IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vijayalakshmi Sivaraman
19.03.2024

Outline

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

Executive Summary

Summary of methodologies:

1. Data collection using API and web scrapping
2. Data Wrangling
3. Exploratory data analysis using SQL
4. Data Visualization
5. Interactive visual analysis using Folium
6. Creating a dashboard
7. Predictive Analysis using Machine Learning
8. Data summary and Presentation

Summary of Results:

1. Exploratory data analysis results
2. Interactive analytics demo
3. Predictive analysis results

Introduction

- As a data scientist of Space Y, the mail goal is to predict the cost of Launch for Space Y.
- Data from Space X launch of Falcon 9 is analyzed.
- Predictive analysis is made using machine learning.
- The results are provided in a dashboard for the stakeholder

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using API get requests and web scrapping from Wikipedia
- Perform data wrangling
  - Data was cleaned to remove null sets
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Linear regression, KNN, SVM, and decision tree models were created using GridSearchCV and analysis were made.

Data Collection – SPACEX API

- Data was collected using :
  - SpaceX REST-API requesting
  - Web scrapping
- GitHub Web link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

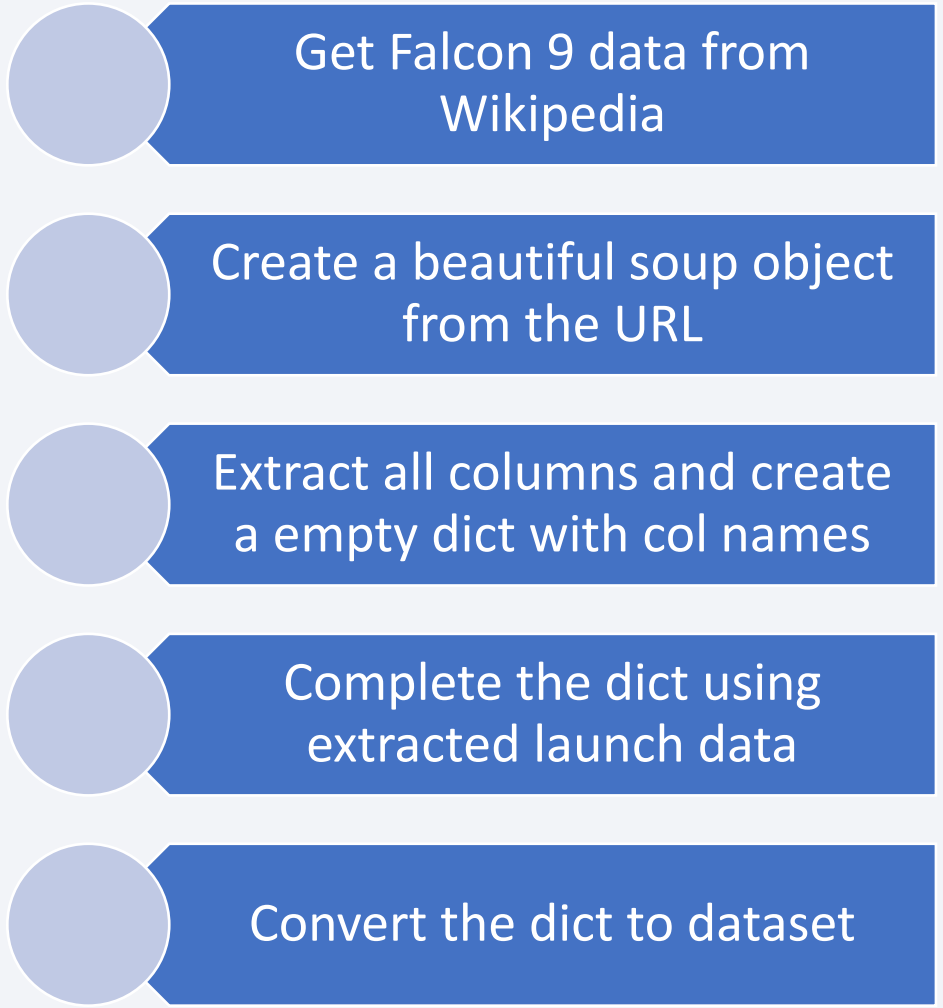
- GET request was sent to SPACEX API and the data obtained was parsed
- The response is decoded as a JSON file
- The JSON file is then changed to a Data Frame using pandas.
- Using web scrapping, the launch data is obtained from the given Wikipedia web ids.

Data Collection - Scraping

Data Scrapping:

- The steps involved in web scrapping is given by the flowchart.
- GitHub link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git



Data Wrangling

- Data Wrangling steps:
  - Calculated no of Launches in each site
  - Calculated the no of occurrence on each orbit
  - Calculated the occurrence number and occurrence of mission outcome per orbit type
  - Created a landing outcome label from outcome column
- GitHub Web link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

EDA with Data Visualization

- The plots plotted:
  - Scatter plot: Flight number vs payload mass, flight number vs launch site, payload mass and launch site, flight number and orbit type, payload mass and orbit type
  - Bar chart: success rate of orbits
  - Line graph : Launch success over the years
  - GitHub link: https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

EDA with SQL

- SQL queries performed:
  - 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 records which will display 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 link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

Build an Interactive Map with Folium

- Folium markers used to show the Launch sites
- Lunch sites were connected to nearest landmarks like railway lines, coastlines, cities and highways using polylines
- RED represented launch failures
- GREEN represented launch success.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
- GitHub link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

Build a Dashboard with Plotly Dash

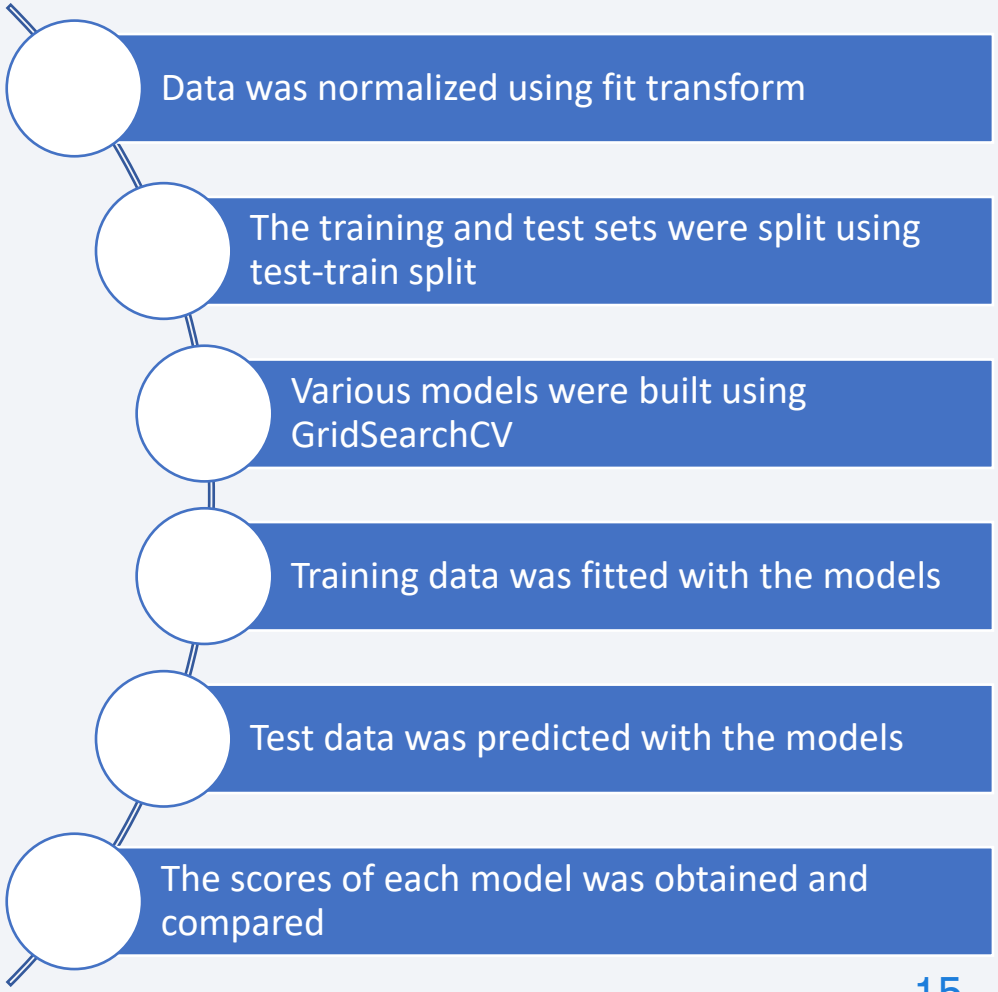
- Building a Dash board:
  - Pie chart: Success count for all launch sites. All launch sites or particular launch sites can be chosen.
  - Scatter plot: Success count of payload mass for all launch sites. Payload mass can be adjusted using a Payload range slider
- GitHub link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git

