Winning Space Race with Data Science

AJAYI GEORGE OLANREWAJU

July 4th 2022

Outline

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

Executive Summary

- Collected the data using SpaceX API ,filtered the data to get only information about Falcon 9 launches, performed data wrangling, exploratory data analysis with SQL , pandas and matplotlib. Created interactive dashboards using plotly and dash , built predictive models using the sklearn module in python.

Introduction

- Project involves predicting 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 to find answers to

  Determining the price of each rocket launch, and also if Spacex will reuse the first stage.

.

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

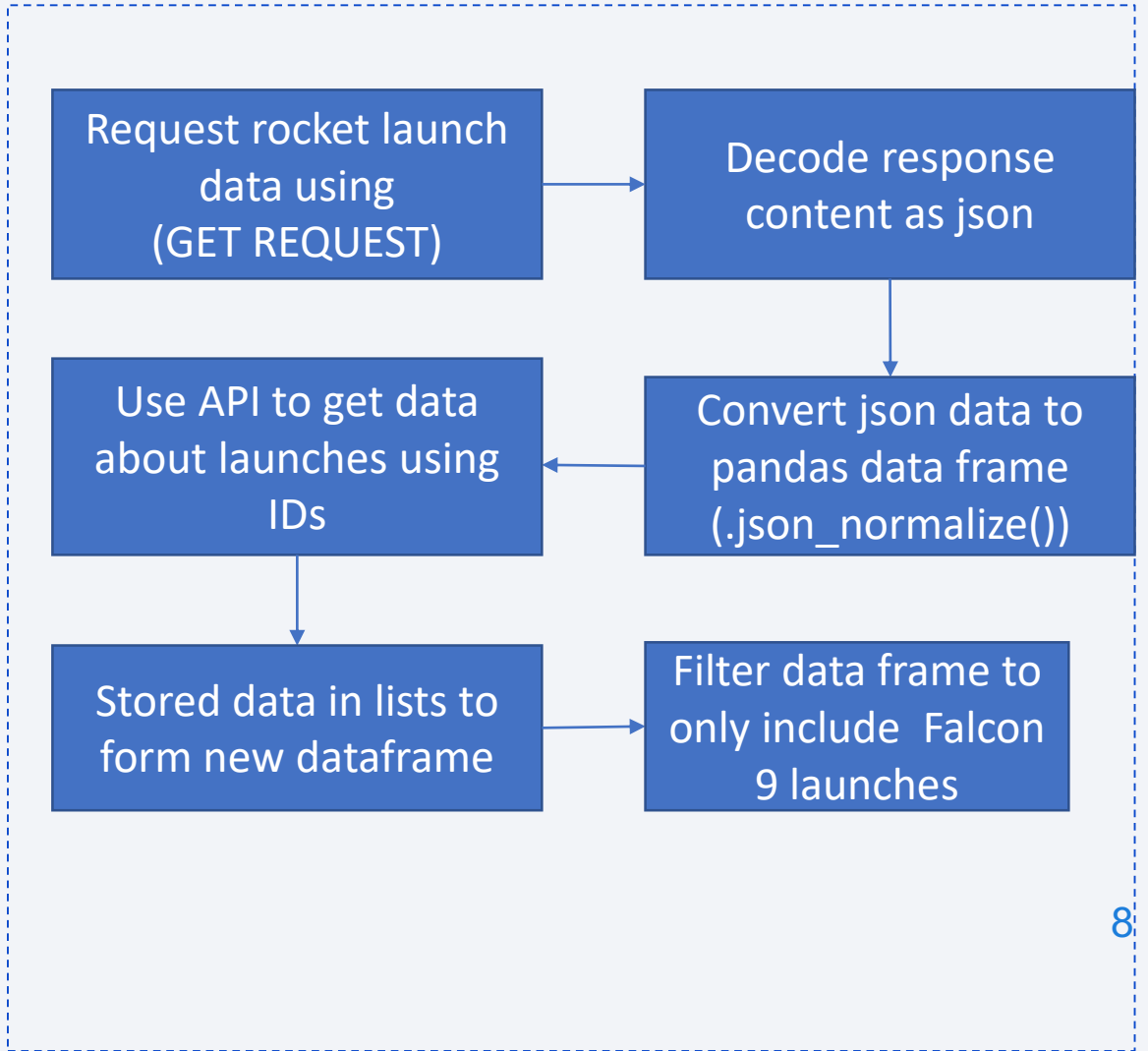
-Data was collected using a get request to the SpaceX API

-Data was collected by performing web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`

Data Collection -SpaceX API

- GitHub URL of the completed SpaceX API calls notebook:

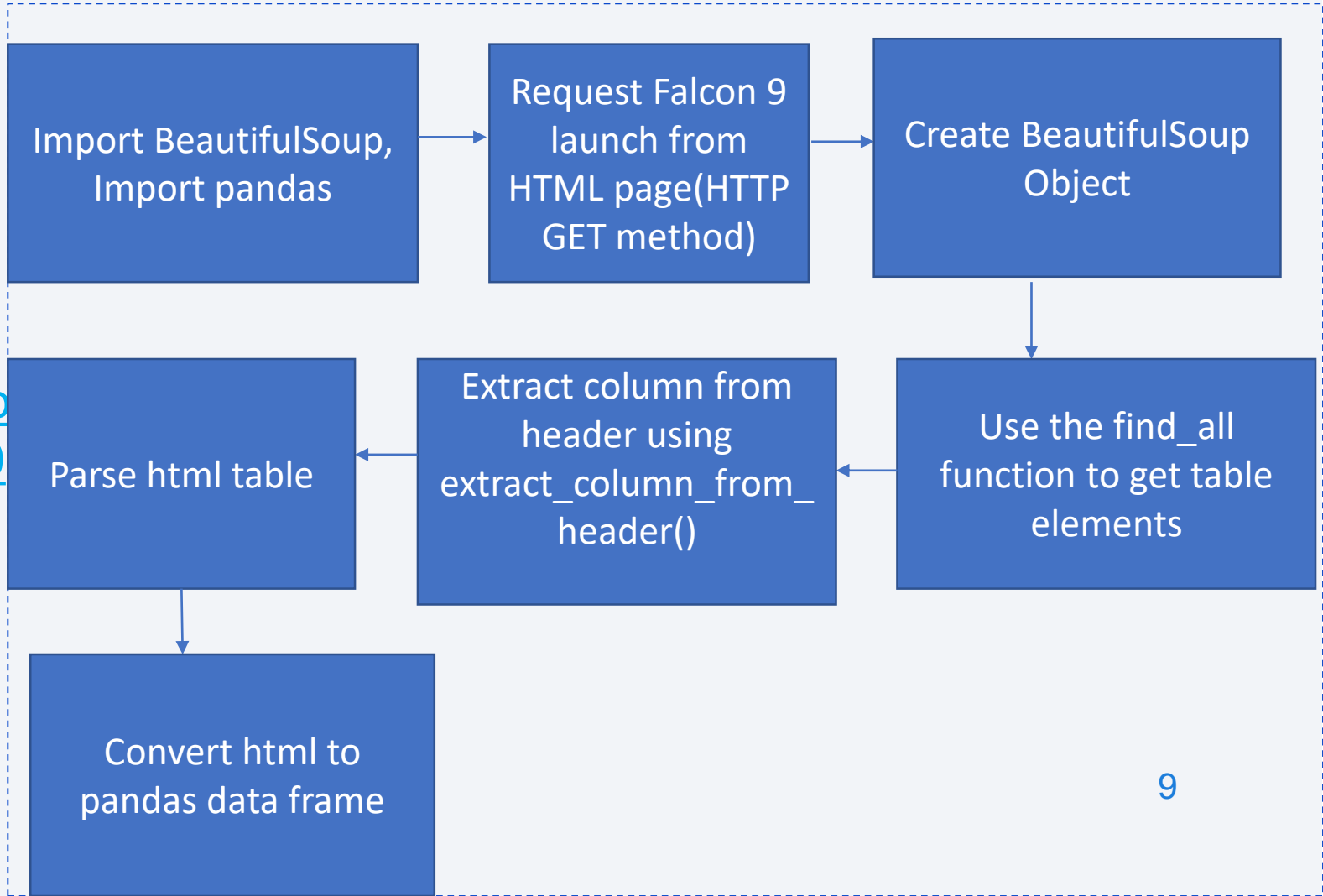
(https://github.com/george-ajayi/datascience_projects/blob/199f77e693d91877900f85ad470bb809d79cf7c3/jupyter-labs-spacex-data-collection-api%20(1).ipynb)



