
# <Falcon 9 launch Analysis and Predictions>

<Kevin Barnett>

<10/04/2023>

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

- Summary of methodologies
  - Data collection - Data wrangling - Exploratory Data Analysis with Data Visualization - Exploratory Data Analysis with SQL - Building an interactive map with Folium - Building a Dashboard with Plotly Dash - Predictive analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis results - Interactive analytics demo in screenshots - Predictive analysis results

# INTRODUCTION

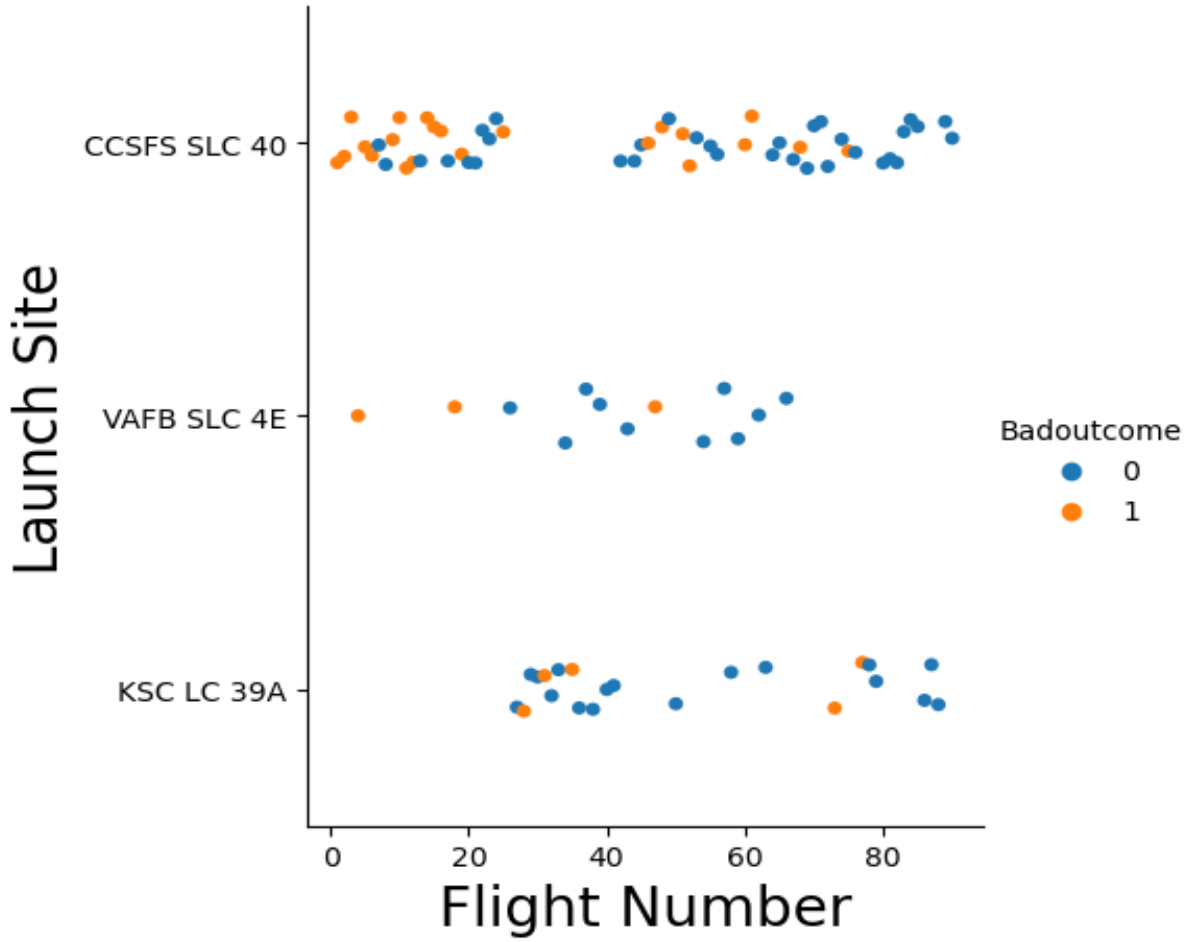
- Project background and context
  - SpaceX is the most successful company of the commercial space age, making space travel affordable. The company advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.
- Questions to be answered
  - How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing? - Does the rate of successful landings increase over the years? - What is the best algorithm that can be used for binary classification in this case?

