IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Taiwo John
09/21/25

Outline

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

Executive Summary

Summary of Methodologies

- Data: **(91)** Falcon 9 SpaceX launches.
- Putting launch sites on an interactive **Folium** map to measure distances to cities and highways **for risk assessment.**
- **Dash** for exploration; **ML models** (LogReg/SVM/DecisionTree/k-NN) tuned with **GridSearchCV** for **prediction**.

Summary of Results

- **Top site: KSC LC-39A** — 10/13 launches successful **(77%).**
- **Payload sweet spot**: (2.5–5k kg) shows a **57% success rate** with **21 launches** while (7.5–10k kg) also looks high (60%) but is based on only **5 launches**.
- **Best model: Decision Tree** (CV_Score=0.89; Test_Score=0.83).

Introduction

Project background and context

We analyze Falcon-9 first-stage landing outcomes to understand what drives success (site, payload, geography) and to build simple tools that help planning and bidding decisions.

Data used: SpaceX course CSVs (spacex_launch_dash.csv, spacex_launch_geo.csv): Launch Site, Payload Mass (kg), Booster Version Category, Class(1,0) , site lat/long.

Approach: EDA (plots + SQL), Folium map, Plotly Dash, and classification models.

Problems you want to find answers

- Which site has the largest successful launches?
- Which site has the highest launch success rate?
- Which payload range(s) has the highest launch success rate?
- Which payload range(s) has the lowest launch success rate?
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?
- Which Model Performed the Best?

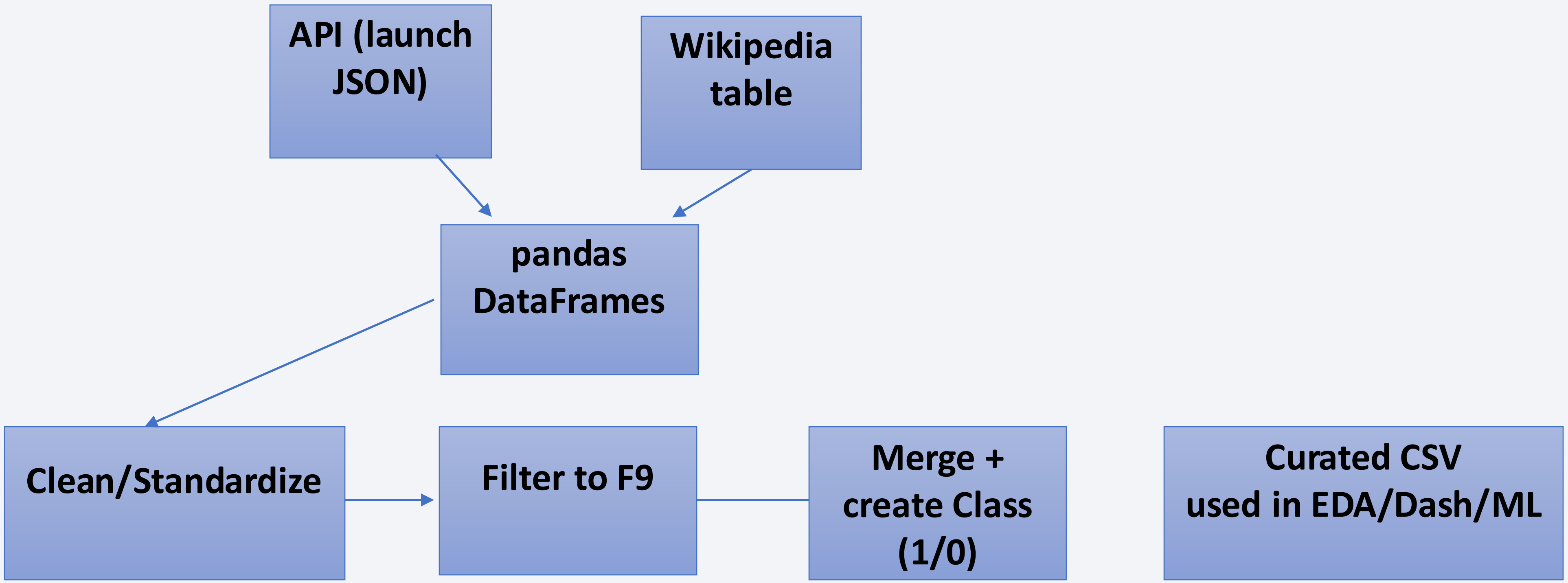
Section 1

Methodology

Methodology

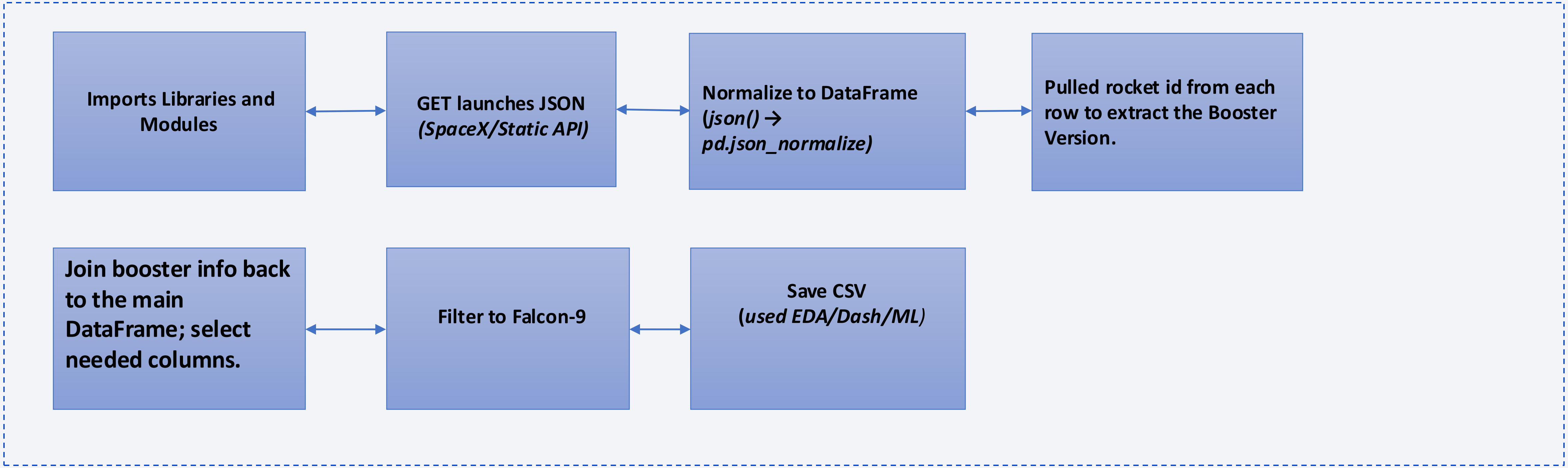
- Collected SpaceX launch records via REST API + Wikipedia web-scrape; cleaned & merged on flight/date to keep Launch Site, Payload Mass (kg), Orbit, Booster, and landing Class (1/0); **scoped to Falcon-9 only** and exported a curated CSV for EDA, Dash, and ML. (DATA COLLECTION)
- Perform Data Wrangling by counting launches by site (CCAFS SLC-40 had the most), grouped by target orbit (ISS led with 21 launches), and converted landing outcomes into a binary **Class** label (1 = success, 0 = failure) for EDA and modeling.
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.
- Perform predictive analysis using ML models by splitting the data into test/train, one hot encoded categoricals, standardized numeric and using LogReg/SVM/Tree/k-NN for the Models, For tuning we used GridSearchCV that picks the best estimator and for evaluation we Computed accuracy and the confusion matrix.

Data Collection



- Collected SpaceX API + Wikipedia + site-coords data, cleaned/standardized in pandas, filtered to Falcon-9, merged and created a binary “Class” label (1=landed, 0=not), then exported a curated 91-row CSV for EDA, Dash, and ML.

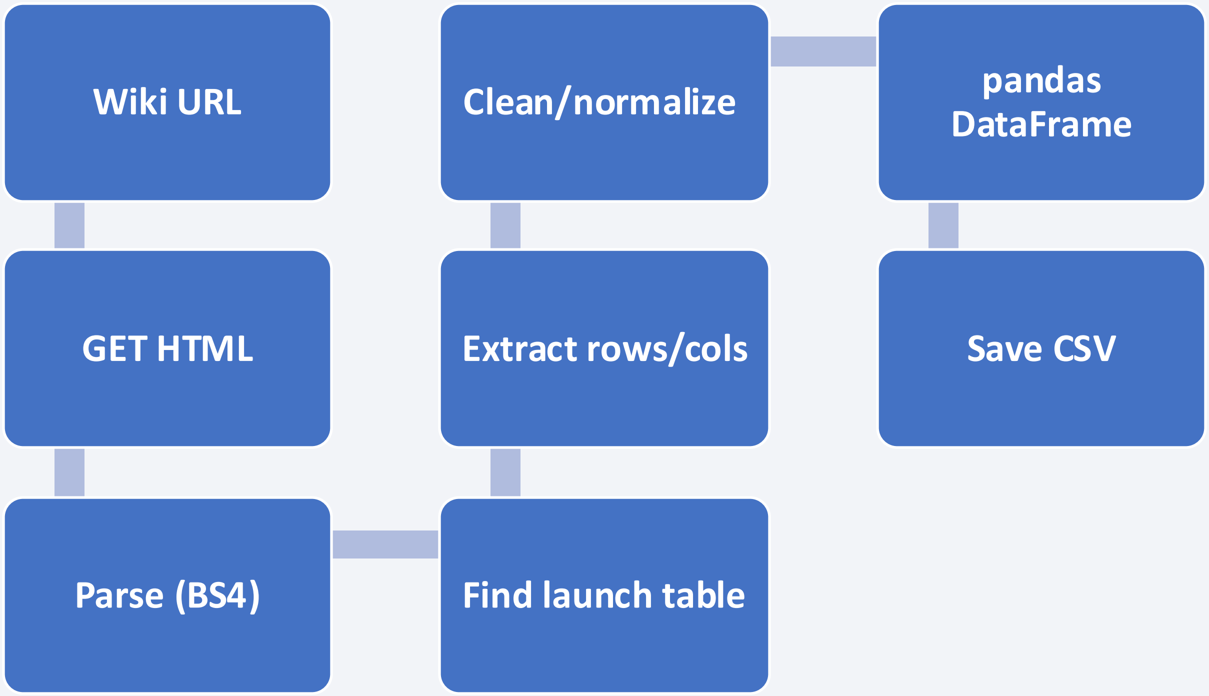
Data Collection – SpaceX API



Using **pandas** and **requests**, we fetched SpaceX launch JSON via GET, normalized it to a DataFrame, looked up each rocket's **BoosterVersion** using its rocket id, merged that back with payload/launch-site fields, **filtered to Falcon-9** records, and saved the curated CSV for EDA.

