IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

ALEXANDER TAMAYO
01/20/2024

Outline

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

Executive Summary

- Summary of methodologies

  - Web scrapping

  - Wrangling

  -EDA

  -Machine learning

- Summary of all results

  -EDA visualizations

  -Predictions using machine learning

Introduction

- Project background and context

Space Y promotes Falcon 9 rocket liftoffs on its official site with a price tag of 62 million dollars; alternative service providers charge over 165 million dollars per launch, with significant savings attributed to Space Y's ability to recycle the initial stage. Consequently, by assessing the predictability of the initial stage's landing, we can estimate the launch cost. This data becomes valuable when a competing firm wishes to submit a bid for a rocket launch against Space Y. The primary objective of this initiative is to establish a machine learning framework for forecasting the success of the initial stage landing.

- Problems you want to find answers

What elements influence the likelihood of a triumphant rocket landing?

The interplay among various factors that contribute to the success likelihood of a landing.

What operational parameters must be established to guarantee a prosperous landing program?

Section 1

Methodology

Methodology

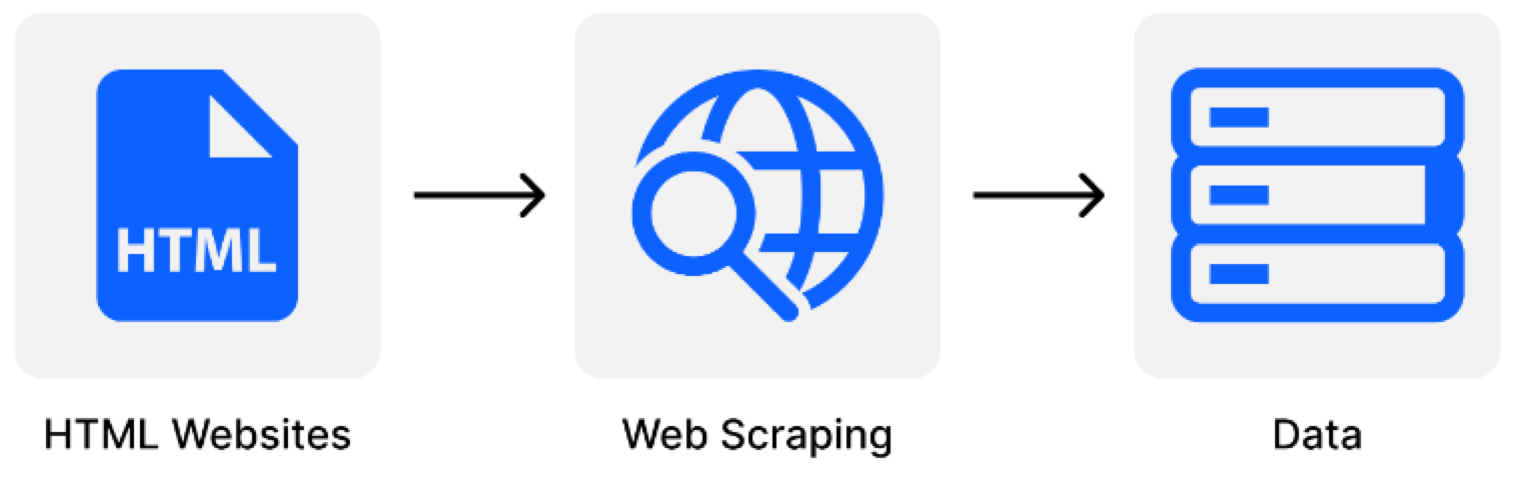
Executive Summary

- Data collection methodology:
  - Data was collected using web scraping in the next link: https://api.spacexdata.com/v4/rockets/
- Perform data wrangling
  - Data was processed using different techniques like dummy variables and more.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different models were using in order to analyze which one had better results.

Data Collection

- Describe how data sets were collected.

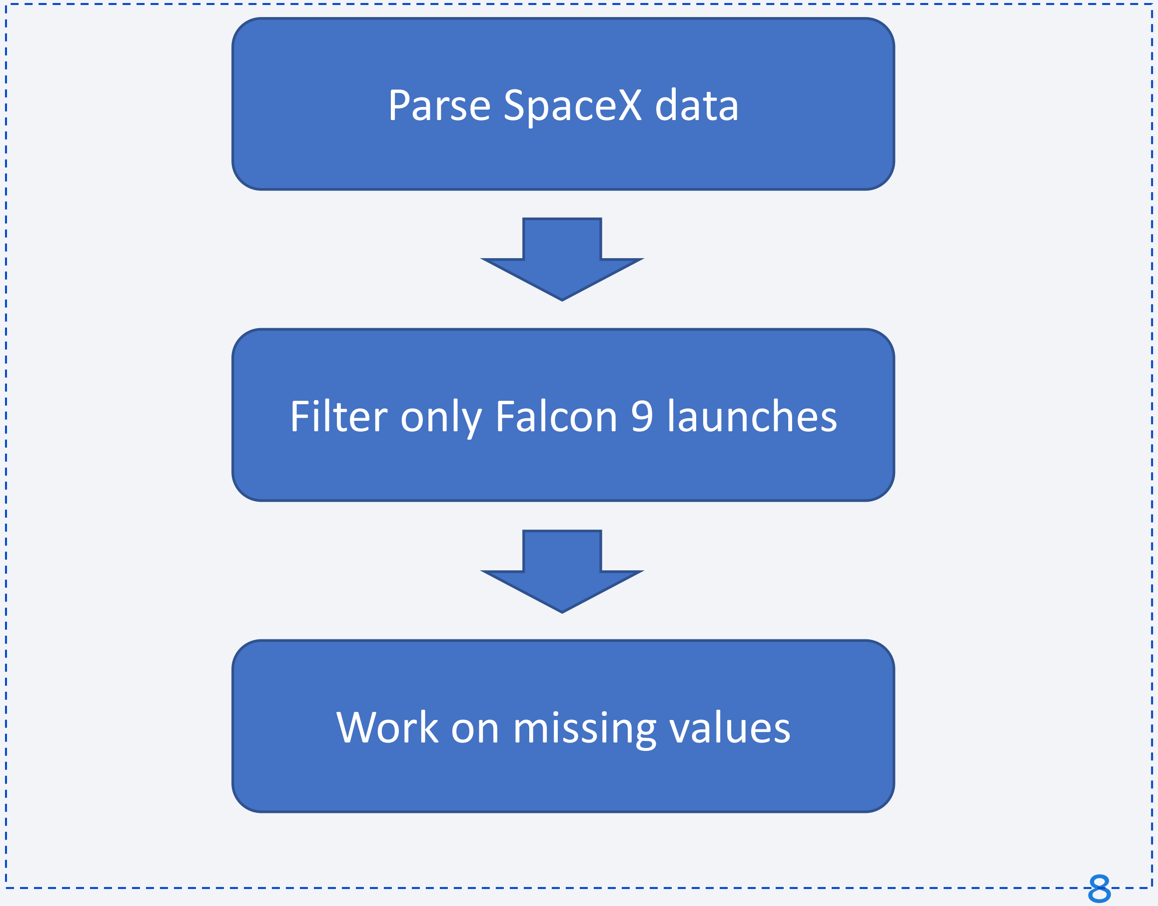
Datasets were collected using api’s and web scrapping.



