Marina Safarova

May 2025

SPACEX FALCON-9 FIRST STAGE LANDING PREDICTION

OUTLINE

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

EXECUTIVE SUMMARY

- Summary of methodologies
  - Data Collection using SpaceX API
  - Data Collection with Web Scraping
  - Data Wrangling
  - EDA using SQL
  - EDA Data Visualization using Python Pandas and Matplotlib
  - Launch Sites Analysis with Folium and Ploty Dash
  - Machine Learning Landing Prediction
- Summary of all results
  - EDA results
  - Interactive Analysis
  - Predictive Analysis

INTRODUCTION

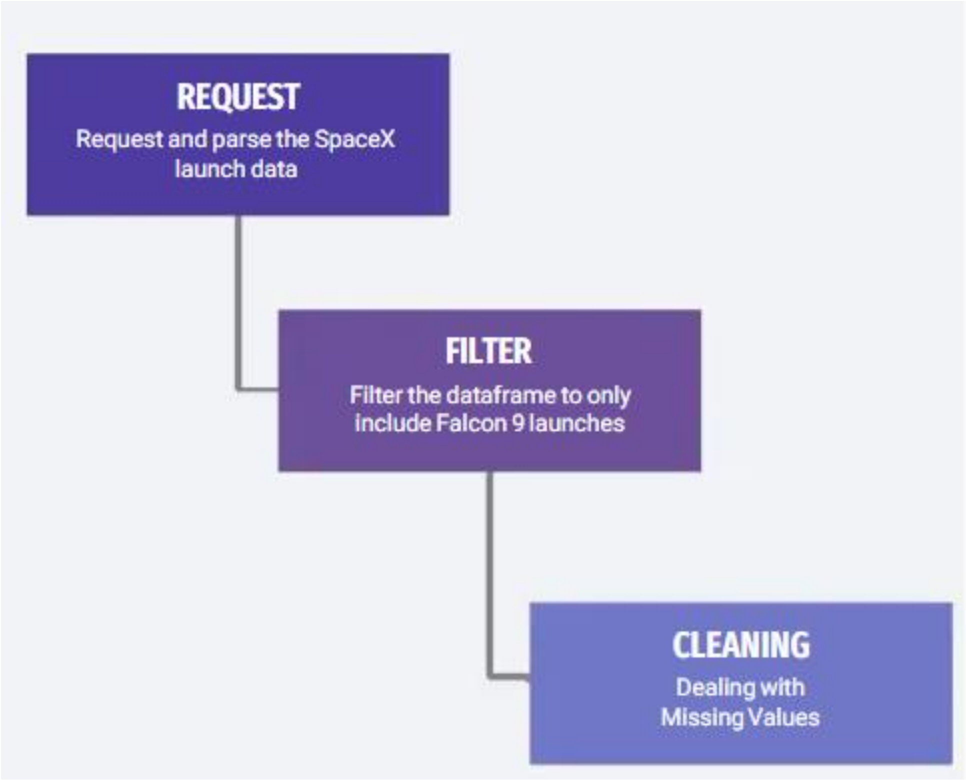
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
- In this analysis we will focus on the study of the first stage of the SpaceX Falcon 9 Rocket in order to obtain conclusions that allow us to make cost projections as well as obtain insight that allows to bid against SpaceX for a rocket launch.
- We will work on finding the following solutions:
  - How to predict if the rocket will land successfully.
  - What parameters can determine the success rate of a successful landing.
  - Project and determine the costs of future launches.

SECTION 1

METHODOLOGY

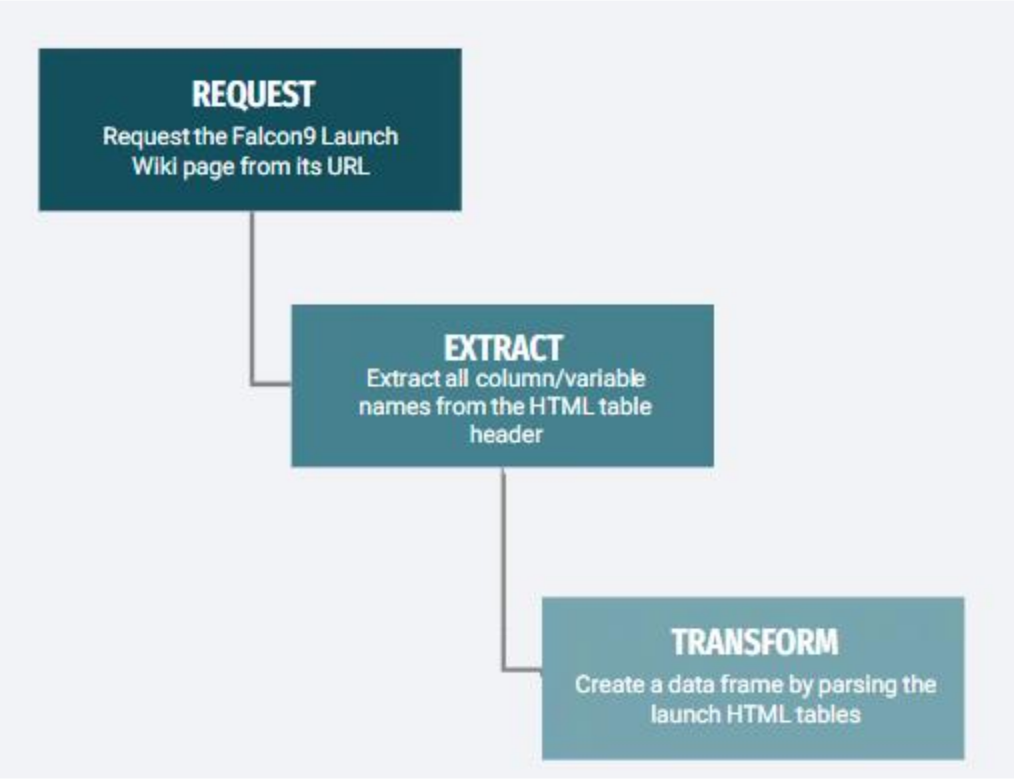
DATA COLLECTION – SPACEX API

- The SpaceX API data are available publically (link)
- Data collection steps:
  - Place call to SpaceX API (using *requests.get()* method)
  - Extract nested data and convert its format (using *.json()* and *.json_normalize()*)
  - Generate specific columns of data (using defined functions)
  - Combine separate columns into Data Frame (using *pd.DataFrame()*)
  - Filter out all launches with rockets other than the Falcon9 (removing other dates with != condition)
  - Handle missing values (using mean values and the *.replace()* function to replace np.nan values)
- For more details follow this link



