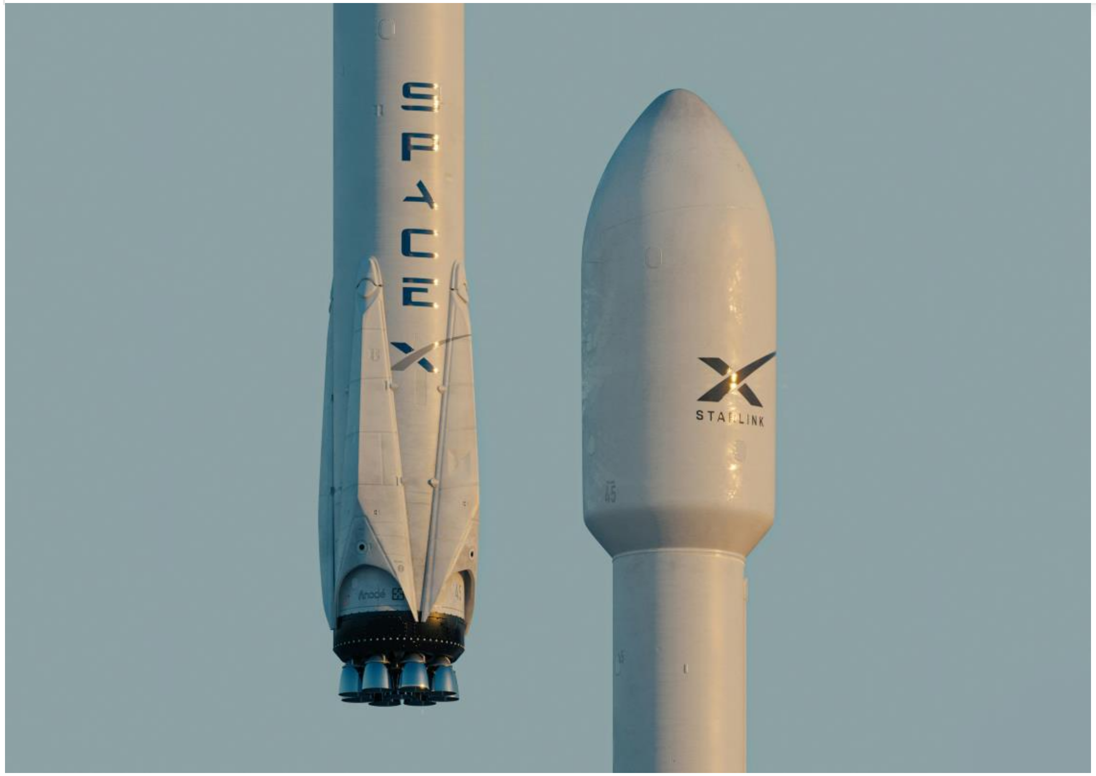
SpaceX Falcon 9 First Stage Landing Prediction

Data Science & D Machine Learning for Launch Cost Optimization

Name: Yusuf Elvan

Date: 01.18.2026

OUTLINE



- Executive Summary
- Introduction
- Methodology (Data Collection, Wrangling, SQL, EDA)
- Results (Launch Site, Payload, Orbit & ML Performance)
- Dashboard Summary
- Discussion & Conclusion

