Winning Space Race with Data Science

Bruno Ilves
2024/02/10

Outline

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

Executive Summary

- For this project, multiple machine learning algorithms were used to predict whether a mission was a success or not. Data was collected using public SpaceX REST API and SpaceX Wikipedia page to perform Exploratory Data Analysis (EDA), train the model via machine learning, produce data visualization and interactive dashboards

- Multiple machine learning algorithms were trained to predict successfulness of the mission. Model evaluations showed that they have very similar accuracy score with false positive results.

Introduction

- Due to ability to land and reuse first stage of rocket, SpaceX`s costs are much lower compared to other private space launch companies. SpaceY is interested in evaluating the viability of competing with SpaceX using their public data and machine learning.

- The objective of the project is to train and evaluate the machine learning algorithms, using public launch data from SpaceX, to predict successfulness of landing the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection:
  - SpaceX data was obtained:
    - From public SpaceX REST API (https://api.spacexdata.com/v4/rockets/)
    - By scraping SpaceX Wikipedia page (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches)
- Data wrangling
  - Landing outcome label “class” was created to represent outcome of each launch. If the value is zero, the first stage did not land successfully, if the value is one - first stage landed successfully.
- Exploratory data analysis (EDA) using visualization and SQL was performed

Methodology

Executive Summary

- Interactive visual analytics using Folium and Plotly Dash were created
- Performed predictive analysis using classification models
    - Collected data was normalized, split between training and testing data sets and used for training and evaluating four different machine learning algorithms

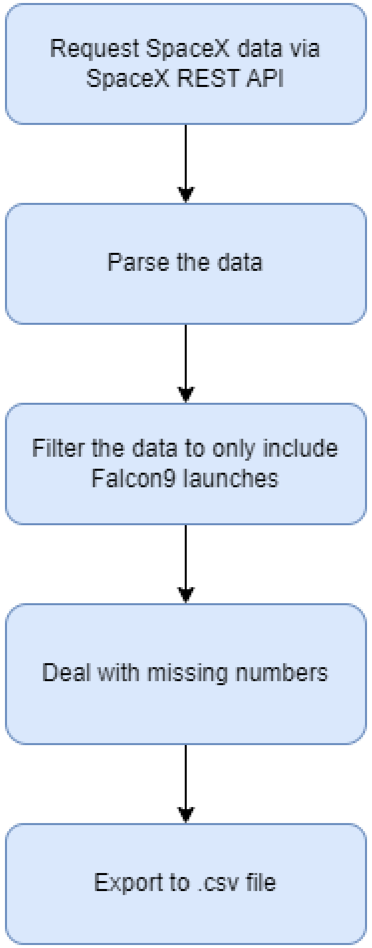
Data Collection

- SpaceX data was obtained from public SpaceX REST API (https://api.spacexdata.com/v4/rockets/) and by scraping SpaceX Wikipedia page (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- Data was collected via accessing public SpaceX REST API
- Data collection steps can be seen on the flowchart

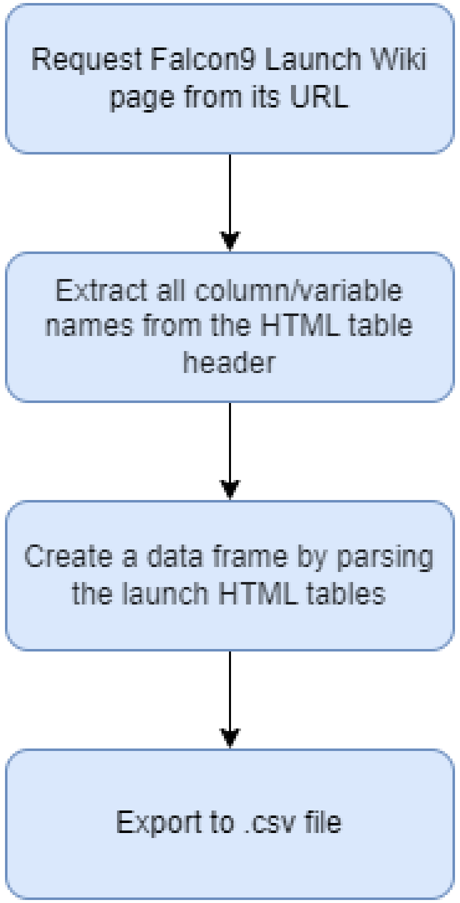
- GitHub



Data collection flowchart – SpaceX API

Data Collection - Scraping

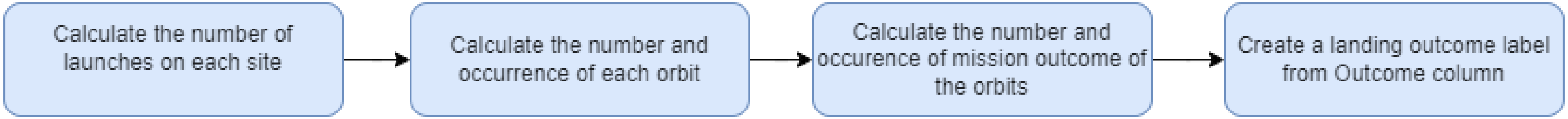
- Data was collected by scraping SpaceX Wikipedia page (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches)
- Data collection steps can be seen on the flowchart
- GitHub



Data collection flowchart – webscraping

Data Wrangling

- Landing outcome label “class” was created to represent outcome of each launch. If the value is zero, the first stage did not land successfully, if the value is one - first stage landed successfully.



Data wrangling flowchart

- GitHub

EDA with Data Visualization

- Scatterplots and barplots were used to analyze relationships between different features, such as :
    - Flight Number vs Payload Mass, Flight Number vs Launch Site, Payload Mass vs Launch Site, Orbit vs Class, Flight Number vs Orbit, Payload Mass vs Orbit, Date vs Class

- GitHub

EDA with SQL

- The following SQL queries were performed:
  - **Display the names of the unique launch sites in the space mission**
  - **Display 5 records where launch sites begin with the string 'CCA'**
  - **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 between the date 2010-06-04 and 2017-03-20, in descending order.**
- [GitHub]()

Build an Interactive Map with Folium

- Folium map objects (e.g., lines, circles, markers and marker clusters) were used to indicate points of interest, such as:
  - Markers to indicate specific launch sites
  - Circles to highlight launch site areas
  - Marker clusters to indicate multiple launches at a launch site
  - Lines to indicate distance between launch site and closest shore point

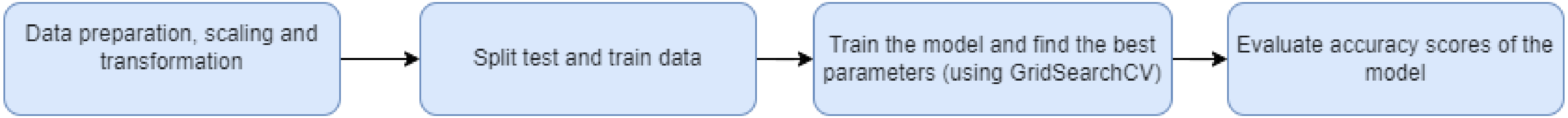
- GitHub
- Link with rendered maps

Build a Dashboard with Plotly Dash

- Pie charts and scatter plots were used to visualize the following data:
  - Percentage of successful launches in a specific launch site and total in all launch sites combined
  - Payload mass vs success rate for specific launch site and for all launch sites combined
- Scatterplots have a slider for maximum and minimum payload mass to help filter the data
- GitHub

