IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Elizabeth García
March 2022

Outline

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

Executive Summary

- Summary of methodologies
  - ✓ Data collection through the SpaceX API and using Web Scrapping.
  - ✓ Exploratory Data Analysis with SQL, Data visualization and Folium.
  - ✓ Machine Learning Prediction.
- Summary of all results
  - ✓ Exploratory Data Analysis result.
  - ✓ Interactive analytics in screenshots.
  - ✓ Predictive Analytics result.

Introduction

- Project background and context

  Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

- Problems you want to find answers

  ✓ What factors determine if the rocket will land successfully?

  ✓ The interaction amongst various features that determine the success rate of a successful landing.

  ✓ What operating conditions needs to be in place to ensure a successful landing program.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from the SpaceX API
- Perform data wrangling
  - It was created a landing outcome label based on outcome data after summarizing and analyzing features.
- 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 was divided into training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

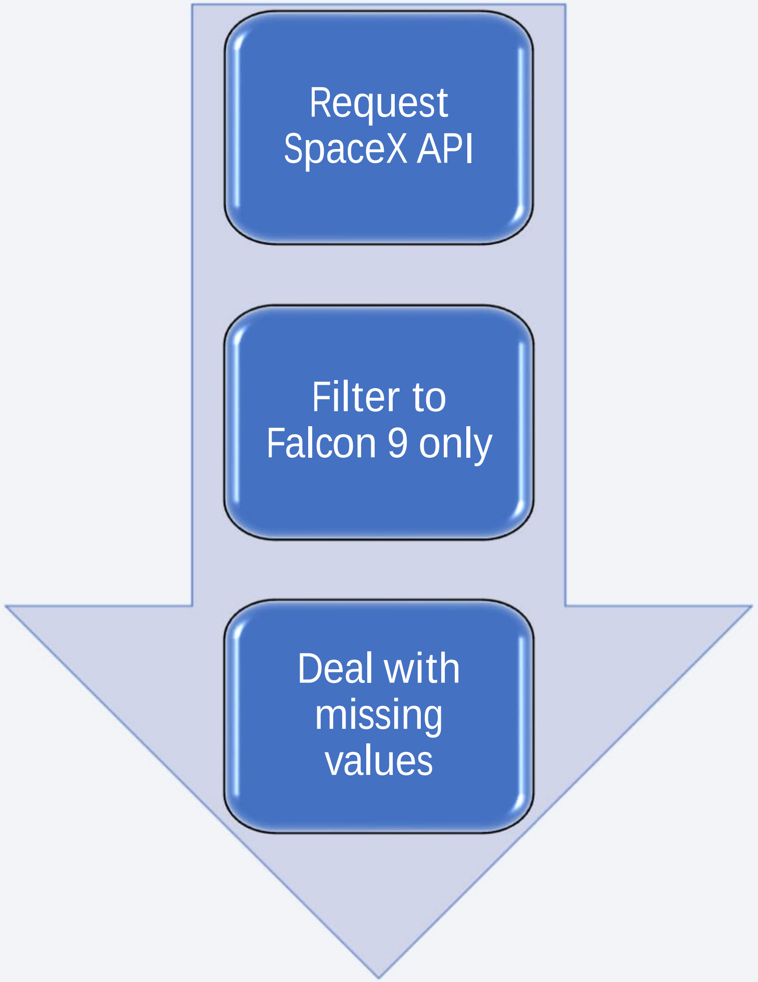
Data Collection

- Describe how data sets were collected.
  - Data collection was done using get request to the SpaceX API.
  - Next, we decoded the response content as a Json using .json() function call and turn it into a pandas dataframe using .json_normalize().
  - We then cleaned the data, checked for missing values and fill in missing values where necessary.
  - In addition, we performed web scraping from Wikipedia for Falcon 9 launch records with BeautifulSoup.
  - The objective was to extract the launch records as HTML table, parse the table and convert it to a pandas dataframe for future analysis.

Data Collection – SpaceX API

- The public API offered by Space X was used to collect the data.

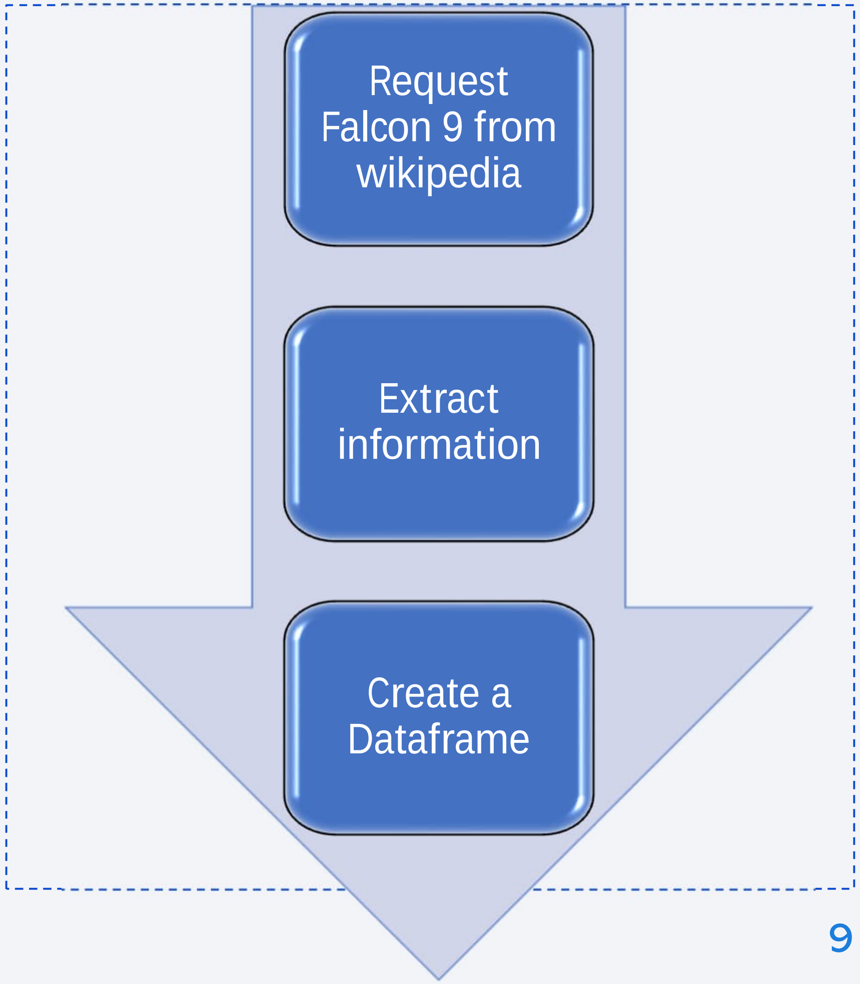
- The link to the notebook is: https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/1.%20Data%20Collection.ipynb



