Winning Space Race with Data Science

Sai Gowtham
27th May, 2023

Outline

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

Executive Summary

- The classifications methods used are logistic regressions, support vector machines, decision tree algorithm and k nearest neighbors.
- The f1 scores we got by these methods are 0.833, 0.833, 0.777, 0.777 respectively. The accuracies are 84.64 %, 84.82 %, 90%, 84.82 % respectively.

Introduction

- This project is all about predicting the first stage landing outcome of the falcon 9 rockets launched by SpaceX.
- We want to find the answer that the first stage will land successfully or not depending on the many parameters like launch site, functions of landing gears, date and flight number.

Section
1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is collected using the API and web scraping from the wiki page.
- Perform data wrangling
  - Data collected is cleansed and a new column, ‘Class’ is added declaring whether the landing was successful or not.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models using logistic regression, svm, decision tree and KNN.

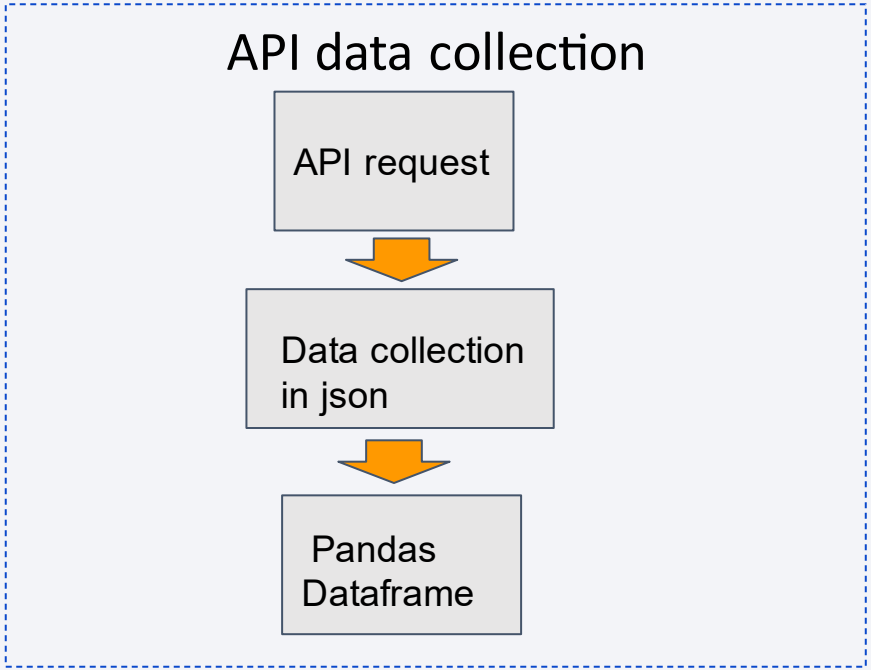
Data Collection

- Data sets are collected by SpaceX API and by web scraping from the wikipedia page.

Data Collection – SpaceX API

- Data is collected from the spaceX api and then converted into a json. Then the data is finally converted into a pandas dataframe.
- GitHub URL :

  link



Data Wrangling

- Data were processed in a dataframe.
- The main objective of this data wrangling is to get a new column ‘Class’ which reflects what was the result of the landing. 0 for failure and 1 for success.
- We also obtain a orbit wise success and failure overview.
- GitHub URL : link

EDA with Data Visualization

- After 52nd flights the successful landings gets more prominent.
- After 63rd flights the payload capacity increased drastically.
- CCAFS 40 SLC launched most of the missions.
- SSEO, VLEO have better success rate with comparatively higher missions.
- After 2013 the success rate increases gradually and it is around 80% in 2019

- GitHub URL : link

EDA with SQL

- There are FOUR distinct launch sites.
- Average payload mass carried by **booster version F9 v1.1** 2928.4 kg.
- Successful landing first happened in 01-05-2017
- There are FOUR boosters those have a payload mass between 4000 and 6000 kg.

- GitHub URL : link

Build an Interactive Map with Folium

- The launch sites are marked with circles.
- Out of four launch sites , three are in the eastern coast and nearer to the equator.
- The launch sites are nearer to airport, rail, road and sea ports.
- GitHub URL : link

Build a Dashboard with Plotly Dash

- Most of the successful launches are from KSC LC 39A
- The sunburst chart shows the number of launches from each launching centres and their shares of success and failure landings.
- Then there is the pie chart for the launch sites with the number of launches and the payload selector.

- GitHub URL : link

Predictive Analysis (Classification)

