IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Javier Castellanos
June 30, 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using SpaceX API and web scraping.
  - Exploratory Data Analysis (EDA) using data wrangling, data exploration using SQL, and data visualization and interactive visual analytics.
  - Machine Learning Prediction.
- Summary of all results
  - It was possible to collect necessary data.
  - EDA exposed which features are ideal to use to predict successful landings.
  - Machine Learning Prediction exposed the best classification model to use to predict which characteristics are most important to solve our problems using all collected data.

Introduction

- The project objective is to evaluate the viability for a new (imaginary) company SpaceY to compete with SpaceX.
- Desired answers
  - Find the best way to estimate the total cost of launches by predicting successful landings of the first stage of rockets.
  - Find where the best launch site to use is.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Web Scraping
- Perform data wrangling
  - Data was enriched by creating an outcome label based on outcome data found 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
  - Data was standardized, divided into training and test sets, and evaluated by four different machine learning classification models. Each model tested with different combinations of parameters to find best possible accuracy.

Data Collection

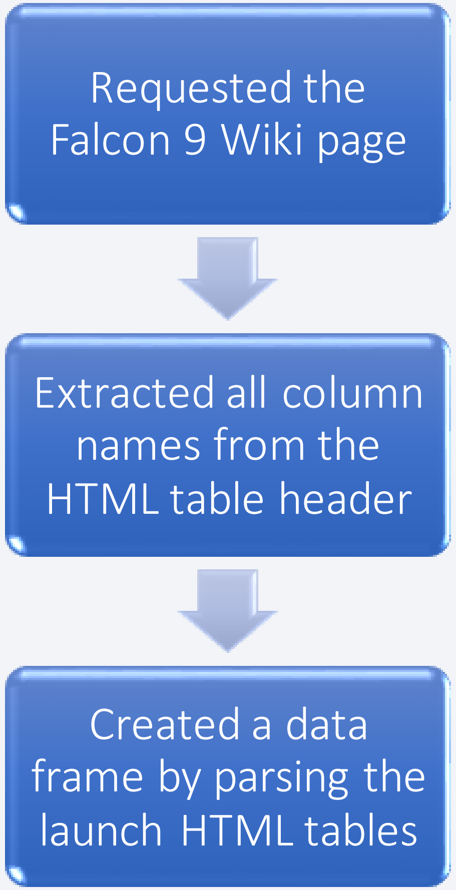
- Data was gathered directly from a SpaceX API and Wikipedia tables using web scraping.
  - SpaceX API: https://api.spacexdata.com/v4/rockets/
  - Wikipedia: https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained from and used.
- The API was used as shown by the flow chart to the right.
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/Data-Collection-API.ipynb

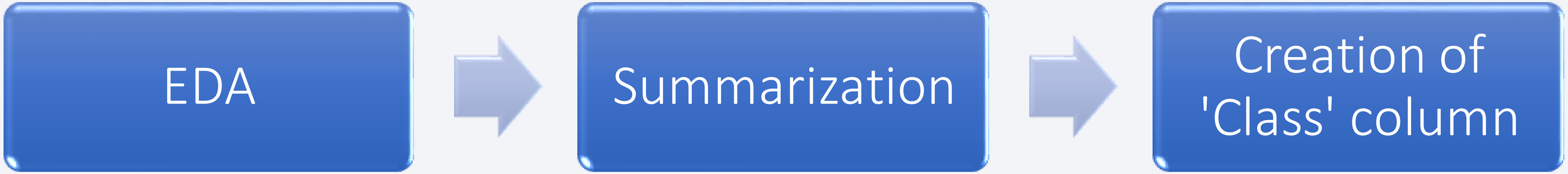
Data Collection - Scraping

- SpaceX data could also be downloaded from Wikipedia.
- The data was read from the website and handled as shown by the flow chart.
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/Data-Collection-With-Web-Scraping.ipynb



Data Wrangling

- Initial Exploratory Data Analysis (EDA) was performed on dataset.
- Calculated summaries of the dataset involving launches per site, occurrence of each orbit, and occurrence of mission outcome per orbit type.
- Classified data using 'Outcome' column showing whether or not a booster landed successfully or not.
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/Data-Wrangling.ipynb



EDA with Data Visualization

- Different combinations of feature sets were used to create multiple graphs:
- Graphs used:
  - Scatter Plot, Bar Chart, Line Chart
- Features used:
  - Flight number, Payload mass, launch site, orbit, success outcome, success rate, year
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/EDA-Using-Pandas-and-Matplotlib.ipynb

EDA with SQL

- SQL queries performed:
  - Names of the unique launch sites in the space mission
  - 5 records where launch sites 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 with success in drone ship and have payload mass greater than 4000 but less than 6000
  - 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 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 Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/EDA-Using-SQL.ipynb

Build an Interactive Map with Folium

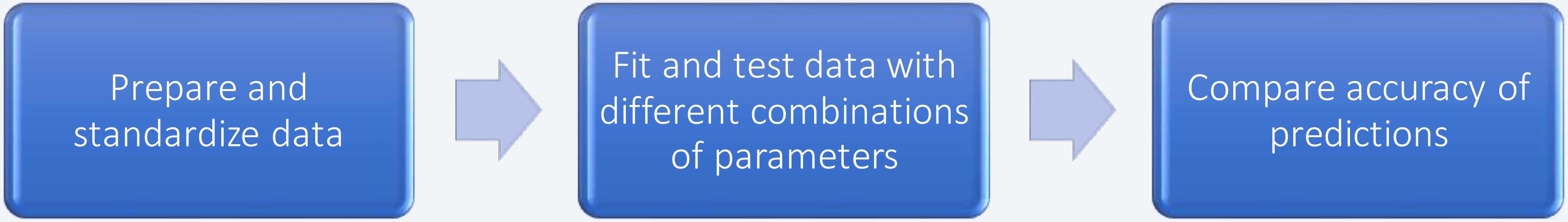
- Markers, circles, lines, and marker clusters were using with Folium Maps
  - Markers were used to indicate launch sites
  - Circles were used to highlight area around launch sites
  - Lines were used to indicate distance between launch sites and nearest city, railway, coastline, and highway.
  - Marker clusters were used to group successful landings in specific coordinates.
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/Interactive-Visual-Analytics-and-Dashboard.ipynb

Build a Dashboard with Plotly Dash

- The following interactive graphs and plots were used to visualize data
  - Pie Chart – Total Successful Launches by Site
  - Scatter Plot – Correlation Between Payload and Success by Launch Site
- These plots showed us the relationship between payload and launch sites, showing us the ideal site to use according to payload amount in order to replicate successful landings.
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/SpaceX_Dash_App.py

Predictive Analysis (Classification)

- In total, four classification models were compared:
  - K-Nearest Neighbors, Decision Tree, Support Vector Machine, Logistic Regression
- Source Code: https://github.com/JavierACM/IBM-Applied-Data-Science-Capstone/blob/master/Machine-Learning-Prediction.ipynb

