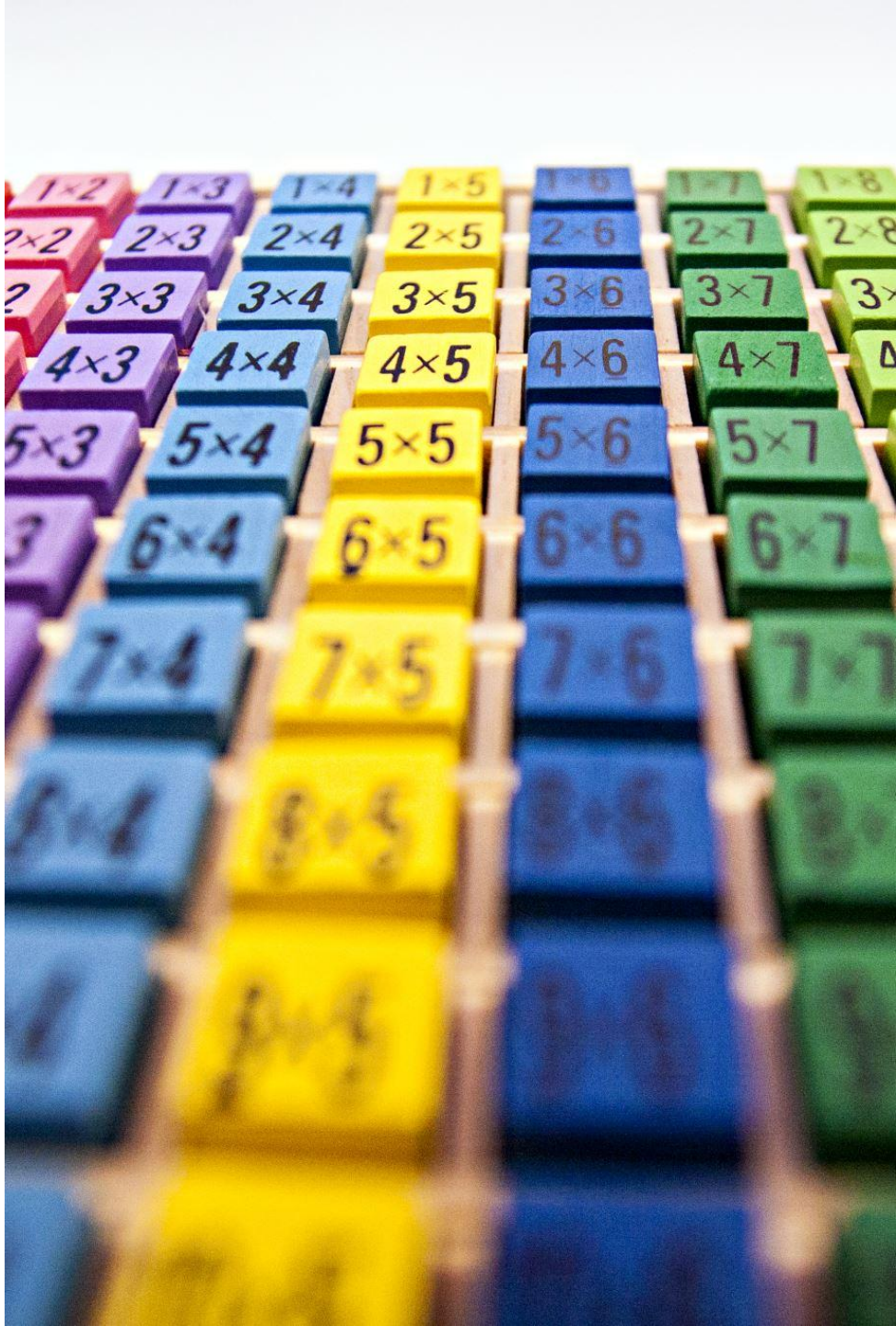
SPACEX LAUNCH ANALYSIS: DATA SCIENCE CAPSTONE PRESENTATION

1. EXECUTIVE SUMMARY

- Explored SpaceX launch performance using multiple data sources (API, Wikipedia, CSV).
- Applied web scraping, SQL, Folium, and machine learning.
- Developed interactive dashboard with Plotly Dash.
- Achieved 83.33% accuracy across all classifiers.
- Delivered key insights on launch site performance, success patterns, and booster versions.

