IBM SpaceX Rocket Reusability

Mike Piatek

04/20/2023

IBM Developer SKILLS NETWORK

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Research aimed to identify factors for successful rocket landings, using the following methodologies:

Data collected via SpaceX API and web scraping

Data structured for success/fail outcome variable

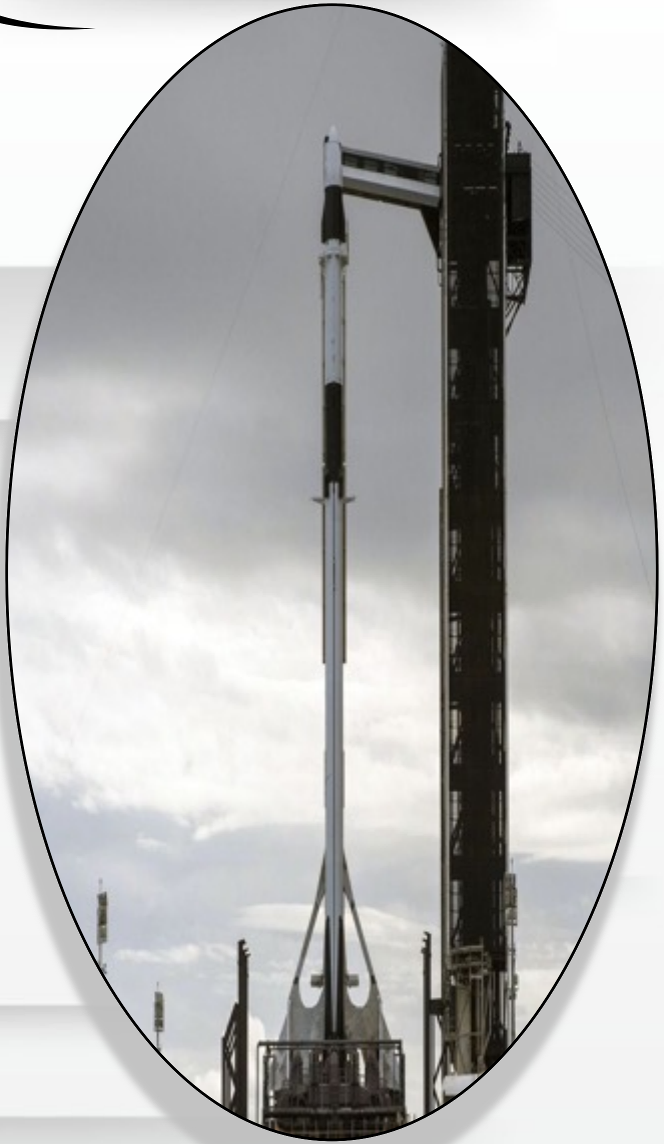
Data patterns visualized for payload, launch site, flight number, and yearly trends

Data analyzed with SQL for total payload, payload range, and success/failure outcomes

Launch site success rates investigated in relation to geographical markers

Visualizations created for successful launch sites and payload ranges

Models built with logistic regression, SVM, decision tree, and KNN algorithms for outcome prediction



EXECUTIVE SUMMARY

Summary of Results

Exploratory Data Analysis	Predictive Analytics	Visualization/Analytics
Findings suggest improvement in launch success over time KSC LC-39A demonstrated the highest success rate among all landing sites Orbits ES-L1, GEO, HEO, and SSO all achieved a 100% success rate	All models showed comparable performance on the test set Decision tree model outperformed the other models slightly	Most launch sites are situated near the equator and coastal areas

INTRODUCTION

SpaceX is a leading force in making space travel accessible and affordable.

Their impressive track record includes launching satellites for internet access, sending spacecraft to the ISS, and manned missions to space.

How do they keep launch costs at $62 million when competitors charge upwards of $165 million?

By reusing the Falcon 9's first stage, which they can do because they successfully land it.

Predicting a successful landing can determine launch price, and machine learning models and public data can be used for this.

This approach leads to greater pricing accuracy and transparency in the industry.

INTRODUCTION

1. Predicted rocket landing success rates in space exploration
2. Explored factors impacting landing success
3. Identified contributing factors and examined success rate trends
4. Determined best predictive model through binary classification
5. Improve space exploration efficiency and contribute to cost-effectiveness
6. Our work supports scientific community's push for expansion

PROJECT PROCEDURES

PROJECT PROCEDURES

Collect

Utilize SpaceX's REST API and advanced web scraping techniques to collect comprehensive data

Wrangle

Streamline the data by filtering and handling missing values, and optimize it for analysis and modeling through the implementation of one-hot encoding

Explore

Conduct extensive exploratory data analysis (EDA) using SQL and employ various data visualization techniques to gain valuable insights

Visualize

Create dynamic and interactive visualizations of the data using advanced tools such as Folium and PlotlyDash

Machine Learning

Develop powerful classification models to accurately predict landing outcomes, fine-tuning and evaluating them to determine the optimal model and parameters

DATA COLLECTION - SPACEX API

Retrieving response from API endpoint

Parsing response to JSON format

Applying customized data cleaning functions

Converting list to dictionary and then to data frame

Filtering data frame and exporting it to a flat file format (e.g., .csv)

GitHub

DATA COLLECTION - WEB SCRAPPING

1. Retrieving HTML response from server
2. Instantiating a BeautifulSoup object to parse the HTML response
3. Locating tables within the HTML response
4. Extracting column names from the table headers
5. Creating a dictionary object to hold the table data
6. Appending data rows to their respective dictionary keys
7. Converting the dictionary object to a pandas data frame
8. Exporting the data frame to a .csv file format

GitHub

DATA WRANGLING

Landing Outcome	Landing Clarity	Data Wrangling Process
True Ocean	Mission success: Accurately landed in designated ocean region	Conducted an Exploratory Data Analysis (EDA) to determine the appropriate data labels
False Ocean	Mission failure: Unsuccessful landing within ocean region	Calculated number of launches for each launch site
True RTLS	Mission success: Landed successfully on a ground pad	Calculated frequency of orbit occurrence
False RTLS	Mission failure: Unsuccessful landing on a ground pad	Calculated frequency of mission occurrence
True ASDS	Mission success: Successfully landed on a drone ship	Calculated outcomes per orbit type
False ASDS	Mission failure: Unsuccessful landing on a drone ship	Created a binary landing outcome column as the dependent variable to facilitate predictive modeling
Outcomes Converted	Encoding landing outcomes as 1 (successful) or 0 (unsuccessful)	Exported the cleaned and refined data to a CSV file for further analysis and modeling

GitHub

EDA - VISUALIZATION

Throughout the course of the research study, various graphical representations were employed to effectively communicate the data. Specifically, the three types of graphs utilized were bar graphs, line graphs, and scatter graphs, each serving a distinct purpose in visually depicting the findings

Bar Graph

Success Rate vs Orbit

Bar graphs are effective for displaying and comparing categorical data

Line Graph

Success Rate vs Year

Line graphs are a preferred method of data visualization due to their ability to effectively communicate trends and patterns over time or across a continuous variable

Scatter Graph

Flight Number vs Launch Site
Payload vs launch Site
Flight Number vs Orbit
Payload vs Orbit

Scatter graphs are useful for identifying the relationship between two continuous variables

GitHub

EDA - SQL

Below are the SQL queries that were executed during the course of the project:

1. **Displayed** a comprehensive list of unique launch sites utilized in the space mission
2. **Displayed** records of five launch sites beginning with the designated string "KSC"
3. **Displayed** the total payload mass carried by NASA (CRS) boosters launched
4. **Displayed** the average payload mass carried by booster version F9 v1.1
5. **Listed** the dates of successful drone ship landing outcomes
6. **Listed** the names of boosters achieving success in ground pad with a payload mass greater than 4000 but less than 6000
7. **Listed** a tally of the number of successful and failed mission outcomes
8. **Listed** the booster versions that have carried the maximum payload mass
9. **Listed** records displaying the month names, successful landing outcomes in ground pad, booster versions, and launch site for months in the year 2017
10. **Ranked** the count of successful landing outcomes between the dates of 2010-06-04 and 2017-03-20 in descending order