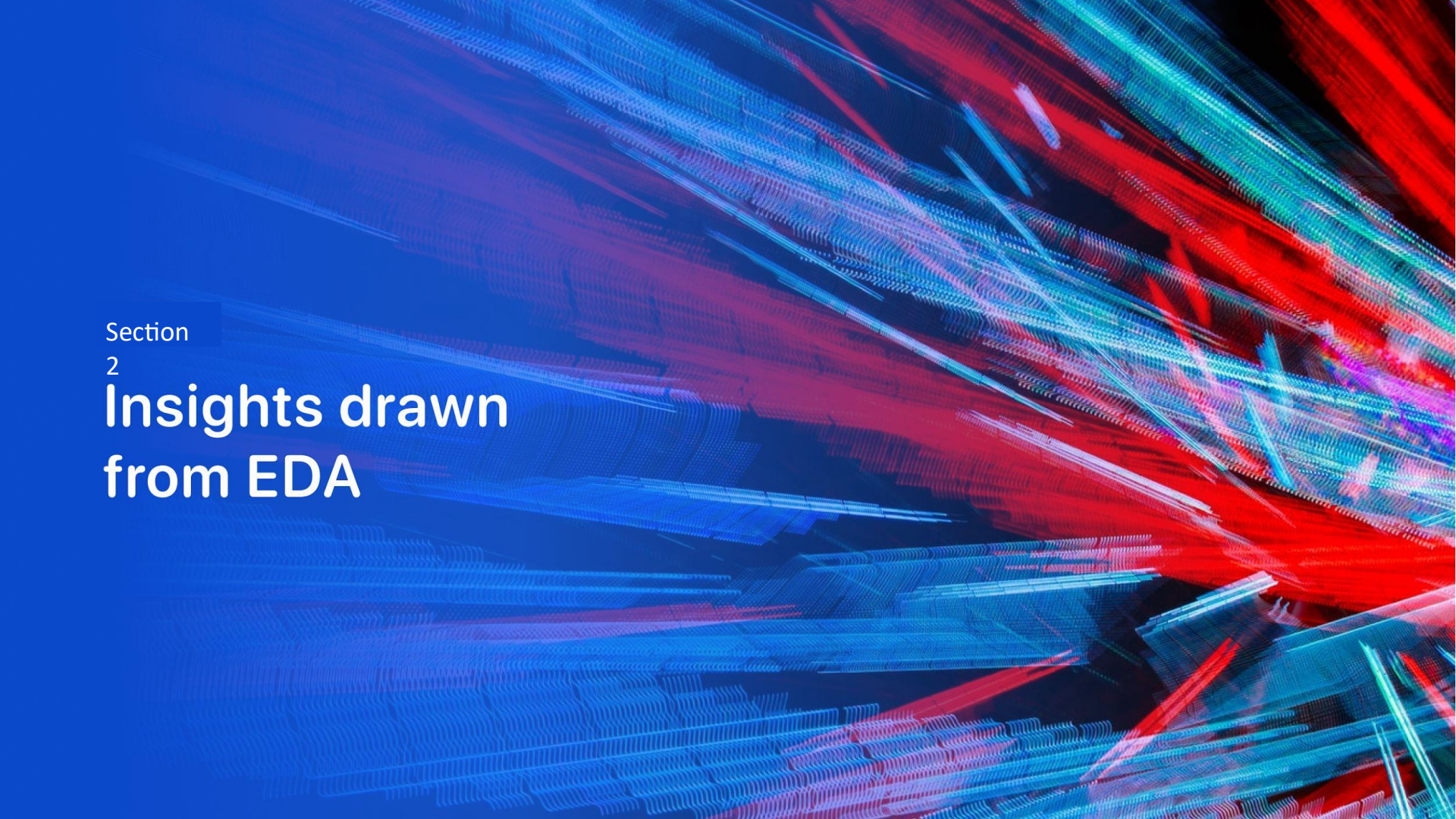
- Predicting successful landing here is a classification problem. So we used logistic regressions, support vector machines, decisions tree classifier and k nearest neighbors methods for the predictive analysis.
- The accuracy in predicting is the best in decision tree model with 90%
- The f1 score is better in logistic regression and support vector machines with 0.83.
- GitHub URL : link

Results

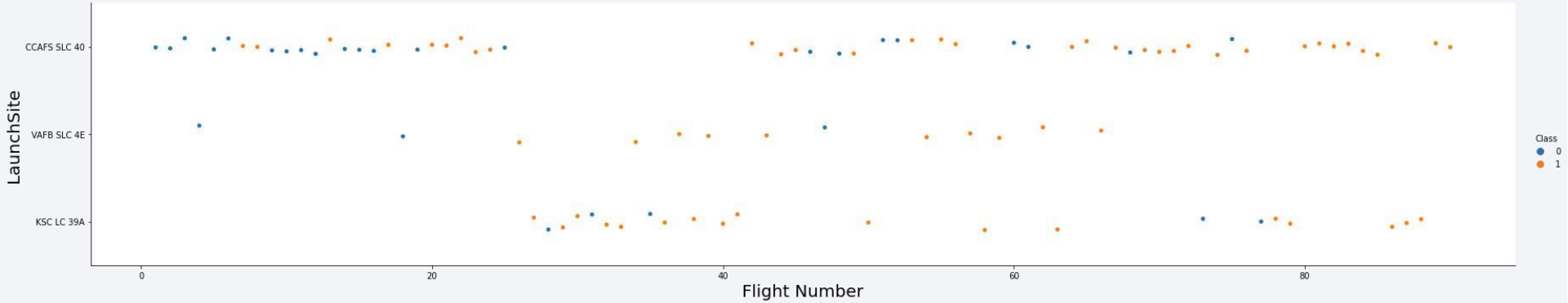
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

