Winning Space Race with Data Science

Abir Dey
10/11/2022

Outline

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

Executive Summary

- The following methodologies were used to analyze the data:
  - Data was collected using SpaceX API and web scraping
  - Data Wrangling, Exploratory Data Analysis (EDA) and Data Visualization, Interactive Visual Analytics
  - Machine Learning
- Summary of all results
  - Correct public sources were used to obtain data.
  - Removal of errors in the dataset using data wrangling and identifying the correct features to be used to predict upcoming launches using EDA
  - Found the best model to predict the results using Machine Learning

Introduction

- The objective is to evaluate whether a new company Space Y is viable to compete with Space X in terms of cost.
- Predict the cost of the launch by predicting the success rate of each launch.
- Predict the best launch site for launching the rocket to ensure maximum success rate.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using Space X API (https://api.spacexdata.com/v4)
  - And Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - More insights about the data were found by comparing the landing outcome with different features to find out which features are to be selected for training the model for prediction.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Data collected were normalized, divided into train and test sets and were evaluated by four different algorithms by using different parameters for each to determine the model and the parameter to get the best accuracy.

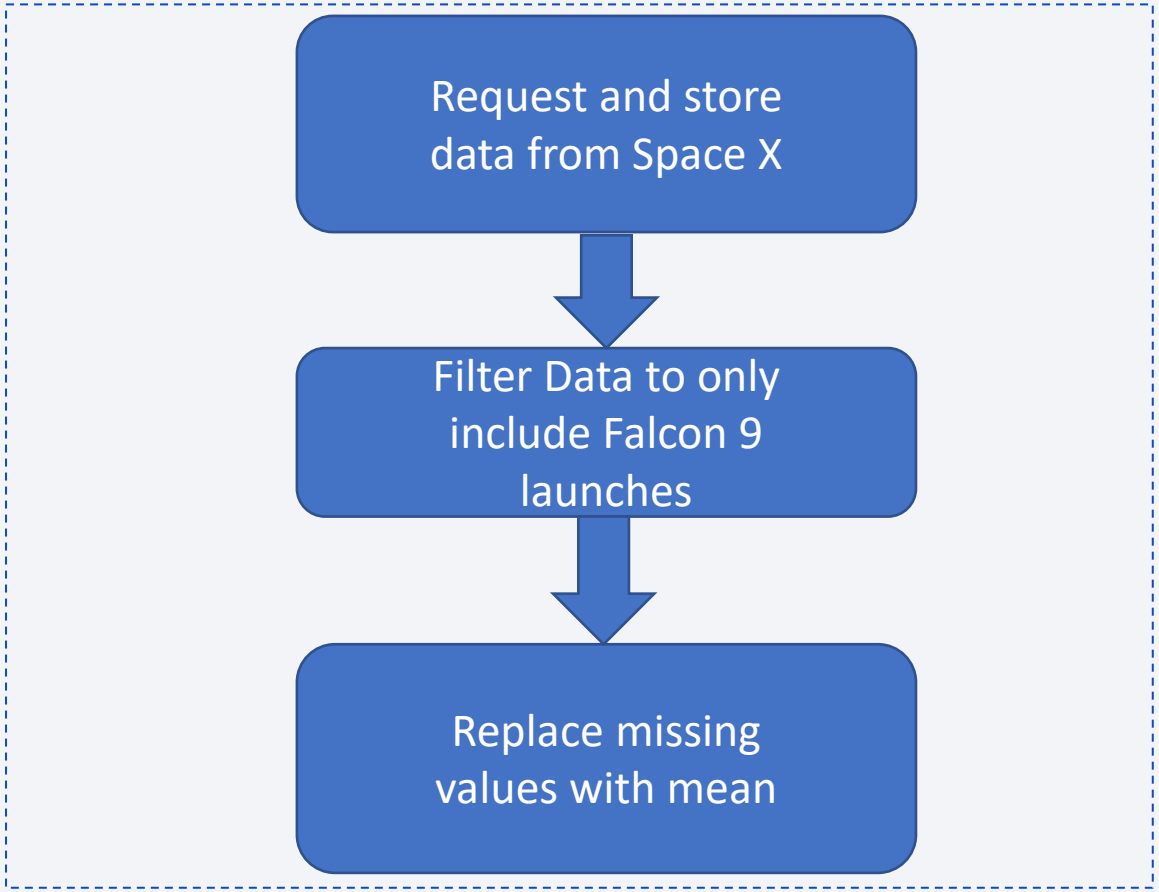
Data Collection

- Data were collected using Space X Public API v4: https://api.spacexdata.com/v4/
- Additional data were also collected by using web scraping to scrape the data publicly available on the web: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches/

Data Collection – SpaceX API

- Space X provides an Open-Source REST API for all its launch, rocket, core, capsule, starlink, launchpad, and landing pad data
- GitHub URL:

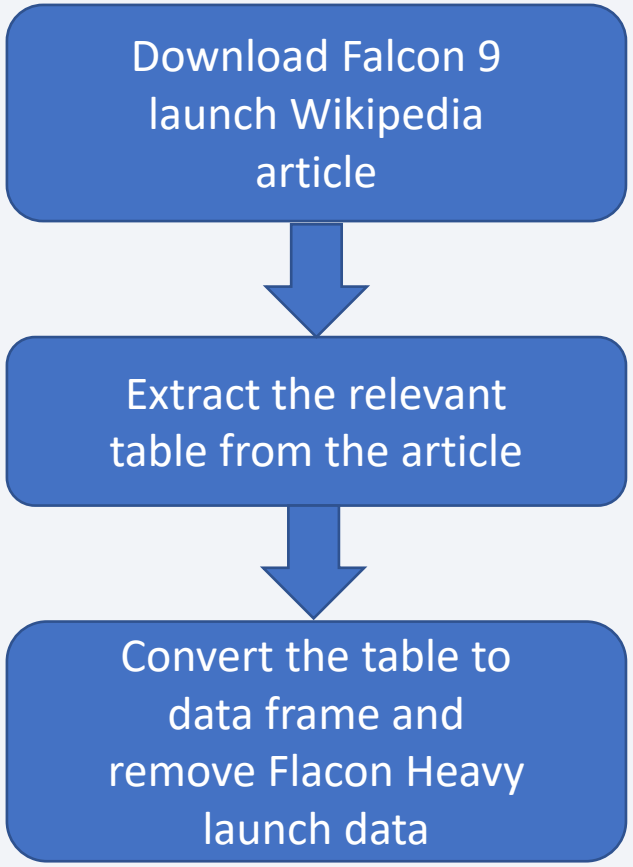
Data Collection — SpaceX API



Data Collection - Scraping

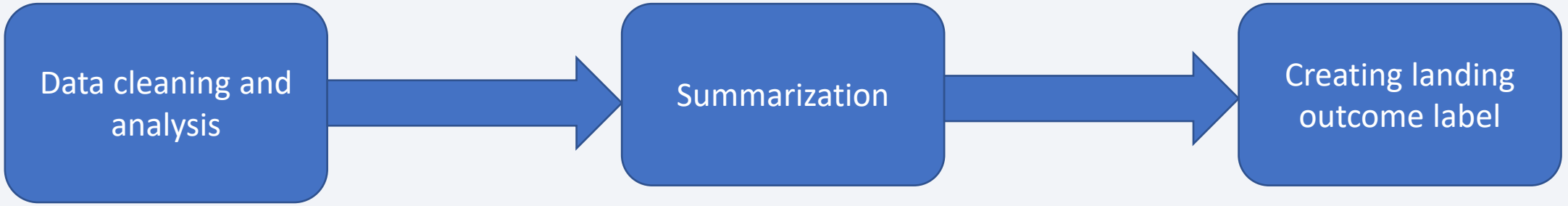
- Additional Falcon 9 launch data was extracted from Wikipedia article: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches/ with the help of web scraping.
- GitHub URL:

Data Collection - Scraping



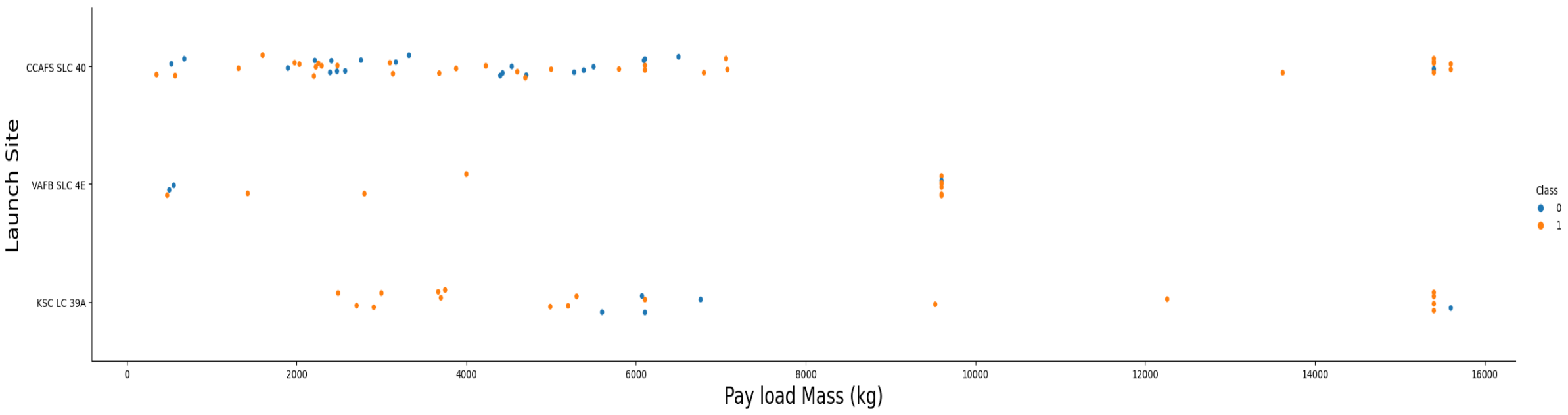
Data Wrangling

- Removed null elements from the dataset and then the data was analyzed
- Summarized the data based on launches per site, occurrences of each orbit and mission outcome based on orbit
- Finally, landing outcome label was created from the Outcome column
- GitHub URL: Data Wrangling



EDA with Data Visualization

- Scatterplots and Barplots were created between:
  - Flight Number and Launch Site, Payload and Launch Site, Payload and Launch Site, Flight Number and Orbit type, Success rate of each orbit to find how each important variable would affect the success rate.



- GitHub URL: [EDA with Data Visualization]

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; and
  - 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.
- GitHub URL: EDA with SQL

Build an Interactive Map with Folium

- Markers, circles, lines, etc. you created and added to the Folium Map
  - Circles are used to indicate the launch sites
  - Markers were used to mark the success/failed launches for each site on the map.
  - Lines are used to measure the distance from each launch site to the nearest coastline, city, railway and highway

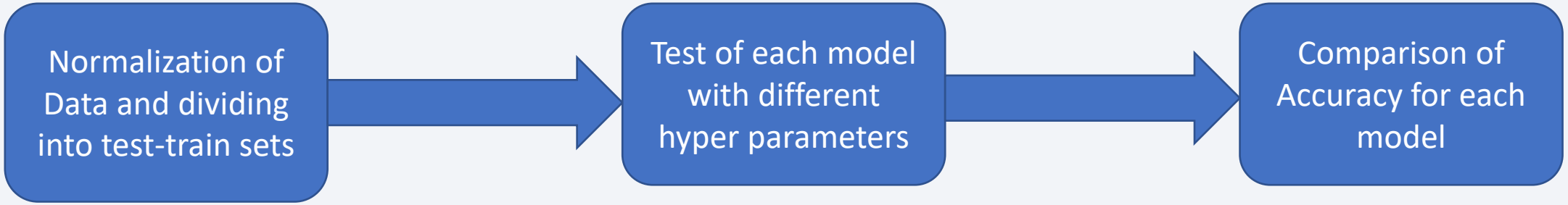
- GitHub: Interactive Map with Folium

Build a Dashboard with Plotly Dash

- The following elements were added to the dashboard:
  - Dropdown menu with the option to select a particular launch site
  - Pie chart showing the percentage of successful and failed launches from the selected site
  - A slider used to select the range of payload carried by the rocket
  - Scatter plot indicating the number of failed and successful rockets launched by the selected launch site carrying payload range indicated by the slider
- The dashboard allowed to quickly find the best place to launch a rocket with a particular amount of payload
- GitHub URL: Dashboard with Plotly Dash

Predictive Analysis (Classification)

- Four different model with different hyperparameters for each were compared to find the most optimal accuracy:
  - logistic regression
  - support vector machine
  - decision tree
  - k nearest neighbours



- GitHub URL: [Prediction Analysis]

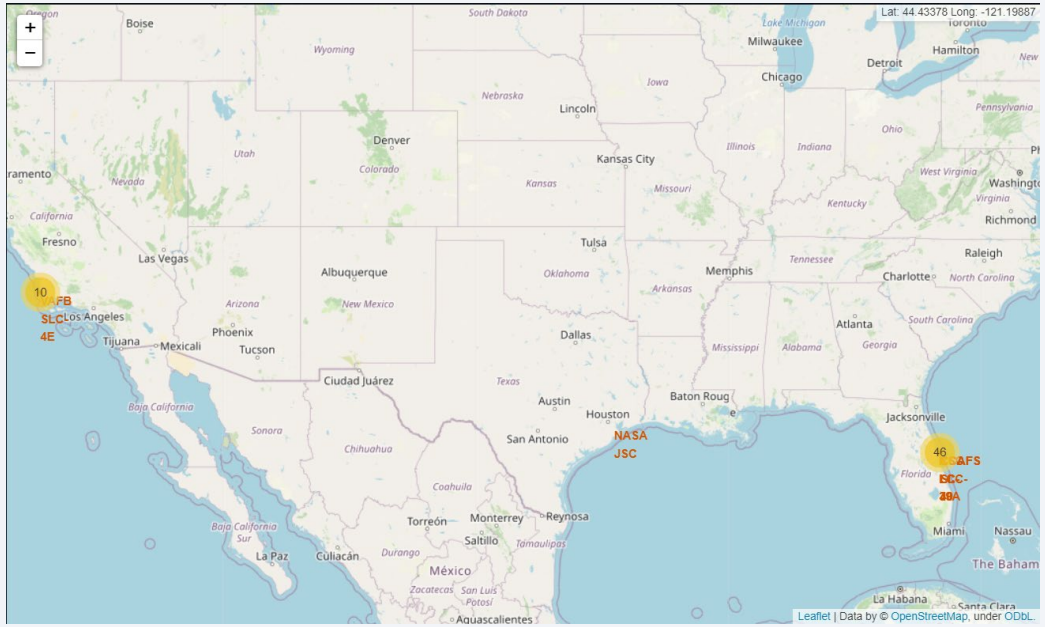
Results

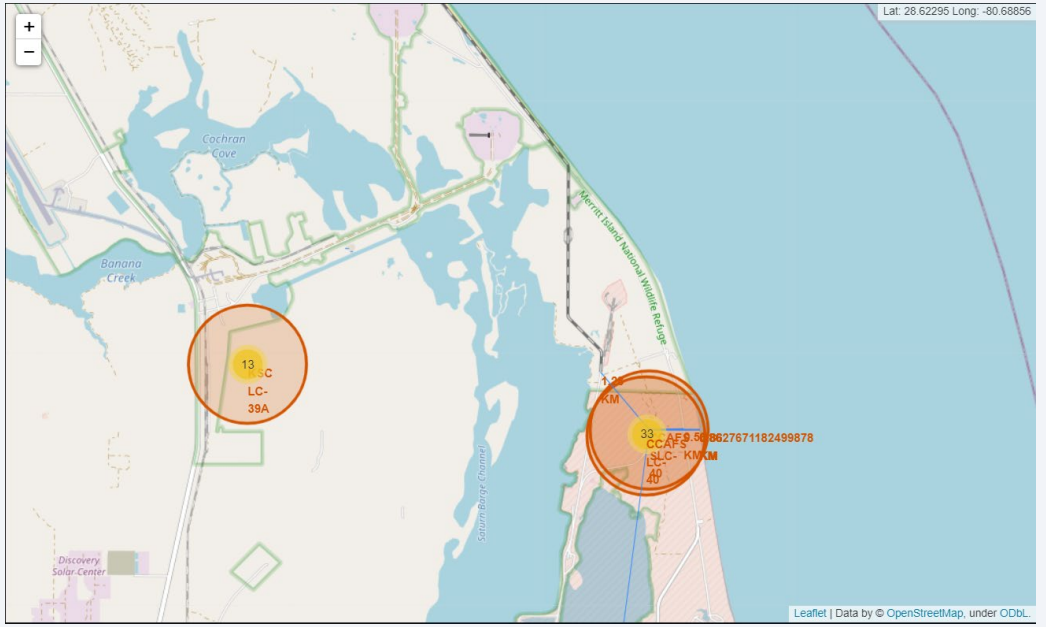
- Exploratory data analysis results:

  - 4 sites are used by Space X to launch its rockets.
  - VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
  - With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS, but the same cannot be said for GTO
  - Almost 100% of mission outcomes were successful
  - The first successful landing was in 2015, which is 5 years since the first launch in 2010
  - The success rate of launches have been increasing since 2013 till 2020

Results

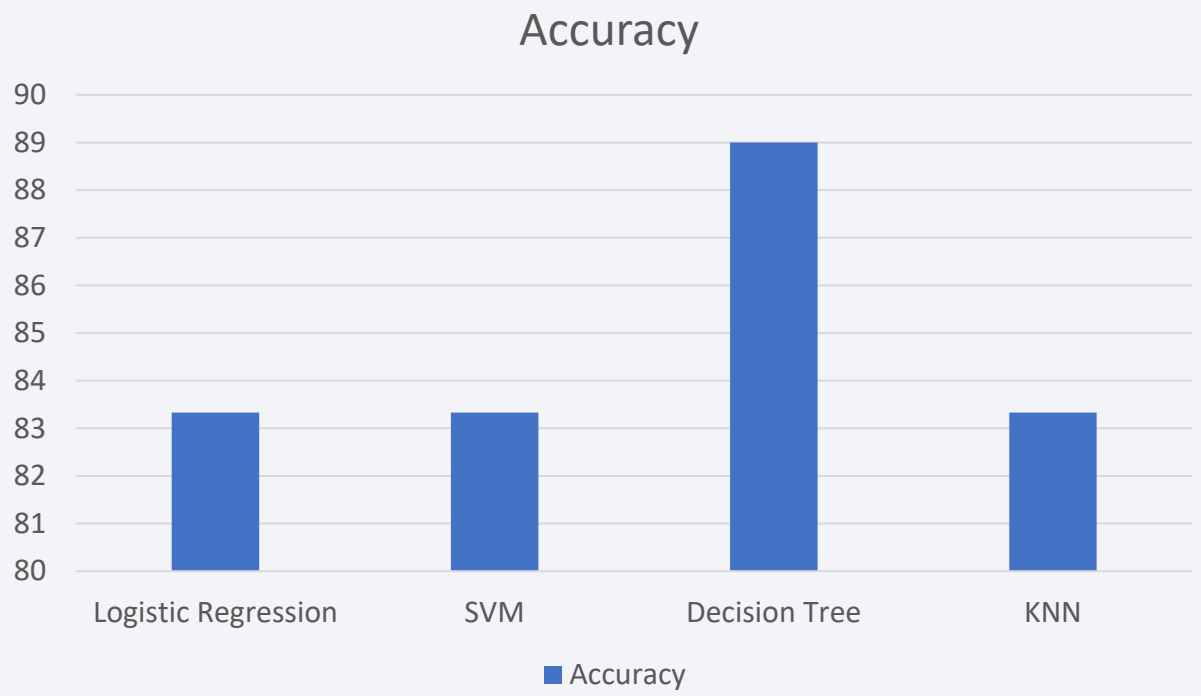
- Using the interactive maps, analysis was possible for the launch sites to identify the nearest coastline, railways and highway so as to have a good logistic infrastructure around.
- Analysis also revealed that more launches happened in the east coast of USA than the west coast.





Results

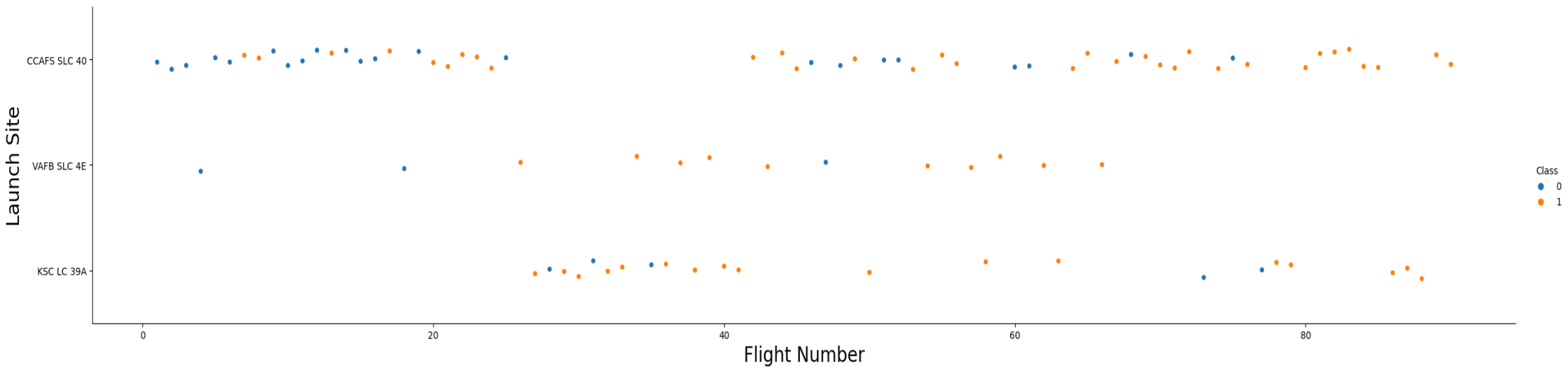
- Predictive analysis results:
  - Predictive analysis revealed that Decision Tree Classifier is the best model to predict the landings.
  - The Decision Tree Classifier had an accuracy of 88% while the other models had an accuracy around 83%.



Section 2

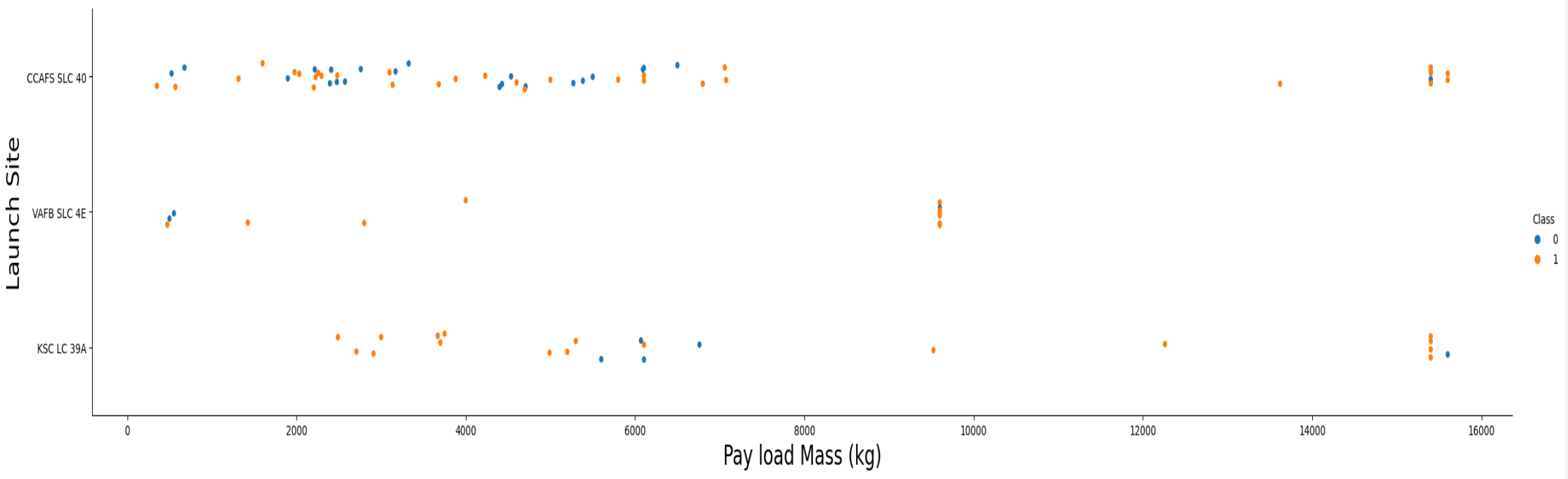
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot, most of the rockets were launched at site CCAFS SLC 40
- It is also found that the success rate also increased with the increase in the number of flights.

