Winning Space Race with Data Science

Vivian Tan Wei Xin

9 May 2022

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- Results from EDA
- Interactive analytics results based on variables such as locations, payload mass and booster version
- Predictive analysis results using machine learning methods

Introduction

Space X Falcon 9 rocket launches cost 62 million dollars, whilst other providers cost 165 million dollars and above. The major cost saving is due to Space X's reusable first stage. Therefore, by determining if the first stage will land successfully, we can determine the cost of a launch.

The problems we need to solve:

1) What are the factors that determine if the first stage will land successfully
2) The relationships between each variables, such as flight number, payload mass, orbit type, launch site and success rate
3) What is the best model to predict the first stage's success rate of landing

Section 1

Methodology

Methodology - Executive Summary

Data Collection

- SpaceX API
- Web scraping from Wikipedia using BeautifulSoup

Data wrangling

- Finding out the number of launches on each site, occurrence at each orbit and the landing outcome
- Creating landing outcome labels

Perform exploratory data analysis (EDA)

- Finding out the relationships among flight number, launch site, payload, orbit type and success rate via data visualization
- Understanding the dataset using SQL queries

Perform interactive visual analytics

- Analyzing Launch Sites Locations with Folium
- Building a Dashboard Application with Plotly Dash

Perform predictive analysis using classification models

- Building models such as SVM, Classification Trees and Logistic Regression
- Determine the method that performs best

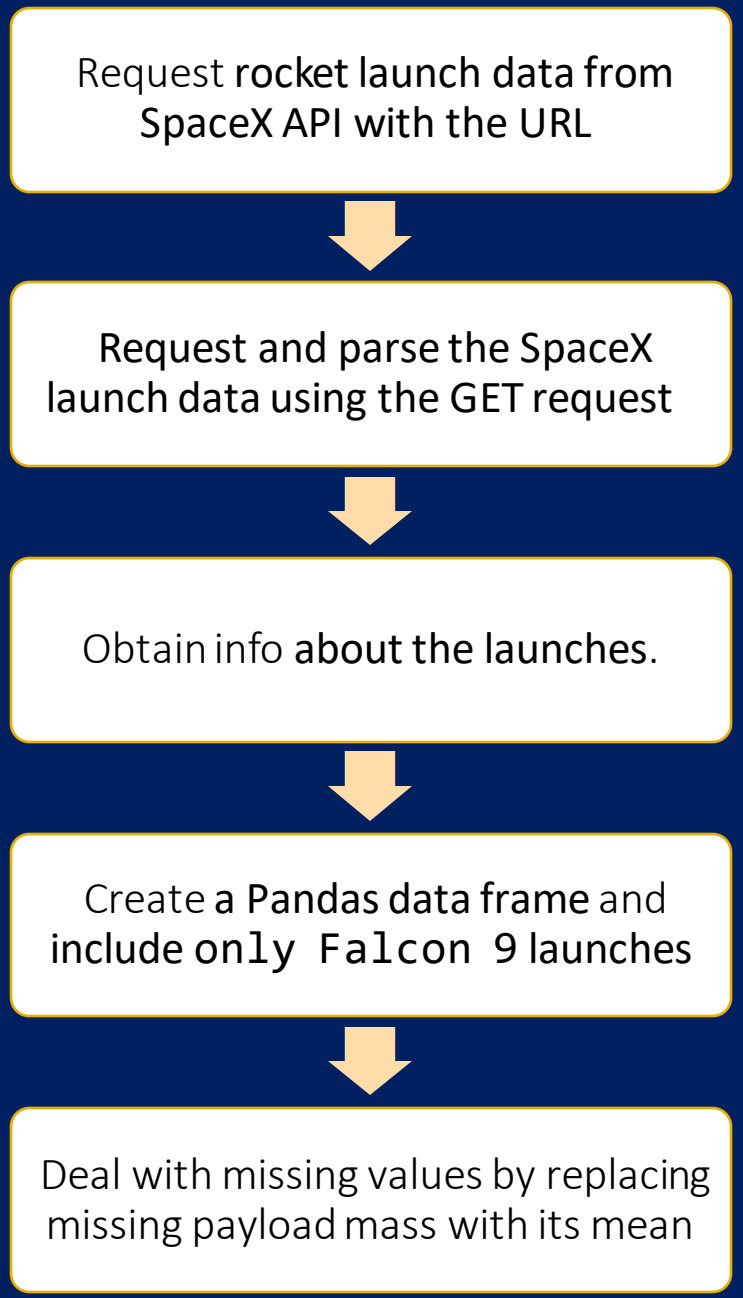
Data Collection

- SpaceX API

Our objectives:

- Obtain information about
  - booster name
  - mass of the payload
  - the orbit that it is going to
  - name and location of the launch sites used
  - outcome of the landing
- Clean the requested data

Reference : Link to API call notebook



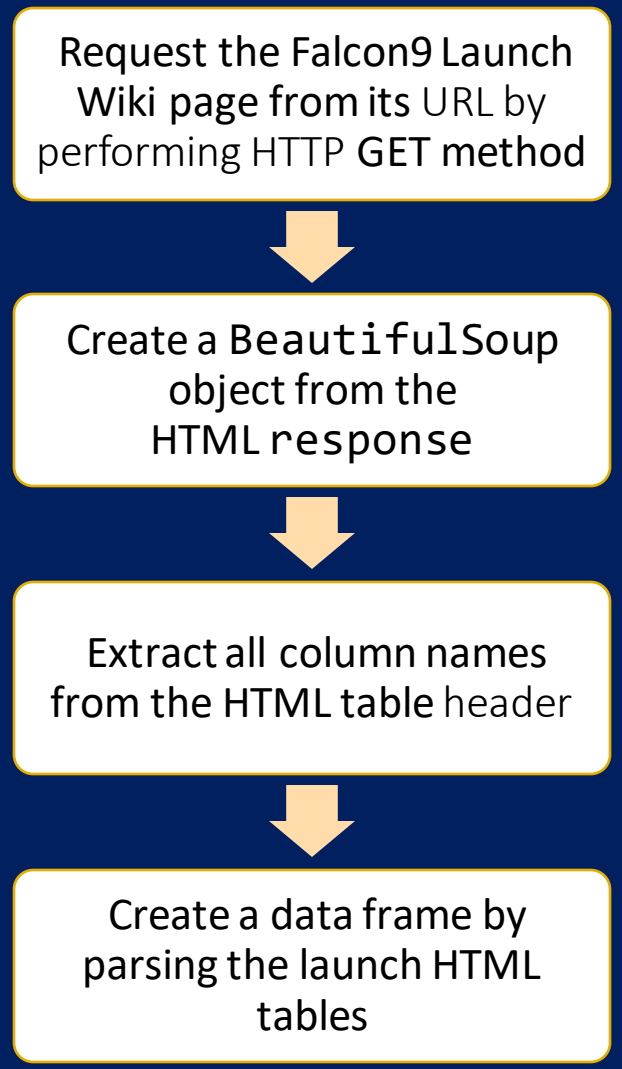
Data Collection

- Web Scraping

Our objectives:

- Collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches with Beautiful Soup
- Create a data frame by parsing the launch HTML tables

