IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Saharat Pholcharoen
2022-02-02

Outline

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

Executive Summary

Methodologies:

- Collect data on Falcon 9 first-stage landing using a RESTful API and web scraping
- Perform exploratory data analysis (EDA) using visualization and SQL
- Build an interactive dashboard with Plotly Dash and interactive map with Folium
- Use machine learning to determine if the first stage of Falcon 9 will land successfully

Results:

- KSC LC-39A is a launch site which has the largest successful launches
- Falcon 9 with Booster Version FT has the highest launch success rate
- All machine learning models perform best with 83.33% accuracy using test data

Introduction

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of savings is because SpaceX can reuse the first stage.

Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

In this project, we will predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

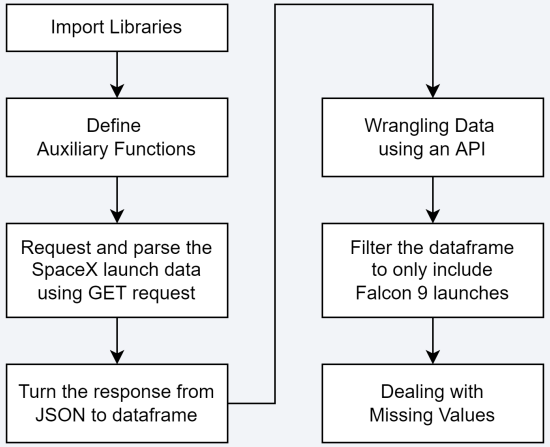
- Data collection methodology:
  - Collect data with SpaceX REST API and web scraping
- Perform data wrangling
  - Convert landing outcomes to labels (either 0 or 1)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build models the sklearn library, perform Grid Search to tune hyperparameters, and evaluate the classification models with confusion matrix

Data Collection

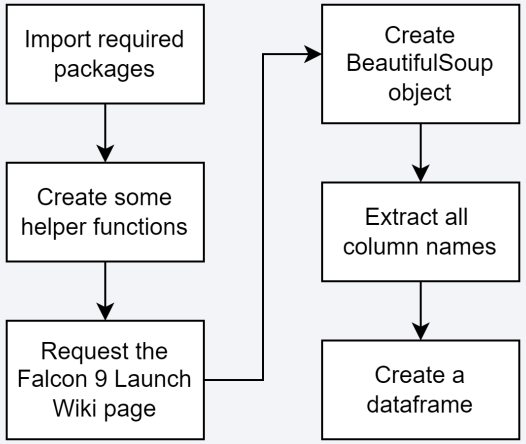
SpaceX Falcon 9 launch data was collected by 2 methods:

- With SpaceX REST API
- Web Scraping from a Wikipedia page

SpaceX API



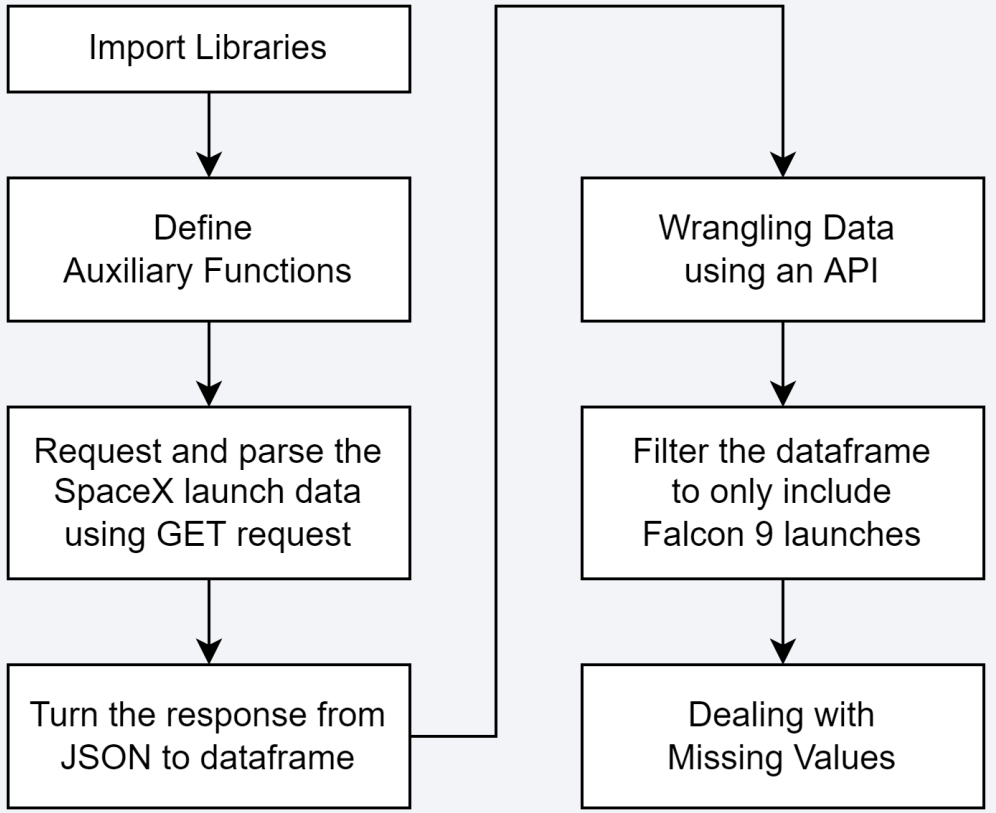
Scraping



Data Collection – SpaceX API

- The process of the data collection with SpaceX API is illustrated by the following flowchart