Data Collection – SpaceX API

- Source code:

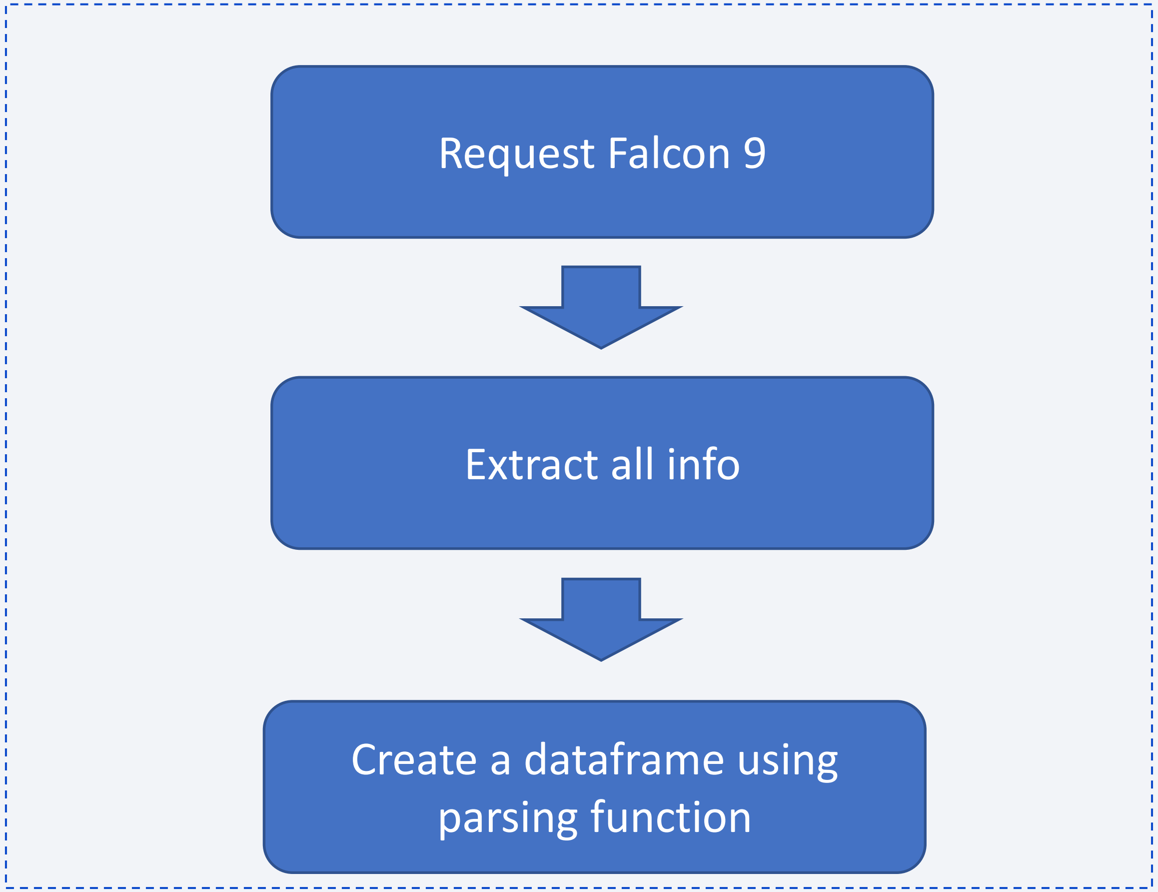
https://github.com/AlexanderTamayo/space_x_final_project/blob/main/1-jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

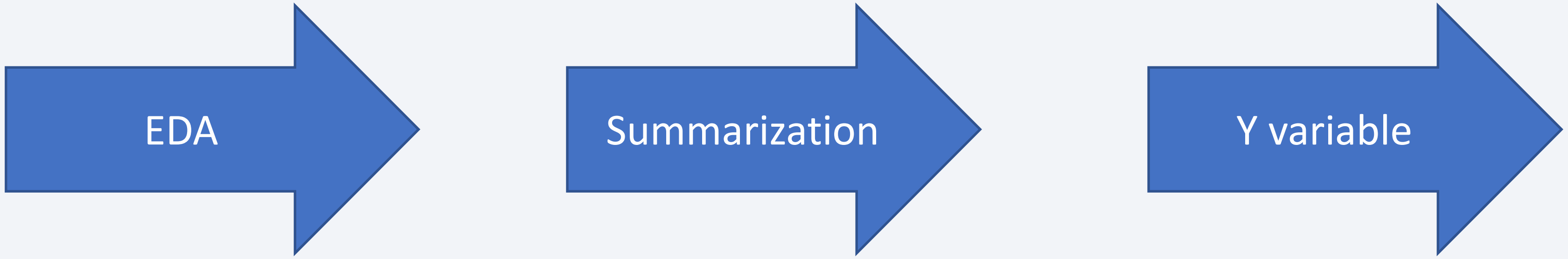
- Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/2-jupyter-labs-webscraping.ipynb



