



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adiel Wiryadinata
25 November 2025



Outline

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

Executive Summary

In this project, a competitor of SpaceX — referred to as *SpaceY* — analyzes Falcon 9 rocket data to assess the success rate of first-stage landings and estimate the cost per launch. The following sections provide a summary of the methodologies and results presented in this report.

Summary of methodologies

- Data Collection
- Data wrangling
- Exploratory data analysis with data visualization and SQL
- Building an interactive map with Folium
- Building Dashboard with Plotly Dash
- Predictive analysis (classification)

Summary of all results

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

Introduction

This capstone project is part of the **IBM Data Science Professional Certificate** program. The objective is to demonstrate proficiency in **data science and machine learning techniques** by applying them to a real-world dataset and summarizing the findings in a comprehensive report.

In this project, a **rival company to SpaceX**, referred to as *SpaceY*, analyzes **Falcon 9 rocket data** to evaluate the success of first-stage landings and estimate the **cost per launch**. The purpose of this analysis is to enable SpaceY to **compete with SpaceX** in future rocket launch bids. SpaceX advertises its Falcon 9 launch cost at **USD 62 million**, while other companies face costs exceeding **USD 165 million** per launch.

All data collection and analysis are conducted using **Python** within **Jupyter Notebooks**. These notebooks, along with the final report in PDF format, are stored in the author's **GitHub repository**.

This report covers the complete data science workflow, including **data collection methodology, data wrangling, exploratory data analysis (EDA), interactive data visualization, machine learning (ML) model development, and model evaluation**. The performance of multiple ML algorithms is compared to determine their **accuracy in predicting the likelihood of successful Falcon 9 first-stage landings**.

Section 1

Methodology

Methodology

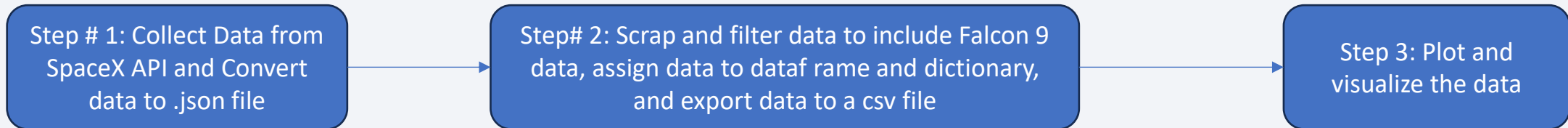
The data used in this project were collected from the **SpaceX REST API** and the **Wikipedia Launch Table**. The data wrangling process involved **cleaning, structuring, and preparing the dataset** for visualization and for use in **machine learning (ML) predictive models**. Several ML algorithms were applied, including **Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)**.

An **Exploratory Data Analysis (EDA)** was conducted using a combination of **data visualization** techniques and **SQL queries** to uncover trends and relationships within the data. Tools such as **Folium** and **Plotly Dash** were used to create **interactive visualizations** and **geographical representations** of the launch data.

Finally, **predictive modeling** was performed using **classification algorithms** implemented in **Scikit-learn** to predict whether the **Falcon 9 first stage** would successfully land. The performance of these models was evaluated and compared based on their **accuracy scores** and predictive capability.

Data Collection

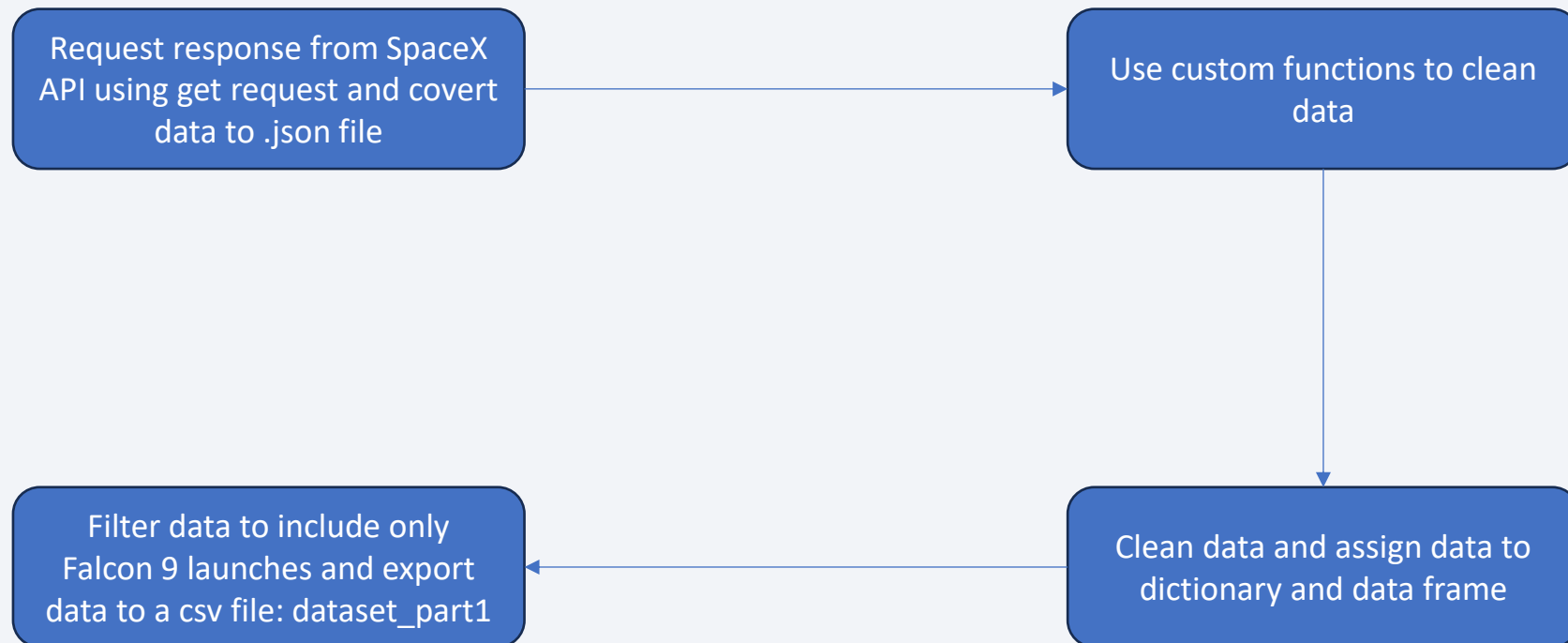
Data collection and visualization major steps:



FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount
1	2010-06-04	Falcon 9	6123.547647058824	LEO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False		1.0	0
4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False		1.0	0
5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
6	2014-01-06	Falcon 9	3325.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
7	2014-04-18	Falcon 9	2296.0	ISS	CCSFS SLC 40	True Ocean	1	False	False	True		1.0	0
8	2014-07-14	Falcon 9	1316.0	LEO	CCSFS SLC 40	True Ocean	1	False	False	True		1.0	0
9	2014-08-05	Falcon 9	4535.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
10	2014-09-07	Falcon 9	4428.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0
11	2014-09-21	Falcon 9	2216.0	ISS	CCSFS SLC 40	False Ocean	1	False	False	False		1.0	0
12	2015-01-10	Falcon 9	2295.0	ISS	CCSFS SLC 40	False Ocean	1	True	False	True	5-B-2022202-1-3761621-7th	1.0	0

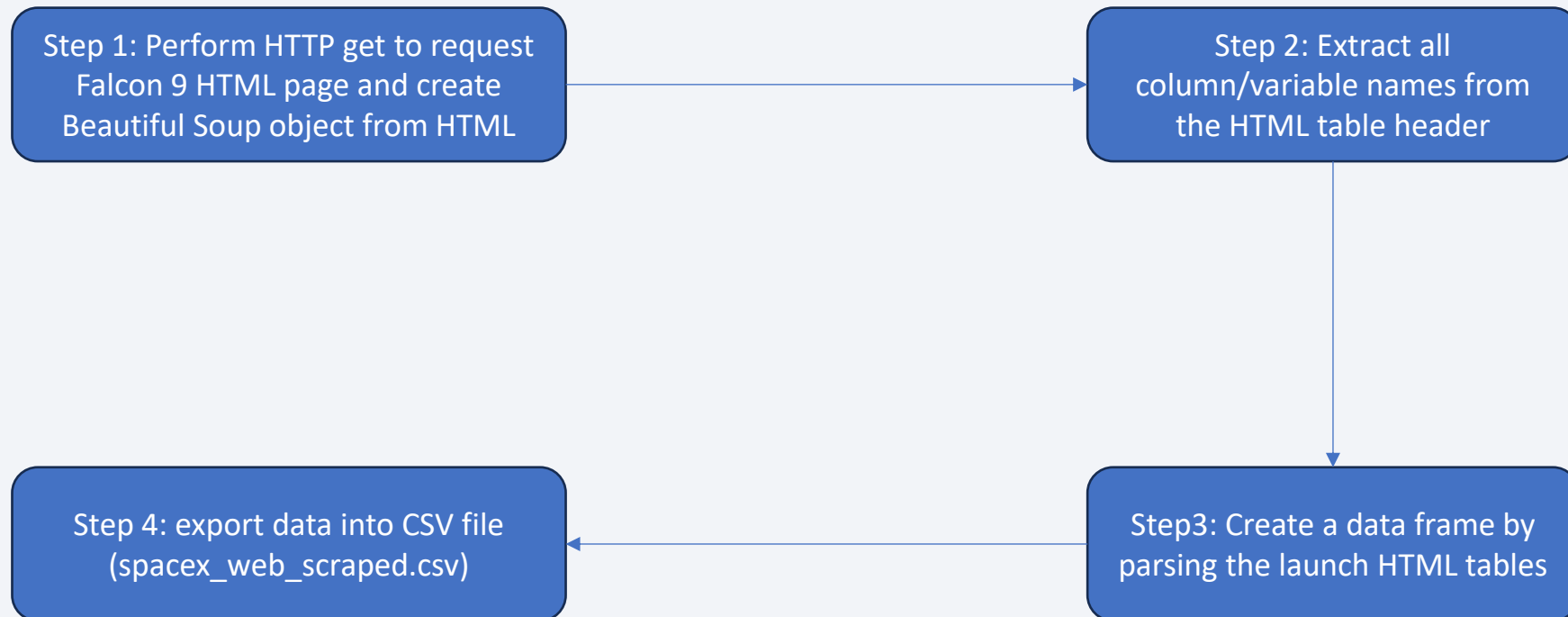
https://github.com/adielgithub123/DataScienceCapstone/blob/main/dataset_part_1.csv

