IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hugo Hernández
02/16/2022

Outline

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

Executive Summary

The methods used in this project include:

- Data collection using APIs, SQL and Web Scraping.
- Data Wrangling
- EDA with Data Visualization and SQL
- Creating interactive maps using Folium
- Building interactive dashboards using Dash
- Classification using ML

Project's main results:

- The best classifier was the Decission Tree.
- It had an accuracy of 88% on the test set.
- Low weighted payloads perform better than the heavier payloads
- The success rate improved during the years 2013-2020.

Executive Summary

The methods used in this project include:

- Data collection using APIs, SQL and Web Scraping.
- Data Wrangling
- EDA with Data Visualization and SQL
- Creating interactive maps using Folium
- Building interactive dashboards using Dash
- Classification using ML

Project's main results:

- The best classifier was the Decission Tree.
- It had an accuracy of 88% on the test set.
- Low weighted payloads perform better than the heavier payloads.
- The success rate improved during the years 2013-2020.

Introduction

- As a company that wants to bid against SpaceX for rocket launches, it would be useful to predict the cost of Falcon 9 launches.
- SpaceX advertises Falcon 9 rocket launches much cheaper than competitors, much of the savings are 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.

Mail goals:

- Predict whether a new launch will land succesfuly.
- Find the correlation between launch attributes and outcome of landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using SpaceX Rest API. [1]
  - Web Scraping from Wikipedia. [2]
- Perform data wrangling
  - Irrelevant columns were dropped.
  - Data types were assesed and corrected.
  - NaN or null values were replaced with column mean value.
  - One Hot Encoding was used to prepare categorical columns for ML.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

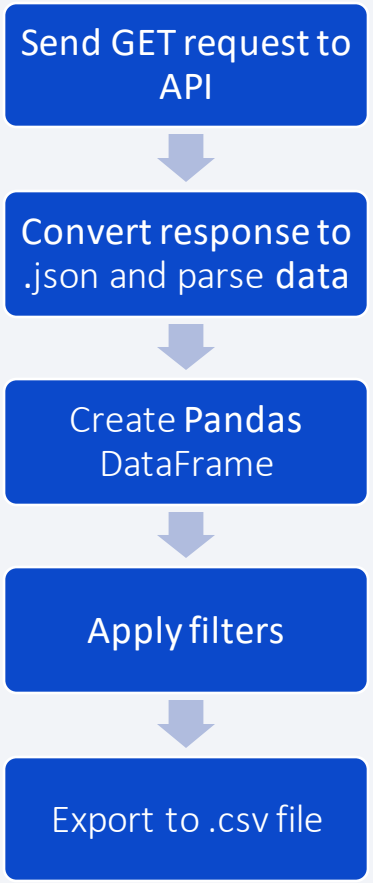
- Perform predictive analysis using classification models
    - SVM, Logistic Regression, Decission Tree, KNN.
    - Hyperparemeter tuning using GridSearchCV on a dev set.
- Best model found using accuracy on a test set.

Data Collection

- Data sets were downloaded from the internet using the SpaceX Rest API and using Web Scraping.
- Filtering and cleaning were applied.
- The data was stored as .csv files.

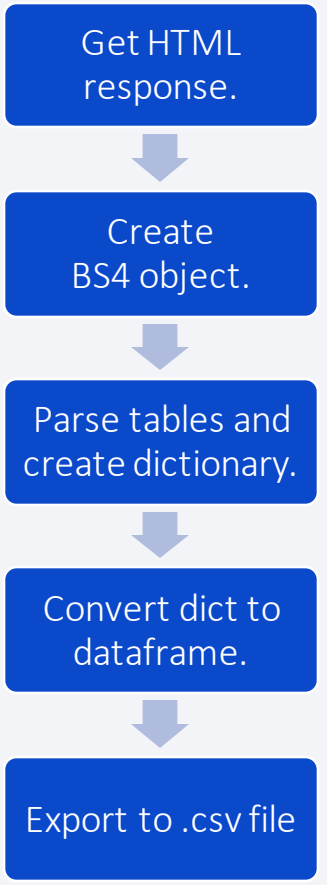
Data Collection – SpaceX API

- The launch data was gathered from the SpaceX REST API using GET.
- Receive response as a .json file.
- Convert json response to Pandas DataFrame.
- Filtering and Data Wrangling.
- Store data as .csv file.
- A Jupyter Notebook was used for the process. [3]



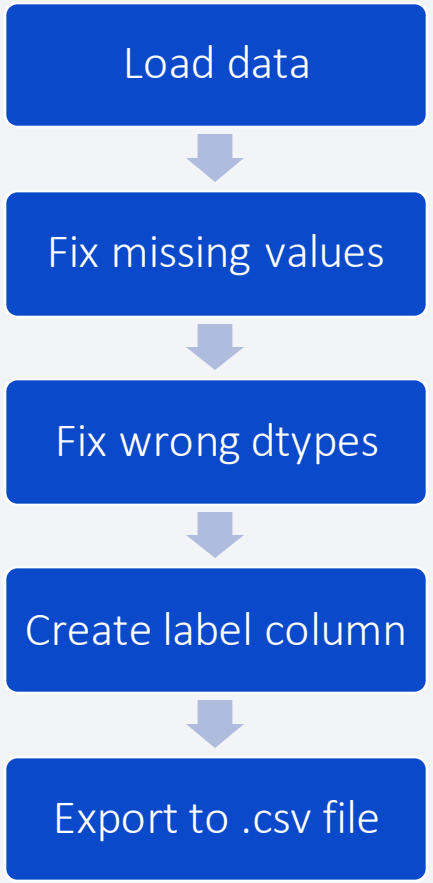
Data Collection - Scraping

- The launch data was gathered from Wikipedia. [2]
- Create BeautifulSoup object with response.
- Find all tables within the object.
- Create dictionary with the table data.
- Create Pandas DataFrame using the dictionary.
- Store data as .csv file.
- A Jupyter Notebook was used for the process. [4]



Data Wrangling

- Data was loaded from a .csv file and stores as a Pandas DataFrame.
- Dealing with null values.
- Fixed columns with wrong data types.
- Convert Landing Outcome as Boolean type.
- Export to .csv file.
- A Jupyter Notebook was used for the process. [5]



EDA with Data Visualization

- **Scatter plot between Flight Number and Launch Site, with Outcome as label:** helps to visualize the relationship between SpaceX flight experience and Success.
- **Scatter plot between Payload and Launch Site:** helps to visualize the relationship between the Payload mass carried and Success.
- **Bar plot of success rate for each orbit type:** checks if there is any relationship between success rate and orbit type.
- **Scatter plot between Flight Number and Orbit type:** helps to visualize the relationship between SpaceX flight experience and Success.
- **Scatter plot between Payload and Orbit type:** helps to visualize the relationship between the Payload mass carried and Success.
- **Line plot for launch success yearly trend:** to visualize change in success rate over the years.
- A Jupyter Notebook was used for the process. [6]

EDA with SQL

- Display the names of the unique launch sites in the space mission.
- Display five records where launch sites begin with the string 'KSC’.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Display average payload mass carried by booster version F9 v1.1.
- List the date where the successful landing outcome in drone ship was achieved.
- List the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.

