Winning Space Race with Data Science

Fabio Ferri
February 2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
  - 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. The main goal of project is predict best Machine Learning method first stage land Successfully.
- Problems you want to find answers
  - Find the best landing rate and Success Landing and Outcome.
  - Find the best labels to best approach model.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection API Xpace4
  - Data Web Scrapping Wikipedia
- Perform data wrangling
  - Enrich, normalize, standardized Data Collected. ,

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Data were collected, normalized, standardized and split in training and test data sets and evaluated by four different classification models, to get the accuracy of each model evaluated.

Data Collection

- Describe how data sets were collected.
  - Datasets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) from python client API code.
    - Elementary labels columns
  - Dataset from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping python BeatfullSoup api.
    - Complement Data labels columns

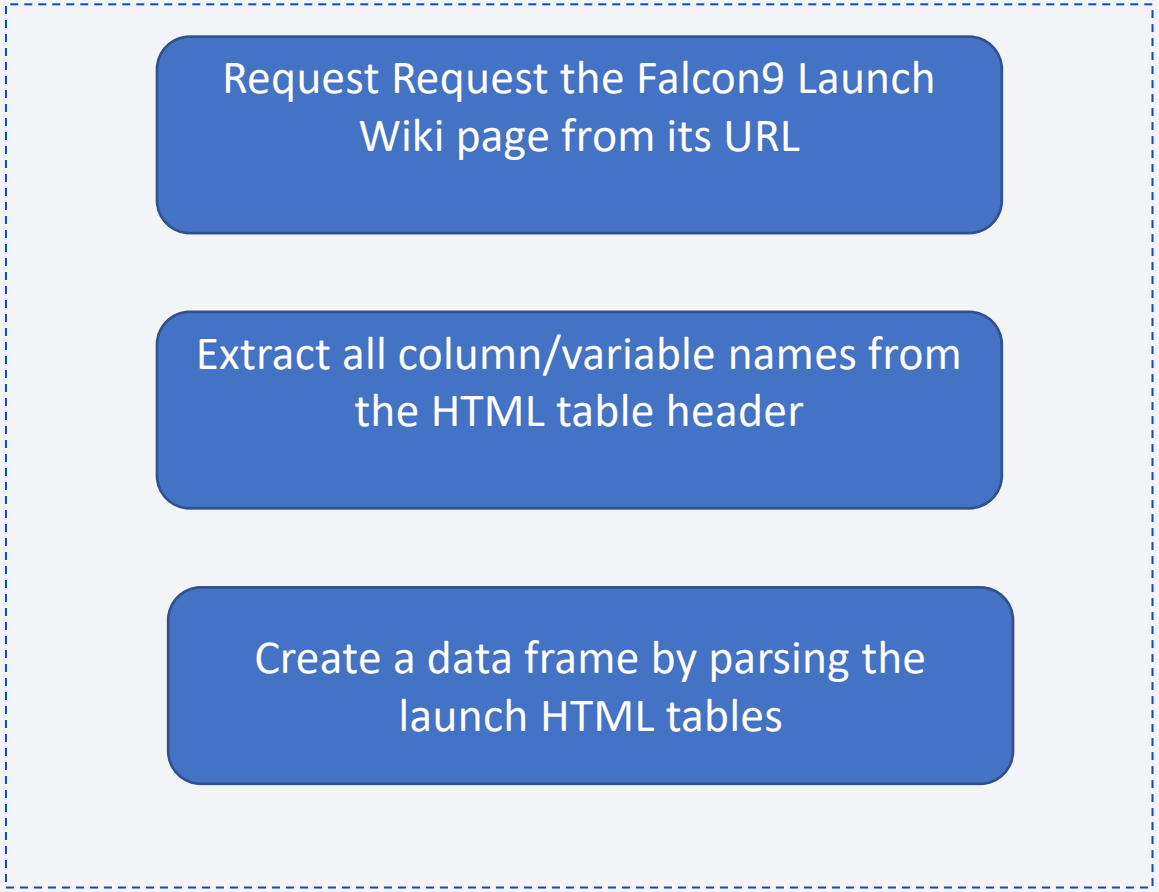
Data Collection – SpaceX API

- SpaceX offers a public API from where data do be collected.
- GitHub URL of the completed SpaceX API calls notebook: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week1/Lab1-DataCollecting-API.ipynb

Data Collection - Scraping

- Web Scracping Data Collection to add more complement labels features.

- GitHub URL of the completed web scraping notebook: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week1/Lab1-DataCollecting-API.ipynb



Data Wrangling

- Normalization Data and Format
- Missing null values , replace by mean value labels.
- Transform Labels to Outcome Labels to Success or Fail , land, payload mass, orbit, e.g)

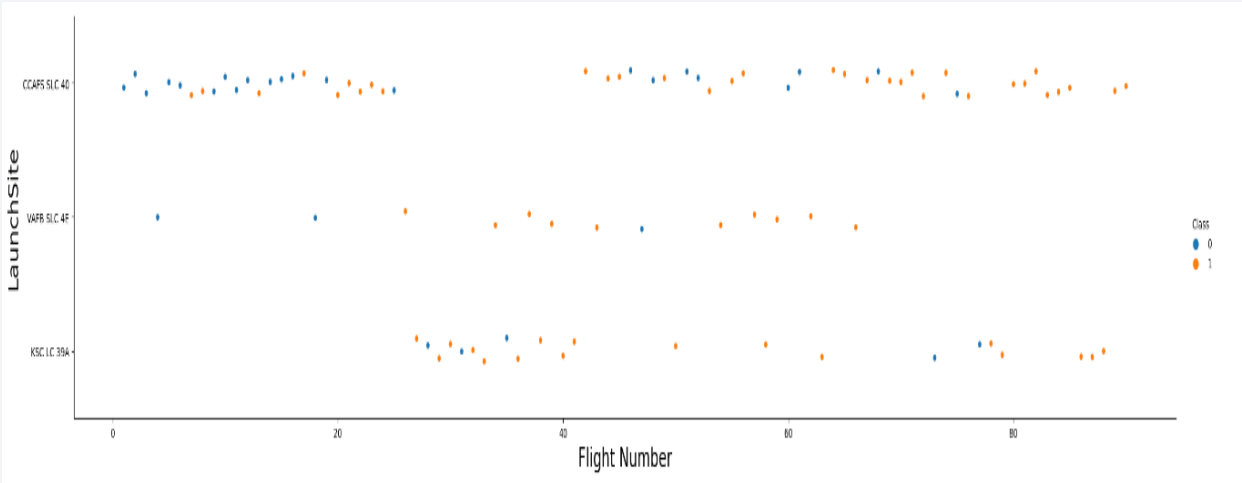
- GitHub URL:
  - https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week1/Lab_DataWrangling_FF.ipynb

EDA with Data Visualization

- Exploratory Data Analysis and Feature Engineering have got:
  - LEO orbit the Success appears related to the number of flights.
  - Success rate since 2013 kept increasing till 2020.
  - With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
  - Payload mass have been increased since begin flights from aproxly < 2000 KG to today above 16.000 kg.
  - CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- GitHub URL: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week2/Lab-FF-EDAwithDataVisualization.ipynb

EDA with SQL

- Compare Scatterplots using :
  - Payload Mass X Flight Number
  - Launch Site X Flight Number
  - Launch Site X Payload Mass
  - Orbit and Flight Number
  - Payload and Orbit



- GitHub URL: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week2/Lab-FF-EDAwithDataVisualization.ipynb

Build an Interactive Map with Folium

- All Launch Sites were marked with circle
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

- GitHub URL: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week3/Lab3_Dashboards_FF_feito.ipynb

Build a Dashboard with Plotly Dash

- It have been added and Dropdown for all Lauch sites.
- A pie chart to show the total successful launches count for all sites.
- We plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.

