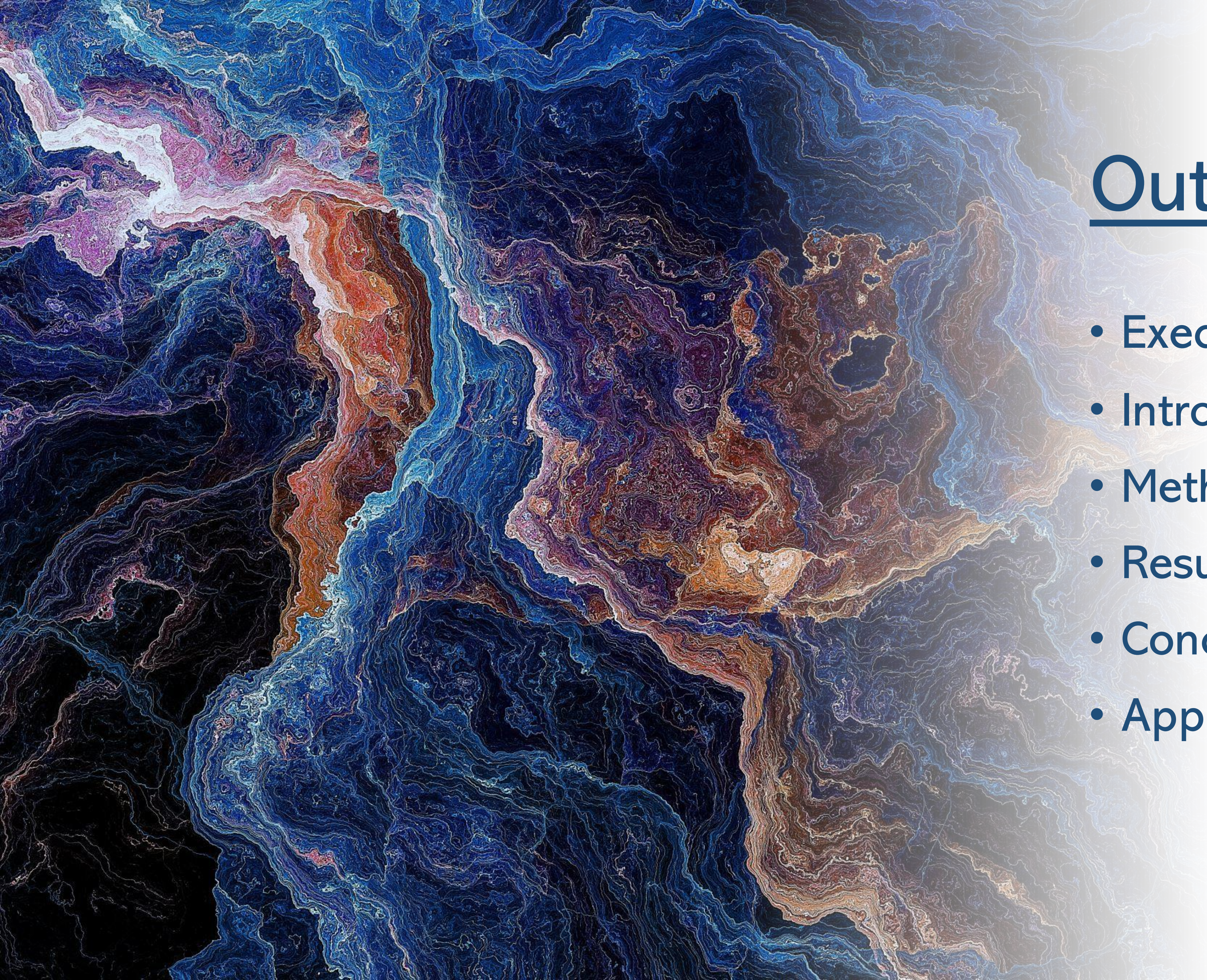
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

German Mercader
April, 2023



Outline

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

Executive Summary

- **Summary of methodologies:**
  - The data used for this analysis was obtained from the SpaceX API and by web scraping Wikipedia, it contains information about every Falcon 9 launch. The data was cleaned, processed, and analyzed using Python and various data science libraries such as pandas, NumPy, and Matplotlib.
- **Summary of results:**
  - In recent years, Falcon 9 first stage landings have a success rate of over 80%.
  - The success in landing is influenced by a variety of factors, we found that the type of mission and the landing location were significant factors affecting the success rate of the first stage landing. However, improvements made over time and based on previous launch data are presented as the main feature of success. Successful landings started in 2013 and have continued to increase until now.

Introduction

- **Project background and context:** The Falcon 9 rocket is a reusable launch vehicle developed by SpaceX. One of the most significant features of this rocket is the ability to land the first stage after launch, which translates into significant cost savings that enable SpaceX to conduct launches at a much lower price than its competitors. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars while other providers cost upward of 165 million dollars each.

- **Problems we want to find answers:** We want to identify which factors have a greater impact on the outcome of the missions. The goal of this data analysis is to evaluate the success of Falcon 9 first stage landings and identify any trends or patterns in the data. The data-driven evaluation of Falcon 9 first stage landings provides valuable insights into the factors that influence landing accuracy and success rates and can inform future improvements to Falcon 9 landing technology. Also, this information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary:

- Data collection methodology:
  - The data was obtained from the SpaceX API and by performing web scraping from a Wikipedia page. The data was cleaned, processed, and analyzed using Python and various data science libraries.
- Perform data wrangling:
  - Exploratory Data Analysis was performed to find some patterns in the data and determine what would be the label for training supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We standardize the data, split it into training data and test data and find the best Hyperparameter for Logistic Regression, Support Vector Machine, Decision Tree Classifier and K-Nearest Neighbor in order to find the method that performs best using test data.

Data Collection

Request to the SpaceX API, Web Scraping and Clean the requested data:

- We request rocket launch data from SpaceX API using the GET request, decode the response content as a Json and turn it into a Pandas dataframe. We also perform web scraping from a Wikipedia.
- We use the API again to get information about the launches using the IDs for each launch, such as:
  - From the rocket we got the booster name.
  - From the payload we got the mass of the payload and the orbit that it is going to.
  - From the launchpad we got the name of the launch site being used, the longitude, and the latitude.
  - From cores we got the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, whether the core is reused, whether legs were used, the landing pad used, the block of the core, the number of times this specific core has been reused, and the serial of the core.
- The data from these requests was stored in lists and was used to create a new dataframe.
- After that, we had to filter the dataframe to only include Falcon 9 launches a deal with some missing values, for example, for the missing values of the Payload column we calculated the mean and use it to replace the missing values.

Data Collection – SpaceX API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

GET request to the SpaceX API

API returns data in JSON

Turn JSON into a Pandas DF and clean de data

- GitHub link of the completed SpaceX API calls notebook: [Falcon 9 API Data Collection]

Data Collection – Scraping

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA (COTS)\nNRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
116	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	06:42
117	118	KSC	Starlink	~14,000 kg	LEO	SpaceX Capella Space and Tyvak	Success\n	F9 B5B1058.8	Success	15 May 2021	22:56
118	119	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1063.2	Success	26 May 2021	18:59
119	120	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA (CRS)	Success\n	F9 B5B1067.1	Success	3 June 2021	17:29
120	121	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	F9 B5	Success	6 June 2021	04:26

- GitHub link of the completed web scraping notebook: [Falcon 9 Web Scraping Data Collection]()

HTTP GET request to the **Falcon9** Wiki page.