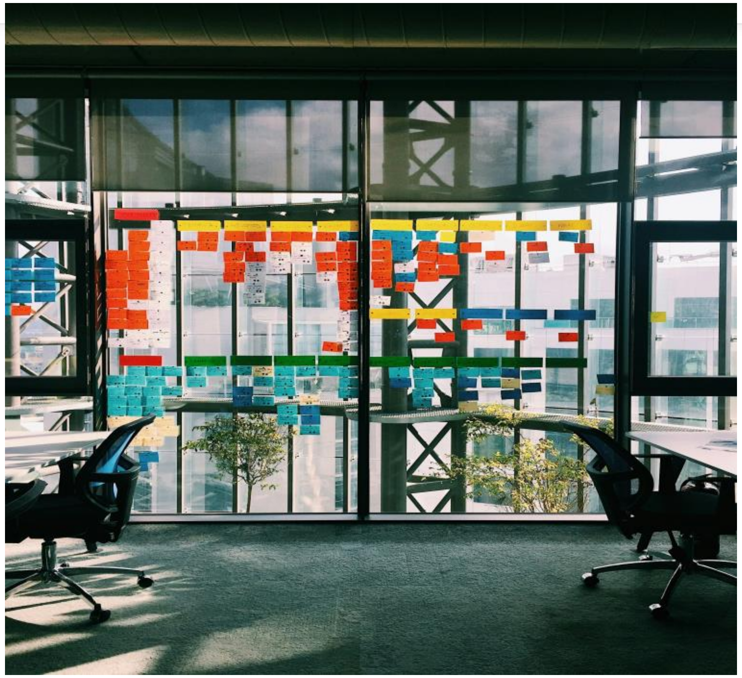
EXECUTIVE SUMMARY

- **Point 1:** Goal is to predict if the Falcon 9 first stage will land successfully.
- **Point 2:** Successful landings reduce launch costs from $165M to $62M.
- **Point 3:** Data was integrated from SpaceX API and Wikipedia web scraping.
- **Point 4:** Four ML models were trained: Logistic Regression, SVM, Decision Tree, and KNN.
- **Point 5:** All models achieved an accuracy of 83.3% on the test set.

INTRODUCTION

- **Point 1:** SpaceX is the leader in reusable rocket technology.
- **Point 2:** Reusability is the key factor in SpaceX's market dominance.
- **Point 3:** Predicting landing outcomes helps competitors estimate launch prices.
- **Point 4:** Analysis covers launch sites, orbits, payload mass, and flight history.

METHODOLOGY



- **Point 1: Data Collection:** Used REST API (SpaceX) and BeautifulSoup (Web Scraping).
- **Point 2: Data Wrangling:** Handled missing values and created a binary "Class" column (1 = Success, 0 = Failure).
- **Point 3: Exploratory Data Analysis:** Used SQL for deep queries and Folium for geographical mapping.
- **Point 4: Machine Learning:** Standardized data and optimized hyperparameters using GridSearchCV.