Data Collection - WebScraping

- GitHub URL of the completed web scraping notebook:

https://github.com/george-ajayi/datascience_projects/blob/b8a1de8bf77679142724cc09d3d95a8f60a7ef0d/jupyter-labs-webscraping.ipynb



Data Wrangling

Performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the labels for training supervised models.

- GitHub URL of completed data wrangling related notebooks:

https://github.com/george-ajayi/datascience_projects/blob/5b8a8aad60e9312facf0f49afae029aeff16ce91/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- GitHub URL of the completed EDA with data visualization notebook:

https://github.com/george-ajayi/datascience_projects/blob/97403e90e376647d38ec3c625bed780f559e5814/jupyter-labs-eda-dataviz.ipynb

Used the seaborn and matplotlib library to plot scatter plots to visualize the relationship between different variables in the data and bar charts to visualize Orbits with high success rates.

EDA with SQL

- SQL queries performed

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

- Display the total payload mass carried by boosters launched by NASA (CRS).

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.

- Display average payload mass carried by booster version F9 v1.1

- List the date when the first successful landing outcome in ground pad was achieved.

- List the total number of successful and failure mission outcomes

- List the names of the booster versions which have carried the maximum payload mass

- GitHub URL of the completed EDA with SQL notebook:

https://github.com/george-ajayi/datascience_projects/blob/97403e90e376647d38ec3c625bed780f559e5814/jupyter-labs-eda-dataviz.ipynb

Build an Interactive Map with Folium

- Used folium markers and folium circle objects

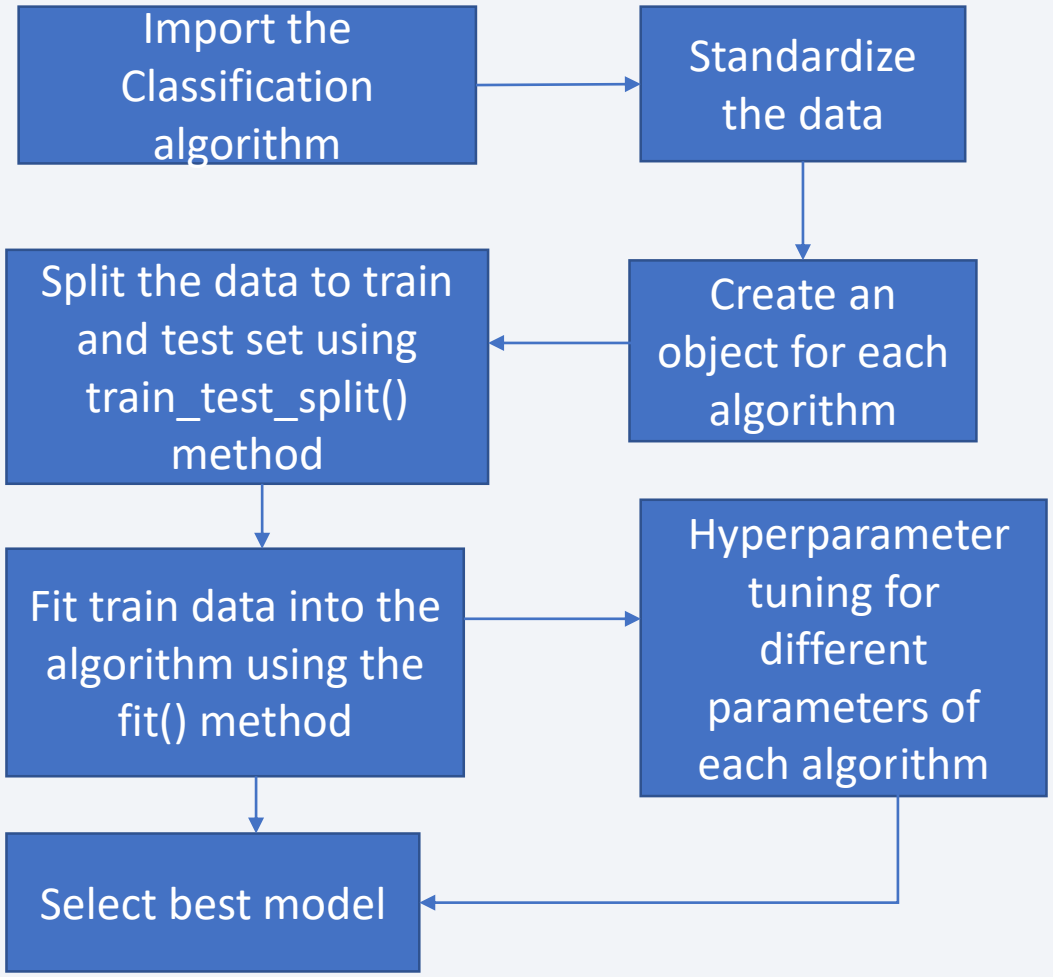
- Used folium circle to add a highlighted circle area with a text label on a specific coordinate.
- Used folium marker for each launch site on the site map
- GitHub URL of completed interactive map with Folium map: https://github.com/george-ajayi/proj/blob/7caf285e0ae41d092c842732de5d36d3aa9d9d86/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dashboard consists of a pie chart and scatter plot, a dropdown menu option and a slider.
- Pie chart shows the total successful launches count for all sites, scatter plot to show the correlation between payload and launch success.The dropdown menu to select the sites and slider to select the payload range.
- GitHub URL of the completed Plotly Dash lab:https://github.com/george-ajayi/datascience_projects/blob/54cab580cfbbf102522650162ca386825ff15633/interactive%20dashboard.ipynb

