IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamed LAMOUCHI

February 2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
  - The project aims to explore and analyze the SpaceX launch data to gain insights into the company's launch history, mission outcomes, and booster performance. The data set used for this project is a comprehensive collection of SpaceX launch records from 2010 to 2020, obtained from the SpaceX API and web scraping.
- Problems you want to find answers
  - What are the trends and patterns in SpaceX's launch history?
  - What are the factors that influence the success or failure of a SpaceX launch?
  - How has SpaceX's booster performance improved over time?
  - What are the potential areas for future improvement in SpaceX's launch operations?

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
- How to build, tune, evaluate classification models

Data Collection

The data for this project was collected from two sources:

- SpaceX API: The SpaceX API provides access to a wide range of data about SpaceX launches, including launch dates, launch sites, mission outcomes, and booster performance.

  (https://api.spacexdata.com/v4/rockets/)

- Web scraping: Web scraping was used to collect additional data about SpaceX launches, such as payload mass, orbit type, and landing outcomes.

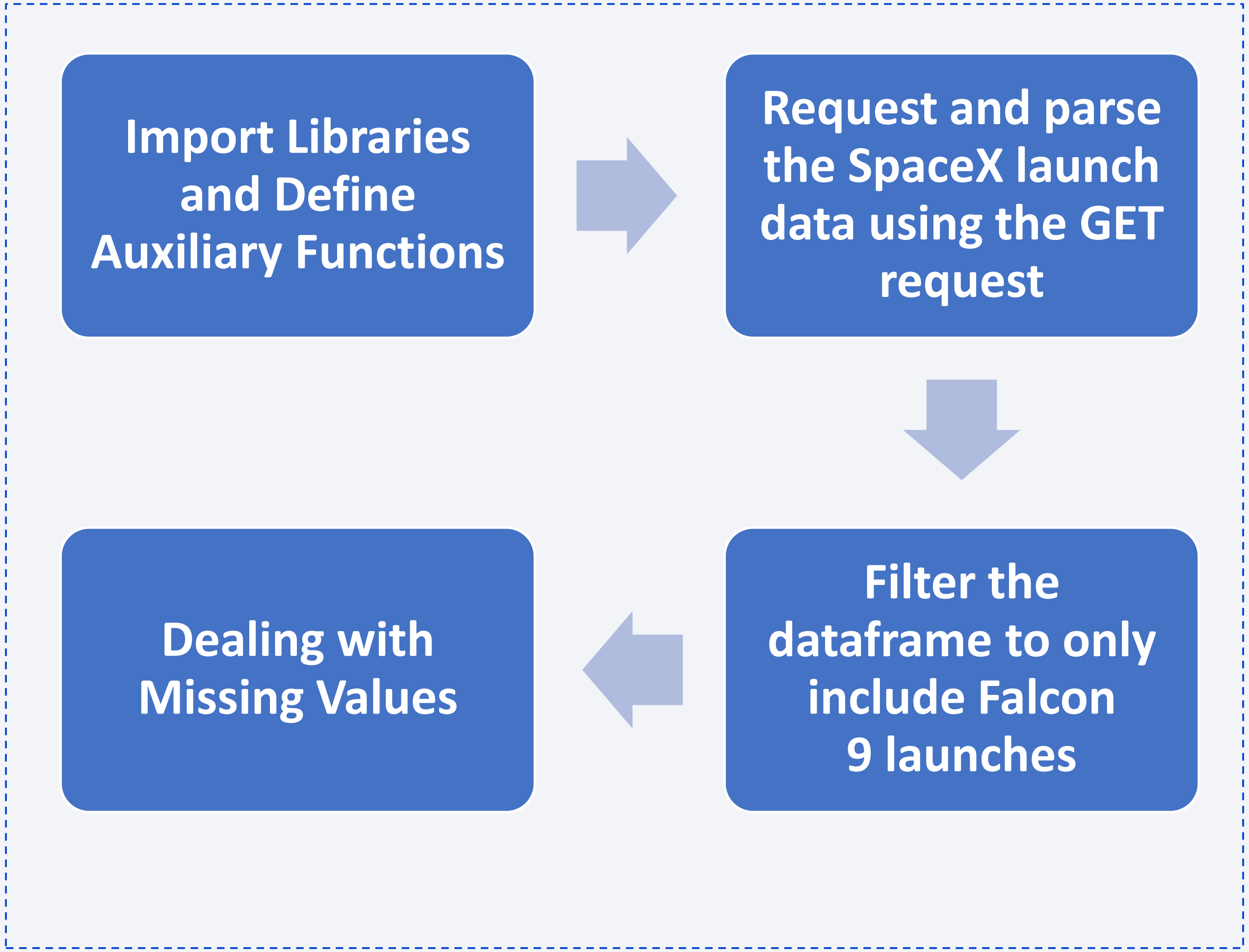
  (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained

- GitHub URL:

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

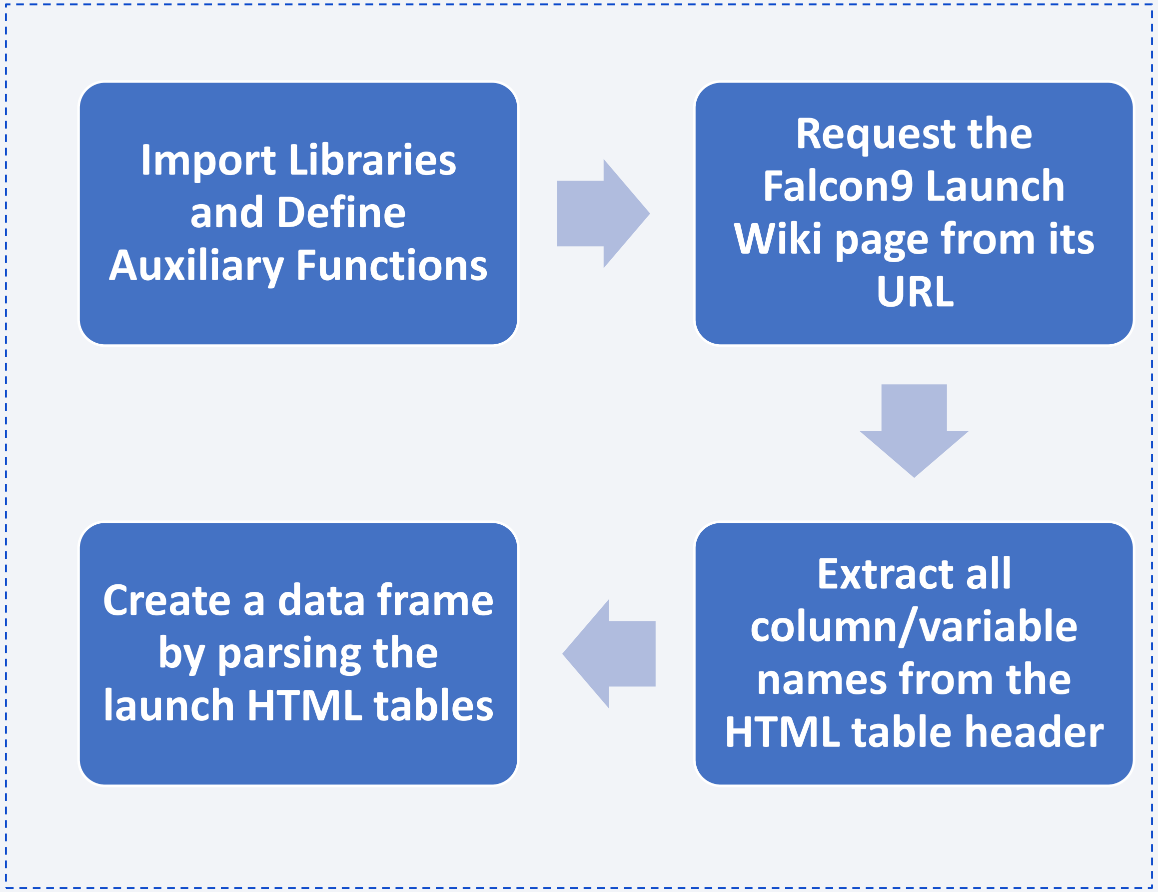


Data Collection - Scraping

- **Web scrap Falcon 9 launch records with BeautifulSoup:**
- **Extract a Falcon 9 launch records HTML table from Wikipedia**
- **Parse the table and convert it into a Pandas data frame**

