IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Byron Vasquez
12.04.2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory Data analysis with SQL
  - Exploratory Data analysis with Data visualization
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive analytics demo
  - Predictive analysis results

Introduction

- The objective of this project is to evaluate the ability of a new company, Space Y, to complete with Space X in selling affordable space travel for everyone.

- Expected result of this project:

  - Find the most effective method of estimating the total launch cost by predicting the success of first-stage rocket landings.

  - Identify where is the best place to make launches

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- The data for Space X was collected from these two sources:

https://api.spacexdata.com/v4/launches/past

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Perform data wrangling

- The data was processed by creating a landing outcome label based on outcome data after summarizing and analyzing features

Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

- The data were normalized, divided into training and test data sets and evaluated by four different classification models and the accuracy of each model was evaluated using different combinations of parameters.

Data Collection

Data sets were collected using an API and web scraping technics.

The Space X data was collected using the API

https://api.spacexdata.com/v4/launches/past

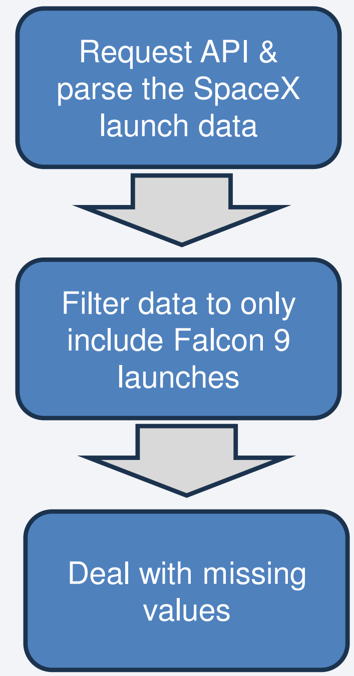
And from Wikipedia using web scraping

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

- SpaceX offers a public API where the data can be collected

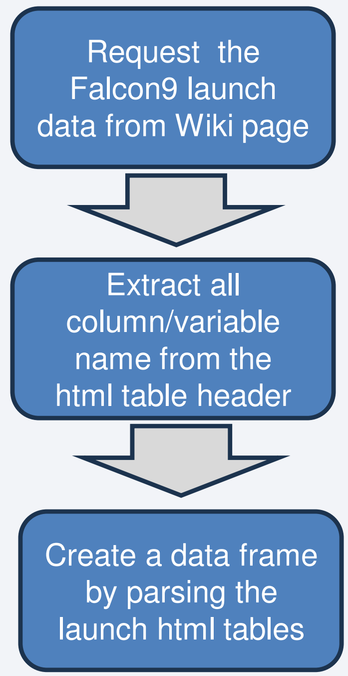
- Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- SpaceX data can also be extracted from Wikipedia using web scraping.

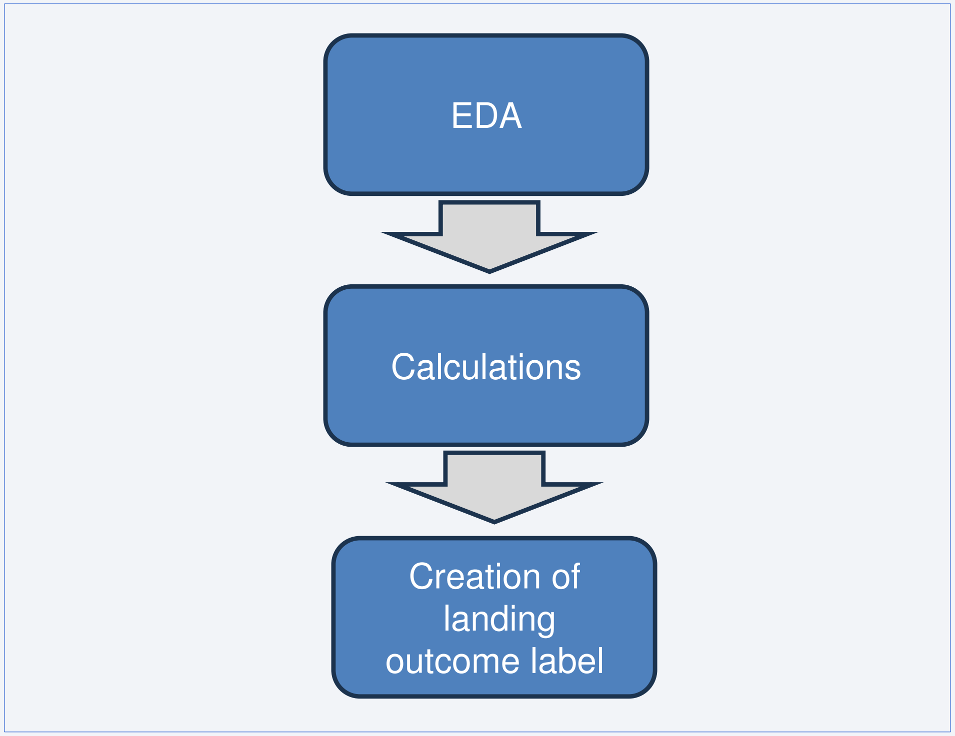
- Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

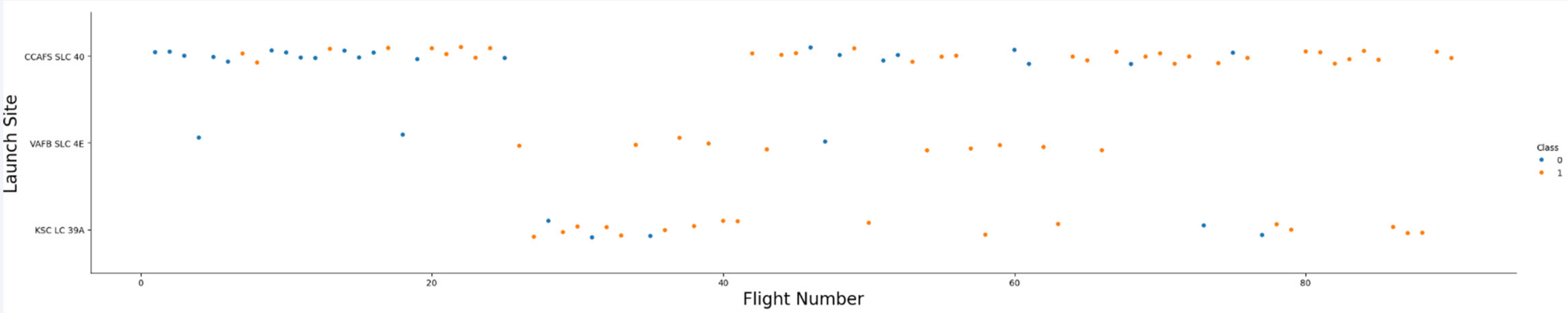
- Data wrangling process began by doing some Exploratory Data Analysis (EDA) on the data.
- Some calculations on the number of launches by site, orbits and mission outcomes were done.
- The landing outcome label was created from the outcome column in the data set.

