IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Khushi Tibrewal
25th Jan 2024

Outline

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

Executive Summary

- Summary of methodologies

-Data collection methodology

-Perform data wrangling

-Perform exploratory data analysis (EDA) using visualization and SQL

-Perform interactive visual analytics using Folium and Plotly Dash

-Perform predictive analysis using classification models

- Summary of all results

-The success rate has increased since 2013 till 2020.

-The result gives 2015 as the earliest ground landing year.

-Launch Site KSC LC-39A has 76.9% of success percentage.

- The Decision tree model gave an accuracy of around 92% which is highest among all other models used.

Introduction

- Project background and context is:

  - To determine the outcome of first stage launch of rockets for company SpaceX . It can help to 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 to find answers to are:

  - Dependence of success on launch site, booster version, payload mass and Orbit type.

  - Best model to predict the outcome.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using Beautiful Soup method, Falcon 9 launch records HTML table is extracted from Wikipedia.
- Perform data wrangling
  - Exploratory data analysis (EDA) is used to find patterns in data and determine labels to train supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - To build the model:- Logistic regression, SVM, Tree, KNN models are built on training datasets.
  - To tune the model:- Use GridSearchCV() for optimizing the parameters of models.
  - To evaluate the model:- Select the model with highest accuracy.

Data Collection

- Data sets were collected using two methods.

  1) SpaceX API

  2) Web scraping.

Data Collection – SpaceX API

- Data collection with SpaceX REST calls is given in flowchart.

- GitHub URL of the completed SpaceX API calls notebook is:

- Data-Science-project/jupyter-labs-spacex-data-collection-api.ipynb at main · khushitibrewal/Data-Science-project (github.com)

- Request data using URL.
- Decode response content as json using .json() and turn it into pandas dataframe using .json_normalize()
- call Booster name
- Call launch site
- call payload mass
- call core data

Data Collection - Scraping

- The web scraping process is presented using flowcharts

- GitHub URL of the completed web scraping notebook is:

- Data-Science-project/jupyter-labs-webscraping (3).ipynb at main · khushitibrewal/Data-Science-project (github.com)

- Request Falcom9 Launch wiki page from URL using Beautifulsoup.

- Extract variable name from HTML table header.

- Create dictionary by parsing the launch HTML tables.

- Convert the extracted data from dictionary to pandas dataframe.

Data Wrangling

- Data wrangling is done in following manner:
- 1) Exploratory data analysis
- 2) Determine training labels
- Github link of data wrangling notebook is as follows:

Data-Science-project/labs-jupyter-spacex-Data wrangling.ipynb at main · khushitibrewal/Data-Science-project (github.com)

EDA with Data Visualization

- Charts plotted were as follows:

  1) Cat plot : flight number vs. payload mass. To see how payload mass affects the outcome of the launch.

  2) Scatterplot : flight number vs. launch site. To see how launch site affects the outcome of the launch.

  3) Scatterplot : launch site vs. payload mass. To see how payload mass and launch site affect the outcome of the launch.

  4) Bar plot : Success rate for Orbit types.

  5) Scatterplot : flight number vs. Orbit types. To see how Orbit types affects the outcome of the launch.

  6) Scatterplot : Payload mass vs. Orbit types. To see how Orbit types and payload mass affect the outcome of the launch.

  7) To know year by year success rate of launches.

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

Data-Science-project/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · khushitibrewal/Data-Science-project (github.com)

EDA with SQL

- The SQL queries performed are:

  - Name of unique launch sites

  - Total payload mass carried by NASA(CRS)

  - Date when first successful landing in ground pad was achieved.

  - Booster version that carried payload mass between 4000 kg and 6000 kg.

  - Total number of successful and failure missions.

  - Booster versions that carried maximum payload mass.

  - Failure outcomes in the year 2015.

- GitHub URL of your completed EDA with SQL notebook is:

Data-Science-project/jupyter-labs-eda-sql-coursera_sqllite (1).ipynb at main · khushitibrewal/Data-Science-project (github.com)

Build an Interactive Map with Folium

- Objects added to a folium map are:

- Circle is used to show various launch sites.

- color markers show outcomes of launches from launch sites.

- Mouse position is used to get the coordinated of points on the map.

- Polyline shows distance between launch sites and places such as coastline , railway etc.

GitHub URL of completed interactive map with Folium map is:

Data-Science-project/lab_jupyter_launch_site_location.jupyterlite (1).ipynb at main · khushitibrewal/Data-Science-project (github.com)

Build a Dashboard with Plotly Dash

- The dashboard includes the following:

  - A drop down to select name of launch site to set the launch site for which a pie chart represents success rate.

  - A drag bar to select payload mass for which a scatter plot represents outcome based on payload mass and booster version.

- GitHub URL of completed Plotly Dash lab is:

Data-Science-project/final_dashboard.py at main · khushitibrewal/Data-Science-project (github.com)

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- The models Logistic Regression, SUM, Decision Tree and KNN were built on training datasets.
- The models were evaluated by testing them on test dataset and finding accuracy score.
- The models were improved by finding best parameters using GridSearchCV() function.
- The best performing classification model was selected based on confusion matric and accuracy level.
- GitHub URL of completed predictive analysis lab is:

Data-Science-project/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb at main · khushitibrewal/Data-Science-project (github.com)

Results

- Exploratory data analysis results

  -The success rate has increased since 2013 till 2020.

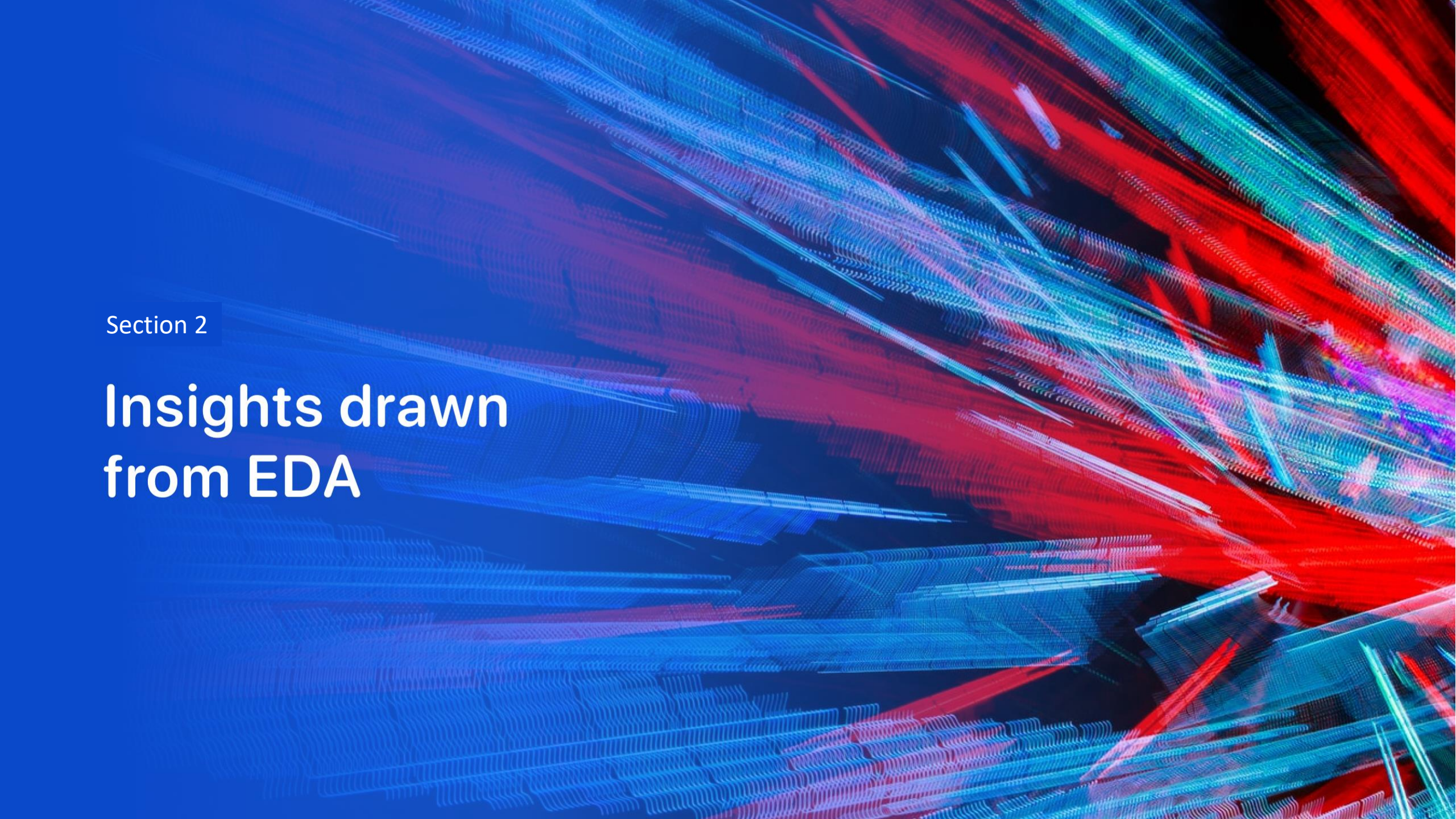
  -The result gives 2015 as the earliest ground landing year.

