IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Dr. Chun Kong Mak>
<18th July 2024>

Outline

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

Executive Summary

- Summary of methodologies:

In this project, data collection is achieved by SpaceX API. Data is cleaned up. Exploratory data analysis (EDA) using visualization SQL are performed. Interactive visual analytics using Folium and Plotly Dash are carried out. At the end, predictive analysis using classification models is done.

- Summary of all results

➢ From EDA through visualization:

➢ At CCAFS LC-40, success rate increases when flight number is more than 40. At VAFB SLC 4E, success rate increases when flight number is more than 20.

➢ At VAFB-SLC launch site, there are no rockets launched for heavy payload mass (greater than 10000kg).

➢ Which orbits have high success rate? VLEO: > 0.8. SSO, HEO, GEO and ES-L1 gives 100% success rate.

➢ In the LEO orbit, the success appears related to the number of flights. Success increases when flight number increases. In the GTO orbit, there seems to be no relationship between flight number and success.

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

➢ Success rate kept increasing since 2013 till 2020.

Executive Summary

- Summary of all results

From EDA through SQL:

- There are four unique sites namely CCAFS LC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40.
- Total payload mass carried by boosters launches by NASA (CRS) is 45596kg.
- The average Payload mass from Booster Version F9 v1.1 is 2928.4kg.
- The first successful landing was on 8th April 2016.
- There are three unique Booster versions namely F9 FT B1032.1, F9 B4 B1040.1 and F9 B4 B1043.1 when the landing was successful and the payload mass was between 4000 and 6000 kg.
- The number of successful mission outcomes was 100 while the number of failed outcome was 1.
- Booster Versions with maximum payload mass were F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3, F9 B5 B1056.4, F9 B5 B1048.5, F9 B5 B1051.4, F9 B5 B1049.5, F9 B5 B1060.2, F9 B5 B1058.3, F9 B5 B1051.6, F9 B5 B1060.3, F9 B5 B1049.7.

Executive Summary

- Summary of all results

Folium map:

➢ There are in total 56 launch sites. All of them are in United States. 10 of them are on the west coast while 46 of them are on the east coast.

➢ All launch sites are close to the coast line.

Plotly Dashboard:

➢ Launch site KSC LC-39A has the highest successful rate i.e. 41.7% among all the four sites.

➢ In the site KSC LC-39A, there are 76.9% success outcome and 23.1% failed outcome.

➢ Site CCAFS LC-40: 26.9% success, 73.1% fail. Success occurs only with FT Booster Version with PayLoad Mass over 2000kg.

➢ Site CCAFS SLC-40: 42.9% success, 57.1% fail. Success rate seems higher when PayLoad Mass is under 4000kg and it doesn’t seem to be dependent on Booster Version.

➢ Site KSC LC-39A: 23.1% success, 76.9% fail. Success occurs only with PayLoad Mass under 5500kg and independent of Booster Version

➢ Site VAFB SLC-4E: 40% success, 60% fail. Success and failure rate seem to be unrelated to PayLoad Mass nor the Booster Version.

Introduction

- Project background and context

In this capstone project, 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.

- Problems you want to find answers

What are the factors that increase the successful landing rate?

Section 1

Methodology

Methodology

Executive Summary

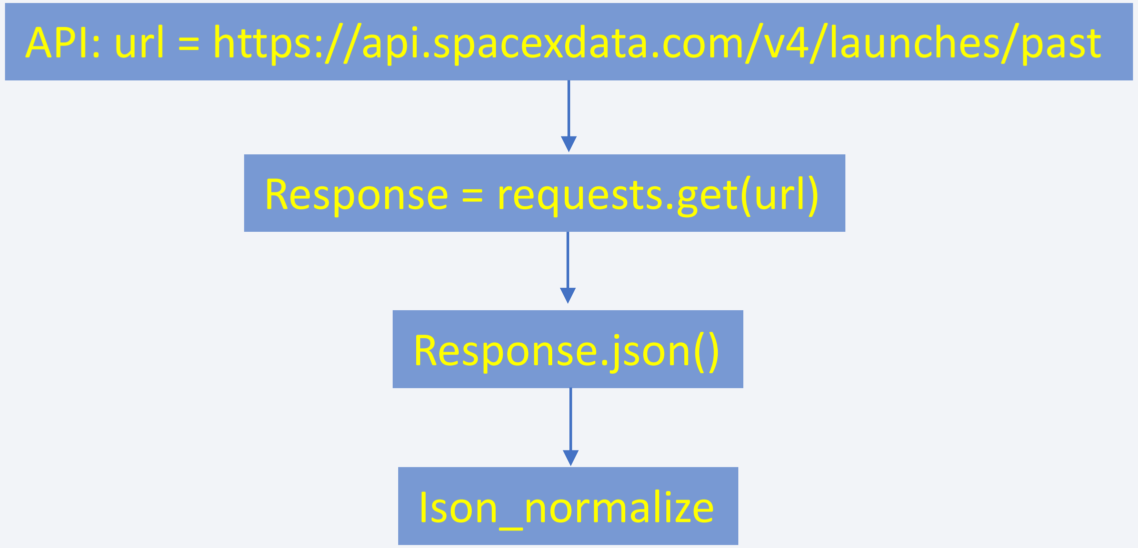
- Data collection methodology:
  - SpaceX launch data was collected by making a get request to SpaceX API.
- Perform data wrangling
  - Data was filtered to have only records from BoosterVersion = Falcon 9. Missing data of the PayLoadMass was replaced by the mean of the PayLoadMass. Number of launches is calculated on each launching site. Number and occurrence of each orbit is calculated. The number of occurrence of mission outcome per orbit type is calculated. When the landing outcome is successful, “1” is assigned to the class. If not successful, “0” is assigned to the class. Hence, success rate 67% is calculated among 90 landings while 30 failed landings occurred.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Class of “1” or “0” is created as a column to represent the successful outcome as “1” and unsuccessful outcome as “0”. The data was standardized and split into training data and test data. The best Hyperparameter was determined for four models i.e. SVM, Classification Trees, Logistic Regression and K-nearest neighbors. All models achieved the same test accuracy of 83.3%. Classification Tree model provides the highest validation accuracy 90.4%, which suggests it performed better on the validation data during hyperparameter tuning. It performs the best based on the hyperparameter tuning results and validation accuracy.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose: https://github.com/ckmak9999/ibm_datascience_capstone/blob/cd10996168de6951d98f7e22d00d9679bf702a5b/1_1_jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Here is the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose: https://github.com/ckmak9999/ibm_datascience_capstone/blob/cd10996168de6951d98f7e22d00d9679bf702a5b/1_2_jupyter-labs-webscraping.ipynb

Request Wikipedia URL
https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

↓

Web scraping with BeautifulSoup

↓

Convert to Panda dataframe

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts

Filter data only Falcon 9 → Replace missing data of PayLoadMass by Mean of PayLoadMass → Calculate number of launches on each site → When outcome is successful, class = "1"
When outcome is not successful, class = "0"

- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose:

https://github.com/ckmak9999/ibm_datascience_capstone/blob/cd10996168de6951d98f7e22d00d9679bf702a5b/1_3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

1. Scatter plot: Flight Number versus Pay Load Mass (kg), overlay outcome of launch → Flight number increases, land success increases. Payload mass increases, success decreases.
2. Scatter plot: Launch Site versus Flight Number, overlay outcome of launch → Different launch sites have different success rates.
3. Scatter plot: Launch Site versus Pay Load Mass (kg), overlay outcome of launch → VAFB-SLC launchsite, no rockets launched for heavy payload mass (> 10000kg).
4. Bar chart: Orbit Type versus Success Rate → which orbits have high success rate?
5. Scatter plot: Orbit versus Flight Number, overlay outcome of launch→ Relationship between Flight Number and Orbit Type
6. Scatter plot: Payload Mass versus Orbit Type, overlay outcome of launch→ Relationship between PayLoad Mass and Orbit Type
7. Line plot: Success rate versus Date → success rate increases since 2013 until 2020.

EDA with Data Visualization

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

https://github.com/ckmak9999/ibm_datascience_capstone/blob/cd10996168de6951d98f7e22d00d9679bf702a5b/2_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

- Task 1: %sql select DISTINCT Launch_Site from SPACEXTBL
- Task 2: %sql select * from SPACEXTBL where Launch_Site like "KSC%" LIMIT 5
- Task 3: %sql select SUM("PAYLOAD_MASS__KG_") AS total_payload_mass from SPACEXTBL where Customer = "NASA (CRS)“
- Task 4: %sql select AVG(PAYLOAD_MASS__KG_) as Average_Payload_Mass from SPACEXTBL where Booster_Version = "F9 v1.1“
- Task 5: %sql select min(Date) as date_successful_landing from SPACEXTBL where "Landing_Outcome" = "Success (drone ship)“
- Task 6: %sql select DISTINCT Booster_Version from SPACEXTBL where "Landing_Outcome" = "Success (ground pad)" and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000
- Task 7: %sql select count(*) as num_successful_outcomes from SPACEXTBL where "Mission_Outcome" like "Success%“
- %sql select count(*) as num_failure_outcomes from SPACEXTBL where "Mission_Outcome" like "Failure%“
- Task 8: %sql select distinct Booster_Version from SPACEXTBL where "PAYLOAD_MASS__KG_" = (Select MAX("PAYLOAD_MASS__KG_") from SPACEXTBL)
- Task 9: %sql SELECT substr("Date", 6, 2) AS Month_Number, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr("Date", 0, 5) = '2017'AND Landing_Outcome LIKE 'Success%‘
- Task 10: %sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' group by Landing_Outcome ORDER BY Outcome_Count DESC

EDA with SQL

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

https://github.com/ckmak9999/ibm_datascience_capstone/blob/766566d290a157276c54dee35c6dfb8ebf0b1f31/2_1_jupyter-labs-eda-sql-edx_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects

Folium.Circle: create for NASA Johnson Space Center

Folium.map.Marker: to add coordinate and information of the spot. Add color of the marker to show successful outcome or not.