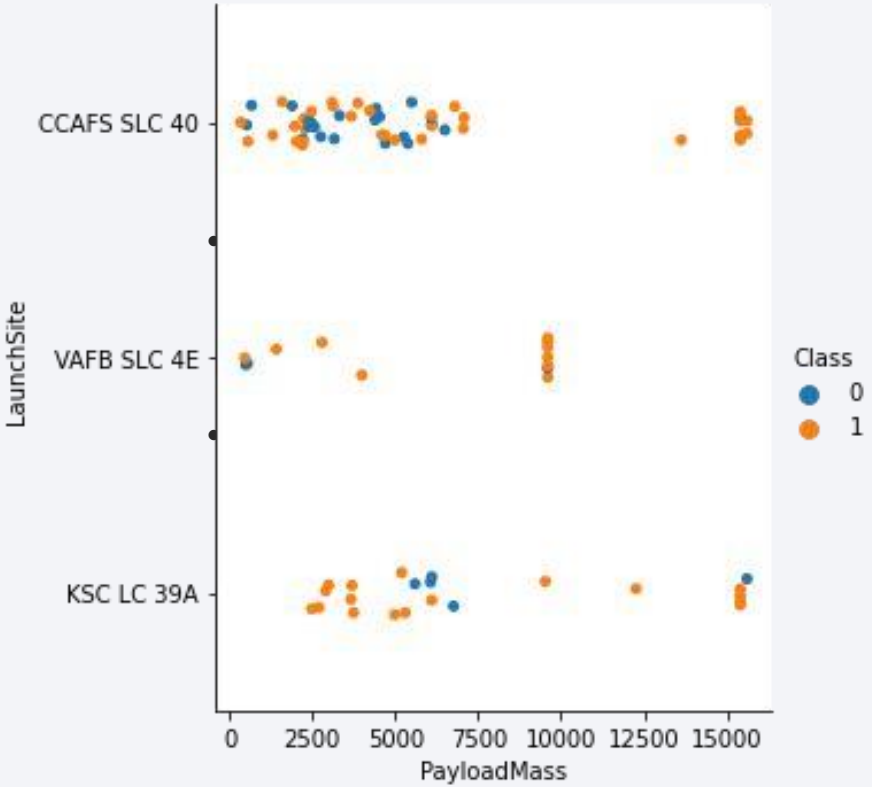


Flight Number vs. Launch Site



Maximum launches happened in CCAFS SLC 40. Then KSC LC 39A.
Though less launches are from VAFB SLC 4E, success rate is better here.

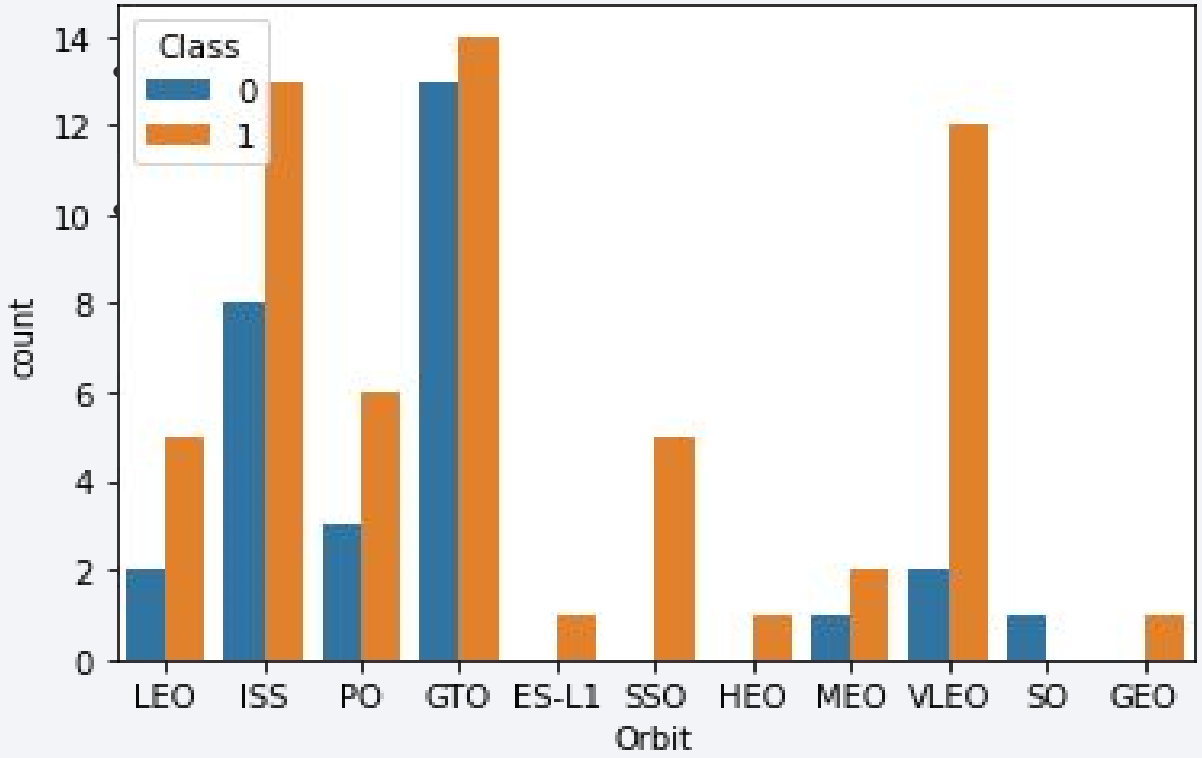
Payload vs. Launch Site



Payloads with more than 10000 kgs are not launched from VAFB SLC 4E.

Higher payload masses have a better successful landing probability. It is because the higher payload masses are launched when the technology is perfected.

