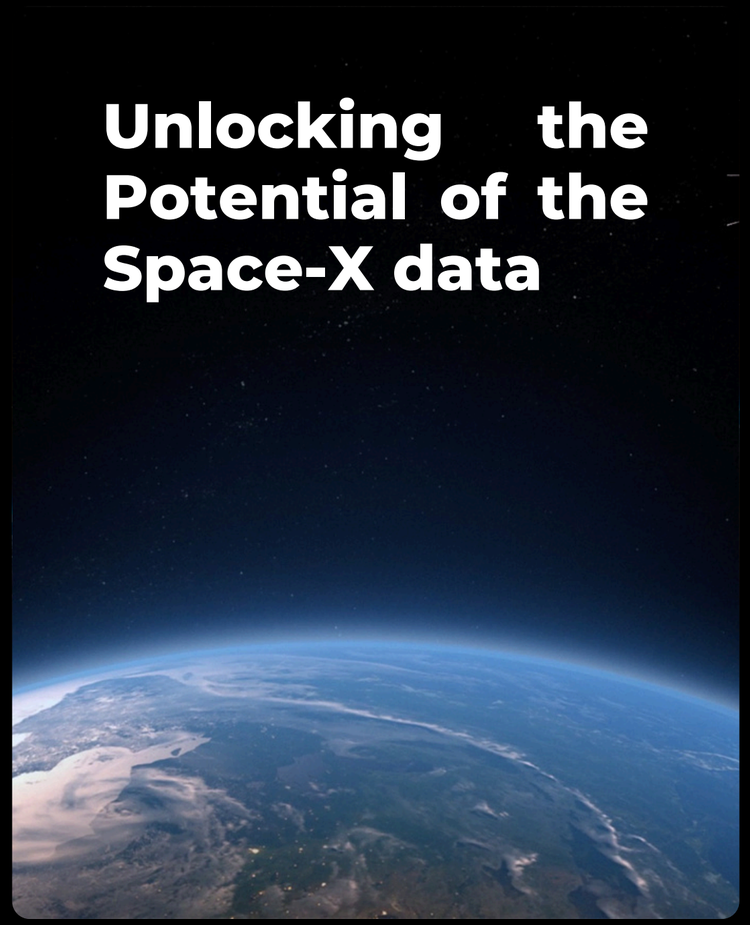
# NAVIGATING THE COSMOS WITH DATA SCIENCE