DATA COLLECTION – WEB SCRAPING

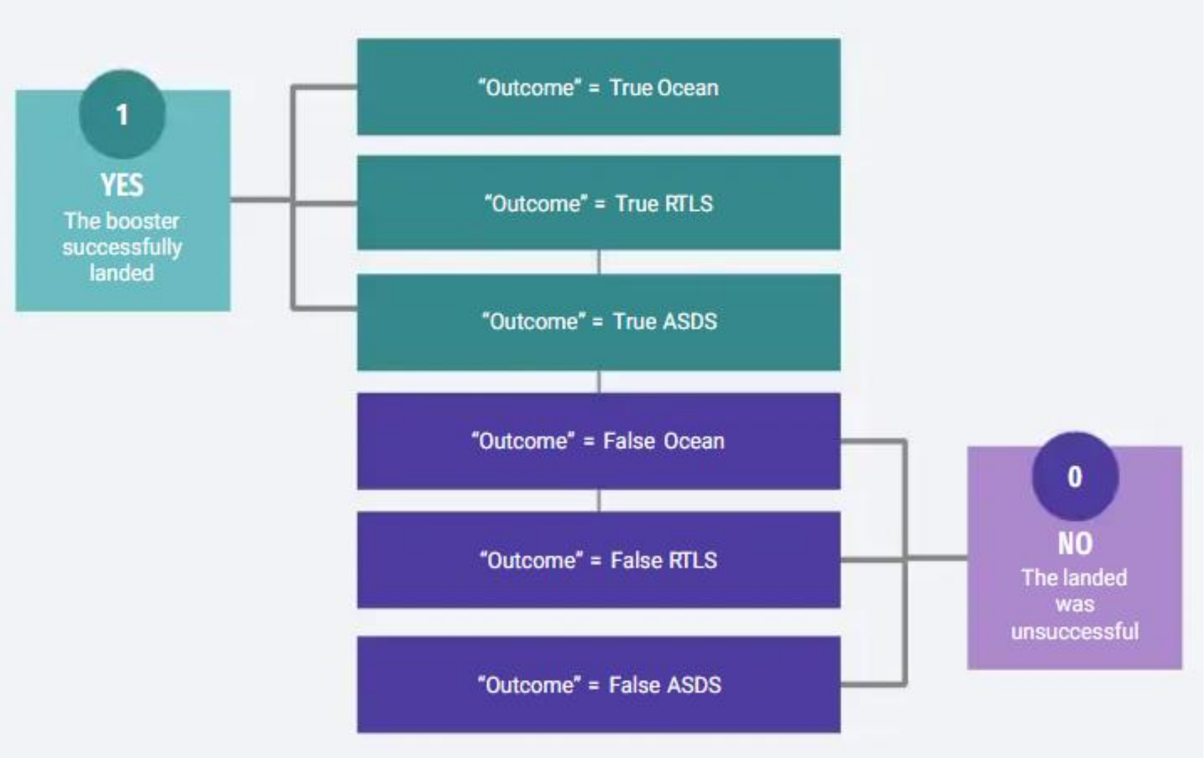
- It's possible to collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches
- Data collection steps:
  - Web Scrape the page to get the HTML text (using *requests.get()* method)
  - Create a Beautiful Soup object from the response text content (using *BeautifulSoup()*)
  - Find the tables (using *.find_all()* function)
  - Extract the column names from the tags from the launch table (using defined functions)
  - Create DataFrame by parsing the launch tables (creating an empty dictionary, filling it with launch records extracted from table rows and creating a dataframe from that dictionary using *pd.DataFrame()*)
- For more details follow this link



DATA WRANGLING

Data wrangling steps:

- Calculate (using *.value_counts()*):
  - the number of launches on each site
  - the number and occurrence of each orbit
  - the number and occurence of mission outcome of the orbits
- Create a landing outcome label from Outcome column:
  - Creating a set of outcomes where the second stage did not land successfully (using *.keys()*)
  - Creating a list where the element is zero if the corresponding row in unsuccessful landing set, otherwise, it's one, and assign it to the variable landing_class (using *.apply()*)



For more details follow this link

EDA WITH SQL

- Information extracted using queries:
  - Launch sites
  - Payload masses
  - Dates
  - Booster types
  - Mission outcomes
- For more details follow this link

EDA WITH DATA VISUALIZATION

- Performed EDA and Feature Engineering using Pandas and Matplotlib
- Used scatter plots (*sns.catplot()*) to visualize relationship between flight number and launch site, payload and launch site, flight number and orbit type, payload and orbit type
- Used bar chart (*sns.barplot()*) to visualize success rate of each orbit type
- Used line plot (*sns.lineplot()*) to visualize the launch success yearly trend
- For more details follow this link

INTERACTIVE VISUAL ANALYTICS WITH FOLIUM

- Mark all launch sites on a map (using *folium.Circle()*, *folium.map.Marker()* and *.add_child()*)
- Mark the successful/failed launches for each site on the map (using *folium.Marker()* and *.add_child()*)
- Calculate the distances between launch site and neares railway, highway, town and coastline (using *folium.Marker()*, *folium.PolyLine()* and *.add_child()*)
- For more details follow this [link]

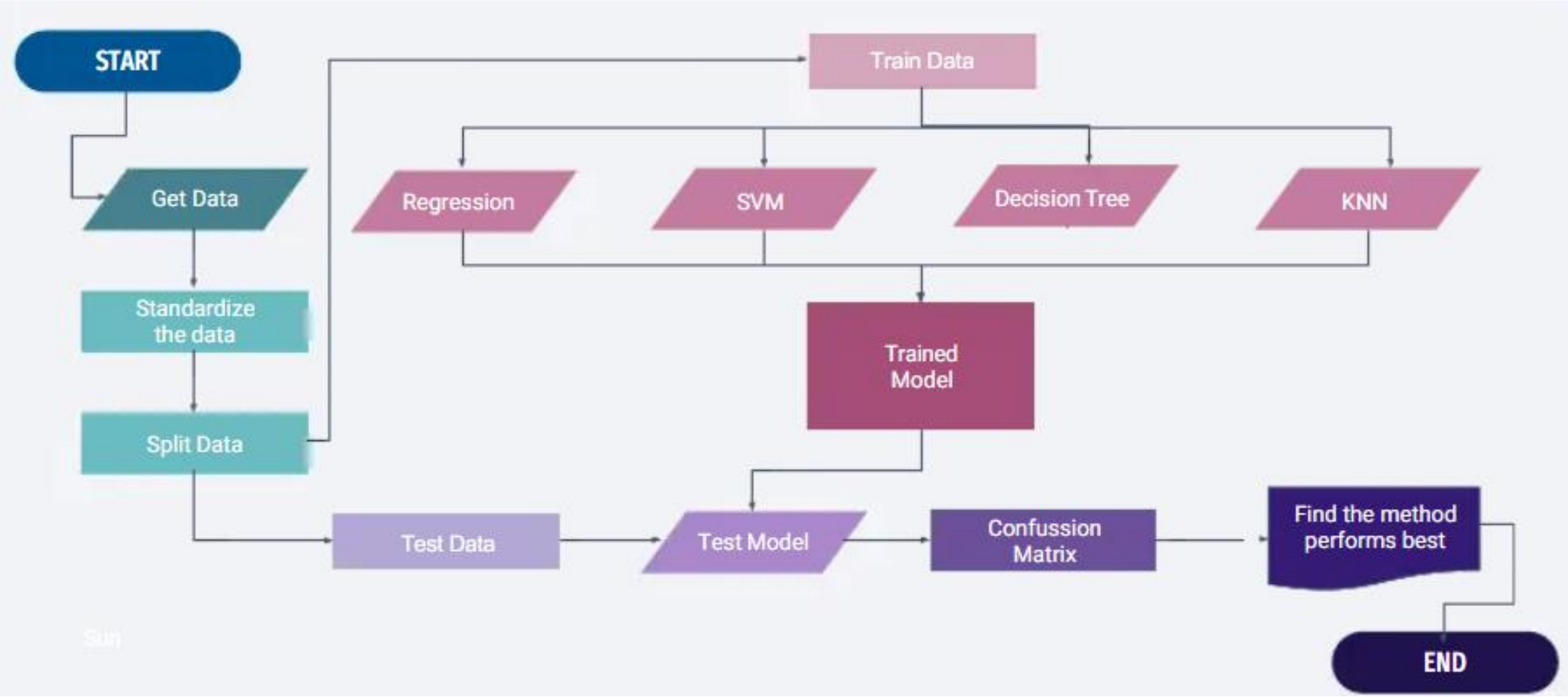
BUILD A DASHBOARD WITH PLOTLY DASH

- Built an interactive dashboard application with Plotly Dash by:
  - adding a launch site drop-down inputcomponent (using *dcc.Dropdown()*)
  - adding a callback function to render success-pie-chart based on selected site dropdown (using *@app.callback()*)
  - adding a range slider to select payload (using *dcc.RangeSlider()*)
  - adding a callback function to render the success-payload-scatterchart scatter plot (using *@app.callback()*)
- For more details follow this link

MACHINE LEARNING LANDING PREDICTION

- In order to find the best ML model (LogisticRegression, SVM, Classification Tree, KNN) that would perform best using the test data:
  - first create an object for each of the algorithms, then create GridSearchCV object and assigne them a set of parameters forneach model
  - for each of the models under evaluation, the GridSearchCV object was createdwith cv=10, then fit the training data into GridSearch object for each model to find best hyperparameter
  - after fitting the training set, we output GridSearchCV object for each model, then display the best parameters using data attribute *best_params_* and the accuracy on the validation data using *best_score_* attribute.
  - finally using *.score()* to calculate the accuracy on the test data for each model and plot confussion matrix for each model using test and predicted outcomes
- For more details follow this link

