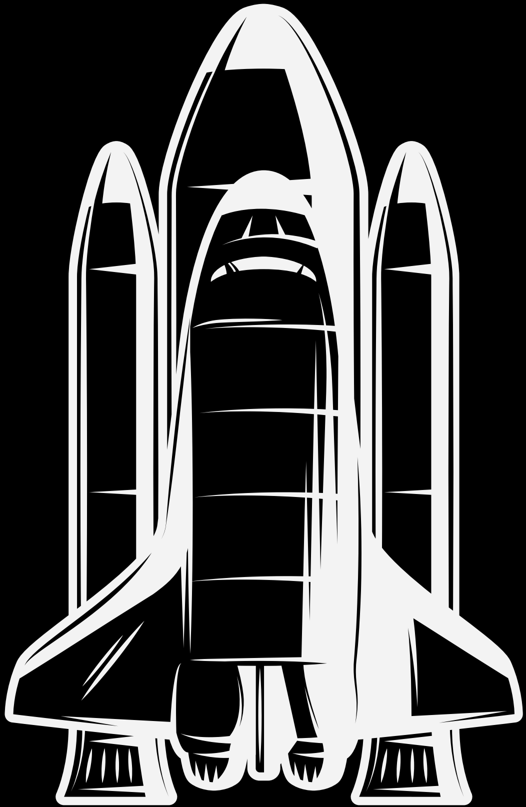


0001

Predicting Falcon 9 First Landing Success: A Data-Driven Approach

Executive Summary

01 Project Objective

The goal of this project is to predict the successful landing of the Falcon 9 first stage. SpaceX's Falcon 9 rockets are reusable, leading to significant cost savings. By determining the success of the first stage landing, we can estimate the cost of a launch and provide valuable information for competitive bidding.

02 Data Collection

Collect historical launch data from SpaceX using their API. This data includes details about the rocket, payloads, launch site, cores, and launch outcomes.

03 Data Processing

The data was cleaned and processed to focus on Falcon 9 launches. We filtered out non-relevant data and extracted useful features to build a comprehensive dataset for analysis.

04 Outcome

The processed dataset contains key features that will be used to develop predictive models for Falcon 9 landing success, enabling better cost estimation and competitive analysis.

01

INTRODUCTION

BACKGROUND

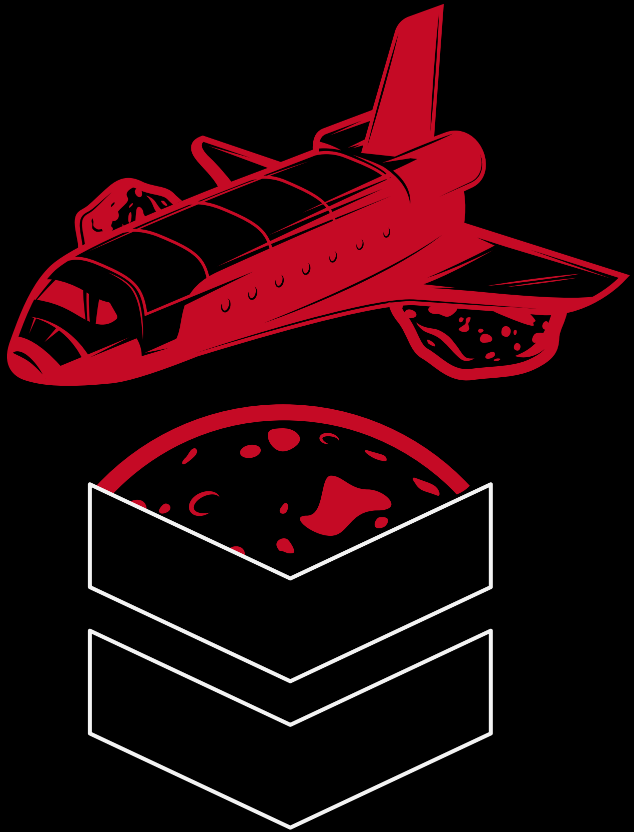
SpaceX has revolutionized the space industry by developing reusable rockets. The Falcon 9 rocket is a key part of this innovation, designed to return and land safely after launch, significantly reducing the cost of space missions.

OBJECTIVE

The main objective of this project is to predict whether the first stage of the Falcon 9 rocket will land successfully. This prediction is crucial for estimating the cost savings and for planning future missions.

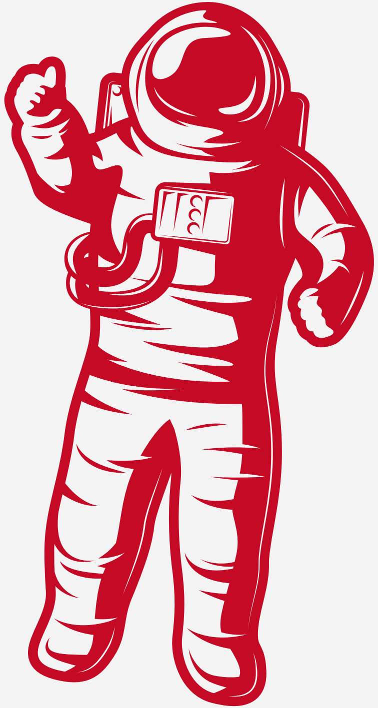
IMPORTANCE

Understanding and predicting the landing success of Falcon 9 can help in optimizing launch strategies, improving cost efficiency, and enhancing the competitiveness of space missions.



