IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmed AbuBakr Mekki
10/09/2021

Outline

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

Executive Summary

First, we started with collecting data using the SpaceX API. Then, we scraped the web for collecting data about the Falcon 9 and their outcomes. Afterwards, the data was wrangled and re-arranged. The collected data was visualized in order to find insights and patterns. At last, classification algorithmswere used to find our success rate.

There are many factors affecting the successful return of the 1st stage of the launcher. Relationships between payload mass, launch site location, and orbit type were very influencing. Visualizing the datasets and using the folium map allowed us to have a better understanding of the relationships that exists between the various factors. Different classification methods were used to predict the success of the 1st stage return that resulted in accuracy of %83.33.

Introduction

The commercial space age is here and SpaceX has changed the game by reusing the first stage of launchers, downsizing the expenses to a third (62$ million).

What we would like to do is to predict the successful reuse of the 1st stage of the launcher in order to estimate the cost of space missions.

Section 1

Methodology

Methodology

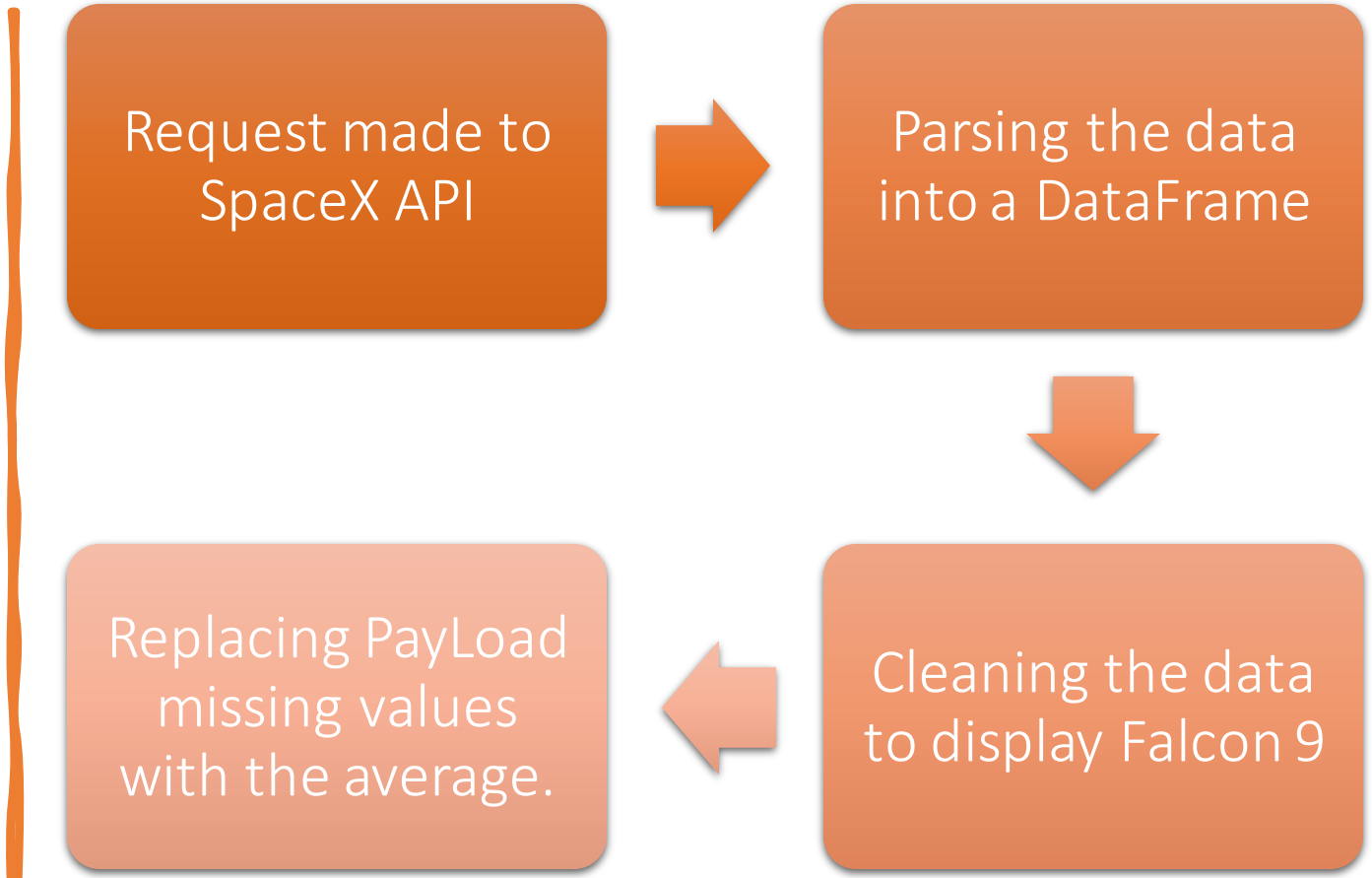
- Executive Summary
- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

- A request was made to the SpaceX API in order to get information about the rocket launches.
- The data was cleaned and arranged in a matter to display the Falcon 9 usage.
- A DataFrame containing various information such as the payload mass, launch site and the location was made as it will help in the analysis.

