IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ryan Smith
June 2024

Outline

Executive Summary

- Overview of the analysis on SpaceX launch data.
- Identification of key factors influencing the successful retrieval of the first stage.
    - Variables considered (e.g., launch site, weather conditions, rocket type).
    - Data sources and analysis techniques used.
    - Summary of findings.
- Implications for SpaceY's strategy to compete with SpaceX.
- Recommendations for future launches.
- Conclusion and next steps.

Introduction

- Reusable rocket technology.
- SpaceX’s achievements in first stage retrieval.
- Objectives of the analysis.
  - Location of launch.
  - Optimal payload weight.
- Importance for SpaceY.
  - Competitive landscape.
  - Strategic goals.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Webscraping
  - API
- Perform data wrangling
  - Cleaned data for Falcon 9
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Determining best ML model

Data Collection

- SpaceX API
  - Compile list of necessary items
  - Filtering on booster version
  - Handling missing values
- Webscraping SpaceX wiki for tables

Data Collection – SpaceX API

- Collect data through SpaceX data API
  - Rockets
  - LaunchPads
  - Payloads
  - Cores
- Convert .json to pandas dataframe to work with the data
- Extract necessary information in lists to create a new dataframe of the desired data
- A minimal amount of data is missing
  - The missing data, because it is reasonably small, is replaced with a mean of existing data

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Wikipedia charts contain useful information to use in our project
- Webscraping the information using get requests and beautiful soup

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/jupyter-labs-webscraping-bak-2024-06-26-03-19-23Z.ipynb

Data Wrangling

- Data is first analyzed
  - calculating number of launches on each site
  - calculating number and occurrences of each orbit
  - calculating the number and occurrence of mission outcome of the orbits
- Adding column for the landing outcome success or failure
- Populating the new column with 0 and 1 for failure and success respectively

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- EDA visualization lab
    - scatter plot payload mass flight number
    - scatterplot flight number launch site
    - scatter plot payload mass launch site
    - barchart orbit success rate
    - scatter plot orbit flight number
    - scatter plot payload mass orbit
    - line chart year success rate
    - change categorical data to numerical
    - cast dataframe to float64 to export to csv
- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/edadataviz.ipynb

EDA with SQL

- Gather launch site locations
- Display total payload mass carried by boosters launched by NASA
- Display average payload mass carried by booster version F9
- Discover first date of successful landing
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Created Folium map of Launch Site Locations
  - Clickable locations and Launch data
  - Distances to major infrastructure and ocean
    - Launch hazard potential

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Interactive dash application
  - Successful launches by Launch Site
  - Launch outcome by Payload weight

- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/dashapp.PNG

Predictive Analysis (Classification)

- Predictive Analysis of the data with Classification Machine Learning
- Tested between multiple classification algorithms
  - Logistic regression, Support Vector Machine, Decision Tree Classifier, K Nearest Neighbors
  - Scored each of them respectively after testing the data
  - Decision Tree Classifier offered the best accuracy among the different algorithms
- https://github.com/Hamstercrumbs/DataScienceCapstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

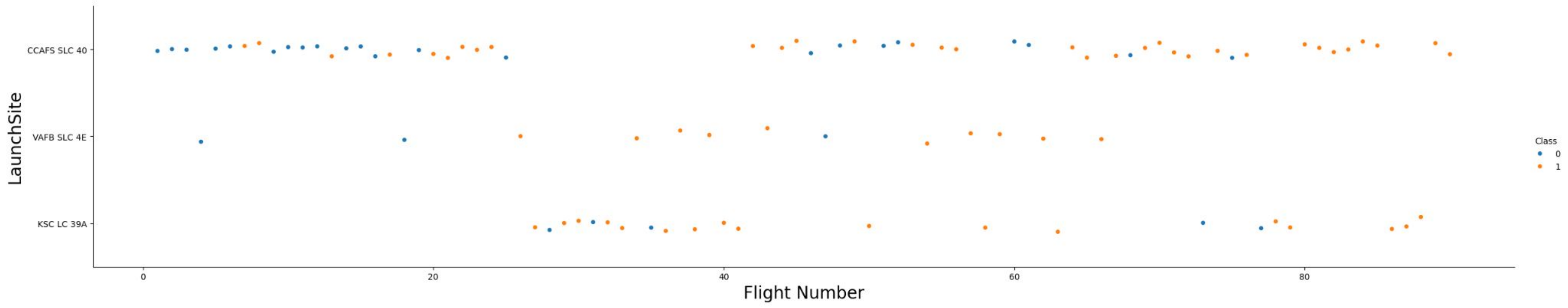
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