Data Collection - Scraping

- Data from Wikipedia was also obtained and parsed into a data frame

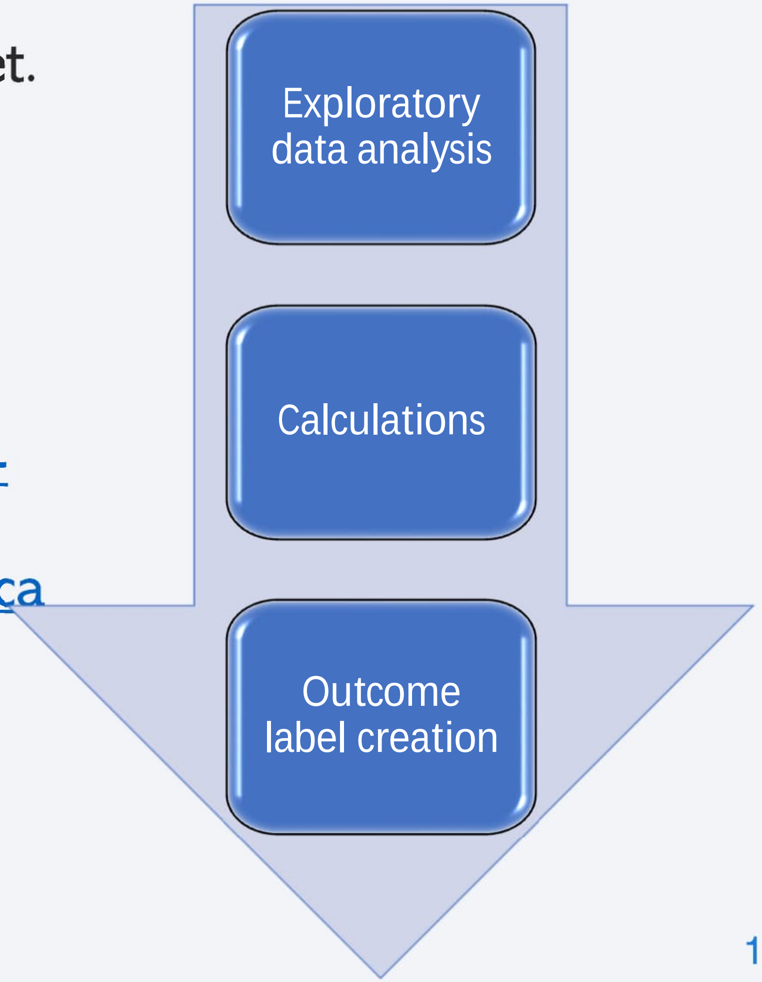
https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/2.%20Data%20collection%20with%20web%20scrapping.ipynb



Data Wrangling

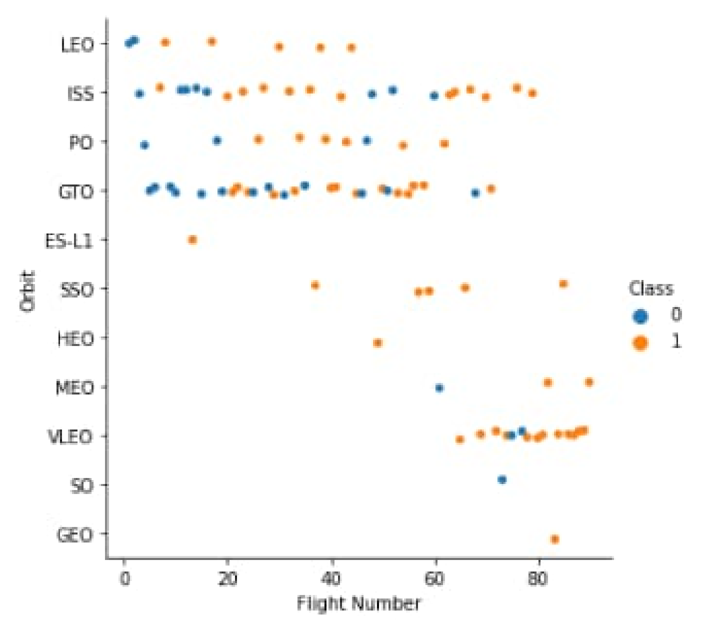
- Exploratory data analysis was performed in the data set.
- Number of launches from each site and the occurrence of the different orbits was calculated
- The outcome label was created, and a csv file was created with the results.

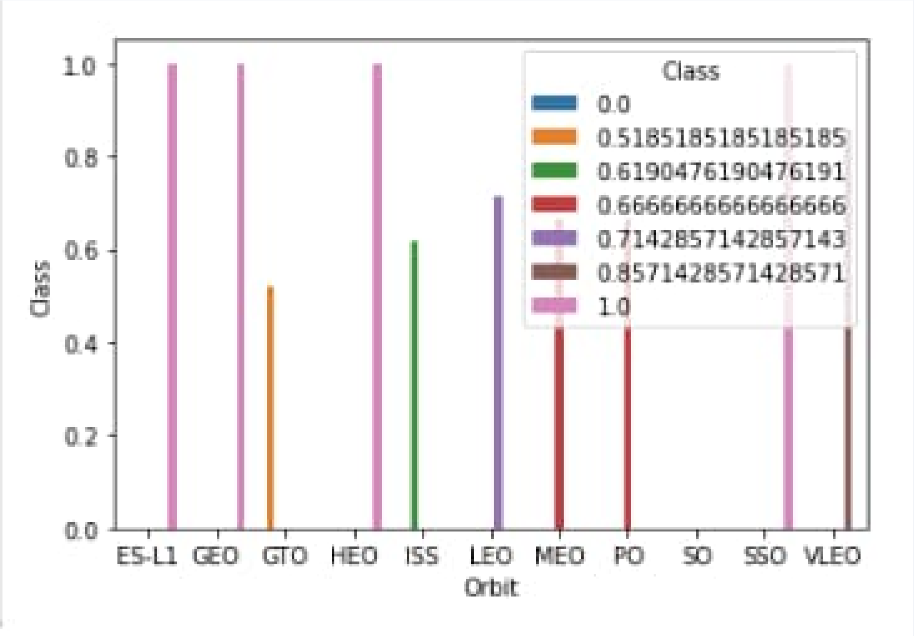
https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/3.%20Data%20Wrangling.ipynb



EDA with Data Visualization

- Scatter and bar plots were used to try to find the relationships between the different features.





https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/4.%20EDA.ipynb

EDA with SQL

- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission;
  - Top 5 launch sites whose name begin with the string 'CCA';
  - Total payload mass carried by boosters launched by NASA (CRS);
  - Average payload mass carried by booster version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  - Total number of successful and failure mission outcomes;
  - Names of the booster versions which have carried the maximum payload mass;
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
  - Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/5.%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- A folium map was created where the following objects were included:
- Circle object on the launch site with a popup label to show the launch site name
- Color labels to identify the success
- Lines to calculate the distance between sites and proximities

https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/6.%20Launch%20sites%20locations%20analysis%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

