IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Avie Arellano
October 2024

Outline

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

Executive Summary

- **Summary of methodologies**

- Collected and cleaned data from the SpaceX API by using Pandas and dealing with missing values (replace with mean).
- Web scraped Falcon 9 historical launch records from Wikipedia using HTTP requests, BeautifulSoup and Pandas.
- Understood the SpaceX Dataset using a Db2 database and SQL queries.
- Explored and prepared data via Exploratory Data Analysis (EDA) and Feature Engineering using Seaborn, Pandas and Matplotlib.
- Performed EDA and determined training labels using Pandas and Numpy and by standardising data, logistic regression, confusion matrices, decision tree classifiers, support vector machines, and k nearest neighbours.
- Performed visual analytics on existing launch site locations using Folium.

- **Summary of all results**

- As the flight number increases, the first stage is more likely to land successfully. Even as the payload mass increases, the first stage still often return successfully.
- The Falcon 9 first stage will land successfully.
- If I can determine if the first stage will land, I can determine the cost of the launch.
- Effect of launch site locations on launch success rate.
- All launch sites are in proximity to the Equator line and are in very close proximity to the coast. Launch sites are in close proximity to railways ,highways and coastlines. Launch sites keep a certain distance away from cities.
- The method that performs best on test data is the Decision Tree classification algorithm.

Introduction

- **Project background and context**

- SpaceX is the only private company ever to return a spacecraft from low-earth orbit. It advertises Falcon 9 rocket launches with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of which the savings is because SpaceX can reuse the first stage.

- I will determine if the first stage will land, and thus 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**

- If the Falcon 9 first stage will land successfully?
- Pattern in flight number versus launch site
- Relationship between launch sites and their payload mass
- Relationship between success rate and each orbit type
- Relationship between flight number and each orbit type
- Relationship between payload mass and each orbit type
- Yearly trend of launch success
- Factors involved with optimal launch site location and launch success rate

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Made HTTP requests to SpaceX API and parsed requested data using Pandas
  - Web scraped and parsed Falcon 9 historical launch records from a Wikipedia page using BeautifulSoup, HTTP requests and Pandas
- Perform data wrangling
  - Cleaned/Processed requested data by dealing with missing values (replace with mean)
- 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

Version	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

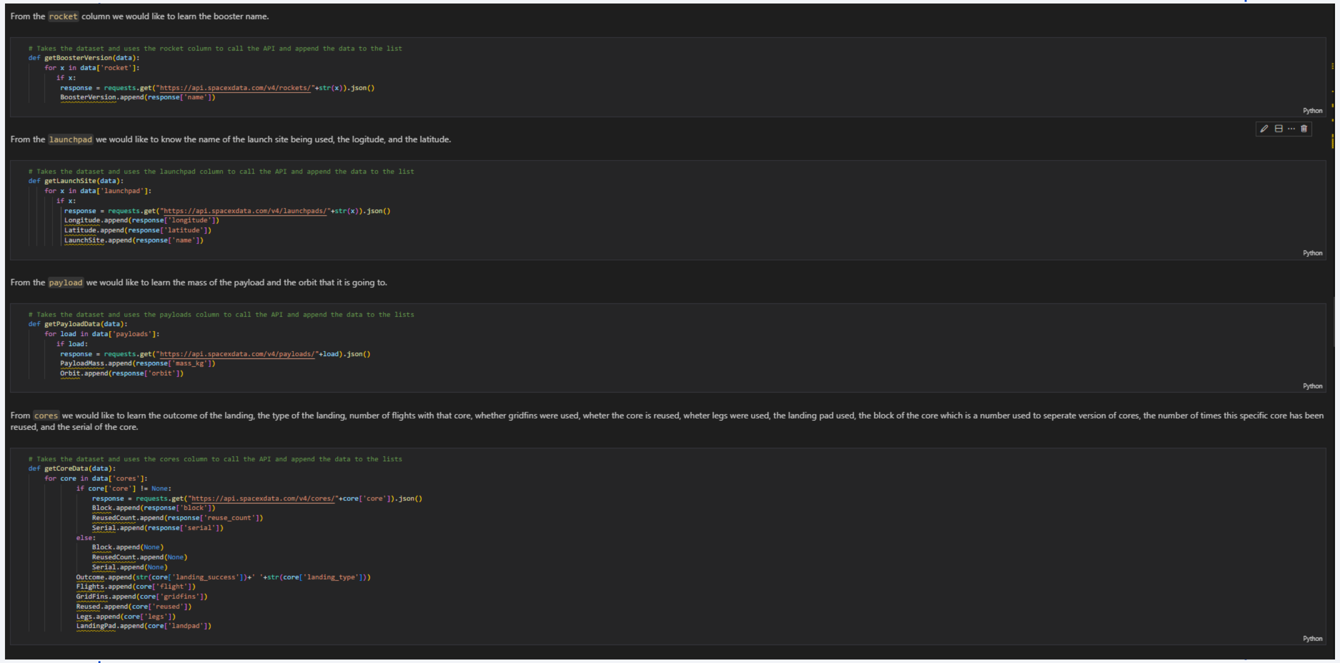
Data Collection

- **Describe how data sets were collected.**
- Made a HTTP GET request to SpaceX API to request rocket launch data.
- Parsed the requested data by turning it from JSON format to a Pandas data frame.
- Filtered the data frame to only include Falcon 9 launches.
- **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

- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-spacex-data-collection-api.ipynb



From the `rocket` column we would like to learn the booster name.

```python
# Takes the dataset and uses the rocket column to call the API and append the data to the list
def getBoosterVersion(data):
    for x in data['rocket']:
       if x:
        response = requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
        BoosterVersion.append(response['name'])
```

Python

From the `launchpad` we would like to know the name of the launch site being used, the logitude, and the latitude.

```python
# Takes the dataset and uses the launchpad column to call the API and append the data to the list
def getLaunchSite(data):
    for x in data['launchpad']:
       if x:
         response = requests.get("https://api.spacexdata.com/v4/launchpads/"+str(x)).json()
         Longitude.append(response['longitude'])
         Latitude.append(response['latitude'])
         LaunchSite.append(response['name'])
```

Python

From the `payload` we would like to learn the mass of the payload and the orbit that it is going to.

```python
# Takes the dataset and uses the payloads column to call the API and append the data to the lists
def getPayloadData(data):
    for load in data['payloads']:
       if load:
        response = requests.get("https://api.spacexdata.com/v4/payloads/"+load).json()
        PayloadMass.append(response['mass_kg'])
        Orbit.append(response['orbit'])
```

