Winning Space Race with Data Science

JIBRIN HARUNA KASIM

10TH SEPT. 2021

EXECUTIVE SUMMARY

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

Executive Summary

➢Summary of methodologies

- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

➢Summary of all results

Exploratory data analysis results, Interactive demo in screenshorts and Predictive analysis results

Introduction

- **Project background and context**

We predicted if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch

- **Problems you want to find answers**

- What influences if the rocket will land successfully?
- The effect each relationship with certain rocket variables will impact in determining the success rate of a successful landing.
-What conditions does SpaceX have to achieve to get the best results and ensure the best rocket success landing rate.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology

-SpaceX API (webscraping) from wikipidia

- Perform data wrangling

 - One Hot Encoding data fields for Machine Learning and dropping irrelevant columns

 Performed exploratory data analysis (EDA) using visualization and SQL

  - Plotting : Scatter Graphs, Bar Graphs to show relationships between variables to show patterns of data

- 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 following datasets were collected

We worked with SpaceX launch data that is gathered from the SpaceX REST API.
This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

Our goal is to use this data to predict whether SpaceX will attempt to land a rocket or not.

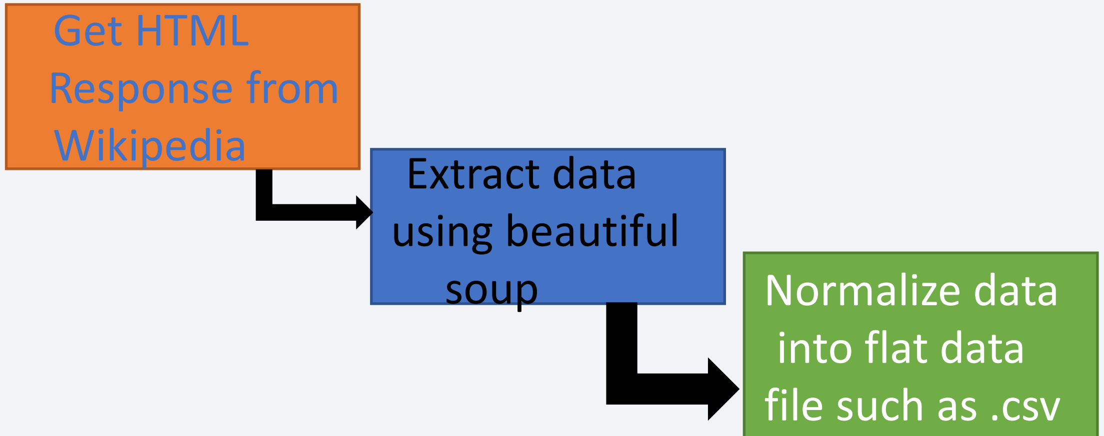
- The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.

Another data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using

BeautifulSoup.

SpaceX API

Web Scrapping



EDA with SQL

Performed SQL queries to gather information about the dataset.

For example of some questions we were asked about the data we needed information about. Which we are using SQL queries to get the answers in the dataset :

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'KSC'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster_versions which have carried the maximum payload mass.
- Listing the records which will display the month names, successful landing_outcomes in ground pad ,booster_versions, launch_site for the months in year 2017.

Build an Interactive Map with Folium

To visualize the Launch Data into an interactive map. We took the Latitude and Longitude Coordinates at each launch site and added a *Circle Marker around each launch site with a label of the name of the launch site.*

We assigned the dataframe launch_outcomes(failures, successes) to *classes*

0 and 1 with Green and Red markers on the map in a MarkerCluster()

- **Using Haversine's formula we calculated the distance** from the Launch Site to various landmarks to find various trends about what is around the Launch Site to measure patterns. **Lines** are drawn on the map to measure distance to landmarks

- **Example of some trends in which the Launch Site is situated in**

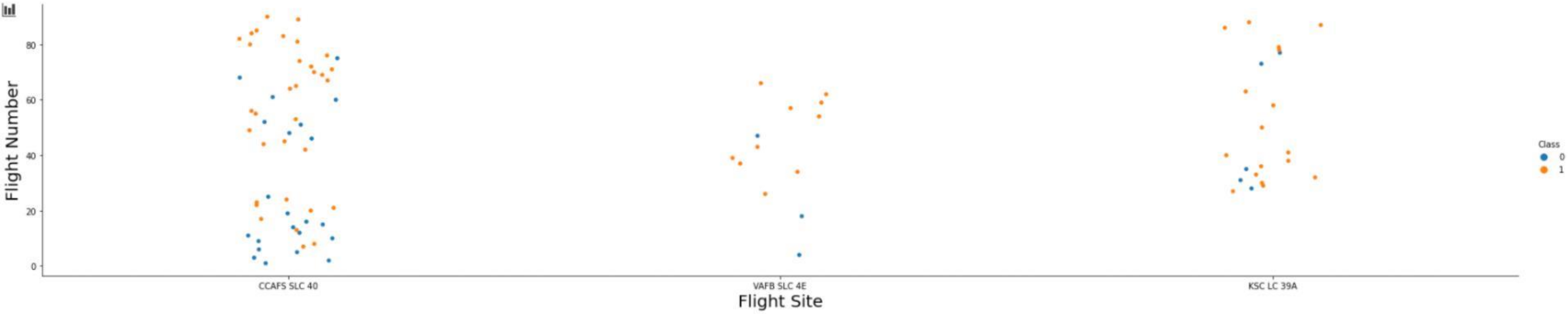
- **Example of some trends in which the Launch Site is situated in.**
- •Are launch sites in close proximity to railways? No
- •Are launch sites in close proximity to highways? No
- •Are launch sites in close proximity to coastline? Yes
- •Do launch sites keep certain distance away from cities? Yes

