IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vivek Kumar Rai
9th July 2023

Outline

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

Executive Summary

- **Summary of methodologies:-**

➢ Data Collection using API and Web Scrapping.

➢ Data Wrangling using Pandas and NumPy.

➢ EDA with SQL and Seaborn.

➢ Building interactive map with Folium.

➢ Building dashboard application with Plotly Dash.

➢ Predictive Analytics by Classification methodologies.

- **Summary of all results:-**

➢ Results derived from EDA.

➢ Results derived from Folium Maps and Interactive Dashboard.

➢ Results derived from Predictive Analytics.

Introduction

- **Project background and context:-**

Space X 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 Space X 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 space X for a rocket launch.

- **Problems you want to find answers:-**

➢ What factors influence the successful landing of the first stage?

➢ The magnitude of effect each individual factor has on the successful landing of the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API.
  - Web Scrapping.
- Perform data wrangling
  - Data was processed using Pandas and NumPy.
- Perform exploratory data analysis (EDA) using Visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardizing the data, splitting, finding the best parameters and accuracy score.

Data Collection

The Data was collected from the following two sources:-

- Publicly available SpaceX REST API.
- Web Scrapping a Wikipedia web page titled – "List Of Falcon 9 and Falcon Heavy Launches".

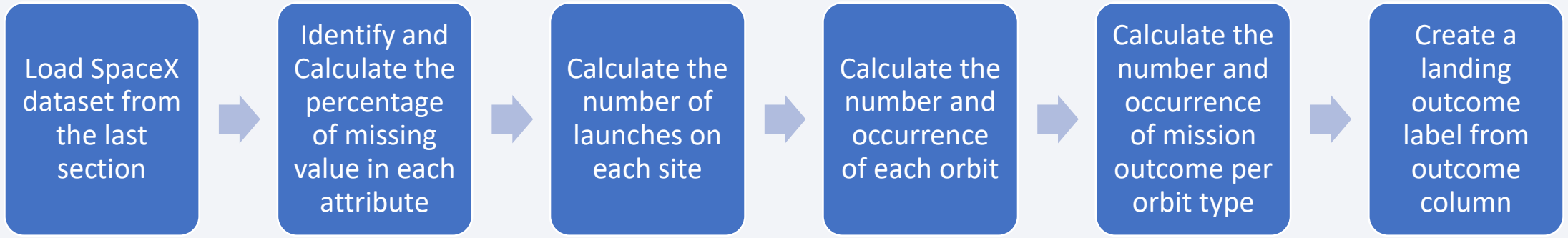
Data Collection – SpaceX API

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/jupyter-labs-webscraping.ipynb

Data Wrangling



GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

The following charts were plotted using Seaborn:-

- Scatter Plot – To observe relationship between variables.
- Bar Plot – To compare categorical data.
- Line Plot – To identify trends.

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

The following SQL Queries were performed:-

- Names of unique launch sites.
- 5 records were launch sites begin with string ‘CCA’.
- Total Payload mass carried by boosters launched by NASA(CRS).
- Average payload mass carried by booster F9 v1.1
- Date when first successful landing on Ground Pad was achieved.
- Boosters with success in drone ship and payload mass greater than 4000 and less than 6000.
- Total successful and failure mission outcomes.
- Booster versions which have carried the maximum payload mass.
- The month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months 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.

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

The following map objects were created on the Folium Map:-

- Circle Object – To add a highlighted circle area with a text label for a specific co-ordinate.
- Marker – To present data on map using its location co-ordinates.
- MousePosition – To get a co-ordinate of a point on mouse over.
- PolyLine – To draw a line between two co-ordinates.
- Marker Clusters – To represent many points having the same co-ordinates.

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/lab_jupyter_launch_site_location.jupyterlite.ipynb

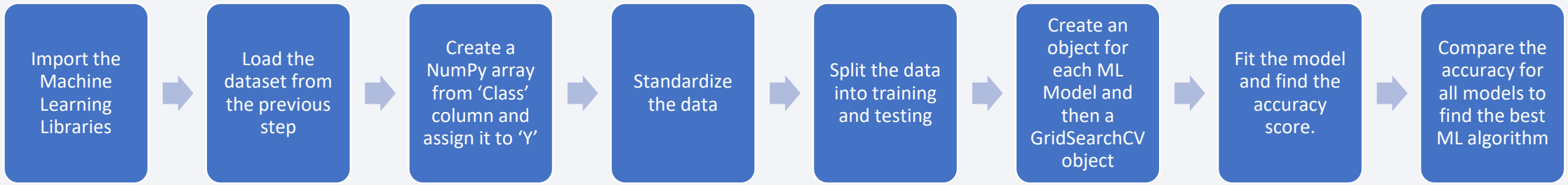
Build a Dashboard with Plotly Dash

The following graphs/interactions were added to the dashboard:-

- Dropdown list – To enable launch site selection.
- Pie Chart – To show successful launch count for all sites/Successful-Failed count if a site is selected.
- Slider – To select Payload range.
- Scatter Chart – To show correlation between payload and launch success.

GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/Plotly%20Dash%20App.txt

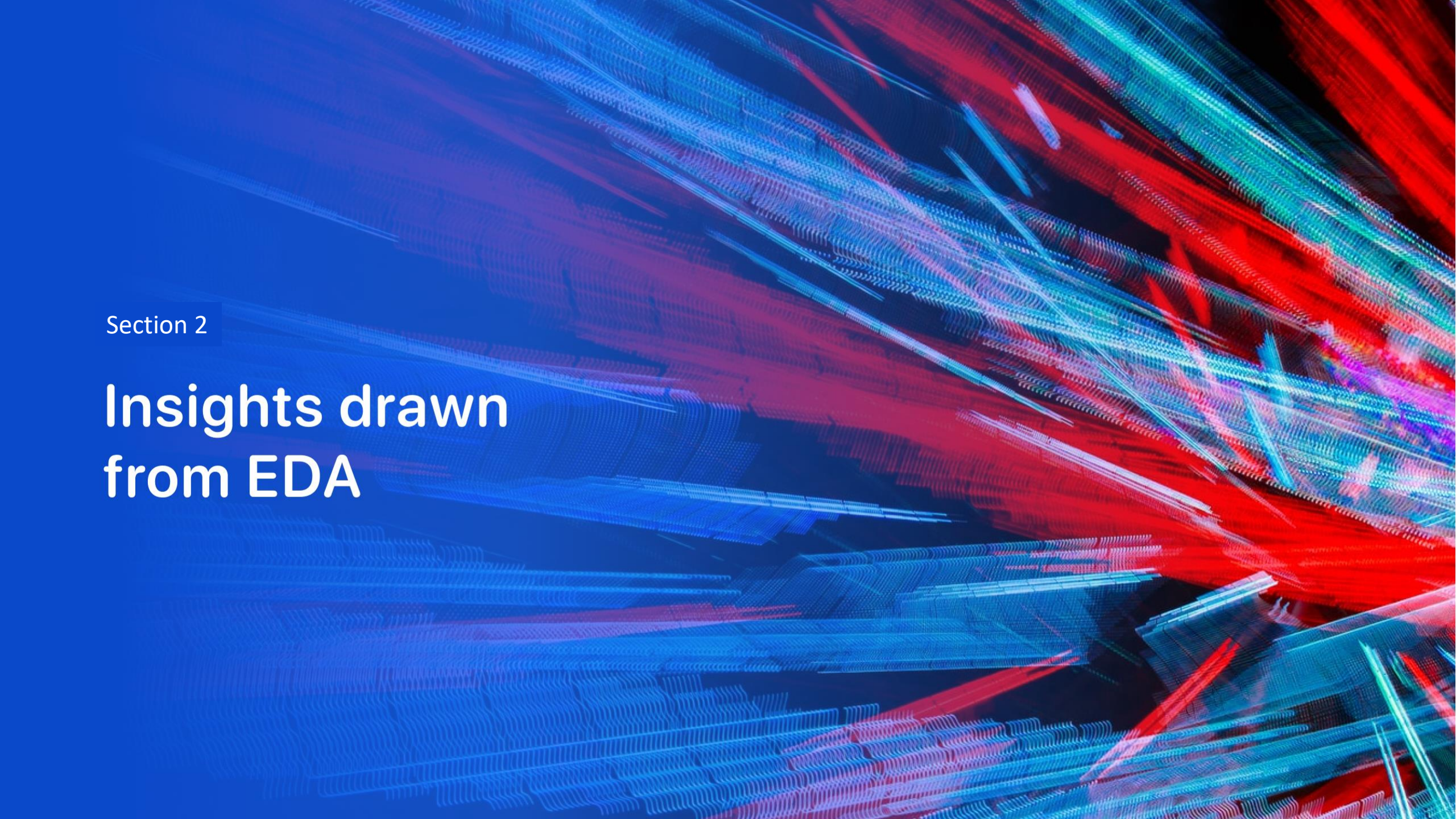
Predictive Analysis (Classification)



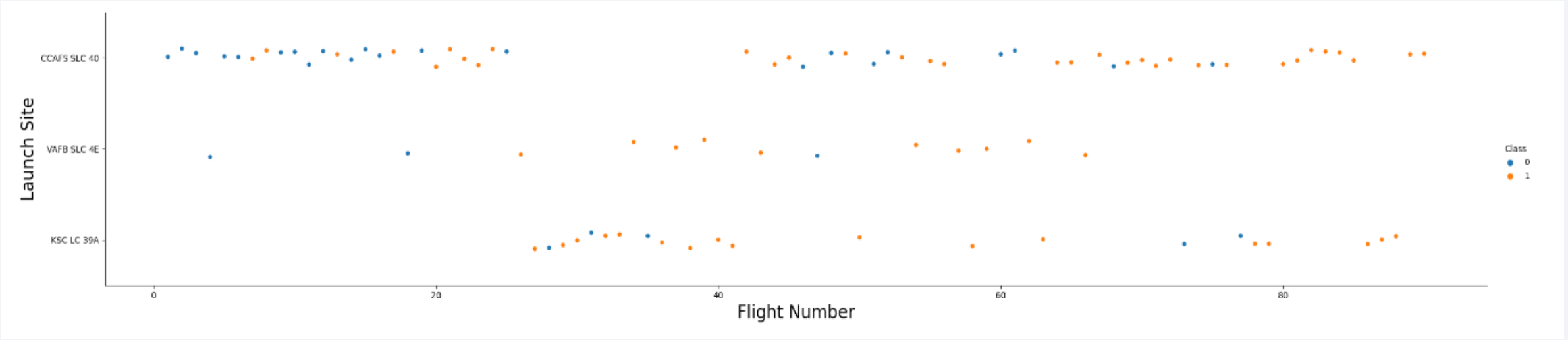
GitHub URL :- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/blob/main/Capstone%20Project/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

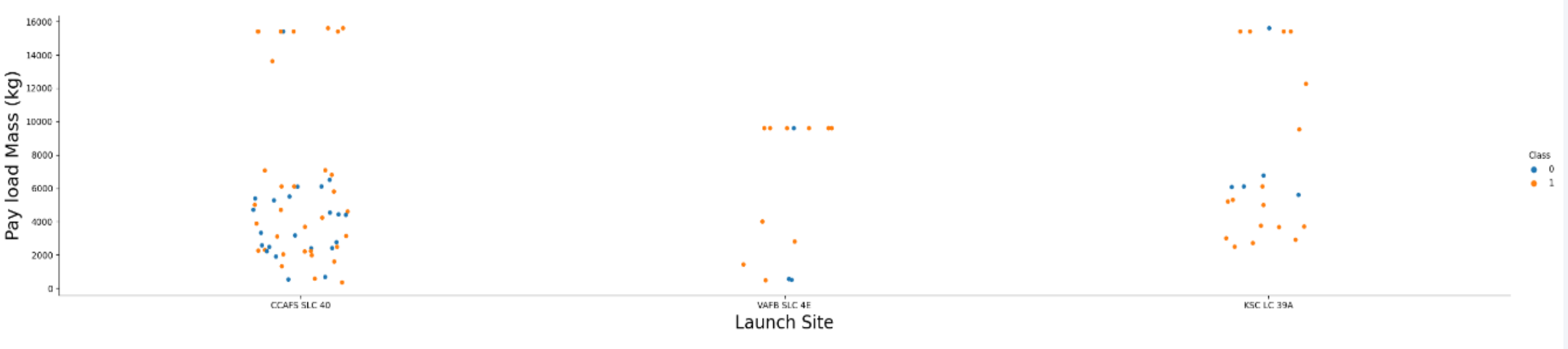

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



We can see that CCAFS SLC 40 has the highest number of launches and VAFB SLC 4E has the lowest among the three.