02

DATA COLLECTION METHODOLOGY



DATA SOURCE

The primary data source for this project is the SpaceX API, which provides detailed information about past launches, including rocket details, payloads, launch sites, and outcomes.

API REQUEST

I made GET requests to the SpaceX API to retrieve historical launch data. The data was fetched in JSON format and converted into a Pandas DataFrame for further processing.

DATA CLEANING

Initial data contained various IDs for rockets, payloads, launch sites, and cores. We performed additional API requests using these IDs to get detailed information. Rows with multiple cores or payloads were removed to maintain consistency.

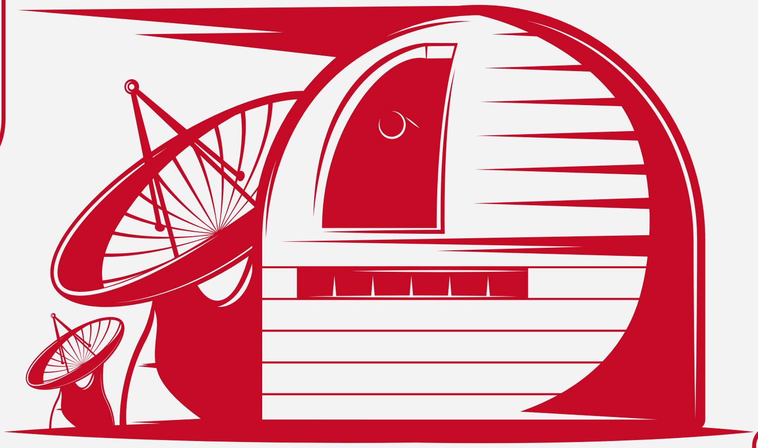
EXAMPLE DATA AND SUMMARY

```
   FlightNumber        Date BoosterVersion  PayloadMass Orbit    LaunchSite  \
4             6  2010-06-04       Falcon 9          NaN   LEO  CCSFS SLC 40
5             8  2012-05-22       Falcon 9        525.0   LEO  CCSFS SLC 40
6            10  2013-03-01       Falcon 9        677.0   ISS  CCSFS SLC 40
7            11  2013-09-29       Falcon 9        500.0    PO   VAFB SLC 4E
8            12  2013-12-03       Falcon 9       3170.0   GTO  CCSFS SLC 40

      Outcome  Flights  GridFins  Reused   Legs LandingPad  Block  \
4   None None        1     False   False  False       None    1.0
5   None None        1     False   False  False       None    1.0
6   None None        1     False   False  False       None    1.0
7 False Ocean        1     False   False  False       None    1.0
8   None None        1     False   False  False       None    1.0

   ReusedCount Serial   Longitude   Latitude
4            0  B0003  -80.577366  28.561857
5            0  B0005  -80.577366  28.561857
6            0  B0007  -80.577366  28.561857
7            0  B1003 -120.610829  34.632093
8            0  B1004  -80.577366  28.561857
       FlightNumber   PayloadMass    Flights      Block  ReusedCount  \
count     90.000000     85.000000  90.000000  90.000000    90.000000
mean      56.477778   6123.547647   1.788889   3.500000     3.188889
std       29.232977   4870.916417   1.213172   1.595288     4.194417
min        6.000000    350.000000   1.000000   1.000000     0.000000
25%       32.250000   2482.000000   1.000000   2.000000     0.000000
50%       55.500000   4535.000000   1.000000   4.000000     1.000000
75%       82.750000   9600.000000   2.000000   5.000000     4.000000
max      106.000000  15600.000000   6.000000   5.000000    13.000000

        Longitude   Latitude
count   90.000000  90.000000
mean   -86.366477  29.449963
std     14.149518   2.141306
min   -120.610829  28.561857
25%    -80.603956  28.561857
50%    -80.577366  28.561857
75%    -80.577366  28.608058
max    -80.577366  34.632093
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None
...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca

03

DATA WRANGLING AND PROCESSING



DATA WRANGLING AND PROCESSING

LOAD AND INSPECT DATA

Loaded dataset from SpaceX API CSV file and displayed initial rows to understand the structure.

```
Number and Occurrence of Mission Outcomes:
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: count, dtype: int64
```

HANDLING MISSING VALUES

Calculated the percentage of missing values for each attribute

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: count, dtype: int64
```

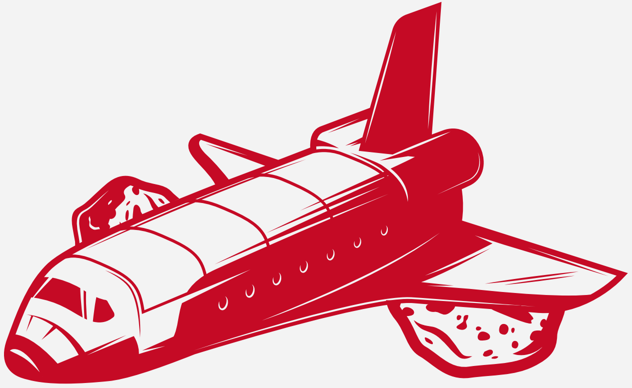
IDENTIFY DATA TYPE

Determined which columns are numerical and which are categorical.

