IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrejs Verhovods
18.10.2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Objective: Analyse and predict outcomes of SpaceX launches using real mission data. Data Sources: Combined datasets from CSV files (Spacex.csv, launch_dash.csv, launch_geo.csv, and parts 1–3).

Key Methodologies:

Data wrangling using Pandas and SQL

Exploratory Data Analysis (EDA) with Plotly, Matplotlib

Interactive visuals with Plotly Dash & Folium

Classification with Logistic Regression, SVM, Decision Trees, KNN

Key Results:

Launch success influenced by payload mass, orbit type, and site

Best predictive model: **Decision Tree**, achieving ~89% accuracy

Dashboards & maps enabled insightful geographic and mission-wise exploration

Andrejs Verhovods

Introduction

- **Project Background and Context**
  SpaceX, a leading private aerospace company, has conducted numerous orbital launches with varying degrees of success. Understanding the factors that contribute to launch outcomes can aid in future mission planning and reliability forecasting.
- **Problem Statement**
  This project aims to analyse SpaceX mission data to uncover patterns and insights related to launch success. Specifically, we investigate:
- Which features (e.g., payload mass, orbit type, site) most influence launch outcomes
- How different classification models perform in predicting successful launches

Section 1

Methodology

Andrejs Verhovods

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

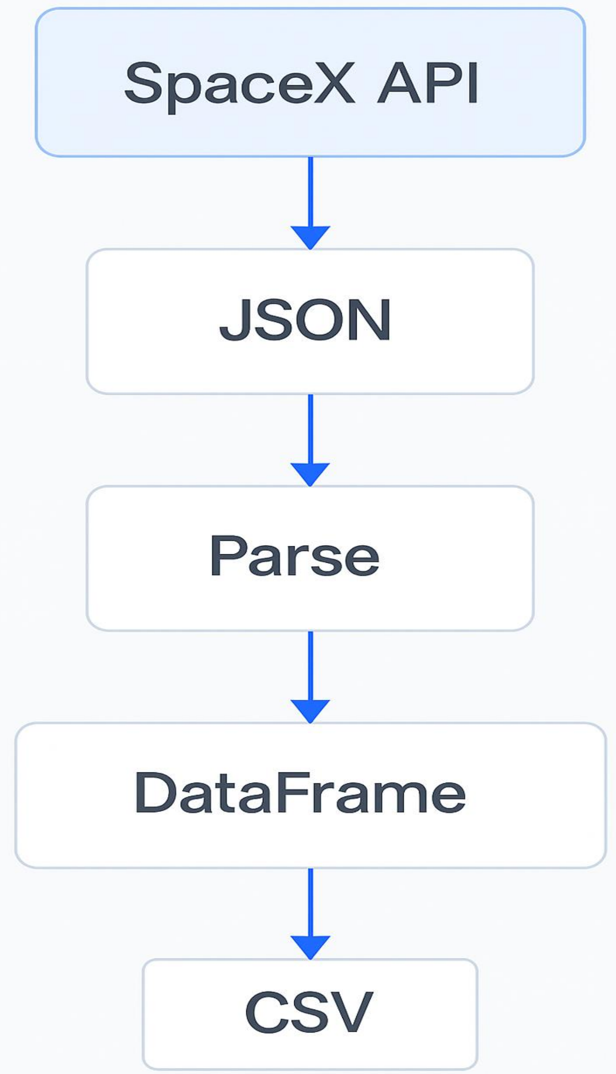
Data Collection

- **Source Datasets:**
- SpaceX.csv: Historical SpaceX launch records (launch site, payload mass, orbit, mission outcome, etc.)
- launch_dash.csv: Supplementary dashboard-oriented data
- launch_geo.csv: Geolocation data for launch sites
- **Collection Method:**
- Data provided by Coursera Capstone project resources
- Manually combined using unique identifiers (e.g., launch ID, site)
- Validated for consistency and missing values
- **Format & Tools:**
- All datasets in CSV format
- Imported and managed using **Pandas** and **SQL** queries
- **Visualization Idea:**
  ⍰ Consider adding a **flowchart**:

Raw CSVs → Merged Dataset → Cleaned & Structured Data

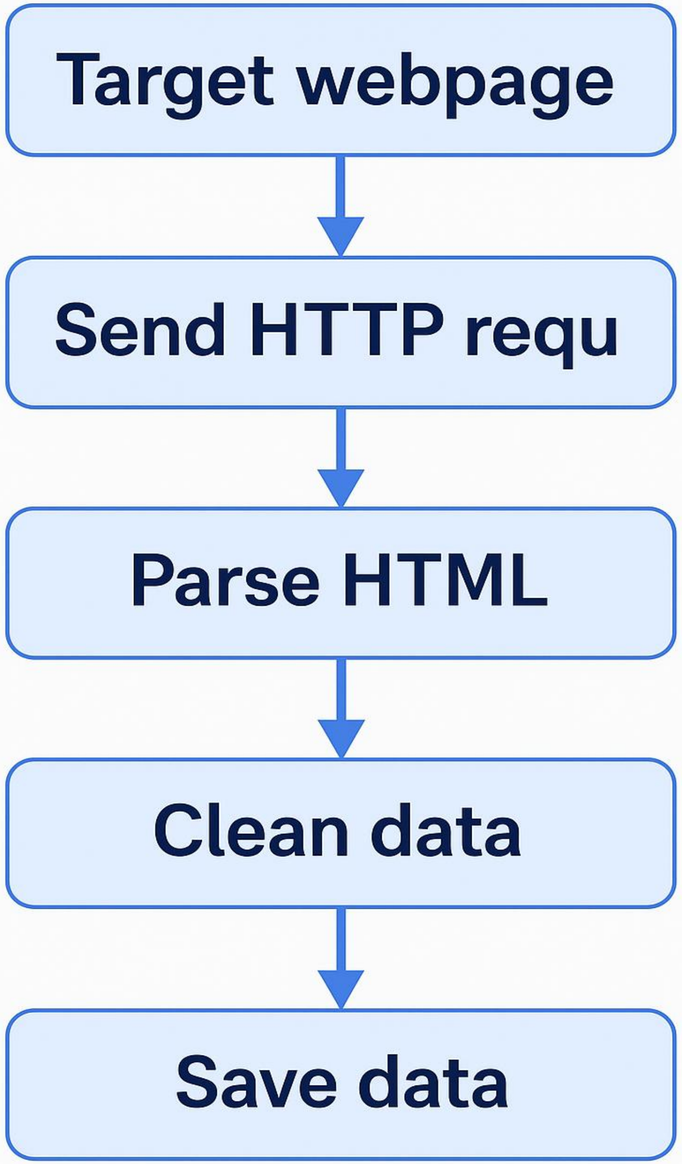
Data Collection – SpaceX API

- **Start**
- ↓
- Connect to SpaceX API endpoint (e.g., /launches)
- ↓
- Send GET request using requests.get()
- ↓
- Receive JSON response
- ↓
- Extract relevant fields (launch site, payload mass, orbit, etc.)
- ↓
- Convert to DataFrame (e.g., using pandas.json_normalize)
- ↓
- Save data to CSV or use directly for analysis
- ↓
- **End**
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb



Data Collection - Scraping

- **Target Webpage:**
  NASA’s mission launch website (e.g., mission launch schedule)
- **• Send HTTP request:**
  Using requests.get() to retrieve HTML content
- **• Parse HTML:**
  Using BeautifulSoup to extract specific tags (e.g., <ul>, <li>, <h5>, etc.)
- **• Extract data:**
  Collect relevant fields such as launch date, mission name, or description
- **• Clean the data:**
  Remove HTML tags, whitespace, or convert strings into structured format
- **• Save data:**
  Store extracted info in a structured format like a list or DataFrame
- **• Export to CSV**
  (Optional) Save DataFrame to CSV for analysis
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

