IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tonatiuh Hernández del Toro
October, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - In this project, we analyze a SpaceX dataset containing information about successful/unsuccessful landings to predict whether the next launch will land successfully. We followed the traditional methodology schema (Data collection, Data cleaning, EDA, Visualization, Model building)
- Summary of all results
  - The outcome of the analysis tells us that the success rate has improved with time, the lower the orbit the higher the success rate, and the models were able to predict the class with 83.3% of accuracy

Introduction

- Project background and context
  - SpaceX is a brand of commercial space launches. Lately they have been trying to implement a launch system where the first stage of launch is reusable, and with this save money, although this has not been successfully completed yet since many of the launches end up failing.
  - In this project we present SpaceY, a company which attempts to use SpaceX open information bout launches to predict if a launch is going to be successful or not
- Problems you want to find answers
  - Do an EDA over the dataset of find insights
  - Develop models to predict if a launch is going success based on its features

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is obtained from SpaceX open API
- Perform data wrangling
  - Data came in a dictionary which was extracted as information into a dataframe
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We cleaned, standardized, trained and tuned 4 models (LogReg, SVM, Tree, kNN)

Data Collection

- Datasets were collected via SpaceX open API https://api.spacexdata.com/v4/launches/past
- Data was collected using web scrapping with the response function
- Data was extracted from the json file into a dataframe
- Data was cleaned and filtered

Data Collection – SpaceX API

- Datasets were collected via SpaceX open API https://api.spacexdata.com/v4/launches/past and then, data was collected using web scrapping with the response function

- The github containing the notebook can be found here https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/01%20-%20spacex-data-collection-api.ipynb

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/AI
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

```
# Get the head of the dataframe
data.head()
```

Data Wrangling

- Data was extracted from the json file in order to transform it into a pandas DataFrame
- The data was filtered to include only falcon9 launches
- Then data was cleaned removing of replacing missing values
- Finally, data as exported into a csv file

EDA with SQL

- The queries performed were to extract the following information in a nushell:
  - Unique launches
  - Average Payloads
  - Total success and failed missions
  - Specific information within dates
- The github with the notebook of this exercise can be found in https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/04%20-eda-sql.ipynb

EDA with Data Visualization

- Several charts were plotted:
  - Scatter plots to see relationship among variables
  - Bar plots to see relationship about success rate and orbit type of the mission
  - Line plots to see the success rate over time
- The github with the notebook of this exercise can be found in https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/05%20-%20eda-dataviz.jupyterlite.ipynb

Build an Interactive Map with Folium

- Several folium maps were created:
  - Mark all launch sites on a map to see where they are located
  - Create and add folium circle and folium marker for each launch site on the site map to get better visualization
  - Mark the success/failed launches for each site on the map
  - Draw a PolyLine between a launch site to the selected coastline point
- The github with the notebook of this exercise can be found in https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The dashboards created were:
  - Number of launches per site in pie graph
  - Success launches for specific sites
  - Scatter plot for payload and success
- The github with the code of this exercise can be found in https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/07%20-%20plotly.py

Predictive Analysis (Classification)

- In the predictive analysis part, we:
  - Selected relevant features
  - Normalized dataset
  - Created class
  - Encoded features in one hot fashion
  - Trained and tested 4 models
- The github with the notebook of this exercise can be found in https://github.com/tonahdztoro/IBM_Certification_Capstone/blob/main/08%20-%20Prediction.ipynb

