SPACE X PROJECT MISSION

SHAULA M. MARQUEZ

DECEMBER 25, 2024

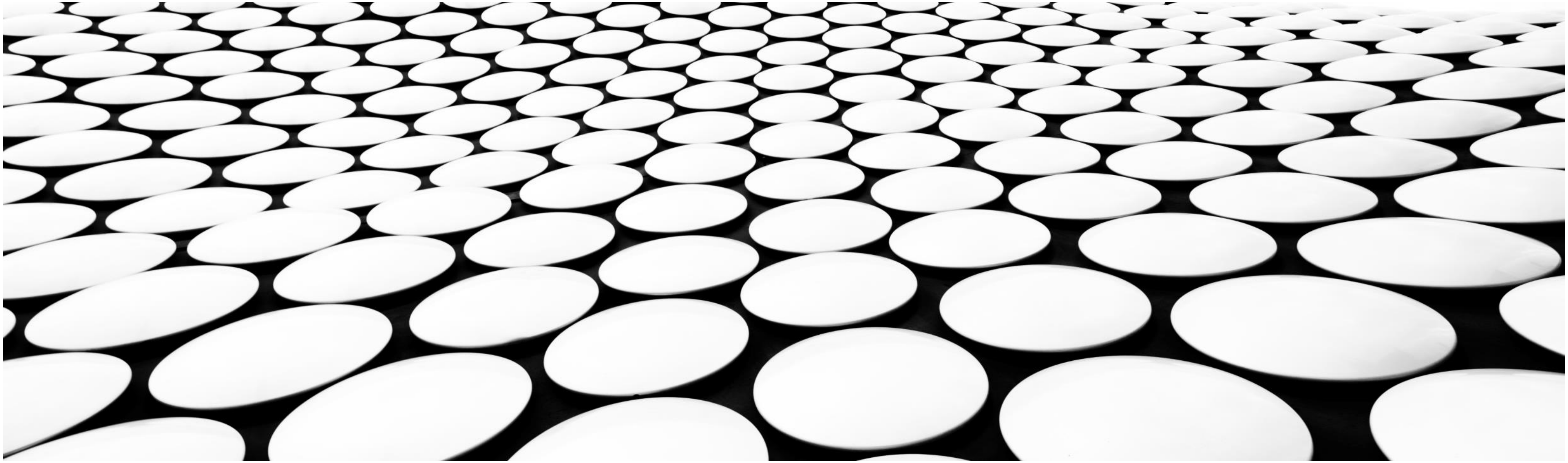


TABLE OF CONTEXT

EXECUTIVE SUMMARY

- Space Y would like to compete with SpaceX founded by Billionaire industrialist Allon Musk.
- 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 because SpaceX can reuse the first stage.
- This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

INTRODUCTION

- This is an Applied Data Science Capstone Project by IBM in Coursera.
- In this capstone, we will predict if the Falcon 9 first stage will land successfully.
- 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 because SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

METHODOLOGY

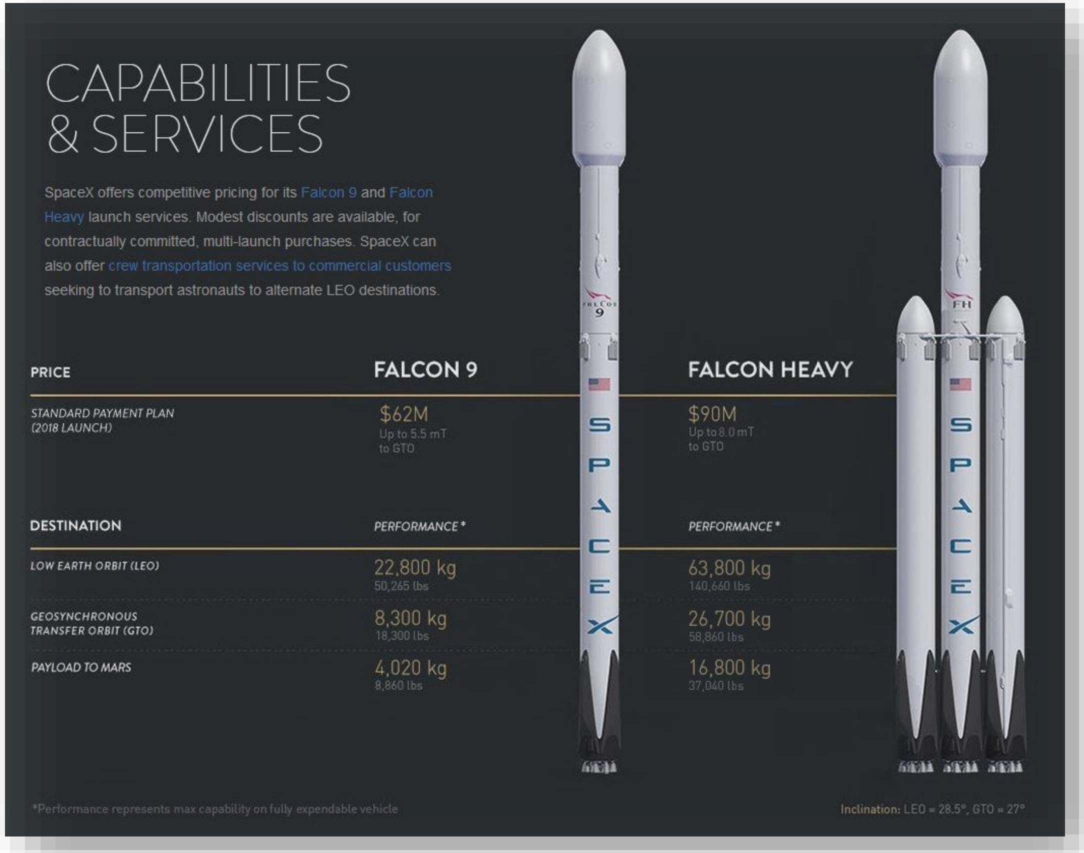
Data Collection

API

- Request and parse the SpaceX launch data using the GET request
- Filter the data frame to only include Falcon 9 launches
- Dealing with Missing Values by replacing it with mean values for PayloadMass