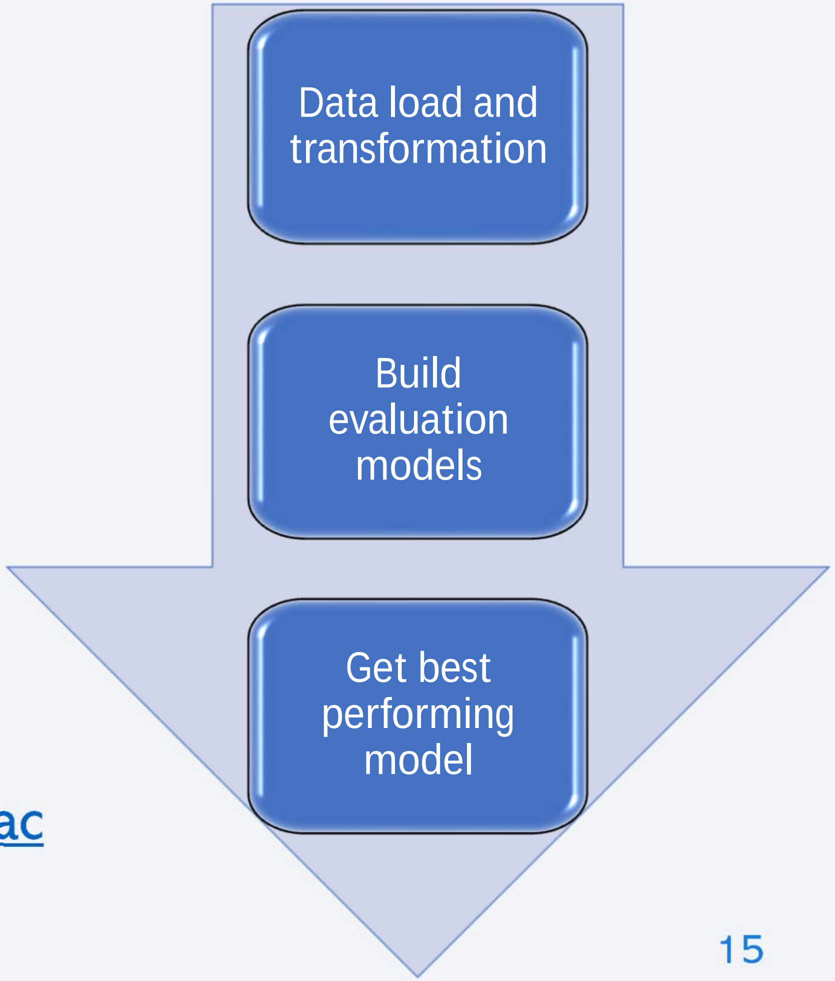
- An interactive dashboard with Plotly was created.

https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/f5d184b1a51e923d69b44d2bb800d4eb6d4310c2/Dash

Predictive Analysis (Classification)

- The data was loaded and transformed.
- It was split into training and testing data sets.
- Different classification models were used.
- Accuracy was used as metric.
- Results were compared to get the best performing model.

https://github.com/ElizabethGGgithub/IBM-Data-Science-Capstone-SpaceX/blob/eb0d80344e10d9cb80b3d69a535ebdc5cacd507a/8.%20Machine%20Learning%20Prediction.ipynb



Results

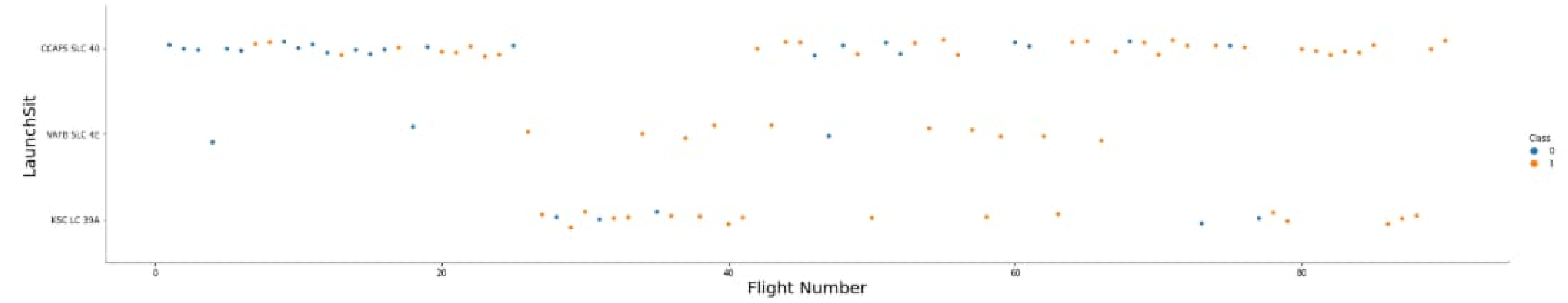
- Exploratory data analysis results:

- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

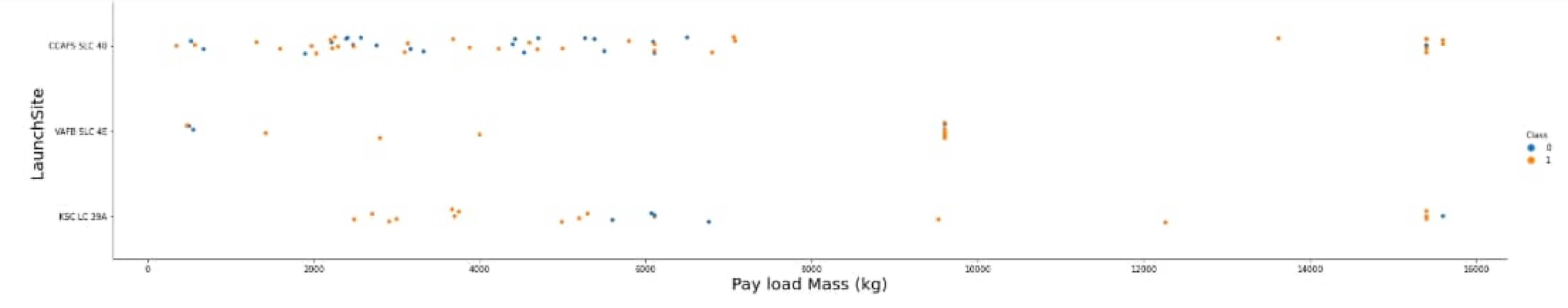
Insights drawn from EDA

Flight Number vs. Launch Site



- From the plot it can be seen that the launch site that has more successful launches is CCAFS SLC 40.
- Launch site CCAFS SLC 40 is also the most used launch site.