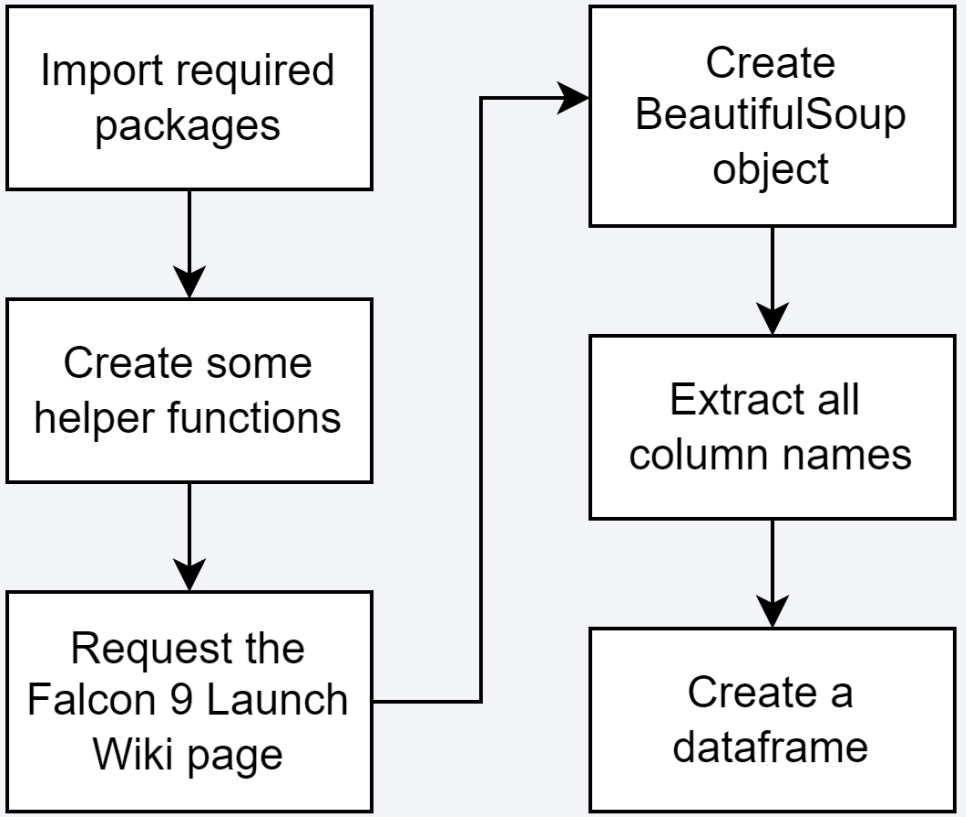
- Reference: [spacex-falcon9-ml-prediction/SpaceX - Data Collection with SpaceX API.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)]



Data Collection - Scraping

- The process of data collection using web scraping is illustrated by the following flowchart

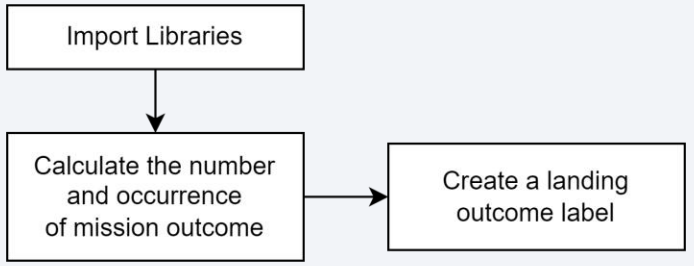
- Reference: [spacex-falcon9-ml-prediction/SpaceX - Data Collection with Web Scraping.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)]



Data Wrangling

Convert the column "Outcome" which indicates if the first stage successfully landed to labels:

- 1 means the booster successfully landed
- 0 means it was unsuccessful



Reference: [spacex-falcon9-ml-prediction/SpaceX - Data Wrangling.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)]

EDA with Data Visualization

To obtain some preliminary insights about how each important attribute would affect the success rate then these features can be used in success prediction with machine learning

Perform EDA to visualize the relationship between each variable as follows:

- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Success Rate vs. Orbit Type
- Flight Number vs. Orbit Type
- Payload vs. Orbit Type
- Launch Success Yearly Trend

Reference: spacex-falcon9-ml-prediction/SpaceX - EDA with Data Visualization.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)

EDA with SQL

Understand the SpaceX dataset better as follows:

- All launch site names are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A and VAFB SLC-4E
- Total pay load mass carried by boosters launched by NASA (CRS) is 45,596 kg
- Average payload mass carried by booster version F9 v1.1 is 2,928 kg
- First successful landing in ground pad was achieved is 22 December 2015
- Total number of successful mission outcome is 100 for failure is 1

Reference: spacex-falcon9-ml-prediction/SpaceX - EDA with SQL.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)

Build an Interactive Map with Folium

Created map objects such as markers, circles, lines and added them to a folium map to analyze the launch site geo data map as follows:

- Mark the locations and proximities of launch sites
- Discover patterns via exploring the map
- Explain how to choose an optimal launch site location

Reference: spacex-falcon9-ml-prediction/SpaceX - Folium Interactive Map.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)

Build a Dashboard with Plotly Dash

- Added an interactive pie chart to see the proportion of successful launches for each launch site to find the launch site that has the highest success rate
- Built a scatter plot relationship between Payload Mass and Launch Outcome to see which range of payload mass and booster version perform the best success rate

Reference: spacex-falcon9-ml-prediction/SpaceX - Plotly Dash Dashboard.py at main · rockslpy/spacex-falcon9-ml-prediction (github.com)

Predictive Analysis (Classification)

- Created a machine learning pipeline including:

- Standardize the data (Preprocessing)

- Split data into training and testing data

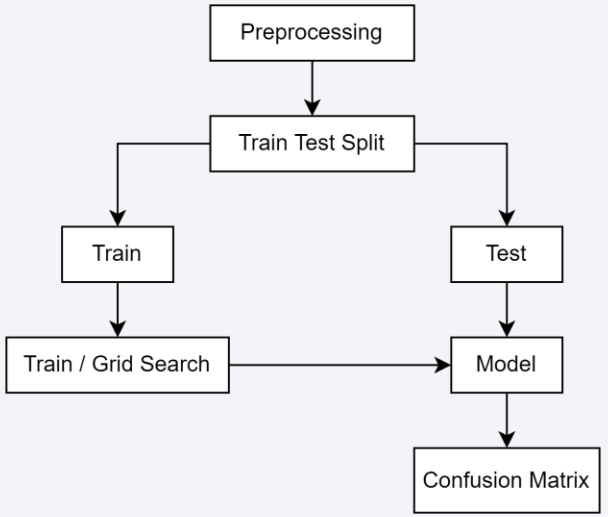
- Train models and perform Grid Search to find the best hyperparameters

- Evaluate models with confusion matrix using test data

- The best performing classification model is the Decision Tree Classifier on the validation data and are all models calculated on the test data

