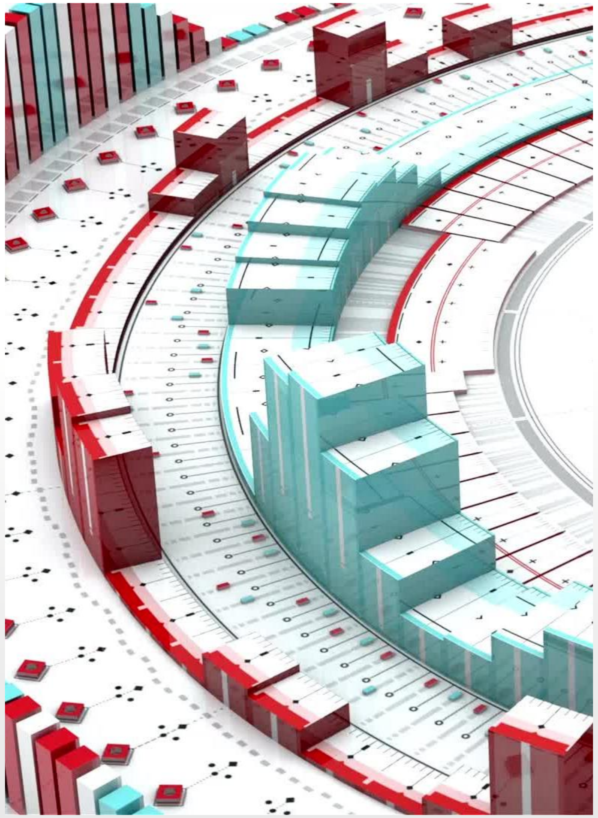
SPACEX FALCON 9 FIRST STAGE LANDING PREDICTION

DATA SCIENCE CAPSTONE PROJECT

Presented by: Nathaniel Attoh
09/09/2025

•**Objective:** To predict the success of SpaceX Falcon 9 first-stage landings, a critical factor in their significantly lower launch costs compared to competitors.

•**Method:** Utilized a comprehensive data science methodology involving data collection from the SpaceX API and web scraping, followed by data wrangling, exploratory data analysis (EDA) with SQL and visualization, interactive mapping, dashboard creation, and predictive modeling using machine learning classification algorithms.



INTRODUCTION / BACKGROUND

- SpaceX has revolutionized the space launch industry with a cost of $62 million per Falcon 9 launch, compared to upwards of $165 million for other providers.
- This cost advantage is largely due to the reusability of the Falcon 9 rocket's first stage.
- **Problem Statement:** Can we accurately predict the success of a first-stage landing based on historical launch data?
- **Data Sources:** Data was collected from two primary sources:
  - **SpaceX REST API:** For detailed launch records.
  - **Wikipedia:** Via web scraping for additional launch history details.

DATA COLLECTION & WRANGLING

Data Collection:

•Made GET requests to the SpaceX API to collect structured launch data in JSON format.

•Used Python's BeautifulSoup library to scrape and parse HTML launch tables from Wikipedia.

Data Wrangling:

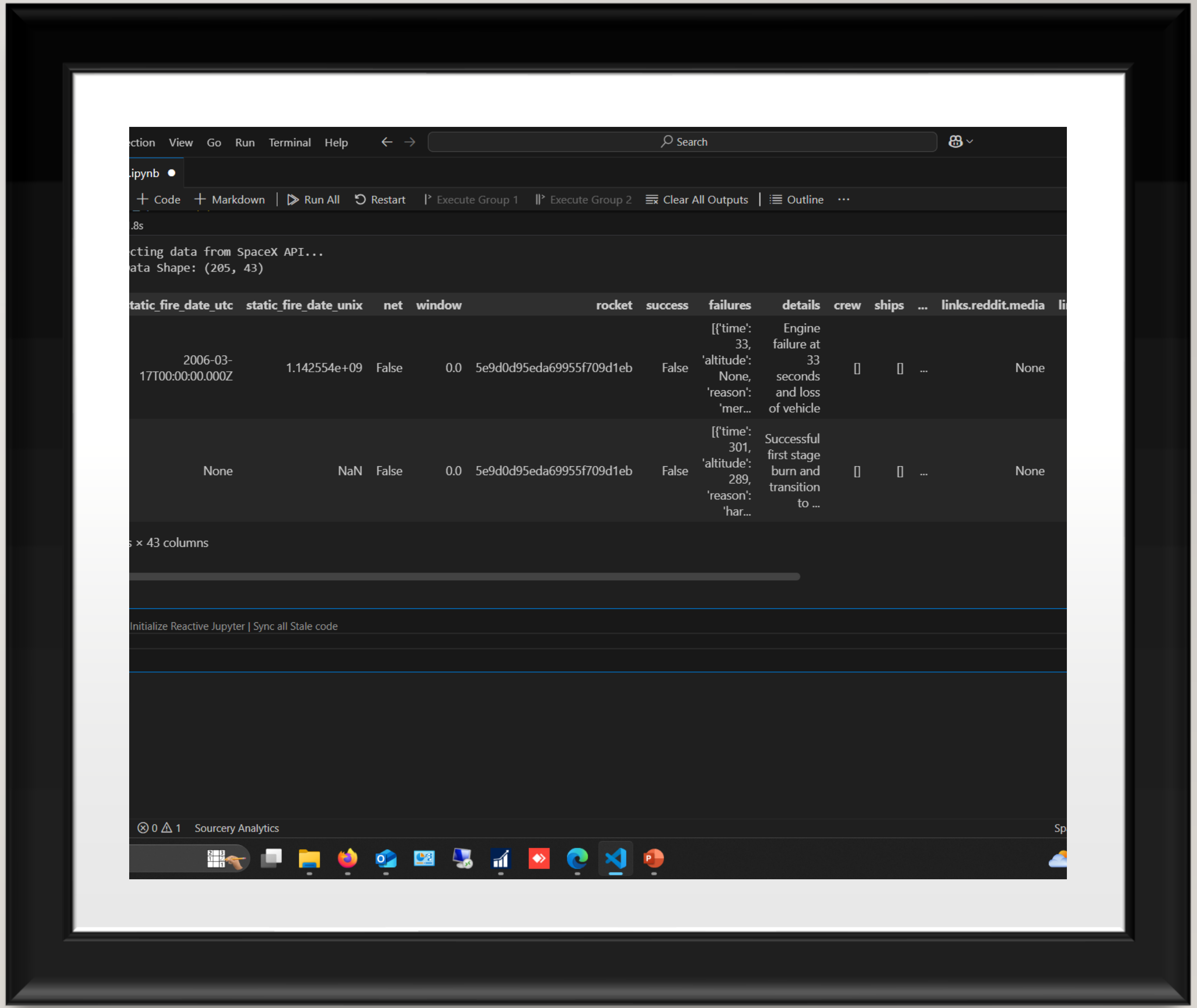
•Loaded data into Pandas DataFrames for cleaning and manipulation.

•Handled missing values in key columns like landing_pad.

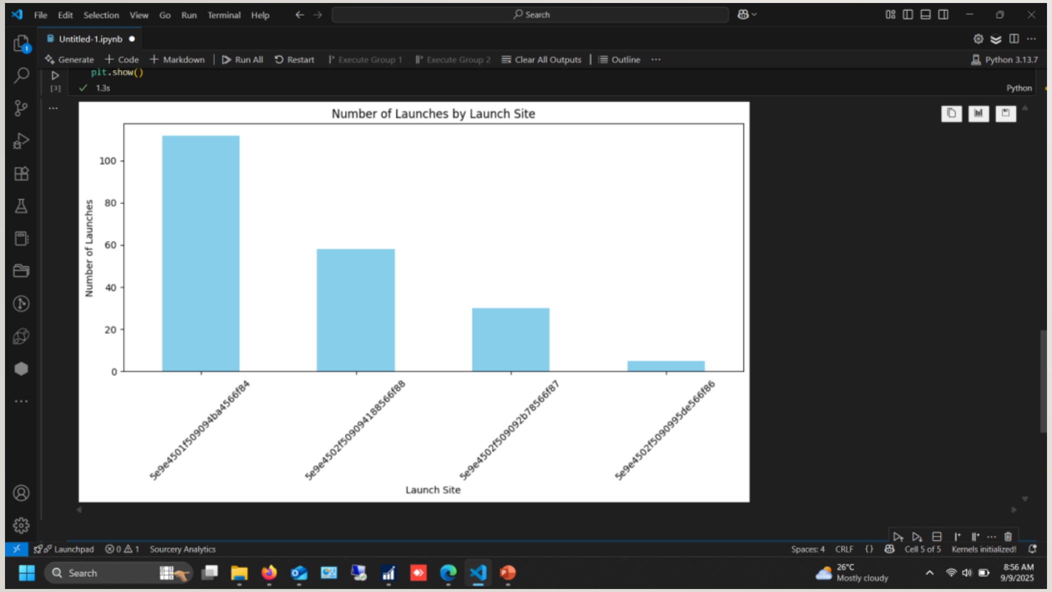
•Filtered the dataset to include only Falcon 9 launches.

