IBM Applied Data Science Capstone

SpaceX Falcon 9 First Stage Landing Prediction

PRESENTED BY: DANIEL GRILLO

2021-10-15

Table of Contents

Executive Summary

This applied data science capstone project attempts to predict if the SpaceX Falcon 9 first stage rocket launch will land successfully.

Several techniques were implemented in order to make predictions with the highest accuracy possible, some of these techniques being EDA (exploratory data analysis), web scrapping and data wrangling, visual analytics tools, and machine learning algorithms for predicting outcomes.

Introduction

The goal for this capstone project was to successfully predict the outcome of the SpaceX Falcon 9 first stage rocket launch to determine the probability of a successful rocket landing.

Considering the high costs of such procedures, utilizing data science to determine the landing outcome of a Falcon 9 rocket can significantly reduce the monetary costs for SpaceX.

The average cost of a Falcon 9 rocket launch is $62 million dollars, therefore if we can predict the landing outcome of a rocket launch before launching, this could benefit the people in charge of such operations.

How can we use data science to predict the landing outcome of SpaceX Falcon 9 rocket launches?

Methodology

Several labs were completed in order to obtain the data utilized throughout this entire capstone project, perform exploratory data analysis (EDA), better visualize information and relationships presented in the dataset, as well as determine the best machine learning algorithm and hyperparameters in order to make accurate predictions regarding the first stage rocket launch of SpaceX Falcon 9 rocket.

These labs are:

- Data Collection Lab
  - Request and parse the SpaceX launch data using GET request
  - Turn this data into JSON format, to then be converted into a data frame
  - Filter the data to only include Falcon 9 launches
  - Handle missing values

- Data Wrangling Lab
  - Calculate number of launches per site
  - Calculate number of occurrences per orbit
  - Calculate number and occurrence of mission outcome per orbit type
  - Create a landing outcome label from Outcome column

Methodology

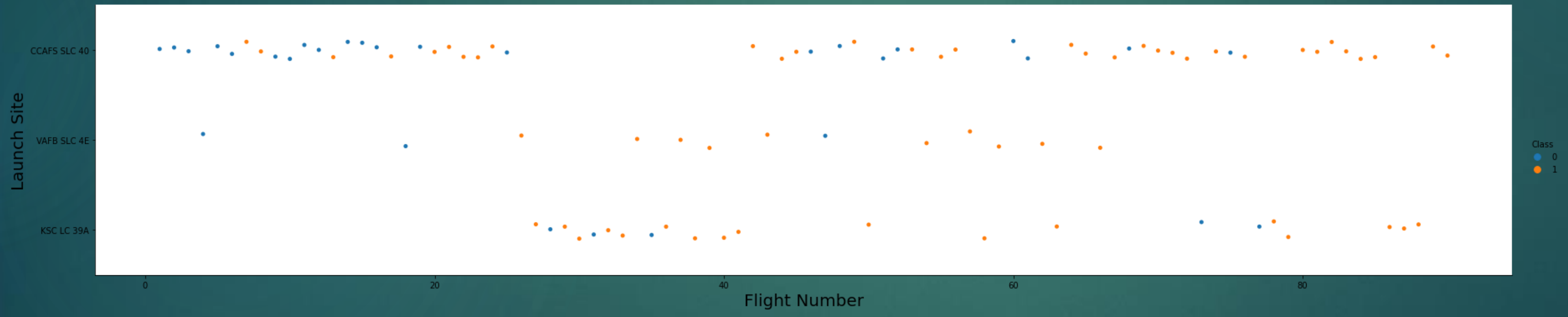
- EDA with Visualization Lab
  - Visualize the relationship between Flight Number and Launch Site
  - Visualize the relationship between Payload and Launch Site
  - Visualize the relationship between Success Rate of each Orbit Type
  - Visualize the relationship between Flight Number and Orbit Type
  - Visualize the relationship between Payload and Orbit Type
  - Visualize the Launch Success yearly trend

- Machine Learning Predictions Lab
  - Standardize the data
  - Split the data into training data and testing data
  - Create a Logistic Regression object, SVM object, Decision Tree Classifier object, and a KNN object
  - Fit a GridSearchCV object to all these objects, and fit data to find best parameters

