IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Khoi Huynh
Date: 16/01/2025

Outline

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

Executive Summary

- Summary of methodologies
  - Data collected via SpaceX API endpoints and web scraping
  - After data wrangling, exploratory data analysis (EDA) is performed using Python visualization libraries and SQL
  - Interactive visual analytics is also performed using Folium and Plotly Dash
- Summary of all results
  - Decision Tree classification model yields the highest accuracy score, however, due to the test data set sample size being relatively small, the current accuracy score might not fully reflect the model's performance
  - Launch site KSC LC 39A is the launch site with the highest launch success rate
  - The following orbits have a 100% launch success so far: ES-L1, GEO, HEO, and SSO
  - The orbit SO has never had a successful launch before
  - Most of the success launches have the payload mass within the range 2000 – 6000 kg
  - Booster version FT has the highest launch success rate

Introduction

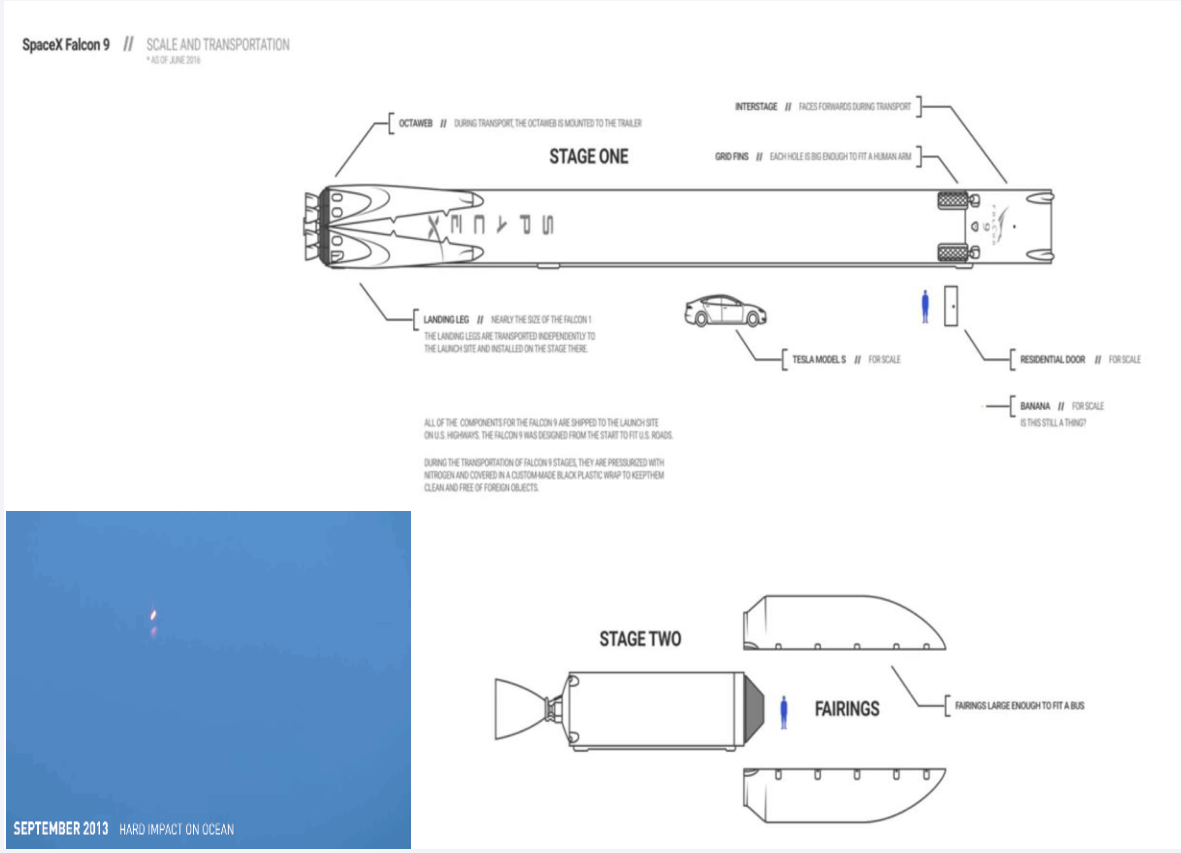
- Project background and context

1. One of SpaceX's competitive advantage is that its **rocket launches** are **relatively inexpensive**.
2. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can **reuse the first stage**.
3. Unlike other rocket providers, SpaceX's Falcon 9 can **recover the first stage**. Sometimes the first stage does not land. Sometimes it will crash as shown in the clip below.

Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- Problems you want to find answers

1. What attributes influence whether or not the Falcon 9 first stage will land successfully?
2. Can we predict whether or not the Falcon 9 first stage will land successfully to a certain level of accuracy?
3. Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?
4. Which payload range(s) has the highest/lowest launch success rate?



Section 1

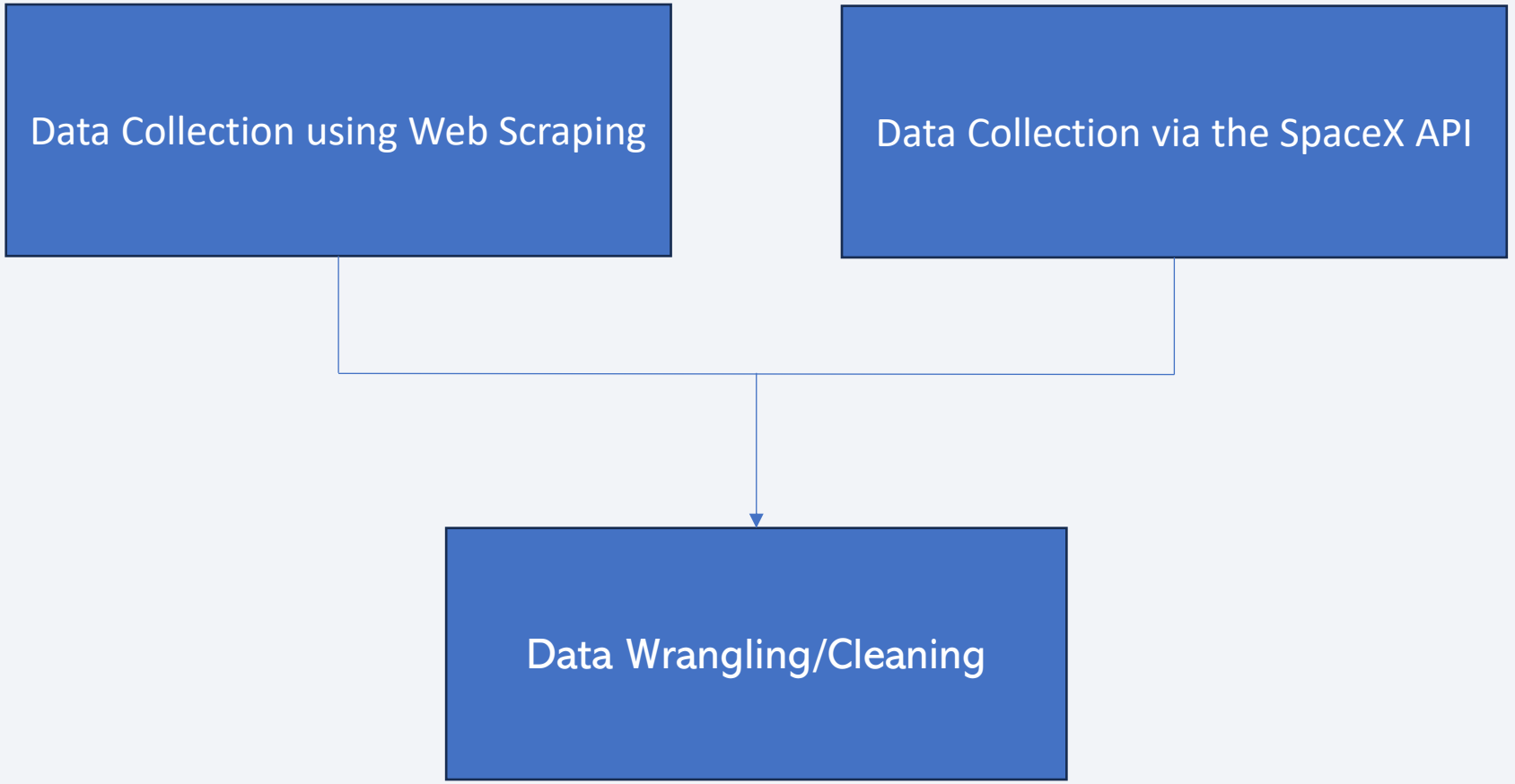
Methodology

Methodology

Executive Summary

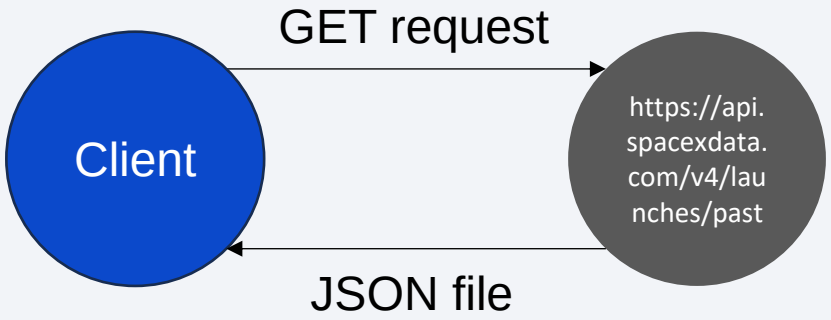
- Data collection methodology:
    1. Data Collection via the SpaceX API
    2. Data Collection using Web Scraping
- Perform data wrangling
    1. Select relevant subset of columns to retain for analysis
    2. Remove rows with multiple cores and payloads
    3. Filter the data frame to only include Falcon 9 launches
    4. Handle missing values by replacing them with the respective mean/mode
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    1. Logistic Regression
    2. Support Vector Machine
    3. Decision Tree
    4. K-Nearest Neighbors

Data Collection

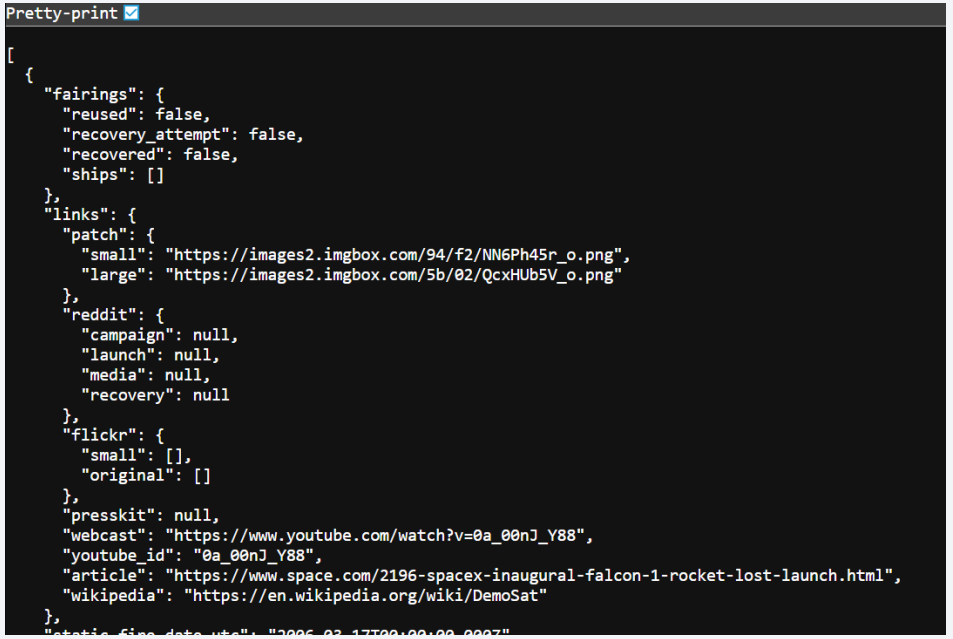


Data Collection – SpaceX API

- Make **HTTP GET requests** to multiple SpaceX **API endpoints** to collect data in the form of JSON files regarding past launches, launch sites, payloads, cores, booster versions, orbits, etc



- GitHub URL of the completed SpaceX API calls notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

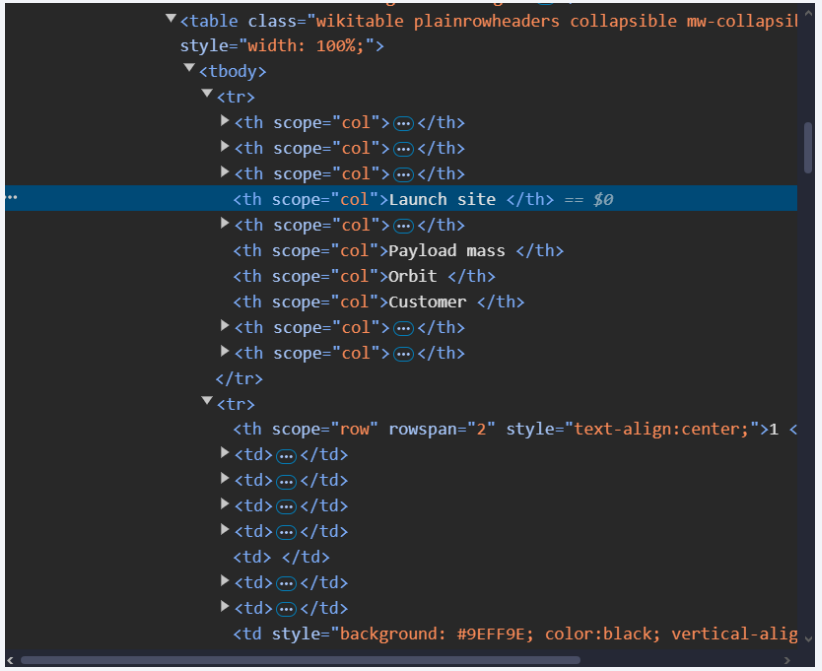
- Web Scraping process:

1. Extract the HTML table using the requests and BeautifulSoup libraries
2. Extract all column/variable names from the HTML table header by locating and iterating through the list of <th> tags
3. Parse the HTML table and convert it into a data frame

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule,[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

