IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christopher Chan

Outline

Executive Summary

- This capstone project is to determine if SpaceX's Falcon 9 launch will successfully land using different machine learning models.
- The methodology to acquire, clean, and analyze data is as follows: Data collection (API & Webscraping), Data wrangling (pandas), Exploratory data analysis (SQL and pandas), Data visualization (plotly, folium, and seaborn), and Machine Learning (pandas and sklearn)
- There were several features that was correlated to the success of a rocket launch, these features were used during modeling
- Based on confusion matrix and score analysis of the different ML models, we conclude that the best model is decision tree with an accuracy of 0.89

Introduction

- The goal of the capstone project is to determine if Falcon 9 launch will be successful. The inspiration behind this project is the ability to reuse the rocket after launch, which will save operating cost and enable rocket reusability. We can significantly reduce the cost of SpaceX by helping them predict the success rate of their rocket launches.
- The problem we hope to answer with our analysis is whether Falcon 9 launch will be successful based on several features

Section 1

Methodology

Executive Summary

Data was collected through requesting data from SpaceX API and webscraping launch data. The data sets were cleaned and transformed using pandas. After cleaning the data, exploratory data analysis was carried out using python data analysis (pandas) visualization (seaborn, plotly dash, and folium) libraries and SQL to grasp a better understanding of the dataset.

Logistic regression, support vector machine, decision tree, and k-nearest neighbor models were used for predicative analysis. Each model was trained using the same training and testing dataset and optimized using GridSearchCV. The accuracy of each model was determined using a confusion matrix and score function

Summary

1. API
2. Webscraping
3. Data Wrangling
4. EDA
5. Folium
6. Plotly Dash
7. Machine Learning Model Optimization

Github link: Github

Data Collection – SpaceX API

1. Request SpaceX launch data using GET request
2. Normalize JSON response into a dataframe, data
3. Extract useful columns and move them to dictionary for new dataframe
4. Filter dataframe to include only Falcon 9 launches
5. Handle missing data
6. Export data to CSV file

Data Collection - Scraping

1
- Request SpaceX launch data using GET request

2
- Normalize JSON response into a dataframe, data

3
- Extract useful columns and move them to dictionary for new dataframe

4
- Filter dataframe to include only Falcon 9 launches

Github link: Github

Data Wrangling

1
- Calculate the number of launches on each launch site

2
- Calculate the number and occurrence of each orbit

3
- Calculate the number and occurrence of mission outcome per orbit type

4
- Create a landing outcome label from Outcome column using replace

5
- Export data to CSV file

Github link: [Github]

EDA with Data Visualization

- Scatterplot – Visualize the relationship between two features.
    - Flight Number and Payload
    - Flight Number and Launch Site
    - Payload and Launch Site
    - Flight Number and Orbit Type
    - Payload and Orbit Type
- Bar graph – Visualize the relationship between different categorical groups
    - Orbit Type and Success Rate
- Line plot – Visualize the trend of a variable over time
    - Years and Success Rate

Github link: Github

EDA with SQL

- Performed various SQL queries to determine the following:
  - Types of landing outcomes
  - Total payload mass by NASA
  - Average payload mass by F9 booster
  - Successful launches with a specific payload (4000kg to 6000kg)
  - Boosters carrying the maximum payload
  - Successful launches between certain date range
- SQL allows use to filter data and find useful information about the dataset
- Github link: Github

Build an Interactive Map with Folium

- Objects were created and added to the Folium map
  - Marker objects
    - Used to show launch sites
    - Number of success/failed launches
  - Line objects
    - Calculate the distance between launch site and geographical feature
- Adding these objects allowed us to answer the following questions
  - Are launch sites in close proximity to railways? Yes
  - Are launch sites in close proximity to highways? Yes
  - Are launch sites in close proximity to coastline? Yes
  - Do launch sites keep certain distance away from cities? Yes
- Github Link: Github

Build a Dashboard with Plotly Dash

- The dashboard contains two visuals:
    - Pie Chart
        - Shows the percentage of successful launches from each site.
        - Visualize the success rate of each site to make appropriate insights
    - Scatterplot
        - Shows the relationship between landing outcome and payload mass
        - Filter booster type with dropdown
        - Slider allows user to see data of different payload ranges
- Github link: Github

Predictive Analysis (Classification)

1. Import data and create a label using 'Class' column
2. Preprocess feature dataset
3. Create train and test dataset
4. Determine the hyperparameter of different models (LogR, Tree, KNN, and SVM) using GridSearchCV
5. Determine accuracy of the different models
6. Plot the confusion matrix

