IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Keerthi Vijay
27-12-2021

# Outline

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

# Executive Summary

- Data has been collected and cleaned by data wrangling and performed exploratory analysis using visualization and SQL. Next, performed interactive visual analytics using Folium and Plotly Dash. Finally predictive analysis using classification model has been performed
- Summary of all results

# Introduction

- Project background and context
  - 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, this project is done to determine if the first stage will land, then 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
  - Determine if the first stage will land, then can determine the cost of a launch.

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Using the SpaceX API, rocket launch data has been collected. Response content using the GET request has been decoded as Json.
- Perform data wrangling
  - Replaced the missing values with mean in the PayloadMass and calculated the number of launches on each site, number and occurrence of each orbit and mission outcome per orbit type.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split the dataset into training and testing data with test size 0.2. Created GridSeach for different classification models and calculated the accuracy.

# Data Collection

- Describe how data sets were collected.
    - Request and parse the SpaceX launch data using the GET request and created a dataframe. Some of the data still contained id. Next used APIs to get the real value for those ids. Finally, a dataframe is created with required data.
- You need to present your data collection process use key phrases and flowcharts

# Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose
  - https://github.com/kvi24/IBM-Cousera/blob/master/Final%20Course%20-%20Week%201.ipynb

Place your flowchart of SpaceX API calls here

# Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
  - https://github.com/kvi24/IBM-Cousera/blob/master/Data%20Collection%20with%20web%20scrapping.ipynb

Place your flowchart of web scraping here

# Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
  - https://github.com/kvi24/IBM-Cousera/blob/master/Final%20Course%20-%20Week%201.ipynb
  - https://github.com/kvi24/IBM-Cousera/blob/master/EDA.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
  - https://github.com/kvi24/IBM-Cousera/blob/master/EDA%20Week%202%20-%20Part2.ipynb

# EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://github.com/kvi24/IBM-Cousera/blob/master/EDA%20Week%202%20-%20Part2.ipynb

# 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
    - https://github.com/kvi24/IBM-Cousera/blob/master/Week%203%20-%20Interactive%20Visual%20Analytics%20with%20Folium%20.ipynb

# 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
  - https://github.com/kvi24/IBM-Cousera/blob/master/spacex_dash_app.py

# 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
  - https://github.com/kvi24/IBM-Cousera/blob/master/Machine%20Learning%20Prediction%20Lab.ipynb

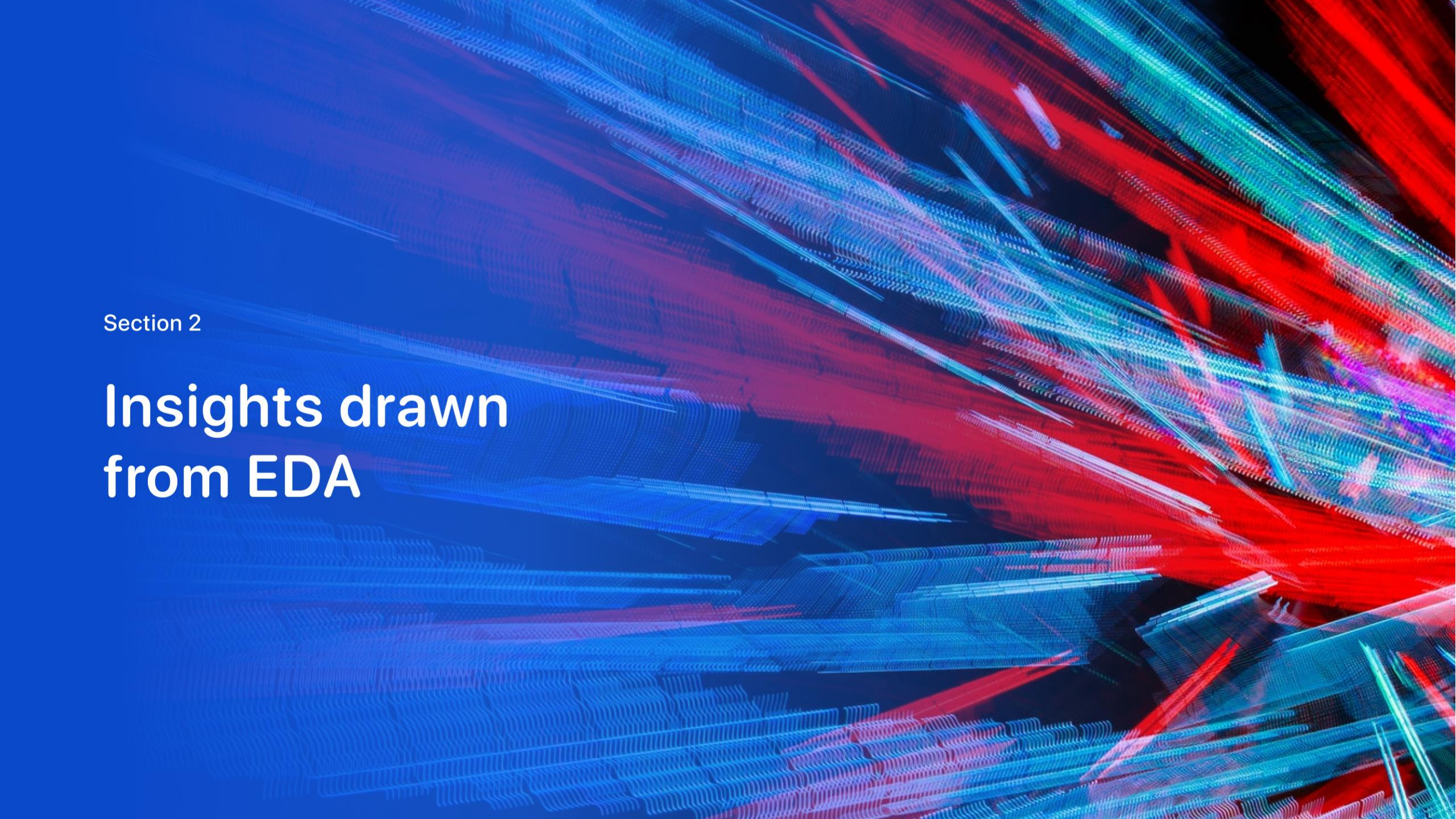
# Results

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

| Model | Accuracy | Accuracy using Score |
|---|---|---|
| Logistic Regression | 84.64% | 83.34% |
| SVM | 84.82% | 83.34% |
| Decision Tree | 87.68% | 88.89% |
| KNN | 84.82% | 83.34% |

