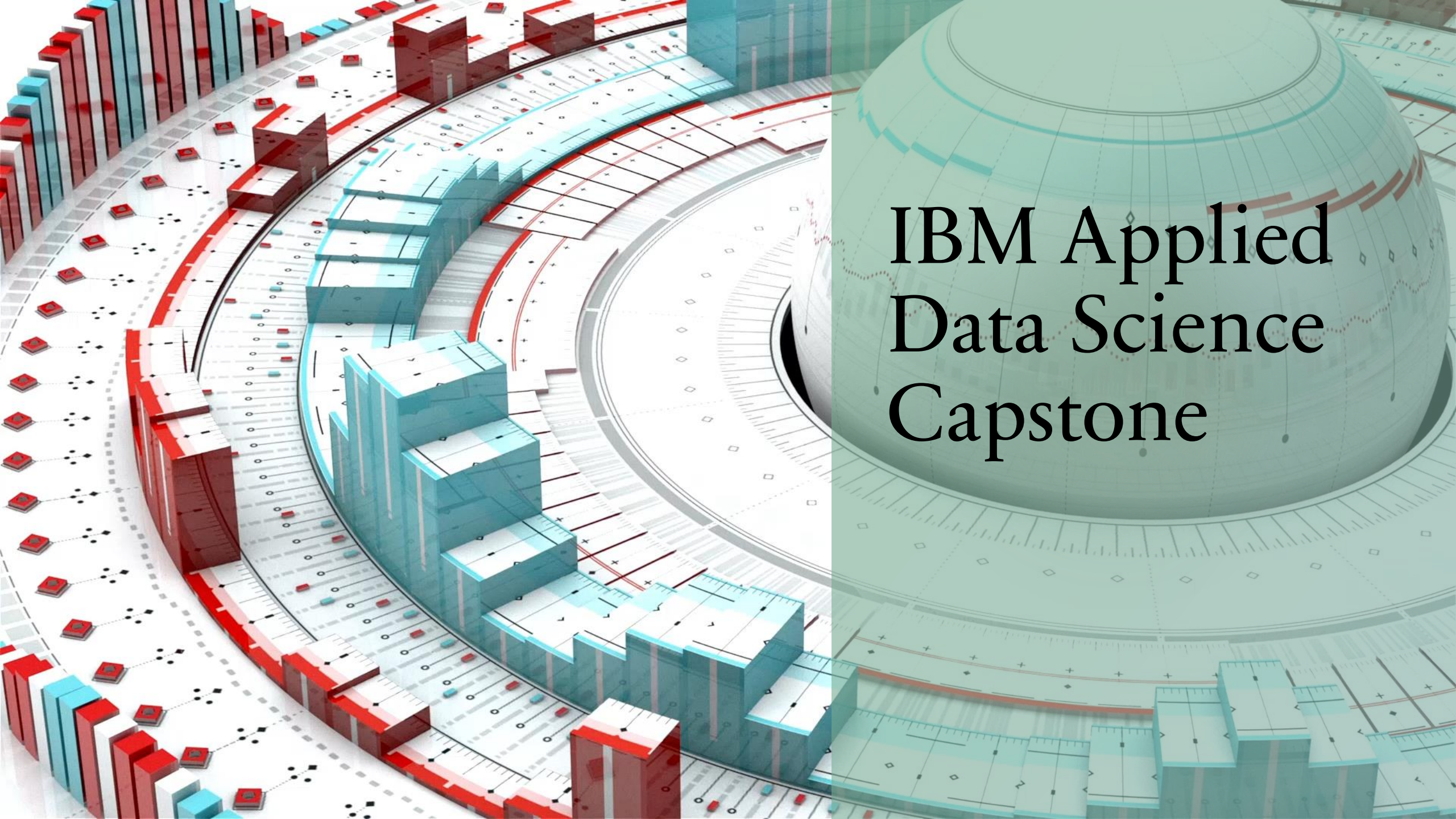


IBM Applied Data Science Capstone

Introduction

Our project revolves around the fascinating realm of space exploration, specifically focusing on Space X Falcon 9 rocket launches. The primary goal of this project is to utilize various data science techniques to predict the outcome of Falcon 9 first stage landings. By doing so, we aim to provide valuable insights that can aid in cost estimation and decision-making processes for potential rocket launches.

