IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

Project background and context

The space industry is on the cusp of a revolution, with private companies like SpaceX, Blue Origin, and Virgin Galactic leading the charge. As the demand for space travel and satellite launches continues to grow, the cost of access to space remains a significant barrier. Among the key players, SpaceX has disrupted the market with its reusable rockets, significantly reducing the cost of launches. The ability to reuse the first stage of a rocket is a game-changer, and other providers, like SpaceY, are eager to follow suit. However, the success of reuse hinges on various factors, including the condition of the rocket, atmospheric conditions, and technological capabilities.

Problems you want to find answers

This machine learning project aims to develop a predictive model that determines whether SpaceY will successfully reuse its first stage, enabling the company to estimate the cost of launches accurately. By analyzing historical data on rocket launches, weather conditions, and technological advancements, our model will identify patterns and correlations that influence the likelihood of successful reuse. The project's objectives are:

1. Predict Reusability: Develop a machine learning algorithm that predicts the probability of successful first-stage reuse for SpaceY rockets.
2. 2. Cost Estimation: Use the predicted reusability to estimate the cost of launches, enabling SpaceY to remain competitive in the market.
3. 3. Identify Key Factors: Determine the most significant factors influencing reusability, providing insights for SpaceY to optimize its technology and launch strategies.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

- Describe how data sets were collected.

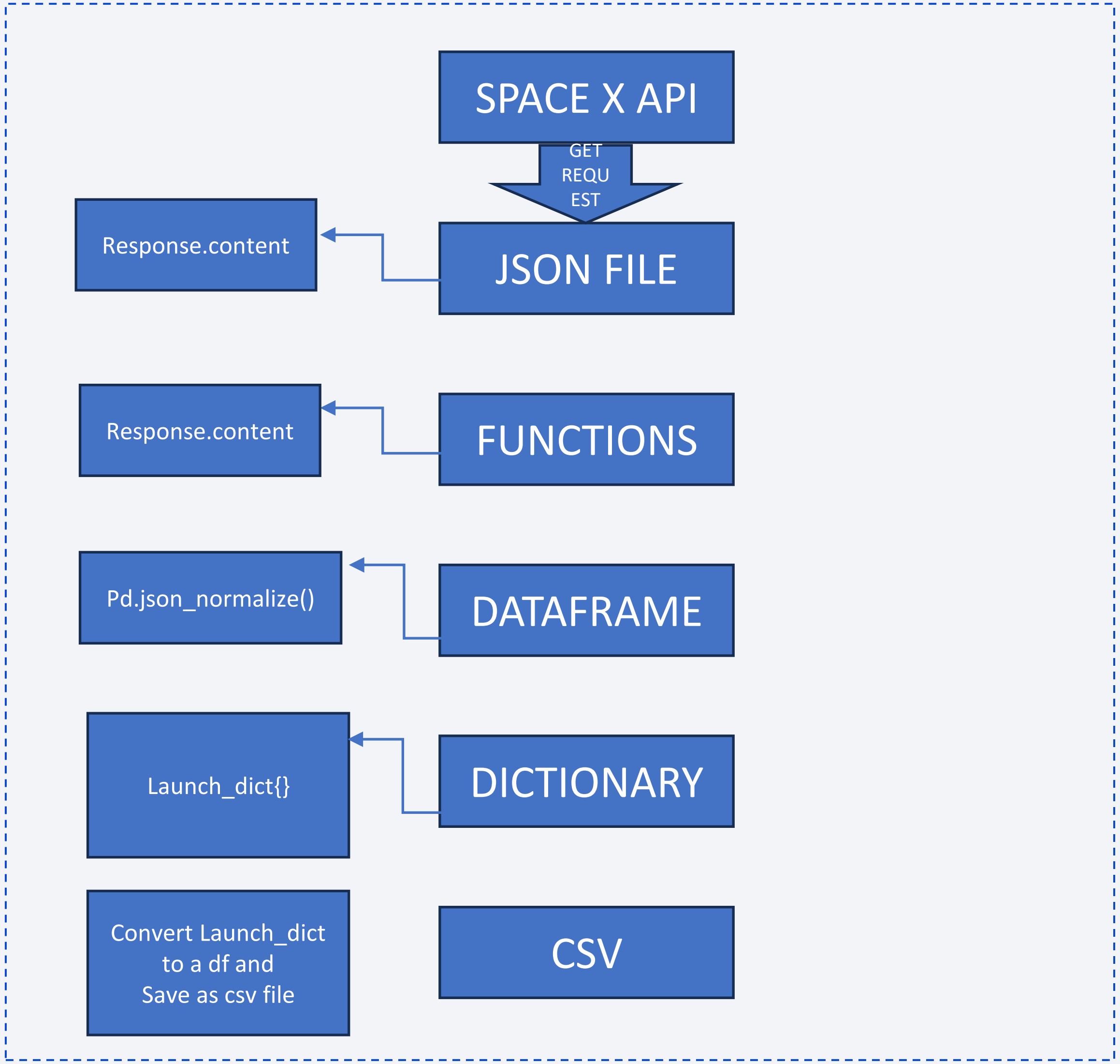
    Data was collected from space X API using python script and

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Wrote a GET request to extract data from SpaceX API and process the returned byte object into a data Frame.

