01/05/2026

Winning Space Race with Data Science

IBM Developer
SKILLS NETWORK

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion
6. Appendix

Executive Summary

- Summary of methodologies:
  - Data was collected using the SpaceX API and Wikipedia web scraping, cleaned and structured through data wrangling, explored using SQL and visualizations, analyzed with interactive tools (Folium and Dash), and used to build machine learning models for predicting landing success.

- Summary of all results:
  - The analysis revealed key patters in launch success, with improved performance over time and variation across launch sites. Machine learning models were able to predict landing success with measurable accuracy, with tuned models providing the best performance.

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Introduction

- This project analyzes SpaceX launch data to understand factors affecting rocket landing success using data science techniques including API data collection, web scraping, visualization and machine learning.

- The project aims to identify what influences launch success and build a model capable of predicting whether a rocket landing will be successful based on historical launch data.

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

SECTION 1

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using SpaceX API (HTTP requests).
  - Additional data was extracted using web scraping from Wikipedia.
  - Both sources were combined to build the final dataset.
- Perform data wrangling:
  - Raw data was cleaned and structured, missing values were handled, a new variable Class was created:
    - 1 = Successful landing.
    - 0 = Failed / No landing.
  - Final dataset was saved for analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.
- Perform predictive analysis using classification models.

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Data Collection

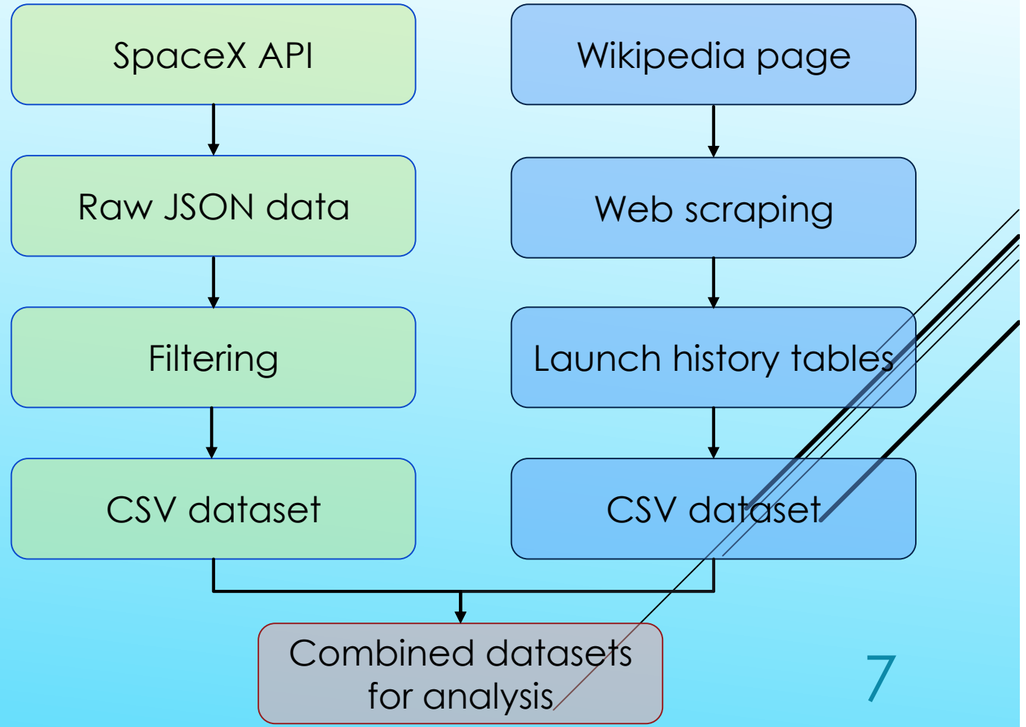
Data collected from SpaceX API (REST requests)

Additional historical data extracted via Wikipedia web scraping

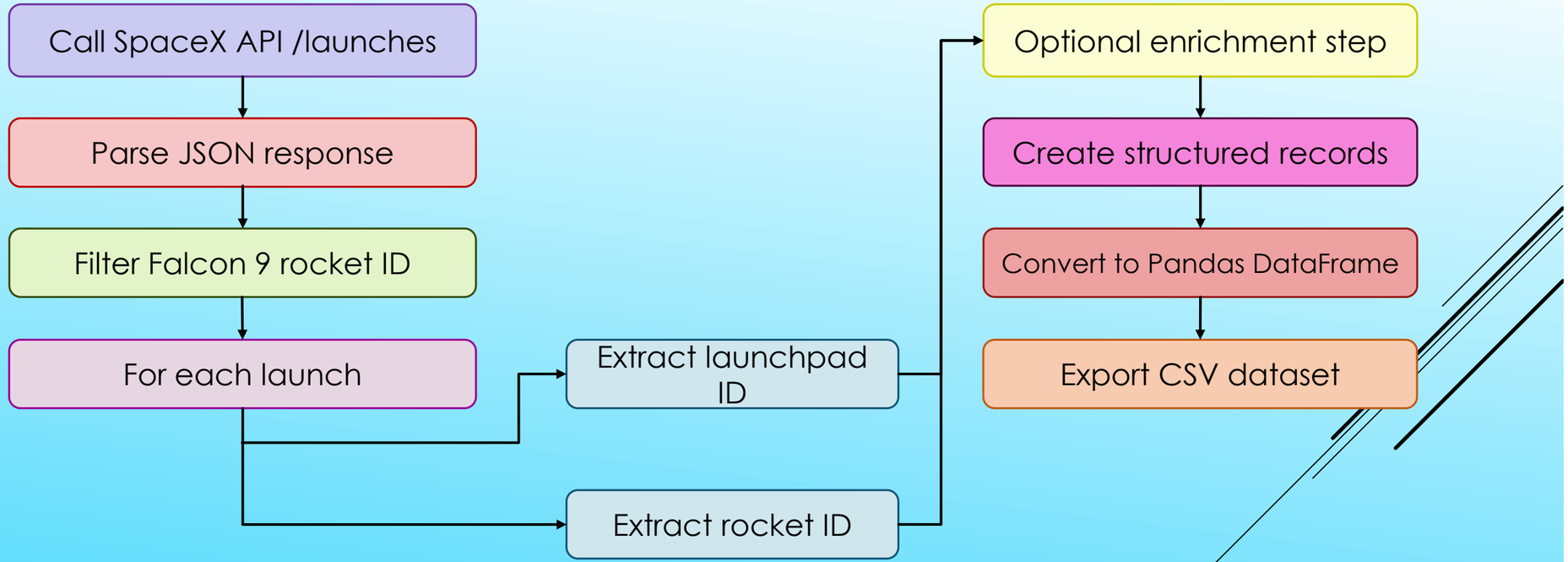
Combined structured and unstructured data sources