- Interactive analytics demo in screenshots

  -Launch Site KSC LC-39A has 76.9% of success percentage.

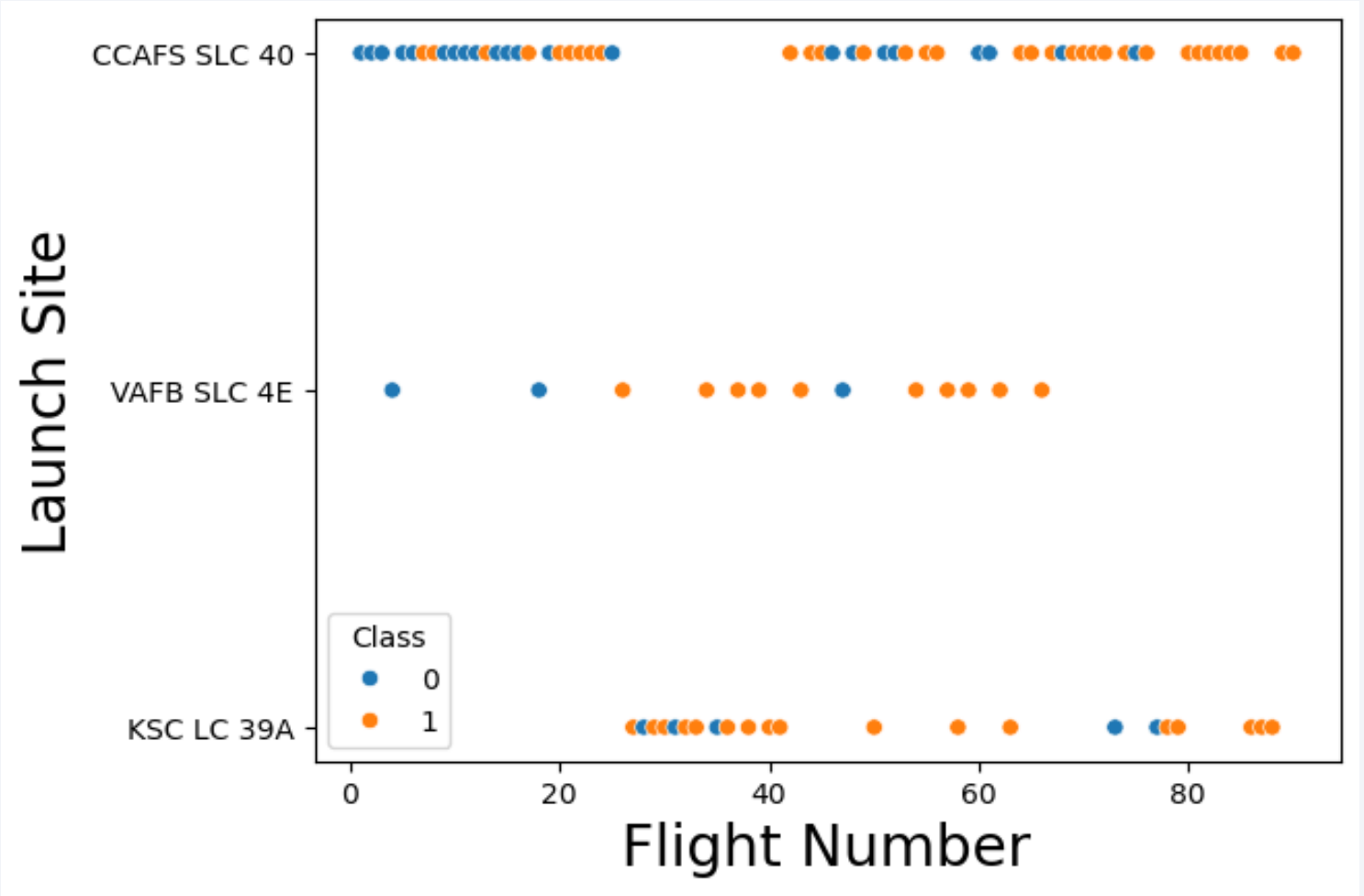
- Predictive analysis results:

  - The Decision tree model gave an accuracy of around 92% which is highest among all other models used.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

