
# Winning Space Race with Data Science

Justin Ho
25th Jan 2023

# Outline

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

# Executive Summary

- The methodologies that were used to analyze the data are as follows:
  - Data was collected via the SpaceX API and web scraping.
  - Exploratory Data Analysis (EDA) included data wrangling and interactive visualization analytics.
  - Employment of Machine Learning for predictions.
- Summary of all results
  - To predict the success of launches, EDA identified the best features to do so.
  - Valuable data was procured from public sources.
  - With all the collected data, Machine Learning produced the best possible optimum model for predictions.

# Introduction

- This is a feasibility study to evaluate the potential of the new company Space Y to compete head on with Space X.

- Two metrics of measurement to define a favorable outcome are as follows :
  - By estimation of the total cost of the launches, it is necessary to predict the successful landings of the first stage of the rockets.
  - The best possible location to implement the launches.

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology :
  - Data sources were obtained from Space X :
    - Space X API
    - Web scraping
- Perform data wrangling
  - After summarizing the analyzed features based on the outcome data, a favorable landing outcome label was created.
- Perform exploratory data analysis (EDA) using visualization and SQL

# Methodology

## Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - After normalization of data, it was split into train and test data sets and evaluated by 4 different classification models. The accuracy of each model was determined by choosing various combinations of the parameters specified.

# Data Collection

- The data collected from these sources are as follow :
  - Space X API
    - https://api/spacexdata.com/v4/rockets/
  - Wikipedia
    - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

# Data Collection – Space X API

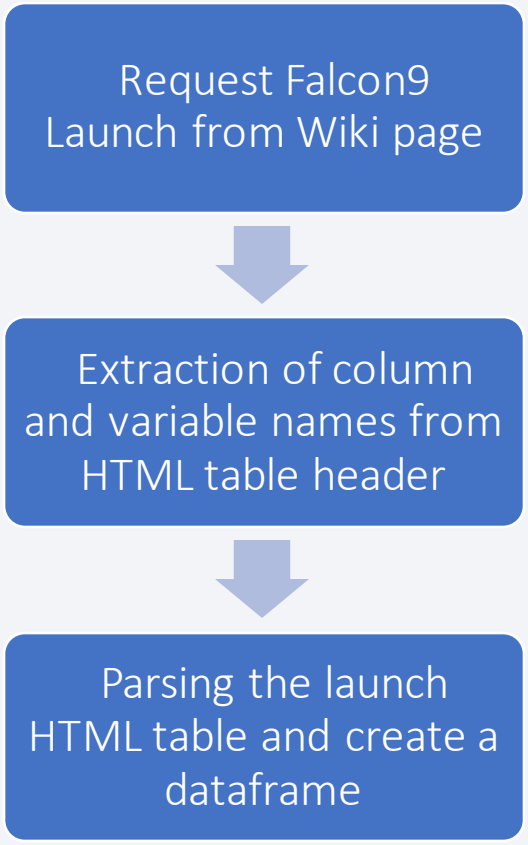
- An API that is accessible to the public from Space X was utilized.
- The API was called as depicted in the flowchart and data preparation ensued.


Source :

https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

# Data Collection - Scraping

- Additional data from SpaceX launches were obtained from Wikipedia.




Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/jupyter-labs-webscraping.ipynb

# Data Wrangling

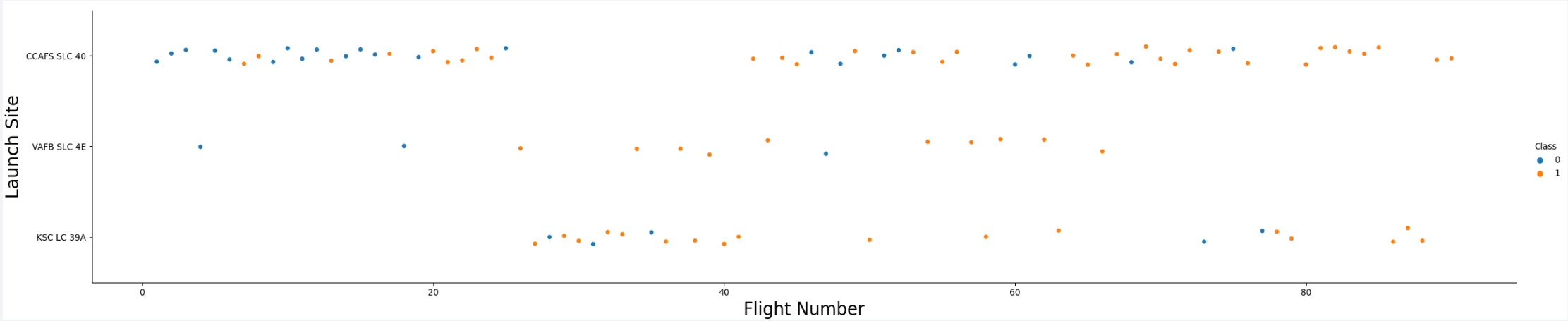
- Exploratory Data Analysis (EDA) was carried out with the obtained dataset.
- Calculations were determined by the summaries of launches per site, occurrences of mission outcome per orbit type as well as occurrences of each orbit.
- The outcome label was created from the outcome column.


Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

# EDA with Data Visualization

- After exploration of data, bar plots and scatterplots were utilized to present the relationship between these features and certain pairing examples:
    - Launch Site Flight Number, Payload Mass X Flight Number, Launch Site X Payload Mass, Payload and Orbit, Orbit and Flight Number.




Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

# EDA with SQL

- The SQL queries executed are as follows :
  - Names of unique space mission launch sites.
  - Top 5 launch site names that begin with string 'CCA'.
  - Total payload mass carried by boosters launched by NASA (CRS).
  - The average payload mass by booster version F9 c1.1.
  - Data of first successful landing outcome in ground pad.
  - Names of boosters that were successful with drone ship and a payload mass between 4000 & 6000 kg.
  - Total number of mission outcomes that consists of success or failures.
  - Names of booster versions with carried maximum payload mass.
  - Names of booster versions from failed landing outcomes in drone ship and launch site names in 2015.
  - Ranking count of landing outcomes between date 2010-06-04 and 2017-03-20.

Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

# Build an Interactive Map with Folium

- Circles, lines, markers and marker clusters were utilized with Folium Maps.
  - Circles showed highlighted areas with coordinates specified.  e.g. NASA Johnson Space Center.
  - Launch sites were indicated by markers.
  - Lines were drawn to show the distances between two coordinates.
  - Grouping of events in each coordinate pair such as launches from a launch site were indicated by marker clusters.

Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

# Build a Dashboard with Plotly Dash

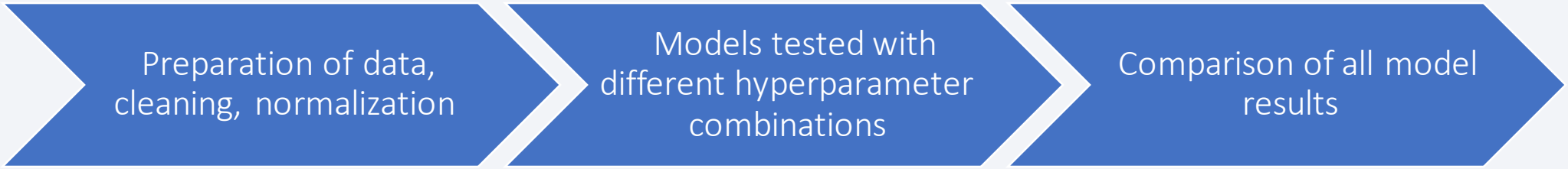
- Data visualizations with graphs and plots are as follow :
  - Payload range
  - Percentage of launches by site.
- This dashboard method showed a quick and clear analysis between the relationships of payloads and launch sites.  It also identified the best launch sites.

Source :

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/spacex_dash_app.py

# Predictive Analysis (Classification)

- These are the 4 classification models were built and compared :
  - Logistic regression
  - Support vector machine
  - Decision Tree
  - K Nearest Neighbors




Source

- https://github.com/thisSELFmySELF/IBM-applied-data-science-capstone-project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