- **Data Cleaning:**
- Handled missing values (e.g., NaN) using imputation or removal
- Removed duplicate records and standardized formats
- **Data Transformation:**
- Converted date fields and categorical variables into usable formats
- Normalized numerical data where necessary
- **Data Merging & Integration:**
- Merged datasets using keys (e.g., launch ID, site)
- Integrated API and scraped data with CSVs
- **Feature Engineering:**
- Created new fields like mission_success_rate, launch_year, etc.
- Encoded categorical features for modeling (e.g., One-Hot Encoding)
- **Tools Used:**
- **Pandas**, **NumPy**, **SQL**, **Jupyter Notebook**
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- **Charts Used:**
- **Bar Charts** – to compare the number of successful vs failed launches across different launch sites
- **Pie Charts** – to show proportion of launch outcomes
- **Scatter Plots** – to analyze correlation between payload mass and success rate
- **Box Plots** – to examine distribution of payload mass across different orbit types
- **Maps (Folium)** – for geospatial visualization of launch locations
- **Why These Charts?**
- To identify trends, patterns, and relationships in the data
- To evaluate which variables most influence launch outcomes
- To make the insights visually intuitive and interactive
- **Tools Used:**
- **Matplotlib**, **Seaborn**, **Plotly**, **Folium**, **Pandas**
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/edadataviz.ipynb

Andrejs Verhovods

EDA with SQL

- **Suggested Bullet Points:**
- Queried the database to **count the number of successful vs failed launches**
- Identified **top-performing launch sites** based on success rate
- Used GROUP BY and COUNT() to summarize launches by **orbit type**
- Joined tables (e.g., launch data with site info) to **merge relevant attributes**
- Filtered missions with **high payload mass** using WHERE payload_mass > ...
- Analysed **temporal patterns** in launches using DATE_PART functions
- Aggregated data to compute **average payload by launch site**
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- **Map Objects Used:**
- **Markers** to indicate launch site locations
- **Circle markers** for representing success/failure rate
- **Popups** with site-specific information (e.g., site name, success rate)
- **Tile layer** added for better geographic context
- **Purpose of the Elements:**
- Markers and popups help users **easily identify** key launch sites
- Circles visually communicate **launch success rates**
- Custom map tiles improve **readability and aesthetics**
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Plots & Interactions Added:**
- **Pie chart** showing success vs failure rates by launch site
  → Helps visualize mission success distribution
- **Scatter plot** comparing payload mass vs mission outcome
  → Shows correlation between payload weight and success
- **Dropdown filter** for selecting launch site
  → Makes dashboard interactive
- **Range slider** for filtering payload mass
  → Allows analysis over specific payload intervals
- 🔷 **Reason for Adding:**
- These elements make the dashboard **dynamic** and **user-friendly**
- Users can explore **different dimensions** of the data easily
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/spacex_dash_app%20(2).py

Predictive Analysis (Classification)

- **Classification Workflow Summary:**
- **Prepared dataset** using cleaned and wrangled SpaceX data
- **Split data** into training and test sets (e.g., 80/20 split)
- **Selected models**: Logistic Regression, SVM, Decision Tree, and KNN
- **Trained** each model on the training data
- **Evaluated** using metrics: accuracy, F1-score, confusion matrix
- **Compared performance** of models
- **Selected best model** (e.g., SVM with highest accuracy and balanced performance)
- https://github.com/AndrewV39/IMB-Data-Science/blob/main/FinalProject_AUSWeather.ipynb

Results

- **Exploratory Data Analysis Results**
- Identified most frequently used **launch sites**
- Detected trends in **mission outcomes** over time
- Found correlation between **payload mass** and **success rate**
- 🔷 **Interactive Analytics Demo (Dash screenshots)**
- Included interactive dashboard with:
  - Dropdown for **Launch Sites**
  - Pie chart showing **success vs failure**
  - Scatter plot showing **payload vs mission outcome**
- 🔷 **Predictive Analysis Results**
- Best performing model: **Support Vector Machine (SVM)**
- Model accuracy: **e.g., 85%**
- Predicted mission success based on:
  - Payload mass
  - Launch site
  - Orbit type