GitHub

MAP - FOLIUM

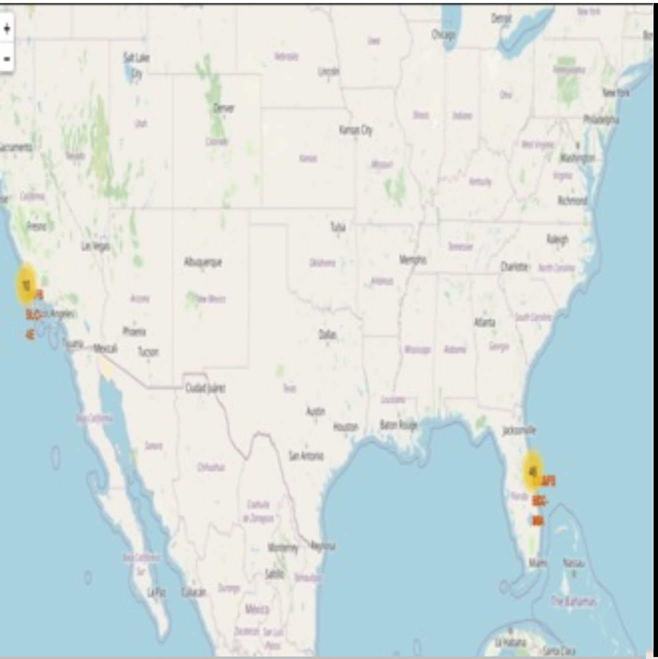


Image Two

As part of our data visualization efforts, we assigned the launch outcomes of failures and successes to respective classes of 0 and 1. These classes were then visually represented on the map using marker clusters, with green markers denoting successes and red markers indicating failures

Questions?

Are launch sites in close proximity to railways? No
Are launch sites in close proximity to highways? No
Are launch sites in close proximity to coastline? Yes
Do launch sites keep certain distance away from cities? Yes

Image One

In order to visually represent the launch sites, latitude and longitude coordinates were collected and utilized to generate circular markers around each site. Each marker was accompanied by a label indicating the name of the corresponding launch site

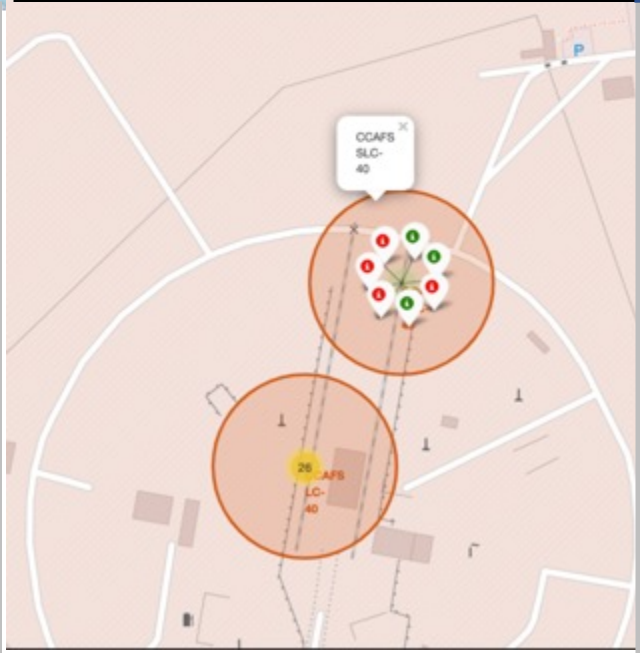


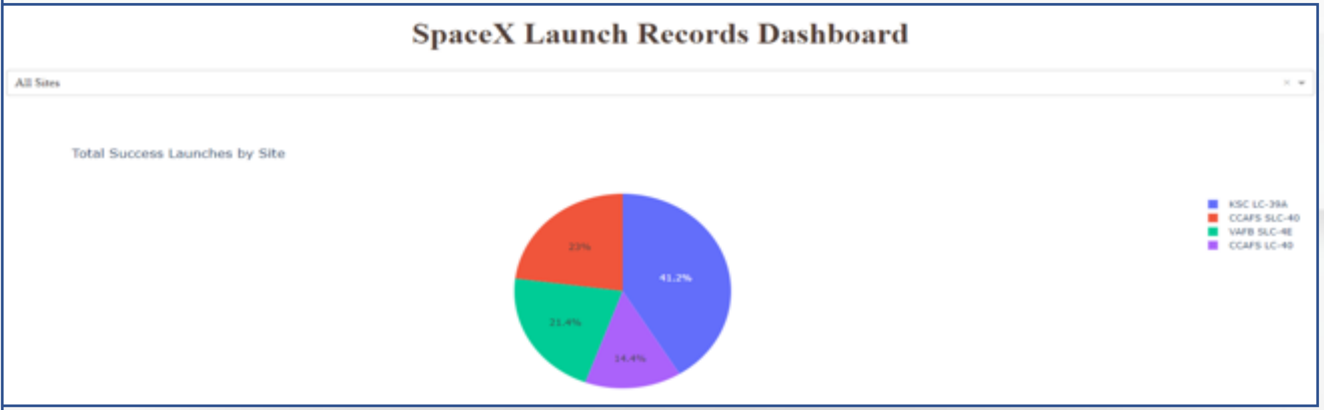
Image Three

In order to gain insights into the surrounding environment of the launch sites, we employed Haversine's formula to calculate the distances to various landmarks. These calculations allowed us to identify trends and patterns in the surrounding area. To visually represent these findings, lines were drawn on the map to illustrate the distances to the corresponding landmarks

DASHBOARD - PLOTLY

Improve user experience by providing a dropdown list of launch sites, allowing selection of all sites or a specific one

Visualize launch success rates with a pie chart for easy comparison of successful and unsuccessful launches to the total:



Add a slider for payload mass range selection to allow the user to adjust and choose the desired range

GitHub

PREDICTIVE ANALYTICS

1 Building Model

Constructed a NumPy array utilizing the column labeled "class”

2 Building Model

Utilizing the Standard Scaler method, both fitting and transforming of the data

3 Building Model

Partitioning the Data into Training and Testing Sets

4 Building Model

Ran logistic regression with GridSearchCV, showing optimized parameters and validation accuracy with 'best params' and 'best score'

5 Evaluating Model

Computed the accuracy score of the test data utilizing the 'score' method

6 Evaluating Model

Obtained optimized hyperparameters for each algorithmic type

7 Evaluating Model

Generate a graphical representation of the Confusion Matrix

8 Improving Model

Conduct Feature Engineering and Refine Algorithmic Models via Tuning

9 Best Model

Determine the Optimal Model based on Jaccard Score, F1 Score, and Accuracy Metrics

GitHub

RESULTS SUMMARY

Exploratory Data Analysis (EDA)

The success rate of launches has shown an upward trend with the passage of time

KSC LC-39A is the landing site with the highest success rate, indicating its reliability

Orbits ES-L1, GEO, HEO, and SSO stand out with a 100% success rate, showcasing their exceptional performance

Visual Analytics

Most of the launch sites are situated in close proximity to the equator, with all being located near the coast

The launch sites are carefully chosen to ensure that a failed launch does not cause damage to any nearby structures, such as cities, highways, and railways

