IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Richard Lee
September 23, 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
    - Data Collection using Webscraping and REST API queries
    - Data Wrangling to Classify Lauches based on Success and transform data into standardized numeric form
    - Exploratory Data Analysis using SQL and Visualization packages for Python
    - Interactive Plotly Web App to visualize payload and success launch data at each Launch Site
    - Exploring Launch Sites using interactive Folium Maps
    - Predictive analysis for classification of Rocket Landing Success
- Summary of all results
    - Exploratory Data Analysis Results
    - Predictive Analysis Results

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 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. Thus it is advantageous to be able to predict whether the Falcon 9 first stage will land successfully for new missions.
- To make valuable predictions we must solve the following:
  - What factors of a mission influence Falcon 9 launch success?
  - What conditions must be met by SpaceX to ensure the highest probability of success for a given mission?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Requested past launch data from SpaceX's Rest API https://api.spacexdata.com/v4
  - Webscraped tabular data on SpaceX launches from
- Perform data wrangling
  - Dropped data on non Falcon 9 launches
  - Used One Hot Encoding to transform categorical variables into factors
  - Transformed factors to integers
  - Replaced missing numerical data for payload masses with the sample mean
  - Classified data as 1 for Successful Landing or 0 for failed landing

Methodology

Executive Summary

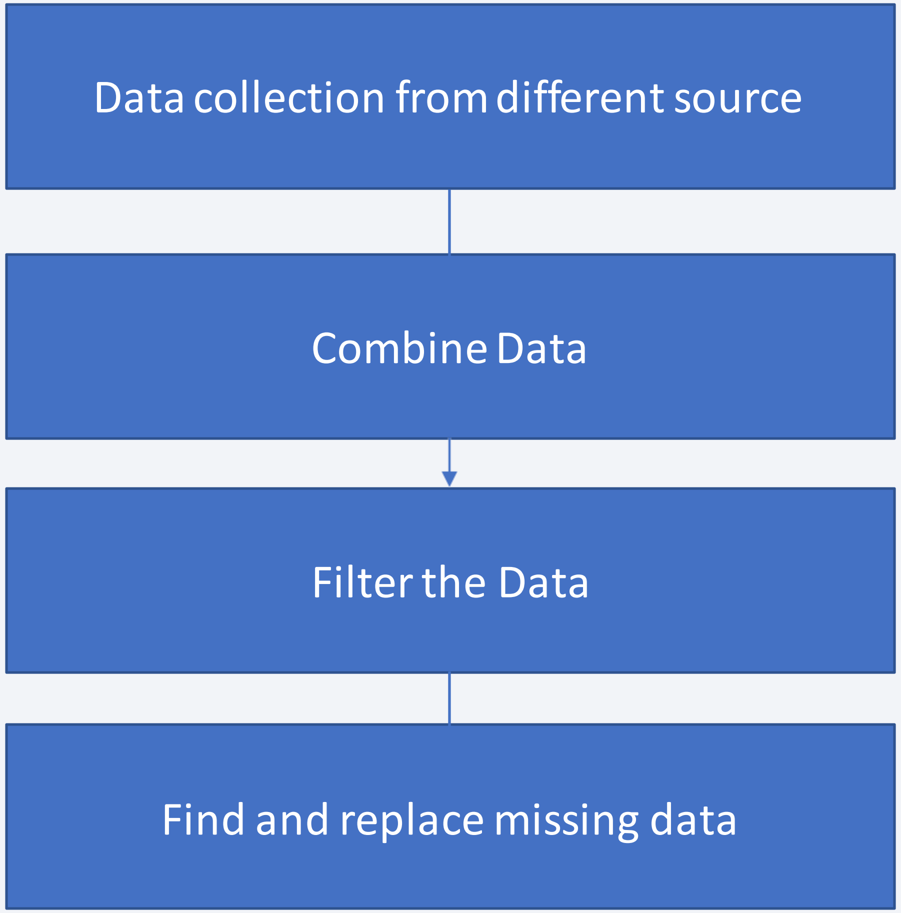
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Used Scatter Plots and Bar Graphs to visualize relationships between variables
  - Used SQL Queries to understand the data collecte
- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive Plotly Web App to visualize payload and success launch data at each Launch Site
  - Exploring Launch Sites using interactive Folium Maps
- Perform predictive analysis using classification models
  - Built and tuned multiple classification models to predict landing success
  - Used Grid Search and Cross Validation to find the best model parameters for each model tested (Logistic, SVM, Decision Tree, and KNN)
  - Split Data into testing and training to test model accuracy resilience on Out of Sample Data
  - Selected top performing Model on both testing and training set based on accuracy of predictions
  - How to build, tune, evaluate classification models

Data Collection

- Used SpaceX REST API to gather data on rocket launches:
  - https://api.spacexdata.com/v4
- API provides data on rockets used, launch dates, payload masses, launch success or failure, launch site name and location (latitude and longitude), booster version (note for this experiment we are only interested in Falcon V9 boosters), landing outcome, etc. (47 columns of data for each launch in total)
- Our Goal is to Predict the Landing outcome using the other variables
- Falcon 9 launch data was also collected via Webscraping Wikipedia using BeautifulSoup from the page below:
  - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection - SpaceX API

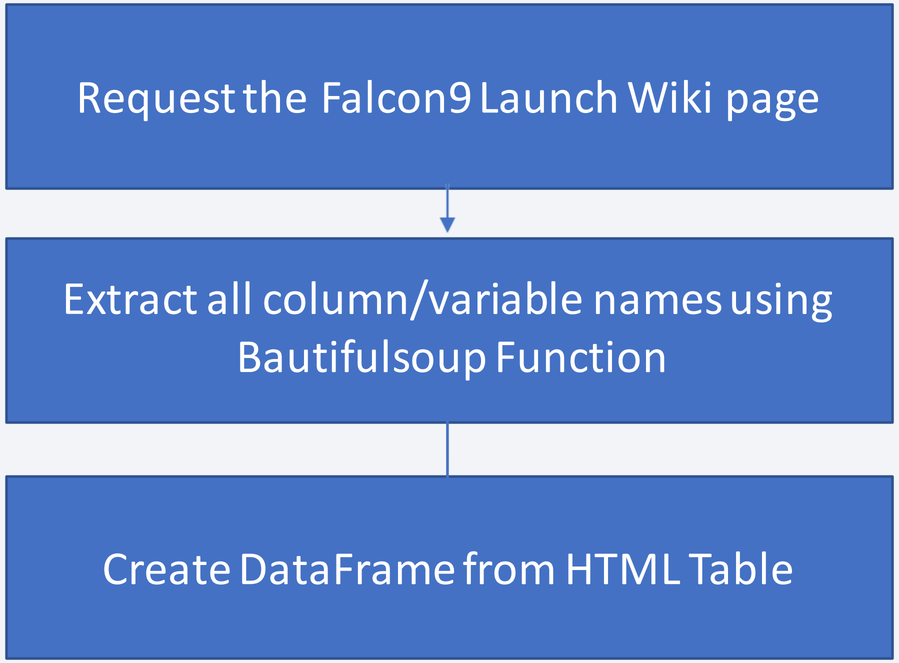
- Calling Rest API we will collect data of Rocket, Launchpad, Payloads, Landing Outcome and pas history of Launches
- Combine the data in single DataFrame
- Filter the data for Falcon 9 Launches
- Find the missing data
- Replace the missing data with mean.