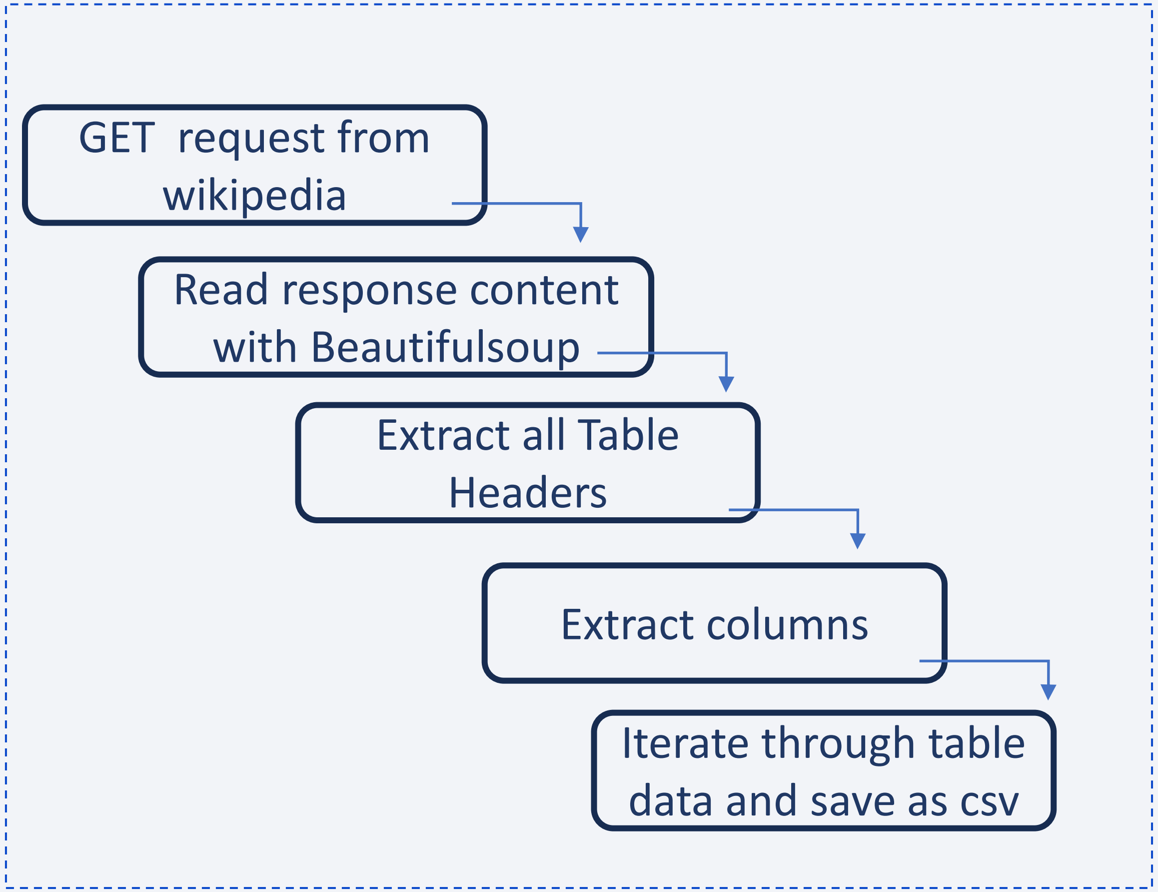
- https://github.com/Hor-layinka/Abioye-Data_projects/blob/HOBS/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data scraped from Wikipedia using BeatifulSoup, parsed the html object to dictionary of list then converted to a df.

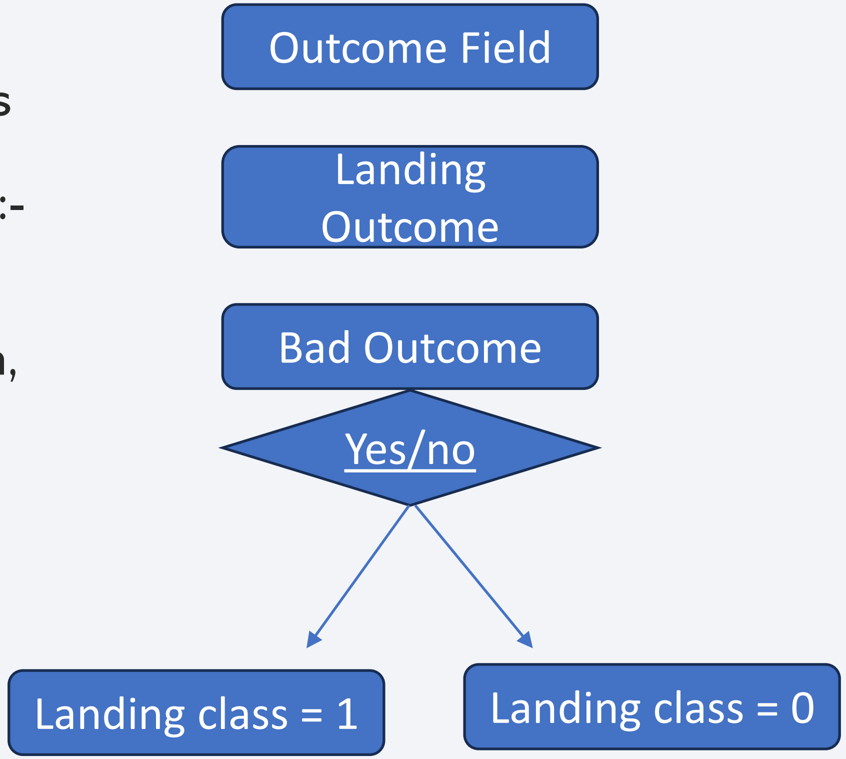
- https://github.com/Hor-layinka/Abioye-Data_projects/blob/HOBS/jupyter-labs-webscraping-bak-2024-05-25-12-31-03Z.ipynb



Data Wrangling

To prepare the data for training, we extracted the necessary labels from the Outcome column, which indicates whether a mission landed successfully (True) or not (False). We created two variables:- Landing Outcomes (True): Representing successful landings- Bad Landing Outcomes (False): Representing unsuccessful landings, equivalent to 0 and 1 respectively. We then created a new column, Landing Class which consist of 0 and 1.

https://github.com/Hor-layinka/Abioye-Data_projects/blob/HOBS/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites begin with the string 'CCA’
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listed the   names of the booster versions which have carried the maximum payload mass
- Listed the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- Ranked 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.

