



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vijay R. Shinde
02nd August 2025



Executive Summary

- Summary of methodologies

- Data collection via API
- Data collection with Web Scraping
- Data wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Visualization
- Interactive Visual Analytics with Folium
- Interactive Dashboard with Plotly Dash
- Machine Learning Prediction

- Summary of all results
- **Temporal Success Rate:** A clear upward trend in success rate has been observed over the dataset's timeline, indicating a positive temporal evolution of launch performance.
- **Launch Site Performance:** A stratified analysis by launch site reveals that KSC LC-39A consistently exhibits the highest success rate, outperforming all other launch locations.
- **Orbital Regime Analysis:** The orbital destinations of ES-L1, SSO, HEO, and GEO are correlated with the highest overall success rates among all mission types.
- **Payload Mass Impact:** Early in the mission history, a high failure rate was observed for heavier payloads. However, this has been significantly mitigated over time, showing a strong improvement in handling higher payload mass.
- **Predictive Modeling:** The DecisionTreeClassifier algorithm proved to be a highly effective model for predictive analysis, demonstrating strong performance and accuracy in forecasting landing outcomes.

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
- Other providers cost upward of 165 million dollars each.
- Much of the savings is due to SpaceX's ability to reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch.
- Our company, IBM Developer Skills Network, can use this information to successfully bid against SpaceX for a rocket launch.

Introduction

- The business case is to use predictive modeling to determine if a first-stage rocket will land. This information can be used to determine launch costs and help IBM Developer Skills Network successfully bid against SpaceX.
- **The project will involve:**
 - Collecting and cleaning publicly available data.
 - Exploring the data with SQL and visualizations.
 - Building, evaluating, and refining predictive models to predict landing outcomes.

Problem Statement

- Predict the successful landing of SpaceX's Falcon 9 first stage, a critical factor in determining the cost-efficiency of reusable rocket launches.

Why we should solve this problem?

- Solving this problem allows companies like IBM Developer Skills Network to make competitive bids against SpaceX by accurately estimating launch costs.

Questions

- What is the historical success rate of Falcon 9 first stage landings?
- What factors most significantly influence the success or failure of a landing?
- Can we develop a predictive model that accurately forecasts the outcome of a landing?

Methodology

Executive Summary

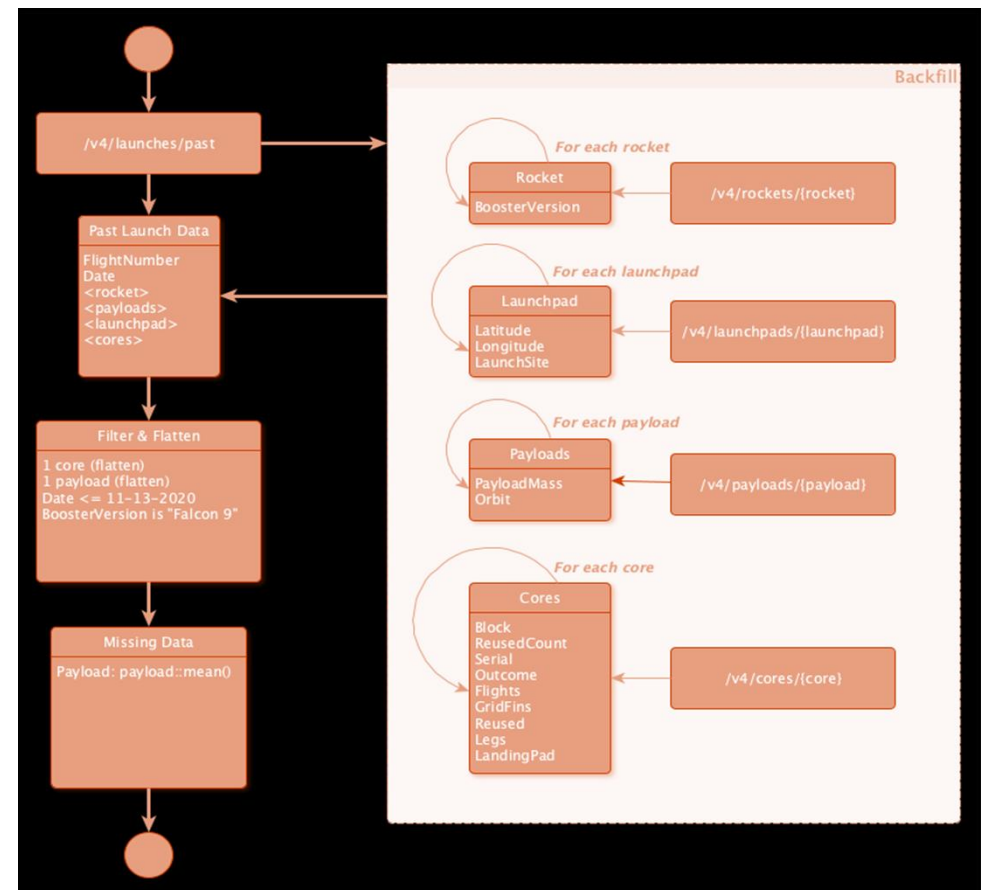
- Data collection methodology
 - Retrieval and consolidation from multiple [SpaceX API](#) endpoints
 - Web scraping tabular data from [Wikipedia](#)
- Perform data wrangling
 - Extracted relevant records
 - Flattened fields and resolved missing values
- Perform exploratory data analysis (EDA) using visualization and SQL
 - Visualize variable relationships
 - Look at the data in aggregate
- Perform interactive visual analytics using Folium and Plotly Dash
 - Mark all launch sites on a map
 - Mark successful and failed launches
 - Calculate distances to proximate locations
 - Provide for interactive exploration of the data
- Perform predictive analysis using classification models
 - Build, evaluate, and compare several predictive classification models

Data Collection

- Data was collected using a combination of retrieval techniques:
 - HTTP requests against various [SpaceX API](#) endpoints:
 - Initial launch data was obtained from: `/v4/launches/past`
 - Additional data was backfilled from:
 - `/v4/rockets`
 - `/v4/launchpads`
 - `/v4/payloads`
 - `/v4/cores`
 - Tabular data from the [List of Falcon 9 and Falcon Heavy launches](#) Wikipedia page.

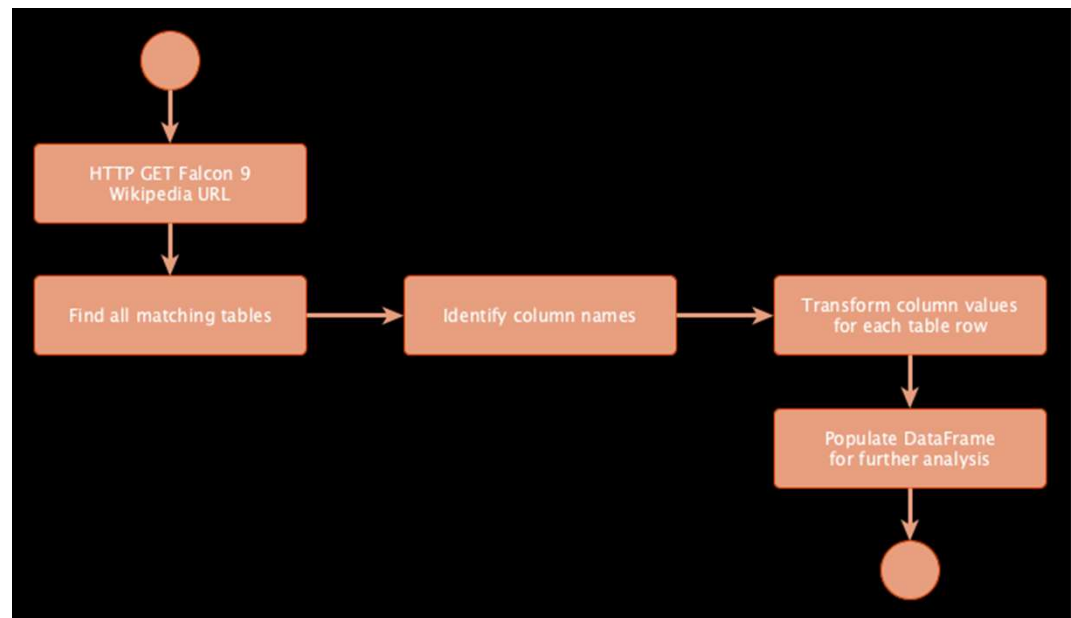
Data Collection – SpaceX API

- The initial data was obtained from the `/v4/launches/past` API endpoint.
- Additional data was backfilled from the `rocket`, `launchpad`, `payloads`, and `cores` API endpoints, for records with extant corresponding IDs.