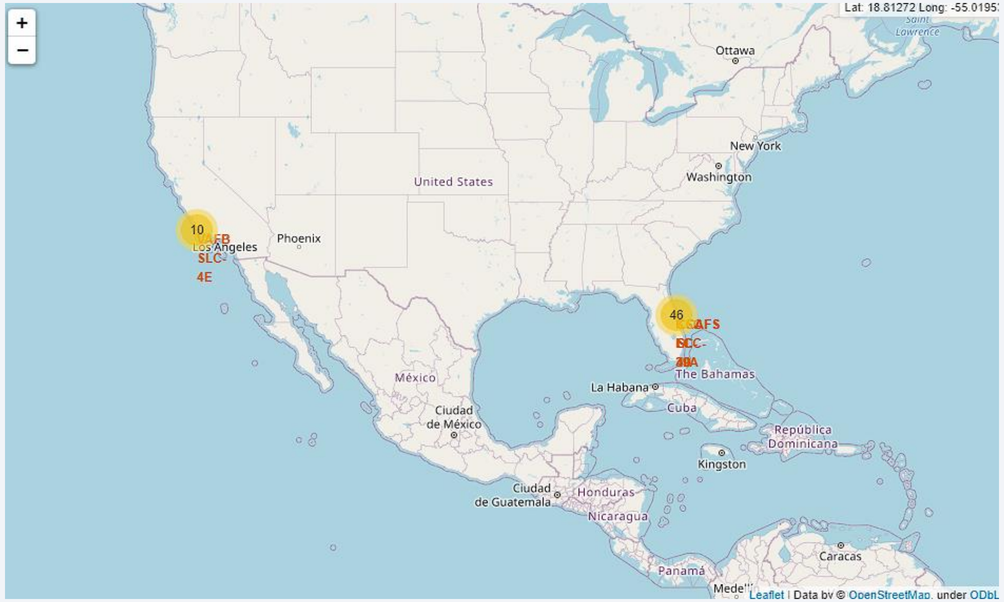


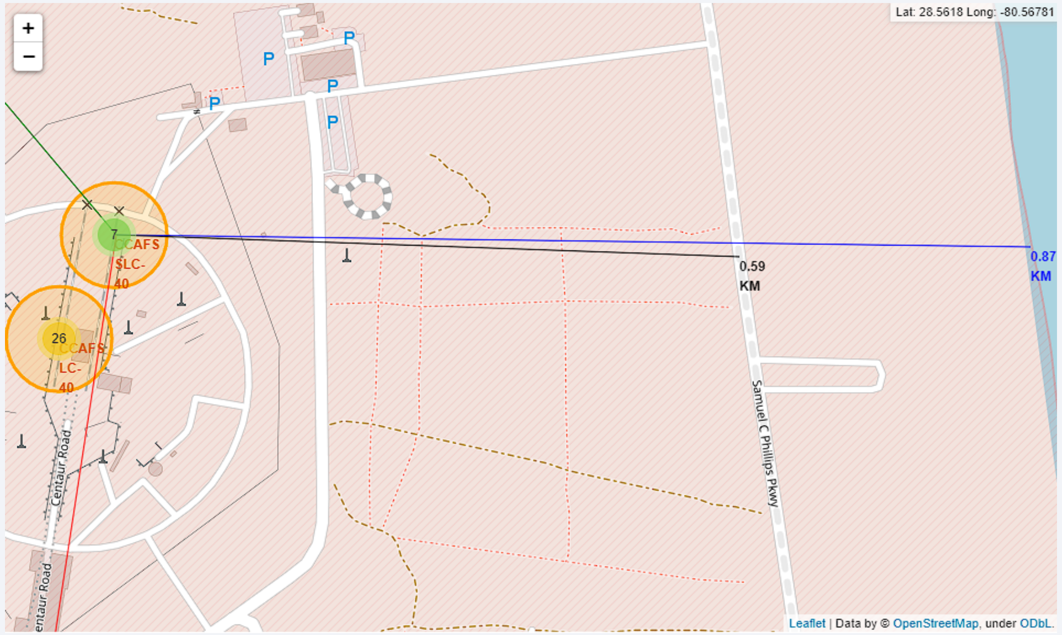
Exploratory Data Analysis Results

- SpaceX uses 4 different launch sites
- The average payload carried by booster version F9 v1.1 is 2,928 kg
- First successful landing outcome achieved about 5 years after first successful launch
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
- Mission outcomes were almost 100% successful; only 1 mission failed
- In 2015, F9 v1.1 B1012 and F9 v1.1 B1012 failed to land on drone ship
- Landing outcome improved with time

Interactive Analytics Demo

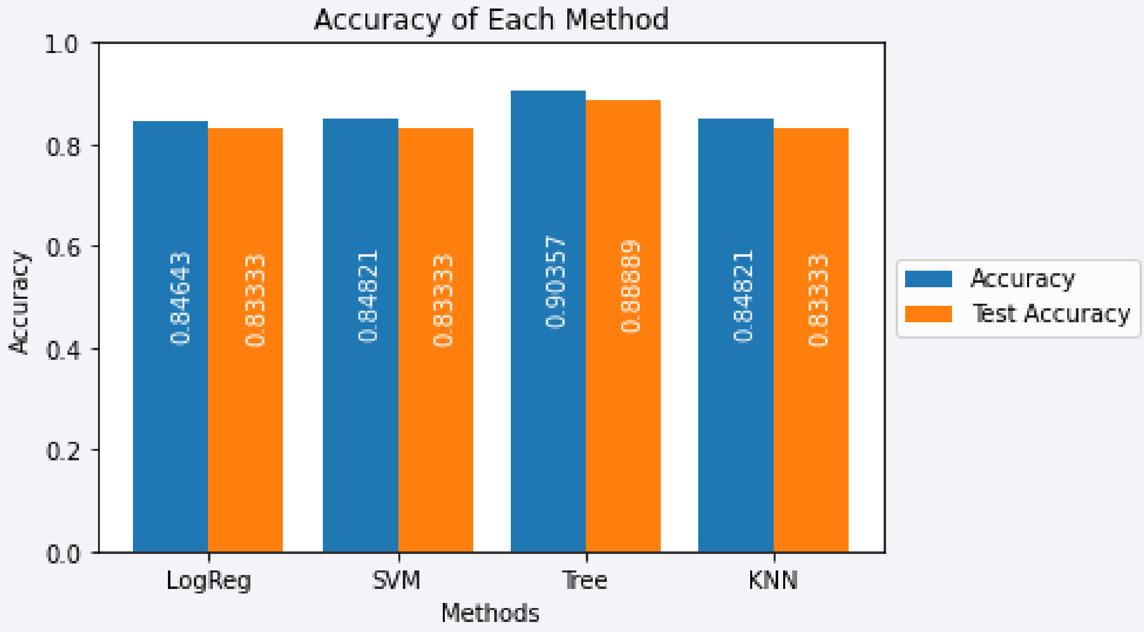
- Launch sites are located close to coastal areas far away from cities and are in southern United States
- Most launches happened on the east coast launch sites





Predictive Analysis Results

- After testing and comparing different models, Decision Tree Classification resulted being the best model to use for prediction as it had both training and testing accuracy of over 88%.