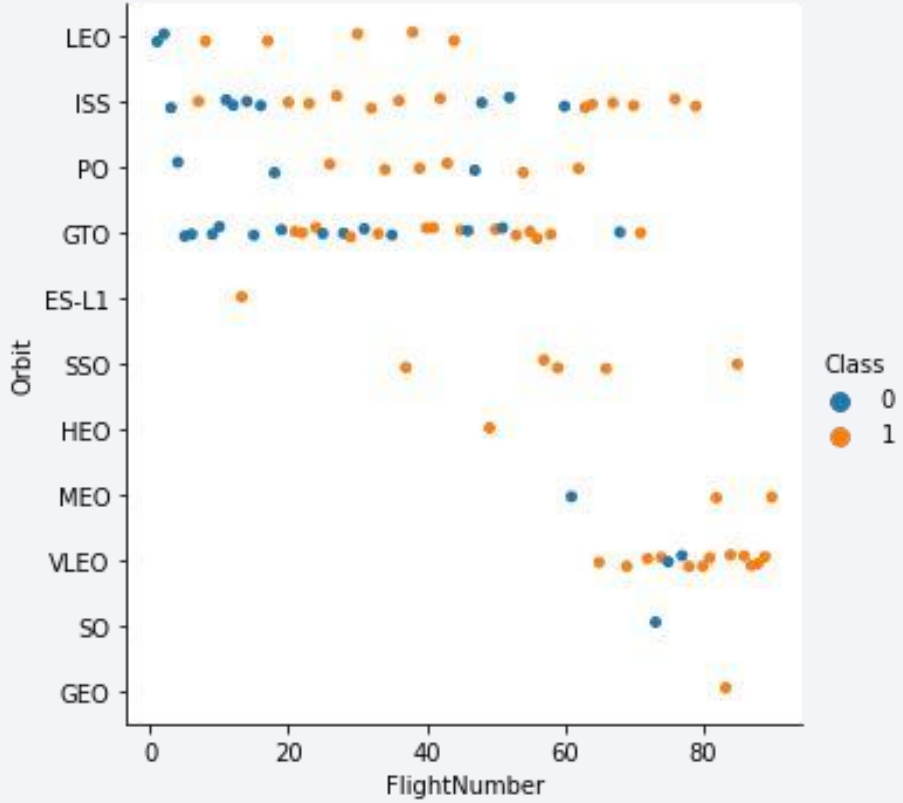
Success Rate vs. Orbit Type



GTO has nearly 50% success rate with a considerable amount of launches.

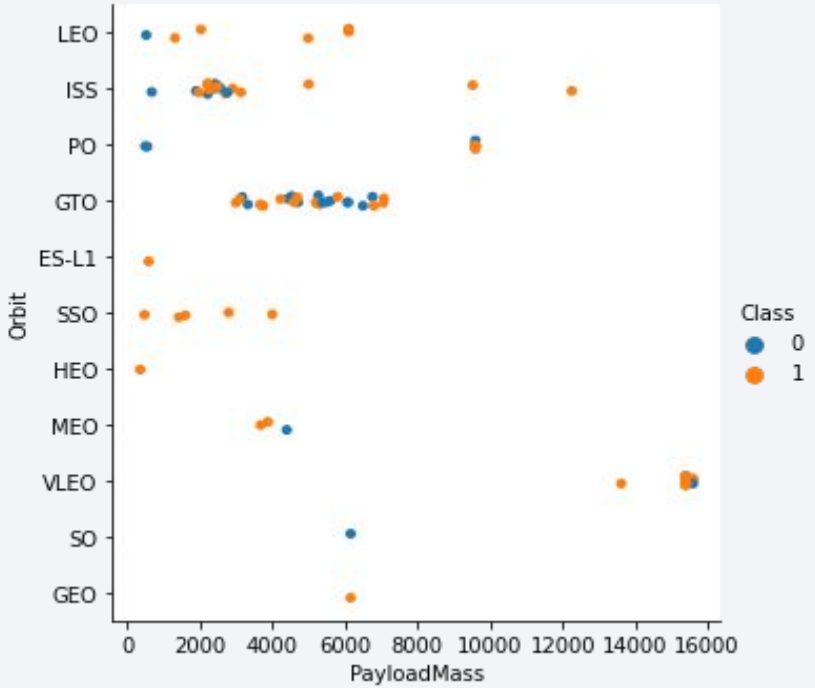
VLEO has fantastic success rate of 12:2.

Flight Number vs. Orbit Type



As more and more flights are experimented, irrespective of successful or unsuccessful landing, the launch has been towards the VLEO orbit.

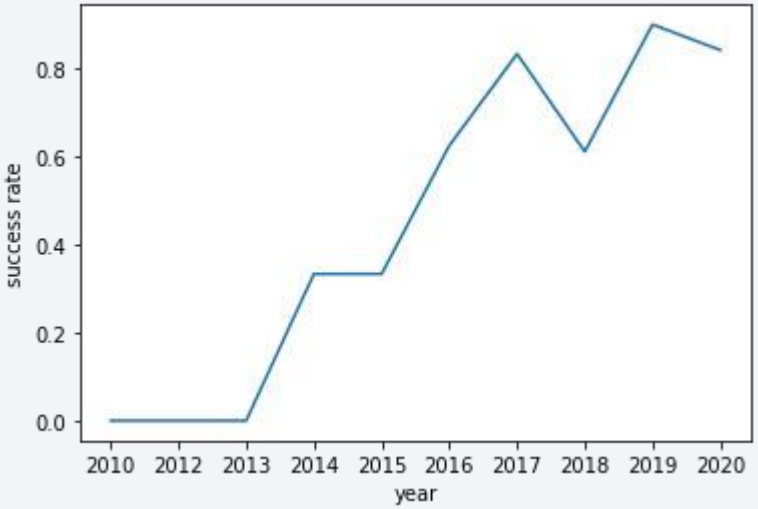
Payload vs. Orbit Type



In the orbit GTO the payload masses ranges between 2000 and 6000 kg.