ANALYZE LAUNCH SITES

Counted the number of launches at each site

04



EDA with SQL Results

OVERVIEW AND LAUNCH SITES ANALYSIS

OBJECTIVE

Perform Exploratory Data Analysis (EDA) on SpaceX dataset using SQL queries

KEY TASKS

1. identify unique launch sites.
2. Extract specific records based on criteria.
3. Analyze payload mass and mission outcomes.

UNIQUE SITES ANALYSIS

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

PAYLOAD AND LANDING ANALYSIS

Total Payload Mass by NASA (CRS)

```
# Task 3
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE 'NASA (CRS)%'
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
None

Average Payload Mass by Booster Version F9 v1.1

```
 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Pad Landing

```
 * sqlite:///my_data1.db
Done.
```

MIN("Date")
2015-12-22

PAYLOAD AND LANDING ANALYSIS

Launch Sites Beginning with 'CCA'

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

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters with Success in Drone Ship

Ranked by count in descending order:

05

EDA with Visualization

Overview and Key Insights from EDA with Visualization

OBJECTIVE

Perform Exploratory Data Analysis (EDA) and visualize key relationships in the SpaceX dataset.

KEY TASKS

Flight Number vs. Launch Site:
Observation: As the number of flights increases, the success rate tends to improve.
Success Rates: KSC LC-39A and VAFB SLC-4E have higher success rates compared to CCAFS LC-40.

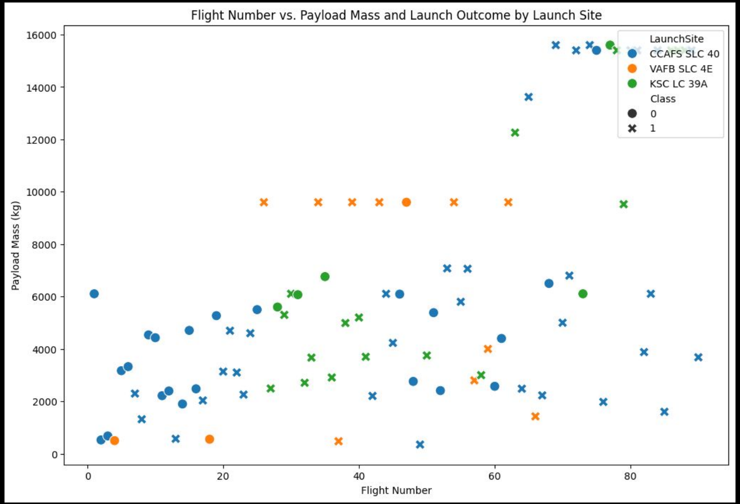
Payload Mass vs. Launch Site:
Observation: VAFB SLC-4E has no launches with payloads greater than 10,000 kg.

Orbit Success Rates:
Observation: LEO and ISS orbits have higher success rates, while GTO has a lower success rate.

Detailed Visualizations

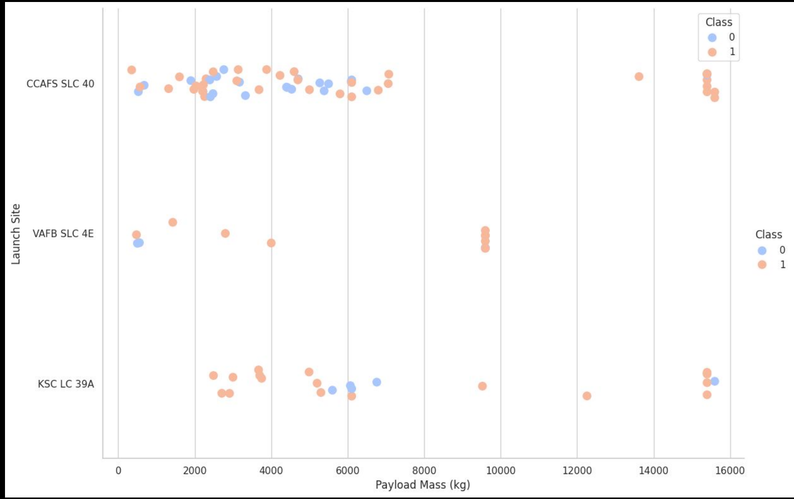
Flight Number vs. Payload Mass by Launch Site

1. Higher flight numbers correlate with higher success rates.
2. Different launch sites show varying success rates.



Payload Mass vs. Launch Site

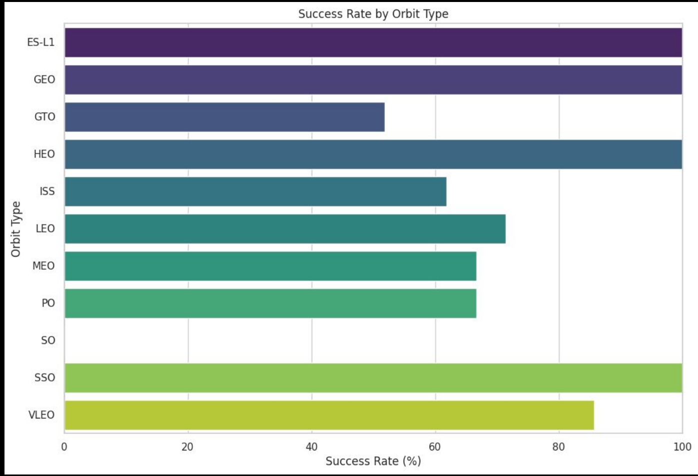
VAFB SLC-4E launch site does not handle heavy payloads (>10,000 kg).