Predictive Analysis (Classification)

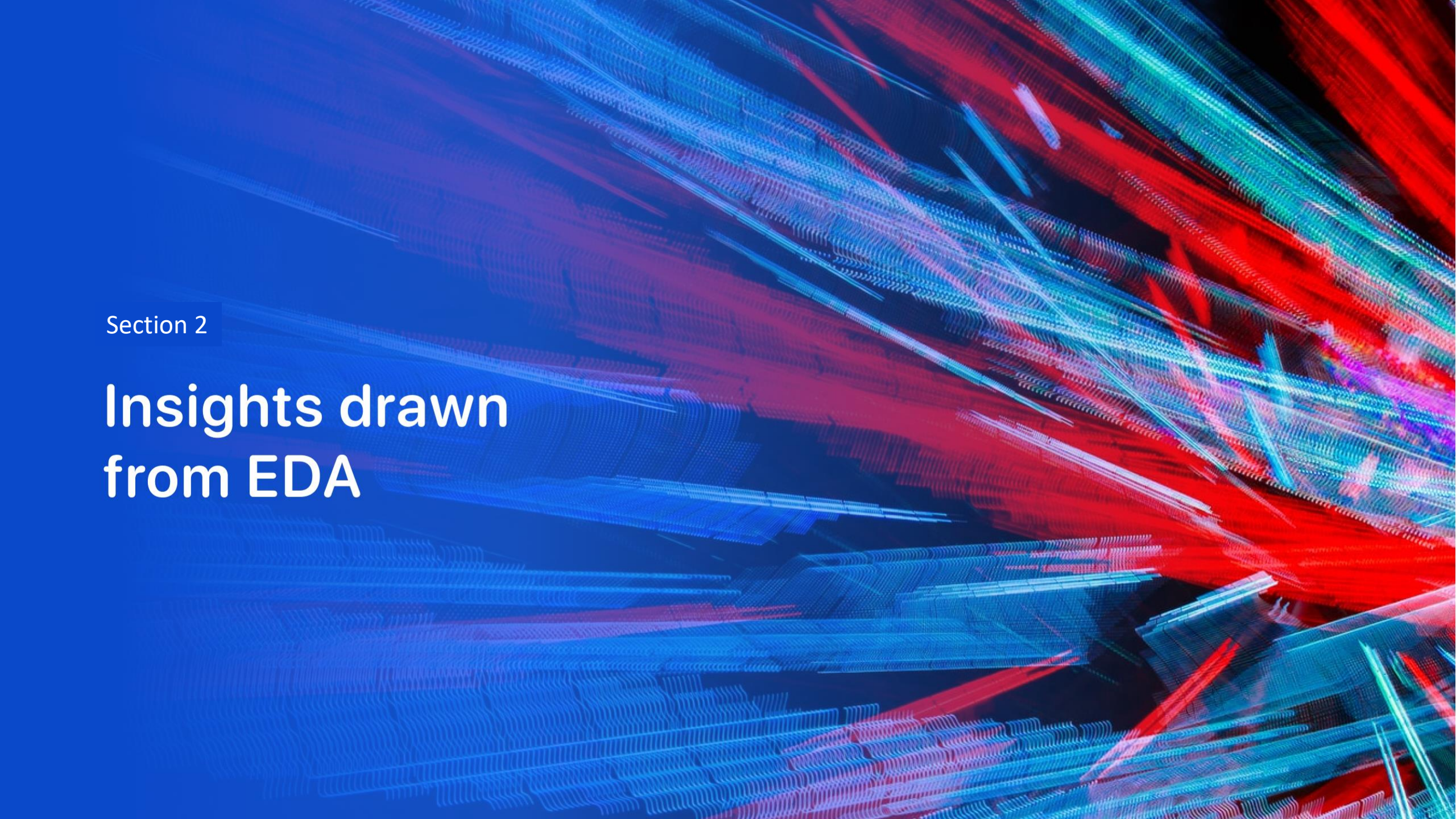
- Predictive Analysis using Machine learning:
  - In order to identify the best model, various models were trained and the scores were compared.
  - Scikit-learn python package was used
- GitHub link:

https://github.com/Vijayalakshmi2411/Final-capstone-project-for-DS-with-PY-course.git



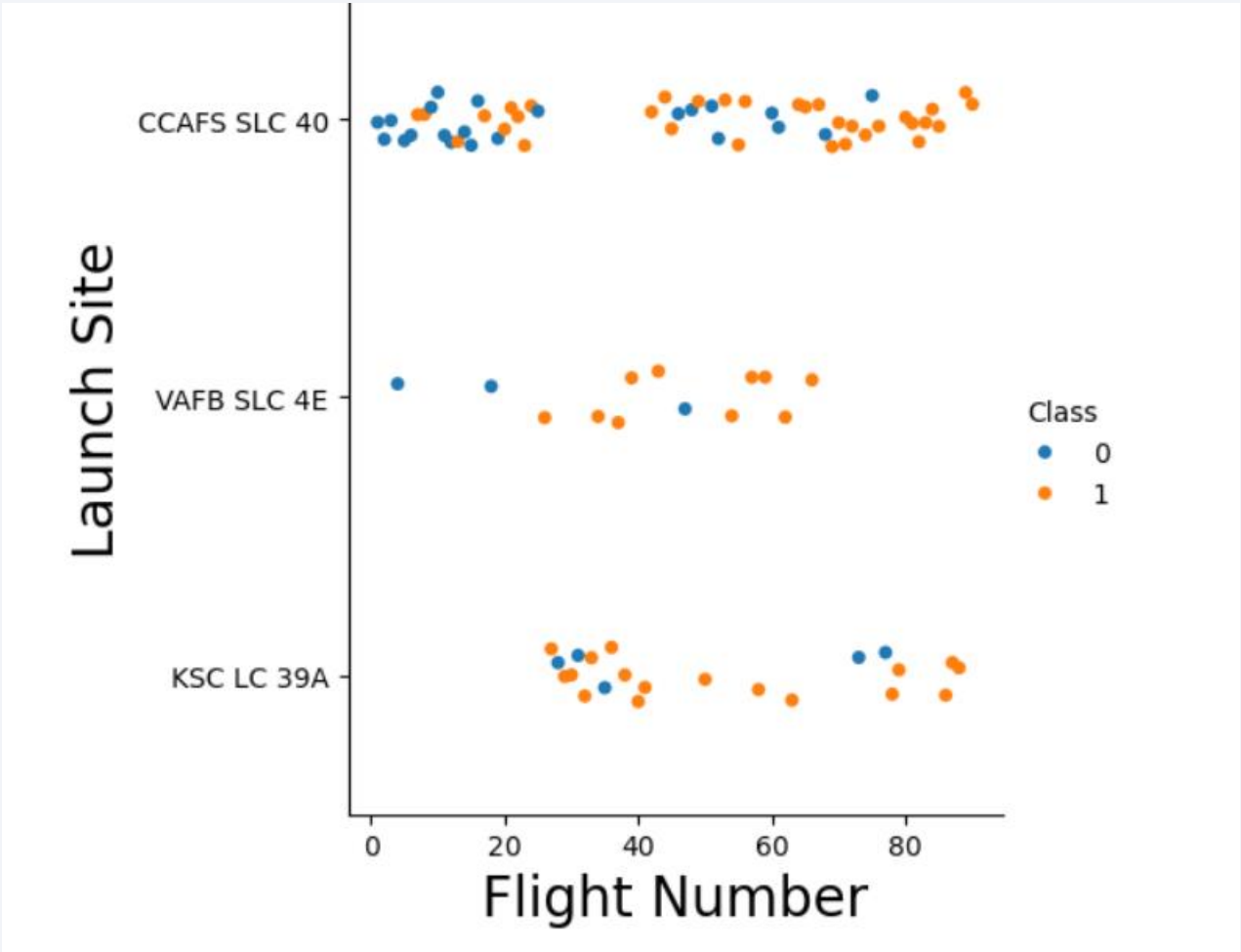
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

