IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ismail Shareef
26 March 2022

Outline

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

Executive Summary

- SpaceX advertises launch services at a cost of $62 million for missions that allow some fuel to be reserved for landing the Falcon 9 1st stage rocket booster, so that it can be reused.
- The main goal of this project is to predict whether the Falcon 9 first stage of SpaceX will land successfully or not, so that it can help SpaceY to determine the cost of a launch.
- Based on the given mission data, our machine learning model were able to predict the first stage rocket booster landing successfully with a reasonable degree of accuracy of 83.33%

Introduction - Project Background

- This report has been prepared as part of the ‘Data Science and Machine Learning Capstone Project’ course.
- In this capstone, I am taking the role of a data scientist working for a new rocket company called SpaceY.

Introduction - problems that need answers

- What factors that affects the success or failure of landing?
- Relationship with rocket variables in determining the success rate of a landing?
- Conditions needed for highest success landing rate?

Section 1

Methodology

Methodology

Executive Summary

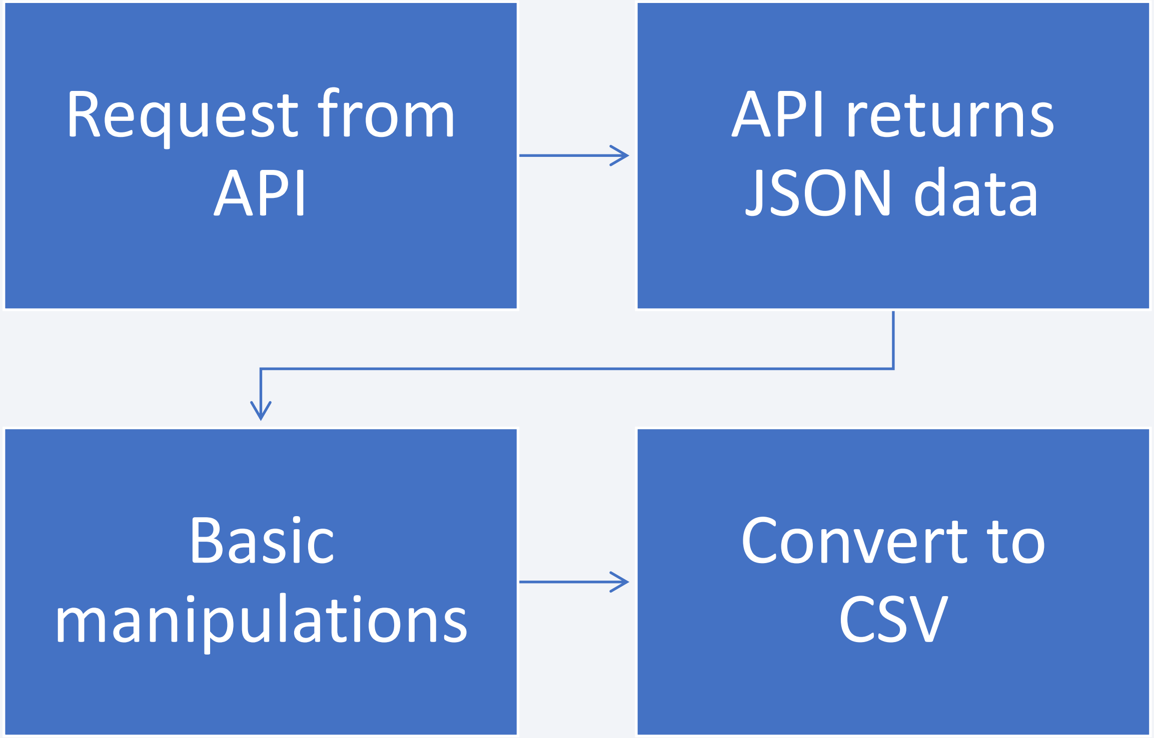
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- SpaceX REST API
  - Launch data that provided data about launches, rocket use, payload, launch specs, landing specs and learning outcome.
  - https://api.spacexdata.com/v4/launches/past
- Web Scraping
  - Historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

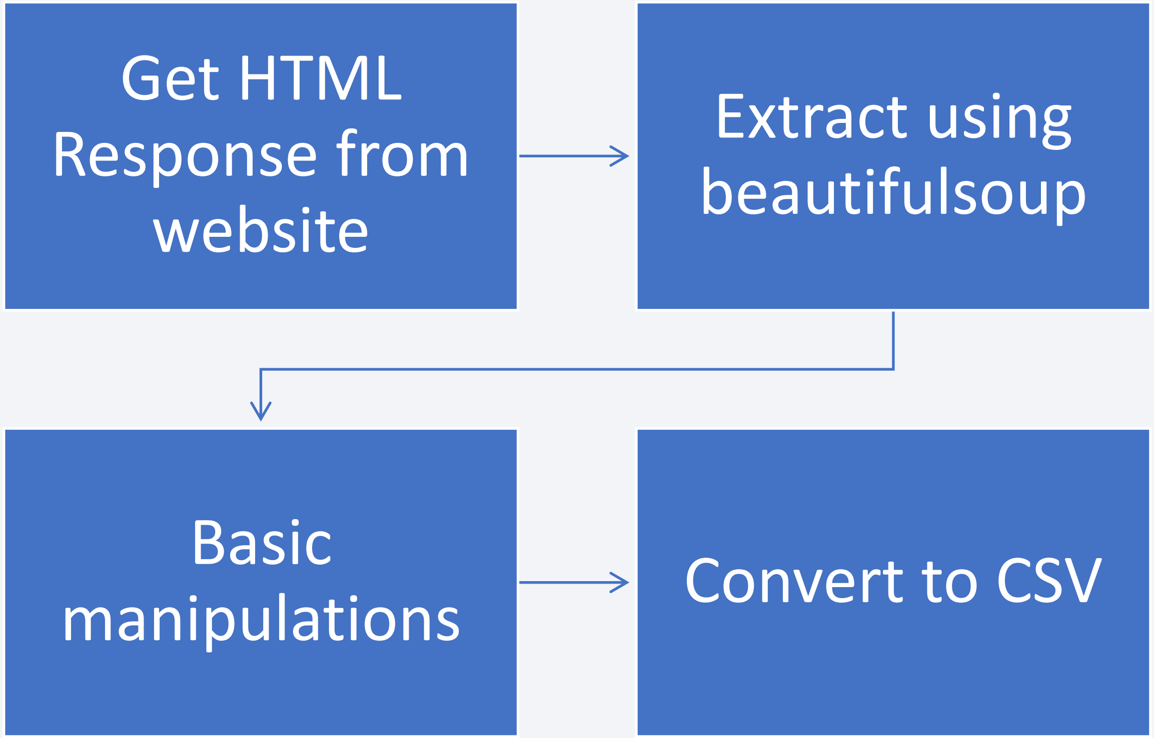
Data Collection – SpaceX API

- Acquired historical launch data from open source REST API for SpaceX
- GitHub: Link