Data Collection - Github Link - Click Here

Data Collection - Scraping

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



Web Scraping - Github Link - Click Here

Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad. True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.
- To convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful

[Data Wrangling - Github Link - Click Here]

EDA with Data Visualization

We will create different charts for Exploratory Data Analysis and Feature Engineering

- Caterplot to plot Filghtnumber Vs. Launchsite
- Scatterplot to plot Payloadmass Vs. Launchsite
- Barchart for the success rate of each orbit
- Scatterplot to plot FlightNumber Vs. Orbit
- Scatterplot to plot Payload Vs. Orbit
- Linechart to Visualize the launch success yearly trend

Data Visualization - Github Link - Click Here

EDA with SQL

SQL queries performed

- %sql Select distinct Launch_site from SPACEXTBL
- %sql select * from spacextbl where Launch_site like 'CCA%' limit 5
- %sql select sum(PAYLOAD_MASS__KG_) from spacextbl where customer = 'NASA (CRS)'
- %sql select avg(PAYLOAD_MASS__KG_) from spacextbl where Booster_Version = "F9 v1.1"
- %sql select min(substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2)) as MinDate from spacextbl where "Landing _Outcome" = "Success (ground pad)"

[EDA with SQL - Github Link - Click Here]

EDA with SQL

SQL queries performed

- %sql select distinct Booster_Version from spacextbl where "PAYLOAD_MASS__KG_" between 4000 and 6000 and "Landing _Outcome" = 'Success (drone ship)’
- %sql select "Mission_Outcome", count(*) as Result from spacextbl group by "Mission_Outcome"
- %sql select Booster_Version, sum(PAYLOAD_MASS__KG_) from spacextbl as a group by Booster_Version having sum(PAYLOAD_MASS__KG_) in (select sum(PAYLOAD_MASS__KG_) as max_payload from spacextbl group by Booster_Version order by 1 desc limit 1)
- %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" = 'Failure (drone ship)'
- %sql select RANK() OVER (ORDER BY Date desc) AS 'Rank', * from spacextbl where substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2) between '20100604' and '20170320’ and "Landing _Outcome" like '%Success%'

[EDA with SQL - Github Link - Click Here]

Build an Interactive Map with Folium

- We have created
  - Markers - to point out the launchsite
  - marker Cluster - To point Successrate for each sites
  - Circles - To highlight Circle area
  - Lines - To draw a line to highlight two different points/location
  - Mouseposition - to get coordinates for a mouse over point

Interactive Map with Folium - Github Link - Click Here

Build a Dashboard with Plotly Dash

We have created interactive object as per below to interact with data and find some insights from the data

- Dropdown Menu
- PieChart
- Sliderbar
- Scatterplot

Dashboard with Plotly Dash - Github Link - Click Here

Predictive Analysis (Classification)

- Perform exploratory Data Analysis and determine Training Labels
  - create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Find the method performs best using test data