EDA with SQL

- List the names of the booster versions which have carried the maximum payload mass.
- List the records which will display the month names, successful landing outcomes in ground pad ,booster versions, launch site for the months in year 2017.
- Rank the count of successful landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.
- A Jupyter Notebook was used for the process. [7]

Build an Interactive Map with Folium

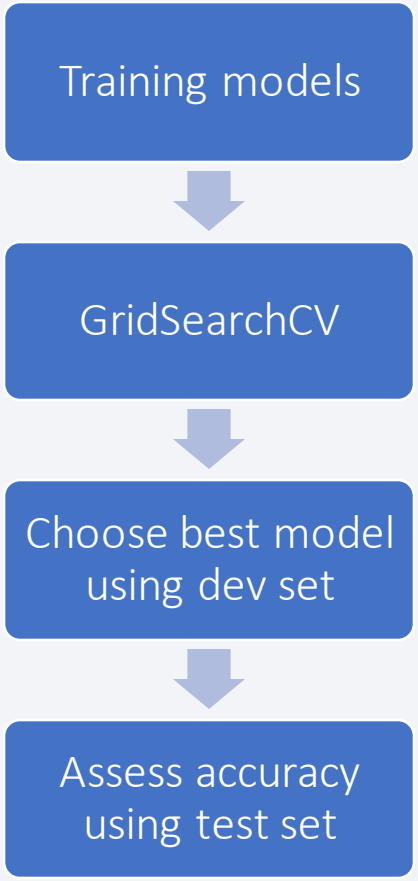
- To visualize the data of each launch site we added a Circle Marker around each launch site with a label of the name of the launch site.
- Different color was used for launch sites with success or failed landings.
- Created Marker Clusters to simplify the map.
- Created lines and measured the distance from launch site to certain landmarks.
- A Jupyter Notebook was used for the process. [8]

Build a Dashboard with Plotly Dash

- A dashboard was created for interactive visualization:
- The dashboard contained:
  - Pie Chart showing the percentage of launches in each site
  - Pie Chart showing success rate of each site
  - Scatter Graph for the Outcome vs. Payload Mass (Kg) for different Booster Versions
- A Jupyter Notebook was used for the process. [9]

Predictive Analysis (Classification)

- Perform predictive analysis using four classification models
    - SVM, Logistic Regression, Decission Tree, KNN.
- Check accuracy for each model.
- Hyperparemeter tuning using GridSearchCV on a dev set.
- Best model found using accuracy on a dev set.
- Assess model score using accuracy on a test set.
- A Jupyter Notebook was used for the process. [10]