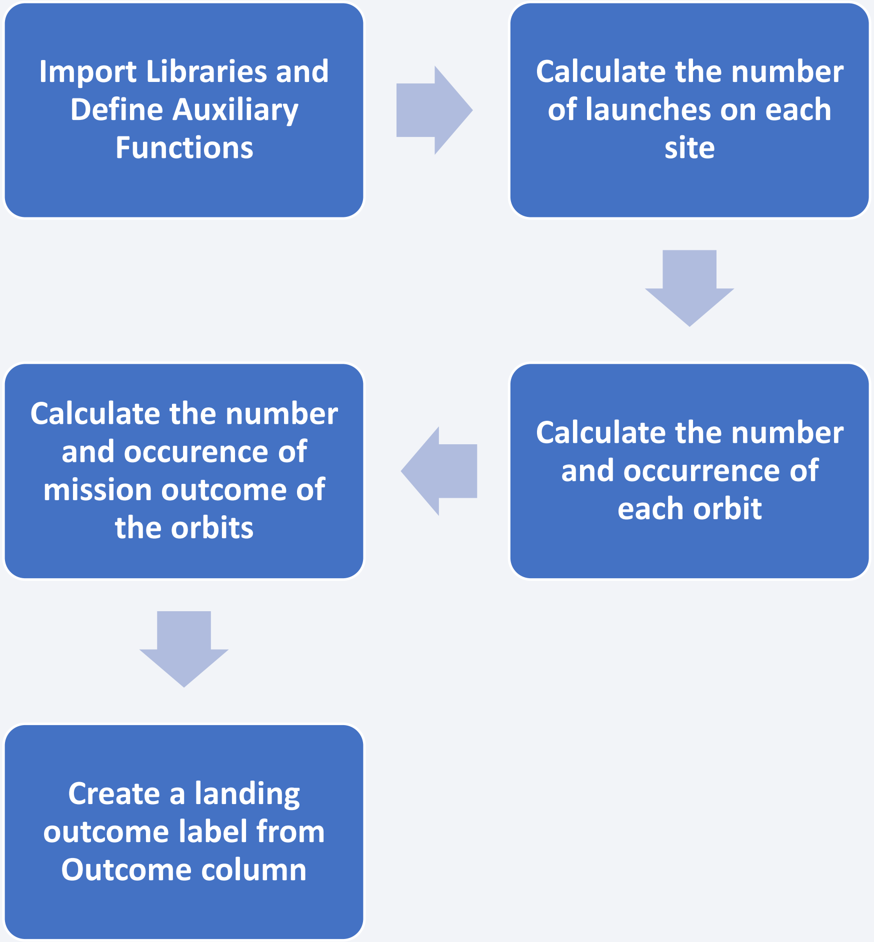
- GitHub URL:

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.



- GitHub URL:

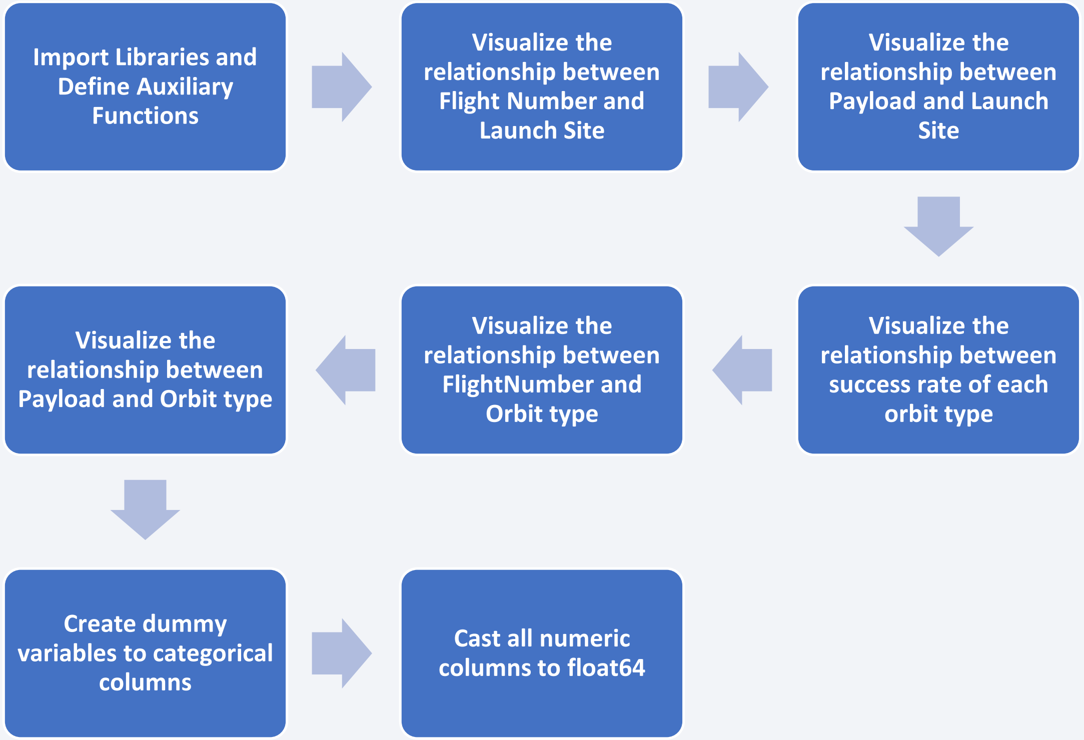
https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- EDA was used to explore the data and identify patterns and trends. This involved creating visualizations such as scatter plots, bar charts, and line charts.

- GitHub URL:

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb



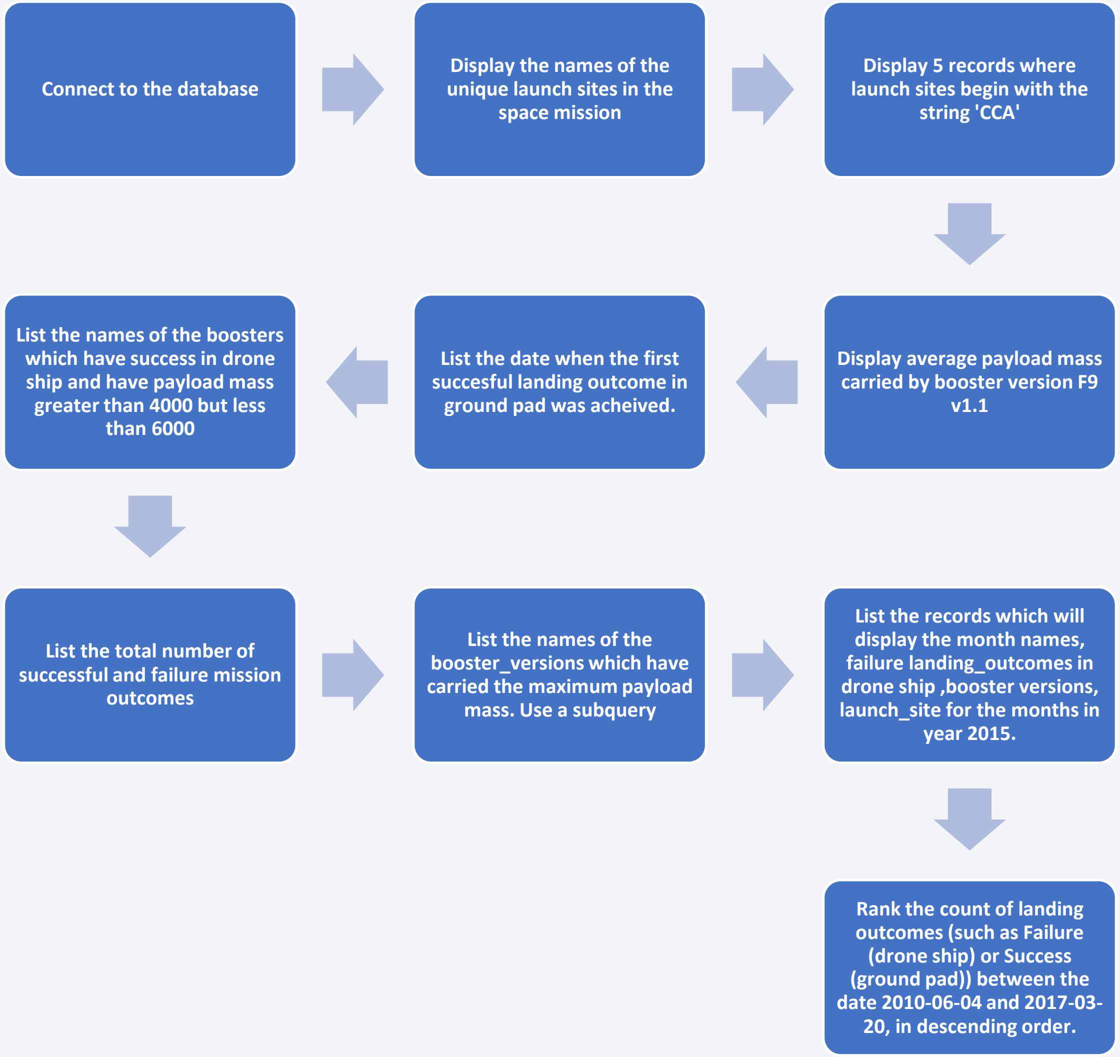
EDA with SQL

Using this Python notebook you will:

- Understand the Spacex DataSet
- Load the dataset into the corresponding table in a Db2 database
- Execute SQL queries

GitHub URL:

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb



Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps

Markers indicate points like launch sites

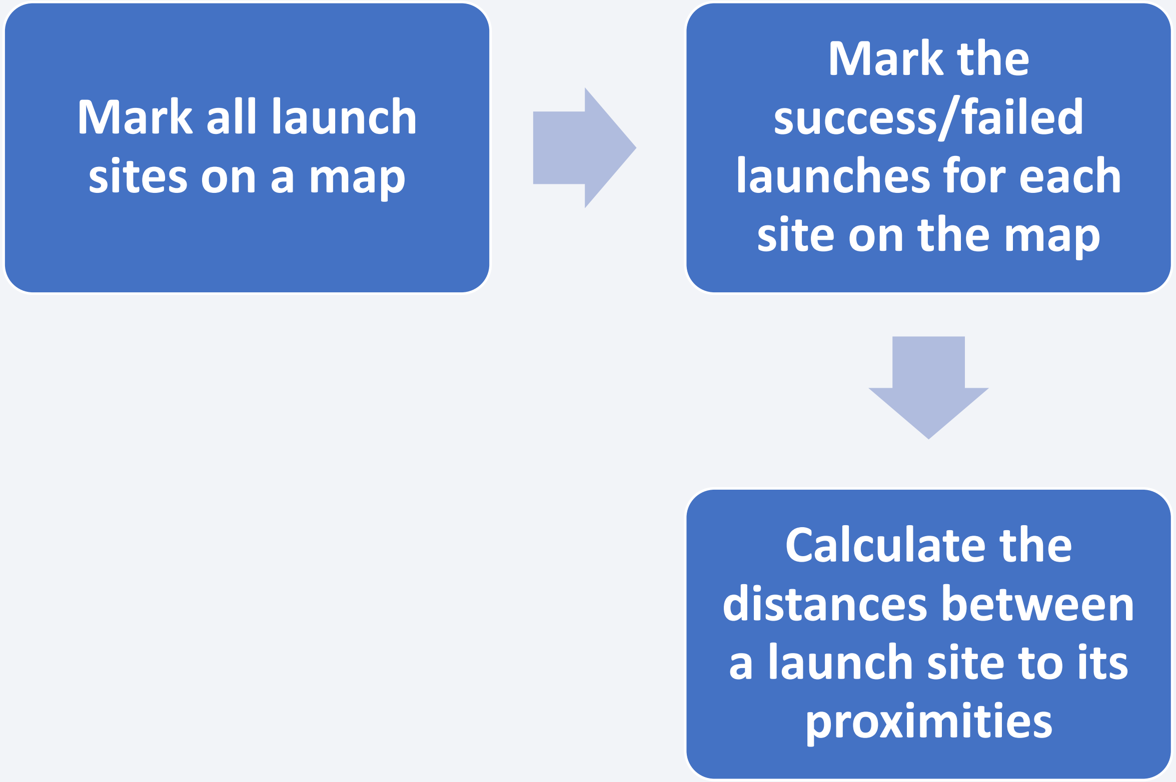
Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;

Marker clusters indicates groups of events in each coordinate, like launches in a launch site

Lines are used to indicate distances between two coordinates.

GitHub URL:

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb



Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data

  Percentage of launches by site

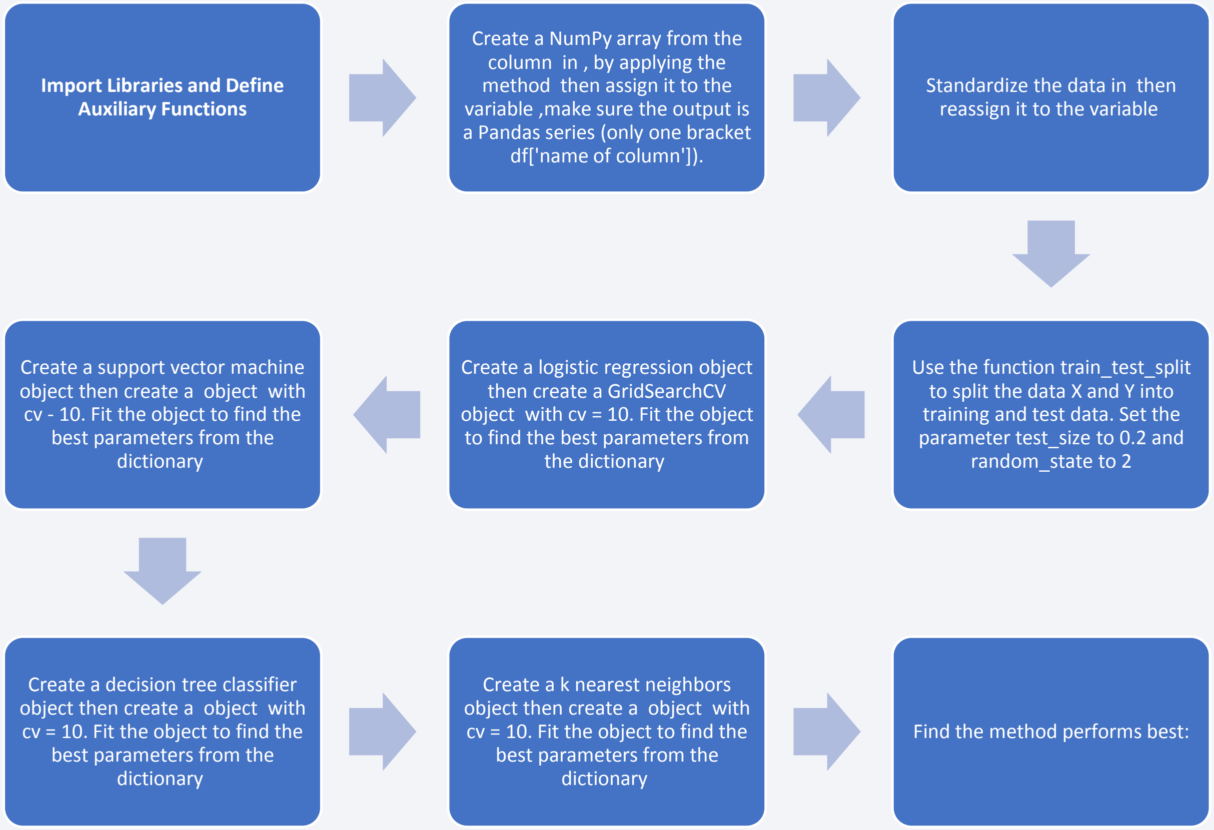
  Payload range

- GitHub URL

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



GitHub URL

https://github.com/mohamedlamouchi/IBM-Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- **Exploratory data analysis results**
- The EDA revealed several interesting patterns and trends in the data. For example, the EDA showed that:

  - The success rate of SpaceX launches has increased over time.
  - FlightNumber and PayloadMass are important factors in predicting the outcome of a SpaceX launch..
  - The scatter plot also shows that there is a cluster of unsuccessful landings at low FlightNumbers and high PayloadMasses. This suggests that SpaceX had some early challenges with landing heavy payloads..
  - However, as FlightNumber increases, the cluster of unsuccessful landings at high PayloadMasses becomes smaller. This suggests that SpaceX has made progress in overcoming these challenges.

Interactive analytics demo in screenshots

- **Predictive analysis results**

Results

- Interactive analytics demo in screenshots
  - Using interactive analytics was possible to identify that launch use to be in safe place near the sea and have good infrastructure around
  - Most launches sites happens at east cost launch sites

Results

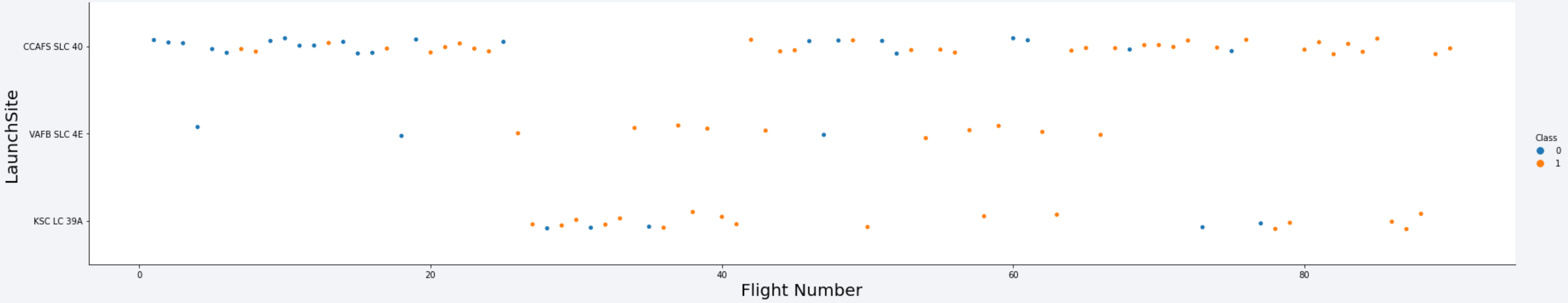
- Predictive analysis results

Predictive Analysis showed that Decision Tree is the best model to predict successful landings, having is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

Section 2

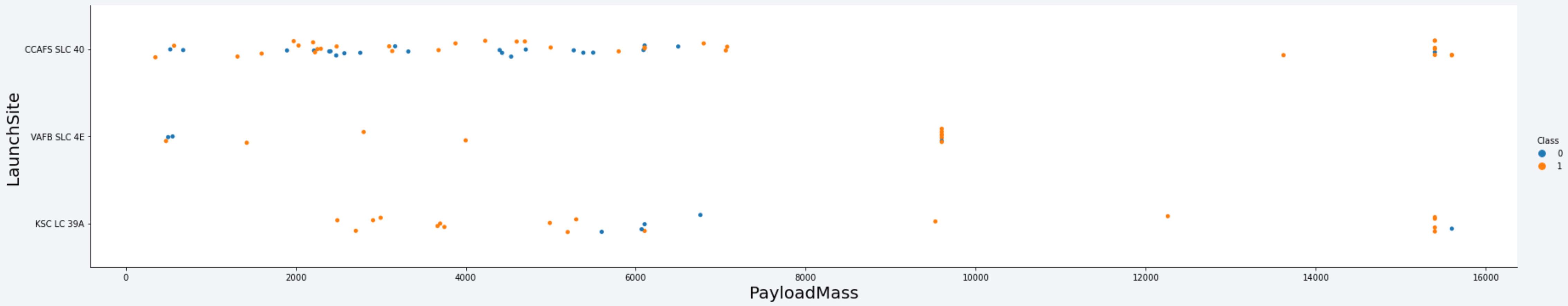
Insights drawn from EDA

Flight Number vs. Launch Site



- It’s also possible to see that the general success rate improved over time.
- CCAF5 SLC 40, is the most of successful launch site