Reference : Link to Web Scraping notebook

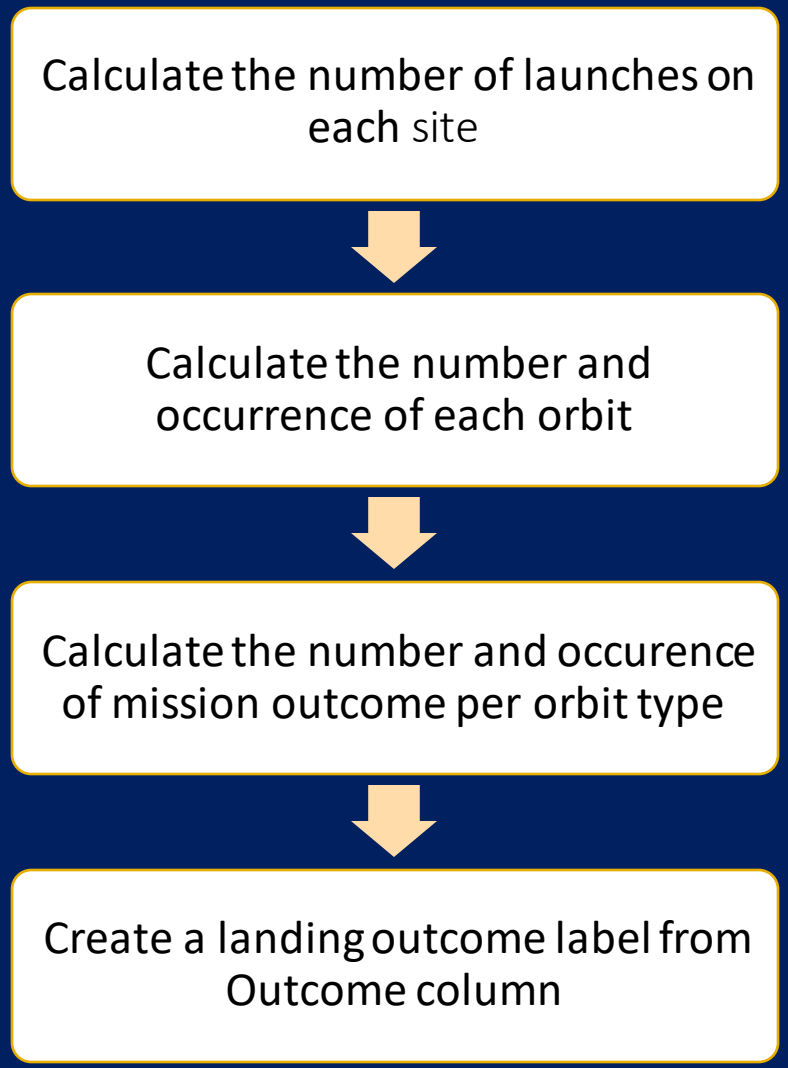


Data Wrangling

Our objectives:

- Find patterns in the data and determine what would be the label for training supervised models.
- Convert the landing outcomes into Training Labels with `1` means the booster successfully landed `0` means it was unsuccessful.

Reference : Link to Data Wrangling notebook



EDA with Data Visualization

The following charts were plotted to determine the respective relationships:

- Scatter Plot
  - Flight Number & Launch Site
  - Payload & Launch Site
  - Orbit Type & Success Rate
  - Flight Number & Orbit type
- Bar Chart
  - Payload & Orbit type
- Line Chart
  - Success Rate & Year

Reference: Link to Data Visualization Notebook

EDA with SQL

Our objectives:

- Understand the SpaceX dataset by executing SQL queries
- Load the dataset into the corresponding table in a Db2 database

Reference: Link to EDA with SQL Notebook

- Display the names of the unique launch sites.
- 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 successful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass between 4000 and 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 failed landing_outcomes in drone ship, their booster versions, and launch site names for 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.

Build an Interactive Map with Folium

Mark all launch sites on a map with makers and circles

- Visualize the locations by pinning them on a map.

Mark the success launches for each site using green markers and failed ones using red markers

- Identify which launch sites have relatively high success rates.

Calculate the distances between a launch site to its proximities and use markers and lines to show its distances

- Identify if the launch site is in close proximity to railways, highways, coastline and cities.

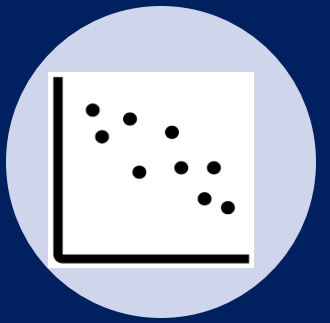
Reference: Link to Folium Notebook

Build a Dashboard with Plotly Dash

ADD A LAUNCH SITE DROP-DOWN INPUT COMPONENT

ADD A CALLBACK FUNCTION TO RENDER SUCCESS-PIE-CHART BASED ON SELECTED SITE DROPDOWN

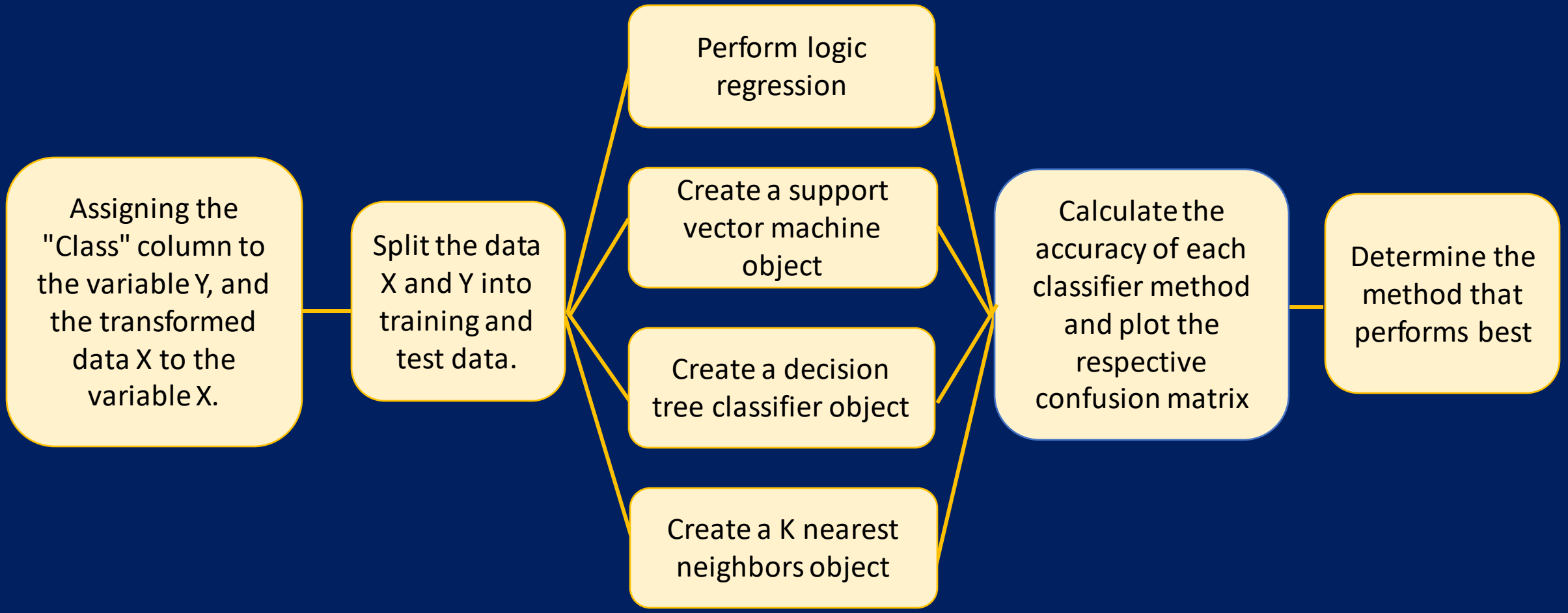
ADD A RANGE SLIDER TO SELECT PAYLOAD



ADD A CALLBACK FUNCTION TO RENDER THE SUCCESS-PAYLOAD-SCATTER-CHART SCATTER PLOT

Reference: Link to Plotly Dash Notebook

Predictive Analysis (Classification)

