Winning Space Race with Data Science

Eric Szumlas
5/28/2022

Outline

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

Executive Summary

- Data on past launches were collected from SpaceX API's and Wikipages
- Exploratory Data Analysis was performed to examine relationships between success rates, flight number, payload mass, orbit, launch site, booster rocket version, core data, and yearly trends.
- Models were created to predict success outcomes of future launches.
- We can predict the success outcome for a launch with a good amount of accuracy.
- As the number of launches increased, so did the success outcomes. This shows that future outcomes are predicted to be better than past ones.
- Launches with heavier payload masses have a much higher success rate.
- Launches for orbits ES-L1, GEO, HEO, SSO have the best outcome for success.

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- We will predict if the Falcon 9 first stage will land successfully.

Section 1

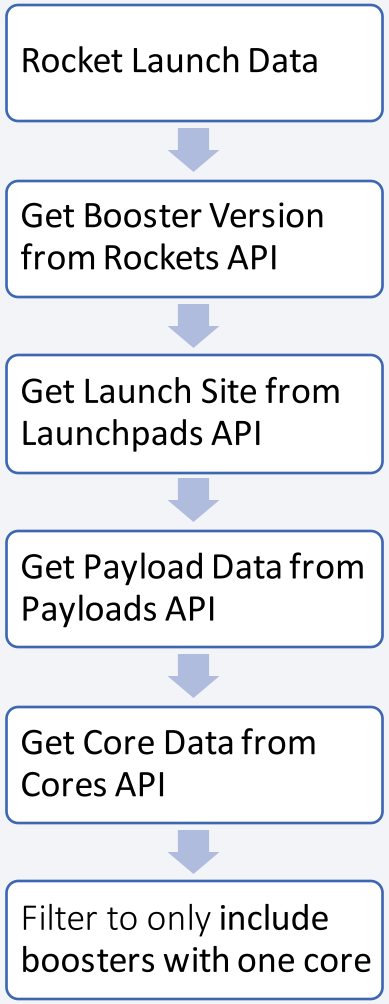
Methodology

Data Collection – SpaceX API

Data was collected from the SpaceX API's.

`The various API's were located at https://api.spacexdata.com/v4/`

[SpaceX API calls notebook]



Data Collection - Scraping

Falcon 9 launch records were extracted from HTML tables from Wikipedia

Wikipedia page
https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Web scraping notebook

Falcon9 Launch Data URL page

↓

Extract column/variable names from tables

↓

Create data frame by parsing launch HTML tables

EDA with Data Visualization

The following were explored to see how they would affect the launch outcome

- Flight Number (indicating the continuous launch attempts.) was plotted against Payload
- Flight Number was plotted against Launch Site
- Payload was plotted against Launch Site
- Success Rate of each Orbit Type visualized as a bar graph
- Flight Number was plotted against Orbit Type
- Payload was plotted against Orbit Type
- Launch Success yearly trend

EDA with data visualization notebook

EDA with SQL

SQL queries were performed on the dataset which was moved to a Db2 server

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites begin with the string 'CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome in ground pad was achieved
- Listed the names of the boosters which have success in drone ship and had payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of the booster_versions which have carried the maximum payload mass using a subquery
- Listed the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranked 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

EDA with SQL notebook

Build an Interactive Map with Folium

Mark all launch sites on a map

- Used circle and marker for each site

Mark the success/failed launches for each site on the map

- Used a Marker Cluster for each site

Calculate the distances between a launch site to its proximities

- Exercise to demonstrate the use of a line and distance to markers for nearby objects
  - Railway
  - Highway
  - City

Interactive map with Folium notebook

Build a Dashboard with Plotly Dash

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.

- Launch site drop-down input component
- Callback function to render pie chart based on selected site dropdown showing success rates
- Range slider to select payload
- Callback function to render the scatter plot showing success rates

Plotly Dash lab

Predictive Analysis (Classification)

Perform exploratory Data Analysis and determine Training Labels

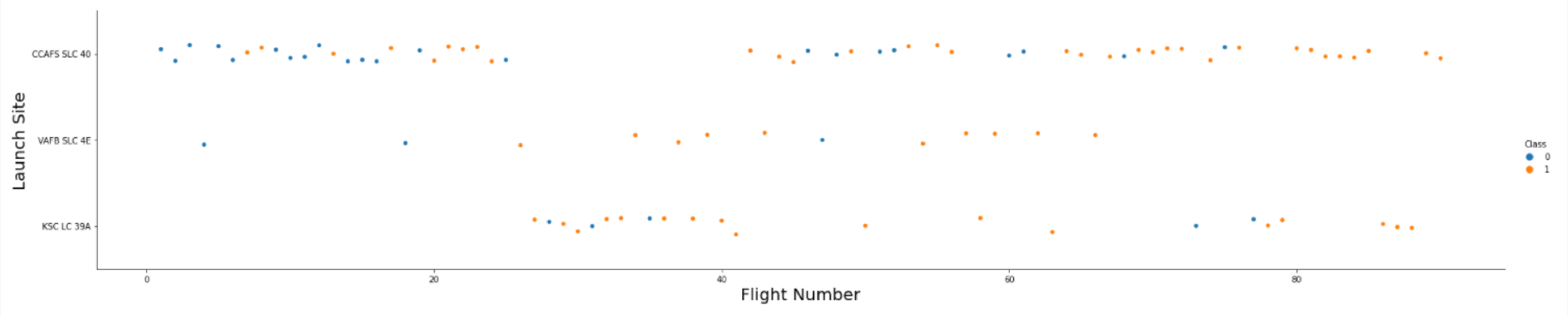
Created a column for the class → Standardized the data → Split into training data and test data → -Found best Hyperparameter for SVM, Classification Trees and Logistic Regression → Found the method that performs best using test data