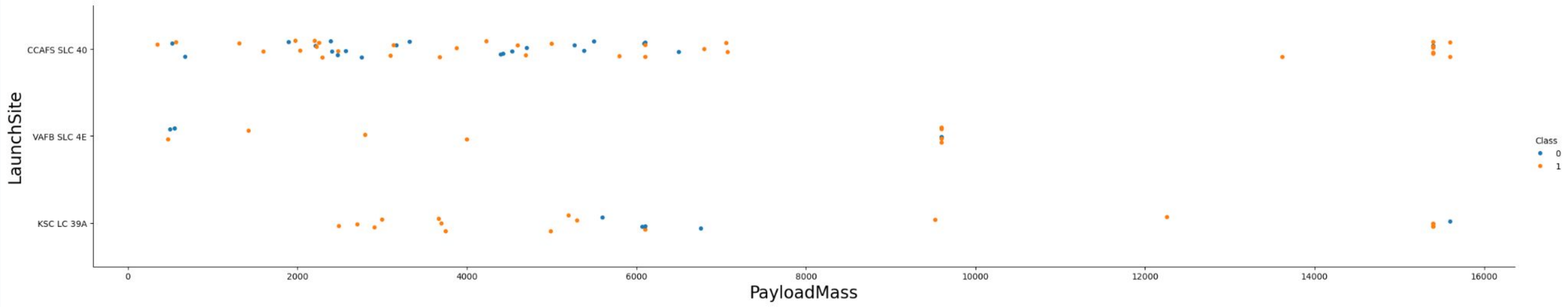
Flight Number vs. Launch Site

- FLight number vs Launch site
    - Class color for success and failure



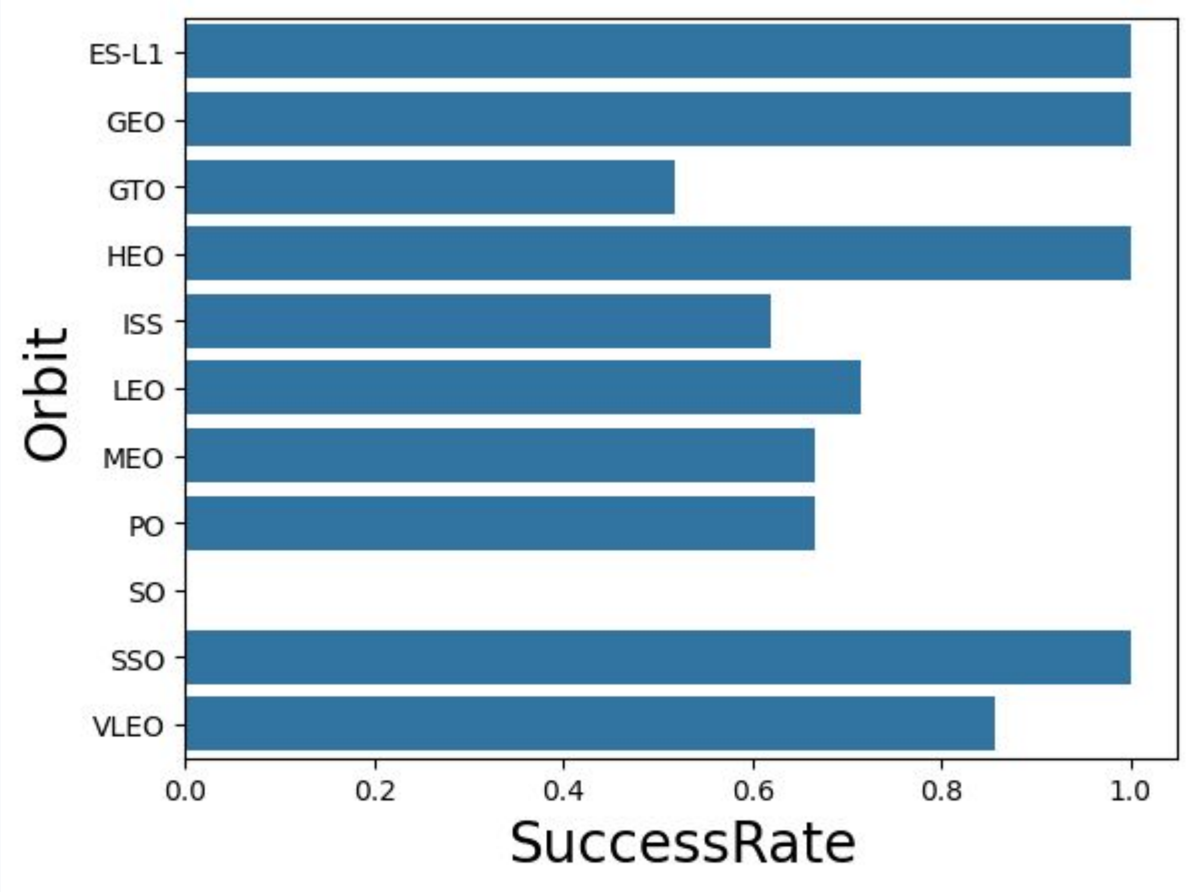
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
  - Class color for success and failure



