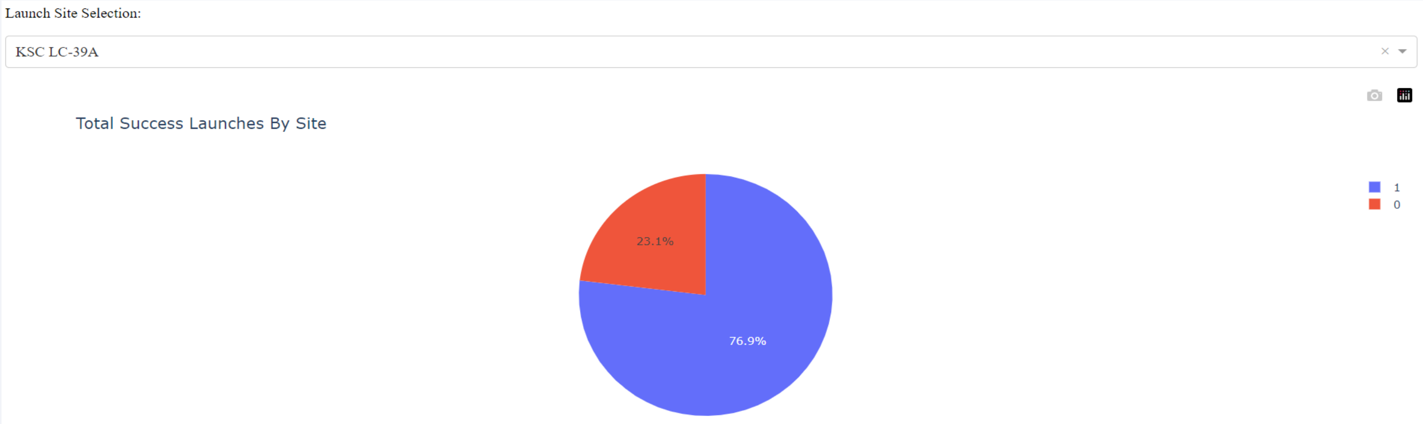


Figure 15. Dashboard's Pie Chart with Success Rate per Site. Own elaboration.

Results

- Interactive analytics demo.

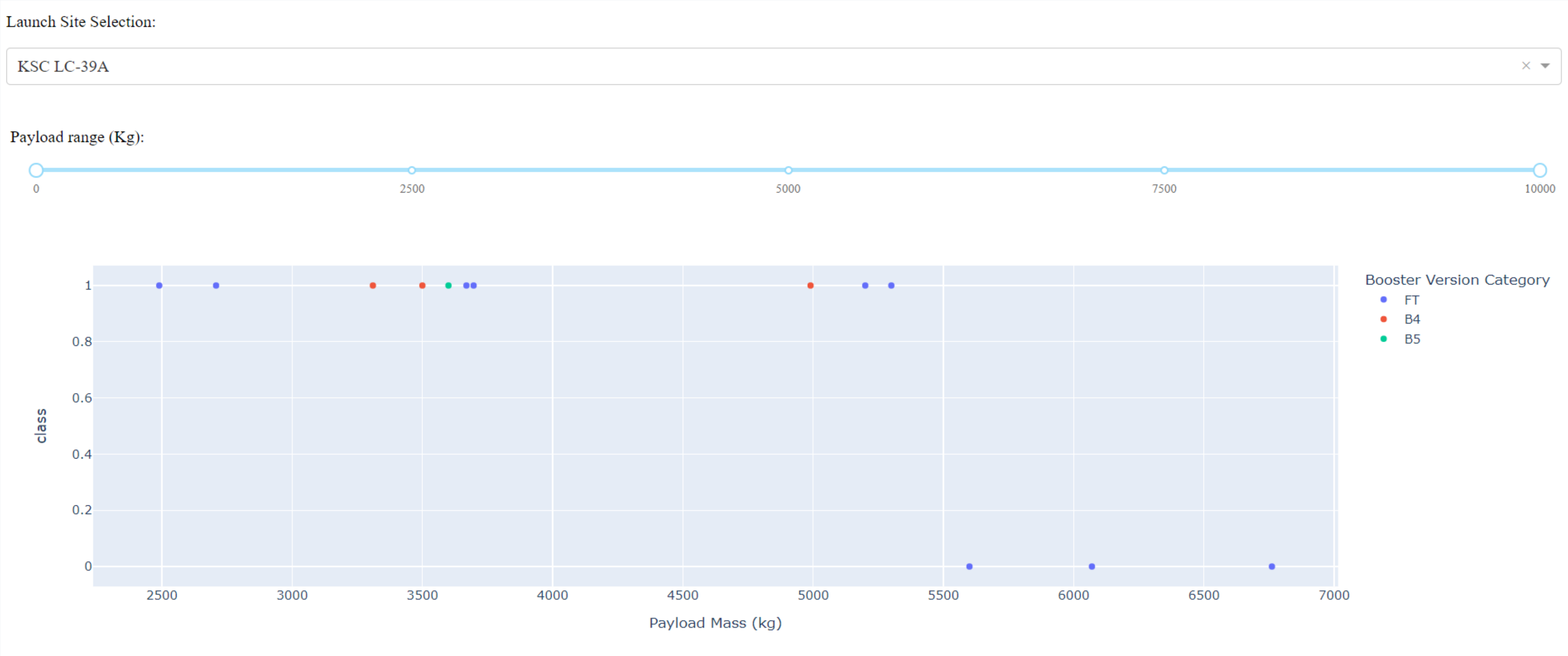


Figure 16. Dashboard's Scatter Plot of Payload Mass vs Class with Range Slider for Payload Mass. Own elaboration.

Results

- Predictive analysis results

Figure 17. Logistic regression confusion matrix. Own elaboration.

Figure 18. SVM confusion matrix. Own elaboration.

Results

- Predictive analysis results

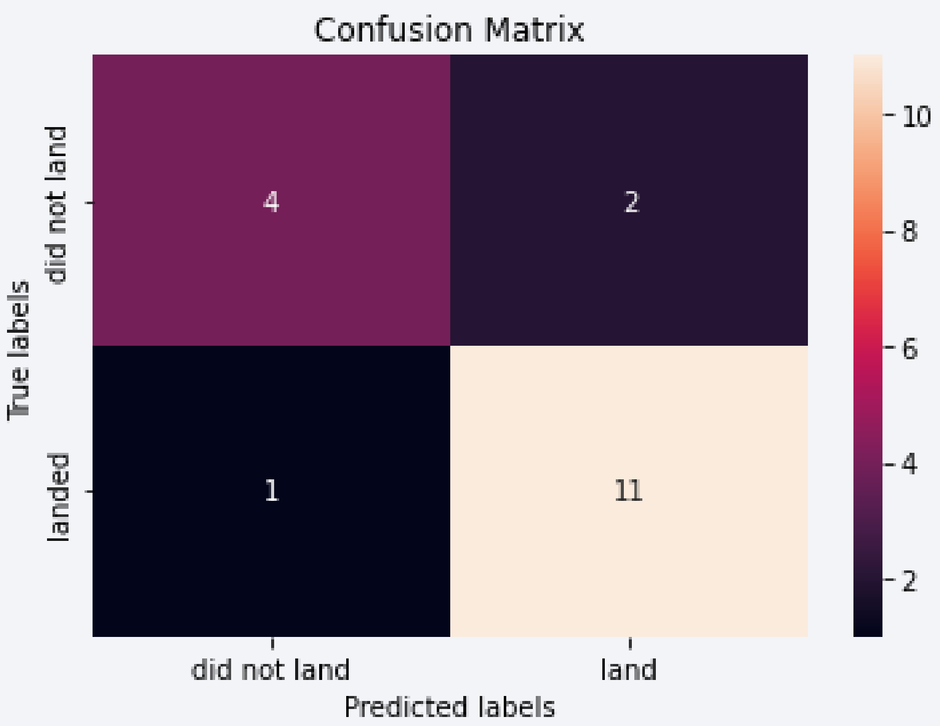


Figure 19. Decision trees confusion matrix. Own elaboration.

Figure 20. KNN confusion matrix. Own elaboration.

