IBM DATA SCIENCE

APPLIED CAPSTONE PROJECT
ON SPACEX

07/06/2024

EXECUTIVE SUMMARY

- This research aims to identify the factors contributing to a successful rocket landing for SpaceX.
- By leveraging public data and machine learning models, we predict the likelihood of first-stage landings, which directly impacts launch costs.
- The study includes data collection through APIs and web scraping, data wrangling, exploratory data analysis (EDA), and predictive modeling using logistic regression, support vector machine (SVM), decision tree, and K-nearest neighbor (KNN) algorithms.
- Our findings highlight that launch success has improved over time, with certain launch sites and orbits showing higher success rates.
- Among the models, the decision tree slightly outperformed the others

INTRODUCTION

- This study aims to predict the likelihood of first-stage landings using historical launch data. Understanding the factors that influence landing success can help SpaceX and other space companies optimize their operations and reduce costs.

Objectives

- Determine how payload, launch site, number of flights, and orbits affect first-stage landing success.
- Analyze the rate of successful landings over time.
- Identify the best predictive model for successful landing outcomes.

METHODOLOGY

Data Collection – API

- SpaceX API: Data on rocket launches was collected using the SpaceX REST API.

Data Collection - Web Scraping

- Wikipedia: Falcon 9 launch data was scraped from Wikipedia.

Data Wrangling

- Outcomes were converted into binary values: 1 for a successful landing and 0 for an unsuccessful landing

METHODOLOGY Cont.

Exploratory Data Analysis (EDA)

- Visualization: Relationships and comparisons were analyzed using various charts.
- SQL Analysis: Queries were executed to gain deeper insights into the data.

Maps with Folium

- Visualization: Maps were created to visualize launch sites, outcomes, and distances to geographical proximities.

Dashboard with Plotly Dash

- A dashboard was developed to present key metrics

Predictive Analytics

- Algorithms applied: logistic regression (`LogisticRegression()`), support vector machine (`SVC()`), decision tree (`DecisionTreeClassifier()`), and K-Nearest Neighbor (`KNeighborsClassifier()`).
- Model accuracy was calculated using `.score()` on the test data