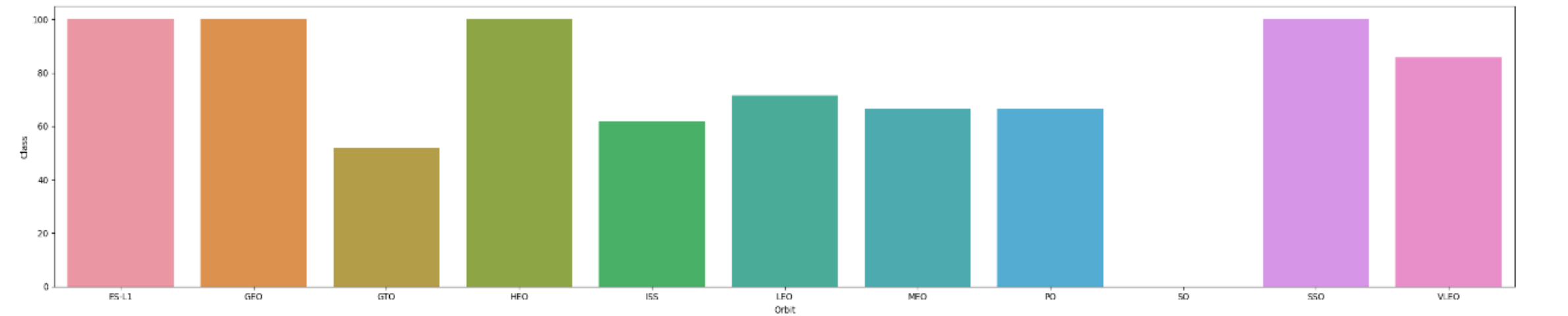
Payload vs. Launch Site



We can observe that –

- VAFB SLC4E has no launches for more than 10000 KG of payload mass.
- KSC LC 39A has a very high success ratio for payload mass greater than 9000 KG.

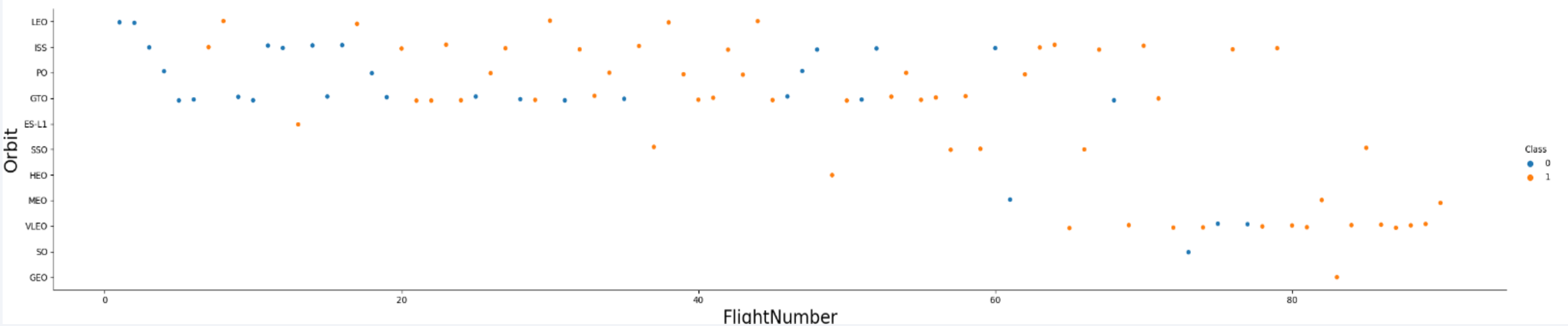
Success Rate vs. Orbit Type



We can observe the following –

- ES L1, GEO, HEO and SSO has success rate of 100%
- SO has 0% success rate.

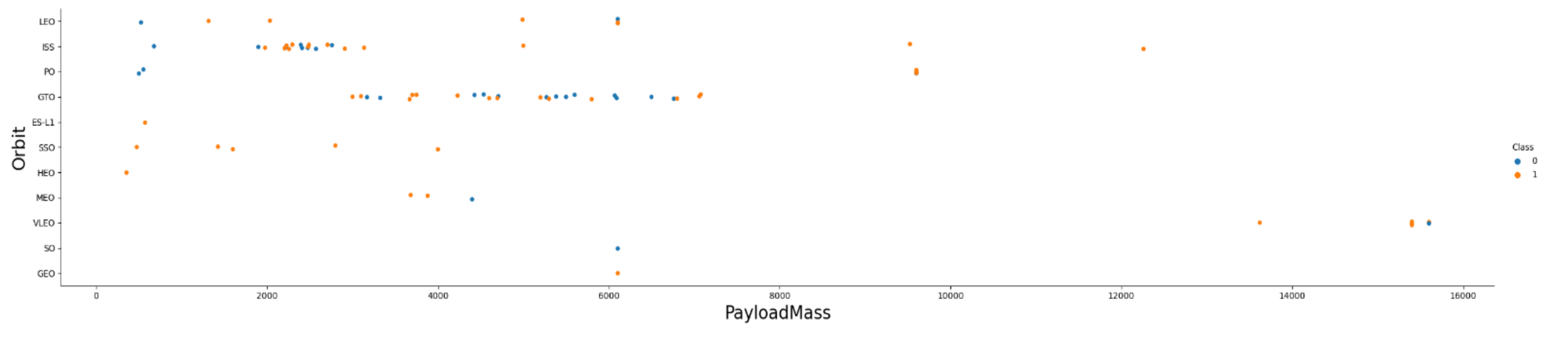
Flight Number vs. Orbit Type



We can observe the following –

- Most of the recent launches are into the VLEO orbit.