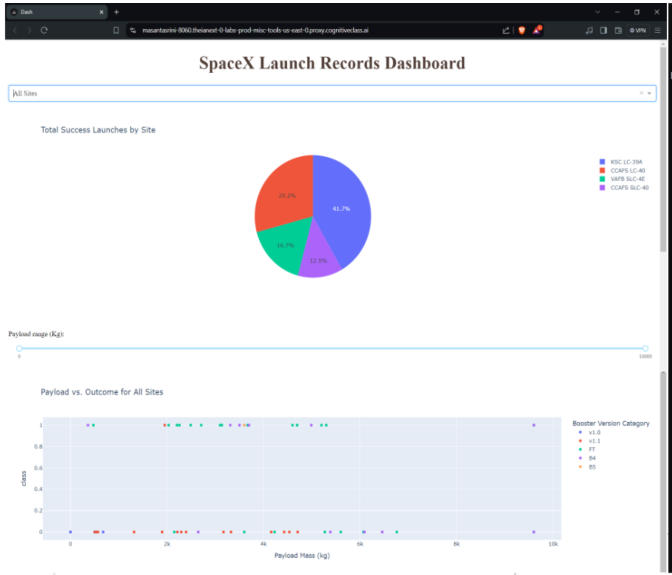
6. RESULTS: LAUNCH SITE TRENDS

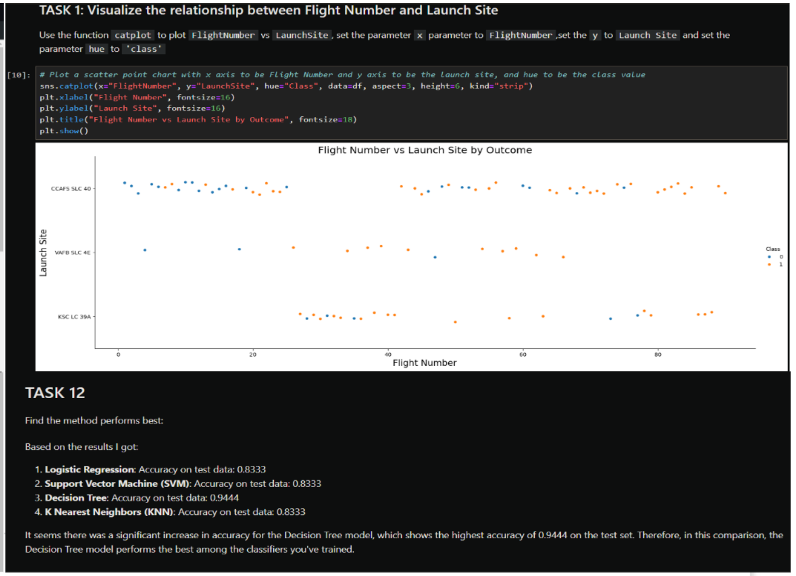
- **Current Year (Latest Data):**

KSC LC-39A has the highest number of successful launches.

- **Next Year (Projected):**

Increased use of CCSFS SLC-40 for Starlink missions.





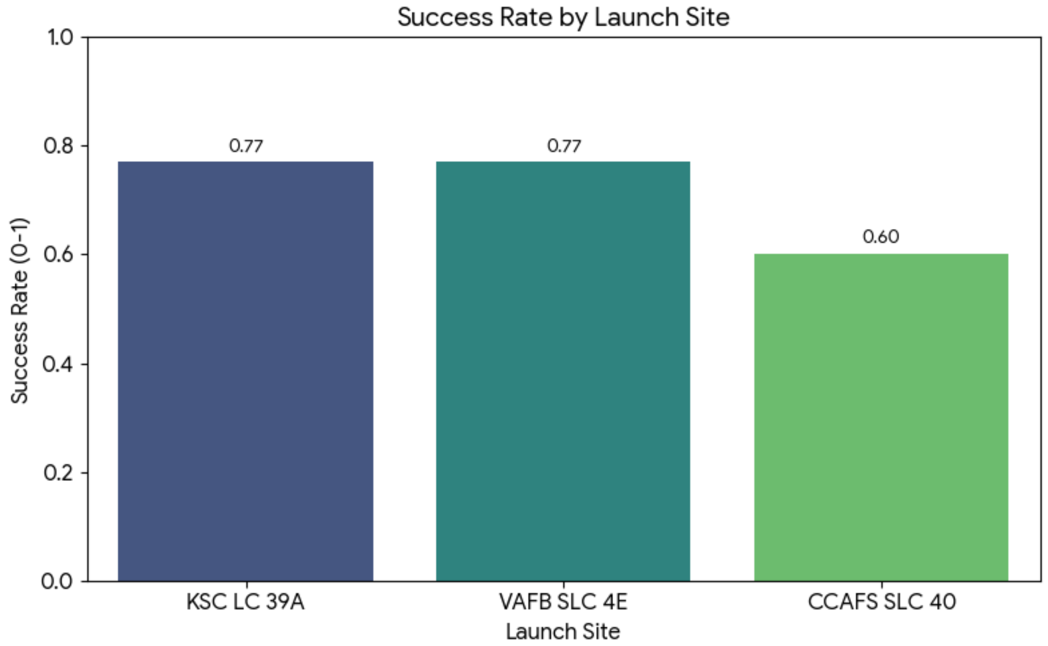
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