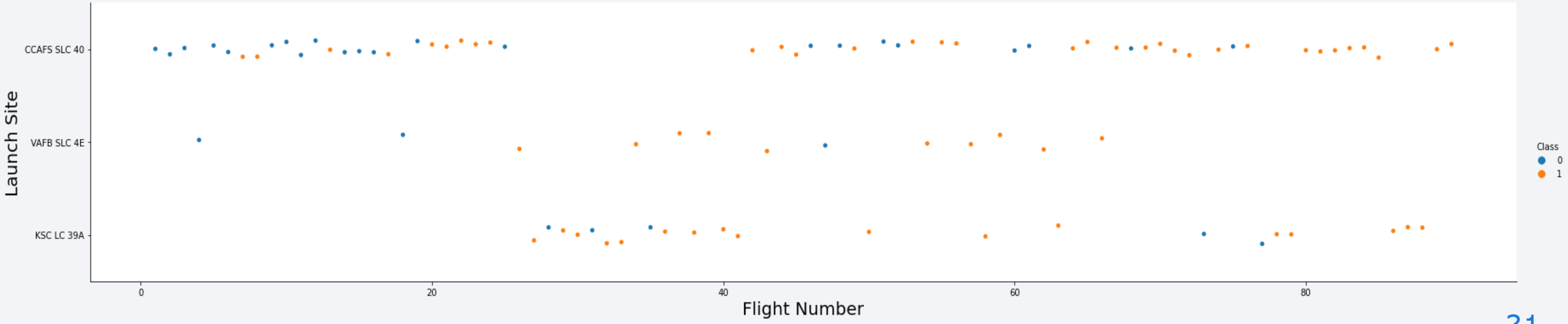


Section 2

Insights drawn from EDA

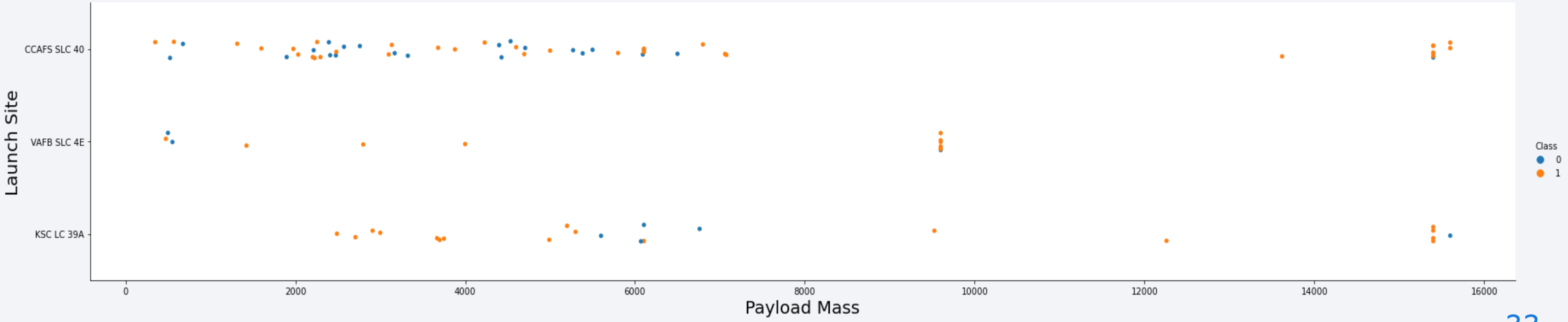
Flight Number vs. Launch Site

- Launch Site CCAFS SLC 40 is the current best site to use as it has the most recent successful launches.
- Generally speaking, the success rate of landing improved over time.



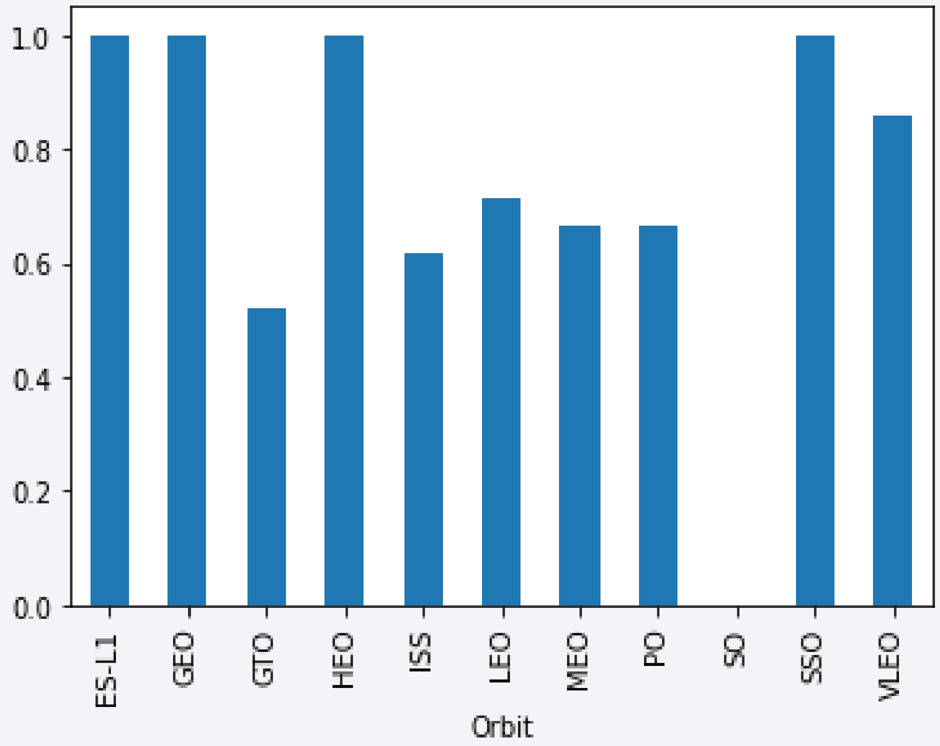
Payload vs. Launch Site

- Payloads over 7,000kg have near perfect success rates.
- It seems to be impossible (or yet to be done) to launch above a payload mass of 10,000kg at launch site VAFB SLC 4E.



