IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Felipe Seleme Ribeiro
Jun/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Data Collection through Application Programming Interface - API
  - Web Scraping Data Collection
  - Data Wrangling
  - Exploratory Data Analysis with Structured Query Language – SQL EDA
  - Exploratory Data Analysis with Data Visualization – DataViz EDA
  - Interactive Dashboard Analysis with Folium
  - Machine Learning Predictions
- Summary of all results
  - Exploratory Data Analysis with Charts and Tables
  - Predictive Models Results for SpaceX Falcon 9 First Stage Landing Success

Introduction

- Project background and context

  In the project's introduction, we highlight the background and context of the study, focusing on SpaceX and its Falcon 9 rocket launches. SpaceX stands out by offering significantly lower prices, with a launch cost of $62 million compared to competitors' prices exceeding $165 million. This cost reduction stems from SpaceX's ability to recover and reuse the Falcon 9's first stage. If we can accurately predict the success of the first stage landing, it becomes possible to determine the cost associated with each launch. This valuable information can empower alternative companies to compete against SpaceX by making informed decisions when bidding for rocket launches, potentially challenging SpaceX's market dominance.

- Problems
  - Understand the factors influencing successful rocket landing.
  - Correlation between features and landing success.
  - Optimal operational conditions for successful landings.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Wikipedia web scaping with Python
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- The Falcon 9 first stage data collection was done with a get request to the SpaceX API and by web scraping the Falcon 9 Wikipedia page.

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Collection – SpaceX API

- Request to the SpaceX API
- Clean the requested data

[GitHub URL to Notebook]

1. Request and parse the SpaceX launch data

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
data = pd.json_normalize(response.json())
```

2. Filter the dataframe

```
data_falcon9 = launch_df[launch_df['BoosterVersion']!='Falcon 1']
```

3. Dealing with missing values

```
mean = data_falcon9['PayloadMass'].mean()
data_falcon9['PayloadMass'].fillna(value=mean, inplace=True)
```

4. Export to CSV

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection – Scraping

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

[GitHub URL to Notebook]

1. Request the Falcon9 Launch Wiki page

```
static_url = "https://en.wikipedia.org/w/index.php?title=
response = requests.get(static_url)
soup = BeautifulSoup(response.text, 'html.parser')
```

2. Extract all column/variable names