Predictive Analysis (Classification)

- Four different machine learning algorithms (logistic regression, SVM, decision trees and KNN) were trained and their accuracy scores were evaluated.
- The following flowchart depicts main steps for predictive analysis



Predictive analysis flowchart

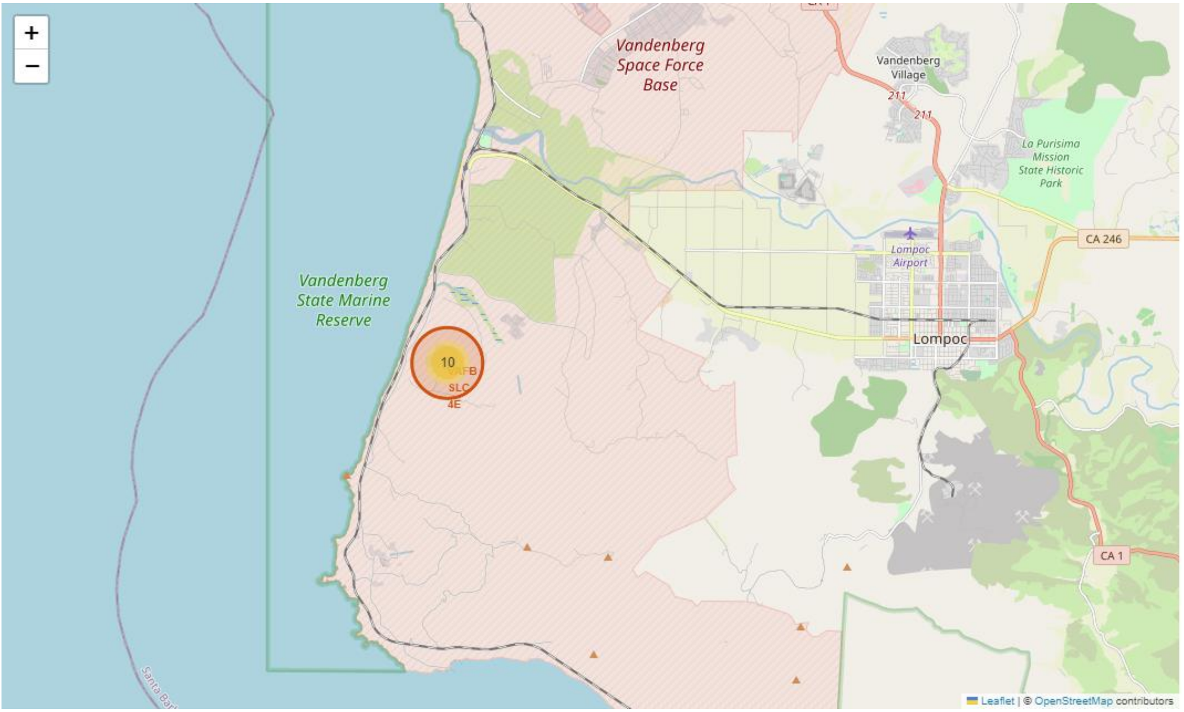
- GitHub

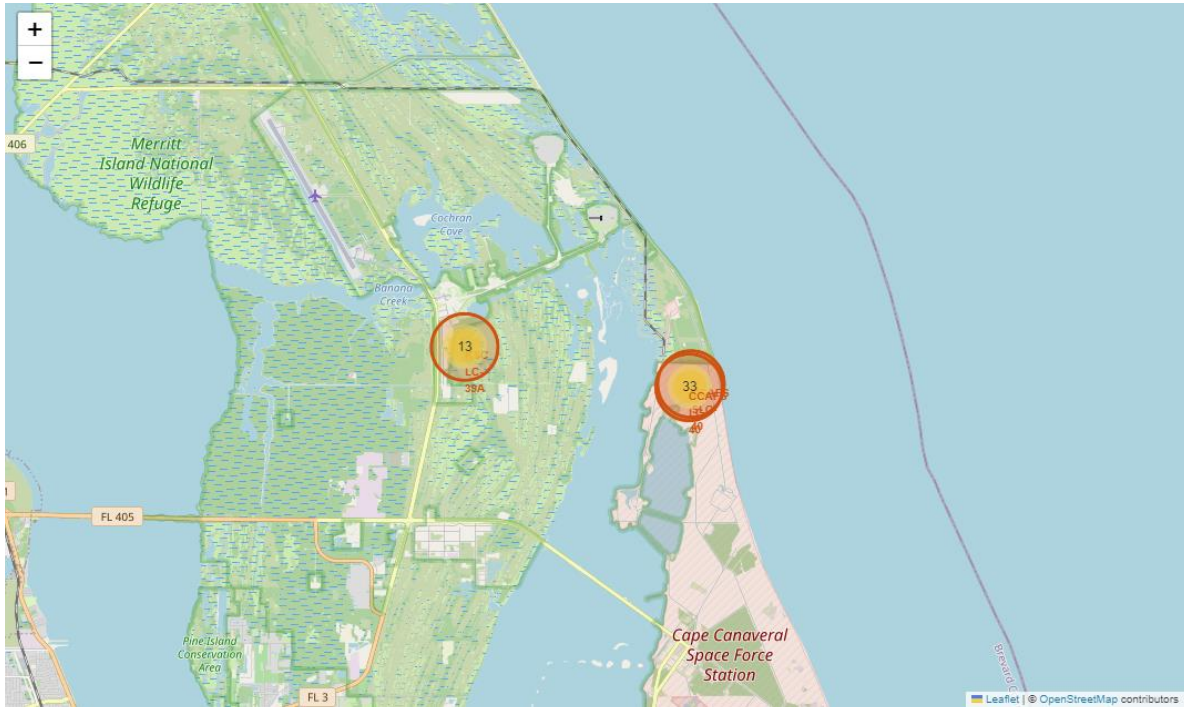
Results

- Exploratory data analysis results:
  - 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 77%.
  - 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.
  - With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS orbits.
  - Launch success rate kept increasing since 2013 till 2020

