IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Joseph Harmonson
10/17/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA) with Data Analysis
  - EDA with SQL
  - Interactive Maps with Folium
  - Interactive Dashboard
  - Predictive Analysis with Machine Learning
- Summary of all results
  - EDA results
  - Interactive Dashboard with screenshots
  - Machine Learning Results

Introduction

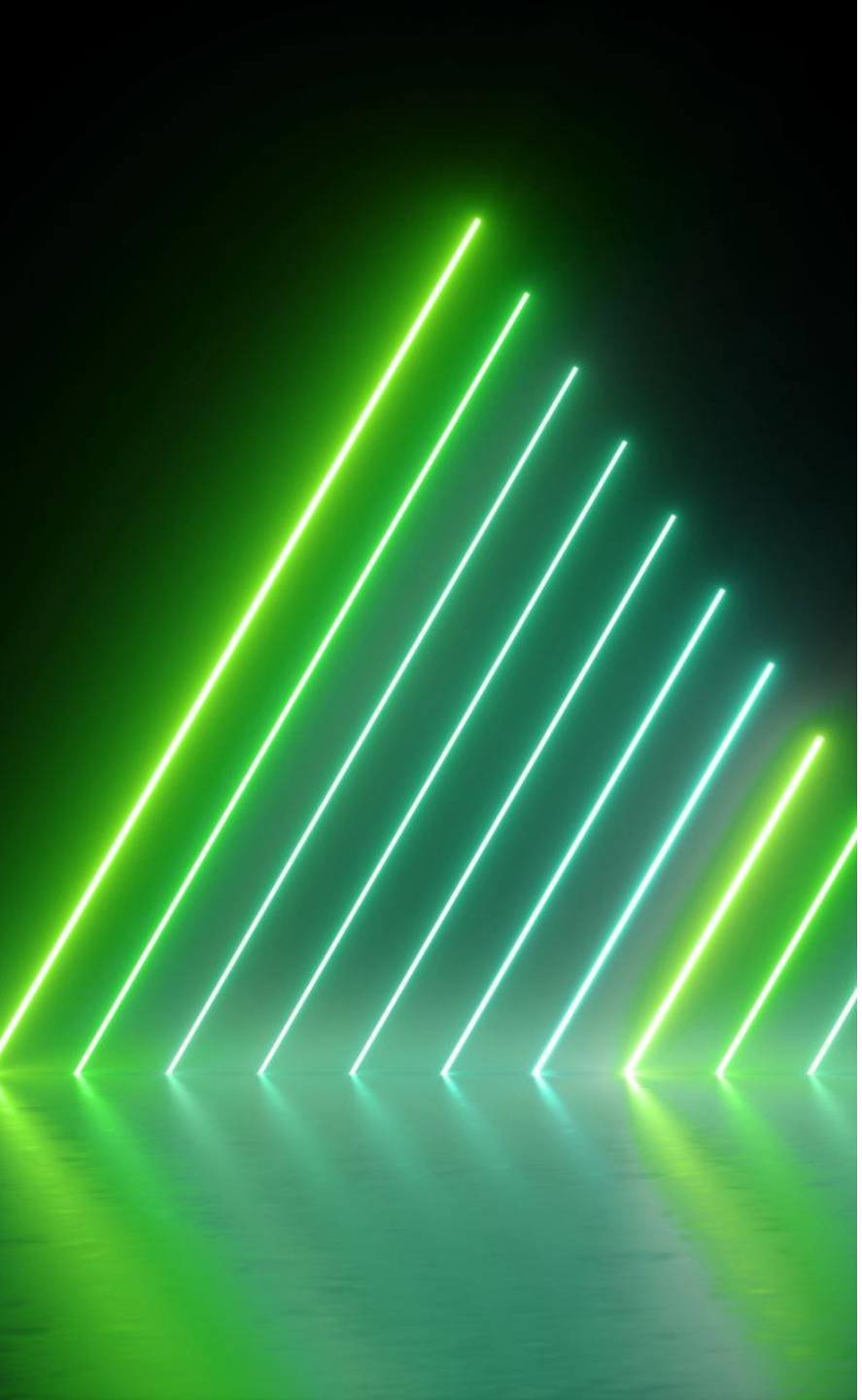
- Project background and context
    - SpaceX has shaken space exploration with its ability to create cheaper rockets. This cost reduction is made possible due to its rocket reusability. This cuts the cost by approximately $100 million. If we can determine whether the first stage can be reused due to a successful landing, we can determine the cost to continue a program similar to SpaceX.
- Problems you want to find answers
    - What factors determine a successful landing of the first stage?
    - What variables impact the success rate?
    - What conditions provide the best results?

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
    - Get request to the SpaceX API
    - Web scrapping from Wikipedia
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Built, tuned, and evaluated classification models
        - Logistic Regression, K Nearest Neighbor, Data Classification Tree, and Support Vector Machine



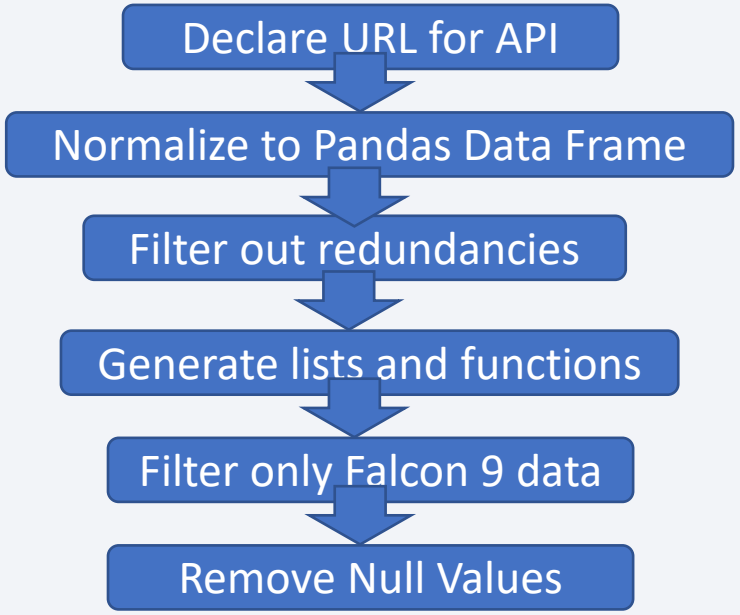
Data Collection

- Request to the SpaceX API
- Requested data was cleaned and missing values were removed
- Data was scraped from Wikipedia Launch Records
- Data was parsed and converted to a Pandas data frame

Data Collection – SpaceX API

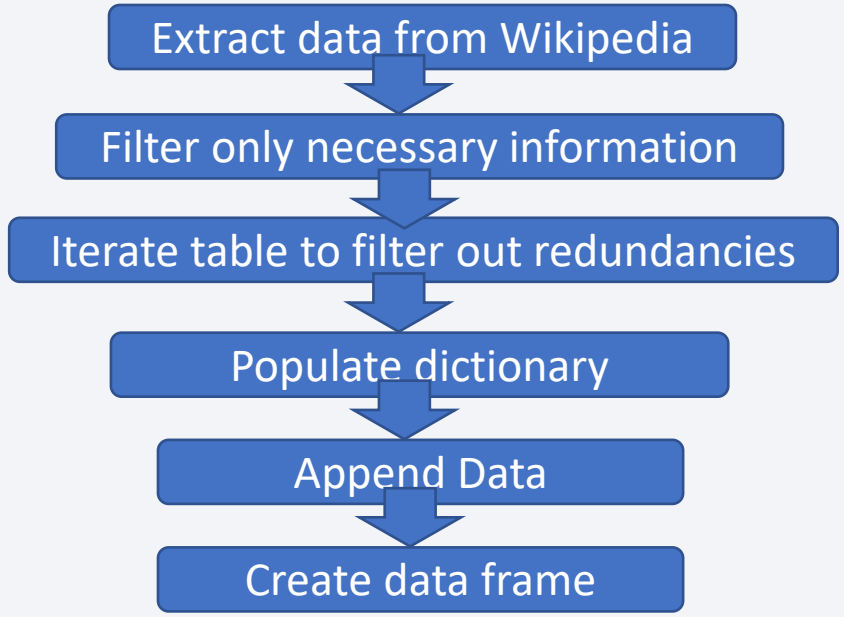
- [SpaceX API notebook with outcomes]

ghtNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
8	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
10	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
11	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
12	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857



Data Collection - Scraping

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame
- [Web Scrapping Notebook GitHub URL]