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

Section 2

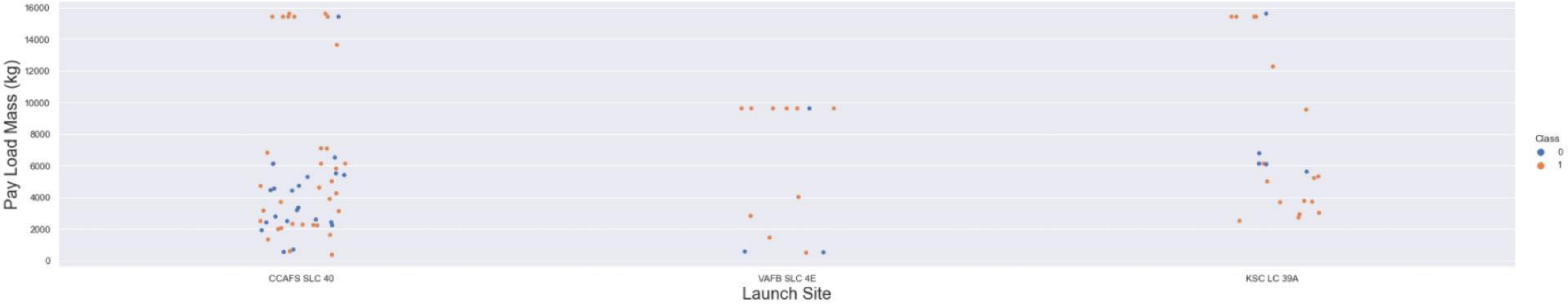
Insights drawn from EDA

Flight Number vs. Flight Site



The more amount of flights at a launch site the greater the success rate at a launch site.

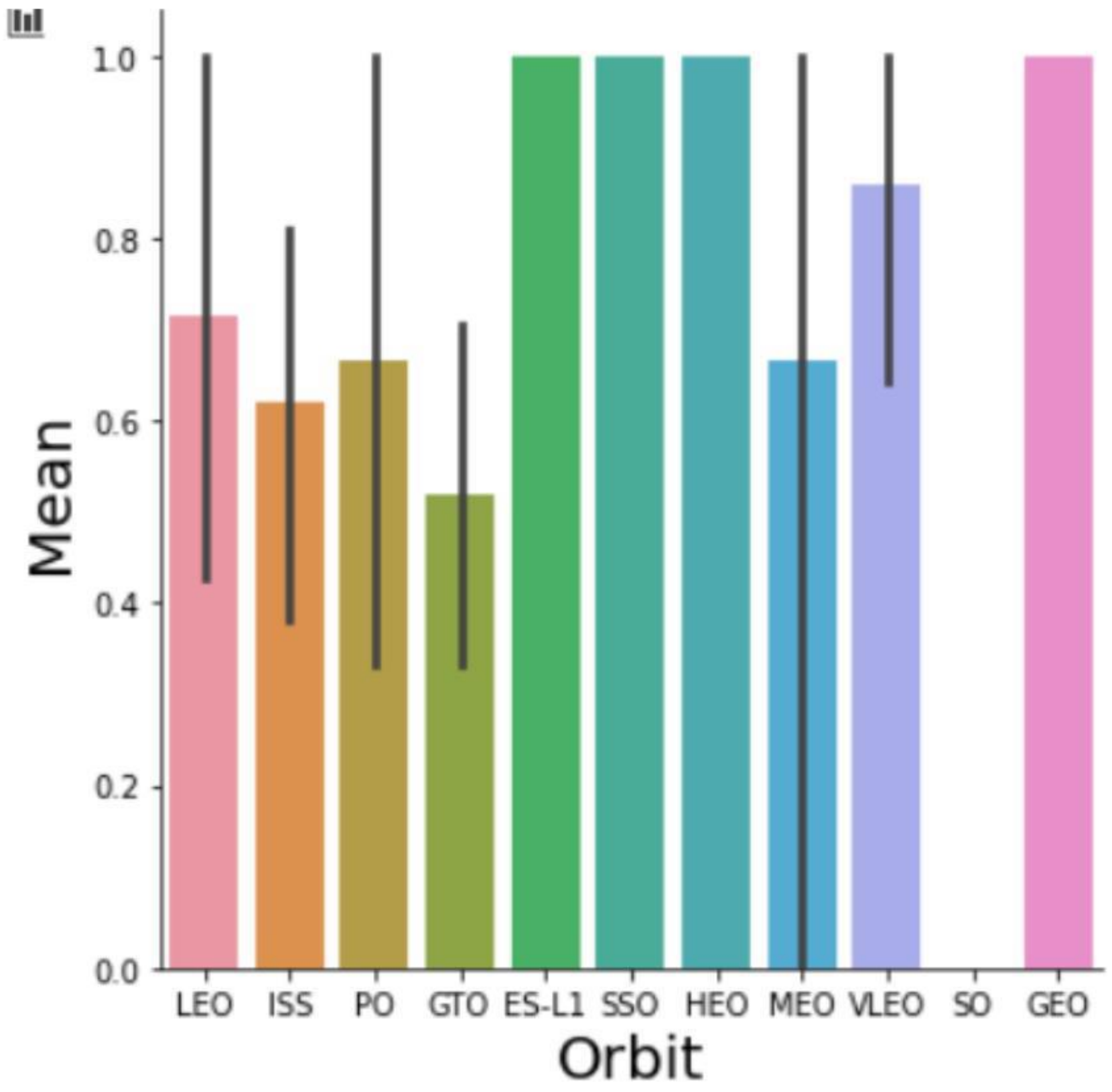
Payload Mass vs. Launch Site



The greater the payload mass for Launch Site CCAFS SLC 40 the higher the success rate for the Rocket. There is not quite a clear pattern to be found using this visualization to make a decision if the Launch Site is dependant on Pay Load Mass for a success launch.

