Winning Space Race with Data Science

Papu Chetry
6th Feb’22

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies

  Collecting the data

  Data Wrangling

  Exploratory data analysis

  Predictive analysis (classification)

- Summary of all results – EDA results, Interactive and Predictive results
- The past data of Space X launches and the variables could be known
- The success rate analysed and effort was made to predict the landing outcome

Introduction

- Project background and context
- Out of the several companies working on space travel, Space X is the most successful as on date.
- One major reason for the success is the least expensive methods used by Space X where the first stage of the rocket can be re-used.
- However, not always the re-use is possible.
- As a data scientist for the competitor Space Y, an attempt is made to find the favorable scenarios, where the inexpensive methodology can be adopted with highest possibility of success
- Problems we want to find answers
- The parameters in the launches executed by Space X.
- Based on the parameters, prepare a model which can effectively predict when the first stage is re-used by Space X.
- More favorable parameters will increase possibility to provide the launches at lesser cost , on more occasions thereby providing a perfect competition to Space X

Section 1

Methodology

Methodology
Executive Summary

- Data collections – using API and webscraping
- Data Wrangling
- Exploratory data analysis using visualisation and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models – LR,KNN,SVM,DT models evaluated

Data Collection

Request to SPACE API

Request and parse the SpaceX launch data

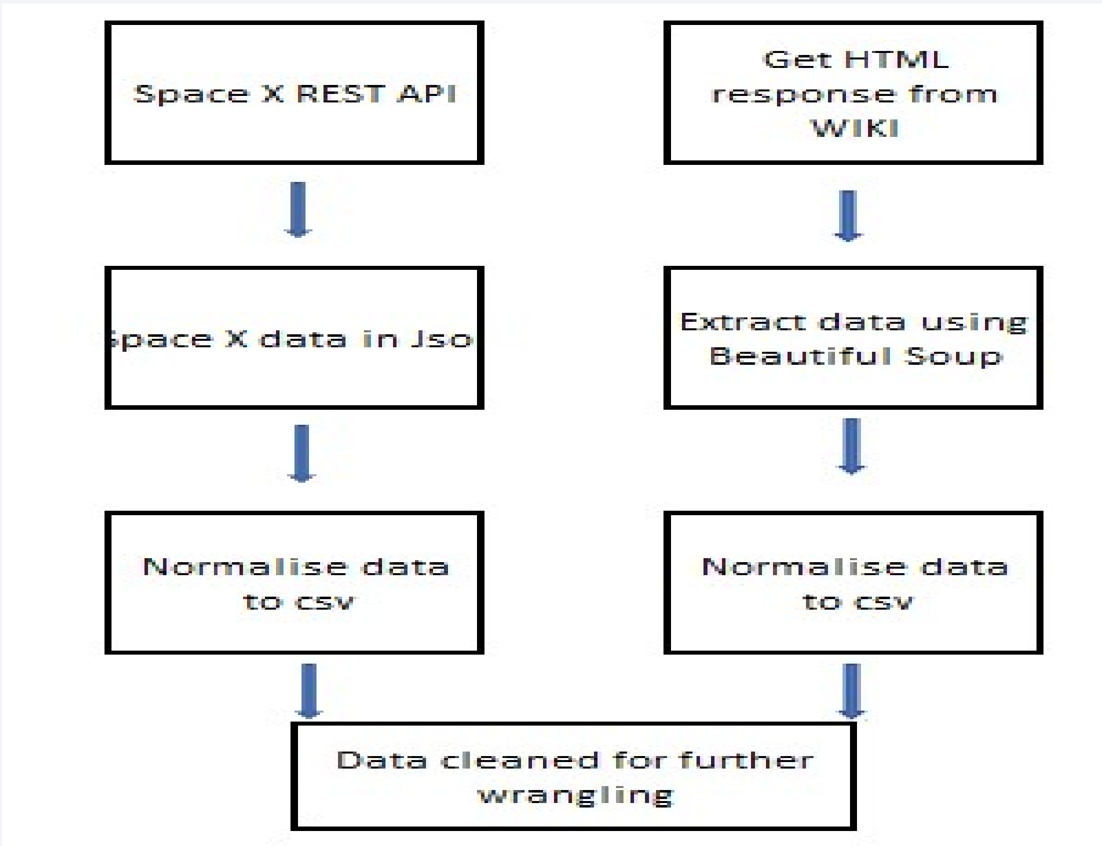
Web scrapping using Beautiful Soup

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
data =pd.json_normalize(response.json())
data.head()
```

Data Collection – SpaceX API




```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
data =pd.json_normalize(response.json())
data.head()
```

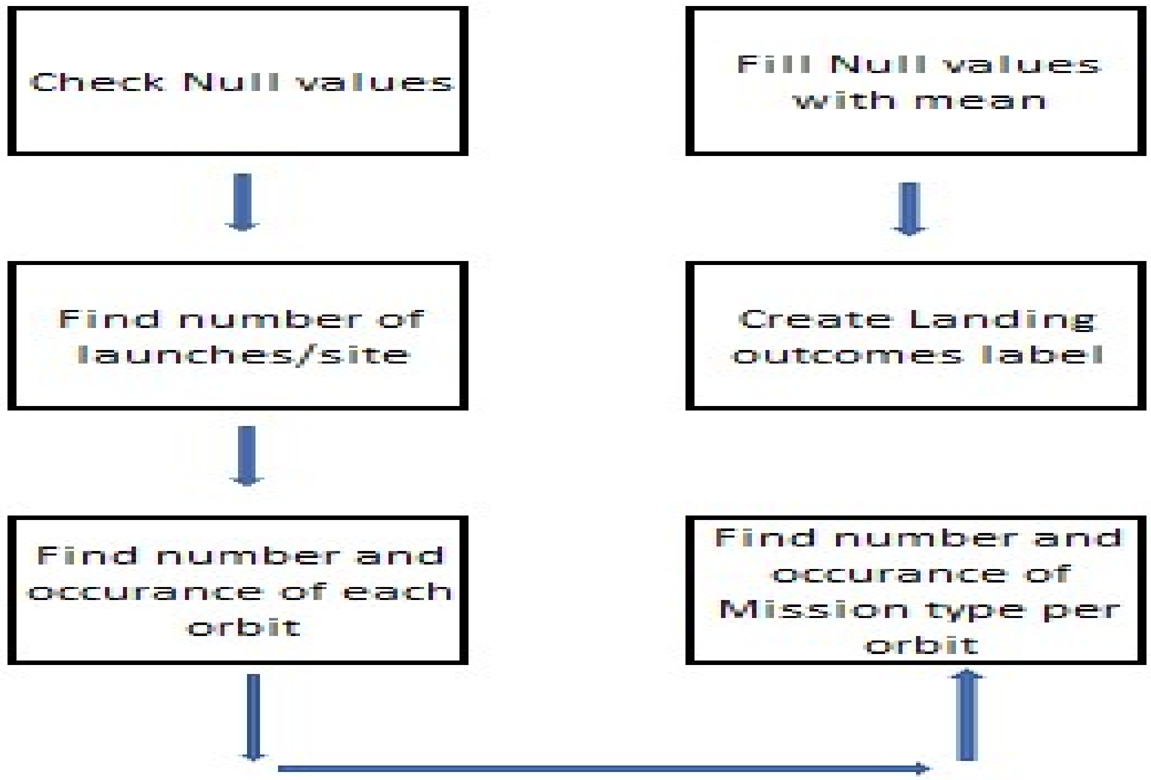
```
# Call getBoosterVersion
getBoosterVersion(data)
```

