Space X

IBM Coursera Capstone

CHEN PIN SYUE

2023/12/01

Executive Summary

This study aims to uncover the key elements contributing to a prosperous rocket landing. In pursuit of this objective, the research employs the following methodologies

Data Collection: Utilizing the SpaceX REST API and employing web scraping techniques to gather relevant data.

Data Wrangling: Organizing collected data to establish a success/fail outcome variable.

Data Exploration: Employing data visualization techniques to scrutinize factors such as payload, launch site, flight number, and annual trends.

Statistical Analysis: Utilizing SQL to analyze data and compute statistics including total payload, payload range for successful launches, and the overall number of successful and failed outcomes.

Site-specific Analysis: Investigating launch site success rates and their proximity to geographical markers.

Visualization: Representing launch sites with the highest success rates and optimal payload ranges through graphical means.

Predictive Modeling: Constructing models for predicting landing outcomes, employing logistic regression, support vector machine (SVM), decision tree, and K-nearest neighbor (KNN) algorithms.

Introduction

Background:

SpaceX, a pioneering force in the space industry, is committed to democratizing space travel by making it more cost-effective. Notable achievements include deploying spacecraft to the international space station, establishing a satellite constellation for global internet access, and conducting manned space missions. SpaceX's ability to achieve affordability is attributed to the innovative reutilization of the first stage of its Falcon 9 rocket, resulting in relatively low launch costs at $62 million per launch. In contrast, competitors lacking first stage reusability incur expenses exceeding $165 million per launch. The key determinant influencing launch costs is the successful landing of the first stage.

Objective:

The **primary objective** is to ascertain the predictability of first-stage landings, influencing the overall launch cost. This involves leveraging publicly available data and employing machine learning models to forecast the reusability of the first stage, not only for SpaceX but also for competing companies.

Methodology

1. Data Collection: Utilize SpaceX REST API alongside web scraping techniques to gather comprehensive data.

2. Data Wrangling: Enhance data quality by filtering, addressing missing values, and applying one-hot encoding, ensuring the data is well-prepared for subsequent analysis and modeling.

3. Exploratory Data Analysis (EDA): Employ SQL and data visualization techniques to delve into the data, uncovering patterns, correlations, and key insights.

4. Data Visualization: Leverage tools like Folium and Plotly Dash to create dynamic and informative visualizations, enhancing the understanding of the dataset and its nuances.

5. Model Building: Develop predictive models for landing outcomes using classification algorithms, considering features identified during the EDA phase.

6. Model Tuning and Evaluation: Fine-tune the models to optimize their performance and evaluate them rigorously. This involves adjusting parameters to enhance accuracy, precision, and recall.

7. Identify Optimal Model and Parameters: Systematically compare the tuned models, assessing their effectiveness in predicting landing outcomes. Identify the best-performing model and its optimal parameters for further utilization.

Data Collection

1. Data Retrieval: Initiate a request for Falcon 9 launch data from Wikipedia, aiming to obtain a comprehensive dataset.

2. HTML Parsing: Construct a BeautifulSoup object from the HTML response, facilitating the extraction of relevant information.