Folium.Polyline: to show the calculated distance between the launch site and other objects such as closest city, railway, highway, coastline.

- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

https://github.com/ckmak9999/ibm_datascience_capstone/blob/766566d290a157276c54dee35c6dfb8ebf0b1f31/3_1_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard

Pie chart of total success launches for all sites, indicating the relative ratio of success rates among four launch sites. Hovering on the chart displays the launch site name and number of data points

Pie chart displaying the total success launches for each of the four individual sites. It shows % of success and fail. Hovering on the chart displays the value of class and number of data points.

Range slider: x-axis is PayLoad Mass (kg). y-axis is class (1 or 0). The color code legend indicates Booster Version.

- Explain why you added those plots and interactions

The plots and interactions provide us insights of: which launch site has more success rate? In each launch site, are Booster Version and PayLoad Mass (kg) related to the success rate?

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

https://github.com/ckmak9999/ibm_datascience_capstone/blob/766566d290a157276c54dee35c6dfb8ebf0b1f31/3_2_spacex_dash_app.py

Predictive Analysis (Classification)

Load dataset generated from EDA

→ Split data into training data and testing data sets. Test data set is 20%.

→ Apply four models separately

i. Logistic Regression
ii. Support vector machine (SVM)
iii. Decision tree classifier
iv. K nearest neighbors

→ Determine and compare

1. Tuned hyperparameters
2. Accuracy on validation data
3. Accuracy on the test data
4. Confusion matrix

→ Which model is the best?

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

https://github.com/ckmak9999/ibm_datascience_capstone/blob/766566d290a157276c54dee35c6dfb8ebf0b1f31/4_1_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

Section 2

Insights drawn from EDA

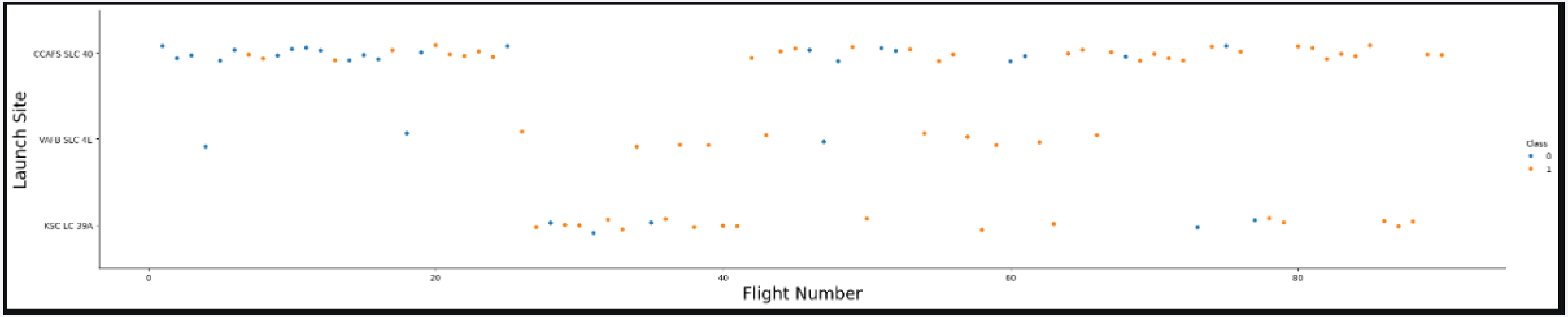
Flight Number vs. Launch Site

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

Different launch sites have different success rates dependent on the Flight Number.

At CCAFS LC-40, success rate increases when flight number is more than 40.

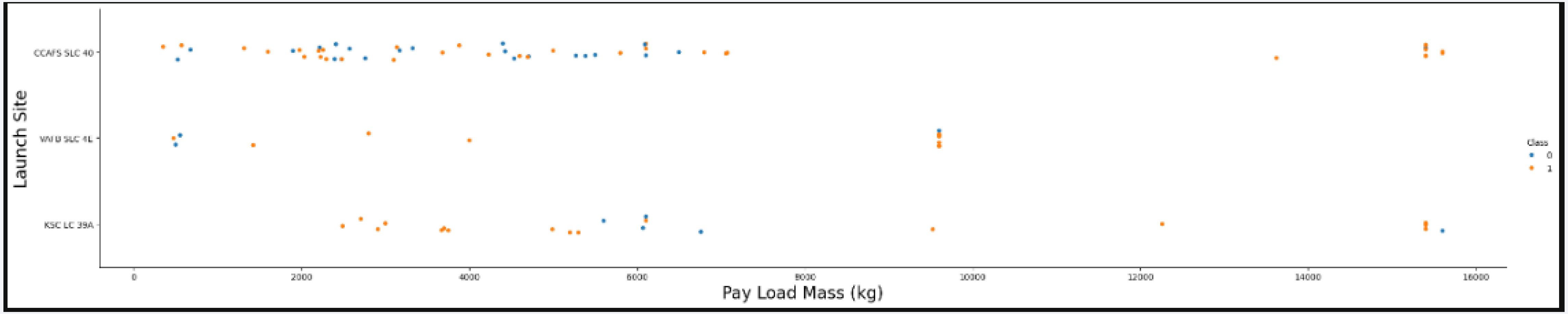
At VAFB SLC 4E, success rate increases when flight number is more than 20.



Payload vs. Launch Site

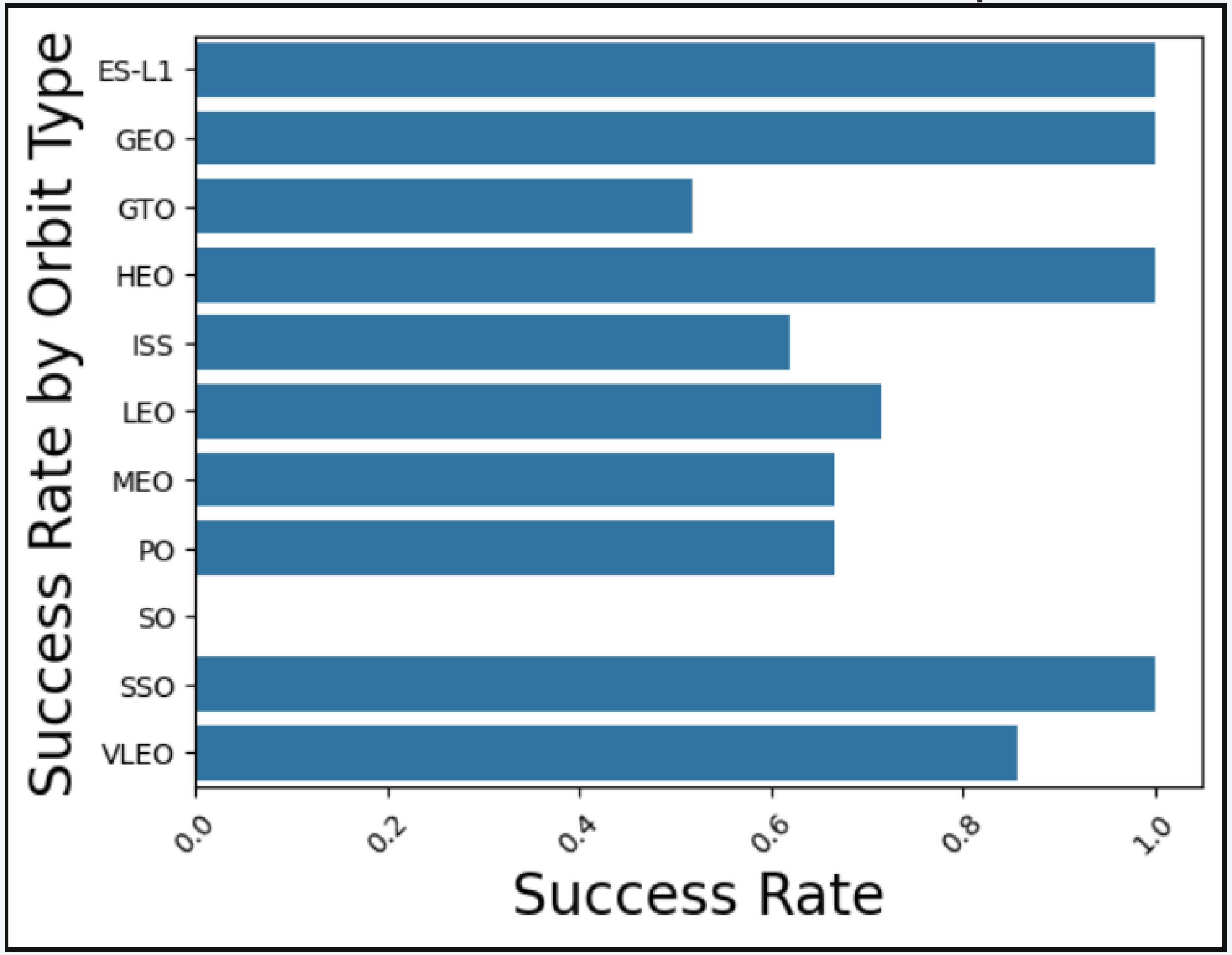
- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations

At VAFB-SLC launch site, there are no rockets launched for heavy payload mass (greater than 10000kg).



