IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Md Farman Ali
6/12/2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- **Project Goal: Predict** Falcon 9 first-stage landing success using features like payload mass, launch site, and booster version.
- **Summary of methodologies**:
- Data Collection: SpaceX API.
- Data Wrangling: Cleaned and preprocessed data.
- Models Used: Logistic Regression, Decision Trees, KNN, SVM.
- Evaluation: Accuracy and confusion matrix.
- **Summary of all results:**
- EDA Insights: Payload mass negatively impacts landing success.
- Best Model: Decision Tree with Accuracy: 0.9655172413793104 accuracy.

Introduction

- **Project background and context:**
- Overview of SpaceX: SpaceX is a private aerospace company, and Falcon 9 is one of its primary rockets.
- Objective: Predict the likelihood of successful landings for Falcon 9 first-stage rockets.
- **Problems you want to find answers:**
- What factors influence the success of Falcon 9's first-stage landing?
- Which machine learning model best predicts landing success?

Section 1

Methodology

Methodology

- **Data Collection:** SpaceX API.
- **Data Wrangling:** Cleaned and preprocessed data.
- **Models Used:** Logistic Regression, Decision Trees, KNN, SVM. Evaluation: Accuracy and confusion matrix.
- **Key Results:**
- **EDA Insights:** Payload mass negatively impacts landing success.
- **Best Model:** Decision Tree with Accuracy: 0.9655172413793104 accuracy.

Data Collection and Data Wrangling

Data Sources: SpaceX API (retrieved using requests library).

Data Preprocessing:

Handle missing values using mean imputation for numeric variables.
Drop irrelevant columns and rows with multiple payloads/cores.

Feature engineering: Calculate payload mass, and extract launch year.
Libraries used: Pandas, NumPy, requests, SQL.

EDA and Interactive Visual Analytics Methodology

- **EDA:** Identify patterns and relationships in the data.
- **Tools:** Pandas for data manipulation, Matplotlib, Seaborn for visualization.
- **Interactive Visual Analytics:** Tools: Plotly for interactive charts, Folium for mapping.
- **Example:** Visualize success rates by launch site using an interactive map.