Results

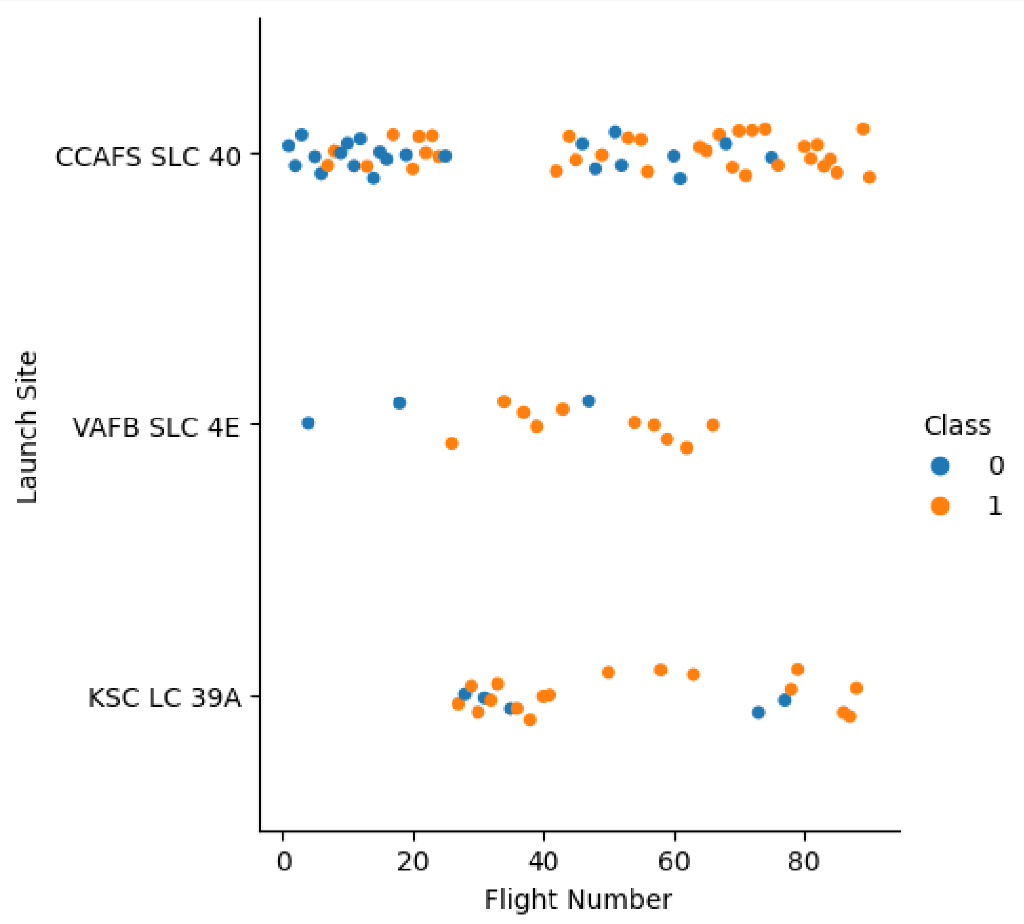
- From the EDA, the success rate of Falcon 9 landings were on average 66.6%
- The predictive analysis shows that the Decision tree model was the most optimal model with an accuracy of 89% and good confusion matrix

Section 2

Insights drawn from EDA

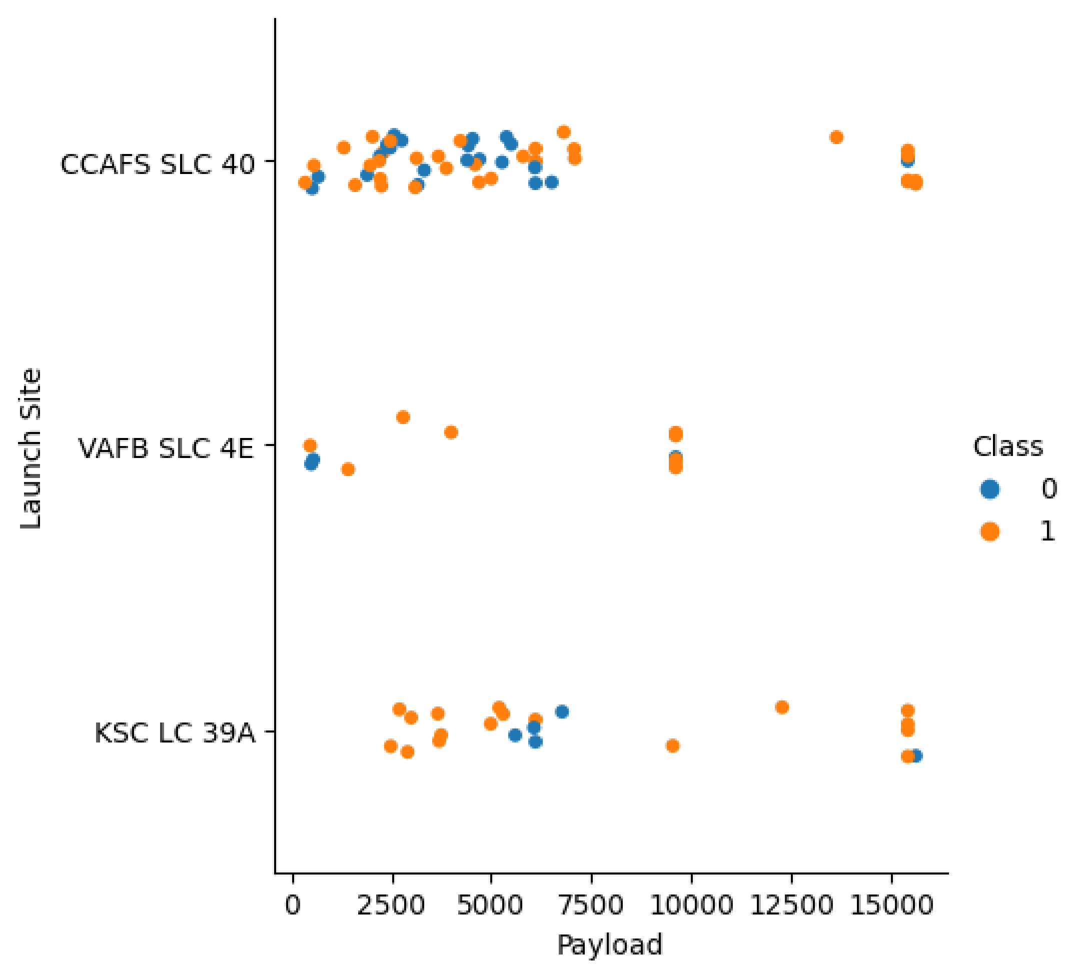
Flight Number vs. Launch Site

- The visualization shows that the success rate increased as the flight number increased
- There was an increase in the probability of successful flights after the 40^{th} launch