Web Scraping

- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from HTML table header
- Create a data frame by parsing the launch HTML tables



METHODOLOGY

Data Wrangling

- **Calculate the number of launches**
  - The data contains several SpaceX: Cape Canaveral Space Launch Complex 40 **VAFB SLC 4E**, Vandenberg Air Force Base Space Launch Complex **(SLC-4E)**, Kennedy Space Center Launch Complex 39A **KSC LC 39A.**
- **Calculate the number and occurrence of each orbit**
  - Determine the number and occurrence of each orbit in the column **Orbit.**
- **Calculate the number and occurrence of mission outcome of the orbits**
  - Determine the number of landing outcomes.
- **Create a landing outcome label from outcome column**
  - This variable will represent the classification variable that represents the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed Successfully

Number of Landing outcomes

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: count, dtype: int64
```

METHODOLOGY

Exploratory Data Analysis (SQL)

Performed SQL queries:

- Display the names of the unique launch sites in the space mission
- Display records where launch sites begin with the string ‘CCA’
- Display the total payload mass caried by boosters launched by NASA (CRS)
- Display average payload mass carried booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad achieved
- List the name of the boosters which have a success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass.
- 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.
- 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

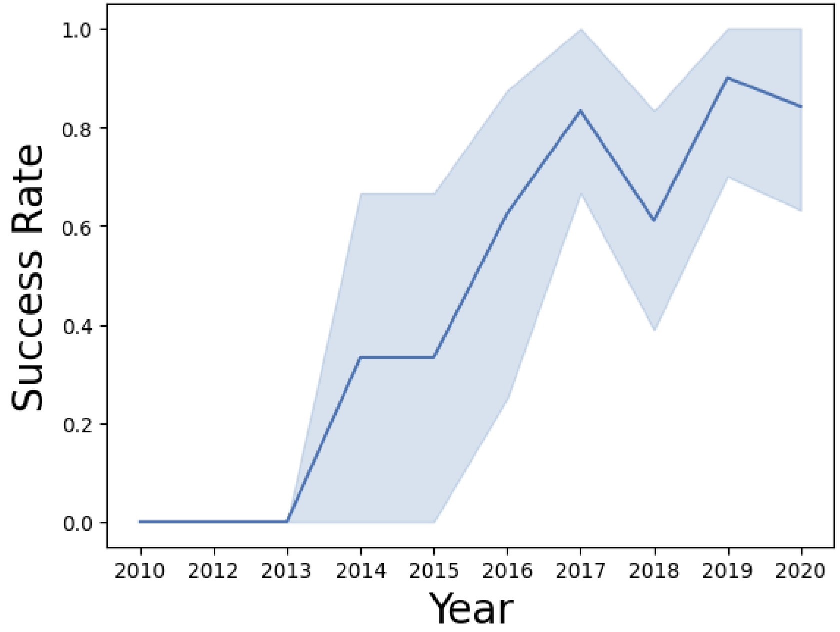
Landing_Outcome	TOTAL_LANDING_OUTCOMES
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

METHODOLOGY

Exploratory Data Analysis (Visualization Lab)

