IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Radoslav Mitrov
31.07.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using Space X API
  - Data Collection using Python Beautiful Soup for Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis using SQL
  - Exploratory Data Analysis using Python Pandas and Matplotlib
  - Space X Launch Sites Visual Analytics and Dashboard
  - Machine Learning Landing Prediction
- Summary of all results
  - EDA results
  - Visual Analytics and Dashboards
  - Predictive Analysis (Classification)

Introduction

- Project background and context
  - Nowadays there are a great deal of companies such as SpaceX, Virgin Galactic, Rocket Lab, Blue Origin etc. that aim to make space travel affordable for everyone. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, whereas other providers cost upwards of 165 million dollars each. The main advantage of SpaceX is that it can reuse the first stage. As a result, if we can determine whether the first stage will land, we can determine the cost of a launch.
- Problems you want to find answers
  - The goal of this Projects is to find out whether Falcon 9 first stage will land successfully using Space X launch data.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Description of how the data was collected:
  - Using SpaceX API, specifically the SpaceX REST API that will give us data about launches, including information about the rocket used, payload delivered, launch specifications and landing outcome. We will perform a get request using the requests library to obtain the launch data, which we will use to get the data from the API.
  - The response will be a list of JSON objects, which each represents a launch. In order to make the data more consistent, we will convert the JSON to a Pandas DataFrame.
  - Another Data source for obtaining Falcon 9 launch data is Web scraping related Wiki pages. I will use Python BeautifulSoup package to web scrape some HTML tables, parse the data from those tables and convert them into a Pandas DataFrame for further visualization and analysis.

Data Collection – SpaceX API

- Data collection using SpaceX REST API and converting the JSON objects a Pandas DataFrame

- GitHub URL of the completed notebook here.

Task 1: Request and parse the SpaceX launch data using the GET request ¶

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
df1 = response.json()
data = pd.json_normalize(df1)
```

Data Collection - Scraping

- Performing Web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled “List of Falcon 9 and Falcon Heavy launches” using Beautiful Soup and converting them into a Pandas DataFrame
- GitHub URL of the completed notebook here

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

Next, request the HTML page from the above URL and get a `response` object

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
soup = BeautifulSoup(response.content,'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly soup.title

```
soup.title
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

Data Wrangling

- Performing some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- Creating a landing outcome label from Outcome column to represent if Falcon 9 first stage will land successfully or not.
- GitHub URL of the completed notebook [here]

TASK 4: Create a landing outcome label from Outcome column

Using the `Outcome`, create a list where the element is zero if the corresponding row in `Outcome` is in the set `bad_outcome`; otherwise, it's one. Then assign it to the variable `landing_class`:

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
df['Class'] = df['Outcome'].apply(lambda x: 0 if x in bad_outcomes else 1)
df['Class'].value_counts()
```

```
1    60
0    30
Name: Class, dtype: int64
```

This variable will represent the classification variable that represents the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed Successfully

```
# df['Class']=landing_class
df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

EDA with Data Visualization

- Performing Exploratory Data Analysis (EDA) and Feature Engineering using Pandas and Matplotlib
- Scatter plots were used to see how Flight Number, Payload, Launch sites and Orbit type would affect the Launch outcome
- Bar chart was created to check the relationship between success rate and orbit type
- Line chart was used to observe the success rate per each year.
- GitHub URL of the completed notebook here

EDA with SQL

- Performed SQL queries displaying:
  1. The names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT LAUNCH_SITE as "LaunchSite" FROM SPACEXTBL;
```

  2. 5 records where launch sites begin with the string 'CCA'

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

  3. The total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "TOTAL PAYLOAD MASS IN KG", CUSTOMER FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
```

  4. Average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS "AVG PAYLOAD MASS IN KG", Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version = 'F9 v1.1
```

  5. The date when the first successful landing outcome in ground pad was achieved

```
%sql SELECT MIN(DATE) as "Date of first succesful landing", Landing_Outcome FROM 'SPACEXTBL' WHERE Landing_Outcome = 'Success(ground pad)';
```

EDA with SQL

6. 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 'SPACEXTBL' WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

7. The total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT ("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

8. The names of the booster_versions, which have carried the maximum payload mass using a subquery

```
%sql SELECT Booster_Version,Payload, PAYLOAD_MASS__KG_ FROM 'SPACEXTBL' WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM 'SPACEXTBL');
```

9. The records displaying the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015

```
%sql SELECT  substr(Date,7,4), substr(Date, 4, 2), Booster_Version,PAYLOAD_MASS__KG_, Launch_site, Landing_Outcome FROM
'SPACEXTBL' WHERE substr(Date,7,4)='2015' AND Landing_Outcome = 'Failure (drone ship)';
```

10. 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 * from 'SPACEXTBL' WHERE "Landing_Outcome" LIKE "Success (ground pad)%" and (Date BETWEEN '04-06-2010' AND '20-03-2017') ORDER BY Date DESC;
```

- GitHub URL of the completed notebook here

Build an Interactive Map with Folium

- Map objects such as markers, circles, lines, were created and added to a folium map in order to mark all launch sites on a map, to mark the success/failed launches for each site on the map and to calculate the distances between a launch site to its proximities
- GitHub URL of the completed notebook here

Build a Dashboard with Plotly Dash

- Plotly Dash application was built to perform interactive visual analytics on SpaceX launch data by:
  - Adding a Launch Site Drop-down Input Component
  - Adding a callback function to render success-pie-chart based on selected site dropdown
  - Adding a Range Slider to Select Payload
  - Add a callback function to render the success-payload-scatter-chart scatter plot
- GitHub URL of the completed notebook here

Predictive Analysis (Classification)

In order to predict if the first stage of the Falcon 9 will land successfully, I created a Machine learning pipeline including:

- Preprocessing, allowing to standardize the data
- Train_test_split, allowing to split the data into training and testing data setting the parameter test_size to 0.2 and random_state to 2.
- Train the model and perform Grid Search, allowing to find the hyperparameters that allow a given algorithm to perform best. Using the best hyperparameter values, will be determined the model with the best accuracy using the training data.
- Use of Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors. Finally, we will output the confusion matrix.
- GitHub URL of the completed notebook here

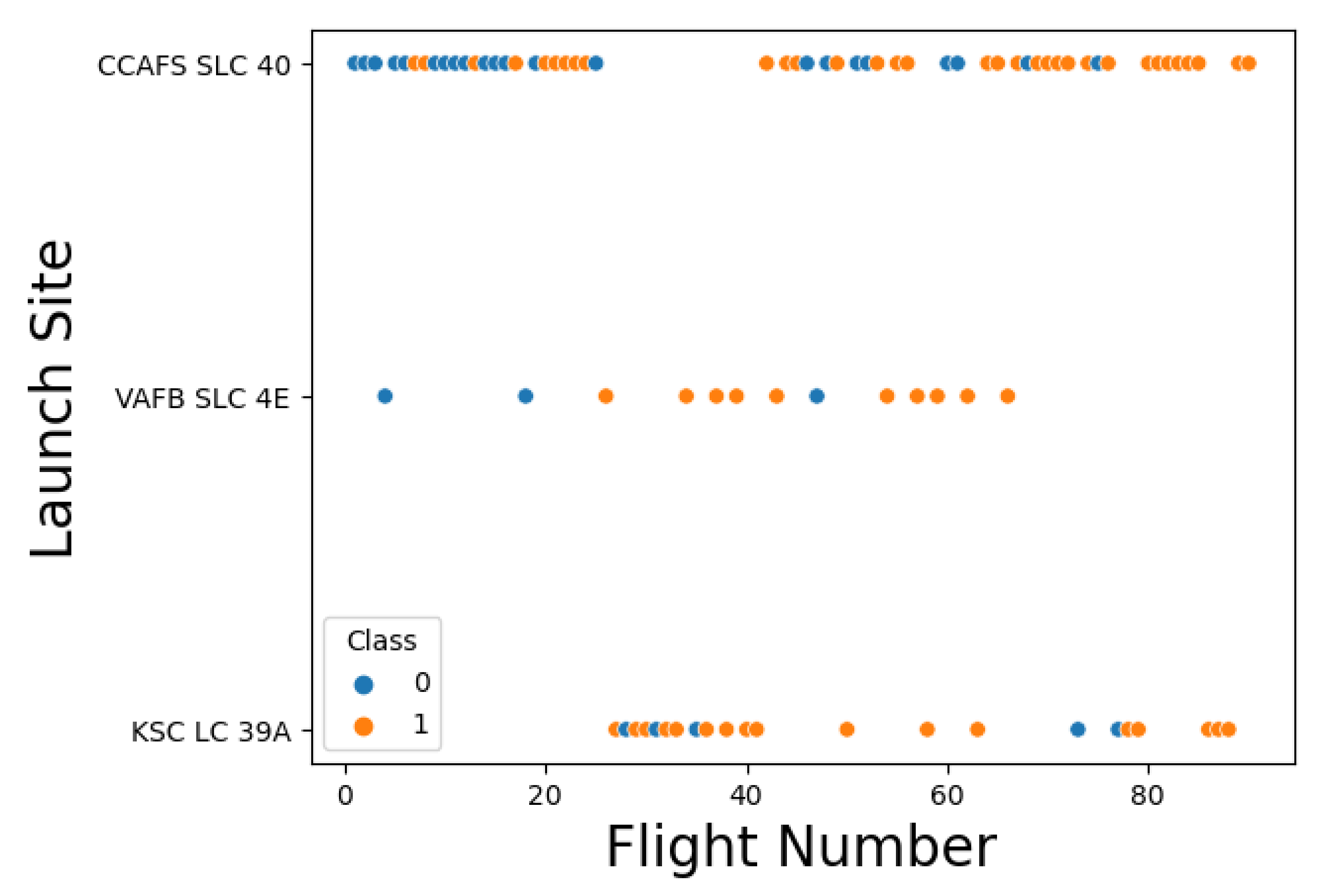
Results

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

Section 2

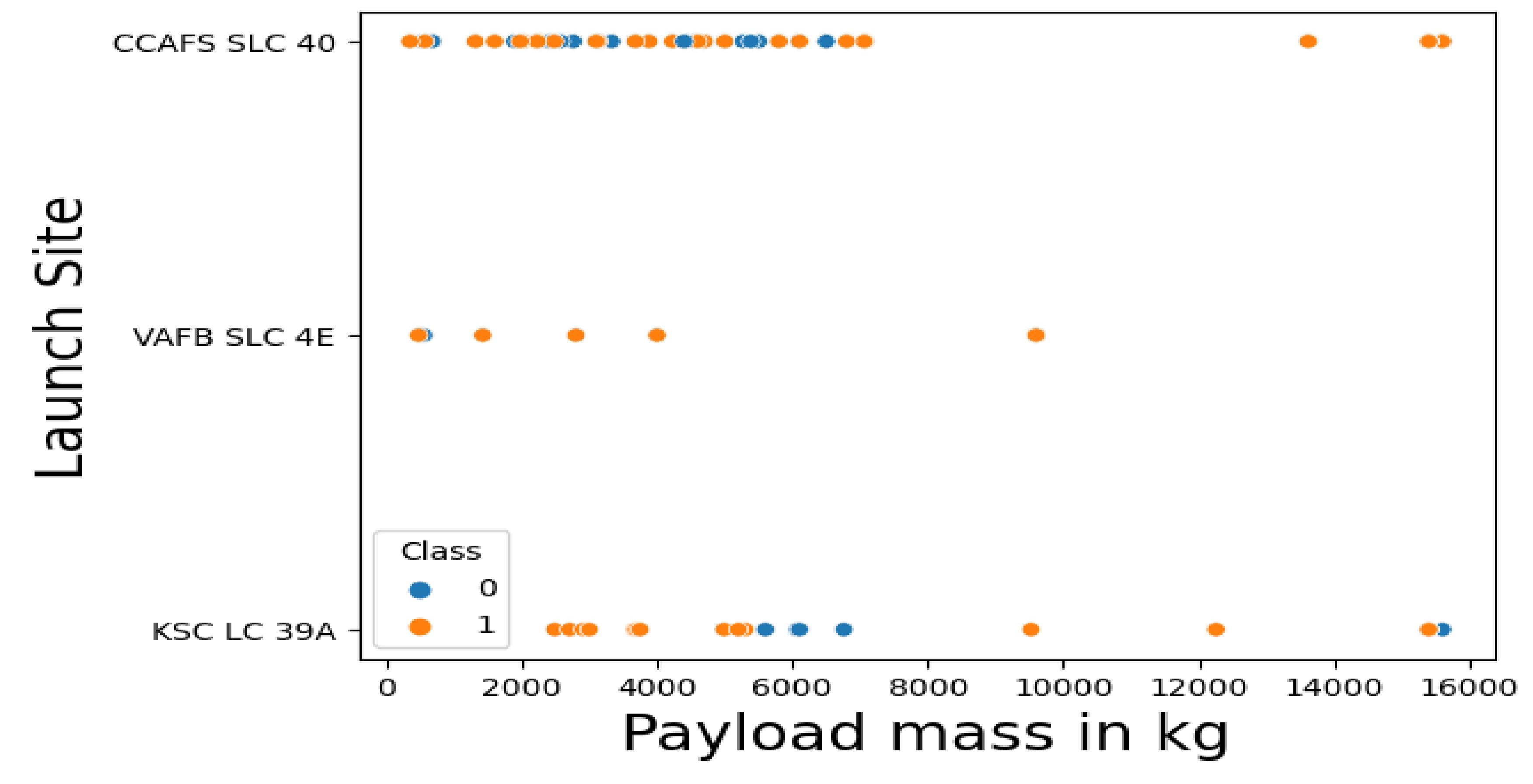
Insights drawn from EDA

Flight Number vs. Launch Site



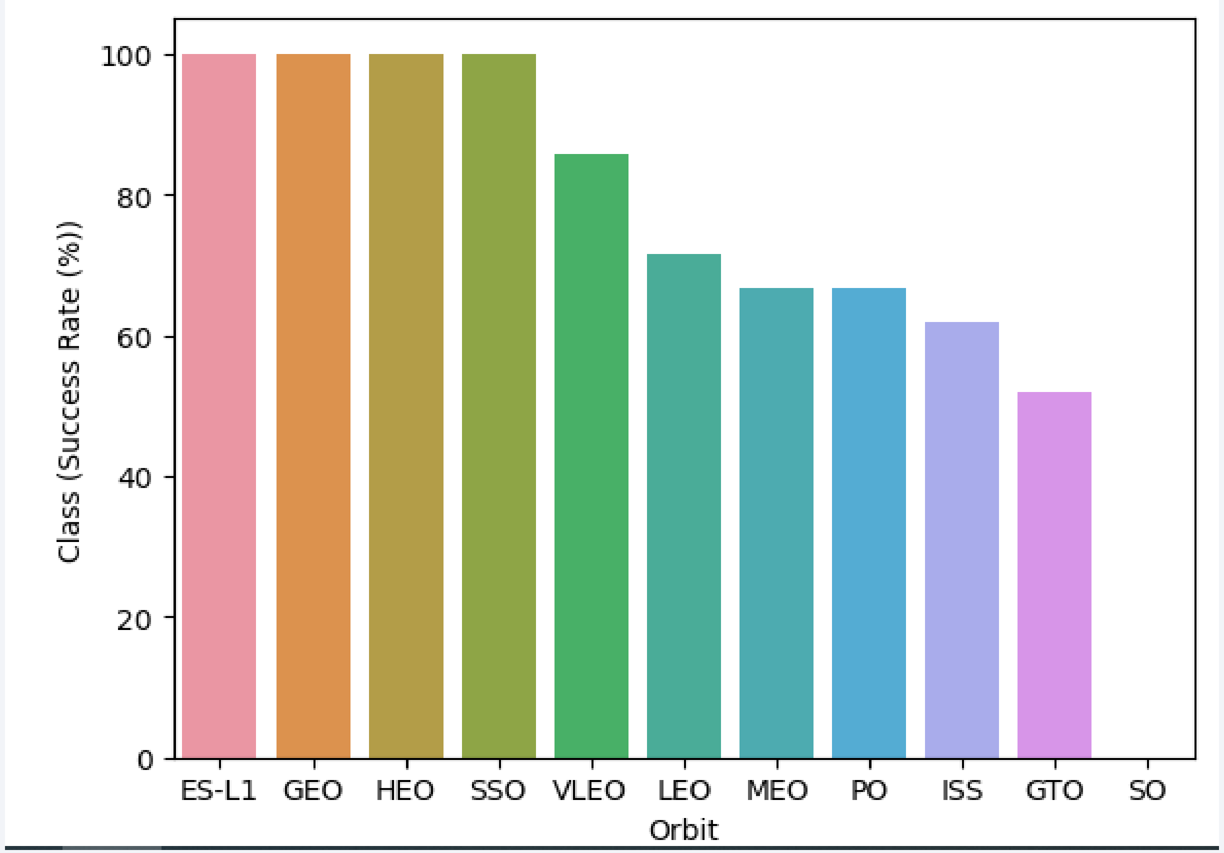
The scatter plot shows that with the increased number of flights, increases also the success rate from all Launch sites