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

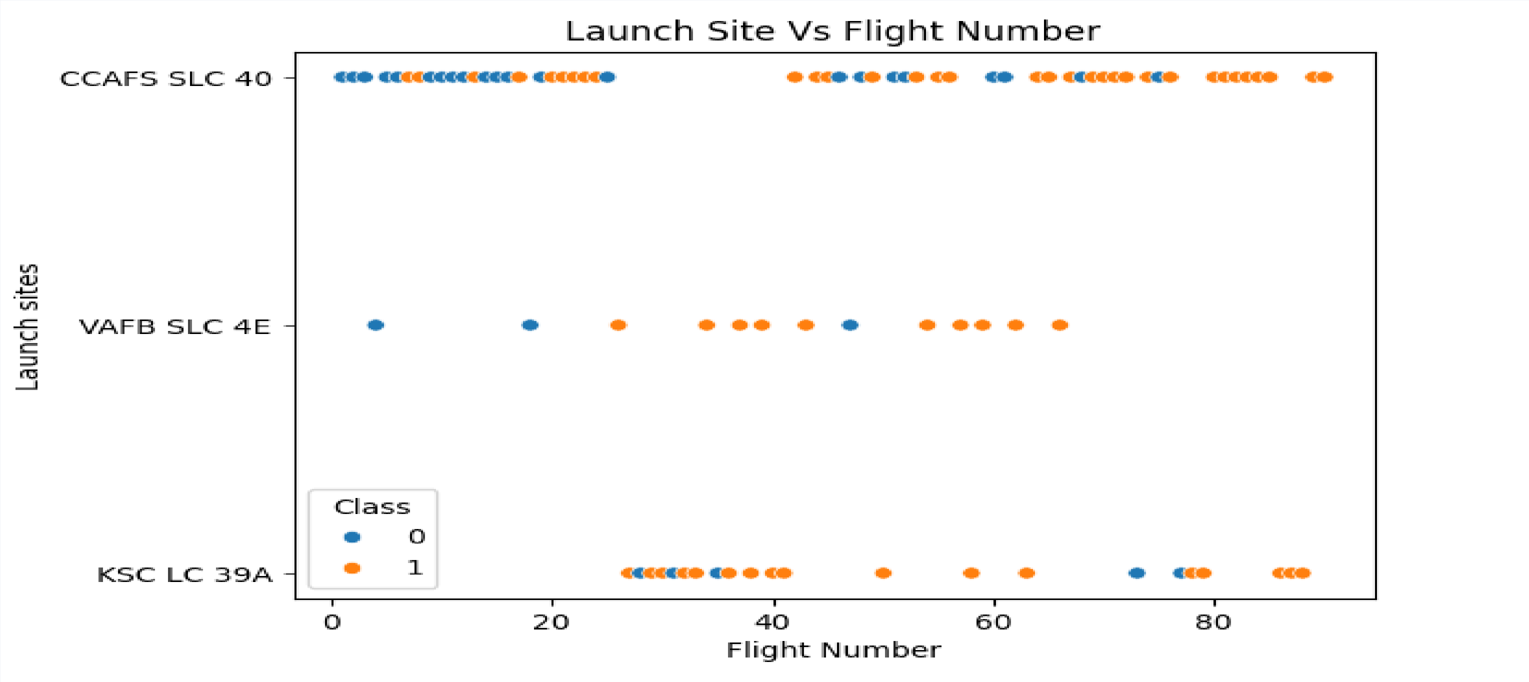
Results

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

Section 2

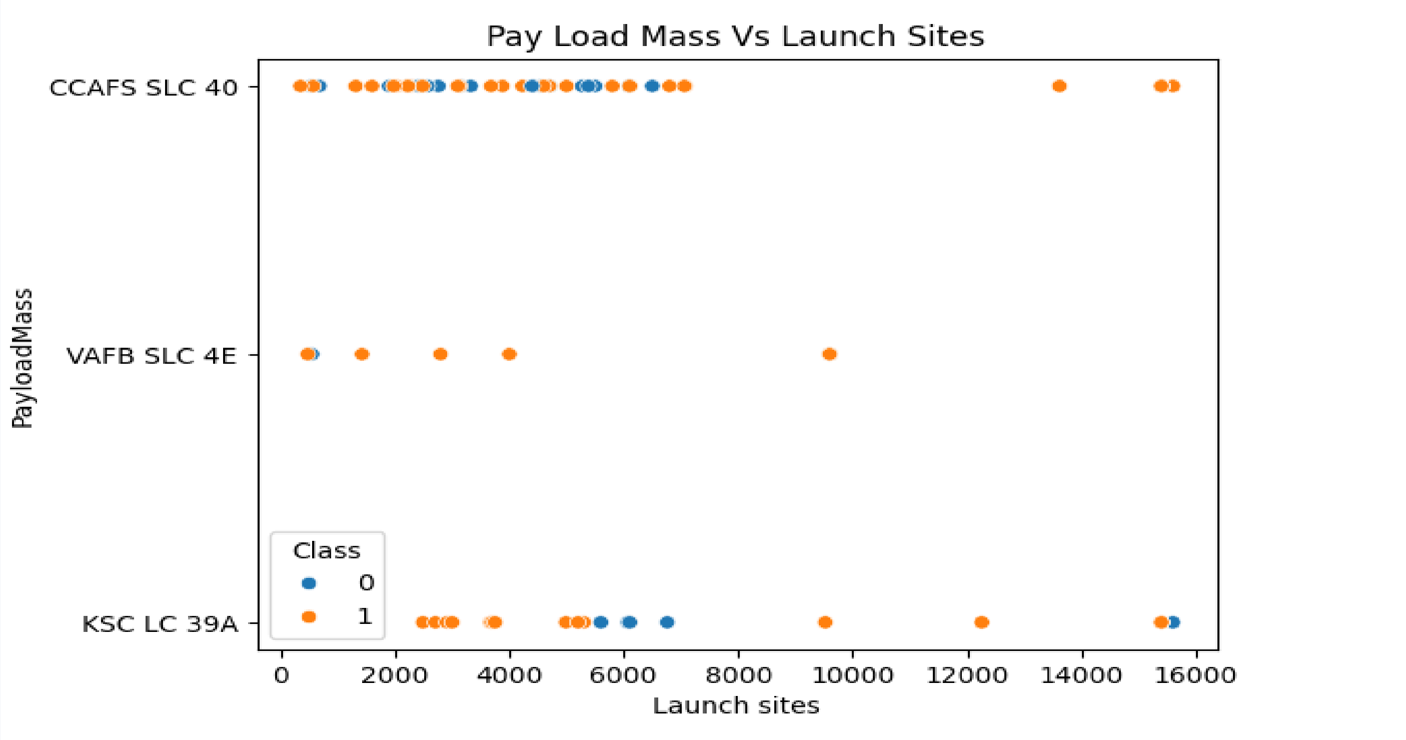
Insights drawn from EDA

Flight Number vs. Launch Site



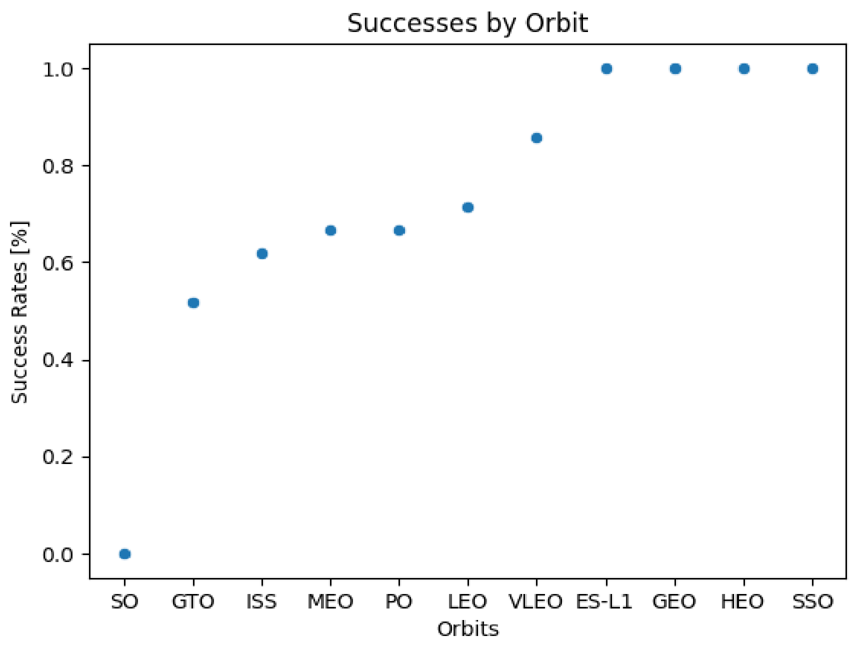
- The relationship between Light number and Launch Sites appears to be very weak almost all the launch sites recorded more successes than failure.