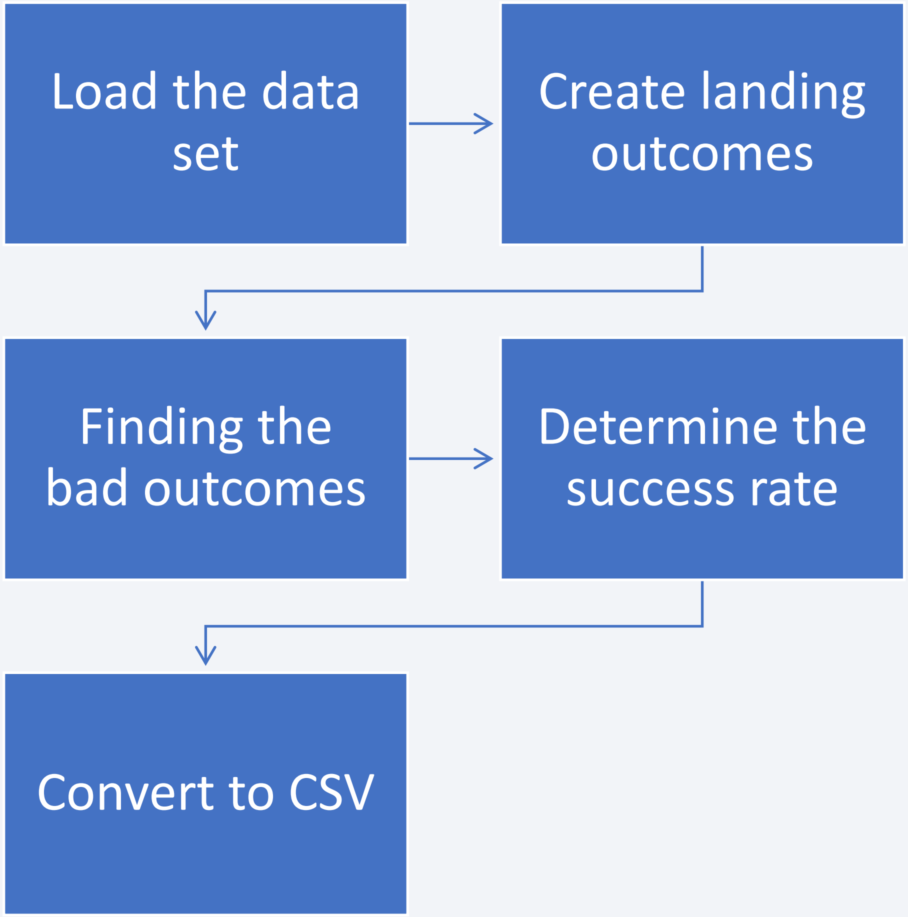
Data Collection - Scraping

- Some of the essential data was collected from Wikipedia using web scrapping with the help of beautiful soup framework.
- GitHub: Link



Data Wrangling

- To determine the labels for training models the following was calculated
  - Number of launches on each site
  - Number and occurrences of each orbit
  - Number and occurrences of mission outcome per orbit
- Landing outcome label was created from the outcome column
- GitHub: Link



EDA with Data Visualization

- Following relationships between variables were plotted:
  - Flight Number vs. Payload (Cat plot)
  - Flight Number vs. Launch Site (Cat plot)
  - Launch Site vs. Payload (Scatter plot)
  - Success Rate vs. Orbit type (Bar plot)
  - Orbit type vs. Flight Number (Scatter plot)
  - Orbit type vs. Payload (Scatter plot)
  - Success rate vs. Time in years (Line plot)

- Further explanations on these various plots are shown in Section 2
- GitHub: Link

EDA with SQL

- Loaded data into an IBM db2 instance and ran SQL queries to display and list the information about the following
  - Launch Sites
  - Payload Mass (kg)
  - Mission Outcomes
  - Booster Version
- GitHub: Link

Build an Interactive Map with Folium

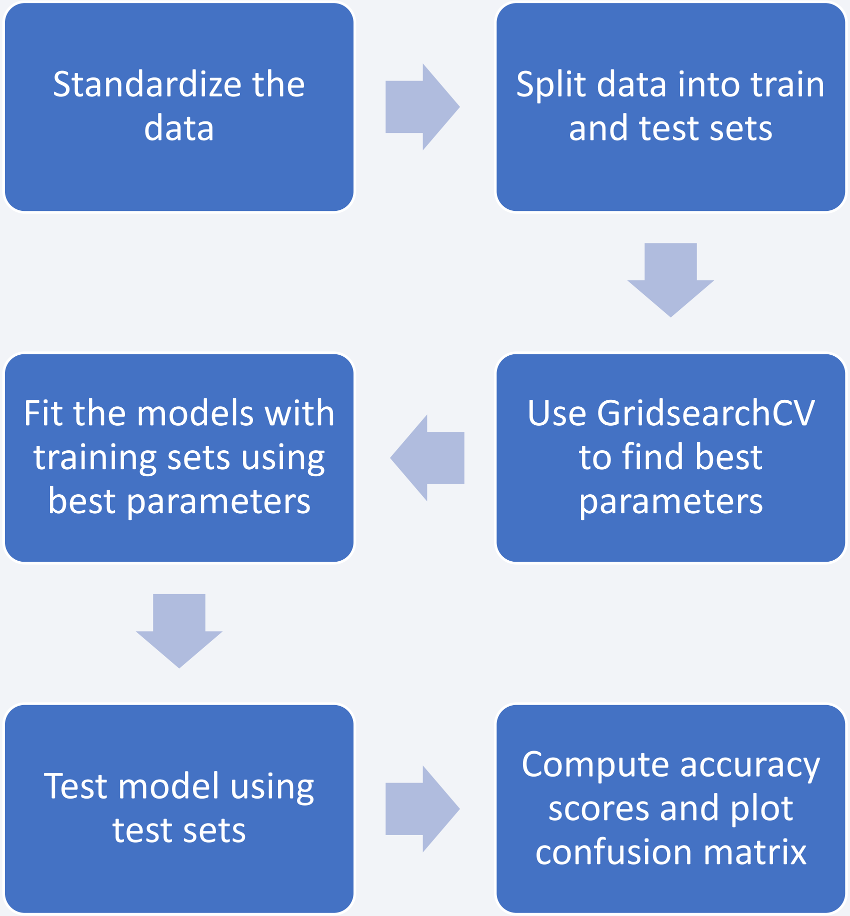
- Using the Python package Folium, we created an interactive map where one can:
  - View where each Falcon 9 launch site is located, represented by a circle
  - Learn how many launches occurred at each location, represented by markers. Green markers represent a successful recovery while red markers represent unsuccessful one
  - Determine distances to the closest coastline, city, railway, and highway, each represented by a blue line.
- GitHub: Link