Additional Insights and Yearly Trends

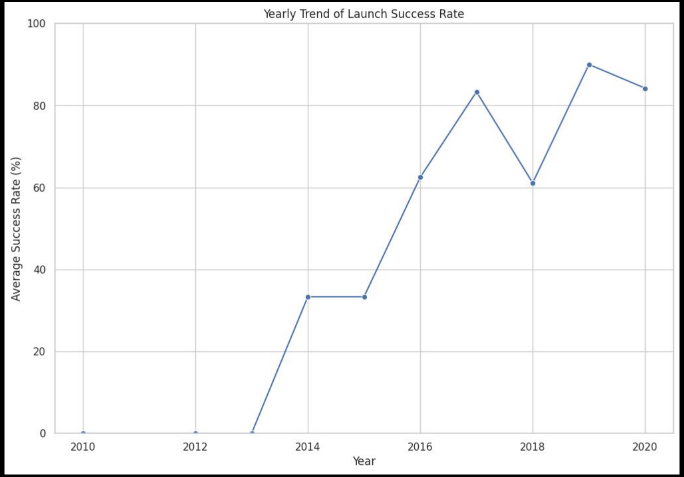
Orbit Type Success Rates

1. LEO and ISS orbits have higher success rates.
2. GTO orbit has lower success rates, indicating challenges with high-altitude launches.



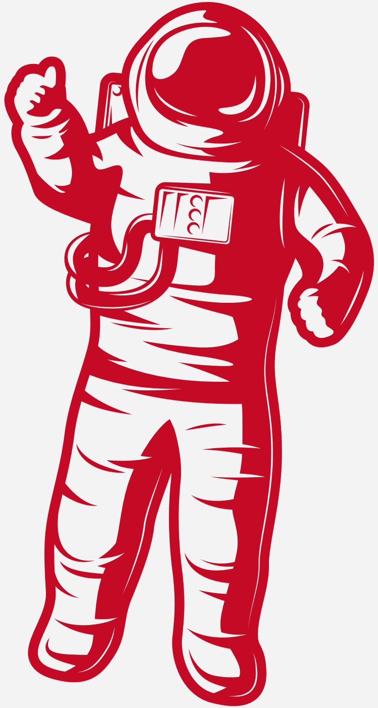
Yearly Trend of Launch Success Rate

Success rates have been steadily increasing since 2013, reaching nearly 100% in recent years.



06

EDA and Interactive Visual Analytics – Methodology



Exploratory Data Analysis and Interactive Visual Analytics

OBJECTIVE

Perform comprehensive EDA and interactive visual analytics to understand patterns in SpaceX launch data and determine factors affecting launch success.

METHODOLOGY

Data Collection:
1. Downloaded and processed SpaceX dataset with added latitude and longitude coordinates for launch sites.

2. Utilized CSV files to read and preprocess data using Pandas.

Exploratory Data Analysis:
1. Visualized relationships between flight number, payload mass, launch site, and launch outcomes using Seaborn and Matplotlib.

2. Key Findings:
- Higher flight numbers correlate with higher success rates.
- Different launch sites have varying success rates.
- Certain orbits like LEO and ISS have higher success rates compared to GTO.

Interactive Visual Analytics with Folium

Mapped all launch sites and outcomes using Folium.
Calculated distances to proximities (e.g., coastline, highway, railway, city).
Key Findings:

- Launch sites are near the Equator and coastlines.
- Proximity to railways and highways for transportation purposes.
- Safe distances maintained from cities to avoid accidents in populated areas.

07

Interactive Map with Folium

Interactive Map with Folium - Launch Sites Analysis

OBJECTIVE

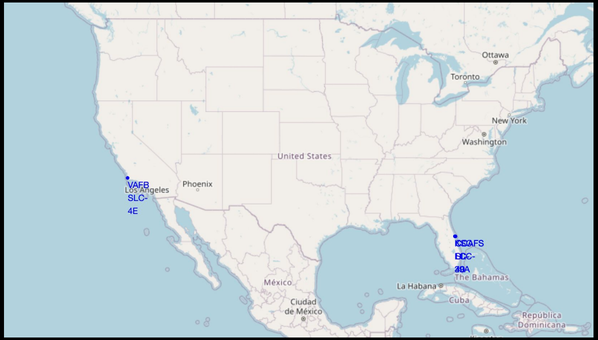
Perform interactive visual analytics using Folium to explore the geographical patterns of SpaceX launch sites.