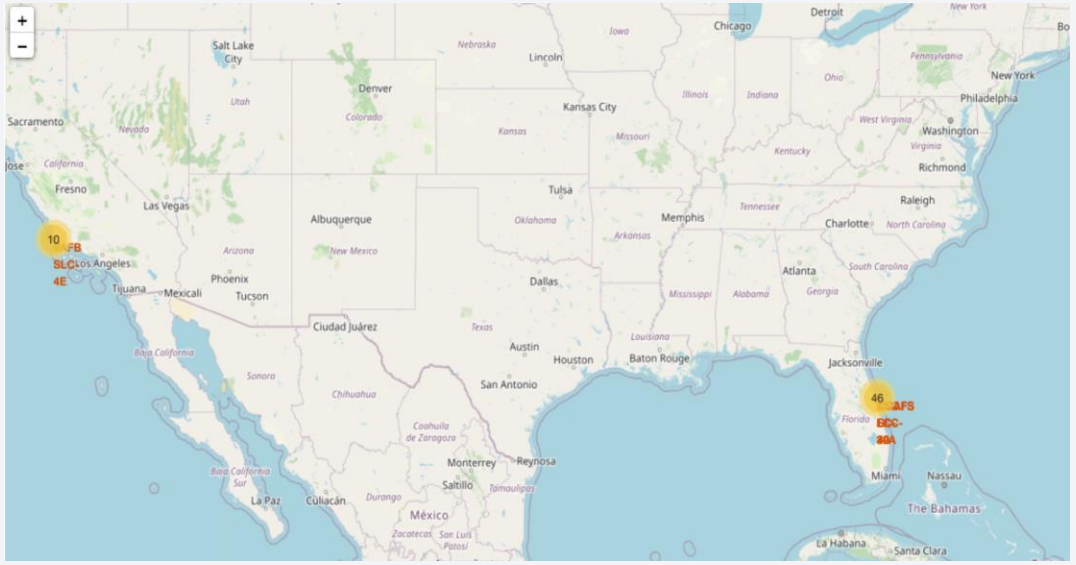
# Results

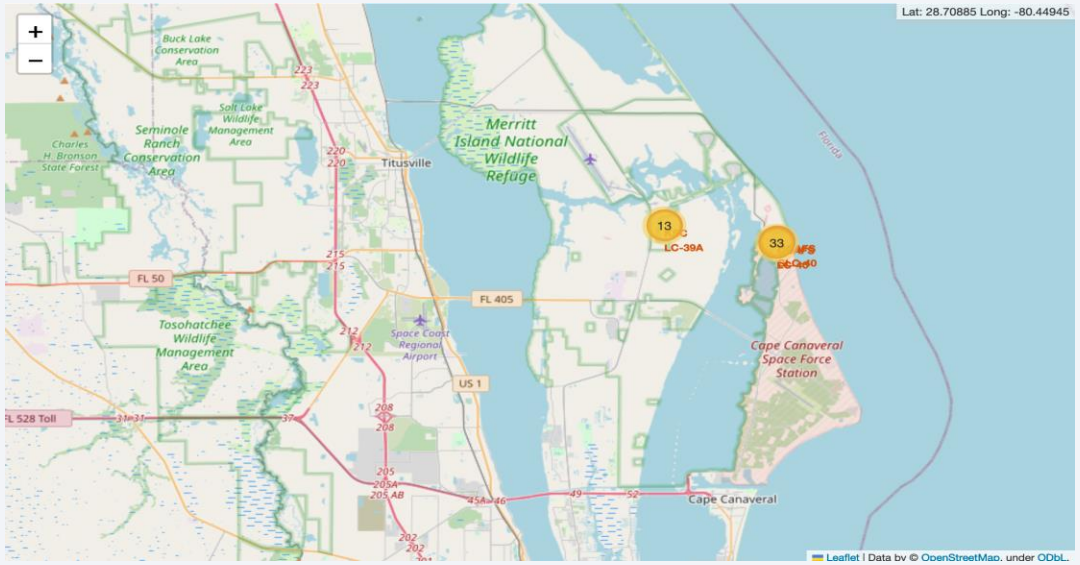
- 4 different launch sites were used by Space X.
- The F9 v1.1 booster carries an average payload of 2.928kg.
- After the first launch, the following successful landing outcome occurred in 2015.
- With an above average payload, numerous Falcon 9 booster versions succeeded in landing in drone ships.
- A high percentage as close to 100% attributed towards successful mission outcomes.
- 2 booster versions that failed in landing on drone ships in 2015 were F9 v1.1 B1012 and F9 v1.1 B1015.
- As the years passed, the number of landing outcomes progressively improved.

- Interactive analytics demo in screenshots
- Predictive analysis results

# Results

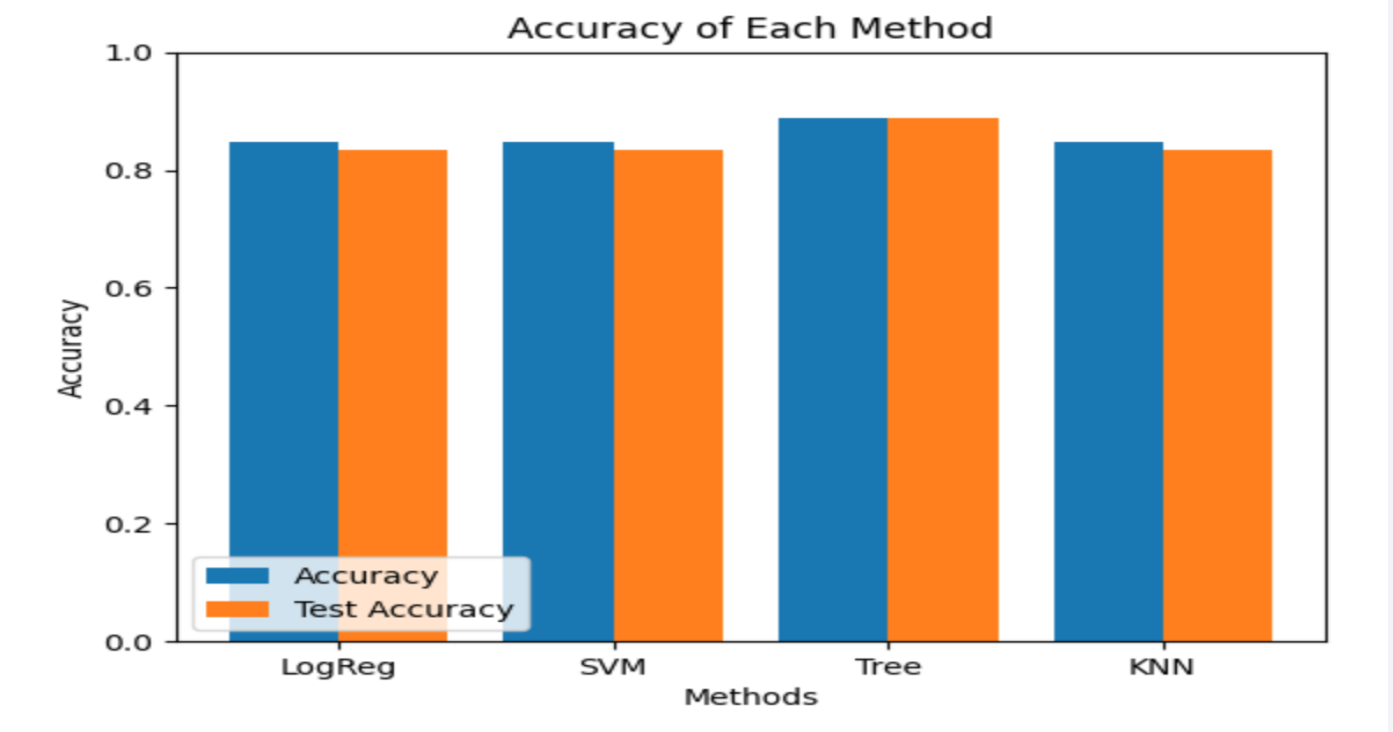
- Visualizations with interactive analytics showed that launch sites deemed safe were located near the coastal area and surrounded with adequate logistical infrastructure.
- Majority of the launch sites were situated along the east coast.