Python

From `cores` we would like to learn the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, wheter the core is reused, wheter legs were used, the landing pad used, the block of the core which is a number used to seperate version of cores, the number of times this specific core has been reused, and the serial of the core.

```python
# Takes the dataset and uses the cores column to call the API and append the data to the lists
def getCoreData(data):
    for core in data['cores']:
            if core['core'] != None:
                response = requests.get("https://api.spacexdata.com/v4/cores/"+core['core']).json()
                Block.append(response['block'])
                ReusedCount.append(response['reuse_count'])
                Serial.append(response['serial'])
            else:
                Block.append(None)
                ReusedCount.append(None)
                Serial.append(None)
            Outcome.append(str(core['landing_success'])+' '+str(core['landing_type']))
            Flights.append(core['flight'])
            GridFins.append(core['gridfins'])
            Reused.append(core['reused'])
            Legs.append(core['legs'])
            LandingPad.append(core['landpad'])
```

Python

Data Collection - Scraping

- **Present your web scraping process using key phrases and flowcharts**

- Performed a HTTP GET request method to request the Falcon 9 Launch HTML page as a HTTP response.

- Created a BeautifulSoup object from the HTML response.

- Extracted all column/variable names from the HTML table header.

- Created a Pandas data frame by parsing the launch HTML tables.

- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-webscraping.ipynb

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.07B0004.18	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.07B0005.18	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.07B0006.18	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.07B0007.18	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
343	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX Capella Space and Tyvak	Success\n	F9 B5B1051.10657	Success	9 May 2021	06:42
344	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8660	Success	15 May 2021	22:56
345	119	CCSFS	Starlink	15,600 kg	LEO	NASA (CRS)	Success\n	F9 B5B1063.2665	Success	26 May 2021	18:59
346	120	KSC	SpaceX CRS-22	3,328 kg	LEO	Sirius XM	Success\n	F9 B5B1067.1668	Success	3 June 2021	17:29
347	121	CCSFS	SXM-8	7,000 kg	GTO	NaN	NaN	F9 B5	NaN	6 June 2021	04:26

348 rows × 11 columns

Data Wrangling

- **Describe how data were processed**
- Performed EDA to find some patterns in the data and determine what would be the label for training supervised models.
- Calculated the number of launches on each site.
- Calculated the number and occurrence of each orbit.
- Calculated the number and occurrence of mission outcome of the orbits.
- Created landing outcome label from Outcome column.
- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/labs-jupyter-spacex-Data%20wrangling.ipynb

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: count, dtype: int64
```

```
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
HEO       1
ES-L1     1
SO        1
GEO       1
Name: count, dtype: int64
```

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: count, dtype: int64
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

EDA with Data Visualization

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

- Cat plot of Flight Number and Payload mass to see how these 2 variables would affect the launch outcome.
- Scatter point chart of Flight Number and Launch Site to see how they would affect launch outcome.
- Scatter point chart of Payload Mass and Launch Site to see how they would affect launch outcome.
- Bar chart of Orbit Types and Success Rate to visually check if they have a relationship.
- Scatter point chart of Flight Number and Orbit Types too see if they have any relationship.
- Scatter point chart of Payload Mass and Orbit Types to reveal their relationship.
- Line chart of Average Success Rate and Year to visualise the launch success yearly trend.

- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/edadataviz.ipynb

EDA with SQL

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

- SELECT … FROM … WHERE … AND … LIKE … LIMIT …
- Sum(…)
- Avg(…)
- Min(…)
- SELECT DISTINCT … COUNT(…) AS … FROM … GROUP BY …
- MAX(…)
- SUBSTR(…, …, …)
- ORDER BY …
- BETWEEN … AND …

- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-eda-sql-coursera_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**

- Circle and icon name marker at NASA Johnson Space Centre to initialise it as a centre location.
- Circle and icon name marker for each launch site to check their proximity to the Equator line and to the coast.
- Colour-labelled markers for all launch records to identify which launch sites have relatively high success rates.
- PolyLine between a launch site and nearest coastline/railway/highway/city point to check the proximity to railways, highways, coastlines and cities.

- **https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/lab_jupyter_launch_site_location.ipynb**

Build a Dashboard with Plotly Dash

- **Summarize what plots/graphs and interactions you have added to a dashboard**
- **Explain why you added those plots and interactions**

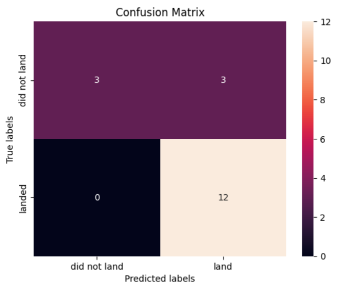
- Added a dropdown list to enable Launch Site selection.
- Added a pie chart to show the total successful launches count for all sites and for each site, showed the Success vs. Failed counts.
- Added a scatter chart to show the correlation between payload and launch success.

- **https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/spacex_dash_app.py**

```
array([0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1], dtype=int64)
```

