IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Marc Muijzer
03-07-2024

Outline

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

Executive Summary

Using open source data the success rate of the Falcon 9 first-stage landing was collected via web scraping and the SpaceX API.

The data was analyzed and presented in a dashboard. The exploratory data analysis identified the best predictors to predict a successful landing. A predictive model was developed to predict if the first-stage Falcon 9 would land. Various models were evaluated with the best performing model having an accuracy of 83%.

Introduction

SpaceX Falcon 9 is a rocket that can land the first stage for re-use, significantly reducing costs of a launch.

Not every time the rocket lands and in an unsuccessful landing the first-stage is lost. Thereby increasing the costs of spaceflight.

A new company SpaceY want compete with SpaceX.

The objective of this project is to investigate the success of landing the first stage of the Falcon 9. To this end the following will be investigated:

- success rate of the Falcon 9 launch / landing
- Identify the best place to make launches
- predicting successful landings of the first stage of rockets

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
  - Data from Space X was obtained from Space X API and wikipedia using webscraping
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- 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 that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection – SpaceX API

SpaceX offers a public API from where data can be obtained and then used;

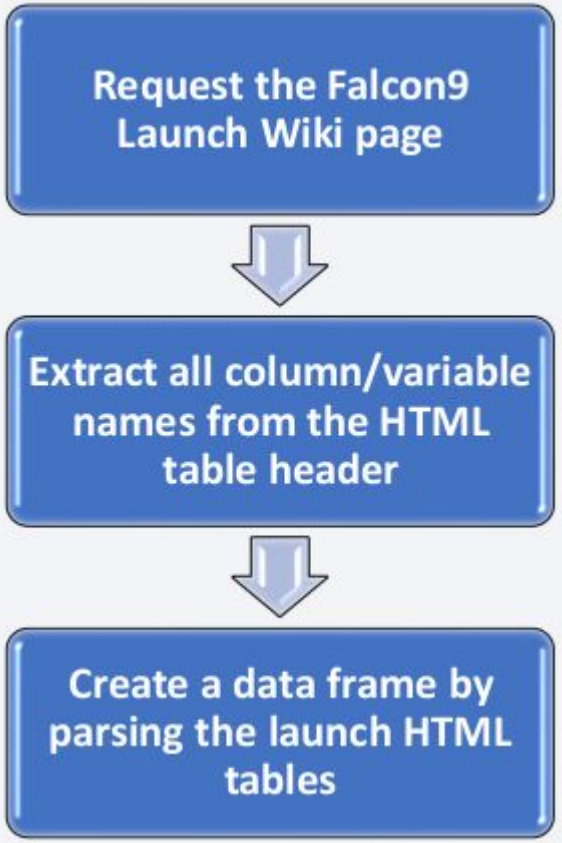
- https://api.spacexdata.com/v4/rockets/

This API was used according to the flowchart on the right hand side.

Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module1/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
  - https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches
- Data are downloaded from Wikipedia according to the flowchart on the right hand side.



Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module1/jupyter-labs-webscraping.ipynb

Data Wrangling

- Initially Exploratory Data Analysis (EDA) was performed on the dataset.
- Various summaries were calculated, including numbers per launch sites, occurrences per orbit type, and mission type.
- Finally, the landing outcome label was created from Outcome column.
- source:

Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module1/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:
  - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass,
  - Orbit and Flight Number, Payload and Orbit

Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module2/jupyter-labs-eda-dataviz.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: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module2/jupyter-labs-eda-sql-coursera_sqllite.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;
  - Lines are used to indicate distances between two coordinates.

Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module3/lab_jupyter_launch_site_location.ipynb

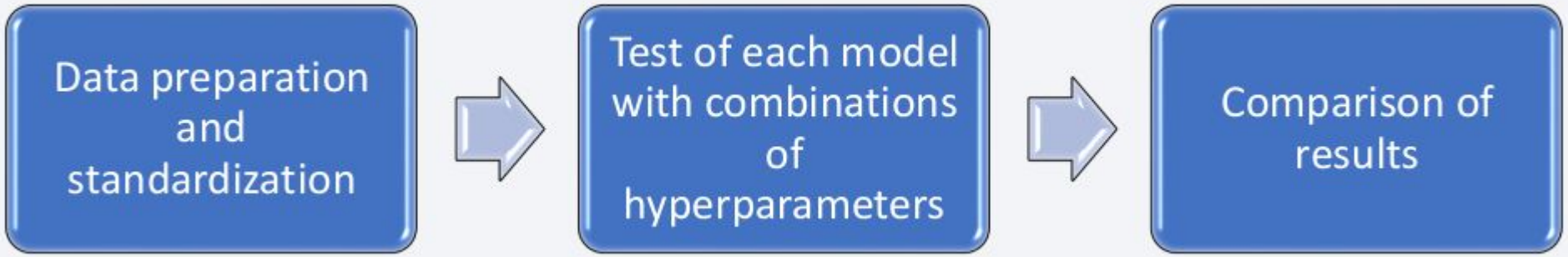
Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- The results were combined to analyze the relation between payload and launch sites, helping to identify the best launch site for each payload.

Source:https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module3/spacex_dash_app.py

Predictive Analysis (Classification)

- Data was divided in a training and test set. The data was normalize to enable development of predictive models.
- the following classification models were investigated: logistic regression, support vector machine, decision tree and k nearest neighbors.
- The accuracy was calculated and confusion matrices created.



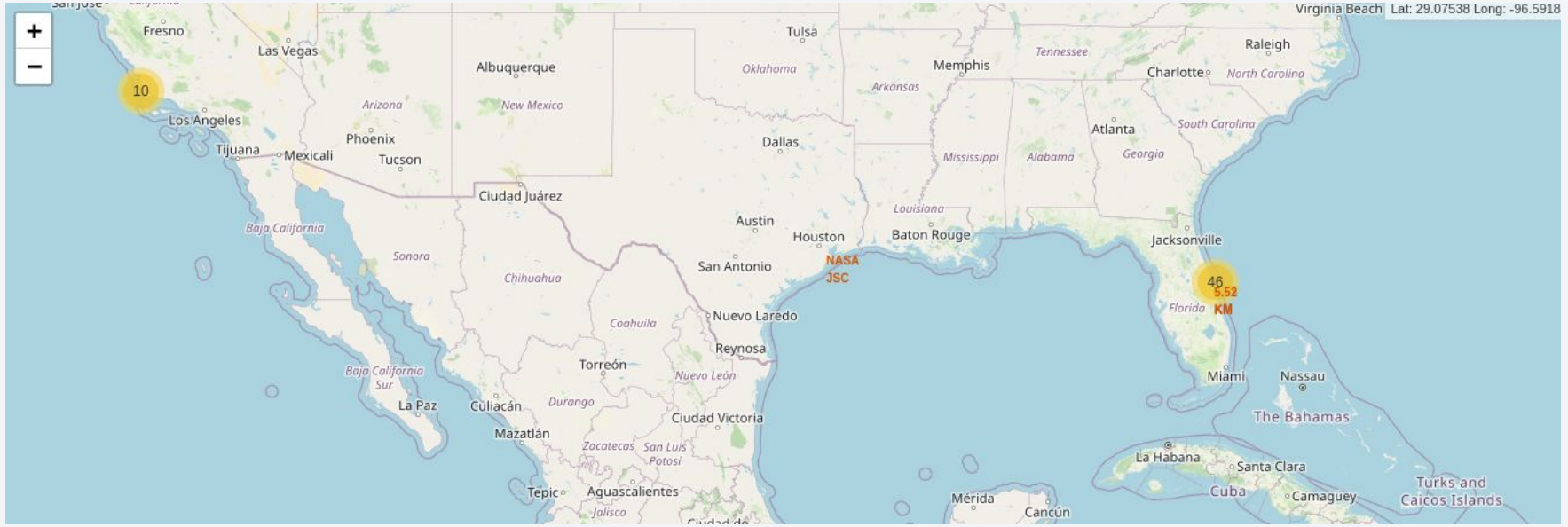
Source: https://github.com/snipemuis/IBM-data-science-course/blob/main/capstone_project/module4/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results - EDA

- 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.