Payload vs. Launch Site

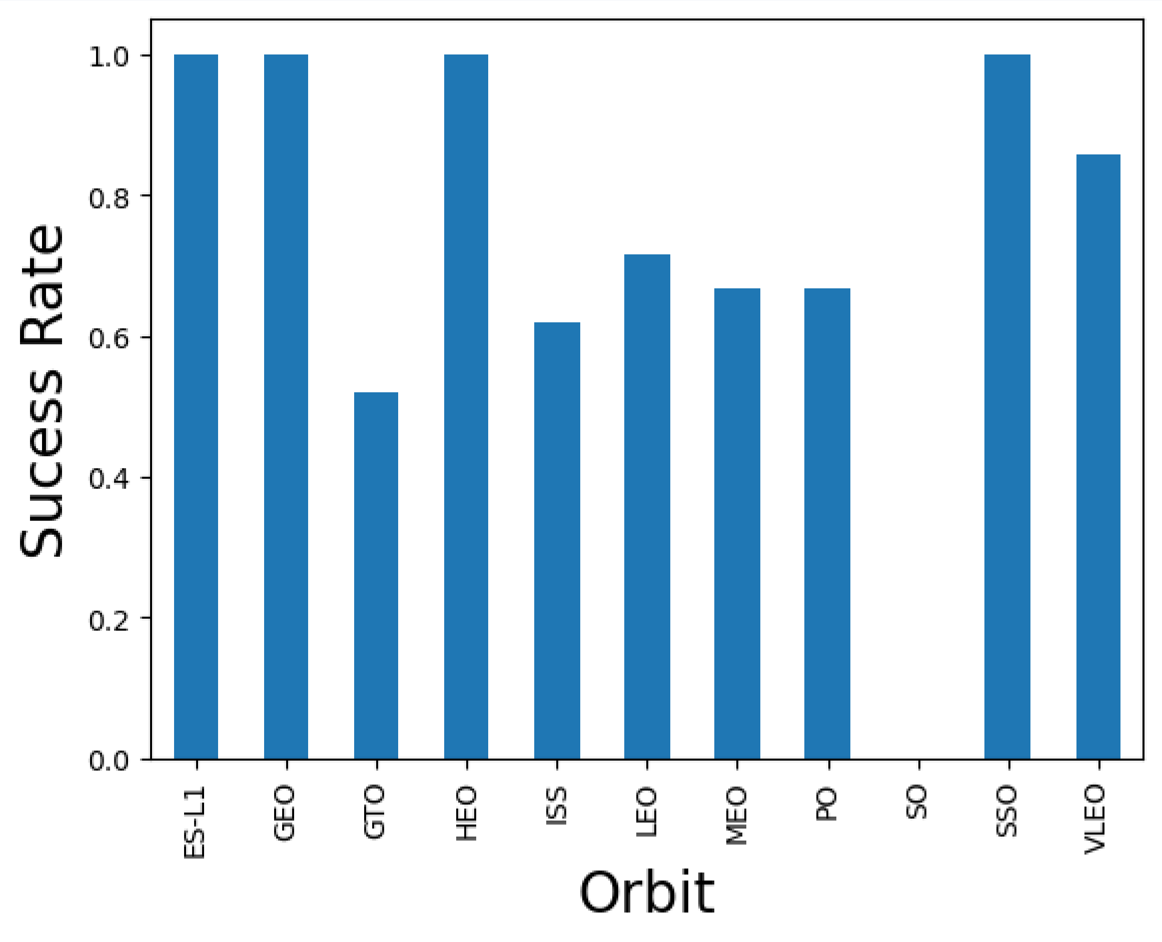


- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

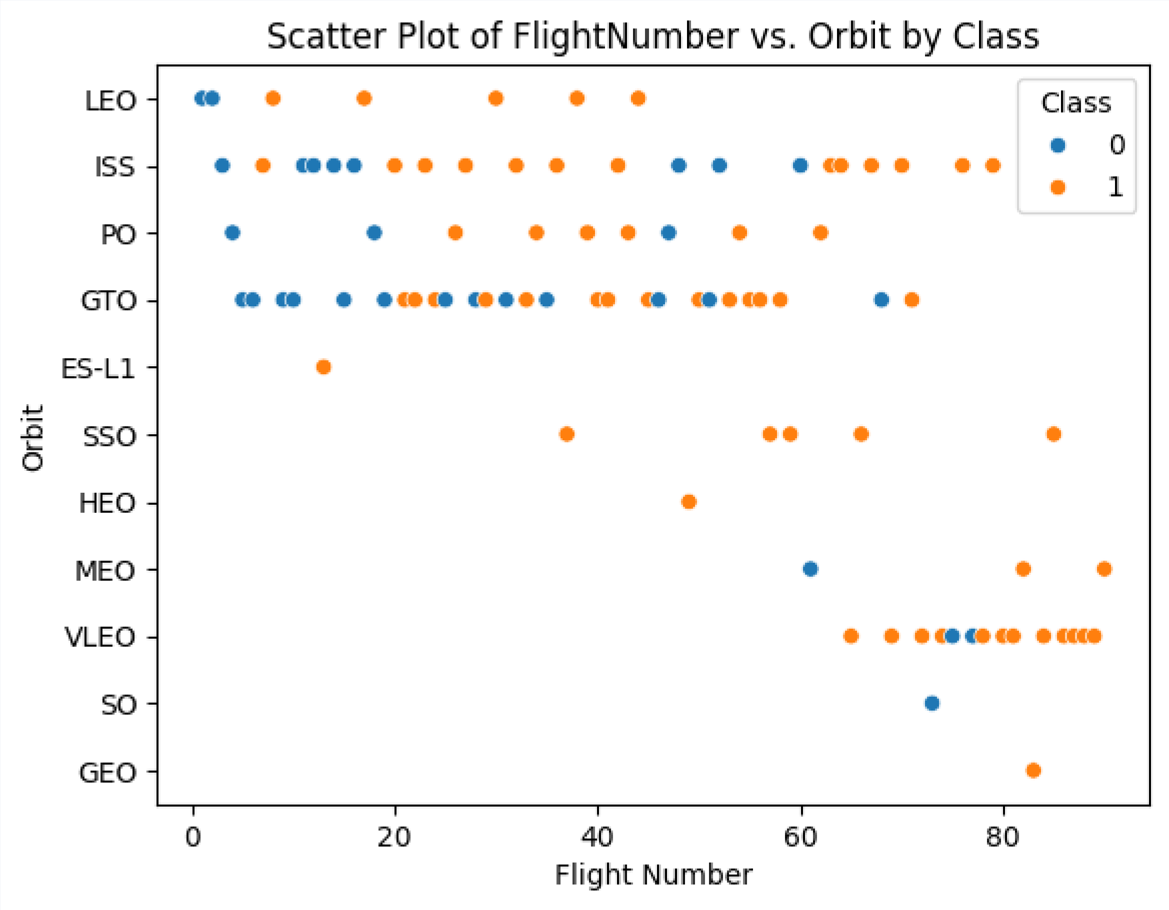
Success Rate vs. Orbit Type

·The biggest success rates happens to orbits:

- ES-L1
- ·GEO
- HEO
- ·SSO
- GTO site has the lowest success rate around 55%



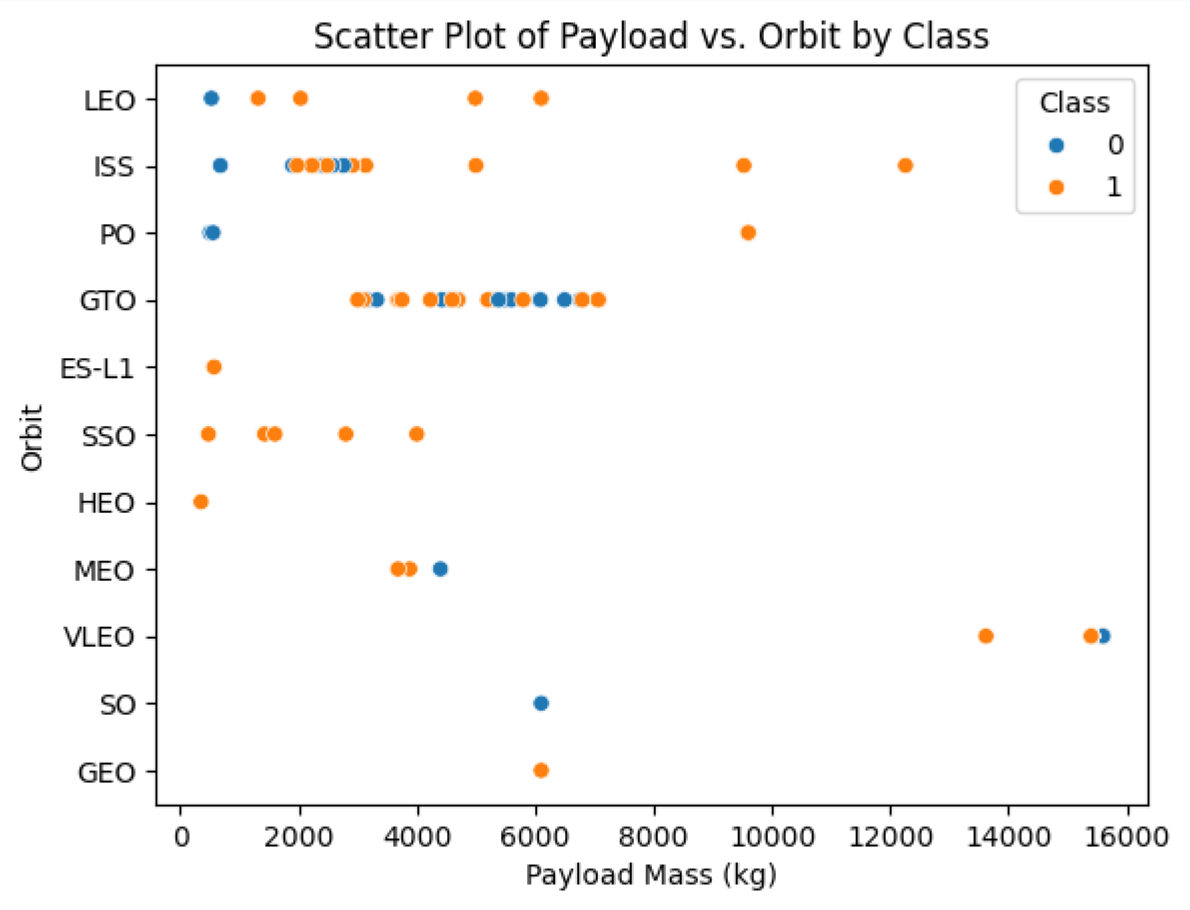
Flight Number vs. Orbit Type



- success rate improved over time to all orbits
- SSO orbit has the best success rate but relatively fewer flight number