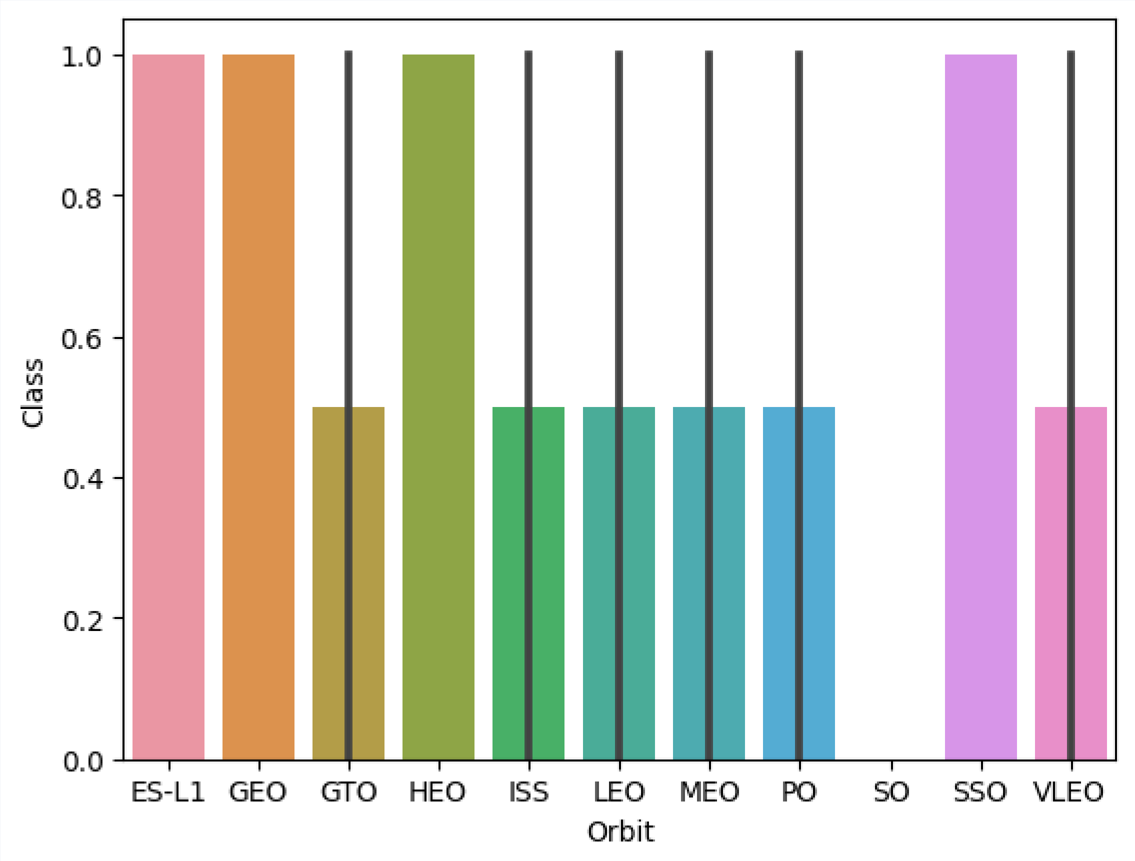
Payload vs. Launch Site

- The visualization shows that the success rate increased as the payload increased
- No rocket launches with >10000kg payload was carried out on VAFB-SLC launch



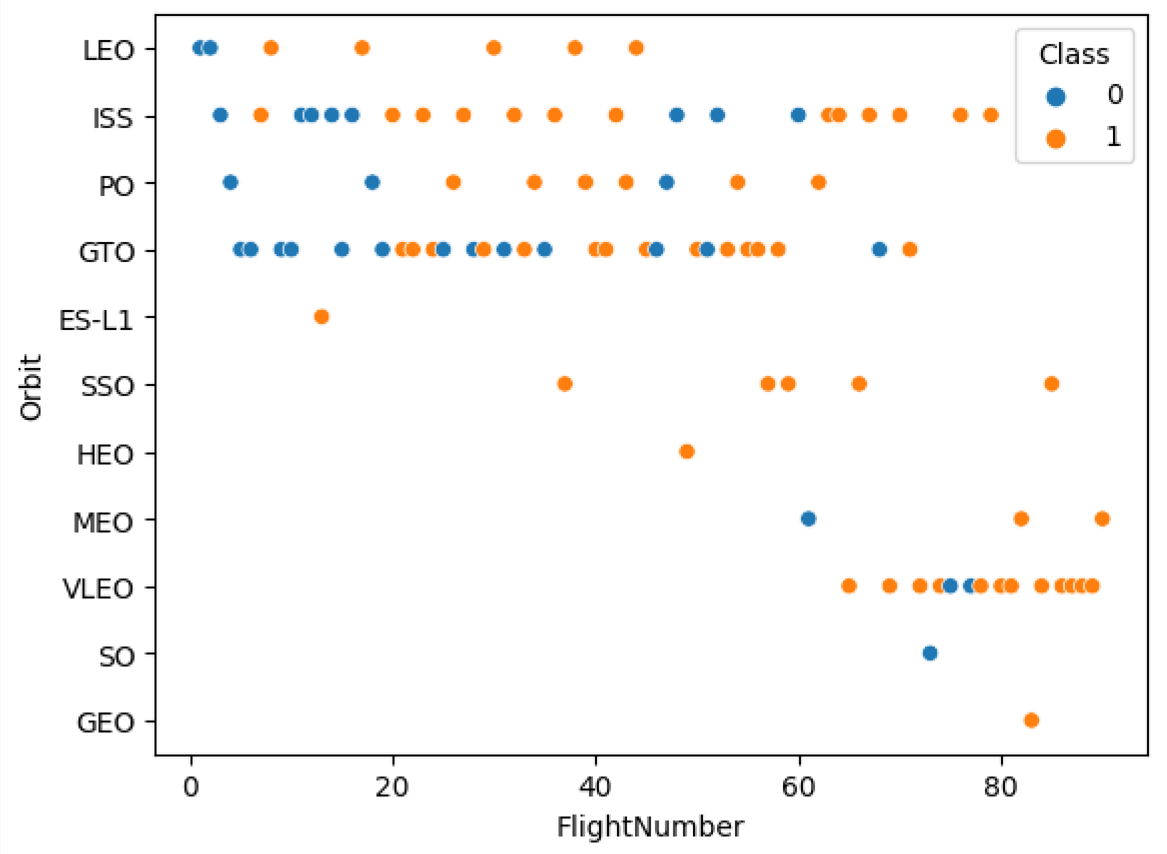
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbit launches had 100% successful launches
- SO orbit launches did not have any successful launches
- From this, Falcon 9 rocket may not be suitable for certain orbit launches