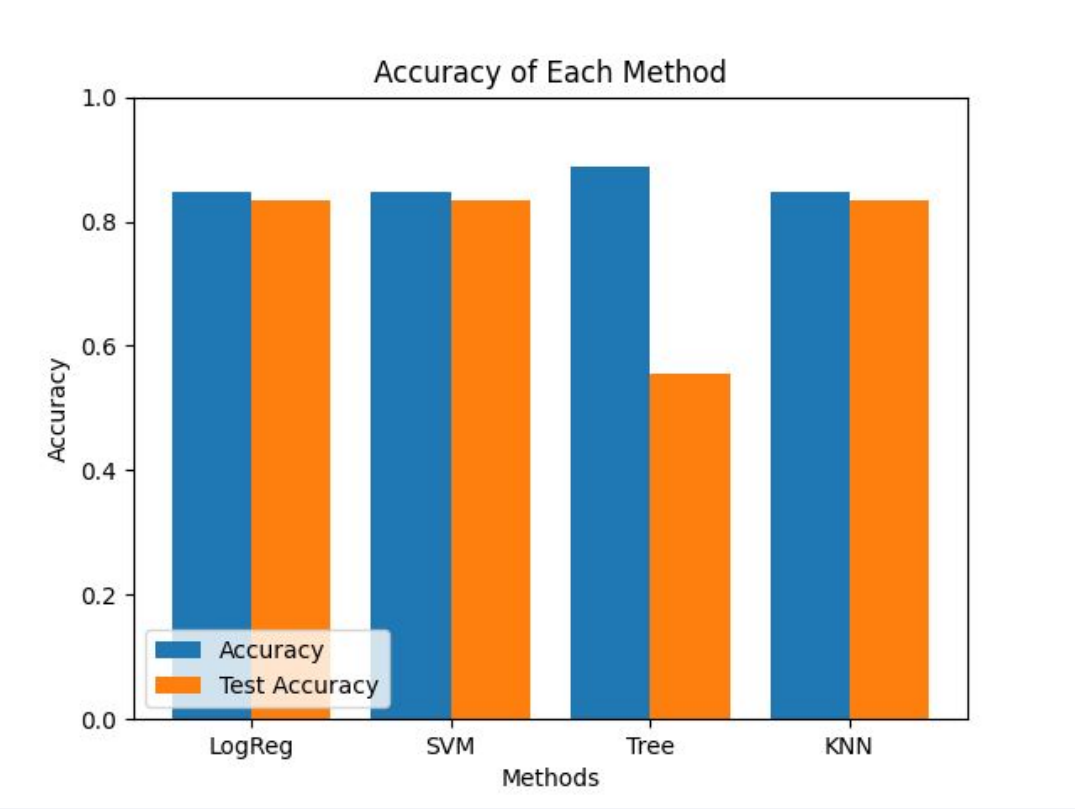
Results - EDA

- 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 happens at east coast launch sites.



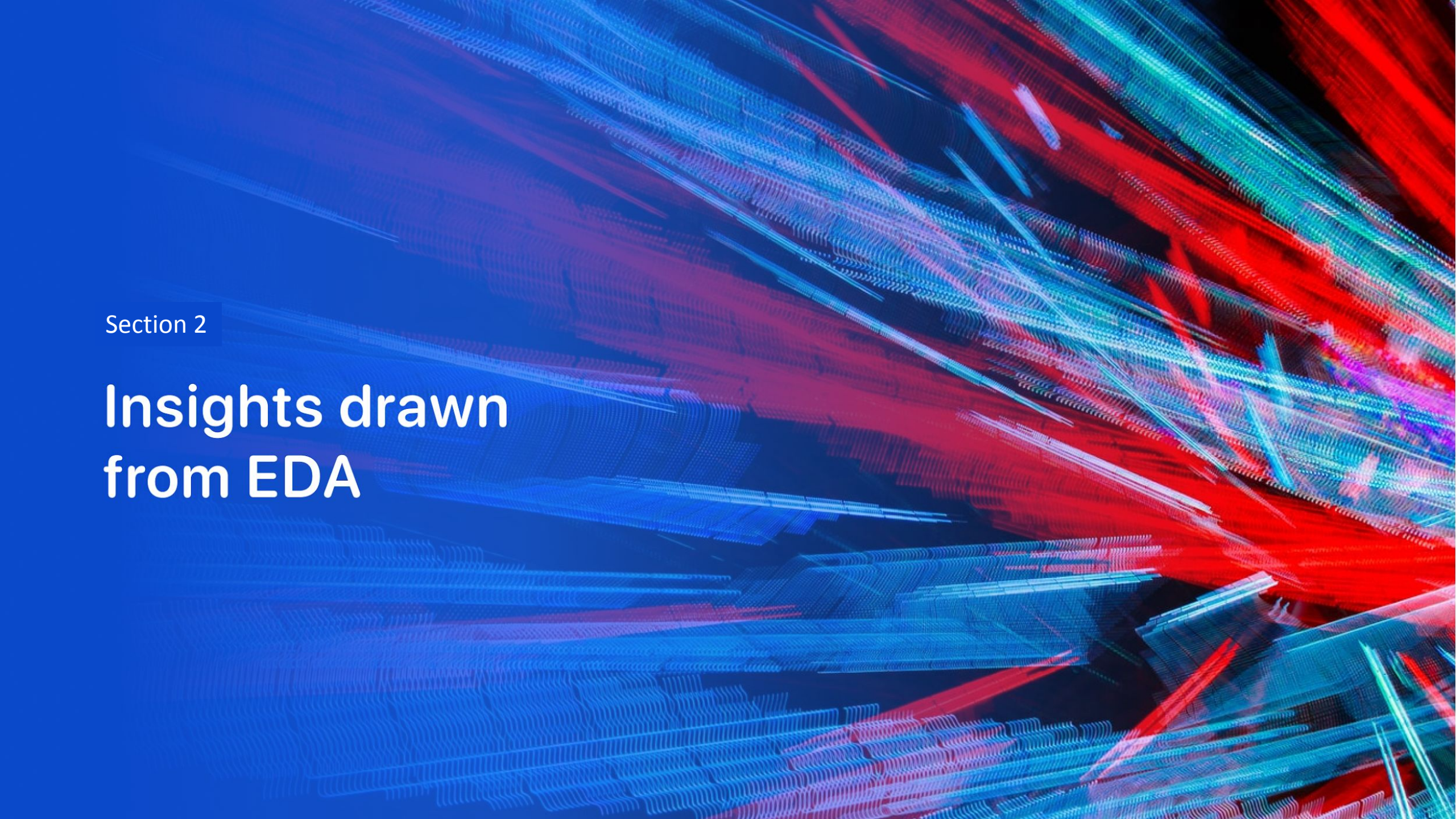
Results - classification models

- Predictive Analysis showed that SVM, logistic regression, and KNN are the best models to predict successful landings, performing similarly and having a training set and test set accuracy over 80%