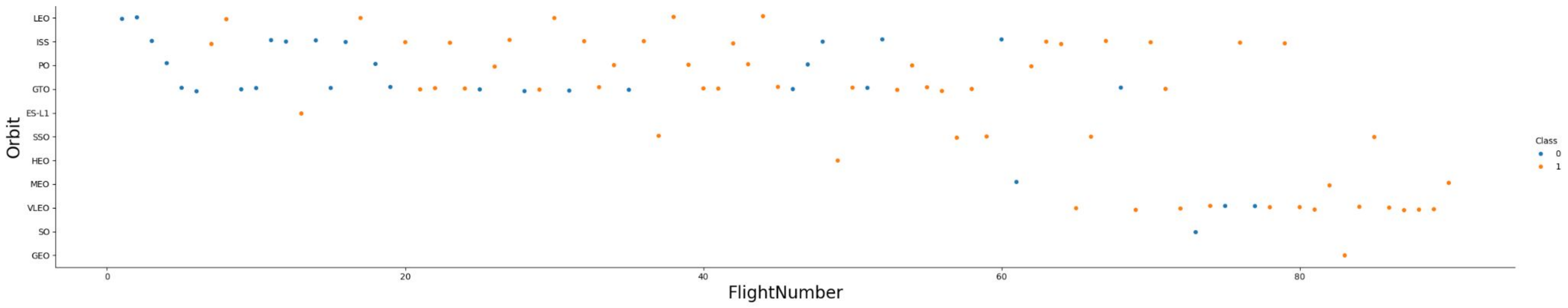
Success Rate vs. Orbit Type

- Bar chart of the success rate of each orbit type



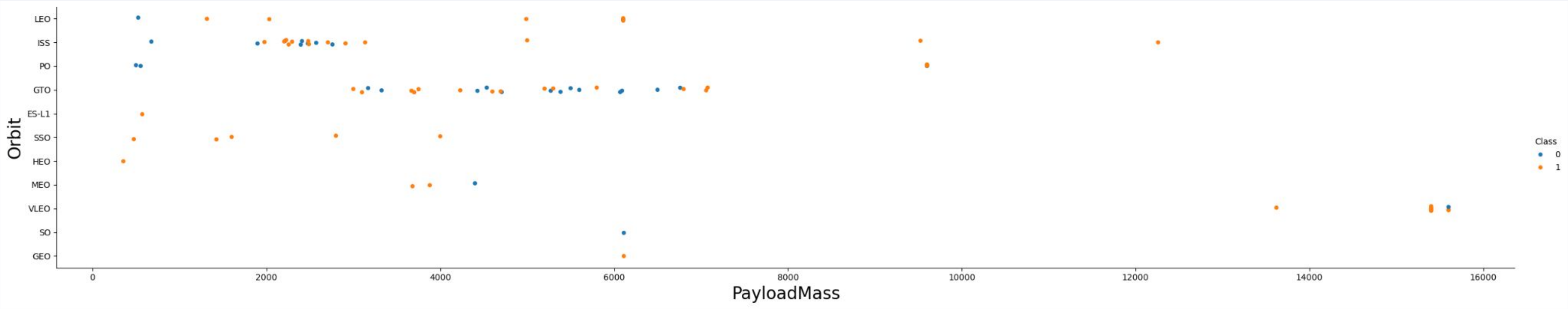
Flight Number vs. Orbit Type