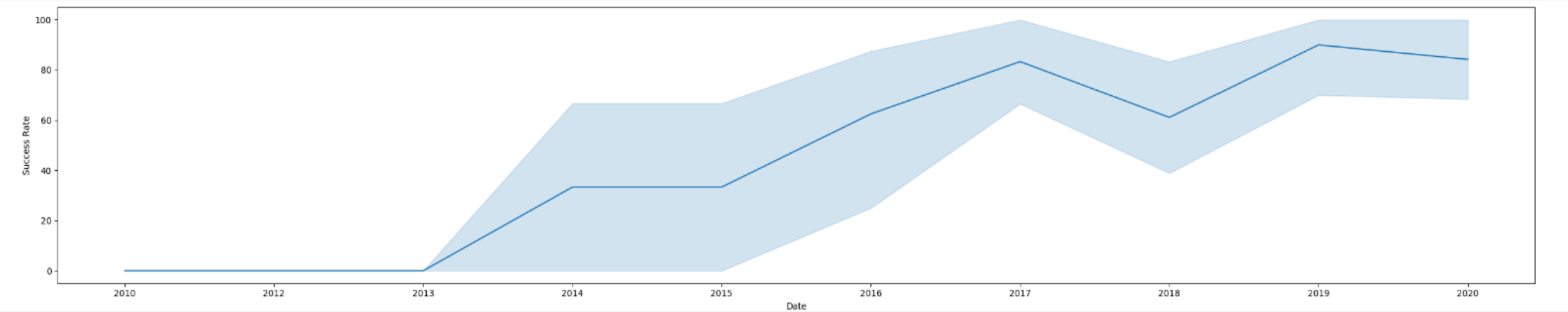
Payload vs. Orbit Type



We can observe the following –

- Payload mass greater than 13000 KG are launched in VLEO orbit.
- Except ISS, PO and VLEO all other orbits are used for Payload mass less than 7000 KG only.

Launch Success Yearly Trend



Launch success has been increasing steadily from 2013 to 2020.

All Launch Site Names

There are a total of 4 unique launch sites in our dataset.

Some rows are there which do not have the date available for the launch site.

```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

Here, we can see the top 5 rows from the dataset where the launch site name begins with ‘CCA’.

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (paracl
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (paracl
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No atte
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No atte
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No atte

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS SUM from SPACEXTBL WHERE CUSTOMER LIKE 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

SUM
45596.0

We can use the SUM aggregate function to find total Payload Mass for NASA(CRS).

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

AVERAGE
2534.6666666666665

We can use the AVG aggregate function to find average Payload Mass for F9 V 1.1 Booster Version.

First Successful Ground Landing Date

```
%sql SELECT MIN(DATE) AS DATE FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE 'SUCCESS'
```

```
 * sqlite:///my_data1.db
Done.
```

DATE
01/06/2014

We have used the MIN function on the Date attribute for first successful landing date.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE (LANDING_OUTCOME LIKE 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEE
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

We have 4 Booster Versions with successful landings on drone ship carrying payload mass between 4000 and 6000 KG.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS COUNT FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	COUNT
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

We have 1 Failure outcome, 100 Success outcome and 898 rows of data where the information is missing.

Boosters Carried Maximum Payload

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

We have 12 booster versions that carried the maximum payload mass.

2015 Launch Records

```
%sql SELECT substr(Date, 4, 2) as Month, landing_outcome, booster_version, launch_site from SPACEXTBL WHERE substr(Date,7,4)
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

We have two missions where there was a failure in landing on drone ship in the year 2015.

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

```
%sql select landing_outcome, count(*) AS Count1 from SPACEXTBL where Date >= '04-06-2010' AND Date <= '20-03-2017' GROUP BY
```

```
 * sqlite:///my_data1.db
Done.
```

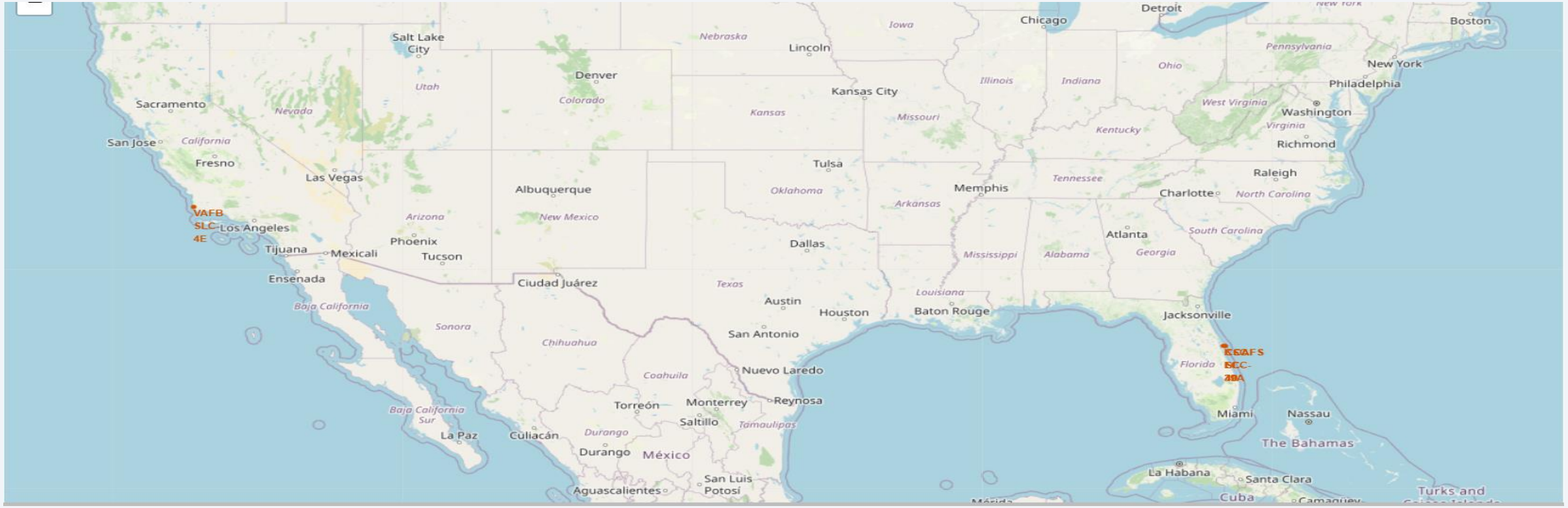
Landing_Outcome	Count1
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

We can see there are 35 successful landing outcomes between the given dates.