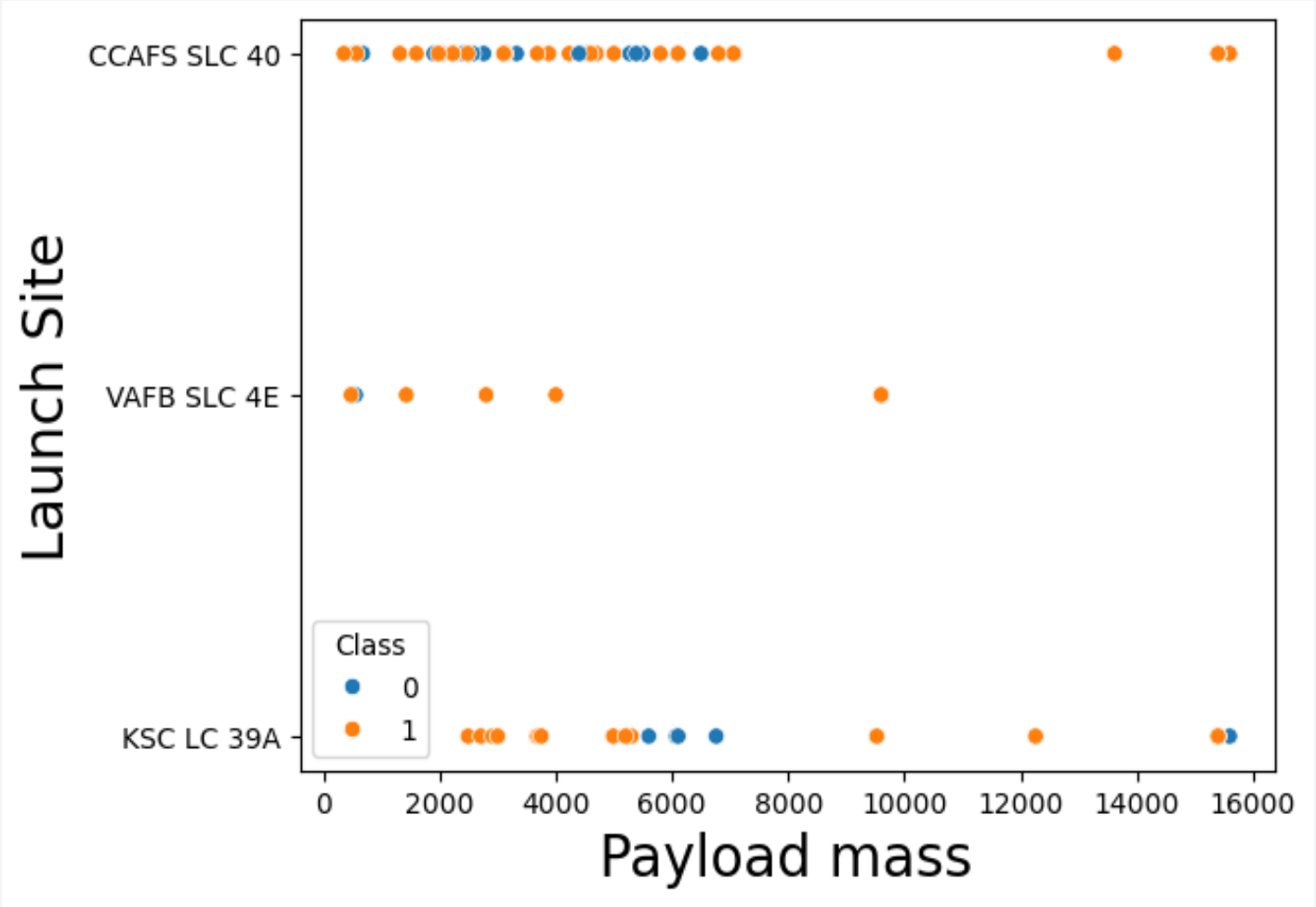
- The plot represents the scatter plot of Flight Number vs. Launch Site.
- After 43rd flight, Launch Site VAFB SLC 4E has launched all the flights successfully.
- For other Launch Site a clear cut conclusion cannot be reached.



Payload vs. Launch Site

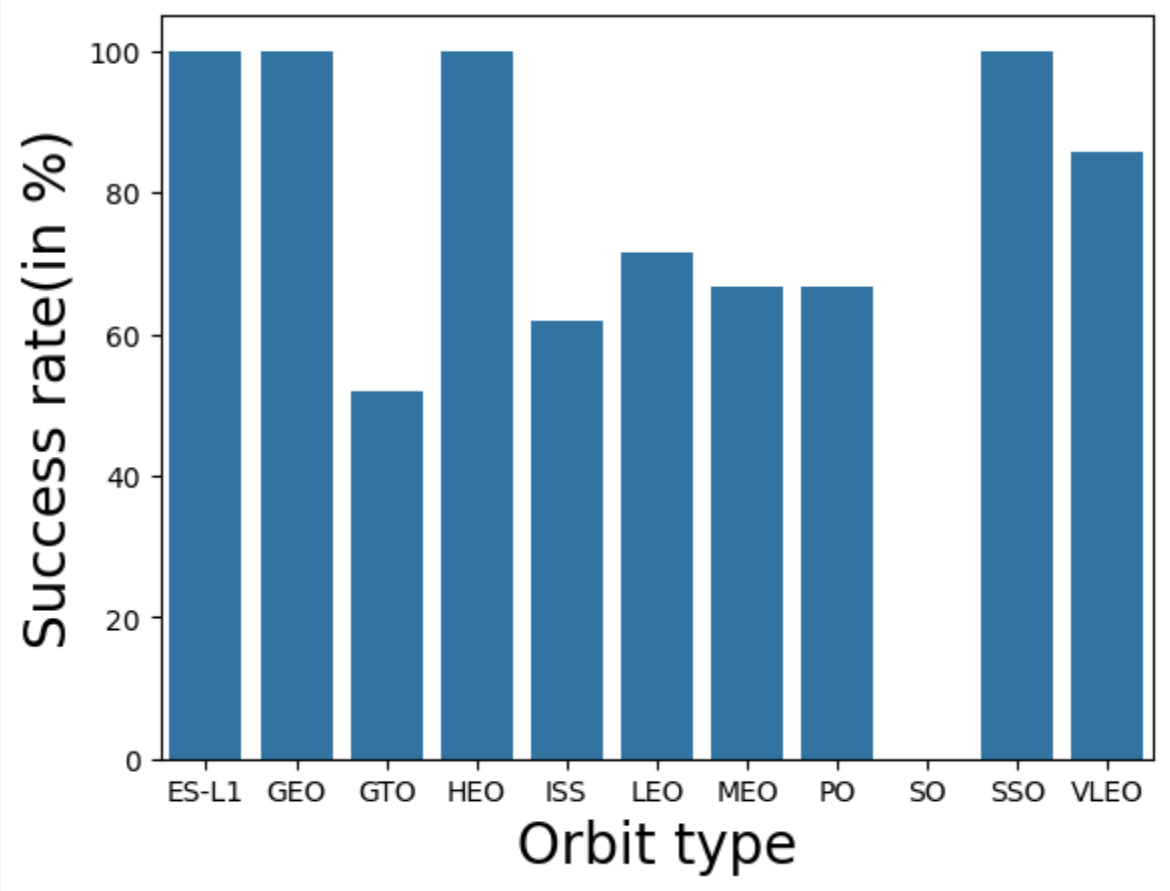
- The plot represents the scatter plot of Payload vs. Launch Site

- VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000)