Results

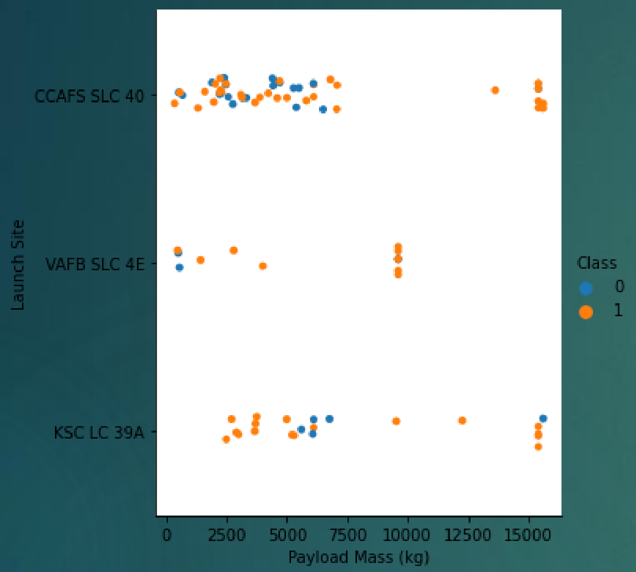
We started this capstone project by requesting the rocket launch data from the SpaceX API, followed by converting this JSON data into a data frame for easier handling, and lastly, we filtered the data to the launches relevant to Falcon 9.

From the Exploratory Data Analysis with Visualization Lab, we obtained the following results:

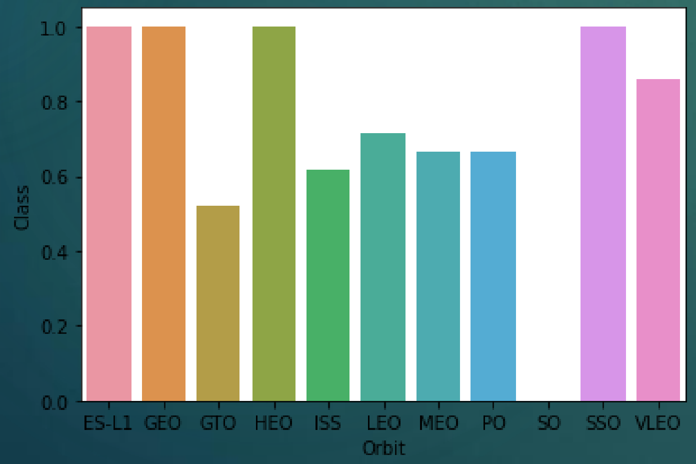


Relationship between Flight Number and Launch Site

Results

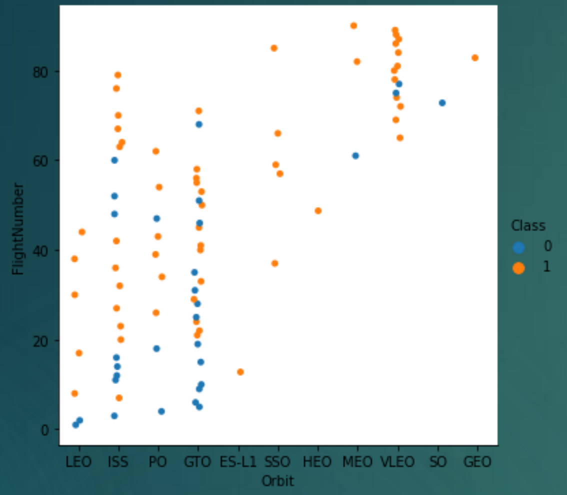


Relationship between Payload and Launch Site

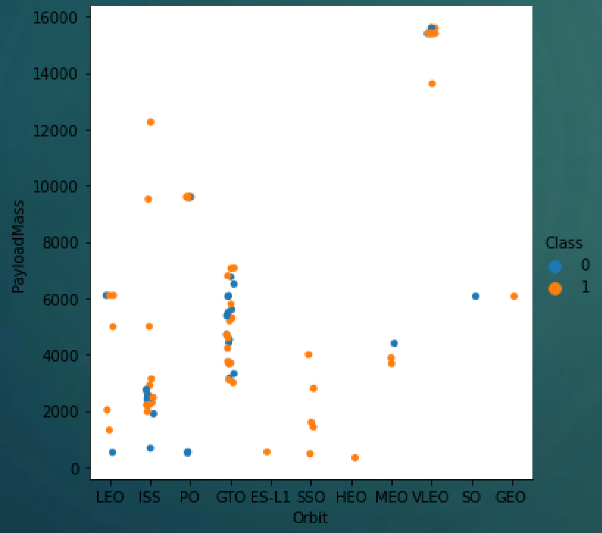


Relationship between Success Rate of each Orbit Type

Results

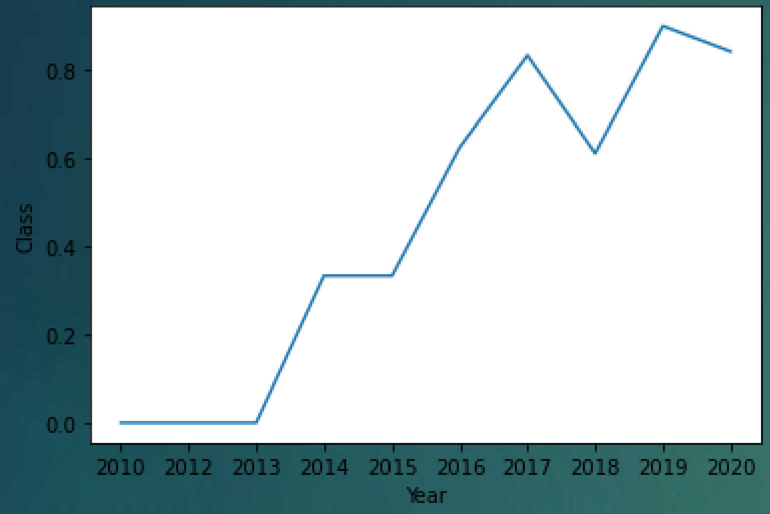


Relationship between Flight Number and Orbit Type



Relationship between Payload and Orbit Type

Results



Launch Success yearly trend visualization

From the Exploratory Data Analysis with SQL Lab, we obtained the following results:

['CCAFS LC-40', 'VAFB SLC-4E', 'KSC LC-39A', 'CCAFS SLC-40']

Unique launch sites in space mission

Results

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

5 records where launch sites begin with the string ‘CCA

45596 kg Total payload mass carried by boosters launched by NASA (CRS)

2928.4 kg Average payload mass carried by booster version F9 v1.1

2010-04-06 00:00:00 Date when first successful landing outcome in ground pad was achieved

Results

['F9 FT B1022', 'F9 FT B1026', 'F9 FT B1021.2', 'F9 FT B1031.2']

Names of boosters which have success in drone ship and have payload greater than 4000kg, but less than 6000kg

Successful Mission Outcomes: 98
Failed Mission Outcomes: 1

Total number of successful and failure mission outcomes

['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']

Names of the booster versions which have carried maximum payload mass

Results

	Booster_Version	Launch_Site
13	b'F9 v1.1 B1012'	CCAFS LC-40
16	b'F9 v1.1 B1015'	CCAFS LC-40

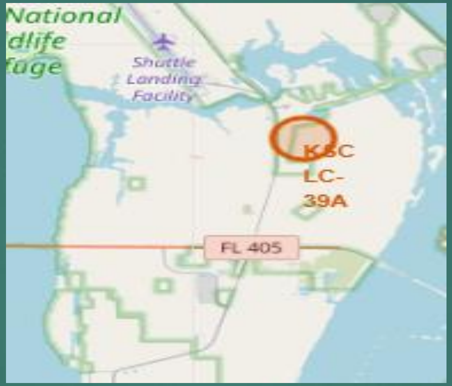
Booster versions and launch sites for failed landing outcomes in drone ship, for 2015

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

Count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order

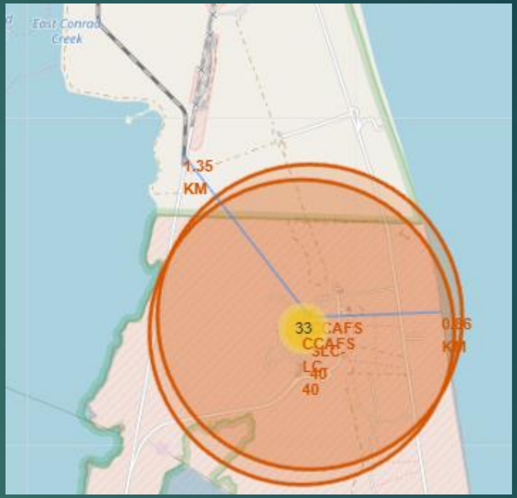
Results

From the Launch Sites Location Analysis with Folium Lab, we obtained the following results:



Launch Sites Locations on Map

Results

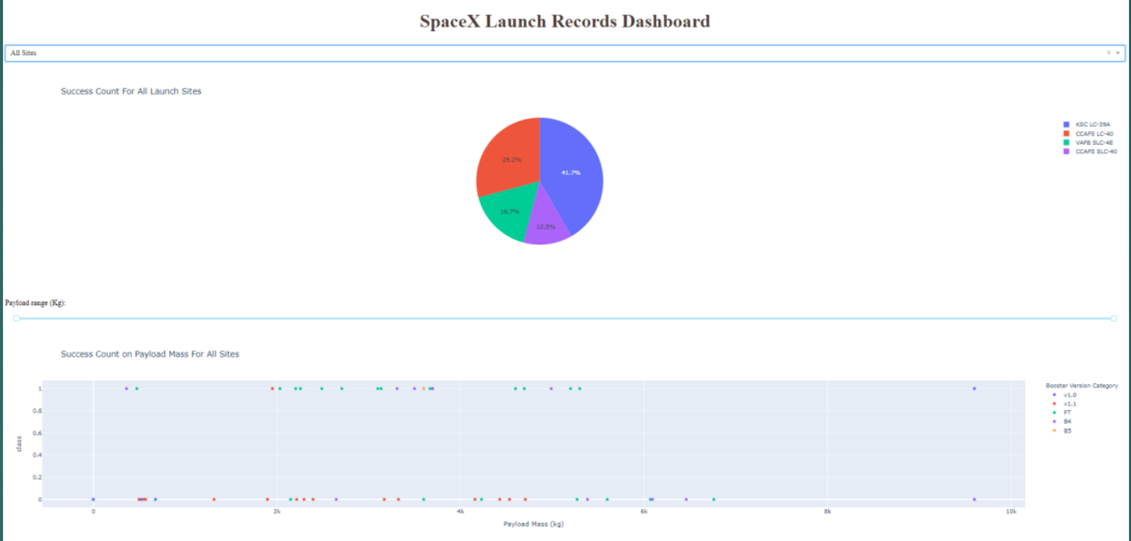


All the successful/failed launches for each site

Distance to closest railroad and coastline

Results

From the Plotly Dash Dashboard Lab, we obtained the following results:



This in an interactive dashboard generated utilizing Dash and Plotly (Python libraries). It contains a dropdown menu at the top of the dashboard, which allows the user to select the specific launch site and view its statistics or choose to view all the launch sites and compare them. It contains a range slider at the bottom to change the payload amount carried during the launch, as well as a scatter plot showing the success rate based on the payload amount selected.

Results

From the SpaceX Falcon 9 First Stage Landing Prediction Lab, we obtained the following results:

	Logistic Regression	**Support Vector Machine**	**Decision Tree Classifier**	**K Nearest Neighbors**
Training data accuracy	0.84643	0.84821	0.88750	0.84821
Testing data accuracy	0.83333	0.83333	0.55556	0.83333

Discussion

The data collection procedures were necessary to obtain the launch data from the SpaceX API, as well as the techniques utilized to transform the JSON data into a data frame for easier handling and visualization, and the filtering procedures to only work with the data relevant to the subject.

Data wrangling was a very important part of this project in order to analyze and better observe the trends within this data, this allowed noticing key factors that contribute to success/failure of Falcon 9 first stage rocket launch.

Creating visualizations is a great tool to better understand the data we are working with, in addition to graphically represent relationships between variables in the data set obtained.

Lastly, the different machine learning models used allowed us to see how different models perform predictions with the training and testing data, therefore we were able to choose the one with the highest overall accuracy to further predict the outcome of Falcon 9 first stage rocket launching.

Conclusion

This applied capstone project was a great learning experience to better understand the key characteristics of a data science project and how to tackle problems with a data scientist mentality.

The tools implemented in this capstone are very useful in several different ways, from collecting data necessary to find solutions to problems, filtering the data to use relevant information, to graphically represent the insights found on this data and output it in a graphical and user-friendly manner so that third parties can easily understand the relationships and key features observed.

The content in this course will be very useful when approaching different problems since this is a very detailed breakdown of a topic (Falcon 9 first stage rocket launching prediction), and how to use all the tools learned throughout the entire IBM Data Science Professional Certificate.

Thank you for your time