Final Presentation for Data Science Capstone

Veronika Grundmane

01/08/2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

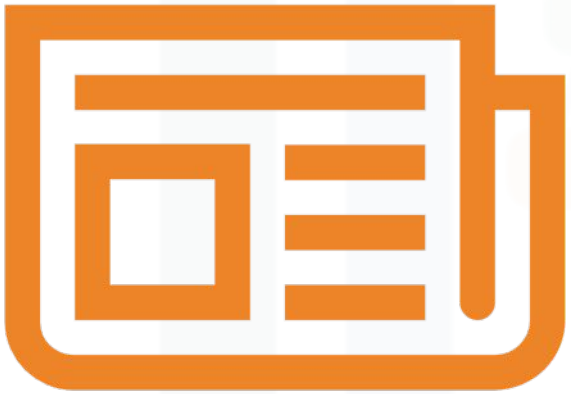
EXECUTIVE SUMMARY

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

INTRODUCTION

- Project background and context
- Space X promotes Falcon 9 rocket launches on its website at a price of 62 million dollars, while other providers charge over 165 million dollars for each launch. A significant portion of these savings is attributed to Space X's ability to reuse the first stage. Thus, by predicting the successful landing of the first stage, we can ascertain the cost of a launch. The project's objective is to develop a machine learning pipeline that can forecast whether the first stage will land successfully. This predictive capability can be valuable for a competing company that wishes to bid against Space X for a rocket launch.
- Problems you want to find answers
  - What elements decide the success of the rocket landing?
  - The interplay among different characteristics that dictate the rate of success in achieving a proper landing.
  - What specific operational conditions need to be fulfilled to ensure the success of the landing process?

METHODOLOGY



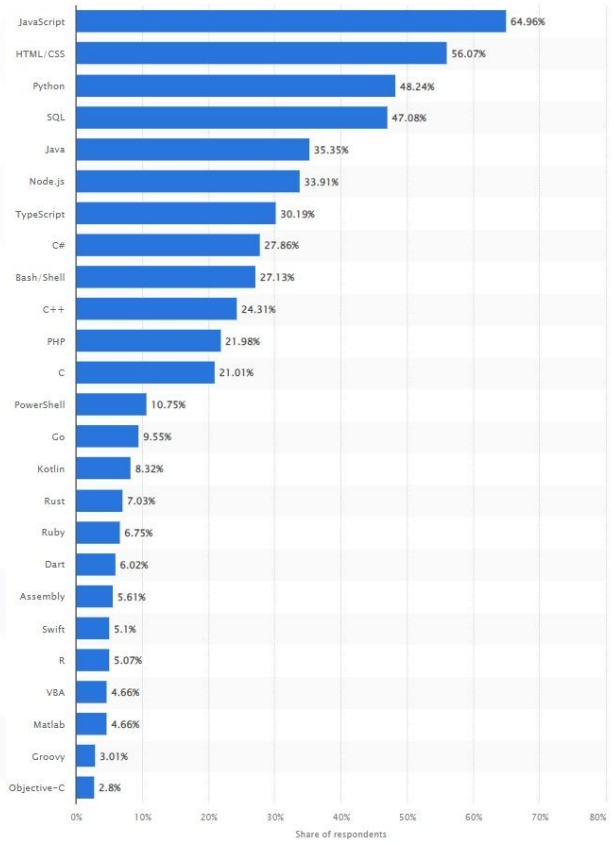
- Executive Summary
- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia
- Perform data wrangling
  - One-hot encoding was applied to categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