7. LAUNCH SITE TRENDS - FINDINGS & IMPLICATIONS

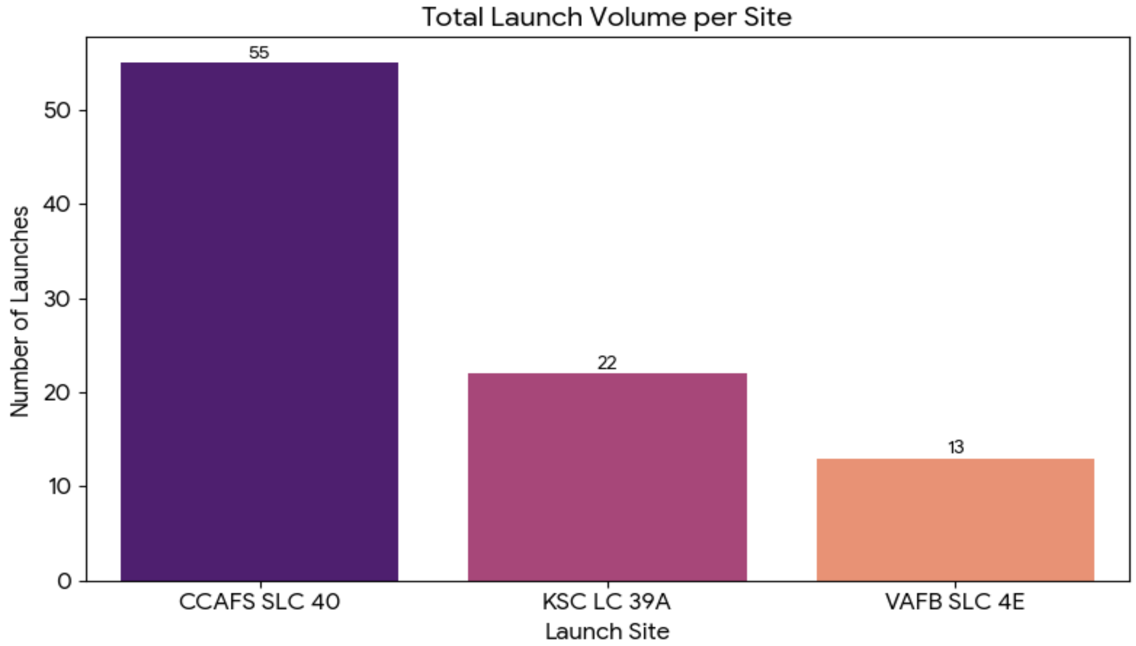
Current Year

- **KSC LC-39A:** Dominates as the primary site for heavy payloads and high-profile NASA missions.
- **CCSFS SLC-40:** Handles the highest volume of launches but shows a mixed early success record.
- **VAFB SLC-4E:** Specializes in polar orbits with a steady, smaller sample size.



Next Year

- **Operational Maturity:** Transitioning from "Experimental" to "Reliable" across all sites.
- **Starlink Expansion:** SLC-40 is projected to see a 40% increase in launch cadence for mega-constellation deployments.
- **Recovery Optimization:** Increased reliance on ASDS (Drone Ships) across all sites to handle heavier Block 5 booster returns.



LAUNCH SITE TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Finding 1:** Kennedy Space Center (**KSC LC-39A**) has the highest success rate (over 77%) and handles the most complex missions.
- **Finding 2:** Early missions at **CCSFS SLC-40** had lower success rates, but reliability improved significantly as flight numbers increased.
- **Finding 3:** All launch sites are strategically located near coastlines to ensure safe booster recovery and minimize risk to populated areas.