KEY TASKS

1. Mark all launch sites on a map.
2. Mark the success/failed launches for each site on the map.
3. Calculate the distances between a launch site and its proximities (e.g., coastline, highway, railway, city).

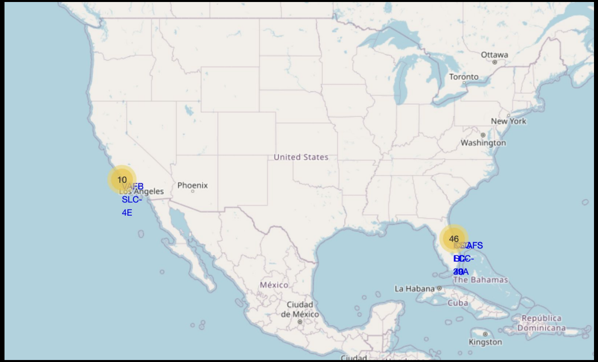
2024

Marking Launch Sites and Outcomes



All Launch Sites Marked

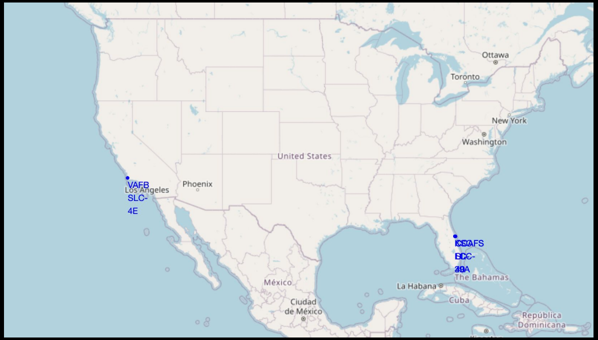
1. Each launch site is marked with a blue circle and a label.
2. NASA Johnson Space Center at Houston, Texas used as the initial center location.

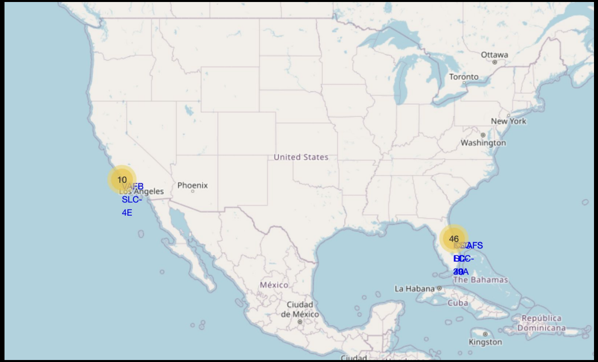


Launch Outcomes

1. Successful launches are marked with green markers.
2. Failed launches are marked with red markers.

Marking Launch Sites and Outcomes





Questions & Answer

Q: Are all launch sites in proximity to the Equator line?
A: Yes, most launch sites are near the Equator to maximize speed gain from Earth's rotation.

Q: Are all launch sites in very close proximity to the coast?
A: Yes, all launch sites are close to the coast for safe disposal of rocket debris.

Proximity Analysis

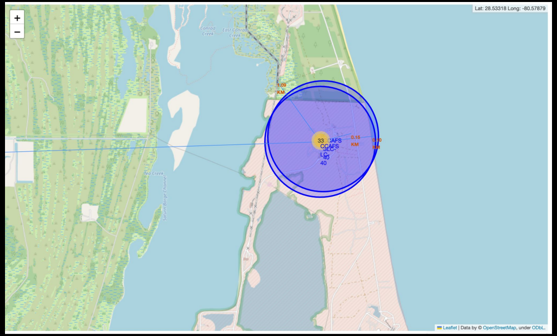
Distance Calculations

1. Added MousePosition plugin to get coordinates of points of interest (railway, highway, coastline, etc.).

2. Used Haversine formula to calculate distances.

Visualizations

Markers and lines drawn for proximity points (e.g., closest coastline, railway, highway, city).



Findings

Railways: Launch sites are in close proximity to railways for transportation of large rockets and equipment.

Highways: Launch sites are near highways to ensure convenient transportation of equipment and personnel.

Coastline: All launch sites are close to the coastline for safe rocket debris disposal in case of failure. Cities: Launch sites are usually kept at a safe distance from cities to avoid accidents in populated areas.

09

Plotly Dash Dashboard

Interactive Visual Analytics with Plotly Dash

Introduction to Plotly Dash Application

1. Developed a dashboard to perform interactive visual analytics on SpaceX launch data.

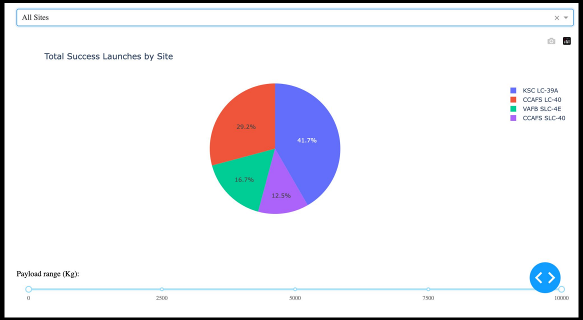
2. Utilized Plotly Dash for real-time data interaction and visualization.