Focus on Falcon 9 launch records

Data stored in CSV format for analysis

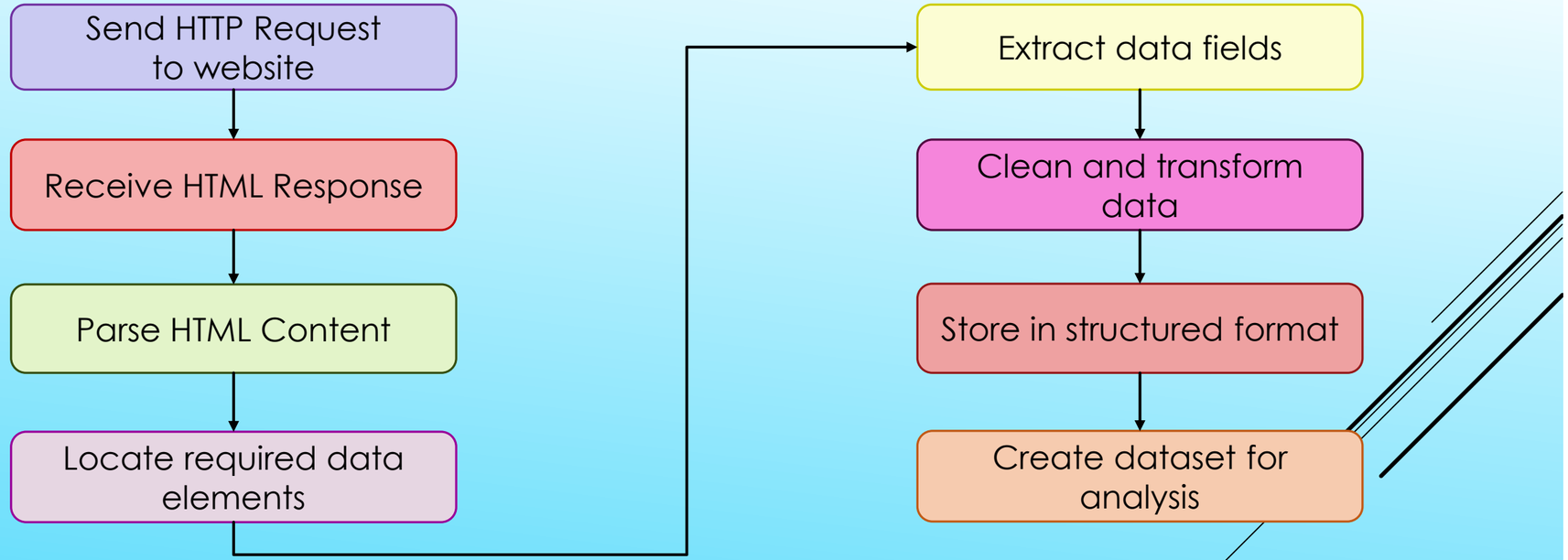


Data Collection – SpaceX API



APICollection

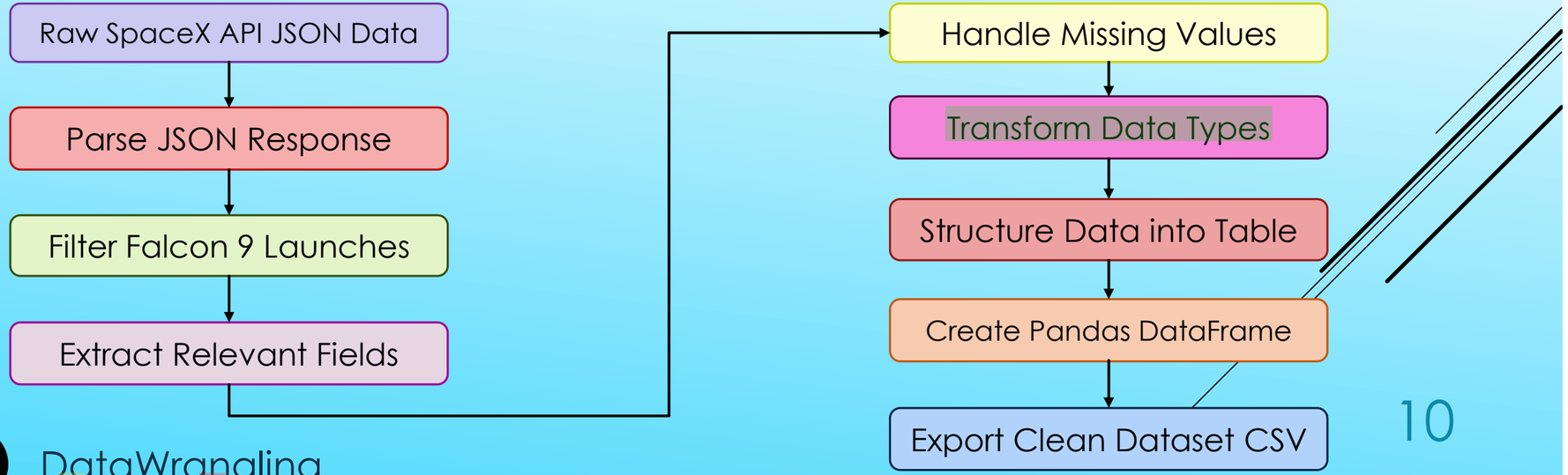
Data Collection - Scraping



WebScraping

Data Wrangling

The raw data collected from the SpaceX REST API was processed and cleaned to create a structured dataset suitable for analysis. This included filtering transforming and organizing the data into a tabular format using Python and Pandas.