- Scatter plot of Flight number vs. Orbit type
    - Class color for success and failure



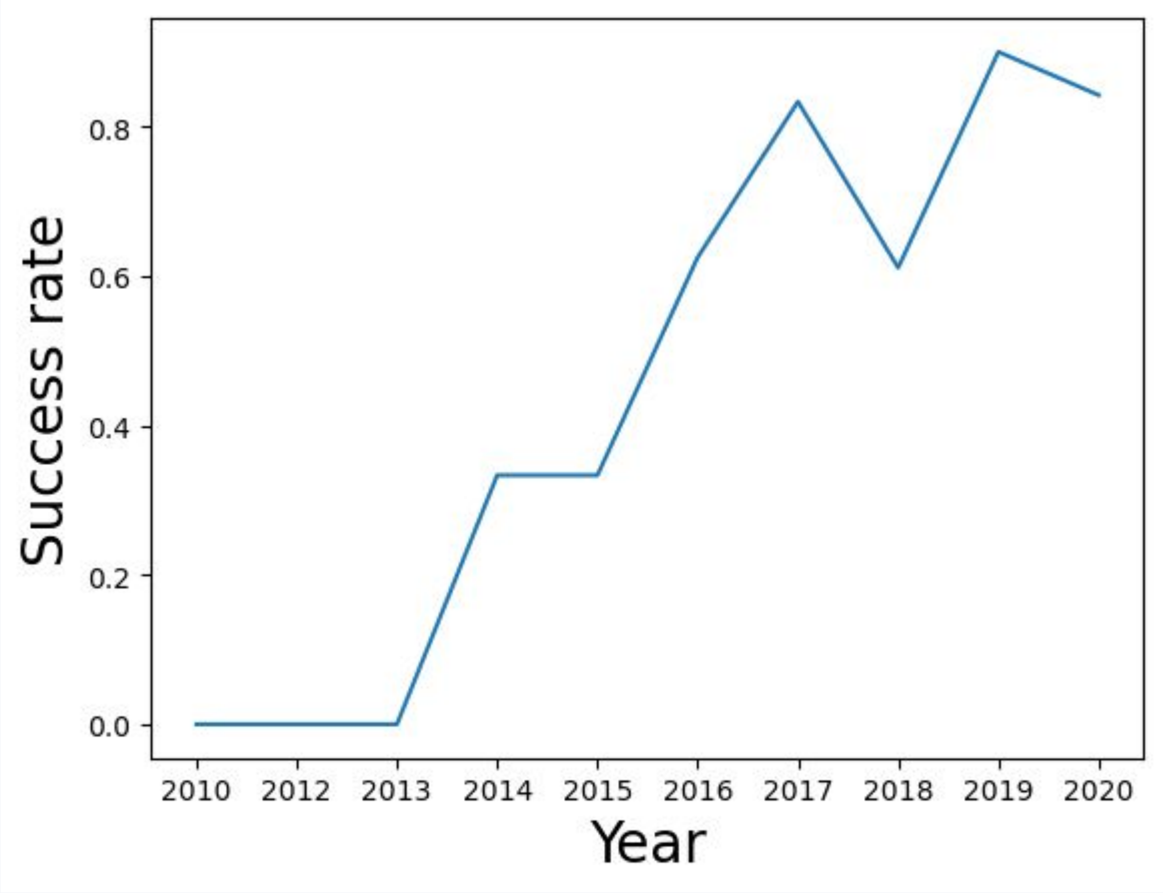
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type
    - Class color for success and failure