# METHODOLOGY

- Data collection methodology:
  - Using SpaceX Rest API
  - Using Web Scrapping from Wikipedia
- Performed data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Building, tuning and evaluation of classification models to ensure the best results

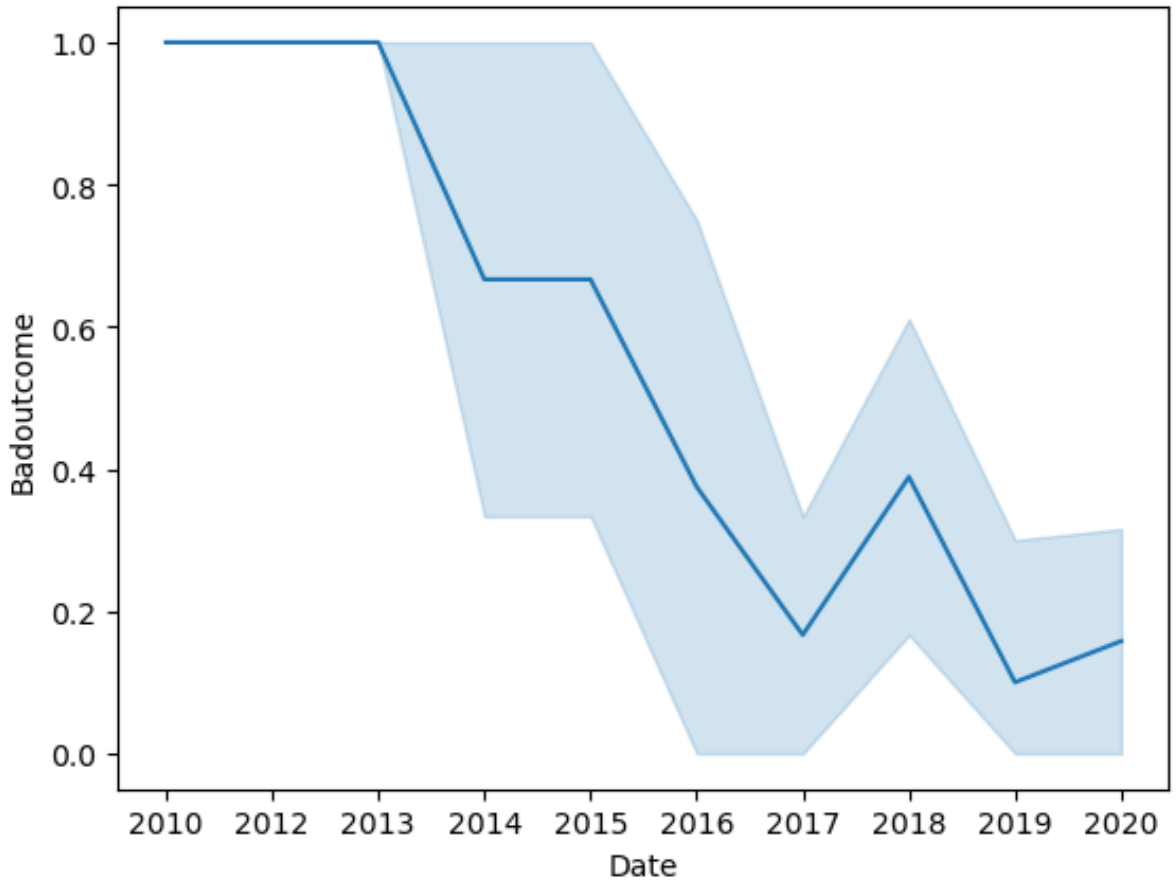
# Exploratory Data Analysis Insights

# Launch Site Vs Flight Number in Relation to Outcome



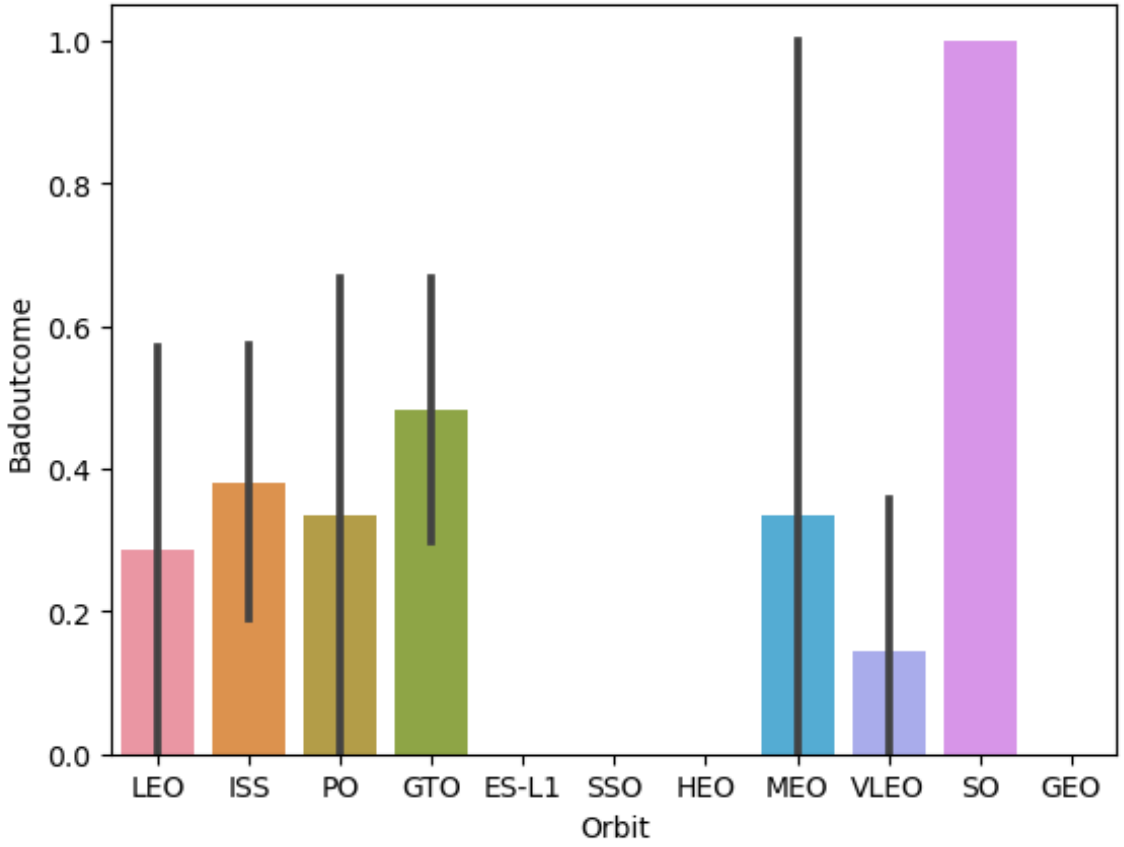

- For all three launch sights we can see the number of bad outcomes tends to go down as the number of flights increase
- The VAFB SLC 4E launch site seems to have the best outcomes but also has the least flight numbers overall

# The trends of Bad Outcomes from 2010-2020




- Overtime the number of bad outcomes has dropped significantly, meaning the flights of gotten more successful