Predictive analysis lab

Section 2

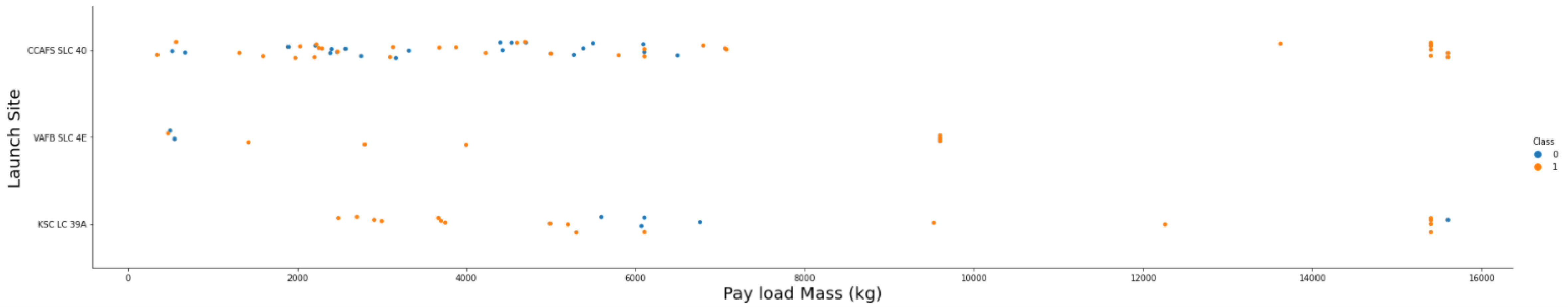
Insights drawn from EDA

Flight Number vs. Launch Site



For every launch site, the success rate generally improved as the flight numbers increased.

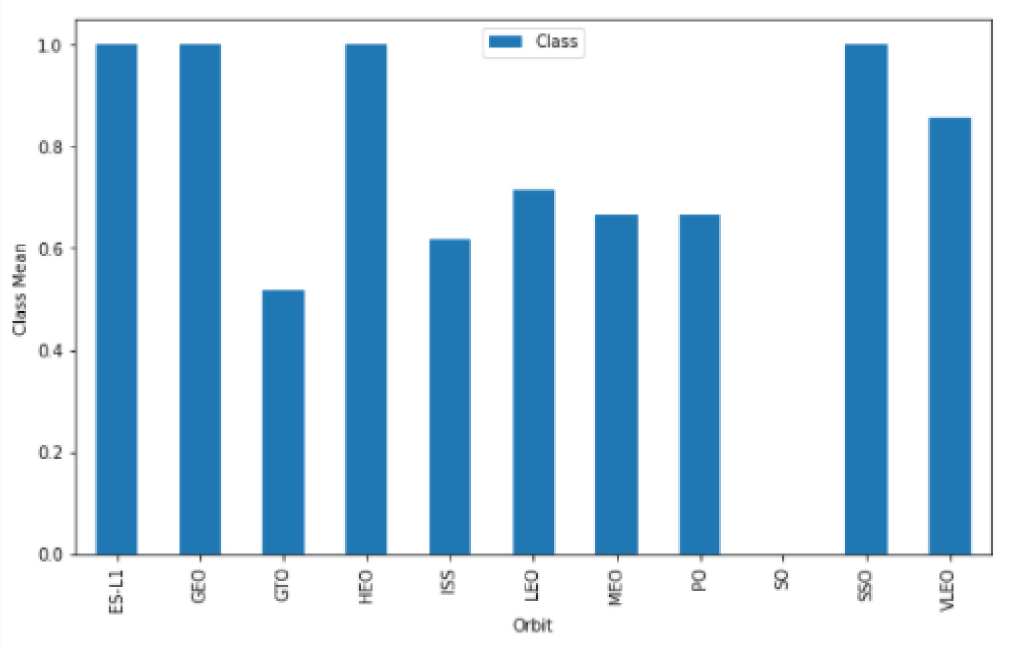
Payload vs. Launch Site



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

As the payload mass increased, the success rate increased.