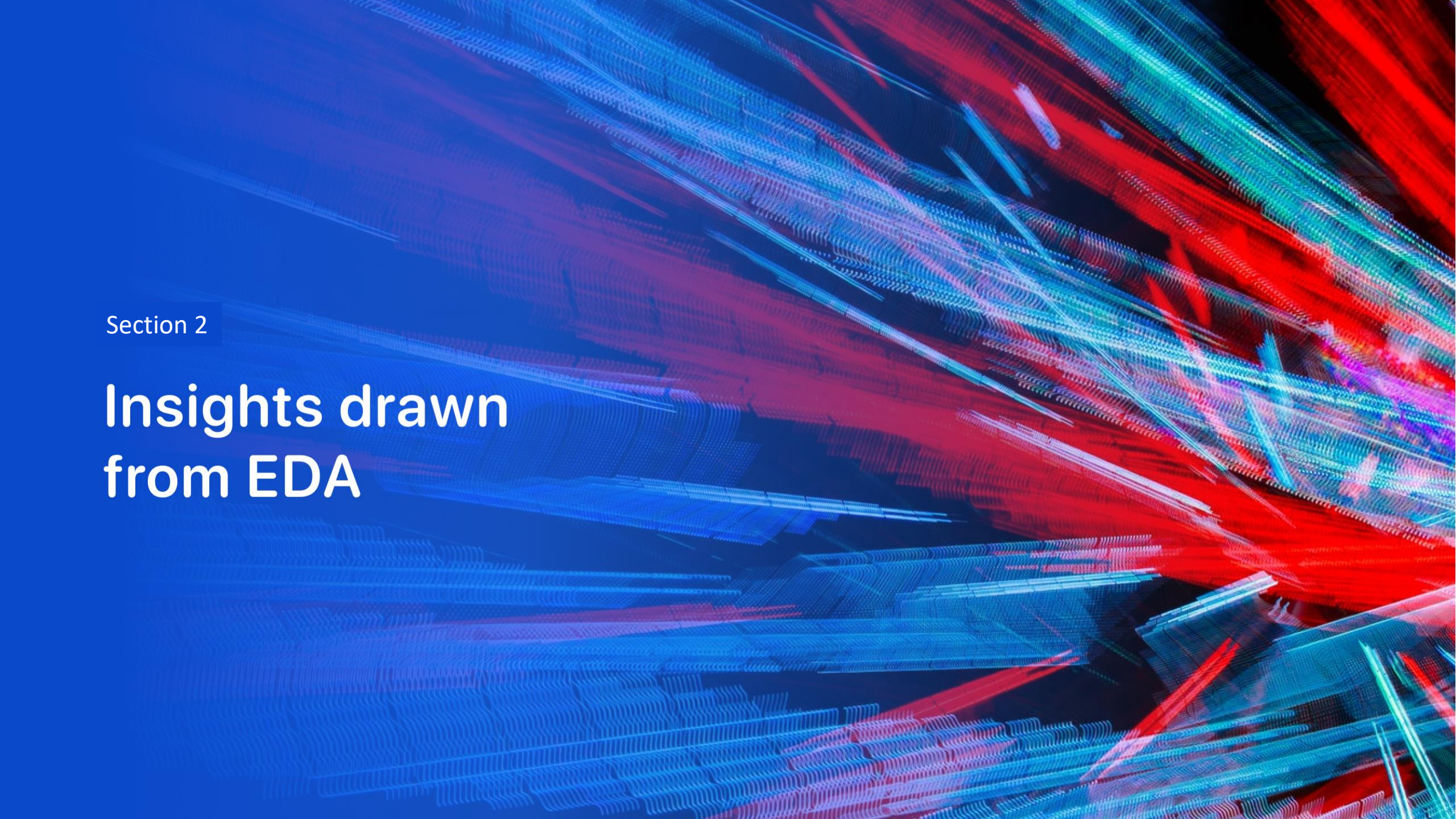
Predictive Analysis (Classification)

- Imported the Decision tree, Logistic regression, K Nearest Neighbors, Support Vector Machine algorithm from sklearn module.
- Standardized the data
- Passed the data to each algorithm
- Split the data to training and testing sets.
- Performed hyper parameter tuning on each of the classification models using GridSearchCV.
- Selected the model with highest accuracy.
- GitHub URL of completed predictive analysis lab: https://github.com/george-ajayi/proj/blob/f3cc3c409253791a3f09cb3bfc314076ba2fdb19/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



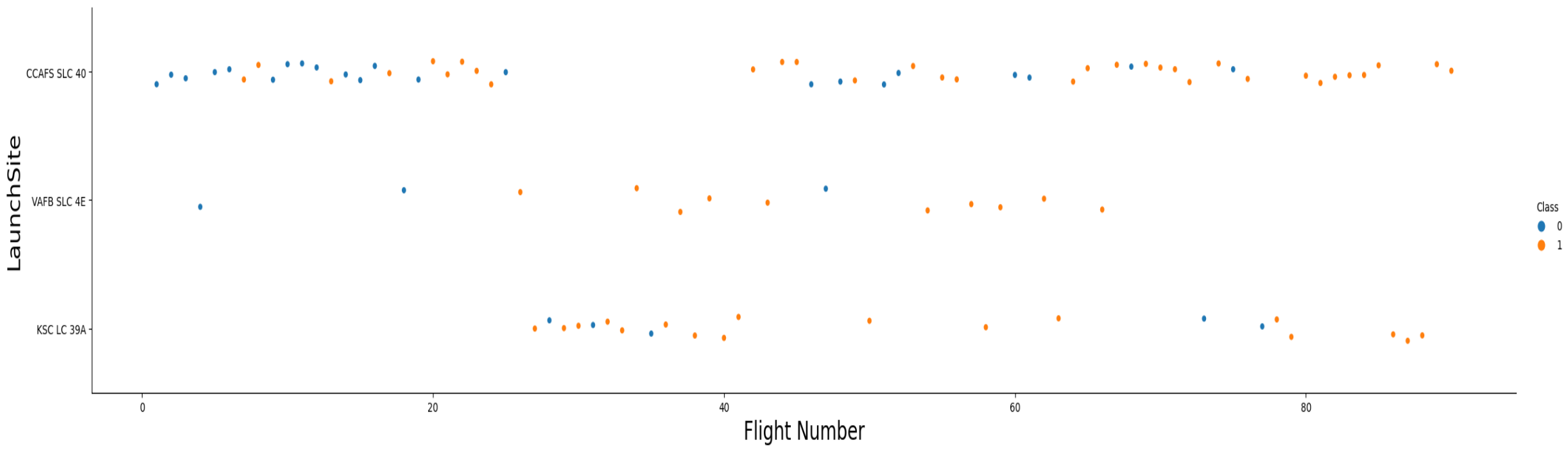
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site