Section 3

Launch Sites Proximities Analysis

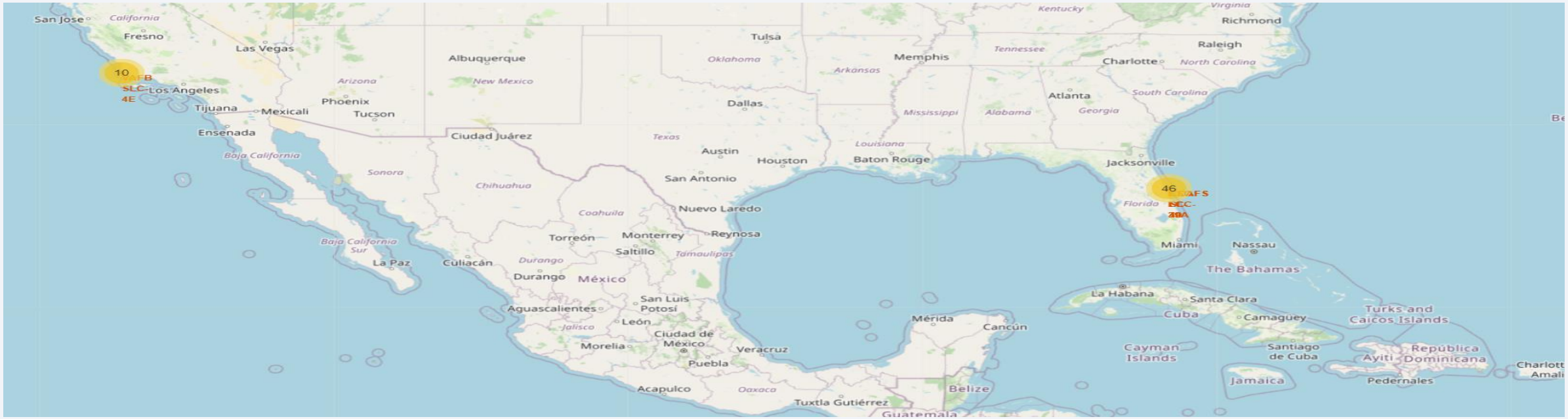
Launch Sites Map



We can make the following observations:-

- All launch sites are located on the coast.
- All launch sites are close the equator.