```
Train set: (72, 83) (72,)
Test set: (18, 83) (18,)
```

```
GridSearchCV
GridSearchCV(cv=10, estimator=LogisticRegression(),
             param_grid={'C': [0.01, 0.1, 1], 'penalty': ['l2'],
                         'solver': ['lbfgs']})
estimator: LogisticRegression
LogisticRegression()
LogisticRegression
LogisticRegression()
```



Predictive Analysis (Classification)

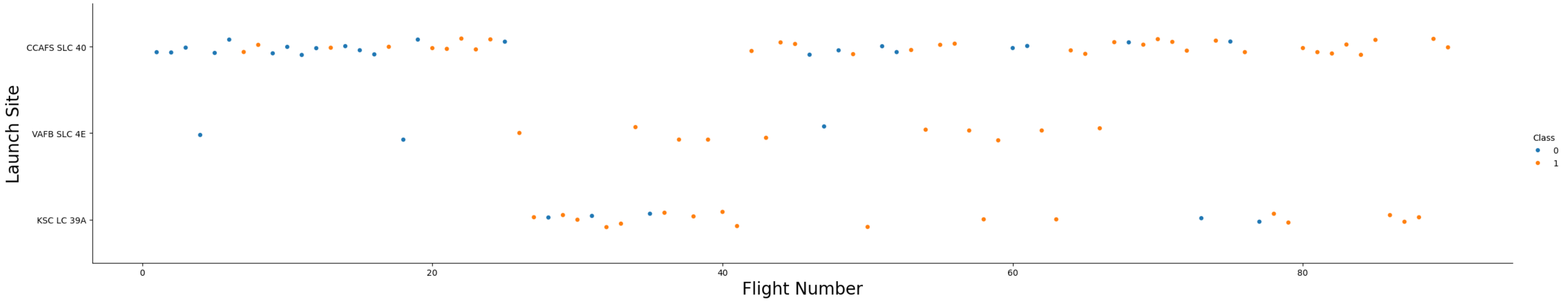
- **Summarize how you built, evaluated, improved, and found the best performing classification model**
- Converted Class column to a NumPy Array.
- Standardised data using a transform method.
- Split the data into training and test data.
- Used logistic regression, support vector machine (SVM), decision tree and k nearest neighbours classification models.
- Calculated the accuracy on the test data using the score method and confusion matrices.
- https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

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

Section 2

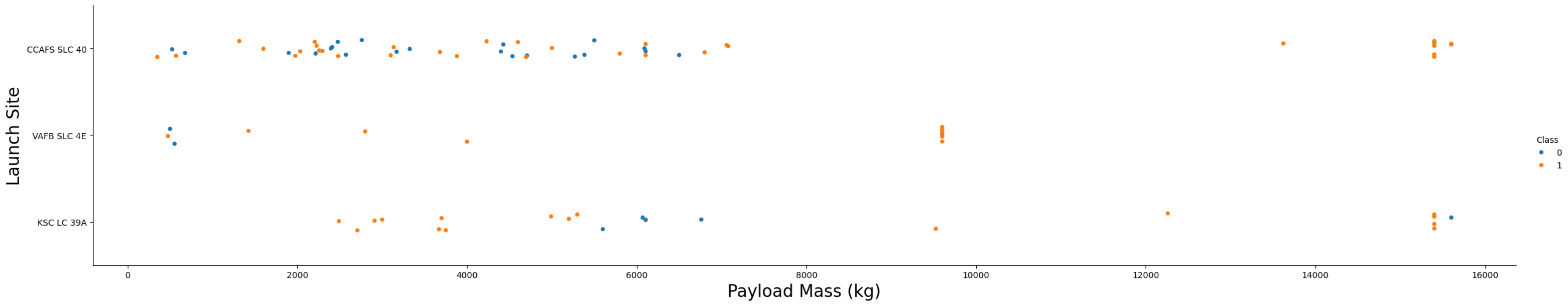
Insights drawn from EDA



Flight Number vs. Launch Site

- **Show the screenshot of the scatter plot with explanations**

- In general, as flight number increases, the number of successful launches increases for each launch site.
- Most number of failed launches at earlier flights took place at launch site CCAFS SLC-40.