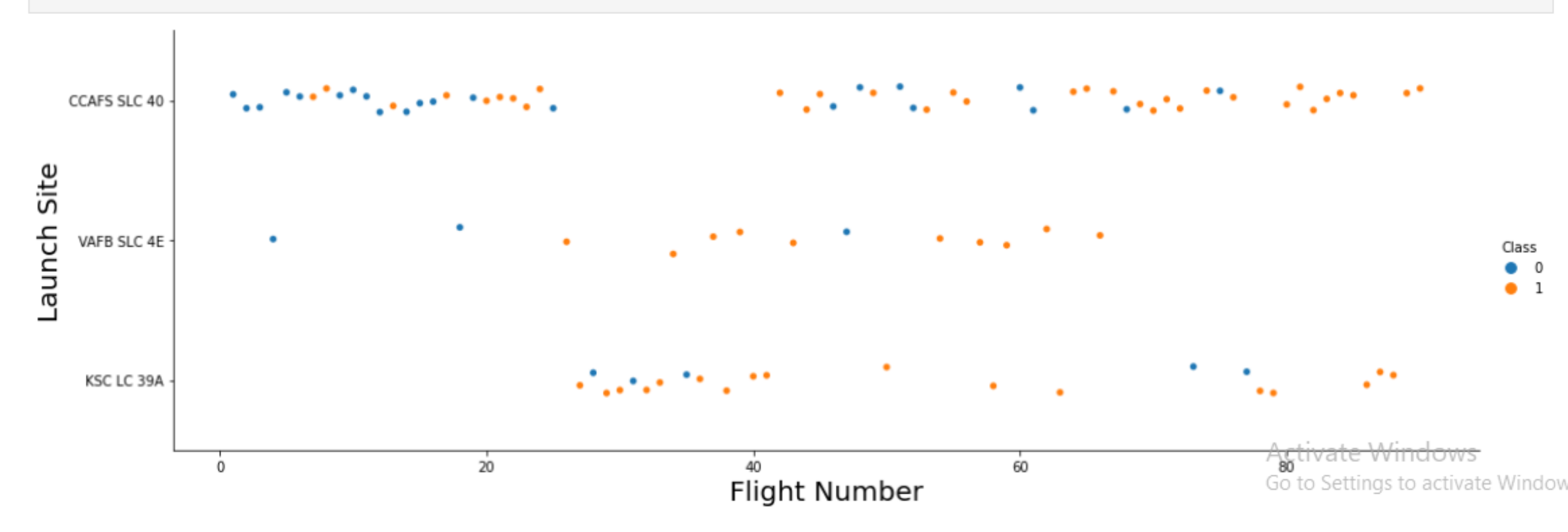


Reference: Link to Predictive Analysis Notebook

Section 2

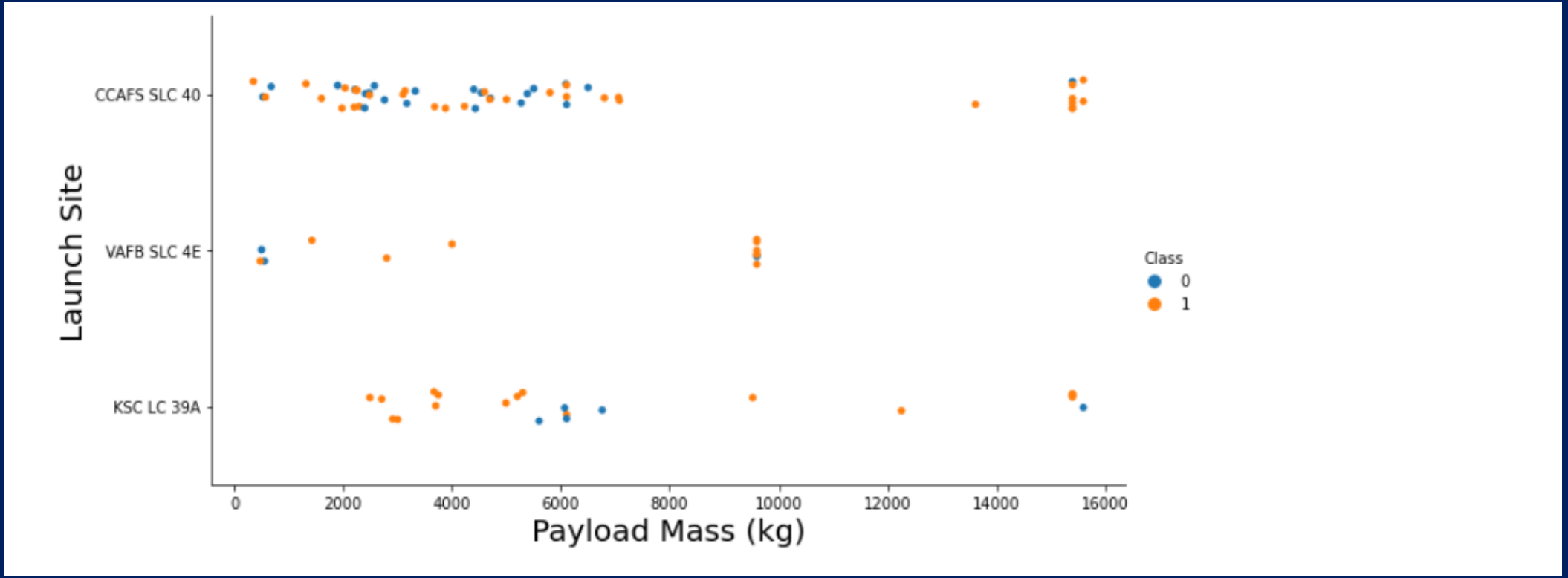
Insights drawn from EDA

Flight Number vs. Launch Site



Here we observe that the higher the flight number at a launch site, the higher the success rate.

