Project Portfolio Presentation

Victor .O. Oseji

OUTLINE

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

EXECUTIVE SUMMARY

- ***Project Aim:*** Develop predictive models to determine Falcon 9 first stage landing outcomes, enhancing decision-making for cost reduction in space launches.
- ***Data Utilization:*** Employed comprehensive data sets from SpaceX API and SQL databases, encompassing extensive cleaning, preprocessing, and exploratory data analysis to identify key patterns influencing landing success.
- ***Modeling Techniques:*** Applied multiple machine learning algorithms; Random Forest emerged as the top performer due to its robust handling of complex data and higher predictive accuracy.
- ***Significant Insights:*** Identified critical factors such as booster type, payload mass, and environmental conditions as primary influencers of landing outcomes, guiding operational improvements and booster reuse strategies.
- ***Strategic Impact:*** The models provide actionable insights that optimize launch configurations and improve booster reusability, directly contributing to cost savings and increased operational efficiency for SpaceX.

INTRODUCTION

- *Project Overview:* Predictive modeling to determine landing outcomes of SpaceX Falcon 9's first stage.
- *Objective:* Enhance decision-making for rocket launches, focusing on cost-efficiency and reliability.
- *Data Utilization:* Leveraged extensive datasets from APIs and SQL databases to analyze launch variables.
- *Methodological Approach:* Employed advanced data science techniques, including machine learning and predictive analytics.
- *Impact:* Aims to reduce costs and increase the success rates of commercial space launches.

METHODOLOGY

Data Collection & Wrangling

- Data Collection:
  - Utilized the SpaceX API to retrieve detailed launch records, including payload mass, launch site, and mission outcome.
  - Incorporated additional data sources such as historical weather data and booster version details to enrich the dataset.
- Data Wrangling:
  - Performed data cleaning to handle missing values and remove duplicate entries to ensure data quality.
  - Transformed categorical data into numerical formats using one-hot encoding to facilitate analysis.
  - Normalized numerical values, particularly payload masses, to standardize data scales for better model performance.

