IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jakub Sisma
2024/04/03

Outline

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

Executive Summary

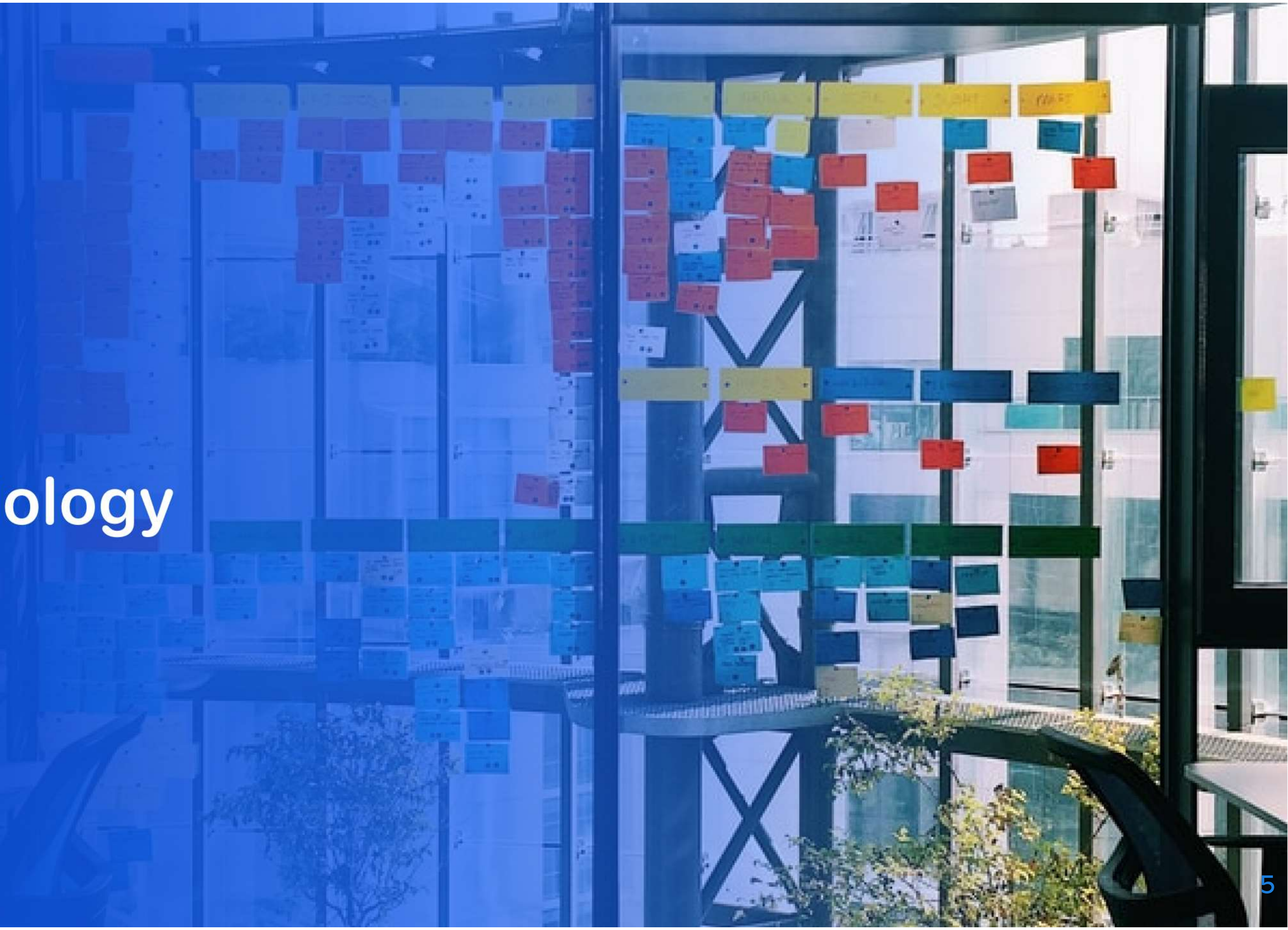
- This project uses data science methods to get insights into rocket launch costs and the success of landing stage 1.
- Based on created models using data science we can predict the success of landing the first stage in 83.3% of the time.

Introduction

- In this project we are going to take the role of a data scientists working for a new rocket company. Our job is to determine the price of each launch by gathering information about Space X and creating dashboards for our team. We will also determine if SpaceX will reuse the first stage.
- We will attempt to answer the following problems: where are the launches executed and on what scale? What is the rate of successful launches? Can we predict if the landing of stage 1 will be successful?

Section 1

Methodology



Methodology

Executive Summary

- Data collection methodology:
  - APIs, Webscraping
- Perform data wrangling
  - Dropp null values, create ‘Class‘ variable
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - SVM, Logistic Regression, Decision Tree,

Data Collection – SpaceX API

Find API

- Find the API address

Get data

- Use get function to get the data

Transform

- Transform the data file into a pandas dataframe

https://github.com/DonKubini/ibm-capstone/blob/main/Module%201/jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection - Scraping

Find Wikipedia artice

- Find the Wikipedia article url

Get data

- Use get function to get the data

Transform

- Use beautiful soup library to extract the data into a dataframe

https://github.com/DonKubini/ibm-capstone/blob/main/Module%201/jupyter-labs-webscraping.ipynb

Data Wrangling

- Important to create necessary features and deal with null values

Understand your data

- Use pandas functions to get basic understanding of the data structure

Deal with NA values

- Drop unnecessary null values or transform them into averages

Create class column

- Create a new column that contains the info on success of launch

https://github.com/DonKubini/ibm-capstone/blob/main/Module%201/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

