IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Heramb Devarajan
23 April 2023

Outline

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

Executive Summary

- **Summary of all methodologies:**
  - Data Collection using SpaceX API.
  - EDA(Exploratory Data Analysis) including data wrangling, data visualization and interactive visual analytics.
  - Machine Learning Prediction using various classification models.
- **Summary of all results:**
  - It was possible to collected valuable data from public sources.
  - EDA allowed to identify which features are the best to predict success of launchings;
  - Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- The objective of this project is to predict the landing of the SpaceX launches.
- Problems you want to find answers:
    - Where is the best place to launch the rocket.
    - Predicting the percentage of successful landing of the first stage of rockets to maximize profit.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from SpaceX acquired from SpaceX API(https://api.spacexdata.com/v4/rockets/)
- Performed data wrangling
  - Collected Data was cleaned and wrangled and nan values were replaced with mean.
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Data was normalized and then distributed into training and test set and evaluated by 4 classification models and accuracy were determined.

Data Collection

- Data sets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) as json file.
- The file was then converted to a CSV file for easier implementation of pandas functions.

Data Collection – SpaceX API

- SpaceX offers a public API where we can obtain the landing data from and use it for our data science project.
- The way SpaceX API was used is described using the flowchart on the right.

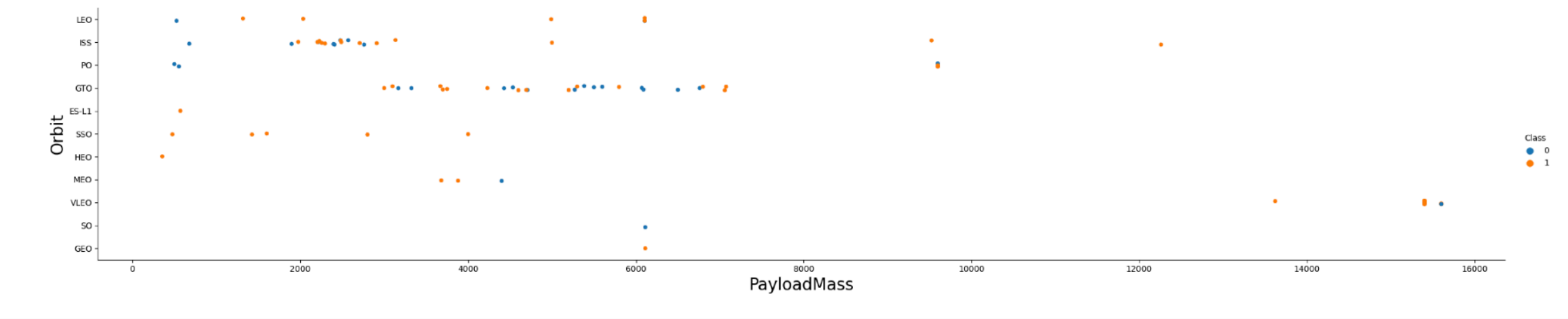
- Source Code: https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/SpaceX%20Data%20Collection.ipynb

Data Wrangling

- Data was preprocessed and cleaned.
- Data which contained nan values had them replaced with the mean.
- Columns which were not required were removed and some columns were added.
- Source Code: https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/SpaceX%20Data%20Wrangling.ipynb

EDA with Data Visualization

- Bar plots and Scatter plots were used to visualize the relationship between categories and also the launch sites were compared.
  - For example Orbit and Payload were compared to see if they have any relationship.



- Source code: https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/EDA%20with%20data%20Visualization.ipynb

EDA with SQL

- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission.
  - Top 5 launch sites whose name begin with the string ‘CCA’.
  - Total payload mass carried by boosters launched by NASA (CRS).
  - Average payload mass carried by booster version F9 v1.1.
  - Date when the first successful landing outcome in ground pad was achieved.
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg.
  - Total number of successful and failure mission outcomes.
  - Names of the booster versions which have carried the maximum payload mass.
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
  - Rank 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.
  - Source code:https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/SpaceX%20Exploratory%20Data%20Analysis.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps

- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and
- Lines are used to indicate distances between two coordinates.

- Source code:https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

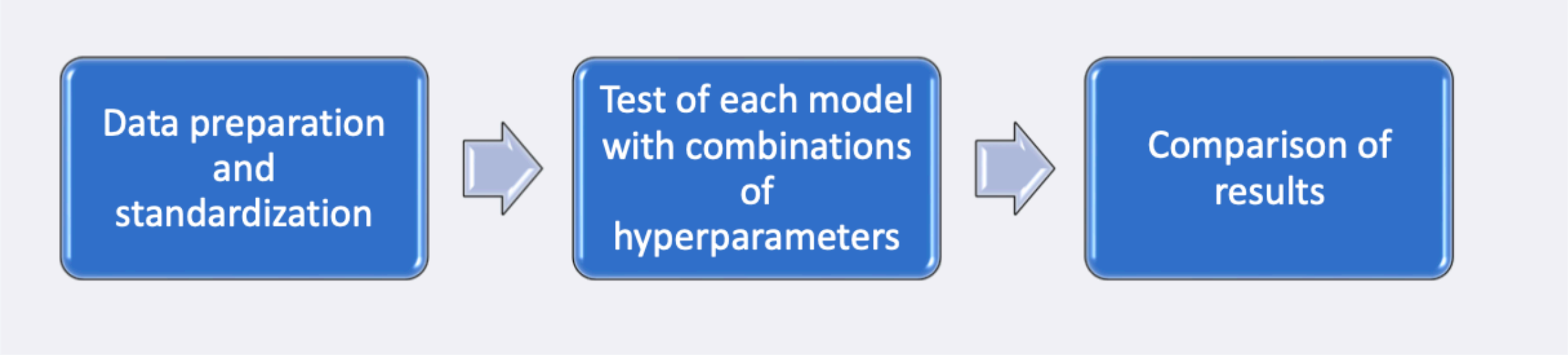
Build a Dashboard with Plotly Dash

- The following plots and graphs were used to visualize the given data.
    - Payload Range
    - Percentage of launches by launch site
- These plots and graphs helped me to quickly analyze the relationship between payload ranges and launch sites and helped determine the success rate of each payload in a given launch site. This led to identification of best places to launch a given payload.

- Source code:https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/spacex_dash_app.py

Predictive Analysis (Classification)

- Four different classification models were compared: logistic regression, decision tree, k nearest neighbors and support vector machine.