Success Rate vs. Orbit Type

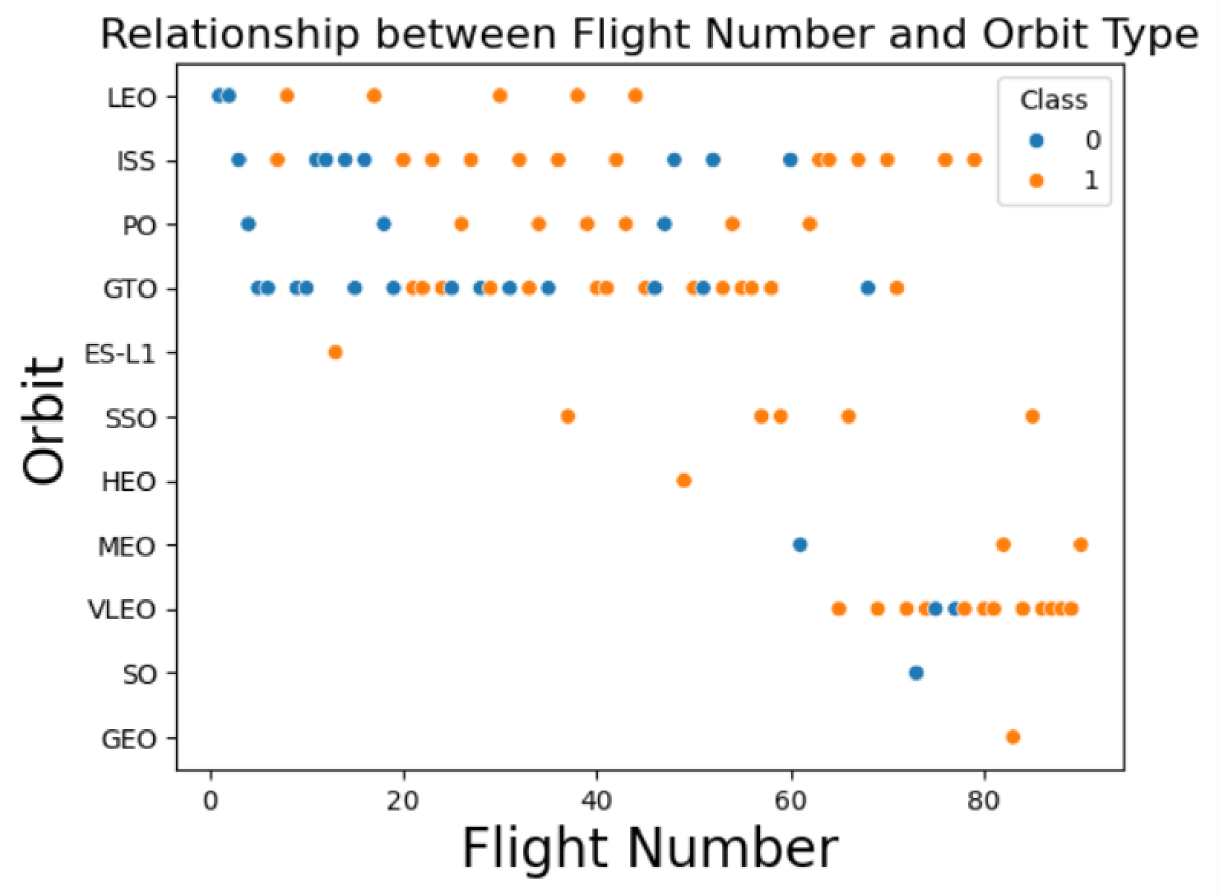
- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations



Which orbits have high success rate?
VLEO: > 0.8
SSO, HEO, GEO and ES-L1 gives 100% success rate.

Flight Number vs. Orbit Type

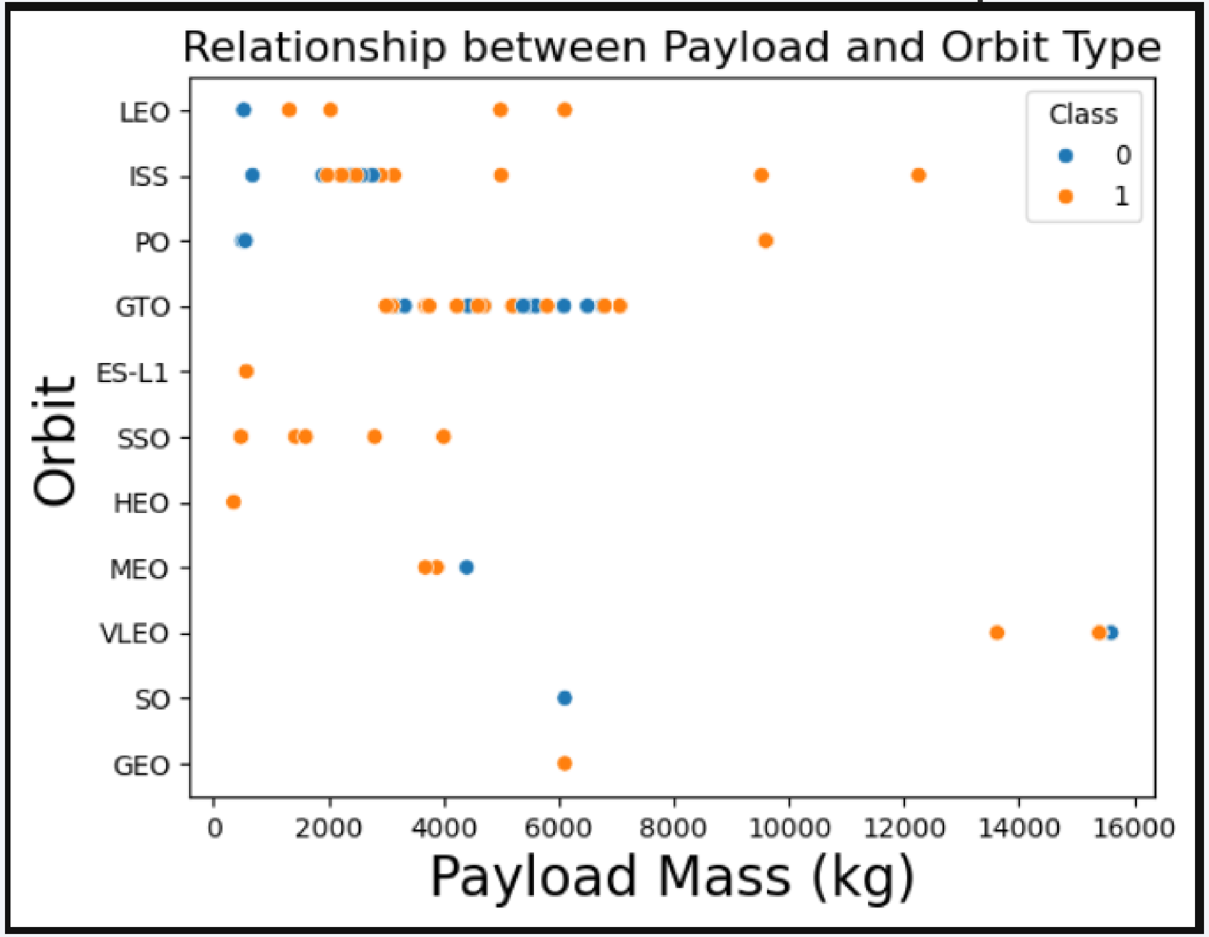
- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



In the LEO orbit, the success appears related to the number of flights. Success increases when flight number increases. In the GTO orbit, there seems to be no relationship between flight number and success.