EDA with Data Visualization

- We used multiple scatterplots to understand relationships between different variables
- We used a bar plot to visualize the success rate of launch per orbit
- We used a line chart to visualize the succes rate trend over time

https://github.com/DonKubini/ibm-capstone/blob/main/Module%202/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- We used SQL queries to display the following:
  - Unique launch sites in the space mission
  - Total payload mass of NASA (CRS) boosters
  - Booster versions that were successful with specific payload mass range
  - Total number of successful and unsuccessful mission outcomes
  - Boosters that carried the maximu payload mass
  - And more…

https://github.com/DonKubini/ibm-capstone/blob/main/Module%202/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We used interactive circles, markers and lines to identify launch sites and their respective successful and unsuccessful attempts
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

https://github.com/DonKubini/ibm-capstone/blob/main/Module%203/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- We used a drop down menu so the user can choose a location. The location will display a success rate pie chart
- We used a slider to select payload range. This will display the relationship between payload and success based on selected location

https://github.com/DonKubini/ibm-capstone/blob/main/Module%203/spacex-dash-app.py

Predictive Analysis (Classification)

Standardize and spilt the data

- Use StandardScaler to standardize the predictors
- Split the data into training and testing splits

CV and Grid Search

- Use grid search cross-validation to find the best parameters on the training data (for each model)

Test and evaluate the models

- Get predictions, accuracy score and confusion matrix for each model
- Evaluate different models

https://github.com/DonKubini/ibm-capstone/blob/main/Module%204/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

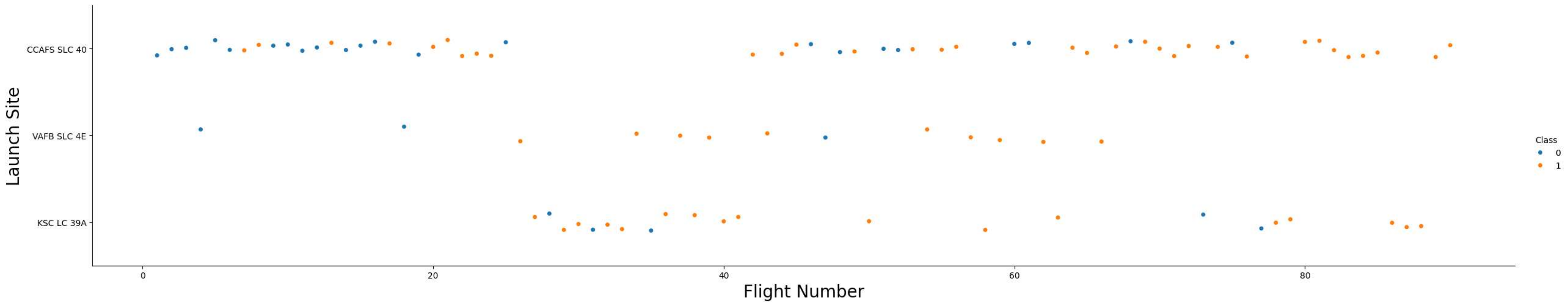
Results

- Models ranked on accuracy:
  - Logistic Regression: 83%
  - K-Nearest Neighbours: 83%
  - Support Vector Machines: 83%
  - Decision Tree: 77%