MACHINE LEARNING LANDING PREDICTION



SECTION 2
RESULTS

EDA RESULTS
USING SQL

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

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

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

There are four unique launch sites

EDA RESULTS

USING SQL

```
%sql SELECT * FROM SPACEXTBL 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

This gives us sense of the data containsed in database table

EDA RESULTS

USING SQL

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

The total payload carried by boosters from NASA(CRS) is 45596 kg

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Average Payload Mass(Kgs)", Booster_Version FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1'
```

Average Payload Mass(Kgs)	Booster_Version
2928.4	F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2928 kg

EDA RESULTS

USING SQL

```
%sql SELECT MIN(DATE), Landing_Outcome FROM 'SPACEXTBL' WHERE Landing_Outcome = 'Success (ground pad)';
```

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

MIN(DATE)	Landing_Outcome
2015-12-22	Success (ground pad)

First successful landing outcome on ground pad occurred in 22th December 2015.

```
%sql SELECT DISTINCT Booster_Version, PAYLOAD_MASS__KG_, Landing_Outcome FROM SPACEXTBL WHERE "Landing_Outcome" =
 "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Four booster versions with payload mass between 4000 and 6000 kg that have successfully landed

EDA RESULTS

USING SQL

```
%sql SELECT Mission_Outcome, COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome"
```

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

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

There was 100 successful missions and 1 failed.

EDA RESULTS

USING SQL

```
%sql SELECT "Booster_Version", Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" =(SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

Tweleve Falcon9 boosters carried maximal amount of payload mass (15600 kg)

EDA RESULTS

USING SQL

```
%sql SELECT substr(Date, 6,2), substr(Date,0,5), "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL
                                                         WHERE substr(Date,0,5)='2015' AND "Landing_Outcome" = "Failure (drone ship)";
```

substr(Date, 6,2)	substr(Date,0,5)	Landing_Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

There were two failed landing outcomes with drone ship in 2015 in January and April. Both launched from CCAFS LC-40

```
%sql SELECT COUNT("Landing_Outcome") AS Total, "Landing_Outcome" FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (ground pad)" OR "Landing_Outcome" =
"Failure (drone ship)" AND DATE BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY "Landing_Outcome" ORDER BY Total DESC;
```

Total	Landing_Outcome
9	Success (ground pad)
5	Failure (drone ship)

Count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

EDA RESULTS

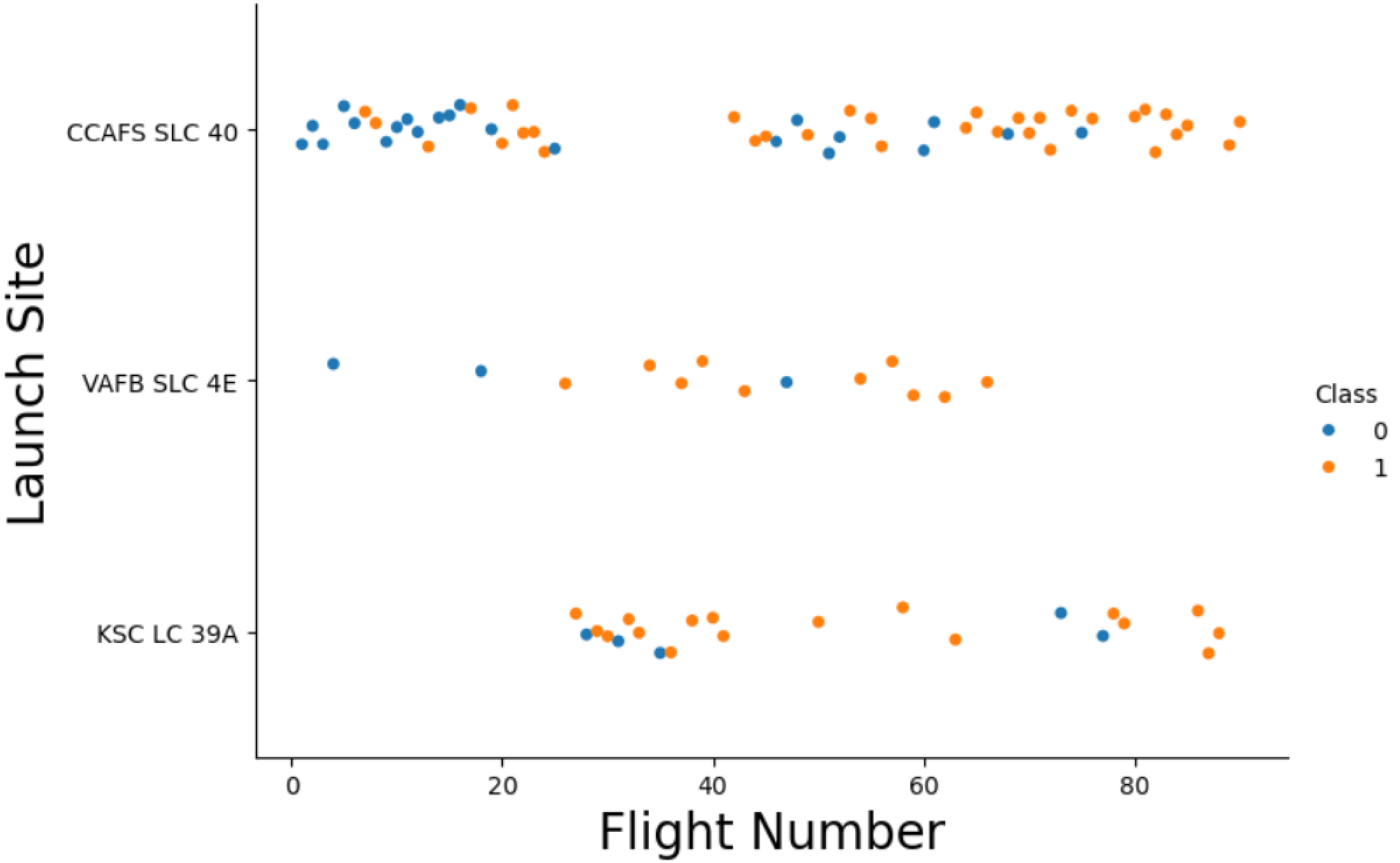
DATA VISUALIZATION

Flight Number vs Launch Site

From this scatter plot we can see that launch site CCAFS SLC 40 has the most number of launches.

Failed landings (Class 0) are often founded ar early flight attempts.

As flight number increases so does the succes rate (Class1). Success rate for VAFB SLC 4E launch site is 100% after 50th flight, and for KSC LC 39A and CCAFS SLC 40 – after 80th flight.