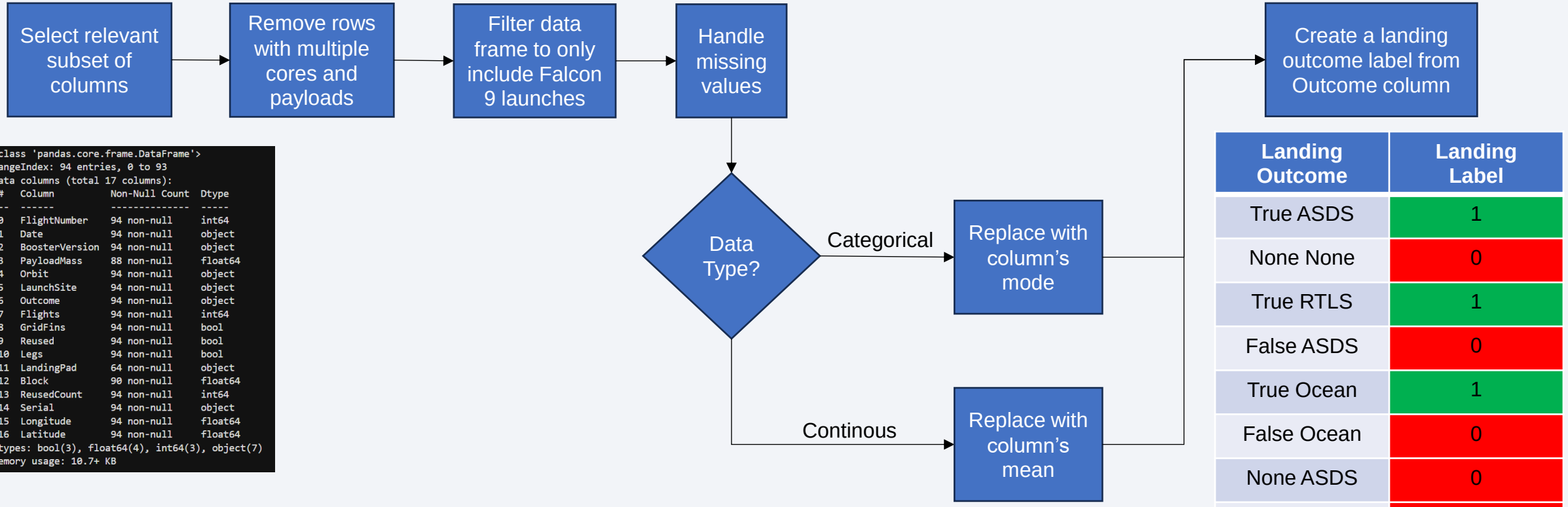


```
<table class="wikitable plainrowheaders collapsible mw-collapsi
style="width: 100%;">
 <tbody>
  <tr>
   <th scope="col">…</th>
   <th scope="col">…</th>
   <th scope="col">…</th>
   <th scope="col">Launch site </th> == $0
   <th scope="col">…</th>
   <th scope="col">Payload mass </th>
   <th scope="col">Orbit </th>
   <th scope="col">Customer </th>
   <th scope="col">…</th>
   <th scope="col">…</th>
  </tr>
  <tr>
   <th scope="row" rowspan="2" style="text-align:center;">1 <
   <td>…</td>
   <td>…</td>
   <td>…</td>
   <td>…</td>
   <td> </td>
   <td>…</td>
   <td>…</td>
   <td style="background: #9EFF9E; color:black; vertical-alig
```

- GitHub URL of the completed web scraping notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Data Processing Stage