Highest payload masses are launched in the VLEO orbit

Launch Success Yearly Trend



2013 onwards the success rate starts climbing up.

2014-15 the success rate flattens.

2015-17 the success rate climbs drastically and tops at 2017.

2018 the rate droops down temporarily but it regains and breaches it previous record and reaches top at 2019.

All Launch Site Names

The names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

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

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

```
[ ]  1  %%sql
     2  SELECT SUM(PAYLOAD_MASS__KG_)
     3  FROM SPACEXTBL
     4  WHERE Customer='NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
SUM(PAYLOAD_MASS__KG_)
45596
```

Average Payload Mass by F9 v1.1

```
[ ]  1  %%sql
     2  SELECT AVG(PAYLOAD_MASS__KG_)
     3  FROM SPACEXTBL
     4  WHERE Booster_Version = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
[ ]  1  %%sql
     2  SELECT MIN(Date)
     3  FROM SPACEXTBL
     4  WHERE "Landing _Outcome" LIKE 'success%'
```

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

MIN(Date)
01-05-2017

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

```
[ ]  1  %%sql
     2  SELECT DISTINCT "Booster_Version"
     3  FROM SPACEXTBL
     4  WHERE "Landing _Outcome" LIKE '%success%' AND
     5  "Landing _Outcome" LIKE '%drone ship%' AND
     6  "PAYLOAD_MASS__KG_" >4000 AND
     7  "PAYLOAD_MASS__KG_" <6000
```

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

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

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

```
[ ]  1  %%sql
     2  SELECT (SELECT COUNT("Mission_Outcome")
     3  FROM SPACEXTBL
     4  WHERE Mission_Outcome LIKE '%success%') AS 'SUCCESS',
     5  (SELECT COUNT("Mission_Outcome")
     6  FROM SPACEXTBL
     7  WHERE Mission_Outcome LIKE '%failure%') AS 'FAILURE';
```

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

SUCCESS	FAILURE
100	1

Boosters Carried Maximum Payload

```
[ ]  1  %%sql
     2  SELECT "Booster_Version", "PAYLOAD_MASS__KG_"
     3  FROM SPACEXTBL
     4  ORDER BY "PAYLOAD_MASS__KG_" DESC
     5  LIMIT 12 ;
```

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

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

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

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

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

```
[ ]  1  %%sql
     2  SELECT SUBSTR(Date,4,2) as 'Month' , "Landing _Outcome", Booster_Version, Launch_Site
     3  FROM SPACEXTBL
     4  WHERE SUBSTR(Date,7,4)=='2015' and
     5  "Landing _Outcome" LIKE '%failure%' and
     6  "Landing _Outcome" LIKE '%drone ship%';
```