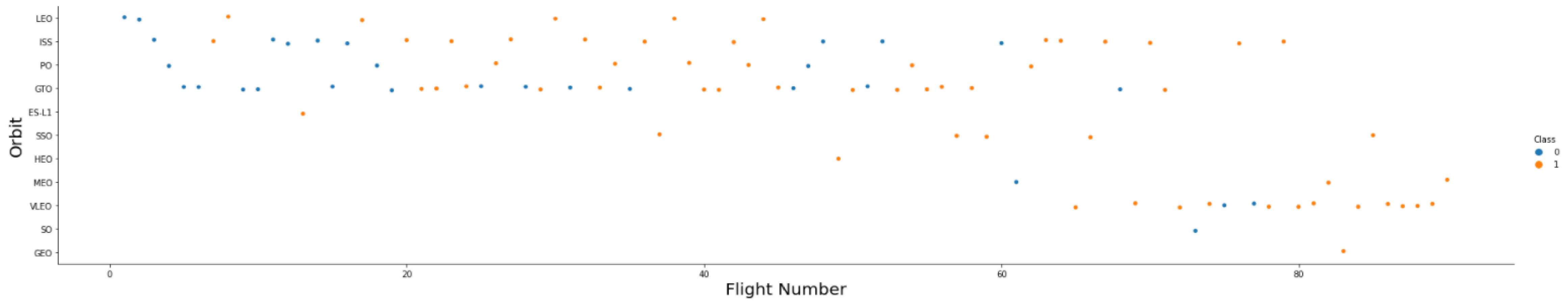
Success Rate vs. Orbit Type



The following orbit types had the highest success rates of 100%

- ES-L1
- GEO
- HEO
- SSO

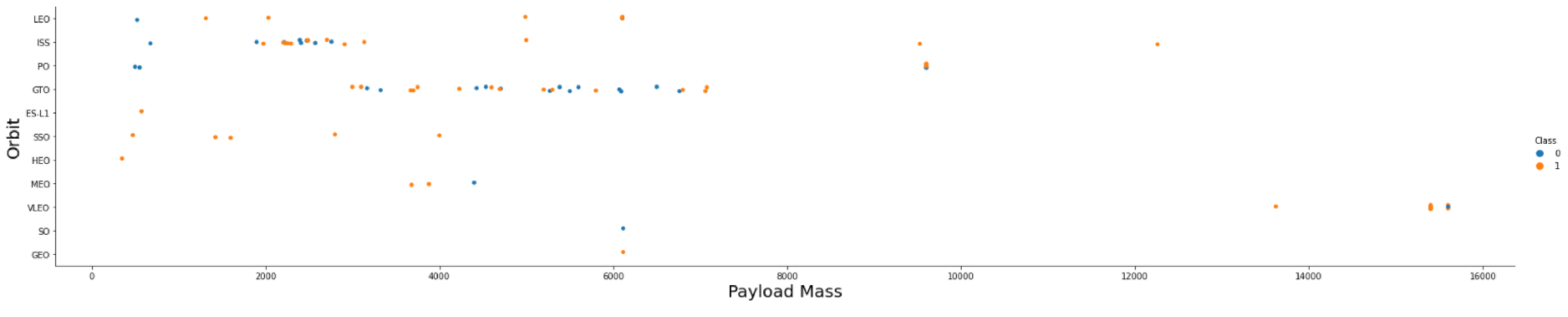
Flight Number vs. Orbit Type



The LEO orbit the Success appears related to the number of flights.

There seems to be no relationship between flight number when in GTO orbit.

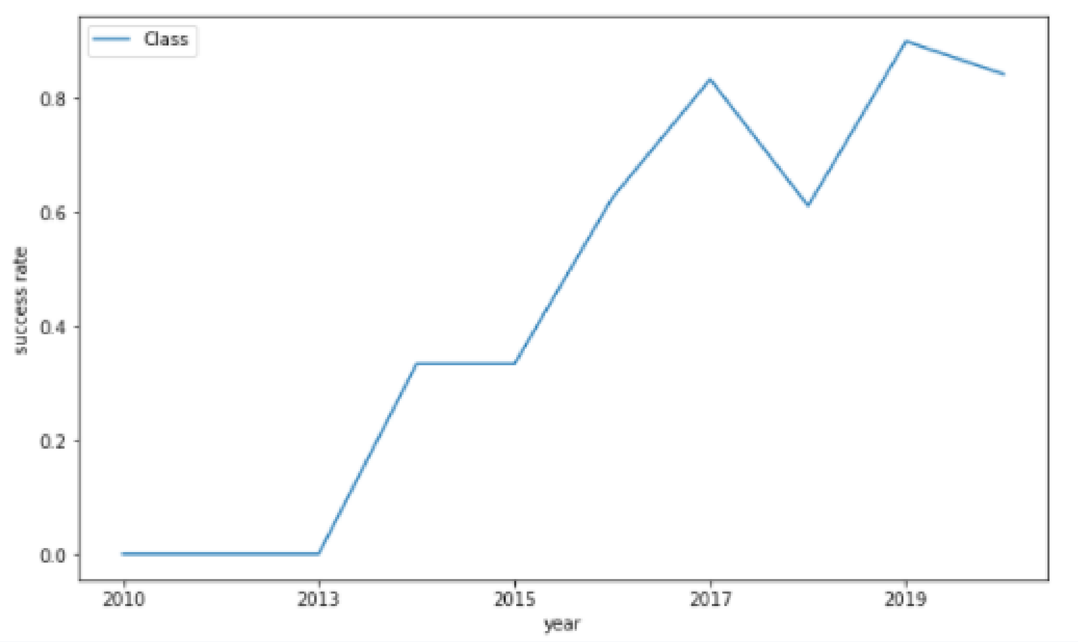
Payload vs. Orbit Type



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

However 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



The sucess rate since 2013 kept increasing until 2020

All Launch Site Names

Unique launch site names were obtained from the database

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

Launch Site Names Begin with 'CCA'

The first 5 launch sites that being with name '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

- Total payload carried by boosters from NASA

total_payload_mass
45596

Average Payload Mass by F9 v1.1

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

average_payload_mass
2928

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

DATE
2015-12-22

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

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
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