Results

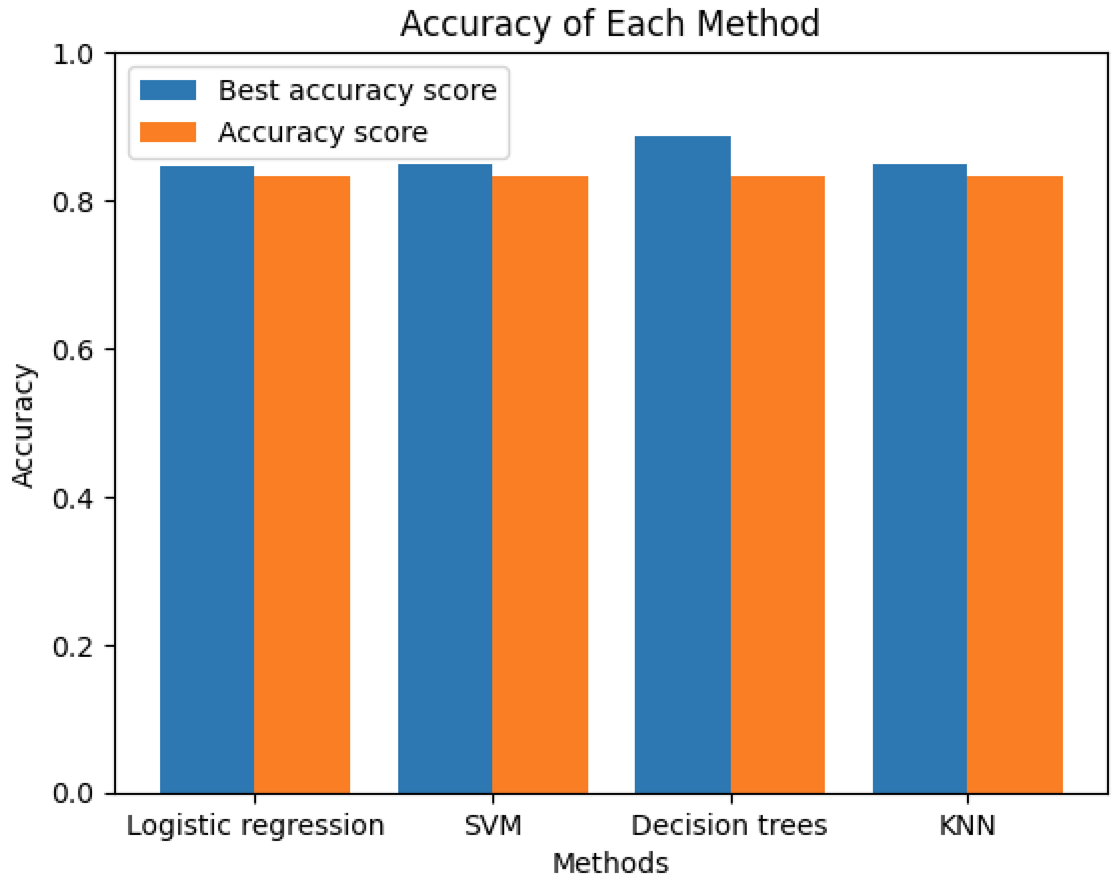
- Using interactive analytics, the launch sites were identified to be located close to equator, near the shoreline and with a good infrastructure around.





Results

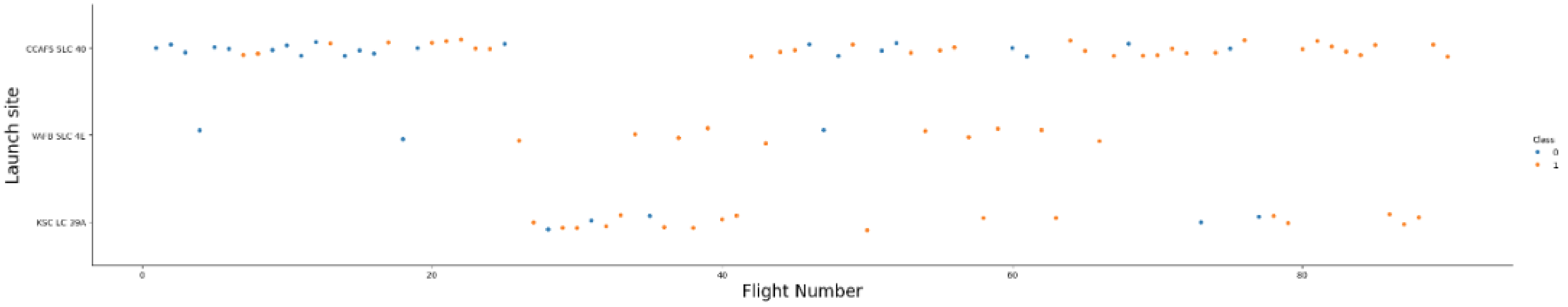
- Predictive analysis showed that out of the four tested methods, decision trees have the best accuracy score when used with best hyperparameters



Section 2

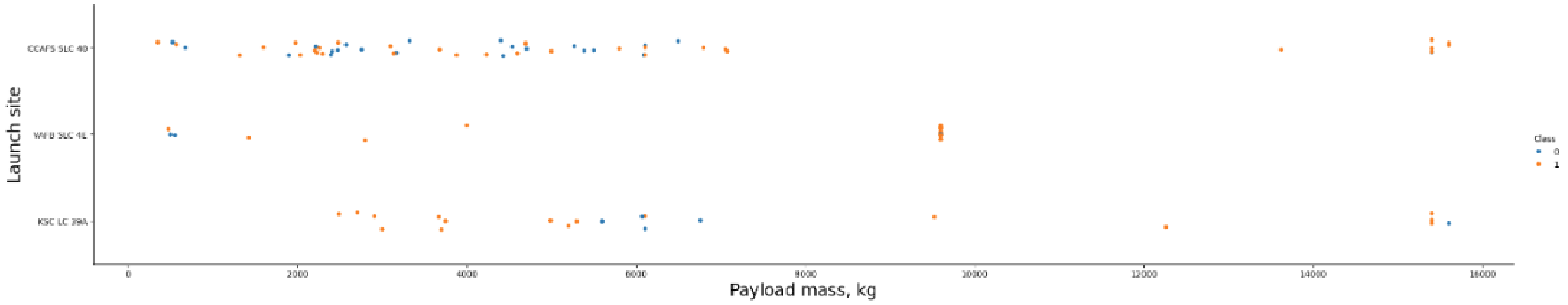
Insights drawn from EDA

Flight Number vs. Launch Site



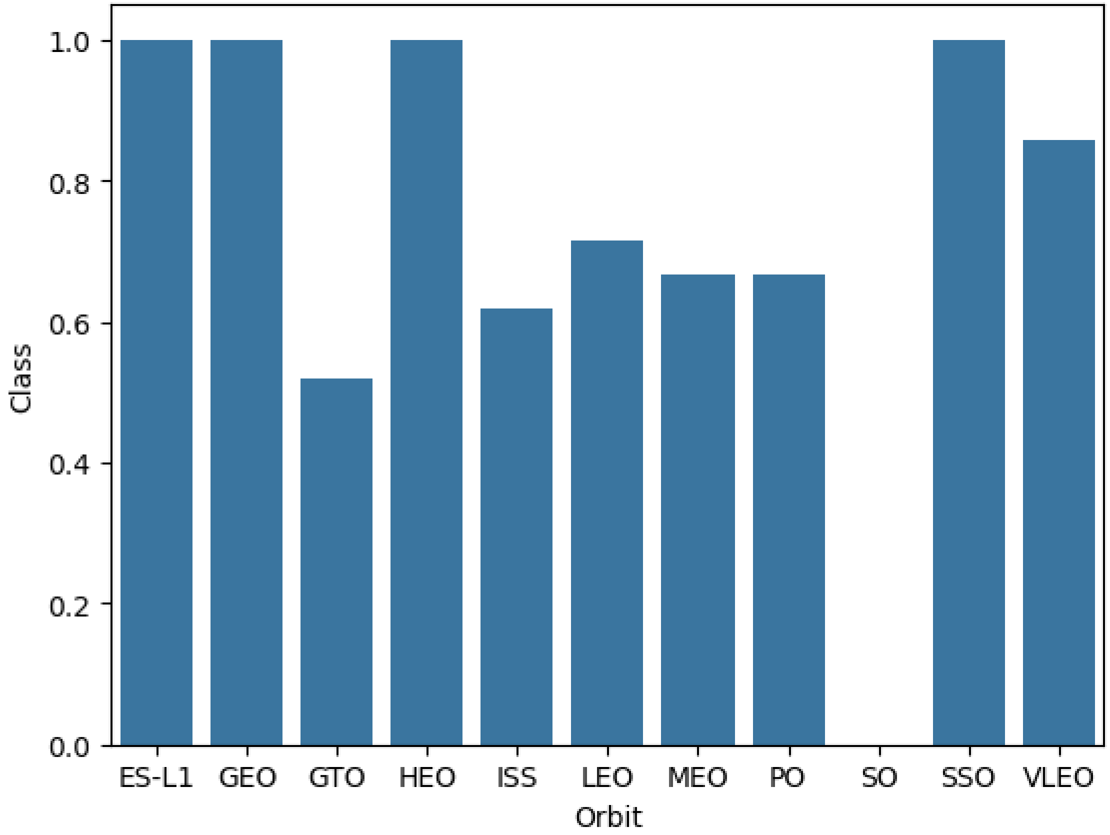
- We see that different launch sites have different success rates. CCAFS SLC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

