IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Julio Alburquerque
Quintanilla

03-06-2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQLBuilding an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

Background

This project is the capstone of our learning journey—the final course in the IBM Data Science specialization. Imagine us as the analytical minds behind SpaceY, a new aerospace company with aspirations to compete with giants like SpaceX.

In this project, we'll apply cutting-edge data science techniques to help SpaceY make informed decisions about rocket launches—and maybe even outperform the industry leaders.

Business Problem

Here's the challenge: SpaceX revolutionized the space industry by dramatically lowering launch costs. A typical Falcon 9 launch costs around $62 million—significantly less than the competition—largely because they reuse the rocket's first stage. That stage alone can cost over $15 million to build, not including development expenses.

However, reusability isn't always possible. If the payload is particularly heavy or needs to reach a unique orbit, the rocket's first stage may not be recovered—meaning the cost-saving advantage is lost.

Our mission is to predict whether a rocket's first stage will be successfully recovered. Accurate predictions could translate into substantial savings—a figure any stakeholder would be eager to see

Section 1

Methodology

Methodology

Executive Summary

- **Data Collection Methodology**
- **Data Sources**:
  - SpaceX Open Source REST API
  - Web scraping from the Wikipedia page *"List of Falcon 9 and Falcon Heavy Launches"*
- **Data Preparation**:
  - Conducted data wrangling to clean and structure the dataset.
  - Applied One-Hot Encoding to transform categorical variables for compatibility with machine learning algorithms.
  - Removed missing values and irrelevant information to ensure data quality.
- **Exploratory Data Analysis (EDA)**:
  - Performed EDA using visualizations and SQL queries to uncover insights and patterns in the data.
- **Interactive Visual Analytics**:
  - Developed interactive dashboards using **Folium** (for geographic visualizations) and **Plotly Dash** (for dynamic data exploration).
- **Predictive Modeling**:
  - Built and evaluated several classification models, including:
    - Logistic Regression
    - K-Nearest Neighbors (KNN)
    - Support Vector Machines (SVM)
    - Decision Trees
  - The goal was to identify the most effective model for predicting the success of rocket stage recovery.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