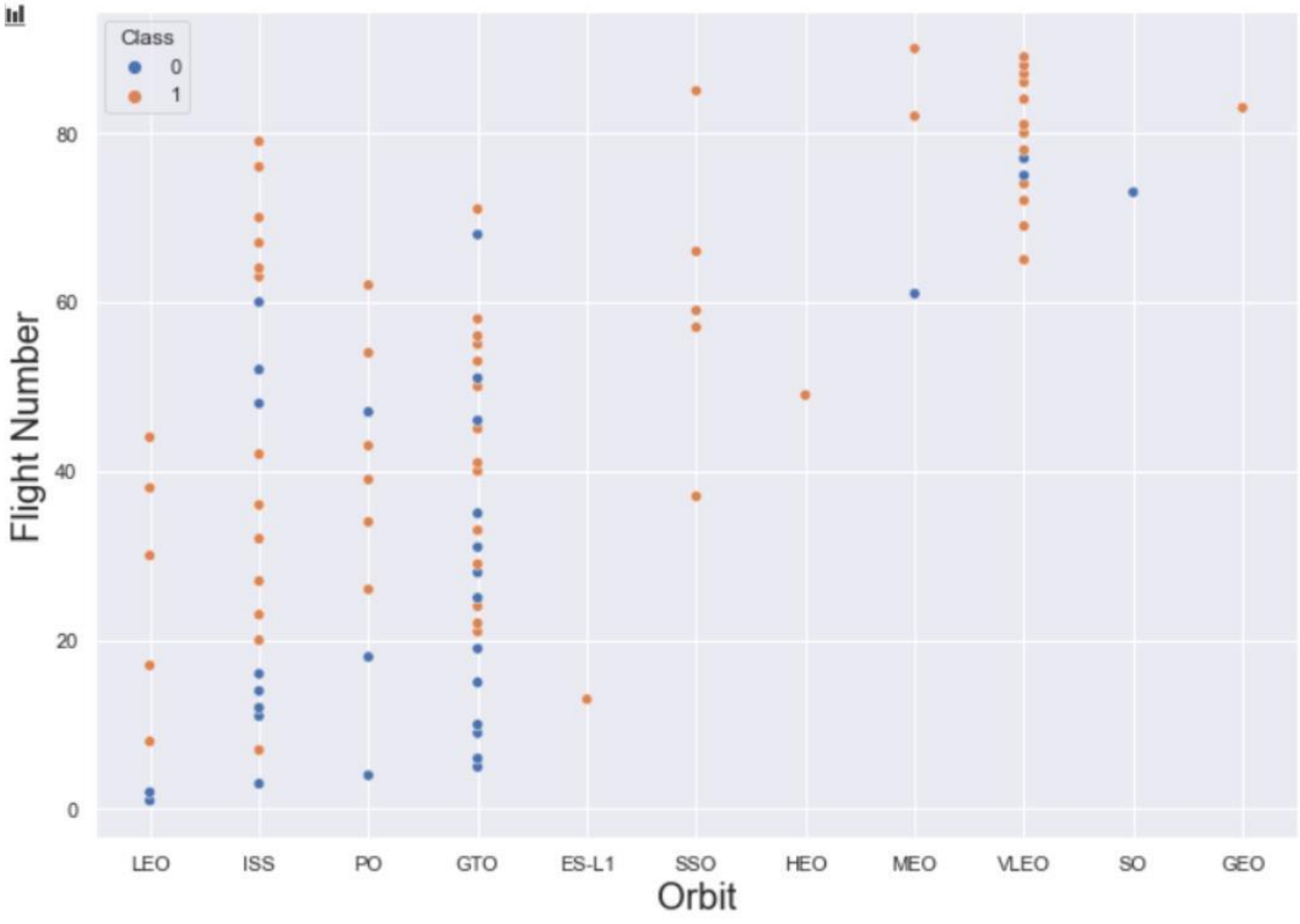
Success rate vs. Orbit type

Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate