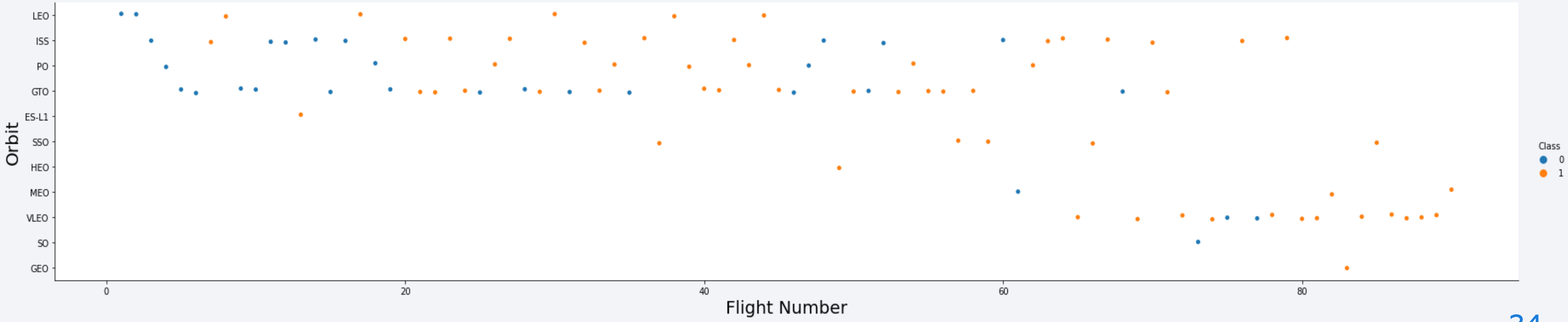
Success Rate vs. Orbit Type

- Orbits with perfect success rates include:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LEO (above 70%)



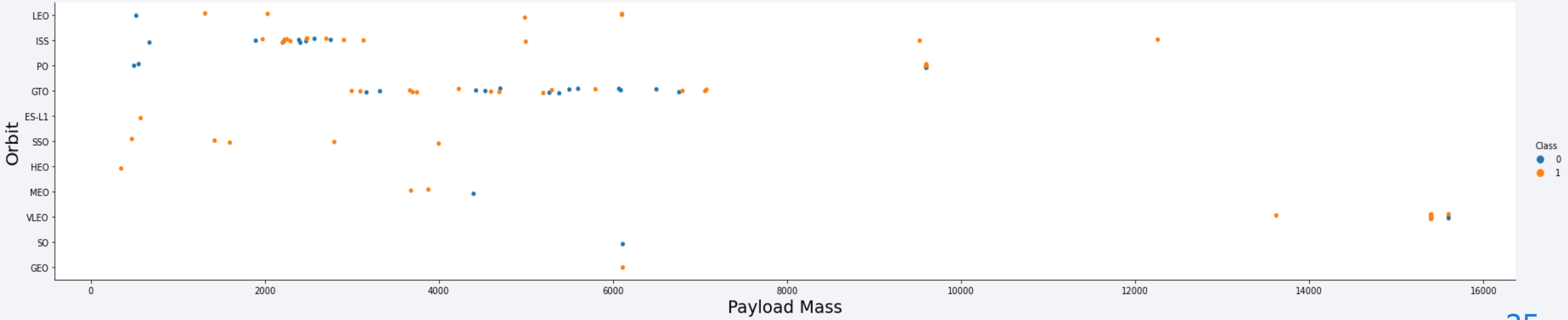
Flight Number vs. Orbit Type

- Success rate improved over time to all orbits.
- VLEO seems to currently be the most optimal orbit to launch to, as the most recent launches tend to favor that orbit over others.



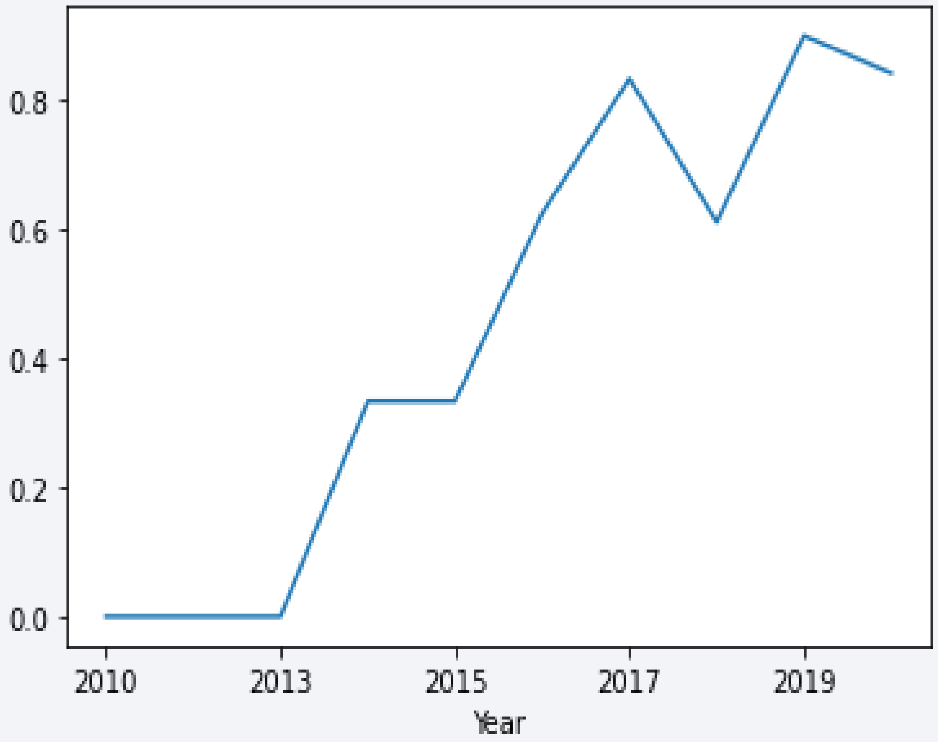
Payload vs. Orbit Type

- Orbit GTO seems to not have a correlation between payload mass and success rate.
- There seems to be only 1 single launch to each SO and GEO.
- Orbit VLEO seems to be the go-to orbit for almost all payloads over 10,000kg.
- Orbit ISS seems to have a wide range of payloads with a decent success rate.



Launch Success Yearly Trend

- It is clear that success rate increased over time.
- There was a dip in the success rate in 2018 and 2020.



All Launch Site Names

- In total, there are 4 launch sites:

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

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- This number represents the total payload carried by boosters from NASA.
- NASA has code 'CRS' under payload column.

Total Payload (kg)
111,268

Average Payload Mass by F9 v1.1

- This number represents the average payload mass carried by ONLY booster version F9 v1.1.

Average Payload (kg)
2928

First Successful Ground Landing Date

- This date represents the first successful landing on a ground pad (gp).
- December 22, 2015

First Successful GP
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4,000kg but less than 6,000kg.

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

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes.

Mission Outcome	Total
Success	99
Failure (in flight)	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- This is a list of boosters which have carried the maximum payload mass.

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

- The following are the boosters which failed to land on a drone ship in 2015.

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- The ranking of all landing outcomes between June 4, 2010 and March 20, 2017.