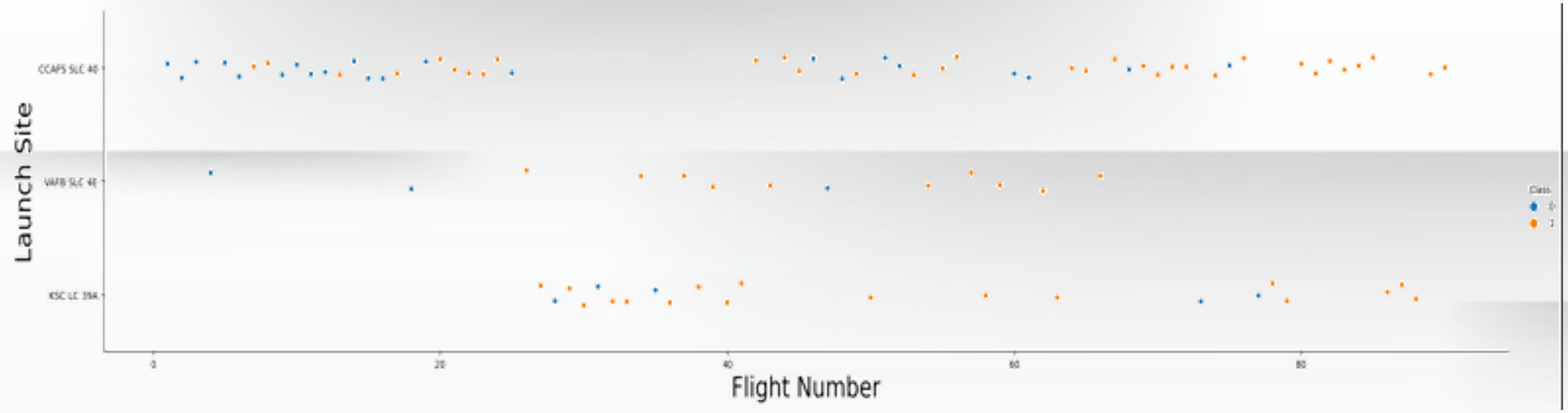
However, they are strategically positioned to facilitate the transportation of essential resources and personnel required for successful launch

Predictive Analytics

Based on our analysis, the Decision Tree model has been identified as the optimal predictive model for the given dataset


EDA DISCOVERIES

FLIGHT NUMBER VS LAUNCH SITE



Previous flights experienced a lower success rate, as indicated by the blue marking for failure

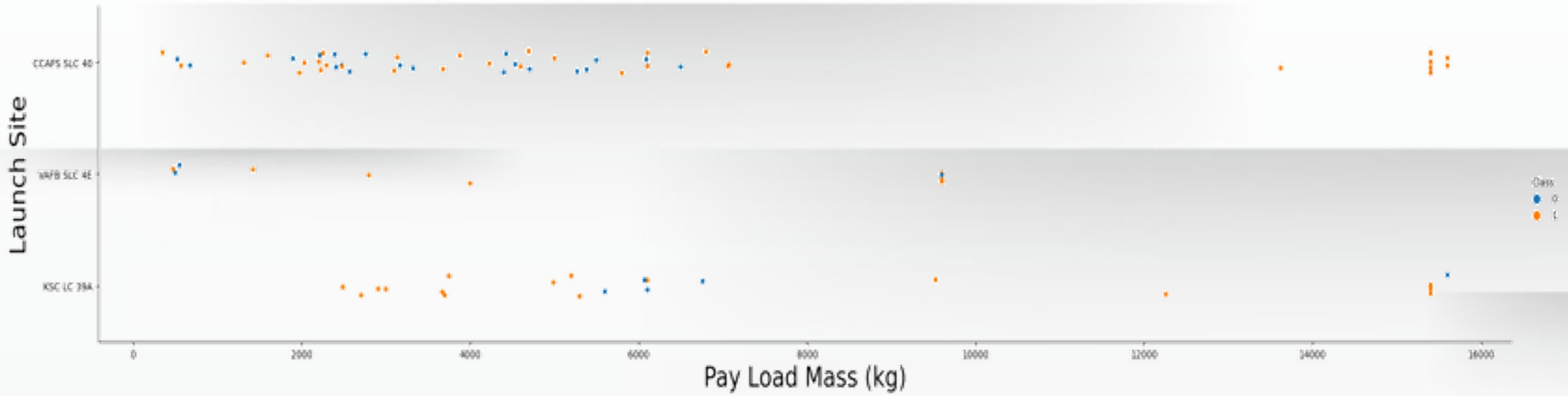
In contrast, subsequent flights demonstrated an increased success rate, as indicated by the orange marking for success

Approximately 50% of all launches originated from the CCAFS SLC 40 launch site

Launches from the VAFB SLC 4E and KSC LC 39A sites exhibited a higher success rate compared to others

Based on our observations, we can infer that newer launches are more likely to achieve success

PAYLOAD VS LAUNCH SITE



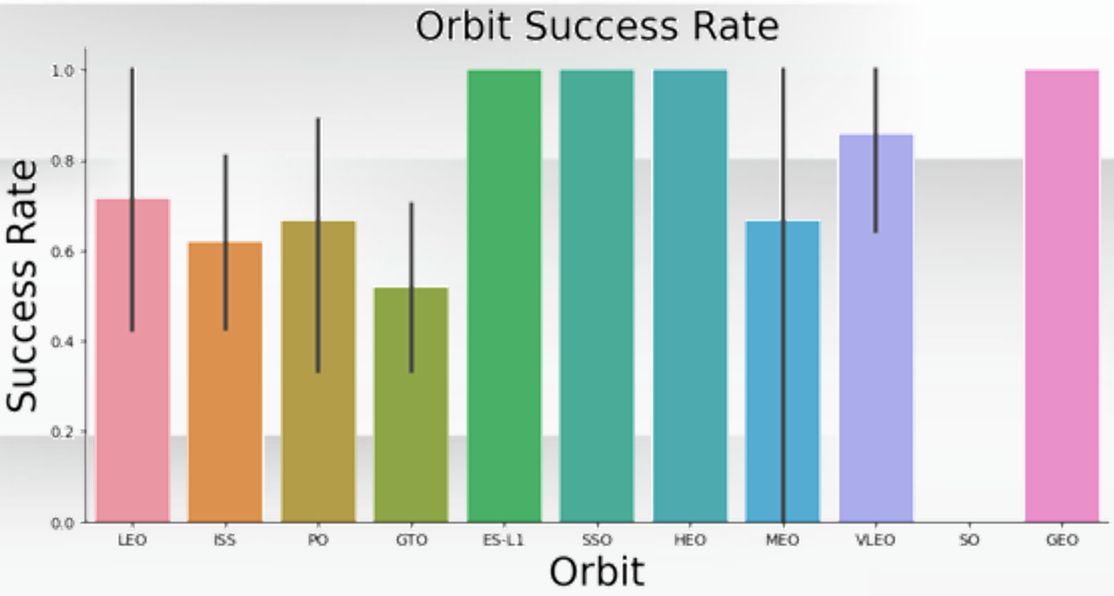
In general, there is a positive correlation between the mass of a payload (measured in kilograms) and the likelihood of a successful launch

The majority of launches that carried payloads weighing more than 7,000 kg experienced successful outcomes

Launches from the KSC LC 39A site with payloads less than 5,500 kg have achieved a 100% success rate

To date, the VAFB SKC 4E launch site has not conducted a launch with a payload exceeding approximately 10,000 kg