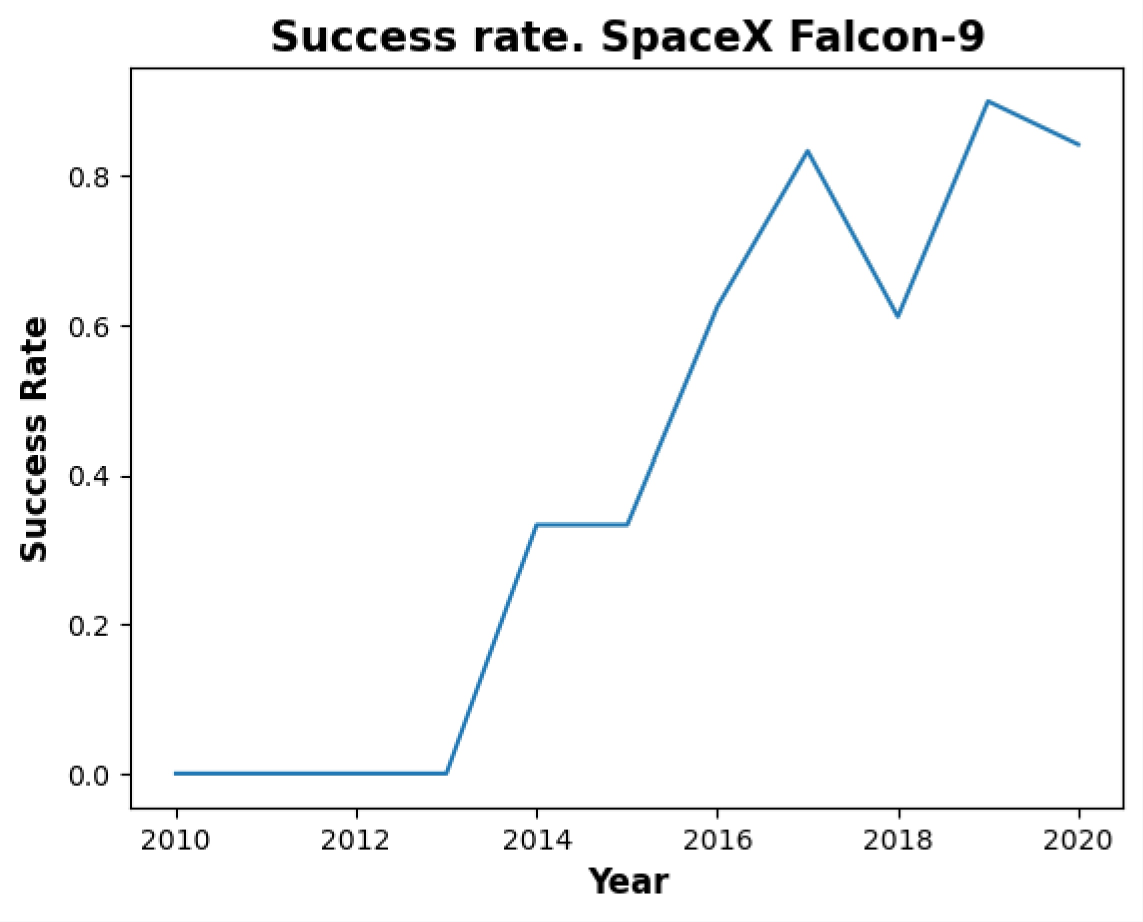
Payload vs. Orbit Type

- ISS (International Space Station) and GTO (GEOSYNCHRONOUS TRANSFER ORBIT) have each one a specific Payload Mass Between 2000 and 4000kg for ISS and 2500 to 7500kg for GTO



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020
- For the period between 2010 and 2013 SpaceX experiment and learn to ride the success curve



All Launch Site Names

- Names of the unique launch sites

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

- Obtained using "SELECT DISTINCT Launch_Site FROM SPACEXTBL"

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

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

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

Total Payload Mass

- Total payload carried by boosters from NASA

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

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

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

```
SELECT AVG(PAYLOAD_MASS__KG_) AS AveragePayloadMass FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.1%' ;
```

AveragePayloadMass
2534.6666666666665

First Successful Ground Landing Date

- First successful landing outcome on ground pad

```
SELECT MIN(DATE) AS FirstSuccessfulGroundPadLandingDate FROM SPACEXTBL
WHERE Landing__Outcome = 'Success (ground pad)';
```

FirstSuccessfulGroundPadLandingDate
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = ‘Success (drone ship)’ AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

SELECT SUM(CASE WHEN Landing_Outcome LIKE '%Success%' THEN 1 ELSE 0 END) AS TotalSuccess, SUM(CASE WHEN Landing_Outcome LIKE '%Failure%' THEN 1 ELSE 0 END) AS TotalFailure FROM SPACEXTBL;