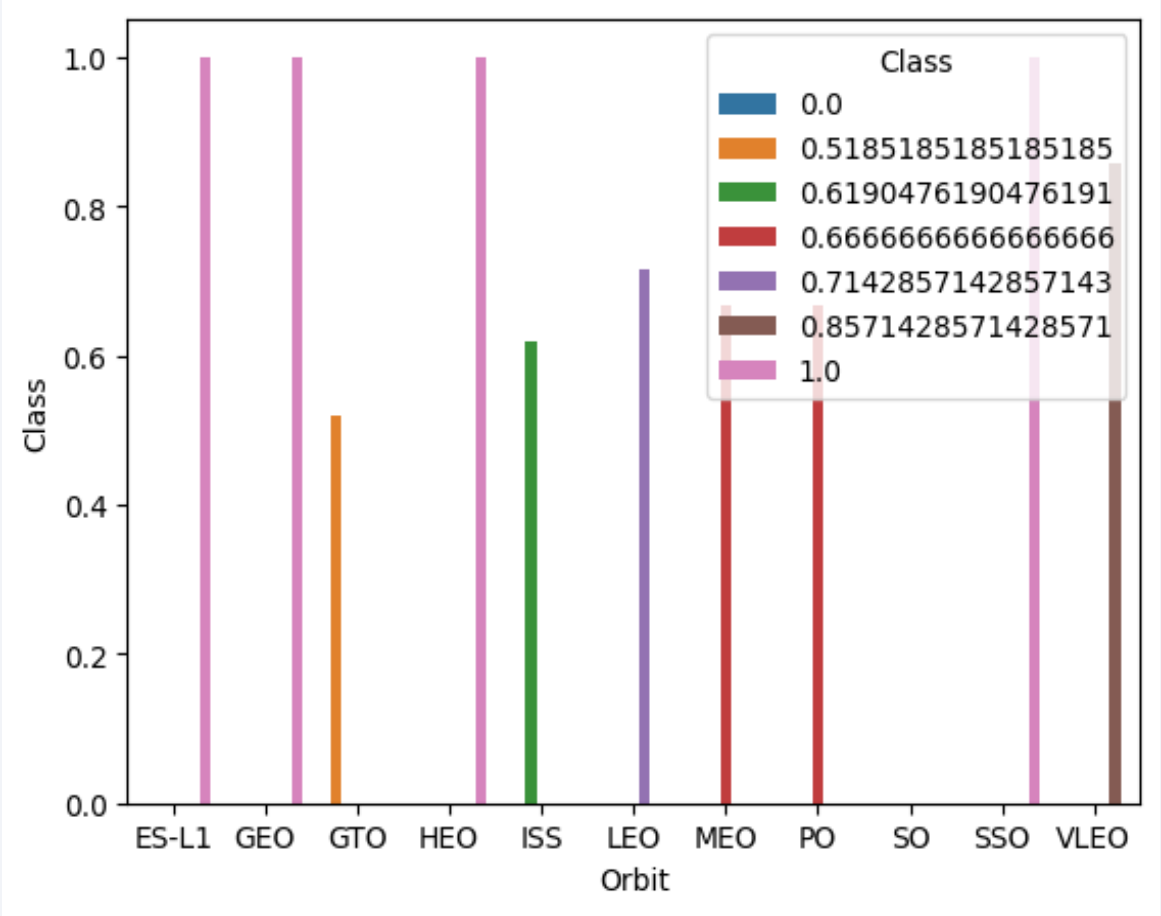
Payload vs. Launch Site



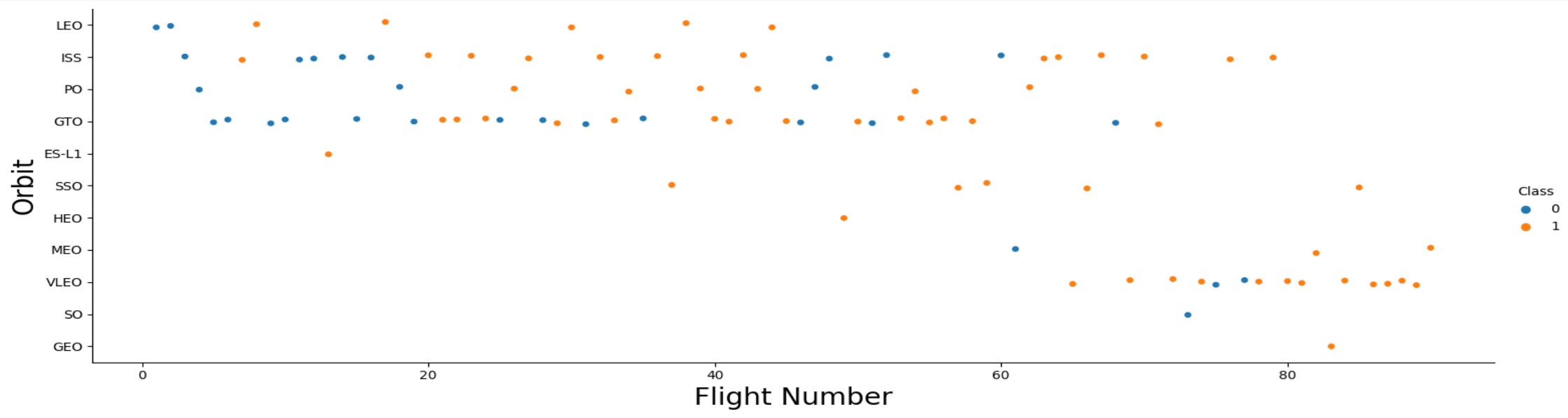
- Rockets carrying payloads over 8000kg are having an excellent success rate
- Payloads greater than 1000 kg are not launched from site VAFB SLC 4E and are only possible from sites CCAFS SLC 40 and KSC LC 39A

Success Rate vs. Orbit Type

- The launches to orbits having 100% success rates :
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)

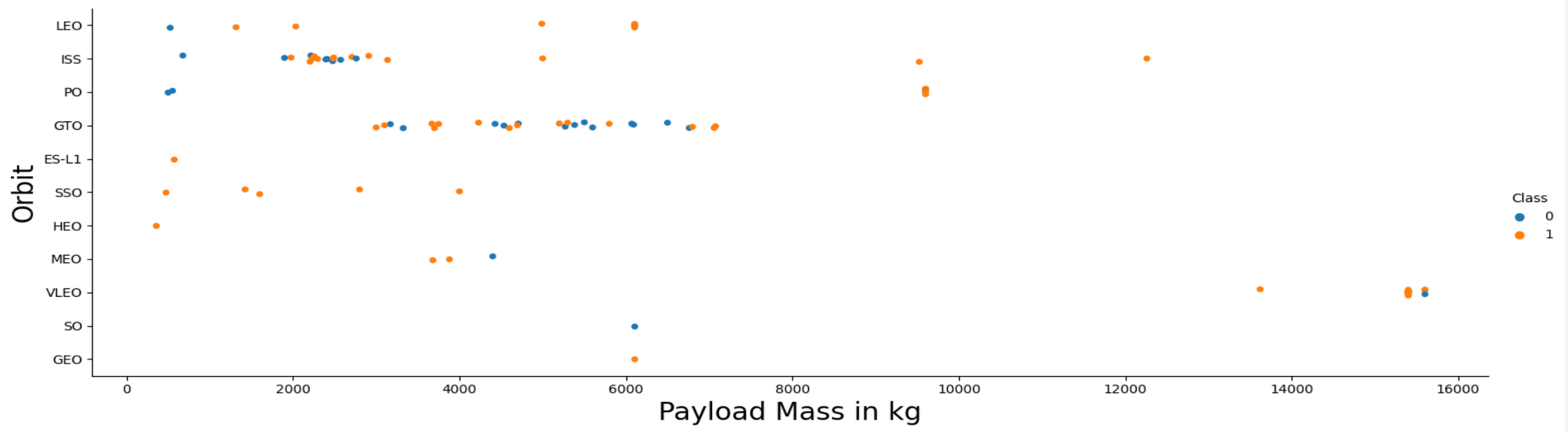


Flight Number vs. Orbit Type



- The success rate has improved in the recent years around all orbits
- Launches to new orbits such as VLEO, GEO, MEO have been possible with VLEO having more frequent launches