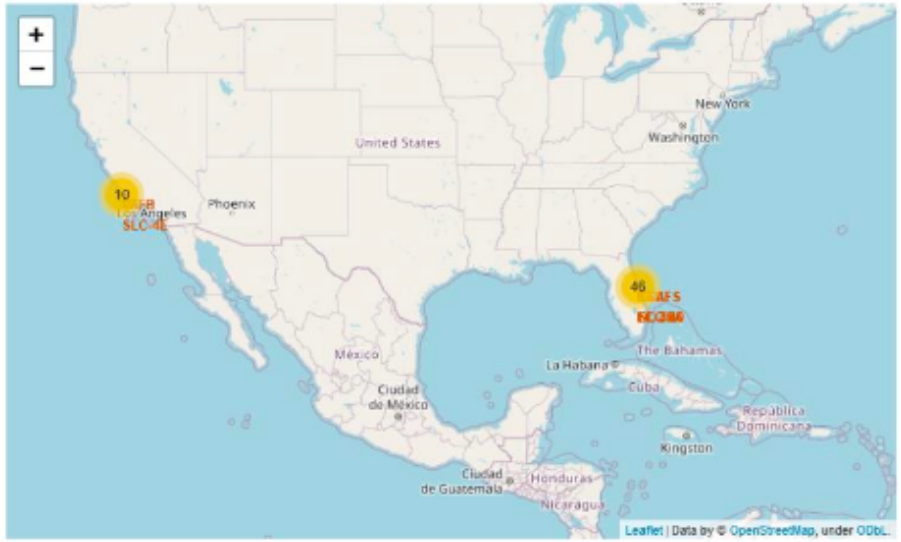
- Source code: https://github.com/HerambDevarajan/SpaceX-Landing-Prediction/blob/6ba572e16cf08ce36a1d86000f0e7b0914e27e1d/Machine%20Learning%20Prediction.ipynb

Exploratory Data Analysis Results

- **Exploratory data analysis results:**
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

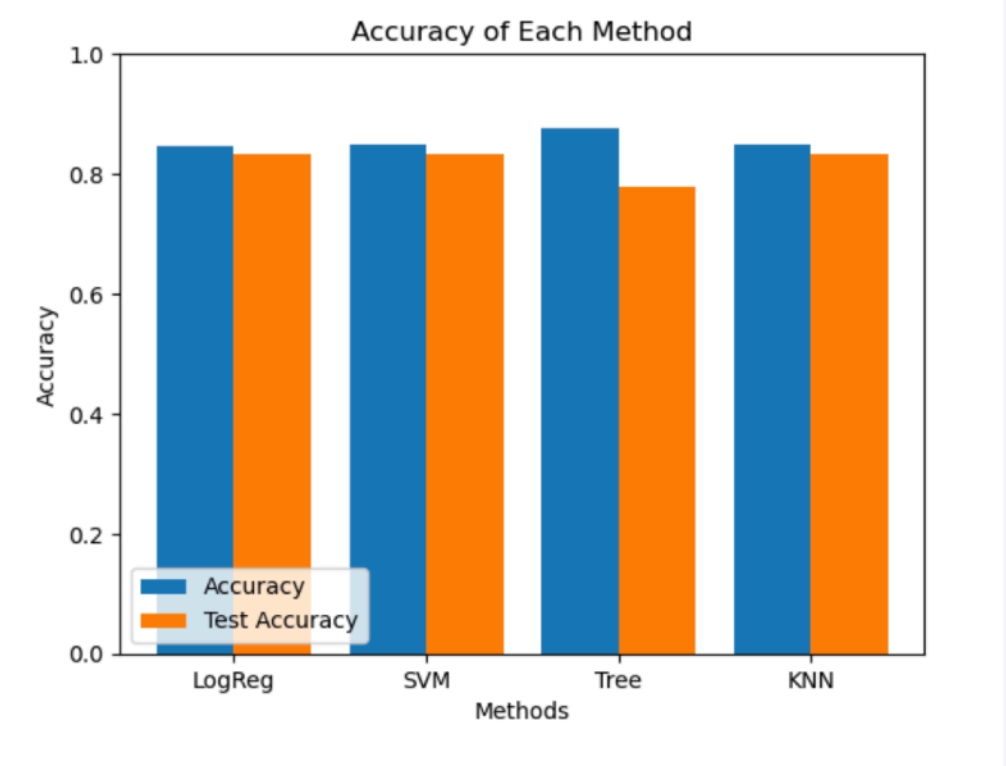
Interactive Analytics Results

- **Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.**
- **Most launches happen in the east coast launch sites.**



Predictive Analysis Results

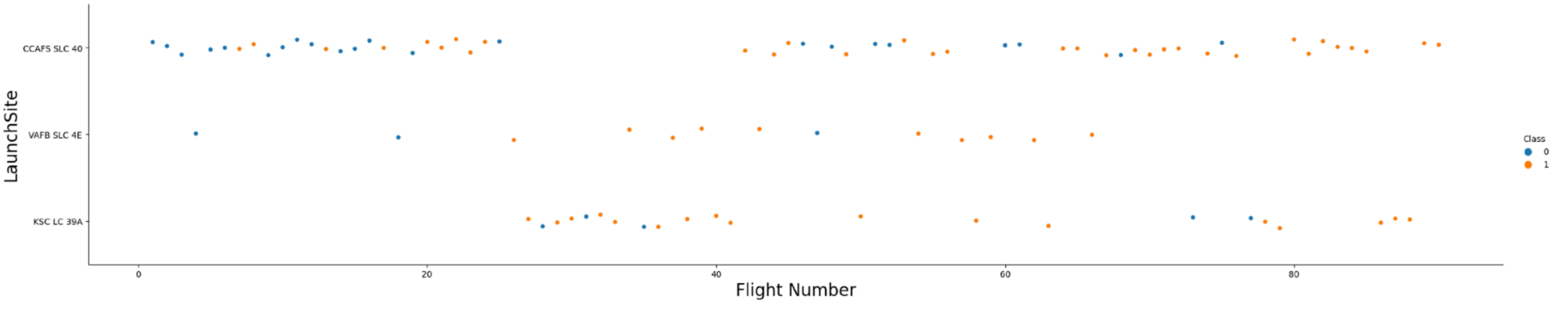
- **Predictive Analysis shows that Decision tree classifier is the best model to predict the landing as it has 88% accuracy compared to other models having 84%.**



Section 2

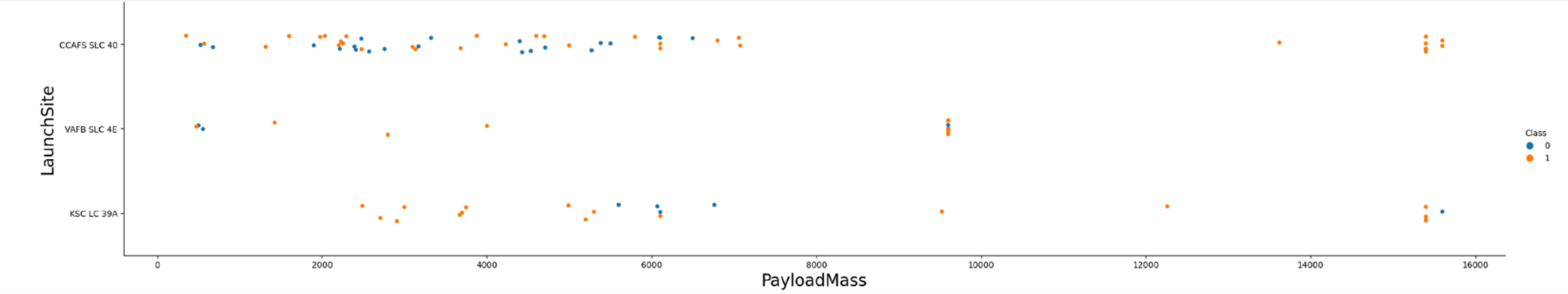
Insights drawn from EDA

Flight Number vs. Launch Site



- This is a scatter plot of Flight Number vs. Launch Site
- We can see that the recent launches of CCAFS are more successful making it the best place to launch rockets.
- We can see that VAFB is the next best alternative.