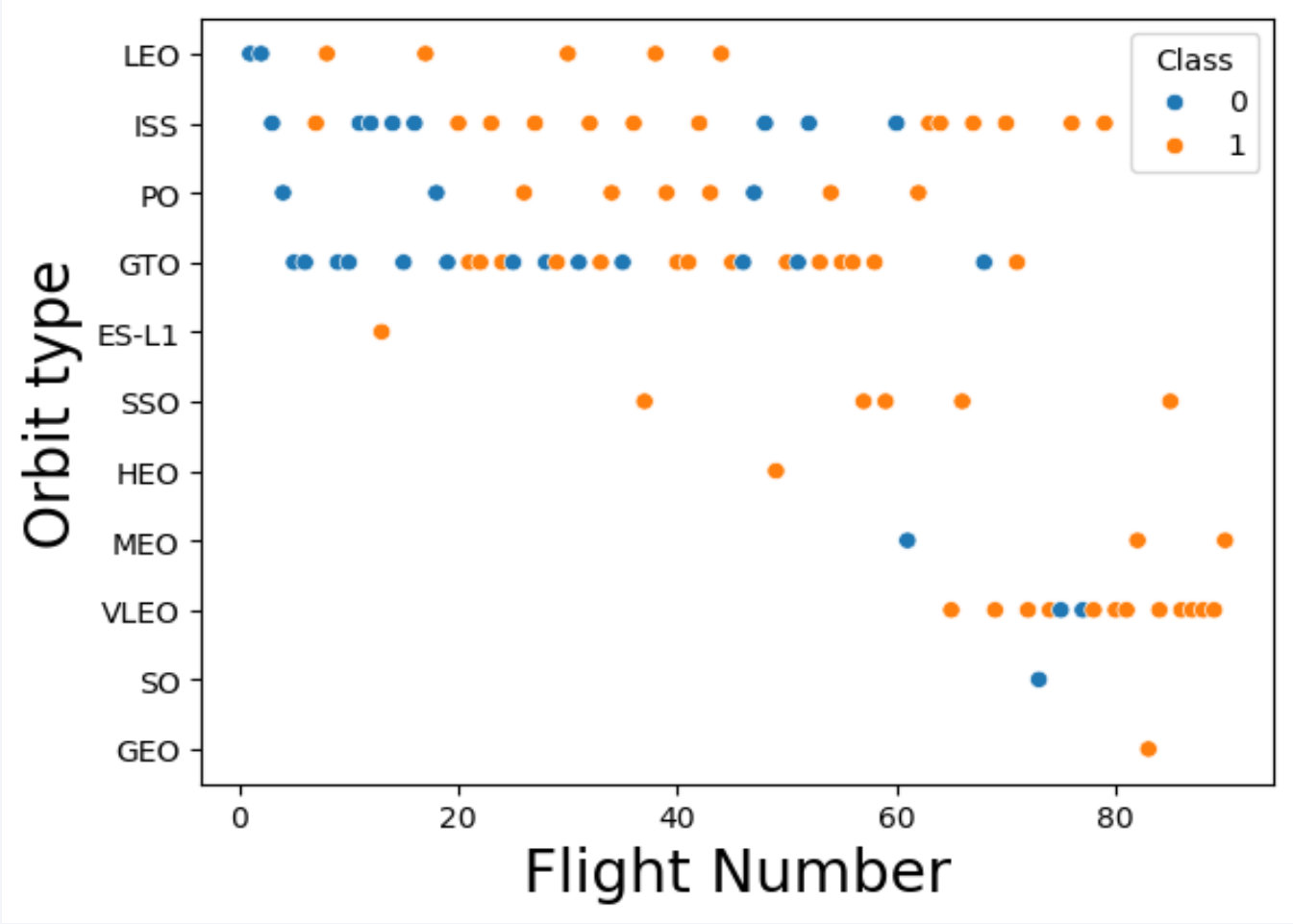
Success Rate vs. Orbit Type

- The graph represents a bar chart for the success rate of each orbit type
- Orbits ES-L1, GEO, HEO and SSO have highest success rate.



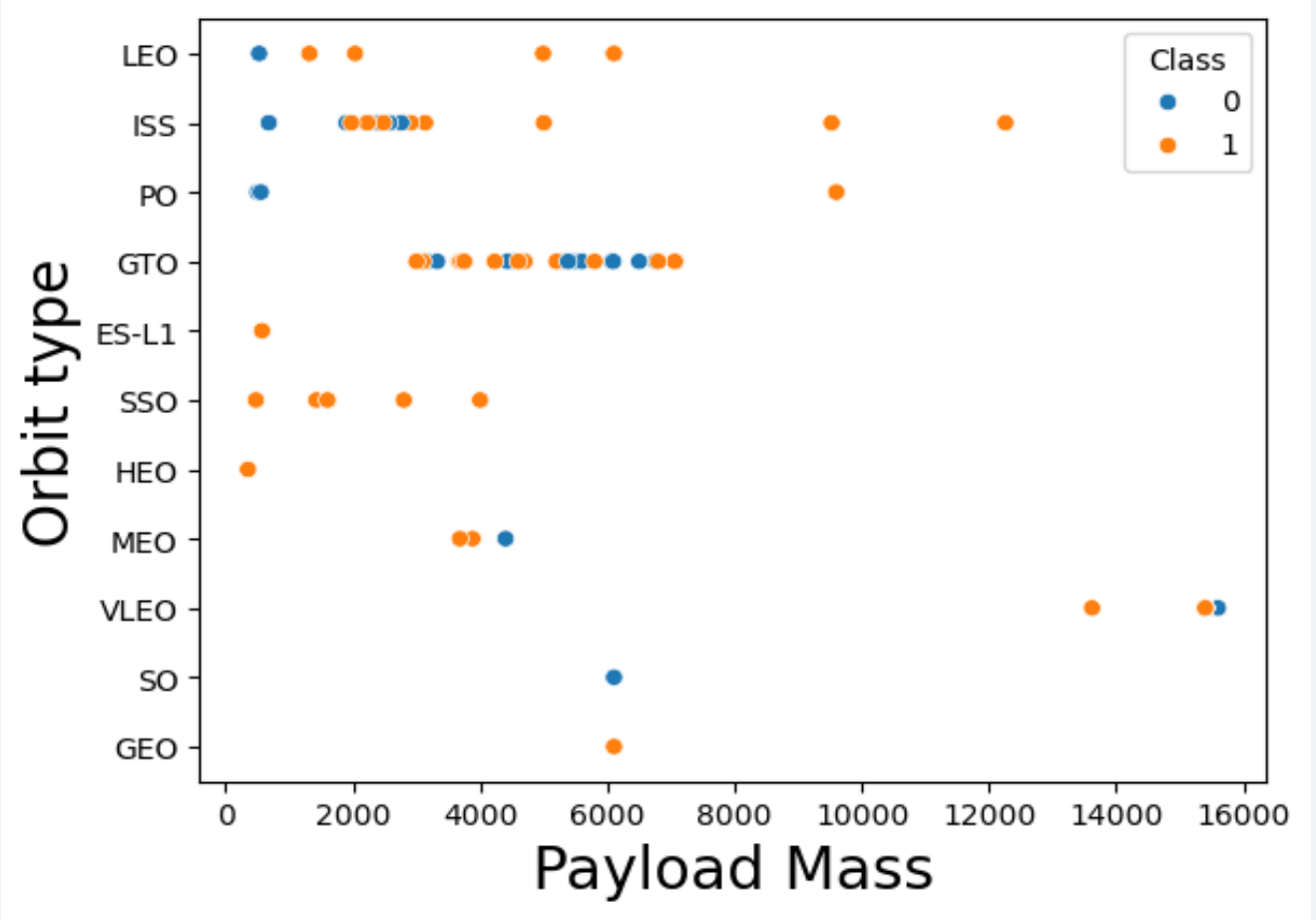
Flight Number vs. Orbit Type

- The plot represents a scatter point of Flight number vs. Orbit type
- 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