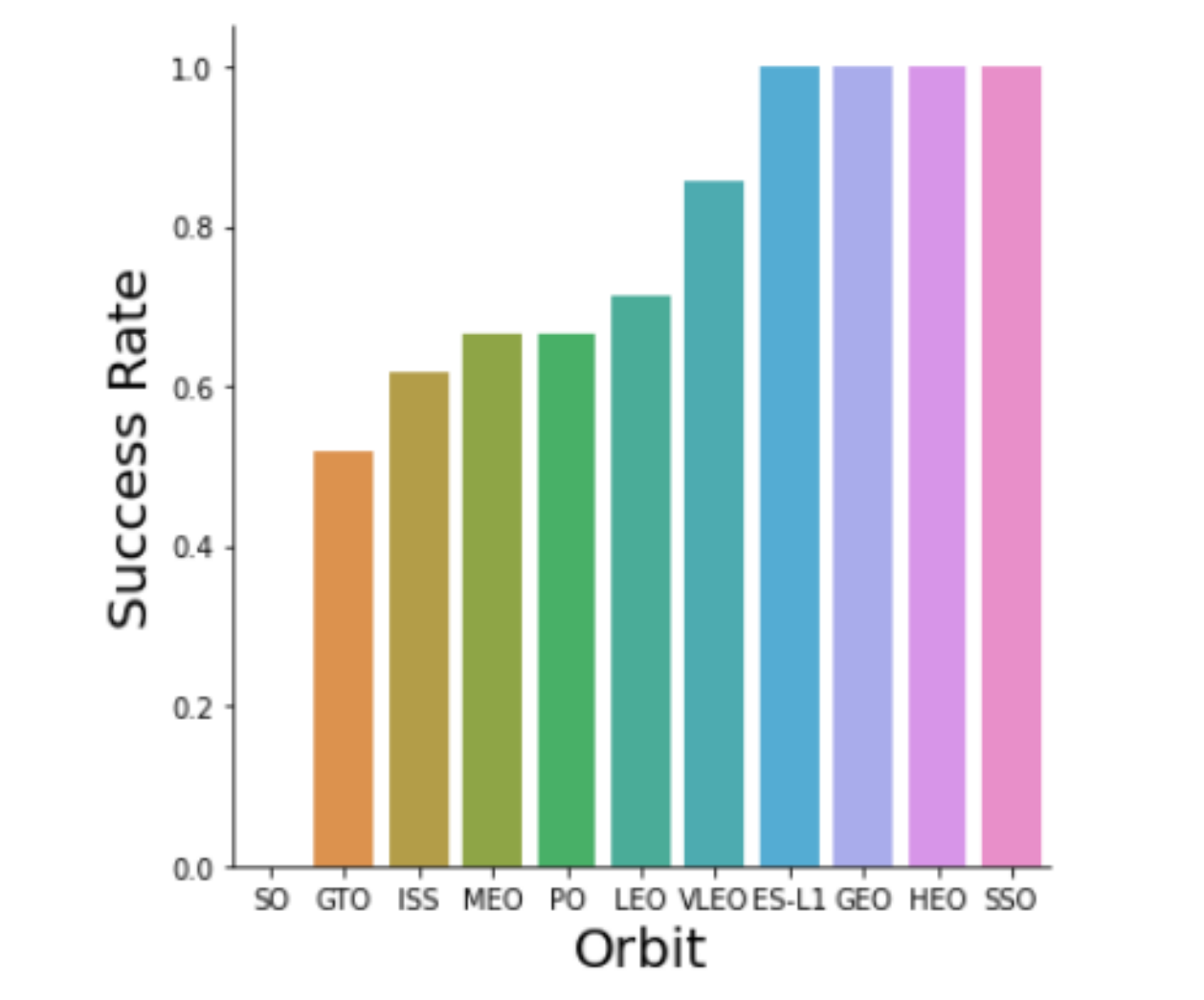
Payload vs. Launch Site



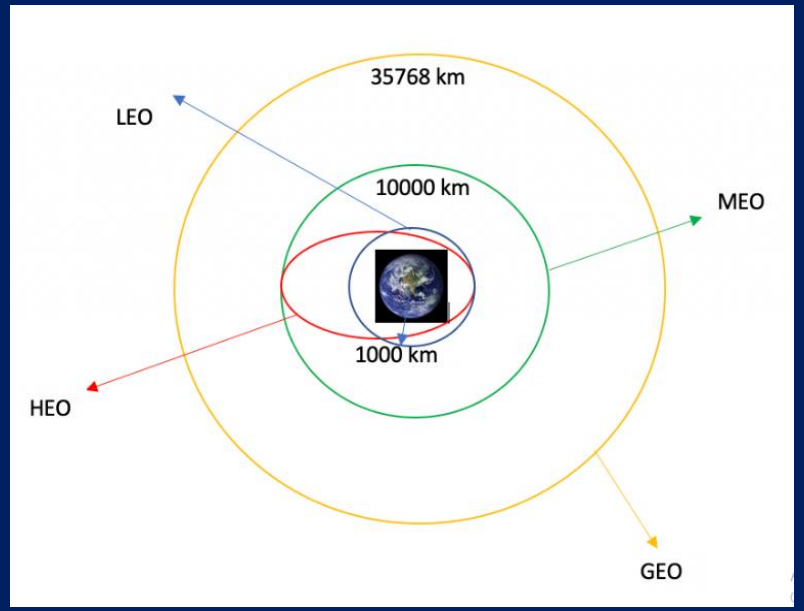
For the VAFB-SLC launch site there is no rocket with heavy payload mass (>10000) launched.

There is also no clear pattern that indicates the success rate of the launch at the sites is dependent on the payload mass.

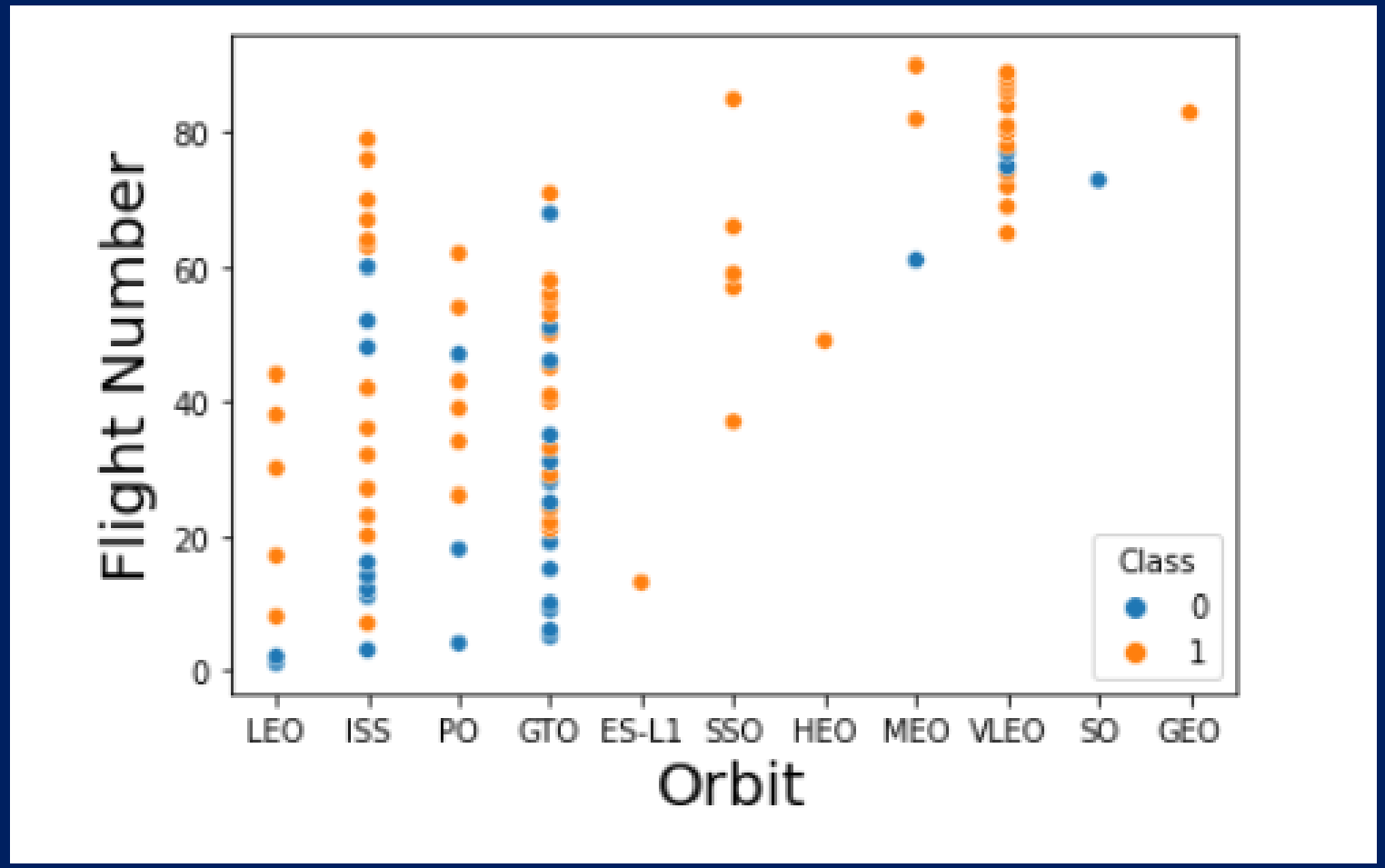
Success Rate vs. Orbit Type



The rockets that were launched to orbit ES-L1, GEO, HEO and SSO have the highest success rate.