Data Collection – SpaceX API



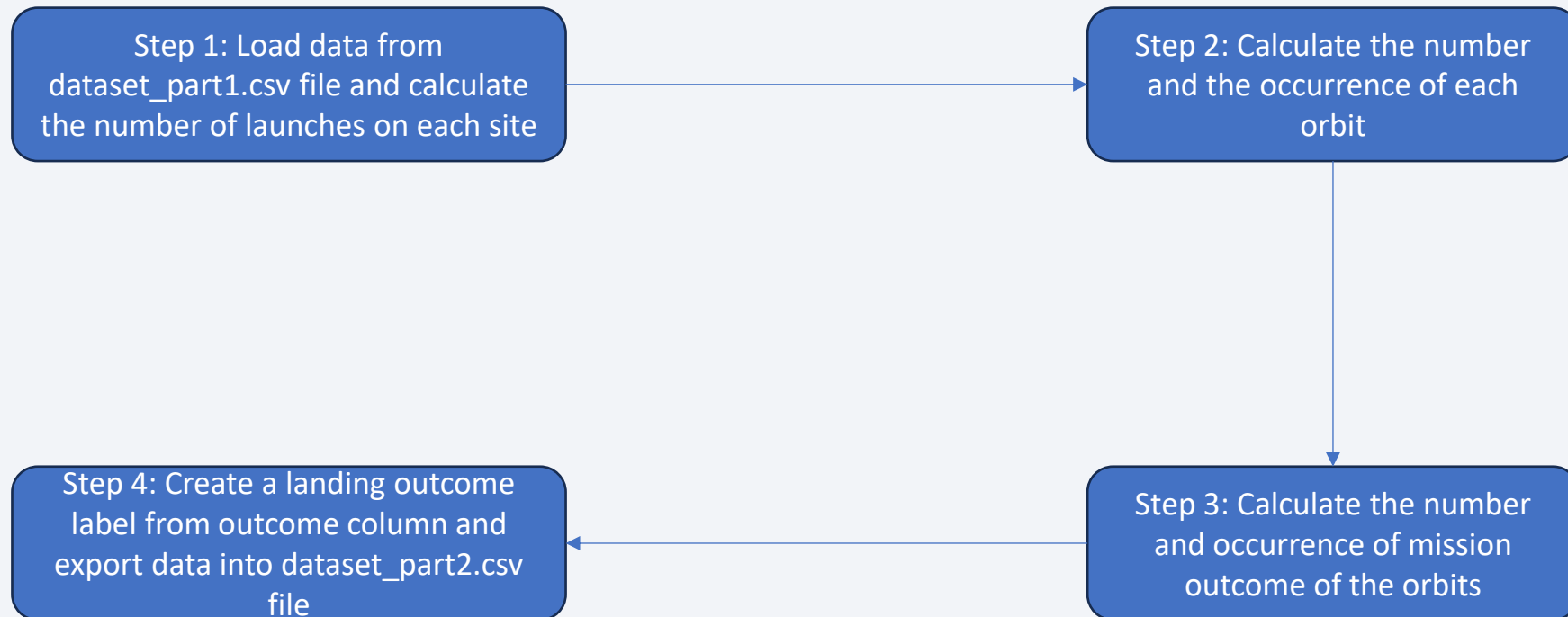
https://github.com/adielgithub123/DataScienceCapstone/blob/main/dataset_part_1.csv

Data Collection - Scraping



https://github.com/adielgithub123/DataScienceCapstone/blob/main/spacex_web_scraped.csv

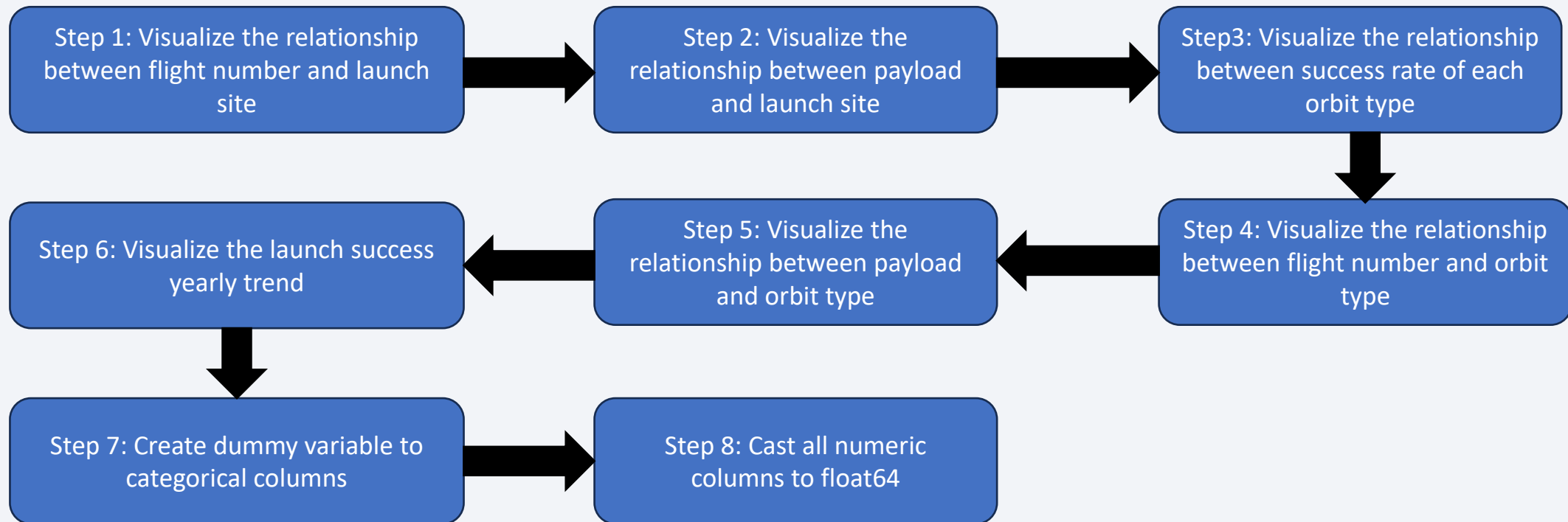
Data Wrangling



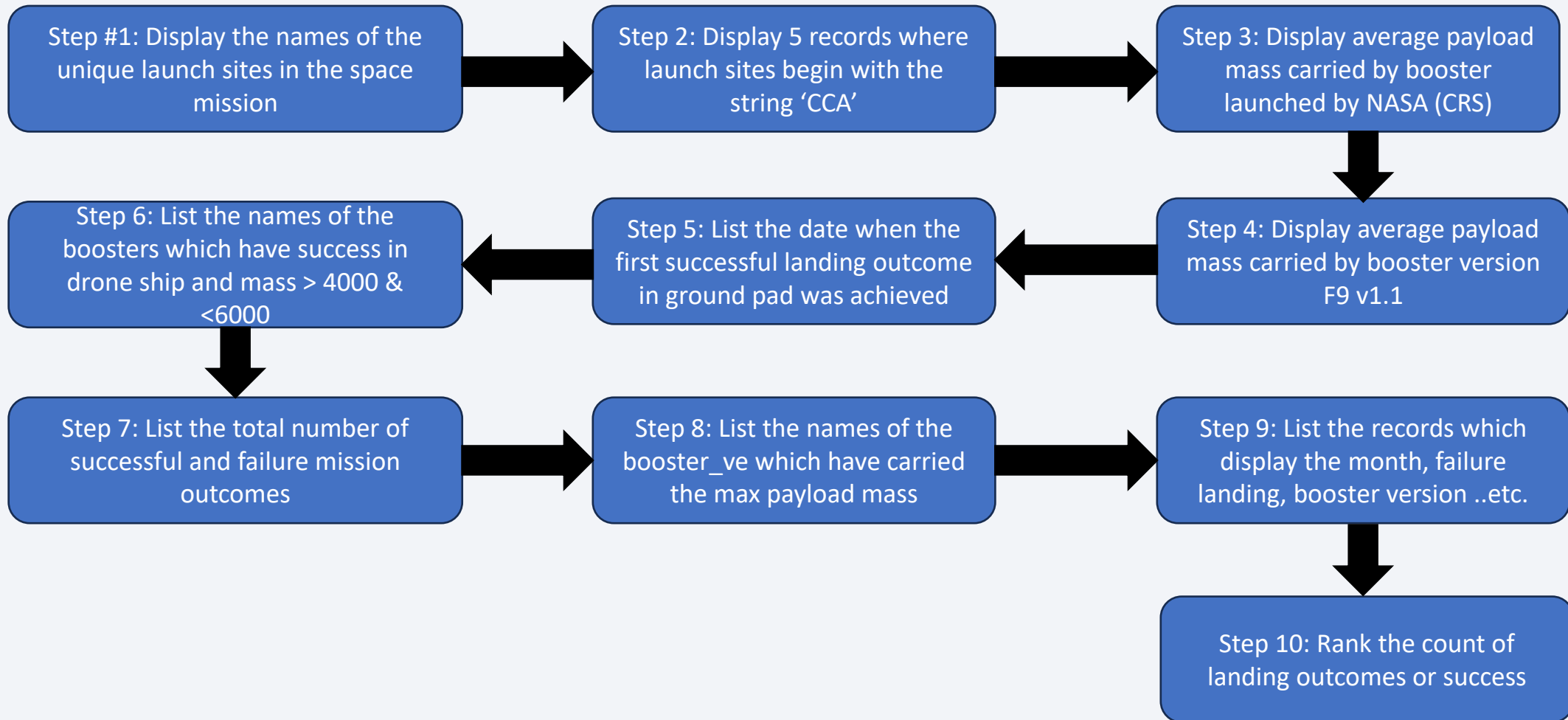
https://github.com/adielgithub123/DataScienceCapstone/blob/main/dataset_part_3.csv