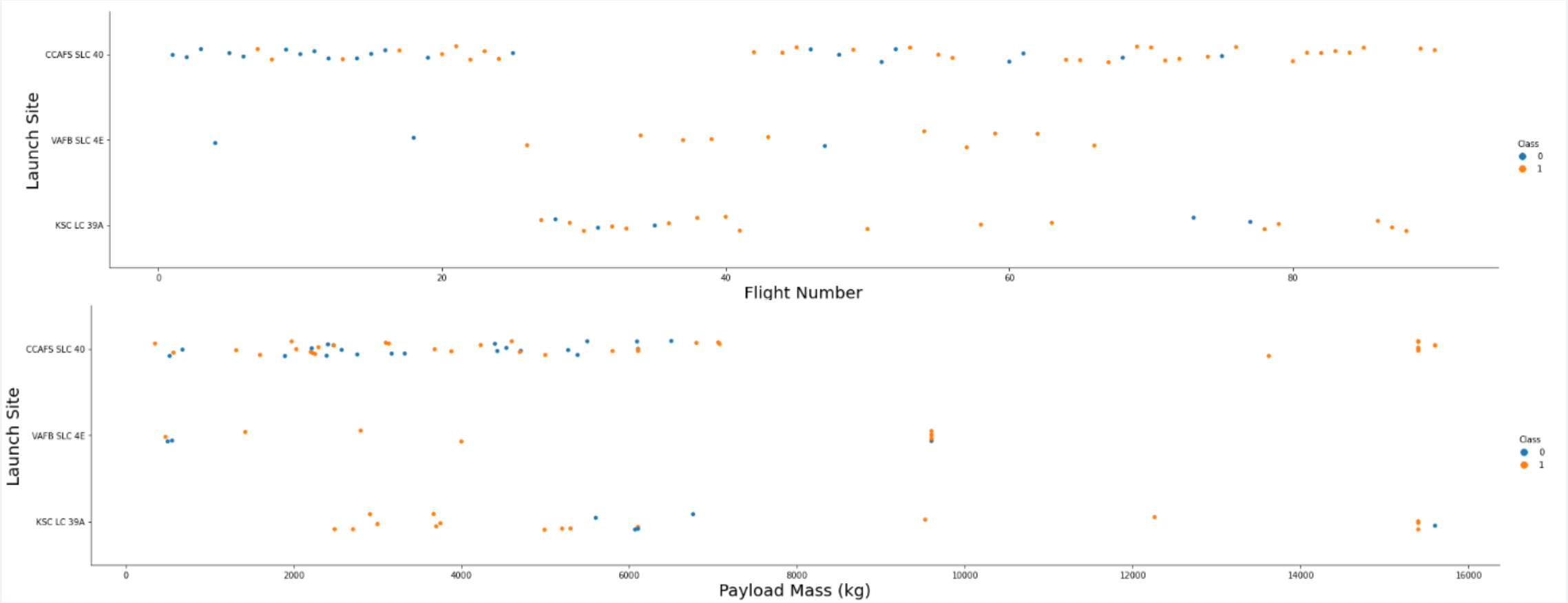
	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA (COTS)\nNRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Data Wrangling

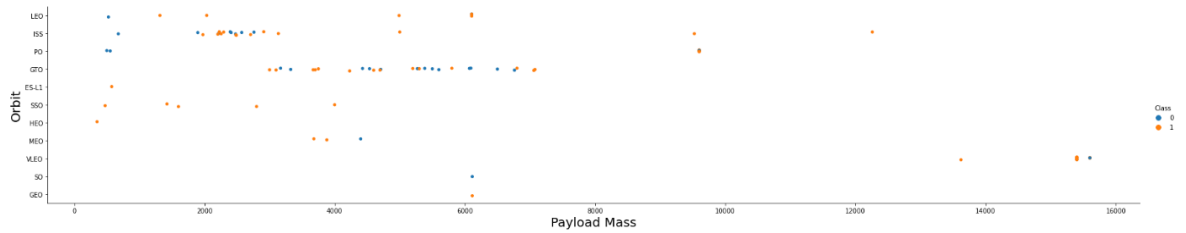
- Exploratory Data Analysis
- Determine Training Labels
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit
- Create a landing outcome label from outcome column
- Data Wrangling Notebook GitHub URL

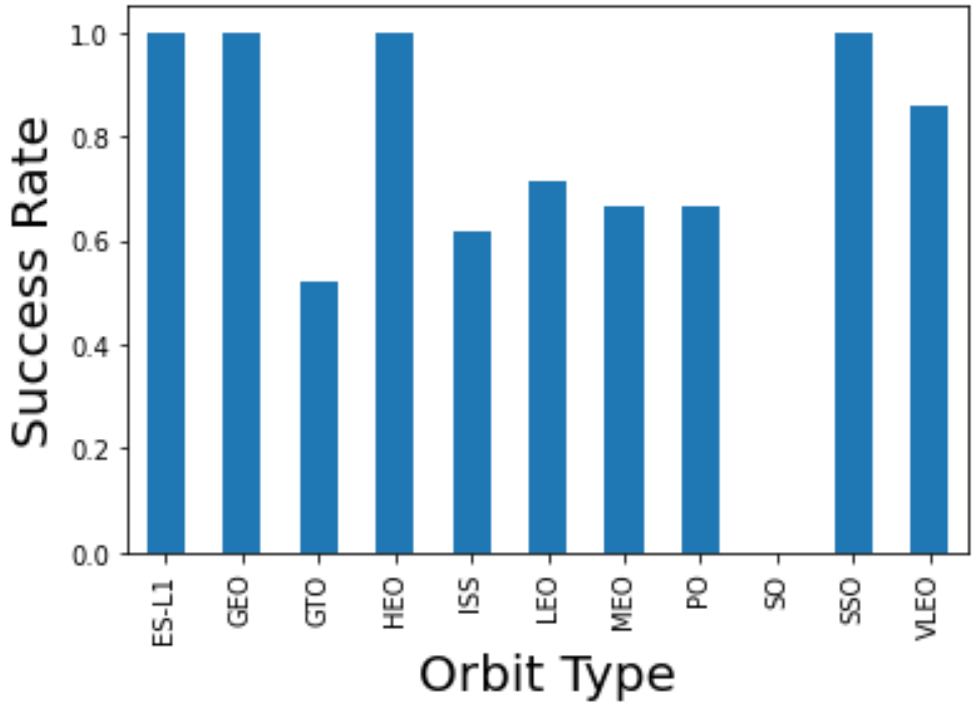
EDA with Data Visualization

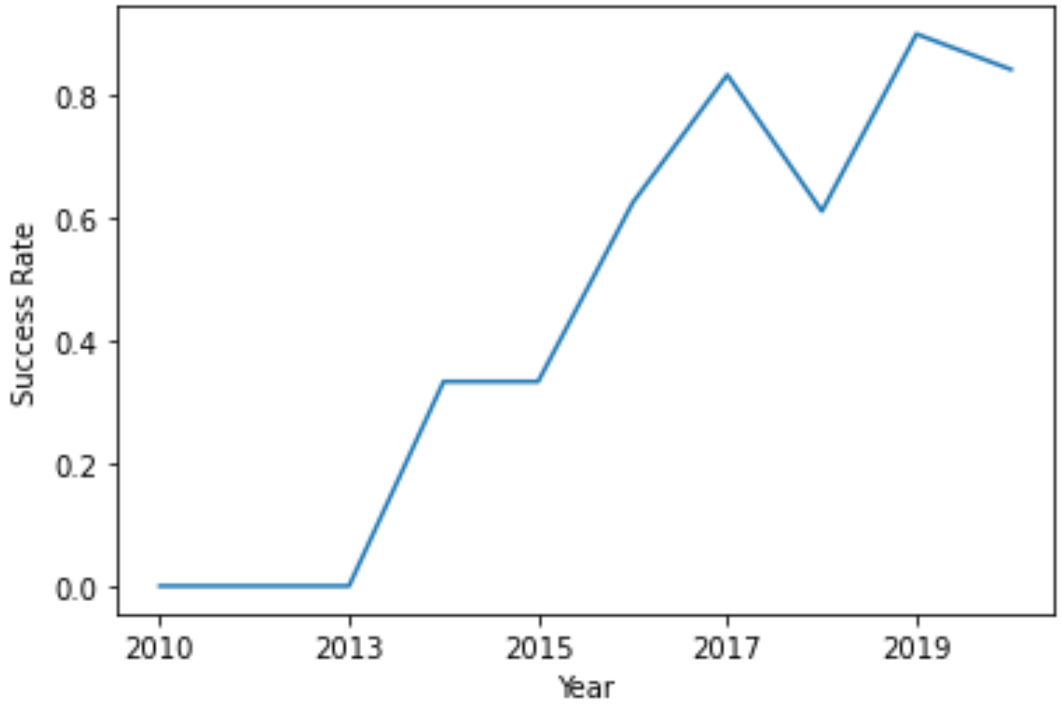
- EDA with Data Visualization GitHub URL
- Perform exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib using visuals