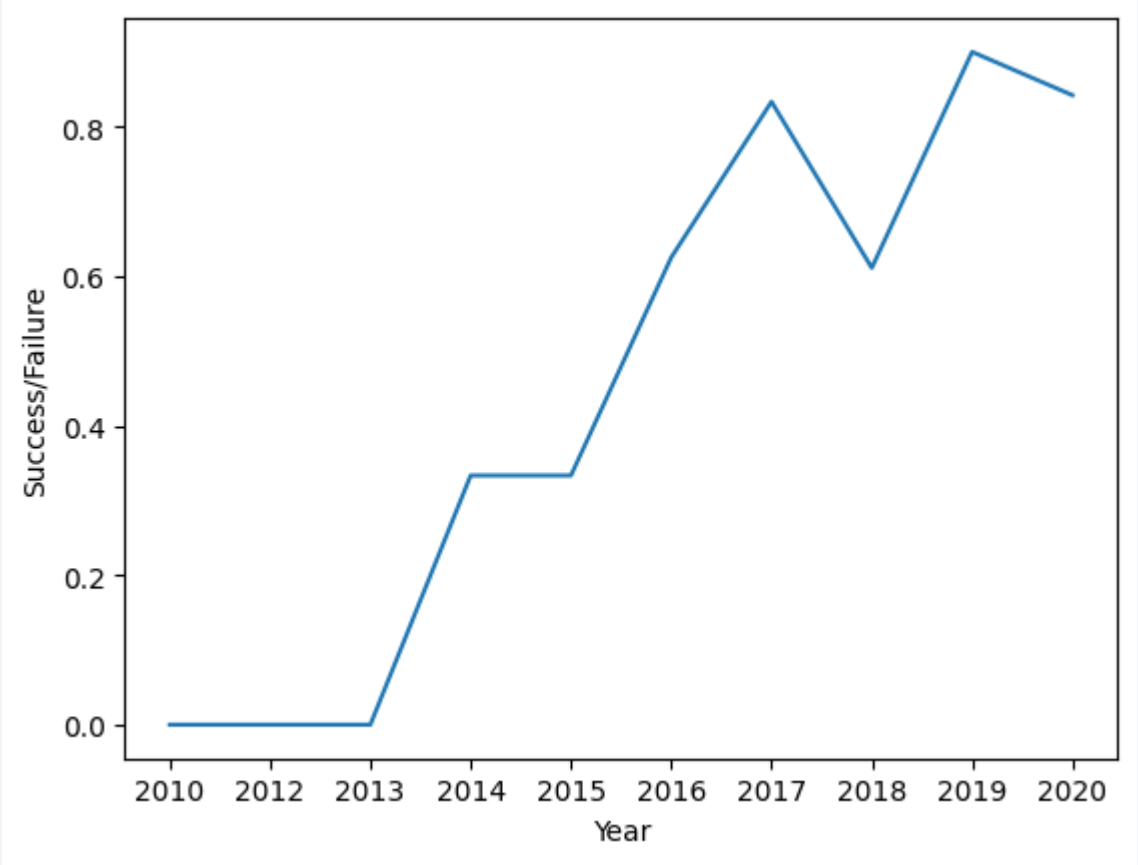
Payload vs. Orbit Type



- Launches to orbit GTO have the widest payload range
- Very heavy payloads have been delivered to the orbit VLEO

Launch Success Yearly Trend

- The success rates of launches have been increasing since 2013
- There is a decrease in success rate during the year 2018
- The years 2010 – 2013 were a period of improvement and adjustments



All Launch Site Names

- According to the dataset, 4 launch sites were used by Space X to launch their rockets:

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745

- Three launch sites CCAFS LC-40, CAFS SlC-40, KSC LC-39A are located in the east coast of United States
- Only one launch site VAFB SLC-4E is located in the west coast.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA` :

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- All the launches happened in Cape Carnival.

Total Payload Mass

- The total payload carried by boosters from NASA was:
- 45596

- Total payload calculated above, by summing all payloads whose codes contains ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1:
- 2928

- Filtering data by booster version and calculating the average of all payload mass was found to be 2,928 kg

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

  2010-06-04

- By filtering the data to select only successful mission outcomes, the first date of successful landing was on 2010-06-04

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:
  - F9 FT 81022
  - F9 FT 81026
  - F9 FT 81021.2
  - F9 FT 81031.2
- Selecting the boosters which had successful drone ship landing and had a payload between 4000 and 6000, we get the above four results.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes:

Mission Outcome	Occurances
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Almost all the mission outcomes were successful

Boosters Carried Maximum Payload

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

- These are the results found after searching for boosters carrying the maximum payload mass in the databas

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

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- There were only two failed landings in drone ship in the year 2015

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

- Ranking of landing outcomes between the date 2010-06-04 and 2017-03-20

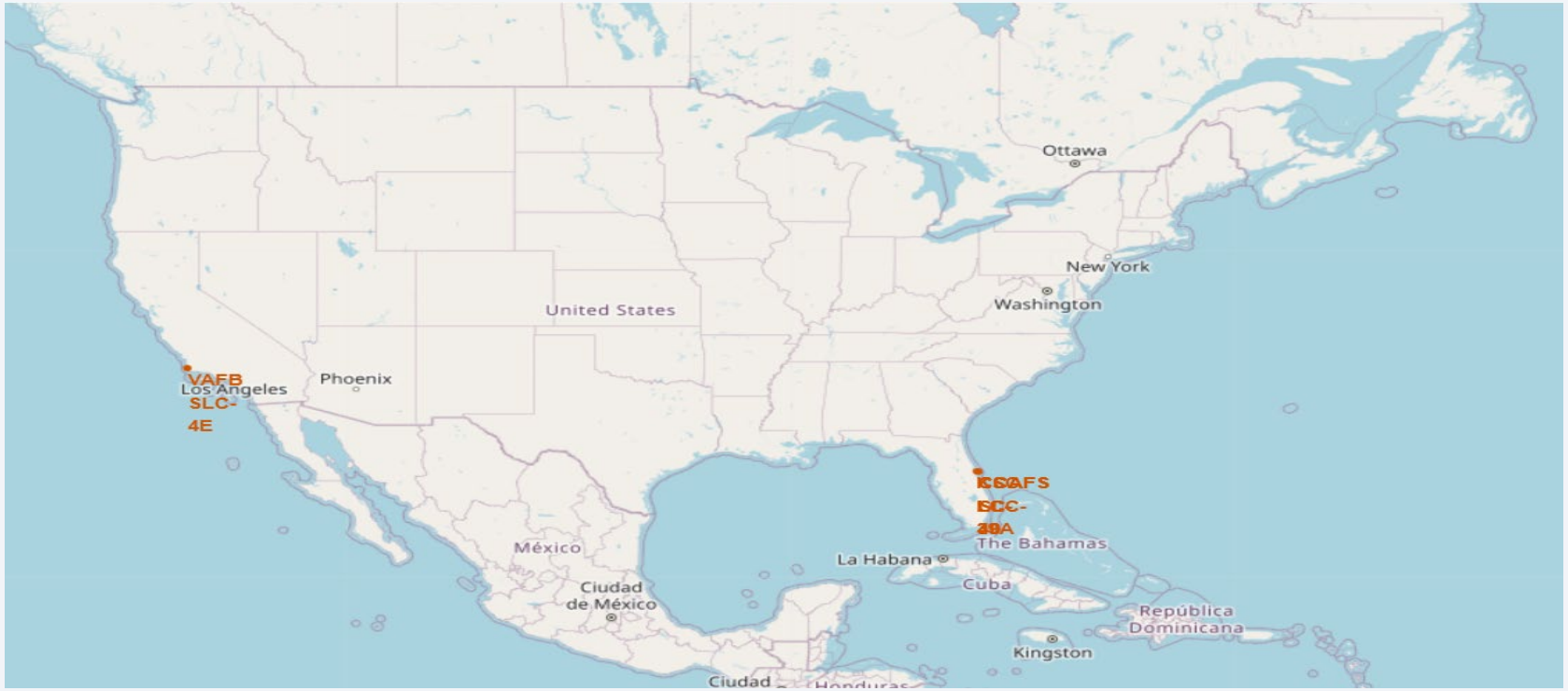
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This data indicates that most of the landings were not attempted during that period.