Build a Dashboard with Plotly Dash

- An interactive web application was created using Plotly dash that shows
  - A pie chart showing the proportion of successful recoveries to unsuccessful ones for each site (selected from a dropdown menu)
  - A recovery outcome vs Payload mass scatter plot with a range of (0-10000 kg) with a range slider that can changed by the user
- This dashboard provides insight into the launch sites' and payload masses' relationships with the recovery outcomes.
- GitHub: Link

Predictive Analysis (Classification)

- The model development process in the diagram in the right was used for all of the following methods
  - Logistic Regression
  - Support Vector Machines
  - Decision Tree
  - KNN
- GitHub: Link



Results

- Exploratory data analysis results
  - Launches are most successful when launched in 2017 or later
  - Light payloads are easier to recover, as most successful recoveries occur when the payload has a mass between 2000kg and 4000kg)
  - Each launch site is reasonably successful, but site KSC LC-39A appears to be ideal as it has a success rate of over 75%
  - The best recovery method appears to be via drone ships. SpaceX has much more control over the recovery maneuvering a ship in the ocean as opposed to a stationary landing pad on land

Results

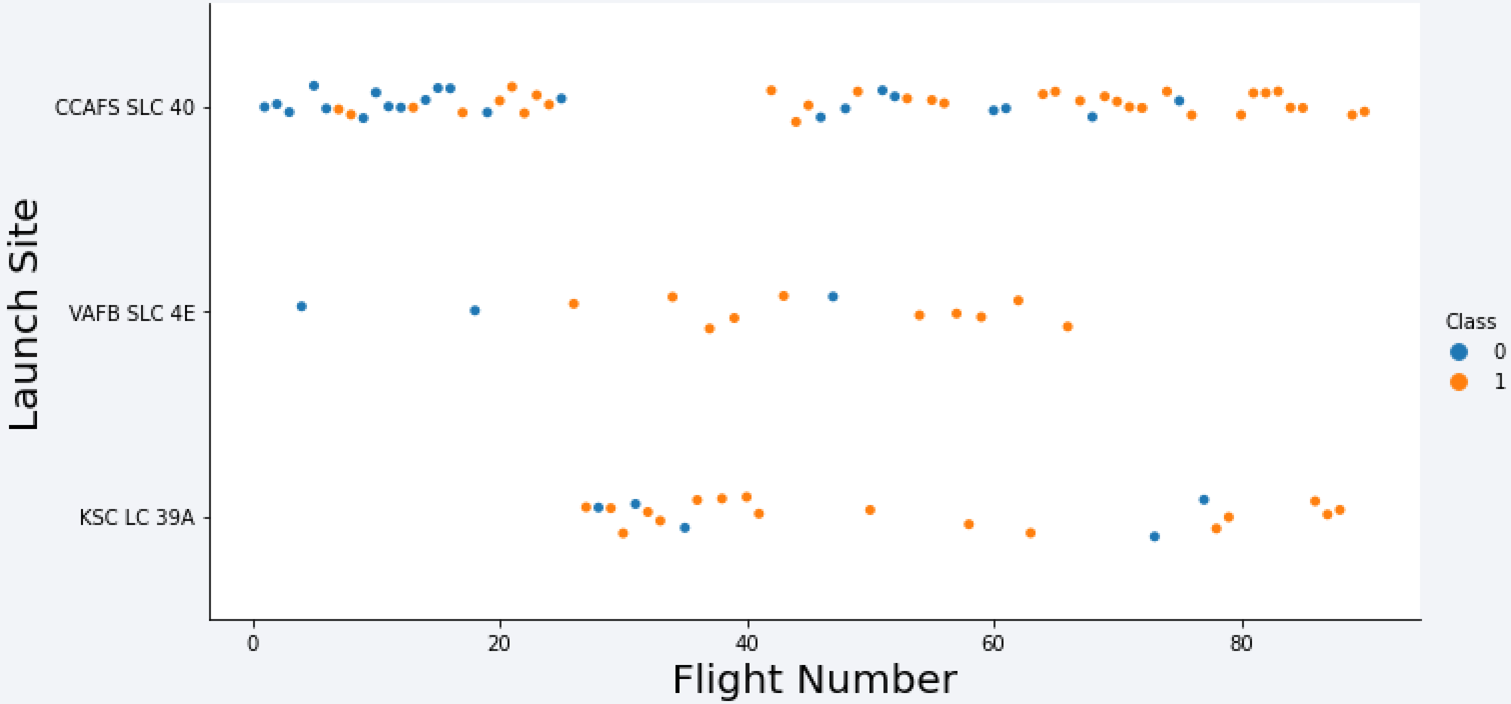
- Interactive analytics demo in screenshots
  - The py file can be downloaded from page 15
  - Refer to Section 5 for screenshots of the dashboard
- Predictive analysis results
  - Each model performed about equally, correctly predicting a recovery outcome at a rate of 83.33%