Data Collection – SpaceX API

Data Collection - SpaceX API

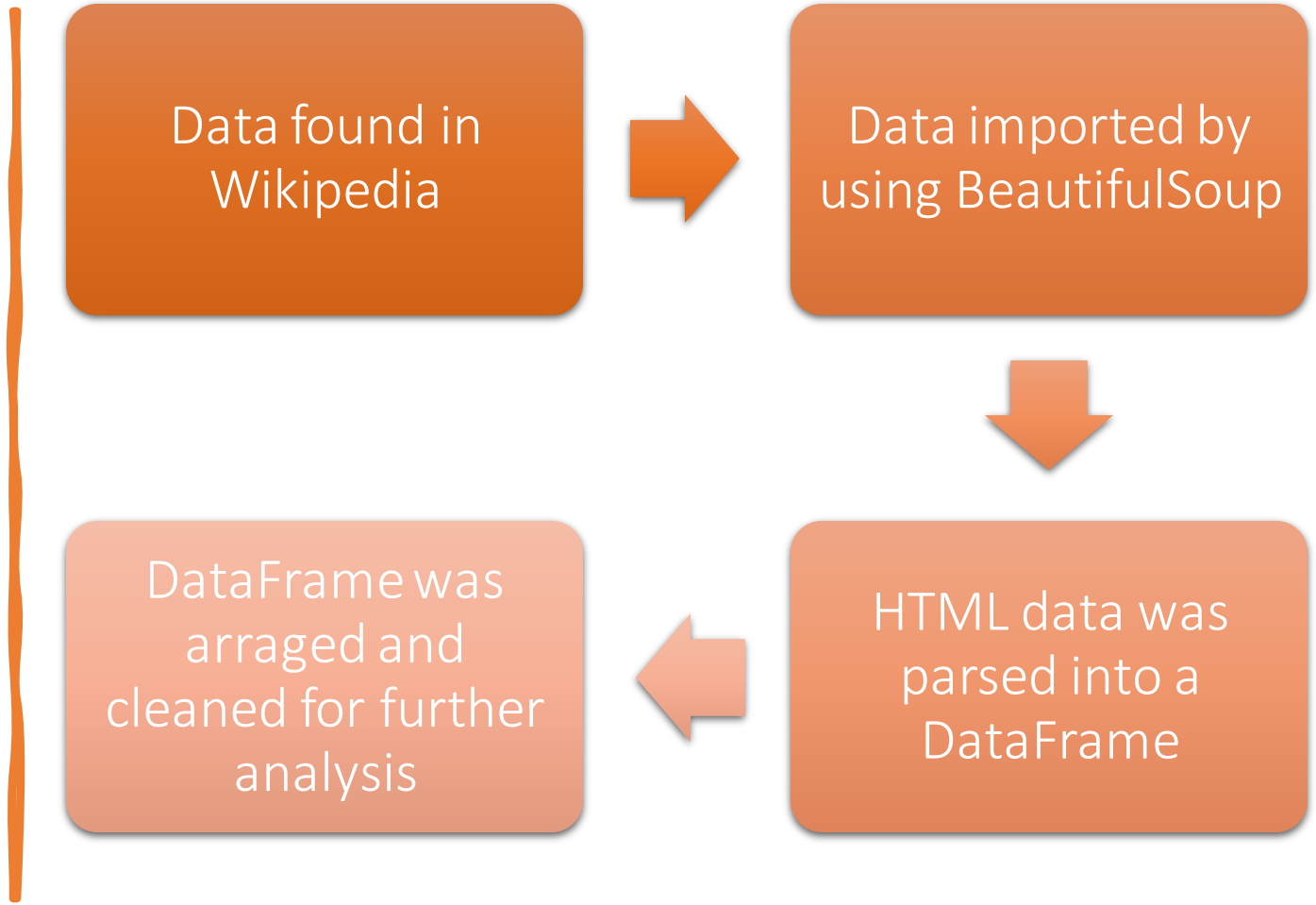


Data Collection – Scraping

- The historical launch records of the Falcon 9 was scraped from HTML Wikipedia.
- Then the HTML data was parsed into a Pandas DataFrame.
- The DataFrame was cleaned and arranged to display the features required.

Data Collection – Scraping

Data Collection - Scraping

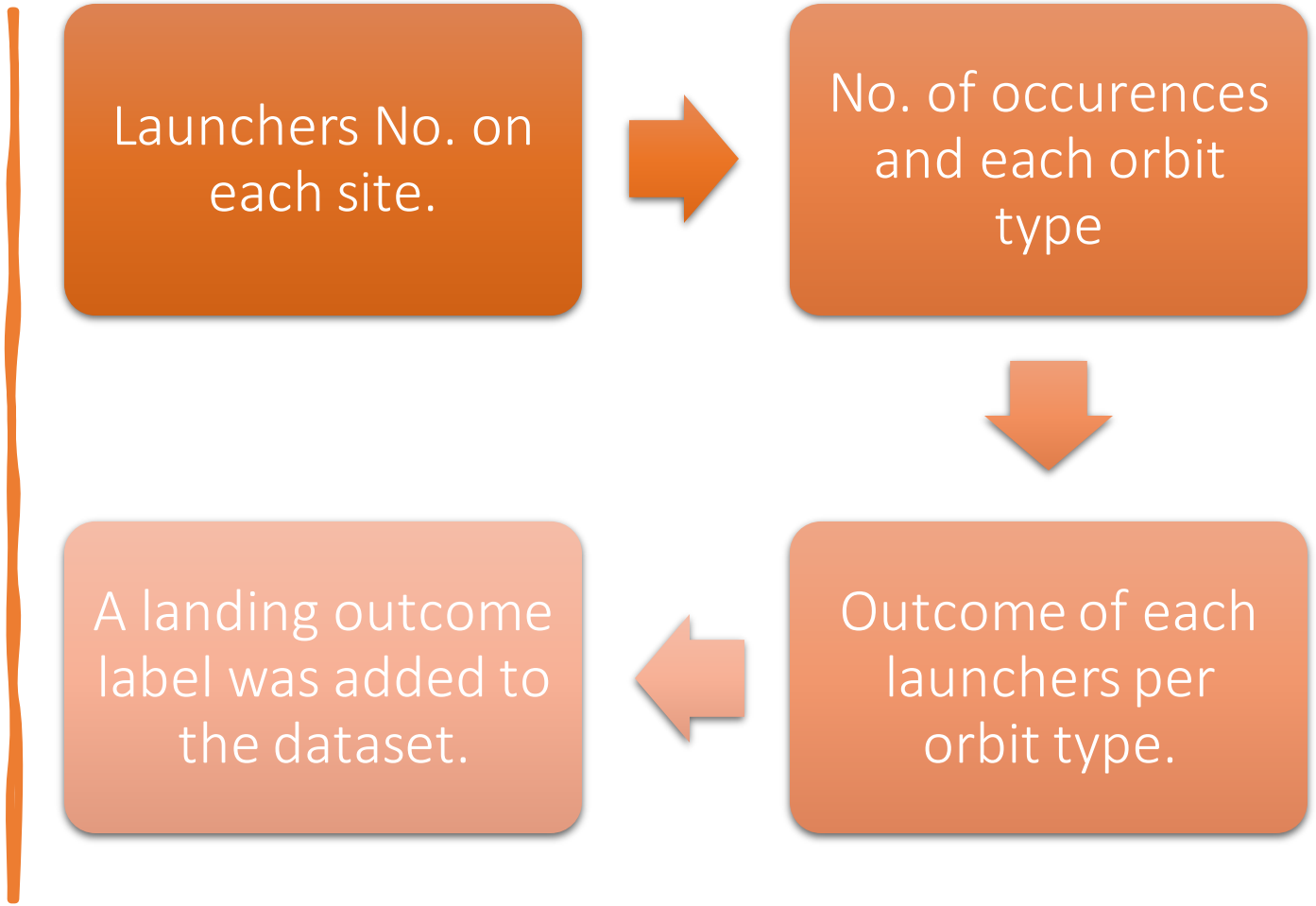


Data Wrangling

- The goal is to find some patterns in the data and determine what would be the label for training supervised models.
- Caluclations of the number of launches of each site and their landing outcome were made.
- A display of the launchers outcome in a labeled manner was added to the data to help predicting.

Data Wrangling

Data Wrangling



EDA with Data Visualization

- A scatter-plot of the number of flights and payload was created as it showed when the flight number increases, the first stage is more likely to land successfully.
- Also, when the payload mass increases, the more likely the first stage won't return.
- Each launcher and launch site has different success rates.
- A line plot displayed the imporvement in successful return of the 1st stage of launchers since 2013.
- EDA with Data Visualization

EDA with SQL

- Display of the different launch sites that are used.
- The date of the first successful landing on a ground pad.
- Which booster version was success in drone ship landing with a mass between 4000 and 6000 kg.
- Booster versions that carried the most payload.
- EDA with SQL

Build an Interactive Map with Folium

- Markers of all the launch sites were created on the folium map.
- The launch outcome (fail/success) were added to each site on the folium map.
- Then a calculation of the distance between each site and the railway and coastline were made.
- All the launch sites are on the coastlines.
- Folium Map