Results

- Predictive analysis results

The Accuracy, Recall & F1-score score of the Logistic Regression is:

	precision	recall	f1-score	support
0	1.00	0.50	0.67	6
1	0.80	1.00	0.89	12
accuracy			0.83	18
macro avg	0.90	0.75	0.78	18
weighted avg	0.87	0.83	0.81	18

The Accuracy, Recall & F1-score score of the SVM Classification is:

	precision	recall	f1-score	support
0	1.00	0.50	0.67	6
1	0.80	1.00	0.89	12
accuracy			0.83	18
macro avg	0.90	0.75	0.78	18
weighted avg	0.87	0.83	0.81	18

Figure 21. Logistic regression and SVM models classification reports. Own elaboration.

Results

- Predictive analysis results

The Accuracy, Recall & F1-score score of the Decision Tree Classification is:

	precision	recall	f1-score	support
0	1.00	0.50	0.67	6
1	0.80	1.00	0.89	12
accuracy			0.83	18
macro avg	0.90	0.75	0.78	18
weighted avg	0.87	0.83	0.81	18

The Accuracy, Recall & F1-score score of the KNN Classification is:

	precision	recall	f1-score	support
0	1.00	0.50	0.67	6
1	0.80	1.00	0.89	12
accuracy			0.83	18
macro avg	0.90	0.75	0.78	18
weighted avg	0.87	0.83	0.81	18

Figure 22. Decision trees and KNN models classification reports. Own elaboration.

Section 2

Insights drawn from EDA

Launch Site vs. Flight Number

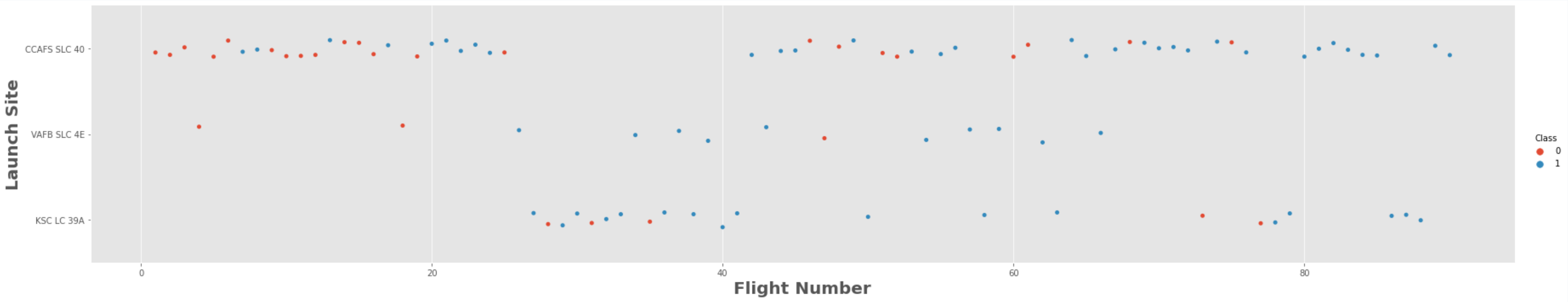


Figure 7. Launch Site vs. Flight Number. Own elaboration.

- **CCAFS SLC 40** is the most used Launch Site and the **VAFB SLC-4E** is the least used one.
- The success rate has improved over the years as represented by the Flight number.
- **KSC LC-39A** is the launch site with the highest number of successful missions.

Launch Site vs. Payload

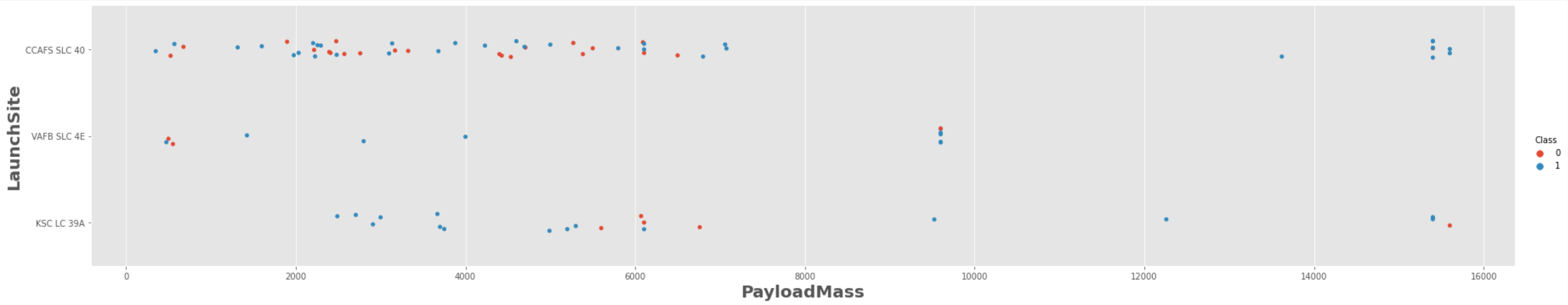


Figure 10. Launch Site vs. Payload. Own elaboration.

- **CCAFS SLC 40** and **KSC LC-39A** are used for a broad range of payload masses, from light to extra heavy ones.
- **VAFB SLC-4E** is only used for light and medium payloads, which might explain why is the less used launch site.
- The missions tend to be more successful with heavier payloads.

Success Rate vs. Orbit Type

- **ES-L1, GEO, HEO** and **SSO** are the orbit types with the **highest success rates** (100%).
- **SO** is the orbit type with the **lowest success rate** (0%).
- The rest of the orbit types have a success rate ranging **from 50% to 85%.**

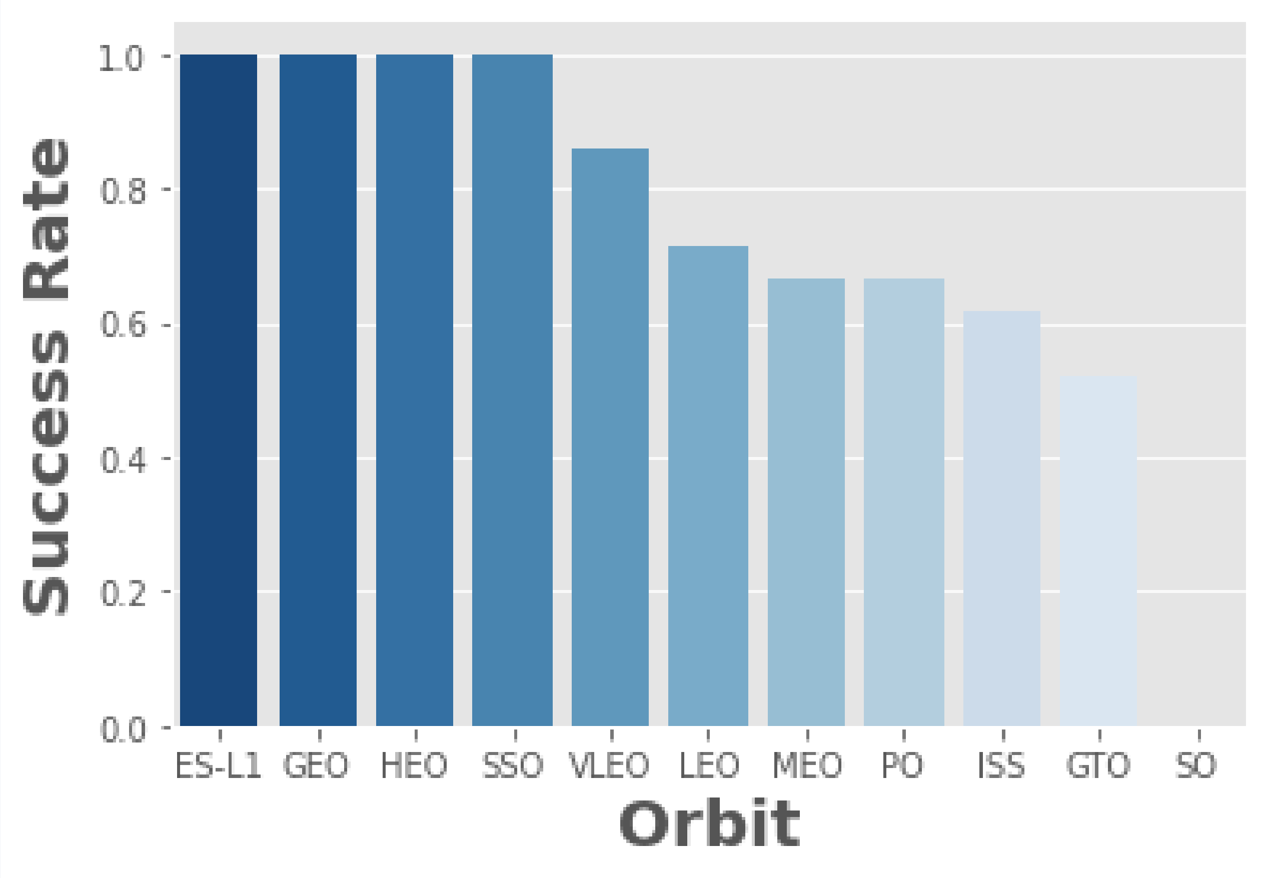


Figure 11. Orbit Success Rate. Own elaboration.