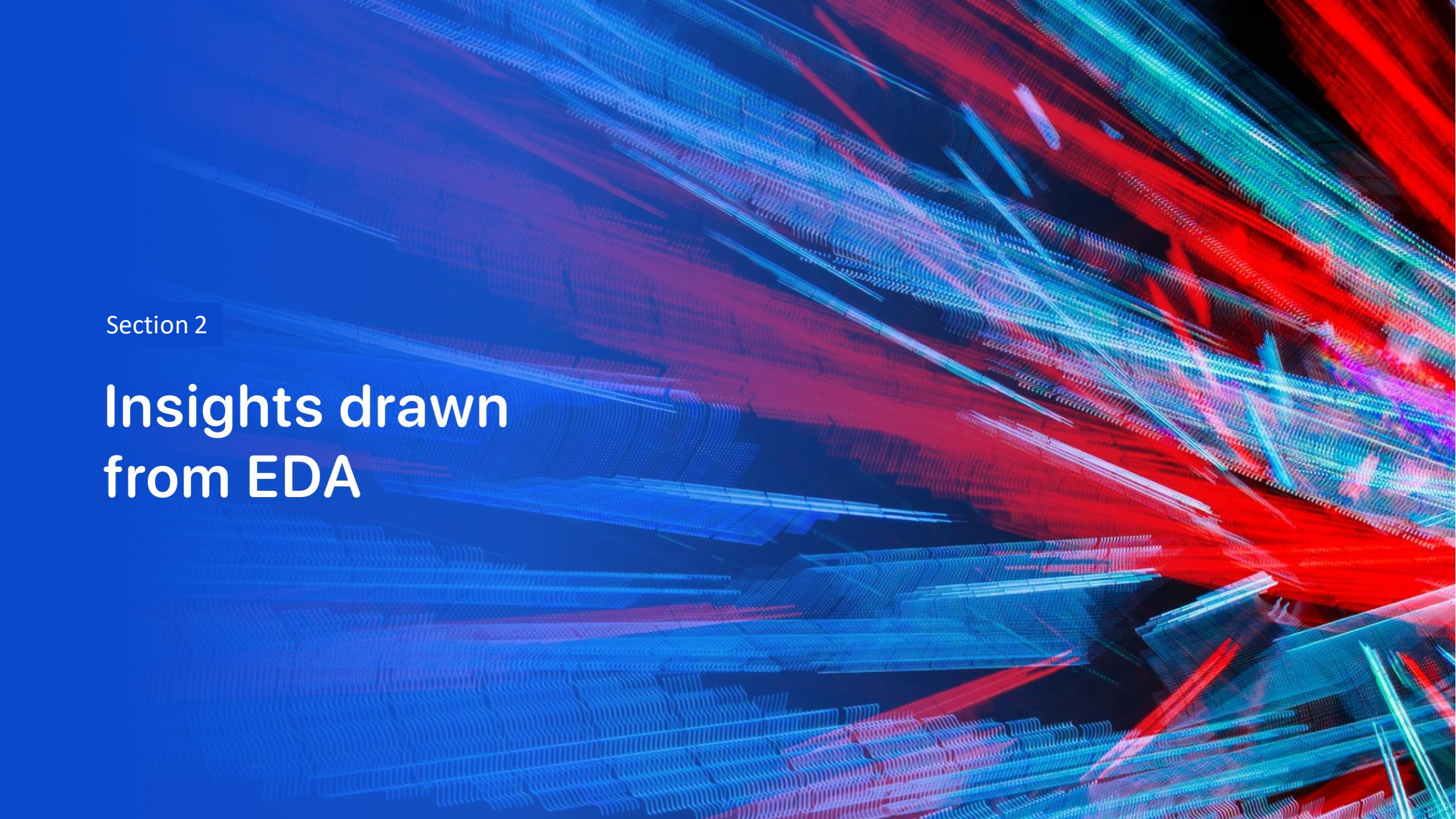


Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

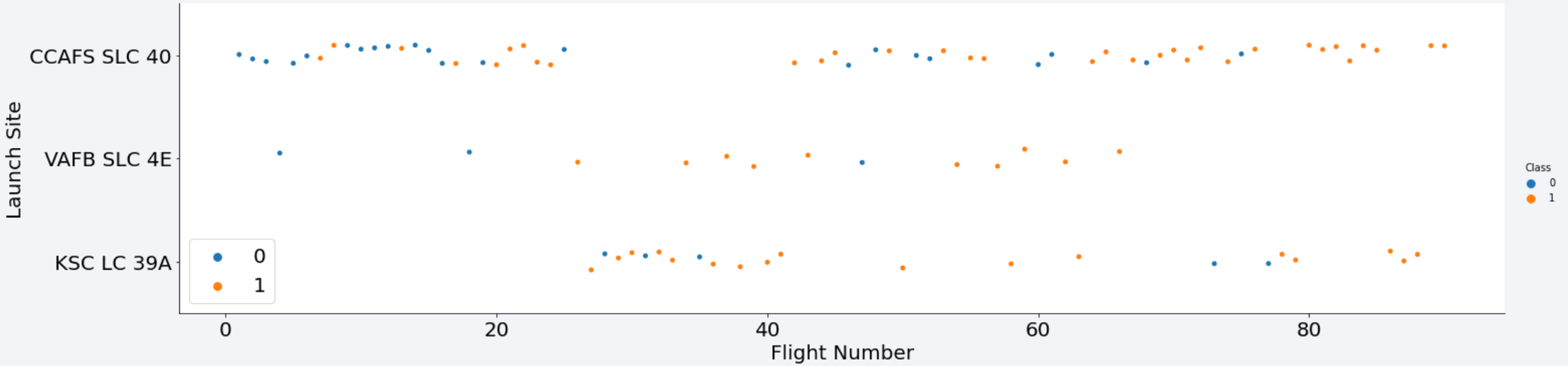
Section 2

Insights drawn from EDA



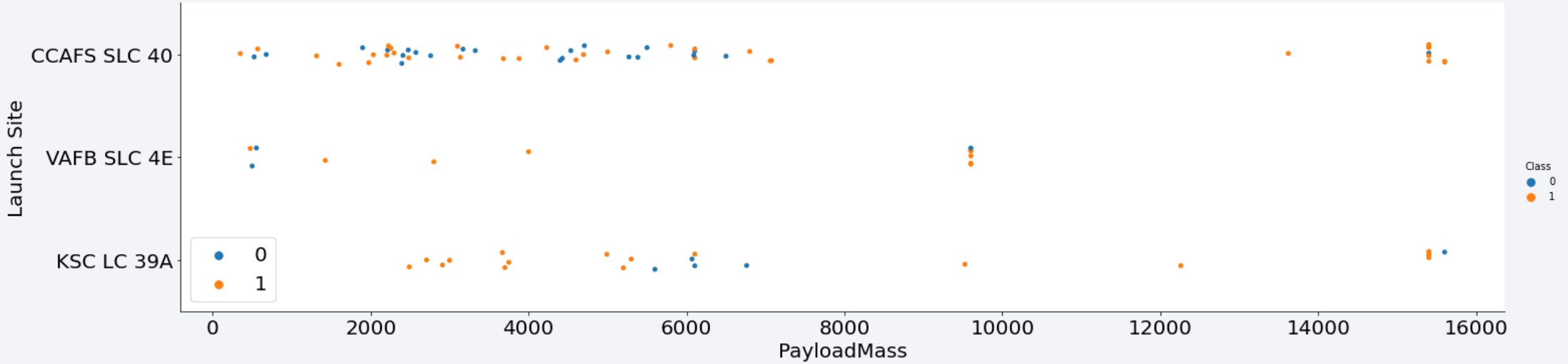
Flight Number vs. Launch Site

- Launches from the site of CCAFS SLC 40 are significantly higher than launches form other sites.



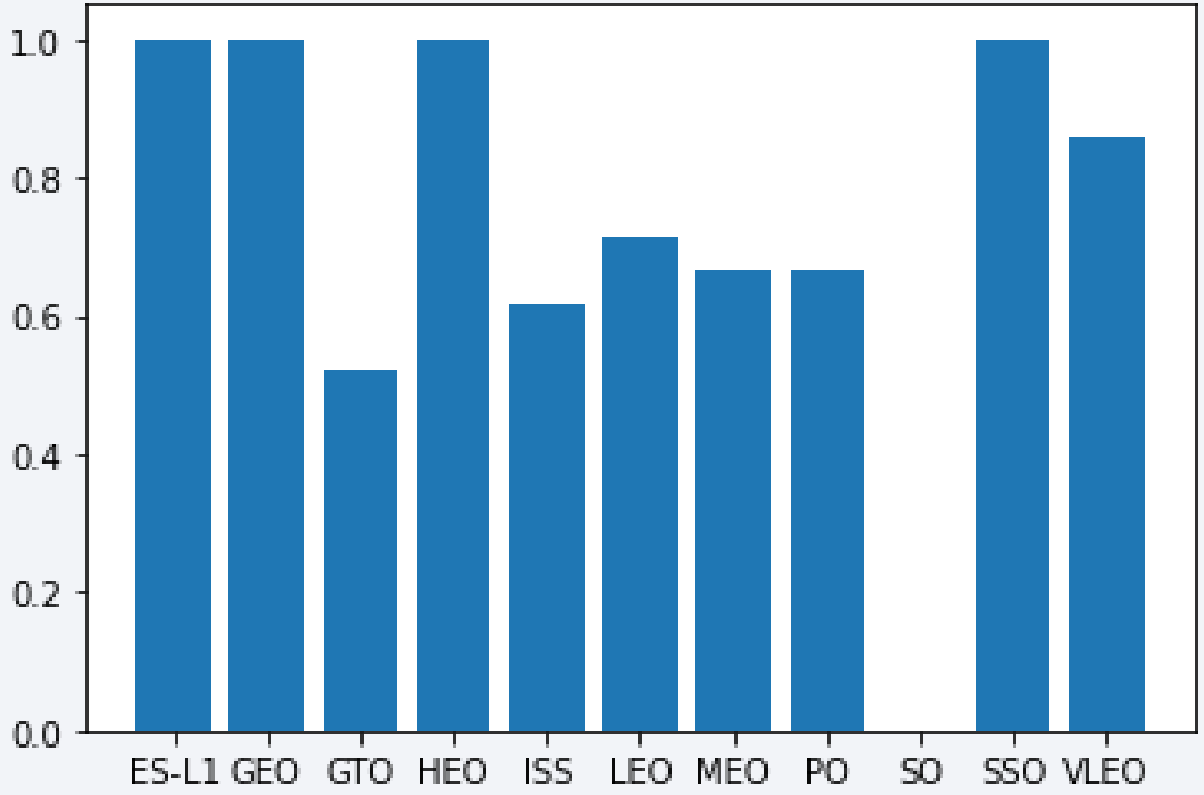