```
html_tables = soup.find_all('table')
```

3. Create a data frame

```
df=pd.DataFrame(launch_dict)
```

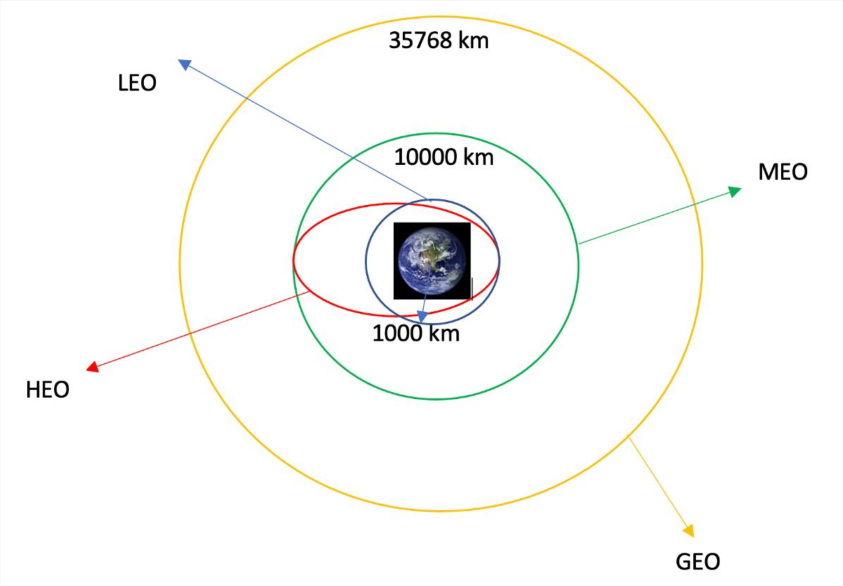
4. Export to CSV

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Wrangling

- Exploratory Data Analysis (EDA)
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the occurrence of mission outcome

- Determine Training Labels
  - Create a landing outcome label



Each launch aims to a dedicated orbit.

[GitHub URL to Notebook]

EDA with Data Visualization

- Scatter Graphs
  - Flight Number vs Payload Mass
  - Flight Number vs Launch Site
  - Payload Mass vs Launch Site
  - Flight Number vs Orbit
  - Payload Mass vs Orbit
- Bar Chart
  - Orbit vs Success Rate
- Line Chart
  - Year vs Sucess Rate

[GitHub URL to Notebook]

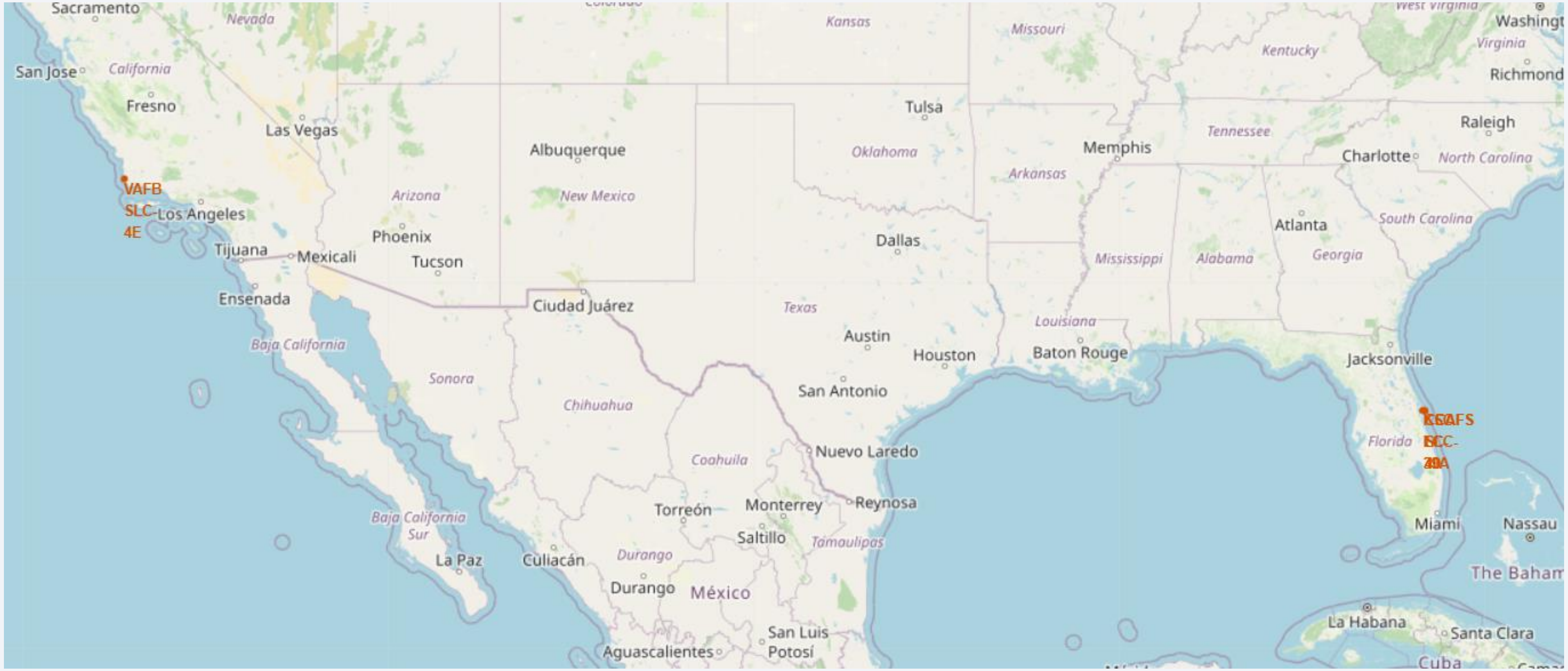
EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was achieved
- List the names of the boosters which have 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

GitHub URL to Notebook

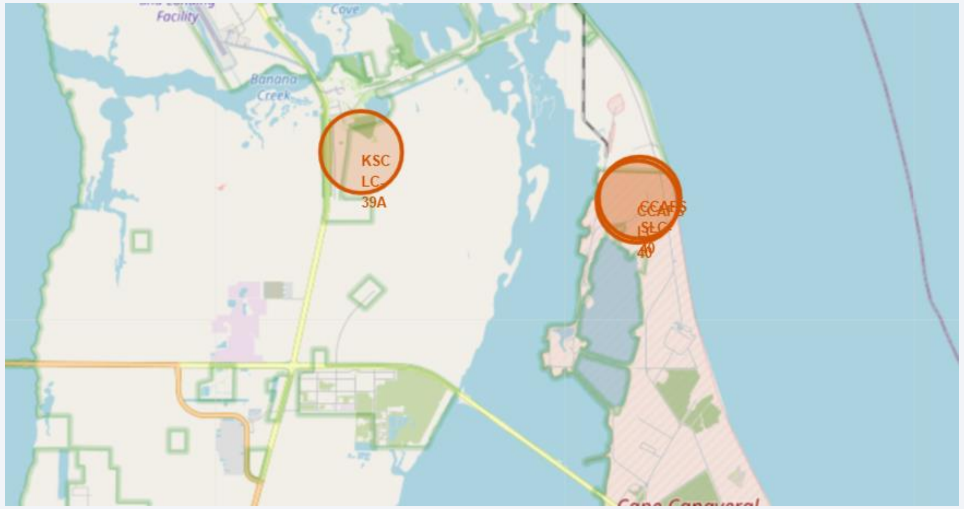
Build an Interactive Map with Folium

- Mark all launch sites on a map



Build an Interactive Map with Folium

- Mark all launch sites on a map

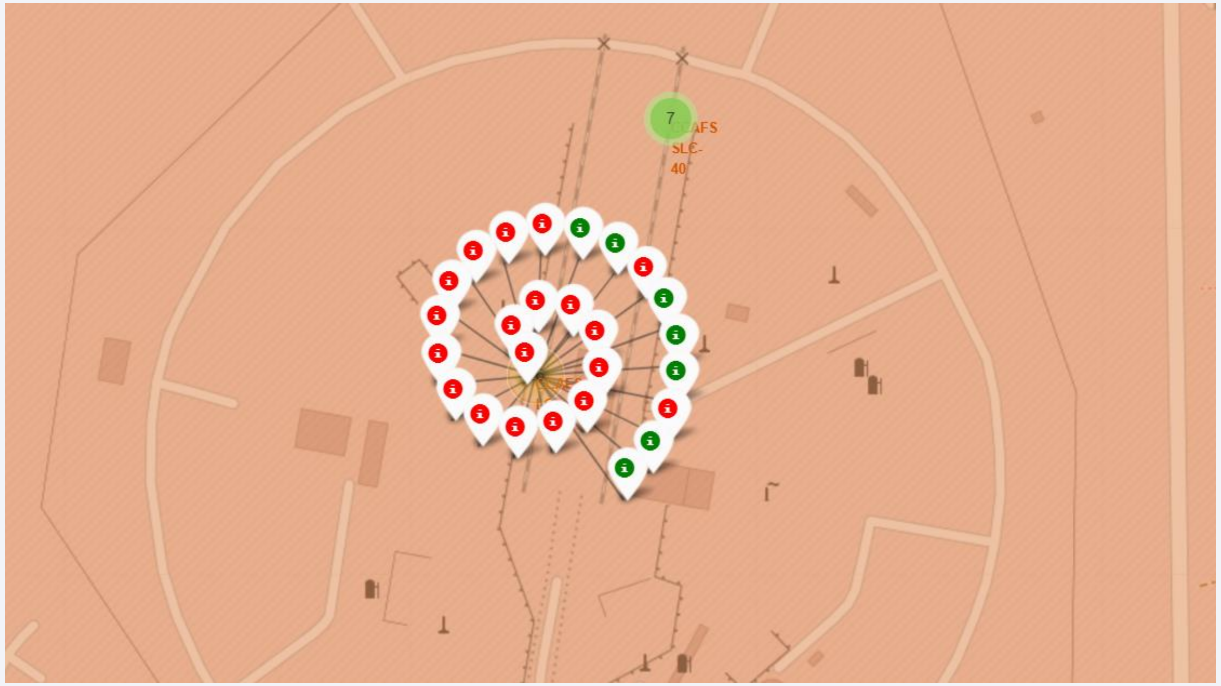


We can observe that all launch sites are in proximity to the Equator line and very close proximity to the coast.

[GitHub URL to Notebook]

Build an Interactive Map with Folium