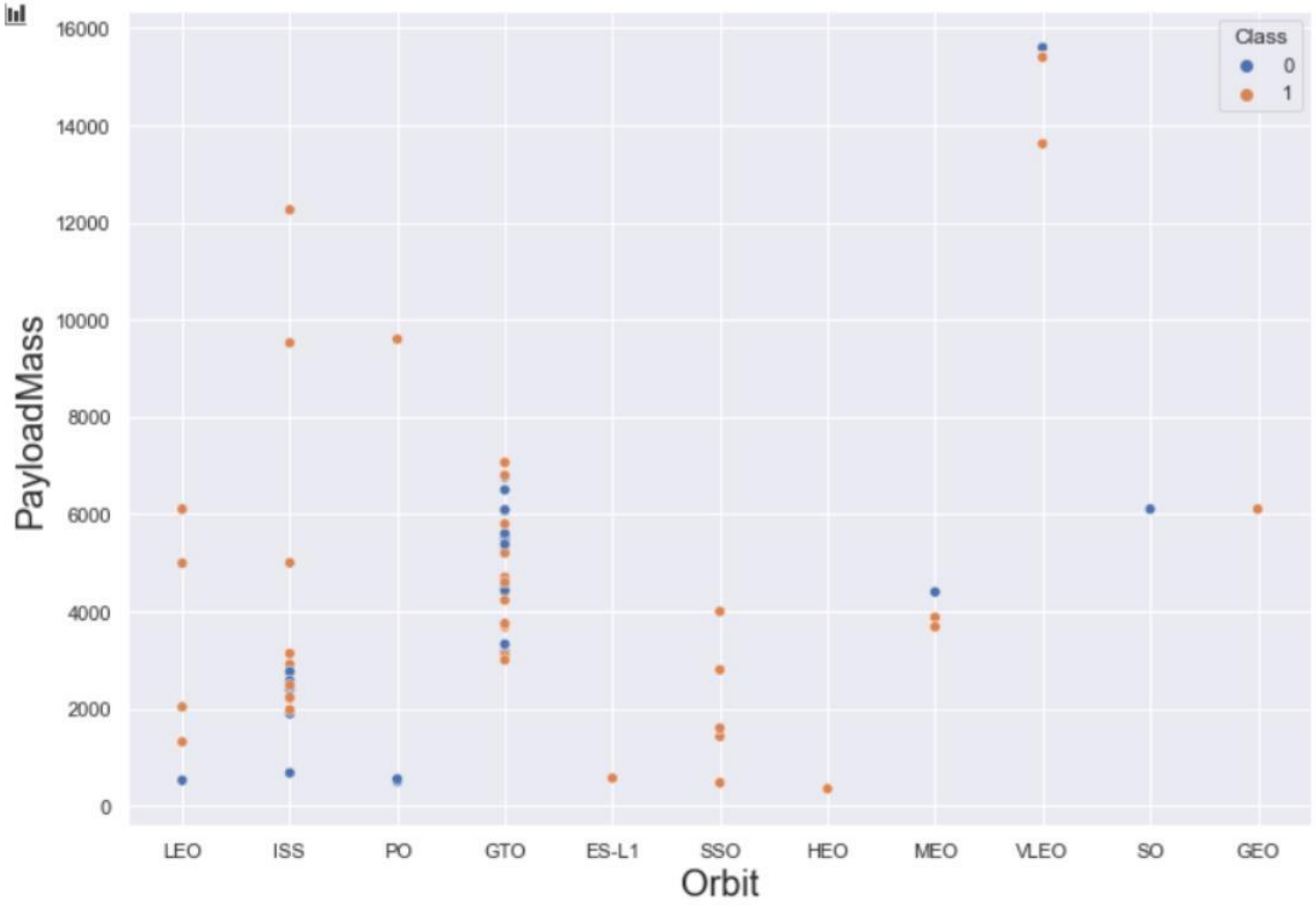


Flight Number vs. Orbit type

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.