2. INTRODUCTION & OBJECTIVE

- Objective: Analyze SpaceX launches to identify success drivers.
- Tools: Python, Pandas, BeautifulSoup, Folium, SQL, Plotly, scikit-learn.
- Scope: From data extraction to predictive modeling.



3. DATA COLLECTION

NASA/SpaceX API used to extract booster-related metadata.

Wikipedia scraped for launch records (launch site, date, outcome).

CSV files used for integration and formatting.

4. DATA WRANGLING

01

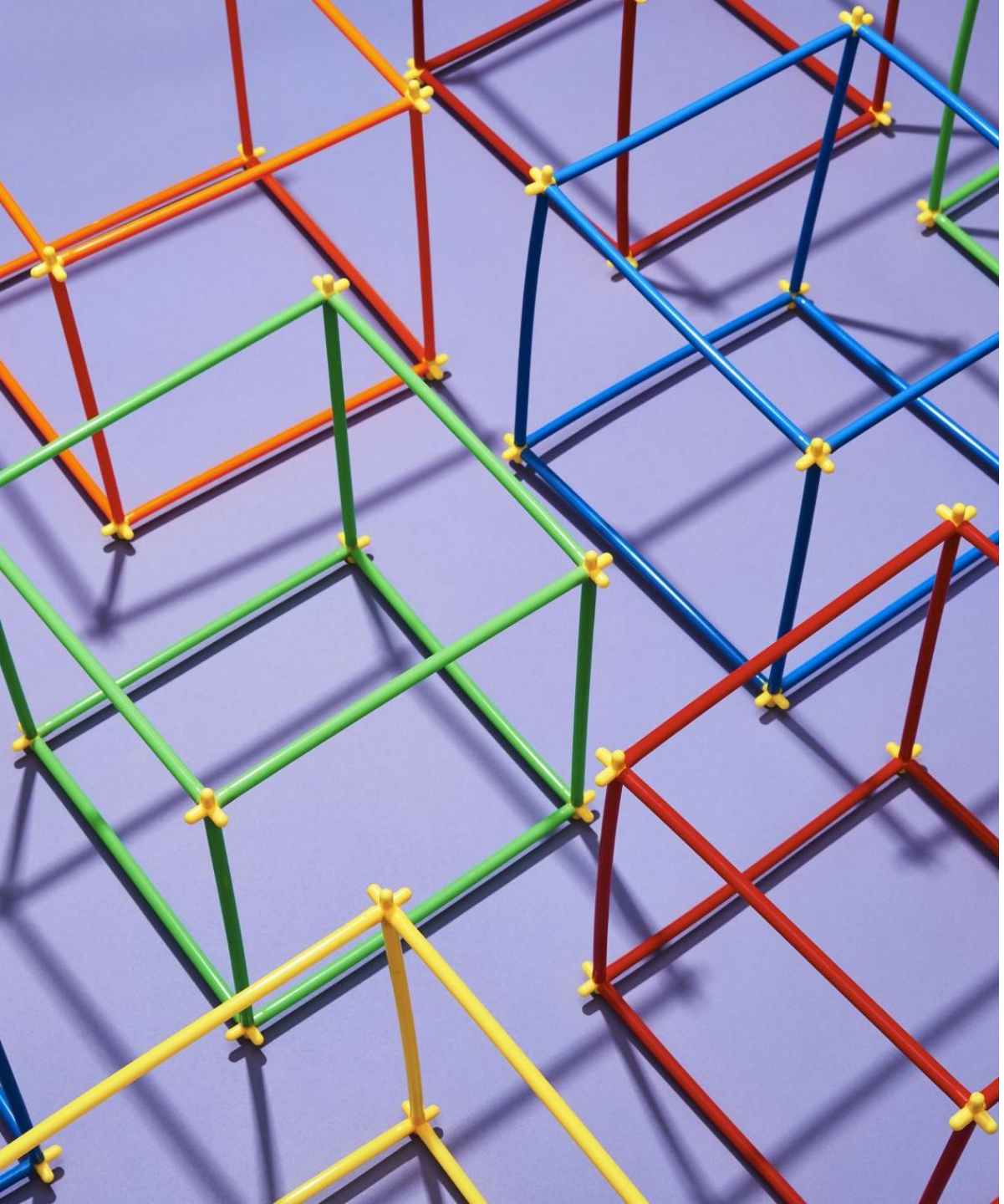
Cleaned column names, handled missing values.