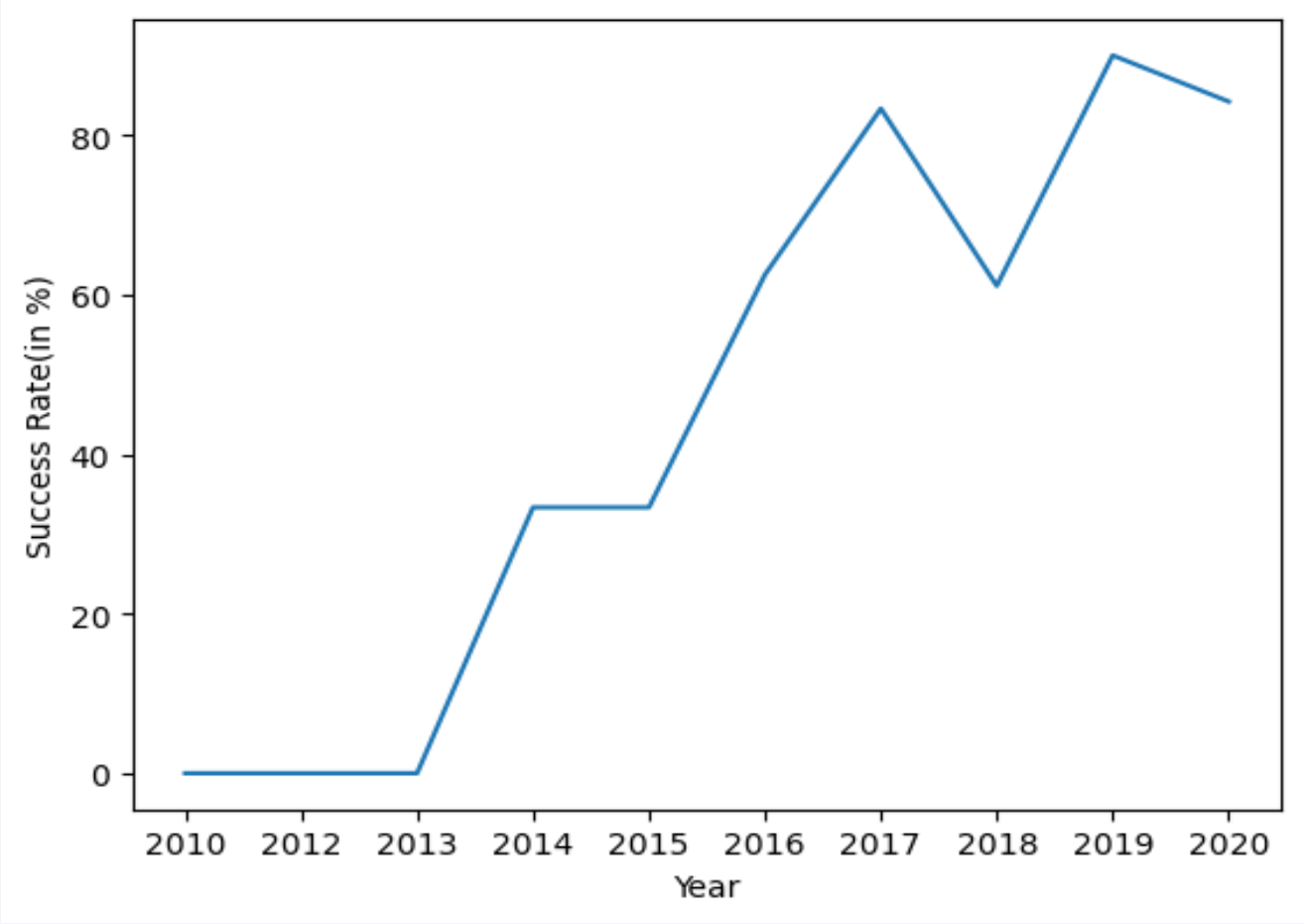
- The plot represents a scatter point of payload vs. orbit type
- With heavy payloads the successful landing or positive landing rate are more for Polar , LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both present here.



Launch Success Yearly Trend

- The graph represents a line chart of yearly average success rate.

- The success rate has increased since 2013 till 2020



All Launch Site Names

- The code to extract unique launch site is:

  > %sql SELECT DISTINCT("Launch_Site") FROM SPACEXTABLE;

- The result of the query is:

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

- The table shows all the unique launch sites name present in the dataset.

Launch Site Names Begin with 'CCA'

- The query is:

```
> %%sql

  select “Launch_Site” from SPACEXTABLE

  where “Launch_Site” like (‘CCA%’) limit 5;
```

- The result gives the Launch Site which is CCAFS LC-40 or CCAFS SLC-40
- The limit 5 parameters returns top five results.

Total Payload Mass

- The query to calculate the total payload carried by boosters from NASA is:

```
%%sql
select “Customer”, sum(“Payload_Mass_KG_”) as payload from SPACEXTABLE
where “Customer” like (‘NASA (CSR)’);
```

- The result gives the sum of payload mass carried by boosters from NASA.

Customer	payload
NASA (CSR)	45596

Average Payload Mass by F9 v1.1

- The query to calculate the average payload carried by F9 v1.1 is:

```
%%sql

select “Booster_Version”, AVG(“Payload_Mass_KG_”) as payload from SPACEXTABLE

where “Booster_Version r” like (‘F9 v1.1’);
```

- The result gives the average of payload mass carried by F9 v1.1:

Booster _Version	AVG(“Payload_Mass_KG_”)
F9 v1.1	2928.4

First Successful Ground Landing Date

- The query to find the dates of the first successful landing outcome on ground pad is:

```
%%sql

select min("Date") from SPACEXTABLE

where "Landing_Outcome" like ('SUCCESS (ground pad)');
```

- The result gives 2015 as the earliest ground landing year.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.6
F9 FT B1031.2

- Conditional function is used to filter the result.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

Success	Failure
100	1

- The mission outcome shows that, 100 cases are successful and 1 case in failure.

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass is:

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

- Thos is the list of boosters that carried maximum payload mass.

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are:

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

- The month and year is first extracted from Date column using substr() function.

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

- The ranking of 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 is :

- Maximum landing outcome is No attempt.

Landing_Outcome	QTY
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

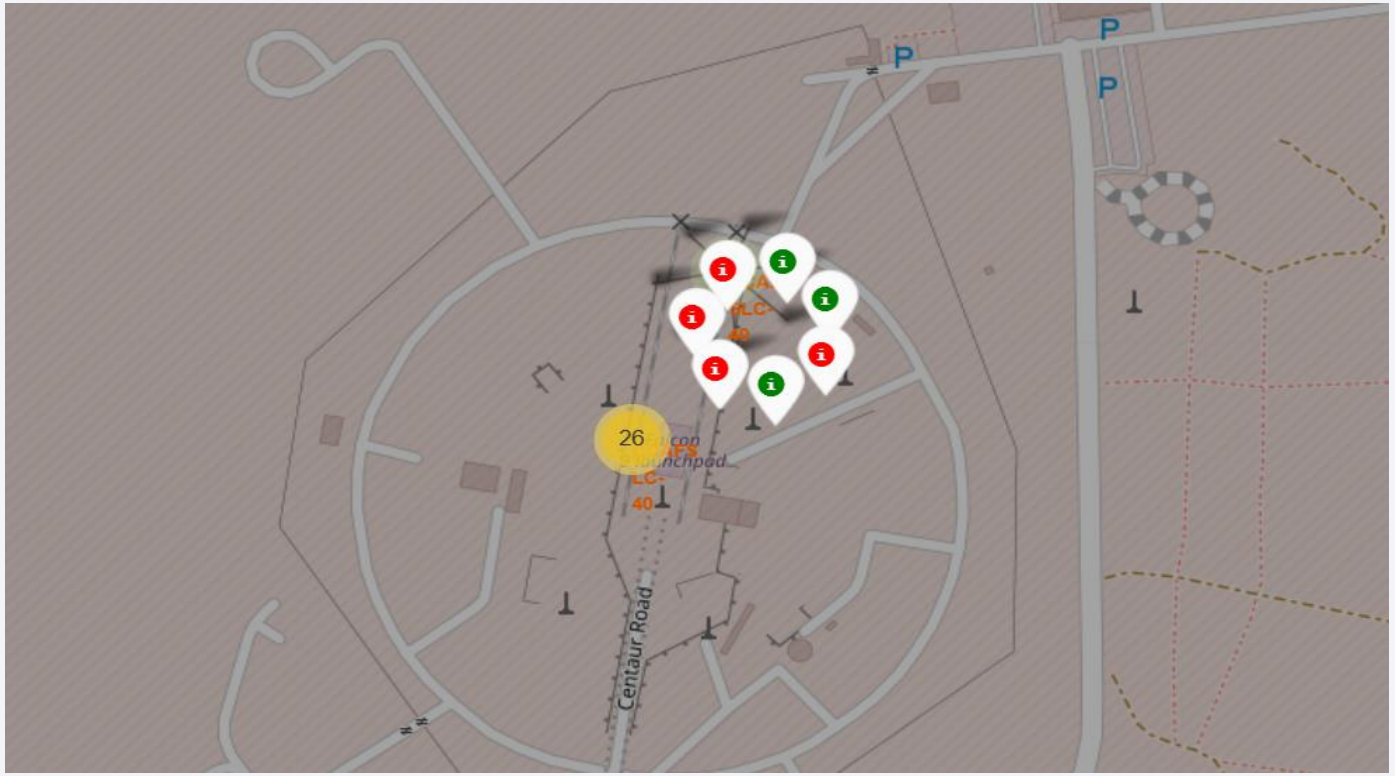
Launch Sites Proximities Analysis

Location of Launch Sites

- VAFB SLC-4E is on the western coast while CCAFS LC-40, CCAFS SLC-40 and KSC LC-39A are on the eastern coast.