Payload vs. Launch Site

- **Show the screenshot of the scatter plot with explanations**

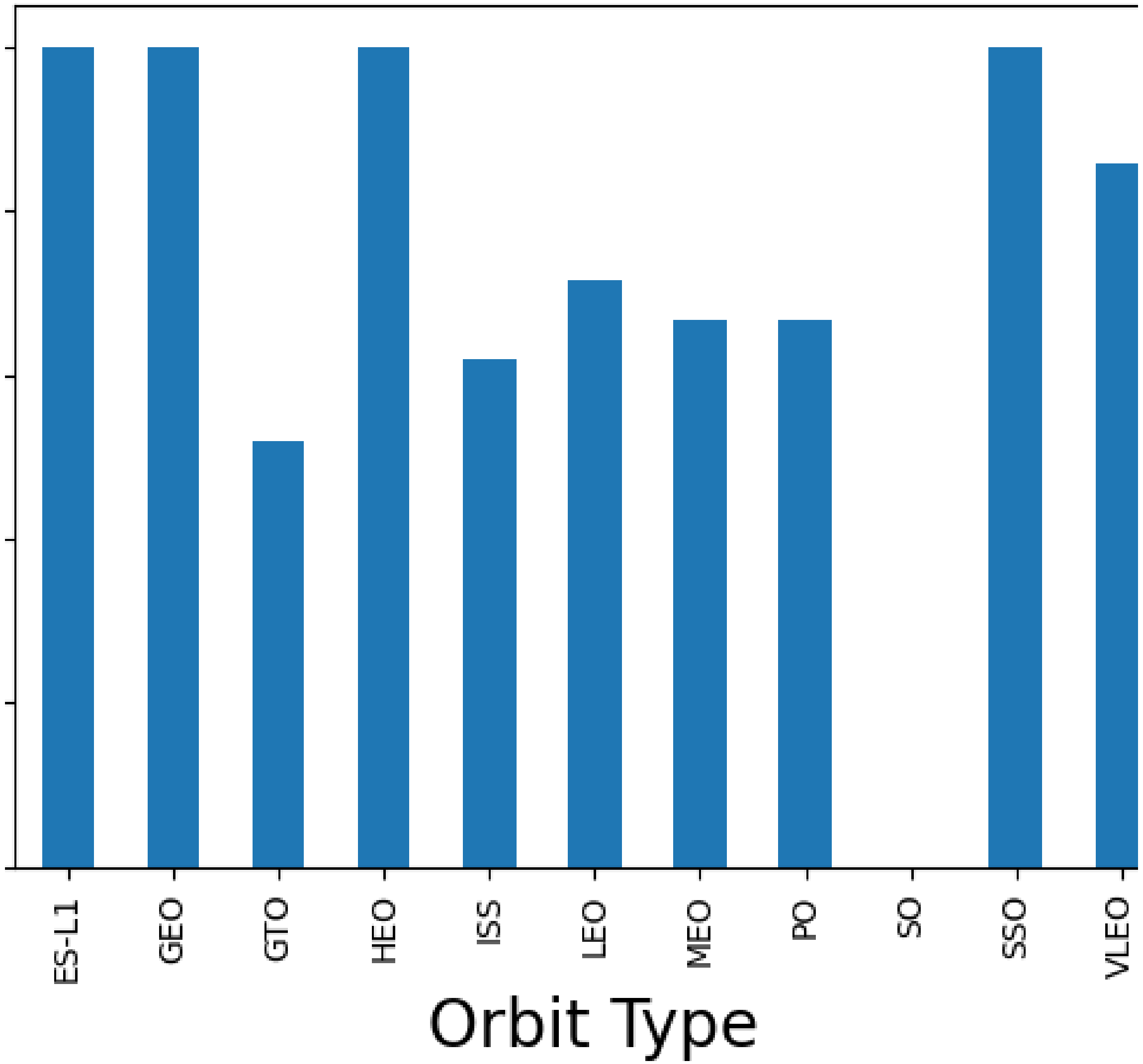
- At the VAFB-SLC launch site, there are no rockets launched for heavy payload mass greater than 10,000 kg.

- Launch sites CCAFS-SLC and KSC-LC handle heavier payload mass (> 14,000 kg) better than VAFB-SLC.

Success Rate vs. Orbit Type

- **Show the screenshot of the scatter plot with explanations**

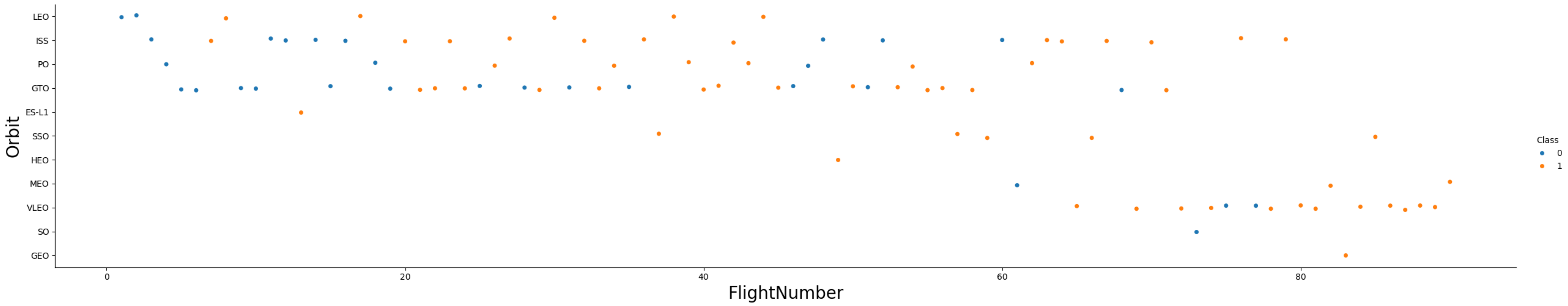
- Orbit type SO has 100% failures.
- Orbit types ES, GEO, HEO and SSO have 100% success rate.
- The rest of the orbit types have at least a 50% chance of success.



Flight Number vs. Orbit Type

- **Show the screenshot of the scatter plot with explanations**

- In the LEO and MEO orbits, success seems to be related to the number of flights.
- Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

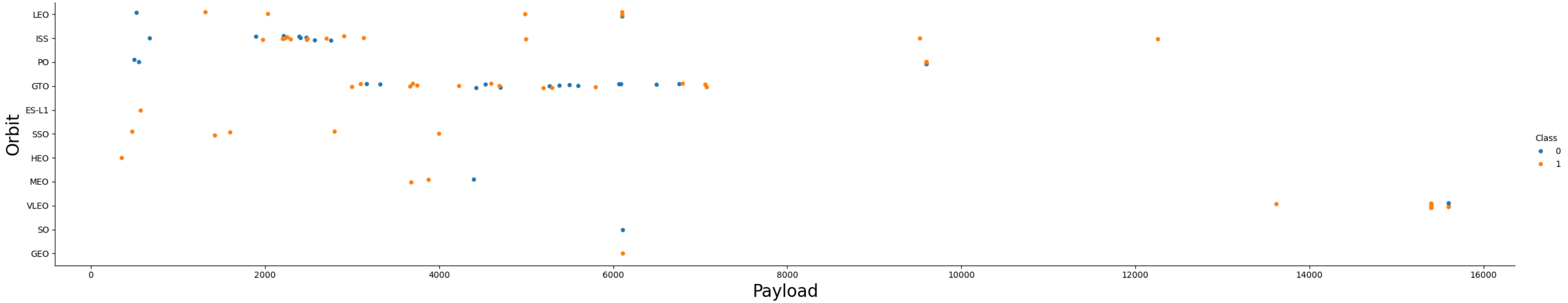


Payload vs. Orbit Type

- **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 orbits.

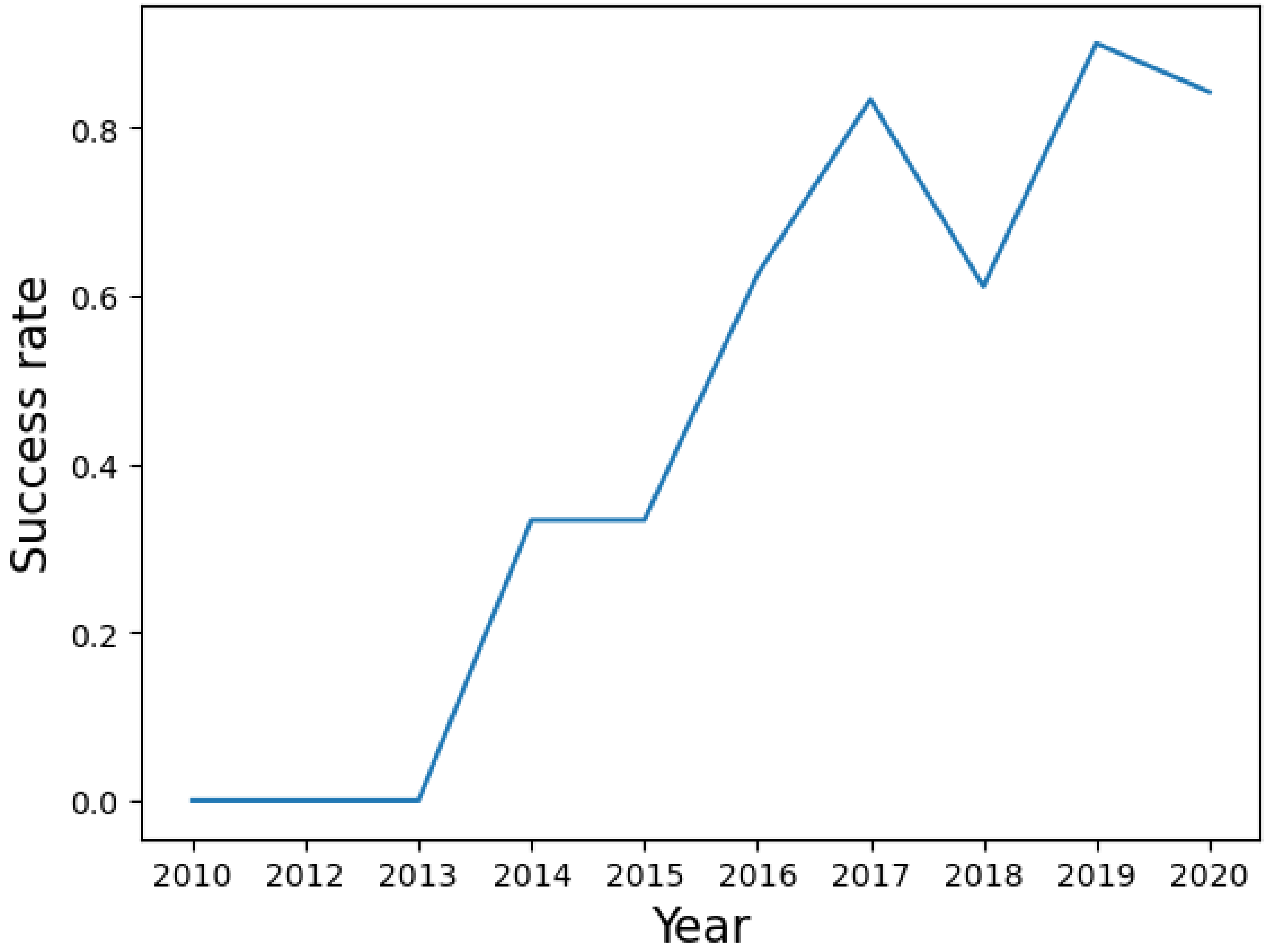
- However, for GTO, it is difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

- **Show the screenshot of the scatter plot with explanations**

- In general, the success rate since 2013 kept increasing until 2020.
- The success rate from 2010 to 2013 remain unchanged.



All Launch Site Names