↓

Parse the HTML with BeautifulSoup

↓

Turn it into a Pandas DF

Data Wrangling

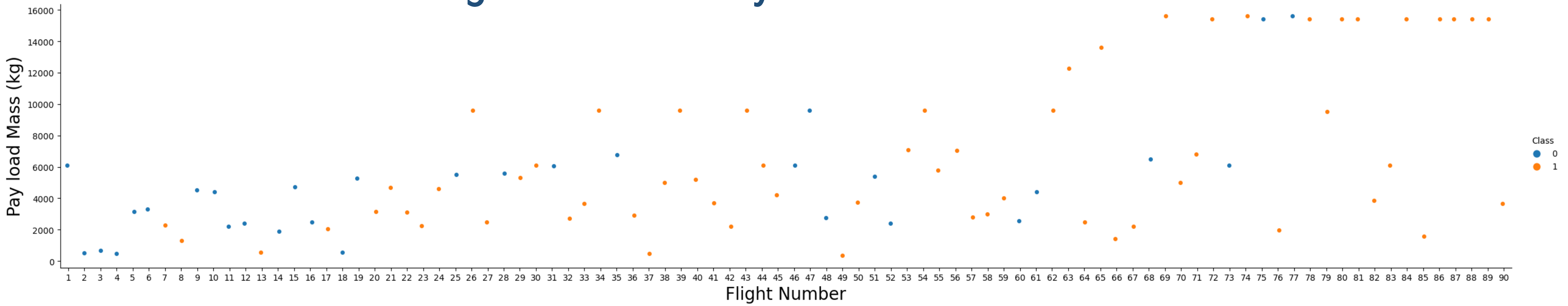
Launches and outcomes by site and orbit → One hot encoding for the outcomes → Determine Training Labels

- We performed some Exploratory Data Analysis to find patterns in the data and determine what would be the label for training supervised models.
- Sometimes a landing was attempted but failed:
  - True Ocean means the mission outcome was successfully landed to a specific region of the ocean, while False Ocean means the mission outcome was unsuccessfully.
  - True RTLS means the mission outcome was successfully landed to a ground pad, while False RTLS means the mission outcome was unsuccessfully.
  - True ASDS means the mission outcome was successfully landed on a drone ship, while False ASDS means the mission outcome was unsuccessfully.
- We calculated the number of launches on each site, the number and occurrence of each orbit and the number and occurence of mission outcome per orbit type.
- Then, we used "one hot encoding" to convert those outcomes into Training Labels with 1 means the booster successfully landed and 0 means it was unsuccessful.
- GitHub link of the completed data wrangling related notebook: Falcon 9 Data Wrangling

EDA with Data Visualization

- Scatterplots, barplots and lineplots were utilized to visualize the correlation between pairs of features during the process of data exploration.
- GitHub link of completed EDA with data visualization notebook: [Falcon 9 EDA with Data Visualization]

Let's see how the FlightNumber and Payload variables affect the launch outcome:



We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

EDA with SQL

<u>In order to further understand the Spacex DataSet, we load the dataset in a Db2 database and execute the following SQL queries:</u>

- 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 the date when the first successful 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 failed landing outcomes in drone ship, their booster versions, and launch site names for 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.
- GitHub link to the completed EDA with SQL notebook: [Falcon 9 EDA whth SQL]

Build an Interactive Map with Folium

- **The objective was mark all launch sites on a map, mark the success/failed launches for each site on the map and calculate the distances between a launch site to its proximities.**

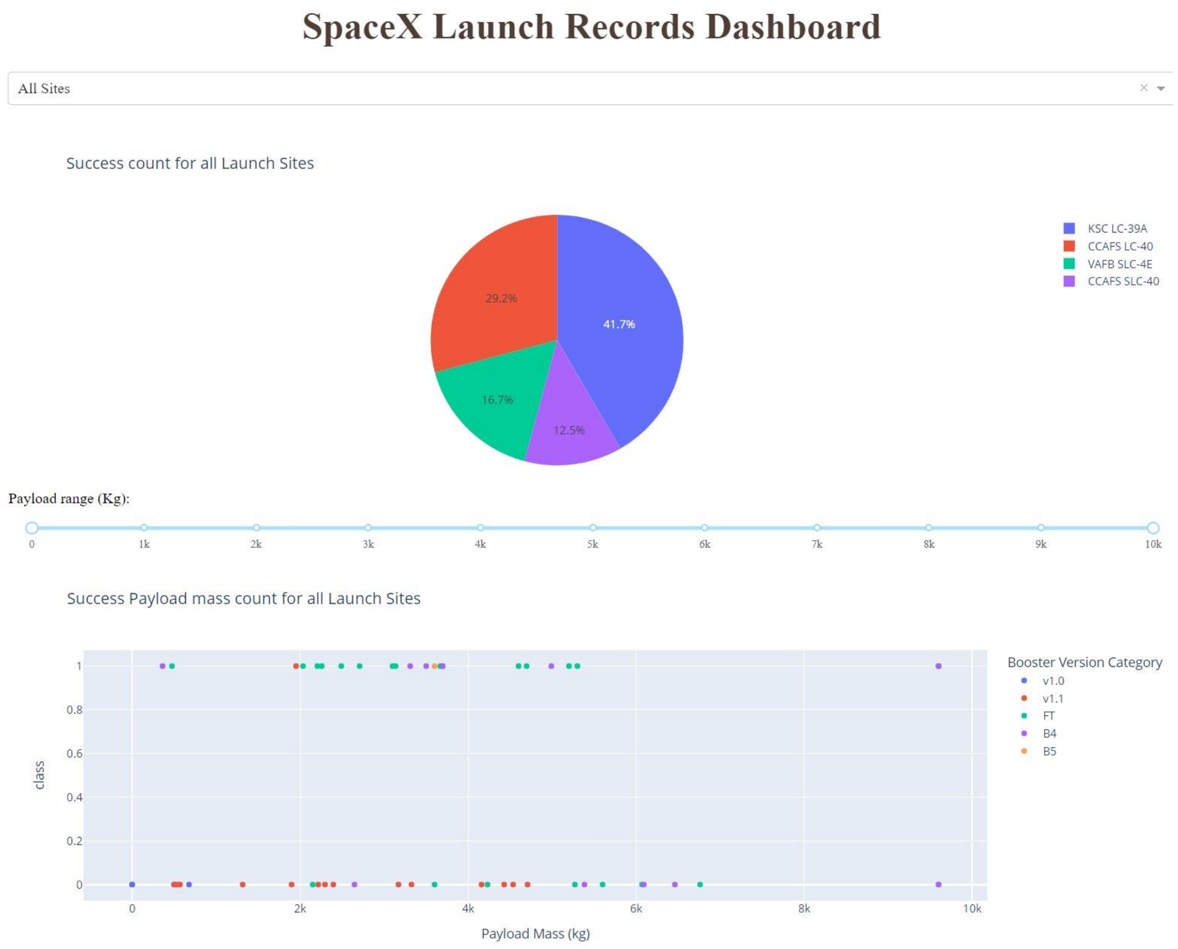
We took the Latitude and Longitude Coordinates of each launch and added a Circle Marker. Then, to visualize the launch outcomes data in a map, we used a Marker Cluster and assigned the failures and successes to the classes 1 and 0, respectively. The markers for the successes were colored green, while the markers for the failures were colored red. We also analyze launch site proximities

GitHub link to the completed EDA with SQL notebook: [Falcon 9 Interactive Map with Folium]

Build a Dashboard with Plotly Dash

- We developed a real-time interactive visual analytics application for SpaceX launch data using Plotly Dash. The application allows users to adjust parameters such as launch site and payload range, generating corresponding graphs and plots for quick analysis of the relationship between payloads and launch sites.