Results

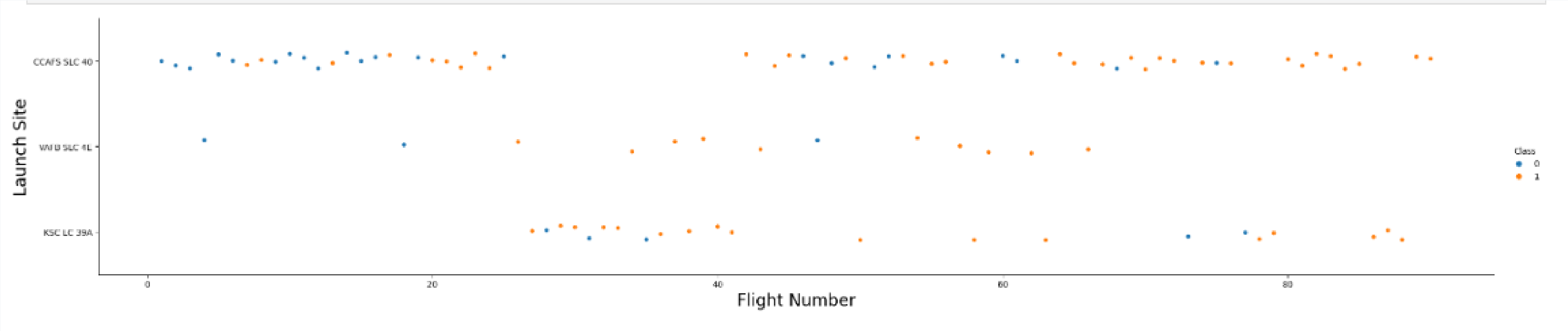
- Exploratory data analysis results
    - There is correlation between some variables (e.g time and success rate)
    - The lower the orbit, the higher the success rate
- Interactive analytics demo in screenshots
    - Launch sites are near the equuator
- Predictive analysis results
    - The four classifiers achieved an accuracy of 0.83

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

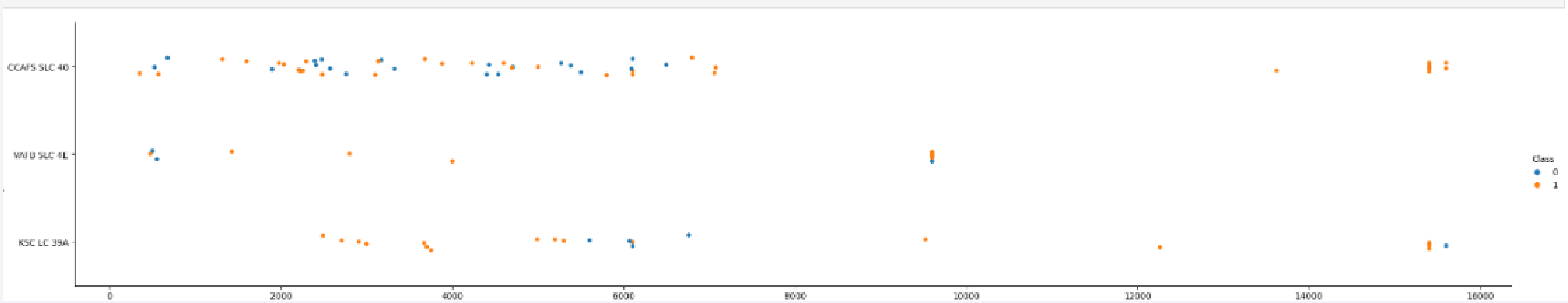
- Scatter plot of Flight Number vs. Launch Site



- We can see that as the flight number increases, there are more launches in the first site, and the success rate increases

Payload vs. Launch Site

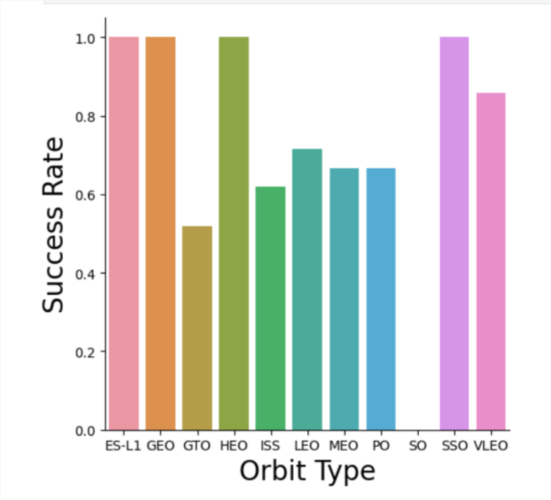
- Scatter plot of Payload vs. Launch Site



- We can see that for low payload mass, the first launch site has very low success rate, and when the payload mass increases, the success rate does it too