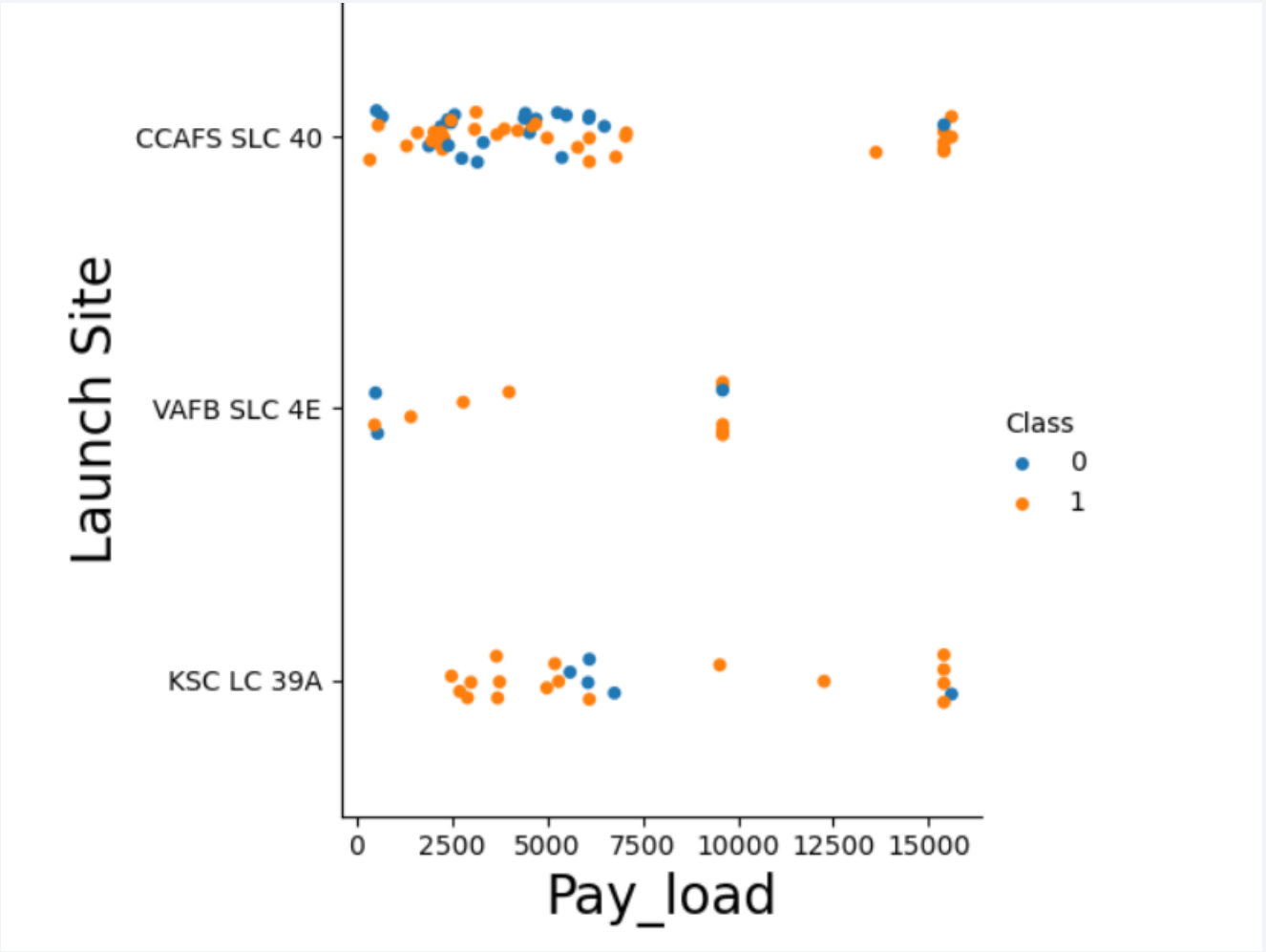
- A scatter plot of Flight Number vs. Launch Site is shown
- Increase in flight no resulted to increased success of the launch site **CCAFS SLC -40**.
- But the same cannot be inferred for others.



Payload vs. Launch Site

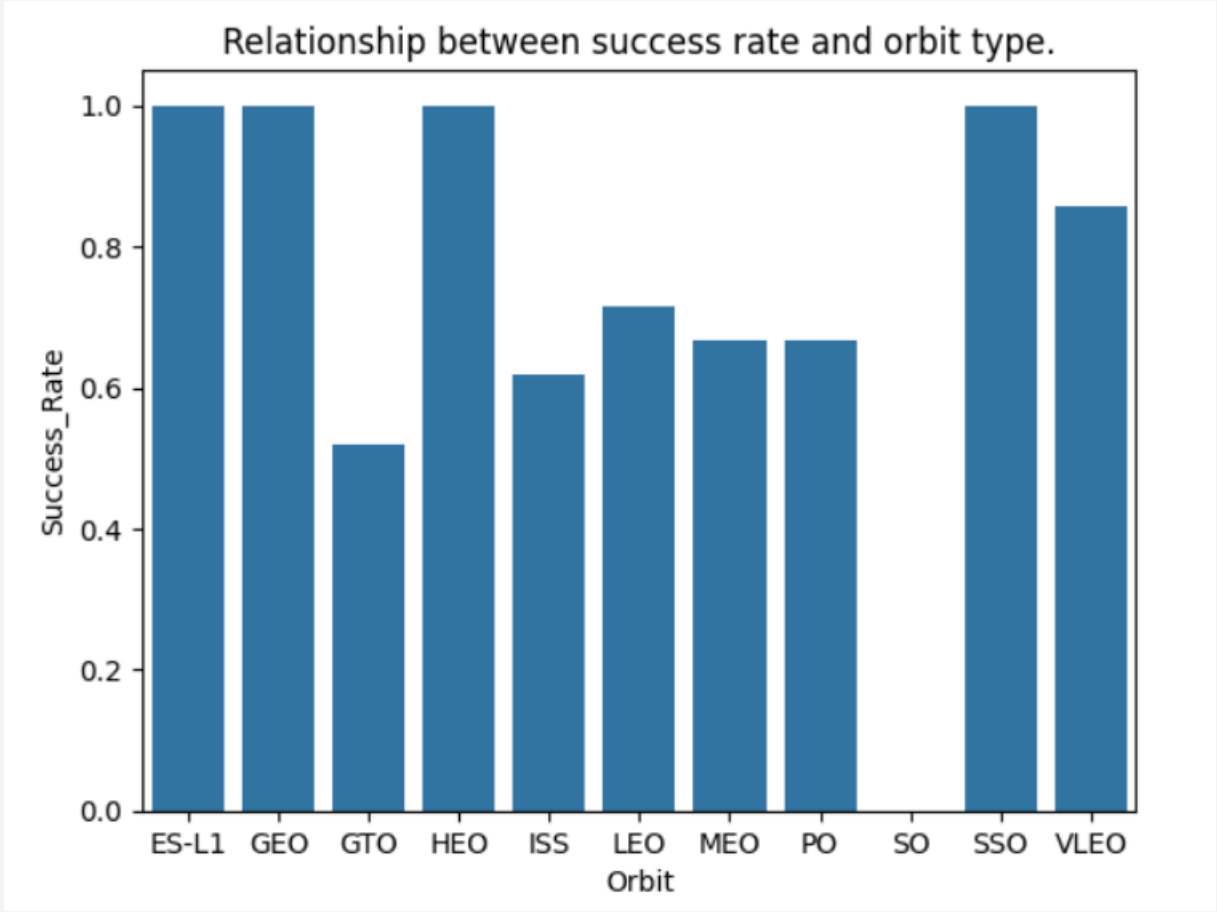
- A scatter plot of Payload vs. Launch Site is shown