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

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

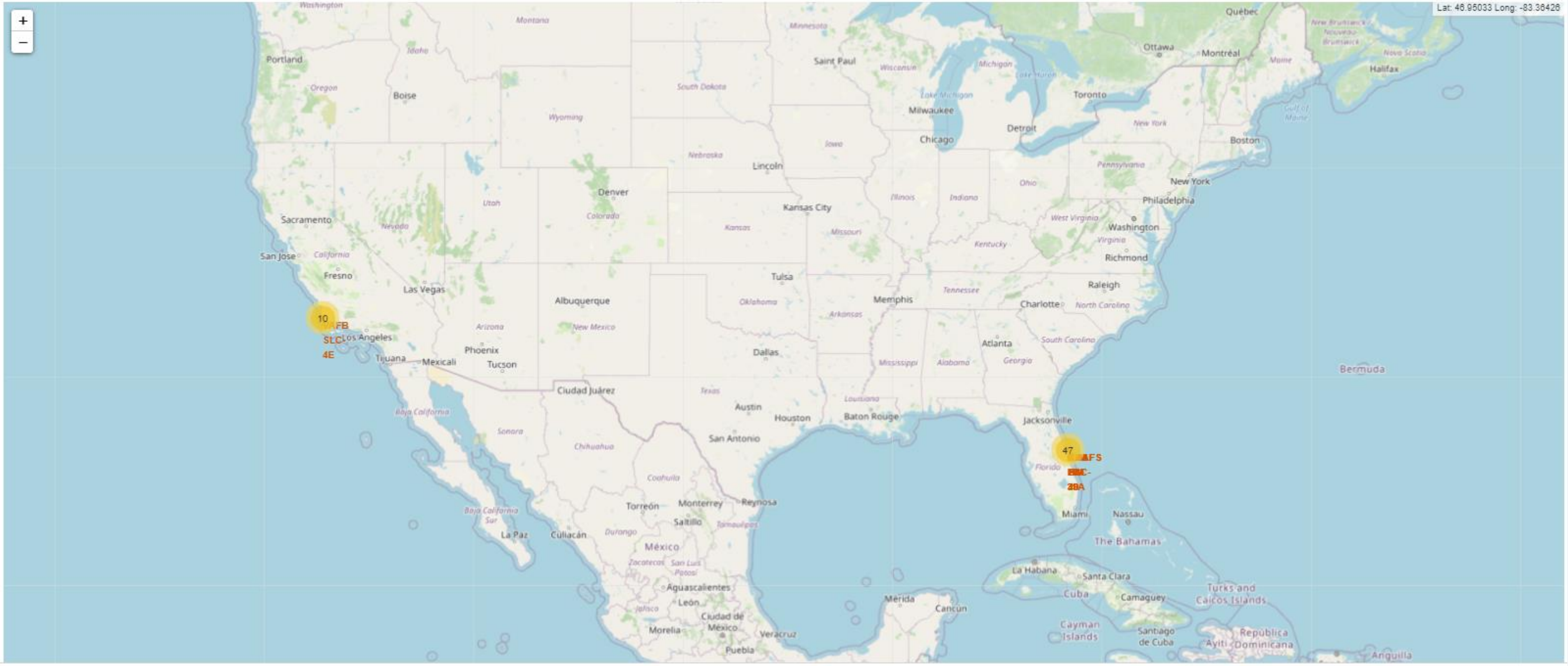
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	COUNT
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

Launch Site Locations

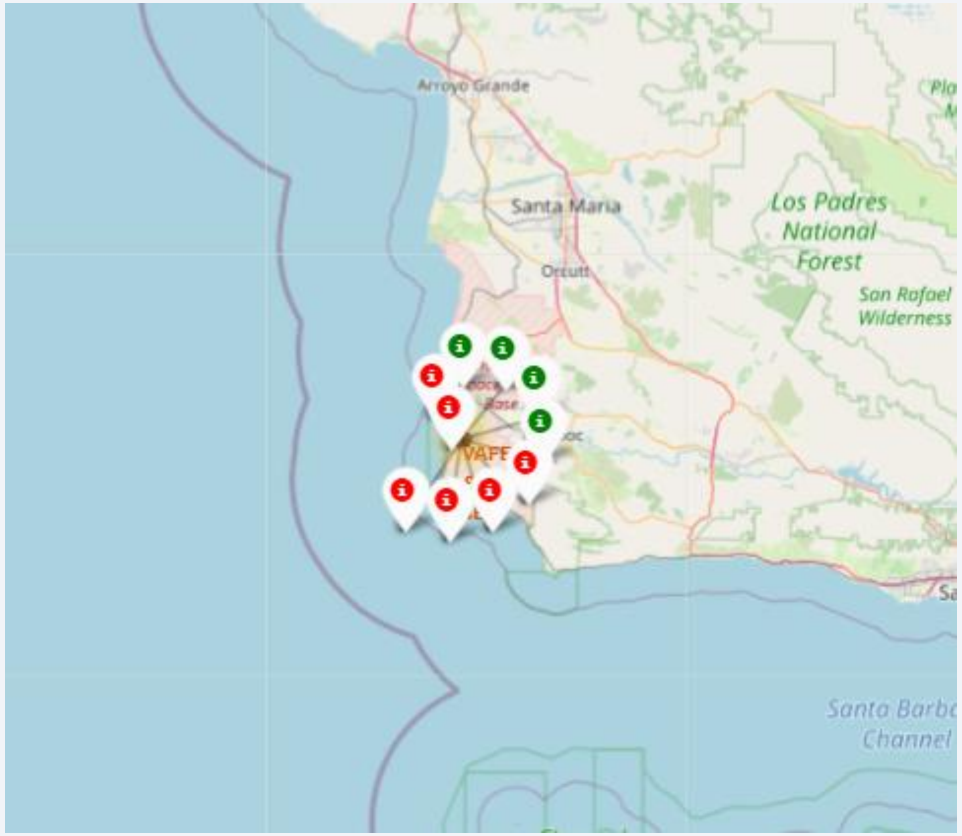


Launch sites are grouped in two spots near the ocean shoreline and far South near the equator for the US.