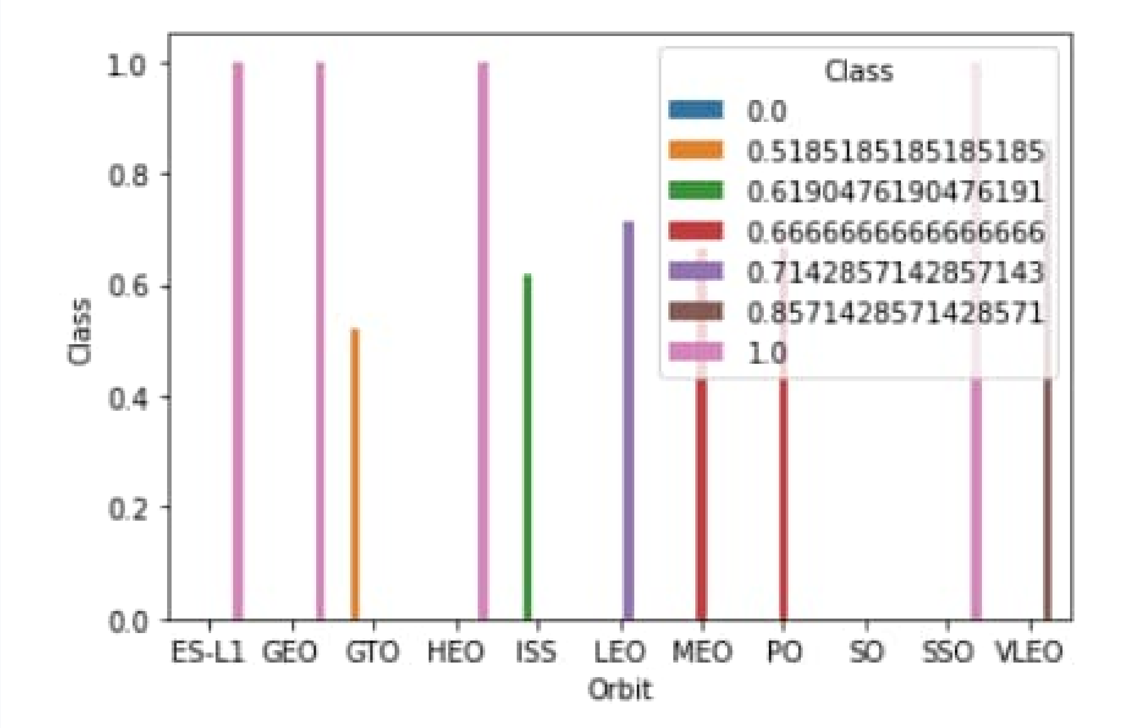
Payload vs. Launch Site



- Payloads around 9500kg and around 15500kg are the most successful ones
- Payloads above 9500 kg are never launched from VAFB-SLC 4E launch site.