SUCCESS RATE VS ORBIT

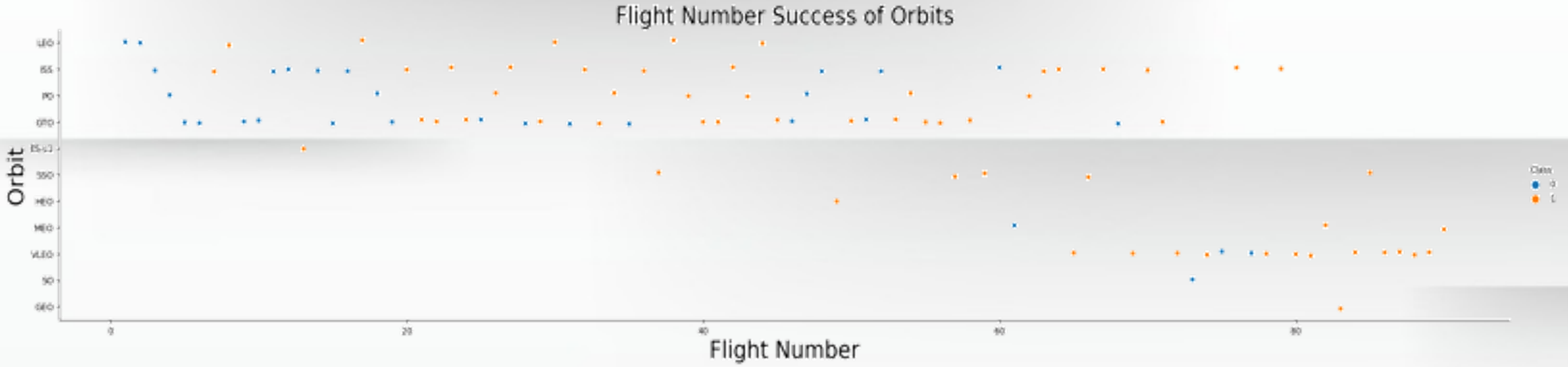


100% success rate:
ES-L1 GEO HEO SSO

50% - 80% success rate:
GTO ISS LEO MEO PO

0% success rate:
SO

FLIGHT NUMBER VS ORBIT

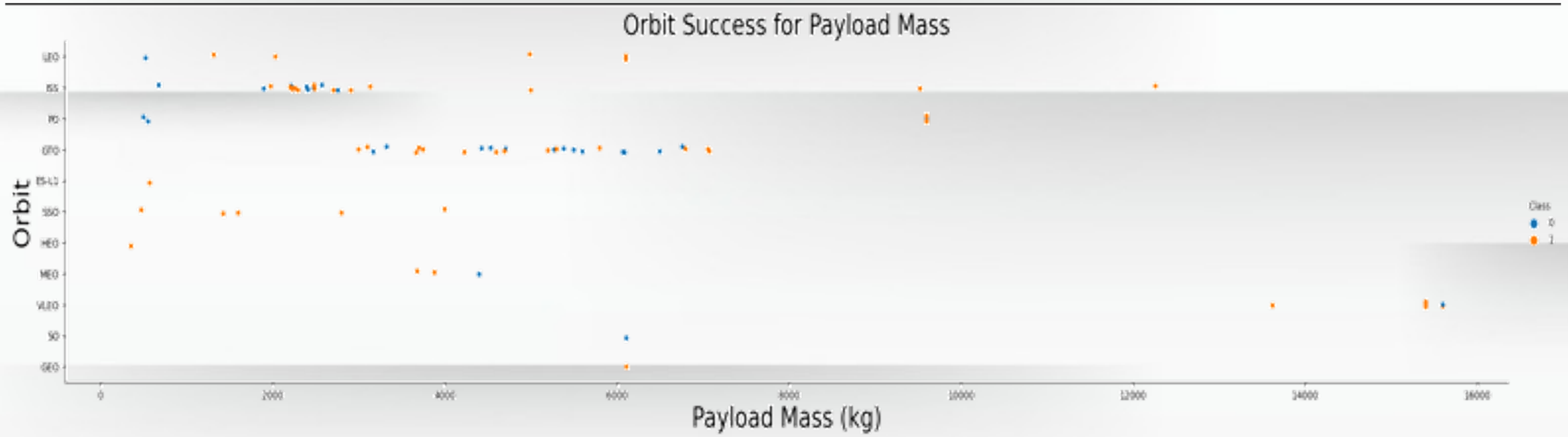


Generally, as the number of flights for a given orbit increases, there is a corresponding increase in the success rate of launches

This trend is particularly pronounced for the Low Earth Orbit (LEO)

However, the Geostationary Transfer Orbit (GTO) does not conform to this pattern, as the success rate is not significantly impacted by the number of flights

PAYLOAD VS ORBIT



Heavy payloads typically exhibit a higher success rate for launches targeting the Low Earth Orbit (LEO), International Space Station (ISS), and Polar Orbit (PO) trajectories

In contrast, the Geostationary Transfer Orbit (GTO) has demonstrated a mixed success rate for launches carrying heavier payloads

Launches carrying heavy payloads to the LEO, ISS, and PO orbits require precise launch vehicle capabilities to achieve optimal success rates, while launches to the GTO orbit may benefit from additional technological advancements to improve outcomes for heavier payloads

LAUNCH SUCCESS YEARLY TREND

The success rate of launches demonstrated notable improvement during two distinct periods: from 2013-2017 and again from 2018-2019

Conversely, the success rate exhibited a decrease from 2017-2018 and again from 2019-2020

In general, the success rate of launches has demonstrated a positive trend since 2013

Launch Site Names

Launch Site Names

CCAFS LC-40 - Cape Canaveral Air Force Station Launch Complex 40

CCAFS SLC-40 - Cape Canaveral Air Force Station Space Launch Complex 40

KSC LC-39A - Kennedy Space Center Launch Complex 39A

VAFB SLC-4E - Vandenberg Air Force Base Space Launch Complex 4E

```
%sql \
SELECT UNIQUE(LAUNCH_SITE) as "Launch Sites" \
FROM SPACEX;
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

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

LAUNCH SITE NAMES START WITH 'CCA'

In our data retrieval efforts, we have successfully obtained a set of 5 records that meet the criteria of having launch sites that begin with the characters 'CCA'. This outcome reflects the effectiveness of our data gathering methods and provides us with valuable information for our research and analysis

```
%sql \
SELECT * \
FROM SPACEX \
WHERE LAUNCH_SITE \
LIKE'CCA%' \
LIMIT 5;
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

DATE	time_utc	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

TOTAL PAYLOAD MASS

The Total Payload Mass

Nasa: 45,596 kg

The NASA Commercial Resupply Services (CRS) program launched a total payload mass of 45,596 kg to the International Space Station (ISS)

```
%sql \
SELECT SUM(PAYLOAD_MASS_KG_) as "Total Payload Mass" \
FROM SPACEX WHERE CUSTOMER = 'NASA (CRS)';
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

Total Payload Mass
45596

AVERAGE PAYLOAD MASS BY F9 V1.1

The Average Payload Mass:

SpaceX: 2,928 kg

The booster version F9 v1.1 had an average payload mass of 2,928 kg. This booster was a previous version of the Falcon 9 rocket, which is designed and manufactured by SpaceX to transport payloads to various orbits

```
%sql \
SELECT AVG(PAYLOAD_MASS_KG_) as "Average Payload Mass" \
FROM SPACEX WHERE BOOSTER_VERSION = 'F9 v1.1';

 * ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

Average Payload Mass
2928

FIRST GROUND LANDING DATE

First Successful Landing On Ground Pad:

SpaceX made history on December 22nd, 2015, by achieving the first successful rocket landing on a ground pad, a significant milestone in the development of reusable rockets

```
%sql \
SELECT MIN(DATE) as "First Successful Outcome In Ground Pad" \
FROM SPACEX WHERE LANDING_OUTCOME = 'Success (ground pad)';

 * ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

First Successful Outcome In Ground Pad
2015-12-22

Drone Ship Landing with Payload 4,000kg - 6,000kg

Booster Drone Ship Landing:

On December 22nd, 2015, SpaceX successfully landed a booster weighing between 4,000 and 6,000 kilograms on a drone ship, carrying payloads for JSCAT-14, JSCAT-16, SES-10, SES-11, and EchoStar 105 satellites

```
%sql \
SELECT BOOSTER_VERSION as "Successful Landing Boosters Between 4000 & 6000 Payload" \
FROM SPACEX WHERE LANDING_OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000;
```

```
 * ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

Successful Landing Boosters Between 4000 & 6000 Payload
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

MISSION OUTCOMES: SUCCESS / FAILUER

Total Number Of Successful And Failed Mission Outcomes:

99 Successful Landings
1 Success (payload status unclear)
1 Failure In Flight

```
%sql \
SELECT MISSION_OUTCOME as "Mission Outcome", \
COUNT(*) as "Total Number" \
FROM SPACEX \
GROUP BY MISSION_OUTCOME;
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

Mission Outcome	Total Number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

BOOSTERS CARRYING MAX PAYLOAD

Names of boosters which have max carrying payload mass using a subquery

SpaceX Falcon 9 Block 5 B1048.4	SpaceX Falcon 9 Block 5 B1048.5	SpaceX Falcon 9 Block 5 B1058.3
SpaceX Falcon 9 Block 5 B1049.4	SpaceX Falcon 9 Block 5 B1051.4	SpaceX Falcon 9 Block 5 B1051.6
SpaceX Falcon 9 Block 5 B1051.3	SpaceX Falcon 9 Block 5 B1049.5	SpaceX Falcon 9 Block 5 B1060.3
SpaceX Falcon 9 Block 5 B1056.4	SpaceX Falcon 9 Block 5 B1060.2	SpaceX Falcon 9 Block 5 B1049.7

```
%sql \
SELECT BOOSTER_VERSION as "Booster Version", \
PAYLOAD_MASS_KG_ as "Max Payload Mass" \
FROM SPACEX \
WHERE PAYLOAD_MASS_KG_ = (SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEX)
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

Booster Version	Max Payload Mass
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 LAUNCH RECORDS

Date / Booster Version	Launch Site	Landing Outcome
10-01-2015 F9 v1.1 B1012	CCAFS LC-40	Failure (Drone Ship)
14-04-2015 F9 v1.1 B1015	CCAFS LC-40	Failure (Drone Ship)

```
%sql \
SELECT DATE, \
BOOSTER_VERSION as "Booster Version", LAUNCH_SITE as "Launch Site", LANDING_OUTCOME as "Landing Outcome" \
FROM SPACEX \
WHERE LANDING_OUTCOME = 'Failure (drone ship)' \
AND SUBSTR(DATE,1,4) = 2015;
```

```
* ibm_db_sa://jxk70787:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

DATE	Booster Version	Launch Site	Landing Outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

LANDING OUTCOMES 2010 - 2017

The descending order of the count of landing outcomes recorded during the period spanning from June 4th, 2010 to March 20th, 2017 is to be presented

Success	20
No Attempt	10
Success (Drone Ship)	8
Success (Ground Pad)	6
Failure (Drone Ship)	4
Failure	3
Controlled (Ocean)	3
Failure (Parachute)	2
No Attempt	1

```
%sql SELECT LANDING_OUTCOME "Landing Outcome", COUNT(*) as "Count Outcomes" FROM SPACEX WHERE DATE BETWEEN '2010-06-04' and '2017-03-20' group by LANDING_OUTCOME order by "Count Outcomes" DESC;
```