Section 2

Insights drawn from EDA

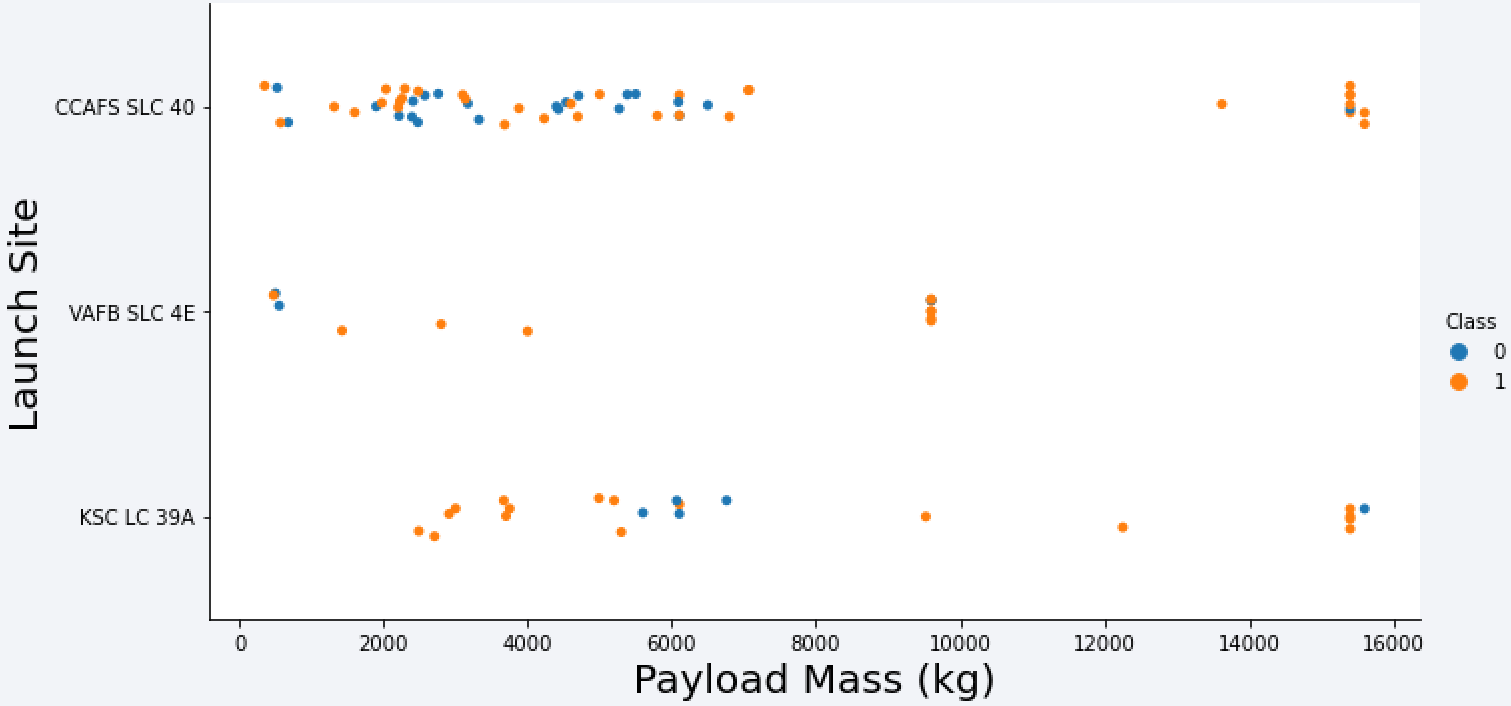
Flight Number vs. Launch Site

- Rate of success has grown over time at each site, though KSC LC-39A seems to be the most consistent