Orbit Type vs. Flight Number

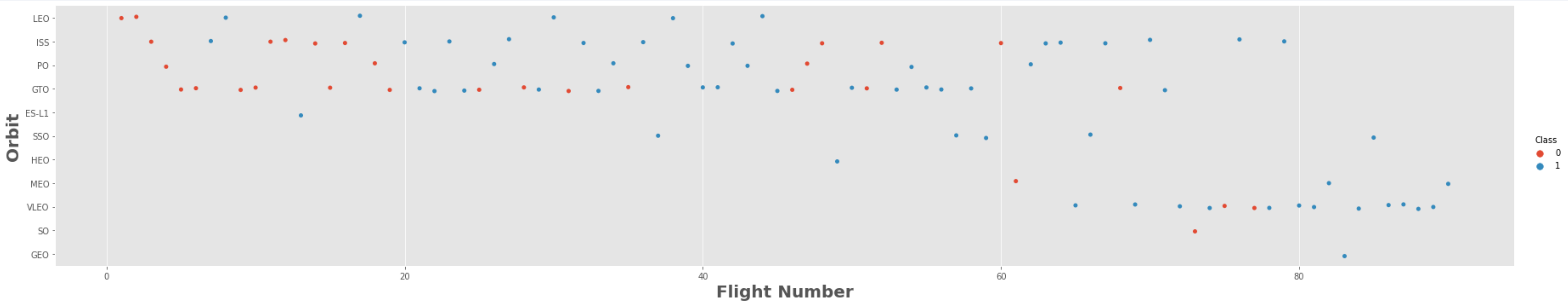


Figure 8. Orbit Type vs. Flight Number. Own elaboration.

- **LEO, ISS, PO, GTO** and **VLEO** are the most common orbit types.
- **EO, PO, GTO** and **VLEO** are less common in recent flight numbers.
- **VLEO** is the most common orbit type in recent flight numbers.
- Earlier flight numbers show a tendency to fail **regardless of the orbit type**.

Orbit Type vs. Payload

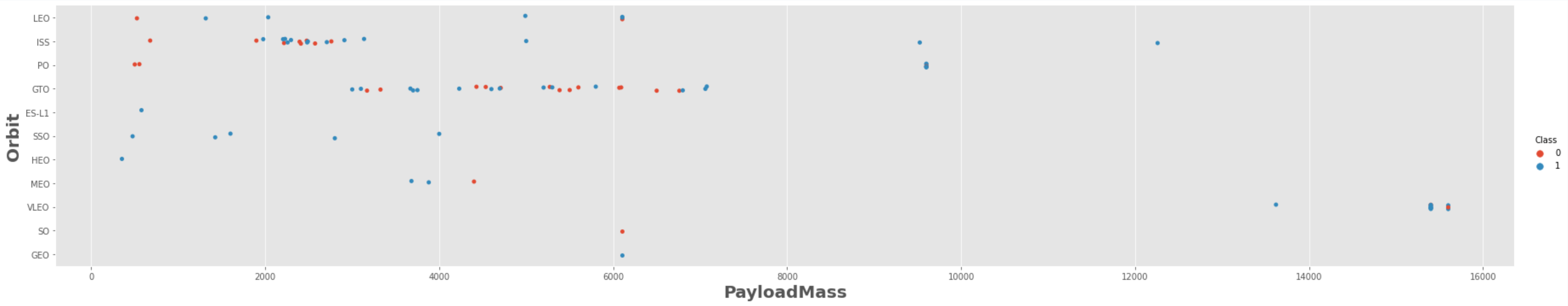


Figure 9. Orbit Type vs. Payload. Own elaboration.

- Heavier payloads are selected for **VLEO, IS** and **PO** orbit types and have a high rate of success.
- It appears that **ES-L1, SSO** and **HEO** are used with light payloads (from 0 to 4000 kg) and have a high rate of success.
- The rest of the orbit types are used with **light and medium payloads (from 0 to 7000 kg).**

Launch Success Yearly Trend

- Definitely, the launch success rate have **improved over the years.**
- In particular, the success rate have improved since 2013, from 0% to an 80% in 2017.
- Since 2017, it appears that the success rate has stabilized in **about 80%.**

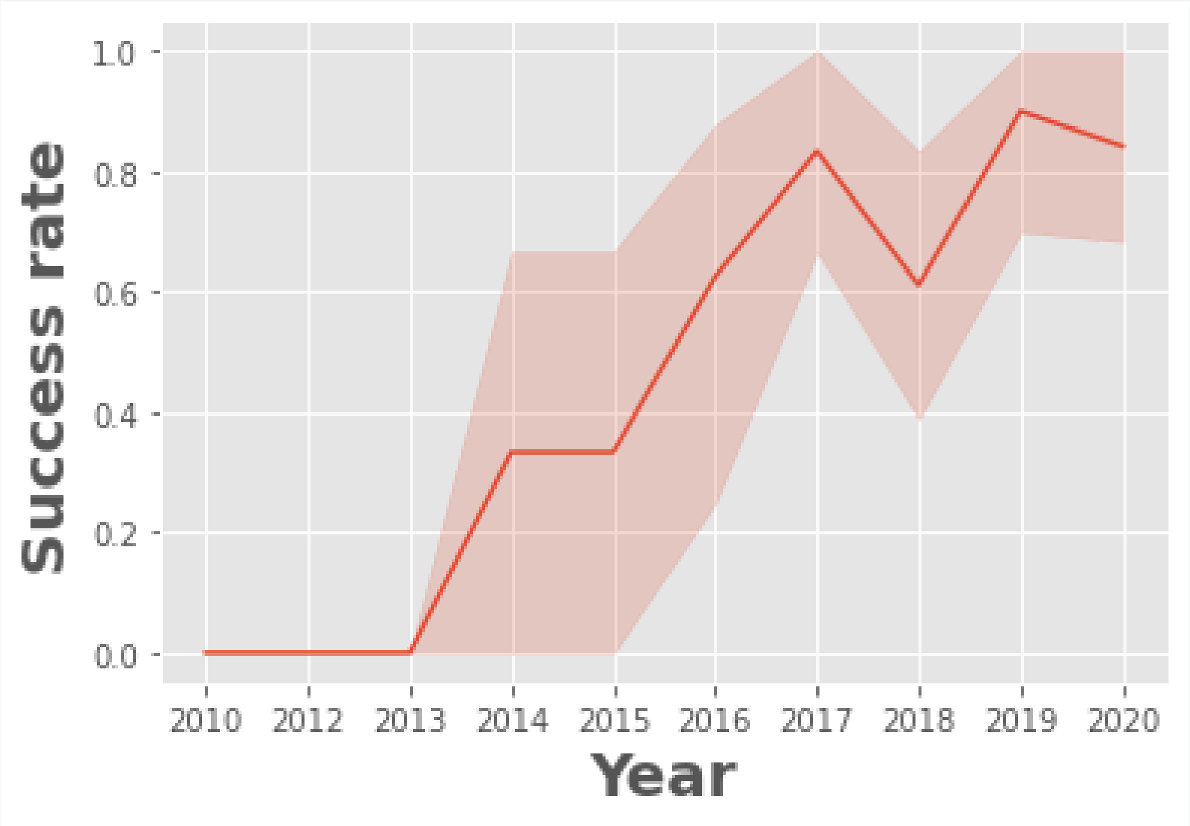


Figure 12. Success Rate Over Time. Own elaboration.

All Launch Site Names

- **CCAFS SLC-40** is the most used Launch Site and the **VAFB SLC-4E** is the least used.

```
In [ ]: %%sql

SELECT Launch_Site, COUNT(Launch_Site) AS Count FROM SPACEXTBL GROUP BY Launch_Site;

 * sqlite:///my_data1.db
Done.
Out[ ]:
```

Launch_Site	Count
CCAFS LC-40	26
CCAFS SLC-40	34
KSC LC-39A	25
VAFB SLC-4E	16

Figure 23. SQL Query with Launch Site Count. Own elaboration.