Here is a Github link : https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

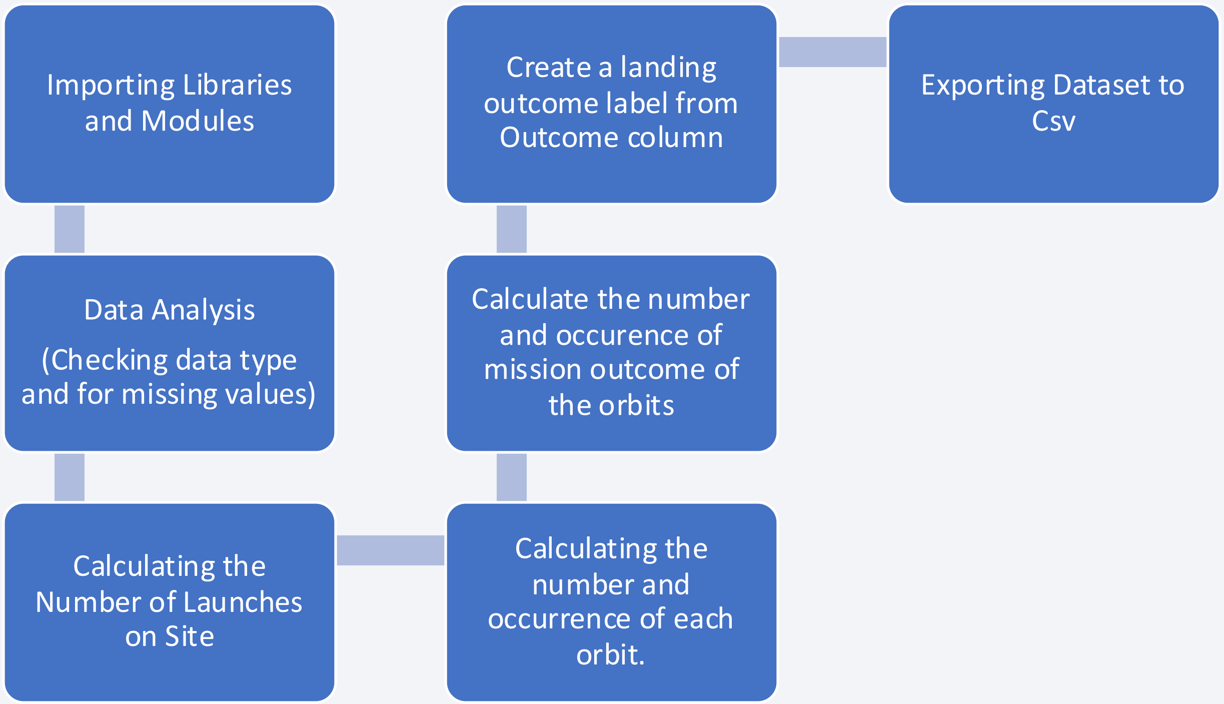
Data Collection - Scraping



We sent a **GET** request to the Falcon-9 launches page, **found the main launch table**, **extracted its rows/columns** with BeautifulSoup into a pandas DataFrame, then **cleaned & normalized** fields (strip refs, parse “kg”→numbers, standardize dates) using helpers like get_mass and booster_version, and saved a curated **CSV** for EDA/Dash/ML.

Github link: https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



imported pandas/numpy, checked data type and missing values, identify categorical fields (Launch Site, Orbit). We then counted activity: **CCAFS SLC-40 launch Site had the most launches (55)**, and the most common **orbit (destination path around Earth) was ISS**.
and the most frequent landing outcome was **successful ASDS (drone-ship) recovery** with **41** cases. Created target label **Class: 1** if landing succeeded (True RTLS/ASDS/Ocean), **0** if failed or no attempt (False…, None…).Finally Exporting Dataset to csv file format.

Github Link: https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

EDA with Data Visualization

- A scatter plot of Flight number(x) against the payload mass and had the points colored by success vs fail, This was used to check two ideas at once do the teams get better with experience(higher flight number = more success) and do heavier payloads make landings harder, and
- A scatter plot of Flight Number(x) against the Launch Site: we used this plot to see how performances changes over time at each site.
- A scatter plot of Payload and Launch Site: we used this to see if site success gaps are driven by the **types of payloads flown** (payload mix) rather than the site itself.
- A bar chart of success rate by orbit to check whether orbit type is associated with landing success.
- A scatter plot of orbit and Flight number: we used this plot To check if later flights do better (experience) within each orbit.
- A scatterplot of payload mass and orbit type was used to see **within each orbit**, how **heavier vs lighter payloads** relate to landing success
- A line chart of average landing success rate by year we used it To show the **trend over time**—i.e., SpaceX's **learning curve**.

Github Link: https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- The Distinct query was used to get the Unique launch site in the space mission by which we had 4 launch sites.
- Displayed 5 Launch Sites records where the launch site begin with ‘CCA’ it gave us record of (CCAFS LC-40, CCAFS SLC-40).
- **Total CRS payload:** Filtered to NASA CRS missions (SQL: WHERE Payload LIKE '%CRS%') and summed PAYLOAD_MASS__KG_ (SQL: SUM) → **111,268 kg**.
- **Booster-specific payload:** Filtered to **F9 v1.1** (SQL: WHERE Booster_Version LIKE 'F9 v1.1%') and averaged PAYLOAD_MASS__KG_ (SQL: AVG) → **≈2,534.7 kg**.
- **First ground-pad success:** Filtered Landing_Outcome for **Success** + **ground pad** (SQL: WHERE … LIKE), then took earliest date (SQL: MIN(Date)) → **2015-12-22**.
- *WHERE Landing_Outcome LIKE ‘%Success%’ AND ‘%drone ship%’ AND payload 4000–6000→ boosters with successful drone-ship landings in that payload range.*
- SELECTED THE MISSION OUTCOME(SELECT),Grouped it(GROUP BY) and counted each of the value count(Count(*) AS total) and rename as Total
- **SQL (2015 drone-ship failures):** Filtered 2015 rows and picked **landing outcomes containing “failure” + “drone”**; using substr(Date,6,2) for month. Result: **Jan (01)** – *F9 v1.1 B1012*, **CCAFS LC-40**; **Apr (04)** – *F9 v1.1 B1015*, **CCAFS LC-40**.

Github Link: https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Created a Folium map object with an initial center location to be NASA Johnson Space Center at Houston, Texas. Folium needed a starting map and a familiar center spot in the U.S so when the map opens you are already near Florida.
- Added a circle for each launch site in the dataframe, we used **Folium Circle** to create the circle and used **Marker** to give a text label. We drew a visible circle to highlight each site's location/area and added a labeled marker so viewers can instantly identify the site and use it for distance/route analyses later.
- We plotted every launch as a green/red marker (**Folium Marker + Icon**) and clustered overlapping pins (**MarkerCluster**), so you can quickly **see success vs. failure** at each site without a crowded map.
- Used **folium.plugins.MousePosition** to read lat/long anywhere on the map, Then Wrote a **Haversine** function (calculate_distance) to measure km from the launch site to **coastline, railway, highway, and nearest city**. Drew a **folium. Polyline** between the site and each picked point. Added a **Marker/DivIcon** label to show the distance. We did this because this shows the sites are near the coast and kept away from cities/roads—good for safety and logistics.
- Example result: CCAFS LC-40 at (28.5623, −80.5774) — drew a 1-km circle (Folium Circle) and a labeled marker (Folium Marker + DivIcon). Used this point as the anchor to measure distances: 0.95 km to coastline, 13.05 km to nearest highway, 21.12 km to railway, and 54 km to the nearest city (via PolyLine from site to each point).

Github Link: https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

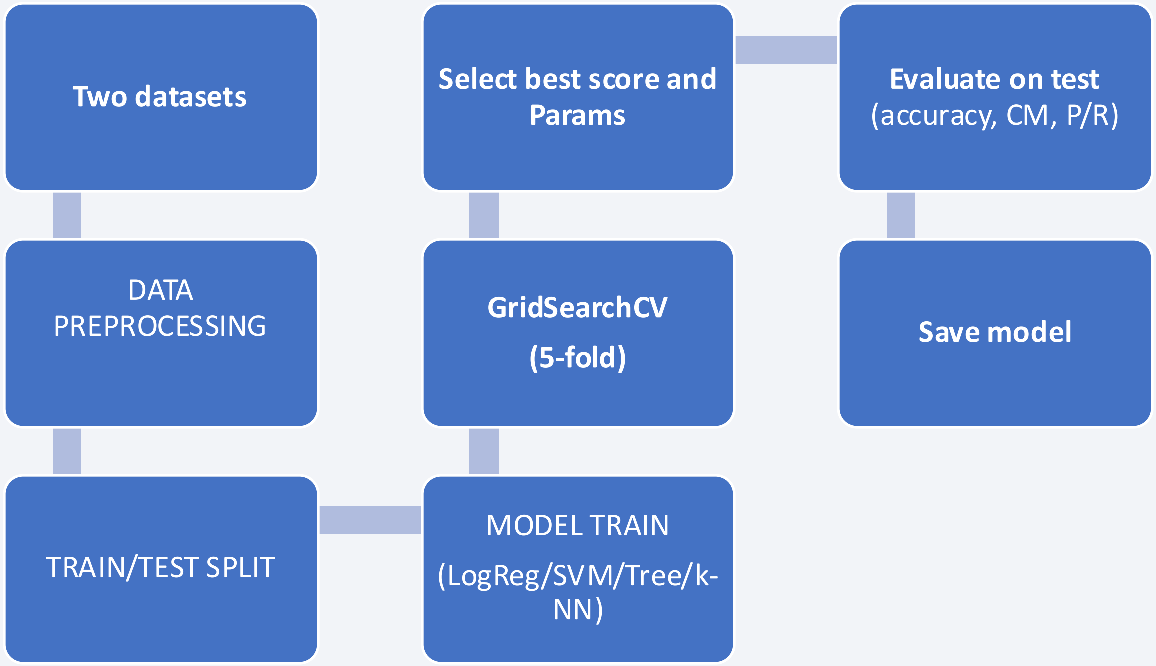
- Built a Plotly Dash app to explore Falcon-9 landings by **site** and **payload**. We did this To let stakeholders **interactively explore site × payload scenarios**, see success rates instantly, and spot the payload.