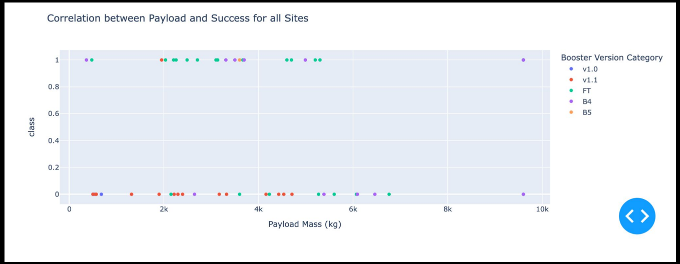
Dashboard Components

Dropdown Menu: Allows selection of different launch sites. Range Slider: Enables filtering of data based on payload mass.

Pie Chart: Visualizes the success counts of launches per site.

Scatter Plot: Displays the correlation between payload mass and launch outcomes, color-labeled by booster version.

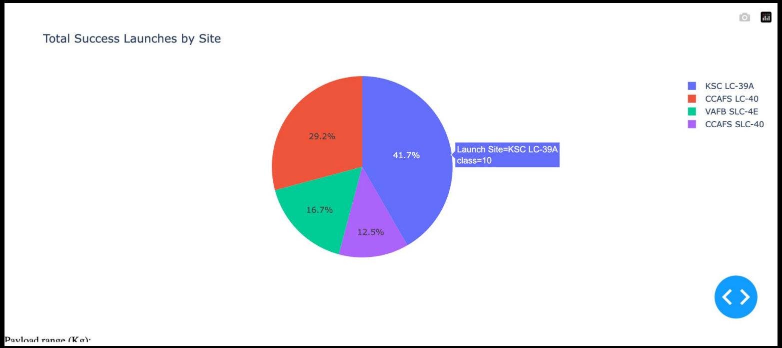




Insights and Findings

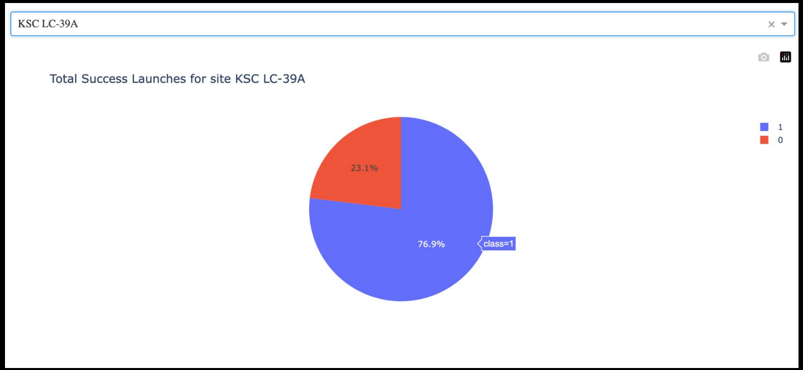
Q: Which site has the largest successful launches?

A: KSC LC-39A has the largest number of successful launches.



Q: Which site has the highest launch success rate?

A: KSC LC-39A has the highest launch success rate.



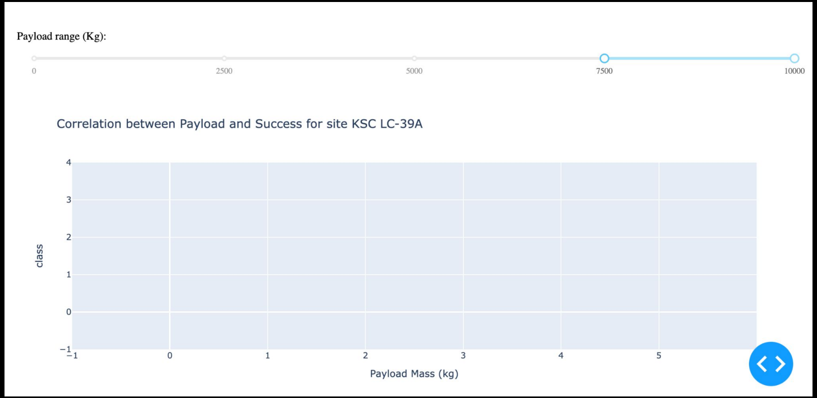
Insights and Findings

Q: Which payload range(s) has the highest launch success rate?

A: Payload ranges between 0 to 5000 kg have the highest success rate.

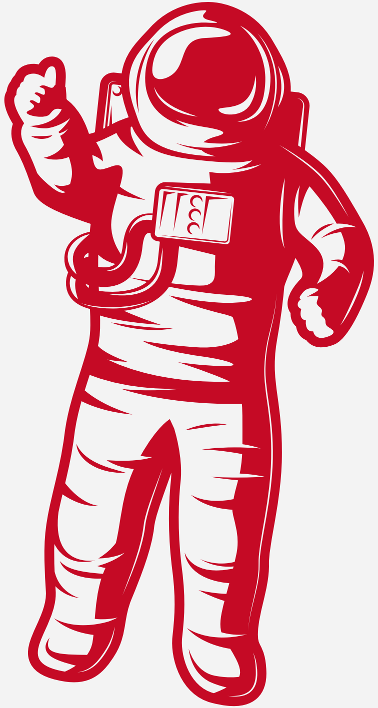
Q: Which payload range(s) has the lowest launch success rate?