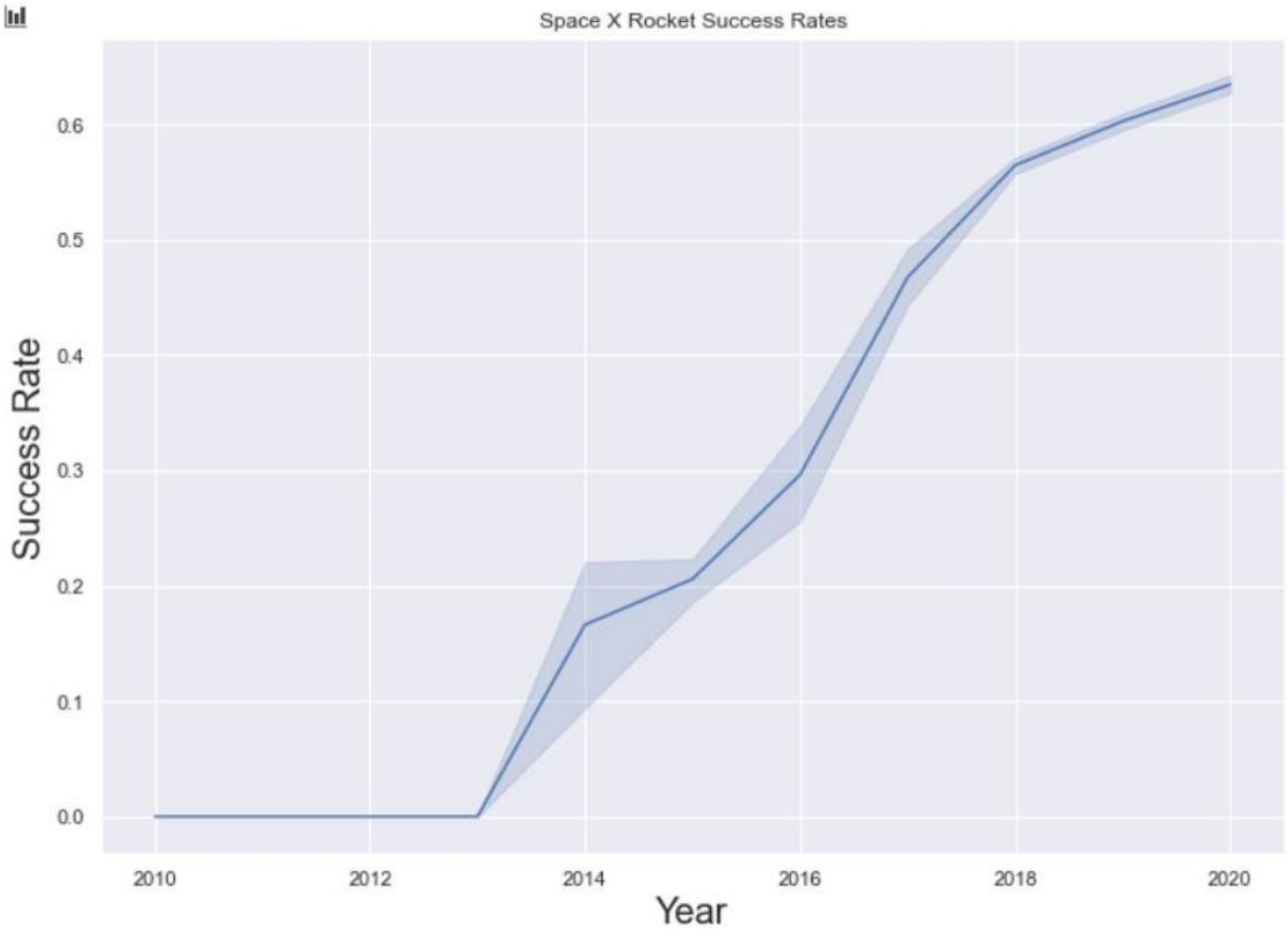


Payload vs. Orbit type



You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch success yearly trend



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

Unique Launch Site

SQL QUERY

SELECT DISTINCT *Launch_Site* from *spacex*

Out[6]:

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

QUERY EXPLAINATION

Using the word ***DISTINCT*** in the query means

that it will only

show Unique values in the ***Launch_Site*** column

from ***spacex***

Launch Site Names Begin with 'CCA'

SQL Query

%sql select * from spacex WHERE Launch_Site LIKE 'KSC%' limit 5

Query Explanation

Using the word ***LIMIT 5*** in the query means that it will only show 5 records from ***spacex*** and ***LIKE*** keyword has a wild card with the words ***'KSC%'*** the percentage in the end suggests that the Launch_Site name must start with KSC.

ut[8]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass by Customer NASA (CRS)

SQL Query

%sql select sum(payload_mass__kg_) as totalpayloadmass from spacex where Customer = 'NASA (CRS)'

totalpayloadmass
22007

Query Explanation

Using the function ***count*** *give* the total number in the column ***PAYLOAD_MASS_KG_***

The ***WHERE*** clause filters the dataset to only perform calculations on ***Customer NASA (CRS)***

Average Payload Mass carried by booster version F9 v1.1

SQL Query

%sql select avg(payload_mass__kg_) from spacex where booster_version like 'F9 v1.1%'

[10]:

1
3226

Using the function ***avg*** works out the average in the column

PAYLOAD_MASS_KG_

The ***WHERE*** clause filters the dataset to only perform calculations on ***Booster_version F9 v1.1***

First Successful Ground Landing Date

SQL Query

%sql select min(date) from spacex where landing_outcome not like 'failure%'

1
2010-04-06

Query Epxlanation

Using the function ***MIN*** works out the minimum date in the column ***Date.***

The ***WHERE*** clause filters the dataset to only perform calculations on ***Landing_Outcome Success (drone ship)***

Successful Drone Ship Landing with Payload between 4000 and 6000

SQL Query

%sql select booster_version from spacex where mission_outcome='Success' and payload_mass__kg_>4000 and payload_mass__kg_ < 6000

booster_version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 FT B1020
F9 FT B1022
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1051.2
F9 B5B1062.1

Query Explanation

The ***WHERE*** clause filters the dataset to ***Landing_Outcome = Success (drone ship)***

The ***AND*** clause specifies additional filter conditions ***Payload_MASS_KG_*** > 4000 AND ***Payload_MASS_KG_ < 6000***

Boosters Carried Maximum Payload

SQL Query

%sql select distinct booster_version from spacex.

Query Explanation

Using the word ***DISTINCT*** in the query means that it will only
show Unique values in the ***Booster_Version*** column from
spacex

2015 Launch Records

SQL Query

%sql select landing__outcome, booster_version, launch_site from spacex where date like '2015%'

Query Explanation

A very simple query ***Select*** launch records using ***where*** clause ***like*** filter the year ***'2015%'***

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
No attempt	F9 v1.1 B1014	CCAFS LC-40