- GitHub URL: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week3/spacex_dash_app.py

Predictive Analysis (Classification)

- Load DataSet into DataFrame Pandas e Numpy array.
- Standarlize the data X variables.
- Set train and set tests 80% to Train/20 % to Test.
- Create a four Machine Learnings methods:

  Logistics Regression

  Support Vector Machine

  Decision tree

  KNN neighbors

  Find the method performs best

- Evaluate the best accuracy four models

- GitHub URL: https://github.com/FabioFerri2907/IBMDataScience/blob/master/Curso10-Applied%20Data%20Science%20Capstone/Week4/SpaceX_Machine%20Learning%20Prediction_Part_5_FF_Labs.ipynb

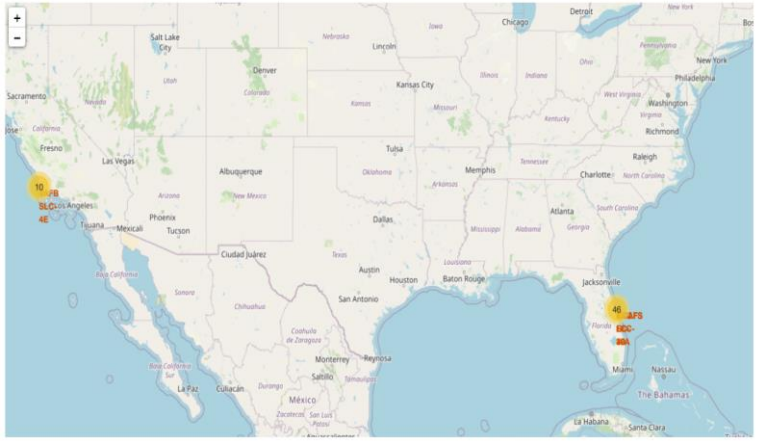
Results

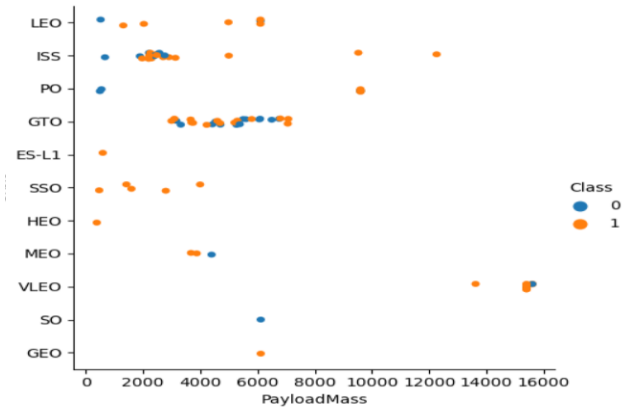
- Exploratory data analysis results

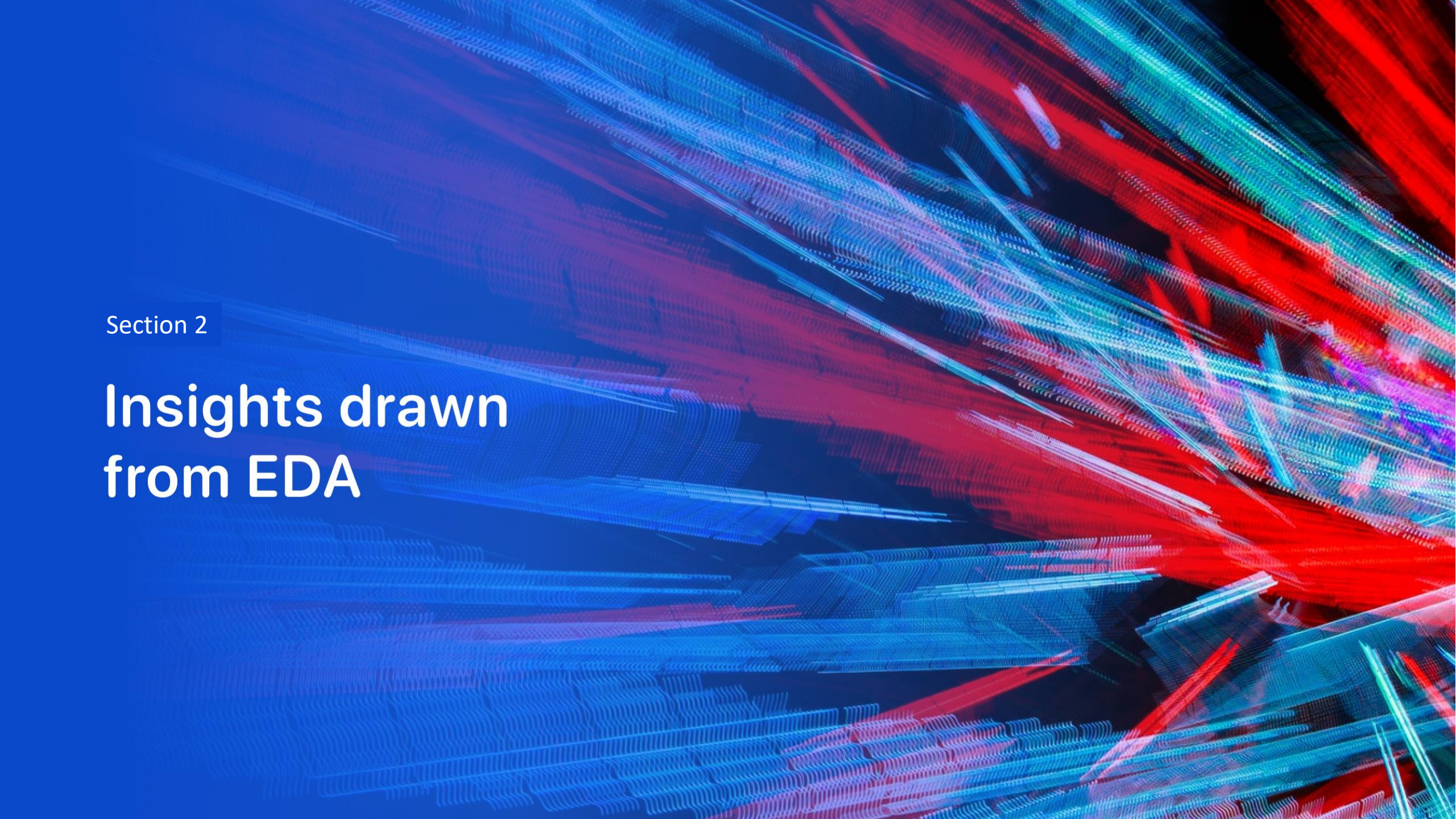
  - Space X uses 4 different launch sites.
  - The first launches were done to Space X itself and NASA.
  - The First Success launch was in 2015.
  - Payload Mass have been increasing last years.
  - KSC LC-39A and VAFB SLC 4E has a success rate of 77%,
    CCAFS LC-40, has a success rate of 60 %.
  - The number of landing outcomes became as better as years passed.
  - They have had positive landing rate are more for Polar,LEO and ISS orbit.

Results

- Most of launch sites are near to sea, cost line and high , because good logistic.
- LEO, VLEO and SSO are the best rate success outcomes.
- Polar,LEO and ISS the highest payload mass.