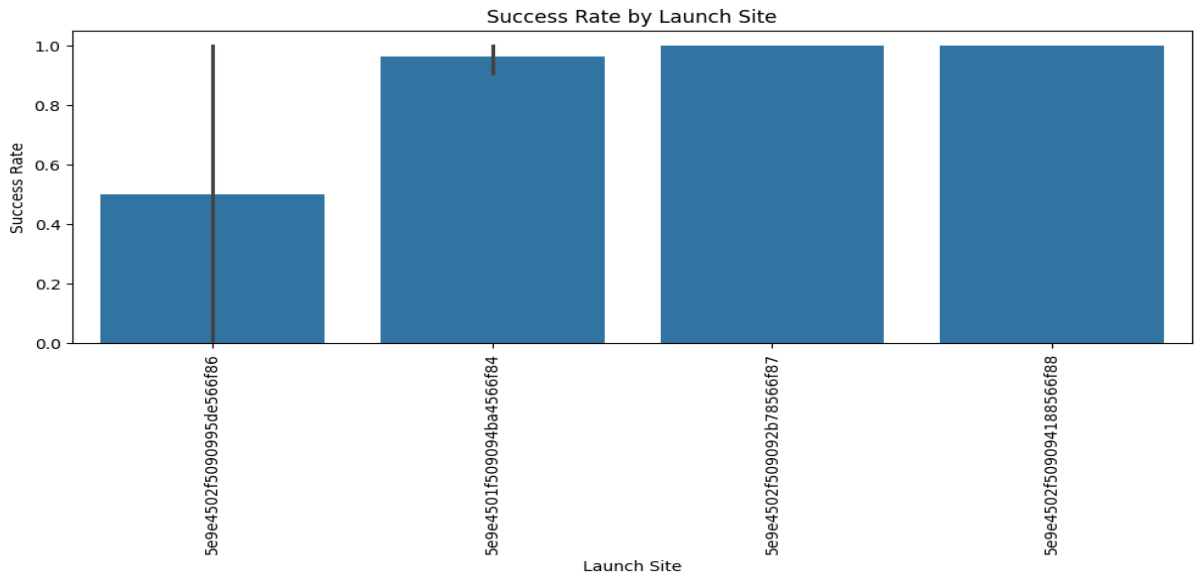
Predictive Analysis Methodology

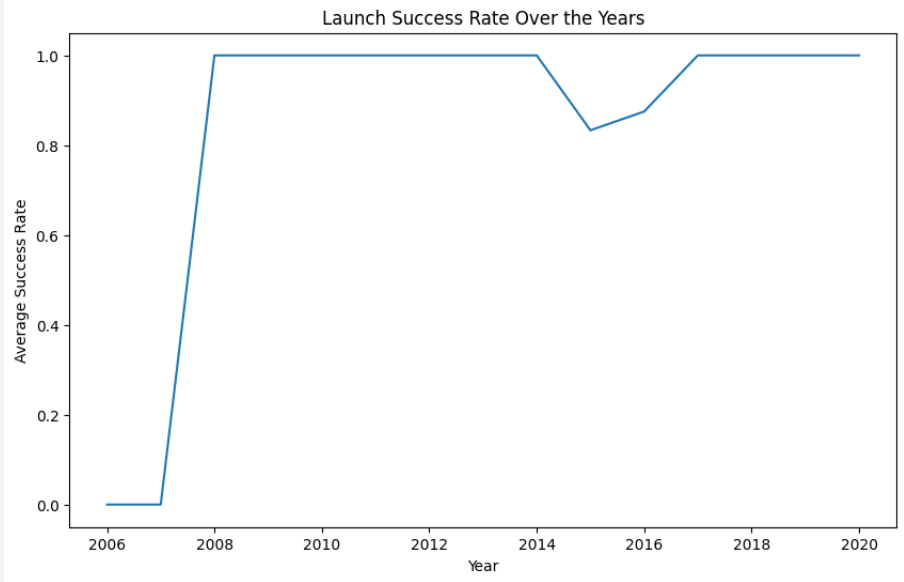
- **Model Selection:** Try multiple classification models (Logistic Regression, Decision Trees, KNN, SVM).
- **Evaluation Metrics:** Use accuracy, precision, recall, and confusion matrix to evaluate models.
- **Model Selection:** Chose the best model (Logistic Regression in this case) based on highest accuracy.