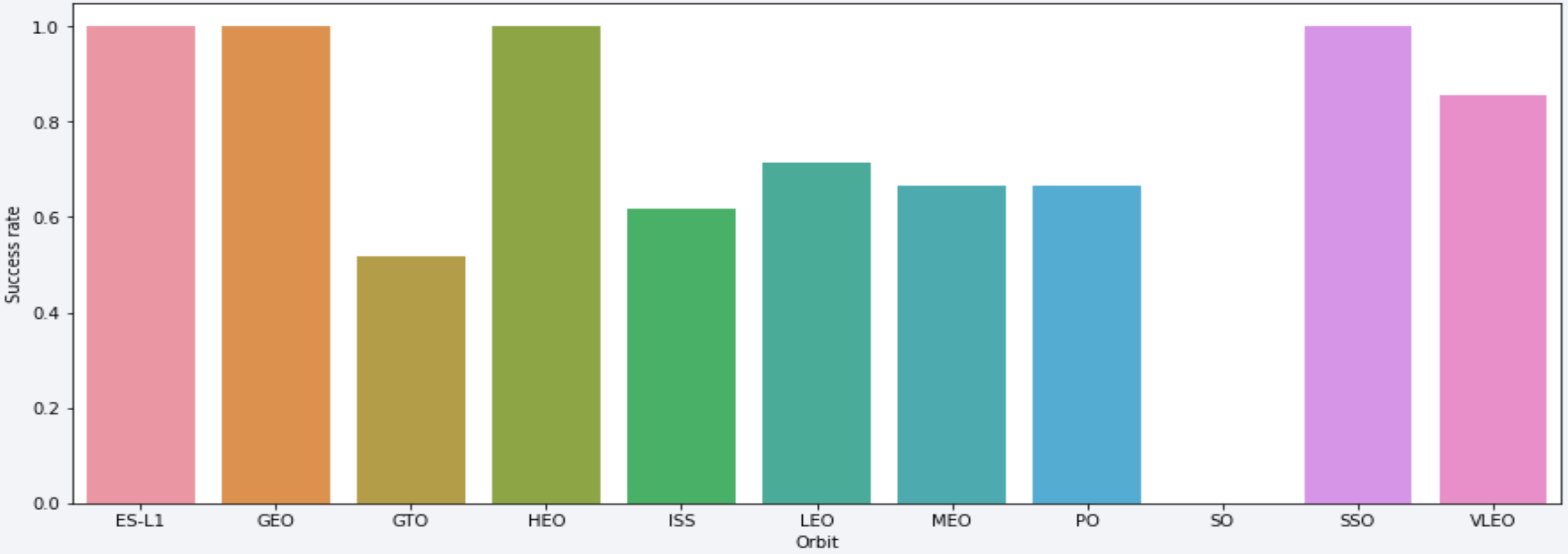
Payload vs. Launch Site

- Scatter plot of 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

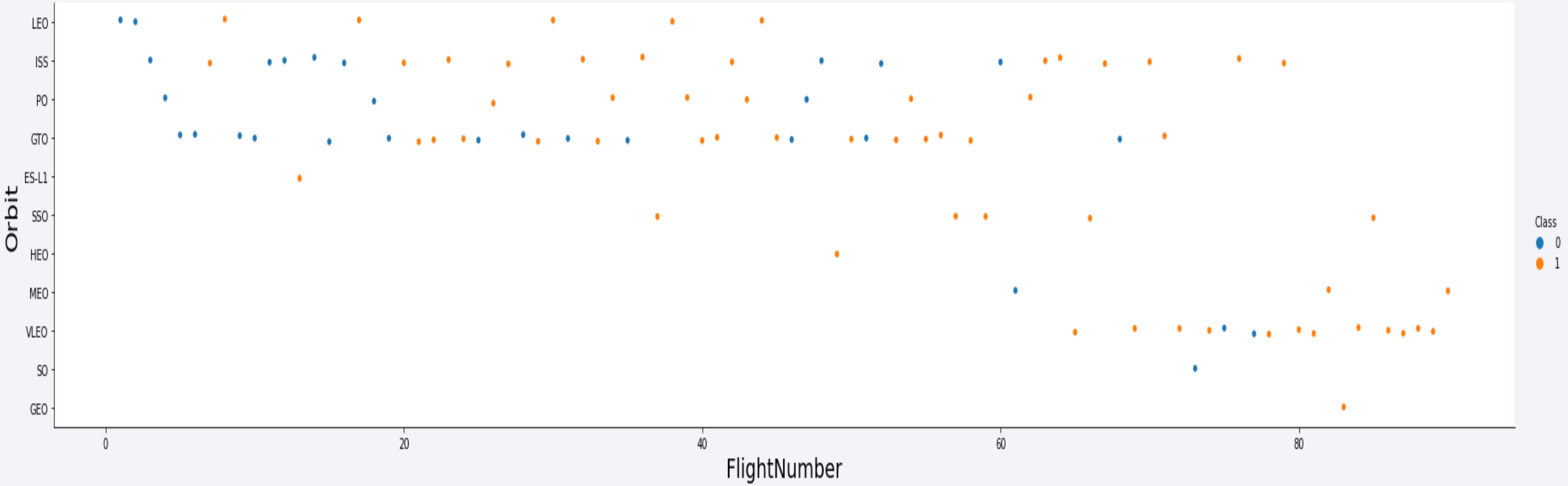
- Bar chart for the success rate of each orbit type



From the bar chart, the orbits with the highest success rate are ES-L1, GEO, HEO, SSO

Flight Number vs. Orbit Type

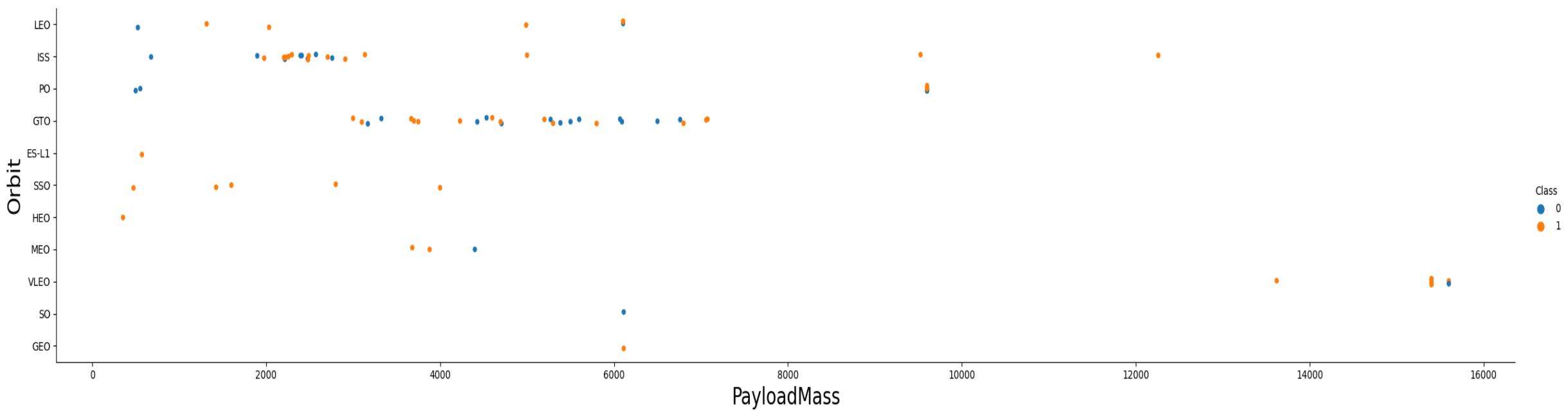
- Scatter point of Flight number vs. Orbit type