Performed Exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib

- Visualized the relationship between Flight Number and Launch Site
- Visualized the relationship between Payload and Launch Site
- Visualized the relationship between success rate of each orbit type
- Visualized the relationship between Flight Number and Orbit type
- Visualized the relationship between Payload and Orbit type
- Visualized the launch success yearly trend
- Created dummy variables to categorical columns
- Casted all numeric columns to `float64`



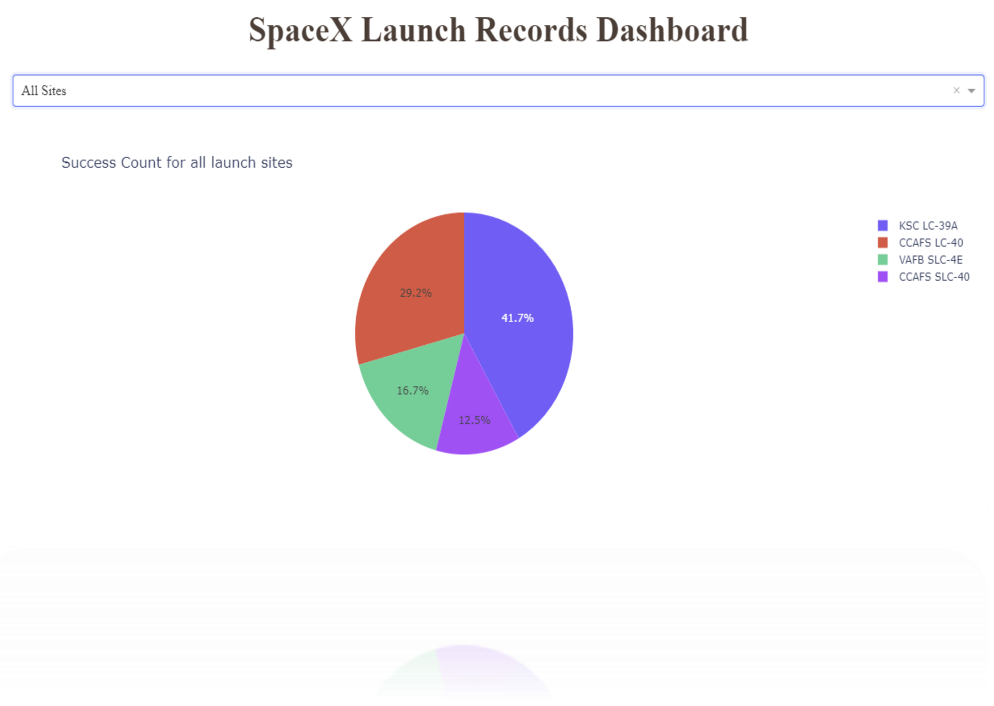
METHODOLOGY

Interactive Dashboard with Ploty Dash

Built a Plotly Dash application for users to perform interactive visual analytics on SpaceX launch data in real-time.

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart. You will be guided to build this dashboard application via the following tasks:

- Added a Launch Site Drop-down Input Component
- Added a callback function to render success-pie-chart based on selected site dropdown
- Added a Range Slider to Select Payload
- Added a callback function to render the success-payload-scatter-chart scatter plot

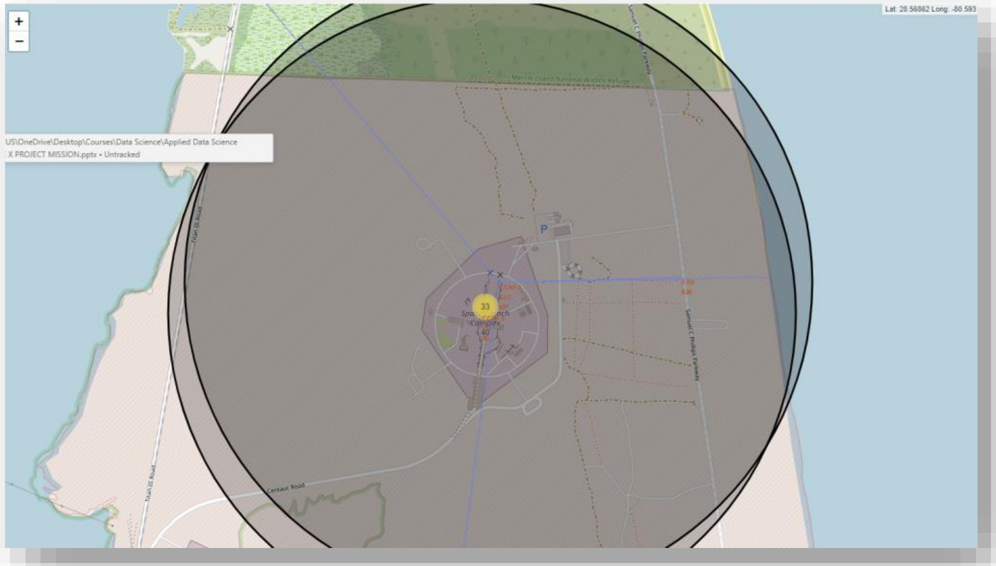


METHODOLOGY

Interactive Visual Analytics with Folium

The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

- Marked all launch sites on a map
- Marked the success/failed launches for each site on the map
- Calculated the distances between a launch site to its proximities

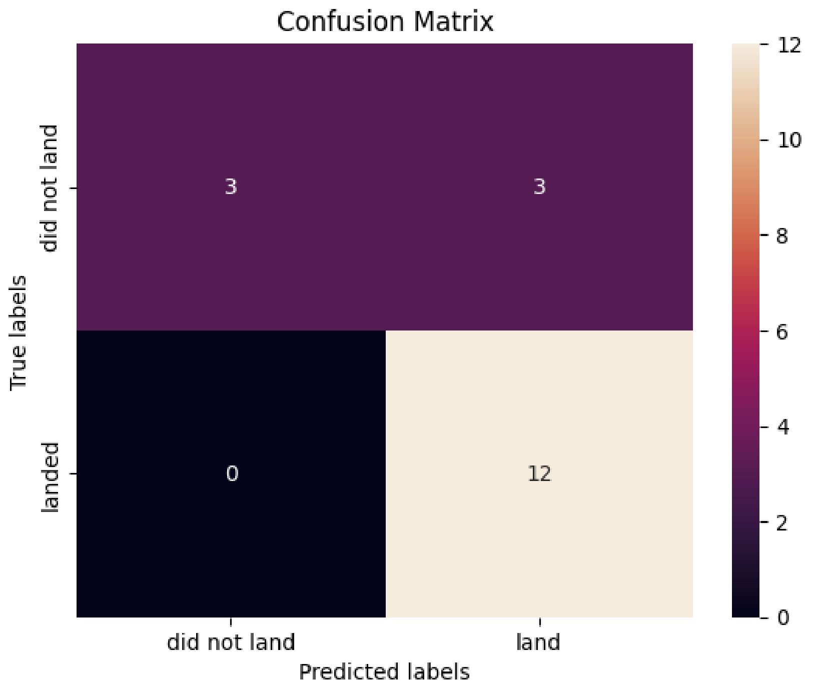


METHODOLOGY

Predictive Analysis

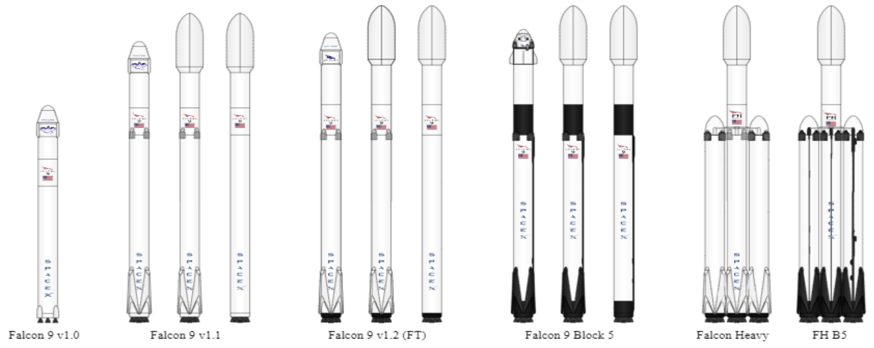
Find the best method performs best using test data Hyperparameter for SVM, Classification Trees and Logistic Regression.

- Imported the following libraries
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn
  - Sklearn
- Loaded the data to the data frame