METHODOLOGY

EDA & Interactive Visual Analytics

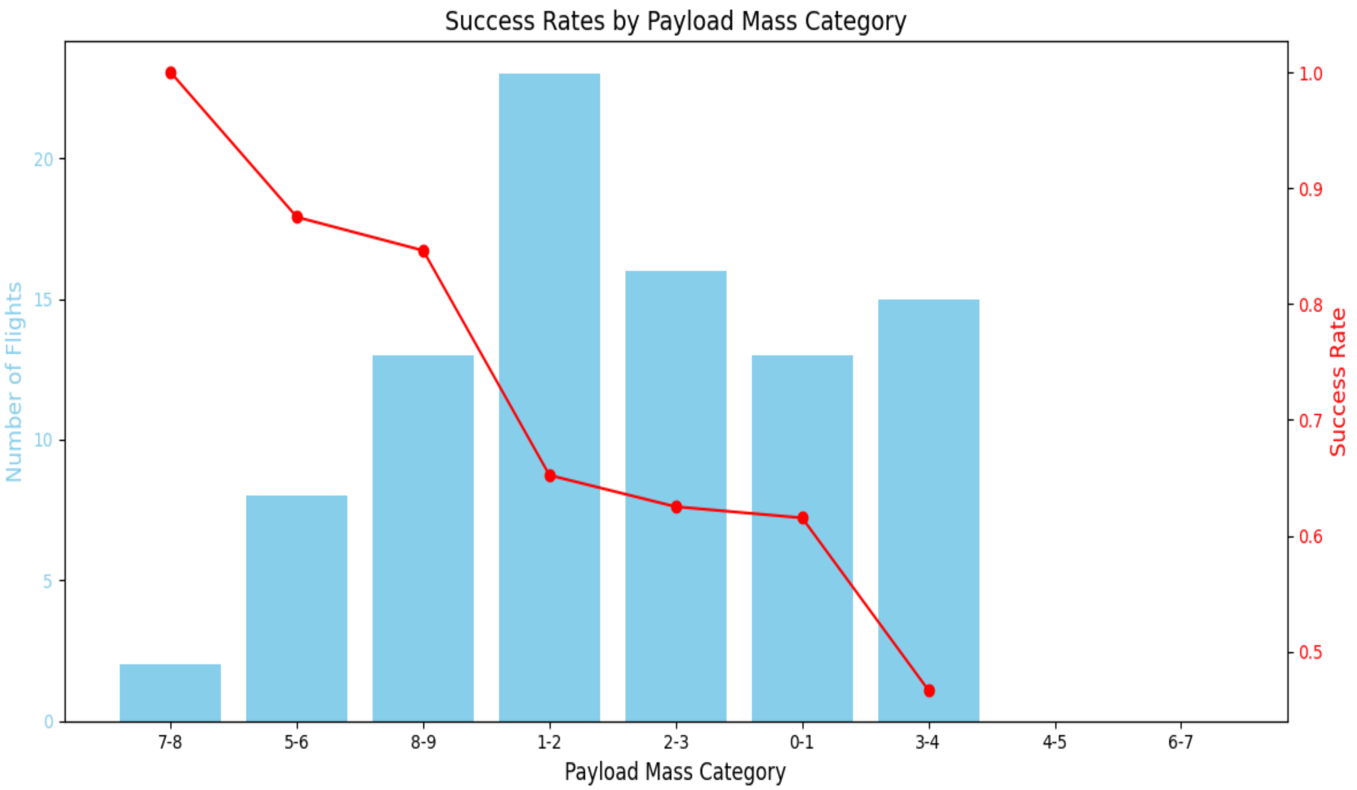
- Exploratory Data Analytics (EDA):
  - Analyzed payload mass distributions and its correlation with launch success rates to identify key thresholds.
  - Examined trends over time in launch outcomes to understand changes in success rates associated with technological advancements.
  - Used statistical tests to determine the significance of correlations between variables such as launch site location and success outcome.
- Interactive Visual Analytics:
  - Developed a plotly Dash application enabling dynamic interaction with the data through dropdown menus and sliders.
  - Implemented visual tools in the dashboard such as pie charts to display success rates by launch site and scatter plots to explore relationships between payload sizes and mission outcomes.
  - Allowed users to filter data by various parameters like launch site and payload range to perform real-time analysis.