```
 * ibm_db_sa://vhb27331:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/BLUDB
Done.
```

Landing Outcome	Count Outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1


LAUNCH SITE ANALYSIS

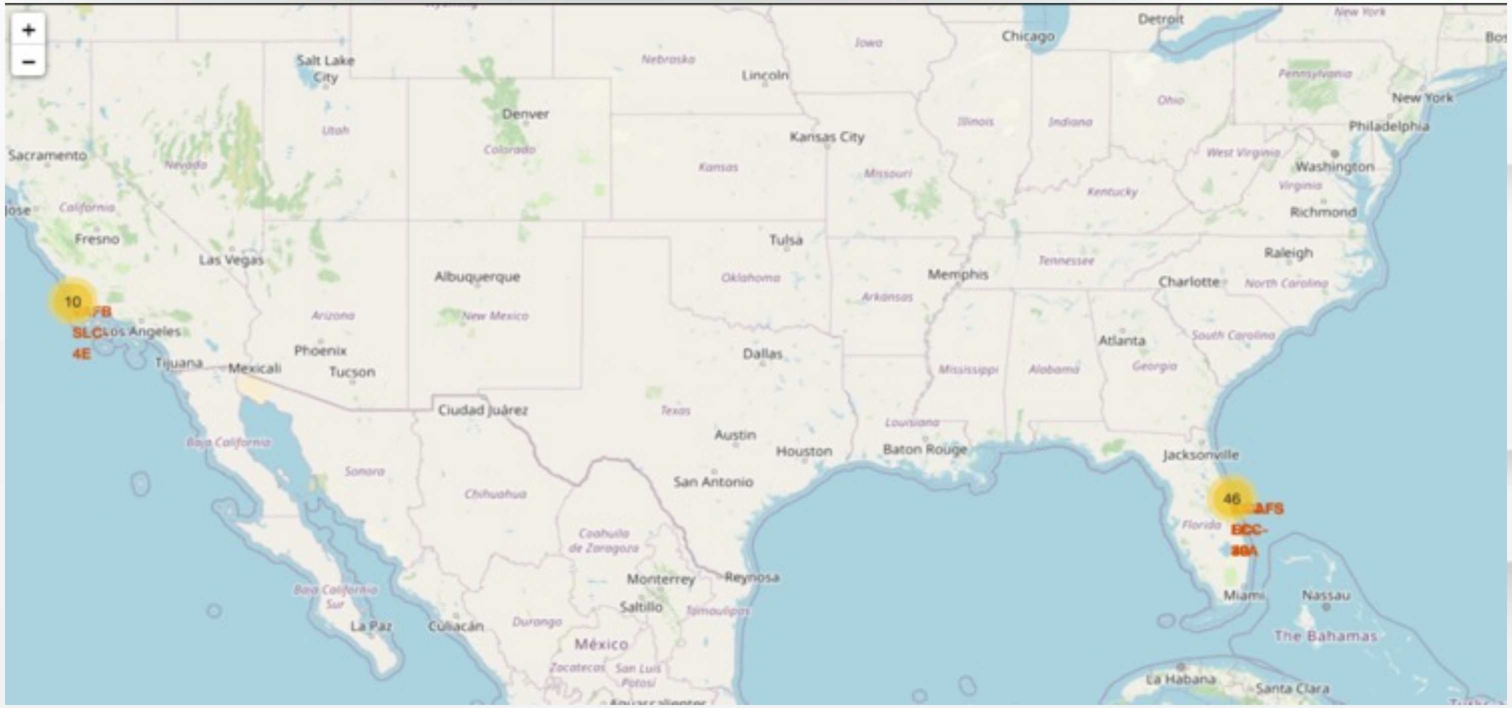
UNITED STATES LAUNCH SITE MAP

The proximity of a launch site to the equator is a crucial factor in optimizing the trajectory for launching a rocket to an equatorial orbit

The Earth's rotational speed provides an additional natural boost, enhancing the efficiency of launching a prograde orbit and achieving an equatorial orbit

Launching from sites situated closer to the equator can significantly reduce the cost associated with additional fuel consumption and the use of booster rockets to attain the desired orbit

Rockets launched from sites near the equator can take advantage of the natural boost provided by the Earth's rotation, thereby improving their overall performance and reducing operational costs

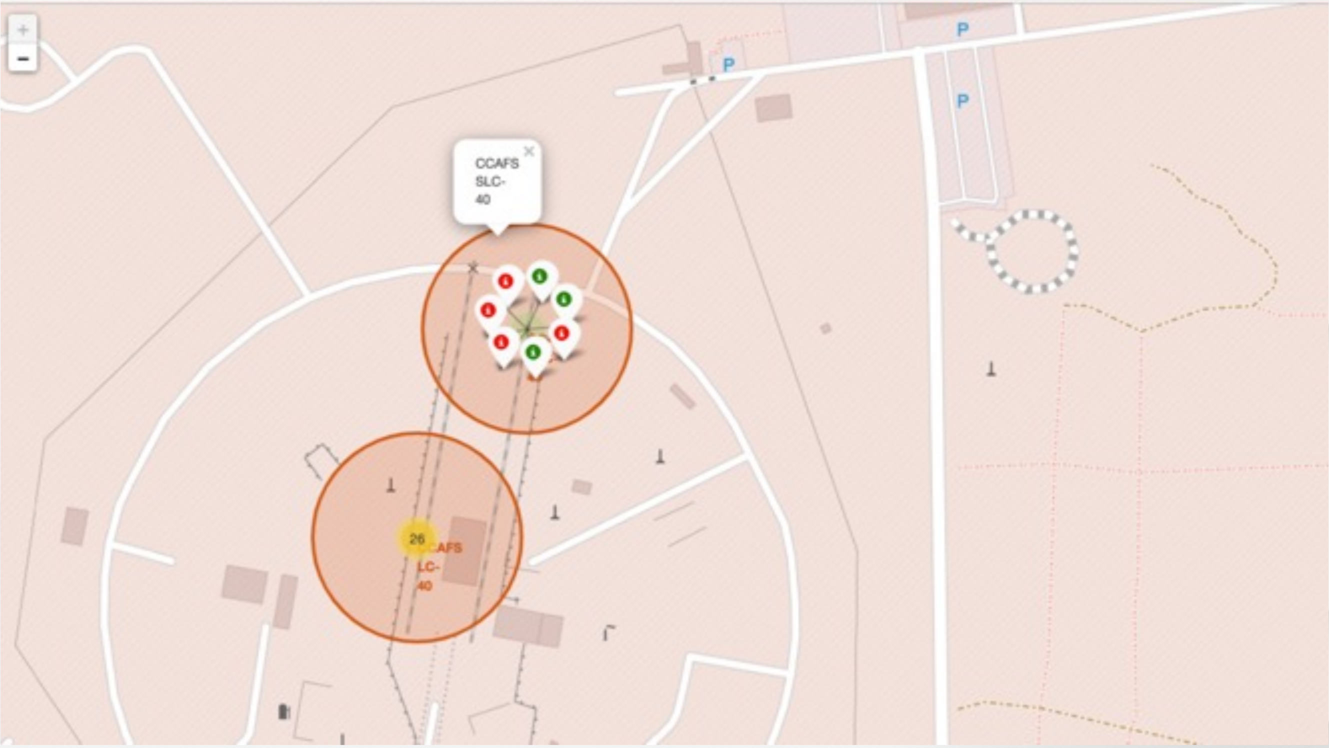


SITE LAUNCH RESULTS MAP

Successful launches are identified by green markers

Unsuccessful launches are denoted by red markers

Launch site CCAFS SLC-40 has a success rate of 3/7 (42.9%)



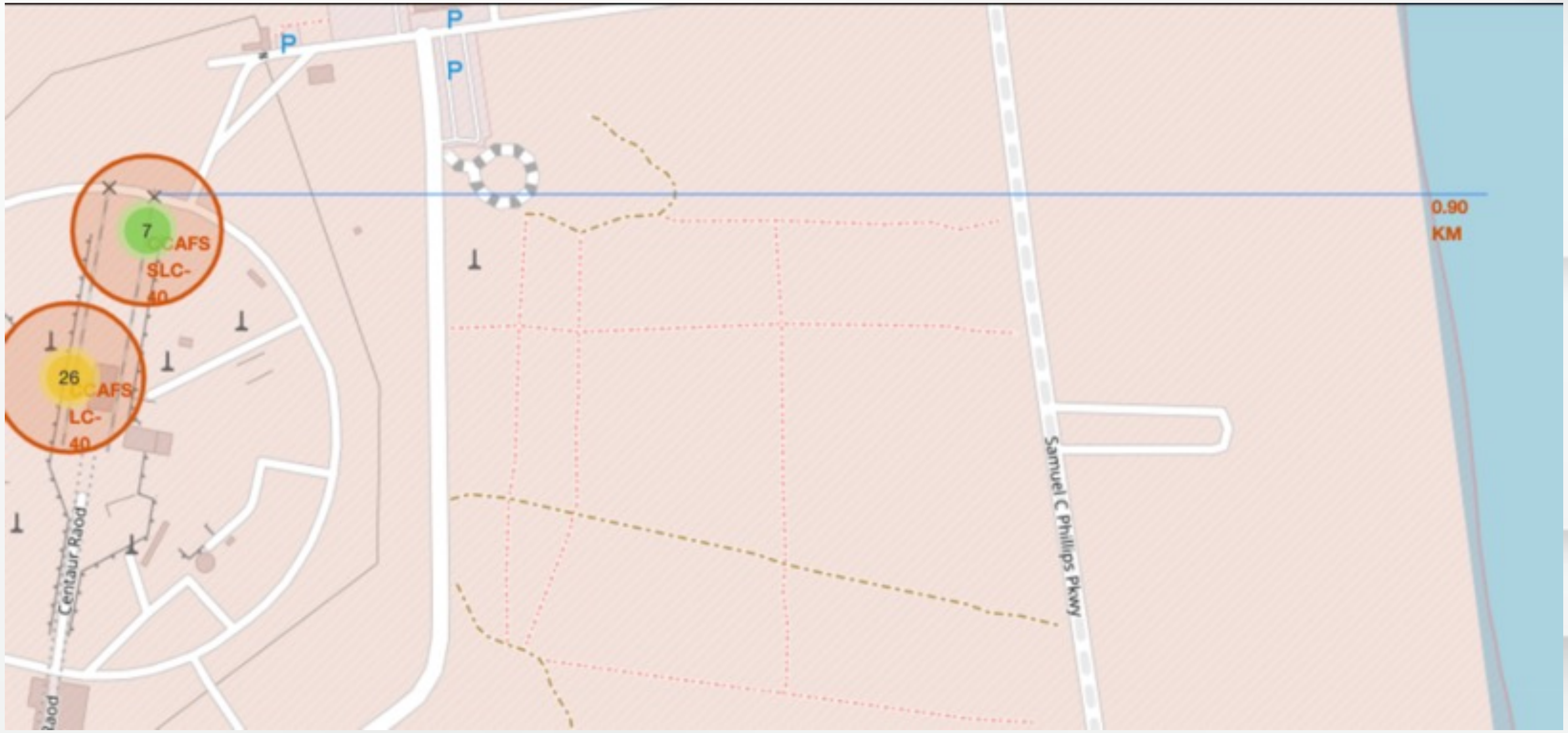
LAUNCH SITE PROXIMITIES MAP

Launch site CCAFS SLC-40 is 0.90km from the coast for easy transportation via waterways

Nearest railway is 21.96km away for minimal train traffic interference during launches

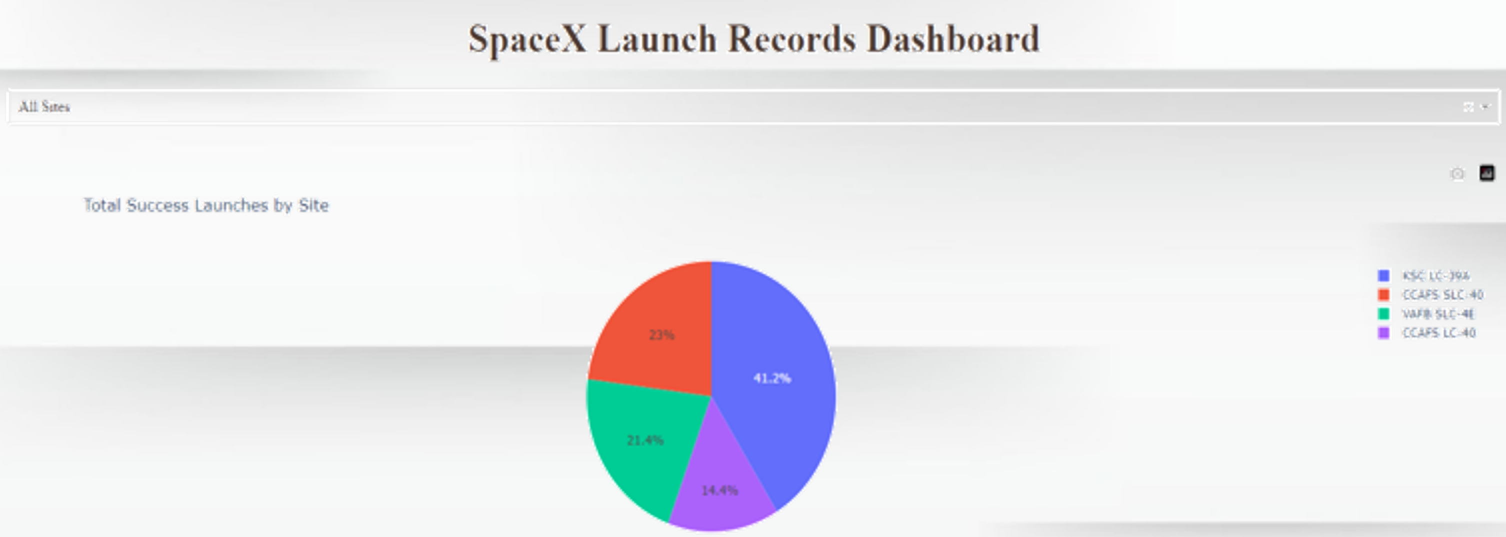
CCAFS SLC-40 is 23.23km from the nearest city for reduced noise and safety concerns while allowing easy access to the site

26.88km from the nearest highway to minimize vehicular traffic disruption during launches



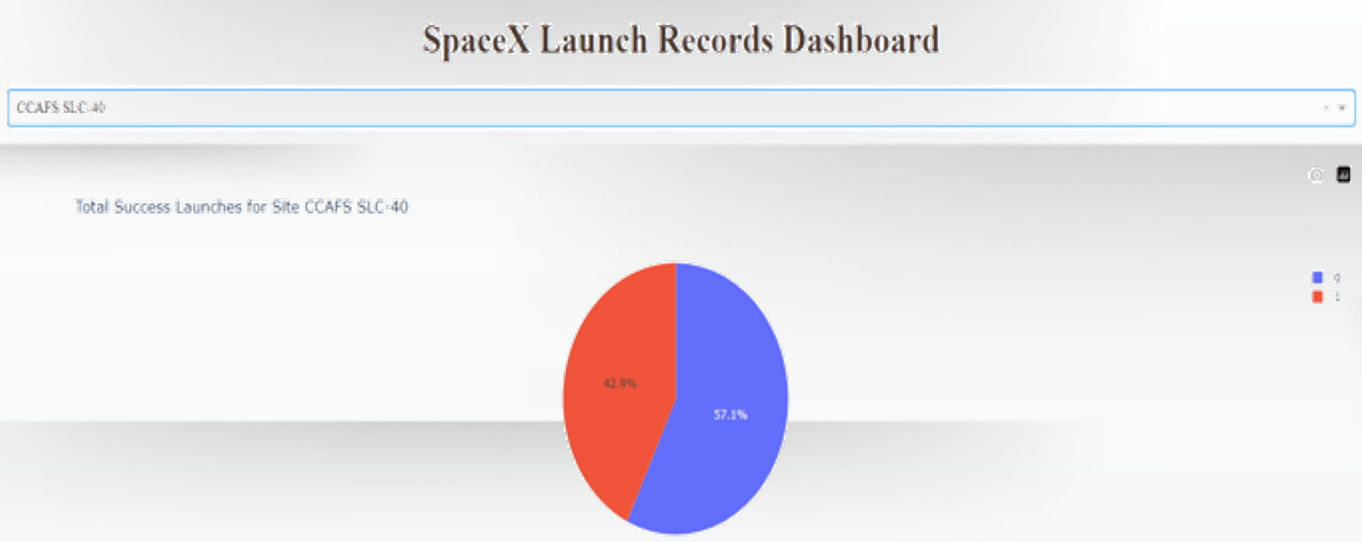

PLOTLY DASHBOARD

SUCCESSFUL LAUNCH OF ALL SITES



