IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Moaz Bannora

7/10/2025

Outline

Executive Summary - Methodologies

Data Collection

- Scraped launch records via SpaceX API and Wikipedia
- Normalized payload and metadata fields for structure

Data Wrangling

- Mapped landing outcomes to binary labels (Class: success/failure)
- Addressed missing data and exported cleaned datasets

Exploratory Data Analysis

- Used seaborn to plot relationships across payload, orbit type, launch site
- SQL queries revealed trends in booster performance and mission reliability

Interactive Analytics

- Created dynamic maps with Folium showing launch clustering and outcome proximity
- Built Dash dashboard with dropdowns, sliders, and scatter plots for filterable insights

Predictive Modeling

- Applied GridSearchCV to tune Logistic Regression, SVM, KNN, Decision Tree
- Selected models based on accuracy (83%) and recall (1.0) for landing prediction

Executive Summary - Insights

Mission Trends & Reliability

- Post-2013 success surge with near-perfect outcomes by 2019
- LEO, ISS, and Polar orbits favored heavier payload recoveries

Booster & Site Insights

- F9 B5 boosters led in max payload handling
- KSC LC-39A showed top-tier reliability and flight success
- VAFB supported lighter missions with limited recoveries

SQL-Driven Discoveries

- ~98 confirmed successful missions via SQL outcome aggregation
- Ground pad landing milestone: Dec 22, 2015
- NASA CRS payload total: 48,213 kg

Geospatial Visualization

- Folium maps revealed launch infrastructure proximity enhancing logistics
- Coastal launch sites like CCAFS proved strategically efficient

Model Selection for Deployment

- Logistic Regression: best for interpretability
- SVM & KNN: ideal for flexibility or rapid prototyping
- All models achieved strong F1-scores (~0.89) with high recall

Introduction

Business Context

- SpaceX offers Falcon 9 launches at $62 million—far below competitors’ $165 million by reusing the first stage.

Problem Statement

- We need to predict whether Falcon 9’s first stage will land successfully, because landing outcome directly determines per-launch cost.

Why It Matters

Accurate landing forecasts enable:

- Cost modeling for mission planning
- Competitive bids by alternative launch providers
- Resource allocation for refurbishment and turnaround

Section 1

Methodology

Methodology

📥Data Collection

- Retrieved launch records from SpaceX API and scraped Falcon 9 missions from Wikipedia
- Normalized and filtered raw data, extracted metadata, and saved cleaned datasets for analysis

🧹Data Wrangling

- Assessed missing data
- Mapped mission outcomes to binary landing labels (Class) for supervised learning
- Exported final structured dataset for training

📊 Exploratory Data Analysis (EDA)

- Visualized correlations between launch outcomes, payload mass, orbit type, and site history
- Created time trends and performance breakdowns using seaborn and SQL-based queries

Methodology - Cont.

🌐 Interactive Visual Analytics

- Mapped launch locations, outcomes, and proximities using Folium
- Built a dynamic dashboard with dropdown filters, payload sliders, and performance plots via Plotly Dash

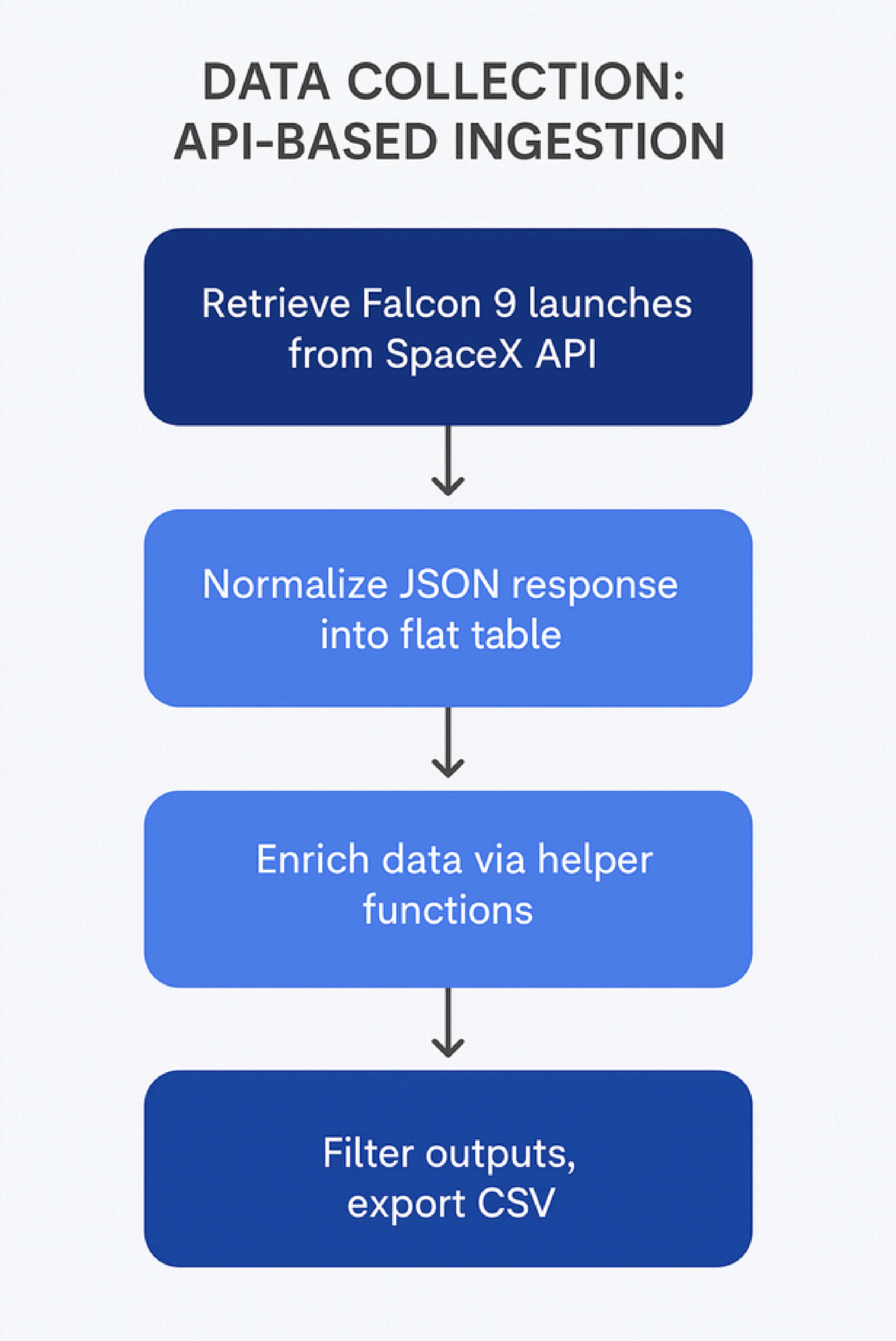
🤖 Predictive Analysis – Classification

- Built, tuned, and evaluated ML models (Logistic Regression, SVM, KNN, Decision Tree) using 10-fold GridSearchCV
- Identified top performers with 83% test accuracy and high recall on landings
- Selected deployment-ready models based on interpretability, scalability, and precision

Data Collection

Workflow Summary

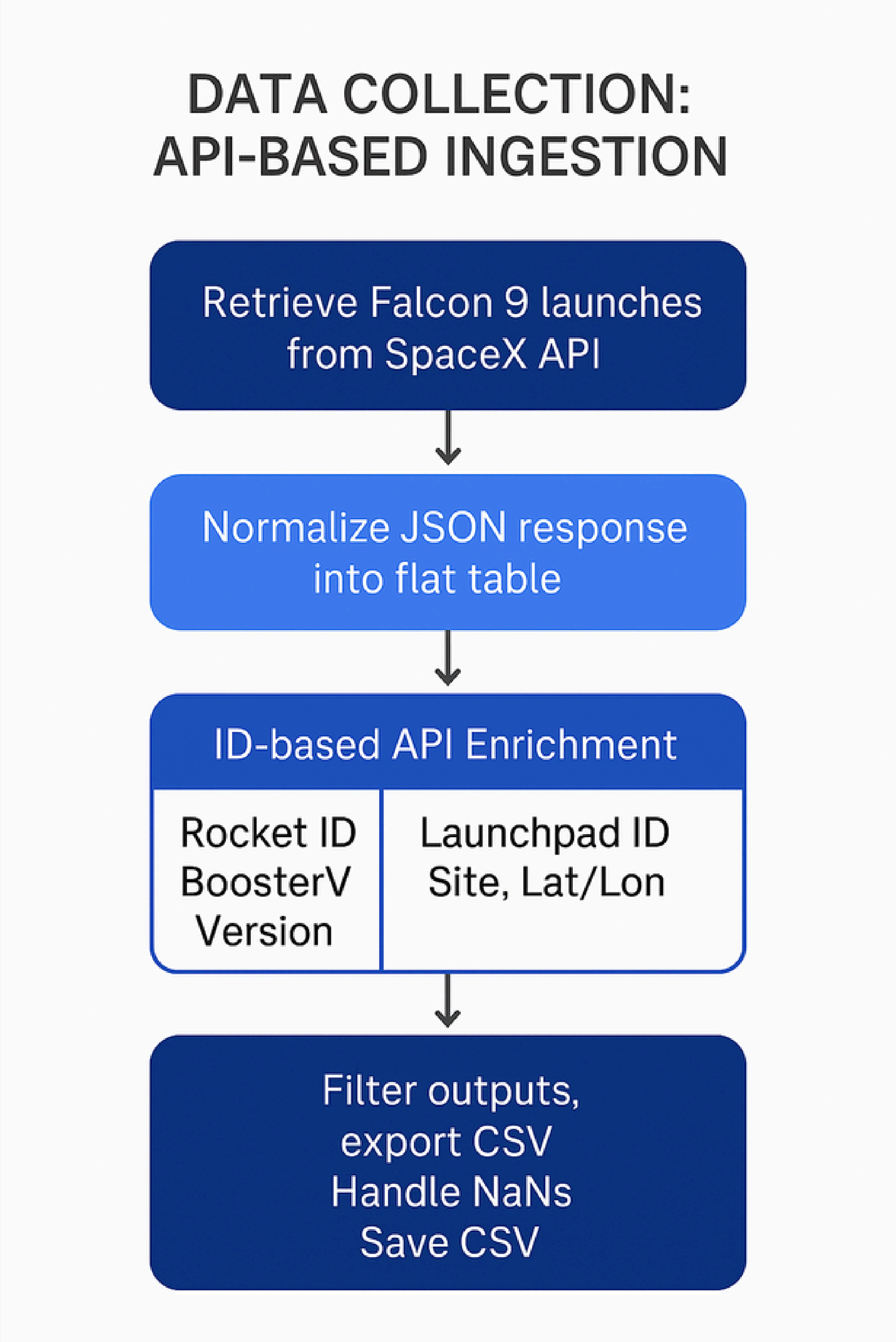
- Fetched historical SpaceX launch records from the API's /v4/launches/past endpoint using a static JSON snapshot for consistency.
- Transformed nested JSON into a flat table using pd.json_normalize(), then filtered out launches with multiple cores or payloads.
- Constructed the enriched launch_df with key features, filtered for Falcon 9 missions only, and reset flight numbers.
- Filled missing PayloadMass values with the mean (~5919.17 kg) and exported the cleaned dataset to dataset_part_1.csv for downstream analysis.



Data Collection – SpaceX API

Workflow Summary

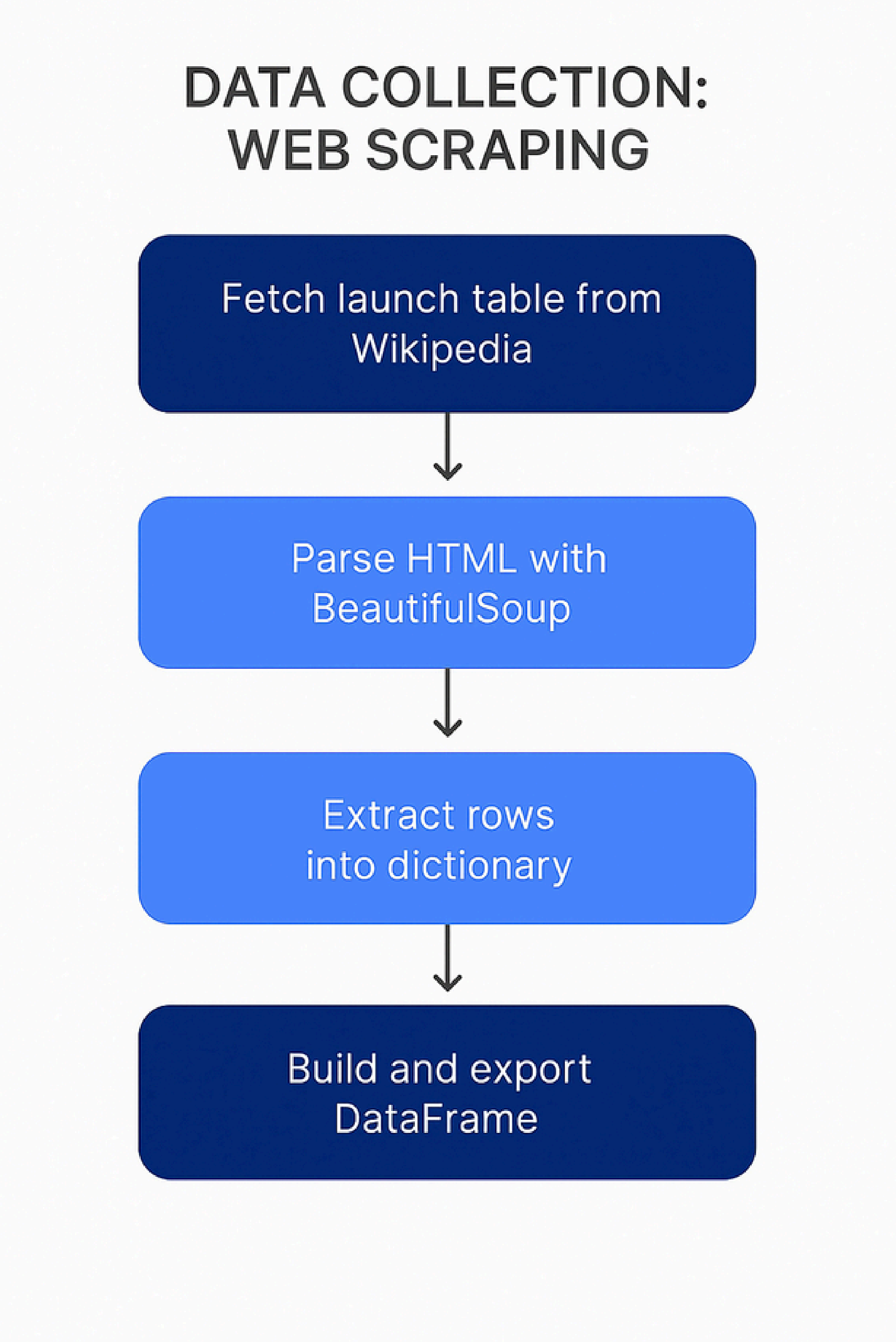
- Fetched historical SpaceX launch records from the API's /v4/launches/past endpoint using a static JSON snapshot for consistency.
- Flattened nested JSON into a structured table. Filtered out multi-core/multi-payload launches and restricted the dataset to missions before November 13, 2020.
- Used ID-based queries (rockets, launchpads, payloads, cores) to extract booster versions, orbit types, site coordinates, landing outcomes, and reuse metrics.
- Handled missing values with imputation, retained only Falcon 9 launches, and exported the final dataset for later analysis.