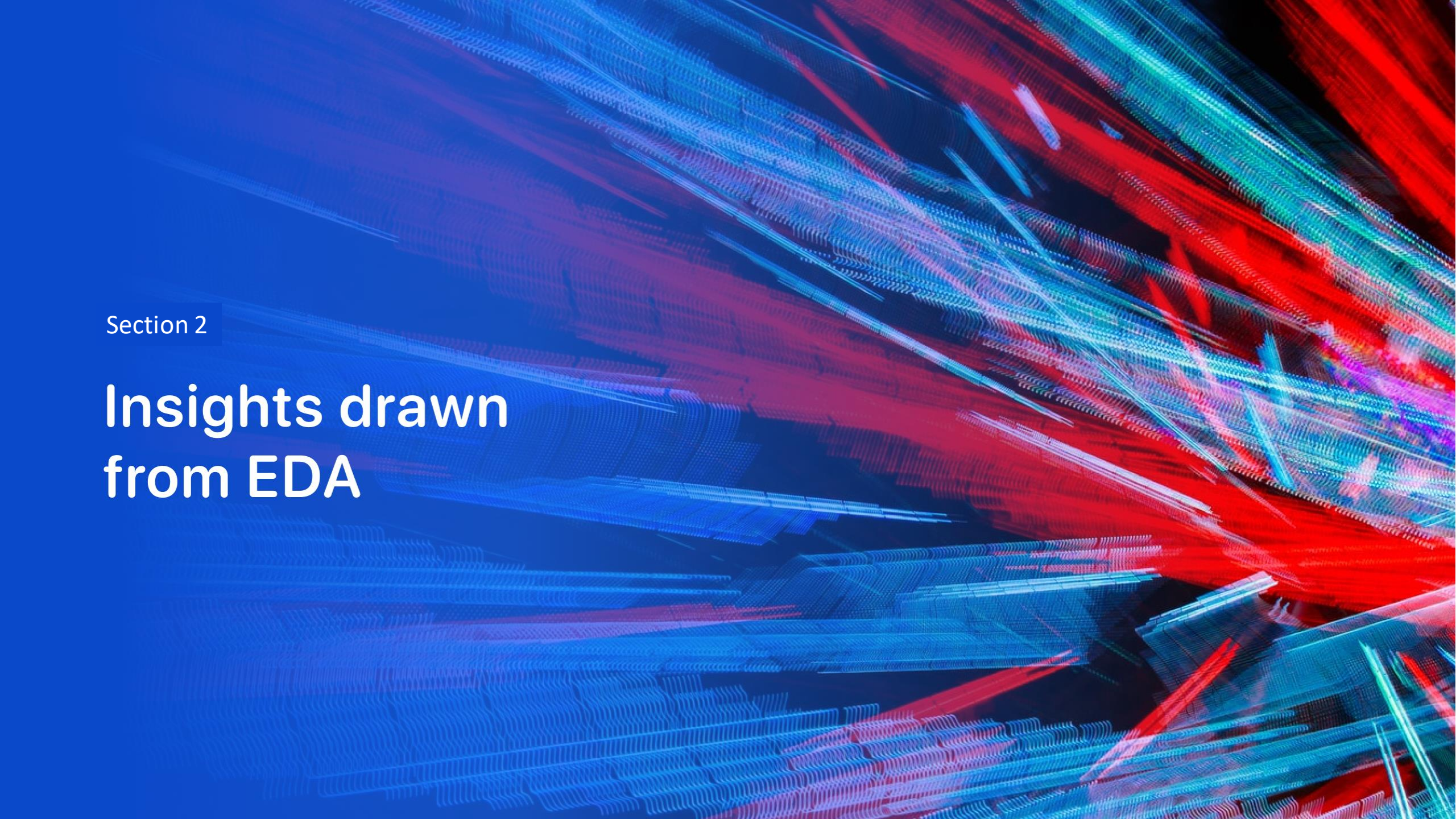
# Results

- After comparing the results of the 4 classification models that were created, the best results for training and test accuracy belonged to the Decision Tree Classifier model.

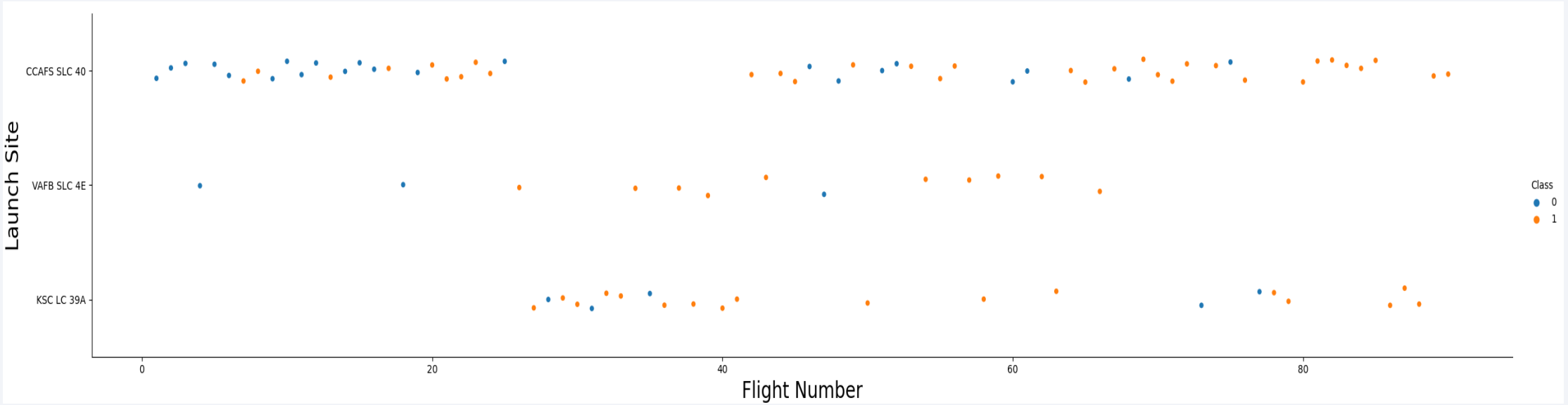


Section 2

# Insights drawn from EDA

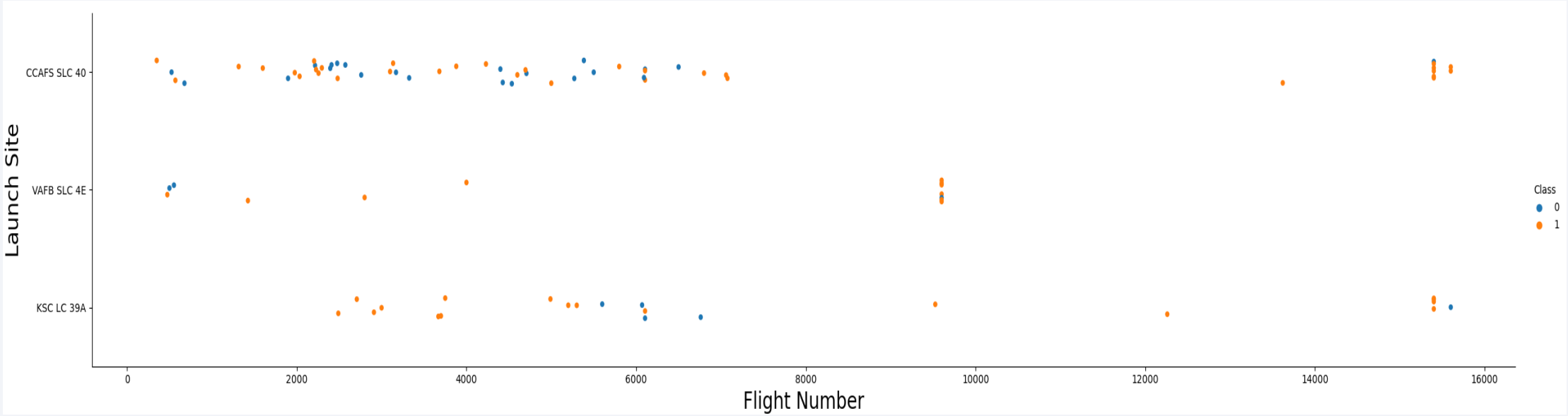


# Flight Number vs. Launch Site



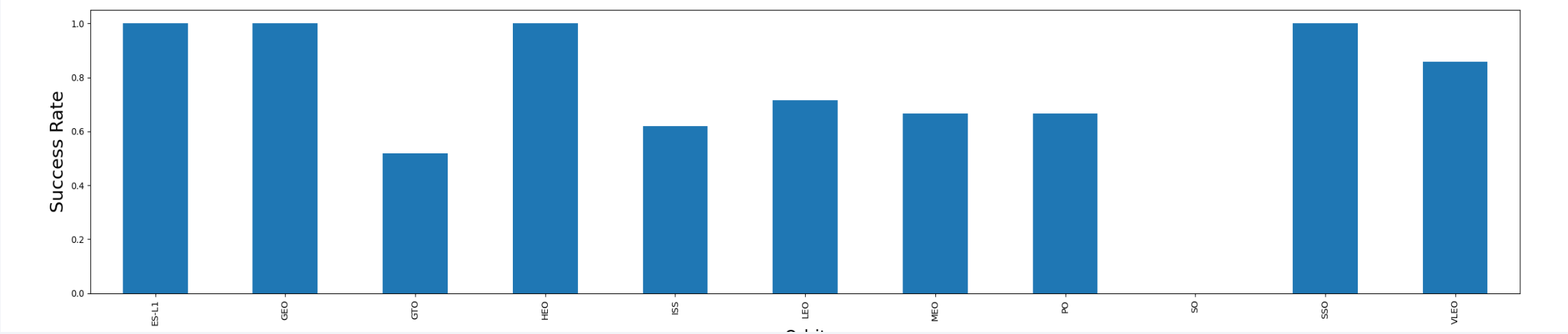

- The best launch site is CCAF5 SLC 40 as most of the recent launches were successful.
- VAFB SLC 4E and KSC LC 39A fall into 2nd and 3rd place respectively.
- There is a possibility that the success rate will improve over time.

# Payload vs. Launch Site



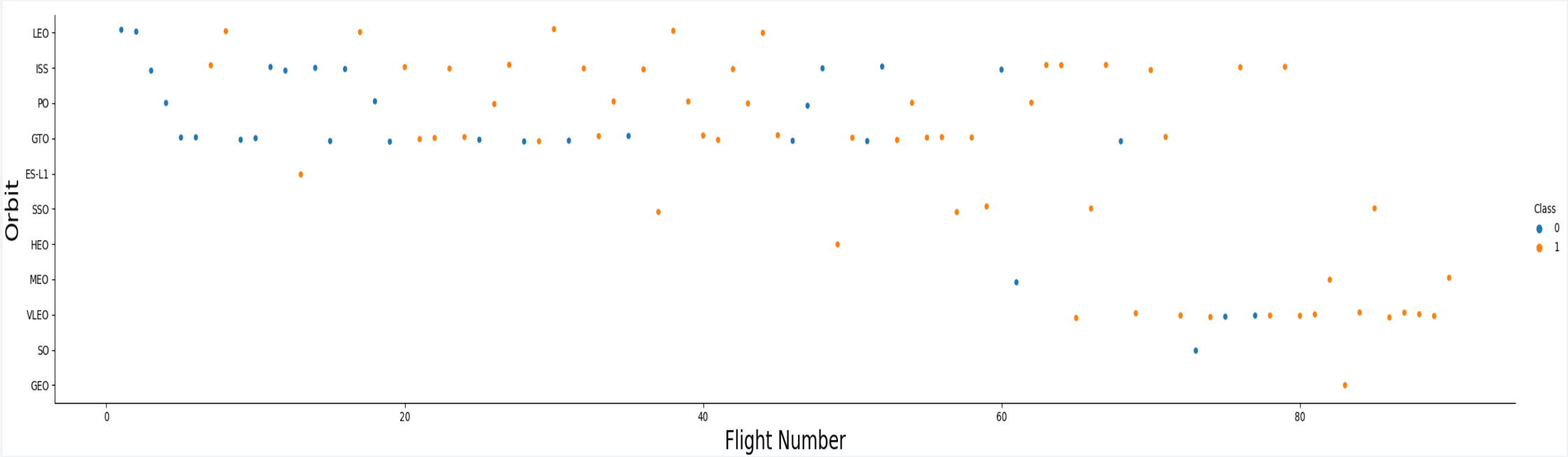

- A high success rate came from payloads over 9000kg.
- CCAFS SLC 40 and KSC LC 39A launch sites consisted of payloads over 120000kg.