Success Rate vs. Orbit Type

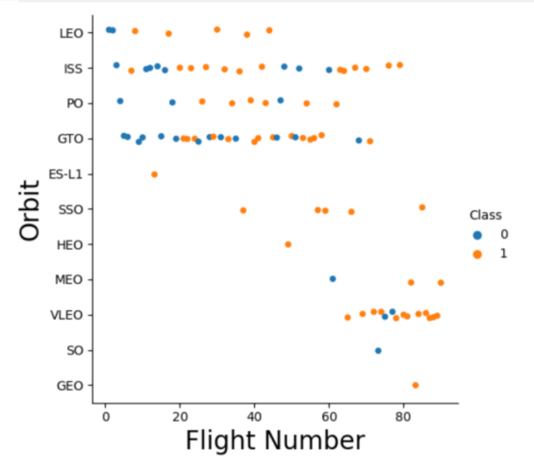
- Bar chart for the success rate of each orbit type



- We can see that the lower the orbit, the higher the success rate

Flight Number vs. Orbit Type

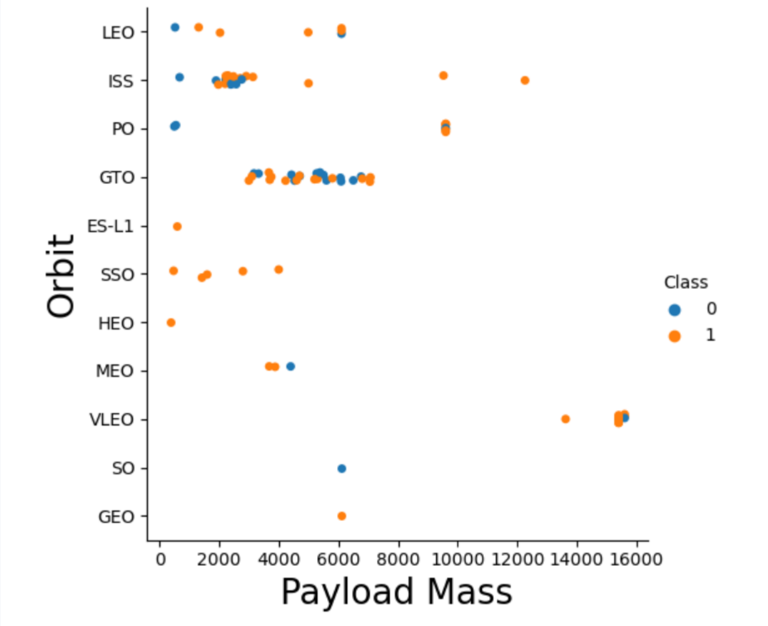
- Scatter point of Flight number vs. Orbit type



- We can see that as the flight number increases, the orbit goes into the higher ones

Payload vs. Orbit Type

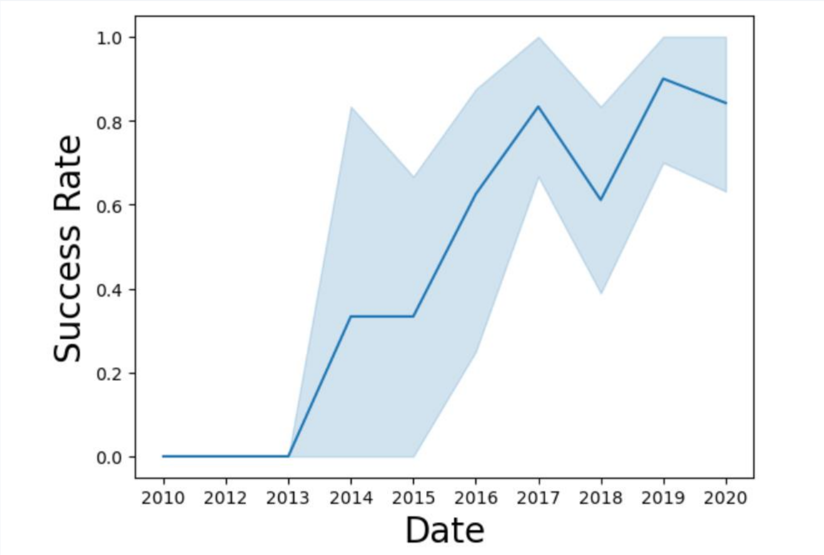
- Scatter point of payload vs. orbit type