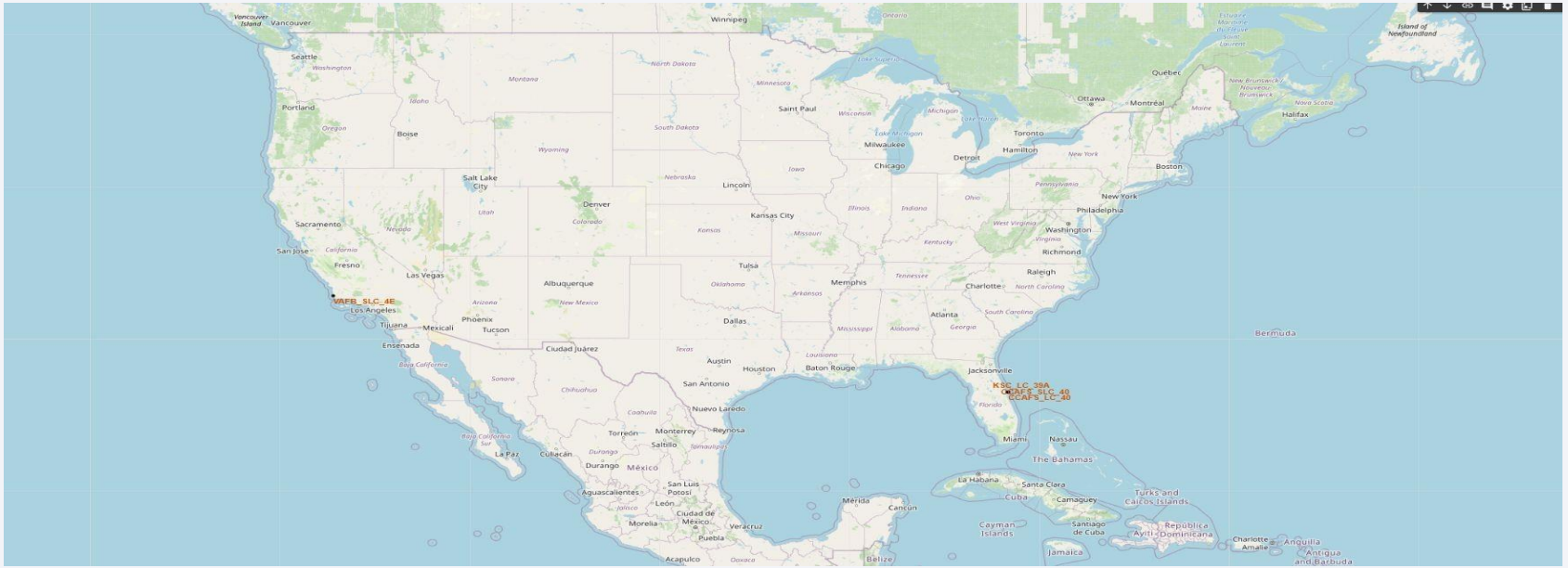
```
 * sqlite:///my_data1.db
Done.
```

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

Section 3

Launch Sites Proximities Analysis

Launch Site Locations



Launch sites are located near the sea in west and east coasts.
Launch sites in the east coast are nearer to the equatorial plane, as compared to the west coast one.

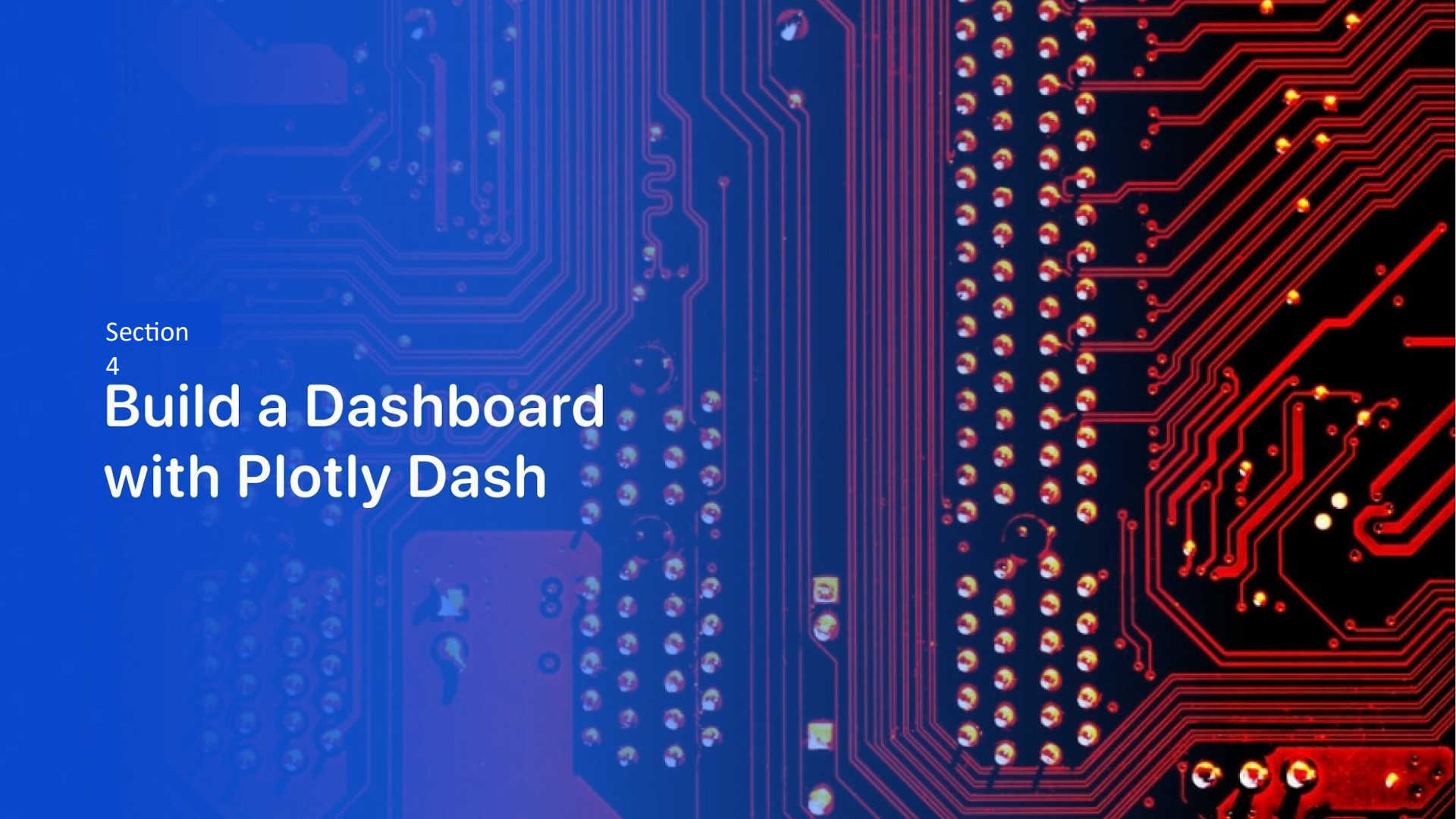
Number of launches from Launch Sites

The launches are more in number in the launch sites in eastern coast.