Data Wrangling

- Explore the Data
- Summarization of launcher per launch site
- Create the variable y called class in order to know if landing was successful
- Source code: https://github.com/AlexanderTamayo/space_x_final_project/blob/main/3-labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Distinct launch sites
- CCA launch sites
- Payload mass carried by boosters from Nasa
- Average payload mass carried by F9 boosters
- Date of the first successful launched
- Name of boosters that have succeed in launching which have between 4,000 and 6,000 kg in payload mass
- Total of successful and failure missions
- Failed landing outcomes in drone ship
- Rank of the counts of landing outcomes

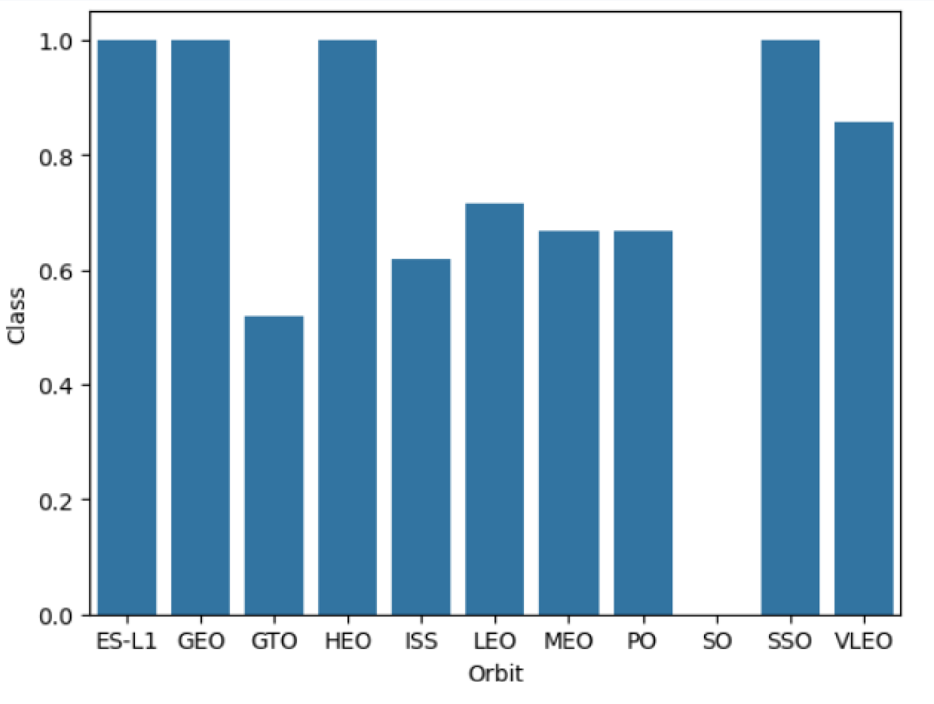
Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/4-jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- Exploration and correlations of data
- Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/5-jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb



Build an Interactive Map with Folium

- Distance from launching sites to railroads
- Markers in the country map to visualize where the launches occur
- Visualization of failing launches

I used these elements to easily understand info in the map.

Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/6-lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Interactive dashboard for visualization of data
- Pie charts for categorization
- Scatter plot for x and y variables
- Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/7-Dash-lab_jupyter_launch_site_location.jupyterlite.ipynb

Predictive Analysis (Classification)

- Load data from df and split it using the function train test split
- Standardize the variables x
- Train and fit different models such as Decision Tree, SVM and Logistic Regression
- Choose the best model based on the results
- Source code:

https://github.com/AlexanderTamayo/space_x_final_project/blob/main/8-SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results

- SpaceX operates from a quartet of distinct launch facilities, with the inaugural launches catering to both SpaceX and NASA
- The F9 v1.1 booster registers an average payload of 2,928 kg
- The initial triumphant landing transpired in 2015, a half-decade subsequent to the maiden voyage
- Several iterations of the Falcon 9 booster have achieved successful landings on autonomous drone ships, exceeding the stipulated average payload
- Nearly all mission endeavors boast a flawless track record, with success rates approaching 100%
- Over the years, the proficiency of landing outcomes has exhibited a marked improvement

Results

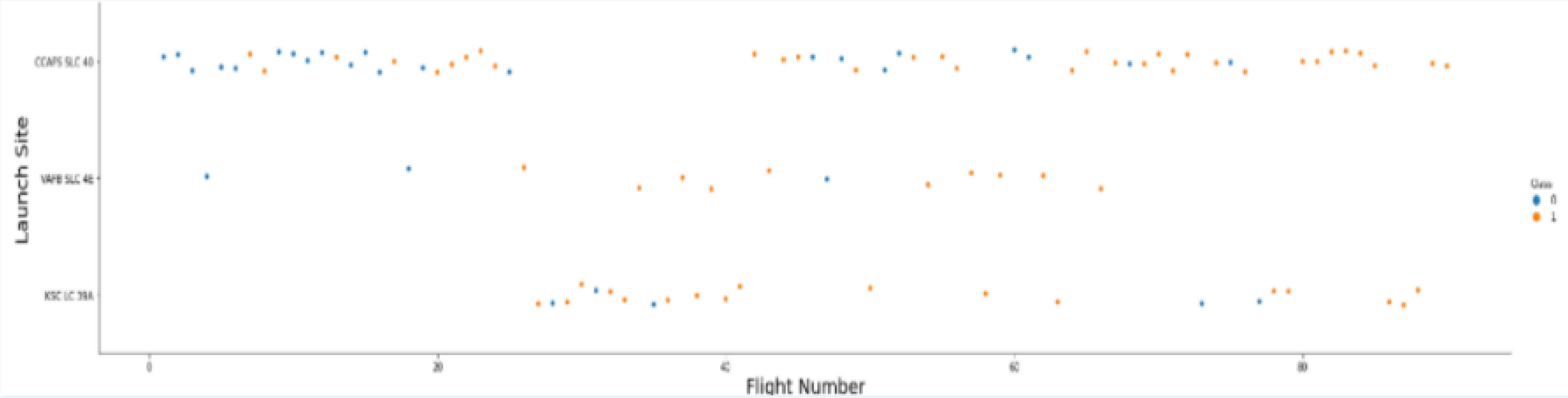
- Coasts are prefered to be the locations of launchings
- Any launch happend so close to a city