Launch Site Names Begin with 'CCA'

- The launch site whose name begins with the string 'CCA' is **CCAFS SLC-40**.

```
%%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
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

Figure 24. SQL Query with Launch Site whose name begins with 'CCA'. Own elaboration.

Total Payload Mass

- The total payload carried by boosters from NASA is **45 596 kg.**

```
In [ ]: %%sql

        SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';

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

Out[ ]:

SUM(PAYLOAD_MASS__KG_)
45596

Figure 25. SQL Query with Total Payload Carried by Booster from NASA. Own elaboration.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is **2534.66 kg**, which is in the range of the light payloads.

```
In [ ]: %%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

Figure 26. SQL Query with average payload mass carried by booster version F9 v1.1. Own elaboration.

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was achieved on **2015-12-22.**

```
In [ ]: %%sql

        SELECT MIN(Date) FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (ground pad)';

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

Out[ ]:

MIN(Date)
2015-12-22

Figure 27. SQL Query with first successful landing outcome on ground pad. Own elaboration.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are **F9 FT B1022, F9 FT B1026, F9 FT B1021.2 and F9 FT B1031.2.**

```
In [ ]: %%sql

        SELECT Booster_Version FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

         * sqlite:///my_data1.db
        Done.
Out[ ]:
```

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

Figure 28. SQL Query with names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000. Own elaboration.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are **100 and 1**, respectively.

In [ ]:
```
%%sql

SELECT Mission_Outcome, COUNT(Payload) FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

Out[ ]:

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

Figure 29. SQL Query with total number of successful and failure mission outcomes. Own elaboration.

Boosters Carried Maximum Payload

- The names of the booster which have carried the **maximum payload mass** are shown in the image below:

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

In [ ]:

```
%%sql

SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

Out[ ]:

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

Figure 30. SQL Query with names of the booster which have carried the maximum payload mass. Own elaboration.

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 happened in **October and April.**

In [ ]:
```
%%sql
-- ,
SELECT DISTINCT (CASE when SUBSTR(Date, 6, 2) = '01' then 'January'
                      when SUBSTR(Date, 6, 2) = '02' then 'February'
                      when SUBSTR(Date, 6, 2) = '03' then 'March'
                      when SUBSTR(Date, 6, 2) = '04' then 'April'
                      when SUBSTR(Date, 6, 2) = '05' then 'May'
                      when SUBSTR(Date, 6, 2) = '06' then 'June'
                      when SUBSTR(Date, 6, 2) = '07' then 'July'
                      when SUBSTR(Date, 6, 2) = '08' then 'August'
                      when SUBSTR(Date, 6, 2) = '09' then 'September'
                      when SUBSTR(Date, 6, 2) = '10' then 'October'
                      when SUBSTR(Date, 6, 2) = '11' then 'November'
                      when SUBSTR(Date, 6, 2) = '12' then 'December'
                 END) AS Month, "Landing _Outcome", Booster_Version, Launch_Site FROM SPACEXTBL WHERE SUBSTR(Date,1,4)='2015' AND "Landing _Outcome" = 'Failure (drone ship)';
```

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

Out[ ]:

Month	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

Figure 31. SQL Query with failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015. Own elaboration.

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

- The rank of 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 is shown in the image below:

In [ ]:
```
%%sql

SELECT SUBSTR(Date,1,4) AS Year, COUNT("Landing _Outcome") AS Successful_landing_outcomes
FROM SPACEXTBL
WHERE DATE > '2010-06-04' AND DATE < '2017-03-20' AND "Landing _Outcome" LIKE 'Success%'
GROUP BY SUBSTR(Date,1,4)
ORDER BY SUBSTR(Date,1,4) DESC;
```

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

Out[ ]:

Year	Successful_landing_outcomes
2017	4
2016	5
2015	1

Figure 32. SQL Query with rank of count of landing outcomes between the date 2010-06-04 and 2017-03-20. Own elaboration.

Section 3

Launch Sites Proximities Analysis

Launch Sites Localization

- Launch sites are located **on the coasts** and **close to the Ecuador.**

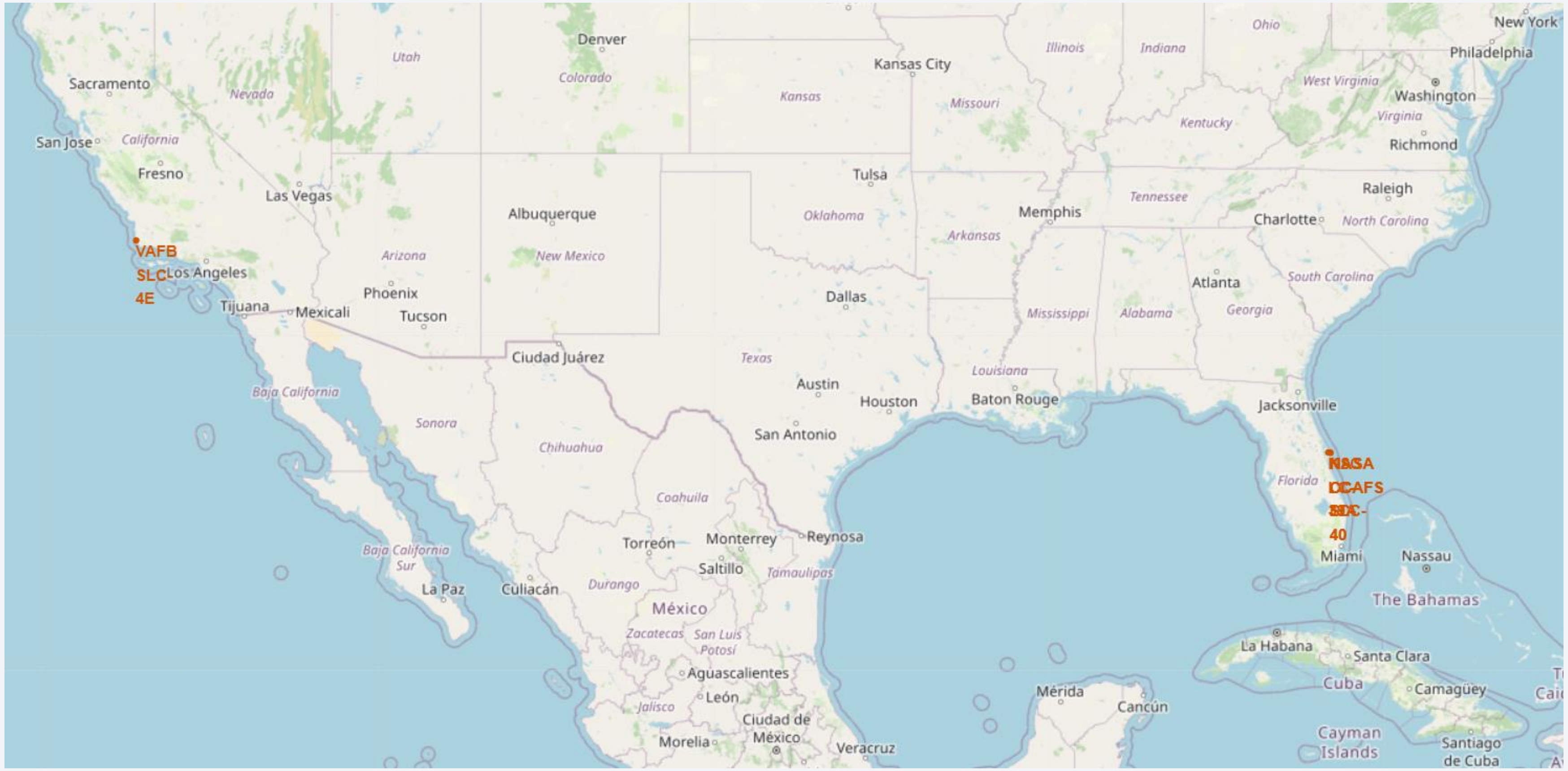


Figure 33. Map with the localization of launch sites. Own elaboration.

Launch Outcomes Per Launch Site

- The Vanderberg (**VAFB SLC-4E**) Launch Site is **the least used** and its number of successful landings is fairly low.