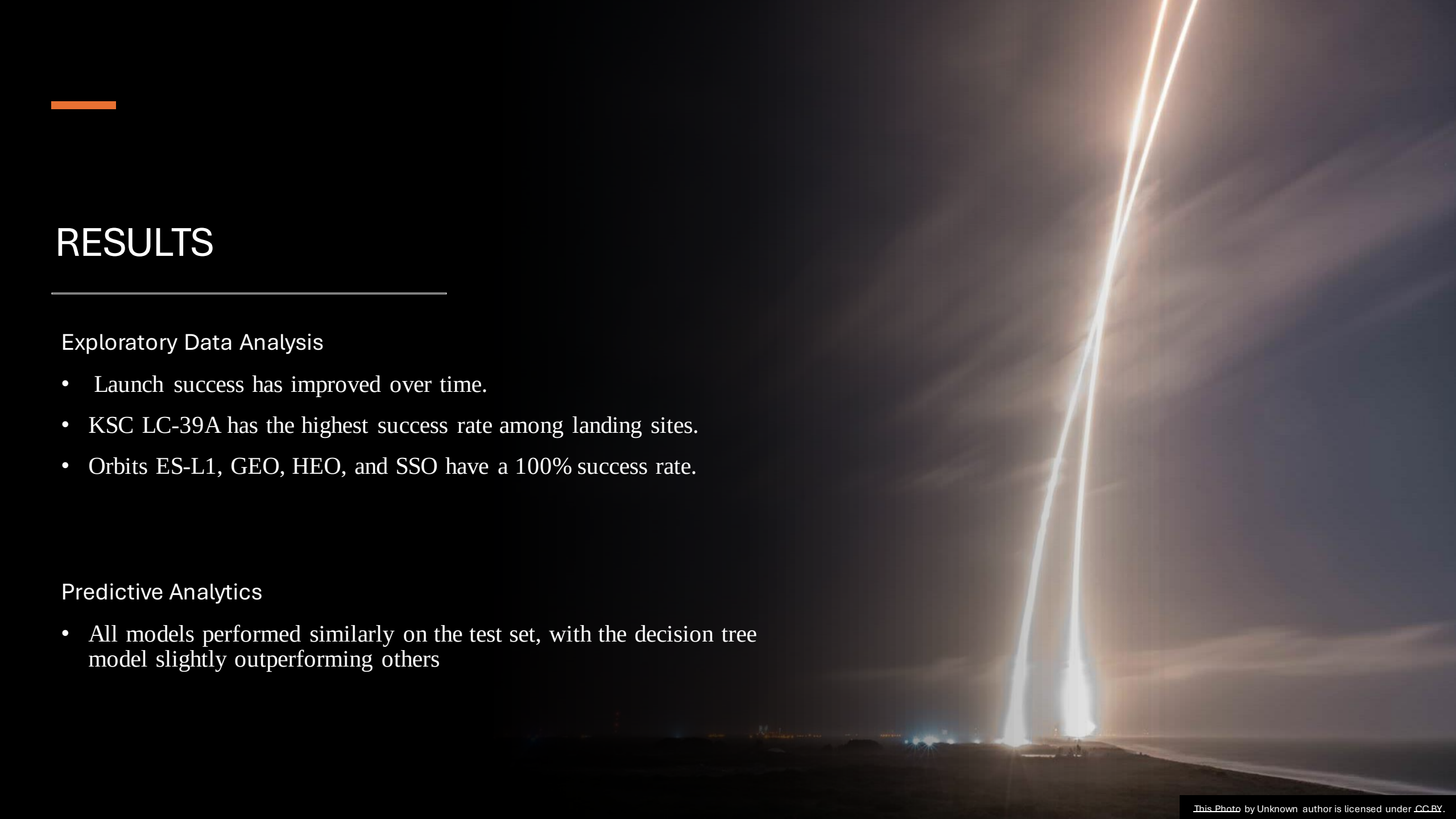
RESULTS

Exploratory Data Analysis

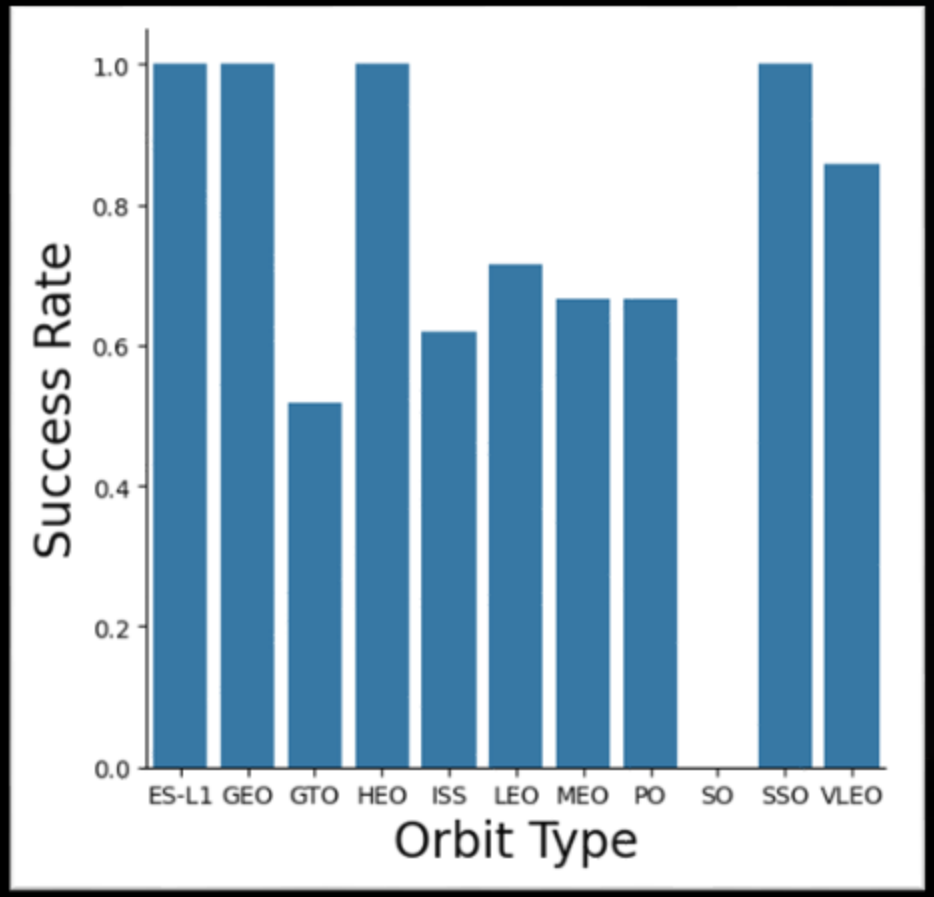
- Launch success has improved over time.
- KSC LC-39A has the highest success rate among landing sites.
- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate.