DataWrangling

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Correct Partial Needs Work

EDA with Data Visualization

- Different charts were used to visualize key aspects such as launch success, frequency, and distribution of variables:
  - Bar Chart (Launch Success Count):
    - Used to compare the number of successful vs failed launches.
  - Bar Chart (Launches per Year):
    - Used to show how the number of launches changes over time.
  - Pie Chart (Success vs Failure Ratio):
    - Used to represent the proportion of successful and failed launches.
  - Histogram (Flight Number Distribution):
    - Used to analyze how launches are distributed over flight numbers.
  - Scatter Plot (Flight Number vs Success):
    - Used to explore relationship between mission order and outcome.

EDA

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

EDA with SQL

- Retrieved and explored all launch records from the data set.
- Filtered data to focus on Falcon 9 launches.
- Selected relevant columns (flight number, date, launch site, outcome).
- Calculated total launches and success vs failure counts.
- Grouped data by year and launch site to identify trends.
- Sorted and filtered data to analyze patterns and performance.

SQL

Build an Interactive Map with Folium

I created and added the following map objects to the Folium Map:

- Markers (Launch Sites) :

Used to visualize where SpaceX launches take place geographically.

- Circle Markers (Launch Frequency / Success) :

Used to represent launch frequency or success rate at each location.

- Lines (Connections) :

Used to illustrate relationships or possible launch paths between locations.

- Map Layers (Interactive View) :

Help users better analyze spatial distribution and patterns.

InteractiveMapsFolium

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Build a Dashboard with Plotly Dash

Plots / Graphs:

- Pie Chart (Success vs Failure) :

To display the proportion of successful and failed launches.

Interactions:

- Dropdown Menu (Launch Site Selection) :

To allow users to select a specific launch site.

- Interactive Callback (Dynamic Update) :

To update the pie chart when a launch site is selected.

LaunchSiteDashApp

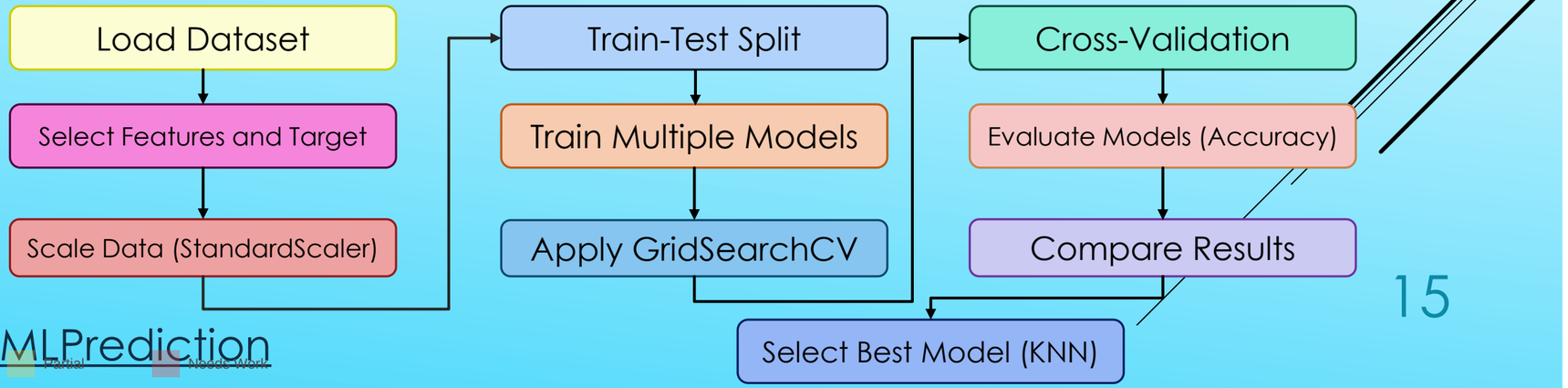
Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Predictive Analysis (Classification)

- The dataset was prepared by selecting features and scaling the data.
- Multiple classification models were trained and optimized using GridSearchCV.
- Models were evaluated using accuracy and the best-performing model (KNN), was selected.
- ¿Why KNN performed best? Because it captures local patterns in the data and achieved the highest accuracy after tuning.



MLPrediction

Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

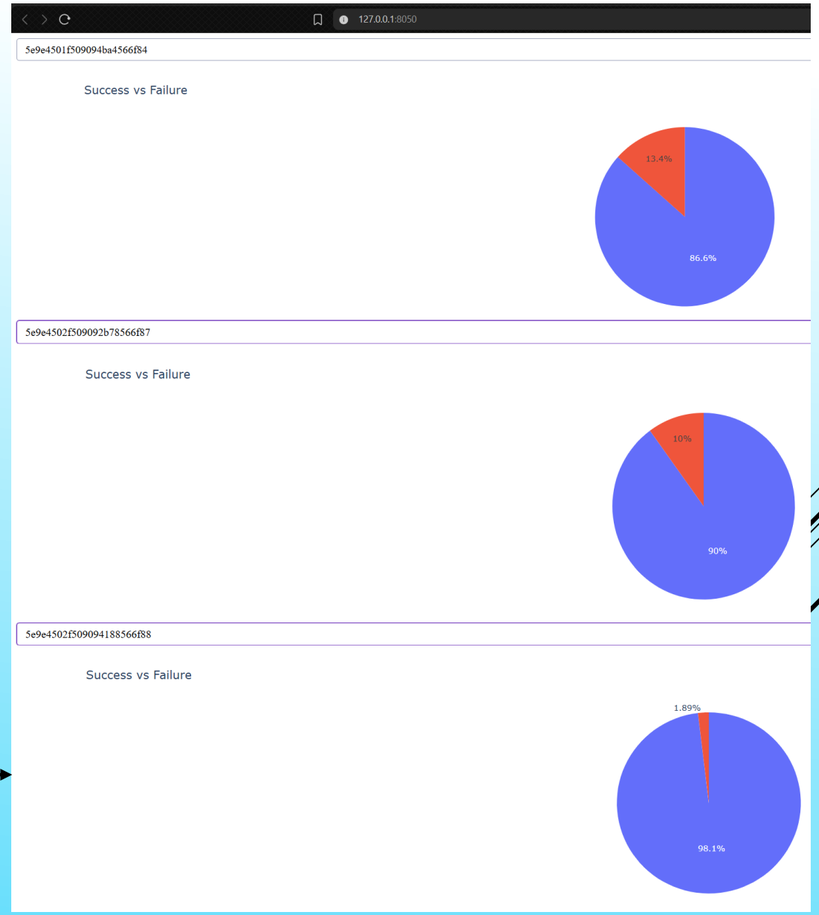
Results

- Exploratory data analysis results:

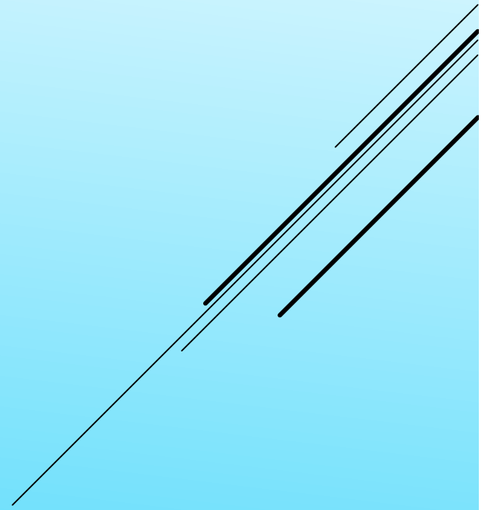
EDA revealed patterns in SpaceX launch data including differences in success rates across launch sites and trends over time. The analysis helped identify key relationships between flight number, launch site, and launch outcome.

- Predictive analysis results:

Multiple classification models were trained to predict launch success. After comparison, the KNN model achieved the highest accuracy (~0.97) and was selected as the best performing model.

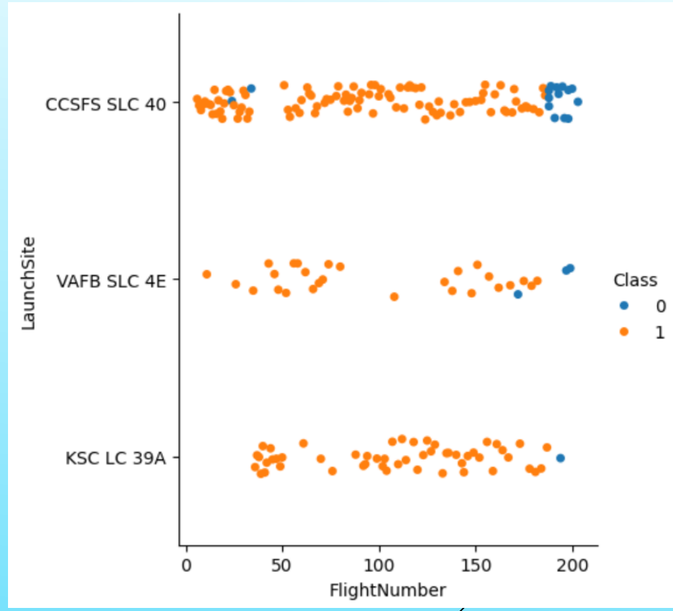


SECTION 2



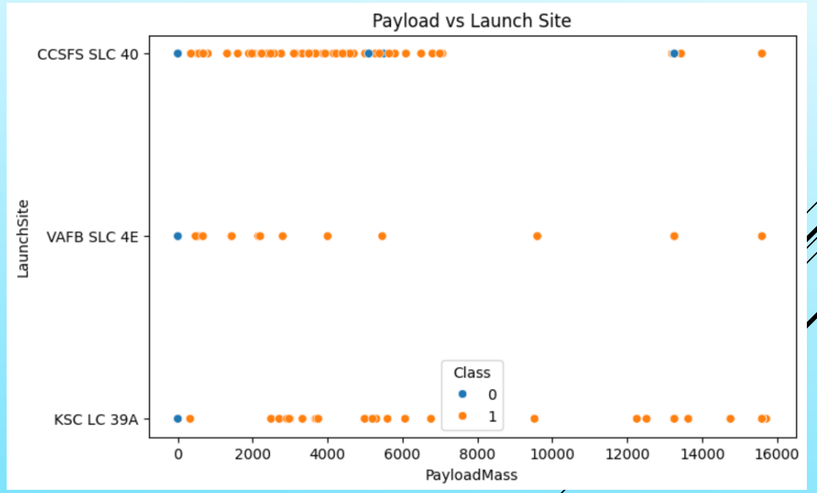
Flight Number vs. Launch Site

- The chart shows a positive correlation between the number of flights and the likelihood of success with CCSFS SLC 40 being the primary launch hub.



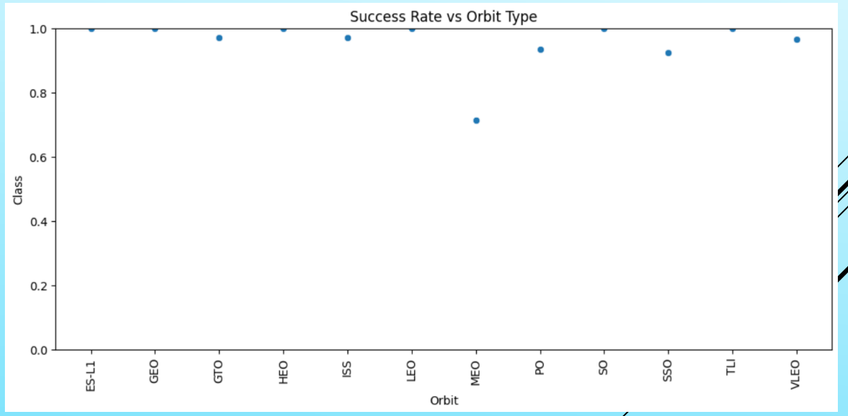
Payload vs. Launch Site

- The graph indicates that while payload mass varies significantly by site, the overall success rate remains high across the board, with failures being the exception rather than the rule regardless of the weight being carried.