Build a Dashboard with Plotly Dash

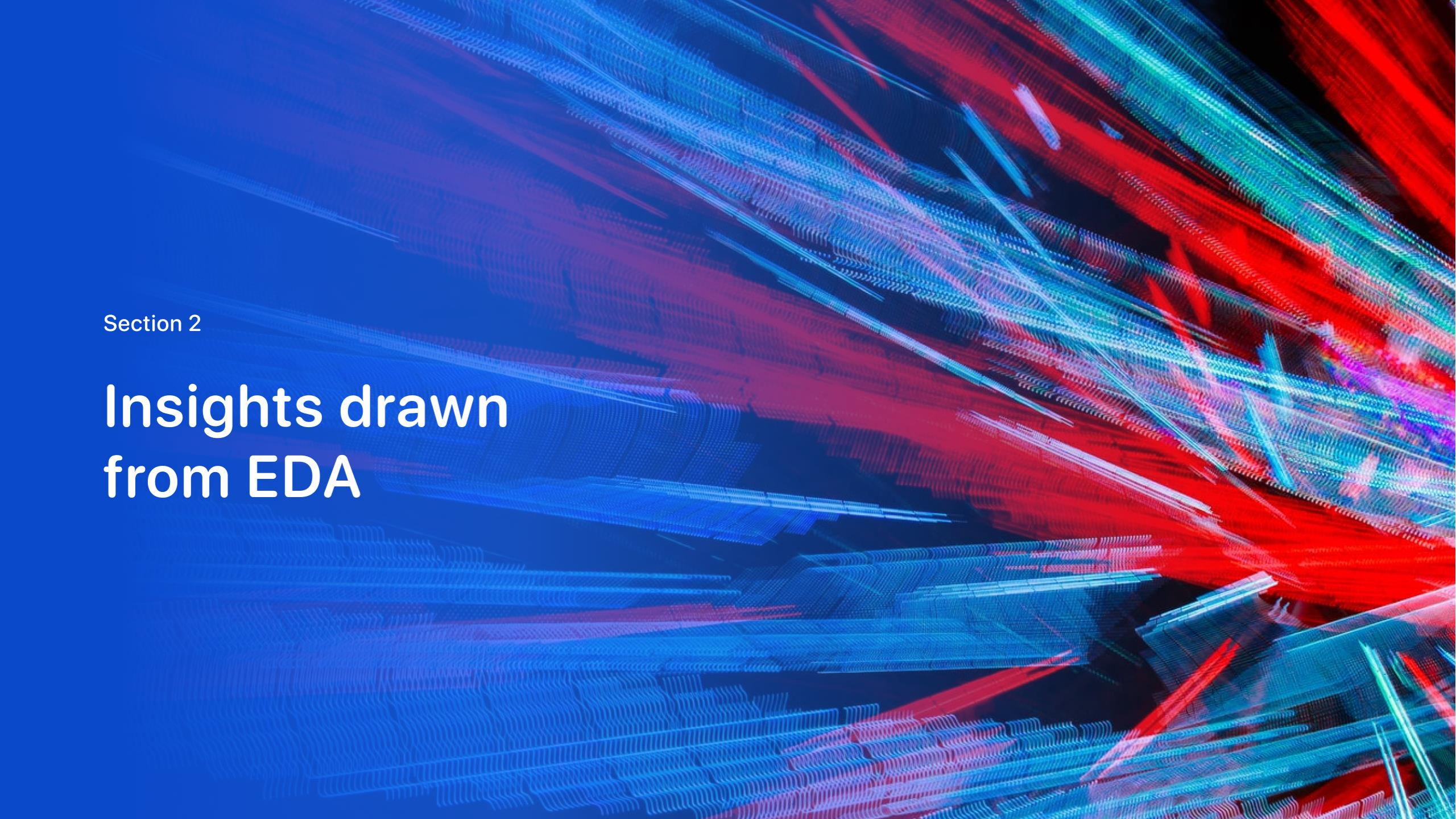
- A dropdown containing the four launch sites was created.
- A pie chart displaying the success of each launch site was added.
- A range-slider to control the payload was added to see how does it affect the mission outcome.
- A scatter plot showing the payload mass and the booster version outcomes from each site.
- Plotly Dashboard

Predictive Analysis (Classification)

- Four different classification methods were tested in order to help predict the success of the launcher's 1st stage return.
- Logistic Regression, Support Vector Machine(SVM), Decision Tree and K-Nearest Neighbors(KNN) were applied.
- An evaluation of each algorithm performance with the best parameters was obtained using GridSearchCV.
- The four methods yielded the same accuracy.
- Predictive Analysis (Classification)

Results

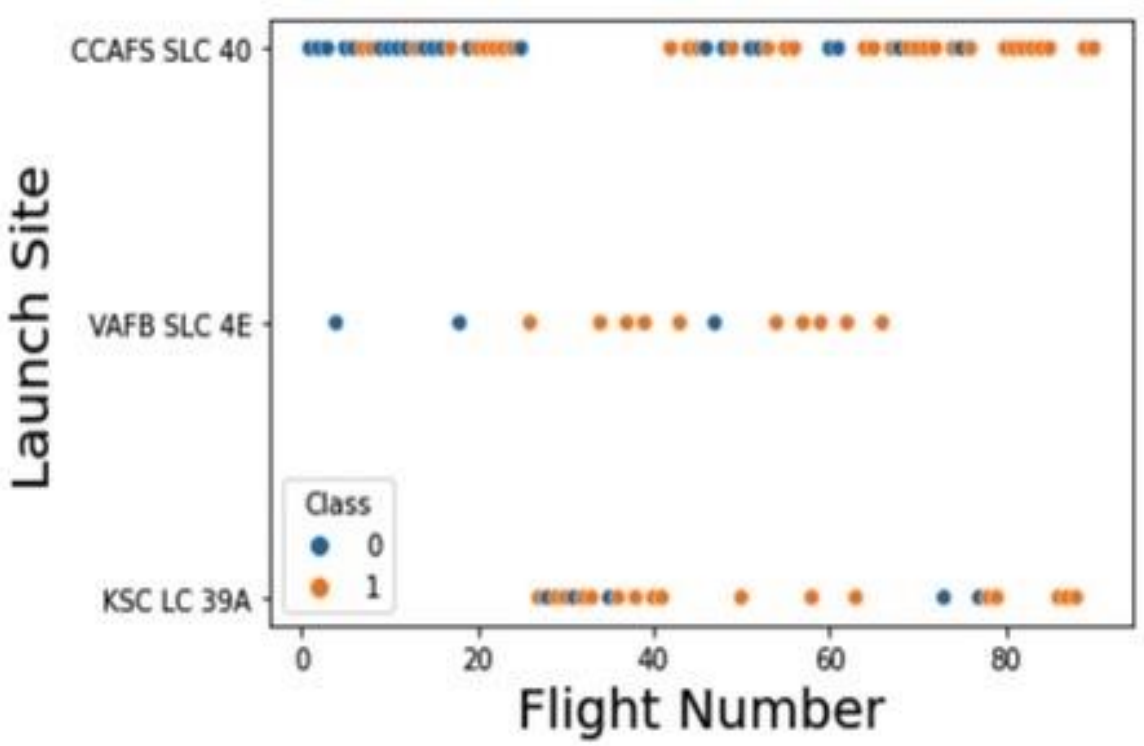
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