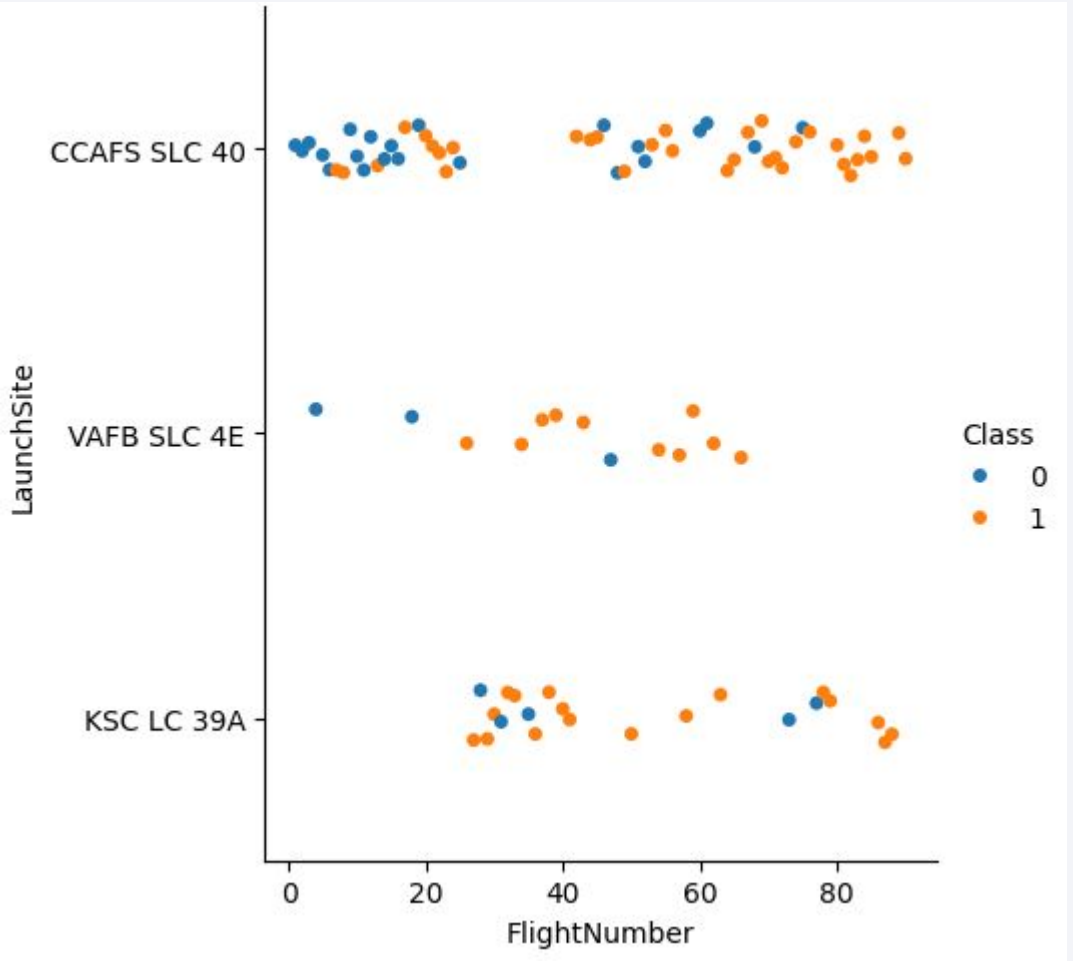
Section 2

Insights drawn from EDA



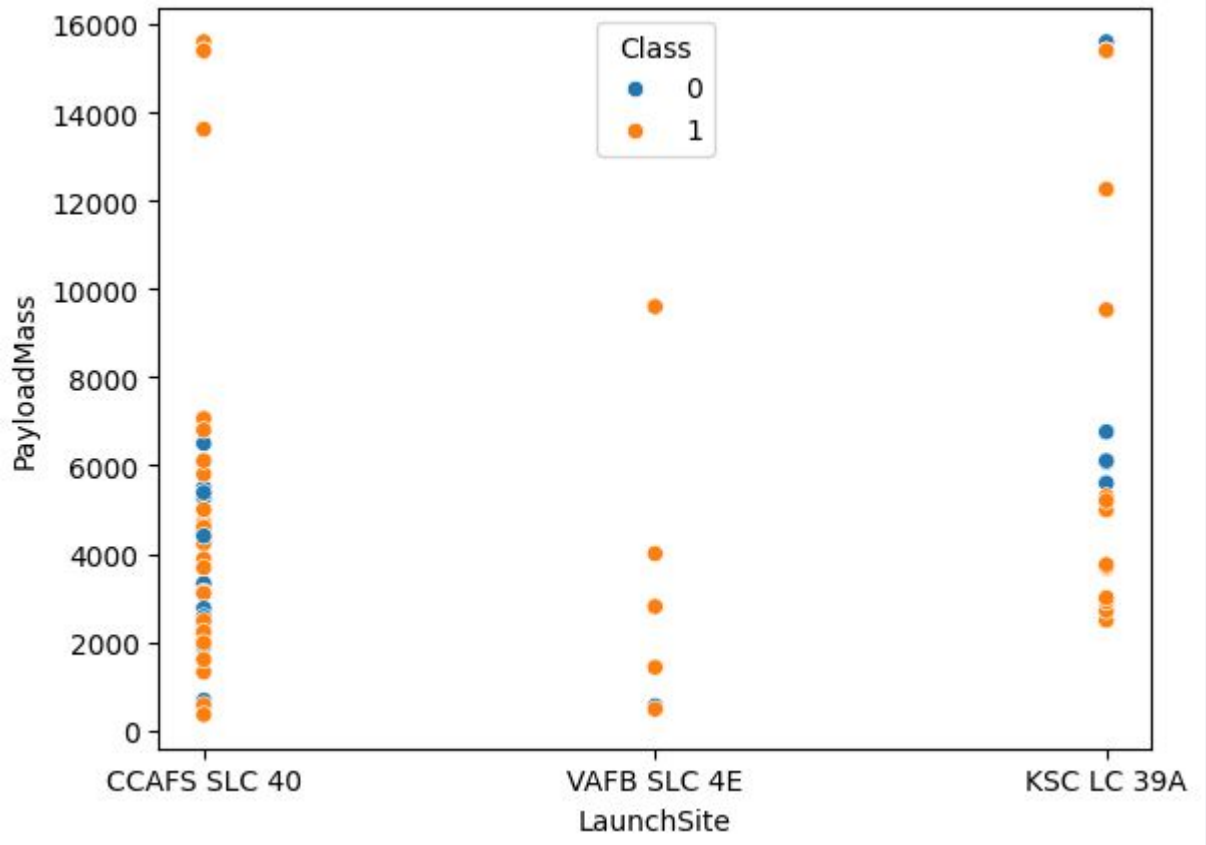
Flight Number vs. Launch Site

- Most launches have been performed from CCAFS SLC40 and next KSC LC 39A. Both launch sites have a high success rate.
- The success rate improved over time