- 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

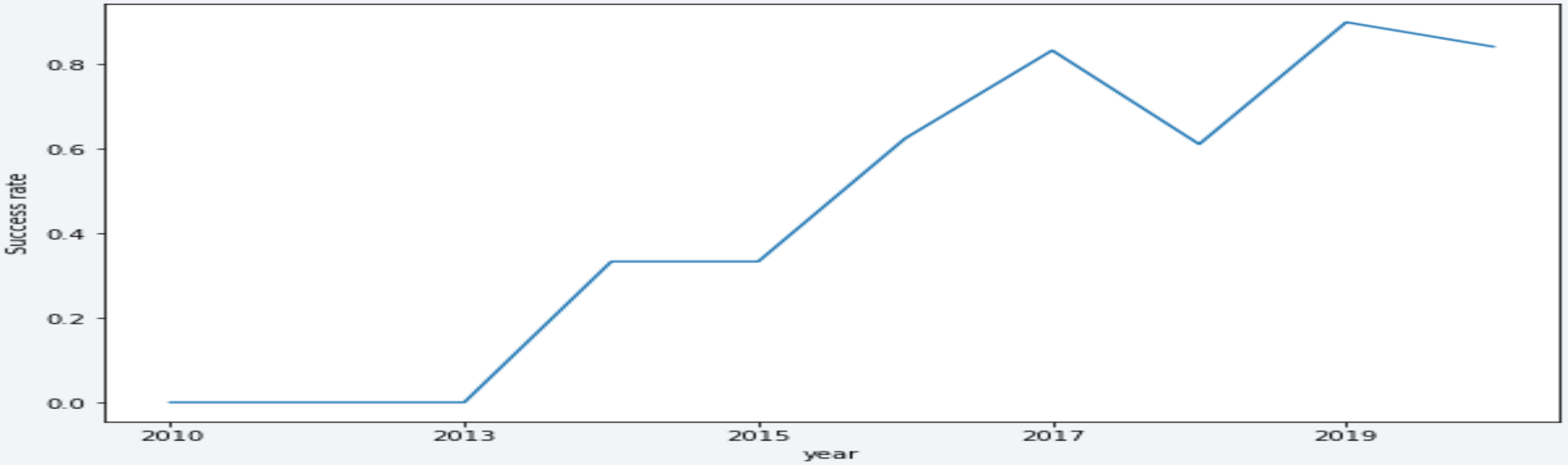
- Scatter point of payload vs. orbit type



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

- Show a line chart of yearly average success rate



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

All Launch Site Names

- Find the names of the unique launch sites

	LaunchSite
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
sql="SELECT * FROM spacextbl WHERE LaunchSite LIKE '%CCA' LIMIT 5"
mycursor.execute(sql)
result=mycursor.fetchall()
result
```

[5] ✓ 0.2s Python

	Date	Time (UTC)	BoosterVersion	LaunchSite	Payload	PAYLOADMASSKG	Orbit	Customer	MissionOutcome	LandingOutcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

- Calculate the total payload carried by boosters from NASA

```
sql3="SELECT SUM(PAYLOADMASSKG) AS TOTAL FROM spacextbl WHERE Customer='NASA (CRS)'"
df1=pd.read_sql_query(sql3,con)
df1
```

Python

	TOTAL
0	45596.0

- Total pay load carried by boosters from NASA is 45596kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
sql4="SELECT AVG(PAYLOADMASSKG) AS Average_payload_mass FROM spacextbl WHERE BoosterVersion='F9 v1.1'"
df2=pd.read_sql_query(sql4,con)
df2
```

Python

	Average_payload_mass
0	2928.4

- the average payload mass carried by booster version F9 v1.1 is 2928kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
sql5="SELECT DATE FROM spacextbl WHERE LandingOutcome ='Success (ground pad)' "
df3=pd.read_sql_query(sql5,con)
df3
```

Python

	DATE
0	22-12-2015
1	18-07-2016
2	19-02-2017
3	01-05-2017
4	03-06-2017
5	14-08-2017
6	07-09-2017
7	15-12-2017
8	08-01-2018

the dates of the first successful landing outcome on ground pad was from the year 2015-2018.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
sql6="""SELECT BoosterVersion FROM spacextbl WHERE LandingOutcome ='Success (drone ship)'
        AND PAYLOADMASSKG >4000 AND PAYLOADMASSKG <6000 """
df4=pd.read_sql_query(sql6,con)
df4
```

Python

	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

- Calculate the total number of successful and failure mission outcomes

```
sql7="SELECT MissionOutcome, COUNT(DISTINCT MissionOutcome)AS Total FROM spacextbl GROUP BY MissionOutcome  "
df5=pd.read_sql_query(sqll,con)
df5
```

Python

	MissionOutcome	Total
0	Failure (in flight)	1
1	Success	99
2	Success (payload status unclear)	1