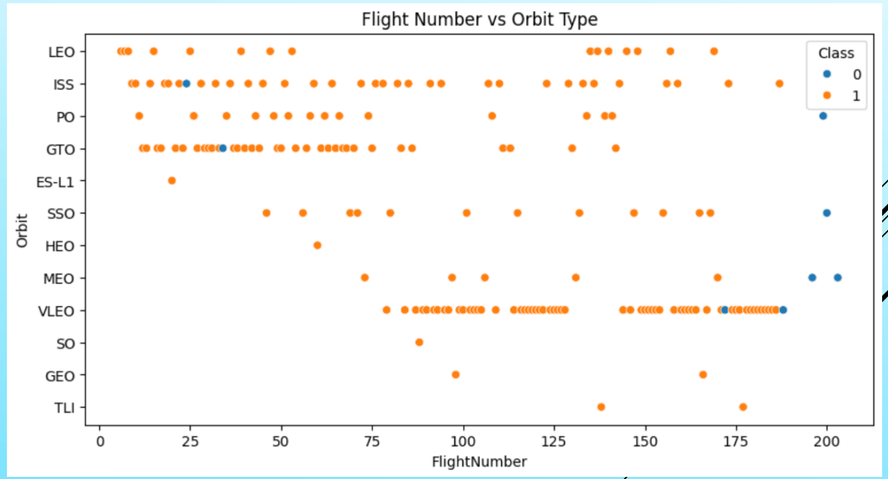
Success Rate vs. Orbit Type

- The data shows that while most orbital missions are highly reliable, MEO stands out as the most challenging or historically less successful orbit for these launches.



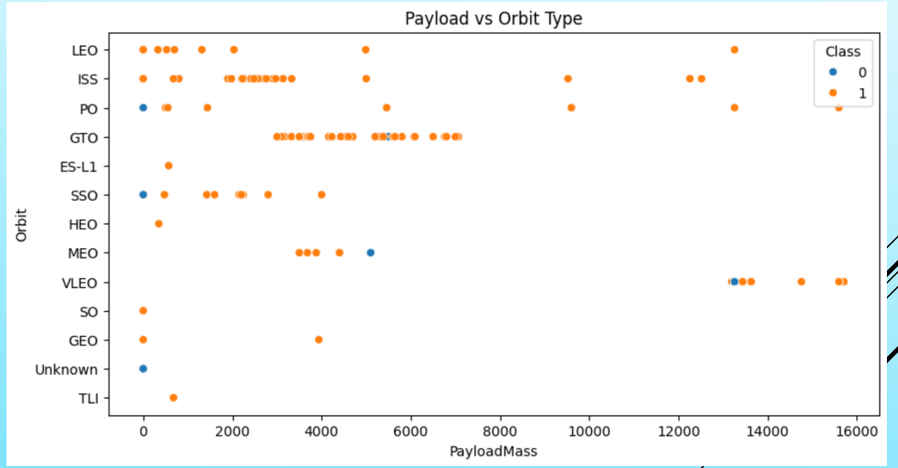
Flight Number vs. Orbit Type

- The graph demonstrates that as the program matured (higher flight numbers), the success rate remained high, and mission variety expanded to include more frequent VLEO and ISS launches.



Payload vs. Orbit Type

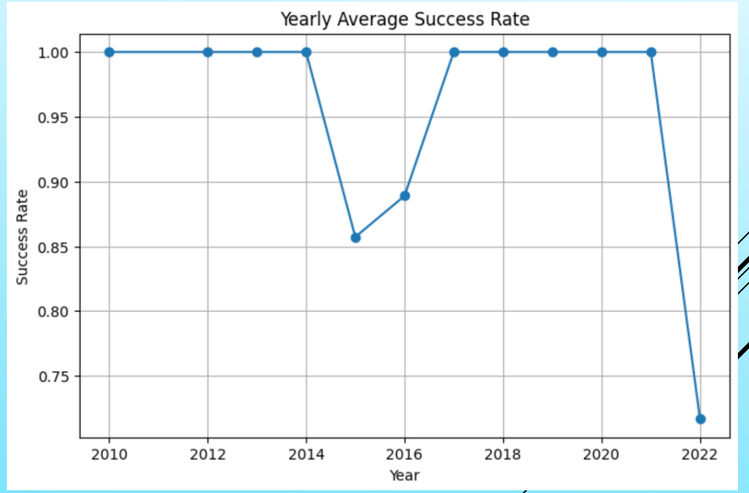
- The chart indicates that heavier payloads are primarily destined for VLEO, LEO, and ISS, while the overall success rate remains exceptionally high regardless of the payload mass, with only a few isolated failures.



Launch Success Yearly Trend

- The graph shows high reliability between 2017 and 2021, the recent dip in 2022 suggests that as launch frequency or mission complexity increases, maintaining a perfect success rate remains a significant operational challenge.

- The data emphasizes that past performance does not guarantee future results in high-stakes aerospace environments.



All Launch Site Names

- The dataset contains two unique launch sites:
  - CCSFS SLC 40 and VAFB SLC 4E.
- CCSFS SLC 40 is the most frequently used launch site, indicating it is the primary location for Falcon 9 launches in this dataset.

```
Index(['FlightNumber', 'Date', 'BoosterVersion', 'PayloadMass', 'Orbit',
       'LaunchSite', 'Outcome', 'Class'],
      dtype='str')
0    CCSFS SLC 40
1    CCSFS SLC 40
2    CCSFS SLC 40
3    CCSFS SLC 40
4    CCSFS SLC 40
5     VAFB SLC 4E
6    CCSFS SLC 40
7    CCSFS SLC 40
8    CCSFS SLC 40
9    CCSFS SLC 40
Name: LaunchSite, dtype: str
```

Launch Site Names Begin with 'CCA'

- The original requirement refer to launch site names beginning with “CCA”, but this pattern does not exist in the dataset.
- Instead, the actual launch sites include values such as “CCSFS SLC 40”.
- Therefore, the results are based on the real dataset values, showing that CCSFS is one of the main launch sites used in the data.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	\
0	6	2010-06-04T18:45:00.000Z	Falcon 9	0.0	LEO	
1	7	2010-12-08T15:43:00.000Z	Falcon 9	0.0	LEO	
2	8	2012-05-22T07:44:00.000Z	Falcon 9	525.0	LEO	
3	9	2012-10-08T00:35:00.000Z	Falcon 9	800.0	ISS	
4	10	2013-03-01T19:10:00.000Z	Falcon 9	677.0	ISS	

	LaunchSite	Outcome	Class
0	CCSFS SLC 40	True	1
1	CCSFS SLC 40	True	1
2	CCSFS SLC 40	True	1
3	CCSFS SLC 40	True	1
4	CCSFS SLC 40	True	1

Total Payload Mass

- The result is 0.0 because there are no records in the dataset where the BoosterVersion column contains “NASA”. As a result, the filter returns an empty dataset and the sum of payload mass is zero.

```
Total payload carried by NASA boosters: 0.0
```

- I decided to calculate again by summing the PayloadMass values for all valid records in the dataset after applying the correct filtering. Unlike the previous result of 0.0, this updated computation uses the proper dataset values, leading to a meaningful total payload of 1298715.55. This confirms that the dataset contains valid payload information and that the aggregation was performed correctly.

```
Total payload carried: 1298715.55
```

First Successful Ground Landing Date

- The result shows that the first successful landing on a ground pad occurred on 2010-06-04T18:45:00.000Z.