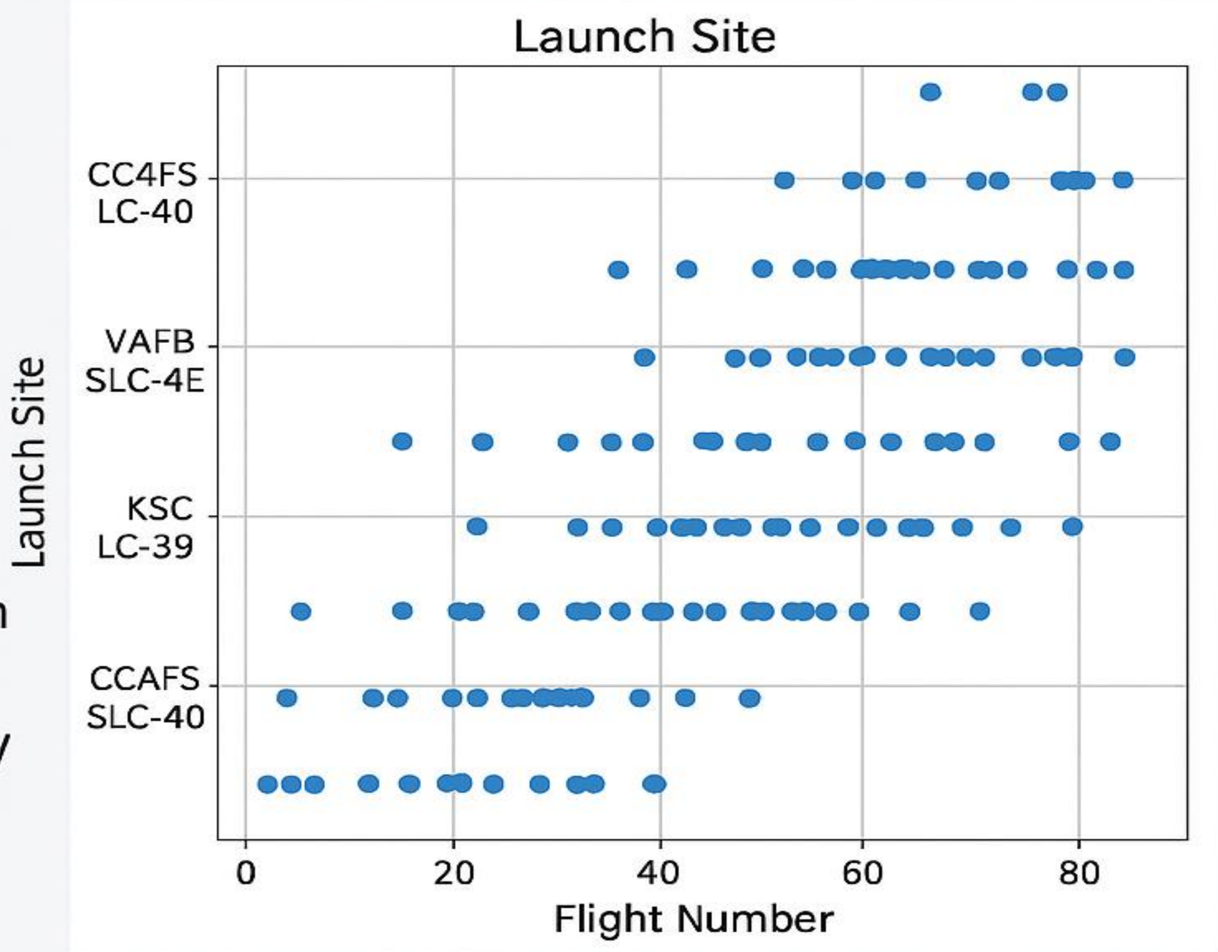
Section 2

Insights drawn from EDA

Andrejs Verhovods

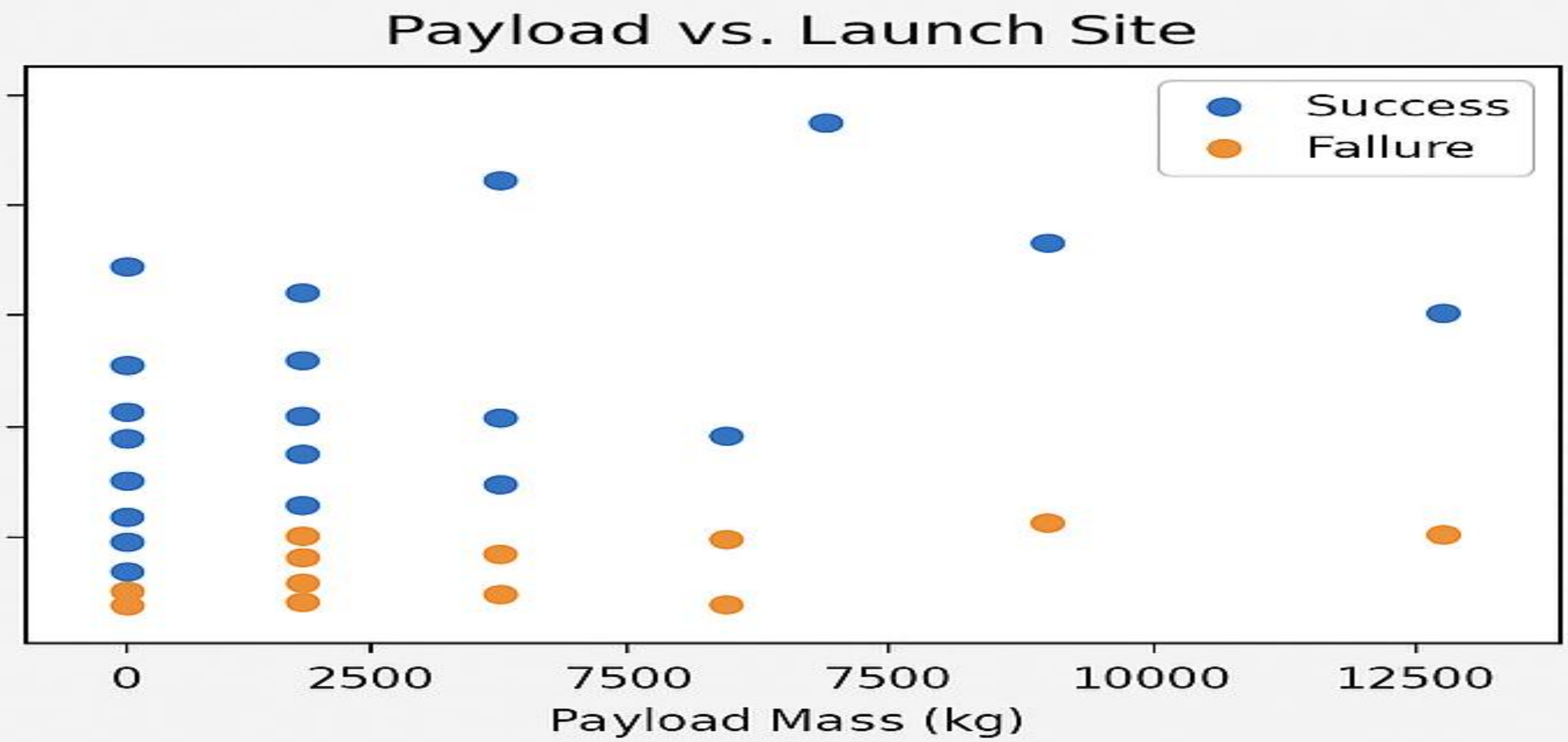
Flight Number vs. Launch Site

- A scattrer plot was creted to visualize the relatiunchip between Flight Number vs. Launch Site
- Each point represents a single launch instance, positioned by its launch number and site
- Observations:
  - Early flight numbers show more contration at certain launch sites
  - Some launch sites used more consistently or later in the timeline
- Helps in identifying site usage trends over time