Plots & Widgets

- **Site dropdown** (dcc.Dropdown id="site-dropdown") – choose *All Sites* or a specific launch site. This help us so users can **filter the visuals by launch site**—compare sites or focus on one updating the pie (success vs. failure for that site) and the scatter (payload vs. outcome) while avoiding cross-site mixing.
- **Payload slider** (dcc.RangeSlider id="payload-slider") – filter to a payload-mass range. This help us so users can **focus on a specific payload range**—see how success changes with mass.
- **Success pie chart** (dcc.Graph id="success-pie-chart") – *All Sites*: pie of **total successes by site** (sum of class), *One site*: pie of **Success vs Failure** for that site. This was done because all site view compare which sites contribute most successes briefly, while single site view shows that site success vs failure.
- **Payload vs Outcome scatter** (dcc.Graph id="success-payload-scatter-chart") – x = payload (kg), y = success (0/1), color = booster/version (or site), with hover details. This was done to see how payload mass relates to landing success and spot patterns by booster/site.

Github Link : https://github.com/taiwotjohn13/SPACEXPROJECT/blob/main/Dash.ipynb

Predictive Analysis (Classification)

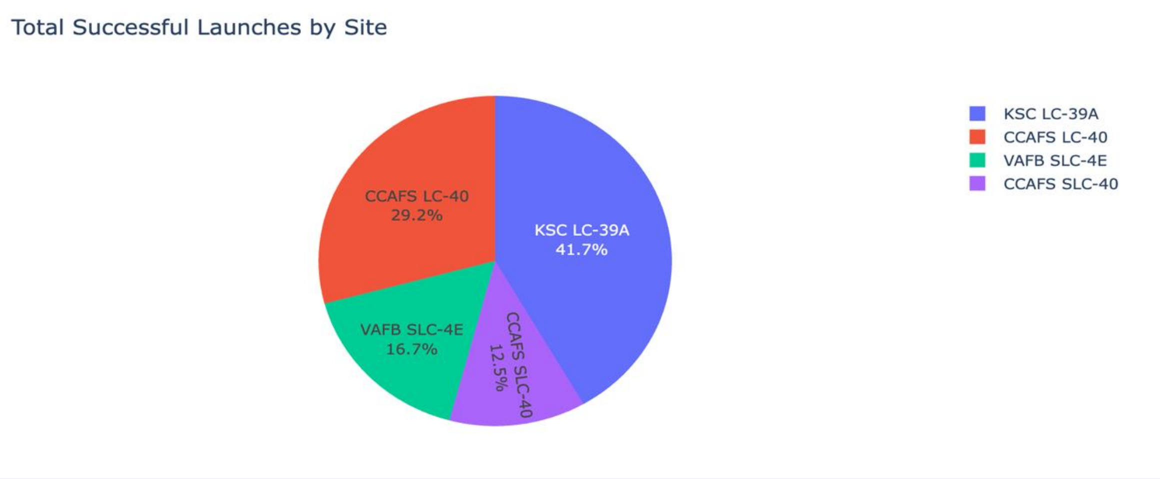


- **Built:** created the target y = Class and a numeric feature matrix X (one-hot encoded categoricals, standardized numerics), then did an 80/20 train–test split and instantiated four scikit-learn classifiers—Logistic Regression, SVM, Decision Tree, and k-NN—inside a pipeline so preprocessing is fit only on the training data.
- **Evaluated:** Tested each tuned model on the held-out test set using **accuracy** and the **confusion matrix** to compare performance and error types.
- **Tuning and Improvment**: For each pipeline, ran GridSearchCV (cv=10, scoring='accuracy', refit=True) over model-specific hyperparameters; GridSearch picks the best settings per model.
- **Decision rule:** Pick the model with the **highest CV accuracy score.**

Results

- Most active site: **CCAFS SLC-40** with **55** Falcon-9 launches.
- Most common orbit**: ISS (21 missions).**
- Landing outcomes: **ASDS (drone-ship) successes = 41.**
- **Flight number vs launch site**: : Early failures at CCAFS SLC-40; success improves after moving to KSC LC-39A and later improves back at CCAFS.
- **Flight Number vs Payload (colored by outcome):** Success increases with experience; very heavy payloads are less likely to land.
- **Success rate by Orbit (bar):** Highest for ES-L1/GEO/HEO, lower for GTO/SO (small-sample orbits noted).
- **Payload vs Launch Site:** VAFB SLC-4E has no >10,000 kg payloads; payload mix differs by site.
- **Yearly trend (line):** Average success climbs from near 0 (pre-2014) to 0.9 by 2019–2020.

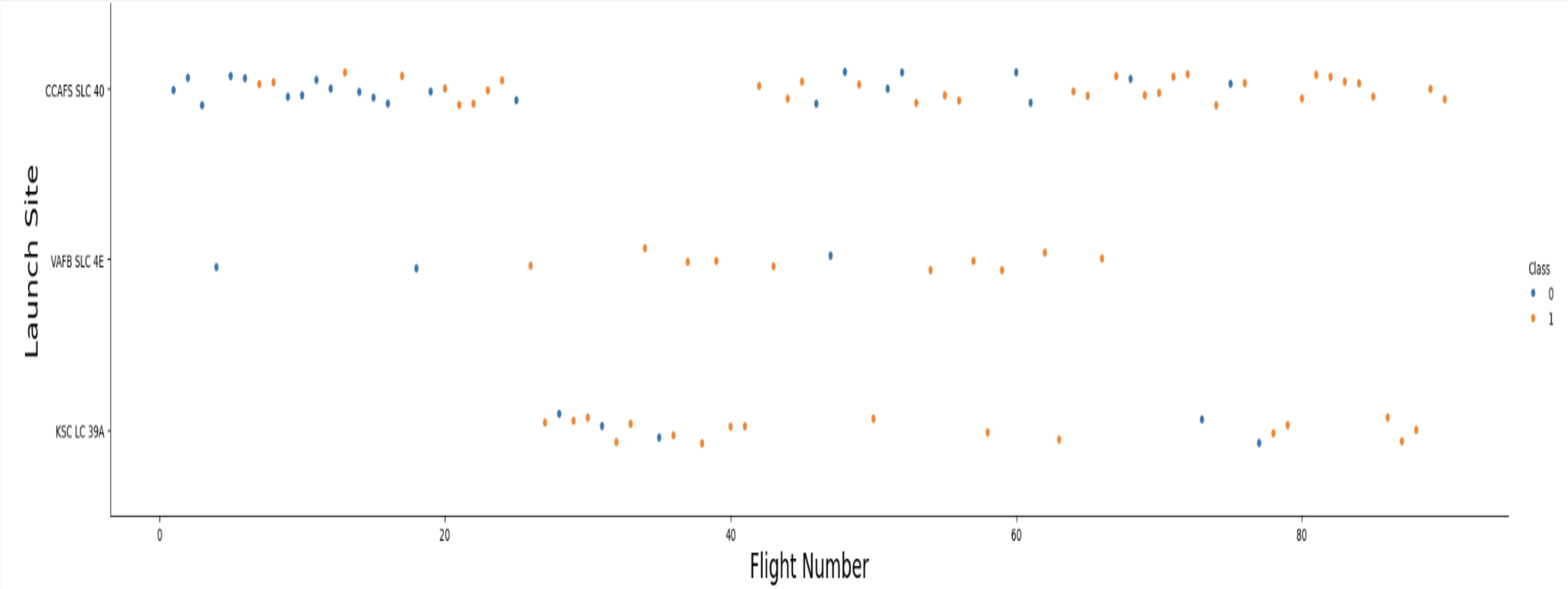
Model	CV accuracy	Test accuracy
Logistic Regression	84.6%	83.3%
SVM	84.8%	83.3%
Decision Tree	88.8%	83.3%
k-NN	84.8%	83.3%



