IBM Data Science Capstone

Neil Kondapalli

August 2024

OUTLINE

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

EXECUTIVE SUMMARY

- Objective:
  - Develop a predictive model to forecast the success of Falcon 9 first stage landings
- Key Steps:
  - Collected and cleaned historical launch data
  - Conducted exploratory data analysis (EDA) to identify key patterns
  - Developed and evaluated machine learning models to predict landing success
- Outcome:
  - Created a model with strong predictive performance
  - Identified significant factors like payload weight and launch site as critical to landing success

INTRODUCTION

- Background
  - SpaceX's Falcon 9 program is a key player in space exploration, known for its innovative approach to reusing rocket stages
  - Successful landings of the Falcon 9 first stage are crucial for reducing costs and enabling the reusability of rockets
- Problem Statement
  - Accurately predicting the success of rocket landings is vital for improving reusability and operational efficiency
- Objective
  - The goal of this project is to develop a machine learning model to predict the success of Falcon 9 first stage landings based on historical launch data

METHODOLOGY

- Data Collection
  - Data was collected from SpaceX's API and supplemented with additional datasets obtained through web scraping
  - The data includes information on launch dates, payloads, launch sites, rocket versions, orbits, and weather conditions
- Data Wrangling/Cleaning
  - Handled missing values, removed duplicates, and corrected inconsistencies in the dataset
  - Normalized and standardized the data to ensure consistency across features
- Feature Engineering
  - Created new features such as payload mass category and launch site success rate
  - Selected important features for modeling based on EDA insights
- Modeling Approach
  - Tested several machine learning models, including logistic regression, decision trees, and random forests
  - Applied cross-validation and hyperparameter tuning to improve model accuracy and reliability

METHODOLOGY

- Data Collection
  - Data was collected from SpaceX's API and supplemented with additional datasets obtained through web scraping
  - The data includes information on launch dates, payloads, launch sites, rocket versions, orbits, and weather conditions
- Data Wrangling/Cleaning
  - Handled missing values, removed duplicates, and corrected inconsistencies in the dataset
  - Normalized and standardized the data to ensure consistency across features
- Feature Engineering
  - Created new features such as payload mass category and launch site success rate
  - Selected important features for modeling based on EDA insights
- Modeling Approach
  - Tested several machine learning models, including logistic regression, decision trees, and random forests
  - Applied cross-validation and hyperparameter tuning to improve model accuracy and reliability

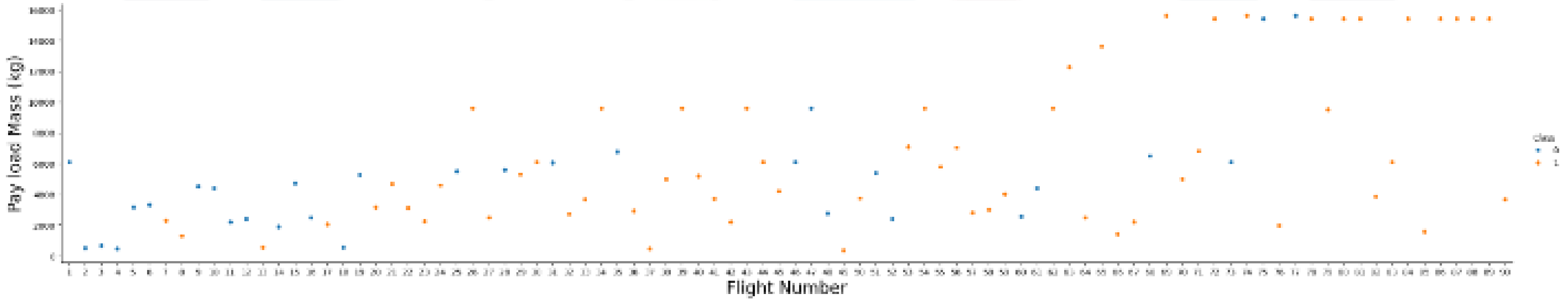
RESULTS – EDA SQL

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

Using SQL, a greater understanding of the data was achieved.