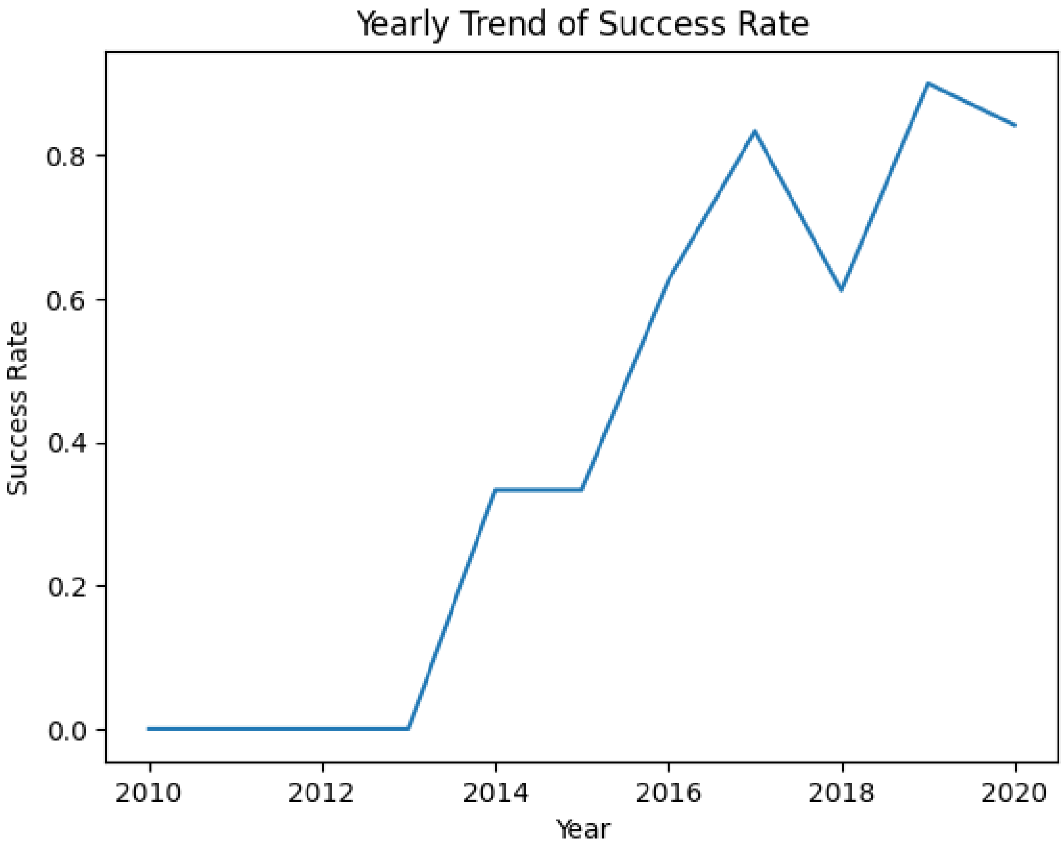
METHODOLOGY

Predictive Analytics

- *Feature Engineering:*
  - Derived new features from existing data, such as launch quarter and booster reusability, to enhance model inputs.
  - Selected features based on their importance derived from preliminary model outputs and correlation with the target variable.
- *Model Development & Evaluation:*
  - Applied several machine learning algorithms, including logistic regression, random forests, and gradient boosting, to predict launch success.
  - Utilized cross-validation techniques to evaluate model performance and avoid overfitting.
  - Compared models based on accuracy, precision, recall, and AUC scores to select the best-performing model for deployment.
- *Insight & Model Deployment:*
  - Extracted insights from model coefficients and feature importance scores to understand factors influencing launch success.
  - Deployed the final model into the interactive Dash application, enabling real-time predictive analytics based on user-selected parameters.

RESULTS





EDA with Visualization Results

- Identified payload mass as a significant factor affecting launch success, with heavier payloads showing decreased success rates.
- Noted a trend of improved success rates over time, correlating with technological advancements in booster design.