- You can find the source code of the application at the following link to conduct your own interactive visual analytics on SpaceX launch data in real-time: [SpaceX Dash App]()



Predictive Analysis (Classification)

- **To find the best performing classification model, we built, evaluated, and enhanced four distinct models: Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors.**

Data Preparation and Standardization

Create a NumPy array from the column Class	Standardize the data	Split the data into training and test datasets

Build and enhance four distinct classification models

Create the models	Enhanced the models by their hyperparameters

Find the best performing classification model

Calculate the accuracy of the models	Comparison of results

- GitHub link to the completed Predictive Analysis notebook: [Falcon 9 Predictive Analysis]

Results

- The "preliminary" results of the predictive analysis indicate that, after comparing the accuracy of the above models, they all performed similarly with an accuracy score of 0.83.

logreg_cv.score(X_test, Y_test)	svm_cv.score(X_test, Y_test)	tree_cv.score(X_test, Y_test)	knn_cv.score(X_test, Y_test)
0.8333333333333334	0.8333333333333334	0.8333333333333334	0.8333333333333334

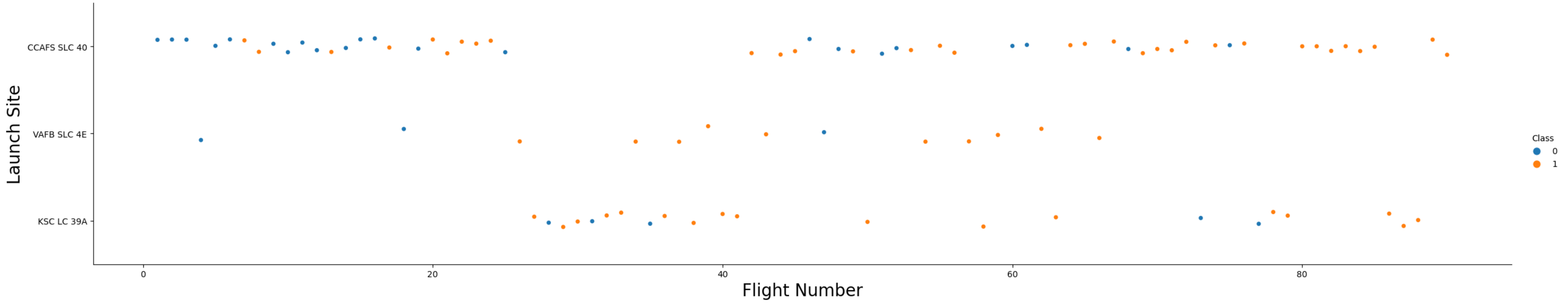
- As for the results of the previous exploratory data analysis, in the next section we will provide detailed insights which are quite interesting.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

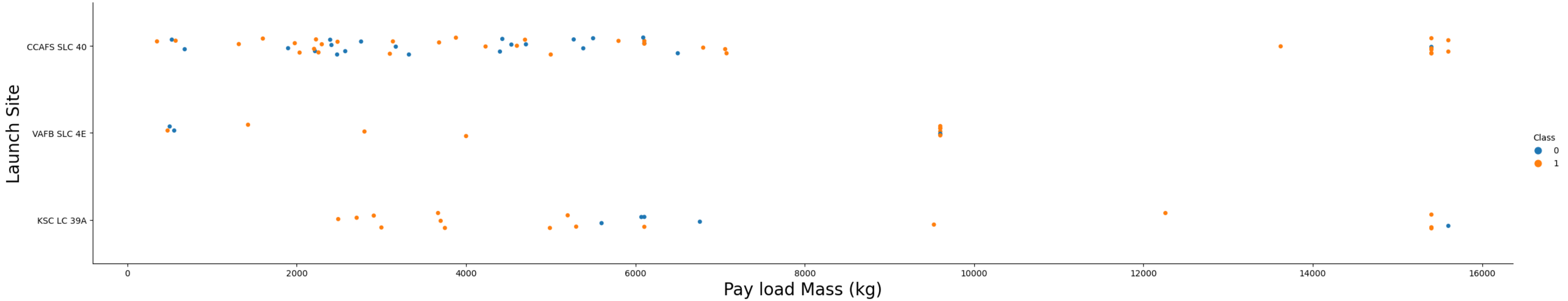
Let's see how the FlightNumber and Launch Site variables affect the launch outcome:



We see that different launch sites have different success rates. CCAFS LC-40 has an average success rate of 60 %, while KSC LC-39A and VAFB SLC 4E have an average success rate of 77%.

Payload vs. Launch Site

Now let's analize the relationship between the Payload and Launch Site:

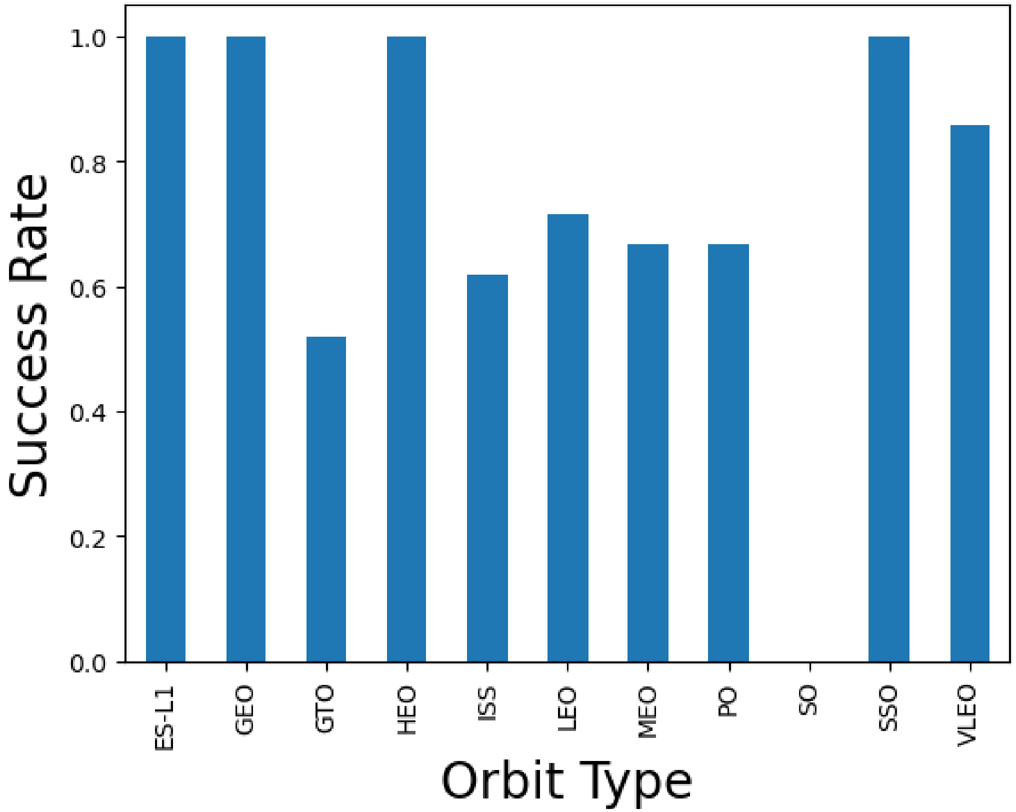


For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass (greater than 10000). Payloads that approached the maximum limit tended to be launched from CCAFS SLC 40 and KSC LC 39A. On the other hand, payloads weighing less than 8000 kg had a higher failure rate when launched from CCAFS SLC 40.