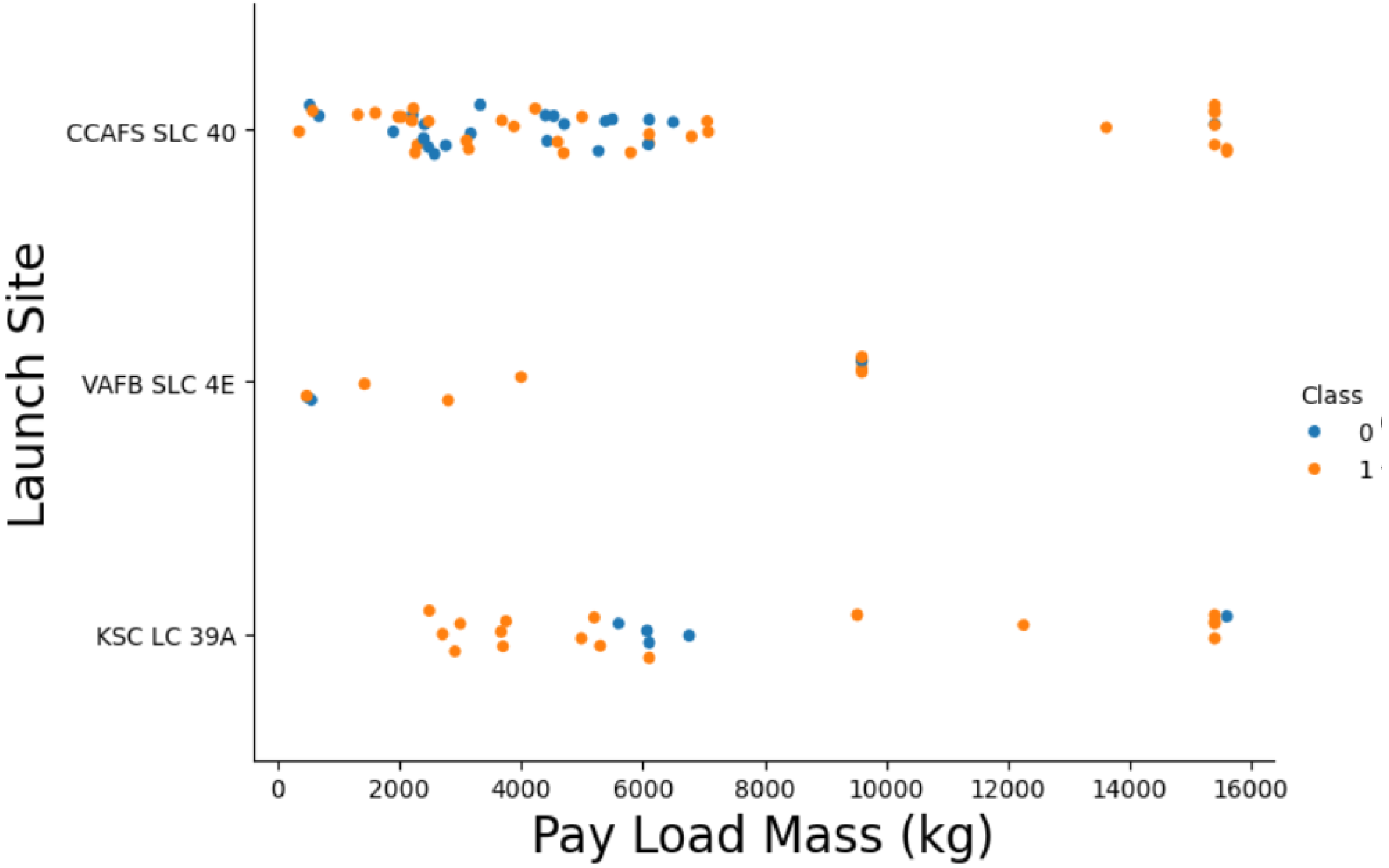
EDA RESULTS

DATA VISUALIZATION

Payload Mass vs Launch Site

Launch site VAFB SLC 4E doesn't have launch that have pay load mass greater than 10000 kg.

The failed landings (Class 0) at the KSC LC 39A launch site are mostly grouped around a narrow band of payload masses.



EDA RESULTS

DATA VISUALIZATION

Success rate of each orbit type

Low Earth orbits:

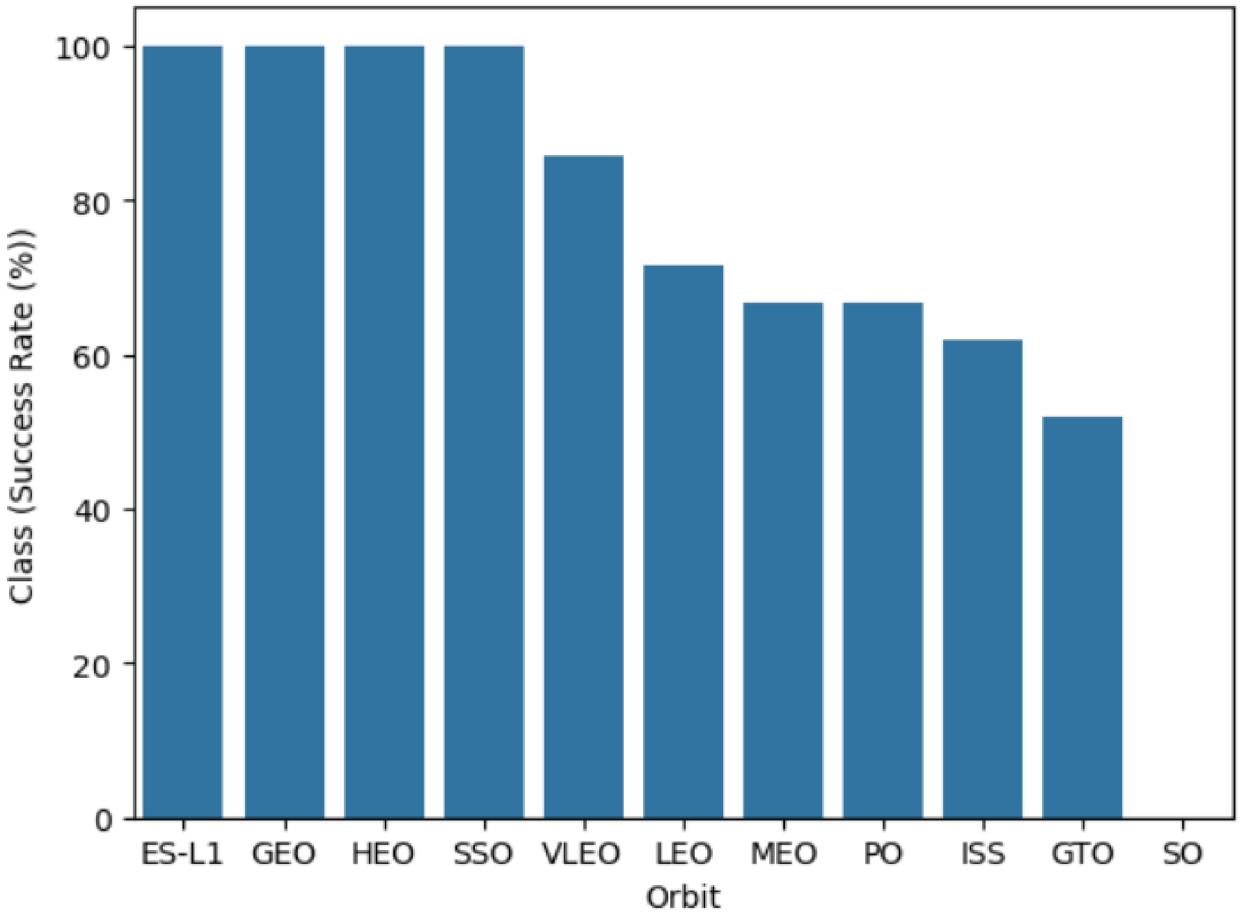
GTO, ISS, LEO, MEO, PO, VLEO

High Earth orbits:

ES-L1, GEO, HEO, SSO

High Earth orbits have the highest success rates at 100%.

SO orbit have no successful first stage landings.