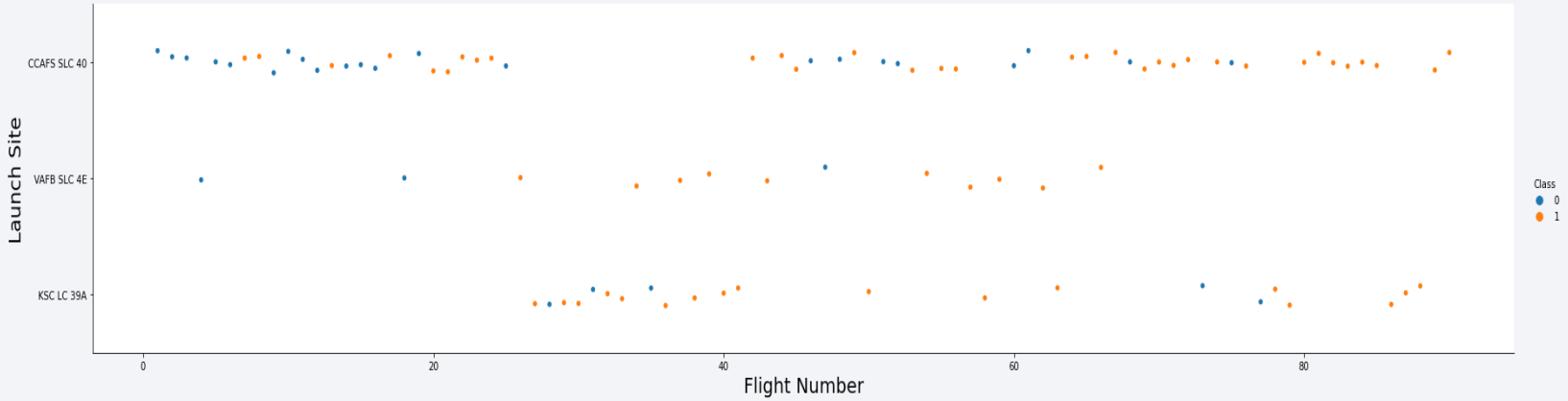
Section 2

# Insights drawn from EDA



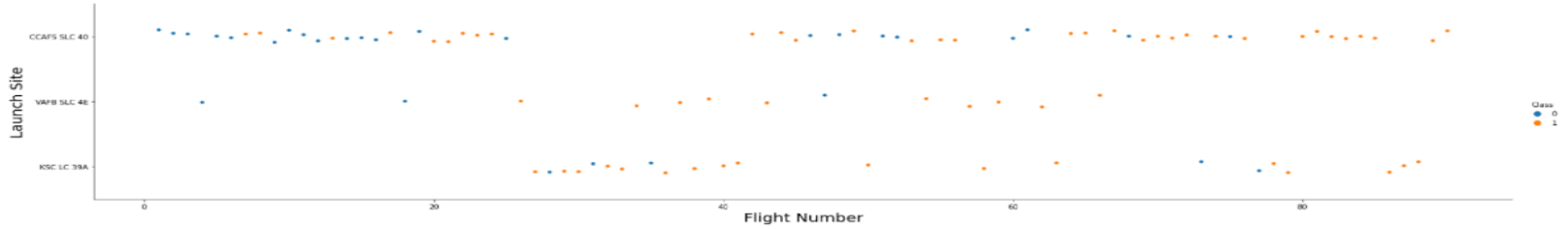
# Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site




- Show the screenshot of the scatter plot with explanations

```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```

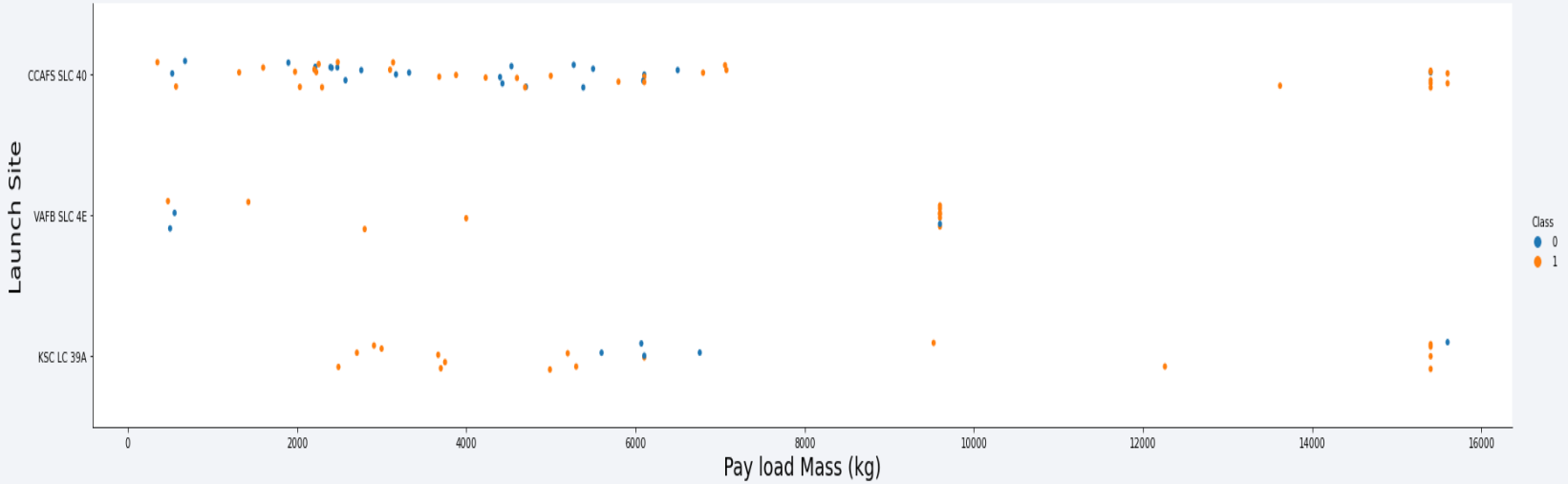



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

# Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



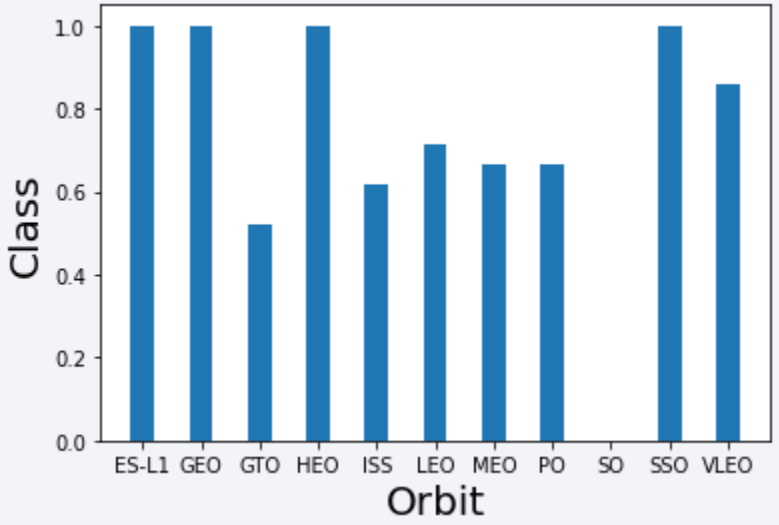

- Show the screenshot of the scatter plot with explanations