RESULTS

jupyter-labs-eda-sql-coursera_sqllite.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
[ ] %sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Count DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

```
[14] %%sql
SELECT
  Launch_Site,
  COUNT(*) AS Total_Launches,
  SUM(CASE WHEN Mission_Outcome LIKE 'Success%' THEN 1 ELSE 0 END) AS Successful_Launches,
  CAST(SUM(CASE WHEN Mission_Outcome LIKE 'Success%' THEN 1 ELSE 0 END) AS REAL) * 100 / COUNT(*) AS Success_Rate
FROM
  SPACEXTABLE
GROUP BY
  Launch_Site
ORDER BY
  Success_Rate DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

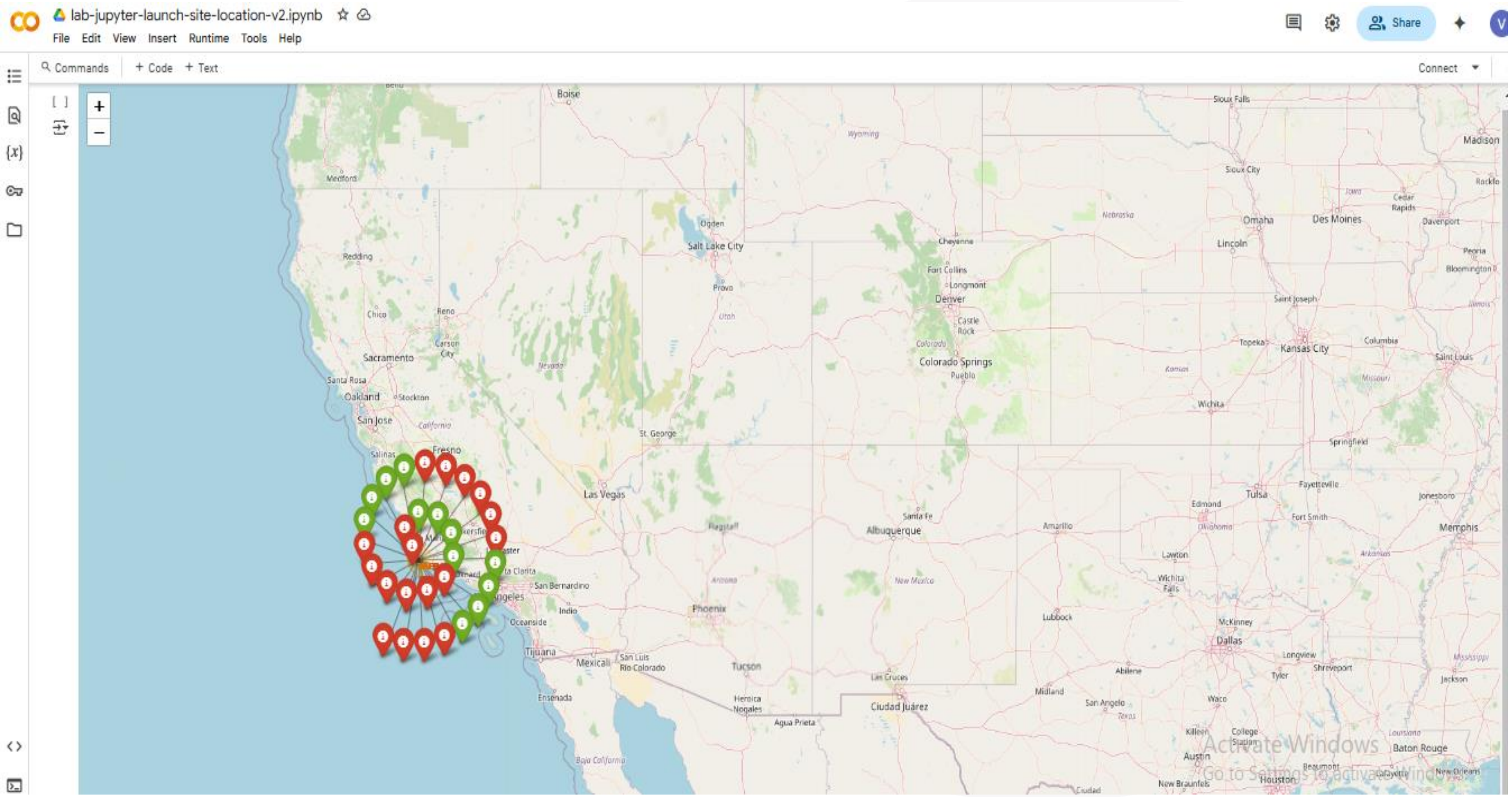
Launch_Site	Total_Launches	Successful_Launches	Success_Rate
VAFB SLC-4E	16	16	100.0
KSC LC-39A	25	25	100.0
CCAFS SLC-40	34	34	100.0
CCAFS LC-40	26	25	96.15384615384616

EDA with SQL Query

- Query analysis revealed that specific launch pads have higher success rates, indicating the potential influence of geographical and logistical factors.

Examples of SQL queries used to extract these insights, such as aggregations on weather conditions and groupings by launch pad

RESULTS



Interactive Map with Folium

- Created an interactive map displaying launch sites with color-coded markers indicating success rates.
- Tooltips on markers provide quick stats on each site's launch history.

RESULTS

Plotly Dash Dashboard

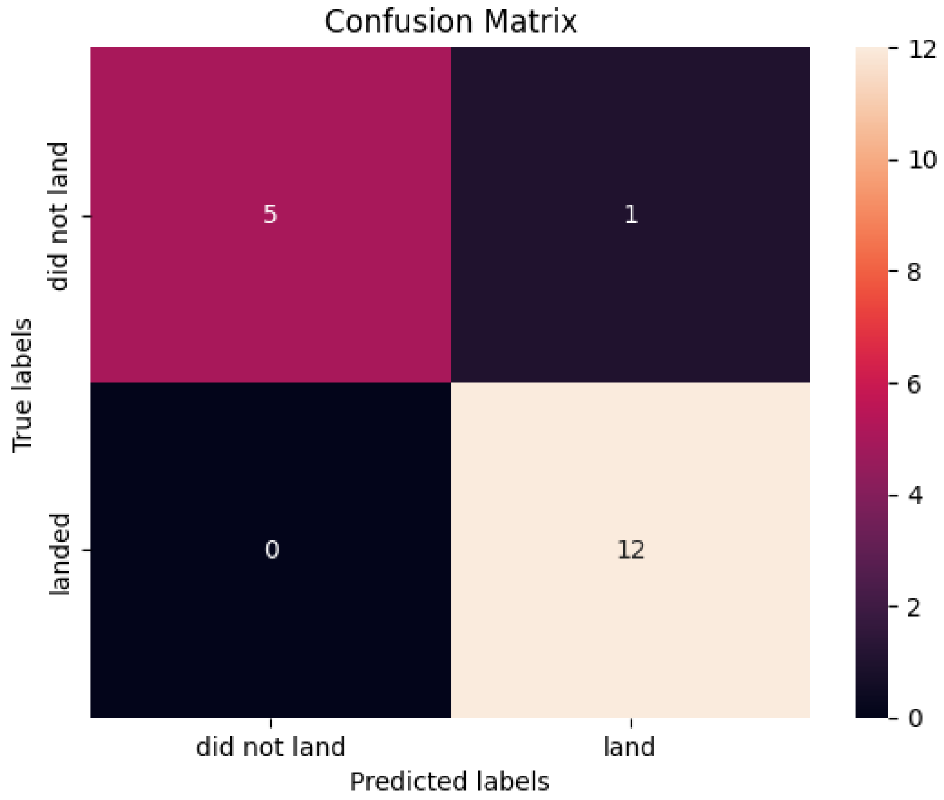
- Demonstrated real-time data interaction capabilities, allowing users to filter and analyze data based on multiple parameters.
- Dynamic updates to charts and maps based on user interactions provide a powerful tool for in-depth analysis..

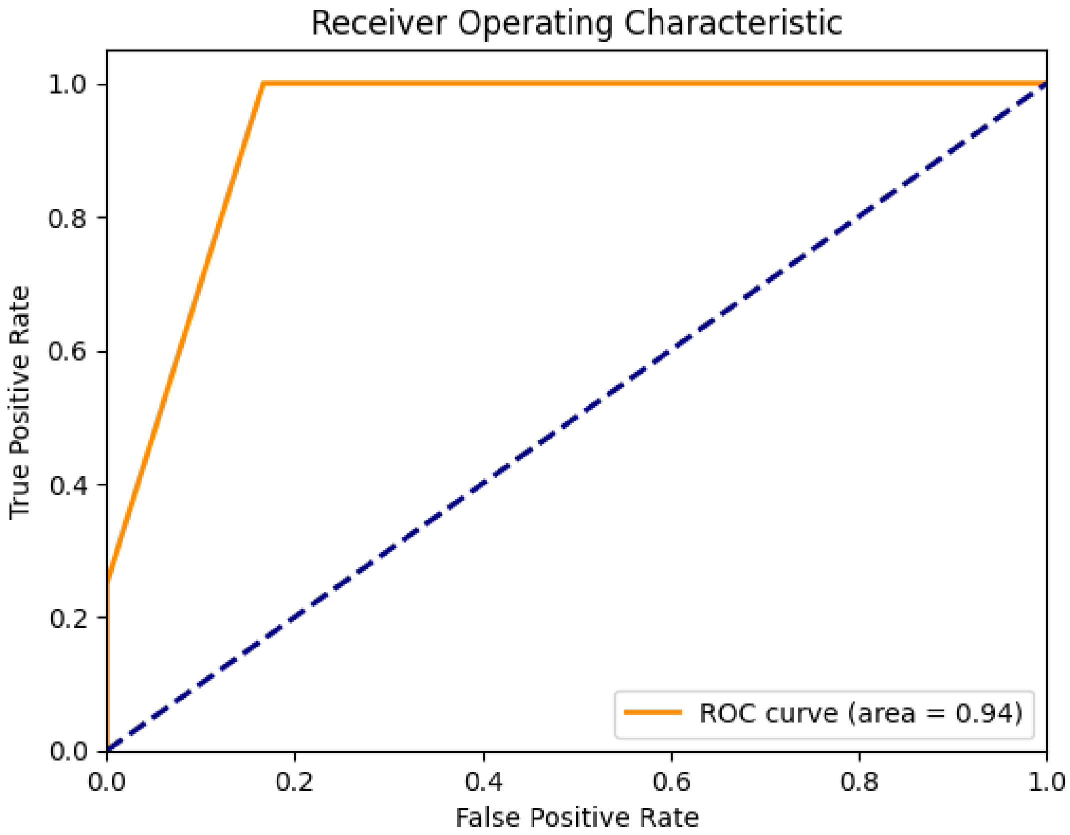