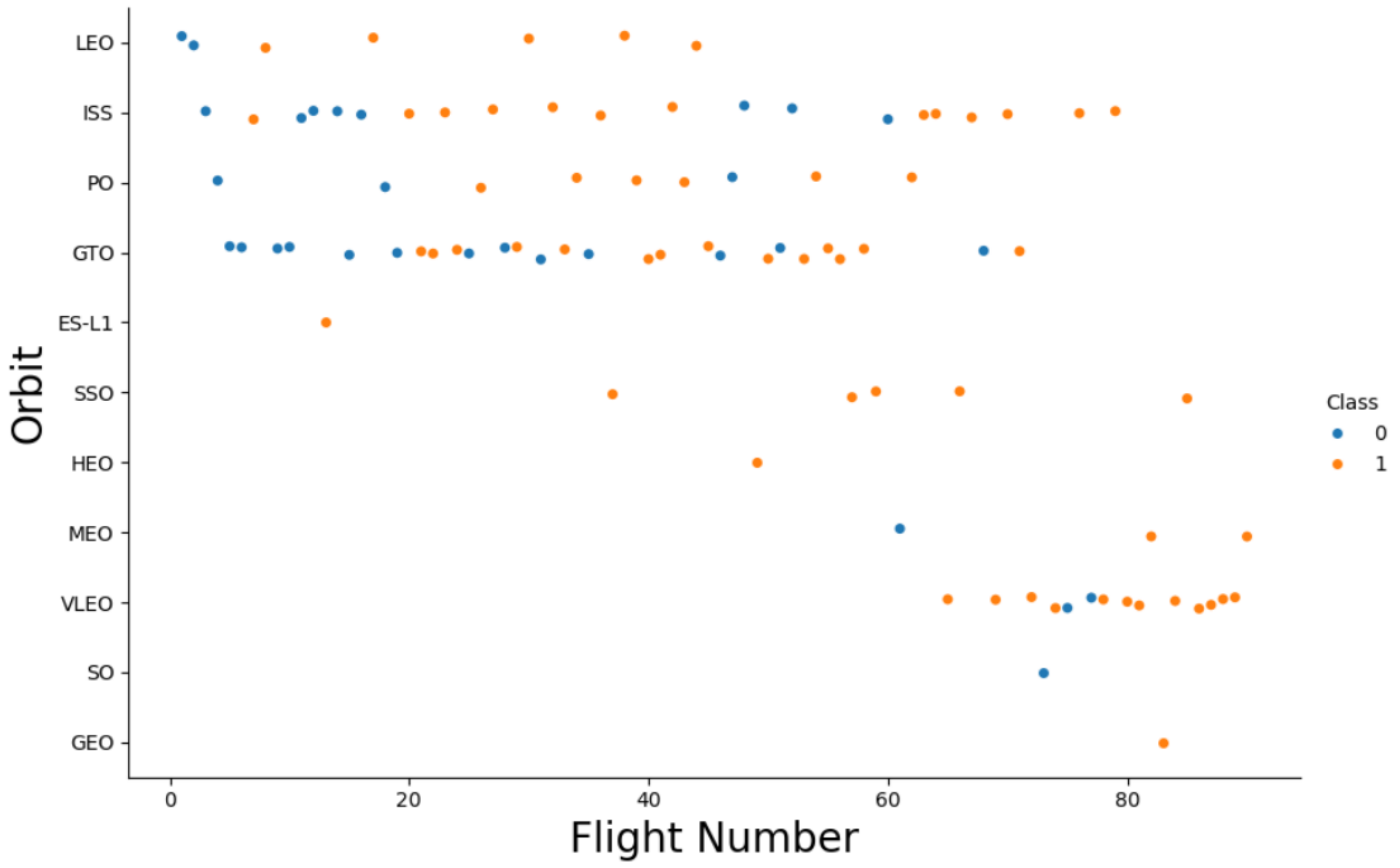
EDA RESULTS

DATA VISUALIZATION

FlightNumber vs Orbit type

In the LEO orbit, success (Class 1) 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.