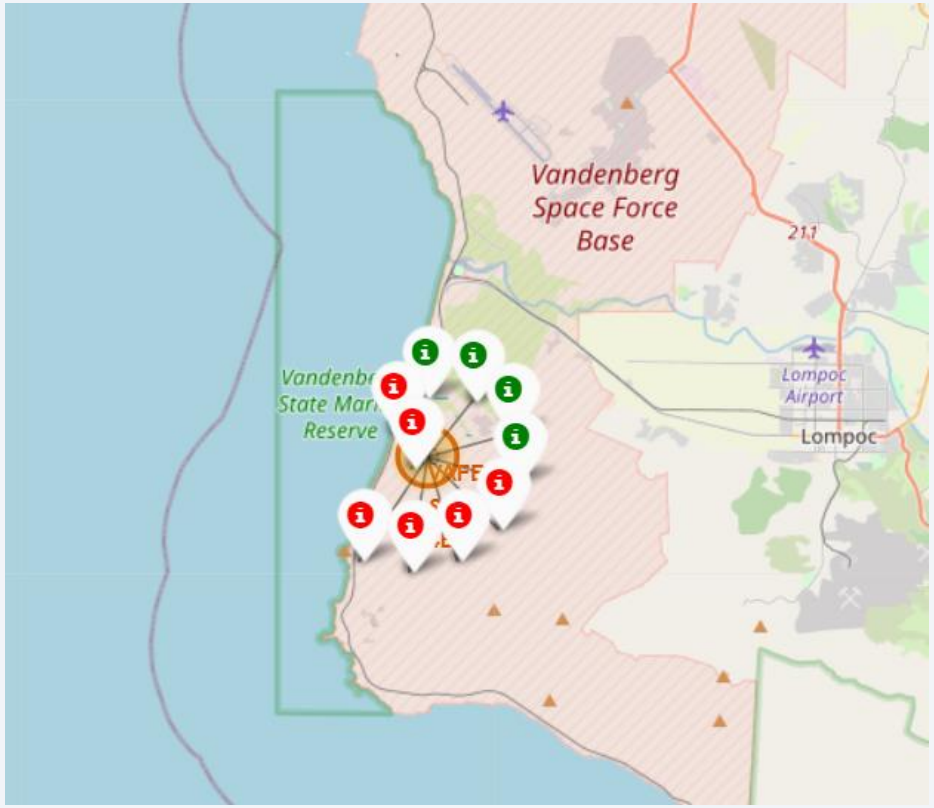


Figure 34. Map with the localization of VAFB SLC-4E launch site and marks of successful and unsuccessful landings. Own elaboration.

Launch Outcomes Per Launch Site

- The Cape Canaveral (**CCAFS LC-40 and CCAFS SLC-40**) is **the most used** Launch Site, even though its number of failed landings is slightly superior to the its number of successful landings.

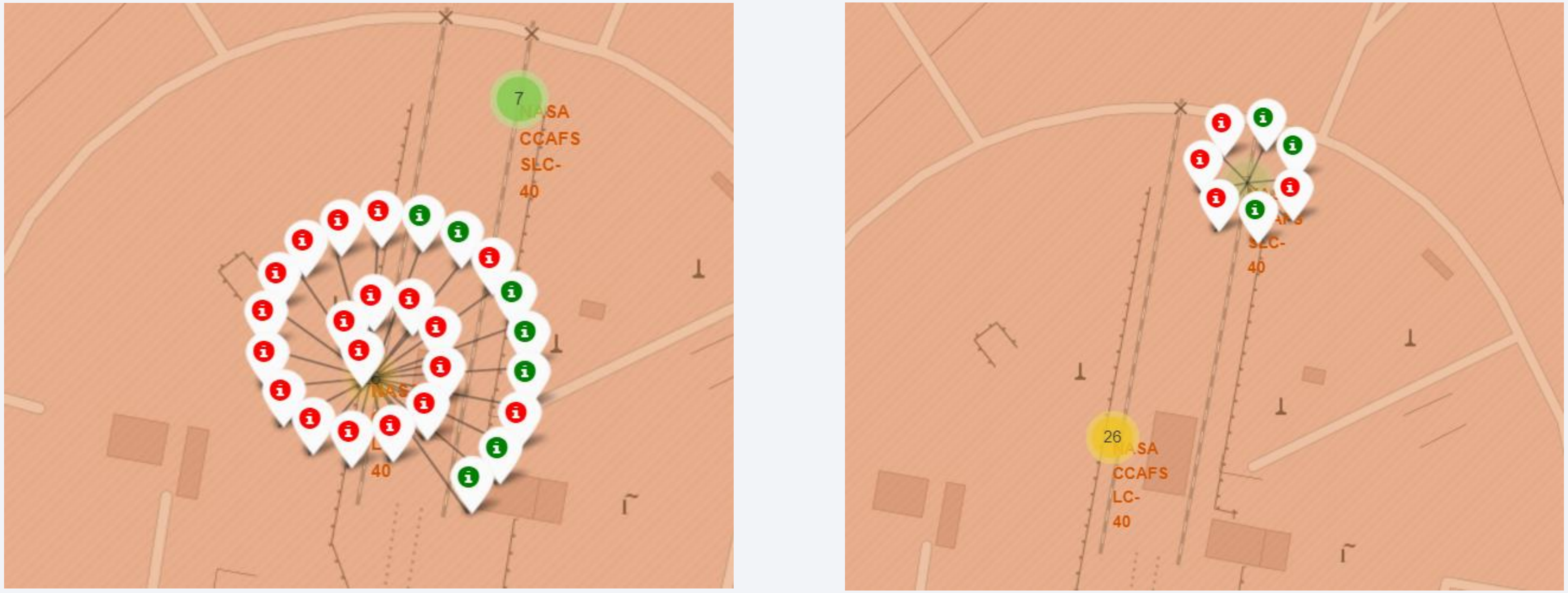


Figure 35. Map with the localization of CCAFS LC-40 and CCAFS SLC-40 launch site and marks of successful and unsuccessful landings. Own elaboration.

Launch Outcomes Per Launch Site

- The Kennedy Space Center (**KSC LC-39A**) has the **highest number of successful missions**. Nonetheless, it is not the most used Launch Site.

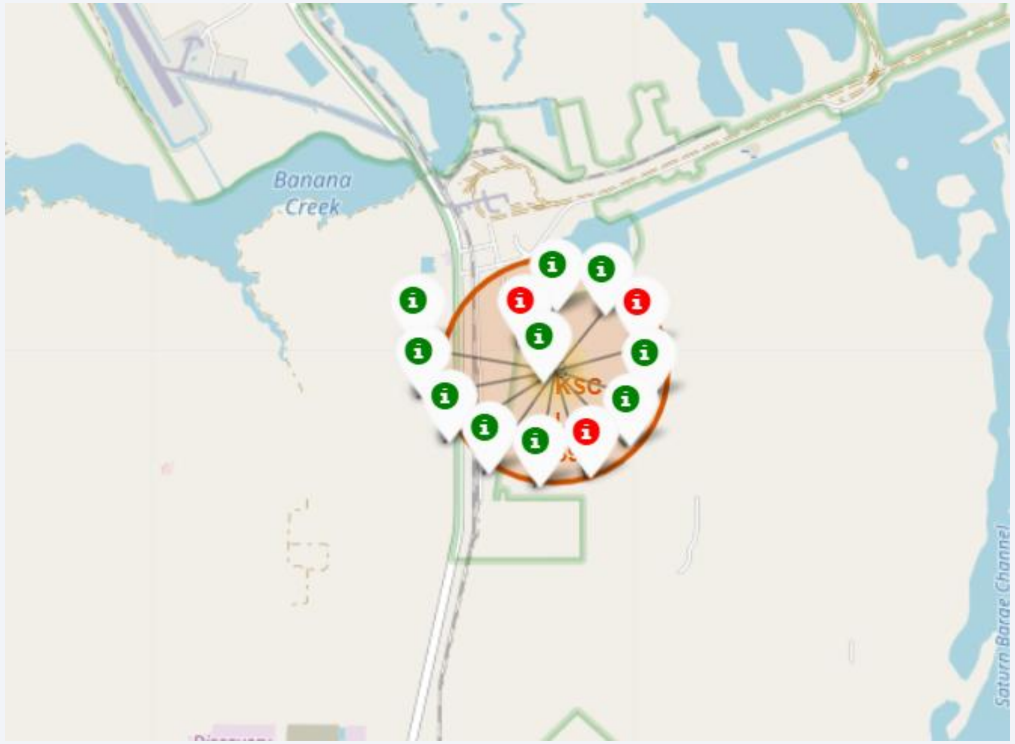


Figure 36. Map with the localization of KSC LC-39A launch site and marks of successful and unsuccessful landings. Own elaboration.