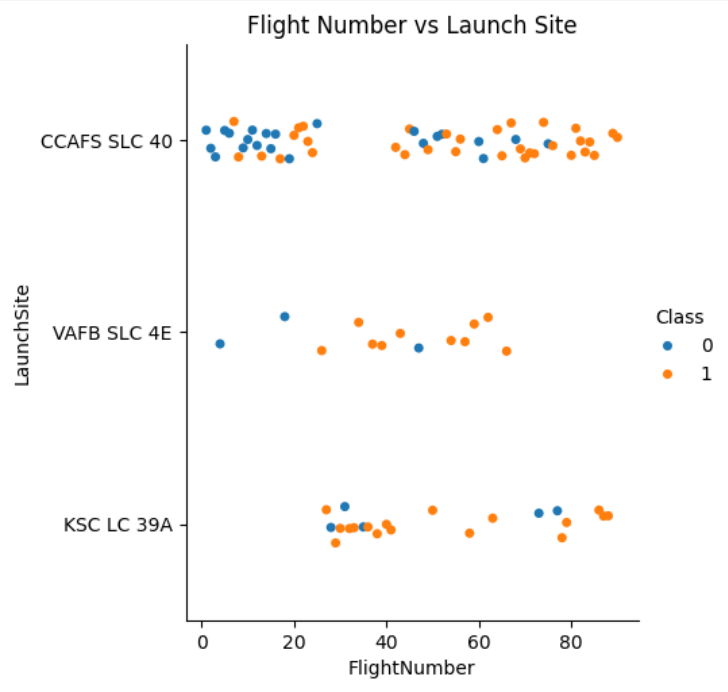

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 94 entries, 0 to 93
Data columns (total 17 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   FlightNumber    94 non-null     int64
 1   Date            94 non-null     object
 2   BoosterVersion  94 non-null     object
 3   PayloadMass     88 non-null     float64
 4   Orbit           94 non-null     object
 5   LaunchSite      94 non-null     object
 6   Outcome         94 non-null     object
 7   Flights         94 non-null     int64
 8   GridFins        94 non-null     bool
 9   Reused          94 non-null     bool
 10  Legs            94 non-null     bool
 11  LandingPad      64 non-null     object
 12  Block           90 non-null     float64
 13  ReusedCount     94 non-null     int64
 14  Serial          94 non-null     object
 15  Longitude       94 non-null     float64
 16  Latitude        94 non-null     float64
dtypes: bool(3), float64(4), int64(3), object(7)
memory usage: 10.7+ KB
```

Landing Outcome	Landing Label
True ASDS	1
None None	0
True RTLS	1
False ASDS	0
True Ocean	1
False Ocean	0
None ASDS	0
False RTLS	0

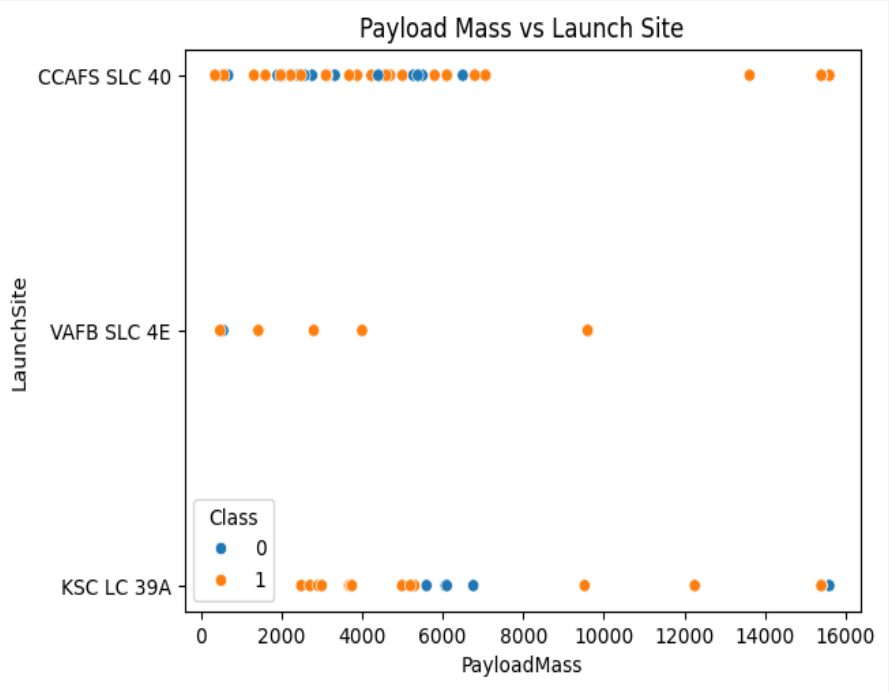
- GitHub URL of the completed data wrangling notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Charts plotted:



To understand which launch site has the highest launch success rate

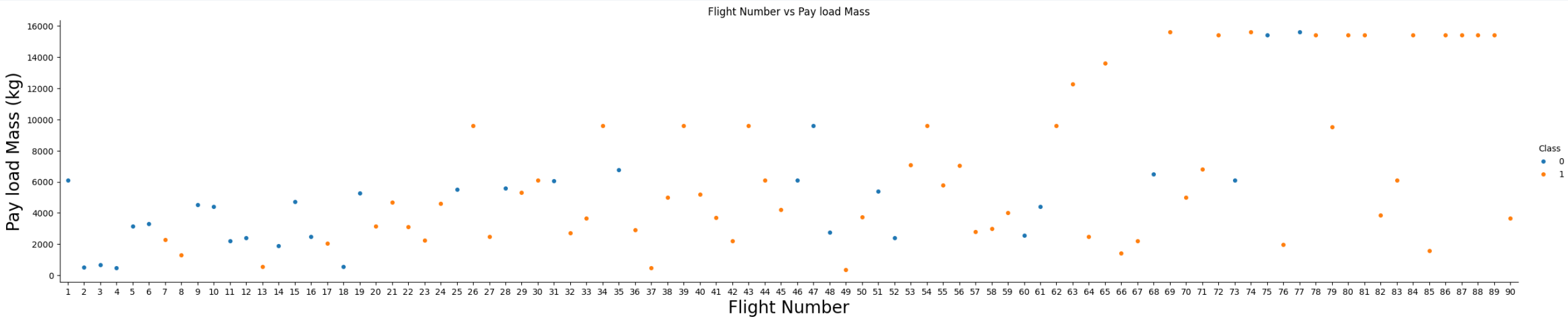


To understand what payload mass ranges yields success for each launch site

- GitHub URL of the completed EDA with data visualization notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with Data Visualization

- Charts plotted:

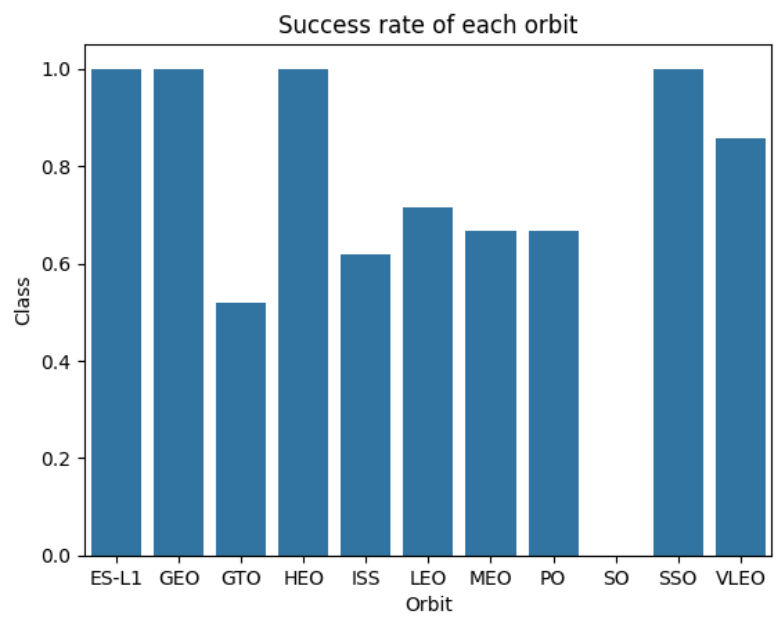


To understand the relationship between flight number and the payload mass

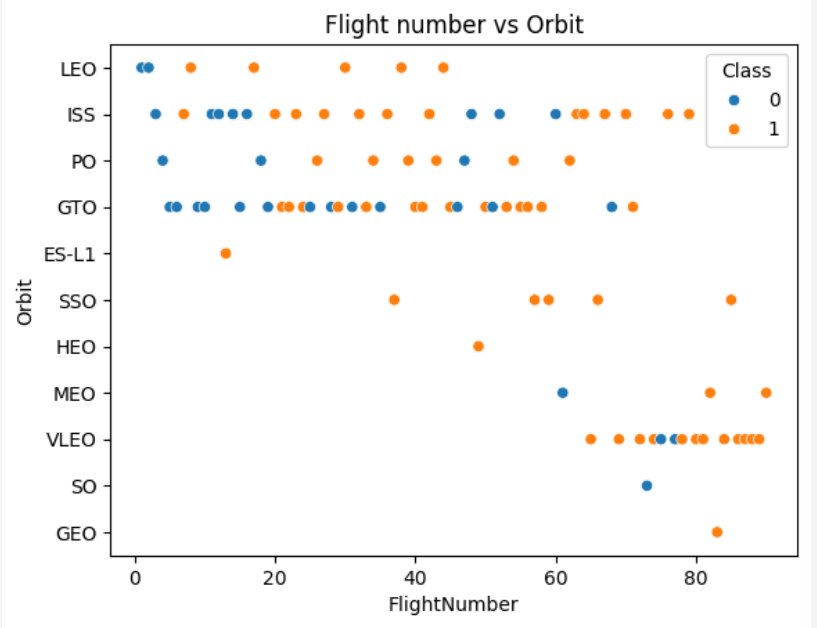
- GitHub URL of the completed EDA with data visualization notebook:
  https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with Data Visualization

- Charts plotted:



To understand which orbit has the highest launch success rate

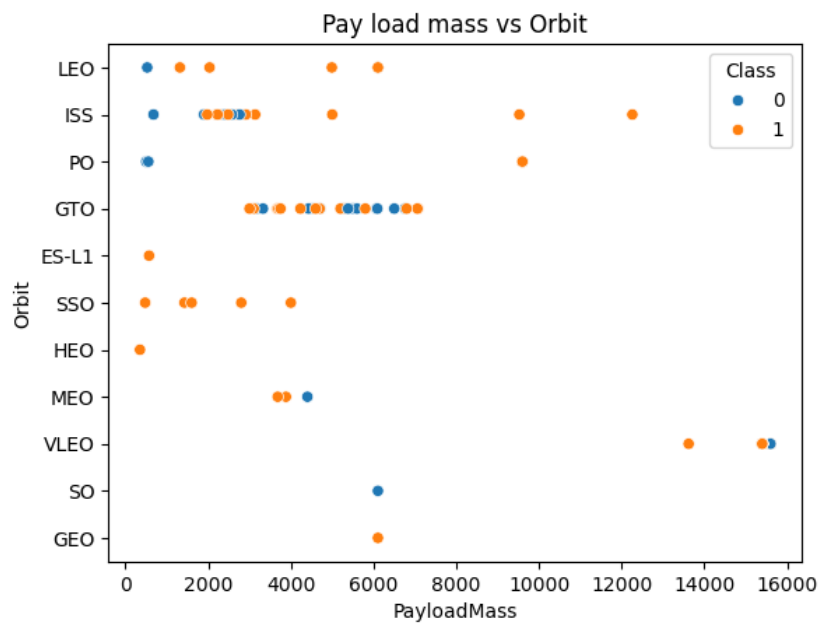


To understand the relationship between flight number and orbit

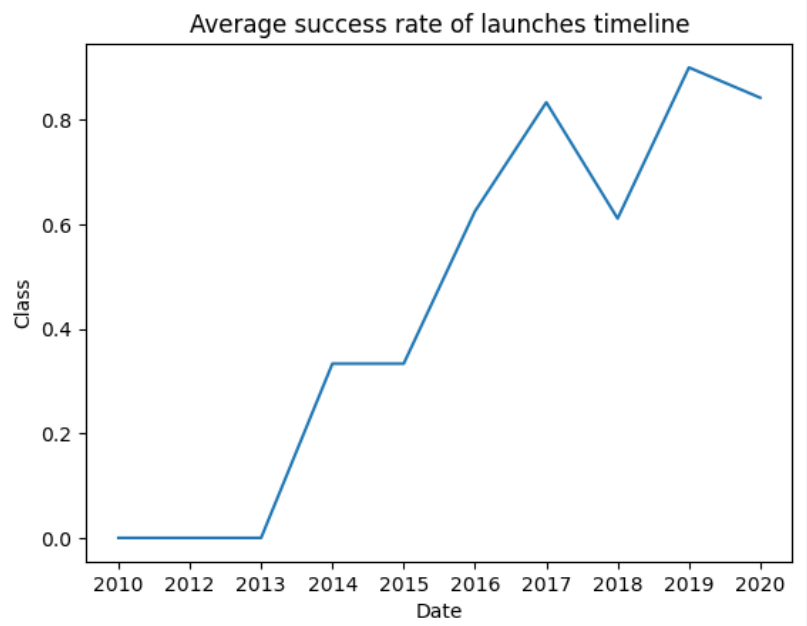
- GitHub URL of the completed EDA with data visualization notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with Data Visualization

- Charts plotted:



To understand what payload mass ranges yields success for each orbit



To understand the trend of the average success rate of launches

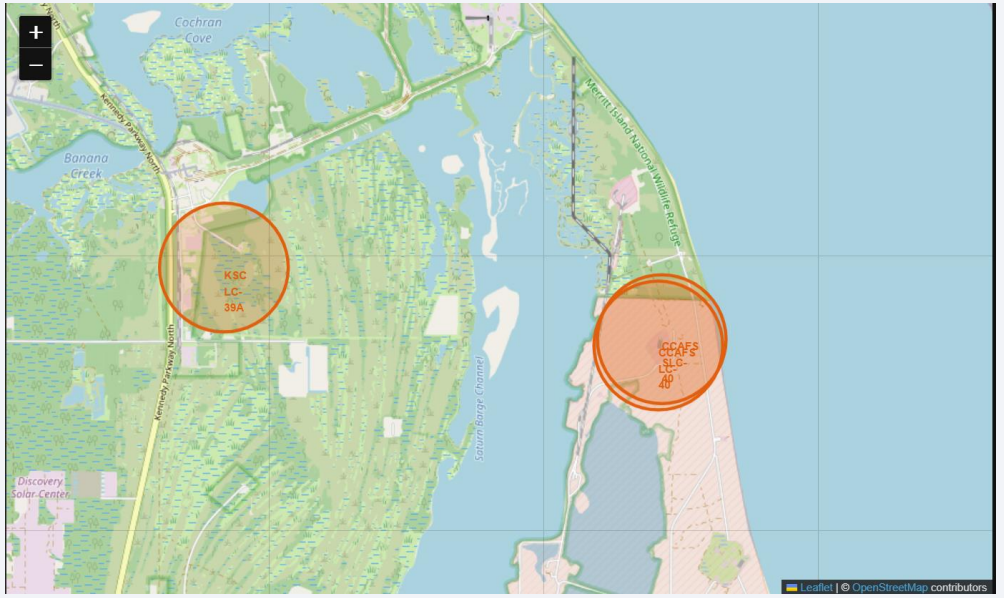