```
%sql select distinct Launch_Site from SPACEXTBL
```

* sqlite:///my_data1.db
Done.

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

- **Find the names of the unique launch sites**
- **Present your query result with a short explanation here**
- CCAFS LC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40 are the names of the unique launch sites in the space mission.

```
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' LIMIT 5
```

Python

* sqlite:///my_data1.db
Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

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

- The first 5 records from the launch site CCAFS LC-40.

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

Pyt

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

sum(PAYLOAD_MASS__KG_)
45596

Total Payload Mass

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

- The total payload mass carried by boosters from NASA is 45,596 kg.

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

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

avg(PAYLOAD_MASS__KG_)
2534.666666666665

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**

\- The average payload mass carried by booster version F9 v1.1 is around 2534.67 kg.

```
%sql select min(Date) from SPACEXTBL where Landing_Outcome = "Success (ground pad)"

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

min(Date)
2015-12-22

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

\- The first successful landing outcome on ground pad happened on 2015-12-22.

First Successful Ground Landing Date

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql

select Booster_Version from SPACEXTBL
where Landing_Outcome = "Success (drone ship)"
    and PAYLOAD_MASS__KG_ > 4000
    and 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

- **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**

\- Booster version F9 FT B1022, F9 FT B1026, F9 FT B1021.2 and F9 FT B1031.2 all have success in drone ship and have payload mass greater than 4000kg but less than 6000kg.

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**

- The total number of failure (in flight) is 1. The total number of successful mission outcomes is 100 where 1 of them has an unclear payload status.

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL);
```

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

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

- **List the names of the booster which have carried the maximum payload mass**
- **Present your query result with a short explanation here**
- Those are the names of the booster versions which have carried the maximum payload mass.
- I used a subquery.

```
%%sql
select substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL
where substr(Date,0,5)='2015' and Landing_Outcome = "Failure (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

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**

- These are the records that have a failed landing outcome for drone ship in the year 2015.

```
%%sql
SELECT LANDING_OUTCOME, COUNT(LANDING_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING_OUTCOME
ORDER BY TOTAL_NUMBER DESC
```

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

Landing_Outcome	TOTAL_NUMBER
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

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**

- This is the rank of the count of landing outcomes between the dates 2010-06-04 and 2017-03-20, in descending order.

Section 3

Launch Sites Proximities Analysis

Map of all Launch Sites

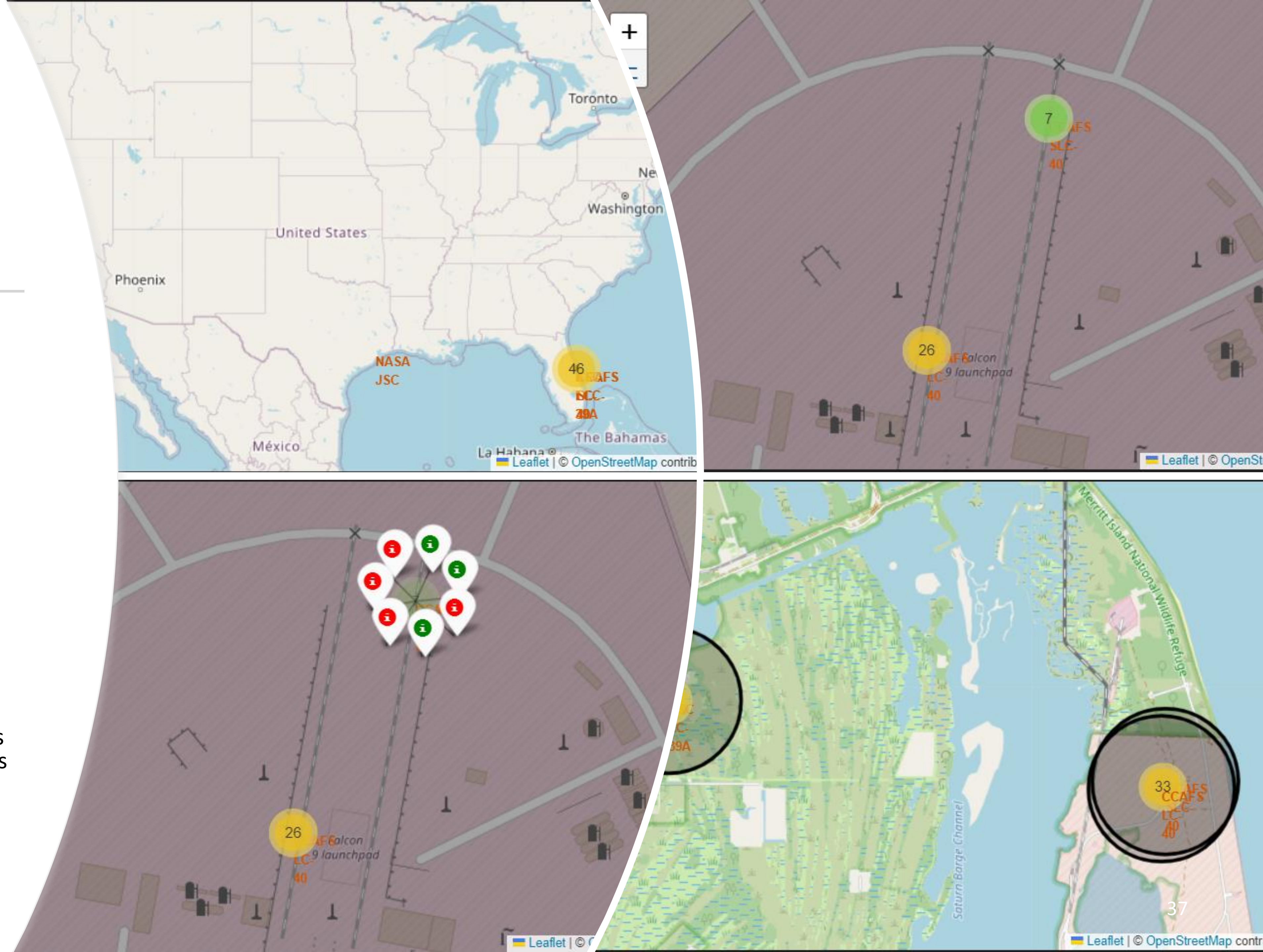
- **Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map**
- **Explain the important elements and findings on the screenshot**

- There are 4 marked launch sites on the map.
- All launch sites are in proximity to the Equator line and are in very close proximity to the coast.