Section 2

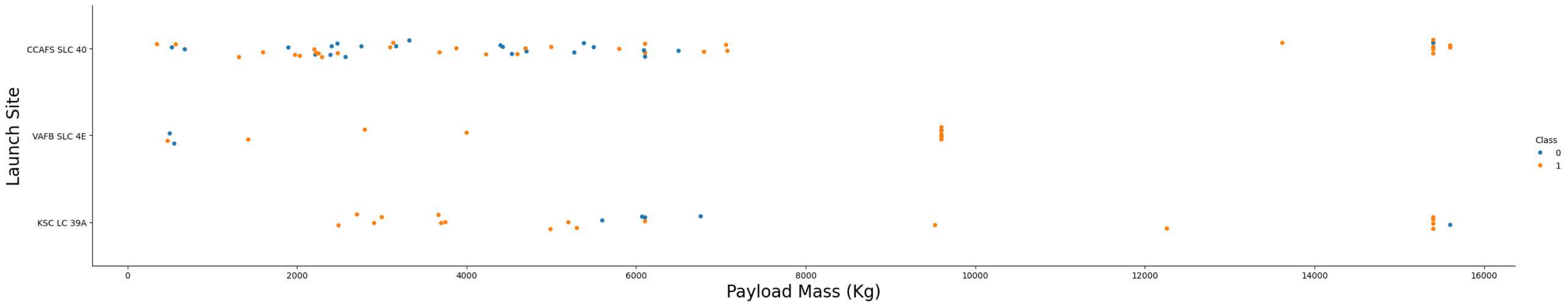
Insights drawn from EDA

Flight Number vs. Launch Site



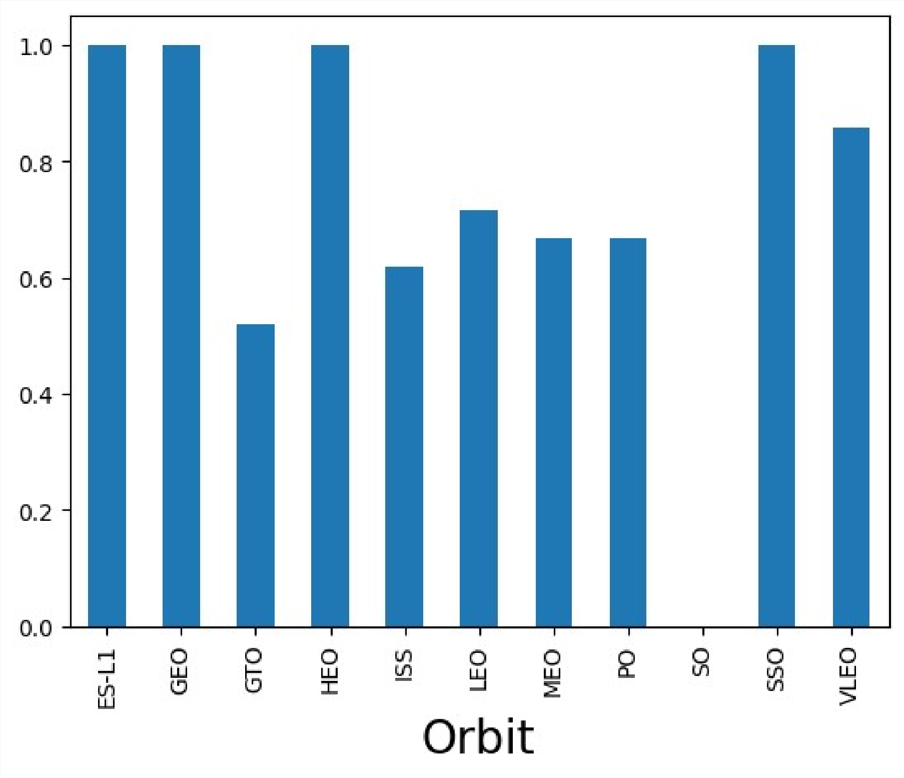
Class: 0 = Failure   Class: 1 = Success

Payload vs. Launch Site

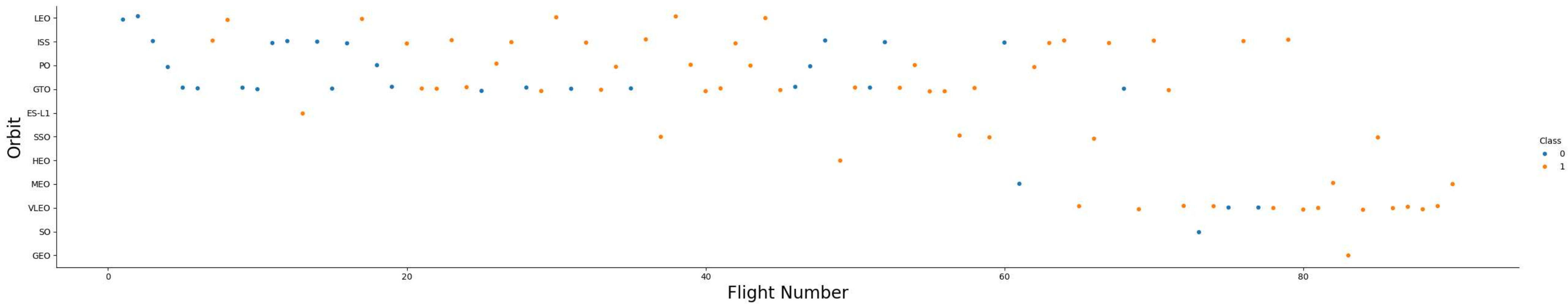


Class: 0 = Failure Class: 1 = Success

Success Rate vs. Orbit Type



Flight Number vs. Orbit Type

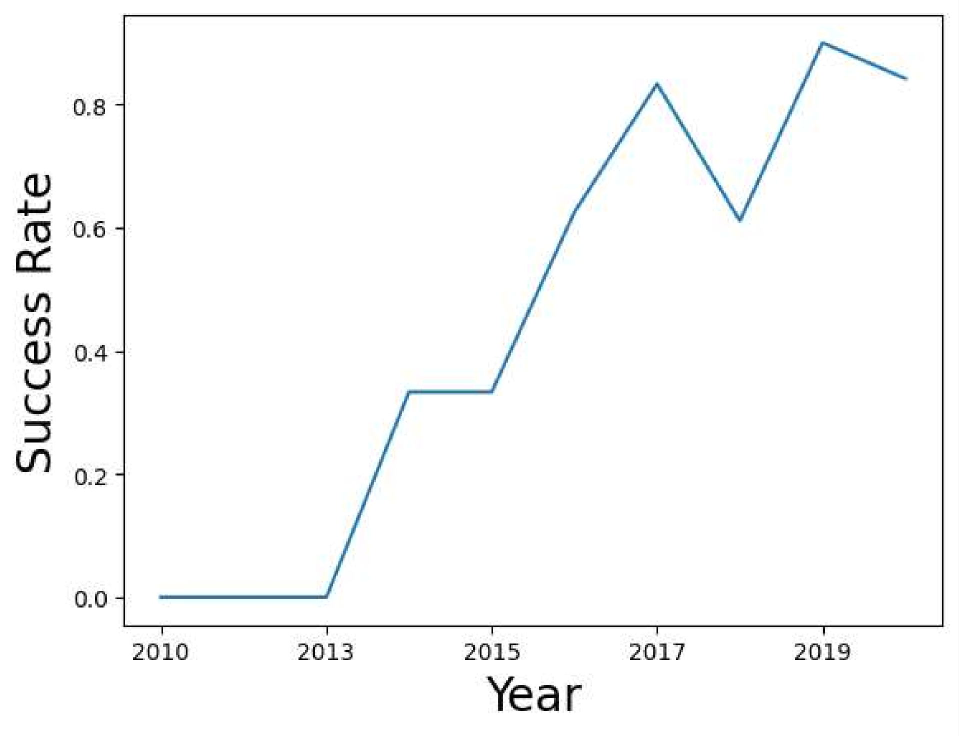


Class: 0 = Failure Class: 1 = Success

Payload vs. Orbit Type

Class: 0 = Failure   Class: 1 = Success

Launch Success Yearly Trend



All Launch Site Names

- Total of 4 launch sites

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

Launch Site Names 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	[bez názvu] 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 is 45,596 Kg