EDA RESULTS

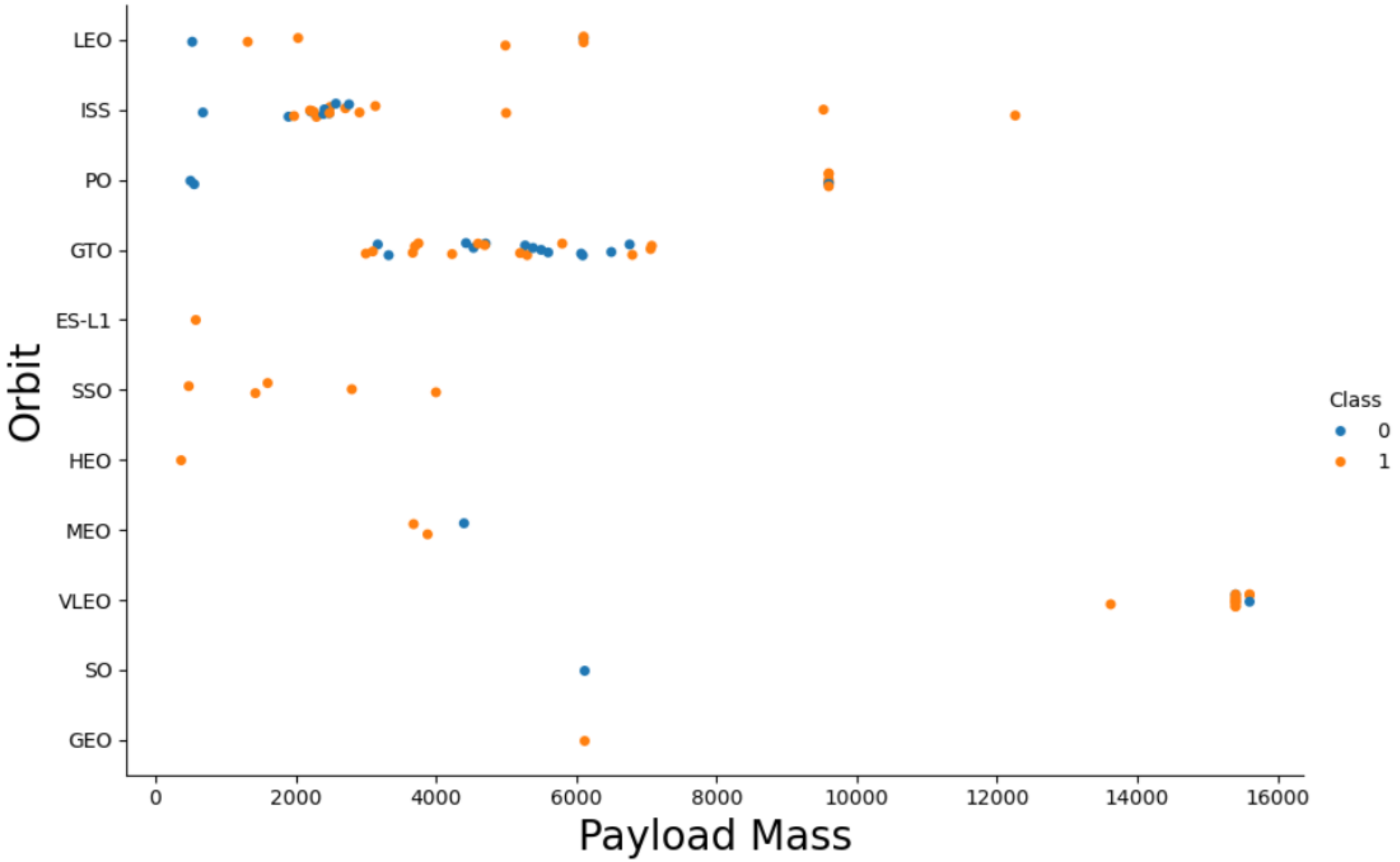
DATA VISUALIZATION

Payload Mass vs Orbit type

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

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

Most payload mass is on VLEO



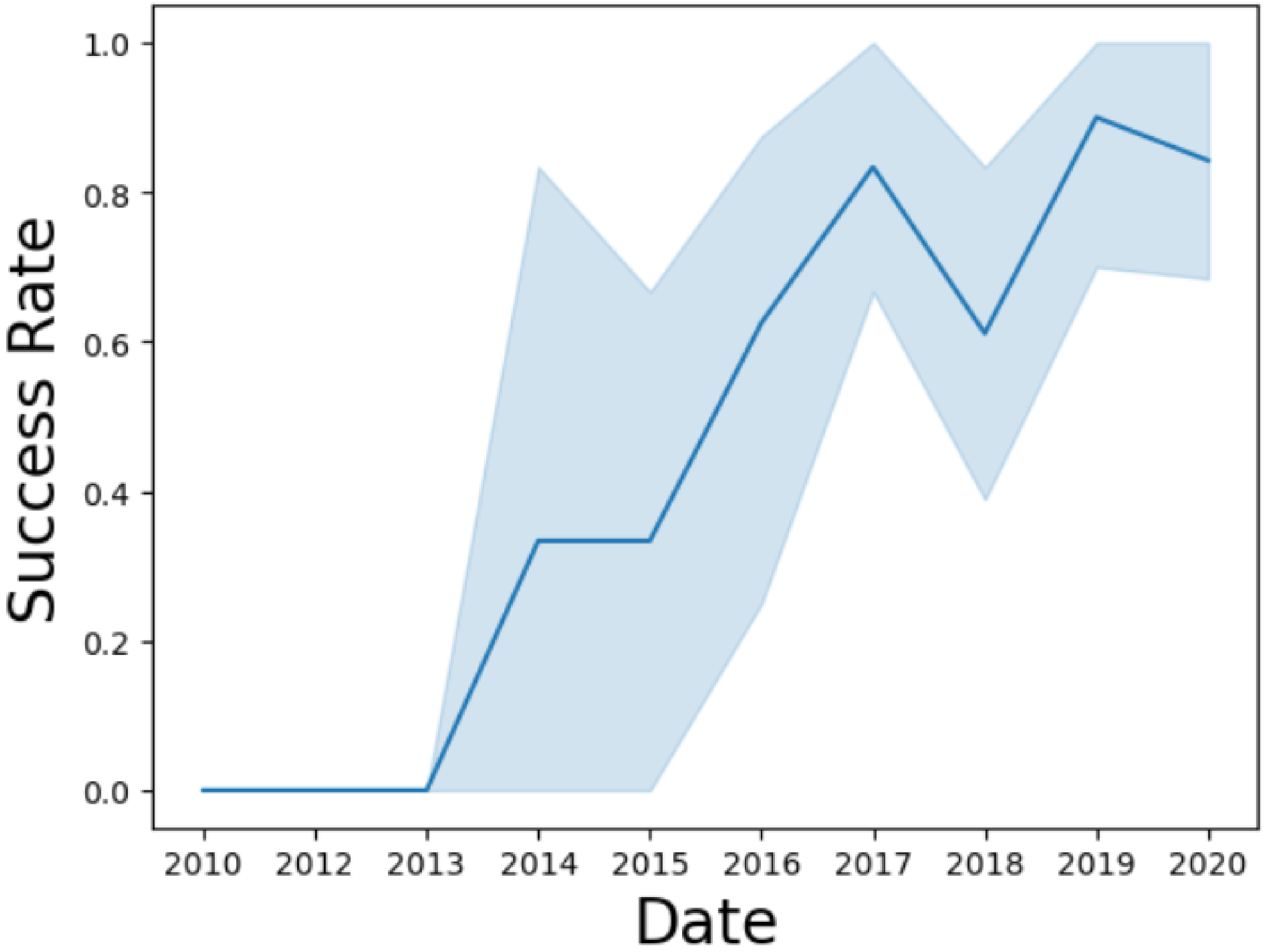
EDA RESULTS

DATA VISUALIZATION

Launch success yearly trend

Success rate since 2013 kept increasing till 2020.

There is 20% decrise in success rate in 2018



RESULTS

INTERACTIVE VISUAL ANALYTICS

Falcon 9 Launch Site Locations

<u>VAFB SLC-4E (California, USA):</u>

Vandenberg Air Force Base Space Launch Complex 4E

<u>KSC LC-39A (Florida, USA):</u>

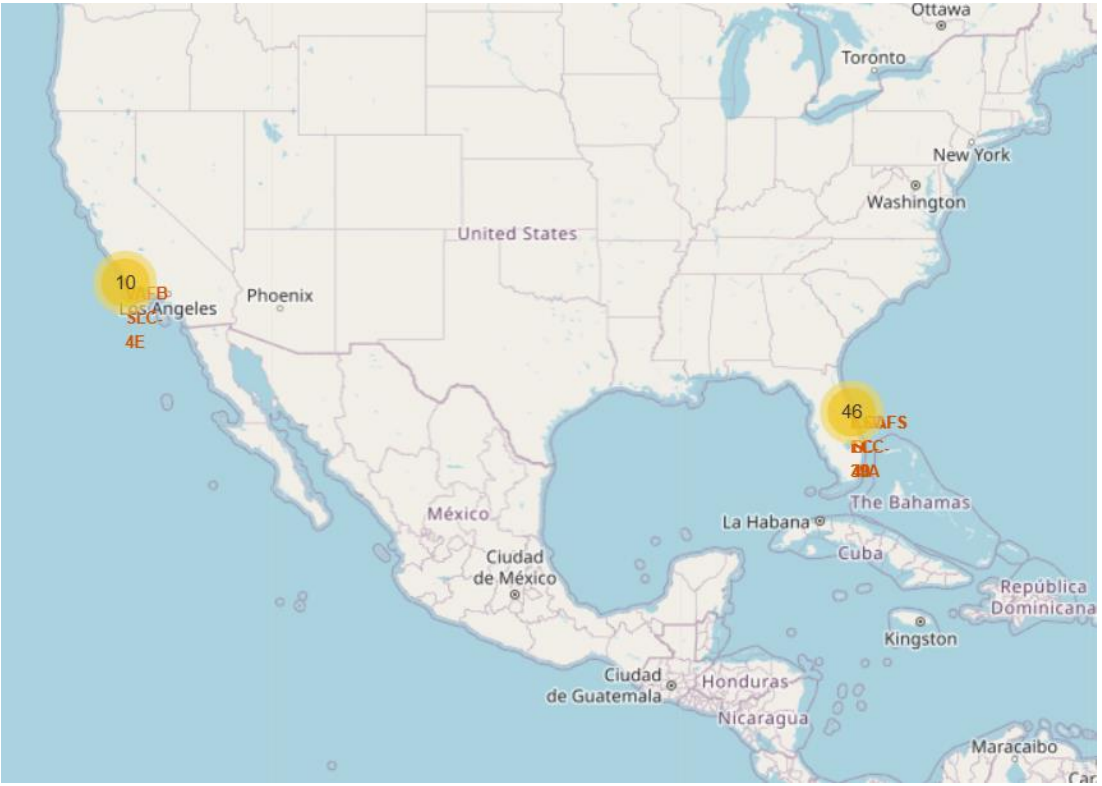
Kennedy Space Center Launch Complex 39A

<u>CCAFS LC-40 (Florida, USA):</u>

Cape Canaveral Air Force Station Launch Complex 40

<u>CCAFS SLC-40 (Florida, USA):</u>

Cape Canaveral Air Force Station Space Launch Complex 40

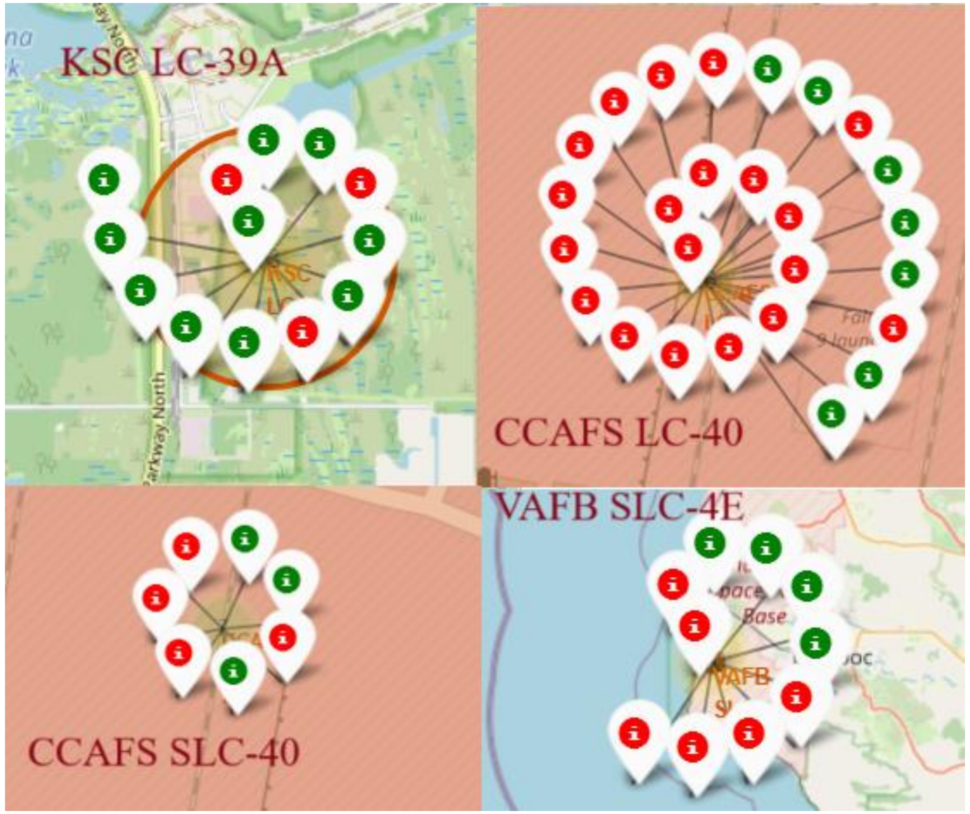


RESULTS

INTERACTIVE VISUAL ANALYTICS

Map Markers of Successful/Failed Landings

- Markers display mission outcomes for Falcon 9 first stage landings. Successful landings marked with green, failed – with red.
- Higher success rate can be seen for KSC LC-39A (76.92%).
- For CCAFS SLC-40 success rate is 42.85%, for VAFB SLC-4E – 40%, and lowest rate – 26.92% - for CCAFS LC-40