- Created a NumPy array from the column in by applying the method to_numpy() then assign it to the variable, make sure the output is a Pandas series.
- Standardized the data x then reassign it to the variable x using the transform provided below.
  - We split the data into training and testing data using the function.
- Used the function train_test_split the data X and Y into training and test data.
- Fit the training data to the following models:
  - logistic regression
  - Support Vector Machine
  - Decision Tree Classifier
  - K Nearest Neighbors Classifier



RESULTS

- Exploratory Data Analysis
- Interactive Analytics Demo in Screenshots
- Predictive Analysis Results

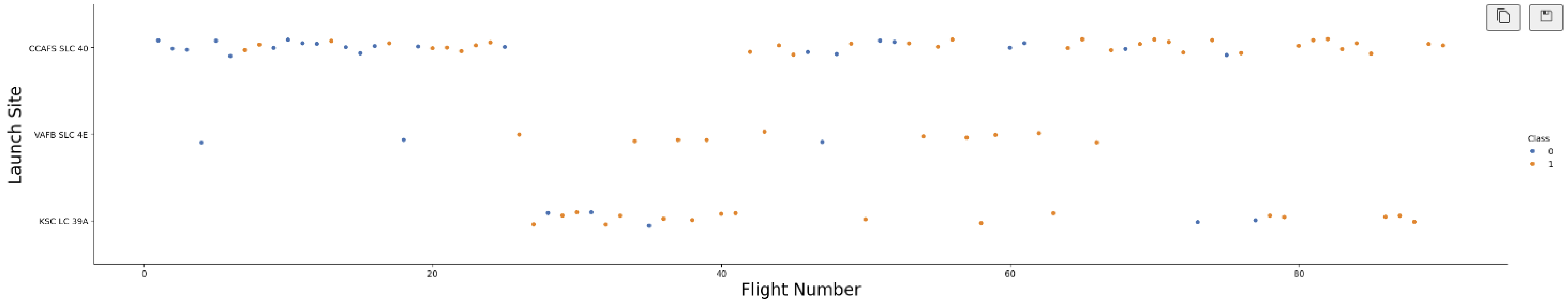


EXPLORATORY DATA ANALYSIS WITH VISUALIZATION

RESULTS

Flight Number VS Launch Site

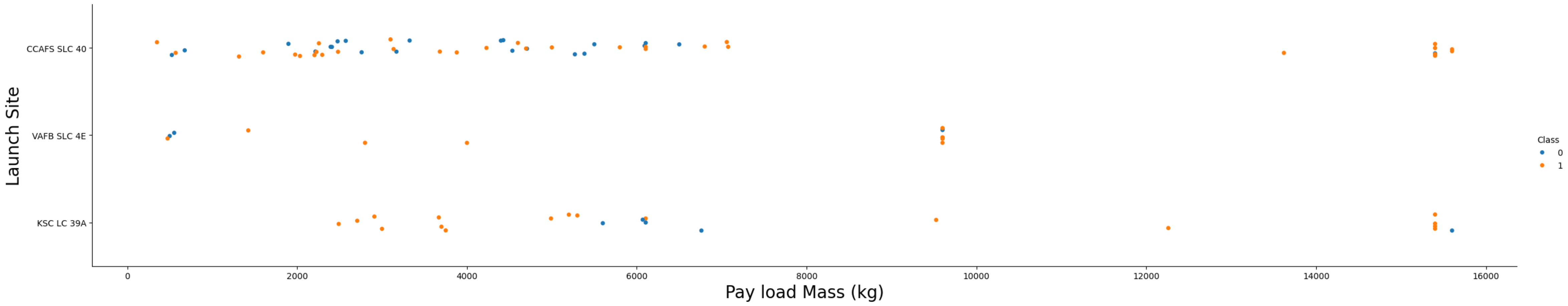
- Early flights tend to have failure while later flights tend to succeed
- Success seems to be consistent across multiple Launch Sites in later stages, reflecting overall improvement in operations and technology.
- The results show a positive trend in flight success over time. This indicates learning and improvement, a hallmark of iterative advancements in aerospace engineering and operational processes.
- **Launch Site Variability:** All launch sites show progress over time, though some (e.g., Kennedy Space Center) may have faced more initial challenges compared to others.



RESULTS

Pay load Mass VS Launch Site

- **Success Rates Improve for Heavier Payloads:** The likelihood of successful launches increases with payload mass, especially at CCAFS SLC 40 and KSC LC 39-A.
- **Launch Site-Specific Trends:** Some sites (e.g., CCAFS SLC 40) manage more diverse payloads, while others (e.g., VAFB SLC 4E) appear to handle lighter payloads with mixed results.