- If you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



Success Rate vs. Orbit Type

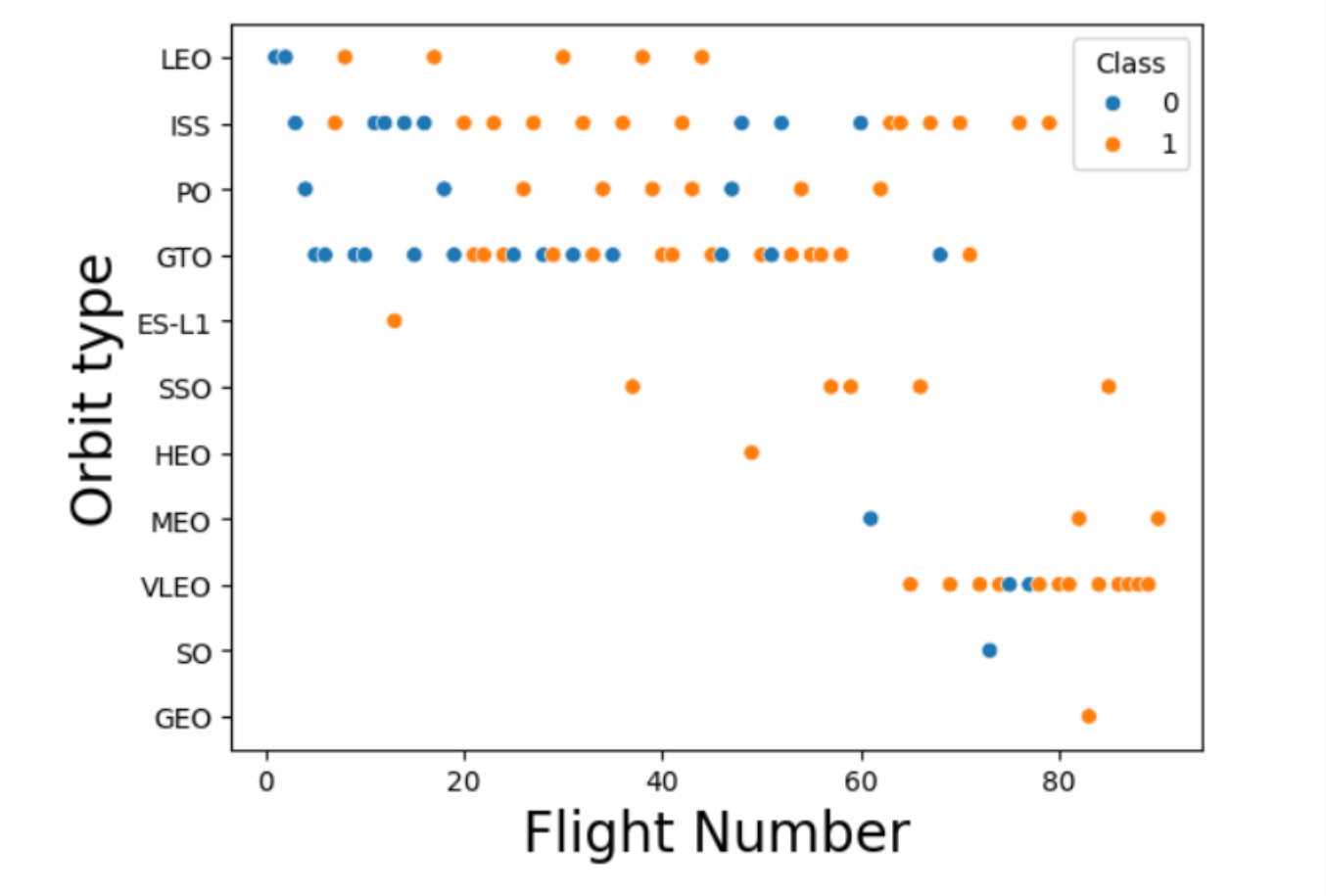
- A bar chart for the success rate of each orbit type is shown
- The Success rate of the orbits **ES-L1**, **GEO**, **HEO** and **SSO** are 100 %



Flight Number vs. Orbit Type

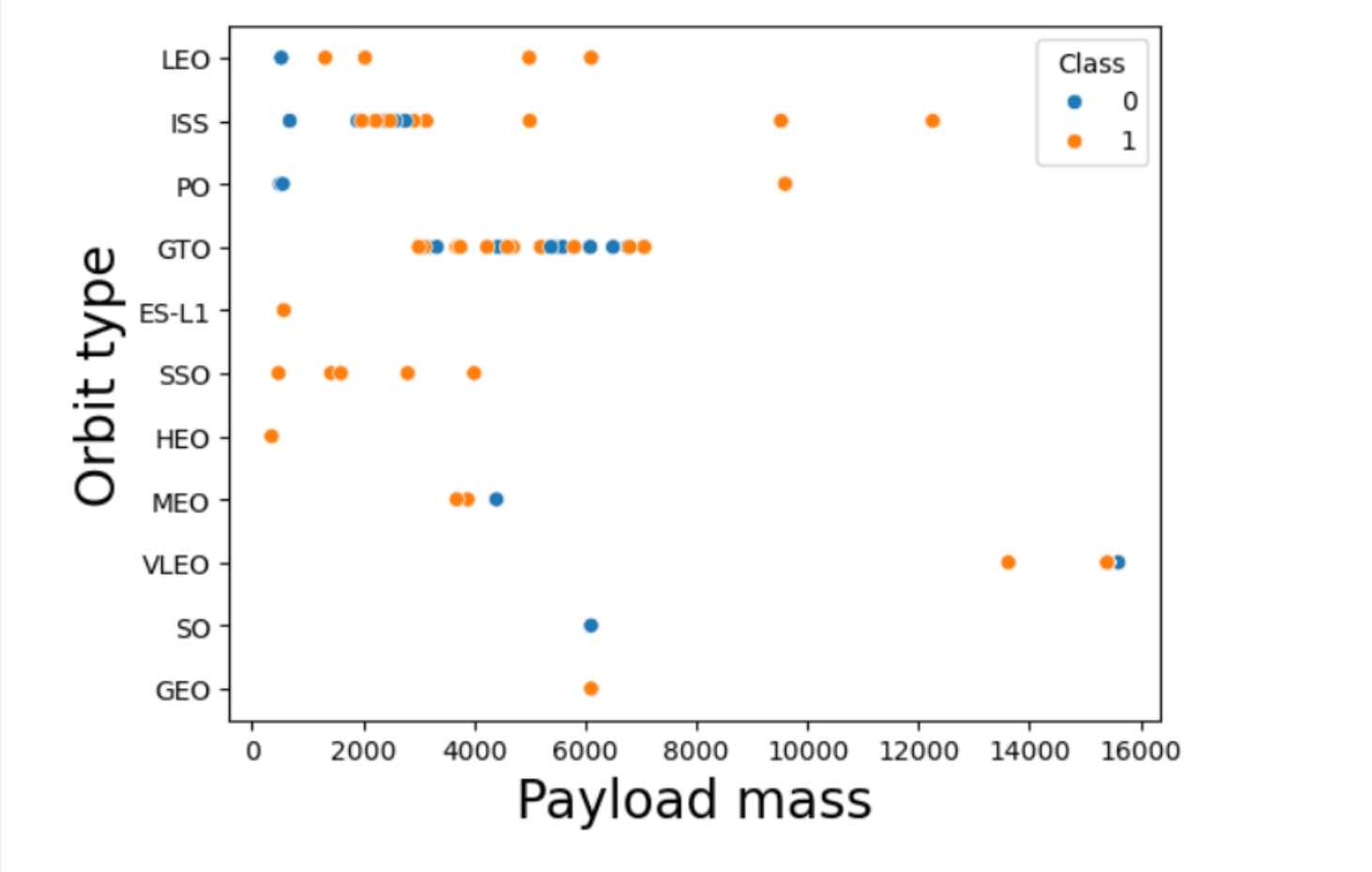
- A scatter point of Flight number vs. Orbit type is shown

- 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.