```
BoosterVersion[0:5]
```

```
['Falcon 1', 'Falcon 1', 'Falcon 1', 'Falcon 1', 'Falcon 9']
```

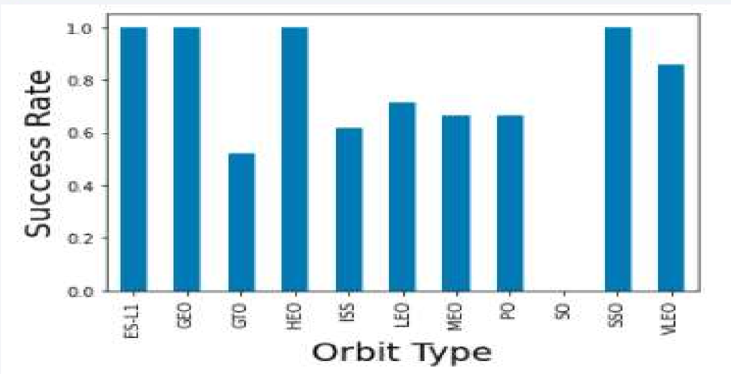

https://gist.github.com/Chetrypapu/1c578b6a58fa52420dc66efe84ce351c

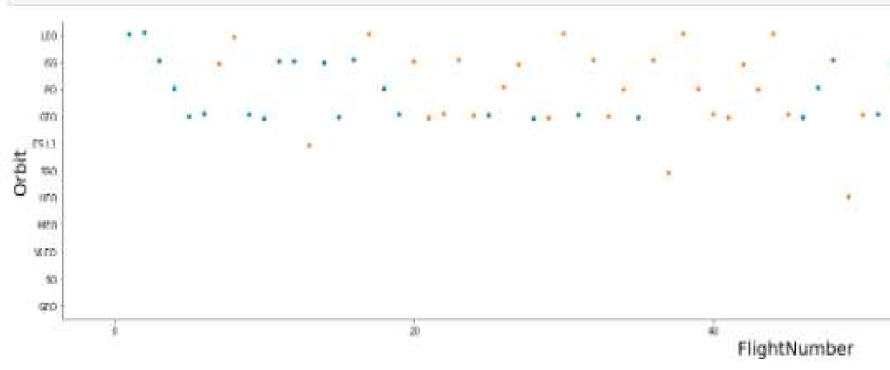
Data Wrangling

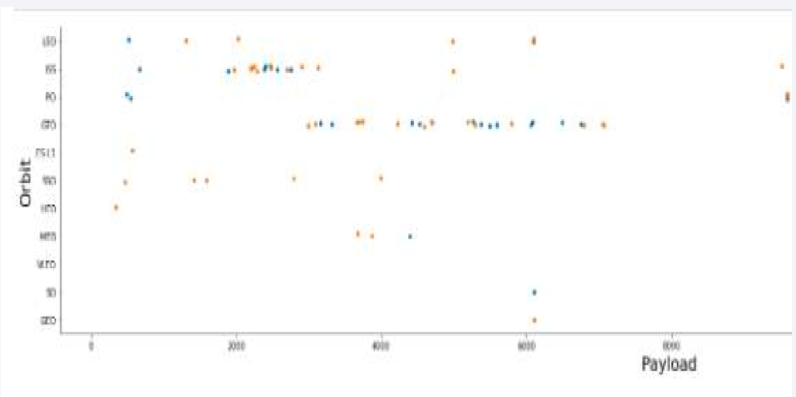


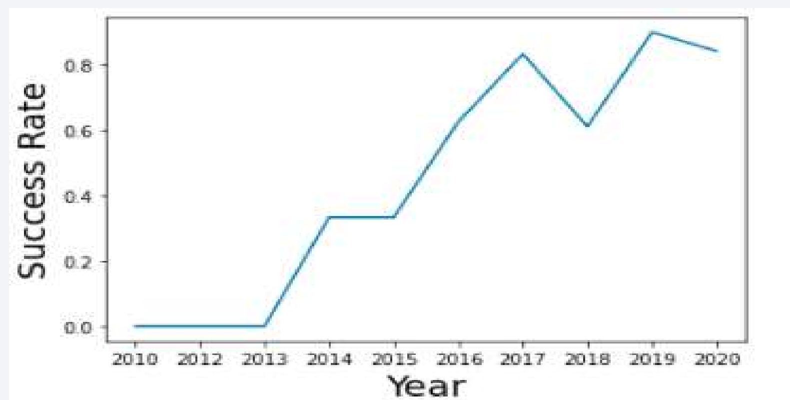
https://gist.github.com/Chetrypapu/c29daed5988f9966857e58971e312889

EDA with Data Visualization









https://gist.github.com/Chetrypapu/91457250fbc5e91e7a844125767d4fd7

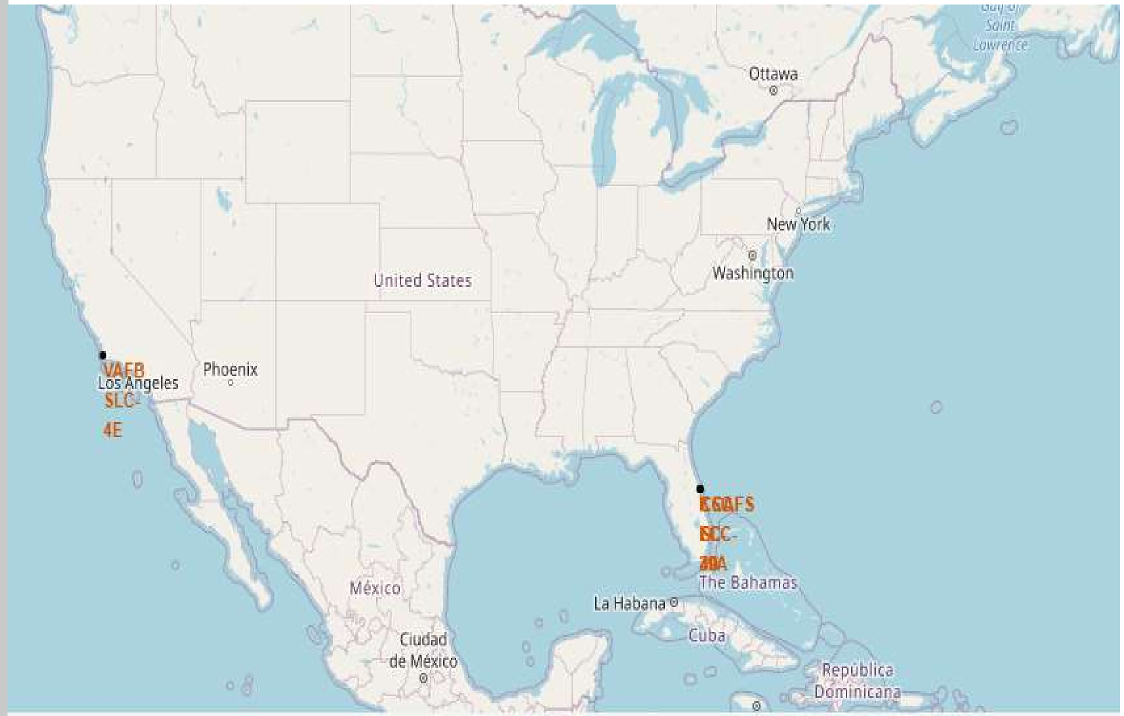
EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

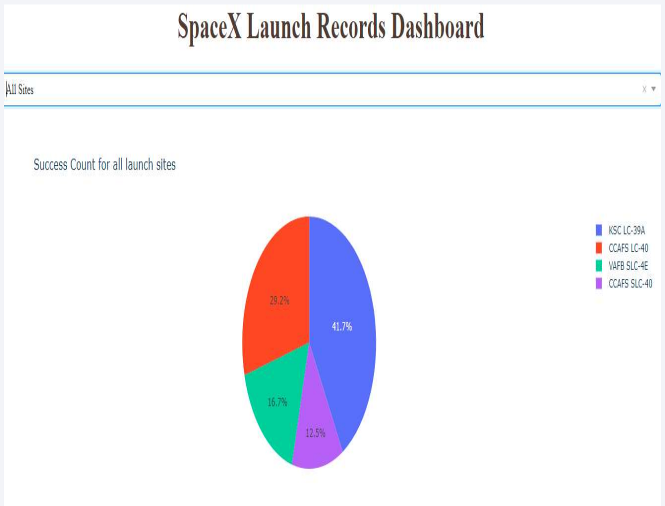
- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first successful landing outcome in ground pad was acheived.**