EDA with Data Visualization

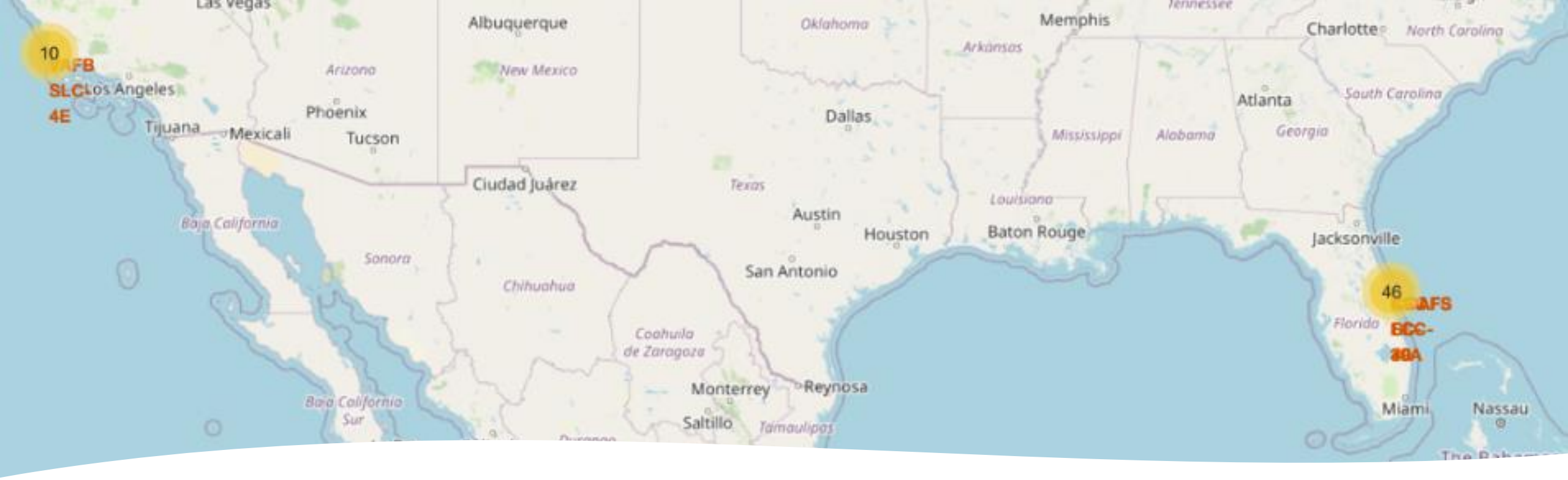






EDA with SQL

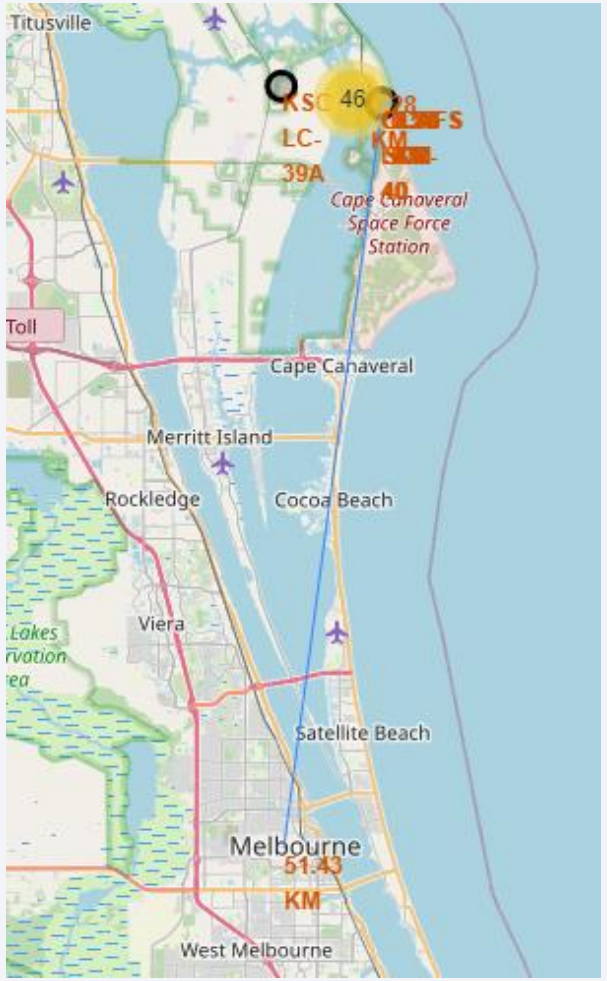
- EDA with SQL GitHub URL
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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

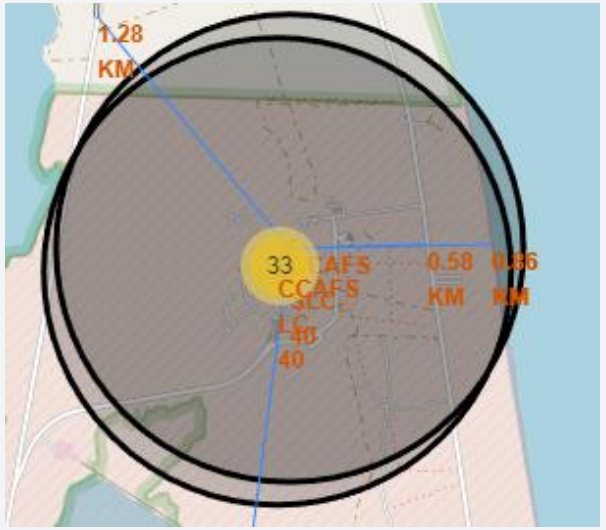


Build an Interactive Map with Folium

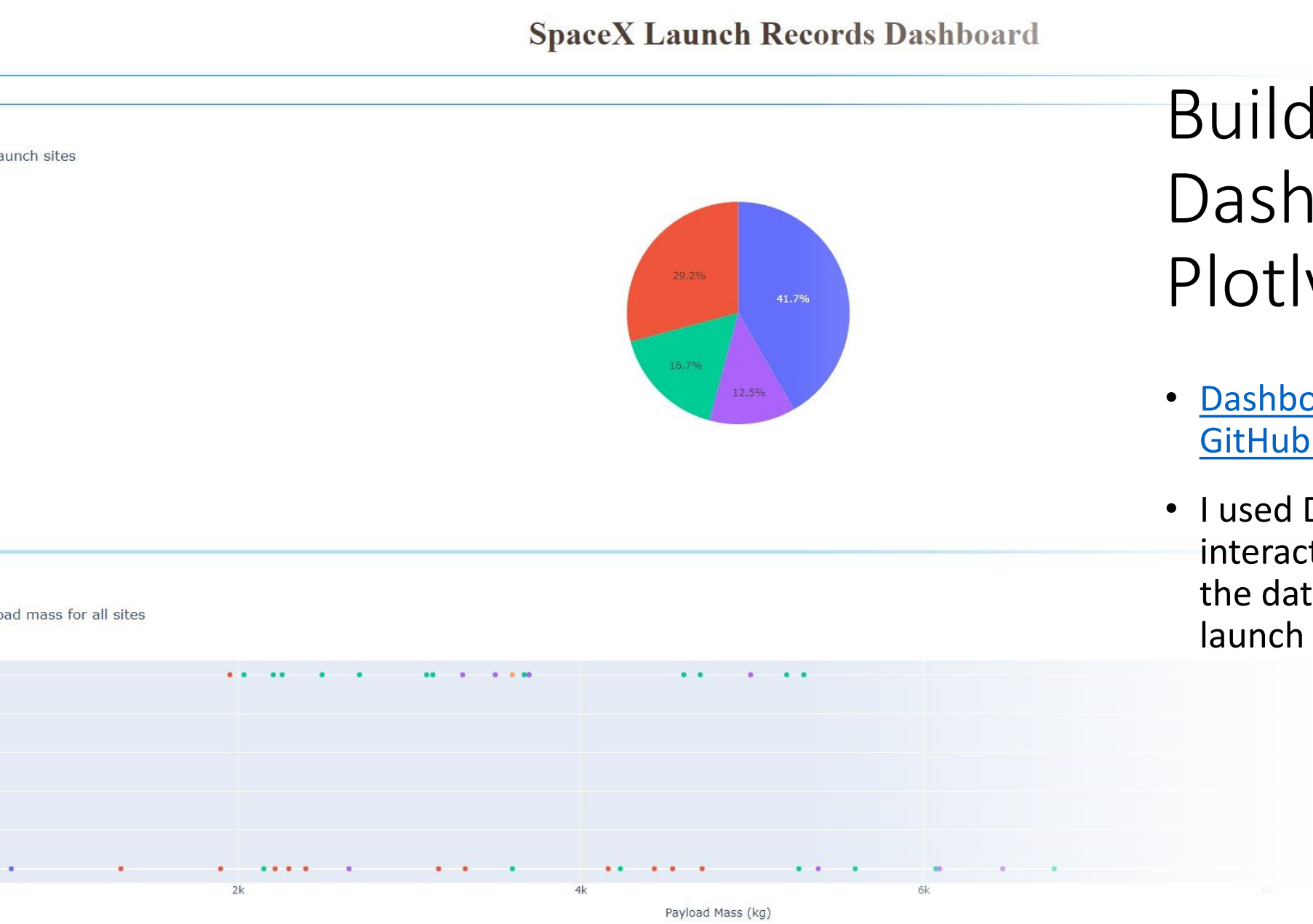
- [Interactive Map with Folium GitHub URL]
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Build an Interactive Map with Folium






```
distance_highway = 0.5834695366934144  km
distance_railroad = 1.2845344718142522  km
distance_city = 51.43416999517233  km
```