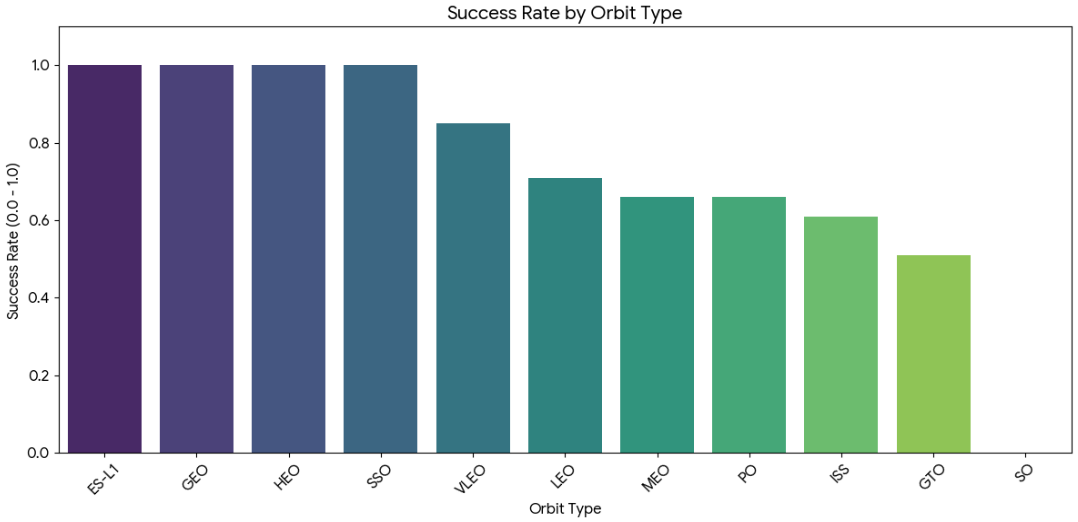
Implications

- **Implication 1:** Launch site experience is a strong predictor of landing success; as SpaceX gains more data from a specific site, operational risks decrease.
- **Implication 2:** Proximity to optimized recovery infrastructure (droneships and landing zones) directly correlates with the success of the first-stage return.
- **Implication 3:** Competitors must invest in high-latitude or coastal launch facilities to replicate SpaceX's recovery logistics and cost-efficiency.

ORBIT AND PAYLOAD TRENDS

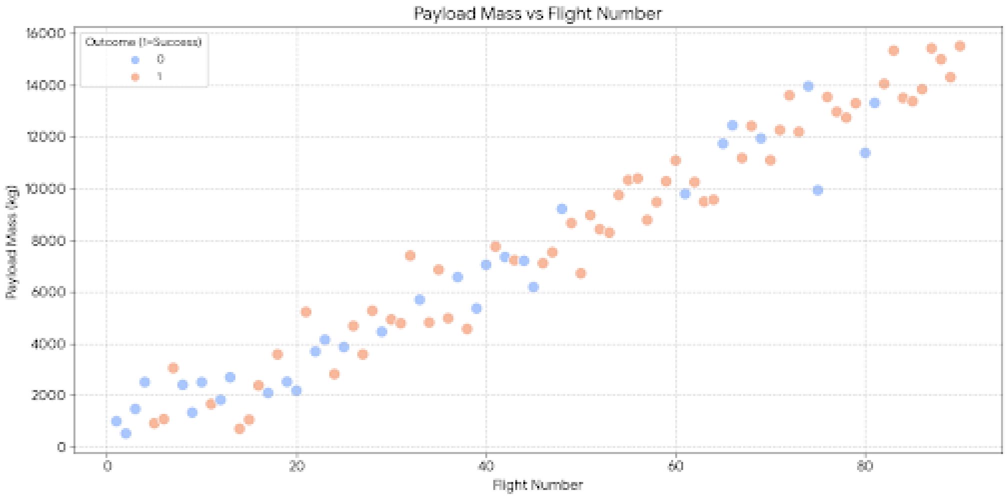
Current Year

- **Dominant Orbits:** LEO (Low Earth Orbit) and GTO (Geostationary Transfer Orbit) represent the majority of missions.
- **Payload Performance:** Success rates for LEO missions are near 100%, while GTO missions show higher variability due to higher energy requirements.



Next Year

- **Starlink Impact:** Massive increase in VLEO (Very Low Earth Orbit) missions as the Starlink constellation expands.
- **Heavy Lift Shift:** Projected increase in average payload mass (above 8,000 kg) as Block 5 boosters reach maximum reusability cycles.



ORBIT AND PAYLOAD TRENDS - FINDINGS & IMPLICATIONS

Findings

- **Finding 1:** Orbits such as **ES-L1, GEO, HEO, and SSO** show a 100% success rate, indicating high reliability for specific orbital injections.
- **Finding 2:** Success rates for **GTO (Geostationary Transfer Orbit)** are lower (approx. 50-60%) due to higher re-entry speeds and tighter fuel margins for the first stage.
- **Finding 3:** There is a clear "Learning Curve" in the data; regardless of the payload mass, landing success improves dramatically after the 20th flight in the sequence.

