SPACEX LAUNCH SITE ANALYSIS: SUCCESS RATES & PREDICTIVE MODELING

EXECUTIVE SUMMARY SLIDE:

- Briefly introduce the project goal: analyzing SpaceX launch data to study success rates across launch sites and predict mission outcomes.
- Highlight the importance: Data-driven insights can improve launch planning and reduce risk.
- Mention the approach: Exploratory Data Analysis, interactive visualization (Folium/Plotly), and Machine Learning.
- End with key outcomes: Identification of the best-performing launch sites and predictive insights on future launches.

INTRODUCTION:

- SpaceX is a leader in commercial space exploration, with multiple launch sites and missions.

Focus of Study:

- Compare success rates across different launch sites.
- Explore payload and booster factors affecting outcomes.
- Predict the likelihood of a successful launch at a given site.

•DATA COLLECTION:

DATASET: SPACEX LAUNCH RECORDS

KEY VARIABLES: LAUNCH SITE, PAYLOAD MASS, BOOSTER VERSION, LAUNCH OUTCOME

•SOURCE: SPACEX REST API & CSV DATASET

FOCUS: LAUNCH SUCCESS VS. SITE, PAYLOAD & BOOSTER TYPE

Data Wrangling:

- Removed irrelevant/unnecessary columns
- Handled missing values & duplicates
- Standardized payload mass & coordinates
- One-hot encoding for categorical features
- Created new variables: marker colors, distance metrics
- Final clean dataset used for visualization & ML

EXPLORATORY DATA ANALYSIS & INTERACTIVE VISUAL ANALYTICS

Pie Chart (Success/Failure by Launch Site)

- Shows overall success/failure distribution by site
- Helps identify which site has the highest success rate
- Example: KSC LC-39A shows the largest share of successes

Scatter Plot (Payload vs Outcome)

- X-axis: Payload Mass (kg)
- Y-axis: Launch Outcome (0 = Fail, 1 = Success)
- Color-coded by Booster Version Category
- Helps analyze if payload and booster type influence success rate

PREDICTIVE ANALYSIS METHODOLOGY

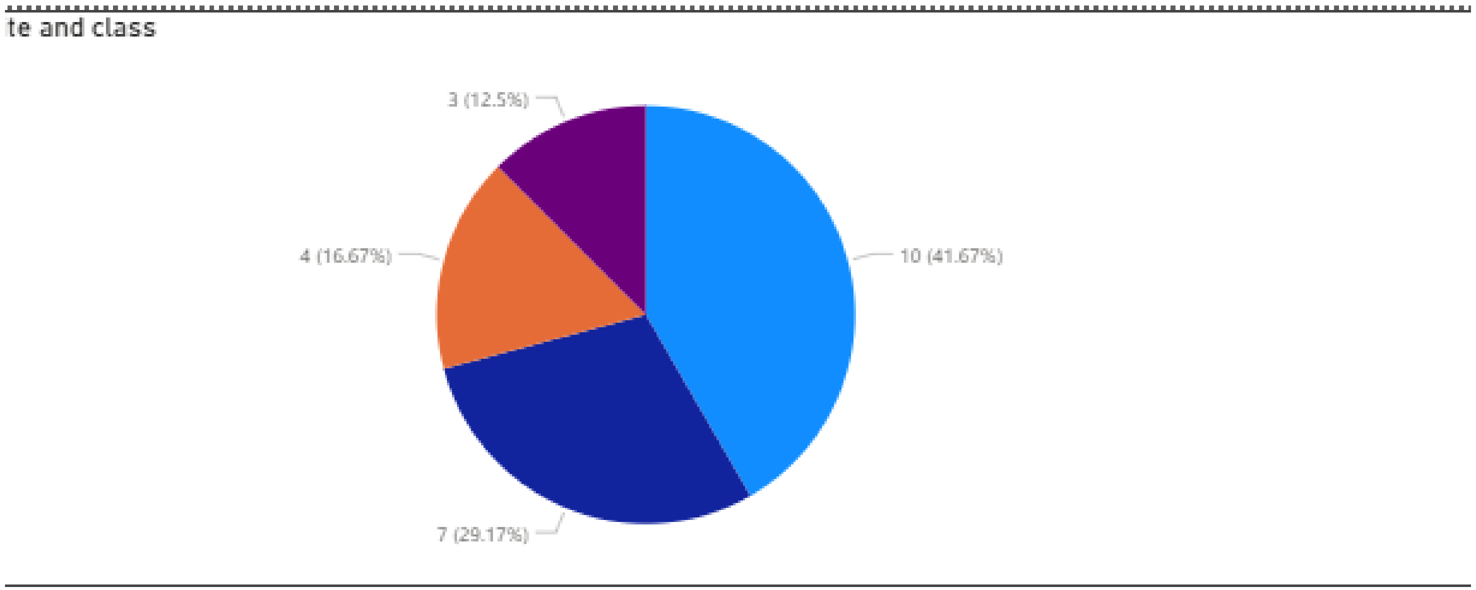
- Applied machine learning models to predict launch success (class 0 or 1)
- Models tested: Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, etc.
- Features used: Launch site, payload mass, booster version category, orbit type
- Performed hyperparameter tuning (GridSearchCV, cross-validation) to improve accuracy
- Evaluated models using accuracy, precision, recall, and F1-score