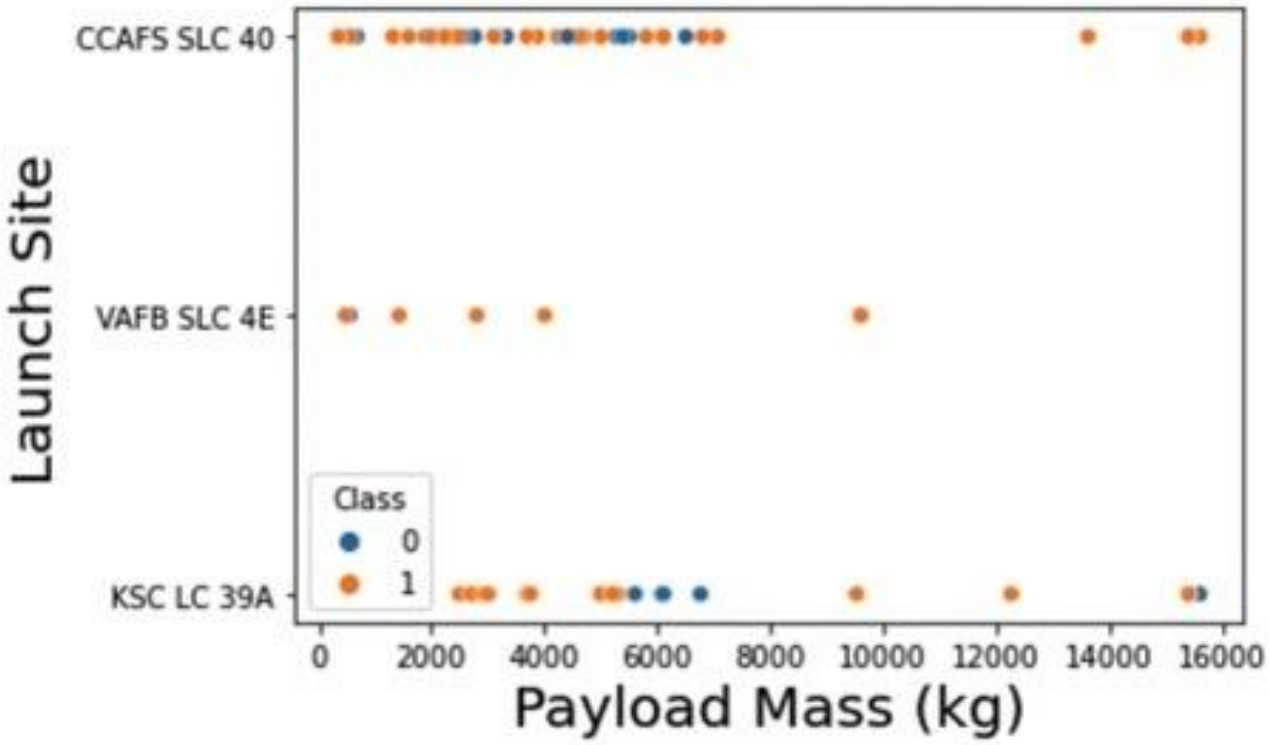


The scatter plot shows a relation between the number flights made from each launch site location and their outcomes.

There's a higher success rate associated with the launch site with higher number of launches.

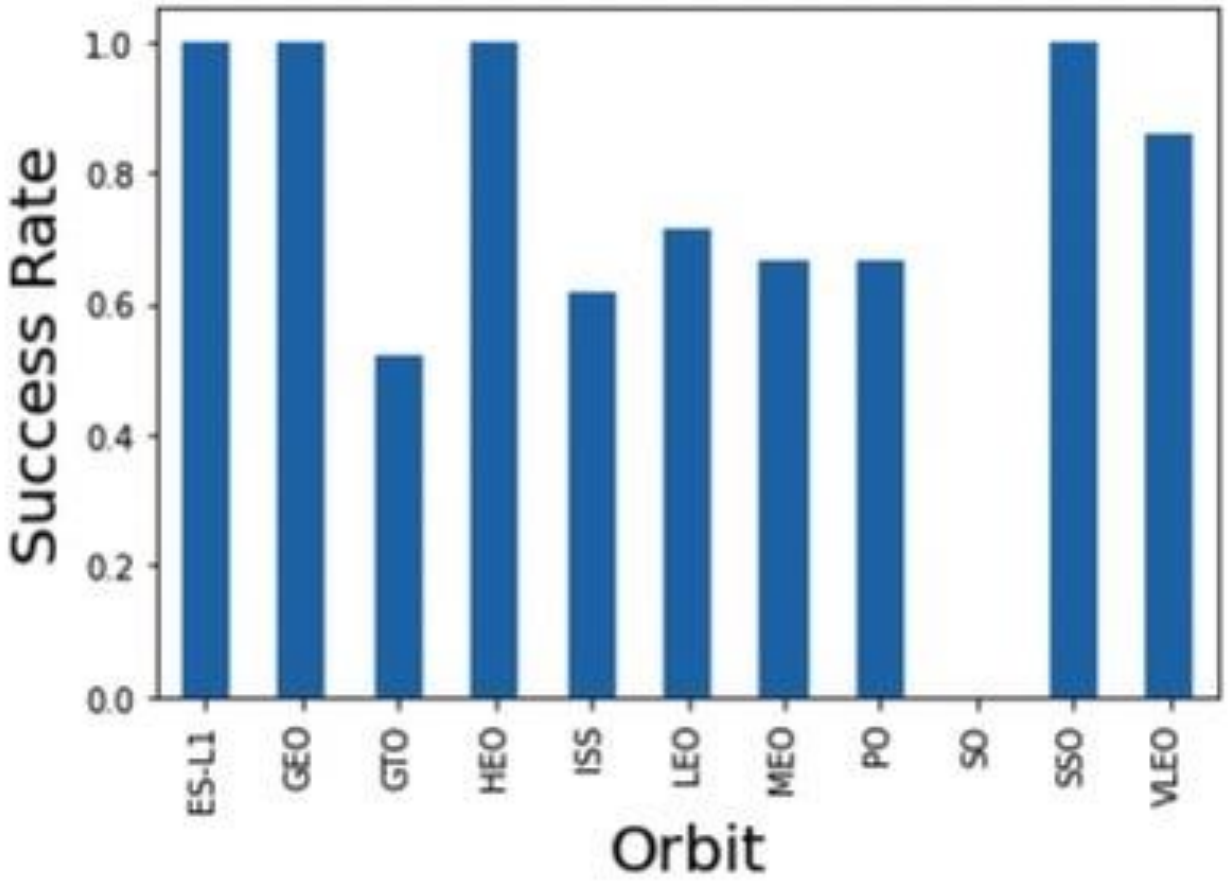
Payload vs. Launch Site

- The scatter plot displays a relation between the payload mass of each launch site and their mission outcome.
- The more massive the payload mass, the less likely the 1st stage will return.