Payload vs. Launch Site



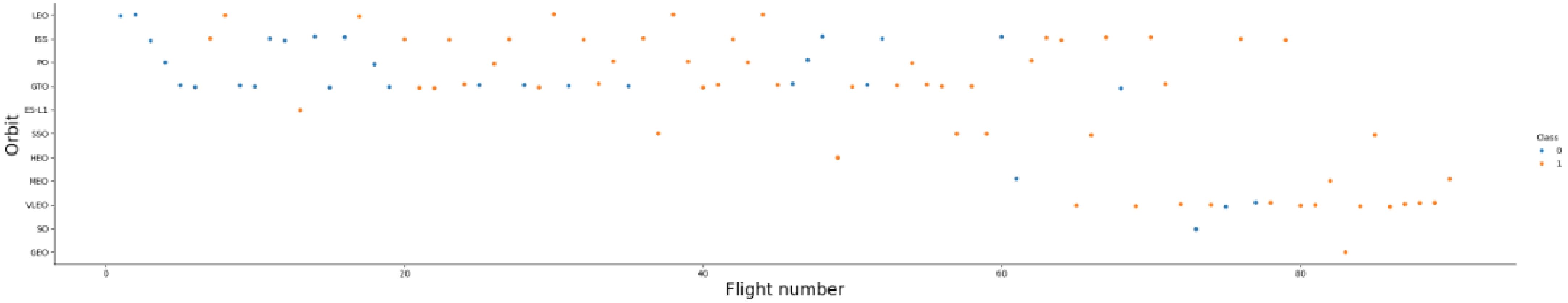
- For VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



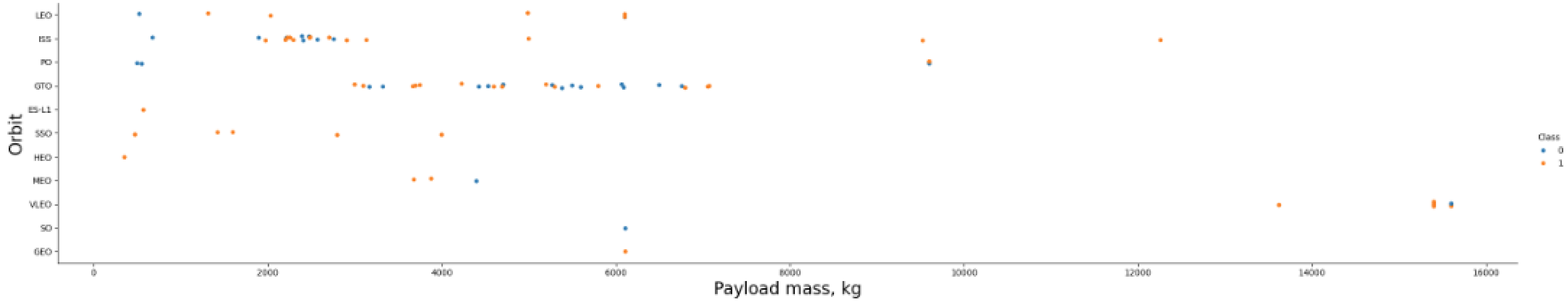
- Orbits with the highest success rate are:
  - ES-L1
  - GEO
  - HEO
  - SSO

Flight Number vs. Orbit Type



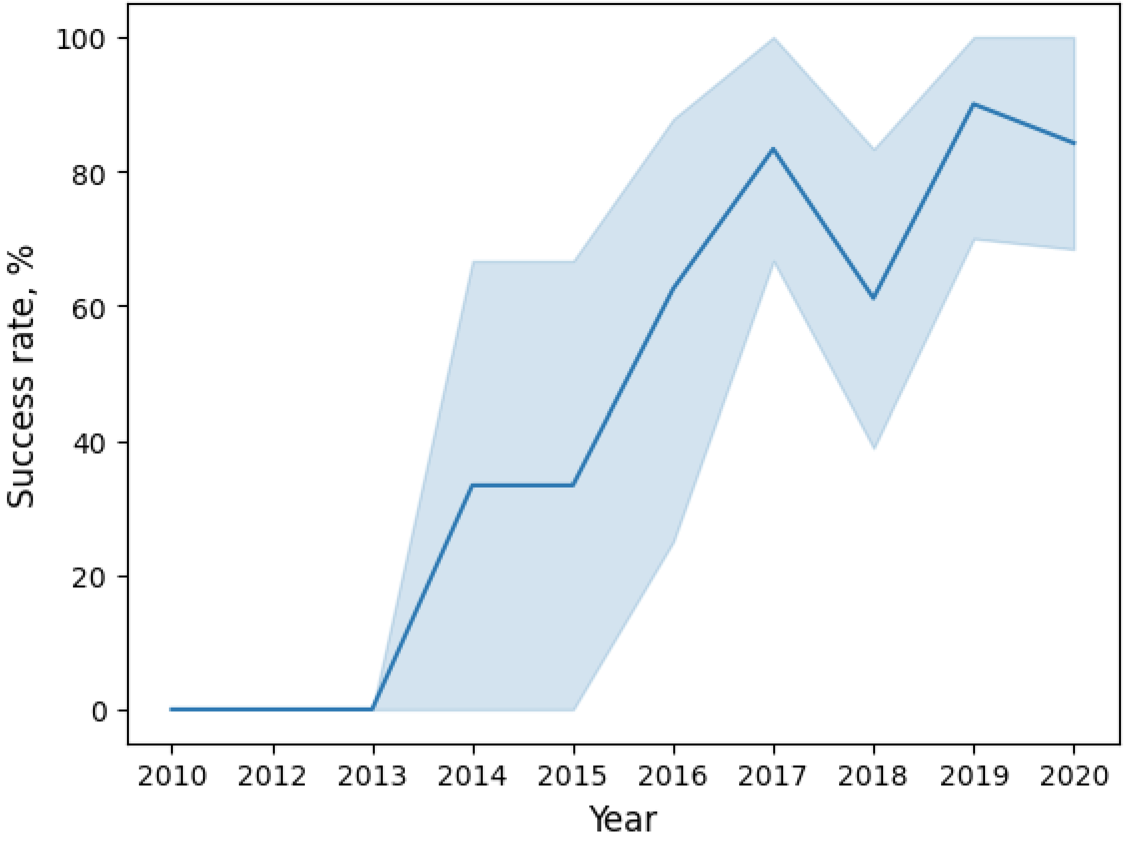
- In the LEO orbit, the success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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