EDA with Data Visualization

Use Matplotlib and Seaborn for data visualization



EDA with SQL



Build an Interactive Map with Folium

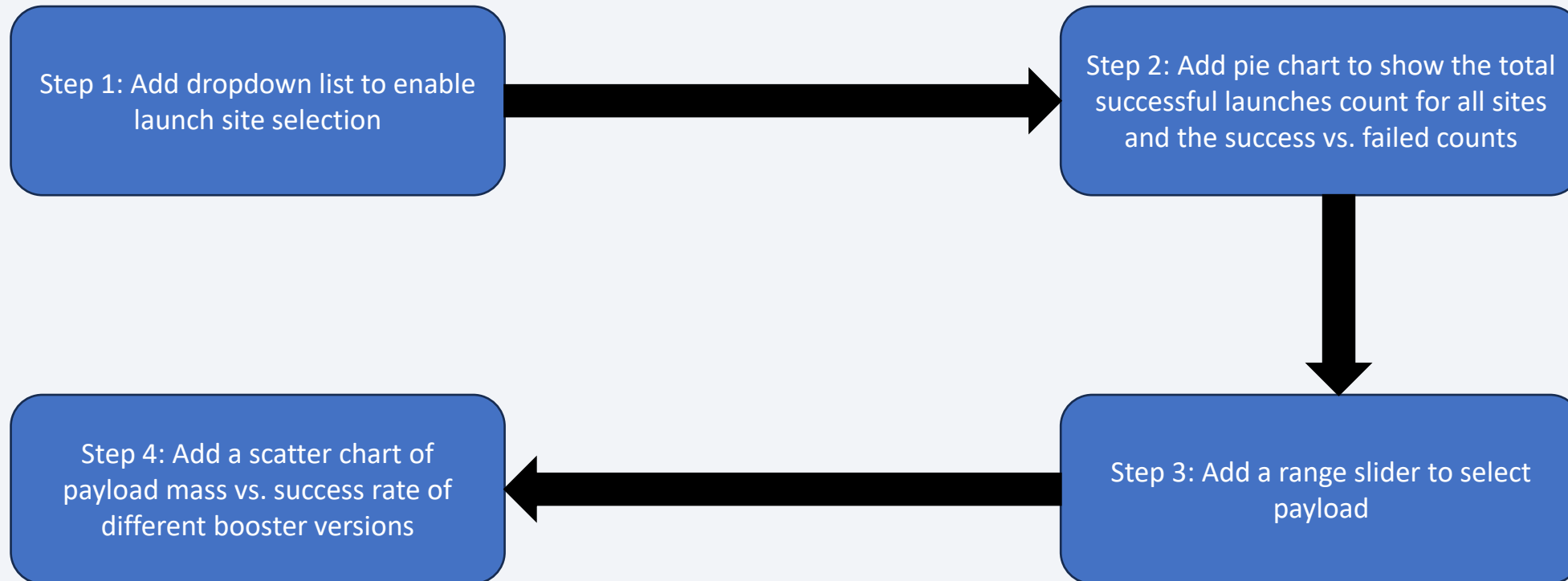


* Explanation:

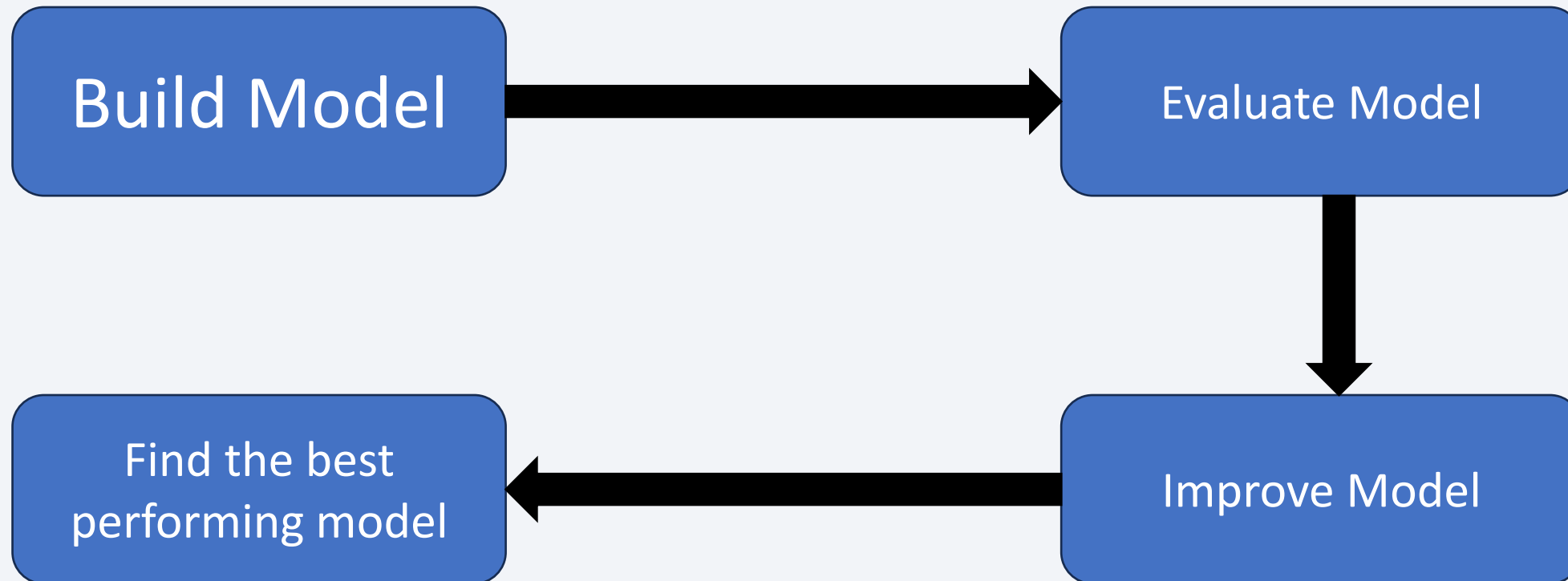
From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:

- relative close to railway (15.23 km)
- relative close to highway (20.28 km)
- relative close to coastline (14.99 km)
- Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km).
- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.