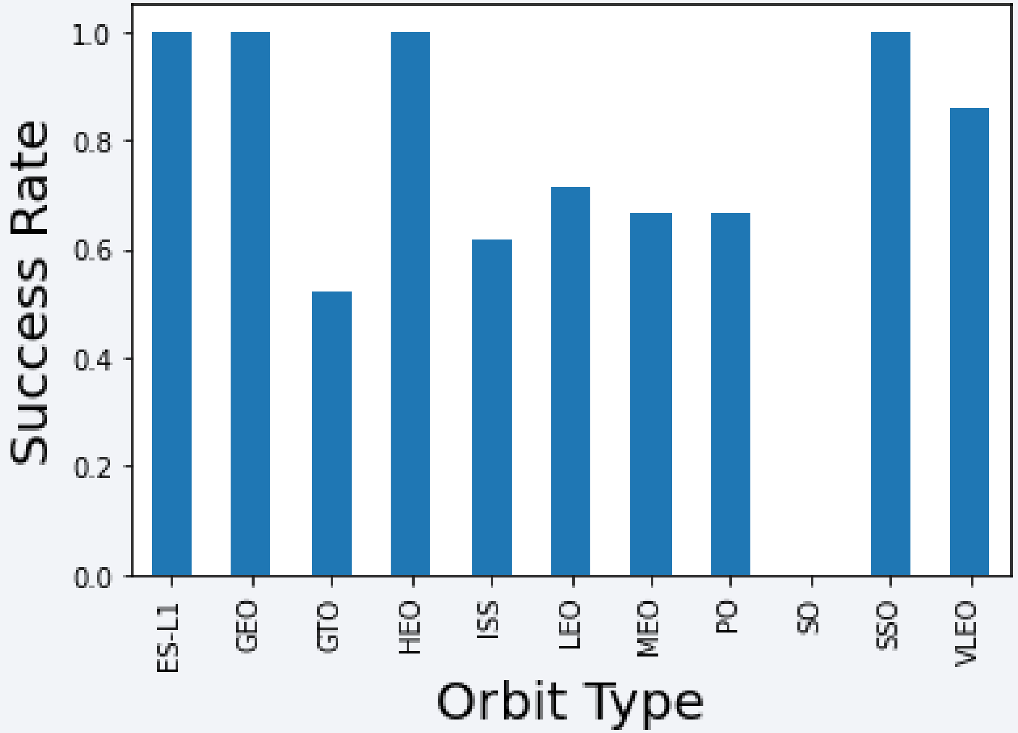
Payload vs. Launch Site

- Smaller payloads (<6000 kgs) seem to correlate with higher success rate



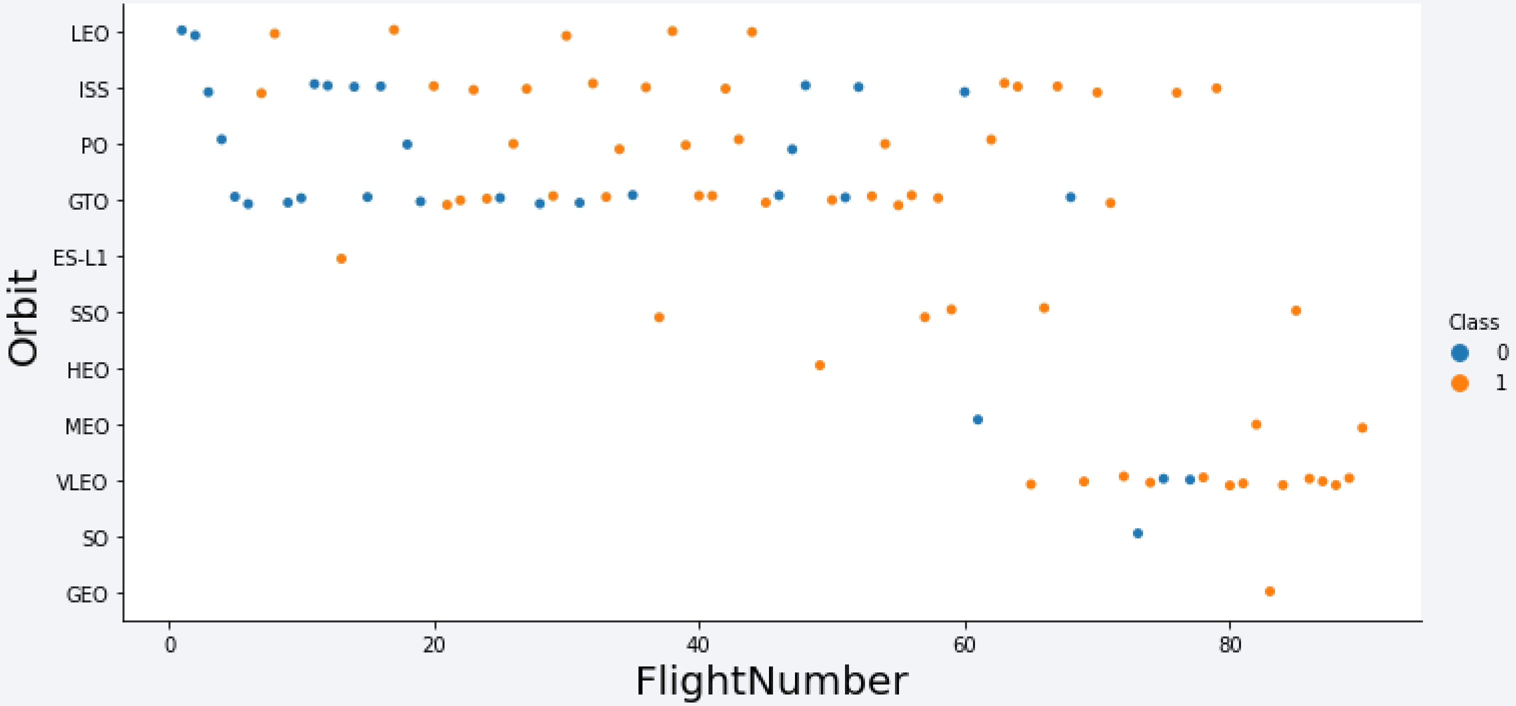
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits are very reliable



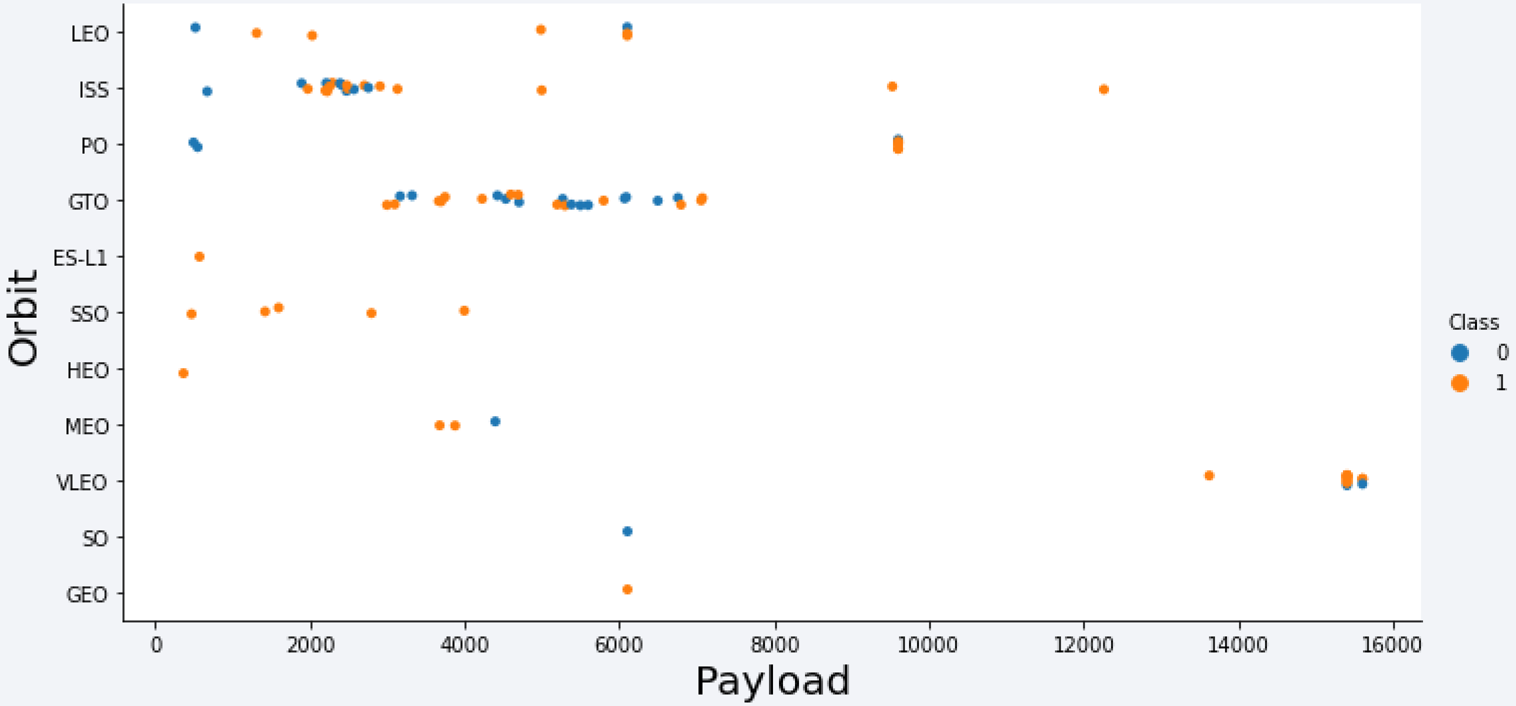
Flight Number vs. Orbit Type

- ES-L1, HEO, and GEO success rate may be skewed due to each having only 1 launch
- LEO, SSO, and VLEO seem all have high success rates while having good sample sizes