02

Created binary success label (Class).

03

Extracted booster name, launch date, payload mass, and site info.

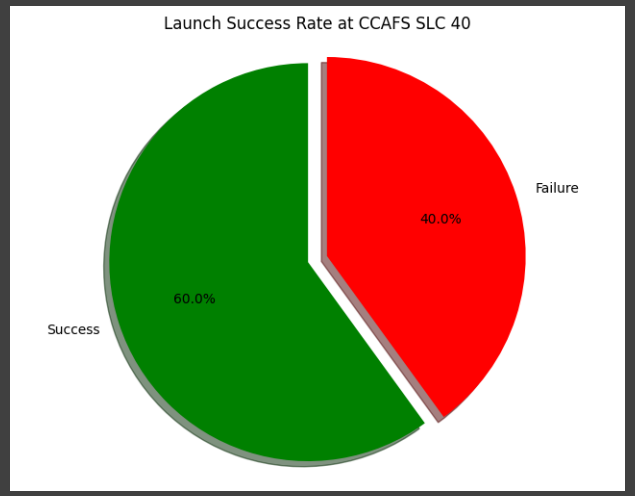


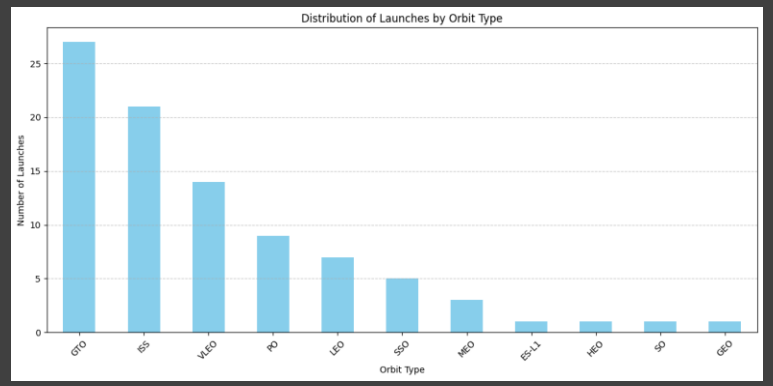
6. PREDICTIVE MODELING: METHODS

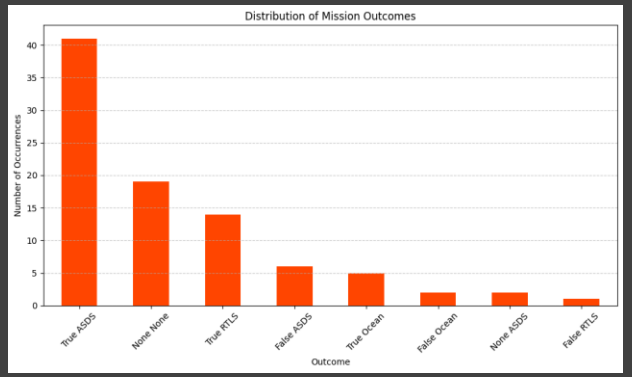
- Standardized features using StandardScaler.
- Trained and validated 4 models: Logistic Regression, SVM, Decision Tree, KNN.
- Used GridSearchCV with 10-fold cross-validation.