Section 3

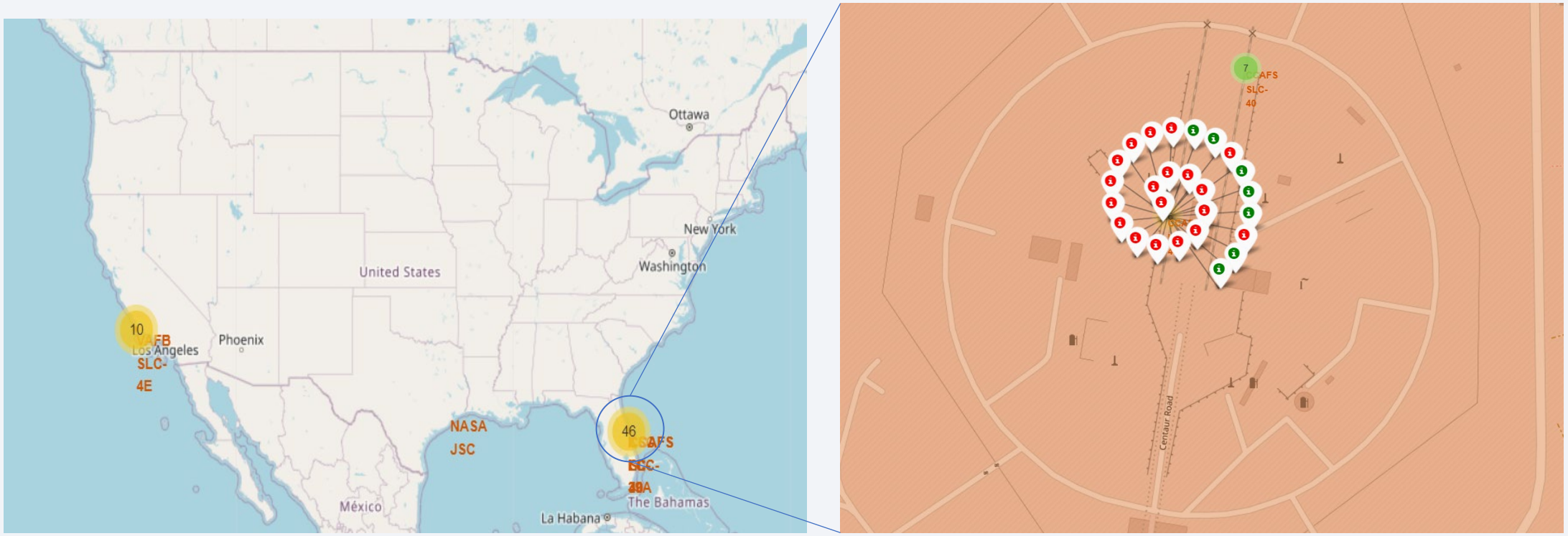
Launch Sites Proximities Analysis

Launch Sites



- All the launch sites are located near coasts for safety reasons.
- The launch sites are also located near to highways and train tracks so that transportation is easy.

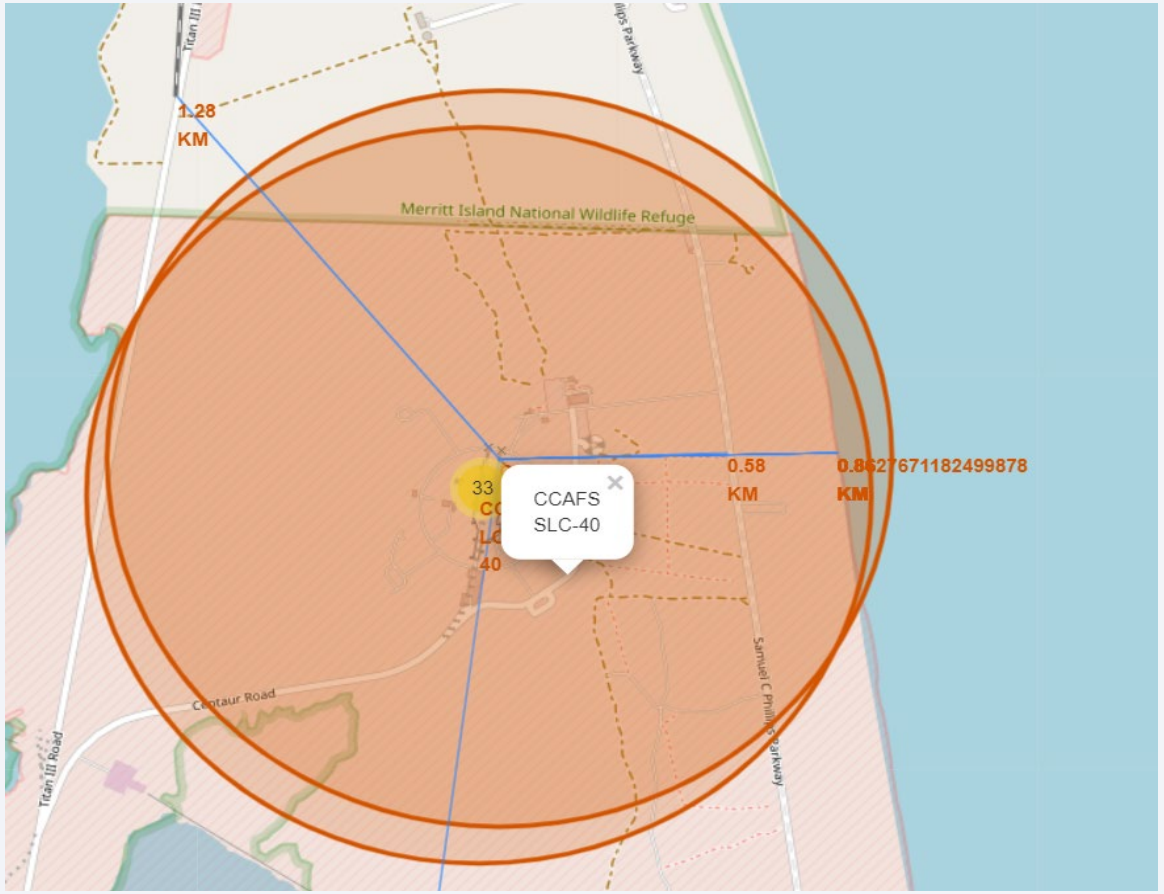
Launch Outcomes by Site

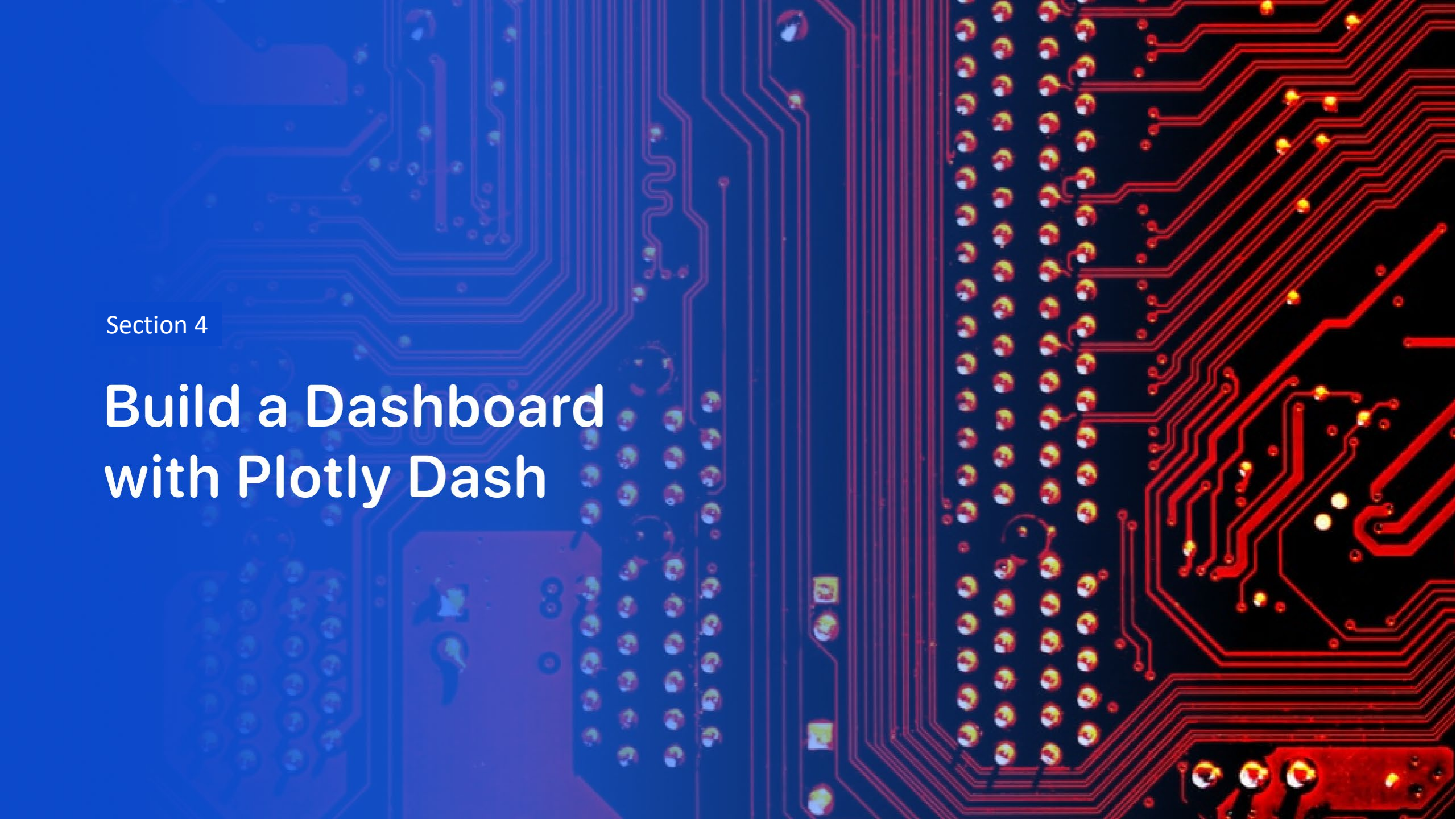


- The green markers indicates success while the red markers indicate failed launch attempts