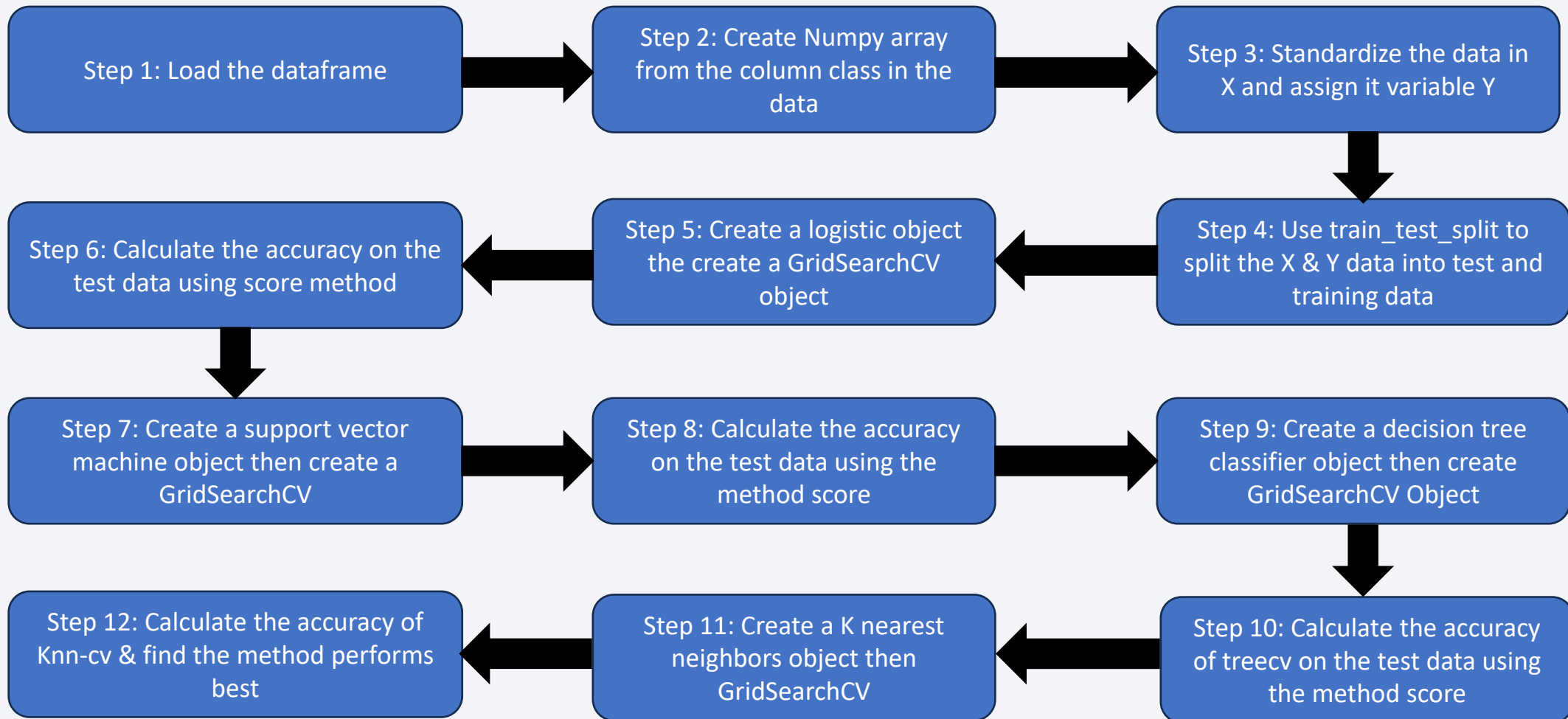
Build a Dashboard with Plotly Dash



Predictive Analysis (Classification) Overview



Predictive Analysis (Classification) Steps



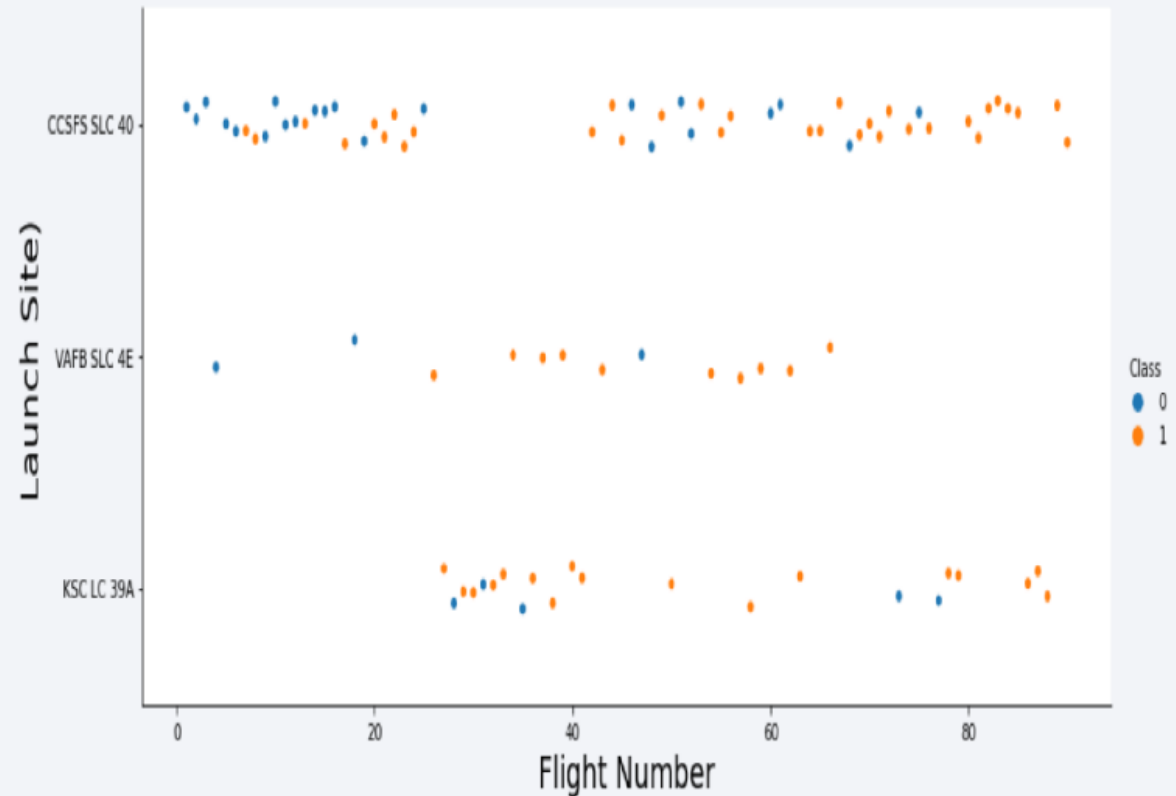
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan, creating a sense of motion and depth. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is high-tech and digital.

Section 2

Insights drawn from EDA

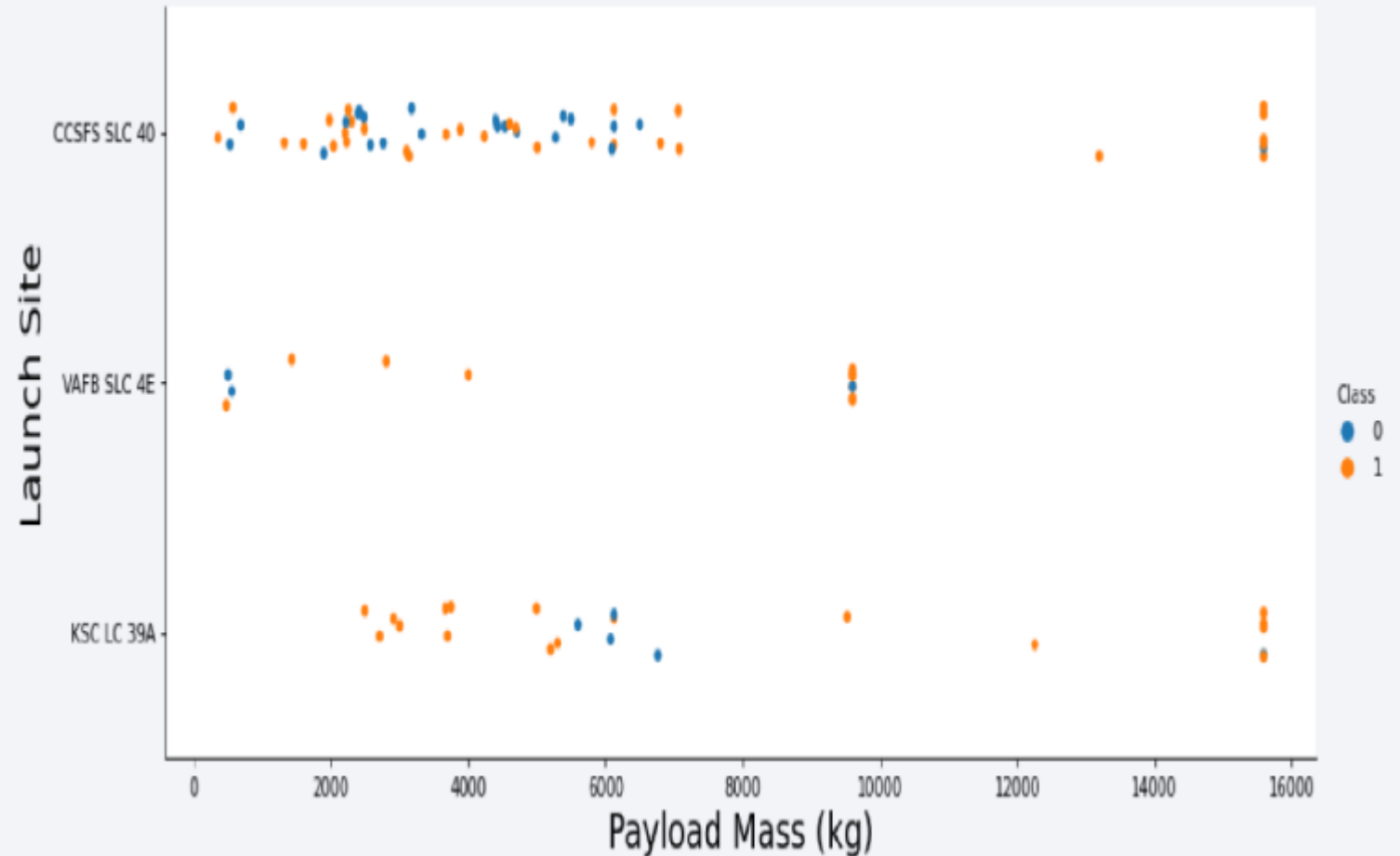
Flight Number vs. Launch Site

- The majority of the flights were launched from the CCAFS SLC 40 sites.
- The VAFB SLC 4E and KSC LC 39A sites have higher success rates than other sites.
- Newer flights have higher success rates than older flights.



