# 🛰️ Space X Launch Success Prediction Capstone

This project serves as a comprehensive demonstration of the full Data Science lifecycle, from data acquisition and preparation through to advanced machine learning modeling, focusing on the prediction of Space X Falcon 9 launch success outcomes.

The goal was to build a robust classification model to predict the final outcome (success or failure) of a launch based on pre-flight parameters and historical data.

## 🎯 Project Goal

To deploy a set of classification models capable of predicting the binary outcome of Space X launches with high accuracy, minimizing risk and optimizing future mission planning.

## 🛠️ Methodology & Data Science Workflow

This capstone was executed using an end-to-end pipeline, showcasing proficiency in all stages of a complex data science challenge:

### 1. Data Acquisition & Sourcing

* **API Collection:** Acquired real-time and historical launch data directly from the official Space X REST API (using Python's requests library).
* **Web Scraping:** Utilized advanced web scraping techniques to supplement API data with specific launch details and information not available in the structured API output.

### 2. Data Wrangling & Cleaning (SQL & Pandas)

* Performed extensive data cleaning, handling missing values, standardizing categorical data, and transforming raw JSON structures into a relational format.
* **SQL EDA:** Employed complex **SQL queries** to structure the raw data, perform initial Exploratory Data Analysis (EDA), and identify correlations and feature relationships.

### 3. Exploratory Data Analysis (EDA) & Visualization

* Conducted thorough univariate and bivariate analysis to identify key predictors of launch success.
* Utilized **Matplotlib** and **Seaborn** for static visualizations and **Folium** for interactive map visualizations of launch sites, demonstrating ability to communicate findings visually.

### 4. Machine Learning Modeling (Prediction)

* Trained and evaluated four distinct classification algorithms:
  + **Logistic Regression**
  + **Decision Tree**
  + **Support Vector Machines (SVM)**
  + **K-Nearest Neighbors (KNN)**
* Used techniques like cross-validation and hyperparameter tuning to optimize model performance.

## ✨ Key Technical Skills Demonstrated

| **Skill Category** | **Specific Techniques Used** |
| --- | --- |
| **Machine Learning** | Classification, Model Evaluation (Confusion Matrix, F1-Score), Hyperparameter Tuning. |
| **Programming** | Python, Pandas, NumPy, Scikit-learn (ML), SQLAlchemy (SQL connection). |
| **Data Manipulation** | Complex SQL Joins, Aggregations, Data Merging, Feature Engineering. |
| **Data Visualization** | Matplotlib, Seaborn, and **Folium** (Interactive Geospatial Mapping). |

## 📦 Repository Contents

* Lab 1 Space X Data Collection API V2.ipynb
* Lab 2 Space X Data Collection with Web Scraping.ipynb
* Lab 3 Space X Data Wrangling.ipynb
* Lab 4 Space X EDA with SQL.ipynb
* Lab 5 Space X EDA with Data Visualization.ipynb
* Final Lab Space X Machine Learning Prediction.ipynb
* Applied Data Science Capstone - Space X Final Report.pdf (Final documentation and findings report)

## 👨‍💻 Created By

**Fabrice Achu Ngando**

* **Email:** [achubrayan62@gmail.com](mailto:achubrayan62@gmail.com)
* **LinkedIn:** <https://www.linkedin.com/in/fabrice-achu-ngando/?trk=opento_sprofile_topcard>