3. Column Identification: Extract column names from the HTML table header to understand the structure of the data.

4. Data Extraction: Collect data by parsing HTML tables, ensuring all pertinent details are accurately captured.

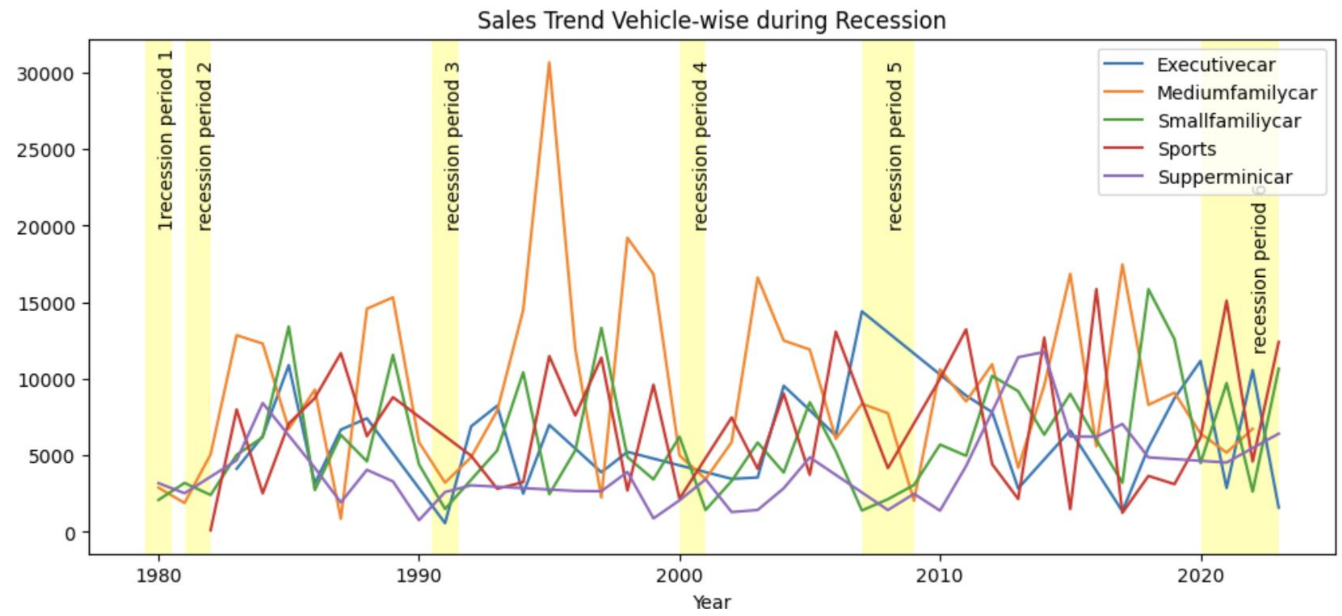
5. Dictionary Formation: Organize the extracted data into a structured dictionary format, promoting clarity and accessibility.

6. DataFrame Creation: Transform the dictionary into a structured DataFrame, allowing for efficient manipulation and analysis of the Falcon 9 launch data.

7. Data Export: Export the curated data to a CSV file, ensuring its availability for future reference or analysis beyond the initial scraping process.