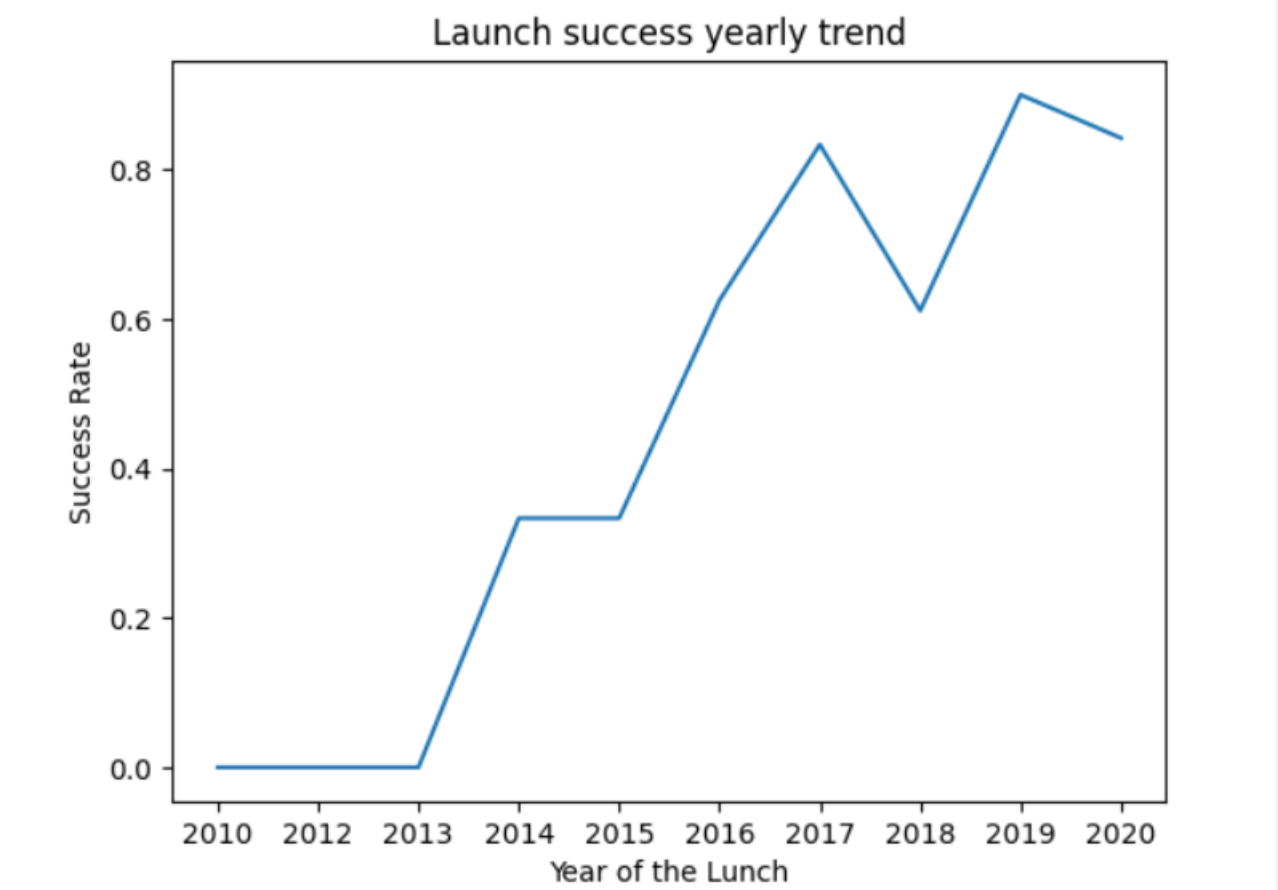
Payload vs. Orbit Type

- Ascatter point of payload vs. orbit type is shown.
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.



Launch Success Yearly Trend

- A line chart of yearly average success rate is shown.
- The launch success increased since 2013.



All Launch Site Names

Launch Site:
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

The above are the different launch sites from where falcon 9 was launched.

```
%sql SELECT Distinct Launch_Site from  SPACEXTABLE
```

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

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

Launch Site Names Begin with 'CCA'

Launch Site

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

- The above five shows launch sites that start with CAA.

```
%sql SELECT Launch_Site from  SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- The total payload mass was found to be **45596** kgs.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) from SPACEXTABLE where customer ='NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass was found to be **2928.4**

```
%%sql SELECT AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE
WHERE Booster_Version = 'F9 v1.1';
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The first successful ground landing was on **22/12/2015.**

```
%%sql
SELECT MIN(Date)
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
```

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- F9 FT B1021.1
- F9 FT B1022
- F9 FT B1023.1
- F9 FT B1026
- F9 FT B1029.1
- F9 FT B1021.2
- F9 FT B1029.2
- F9 FT B1036.1
- F9 FT B1038.1
- F9 B4 B1041.1
- F9 FT B1031.2
- F9 B4 B1042.1
- F9 B4 B1045.1
- F9 B5 B1046.1

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (drone ship)'
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

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

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes was **101.**

Boosters Carried Maximum Payload

- Booster version that carried maximum payload:
- 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

```
%%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

2015 Launch Records

- The failure outcomes:
  - Month : Jan, April
  - Launch site:
    - CCAFS LC-40
  - Booster Version:
    - F9v1.1B1012
    - F9v1.1B1015

```
%%sql select substr(Date,6,2) as month,LANDING_OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXTBL
WHERE substr(Date,0,5)='2015' AND LANDING_OUTCOME='Failure (drone ship)'
```

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

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

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

- Landing outcome:

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

```
%%sql
SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

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

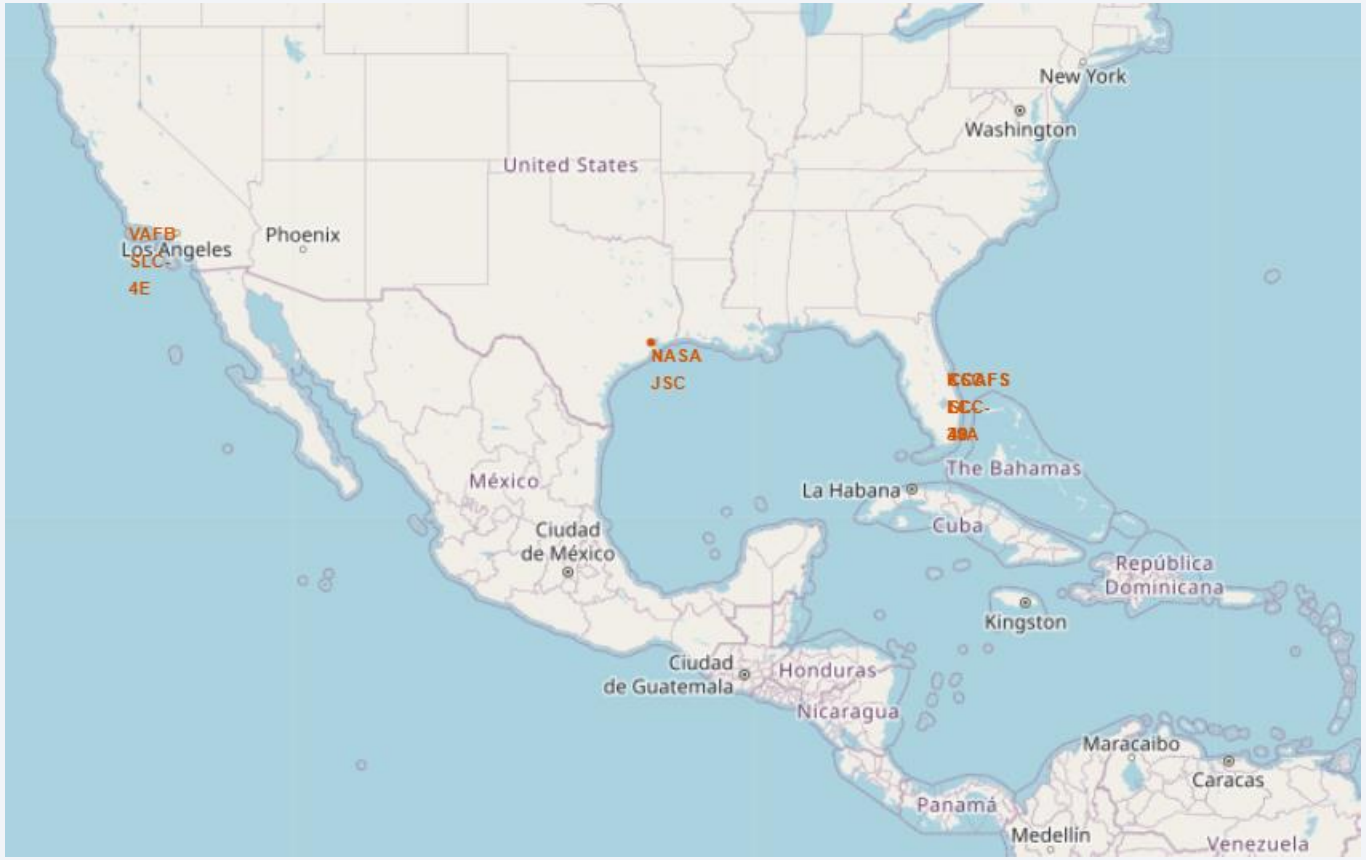
Landing_Outcome	TOTAL_NUMBER
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