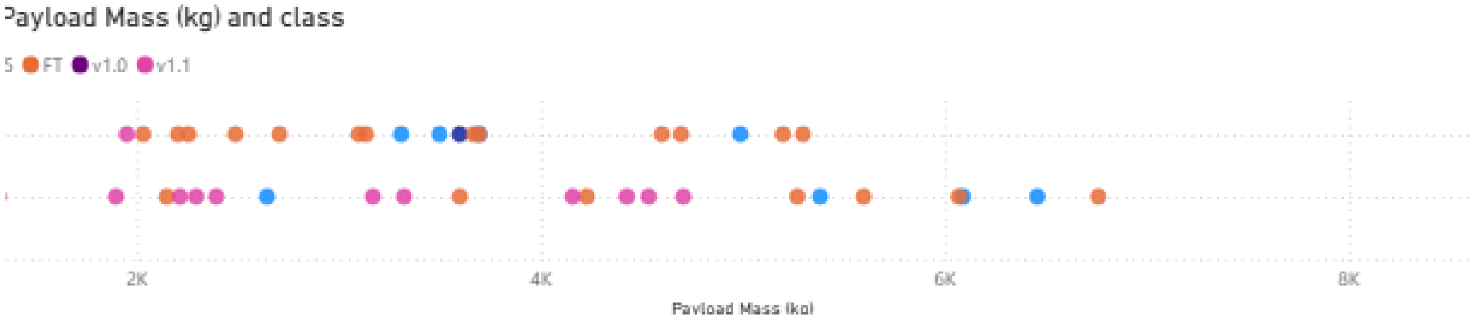
INTERACTIVE VISUAL ANALYTICS

Here we performed exploratory data analysis to understand launch success patterns. The pie chart on top shows how success rates differ by launch site, where we can see that "KSC LC-39A" has the highest number of successful launches.

The scatter plot below analyzes payload mass versus launch outcome, with points color-coded by booster version category.

We also made the visual interactive, allowing filters on payload mass and site selection for deeper insights.





EXPLORATORY DATA ANALYSIS WITH SQL RESULTS:

- Queried dataset using SQL to extract key insights
- Analyzed launch outcomes across sites using aggregate queries
- Examined payload mass distribution vs. success/failure
- Identified success rate per launch site using `GROUP BY`
- Checked relationships between booster version and class outcome

```
%sql SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch
```