Payload vs. Orbit Type

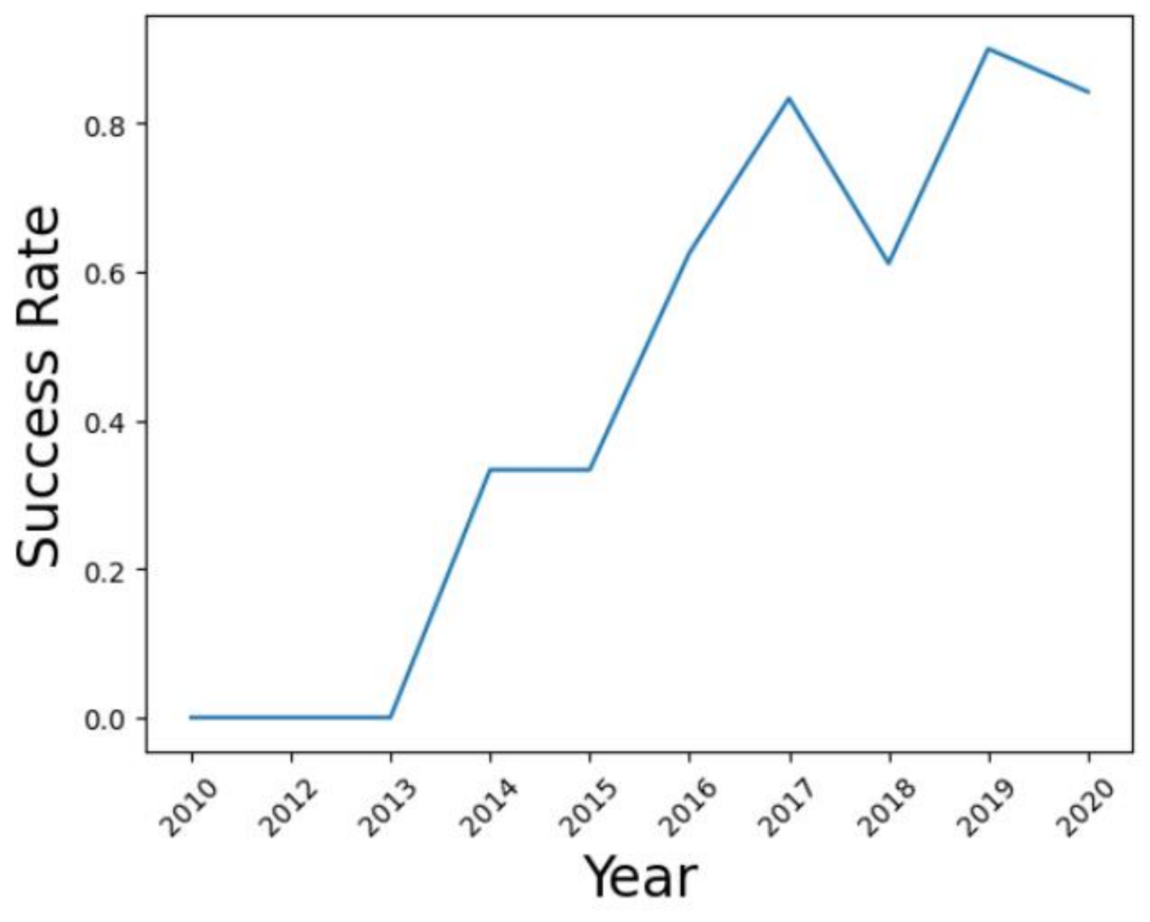
- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



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

Launch Success Yearly Trend

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



Success rate kept increasing since 2013 till 2020.

All Launch Site Names

Use SQL command to find the distinct launch_site record from the SPACEXTBL database

```
[9]: %sql select DISTINCT Launch_Site from SPACEXTBL

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

[9]:

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

Launch Site Names Begin with 'KSC'

Use SQL to search all records from SPACEXTBL when the launch_site has words starting with ‘KSC’. Limit to display the top five entries.

Task 2

Display 5 records where launch sites begin with the string 'KSC'

```
%sql select * from SPACEXTBL where Launch_Site like "KSC%" LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

Sum all the “payload_mass_kg” and save it into a value named total_payload_mass from the SPACEXTBL database. Where the customer has string “NASA (CRS)”

Task 3 ¶

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[21]: %sql select SUM("PAYLOAD_MASS__KG_") AS total_payload_mass from SPACEXTBL where Customer = "NASA (CRS)"
```

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

[21]:

total_payload_mass
45596

Average Payload Mass by F9 v1.1

Use SQL to calculate the average value of “PAYLOAD_MASS__KG_” and store it into a value named Average_Payload_Mass from the database SPACEXTBL where the “Bosster_Version” equals to the string “F9 v1.1”.

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[12]: %sql select AVG(PAYLOAD_MASS__KG_) as Average_Payload_Mass from SPACEXTBL where Booster_Version = "F9 v1.1"

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

[12]:

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

Find the minimum value of Date and store it into a value named date_successful_landing from the database SPACEXTBL where the value “Landing_Outcome” equals to “Success (drop Ship)”

Task 5

List the date where the succesful landing outcome in drone ship was acheived.

Hint:Use min function

[13]:
```
%sql select min(Date) as date_successful_landing from SPACEXTBL where "Landing_Outcome" = "Success (drone ship)"
```

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

[13]:

date_successful_landing
2016-04-08

Successful Drone Ship Landing with Payload between 4000 and 6000

Use SQL to find the distinct unique values of “Booster_Version” from the database SPACEXTBL where the value “Landing_Outcome” equals to string “Success (ground pad)” and “PAYLOAD_MASS__KG” > 4000 and “PAYLOAD_MASS__KG” < 6000

Task 6

List the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000 ¶

```
[14]: %sql select DISTINCT Booster_Version from SPACEXTBL where "Landing_Outcome" = "Success (ground pad)" and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000
```

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

[14]:

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

Use SQL to count the total records, count(*), and store it into a value named "num_successful_outcomes" from the SPACEXTBL database where the value "Mission_Outcome" starts with string "Success". Similarly to the failure outcomes

Task 7

List the total number of successful and failure mission outcomes

```
[15]: %sql select count(*) as num_successful_outcomes from SPACEXTBL where "Mission_Outcome" like "Success%"

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

[15]:

num_successful_outcomes
100

```
[16]: %sql select count(*) as num_failure_outcomes from SPACEXTBL where "Mission_Outcome" like "Failure%"

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

[16]:

num_failure_outcomes
1

Boosters Carried Maximum Payload

Use SQL to search for distinct values of “Booster_Version” from the dataset SPACEXTBL where the “PAYLOAD_MASS__KG_” equals to the maximum value of “PAYLOAD_MASS__KG_”. Use compound SQL query to achieve.

Task 8

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

```
[17]: %sql select distinct Booster_Version from SPACEXTBL where "PAYLOAD_MASS__KG_" = (Select MAX("PAYLOAD_MASS__KG_") from SPACEXTBL)
```

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

[17]:

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

2015 Launch Records

Use SQL to get “substr(“Date”, 6, 2) and store it into value “Month_Number”, and get records of Landing_Outcome, Booster_Version and Launch_Site from the database SPACEXTBL where substr("Date", 0, 5) = '2017‘ and Landing_Outcome starts with string “Success”

Task 9

List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

Note: SQLLite does not support monthnames. So you need to use substr(Date,6,2) for month, substr(Date,9,2) for date, substr(Date,0,5),='2017' for year.

[18]:
```
%sql SELECT substr("Date", 6, 2) AS Month_Number, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr("Date", 0, 5) = '2017'AND Landing_Outcome LIKE 'Success%'
```

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

[18]:

Month_Number	Landing_Outcome	Booster_Version	Launch_Site
01	Success (drone ship)	F9 FT B1029.1	VAFB SLC-4E
02	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
03	Success (drone ship)	F9 FT B1021.2	KSC LC-39A
05	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
06	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
06	Success (drone ship)	F9 FT B1029.2	KSC LC-39A
06	Success (drone ship)	F9 FT B1036.1	VAFB SLC-4E
08	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
08	Success (drone ship)	F9 FT B1038.1	VAFB SLC-4E
09	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
10	Success (drone ship)	F9 B4 B1041.1	VAFB SLC-4E
10	Success (drone ship)	F9 FT B1031.2	KSC LC-39A
10	Success (drone ship)	F9 B4 B1042.1	KSC LC-39A
12	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

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

Use SQL to search records of Landing_Outcome, and count(*). Store count(*) into a value named Outcome_Count. From the database SPACEXTBL. Where "Date" BETWEEN '2010-06-04' AND '2017-03-20'. Group the records by Landing_Outcome. Order by Outcome_Count descendingly.

Task 10

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

[20]:

```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' group by Landing_Outcome ORDER BY Outcome_Count DESC
```

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

[20]:

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

Section 3

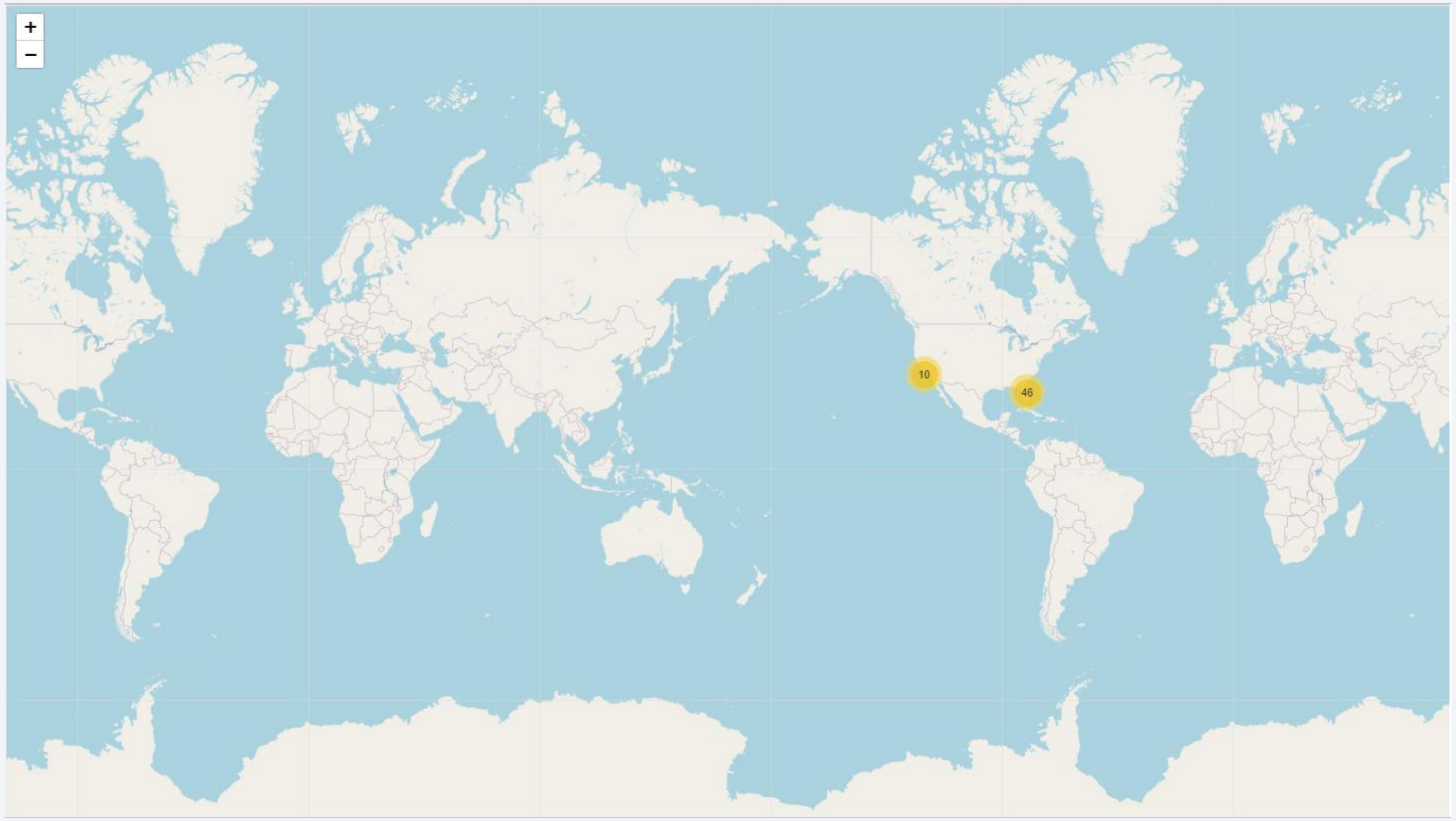
Launch Sites Proximities Analysis

Launch Sites on a Global Map

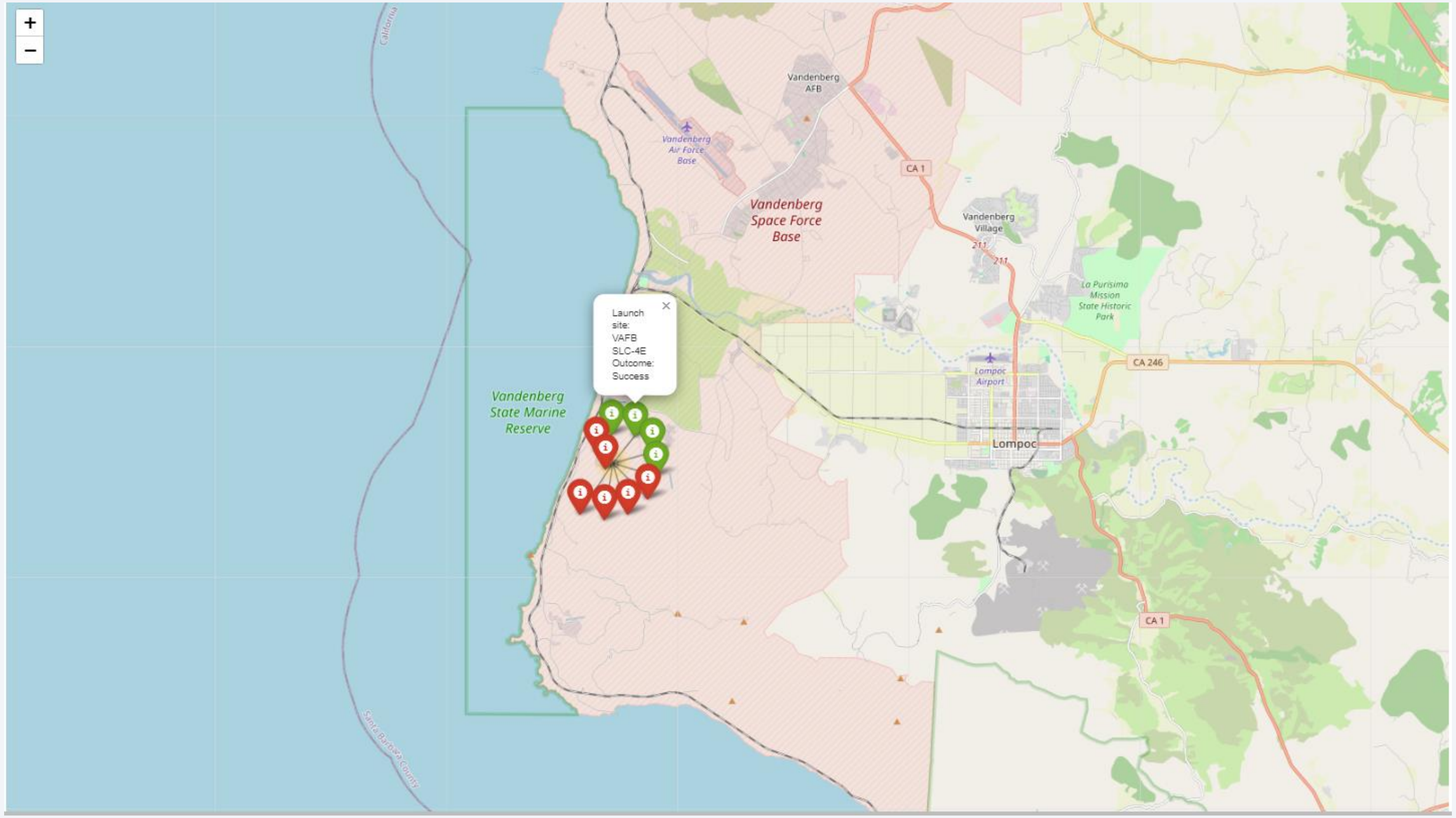
There are in total 56 launch sites. All of them are in United States.

10 of them are on the west coast while 46 of them are on the east coast.

MarkerCluster object is represented by the yellow dots.



Map with color-labeled launch outcomes

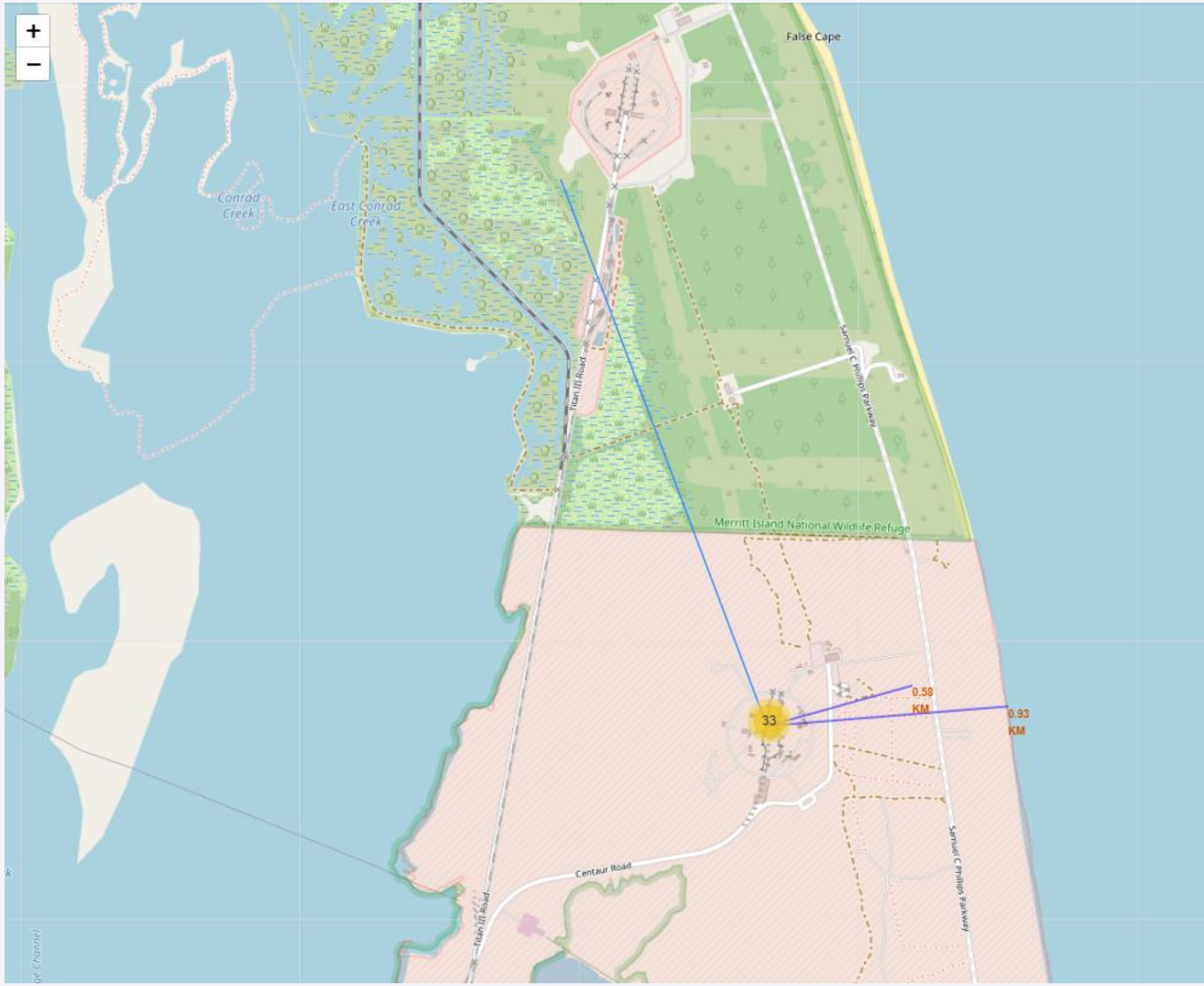


Create a new column called "marker_color".

Assign value "green" when launch_outcome is 1. Assign value "red" when launch_outcome is 0. Create MarkerCluster object. Add the Launch Site name, Lat, Long, marker color to the marker.

In the screenshot, for example, the launch site is VAFB SLC-4E. The dot is in green color when the launch outcome was successful.

Map with calculated distance