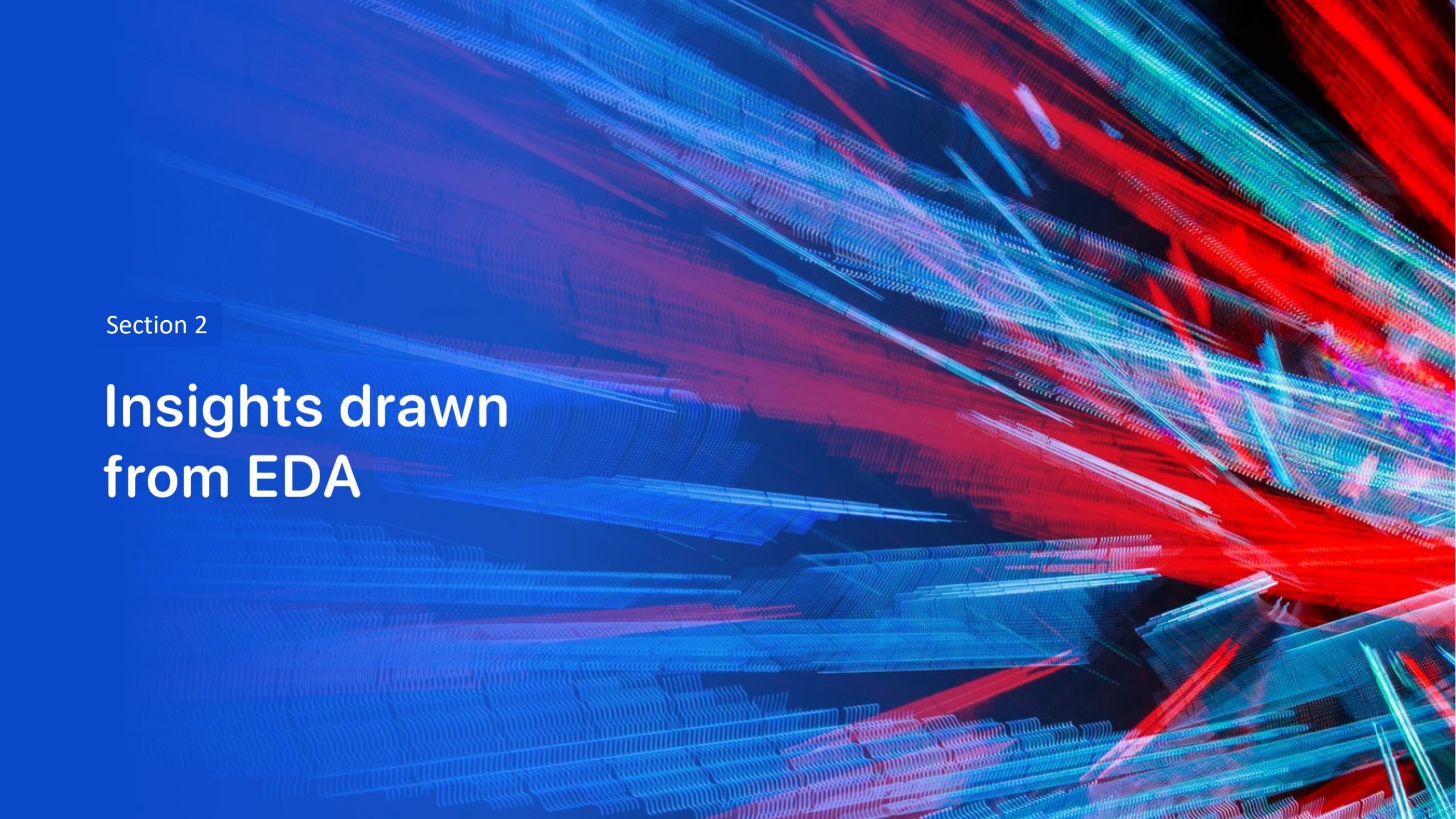


Build a Dashboard with Plotly Dash

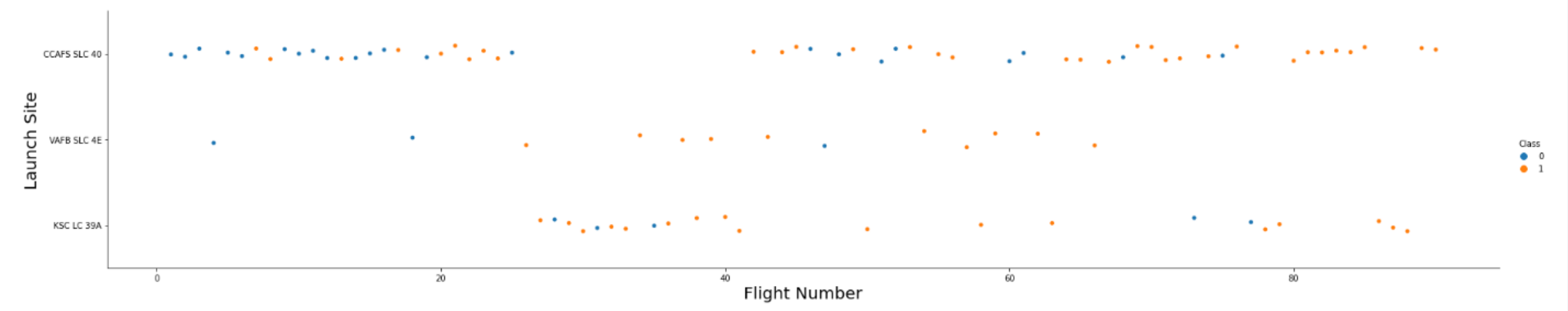
- [Dashboard with Ploly Dash GitHub URL]
- I used Dash to create an interactive dashboard to analyze the data from the launches and launch sites.

Predictive Analysis (Classification)

- Machine Learning Prediction GitHub URL
- Logistic Regression, K Nearest Neighbor, Data Classification Tree, and Support Vector Machine
- Predictions performed at 83%

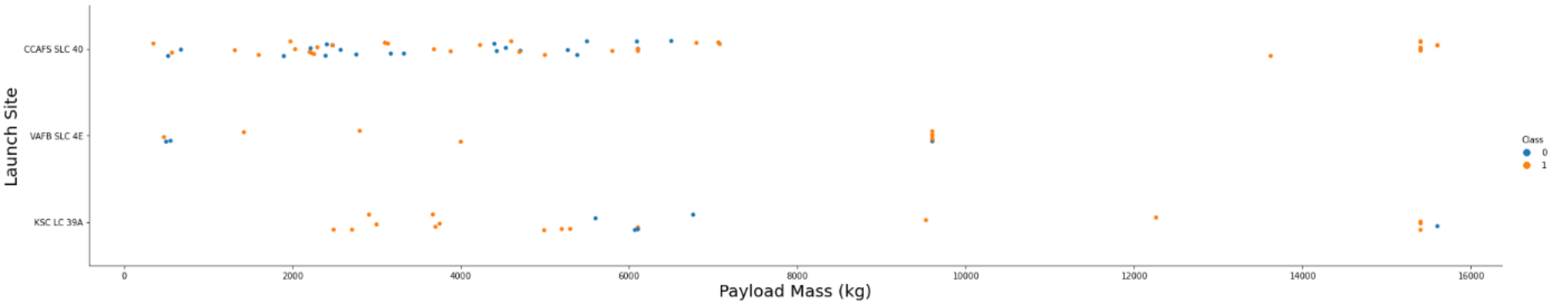

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- The scatter plot above displays the Flight Number vs Launch Site
- Class 1 refers to a success
- Class 0 refers to a failure