Jupyter Server: local Cell 16 of 37

- the total number of successful mission outcomes are 100 and failure mission outcome is 1.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
sql8="SELECT BoosterVersion FROM spacextbl WHERE PAYLOADMASSKG =(SELECT MAX(PAYLOADMASSKG) FROM spacextbl) "
df6=pd.read_sql_query(sql8,con)
df6
```

Python

	BoosterVersion
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

Jupyter Server: local Cell 27 of 37

booster which have carried the maximum payload mass are the F9 B5 B versions

2015 Launch Records

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

	Month_name	Date	BoosterVersion	LandingOutcome	LaunchSite
13	Oct	2015-10-01	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
16	Apr	2015-04-14	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

- The failed landing outcomes in drone ship for the year 2015 were in the month of October and April and the site name is CCAFS LC-40

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

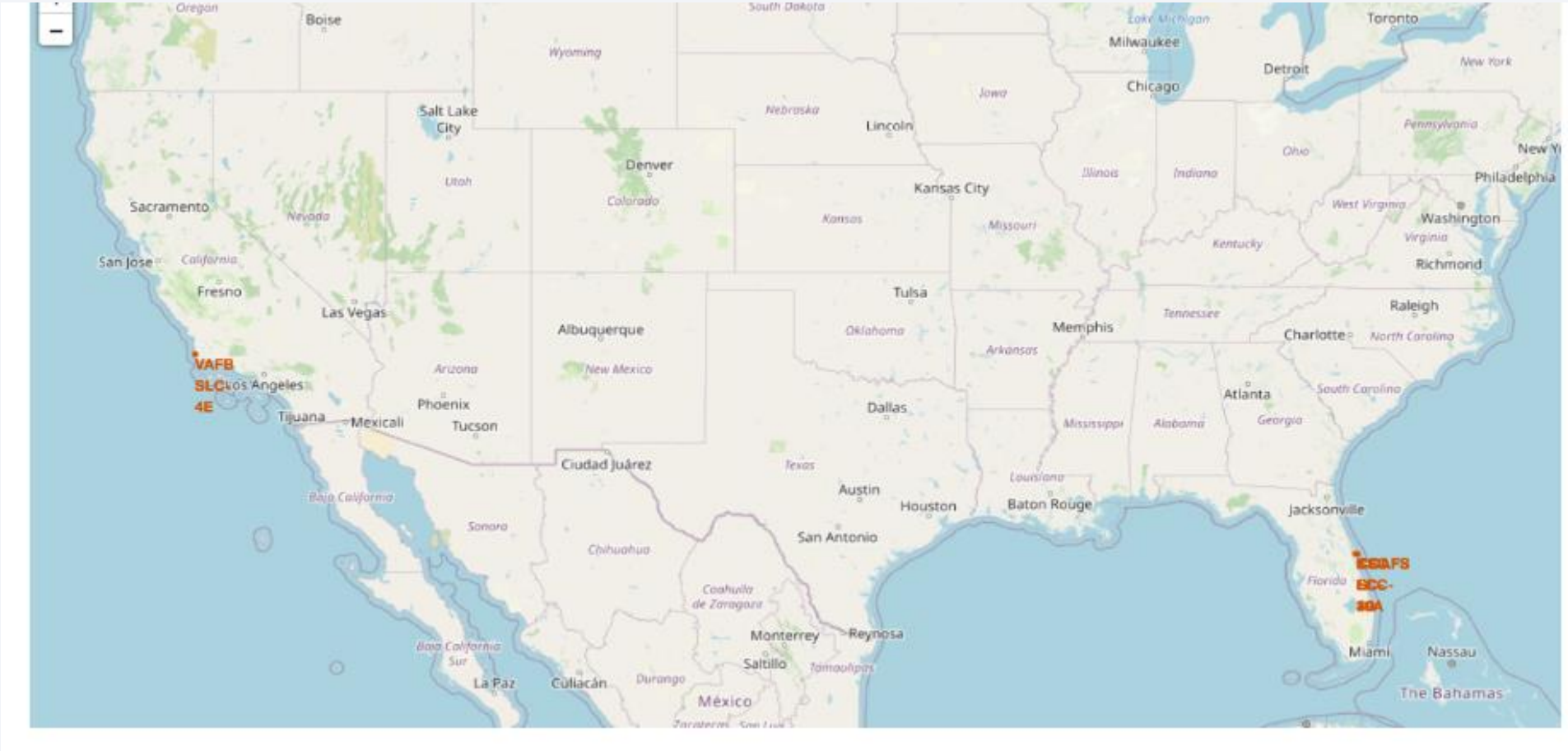
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

	LandingOutcome	COUNT
3	No attempt	10
1	Failure (drone ship)	5
5	Success (drone ship)	5
6	Success (ground pad)	5
0	Controlled (ocean)	3
2	Failure (parachute)	2
7	Uncontrolled (ocean)	2
4	Precluded (drone ship)	1

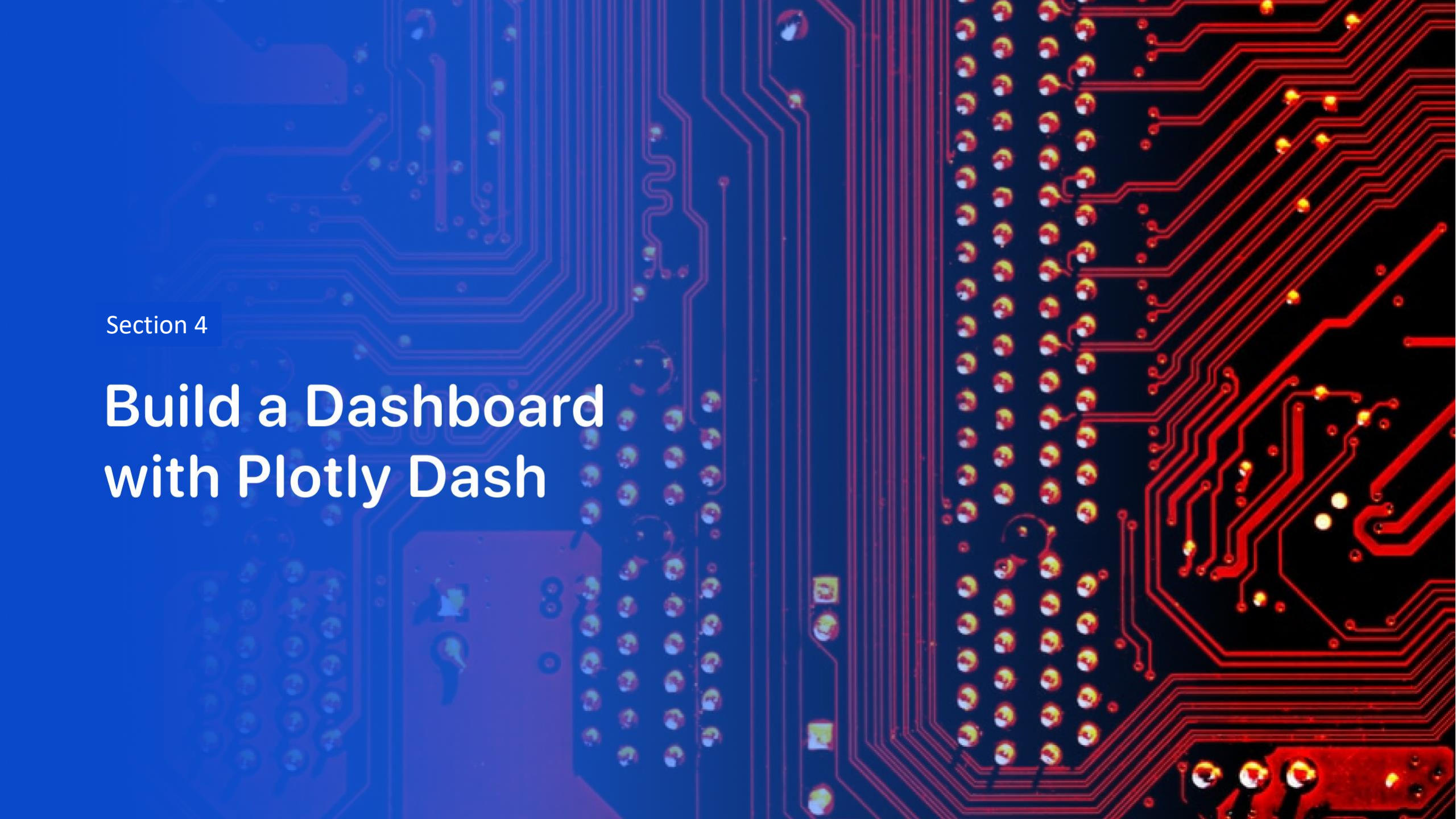
Section 3

Launch Sites Proximities Analysis

Folium Map - all sites location markers

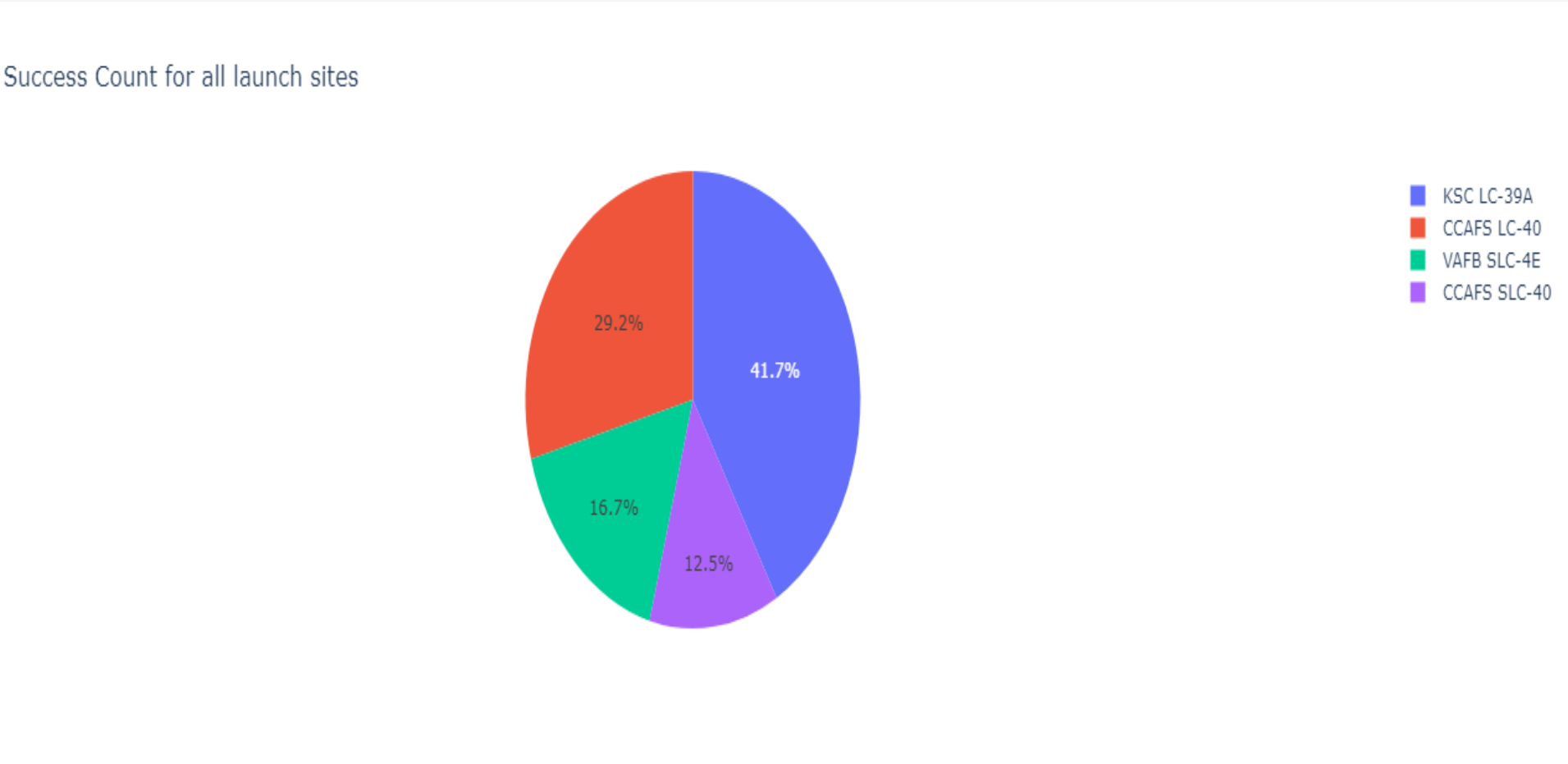


Folium Map -launch site proximities


Section 4
Build a Dashboard with Plotly Dash

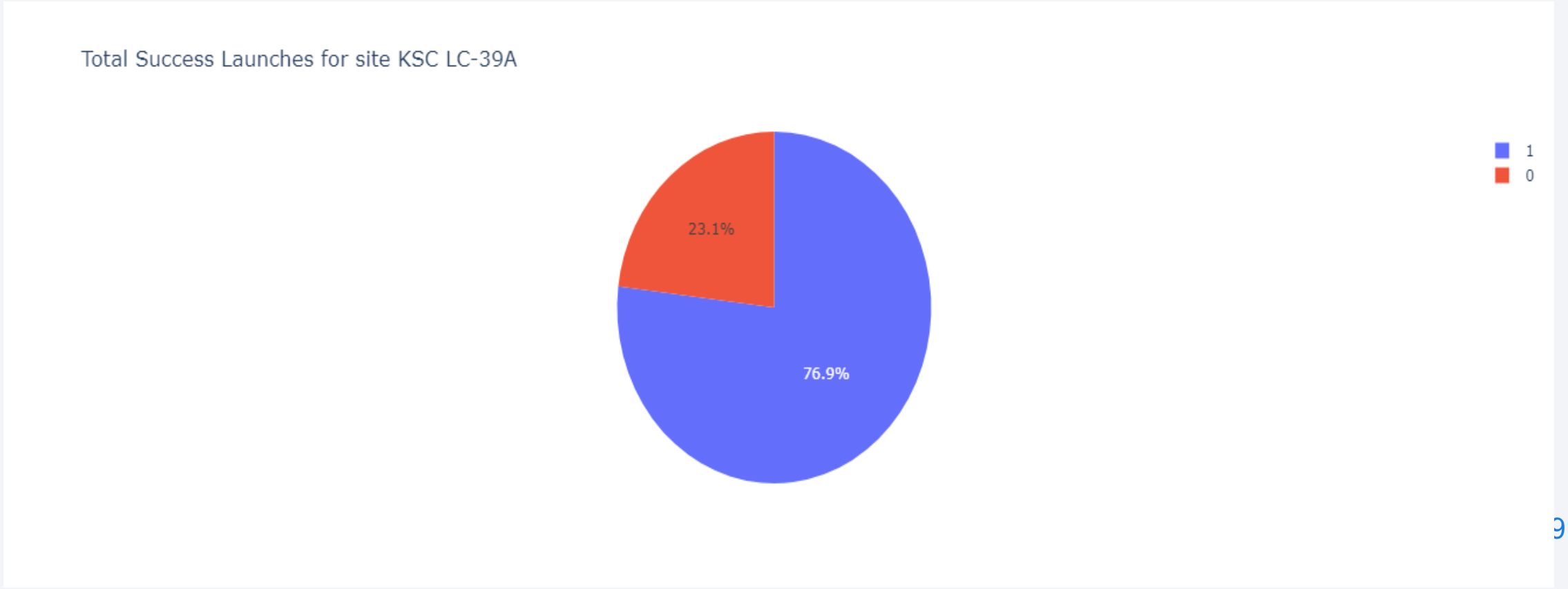
Dashboard(pie chart) Success count for all sites

- Show the screenshot of launch success count for all sites, in a piechart



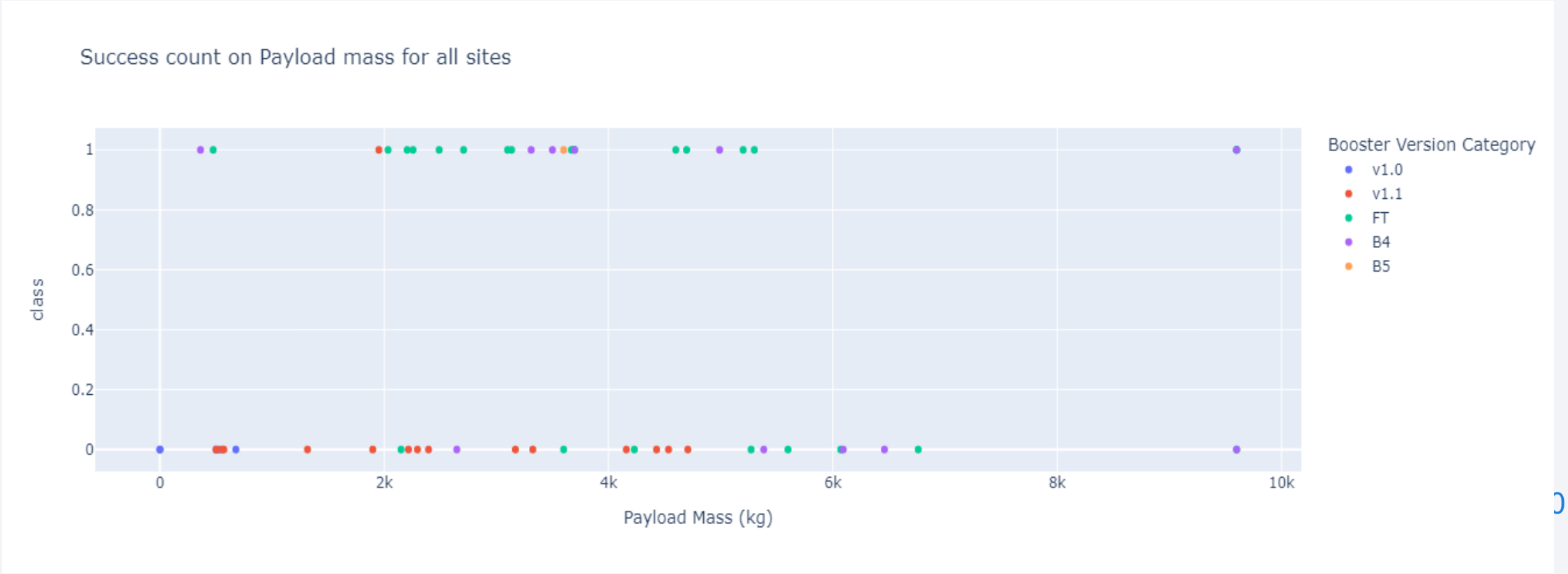
Dashboard(pie chart) highest launch Success ratio

- Show the screenshot of the pie chart for the launch site with highest launch success ratio