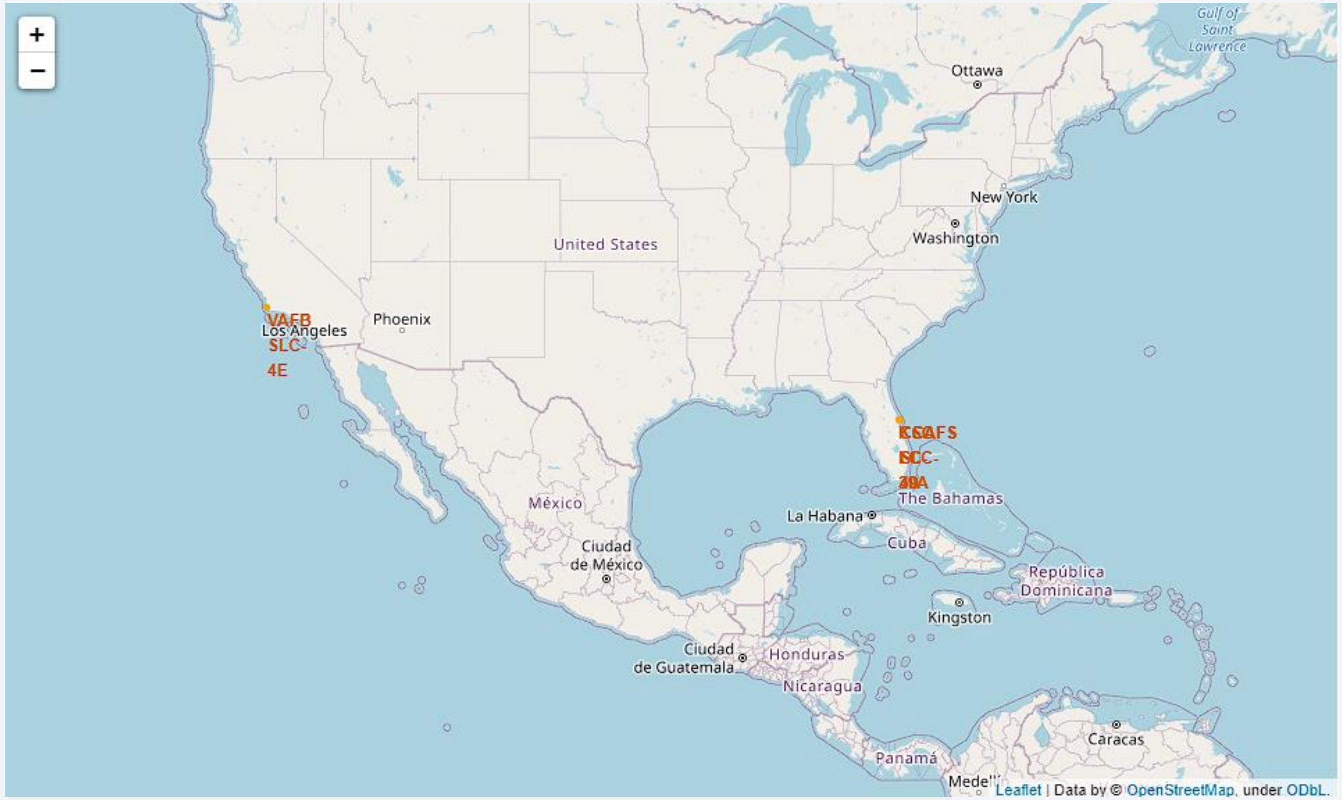
Landing Outcome	Total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