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

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

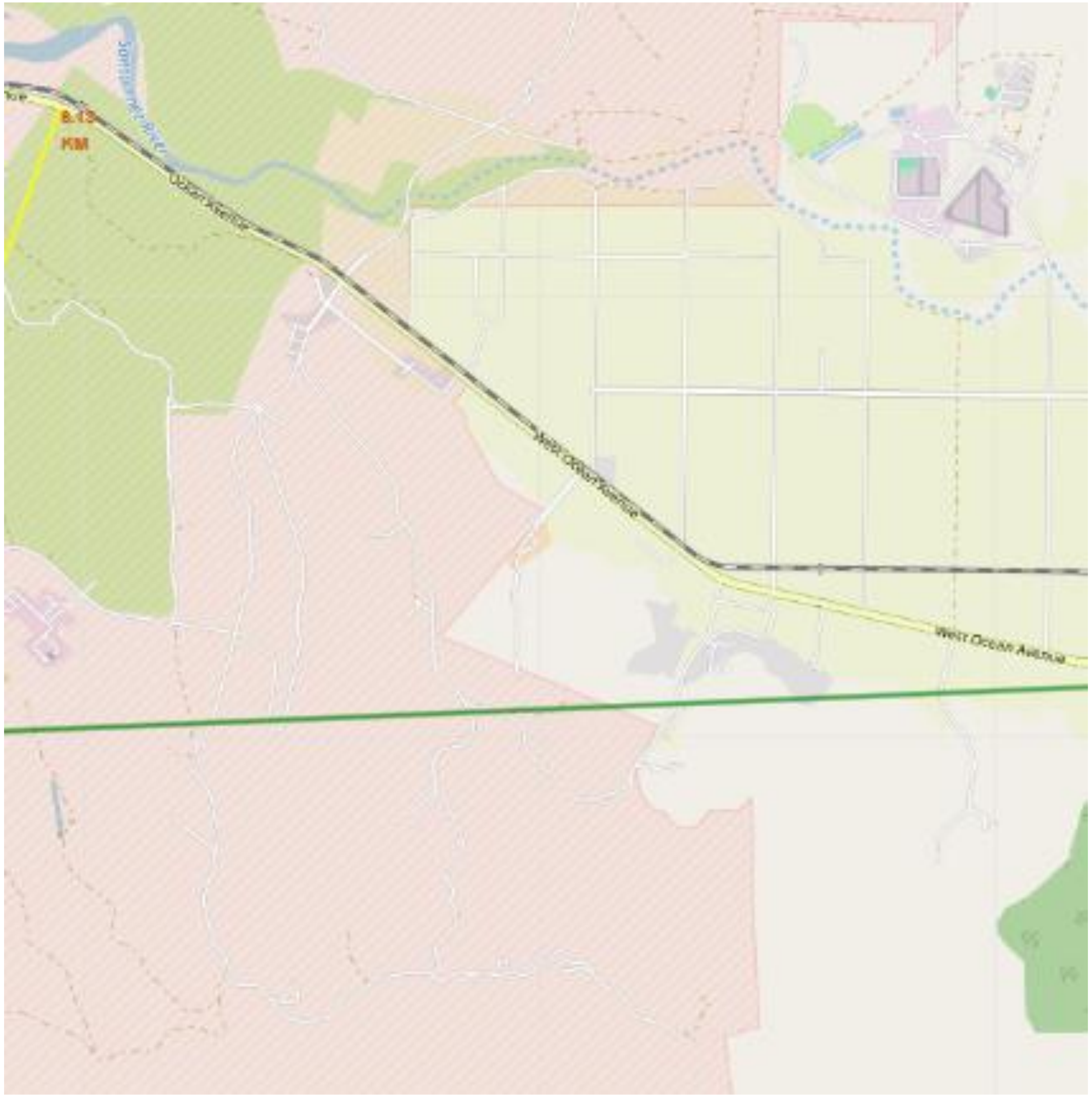
Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad))