- Mark the success/failed launches for each site on the map



Shows the success and failed launches for each site in na interactive map.

[GitHub URL to Notebook]

Build an Interactive Map with Folium

- Calculate the distances between a launch site to its proximities

Shows the distances between a launch site to its proximities in an interactive map.

[GitHub URL to Notebook]()

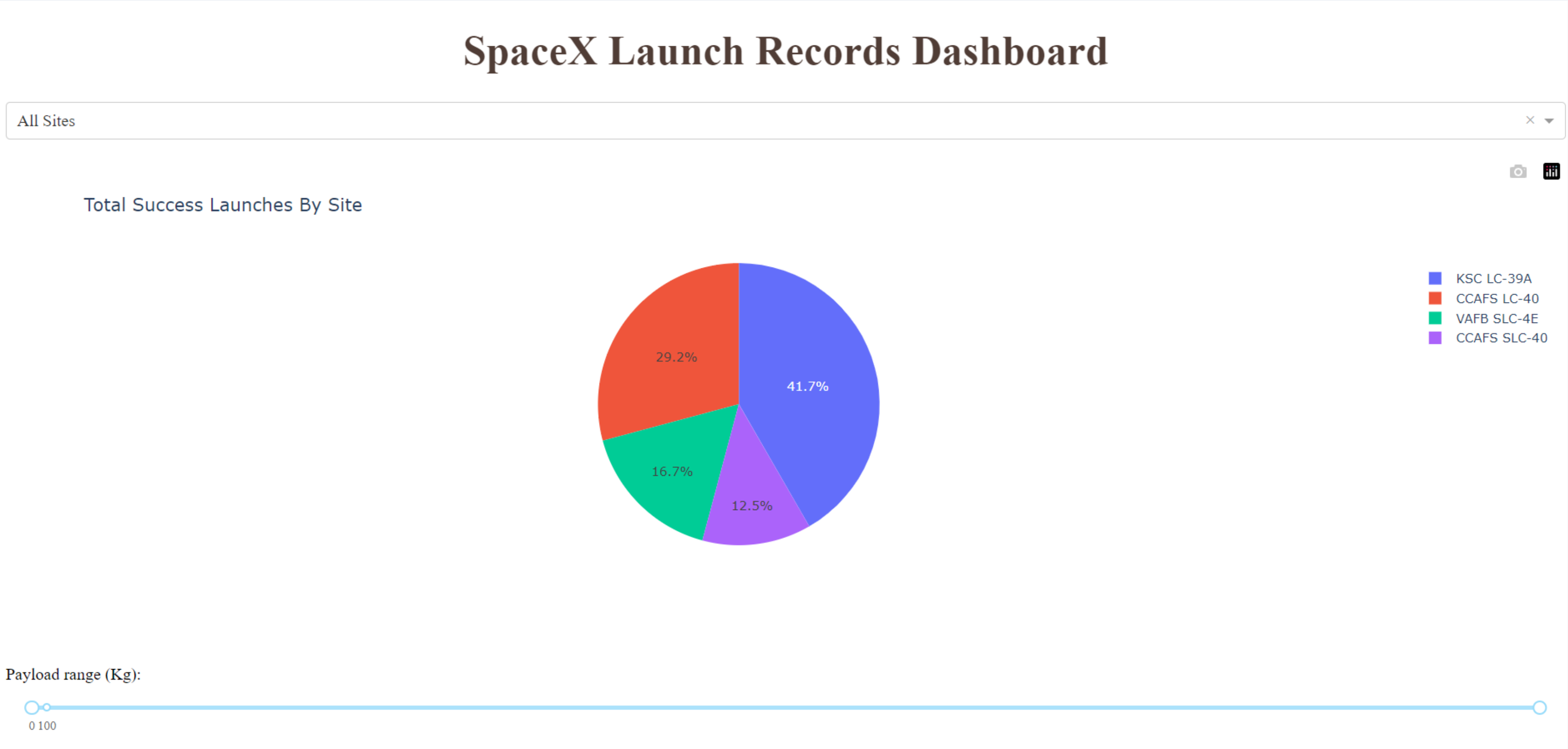
Build a Dashboard with Plotly Dash

- The dashboard is built with Flask and Dash web framework.

- Graphs
  - Pie Chart showing the total launches by a certain site or all sites.
  - Scatter Graph showing the relationship with Outcome and Payload Mass

[GitHub URL to Notebook]

Build a Dashboard with Plotly Dash



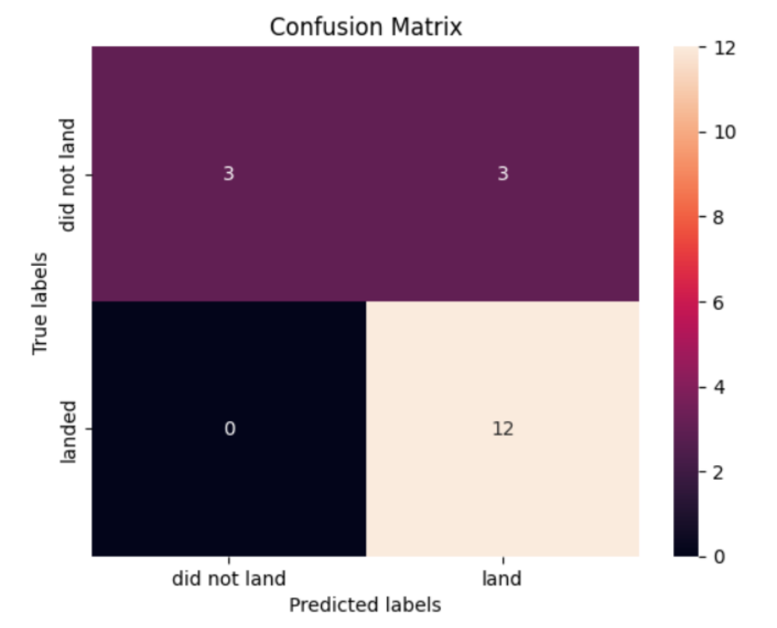
GitHub URL to Notebook

Build a Dashboard with Plotly Dash

[GitHub URL to Notebook]

Predictive Analysis (Classification)

- Load the data
- Standardize the data and fit the model
- Split the data into train and test
- Create different models:
    - Logistic Regression
    - Support Vector Machine (SVM)
    - Decision Tree Classifier
    - K Nearest Neighbors
- Calculate the accuracy on each model
- Build a confusion matrix to visualize the accuracy
- Compare the predictions accuracy for each model



GitHub URL to Notebook

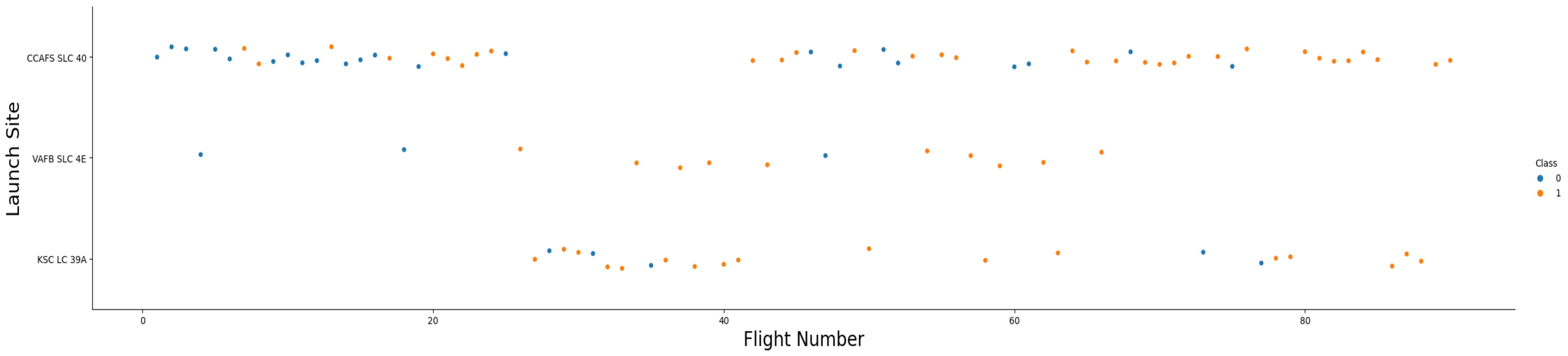
Results

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

Section 2