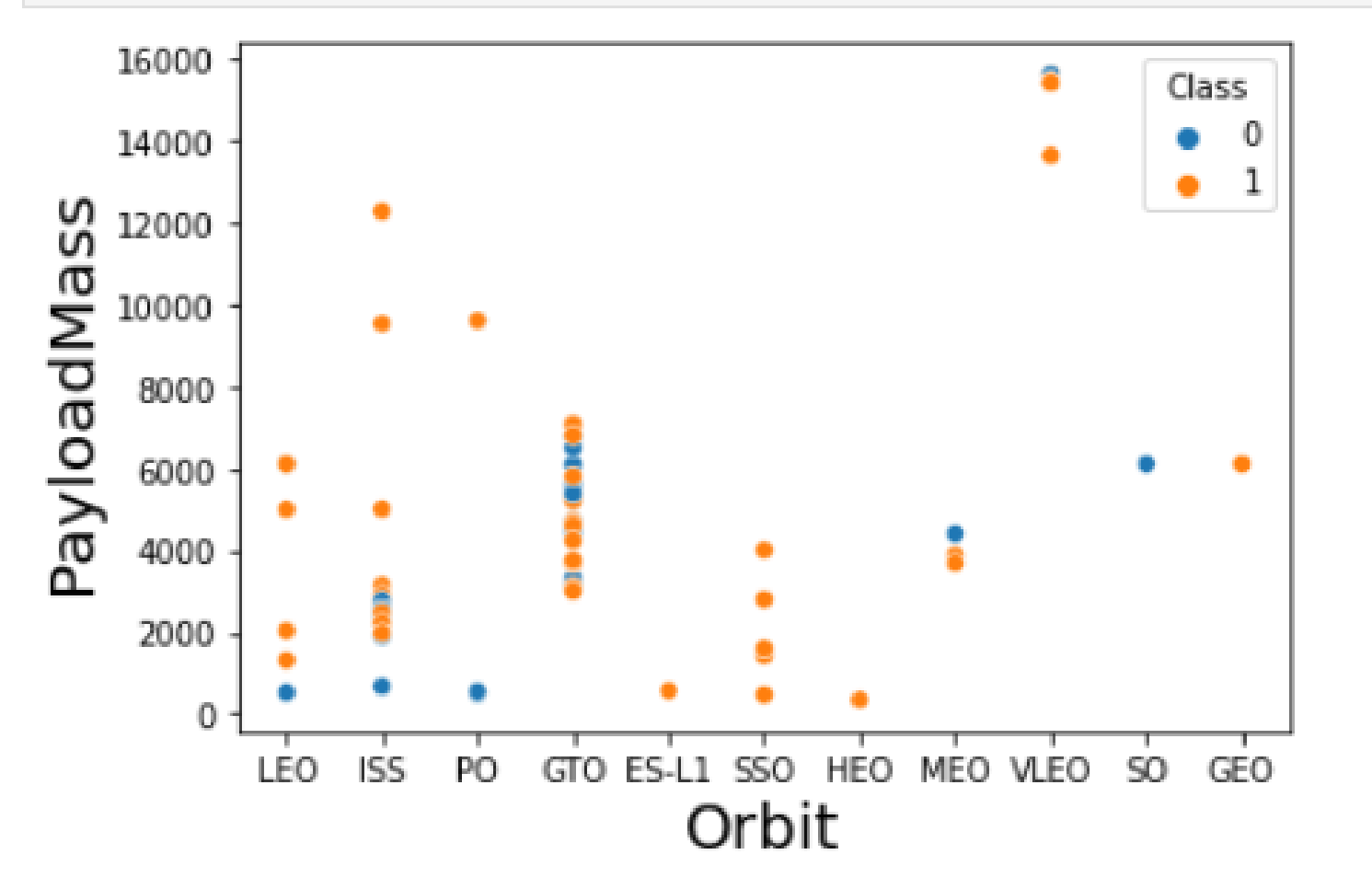


Flight Number vs. Orbit Type



We can observe that the success rate for launches to LEO orbit appears related to the number of flights; whilst there seems to be no relationship between flight number for GTO orbit.

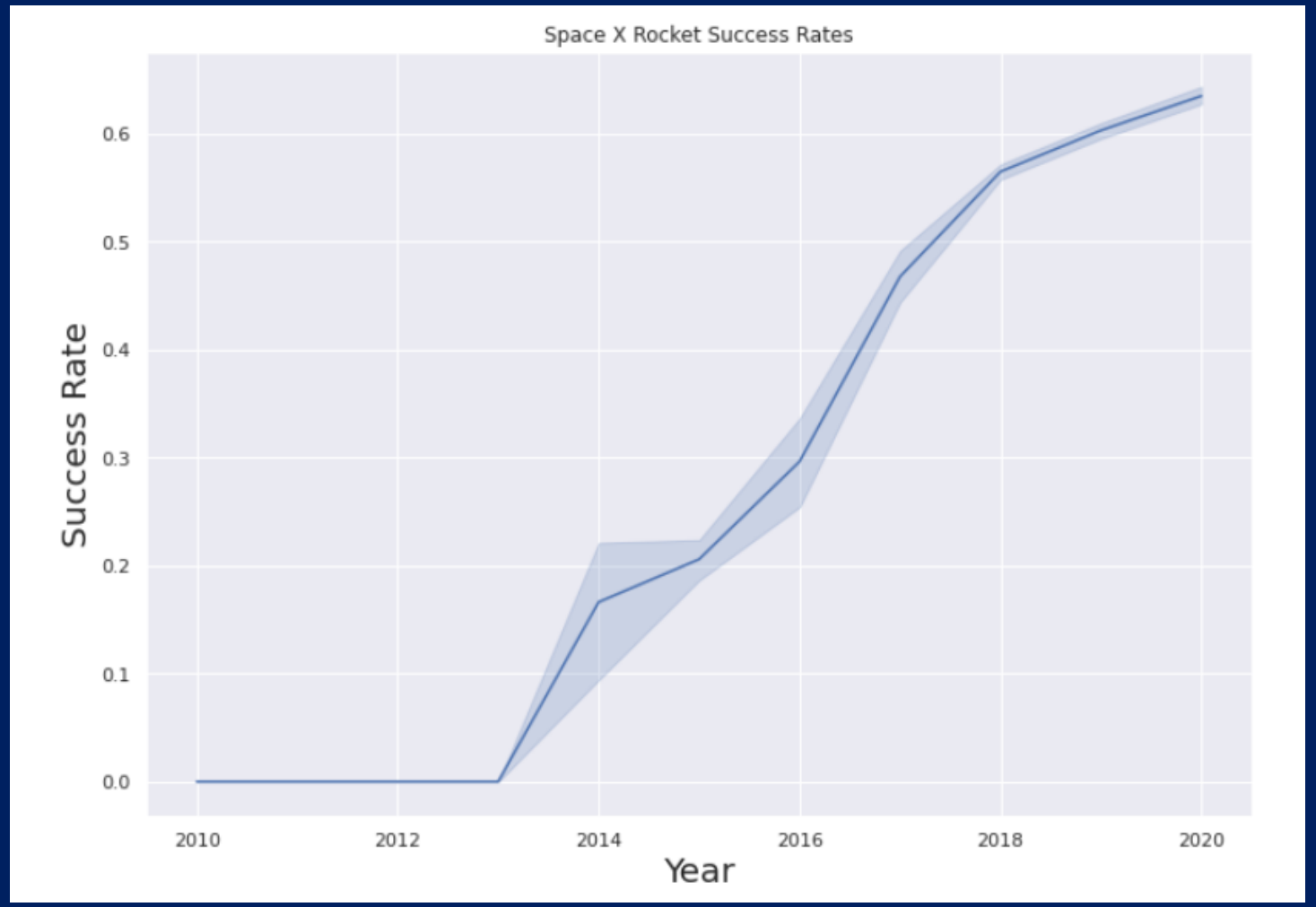
Payload vs. Orbit Type



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

However for GTO there is no clear pattern observed as there were both positive and negative landing at different payload mass.