RESULTS

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

PROGRAMMING LANGUAGE TRENDS

Current Year



PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

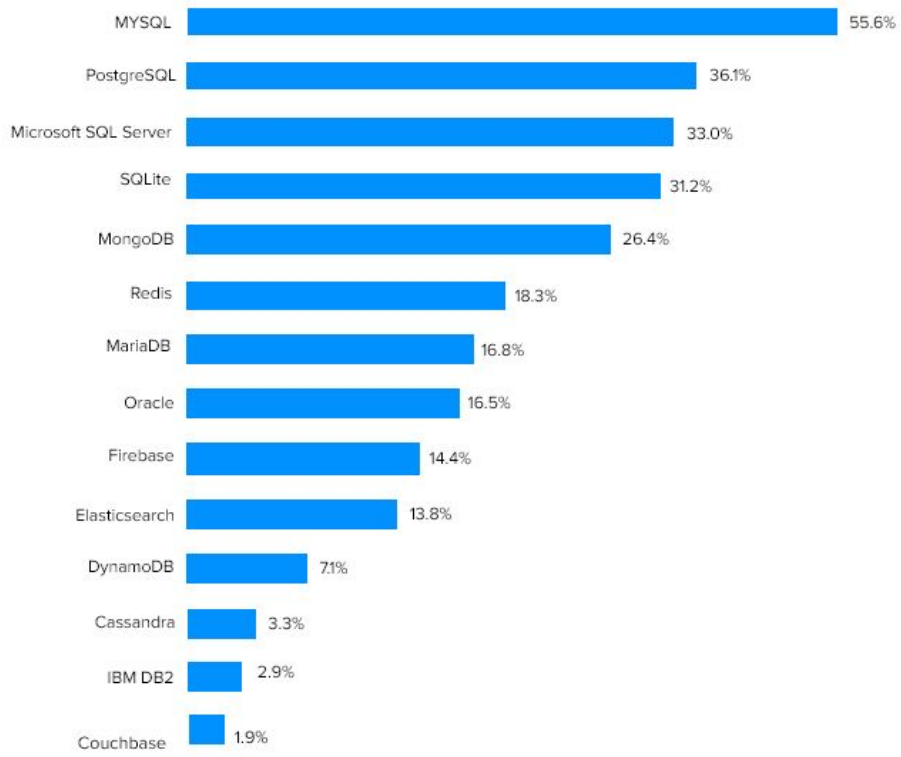
- Dominant Languages
- Emerging Languages
- Specialization and Diversity

Implications

- The top programming languages of the current year, as indicated by their usage and adoption rates, are critical to projects across various domains. Developers must ensure their proficiency in these languages to remain competitive in the job market.
- Alongside the established giants, there is a noticeable emergence of newer programming languages gaining traction. These languages might offer innovative features, enhanced performance, or better compatibility with modern computing paradigms. Staying informed about these newcomers can provide developers with a strategic advantage.
- Different programming languages are increasingly becoming associated with specific tasks or niches. Developers must make informed decisions about language selection based on the requirements of the project. Specialized languages can provide optimized solutions for particular domains.

DATABASE TRENDS

Current Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- Rise of NoSQL Databases
- Big Data and Analytics
- Cloud-Based Databases