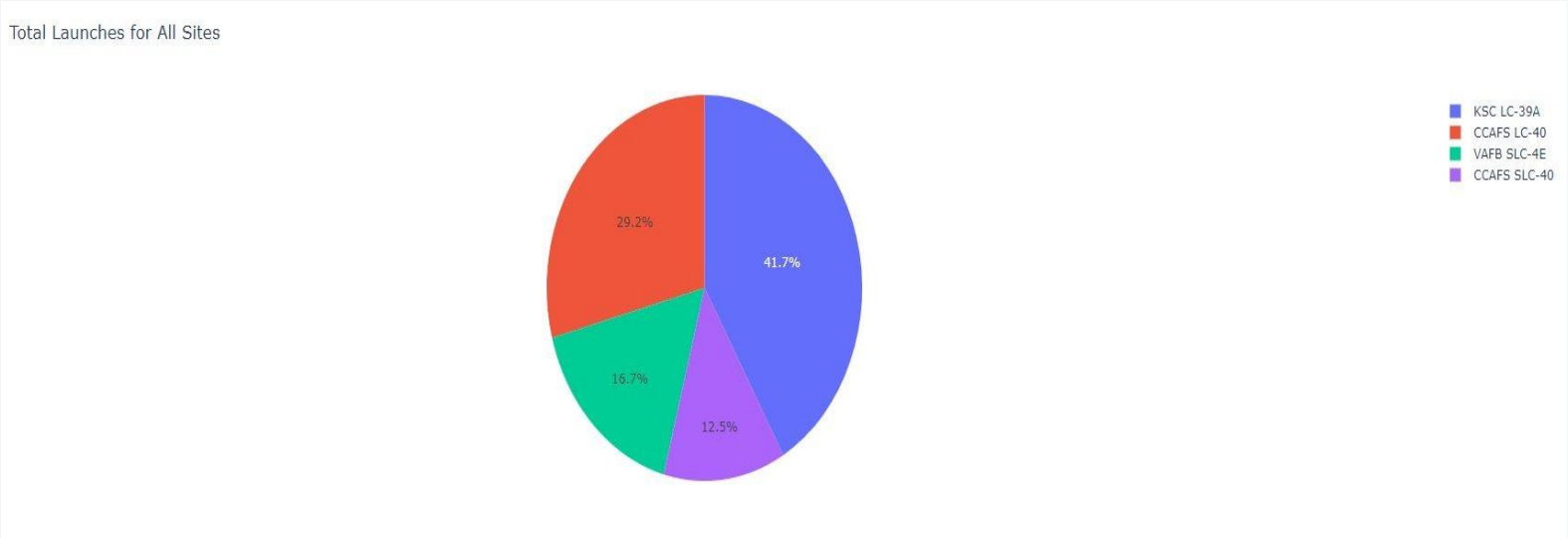
The launch sites in the east coast are close to each other and nearer to the equatorial plane than the west coast one, it helps to get more speed in rocket launching.

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot


Section
4
Build a Dashboard
with Plotly Dash

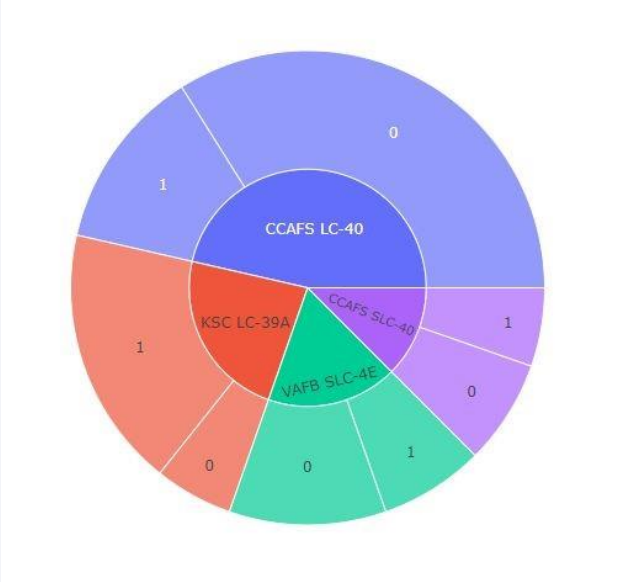
Total launches from all the sites



Pie chart showing percentage of launches from each launch sites.

KSC LC 39A has the highest launches.

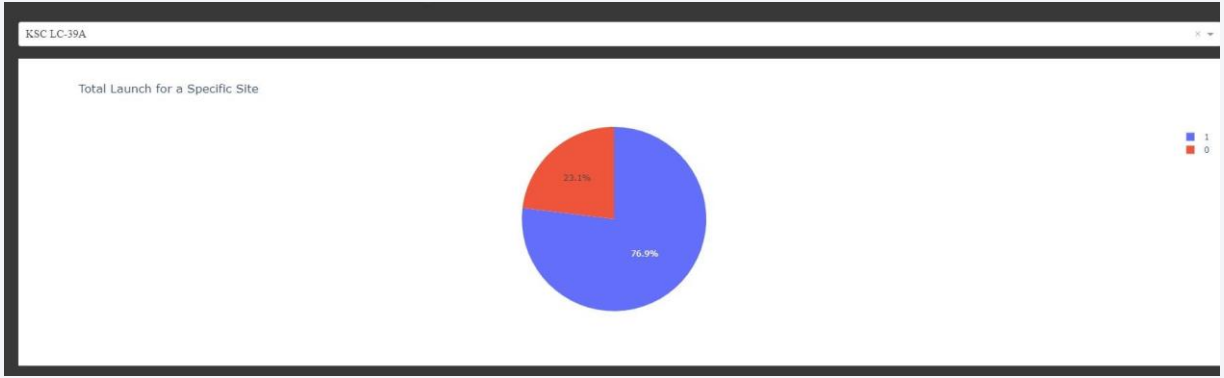
Sunburst chart showing the share of successful landing in each launch location