15-02-2025

MOISES AHUMADA - MOISES.MAT.CUCEI@GMAIL.COM

# Unlocking the Potential of the Space-X data



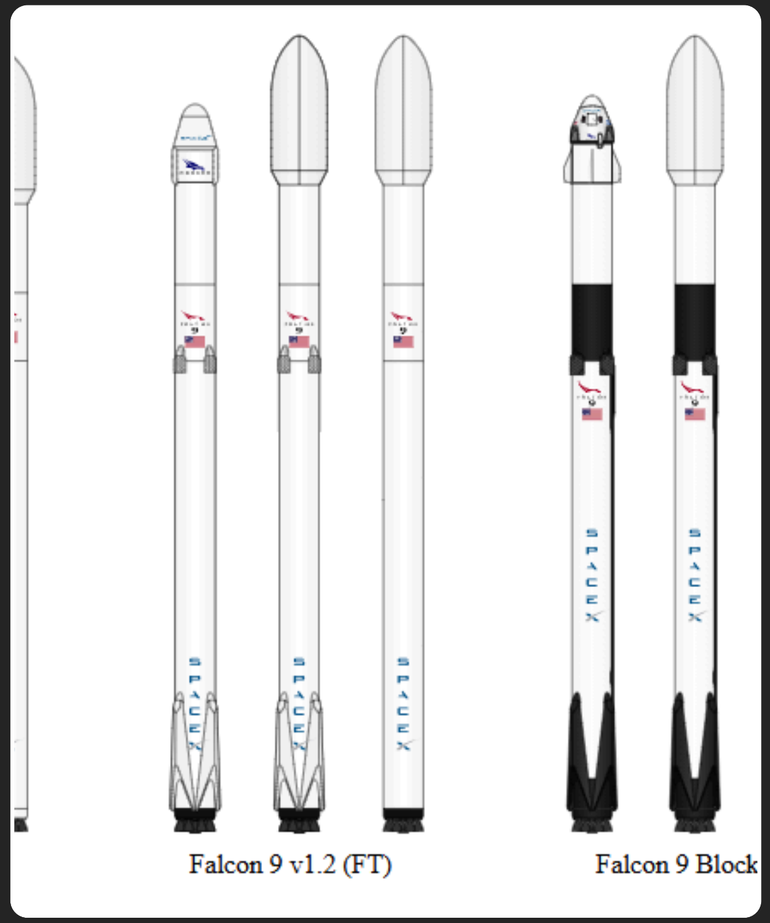
## ABSTRACT

This analysis explores SpaceX launch data using Python to uncover key insights into mission success rates, launch site performance, and payload impact. Using Pandas, Matplotlib, and others, we visualize trends and correlations, providing a data-driven perspective on SpaceX's launch efficiency.

# INTRODUCTION

The analysis of space launch data has become essential for understanding mission success factors and optimizing future launches. SpaceX, a leader in commercial spaceflight, provides valuable data for exploring trends in launch performance, payload impact, and site efficiency. Using Python and data analysis tools, this study uncovers key insights, contributing to the broader understanding of space mission analytics.

NEXT SLIDE



# DATA COLLECTION AND WRANGLING

We will connect to the SpaceX API to retrieve historical launch data from the endpoint:
🔗 https://api.spacexdata.com/v4/launches/past

The data, originally in JSON format, will be converted into a Pandas DataFrame. We will then clean and structure it to ensure consistency, removing irrelevant fields and preparing it for analysis, such as identifying launch trends and success patterns.

| | FlightNumber | Date | BoosterVersion | PayloadMass | Orbit | LaunchSite | Outcome | Flights | GridFins | Reused | Legs | LandingPad | Block | ReusedCount | Serial | Longitude | Latitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **4** | 1 | 2010-06-04 | Falcon 9 | 6123.547647 | LEO | CCSFS SLC 40 | None None | 1 | False | False | False | None | 1.0 | 0 | B0003 | -80.577366 | 28.561857 |
| **5** | 2 | 2012-05-22 | Falcon 9 | 525.000000 | LEO | CCSFS SLC 40 | None None | 1 | False | False | False | None | 1.0 | 0 | B0005 | -80.577366 | 28.561857 |
| **6** | 3 | 2013-03-01 | Falcon 9 | 677.000000 | ISS | CCSFS SLC 40 | None None | 1 | False | False | False | None | 1.0 | 0 | B0007 | -80.577366 | 28.561857 |
| **7** | 4 | 2013-09-29 | Falcon 9 | 500.000000 | PO | VAFB SLC 4E | False Ocean | 1 | False | False | False | None | 1.0 | 0 | B1003 | -120.610829 | 34.632093 |
| **8** | 5 | 2013-12-03 | Falcon 9 | 3170.000000 | GTO | CCSFS SLC 40 | None None | 1 | False | False | False | None | 1.0 | 0 | B1004 | -80.577366 | 28.561857 |

# UNCOVERING DATA INSIGHTS: EXPLORING EDA WITH SQL

## By using SQLite, we can:

- Perform complex queries more efficiently than in-memory Pandas operations.
- Easily filter, aggregate, and join data for deeper exploratory analysis.
- Store results persistently for reproducibility and scalability.

This query identifies which booster versions successfully landed on a drone ship while carrying payloads between 4,000 and 6,000 kg.
By filtering for successful landings and specific payload ranges, we gain insights into reliable boosters for medium-weight missions, helping optimize future launch planning and performance analysis.

```
%sql select distinct Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ between 4000 and 6000 and Landing_Outcome = 'Success (drone ship)'
```

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

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

SpaceX has launched 107,010 kg of payload for NASA (CRS) missions, highlighting NASA's reliance on SpaceX for cargo deliveries.

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer like '%NASA%'
```

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

| sum(PAYLOAD_MASS__KG_) |
|---|
| 107010 |

The F9 v1.1 booster carried an average payload mass of 2,928.4 kg per launch. This provides insights into its capacity and performance in past missions.

**Display average payload mass carried by booster version F9 v1.1**

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1'
```

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

| avg(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

This analysis counts successful and failed missions for SpaceX launches.

- 100 missions were successful (flag = 1).
- 1 mission failed (flag = 0).

This provides a success rate of nearly 99%, highlighting SpaceX's strong reliability in launch outcomes.