Payload Mass vs. Launch Site

- In the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



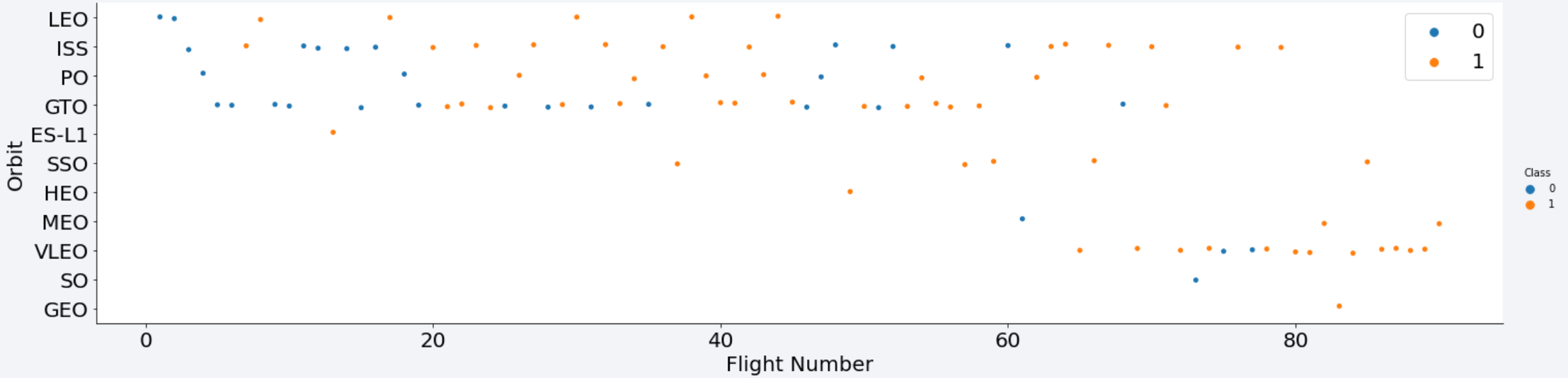
Success Rate vs. Orbit Type

- Different orbits have different mean success rates.



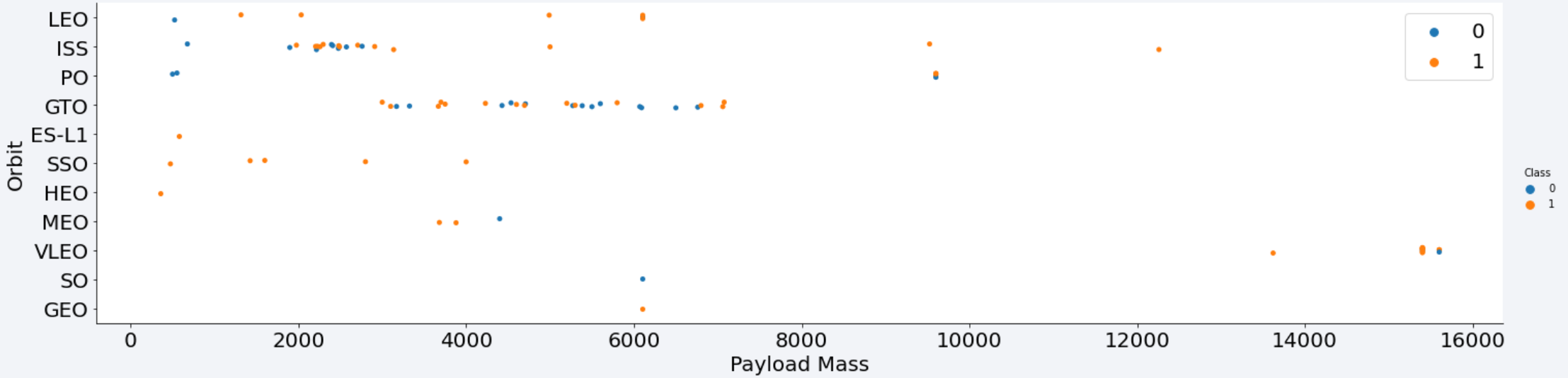
Flight Number vs. Orbit Type

- 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.