Source code:
https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

Scatterplots and bar charts were used to visualize the relationship between features, e.g., Flight number and Launch site.



Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/edadataviz.ipynb

EDA with SQL

The following SQL queries were performed:

- Display the names of the unique launch sites in the space mission
- 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 when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.
- 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.

Source coded: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

The maps include markers, lines and marker clusters

- Markers indicate launch sites.
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center.
- Marker clusters indicate groups of events in each coordinate, e.g., launches in a launch site.
- Line indicate the distance between two coordinates

Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/lab_jupyter_launch_site_location.ipynb

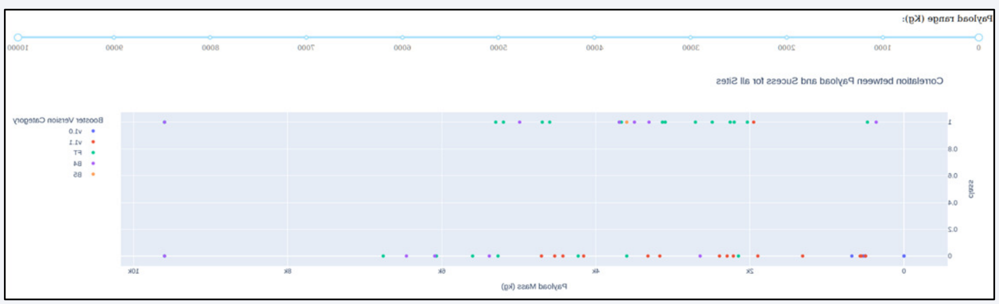
Build a Dashboard with Plotly Dash

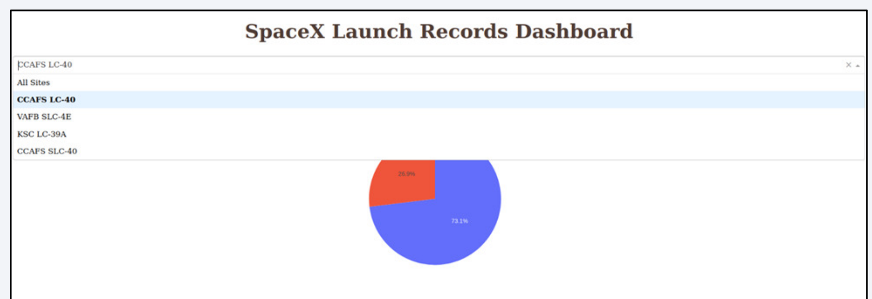
The dashboard contains a scatterplot and a pie chart

The scatterplot shows the relation between Payload and the Booster version class, it also has a slider which makes it easy to change the range of the payload.

The pie chart shows the total successful launches by site, and it has an interactive drop-down list to filter the total by specific site.

These features makes it easier to visualize the data based on specific payload ranges or a specific site.

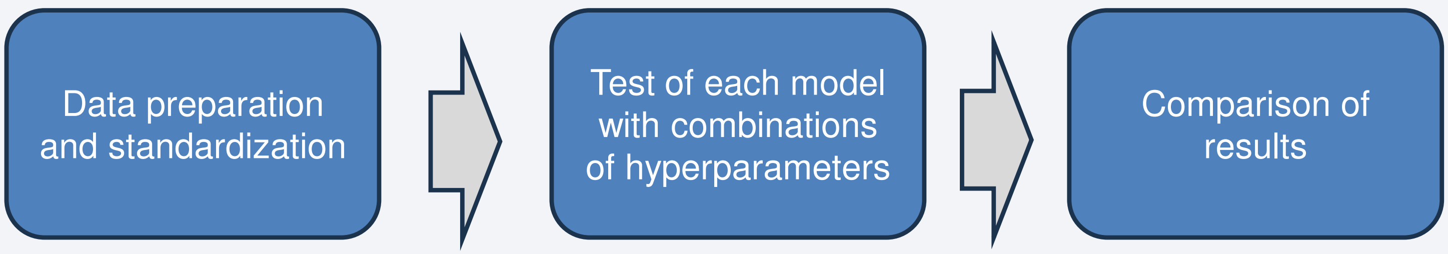




Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

Four classification models were compared: logistic regression, support vector machine (VSM), decision tree and k nearest neighbors (KNN).



Source code: https://github.com/bevg1971popos/AppliedDSCapstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

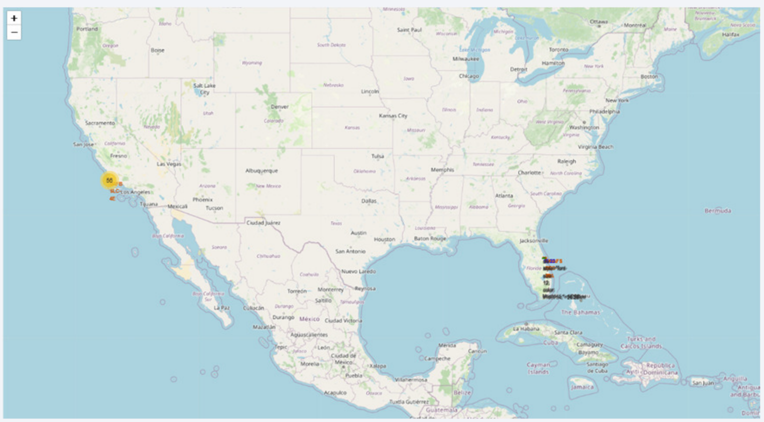
Results

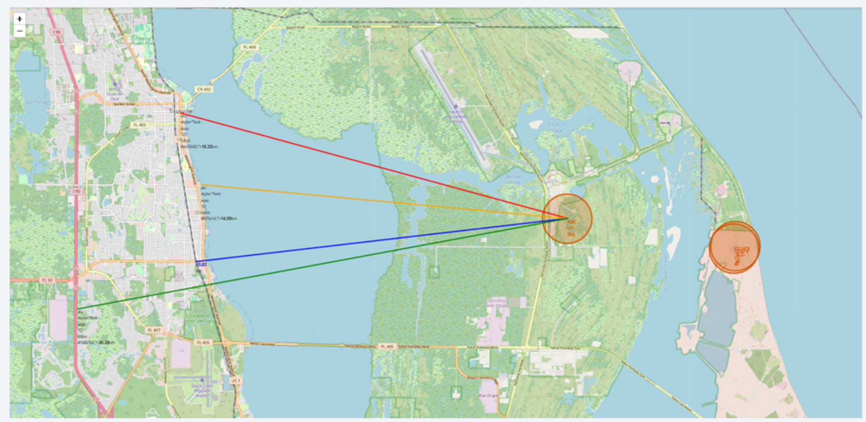
Exploratory data analysis results:

- Space X uses four different launch sites.
- The average payload of F9 v1.1 booster is 2,928 kg.
- The first success landing outcome happened in 2015.
- Falcon 9 booster versions had a higher successful rate at landing in drone ships having payload above the average.
- Two booster version failed at landing in drone ships in 2015:
  - F9 v1.1 B1012
  - F9 v1.1 B1015
- The first four years had a low success landing outcome, since 2015 the success of landing outcomes has improved as the years go by.