- We can see that the lower payload masses are sent into the lower orbit types

Launch Success Yearly Trend

- Line chart of yearly average success rate



- We can see that as the time goes by, the success rate increases. This is quite obvious since the technology advances with time

All Launch Site Names

- The names of the unique launch sites, as well as the query needed is shown below.

```
8]:  %sql select distinct(Launch_Site) from SPACEXTABLE

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

8]:

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` as well as the query needed is shown below

```
%sql select * from SPACEXTABLE 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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload carried by boosters from NASA as well as the query needed is shown below

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer = 'NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 as well as the query needed is shown below

```
%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

- The dates of the first successful landing outcome on ground pad as well as the query are shown below:

```
%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

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 as well as the query are shown below:

```
%sql select Payload from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes as well as the query are shown below:

```
%sql select Mission_Outcome, count(*) as Total from SPACEXTABLE group by Mission_Outcome
```

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

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass as well as the query are shown below:

```
%%sql
select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

```
 * 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

2015 Launch Records

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 as well as the query are sown below:

```
%%sql
select substr(Date,6,2) as month, Date, BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome
from SPACEXTBL
where substr(Date,1,4)='2015' and Landing_Outcome = 'Failure (drone ship)'
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- The ranking of 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 as well as the query are shown below:

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.

```
%sql select Landing_Outcome, count(*) as Total from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by La
```

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

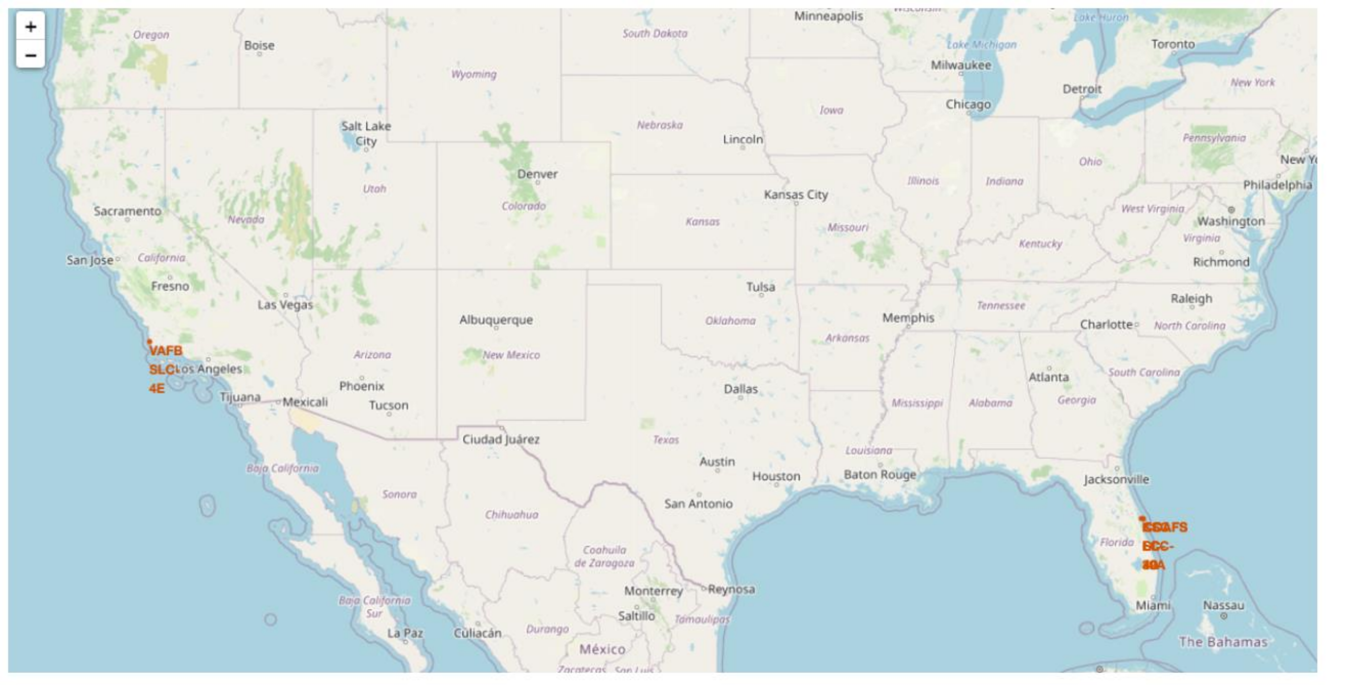
Landing_Outcome	Total
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