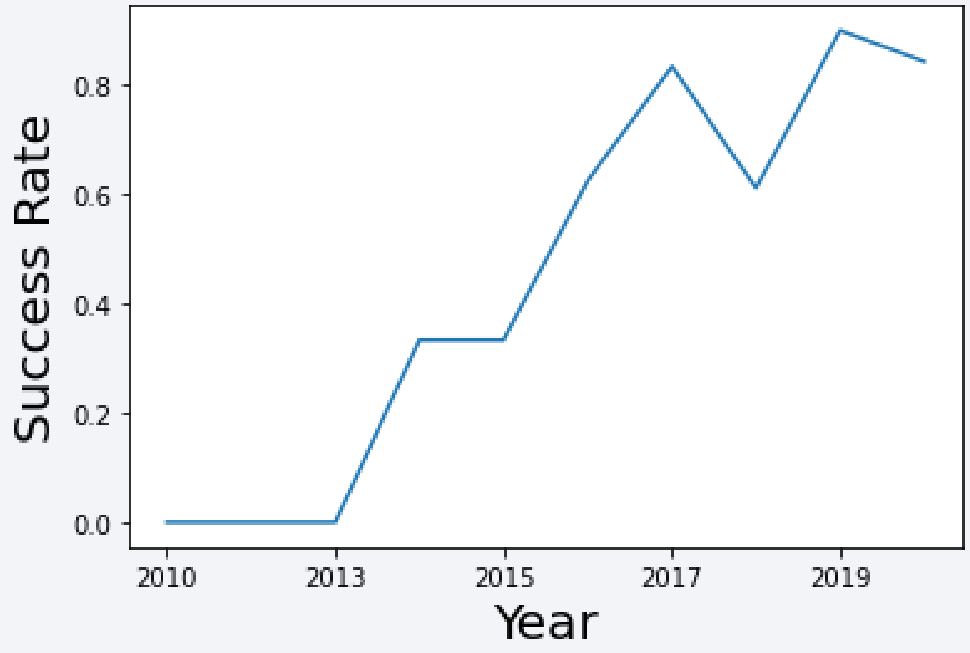
Payload vs. Orbit Type

- LEO and SSO orbits' success may be due to light payloads



Launch Success Yearly Trend

- Overall success rate increased from 2013 to 2017, where it has more or less stagnated



All Launch Site Names

- There were 4 unique launch sire locations according to the SQL query run

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

Launch Site Names Begin with 'KSC'

- These are the first 5 records for launch sites that begin with ‘KSC’:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	06:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

- The total payload mass (in kg) carried by boosters from NASA (CRS) is

1
45596

Average Payload Mass by F9 v1.1

- On average, rockets with by booster version F9 v1.1 carry a mass of:

1
2534

First Successful Drone ship Landing Date

- The first successful Stage One recovery landing in drone ship occurred on

	1
	2016-04-08

Successful Ground pad Landing with Payload between 4000 and 6000

- This is a list of the booster versions which have successfully landed on a ground pad and had payload mass greater than 4000kg but less than 6000kg

booster_version
F9 FT B1019
F9 FT B1025.1
F9 FT B1031.1
F9 FT B1035.1
F9 B4 B1039.1
F9 FT B1035.2

Total Number of Successful and Failure Mission Outcomes

- In total, there were 101 missions recorded in this database. The mission outcomes are as follows

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

Boosters Carried Maximum Payload

- Here is a list of the booster versions 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

2017 Launch Records

- Monthly records for 2017 displaying successful landing outcomes in ground pad, the booster versions, and the launch sites are as follows

month_name	landing__outcome	booster_version	launch_site
February	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
May	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
June	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
August	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
September	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
December	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

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

- This is a rank of the successful landing outcomes (in descending order) between dates 2010-06-04 and 2017-03-20