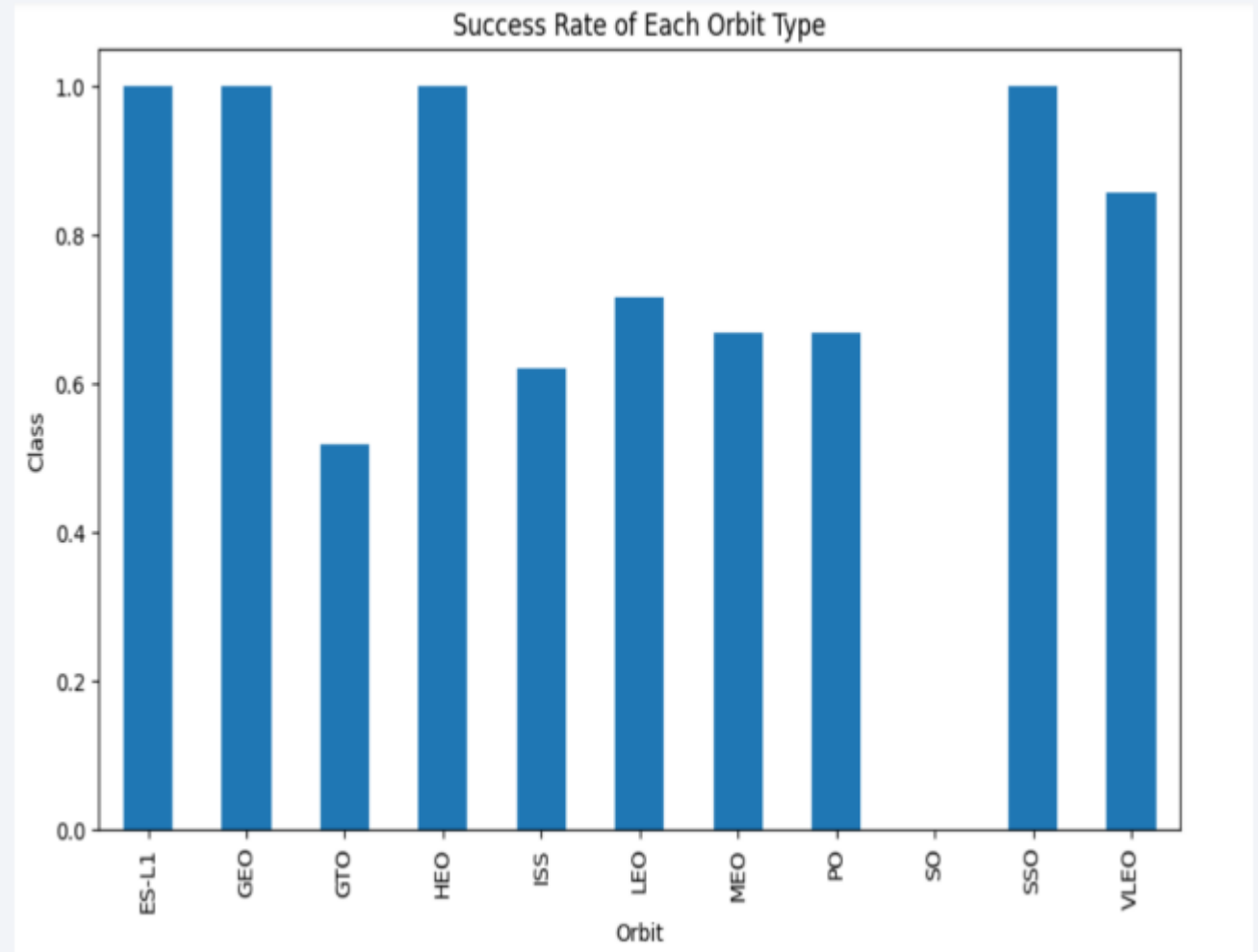
Payload vs. Launch Site

- The majority of the flights with payload mass above 7000 Kg were successful.
- KSC LC 39A success rate for payload mass under 5500 kg is 100%.
- For all launch sites the success rate is proportional to the payload mass.



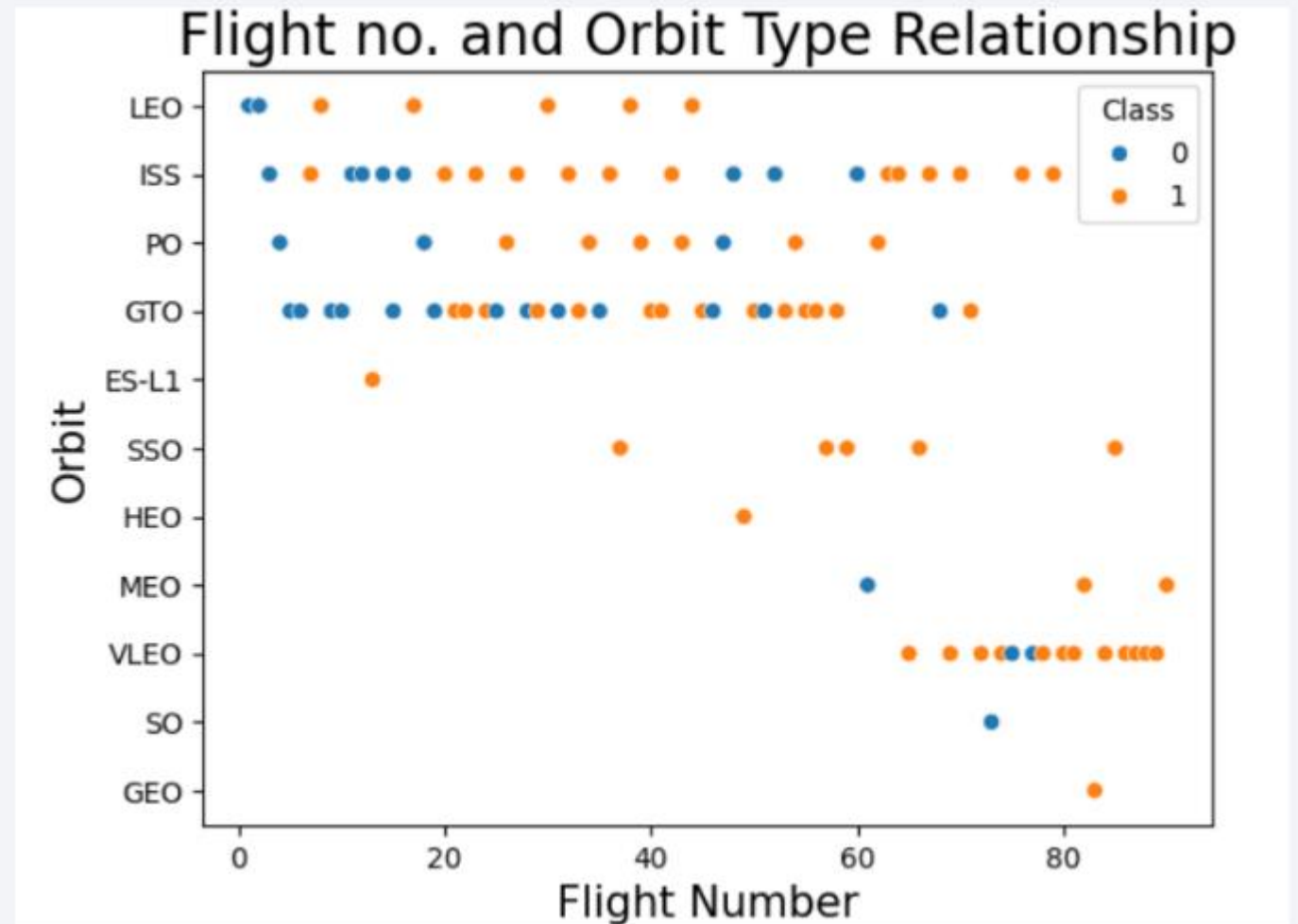
Success Rate vs. Orbit Type

- The OS orbit has 0% success rate.
- The ELS-1, GEO, HEO and SSO orbits have 100% success rate.
- Orbits GTO, ISS, LEO, MEO and PO success rate is higher than 50% and less than 75%.



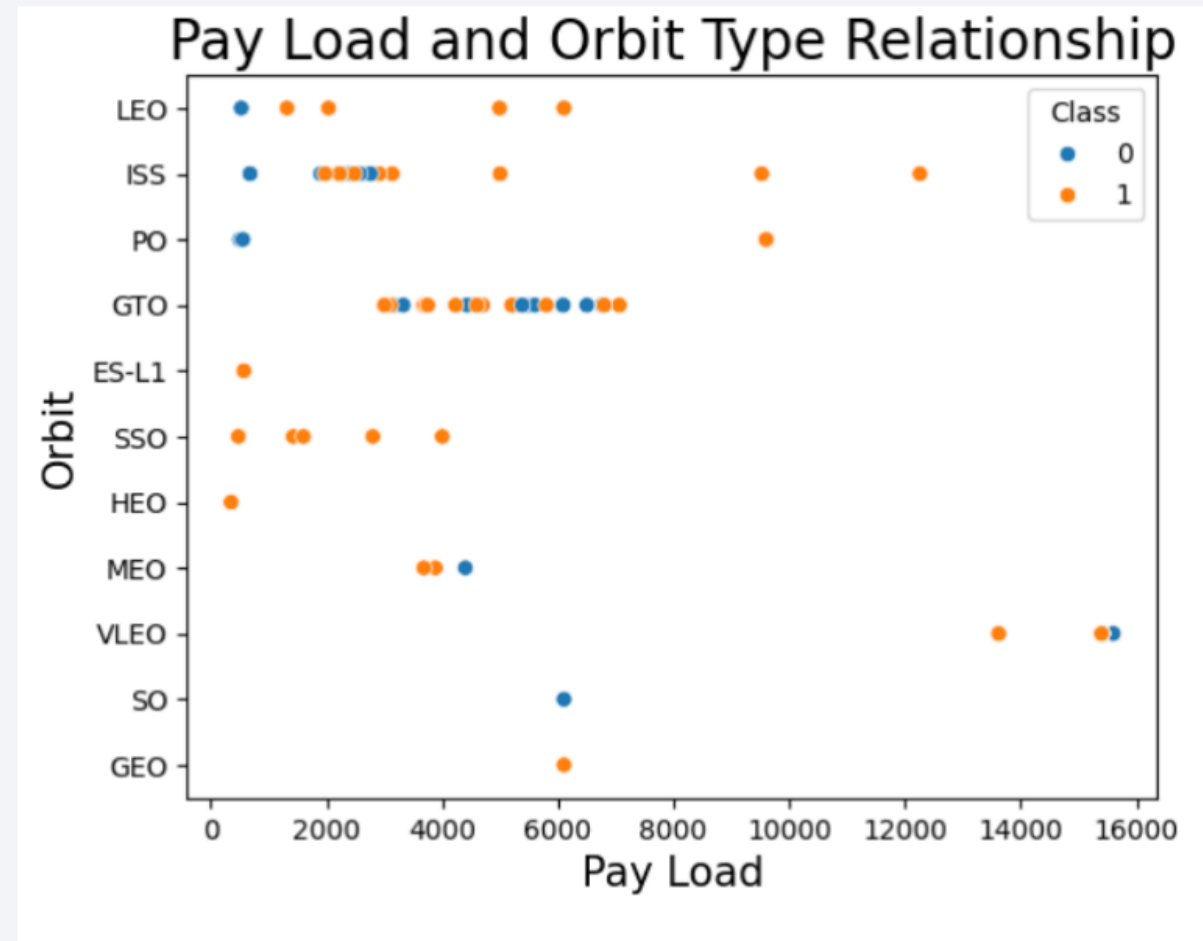
Flight Number vs. Orbit Type

- The majority of the flights were launches to the ISS and GTO orbits.
- The data suggests that there is no relationship between the flight number and the orbit type.



