Applied Data Science Capstone

Sean Lin 03.02.2023

Outline

- Executive Summary
- Introduction
- Applied Methodologies
- Results
- Conclusion

Executive Summary

- Predict successful or not the SpaceX Falcon 9 will be landed
- The research process includes:
  - Data collection and wrangling
  - Exploratory data analysis (EDA)
  - Interactive visual analytics
  - Predictive analysis using Machine Learning
- Conclusion:
  - Decision Tree model is the most affective in this research
  - Lower payload mass has better success rate
  - Launch sites tend to locate near Equator line and by the coast
  - Success rate has been improving
  - KSC LC-39A has the highest success rate

Introduction

- This capstone project aims to predict the success of SpaceX's Falcon 9 first stage landing. The company's low launch cost of 62 million dollars compared to other providers' cost of 165 million dollars or more is largely due to the reusable first stage.
- The project will use public information and machine learning models to determine the success of the first stage landing and its cost, which can be used by other companies bidding against SpaceX for a rocket launch.
- The main question being addressed is whether the set of features such as payload mass, launch site, orbit type, etc., will result in a successful landing of the Falcon 9 first stage.

Applied Methodologies

Data collection and data wrangling methodology

- Data collection
  - SpaceX API
    - Acquire data using SpaceX custom functions
  - Web scraping from Wikipedia
    - Acquire date using BeautifulSoup from HTML response

Applied Methodologies

Data collection and data wrangling methodology

- Data wrangling
  - Process, clean, and combine the data we previously retrieved
  - Create an extra column to label the landing results to "1" as succeed and "0" as failed

Applied Methodologies

EDA and interactive visual analytics methodology

- Exploratory data analysis (EDA)
  - Pandas & NumPy
    - Basic cleaning, calculating, and manipulating data
  - SQL
    - Perform queries to display, list, or count data

Applied Methodologies

EDA and interactive visual analytics methodology

- Interactive visual analytics
  - Matplotlib & Seaborn
    - Using scatter plots, bar charts, and line charts to present the relationships between attributes
  - Folium
    - Plotting interactive maps, furthermore, adding markers and line to display launch sites and their characteristics
  - Dash
    - Create interactive site with dropdown list and slider to exhibit pie charts and scatter plots based on the input data from user

Applied Methodologies

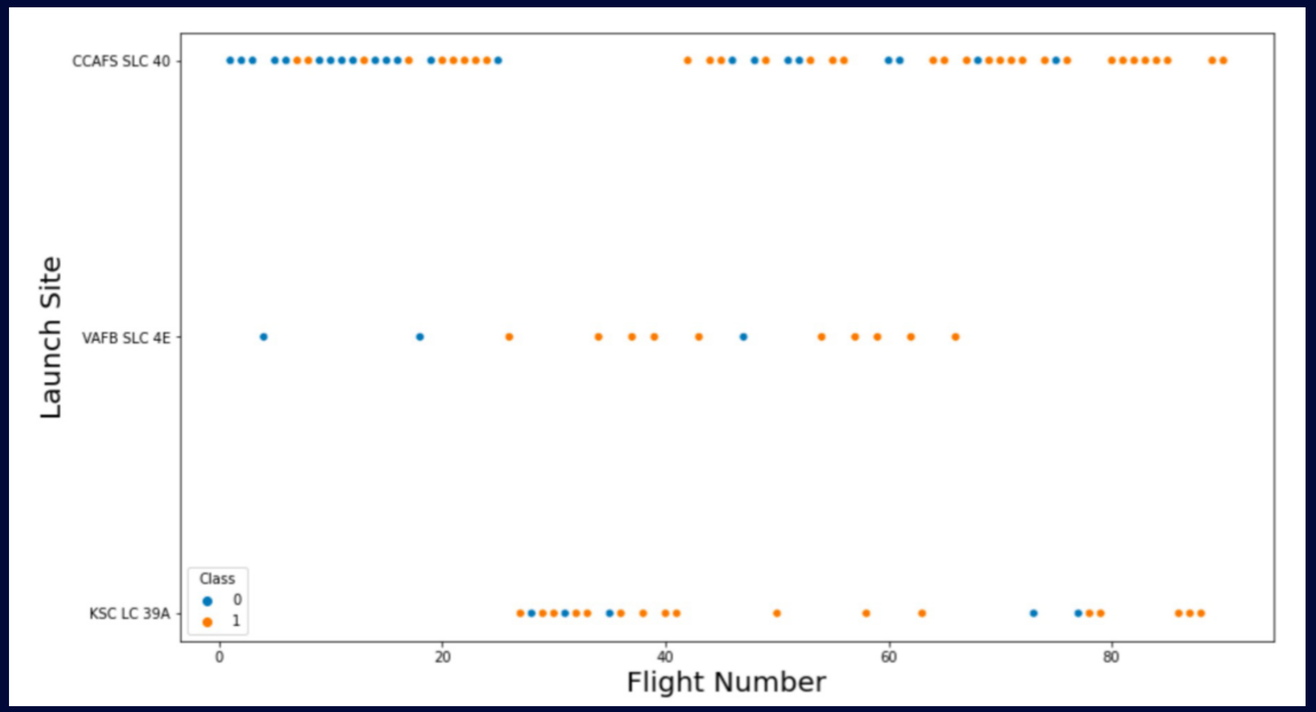
EDA and interactive visual analytics methodology