Payload vs. Launch Site

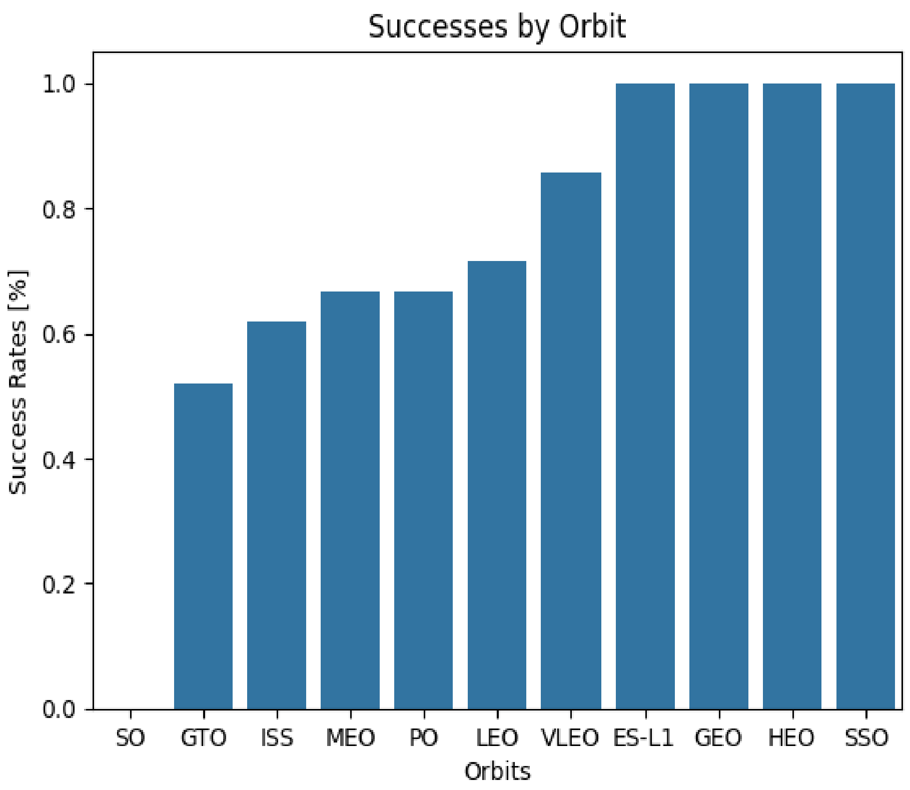


Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

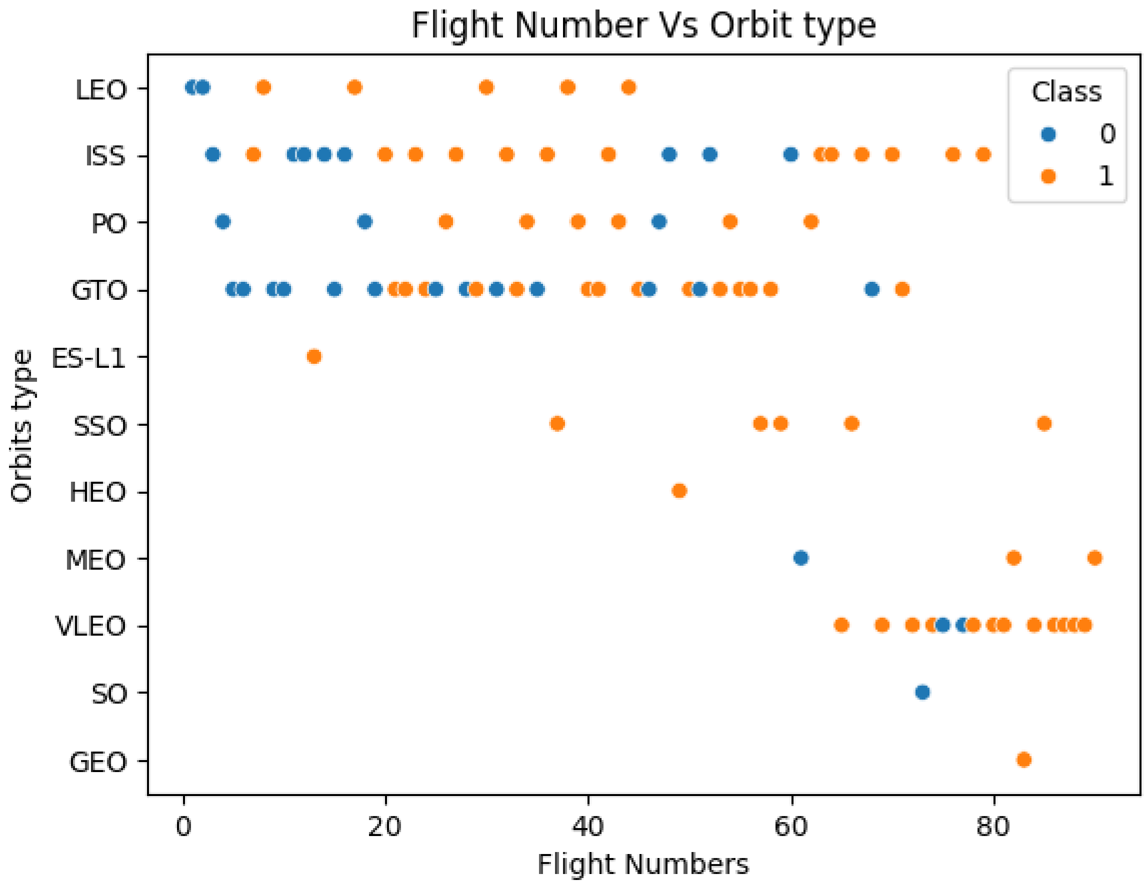
Success Rate vs. Orbit Type



- Orbits ES-L1, GEO, HEO and SSO a 100% success rates with SO orbit having the lowest succes rate of 0.0%