KSC LC-39A distance to the sea and Orlando, FL.

- **KSC LC-39A** is **6.46 KM** far from the Atlantic Sea, and **72.21 KM** far from Orlando, FL.

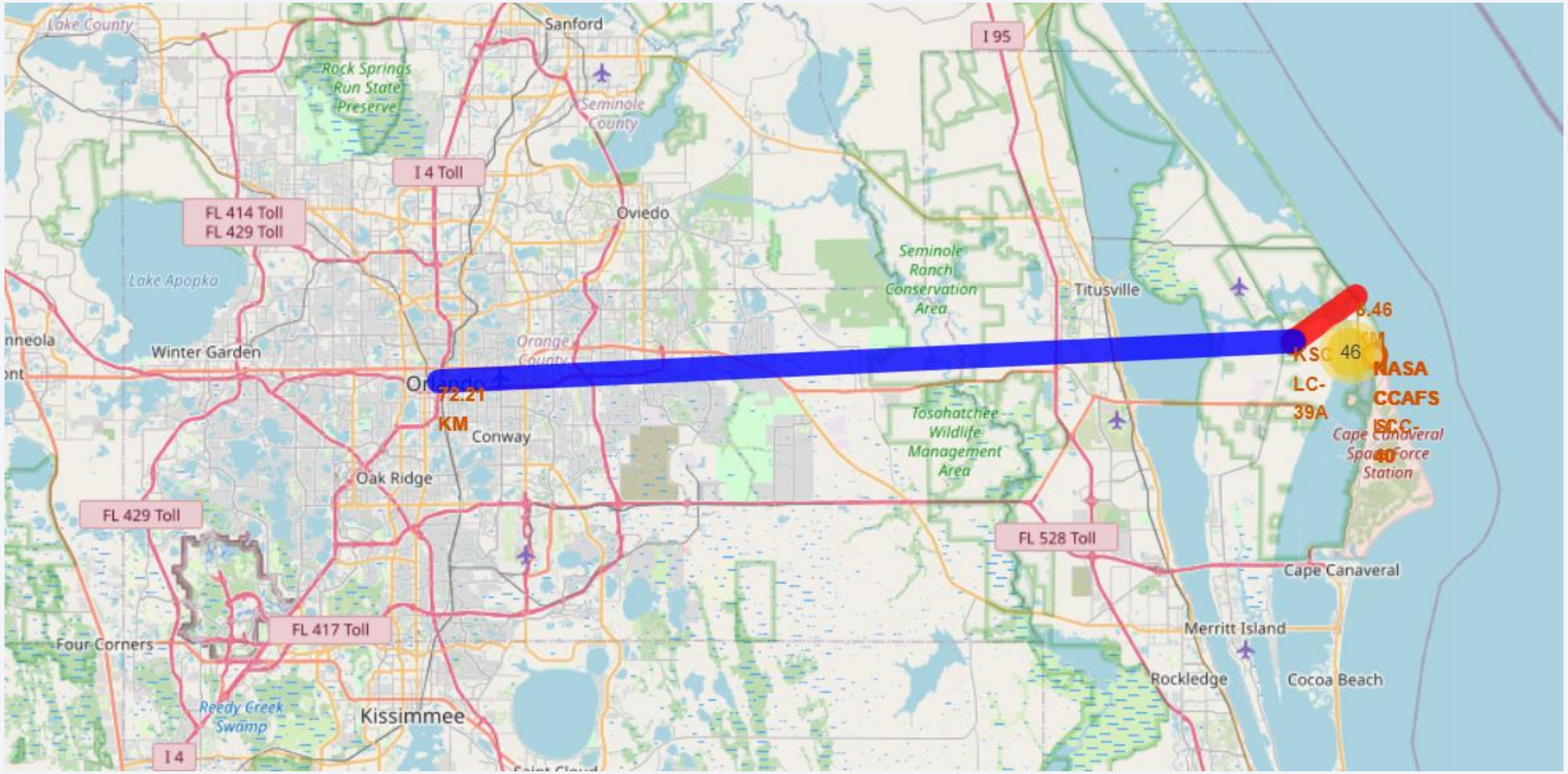


Figure 37. Map with the localization of KSC LC-39A launch site and distances to the sea and Orlando, FL. Own elaboration.

Section 4

Build a Dashboard with Plotly Dash

Launch Success Per Site

- **KSC LC-39A** has the **highest** success launch rates, while **VAFB SLC-4E** holds the **lowest**. It is important to note that CCAFS LC-40 and CCAFS SLC-40 refer to the same facility.

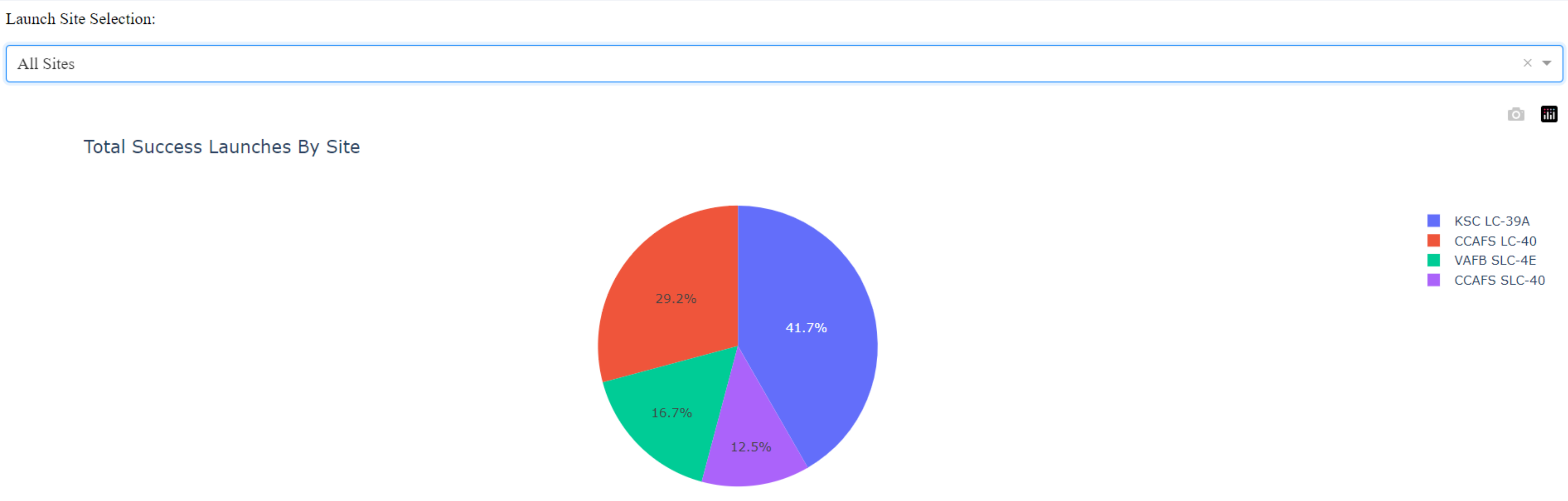


Figure 38. Launch Success Per Site in Dashboard. Own elaboration.

Highest Launch Success Ratio

- The Highest Launch Success Ratio is hold by the **KSC LC-39A** launch site, with a **76.9% of success** and only 23.1% of failure.