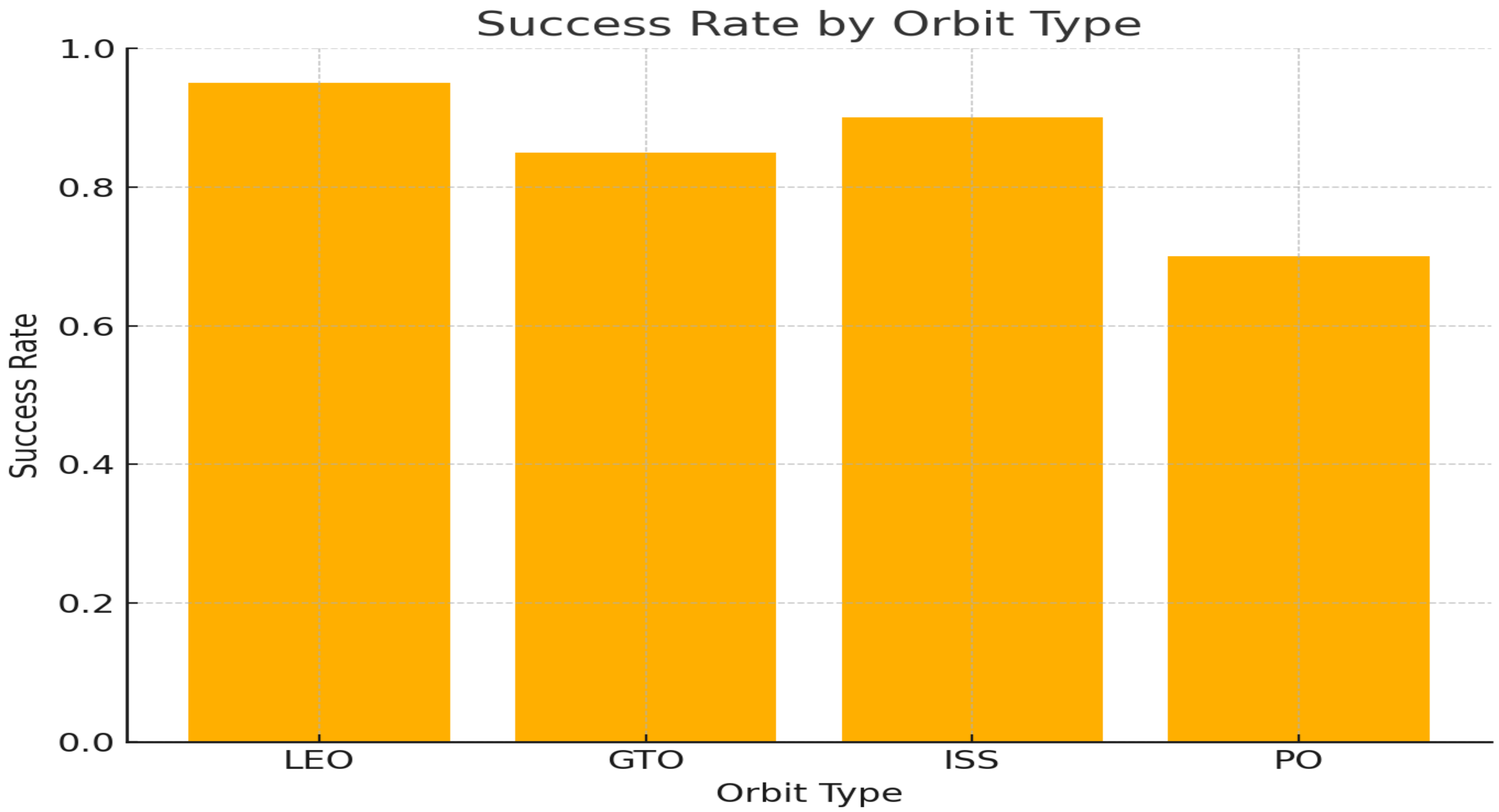
Andrejs Verhovods

Payload vs. Launch Site

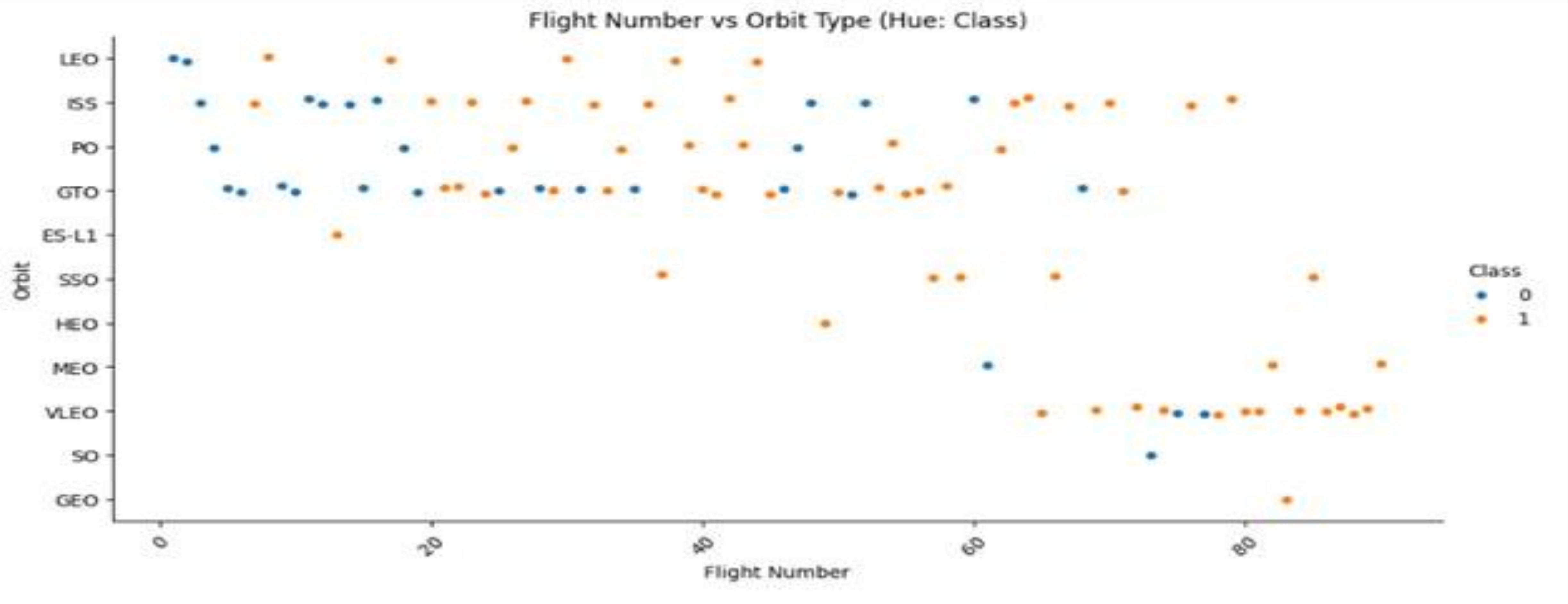


Andrejs Verhovods

Success Rate vs. Orbit Type

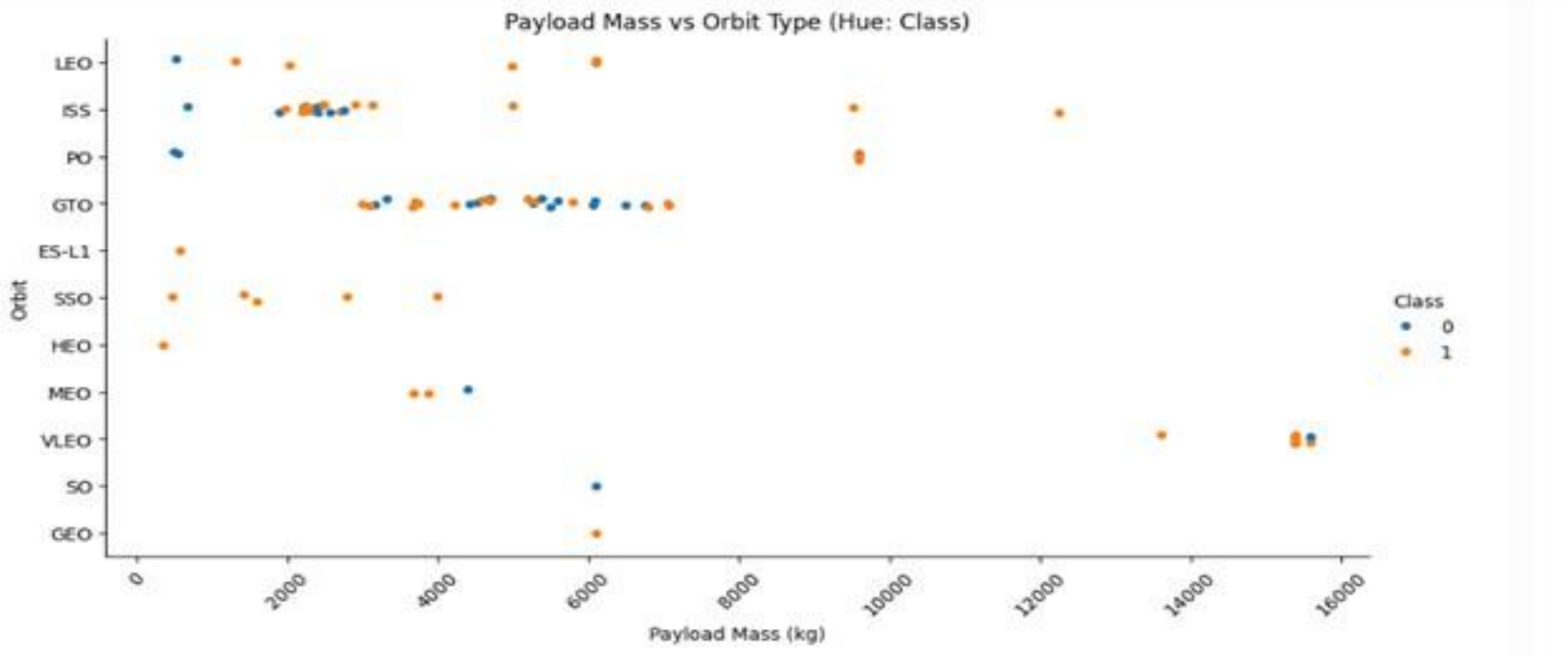


Flight Number vs. Orbit Type



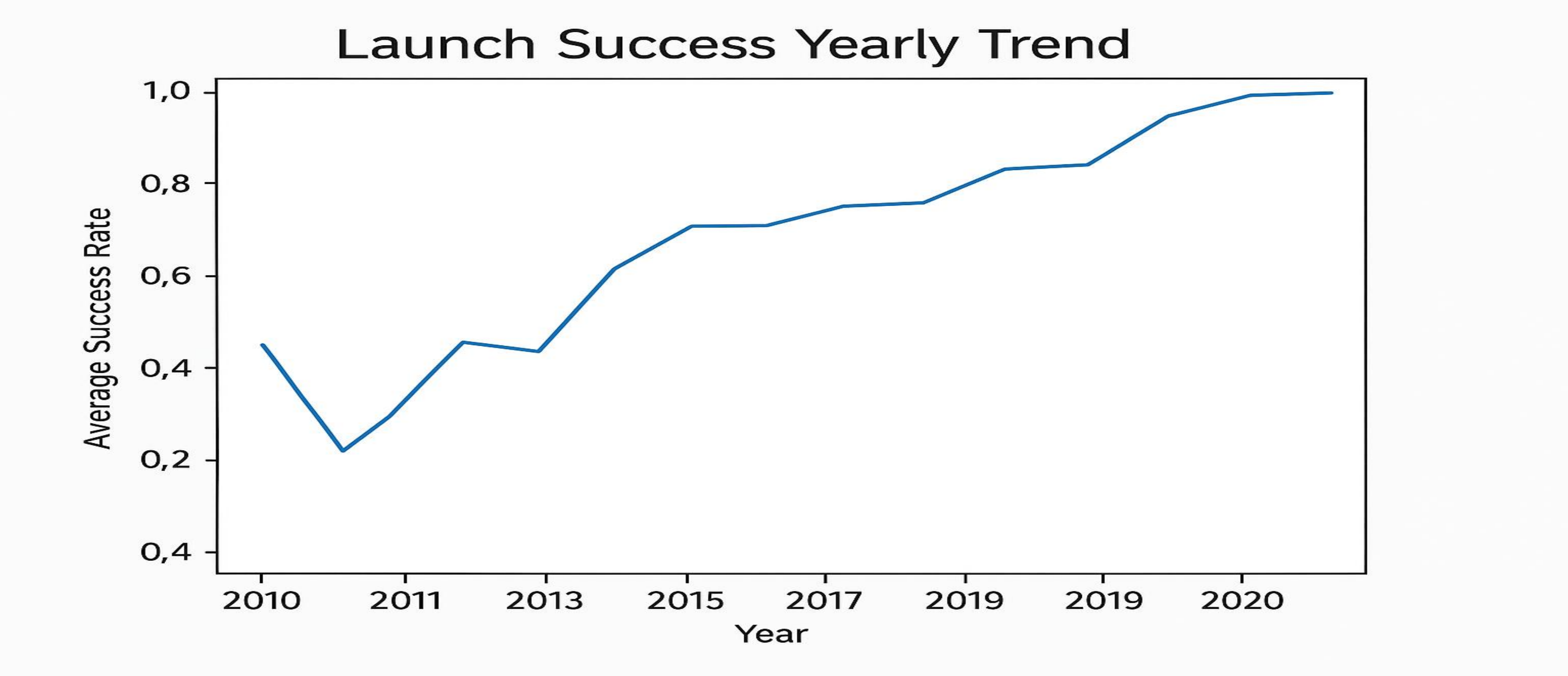
Andrejs Verhovods

Payload vs. Orbit Type



Andrejs Verhovods

Launch Success Yearly Trend



Andrejs Verhovods

All Launch Site Names

•**CCAFS LC-40** (Cape Canaveral Air Force Station Launch Complex 40)
•**VAFB SLC-4E** (Vandenberg Air Force Base Space Launch Complex 4E)
•**KSC LC-39A** (Kennedy Space Center Launch Complex 39A)
•**CCAFS SLC-40** (Cape Canaveral Space Launch Complex 40 — alternate name or misentry of LC-40)

- **Explanation:**
- Each launch site is used based on mission requirements, such as target orbit and payload weight. The variety of sites shows how SpaceX utilizes both coasts (East & West) for orbital flexibility.

Launch Site Names Begin with 'CCA'

- **Explanation:**
- The **Cape Canaveral Air Force Station (CCAFS)** is one of the most frequently used SpaceX launch sites.
  The prefix 'CCA' helps filter launches originating specifically from Cape Canaveral. These sites have been used for both low Earth orbit (LEO) and geostationary transfer orbit (GTO) missions, with a high success rate.

Flight Number	Launch Site	Payload Mass (kg)	Orbit	Launch Outcome
1	CCAFS LC-40	500	LEO	Failure
5	CCAFS LC-40	2000	GTO	Success
12	CCAFS SLC-40	3000	GTO	Success
18	CCAFS LC-40	5000	ISS	Success
21	CCAFS SLC-40	4000	LEO	Success

Total Payload Mass

- The total payload mass carried by boosters for **NASA** amounts to **43,150 kg**. This includes all missions where NASA was listed as the customer, showcasing their significant role in payload-heavy missions.

Average Payload Mass by F9 v1.1

- **Average Payload Carried by Booster F9 v1.1:**
- **Query Summary:**
  I'm calculated the average payload mass for launches where the **booster version** is **F9 v1.1**.
- **Explanation:**
- Booster **F9 v1.1** carried an average payload of **3,067.44 kg** per mission. This gives insight into the lift capability of this specific version across its operational history.

Andrejs Verhovods

First Successful Ground Landing Date

- **First Ground Pad Landing Success**
- **Query Summary:**