A: Payload ranges between 7500 to 10000 kg have the lowest success rate.



10

Predictive Analysis Methodology



Predictive Analysis Methodology for SpaceX Falcon 9 Landing Prediction

OBJECTIVE

Predict if the Falcon 9 first stage will land successfully

DATA PREPARATION

Source: SpaceX launch dataset

Features: FlightNumber, PayloadMass, Orbit, LaunchSite, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial

Target: Class (1 = success, 0 = failure)

STEPS

Data Cleaning: Handled missing values and standardized data Feature

Engineering: Created dummy variables for categorical features Train-Test Split: Split data into 80% training and 20% testing

Predictive Analysis Methodology for SpaceX Falcon 9 Landing Prediction

Models Used

Logistic Regression
Support Vector Machine (SVM)
Decision Tree
K-Nearest Neighbors (KNN)

Model Evaluation

Used GridSearchCV for hyperparameter tuning and confusion matrices for performance evaluation

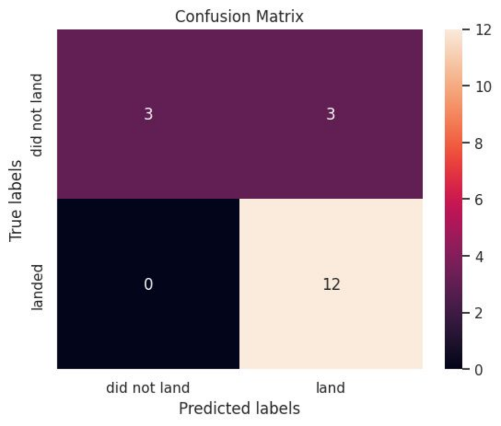
11

Predictive Analysis (Classification) Results

Logistic Regression & SVM Results

Logistic Regression

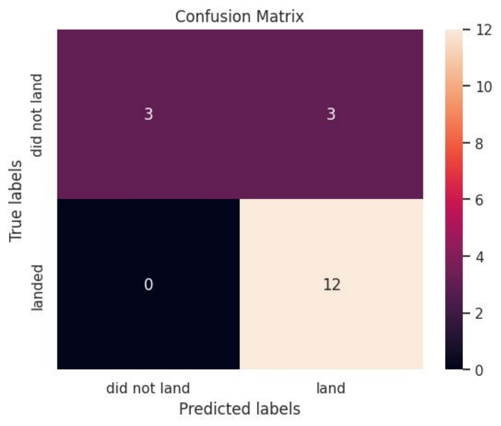
1



Best Parameters:
{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
Validation Accuracy: 0.846
Test Accuracy: 0.833

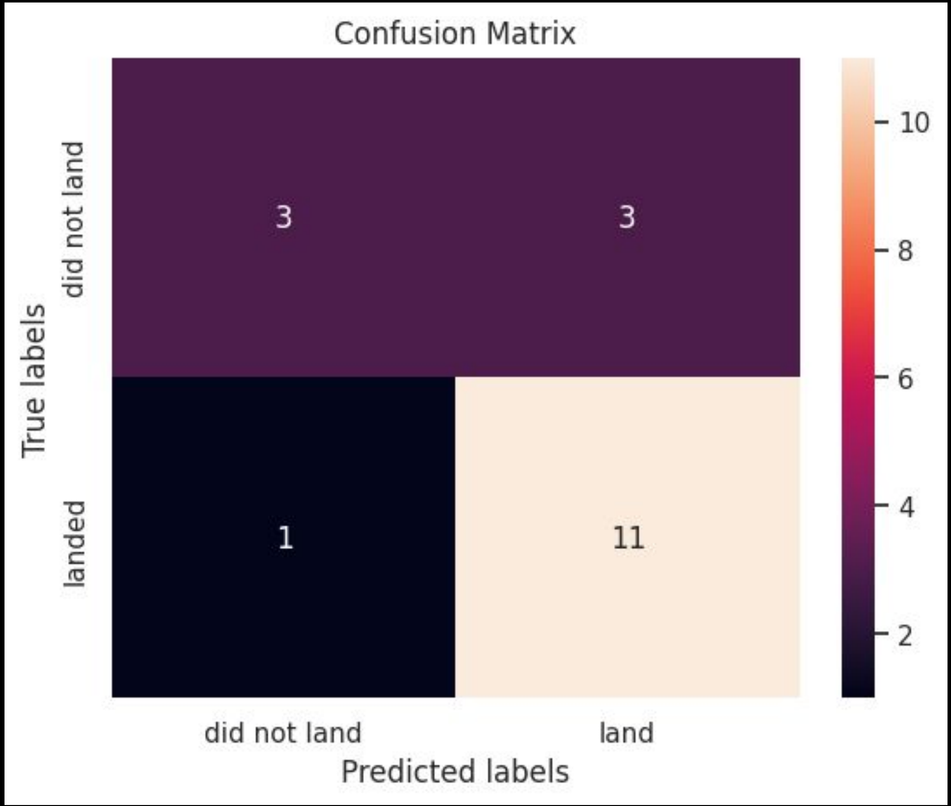
Support Vector Machine (SVM)