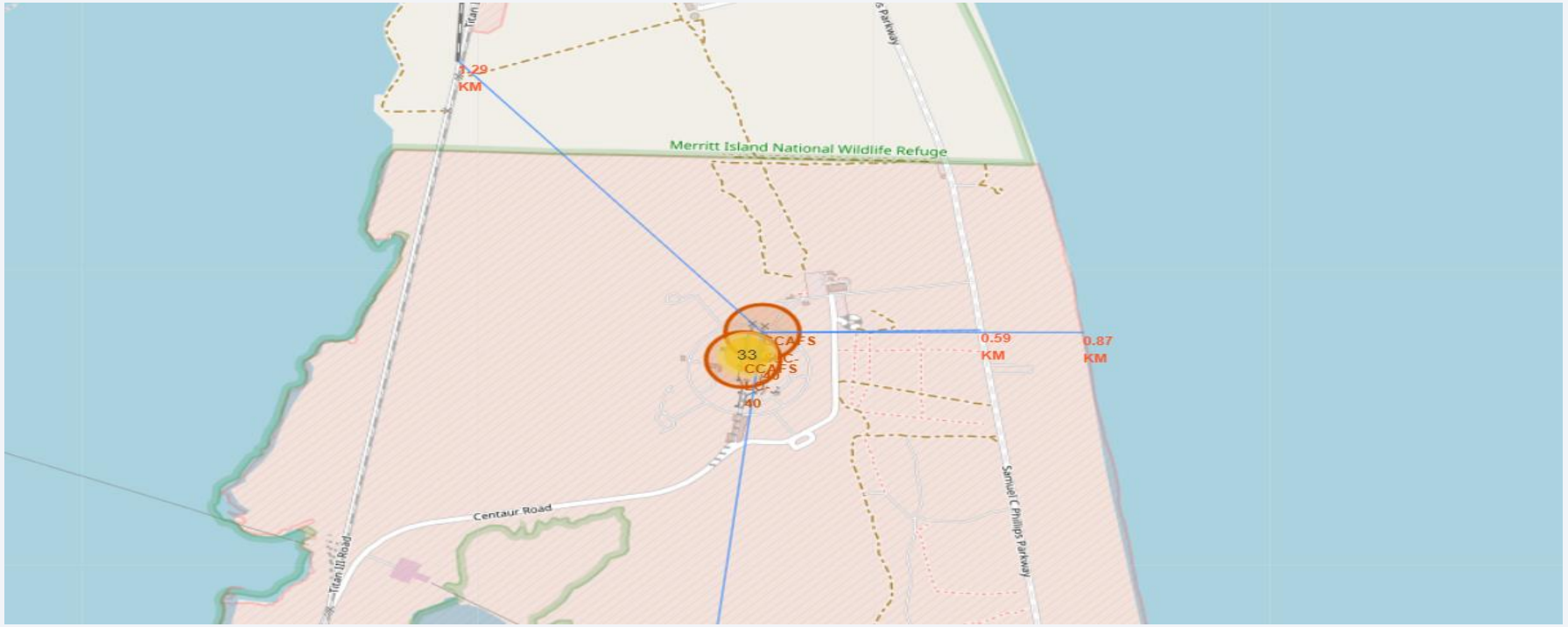
Launch Outcome Map



We can make the following observations:-

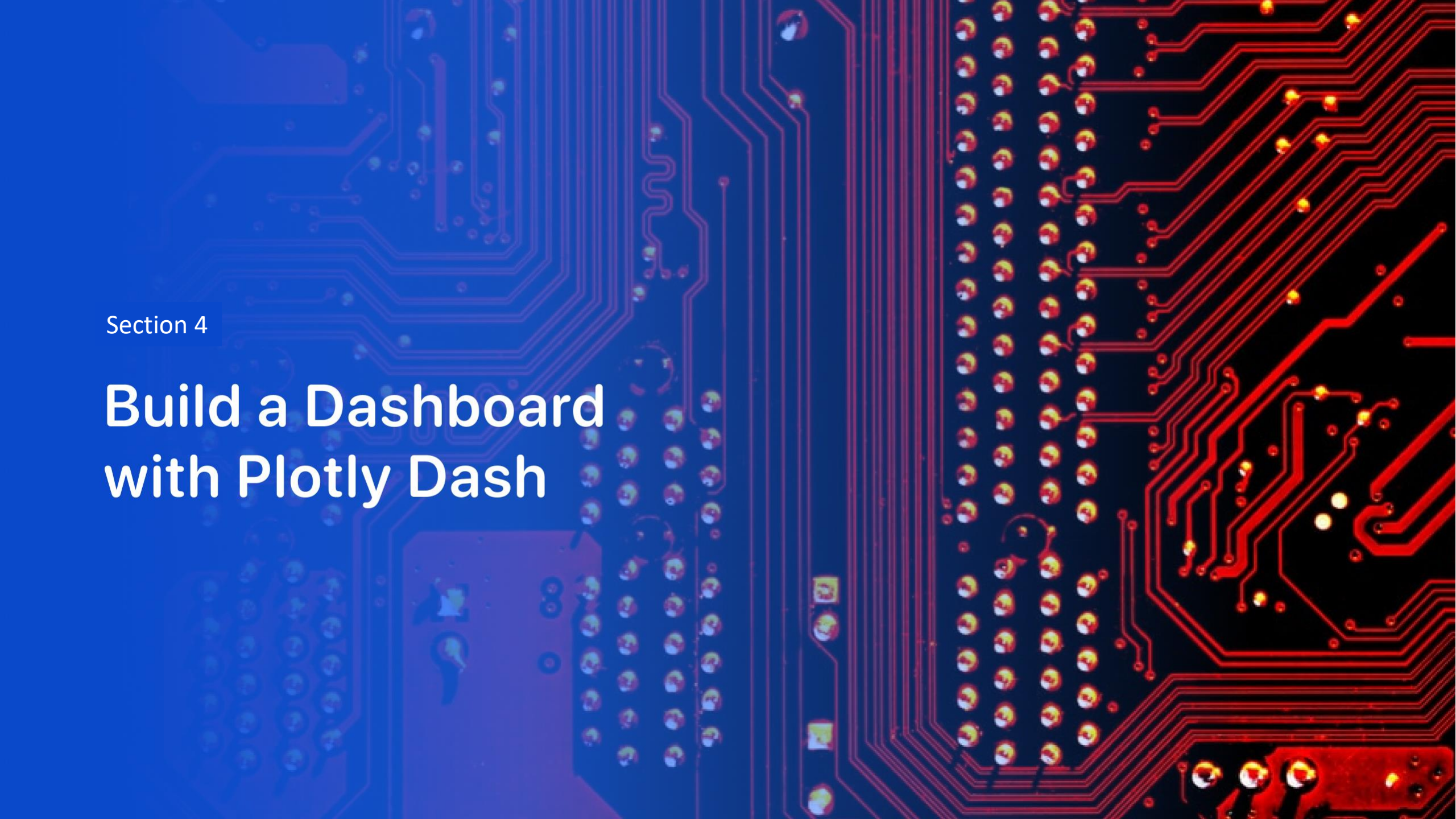
- The three launch sites have 33, 13 and 10 total launches, respectively.
- On zooming in the map, we can see the total successful and failed landing attempts.

Launch site proximity Map

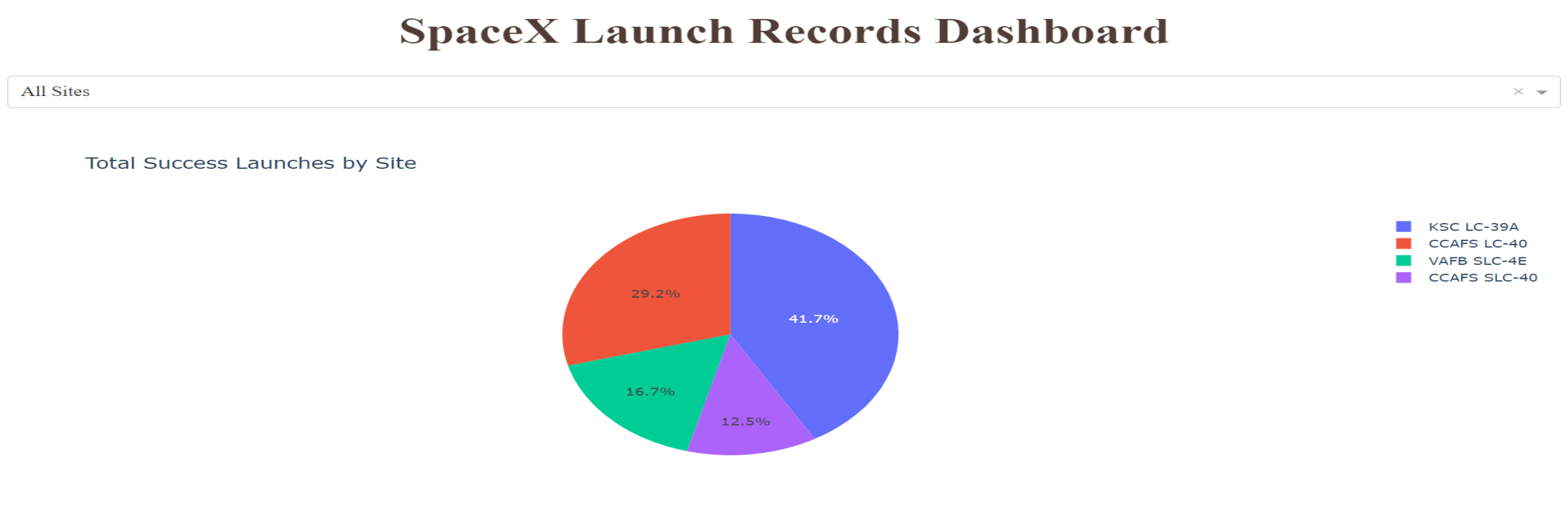


We can make the following observations –

- Launch sites are near coastline, railways and highways.
- Launch sites are always at a certain distance away from the cities.