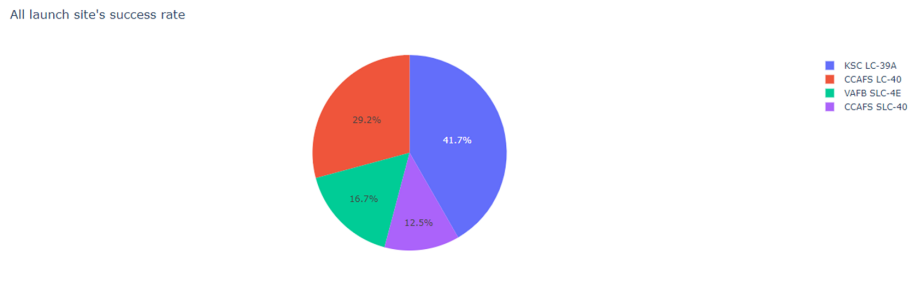
- Model development process is illustrated as the flowchart

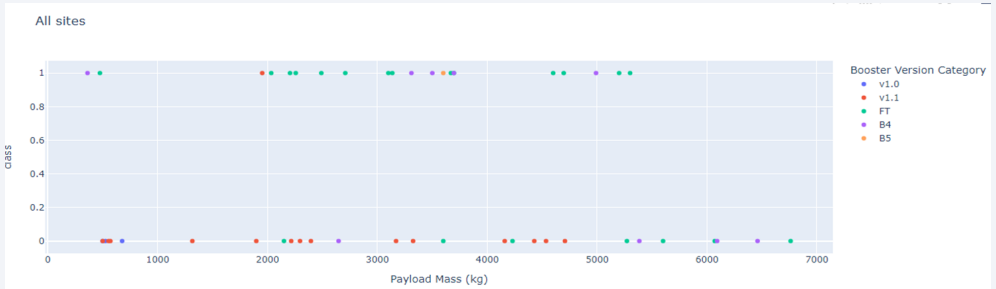
- Reference: [spacex-falcon9-ml-prediction/SpaceX - Machine Learning Prediction.ipynb at main · rockslpy/spacex-falcon9-ml-prediction (github.com)]



Results

- As the flight number increases, the first stage is more likely to land successfully
- Different launch sites have different success rates
- F9 FT boosters with payload 4,000-6,000 kg have successfully landed on drone ship
- KSC LC-39A is a launch site which has the largest successful launches
- Falcon 9 with Booster Version FT has the highest launch success rate