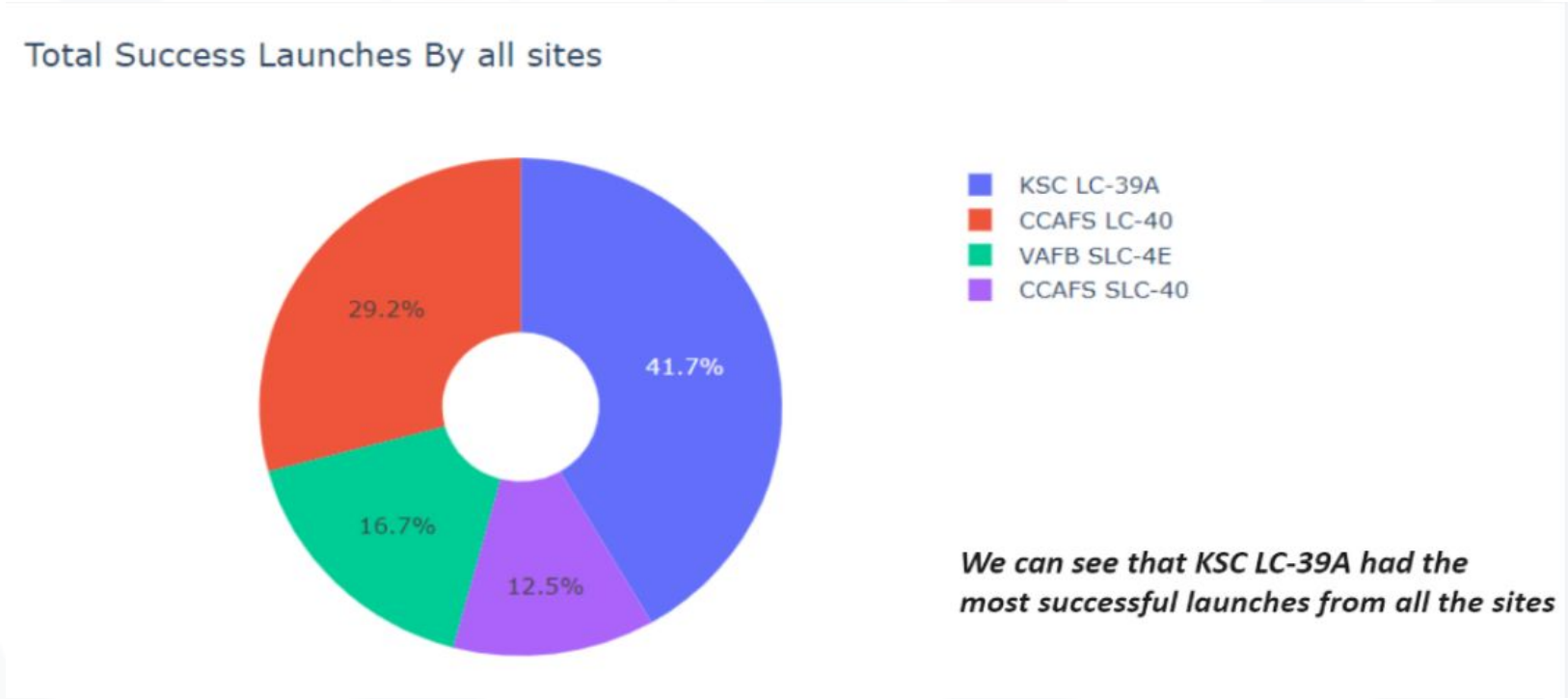
Implications

- NoSQL databases gained popularity due to their ability to handle unstructured and semi-structured data more efficiently than traditional relational databases. This trend implies that businesses are dealing with diverse and massive data types, such as social media posts, sensor data, and multimedia content.
- The growth of big data necessitated databases that can efficiently store, manage, and process vast amounts of data. This trend led to the development of distributed databases and technologies like Hadoop and Spark for large-scale data processing and analysis.
- Cloud computing revolutionized how databases are deployed and managed. Cloud-based databases offer scalability, flexibility, and cost-effectiveness. The implication here is the shift from on-premises to cloud-based infrastructure for database management.

DASHBOARD

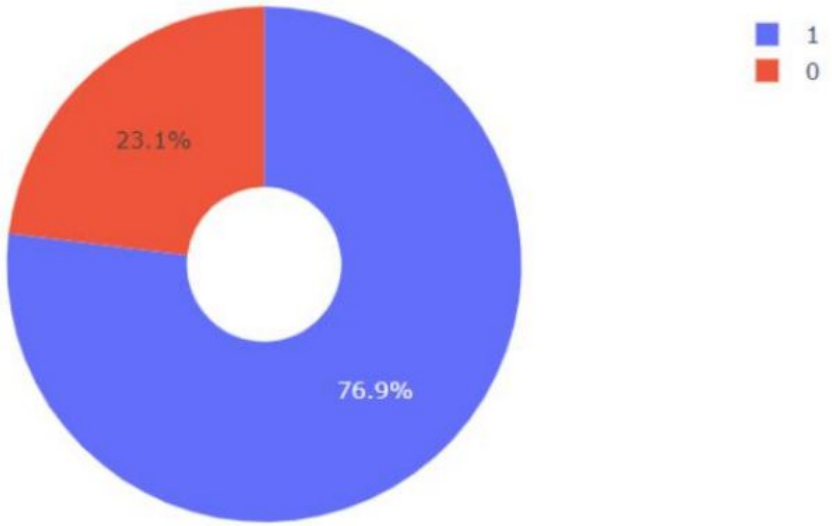
https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=stories-dashboards