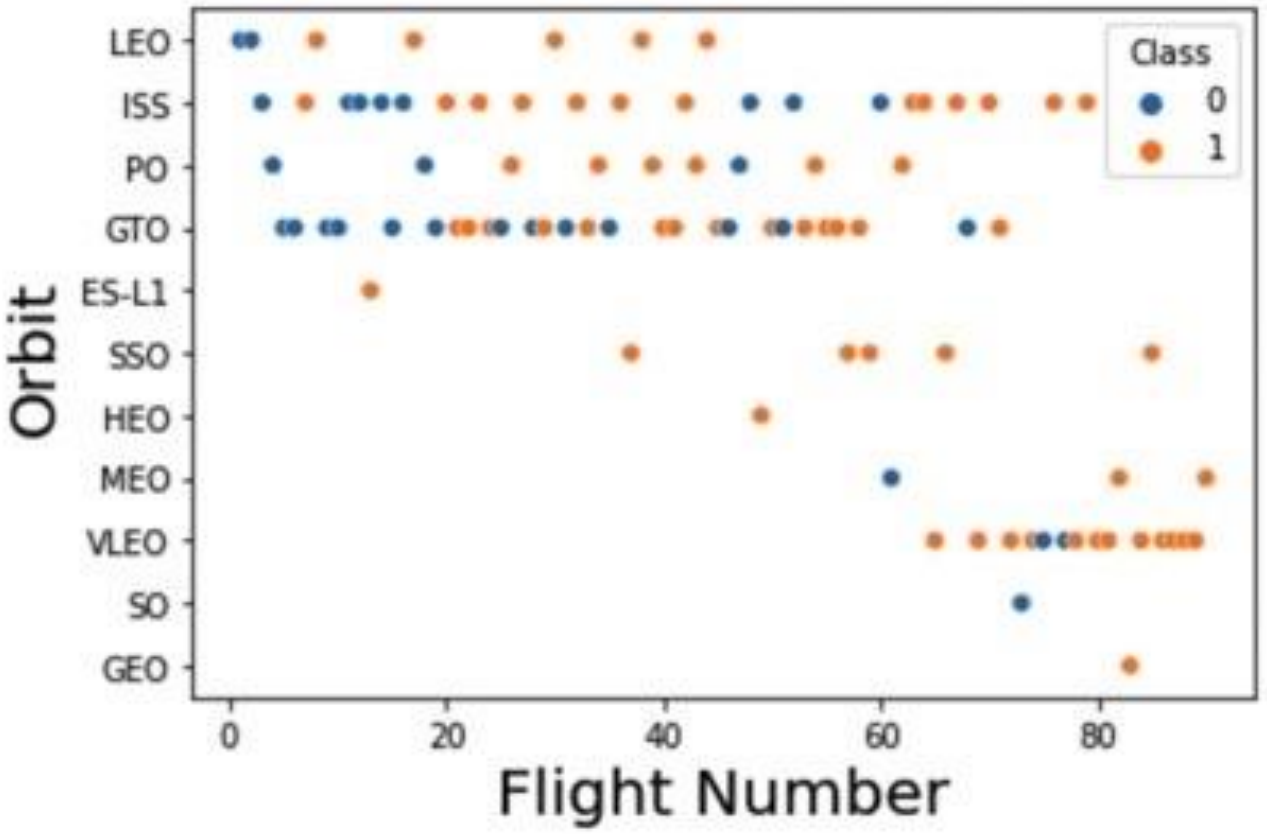
Success Rate vs. Orbit Type

- A bar chart visualizing the relation between the orbit type and the successful return of the 1st stage of the launcher.
- It is displayed clearly that some orbit types have a higher success rate than others.



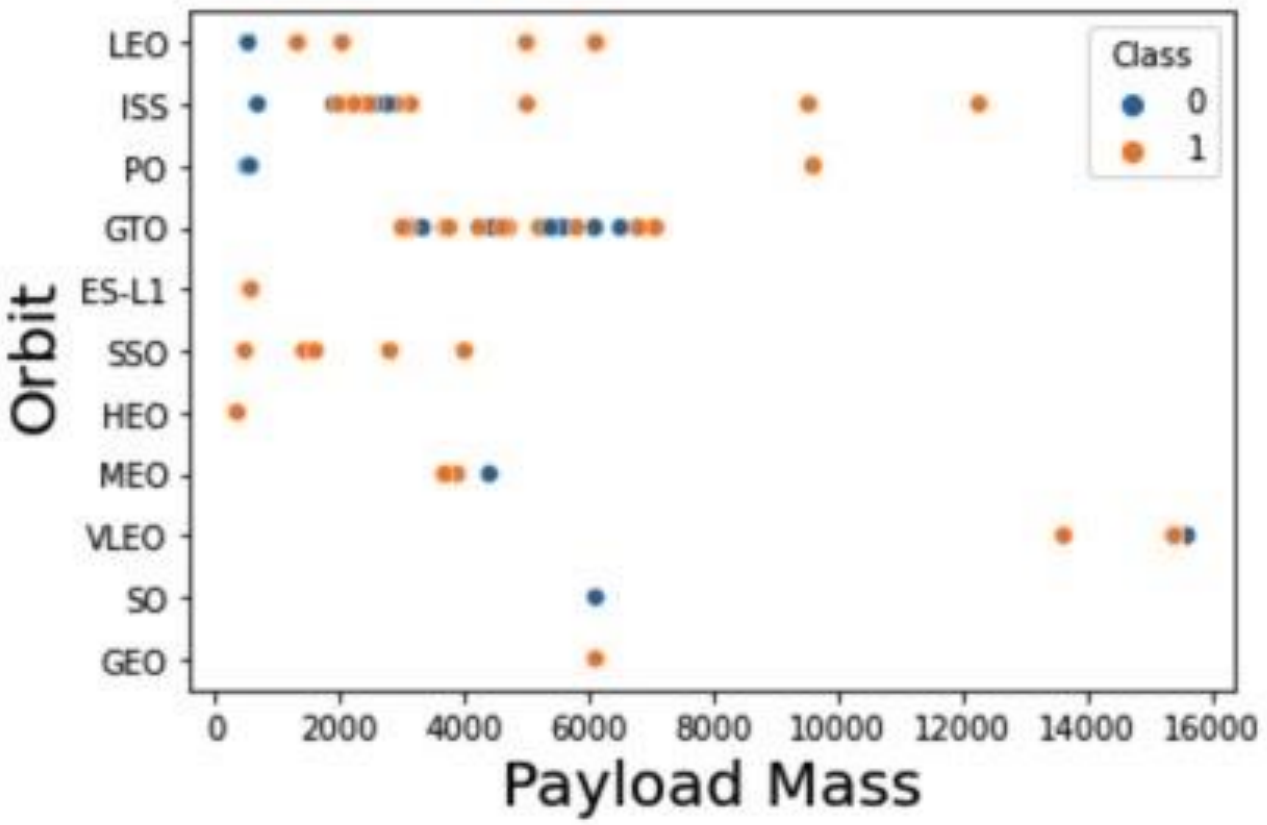
Flight Number vs. Orbit Type

- A scatter-plot showing the relation between the flight number and the orbit type.
- It should be see that 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.