Section 4
Build a Dashboard with Plotly Dash

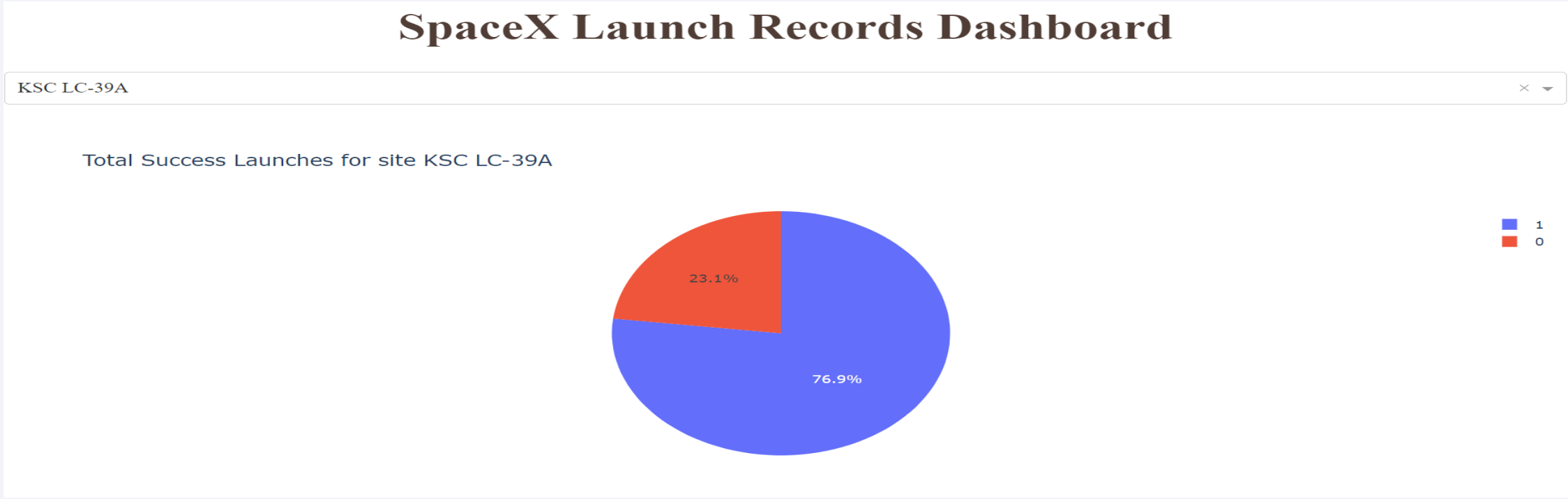
Launch Success Count by Launch Sites



We can observe that KSC LC-39A has the highest success contribution among all sites.

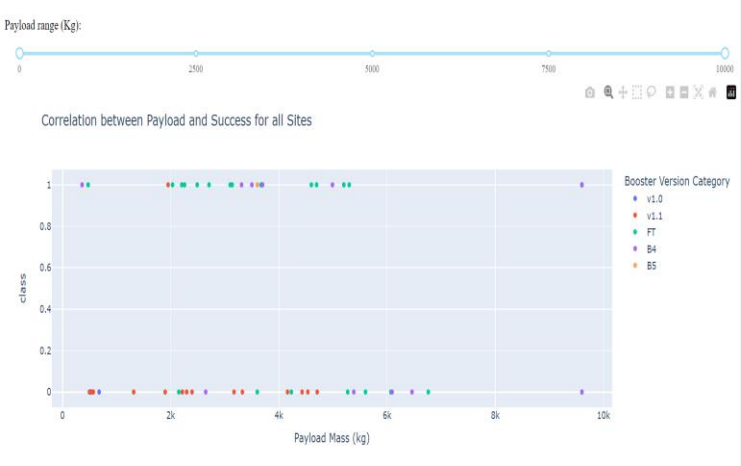
And CCAFS SLC-40 has the lowest contribution in successful launches.

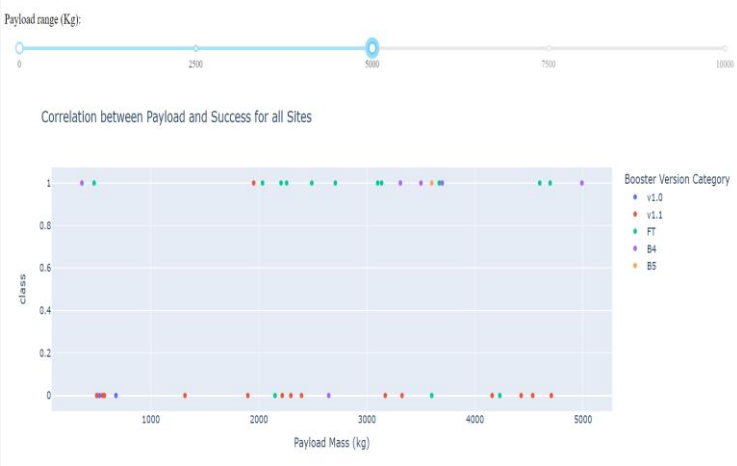
Launch Success Ratio for KSC LC-39A

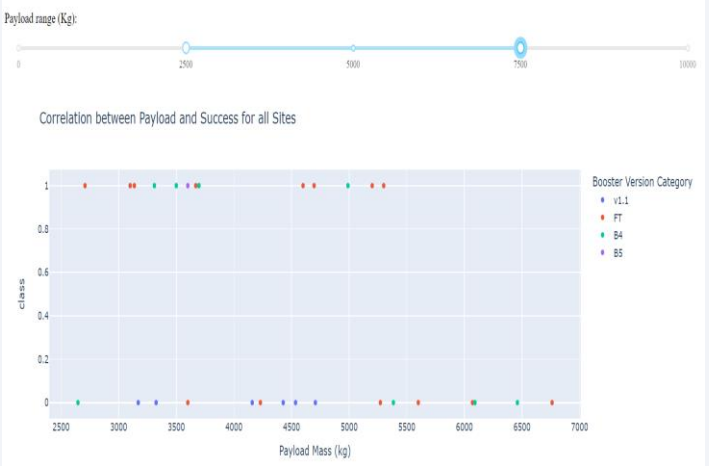


We can observe that nearly 77% of the launches from KSC LC-39A are successful.