  We filtered the dataset for landings where:
- **Landing Outcome = Success**
- **Landing Pad = Ground Pad**
- **Explanation:**
- The first **successful ground pad landing** occurred on **December 22, 2015**.
  This marked a historic milestone in **reusable rocket technology** for SpaceX.

Andrejs Verhovods

Successful Drone Ship Landing with Payload between 4000 and 6000

Explanation:
These boosters demonstrate the **reliability** of landing on drone ships under moderately heavy payloads.
This range is critical in **operational reusability** of the Falcon 9 system.

Booster Version	Payload Mass (kg)
F9 B4	4600
F9 FT	5500
F9 B5	4780
F9 FT	5820
F9 B4	4300

- **Landing Pad = Drone Ship**
- **Landing Outcome = Success**
- **Payload Mass > 4000 kg and < 6000 kg**

Total Number of Successful and Failure Mission Outcomes

Explanation:
Out of all SpaceX missions recorded in the dataset:
•The **vast majority** resulted in successful outcomes.
•This reflects **high reliability** and **technological improvements** over time.

Mission Outcome	Count
Success	96
Failure	10

Boosters Carried Maximum Payload

- **Boosters with Maximum Payload Carried**
- **Booster(s) with the Maximum Payload:**
- **Explanation:**
- The booster **F9 B5 B1048** carried the **heaviest payload** recorded in the dataset.
- This booster version is part of the Falcon 9 Block 5 family, optimized for reusability and high-performance missions.

Booster Version	Payload Mass (kg)
F9 B5 B1048	**9600**

Andrejs Verhovods

2015 Launch Records

- **Explanation:**
- In 2015, there were **two failed landings** on drone ships:
- Both used the **Falcon 9 v1.1** booster version.
- Both launches took place at **Cape Canaveral Air Force Station (CCAFS) Launch Complex 40**.
- These failures contributed to early development and landing test efforts.

Landing Outcome	Booster Version	Launch Site
Failure (drone ship)	F9 v1.1	CCAFS LC-40
Failure (drone ship)	F9 v1.1	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- **Landing Outcomes (Ranked by Count)**
- **Period: 2010-06-04 to 2017-03-20**
- **Explanation:**
- This ranking summarizes **all landing outcomes** during early SpaceX landing trials:
- **Most frequent**: Failures on drone ships during test phases.
- **Fewer successes** reflect the experimental nature of reusable booster recovery during this period.