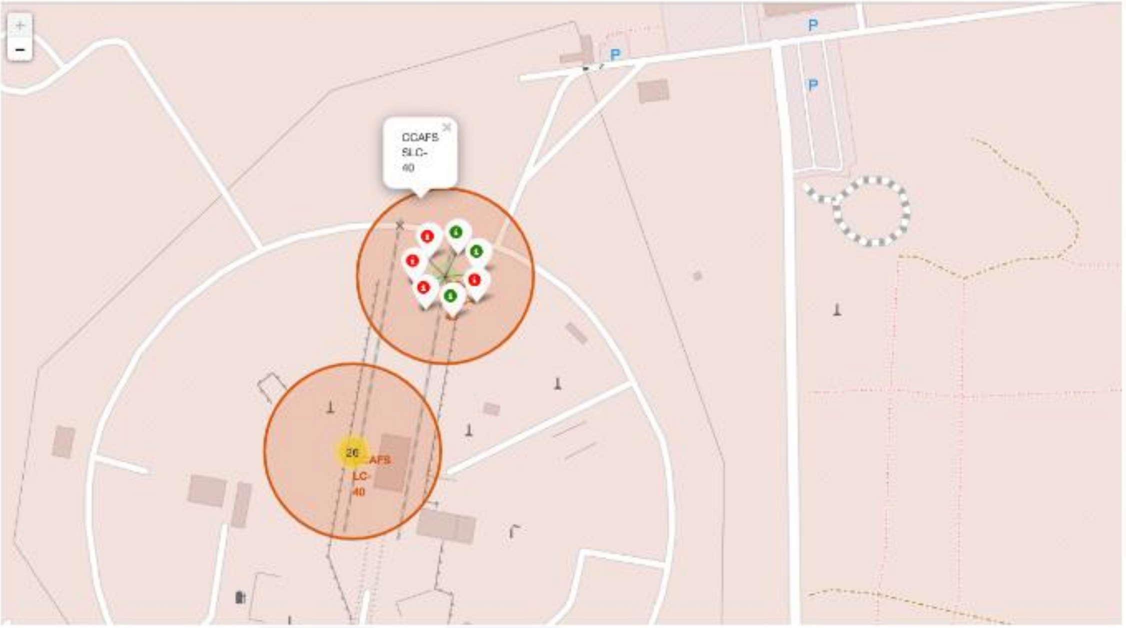
The graphic shown is each of the possible landing outcomes with the amount of times each outcome was encountered. This data is critical when attempting to understand the problem and work to build a solution.

RESULTS – EDA Visualization



The image above shows each flight by its flight number along with the corresponding mass. In addition, it shows the "class" of each flight – a success or a fail.