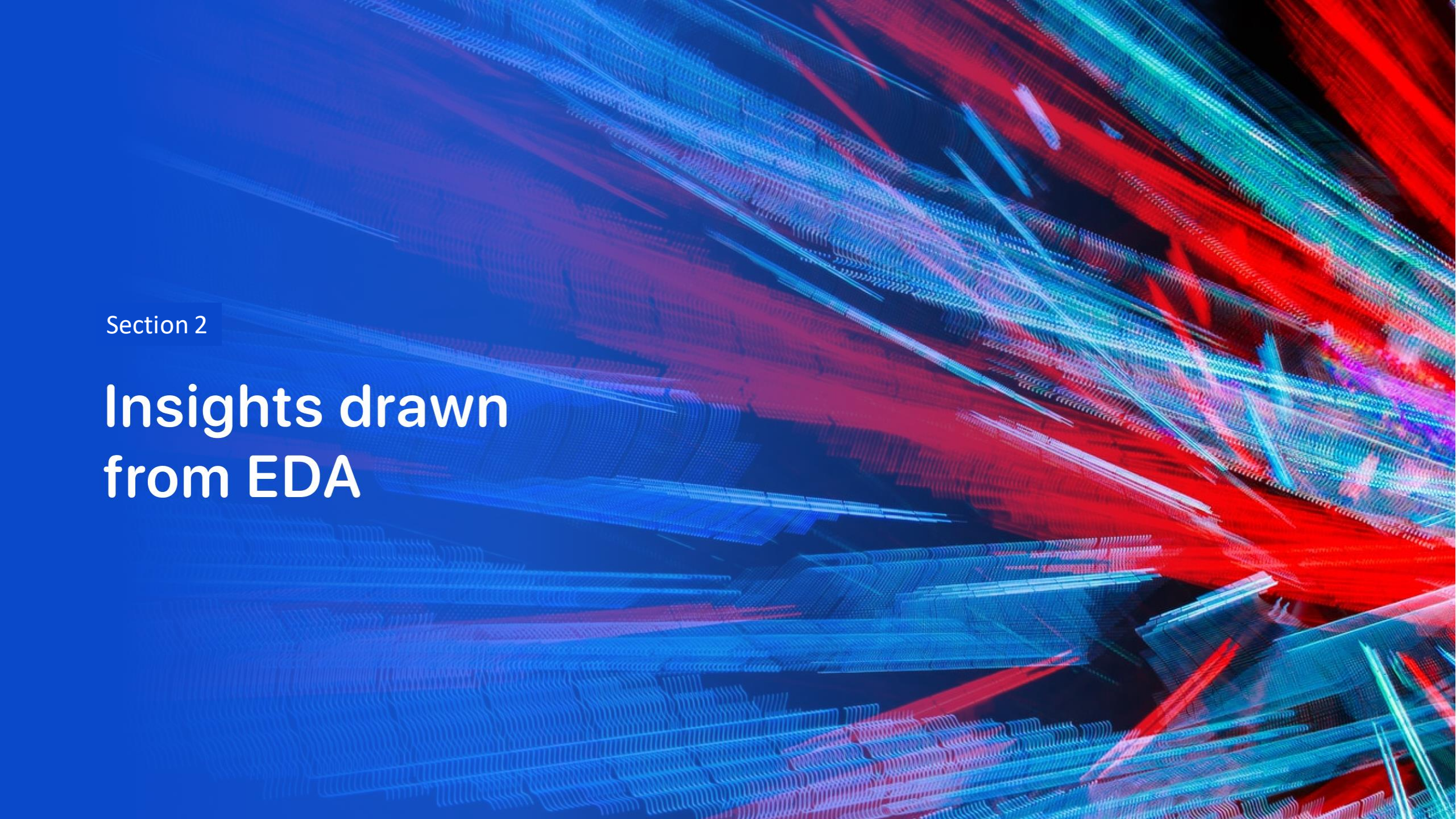




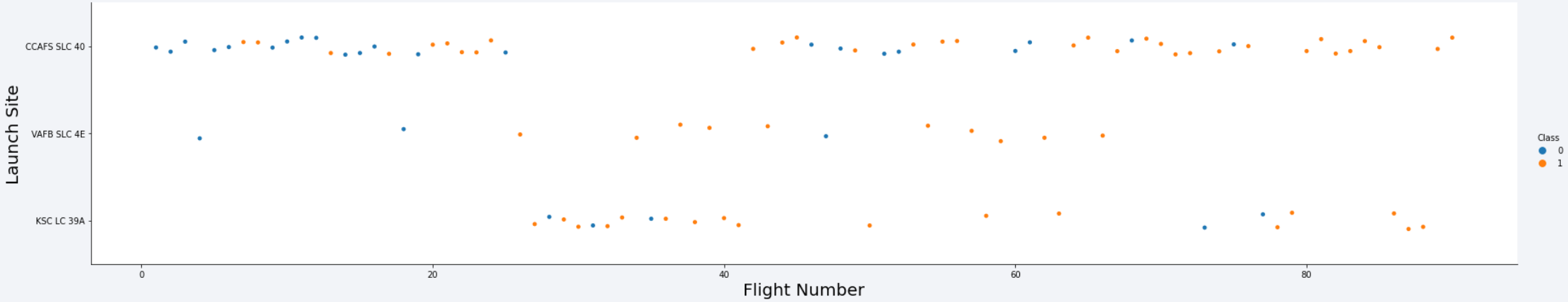
- The method that performs best using test data are all models with 83.33% accuracy

Section 2

Insights drawn from EDA

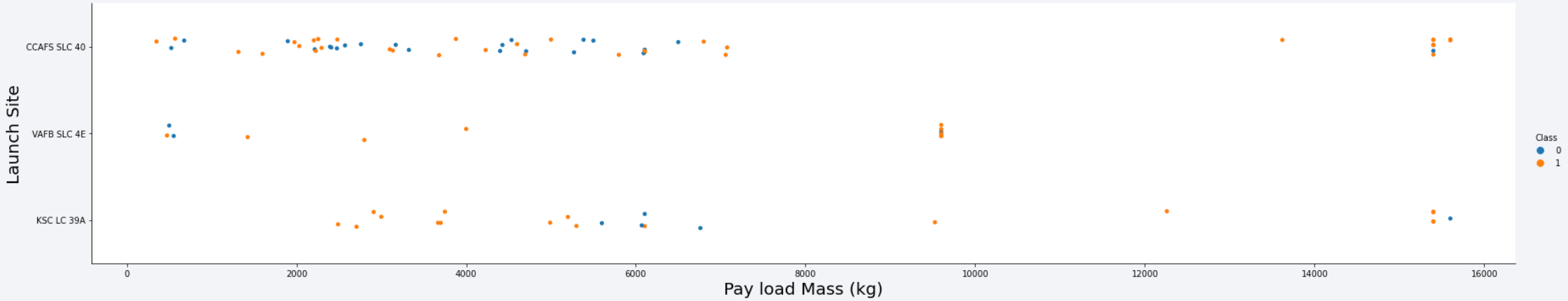


Flight Number vs. Launch Site



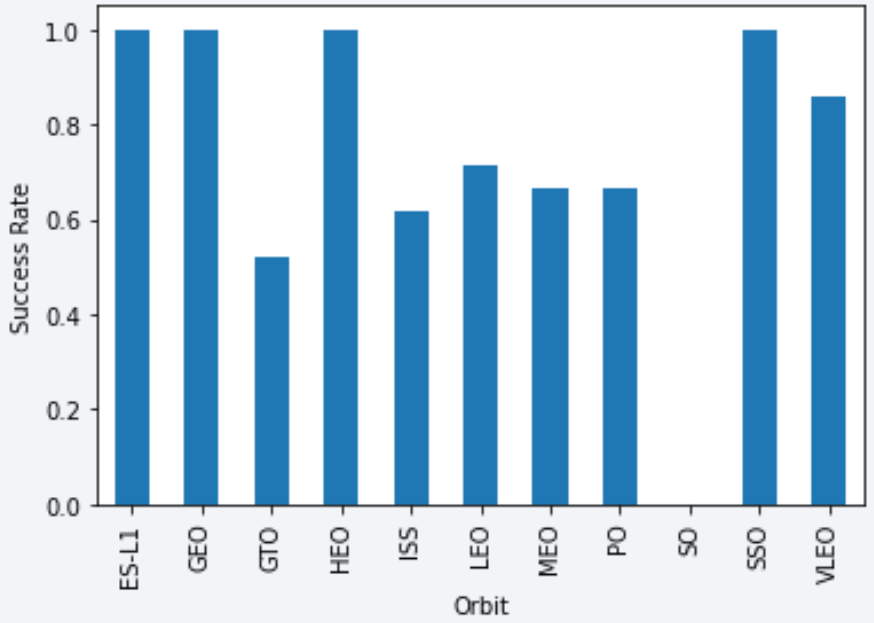
- As the flight number increases, the first stage is more likely to land successfully
- Different launch sites have different success rates
- CCAFS LC-40 has a success rate of 60%
- While KSC LC-39A and VAFB SLC 4E has a success rate of around 77%

Payload vs. Launch Site



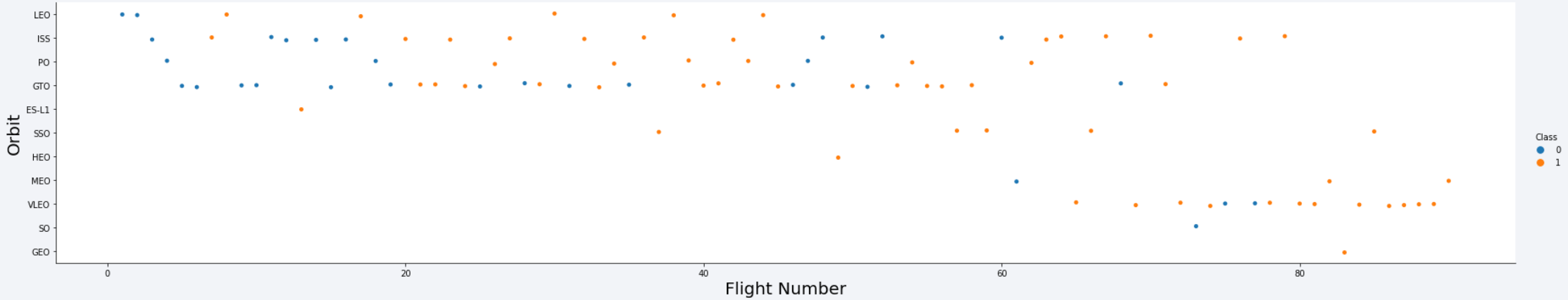
- It seems the more massive the payload, the less likely the first stage will return
- For the VAFB-SLC launch site, there are no rockets launched for heavy payload mass (greater than 10000)
- CCAFS LC-40 has a success rate of 60%, but if the mass is above 10,000 kg, the success rate is 100%