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

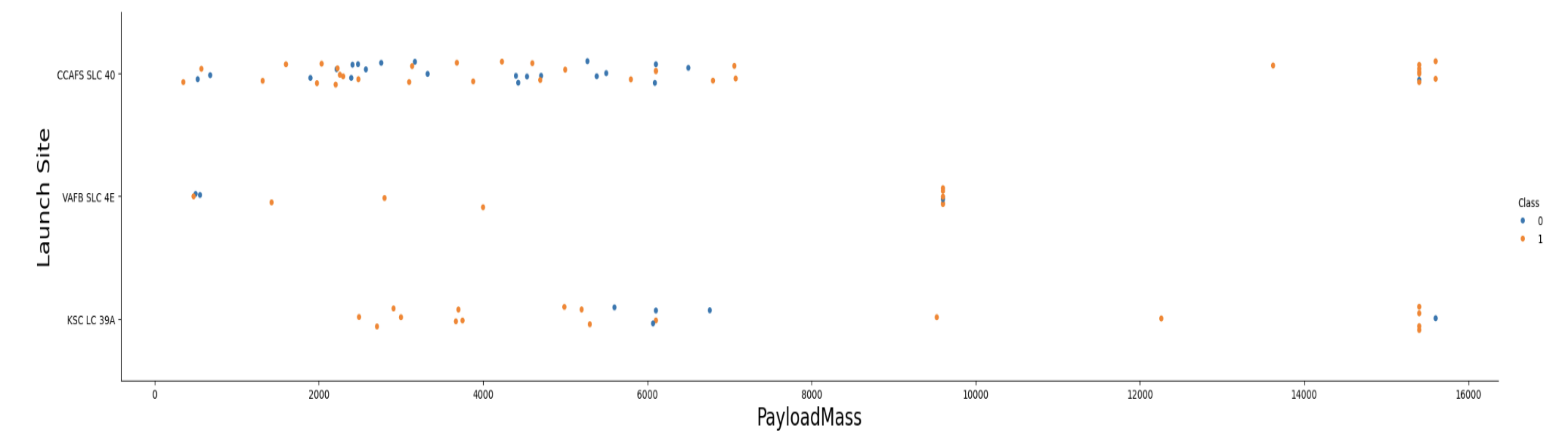


The first 25 launches mostly took place at CCAFS SLC-40 and experienced a high number of failures (Class 0). From Flight 25 to 40, there was a shift to KSC LC-39A, where the success rate (Class 1) noticeably improved. This suggests that changing the launch site may have contributed to better outcomes.

Between Flight 50 and 65, a major improvement occurred with almost no failures, indicating growing success. Later launches between Flight 70 and 90 returned to CCAFS SLC-40, but this time with much higher success, showing that experience and repeated launches led to better performance over time.

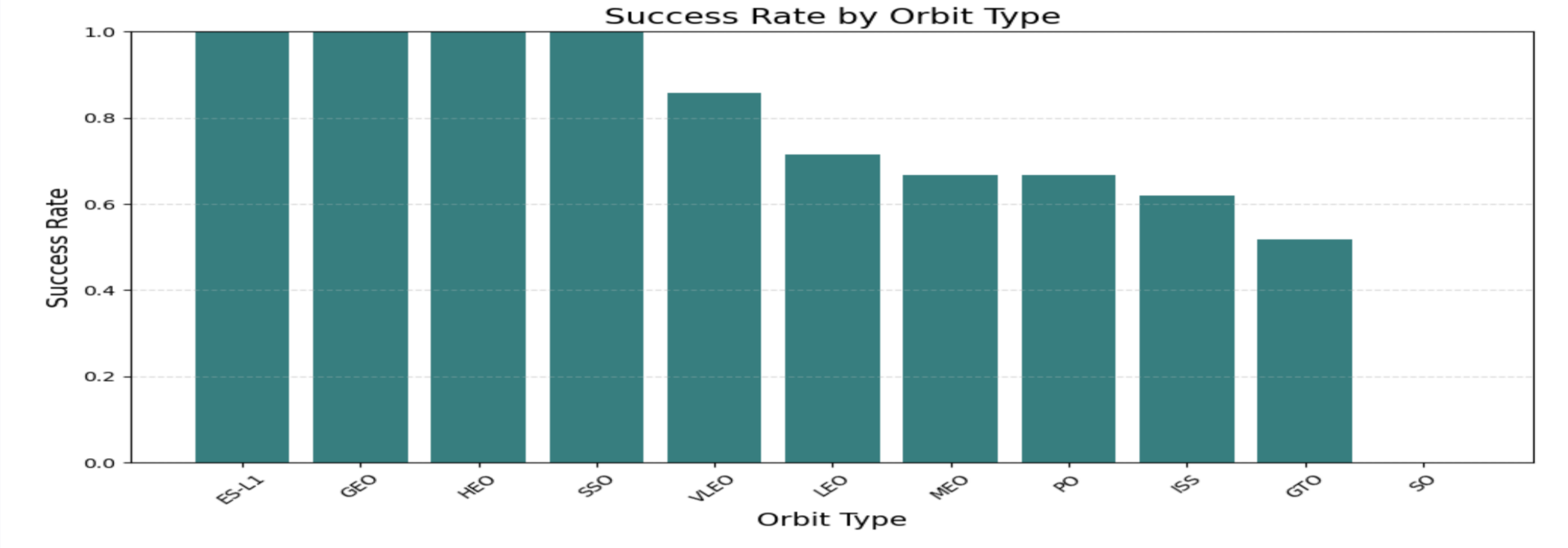
Conclusion: The more launches SpaceX performed, the more successful they became, continuous practice and learning led to consistent improvements, regardless of the launch site.

Payload vs. Launch Site



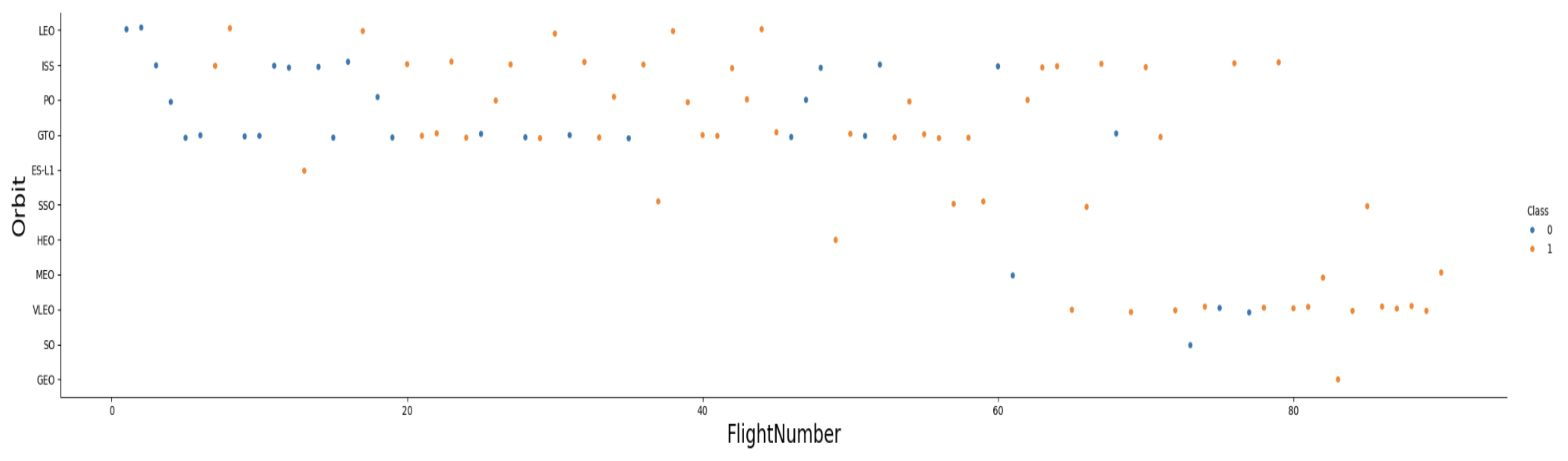
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



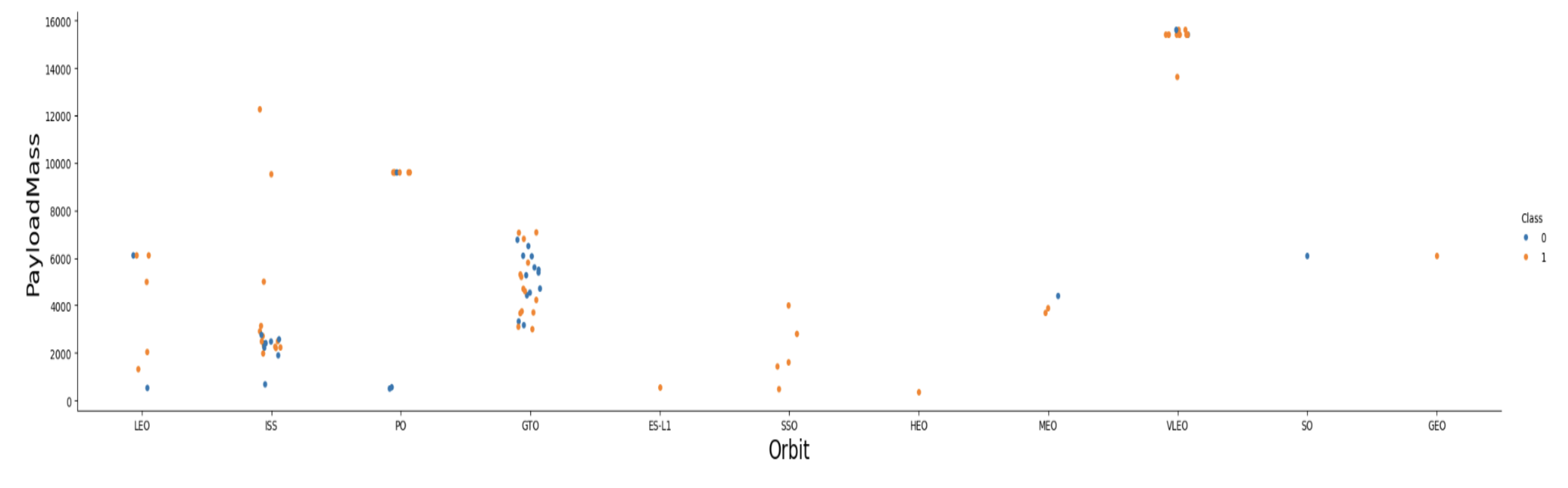
From the bar chart, we observe that orbits like ES-L1, GEO, HEO, and SSO have a perfect success rate, suggesting that missions to these orbits are highly reliable . Meanwhile, orbits like GTO and SO have lower success rates, because they are harder orbits to reach, especially during early launches when SpaceX was still learning. This implies that the orbit type influences launch success, and certain orbits may require more caution or testing.