- GitHub URL of the completed EDA with data visualization notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- SQL queries performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - 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.
- GitHub URL of the completed EDA with SQL notebook: https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

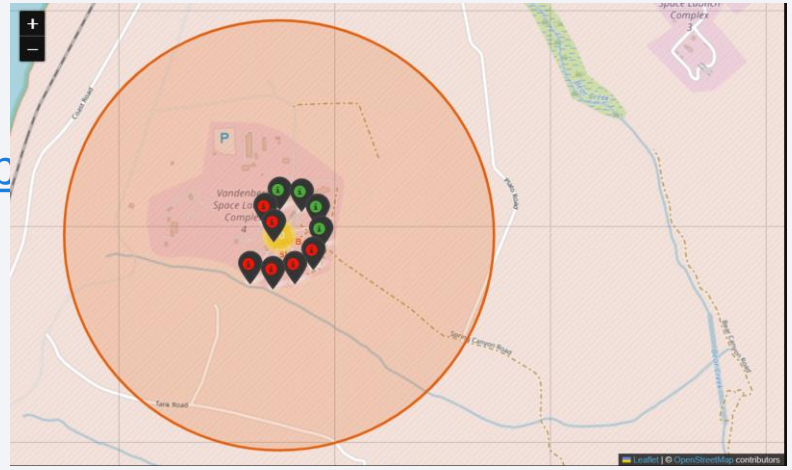
Build an Interactive Map with Folium

- Map objects added:
  - MarkerCluster: To group markers based on similiar coordinates.
  - Marker: To label the name of the launch sites and each rocket launches
  - Circle: To show the general area within a certain radius of the launch sites
  - Lines: To annotate nearby cities, railway, highway along with the distance from the launch sites



- GitHub URL of the completed interactive map with Folium map:

  https://github.com/Khoi-Huynh13/Applied-Data-Science-Cap/blob/main/lab_jupyter_launch_site_location.ipynb

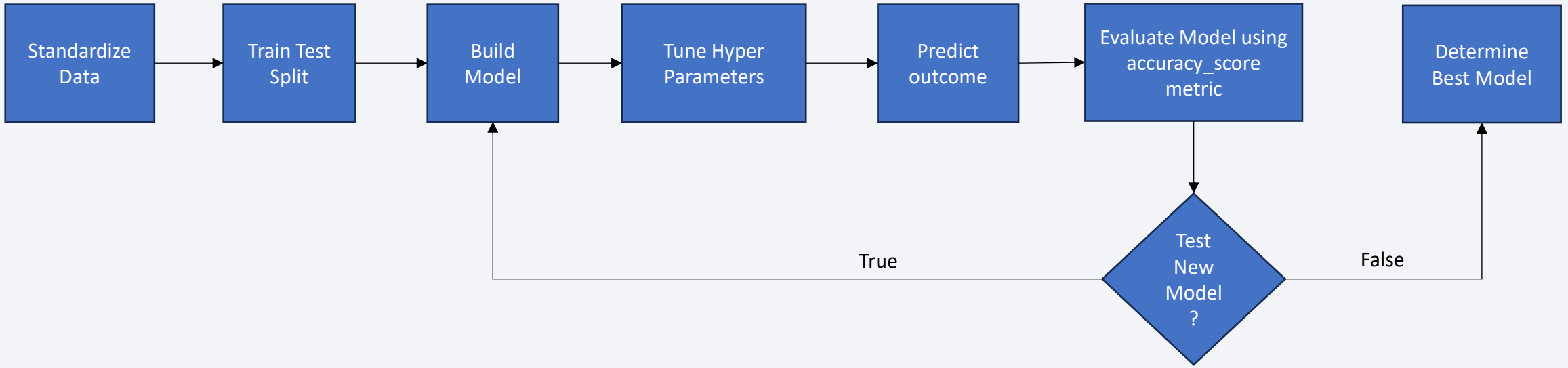


Build a Dashboard with Plotly Dash

- Plots/graphs and interactions added to dashboard:
  - Add a Launch Site Drop-down Input Component, so the user can drill down to each of the launch sites
  - Display a pie chart detailing the launch success count for each launch sites if the user selects the “All Sites” option, else display the proportion of successful and unsuccessful launches for the user selected site
  - Add a Range Slider to Select Payload, so the user can drill down and focus on the range of payload of interest
  - Display a scatter plot detailing the correlation between the payload mass and success based on the user’s selected option in the launch site drop-down component
- GitHub URL of the completed Plotly Dash lab:
  https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

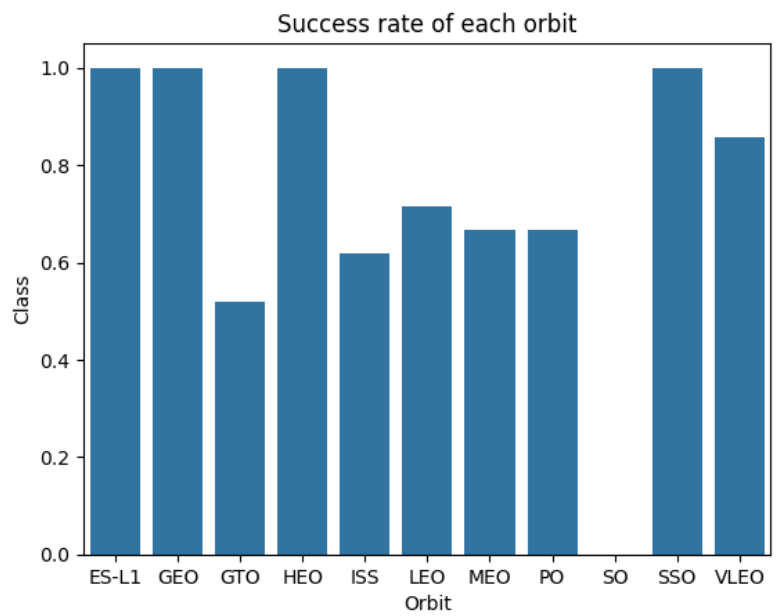
Predictive Analysis (Classification)

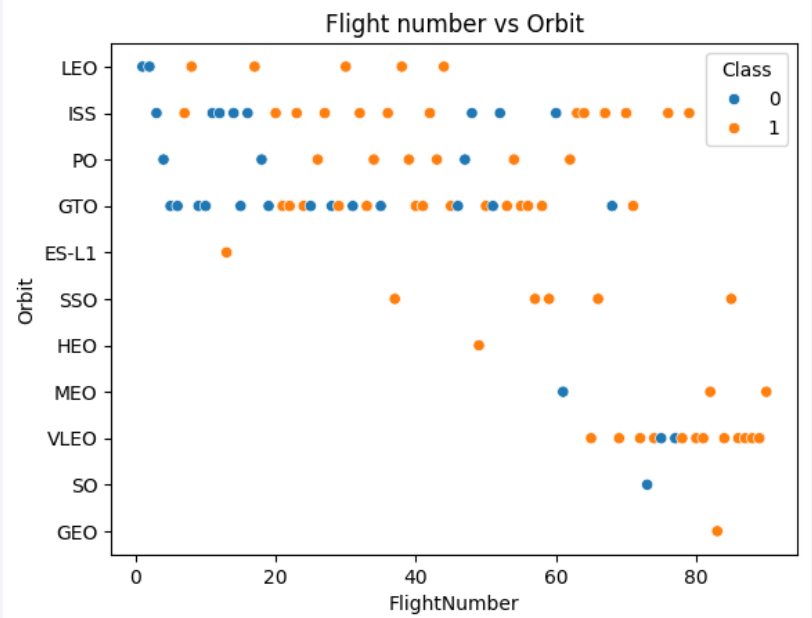
- GitHub URL of the completed predictive analysis lab:
  https://github.com/Khoi-Huynh13/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

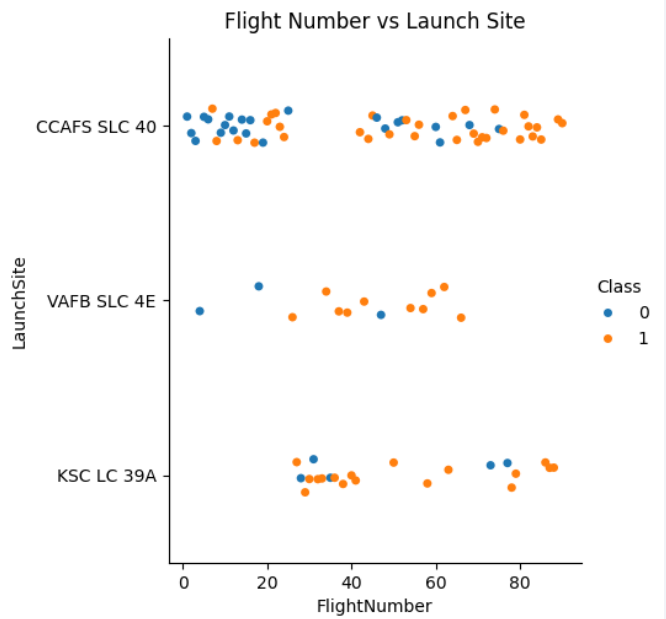


Results

Exploratory data analysis results

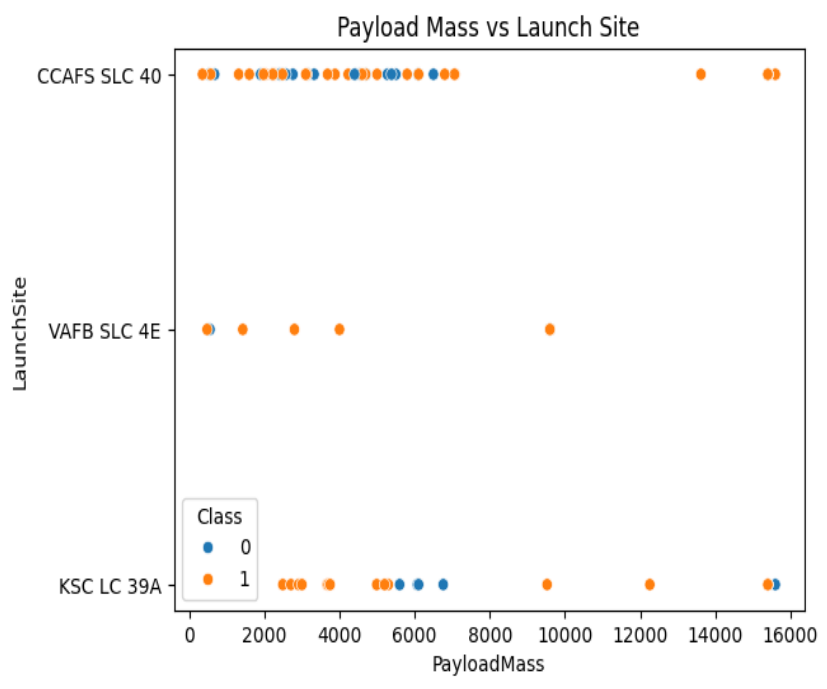


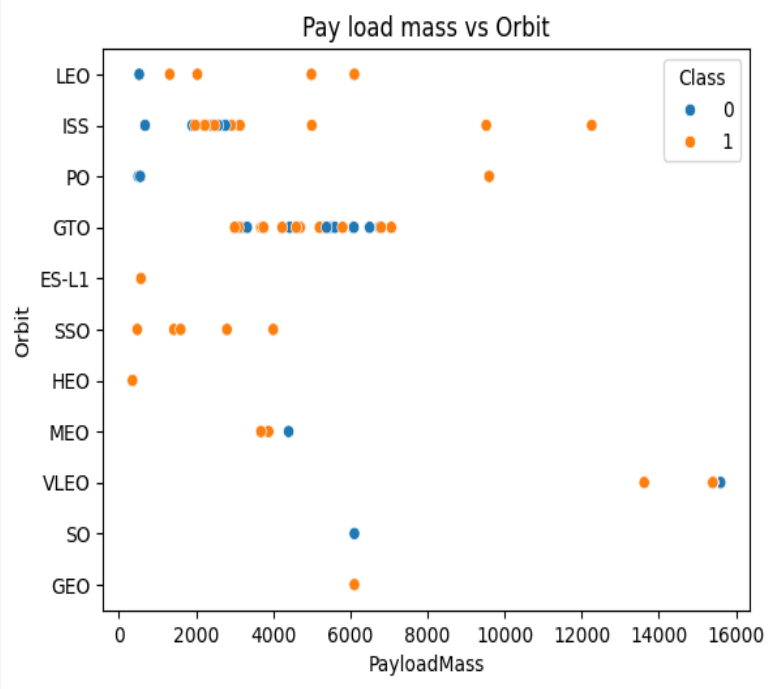


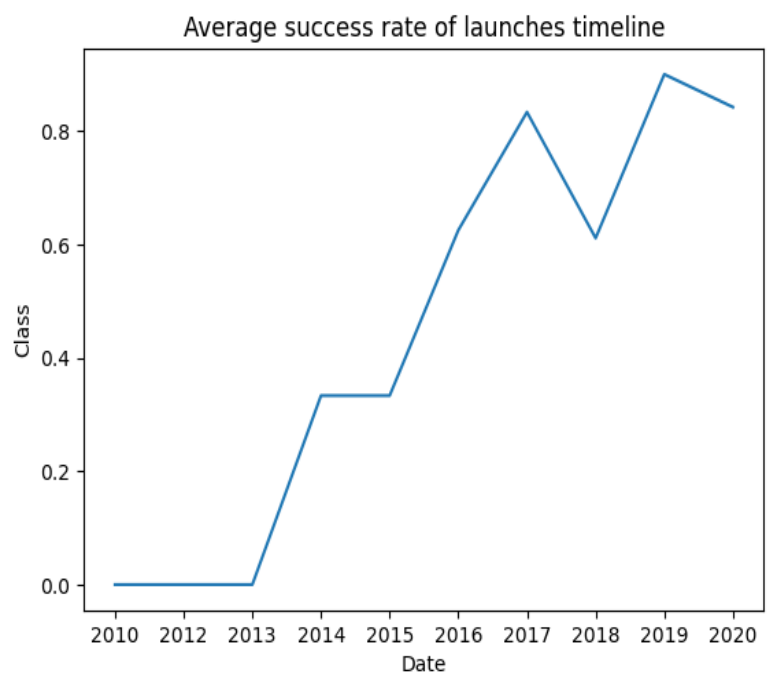


Results

Exploratory data analysis results

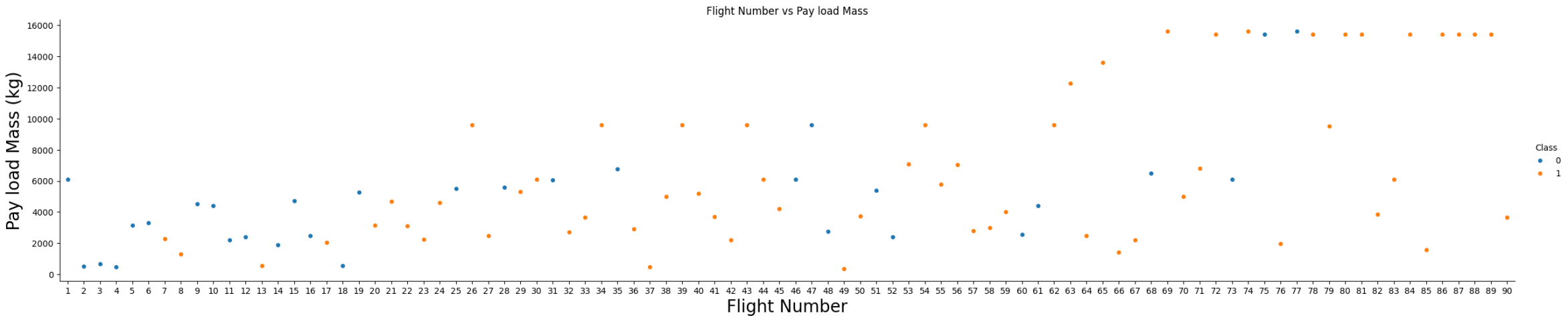






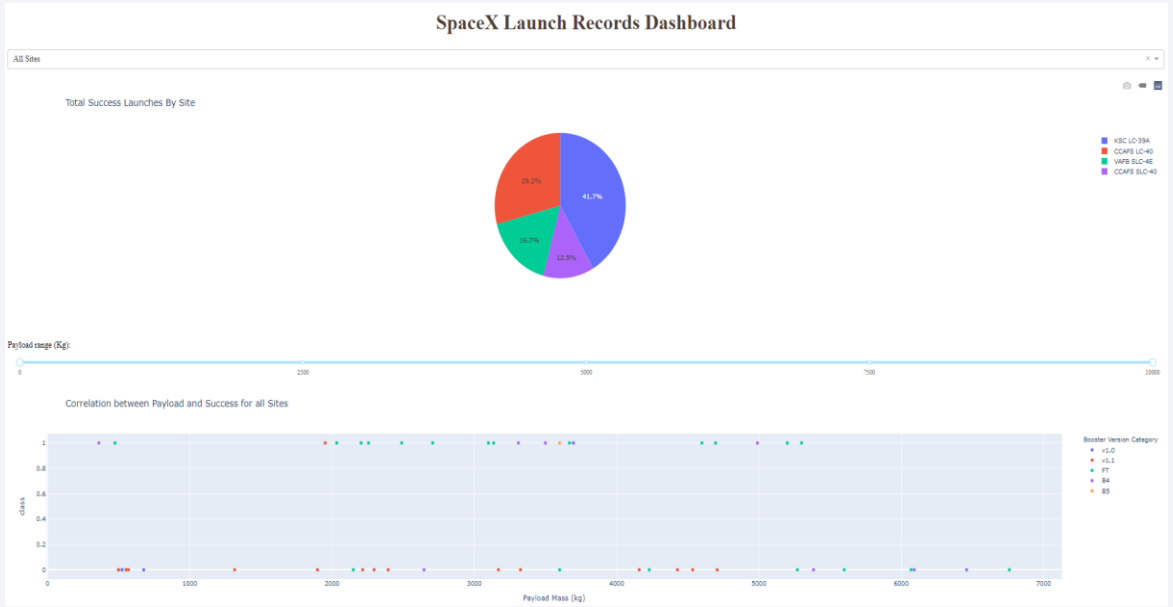
Results

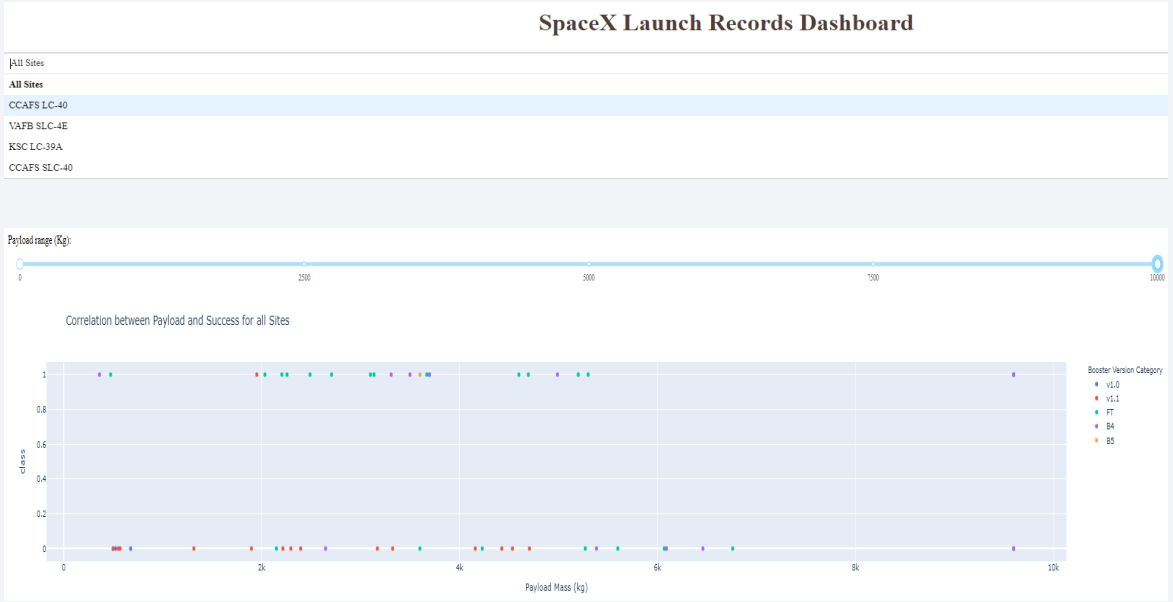
Exploratory data analysis results



Results

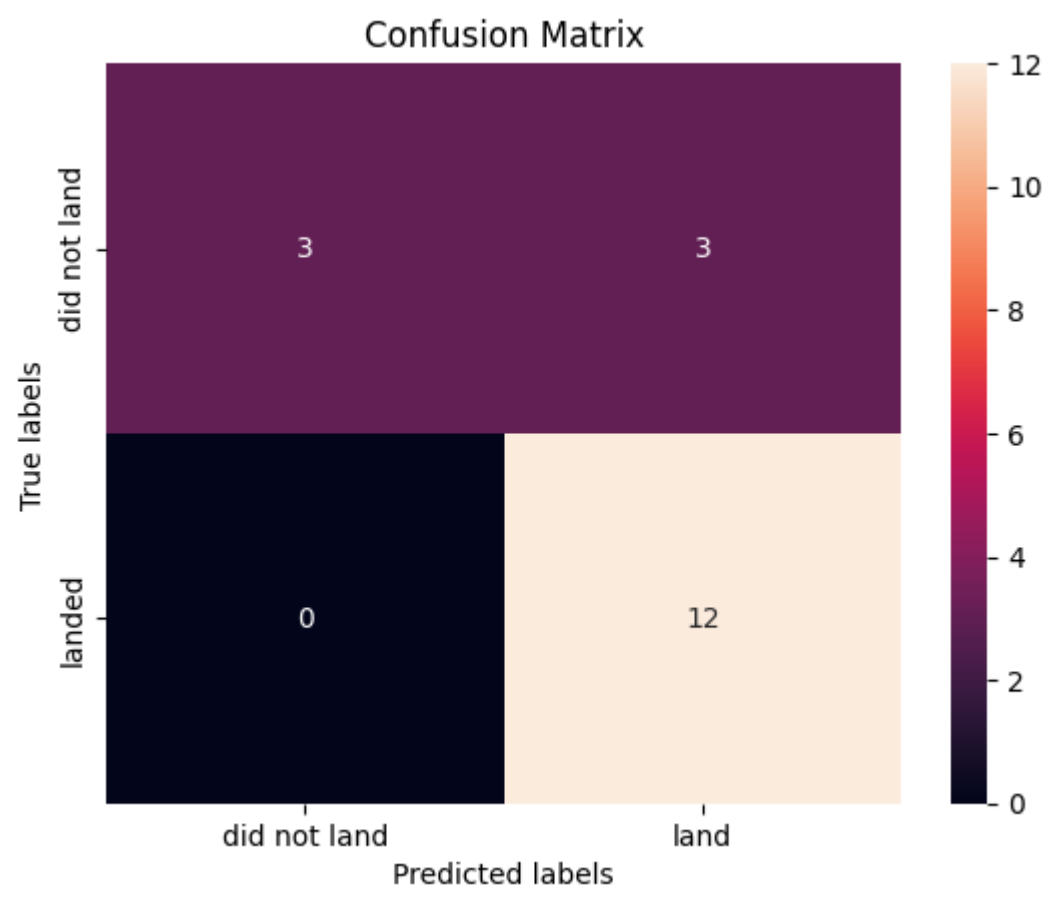
Interactive analytics demo in screenshots





Results

Predictive analysis results