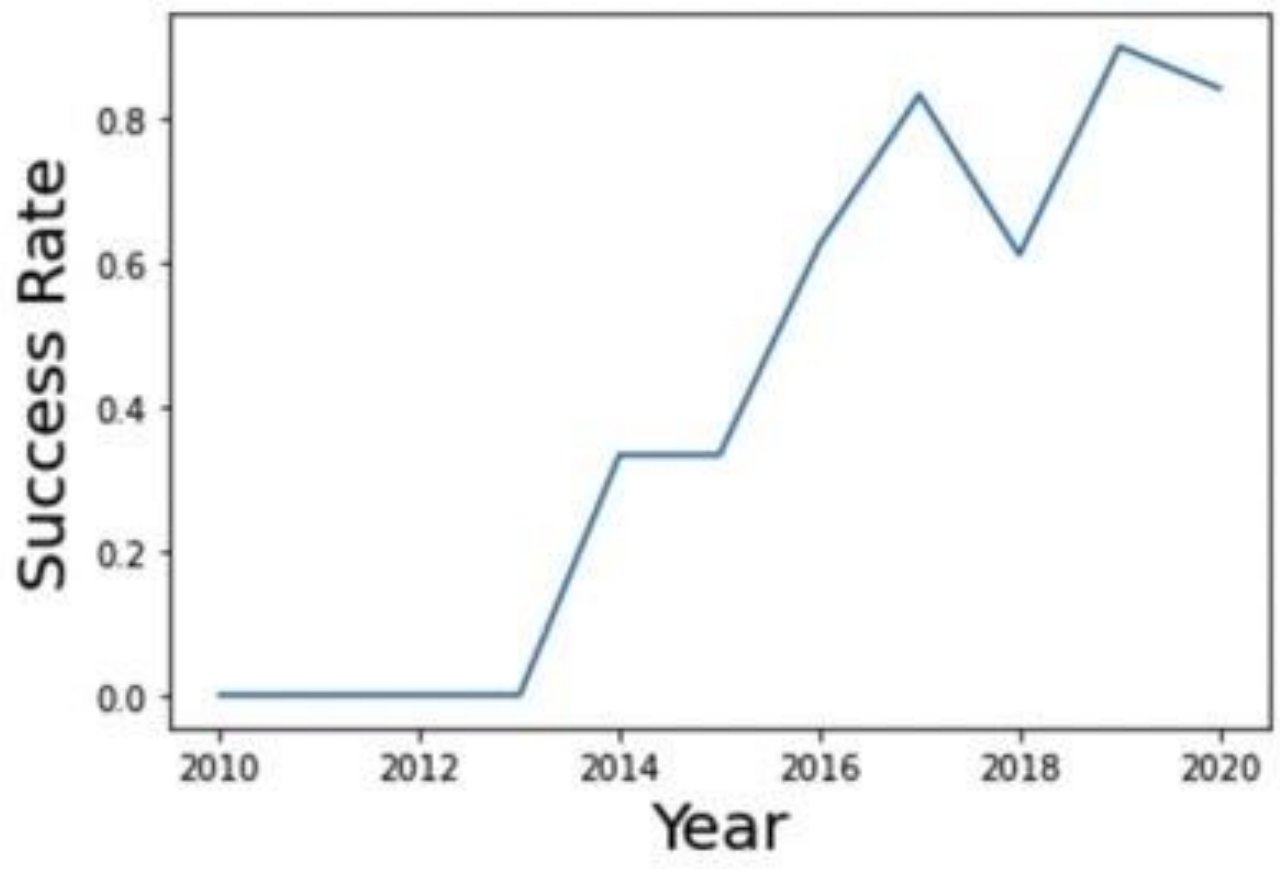
Payload vs. Orbit Type

- A scatter plot visualizing the relation between the payload mass and the orbit type.
- It should be observed that the Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch Success Yearly Trend

- A line plot between the success rate and the years.
- It can be observed that the sucess rate kept on increasing since 2013 till 2020.



All Launch Site Names

- The names of the four different launch sites are shown in the table using SQL.

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

```
In [4]: %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-98
Done.
```

Out[4]:

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

Launch Site Names Begin with 'CAA'

- Information on launch sites with a name that begins with 'CAA' are shown in the table using SQL.

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

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Total Payload Mass

- The total payload mass that was carried by NASA (CRS) is calculated using SQL as the result is shown with the query.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lc
31/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- The average payload mass that was carried by the booster version F9 v1.1 is 2929 kg and it waa calucalted using the shown SQL query.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00
Done.
```

1
2928

First Successful Ground Landing Date

- The first successful ground landing of the launcher's first stage occurred in 22-12-2015 and it was displayed using the shown SQL query.

```
%sql SELECT DATE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)' ORDER BY DATE LIMIT 1;
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdoma
Done.
```

DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- A list of the boosters' names that had a successful drone ship landing with a payload mass between 4000 – 6000 was found using SQL.

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 < 600
0;
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is 101 and it was found using SQL.

```
%sql SELECT COUNT(MISSION_OUTCOME) FROM SPACEXTBL;
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e
Done.
```

1
101

Boosters Carried Maximum Payload

- A list of the boosters which carried the maximum payload mass is displayed in the table which was found using SQL query.

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731
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

- A list of the failed landing outcomes on drone ships, their booster versions, and launch site names in the year 2015 was found using SQL.

```
%sql SELECT LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND
DATE LIKE '2015%';
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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