It is an example of a folium map with calculated distance between the site CCAFS LC-40 and the coastline.

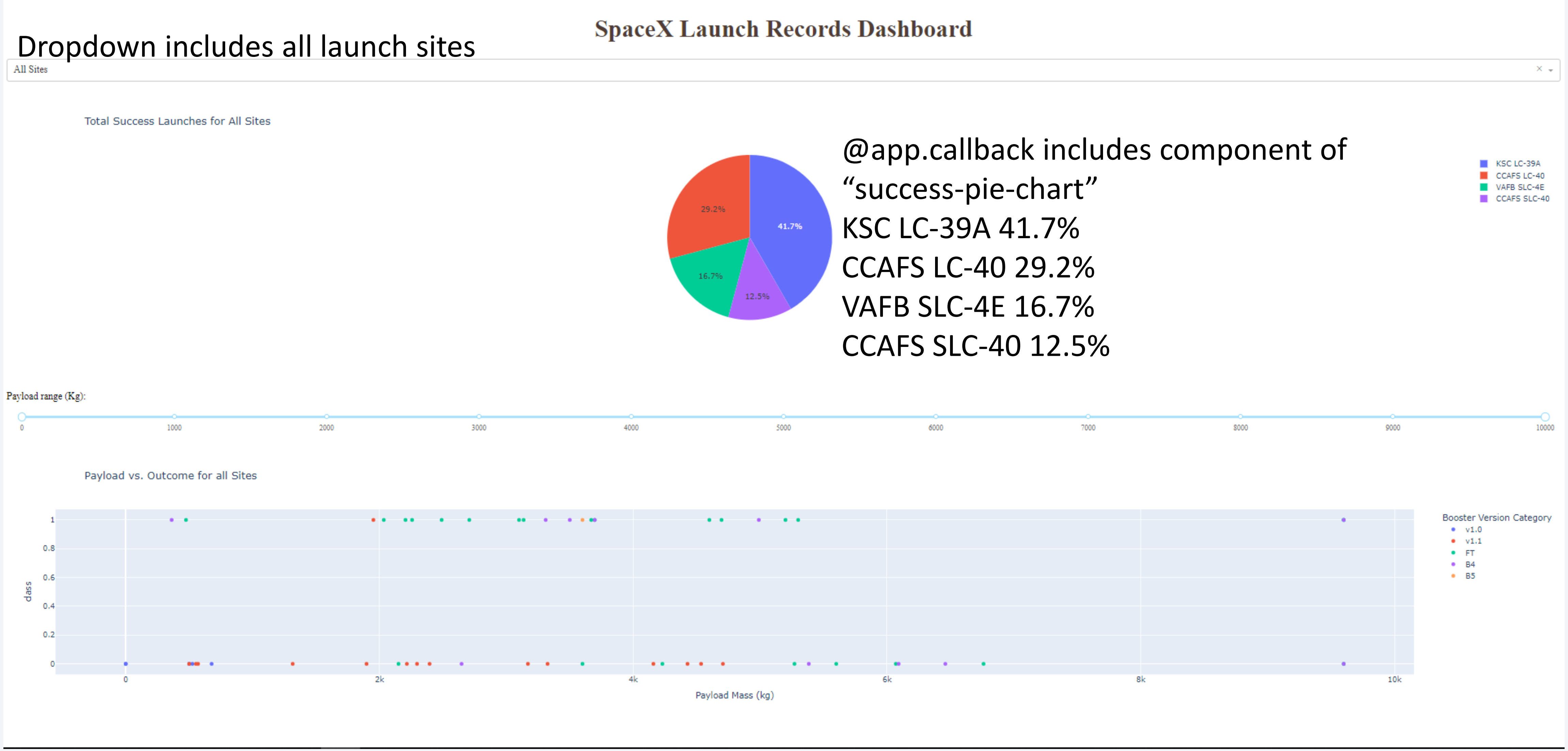
Objective folium.Marker and folium.Polyline are added.

Section 4

Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard – Total Success Launches for All Sites

Dropdown includes all launch sites



@app.callback includes component of “success-pie-chart”

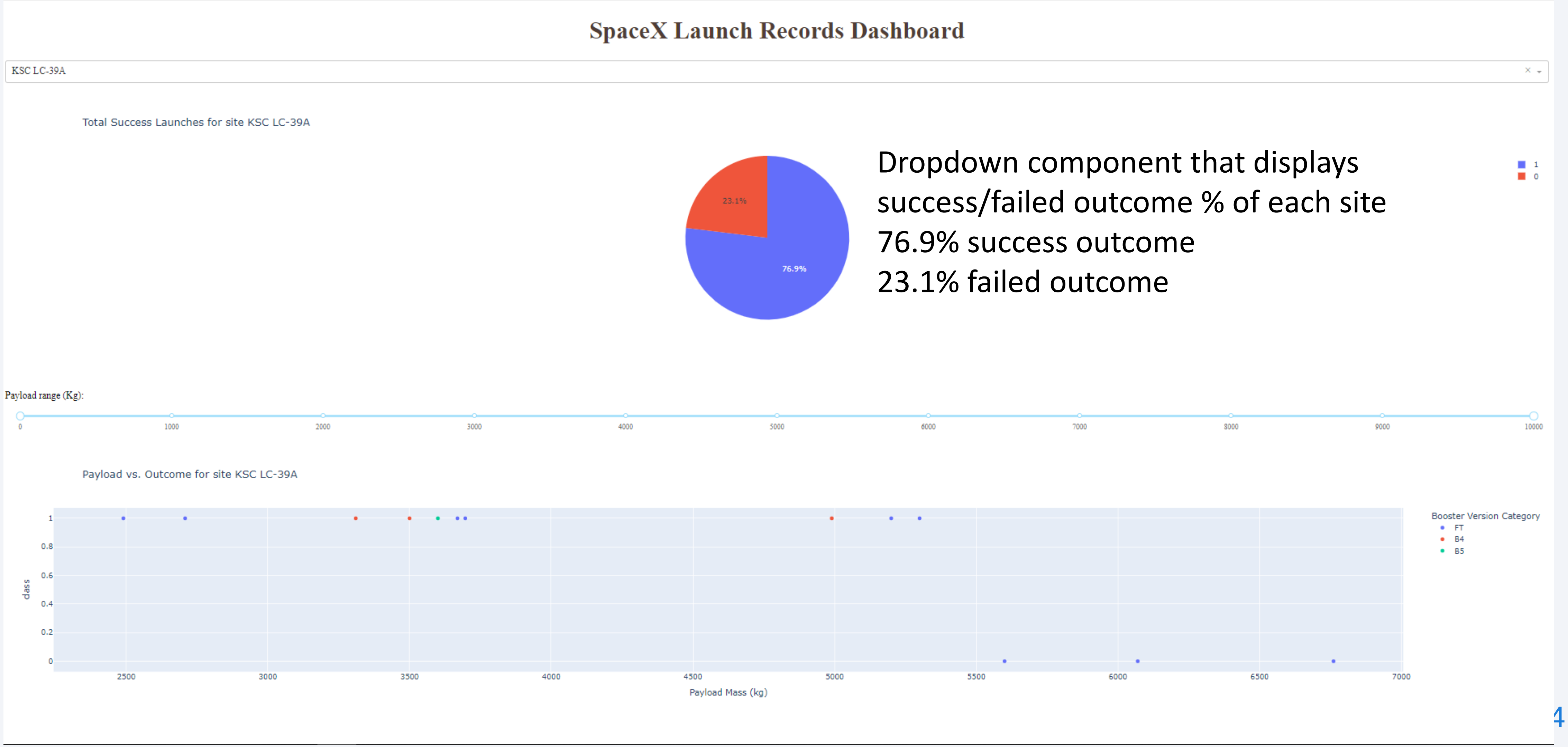
KSC LC-39A 41.7%

CCAFS LC-40 29.2%

VAFB SLC-4E 16.7%

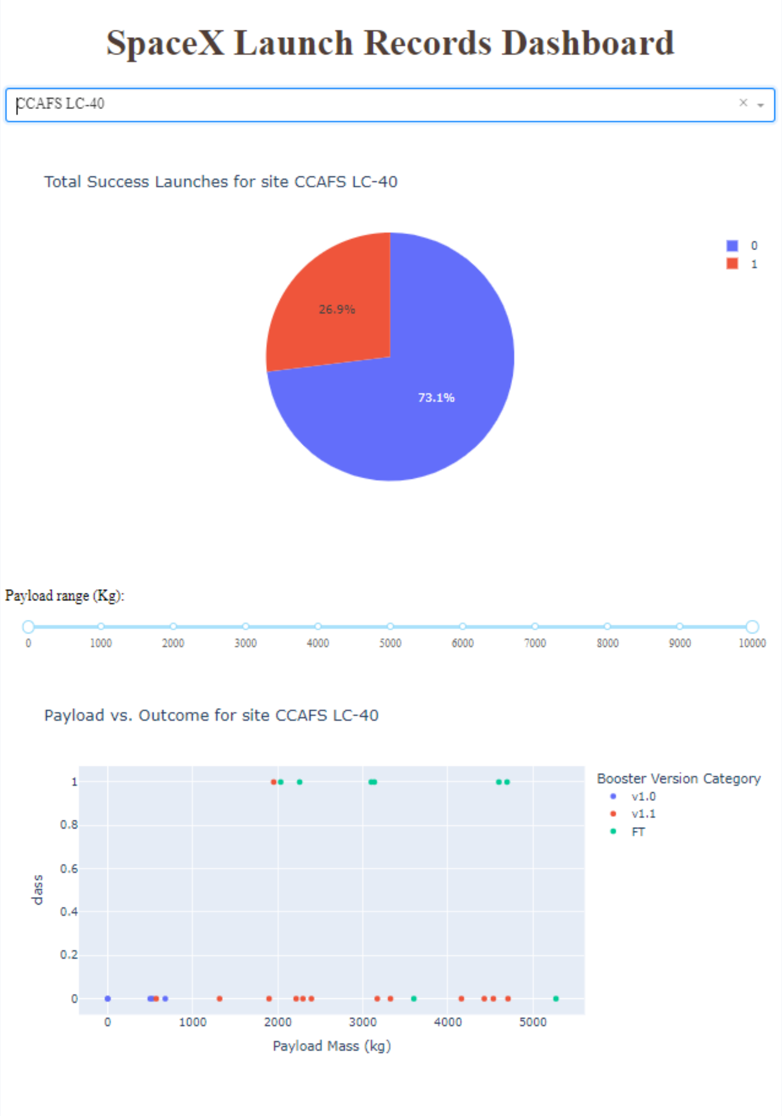
CCAFS SLC-40 12.5%

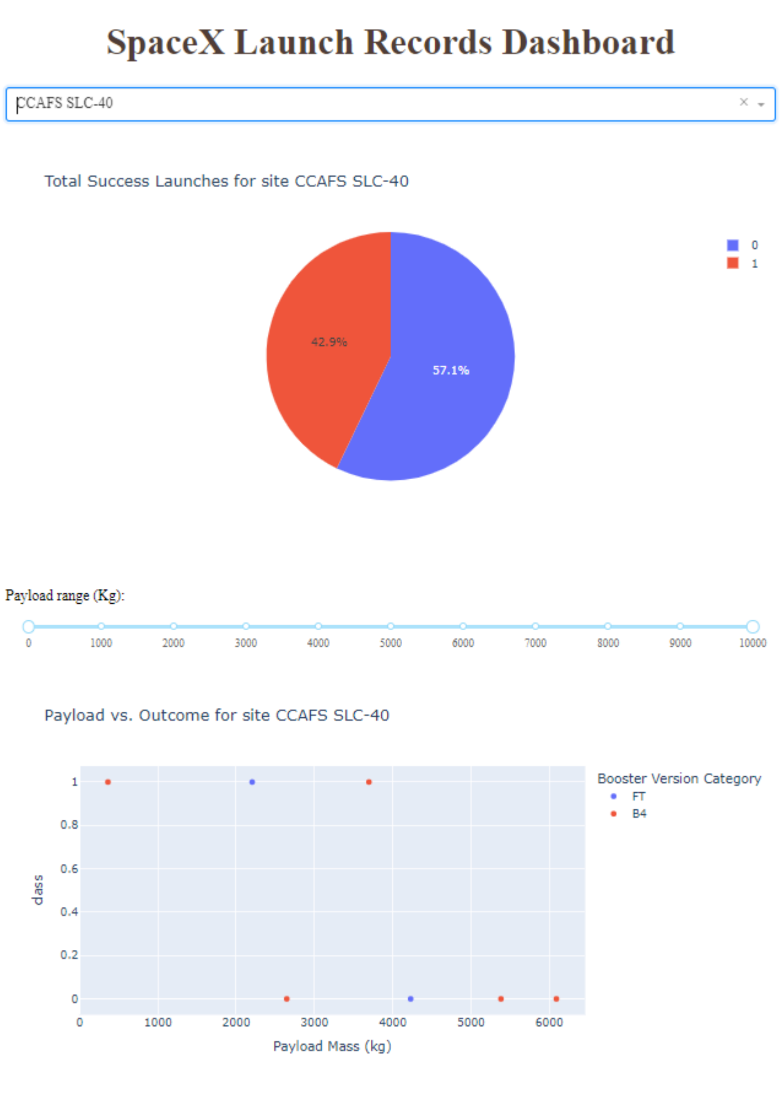
SpaceX Launch Records Dashboard – Site KSC LC-39A has the highest launch success ratio



Dropdown component that displays success/failed outcome % of each site
76.9% success outcome
23.1% failed outcome

SpaceX Launch Records Dashboard – Payload vs Outcome range slider

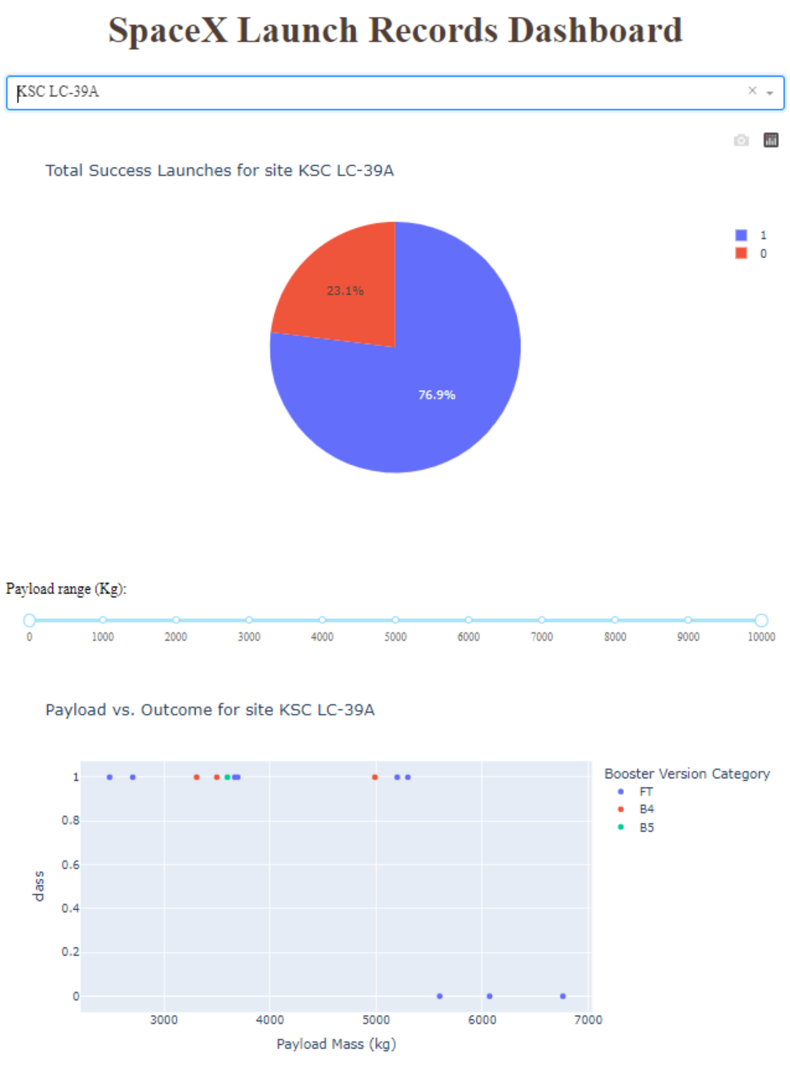


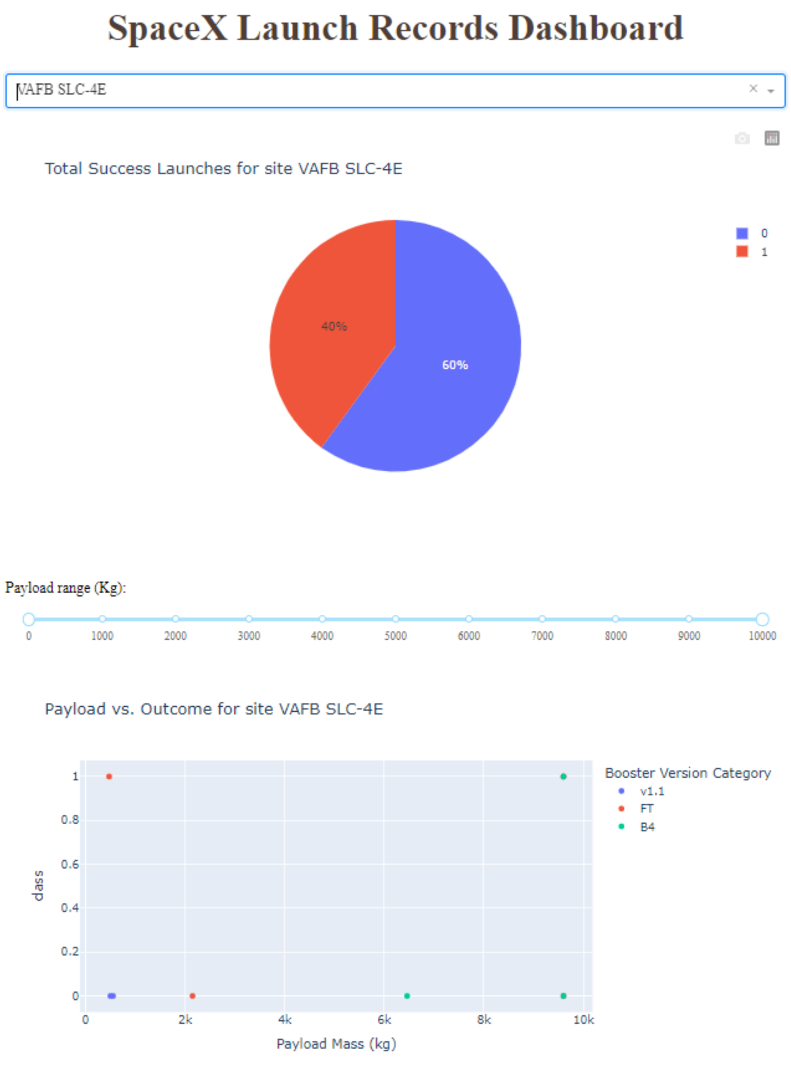


@app.callback includes component of "success-pie-chart"
Site CCAFS LC-40
26.9% success
73.1% fail
Success occurs only with FT Booster Version with PayLoad Mass over 2000kg.