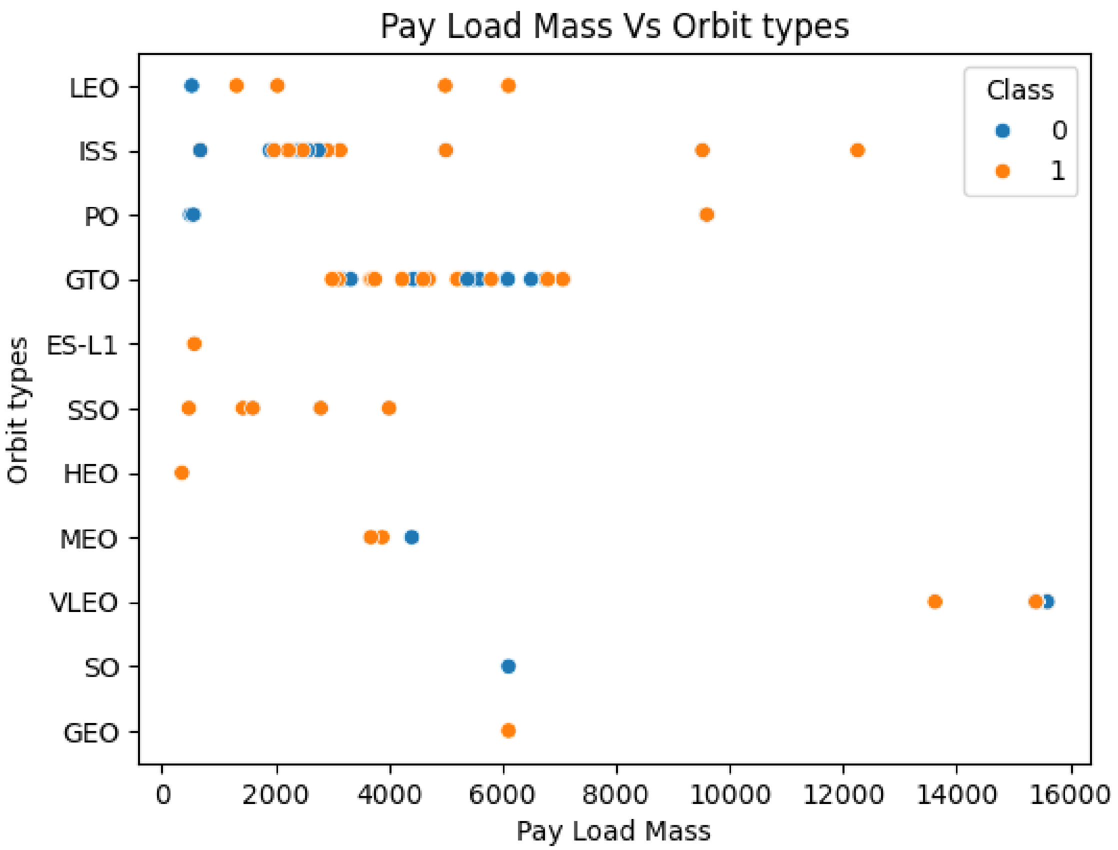


Flight Number vs. Orbit Type



You should 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.