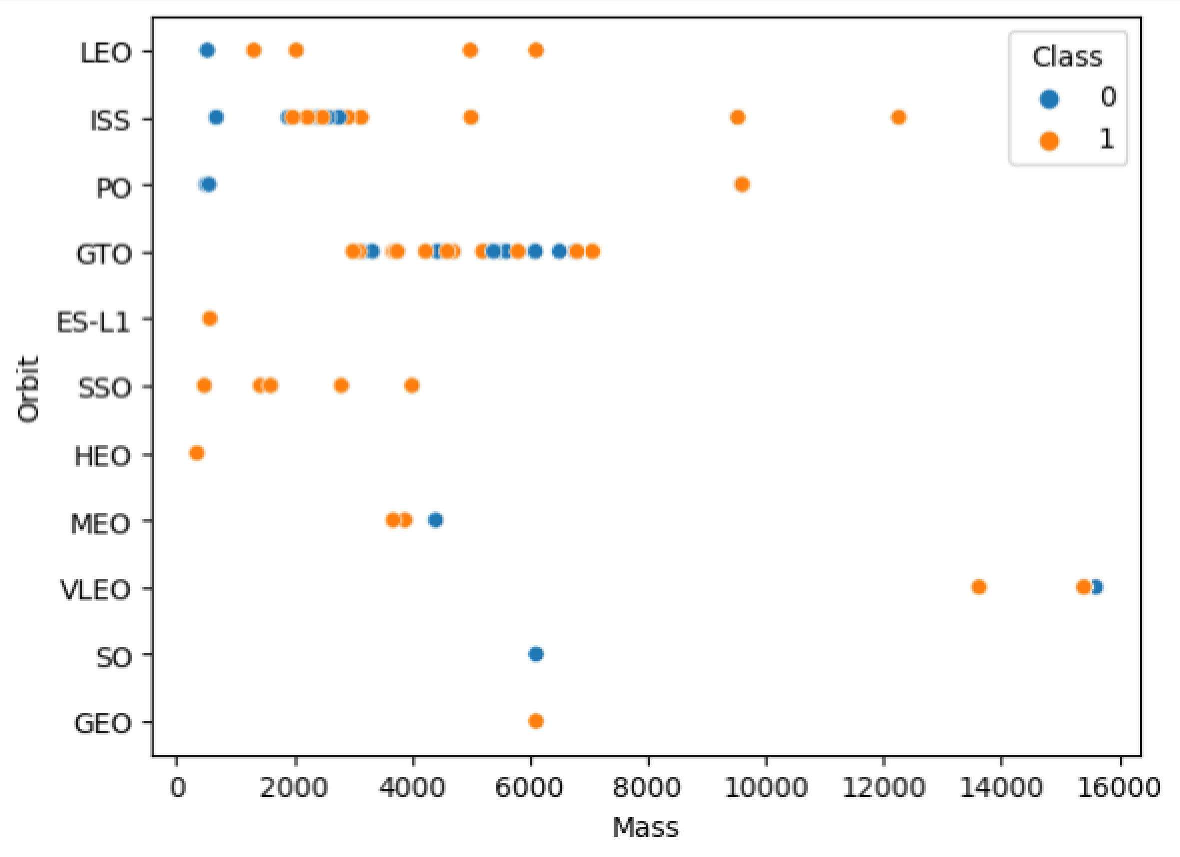
Flight Number vs. Orbit Type

- The visualization shows flight number may be associated to greater success with LEO orbit launches
- However, it seems conclusive that flight number have minimal impact on launch success rate