Success Rate vs. Orbit Type



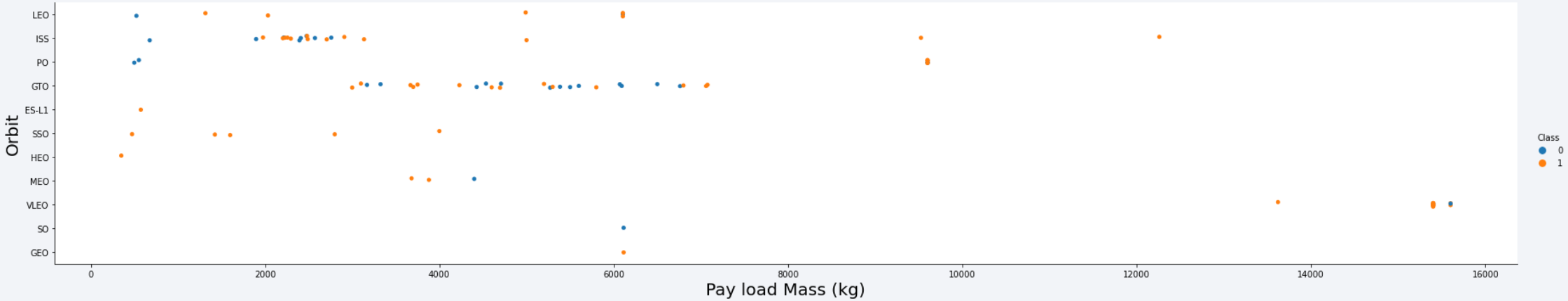
- ES-L1, GEO, HEO and SSO orbits have high success rate around 100%
- There is no successful landing in SO orbit

Flight Number vs. Orbit Type



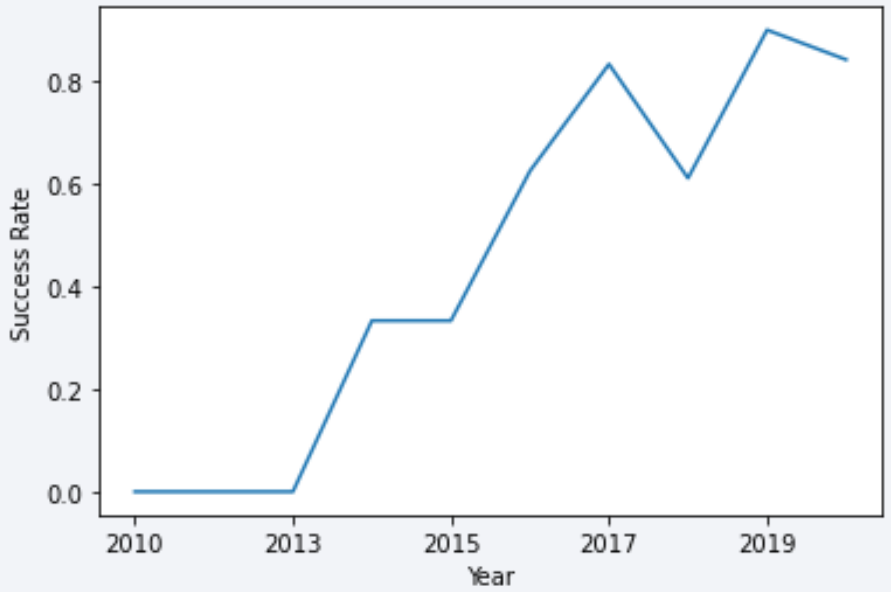
- In the LEO orbit, the Success appears related to the number of flights
- On the other hand, there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- 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 there here

Launch Success Yearly Trend



- The success rate since 2013 kept increasing till 2020

All Launch Site Names

The names of the unique launch sites in the space mission are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A and VAFB SLC-4E

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with 'CCA' as shown below

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

Total Payload Mass

The total payload mass carried by boosters launched by NASA (CRS) is 45,596 kg

total_payload_mass
45596

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1 is 2,928 kg

average_payload_mass
2928

First Successful Ground Landing Date

The date when the first successful landing outcome in ground pad was achieved is 22 December 2015

first_successful_date
2015-12-22

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 are F9 FT B1022, B1026, B1021.2 and B1031.2

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes is shown in the following

mission_outcome	number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass is listed below

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 are listed below

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

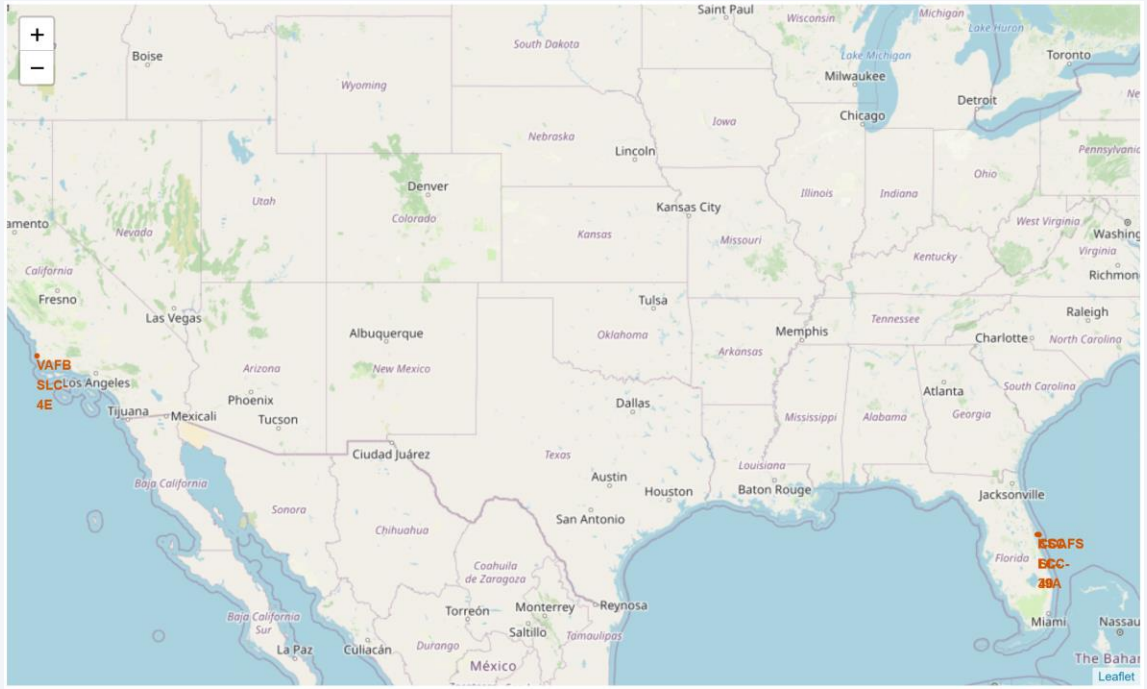
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