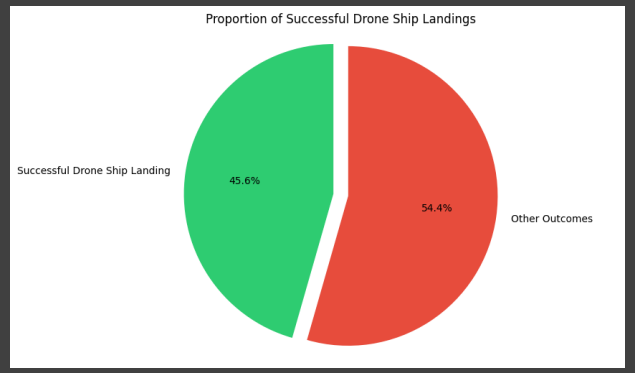
7. EDA: VISUAL ANALYSIS

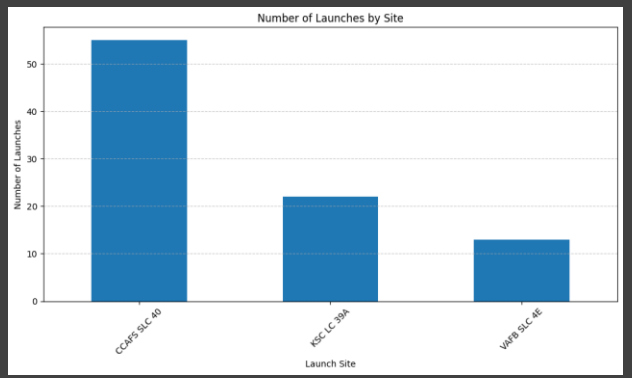
- Launch success increased over time (FlightNumber).
- GTO launches had lower success compared to LEO.
- Payload Mass vs Launch Site scatter showed clustering by success.
- Orbit type was a strong predictor of outcome.