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass by F9 v1.1 (all v1.1s) is 2,534.67 Kg

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%';
```

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The first successful landing date was on 22nd December in 2015

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following booster versions successfully landed with payloads between 4,000 and 6,000 Kg

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

Total Number of Successful and Failure Mission Outcomes

- There were total of 100 successful missions and 1 failure

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

Boosters Carried Maximum Payload

- These booster versions 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

- In 2015 there were 2 failed landings on a drone ship. One in January and the other in April

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

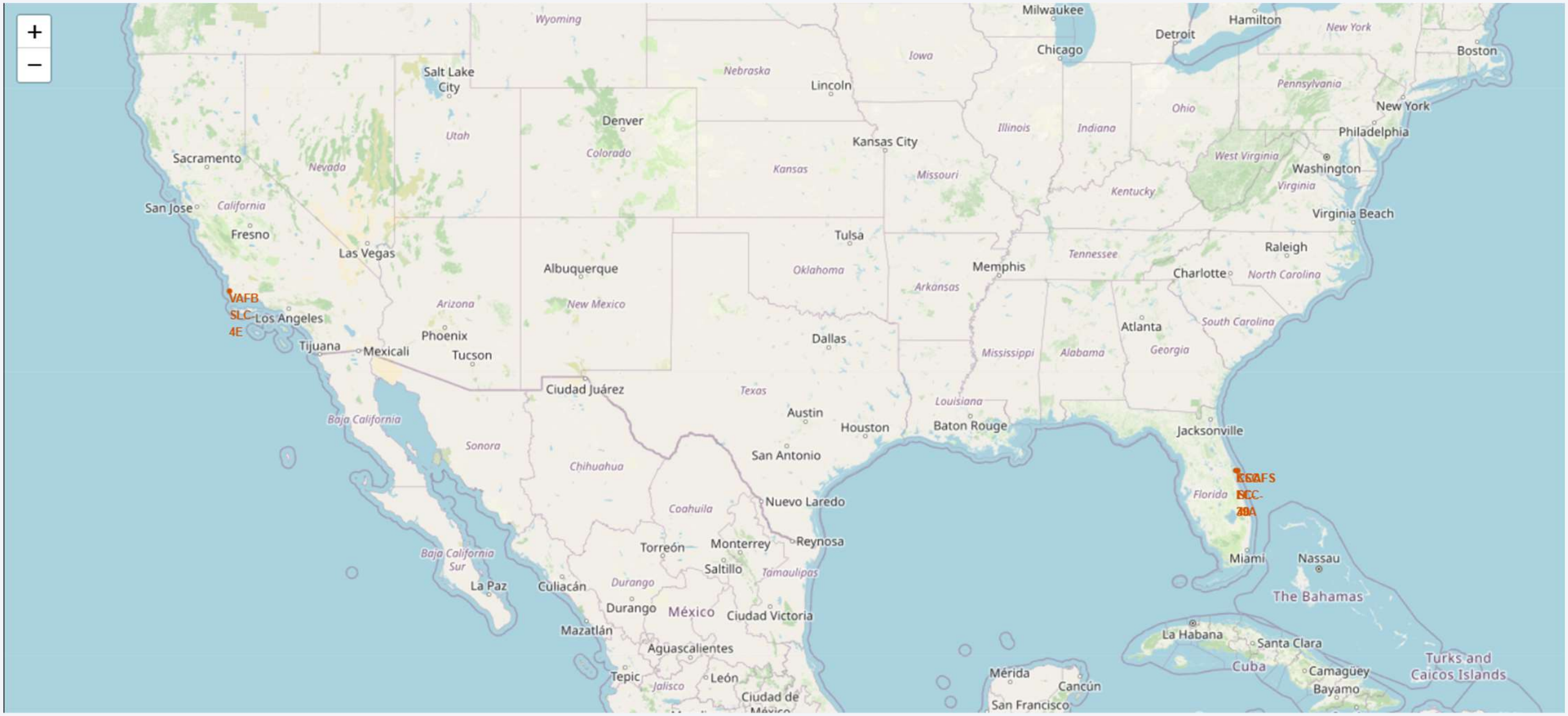
- Landing outcomes ranked:
- Most common was no attempt to land
- Least common was Precluded on drone ship

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

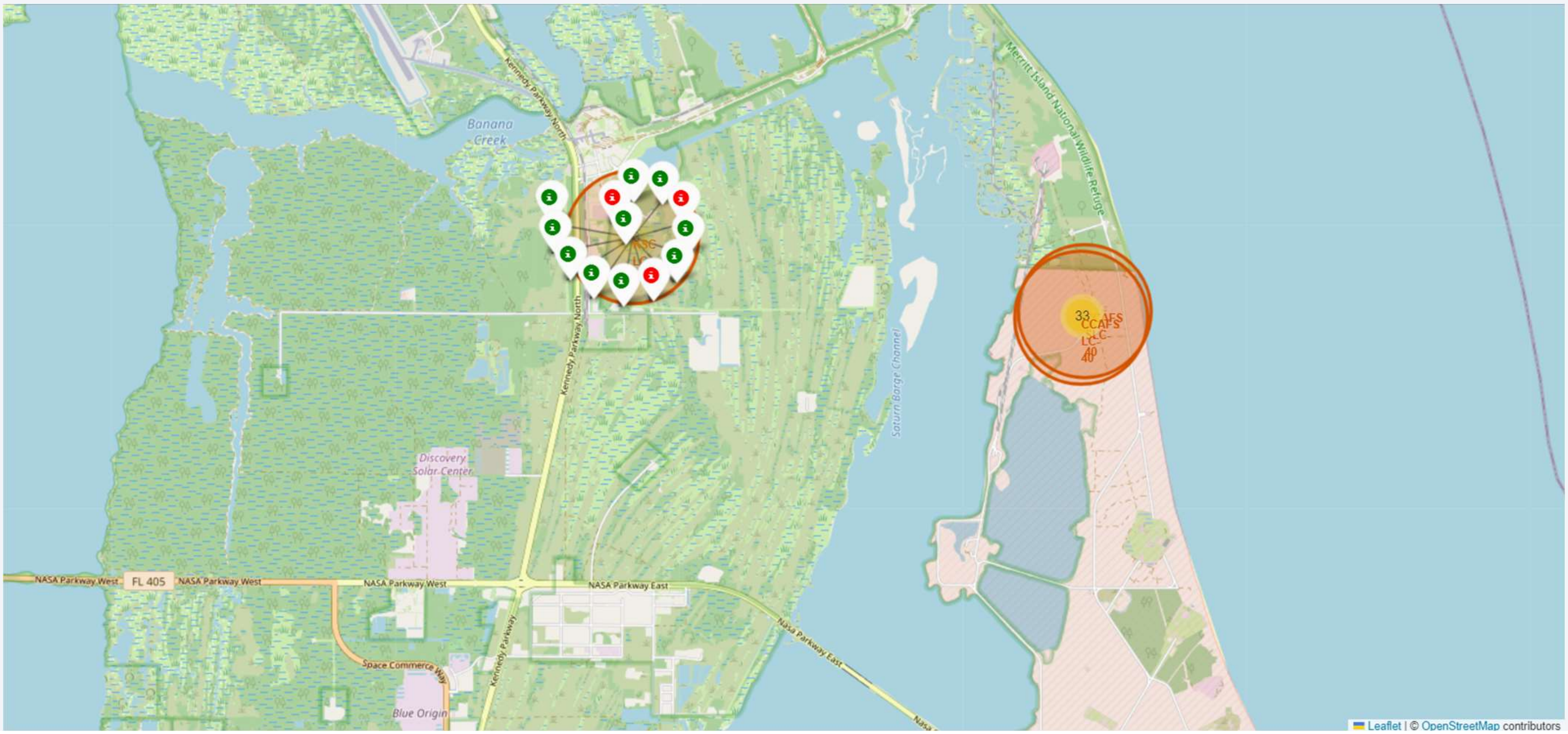
Launch Sites Proximities Analysis

Map of Launch Sites in the US

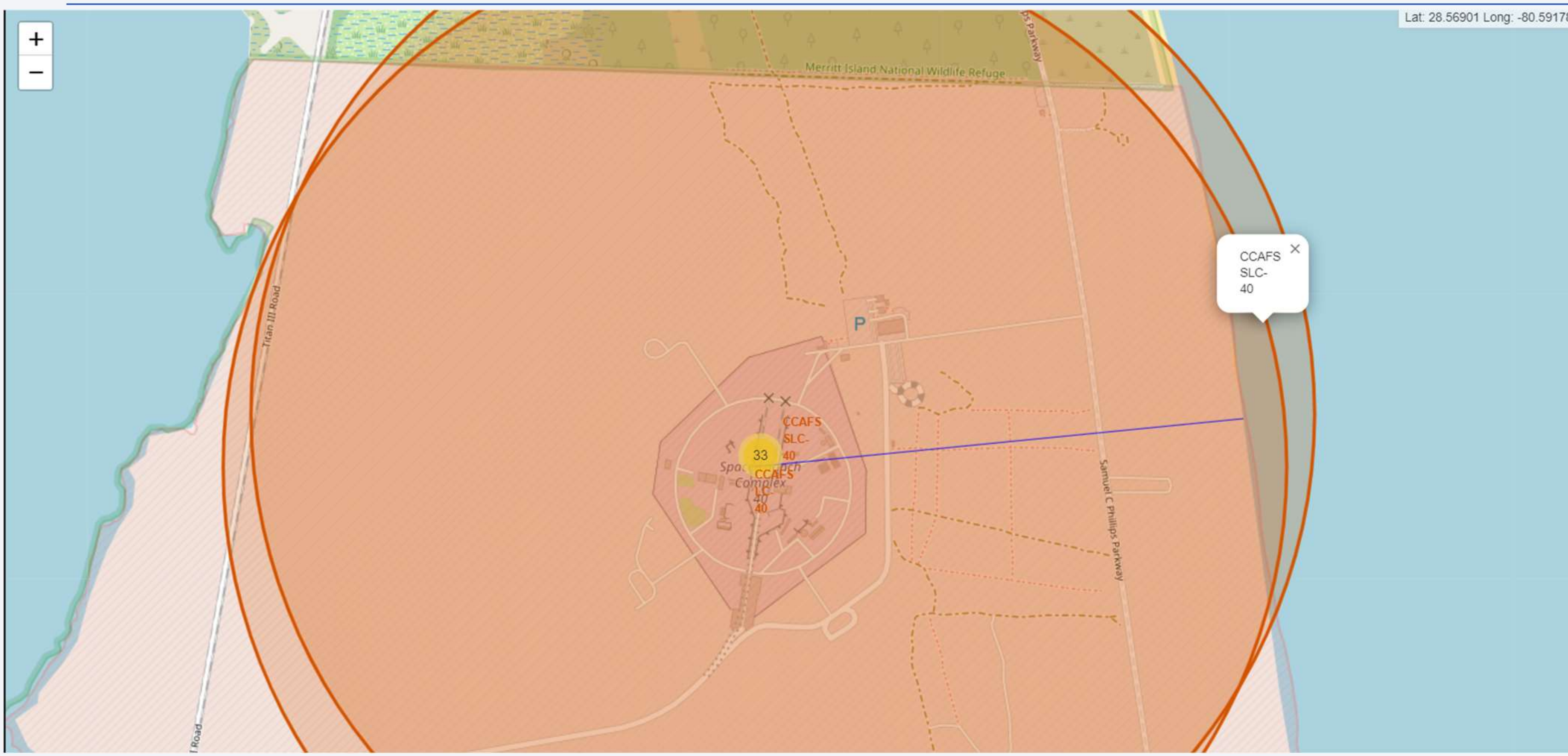


Launch Sites on the East Coast

- Launch outcomes at KSC LC-39A site:



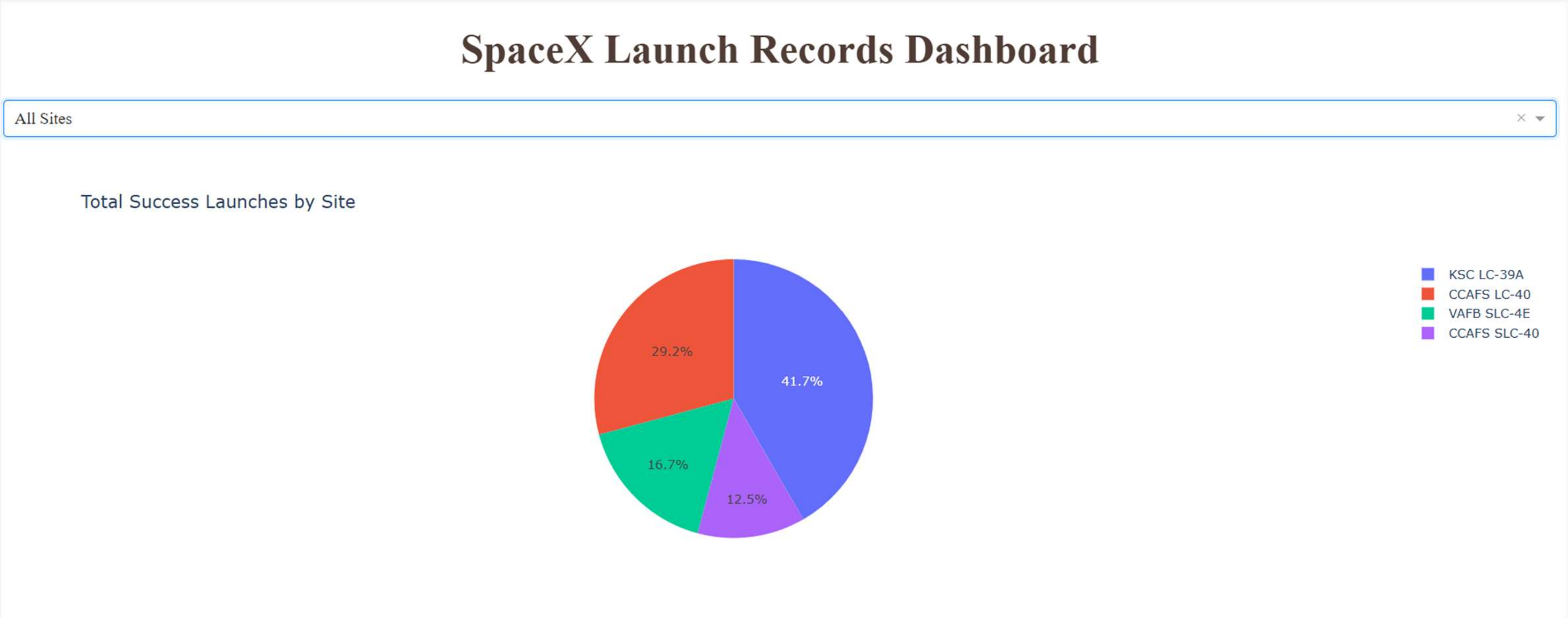
Nearest Coastline to Site CCAFS SLC-40



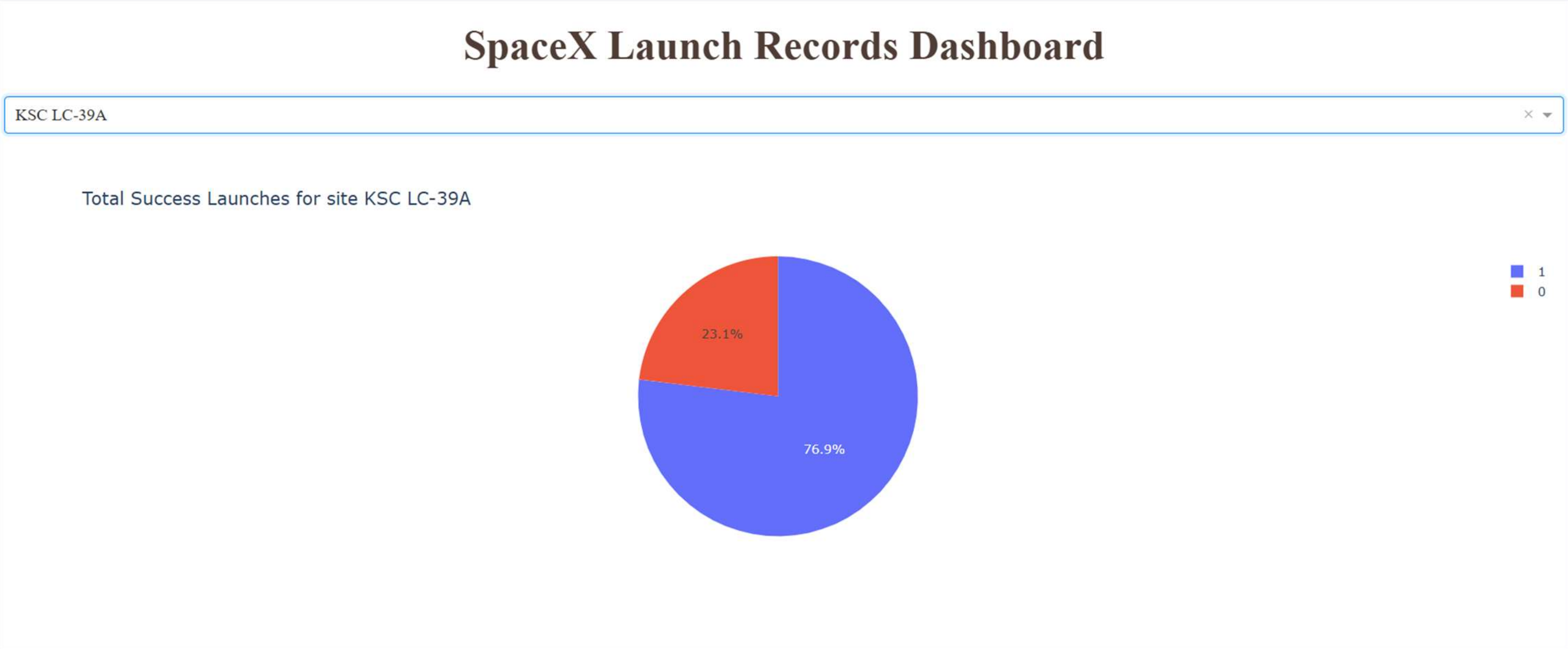
Section 4