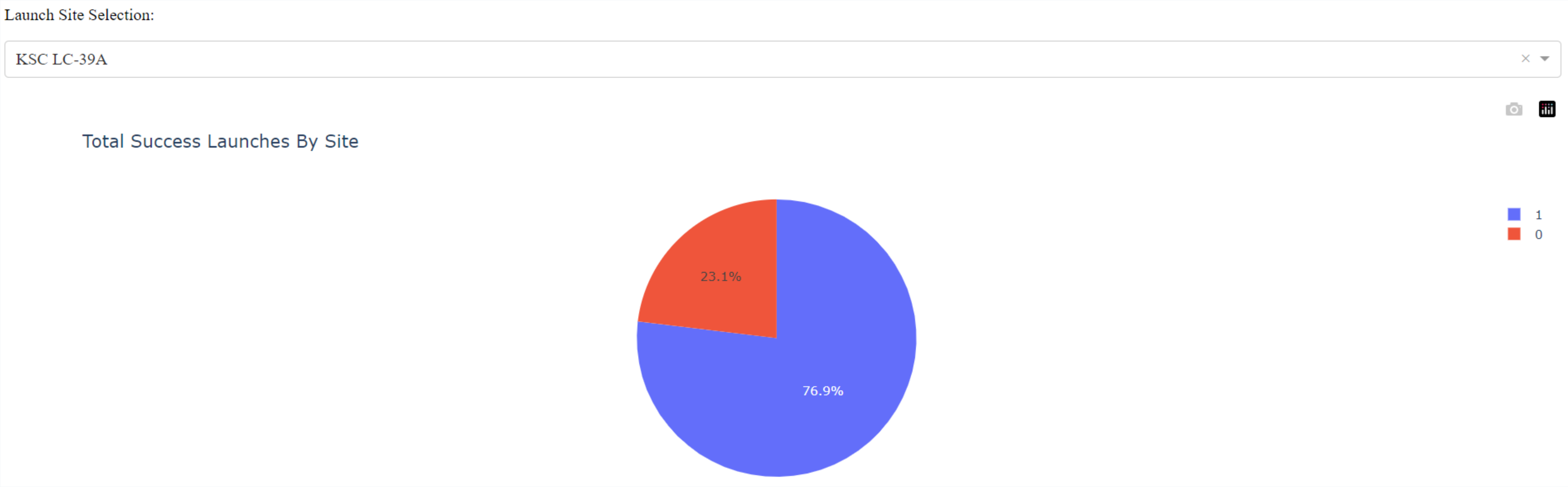


Figure 39. Launch Success for KSC LC-39A in Dashboard. Own elaboration.

Payload vs. Launch Outcome

- The Payload vs. Launch Outcome allowed to identify the payload ranges with the highest and the lowest success rates as well as the more successful F9 Booster versions (see next slides).

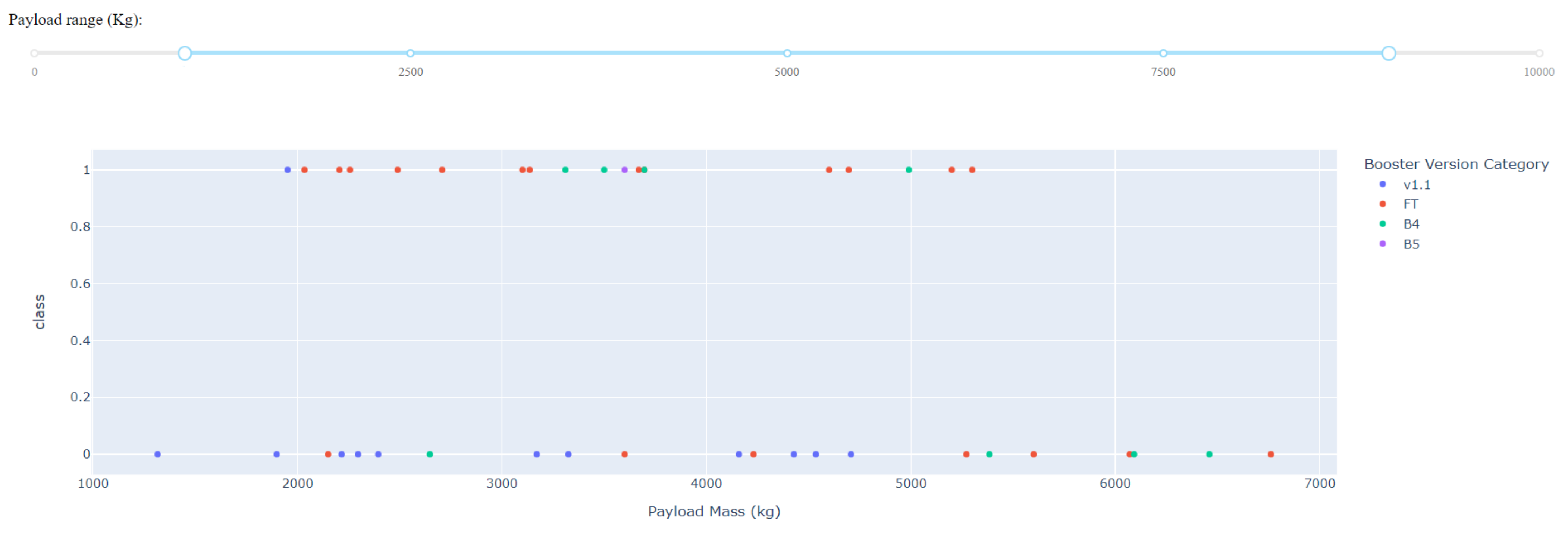


Figure 40. Dashboard's Scatter Plot of Payload Mass vs Class with Range Slider for Payload Mass 2. Own elaboration.

Payload vs. Launch Outcome

- Which payload range(s) has the highest launch success rate? **From 1952 kg to 5300 kg**.

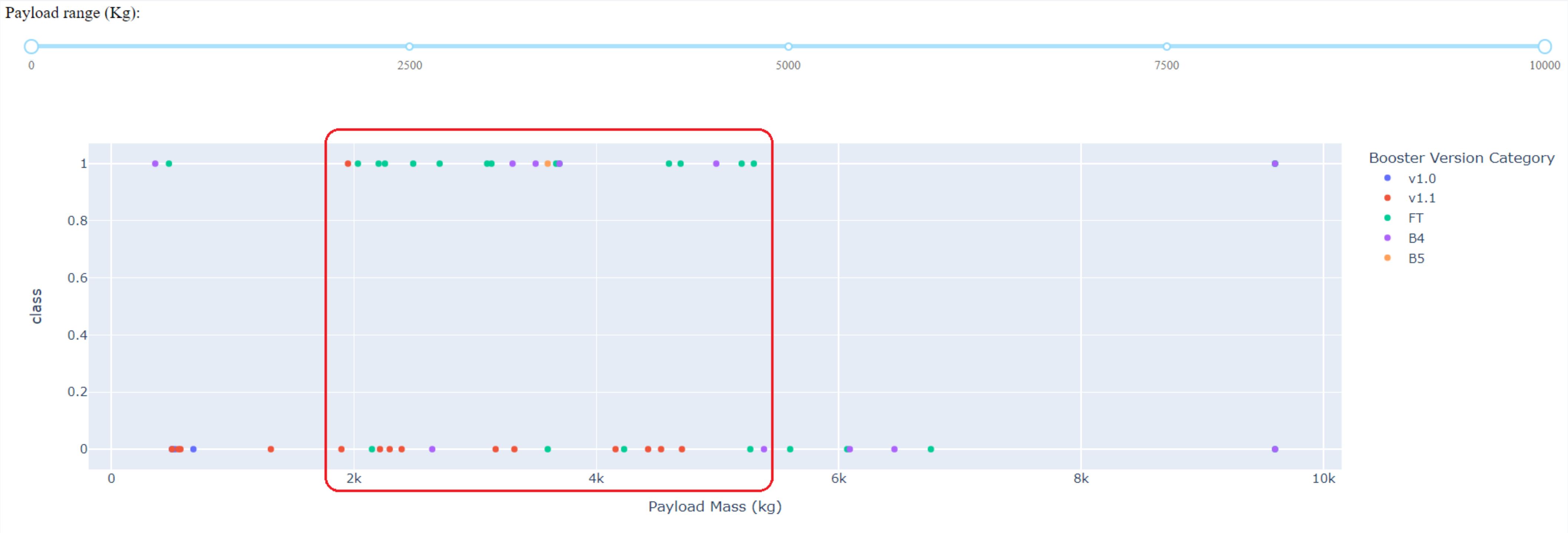


Figure 41. Payload range with the highest launch success rate. Own elaboration.