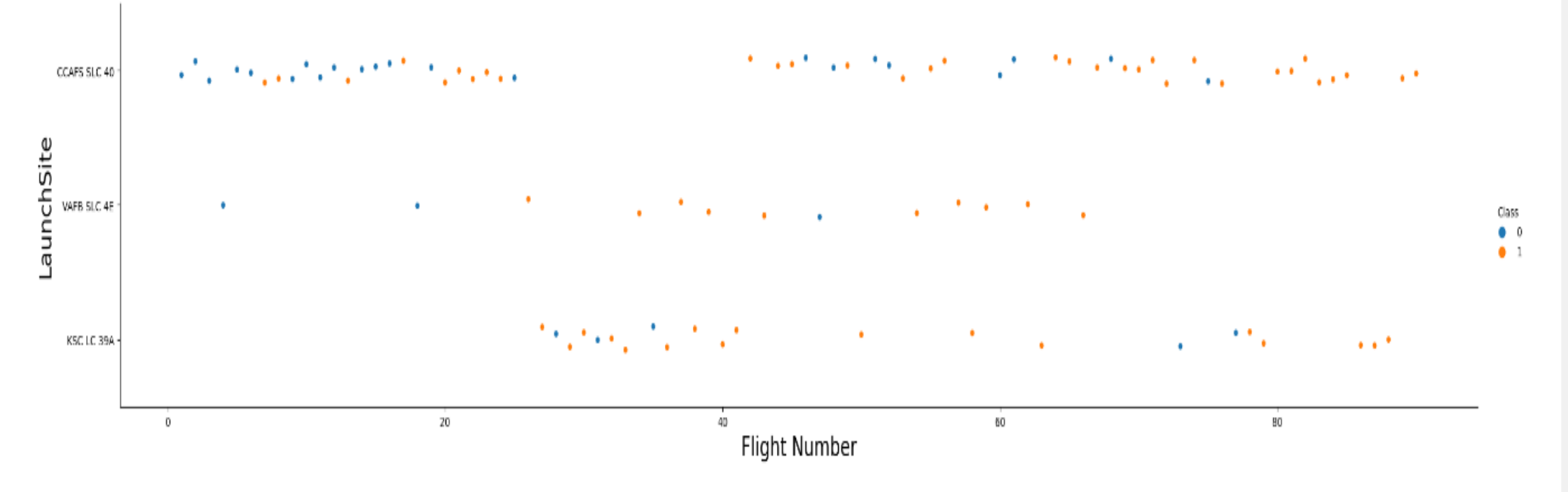





Section 2
Insights drawn from EDA

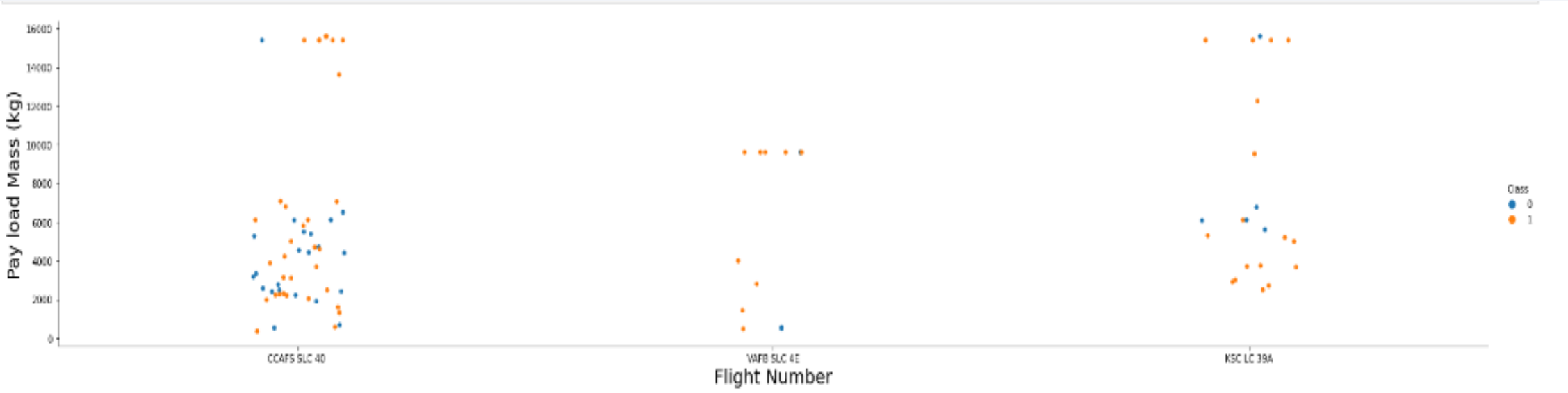
Flight Number vs. Launch Site

- CCAFS SLC 40 have the most of numbers of flight.
- The status success rate have been increase the last flights.



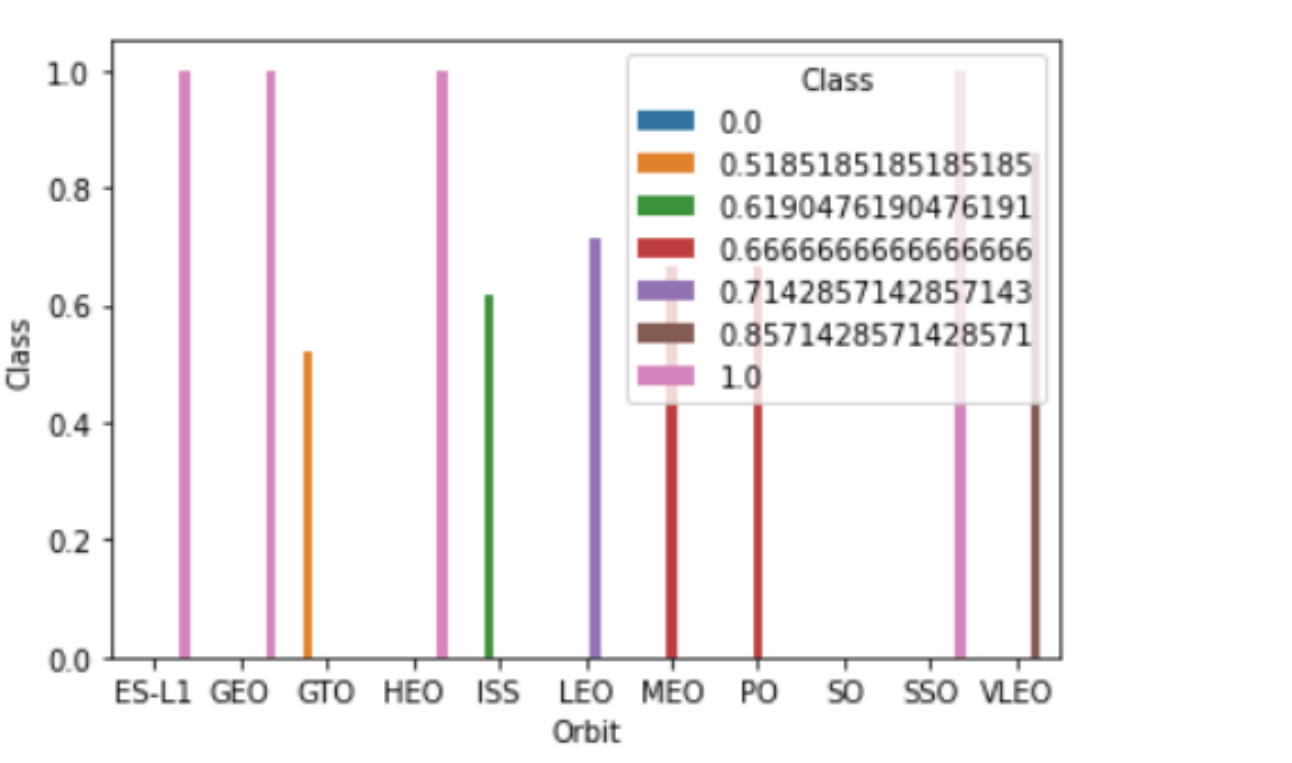
Payload vs. Launch Site

- CCAFS SLC 40 and KSC LC39A site has both have Success or Fail flight, whereas VAFB SLC 4E has more success rate.