Data Collection - Scraping

- Web scraping workflow:
 - Send HTTP Get to Falcon 9 Launch page at Wikipedia, to retrieve the HTML source
 - Create a BeautifulSoup object to extract the tables containing launch data
 - Populate a Pandas DataFrame using the extracted columns



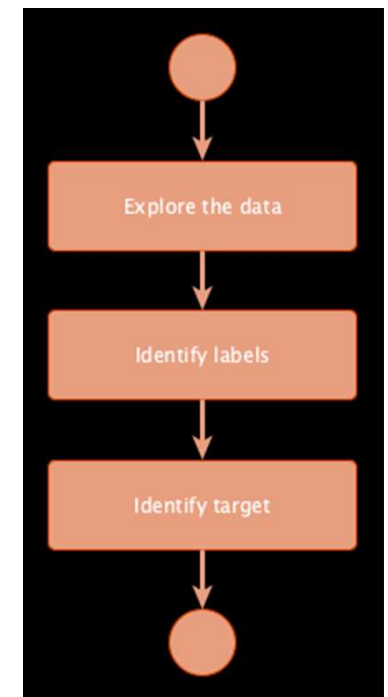
Data Wrangling

Goals:

- Use Exploratory Data Analysis (EDA) to find patterns in the data.
- Determine the labels for training supervised models.

Steps:

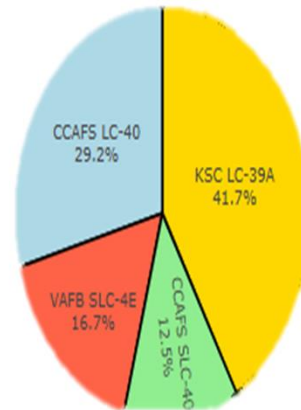
- Find missing values as a percentage of each attribute
- Identify column types, numerical or categorical
- Show launches per Launch Site
- Show distribution of Orbit type in the data set
- Explore outcomes, and group them by binary outcome (success or failure)
- Store outcome as "Class" label, to be used as target value in training



EDA with Data Visualization

- Visualize relationships to gain insight into the importance of each variable:

- Flight Number and Outcome
- Flight Number and Launch Site
- Payload and Launch Site
- Orbit and Outcome
- Flight Number and Orbit
- Payload and Orbit
- Yearly success rate



- Expand categorical variables into "dummy" columns
- Convert numerical columns into `float64`

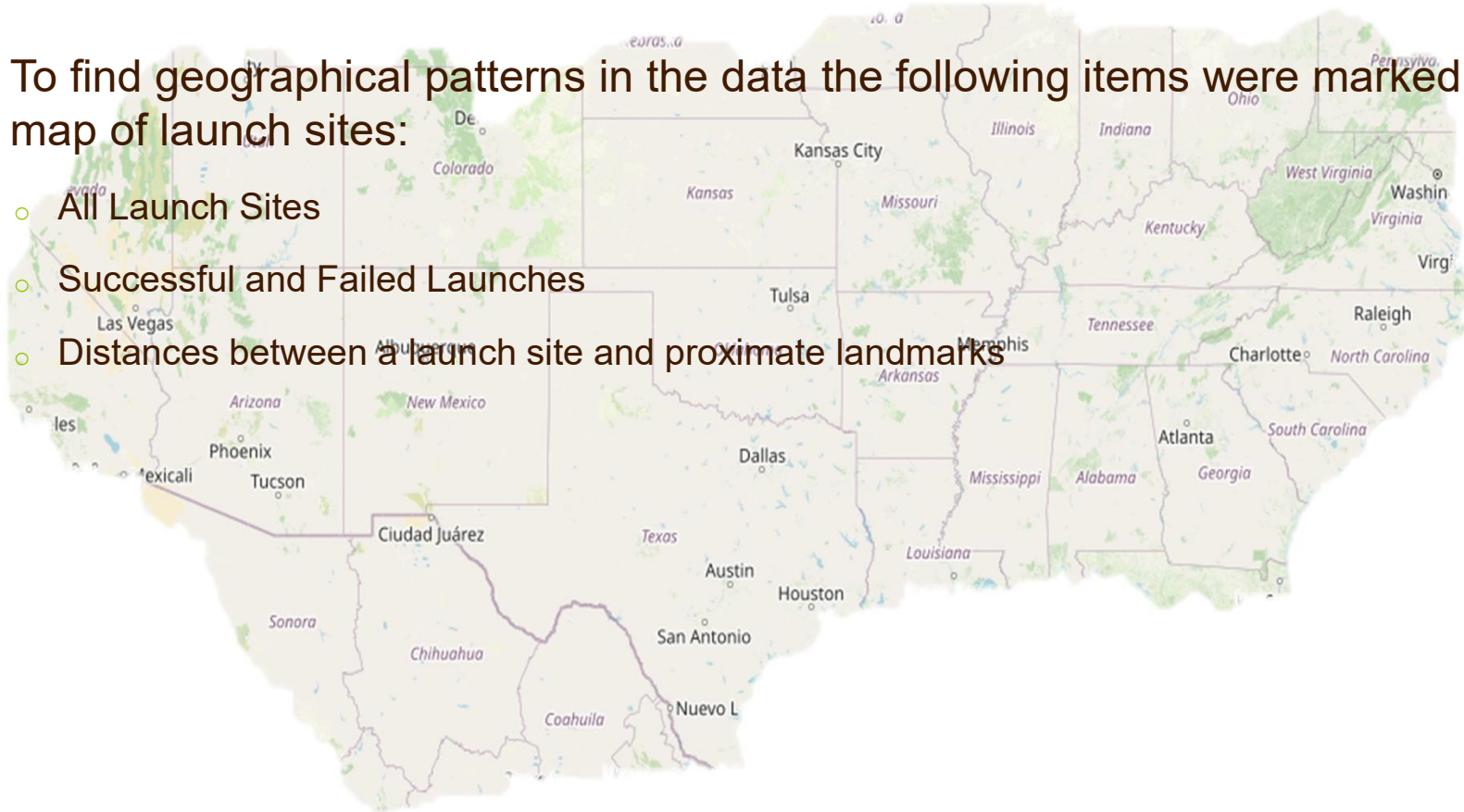
EDA with SQL

- SQL Queries Performed:

- Show each unique launch site
- Show 5 records where launch site names begin with 'CCA'
- Display the total payload mass carried by boosters launched by 'NASA (CRS)'
- Display the average payload mass carried by the v1.1 Falcon 9 booster
- List the date of the first successful ground landing outcome
- List the booster versions with successful outcomes landing on the drone ship with payloads between 4000kg and 6000kg.
- List the number of successful and failed mission outcomes
- List all of the booster versions that carried the max payload mass
- List the month name, outcome, booster version, and launch site for missions with failure outcomes landing on a drone ship in 2015.
- Show the distribution of outcomes between June 4th, 2010 and March 20th, 2017