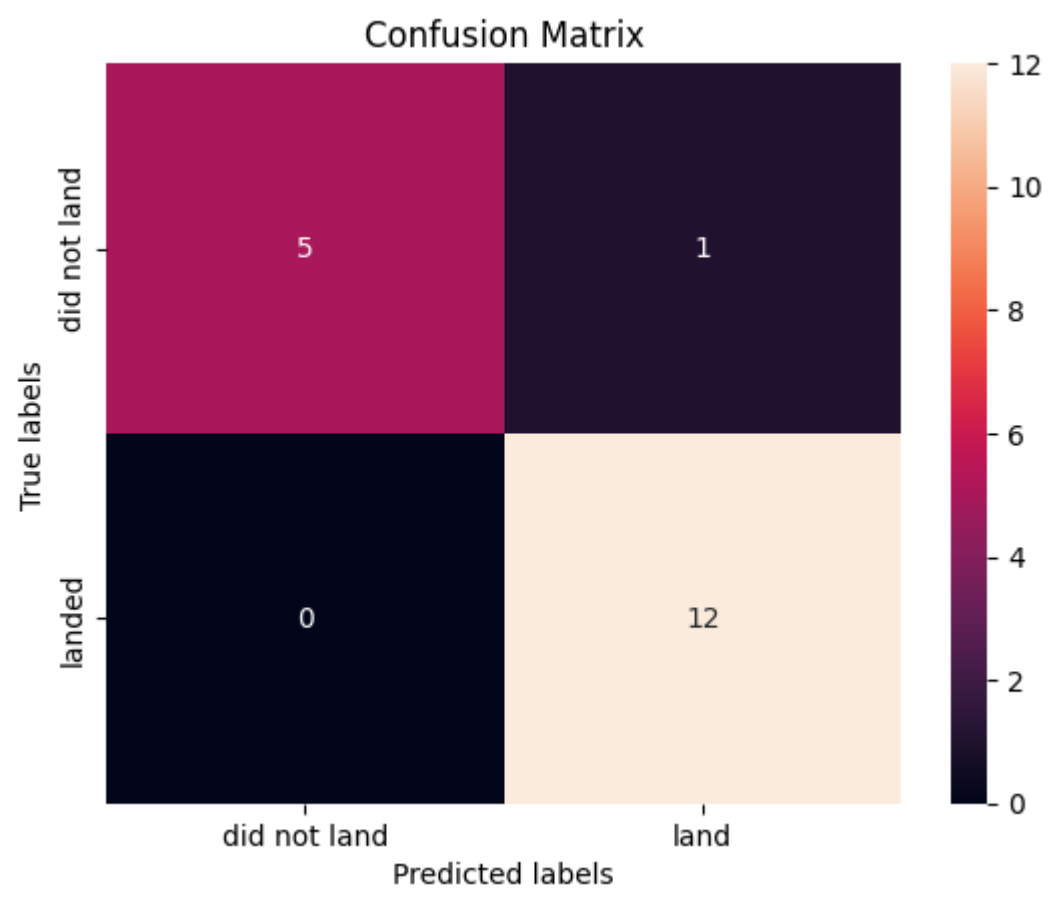


Confusion Matrix for Logistic Regression and Support Vector Machine

Model	Accuracy Score (2 d.p)
Logistic Regression	0.83
Support Vector Machine	0.83
Decision Tree	0.94
K-Nearest Neighbors	0.72

Results

Predictive analysis results

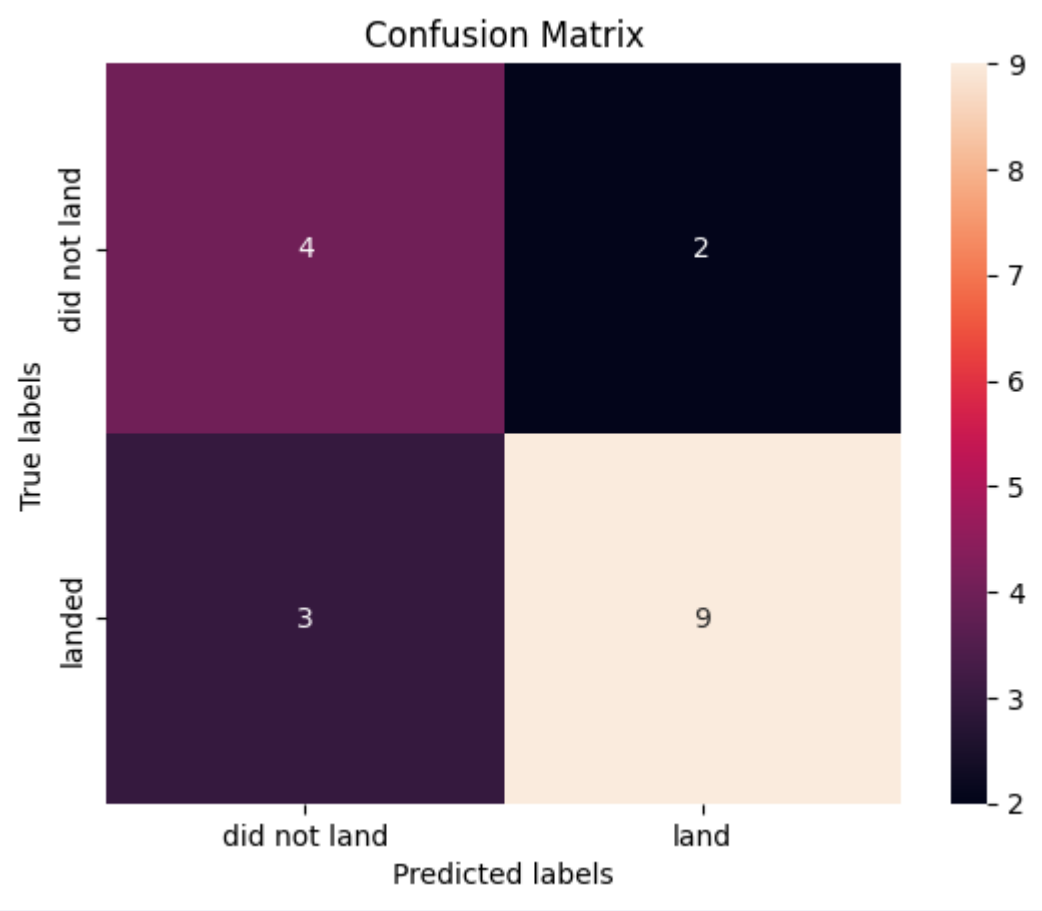


Confusion Matrix for Decision Tree

Model	Accuracy Score (2 d.p)
Logistic Regression	0.83
Support Vector Machine	0.83
Decision Tree	0.94
K-Nearest Neighbors	0.72

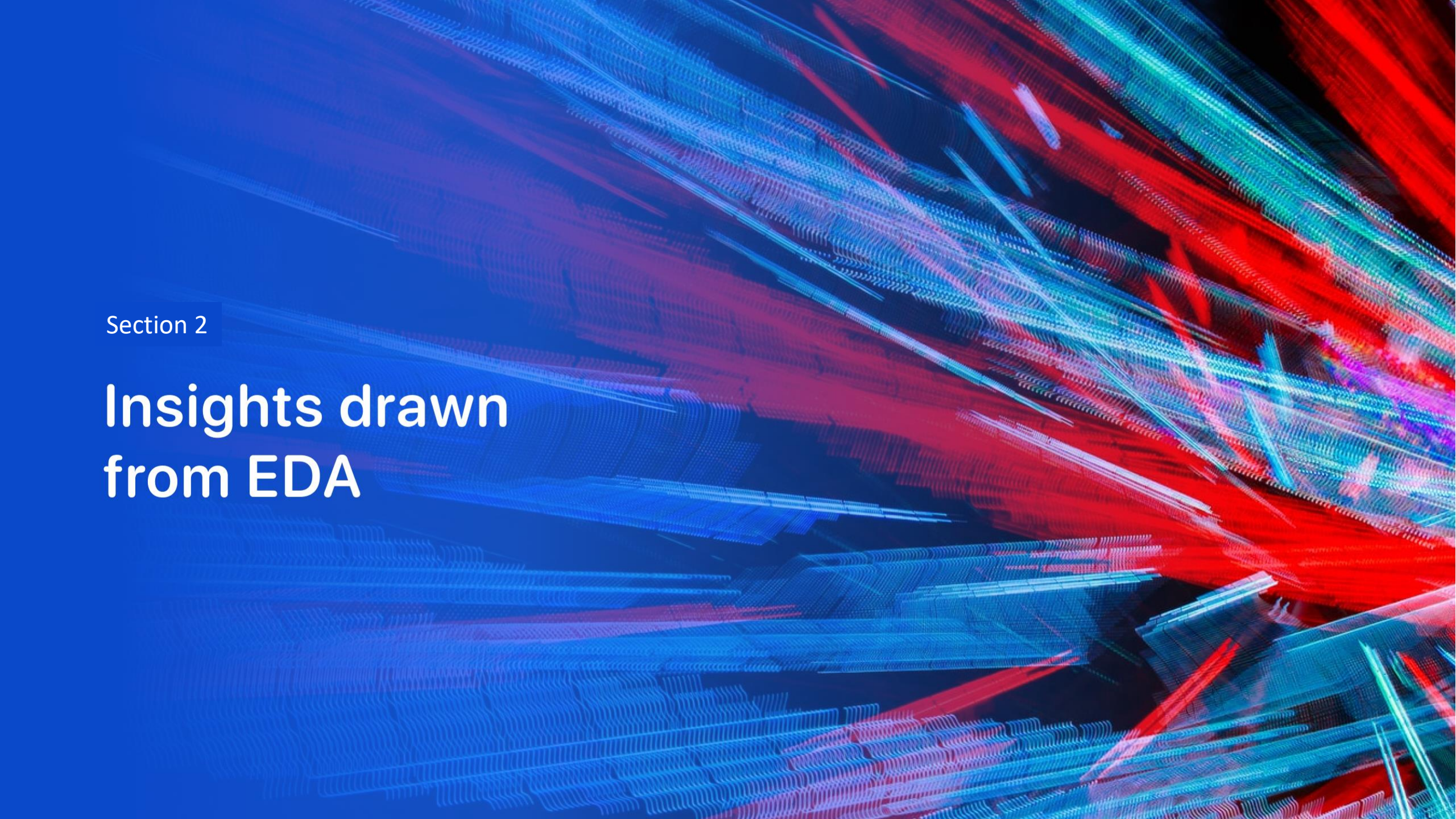
Results

Predictive analysis results



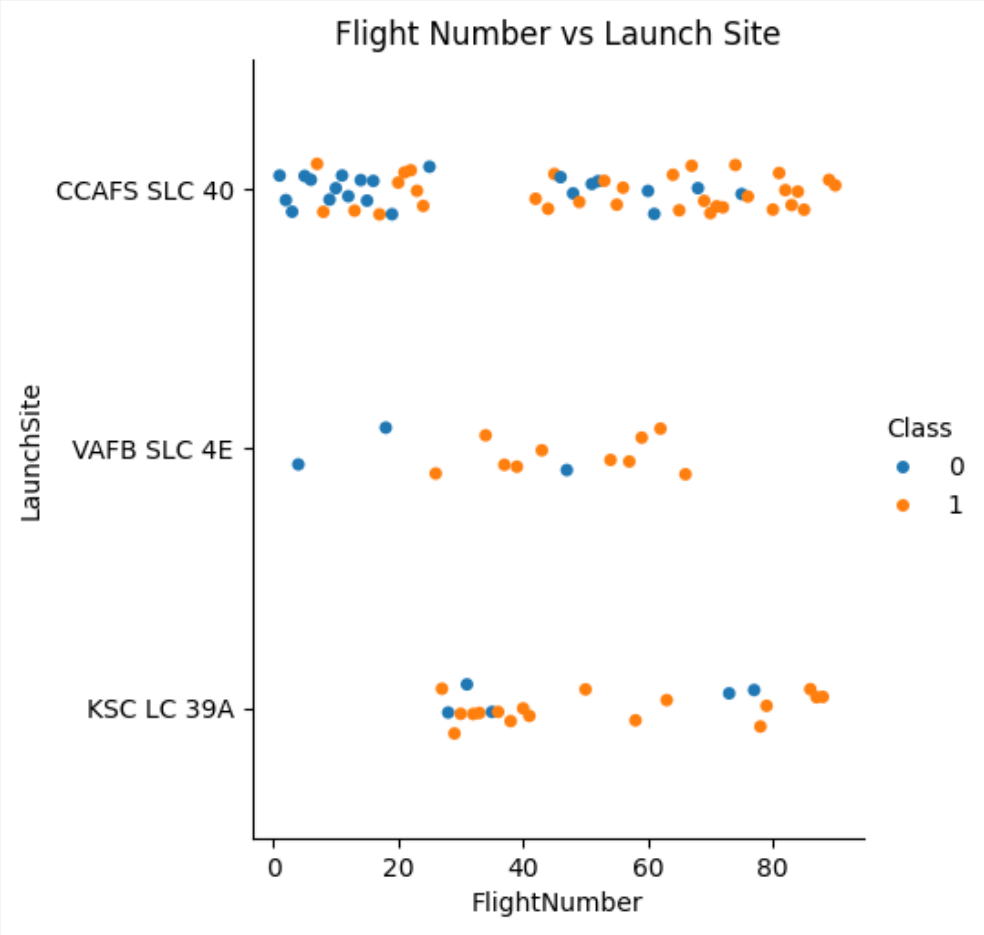
Confusion Matrix for K-Nearest Neighbors

Model	Accuracy Score (2 d.p)
Logistic Regression	0.83
Support Vector Machine	0.83
Decision Tree	0.94
K-Nearest Neighbors	0.72


Section 2
Insights drawn from EDA

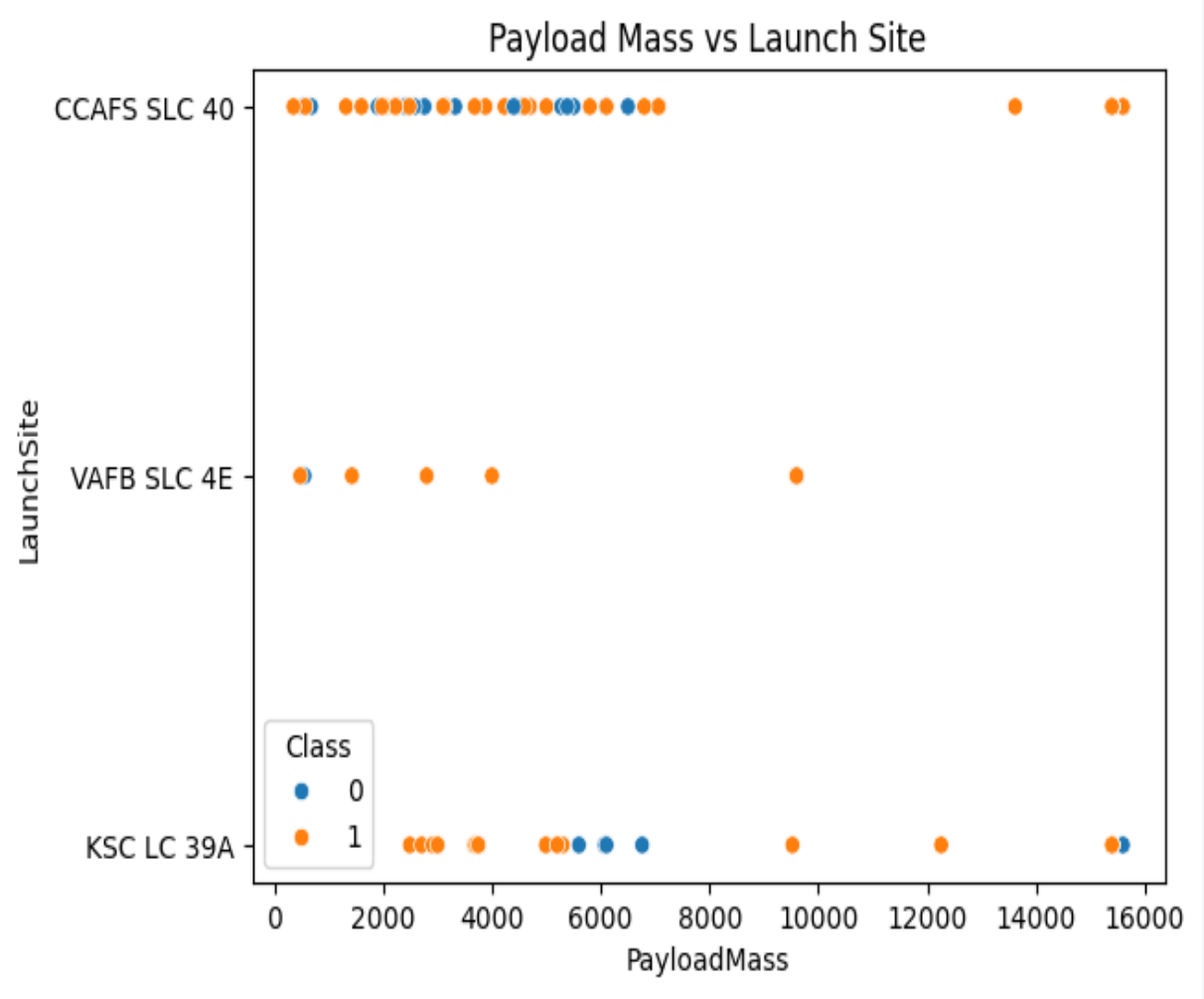
Flight Number vs. Launch Site

- Launch site CCAFS SLC 40 has the most number of rocket launches
- Launch site VAFB SLC 4E has the fewest number of rocket launches
- Launch site KSC LC 39A is the launch site with the highest launch success rate



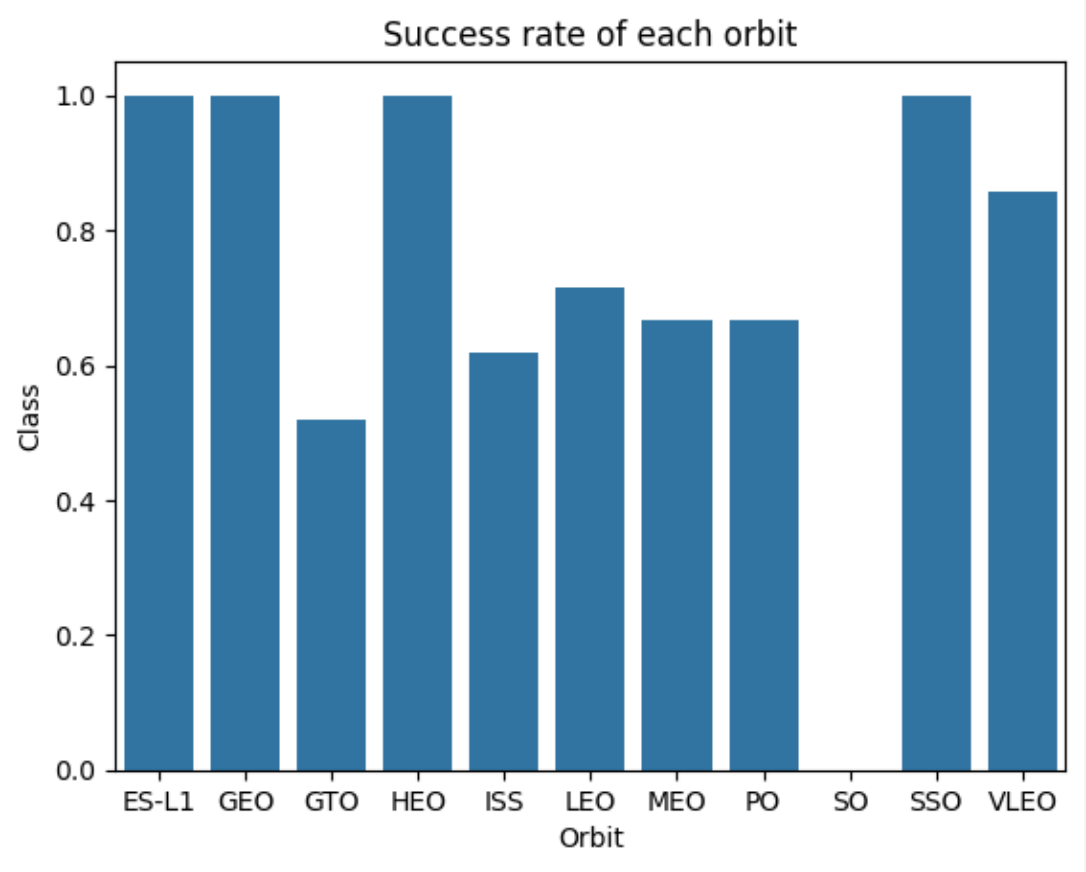
Payload vs. Launch Site

- The majority of the launches at the site CCAFS SLC 40 has a payload mass of less than 7000 kg
- The launch site VAFB SLC 4E doesn't have any launches with payload more than 10000 kg
- The launch site KSC LC 39A doesn't have any launches with less than 2000 kg



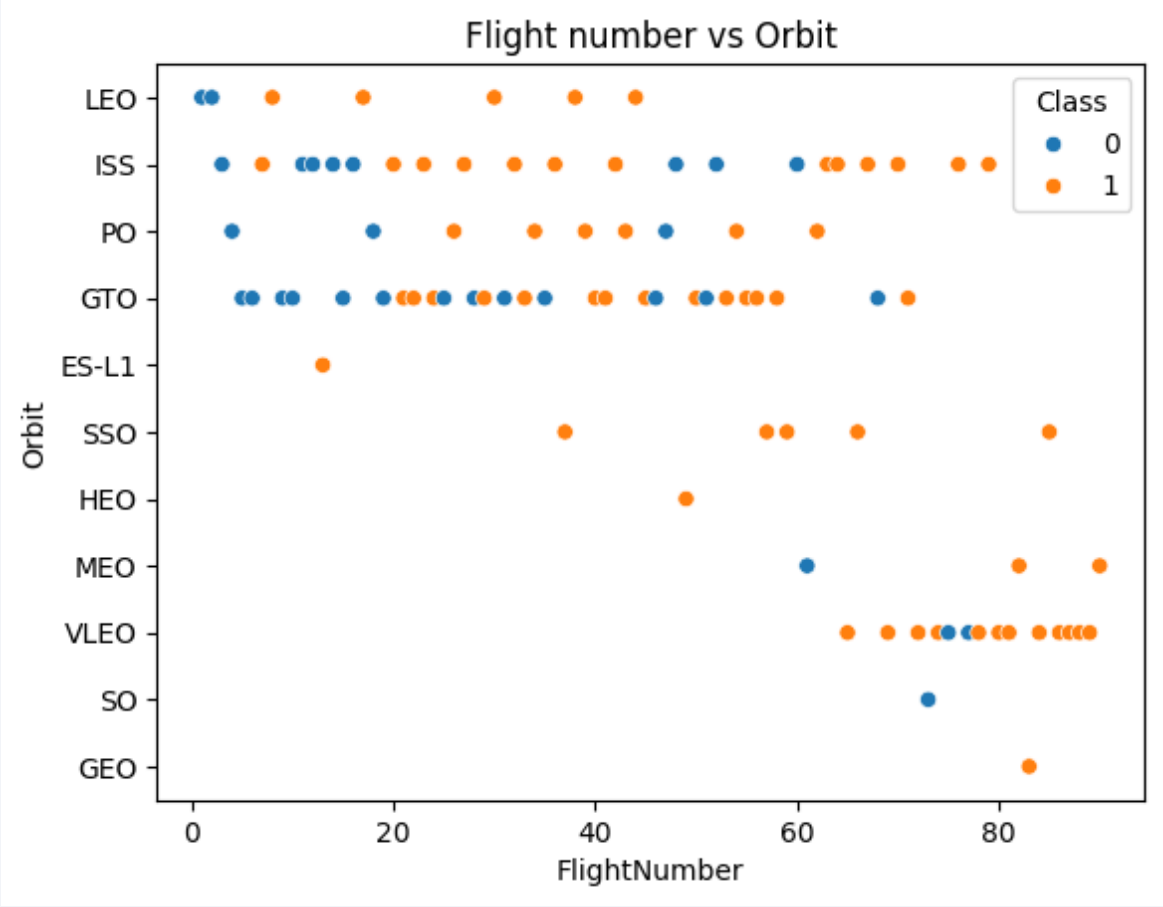
Success Rate vs. Orbit Type

- The following orbits have a 100% launch success so far: ES-L1, GEO, HEO, and SSO
- The orbit SO has never had a successful launch before
- The majority of orbits have a success rate of 0.6 or over



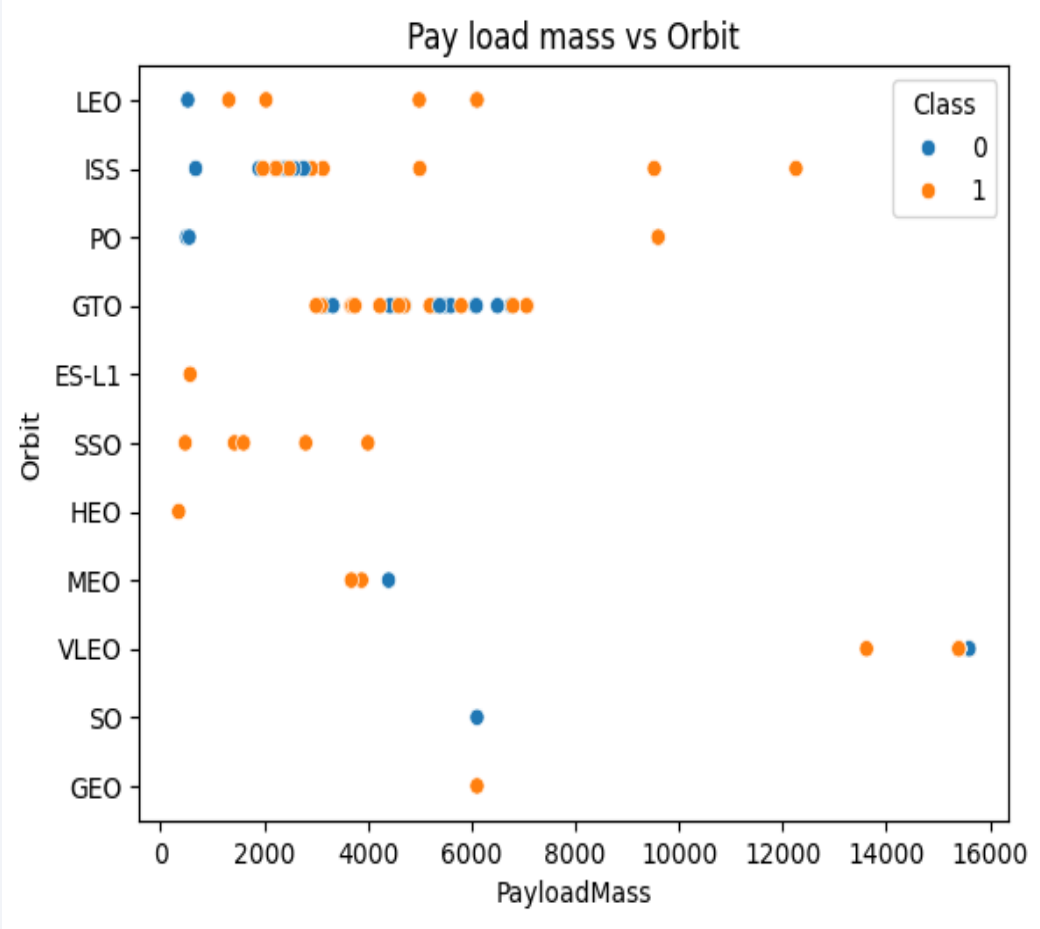
Flight Number vs. Orbit Type

- The following orbits only have 1 rocket launch so far: ES-L1, SO, HEO, and GEO
- The orbits ISS, GTO, and VLEO have the most rocket launches



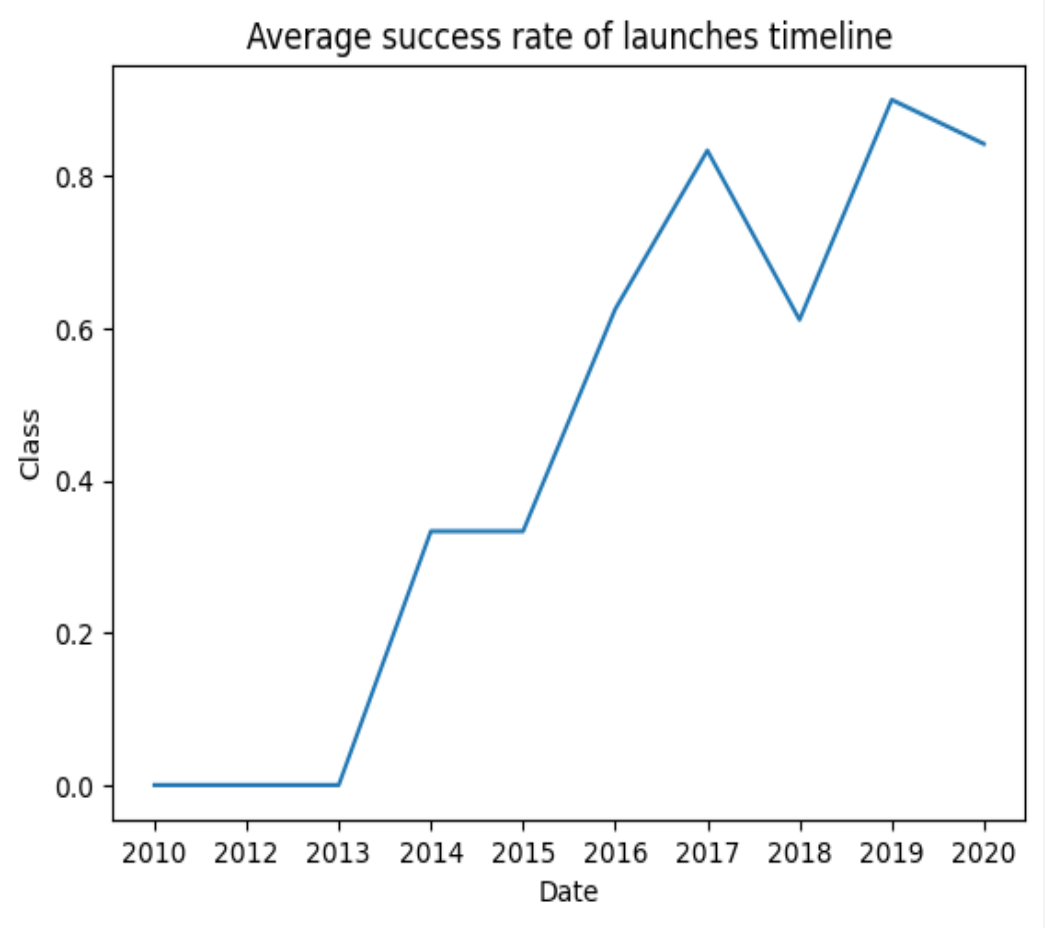
Payload vs. Orbit Type

- The launch site GTO has only had launches with payload mass within the range from 3000 – 7000 kg
- The launch site SSO has had successful launches with payload mass 4000 kg and below so far



Launch Success Yearly Trend

- The average success rate of launches has been increasing ever since 2013
- The interval between 2013 – 2014 has the sharpest rise in success rate
- The success rate reaches a plateau between 2014 – 2015
- The longest interval of rising success rate is between 2015 - 2017



All Launch Site Names

- The use of the DISTINCT method is to ensure there are no duplicates