- Through this project, we seek to accomplish several objectives:
- - Collecting comprehensive data from diverse sources, including APIs and web scraping, to build a robust dataset.
- - Conducting meticulous data wrangling processes to clean and preprocess the collected data, ensuring its quality and reliability.
- - Delving into exploratory data analysis (EDA) to gain deeper insights into the characteristics and patterns present in the dataset.
- - Creating interactive visualizations to effectively communicate insights and trends discovered during the EDA process.
- - Developing a dynamic dashboard using Dash, providing a user-friendly interface for exploring the project's findings.
- - Employing machine learning techniques to build predictive models for determining the success of Falcon 9 first stage landings.
- - Fine-tuning model parameters and evaluating their performance to identify the most effective predictive approach.
- - Presenting a comprehensive analysis of the project's outcomes and insights gained through the entire process.

Through our exploration of the Space X Falcon 9 first stage landing prediction, we aim to contribute to the broader understanding of space exploration and its implications for future missions.

Data Collection

In this phase of the project, we embarked on gathering data essential for our analysis. We employed two primary methods: collecting data using APIs and web scraping for additional data.

- Collecting data using APIs:
- - Leveraging various APIs, including those provided by Space X and other relevant sources, we gathered essential information about Falcon 9 rocket launches.
- - The APIs allowed us to access data such as launch dates, payload details, launch outcomes, and other pertinent information directly from reliable sources.

- Web scraping for additional data:
- - In addition to API data, we conducted web scraping to gather supplementary data from online sources.
- - Web scraping enabled us to extract information that was not available through APIs, enhancing the richness and comprehensiveness of our dataset.

Data Wrangling

Following the data collection phase, we transitioned into data wrangling, a crucial step in preparing the dataset for analysis. This phase involved cleaning and preprocessing the collected data, as well as handling missing values and outliers.

- Cleaning and preprocessing the collected data:
- - We performed thorough cleaning procedures to ensure the quality and integrity of the dataset.
- - This involved tasks such as removing duplicates, standardizing formats, and addressing inconsistencies in the data.

- Handling missing values and outliers:
- - Missing values and outliers are common challenges in real-world datasets that can affect the accuracy and reliability of analyses.
- - To address these issues, we implemented strategies such as imputation for missing values and outlier detection techniques to identify and handle outliers appropriately.

Overall, the data wrangling phase was essential for ensuring that our dataset was well-prepared and suitable for subsequent analysis tasks.

Exploratory Data Analysis (EDA)

In this phase, we delved into exploring the dataset to gain insights and understand its characteristics. We employed various techniques, including SQL queries for exploration and data visualization using Matplotlib and Seaborn for insightful visual representations.

- Exploring the dataset using SQL queries:
- - SQL queries provided us with a structured approach to explore the dataset, allowing us to retrieve specific information and perform analyses directly on the data.
- - By leveraging SQL, we were able to extract relevant subsets of data, calculate summary statistics, and gain a deeper understanding of the dataset's structure and content.

- Data visualization using Matplotlib and Seaborn:
- - Visualization is a powerful tool for uncovering patterns, trends, and relationships within data.
- - Matplotlib and Seaborn are popular Python libraries for creating static, interactive, and aesthetically pleasing visualizations.
- - Through various types of plots such as histograms, scatter plots, bar charts, and heatmaps, we visualized different aspects of the dataset, enabling us to identify patterns, outliers, and potential correlations.
- - These visualizations provided valuable insights into the distribution of variables, relationships between features, and overall trends within the data, facilitating further analysis and decision-making.