Payload vs. Launch Site



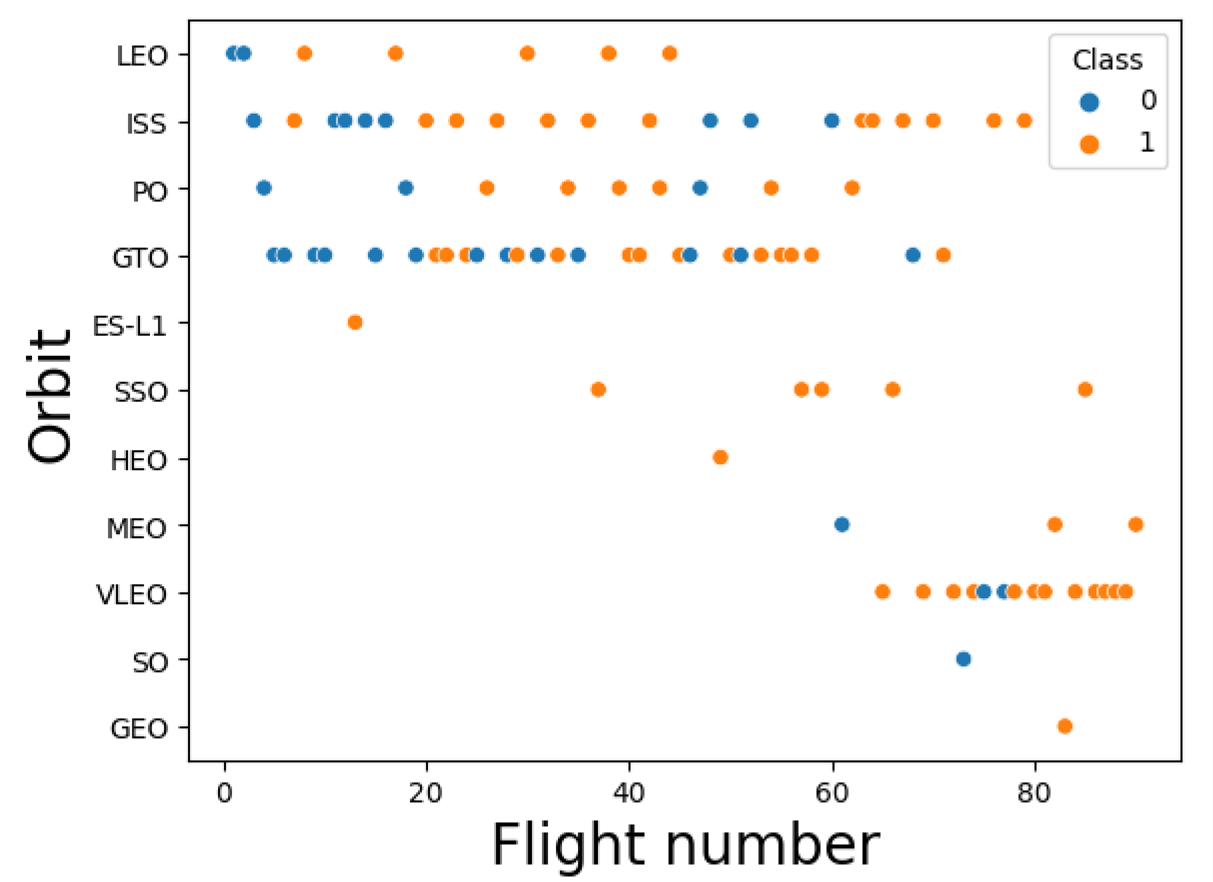
The scatter plot shows that the Launch sites with most Payload mass in kg per launch are CCAFS SLC 40 and KSC LC 39A. However, there can be seen a great deal of unsuccessful launches. On the other hand, the success rate of Launch site VAFB SLC 4E is higher but the Payload mass in kg is considerably lower.

Success Rate vs. Orbit Type



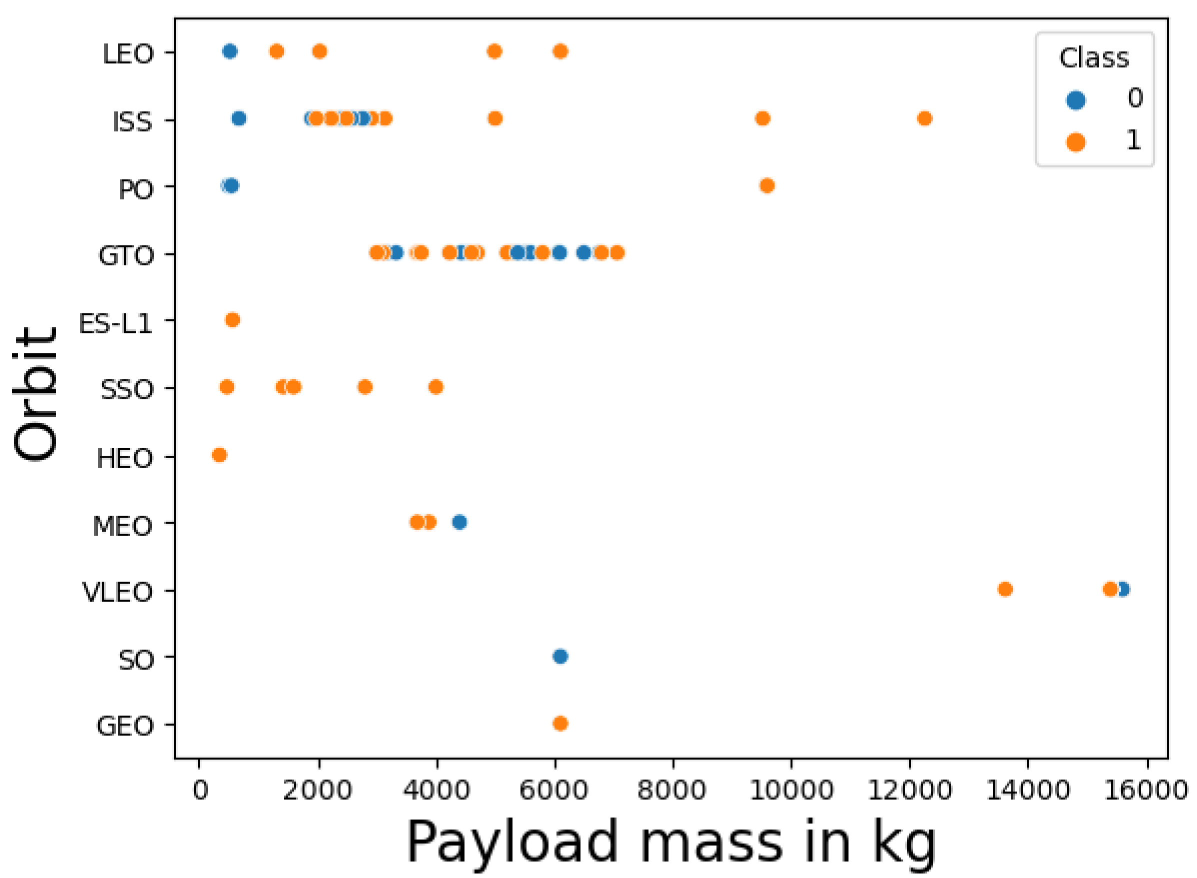
The bar plot shows that the Success rate of 4 Orbits is 100% - ES-L1, GEO, HEO and SSO. The next 6 Orbits's Success rate vary from 80% to 50% and there is one Orbit SO with 0% success.