2



Best Parameters:
{'C': 1.0, 'gamma': 0.001, 'kernel': 'rbf'}
Validation Accuracy: 0.849
Test Accuracy: 0.833

Decision Tree



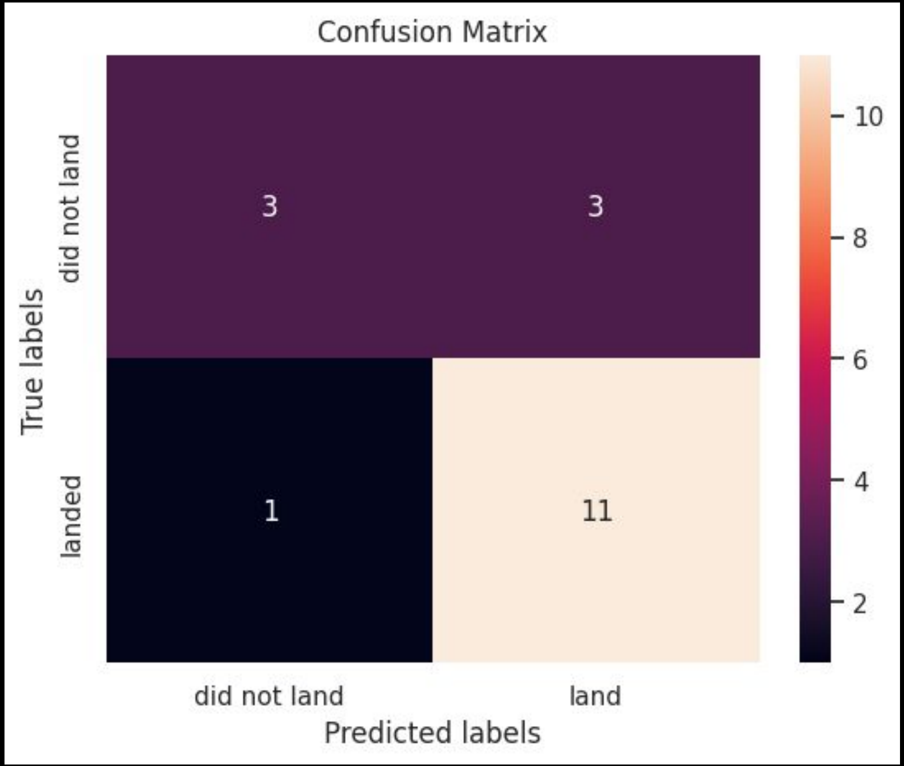
Decision Tree

Best Parameters:
{'criterion': 'gini', 'max_depth': 6,
'max_features': 'auto',
'min_samples_leaf': 1,
'min_samples_split': 2, 'splitter':
'best'}
Validation Accuracy: 0.879
Test Accuracy: 0.778

K-Nearest Neighbors (KNN)

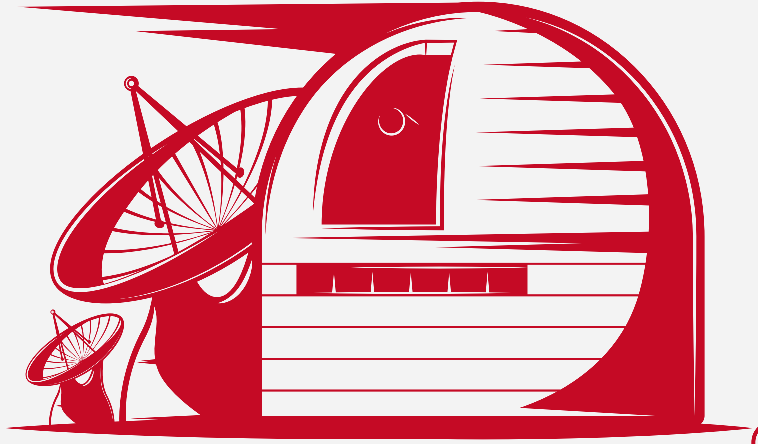
KNN

Best Parameters:
{'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
Validation Accuracy: 0.848
Test Accuracy: 0.833



12

Conclusion



Conclusion

83.33%

Model Performance

Logistic Regression, SVM, and KNN all achieved a test accuracy of 0.833.

Decision Tree had a lower test accuracy of 0.778.

Best Model

Logistic Regression with the highest validation accuracy and robust performance on the test set.

Insights

1. Accurate prediction of landing success can significantly reduce launch costs.

2. Consistent model performance suggests a strong correlation between the selected features and the target variable.

13

Innovative Insights

Innovative Insights

2400

Feature Importance

Importance of categorical features such as LaunchSite, Orbit, and Booster Version in predicting success.

Continuous improvement in launch success rates over time.

Future Work

Ensemble Methods: Explore Random Forest, Gradient Boosting for potentially higher accuracy.

Deep Learning: Implement neural networks to capture complex patterns.

Real-time Prediction: Deploy the best model in a real-time prediction system for future launches.

Business Impact

Potential cost savings by accurately predicting landing success.

Competitive advantage in the aerospace industry by optimizing launch strategies.

THANKS!