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

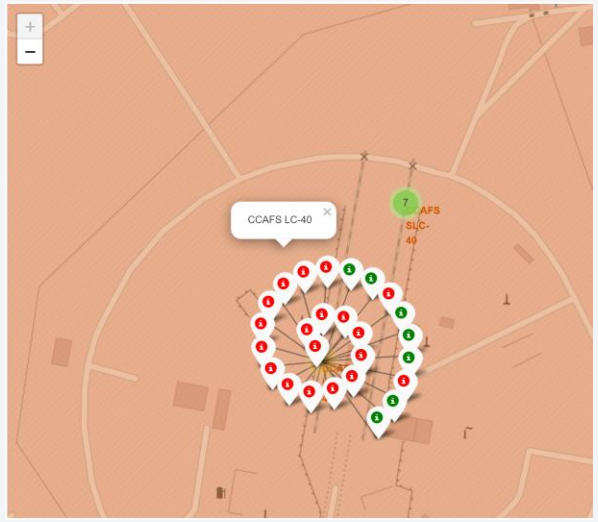

Section 3
Launch Sites
Proximities Analysis

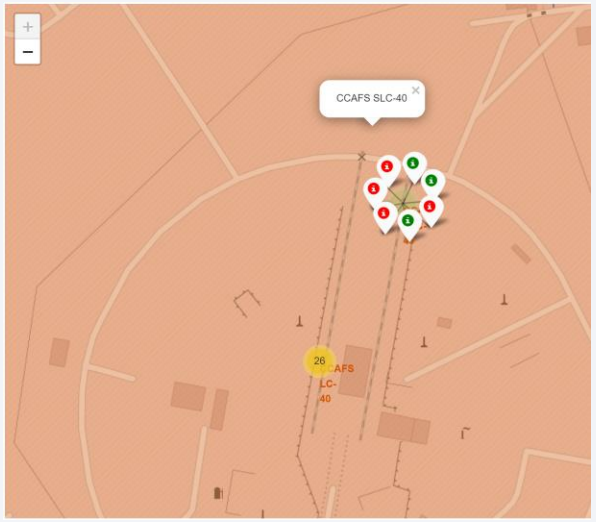
All launch sites' location

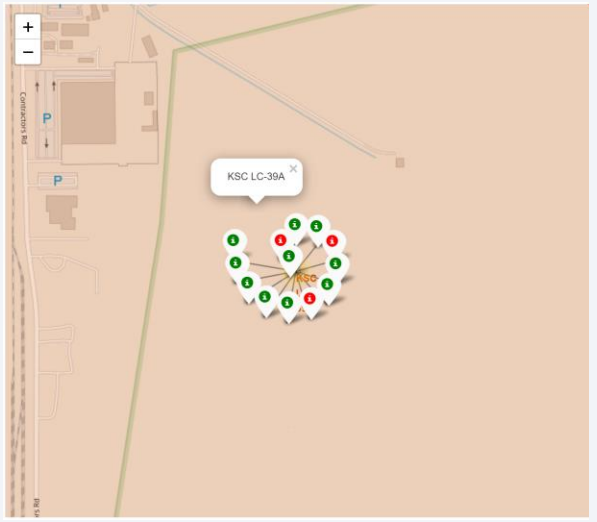


According to the map, all launch sites are in very close proximity to the coast but not in proximity to the Equator line