Predictive Analysis - Github Link - Click Here

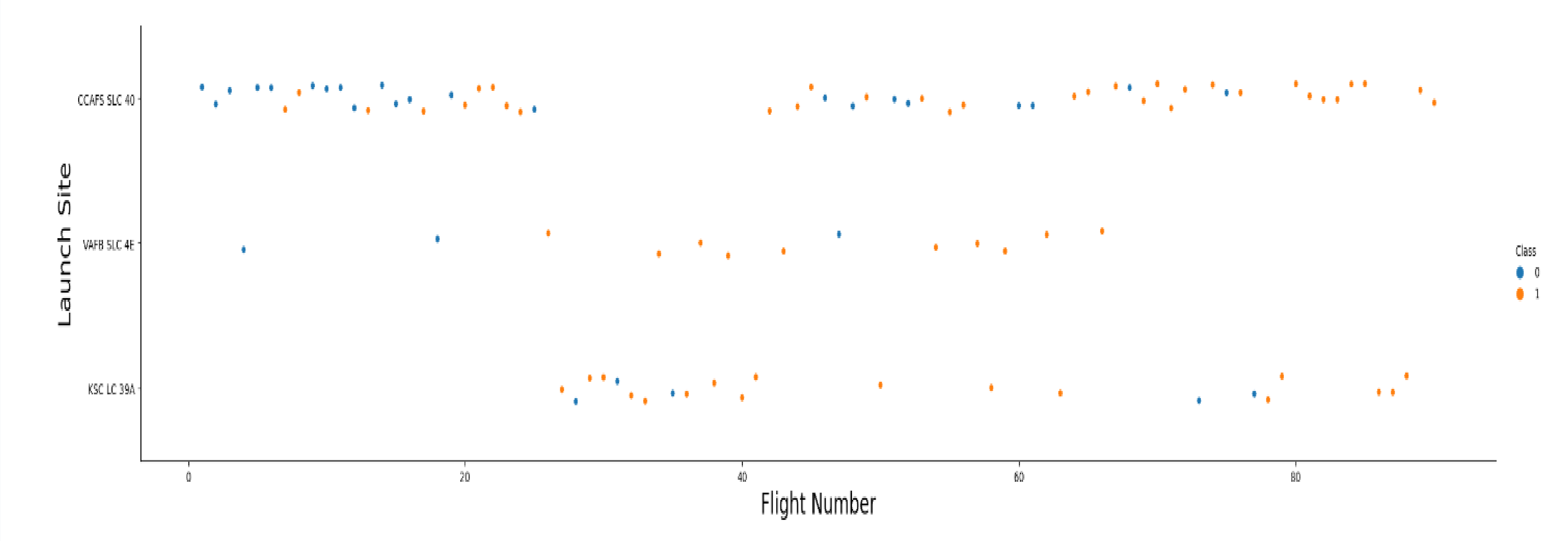
Results

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

Section 2

Insights drawn from EDA

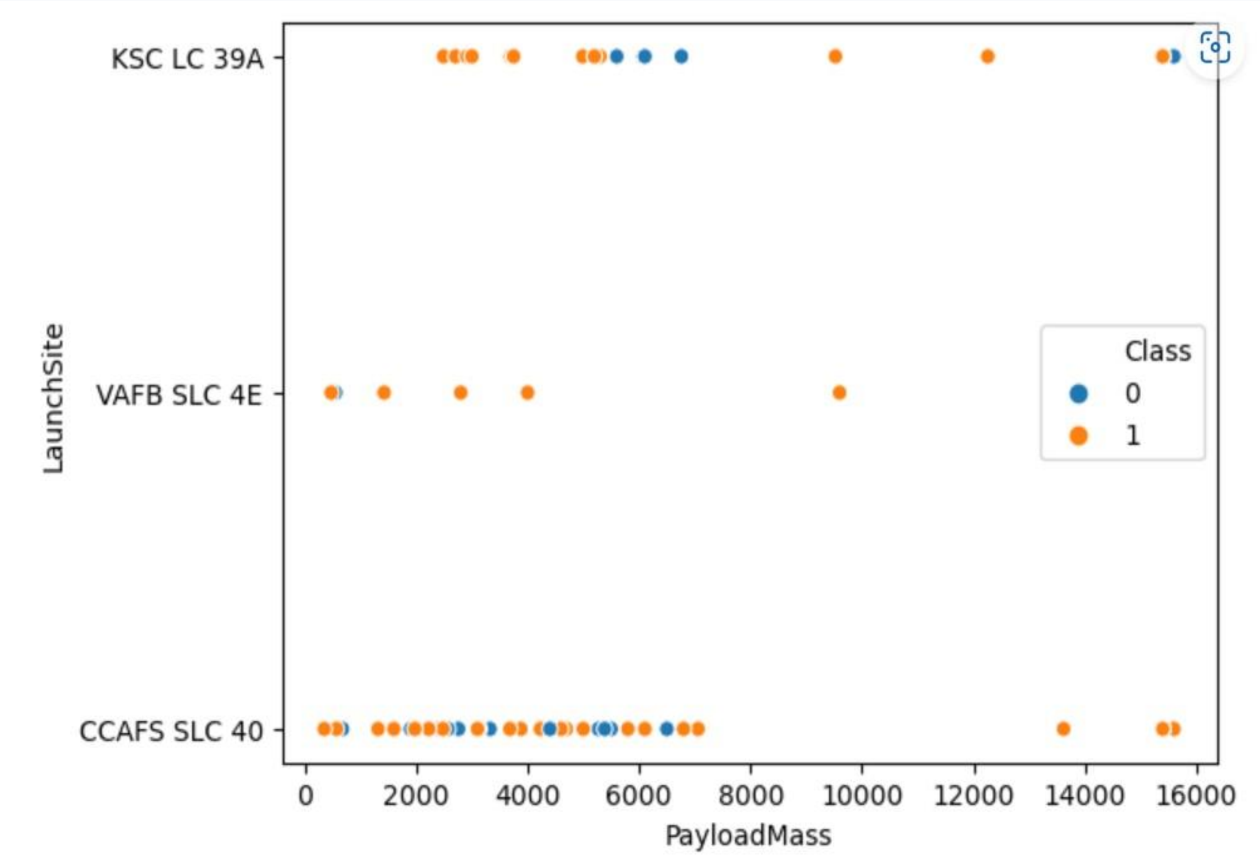
Flight Number vs. Launch Site



Launches from site CCAFS SLC 40 are significantly higher then others and success ration is increased when flight number increased.

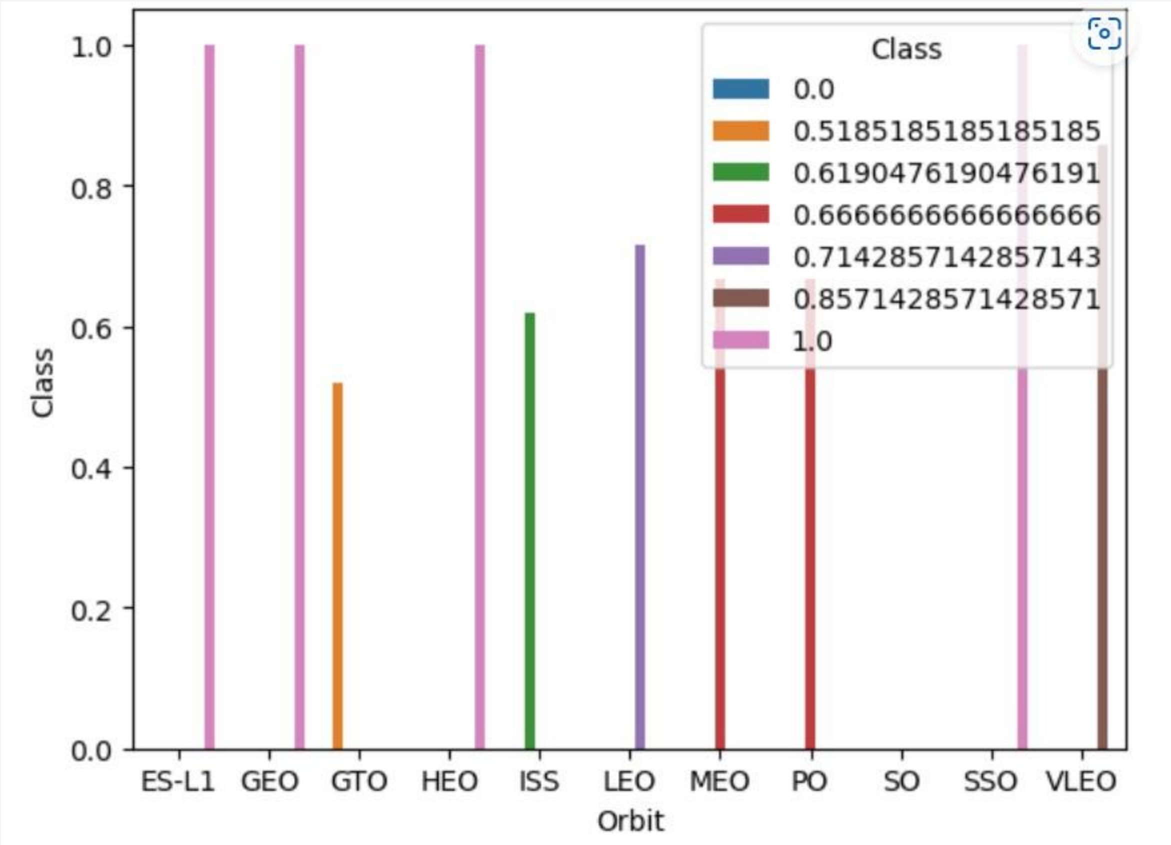
Payload vs. Launch Site

Success ratio is much higher if the launch site is VAFB SLC 4E and Payload is less then 10000