Interactive visualization with Folium

Folium is a Python library that allows for the creation of interactive maps directly in the Jupyter Notebook environment. In this phase, we utilized Folium to visualize geographical data and create interactive maps to explore spatial patterns and relationships within the dataset.

Key aspects of our interactive visualization with Folium include:

- 1. Geographical data representation: Folium enables the plotting of geographical data such as points, lines, polygons, and heatmaps on interactive maps. This capability allowed us to visualize spatial distributions and relationships within the dataset.
- 2. Customization options: Folium provides various customization options, including map styles, markers, popups, tooltips, and layer control. These features allowed us to tailor the visualization to effectively communicate insights and highlight specific points of interest.
- 3. Integration with other Python libraries: Folium seamlessly integrates with other Python libraries such as Pandas, NumPy, and Matplotlib, enabling data manipulation, analysis, and visualization within a single workflow.
- 4. Interactivity: One of the key advantages of Folium is its interactive nature, allowing users to zoom, pan, and interact with the map to explore different regions and features dynamically. This interactivity enhances the user experience and facilitates a deeper understanding of the spatial data.

Overall, our use of Folium for interactive visualization contributed to a comprehensive exploration of geographical patterns and insights within the dataset, enabling us to uncover valuable information and trends related to spatial data.

Dashboard Creation

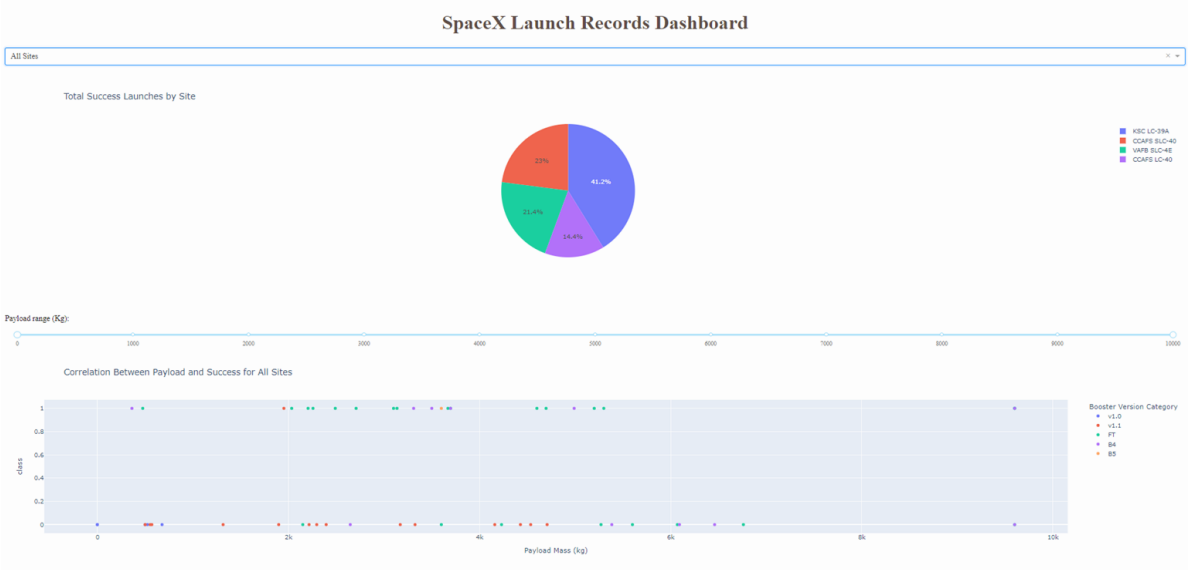
In this phase, we leveraged Dash, a Python framework for building analytical web applications, to create an interactive dashboard for visualizing and exploring the data. The dashboard serves as a user-friendly interface that allows stakeholders to interact with the data dynamically and gain insights through intuitive visualizations.