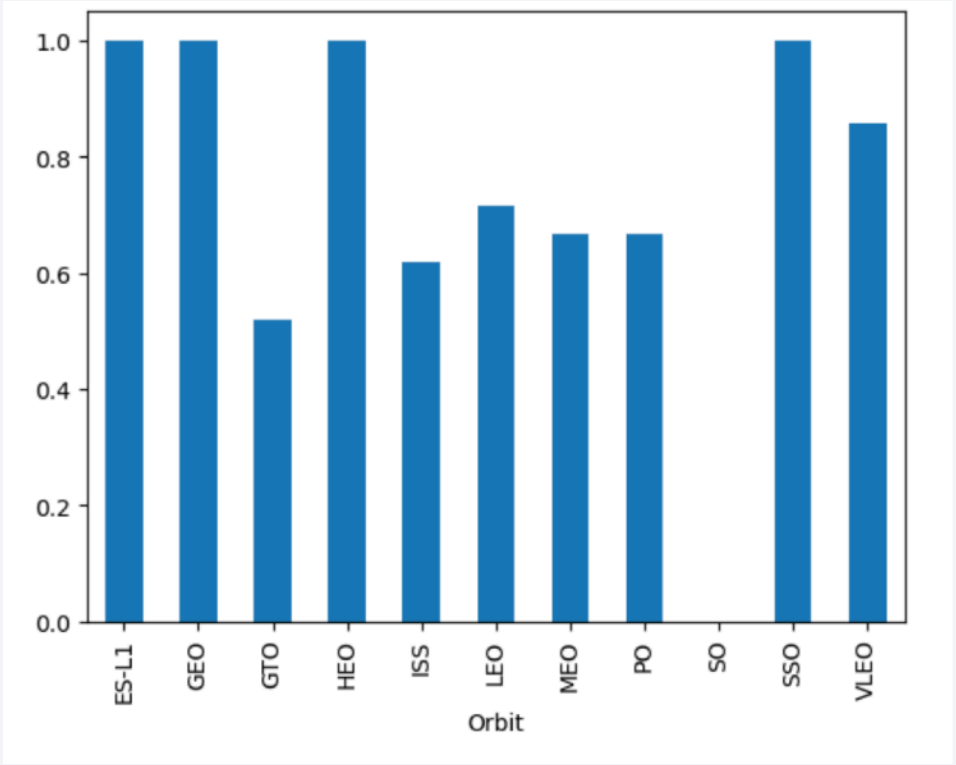
Payload vs. Launch Site



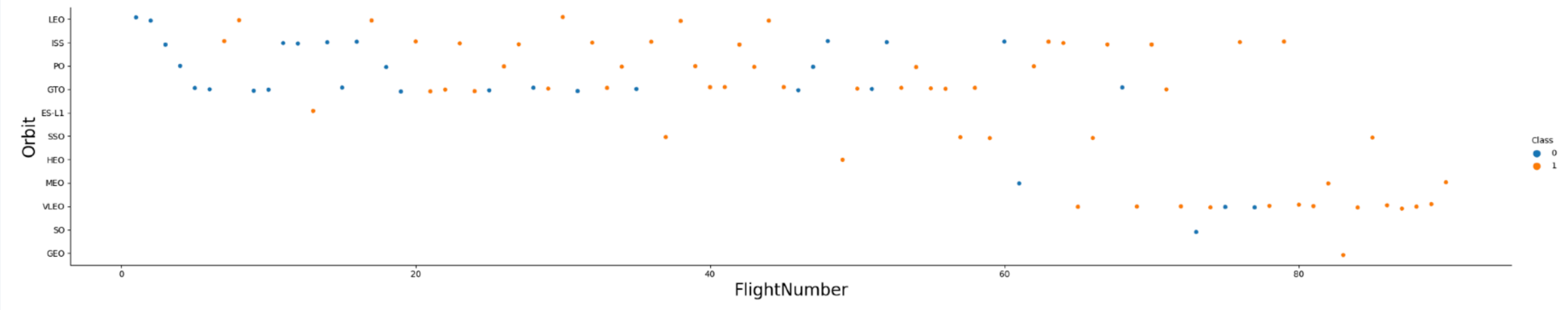
- This is a scatter plot of Payload vs. Launch Site
- We can see that payload with more than 10,000kg is not possible in VAFB.
- We can see that payload with more than 8,000kg have very high success rates.

Success Rate vs. Orbit Type

- This is a bar chart between success rate and orbit to help us determine which orbit is most successful.
- ES-L1, GEO, SSO, and HEO have the biggest success rate.

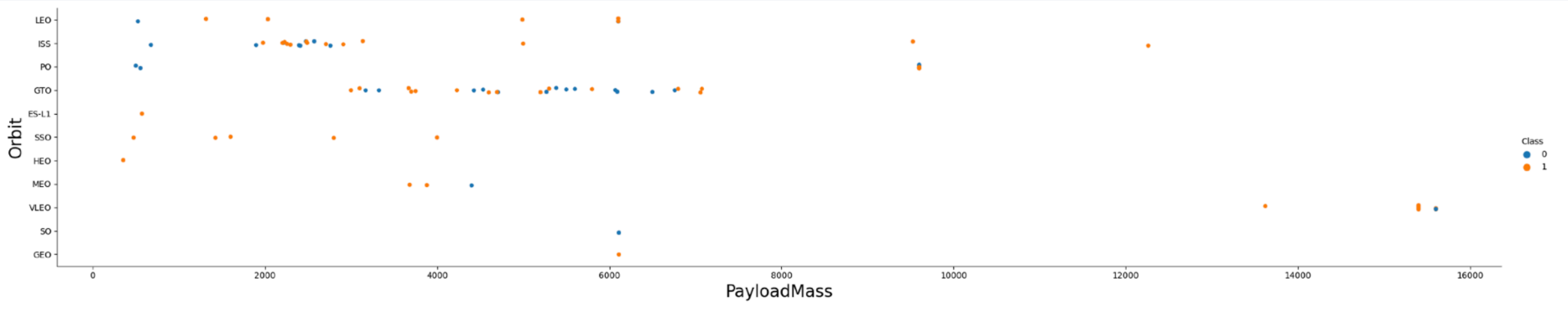


Flight Number vs. Orbit Type



- This is a scatter plot of Flight number vs. Orbit type
- The success rate of each orbit improved over time.
- VLEO is the most sought after orbit recently as it has many new launches.