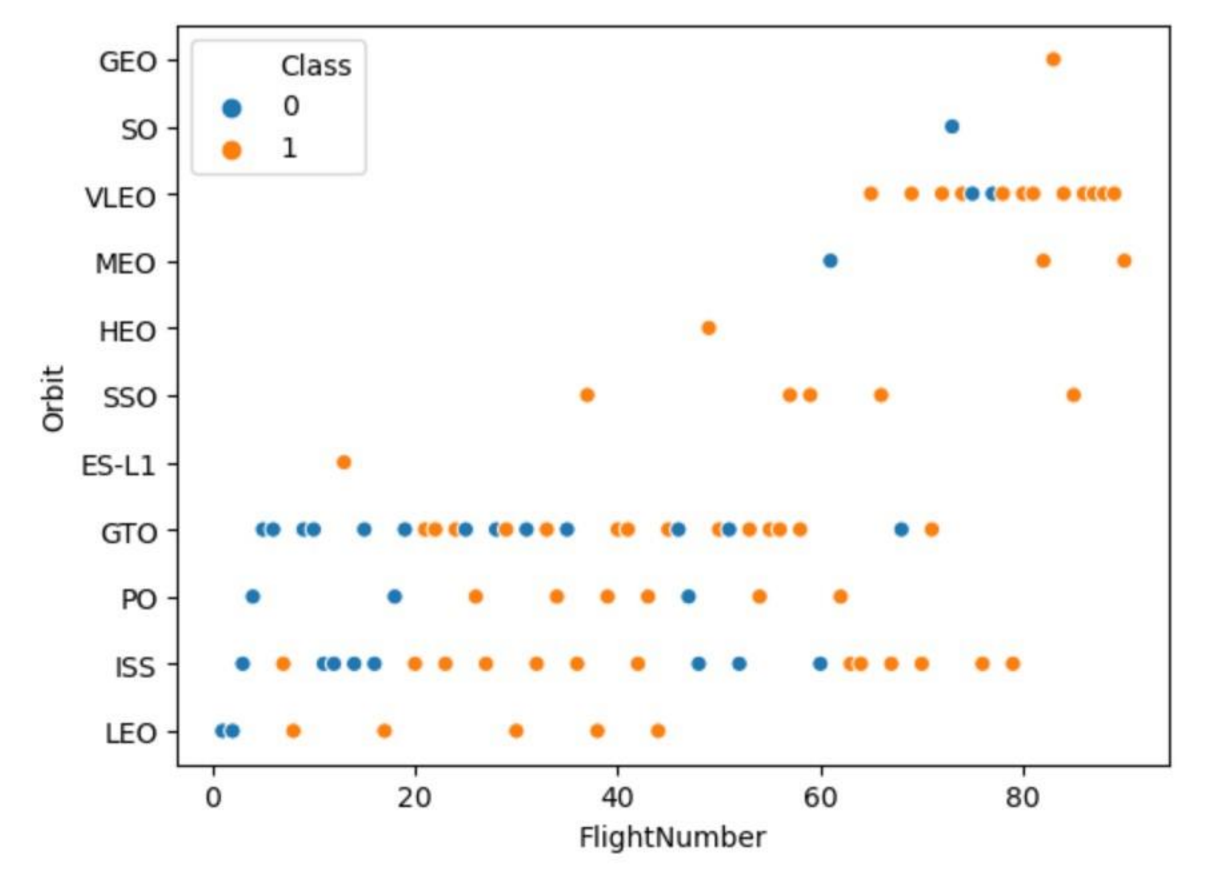
Success Rate vs. Orbit Type

Success rate is higher for the orbit ES-L1, GEO, HEO, SSO



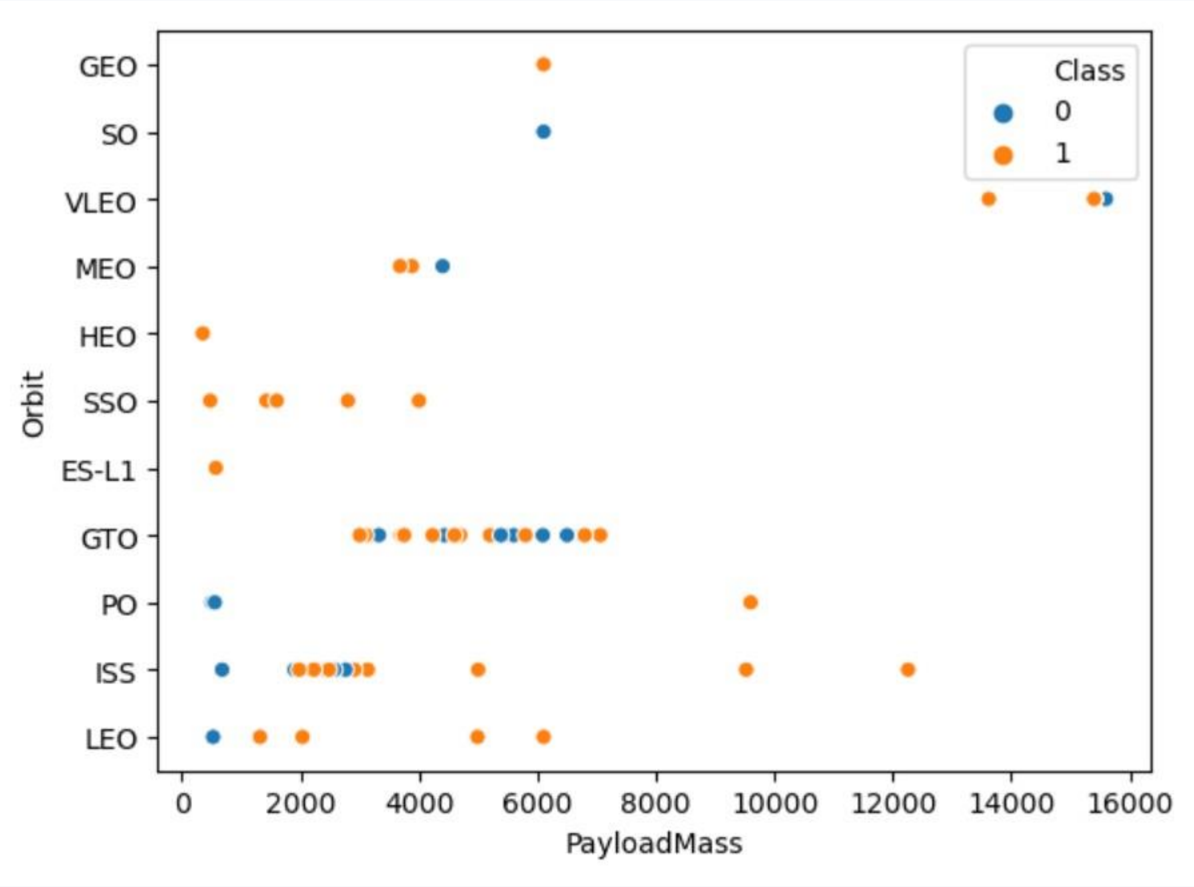
Flight Number vs. Orbit Type

We can see success for orbit VLEO, PO, ISS, LEO when flight is increased.