**List the total number of successful and failure mission outcomes**

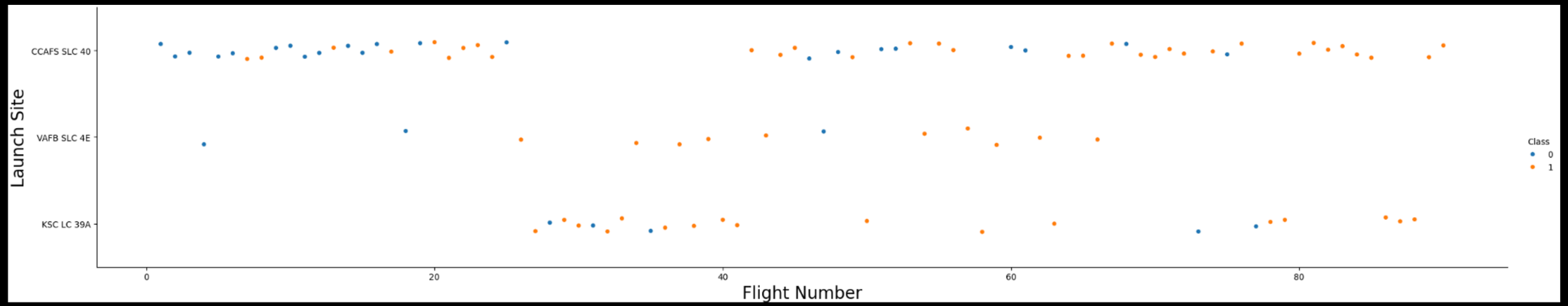
```
#%sql select distinct Mission_Outcome from SPACEXTABLE
%sql select flag, count(*) as totales from (select case when Mission_Outcome like '%Success%' then 1 else 0 end as flag from SPACEXTABLE) group by flag
```

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

| flag | totales |
|---|---|
| 0 | 1 |
| 1 | 100 |

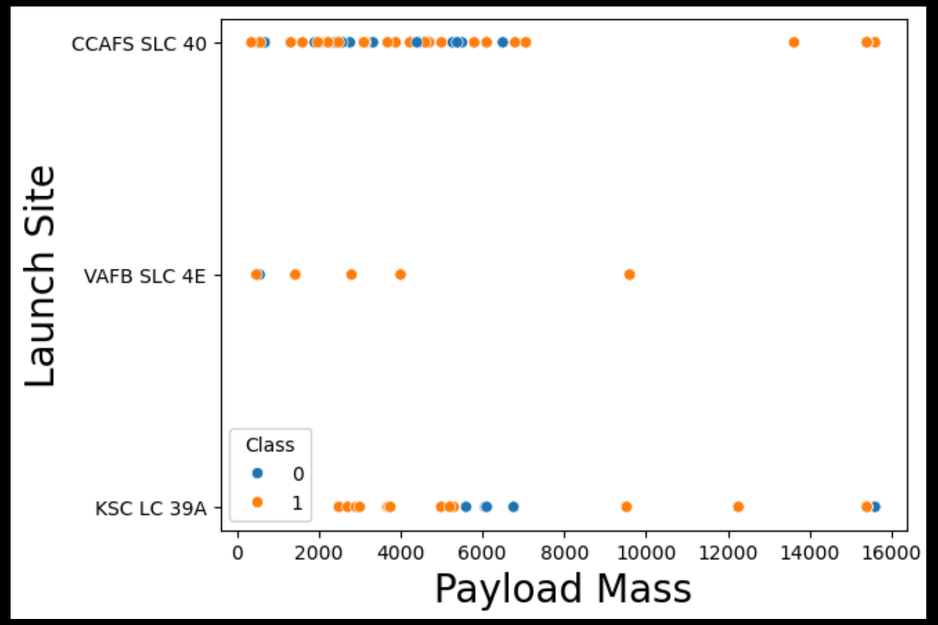
# EDA: UNVEILING INSIGHTS THROUGH VISUALIZATIONS

# THE RELATIONSHIP BETWEEN FLIGHT NUMBER AND LAUNCH SITE




The scatter plot shows Flight Number vs. Launch Site, with color indicating success (1) or failure (0). As flight numbers increase, success rates improve. Different launch sites have varying success distributions.