- **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. Use a subquery**
- **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**

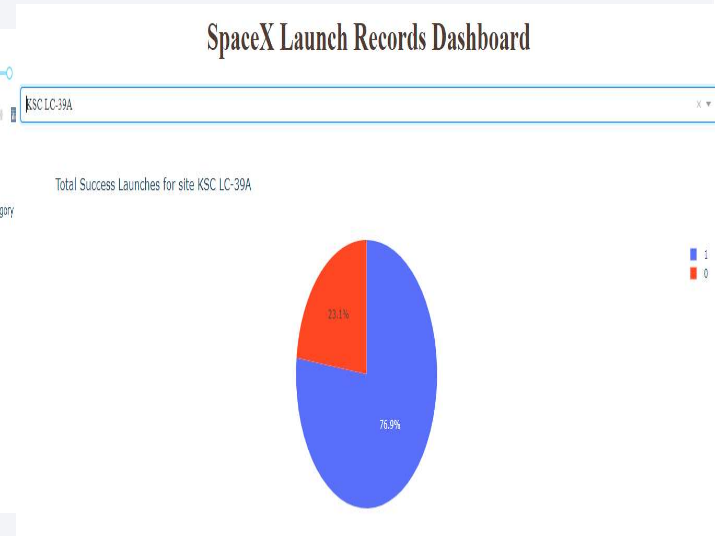
Interactive Map with Folium

- Map markers have been added to the map with aim to finding an optimal location for building a launch site
- https://gist.github.com/Chetrypapu/45e0f8599133e22e09e1e5d6e2ced4db



Build a Dashboard with Plotly Dash

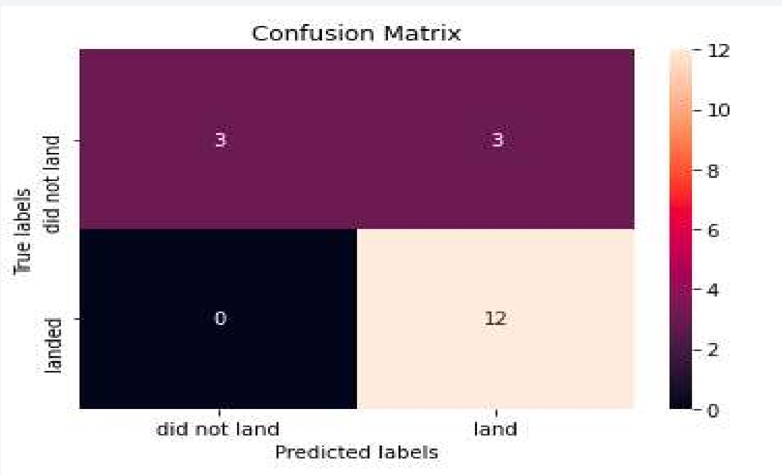




https://gist.github.com/Chetrypapu/37a181d3f53278d9f9ee5934f1e0c395

Predictive Analysis (Classification)

- The SVM, KNN and Logistic Regression model achieved the highest accuracy at 83.33%, while the SVM performs the best in terms of Area Under the curve at 0.958