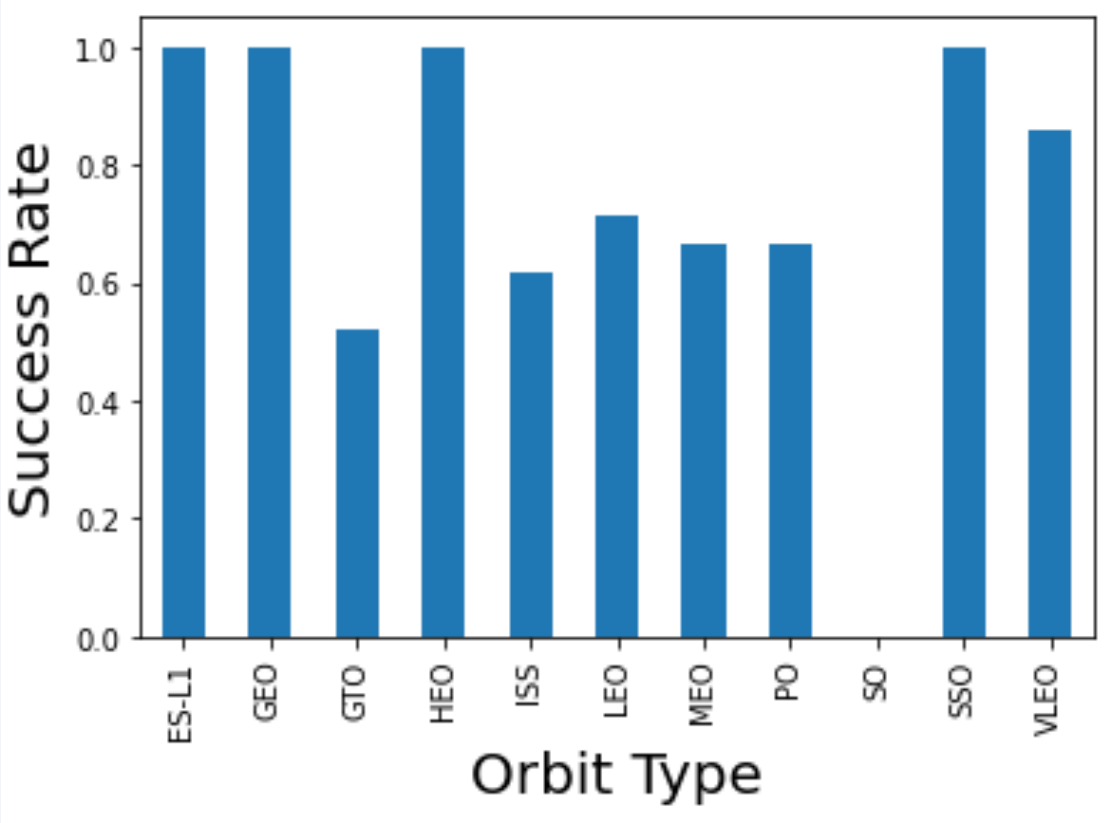
Payload vs. Launch Site



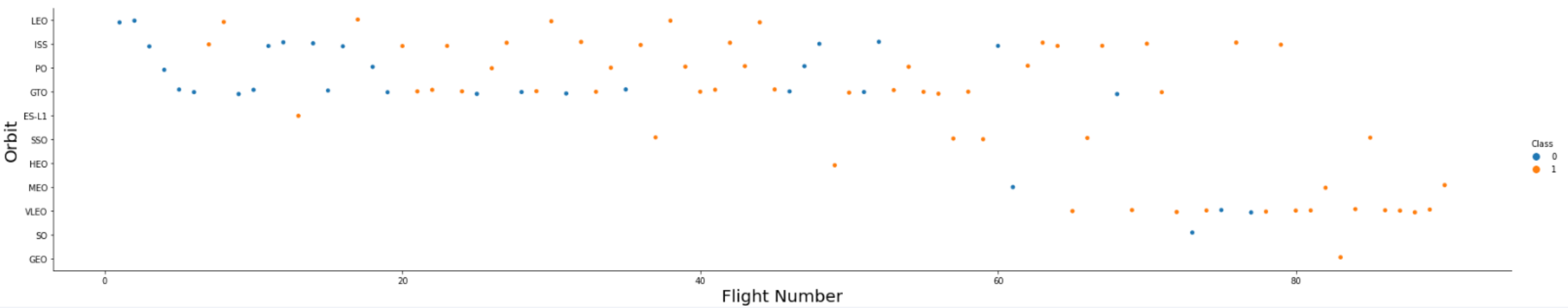
- The scatter plot above displays the Payload vs Launch Site
- Class 1 refers to a success
- Class 0 refers to a failure

Success Rate vs. Orbit Type

- This plot displays the Payload vs Orbit Type

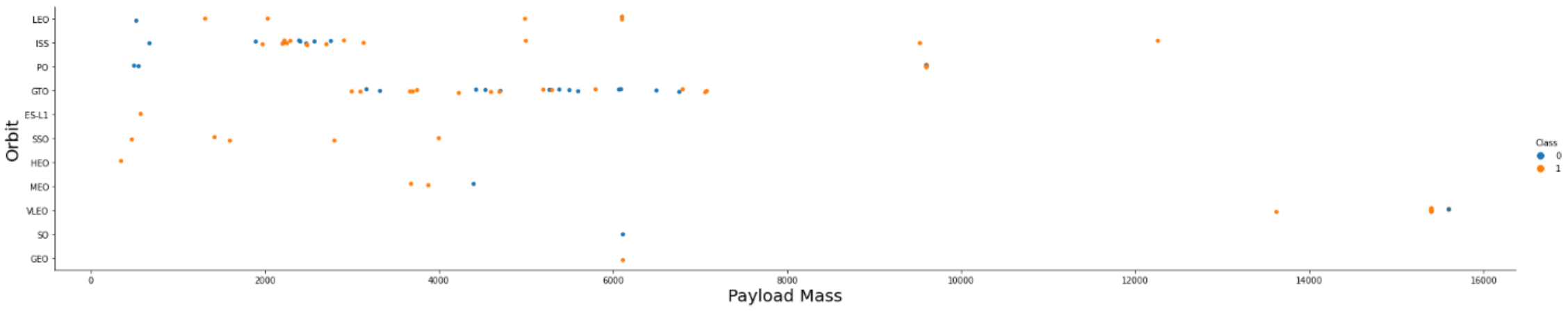


Flight Number vs. Orbit Type



- The scatter plot above displays the Flight Number vs. Orbit Type
- Class 1 refers to a success
- Class 0 refers to a failure