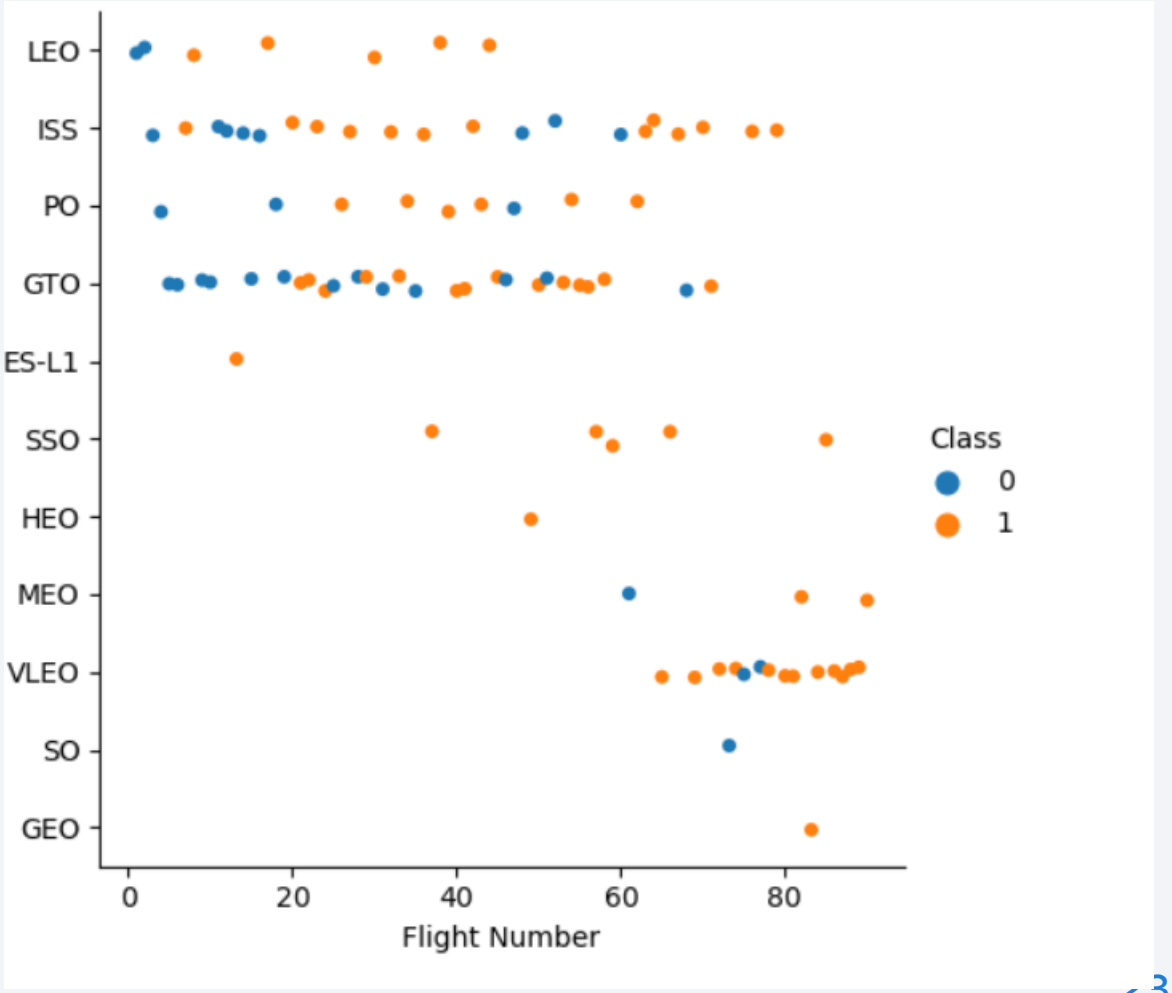
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO, have best success rate.
- GTO, MEO and Polar the worst rate.



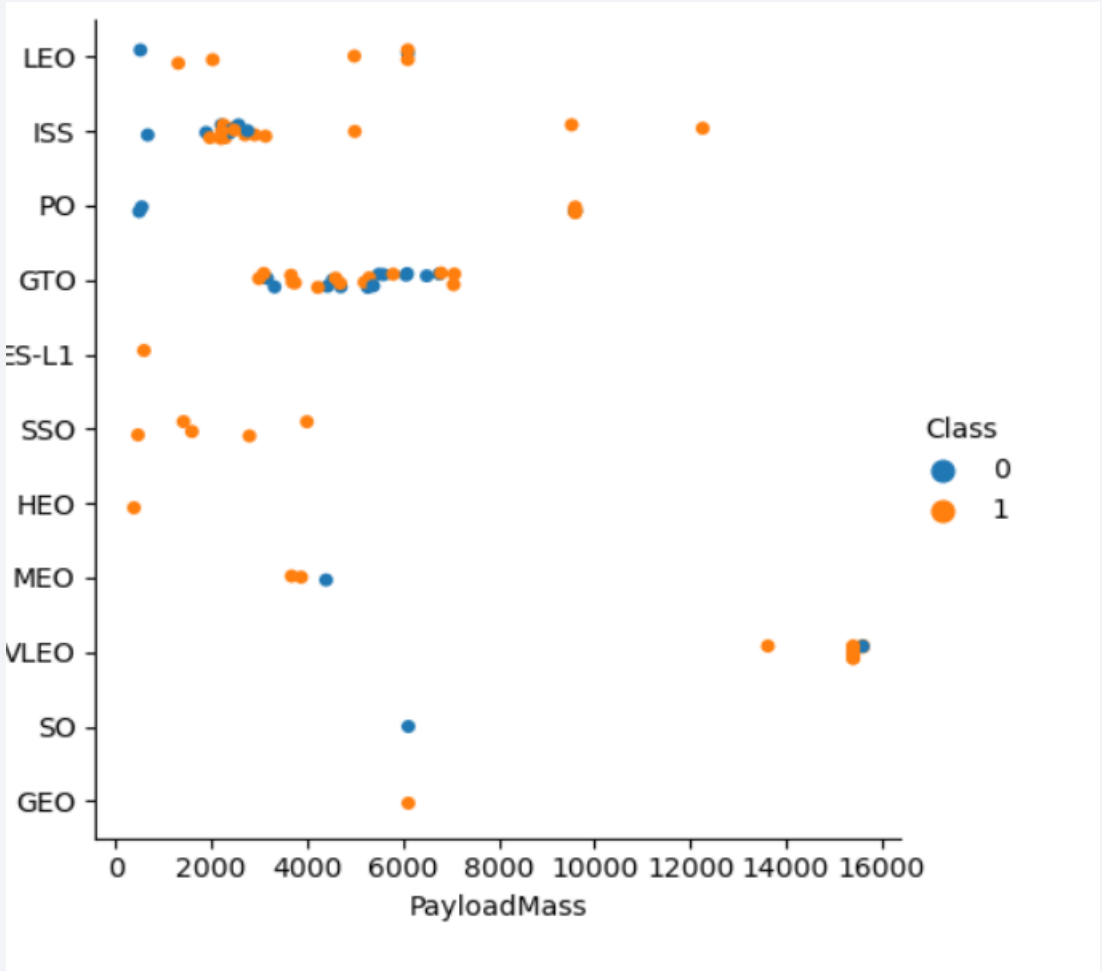
Flight Number vs. Orbit Type

- LEO, SSO and VLEO have had the success rate.
- GTO and ISS have the worst rate.
- SSO have 100% Success rate.