Build an Interactive Map with Folium

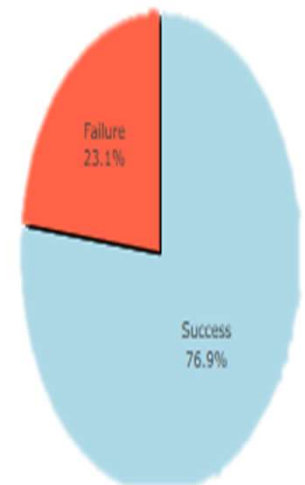
- To find geographical patterns in the data the following items were marked on a map of launch sites:
 - All Launch Sites
 - Successful and Failed Launches
 - Distances between a launch site and proximate landmarks



Build a Dashboard with Plotly Dash

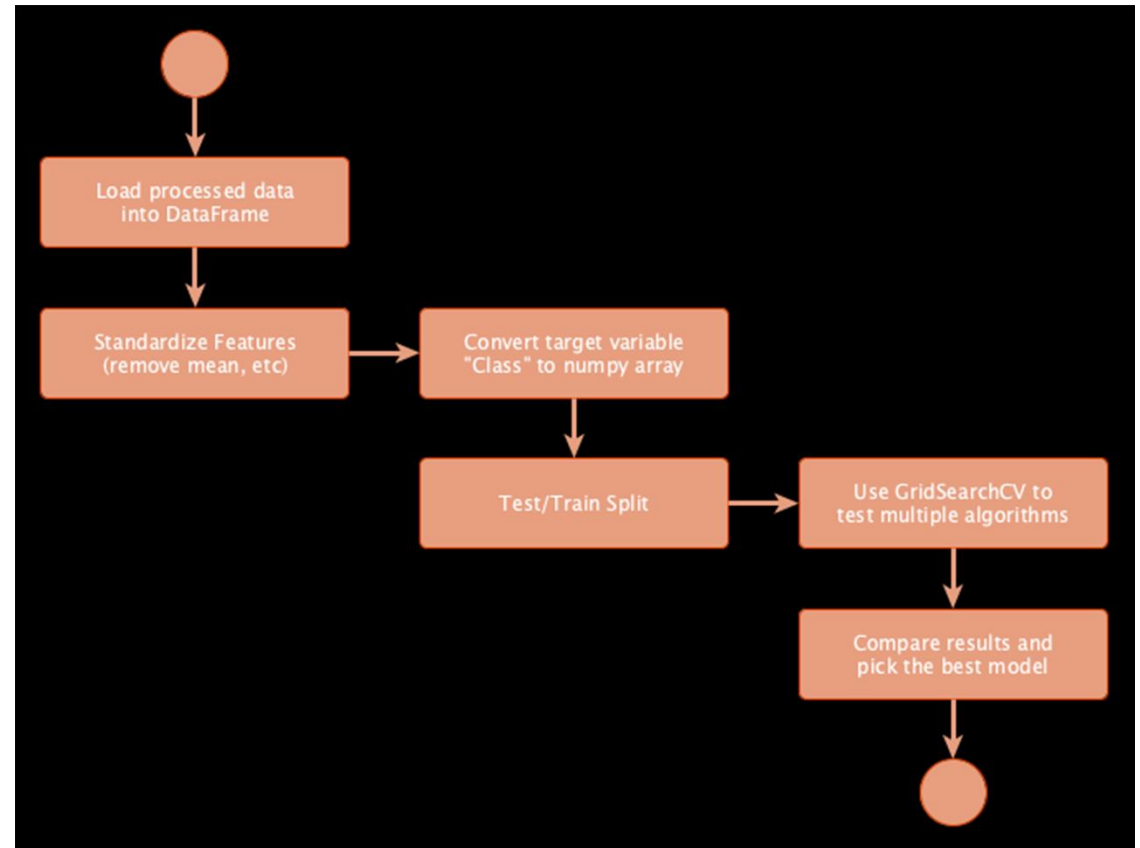
To enable interactive exploration of the data, a Plotly Dash dashboard was developed to include:

- A dropdown selector to choose a Launch Site, affecting:
 - A pie chart
 - All sites selected: shows the breakdown of successful outcomes across all sites
 - A launch site selected: shows the breakdown of successful vs failed launches for the given site
 - A scatter plot
 - All sites selected: shows outcome by payload mass and booster version for all sites
 - A launch site selected: shows outcome by payload mass and booster version for the given site
- A Payload Mass range selector that filters data points on the scatter plot



Predictive Analysis (Classification)

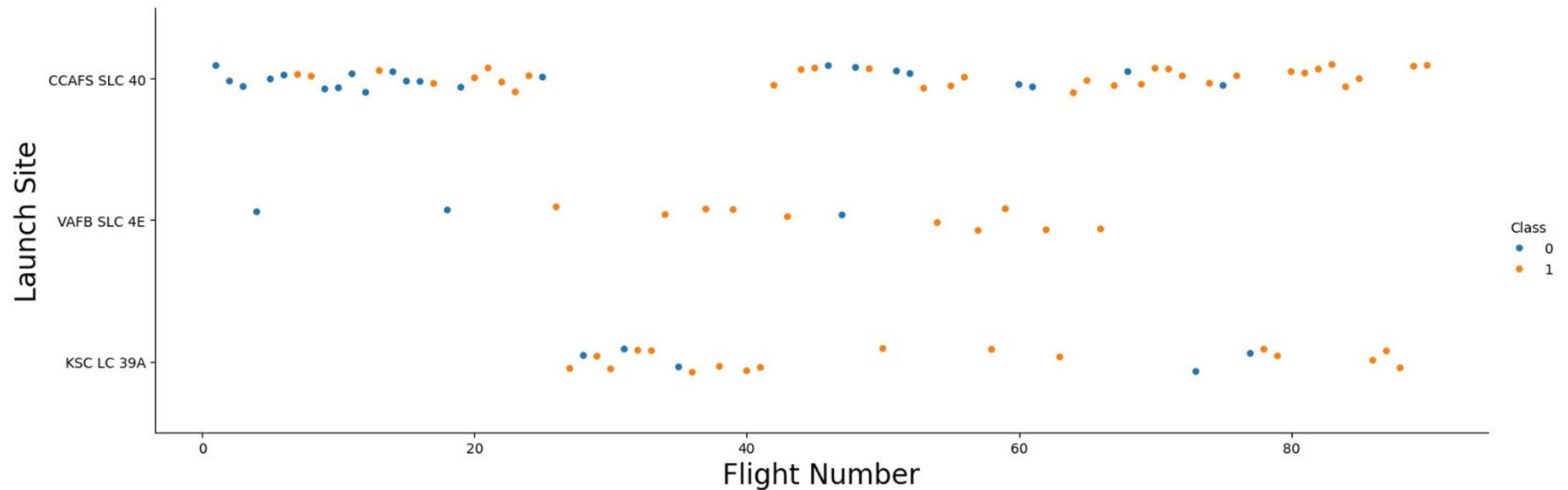
- Load data
- Apply StandardizedScaler on X
- Convert Y to numpy array
- Split training and testing data
- Use GridSearchCV to test
- hyperparameters for multiple algorithms:
 - Logistic Regression
 - SVC
 - Decision Tree Classifier
 - K Neighbors Classifier