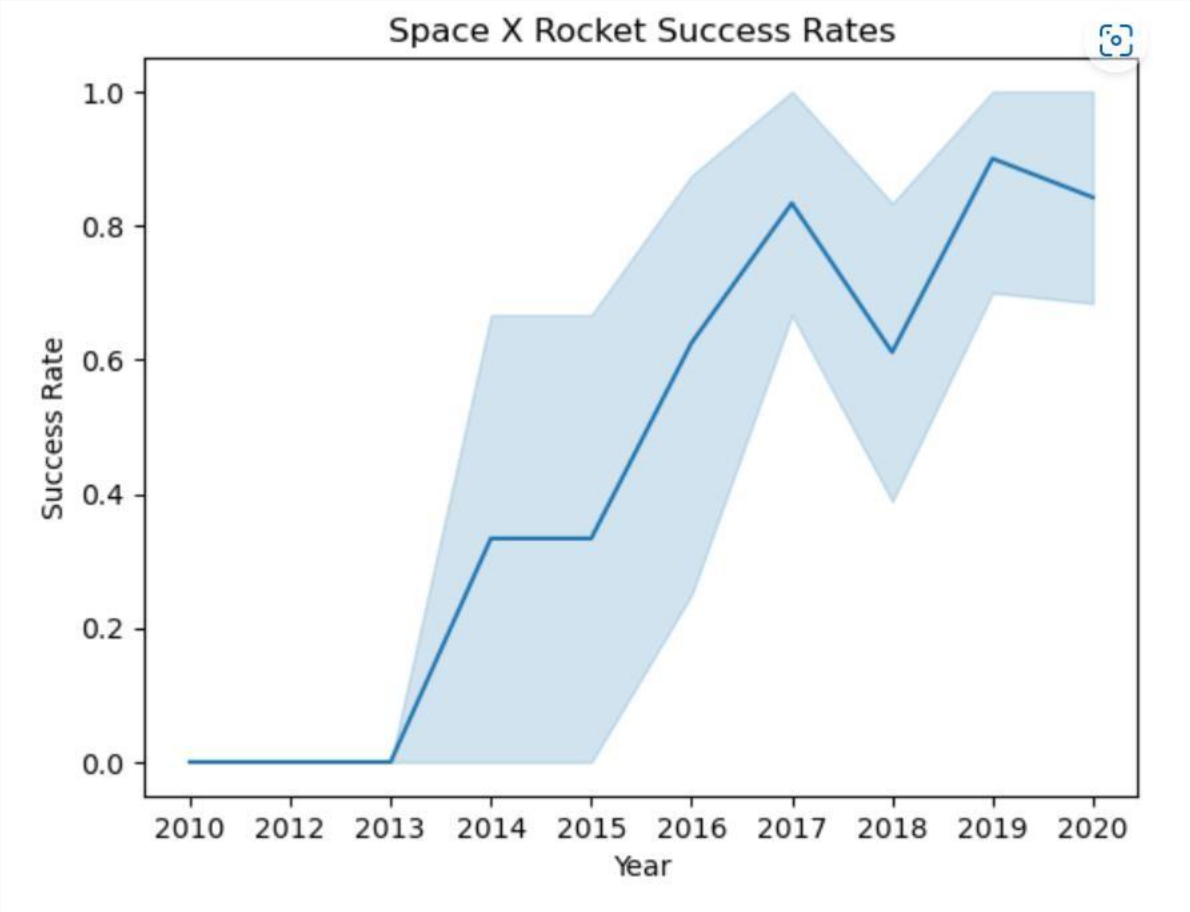
Payload vs. Orbit Type

- We can see that if payload is between 3000 and 14000, success rate is higher for orbit PO, ISS and LEO.
- For the SSO orbit, there is 100% success rate.



Launch Success Yearly Trend

Launch success is started from 2013 and continuous increased till 2019.



All Launch Site Names

There is Total four launch site

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Total payload carried by boosters from NASA

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

First successful landing on ground pad

MinDate
20151222

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes

Mission_Outcome	Result
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Booster which have carried the maximum payload mass

Booster_Version	sum(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

Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

Landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order with Rank

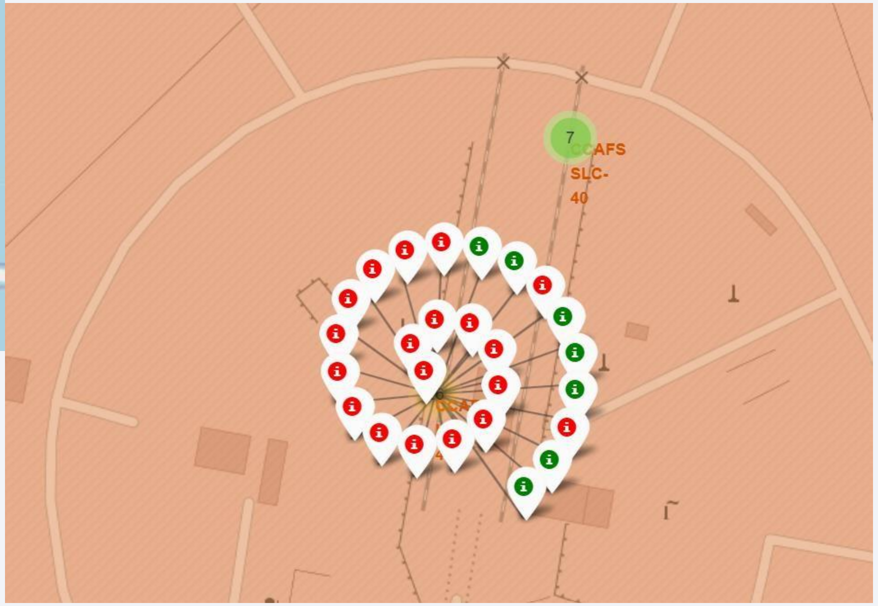
Rank	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
1	27-05-2016	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2	22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
3	19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
4	18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
5	14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
6	14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
7	08-04-2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
8	06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)