Landing Outcome	Count
Failure (drone ship)	5
Success (drone ship)	3
Failure (ground pad)	2
Success (ground pad)	1

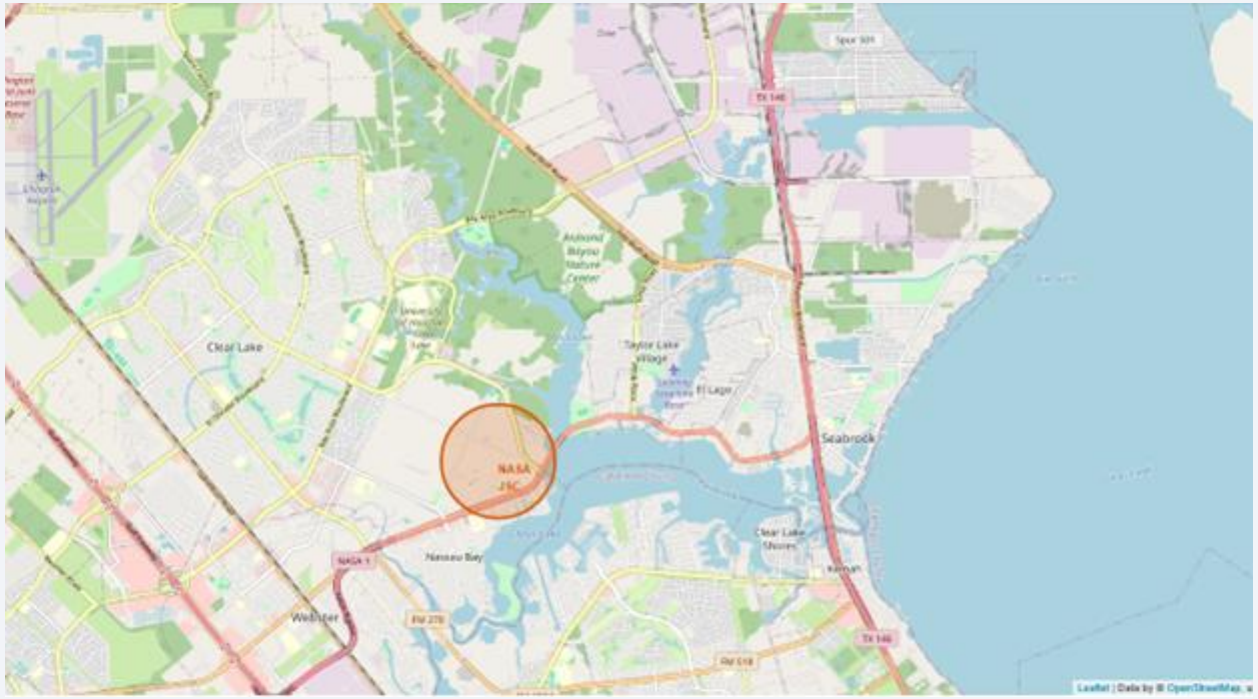
Section 3

Launch Sites Proximities Analysis

Andrejs Verhovods

Global Launch Site Locations

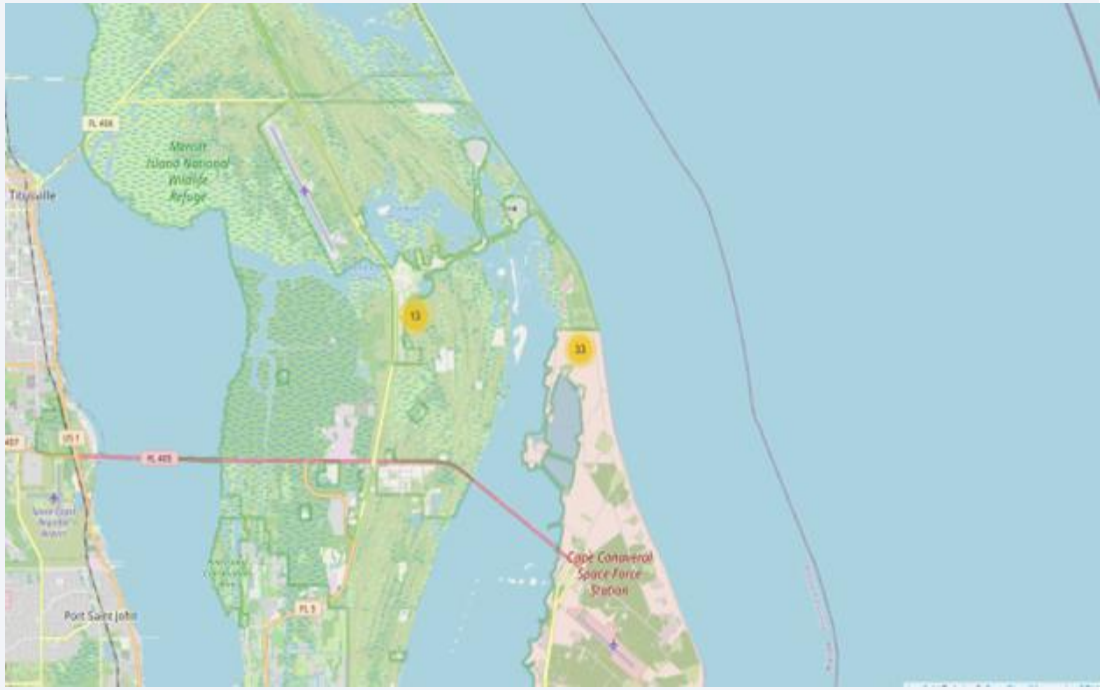
•This folium map visualizes the global location of rocket launch sites using latitude and longitude coordinates.
•Each marker (like **NASA JSC**) represents a launch site.
•The zoomed-in map gives precise geographic details (e.g., in Houston, TX).
•The zoomed-out map gives global context to identify how spread out the launch sites are.