TotalSuccess	TotalFailure
61	10

Boosters Carried Maximum Payload

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

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

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
SELECT 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, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)';
```

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	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

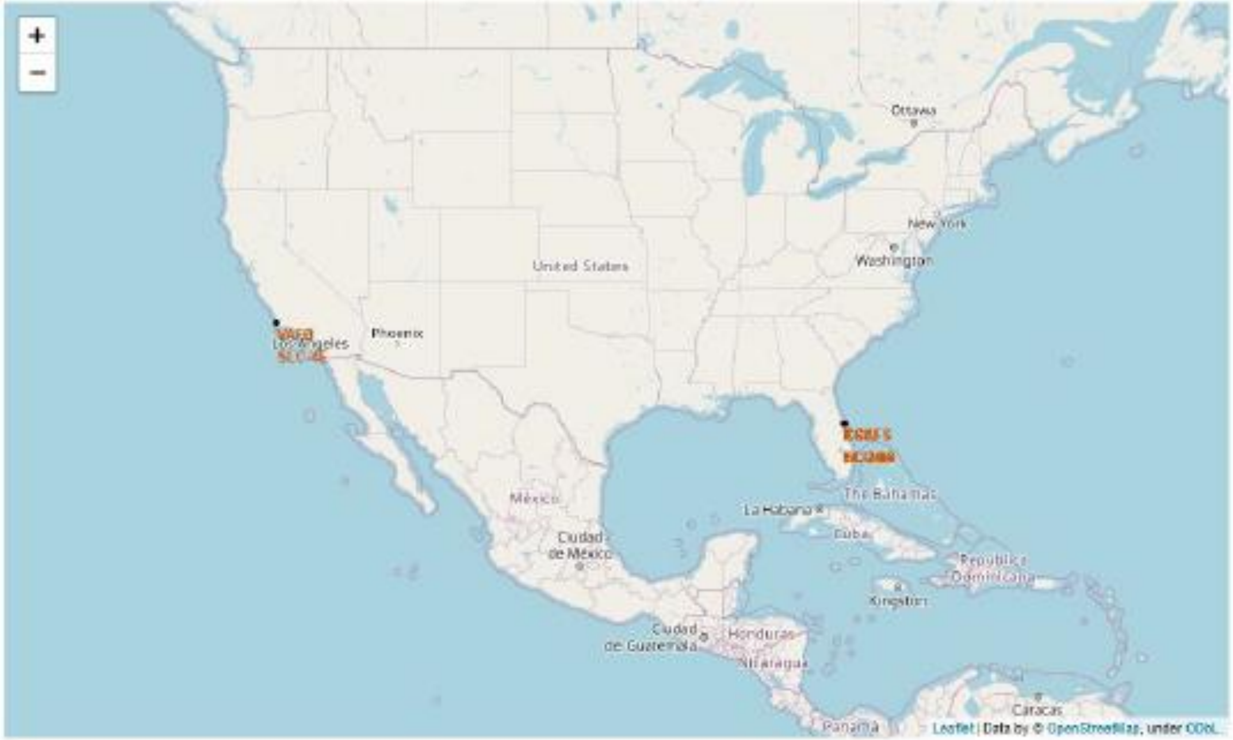
```
%sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTBL WHERE Date
BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY
Count Desc;
```

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

Section 3

Launch Sites Proximities Analysis

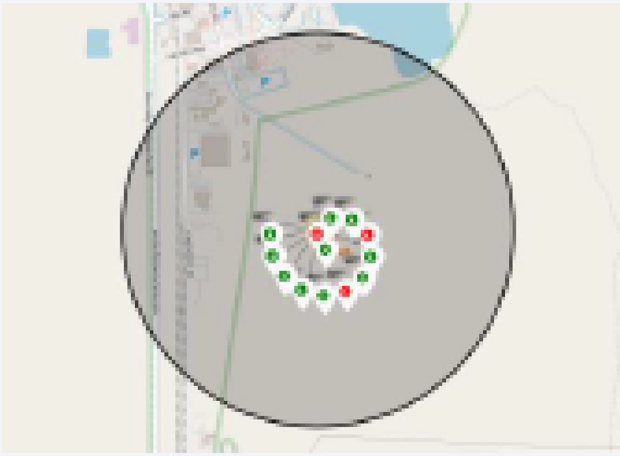
launch sites



- Launch sites are near sea, for safety reason, Launching a rocket from the coast ensures that if something goes wrong during the ascent, the debris will fall into the ocean instead of a densely populated area12. This is a significant safety measure to protect people and property
- Launch sites in proximity to the Equator line Launching from the equator makes it easier to reach certain types of orbits, particularly geostationary orbits, which are useful for communication satellites