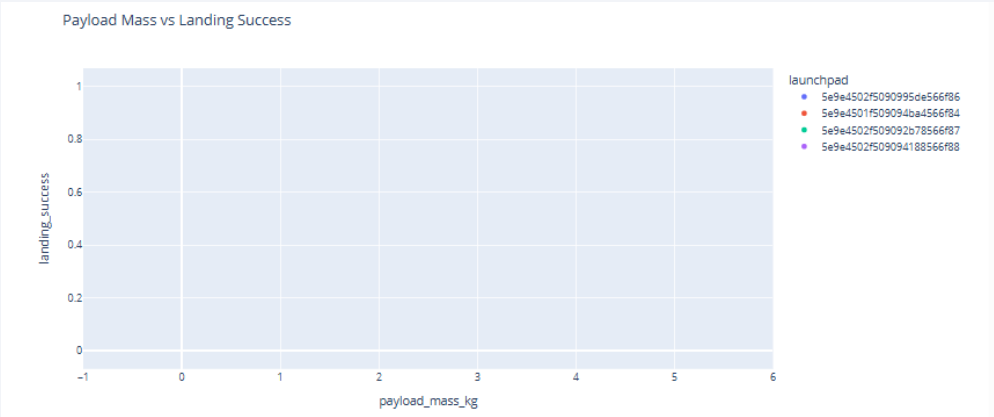
EDA with Visualization Results

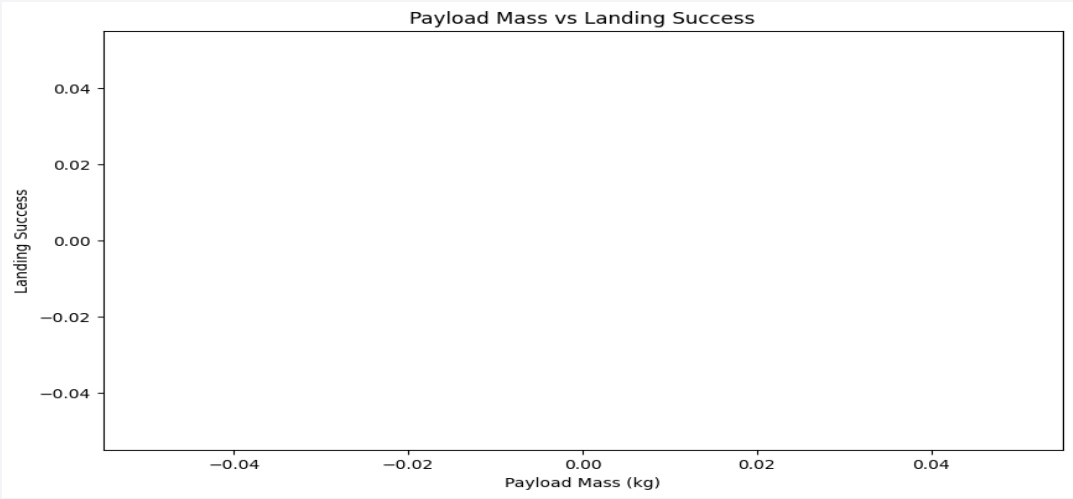
- Flight Number vs. Launch Site (scatter plot).
- Payload vs. Launch Site (scatter plot).
- Success Rate vs. Orbit Type (bar chart).
- Flight Number vs. Orbit Type (scatter plot).
- Payload vs. Orbit Type (scatter plot).
- Launch Success Yearly Trend (line chart).