Data Collection – Scraping

Workflow Summary

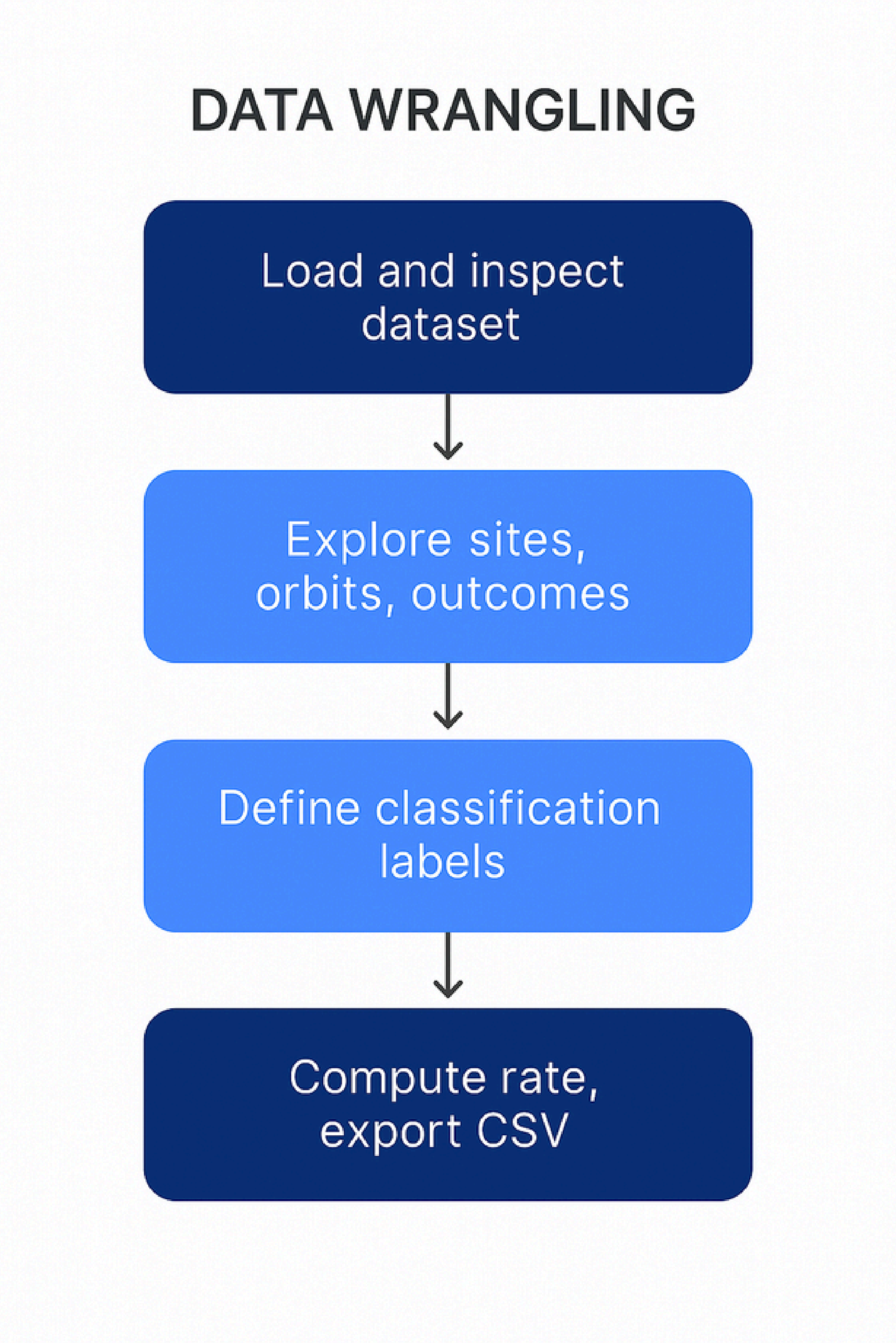
- Requested a static version of the "List of Falcon 9 and Falcon Heavy launches" Wikipedia page using requests.get().
- Identified the third wikitable, extracted headers using helper functions, and cleaned irrelevant column labels to prepare a structured schema.
- Looped through launch rows, collected key fields (date, booster version, site, payload mass, orbit, customer, outcome, landing status), and stored them in a structured launch_dict.
- Converted launch_dict to a Pandas DataFrame, validated contents, and exported the full dataset to spacex_web_scraped.csv for further analysis.



Data Wrangling

Workflow Summary

- Imported Falcon 9 launch records from dataset_part_1.csv, scanned the first few rows, and calculated missing value percentages to assess data quality.
- Used .value_counts() to analyze distribution across launch sites and orbit types. Identified distinct mission outcomes and interpreted their landing success.
- Mapped successful landings (True ASDS, True RTLS, True Ocean) to 1 and unsuccessful/failed landings to 0, forming a new Class column for supervised learning.
- Computed an overall landing success rate of ~66.7% and exported the enriched dataset as dataset_part_2.csv for further analysis and model training.



EDA with Data Visualization

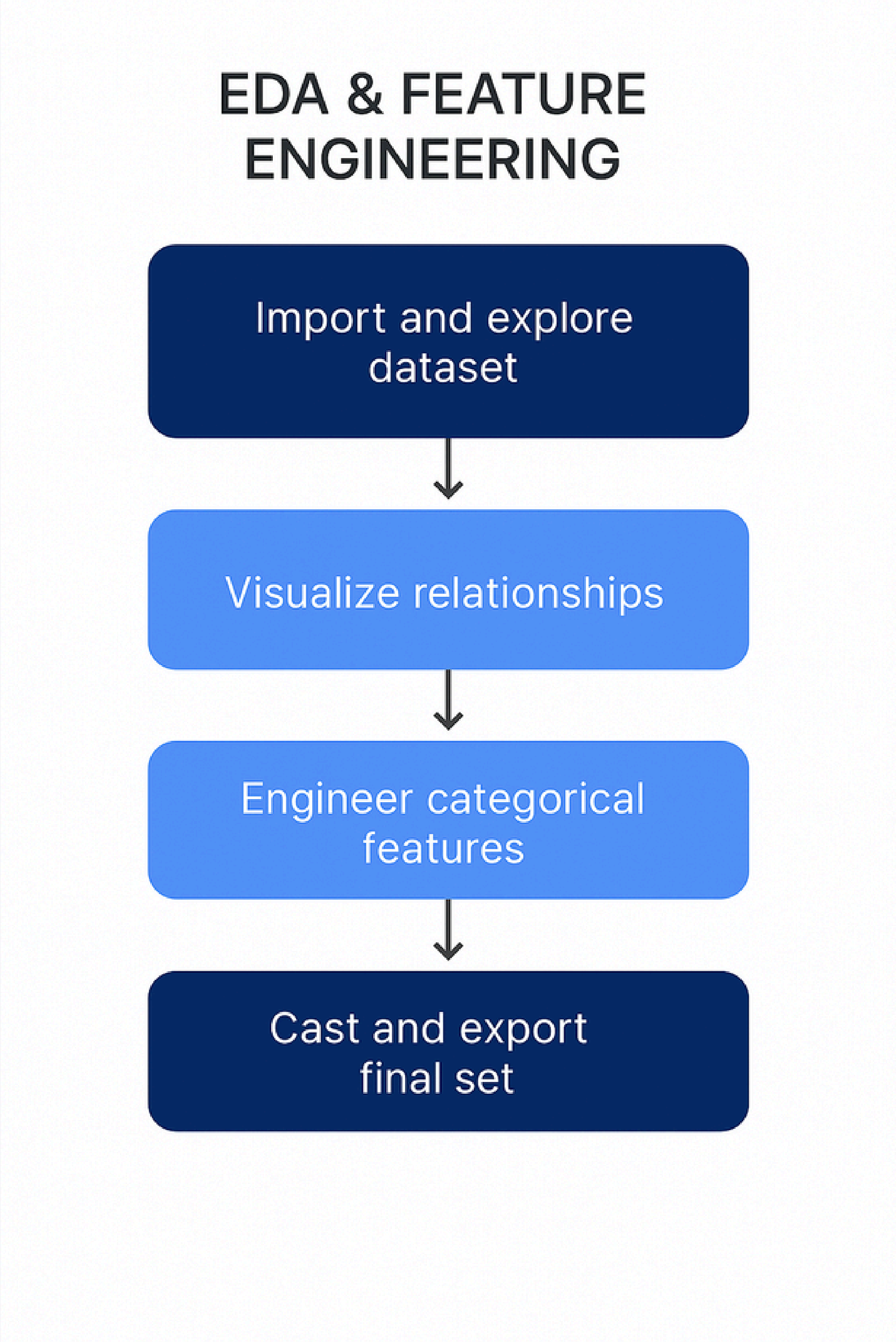
Workflow Summary

- Loaded dataset_part_2.csv, inspected summary stats, and began analyzing how Flight Number and Payload Mass correlate with landing success.

*Used sns.catplot and sns.barplot to explore interactions between:
- Flight Number vs. Launch Site
- Payload Mass vs. Launch Site
- Orbit type vs. Success Rate
- Yearly success trends over time

- Applied pd.get_dummies() to convert columns (Orbit, Launch Site, Landing Pad, Serial) into one-hot encoded features for model compatibility.

- Converted all columns to float64 for modeling consistency and exported the final feature set as dataset_part_3.csv for predictive analysis.



EDA with SQL

Task Summary

- Task 1: Queried unique launch sites using SELECT DISTINCT on Launch_Site
- Task 2: Displayed 5 missions from launch sites starting with 'CCA' using LIKE and LIMIT
- Task 3: Calculated total payload mass carried by NASA (CRS) missions using SUM()
- Task 4: Retrieved average payload mass for booster version F9 v1.1 with AVG()
- Task 5: Identified earliest success on ground pad using MIN(Date) for 'Success (ground pad)'
- Task 6: Listed boosters that landed on drone ships with payload mass between 4000–6000 kg
- Task 7: Grouped mission outcomes and counted total successes/failures via GROUP BY + COUNT()
- Task 8: Retrieved booster versions that carried the maximum payload mass using a MAX() subquery
- Task 9: Extracted failed drone ship landings during 2015 by parsing date string into Month and filtering on Landing_Outcome
- Task 10: Ranked landing outcomes (2010–2017) by frequency using GROUP BY, COUNT(), and ORDER BY DESC