- **Room for Improvement:** The failures for lighter payloads might reflect early-stage missions or experimental technology.

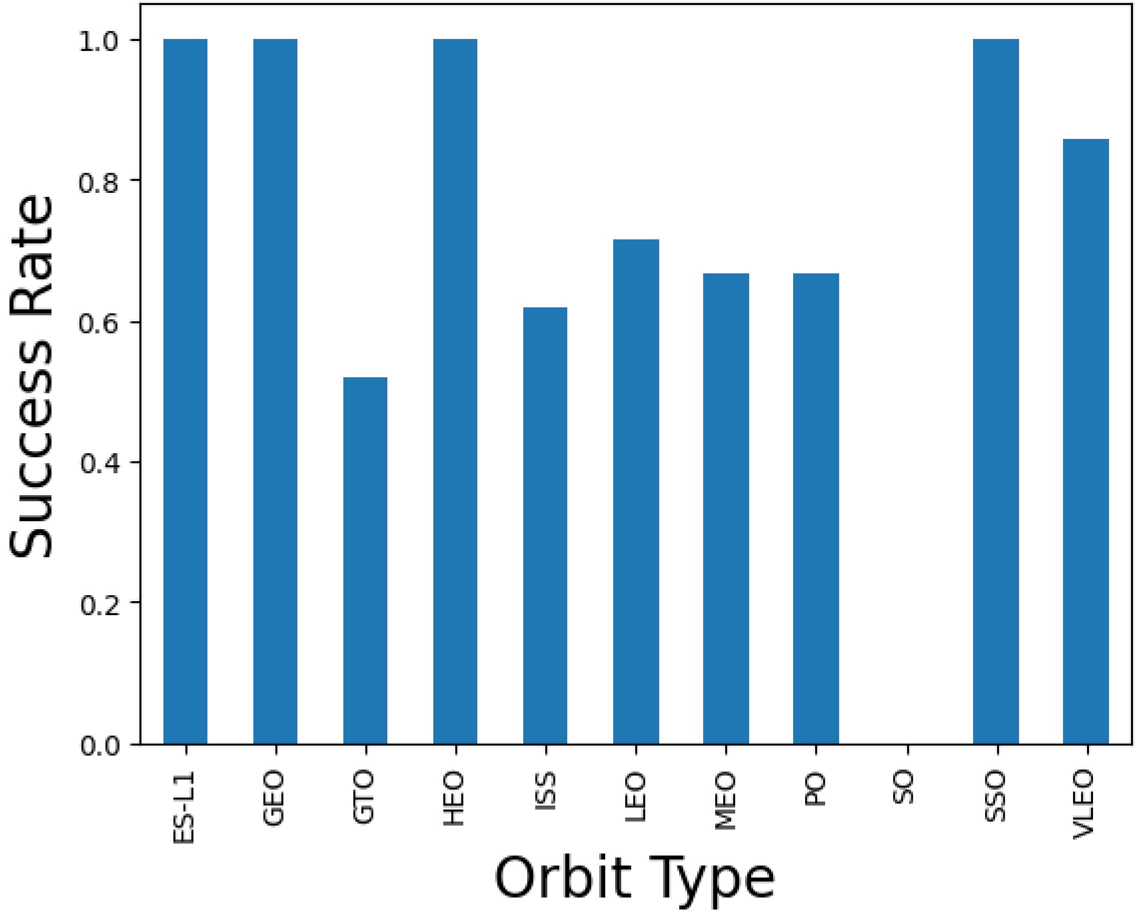


RESULTS

- **High Success Rates (1.0):** Orbits like ES-L1, GEO, LEO, and SSO have achieved a 100% success rate, indicating that all missions targeting these orbit types were successful.
- **Moderate to High Success Rates:** VLEO (Very Low Earth Orbit) and PO (Polar Orbit) have a very high success rate close to 1.0 but slightly below.
- **Moderate Success Rates:** HEO (Highly Elliptical Orbit) and ISS (International Space Station orbit) have a success rate slightly below 0.8, indicating that a significant majority of missions were successful.
- **Low Success Rate:** GTO (Geostationary Transfer Orbit) stands out with a success rate below 0.6, suggesting challenges in achieving mission success for this orbit type.

Insights:

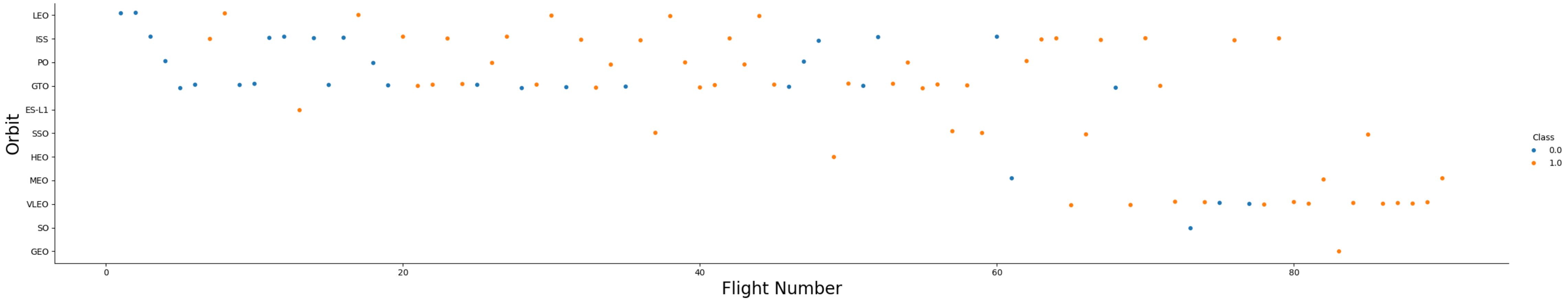
- This chart could be used to assess the reliability and feasibility of space missions depending on the orbit type. The significantly lower success rate for GTO might indicate technical or operational challenges specific to this orbit.



RESULTS

Flight Number VS Orbit

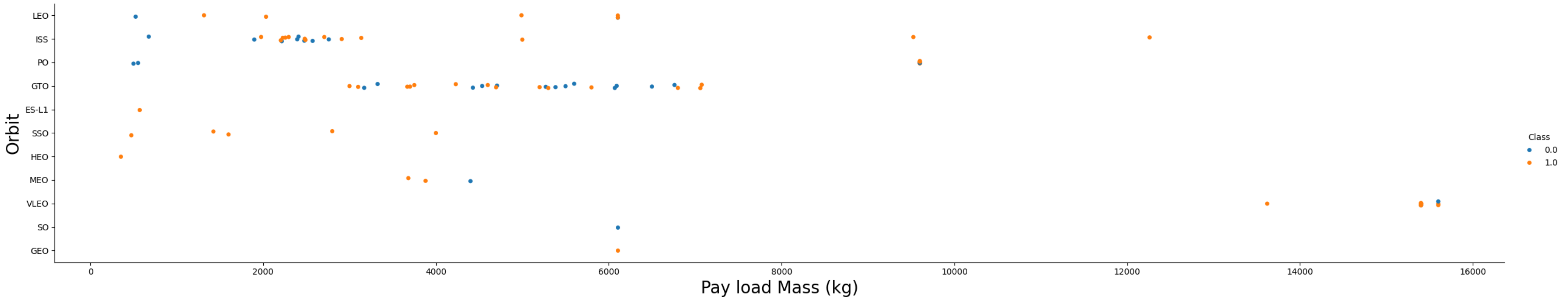
- As the flight number increases, there appears to be a higher proportion of orange dots (successes), suggesting improved mission success rates with experience or advancements in technology.
- Early flight numbers (lower x-axis values) have more blue dots, particularly in orbits like GTO, indicating early challenges.