```
%%sql
SELECT DISTINCT("Launch_Site") FROM SPACEXTABLE;
```

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

Launch Site Names Begin with 'CCA'

- Use LIMIT 5 to only get 5 records
- Use the % wildcard to represent any character sequence

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

```
%%sql

SELECT * FROM SPACEXTABLE
WHERE "Launch_Site" LIKE "CCA%"
LIMIT 5;
```

Total Payload Mass

- Use the SUM aggregate function to calculate the total payload mass
- Group by Customer to extract information related to NASA

```
%%sql

SELECT Customer, SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE
GROUP BY Customer
HAVING Customer = "NASA (CRS)";
```

Customer	SUM("PAYLOAD_MASS__KG_")
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- Use the AVG aggregate function to calculate the average payload mass
- Group by Booster_Version to extract information related to booster version F9 v1.1

```
%%sql

SELECT "Booster_Version", AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE
GROUP BY "Booster_Version"
HAVING "Booster_Version" = "F9 v1.1";
```

Booster_Version	AVG("PAYLOAD_MASS__KG_")
F9 v1.1	2928.4

First Successful Ground Landing Date

- Use the MIN method on the Date column to get the first date
- Filter to only include successful landing outcomes on ground pad

```
%%sql

SELECT MIN(Date) FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Success (ground pad)";
```

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql

SELECT "Booster_Version" FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Success (drone ship)" AND ("PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000);
```

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

Total Number of Successful and Failure Mission Outcomes

- Use of the COUNT(*) method to count all the rows
- Group by Mission_Outcome to segment into different groups of outcomes

```
%%sql

SELECT "Mission_Outcome", COUNT(*) FROM SPACEXTABLE
GROUP BY "Mission_Outcome";
```

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

Boosters Carried Maximum Payload

- Use of the DISTINCT method to ensure there are no duplicates
- Use of inner subquery to find the maximum payload max first before feeding the result to the outer query to find the corresponding booster versions

```
%%sql

SELECT DISTINCT("Booster_Version") FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
```