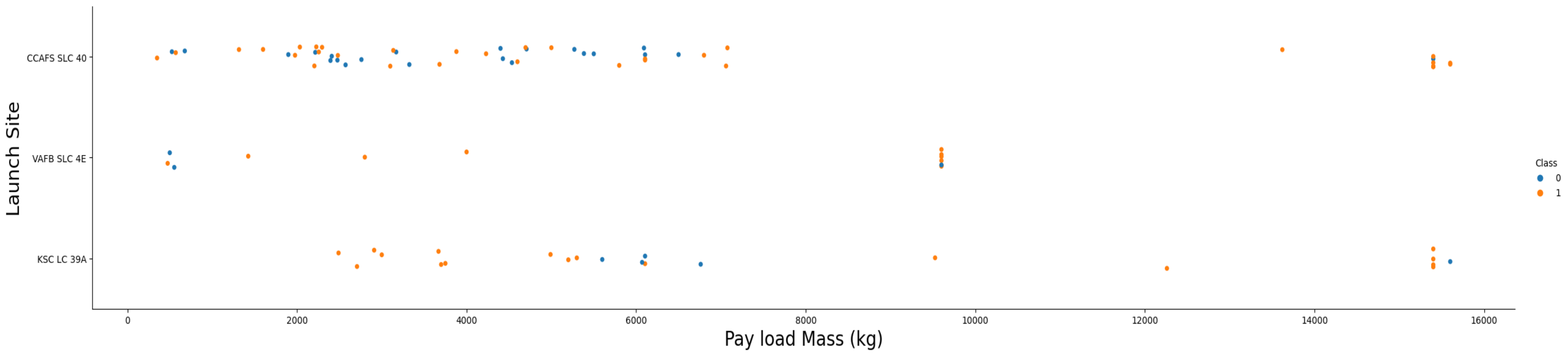
Insights drawn from EDA

Flight Number vs. Launch Site



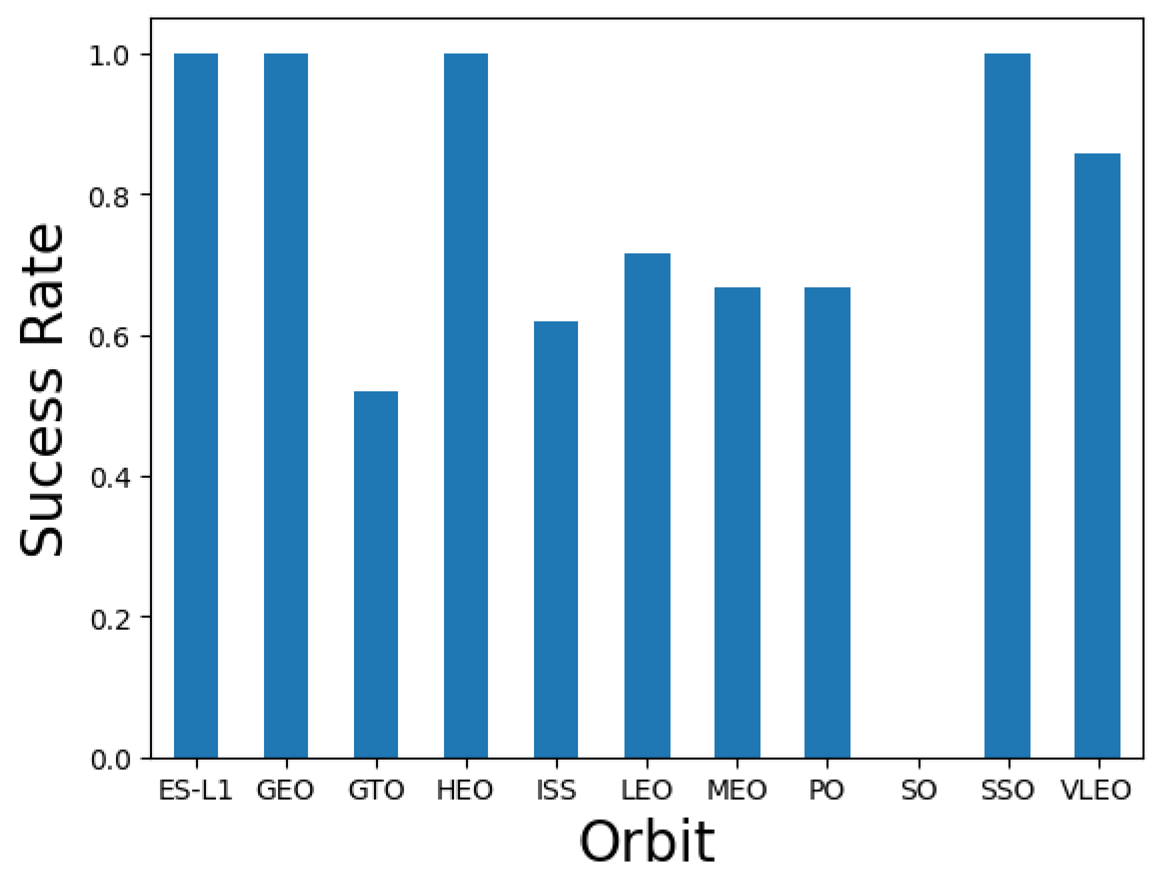
The larger the flight amount at a launch site, the greater the success rate at a launch site.

Payload vs. Launch Site



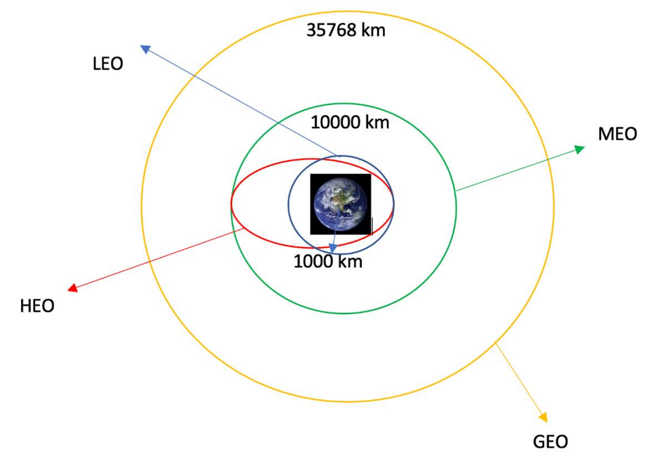
It seems that the greater the payload for Launch Site CCAFS SCL 40, the higher the success rate for the rocket. However, there is not a quite clear pattern for the other Launch Sites.

Success Rate vs. Orbit Type

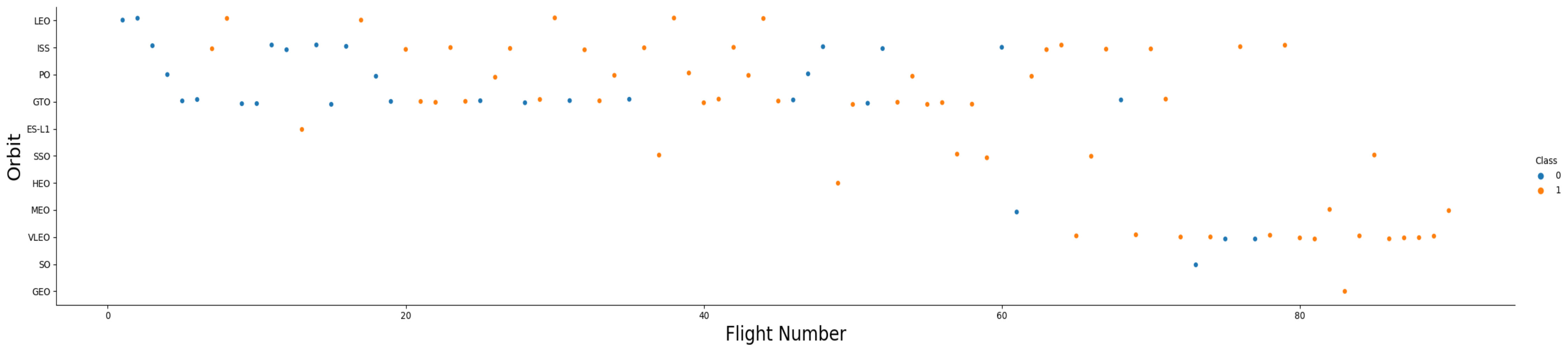


The most success rate are for the orbits:

- ES-L1
- GEO
- HEO
- SSO

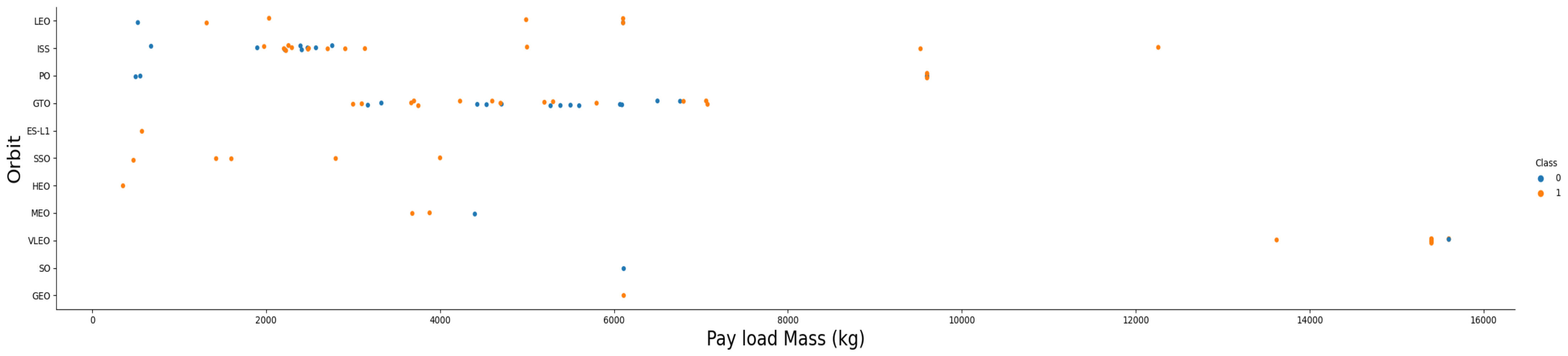


Flight Number vs. Orbit Type



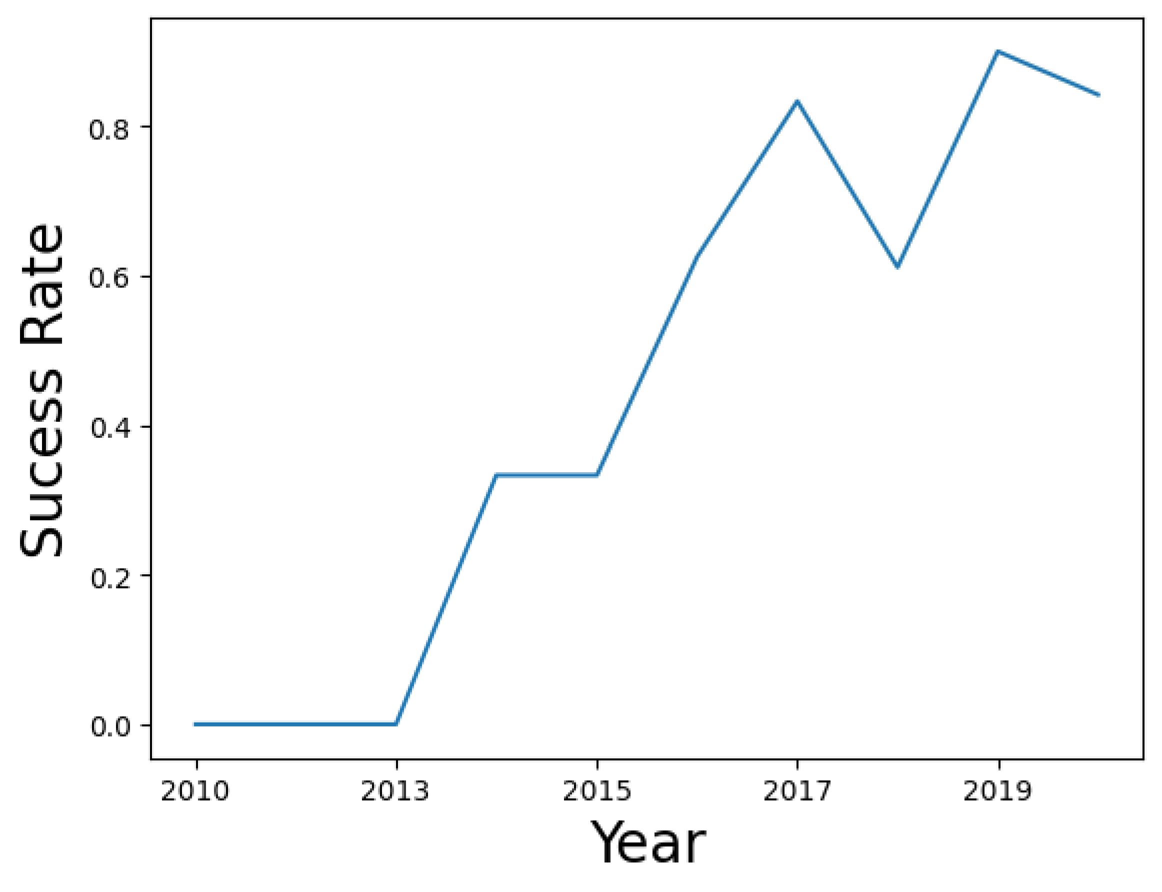
In the LEO orbit, it seems to have a relation between the number of flights and the success. However, the GTO orbit shows no relationship between flight number and success.

Payload vs. Orbit Type



Heavy payloads seems to have a negative influence on GTO orbits and positive on LEO and ISS orbits.

Launch Success Yearly Trend



The success rate kept increasing since 2013 till 2020

All Launch Site Names

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

✓ 0.0s

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

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

Find the names of the unique launch sites

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