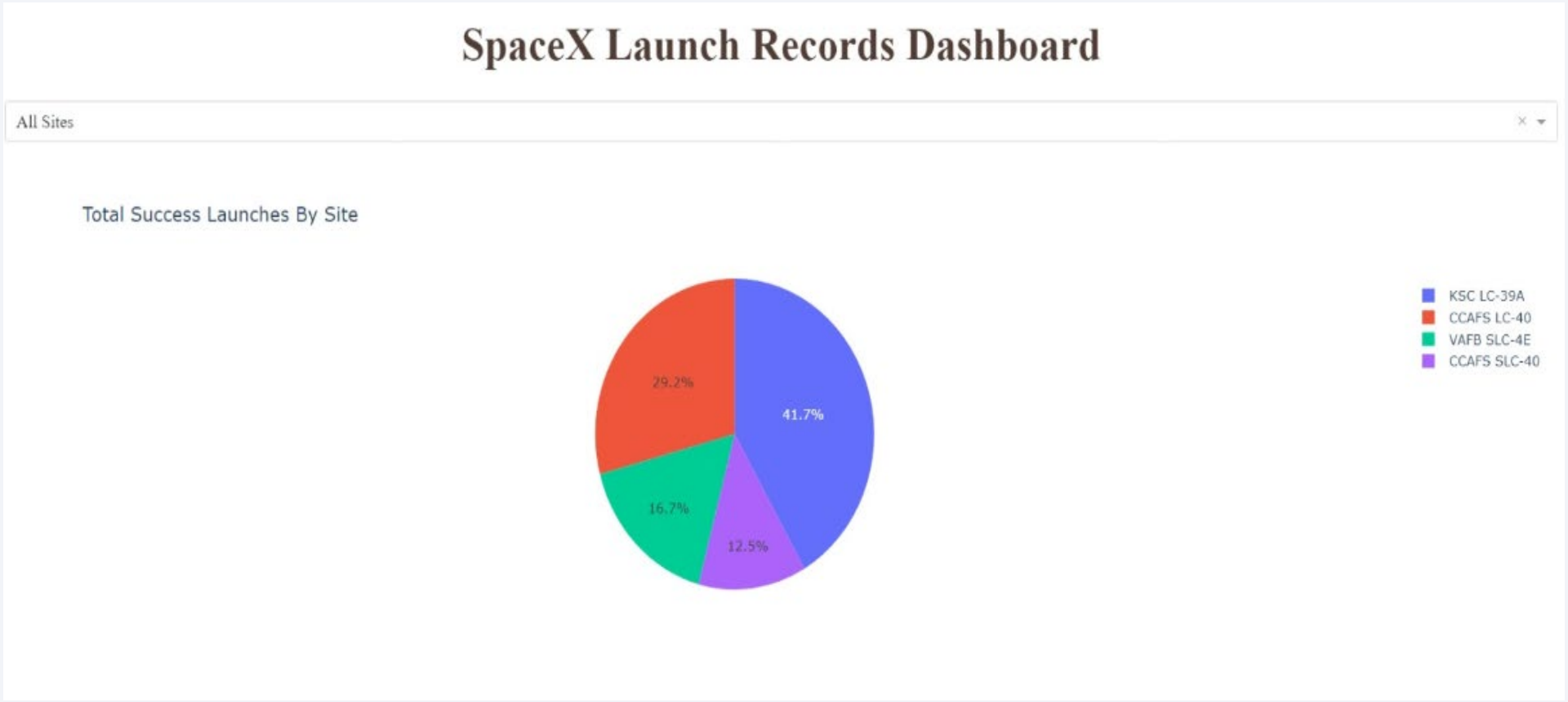
Logistics

- The launch site CCAFS SLC-40 is 0.58 km from highway, 1.28 km away from railway, and just 0.86 km away from coast, making the place away from population areas and highly accessible from all modes of transportation.



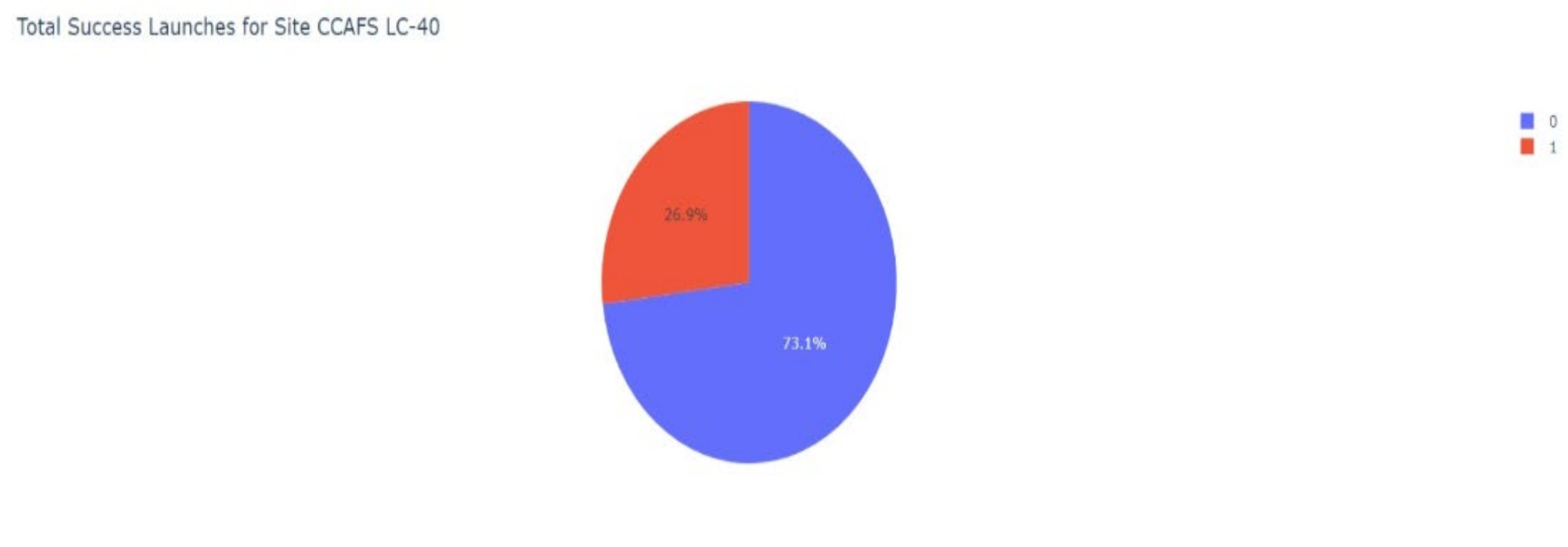

Section 4
Build a Dashboard with Plotly Dash

Successful Launces by Site



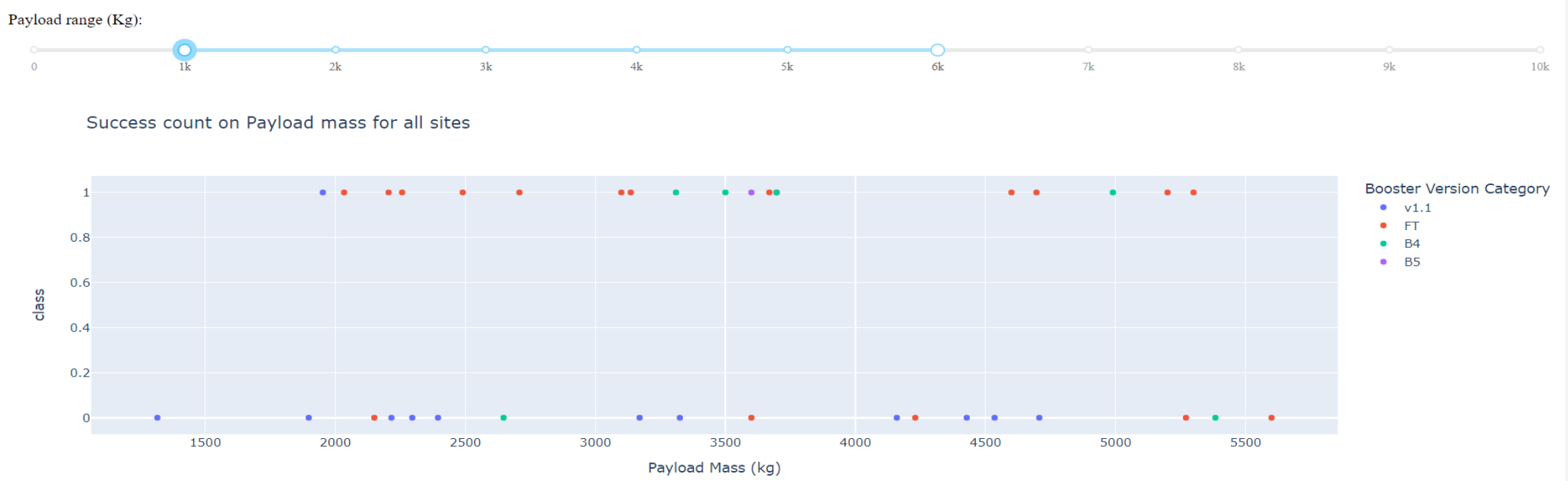
Most of the rockets were launched from KSC LC-39A and CCAFS LC-40 combined.