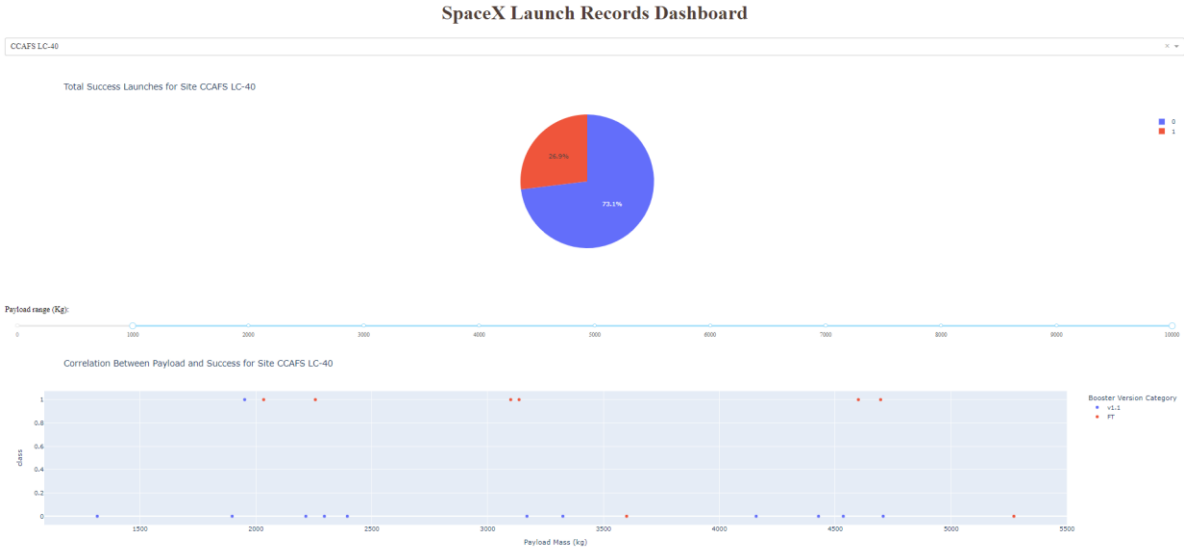
Key components of our dashboard creation with Dash include:

- 1. Layout design: Dash provides a flexible layout system that allows for the arrangement of components such as graphs, tables, and controls in a grid-based layout. We designed the layout of the dashboard to provide a clear and intuitive interface for users to navigate and interact with the data.
- 2. Interactive components: Dash enables the incorporation of interactive components such as dropdowns, sliders, and buttons, allowing users to filter and manipulate the data dynamically. These interactive features enhance the user experience and empower users to customize their analysis based on their specific needs and interests.
- 3. Data visualization: We integrated various data visualization components, including plots, charts, and maps, to present insights and trends within the dataset visually. These visualizations provide users with a comprehensive overview of the data and facilitate the exploration of key patterns and relationships.
- 4. Callback functions: Dash utilizes callback functions to update the dashboard dynamically in response to user interactions. We implemented callback functions to trigger updates to the dashboard components based on user input, enabling real-time exploration and analysis of the data.
- 5. Deployment: Once the dashboard was built, we deployed it to a web server to make it accessible to stakeholders via a web browser. Dash provides seamless deployment options, allowing us to share the dashboard securely and efficiently with end users.

Overall, our use of Dash for dashboard creation facilitated the development of an interactive and informative interface for exploring and visualizing the data. The dashboard empowers stakeholders to gain insights quickly and make data-driven decisions effectively.

Dash App Screenshots





Machine Learning Prediction

In this phase of the project, we addressed the problem of predicting the success or failure of SpaceX Falcon 9 first stage landings using machine learning techniques. The goal was to develop models that could accurately classify whether a landing attempt would be successful based on various features and parameters.