Build an Interactive Map with Folium

Workflow Summary

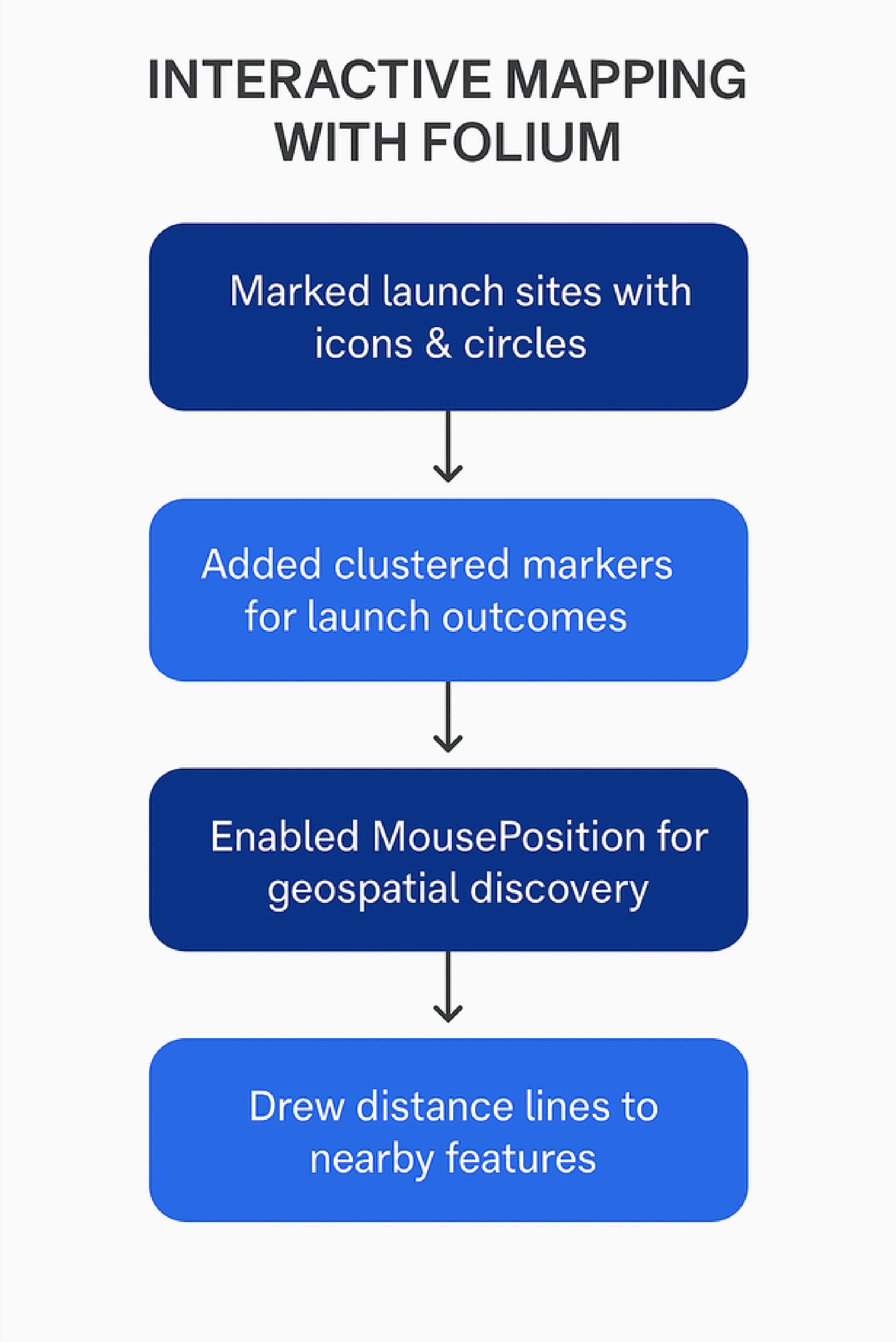
- Plotted each launch site using a green rocket icon marker for easy identification. Wrapped each site in a black folium.Circle to highlight its geographic footprint. Labelled each site using DivIcon for clarity and visual appeal.

*Visualized each launch record using success/failure-based markers:

- Green: Successful (class = 1)
- Red: Failed (class = 0)
- Grouped overlapping points into a MarkerCluster to reduce clutter and improve interactivity.

- Embedded a MousePosition layer to dynamically show coordinates as the user moves the mouse—helpful for identifying and measuring proximity to geographic features like roads or shorelines.

- Connected launch site to nearby targets—Coastline, Railroad, Highway, and Cape Canaveral—using colored PolyLine paths. Calculated haversine distances and displayed them in tooltips. Added labeled markers at endpoints to display distance in kilometers.



Build a Dashboard with Plotly Dash

Workflow Summary

- Added an interactive dropdown menu that lets users select a specific launch site or view all sites. This enables site-level filtering across both visualizations, making comparative analysis and drilldowns seamless.

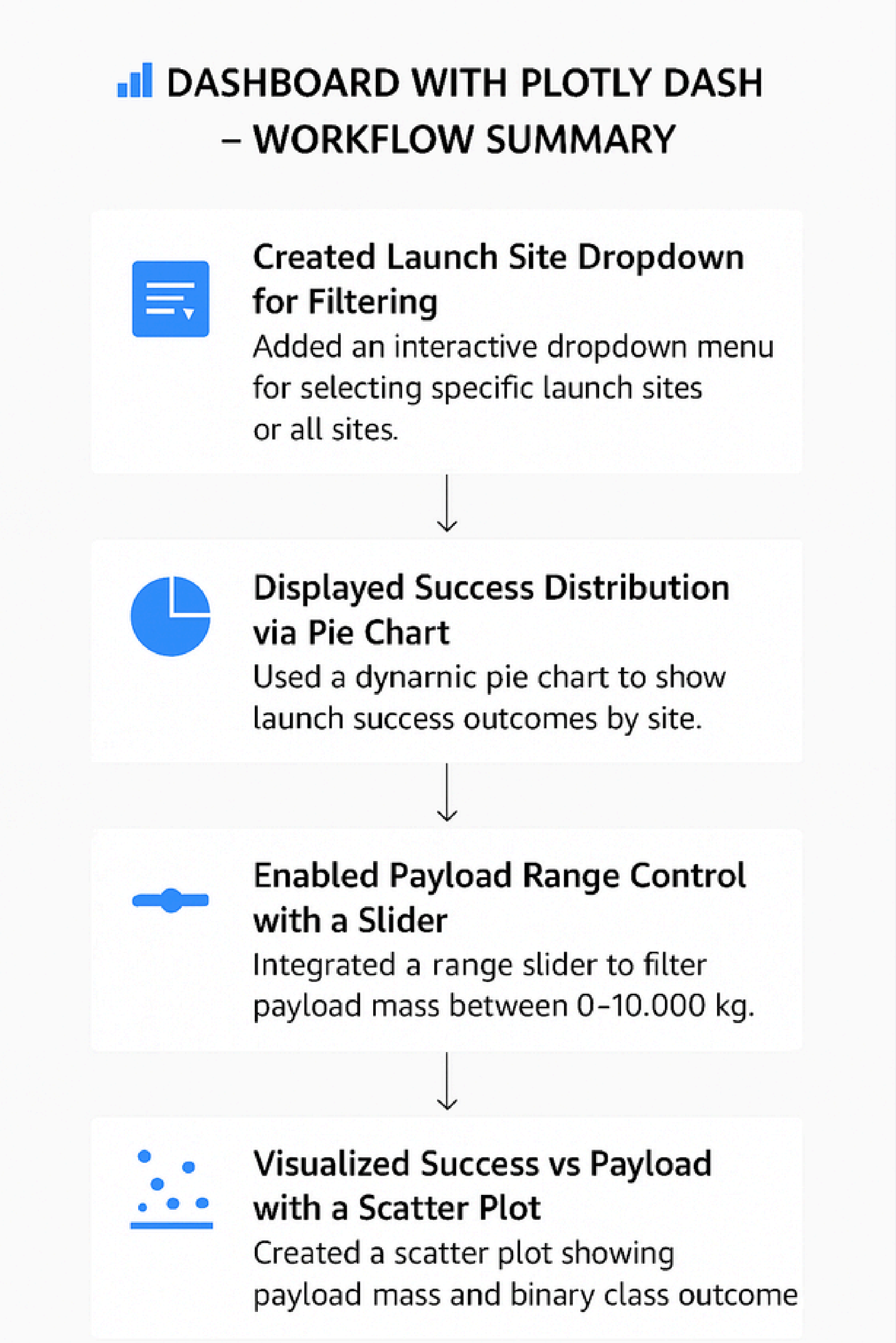
*Used a dynamic pie chart to show launch success outcomes:

- For “All Sites”: displays total success launches per site
- For specific sites: breaks down successes vs failures (this provides an intuitive snapshot of launch reliability across locations).

- Integrated a range slider to filter payload mass between 0–10,000 kg. This allows users to analyze how mission success correlates with payload capacity and select custom mass ranges interactively.

*Created a scatter plot showing:

- x-axis: Payload Mass (kg)
- y-axis: Binary class outcome (success/failure)
- Color-coded: Booster Version Category
- This reveals technical performance patterns and helps assess whether heavier payloads impact success differently across booster generations.



Predictive Analysis (Classification)

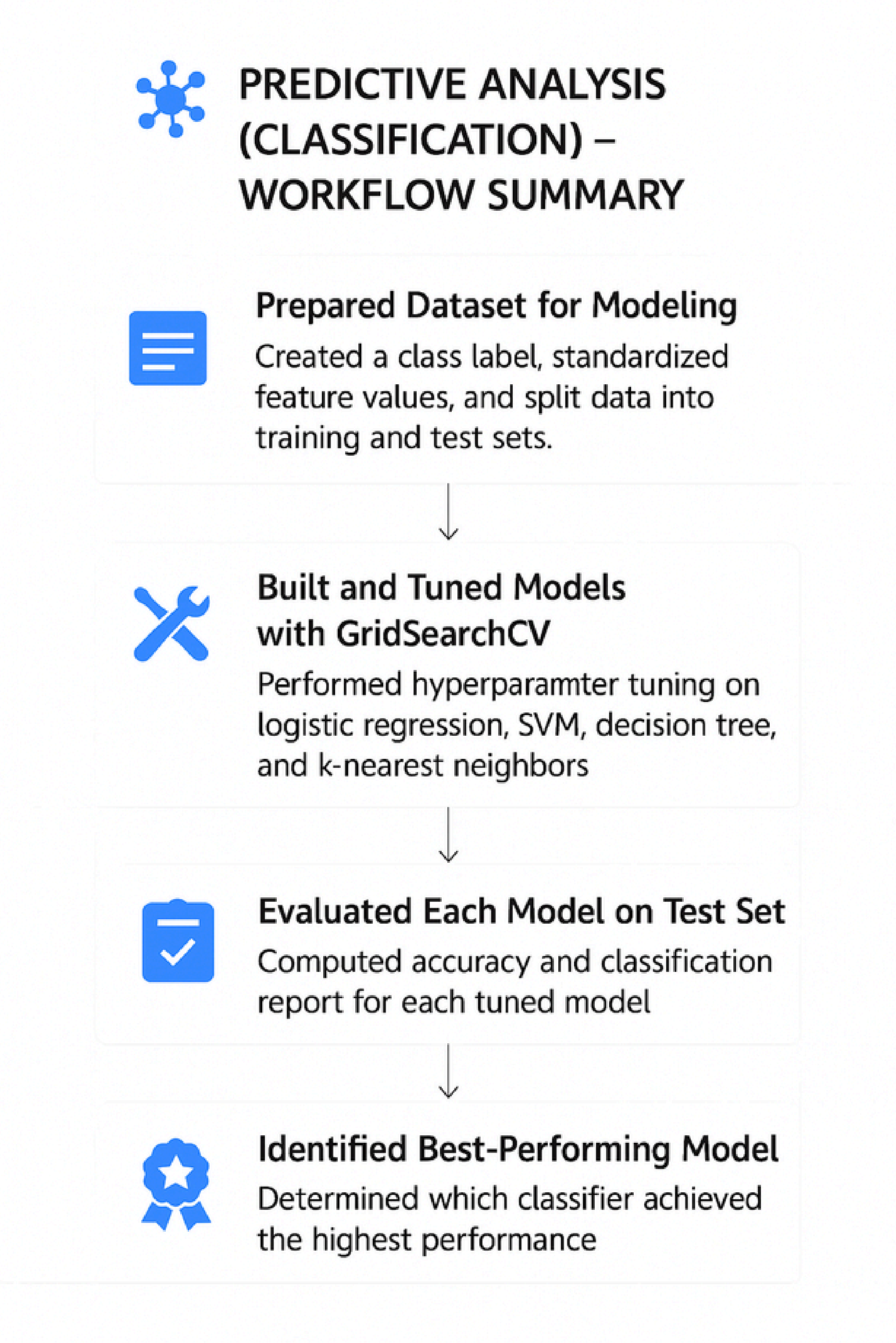
Workflow Summary

- Extracted Class label as target variable and standardized all features using StandardScaler

- Split data into training and test sets (80/20) for fair evaluation and model comparison.

*Implemented 10-fold cross-validation with hyperparameter grids for each classifier:

- Logistic Regression: tuned C and penalty
- Support Vector Machine: tuned C, kernel, and gamma
- Decision Tree: tuned criteria, depth, split rules, and leaf size
- K-Nearest Neighbors: optimized n_neighbors, algorithm, and distance metric p



Predictive Analysis (Classification) - Cont.