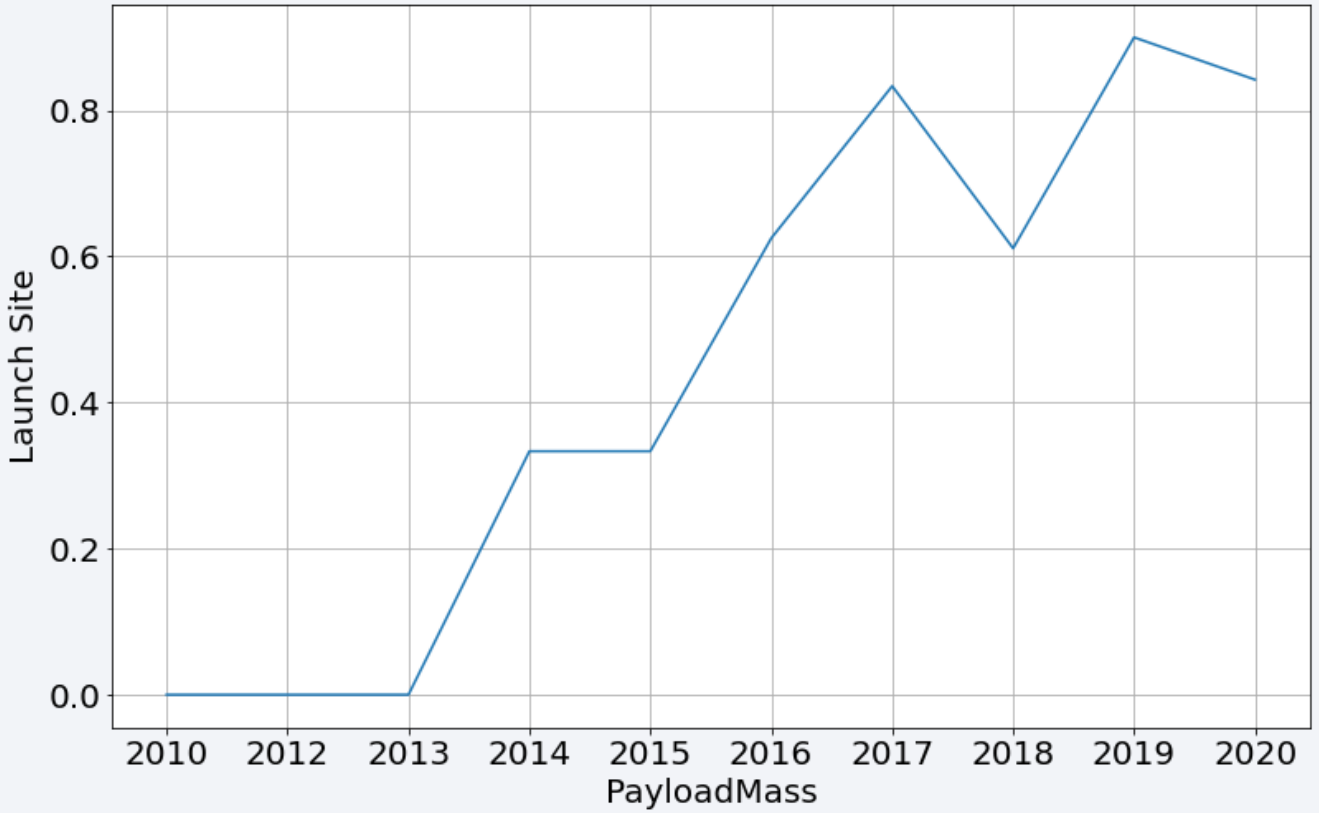
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.



Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020



All Launch Site Names

- All the names of the sites of the Falcon 9 launches.

Launch Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

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

index	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft...	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
48213

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2534.67

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

MIN(Date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Date	Landing_Outcome	PAYLOAD_MASS_KG	Booster_Version
06-05-2016	Success (drone ship)	4696	F9 FT B1022
14-08-2016	Success (drone ship)	4600	F9 FT B1026
30-03-2017	Success (drone ship)	5300	F9 FT B1021.2
11-10-2017	Success (drone ship)	5200	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - Success: 98
  - Failure (in flight): 1
  - Success (payload status unclear): 1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

Booster_Version	PAYLOAD_MASS_KG
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Date	Landing_Outcome	Booster_Version	Launch_Site
10-01-2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

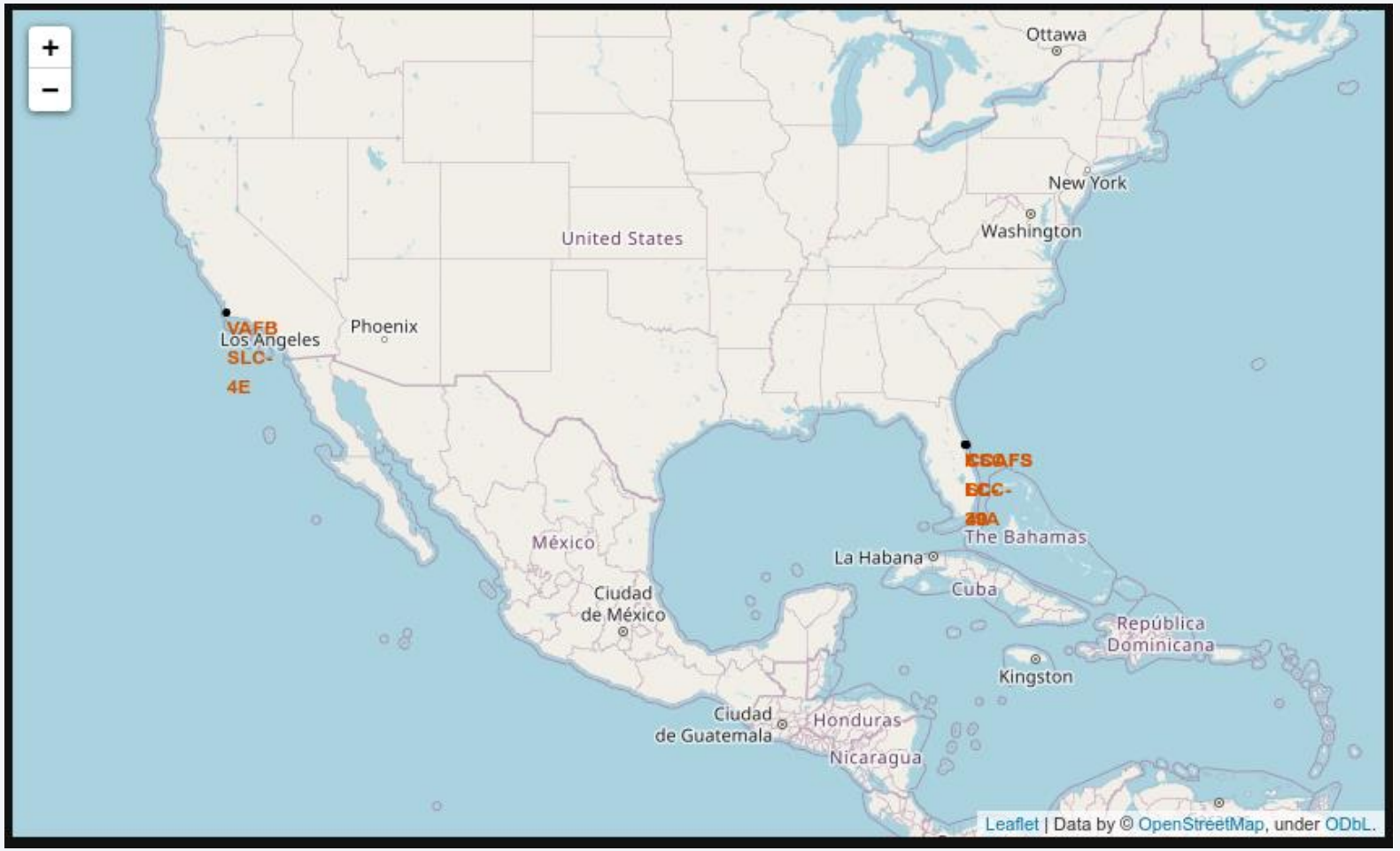
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	Count
Success	19
No attempt	9
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
No attempt	1
Failure (parachute)	1