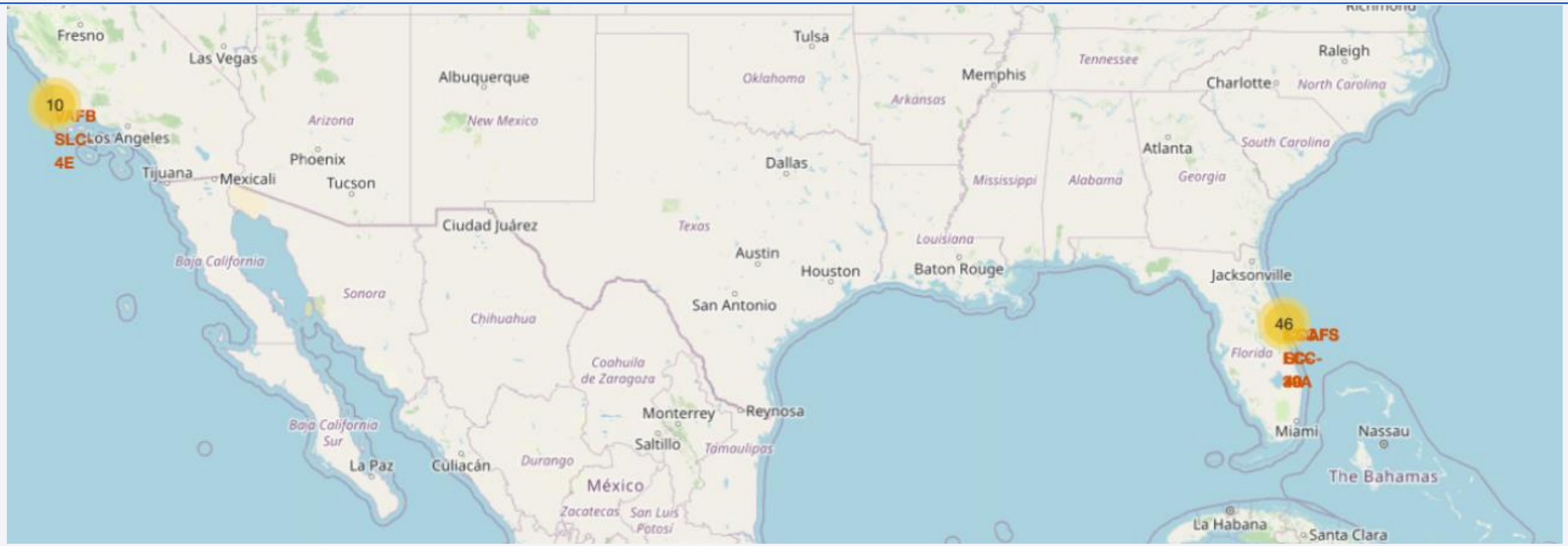
Launch Sites Proximities Analysis

Map with Launch sites

- In the given map the launch sites are marked with appropriate markers.



Success and Failure of Launch site



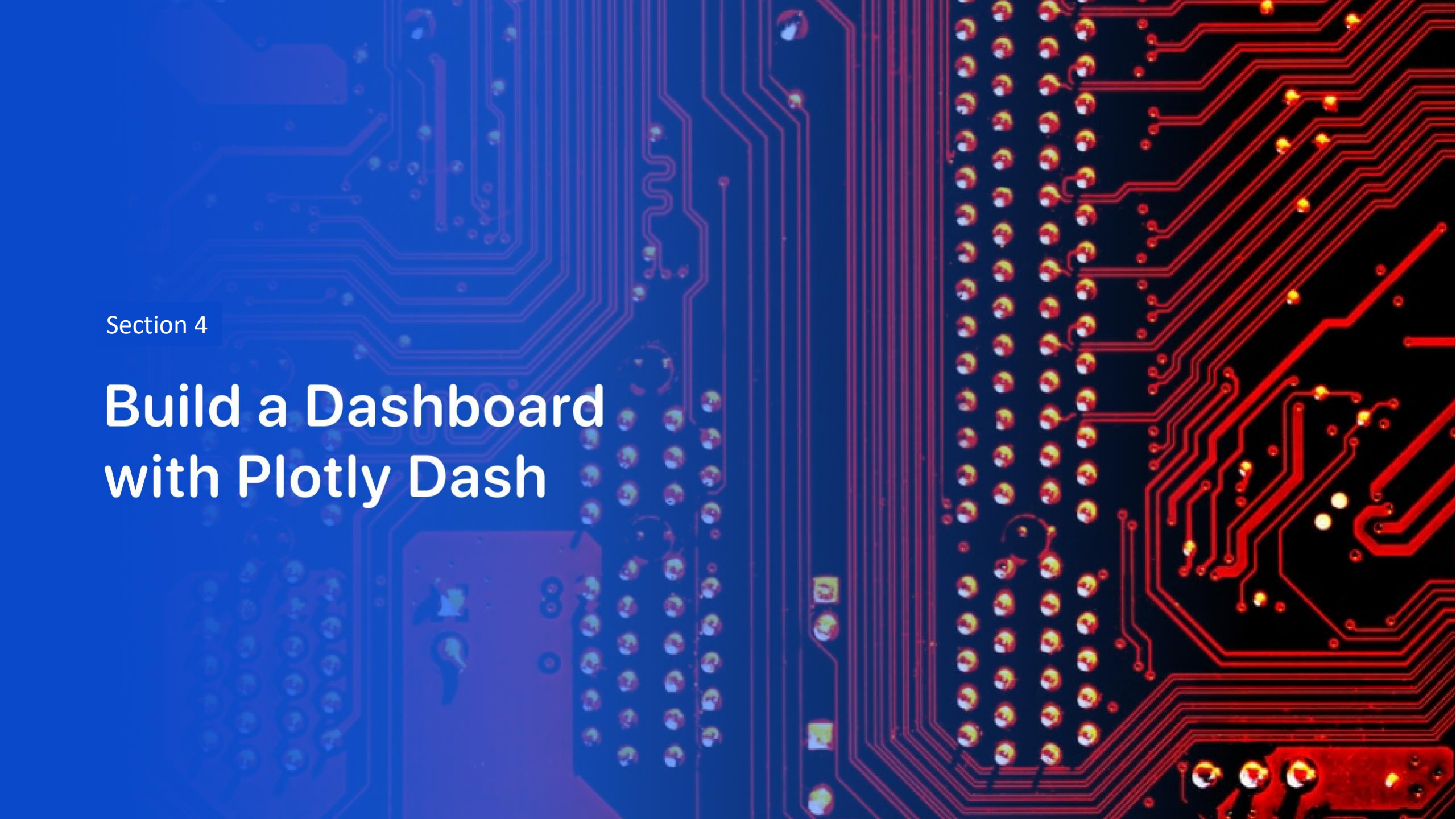
- The success and failures of each launch site is shown