- The count of the landing outcomes whether successful or failure on different platforms between the date 2010-06-04 and 2017-03-20 is found using SQL as shown in the picture.

```
%sql SELECT LANDING__OUTCOME,COUNT(LANDING__OUTCOME) AS LANDING_OUTCOME_COUNT FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' A
ND '2017-03-20' GROUP BY LANDING__OUTCOME;
```

```
 * ibm_db_sa://lbn44437:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

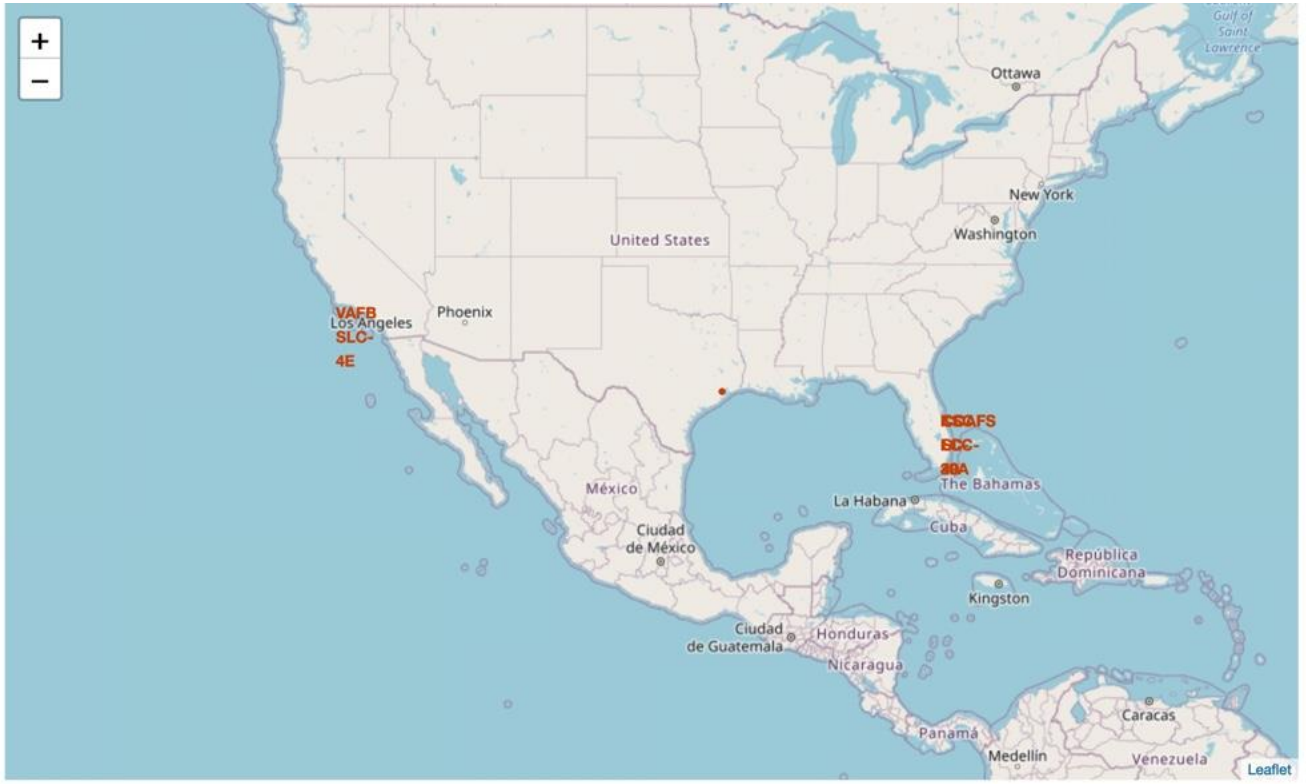
landing__outcome	landing_outcome_count
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 4

Launch Sites Proximities Analysis

Launch Sites Locations

- The launch sites locations are marked with a circle and their names in this folium map using Folium and the information available from the dataset.

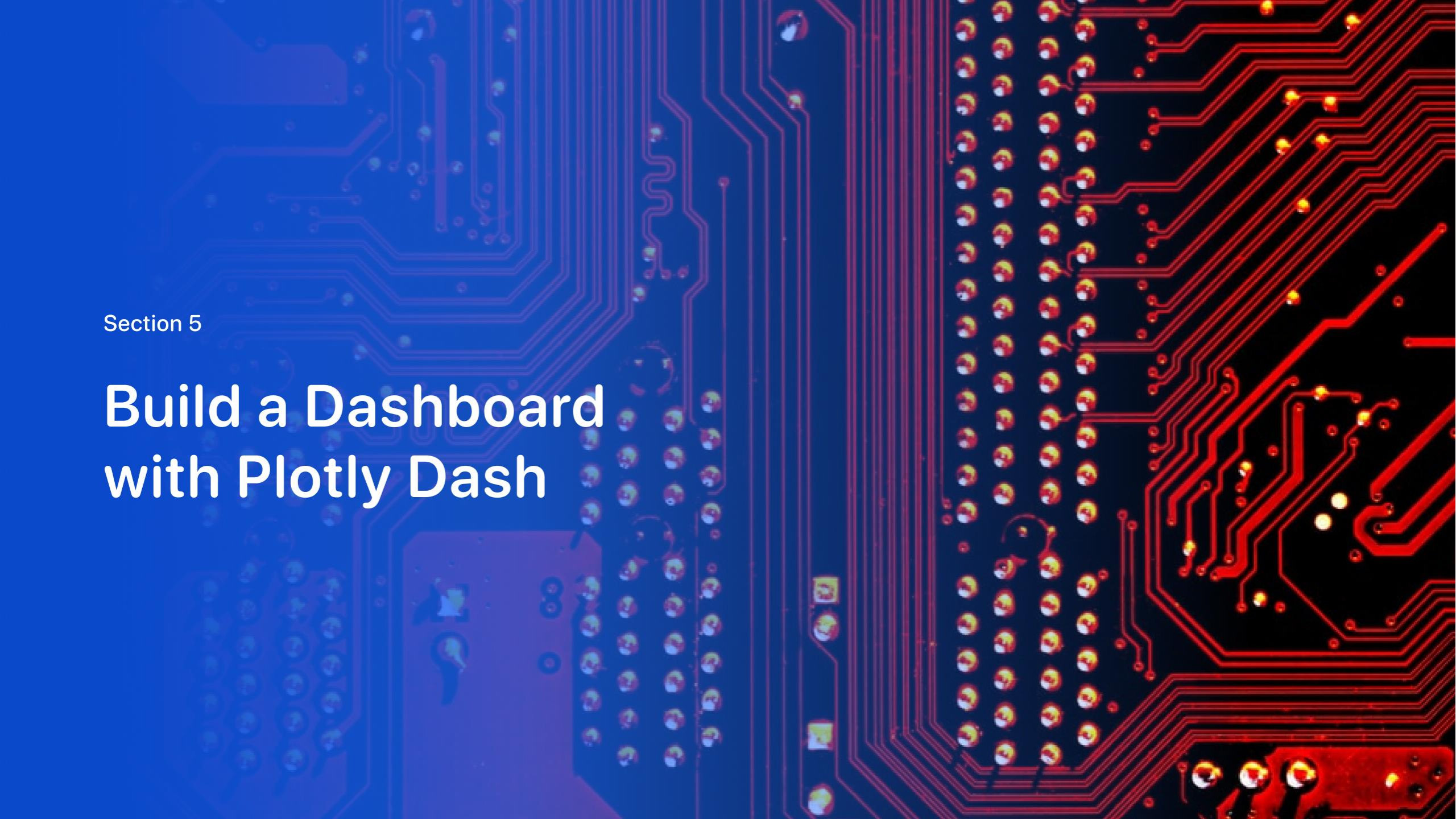


Success/Failed Launches of Each Site

- This folium map is marked with the number of launches, successful or failed, from each site and when it is clicked upon there is a marker cluster with two colors; green (indicating successful) and red (indicating failed).

Launch Sites and Proximities

- The folium map displays the launch site VAFB SLC-4E and its marked with the distance from the railway using a red line with folium features.

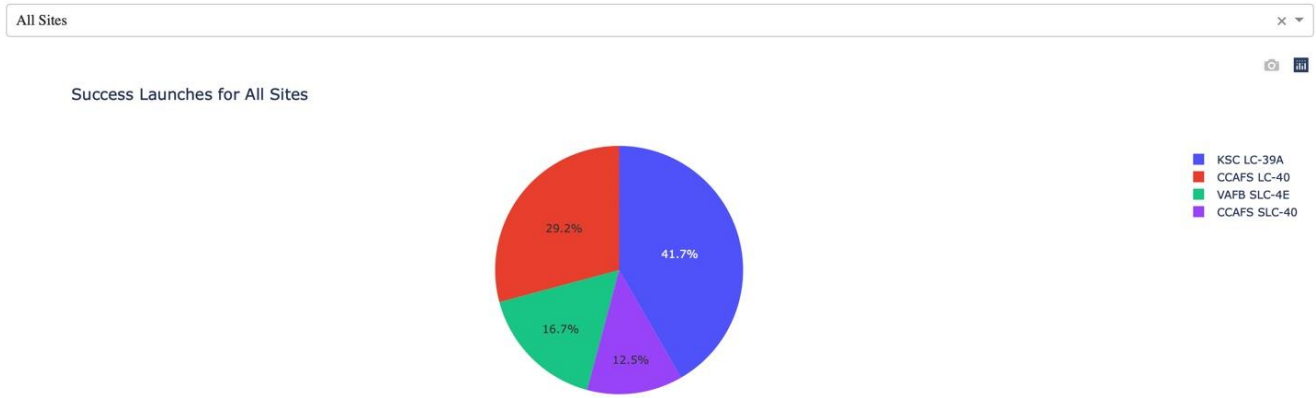


Section 5

Build a Dashboard with Plotly Dash

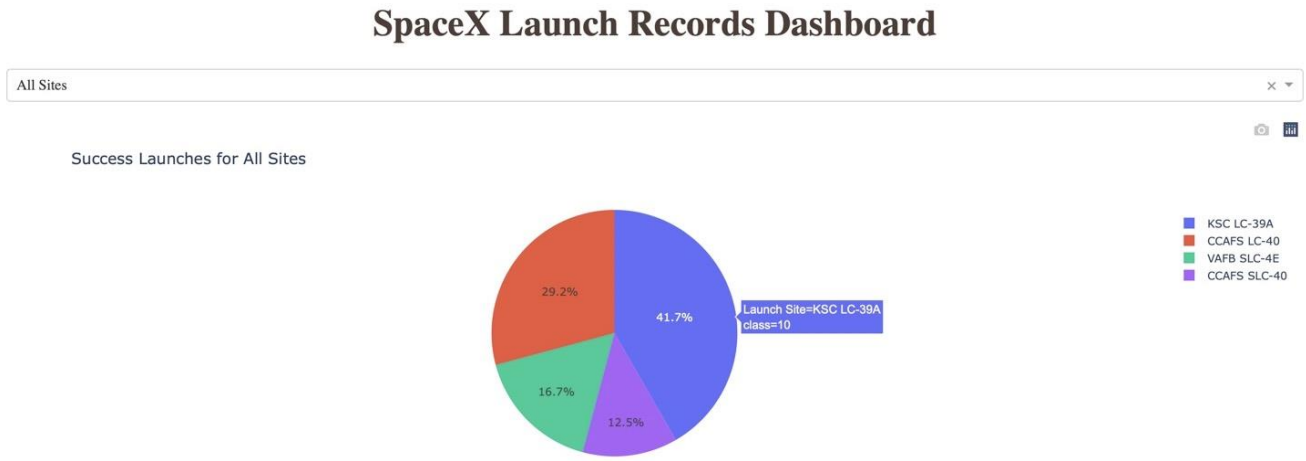
Launches Success Count

- A pie chart was created using Plotly Dash to display the success rate of each launch site in an interactive dashboard.



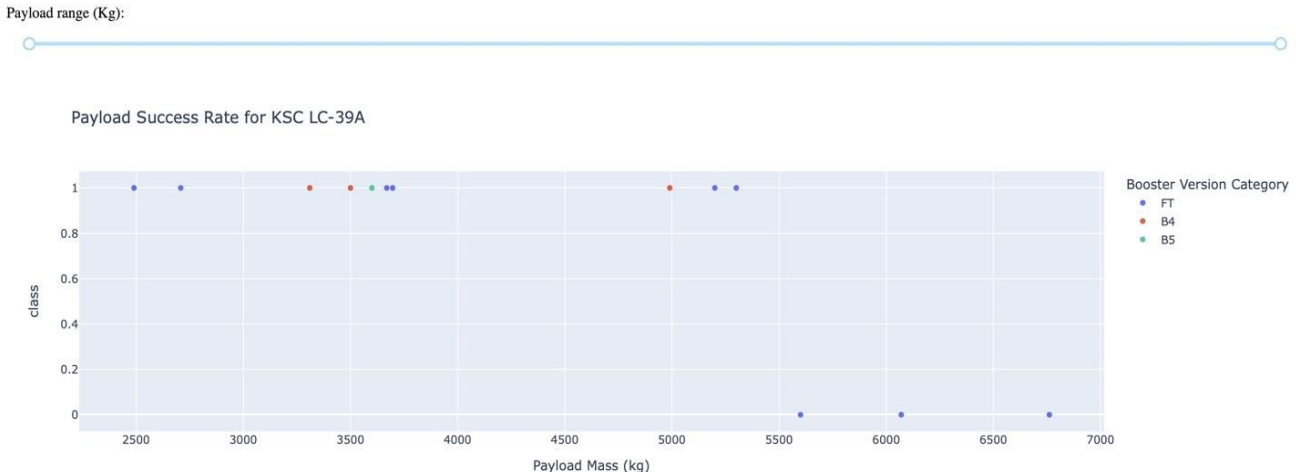
Interactive Launch Records

- The pie chart is inertactive that it displays the name of each site and its success rate.