- LEO (Low Earth Orbit): Consistently has a high number of successes (orange dots), indicating reliability in this orbit type. GTO (Geostationary Transfer Orbit): Shows a mix of failures and successes, with some improvement in success rates as flight numbers increase. SSO (Sun-Synchronous Orbit) and GEO (Geostationary Orbit): Fewer data points are available, but they tend to show success when attempted. PO (Polar Orbit) and ISS (International Space Station): Predominantly successful, especially for later flight numbers.
- The data suggests a learning curve, where success rates improve as more flights are conducted (possibly due to better engineering, processes, or experience).



RESULTS

Orbit VS Payload Mass

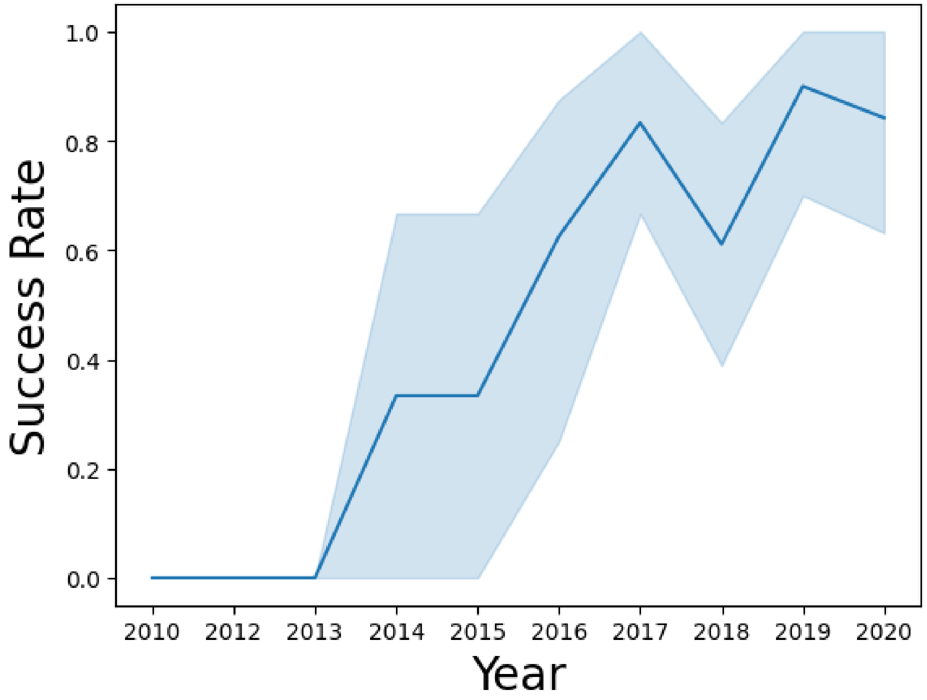
- LEO (Low Earth Orbit): Can handle a wide range of payload masses (from light to very heavy) with a high success rate.
- ISS (International Space Station orbit): Payloads targeting the ISS show high success, even for heavier payloads.
- GTO (Geostationary Transfer Orbit): Has a mix of successes and failures across various payload masses. Notably, there are failures even at moderate payloads (2,000–6,000 kg).
- SSO (Sun-Synchronous Orbit): Handles relatively lighter payloads, with a mix of successes and failures.
- ES-L1 and HEO: Show fewer data points but indicate some challenges with success rates at varying payloads.
- Heavier Payloads (>10,000 kg): Limited attempts (few dots), but some successful missions are visible, suggesting that heavier payloads are more challenging but achievable.
- Payload Mass Impact: Heavier payloads increase the risk of mission failure, likely due to the complexity of launch dynamics.
- Orbit Sensitivity: Certain orbits like LEO and ISS show robust performance across payload masses, while GTO appears more sensitive to payload mass.



RESULTS

Orbit VS Payload Mass

- From 2013 the Success Rate increased until 2018 it decreased. Then it increased from 2019 onwards.



EXPLORATORY DATA ANALYSIS WITH SQL

RESULTS

Task 1

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