- Success rate started to increase in 2013 and kept increasing till 2020

All Launch Site Names

- There are four unique launch sites

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

Launch Site Names Begin with 'CCA'

- Below are five records for launch sites that begin with “CCA”
- All mission outcomes were success, despite landing failure or not attempt at landing

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

- Total payload mass carried by boosters launched by NASA

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- First successful ground landing in ground pad was achieved in 2015/12/22:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following booster versions 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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

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

Boosters Carried Maximum Payload

- The following list of booster versions 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
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

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

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

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

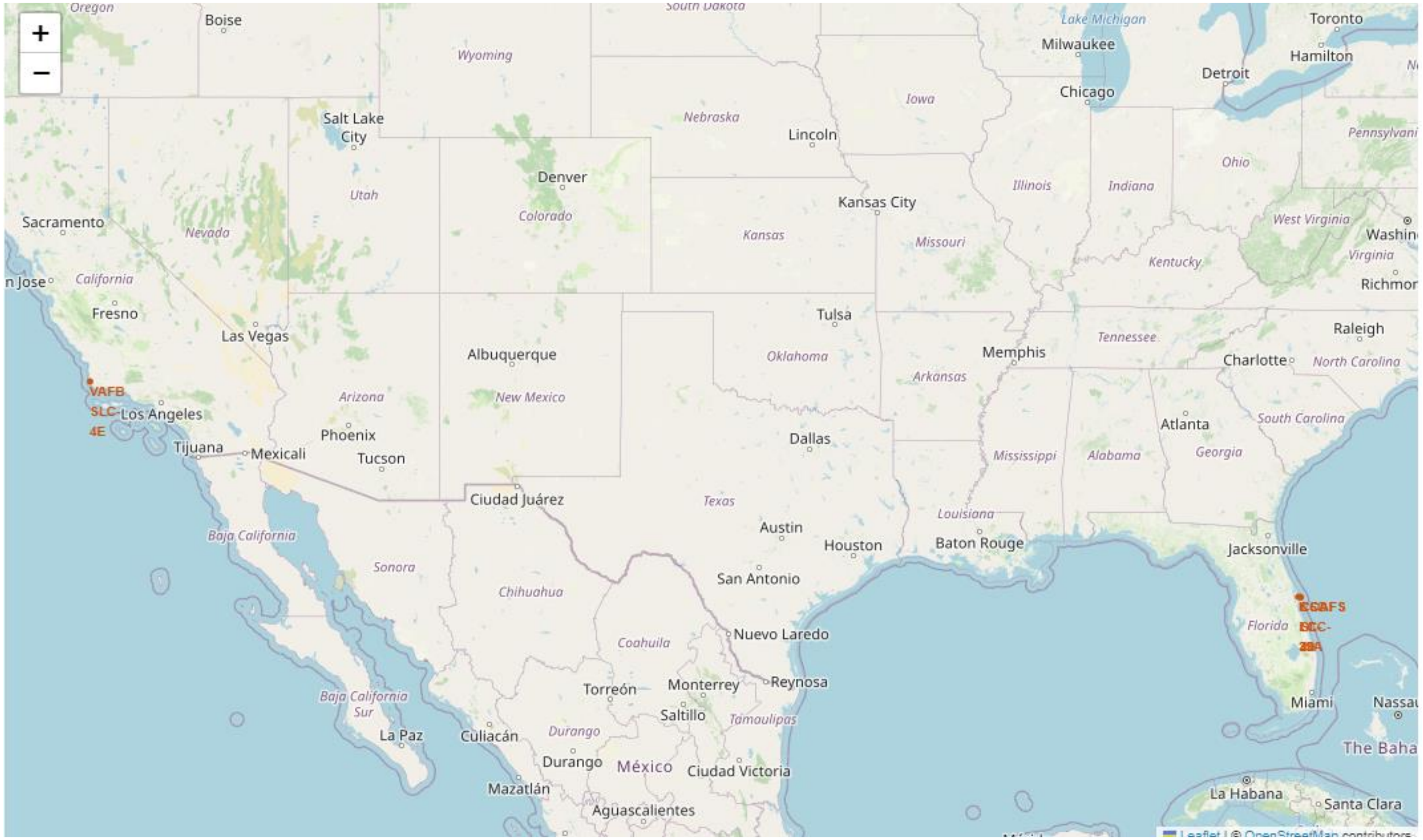
- List of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	Number_of_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