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 of substr method to extract the month and date
- Select only records in the year 2015 and has a landing outcome of failure in drone ship

```
%%sql

SELECT "Booster_Version", "Launch_Site", "Landing_Outcome", substr(Date, 6, 2) AS "Month" FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = "2015" AND "Landing_Outcome" = "Failure (drone ship)";
```

Booster_Version	Launch_Site	Landing_Outcome	Month
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)	01
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)	04

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

- Use of ORDER BY 2 DESC to rank the count of landing outcomes in descending order
- Count of the number of launches with different landing outcomes between the date 2010-06-04 and 2017-03-20

```
%%sql

SELECT "Landing_Outcome", COUNT(*) FROM SPACEXTABLE
GROUP BY "Landing_Outcome"
HAVING Date BETWEEN "2010-06-04" AND "2017-03-20"
ORDER BY 2 DESC;
```

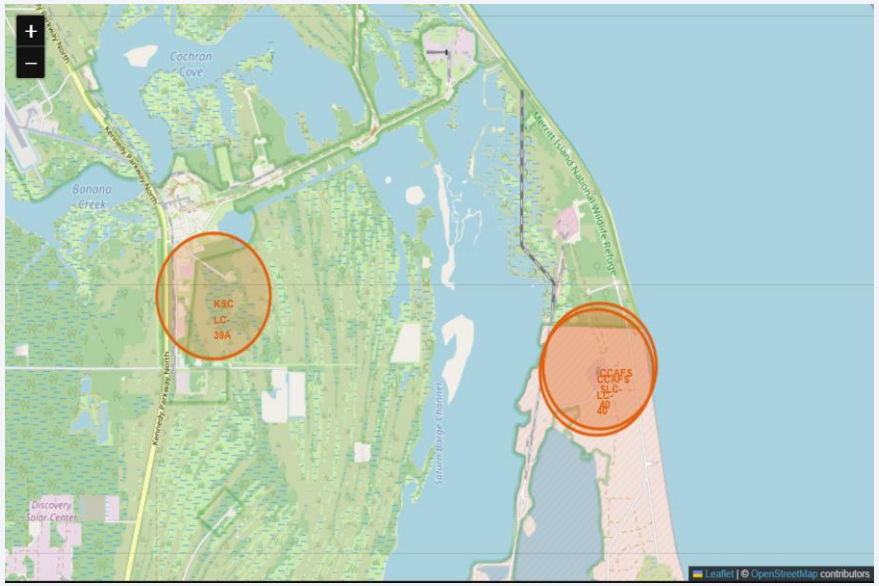
Landing_Outcome	COUNT(*)
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

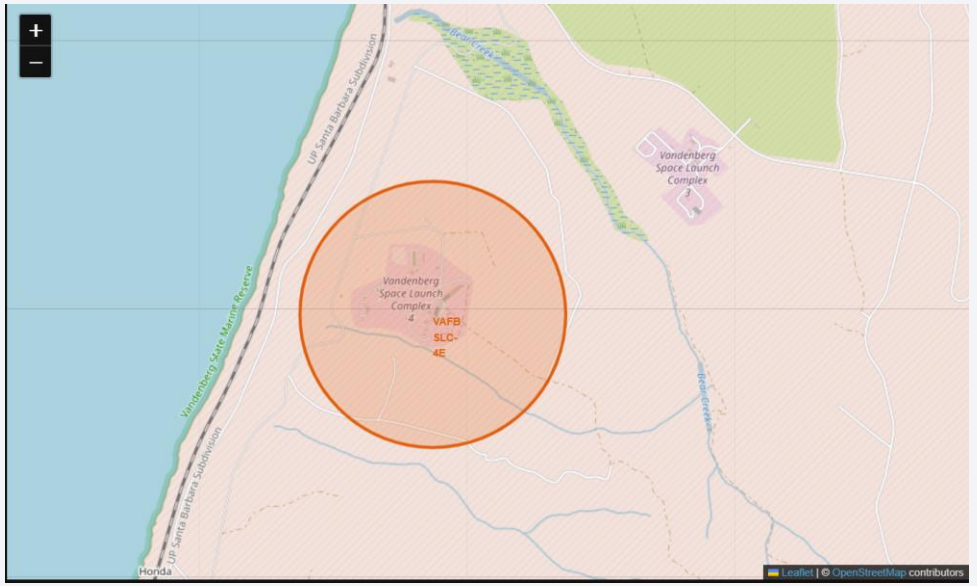
Section 3

Launch Sites Proximities Analysis

Locations of different launch sites

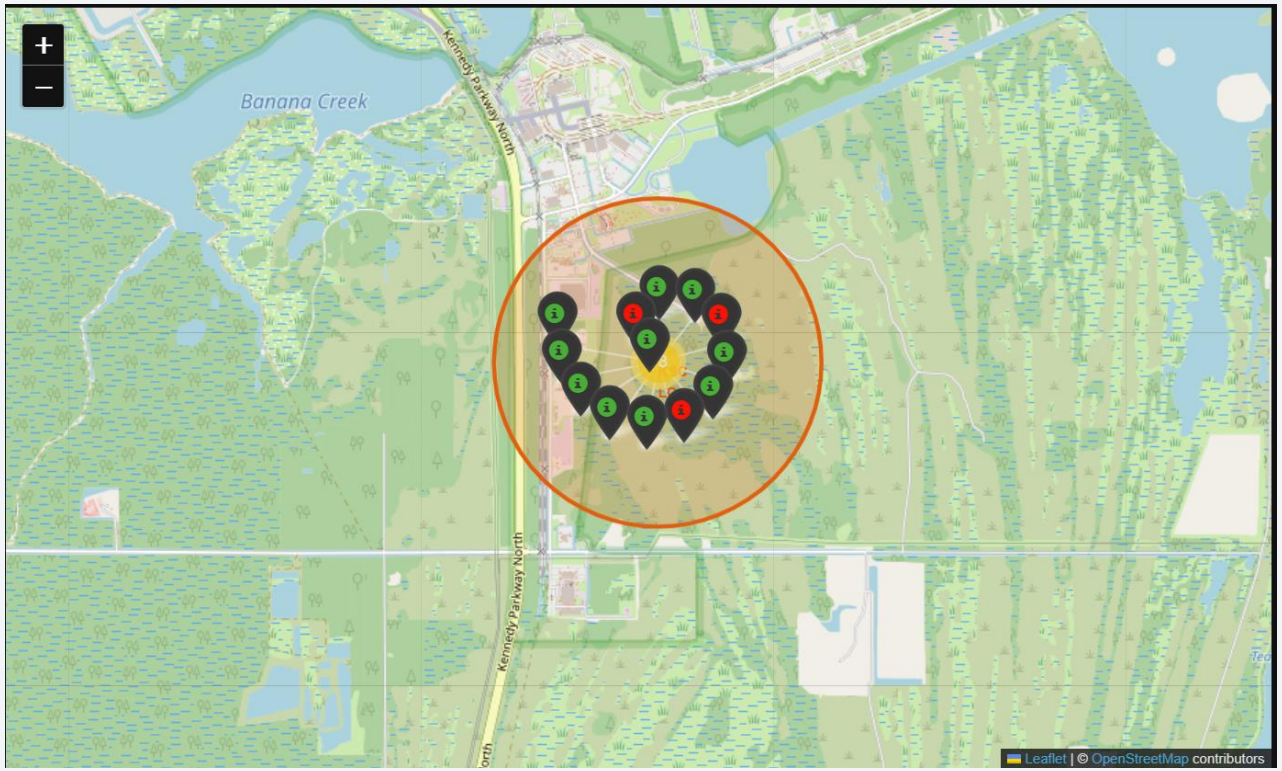
- There are a total of 4 launch sites, one on the west coast and the other three on the east coast
- Launch site CCAFS LC-40 and CCAFS SLC-40 share approximately the same location
- All launch sites are very close in proximity to the coast





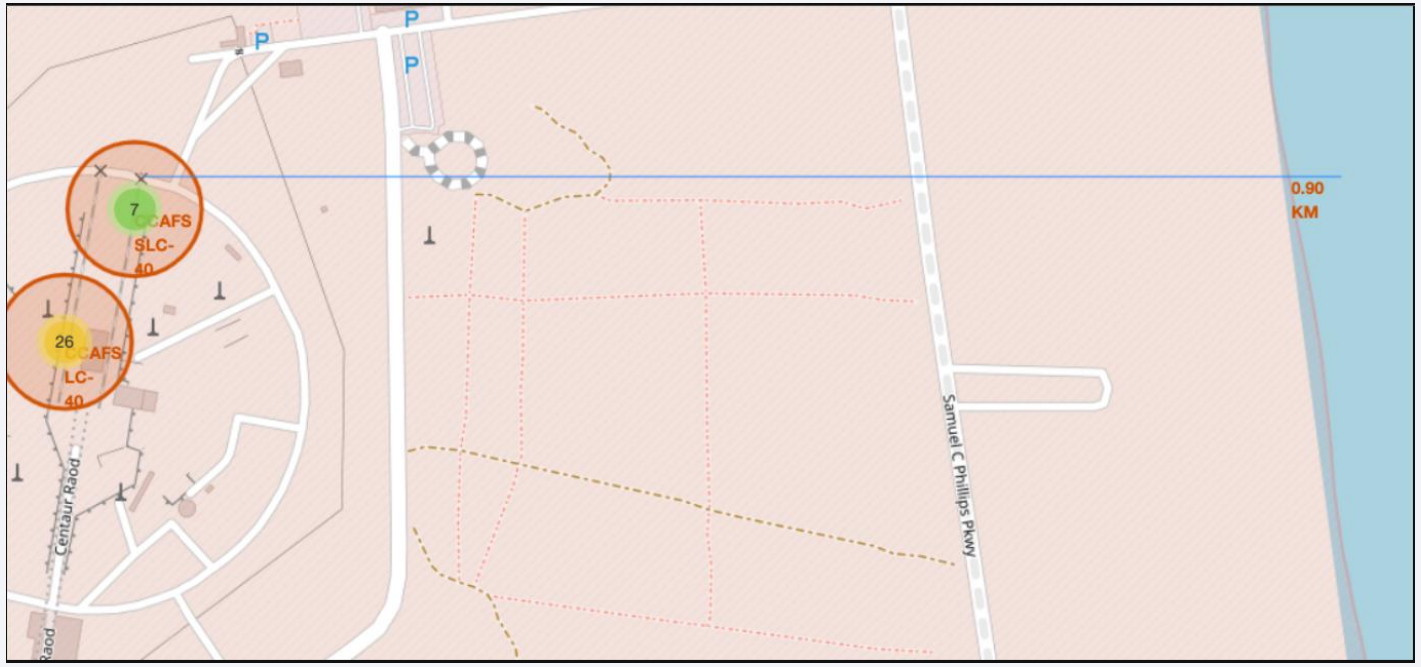
Launch records at each launch site

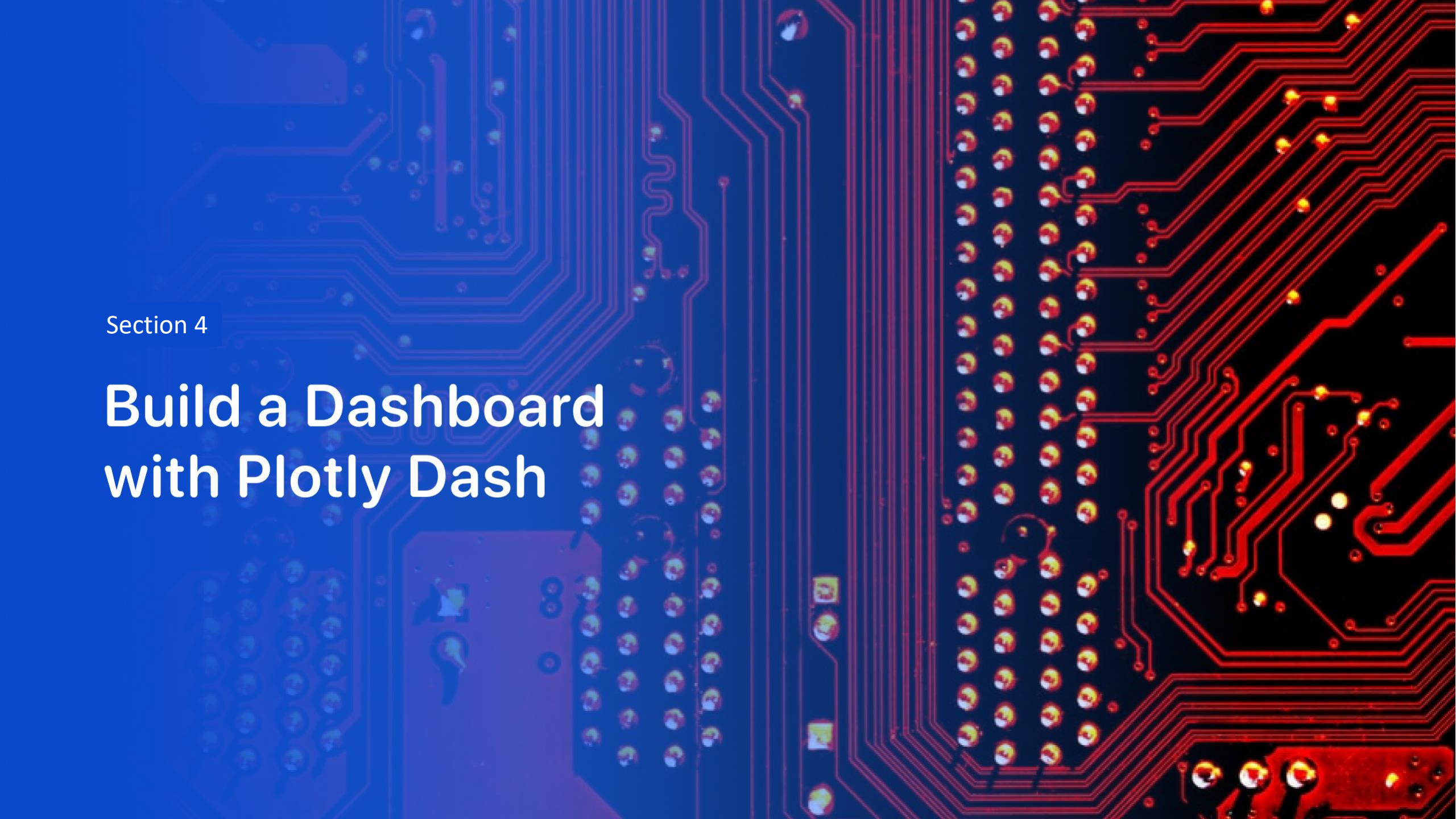
- Green markers denote successful launches while red markers denote unsuccessful launches
- Launch records at each launch site represented by markers