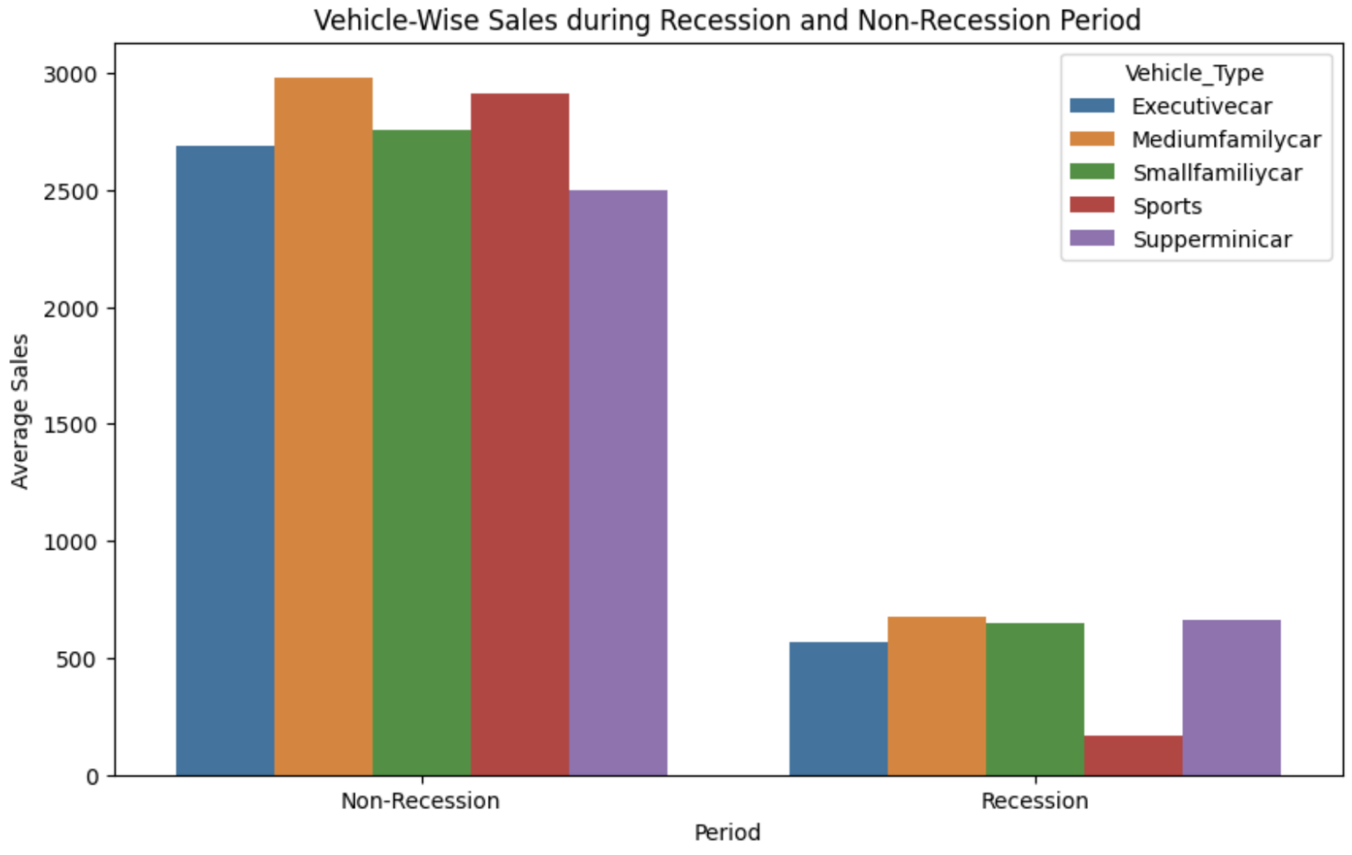
Data Visualization

We can see that sports car (red line) selling declines much more than other kinds of car during recession.