Section 3

Launch Sites Proximities Analysis

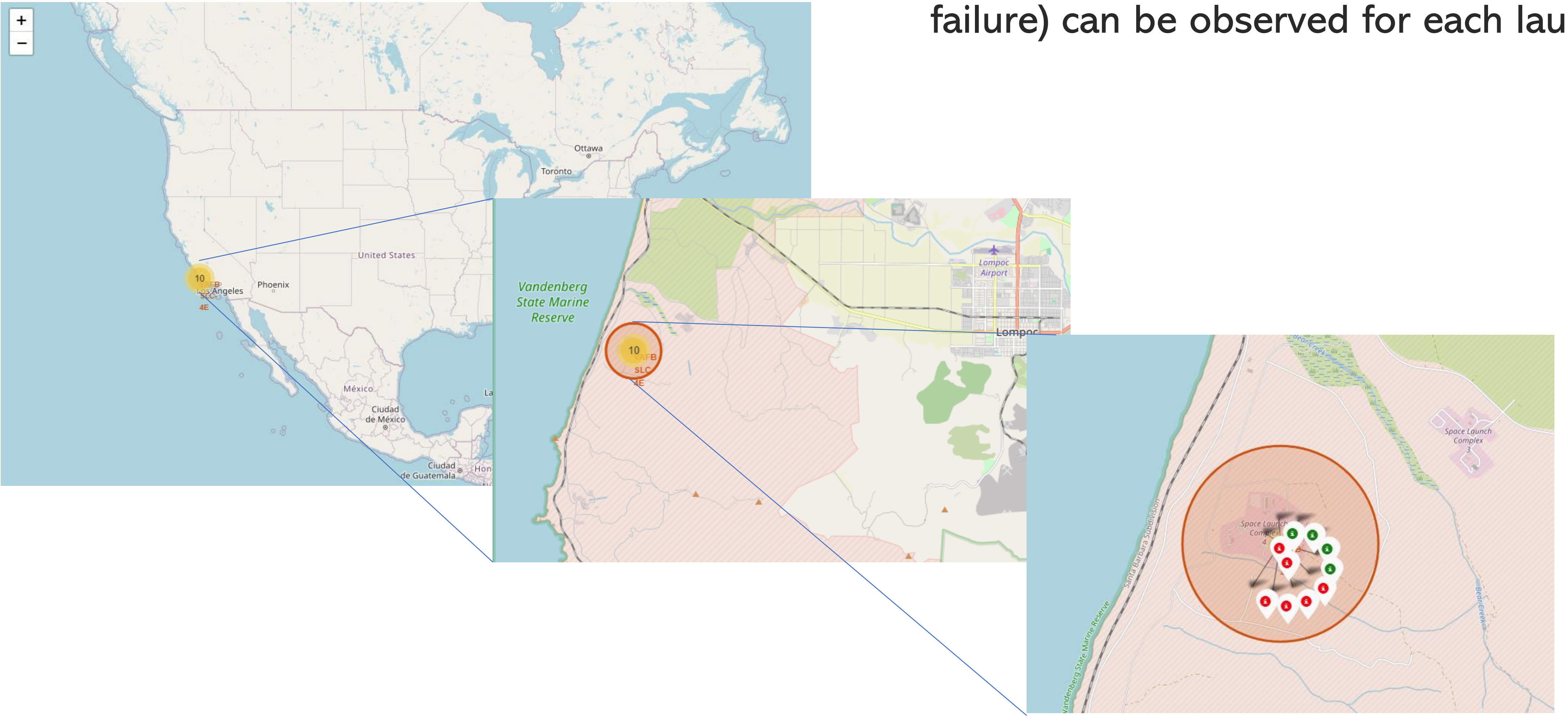
Map of all launch sites



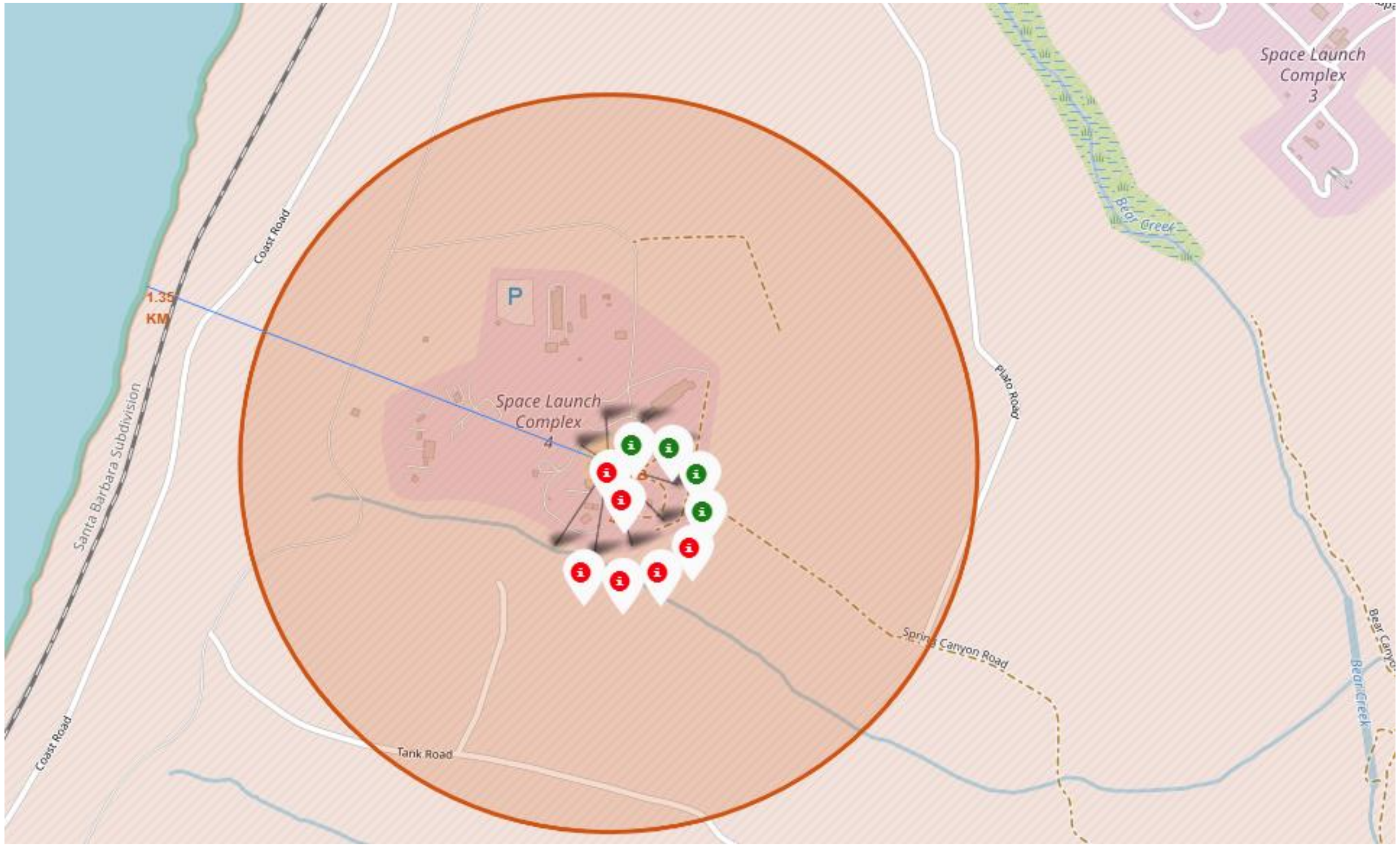
- All launch sites are located close to equator and near the shoreline

Launch outcomes by site

- Launch outcomes (green – success, red – failure) can be observed for each launch site



Distance to shoreline

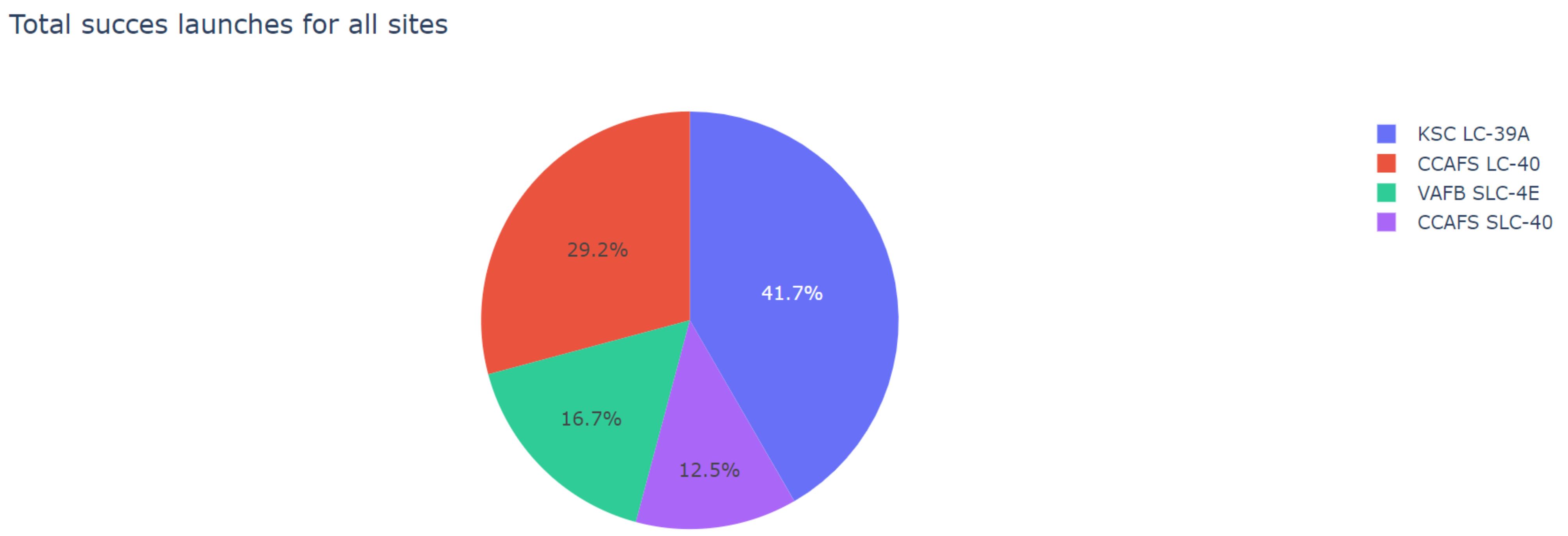


- We can observe the launch site with calculated distance to the nearest shoreline