landing__outcome	counts
Success (drone ship)	5
Success (ground pad)	3

Section 3

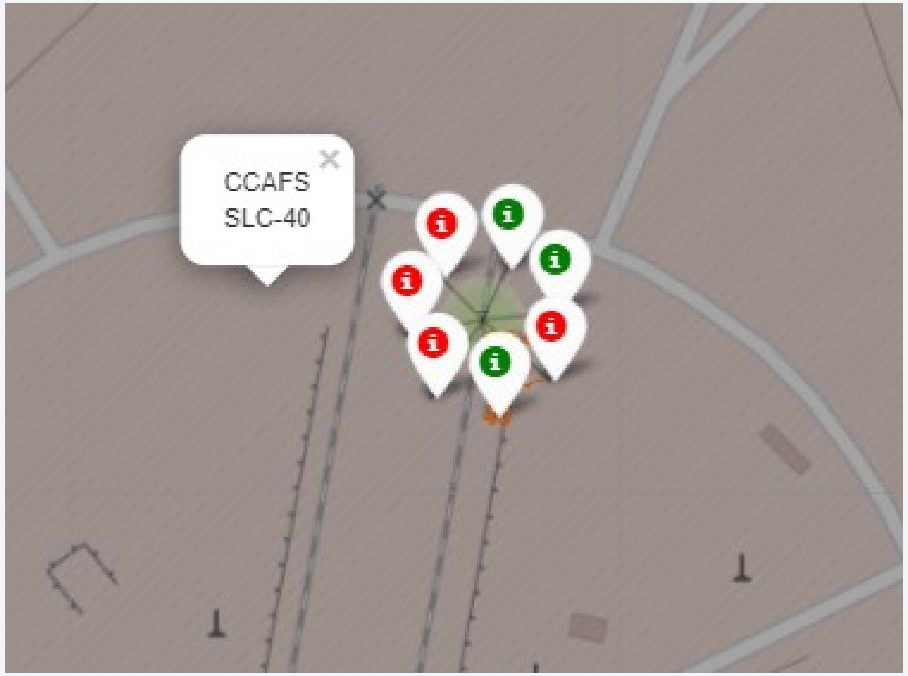
Launch Sites Proximities Analysis

Map of Launch Site Locations

- SpaceX has one launch site on the Pacific coast and Atlantic coast of Florida

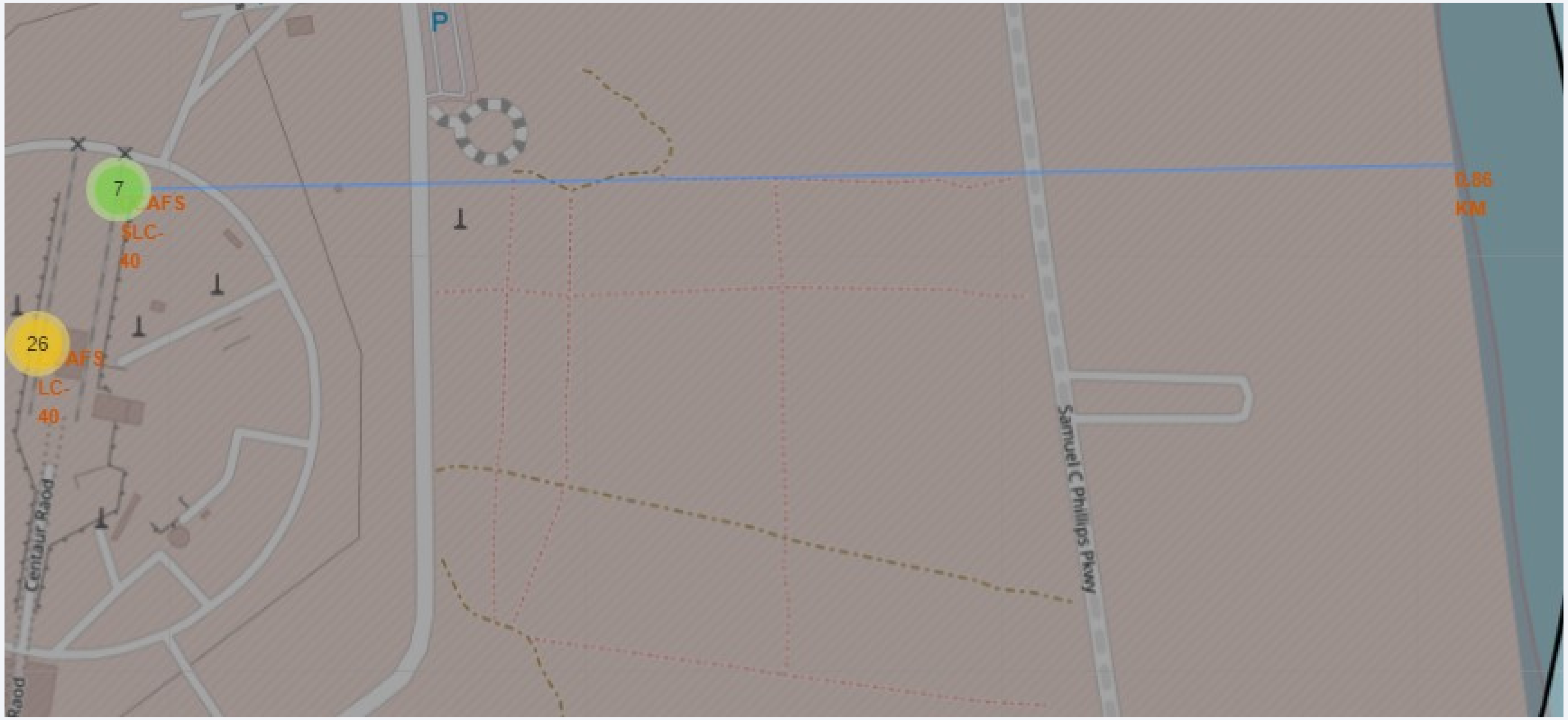
Launch Outcomes

- The map shows the launch outcomes for site CCAFS SLC-40
- Green markers indicate a successful landing while red indicates a failed landing



Distances between a launch site to its proximities

- The blue line represents the distance from site CCAFS SLC-40 to its nearest coastline, with distance calculated and displayed