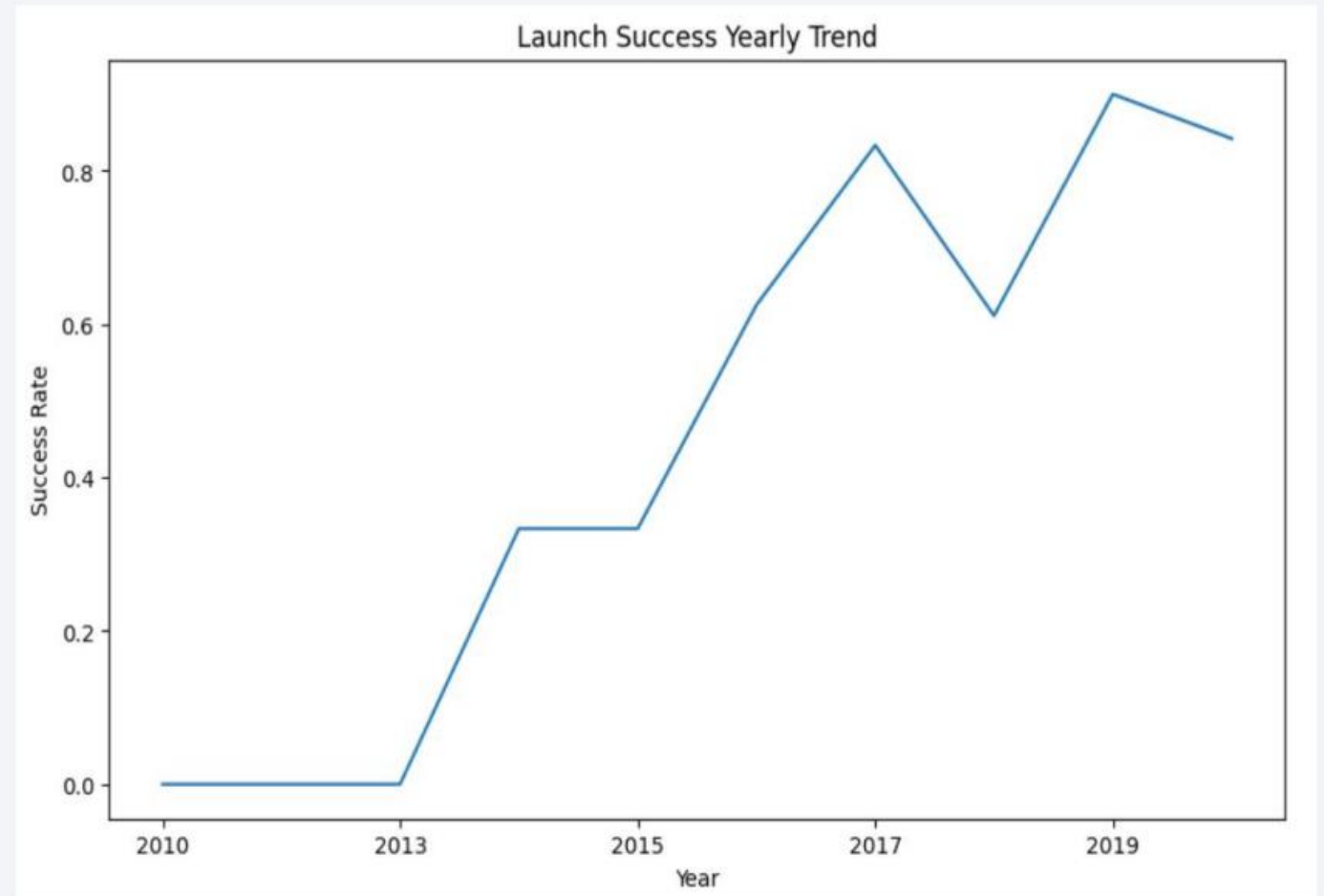
Payload vs. Orbit Type

- Payload masses above 10000 Kg were placed in PO, ISS and LEO orbits.
- Payload masses above 4000 and less than 8000 Kg were placed in the GTO orbit.



Launch Success Yearly Trend

- The launches success rate increased steadily since 2013.
- The increase in the success rate between 2013 and 2017 was linear.
- During 2018 there was a drop in the launches success rate.



All Launch Site Names

The names of the unique launch sites and the query structure for obtaining these sites is shown below.

```
1 features= df[['LaunchSite']]
2 features.head()
```

	LaunchSite
0	CCAFS SLC 40
1	CCAFS SLC 40
2	CCAFS SLC 40
3	VAFB SLC 4E
4	CCAFS SLC 40

Launch Site Names Begin with 'CCA'

5 records for launch sites begin with the string 'CCA' and the query used for obtaining the information is shown below.

```
[ ] 1 %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
2010-06-04	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

Total Payload Mass

- The calculated total payload mass carried by boosters from NASA site =45596 Kg.
- The query for obtaining the total payload mass is shown below.

```
%sql select SUM(PAYLOAD_MASS_KG_) from SPACEXTABLE where "Customer" like 'NASA (CRS)%'
```

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

```
Done.
```

```
SUM(PAYLOAD_MASS_KG_)
```

```
48213
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1=2534.7 Kg.
- Furthermore, the query used to calculate the average payload mass carried by booster F9 v1.1 is shown below.

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where "Booster_Version" LIKE 'F9 v1.1%'
```

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

```
Done.
```

```
AVG(PAYLOAD_MASS__KG_)
```

```
2534.6666666666665
```

First Successful Ground Landing Date

- The first successful landing outcome on a ground pad was in 2015-12-22.
- The query for obtaining this result is shown below.

```
[ ] 1 %sql SELECT MIN(Date) AS first_successful_landing_date FROM SPACEXTABLE WHERE landing_outcome = 'Success (ground pad)';
```

```
↗ * sqlite:///my_data1.db  
Done.  
first_successful_landing_date  
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 is shown below.
- The query used in obtaining this information is shown below.

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[ ] 1 %sql SELECT Booster_Version FROM SPACEXTABLE 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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failed missions is as follows:
- Failure (in flight)= 1 • Successful number of flights= 98 • The query result is shown below.

```
[ ] 1 %sql SELECT mission_outcome, COUNT(*) AS total_count FROM SPACEXTABLE WHERE mission_outcome IN ('Success', 'Failure (in flight)') GROUP BY mis
    2
```



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

```
Done.
```

```
Mission_Outcome total_count
```

```
Failure (in flight) 1
```

```
Success 98
```

Boosters Carried Maximum Payload

- List of the boosters which have carried the maximum payload mass are shown below.
- The query used in obtaining the booster names is shown below

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

```
[ ] 1 %sql SELECT booster_version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX (PAYLOAD_MASS__KG_) FROM SPACEXTABLE);  
2
```

```
↳ * 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

2015 Launch Records

- List of the failed "landing_outcomes" in drone ship, their booster version, and the launch site name during year 2015 is shown below.
- The query used in obtaining the information is shown below

```
[ ] 1 %sql SELECT CASE WHEN substr(Date, 6, 2) = '01' THEN 'January' WHEN substr(Date, 6, 2) = '02' THEN 'February' WHEN substr(Date, 6, 2) = '03' TH  
2
```



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

```
Done.
```

month_name	Booster_Version	Launch_Site	Landing_Outcome
------------	-----------------	-------------	-----------------

January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
---------	---------------	-------------	----------------------

April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
-------	---------------	-------------	----------------------

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

- A rank of the count of landing outcomes (such as Failure (drone ship) or success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order is shown below.
- The query used to obtain the results is shown below.