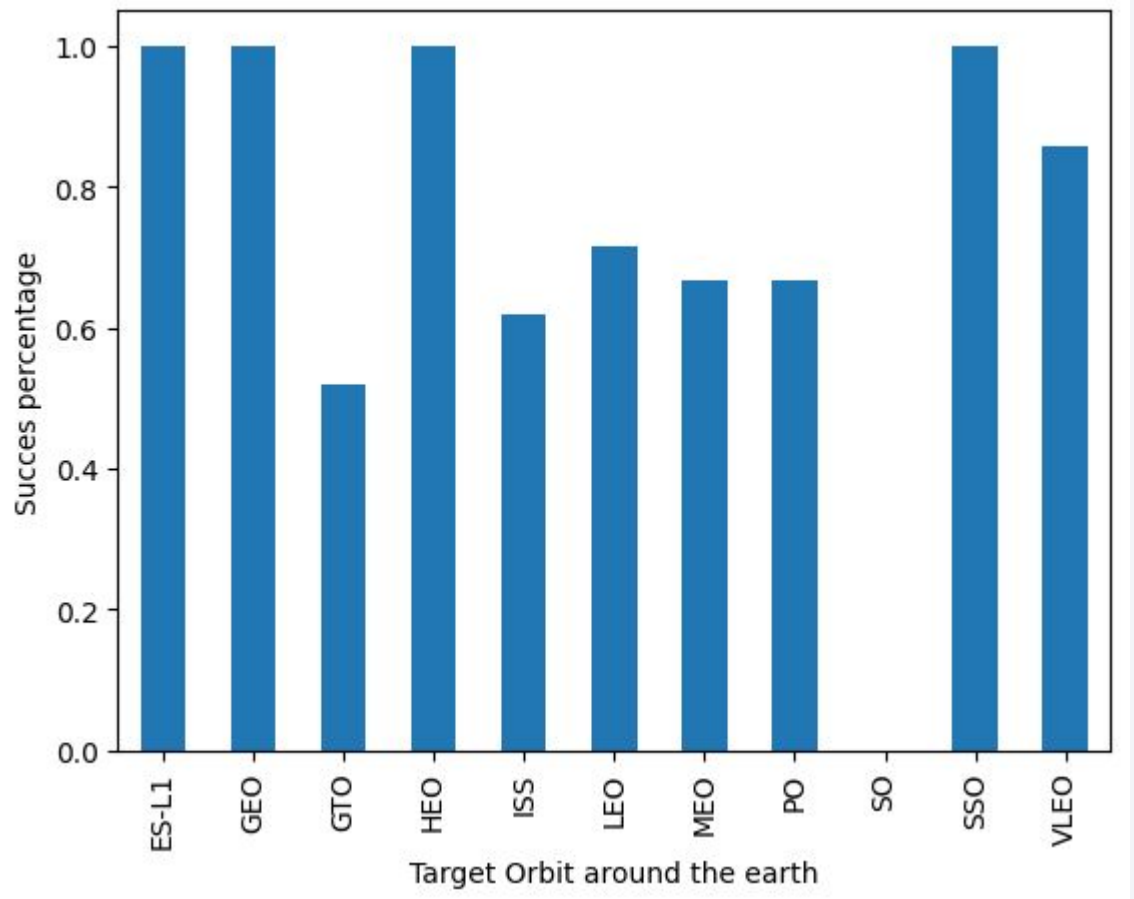
Payload vs. Launch Site

- CCAFS SLC 40 and VAFB SLC 4E are generally used for rockets with lighter payloads.
- No clear relationship between payload, launch site and success rate can be observed.



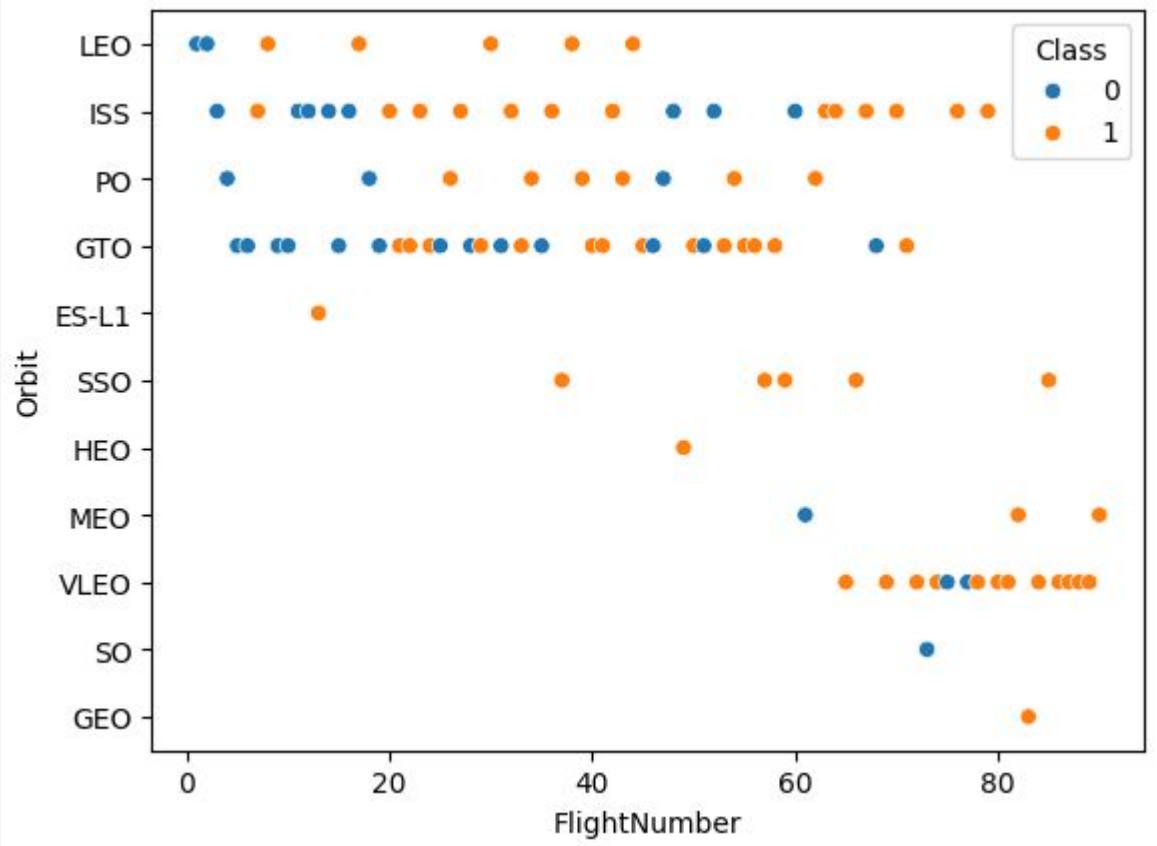
Success Rate vs. Orbit Type

- ESL1, GEO, HEO, SSO, VLEO have the highest success percentages.
- GTO the lowest.



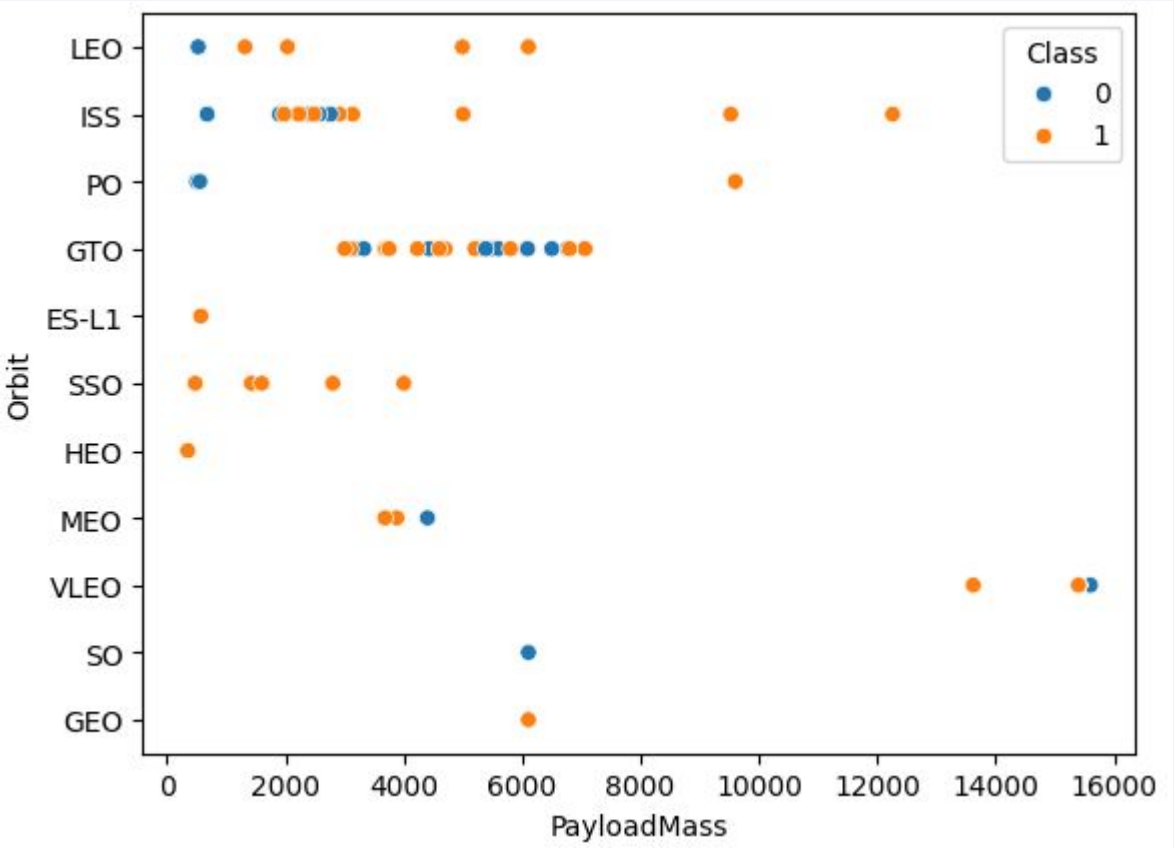
Flight Number vs. Orbit Type

- Early launches aimed mostly at lower earth orbits.
- Later launches aim at higher earth orbits.
- For all types of orbits success rate improved over time.