```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class valu
e
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("Pay load Mass (kg)",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```


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).

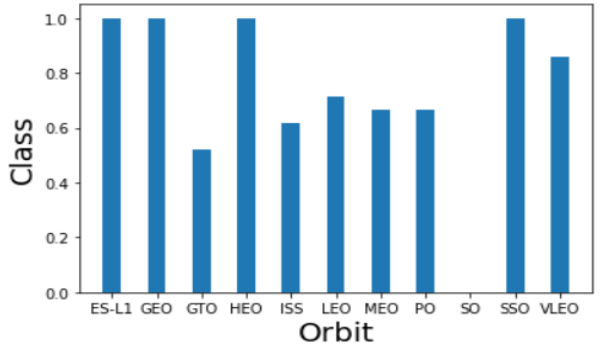
# Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type




- Show the screenshot of the scatter plot with explanations

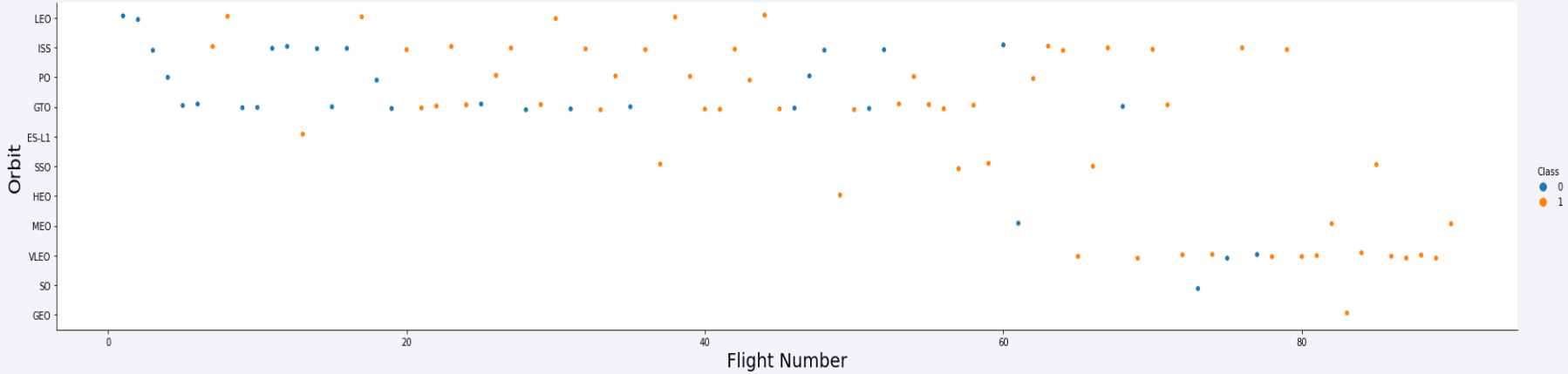
```
# HINT use groupby method on Orbit column and get the mean of Class column
#df[['Orbit','Class']].groupby('Orbit').mean().sort_values('Class')
plt.bar(np.unique(df["Orbit"]),df.groupby(['Orbit']).mean()['Class'],width = 0.4)
plt.xlabel("Orbit", fontsize = 20)
plt.ylabel("Class", fontsize = 20)
plt.show()
```




Analyze the ploted bar chart try to find which orbits have high sucess rate. Orbits SSO, ES-L1, GEO, HEO and VLEO are the ones with 80% or more success rate

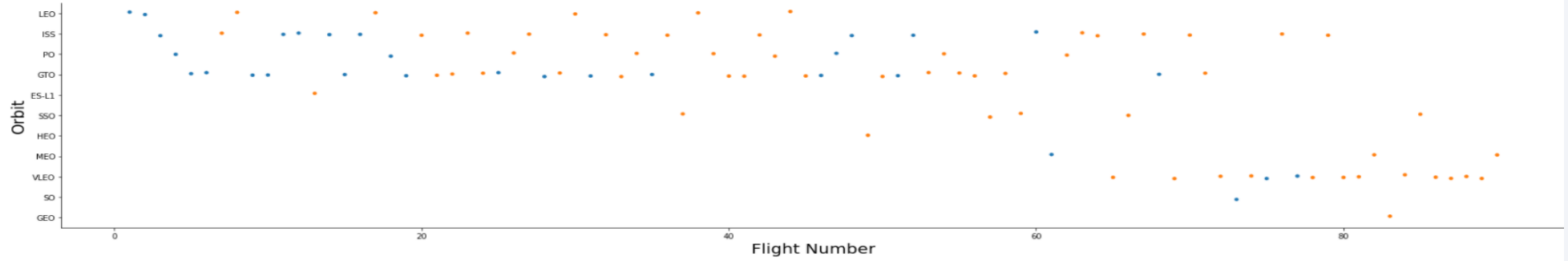
# Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type