Successful Launch Outcomes

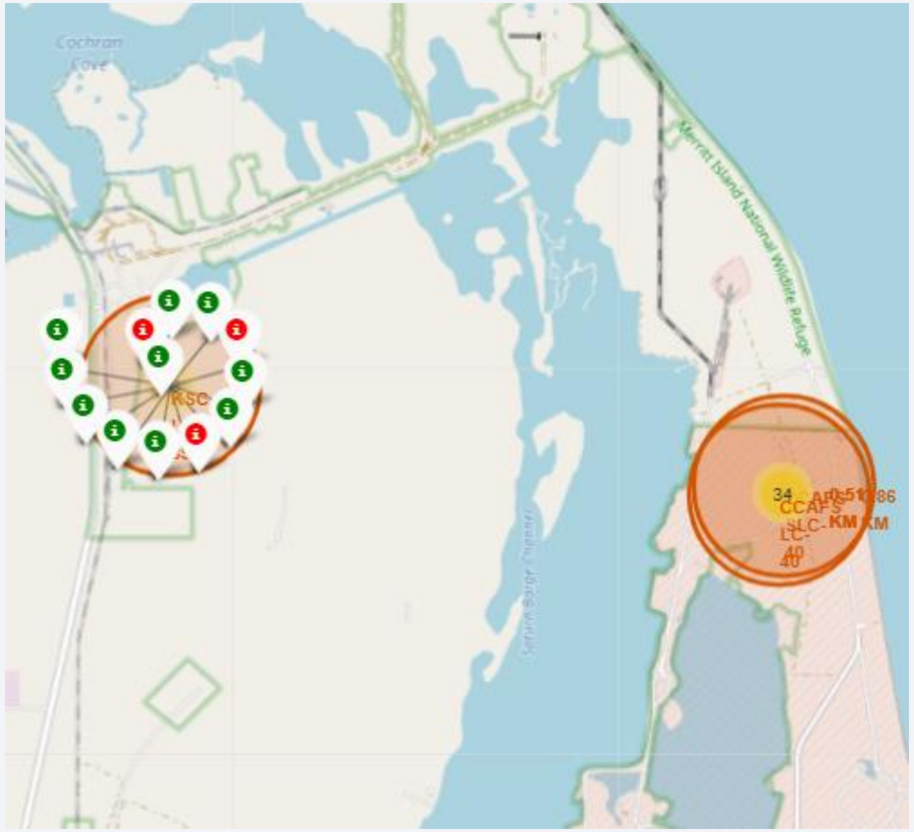
Success = Green

Failure = Red



VAFB SLC-4E

Shows more failures than successes

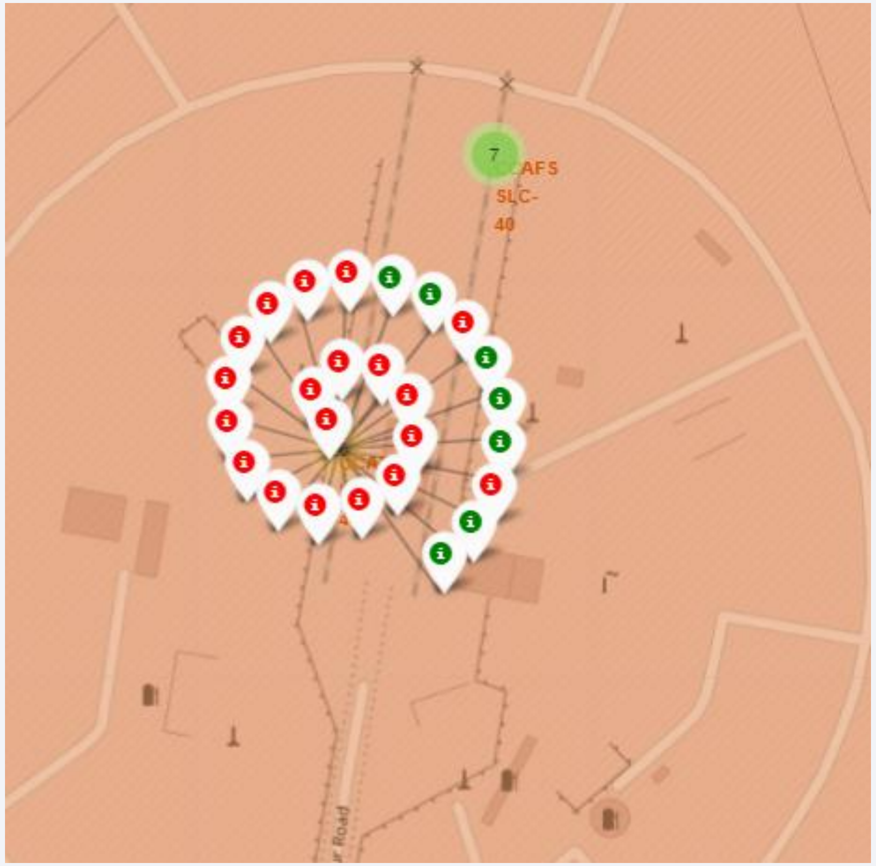


KSC LC-39A

Shows more successes than failures

Successful Launch Outcomes

Success = Green Failure = Red



CCAFS LC-40