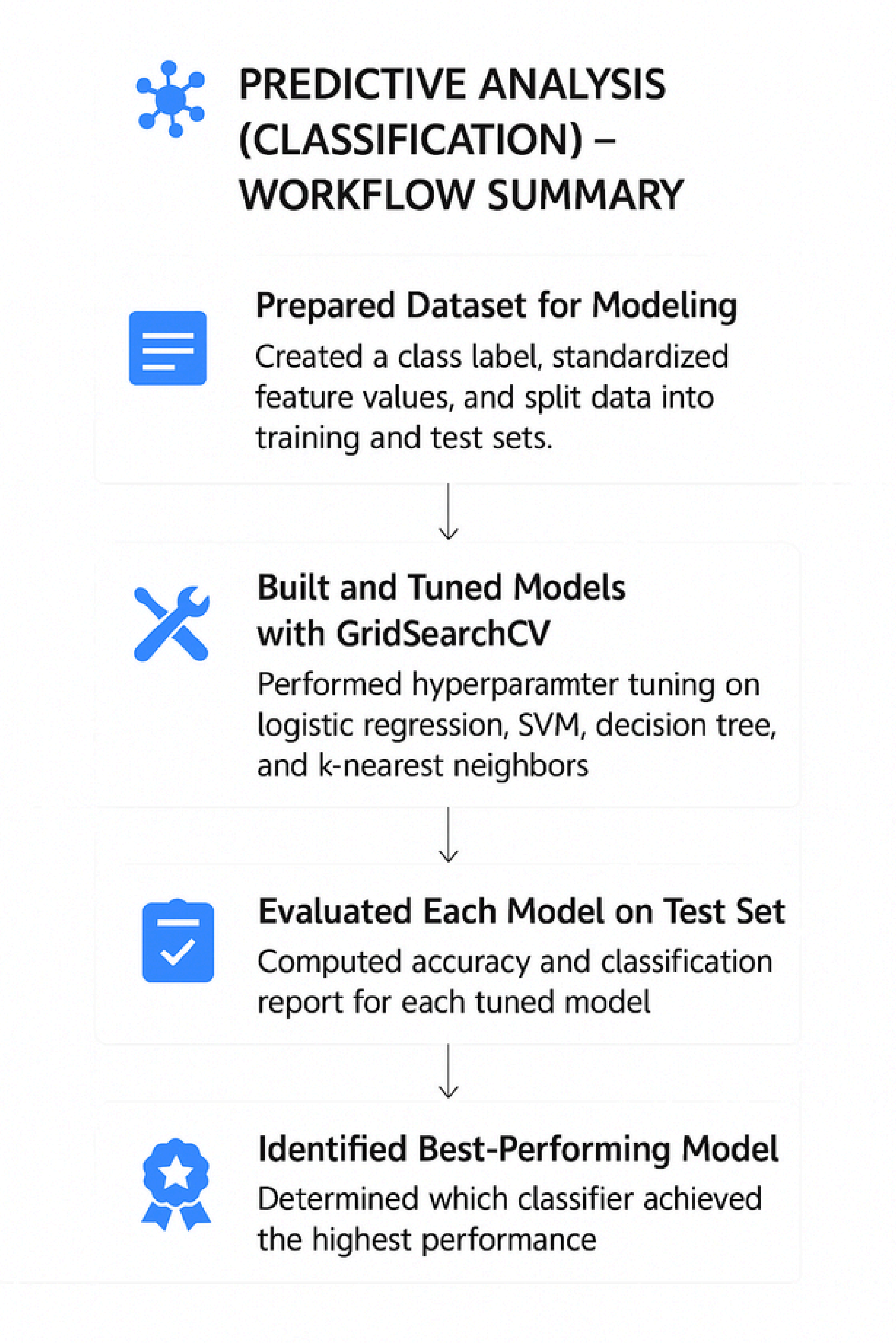
Workflow Summary

*Evaluated Each Model on Test Set:

- Accuracy scores
- Confusion matrix (visualized via sns.heatmap)
- Precision, Recall, and F1-score (via classification_report)
- All three top-performing models—Logistic Regression, SVM, and KNN—achieved 83% test accuracy with high recall and F1-score for successful landings.

*Summarized model strengths for future deployment:

- Logistic Regression – clear and interpretable coefficients
- SVM – effective with nonlinear decision boundaries
- KNN – fast and intuitive for prototyping
- Final model selection may vary based on scale, explainability, and production constraints.



EDA Results

Trend Over Time

- Launch Success steadily climbed from 2013 to 2020, peaking with near-perfect reliability in 2018–2019.

Impact of Orbit Type & Payload

- Heavy payloads favor LEO, Polar, ISS missions for successful recovery.
- GTO shows mixed results — recovery isn't tied to payload size or flight experience.

Flight Experience

- Higher Flight Numbers boost success in LEO missions.
- GTO missions remain unpredictable despite growing experience.

Launch Site Performance

- KSC LC 39A leads in success and payload range.
- VAFB SLC 4E handles lighter missions — fewer recoveries overall.

EDA with SQL Results

Mission Outcome Summary

- Between 2010–2017, most launches had no landing attempt, while drone ship recoveries dominated the effort.
- Overall, ~98 missions succeeded, confirming SpaceX's strong reliability across trials.

Booster & Payload Analysis

- F9 B5 boosters carried SpaceX's heaviest payloads, showing consistent high-performance capacity.
- Booster versions B1022–B1031.2 succeeded with mid-range payloads (4000–6000 kg) on drone ships.
- F9 v1.1 averaged ~2928.4 kg per launch — indicative of early-stage booster capability.

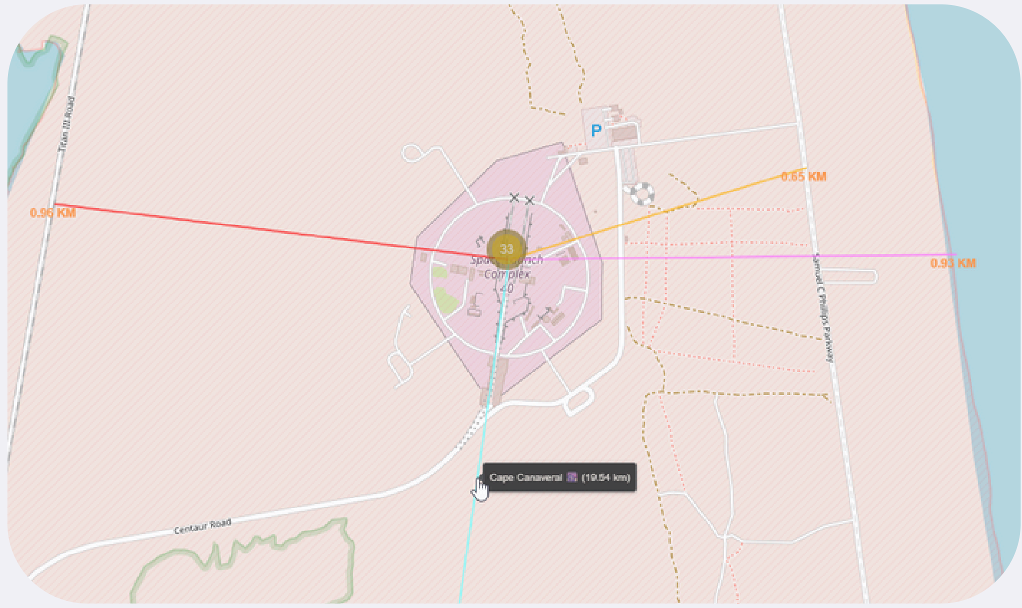
Launch Sites & Timeline

- First ground pad success: Dec 22, 2015 — a landmark moment for reusability.
- NASA CRS missions totaled 48,213 kg in cargo — reflecting strong agency collaboration.
- Launches starting with "CCA" filtered for targeted analysis; 4 unique launch sites identified via SQL.

Interactive Analytics Demo Results

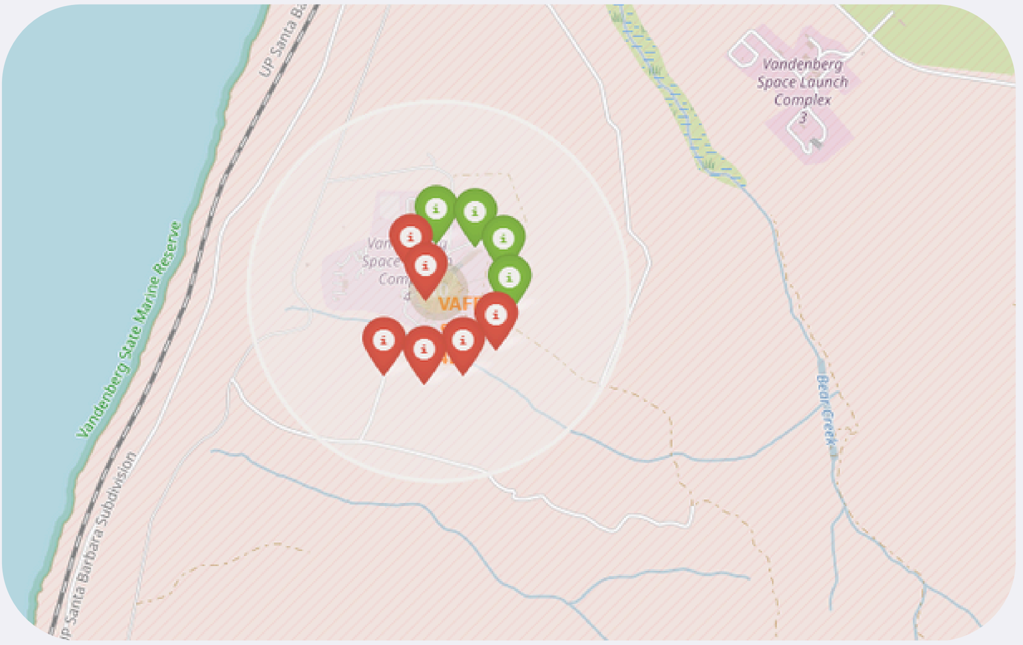
Launch Site Distribution

- Folium maps reveal strategic clustering of SpaceX launch facilities across California (VAFB) and Florida (KSC, CCAFS, SLC-40A), enabling geographic and logistical insights through spatial overlays.



Mission Outcomes Visualization

- At VAFB SLC-4E, mission results are displayed with green (success) and red (failure) markers — making recovery trends and site performance instantly visible.



Infrastructure Proximity

- CCAFS SLC-40's map showcases its close alignment with logistics infrastructure:
- Highway (~0.65 km)
- Coastline (~0.93 km)
- Railroad (~0.96 km)
- City (~14.5 km)

These spatial insights illustrate why coastal and well-connected launch sites support operational efficiency.

Predictive Analysis Results

Result & Deployment Insights

All top models showed strong predictive performance for identifying successful landings

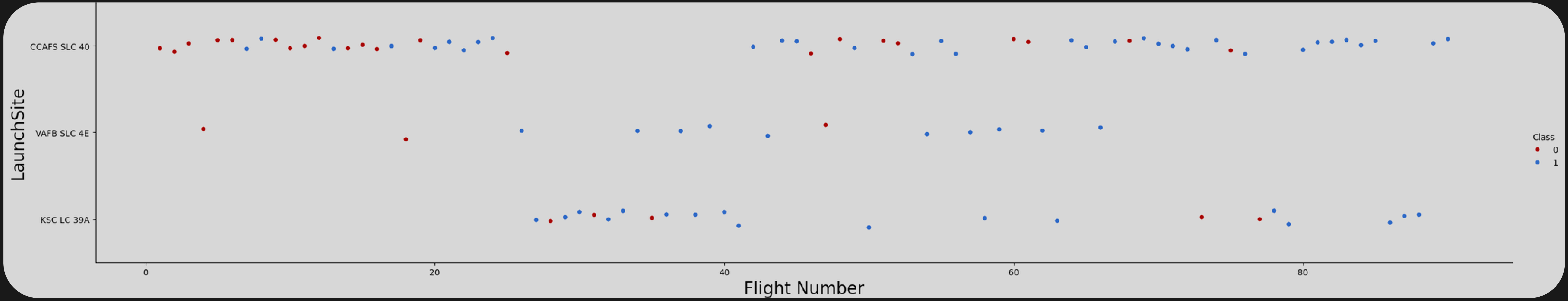
- **Logistic Regression → High transparency, ideal for business cases**
- **SVM → Flexible with complex decision boundaries**
- **KNN → Lightweight, great for quick prototyping**
- **Model choice may depend on application scale, interpretability needs, and production speed**

Model	Accuracy	Recall	F1-Score
Logistic Regression	0.83	1	0.89
Support Vector Machine	0.83	1	0.89
K-Nearest Neighbors	0.83	1	0.89
Decision Tree	~ 0.78	1	0.89

Section 2

Insights drawn from EDA

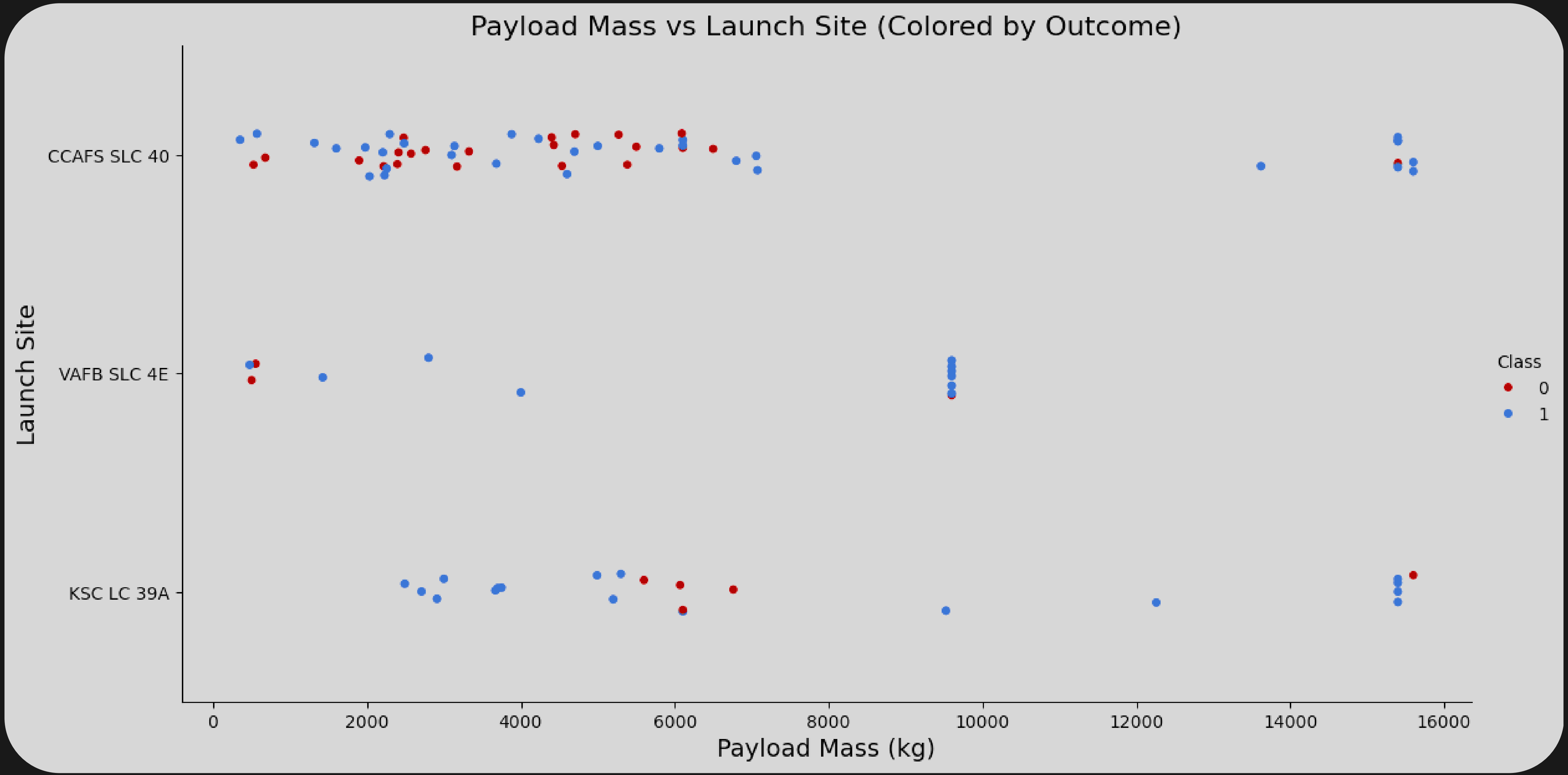
Flight Number vs. Launch Site



*(Success = Blue , Failure = Red)