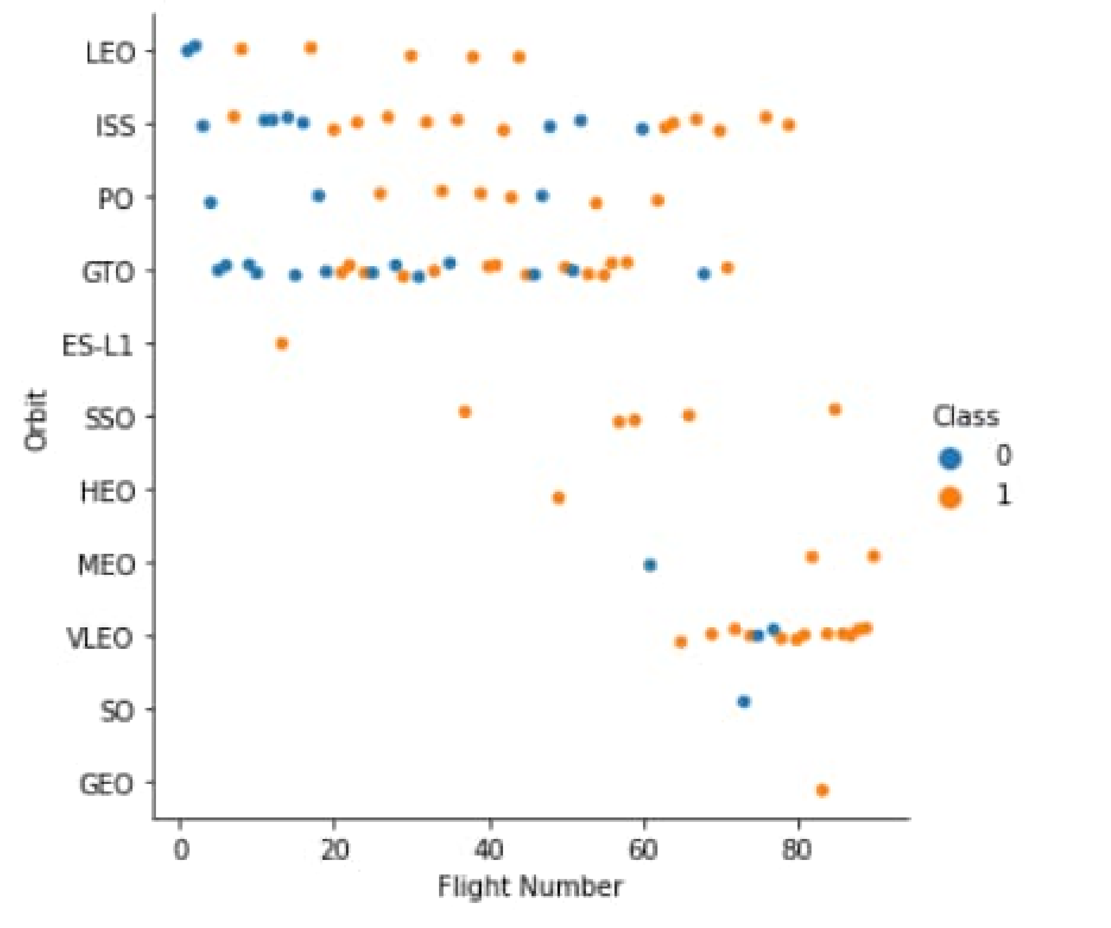
Success Rate vs. Orbit Type

- The following orbits have the highest sucess rate:
- ES-L1
- GEO
- HEO
- SSO



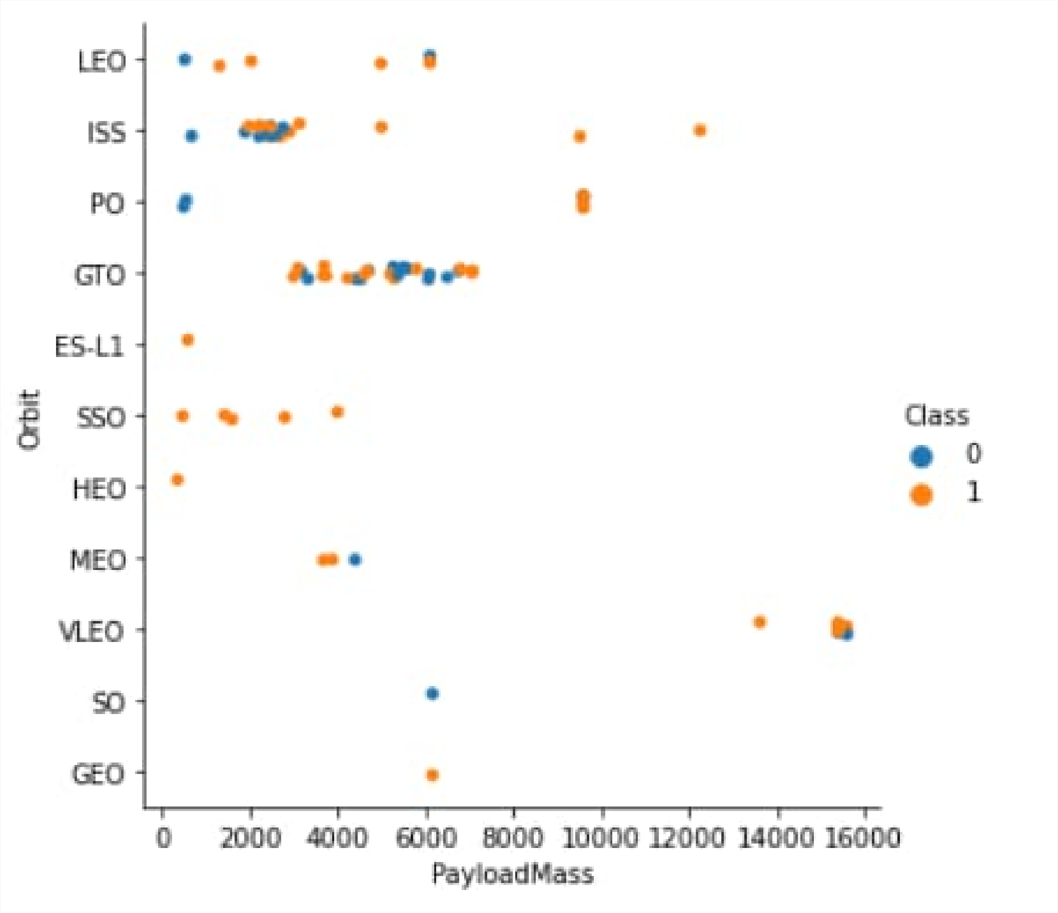
Flight Number vs. Orbit Type

- All filghts in SSO orbit have been successful.
- VLEO and LEO orbits are also very successful in relation to the number of flights



Payload vs. Orbit Type

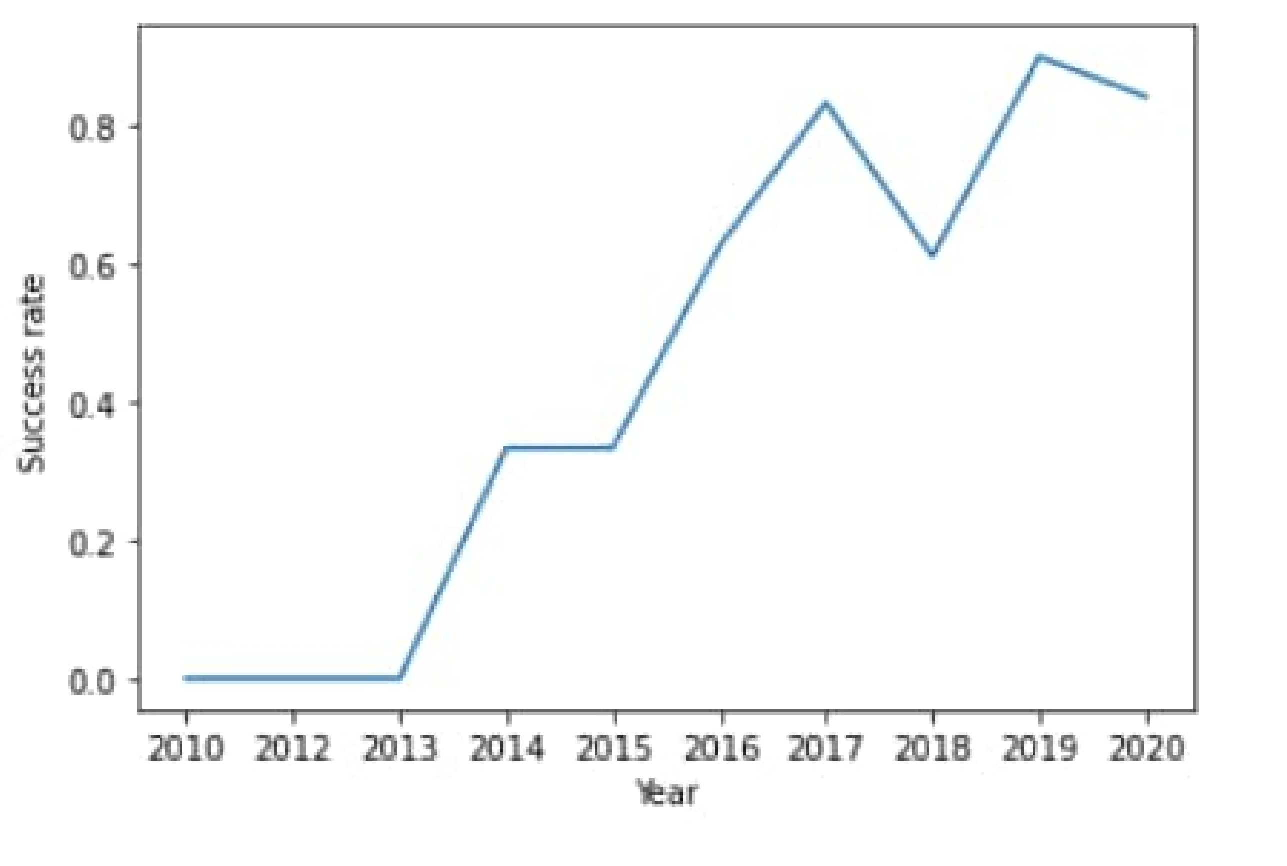
- At lower payloads it seems that the most successful flights have been those at SSO orbit.
- At higher payloads it seems that the most successful orbits are ISS, PO and VLEO.
- PO orbit has 50% of success but it can be clearly seen how the success depends on the payload. At higher payloads better success rate.



Launch Success Yearly Trend

- In 2014, the first flight was successful.

- Since then, the success rate has increased .



All Launch Site Names

- There are four launch sites:

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried out by boosters is 619967 kg

payloadmass
619967

Average Payload Mass by F9 v1.1

- Average payload mass carried out the booster version F9

average_payload_mass
2928.4

First Successful Ground Landing Date

- The first successful landing in ground pad occurred on:

min(DATE)	Landing_Outcome
01-05-2017	Success (ground pad)

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

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

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

boosterversion	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015

Year	month	Mission_Outcome	Booster_Version	Launch_Site
2015	01	Success	F9 v1.1 B1012	CCAFS LC-40
2015	02	Success	F9 v1.1 B1013	CCAFS LC-40
2015	03	Success	F9 v1.1 B1014	CCAFS LC-40
2015	04	Success	F9 v1.1 B1015	CCAFS LC-40
2015	04	Success	F9 v1.1 B1016	CCAFS LC-40
2015	06	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
2015	12	Success	F9 FT B1019	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

Landing_Outcome	COUNT(Landing_Outcome)
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