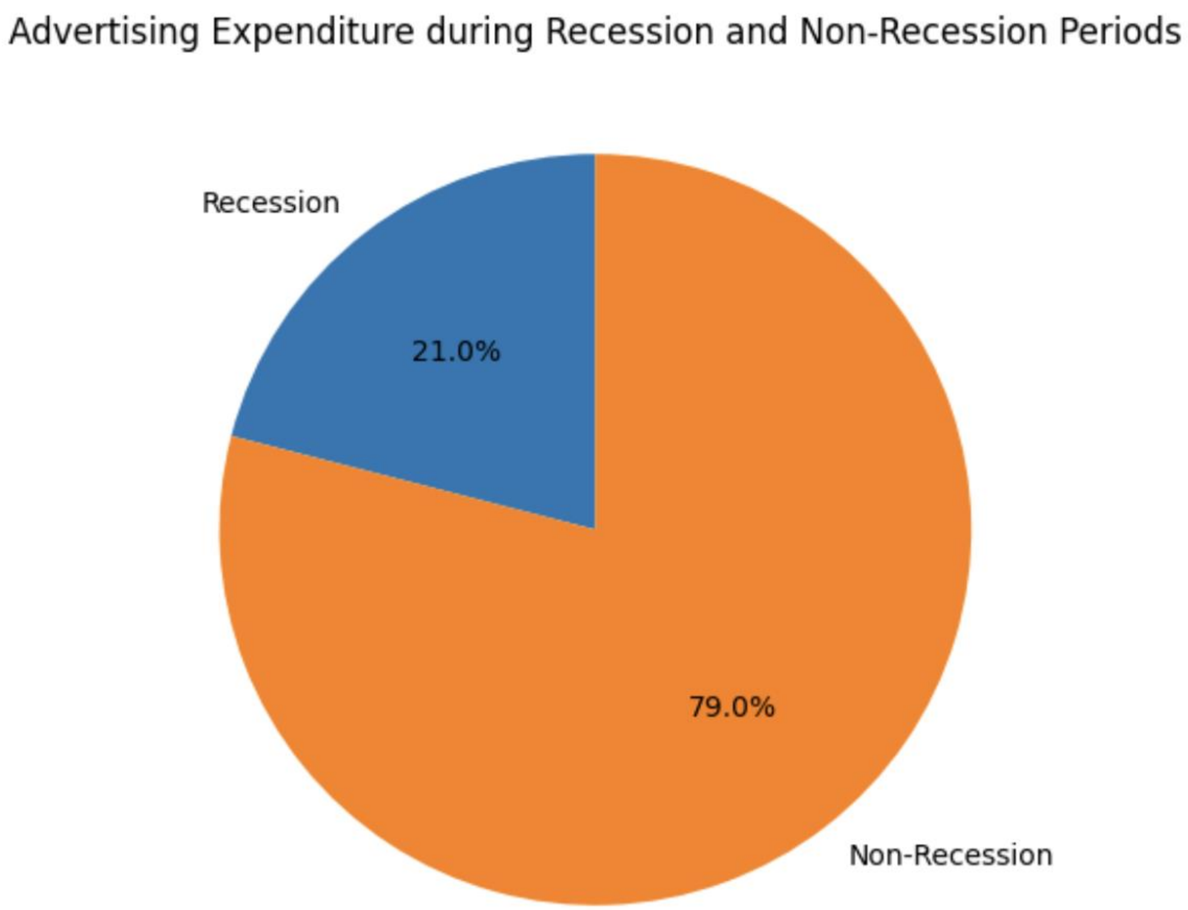
Data Visualization

By using bar chart, we can again prove sports car (red line) selling declines much more than other kinds of car during recession.



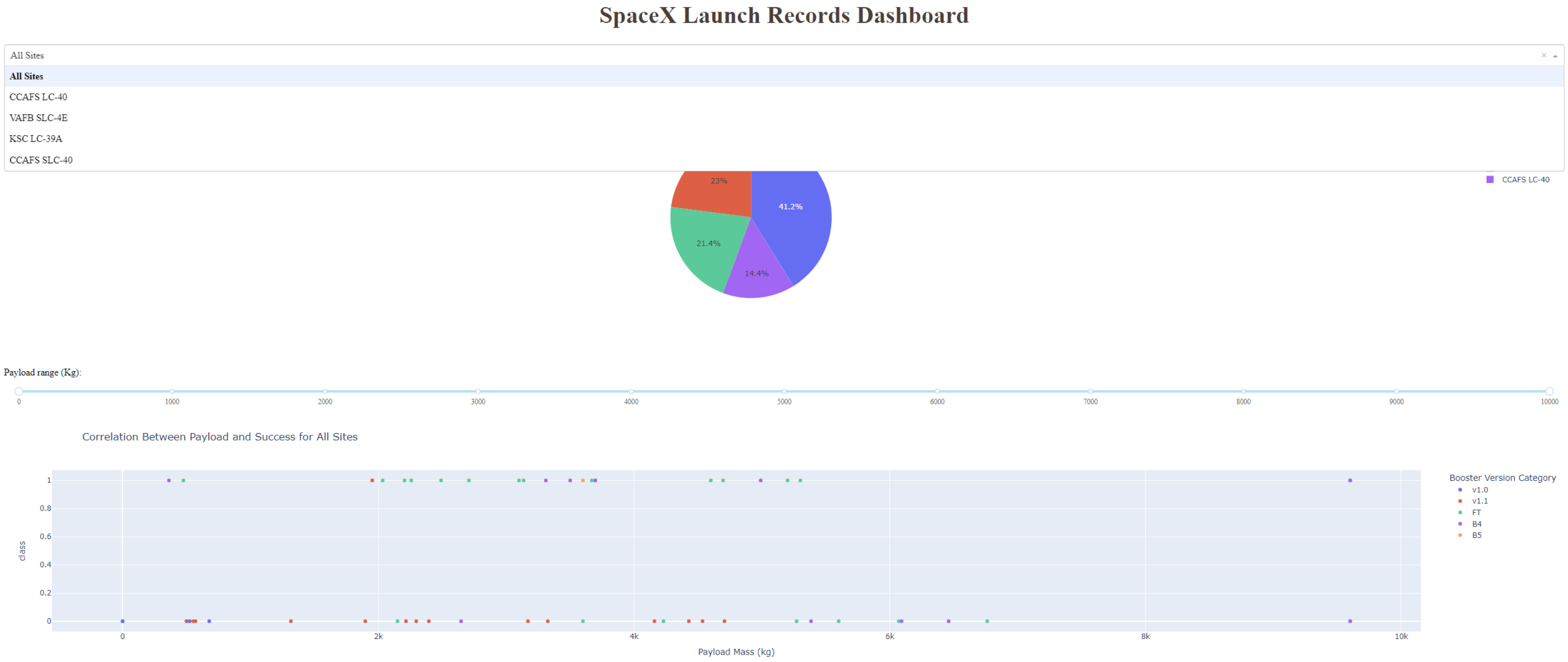
Data Visualization

During recession, the advertising Expenditure in much lesser than usual.