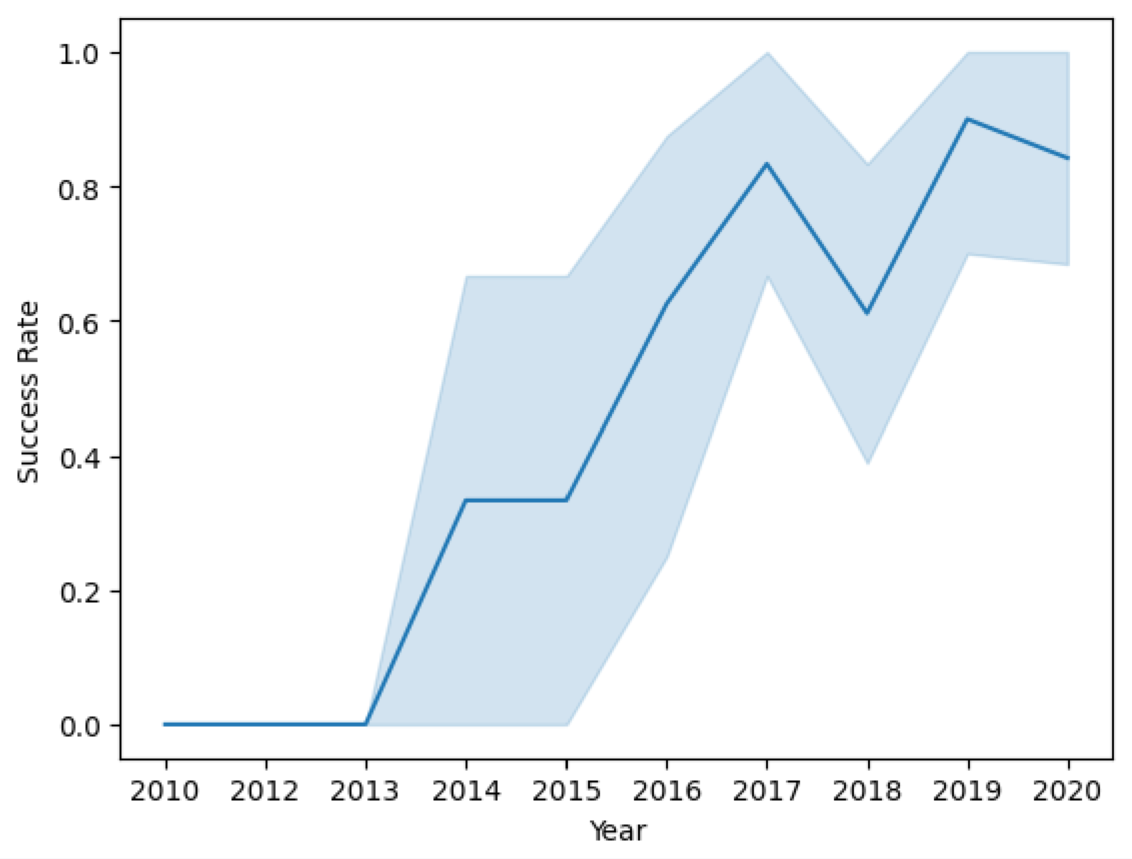
Payload vs. Orbit Type

- The visualization shows that higher payload in LEO, ISS, PO, and SSO led to more successful launches
- There appears to be no relationship between payload mass and success launches for GTO



Launch Success Yearly Trend

- The visualization shows that success rate has been increasing with respect to time until 2019.
- This is likely due to advancements in technology and deeper understanding of materials used in the rockets



All Launch Site Names

- The DISTINCT function was used to return unique launch site from the launch_site column
- The unique launch sites are provided in the image below.

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

Launch Site Names Begin with 'CCA'

- The LIKE and LIMIT functions were used to display the first 5 launch events in the dataset with a launch site name that begin with ‘CCA’

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The SUM function was used to calculate the total kilograms of payload that was carried by NASA boosters

sum(PAYLOAD_MASS__KG_)
111268

Average Payload Mass by F9 v1.1

- The AVG function was used to determine the average kilogram of payload carried by F9 v1.1 boosters
- The WHERE and LIKE functions was used to filter boosters with F9 v1.1

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The MIN function was used to find the data of the first successful landing on ground pad
- The LIKE function was used to filter out the data with only values containing ground pad

min(date)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- The BETWEEN function was used to retrieve data in the dataset where the payload was between 4000kg and 6000kg
- The WHERE function was used to filter out only successful launches

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- The COUNT function is used to count the number of occurrences of different mission outcome, which are grouped together with GROUPBY function.
- Out of 101 launches, 99 of the missions were sucessful

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

Boosters Carried Maximum Payload

- The MAX function was used in a subquery to retrieve a list of the booster that have carried the maximum payload of 15600kg

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 YEAR function was used to isolate the year from the date column and WHERE function was used to isolate 2015 entries.

01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
06	2015	F9 v1.1 B1018	CCAFS LC-40	Precluded (drone ship)

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

- The WHERE and BETWEEN Function was used to isolate mission between 2010-06-04 and 2017-03-20

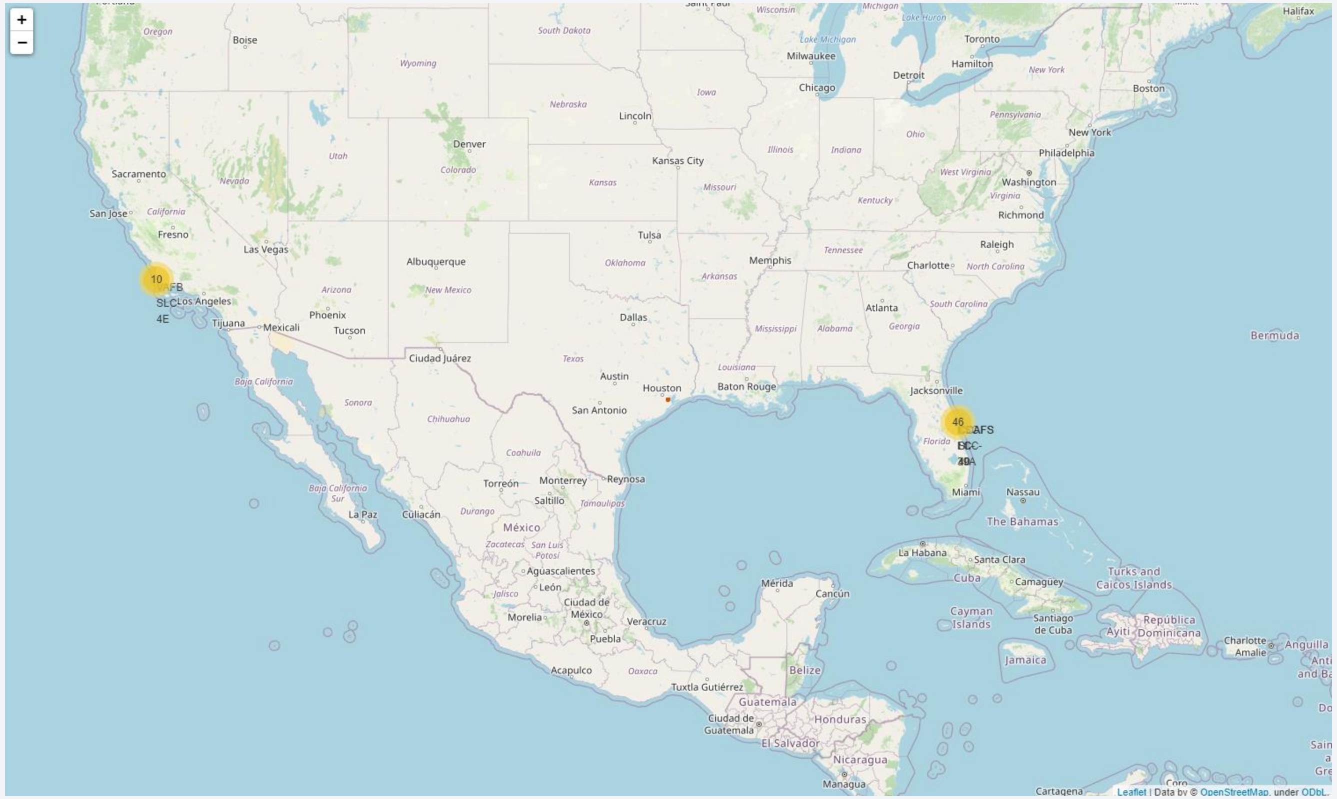
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
08-04-2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
07-09-2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (ground pad)
09-10-2017	12:37:00	F9 B4 B1041.1	VAFB SLC-4E	Iridium NEXT 3	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
08-01-2018	01:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000	LEO	Northrop Grumman	Success (payload status unclear)	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)
11-05-2018	20:14:00	F9 B5 B1046.1	KSC LC-39A	Bangabandhu-1	3600	GTO	Thales-Alenia/BTRC	Success	Success (drone ship)
07-08-2018	05:18:00	F9 B5 B1046.2	CCAFS SLC-40	Merah Putih	5800	GTO	Telkom Indonesia	Success	Success
10-09-2018	04:45:00	F9 B5B1049.1	CCAFS SLC-40	Telstar 18V / Apstar-5C	7060	GTO	Telesat	Success	Success