Launch Success Yearly Trend

- Line chart of yearly average success rate



All Launch Site Names

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

Launch Site Names Begin with '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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Calculate the total payload carried by boosters from NASA
- Total Payload Mass 48,213 kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Average Payload Mass 2,928.4 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- List 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

- Calculate the total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
  - 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

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

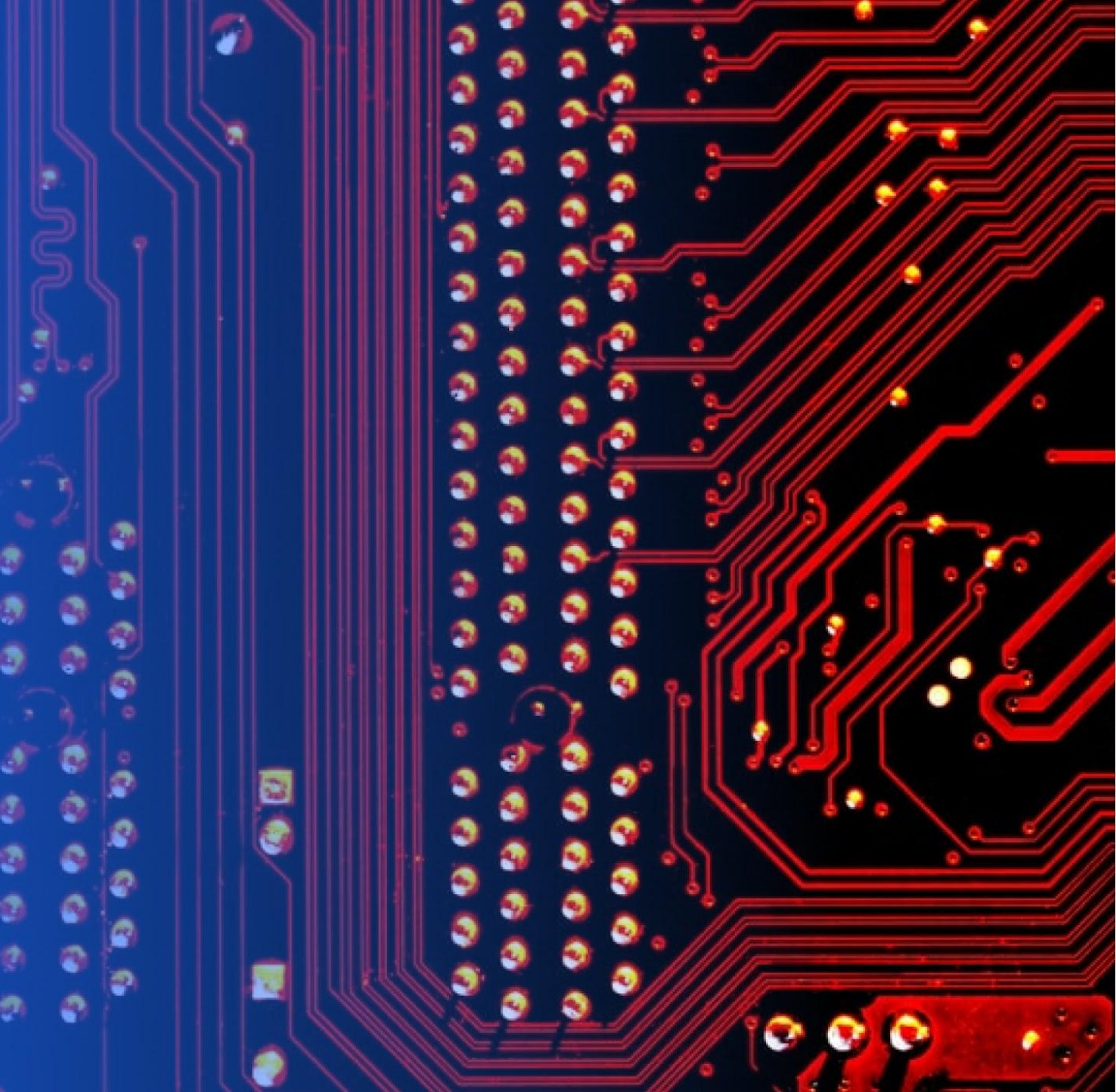
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

