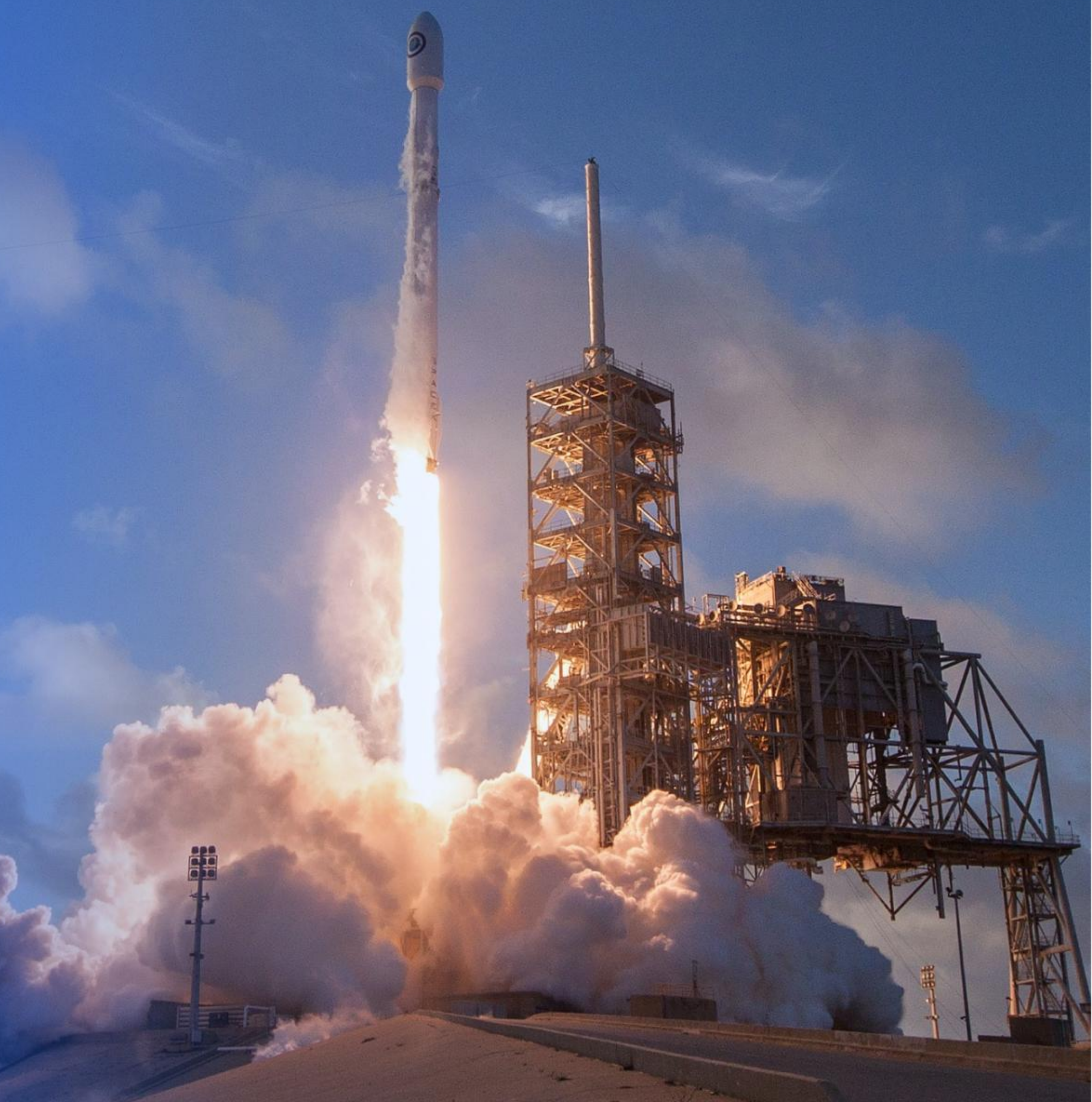
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

IBM Data Science Capstone Project
SpaceX Falcon 9 Launch Analysis

Author: Chathuranga Sudusinghe



Outline

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

Executive Summary

This project analyzes historical SpaceX Falcon 9 launch data to understand the key factors affecting first-stage landing success.

Using data collected from SpaceX APIs, web scraping, SQL analysis, data visualization, interactive maps, dashboards, and machine learning classification models, we identify patterns that influence launch outcomes.