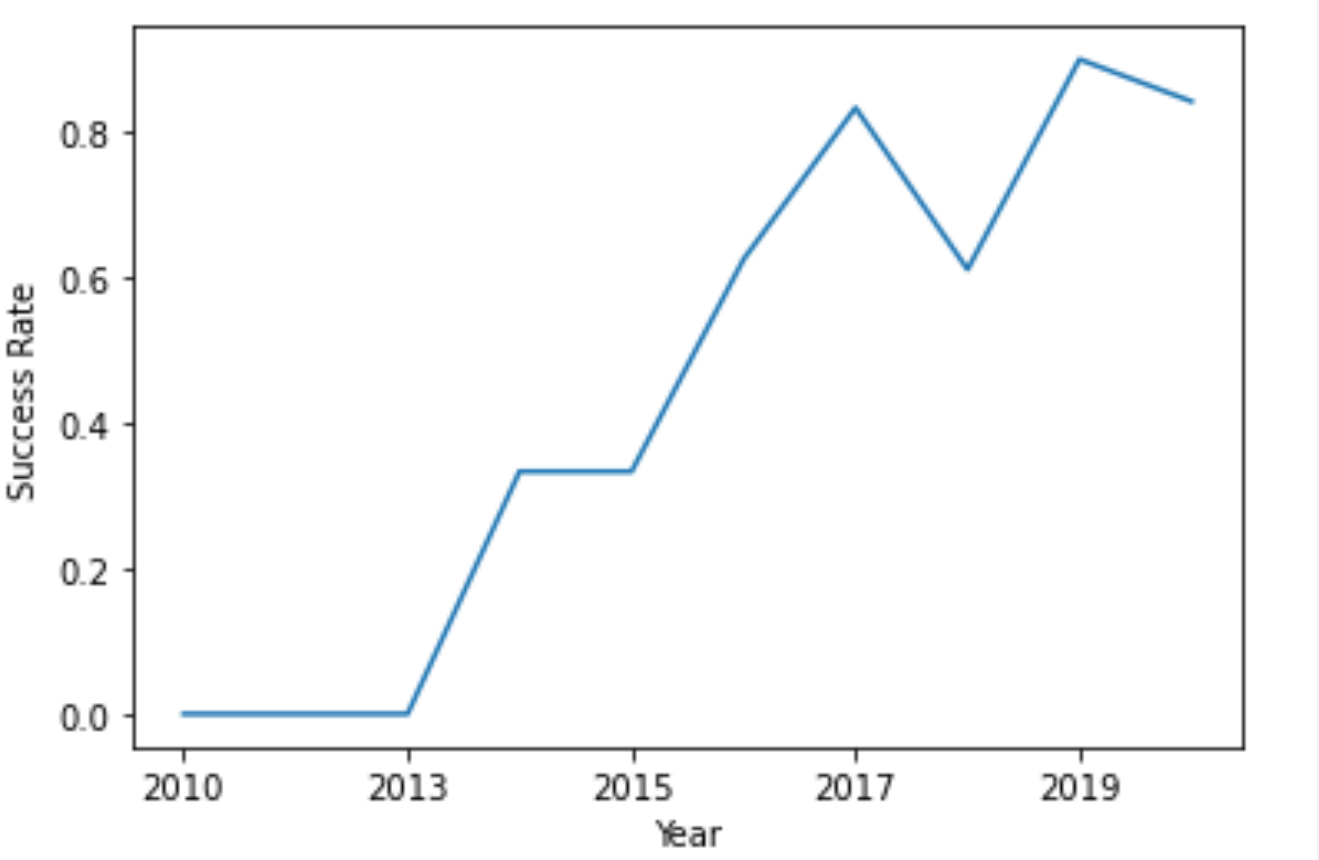
Payload vs. Orbit Type



- The scatter plot above displays the Payload vs. Orbit Type
- Class 1 refers to a success
- Class 0 refers to a failure

Launch Success Yearly Trend

- This graphs shows the yearly trend of launch successes.



All Launch Site Names

- Display the names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

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

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- 45,596 KG

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- 2,534 KG

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- 12/22/2015

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

booster_version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

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

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

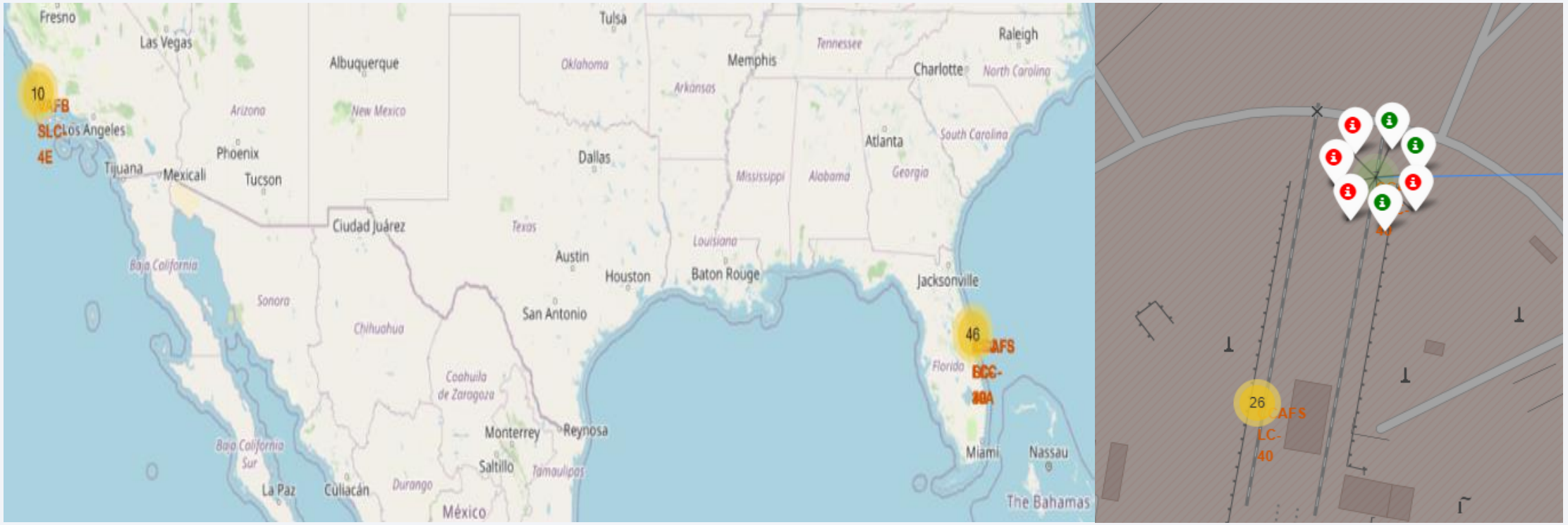
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

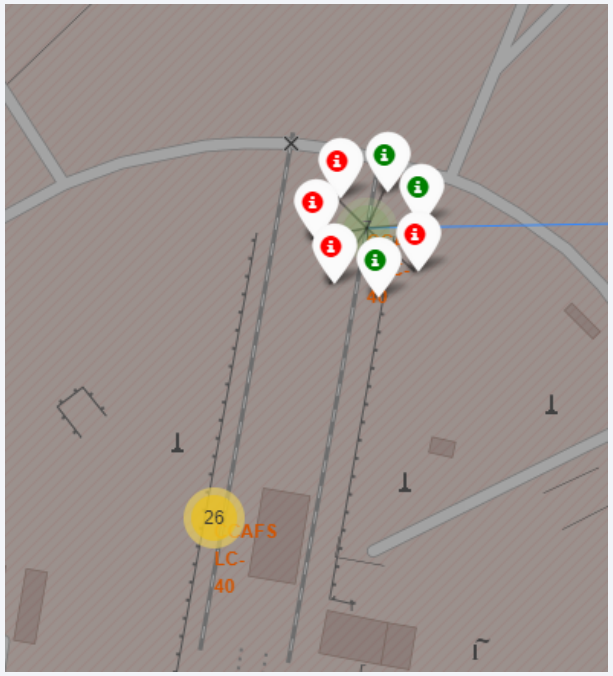
Launch Sites Proximities Analysis

Cluster Map of Launch Sites



- This map is a cluster of launch sites. When you zoom in on each cluster, it will show you specific launch sites with outcomes.

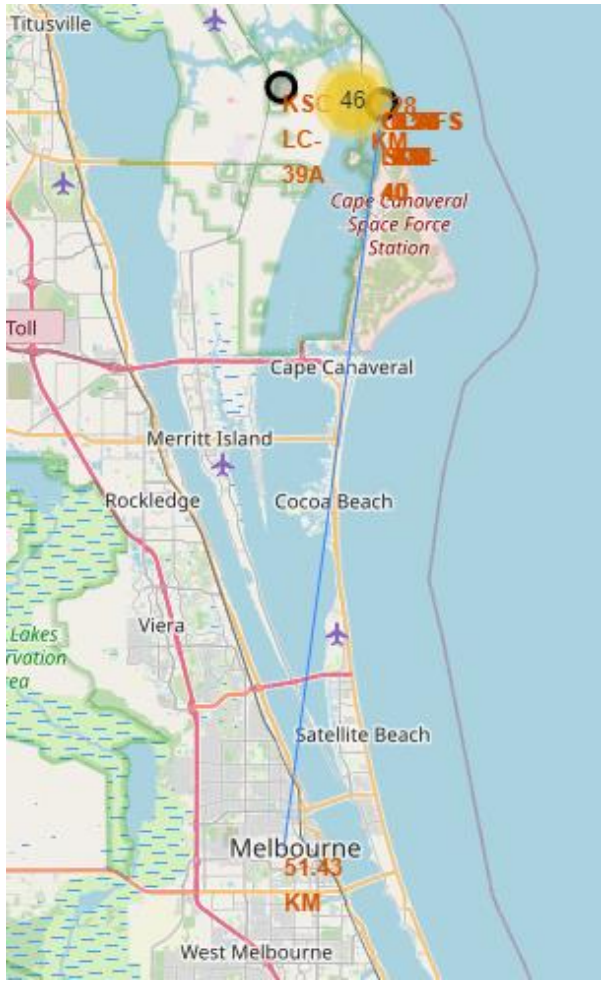
Luanch Map with Outcomes and Proximities