Results

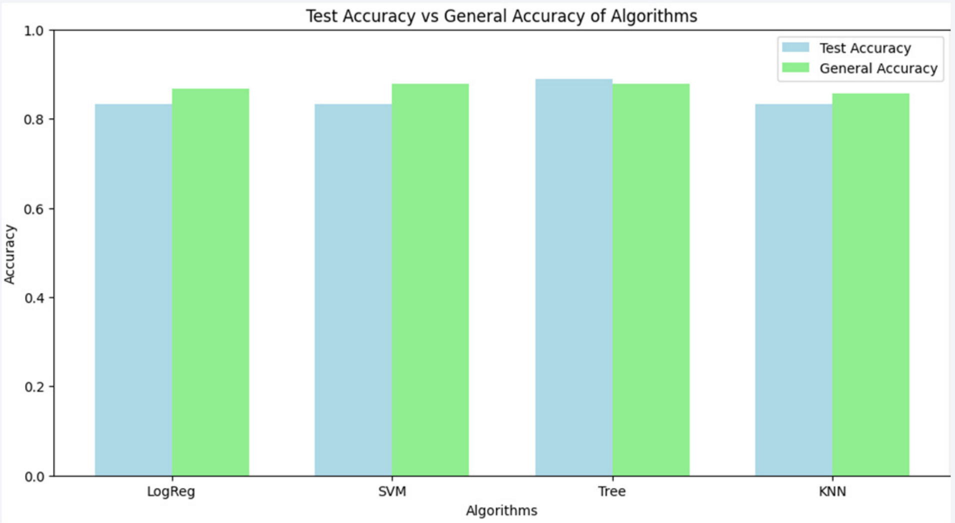
- Interactive analytics were useful to identify what launch sites to use due to their proximity to the sea, safer for the population around those areas and good logistics infrastructure.
- Most launches take place on the east cost launch sites.





Results

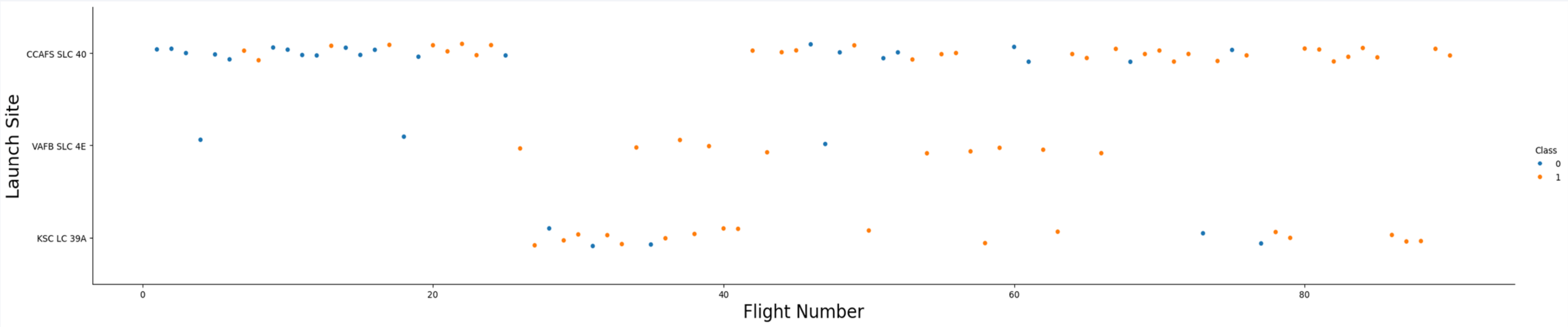
Predictive analysis results show that the Decision Tree classifier is the best model to predict successful landings, having an accuracy of 87.7% and an accuracy of 88.88% on the test data.



Section 2

Insights drawn from EDA

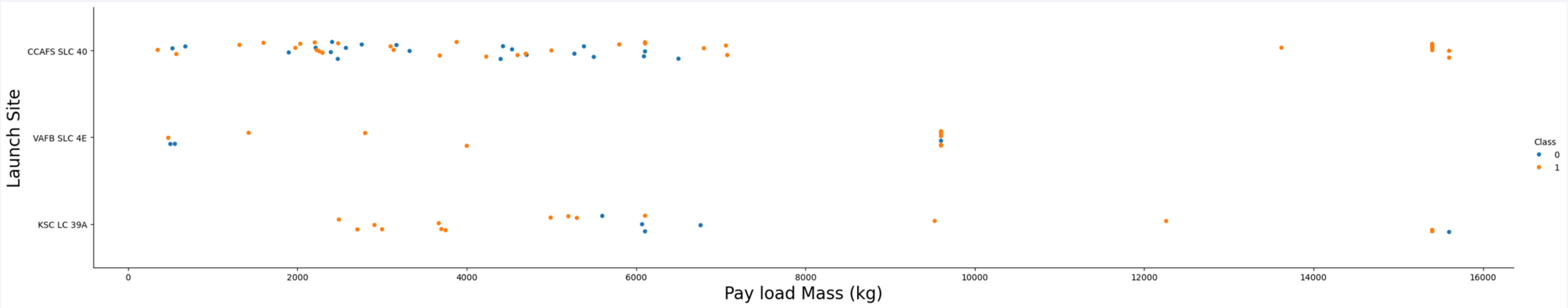
Flight Number vs. Launch Site



The scatter plot shows that:

- The CCAFS SLC 40 launch site has about a half of all the launches.
- The launch sites with higher success rate are VAFB SLC and KSC LC 39A.
- Earlier flight numbers had more failure while latest flight numbers has more success.