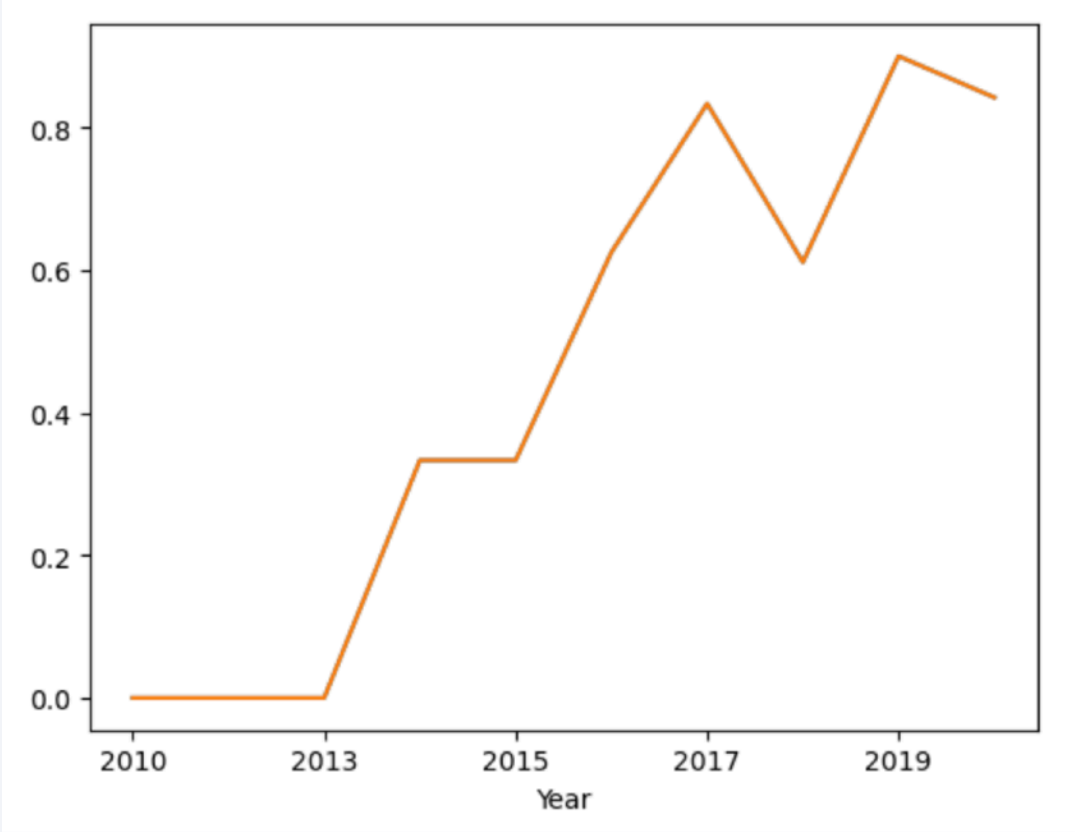
Payload vs. Orbit Type



- This is a scatter plot of payload vs. orbit type
- There is no relation between payload and orbit type.
- GTO and ISS have the most payloads.

Launch Success Yearly Trend

- The success rate has been increasing since 2013.
- We can also see that the success rates have fallen a little bit in 2018.
- We can also see that there was no success from 2010 to 2013.



All Launch Site Names

- There are 4 launch sites:

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

- These were obtained by selecting the distinct occurrences of launch sites.

Launch Site Names Begin with 'CCA'

- These are 5 records where launch sites begin with `CCA`

DATE	TIME_UTC	BOOSTER_VERS	LAUNCH_SIT	PAYLOAD	PAYLOAD_	ORBIT	CUSTOMER	MISSION_	LANDING_OUTCOME
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spa	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon der	0	LEO (IS	NASA (COTS)	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon der	525	LEO (IS	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CR	500	LEO (IS	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CR	677	LEO (IS	NASA (CRS)	Success	No attempt

- These were obtained by selecting the launch sites which started with CCA.

Total Payload Mass

- The total payload carried by boosters from NASA:

TOTAL_PAYLOAD
111268

- This was obtained by calculating the sum of all payload which contained CRS (NASA).

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1:

AVG_PAYLOAD_MASS
2928

- This was obtained by calculating the mean of all mass carried by F9 v1.1.

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad:

FIRST_SUCCESS
2015-12-22

- This was found by finding the minimum date where successful landing was registered.

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:

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

- These were obtained by selecting the boosters between 4000 and 6000 kg which landed on drone ship.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes:

MISSION_OUTCOME	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- These were obtained by calculating the count of grouped mission outcome.

Boosters Carried Maximum Payload

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

BOO	STE	R_VERSION
F9	B5	B1048.4
F9	B5	B1048.5
F9	B5	B1049.4
F9	B5	B1049.5
F9	B5	B1049.7
F9	B5	B1051.3

BOO	STE	R_VERSION
F9	B5	B1051.4
F9	B5	B1051.6
F9	B5	B1056.4
F9	B5	B1058.3
F9	B5	B1060.2
F9	B5	B1060.3

- These were obtained by checking which boosters had the max payload.

2015 Launch Records

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

BOOSTER_VERSION	LAUNCH_SITE
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- These were obtained by selecting failed outcomes in 2015.

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

- The ranked 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
- These were obtained by ordering the grouped landing outcome in descending.

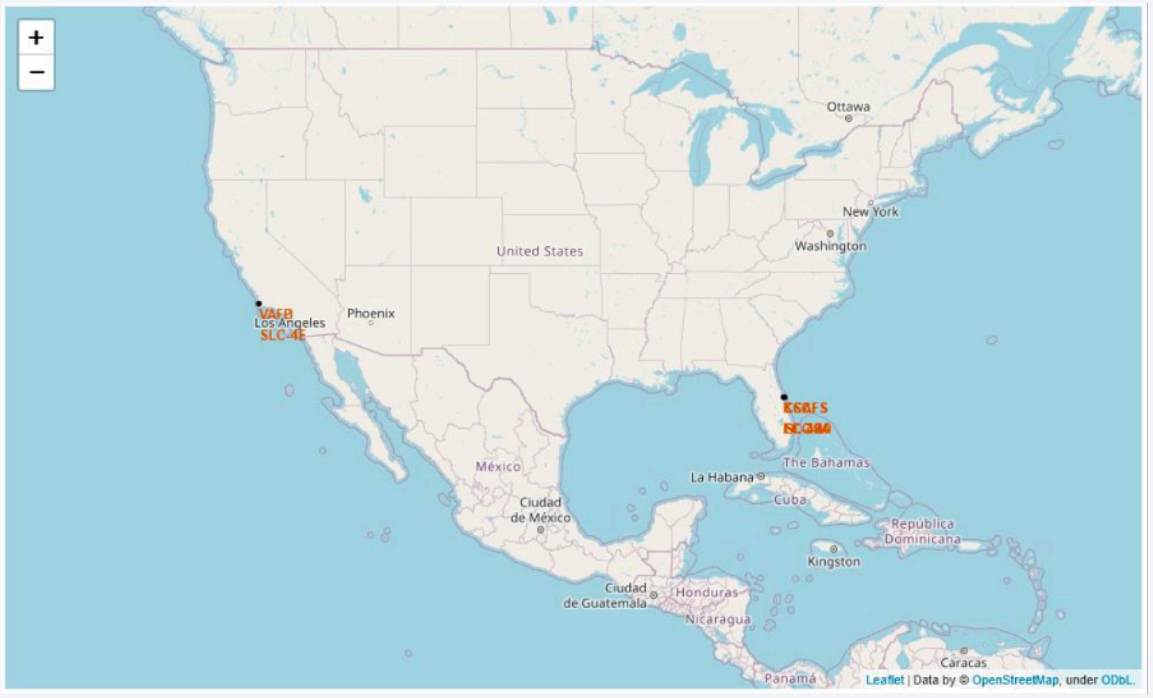
LANDING_OUTCOME	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Sites in USA