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

Magic SQL Query

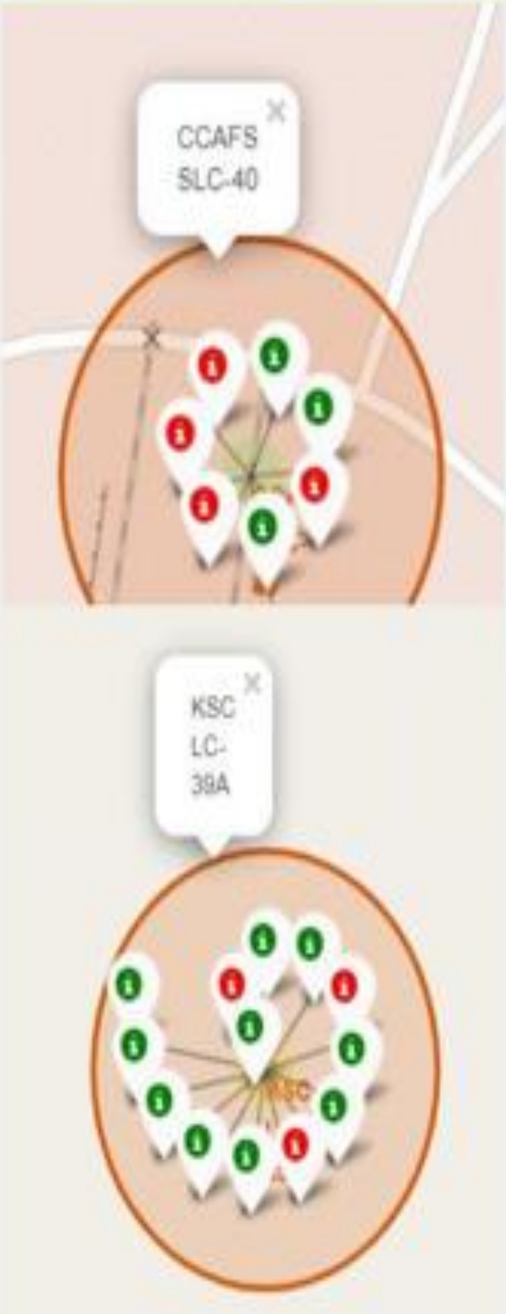
%sql select landing__outcome, count(landing__outcome) from spacex where landing__outcome='Failure (drone ship)' or landing__outcome='Success (ground pad)' and date between '2010-06-04' and '2017-03-20' group by landing__outcome