# Success Rate vs. Orbit Type




- The highest success rates that orbited are : ES-L1, GEO, HEO and SSO.
- LFO achieved a score of above 70%, followed by VLEO that scored above 80%.

# Flight Number vs. Orbit Type

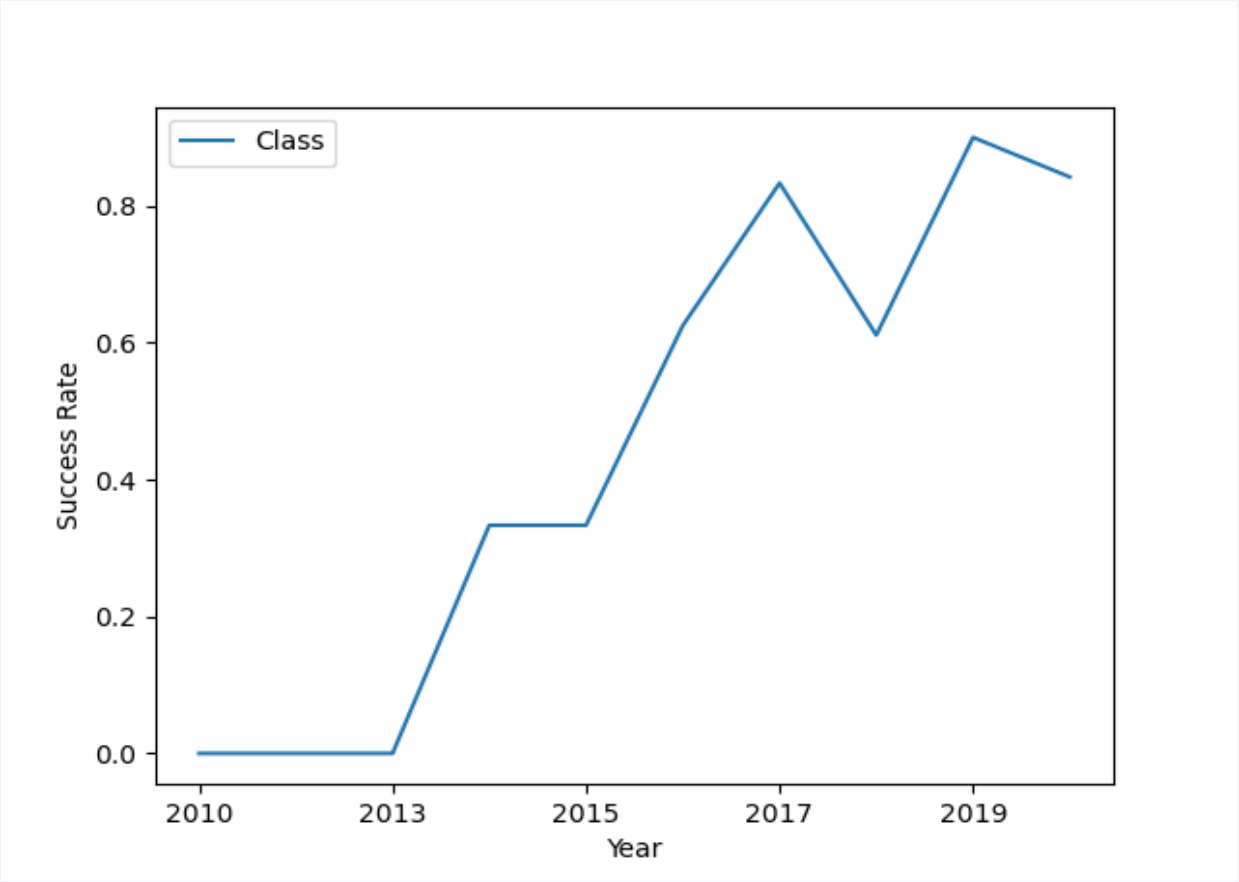



- Over a period of time, the success rate improved..
- Due to a recent increase in frequency for the VLEO orbit, this presents a new business opportunity.

# Payload vs. Orbit Type


- There isn't a correlation between payload and the success rate of GTO orbit.
- The ISS orbit has the widest range of payload that indicates a good rate of success.
- There are minimal launches to the orbits SO and GEO.

# Launch Success Yearly Trend




- The rate of success kept increasing from 2013 to 2020.
- There is an improvement in technology after the first three years of period adjustment.

# All Launch Site Names

- These are all the Launch Site Names.
- %sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL;

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

# Launch Site Names Begin with 'CCA'

- The 5 records where launch sites begin with `CCA` that are launched from Cape Canaveral.
- %sql SELECT LAUNCH_SITE FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

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

# Total Payload Mass

- The Total payload carried by boosters from NASA.
- %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';

| SUM(PAYLOAD_MASS__KG_) |
| --- |
| 45596 |

# Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1
- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';

| AVG(PAYLOAD_MASS__KG_) |
| --- |
| 2928.4 |

# First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad.
- %sql SELECT min(DATE) FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (ground pad)";

| min(DATE) |
| --- |
| 01-05-2017 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.
- %sql SELECT Booster_Version FROM SPACEXTBL WHERE "Landing _Outcome" = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

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

# Total Number of Successful and Failure Mission Outcomes

- This is the total number of successful and failure mission outcomes.
- %sql SELECT COUNT(Mission_Outcome) FROM SPACEXTBL;

| COUNT(Mission_Outcome) |
| --- |
| 101 |

# Boosters Carried Maximum Payload

- These are the names of the booster which have carried the maximum payload mass.
- %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

| 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

- This is the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- %sql SELECT "DATE","Landing _Outcome",count("Landing _Outcome")as LANDING_OUTCOME_COUNT,DATE from SPACEXTBL where substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2) between '20100604' and '20170320' group by "Landing _Outcome" order by count("Landing _Outcome") desc;

| Date | Landing _Outcome | LANDING_OUTCOME_COUNT | Date_1 |
|---|---|---|---|
| 22-05-2012 | No attempt | 10 | 22-05-2012 |
| 08-04-2016 | Success (drone ship) | 5 | 08-04-2016 |
| 10-01-2015 | Failure (drone ship) | 5 | 10-01-2015 |
| 22-12-2015 | Success (ground pad) | 3 | 22-12-2015 |
| 18-04-2014 | Controlled (ocean) | 3 | 18-04-2014 |
| 29-09-2013 | Uncontrolled (ocean) | 2 | 29-09-2013 |
| 04-06-2010 | Failure (parachute) | 2 | 04-06-2010 |
| 28-06-2015 | Precluded (drone ship) | 1 | 28-06-2015 |

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

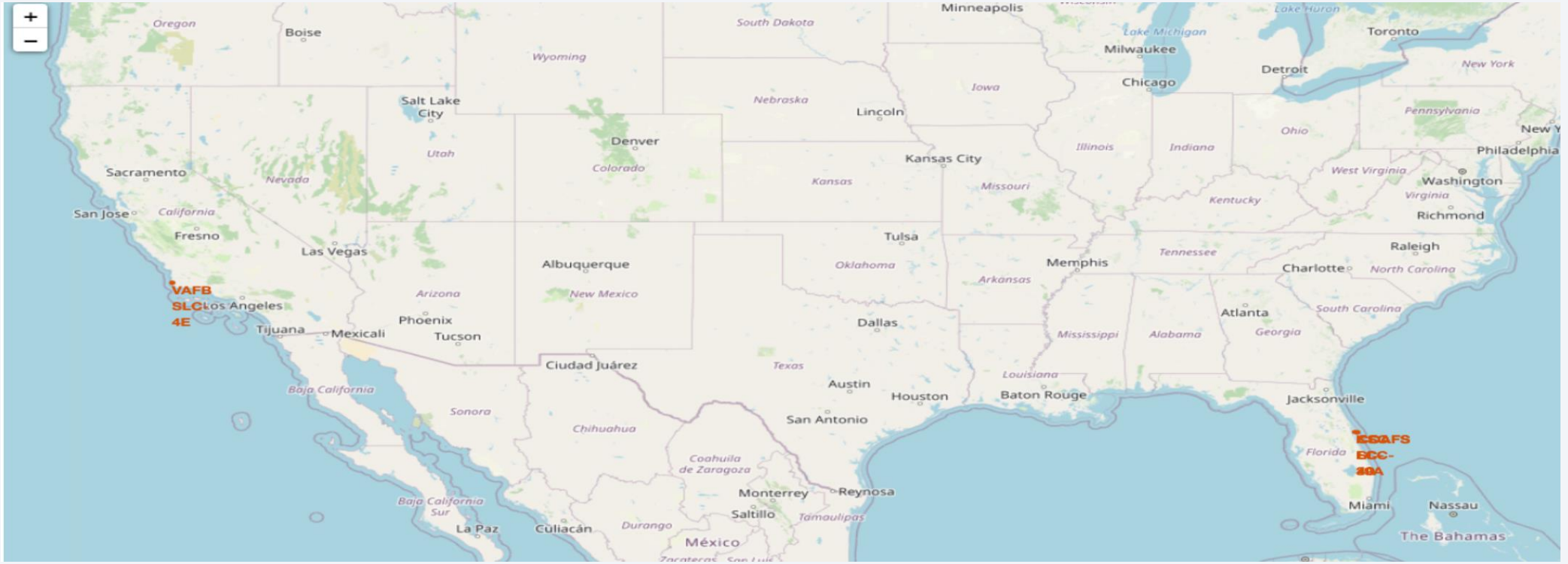
- This is the rank count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- %%sql SELECT "Landing_Outcome",count("Landing_Outcome")as LANDING_OUTCOME_COUNT from SPACEXTBL where DATE between '04-06-2010' and '20-03-2017' group by "Landing_Outcome" order by count("Landing_Outcome") desc;