This scatter plot shows how Falcon 9 landing success varies across launch sites and mission sequence. As flight numbers increase, more missions land successfully, especially from KSC LC 39A. The trend highlights SpaceX's improving recovery rate over time. Launch site also plays a role in the likelihood of success.

Payload vs. Launch Site

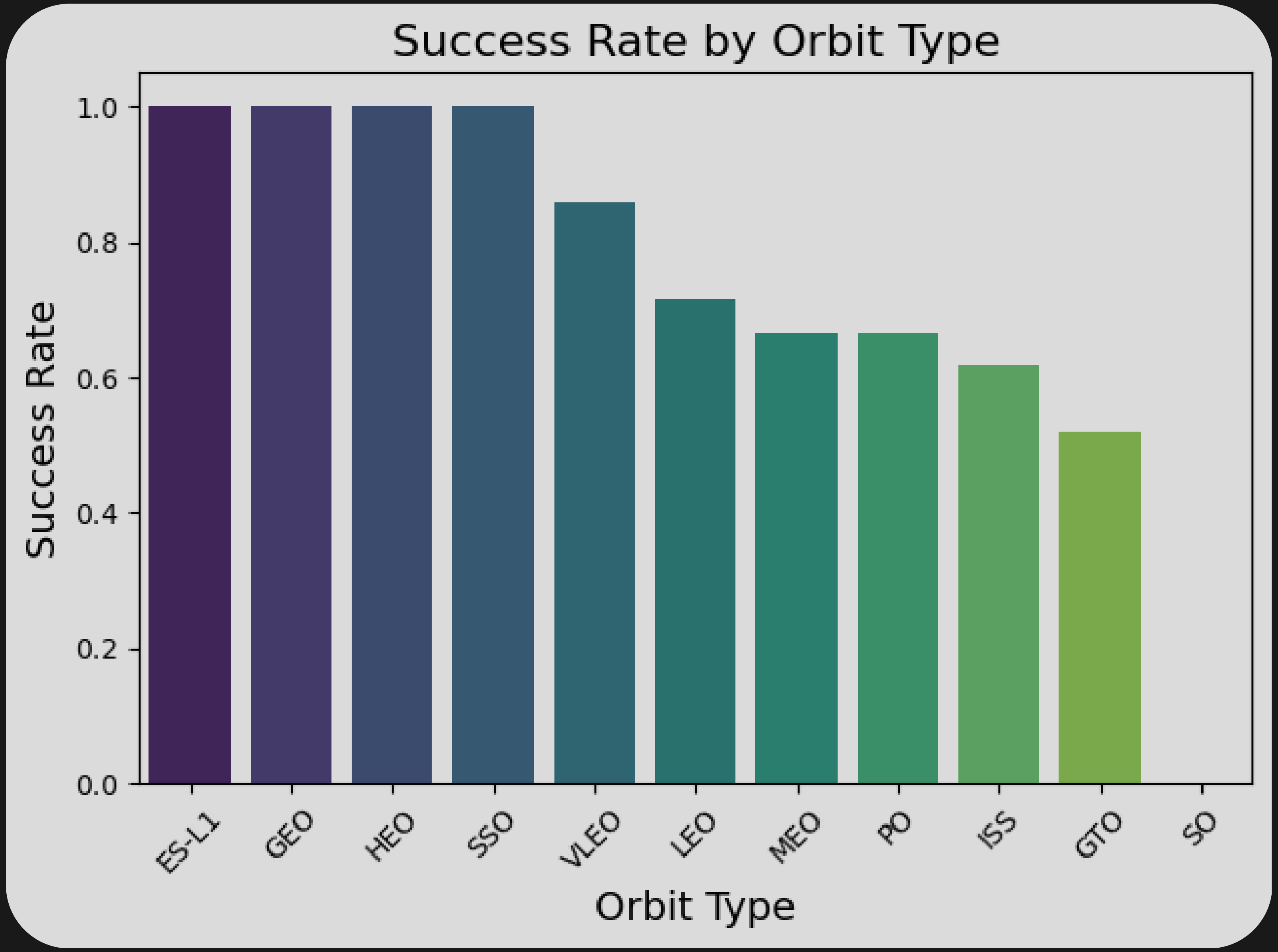


*VAFB SLC 4E lacks launches with heavy payloads (>10,000 kg), unlike other major sites.

*(Success = Blue , Failure = Red)

The scatter plot highlights how launch outcomes vary across sites based on payload mass. KSC LC 39A and CCAFS SLC 40 supported a wide range of payloads, including those above 10,000 kg. In contrast, VAFB SLC 4E shows only lighter missions, suggesting operational or logistical constraints. Successful launches appear across all sites but skew toward higher payloads at KSC.

Success Rate vs. Orbit Type

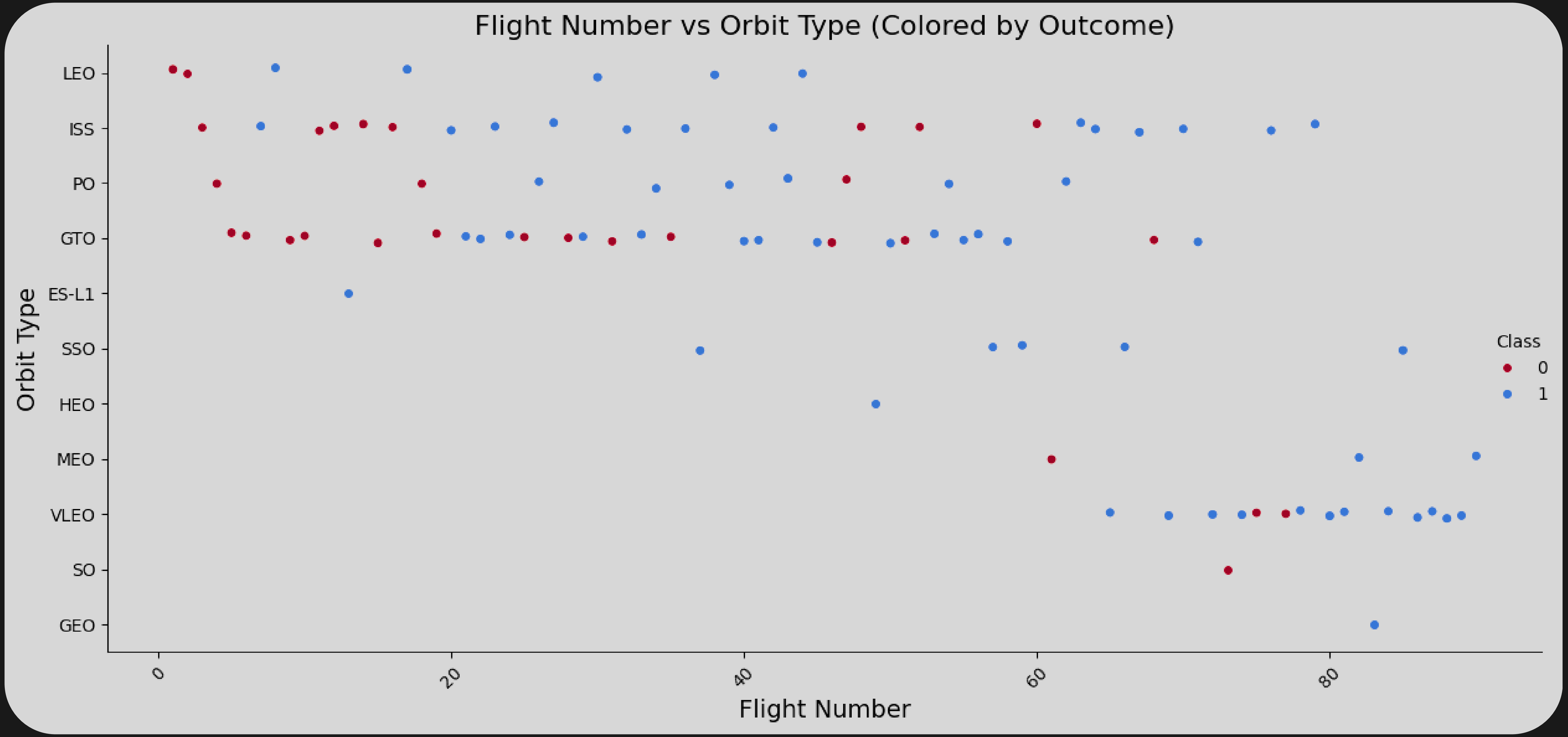


* Success Rate by Orbit Type — Highest at ES-L1, GEO, HEO, and SSO; Lowest at GTO and SO

The bar chart reveals that orbits such as ES-L1, GEO, HEO, and SSO have the highest success rates—indicating strong reliability for those mission types. Lower success rates for GTO and SO suggest these orbits may pose greater challenges for booster recovery. This variation highlights how orbital destination influences landing outcome.

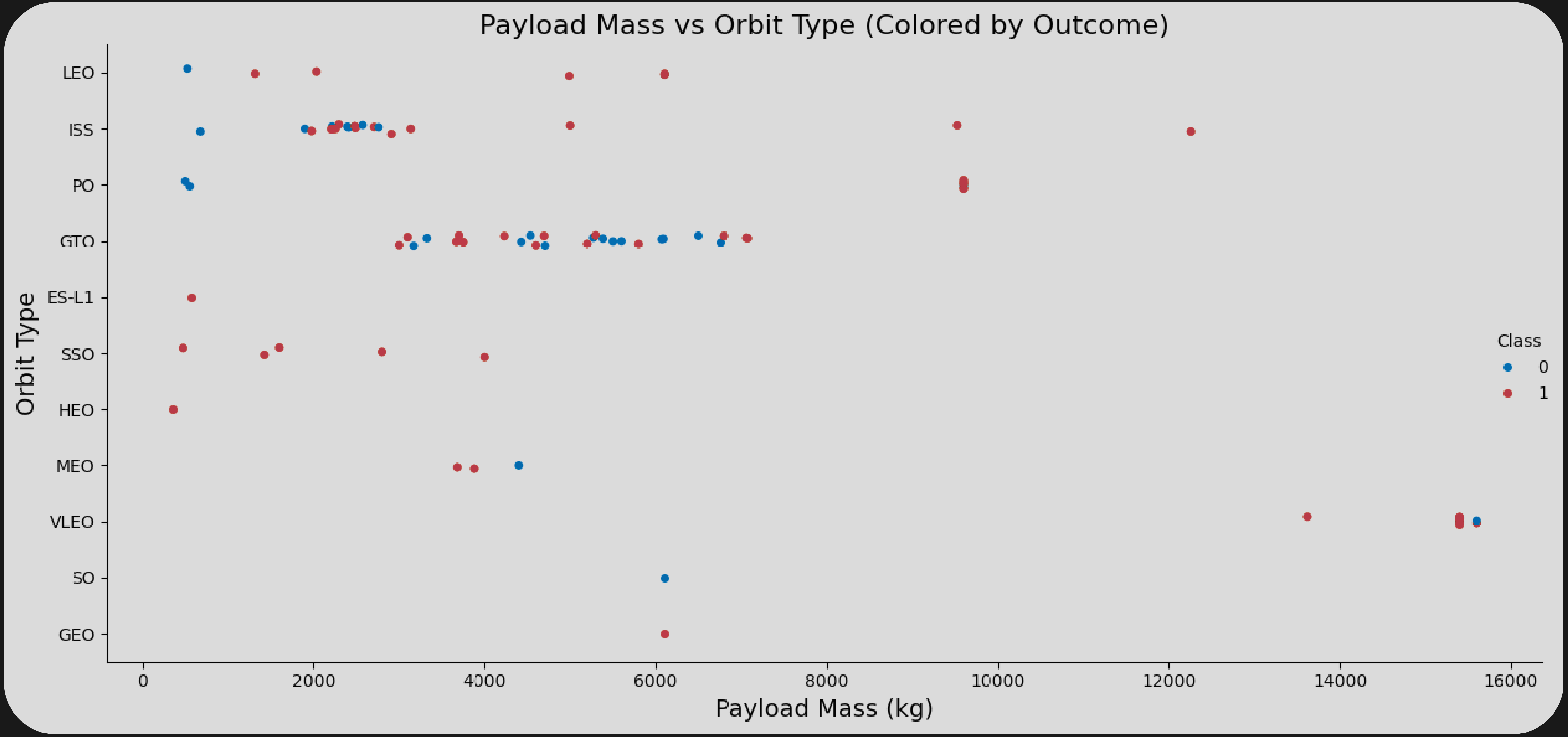
Flight Number vs. Orbit Type

*Flight Number vs Orbit Type — LEO shows increasing success with experience; GTO remains unpredictable



This plot explores how launch success varies with flight experience across different orbit types. In LEO, success improves noticeably as flight number increases, hinting at technical maturity and reliability. On the other hand, GTO missions show mixed outcomes regardless of flight count, suggesting higher complexity or variability. Orbit type clearly affects recoverability trends over time.