The success rate of rocket launches is a critical performance metric for any launch site. Here are some insights about the "Success as a Percentage of Total" for various launch sites:	KSC LC-39A has the highest success rate at 41.2%, proving the team's technical expertise and operational excellence in safe launches.

HIGHEST SUCCESS LAUNCH SITE



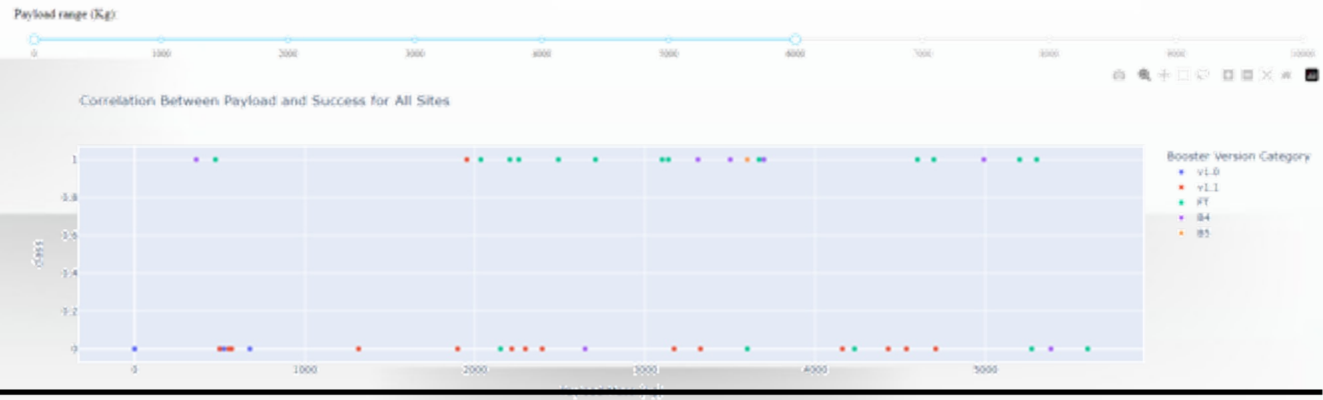
Rocket launch success rates are critical for launch sites. See "Success as a Percentage of Total" for different sites:

KSC LC-39A had 10 successful launches out of 13, proving its ability for reliable and safe launches

KSC LC-39A has a success rate of 76.9%, showcasing its excellence in technical capabilities, safety protocols, and operational efficiency

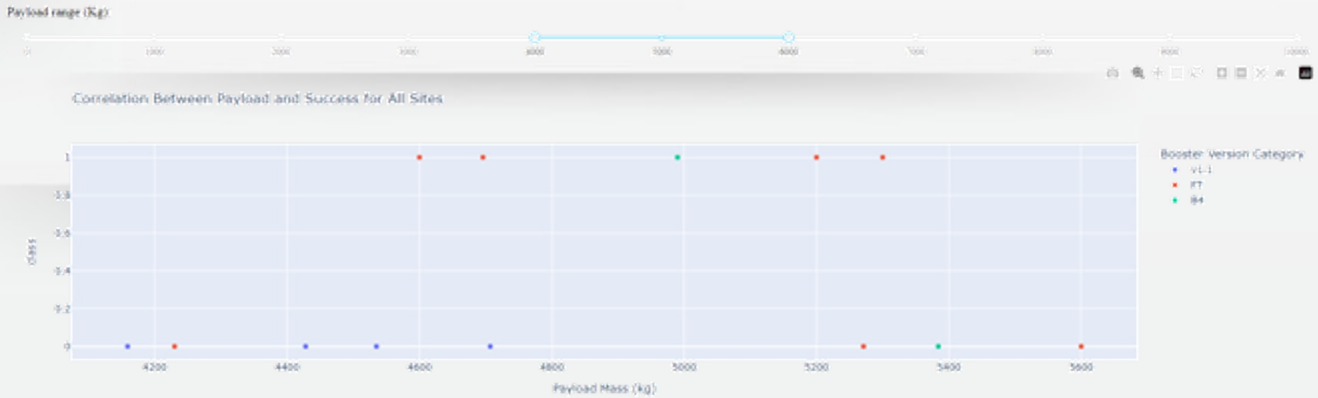
PAYLOAD VS LAUNCH SCATTER PLOT

The analysis of rocket launches by booster version is a crucial exercise to gain insights into the performance and reliability of rocket boosters for different payload categories. Here are some observations related to the "By Booster Version" analysis:



Rocket launches by booster version indicate that payloads in the 2,000-5,000 kg range have the highest success rate, implying that the rocket boosters used for this payload category are the most reliable and efficient

The outcome of each rocket launch is indicated by a binary system, with 1 representing a successful outcome and 0 representing an unsuccessful outcome



PREDICTIVE ANALYSIS

CLASSIFICATION ACCURACY

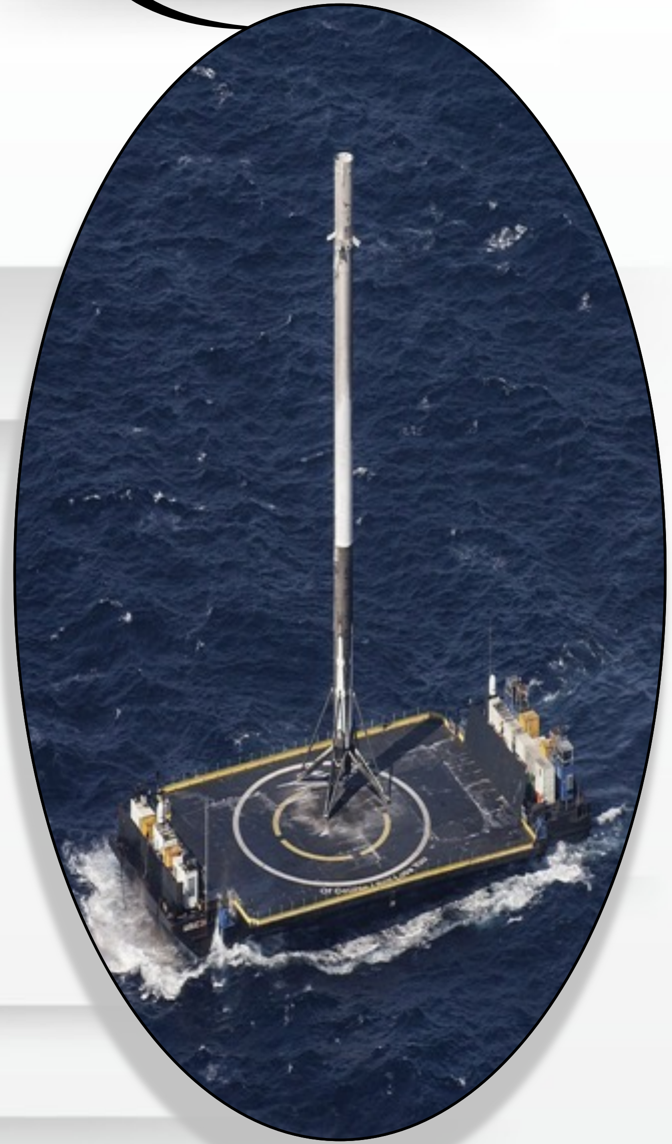
The models exhibited comparable performance and accuracy, likely attributed to the limited dataset

However, the Decision Tree model demonstrated slightly better performance in terms of best score, which represents the mean score of all cross-validation folds for a particular parameter combination

	LogReg	SVM	Tree	KNN
Jaccard Score	0.800000	0.800000	0.800000	0.800000
F1 Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