Results

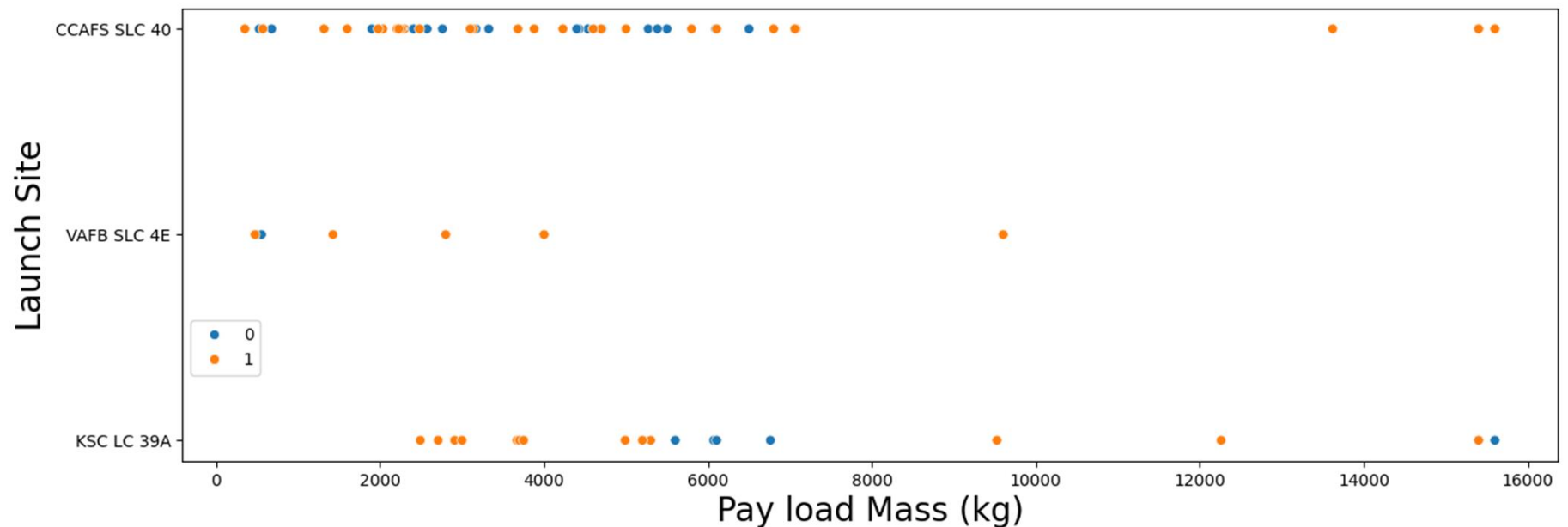
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Flight Number vs. Launch Site



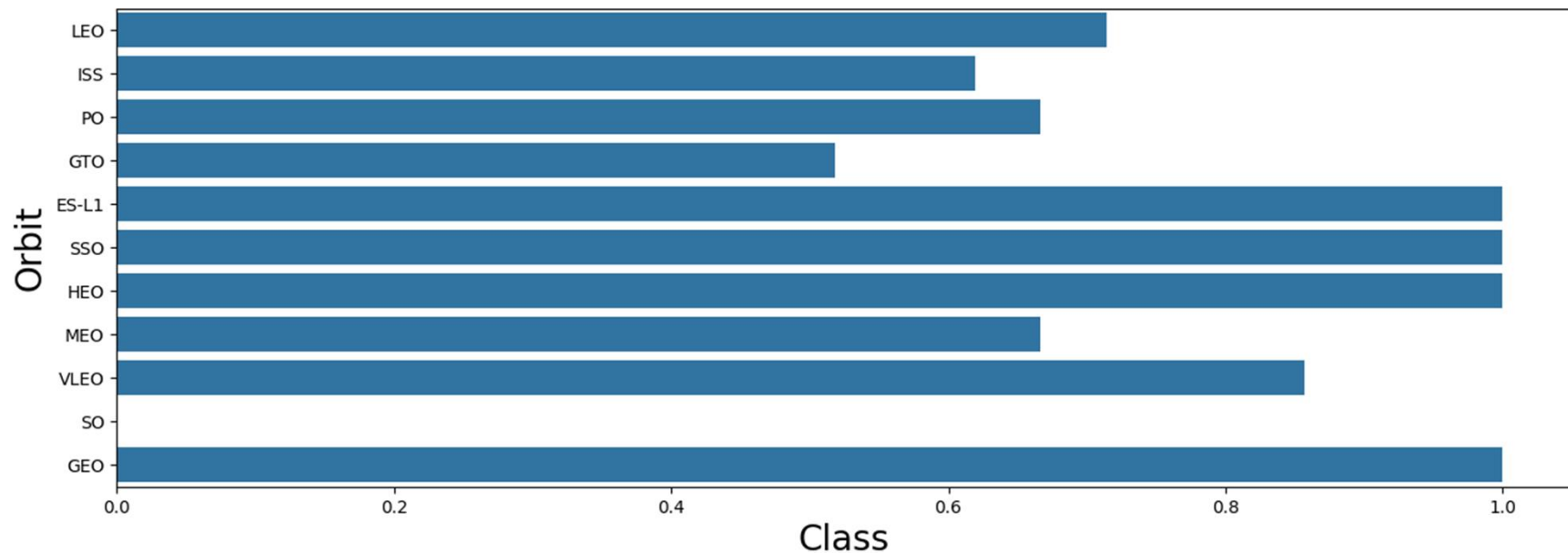
- All sites show a mix of first stage landing successes and failures, with successes increasing over time.
- Early flights predominantly resulted in failures, indicating improvements to technology or process.
- While CCAFS SLC 40 has the most total flights, VAFB SLC 4E appears to have a relatively higher proportion of successful landing outcomes.

Payload vs. Launch Site



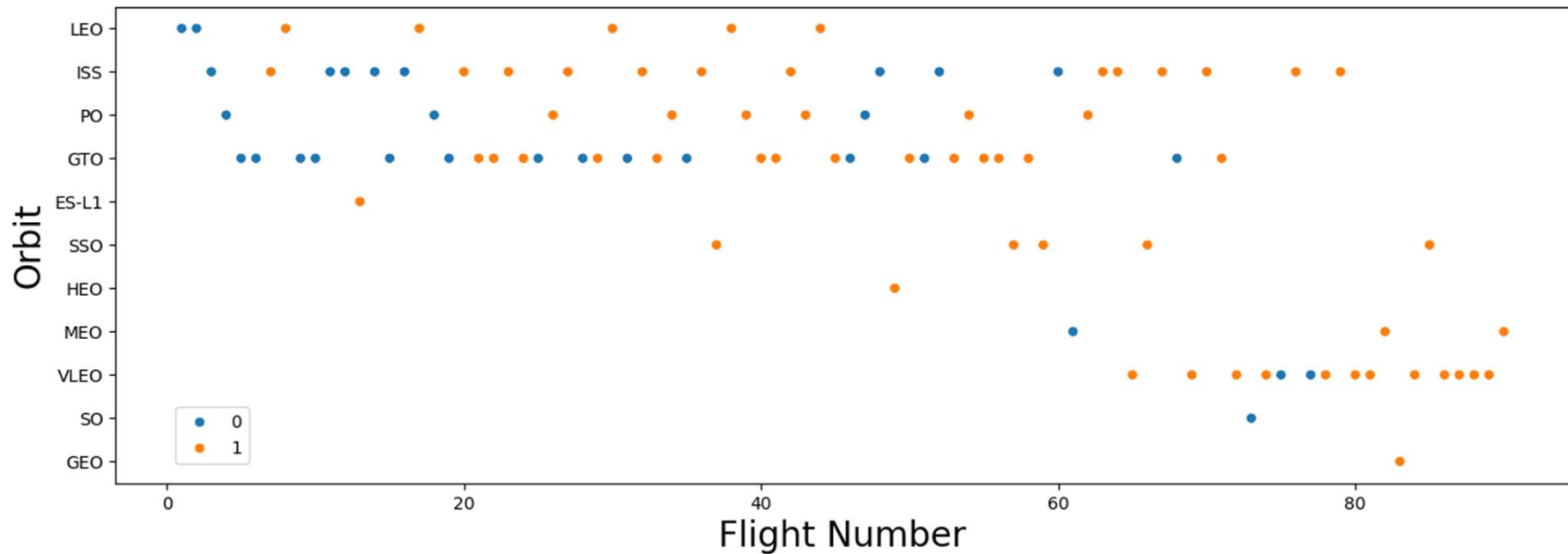
- All sites show a wide range of payload weights, from light to heavy payloads.
- Early flights trend toward lighter payloads, representing the bulk of the landing failures.
- This suggests technological or operational improvements lead to a greater rate of success with heavier payloads.