Flight Number vs. Orbit Type



You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

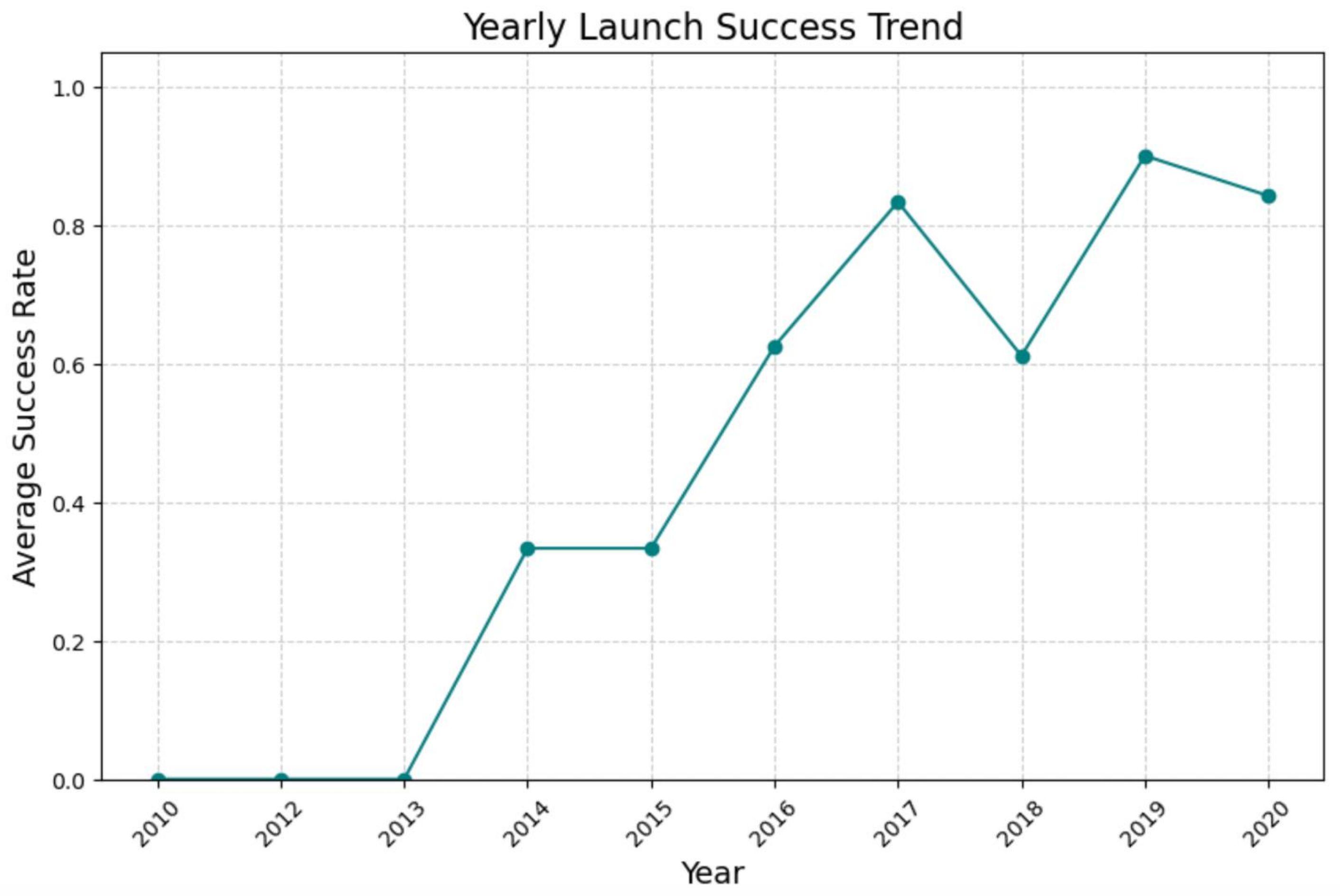
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

Name of All Launch Site Names

1. CCAFS SLC-40 (Cape Canaveral)
2. KSC LC-39A (Kennedy Space Center)
3. VAFB SLC-4E (Vandenberg)
4. CCAFS LC-40 Cape Canaveral Space Force Station — Launch Complex 40.

```
SELECT DISTINCT "Launch_Site"
FROM SPACEXTABLE;
```

- Used **DISTINCT** to de-duplicate and confirm there are 3 launch sites in the dataset.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcom
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 attemp
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

```
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
```

- Returns the first 5 records from **SPACEXTABLE** where **Launch_Site** starts with "CCA" (i.e., Cape Canaveral sites).

Total Payload Mass

- TOTAL_NASA_CRS_PAYLOAD: **111268**

```
SELECT SUM("PAYLOAD_MASS__KG_") AS TOTAL_NASA_CRS_PAYLOAD
FROM SPACEXTABLE
WHERE "Payload" LIKE '%CRS%’
```

- the total payload mass (kg) for all missions whose Payload field contains “CRS” is 111268

Average Payload Mass by F9 v1.1

- AVG_PAYLOAD_F9_V1_1 :2534.67
- SELECT AVG("PAYLOAD_MASS__KG_") AS AVG_PAYLOAD_F9_V1_1

FROM SPACEXTABLE

WHERE "Booster_Version" LIKE "F9 V1.1%"

- Computes the average payload mass for launches whose Booster_Version starts with “F9 V1.1”; the result (2,534.7 kg) means Falcon 9 v1.1 flights carried about 2.5 tons on average

First Successful Ground Landing Date

- SELECT MIN(Date) AS First_Successful_Ground_Landing

FROM SPACEXTABLE

WHERE

"Landing_Outcome" LIKE '%Success%’

AND "Landing_Outcome" LIKE '%ground pad%’

- First_Successful_Ground_Landing: 2015-12-22

- Finds the earliest (MIN) launch date with a successful ground-pad landing; the result 2015-12-22 tells us SpaceX’s first on-land booster landing happened on **Dec 22, 2015.**

Successful Drone Ship Landing with Payload between 4000 and 6000

- SELECT "Booster_Version"

FROM SPACEXTABLE

WHERE "Landing_Outcome" = 'Success'

AND "PAYLOAD_MASS__KG_" > 4000

AND "PAYLOAD_MASS__KG_" < 6000

Booster_Version
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

- Returns the **booster versions** that successfully landed on missions with payload mass between 4,000 and 6,000 kg; the results (e.g., F9 B5 B1046.2, B1047.2, B1048.3…) show several Falcon 9 Block 5 cores achieving success in that payload band.

Total Number of Successful and Failure Mission Outcomes

```
SELECT "Mission_Outcome", Count(*) AS total

FROM SPACEXTABLE

GROUP BY "Mission_Outcome”
```

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

- Counts missions by outcome: 98 “Success,” 1 “Failure (in flight), plus a few rare/duplicate labels—overall high success, with a small labeling inconsistency.

Boosters Carried Maximum Payload

- SELECT "booster_version" FROM
- SPACEXTABLE
- WHERE "PAYLOAD_MASS__KG_" = (
-     SELECT MAX("PAYLOAD_MASS__KG_")
-     FROM SPACEXTABLE
-)

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

- This query finds all boosters whose flights carried the maximum payload mass in the table; the list of F9 B5 variants returned means multiple launches share that same highest payload value, so they all appear.

2015 Launch Records

```
SELECT
  substr(Date, 6, 2) AS Month,
  "Landing_Outcome",
  "Booster_Version",
  "Launch_Site"
FROM SPACEXTABLE
WHERE
  substr(Date, 0, 5) = '2015' AND
  "Landing_Outcome" LIKE '%failure%' AND
  "Landing_Outcome" LIKE '%drone%'
```

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

This query filters 2015 launches for drone-ship landing failures and lists the month, booster version, and site; the result shows two failures—Jan (01) and Apr (04) 2015—both F9 v1.1 at CCAFS LC-40.

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

```
SELECT
"Landing_Outcome",
COUNT(*) AS outcome_count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY outcome_count DESC;
```

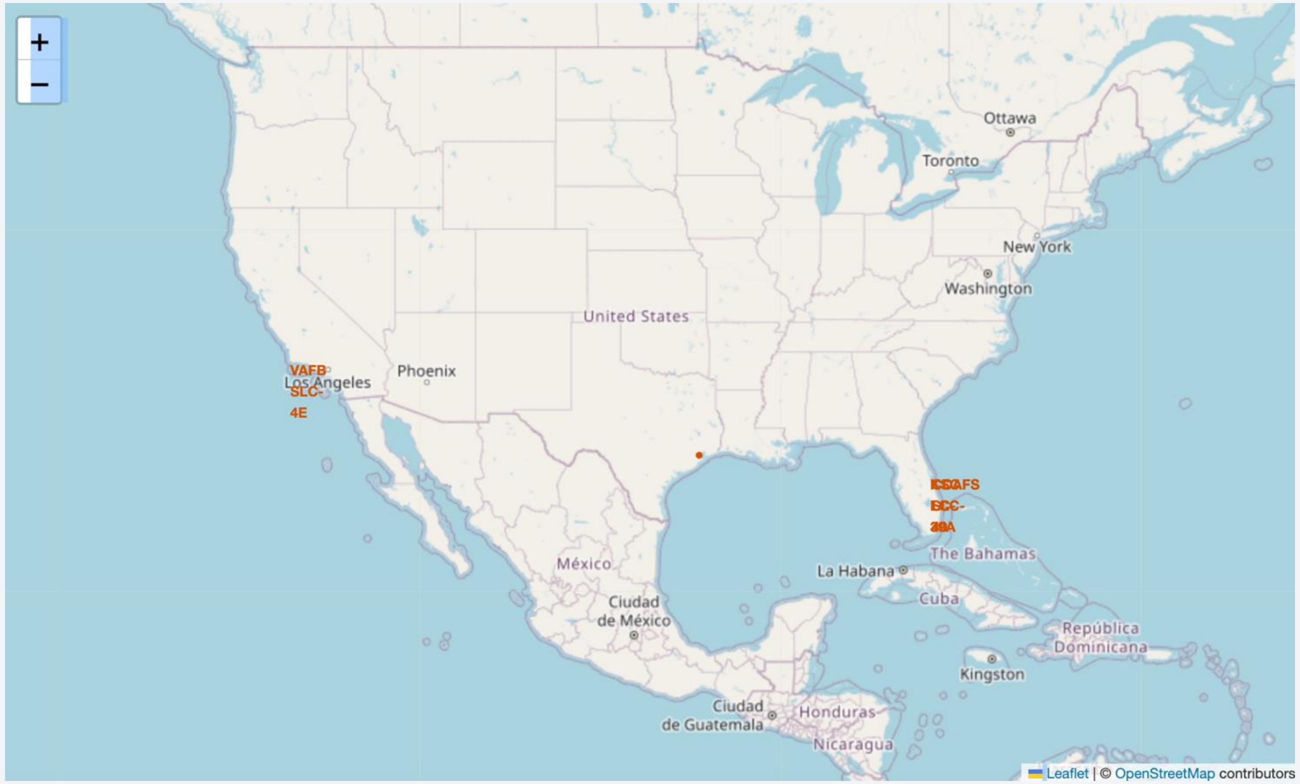
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Counts and ranks landing outcomes between 2010-06-04 and 2017-03-20; results show "No attempt" is most common (10), followed by Success (drone ship) = 5, Failure (drone ship) = 5, Success (ground pad) = 3, etc.—i.e., many missions in that window didn't attempt a landing, with a mix of successes and failures among attempts.