Payload vs. Launch Site



The scatter plot shows that:

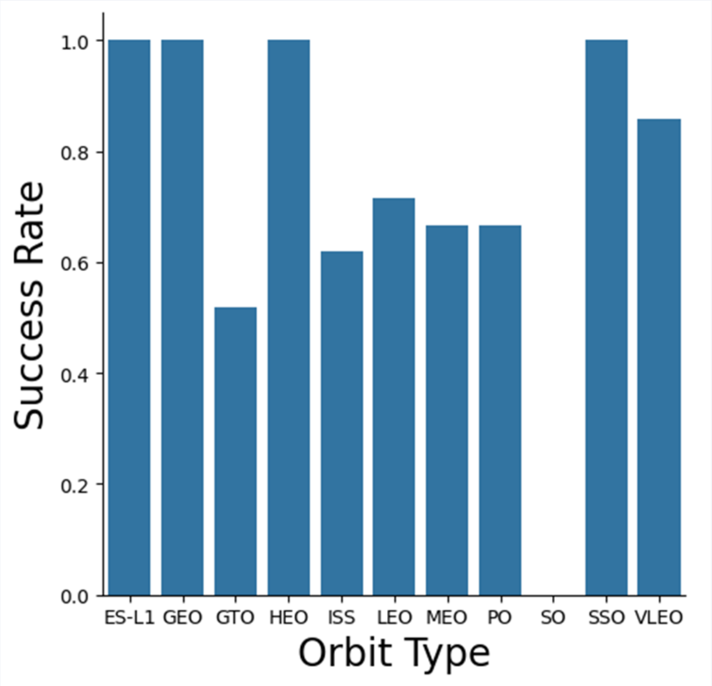
- Most of the launches with payload mass over 7000 kg. were successful.
- Launches with higher payload had a higher success rate.
- KSC LC 39A site has higher success rate for payload mass under 5500 kg.

Success Rate vs. Orbit Type

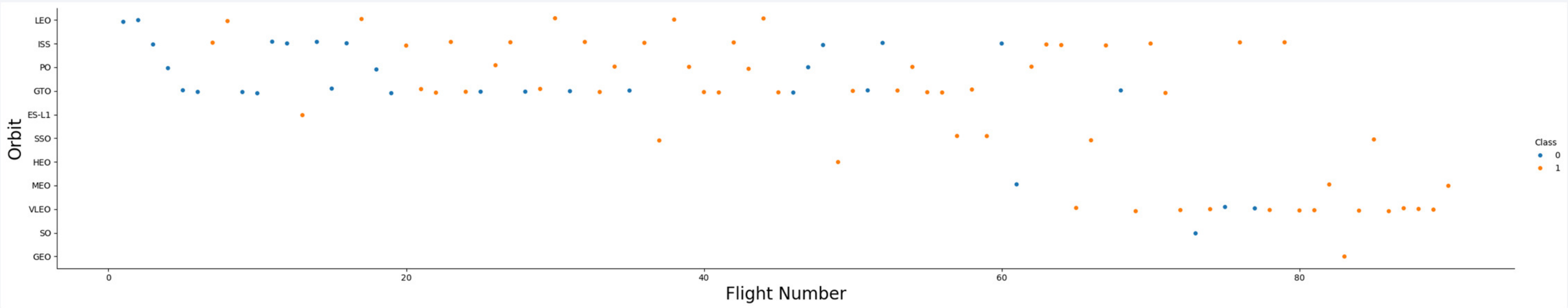
The bar chart shows the success rate of each orbit type

Orbits with the highest success rates are:

- ES-L1
- GEO
- HEO
- SSO



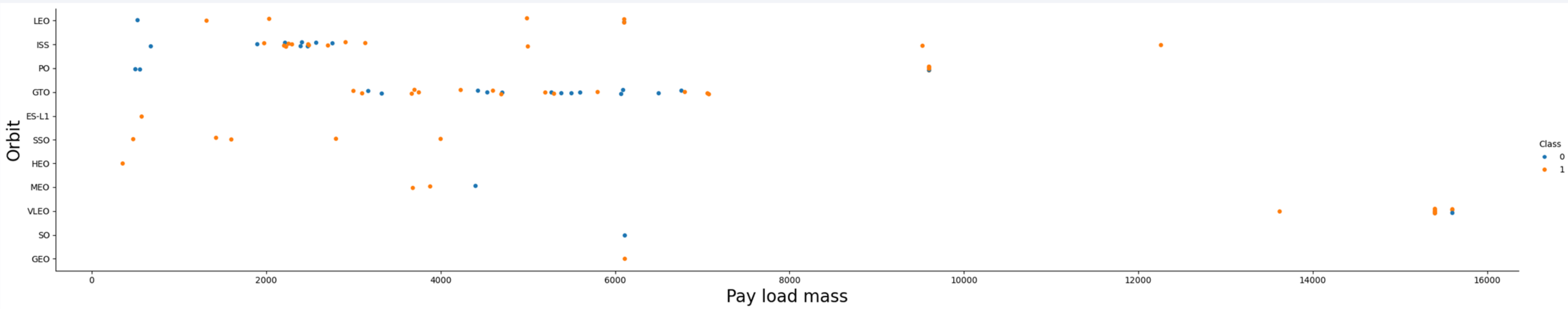
Flight Number vs. Orbit Type



The scatter plot shows that:

- In the LEO orbit, success seems to be related to the number of flights.
- In the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

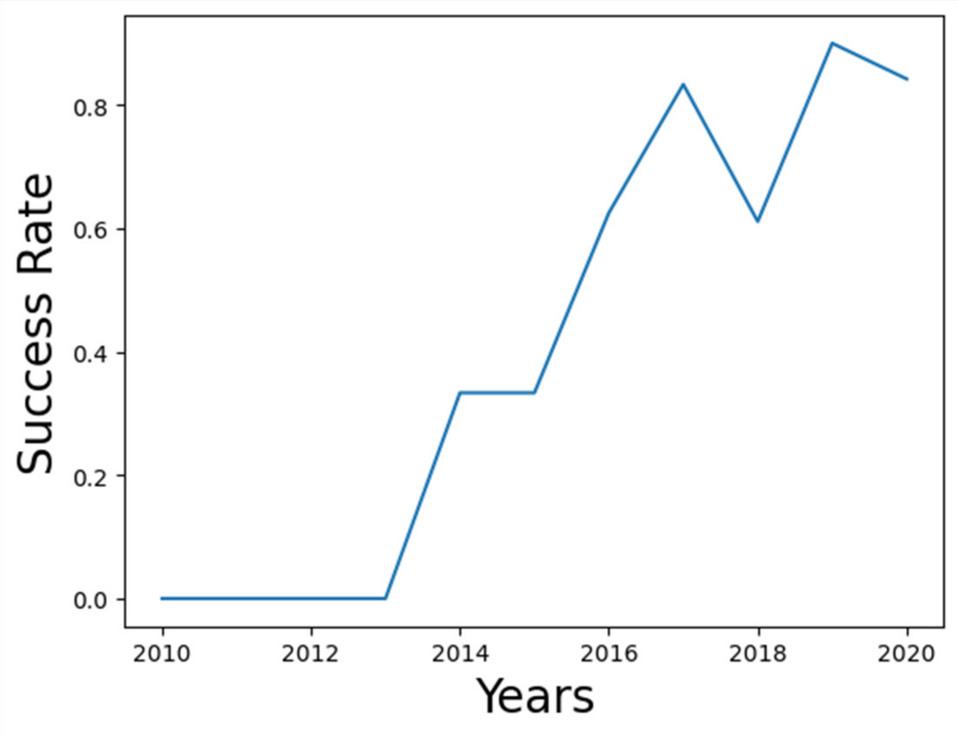


The scatter plot shows that:

- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- For GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

The line chart shows that the success rate since 2013 kept increasing till 2020 of yearly average success rate



All Launch Site Names

- The data shows that these are four launch sites:

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

- The information was obtained by using the SQL query select distinct launch_site values from the SPACEXTABLE table.

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	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

- The information was obtained by using the SQL query Select all the records from the SPACEXTABLE table where launch_site is like ‘CCA%’ and we limited to 5 records.

Total Payload Mass

- The total payload mass carried by boosters launched by NASA (CRS)

total_payload_mass
45596

- The total was obtained by using the SQL query to sum all the payload mass from the SPACEXTABLE table where the customer column contain ‘NASA (CRS)’.

Average Payload Mass by F9 v1.1

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

average_payload_mass
2928.4

- The average was obtained by using the SQL query to average all the payload mass from the SPACEXTABLE table where the booster version is like ‘’%F9 v1.1'

First Successful Ground Landing Date

- This is the date of the first successful landing outcome on ground pad

first_successful_landing
2015-12-22

- The date was obtained by using the SQL query to select the minimum date from the SPACEXTABLE table where the landing outcome is equal to 'Success (ground pad)'

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

- The names were obtained by using the SQL query to select the booster version from the SPACEXTABLE table where the landing outcome is equal to 'Success (ground ship) and the payload mass is between 4000 and 6000.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

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

- The total was obtained by using the SQL query to select the mission outcome and count them from the SPACEXTABLE table and then group them by the mission outcome.

Boosters Carried Maximum Payload

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

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

Booster_Version
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- The information was obtained by using the SQL query to select the booster version from the SPACEXTABLE table where the payload mass is the maximum from the SPACEXTABLE table.

2015 Launch Records

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

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

- The information was obtained by using the SQL query to select the month, booster version, launch site and landing outcome from the SPACEXTABLE table where the landing outcome was equal to Failure (drone ship) and year is equal to 2015.

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

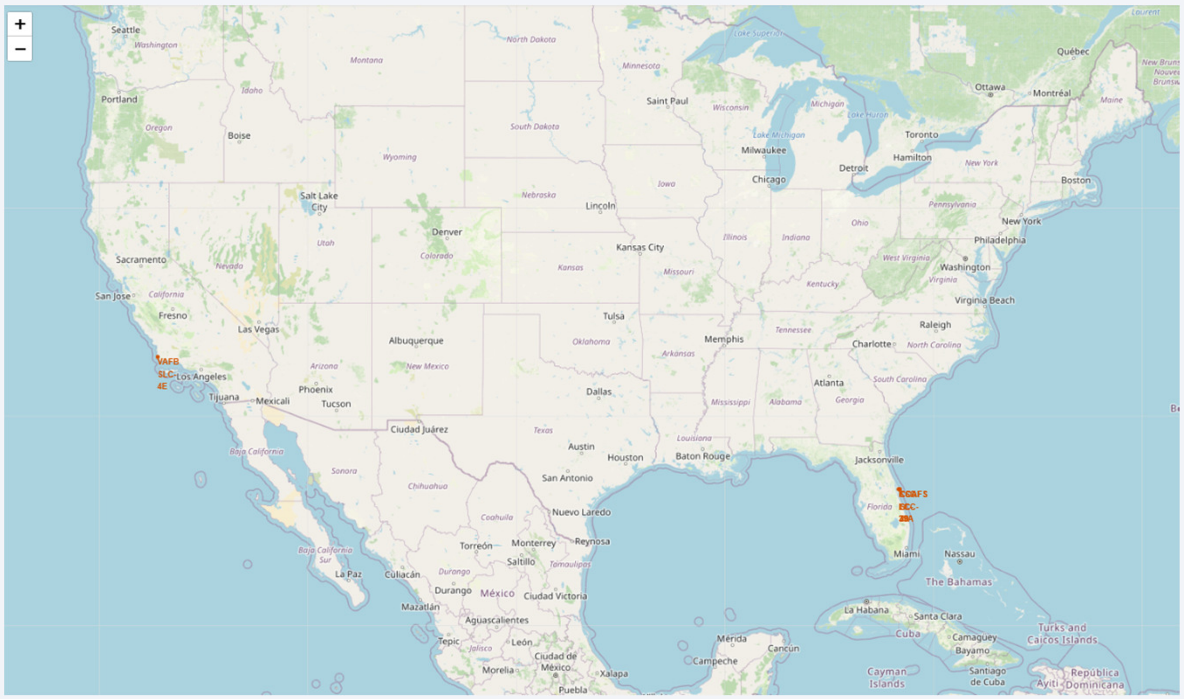
The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	count_outcomes
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

- The information was obtained by using the SQL query to select the landing outcome and do a count from the SPACEXTABLE table where the date is between 2010-06-04 and 2017-03-20 and group by landing outcome then ordered by count outcomes in a descending order.