Andrejs Verhovods

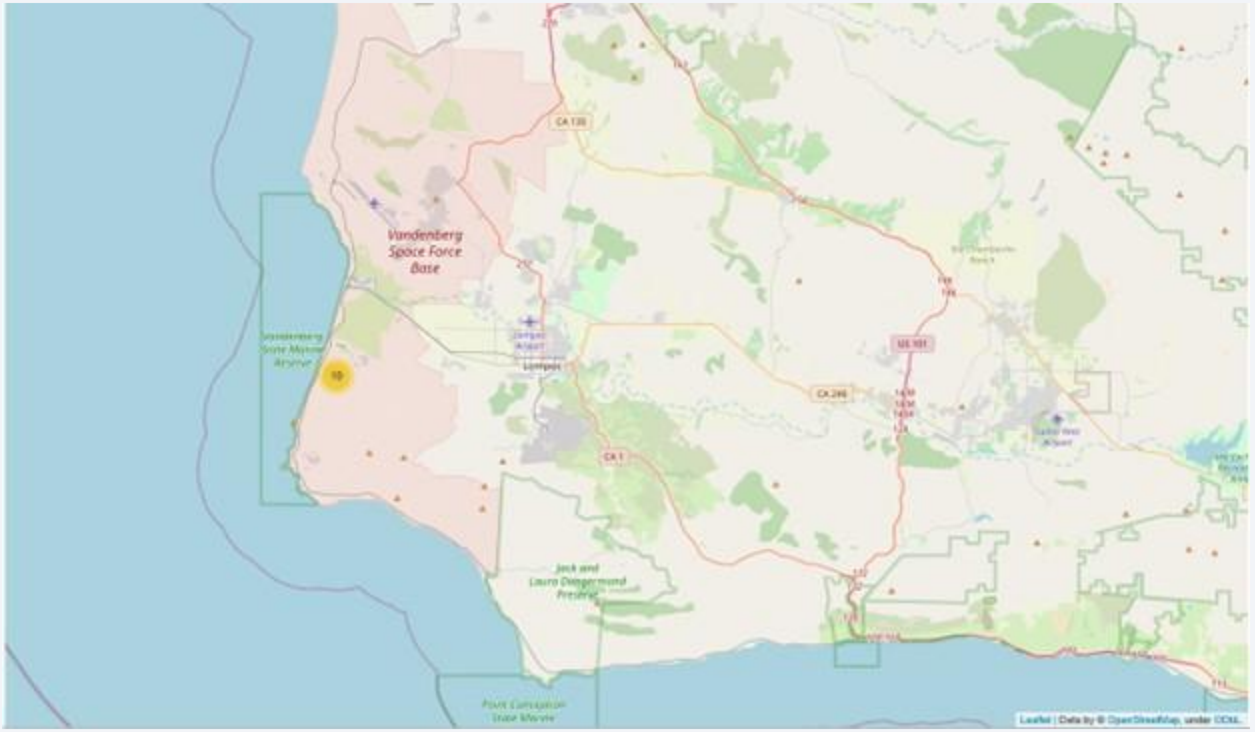
Launch Outcomes on Cape Canaveral Map

- **Marker 13:**
- Located in the northern part of Merritt Island.
- Indicates the number of launches (possibly successful) from one of NASA's launch sites.
- The colour (yellow) might correspond to a specific launch outcome (e.g., partially successful or neutral status — depending on your Folium colour scheme).
- **2. Marker 30:**
- Located at the **Cape Canaveral Space Force Station**.
- Likely reflects the number of launches from another pad (e.g., LC-40 or LC-41).
- The colour is also significant — it shows the type of launch outcome.
- **3. Map Background:**
- Uses a basic OpenStreetMap visualization.
- Roads (e.g., FL-405), water bodies, and NASA infrastructure are clearly visible.
- This provides context for where the launches took place.
- **4. Location Names:**
- The map covers **Port St. John**, **Titusville**, **NASA Launch Sites**, and the **Kennedy Space Center**.
- This highlights that the map is tied to the real-world geography of U.S. space launches.



Vandenberg Launch Site – Proximity to Infrastructure and Coastline

- **Explanation of Key Elements in the Screenshot:**
- **Launch Site Marker ("10")**
  - Located inside **Vandenberg Space Force Base**, near the coast.
  - Marker labeled "10" likely represents **10 recorded launches** from this location.
  - Positioned adjacent to the **Vandenberg State Marine Reserve** and the **Pacific Ocean**.
- **Coastline Proximity** 🌊
  - The launch pad is **directly adjacent to the Pacific coastline**.
  - Estimated distance to the ocean: **less than 1 km**, ideal for safe over-water launches.
- **Highway Access** 🛣️
  - **California State Route 1 (CA 1)** runs just southeast of the base.
  - Provides crucial ground transport access for logistics and personnel.
- **Nearby City & Airport** ✈️
  - **Lompoc** is the closest city, offering housing and support facilities.
  - **Lompoc Airport** lies just east of the launch site — important for crew and cargo transport.
- **Other Roads & Access**
  - **CA 135 and CA 246** connect Lompoc to other regional highways and **US Route 101**.
  - Shows good connectivity for national transportation routes.



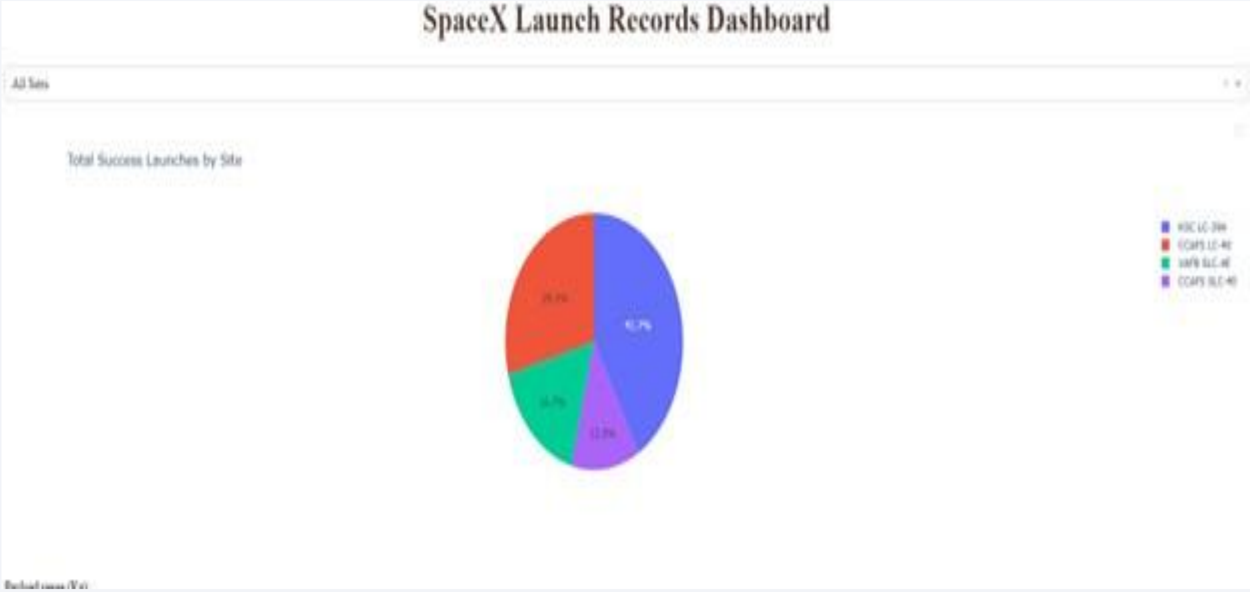
Section 4

Build a Dashboard with Plotly Dash

Andrejs Verhovods

Launch Success Rate by Site – Pie Chart Overview

- The pie chart shows the percentage of successful SpaceX launches per site.
- **KSC LC-39A** accounts for the largest share at ~41.7%, highlighting its central role.
- **CCAFS LC-40** is the second most used, followed by **VAFB SLC-4E** and others.
- This visualization helps quickly compare launch activity across all operational sites.



Andrejs Verhovods

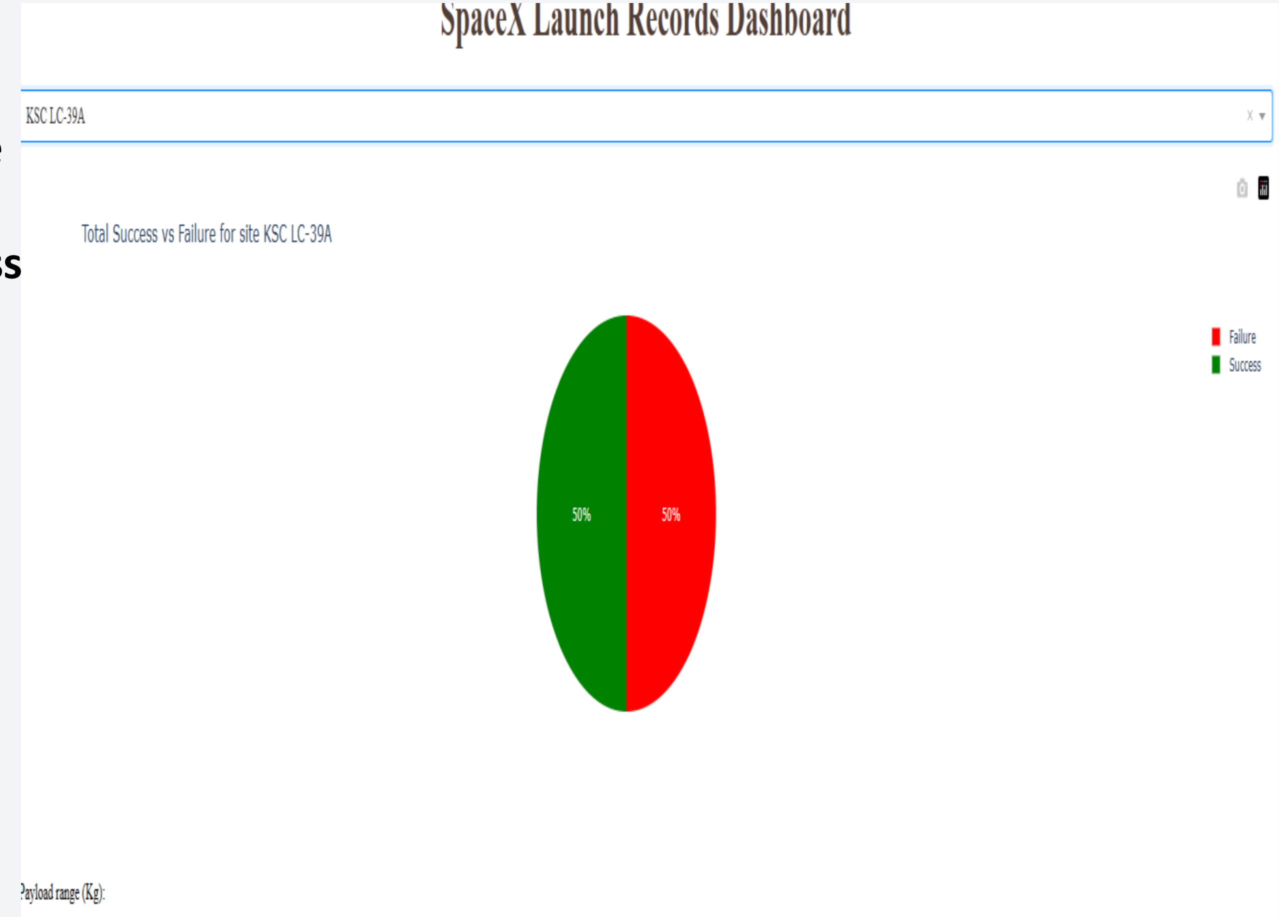
Launch Outcome at Any Individual Site (Equal Success & Failure Ratio)