Flight Number vs. Orbit Type



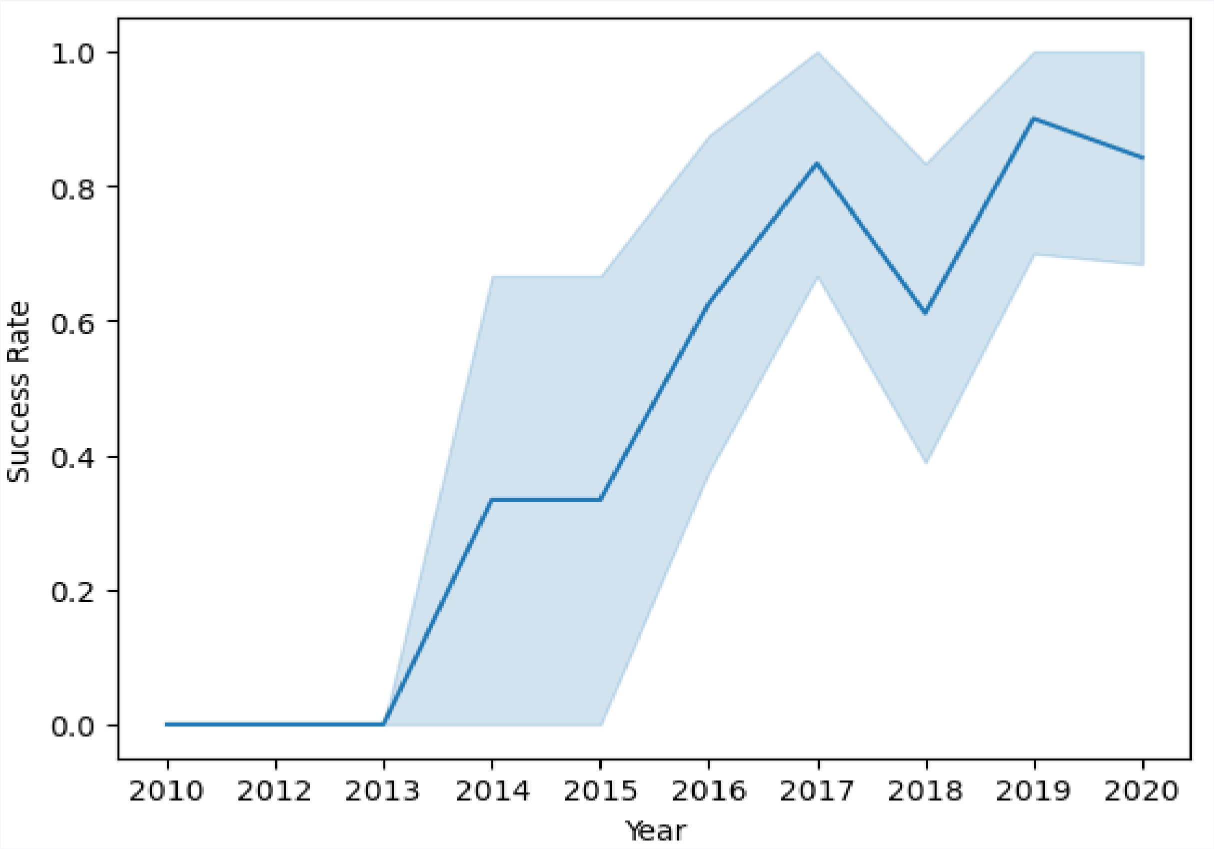
The scatter plot shows that there are orbits such as LEO, where the increased number of flights leads to increased success rate. Alternatively, there are orbits like GTO, where appears to be no relationship between the orbit and the

Payload vs. Orbit Type



The scatter plot shows that apart from orbits GTO, MEO, VLEO and SO, the increased Payload mass in kg the other orbits generate higher success rate.

Launch Success Yearly Trend



The line chart shows that since 2013 the Success rate of launches grows significantly.

All Launch Site Names

```
%sql SELECT DISTINCT LAUNCH_SITE as "LaunchSite" FROM SPACEXTBL;
```

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

LaunchSite
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

The query shows all launch site names.

Launch Site Names 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_Outc
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (paracl
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (paracl
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No att
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No att
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No att

The query shows 5 launch site names that begin with ‘CCA’.