RESULTS

INTERACTIVE VISUAL ANALYTICS

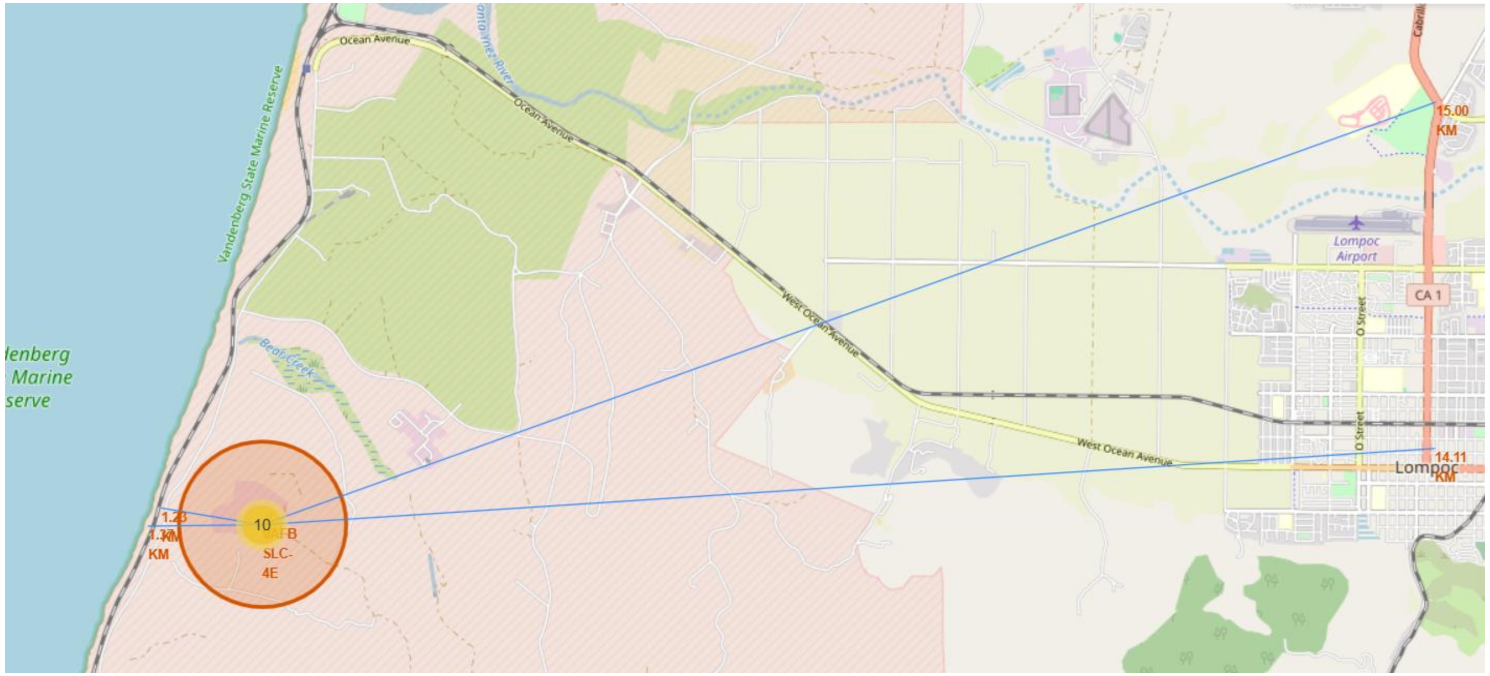
Distances between a launch site to its proximities

- CCAFS LC-40 and CCAFS SLC-40 launch sites are very close to each other, so all proximities was shown for both of them together.
- CCAFS LC-40, CCAFS SLC-40 and KSC LC-39A are relatively far from inhabited areas.
- KSC LC-39A distances to closest:
    - coast line – 7.2 km
    - railway – 15.2 km
    - highway – 14.9 km
    - town – 56.6 km
- CCAFS LC-40 and CCAFS SLC-40 distances to closest:
    - coast line – 0.9 km
    - railway – 22 km
    - highway – 21.6 km
    - town – 55.9 km

RESULTS
INTERACTIVE VISUAL ANALYTICS

Distances between a launch site to its proximities

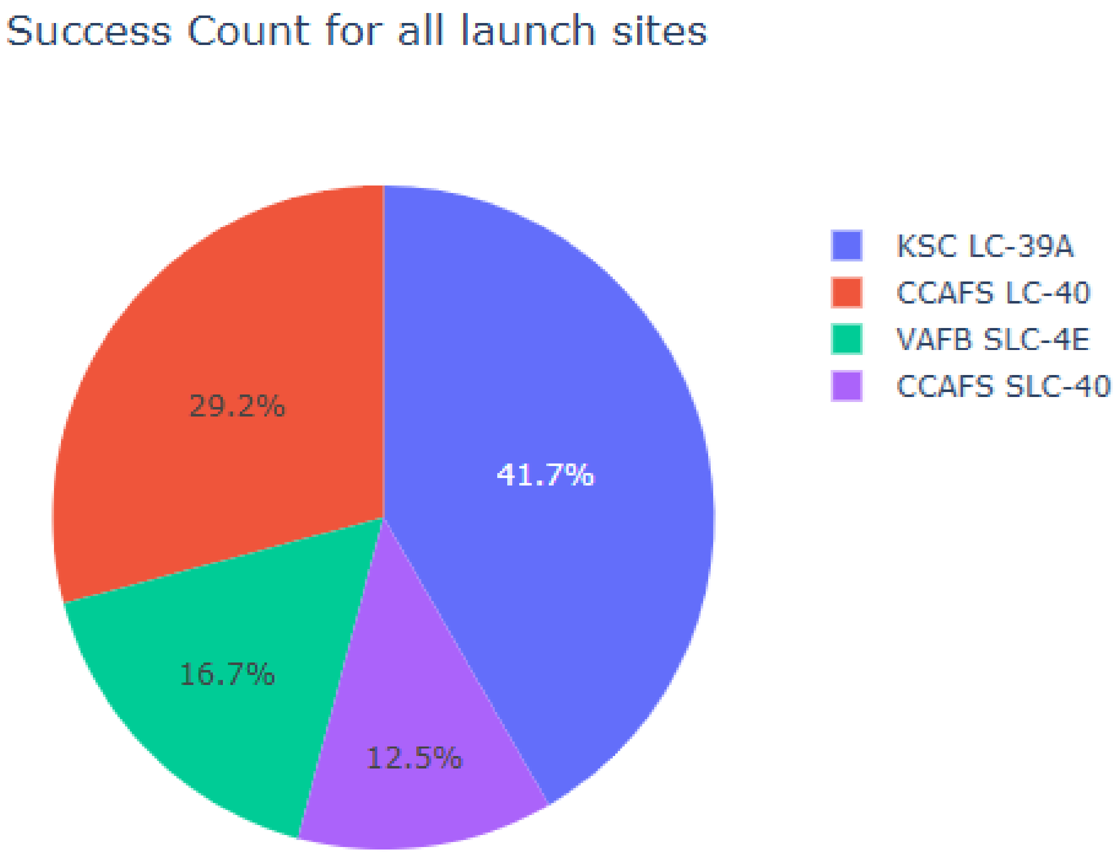
- VAFB SLC-4E distances to closest:
  - coast line – 1.3 km
  - railway – 1.2 km
  - highway – 15 km
  - town – 14.1 km
- VAFB SLC-4E is closer to inhabitate areas
- All launch site have good logistic aspects