Section 3

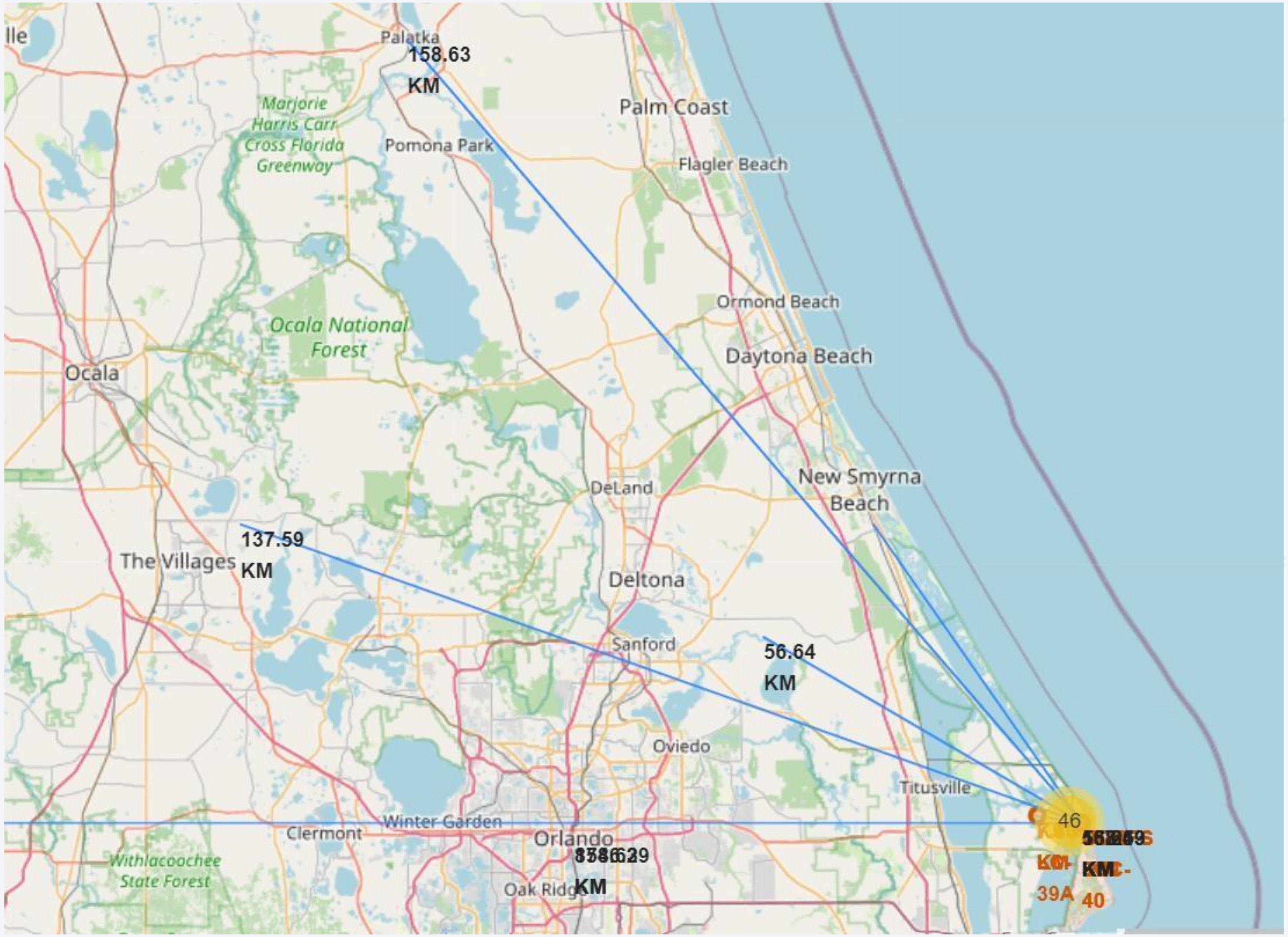
Launch Sites Proximities Analysis

All Launch Site

Successful and Failed launch



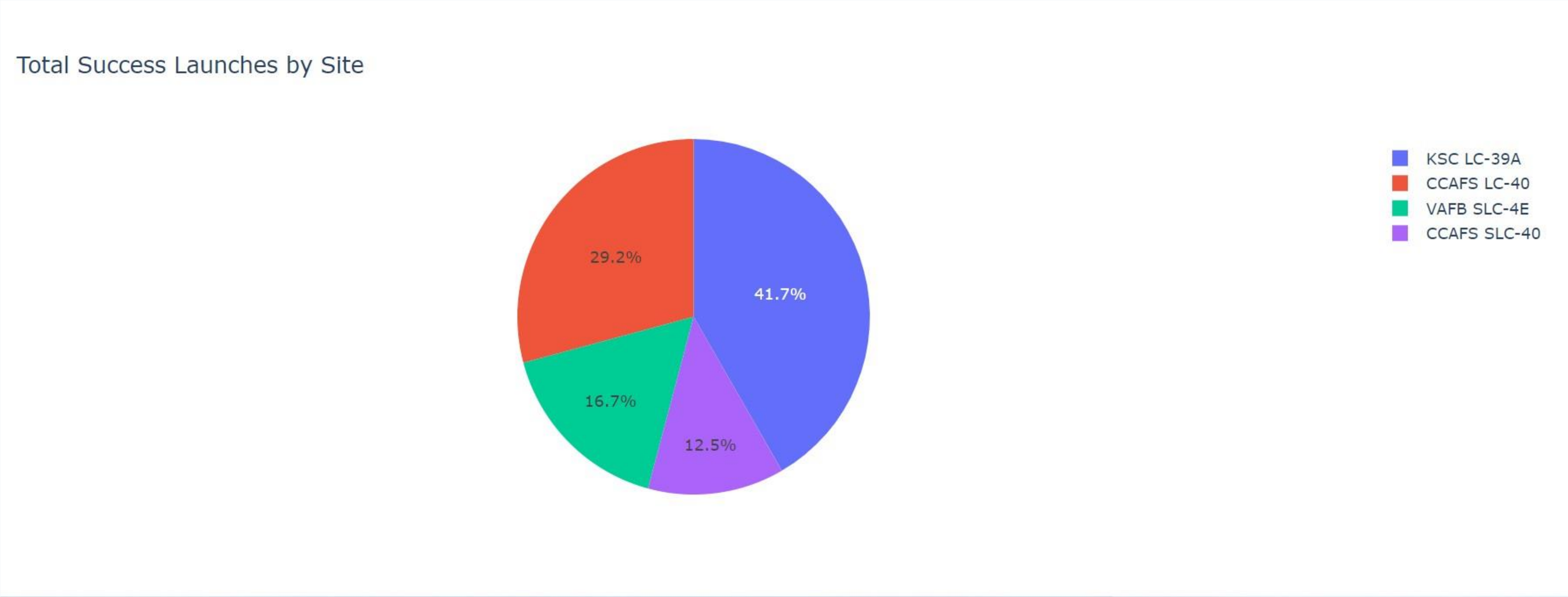
Distance between Launch site to its proximities