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "TOTAL PAYLOAD MASS IN KG", CUSTOMER FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
```

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

TOTAL PAYLOAD MASS IN KG	Customer
45596.0	NASA (CRS)

The query shows total payload mass in kg.

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS "AVG PAYLOAD MASS IN KG", Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version = 'F9 v1.1
```

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

AVG PAYLOAD MASS IN KG	Booster_Version
2928.4	F9 v1.1

The query shows that the average payload mass in kg by Falcon 9 v1.1 is 2928.4kg.

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) as "Date of first succesful landing", Landing_Outcome FROM 'SPACEXTBL' WHERE Landing_Outcome = 'Success (ground pad)';
```

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

Date of first succesful landing	Landing_Outcome
01/08/2018	Success (ground pad)

The query shows that the first successful ground landing was on 01.08.2018.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Booster_Version FROM 'SPACEXTBL' WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_BETWEEN 4000 AND 6000;
```

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

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

The query shows Booster version of successful drone ship landing with payload mass in kg between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT "Mission_Outcome", COUNT ("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

Mission_Outcome	Total
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

The query shows total number of successful and failure mission outcomes.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version,Payload, PAYLOAD_MASS__KG_ FROM 'SPACEXTBL' WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM 'SPACEXTBL');
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600.0
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600.0
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600.0
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600.0
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600.0
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600.0
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600.0
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600.0
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600.0
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600.0
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600.0
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600.0

The names of the Booster versions, which have carried the maximum payload mass using a subquery

2015 Launch Records

```
%sql SELECT  substr(Date,7,4), substr(Date, 4, 2), Booster_Version,PAYLOAD_MASS__KG_, Launch_site, Landing_Outcome FROM
'SPACEXTBL' WHERE substr(Date,7,4)='2015' AND Landing_Outcome = 'Failure (drone ship)';
```

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

substr(Date,7,4)	substr(Date, 4, 2)	Booster_Version	PAYLOAD_MASS__KG_	Launch_Site	Landing_Outcome
2015	10	F9 v1.1 B1012	2395.0	CCAFS LC-40	Failure (drone ship)
2015	04	F9 v1.1 B1015	1898.0	CCAFS LC-40	Failure (drone ship)

Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

-
```
%sql SELECT * from 'SPACEXTBL' WHERE "Landing_Outcome" LIKE "Success (ground pad)%" and (Date BETWEEN '04-06-2010' AND '20-03-2017') ORDER BY Date DESC;
```

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

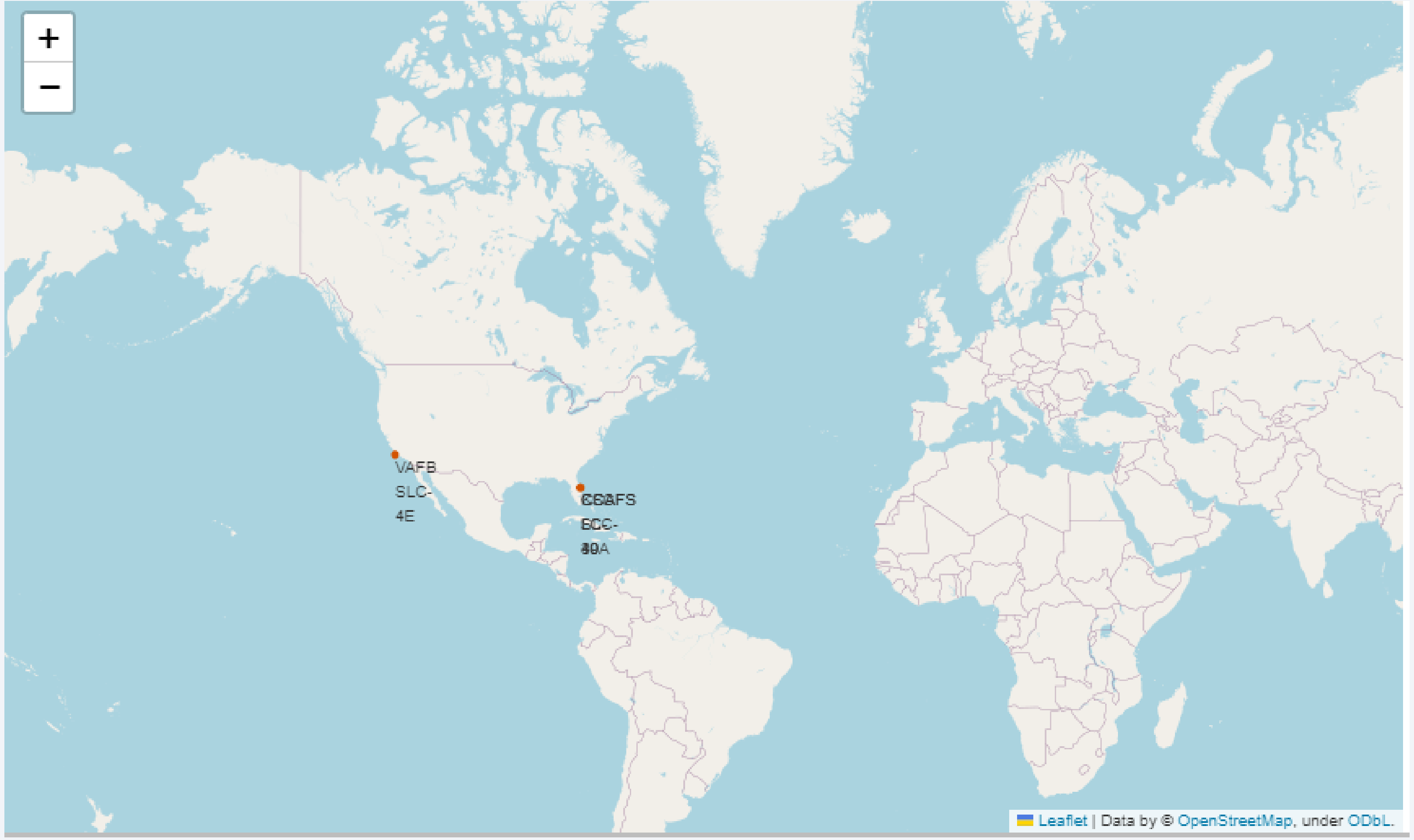
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
19/02/2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18/07/2016	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
15/12/2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14/08/2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
09/07/2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990.0	LEO	U.S. Air Force	Success	Success (ground pad)
06/03/2017	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LEO	NRO	Success	Success (ground pad)