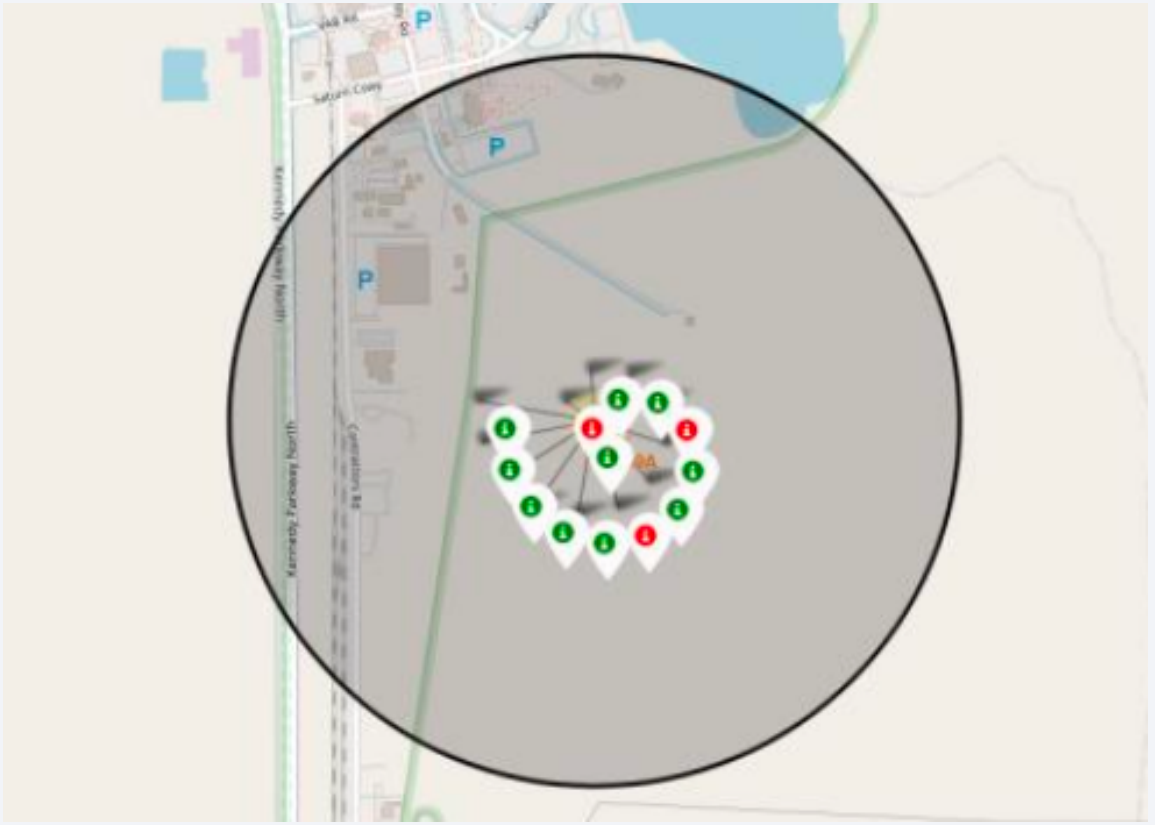
- The launch sites are in the west and east coast of USA nearby sea.
- They are well connected with roads and railway system.
- They majorly used launch site is in the east.



Launch Outcomes in site KSC LC-39A

- We can see that in site KSC LC-39A, the successful launches are marked by the color green and unsuccessful launches are marked by the color red.

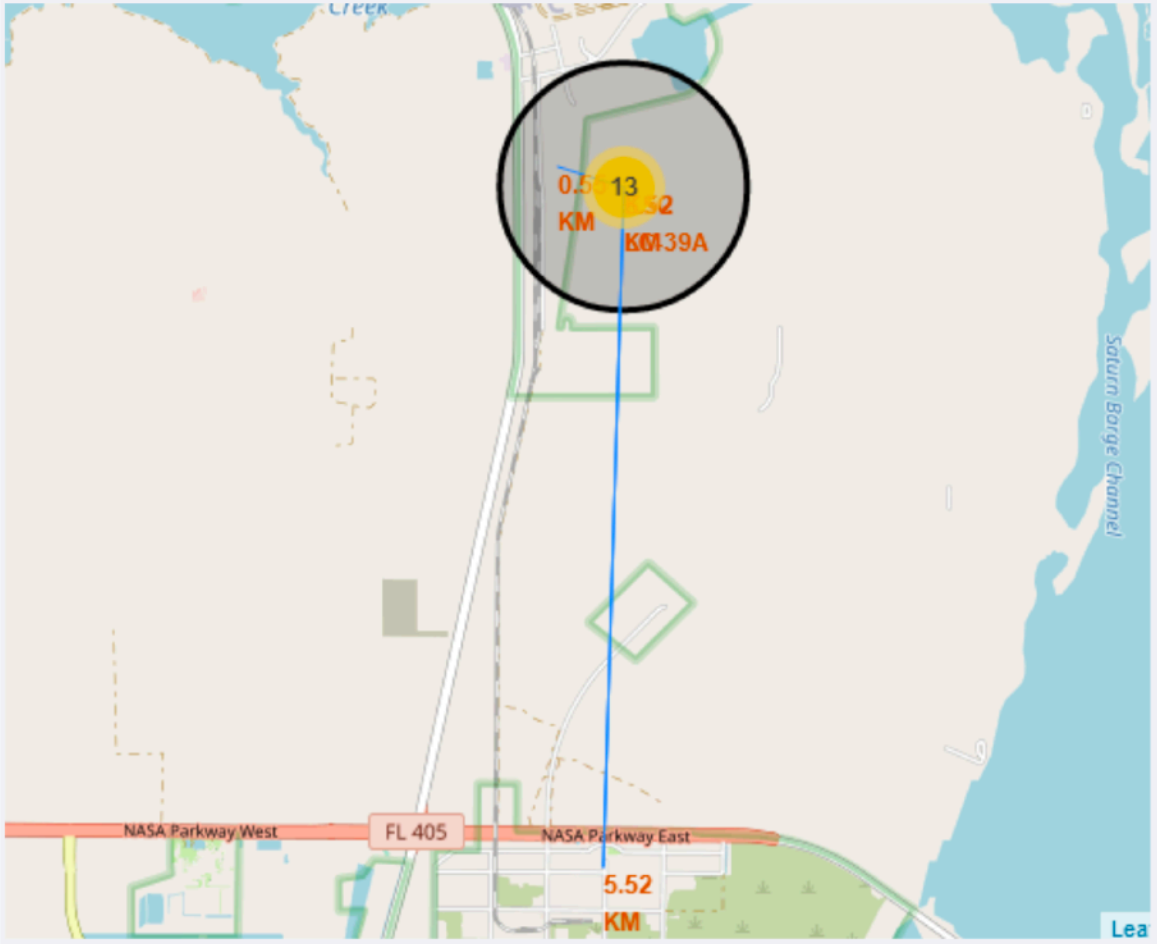
- We can see that the maximum of launched are successful in this site.