Section 4

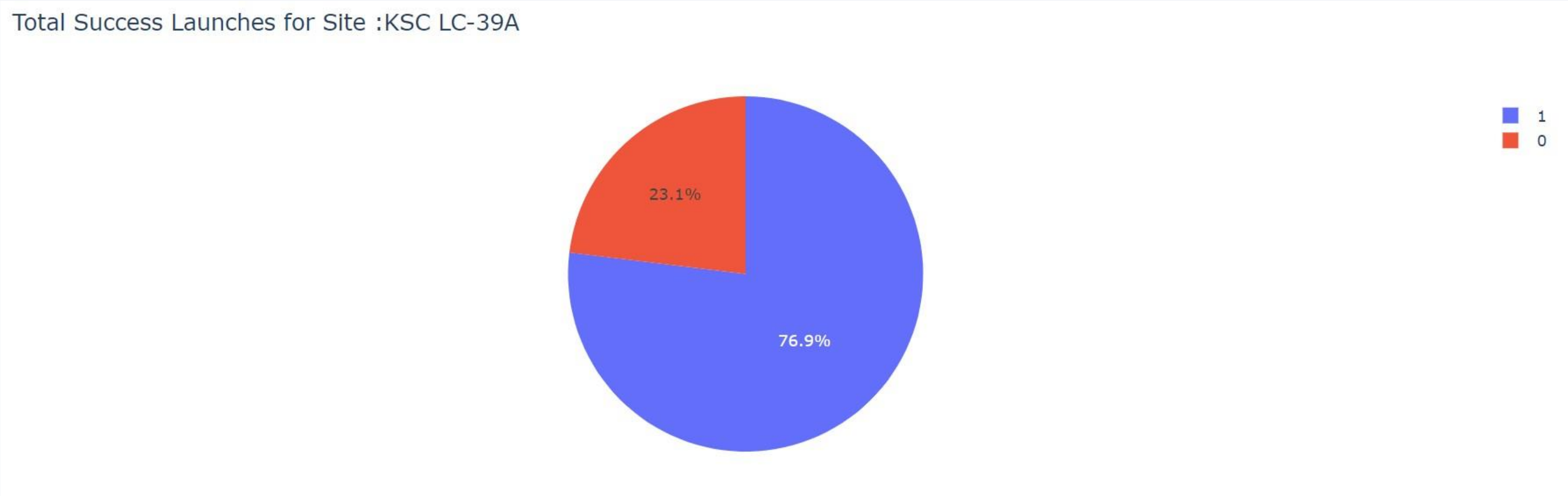
Build a Dashboard with Plotly Dash

Total Success Launches from All Site



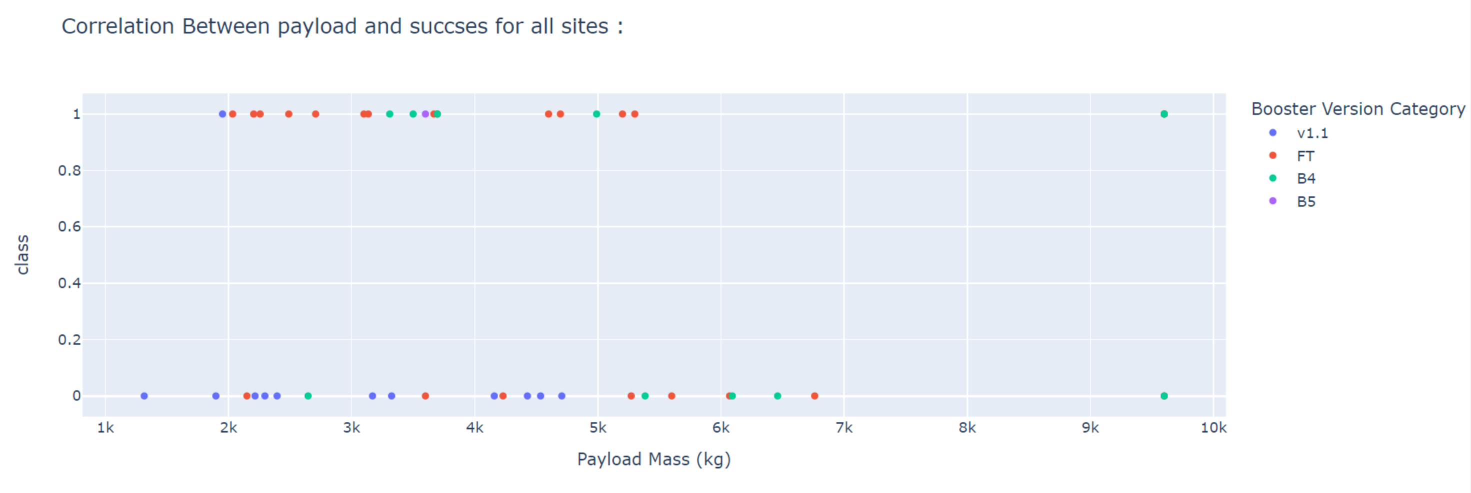
Launch site KSC LC 39A has more success ration then other sites