Build a Dashboard with Plotly Dash

Successful Launches by Launch Site



Succes rate of KSC LC-39A



Correlation between Payload and Success

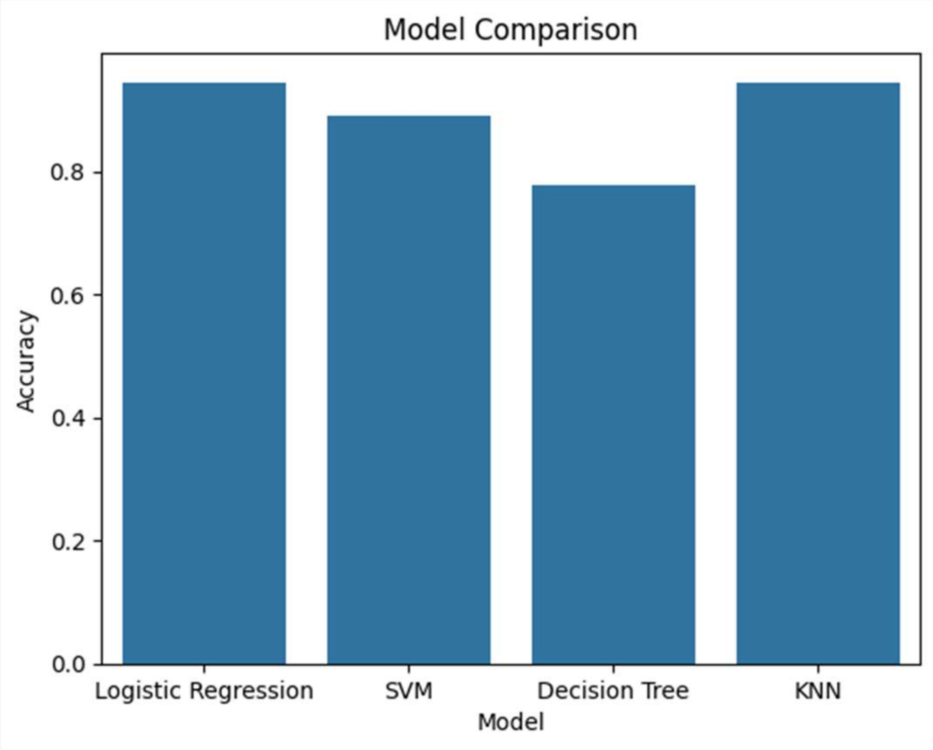
- For payload higer than 5,000 Kg only two booster versions attempted landings
- Only three landings were successful

Section 5

Predictive Analysis (Classification)

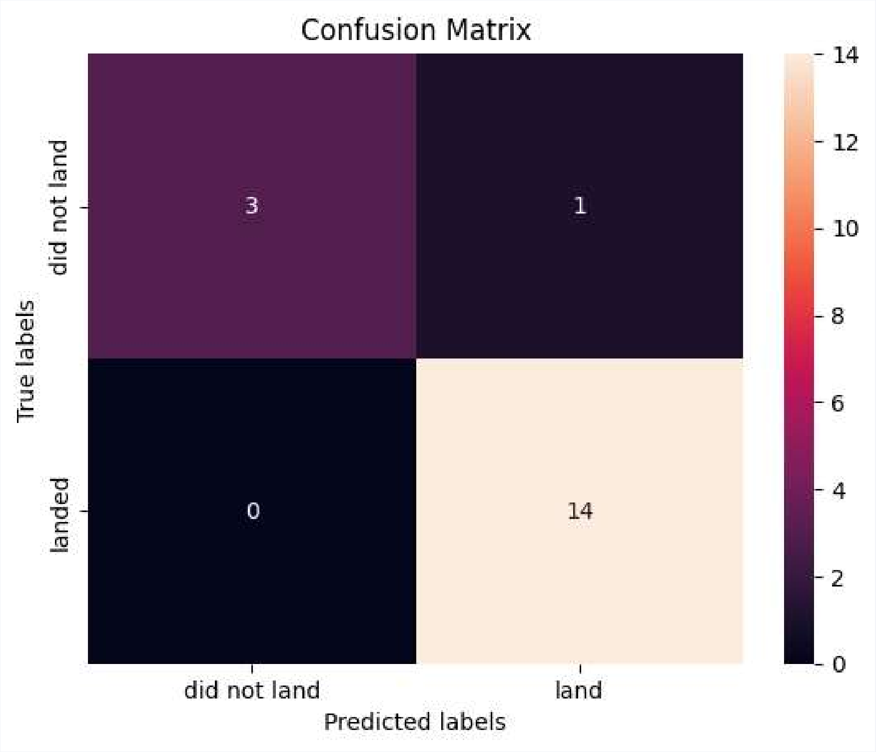
Classification Accuracy

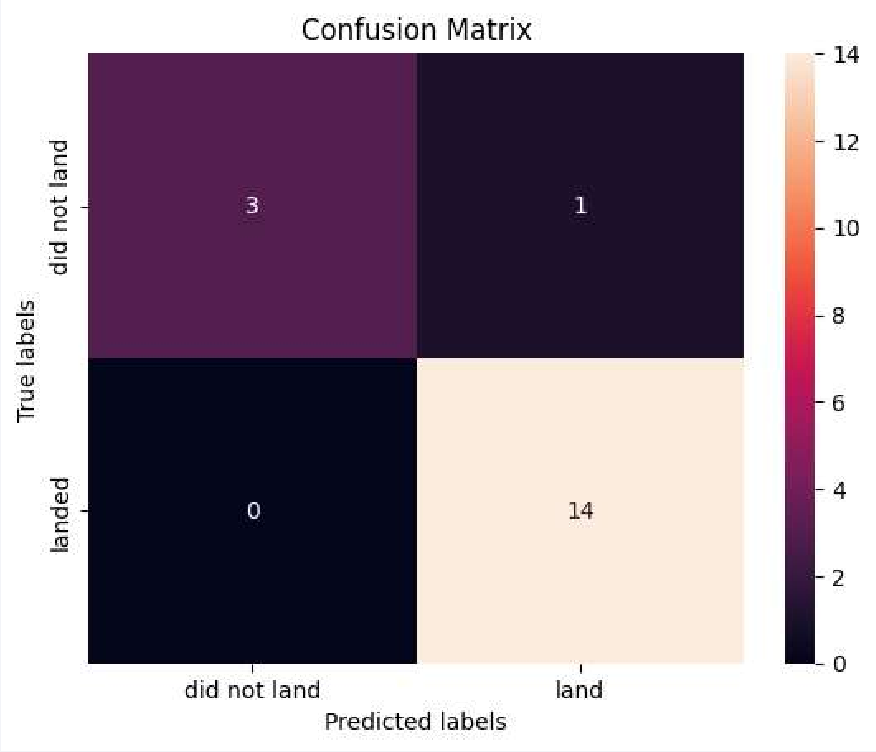
- Both KNN and Logistic Regression have the highest accuracy of 94 %