Locations nearby launch site

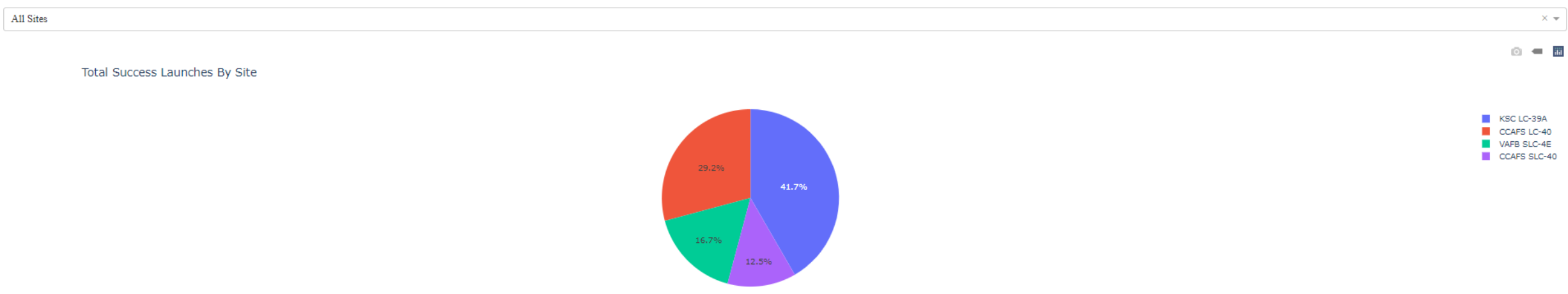
- Nearby locations like highway, coastline, railway is denoted on the map along with the distance from the center of the launch sites
- The distance is calculated based on coordinates (latitude and longitude)
- Coordinates of nearby locations are located using MousePosition




Section 4
Build a Dashboard with Plotly Dash

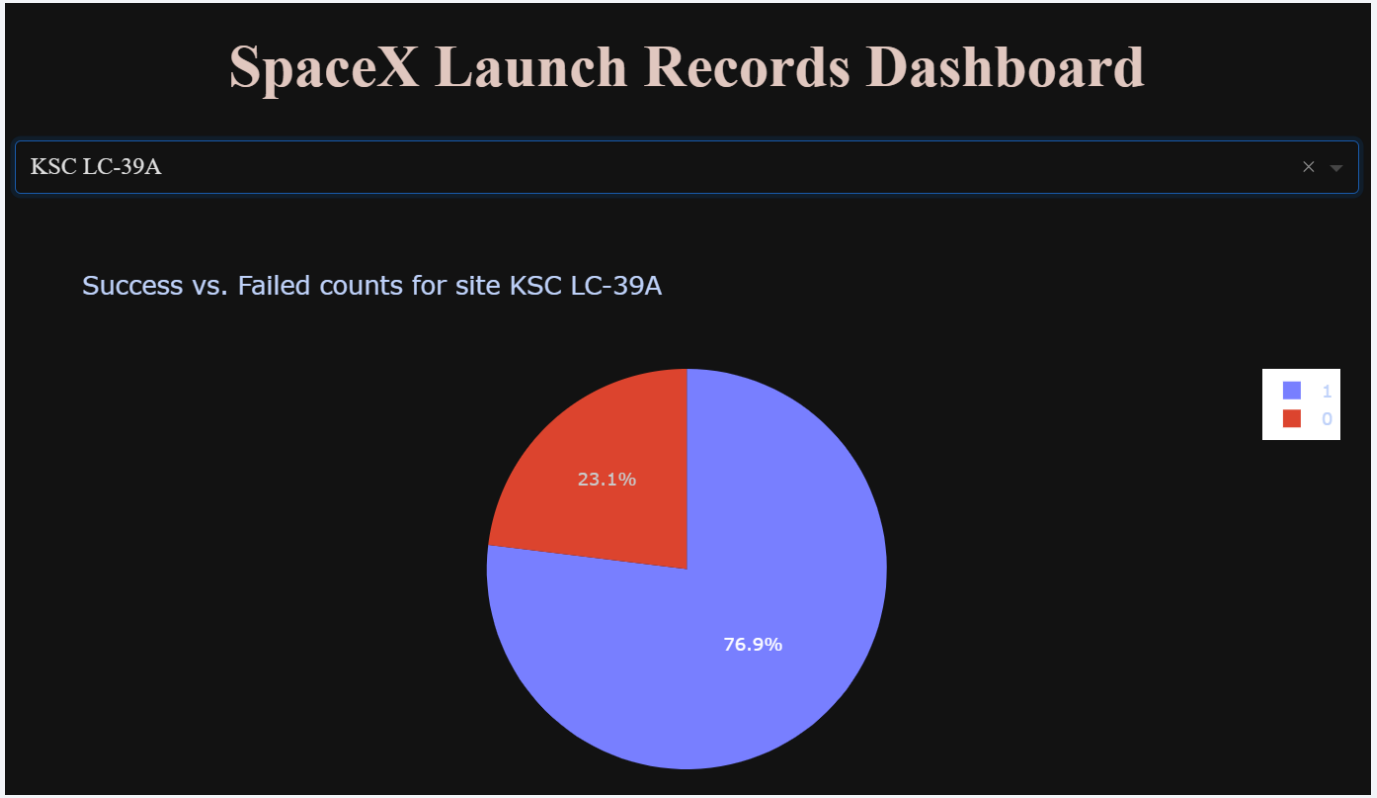
Success launch count for all sites

- Launch site KSC LC-39A has the highest launch count, followed by launch site CCAFS LC-40. Both take up more than 30% of the total success launch count
- Launch site CCAFS SLC-40 has the lowest launch count



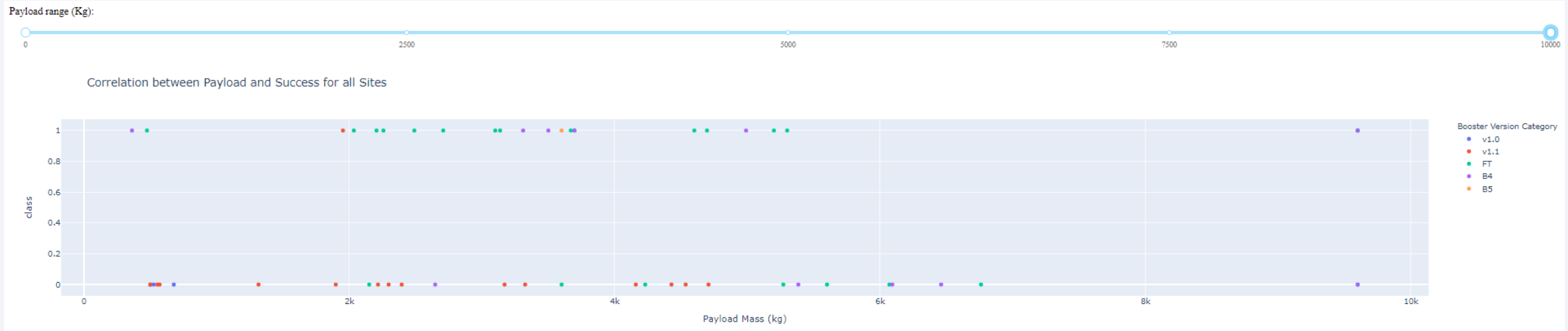
Launch site with highest success ratio

- Launch site KSC LC-39A has the highest success ratio of more than ¾ of the time



Correlation between Payload and Success for all sites

- Booster version FT has the highest launch success rate
- Most of the success launches have the payload mass within the range 2000 – 6000 kg
- Only booster version v1.0 has ever attempted and successfully launched with a payload mass of more than 8000 kg

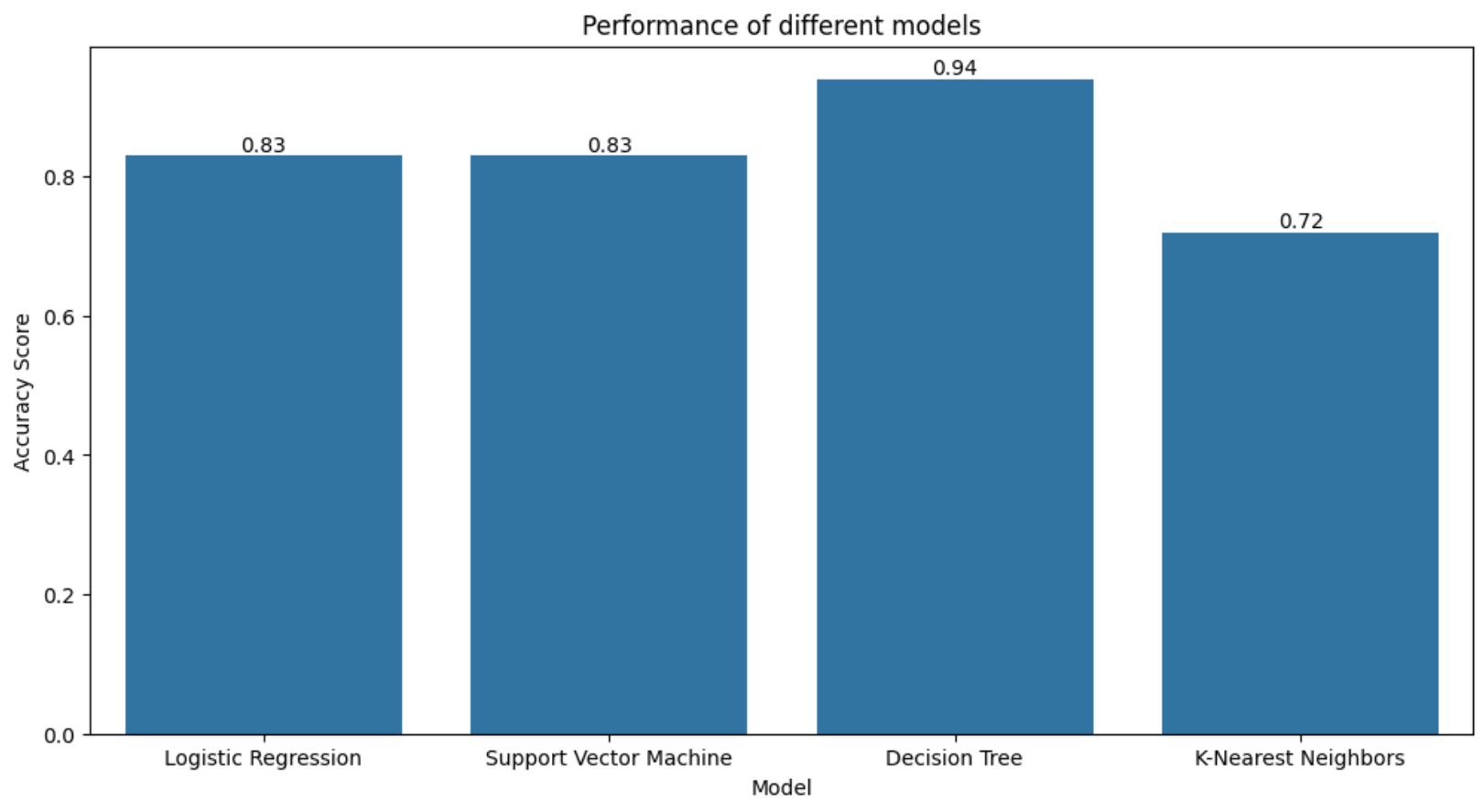


Section 5

Predictive Analysis (Classification)

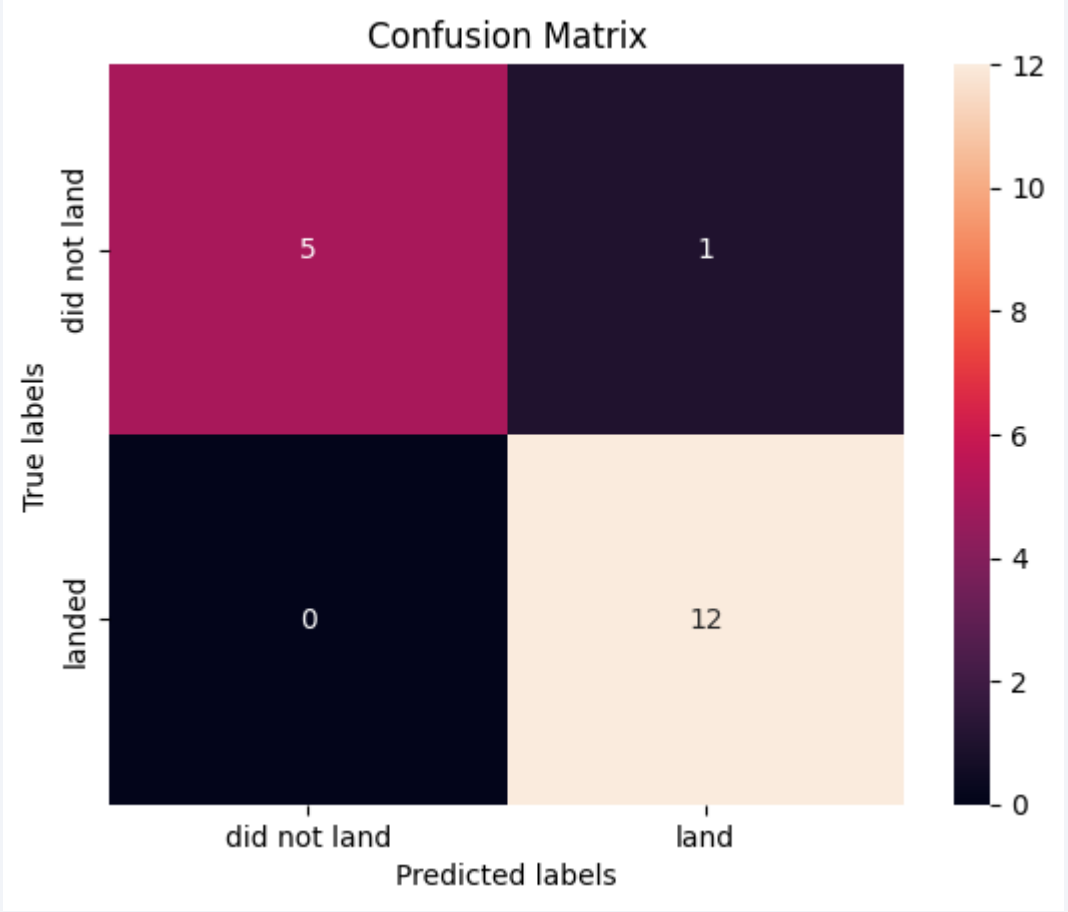
Classification Accuracy

- The decision tree model has the highest accuracy score of 0.94
- The K-Nearest Neighbors has the lowest accuracy score of 0.72



Confusion Matrix

- The decision tree model has the highest accuracy score of 0.94
- Based on the test data set, the model has shown to always correctly predict whether a launch landed
- Based on the test data set, the model correctly predicts whether a launch is unsuccessful 83.3% of the time
- Due to the test data set sample size being relatively small, the current accuracy score might not fully reflect the model's performance



Confusion Matrix for Decision Tree

Conclusions

- Decision Tree classification model yields the highest accuracy score, however, due to the test data set sample size being relatively small, the current accuracy score might not fully reflect the model's performance
- Launch site KSC LC 39A is the launch site with the highest launch success rate
- The following orbits have a 100% launch success so far: ES-L1, GEO, HEO, and SSO
- The orbit SO has never had a successful launch before
- Most of the success launches have the payload mass within the range 2000 – 6000 kg
- Booster version FT has the highest launch success rate

Appendix

- All relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets created/used during this project can be found in the GitHub URL

Thank you!