Section 3
Launch Sites
Proximities Analysis

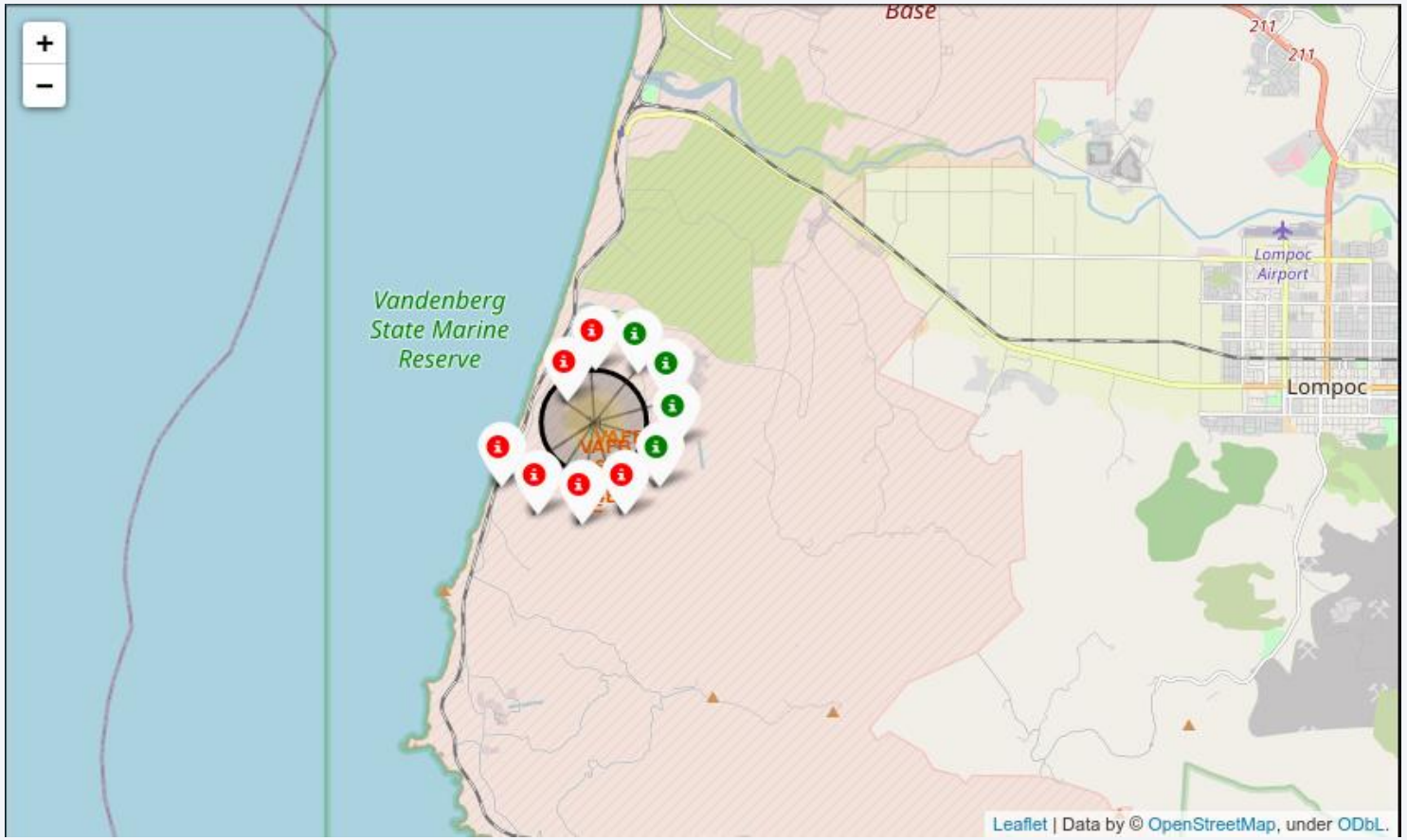
Falcon 9 Launch Sites

- Map showing Falcon 9 launch sites



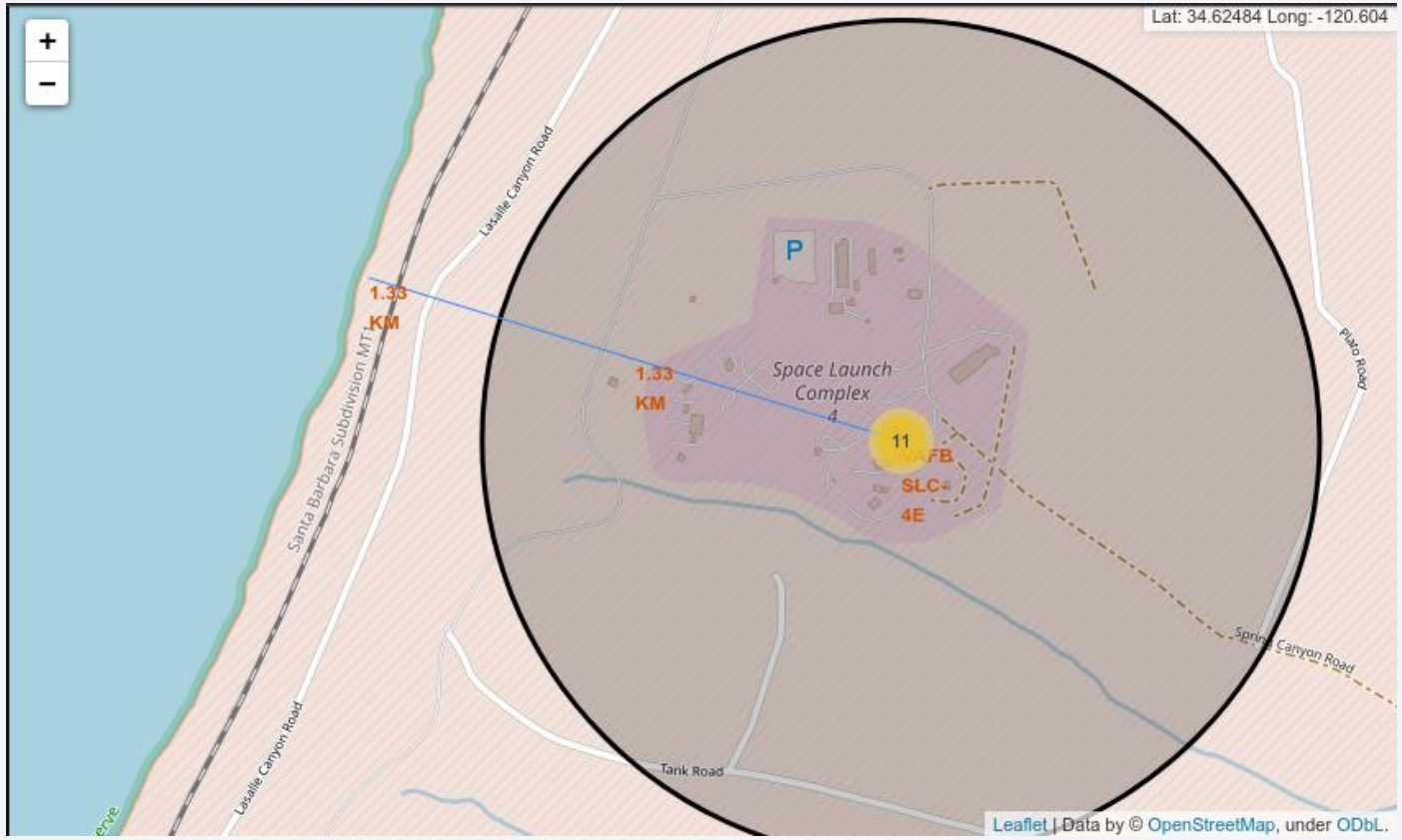
Success and failed Launch Sites