Launch Success Rate at CCAFS SLC 40
Failure
40.0%
60.0%
Success

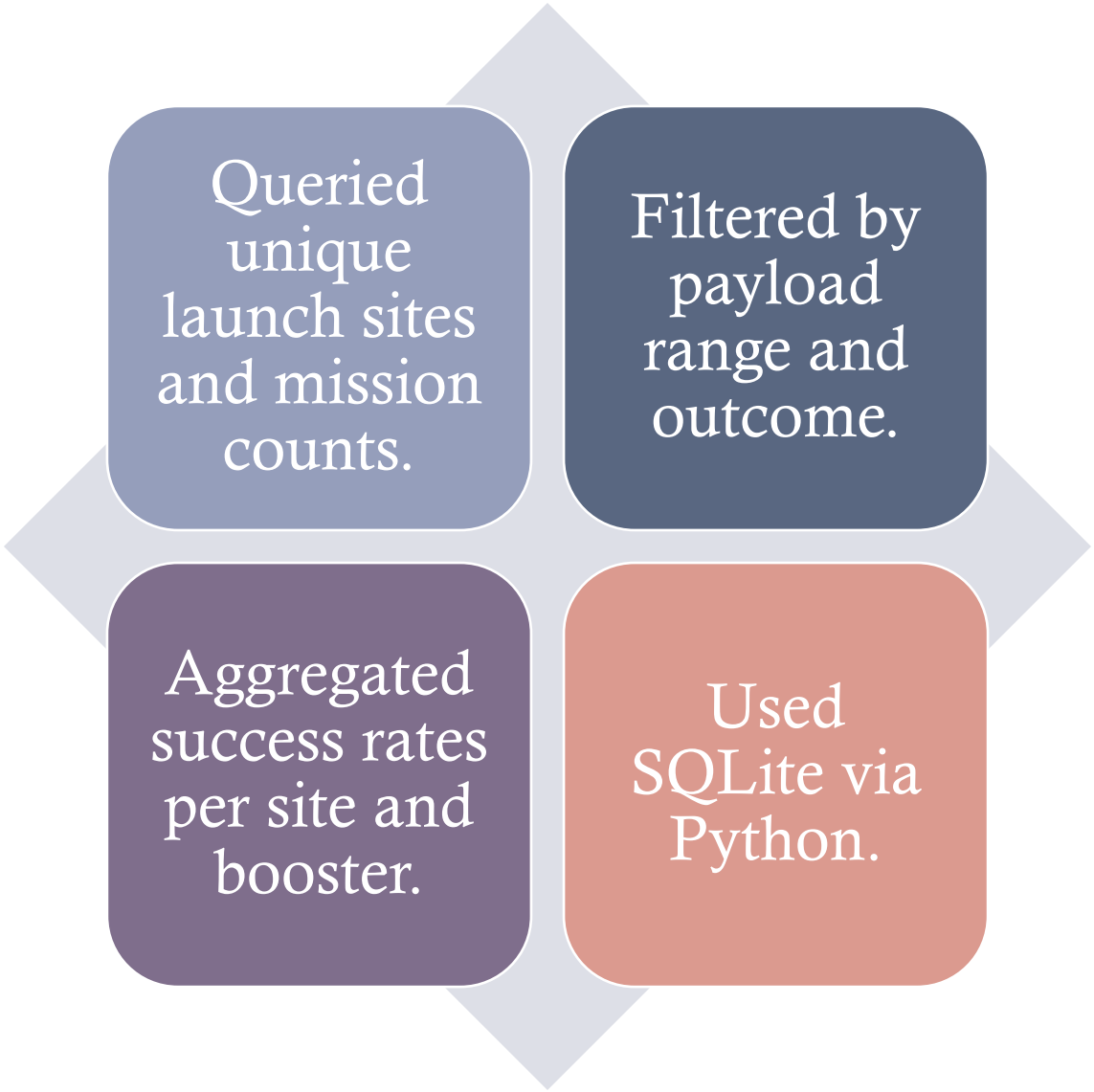

Distribution of Launches by Orbit Type
Number of Launches
0
5
10
15
20
25
GTO
ISS
VLEO
PO
LEO
SSO
MEO
ES-L1
HEO
SO
GEO
Orbit Type


Distribution of Mission Outcomes
Number of Occurrences
0
5
10
15
20
25
30
35
40
True ASDS
None None
True RTLS
False ASDS
True Ocean
False Ocean
None ASDS
False RTLS
Outcome


Proportion of Successful Drone Ship Landings
Successful Drone Ship Landing
45.6%
54.4%
Other Outcomes