Section 3

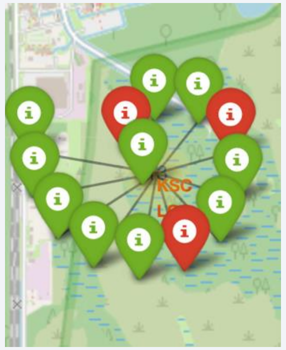
Launch Sites Proximities Analysis

<SpaceX Launch Sites & Outcomes-Folium Overview>



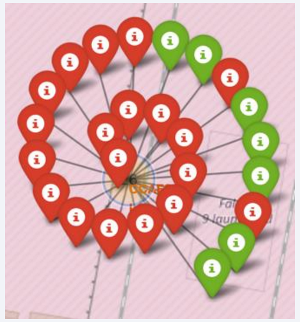
- Labeled site markers (orange text): VAFB SLC-4E (California) and KSC LC-39A / CCAFS SLC-40 (Florida).
- All sites are on the coast so launches go over the ocean (safer) and ships can handle recoveries.
- The three Florida pads (28.6°N) are closer to the equator; VAFB SLC-4E (34.6°N) is farther north.

<Launch Outcomes by Site(Folium)>

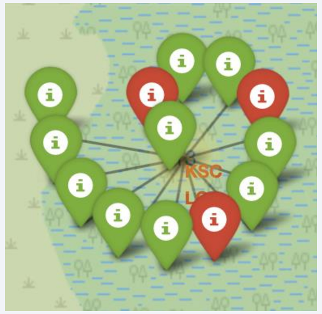


KSC cluster = lots of launches, and most pins are green → mostly successful**.**

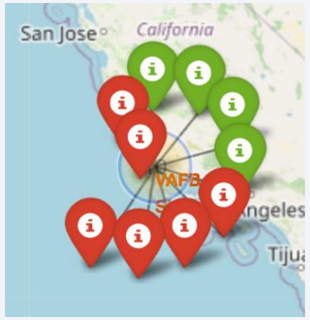
CCAFS SLC-40: only 7 launches, with several red pins (failures) and fewer greens (successes) → mixed results and lower reliability here compared to KSC.



CCAFS LC-40 (~26 launches): mostly failures (many red pins), only a few successes (greens).

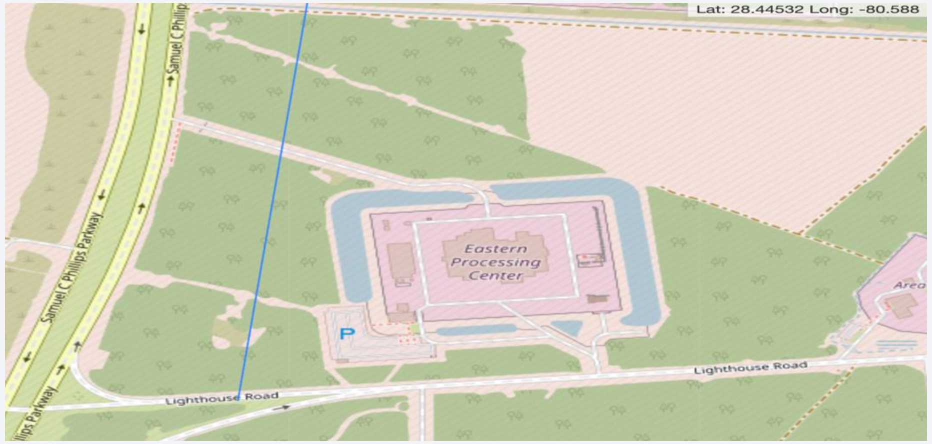


KSC LC-39A: mostly green pins with only a few red → high success rate at this pad compared to others.



VAFB SLC-4E: about half green/half red → mixed results (lower success rate than the Florida pads).

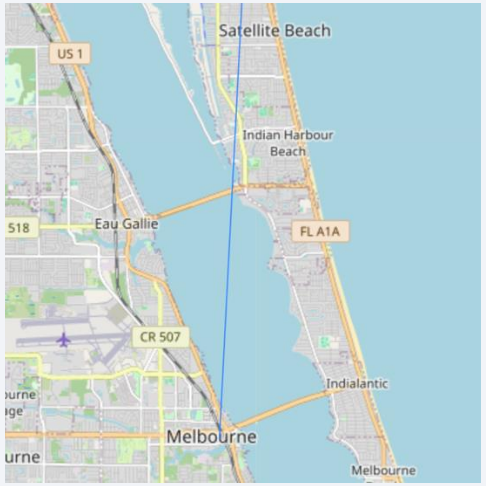
<Launch Site Proximity(Folium)>



The pad is 0.95 km from the ocean, so rockets fly out over water, keeping people and cities safer—this is why launch sites are usually on the coast.

The pad is 13 km from a highway—close enough for easy access, far enough to keep the public safely away.

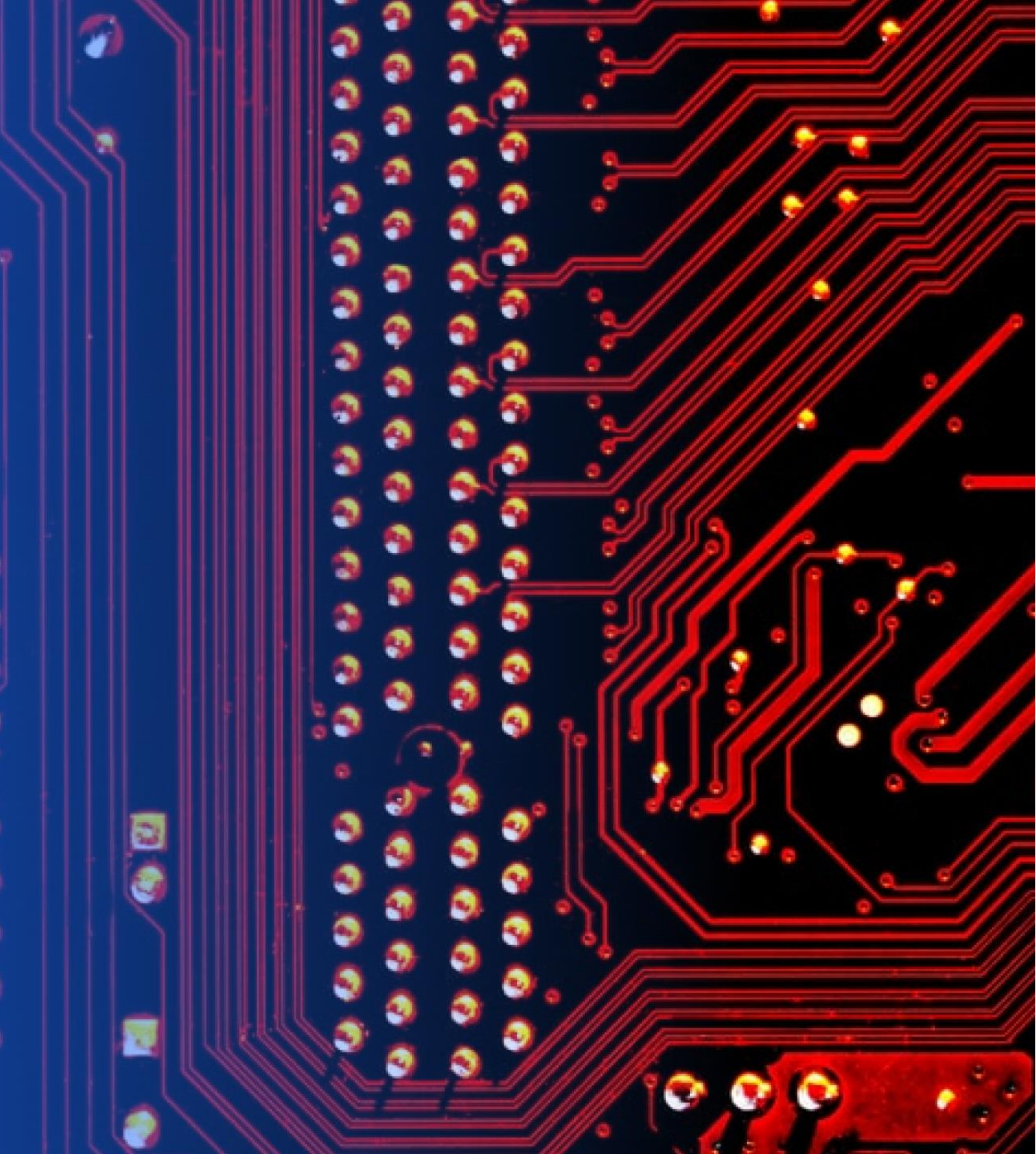
The pad is 21 km from a railway far but workable: ship heavy gear by rail to a nearby hub, then truck it to the site.



