Winning Space Race with Data Science

Gregory Derousselle
10/10/2024

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

Data Collection & Preparation:	2. Data Preprocessing:	3. Machine Learning Models:	4. Evaluation & Analysis:	5. Model Performance Visualization:
• Utilized public SpaceX API and Wikipedia page • Created 'class' column for successful landing classification • Explored data using SQL, visualization, Folium maps, and dashboards • Selected relevant features for machine learning.	• Applied one hot encoding to categorical variables • Standardized data for uniform scale • Optimized model parameters using GridSearchCV	• Logistic Regression • Support Vector Machine • Decision Tree Classifier • K Nearest Neighbors	• Models tended to over predict successful landings • Identified need for more data to enhance accuracy	• Visualized accuracy scores to compare model performance

Introduction

PROJECT BACKGROUND AND CONTEXT

PROBLEMS YOU WANT TO FIND ANSWERS

Section 1

Methodology

Methodology

Data collection:

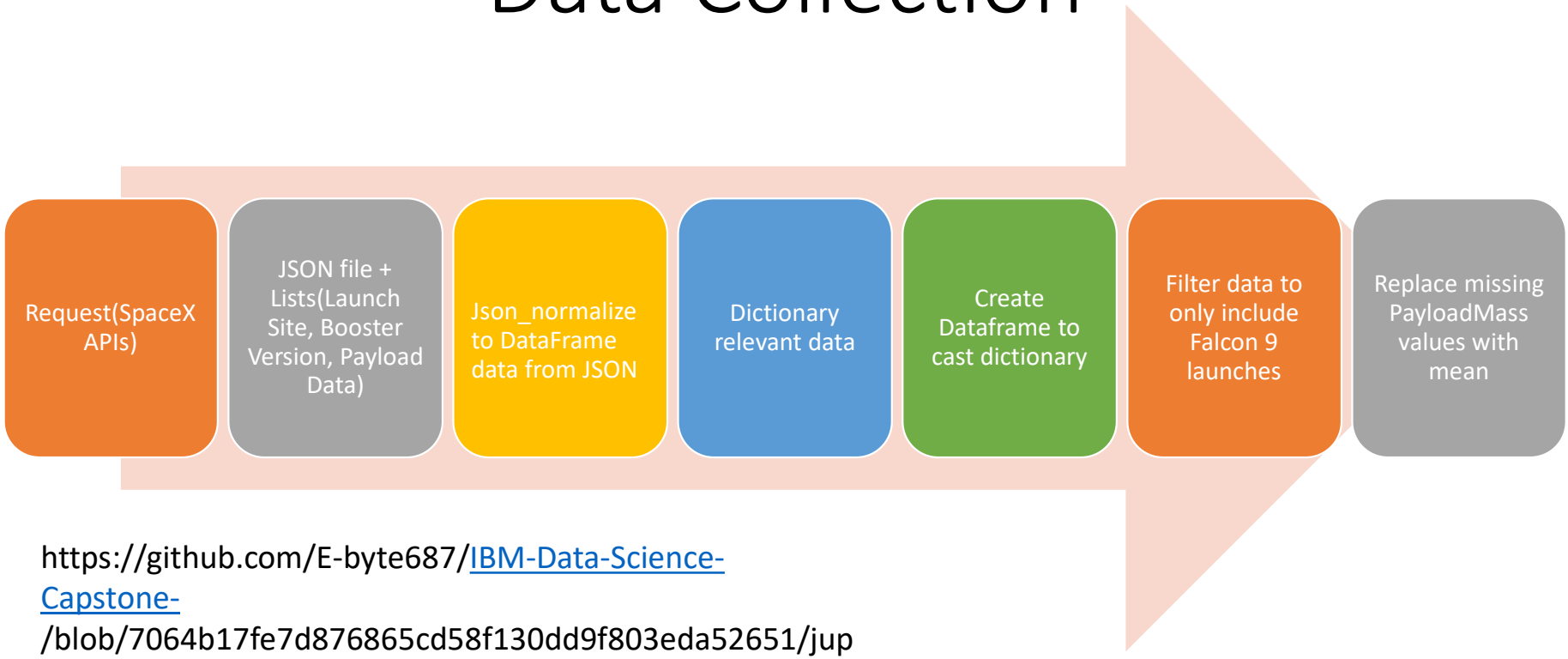
Perform data wrangling

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

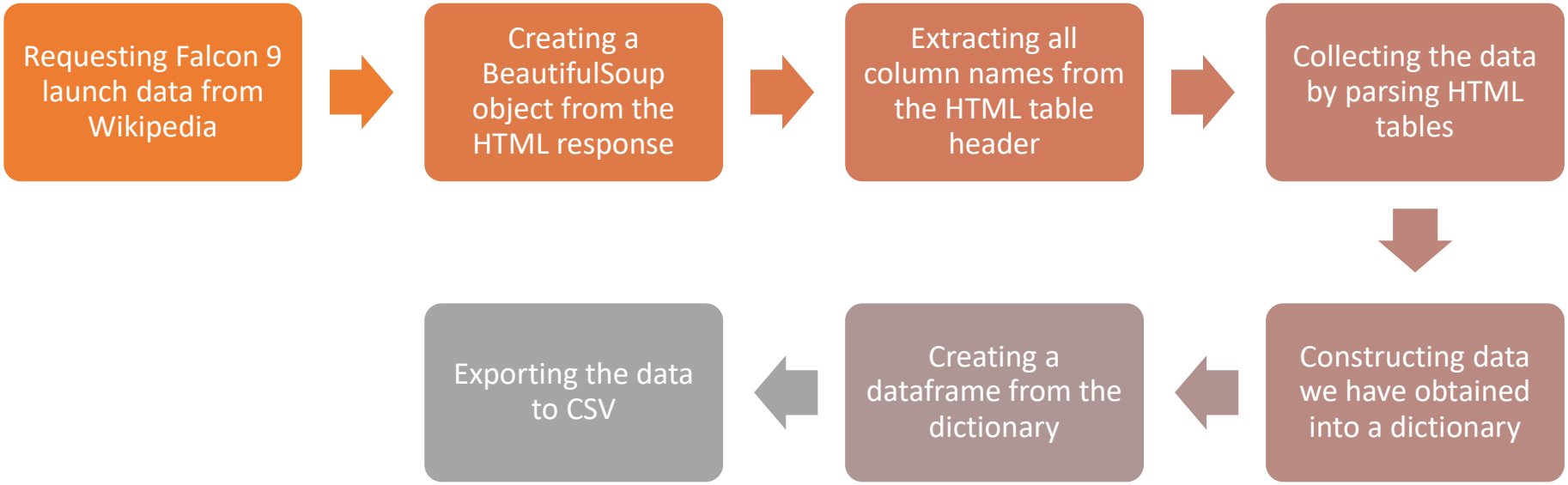
Perform predictive analysis using classification models