Section 3
Launch Sites
Proximities Analysis

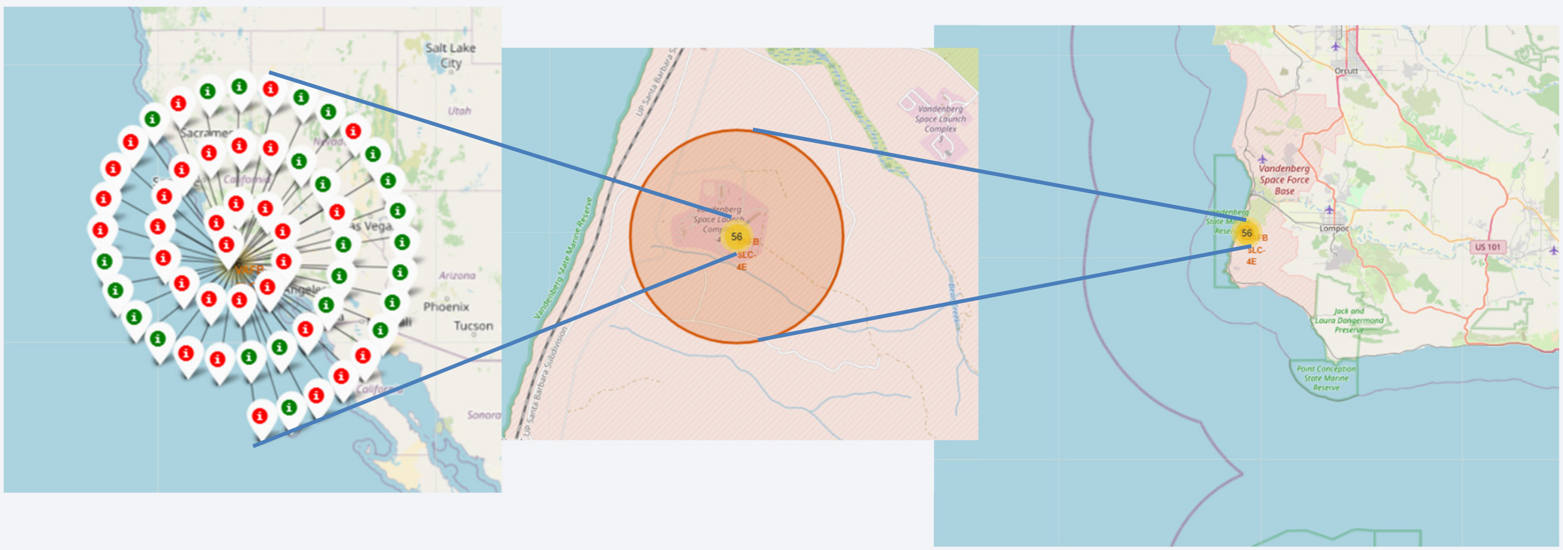
All launch sites



- Launch sites are near the sea, not too far from roads and railroads.
- Proximity to the sea increases the level of safety for the population around the sites in the event of an accident.

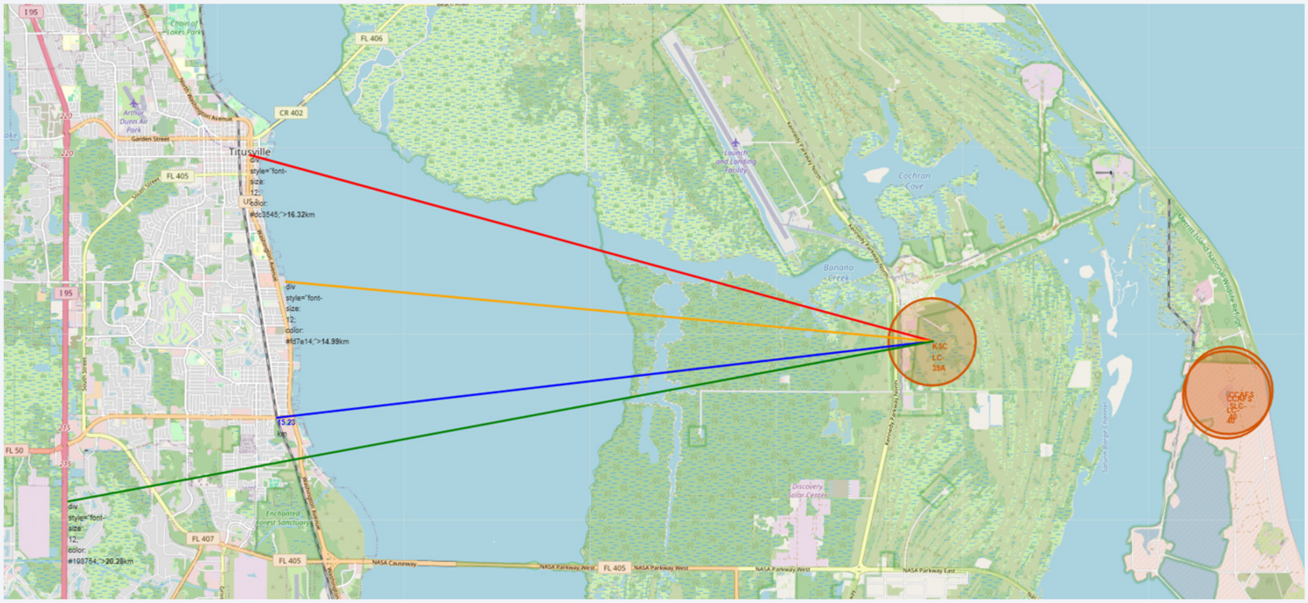
Launch Outcome by site

Example of a launch site launch outcomes



Green markers indicate successful and Red markers indicate failure

Logistics and Safety

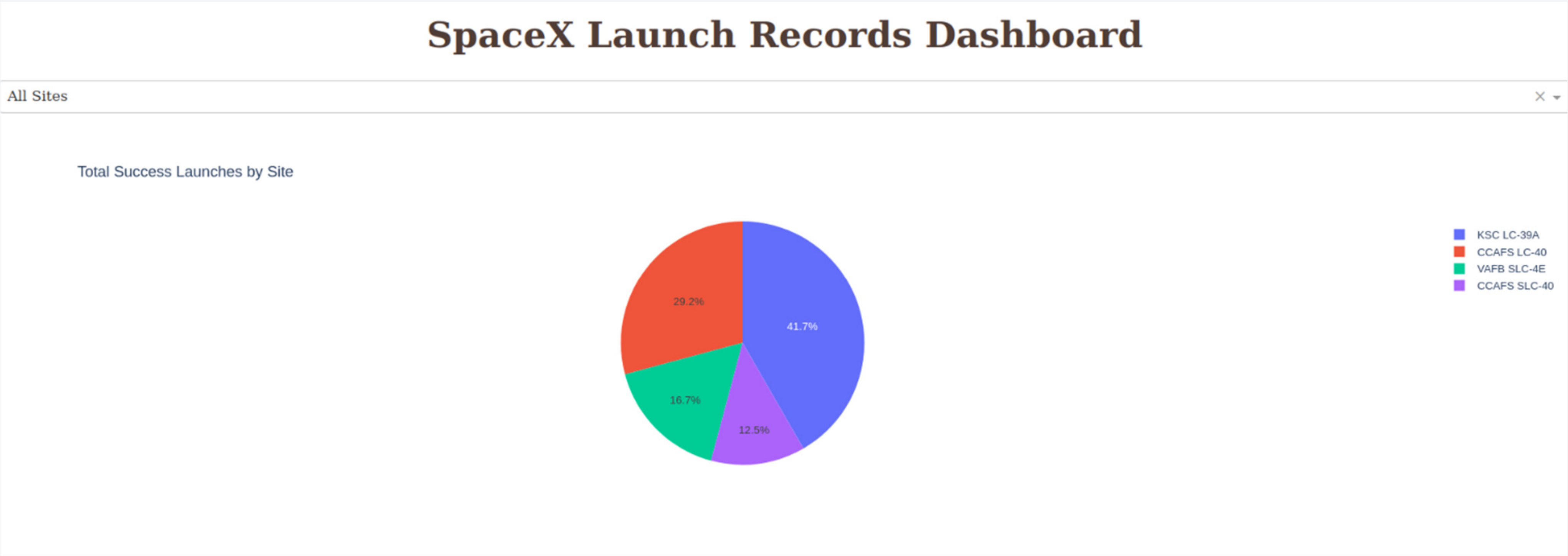


- This launch site has good logistic infrastructure such as railway, highway, coastline, with distance calculated and displayed
- Proximity to the sea increases the level of safety for the population around the sites in the event of an accident.