Interactive visual analytics

Use the menu to observe chart changes.



EDA with SQL

DISPLAY 5 RECORDS WHERE LAUNCH SITES BEGIN WITH THE STRING 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Lar
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Fai
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Fai
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

DISPLAY THE NAMES OF THE UNIQUE LAUNCH SITES IN THE SPACE MISSION.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

EDA with SQL

DISPLAY THE TOTAL PAYLOAD MASS CARRIED BY BOOSTERS LAUNCHED BY NASA (CRS).

total
45596

DISPLAY AVERAGE PAYLOAD MASS CARRIED BY BOOSTER VERSION F9 V1.1.

total
2928.4

EDA with SQL

LIST THE DATE WHEN THE FIRST SUCCESFUL LANDING OUTCOME IN GROUND PAD WAS ACHEIVED.

Date
2015-12-22
2016-07-18
2017-02-19
2017-05-01
2017-06-03
2017-08-14
2017-09-07
2017-12-15
2018-01-08

LIST THE NAMES OF THE BOOSTERS WHICH HAVE SUCCESS IN DRONE SHIP AND HAVE PAYLOAD MASS GREATER THAN 4000 BUT LESS THAN 6000.

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

EDA with SQL

LIST THE TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES.

COUNT(*)
100

LIST THE NAMES OF THE BOOSTER_VERSIONS WHICH HAVE CARRIED THE MAXIMUM PAYLOAD MASS. USE A SUBQUERY

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

EDA with SQL

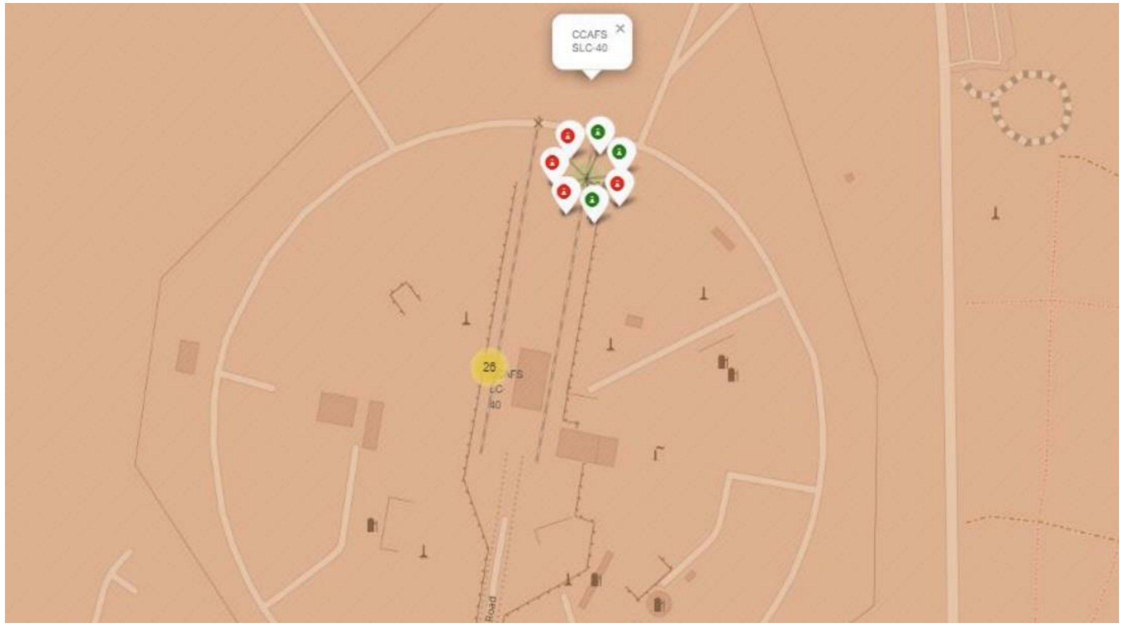
LIST THE RECORDS WHICH WILL DISPLAY THE MONTH NAMES, FAILURE LANDING_OUTCOMES IN DRONE SHIP ,BOOSTER VERSIONS, LAUNCH_SITE FOR THE MONTHS IN YEAR 2015.