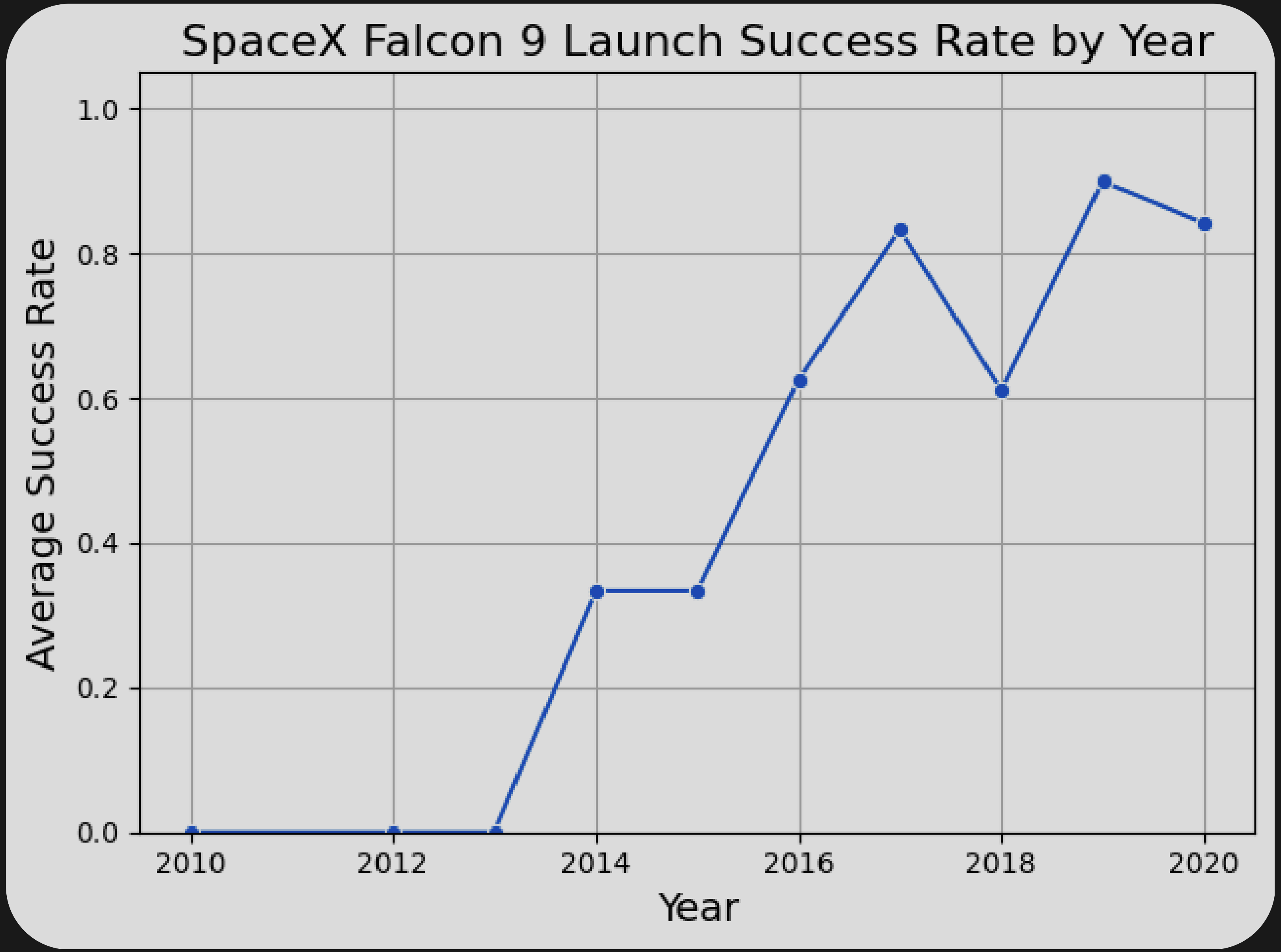
Payload vs. Orbit Type



*Payload Mass vs Orbit Type — Success clusters in LEO, Polar & ISS; GTO outcomes remain mixed regardless of payload.

This scatter plot displays how launch success varies with payload mass across orbit types. For heavier payloads, successful landings are more frequent in Polar, LEO, and ISS missions. In contrast, GTO missions show no clear success pattern—both outcomes appear regardless of payload size. Orbit selection clearly influences recoverability trends.

Launch Success Yearly Trend



*Launch Success Rate by Year — Strong improvement post-2013, peaking around 2019–2020.

The line chart highlights a consistent rise in Falcon 9's launch success rate beginning in 2013. By 2018 and 2019, SpaceX achieved near-perfect reliability, reflecting advancements in landing technology and operational precision. This upward trend confirms their growing dominance in reusable launch systems.

All Launch Site Names

Display the names of the unique launch sites in the space mission

[10]:
```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

[10]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

*Query Output — Distinct Launch Sites from SPACEXTABLE

The query retrieves the distinct launch site names used in SpaceX missions, helping identify the geographic distribution of launches. The result shows four unique sites, confirming where Falcon 9 operations have historically occurred.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

[11]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

[11]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

*Filtered Launch Records from CCAFS Sites ,
First 5 missions starting with ‘CCA’.

This SQL query filters launch records from sites starting with “CCA,” returning the first five missions from Cape Canaveral. The output includes key details such as payloads, orbits, and outcomes—useful for narrowing analysis to specific locations.

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[12]: %sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass_KG FROM SPACEXTABLE WHERE "Customer" LIKE '%NASA (CRS)%';

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

[12]:

Total_Payload_Mass_KG
48213

*Total Payload Mass by NASA (CRS) Missions —
48,213 kg launched across all relevant flights.

The query calculates the combined payload mass of all missions operated by NASA (CRS), yielding a total of 48,213 kg. This highlights NASA's substantial cargo contribution through its collaboration with SpaceX.

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
[13]: %sql SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass_KG FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';

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

[13]:

Average_Payload_Mass_KG
2928.4

*Average Payload for F9 v1.1 Booster
2928.4 kg across recorded launches.

The query calculates the average payload mass carried by SpaceX’s F9 v1.1 booster, resulting in approximately 2928.4 kg. This gives insight into the typical capacity handled by this older booster version across missions.

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

[14]:
```
%sql SELECT MIN("Date") AS First_Successful_GroundPad_Landing FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

[14]:

First_Successful_GroundPad_Landing
2015-12-22

*First Successful Ground Pad Landing December 22, 2015, via MIN(Date) SQL query.

The query uses the MIN function to extract the earliest recorded date of a successful ground pad landing. Based on the result, SpaceX achieved its first successful ground pad landing on 2015-12-22, marking a major milestone in reusability.

Successful Drone Ship Landing with Payload between 4000 & 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[15]: %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;

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

[15]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

*Boosters with Drone Ship Landings and Mid-Range Payloads

— Includes F9 FT B1022, B1026, B1021.2, and B1031.2

This query identifies boosters that successfully landed on drone ships while carrying payloads between 4000 kg and 6000 kg. The result includes four booster versions, confirming their capability under these mission conditions.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
[16]: %sql SELECT "Mission_Outcome", COUNT(*) AS Total_Count FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

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

[16]:

Mission_Outcome	Total_Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

*Mission Outcome Summary — ~ 98 successes, with minimal failure and uncertain payload statuses.

This query aggregates all mission outcomes and counts how often each type occurred. Most missions were marked as successful (around ~ 98 total), with only a handful categorized as partial or failed—offering insight into SpaceX's strong launch reliability.

Boosters Carried Maximum Payload

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

[17]:
```
%%sql
SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE
WHERE "Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_")
FROM SPACEXTABLE);
```

 * sqlite:///my_data1.db
Done.

[17]:

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

*Booster Versions for Max Payload — F9 B5 units led the way in lifting SpaceX's heaviest cargo(15600kg for all boosters).

This query finds all booster versions that carried the heaviest payload by using a subquery to extract the maximum payload value. The result highlights multiple F9 B5 boosters, showing their consistent performance with SpaceX's largest missions.

2015 Failed Launch Records

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

[18]:
```
%%sql SELECT substr("Date", 6, 2) AS Month,"Landing_Outcome","Booster_Version","Launch_Site"
FROM SPACEXTABLE WHERE substr("Date", 1, 4) = '2015' AND "Landing_Outcome" = 'Failure (drone ship)';
```

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

[18]:

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

*Failed Drone Ship Landings in 2015 — Includes month, booster version, and site for unsuccessful recoveries.

This SQL query isolates all failed drone ship landings that occurred in 2015, using substring functions to extract month and year from launch dates. The output reveals two such failures, including booster versions and launch site details.

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[19]:
```
%%sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count
FROM SPACEXTABLE WHERE DATE("Date") BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome" ORDER BY Outcome_Count DESC;
```

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

[19]:

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

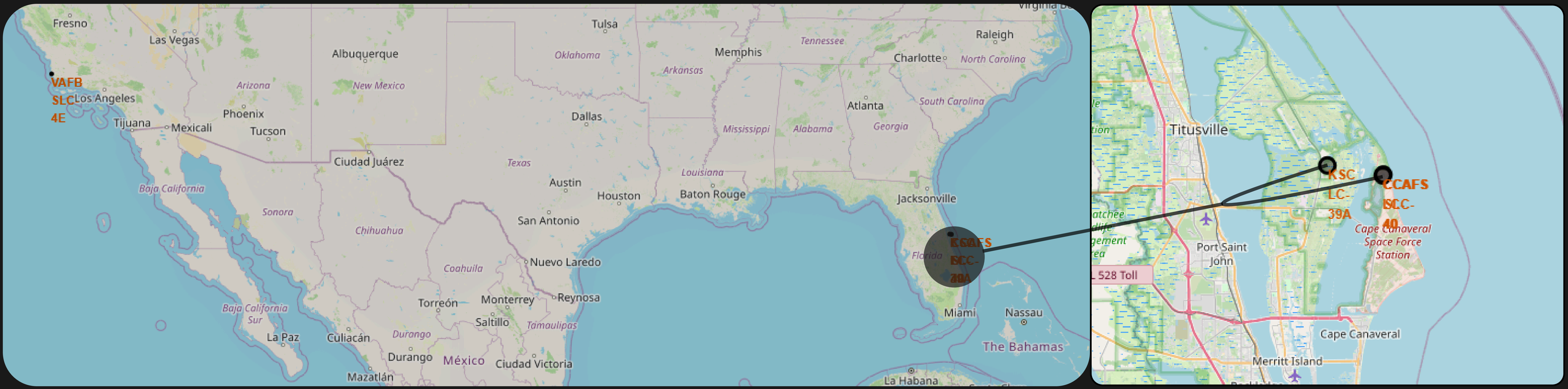
*Landing Outcome Frequency (2010–2017) — Majority had no attempt; drone ship landings dominate recovery trials.

The query ranks SpaceX landing outcomes by frequency between mid-2010 and early 2017. Most missions in that period had no landing attempt, while drone ship outcomes—both successful and failed—were among the most common recovery attempts.