Launch Outcomes for Each Site

- **Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map**
- **Explain the important elements and findings on the screenshot**

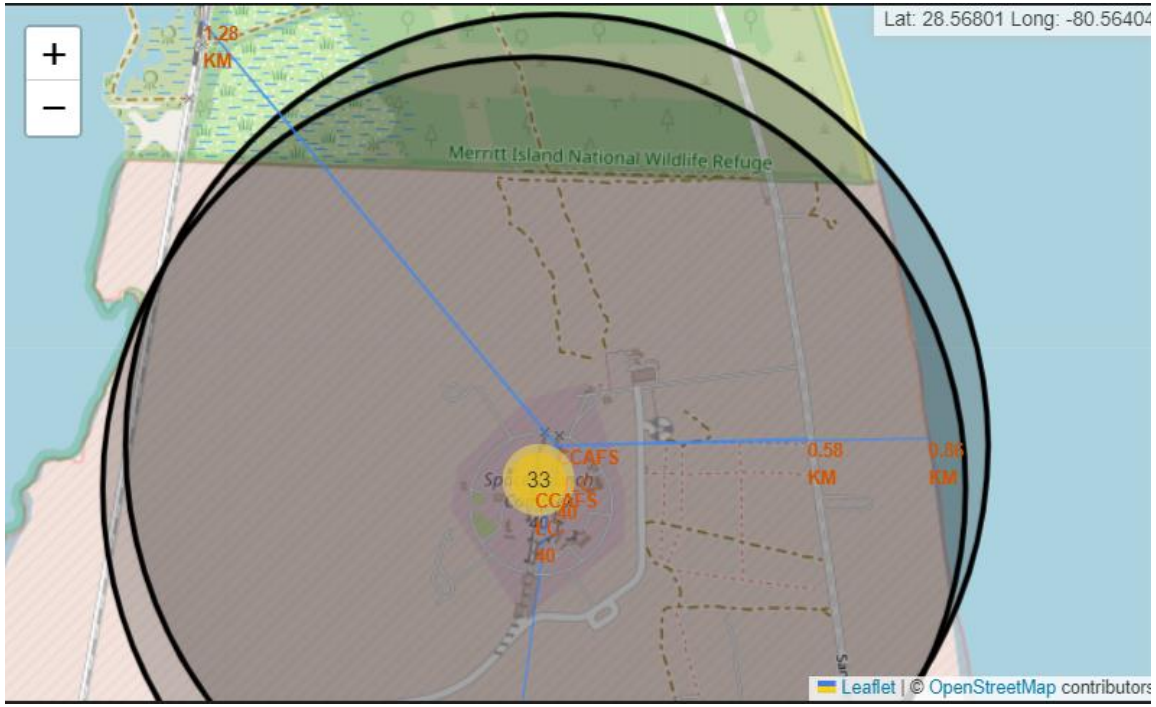
- If a launch was successful, a green marker was used and if a launch was failed, a red marker was used.
- From the colour-labelled markers in marker clusters, I can easily identify which launch sites have relatively high success rates (KSC LC-39A).



Launch Site Proximity to Railway/Highway/Coastline Map

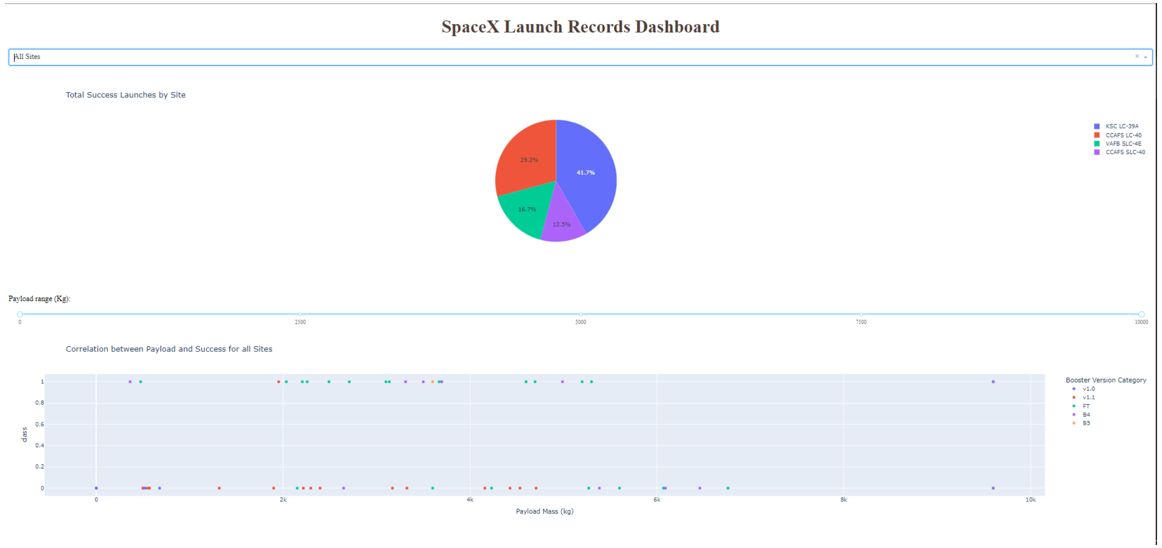
- **Explain the important elements and findings on the screenshot**

- Some launch sites are in close proximity to railways, highways and coastlines.
- Some launch sites keep a certain distance away from cities.
- For example, launch site CCAFS SLC-40 is 1.28km away from a railway to transport heavy cargo.
- It is 0.58km from a highway to transport personnel and equipment.
- It is 0.86km away from a coastline to allow launch abortions, water landings, and minimize risk from falling debris.
- It is 51.43km away from a city to minimise danger to population dense areas.