# Bad Outcomes by Orbit Type

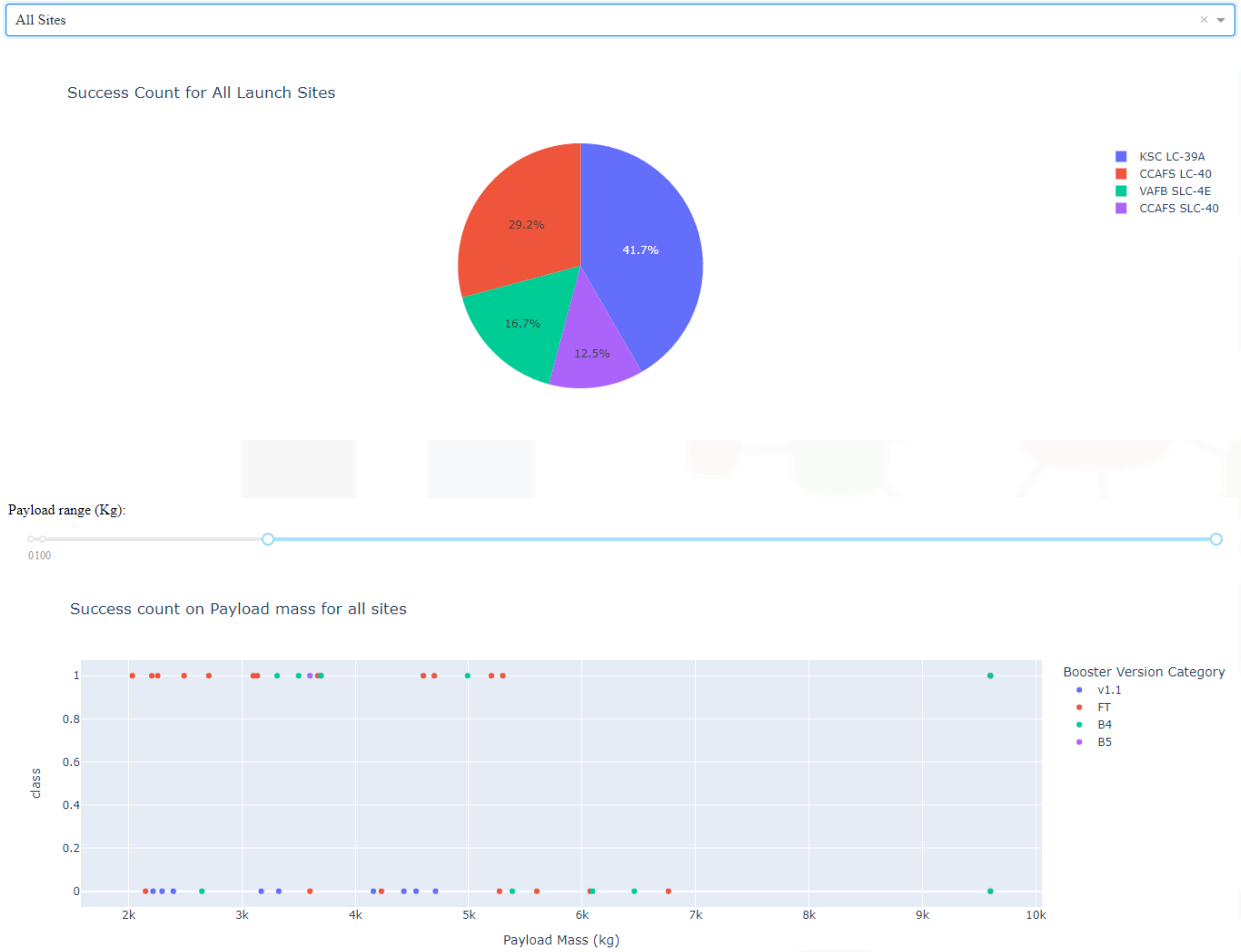



- There are four orbit types that have no bad outcomes (ES-L1, SSO, HEO, GEO)
- The SO orbit type has the most bad outcomes
- There also appears to be a significant number of outliers

# Exploratory Data Analysis with SQL

- The SQL queries performed on the data set were used to:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string ‘CCA’
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display the average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome on a ground pad was achieved
  - List the names of the boosters which had success on a drone ship and a payload mass between 4000 and 6000 kg
  - List the total number of successful and failed mission outcomes
  - List the names of the booster versions which have carried the maximum payload mass
  - List the failed landing outcomes on drone ships, their booster versions, and launch site names for 2015
  - 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

# Interactive Dashboard with Dash




- A drop down menu was created to select launch sites and show the percentage of successful vs not successful outcomes

- A slider to select the payload mass of a flight was created to show the outcomes of flights by booster version