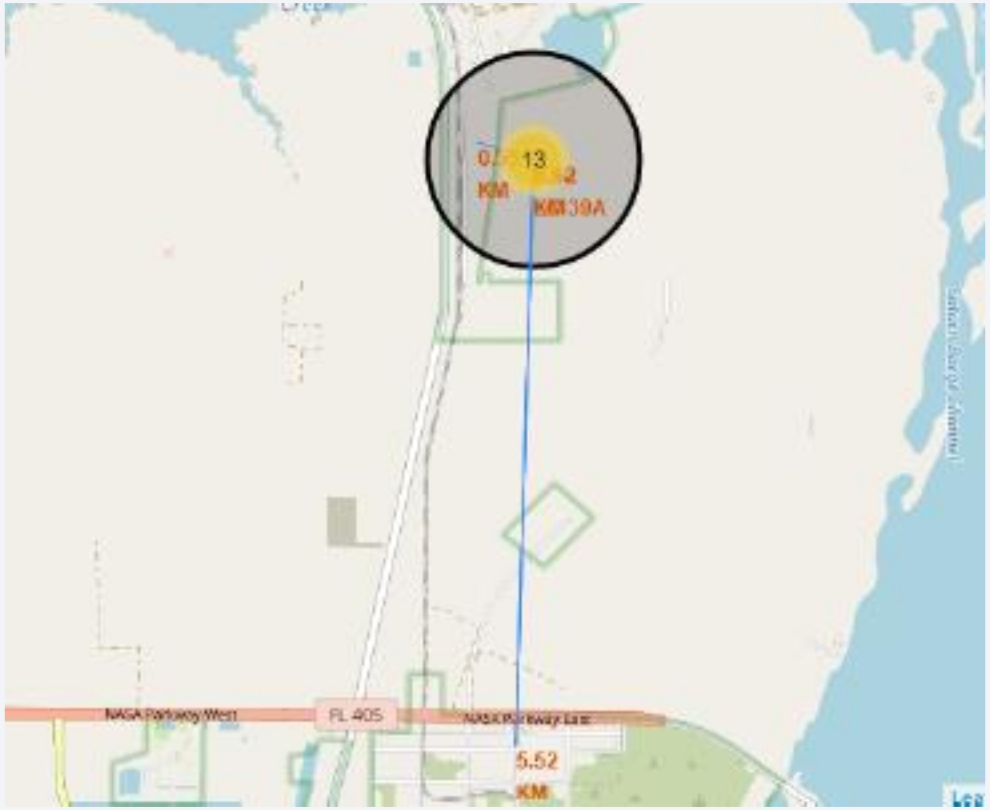
Launch Outcomes by Site

- From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.



Logistics and Safety

- Proximity to infrastructure such main roads, railways and ports and airports, is an important factor to chose a launching site

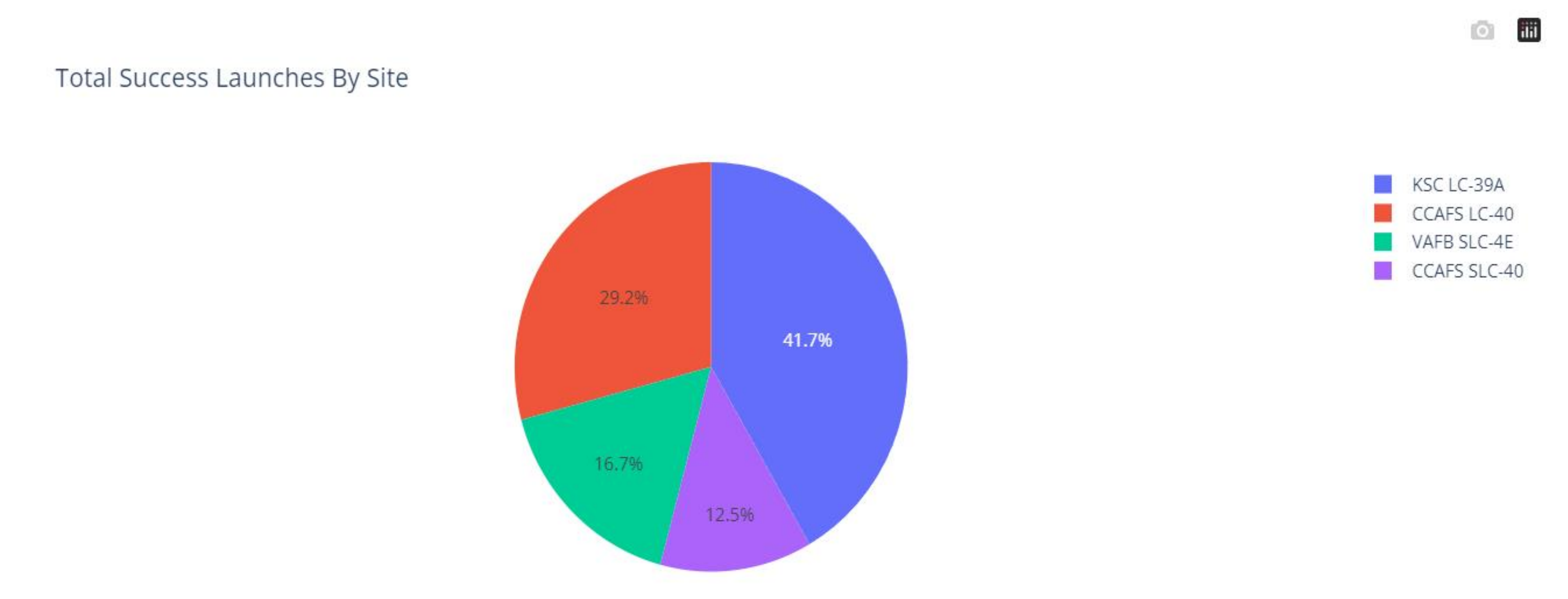


Section 4

Build a Dashboard with Plotly Dash

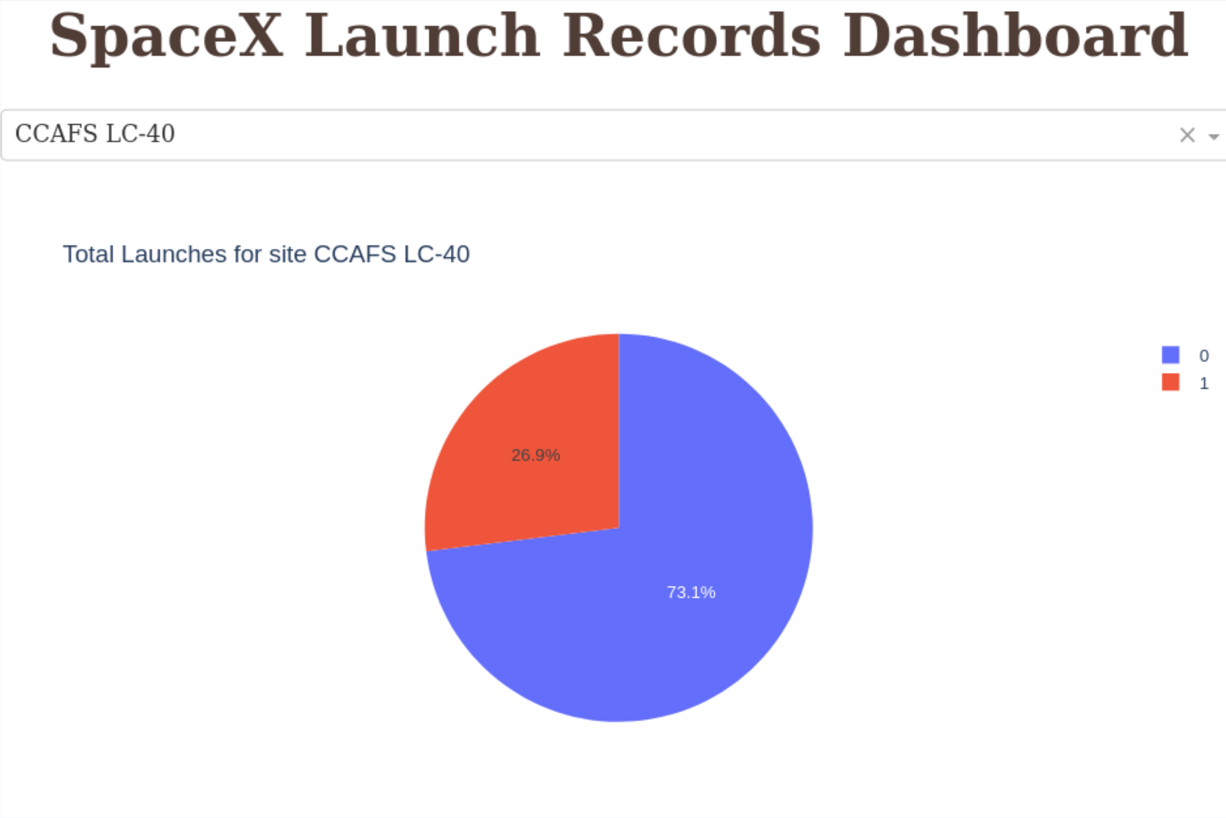
Successful Launches by Site

- KSC LC-39A is the launche site with the most successful launches



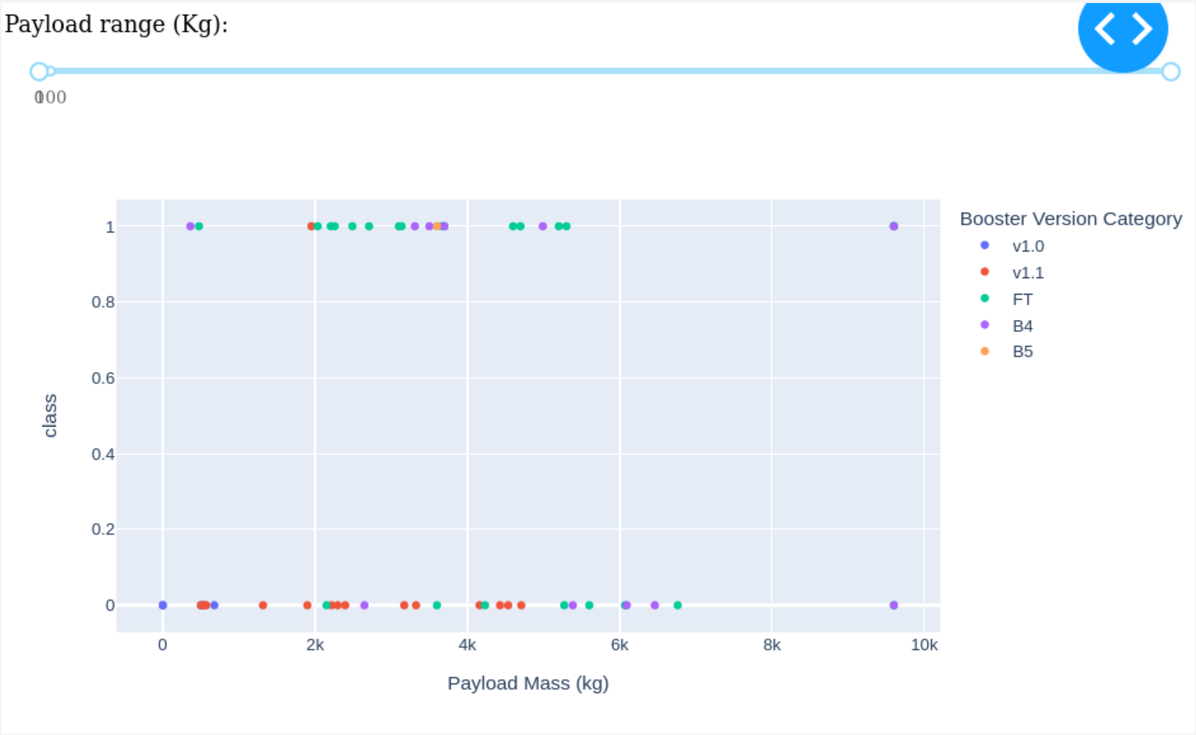
Launch Success Ratio for KSC LC-39A

- 76.9% of launches are successful in this site.



Payload vs. Launch Outcome

- Payloads under 6,000kg and FT boosters are the most successful combination.

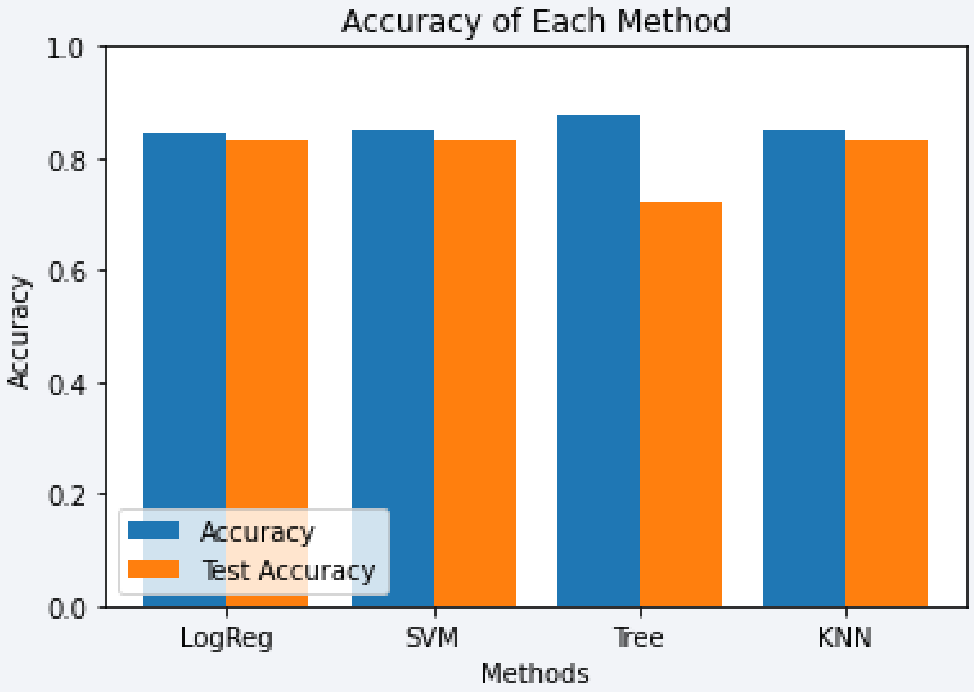


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- ·Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets

- The best launch site is KSC LC-39A
- Decision Tree Classifier can be used to predict successful landings and increase profits

Appendix

- Folium didn't show maps on Github, so I took screenshots.

Thank you!