```
test set accuracy : 0.8333333333333334
```

https://gist.github.com/Chetrypapu/50c512bc7980af9de217524fdf5a2579

Results

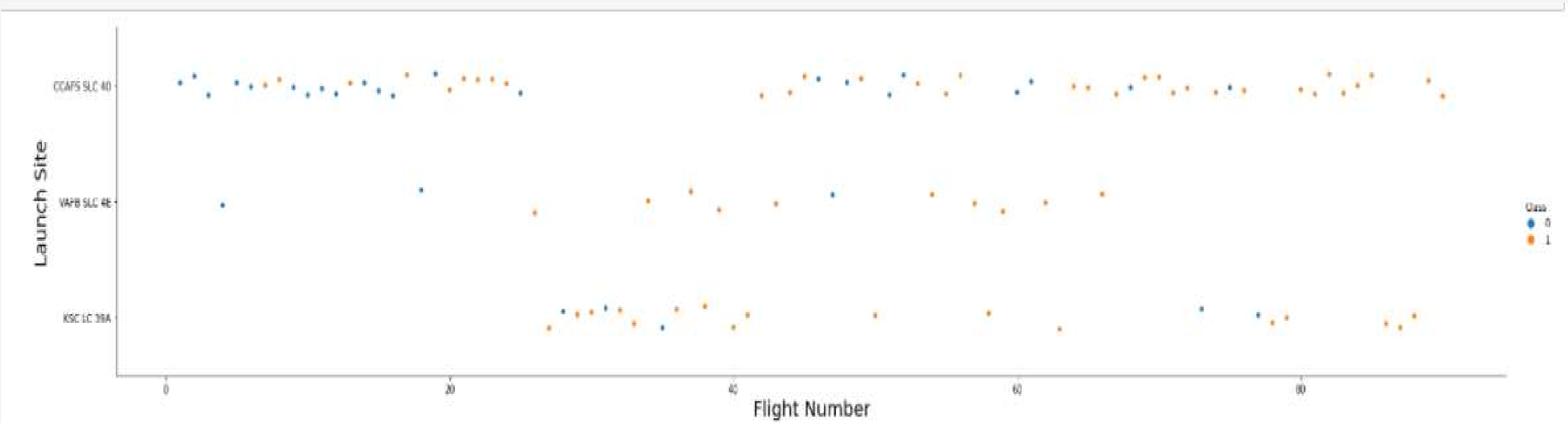
- The SVM, KNN and Logistic Regression models are the best in terms prediction accuracy for this dataset
- Low weighed payloads perform better than the heavier payloads
- The success rate for SpaceX launches is directly proportional time in years they will eventually perfect the launches
- KSC LC 39A had the most successful launches from all the sites
- Orbit GEO , HEO, SSO, ES L1 has the best success rate

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

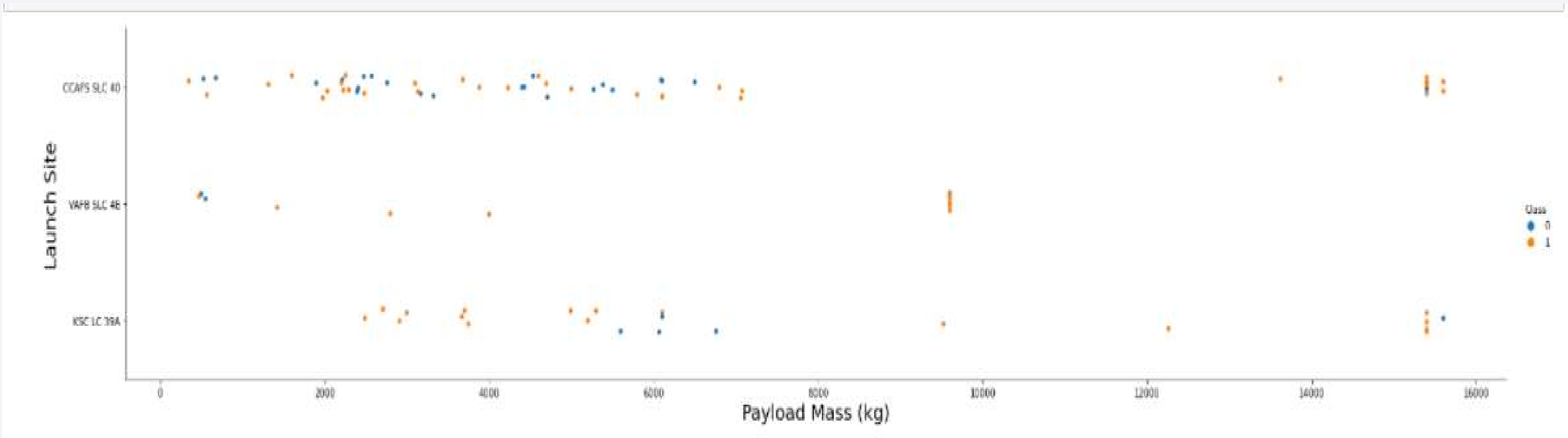
- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites

Payload vs. Launch Site

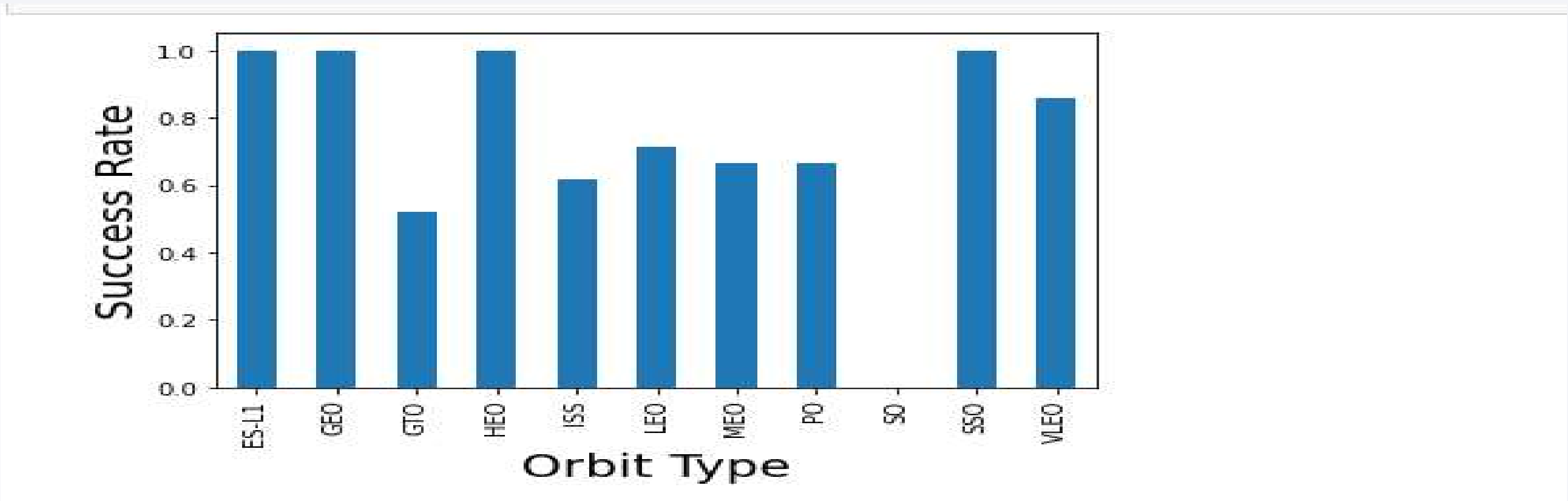
- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Most of the payloads with lower mass have been launched from CCAFS SLC 40

Success Rate vs. Orbit Type

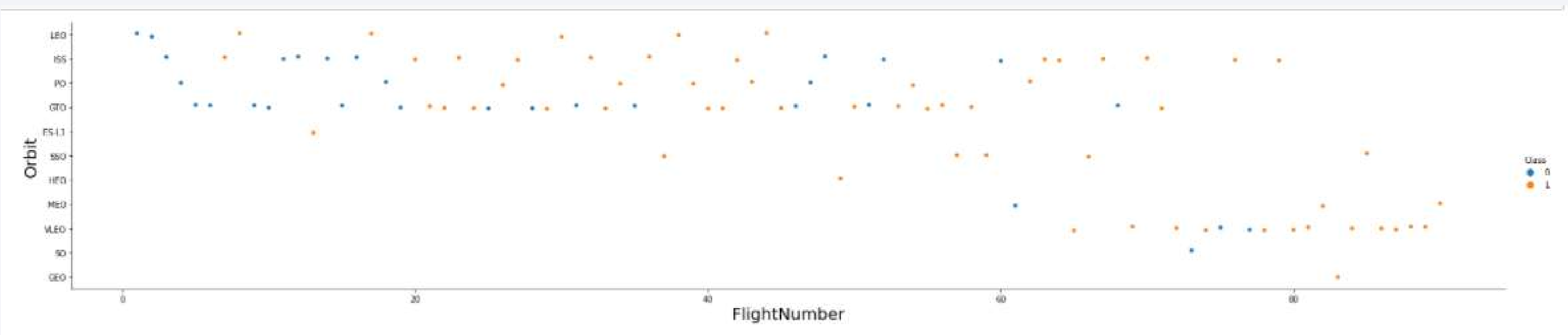
- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations



The orbits ES L1, GEO, HEO and SSO have higher success rates

Flight Number vs. Orbit Type

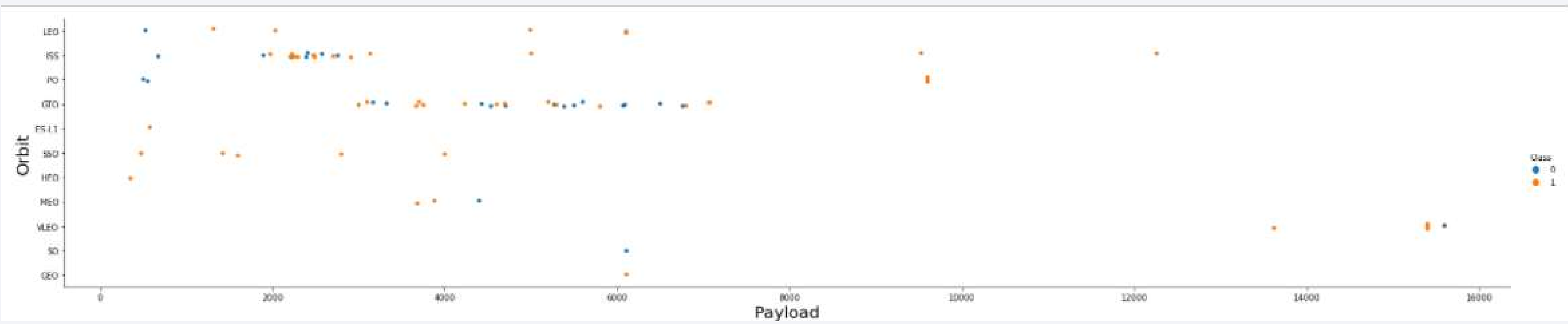
- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



A trend can be observed of shifting VLEO launches over the years