Section 4

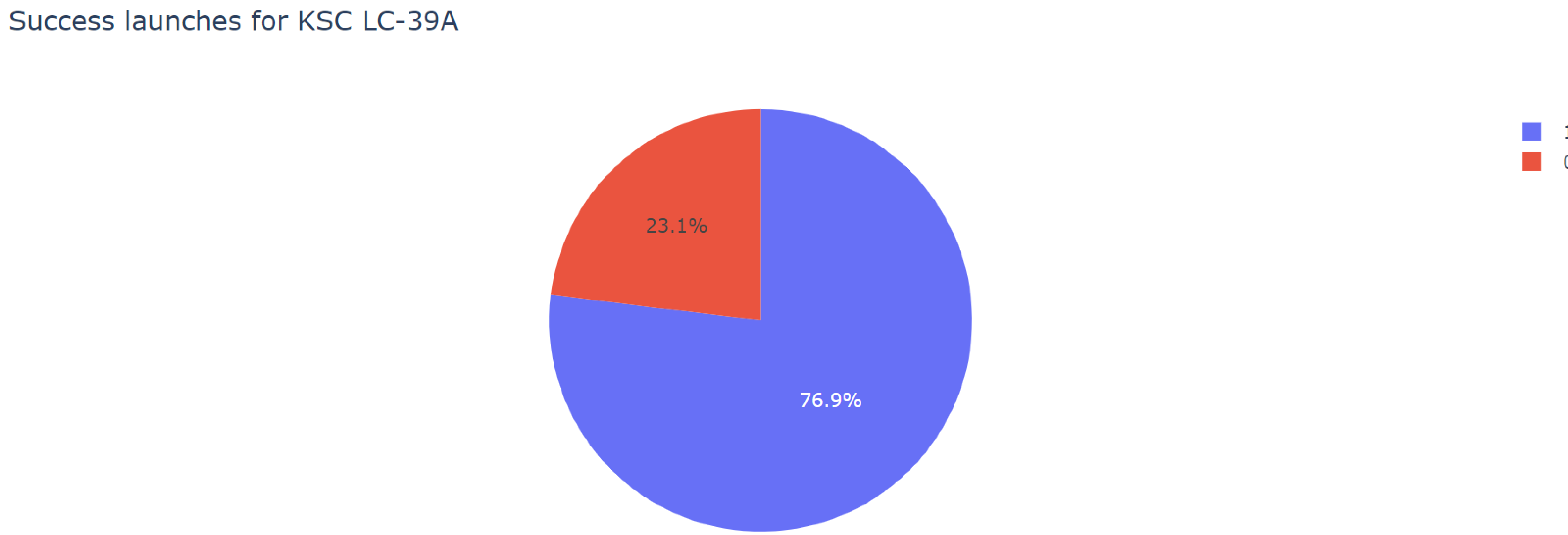
Build a Dashboard with Plotly Dash

Total success of launches for all sites



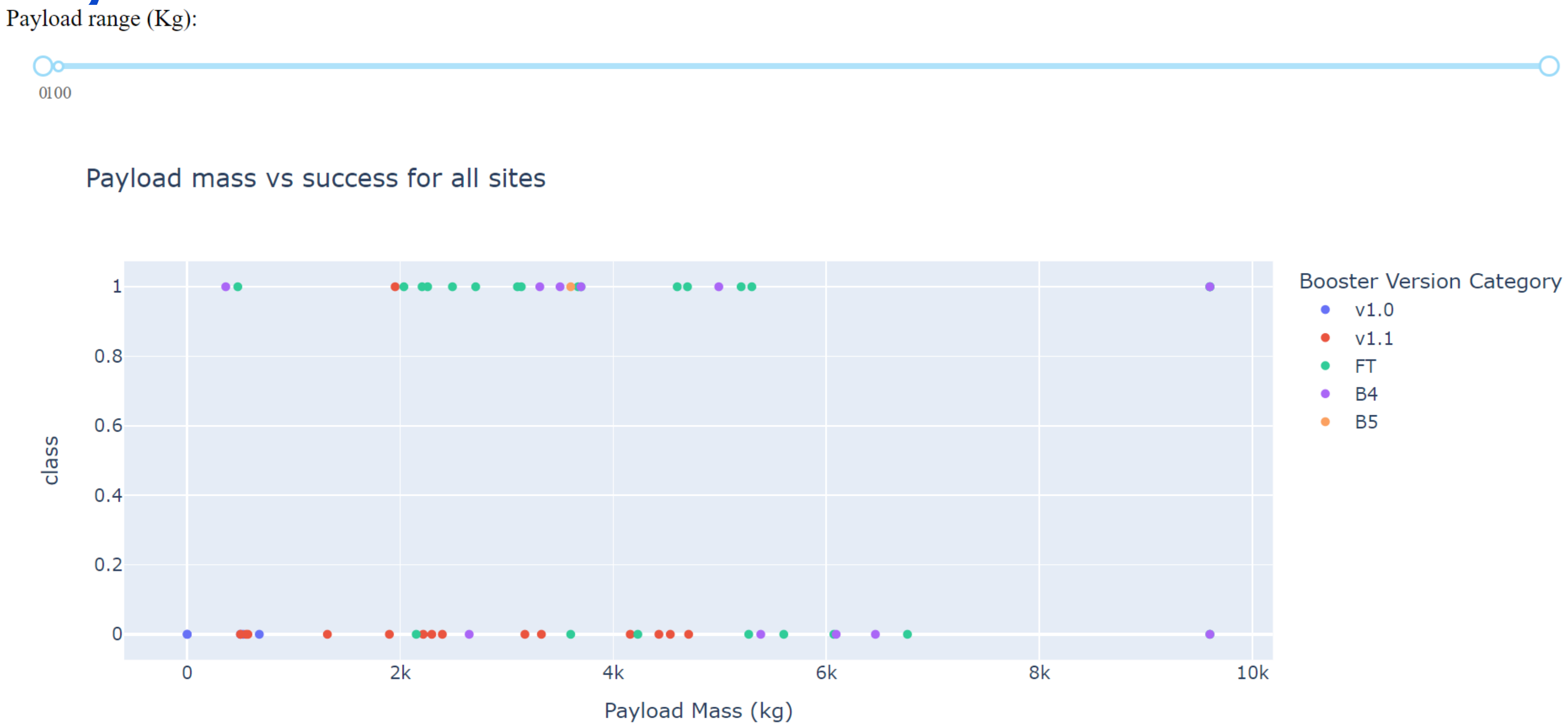
- Launch site KSC LC-39A seems to be the most successful