Success Rate vs. Orbit Type

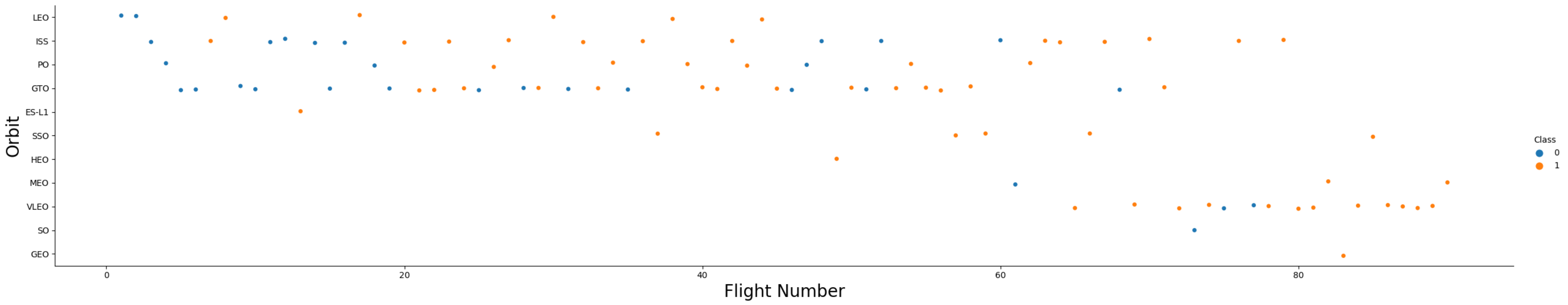
Here we can see the relationship between success rate of each orbit type:

Based on the analysis of the data, it was found that Earth-Sun L1 (ES-L1), Geostationary Earth Orbit (GEO), High Earth Orbit (HEO), and Sun-Synchronous Orbit (SSO) have the highest success rates. Very Low Earth Orbit (VLEO) follows closely as the second highest. On the other hand, Geostationary Transfer Orbit (GTO) has the lowest success rate among all the orbits analyzed. The remaining orbits, namely International Space Station (ISS), Low Earth Orbit (LEO), Medium Earth Orbit (MEO), and Polar Orbit (PO), were found to have an average success rate.



Flight Number vs. Orbit Type

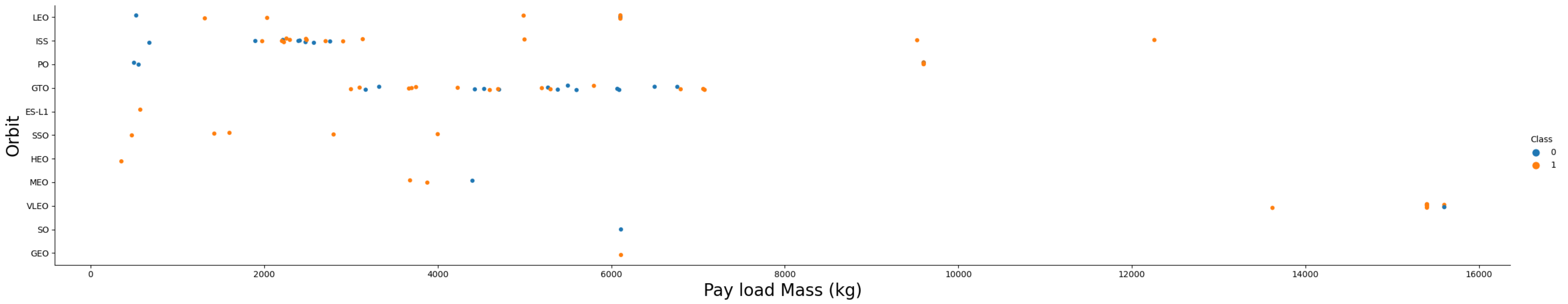
Next we analize if there is any relationship between Flight Number and Orbit type:



In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

Here we can visualize if there is any relationship between Payload and Orbit type:

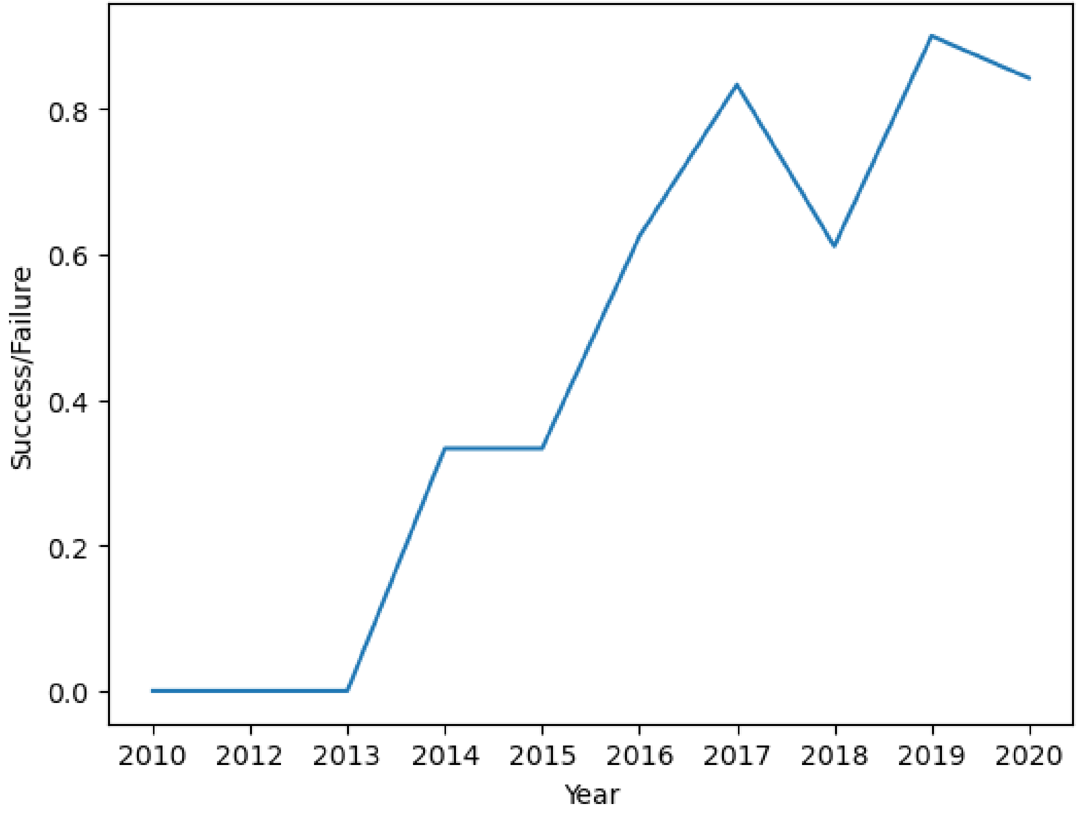


With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are both there here.

Launch Success Yearly Trend

And here we see the launch success yearly trend:

We can observe that the sucess rate since 2013 kept increasing. The success in landing is influenced by a variety of factors, we found that the type of mission and the landing location were significant factors affecting the success rate of the first stage landing. However, improvements made over time and based on previous launch data are presented as the main feature of success. Successful landings started in 2013 and have continued to increase until now.