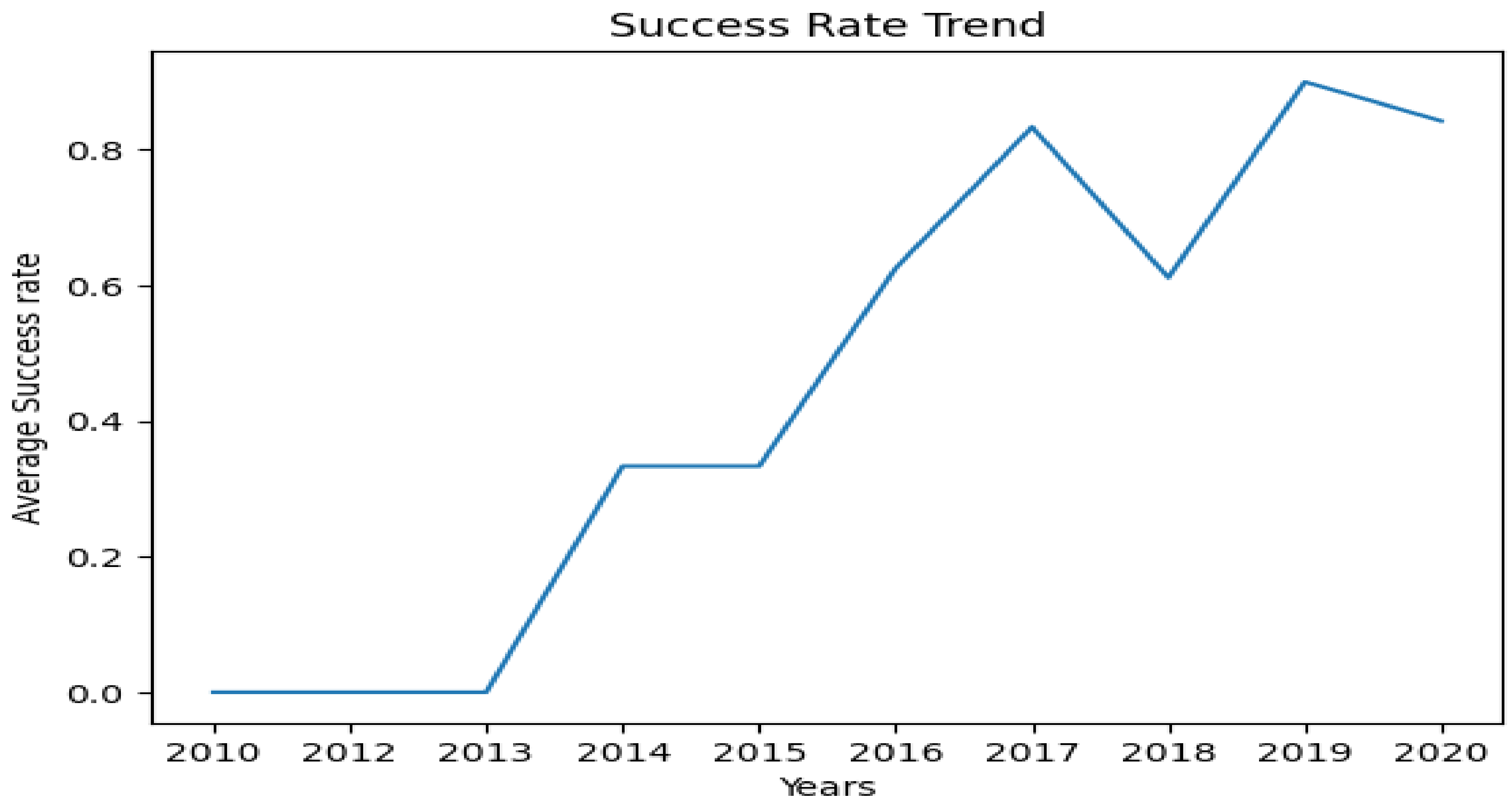
Payload vs. Orbit Type



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



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

select distinct("Launch_Site")
from SPACEXTABLE

The query returns Unique Launch site for SPACEX rockets

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

Launch Site Names Begin with 'CCA'

```
select *
from SPACEXTABLE
where "Launch_Site" LIKE "CCA%"
Limit 5
```

This query returns the first five records of SPACEX Launch sites starting with the word CCA

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

```
SELECT sum(PAYLOAD_MASS__KG_ ) as Total_payload
FROM SPACEXTABLE
WHERE "customer" == "NASA (CRS)“
```

Total_payload
45596

Returns the total payload mass for customer NASA (CRS)

Average Payload Mass by F9 v1.1

SELECT avg(PAYLOAD_MASS__KG_) as avg_payload

FROM SPACEXTABLE

WHERE "Booster_Version" == "F9 v1.1“

2928.4

Return the Average payload mass for boster version F9 v1.1

First Successful Ground Landing Date

SELECT min(Date) as First_landing_groundPad

FROM SPACEXTABLE

WHERE Landing_Outcome == "Success (ground pad)“

Return the date for the first successful Launch for Ground Landing

First_landing_groundPad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
SELECT Booster_Version
FROM SPACEXTABLE
WHERE "Landing_Outcome" == "Success (drone ship)"
AND PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000
```

The query returns the booster versions of successful drone ship Landing by filtering the table to only success recorded for drone ship and using the between to get payload within 4000 and 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

SELECT (select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome like "success%") as successes, count(Mission_Outcome) as failures

FROM SPACEXTABLE

where Mission_Outcome like "failure%"

successes	failures
100	1

Boosters Carried Maximum Payload

```
select Booster_Version
from (
   select Booster_Version, substr(Date,0,5) Years,
max(PAYLOAD_MASS__KG_)
   from SPACEXTABLE
   group by 2
   )
```

The returns maximum payloads carried by boosters, the inner query returns maximum payloads group by Years and the outer query returns the name of the bosters

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0005
F9 v1.1
F9 v1.1
F9 v1.1 B1016
F9 FT B1020
F9 FT B1029.1
F9 B4 B1041.2
F9 B5 B1048.4
F9 B5 B1049.4

2015 Launch Records

select "Date", substr(Date,6,2) as Months, substr(Date,0,5) as Years, Landing_Outcome,Booster_Version,Launch_Site
from SPACEXTABLE
where Years =="2015“

substr() was used to extract the year from date column, then the table is filtered to year 2015 only to get Launch records for 2015

Date	Months	Years	Landing_Outcome	Booster_Version	Launch_Site
2015-01-10	01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-02-11	02	2015	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
2015-03-02	03	2015	No attempt	F9 v1.1 B1014	CCAFS LC-40
2015-04-14	04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
2015-04-27	04	2015	No attempt	F9 v1.1 B1016	CCAFS LC-40
2015-06-28	06	2015	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
2015-12-22	12	2015	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

```
select "Date", substr(Date,6,2) as Months,
substr(Date,0,5) as Years,
Landing_Outcome,Booster_Version,Launch_Site

from SPACEXTABLE

where Years =="2015"