Success Rate vs. Orbit Type



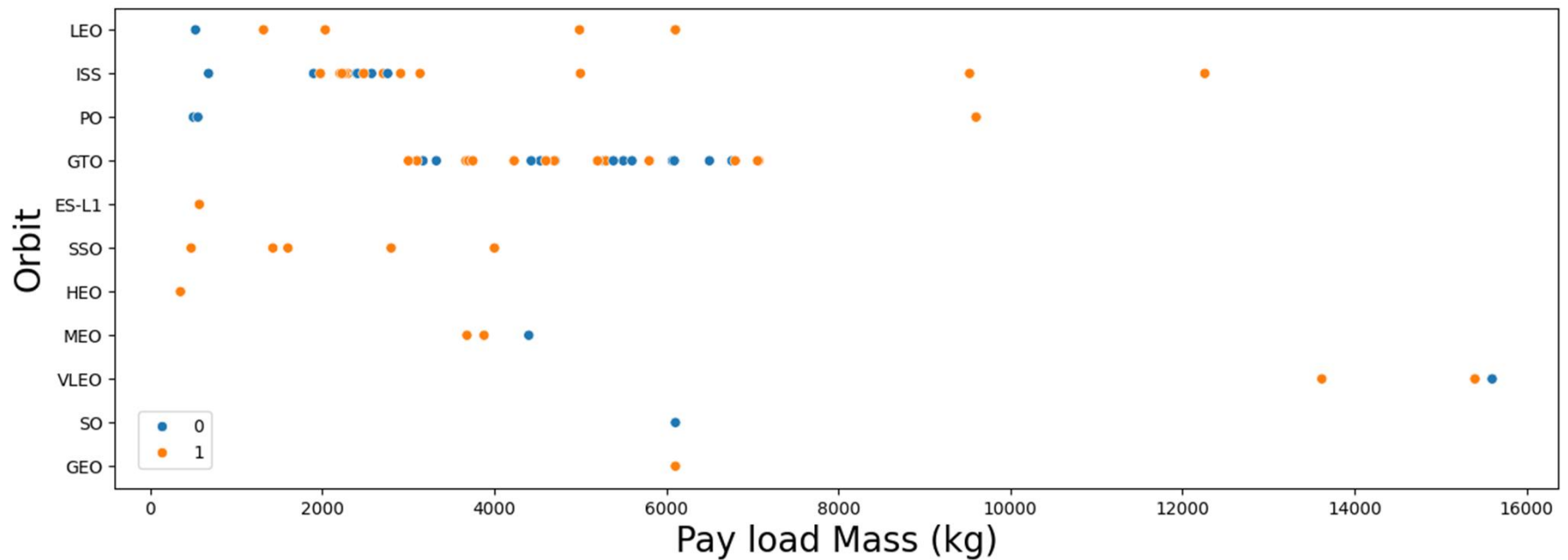
- Some orbits, such as ES-L1, SSO, HEO, and GEO consistently show high success rates.
- Others such as GTO show more mixed outcomes, suggesting some orbit types may introduce operational or technological challenges.
- With only one launch, there is not enough data for the SO orbit type to provide an accurate analysis.

Flight Number vs. Orbit Type



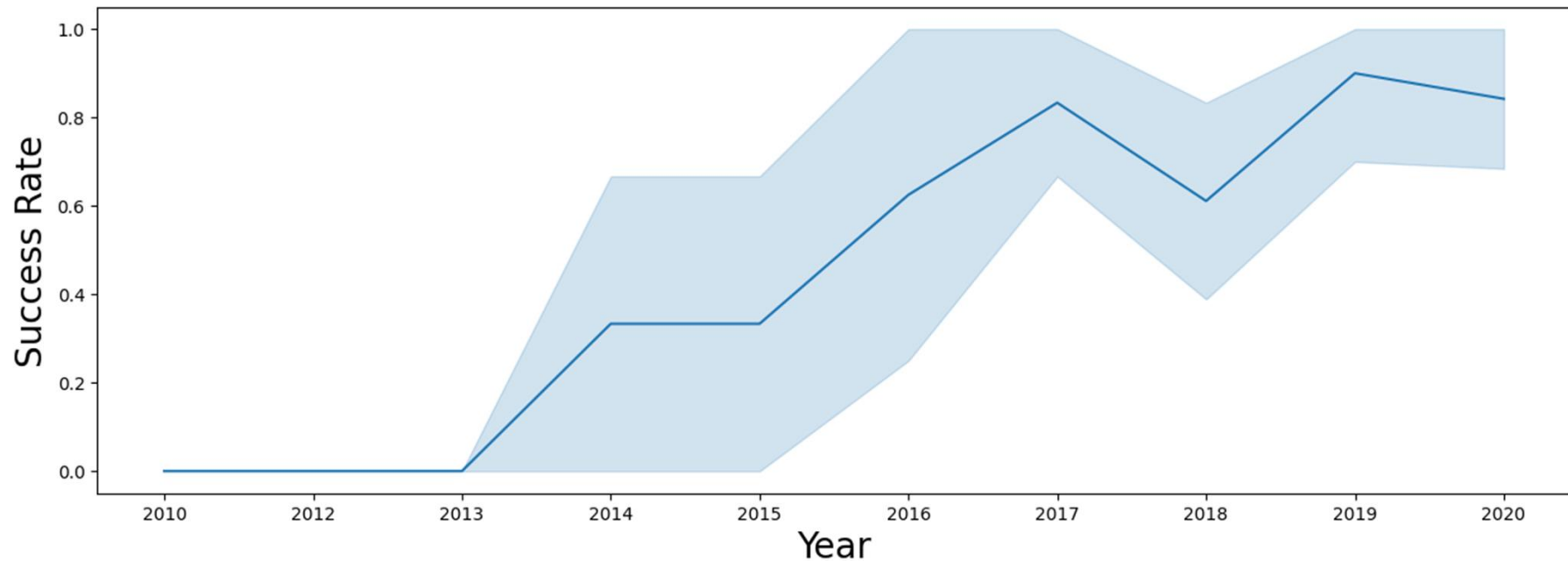
- Many orbits are represented throughout the flight number range, some orbits are not attempted until later flights.
- There is a noticeable improvement in landing success as flight numbers increase, indicating an accumulation of experience and ongoing improvements.

Payload vs. Orbit Type



- Many orbits are represented across a wide range of payload masses, but others like SSO, MEO, HEO and GEO show a generally lower range.
- Orbits with a constrained payload range, tend to show a higher rate of landing success.
- While payload mass does not appear to directly determine mission success, its interplay with orbit suggests a significant correlation.

Launch Success Yearly Trend



- Yearly trend shows a consistent progression from early challenges to high reliability in first stage landings over time.
- From 2016 onward, SpaceX experienced year over year improvement in success rate with a minor setback in 2018.