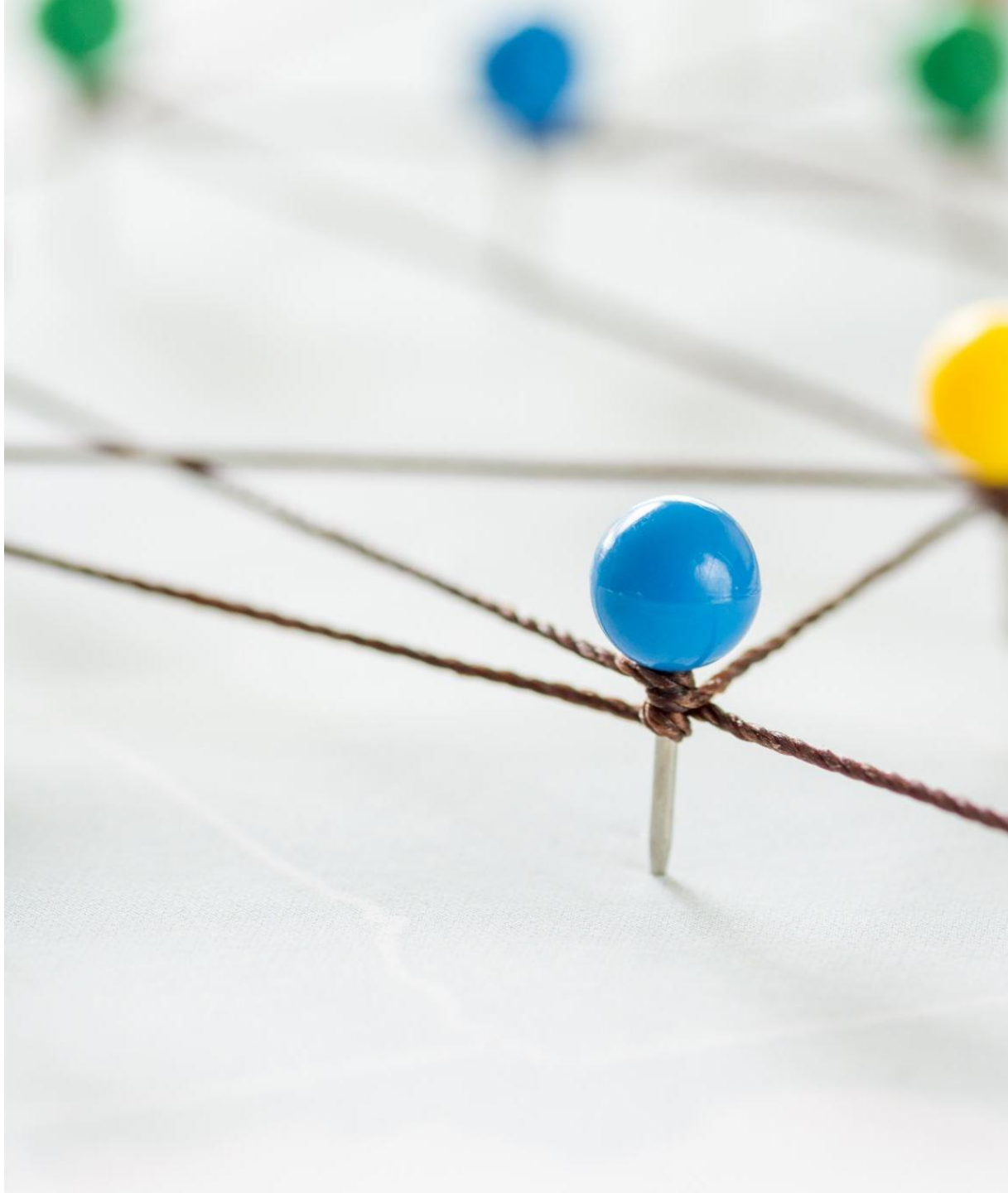

Number of Launches by Site
Number of Launches
0
10
20
30
40
50
CCAFS SLC 40
KSC LC 39A
VAFB SLC 4E
Launch Site