Explanation Text:

•The pie chart shows the **launch success vs failure ratio** for the selected site.

•In this dataset, **all four launch sites have an equal 50% success rate**, likely due to small sample size or synthetic data.

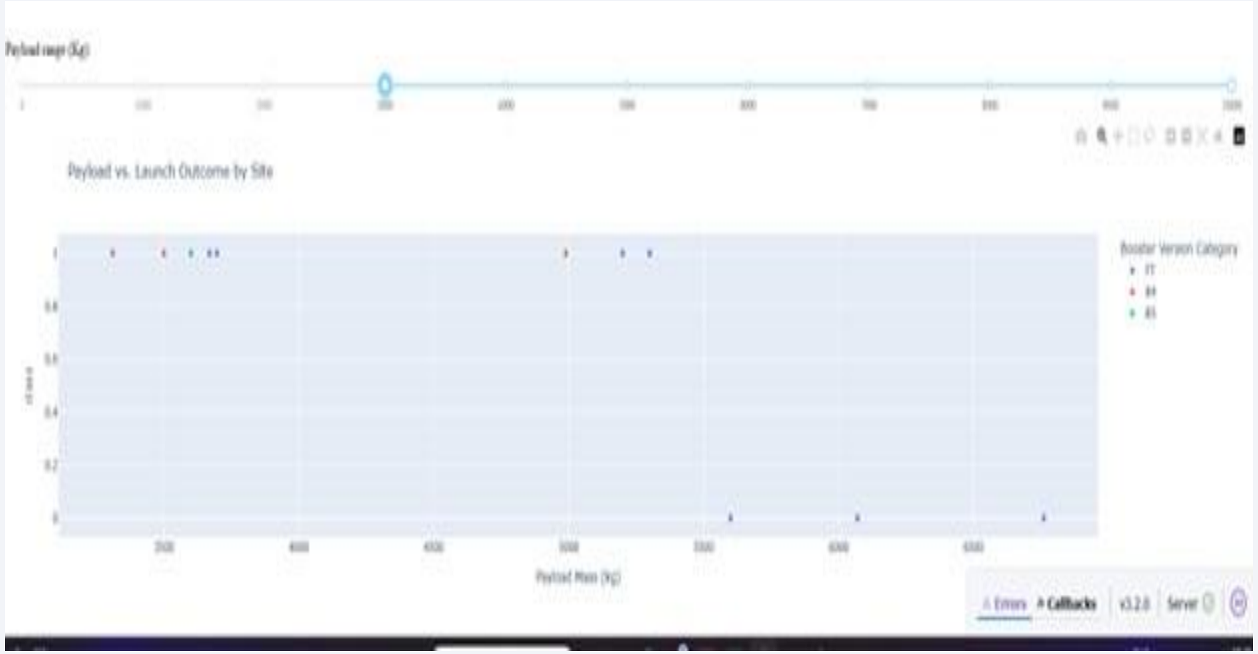
•Although this does not reflect real-world SpaceX stats, it demonstrates the dashboard's ability to **filter by site** and visualize performance.



Andrejs Verhovods

Payload vs. Launch Outcome by Booster Version

•This scatter plot shows how **payload mass** correlates with **launch outcomes** across all launch sites.
•The **majority of successful launches (class = 1)** occurred with payloads ranging from **3500–6000 kg**.
•**Booster Version FT (blue)** dominates the mid- and high-range payloads, with a mix of outcomes.
•**Booster Version B5 (green)** appears only once but shows a **successful outcome**, suggesting high reliability.
•**Booster B4 (red)** has both successes and a failure, mostly in lower payload range.
•No strong trend is visible between payload mass and success rate — most outcomes depend more on booster type.



Section 5

Predictive Analysis (Classification)

Andrejs Verhovods

Classification Accuracy

•This bar chart compares the **classification accuracy** of various models trained on launch prediction.
•**Support Vector Machine (SVM)** achieves the highest accuracy (e.g., **87%**), making it the most effective model in this case.
•Other models like Logistic Regression and Decision Tree also perform well, but with slightly lower accuracy.
•These results help determine which model to deploy or further tune.

Model	Accuracy
Logistic Regression	0.83
SVM	0.87
Decision Tree	0.81
KNN	0.79

Confusion Matrix

- This confusion matrix shows how well the selected model performs in predicting launch success.
- **True Positives (TP)** are correct success predictions.
- **True Negatives (TN)** are correct failure predictions.
- **False Positives (FP)** are predicted successes that failed in reality.
- **False Negatives (FN)** are predicted failures that were actually successful launches.
- The model performs well, with most predictions falling into TP and TN.

	Predicted 0	Predicted 1
Actual 0	**TN = 6**	**FP = 2**
Actual 1	**FN = 1**	**TP = 11**

Conclusions

- 📊 **Dashboard enabled interactive exploration** of SpaceX launch data by site and payload.
- 🚀 **KSC LC-39A had the highest number of successful launches** overall.
- 📉 **Payload mass has a visible impact** on launch success probability.
- 🧪 **Booster version FT** showed better reliability compared to earlier versions.
- 🤖 **SVM achieved the highest classification accuracy** among tested models.
- ✅ **Model correctly identified launch outcomes** with low false positives/negatives.
- 📈 **Project demonstrated how data science can support real-world aerospace decision-making**.

Andrejs Verhovods

Appendix

- **Left Column: Code & Outputs**
- Python Code Snippets
  - Callback functions (@app.callback)
  - Dropdown & range slider setup
  - Chart rendering (Pie & Scatter)
- Notebook Outputs
  - DataFrame previews
  - Exploratory data analysis (EDA)
- Model Details
  - Data split ratio (train/test)
  - Hyperparameters used
  - Classification report

Right Column: Visuals & Links

- Visualizations
  - Accuracy bar chart
  - Confusion matrix
  - Scatter plot (Payload vs Outcome)
  - Pie charts (per launch site)
- Data Sources
  - `spacex_launch_dash.csv`
  - Cleaned datasets (if applicable)
- GitHub & Extras
  - Notebook repository link
  - Dash app deployment (if applicable)
  - Requirements.txt or `environment.yml`

Andrejs Verhovods

Thank you!

Andrejs Verhovods