The count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Section 3

Launch Sites Proximities Analysis

Launch sites' location markers on a global map

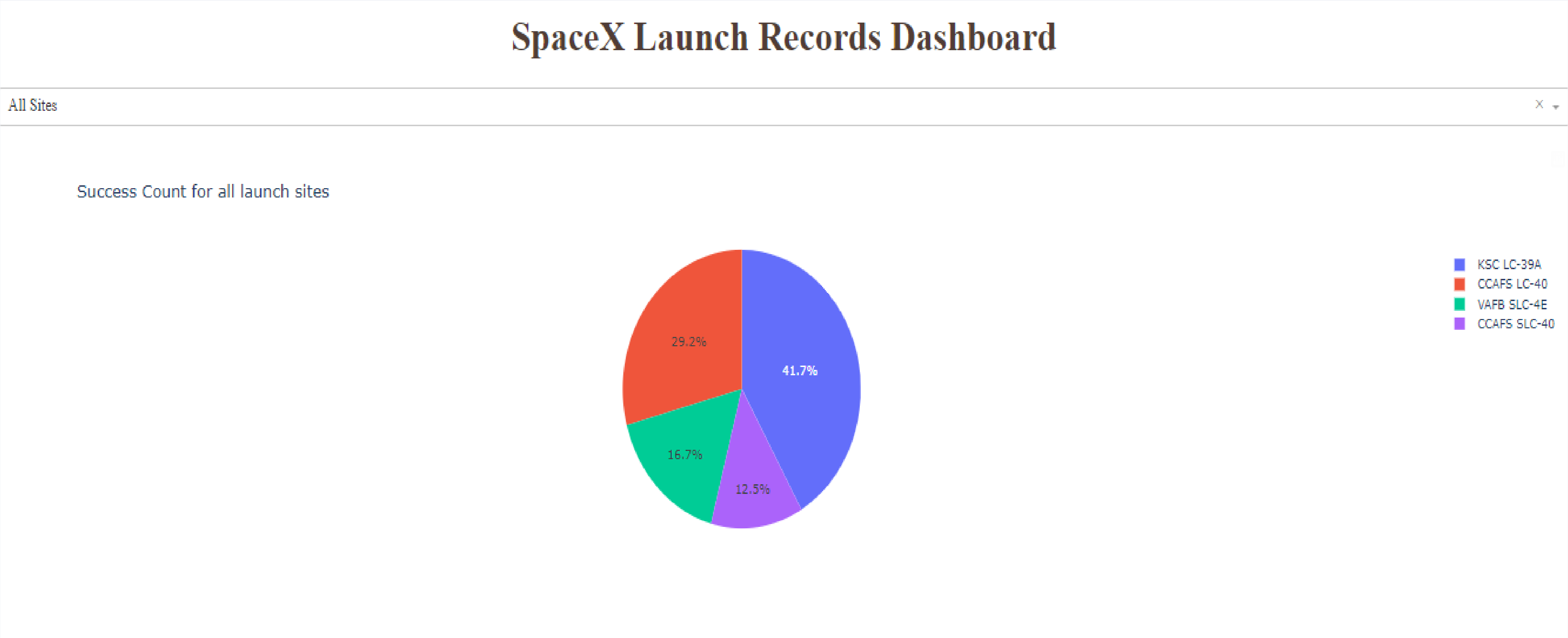


We can observe that all launch sites are located very close to the coast and to the Equator line