The final objective is to build and evaluate predictive models that can estimate the probability of successful Falcon 9 first-stage landings.

Introduction

SpaceX aims to reduce launch costs by reusing the first stage of Falcon 9 rockets.

Accurately predicting whether a launch will successfully land helps:

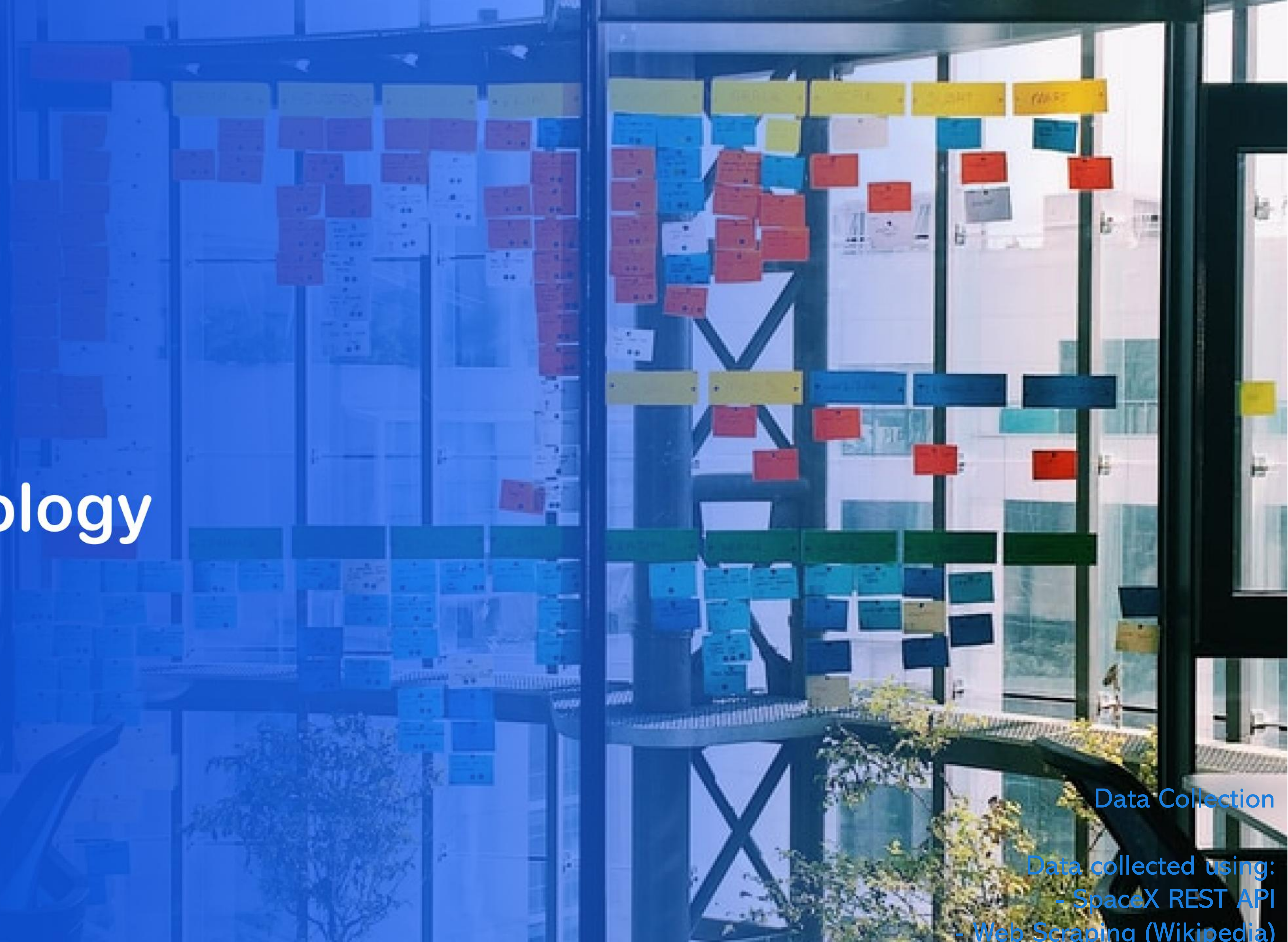
- Optimize mission planning
- Reduce operational risks
- Improve cost efficiency

Key Questions:

- Which launch sites have the highest success rates?
- How do payload mass and orbit type affect launch success?
- Which machine learning model performs best for prediction?