All Launch Site Names

There are four unique Launch Sites

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

```
SELECT DISTINCT Launch_Site from SPACEXTABLE;
```

Launch Site Names Begin with 'CCA'

- First five records where launch sites begin with 'CCA'

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

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

Total Payload Mass

The total payload carried by boosters from NASA (CRS) is **45,596kg**.

```
SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD  
FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';
```


Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is **2,534.67kg**.

```
SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS  
FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%';
```

First Successful Ground Landing Date

The first successful landing outcome on ground pad occurred on **December 22nd, 2015**.

```
SELECT MIN(Date) as LaunchDate
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (ground pad)';
```

Successful Drone Ship Landing with Payload between 4000 and 6000

The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:

Booster	Payload Mass
F9 FT B1022	4,696kg
F9 FT B1026	4,600kg
F9 FT B1021.2	5,300kg
F9 FT B1031.2	5,200kg

```
SELECT Booster_Version, PAYLOAD_MASS__KG_  
FROM SPACEXTABLE  
WHERE Landing_Outcome = 'Success (drone ship) '  
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

Mission Status	Count
Failure	1
Success	100

```
SELECT CASE  
WHEN Mission_Outcome LIKE 'Success%' THEN 'Success'  
END as Mission_Status, COUNT(*)  
FROM SPACEXTABLE  
GROUP BY Mission_Status;
```

Boosters Carried Maximum Payload

- The maximum payload sent was **15,600kg**.
- The boosters that carried the maximum payload are:

```
SELECT
  DISTINCT Booster_Version,
  PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
  SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
)
ORDER BY Booster_Version;
```

Booster Version

F9 B5 B1048.4

F9 B5 B1048.5

F9 B5 B1049.4

F9 B5 B1049.5

F9 B5 B1049.7

F9 B5 B1051.3

F9 B5 B1051.4

F9 B5 B1051.6

F9 B5 B1056.4

F9 B5 B1058.3

F9 B5 B1060.2

F9 B5 B1060.3

2015 Launch Records

- List of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

Month	Outcome	Booster	Launch Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Month	Failure (drone ship)	Booster_Version	Launch_Site	Date
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS	2015-01-01
February				2015-02-01
March				2015-03-01
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS	2015-04-01
May				2015-05-01
June				2015-06-01
July				2015-07-01
August				2015-08-01
September				2015-09-01
October				2015-10-01
November				2015-11-01
December				2015-12-01

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

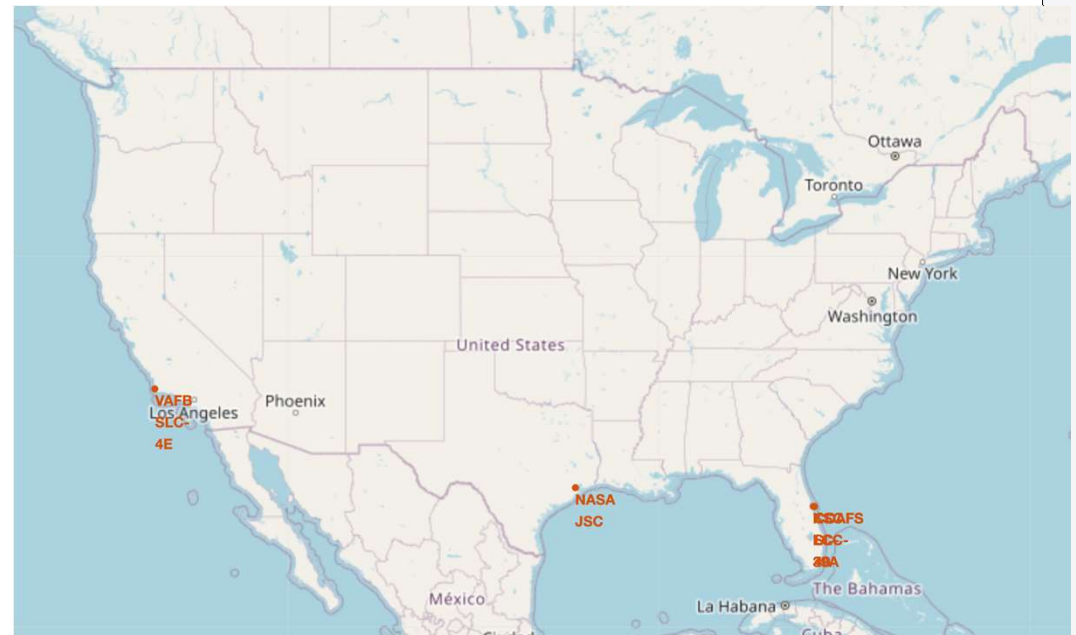
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order:

```
SELECT Outcome, Count
FROM Outcome
WHERE Outcome BETWEEN '2010-06-04' AND '2017-03-20'
ORDER BY Count DESC;
```

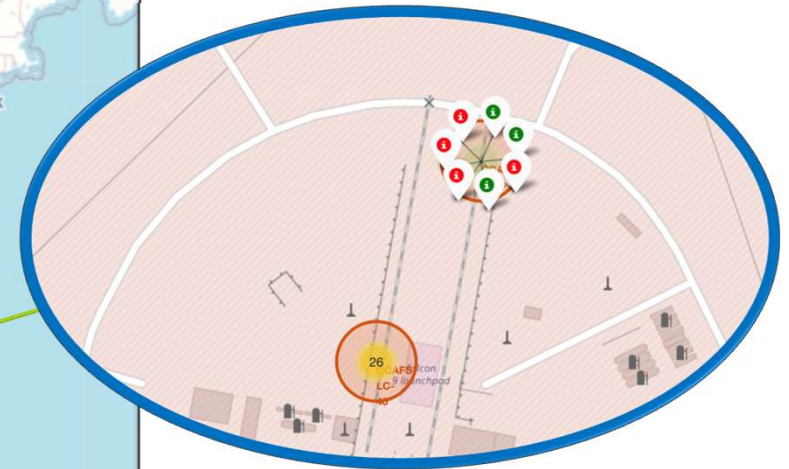
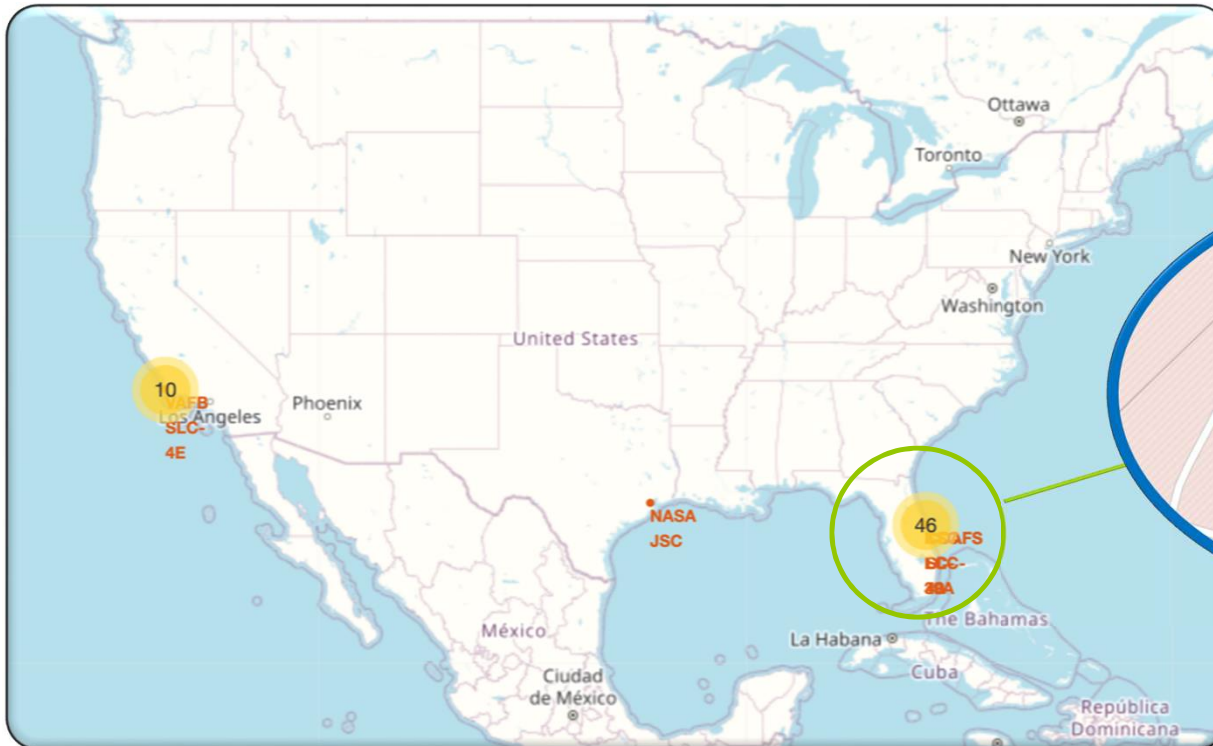
Landing Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Launch Site Locations

- Launch sites are located near coastal regions in Florida and California to reduce risk of catastrophic failures affecting human activities.



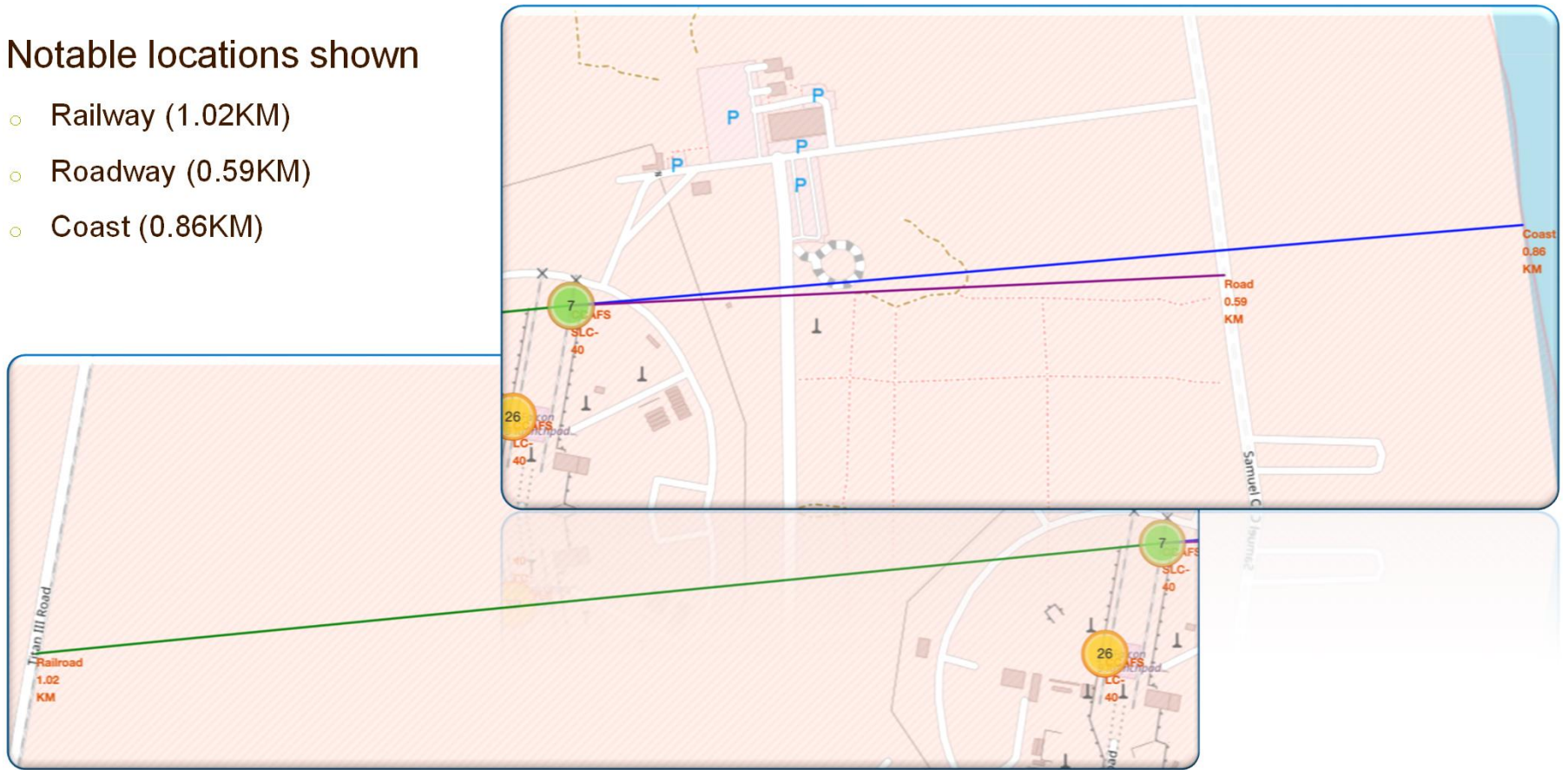
Launch Outcomes



Notable Proximate Locations

- Notable locations shown

- Railway (1.02KM)
- Roadway (0.59KM)
- Coast (0.86KM)



All Launch Sites: Successful Landings

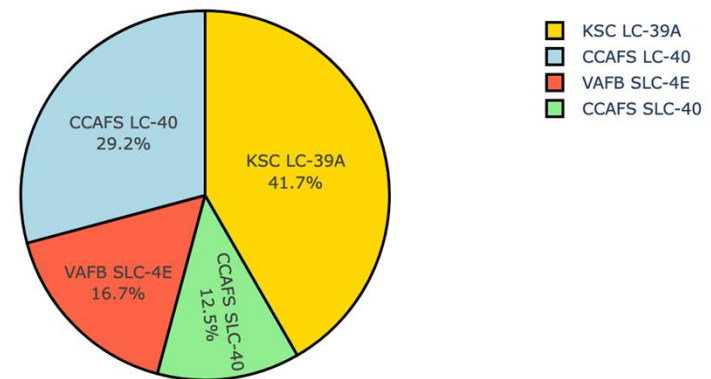
- KSC LC-39A experienced the highest proportion of successful landings, followed by CCAFS LC-40.
- VAFB SLC-4E and CCAFS SLC-40 the lowest.

SpaceX Launch Records Dashboard

All Sites



Total Successful Launches by Site



36

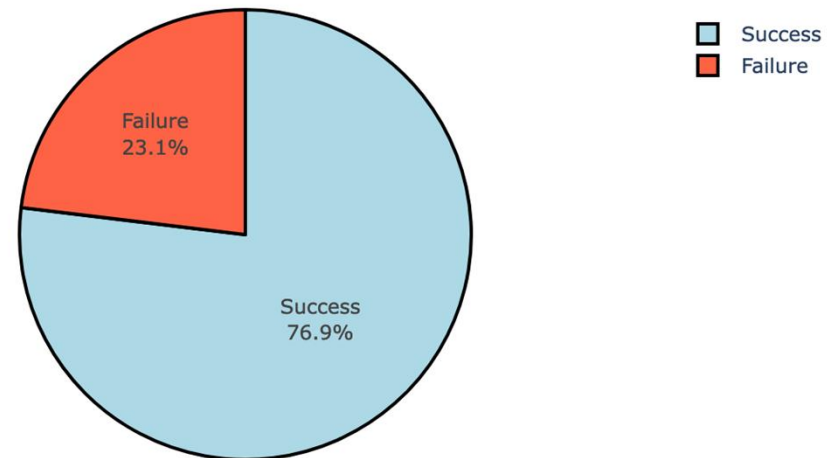
Per-site Launch Success Ratio: High

- KSC LC-39A had the highest ratio of successful landings

KSC LC-39A

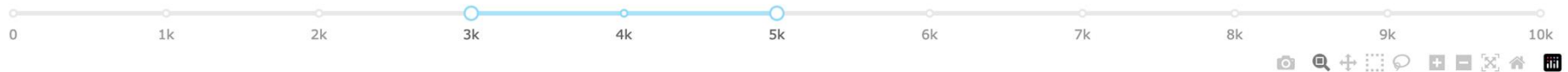


Launch Success vs Failure for site KSC LC-39A

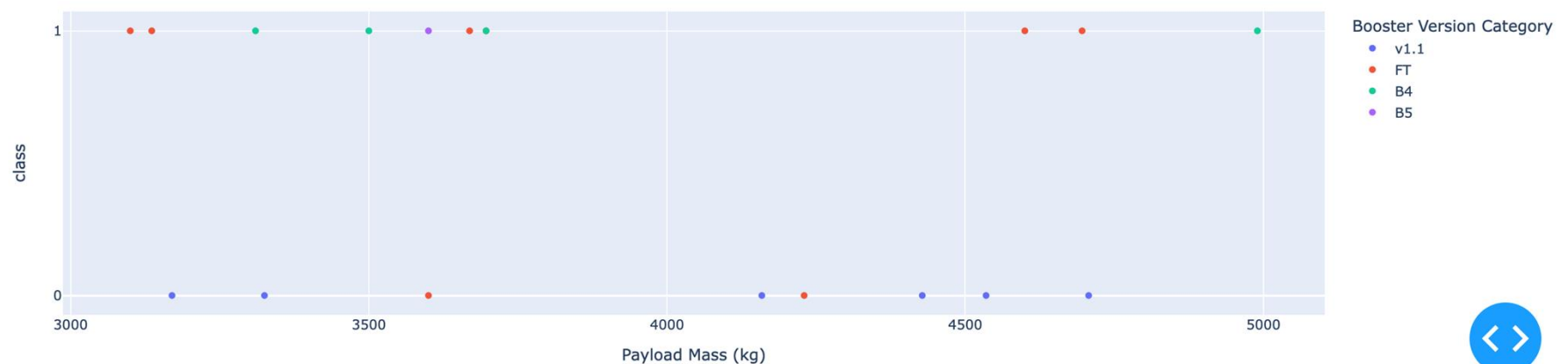


Payload Range

Payload range (Kg):

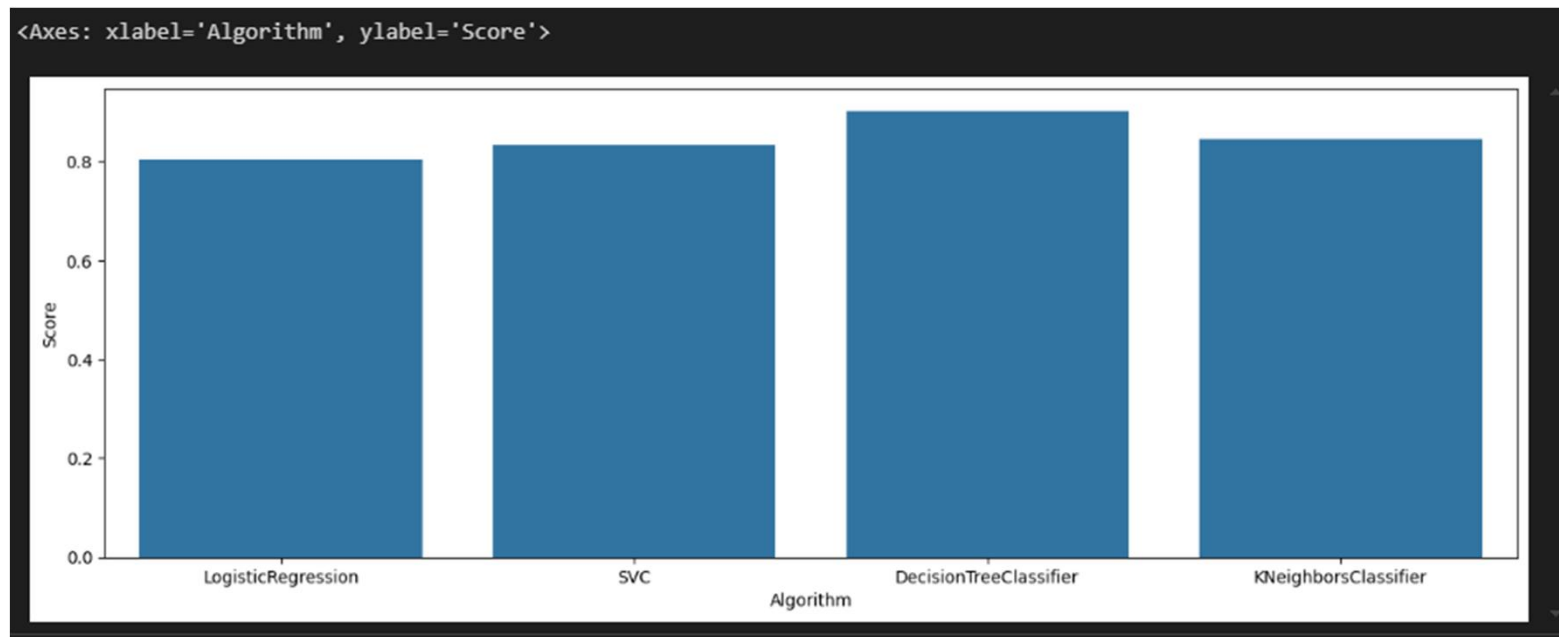


Payload Success Rate for All Sites



- With a payload mass between 3,000kg and 5,000kg, v1.1 boosters performed the worst.
- In the same payload range, B4 and B5 boosters had the best success rate, followed by FT.