RESULTS

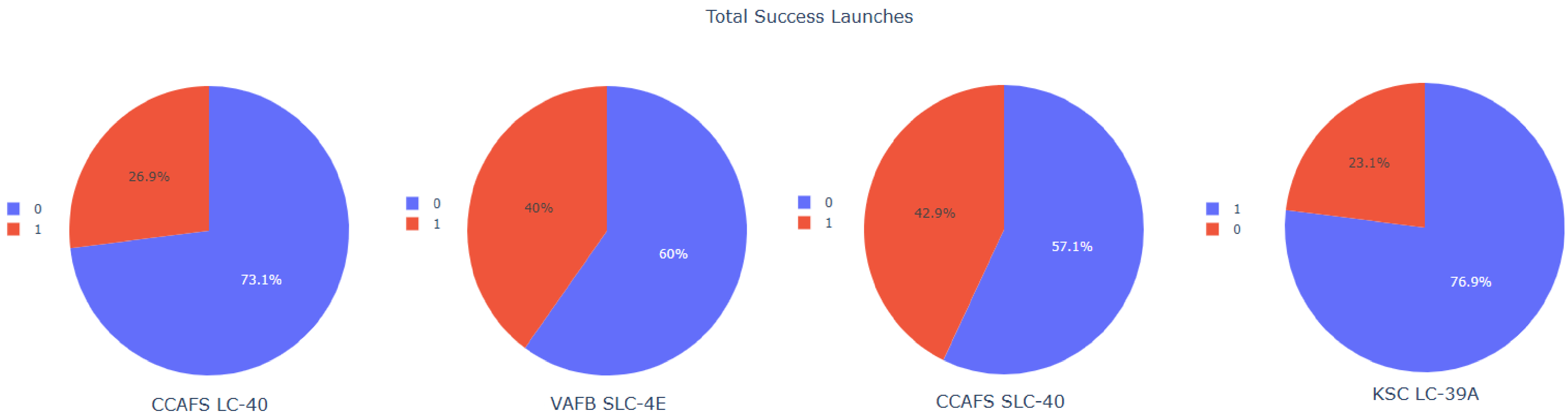
INTERACTIVE DASHBOARD

- Launch site KSC LC-39A has highest launch success rate, followed by CCAFS LC-40, VAFB SLC-4E and CCAFS SLC-40.



RESULTS

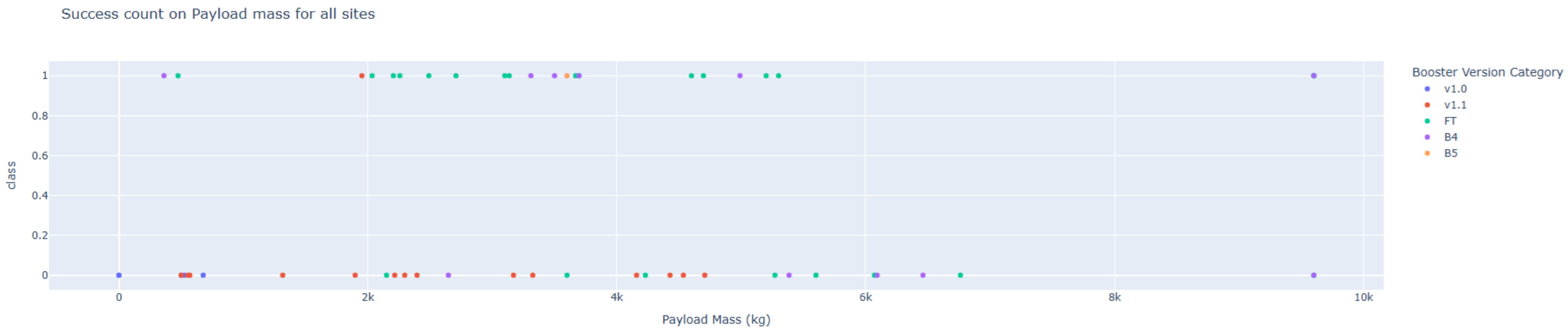
INTERACTIVE DASHBOARD



- KSC LC-39A launch site has highest success rate – 76.9%.
- Please, pay attention, that for first three pie charts success is shown in red color, but for last one – in blue.

RESULTS

INTERACTIVE DASHBOARD



Payload vs Launch Outcome

- Most successful outcomes has payload range about 2000-5000 kg
- Payload under 6000 kg and FT boosters are the most successful combination
- There is not enough data to estimate risk of launches over 7000 kg

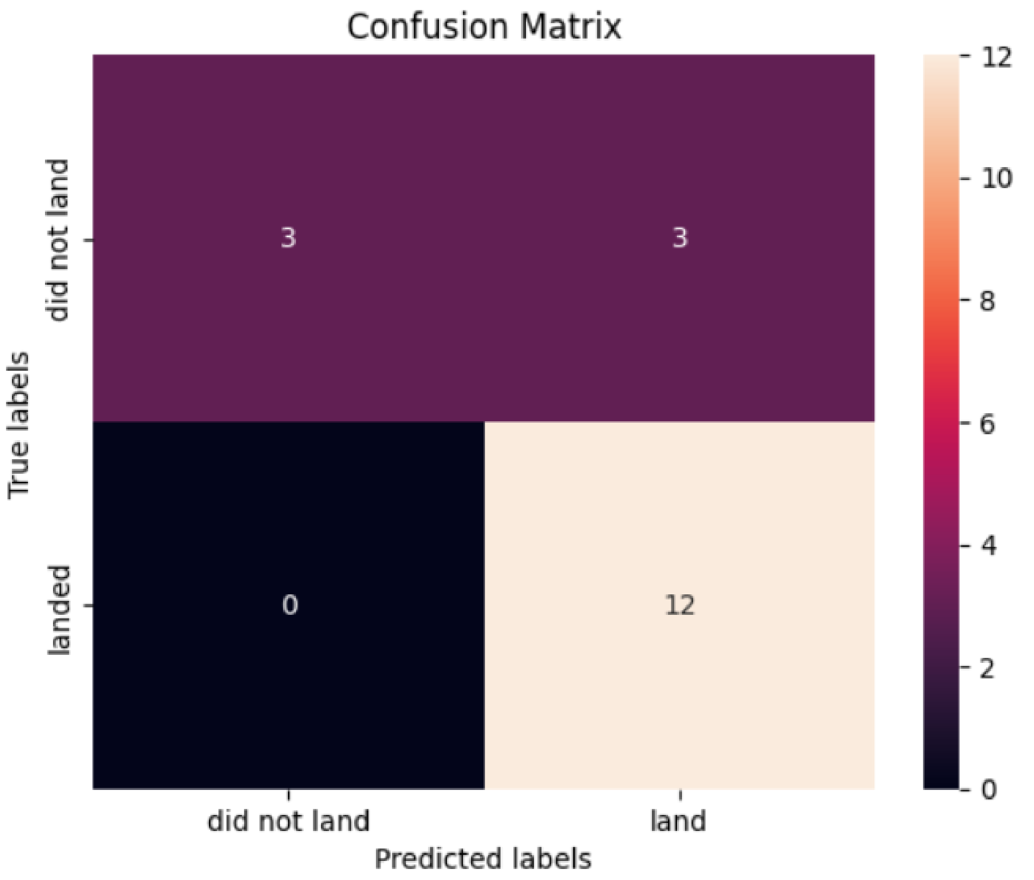
RESULTS

PREDICTIVE ANALYSIS

All methods perform equally on the test data, they all have same accuracy of 0.83 on the test data.

All four classification models have the same confusion matrixes and were able equally distinguish between different classes.

For all models false positive predictions are the major problem.



Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

CONCLUSION

- Falcon 9 first stage landing outcomes have been trending toward greater success as more launches are made.
- Most successful launches are made on KSC LC-39A launch site.
- All launch sites have good logistics and are located relatively far away from inhabitate areas.
- Most successful launches are with payload mass in range from 2000 to 5000 kg.
- FT booster has larges success rate.
- High Eatrh orbits have highest success rate.
- Although for some orbits mission outcome correlates with launch numbers (for LEO orbit for example) or payload mass (for Polar, LEO, ISS orbits successful outcome related with heavy payloads), for GTO orbit we can't find relationship between landing outcome and flight number or payload mass.
- So flight number, payload mass, orbit, booster version and launch site influence successful landing outcome and are important features that should be used in success prediction.
- All four classification models performed equally on the test data and have problem only with false positive predictions.

THANK YOU!