```
[ ] 1 %sql SELECT landing_outcome, COUNT(*) AS outcome_count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing_outcome
```

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

Done.

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

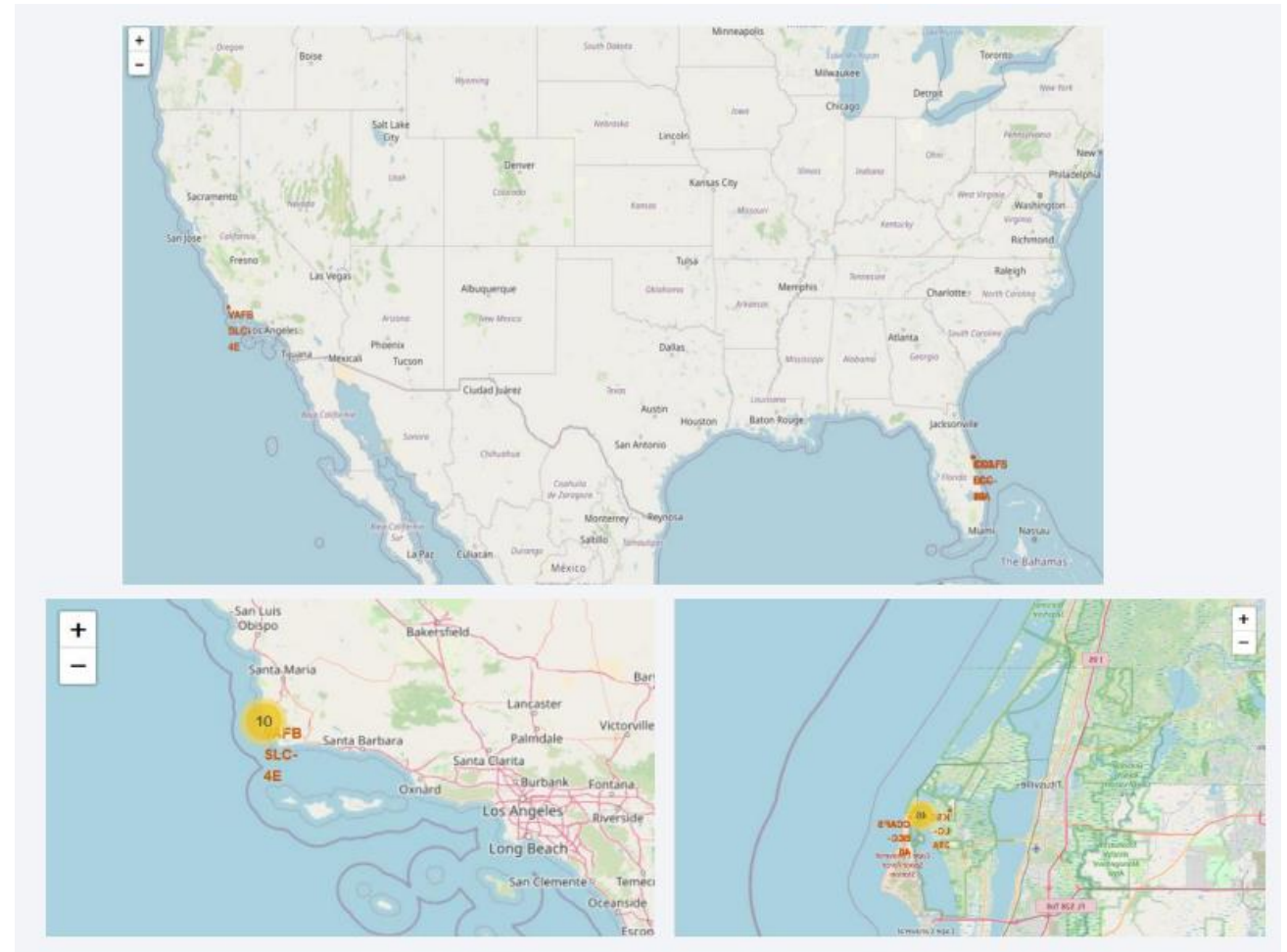
A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

Launch Sites Proximities Analysis

USA Launch Sites in California and Florida

- Most of Launch sites considered in this project are in proximity to the Equator line. Launch sites are made at the closest point possible to Equator line, because anything on the surface of the Earth at the equator is already moving at the maximum speed (1670 kilometers per hour). For example launching from the equator makes the spacecraft move almost 500 km/hour faster once it is launched compared half way to north pole.
- All launch sites considered in this project are in very close proximity to the coast. While starting rockets towards the ocean we minimize the risk of having any debris dropping or exploding near people

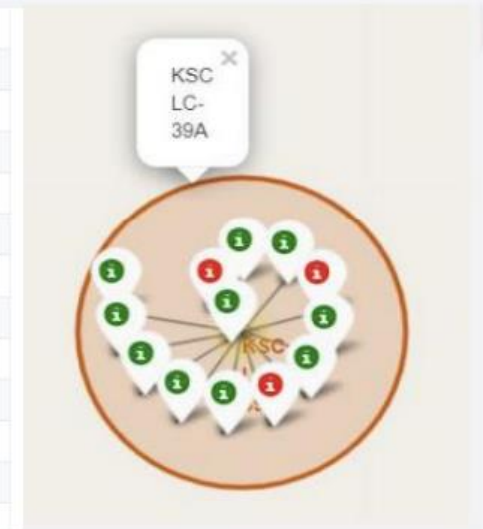


Color Labels Showing the Launch Sites on a Map

	Launch Site	Lat	Long	marker_color
0	CCAFS LC-40	28.562302	-80.577356	red
1	CCAFS LC-40	28.562302	-80.577356	red
2	CCAFS LC-40	28.562302	-80.577356	red
3	CCAFS LC-40	28.562302	-80.577356	red
4	CCAFS LC-40	28.562302	-80.577356	red
5	CCAFS LC-40	28.562302	-80.577356	red
6	CCAFS LC-40	28.562302	-80.577356	red
7	CCAFS LC-40	28.562302	-80.577356	red
8	CCAFS LC-40	28.562302	-80.577356	red
9	CCAFS LC-40	28.562302	-80.577356	red
10	CCAFS LC-40	28.562302	-80.577356	red
11	CCAFS LC-40	28.562302	-80.577356	red
12	CCAFS LC-40	28.562302	-80.577356	red
13	CCAFS LC-40	28.562302	-80.577356	red
14	CCAFS LC-40	28.562302	-80.577356	red
15	CCAFS LC-40	28.562302	-80.577356	red
16	CCAFS LC-40	28.562302	-80.577356	red
17	CCAFS LC-40	28.562302	-80.577356	green
18	CCAFS LC-40	28.562302	-80.577356	green
19	CCAFS LC-40	28.562302	-80.577356	red
20	CCAFS LC-40	28.562302	-80.577356	green



36	KSC LC-39A	28.573255	-80.646895	green
37	KSC LC-39A	28.573255	-80.646895	red
38	KSC LC-39A	28.573255	-80.646895	green
39	KSC LC-39A	28.573255	-80.646895	green
40	KSC LC-39A	28.573255	-80.646895	red