•Engineered the target variable Class, where 1 represents a successful landing and 0 represents a failed or unintended landing.

•Performed One-Hot Encoding on categorical variables (Orbit, LaunchSite, LandingPad, Serial) to prepare data for machine learning.



EDA WITH VISUALIZATION



Exploratory Data Analysis (EDA) through visualization revealed key patterns and relationships

within the data.

Key Insights:

•The launch site **CCAFS SLC-40** has the highest number of launch attempts.

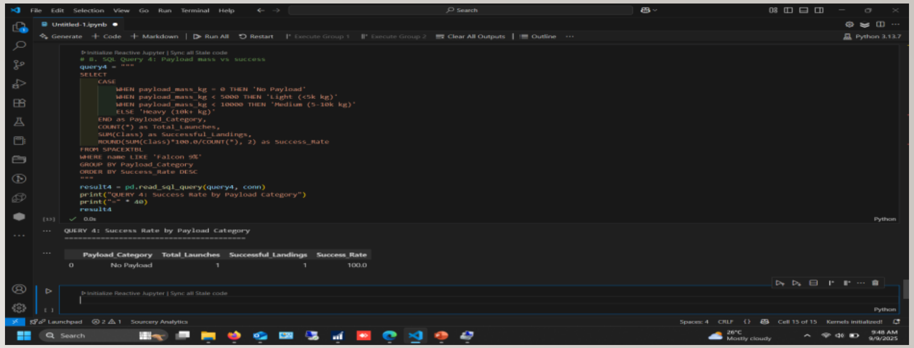
•Success rate has improved significantly over time, as seen in scatter plots of FlightNumber vs.

Outcome.

•Certain orbits, like **ES-L1** and **GEO**, have higher success rates compared to others.

•Heavier payload masses are associated with more challenging landing conditions.

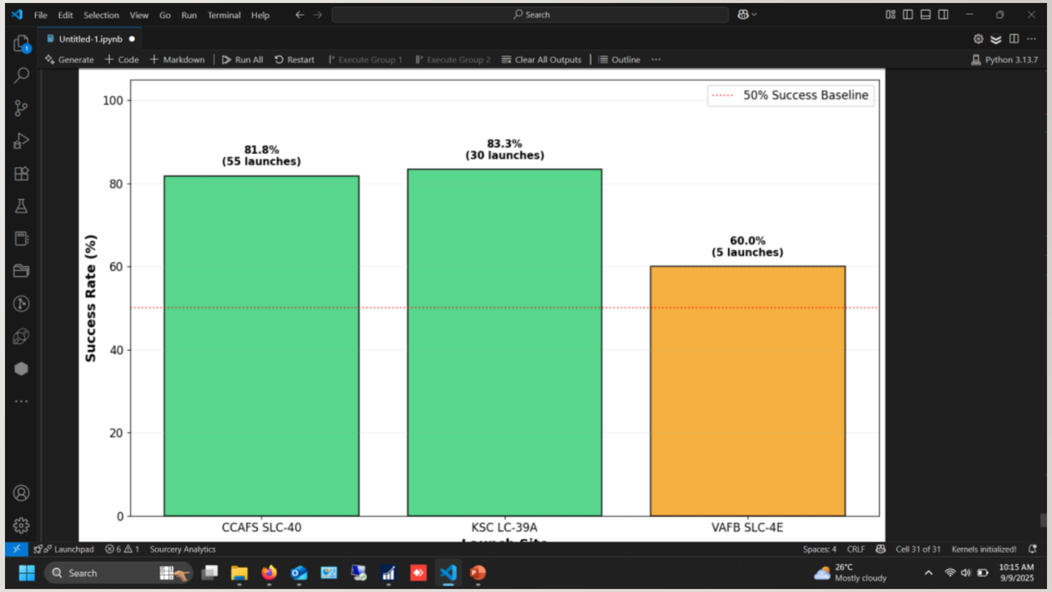
EDA WITH SQL



Objective	SQL Query	Result
Total Falcon 9 Launches	SELECT COUNT(*) FROM SPACEXTBL WHERE name LIKE 'Falcon 9%';	90
Successful Landings	SELECT COUNT("Mission_Outcome") FROM SPACEXTBL WHERE "Mission_Outcome" = 'Success';	60
Payload Mass by Booster	SELECT booster_version, SUM(PAYLOAD_MASS__KG_) AS Total_Mass FROM SPACEXTBL GROUP BY booster_version;	(See result set)
Success Rate by Launch Site	SELECT Launch_Site, COUNT("Mission_Outcome") AS Success_Count FROM SPACEXTBL WHERE "Mission_Outcome" = 'Success' GROUP BY Launch_Site;	(See r