- Show the screenshot of the scatter plot with explanations

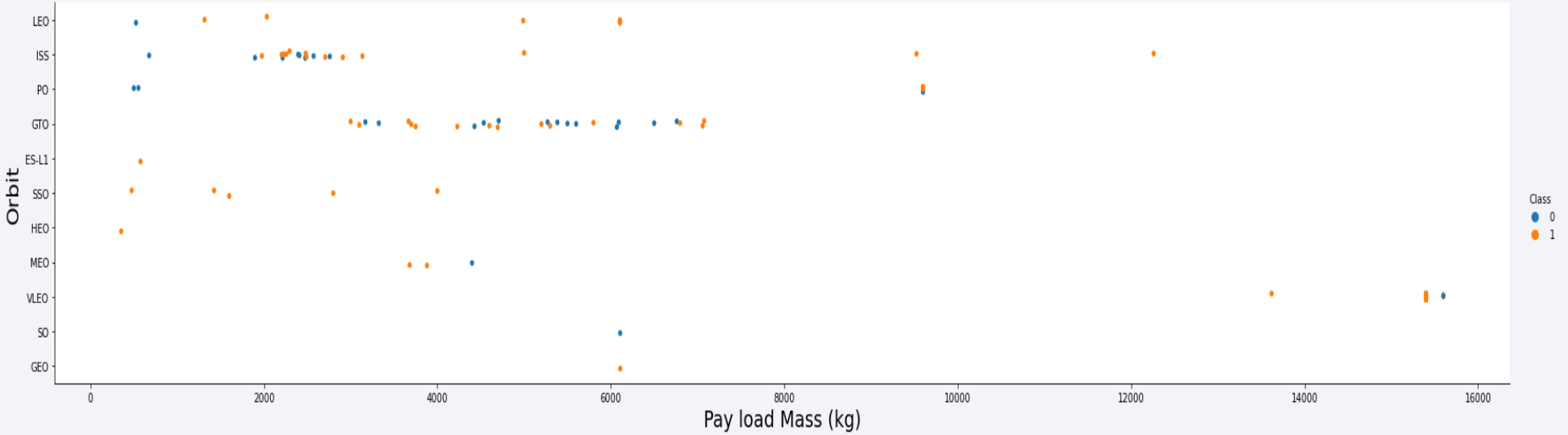
```
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```




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.

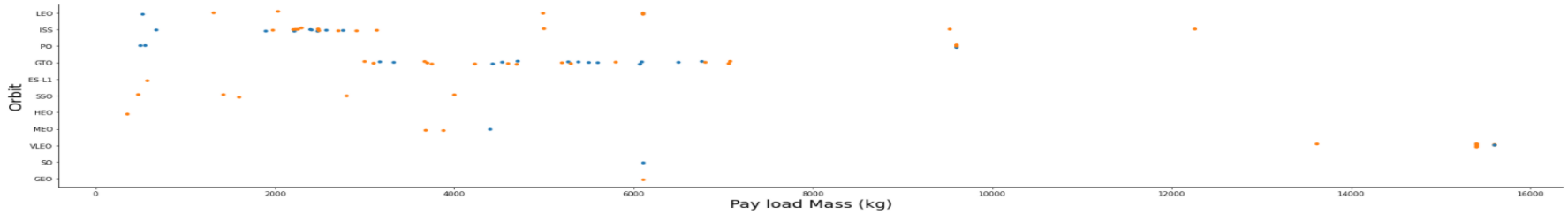
# Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type




- Show the screenshot of the scatter plot with explanations

```
# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("Pay load Mass (kg)",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

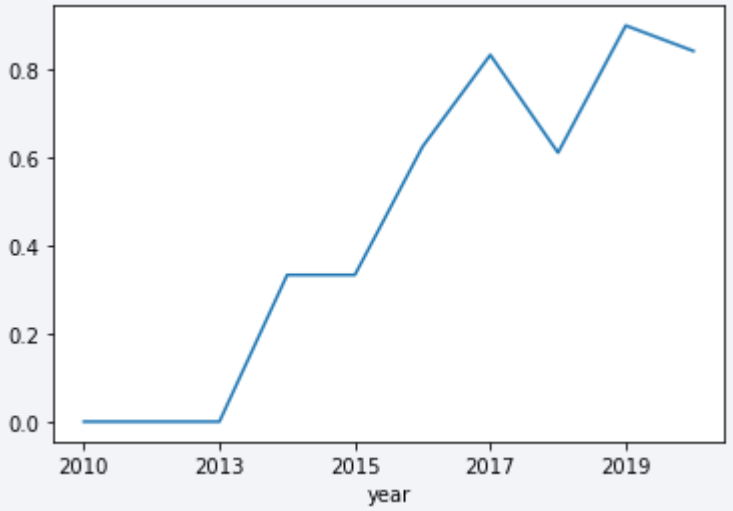



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

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

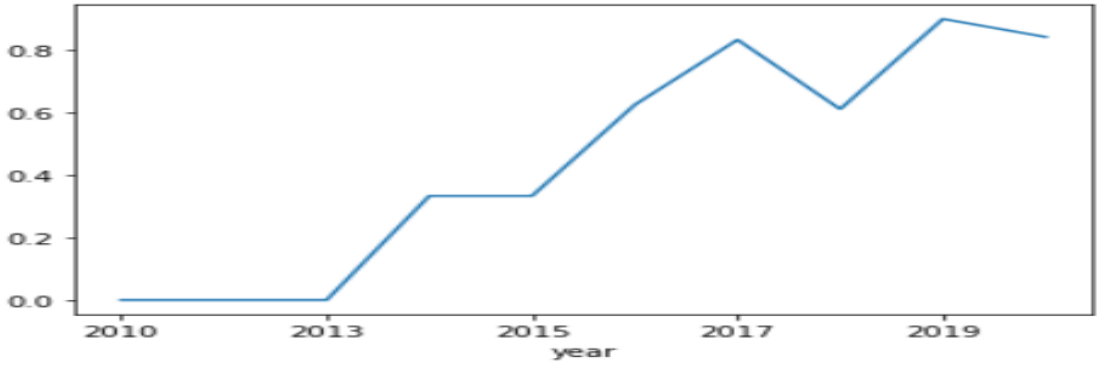
- Show a line chart of yearly average success rate




- Show the screenshot of the scatter plot with explanations

```
# Plot a line chart with x axis to be the extracted year and y axis
df1 = pd.DataFrame(Extract_year(df['Date']),columns = ['year'])
df1['Class'] = df['Class']
df2 =df1.groupby('year')['Class'].mean()
df2.plot(kind='line', y='Success Launch')
```