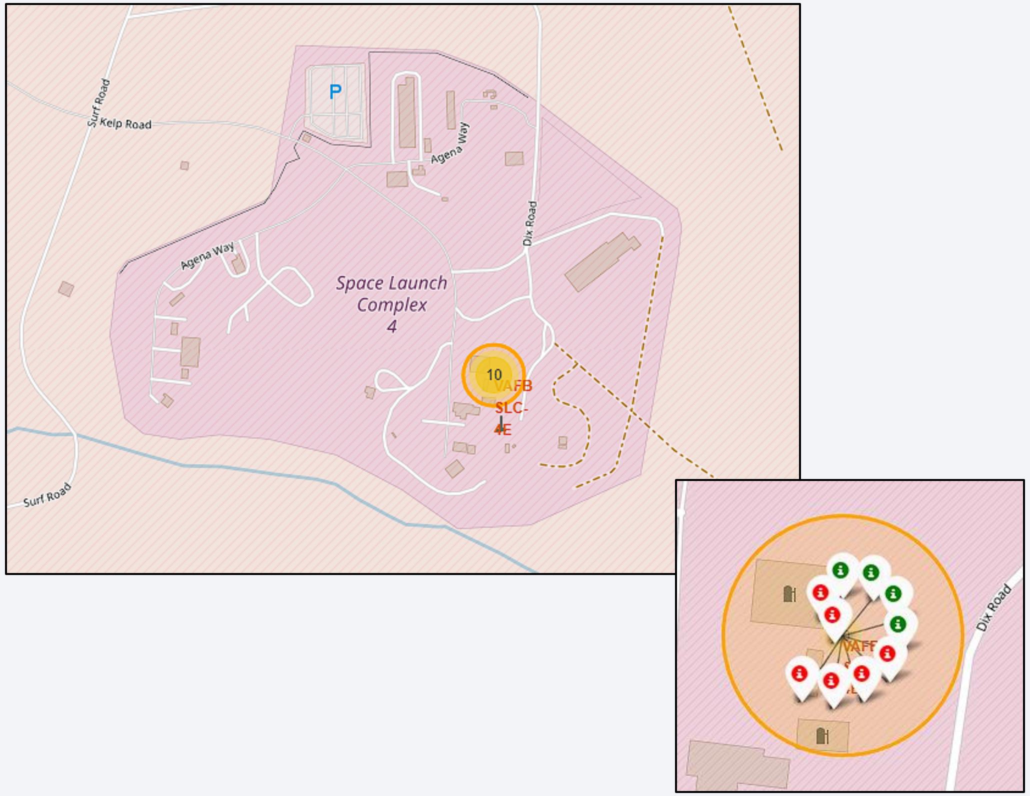
All Launch Sites

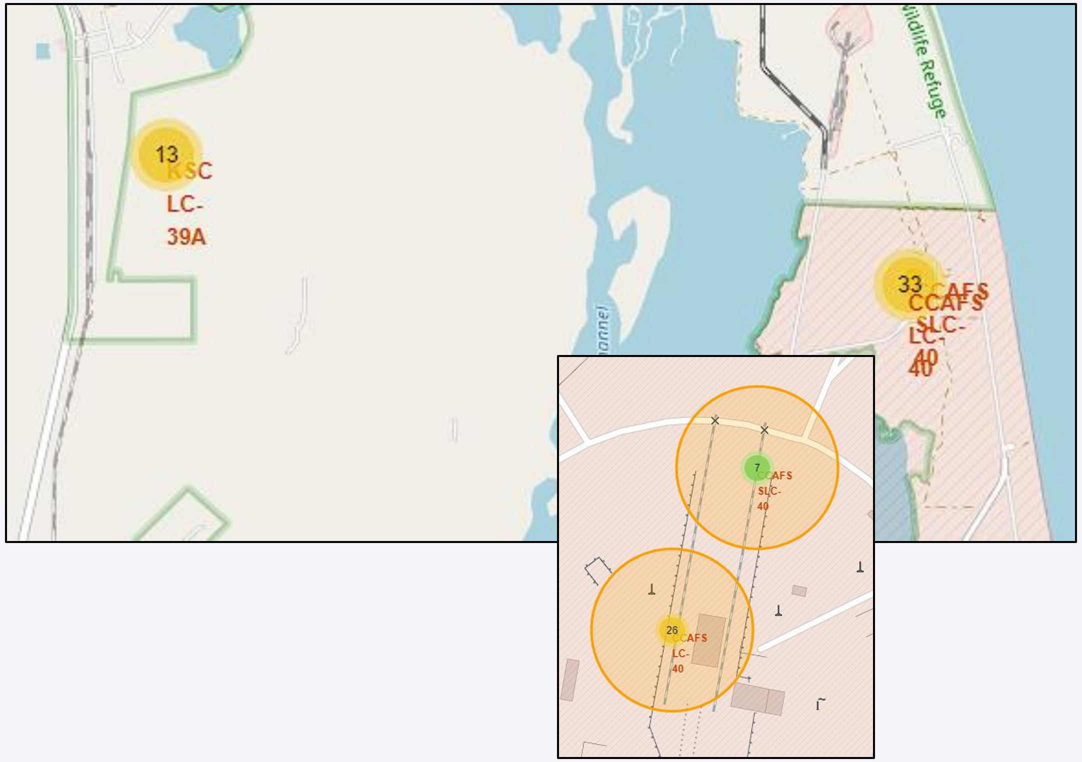
- All launch sites are near sea.



Launch Outcomes per Site

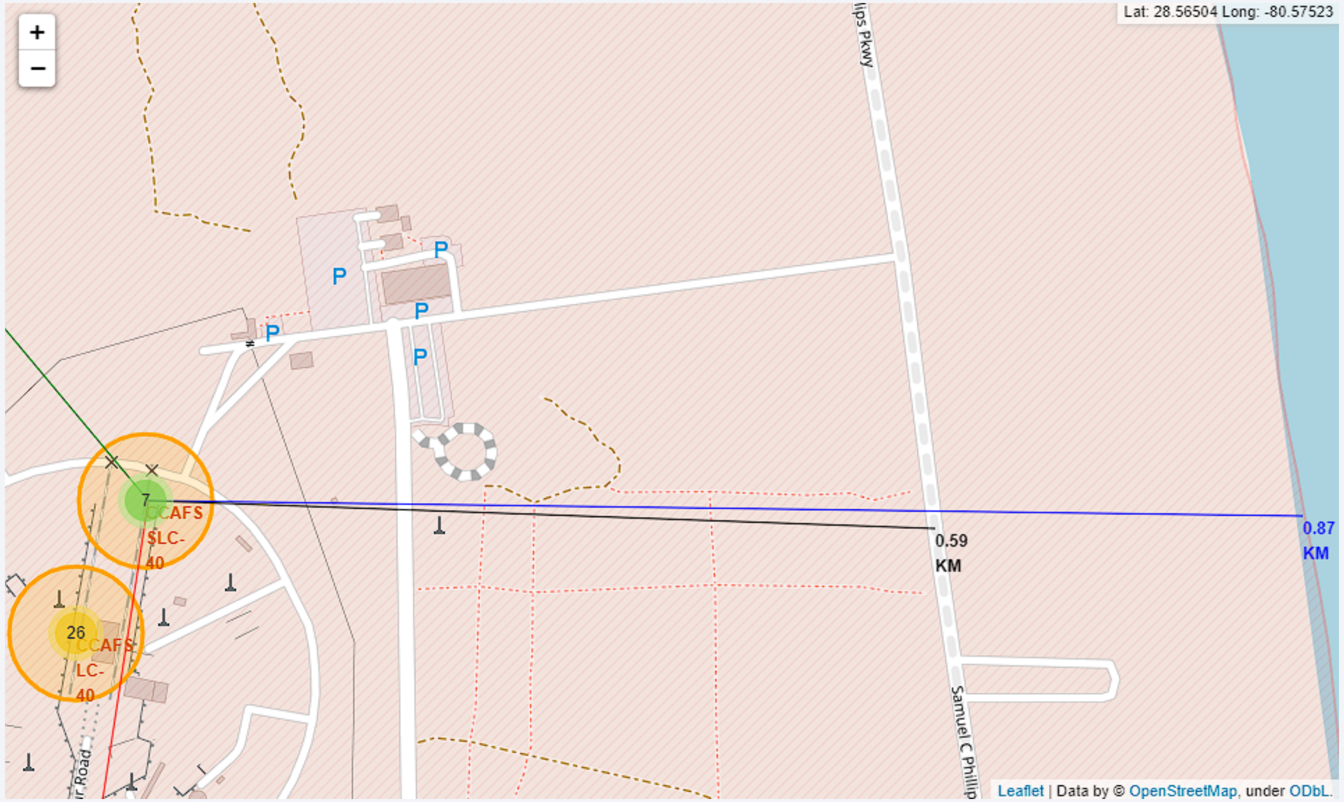
- Numbers represent number of landing outcomes per site.
- Green markers represent successful and red represent failed launch outcomes.





Logistics & Safety

- Launch site CCAFS SLC-40 has good logistics aspects being near coast (0.87km), road (0.59km), and railroad (1.29km), while also being relatively far away from the nearest city (18.16km).