| Landing _Outcome | LANDING_OUTCOME_COUNT |
|---|---|
| Success | 20 |
| No attempt | 10 |
| Success (drone ship) | 8 |
| Success (ground pad) | 6 |
| Failure (drone ship) | 4 |
| Failure | 3 |
| Controlled (ocean) | 3 |
| Failure (parachute) | 2 |
| No attempt | 1 |


Section 3
Launch Sites
Proximities Analysis

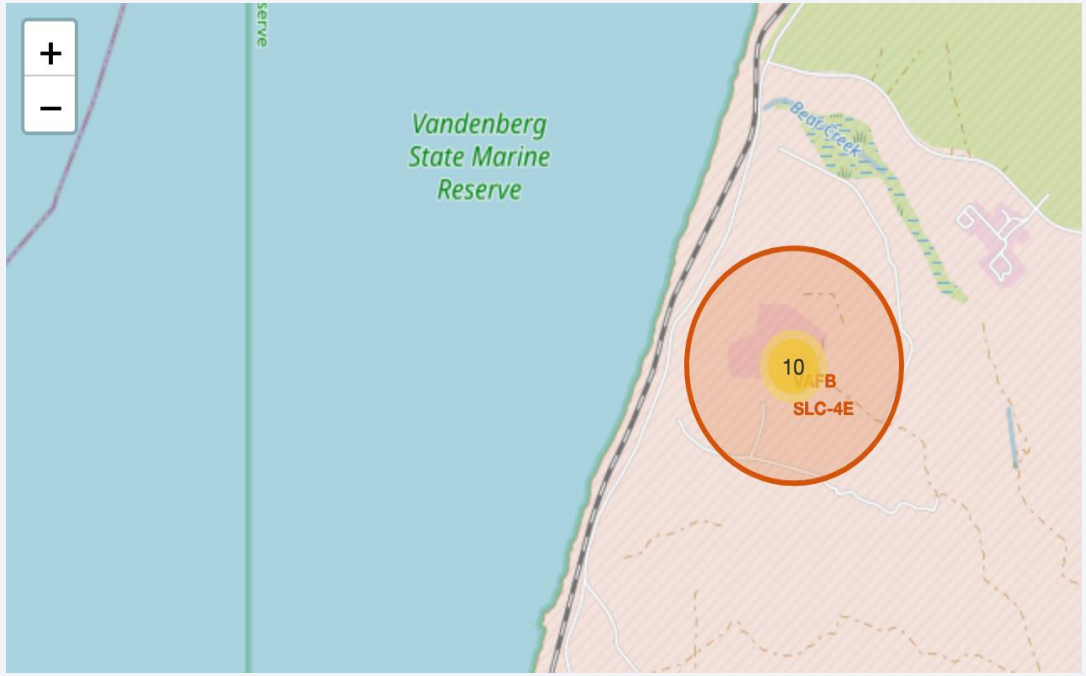
# All launch sites

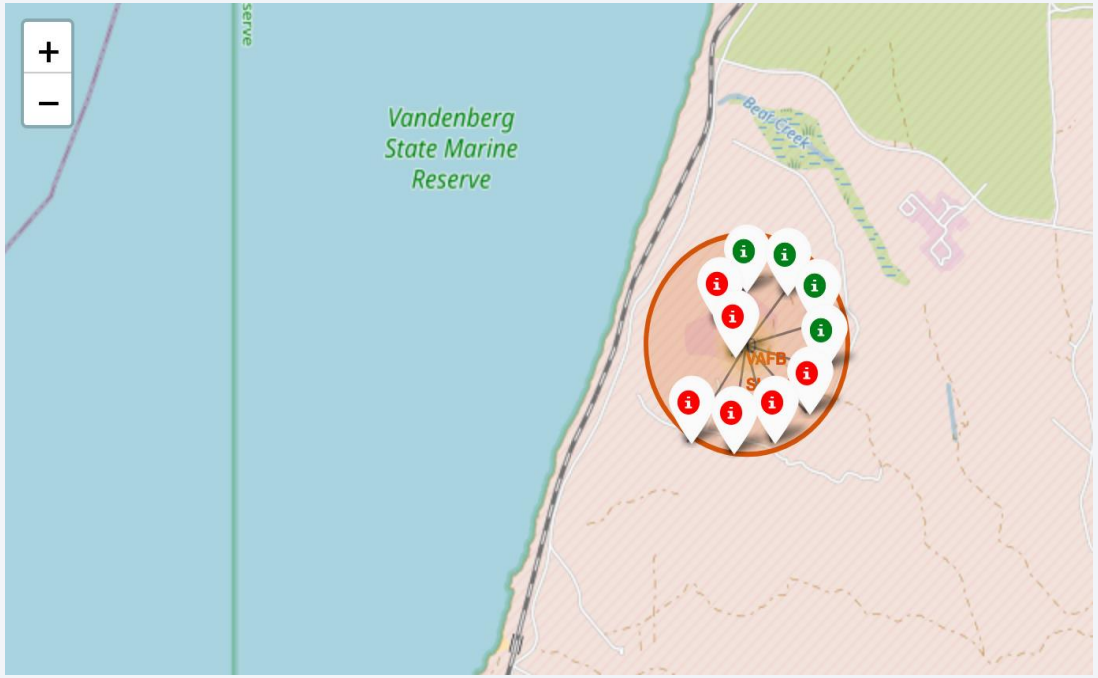
- The launch sites are located near the coast for safety reasons and aren't too far from road and railroad infrastructure.



# Launch Sites by Outcome

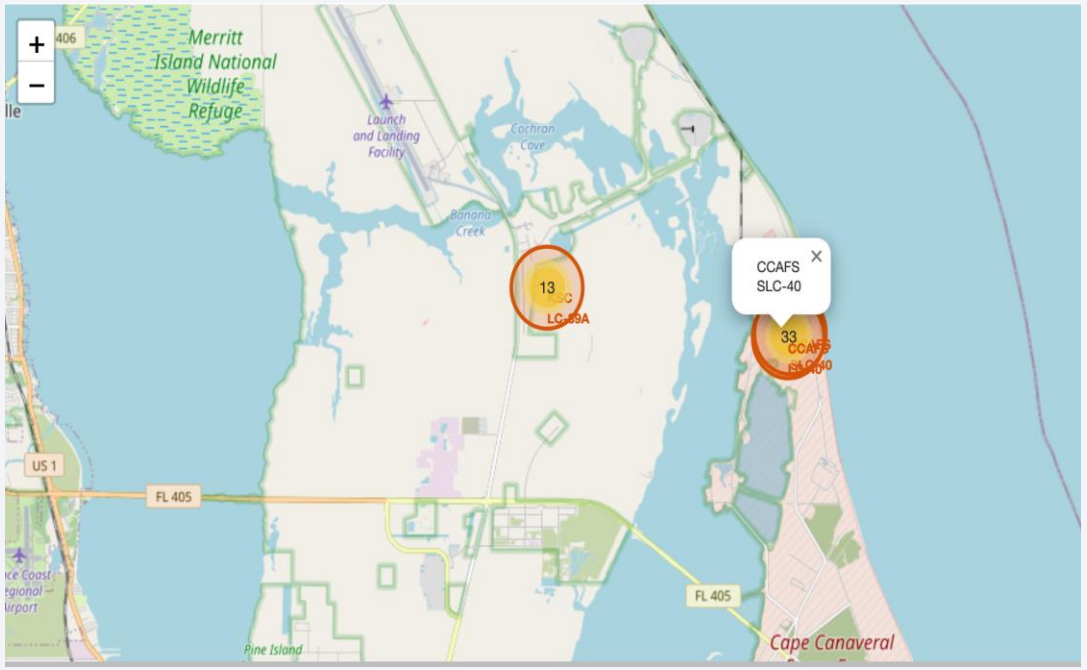
- An example of VAFB SLC 4E launch site.
- Green markers indicate success and red markers indicate failure.