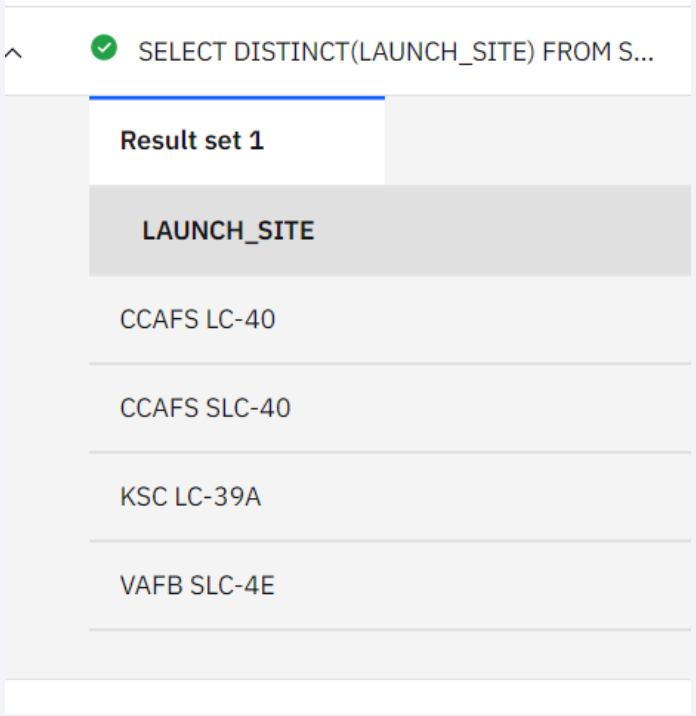
```
3]: <AxesSubplot:xlabel='year'>
```




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

# All Launch Site Names

- Find the names of the unique launch sites
  - CCAFS LC -40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
- Present your query result with a short explanation here



# Launch Site Names Begin with 'CCA'

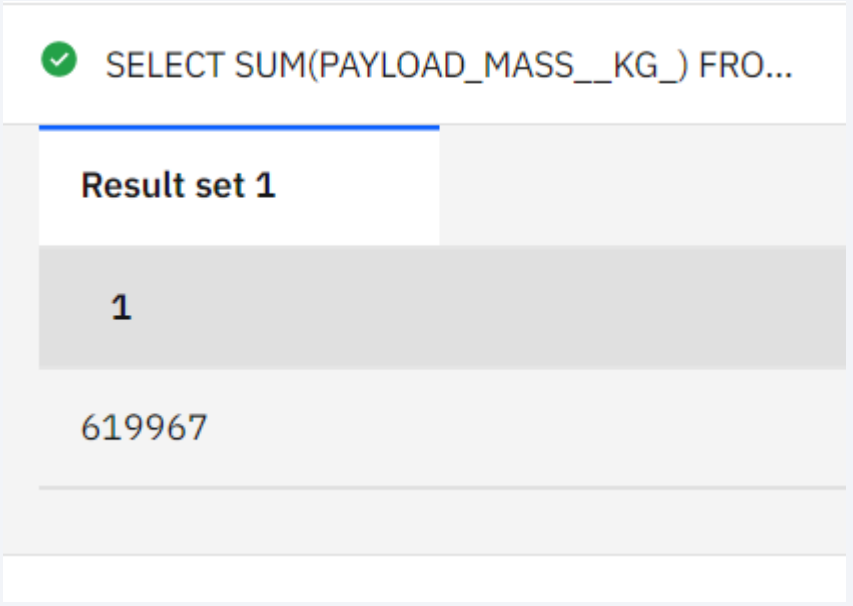
- Find 5 records where launch sites begin with `CCA`

| DATE | Time (UTC) | BOOSTER_VERSION | LAUNCH_SITE | PAYLOAD | PAYLOAD_MASS__KG_ | ORBIT | CUS |
|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | Space |
| 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 |
| 2012-05-22 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA |
| 2012-10-08 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA |

- Present your query result with a short explanation here

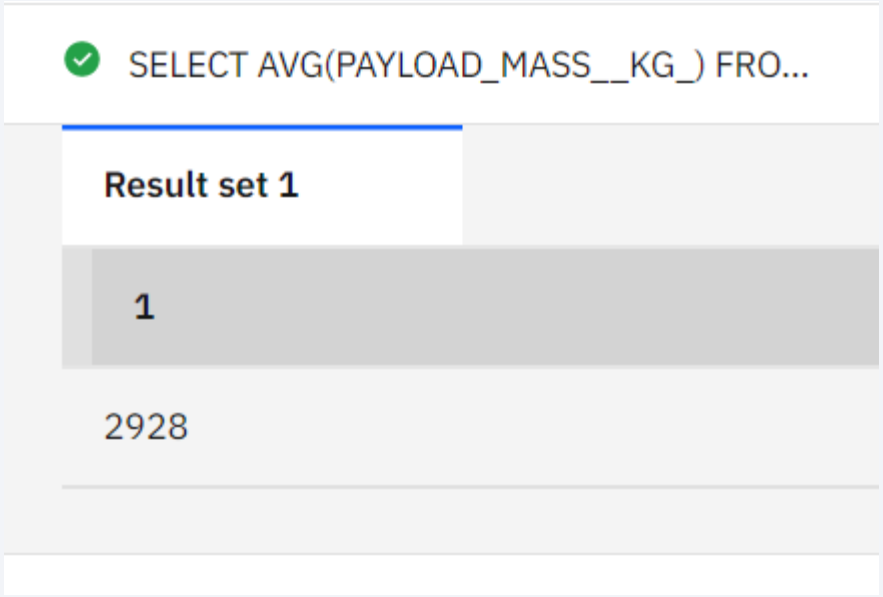
# Total Payload Mass

- Calculate the total payload carried by boosters from NASA.
  - 619967 Kg
- Present your query result with a short explanation here



# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - 2928 Kg
- Present your query result with a short explanation here