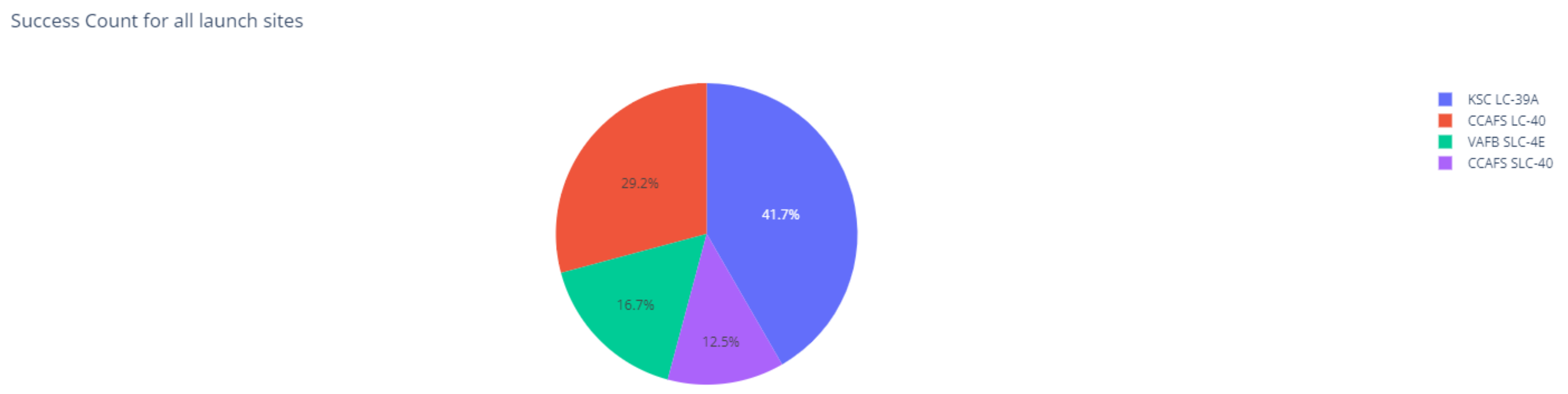
RESULTS – Folium



Launch Site	Proximity	LatLaunchSite	LongLaunchSite	Lat	Long	distance
CCAFS LC-40	coastline	28.562302	-80.577356	28.56340	-80.56796	0.926054
	railway	28.562302	-80.577356	28.57217	-80.58528	1.343093
	highway	28.562302	-80.577356	28.56301	-80.57076	0.649222
	city	28.562302	-80.577356	28.10469	-80.64784	51.365738
CCAFS SLC-40	coastline	28.563197	-80.576820	28.56409	-80.56806	0.861525
	railway	28.563197	-80.576820	28.57217	-80.58528	1.295798
	highway	28.563197	-80.576820	28.56385	-80.57088	0.584818
	city	28.563197	-80.576820	28.10469	-80.64784	51.471476
KSC LC-39A	coastline	28.573255	-80.646895	28.60283	-80.58815	6.613767
	railway	28.573255	-80.646895	28.57314	-80.65398	0.692172
	highway	28.573255	-80.646895	28.57307	-80.65553	0.843713
	city	28.573255	-80.646895	28.10469	-80.64784	52.118441
VAFB SLC-4E	coastline	34.632834	-120.610745	34.63470	-120.62531	1.349004
	railway	34.632834	-120.610745	34.63585	-120.62401	1.259455
	highway	34.632834	-120.610745	34.70480	-120.56940	8.853368
	city	34.632834	-120.610745	34.64253	-120.47331	12.623666

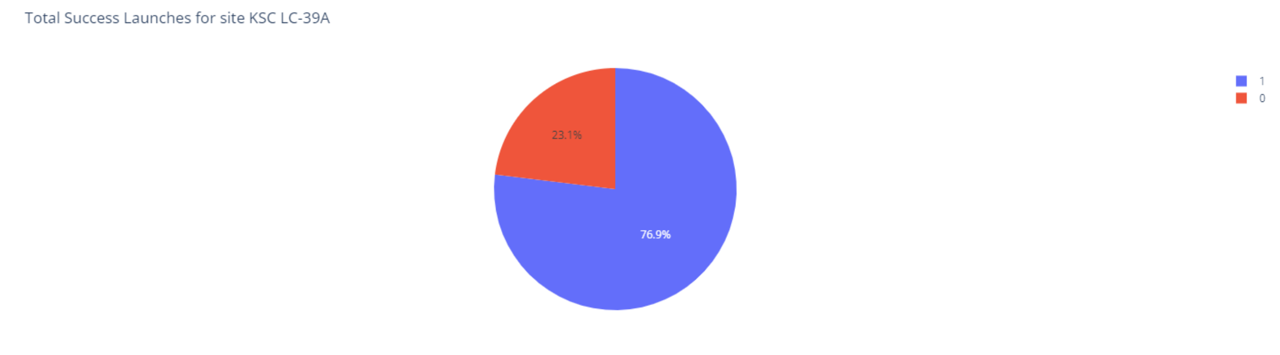
The image above shows one of the launch sites marked on a map. Within that site, it displayed successful flights in green and fails in red. To the right is a generated list of the sites and their distances to landmarks.

RESULTS – Dashboard



From the Plotly Dashboard, the pie chart of success for ALL launch sites is shown above, with each site given its own color.