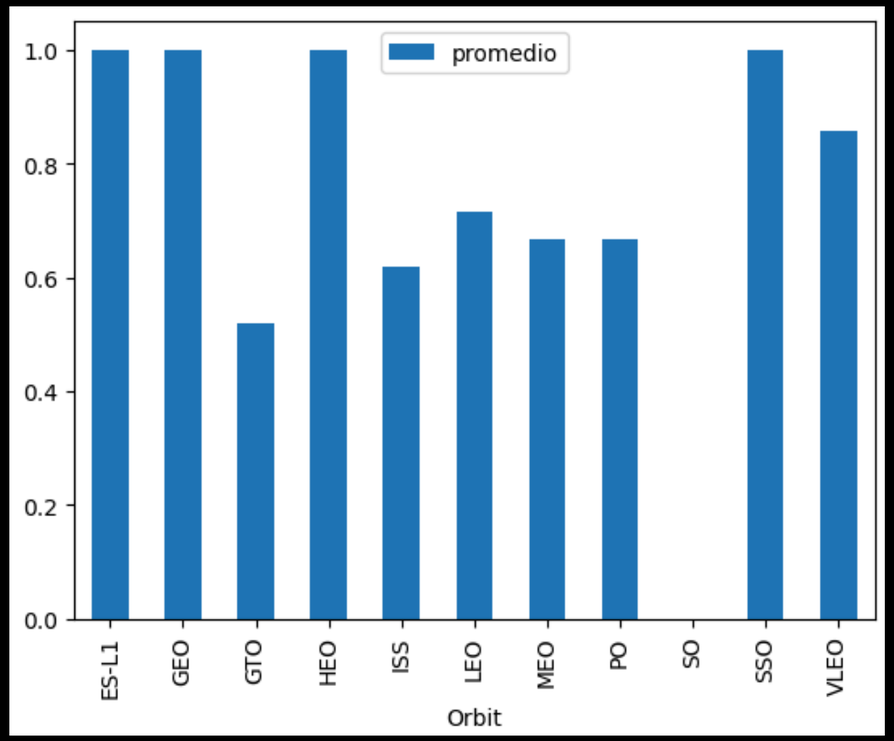
# THE RELATIONSHIP BETWEEN PAYLOAD MASS AND LAUNCH SITE




The plot shows Payload Mass vs. Launch Site, with color indicating success (1) or failure (0).

- VAFB SLC 4E has no launches for payloads above 10,000 kg.
  CCAFS SLC 40 and KSC LC 39A handle both light and heavy payloads.

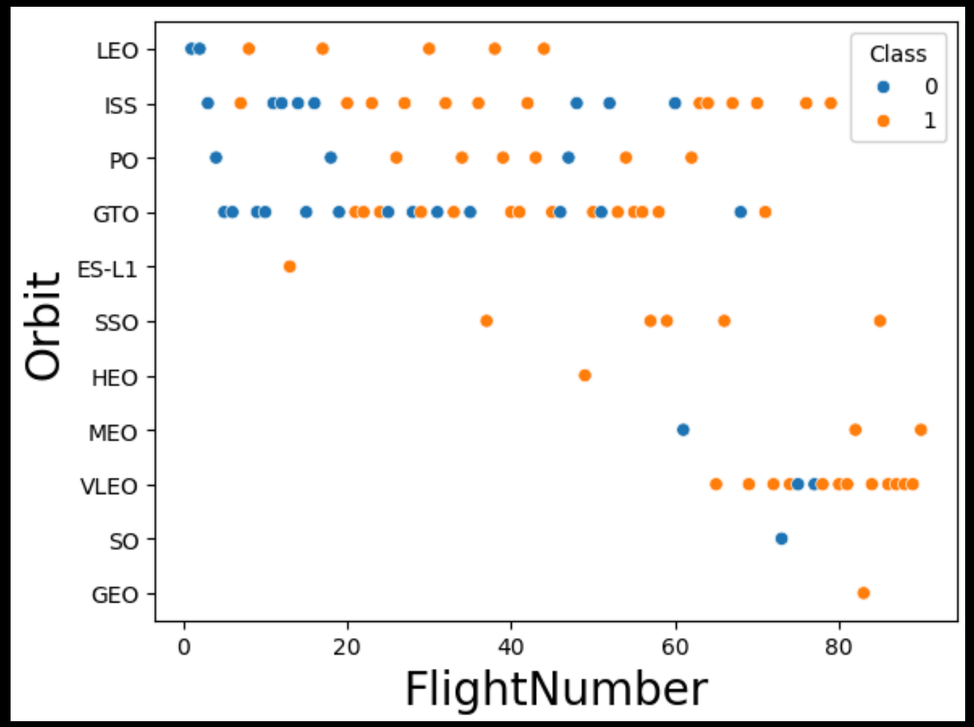
## THE RELATIONSHIP BETWEEN SUCCESS RATE OF EACH ORBIT TYPE




The bar chart shows success rates by orbit type.