Section 3

Launch Sites Proximities Analysis

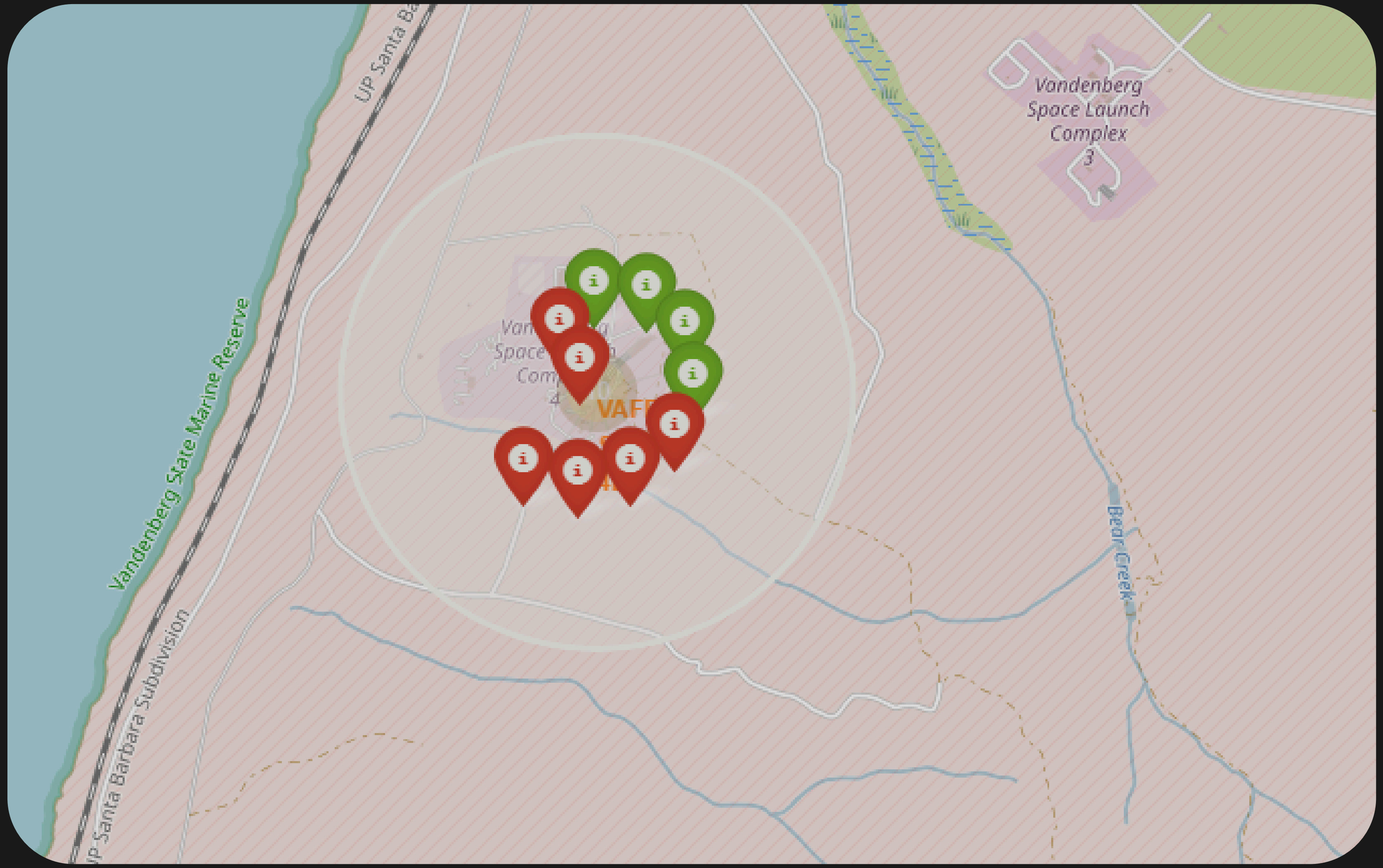
All launch sites' location markers on a global map



*Geographic distribution of SpaceX launch sites, emphasizing coastal access and regional infrastructure advantages.

This map highlights major SpaceX launch facilities across the southern U.S. and northern Mexico, including VAFB in California and KSC, CCAFS, and SLC-40A in Florida, offering a spatial overview of strategic launch site distribution. The labels ensure geographic clarity as you explore launch proximities and regional clustering.

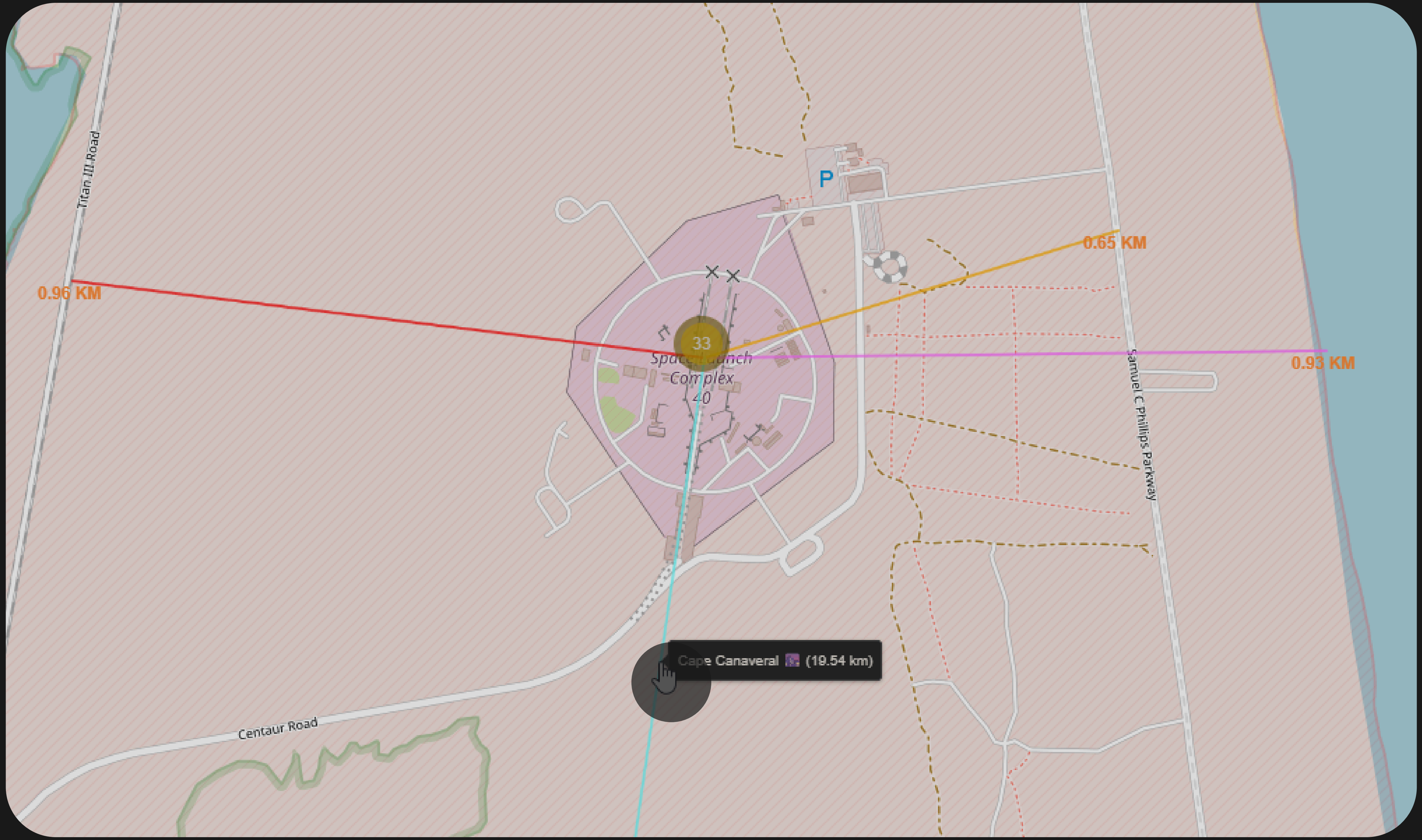
Color-labeled launch outcomes



*Distribution of launch successes and failures from SLC-4E, highlighting operational performance over time.

The map illustrates launch outcomes at Vandenberg Space Launch Complex 4E (SLC-4E), with green markers indicating successful missions and red markers showing failures.

CCAFS SLC- 40 site to its proximities



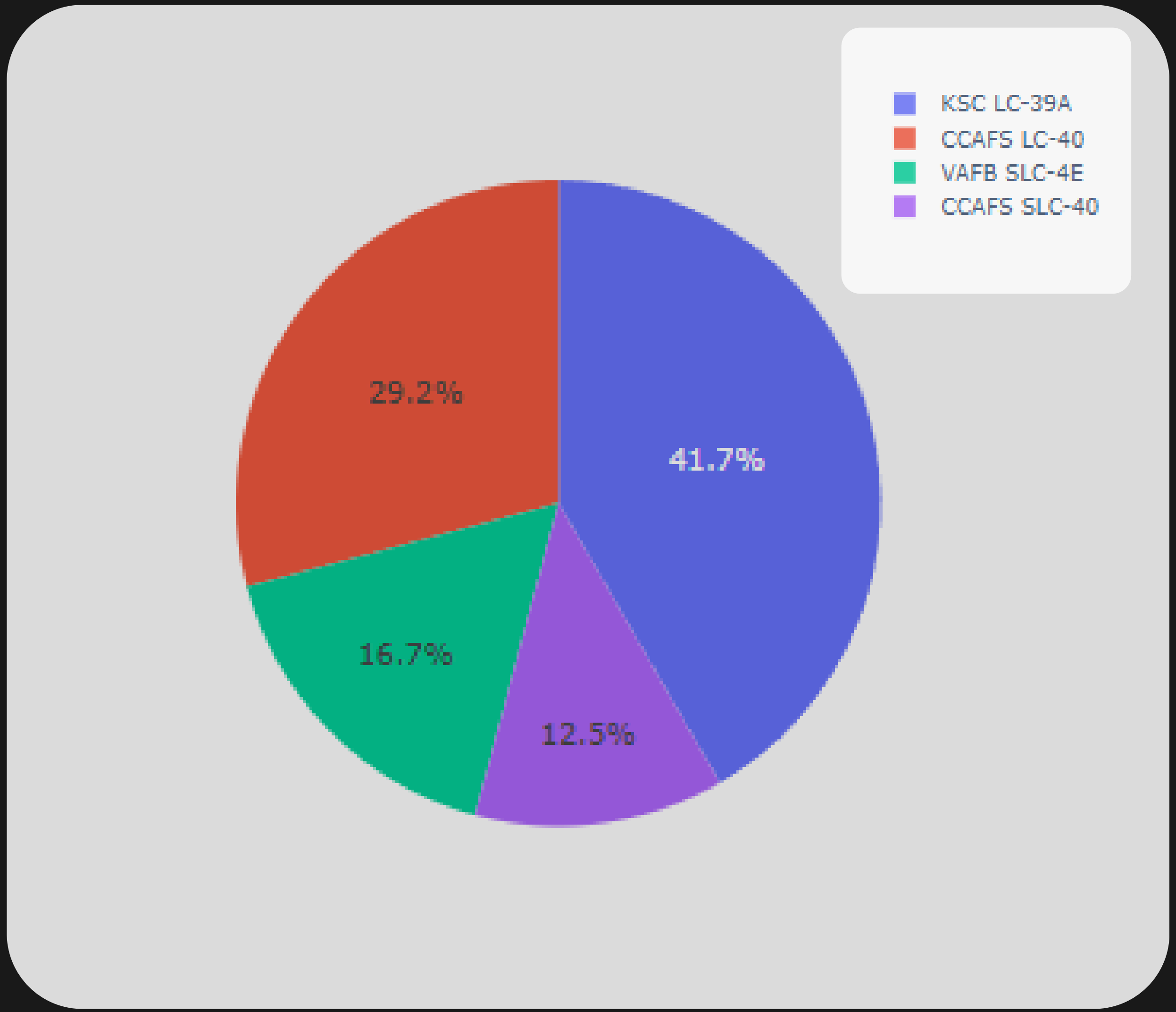
*Strategic placement of SLC-40 relative to essential transport and geographical features, emphasizing its operational accessibility.

The map depicts CCAFS SLC-40's location in relation to key infrastructures, highlighting its close proximity to the highway (~0.65 KM), coastline (~0.93 KM), railroad (~0.96 KM), and a nearby city (~14.5 KM), supporting both logistical accessibility and coastal launch advantages.

Section 4

Build a Dashboard with Plotly Dash

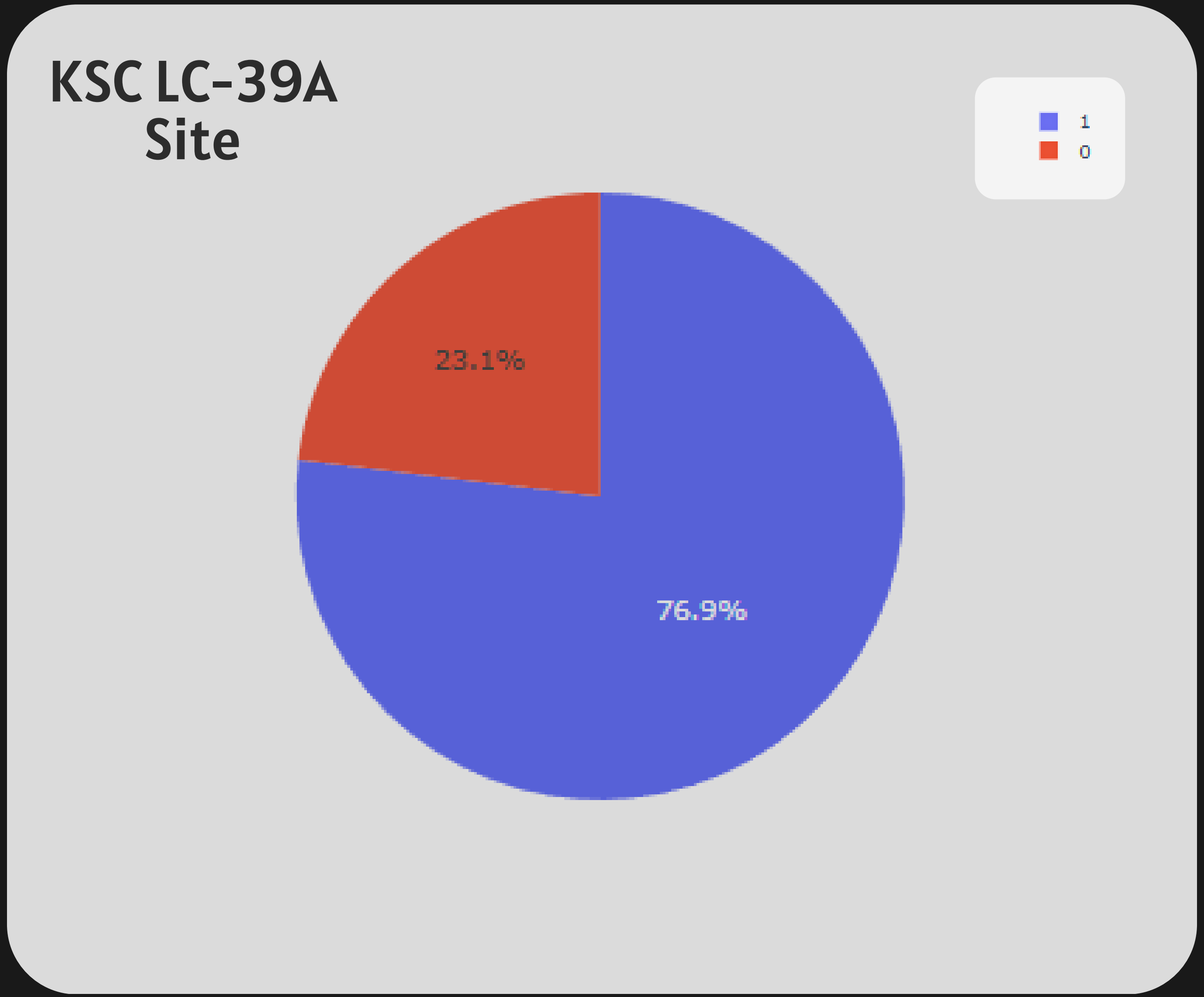
Launch Success Count for All Sites



*Launch Success Count — KSC LC-39A holds the highest share of successful missions.