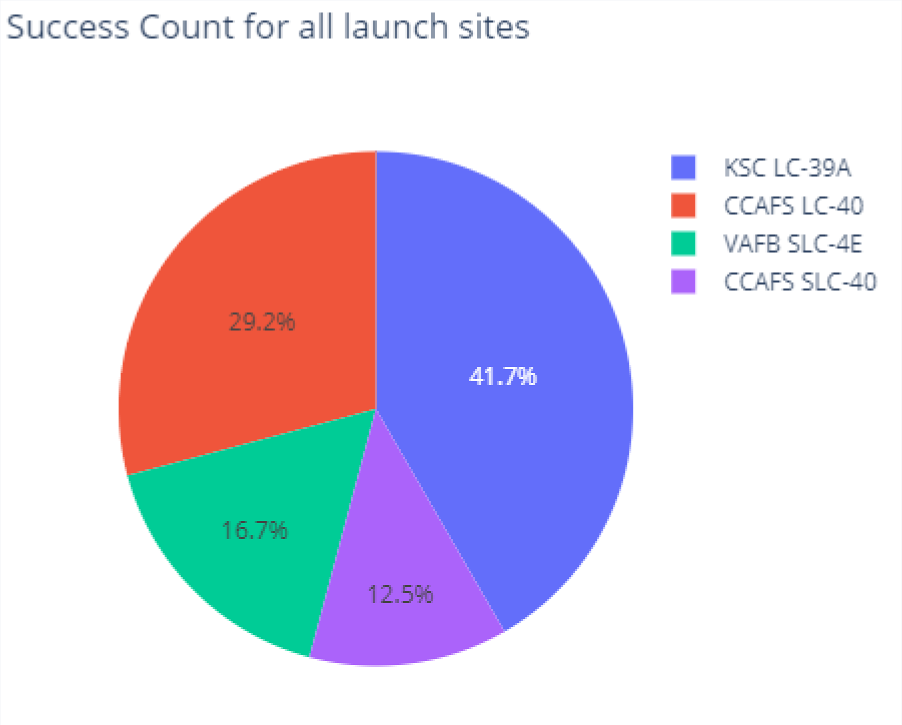


Section 4

Build a Dashboard with Plotly Dash

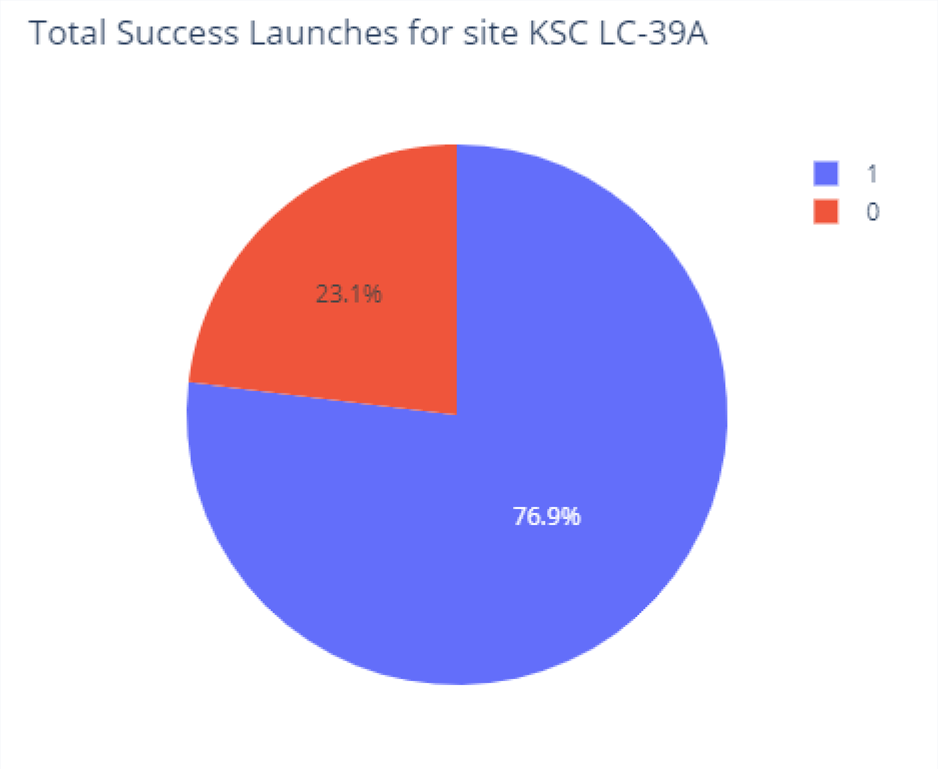
Pie Chart of Successful Launches by Site

- The success count of all sites is show as a pie chart in the image right
- We can see that site KSK LC-39A has the highest successful launches



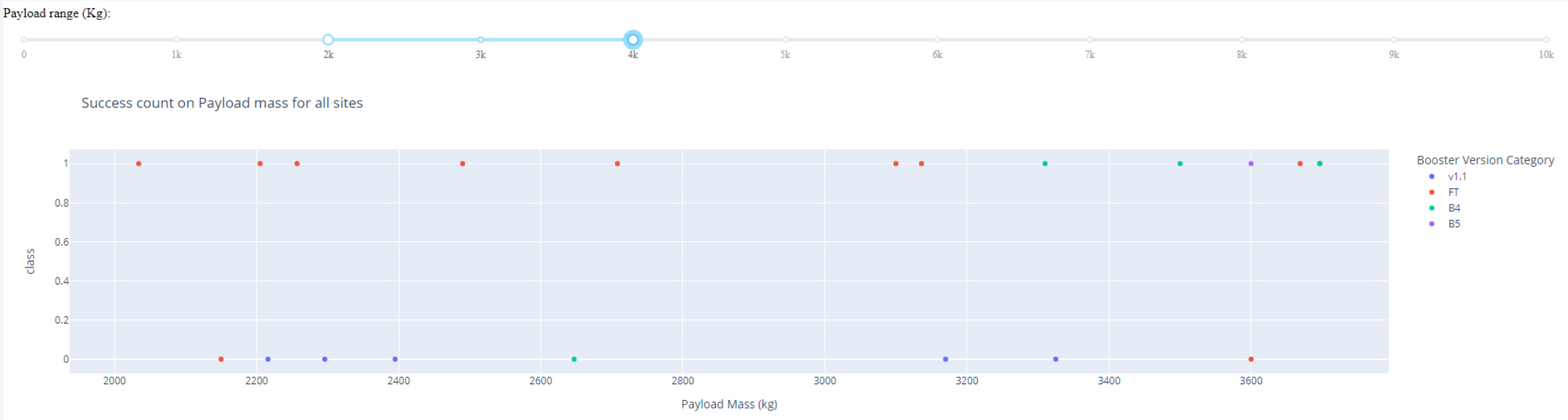
Launch outcome of KSC LC-39A,

- The pie chart shows that from all launches made from site KSC LC-39A 76.9% were succesful



Launch outcome vs payload mass for all sites

- The scatter plot below shows landing launch outcomes for payload mass between 2000 and 4000kg.
- The v1.1 booster seems to have failed the most while the FT booster appears to be very successful

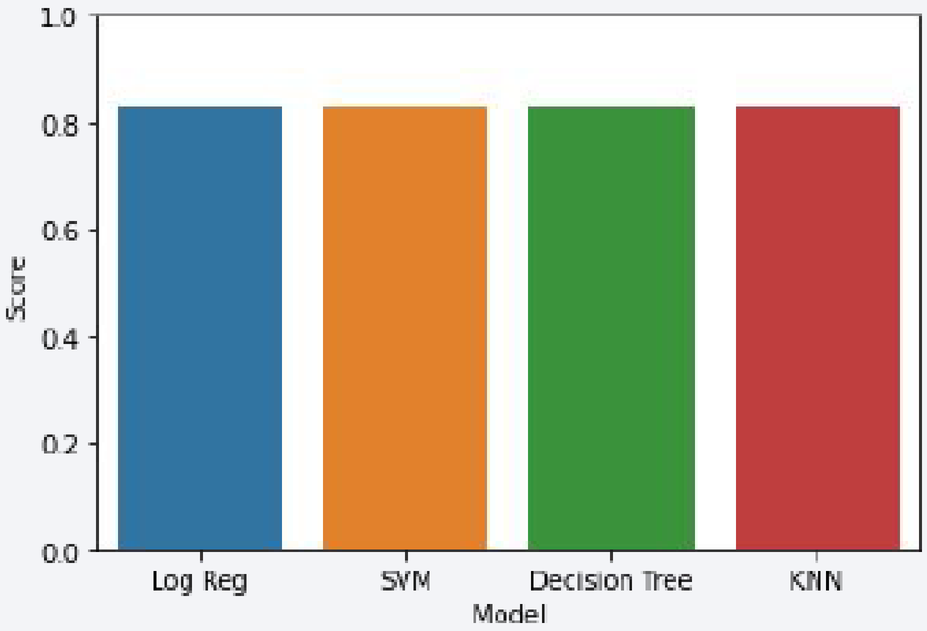


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The bar chart shows the accuracy score of each model.
- As seen, when testing them on the test data, they all exhibit identical scores (83.33%)



Confusion Matrix

- The confusion matrices of the best performing models are the same.
- 3 out of 18 samples was incorrectly predicted to land

Conclusions

- SpaceX's successful recoveries generally have the following properties:
  - A launch date in the year 2017 or later
  - Light payload (in the range 2000-4000kg)
  - Launched from site KSC LC-39A
  - Successfully recovered via drone ship
- Our model can predict the outcome of a given recovery with a reasonable degree of accuracy, 83.33%

Appendix

- All codes can be found on my GitHub
- Github Repo: Link

Thank you!