Correlation between Payload and Success for Sites







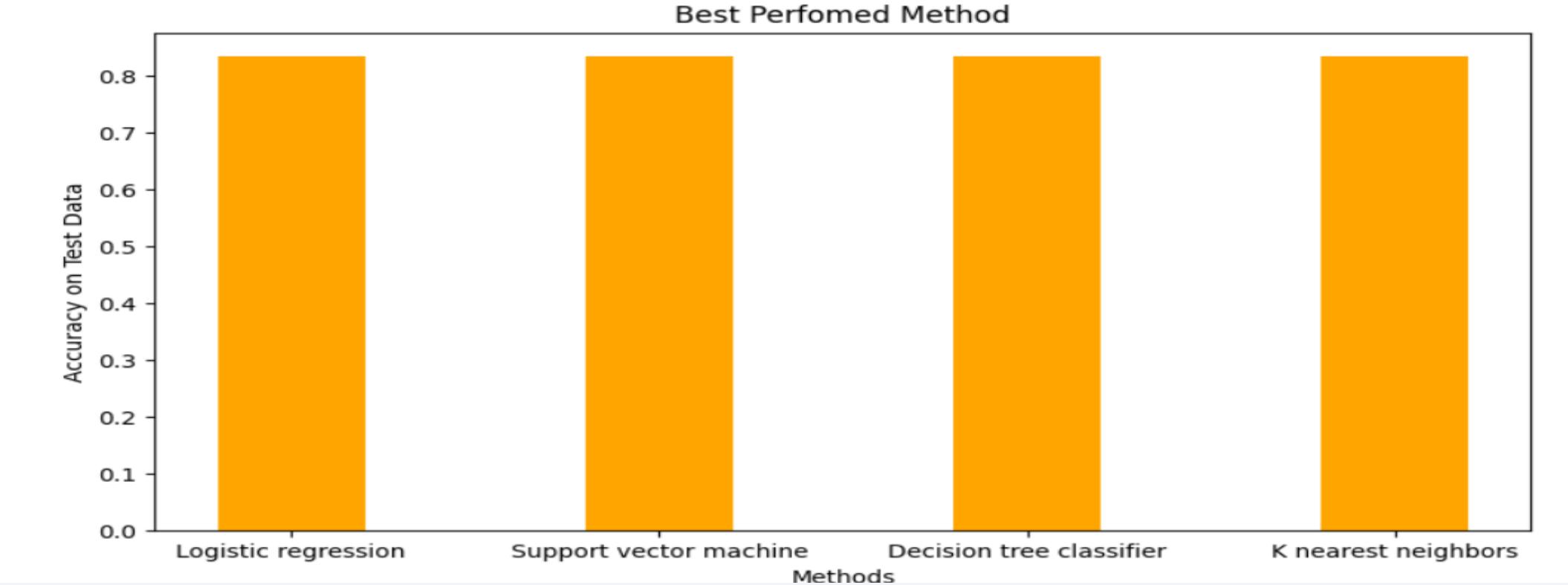
We can observe that –

- Booster Version FT has the highest success rate.
- Failure Rate is very high for Payload mass greater than 6000KG.

Section 5

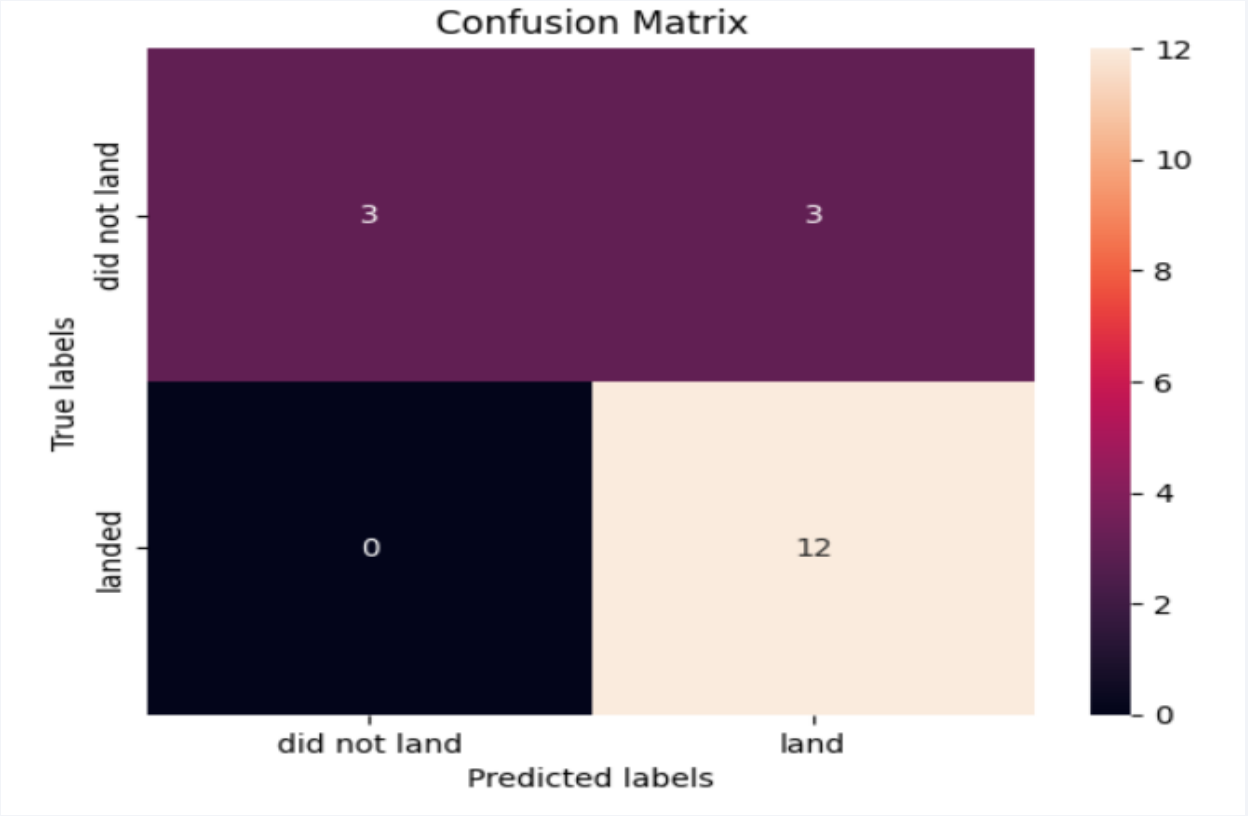
Predictive Analysis (Classification)

Classification Accuracy



All the four classification models have given the same accuracy on the test data.

Confusion Matrix



We can see we have 100% accuracy on the True Positive predictions, while it is 50% on the True Negative predictions.

Conclusions

- The rate of successful launches are steadily increasing with every year since 2013.
- Low weighted payloads have a higher success rate than heavier payloads.
- KSC LC-39A is the most successful launch site.
- ES L1, GEO, HEO and SSO orbits have 100% success rate.
- SO orbit has 0% success rate.

Appendix

Capstone Project GitHub URL:- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/tree/main/Capstone%20Project

IBM Data Science Professional Certificate GitHub Repository:- https://github.com/VivekKumar-Rai/Data-Science-IBM-Professional-Certificate/tree/main

Thank you!