```
models = {'KNeighbors':knn_cv.best_score_,
          'DecisionTree':tree_cv.best_score_,
          'LogisticRegression':logreg_cv.best_score_,
          'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)

print('Best model: ', bestalgorithm,'. Score:', models[bestalgorithm])

if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)

if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)

if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)

if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model:  DecisionTree . Score: 0.8732142857142857
Best params is : {'criterion': 'entropy', 'max_depth': 10, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 10, 'splitte
r': 'random'}
```



CONFUSION MATRIX

Martix Equations

Precision is calculated as TP / (TP + FP), yielding 12 / 15 = .80

Recall is calculated as TP / (TP + FN), yielding 12 / 12 = 1

F1 Score is calculated as 2 * (Precision * Recall) / (Precision + Recall), yielding 2 * (.8 * 1) / (.8 + 1) = .89

Accuracy is calculated as (TP + TN) / (TP + TN + FP + FN) = .833.

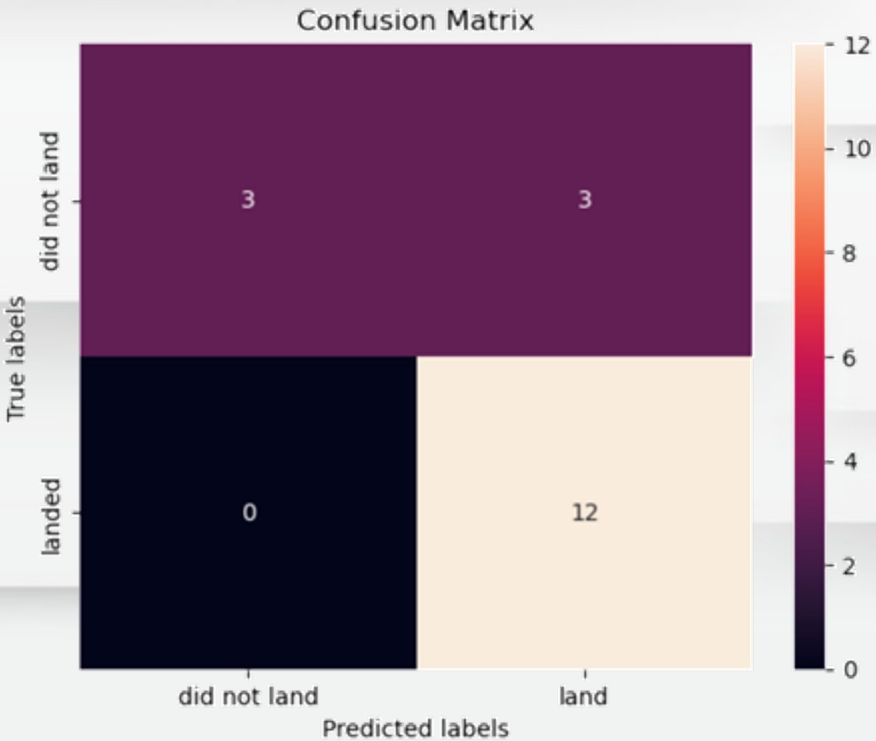
Performance Summary

A confusion matrix provides a comprehensive summary of a classification algorithm's performance

All confusion matrices yielded identical results

The existence of false positives (Type 1 error) is not ideal.

Confusion Matrix Outputs: 12 True Positive, 3 True Negative, 3 False Positive, 0 False Negative



CONCLUSION

Launch Success	Improved technology and operations lead to more successful rocket launches
KSC LC-39A	KSC LC-39A has 100% success rate for launches under 5,500 kg, highlighting their excellence in technology and operations
Orbital Success	100% success rate for launches to ES-L1, GEO, HEO, and SSO orbits demonstrates excellent booster performance
Model Performance	Models had similar performance on the test set, with decision tree slightly better
Proximity to Equtor	Equatorial launch sites save fuel and booster costs due to the earth's rotational speed
Coastal Location	Launch sites near coasts ease component and personnel transportation
Payload Mass	Higher payload masses correlate with higher success rates across launch sites

RECOMMENDATIONS

Increase dataset: Expand dataset and include alternative sources to improve accuracy and generalizability of findings

Evaluate feature analysis: Conduct additional feature analysis, like principal component analysis, to enhance model accuracy

Test XGBoost: Determine the effectiveness of the powerful XGBoost model, unexplored in this study, to outperform other models

Utilize location data: Utilize launch site location data to improve the accuracy of predictions for future launches

Continuously refine models: Update datasets and models regularly, apply them to new launches like Falcon Heavy and others, and explore new tools and models to stay up-to-date with computing developments and improve results


IBM
SPACEX
ROCKET
RESUABILITY