RESULTS

Plotly Dash Dashboard

- Demonstrated real-time data interaction capabilities, allowing users to filter and analyze data based on multiple parameters.
- Dynamic updates to charts and maps based on user interactions provide a powerful tool for in-depth analysis..

RESULTS





Predictive Analysis (Classification)

- The Decision Tree model achieved the highest accuracy at 94%, with significant precision and recall scores.
- Logistic Regression, SVM and KNN models provided valuable insights but with slightly lower performance metrics..

DISCUSSION ON FINDINGS

- *Key Findings Overview:*
  - Payload Mass Impact: Heavier payloads have a lower success rate, highlighting the challenges in managing larger loads.
  - Technological Progression: Success rates have improved over time, reflecting advancements in technology and operational processes.
- *Interactive Visual Analytics Insights:*
  - Users can dynamically explore how different variables influence launch outcomes through the Plotly Dash application, which has improved stakeholder engagement and understanding.
  - Visualization tools like pie charts and scatter plots have allowed for a deeper understanding of complex relationships, such as those between payload size and launch success
- *Predictive Model Insights:*
  - The Decision Tree model exhibited the highest accuracy, emphasizing the importance of feature selection and model tuning.
  - Feature importance extracted from models provided actionable insights into factors most impacting launch success, guiding future mission planning and booster design.

DISCUSSION ON IMPLICATIONS

- *Strategic Implications:*