RESULTS – Dashboard

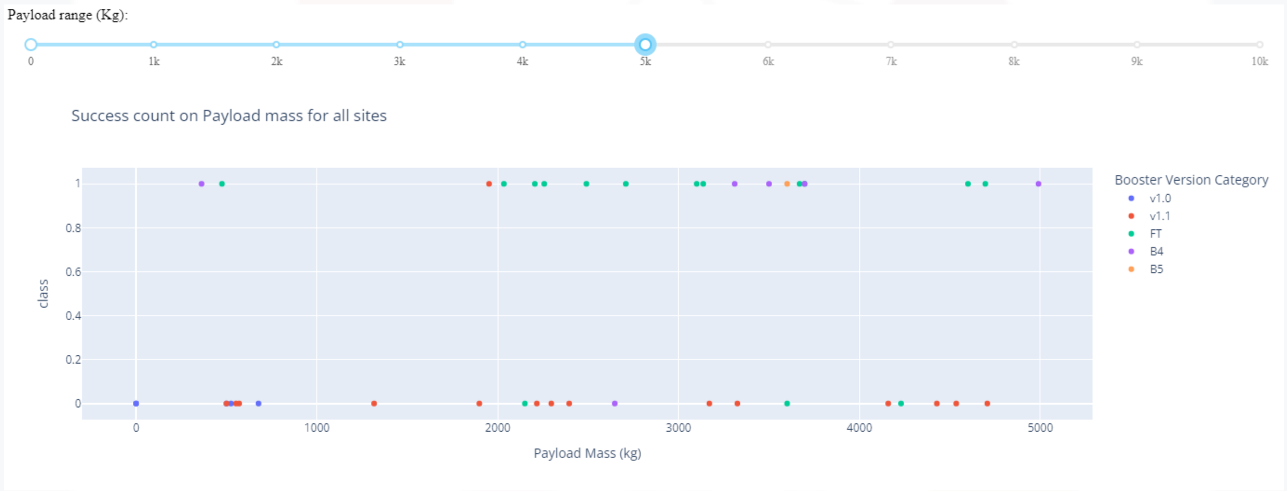


Using the dropdown, you can select a single site to examine individual results.

RESULTS – Dashboard

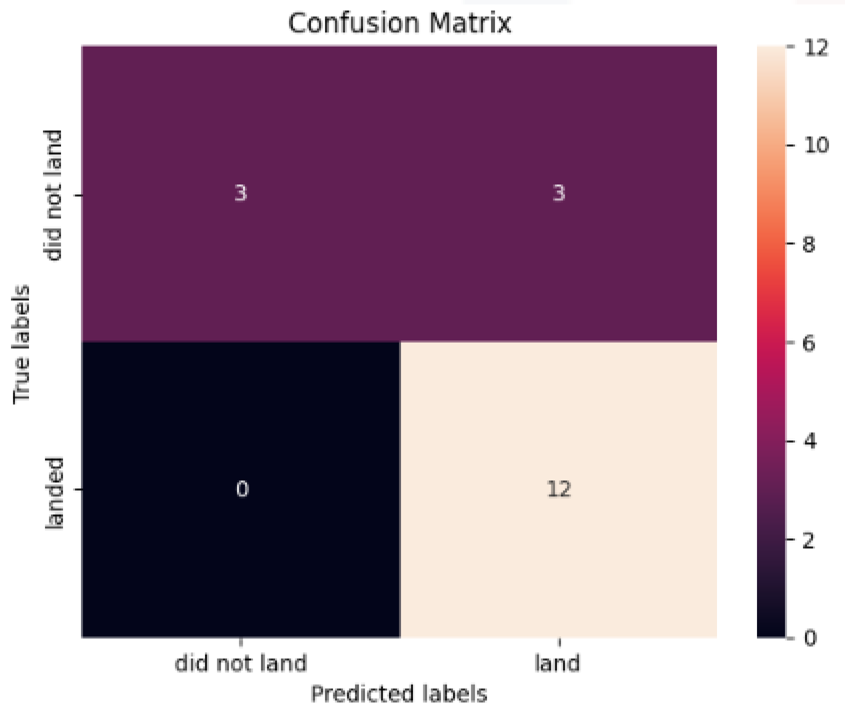
The Payload slider controls the success counts shown on the scatter plot. In this image, the entire payload range is selected.

RESULTS – Dashboard



In this example, only payloads ranging from 0-5000 kg are shown on the scatter plot.

RESULTS – Predictive Analysis




```
best_model = all_models[all_scores.index(max(all_scores))]
best_model
```

```
GridSearchCV(cv=10, estimator=LogisticRegression(),
             param_grid={'C': [0.01, 0.1, 1], 'penalty': ['l2'],
                         'solver': ['lbfgs']})
```

Several confusion matrices were generated using different models and the best model was found to be Logistic Regression with 83% confidence.

CONCLUSION

- Developed a robust predictive model for forecasting Falcon 9 rocket landing success.
- Identified key factors such as payload weight, launch site, and weather conditions as significant predictors.
- Created an interactive dashboard and map that provide valuable tools for real-time analysis and decision-making.
- Enhanced understanding of the variables influencing landing outcomes, supporting improved mission planning.