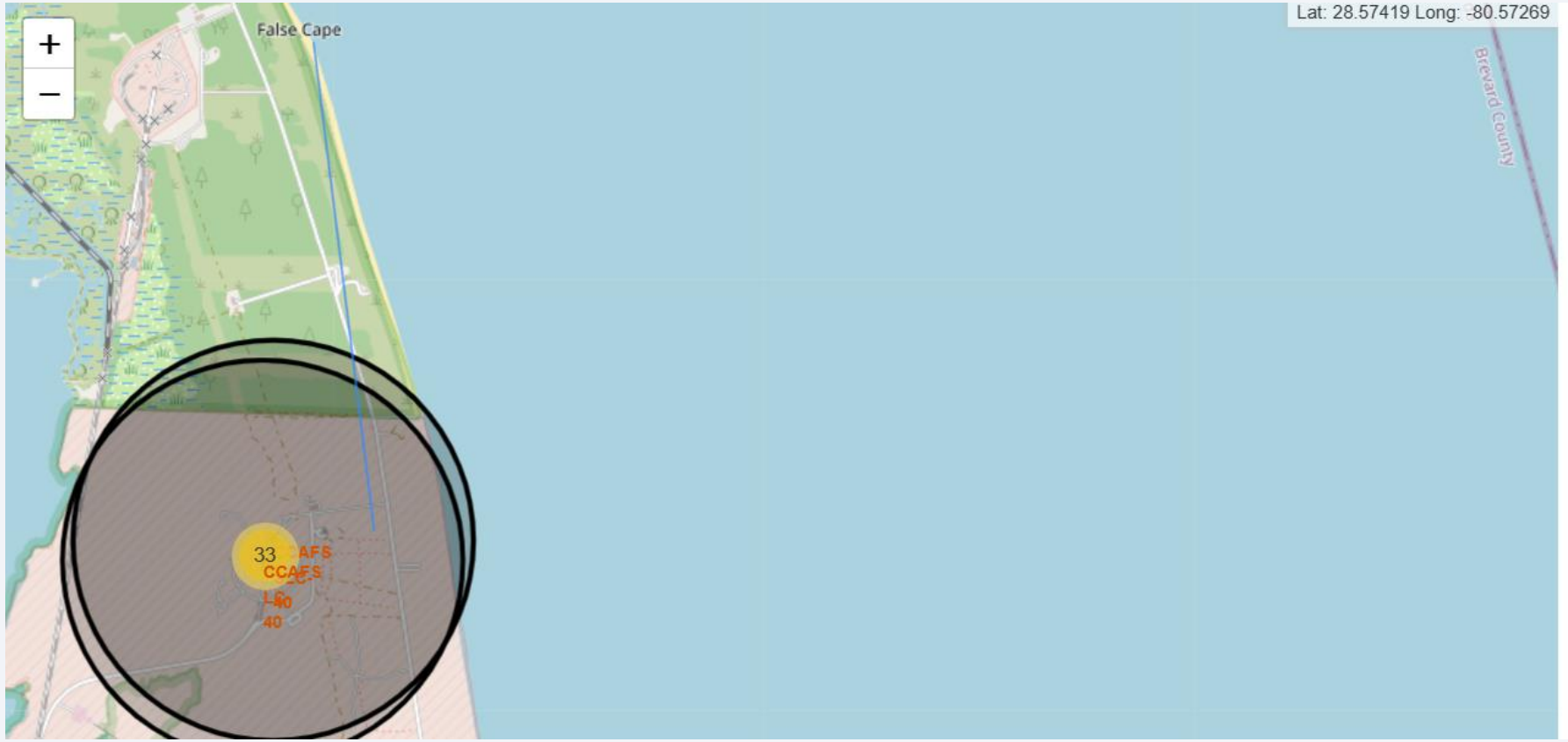
Launch outcomes

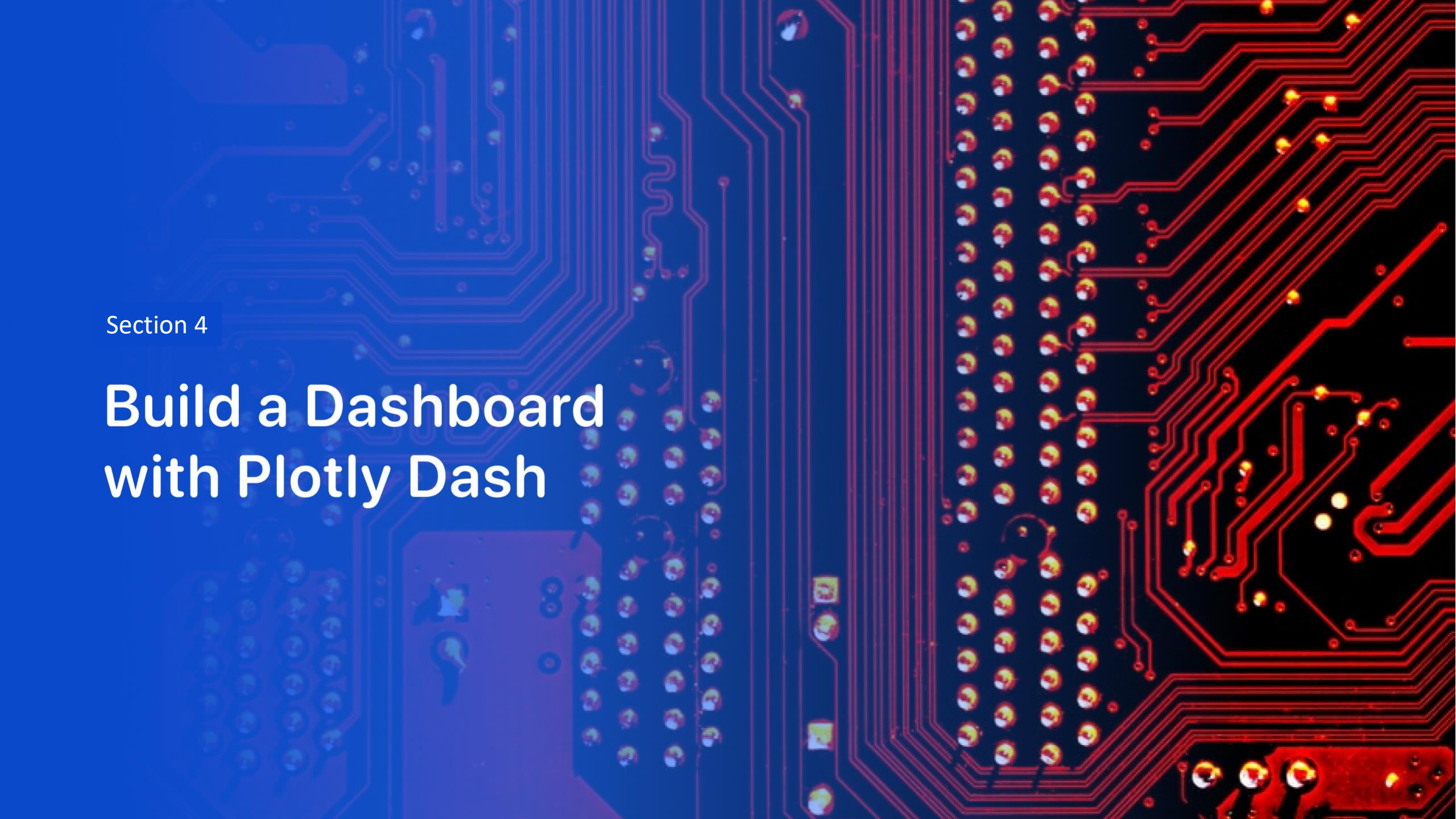
- The red pointers represent launch failure whereas green pointer represents launch success.
- KSC LC-39A has maximum launch success ratio.