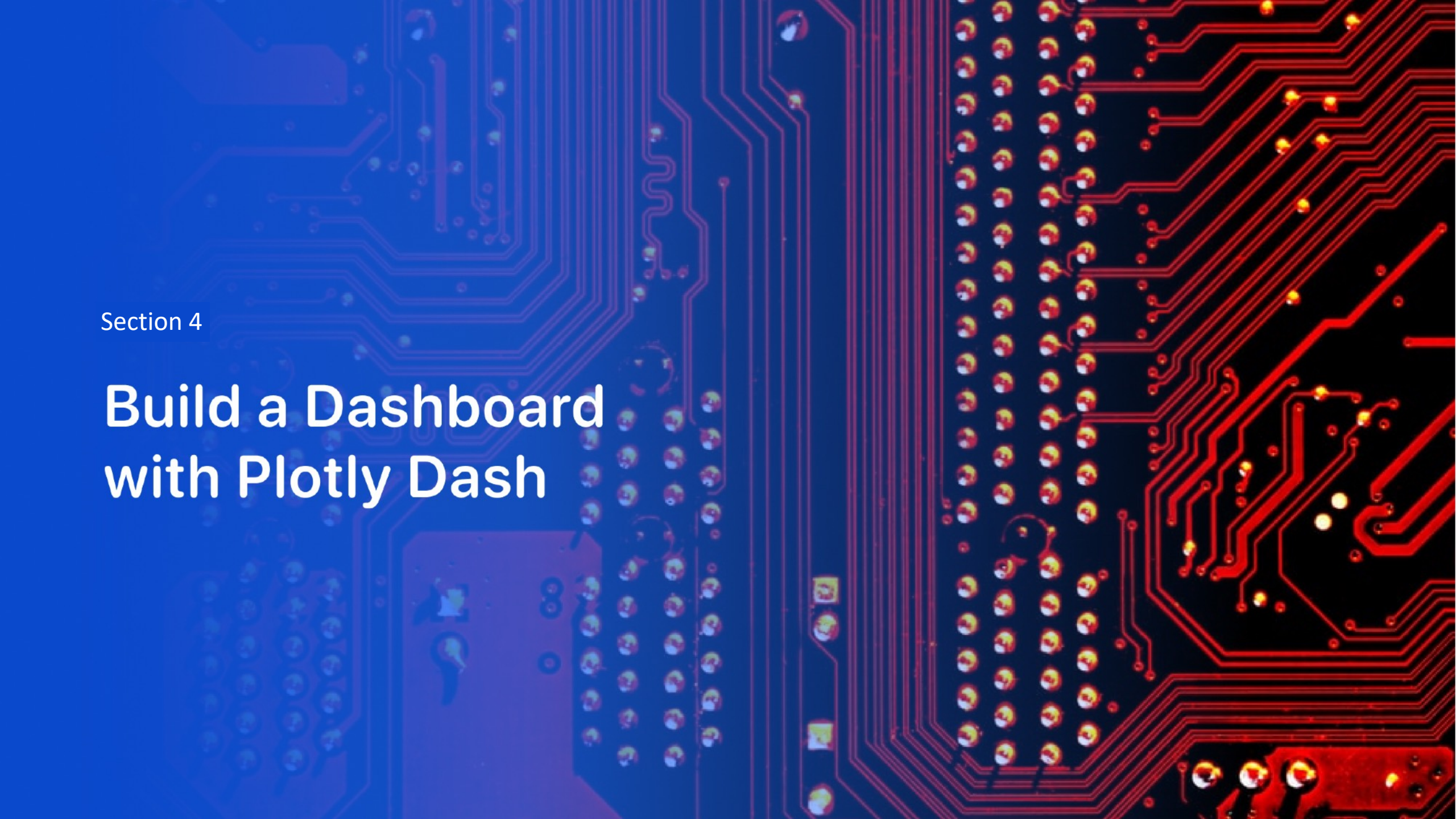


Connectivity of launch site KSC LC-39A

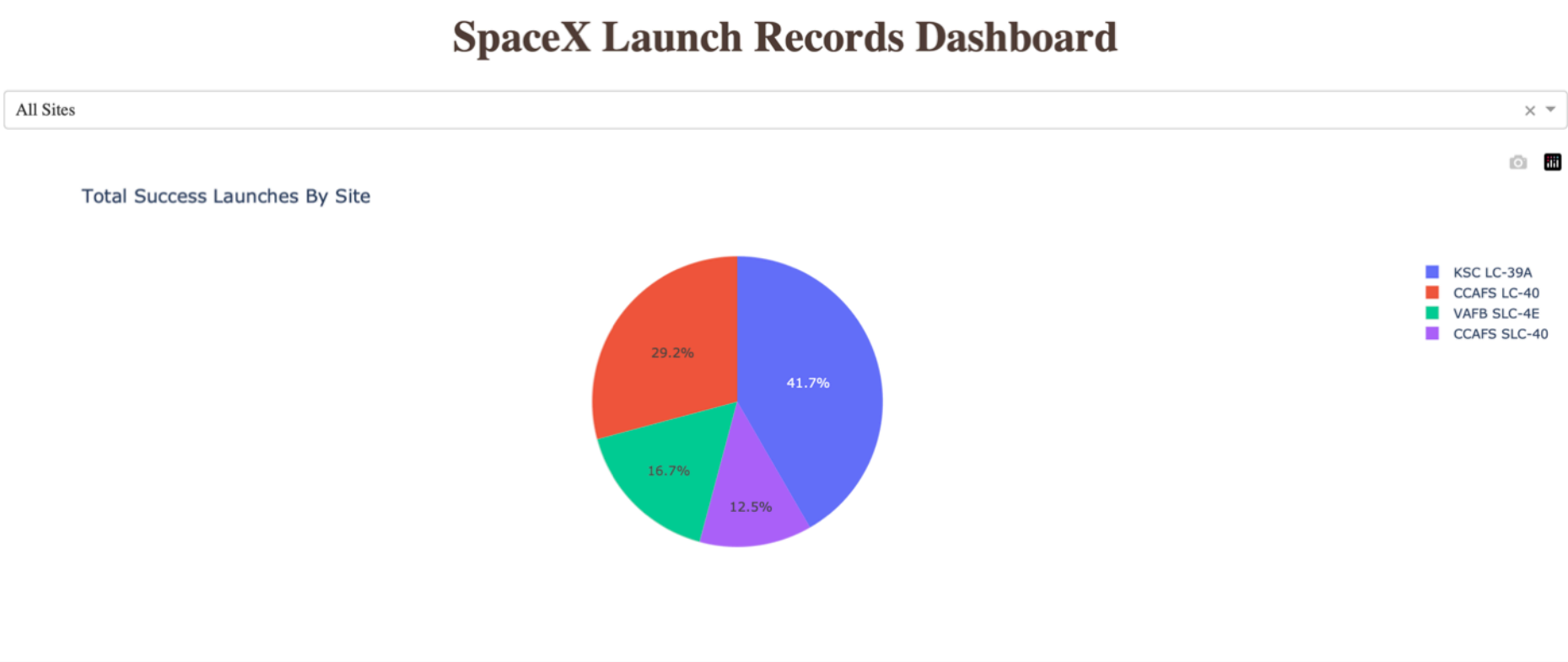
- We can see the distance between roads and railways from the launch site.

- We can see that the launch site is well connected with transport system and also near water.