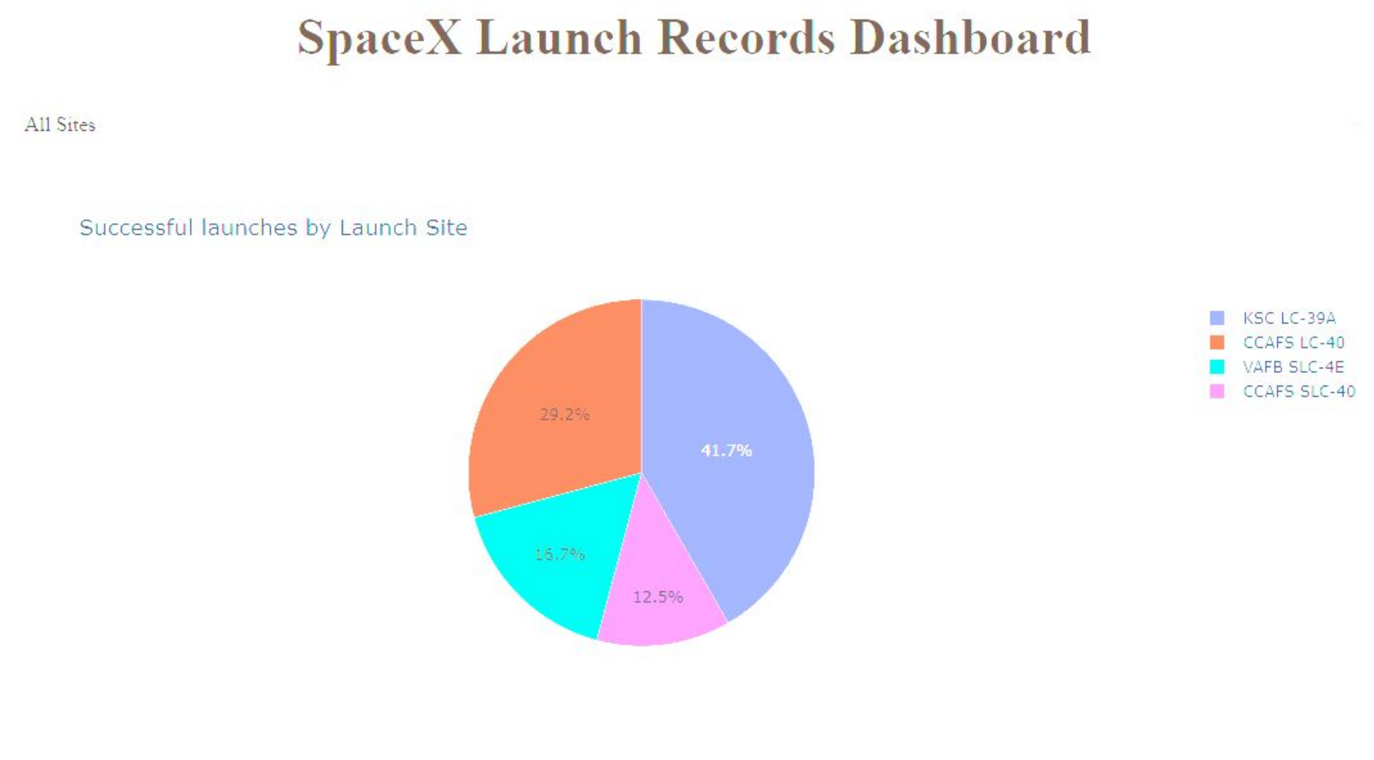
Section 4

Build a Dashboard with Plotly Dash



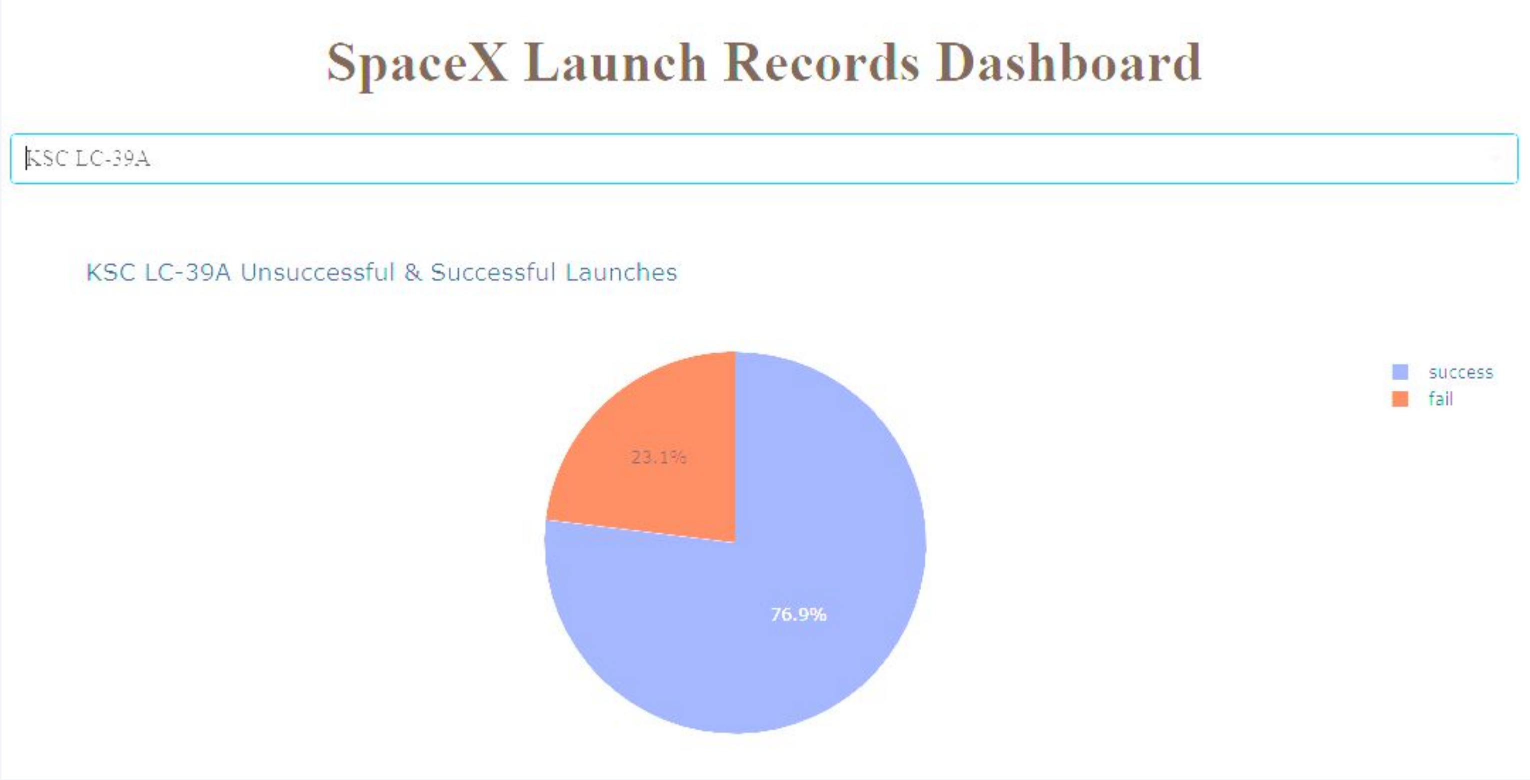
Plotly Dash interactive dashboard

- Piechart of all sites displayed



Plotly Dash interactive dashboard

- Most successful launch location



Plotly Dash interactive dashboard

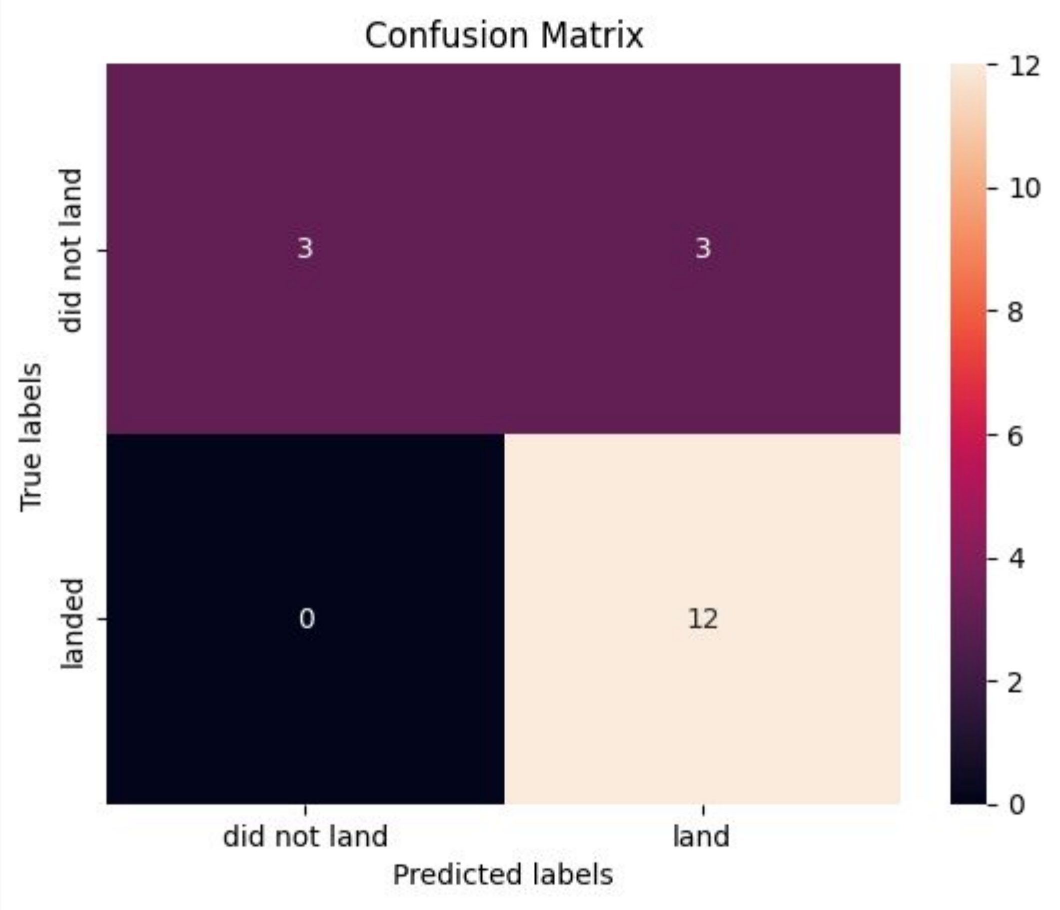
- All payload ranges of all launch locations

Section 5

Predictive Analysis (Classification)

Decision tree Classifier

- With a test accuracy of 0.83 and an accuracy score of 0.875 the decision tree classifier is the most robust model



Conclusions

- Through the work of SpaceX we can use their collected data to make better future decisions for a more profitable company to compete in the market of space travel
- With the data collected and processed we can make appropriate decisions for the best possibility of successful retrievals
- With a higher success in retrieving rockets for reuse we can save significant amounts of money

Appendix

- For additional information please view the github link below:
  - https://github.com/Hamstercrumbs/DataScienceCapstone

Thank you!