limit 10
```

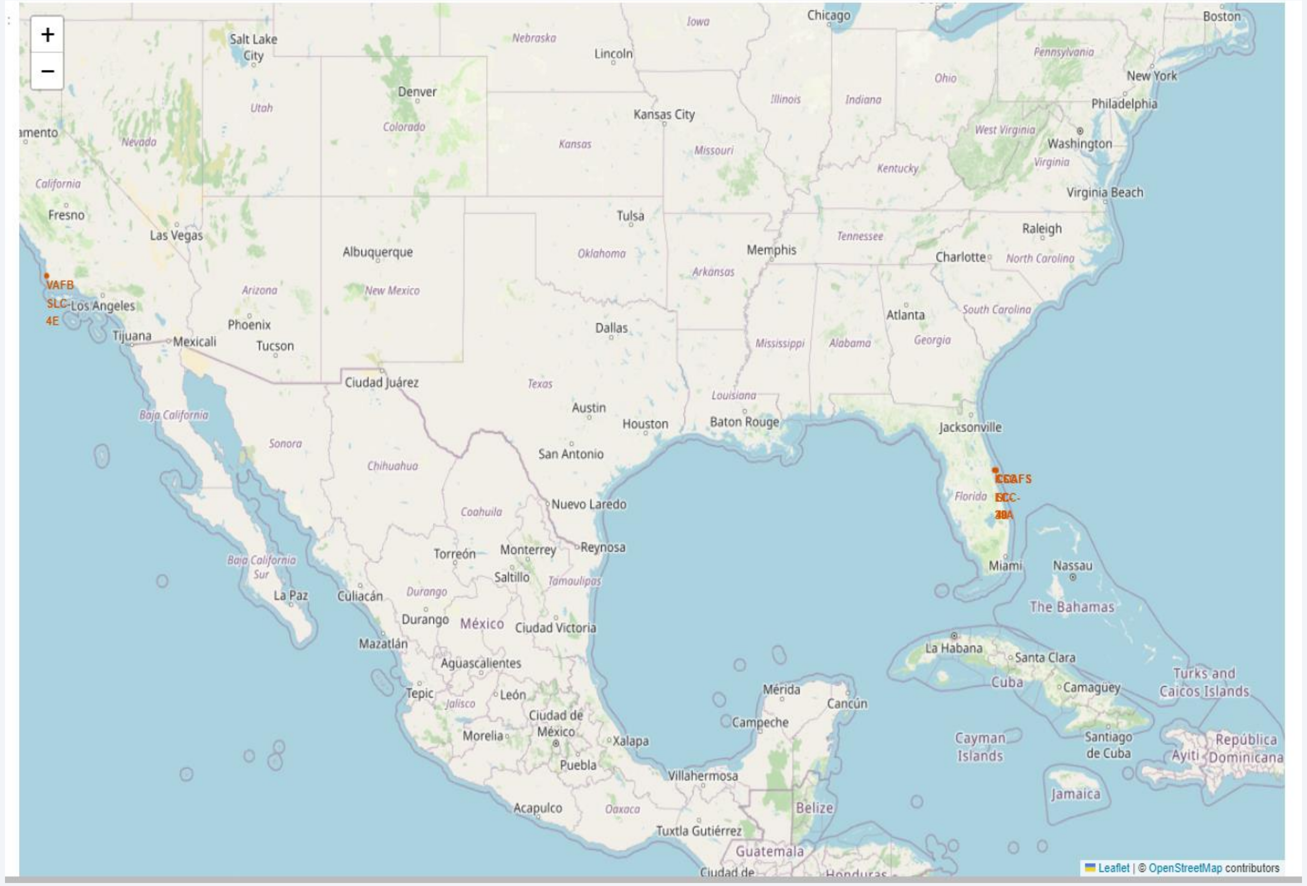
Landing_Outcome	counts
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

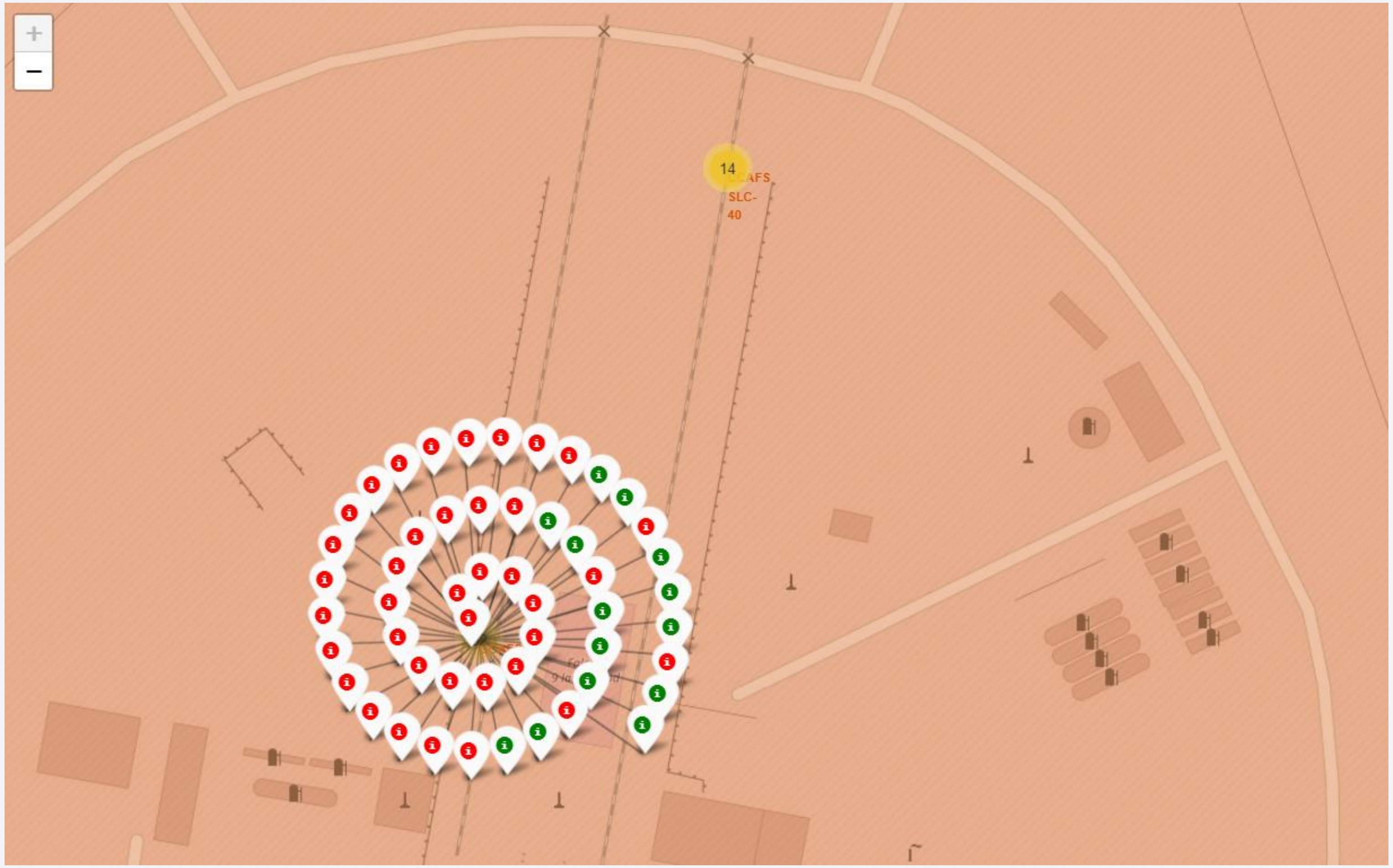
Launch Site Locations

- Explain the important elements and findings on the screenshot



Successful vs Failed Launches

Explain the important elements and findings on the screenshot



<Folium Map Screenshot 3>

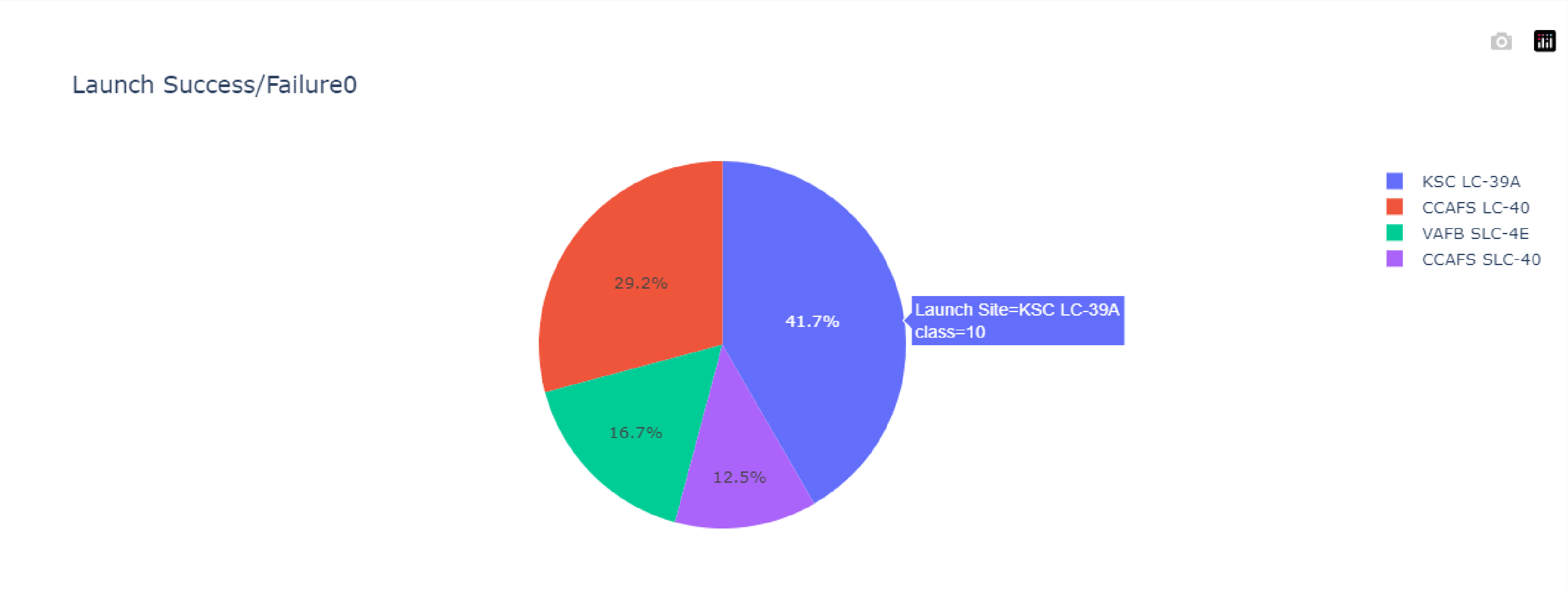