- Predictive analysis methodology using Scikit
  - Standardize the data
  - Split the data into training and testing sets
  - Apply GridSearchCV on:
    - Logistic regression
    - Support vector machine (SVM)
    - Decision tree
    - K nearest neighbors (KNN)
  - Calculate the accuracy on the test sets and exam the confusion matrix
  - Conclude the best ML model based on Jaccard_score and F1_score

Results

EDA with visualization results

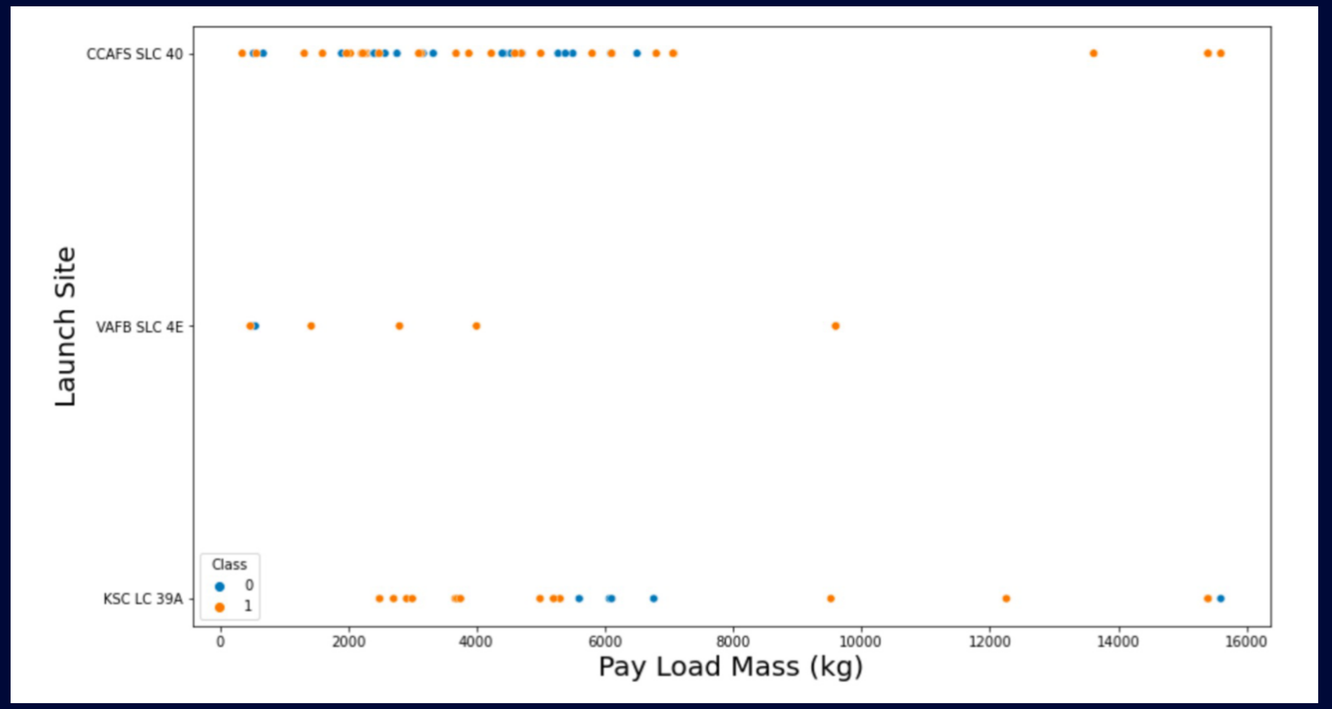
- Launch site & Flight Num.