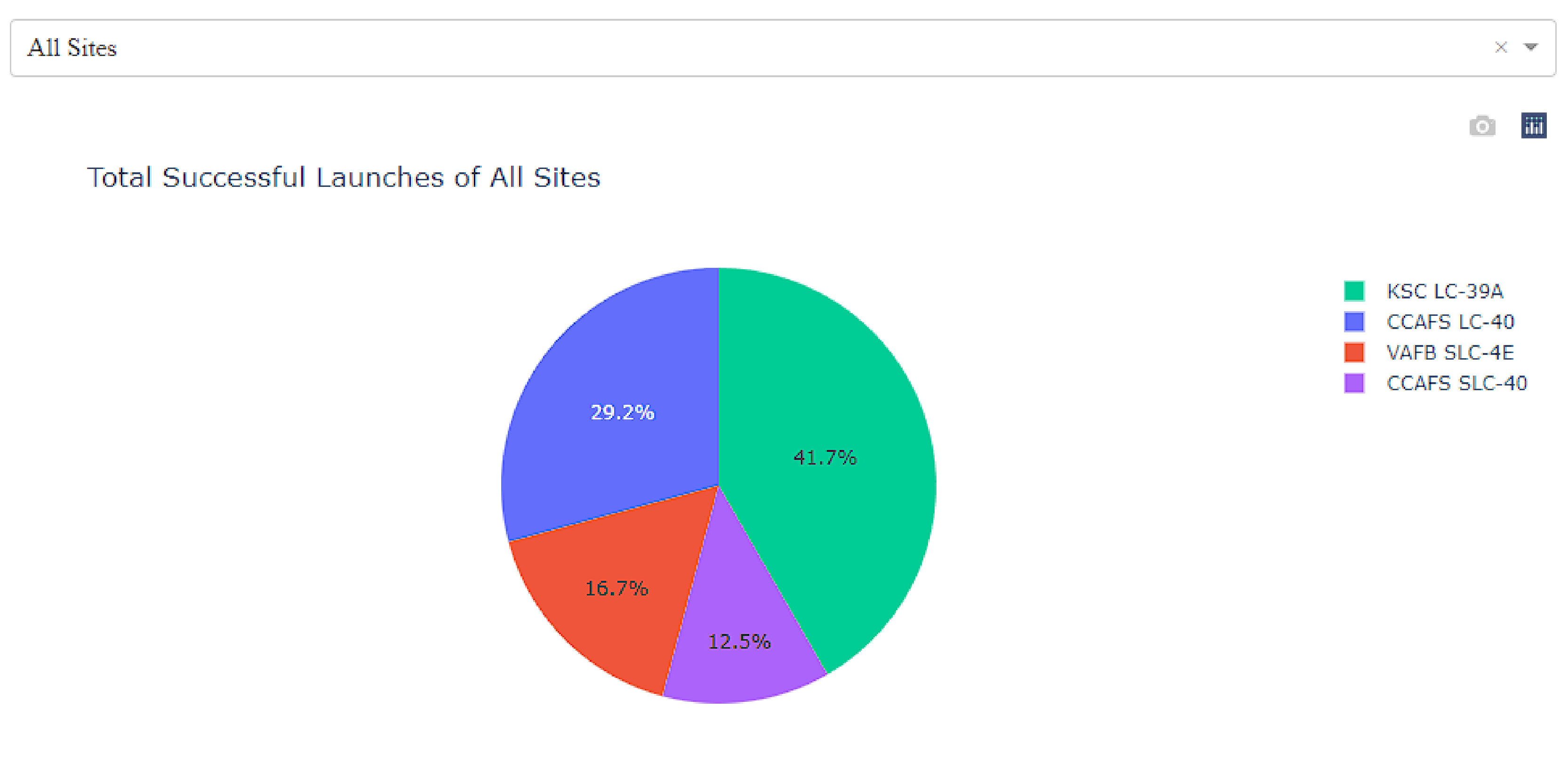


Section 4

Build a Dashboard with Plotly Dash

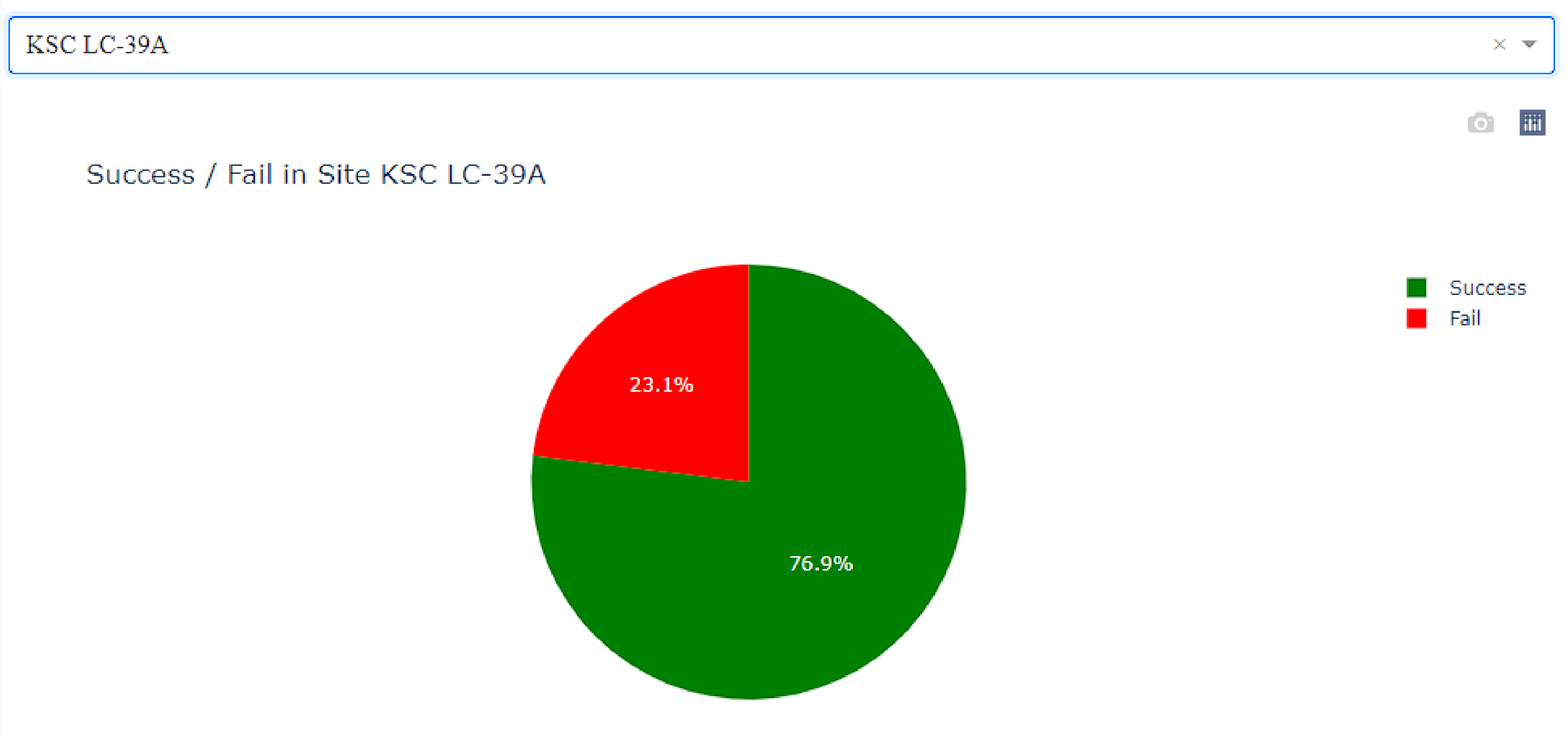
Successful Launches for All Sites

- Launch site seems to be an important factor contributing to the success of a mission.