landing__outcome	2
Failure (drone ship)	2
Success (ground pad)	2

Section 4

Launch Sites Proximities Analysis

Interactive Map with Folium



California Launch Site

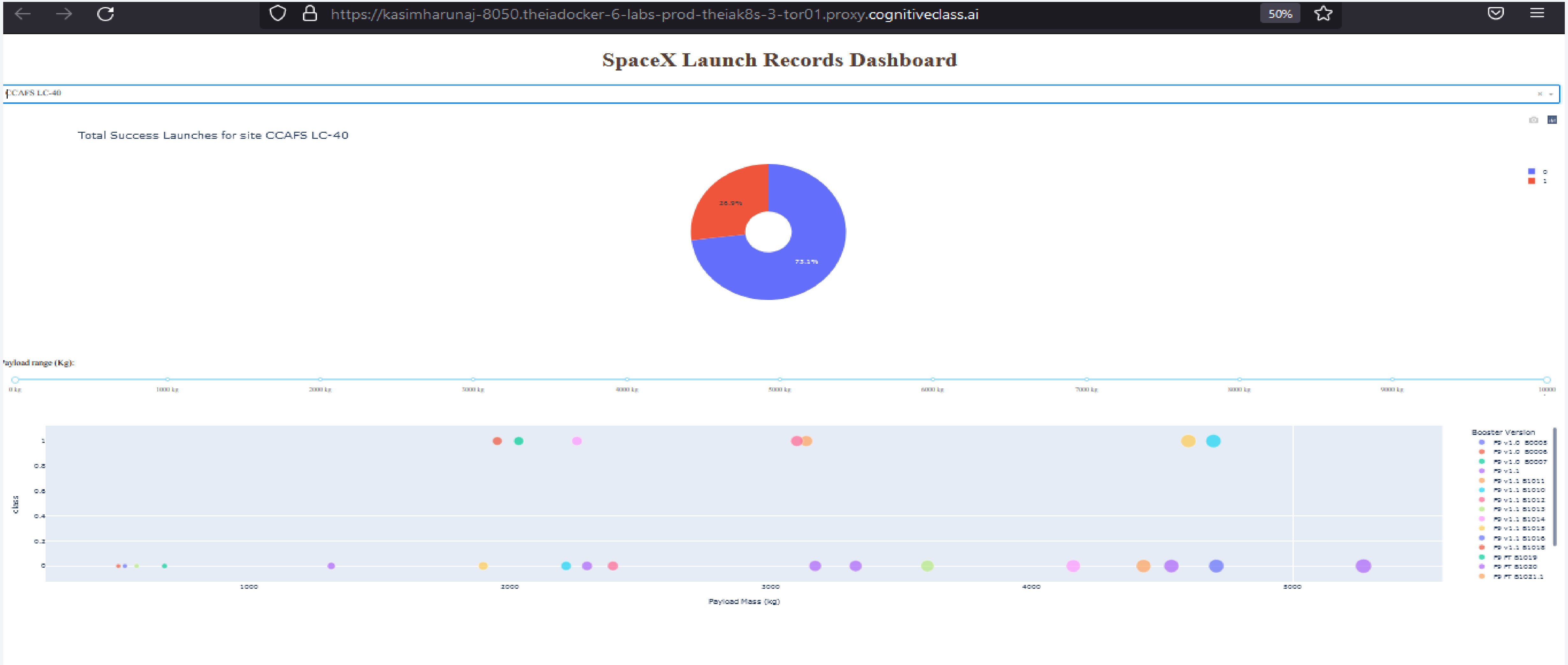
Florida Launch Sites

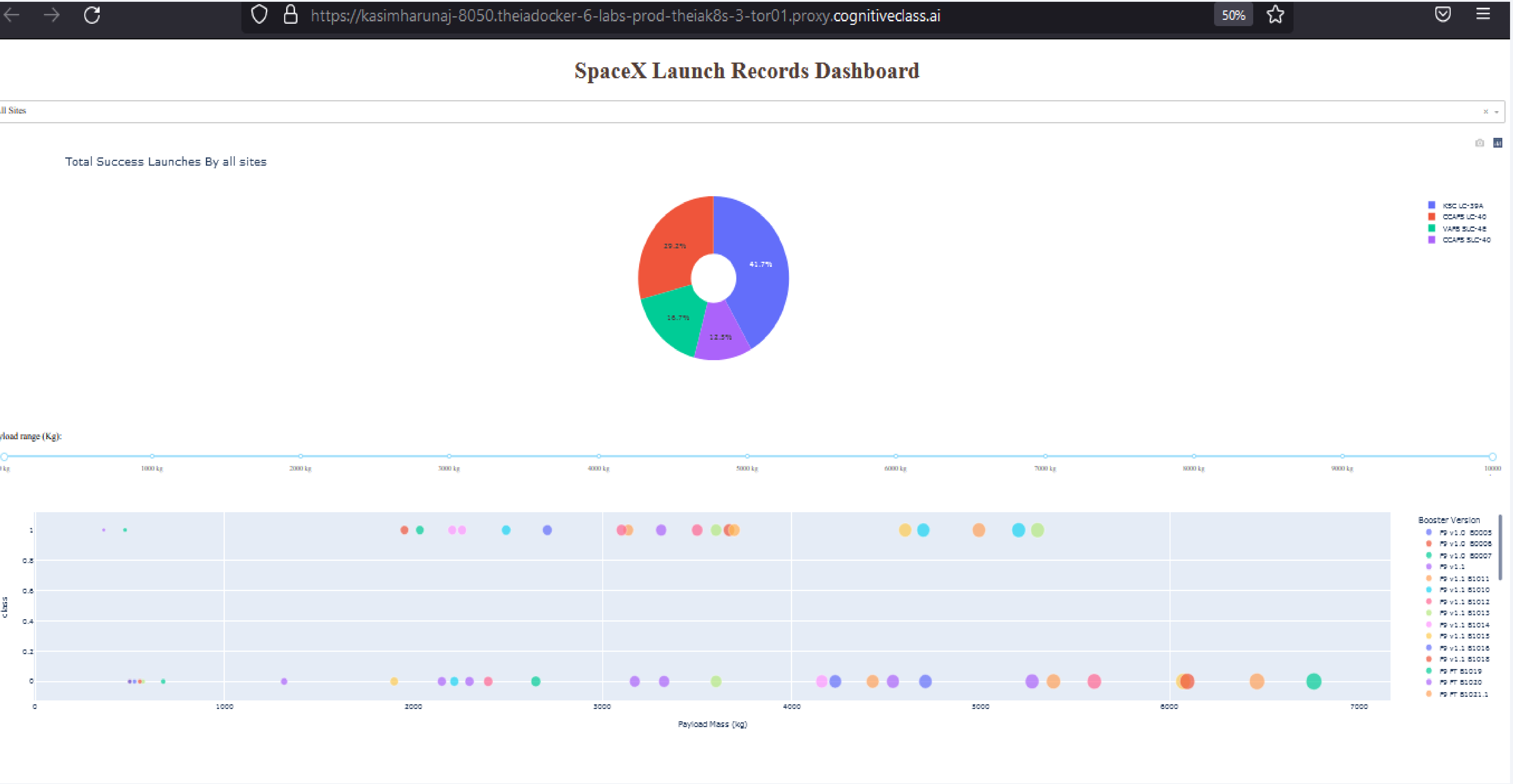
Green Marker shows successful Launches and Red Marker shows Failures