Color-labeled launch outcomes





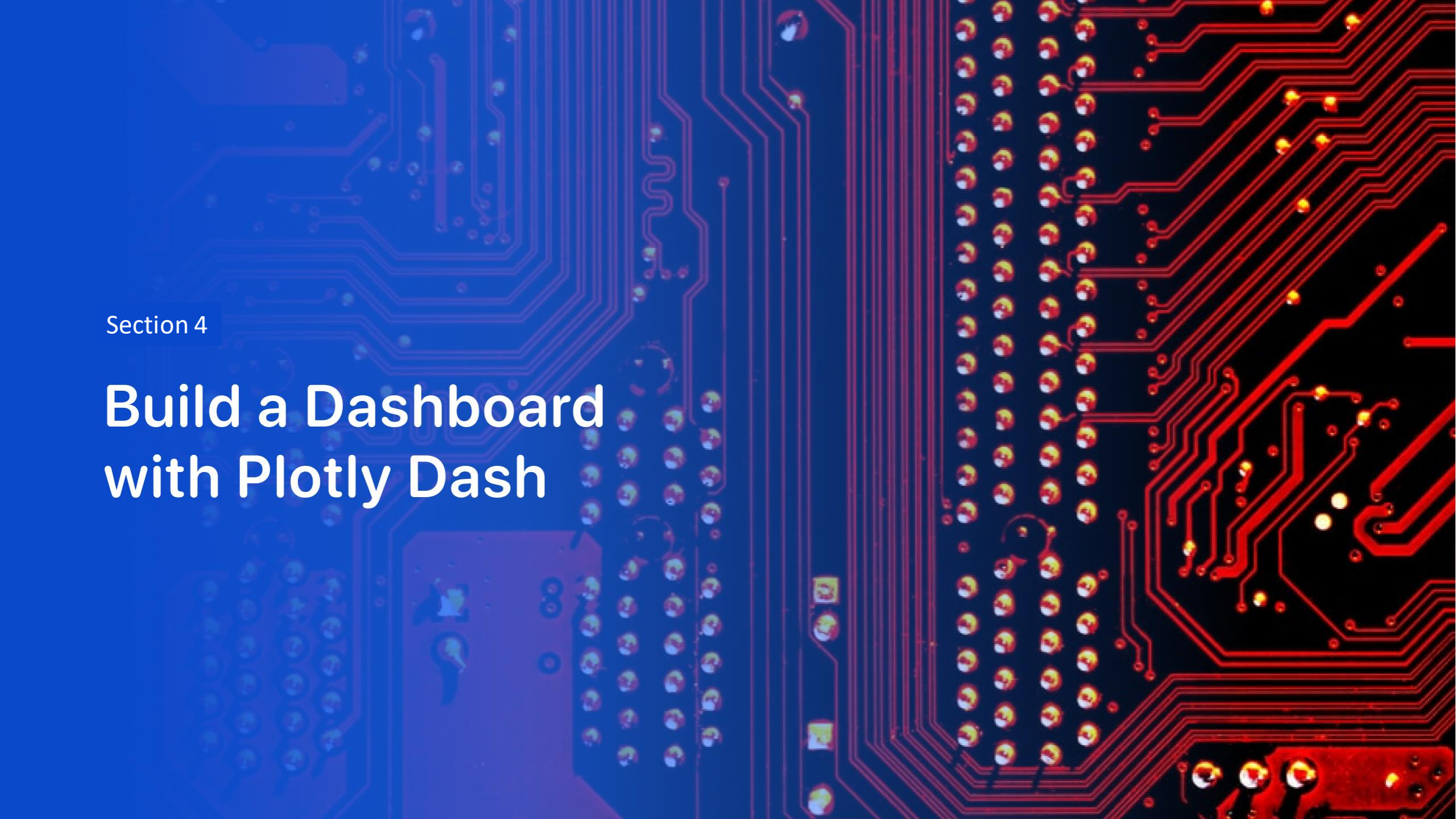


Use a green marker if a launch was successful and a red marker if a launch was failed

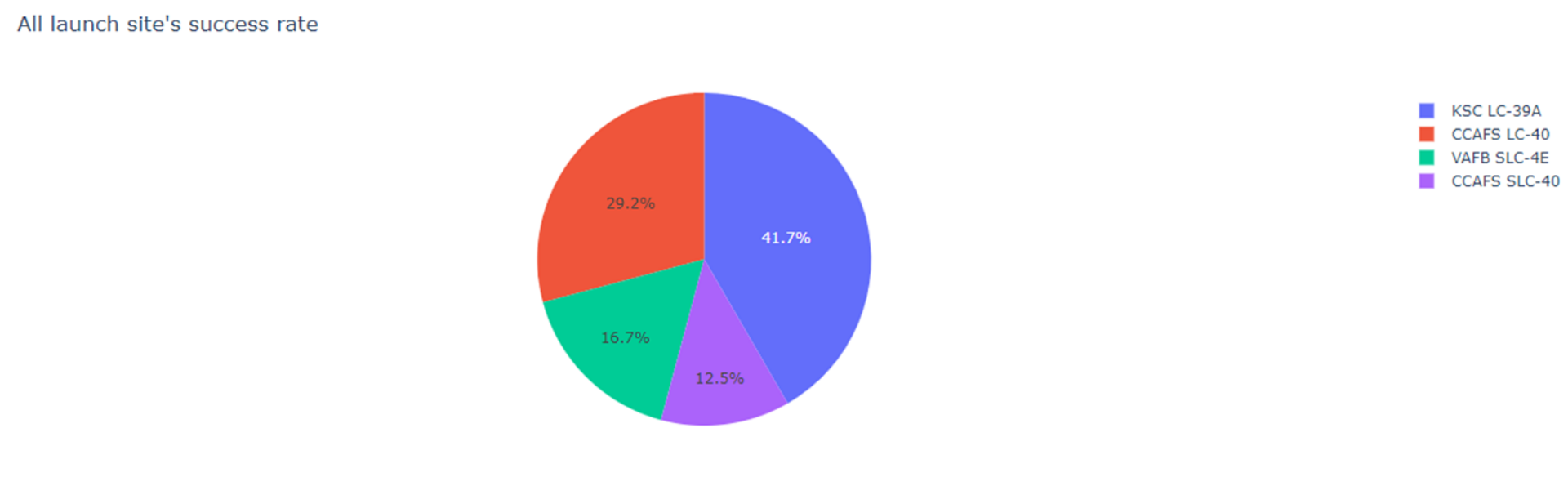
As illustrated on the maps for all launch sites, it can easily identify that KSC LC-39A has the relatively highest success rate

Distance between the launch site to its proximities

For CCAFS SLC-40 launch site, the distance between this launch site and the closet coastline is 0.9 km approximately