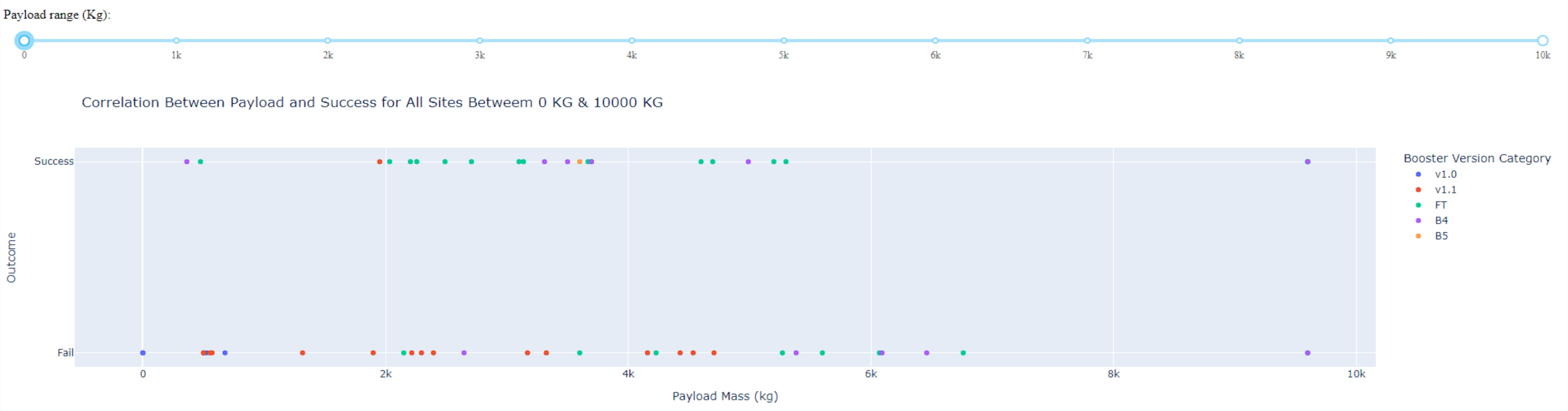
KSC LC-39A Success Ratio

- Success –10 Fail – 3
- This is the only launch site with a success ratio above 50%.



<Dashboard Screenshot 3>

- Payloads under 6,000kg and booster version FT have the largest success rates.

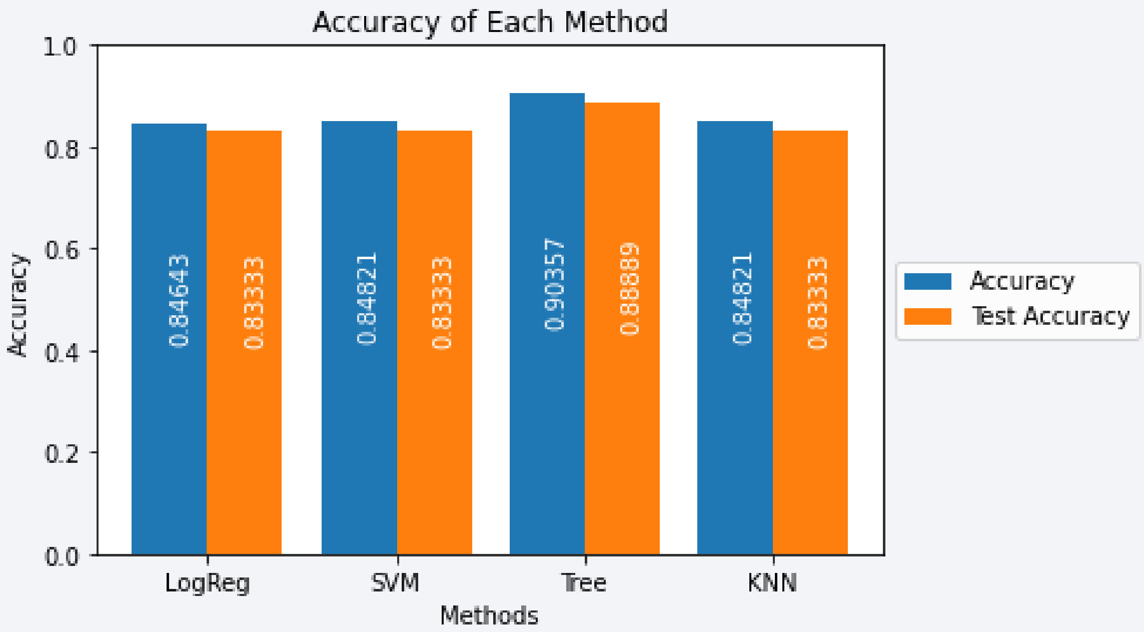


Section 5

Predictive Analysis (Classification)

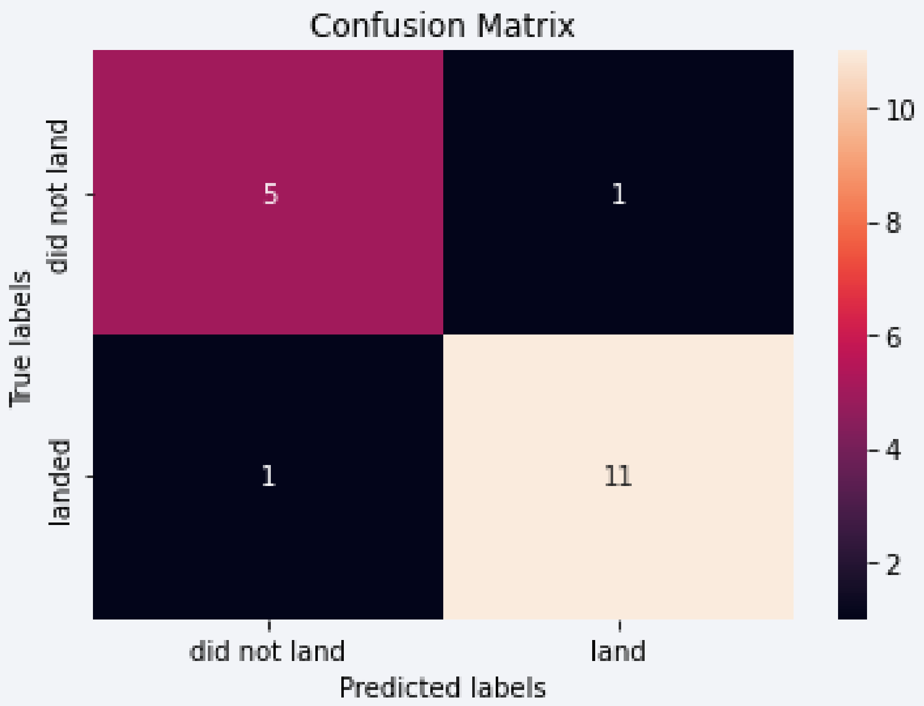
Classification Accuracy

- Four total classification models were tested for accuracy.
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbors
- Decision Tree Classifier is the model with the highest test accuracy (~89%).



Confusion Matrix

- This proves the accuracy of the Decision Tree Classifier, comparing the number of true negatives and true positives cases with false ones.



Conclusions

- The best launch site is KSC LC-39A.
- Launching over a payload mass of 7,000kg has a low chance of failure.
- According to the latest launches, the VLEO orbit seems to be the most favorable orbit to launch to.
- Overall launch success seems to progressively improve over time.
- Decision Tree Classifier should be the ideal classification model to predict successful landings and increase profits.

Appendix

- GitHub did not display maps, so instead I used nbviewer.org with the URL of my notebook to view my maps.
- Random accuracies kept appearing for decision tree model, so to combat this issue, I included `np.random.seed(10)` in my code.

Thank you!