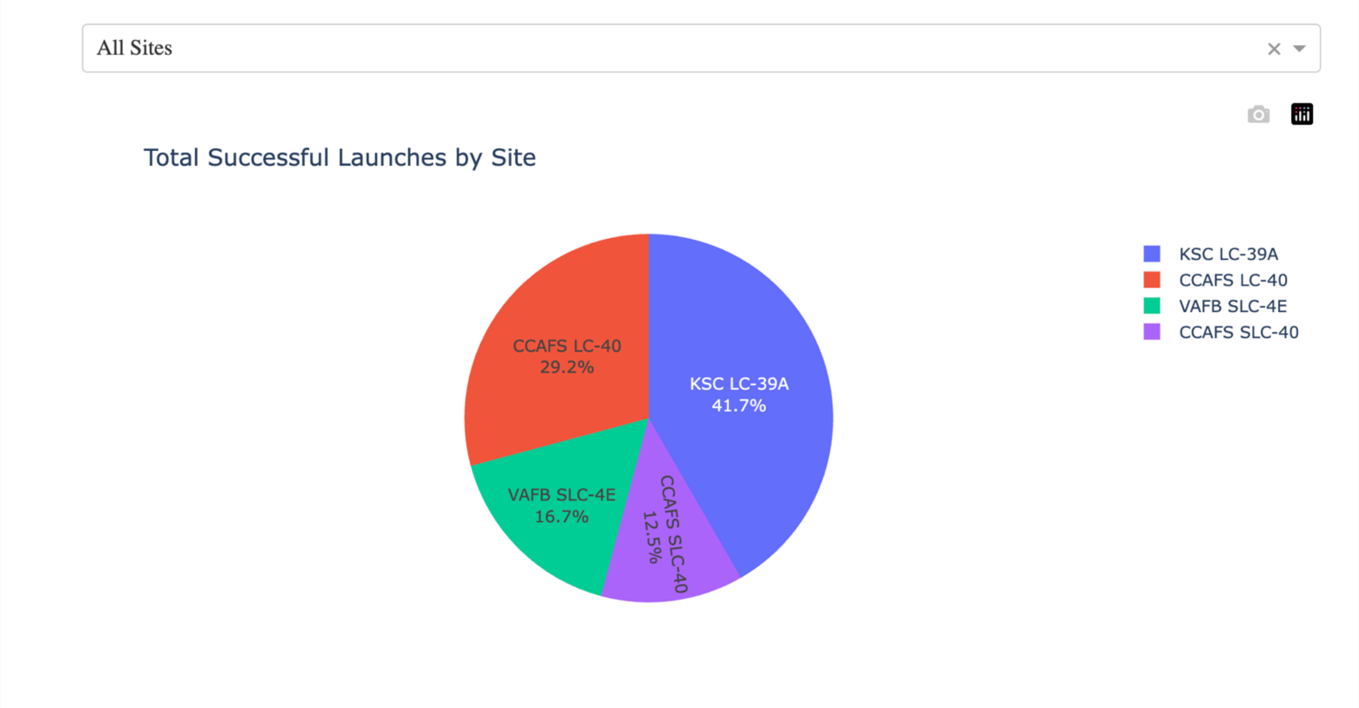
The pad is 54 km from major cities on purpose, to keep people safe and cut noise.

Section 4

Build a Dashboard with Plotly Dash

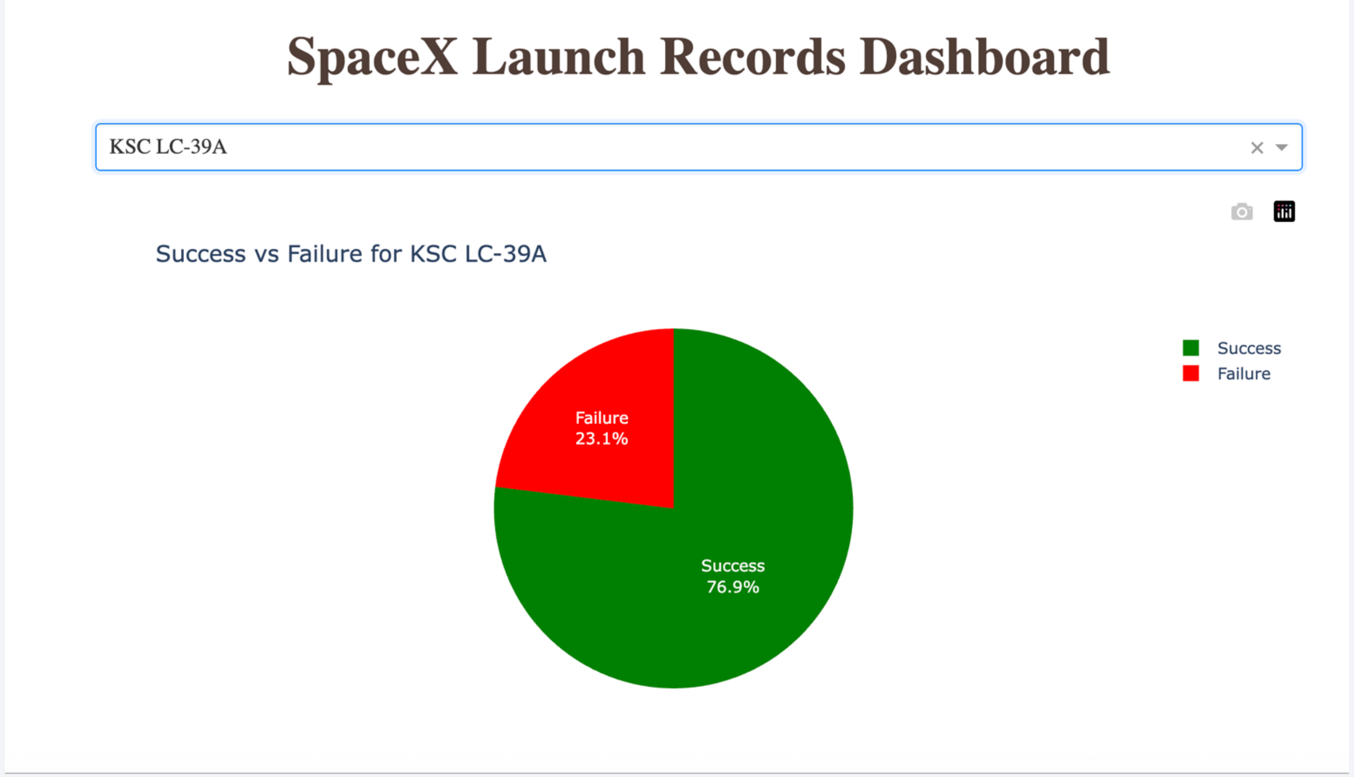


<Launch Successes by Site(Pie-All Sites)>



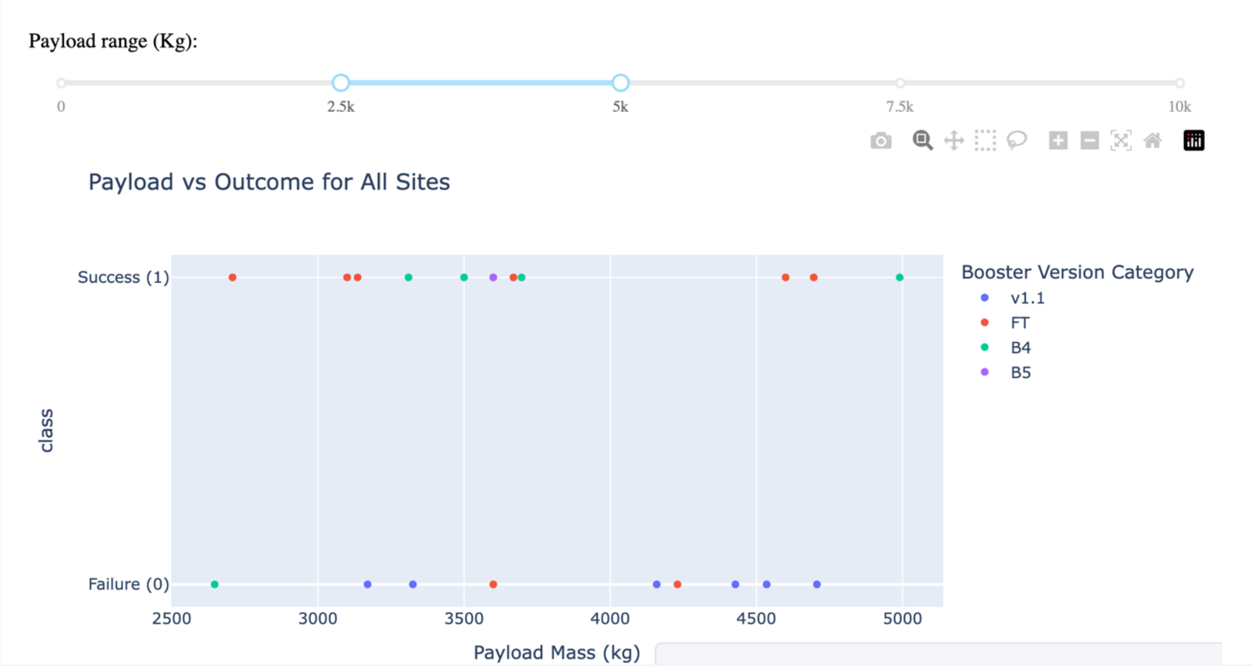
Florida sites (KSC LC-39A and CCAFS SLC-40) account for most successes; VAFB SLC-4E contributes fewer.

<Site With The Highest Success Rate >



KSC LC-39A shows 77% successful launches (green) vs 23% failures (red)—a strong reliability profile for this pad.