Success Ratio for launch site CCAFS LC-40



- Around one fourth of the rockets launched from this site were successful.

Payload vs Launch Outcome

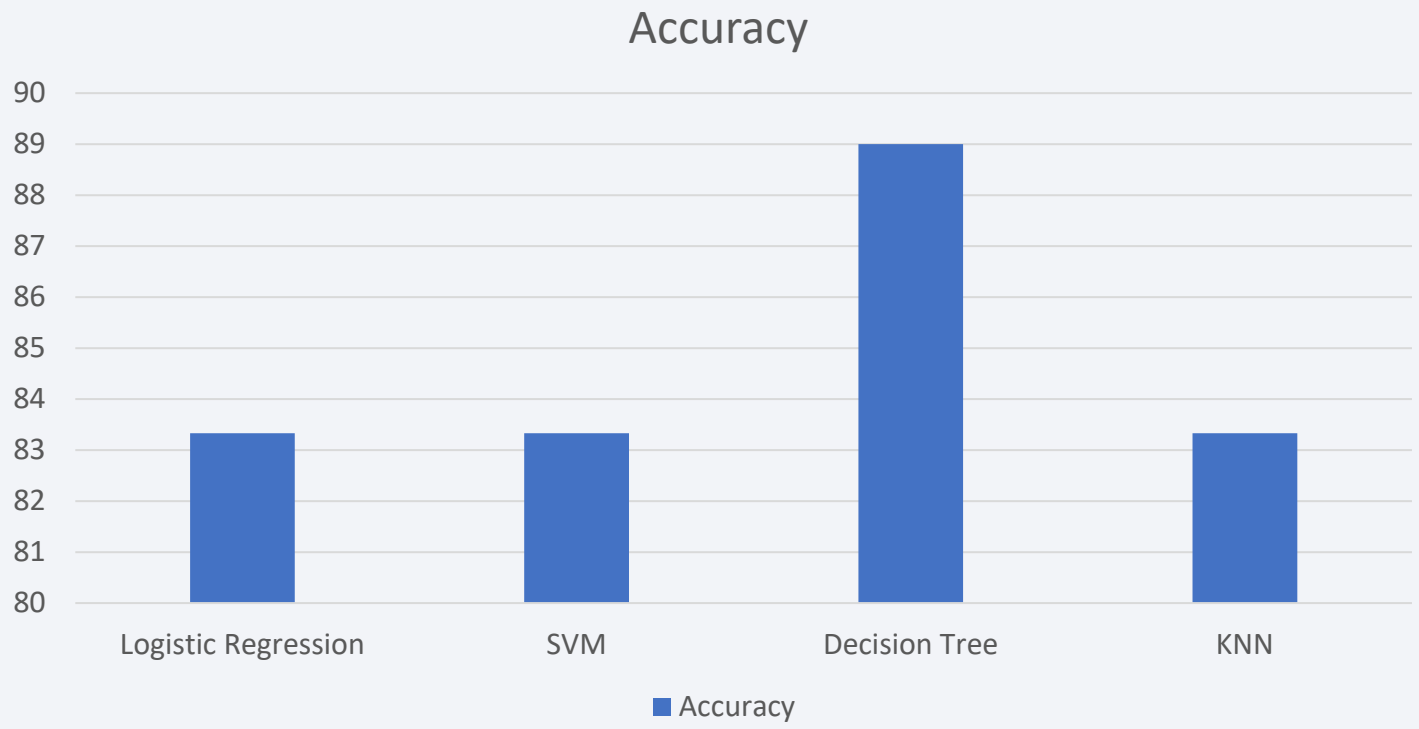


FT and B4 boosters are mostly successful when carrying payloads between 1000 and 6000kg

Section 5

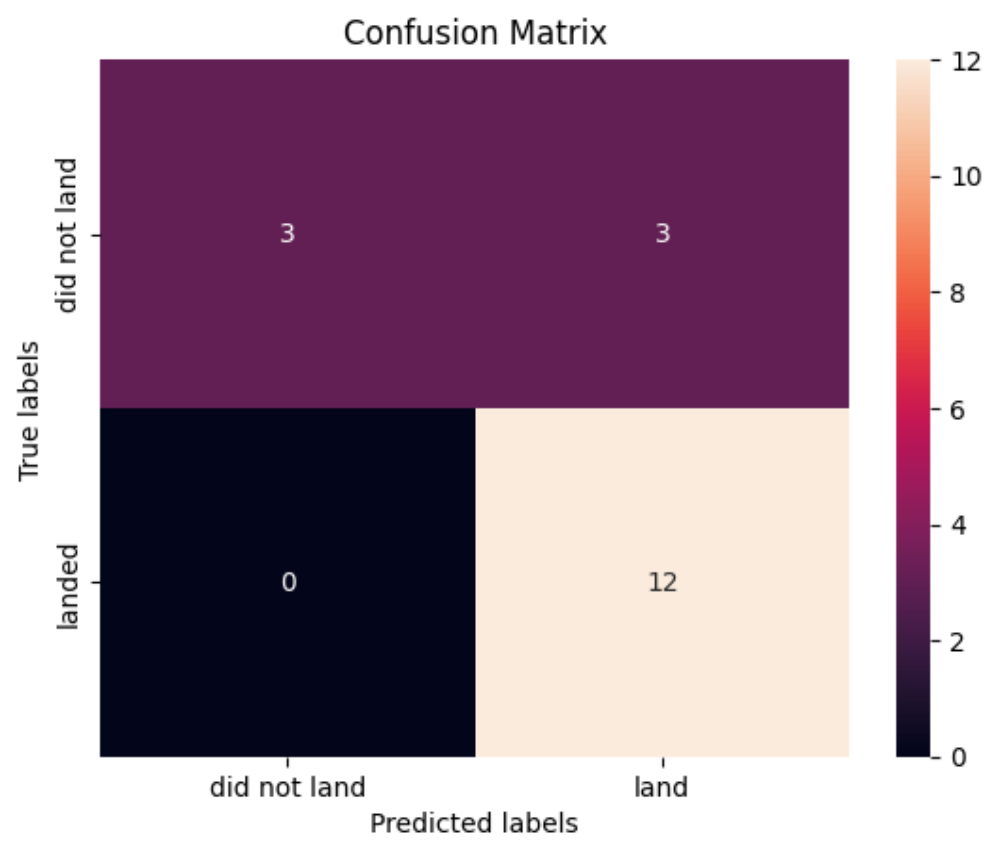
Predictive Analysis (Classification)

Classification Accuracy



- Among all the models used, Decision Tree provided the highest accuracy followed by the other methods giving the same accuracy

Confusion Matrix



- Confusion Matrix indicates the absence of False Negatives and less False Positives, indicating good training of model.

Conclusions

- Different data sources were used to analyze the problem, to increase its reliability of the sources
- The launch site with highest success rate was found to be KSC LC-39A making it the best site for launching rockets
- There are very few failures when launching very heavy payloads (>8000 kg)
- The success rate of the launches has been increasing year over year
- Decision Tree Classifier is found to be the best model to predict mission outcome

Appendix

- All the files used for the project are stored in the following GitHub page:
- https://github.com/AD2011/applied-data-science-capstone-coursera/tree/main/Course%2010%20-%20Applied%20Data%20Science%20Capstone

- Course Link : Applied Data Science Capstone - Coursera

Thank you!