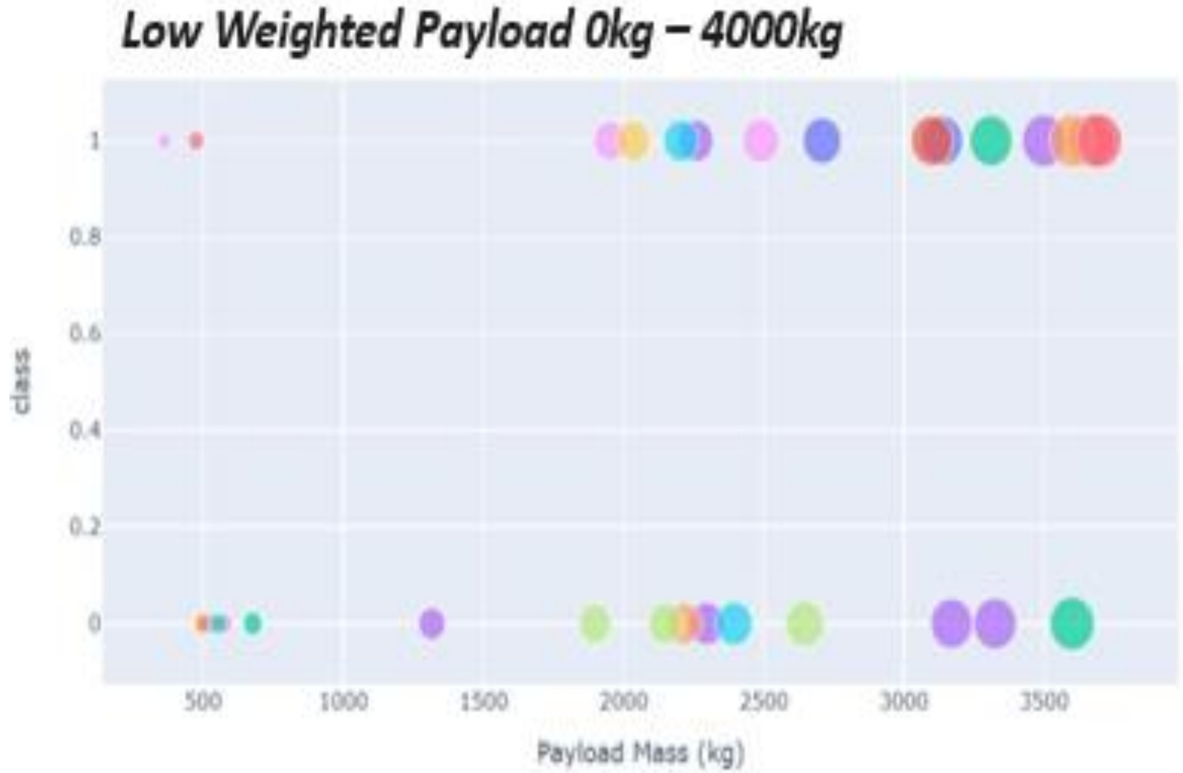
Section 5

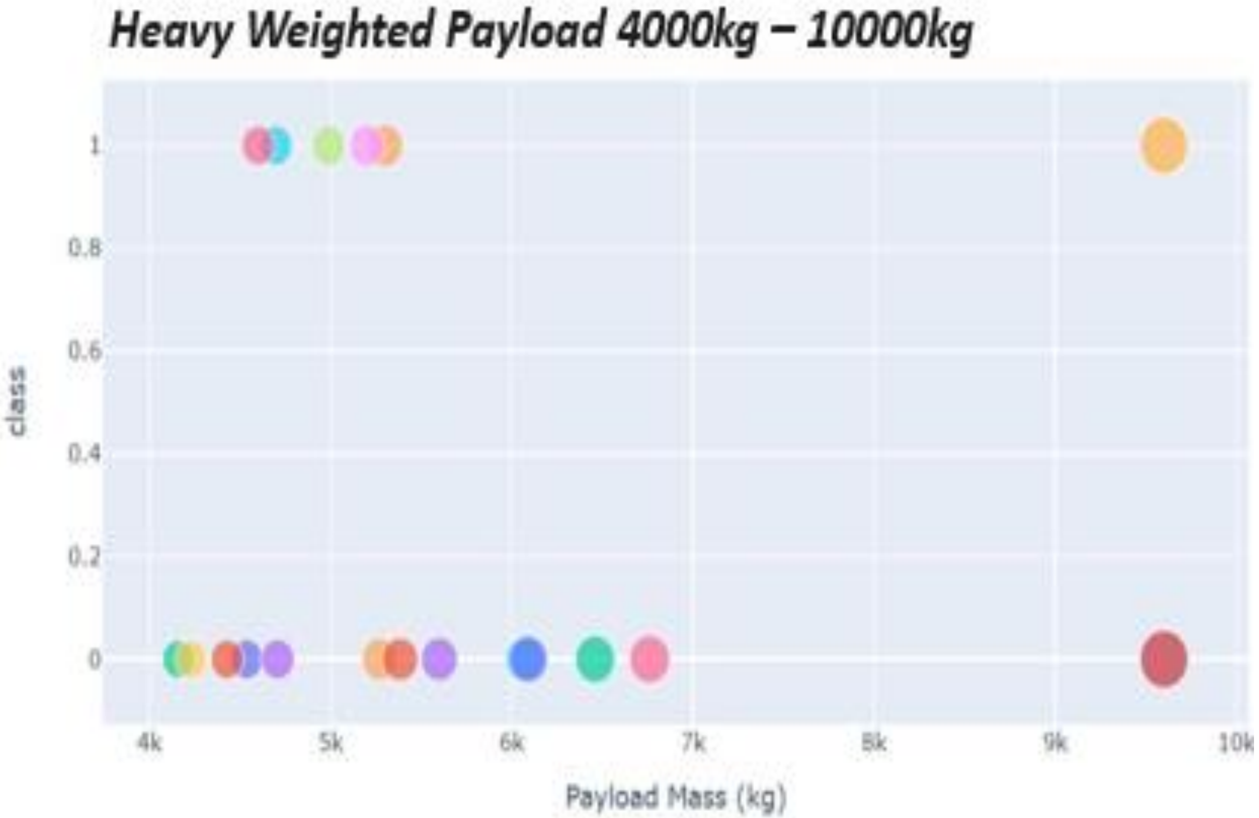
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard




https://kasimharunaj-8050.theiadocker-6-labs-prod-theiak8s-3-tor01.proxy.cognitiveclass.ai
50%
SpaceX Launch Records Dashboard
ll Sites
Total Success Launches By all sites
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
yload range (Kg):
1000 kg
2000 kg
3000 kg
4000 kg
5000 kg
6000 kg
7000 kg
8000 kg
9000 kg
10000
class
Payload Mass (kg)
Booster Version
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014
F9 v1.1 B1015
F9 v1.1 B1016
F9 v1.1 B1018
F9 FT B1019
F9 FT B1020
F9 FT B1021.1





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Conclusions

I was able to predicted that if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

Appendix

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

Thank you!