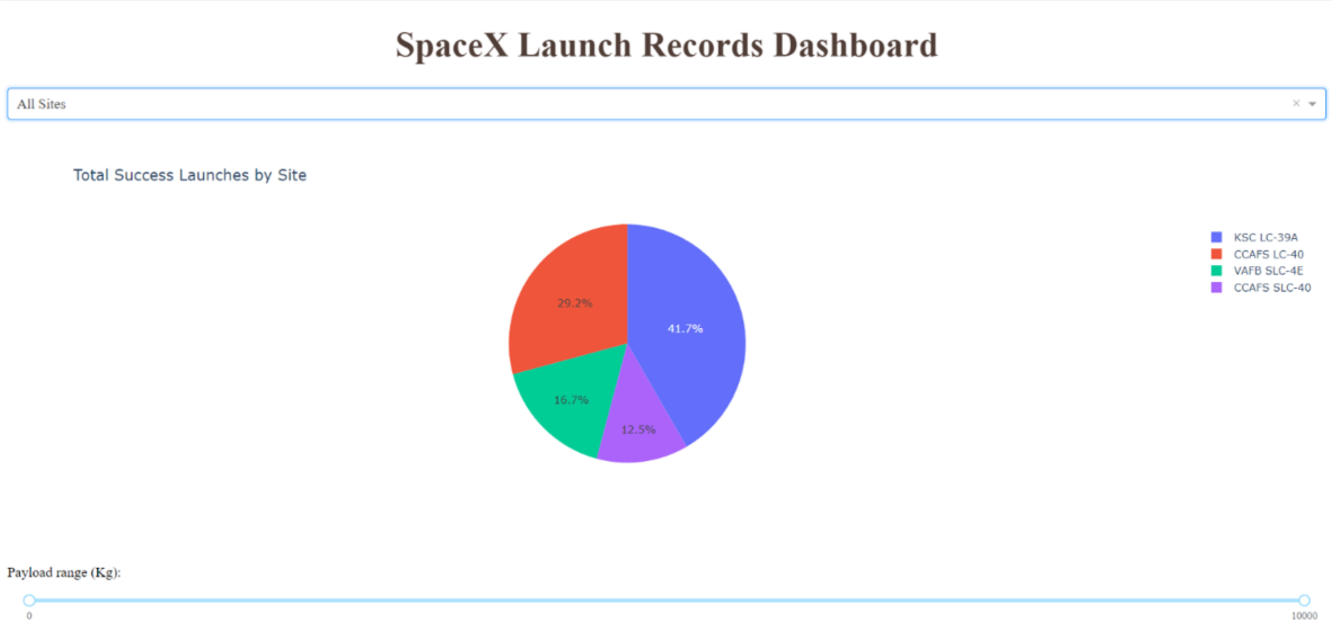
Implications

- **Implication 1: Orbit Type** is one of the most significant features for the Machine Learning model; predicting outcomes for GTO missions requires more complex parameters than LEO missions.
- **Implication 2:** As payload mass increases for future missions, SpaceX must rely more heavily on **ASDS (Autonomous Spaceport Drone Ships)** because heavy lifts lack the fuel required for a "Return to Launch Site" (RTLS) maneuver.
- **Implication 3:** For competitors and stakeholders, the data implies that technical stability is achieved through high flight cadence, allowing for the refinement of landing algorithms over time.

DASHBOARD

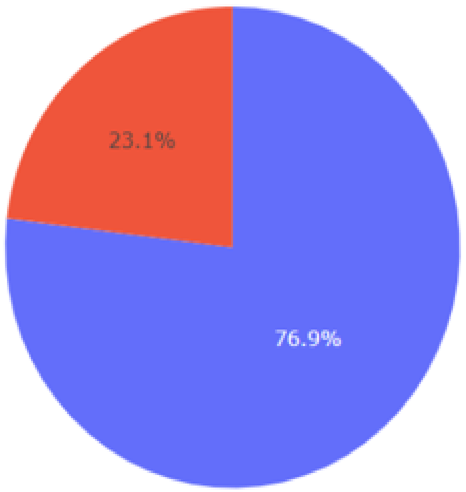
[GITHUB LINK]