Section 2

Insights drawn from EDA

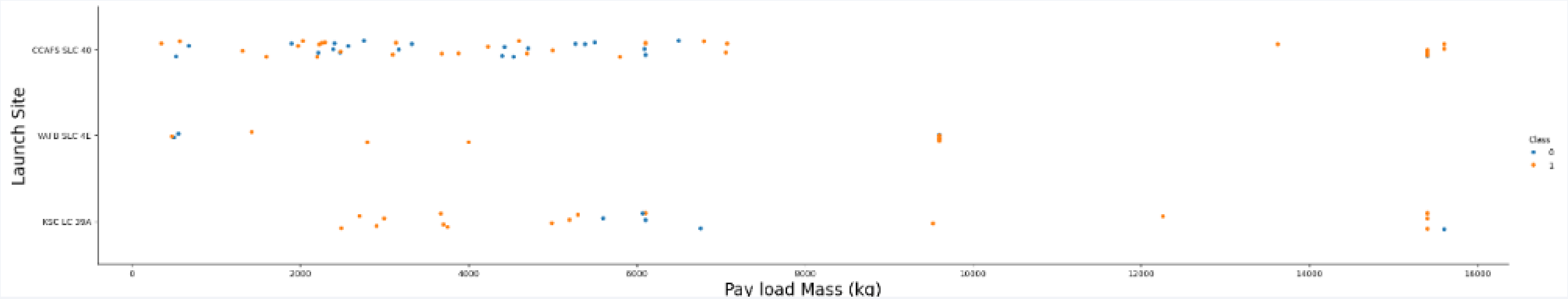
Flight Number vs. Launch Site

- The larger the flight amount the better



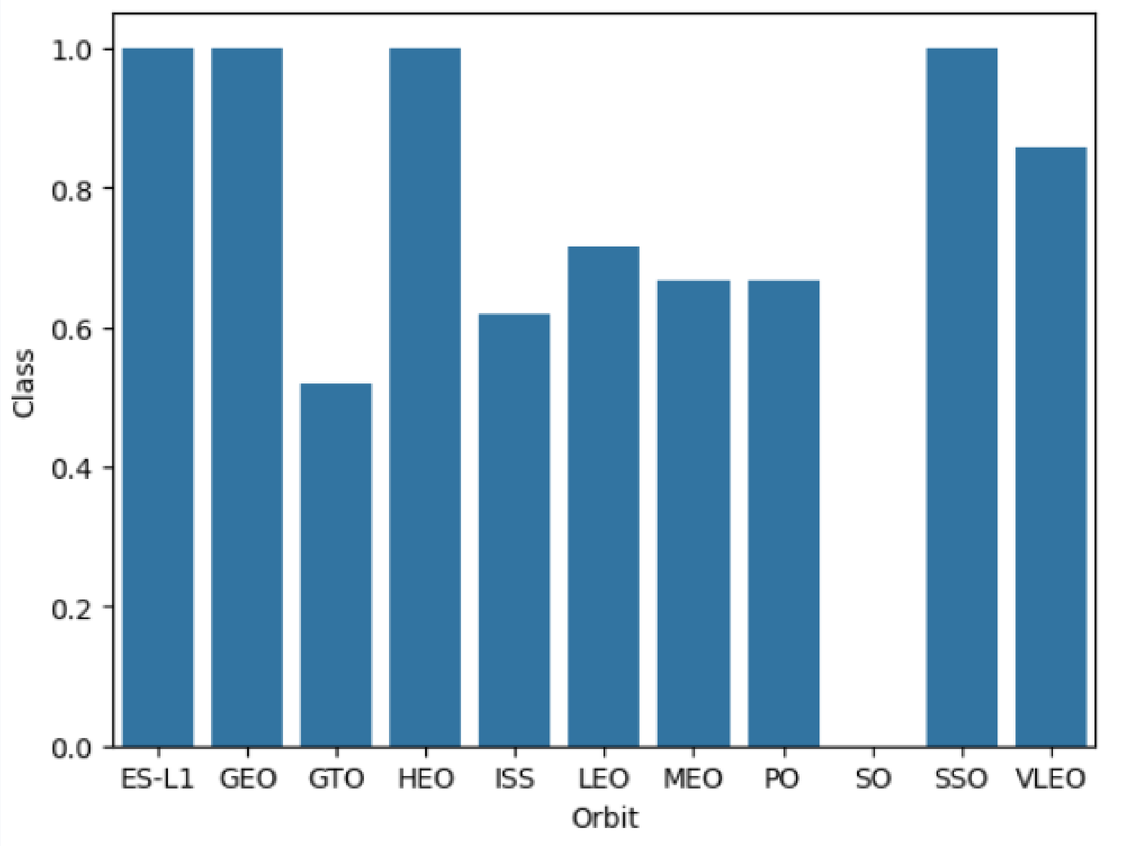
Payload vs. Launch Site

- The more payload mass the less probable to happen any accident.