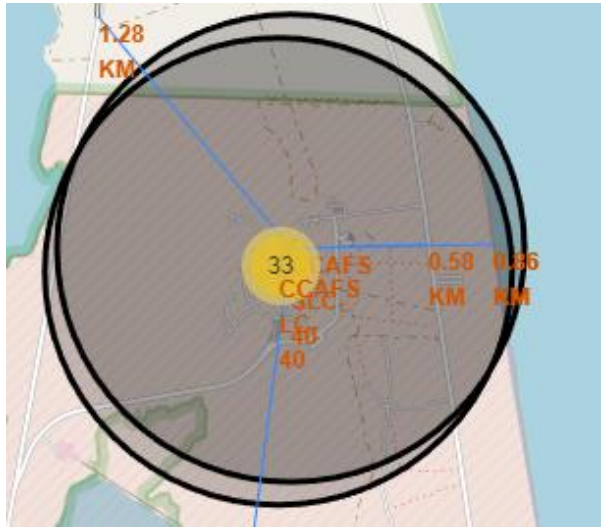


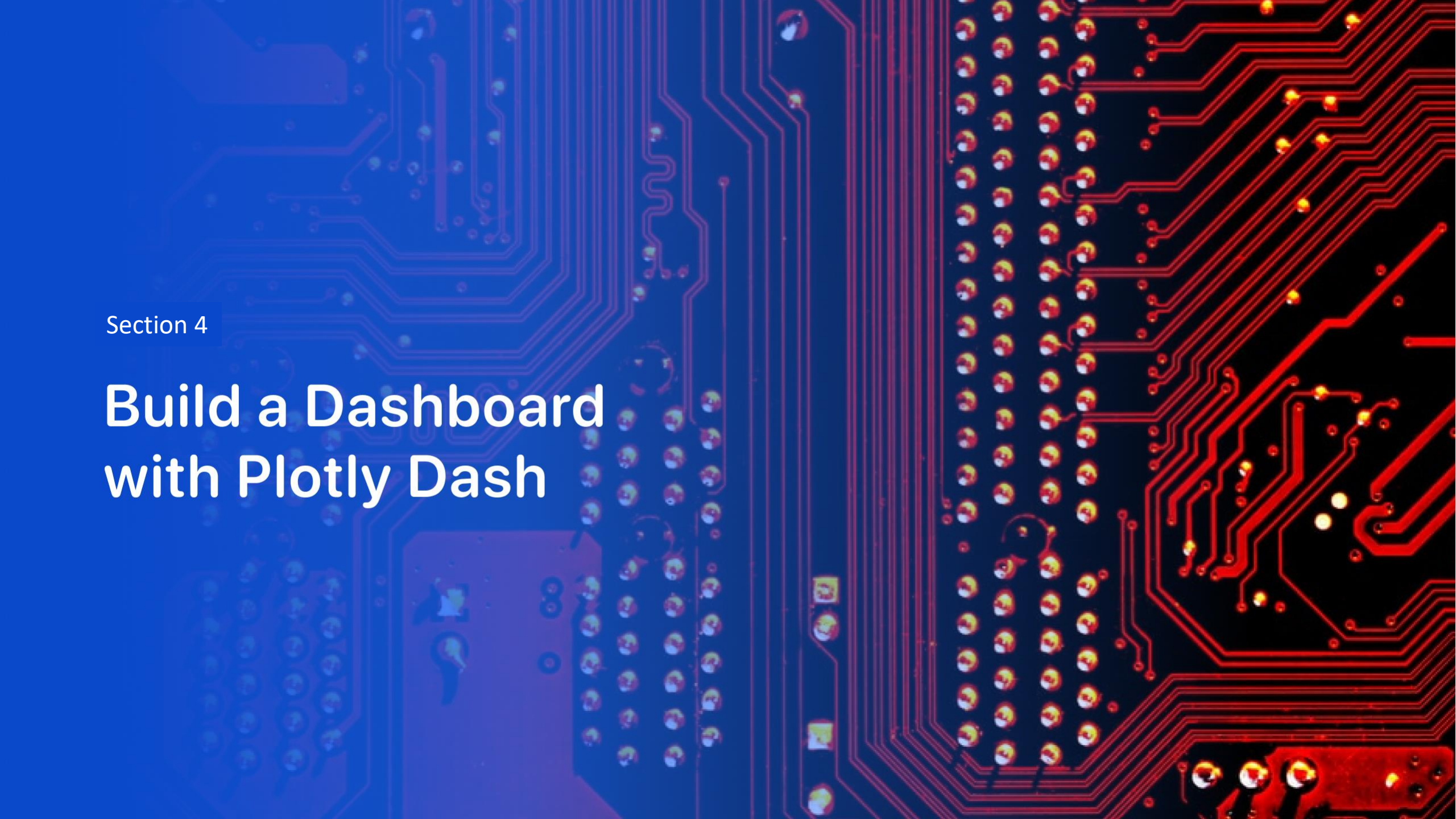
- The map was used to determine the distances from railways, highways, coastline and cities. The launch sites are in close proximity to railways and highways to allow for easy transportation. They are in close proximity to the coast in case of a failure they can touch down in water. Launch sites are away from major cities to keep clear of large populations.

Proximity Map with Folium

- This map shows the distances between the railway, highway, coastline, and the closest major city.



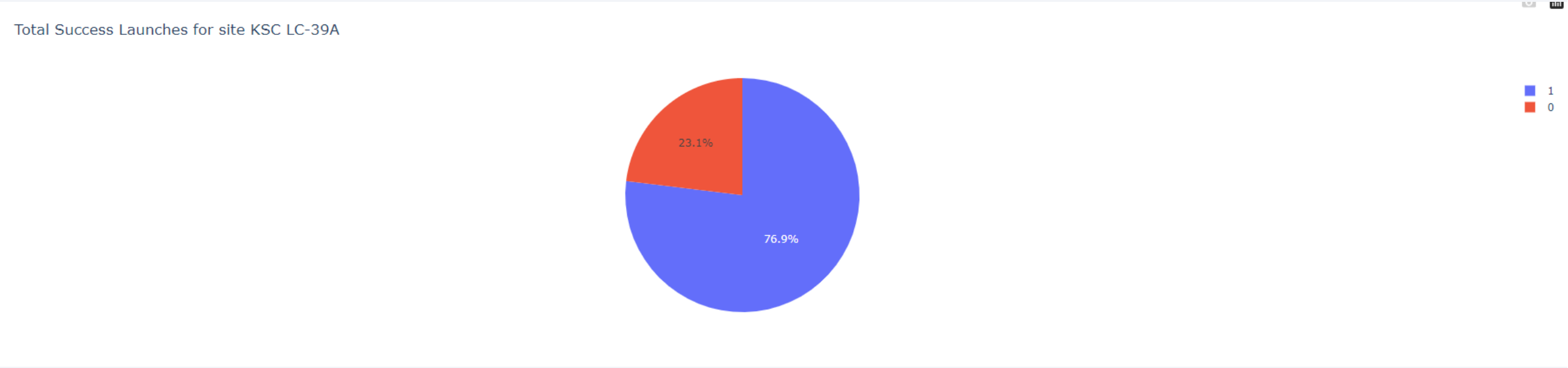



Section 4
Build a Dashboard with Plotly Dash

Dashboard Launch Success Count

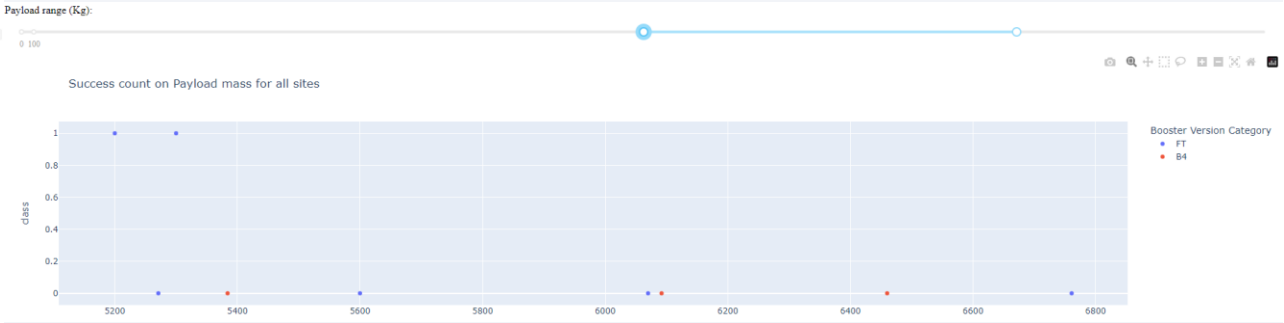
- The screenshot shows the launch success count for all sites