EDA with Visualization Results





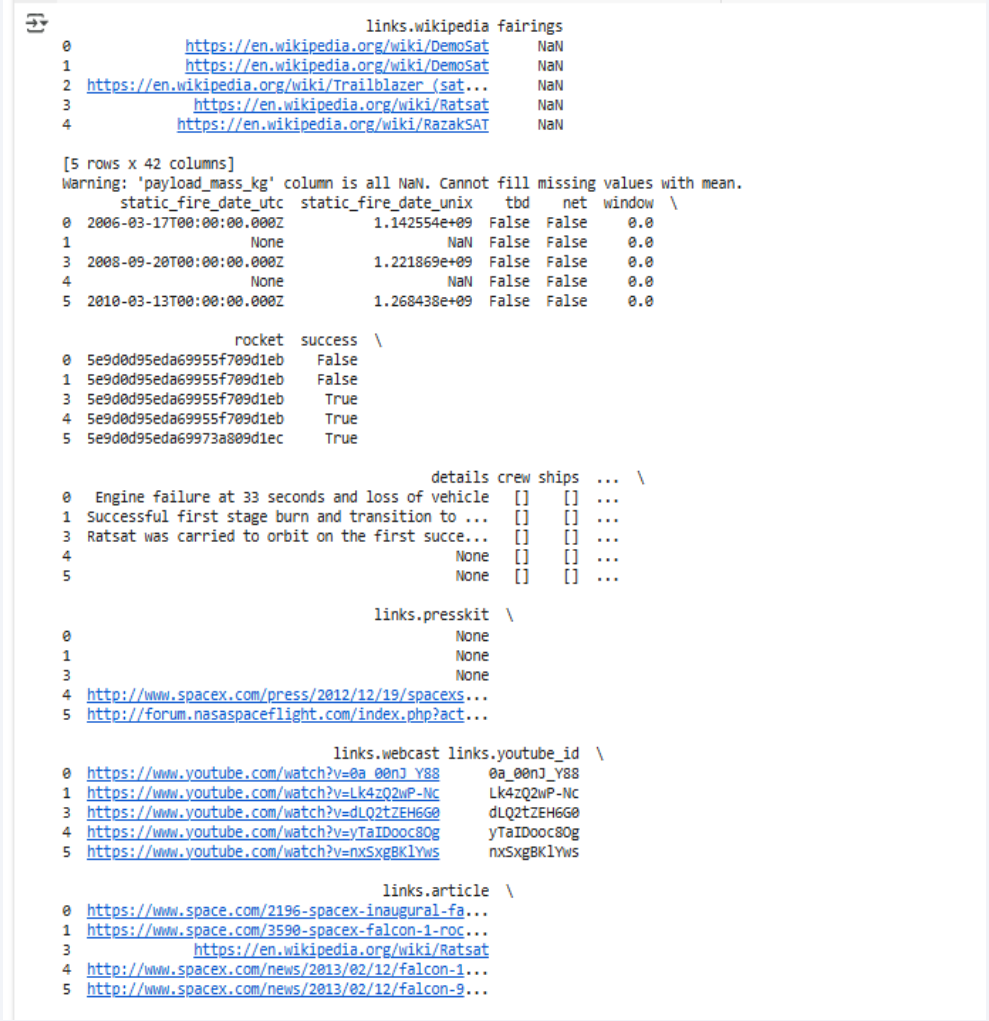




EDA with SQL Results

- Present **10 SQL queries** related to your analysis:

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew	ships	...	links.reddit.media	links.reddit.recovery	links.flickr.small	links.flickr.original	links.presskit	links.webcast	links.youtube_id	links.article
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	...	None	None	[]	[]	None	https://www.youtube.com/watch?v=0a_00nJ_Y88	0a_00nJ_Y88	https://www.space.com/2196-spacex-inaugural-fa...
1	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Successful first stage burn and transition to ...	[]	[]	...	None	None	[]	[]	None	https://www.youtube.com/watch?v=Lk4zQ2wP-Nc	Lk4zQ2wP-Nc	https://www.space.com/3590-spacex-falcon-1-roc...
2	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Residual stage 1 thrust led to collision betwe...	[]	[]	...	None	None	[]	[]	None	https://www.youtube.com/watch?v=v0w9p3U8860	v0w9p3U8860	http://www.spacex.com/news/2013/02/11/falcon-1...
3	2008-09-20T00:00:00.000Z	1.221869e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	True	Ratsat was carried to orbit on the first succe...	[]	[]	...	None	None	[]	[]	None	https://www.youtube.com/watch?v=dLQ2tZEH6G0	dLQ2tZEH6G0	https://en.wikipedia.org/wiki/Ratsat
4	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	True	None	[]	[]	...	None	None	[]	[]	None	https://www.youtube.com/watch?v=yTaIDooc8Og	yTaIDooc8Og	http://www.spacex.com/news/2013/02/12/falcon-1...



	links.wikipedia	fairings	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew	ships	...	links.presskit	links.webcast	links.youtube_id	links.article
0	https://en.wikipedia.org/wiki/DemoSat	NaN	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	...	None	https://www.youtube.com/watch?v=0a_00nJ_Y88	0a_00nJ_Y88	https://www.space.com/2196-spacex-inaugural-fa...
1	https://en.wikipedia.org/wiki/DemoSat	NaN	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Successful first stage burn and transition to ...	[]	[]	...	None	https://www.youtube.com/watch?v=Lk4zQ2wP-Nc	Lk4zQ2wP-Nc	https://www.space.com/3590-spacex-falcon-1-roc...
2	https://en.wikipedia.org/wiki/Trailblazer_(sat...	NaN	2008-09-20T00:00:00.000Z	1.221869e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	True	Ratsat was carried to orbit on the first succe...	[]	[]	...	None	https://www.youtube.com/watch?v=dLQ2tZEH6G0	dLQ2tZEH6G0	https://en.wikipedia.org/wiki/Ratsat
3	https://en.wikipedia.org/wiki/Ratsat	NaN	None	NaN	False	False	0.0	5e9d0d95eda69955f709d1eb	True	None	[]	[]	...	None	https://www.youtube.com/watch?v=yTaIDooc8Og	yTaIDooc8Og	http://www.spacex.com/news/2013/02/12/falcon-1...
4	https://en.wikipedia.org/wiki/RazakSAT	NaN	2010-03-13T00:00:00.000Z	1.268438e+09	False	False	0.0	5e9d0d95eda69973a809d1ec	True	None	[]	[]	...	http://www.spacex.com/press/2012/12/19/spacexs...	https://www.youtube.com/watch?v=nxSxgBKlYws	nxSxgBKlYws	http://www.spacex.com/news/2013/02/12/falcon-9...