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Dat
```

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

Landing_Outcome	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



INTERACTIVE MAP WITH FOLIUM:

- Created an interactive map of SpaceX launch sites using Folium
- Added markers for each launch site with success/failure outcomes
- Integrated popups with site-specific details (success rate, location)
- Drew proximity lines to nearest cities, highways, and railways

PLOTLY DASH DASHBOARD:

- Built a chart using Plotly Dash
- Visualized **success rate trends** over the years
- Dashboard helps identify **patterns and correlations** in launch outcomes

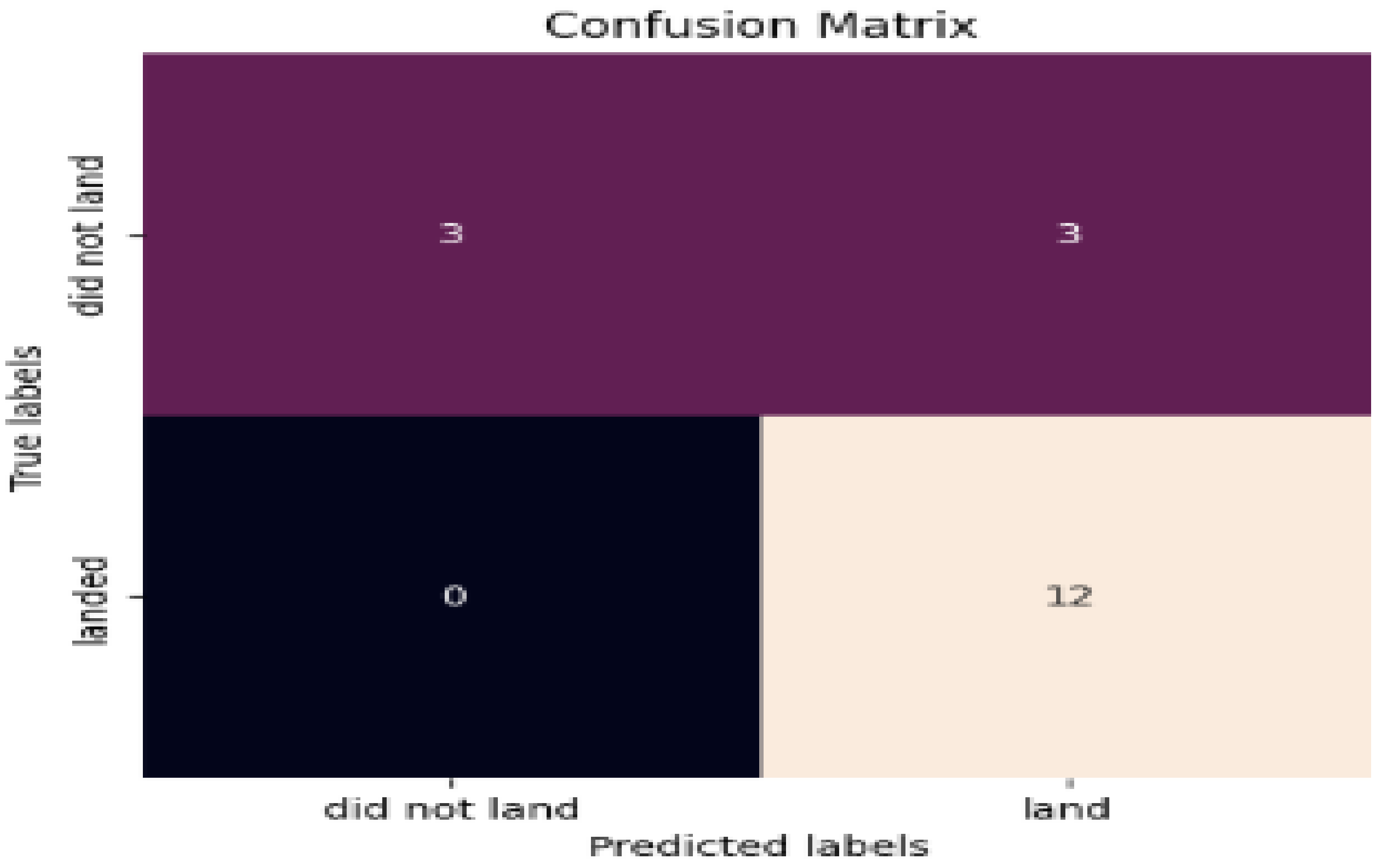
PREDICTIVE ANALYSIS RESULTS:

- **Models Implemented:** Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest
- **Features Used:** Payload Mass, Launch Site, Booster Version, Orbit, etc.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score
- **Best Performing Model:** (e.g., Decision Tree / Random Forest) with highest accuracy and balanced metrics

Test Accuracy: 0.8333333333333334

Lets look at the confusion matrix:

```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



CONCLUSION:

- Conducted exploratory data analysis (EDA) to identify key trends and patterns in launch outcomes.
- Interactive dashboards and visualizations highlighted the role of Launch Site, Payload, and Booster Version in determining success rates.
- Predictive modeling using machine learning demonstrated that Decision Tree / Random Forest (insert your best model) achieved the most accurate predictions..
- The analysis framework can be applied to future launch planning to improve success probability.

THANK YOU