```
%%sql
SELECT *
FROM SPACEXTBL
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
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE CUSTOMER = 'NASA (CRS)';
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

SUM(PAYLOAD_MASS__KG_)
45596.0

Calculate the total payload carried by boosters from NASA

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE BOOSTER_VERSION = 'F9 v1.1';
```

✓ 0.0s

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Calculate the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

```
%%sql
SELECT MIN(DATE) AS 'First succesful landing in ground pad'
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (ground pad)';
```

✓ 0.0s

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

First succesful landing in ground pad
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

✓ 0.0s

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

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

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME)
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

✓ 0.0s

* sqlite:///my_data1.db
Done.

Mission_Outcome	COUNT(MISSION_OUTCOME)
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%%sql
SELECT booster_version, PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL
    );
```

✓ 0.0s

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600.0
F9 B5 B1049.4	15600.0
F9 B5 B1051.3	15600.0
F9 B5 B1056.4	15600.0
F9 B5 B1048.5	15600.0
F9 B5 B1051.4	15600.0
F9 B5 B1049.5	15600.0
F9 B5 B1060.2	15600.0
F9 B5 B1058.3	15600.0
F9 B5 B1051.6	15600.0
F9 B5 B1060.3	15600.0
F9 B5 B1049.7	15600.0

List the names of the booster which have carried the maximum payload mass

2015 Launch Records

```
%%sql
SELECT
    SUBSTR("Date", 4, 2) AS Month,
    LANDING_OUTCOME,
    BOOSTER_VERSION,
    "Launch Site"
FROM
    SPACEXTBL
WHERE
    SUBSTR("Date", 7, 4) = "2015"
    AND LANDING_OUTCOME LIKE "Failure%"
    AND LANDING_OUTCOME LIKE "%drone ship%";
```

✓ 0.0s

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

Month	Landing_Outcome	Booster_Version	"Launch Site"
10	Failure (drone ship)	F9 v1.1 B1012	Launch Site
04	Failure (drone ship)	F9 v1.1 B1015	Launch Site

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

```
%%sql
SELECT
    LANDING_OUTCOME AS Successful_Landing_Outcome,
    COUNT(*) AS Count
FROM
    SPACEXTBL
WHERE
    "Date" BETWEEN "04-06-2010" AND "20-03-2017"
    AND LANDING_OUTCOME LIKE "Success%"
GROUP BY
    LANDING_OUTCOME
ORDER BY
    Count DESC;
```

✓ 0.0s

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

Successful_Landing_Outcome	Count
Success	20
Success (drone ship)	8
Success (ground pad)	7