Section 3

Launch Sites Proximities Analysis

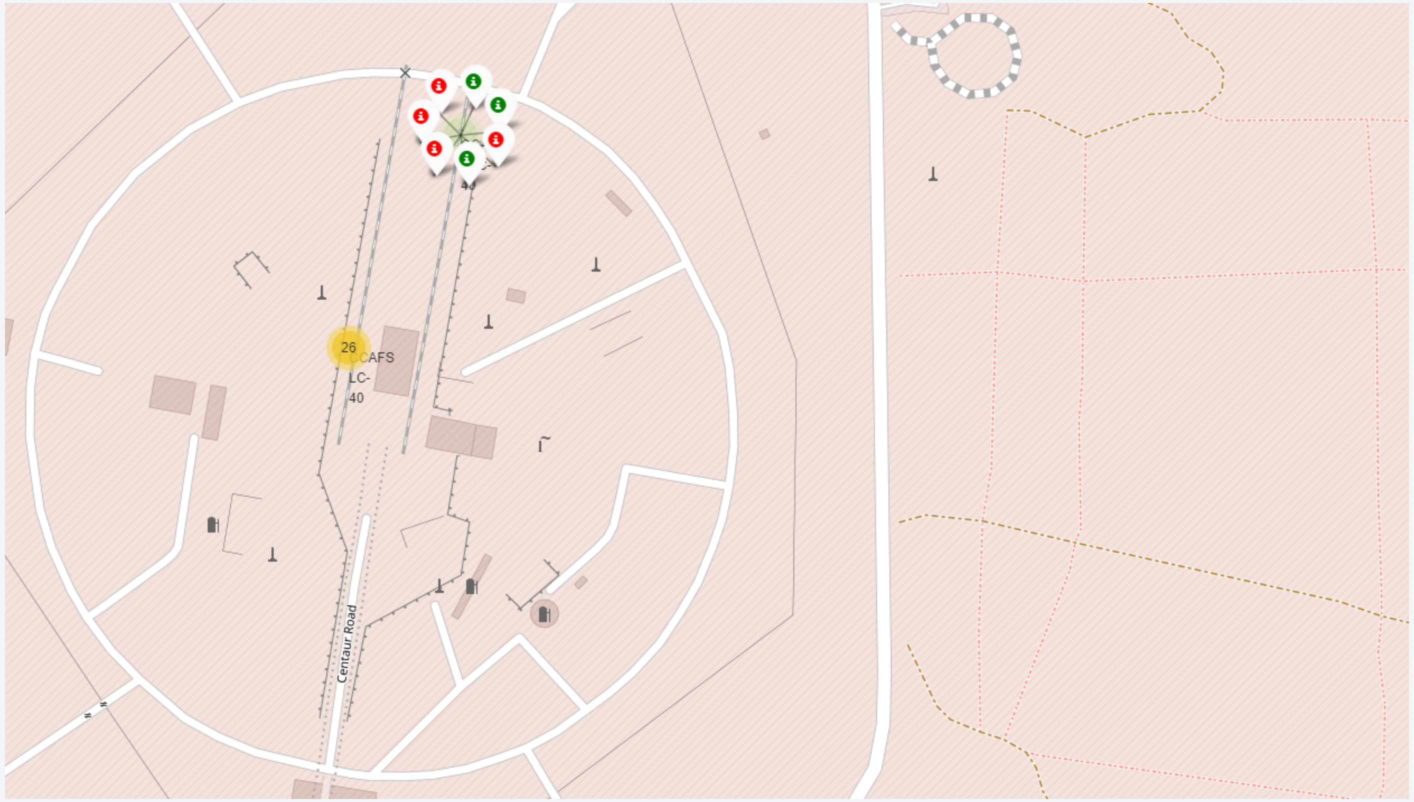
Launch Sites

- The yellow markers locations of the sites
- The number represents the number of launches on each site
- Most of the launches are occurred on the coastlines



Marker Details

- We can click the yellow marker to expand the launches
- The success and fail are highlighted as either red or green



Proximities to geographical features

- The map on the right shows the distance of geographical features from the launch site.
- Launch sites are close to railroad, highways and coastlines, but far from metropolitan areas