All Launch Site Names

- CCAFS LC-40 (Cape Canaveral Air Force Station, Launch Complex 40)
- CCAFS SLC-40 (Cape Canaveral Air Force Station, Space Launch Complex 40)
- KSC LC-39A (Kennedy Space Center, Launch Complex 39A)
- VAFB SLC-4E (Vandenberg Air Force Base, Space Launch Complex 4E)

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

```
%sql select DISTINCT launch_site from Spacex;
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

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

Launch Site Names Begin with 'CCA'

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

```
%sql select * from Spacex WHERE launch_site LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

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

- Here we see an SQL query of five records where launch sites begin with `CCA`.

Total Payload Mass

- The total payload mass carried by boosters from NASA (CRS) is 45596 Kg.

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

```
%sql select SUM(payload_mass_kg_) from Spacex WHERE Customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2534 Kg.

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

```
%sql select AVG(payload_mass_kg_) from Spacex WHERE Booster_Version LIKE 'F9 v1.1%';
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

1
2534

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was 2015-12-22

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

Hint:Use min function

```
%sql select MIN(Date) from Spacex WHERE landing_outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Here we have a list of booster names that have successfully landed on a drone ship with a 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 DISTINCT Booster_Version from Spacex WHERE landing_outcome = 'Success (drone ship)' AND 4000 < payload_mass_kg_ < 6000;
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

booster_version
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1
F9 FT B1029.2
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1038.1

Total Number of Successful and Failure Mission Outcomes

- We have calculated the total number of successful and failed mission outcomes, and the result was 99 successes, 1 failure, and 1 outcome with an unclear status.

List the total number of successful and failure mission outcomes

```
%sql select Mission_Outcome, COUNT(Mission_Outcome) as TOTAL_NUMBER from Spacex GROUP BY Mission_Outcome;
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Here we have a list of booster names 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 DISTINCT booster_version from Spacex WHERE payload_mass_kg_ = (select MAX(payload_mass_kg_) from Spacex);
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Here we have a list of failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.

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
FROM Spacex
WHERE landing_Outcome = 'Failure (drone ship)' AND YEAR(DATE) = 2015;
```

```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

landing_outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	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) AS TOTAL_NUMBER
FROM Spacex
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing_outcome
ORDER BY TOTAL_NUMBER DESC
```

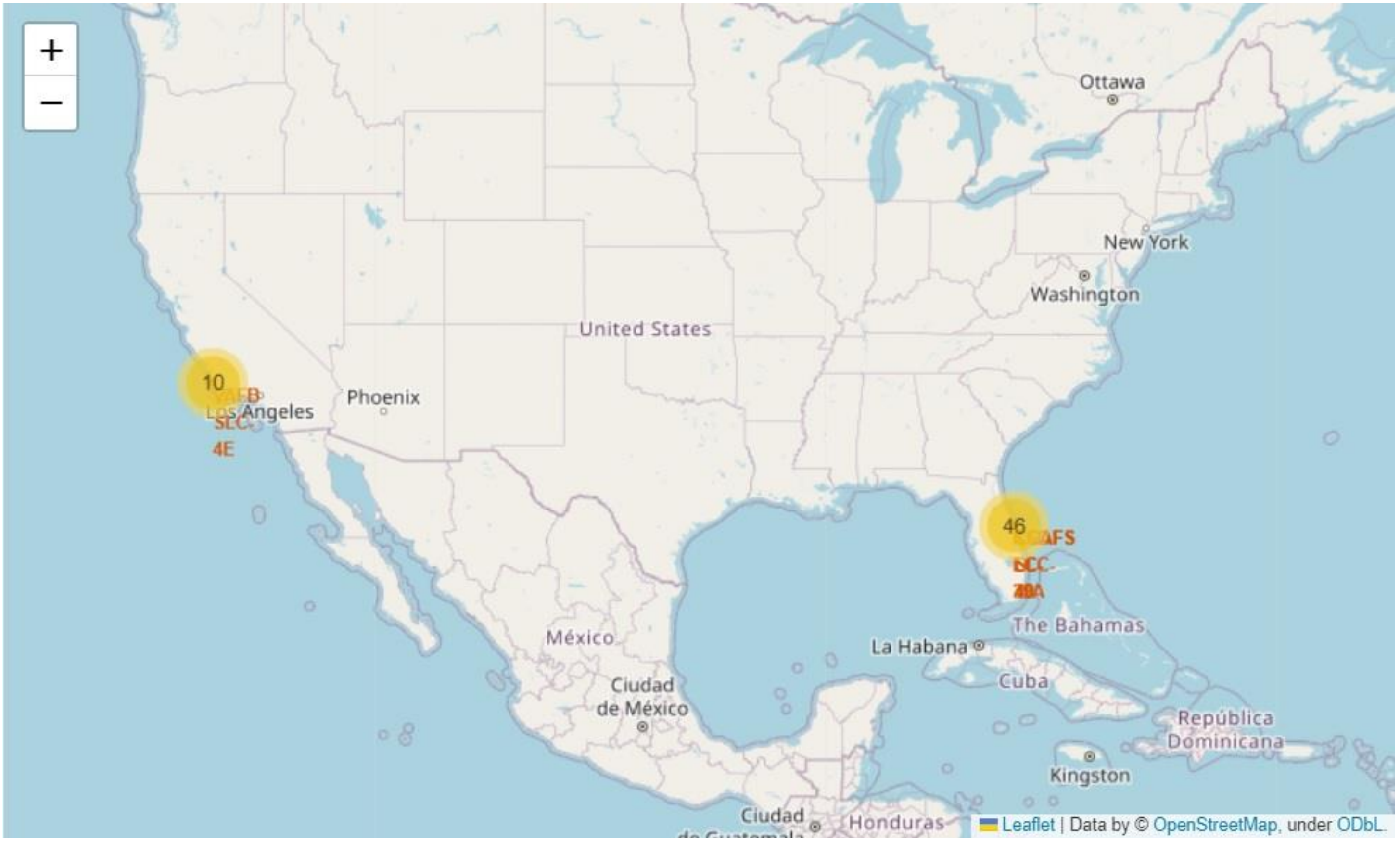
```
 * ibm_db_sa://bjf08488:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

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

Section 3

Launch Sites Proximities Analysis

All launch sites' location markers on a map



- After analyzing all launch sites, it was found that launch sites are typically located near the equator because this allows rockets to take advantage of the Earth's high rotational speed, which provides an extra boost in velocity.
- Additionally, for safety reasons, launch sites are usually located near coastlines to reduce the risk of accidents affecting populated areas.

Color-labeled launch outcomes on the map

- We took the Latitude and Longitude Coordinates of each launch and added a Circle Marker. Then, to visualize the launch outcomes data in a map, we used a Marker Cluster and assigned the failures and successes to the classes 1 and 0, respectively. The markers for the successes were colored green, while the markers for the failures were colored red.