Shows many failures

CCAFS SLC-40

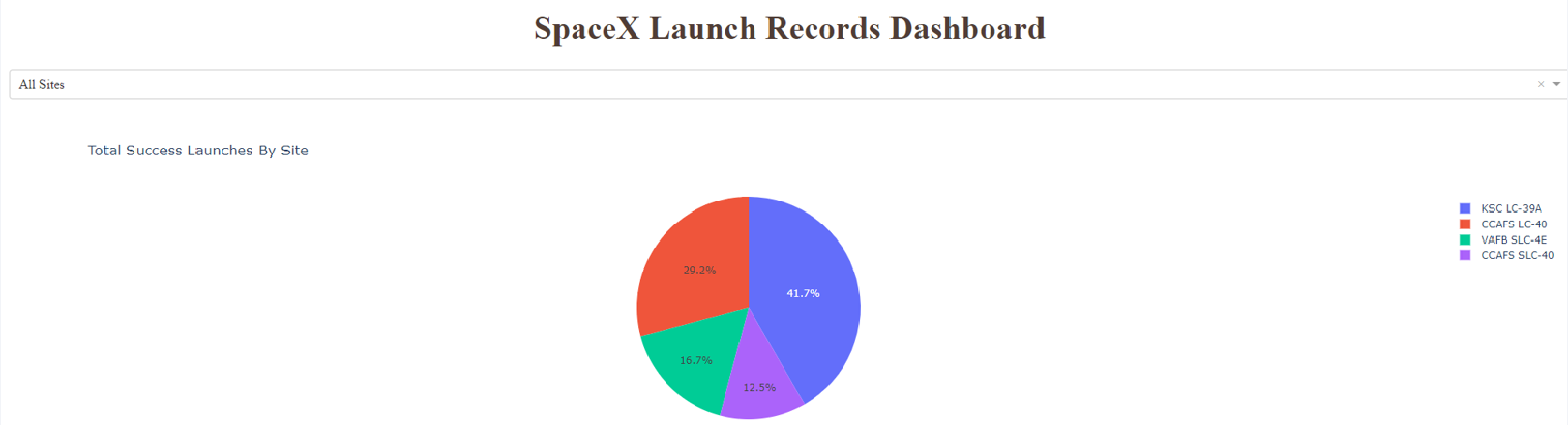
Shows small amount of launches

33

Section 4

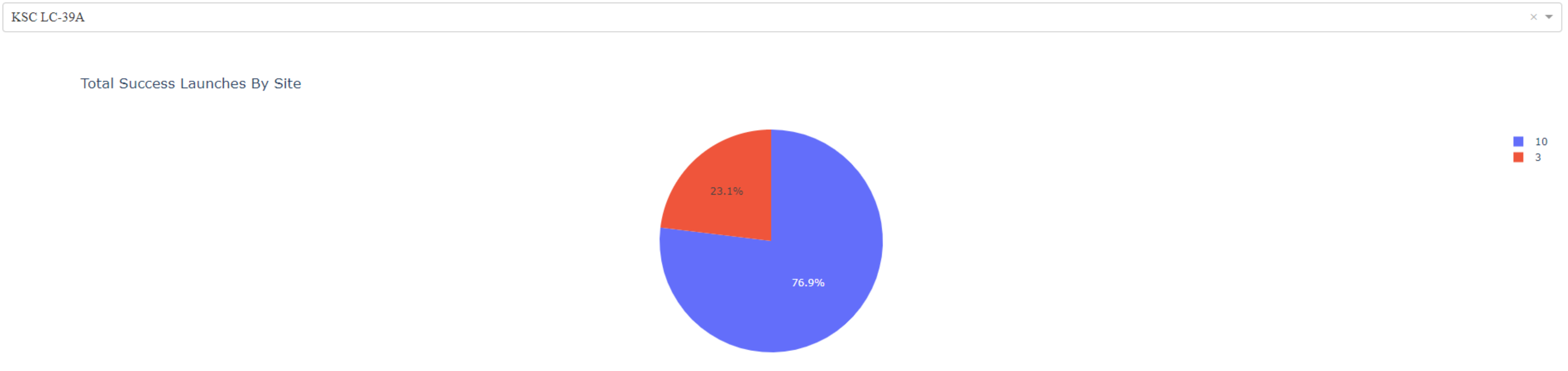
Build a Dashboard with Plotly Dash

Total Success Launches By Site



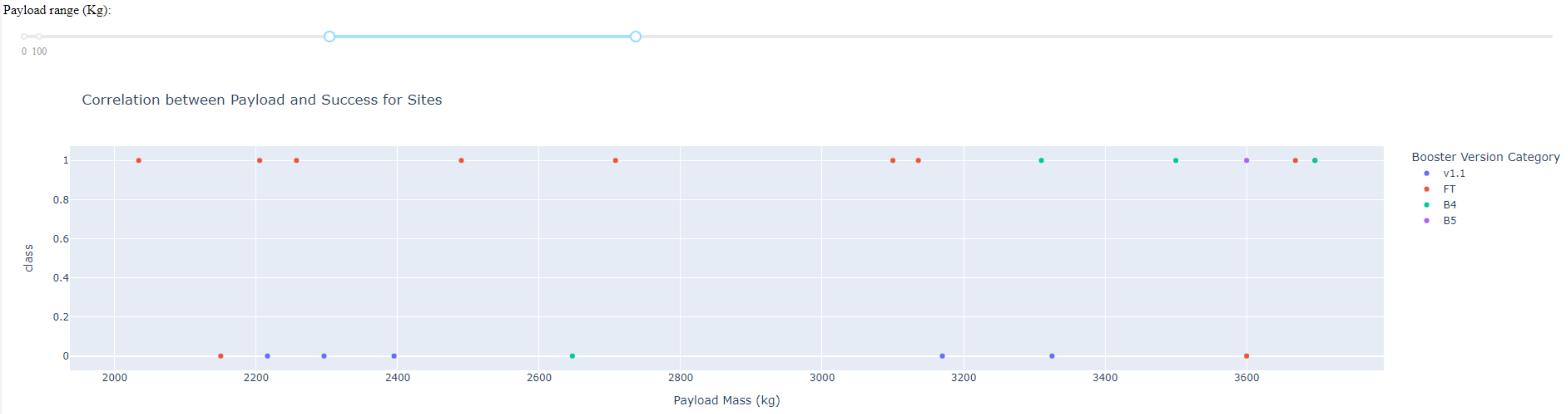
KSC LC-39A has the largest percentage of successful launches.