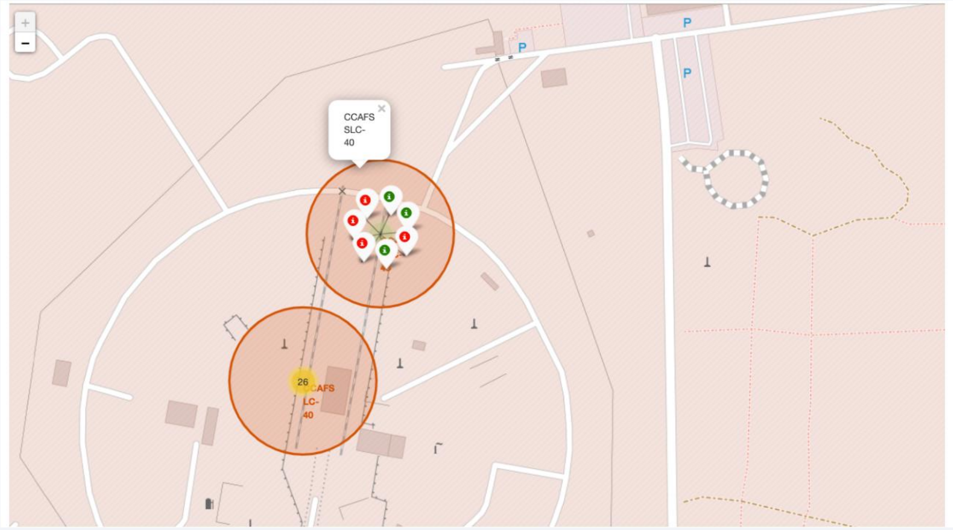
Launch sites

- Here are shown the launch sites, we can see that all of them are located near the east or west coast of the United States



Launch sites proximities

- Here are shown the launch sites as well as the railways, highways, and coasts near of them. We can see that there is not too much near them. This is due to safety reasons