- ES-L1, GEO, ISS, and SSO have 100% success rates, indicating high reliability.
- GTO has the lowest success rate, suggesting higher mission risk.
- LEO, MEO, and VLEO show moderate success rates.

## THE RELATIONSHIP BETWEEN SUCCESS RATE OF EACH ORBIT TYPE



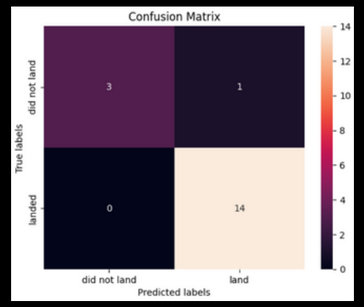

TThe plot shows Flight Number vs. Orbit Type, with LEO improving over time, suggesting learning, while GTO shows no clear trend, likely due to mission complexity.

-

# PREDICTIVE ANALYSIS CLASSIFICATION

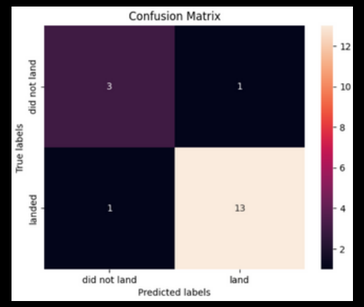
Predictive Analysis Classification involves using machine learning models to classify outcomes based on historical data.
In this case, Logistic Regression, SVM, Decision Trees, and KNN were tested to predict whether the Falcon 9 first stage landed successfully. The models were evaluated on accuracy, with Logistic Regression and KNN achieving the highest performance (94.44%).