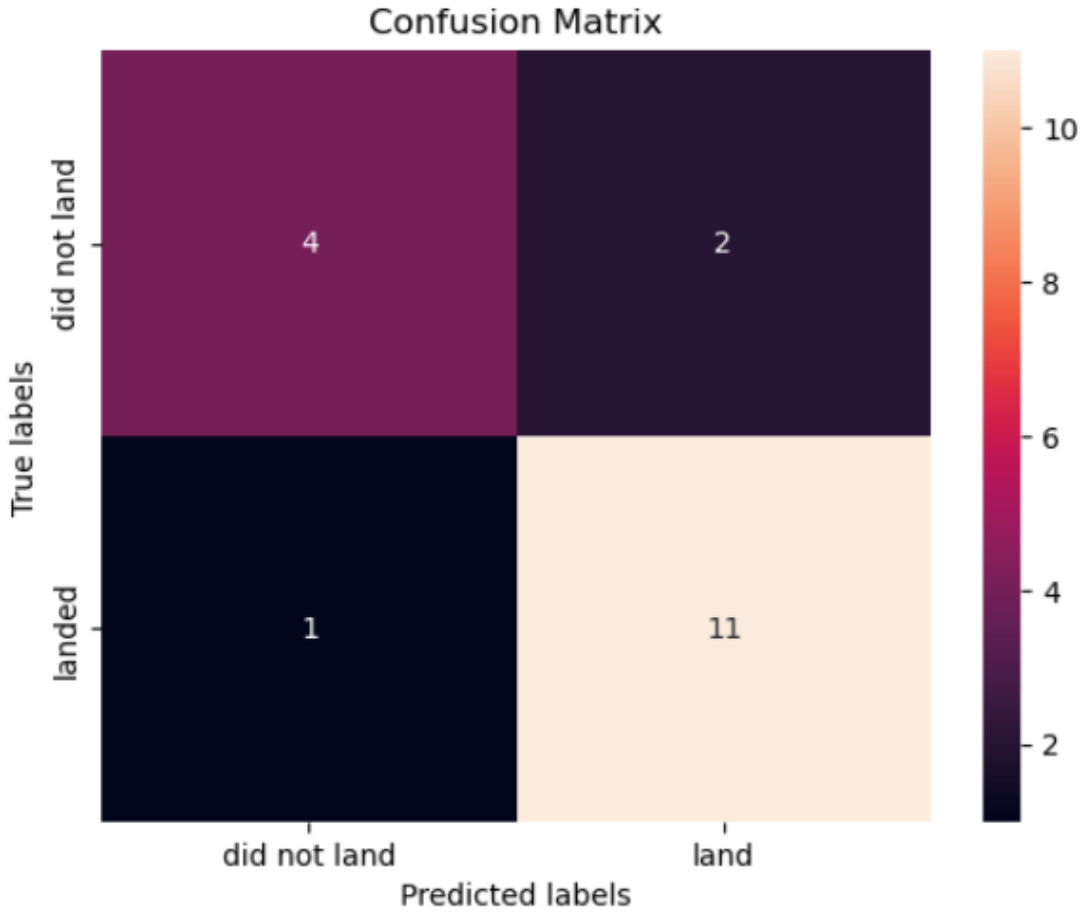
# Machine Learning Predictive Insights

# Classification

- Four classification algorithms were accessed (Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbors)
- Gridsearch CV was used to determine optimal parameters for each model
- Confusion Matrices were used to evaluate Type I & II Errors

# Decision Tree Classifier

- Tuned parameters:
  - Criterion: gini
  - Max depth: 12
  - Min samples leaf: 4
  - Min samples split: 10
  - Splitter: best
- Train Accuracy: 0.87
- Test Accuracy: 0.83
- Type I Error: 1
- Type II Error: 2



# DISCUSSION

# OVERALL FINDINGS & IMPLICATIONS

## Findings

- Finding 1
- Finding 2
- Finding 3

## Implications

- Implication 1
- Implication 2
- Implication 3

# CONCLUSION

- As the number of flights increases, the rate of success at a launch site increases, with most early flights being unsuccessful
- Orbit types ES-L1, GEO, HEO, and SSO, have the highest success rate
- The success for payloads over 4000 kg is lower than that with lower payloads
- The best performing classification model is the Decision Tree Classifier, with an accuracy of 83.3% and the least false positives