Data Collection

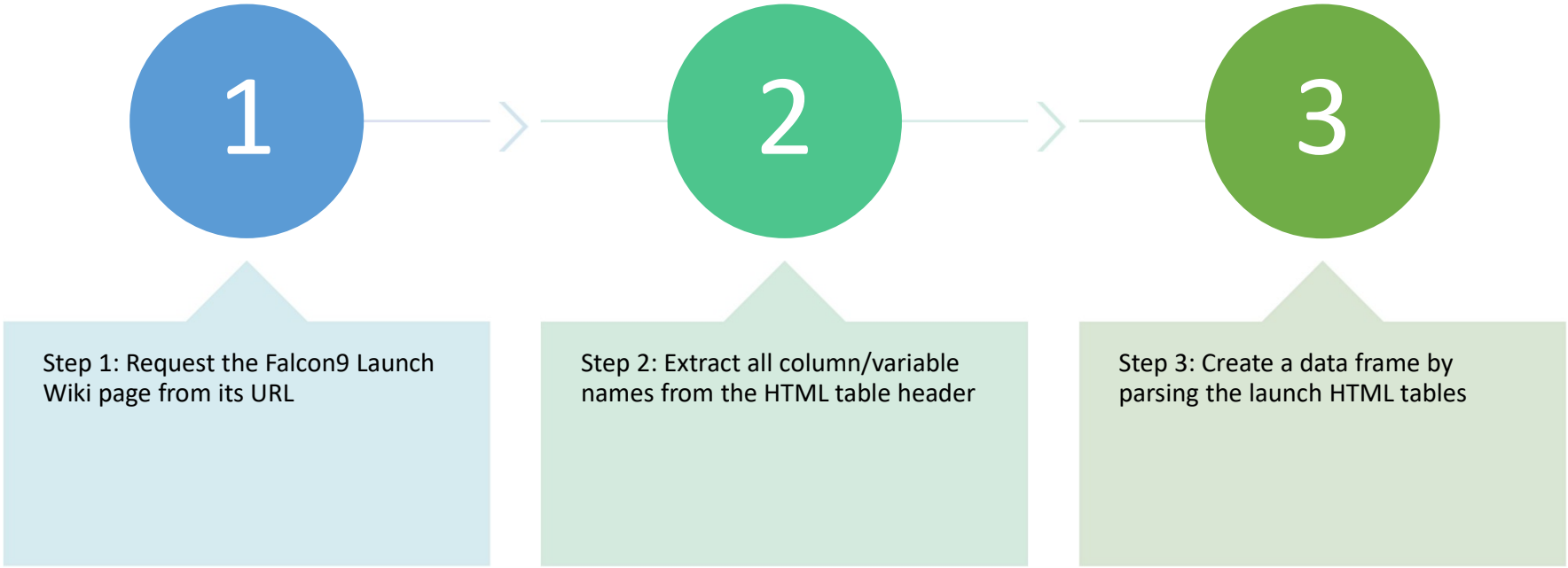


https://github.com/E-byte687/IBM-Data-Science-Capstone-/blob/7064b17fe7d876865cd58f130dd9f803eda52651/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection – SpaceX API



Data Collection - Scraping



Data Wrangling

- Describe how data were processed
- Step 1: Determined the number of launches on each site
- Step 2: Calculate the number and occurrence of each orbit
- Step 3: Calculate the number and occurrence of mission outcome of the orbits
- Step 4: Create a landing outcome label from outcome column
- Data Wrangling Github link

EDA with Data Visualization

Scatter Chart

To show the correlation between

- Payload Mass and Orbit Type
- Flight Number and Orbit Type
- Payload Mass and Launch Site
- Flight Number and Launch Site

Bar Chart

The bar chart represents the relationship between success rate of each orbit type

The goal is to see which orbits has the highest success rates

Line Graph

It is used to see the yearly trend of successful launches.

Data Visualization

EDA with SQL

- Utilized SQL queries to perform comprehensive exploratory data analysis (EDA), extracting valuable insights directly from the dataset
- SQL facilitated efficient querying, aggregation, and manipulation of data, enabling in-depth analysis of various aspects such as distribution, relationships, trends, and outliers
- The EDA with SQL provided a solid foundation for understanding the dataset's characteristics and informing subsequent analytical decisions
- EDA with SQL Github

Build an Interactive Map with Folium

- Utilized Folium, a Python library for creating interactive maps, to perform geospatial analysis and visualization of data. Popup information windows were added to display additional details so that when users interacted with map markers, enhancing the data exploration experience. Interactive features such as zooming, panning, and toggling layers were added to provide users with a better experience.
- Launch Sites Locations with Folium