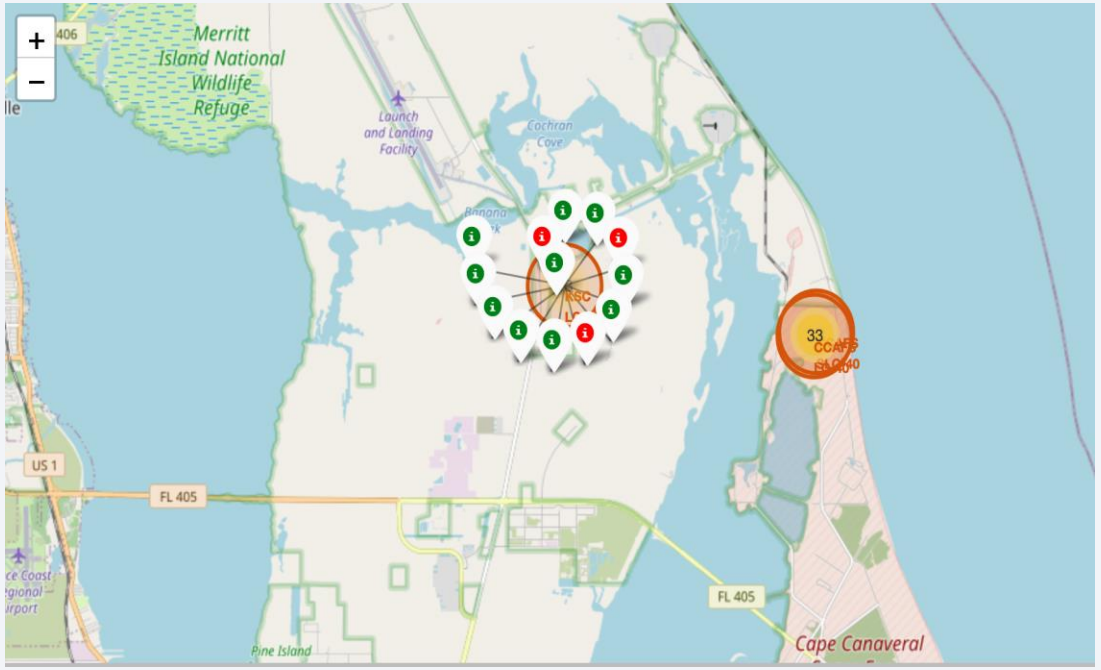


# Launch Sites by Outcome

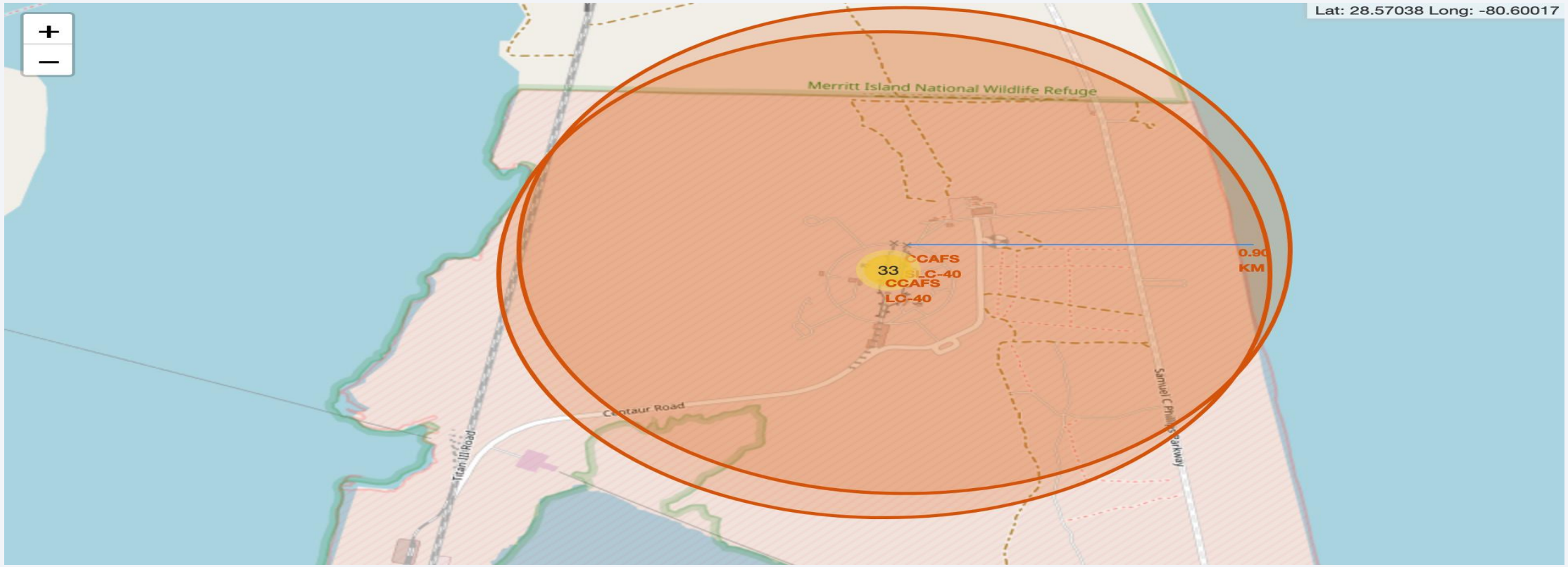
- Examples of launch sites on the east coast.
- Green markers indicate success and red markers indicate failure.