Launch Success Yearly Trend



We can see that the sucess rate has been increasing since 2013 till 2020

All Launch Site Names

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

```
%sql select DISTINCT LAUNCH_SITE from SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

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

```
%sql select LAUNCH_SITE from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

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

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
45596
```

Average Payload Mass by F9 v1.1

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

```
2928
```

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select MIN(DATE) from SPACEXTBL where LANDING__OUTCOME = 'Success (ground pad)'
```

```
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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, PAYLOAD_MASS__KG_ from SPACEXTBL where LANDING__OUTCOME = 'Success (drone ship)' and PAYLOAD_MASS__KG_ be
```

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%%sql SELECT MISSION_OUTCOME, count(*) as OUTCOMES
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME
```

mission_outcome	outcomes
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

```
%sql select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL
```

```
%%sql select DISTINCT BOOSTER_VERSION, PAYLOAD_MASS__KG_
from SPACEXTBL
where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

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

```
%%sql select LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
from SPACEXTBL
where LANDING__OUTCOME = 'Failure (drone ship)'
and DATE like '2015%'
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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

```
%%sql SELECT LANDING__OUTCOME, COUNT(*)
FROM SPACEXTBL
where DATE between '2010-06-04' and '2017-03-20'
group by LANDING__OUTCOME
order by COUNT(*) desc
```

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

Section 3

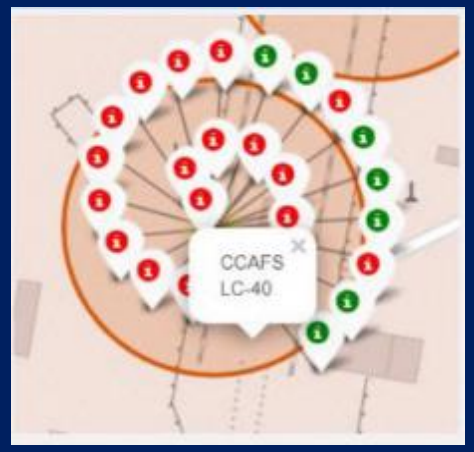
Launch Site Proximities Analysis

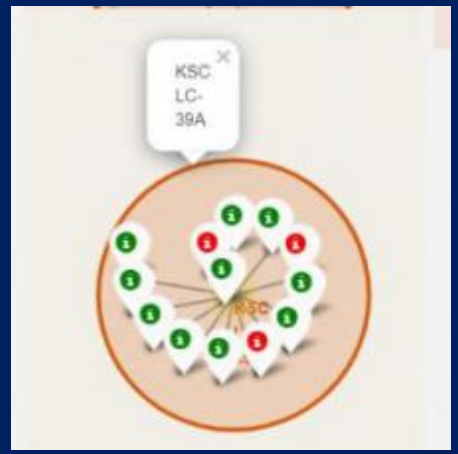
All launch sites on the Map

All of SpaceX launch sites are located in the United States, either in California or Florida.

Success Rates of Different launch sites

Launch site in California



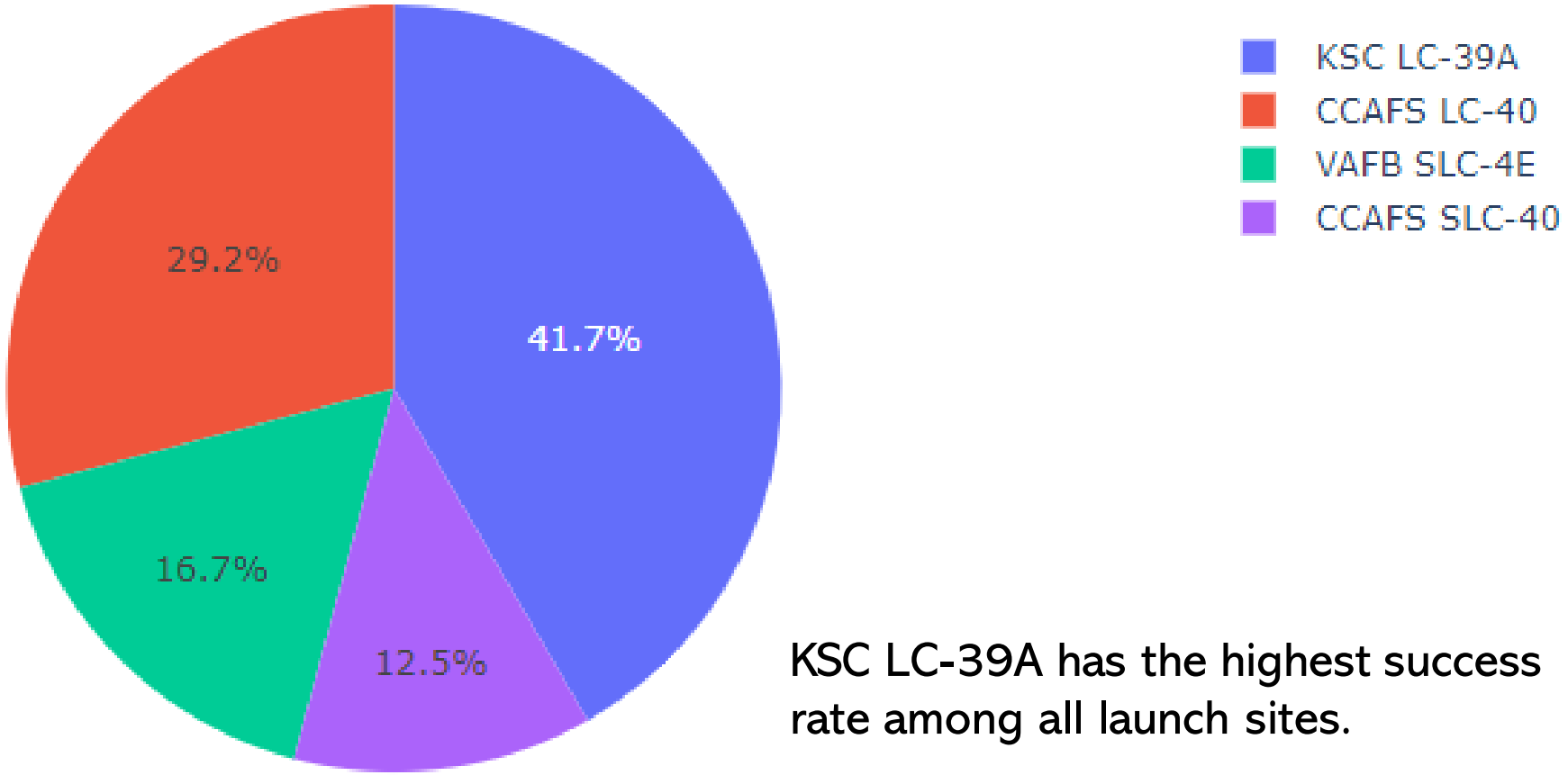


Launch sites in Florida

Section 4

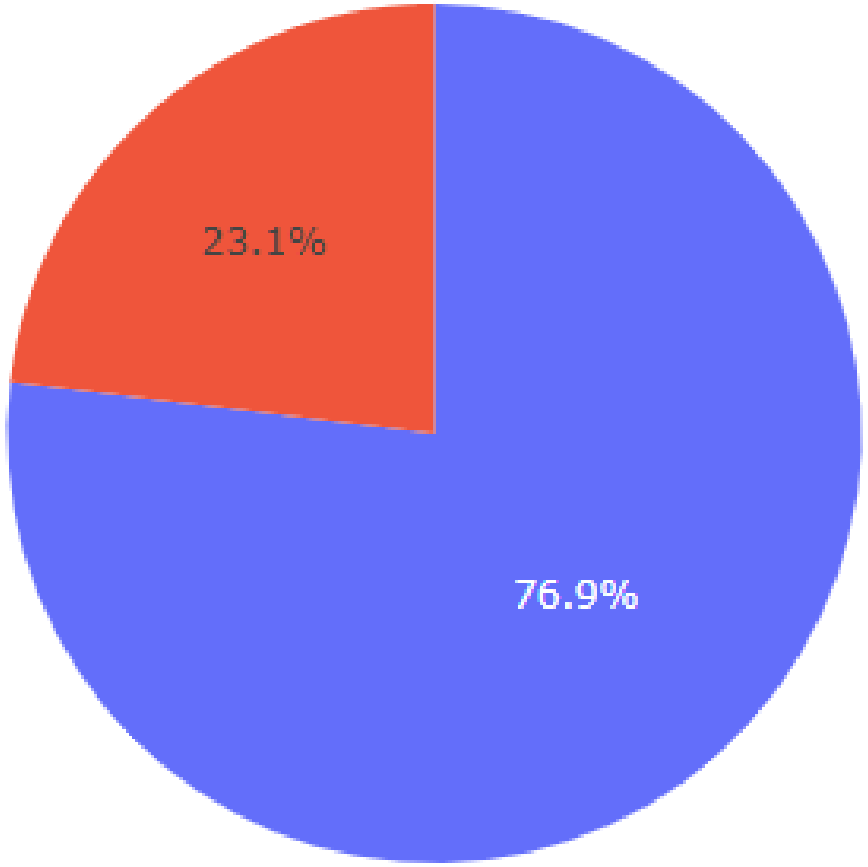
Build a Dashboard with Plotly Dash

Success Rate of All Launch Sites



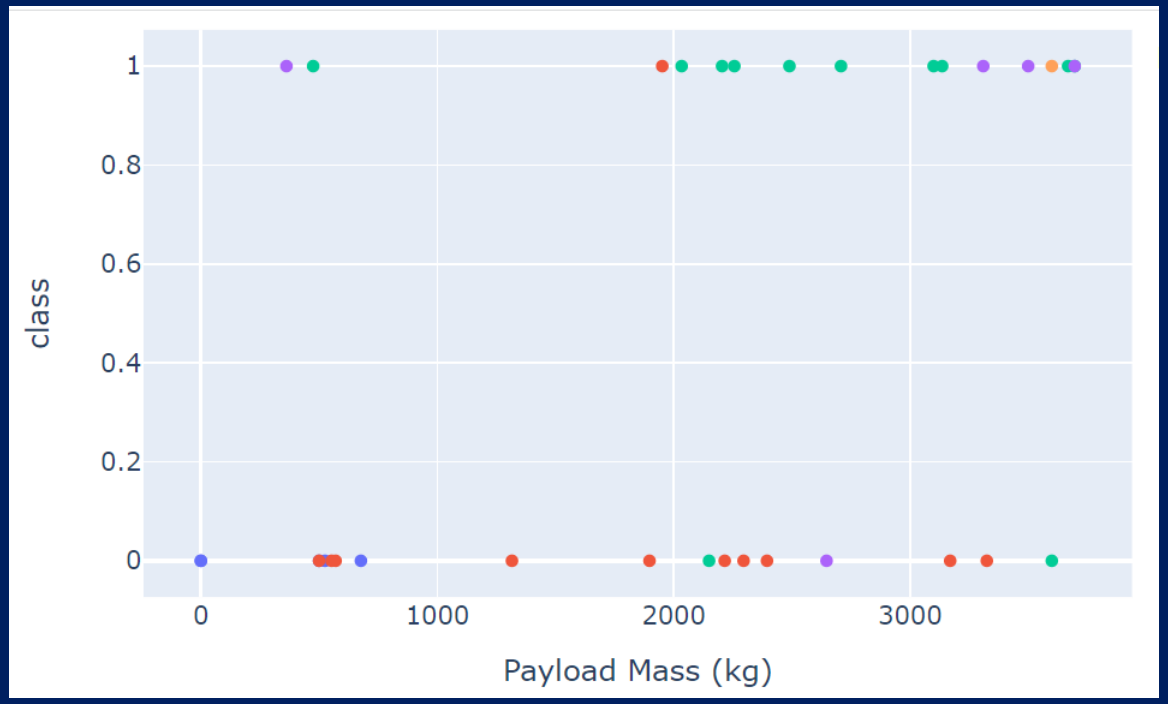
KSC LC-39A has the highest success rate among all launch sites.

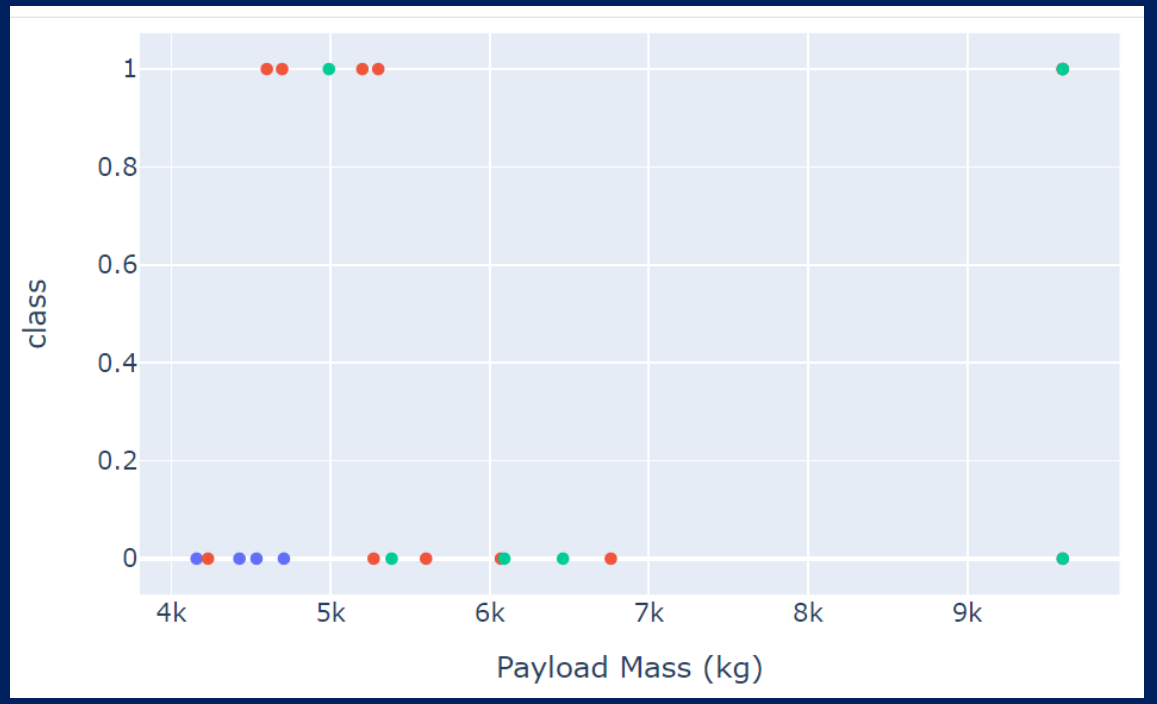
Total Success Launches for Site KSC LC-39A