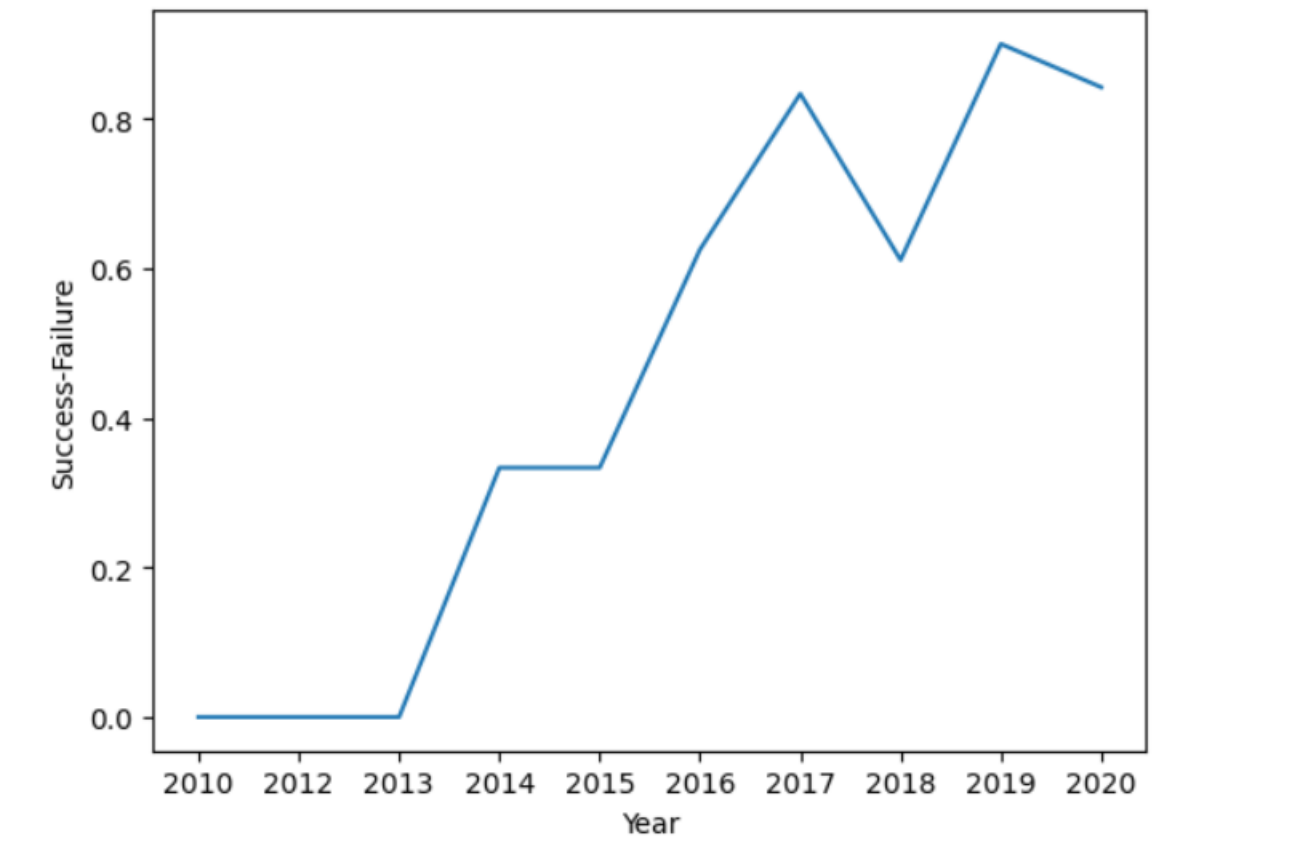
Payload vs. Orbit Type

- ISS, PO and VLEO are the highest heavy payload.



Launch Success Yearly Trend

- We can observe that the success rate since 2013 kept increasing till 2020.



All Launch Site Names

- Fours Launch Sites found.
- DISTINC command used to get unique name.

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

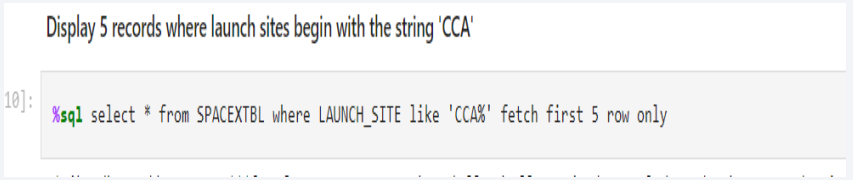
Display the names of the unique launch sites in the space mission

In [8]:
```
%sql select distinct LAUNCH_SITE from SPACEXTBL
```

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- SQL Like command used to filter ‘CCA’

DATE	time__utc_	booster_version	launch_site
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40



Total Payload Mass

- Calculate the total payload carried by boosters from NASA are 45.596 Kg.
- SUM command to PAYLOAD_MASS_KG where NASA (CRS).


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

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928 Kg.
- The AVG command to get mean PAYLOAD_MASS_KG where BOOSTER_VERSION = F9 V1.1

1
2928

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

First Successful Ground Landing Date

- 22th December 2015 of the first successful landing outcome on ground pad
- Min command in Data field where LANDING_OUTCOME = 'Success (ground pad)'

2015-12-22

```
%sql select min(date) from SPACEXTBL where LANDING__OUTCOME = 'Success (ground pad)'
```

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
- Between command 4000 to 6000 in PAYLOAD_MASS_KG to get BOOST_VERSION.

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

```
14]: %sql select BOOSTER_VERSION from SPACEXTBL where LANDING__OUTCOME ='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

Total Number of Successful and Failure Mission Outcomes

- Successful Outcomes 99%
- Failure Outcomes 1%
- Count command MISSION_OUTCOME and GROUP BY used to get Total Number MISSION_OUTCOME field.

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

```
%sql Select MISSION_OUTCOME,count(MISSION_OUTCOME) as count from SPACEXTBL GROUP BY MISSION_OUTCOME
```

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Used a subquery to get MAX payload Mass and used it where in fisrt query to DISTINC COMMAND by BOOSTER_VERSION.

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

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

```
%sql select distinct BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

2015 Launch Records

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

- Just Where SQL by Date and Landing Outcome filter

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

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

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

	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

We selected Landing outcomes and the **COUNT** of landing outcomes from the data and used the **WHERE** clause to filter for landing outcomes **BETWEEN** 2010-06-04 to 2010-03-20.

We applied the **GROUP BY** clause to group the landing outcomes and the **ORDER BY** clause to order the grouped landing outcome in descending order.