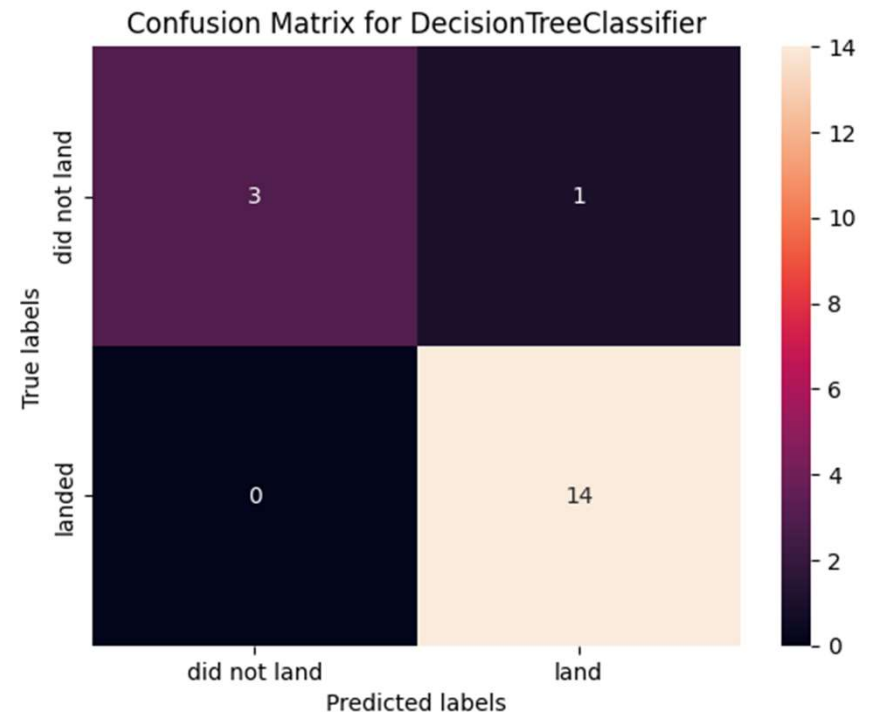
Classification Accuracy



- Of the algorithms tested, the DecisionTreeClassifier was the most accurate.

Confusion Matrix

- 14 observations were correctly predicted as successful landings (true positive)
- 3 observations were correctly predicted as failed landings (true negative)
- 1 observation was incorrectly predicted as a successful landing (false positive)
- No observations were incorrectly predicted as a failed landing (false negative)



Conclusions

- **Temporal Improvements:** A clear upward trend in success rates was observed over time, indicating continuous operational and technological enhancements.
- **Performance by Orbit:** A stratified analysis identified ES-L1, SSO, HEO, and GEO as the orbital regimes with the highest success rates.
- **Launch Site Efficacy:** Launch site was a key predictive feature, with KSC LC-39A and CCAFS LC-40 consistently performing at the highest levels.
- **Model Performance:** The DecisionTreeClassifier was the most effective predictive model evaluated, demonstrating superior performance across accuracy, precision, and recall metrics for forecasting landing outcomes.

Appendix

● Data Sources

- SpaceX API
 - Collected Data: dataset_part_1.csv
- Wikipedia: List of Falcon 9 and Falcon Heavy launches (June 2021)
- Data after wrangling: dataset_part_2.csv
- Geographical data: spacex_launch_geo.csv
- Interactive data source: spacex_launch_dash.csv



Thank You!