Launch Site to coastline distance

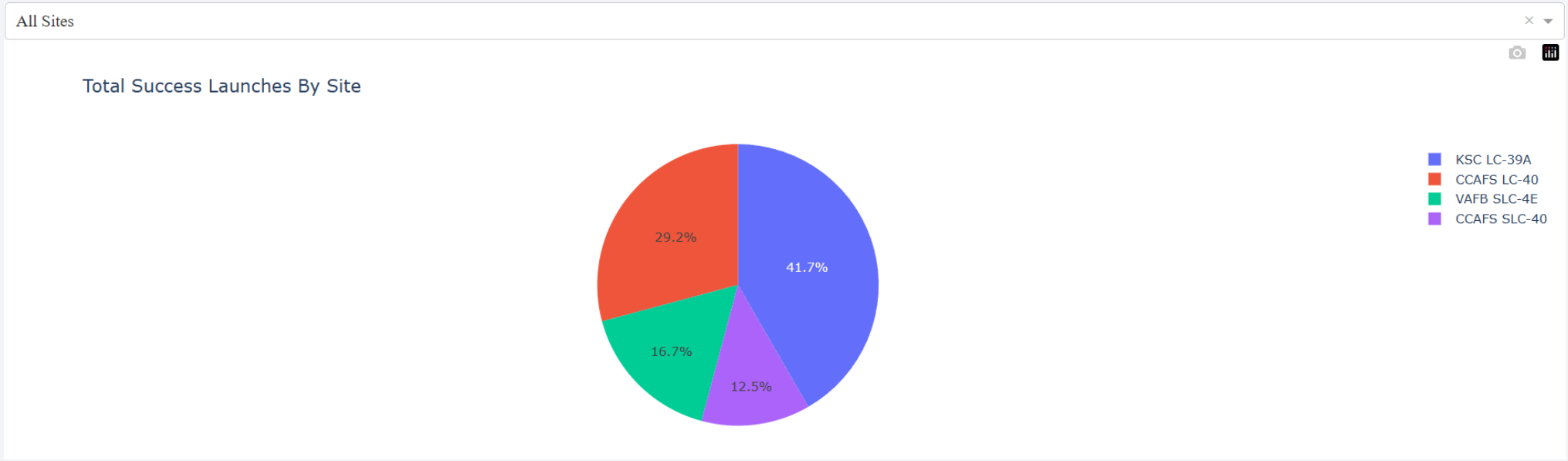
- The map shows distance between launch site and coastline.




Section 4
Build a Dashboard with Plotly Dash

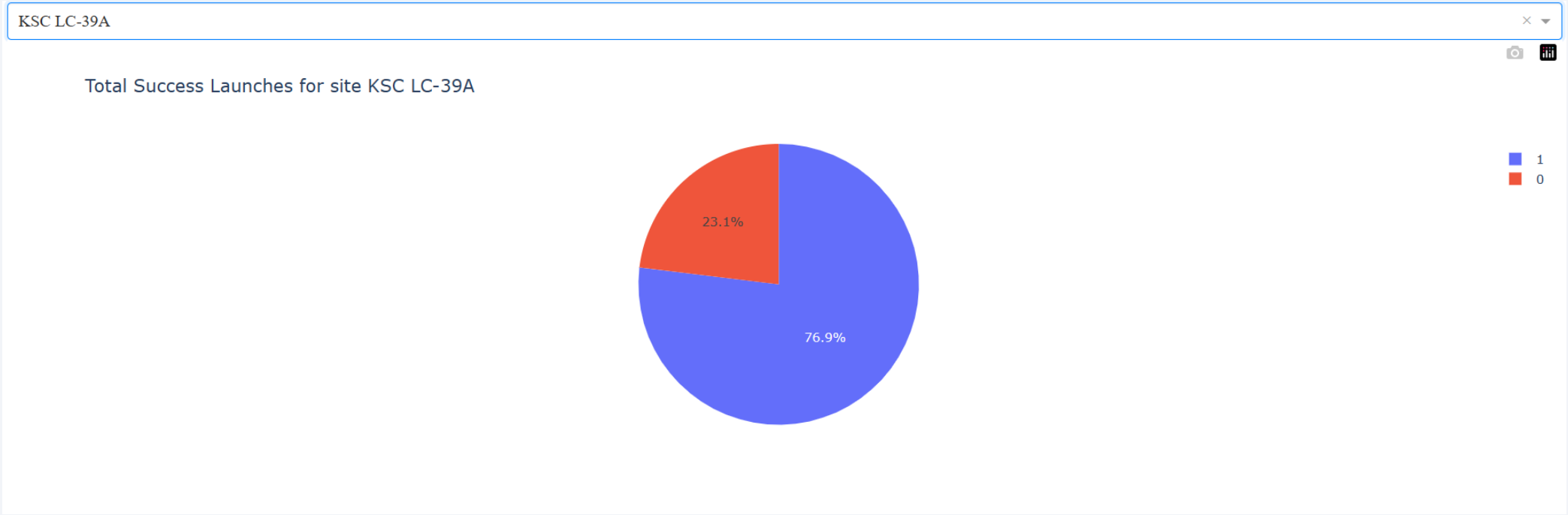
Total Success Launches by Sites

- The Pie Chart represents the proportion of total successful launches from various launch sites.
- The maximum successful launches are from Launch Site KSC LC-39A



Launch Site with maximum Success launches

- The maximum successful launches are from Launch Site KSC LC-39A
- Launch Site KSC LC-39A has 76.9% of success percentage.