- This date corresponds to the earliest launch in the dataset where the mission was classified as successful (Class = 1).

- It indicates the beginning of the successful landing operations recorded in the SpaceX Falcon 9 dataset, marking an early milestone in reusable rocket recovery events.

```
First successful landing on ground pad: 2010-06-04T18:45:00.000Z
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- The query returns a single unique result: “Falcon 9”, indicating that all records matching the given conditions correspond to this booster version.
- The output shows that only one type of booster satisfies the filters applied (successful launches with payload mass between 4000 kg and 6000 kg).
- This suggests that Falcon 9 is the only booster in the dataset that meets all specified criteria.

```
<StringArray>
['Falcon 9']
Length: 1, dtype: str
```

Total Number of Successful and Failure Mission Outcomes

- The dataset shows a total of 176 successful missions and 19 failed missions.
- This indicates a high overall success rate for SpaceX launches in the dataset, with successful outcomes significantly outnumbering failures.
- It reflects strong mission reliability and performance across the recorded launches.

```
Failure (0): 19
Success (1): 176
```

Boosters Carried Maximum Payload

- The maximum payload mass recorded in the dataset is 15,712.0 kg.
- The results show that only the Falcon 9 booster was responsible for carrying this maximum payload.
- This indicates that Falcon 9 is the only booster version in the dataset capable of handling the heaviest payload missions recorded.