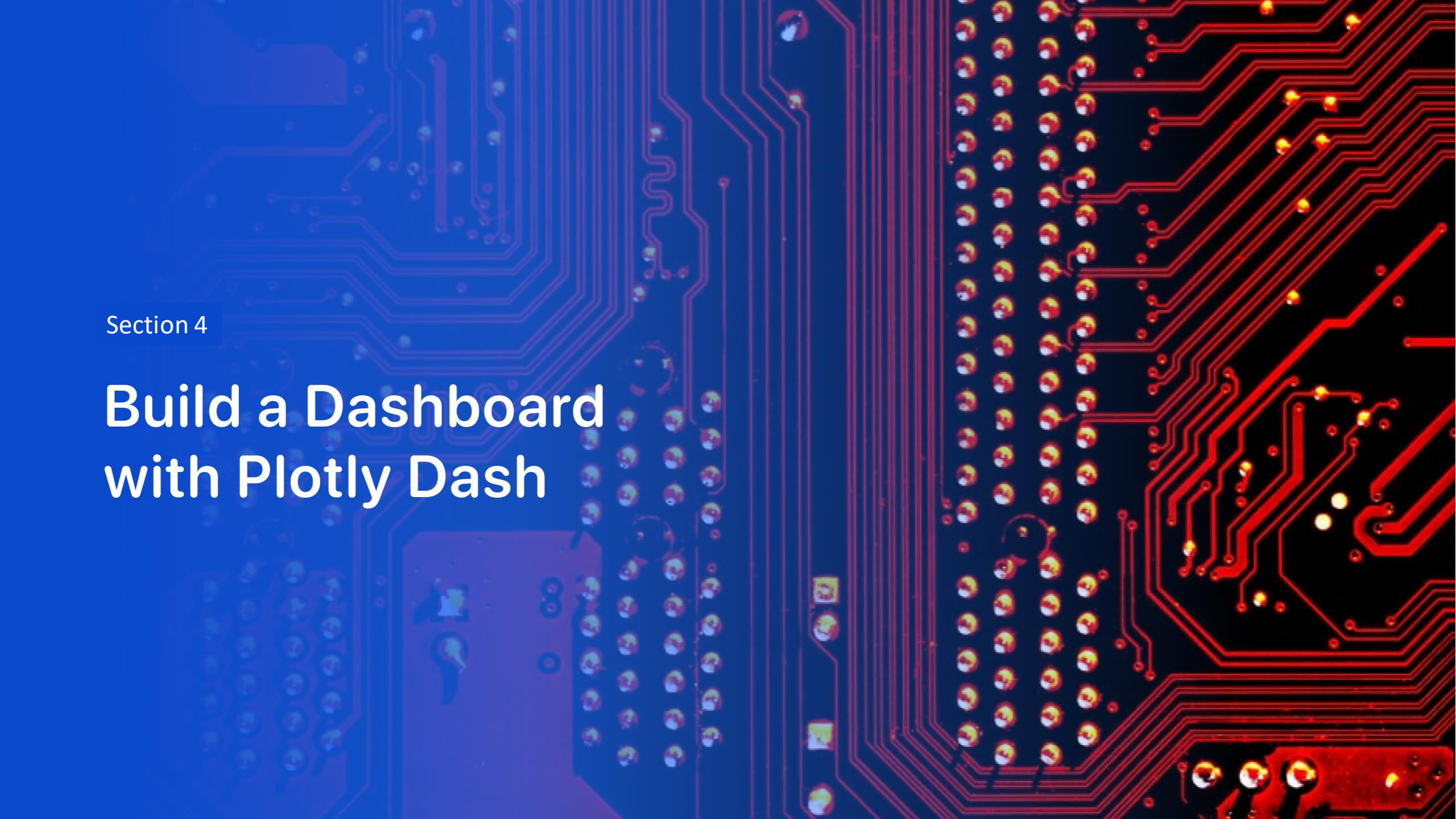


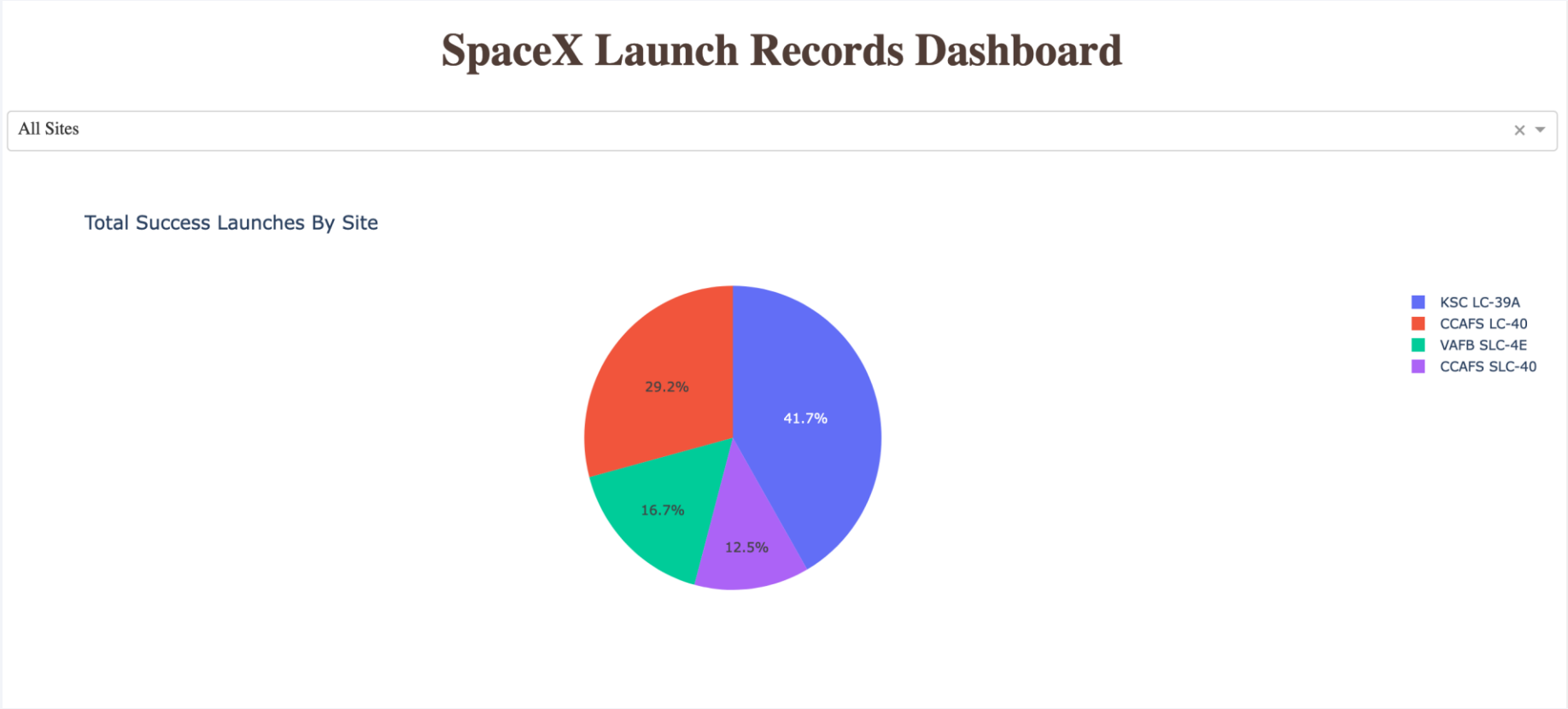
# Logistics and Safety

- A safety distance between the launch site and the coast line



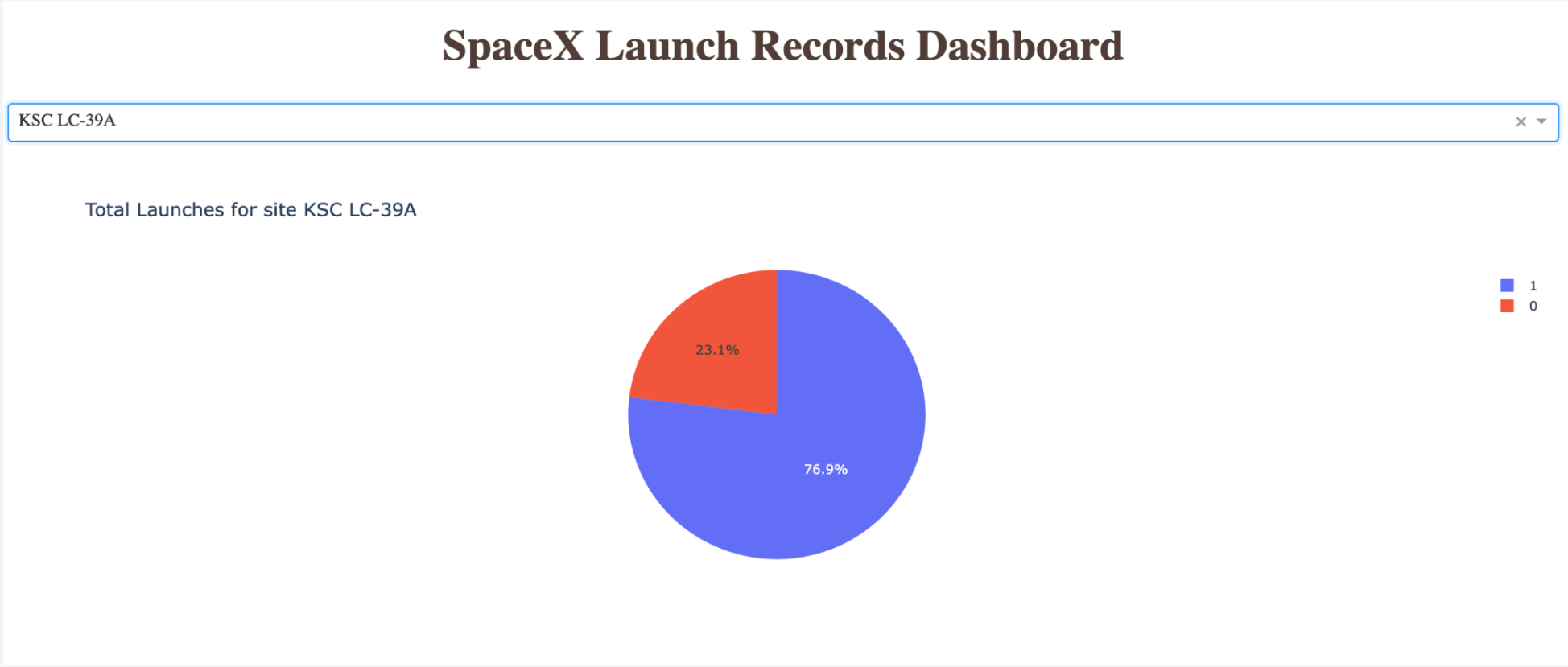

Section 4
Build a Dashboard with Plotly Dash

# Successful Launch by Sites



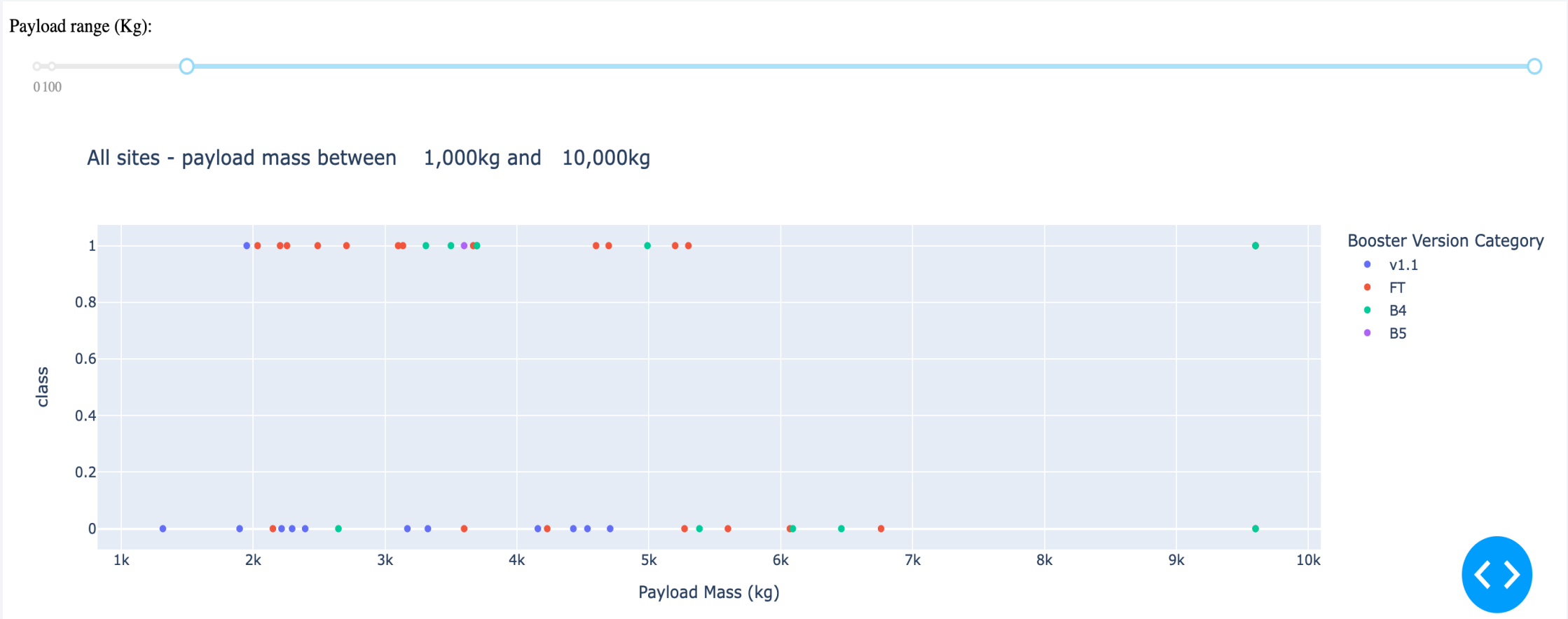

- The location of the launch sites selected is a huge determinant on the success or failure of the outcome.