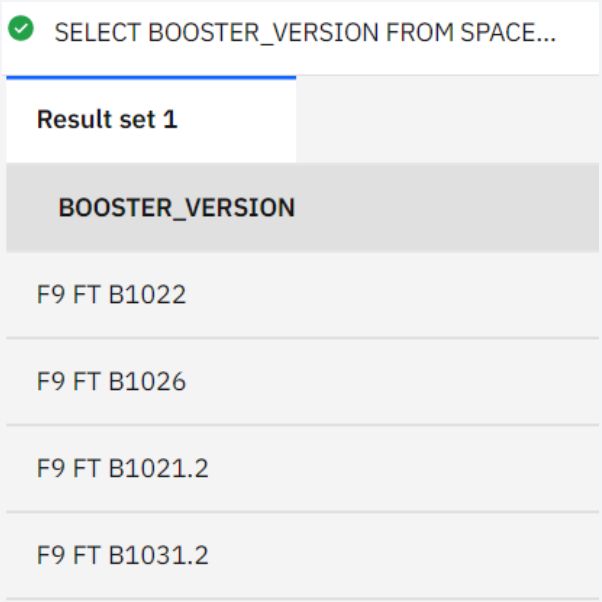
# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - 2015-12-22, 2016-07-18
- Present your query result with a short explanation here

| DATE |
| --- |
| 2015-12-22 |
| 2016-07-18 |
| 2017-02-19 |
| 2017-05-01 |
| 2017-06-03 |
| 2017-08-14 |
| 2017-09-07 |
| 2017-12-15 |
| 2018-01-08 |

# 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
  - F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2
- Present your query result with a short explanation here



# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - Success – 100, Failure - 1
- Present your query result with a short explanation here

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

| BOOSTER_VERSION |
| --- |
| F9 B5 B1048.4 |
| F9 B5 B1048.5 |
| F9 B5 B1049.4 |
| F9 B5 B1049.5 |
| F9 B5 B1049.7 |
| F9 B5 B1051.3 |
| F9 B5 B1051.4 |
| F9 B5 B1051.6 |
| F9 B5 B1056.4 |
| F9 B5 B1058.3 |
| F9 B5 B1060.2 |
| F9 B5 B1060.3 |

# 2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

SELECT BOOSTER_VERSION, LAUNCH_SI...

Result set 1

| BOOSTER_VERSION | LAUNCH_SITE |
|---|---|
| F9 v1.1 B1012 | CCAFS LC-40 |
| 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

- Present your query result with a short explanation here

Section 4

# Launch Sites Proximities Analysis

# Launch Sites

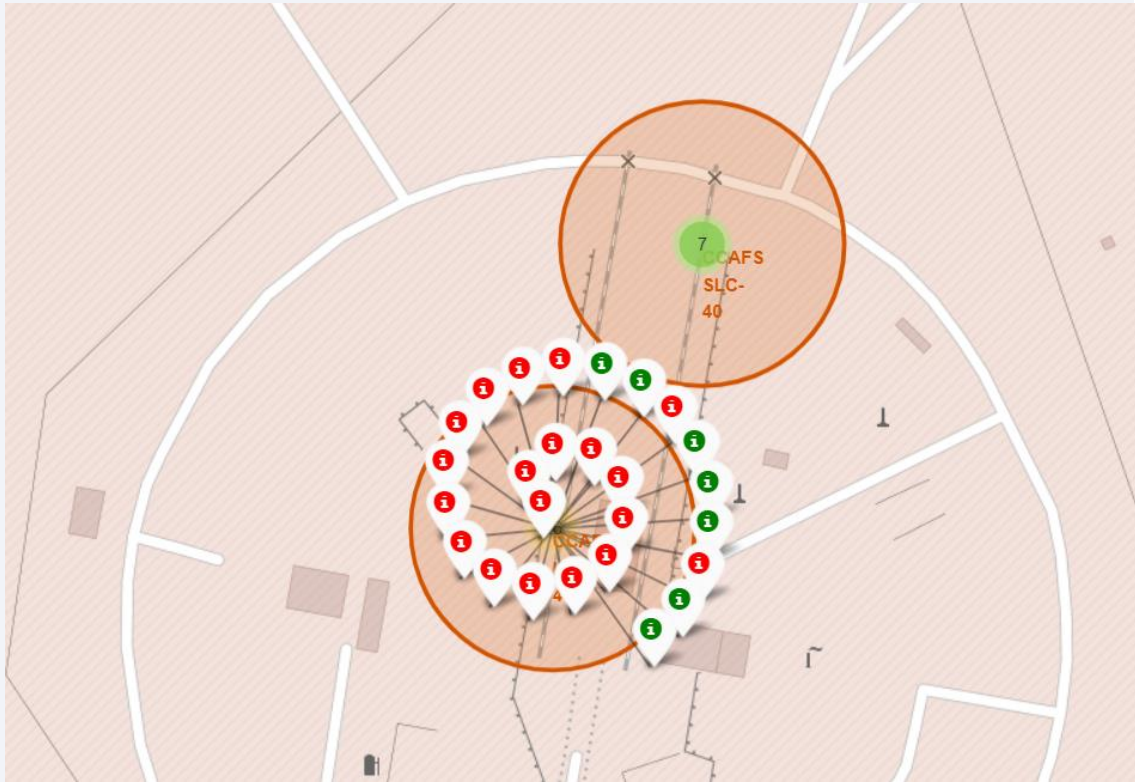
- Replace <Folium map screenshot 1> title with an appropriate title

- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

- Explain the important elements and findings on the screenshot

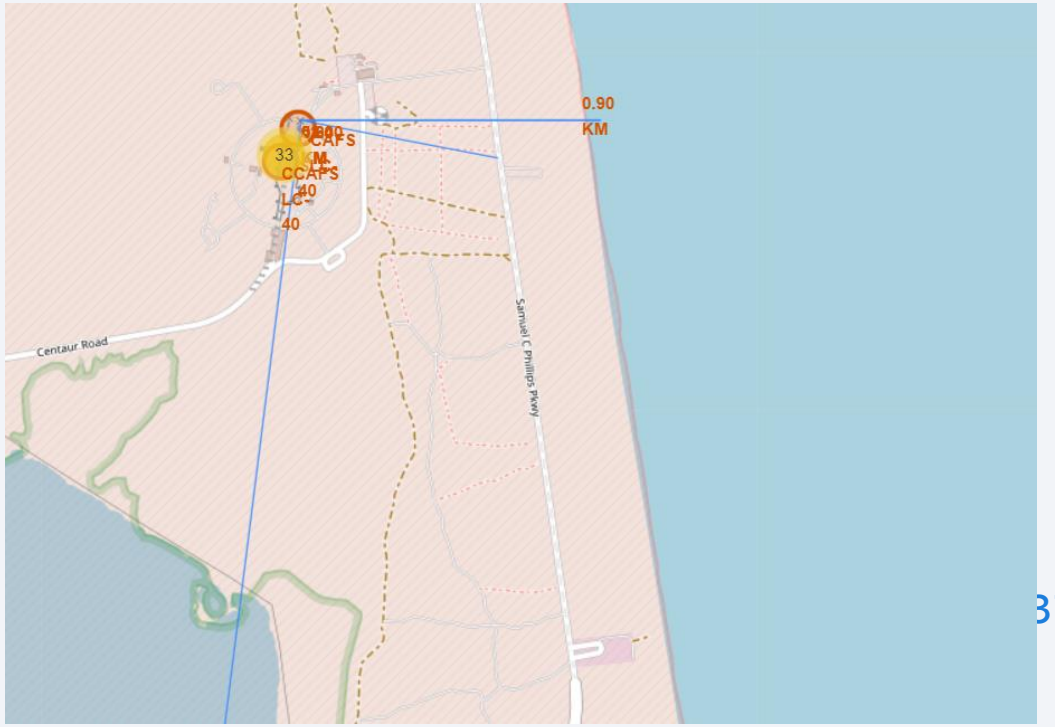
# Launch Site Marker Color

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot



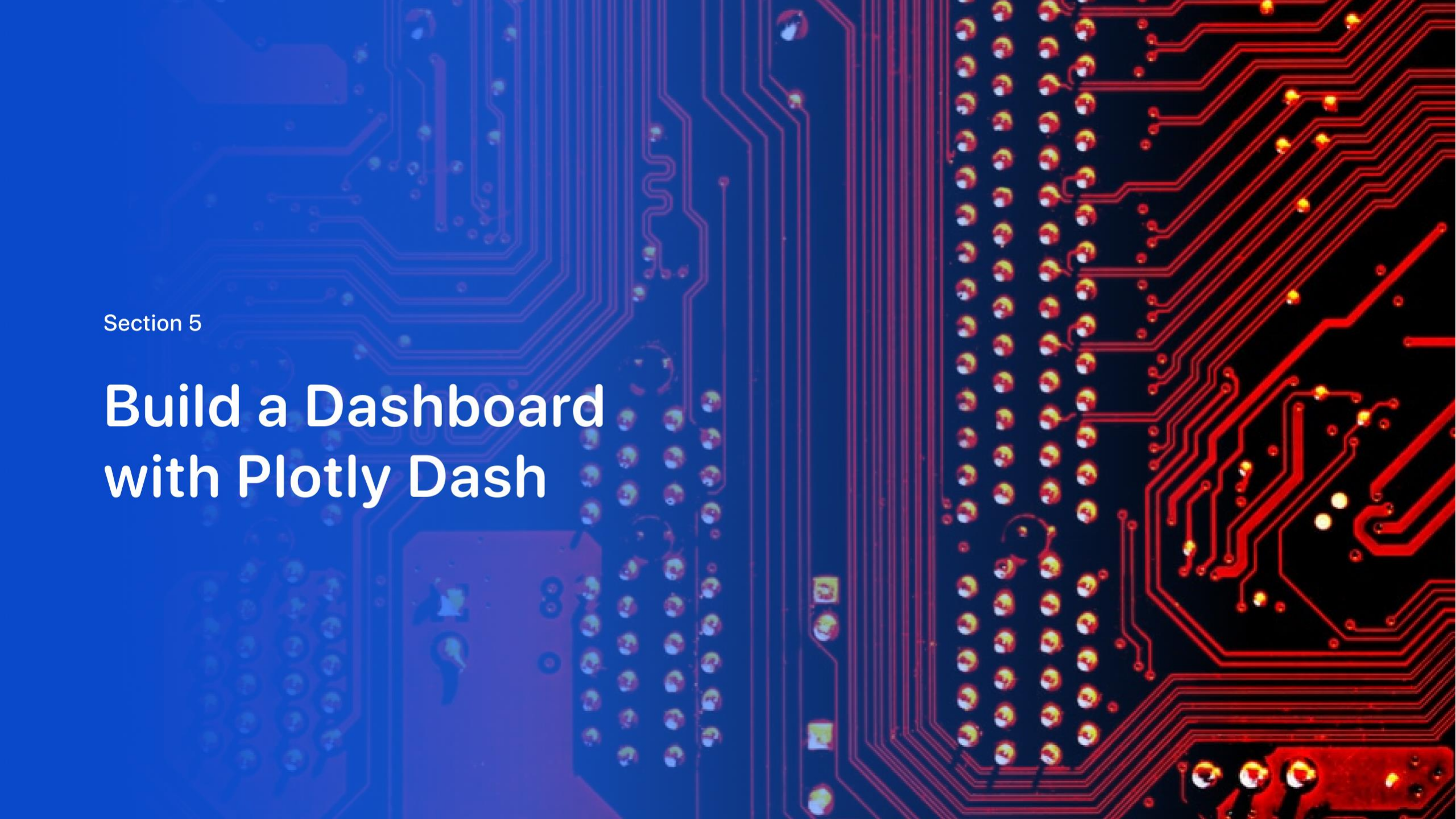
# Polyline

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot



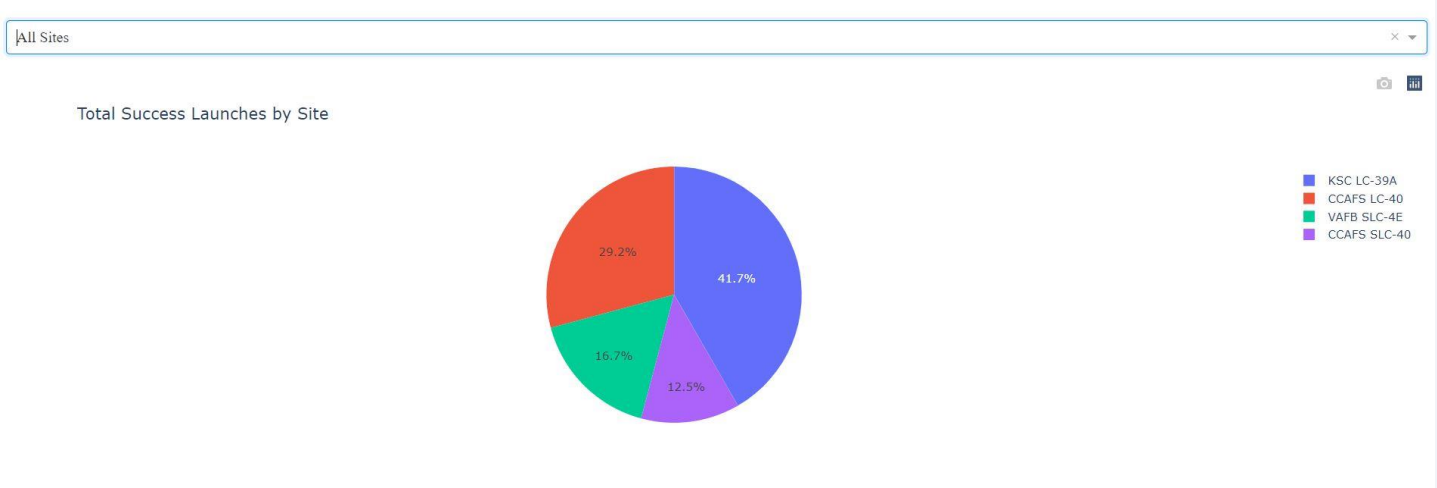
Section 5

# Build a Dashboard with Plotly Dash



# Pie Chart

- 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



# Launch Site with Highest launch success ration

- 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