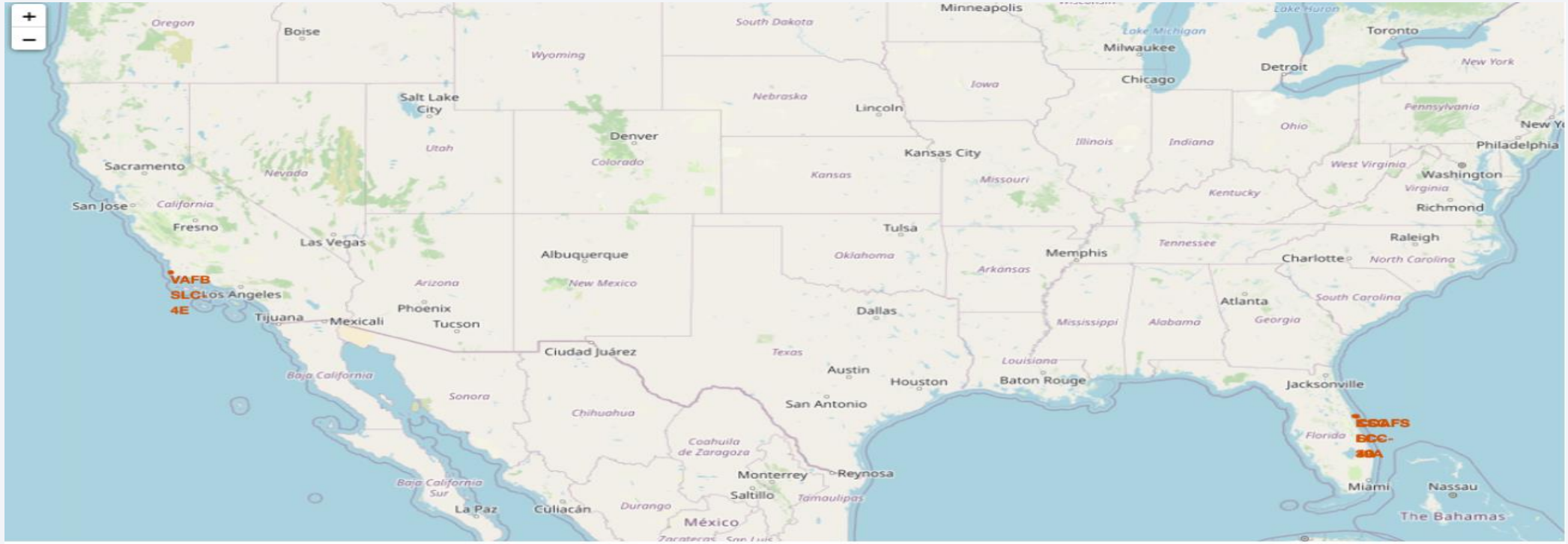
```
%sql select LANDING__OUTCOME, count(LANDING__OUTCOME) as count from SPACEXTBL group by LANDING__OUTCOME having date between '04-06-2010' and '20-03-20
```