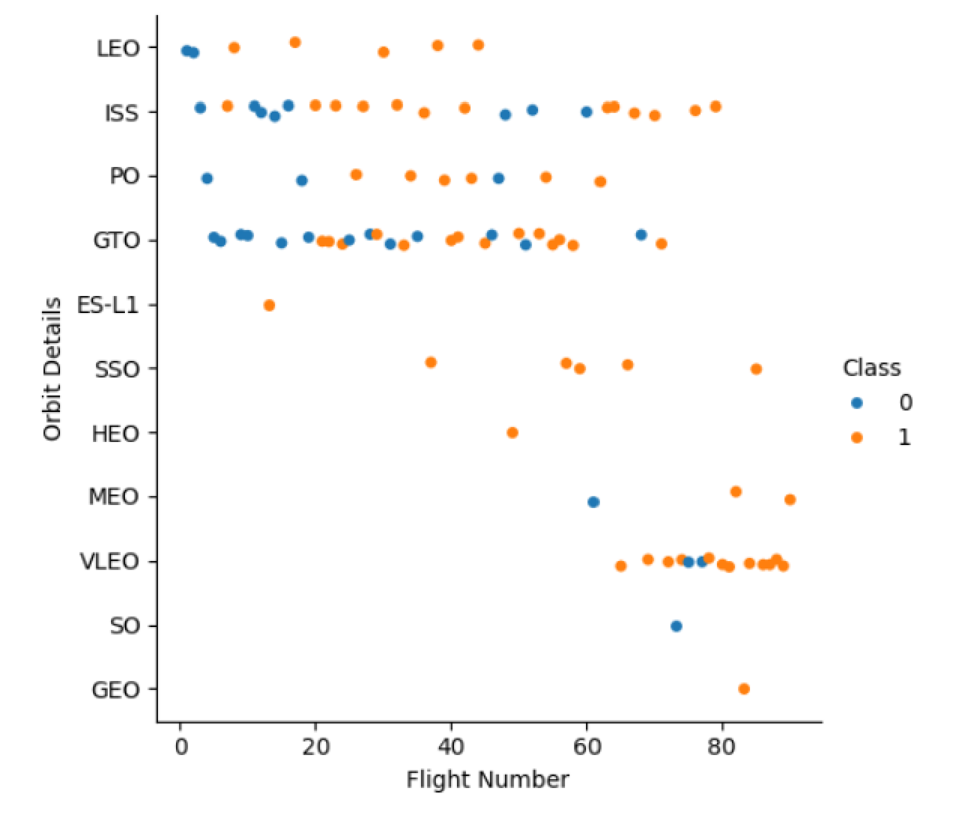
Success Rate vs. Orbit Type

- Orbits like ES-L1 and GEO are one of the best to make success able missions.



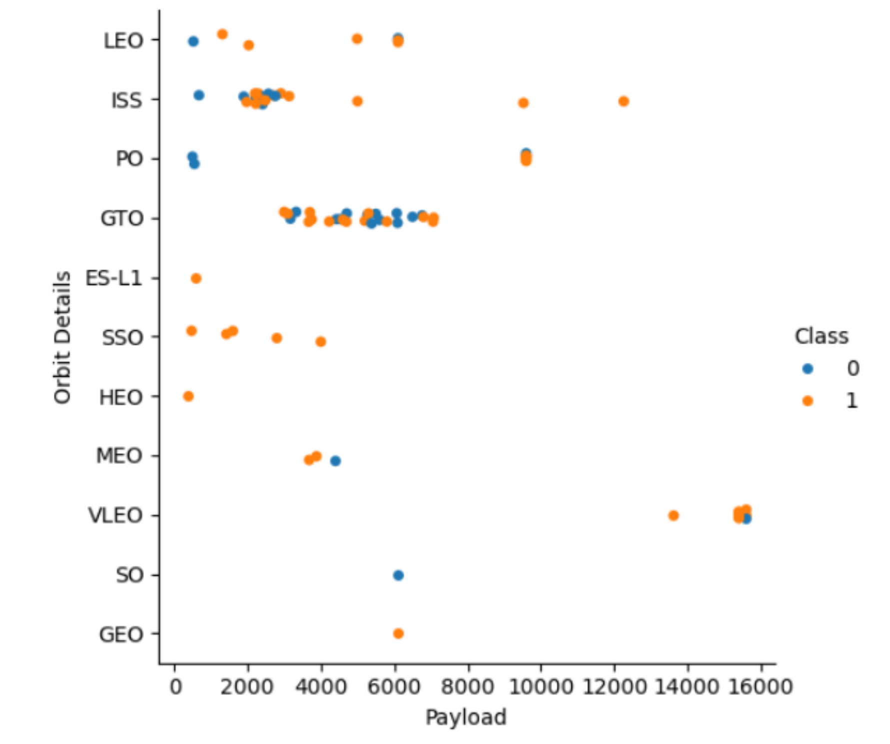
Flight Number vs. Orbit Type

- It looks like LEO is the orbit type with the better amount of success landings.



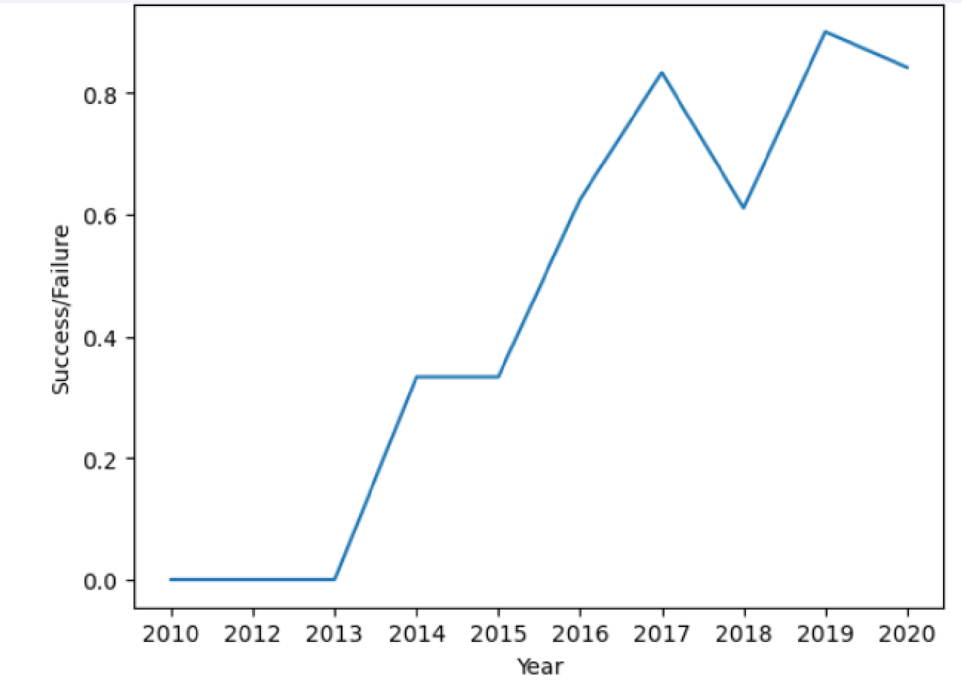
Payload vs. Orbit Type

- SSO has good results with low payload mass.



Launch Success Yearly Trend

- We can appreciate that through the years the landing success rate has been improving.



All Launch Site Names

- There are four launch site names:

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

Launch Site Names Begin with 'CCA'

- Using sql I was able to visualize five registers where the launch site started with “CDA” and be able to visualize some info about it.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Out
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parac
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 (parac
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No att
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No att
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No att

Total Payload Mass

- Also using sql I was able to determine that the total payload mass is 45,596 kg

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as TOTALPLM FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

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

TOTALPLM
45596

Average Payload Mass by F9 v1.1

- And average payload mass is 2928.4 kg

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Exploring the data I realized that the firs date when a landing outcome was successful was in 2015-12-22.