Results

EDA with visualization results

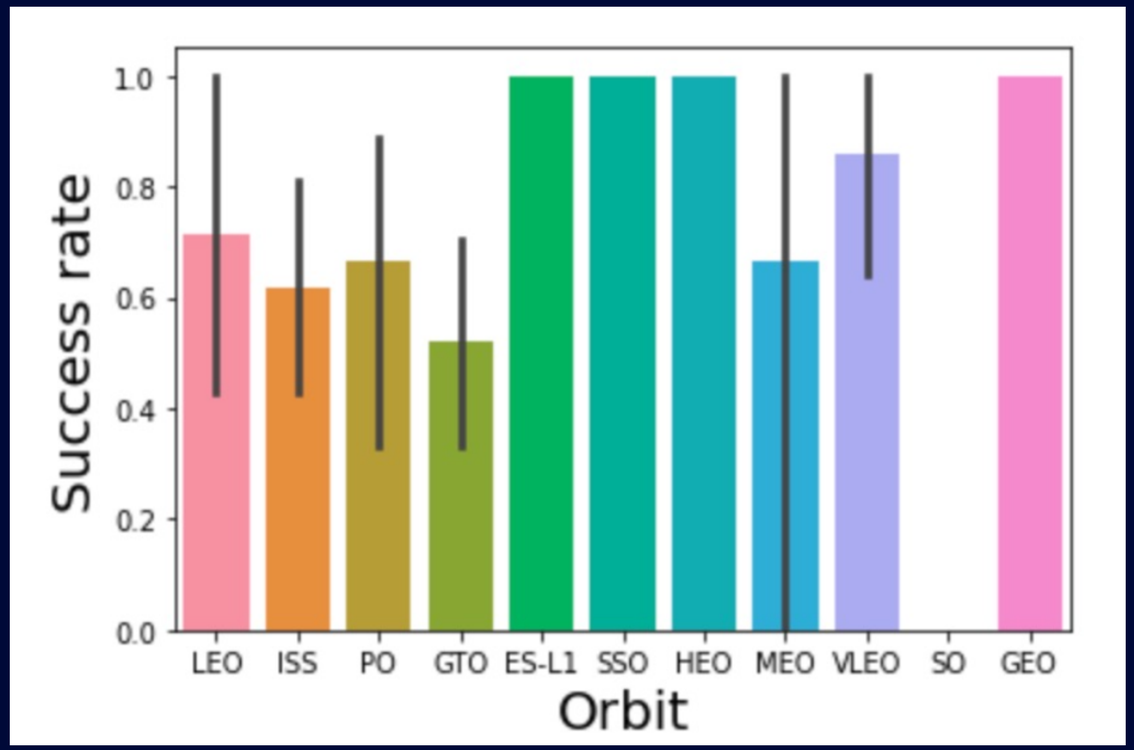
- Launch site & Pay Load (kg)