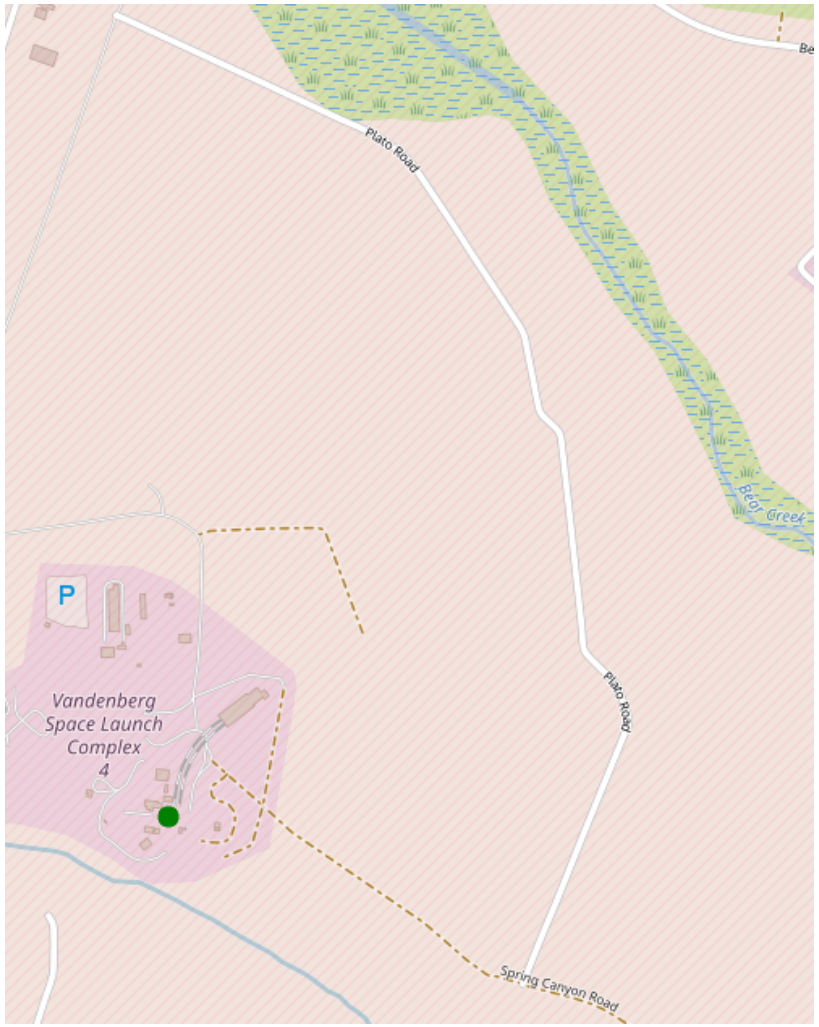
8. EDA: SQL RESULTS

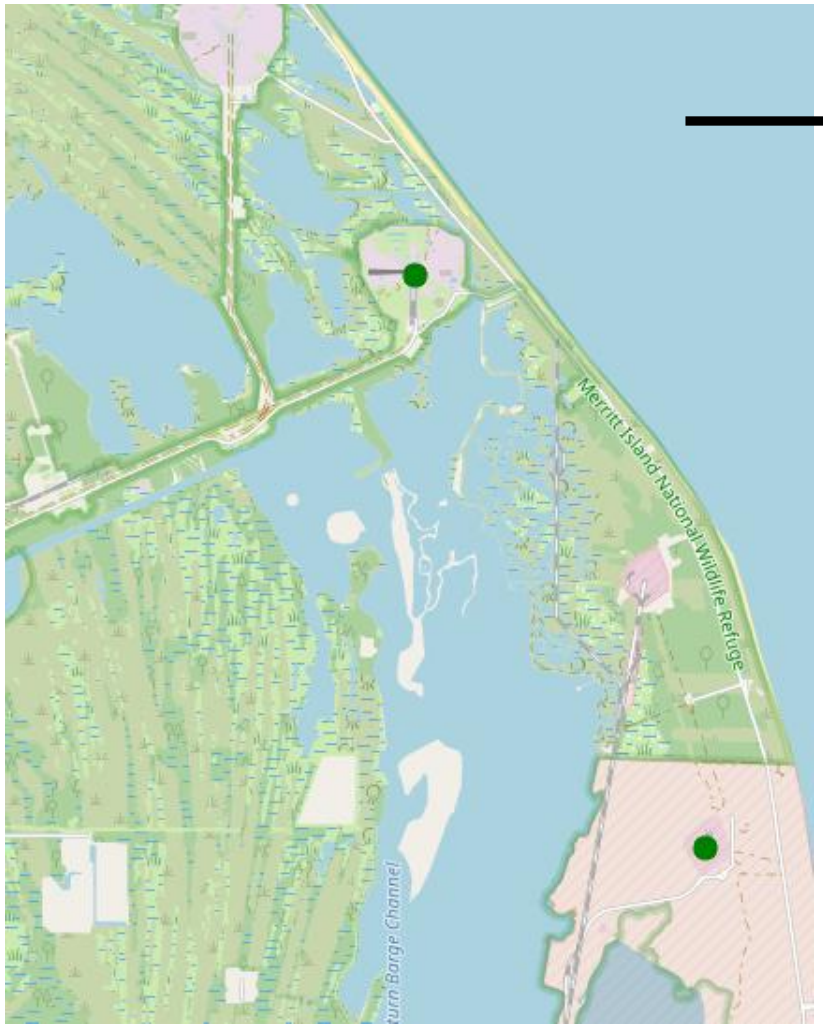


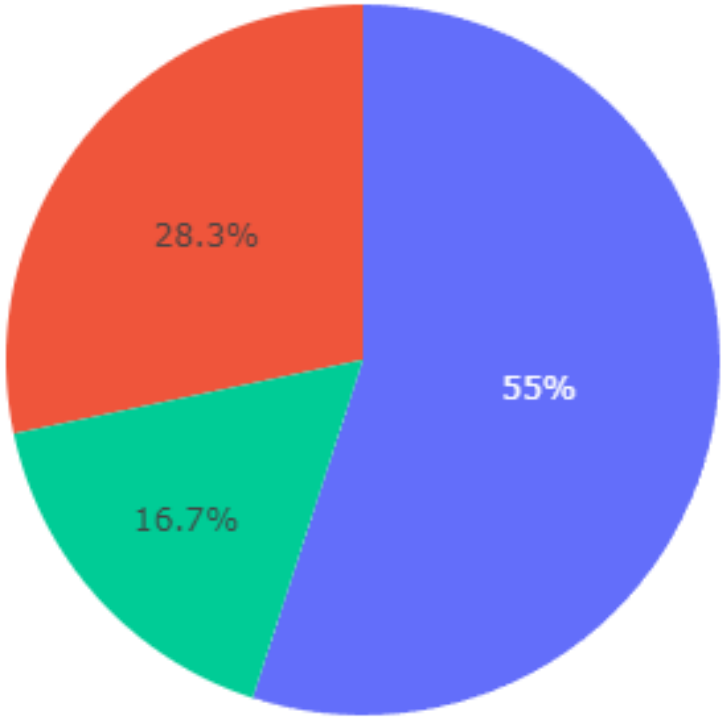
9. INTERACTIVE MAP (FOLIUM)

- Mapped all launch sites with coordinates.
- Colored markers by launch outcome (success/failure).
- Tooltip popups showing site and status.
- Great for spatial pattern recognition.




Plato Road
Bear Creek
P
Vandenberg
Space Launch
Complex
4
Plato Road
Spring Canyon Road

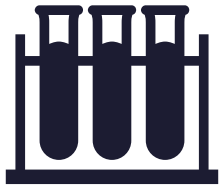

Merritt Island National Wildlife Refuge
Barge Channel


28.3%
16.7%
55%

10. DASHBOARD (PLOTLY DASH)

- Dropdown: Filter by launch site.
- Range slider: Select payload range.
- Pie chart: Launch success distribution.
- Scatter plot: Payload vs outcome by booster version.