Section 4

Build a Dashboard with Plotly Dash

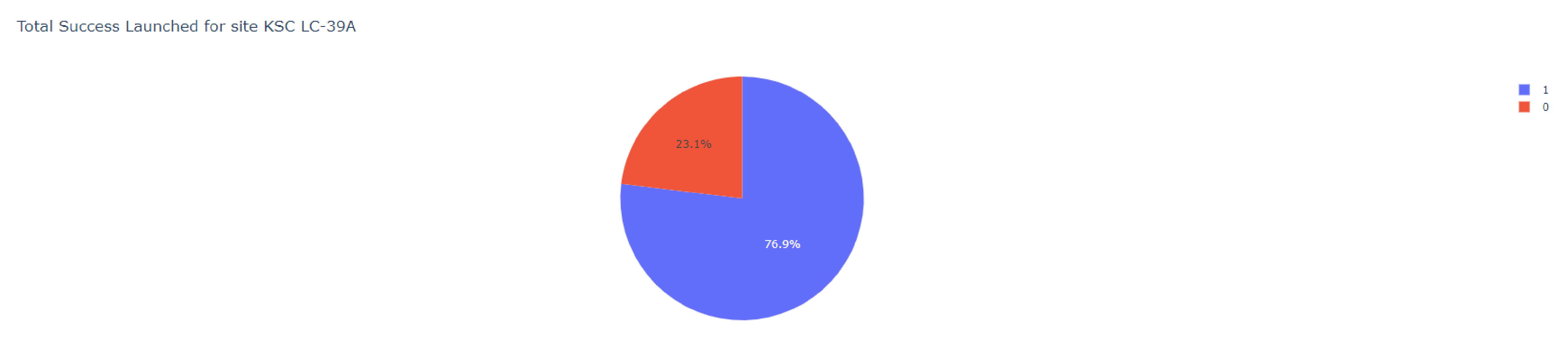
Success Launches by Site

- The visualization below shows the percentages of success from each site
- KSC LC-39A produced the most successful launches

Total Success Launches By Site

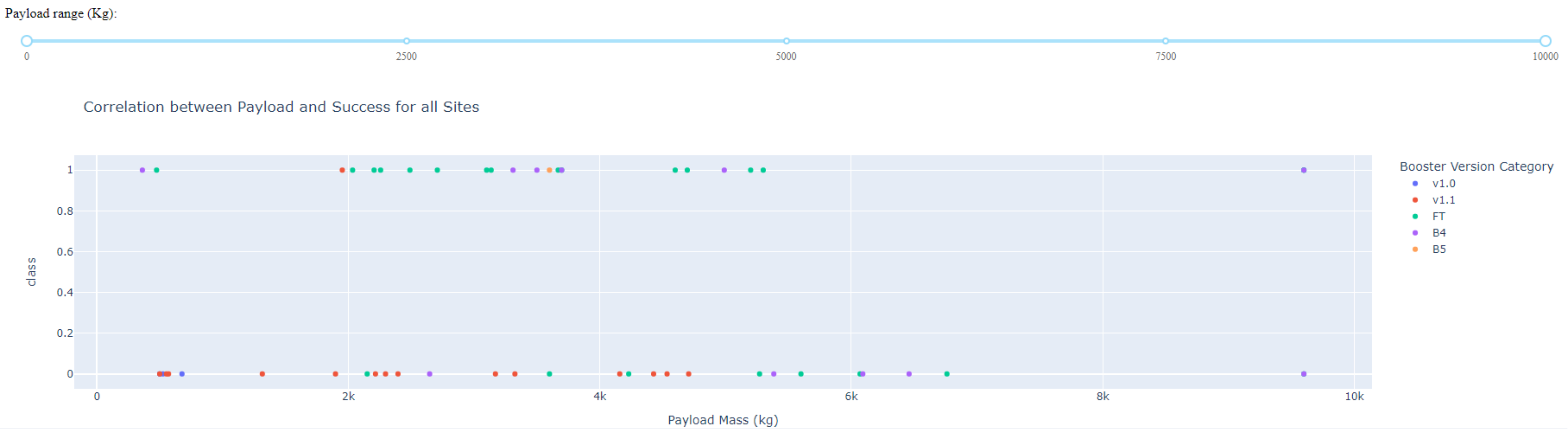
Success Rate of KSC LC-39A

- Diving deeper into KSC LC-39A, the site has a success rate of 76.9%



Scatterplot between Payload Mass and Success

- Successful launches appear to be mostly concentrated with payload ranging between 2000 and 6000kg