Results

EDA with visualization results

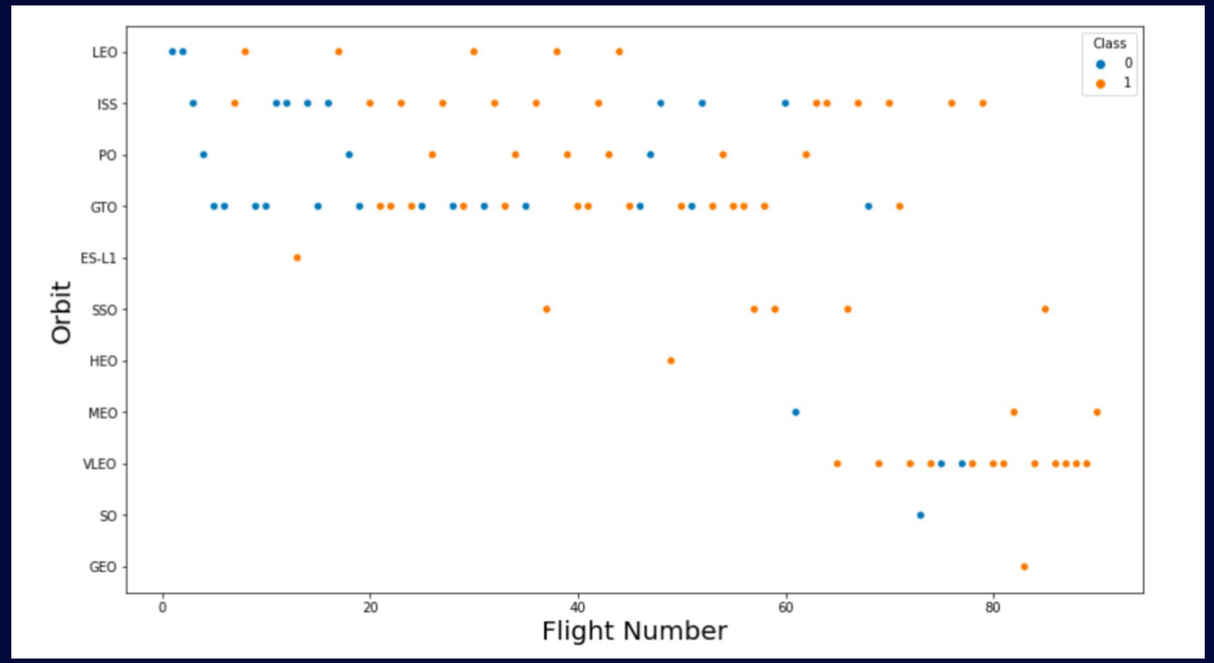
- Success Rate & Orbit type



Results

EDA with visualization results

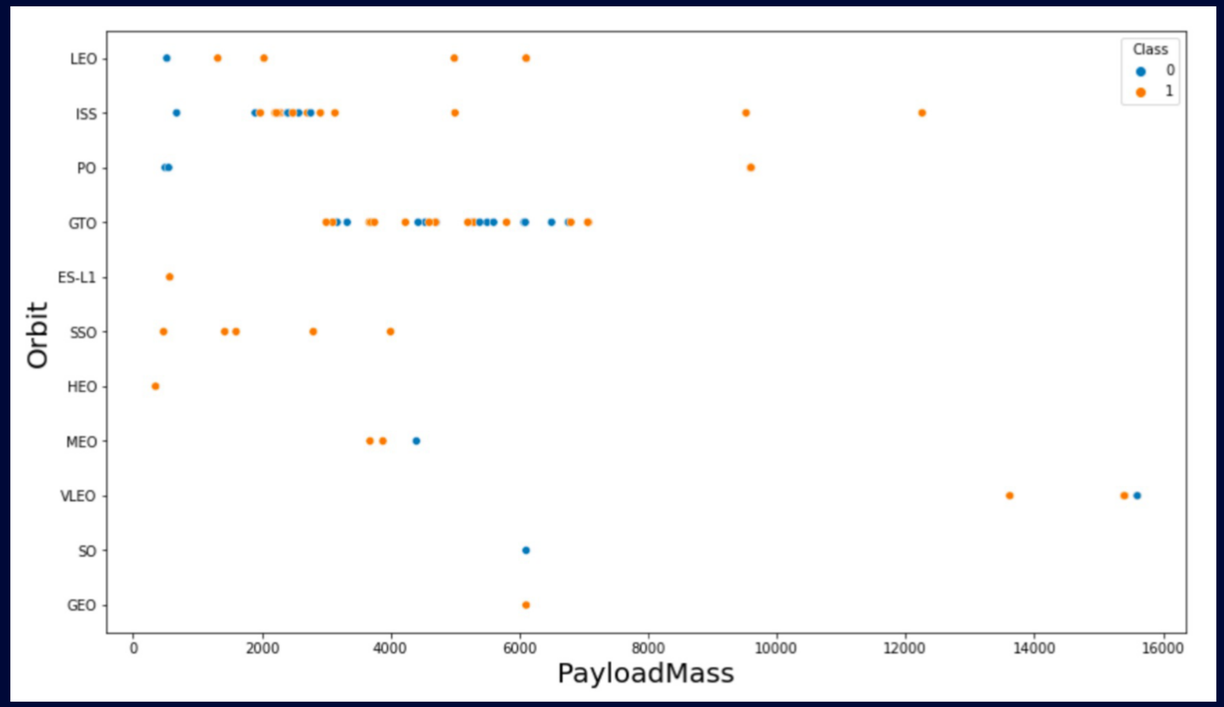
- Flight Num. & Orbit type



Results

EDA with visualization results

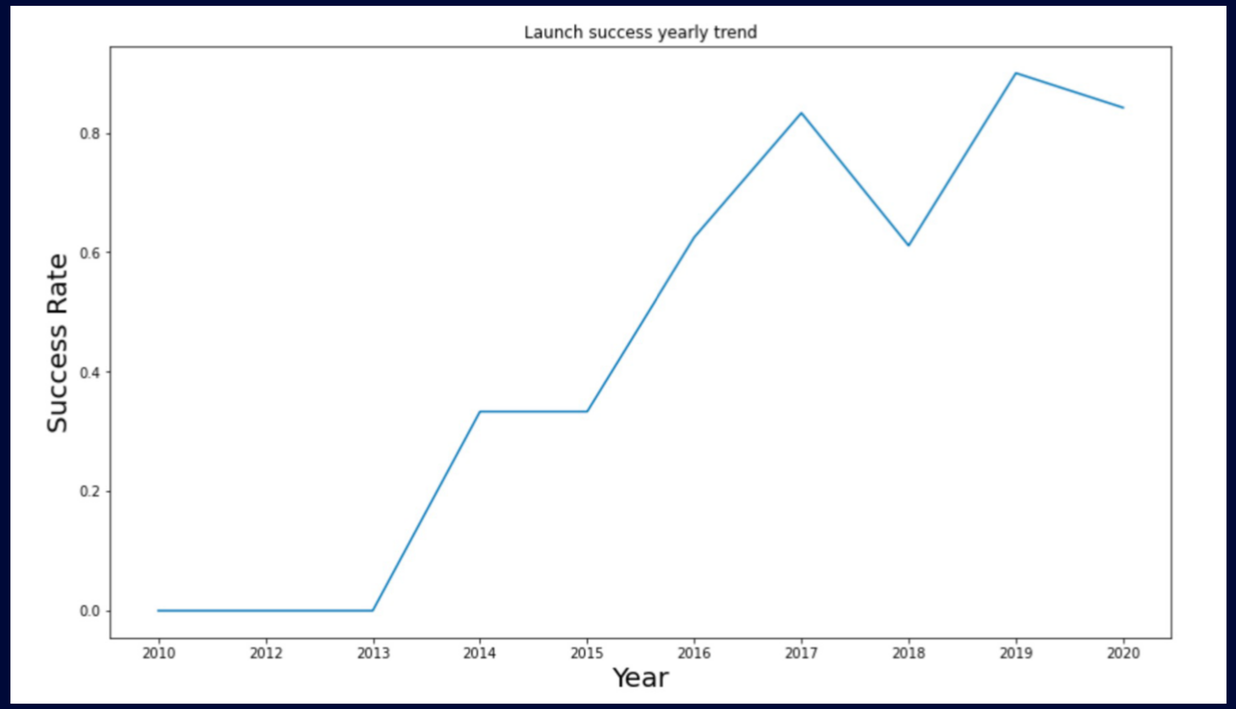
- Pay Load (kg) & Orbit type



Results

EDA with visualization results

- Success Rate by year



Results

EDA with SQL results

- The names of the unique launch sites in the space mission

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