<Payload vs Outcome>

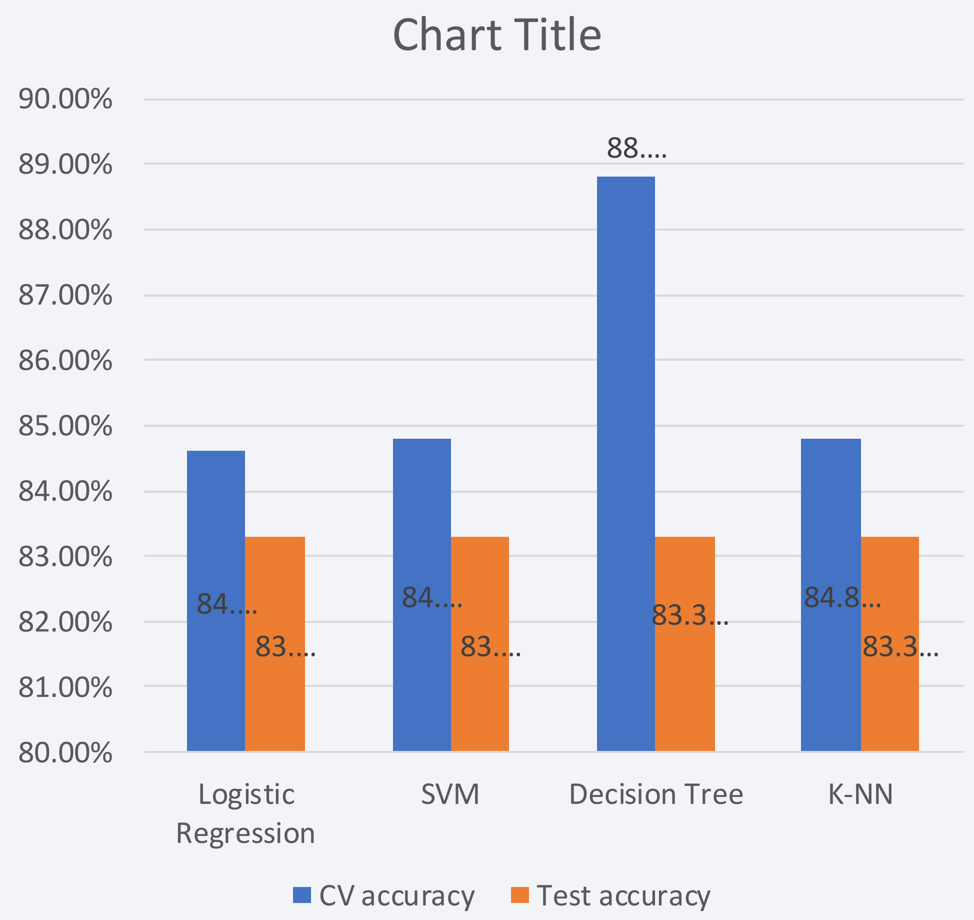


- **Overall success : 57%** (21 successes, 12 failures).
- **Older boosters struggle:** many **v1.1 (blue)** failures, especially near **4000kg–4700kg**.
- **Newer boosters do better: Block 4 (green)** is mostly successful across the band; **Block 5 (purple)** shows success where it appears.
- Pattern: as payload approaches the top of this range, failures rise for older versions, while newer versions maintain higher success.

Section 5

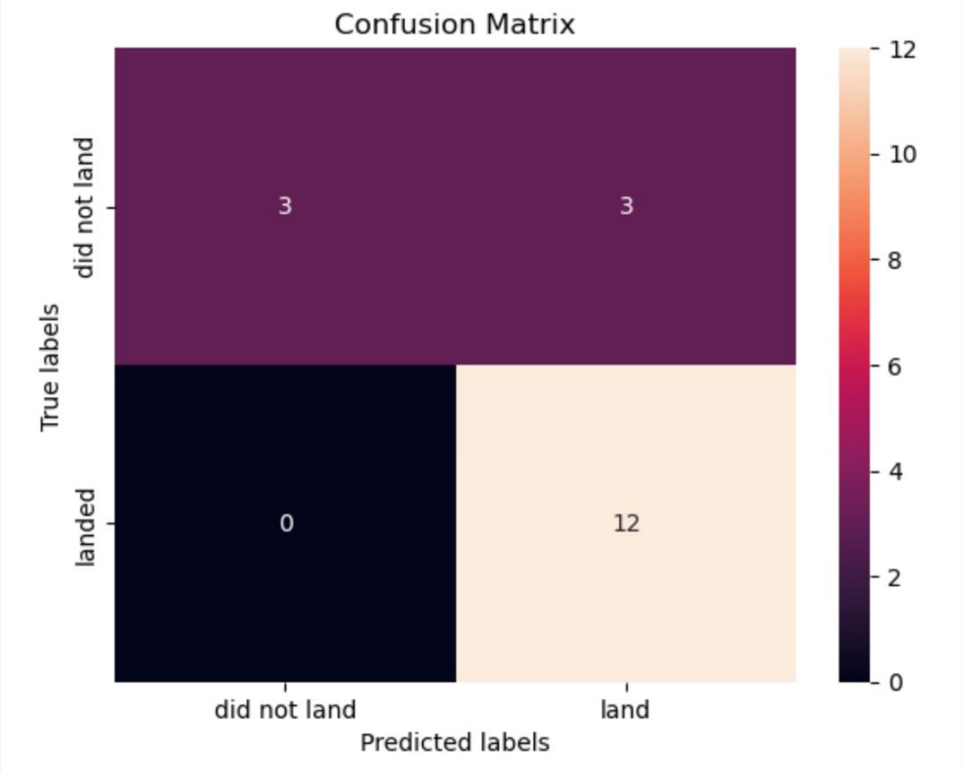
Predictive Analysis (Classification)

Classification Accuracy



- **Finding: Decision Tree** has the highest **CV accuracy (88.8%)**
- **Test accuracy:** all models are **83.3%** (tie).
- The Decision Tree looked best in practice (88.8%), but in the final match all models scored 83.3%, so none proved better in real use. That’s why we say the CV edge didn’t carry over—likely just normal variance or a small test set—so there’s no clear best model from accuracy alone.

Confusion Matrix



- **TN (top-left = 3):** correctly predicted **did not land**.
- **FP (top-right = 3):** predicted **land** but it **did not land** (false alarms).
- **FN (bottom-left = 0):** predicted **did not land** but it **landed** (missed successes) — **none**.
- **TP (bottom-right = 12):** correctly predicted **landed**.

Key metrics from these counts (out of 18 flights):
Accuracy: (3+12)/18 = **83.3%**
Recall/Sensitivity (landed): 12/(12+0) = **100%** → it caught **every landing**.
Precision (landed): 12/(12+3) = **80%** → some predicted landings were wrong.
Specificity (did not land): 3/(3+3) = **50%** → weak at identifying failures.

- The model is **great at never missing a landing** (recall 100%) but **over-predicts landings**, misclassifying about half of the failures as successes.

Conclusions

- **Launch Overview:** CCAFS SLC-40 is the busiest pad, and it is reserved more for crew/heavy/complex payloads, ISS is the top orbit(21) by which NASA pays private companies like SPACEX to routinely fly missions to the ISS. All pads are coastal; at SLC-40 we measured **0.95 km** to the ocean, **13 km** to highway, **21 km** to rail, **54 km** to city—set up for safety and smooth operations.
- **Learning curve dominates:** Early failures clustered at CCAFS SLC-40 referencing the(**Flight Number vs. Launch Site** scatter plot), success jumped after moving to **KSC LC-39A** and stayed high even after returning to CCAFS; the yearly success rate rose to 90% **by 2019** which shows experience > site because If the pad were the main cause, success should have dropped again at CCAFS SLC-40, but it didn't.
- **Payload "sweet spot" to target: 2.5–5 t** has **the best mix of rate and volume → 12/21 successes (≈57%)**. The **7.5–10 t** band shows **60%**, but only **5 flights** (too small to rely on) very heavy payloads are rarer and riskier. Recommendation: **optimize bids in 2.5–5 t**.
- Upgrades pay off (vehicle effect): In the 2.5–5t range, older v1.1 shows more failures near the heavy end, while Block 4/5 maintain higher success. For heavier missions, assign Block 5 when possible.
- **Prediction Takeaway**: Simple features (one-hot + scaled) let four models reach 83.3% test accuracy (tie); Test set had **18 flights**. The models all got **15/18 right** → 83.3%. With so few cases, it's easy for different models to tie on accuracy. Best confusion matrix: On those 18 flights: **12 were real landings** → the model **caught all 12** as "land" → **Recall = 100%** for "land"(it never missed a true landing), **6 were real non-landings** → it only called **3** correctly as "no land" and **mistook 3** as land → **Specificity = 3/6 = 50%(**it had **3 false alarms** ("land" when it didn't)). The data is **imbalanced** (12 landings vs 6 non-landings), and the model is leaning toward predicting **"land"** to avoid missing real landings.

Appendix

- **Data & prep (quick facts):** 90+ Falcon-9 rows → one-hot + scaled features; target = **Class**; split **80/20** (seed=2), **Train = 72, Test = 18.**
- Modules and Libraries Used

1. **Data wrangling:**pandas, numpy
2. **Static charts (EDA):** matplotlib.pyplot (plt), seaborn (e.g., catplot)
3. **Interactive charts & app:** plotly.express (px), dash (and from dash import dcc, html)
4. **Maps:** folium, folium.plugins → MarkerCluster, MousePosition, folium.features → DivIcon, (and basic Folium layers like Circle, Marker, PolyLine).
5. **Machine learning (scikit-learn)**
6. **SQL:** IPython SQL magic: ipython-sql (the %%sql cells), SQLite as the engine (e.g., sqlite:///my_data1.db)

Github Link for the Notebook: https://github.com/taiwotjohn13/SPACEXPROJECT/tree/main

Thank you!