INTERACTIVE MAP WITH FOLIUM

- An interactive geospatial map was created using the Folium library to visualize launch sites and landing outcomes.
- **Map Features:**
- Launch sites are marked on their exact geographical coordinates.
- **Green markers** indicate successful landings.
- **Red markers** indicate failed landings.
- Clicking on a marker displays details about the launch, including date, payload, and mission outcome.



DASHBOARD WITH PLOTLY DASH

AN INTERACTIVE PLOTLY DASH APPLICATION WAS BUILT TO PROVIDE A DYNAMIC AND USER-FRIENDLY TOOL FOR EXPLORING THE LAUNCH DATA.

- **Dashboard Functionality:**
- A dropdown menu allows users to select a specific launch site.
- The dashboard updates in real-time to display:
  - The total number of launches for the selected site.
  - A pie chart showing the ratio of successful vs. unsuccessful landings.
  - A chart showing the trend of outcomes over time (Flight Number).

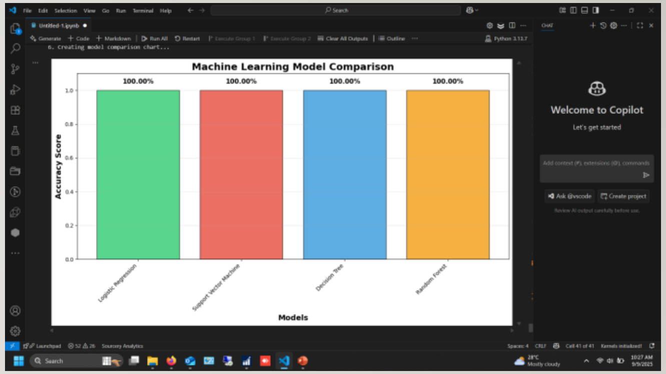
MACHINE LEARNING: METHODOLOGY

•**Objective:** To build a classification model predicting landing success (Class).
•**Algorithms Tested:** Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, K-Nearest Neighbors (KNN).

•**Process:**
1.Data was standardized using Standard Scaler.
2.Data was split into **80% training** and **20% testing** sets.
3.Hyperparameter tuning was performed for each model using GridSearchCV with 10-fold cross-validation to find the optimal settings.
•**Evaluation Metric:** Accuracy (percentage of correct predictions).

MACHINE LEARNING: RESULTS

The tuned machine learning models achieved the following results:



- The **Support Vector Machine (SVM)** model with an **RBF kernel** was selected as the best model due to its high and consistent accuracy.
- The confusion matrix shows the model is effective but occasionally predicts a false positive (predicts success when the landing failed).

CONCLUSION

SUMMARY: THIS PROJECT SUCCESSFULLY CREATED A END-TO-END DATA SCIENCE PIPELINE TO PREDICT SPACEX FALCON 9 FIRST-STAGE LANDING OUTCOMES WITH HIGH ACCURACY.

KEY FINDINGS: LAUNCH SITE, ORBIT, AND PAYLOAD MASS WERE IDENTIFIED AS SIGNIFICANT FACTORS INFLUENCING LANDING SUCCESS. THE COMPANY'S SUCCESS RATE HAS NOTICEABLY IMPROVED OVER TIME.

BUSINESS IMPACT: THIS MODEL CAN BE VALUABLE FOR:

SPACEX: IN MISSION PLANNING AND RISK ASSESSMENT.

COMPETITORS: FOR UNDERSTANDING THE PARAMETERS OF SPACEX'S COST-SAVING REUSABILITY.

INVESTORS & ANALYSTS: FOR EVALUATING THE TECHNICAL AND FINANCIAL PERFORMANCE OF SPACEX.

FUTURE WORK: INCORPORATE REAL-TIME TELEMETRY DATA FOR EVEN MORE ACCURATE PREDICTIONS AND DEPLOY THE MODEL AS A LIVE API.