Confusion Matrix

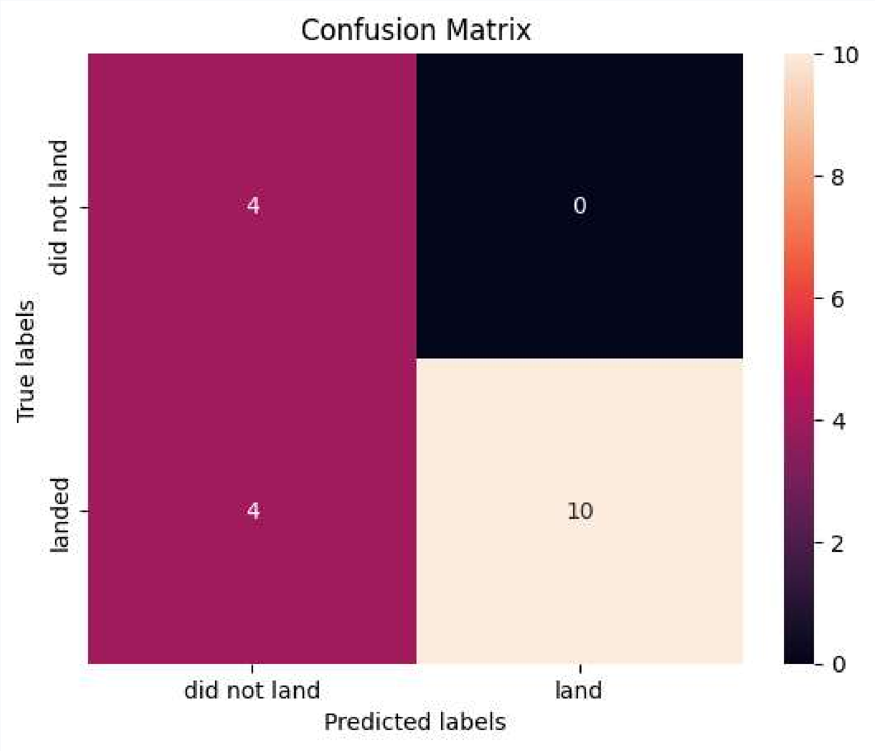
- Confusion matrices of KNN: and Logistic Regression:





Conclusions

- Logistic Regression and KNN are the best predictive models with 94 % accuracy
- Decision Tree is the worst predictor with 78% accuracy
- Decision Tree struggles to recognize some successful landings, but correctly recognizes all unsuccessful ones ->



Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project


Thank you!