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

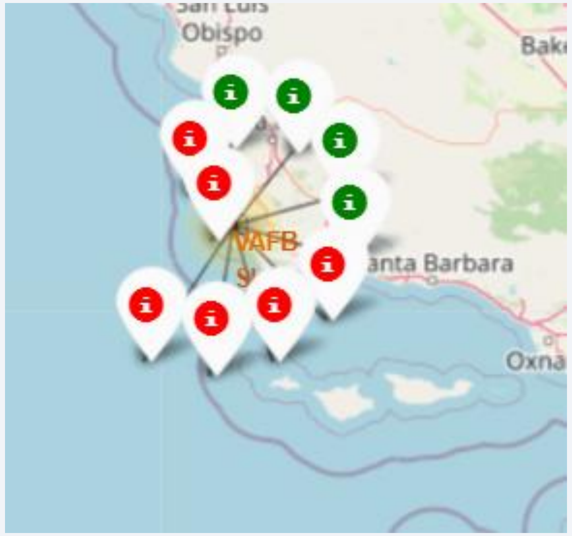
Section 3

Launch Sites Proximities Analysis

All Launch Sites – Global Markers

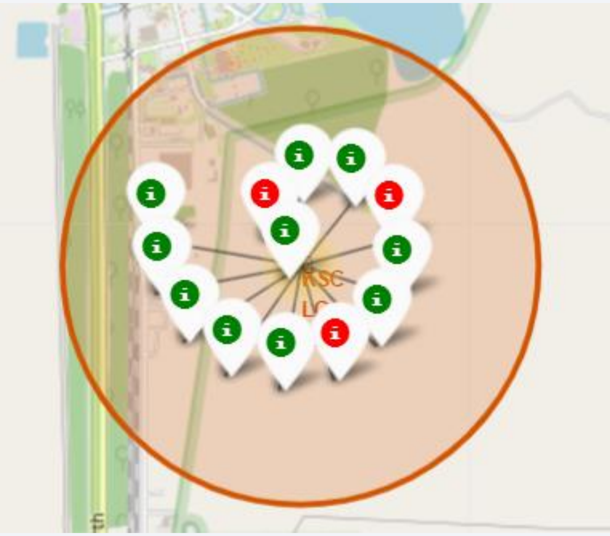
As we can se, the SpaceX launch sites are in the United States of America coasts (Florida and California).

Color Labelled Markers



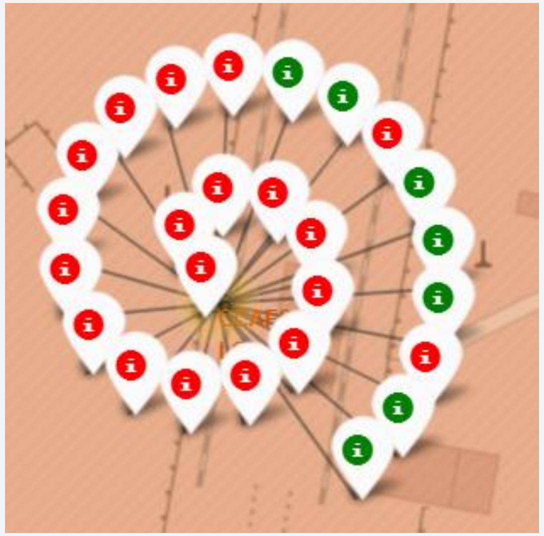
California Launch Site

VAFB-SLC4E



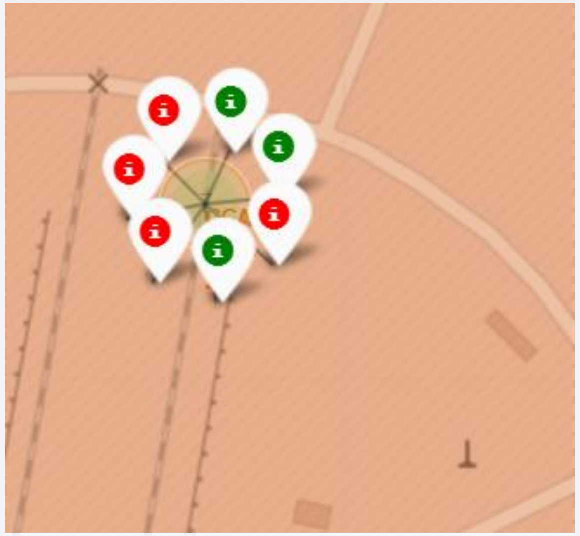
Florida Launch Site

KSC LC-39A



Florida Launch Site

CCAFS LC-40



Florida Launch Site

CCAFS SLC-40

*Green Markers shows successful launches and Red Markers shows failures.

Distance to Landmarks

Distance to coastline

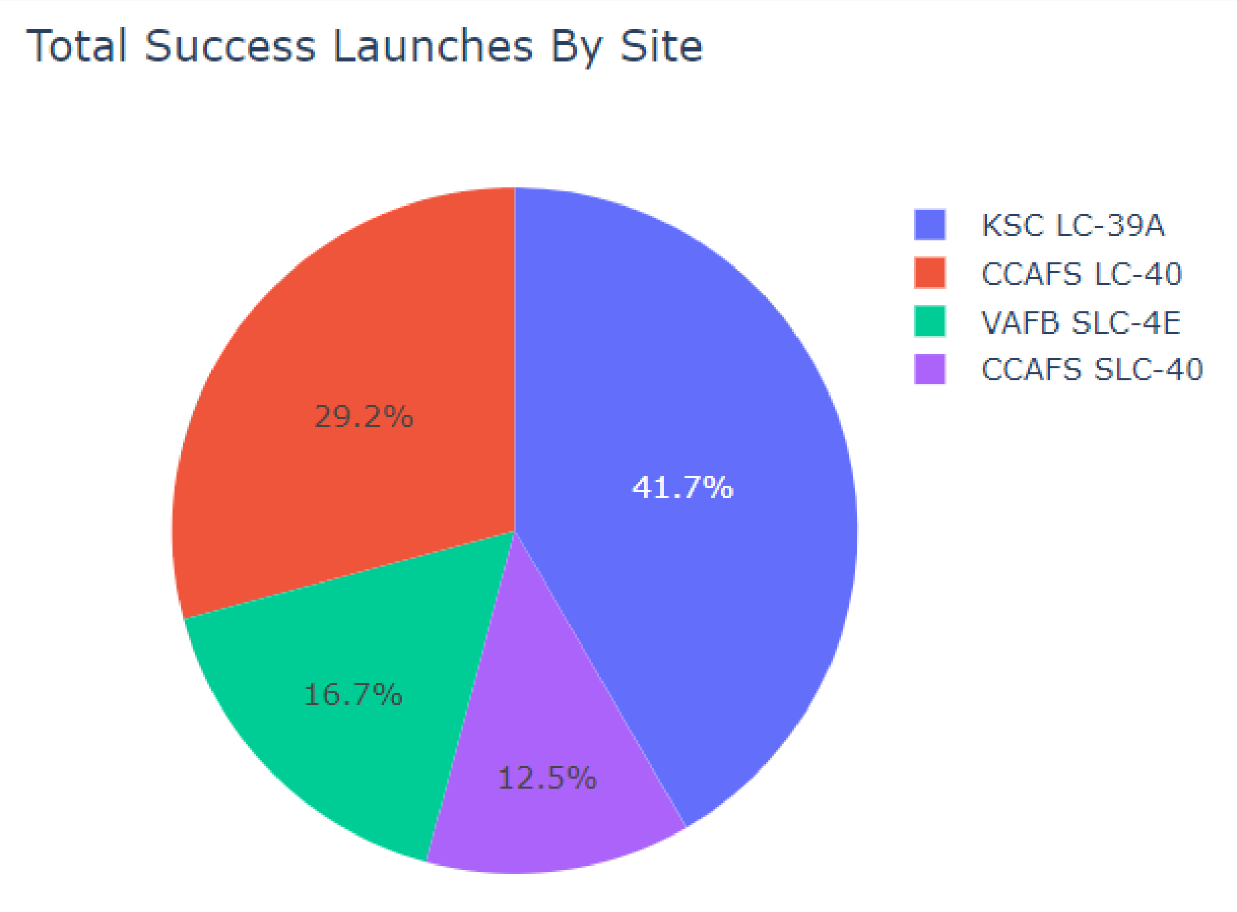
Distance to city

The launch sites are in close proximity to coastline and keep certain distance away from cities, railways and highways.