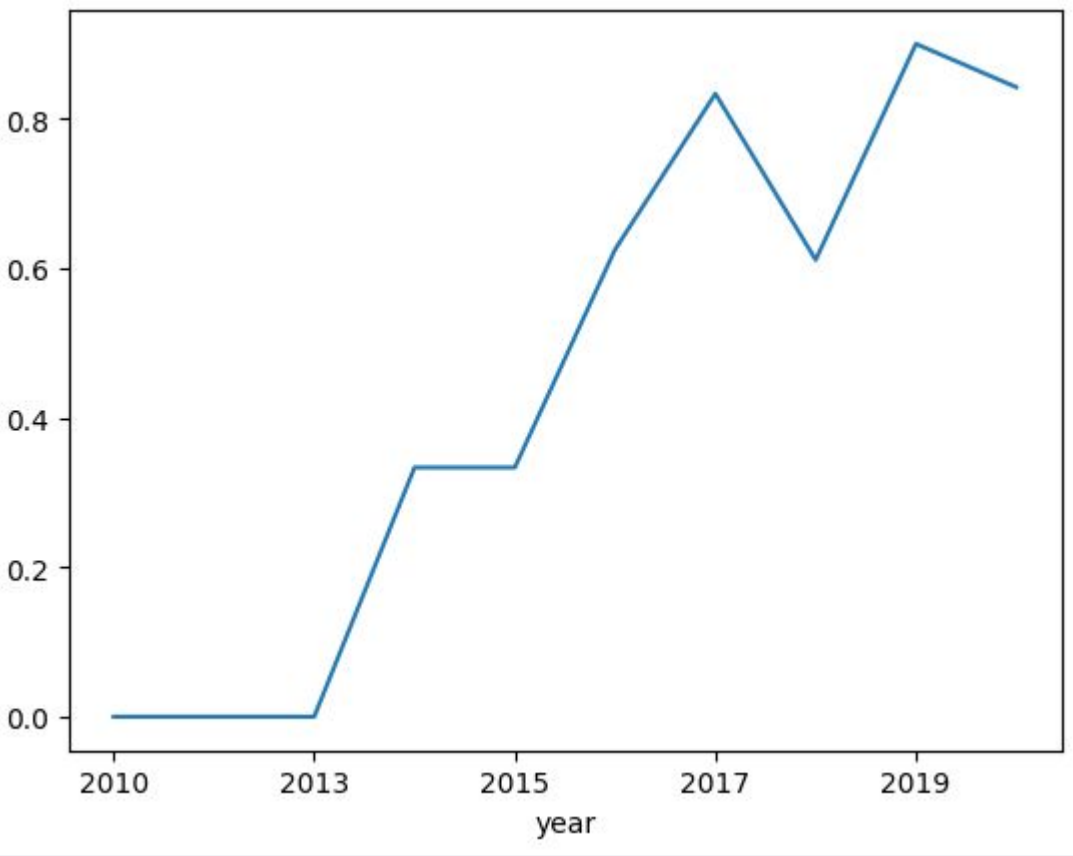
Payload vs. Orbit Type

- There is no relation between payload and success rate to orbit GTO;
  - ISS orbit has the widest range of payload and a good rate of success;
  - There are few launches to the orbits SO and GEO.



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;



All Launch Site Names

- This are the 4 launch sites

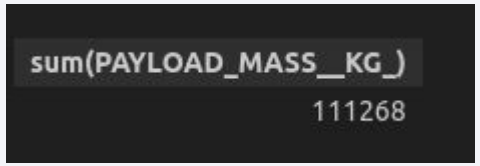
Launch Site Names Begin with 'CCA'

- These are the 5 records beginning 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

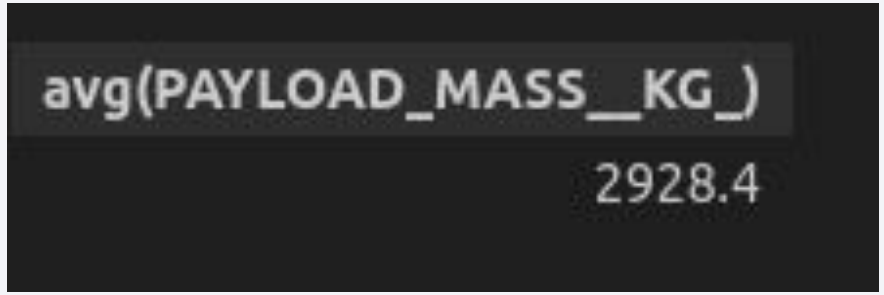
Total Payload Mass

- The total payload mass is 111.268 kg.



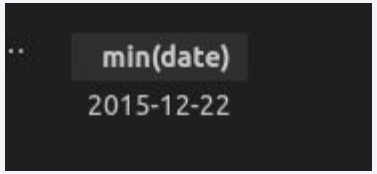
Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg



First Successful Ground Landing Date

- The first successful landing outcome on ground pad was on 22 december 2015.



Successful Drone Ship Landing with Payload between 4000 and 6000

- Below is a list of 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 B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- 99 successful launches, 1 successful launch with payload status unclear and 1 failure.

Mission_Outcome	count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- These are the boosters carried the maximum payload.

Booster_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
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- A list the booster versions, and launch site names for in year 2015 for failed launches from a drone ship

-

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

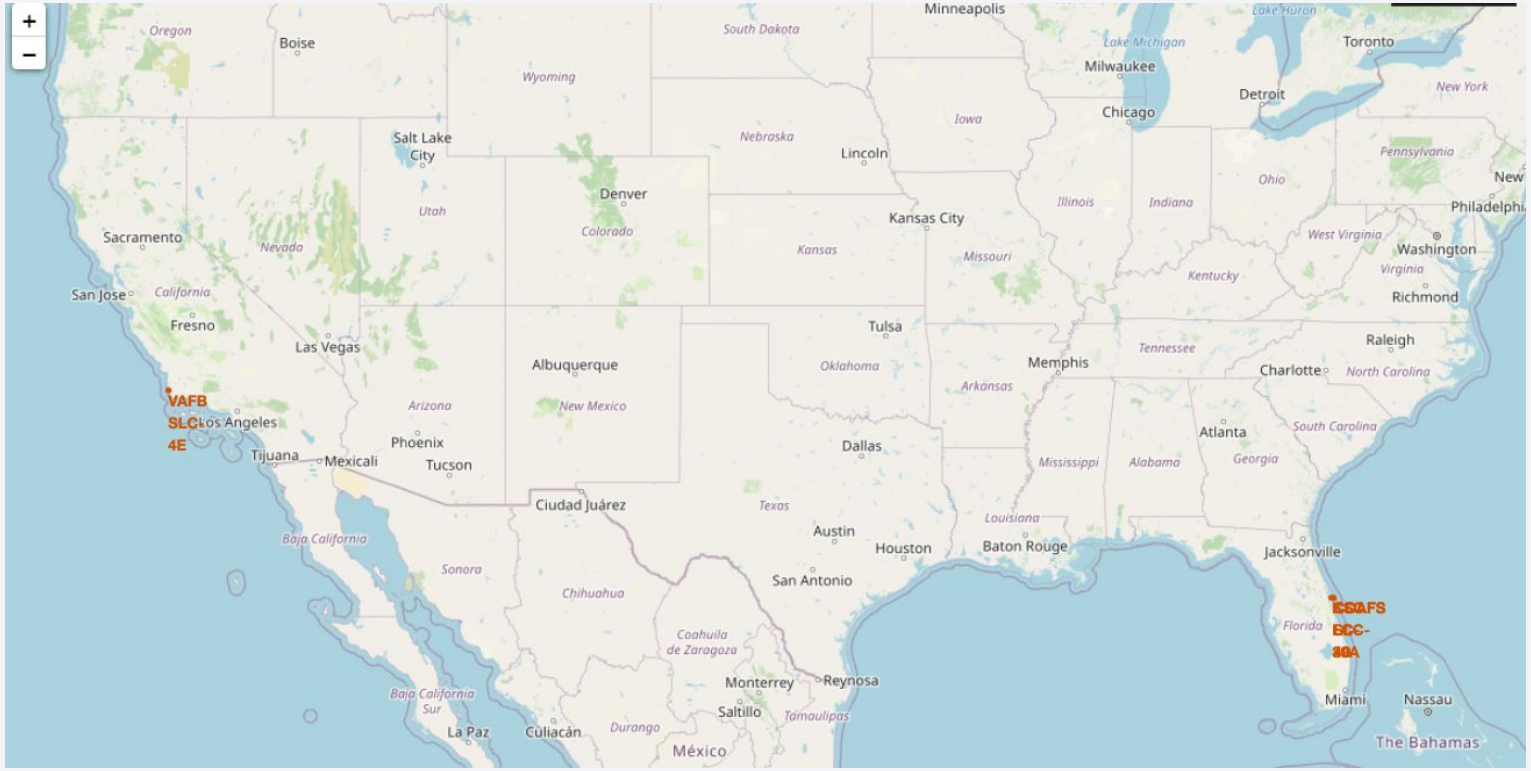
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- A count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