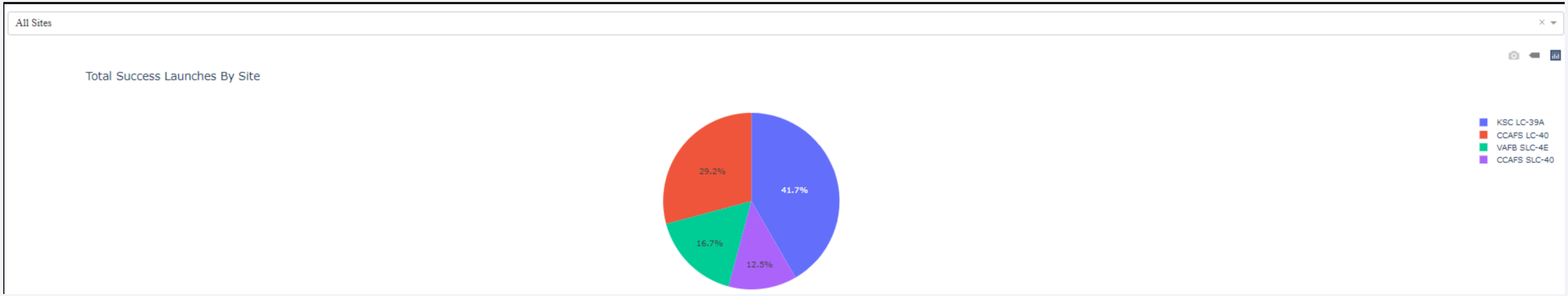


Section 4

Build a Dashboard with Plotly Dash

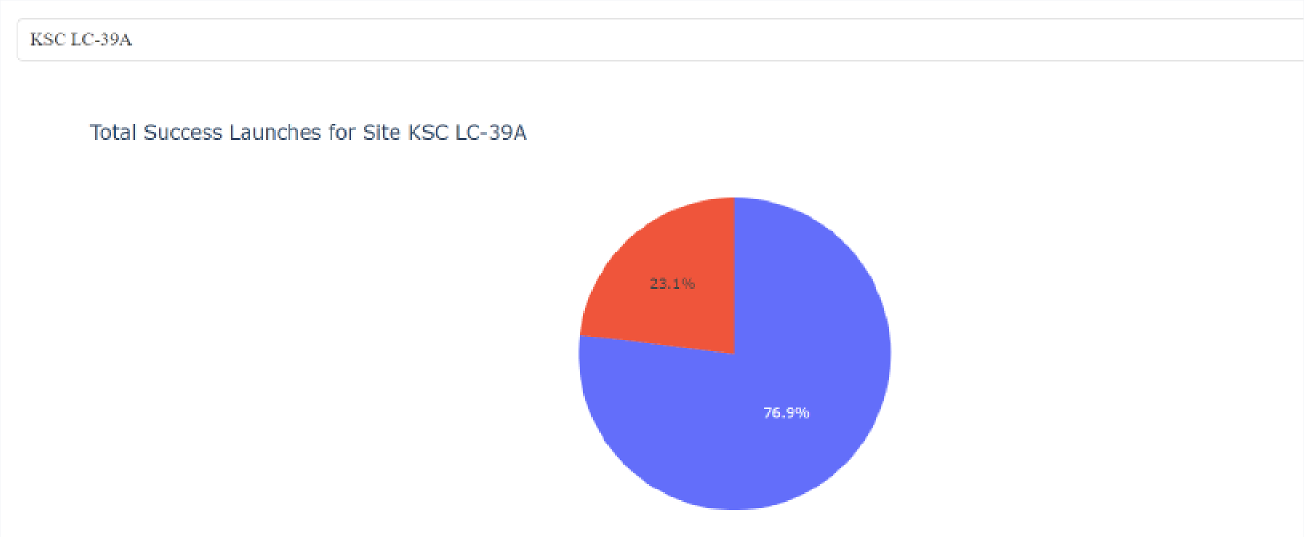
Success rate by launch site

- Here is presented a dashboard which shows a pie chart with the success rate by launch sites



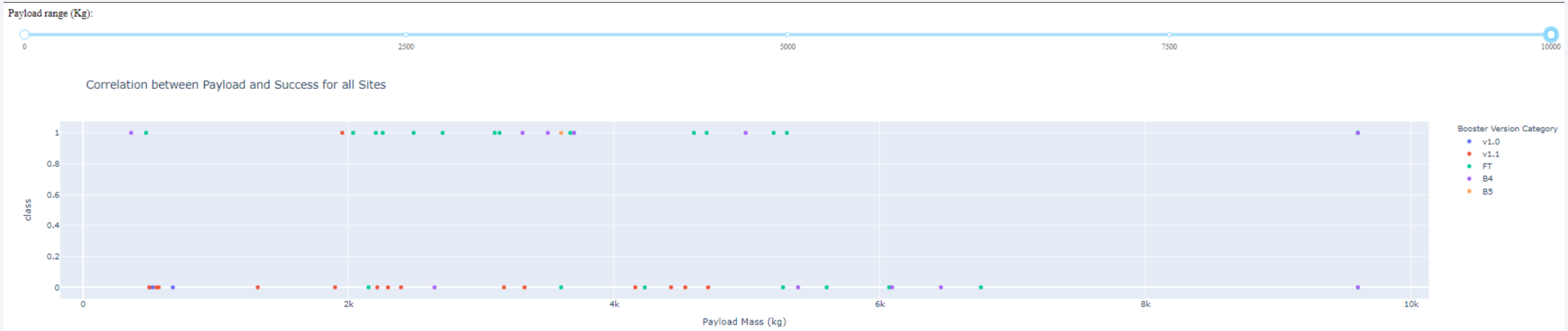
Most successful launch site

- Here is shown a pie chart with the success values of the launch site that achieves the highest success launches



Payload mass vs Launch Outcome

- Here is shown a chart of the dashboard showing a payload vs. launch outcome scatter plot for all sites, with different payload selected in the range slider:

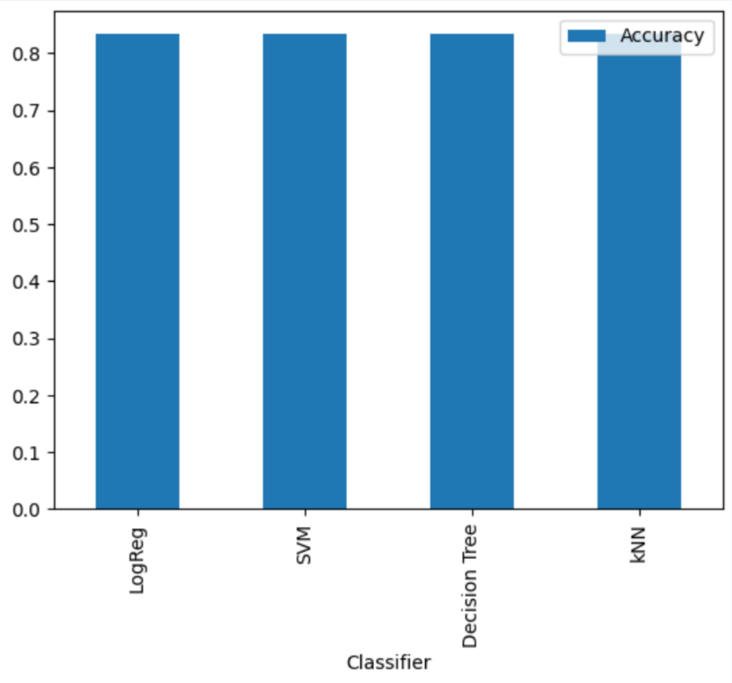


Section 5

Predictive Analysis (Classification)

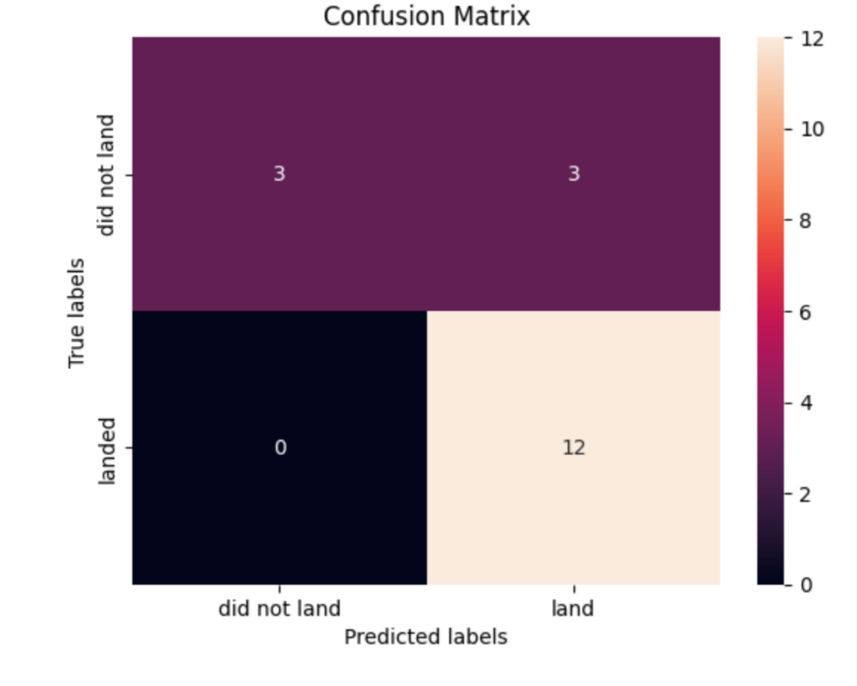
Classification Accuracy

- Here are shown the accuracies of the four models used in this project in a bar chart plot:



Confusion Matrix

- Here is shown the confusion matrix of the best performing model:



Conclusions

- The lower the orbit, the higher the success rate
- The most recent the launch, the higher the success rate
- Most of launch sites are near equator and coast
- All four classifiers achieved an accuracy of 0.83

Thank you!