- 5 records where launch sites begin with ‘CCA’

Total payload mass by NASA (CRS)
45596

- Total payload mass carried by NASA

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

Results

EDA with SQL results

- Average payload mass carried by booster version F9 v1.1

Average payload mass by Booster Version F9 v1.1
2928

- Date when the first successful landing outcome in ground pad was achieved

Date of first successful landing outcome in ground pad
2015-12-22

- The boosters which have success in drone ship and have 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

Results

EDA with SQL results

- The total number of successful and failure mission outcomes

number_of_success_outcomes	number_of_failure_outcomes
100	1

- The names 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

Results

EDA with SQL results

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

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

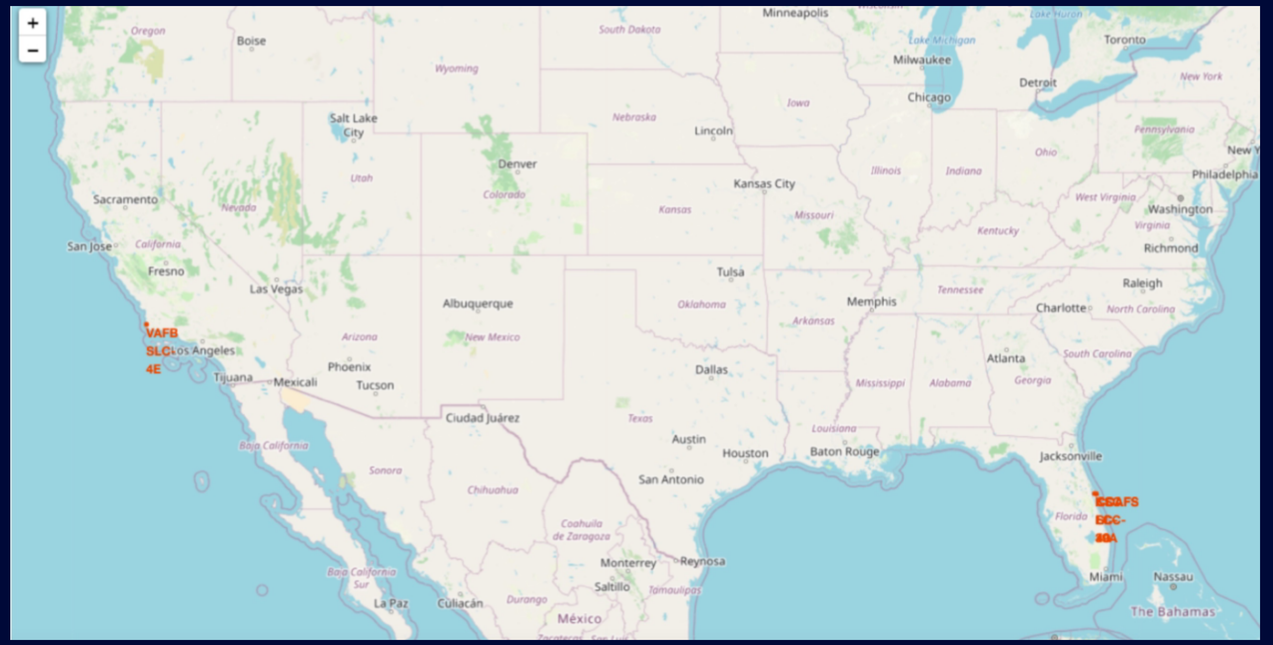
- The count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	landing_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