41	KSC LC-39A	28.573255	-80.646895	green
42	KSC LC-39A	28.573255	-80.646895	green
43	KSC LC-39A	28.573255	-80.646895	red
44	KSC LC-39A	28.573255	-80.646895	green
45	KSC LC-39A	28.573255	-80.646895	green
46	KSC LC-39A	28.573255	-80.646895	green
47	KSC LC-39A	28.573255	-80.646895	green
48	KSC LC-39A	28.573255	-80.646895	green



Green= Successful Launch
Red= Failed Launch

Safe Distance to Launch Site

- The obtained results indicate that all launch sites are at safe distance from railway lines and cities.





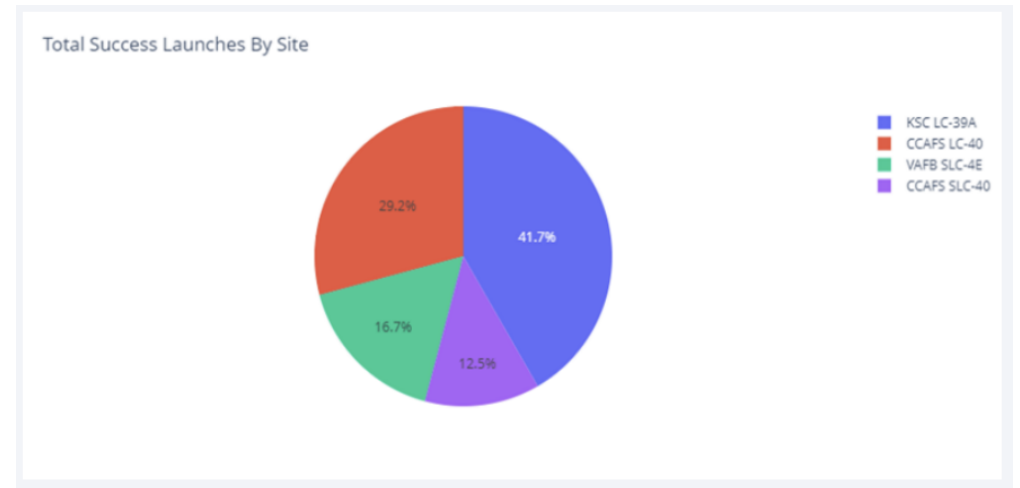
Section 4

Build a Dashboard with Plotly Dash

Total Launch Success for All Sites

The highest success launch rates were recorded at these sites :

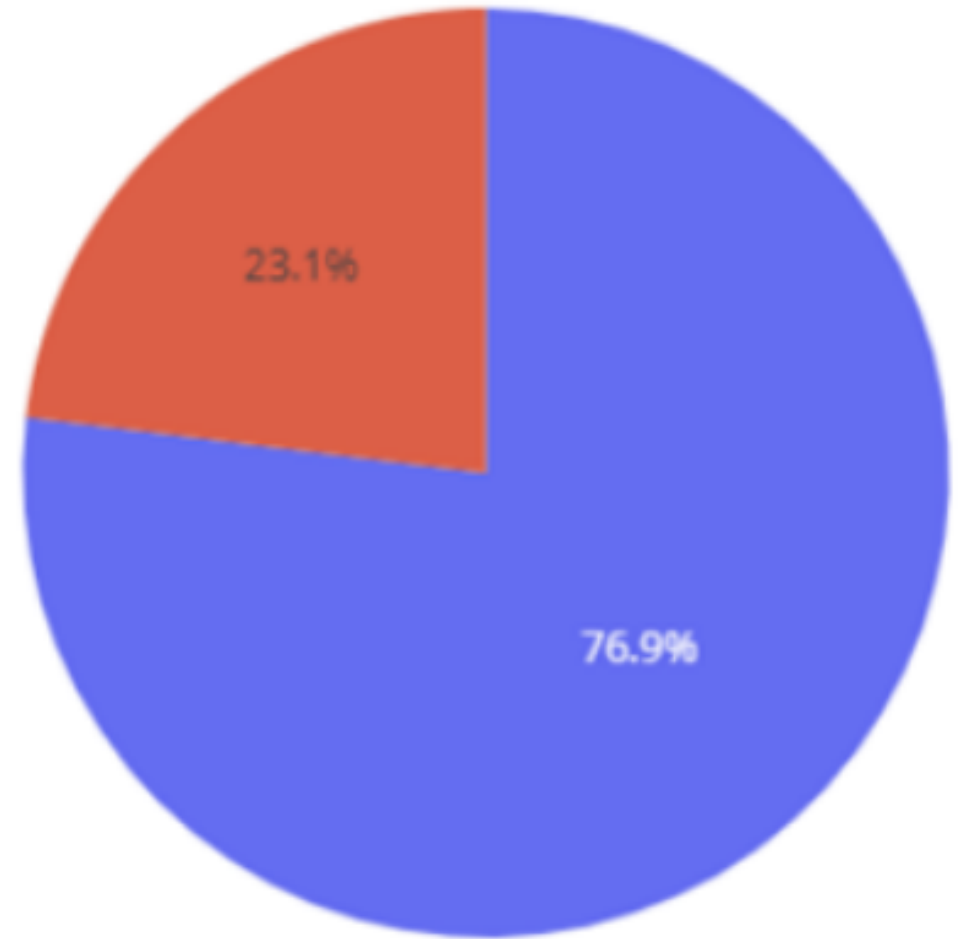
- 1.KSC LC-39A (41.7%)
- 2.CCAFS LC-40 (29.2%)



es for site KSC LC-39A

KSC LC-39 Launch Site Success Rate

Site KSC LC-39 success rate is 76.9%



Payload vs. Launch Outcome for All Sites

Highest success rate for payloads is between 2000 and 5500 Kgs

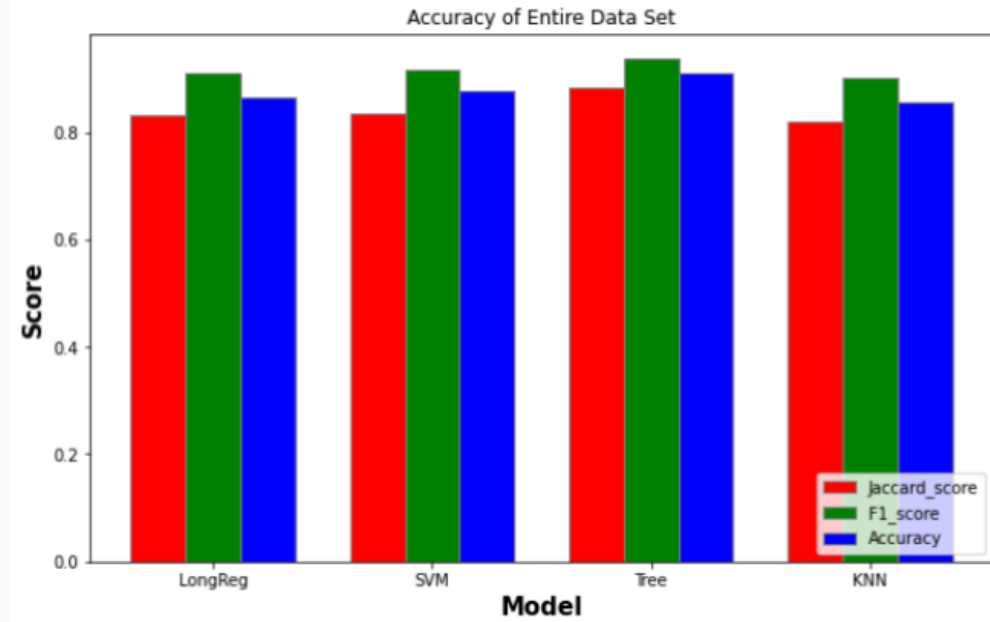
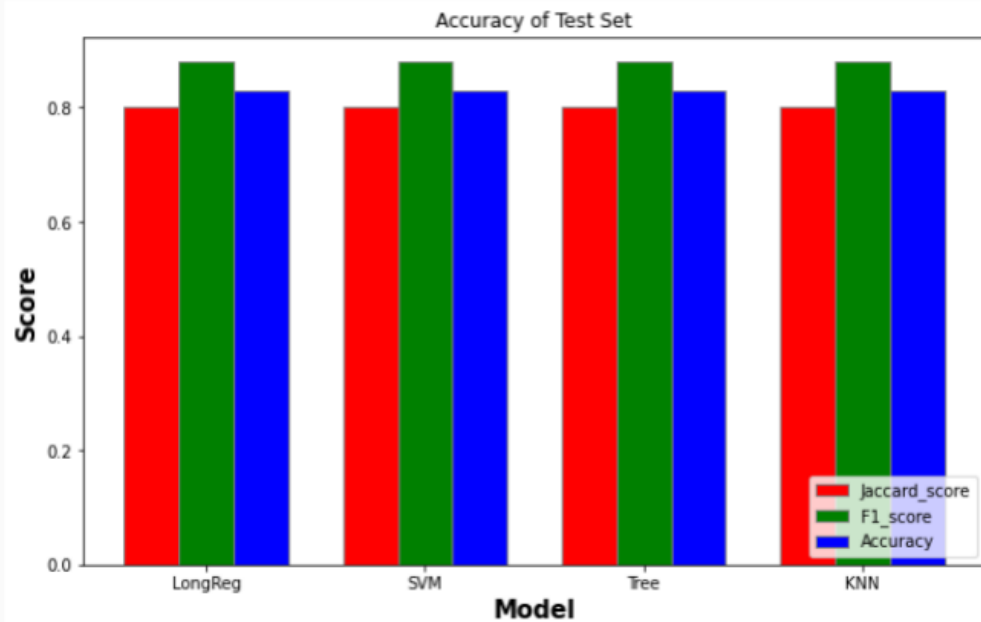


Section 5

Predictive Analysis (Classification)

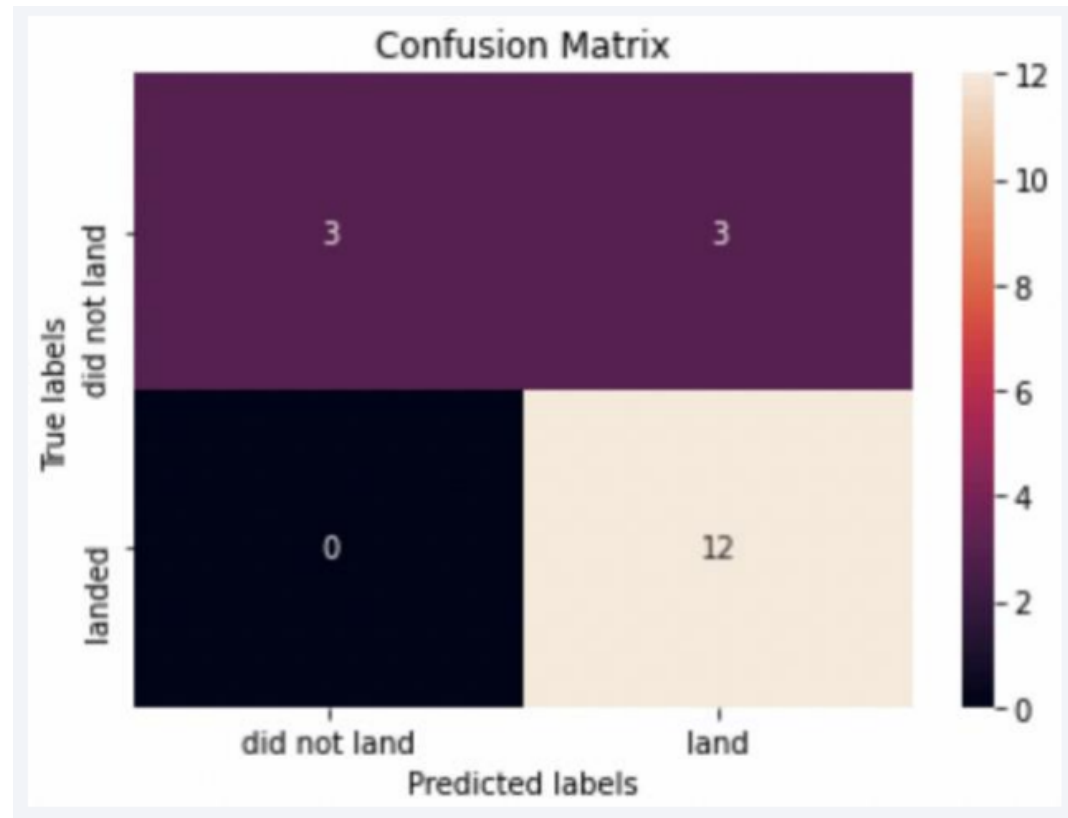
Classification Accuracy

- Using the test set the same accuracy results were obtained from the four models.
- The Tree Model provided the best accuracy results for the entire data set.



Confusion Matrix

- The confusion matrix analysis suggests that the best performing model is the Logistic Regression model.
- The confusion matrix predicts 13 true positives, 3 false positives, 3 true positive, and 0 false negative



Conclusions

- The success rate for the rocket launches increased after 2013.
- Orbits GEO, HEO, ES-L1 and SSO have 100% launch success rate.
- Launch site KSC LC-39A has the highest success rate.
- The Decision Tree model is the best ML algorithm for analyzing the SpaceX data set and provided the best accuracy results.



Appendix

https://www.coursera.org/professional-certificates/ibm-data-science?campaignid=1876641588&adgroupid=70740725700&device=c&keyword=ibm%20data%20science%20professional%20certificate&matchtype=b&network=g&devicemodel=&adposition=&creativeid=347445112274&hide_mobile_promo=&gad_source=1

Thank you!