This dashboard displays the count of successful launches across major SpaceX sites, with KSC LC-39A leading in total recoveries. It provides a quick comparative snapshot for evaluating site-level performance.

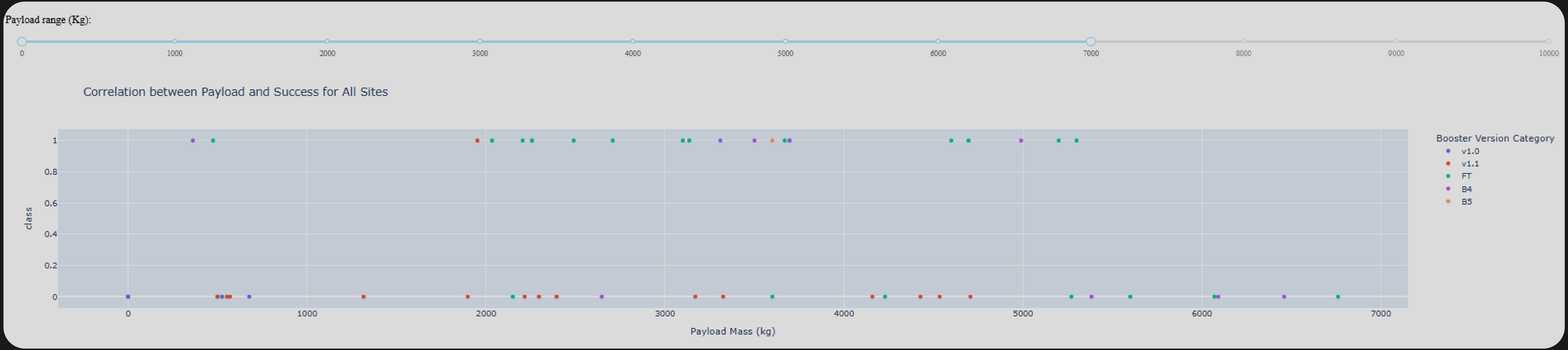
Launch Site with Highest Success Ratio



*Launch Site Success Ratio — Success = 1 (Blue), Failure = 0 (Red)

This pie chart reflects a strong success ratio at KSC LC-39A launch site, with nearly 77% of missions resulting in successful landings. The remaining 23% correspond to failed attempts, based on binary class labels.

Payload vs. Launch Outcome Scatter Plot for All Sites



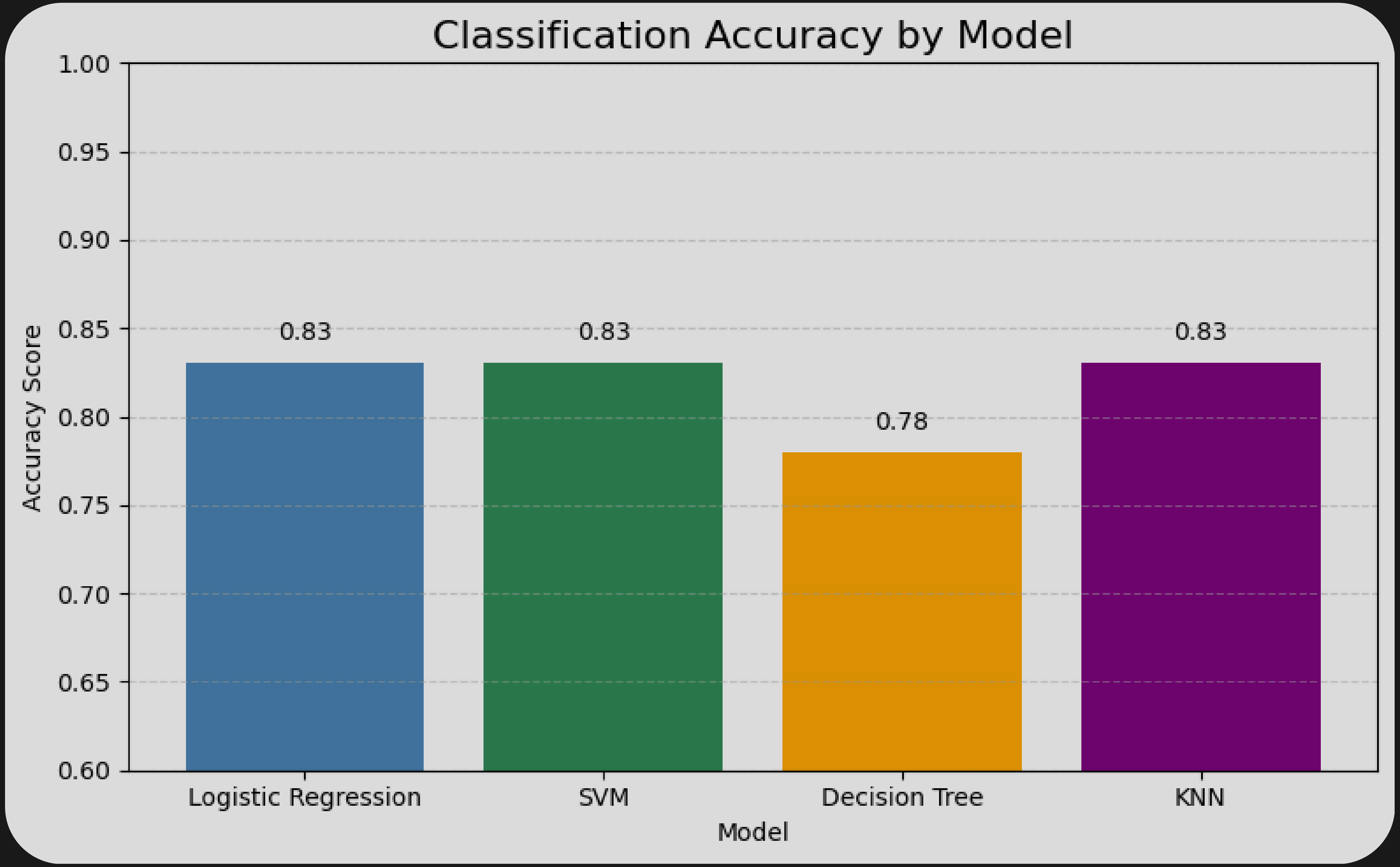
*Payload Mass = 7000 kg — Success is spread across booster versions with no clear leader.

At a payload of 7000 kg, successful launches (class = 1) are scattered across multiple booster versions, with no dominant version consistently outperforming others. The plot suggests that success at this payload level is achievable but not tied to a specific booster

Section 5

Predictive Analysis (Classification)

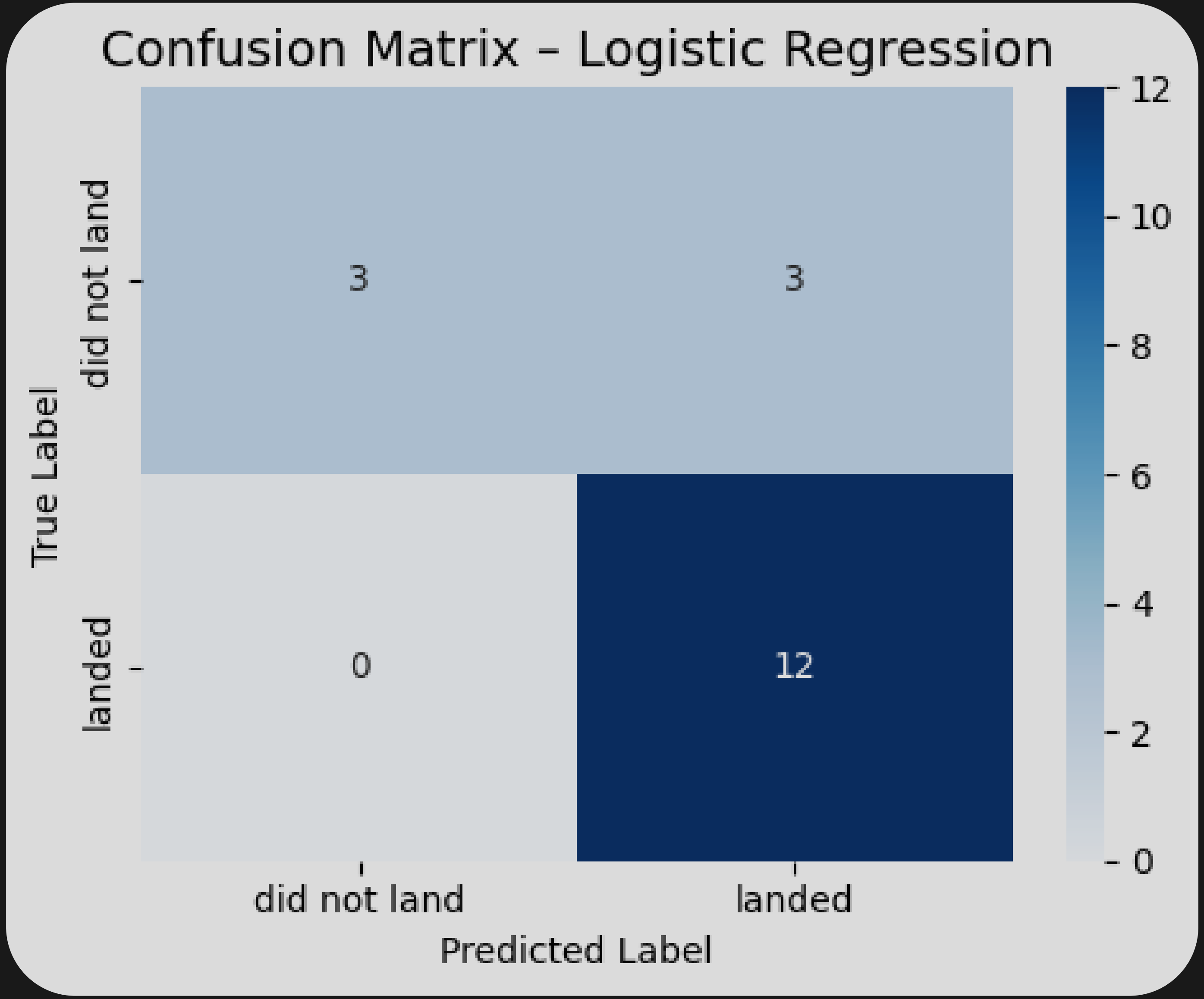
Classification Accuracy



*Classifiers perform comparably , suggesting robustness across linear and nonlinear approaches.

The bar chart summarizes model accuracy for each classifier, with Logistic Regression, SVM, and KNN each achieving 83% accuracy and the Decision Tree slightly lower at ~78% (subject to GridSearchCV tuning variations.)

Confusion Matrix of the Best Performing Model



*Confusion matrix for Logistic Regression, emphasizing prediction strength in identifying successful landings.

The Logistic Regression model accurately differentiates between landed and non-landed launches, with strong diagonal values confirming reliable predictions and minimal misclassifications.

Conclusions

🚀 Performance & Reliability

- SpaceX has consistently improved launch outcomes since 2013, backed by reusable boosters and robust flight experience — especially for LEO, Polar, and ISS missions. KSC LC 39A emerged as the top-performing site for recovery success and heavy payloads.

🗃️ Data-Driven Insights

- SQL-based queries reinforced the dominance of F9 B5 boosters and revealed early reusability breakthroughs, like the first ground pad success in 2015. NASA's heavy collaboration contributed significantly to launch volume and cargo mass.

🗺️ Location Matters

- Spatial analytics pinpoint launch facilities with strong logistical connectivity — particularly CCAFS SLC-40, which sits near key infrastructure. Visual clustering maps support strategic site placement across coasts.

📈 Modeling Success

- Predictive models (Logistic Regression, SVM, KNN) all delivered high recall (1.0) and F1-scores (~0.89), confirming their viability for landing outcome prediction. Model selection should align with project goals, from transparency to scalability.

Appendix

Github URL:

https://github.com/Moaz-Bannora/Testing/tree/main/Data%20Science%20Project

Thanks to my Instructors:

https://www.coursera.org/professional-certificates/ibm-data-science?&instructors

Thank you!