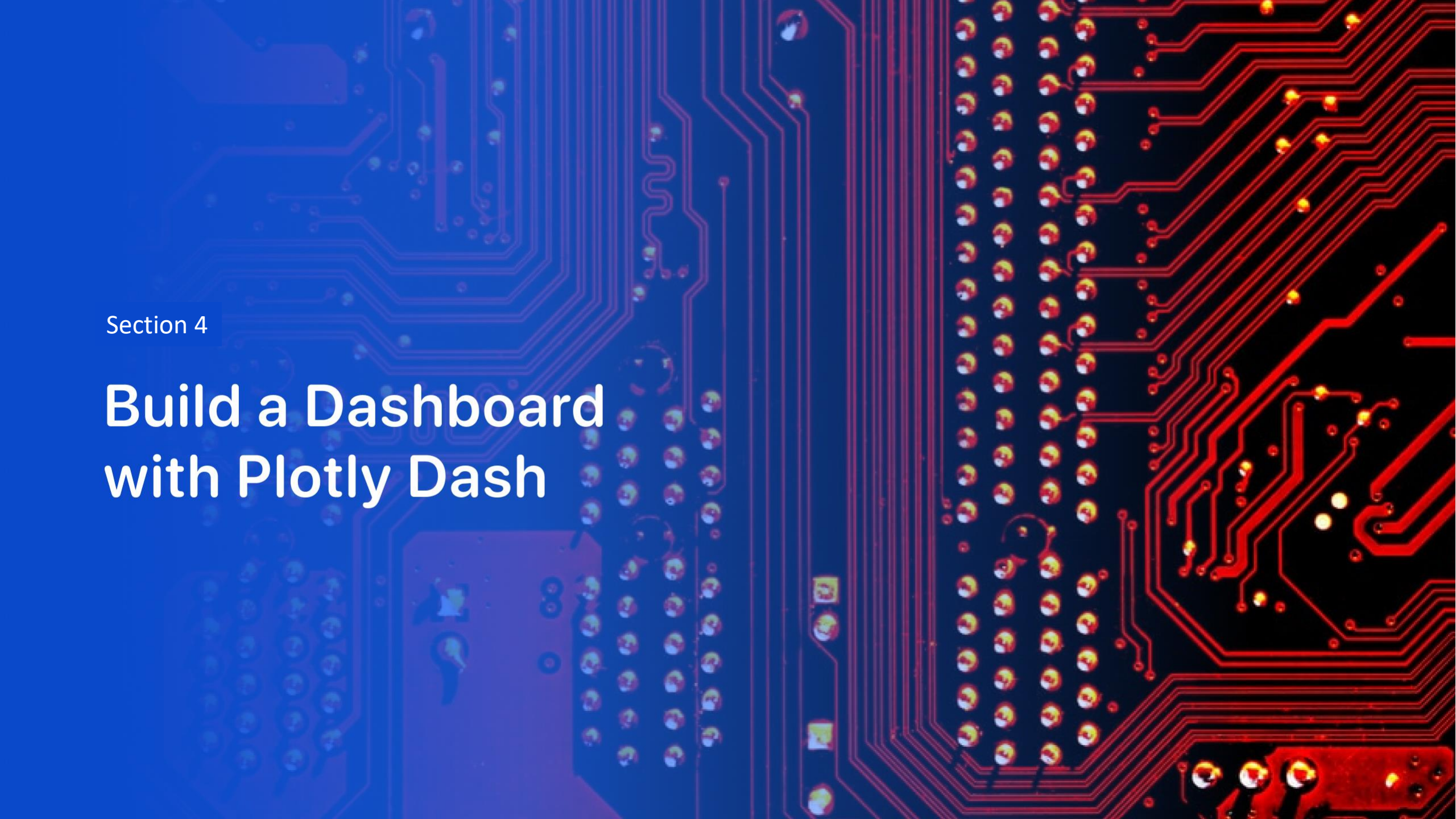
Section 3

Launch Sites Proximities Analysis

All Launch Sites Map

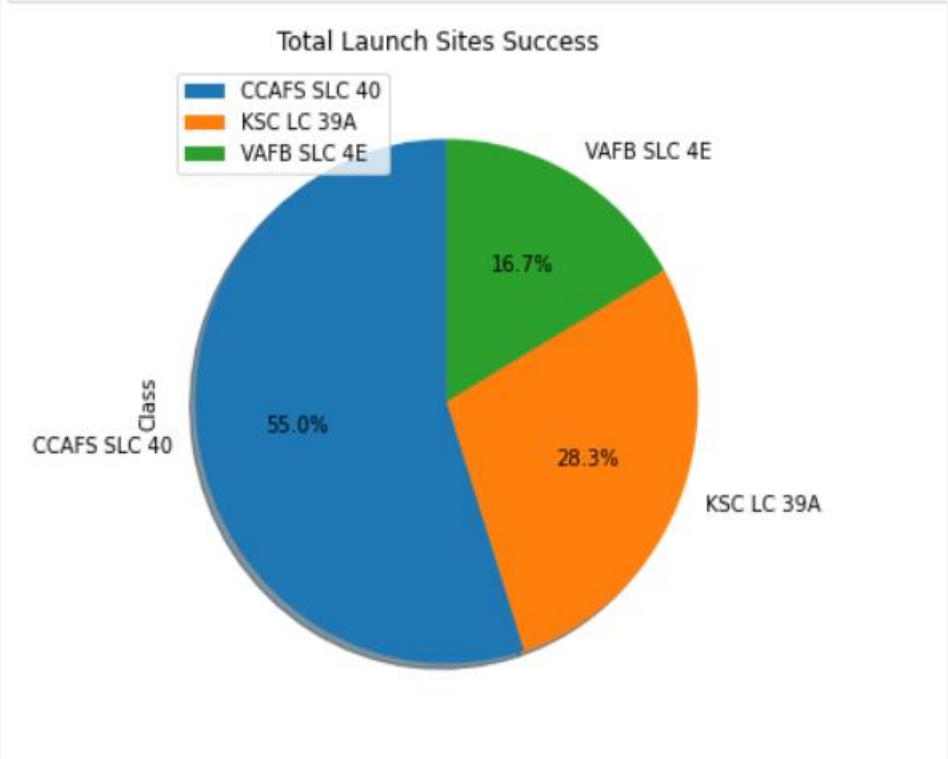
- All the Launch Sites are USA near of sea cost sides.




Section 4
Build a Dashboard with Plotly Dash

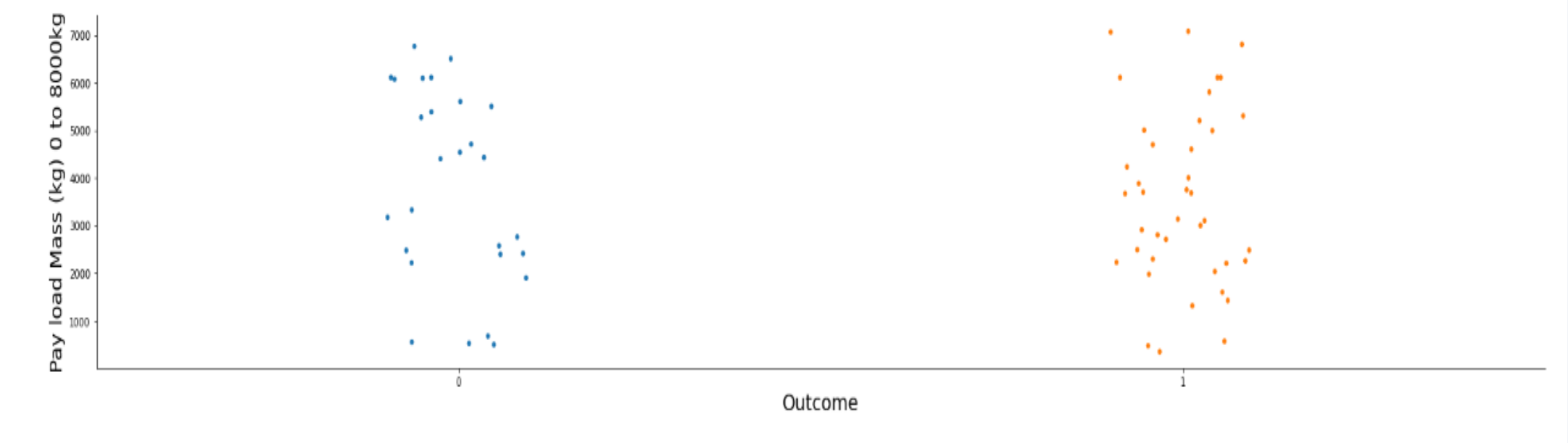
Total Success Launch Sites

- Total Success Outcome rate by Launch Sites



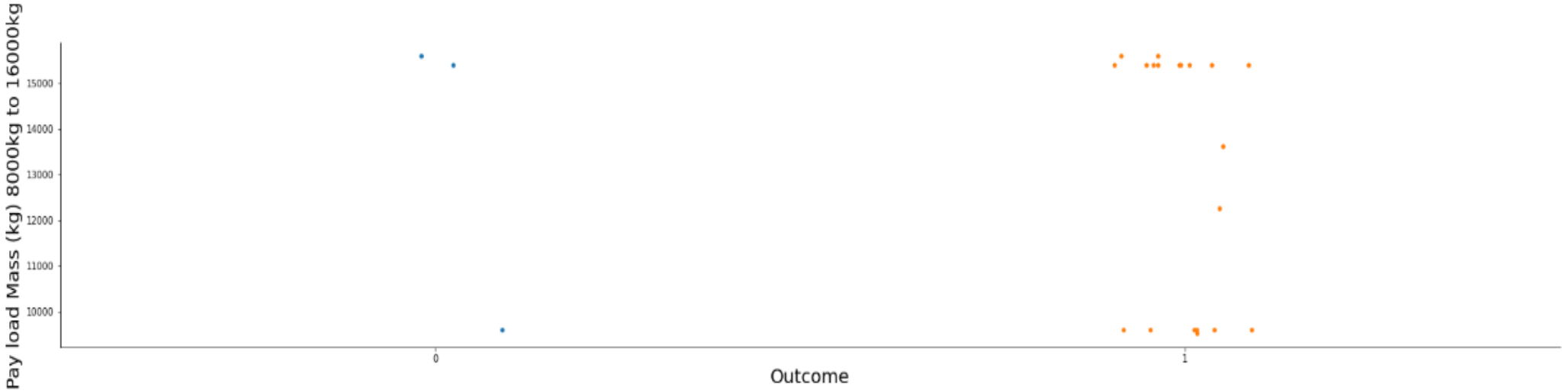
Dashboard Payload X Outcome

- Range Payload Mass from 0 to 8000kg
  - Between these range we have number mix Success Rates for Success Mission and Failed.
  - We have had most Success rate from 2000kg to 6000kg.



Dashboard Payload X Outcome

- Range Payload Mass from 8000kg to 16000kg.
  - Between these range we have biggest number mix Success Rates.
  - We have had most Success rate at 10000kg and 16000kg.