DASHBOARD TAB 1



DASHBOARD TAB 2

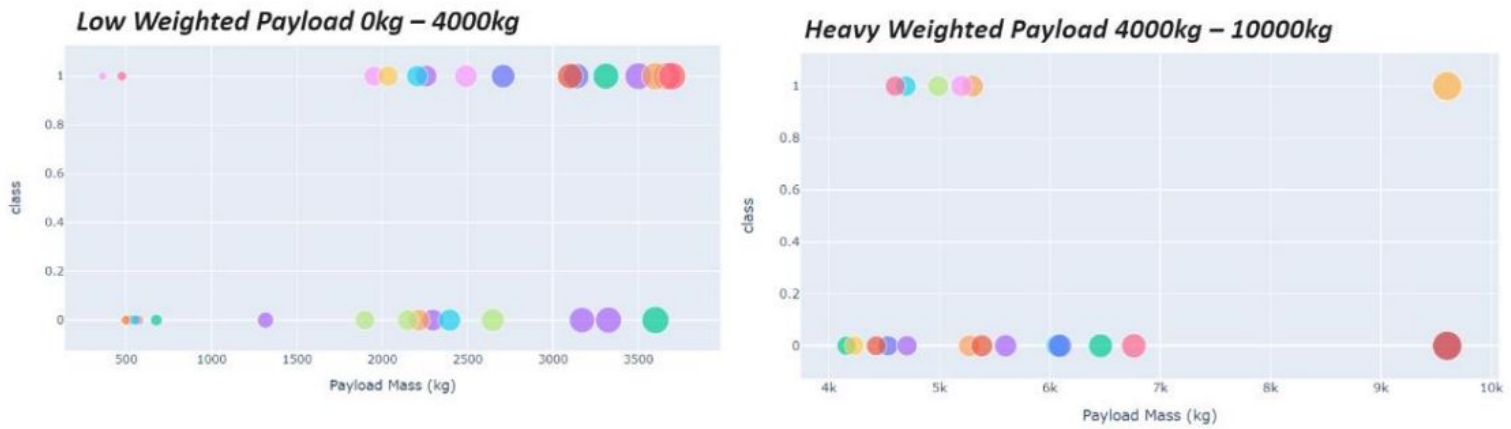
Pie chart showing the Launch site with the highest launch success ratio



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

DASHBOARD TAB 3

Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider



We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

DISCUSSION

Building a reliable predictive model for such a complex and critical task is not without challenges. Noisy or limited data, unforeseen external factors, and the inherent complexity of rocket launches can pose difficulties.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Feature Importance
- Ethical Considerations
- Challenges

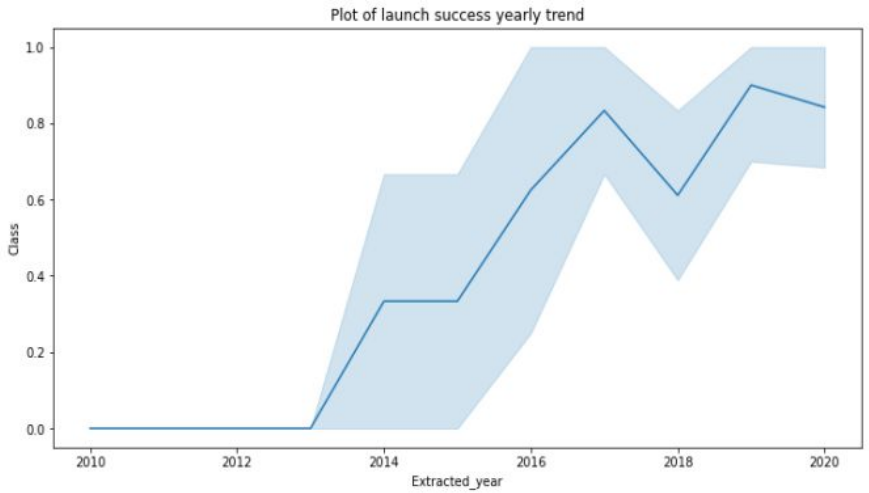
Implications

- Understanding which features contribute most to the prediction can provide insights into what factors are most critical for a successful landing. This information could be valuable not only for predictive purposes but also for improving launch processes.
- It's important to consider the ethical implications of using predictive capabilities in competitive bidding scenarios. Ensuring fair competition and transparency is essential.
- Building a reliable predictive model for such a complex and critical task is not without challenges. Noisy or limited data, unforeseen external factors, and the inherent complexity of rocket launches can pose difficulties.