- Show the screenshot of the scatter plot with explanations

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

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

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

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

- Present your query result with a short explanation here

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

- Present your query result with a short explanation here

Section 3

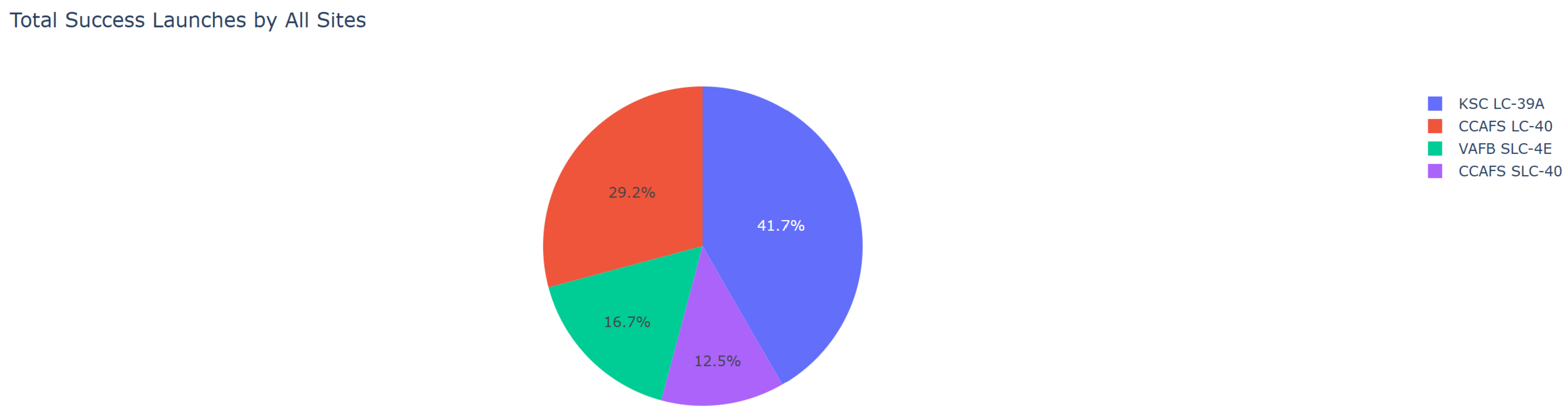
Launch Sites Proximities Analysis

All launch sites

Section 4

Build a Dashboard with Plotly Dash

Sites launches success



Correlation payload vs success

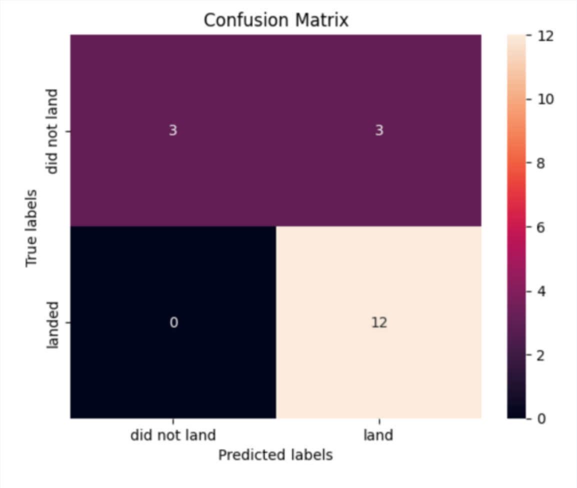
Section 5

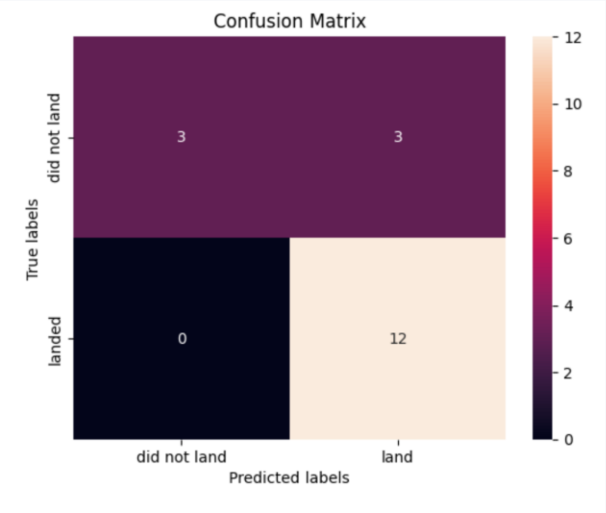
Predictive Analysis (Classification)

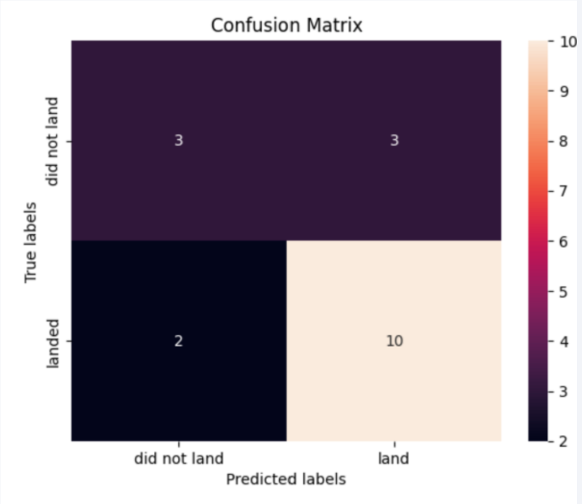
Classification Accuracy

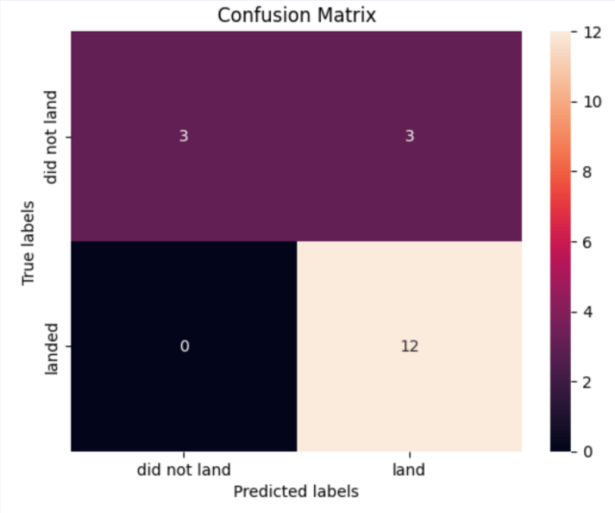
	Best scores
Logistic regresssion	0.846429
SVM	0.848214
Decision tree	0.862500
KNN	0.848214

Confusion Matrix









Conclusions

- **Top Performing Models**: Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) demonstrated the highest accuracy in predicting rocket recovery outcomes.
- **Payload Impact**: Launches with lighter payloads tend to have a higher success rate compared to those carrying heavier payloads.
- **Experience Matters**: The probability of a successful SpaceX launch increases with the company's operational experience, indicating a trend toward greater reliability over time.
- **Launch Site Performance**: Launch Complex 39A at Kennedy Space Center recorded the highest number of successful missions among all launch sites analyzed.
- **Orbit Type Success Rates**: Missions targeting GEO (Geostationary Earth Orbit), HEO (Highly Elliptical Orbit), SSO (Sun-Synchronous Orbit), and ESL1 (Earth–Sun Lagrange Point 1) demonstrated the highest rates of successful launches.

Thank you!