Distance calculation

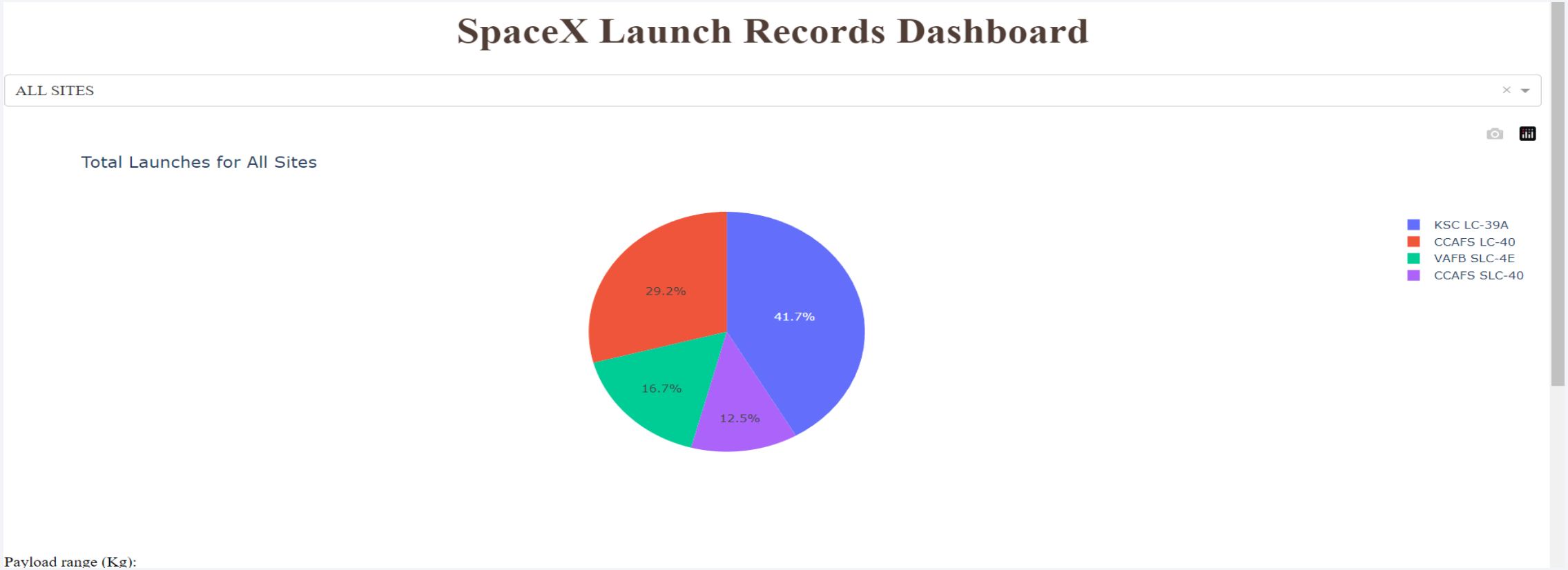
```
distance_highway = calculate_distance(launch_site_lat, launch_site_lon, closest_highway[0], closest_highway[1])
print('distance_highway =',distance_highway, ' km')
distance_railroad = calculate_distance(launch_site_lat, launch_site_lon, closest_railroad[0], closest_railroad[1])
print('distance_railroad =',distance_railroad, ' km')
distance_city = calculate_distance(launch_site_lat, launch_site_lon, closest_city[0], closest_city[1])
print('distance_city =',distance_city, ' km')
```

```
distance_highway = 0.5834695366934144  km
distance_railroad = 1.2845344718142522  km
distance_city = 51.434169995172326  km
```

- The above were the calculated distances.

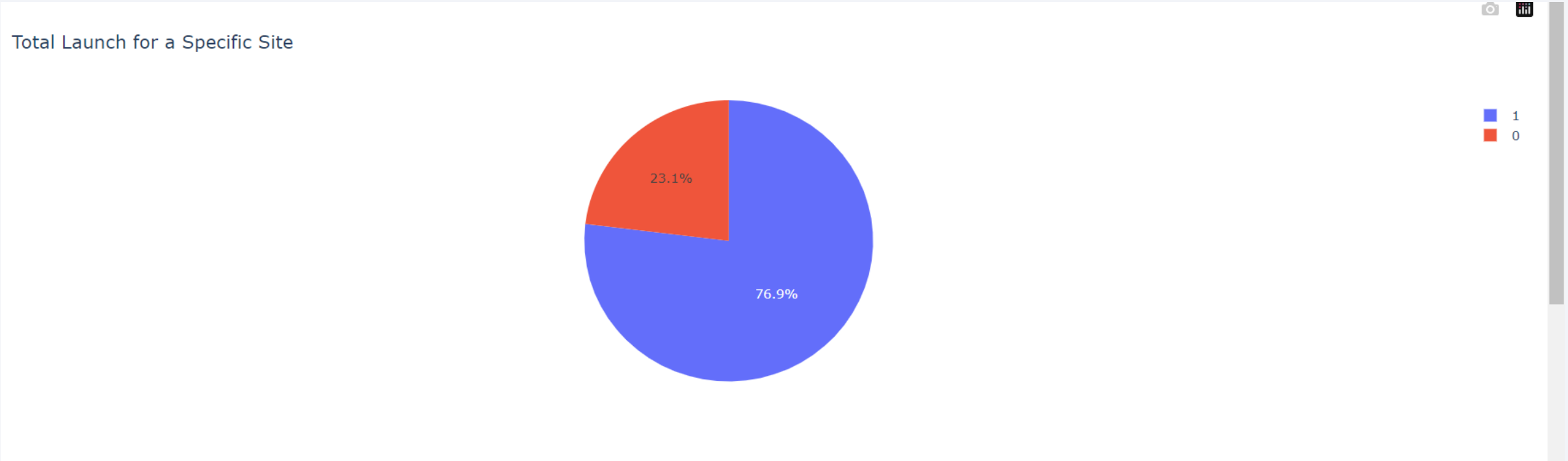

Section 4
Build a Dashboard with Plotly Dash

Dashboard – Success launch of all sites



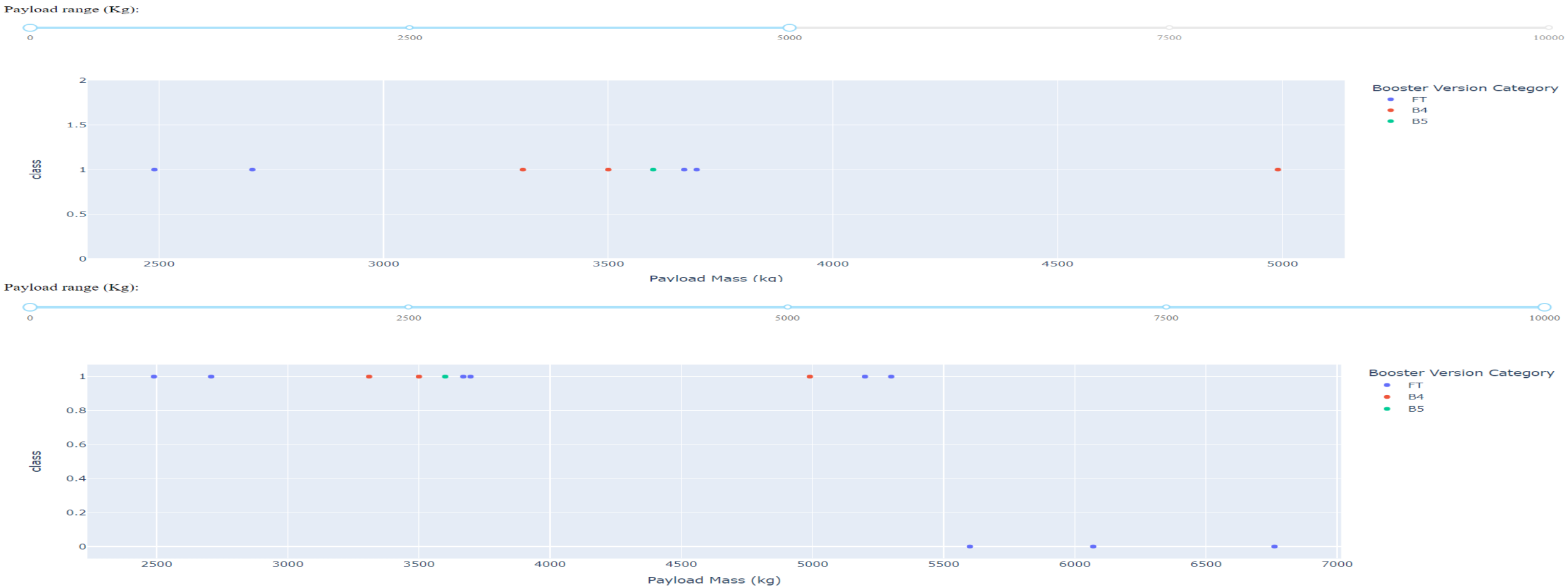
- KSC LC -39 A had the highest success percent.
- CCAFS SLC – 40 had the lowest success percent .