## LogisticRegression



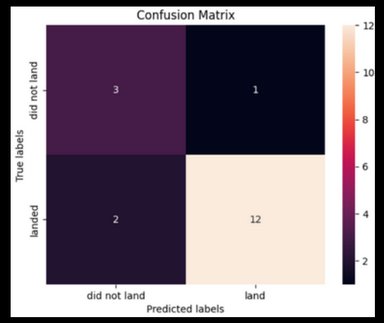

```
Test accuracy: 0.9444
```

## SVM



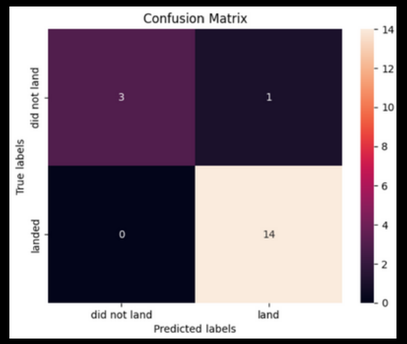

```
Test accuracy: 0.888
```

## DecisionTreeClassifier




```
accuracy : 0.858
```

## KNN




```
0.944444
```

Final Verdict: Choosing the Best Model for Launch Success Prediction 🚀

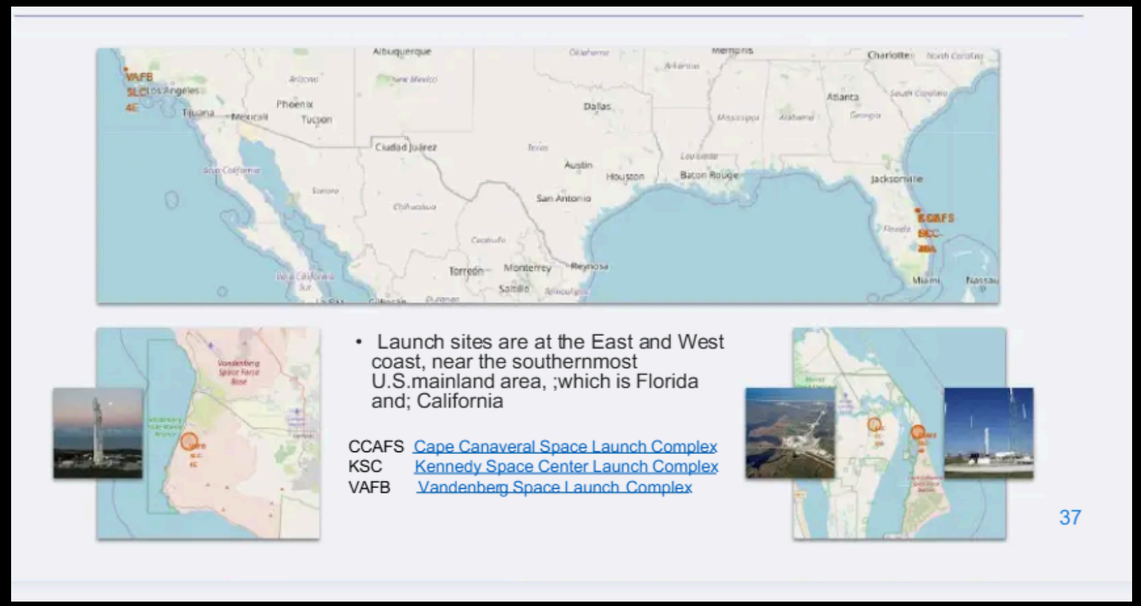
After testing multiple models, Logistic Regression and K-Nearest Neighbors (KNN) achieved the highest accuracy (94.44%), making them the most reliable choices.

- Logistic Regression is ideal for interpretable and fast predictions, making it useful for understanding key factors influencing success.
- KNN excels at pattern recognition, but its computational cost makes it less practical for large datasets.
-

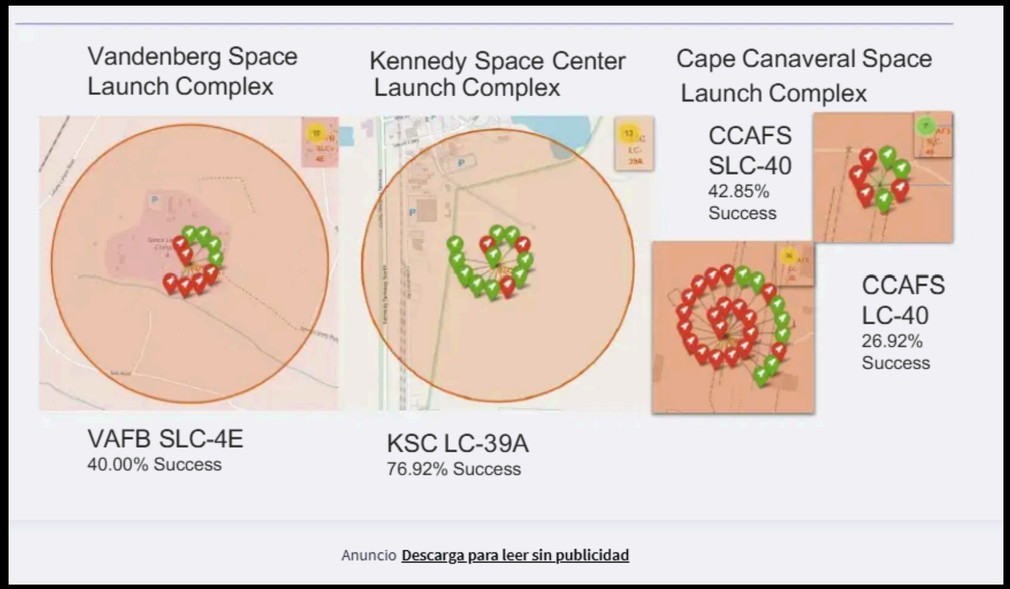
For real-time predictions and operational decision-making, Logistic Regression is the best choice due to its efficiency and interpretability.