Site CCAFS SLC-40
42.9% success
57.1% fail
Success rate seems higher when PayLoad Mass is under 4000kg and it doesn't seem to be dependent on Booster Version.

SpaceX Launch Records Dashboard – Payload vs Outcome range slider





@app.callback includes component of "success-pie-chart"
Site KSC LC-39A
23.1% success
76.9% fail
Success occurs only with PayLoad Mass under 5500kg and independent of Booster Version

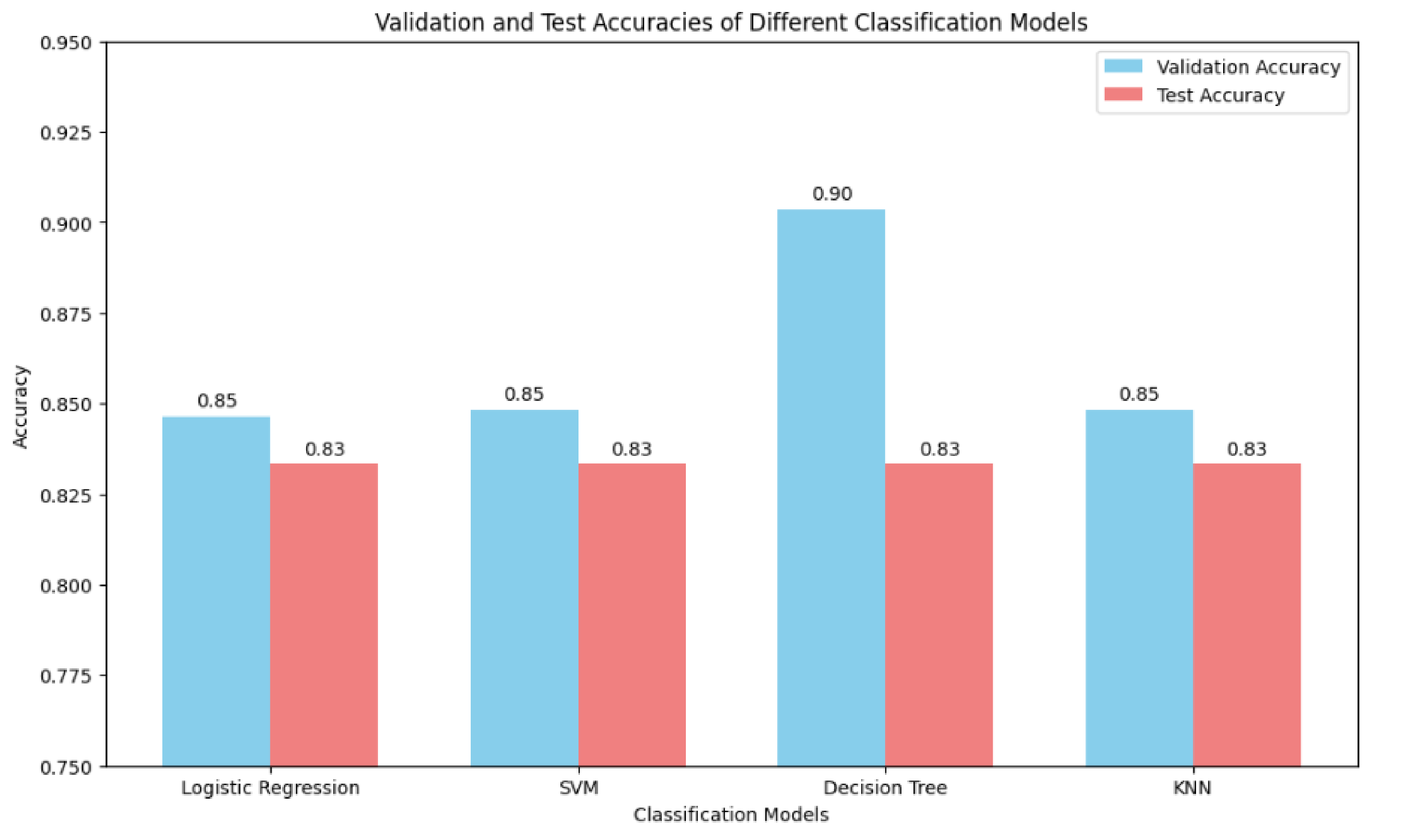
Site VAFB SLC-4E
40% success
60% fail
Success and failure rate seem to be unrelated to PayLoad Mass nor the Booster Version.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart



Classification Accuracy

Find which model has the highest classification accuracy:

All models gives the same model accuracy. However, the Decision Tree model gives the highest validation accuracy and hence is the best model according to the hyperparameters tuning results and validation accuracy.

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

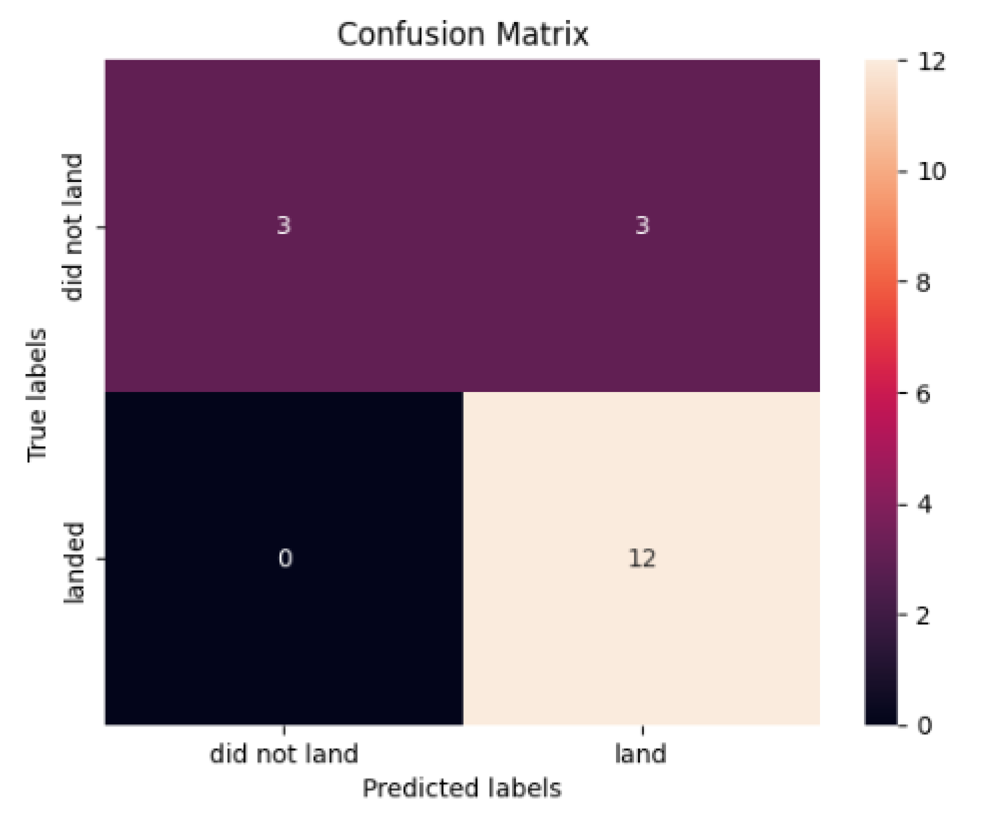
➢ True Negatives (TN): 3 instances where the model correctly predicted "did not land.

➢ "False Positives (FP): 3 instances where the model incorrectly predicted "land" when it was actually "did not land.

➢ "False Negatives (FN): 0 instances where the model incorrectly predicted "did not land" when it was actually "land.

➢ "True Positives (TP): 12 instances where the model correctly predicted "land.“

Accuracy = 15/18 X 100% = 83.3%



Conclusions

1. It is possible to collect the launch data through two ways i. SpaceX API ii. Webscraping with BeautifulSoap from Wikipedia.
2. Through the exploratory data analysis (EDA), the successful landing outcome is dependent on flight number, orbit, payload mass, site and Booster versions.
3. There are totally 56 launch sites and all are in the United States. All are next to the coast line.
4. The site KSC LC-39A has the highest successful rate.
5. Decision Tree model is the best model to predict the outcome of the launch.

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
- This is the GitHub repository including all the completed notebooks and Python files: https://github.com/ckmak9999/ibm_datascience_capstone

Thank you!