Section 4

Build a Dashboard with Plotly Dash



Total Success Launches by Site

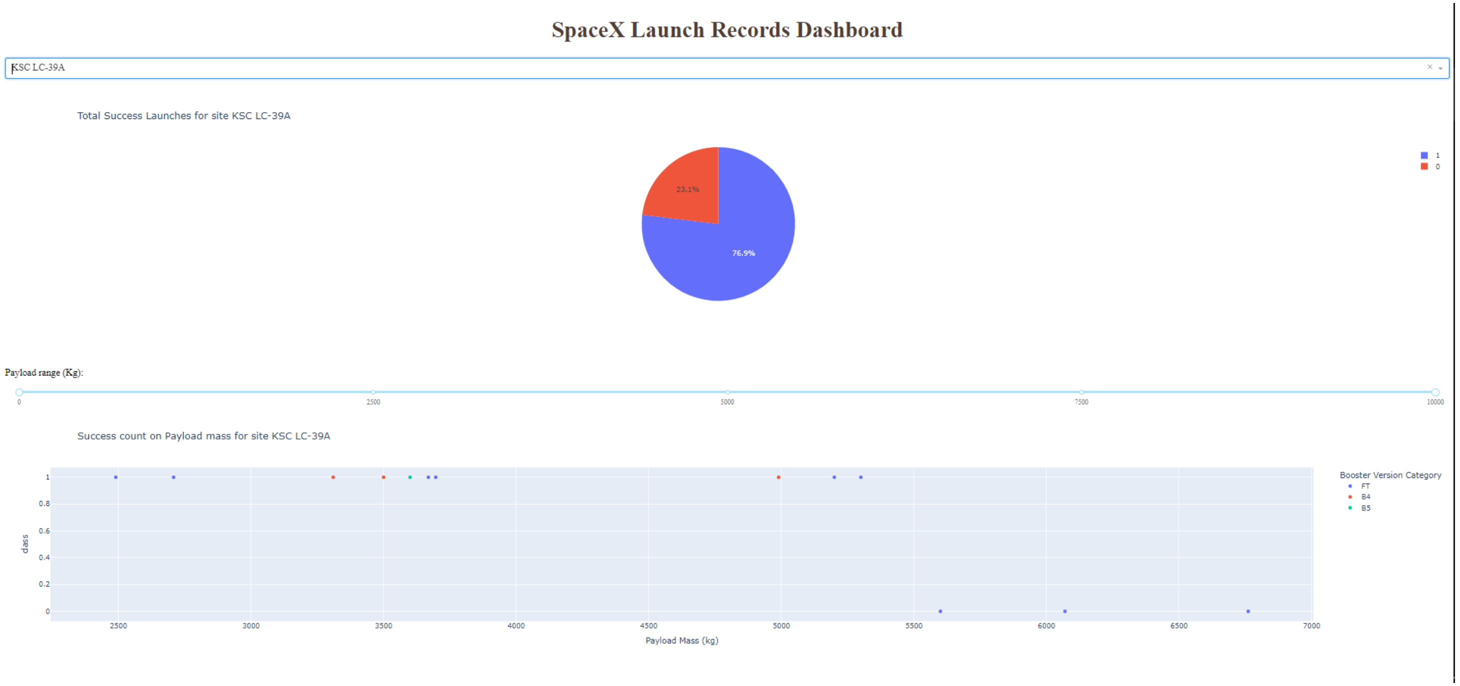
- **Show the screenshot of launch success count for all sites, in a piechart**
- **Explain the important elements and findings on the screenshot**

- Launch site KSC LC-39A had the most success count followed by CCAFS LC-40, VAFB SLC-4E, and CCAFS SLC-40.

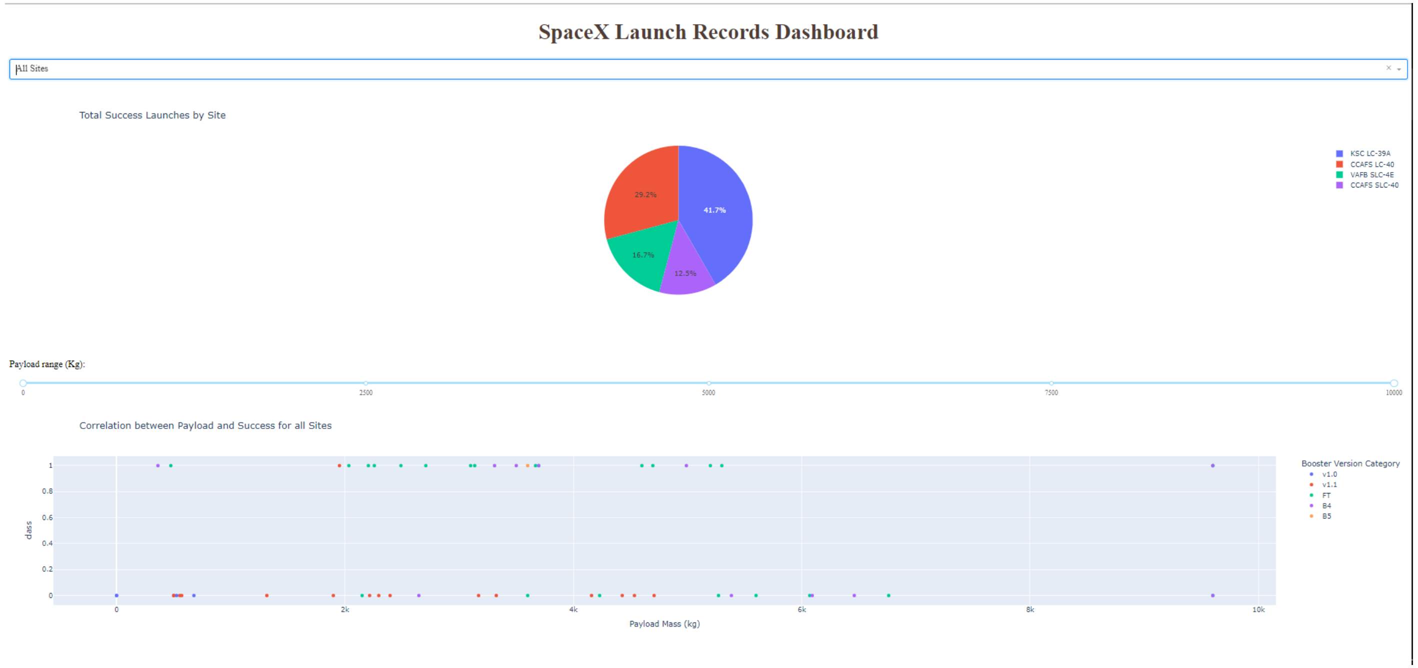
Total Success Launches for site KSC LC-39A

- **Show the screenshot of the piechart for the launch site with highest launch success ratio**
- **Explain the important elements and findings on the screenshot**

- Launch site KSC LC-39A had the highest launch success ratio of 76.9%.