# INTERACTIVE MAPS WITH FOLIUM AND DASHBOARDS

# FOLIUM MAP: LAUNCH SITES

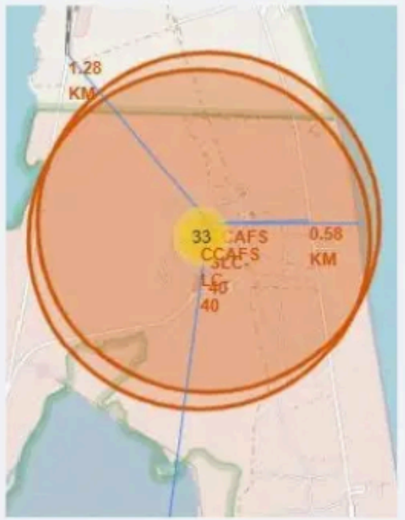



# FOLIUM MAP: LANDING SUCCESS BY LAUNCH SITE

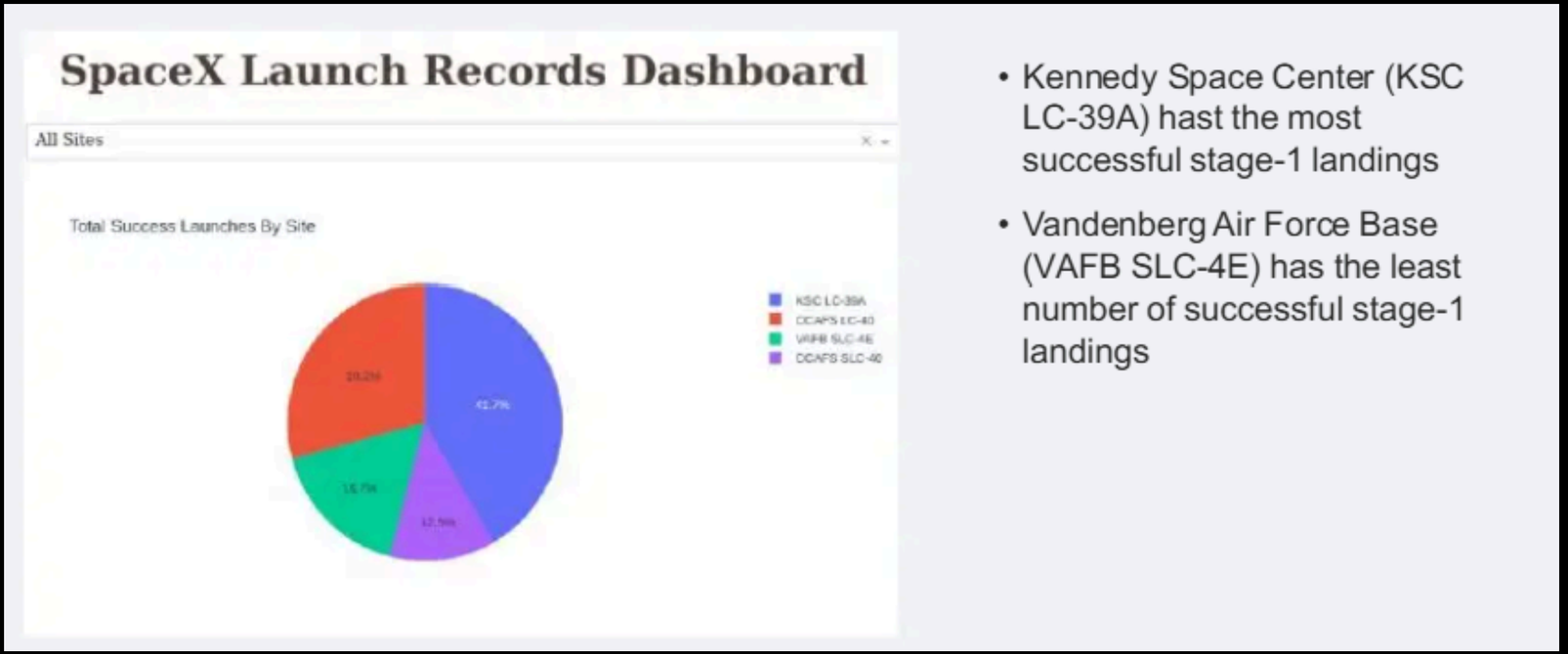


# LOGISTIC AND SAFETY

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and
- relatively far from inhabited areas.



# DASHBOARD: LAUNCH SUCCESS COUNT FOR ALL SITES




- Kennedy Space Center (KSC LC-39A) hast the most successful stage-1 landings
- Vandenberg Air Force Base (VAFB SLC-4E) has the least number of successful stage-1 landings

# CONCLUSION

Through Exploratory Data Analysis (EDA) and visualizations, we uncovered key patterns in Falcon 9 launches, identifying factors that influence mission success.

- EDA and visualizations helped us understand trends, such as the impact of payload mass, launch site, and flight history on success rates.
- Predictive modeling allowed us to estimate the likelihood of a successful landing, with Logistic Regression and KNN achieving the highest accuracy (94.44%).

This process demonstrates that data-driven insights can optimize decision-making in space missions, improving reliability and planning for future launches.