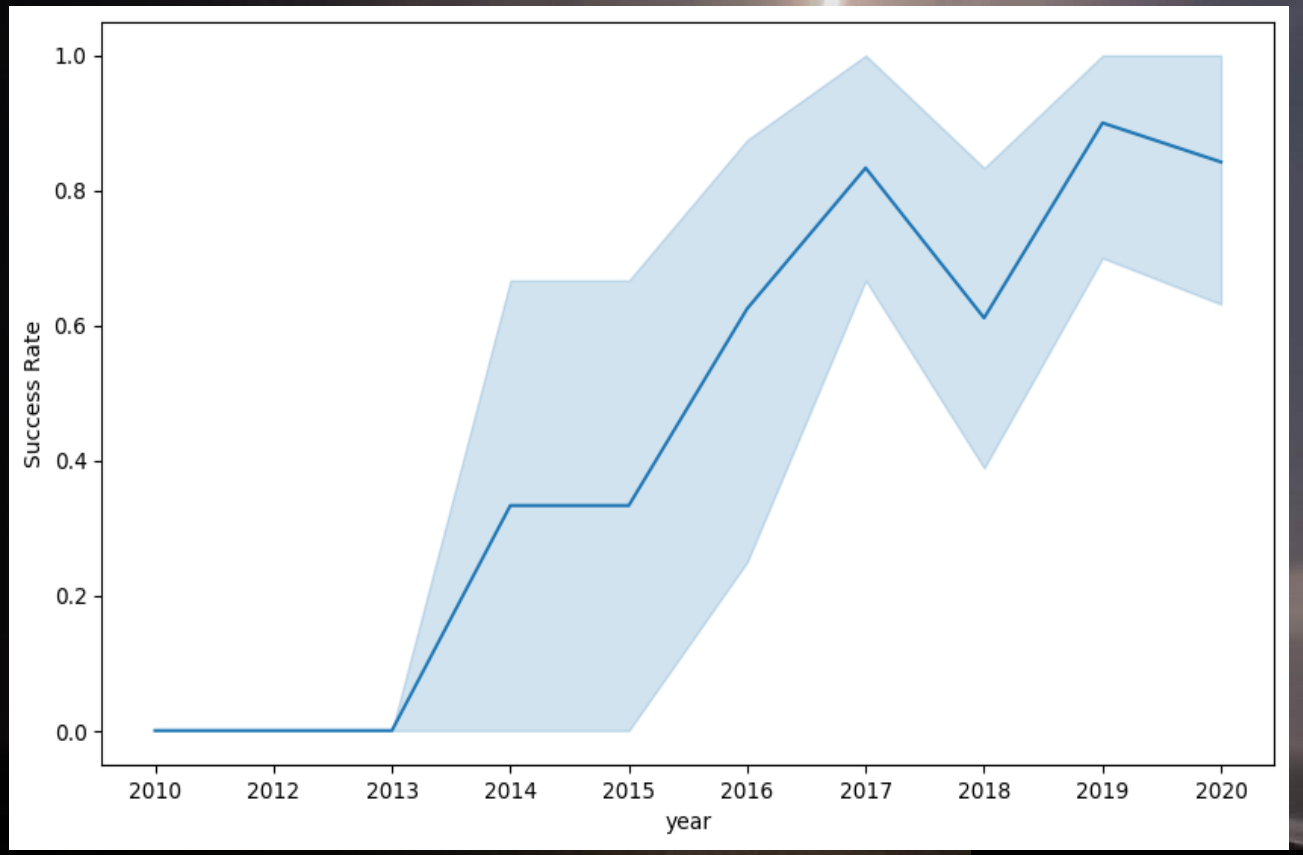
Predictive Analytics

- All models performed similarly on the test set, with the decision tree model slightly outperforming others



RESULTS Cont.





RESULTS Cont.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.769231	0.800000
F1_Score	0.888889	0.888889	0.869565	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

DISCUSSION

- Coastal locations facilitate logistics and safety measures.
- Over time, launch success has increased, highlighting SpaceX's improvements in technology and operations.
- Specific launch sites, such as KSC LC-39A, demonstrated higher success rates, particularly for launches under 5,500 kg.

CONCLUSION

By identifying the landing success,, site, payload, and location factors, SpaceX can further optimize its operations and reduce costs, making space travel more accessible and sustainable

APPENDIX

Launch Sites

- Cape Canaveral Air Force Station (CCAFS)
- Kennedy Space Center (KSC) LC-39A
- Vandenberg Air Force Base (VAFB)
- Boca Chica, Texas
- Cape Canaveral LC-40