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

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Through exploration I could visualize the drone ship landing with payload mass between 4,000 and 6,000.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- As we can see the amount of failing landing is pretty low.

List the total number of successful and failure mission outcomes

```
%sql SELECT(SELECT COUNT(*) FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success') AS Success,(SELECT COUNT(*) FROM SPACE
```

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

Success	Failure
38	3

Boosters Carried Maximum Payload

- We can visualize what booster have been able to carried maximum payload mass.

```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)FROM SPACEX
```

```
 * 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

2015 Launch Records

- Only two launches failed in 2015, both of them were drone ship.

```
%sql SELECT SUBSTR(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE SUBSTR(Date
```

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

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

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

- This info may be the most relevant because we can get to know exactly the total outcome of each try.

Landing_Outcome	OutcomeCount
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

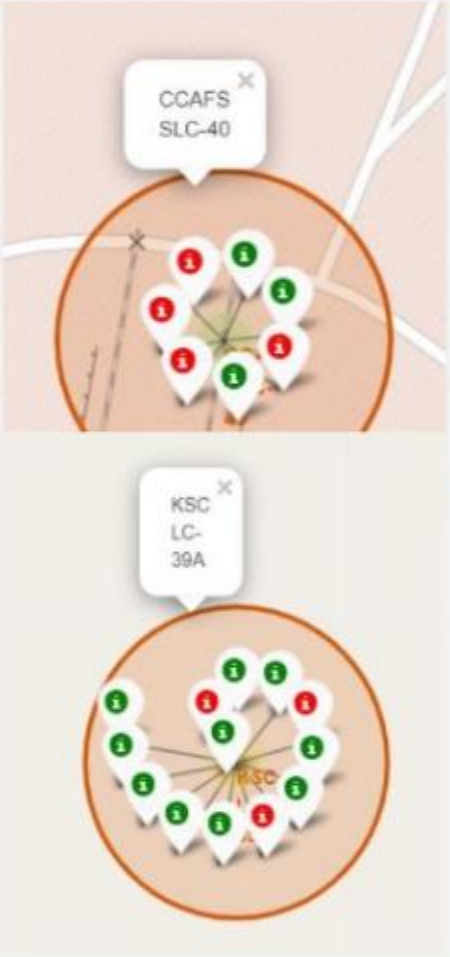
Section 3

Launch Sites Proximities Analysis

Launches locations

- Launchings have only happened in America

Important Locations



Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

37

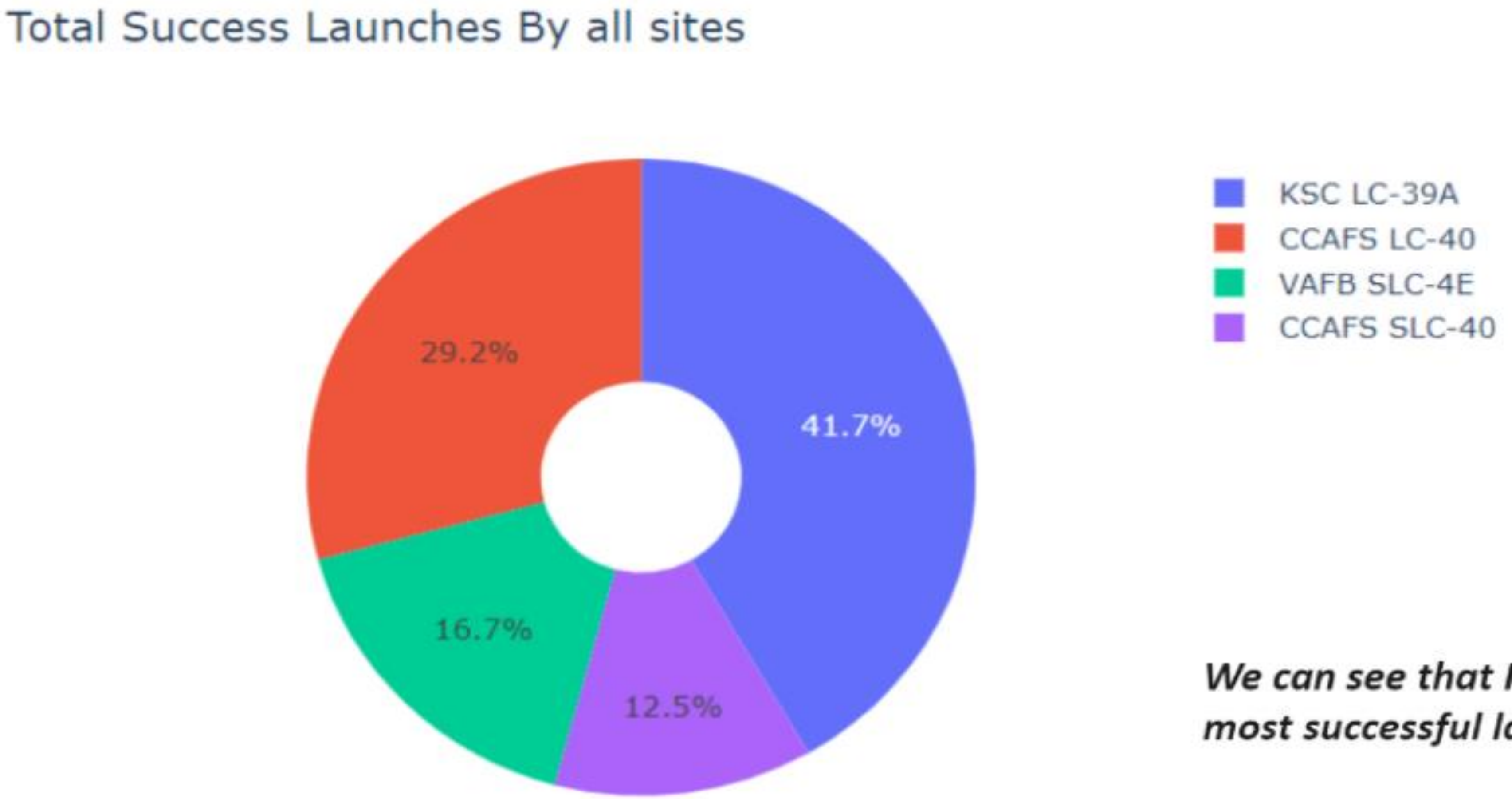
Distance of launchings to different sites

- Launchings are usually execute close to coasts.

Section 4

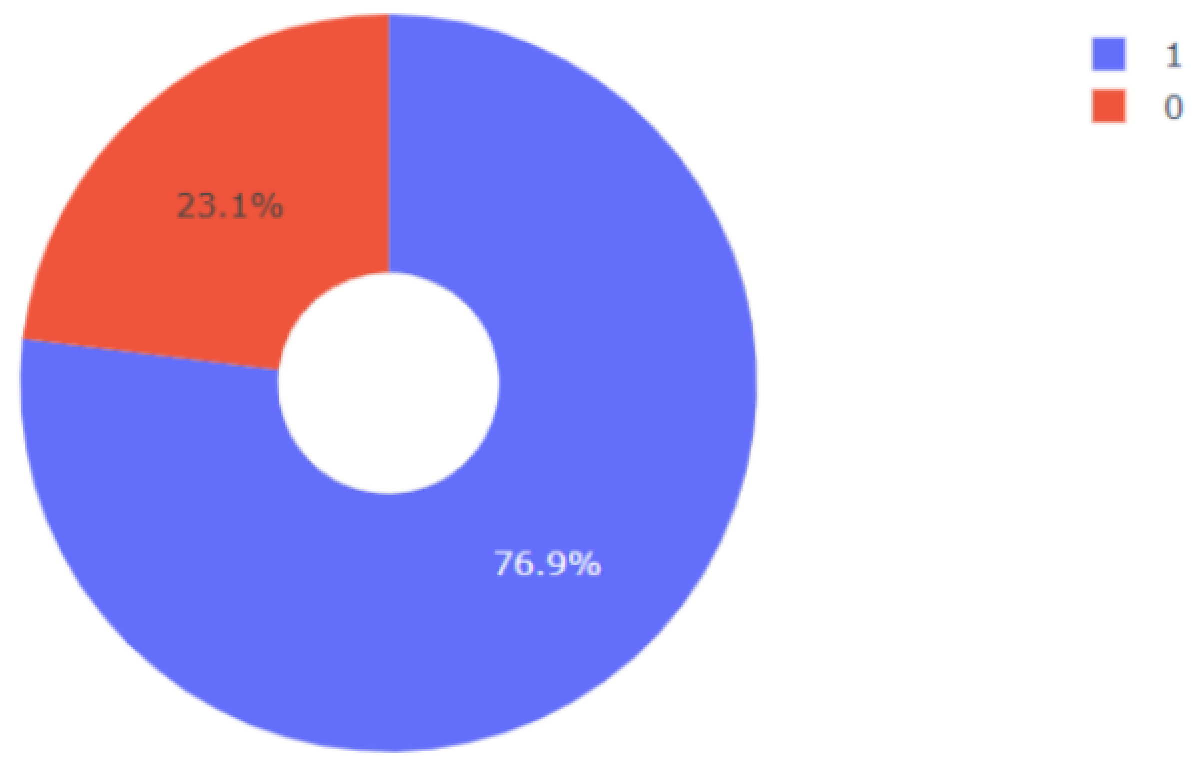
Build a Dashboard with Plotly Dash

Successful Launches by Site



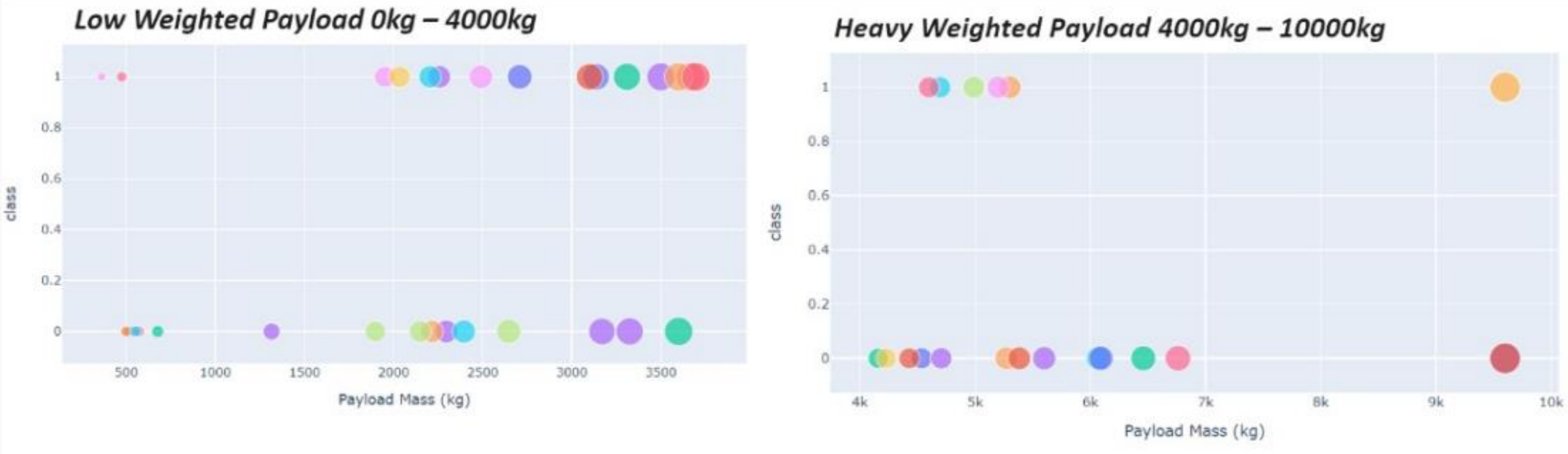
We can see that KSC LC-39A had the most successful launches from all the sites