Launch sites proximities

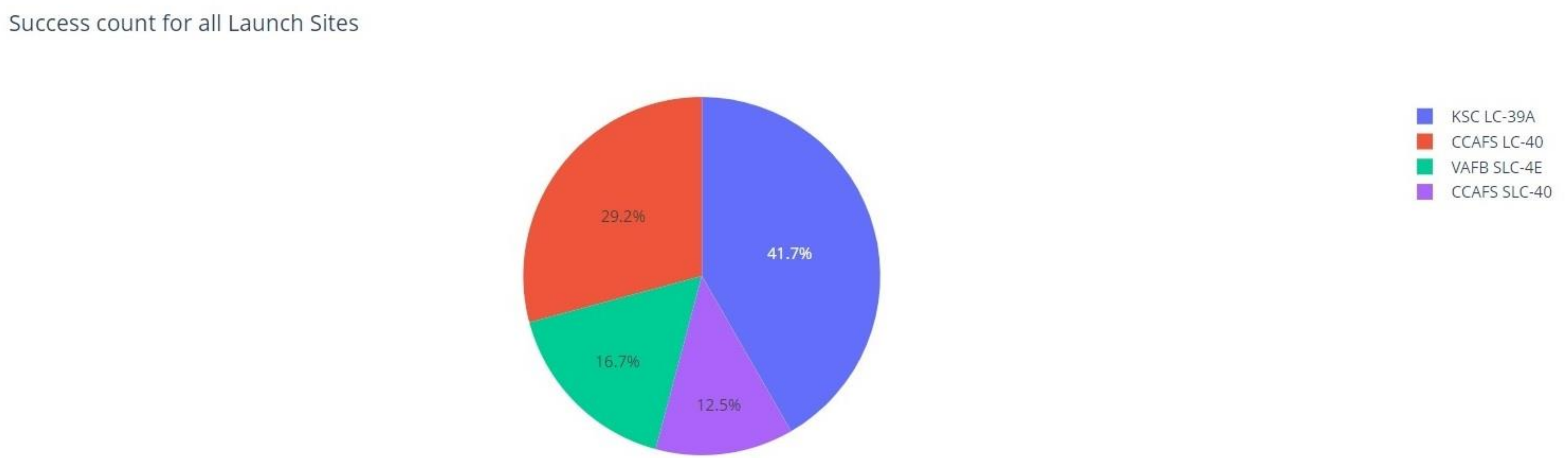
- Furthermore, launch facilities are typically situated in proximity to major transportation routes such as highways and railways, which aids in the movement of large payloads. Moreover, as a safety precaution, these facilities are usually positioned far away from densely populated areas to reduce the potential risk to nearby communities.

- For instance, we have calculated the distance from CCAFS SLC-40 to the nearest coastline, which is 0.86 km away. We have also measured the distance to the closest highway (0.58 km), the closest railroad (0.99 km), and the closest city (51.23 km).

Section 4

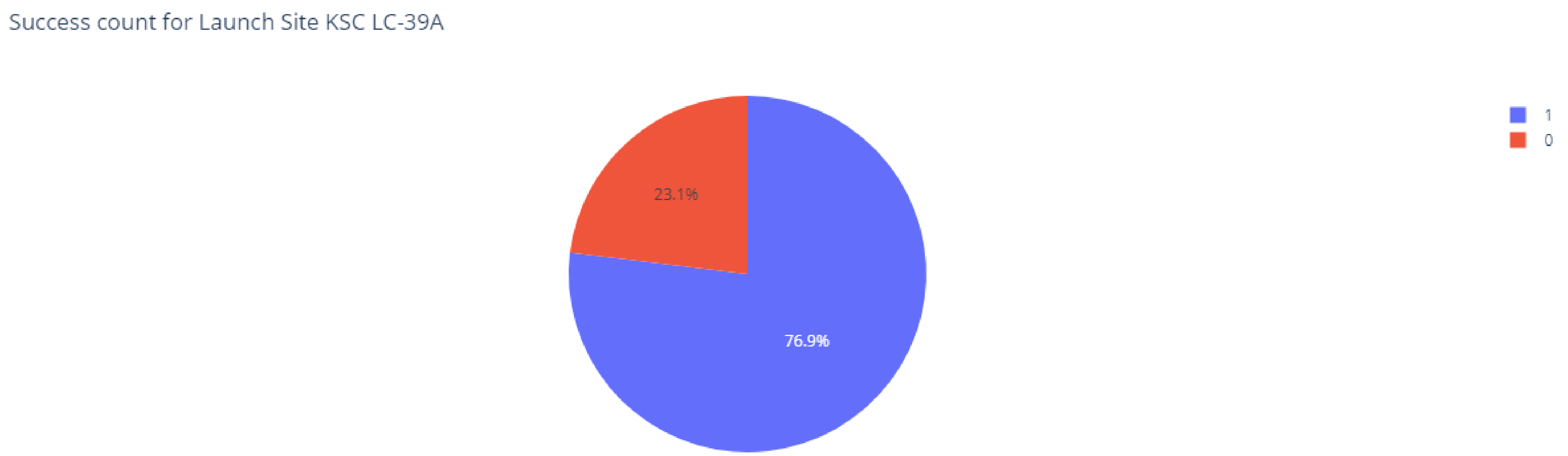
Build a Dashboard with Plotly Dash

Proportional launch success for all sites



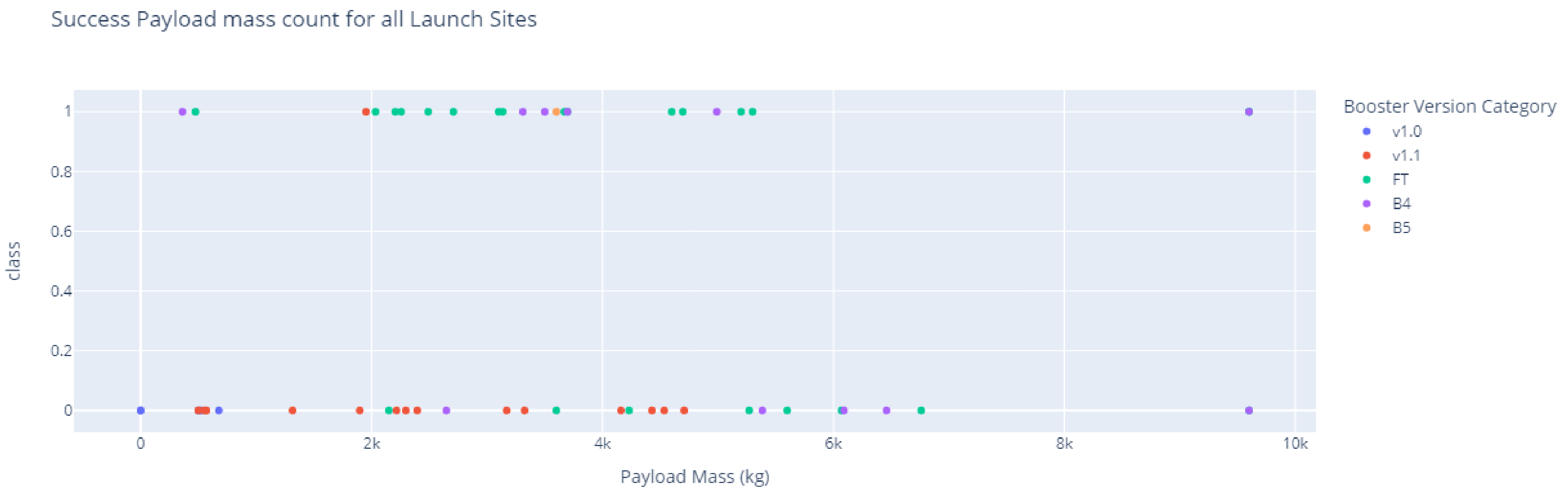
- The chart displays that KSC LC-39A (Kennedy Space Center, Launch Complex 39A) has the highest success rate of 41.7% among all launch sites, whereas CCAFS LC-40 (Cape Canaveral Air Force Station, Launch Complex 40) comes second with a success rate of 29.2%. On the other hand, VAFB SLC-4E (Vandenberg Air Force Base, Space Launch Complex 4E) and CCAFS SLC-40 (Cape Canaveral Air Force Station, Space Launch Complex 40) are the least successful launch sites with a success rate of 16.7% and 12.5% respectively.