Dashboard (Scatter plot)

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

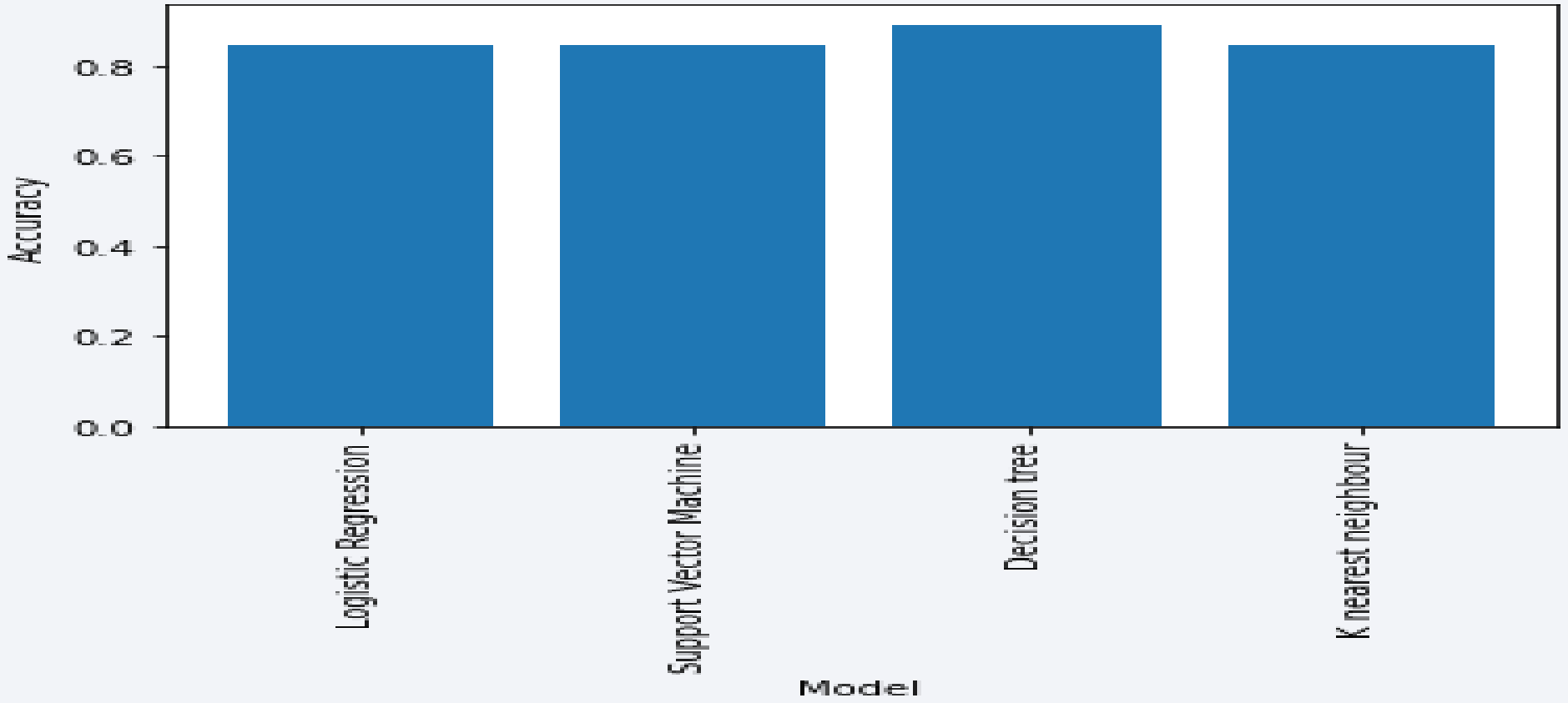


Section 5

Predictive Analysis (Classification)

Classification Accuracy

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



Model with highest accuracy is the Decision tree model

Confusion Matrix

- Confusion matrix (Decision tree)

- Examining the confusion matrix for decision tree we can see that for the positive labels the model classified 5 labels correctly and 1 wrongly , for the negative labels the model classified 10 labels correctly and 2 wrongly.

Conclusions

- The best algorithm for predicting the SpaceX Falcon 9 First stage is the Decision Tree algorithm with 89% accuracy.
- The launch site with the highest launch success ratio is the KSC LC-39A with success rate of 76.9% and failure rate of 23.1%.
- The total number of successful mission outcomes are 100 and failure mission outcome is 1.

Thank you!