Success launches for launch site KSC LC-39A



• Launch site KSC LC-39A has a 76.9% success rate

Payload mass vs success rate for all sites



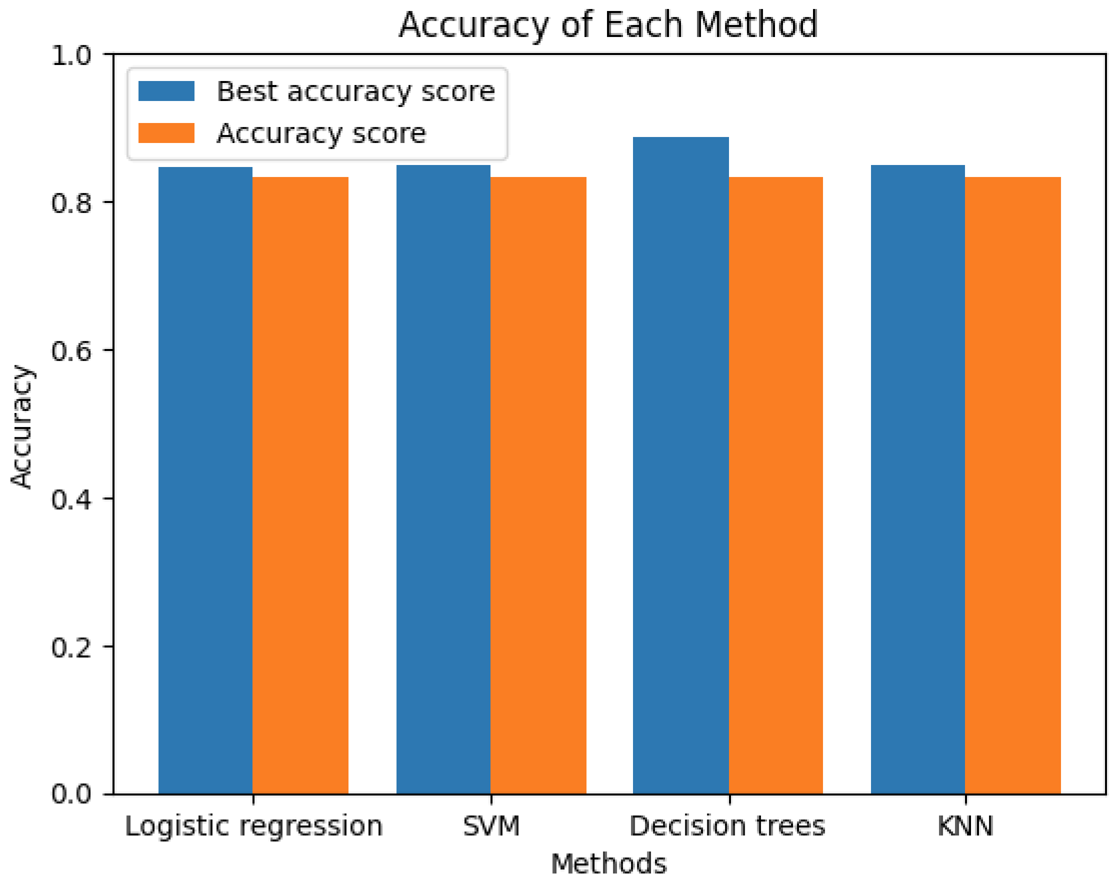
- Payloads under 6000 kg have much better success rate

Section 5

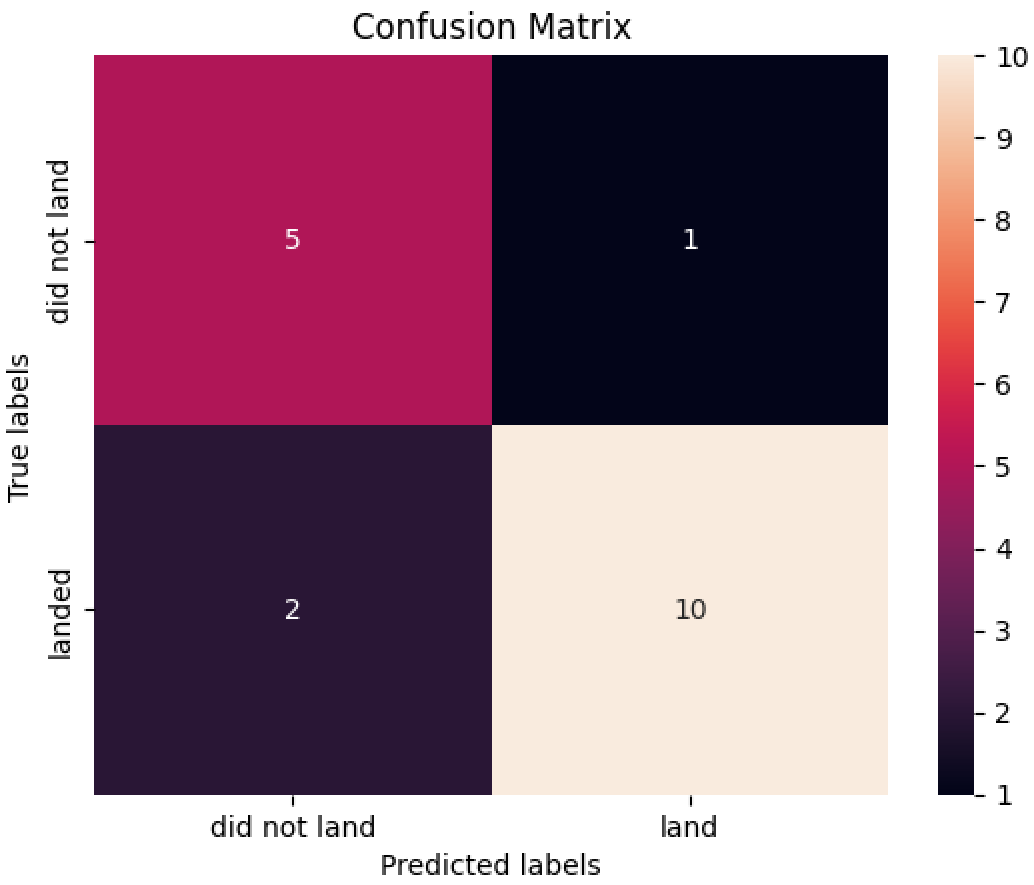
Predictive Analysis (Classification)

Classification Accuracy

- Accuracy evaluation showed that out of the four tested methods, **decision trees** have the best accuracy score when used with best hyperparameters



Confusion Matrix for decision trees



- Confusion matrix for decision trees show lowest number of false positives

Conclusions

- **Best Launch Site**: After analyzing the data, it was determined that Kennedy Space Center's Launch Complex 39A (KSC LC-39A) is the optimal launch site.
- **Launches Above 7,000kg**: Data analysis suggests that launches involving payloads heavier than 7,000kg are associated with lower risk.
- **Improvement in Successful Landings Over Time**: The analysis indicates a trend of improving success rates for landing outcomes over time.
- **Predictive Modeling with Decision Tree Classifier**: The use of a Decision Tree Classifier model has been proposed to predict successful landing outcomes.

Thank you!