Explain the important elements and findings on the screenshot

Section 4

Build a Dashboard with Plotly Dash

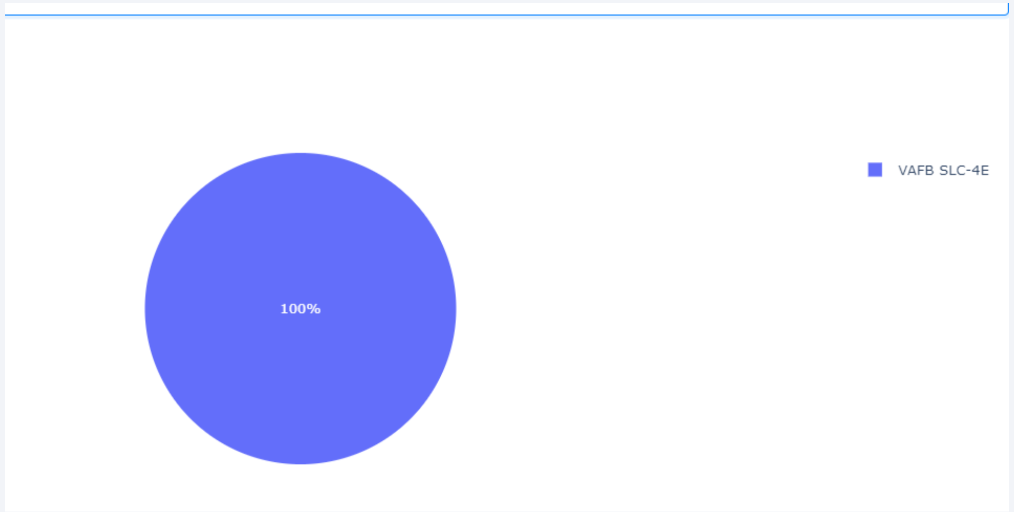
<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot



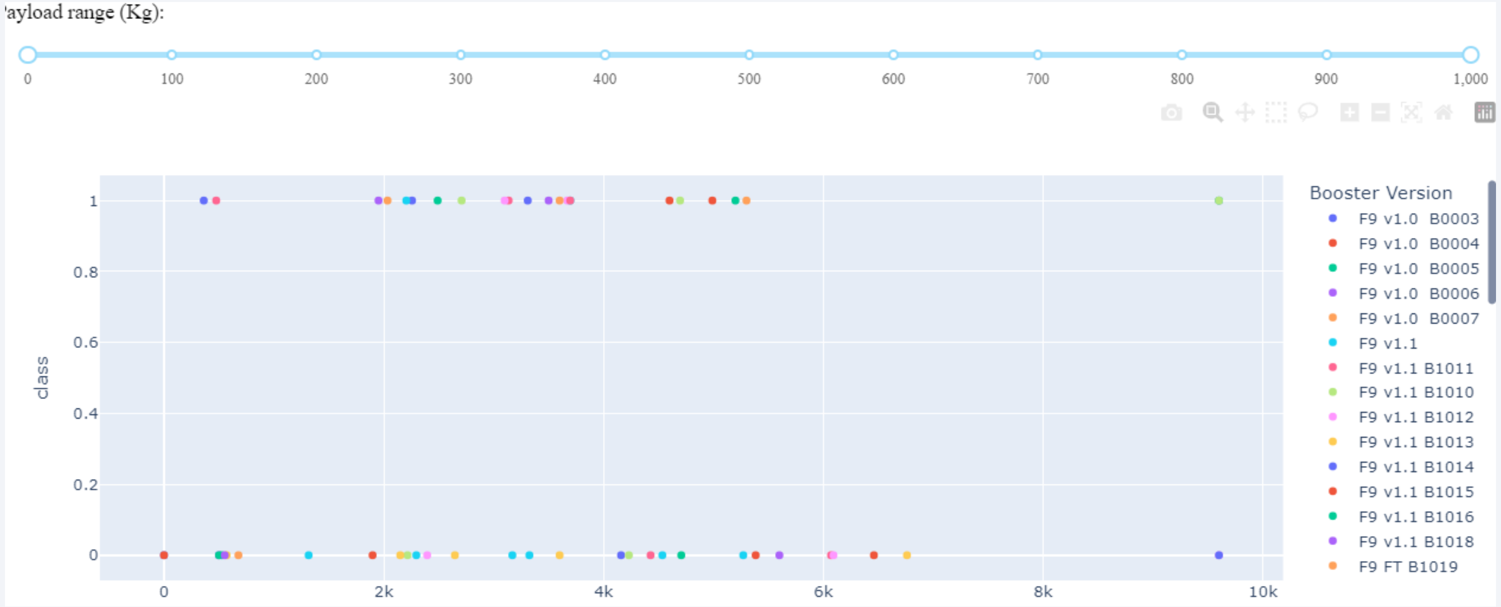
<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot



<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.



Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!