Outcome for various payload mass from various Booster version

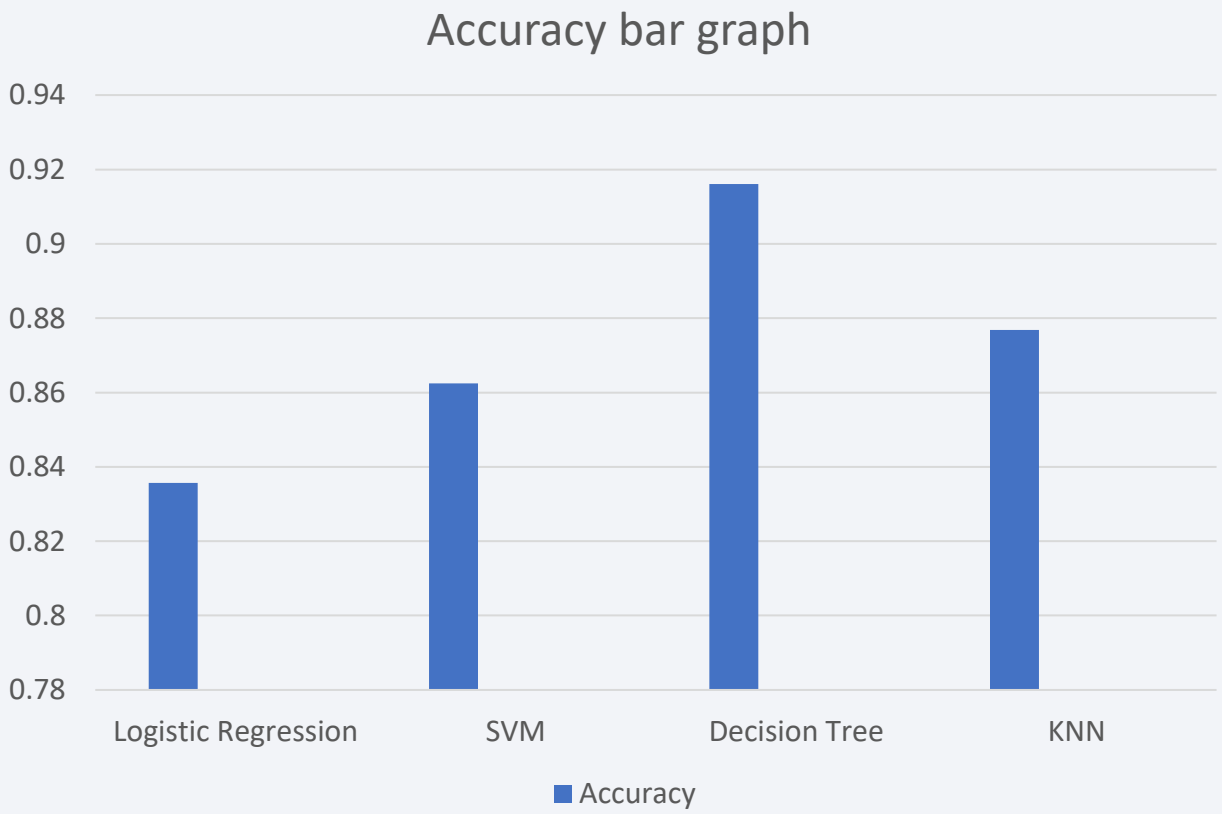
- Booster version B5 has only one successful launch with payload of 3600kg
- No booster version has successful launch for payload more than 5300 kg.

Section 5

Predictive Analysis (Classification)

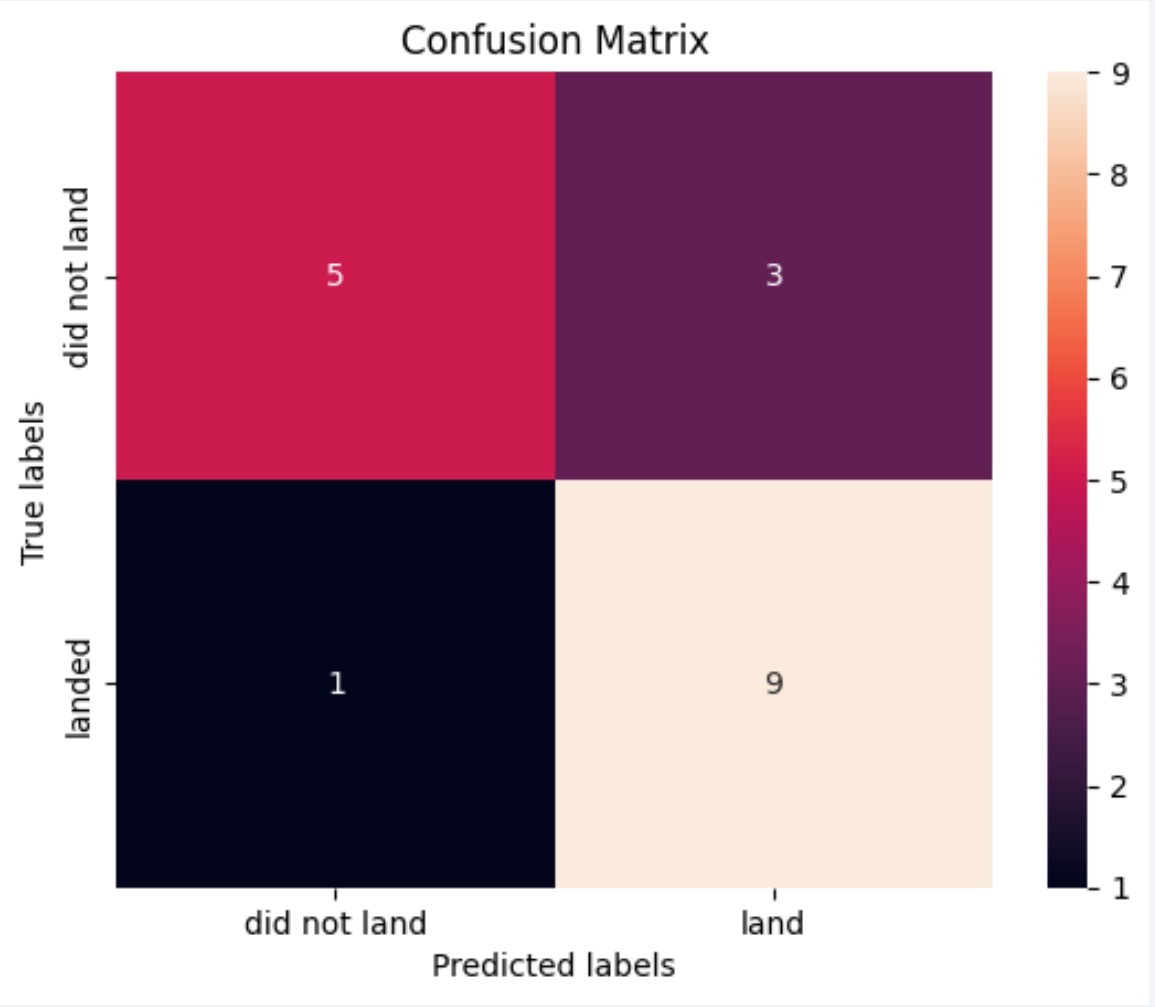
Classification Accuracy

- The bar graph represents the classification Accuracy score of all the models used.
- The model which has the highest classification accuracy is Decision Tree.



Confusion Matrix

- The confusion matrix of the best performing Decision tree model is shown in figure.
- The Decision tree model gave an accuracy of around 92% which is highest among all other models used.



Conclusions

- After 43rd flight, Launch Site VAFB SLC 4E has launched all the flights successfully.
- KSC LC-39A has maximum launch success ratio.
- No booster version has successful launch for payload more than 5300 kg.
- The Decision tree model gave an accuracy of around 92% which is highest among all other models used.

Appendix

- The month and year is first extracted from Date column using substr() function.
- The query to calculate the average payload carried by F9 v1.1 is:

```
%%sql
select “Booster_Version”, AVG(“Payload_Mass_KG_”) as payload from SPACEXTABLE
where “Booster_Version r” like (‘F9 v1.1’);
```

- The query to calculate the total payload carried by boosters from NASA is:

```
%%sql
select “Customer”, sum(“Payload_Mass_KG_”) as payload from SPACEXTABLE
where “Customer” like (‘NASA (CSR)’);
```

Thank you!