# Success of Launch for KSC LC-39A




- A success launch rate of 76.9%.

# Payload vs Launch Outcome



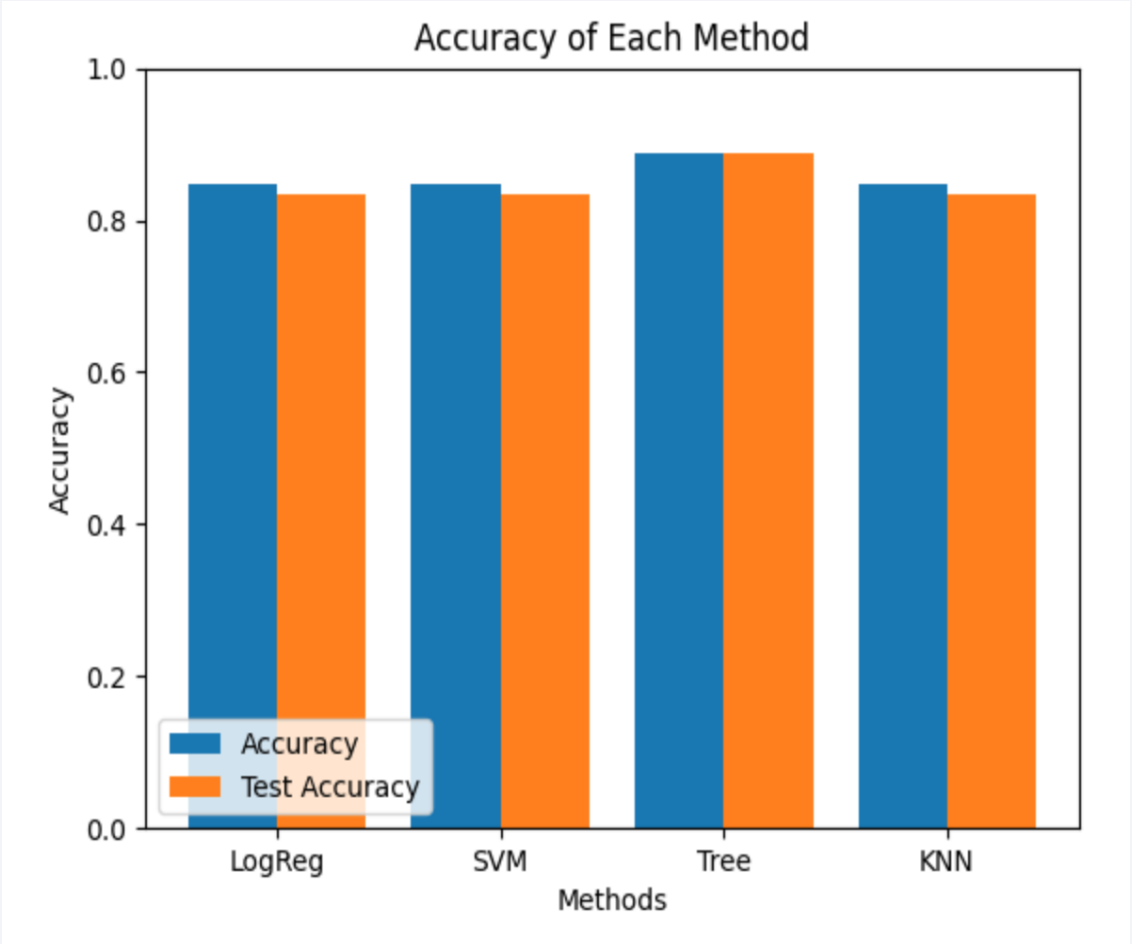

- Payloads under 6000kg with FT boosters show the best combination.

Section 5

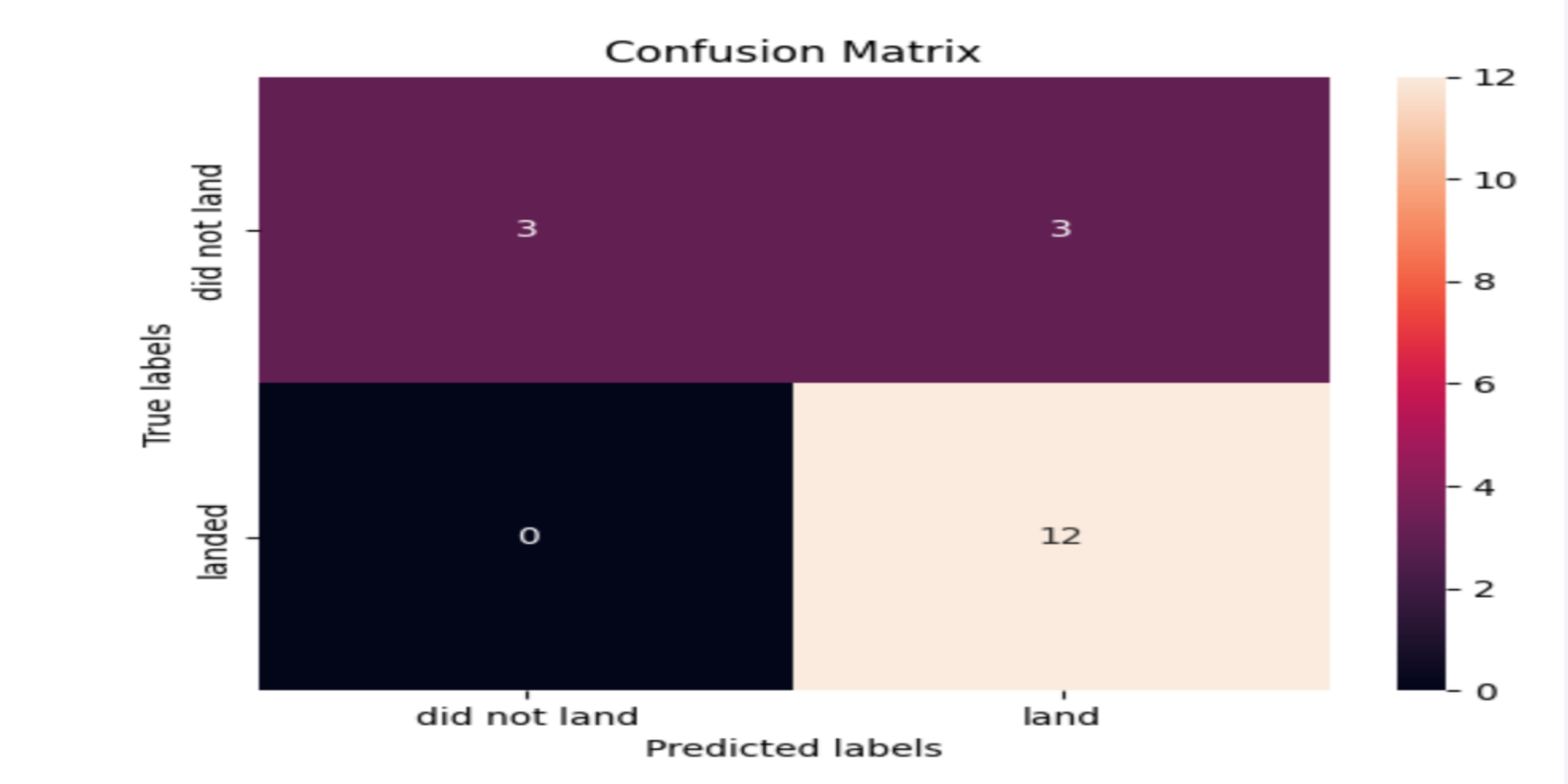
# Predictive Analysis (Classification)

# Classification Accuracy

- 4 model classifiers were built.
- The Decision Tree classifier model has the highest classification accuracy.



# Confusion Matrix




- The best performing model is the Decision Tree Classifier and this is its confusion matrix.

# Conclusions

- All various data sources are collected, cleaned, normalized and analyzed.
- The preferred and best launch site is KSC LC-39A.
- There is less risky approach for launches with payload above 7000kg.
- The best model to be implemented to predict successful landings is the Decision Tree Classifier.
- Given time, technological improvements will evolve as a result of more successful mission landing outcomes.

# Appendix

- Additional folium maps were inserted into this presentation that did not appear in the python script.  These maps were captured via the Skills Network Labs.

# Thank you!