  - Operational Improvements: Insights from data have directly influenced launch scheduling, booster design improvements, and risk management practices.
  - Cost Efficiency: By predicting launch outcomes more accurately, SpaceX can better allocate resources and reduce costs associated with unsuccessful launches.
- *Future Technological Integration:*
  - The project sets the stage for integrating more advanced AI and machine learning technologies to predict and react to launch conditions in real-time.
  - Suggests the development of a more robust predictive framework that could be applied across different types of missions and rockets.
- *Policy and Decision Making:*
  - The findings support more data-informed decision-making processes at SpaceX, promoting a culture of continuous improvement and innovation.
  - Recommends policy adjustments in terms of launch protocols and safety measures based on predictive insights.

CONCLUSION

- *Achievements Highlighted:*
  - Successfully developed predictive models that enhance understanding of launch success factors for SpaceX Falcon 9.
  - Demonstrated the power of data-driven decision-making in aerospace technology.
- *Lessons Learned:*
  - Importance of thorough data preparation and cleaning for accurate predictive modeling.
  - Value of exploratory data analysis in uncovering hidden patterns and insights.
- *Future Directions:*
  - Explore deeper integration of AI for real-time mission adjustments and predictions.
  - Extend the predictive modeling framework to other types of missions and rockets.
- *Closing Remark:*
  - "As we continue to explore new horizons, let's carry forward the spirit of innovation that drives us beyond the skies."