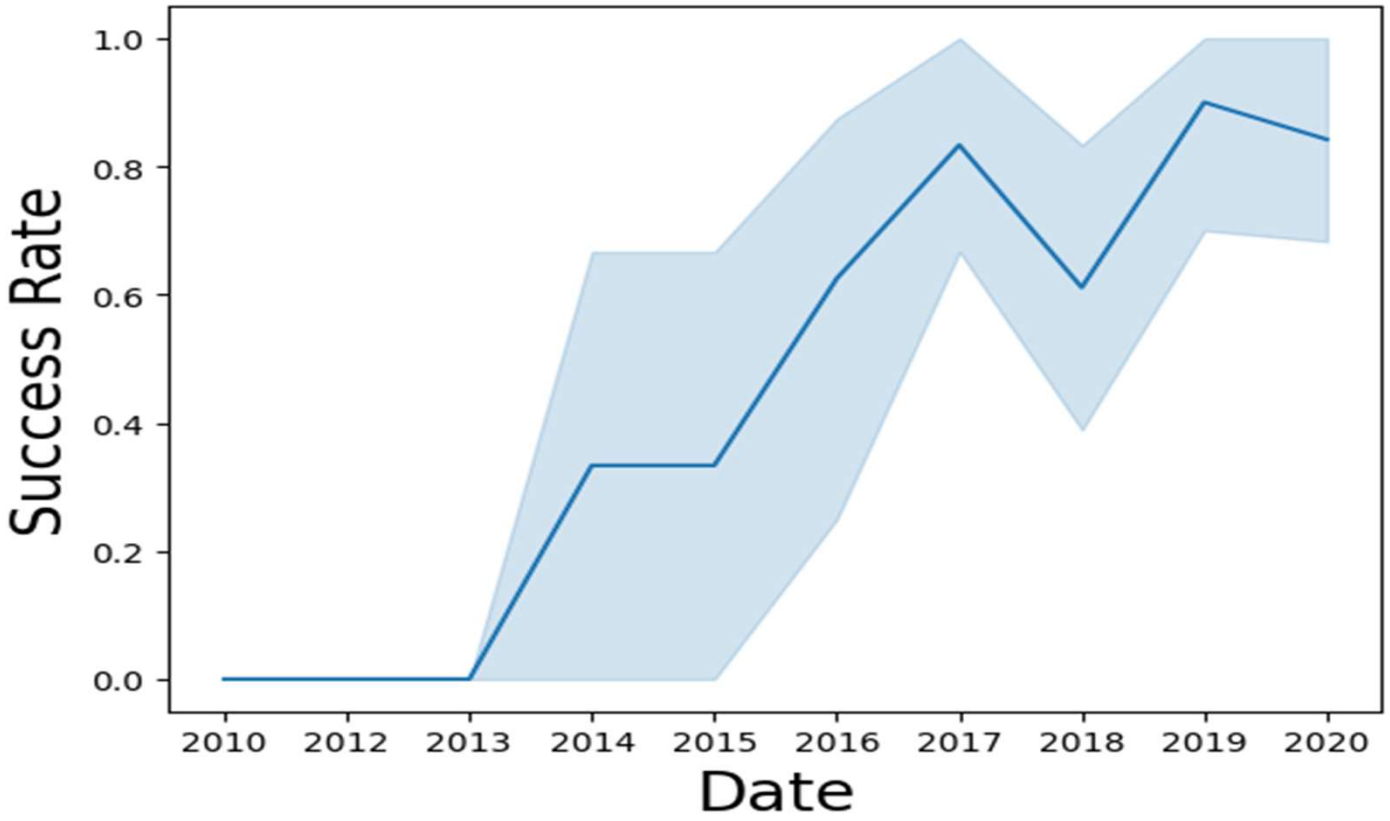
Build a Dashboard with Plotly Dash

- The interactive dashboard built with Plotly Dash offers a dynamic and user-friendly interface for exploring and visualizing data
- Data Visualization:
  - Implemented interactive charts and graphs using Plotly to visualize key insights and trends.
  - Included line charts, bar charts, scatter plots, and heat maps to represent different aspects of the data
  - Integrated dropdown menus, sliders, and date pickers to enable users to filter and customize the displayed data dynamically

Predictive Analysis (Classification)

- The Machine Learning Prediction Lab is dedicated to developing and evaluating predictive models using advanced machine learning techniques
- Model Evaluation:
  - Employed cross – validation techniques to assess model generalization and robustness
  - Identified key factors influencing the target variable based on feature importance analysis
- Predictive Analysis

Results


Success Rate
1.0
0.8
0.6
0.4
0.2
0.0
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020
Date

All Launch Site Names

- Find the names of the unique launch sites
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40
- Present your query result with a short explanation here

Conclusions

- User-friendly interface and intuitive design enable easy creation and customization of dashboards, reducing the learning curve for users
- Seamless data integration capabilities ensure access to comprehensive data from diverse sources, enhancing data analysis and decision-making
- Interactive visualization features empower users to explore data, uncovering insights and trends that drive business outcomes
- Robust collaboration and sharing functionalities facilitate teamwork and communication, fostering a data-driven culture within the organization and driving collective intelligence

Thank you!