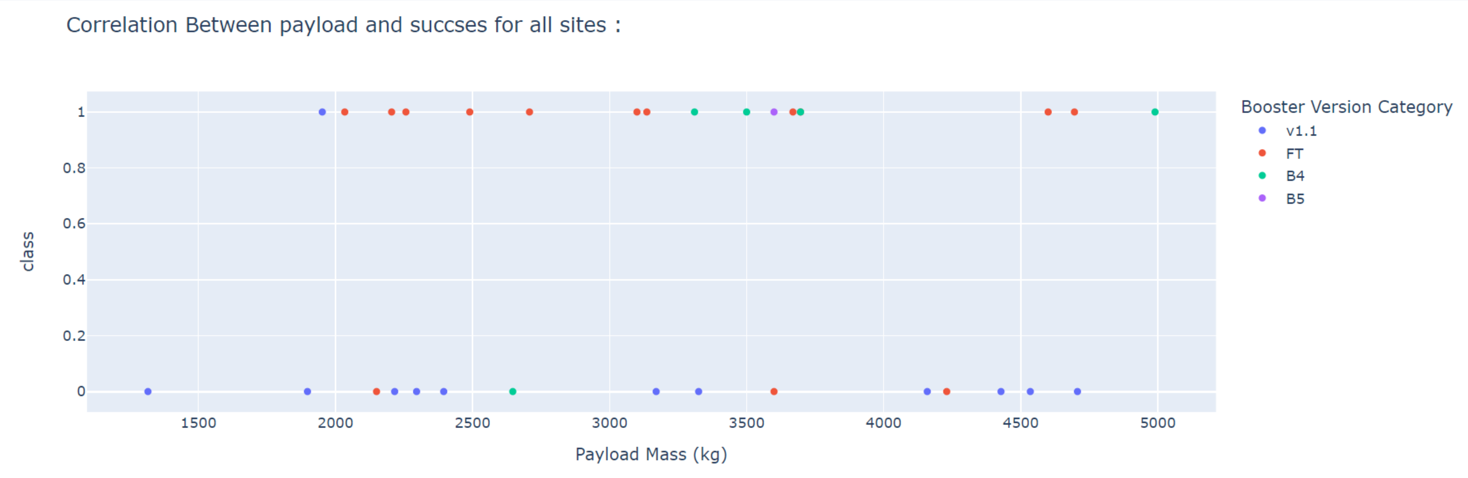
Details from highest launch success Ratio



Success ratio is more then 75% which is a good sign.

Payload Vs. Launch Outcome





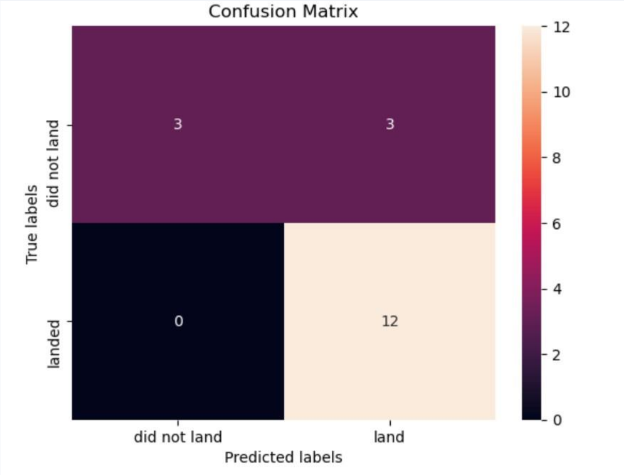
Section 5

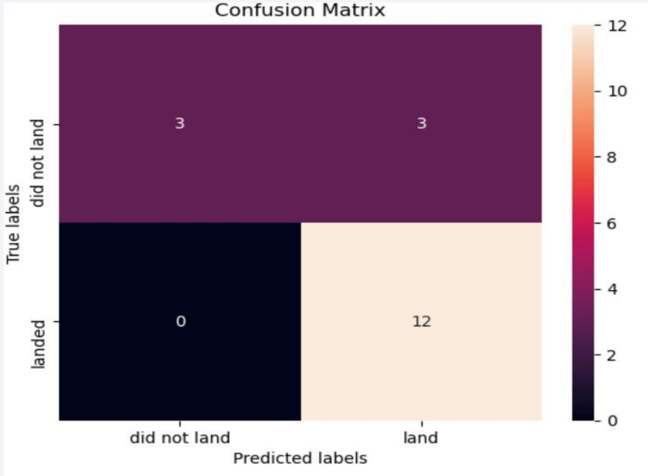
Predictive Analysis (Classification)

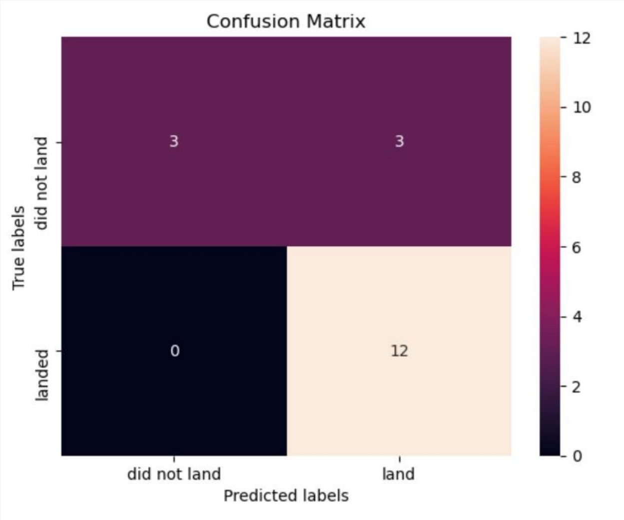
Classification Accuracy

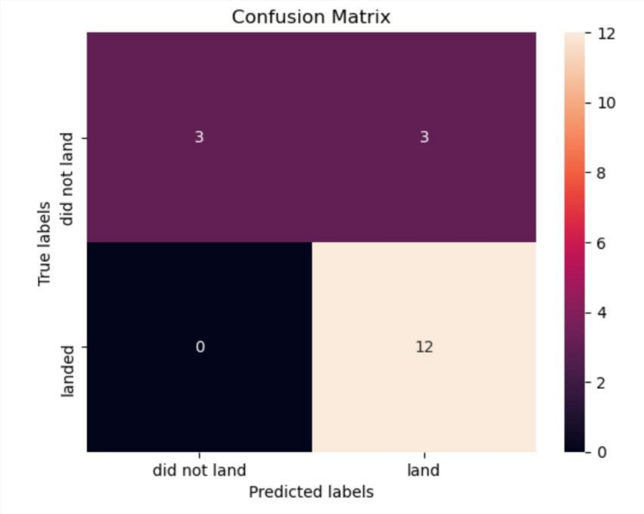
```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.6666666666666666
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Confusion Matrix









Based on the data we have similar outcome from different type of analysis.

Conclusions

- We were able to build a decision tree model that can predict the probability of Falcon 9 Rocket Stage 1 Landing Successfully with an 83.33% accuracy.
- We found that low weight payloads are more likely to land successfully than high payloads.
- We found that SpaceX engineers have been improving the probability of success every year since 2013 but progress has begun to slow down reaching a maximum yearly success percentage of about 63% in 2020 meaning our model will have to be refined as time passes to keep up to date
- The KSC LC-39A Launch Site also has the highest probability of success per launch
- The type of orbit required for the launch has an impact on the landing success, the ES-L1, SSO, HEO, and GEO orbits have the highest rate of success for landing

Thank you!