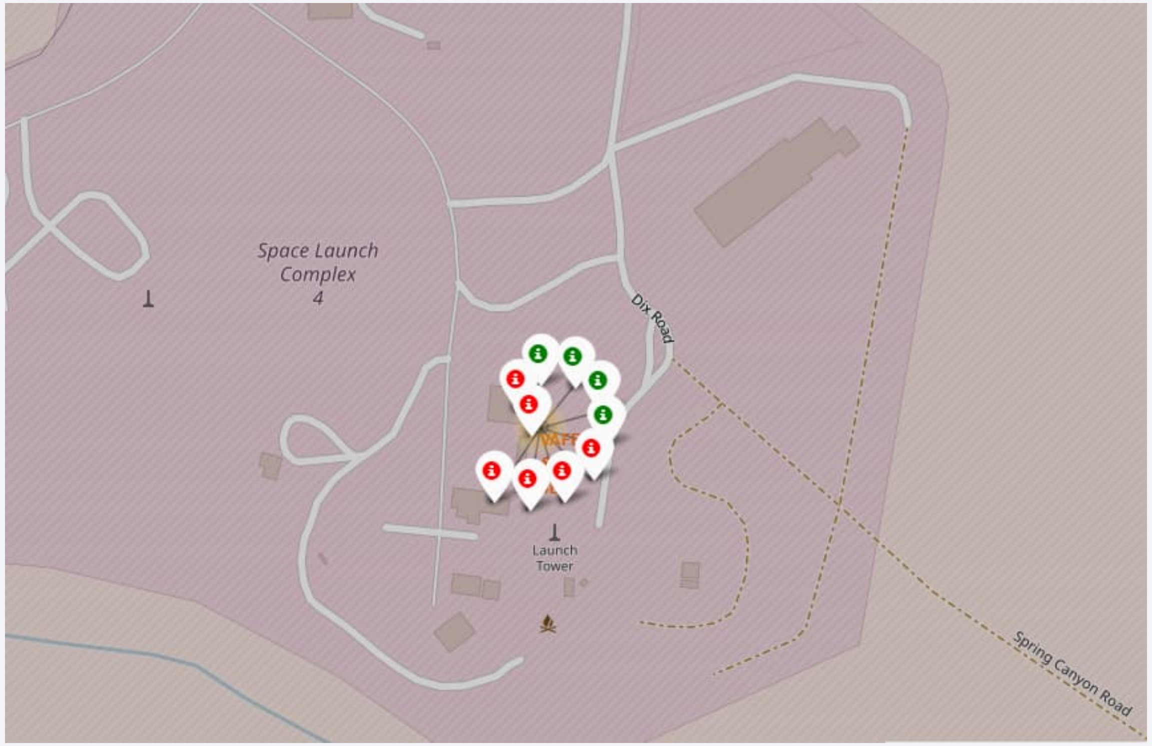
Section 3

Launch Sites Proximities Analysis

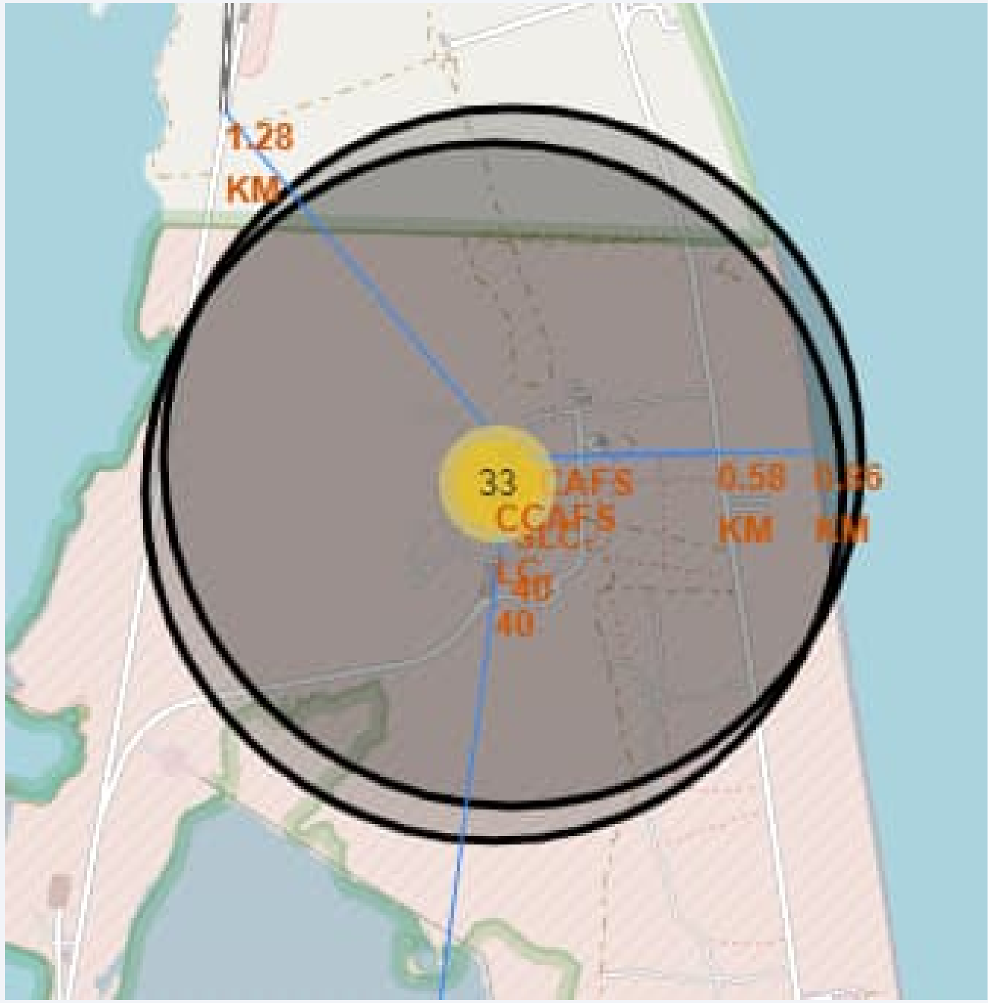
Map of SpaceX launch sites

Launch sites are located in the NASA facilities.