Best Launch Site



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Comparison between Payload and Launch Outcome

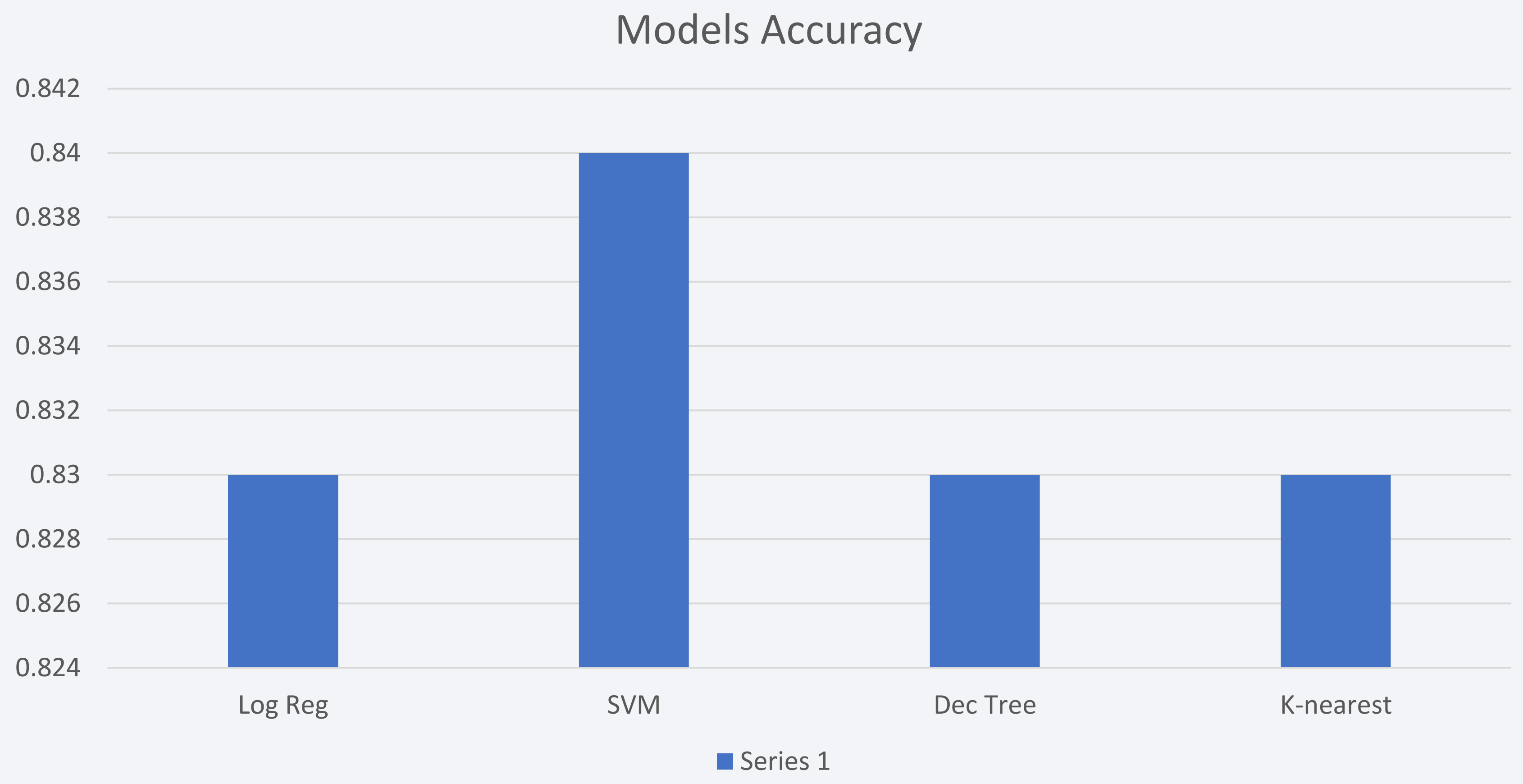


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

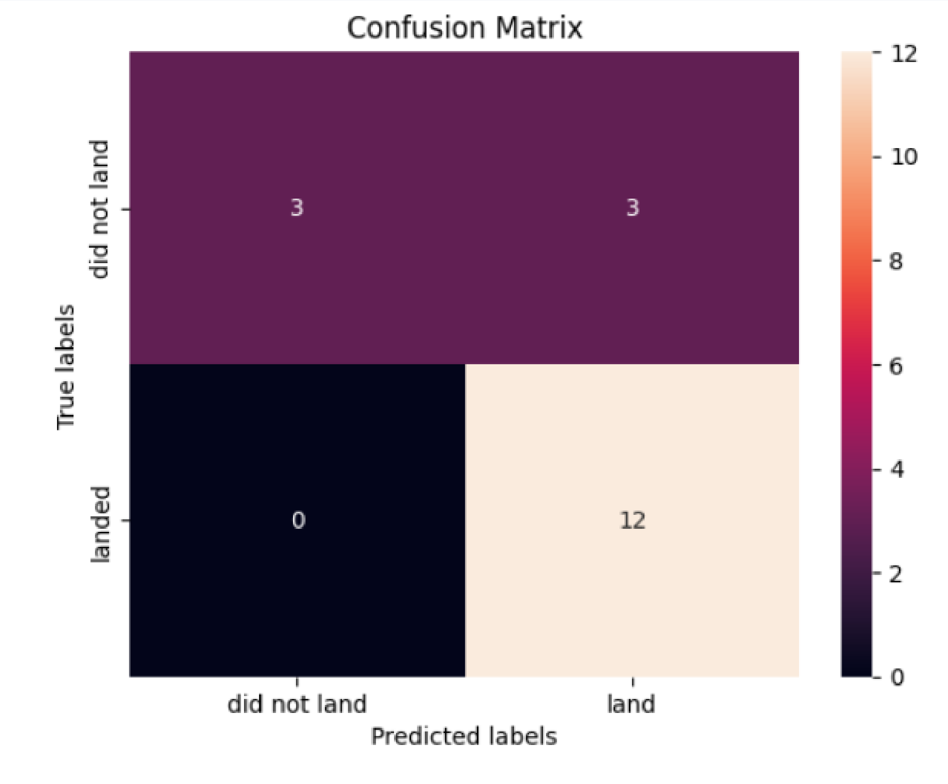
- SVM is the best model to be used and we can visualize the result of model training:

TP: 3

TN: 12

FP: 3

FN: 0



Conclusions

- Inferences can be drawn as follows:
  - A positive correlation seems to exist between the frequency of flights at a given launch site and the corresponding success rate; higher flight volumes tend to coincide with increased success rates.
  - The trend analysis reveals a consistent uptick in launch success rates from 2013 through 2020.
  - Orbits designated as ES-L1, GEO, HEO, SSO, and VLEO exhibit notably high success rates.
  - KSC LC-39A stands out as the launch site with the most triumphs, surpassing other locations in terms of successful launches.

Thank you!