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


Section 3
Launch Sites
Proximities Analysis

All launch sites

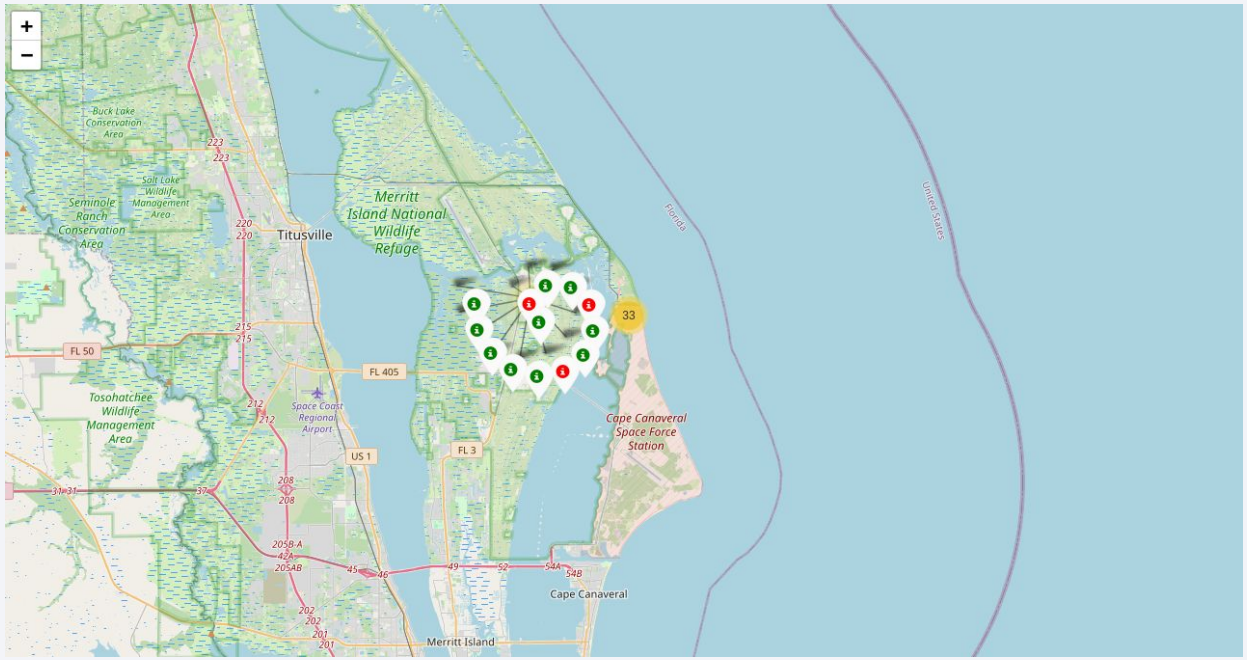


- The 3 launch sites used. the important elements and findings on the screenshot

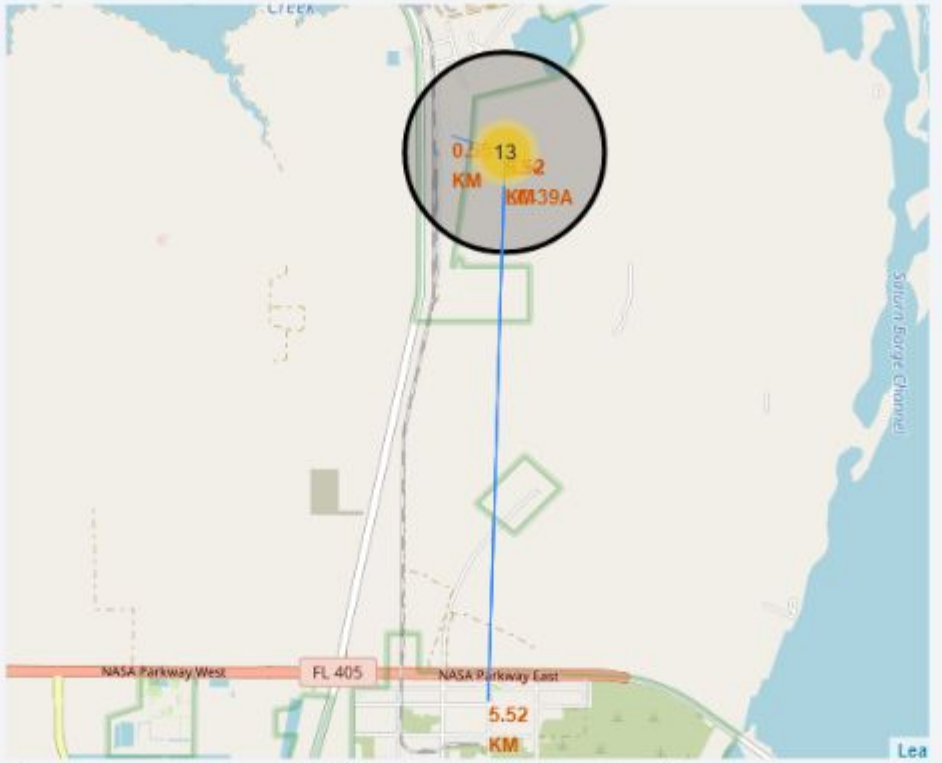
Launch site outcomes

Green markers indicate successful launches,

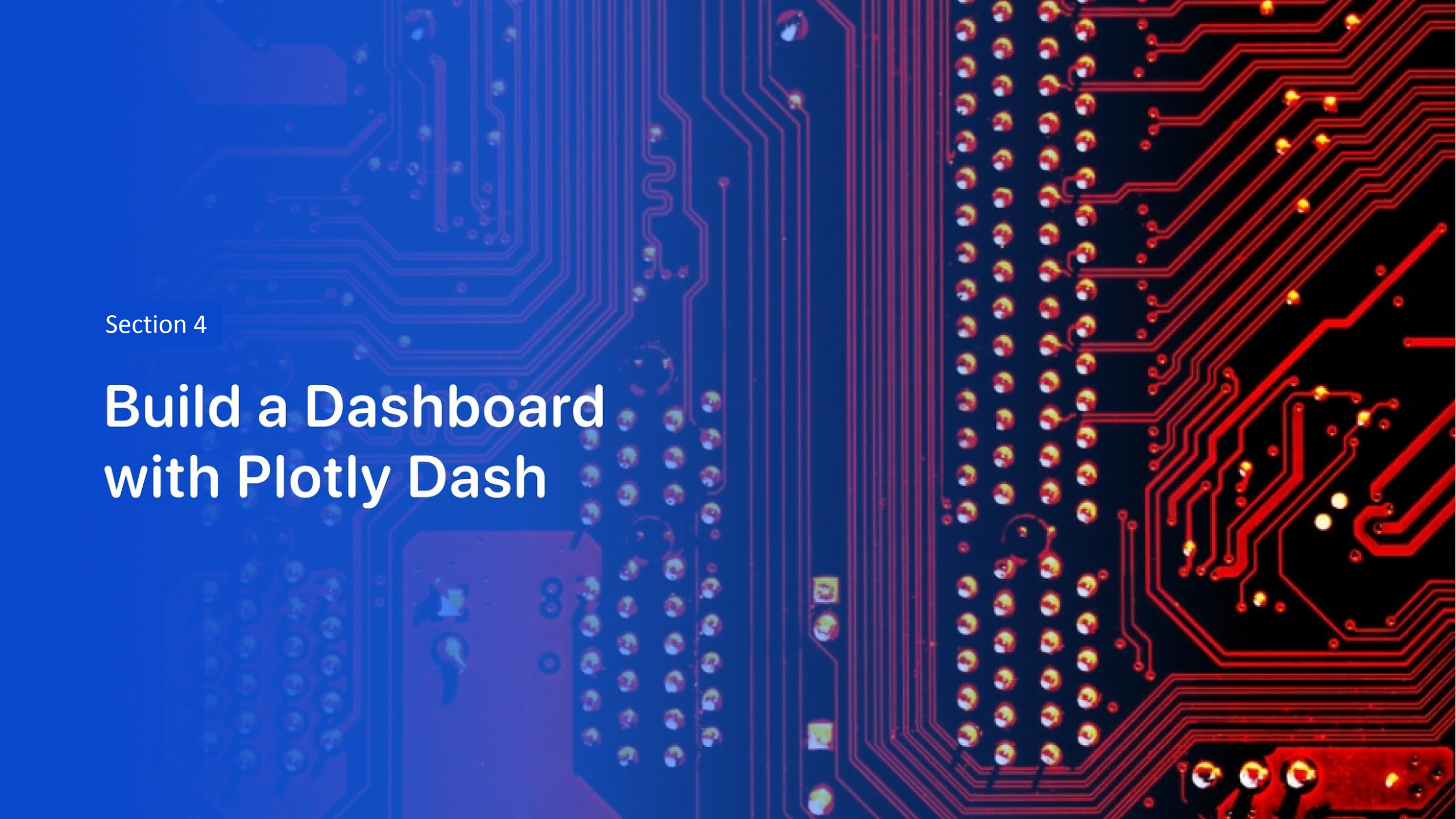
red markers indicate failed launches



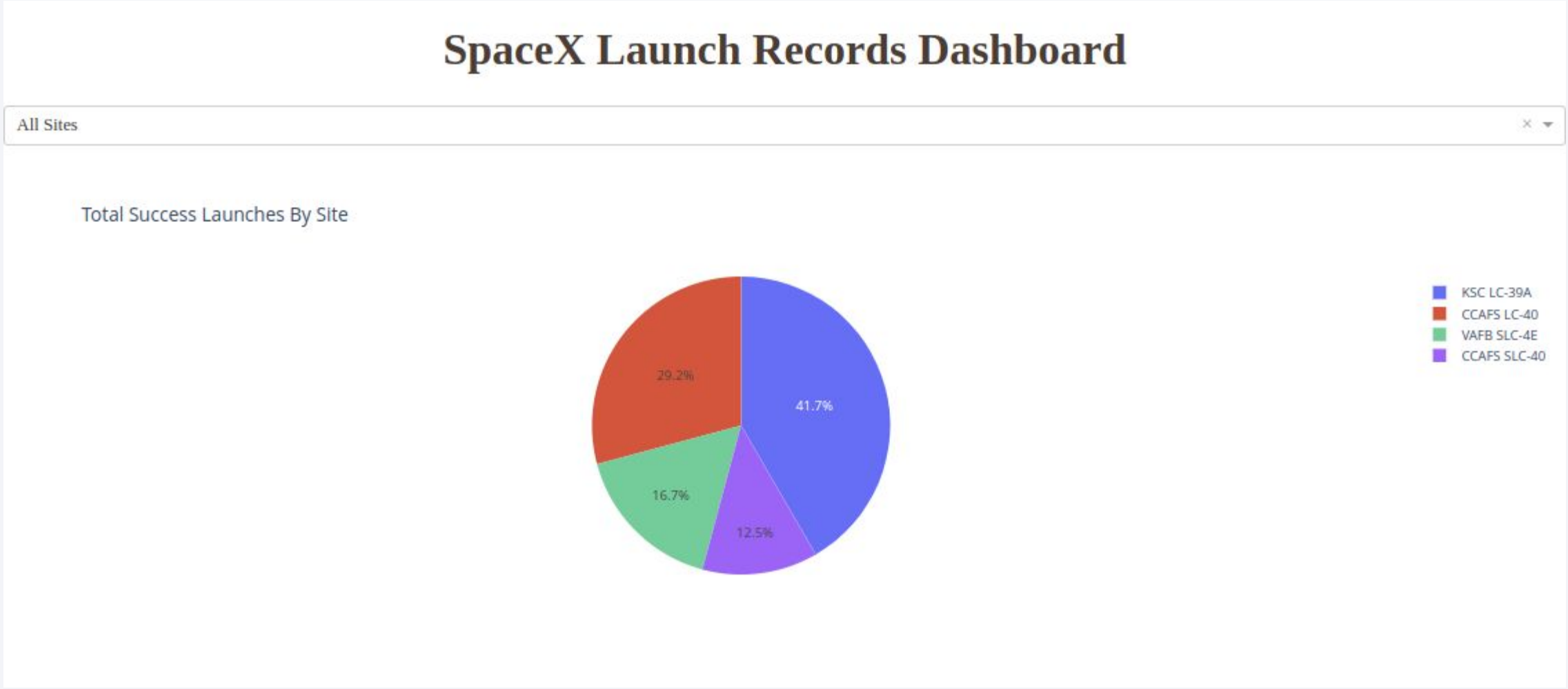
Connections



- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

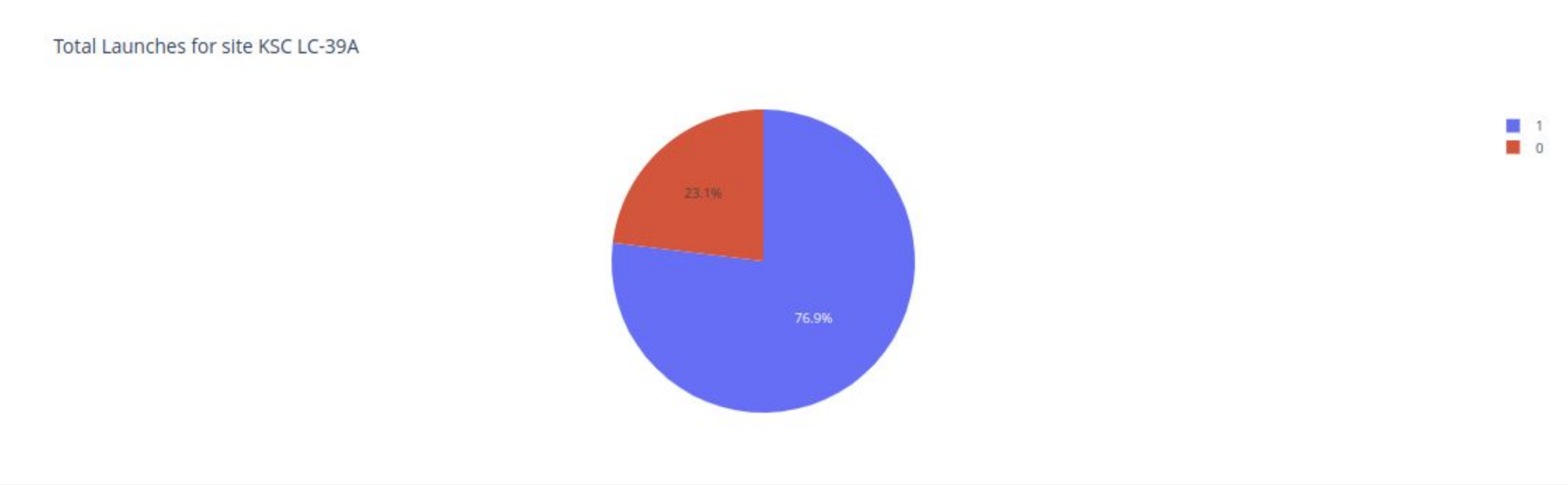