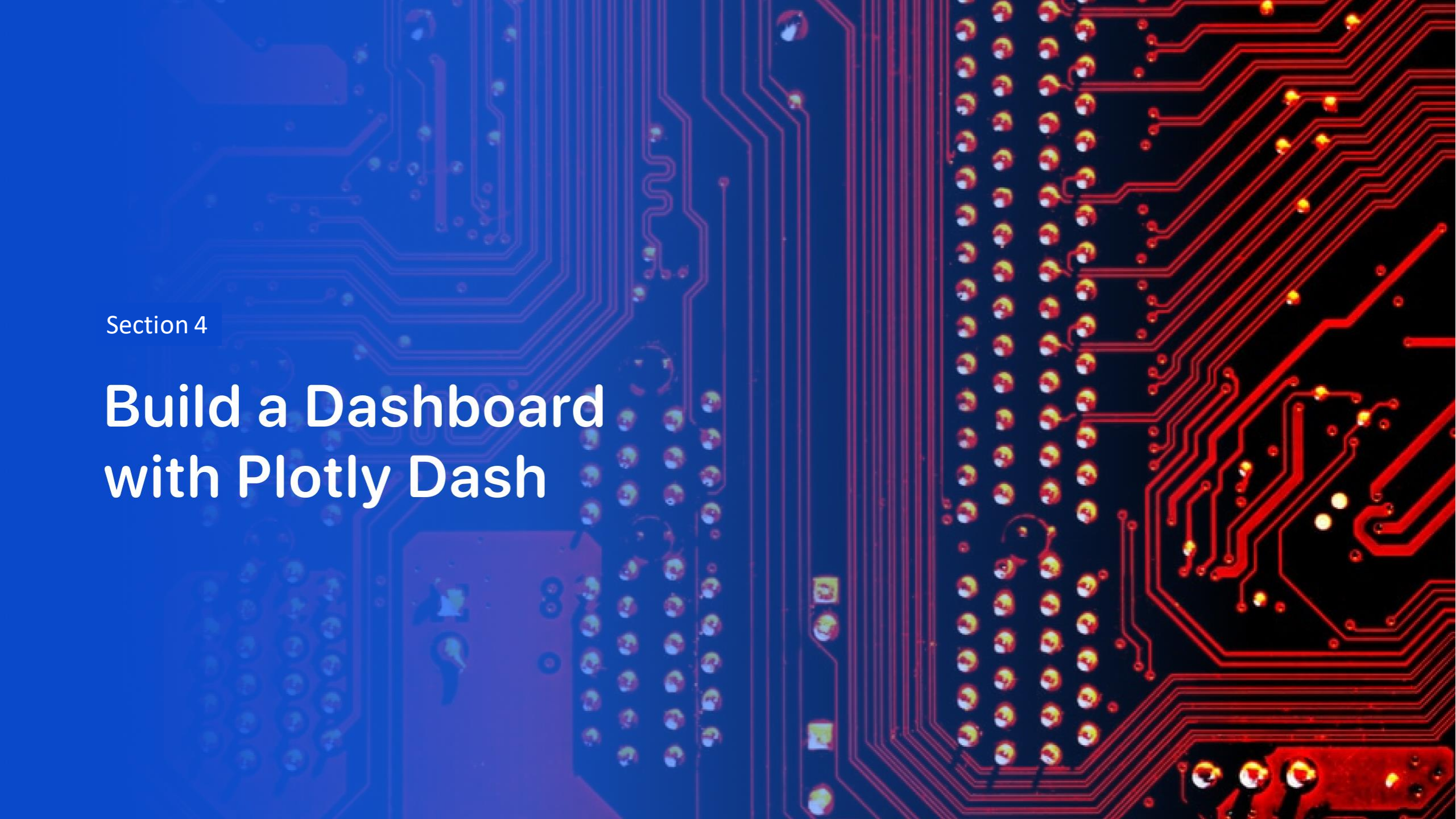
- The successful launch sites are marked in green, and failed in red.



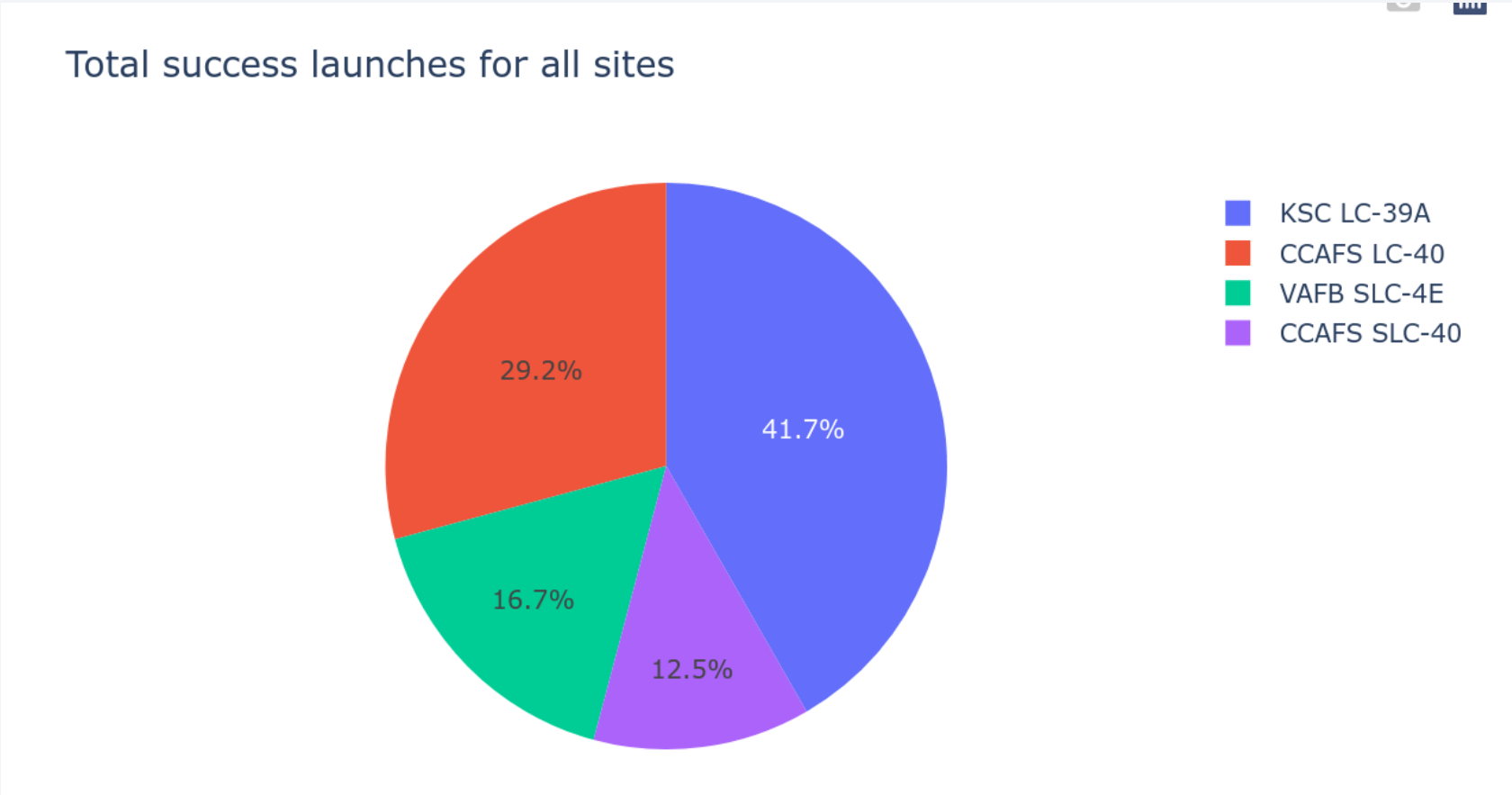
Distance from coastline to launch site.