CONCLUSION

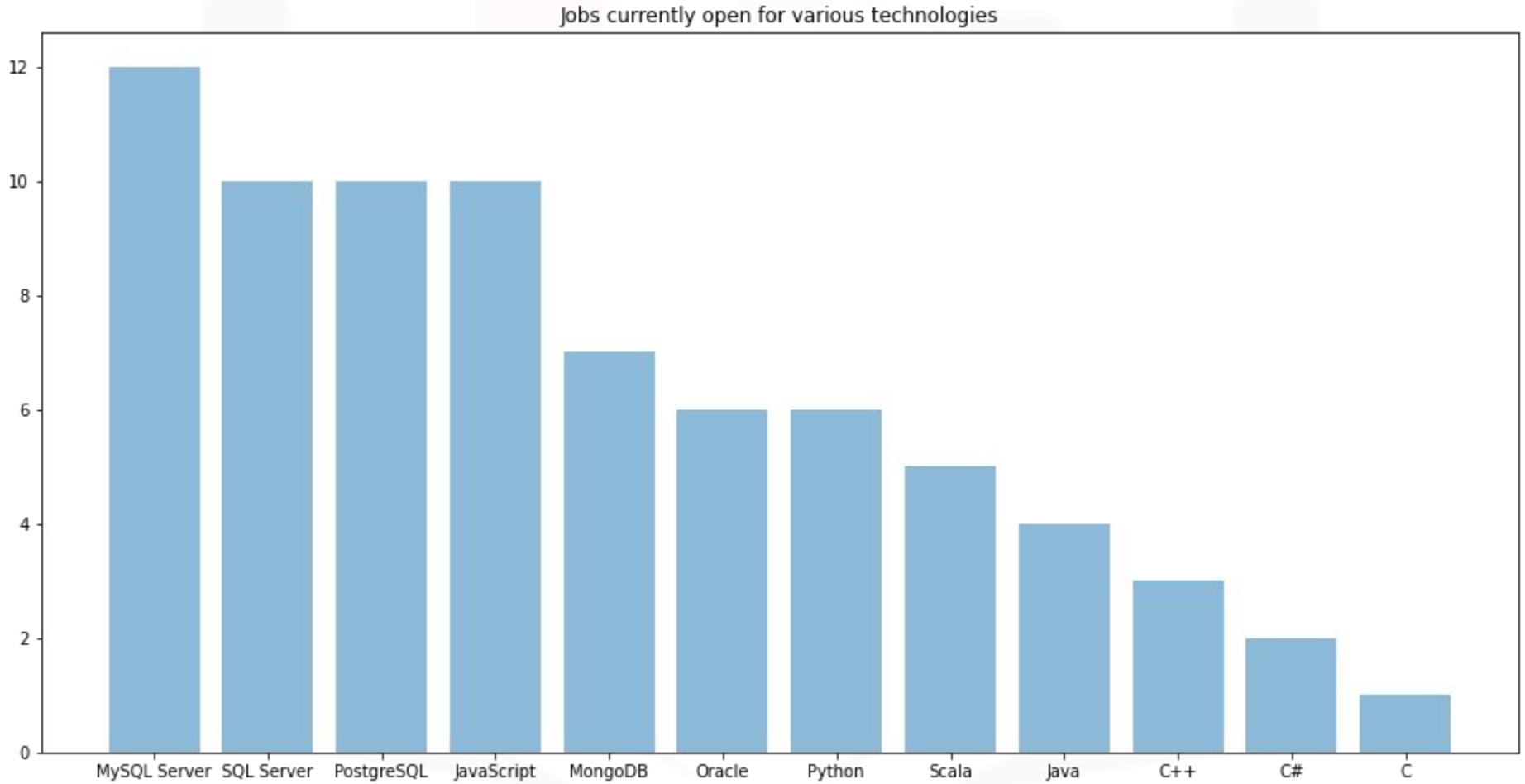
- We can conclude that:
- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

APPENDIX



- From the plot, we can observe that success rate since 2013 kept on increasing till 2020.

JOB POSTINGS



POPULAR LANGUAGES