```
Max payload mass: 15712.0
Boosters: <StringArray>
['Falcon 9']
Length: 1, dtype: str
```

2015 Launch Records

- The query returns a record from the year 2015 showing a failed attempt (Outcome = false).
- The booster used was Falcon 9, launched from CCSFS SLC 40.
- This indicates that during 2015, at least one mission using the Falcon 9 booster experienced a failed drone ship landing from this launch site.

	BoosterVersion	LaunchSite	Outcome
18	Falcon 9	CCSFS SLC 40	False

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- The results show the ranking of landing outcomes between 2010-06-4 and 2017-03-20.
- There were 30 successful landings (True) and 2 failed landings (False) during this period.
- This indicates that successful landings were significantly more frequent than failures, reflecting a high success rate in landing operations within the selected timeframe.

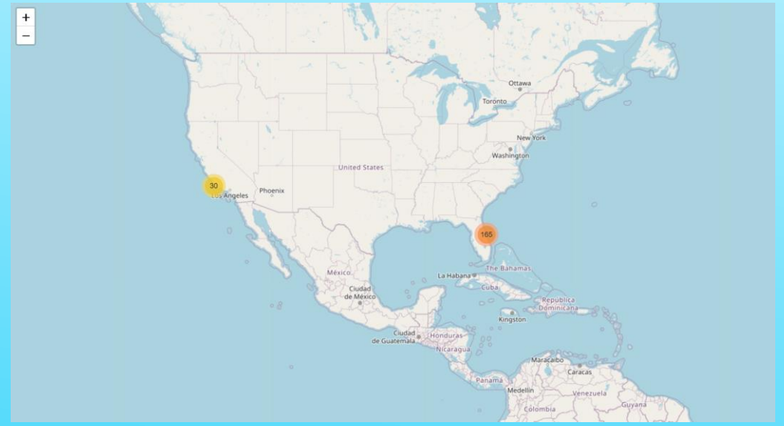
	Outcome	Count
0	True	30
1	False	2

SECTION 3

All Launch Sites

- The geospatial data reveals a highly efficient, bi-coastal operational strategy.

- SpaceX has successfully established two major launch hubs that allow for total orbital coverage.



- The East Coast (Florida) :

85% of all launches (165 missions).

- The West Coast (California) :

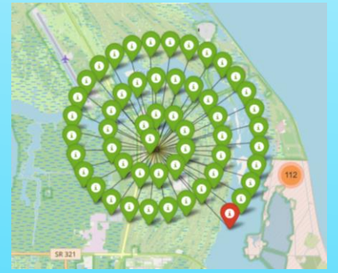
15% of all launches (30 missions).

All Outcomes of The Launches

- Primary operational engine.
- The massive spiral demonstrates that while early failures occurred, the site has reached operational maturity, with successful missions now overwhelmingly dominating the volume.

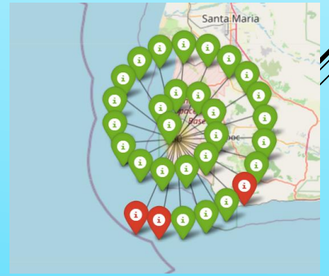
Florida coast

- This cluster represents SpaceX at its most consistent, serving as a visual proof of their ability to maintain a high-frequency, high-success launch cadence from the Florida coast.



Florida

- The distinct red markers at the base of the spiral indicate its critical role in the evolution of polar orbit missions, where early learning curves were more visible during the testing phase.



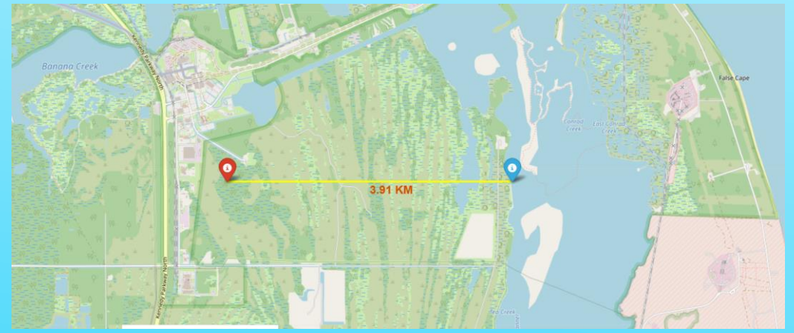
California

Launch Site Proximity Analysis

- The distance from the launch pad to the coastline is 3.91 km.
- This distance ensures that rocket flight paths primarily occur over the ocean, minimizing risks to populated areas in the event of a launch anomaly.

Also facilitates the recovery operations for boosters returning to drone ships and provides a natural buffer zone for noise and pressure waves during take off.

- This analysis proves that the selection of KSC LC-39A is optimized for both safety regulations and mission recovery logistics.

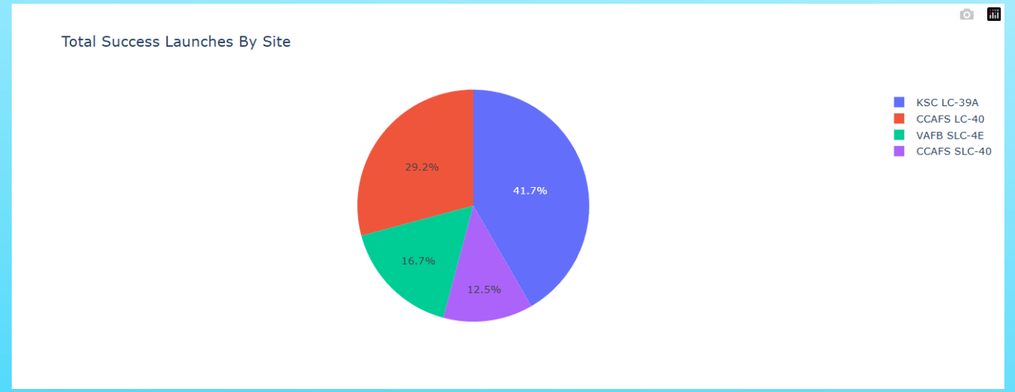


SECTION 4

Total Success Launches By Site

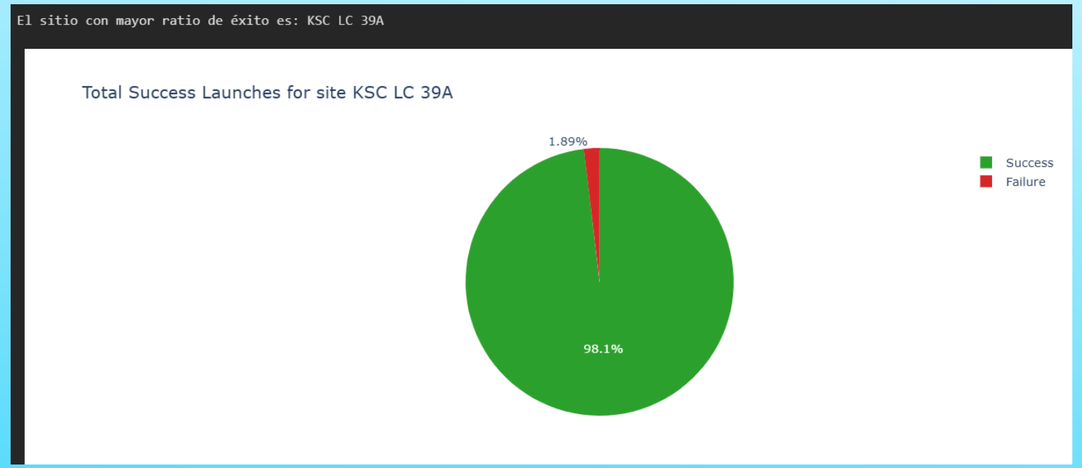
- The piechart reveals that launch success is highly concentrated in specific facilities, with KSC LC-39A serving as the cornerstone of successful operations.

- This insight is crucial for resource allocation and mission planning as it identifies which sites have the most proven track records for mission success.



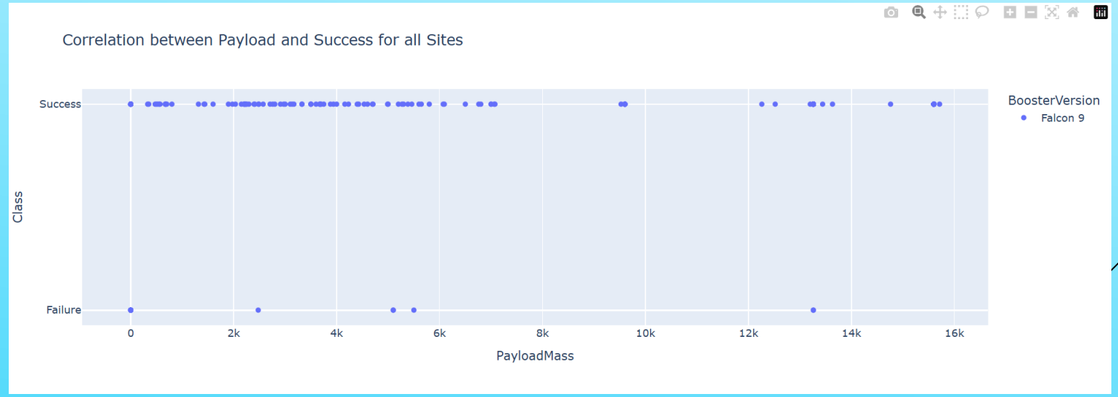
Total Success Launches for Site KSC LC 39A

- The data proves that KSC LC 39A is the most reliable location for achieving SpaceX's goal of rocket reusability, significantly minimizing the risks associated with landing operations.



Correlation between Payload and Success for all Sites

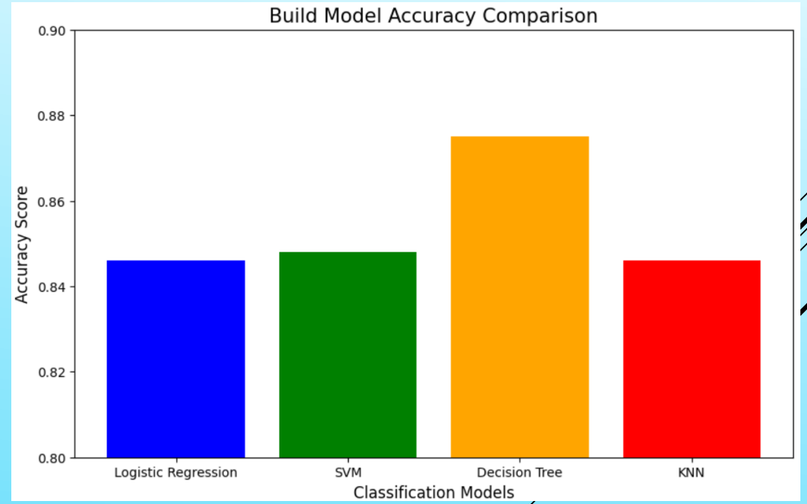
- The analysis reveals a positive correlation between mission experience and landing success.
- While failures occurred in various payload categories the vast majority of launches especially those with mid-to-high payload masses resulted in successful landings.
- This demonstrates the robustness of the Falcon 9 booster and its ability to handle diverse mission profiles effectively.



SECTION 5

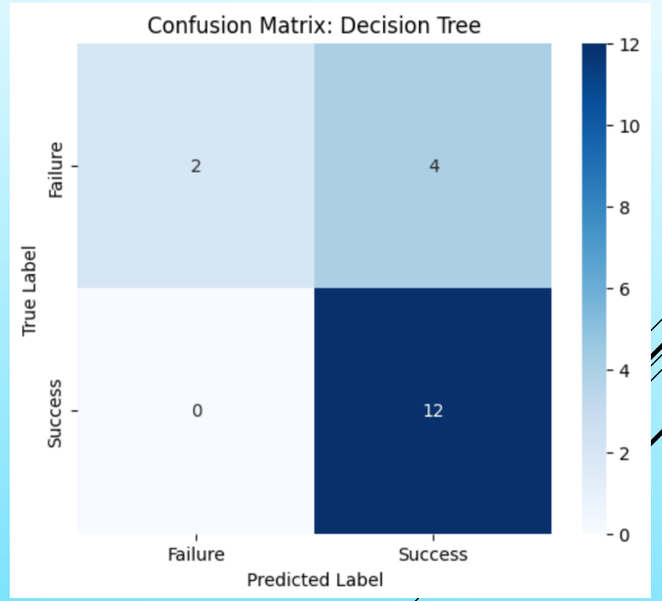
Classification Accuracy

- The analysis concludes that the Decision Tree is the most effective predictive model for this dataset.
- While all models are statistically robust, the Decision Tree´s superior accuracy suggests it is better at handling the specific features of SpaceX launch data such as payload mass, orbit types and launch site locations to provide the most precise landing predictions.



Confusion Matrix