Section 4

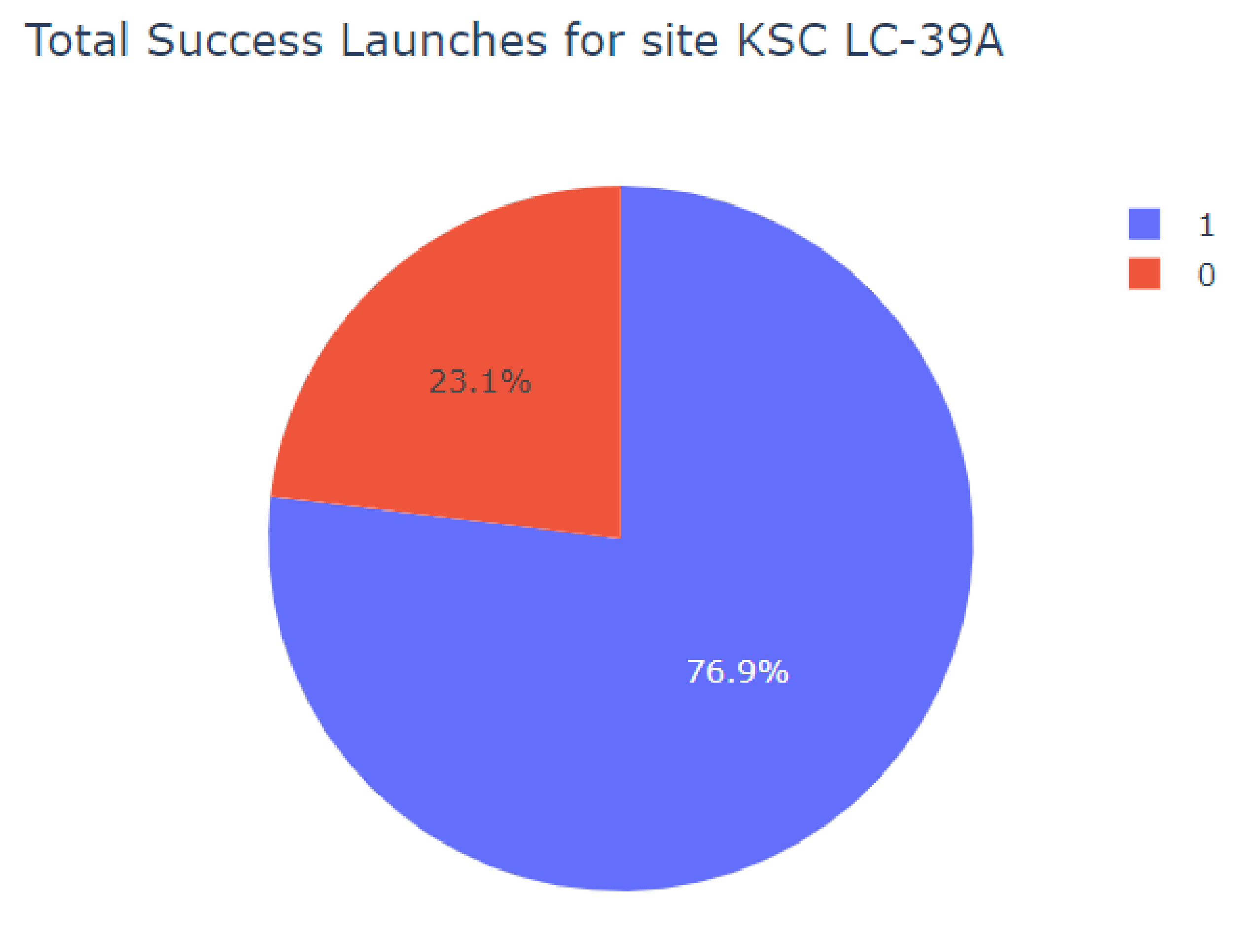
Build a Dashboard with Plotly Dash

Total Success Launches By Site



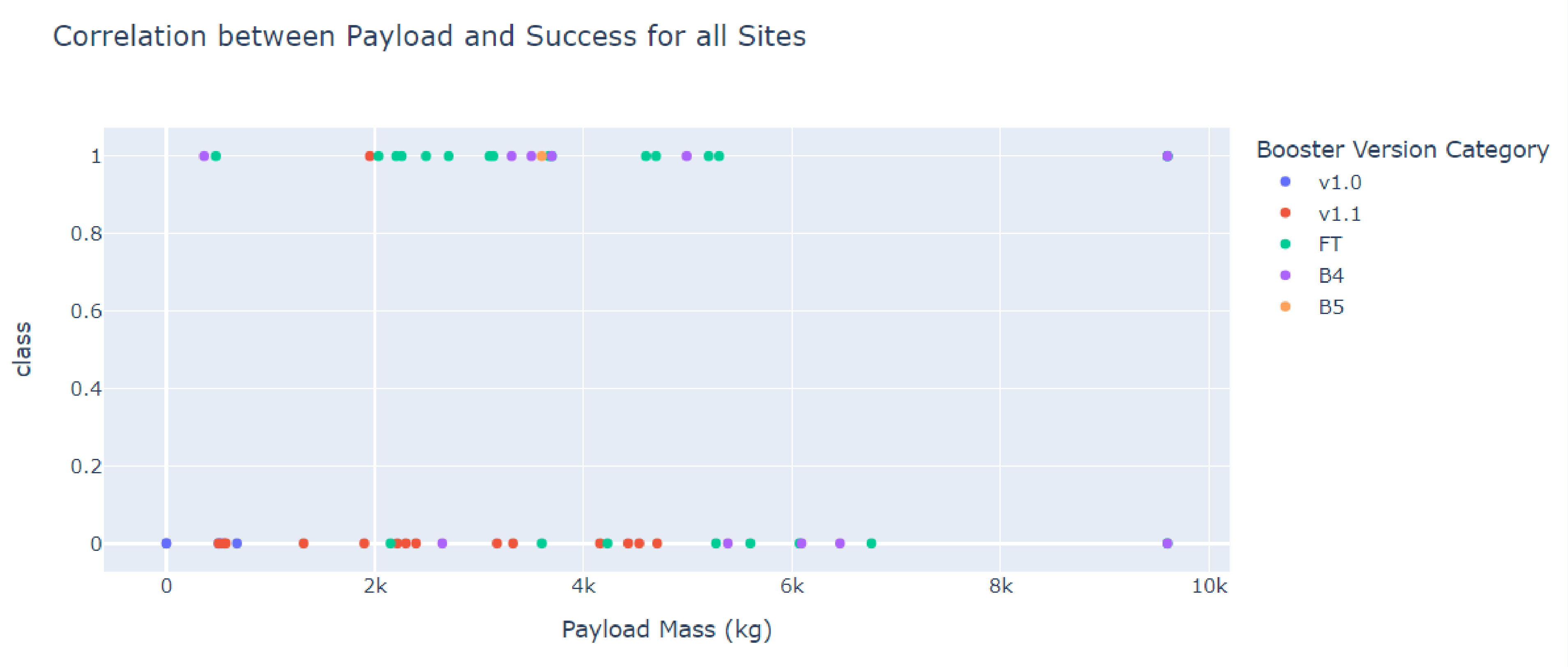
- KSC LC-39 A has the most successful launches

Total Success Launches for KSC LC-39A



- KSC LC-39A achieved a 76,6% success rate.