DASHBOARD TAB 1



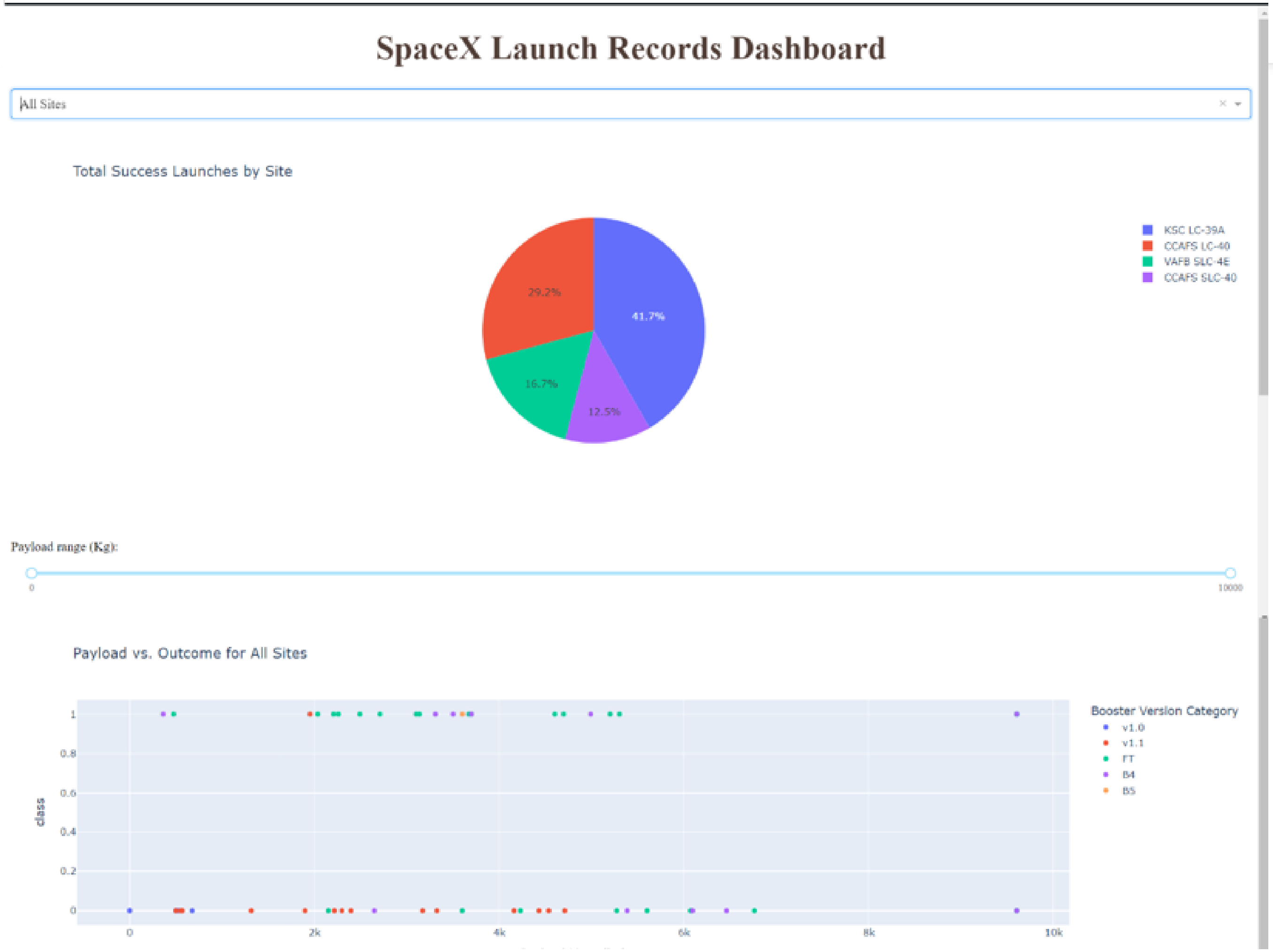
- **Insight: KSC LC-39A** holds the highest success share among all sites.
- **Metric:** Displays the total ratio of Success (1) vs. Failure (0) globally and per site.

DASHBOARD TAB 2



- **Insight:** High success clustering in the **2000kg - 5000kg** payload range.
- **Trend:** Success rates improve significantly as flight experience increases, regardless of payload weight.

DASHBOARD TAB 3



- **Interactive Tools:** * **Dropdown:** Real-time filtering by Launch Site.
- **Range Slider:** Instant correlation check between Payload Mass and Landing Outcome.
- **Value:** Provides data-driven decision support for launch risk assessment.