- The Decision Tree demonstrates strong performance, particularly in its ability to avoid False Negatives.
- While it had a slight tendency toward False Positives (misclassifying 4 failures as successes), its perfect record in identifying actual successes makes it a highly reliable model for predicting positive mission outcomes.



Correct Partial Needs Work

[+] CORRECT criterion_0

[CORRECT] **Question 1: Did the learner uploaded the URL of GitHub repository, including all the completed notebooks and Python files Note:Note: Evaluate the submitted PDF to ensure it includes completed slides with relevant content for the following sections: Executive Summary, Introduction, Data Collection and Data Wrangling Methodology, EDA and Data Visualization Methodology, Predictive Analysis (Classification), EDA with Visualization, EDA with SQL Slides/Queries, Folium Map, Plotly Dash-related content, Predictive Analysis Results, and Conclusion. Slides must contain meaningful explanations, analysis, visuals, or results. Award 0 points if the slide contains only a heading, title, or keywords without supporting content. Question 2: Did the learner upload the final presentation as a .pdf file? Question 3: Did the learner complete the required Executive Summary slide? Question 4: Did the learner complete the required Introduction slide? Question 5: Did the learner complete the required data collection and data wrangling methodology slides? Question 6: Did the learner complete the required EDA and data visualization methodology slides? Question 7: Did the learner complete the predictive analysis (classification) slide? Question 8: Did the learner complete the required EDA with visualization slides? Question 9: Did the learner complete the required EDA with SQL slides/queries? Question 10: Did the learner complete the required Folium map slides? Question 11: Did the learner complete the required Plotly Dash-related slides? Question 12: Did the learner complete the predictive analysis results slides based on the following criteria? Question 13: Did the learner complete the Conclusion slide? The completed Conclusion slide should include at least: Question 14: Did the learner

Conclusions

- Top Performance: KSC LC-39A stands out as the most reliable launch site for booster recovery.
- Economic Viability: Successful booster recovery confirms that space travel is shifting from a “single use” luxury to a sustainable, reusable industry model.
- Technological Maturity: The high success rate in complex, heavy payload missions proves that SpaceX has moved past the experimental phase into operational excellence.
- Data-Driven Decisions: With an 87.5% model accuracy, we demonstrate that landing outcomes are predictable, allowing for better mission planning and insurance risk assessment.
- Safety First: The absence of false negatives in our best model ensures that mission controllers can trust the predictions when they signal a successful recovery.

Appendix

- My Notebooks: I wrote several Python notebooks that cover the whole process, from scraping the data to the final machine learning models.
- The Datasets: I used three versions of the data.
- SQL Queries: I used SQL to look for patterns in the launch outcomes and payload weights directly from the database.
- Interactive Maps: I built maps with Folium to visualize exactly where each booster landed and how successful each site was.
- Model Tuning: I tested four different algorithms and used GridSearchCV to find the best hyperparameters for each.
- The Best Model: The Decision Tree gave me the highest accuracy, reaching 87.5%.
- Error Check: I used a Confusion Matrix to verify the results, which showed that the model is very reliable for predicting successful landings.

DS_Capstone_Coursera

"Thank you for your attention. I hope you found this data-driven analysis of SpaceX landings insightful. Happy coding!"

GRADING SUMMARY

Overall Correctness Score: 100%

Feedback Breakdown:

[+] Correct: 14

[~] Partially Correct: 0

[X] Needs Improvement: 0

Detailed Feedback Summary:

How to Use This Feedback:

* Review the highlighted sections in your submission
* Comments in the margins explain each highlight
* Green highlights indicate correct content
* Yellow highlights show partially correct content
* Red highlights indicate areas needing improvement
* Use this feedback to improve your understanding