Payload vs. Orbit Type

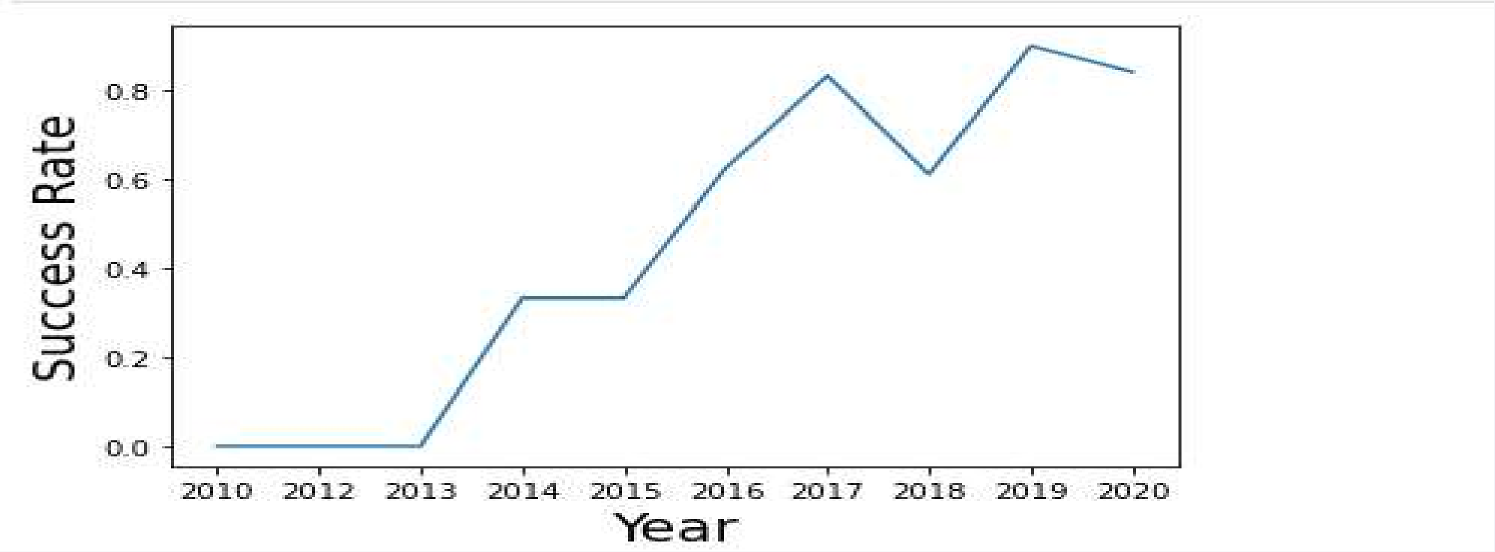
- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



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

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations



The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

- %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'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

- %SQL select * from SPACEXTBL where LAUNCH_SITE like CCA%' limit 5

	Date	Time_UTC	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	19-02-2017	2021-07-02 14:39:00.0000000	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1	16-03-2017	2021-07-02 06:00:00.0000000	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2	30-03-2017	2021-07-02 22:27:00.0000000	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
3	01-05-2017	2021-07-02 11:15:00.0000000	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
4	15-05-2017	2021-07-02 23:21:00.0000000	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

- %SQL select sum(PAYLOAD_MASS_KG_) from SPACEXTBL where CUSTOMER ='NASA (CRS)'

	Total Payload Mass
0	45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

- %SQL select avg(PAYLOAD_MASS_KG_) from SPCEXTBL where BOOSTER_VERSION = ‘F9 v 1.1’