- Distance from launch site to coastline.

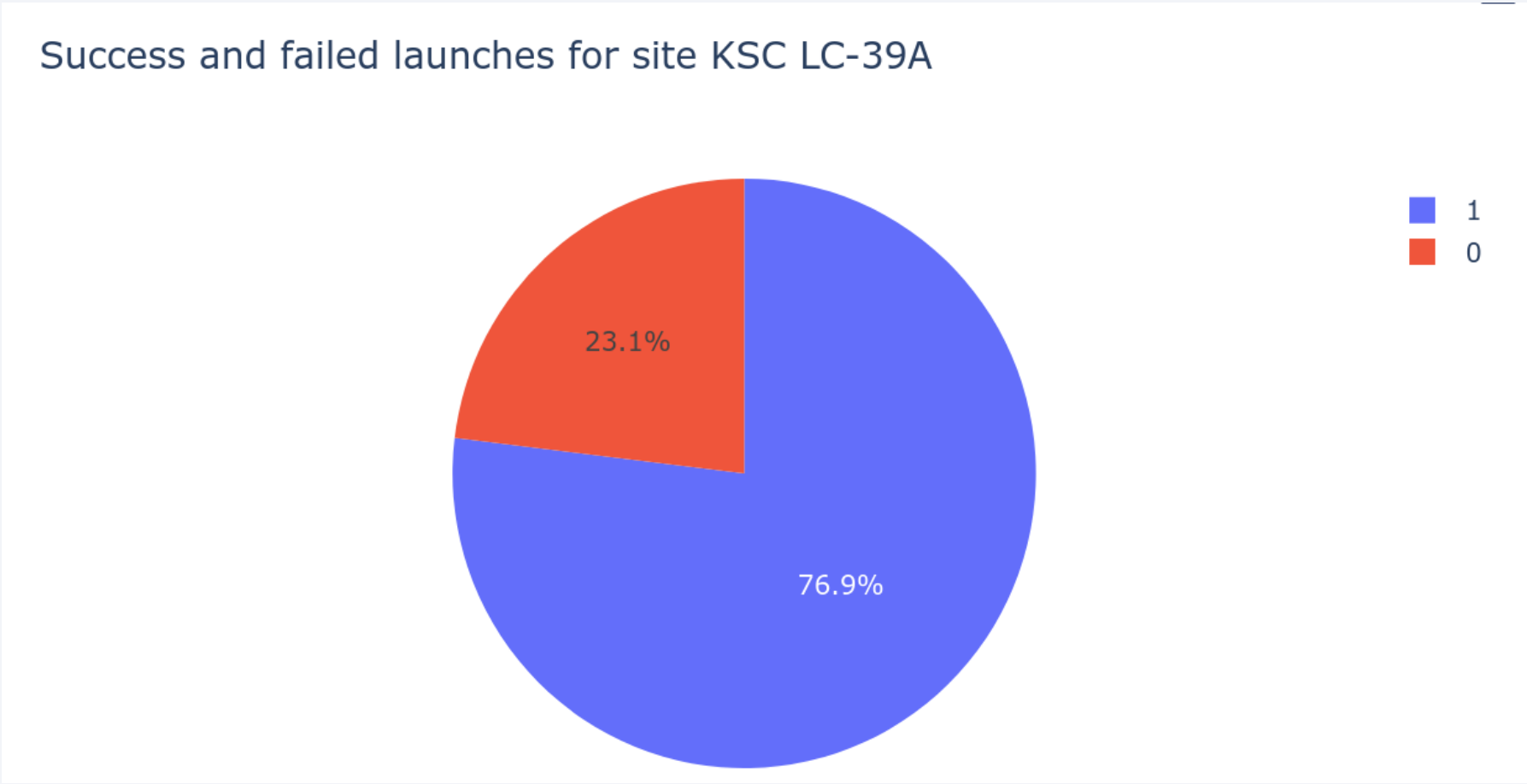



Section 4
Build a Dashboard with Plotly Dash

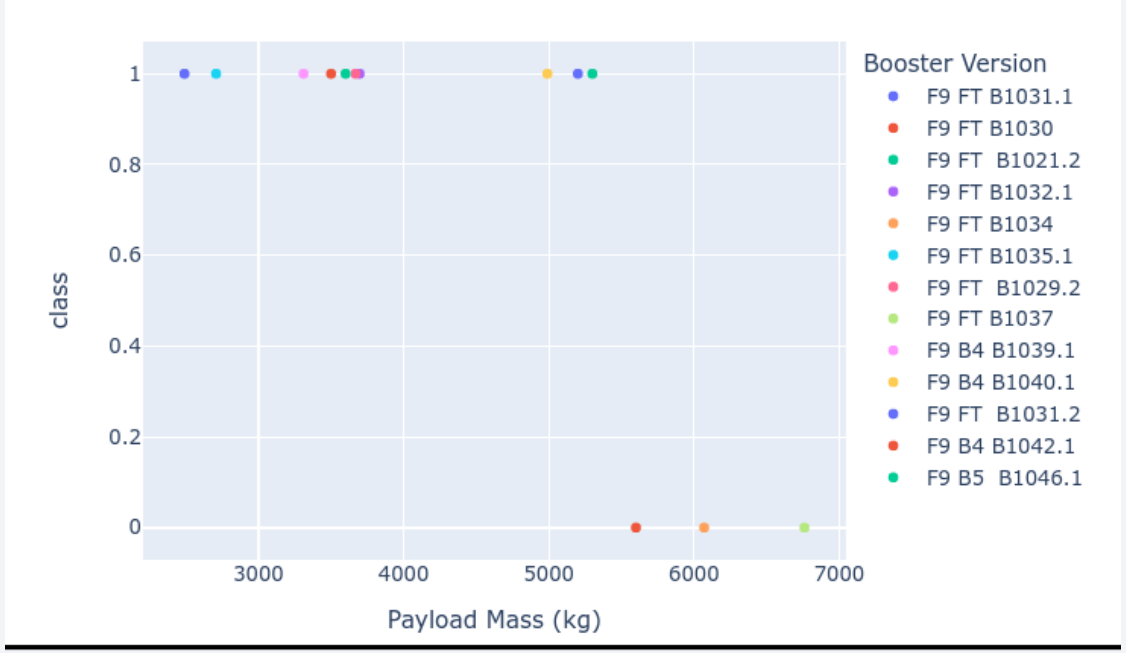
Launch success count for all sites

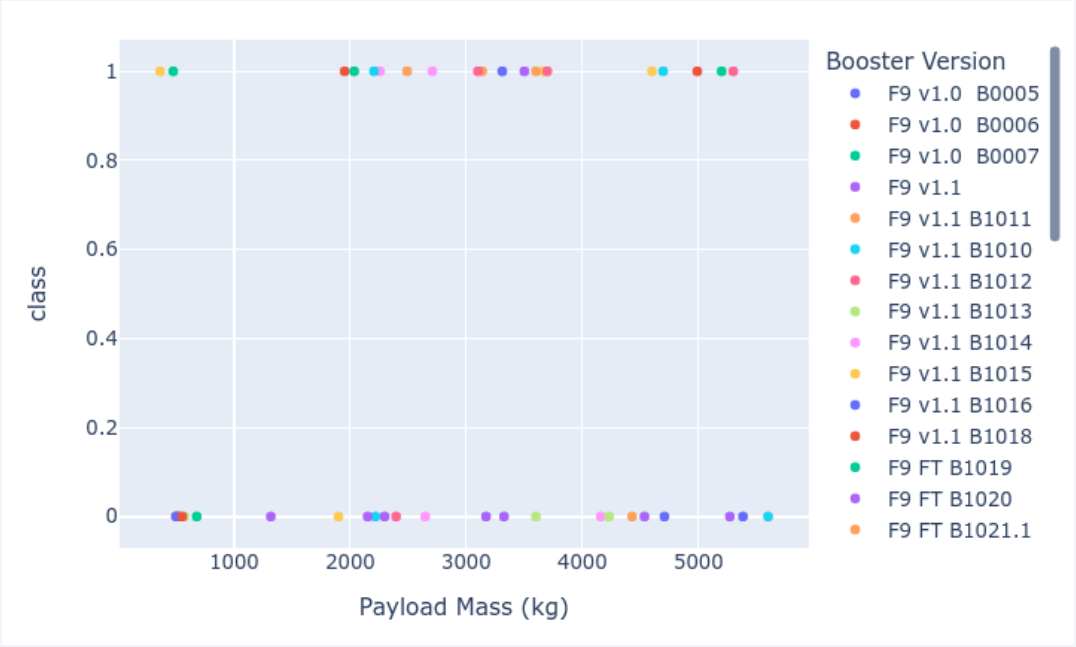


Launch site with highest success



Success vs Payload mass



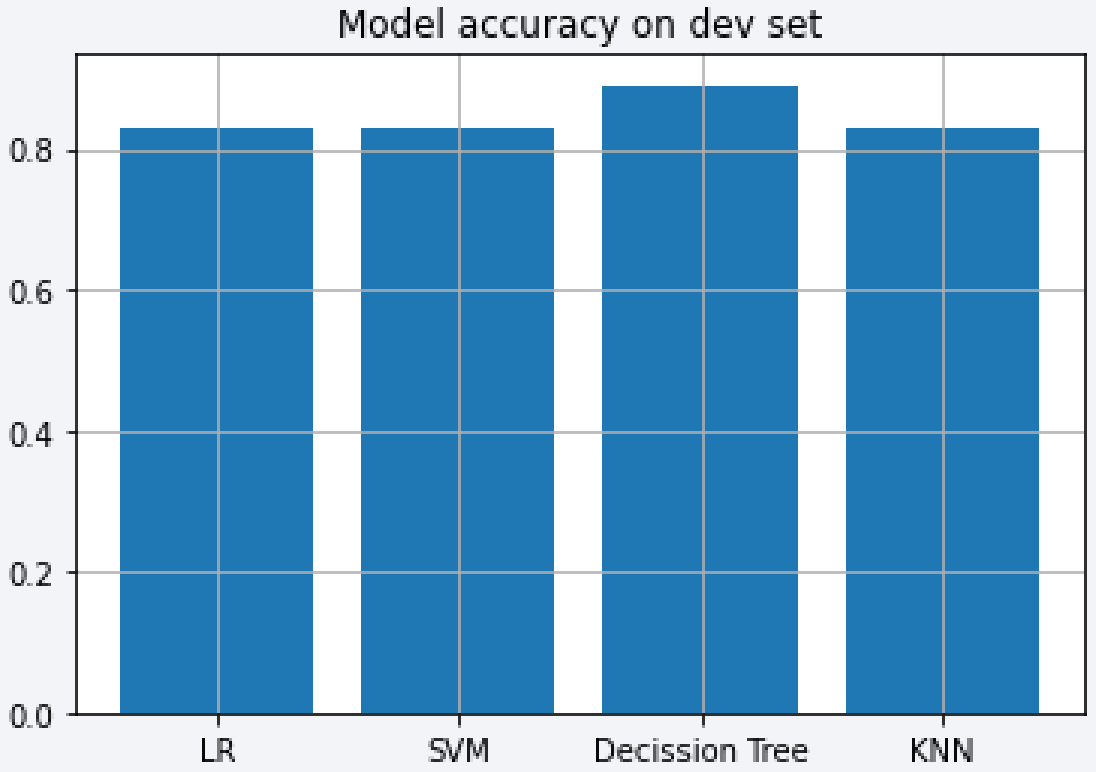


Section 5

Predictive Analysis (Classification)

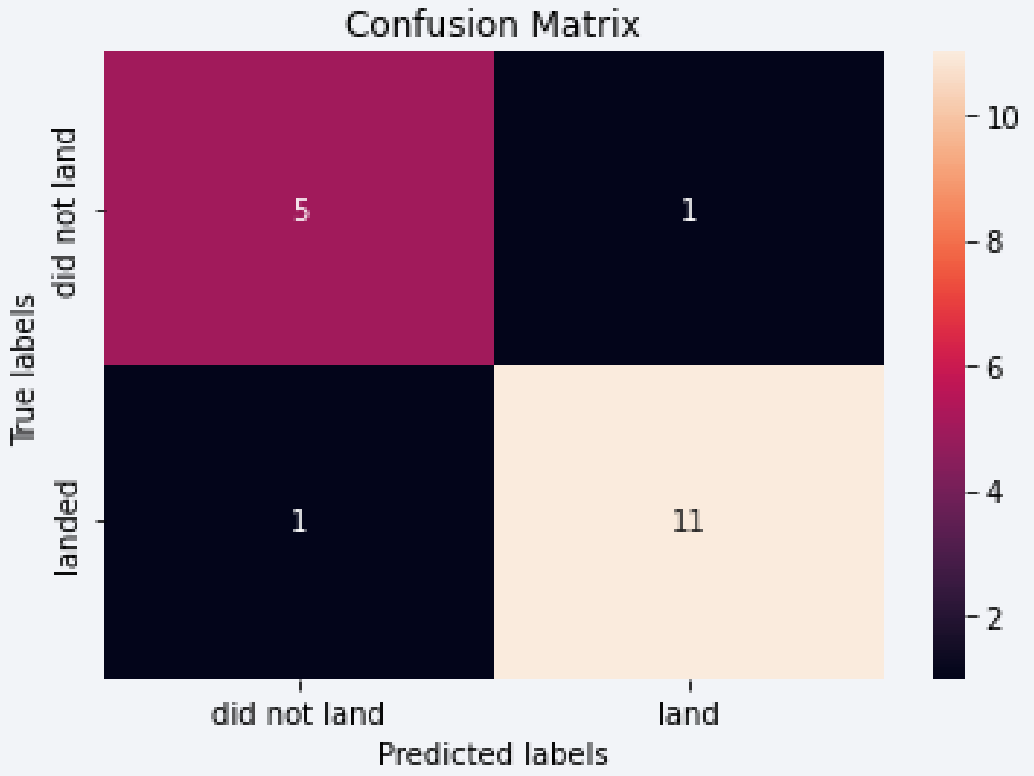
Classification Accuracy

- Different models accuracy was found.
- The model that performed best on the dev set was the Decission Tree Classifier with an accuracy of 0.88.
- On the test set, the Decission Tree's accuracy was 0.89.



Confusion Matrix

- The Decission Tree Classifier classified correctly eleven launches.
- The Decission Tree Classifier had one False Positive and one False Negative.



Conclusions

- The best classifier was the Decission Tree.
- It had an accuracy of 88% on the test set.
- Low weighted payloads perform better than the heavier payloads.
- The success rate improved during the years 2013-2020.

Thank you!