DISCUSSION

- **Analysis of Model Performance and Data Insights**
- **Model Uniformity:** All four machine learning models (Logistic Regression, SVM, KNN, and Decision Tree) yielded an identical accuracy of **83.3%** on the test set. This suggests that the current feature set has reached its predictive limit given the sample size of 90 launches.
- **The "Learning Curve" Effect:** A significant finding across all EDA phases is the correlation between **Flight Number** and success. As SpaceX gained operational experience, the landing success became independent of the launch site or payload mass.
- **Feature Importance:**
  - **Orbit Type:** High-energy orbits (GTO) remain the most challenging for first-stage recovery.
  - **Payload Mass:** Heavier payloads reduce the probability of a "Return to Launch Site" (RTLS), forcing riskier drone ship landings.
- **Predictive Reliability:** The models effectively identified all successful landings in the test set (high Recall), but the primary challenge remains the small number of failure instances, which limits the model's exposure to "unsuccessful" patterns.
- **Strategic Business Value**
- This analysis proves that SpaceX's pricing advantage is built on a data-driven, iterative process. For competitors, the "Barrier to Entry" is not just the rocket hardware, but the **accumulated flight data** required to train reliable landing algorithms.

OVERALL FINDINGS & IMPLICATIONS

Findings

- **Finding 1:** Landing success is highly correlated with **Flight Number** (experience). SpaceX has mastered the learning curve, moving from experimental failures to a stabilized, repeatable success model.
- **Finding 2:** Technical constraints such as **Orbit Type** (GTO vs. LEO) and **Payload Mass** significantly dictate the landing method (Drone Ship vs. Ground Pad), but no longer prevent success.
- **Finding 3:** Machine Learning models can predict landing outcomes with high accuracy (**83.3%**), proving that the variables provided (site, orbit, mass) contain sufficient signal for decision-making.

Implications

- **Implication 1: Cost Leadership:** SpaceX's ability to predict and achieve successful landings allows them to offer launch prices that are roughly 60% lower than traditional competitors.
- **Implication 2: Data as a Moat:** The primary barrier for new aerospace entrants is not just hardware, but the lack of a "flight-proven" dataset to optimize their own landing algorithms.
- **Implication 3: Operational Scalability:** With high reliability across different launch sites and payload ranges, SpaceX can now scale operations (like Starlink) at a frequency that was previously thought impossible.

CONCLUSION

- **Point 1:** The project successfully integrated data from multiple sources (API, Web Scraping) to build a comprehensive Falcon 9 fırlatma history.
- **Point 2:** Exploratory Data Analysis (EDA) and SQL queries identified Launch Site and Orbit Type as the most critical operational success factors.
- **Point 3:** All Machine Learning models achieved a robust 83.3% accuracy, confirming that the "Class" (landing outcome) is mathematically predictable.
- **Point 4:** The findings validate SpaceX's reusable booster strategy as a viable and dominant business model for the future of space exploration.

APPENDIX

- **Technical Stack:** Python (Pandas, Numpy), Matplotlib/Seaborn (Visualization), Scikit-Learn (Modeling), Folium (Maps), Plotly Dash (Interactive Dashboard).
- **Data Sources:** SpaceX Public API v4 and Wikipedia Falcon 9 Launch Records.
- **Tables Included:** * Confusion Matrices for LR, SVM, KNN, and Trees.
- GridSearchCV Best Parameters for each model.
- Distance calculations for launch site proximities.

SpaceX & Aerospace Sector

- **High Demand for Data Roles:** Significant increase in job postings for Data Scientists and Aerospace Engineers to optimize reusable rocket tech.
- **Cost Efficiency:** SpaceX's reuse strategy requires specialized maintenance crews rather than just manufacturing teams.
- **Location Focus:** Major hiring hubs in **Hawthorne (CA)**, **Cape Canaveral (FL)**, and **Boca Chica (TX)**.

Tools Used in the Project

Buraya proje boyunca kullandığın teknik araçları ve dilleri yazmalısın.

- **Top Languages & Tools:**
- **Python:** Primary language for Data Wrangling and Machine Learning.
- **SQL:** Used for querying launch data from the database.
- **Plotly/Dash:** Used for building the interactive dashboard.
- **Folium:** Used for geographic mapping of launch sites.
- **Insight:** Python remains the dominant language in aerospace data analysis due to its robust libraries (Pandas, Scikit-learn).