Results

Interactive map with Folium results

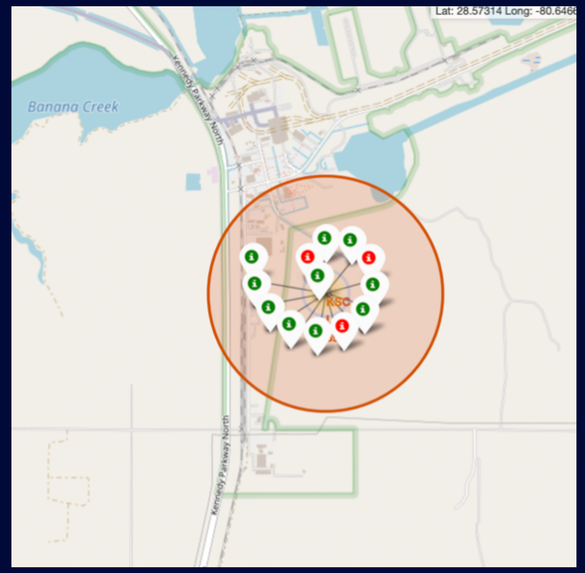
- Launch Sites



Results

Interactive map with Folium results

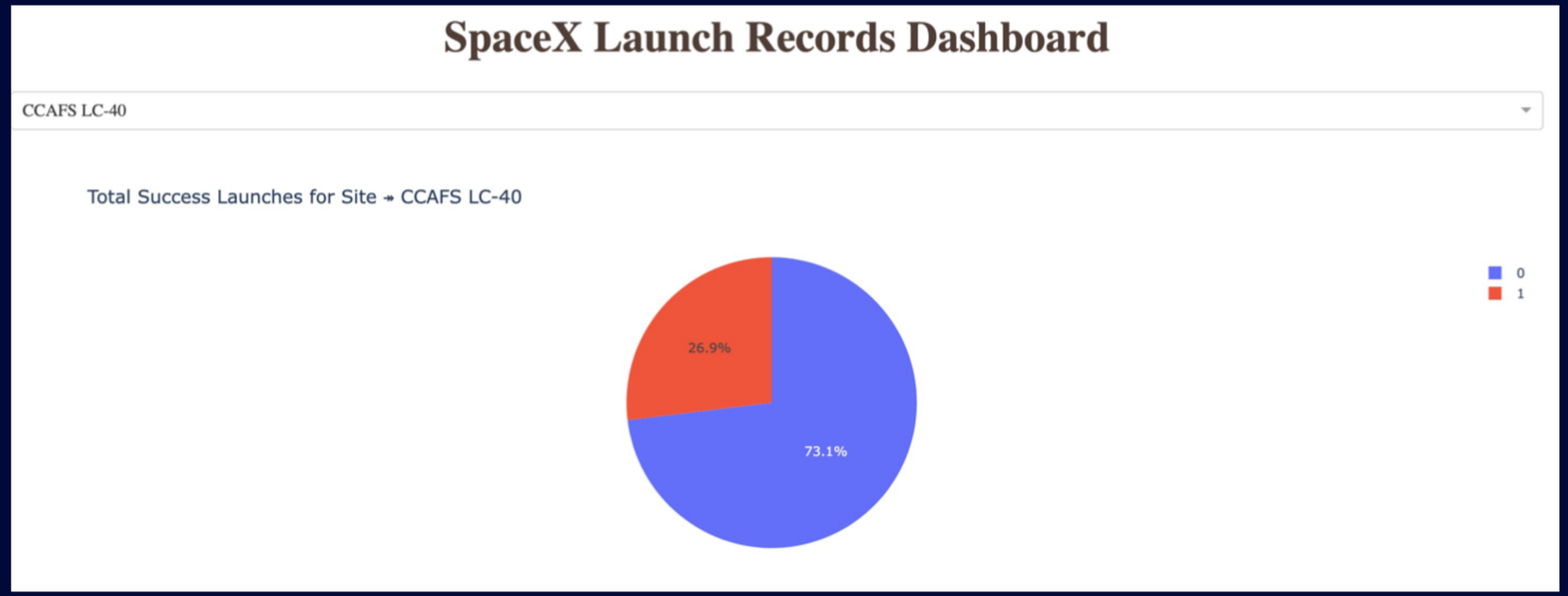
- Launch sites and succeed or not



Results

Plotly Dash dashboard results

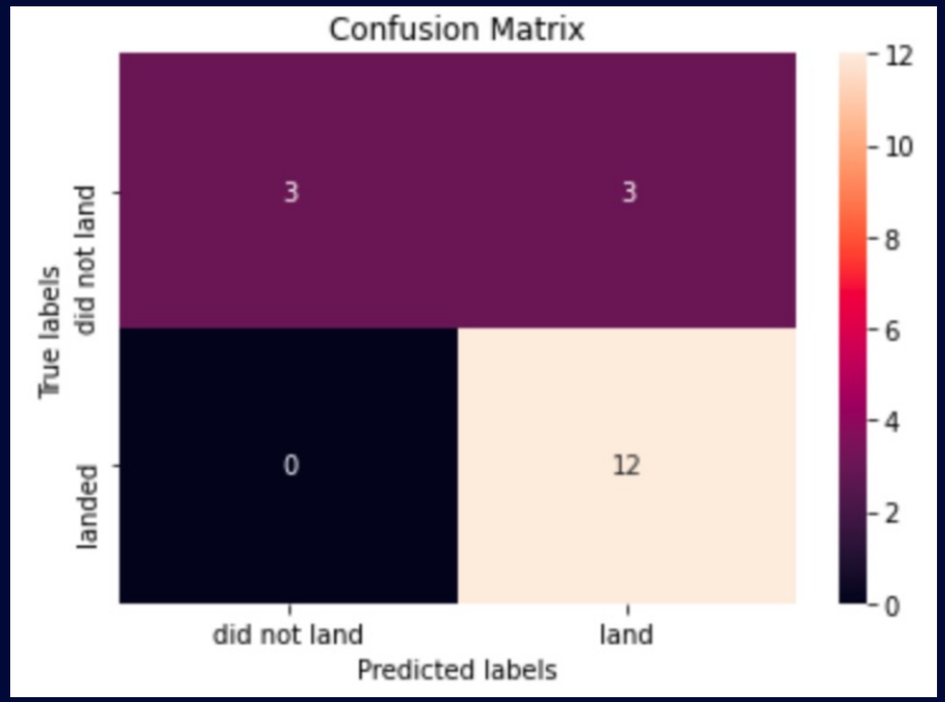
- Dashboard



Results

Predictive analysis (classification) results

- Logistic regression
  - GridSearchCV best score: 0.8464285714285713
  - Accuracy score on test set: 0.8333333333333334
- Support vector machine (SVM)
  - GridSearchCV best score: 0.8482142857142856
  - Accuracy score on test set: 0.8333333333333334
- Decision tree
  - GridSearchCV best score: 0.8892857142857142
  - Accuracy score on test set: 0.8333333333333334
- KNN
  - GridSearchCV best score: 0.8482142857142858
  - Accuracy score on test set: 0.8333333333333334



Conclusion

- In this project, we try to predict if the first stage of a given Falcon 9 launch will land in order to determine the cost of a launch.
- Each feature of a Falcon 9 launch, such as its payload mass or orbit type, may affect the mission outcome in a certain way.
- Several machine learning algorithms are employed to learn the patterns of past Falcon 9 launch data to produce predictive models that can be used to predict the outcome of a Falcon 9 launch.
- Several machine learning algorithms are employed to learn the patterns of past Falcon 9 launch data to produce predictive models that can be used to predict the outcome of a Falcon 9 launch.