Section 4

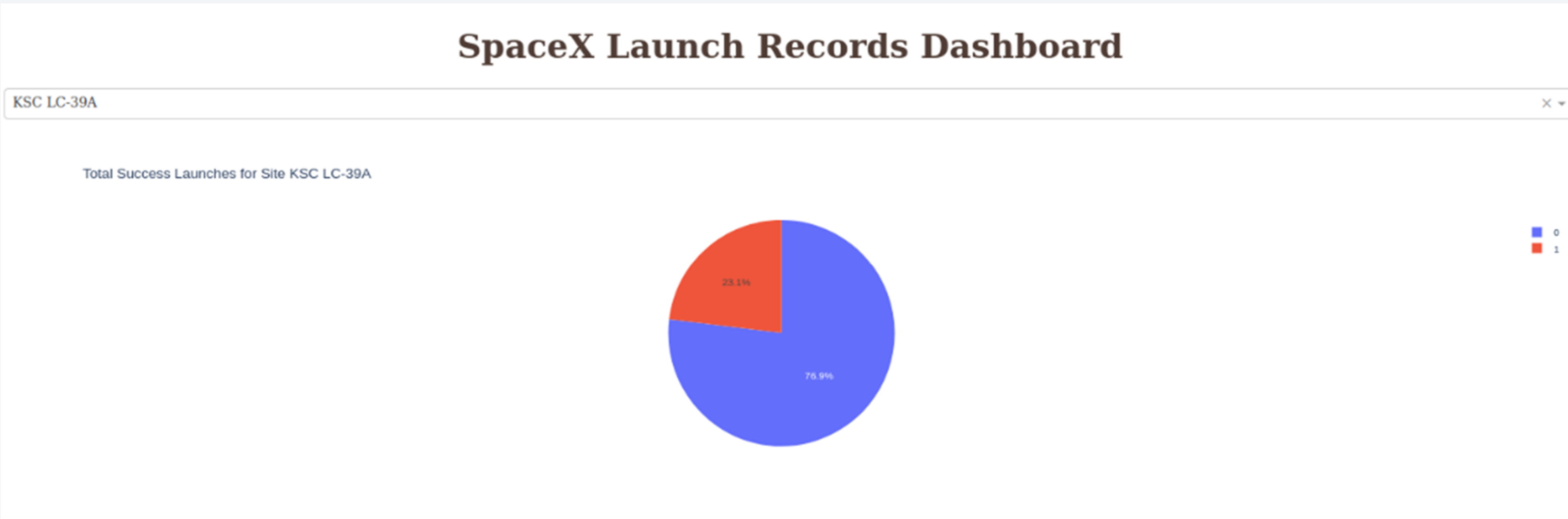
Build a Dashboard with Plotly Dash

Successful Launches by Site



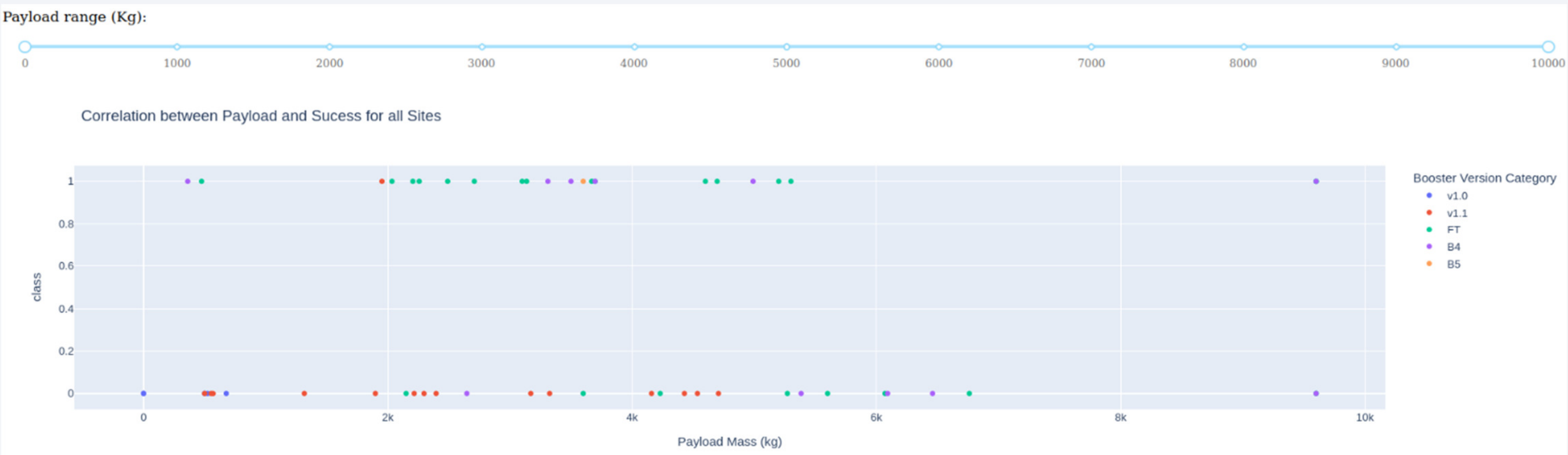
The location of the launch site is an important factor of the success of missions