Section 4
Build a Dashboard with Plotly Dash

Launch site and success rate



- Successful launches per launch site

Launch site KSC LC-39A



- The launch site with the highest success percentage

Payload versus booster version

Payloads under 6,000kg and FT boosters are the most successful combination

Payload versus booster version

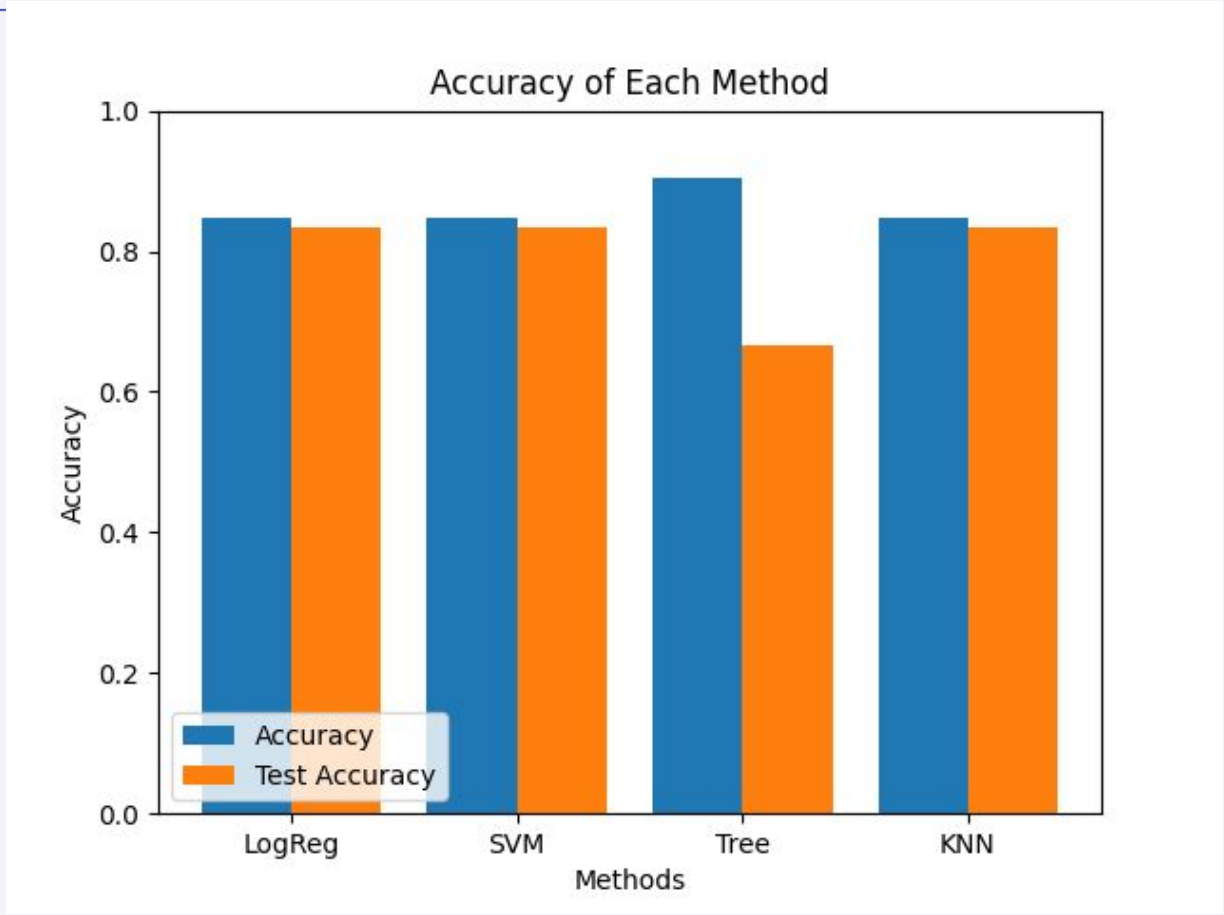
There’s not enough data to estimate risk of launches over 7,000kg

Section 5

Predictive Analysis (Classification)

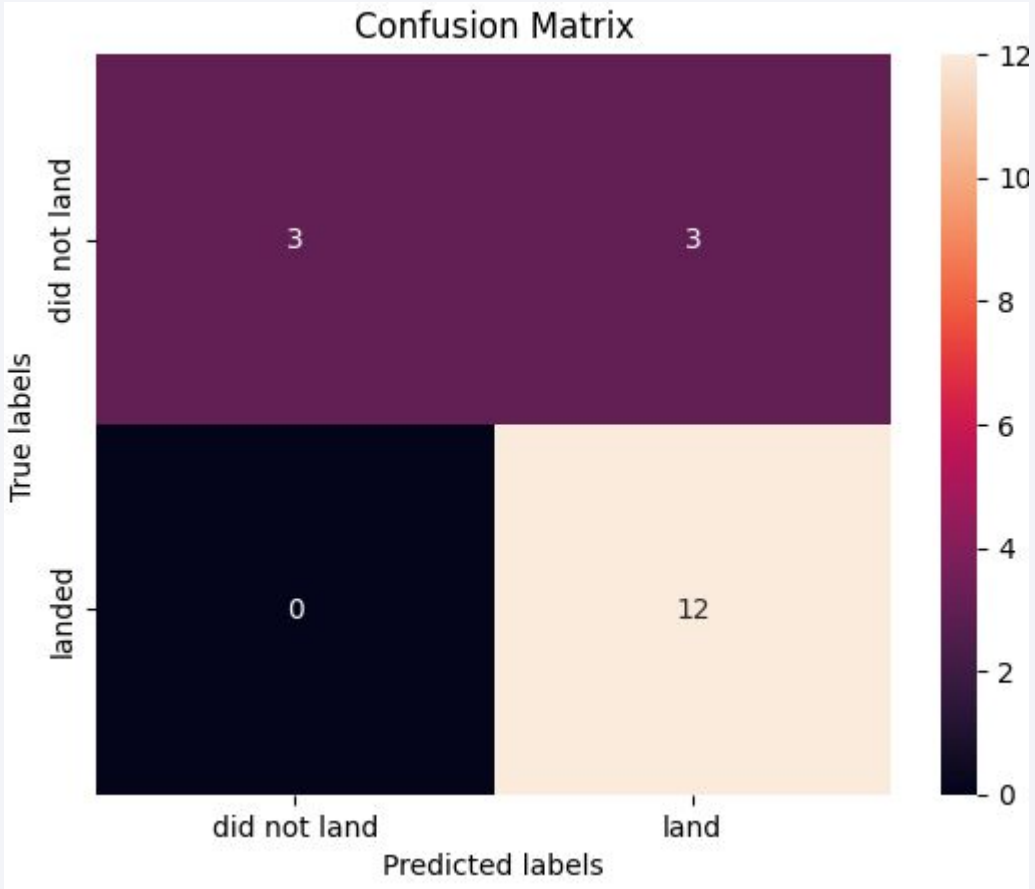
Classification Accuracy

- Logistic regression, SVM, and KNN perform similarly. All achieve a training and test accuracy of ~80%



Confusion Matrix

- The three best performing models all show the same confusion matrix, with 3 incorrect predictions. The model predicted the missile will land, whereas in practice it did not land.



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Although most of mission outcomes are successful,
- successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Logistic regression, svm, and KNN can be used to predict successful landings

Thank you!