Correlation between Payload and Success for all Sites



- **Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.**

- FAFB SLC-4E has the heaviest successful booster landing success.
- Payload mass less than 5,300 kg had the highest booster landing success rate.
- Payload mass greater than 5,300 kg had the lowest booster landing success rate.

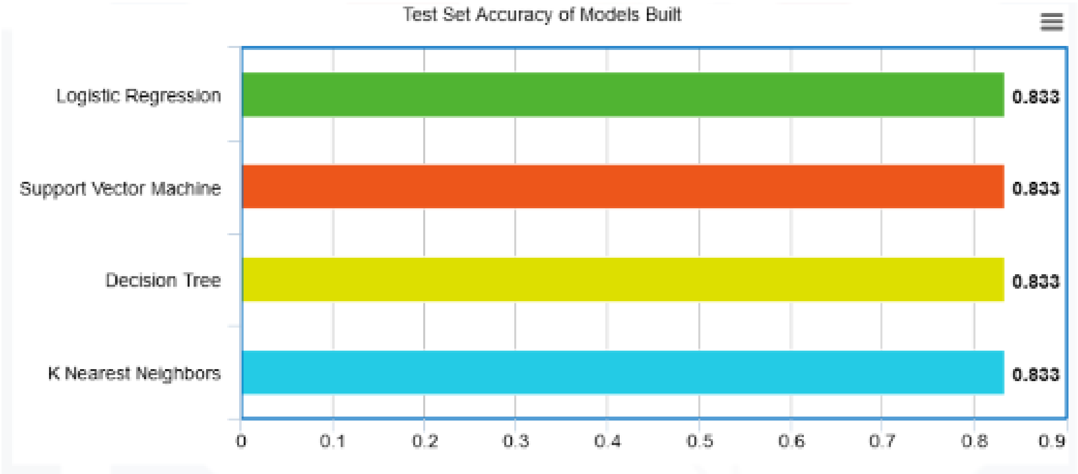
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- **Visualize the built model accuracy for all built classification models, in a bar chart**
- **Find which model has the highest classification accuracy**

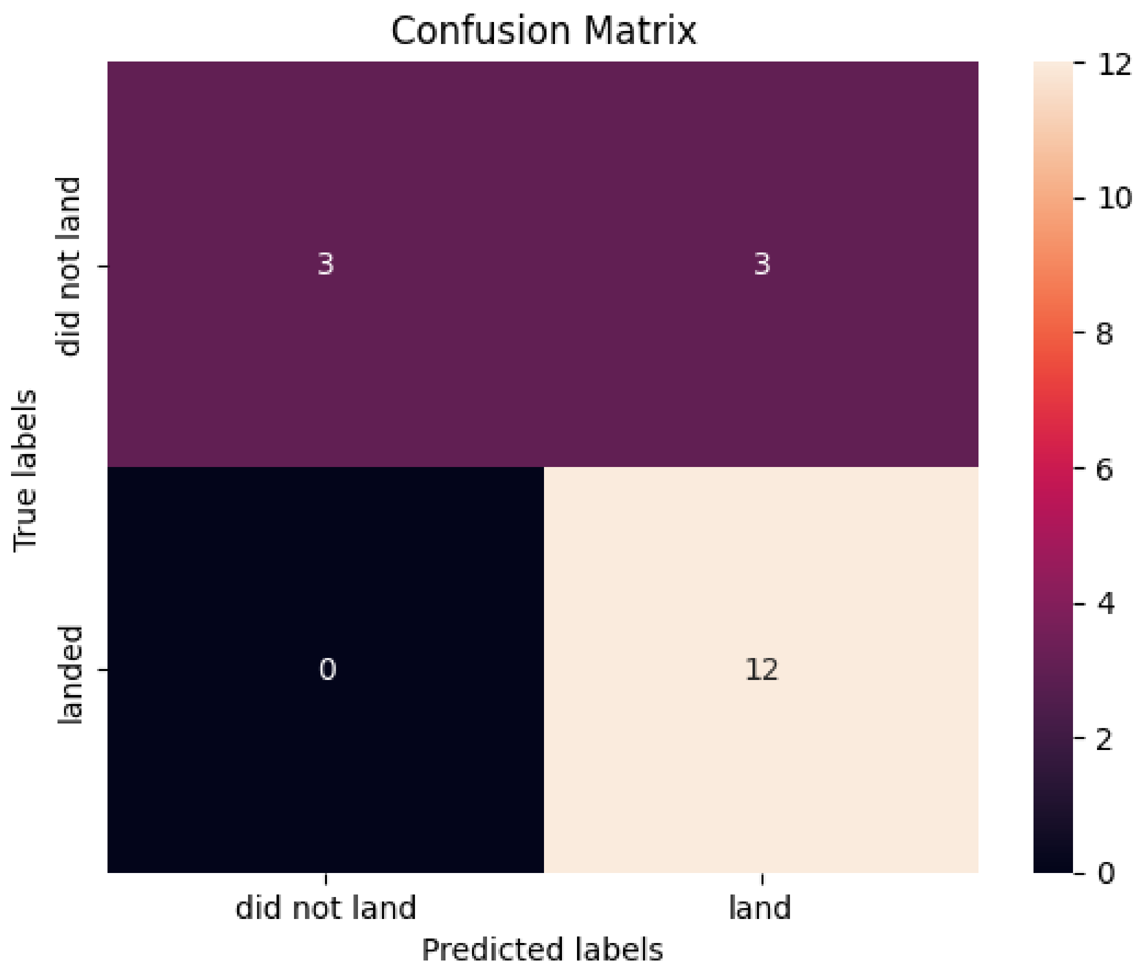
- All models had the same classification accuracy of around 83.3%.



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

The confusion matrix of the best performing model is a 4-way tie. A major problem is false positives as evidenced by the model predicting the first stage booster to land in 3 out of 18 samples in the test set.

Confusion Matrix



Conclusions

- I can predict with about 83.33% accuracy that SpaceX will successfully land the first stage booster.
- The first stage booster costs upward of $15 million to build.
- This will enable me to make more informed bids against SpaceX, since I will have a good idea of when to expect the SpaceX bid to include the cost of a sacrificed first stage booster.
- The SpaceX bid would cost upward of $77 million after sacrificing $15+ million at the first stage and with a list price of $62 million per 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

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-eda-sql-coursera_sqllite.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-spacex-data-collection-api.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/jupyter-labs-webscraping.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/lab_jupyter_launch_site_location.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/labs-jupyter-spacex-Data%20wrangling.ipynb

https://github.com/kyeav/IBM_DS/blob/master/Peer%20Reviewed%20Assignment%20of%20the%20Applied%20Data%20Science%20Capstone%20course/spacex_dash_app.py

Thank you!