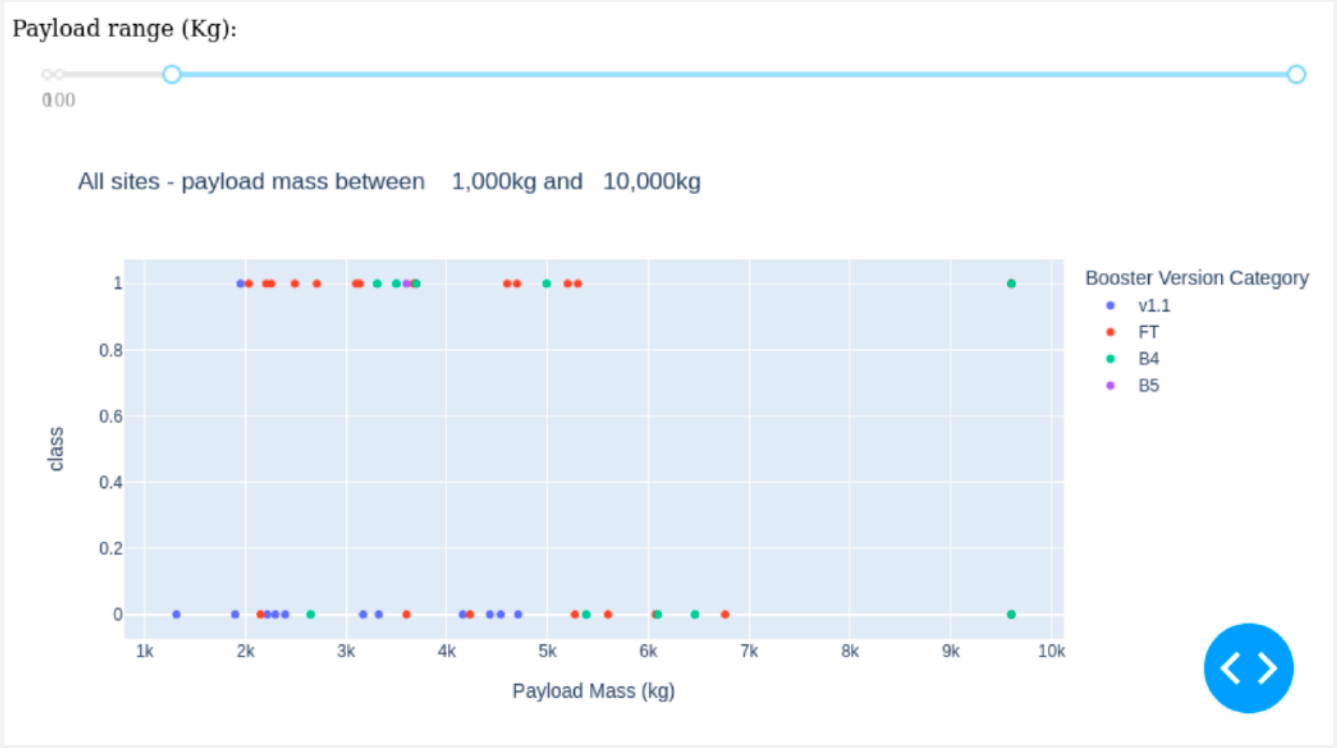

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by site



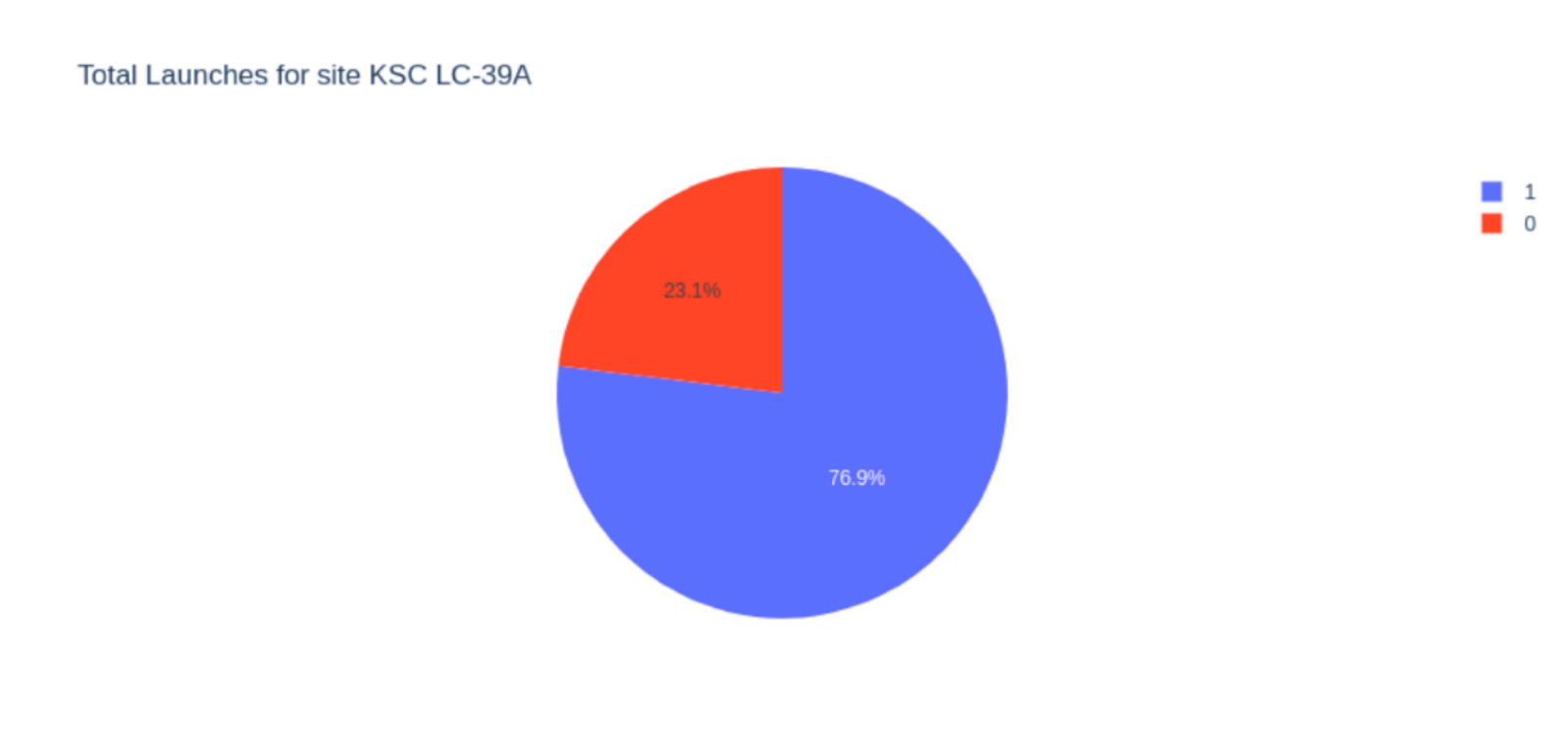
- We can see that KSC LC-39A has the most number of successful launches.

Success of Booster with payload of 1k kg to 10k kg



- We can see that FT booster is the most successful when the payload is between 1,000kg and 10,000kg.

Success rate for KSC LC 39A



- We can see that the most successful site has success rate of 76.9%.