Launch sites with color labels



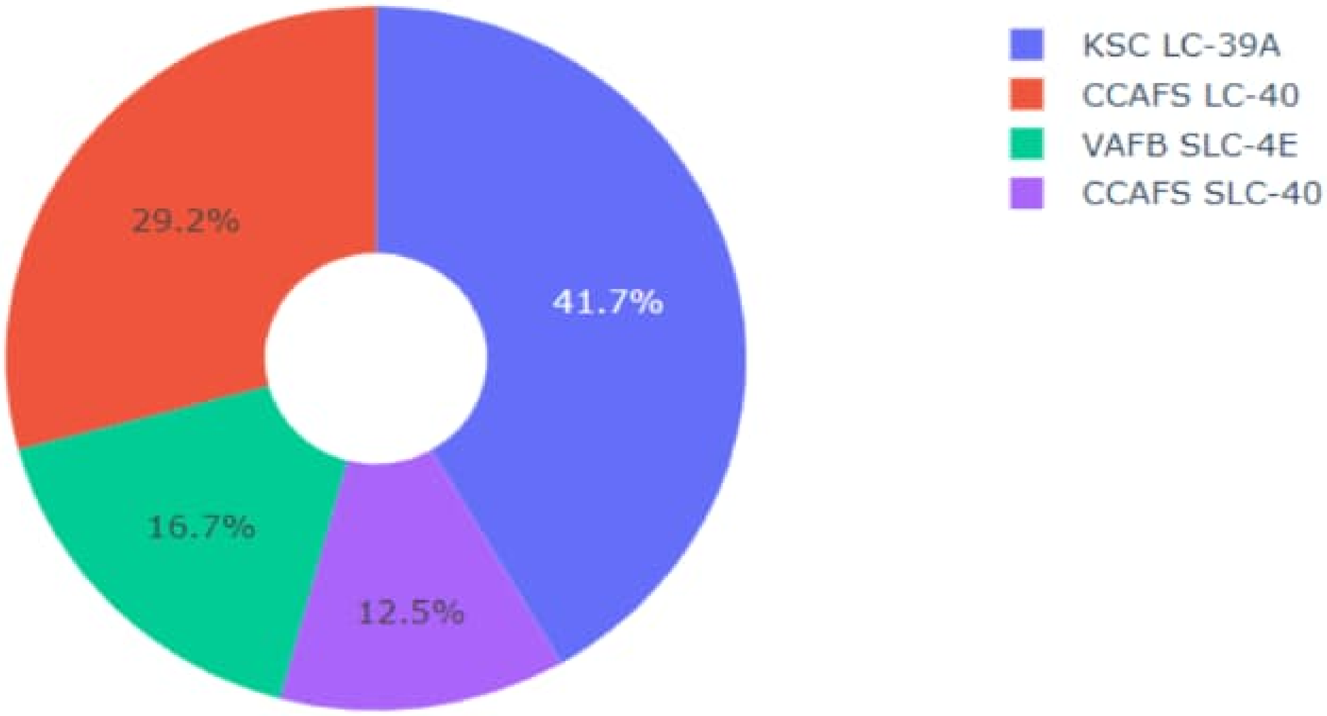
Distance from launch site to landmarks



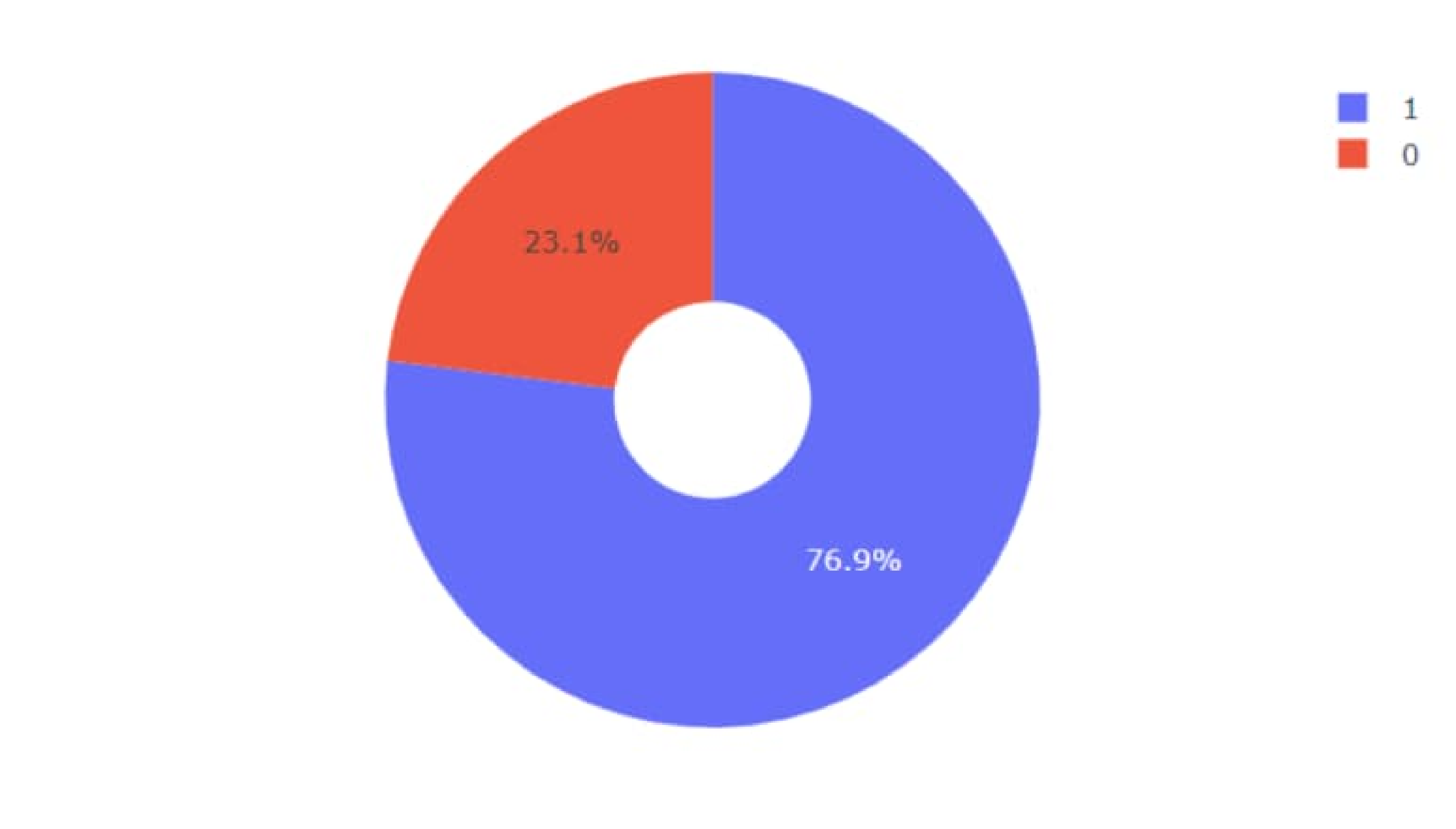
Section 4

Build a Dashboard with Plotly Dash

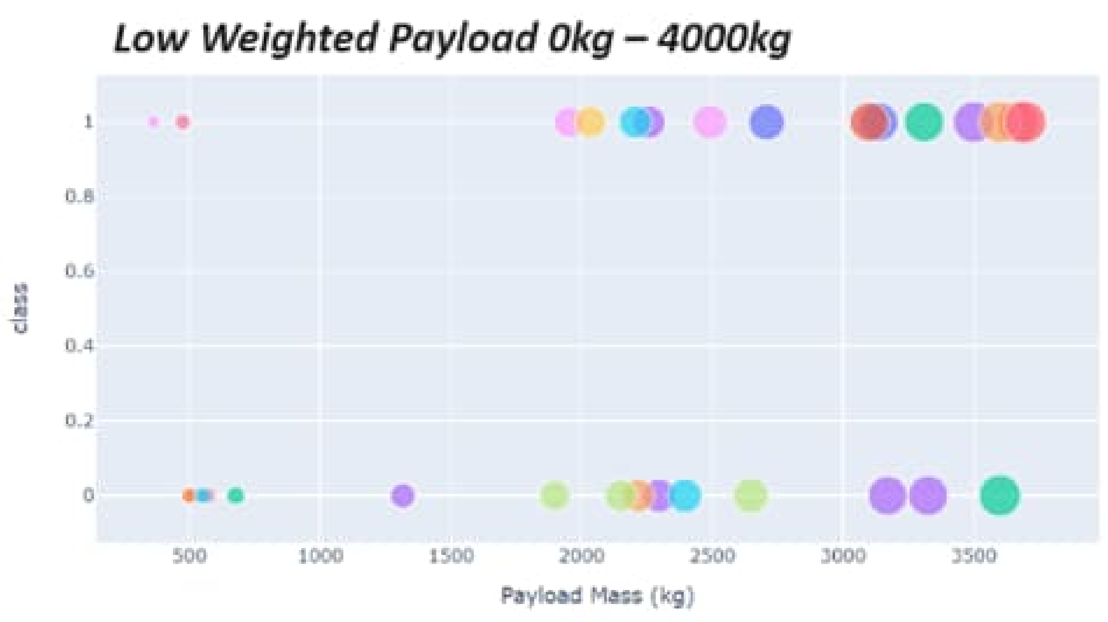
Success per launch site

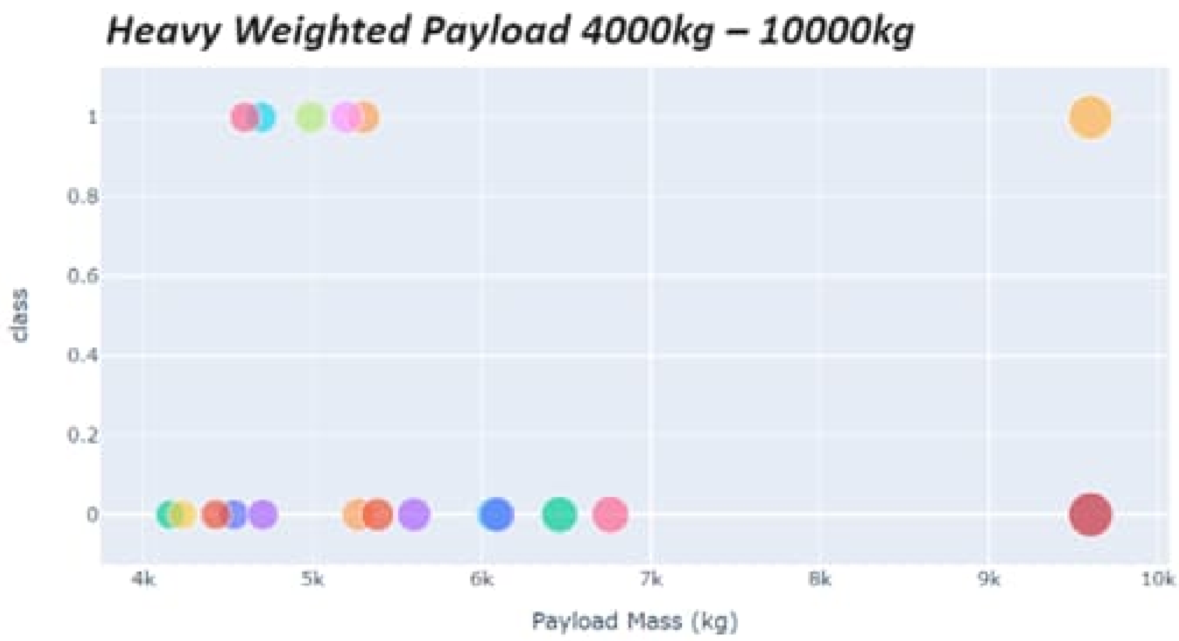


Launch success ratio at KSC LC-39A



Payload vs Launch Outcome





Section 5

Predictive Analysis (Classification)

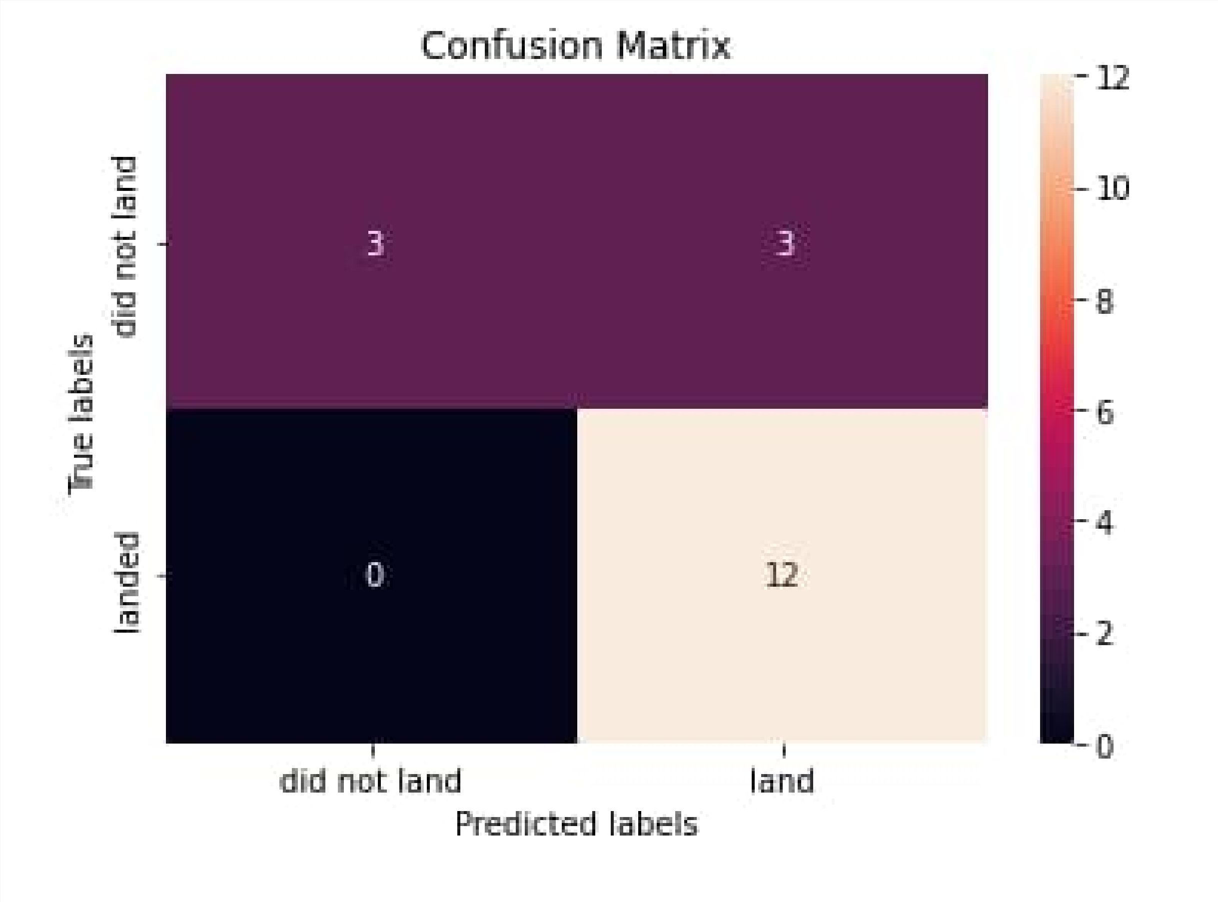
Classification Accuracy

Find the method performs best:

```
print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print( 'Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))
```

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8333333333333334
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Confusion Matrix



Conclusions

- Payloads around 9500kg and around 15500kg are the most successful ones
- Orbits with higher success rate are ES-L1, GEO, HEO and SSO. Being SSO the best orbit for lower payloads and ISS, PO and VLEO for higher payloads.
- The first successful flight occurred in 2014.
- CCAFS SLC 40 is the most successful launch site

Thank you!