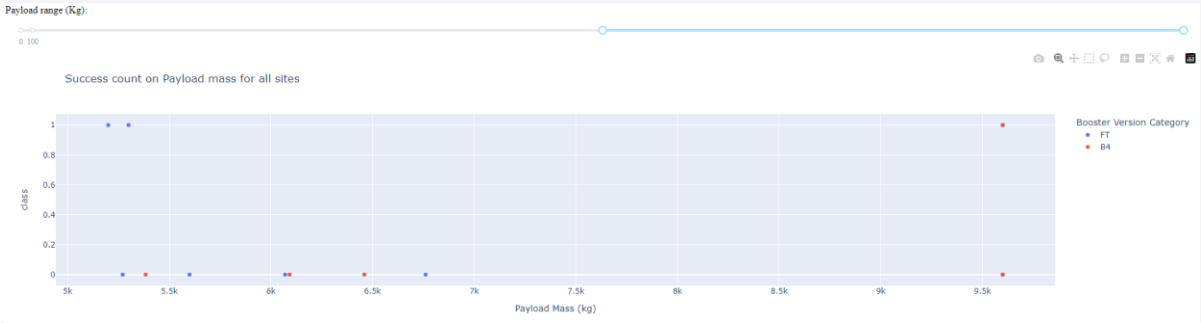
Dashboard Highest Launch Success Ratio



- The screenshot shows the launch site with the highest launch success ratio. The #1 or blue area means a successful launch.

Dashboard Payload vs. Launch Outcome



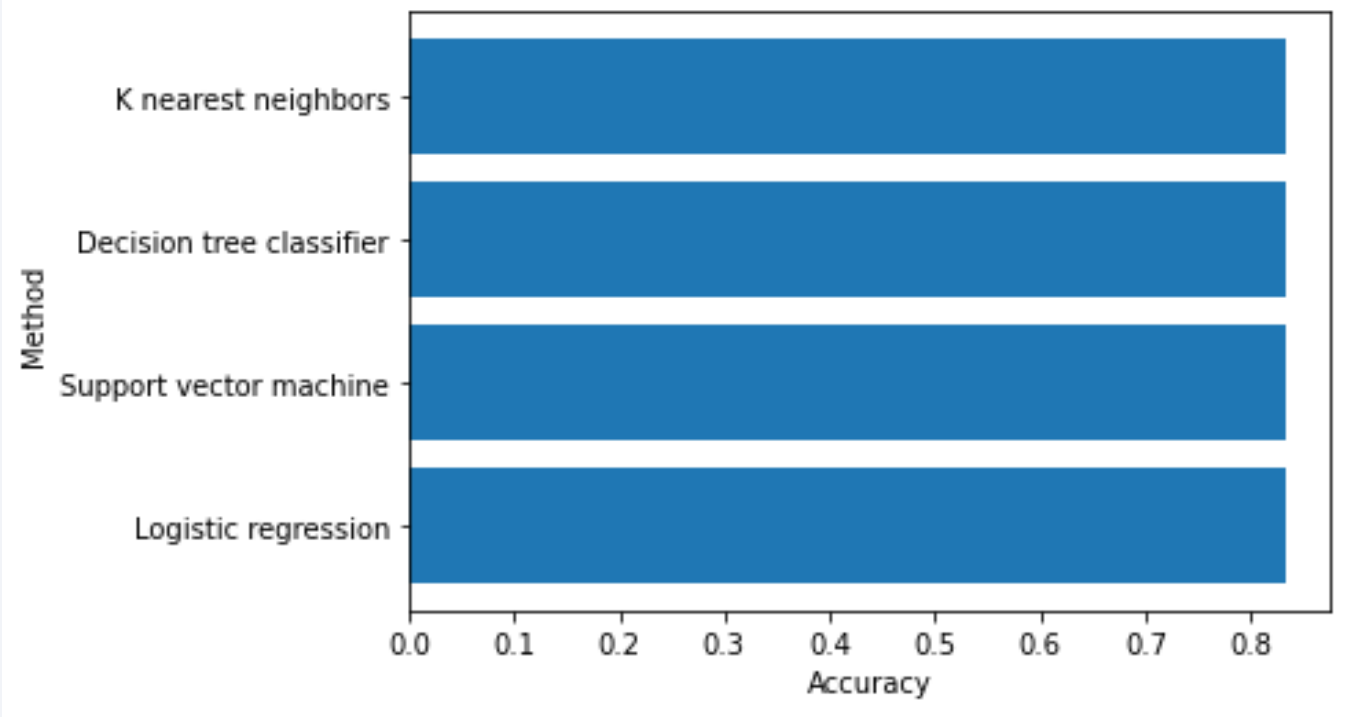


- These screenshots show the Payload vs. Launch Outcome scatter plots for all sites, with different payloads selected in the range slider.

Section 5

Predictive Analysis (Classification)

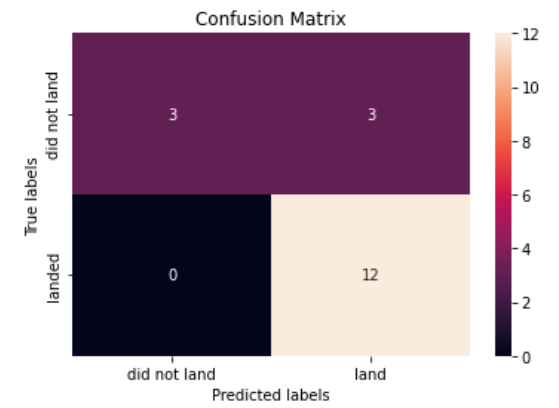
Classification Accuracy



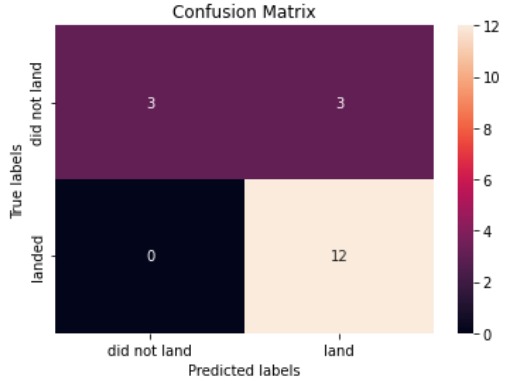
- Based on my prediction analysis, all methods contained the same accuracy of 83%. We can feel confident that our training and testing data can be used for accurate predictions for success.

Confusion Matrix

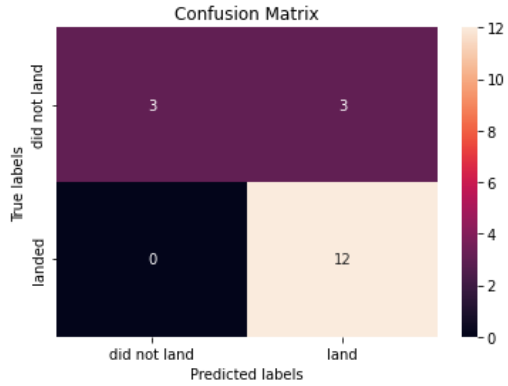
```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



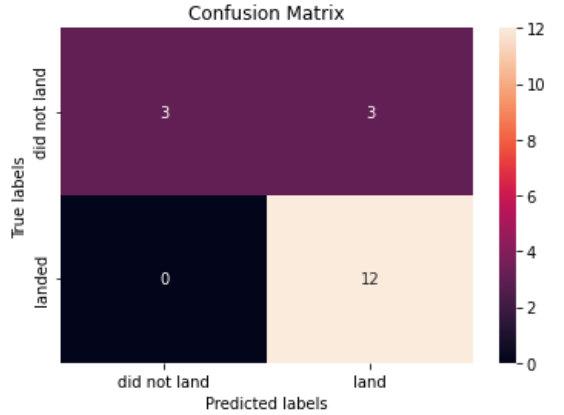

```
yhat=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
yhat = svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



- Since all models performed with the same accuracy, we have very similar confusion matrixes.

Conclusions

- As the number of flights increased, the success rate increased. The more rockets launched the better you will get at successful launches.
- The orbit types impact the success rate. SSO, HEO, GEO, and ES-L1 are preferred orbits for a successful launch.
- Booster version KSLC-39A is the best version. It had the highest number of successful launches and landings.
- The higher the payload the more of an impact it has on the success of the launch. Using maximum payloads is not advised.
- All predictive models achieved 83%. More data needs to be collected to ensure accurate predictions.

Thank you!