Detailed Pie Chart for Most Successful Site



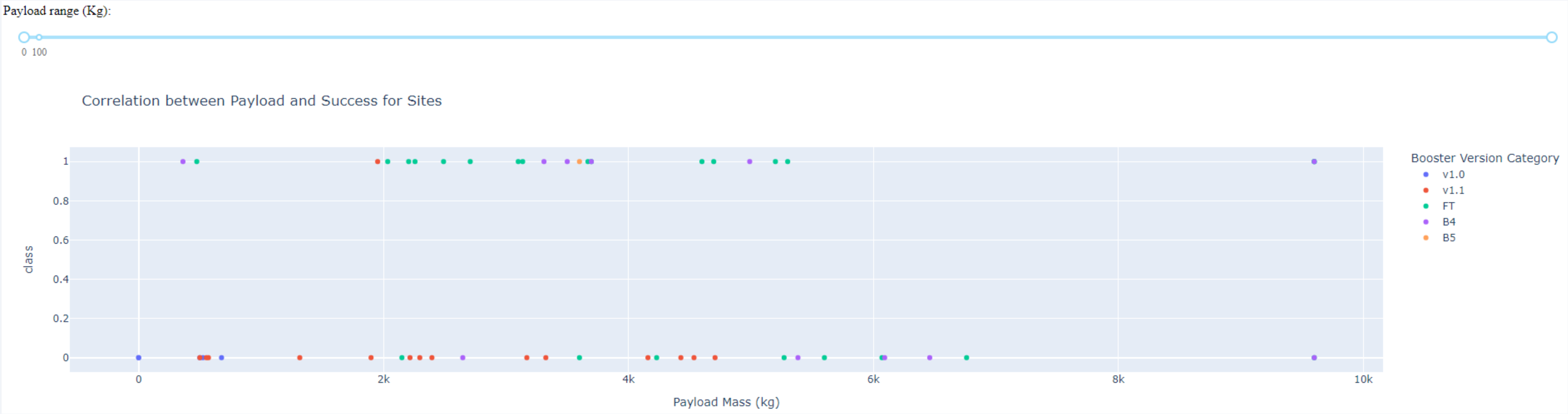
KCS LC-39A had 10 successful launches and 3 failures

Success Rate by Payload Range



The Payload range of 2000 kg to 3800 kg has the largest amount of success rate

Booster Version Very Large Payload



Booster Version B4 is the only one to have a very heavy payload greater than 7000 kg