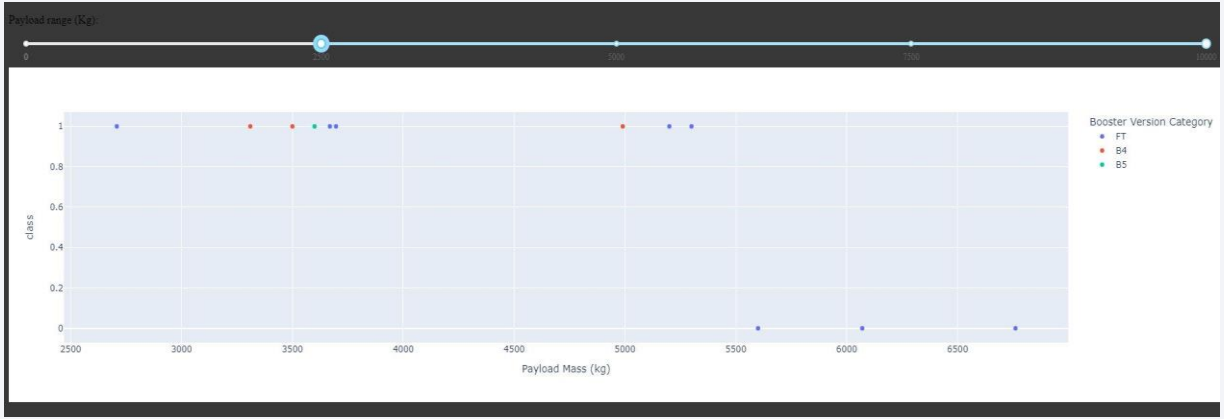


KSC LC 39A, CCAFS LC 40 have the largest share of launches.

KSC LC 39A has better success rate in landing.

Launches from a specific site with selected payload range

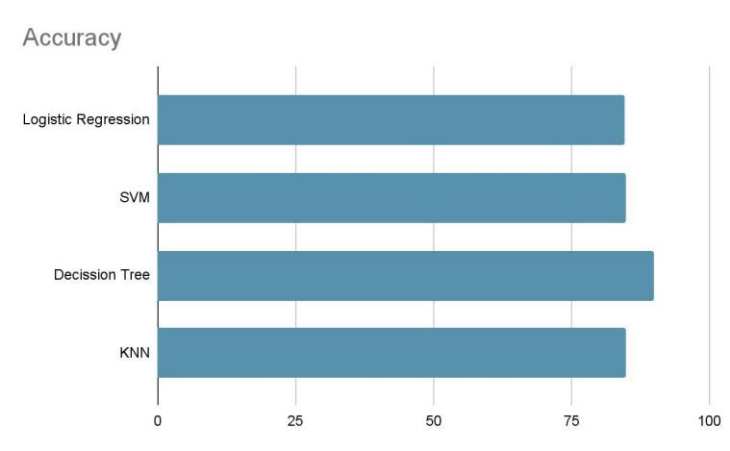




Section 5

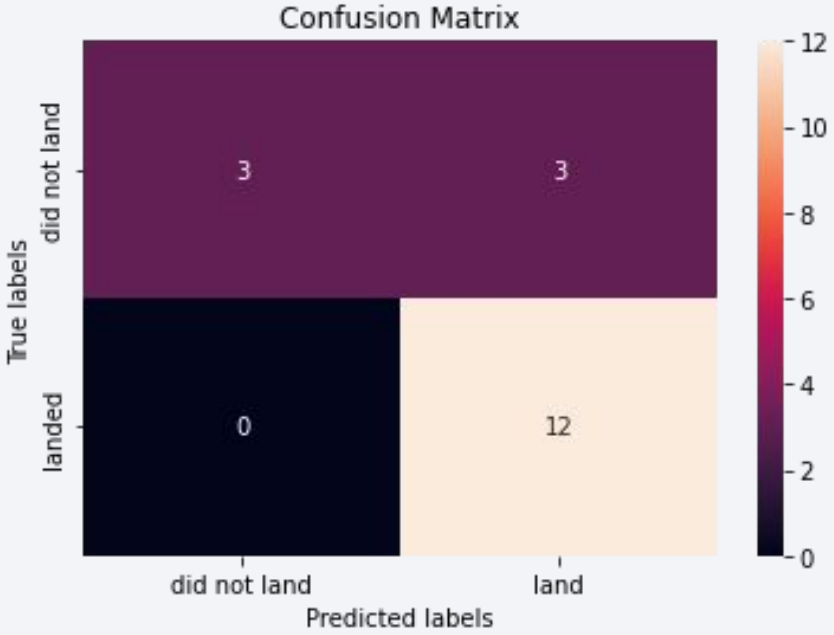
Predictive Analysis (Classification)

Classification Accuracy



Measuring the accuracies in different classification model, decision tree model gives highest accuracy.

Confusion Matrix



Confusion matrix for the decision tree algorithm.

True positive is 12 cases out of 17 test cases.

Conclusions

- The classifications methods used are logistic regressions, support vector machines, decision tree algorithm and k nearest neighbors.
- The f1 scores we got by these methods are 0.833, 0.833, 0.777, 0.777 respectively. The accuracies are 84.64 %, 84.82 %, 90%, 84.82 % respectively.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!