Dashboard – Pie chart of Launch site KSC-LC 39-A



- KSC-LC 39 -A had the highest success rate of 76.9%

Scatter plot for different payload range



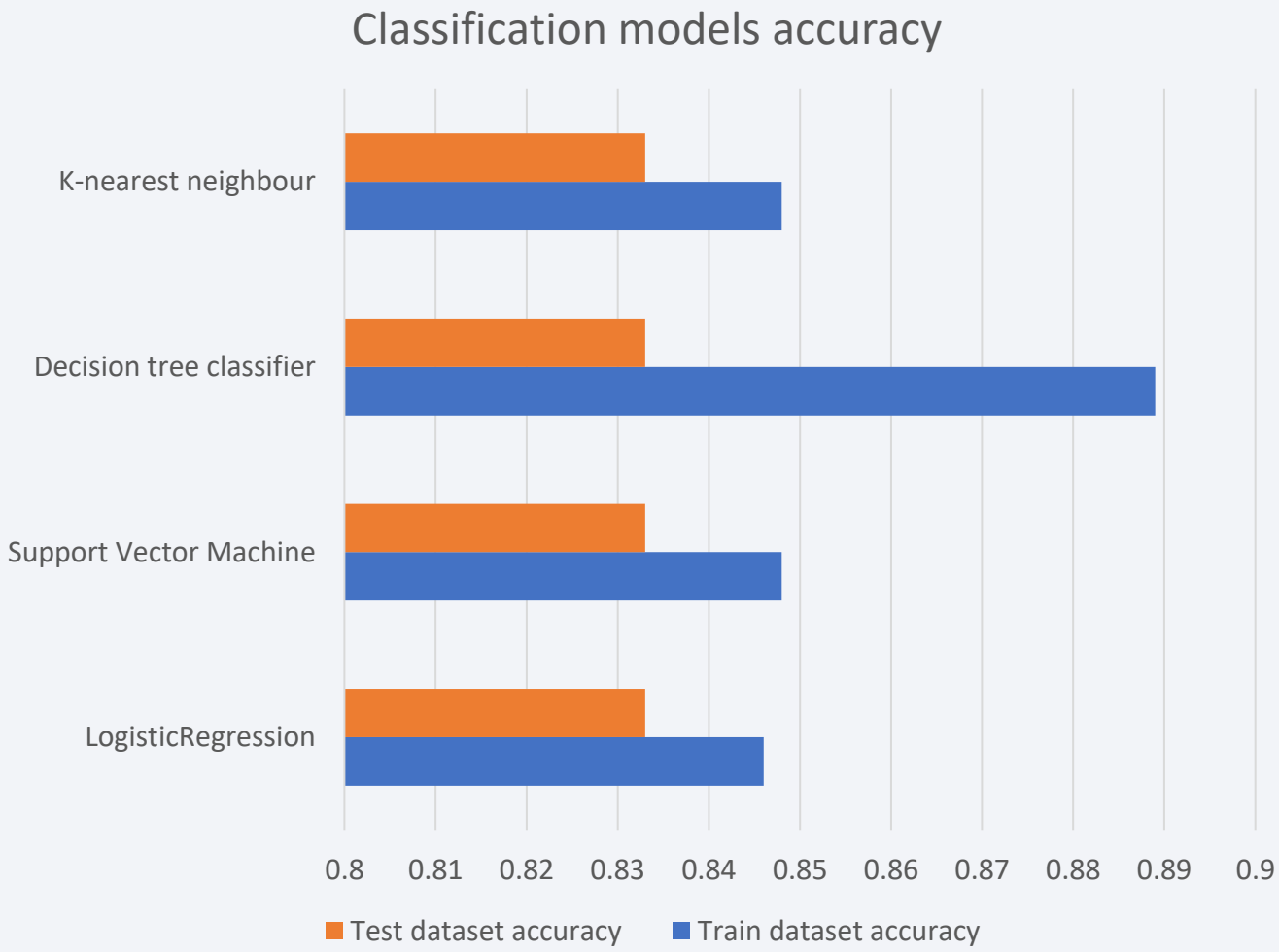
- Scatter plot for the payload range 5000, 10000 is shown.
- For higher payloads the FT booster version was predominantly used.

Section 5

Predictive Analysis (Classification)

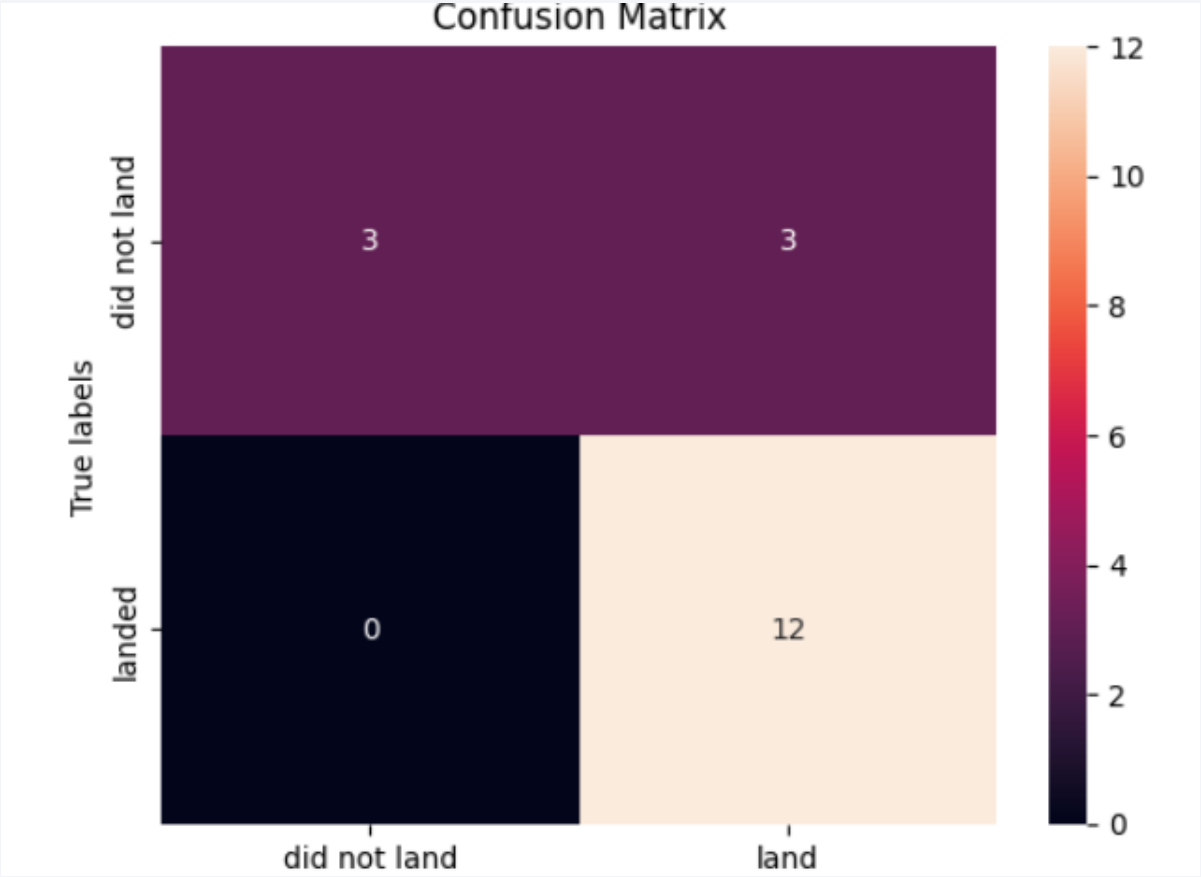
Classification Accuracy

- The test dataset accuracy was similar for all the four models.
- But when comparing the accuracy of the train dataset the decision tree model performed slightly better with an accuracy of 0.89.



Confusion Matrix

- The confusion matrix of decision tree classifier is shown.
- It performed the best when comparing the training datasets.



Conclusions

- The success rate of launches increased constantly from the year 2013.
- **VAFB-SLC** launch site was not preferred for heavy payload mass launches.
- Orbits **ES-L1, GEO, HEO** and **SSO showed 100% success rate.**
- In the **LEO** orbit the Success appears related to the number of flights.
- With heavy payloads the successful landing or positive landing rate are more for **Polar, LEO and ISS.**
- The launch sites were in the costal area with close proximity to Railways and Highways and far from city.
- **KSC-LC 39 -A** launch site had the highest success rate of **76.9%**
- **FT booster version** was predominantly used for higher payloads.
- **Decision tree classifier** had the highest accuracy which predicted a success rate of **83%**

Appendix

- All the Jupiter notebooks used in the lab is uploaded in my GitHub account.
- Link to my GitHub repository

Thank you!