# Payload vs. Launch Outcome Scatter Plot

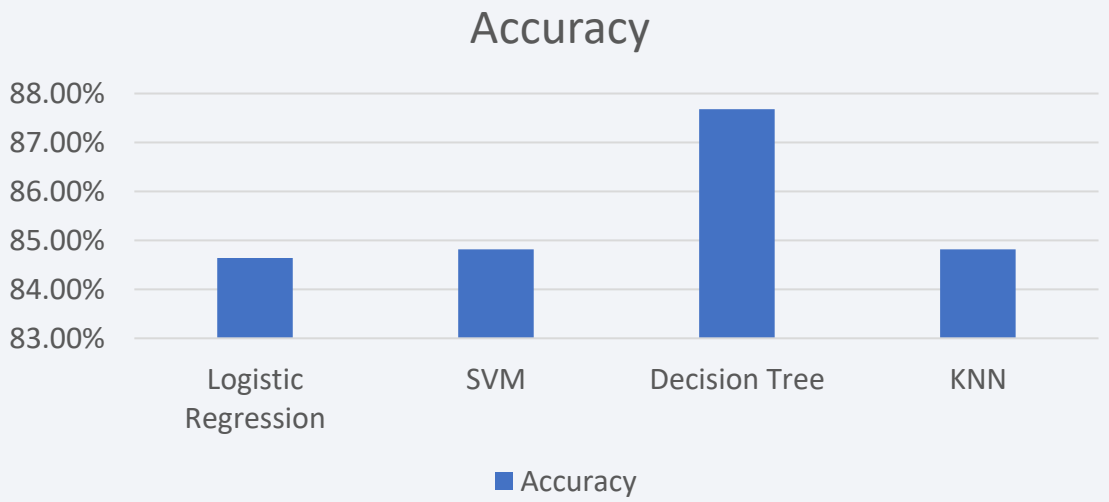
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version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

Section 6

# Predictive Analysis (Classification)

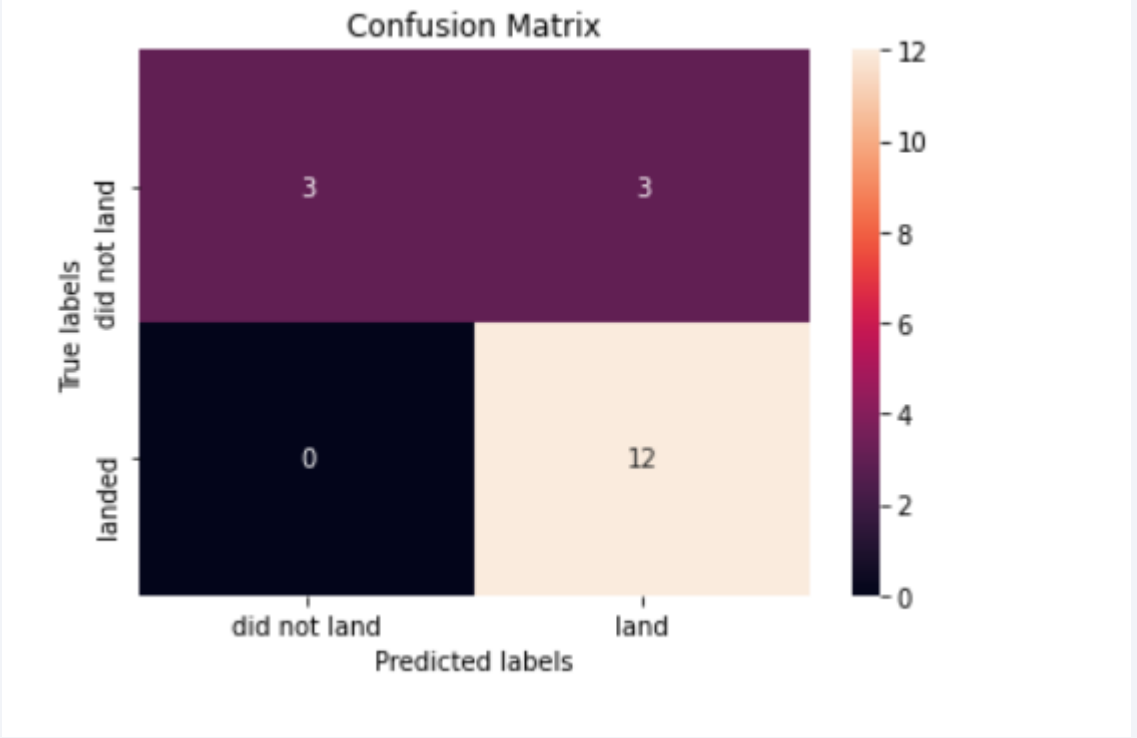
# Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy
  - Decision Tree



# Confusion Matrix

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



# Conclusions

- Point 1
- Decision Tree Model has high classification accuracy

# 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!