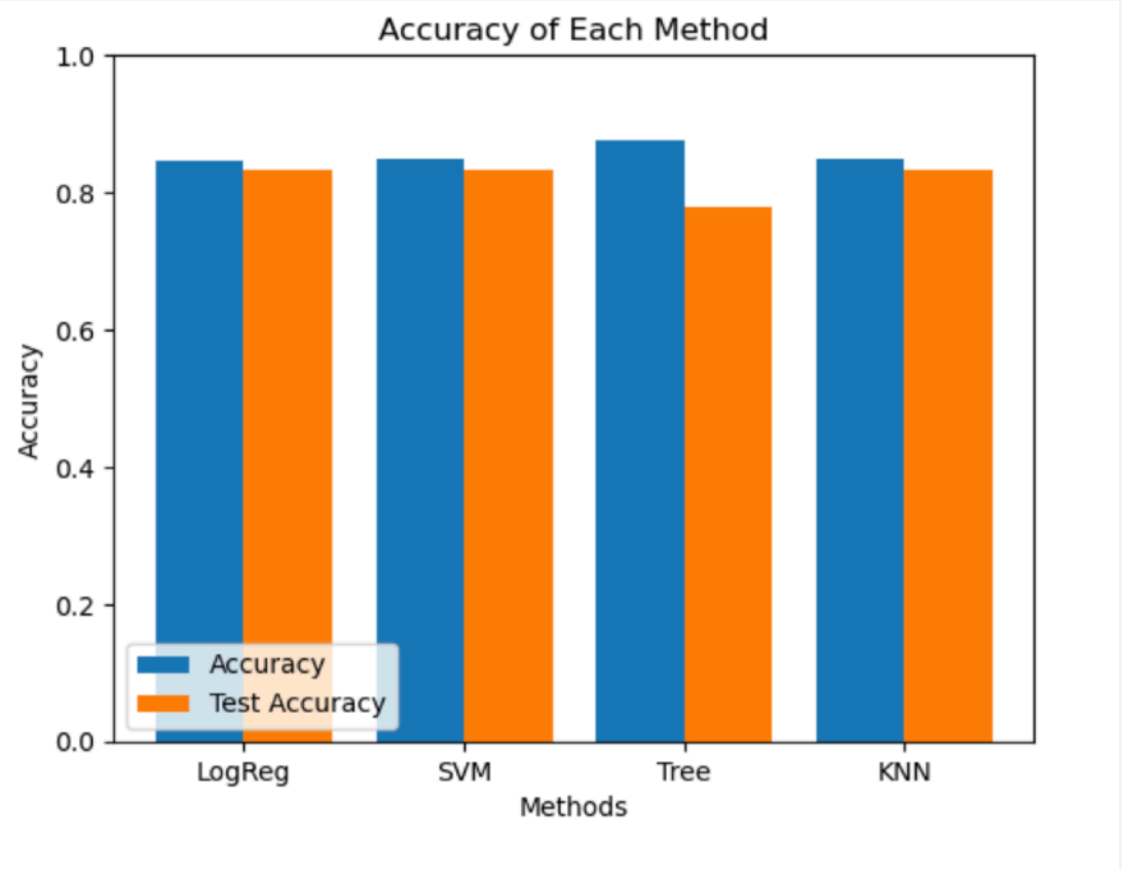
Section 5

Predictive Analysis (Classification)

Classification Accuracy

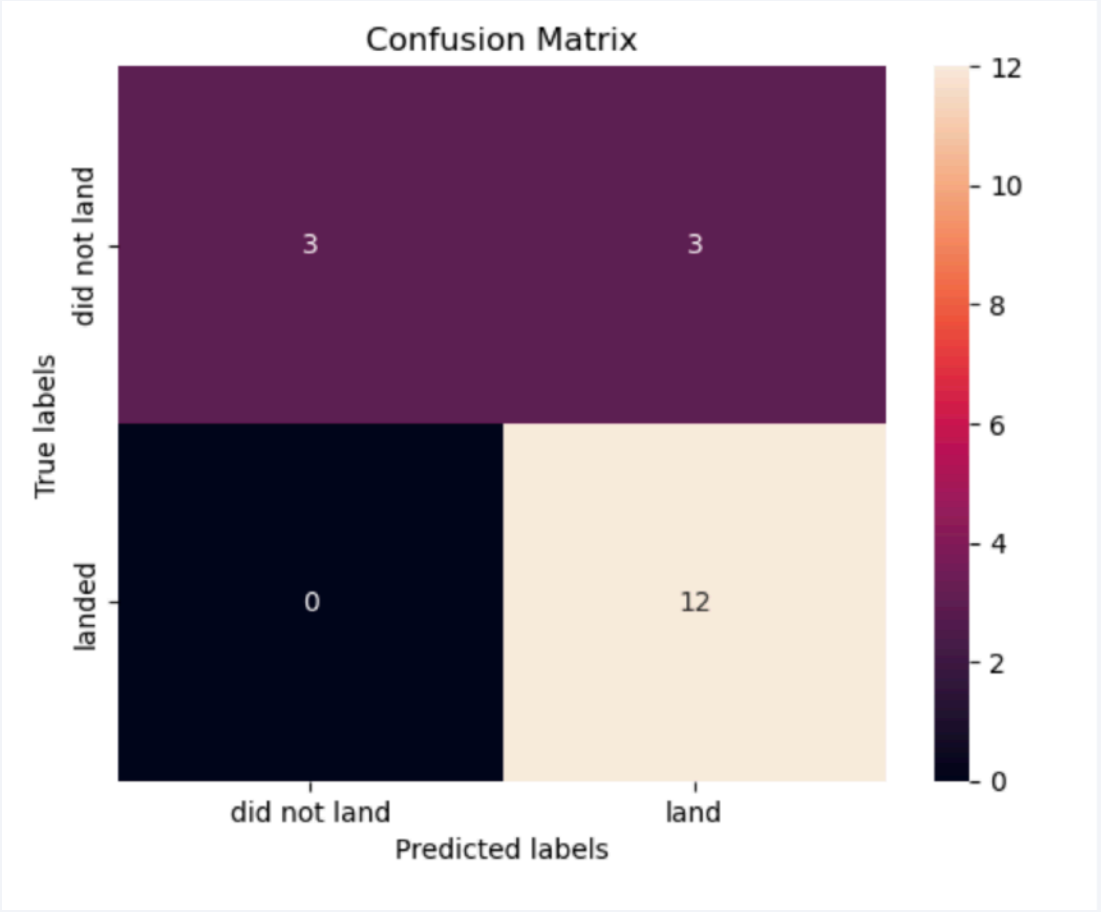
- The accuracy of the four classification model is on the right.

- We can see that the Decision Tree classification has the highest accuracy of 87.5%.



Confusion Matrix

- The confusion matrix of the best performing model Decision tree is on the right.
- We can see that it has a little bit of false positive error.



Conclusions

- We can see that the success rate has been increasing since 2013.
- We can see that the launch site CCAFS SLC 40 is most successful for heavy payloads.
- We can see that the launch site KSC LC-39A is the most successful and has a success rate of 76%.
- We can also see that FT booster are the most successful boosters.
- We can see that the best classification model for this project is Decision tree model as it has the highest accuracy.

Thank you!