Payload vs Success for Booster Category

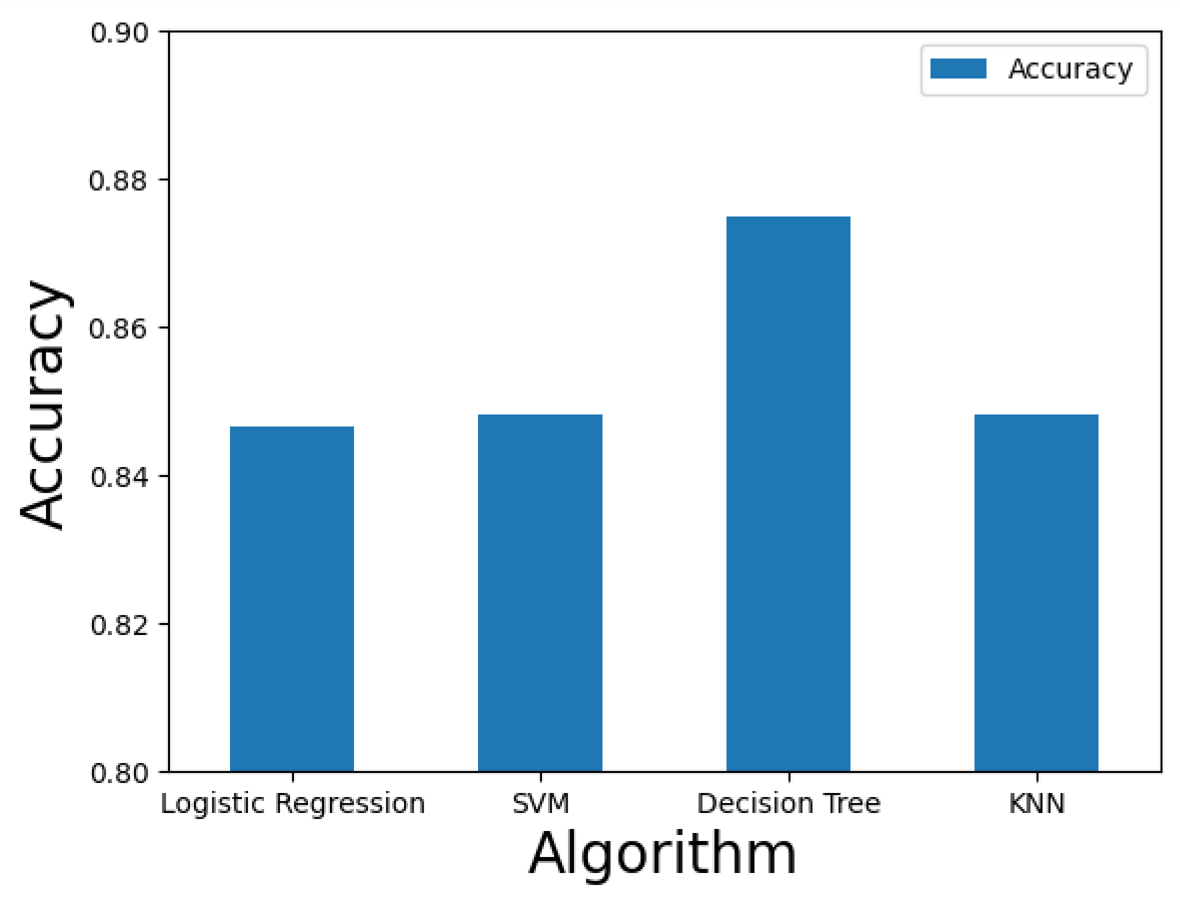


We can see that the success rate is higher for low weighted payloads.

Section 5

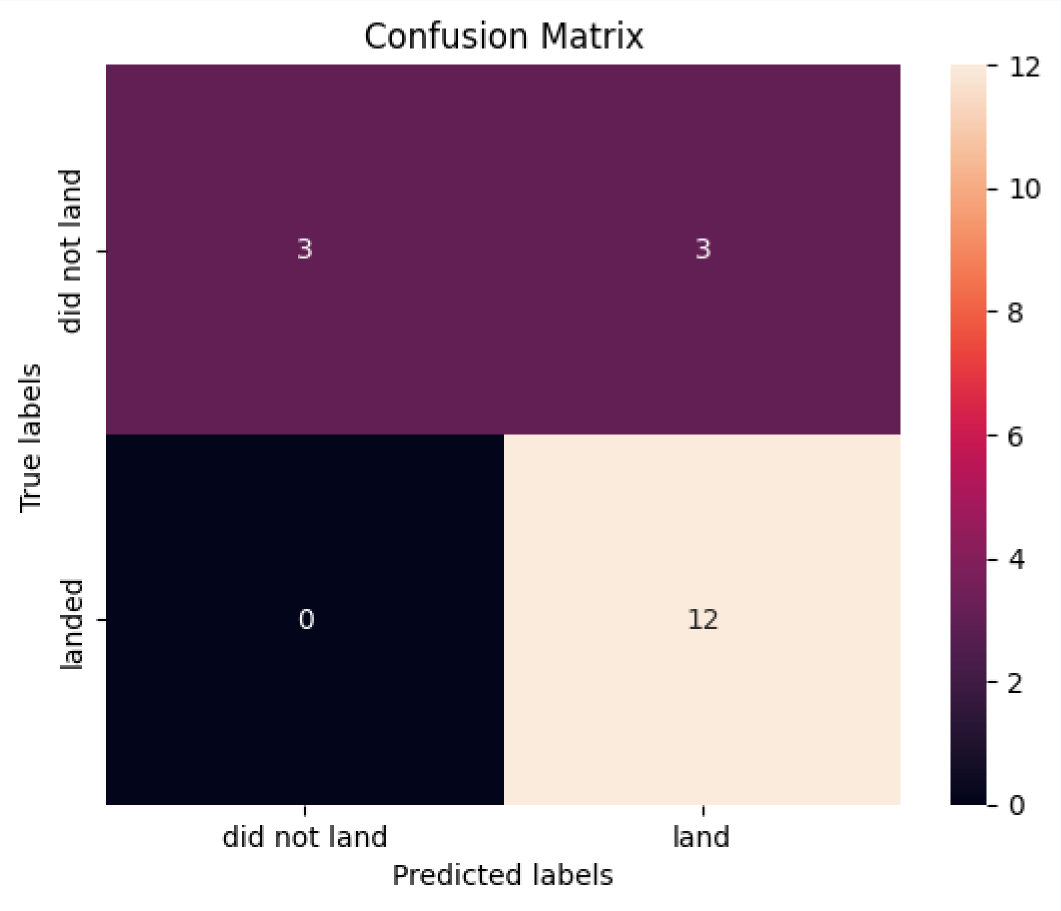
Predictive Analysis (Classification)

Classification Accuracy



- The Decision Tree model has the highest classification accuracy

Confusion Matrix



- The Decision Tree model predicted 12 landed and 3 not landed right. But missed 3 not landed.

Conclusions

- Correlation between launch site activity and success rate: There is a positive relationship between the number of flights conducted at a launch site and the likelihood of a successful launch. As the flight amount increases at a particular site, the success rate also tends to increase.
- Temporal trends in launch success: Starting from 2013 and continuing until 2020, there has been a notable upward trend in launch success rates. Over this period, the probability of a successful launch has been consistently increasing.
- Success rates for different orbits: Orbits such as ES-L1, GEO, HEO, SSO, and VLEO demonstrate the highest success rates compared to other orbits. These specific orbits have shown a greater likelihood of achieving a successful mission.
- Outstanding performance of KSC LC-39A launch site: Among all launch sites, KSC LC-39A stands out as the site with the highest number of successful launches. It has consistently achieved a remarkable track record in terms of launch success.
- Optimal machine learning algorithm: The Decision Tree model has been identified as the most effective machine learning algorithm for this particular task. It outperforms other algorithms in accurately predicting the success or failure of rocket launches.

Thank you!