Section 4
Build a Dashboard with Plotly Dash

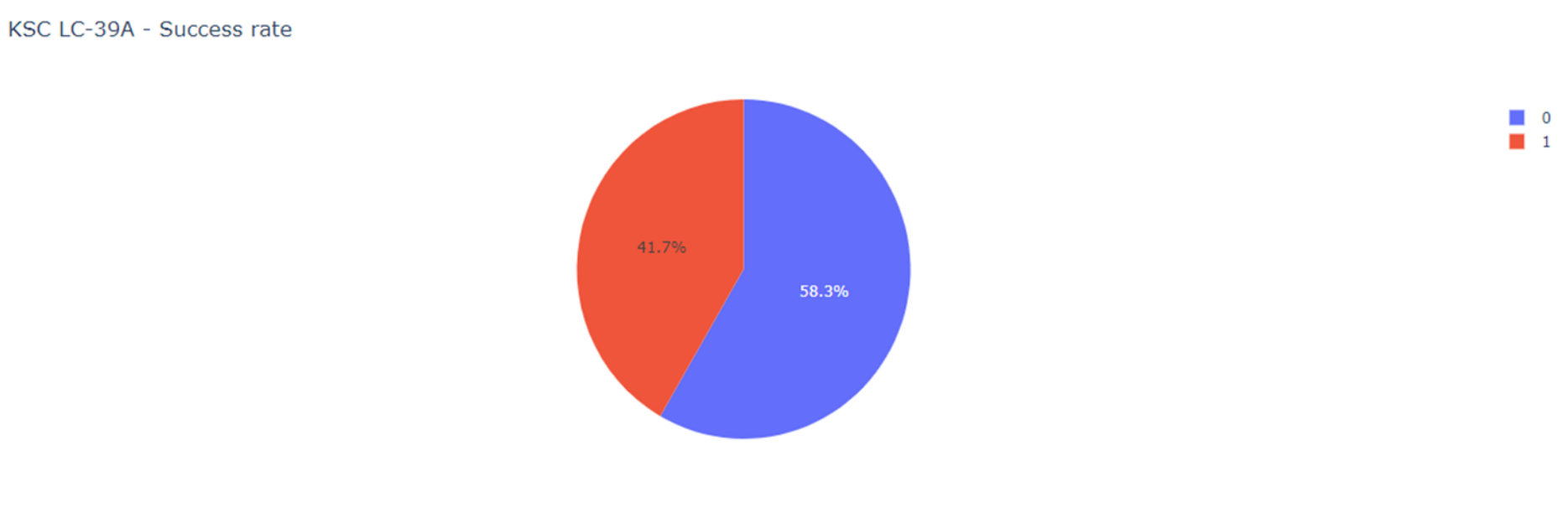
Total Success Launches By Site



As illustrated on the pie chart,

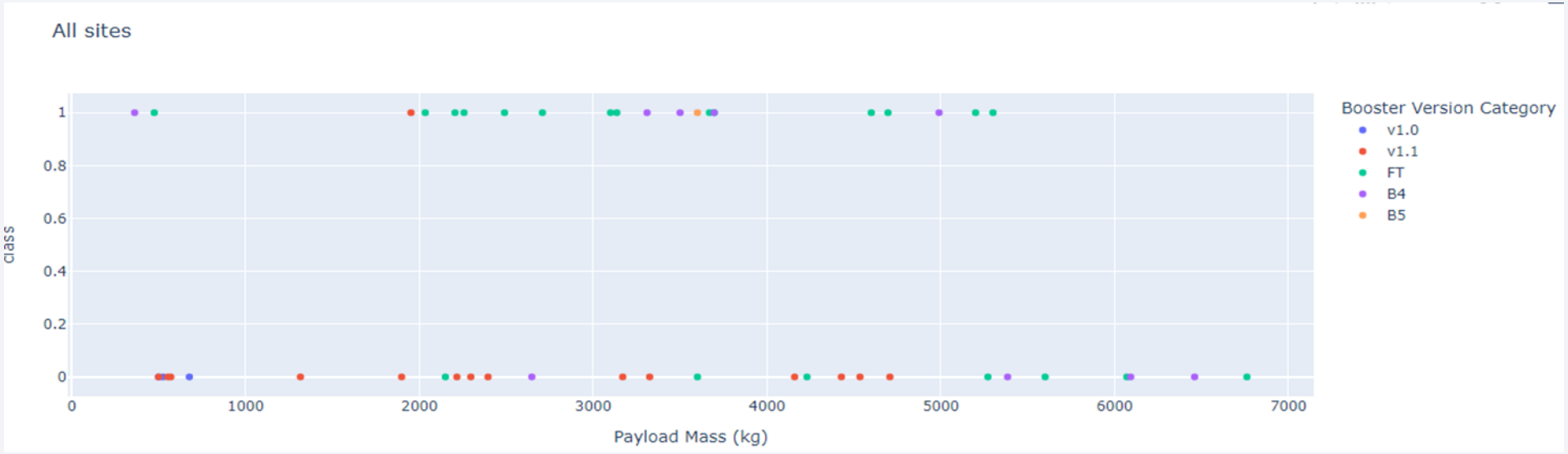
KSC LC-39A is a launch site that has the largest successful launches

Launch Site with Highest Launch Success Ratio



Moreover, the pie chart shows that KSC LC-39A also has the highest success rate among other launch sites equals 41.7%

Payload vs. Launch Outcome



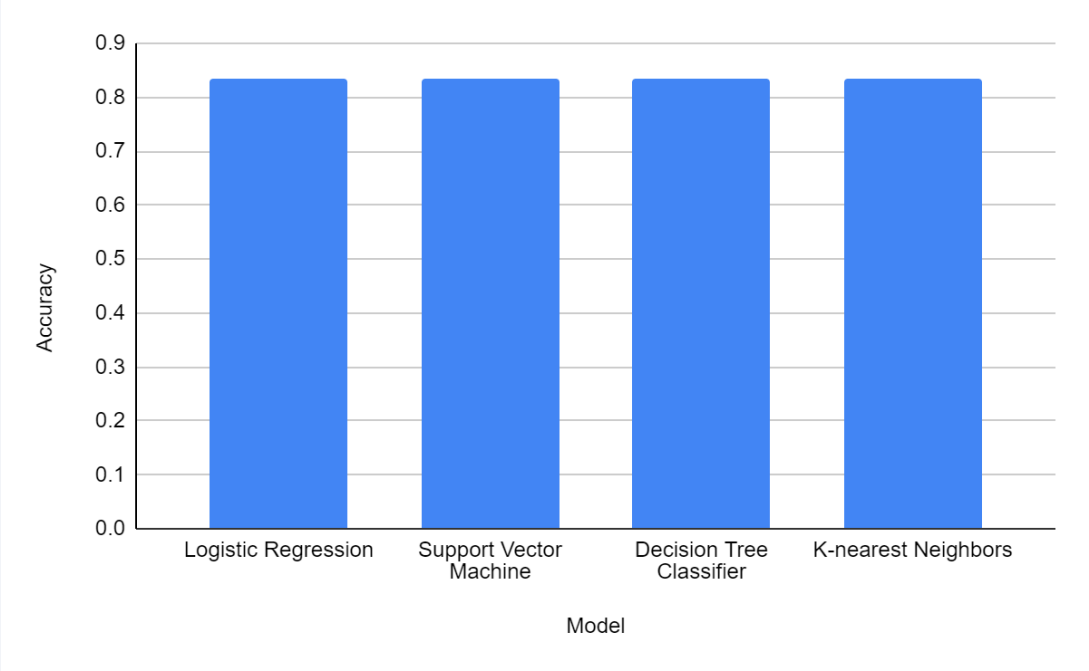
According to the scatter plot,

- Payload Mass in range 2,000 – 4,000 kg has the highest launch success rate
- Payload Mass below 2,000 and above 6,000 has the lowest launch success rate
- F9 Booster Version FT has the highest launch success rate opposite to v1.1

Section 5

Predictive Analysis (Classification)

Classification Accuracy



All models have the highest classification accuracy equals 83.33%

Confusion Matrix

According to the confusion matrix of the best performing model, the major problem is false positives

Conclusions

- As the flight number increases, the first stage is more likely to land successfully
- It seems the more massive the payload, the less likely the first stage will return
- Most of successful landing outcomes are on drone ship
- All launch sites are in very close proximity to the coast
- KSC LC-39A is a launch site that has the highest success rate
- Payload Mass in range 2,000 – 4,000 kg has the highest launch success rate
- F9 Booster Version FT has the highest launch success rate
- All machine learning models perform best with 83.33% accuracy using test data

Appendix

- GitHub URL of the project: rockslpy/spacex-falcon9-ml-prediction (github.com)

Thank you!