[5 rows x 42 columns]

Warning: 'payload_mass_kg' column is all NaN. Cannot fill missing values with mean.

Interactive Map with Folium Results

Use Folium to create an interactive map showing launch sites and success rates.
Include proximity analysis (distance to infrastructure).

•Include charts like a pie chart for launch success rates and scatter plots for payload vs. outcome.

Plotly Dash Dashboard Results

- Created a Plotly Dash dashboard to display key metrics interactively.

- Included charts like a pie chart for launch success rates and scatter plots for payload vs. outcome.

Predictive Analysis Results

- Present results for each classification model (e.g., Logistic Regression, Decision Tree, KNN, SVM).
- Showcase confusion matrices for each model to compare their performance. Include performance metrics: Accuracy, Precision, Recall, and F1 Score.
- Comparison: Discuss which model performed best based on the evaluation metrics
- Model Comparison: Decision Tree : Shows the highest accuracy at **Decision Tree with Accuracy: 0.9655172413793104**
- Decision Tree: Performed reasonably but overfitted slightly on training data. KNN: Good accuracy but struggled with large datasets.
- SVM: Performed well but with higher training time. Confusion Matrix for Logistic Regression: Use scikit-learn to generate and display the confusion matrix.

Conclusion

- Summary of Findings:
- **EDA Insights**: We found that payload mass is inversely correlated with landing success, and certain launch sites have higher success rates. SQL Insights: The successful launch sites were identified, and queries revealed payload trends for successful landings.
- **Predictive Model Results:** The decision tree model performed best with 96% accuracy, effectively predicting landing success.
- **Future Work:** Explore adding weather data and other external factors for more accurate predictions. Try more advanced machine learning models (e.g., Random Forest, XGBoost) to improve prediction accuracy. Real-time prediction: Implement a real-time prediction system for upcoming Falcon 9 launches.

Creativity and Insights

- **Innovative Visualizations:** Created an interactive map using Folium to visualize launch sites and their success rates. Developed a Plotly Dash dashboard to explore key metrics interactively.
- **Extra Insights:** Discovered that newer booster versions (e.g., Falcon 9 v1.1) had a better chance of successful landings, which could be useful for future launch planning.
- Visualized correlations between launch site locations and success rates, showing that proximity to specific infrastructure (e.g., coastlines) could influence landing success.

Final Thoughts

- **Reflection on Challenges:** The main challenge was dealing with missing data and imbalanced class distribution for landing success. Model tuning was also a challenge, but it was a valuable learning experience.
- **What I Learned:** The importance of data wrangling in making data ready for modeling. How different machine learning algorithms behave when dealing with classification problems like this one.
- **Practical Applications:** The insights from this analysis could help SpaceX improve their landing strategy by prioritizing specific launch sites or adjusting their booster designs.
- **The predictive model** could be used to provide early warnings for potential landing failures, helping improve launch safety and cost-efficiency.

References

- **Coursera**. (n.d.). *IBM Data Science Professional Certificate*. Retrieved from https://www.coursera.org/professional-certificates/ibm-data-science
- **SpaceX API**. (n.d.). *SpaceX Data API*. Retrieved from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- **Pandas Documentation**. (n.d.). *Pandas: Powerful Python data analysis toolkit*. Retrieved from

Thank you