Launch site with highest launch success ratio



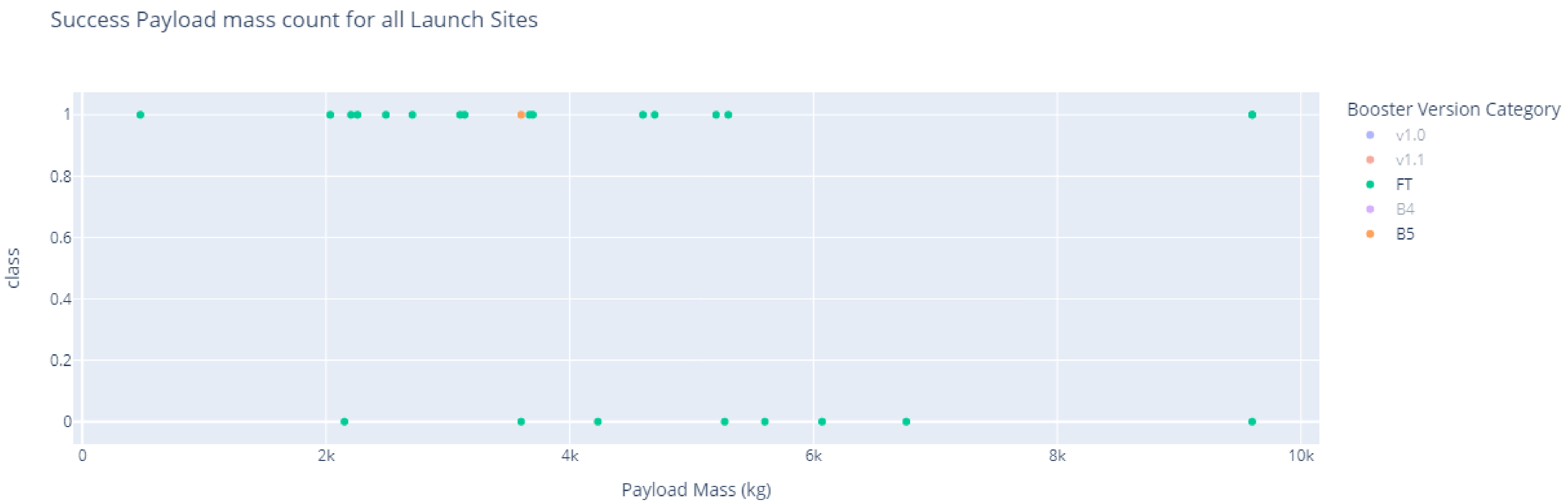
- After analyzing the data with the Dash Application, it was found that the launch site KSC LC-39A (Kennedy Space Center, Launch Complex 39A) has the highest success rate with a ratio of 76.9% success and 23.1% failure.

Payload vs. Launch Outcome scatter plot for all sites



- Additionally, we found that the payload range between 2000 Kg. to 4000Kg. had the highest launch success rate, while the range between 6000 Kg. to 8000Kg. had the lowest success rate.

Isolated FT and B5 Booster versions



- We also analyzed the success rate of different F9 Booster versions (v1.0, v1.1, FT, B4, B5, etc.) and found that the FT version had 15 successes and 8 failures, while the B5 version had only one successful launch and no failures.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

```
print("tuned hpyerparameters :(best parameters) ",svm_cv.best_params_)
print("accuracy :",svm_cv.best_score_)
```

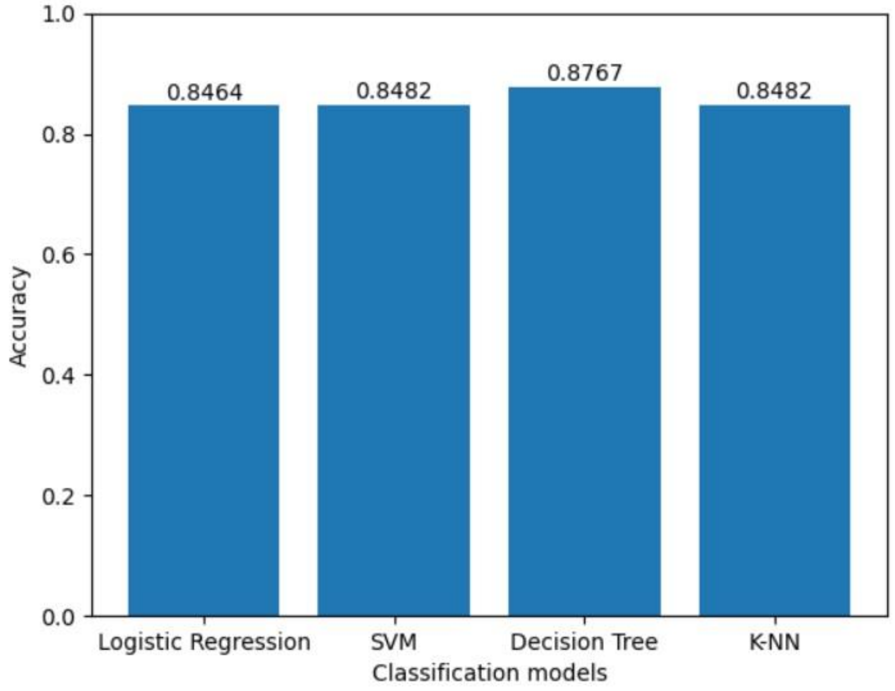
```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

```
print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
print("accuracy :",tree_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split':
10, 'splitter': 'random'}
accuracy : 0.8767857142857143
```