Launch Success Ratio for KSC LC-39A



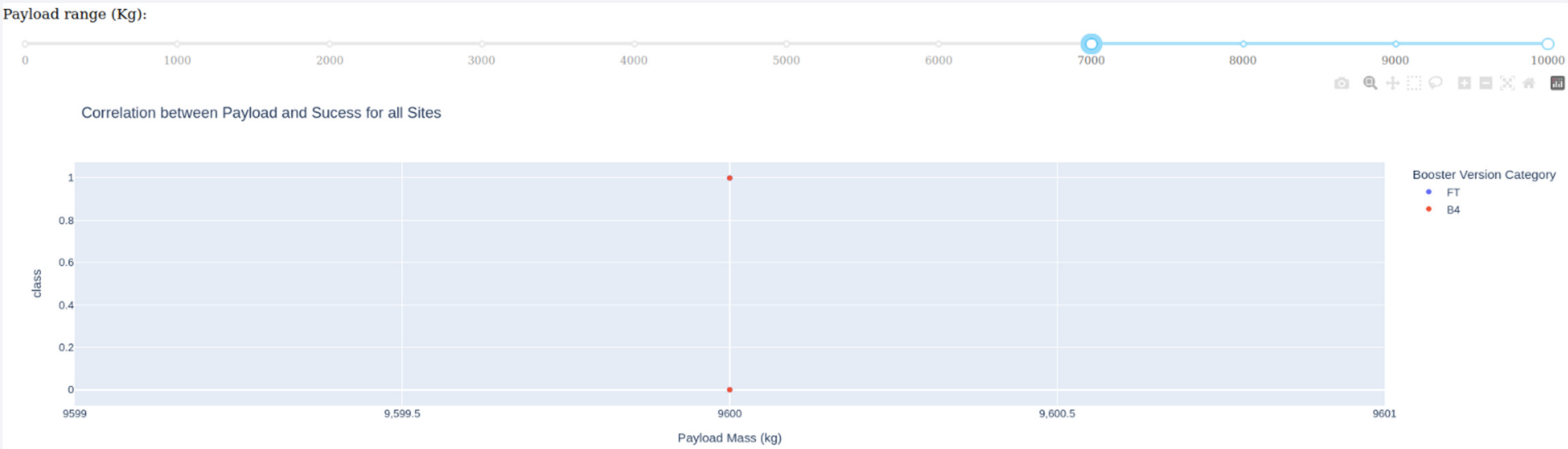
This site has 76.9% of successful launches.

Payload vs. Launch Outcome



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

Payload vs. Launch Outcome



There is not enough data to estimate risk of launches with payloads of over 7,000 kg.

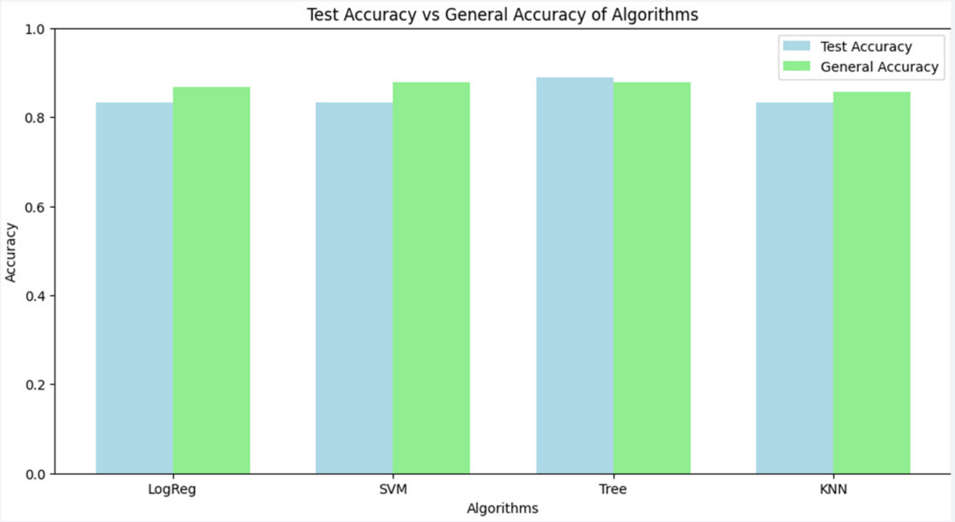
Section 5

Predictive Analysis (Classification)

Classification Accuracy

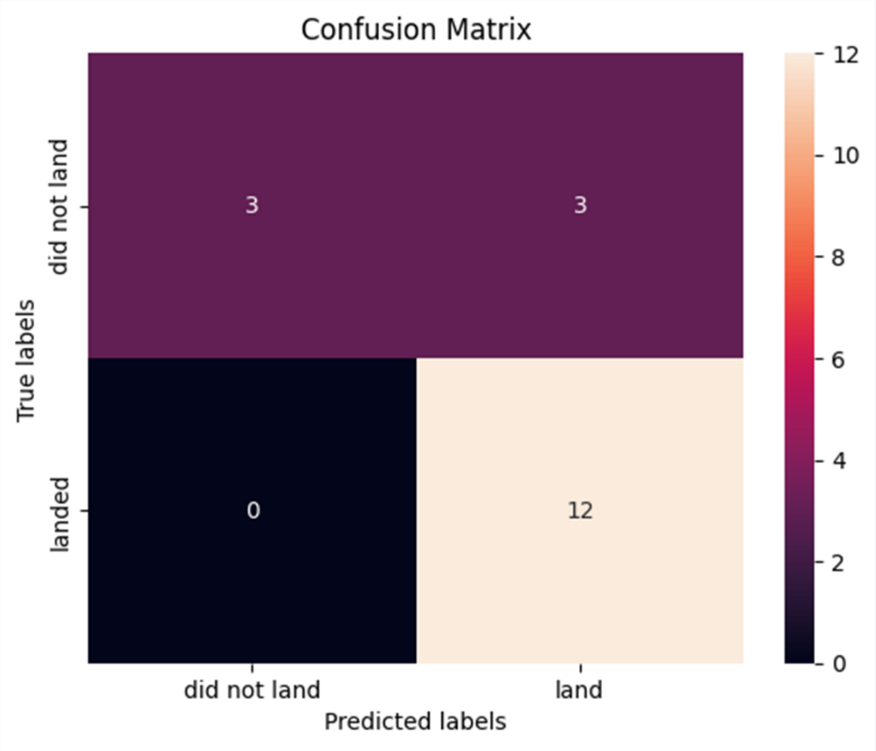
Four classification models were tested.

The Decision Tree classifier is the best model to predict successful landings, having an accuracy of 87.7% and an accuracy of 88.88% on the test data.



Confusion Matrix

The confusion matrix of the decision tree classifier shows its accuracy by showing a large number of true positives and true negatives compared to the false ones.



Conclusions

Different data sources were analyzed, narrowing down to the best model

The best launch site is KSC LC-39A

Launches above 7,000 kg have less risk.

Successful landing outcomes seem to improve over time.

Decision tree classifier can be used to predict successful landings and thus increase profits.

Appendix

This power point presentation should have the format shape set as Do not Autofit as standard.

Folium did not show map on Github.


Thank you!