```
%sql select distinct launch_site from SPACEXTABLE;
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

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

RESULTS

Task 2

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

```
%sql select * from SPACEXTABLE where launch_site like 'CCA%' limit 5;
```

✓ 0.0s

Python

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

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

RESULTS

Task 3

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer='NASA (CRS)';
```

✓ 0.0s

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

sum(PAYLOAD_MASS__KG_)
45596

RESULTS

Task 4

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.1%';
```

✓ 0.0s

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

RESULTS

Task 5

List the date when the first successful landing outcome in ground pad was achieved. average payload mass carried by booster version F9 v1.1

```
%sql SELECT min(date) from SPACEXTABLE where landing_outcome like 'Success (ground pad)';
```

✓ 0.0s

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

min(date)
2015-12-22

RESULTS

Task 6

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

```
%%sql SELECT Booster_Version
    FROM SPACEXTABLE
    WHERE Landing_Outcome LIKE 'Success (drone ship)'
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

✓ 0.0s

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

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

RESULTS

Task 7

List the total number of successful and failure mission outcomes

```
%%sql
SELECT TRIM(MISSION_OUTCOME) AS MISSION_OUTCOME, COUNT(*) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY TRIM(MISSION_OUTCOME);
```

✓ 0.0s

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

MISSION_OUTCOME	TOTAL_NUMBER
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

RESULTS

<u>Task 8</u>

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql SELECT BOOSTER_VERSION
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ =
    (SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL)
```

✓ 0.0s

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

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

RESULTS

Task 9

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.

```
%%sql SELECT substr(Date, 6,2) AS Month,
             substr(Date, 1,4) AS Year,
             BOOSTER_VERSION,
             LAUNCH_SITE,
             LANDING_OUTCOME
      FROM SPACEXTABLE
      WHERE LANDING_OUTCOME LIKE 'Failure (drone ship)'
      AND substr(Date, 1,4) = '2015';
```

✓ 0.0s

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

Month	Year	Booster_Version	Launch_Site	Landing_Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

RESULTS

Task 10

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

```
%%sql
SELECT LANDING_OUTCOME,
        COUNT(LANDING_OUTCOME) AS TOTAL_LANDING_OUTCOMES
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_LANDING_OUTCOMES DESC;
```

✓ 0.0s

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

Landing_Outcome	TOTAL_LANDING_OUTCOMES
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

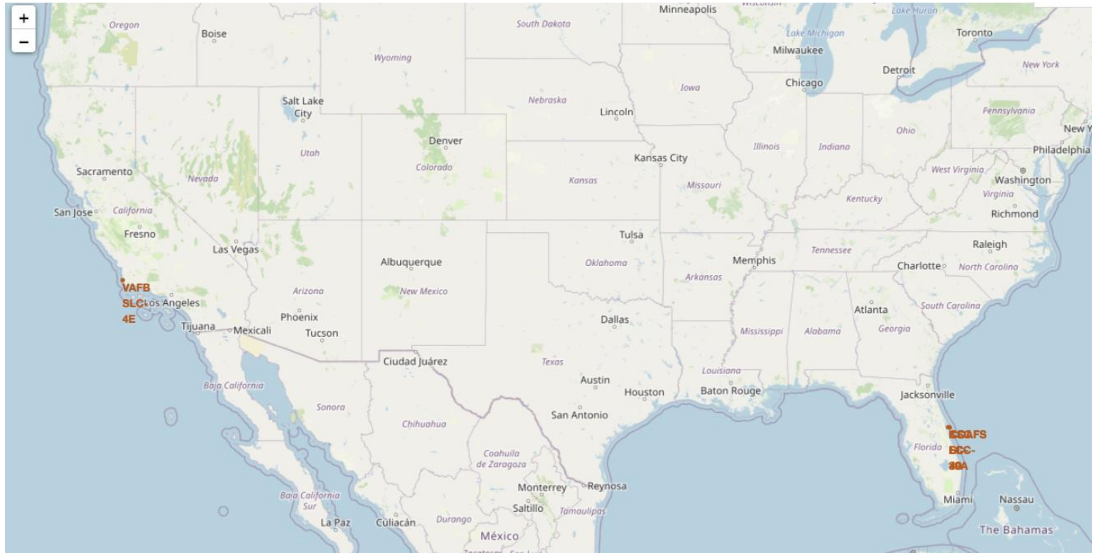
INTERACTIVE MAP WITH FOLIUM

RESULTS

The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.

This lab contains the following tasks:

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities



RESULTS

We could see the successful rate of launch sites based on color markers on the map.

Green Marker – Successful

Red Marker – Failed Launch

KSC LC-39A has a successful launch rate.

RESULTS

Visualizing the railway, highway, coastline, and city proximities for each launch site allows us to see how close each is, for example:

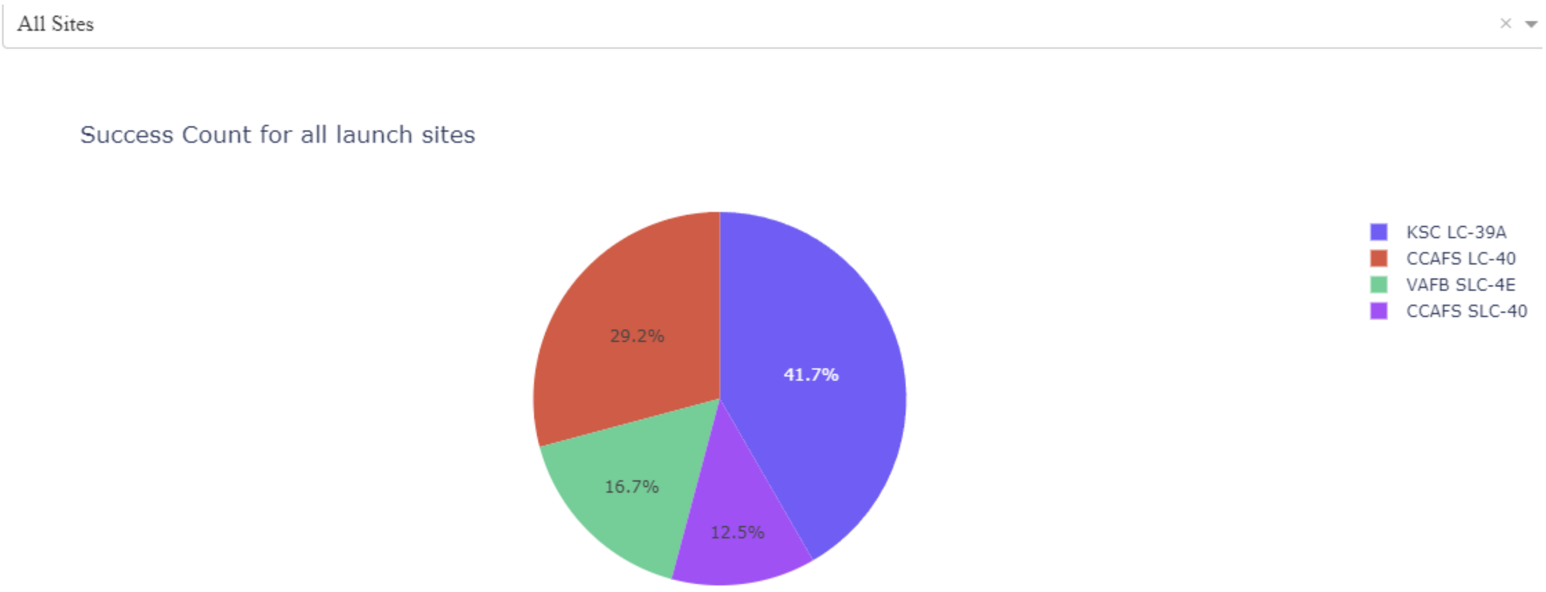
Proximities for CCAFS SLC 40:
railway: 1.28 km
highway: 0.58 km
coastline: 0.86 km
city: 51.43 km

BUILD A DASHBOARD WITH PLOTLY DASH

RESULTS

Launch success count for all sites

The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches.



RESULTS

Launch success count for all sites

Scatter plot showing Payload Mass VS Landing Outcome that are color coded by Booster Versions.

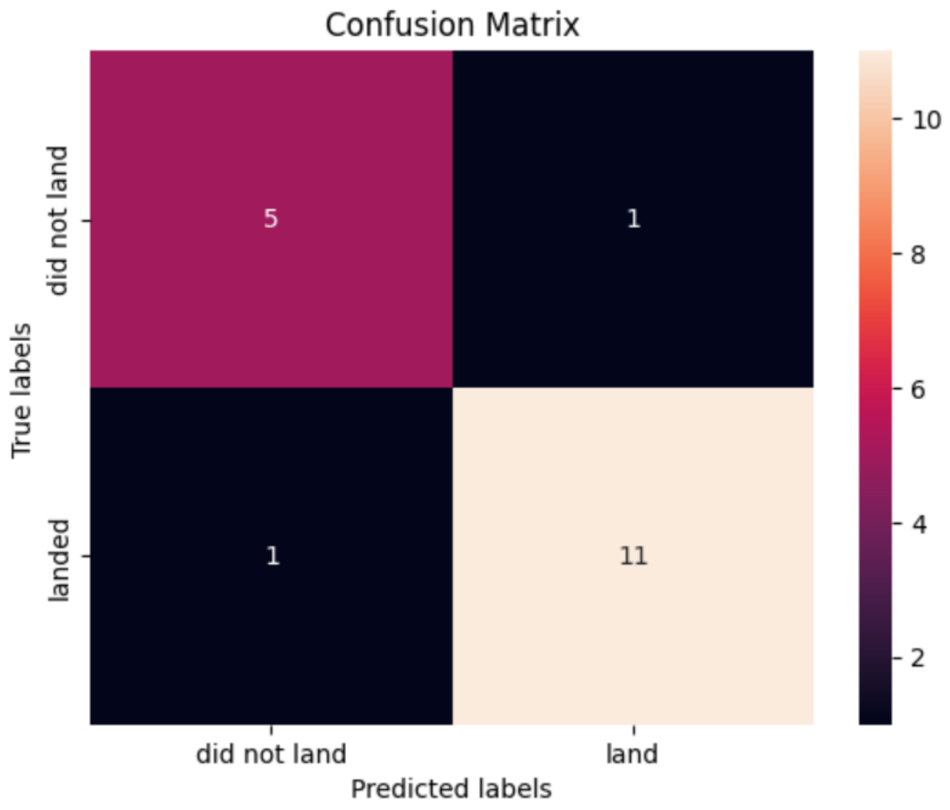
PREDICTIVE ANALYSIS (CLASSIFICATION)

RESULTS

Finding the Method Performs the Best

- The best method to perform is definitely Decision Tree Method.
- The model performs well, with high accuracy, precision, and recall, indicating it is effective at distinguishing between "landed" and "did not land" cases.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.846154	0.800000
F1_Score	0.888889	0.888889	0.916667	0.888889
Accuracy	0.833333	0.833333	0.888889	0.833333



CONCLUSION

- Decision Tree is the best model for the dataset with 88% accuracy.
- The success rate of launches increased through the years.
- KSC LC-39A has the highest success count with 41.7%.
- Success rate with a low payload mass show better results than larger payload mass.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.

APPENDIX

All the analysis are found in my GitHub:

https://github.com/shaulamarquez/IBM-Data-Science

Acknowledgement

Thank you to Instructors, Coursera and IBM.