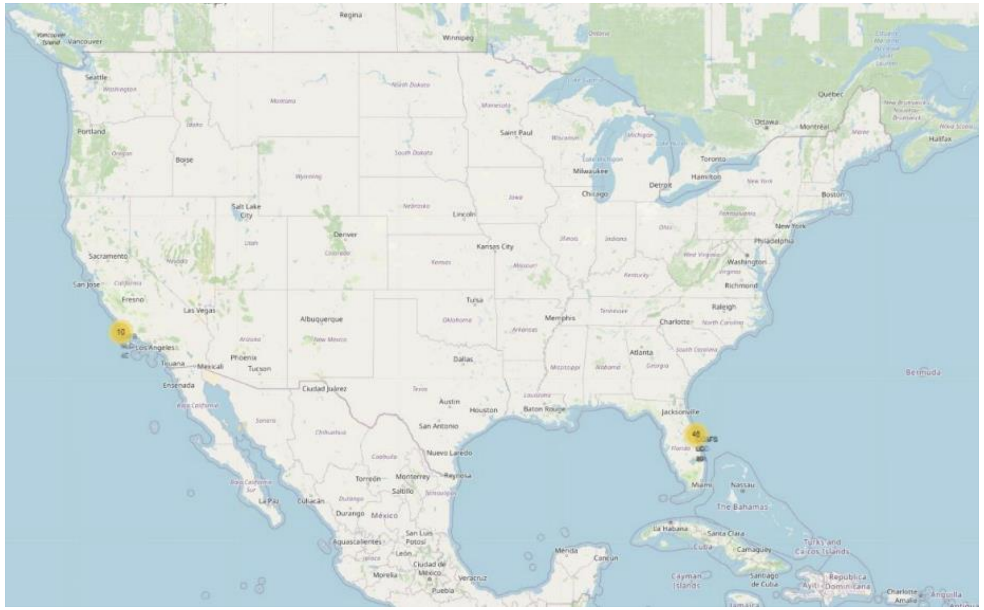
Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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.

Landing_Outcome	outcome_count	outcome_rank
Failure (drone ship)	5	1

Interactive map with Folium





Classification

The right side uses .score() to calculate the accuracy of each machine learning classification algorithm.

Logistic Regression is the best among them

	Model	Accuracy
0	Logistic Regression	0.833333
1	Support Vector Machine	0.777778
2	Decision Tree	0.777778
3	K Nearest Neighbors	0.777778

Conclusion

Environmental Factors:

Coastal Proximity:
All launch sites are situated in close proximity to coastlines, enhancing logistical and operational flexibility.

Trends and Success Rates:

Launch Success Over Time:
An encouraging trend reveals an increase in launch success rates over time, reflecting advancements and improvements in space launch technology.

KSC LC-39A Success:
KSC LC-39A stands out with the highest success rate among launch sites, achieving a perfect 100% success rate for launches below 5,500 kg.

Orbital Success:
Specific orbits, including ES-L1, GEO, HEO, and SSO, consistently demonstrate a flawless 100% success rate.

Conclusion

Model Performance:
The models exhibited comparable performance on the test set, with a slight advantage observed in the decision tree model.

Payload Analysis:

Payload Mass Impact:
Across all launch sites, there is a positive correlation between higher payload masses (kg) and increased success rates, underscoring the significance of payload mass in mission success.

Dataset and Further Analysis:

Dataset Size Consideration:
Expanding the dataset size could bolster the generalizability of predictive analytics results, providing insights into the broader applicability of the findings.

Feature Analysis / PCA:
A more in-depth feature analysis or principal component analysis (PCA) is recommended to explore if additional insights can be derived, potentially enhancing model accuracy.