	Average Payload Mass
0	2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

%SQL select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success(ground pad)'

	Date which first Successful landing outcome in drone ship was acheived.
0	06-05-2016

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
- Present your query result with a short explanation here
- %SQL select BOOSTER_VERSION from SPACEXTBL where Landing_Outcome ='Success(drone ship)' and PAYLOAD_MASS_KG_ > 4000 and PAYLOAD_MASS_KG_<6000

	Date which first Successful landing outcome in drone ship was acheived.
0	F9 FT B1032.1
1	F9 B4 B1040.1
2	F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

- %SQL select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME ='Success' or MISSION_OUTCOME ='Failure(in flight)'

	Successful_Mission_Outcomes	Failure_Mission_Outcomes
0	100	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

- %SQL select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS_KG =(select max(PAYLOAD_MASS_KG_)from SPACEXTBL)

	Booster_Version	Maximum Payload Mass
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
...	...	...
92	F9 v1.1 B1003	500
93	F9 FT B1038.1	475
94	F9 B4 B1045.1	362
95	F9 v1.0 B0003	0
96	F9 v1.0 B0004	0

97 rows × 2 columns

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here
- %SQL select * from SPACEXTBL where Landing_Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

Month	Booster_Version	Launch_Site	Landing_Outcome
January	F9 FT B1029.1	VAFB SLC-4E	Success (drone ship)
February	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
March	F9 FT B1021.2	KSC LC-39A	Success (drone ship)
May	F9 FT B1032.1	KSC LC-39A	Success (ground pad)

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
- Present your query result with a short explanation here
- %SQL select * from SPACEXTBL where landing_Outcome like ‘Success%’ and (DATE between ‘2010-06-04’ and ‘2017-03-20’) order by date desc

	Successful Landing Outcomes Between 2010-06-04 and 2017-03-20
0	34

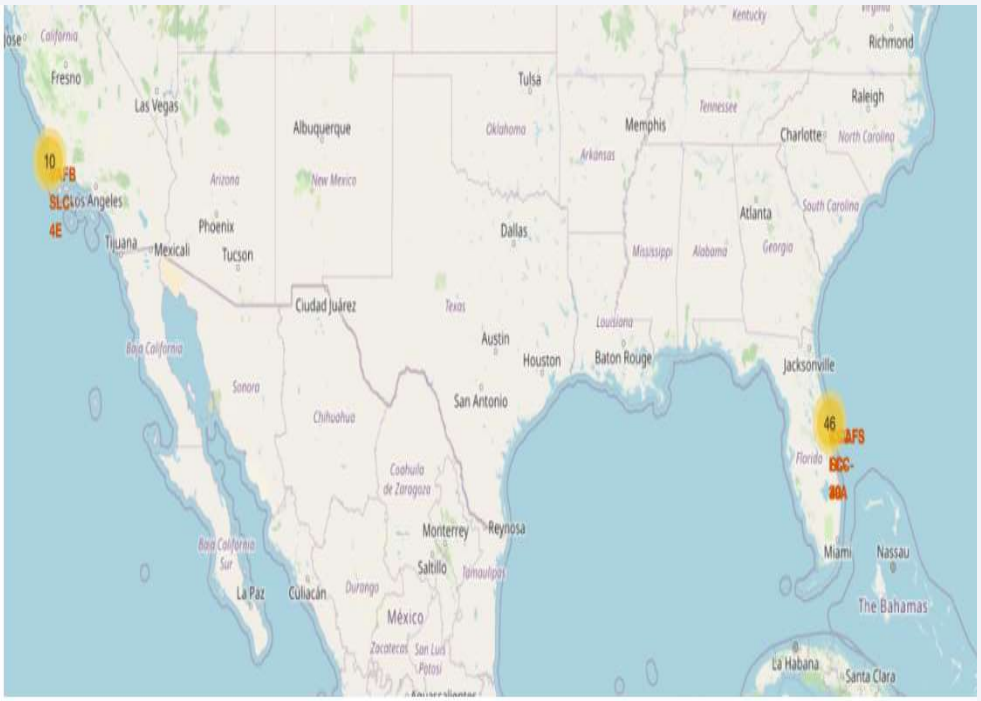
Section 3

Launch Sites Proximities Analysis

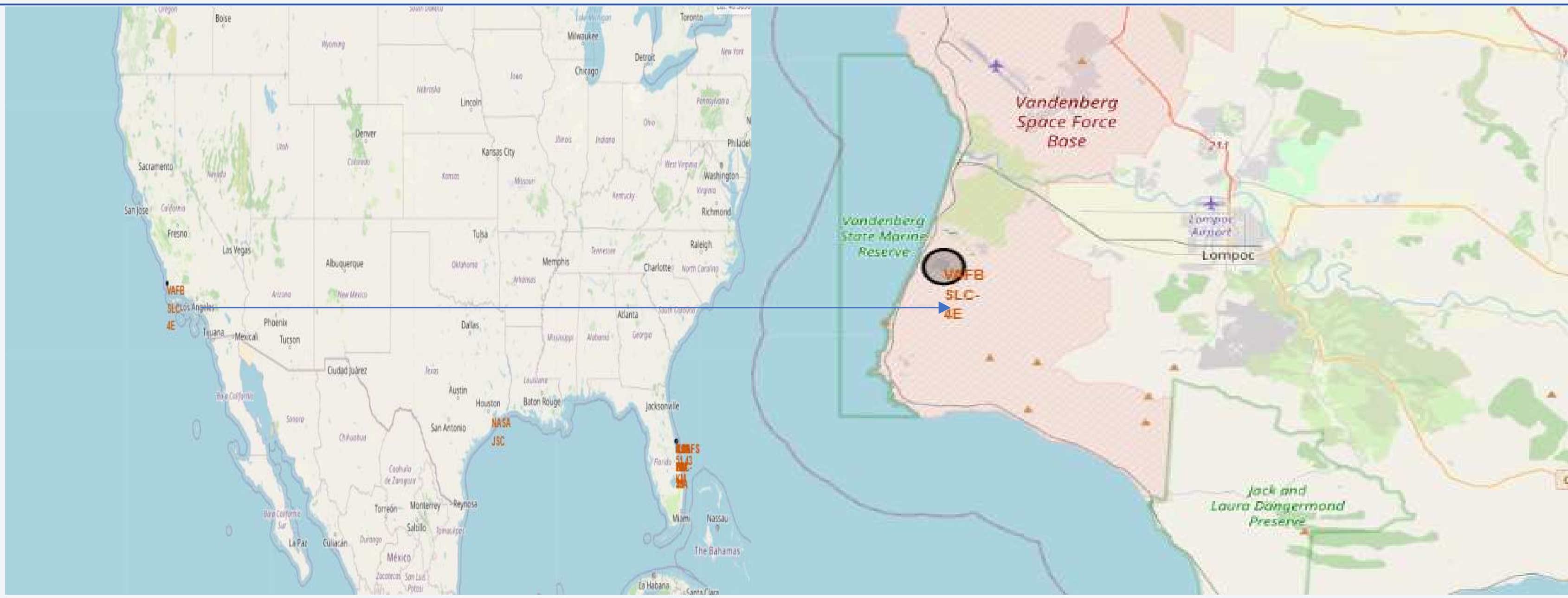
SpaceX Launch Sites marked in Map

The launch sites in the coasts, Florida and California are marked in the map with labels

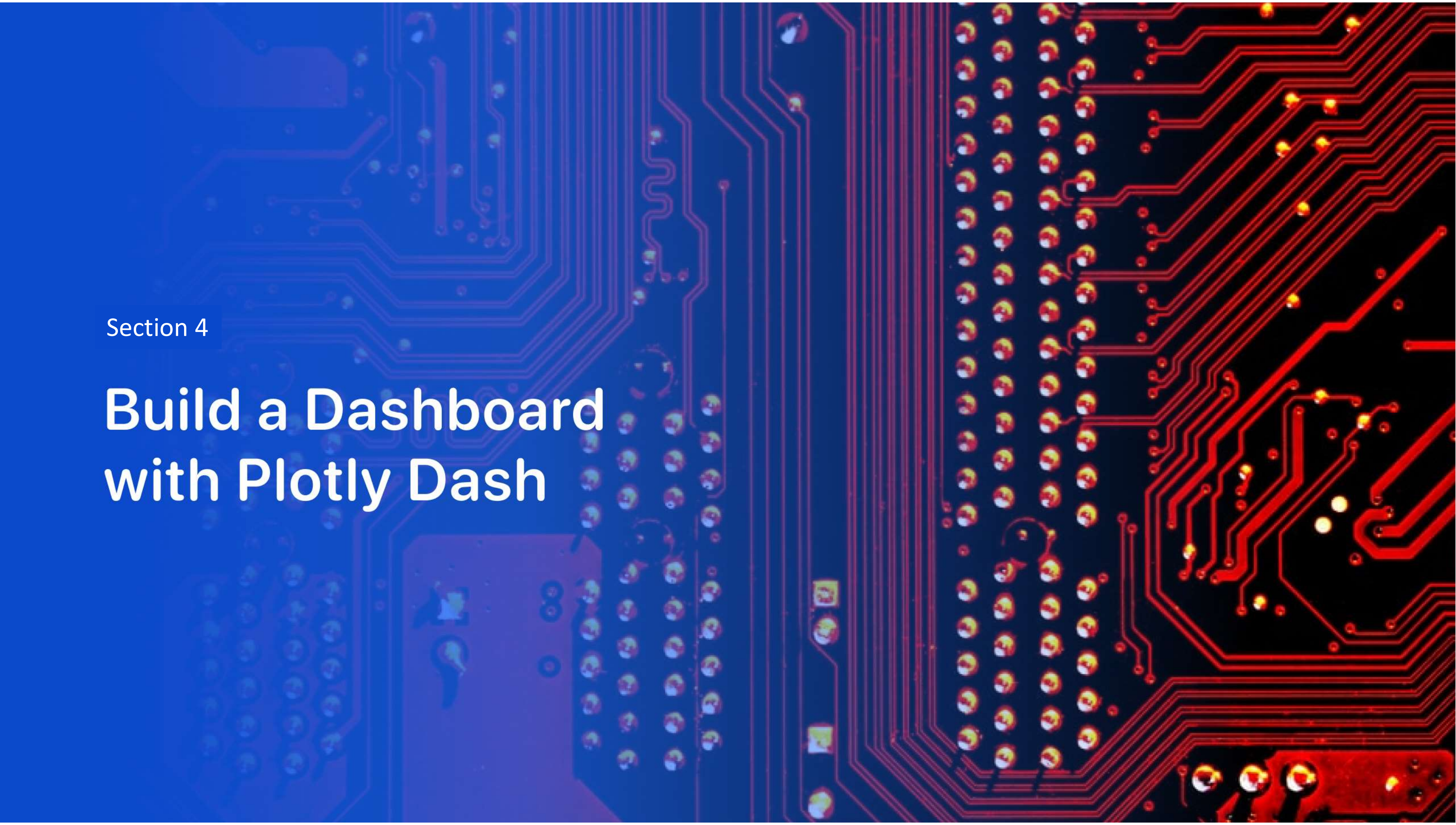
Labelled markers



Launch Sites distance to Landmarks (using Folium maps)



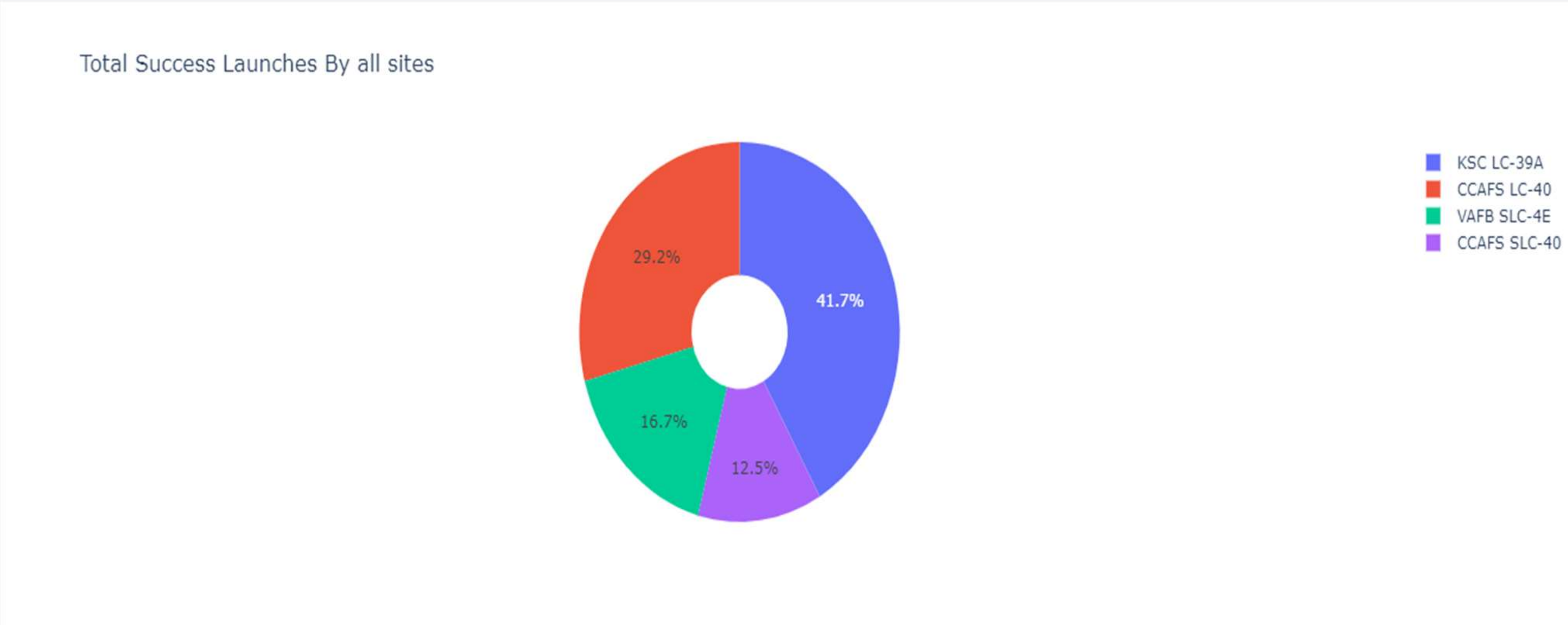
- Are launch sites in close proximity to railways? NO
- Are launch sites in close proximity to highways? NO
- Are launch sites in close proximity to coastline? YES