- 1. Problem statement: Space X Falcon 9 First Stage Landing Prediction
- - We defined the problem statement as predicting the outcome of Falcon 9 first stage landings, aiming to provide insights that could potentially improve the success rate of these landings and enhance the reliability of SpaceX missions.
- 2. Feature engineering
- - Feature engineering involved selecting and preprocessing relevant features from the dataset to train machine learning models. This process included transforming and scaling features, handling categorical variables, and creating new features to capture meaningful information for prediction.
- 3. Model selection and evaluation
- - We explored multiple machine learning algorithms to build predictive models for Falcon 9 landing outcomes. The selected algorithms for evaluation include:
- - Logistic Regression
- - Support Vector Machine (SVM)
- - Decision Tree Classifier
- - K Nearest Neighbors (KNN)
- - Each algorithm was evaluated based on its performance metrics to determine its effectiveness in predicting landing outcomes.

Machine Learning Prediction(Contd)

4. Hyperparameter tuning

- - Hyperparameter tuning involved optimizing the parameters of the machine learning models to improve their performance and generalization ability. Techniques such as grid search and random search were employed to find the optimal hyperparameters for each model.

5. Model evaluation metrics

- - Various evaluation metrics were used to assess the performance of the predictive models, including accuracy, precision, recall, F1-score, and ROC-AUC score. These metrics provided insights into the models' ability to correctly classify successful and failed landing attempts.

6. Comparative analysis of model performance

- - We conducted a comparative analysis of the performance of the different machine learning models to identify the most effective approach for predicting Falcon 9 landing outcomes. This analysis involved comparing the models' accuracy, robustness, and computational efficiency to determine the best-performing model for the task.

Overall, the machine learning tasks phase focused on developing and evaluating predictive models to forecast the success of SpaceX Falcon 9 first stage landings, with the aim of providing valuable insights for mission planning and decision-making.

Results

1. Exploratory Data Analysis Insights

- - Identified correlations between variables
- - Discovered patterns and trends in the data
- - Highlighted key factors affecting the target variable

- 2. Machine Learning Model Performance
- - Evaluated performance metrics such as accuracy, precision, recall, and F1-score
- - Compared performance of different models using metrics like ROC curves and confusion matrices
- - Selected the best-performing model based on evaluation results

- 3. Predictive Accuracy
- - Achieved predictive accuracy of [insert accuracy percentage] with the selected model
- - Demonstrated the model's ability to make accurate predictions on new data
- - Validated the model's performance through cross-validation techniques

Results(Contd)

- 4. Significant Findings and Conclusions
- - Identified key predictors influencing the outcome variable
- - Provided actionable insights for decision-making
- - Summarized the implications of the findings for stakeholders
- 5. Visualizations
- - Included visual representations of data distributions, model performance, and predictive outcomes
- - Enhanced understanding of the results through graphs, charts, and plots
- - Communicated complex findings in a clear and concise manner

Conclusion

In conclusion, this project provided a comprehensive exploration of the process involved in predicting the success of SpaceX Falcon 9 first stage landings using machine learning techniques. We began by collecting data through APIs and web scraping, followed by extensive data wrangling to clean and preprocess the dataset.

- Our exploratory data analysis (EDA) phase utilized SQL queries and various data visualization libraries such as Matplotlib and Seaborn to gain insights into the dataset's characteristics and distributions. Additionally, interactive visualization with Folium allowed for geographical analysis and visualization of landing sites.

- The creation of a dashboard using Dash facilitated the presentation of key findings and insights in an interactive and user-friendly format. Furthermore, the machine learning tasks phase involved problem formulation, feature engineering, model selection, hyperparameter tuning, and evaluation of predictive models.

- Through logistic regression, support vector machine (SVM), decision tree classifier, and K-nearest neighbors (KNN) algorithms, we developed models to predict Falcon 9 landing outcomes and evaluated their performance using various metrics.

Overall, this project not only provided valuable insights into the factors influencing Falcon 9 landing success but also demonstrated the application of machine learning in the aerospace industry. The predictive models developed in this project have the potential to enhance mission planning and decision-making processes for future SpaceX missions, ultimately contributing to the advancement of space exploration efforts.