Section 1

Methodology



Data Collection

Data collected using:
- SpaceX REST API
- Web Scraping (Wikipedia)

Methodology

The project follows a structured data science workflow:

1. Data Collection (API & Web Scraping)
2. Data Wrangling
3. Exploratory Data Analysis (EDA)
4. SQL Analysis
5. Interactive Visual Analytics (Folium & Plotly Dash)
6. Predictive Analysis using Classification Models
7. Model Evaluation and Selection

Project Resources & Repository

This project is fully documented and reproducible.
All source code, notebooks, and deliverables are available in the GitHub repository below.

GitHub Repository:
https://github.com/chathuranga-sudusinghe/IBM-Data-Science-SpaceX-Falcon-9-Launch-Analysis-and-Prediction/tree/main

Contents include:

- Data collection using SpaceX REST API
- Web scraping notebooks
- Data wrangling and preprocessing
- EDA with visualization and SQL
- Interactive Folium maps
- Plotly Dash dashboard
- Machine learning classification models
- Final project presentation (PPTX / PDF)

Data Collection

Data was collected from two main sources:

- **SpaceX REST API**
  Used to retrieve launch details, rocket information, payload mass, orbit type, and landing outcomes.

- **Web Scraping (Wikipedia)**
  Used to extract historical Falcon 9 launch records.

These datasets were merged to create a comprehensive launch dataset for analysis.

Data Wrangling

The collected data was processed using the following steps:

- Handling missing and null values
- Converting categorical variables into numerical format
- Feature selection for modeling
- Standardizing numerical features
- Splitting data into training and testing sets

These steps ensured data quality and model readiness.

EDA with Data Visualization

EDA was performed using visual tools to understand relationships between variables.

Key visualizations included:

- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- Success Rate vs Orbit Type
- Yearly Launch Success Trend

These plots helped identify patterns and trends in launch performance.

EDA with SQL

SQL queries were used to analyze the dataset efficiently.

Key SQL analyses included:

- Identifying unique launch sites
- Calculating total and average payload masses
- Counting successful and failed missions
- Ranking landing outcomes by frequency
- Filtering launch records by year and payload range

SQL provided precise, query-based insights into launch data.

Build an Interactive Map with Folium

An interactive map was built using Folium to visualize launch sites geographically.

Features included:

- Location markers for all launch sites
- Color-coded launch outcomes (success/failure)
- Distance calculations to nearby infrastructure (railways, highways, coastlines)

This helped assess geographic and logistical influences on launch success.

Build a Dashboard with Plotly Dash

An interactive dashboard was created using Plotly Dash.

Dashboard components included:

- Pie chart of launch success counts by site
- Pie chart for the site with the highest success ratio
- Scatter plot of Payload Mass vs Launch Outcome with interactive filters

The dashboard allows dynamic exploration of launch performance.

Predictive Analysis (Classification)

Several machine learning classification models were built and evaluated:

- Logistic Regression
- Decision Tree Classifier
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Hyperparameter tuning and cross-validation were applied to improve model performance.

Results

Model evaluation showed that:

- The **Decision Tree model** achieved the highest classification accuracy.
- Payload mass and orbit type were the most influential features.
- Certain launch sites consistently showed higher success rates.

The models demonstrated strong predictive capability.

Section 2

Insights drawn from EDA

Insights Drawn from Exploratory Data Analysis (EDA)

- Launch success rate improves significantly as **flight number increases**, indicating strong learning and operational maturity over time.
- **Certain launch sites** (e.g., CCAFS LC-40 and KSC LC-39A) consistently show **higher success rates** compared to others.
- **Payload mass** has a noticeable impact on launch success:
  - Very low and very high payload ranges show lower success probability.
  - **Mid-range payloads** tend to have higher success rates.
- **Orbit type** strongly influences mission outcome:
  - Orbits such as **LEO and ISS** demonstrate higher success rates.
  - More complex orbits show relatively lower success probabilities.
- The **yearly trend analysis** reveals a clear improvement in launch success over time, especially after 2016, reflecting technological advancements and experience gains.

Appendix

The appendix includes:

- Python code snippets
- SQL queries
- Additional charts and plots
- Notebook outputs
- Data sources and references

These materials support transparency and reproducibility of the analysis.

Thank you!