```
print("tuned hpyerparameters :(best parameters) ",knn_cv.best_params_)
print("accuracy :",knn_cv.best_score_)
```

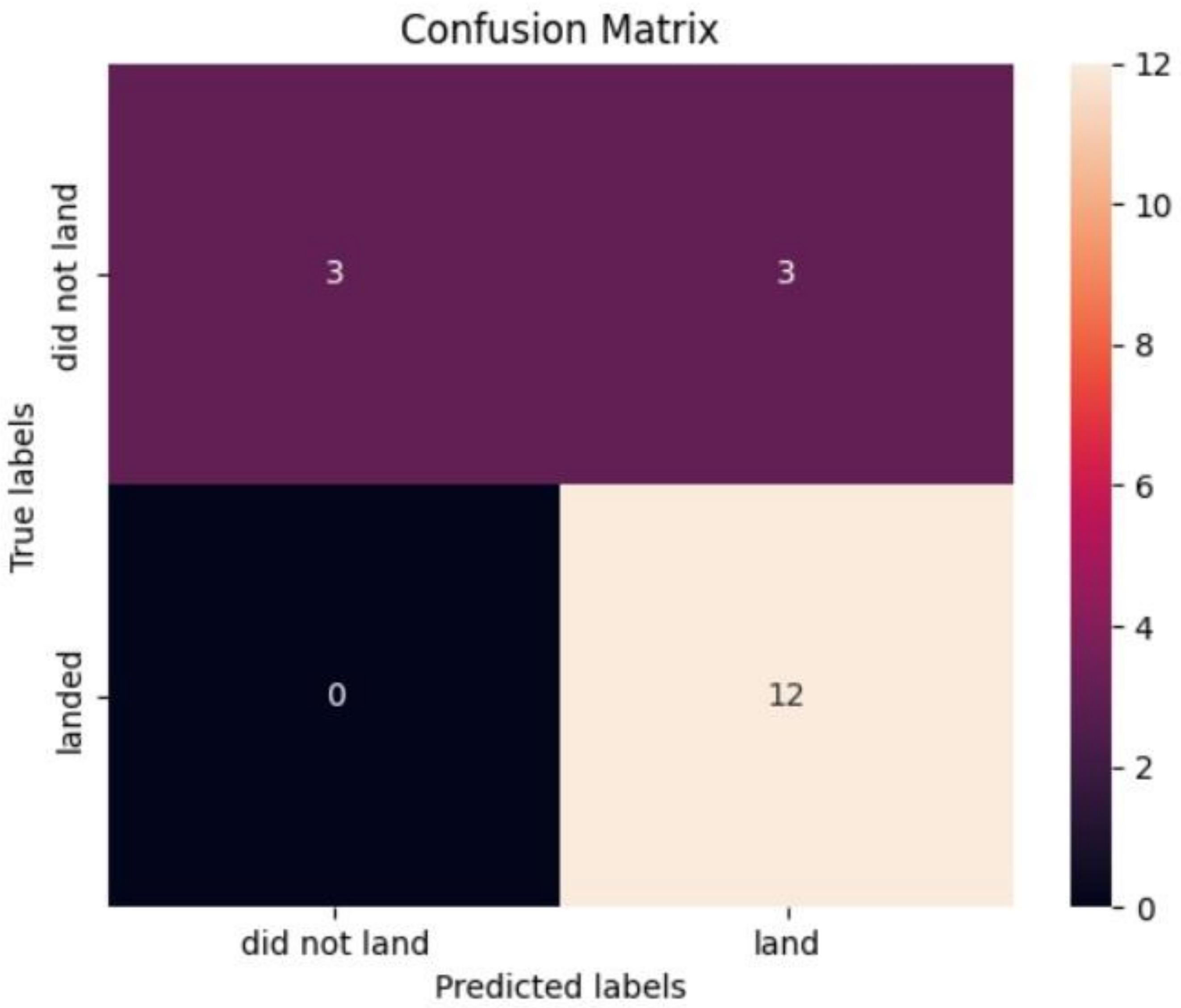
```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```



While the preliminary precision results of the different classification models indicated similar values, after optimizing the hyperparameters of these models, we found that the Decision Tree model has the highest classification accuracy at 0.8767.

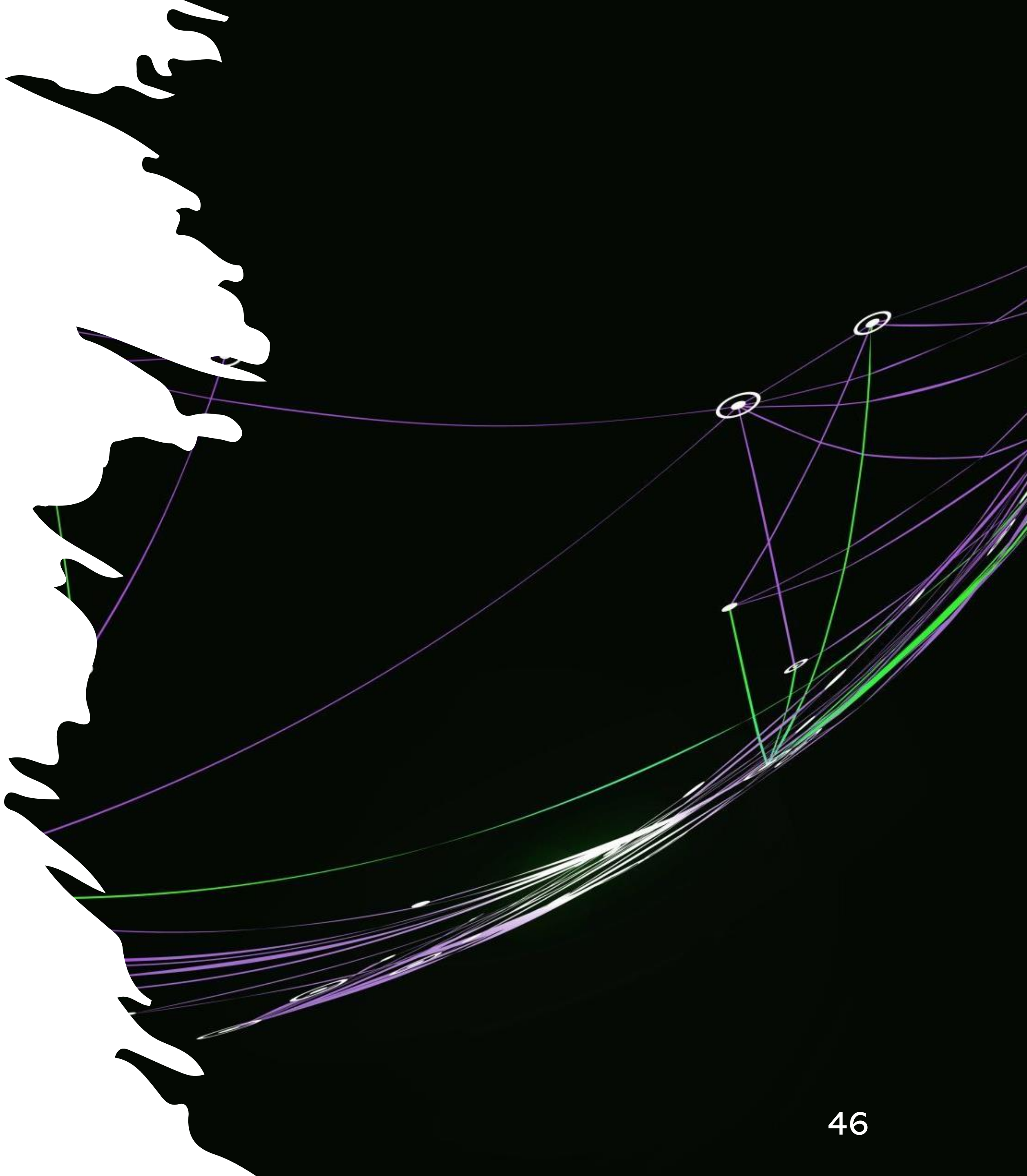
Confusion Matrix

- Upon examining the confusion matrix, we observe that the Decision Tree model is able to differentiate well between the different classes. However, a major issue is the occurrence of false positives.



Conclusions

- The Decision Tree classifier algorithm is determined to be the most suitable machine learning algorithm for this dataset.
- A higher number of flights at a launch site is correlated with a greater success rate at that launch site.
- Launch success rates have been on the rise since 2013, reaching a peak in 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO have the highest success rates.
- KSC LC-39A is the most successful launch site of all the sites studied.
- Low-weighted payloads tend to have better success rates than heavier payloads.



Appendix

- You may encounter issues viewing the maps in the Interactive Map with Folium notebook from GitHub. Where the maps should be displayed, you may see the message: "Make this Notebook Trusted to load map: File -> Trust Notebook". This is not an error in the lab exercise, but rather due to security settings in some web browsers. You can try adjusting your security settings or simply follow the code written in the corresponding notebook cell and refer to the screenshots provided in this presentation.
- The SpaceX Dash App is an interactive dashboard that remains operational while running from a server, but fades away when the connection is closed. Therefore, it cannot be kept operational indefinitely in this exercise. If you want to interact with the SpaceX Dash App, I have left its source code on GitHub so you can deploy it and perform your analysis.

Thank you!