Section 5

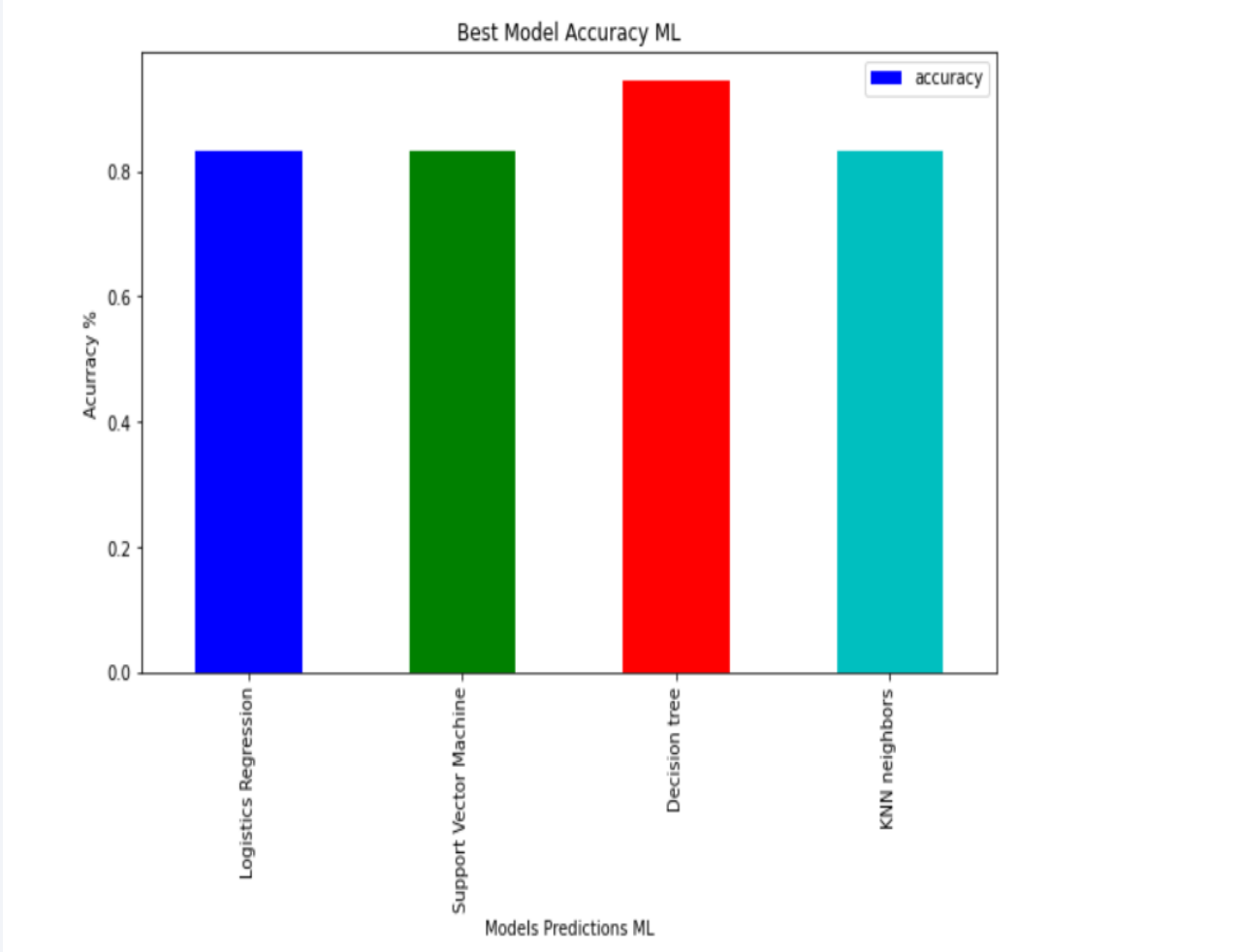
Predictive Analysis (Classification)

Classification Accuracy

- The Best Model Accuracy is Decision Tree.

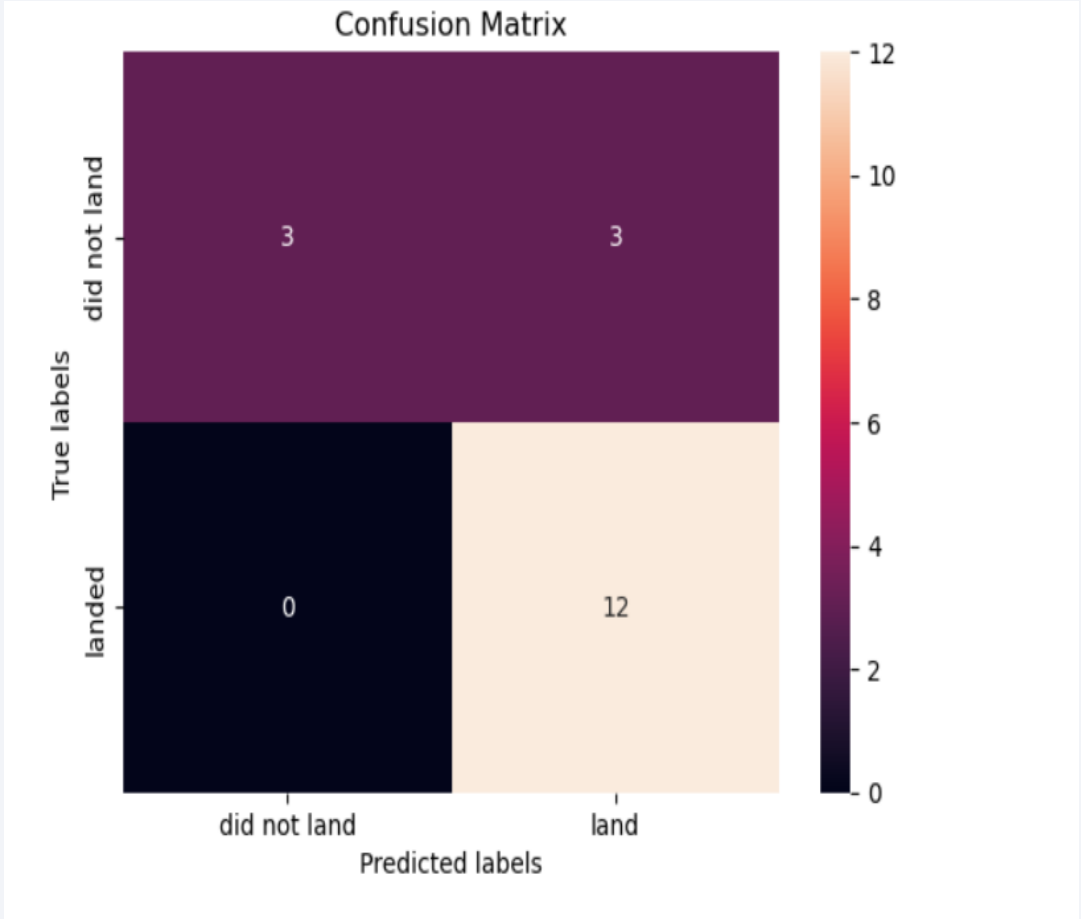
```
In [56]: print('Accuracy for Logistics Regression :', logreg_cv.score(X_test, y_test))
         print( 'Accuracy for Support Vector Machine ', svm_cv.score(X_test, y_test))
         print('Accuracy for Decision tree:', tree_cv.score(X_test, y_test))
         print('Accuracy for KNN neighbors :', knn_cv.score(X_test, y_test))

         Accuracy for Logistics Regression : 0.8333333333333334
         Accuracy for Support Vector Machine  0.8333333333333334
         Accuracy for Decision tree: 0.9444444444444444
         Accuracy for KNN neighbors : 0.8333333333333334
```



Confusion Matrix

- The Decision Tree Model had the Best Model and Confusion Matrix of it, it is good and it can classifier as well.
- Some observations are not correctly predictive value for Unsuccessful Landing for Success Landing.



Conclusions

- CCFA SLC 40 Is the highest Success Launch Site rate.
- Launch success rate started to increase in 2013 till 2020.
- The Payload Mass have been increasing last flight missions.
- Most of launch sites are near to sea, cost line and high , because good logistic.
- LEO, VLEO and SSO are the best rate success outcomes.
- Polar,LEO and ISS the highest payload mass.
- The Descion Tree is chosen model for the support the

  predict the Falcon 9 first stage will land successfully.

Thank you!