Section 4

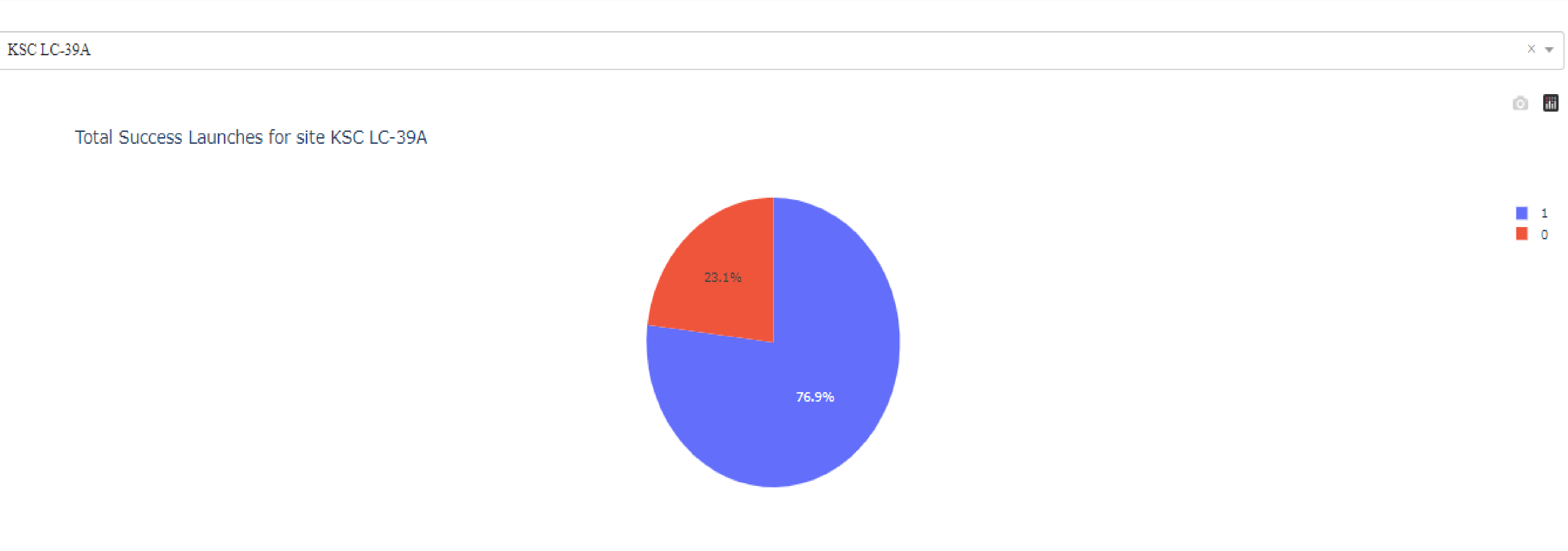
Build a Dashboard with Plotly Dash

Pie chart Success count for all launch sites



The Pie chart for the Success Count for all launch sites shows that the launch site with highest launch success ratio is KSC LC-39A with more than 41% of the total successful SpaceX launches. On the other hand, CCAFS SLC-40 is the launch site with the least successful launces with12% of the total launches.

Pie chart for the launch site with highest launch success ratio



The Pie chart for the launch site with highest launch success ratio KSC LC-39A shows that almost 77% of the launches from this site (10 out of 13) were successful and only 23% (3 out of 13) were not.

Scatter plots on Payload vs Launch Outcome

Scatter plots on Payload vs Launch Outcome

Section 5

Predictive Analysis (Classification)

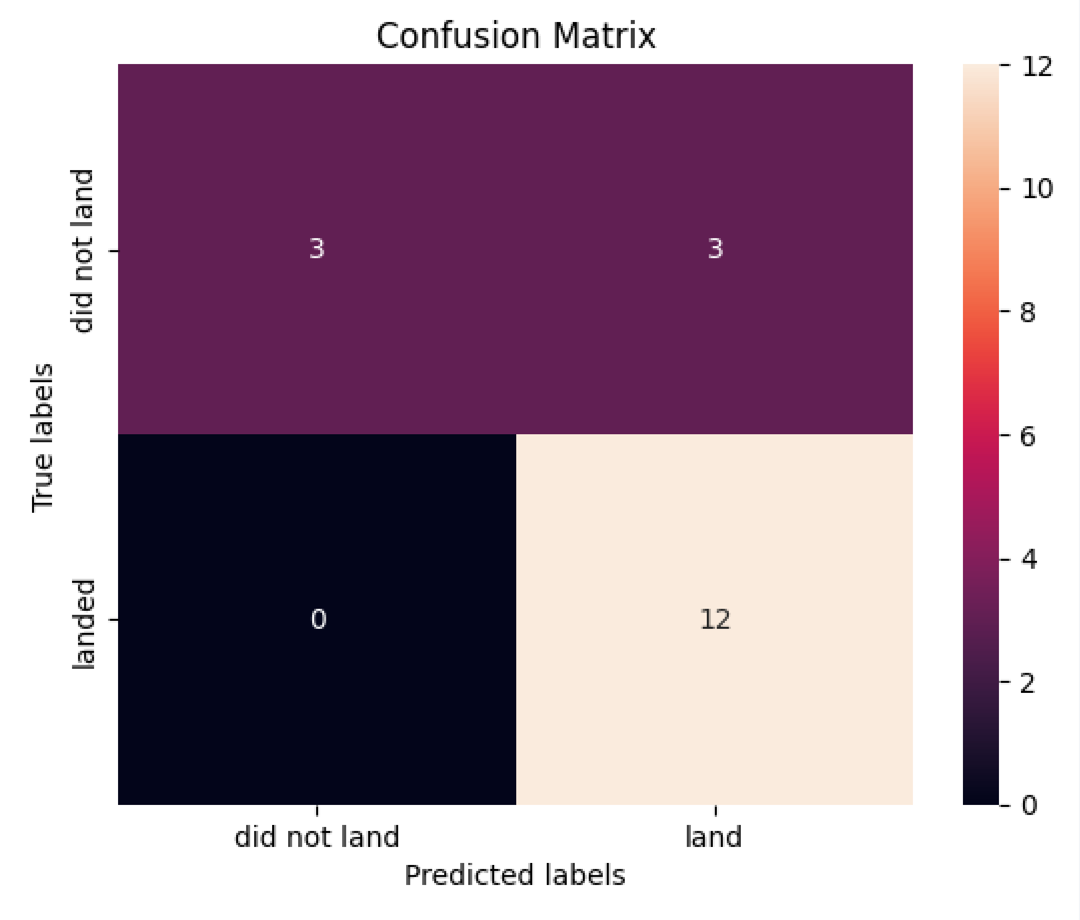
Classification Accuracy

- Logistic Regression, Support Vector Machine (SVM), Decision Tree and K Nearest Neighbors (KNN) Find have the same classification accuracy of 0,83%.

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- All models are performing the same and have the same Confusion Matrix.
- Examining the confusion matrix, it can be seen that the major problem is False positives.



Conclusions

Taking into consideration the performed data analysis on SpaceX Falcon 9 we could draw the following conclusions:

- All launch sites are located very close to the coast and to the Equator line.
- Since 2013 the number of flights increased from all Launch sites and as a result the Success rate from all Launch sites increased significantly. In 2020 the Success rate reached above 80%.
- The Launch sites with most Payload mass in kg per launch are CCAFS SLC 40 and KSC LC 39A with almost 16 000 kg.
- Depending on the Orbit types the Success rates vary from 100% in some of the Orbits such as ES-L1 as GEO to 0% in Orbit SO.
- The created Machine learning pipeline predicted that there is more than 83% probability of successful Space X Falcon 9 first stage landing.

Thank you!