76.9% of the launches at site KSC LC-39A are successful.

Success Rate of various Payload Range & Booster Version



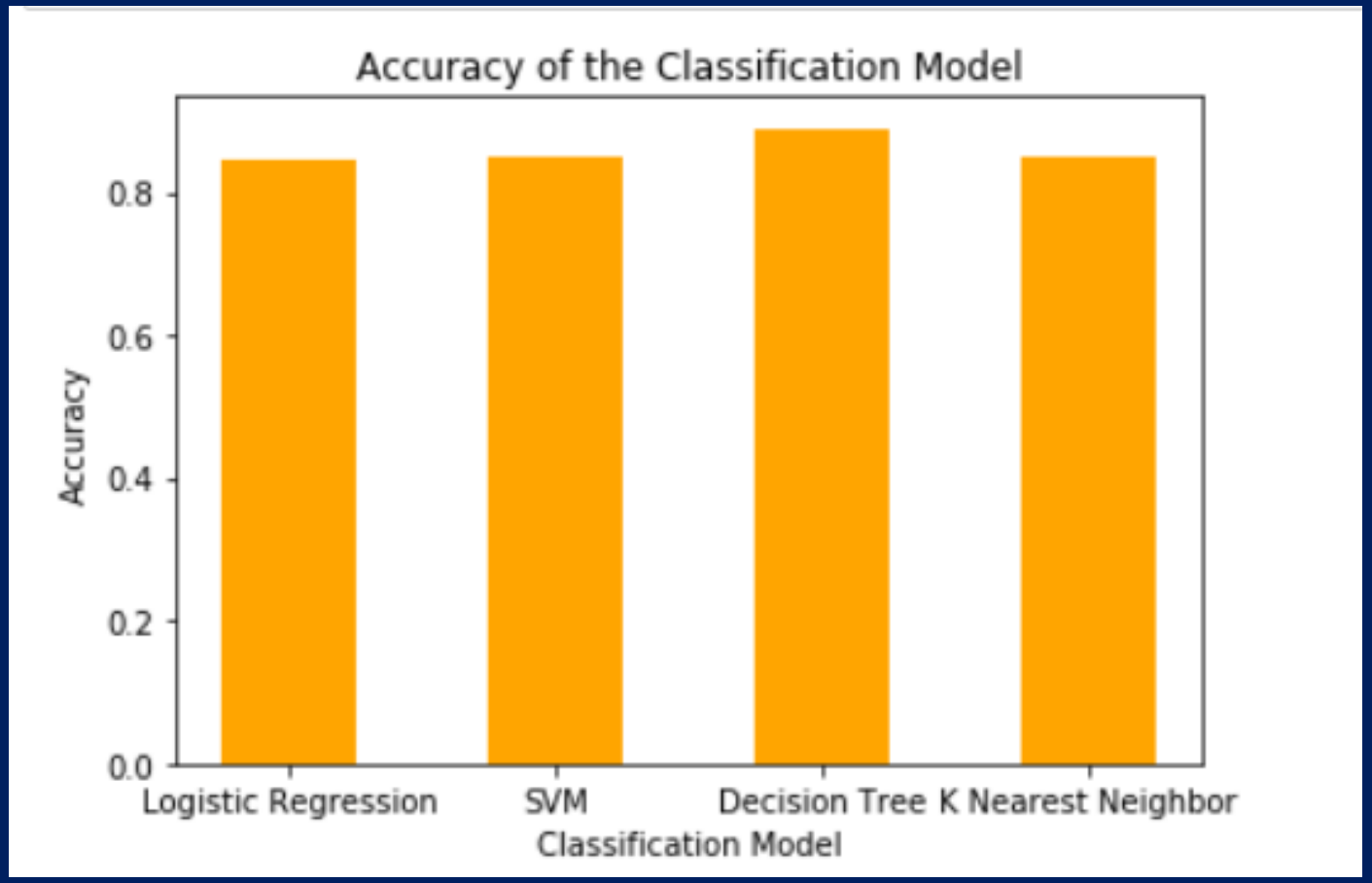


Success rates for rockets with lighter payloads is higher than the ones with heavy payloads.

Section 5

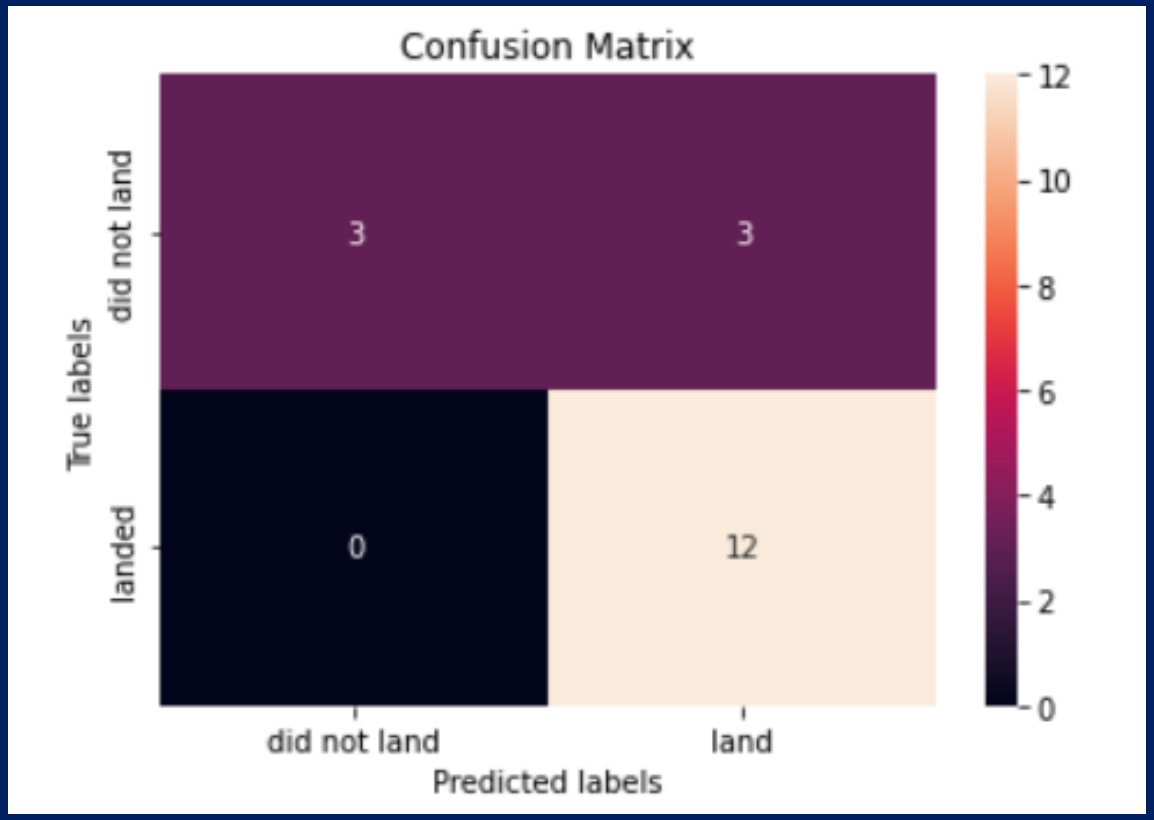
Predictive Analysis (Classification)

Classification Accuracy



The model with the highest classification accuracy is Decision Tree with a score of 0.89.

Confusion Matrix – Decision Tree



We observe that the major problem in our decision tree's performance is the false positives, whereby the predicted landings did not occur in reality.

Conclusions

1) There are various factors that affect the success rate of the first stage landing, such as:
    - The rockets that were launched to orbit ES-L1, GEO, HEO and SSO have the highest success rate to have their first stage landing successfully
    - Launches at site KSC LC-39A has higher rate to have successful first stage landing.
    - Success rate of first stage landing increases as the year increases. Thus, rocket launches will eventually be more affordable and accessible to businesses and people.

2) The best model to predict the success rate of landing is Decision Tree, which has the highest classification accuracy.


IBM Developer
SKILLS NETWORK
Thank You!