- Fully interactive, built with Dash components.

11. PREDICTIVE RESULTS



Test set had 18 samples (20% split).

All models achieved 83.33% test accuracy:

Logistic Regression

SVM (rbf kernel)

Decision Tree

KNN

Suggests high-quality, consistent data — or limited variability.

12. CONCLUSION

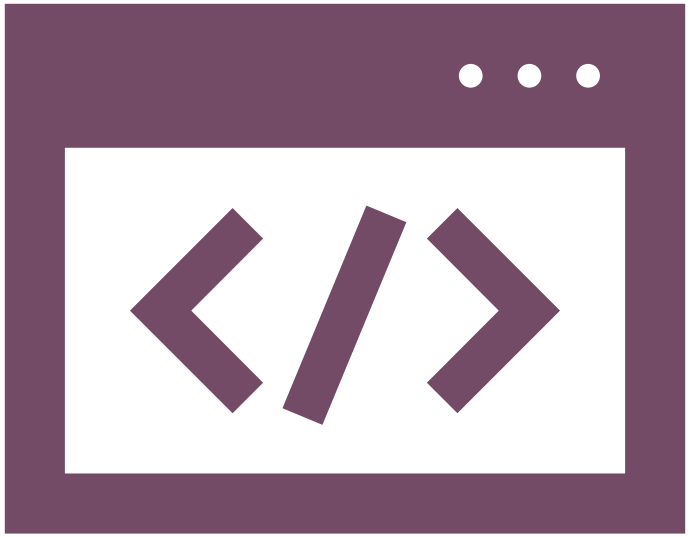
SpaceX launch data supports successful modeling.

Orbit and payload are key predictors of outcome.

Visual tools like Folium and Dash greatly enhance insights.

All models performed equally; future work could explore ensemble methods or time-based prediction.

13. CREATIVITY & INNOVATION



- Used real APIs + scraping + SQL + Dash.
- Highlighted anomaly (equal model accuracy) as an insight.
- Balanced technical rigor with clear storytelling.
- GitHub repo