- Do launch sites keep certain distance away from cities? YES



Section 4

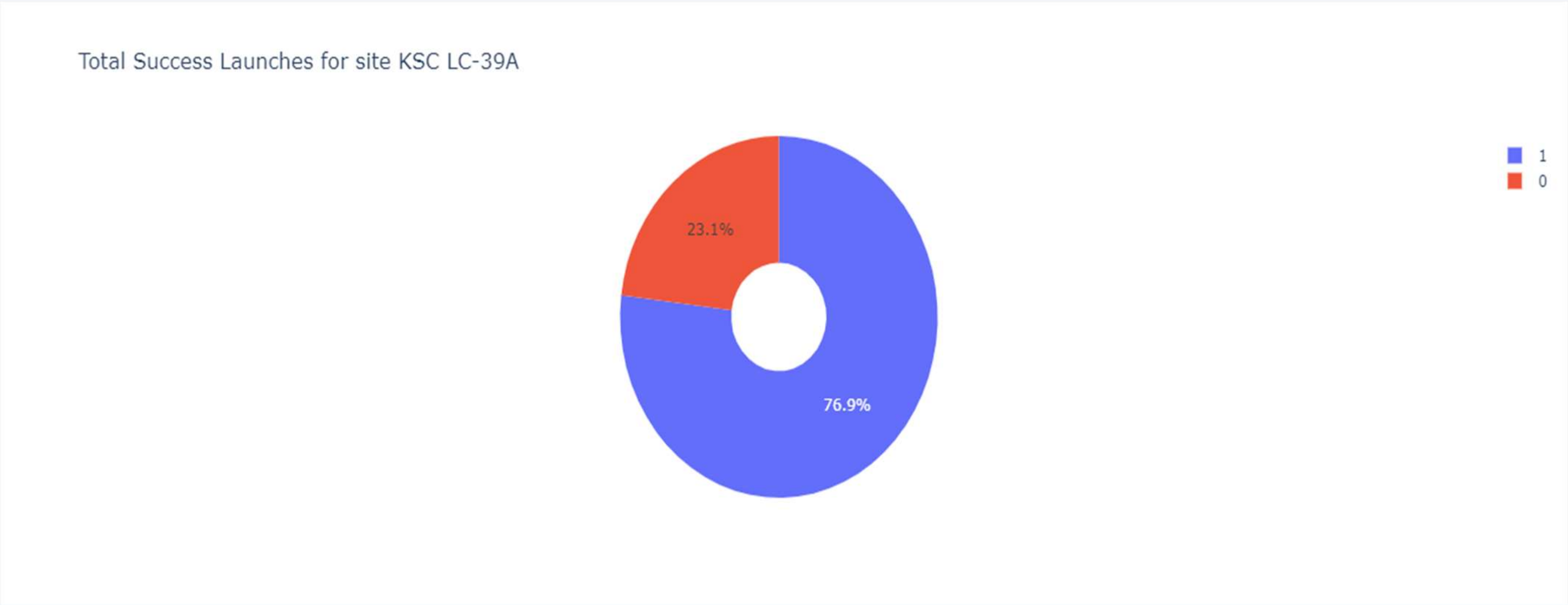
Build a Dashboard with Plotly Dash

Total Launch Success by all Sites



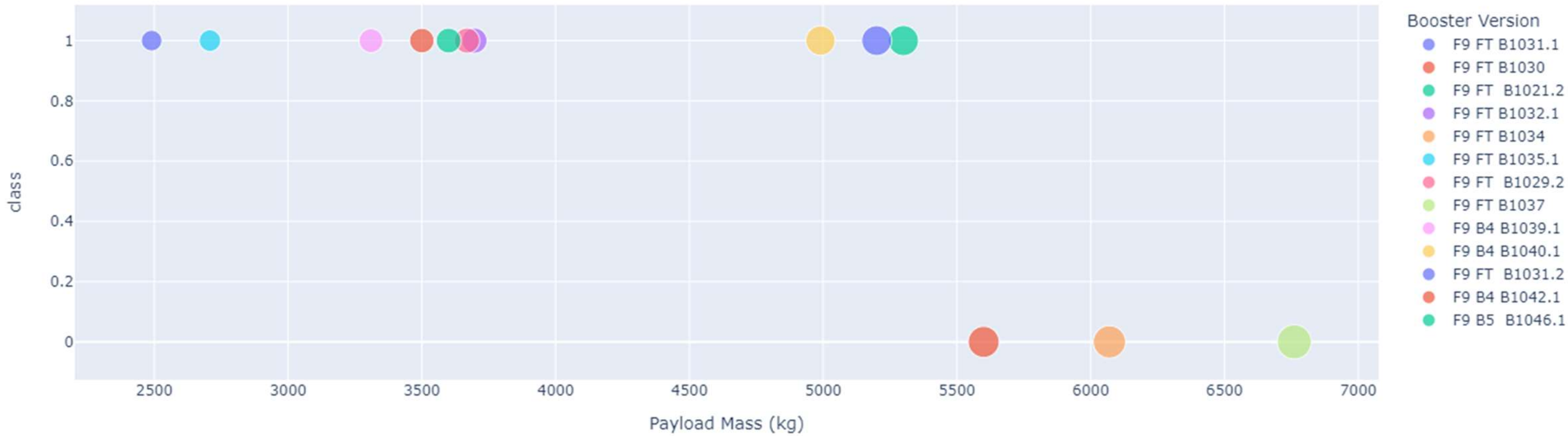
It is evident that KSC LC-39A has the most successful launches among all sites

Contd. Most successful Launch site drill down



Success and Failure rates for launch site
KSC LC-39A which is the most successful launch site

Payload vs Launch outcome analysis

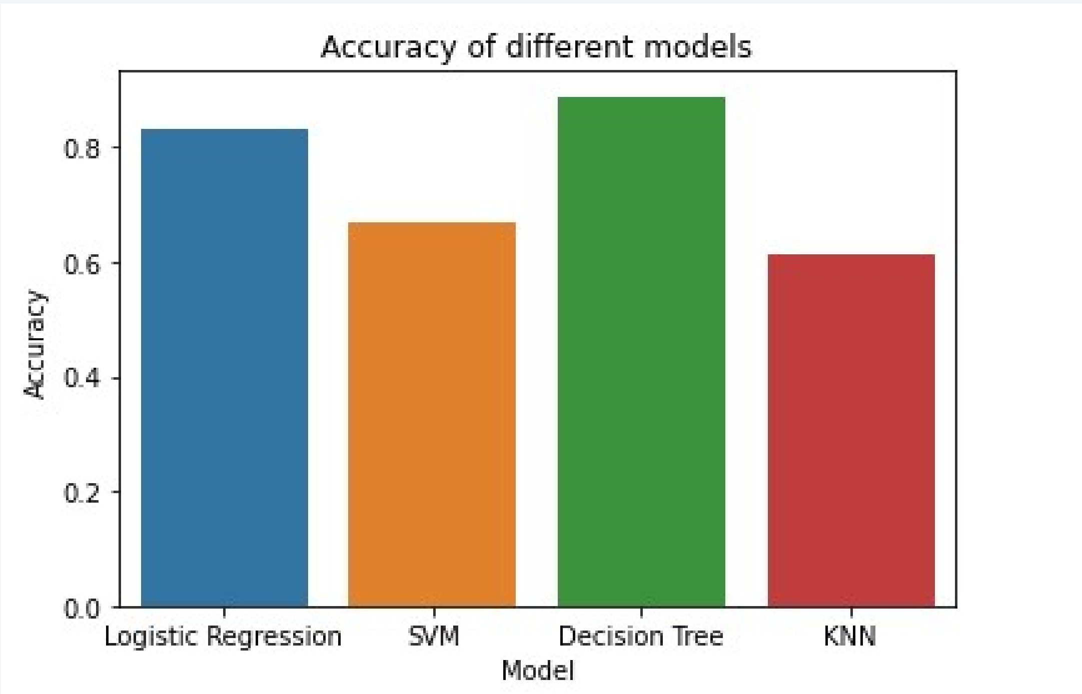


It is evident from the visualization that success rates for low weighed payloads are higher than heavy payloads

Section 5

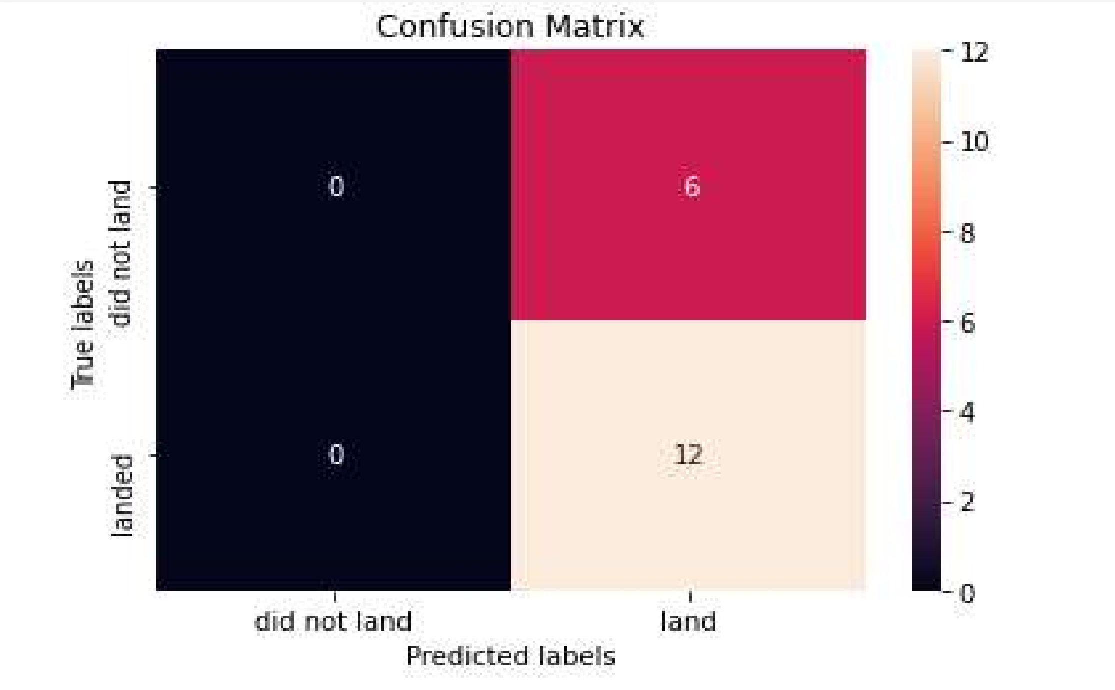
Predictive Analysis (Classification)

Classification Accuracy



The Decision tree is marginally more accurate than the others

Confusion Matrix



- The major problem observed from the confusion matrix is the false positive data points

Conclusions and Take Away

- The TREE classifier is the most accurate model for the dataset analyzed
- Low payloads have higher success rates than heavy payloads
- KSC LC-39A has the most successful launches from all sites
- Orbit GEO, HEO, SSO, ES-L1 has the best success rate
- The success rate of SPACEX launches is improving over time


Thank you!