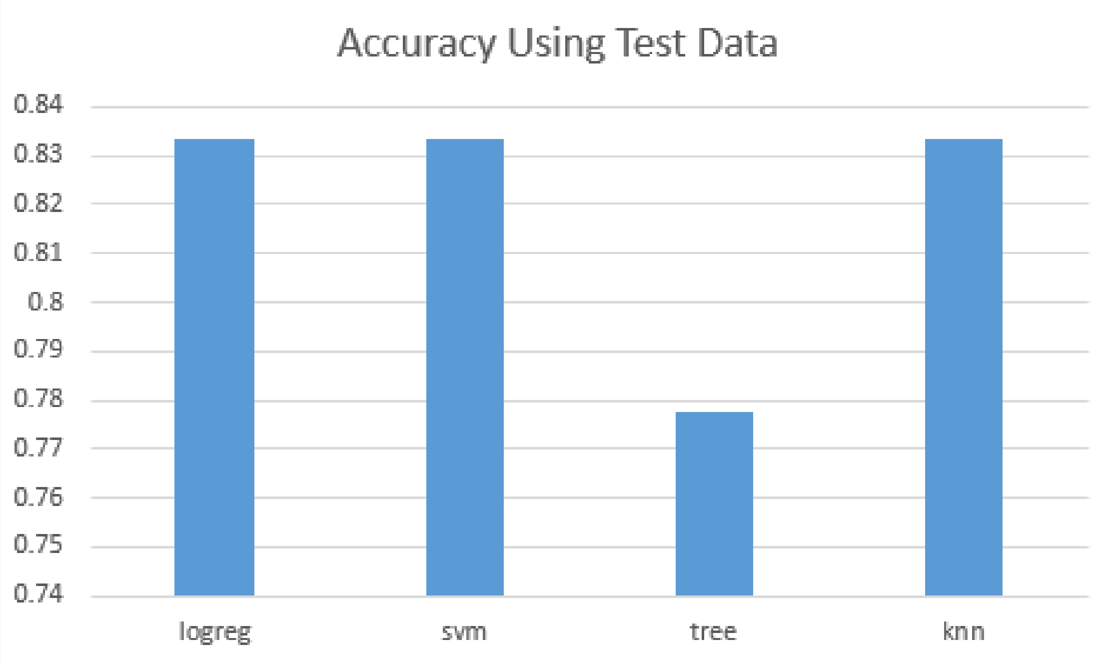
Small Payload Success Rate

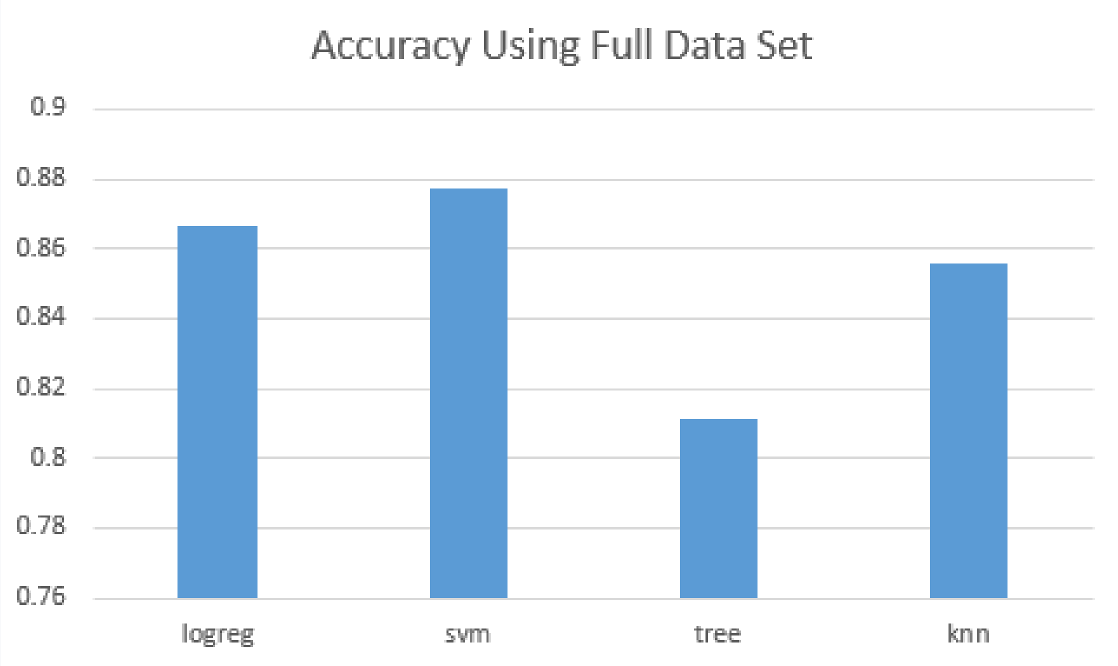
Small payloads have a low success rate

Section 5

Predictive Analysis (Classification)

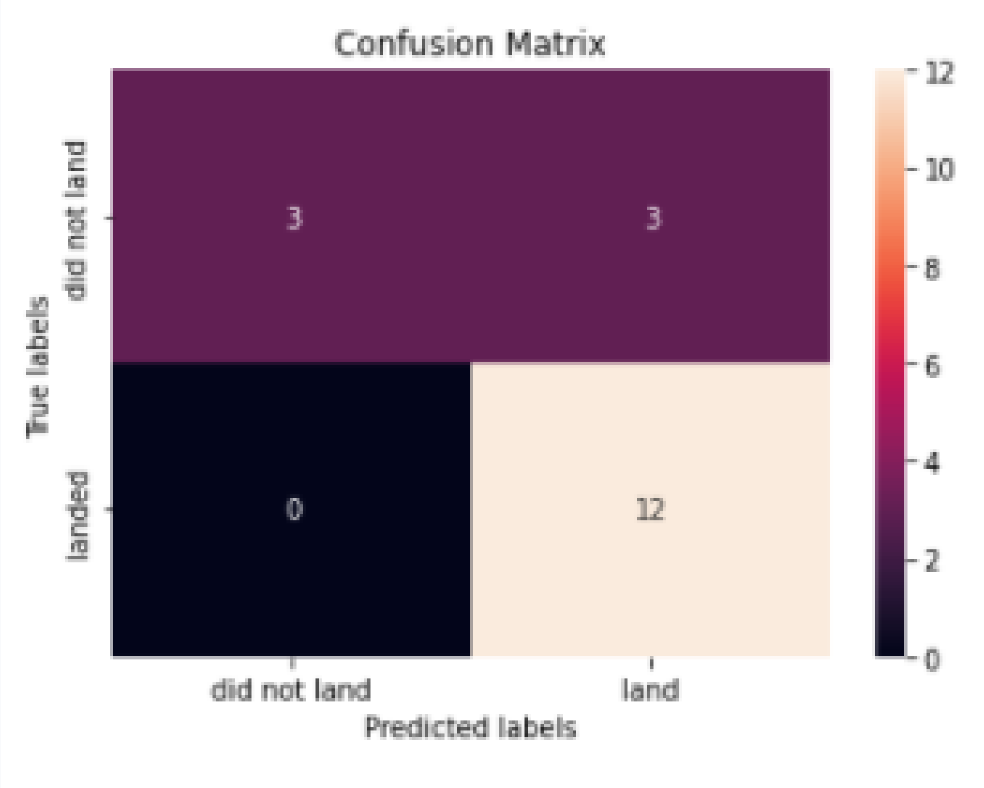
Classification Accuracy





SVM has the best accuracy for this data set

Confusion Matrix



The SVM Confusion Matrix shows the major problem is false positives

Conclusions

- We can predict the success outcome for a launch with a good amount of accuracy of 87%.
- As the number of launches increased, so did the success outcomes. This shows that future outcomes are predicted to be better than past ones.
- Launches with heavier payload masses have a much higher success rate.
- Launches for orbits ES-L1, GEO, HEO, SSO have the best outcome for success.

Appendix

Although the data came from API's and web page scraping, we used shortened versions for the purposes of uniformity for the lab exercises.

- Exploratory Data Analysis dataset 1 https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv
- Exploratory Data Analysis dataset 2 https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv
- Dataset used to run SQL Queries on https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2021-01-01
- Dataset to run API queries on https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Appendix

API sites

- https://api.spacexdata.com/v4/rockets/
- https://api.spacexdata.com/v4/launchpads/
- https://api.spacexdata.com/v4/payloads/
- https://api.spacexdata.com/v4/cores/
- https://api.spacexdata.com/v4/launches/past

Snapshot of Wikipage updated on 9th June 2021

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

Thank you!