Payload vs. Launch Outcome

- Which payload range(s) has the lowest launch success rate? **Below 1952 kg and above 5300 kg.**

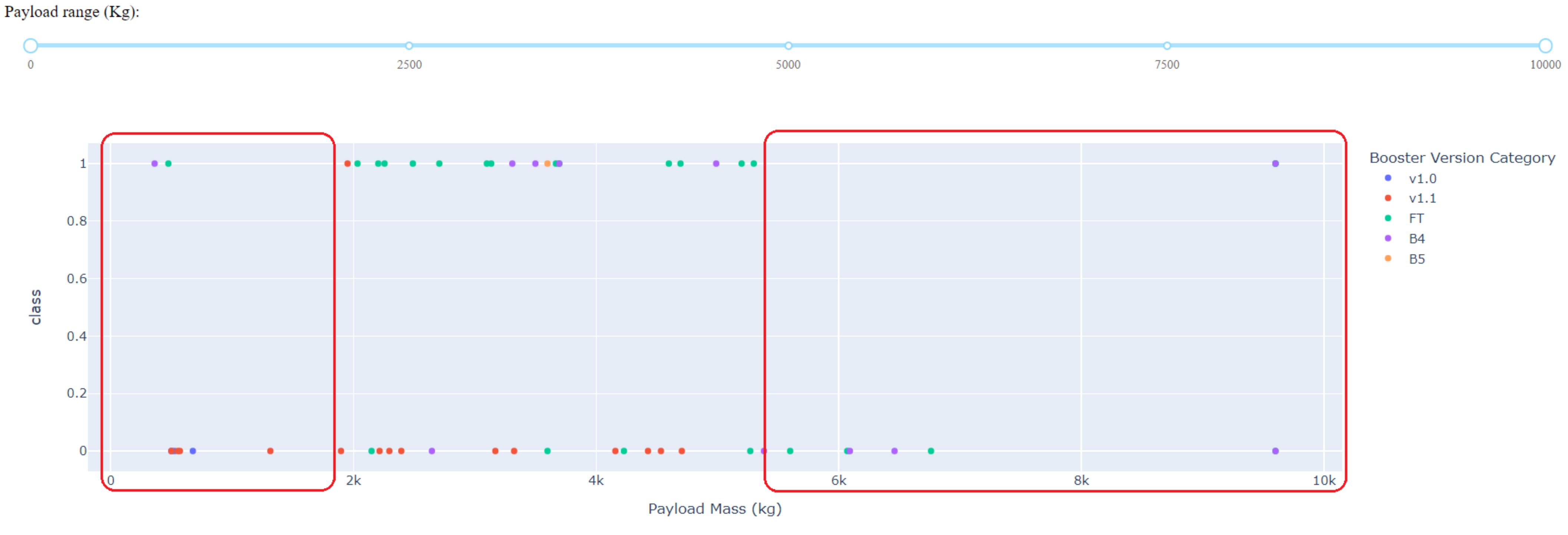


Figure 42. Payload ranges with the lowest launch success rate. Own elaboration.

Best Booster versions

- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? **The FT and the B5 Booster versions.**

Figure 43. F9 Booster versions with the highest launch success rate. Own elaboration.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest classification accuracy was the one built with **Decision Trees with an accuracy of 88.9%, when fitted with the best parameters:**
- {'criterion': 'gini', 'max_depth': 2, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}

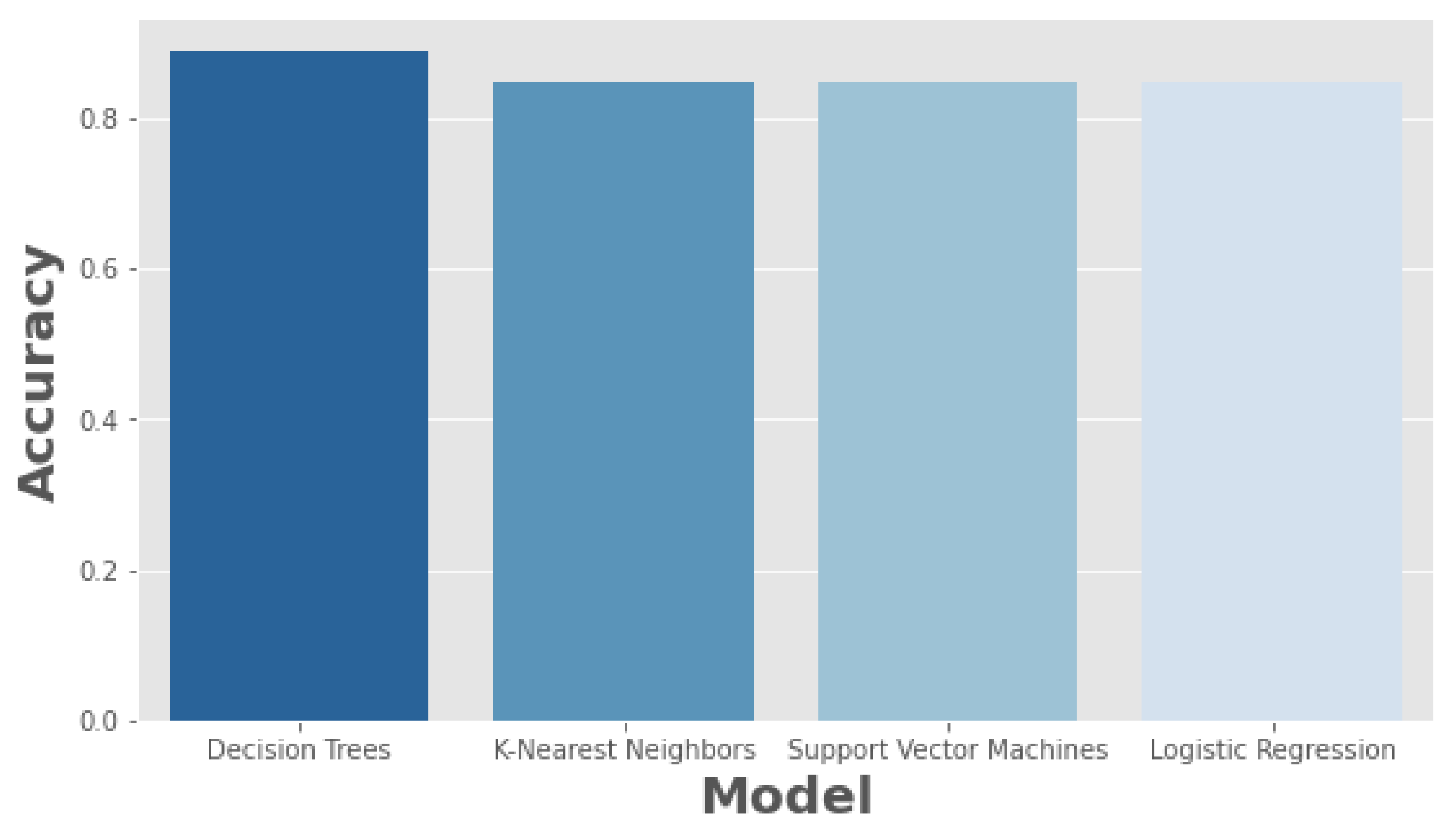


Figure 44. Accuracy of the built classification models. Own elaboration.

Confusion Matrix

Figure 19. Decision trees confusion matrix. Own elaboration.

- The confusion matrix suggests that the built model is better at **predicting successful landings** than unsuccessful landings.
- So, there is room for improvement in terms of reducing both false positives and false negatives.
- In spite of the above, this results suggest that, indeed, **most of the launches from SpaceX will land successfully.**

Conclusions

- Firstly, a classification model using Decision Trees was developed with a precision of **88.9%** and whose outcome indicates that **most of the launches from SpaceX will land successfully**.
- On the other hand, analysis from the historical launching data suggests that SpaceX has gotten better at launching and its success rate has stabilized since 2017 in about **80%**.
- Thus, the present study suggests that the cost of the Falcon 9 rocket launches should be set at **62 million dollars**, as stated by SpaceX.

Conclusions

- Other insights obtained from the present analysis were:
- **KSC LC-39A** is the launch site with the highest number of successful missions and **VAFB SLC-4E** is the one with the lowest.
- **ES-L1**, **GEO**, **HEO** and **SSO** are the orbit types with the highest success rates (100%) but with a small number of missions. On the other hand, **VLEO** is the most common orbit type in recent years.
- The **payload range** with the highest launch success rate is from 1952 kg to 5300 kg.
- **FT** and **B5** are the **booster versions** with the highest launch success rate.

Appendix

- Data set created from the calls to API SpaceX:
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/dataset_part_1_DEL.csv
- Data set created from the web scraping to Wikipedia's article on List of Falcon 9 and Falcon Heavy launches:
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/spacex_web_scraped.csv
- Data set created after Data wrangling:
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/dataset_part_2_DEL.csv
- Dataset created after the Feature Engineering process:
- https://github.com/DanielEduardoLopez/IBM-SpaceX/blob/main/dataset_part_3_DEL.csv

Thank you!