- It can be observed that the KSC LC-39A launch site has the highest success rate compared to the other sites.



Payload vs. Launch Outcome

- The scatter plot describes the relation between the range of payload mass with the range slider and displays the type of booster and its success or fail status. Each launch site can be selected using the site dropdown as it will view the booster versions as well.



Section 6

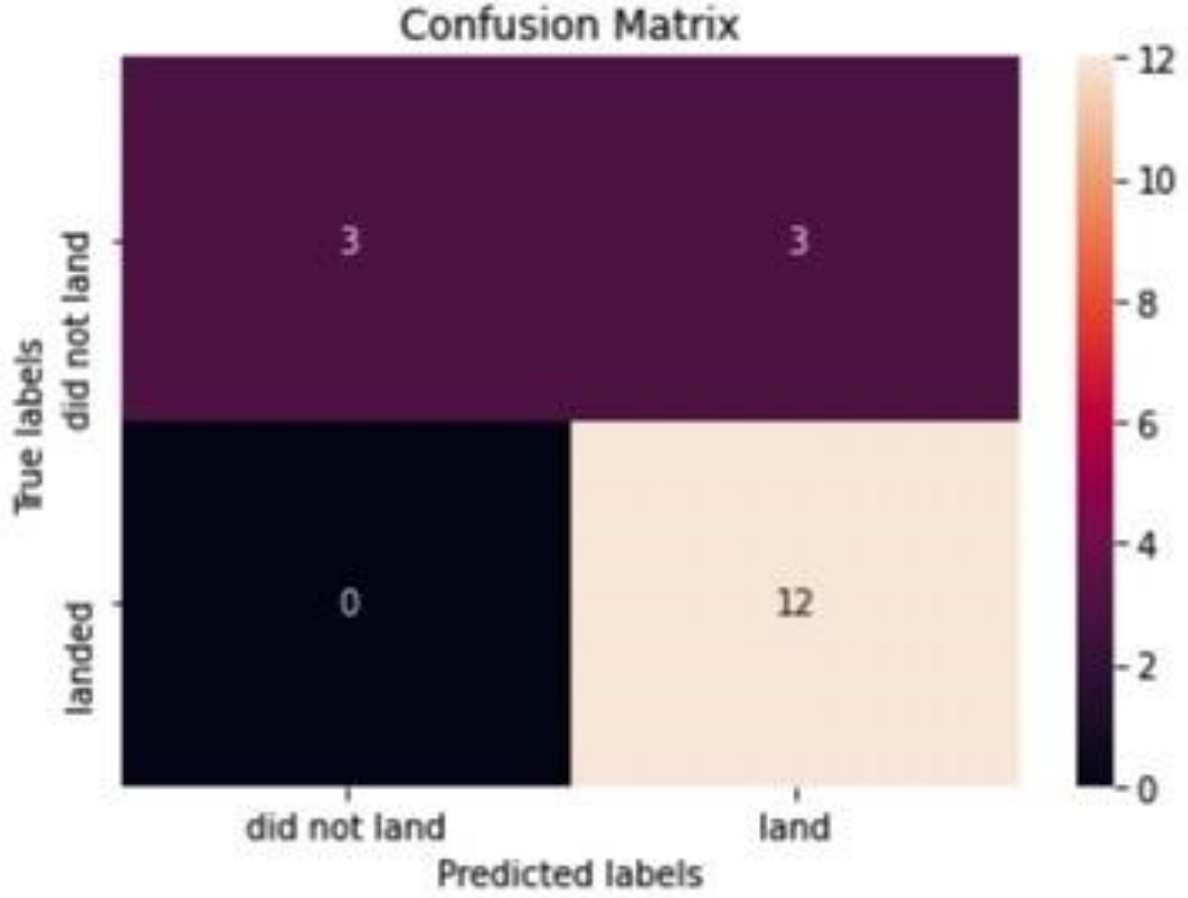
Predictive Analysis (Classification)

Classification Accuracy

- Four classification methods were tested to find out which model will perform the best.
- The four models were Logistic Regression, Support Vector Machine(SVM), Decision Tree and K-Nearest Neighbor(KNN).
- All of the models came back with almost the same accuracy which was %83.33.

Confusion Matrix

- The confusion matrix shown is one of the models since they've all yielded the same accuracy, it is posted for clarification.

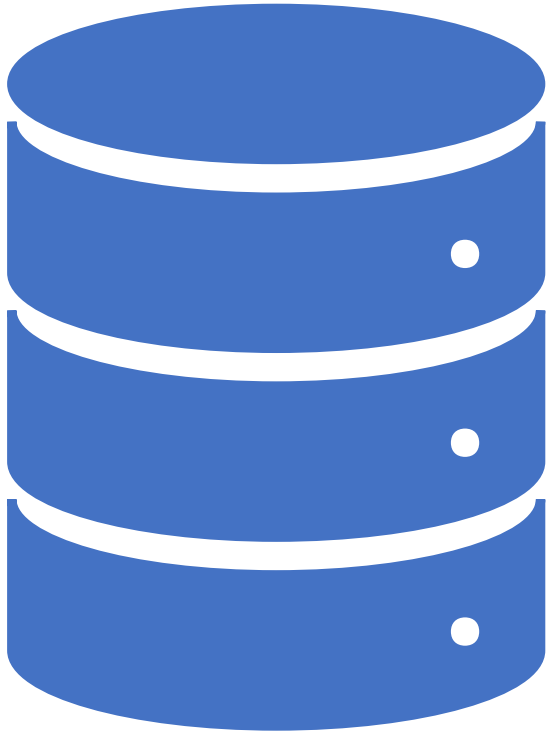


Conclusions

- There are many factors affecting the successful return of the 1st stage of the launcher.
- Relationships between payload mass, launch site location, and orbit type were very influencing.
- Visualizing the datasets and using the folium map allowed us to have a better understanding of the relationships that exists between the various factors.
- Different classification methods were used to predict the success of the 1st stage return that resulted in accuracy of %83.33.

Appendix

- Data Collection - SpaceX API
- Data Collection - Scraping
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Folium Map
- Plotly Dashboard
- Predictive Analysis (Classification)



Thank you!