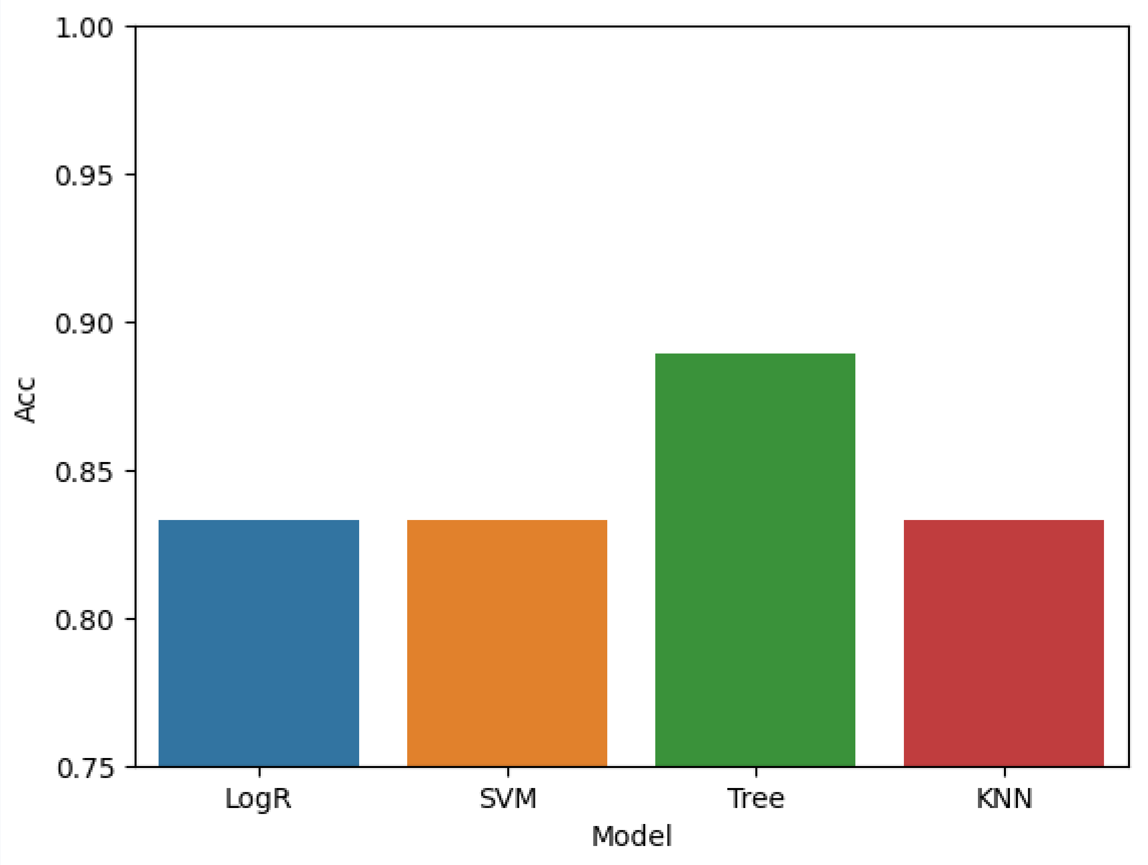
Section 5

Predictive Analysis (Classification)

Classification Accuracy

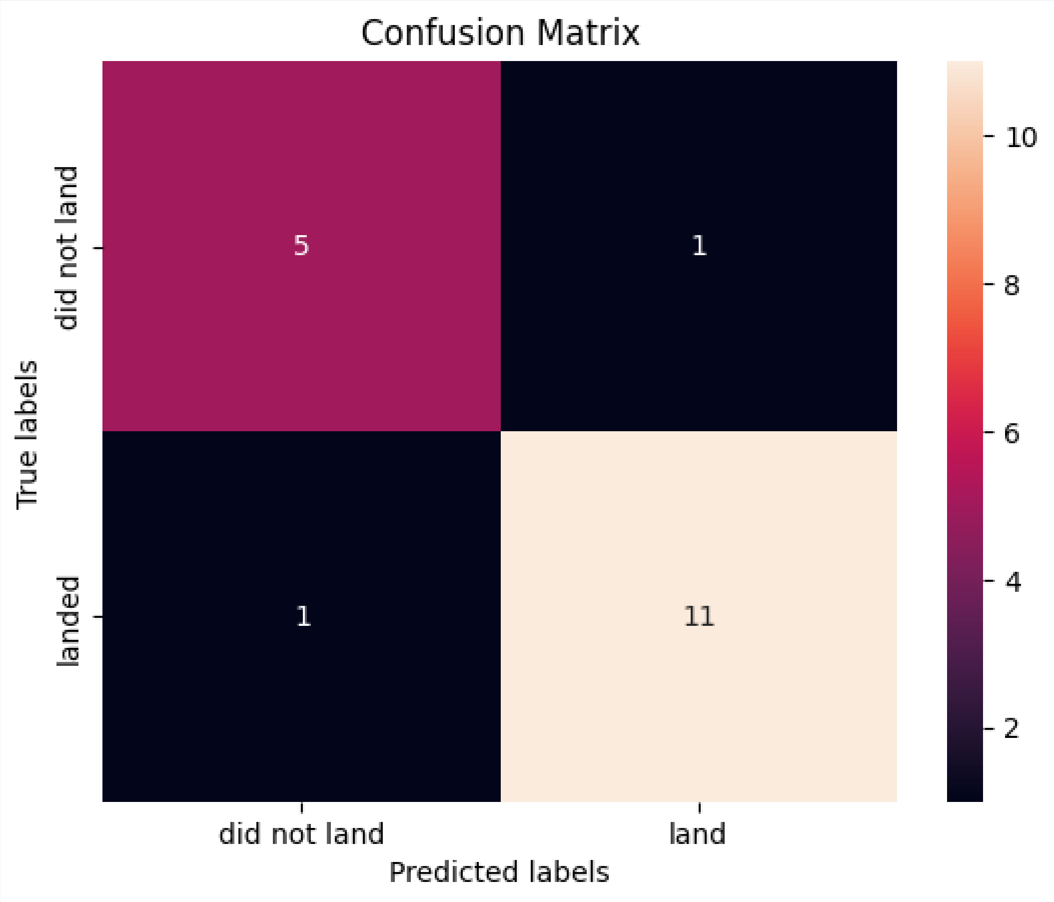
- The Decision Tree model was the most accuracy with 89%

Model	Accuracy
LogR	0.833333
SVM	0.833333
Tree	0.888889
KNN	0.833333



Confusion Matrix

- The confusion matrix of the tree model highlights that the model was extremely good at predicting success or failures with minimal false positives and false negatives



Conclusions

- There was a strong positive correlation between the number of flights and success rate.
- The success rate of launches increased significantly with respect to time. Likely due to technological advances.
- Payloads between 2000kg and 6000kg proved to be most ideal for successful launches
- Falcon 9 rocket proved favorable for LEO, ISS, PO, and SSO orbital launches
- The best model to predict the success of launch is with Decision Tree, with an accuracy of 89%

Appendix

- https://github.com/Christopherc0923/IBM-Data-Science-Cert

Thank you!