IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anand Kramer
1/7/24

Outline

Executive Summary

The commercial space age is here, for that reason our company Space Y was born. Space Y wants to make space travel affordable for all.

Introduction

Space Y is here to compete in the commercial space race. We are making rocket launchers relatively inexpensive for everyone.

Space Y can save millions in every launch of our Eagle rocket because we can reuse it's first stage.

In addition, we can determine if the first stage of our competitor will land and determine the cost of a launch by using data science and machine learning models.

Section 1

Methodology

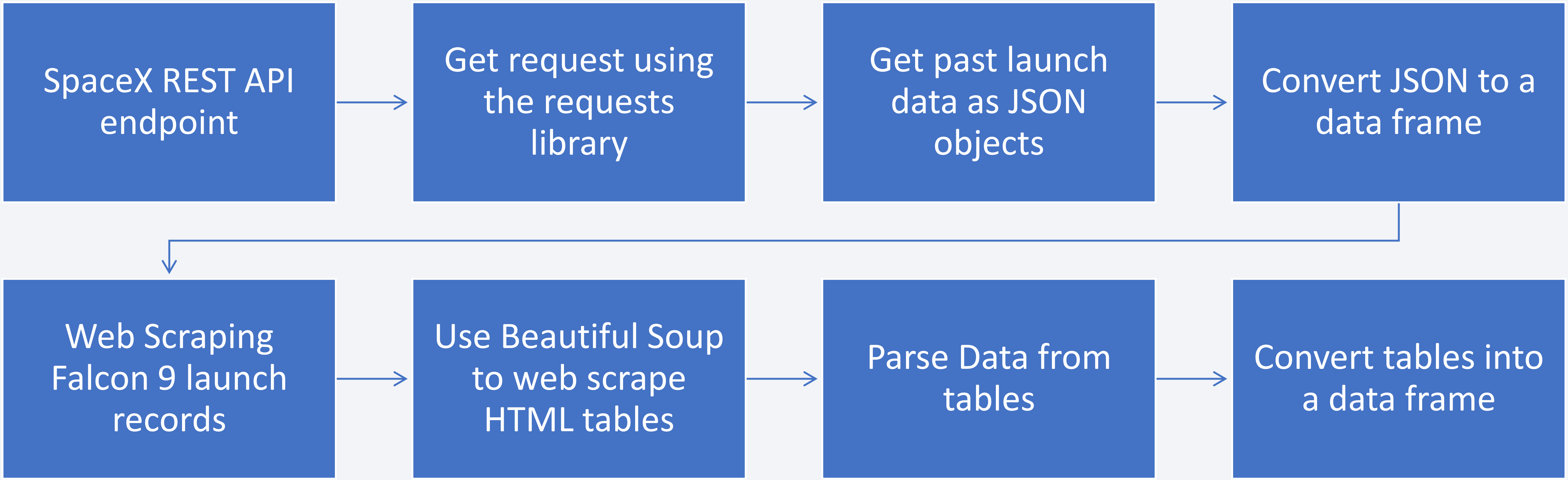
Methodology

Executive Summary

- Data collection methodology:
  - Data was gathered from SpaceX REST API and web scraping from wiki pages.
- Perform data wrangling
  - Data collected in JSON form object and HTML tables which were converted into a Pandas dataframe for visualization and analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use machine learning to determine if the first stage of Falcon 9 will land successfully.

Data Collection

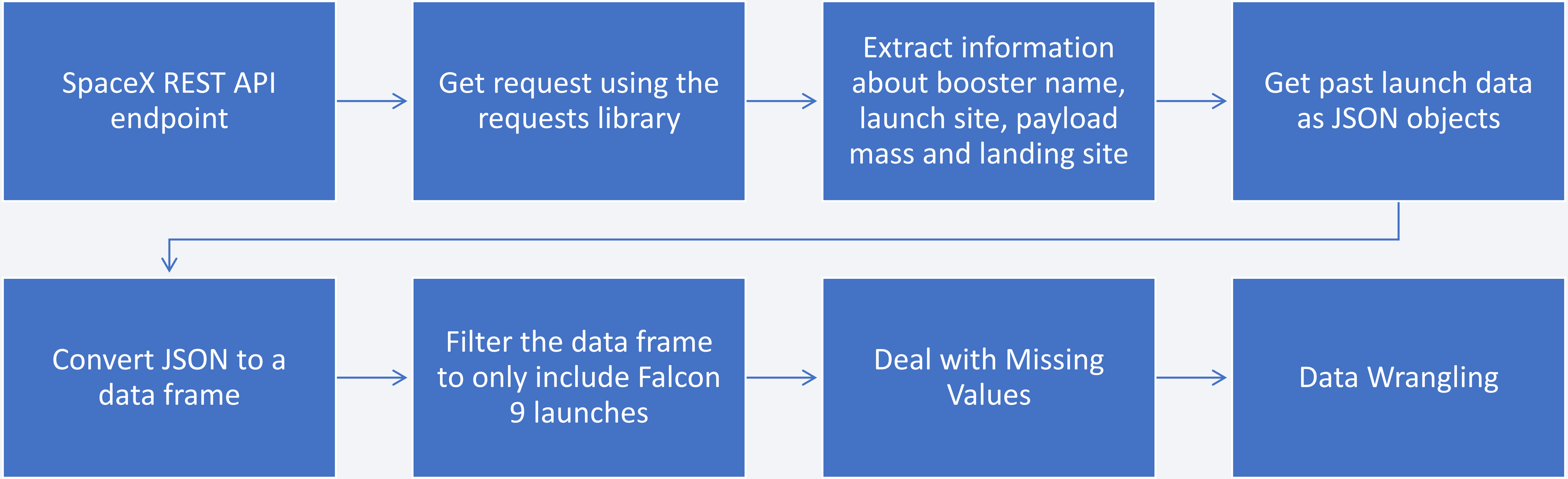
The data was gathered from the SpaceX REST API and web scraped from wiki pages



https://github.com/pinnaclewealth/capstone-space-Y.git

Data Collection – SpaceX API

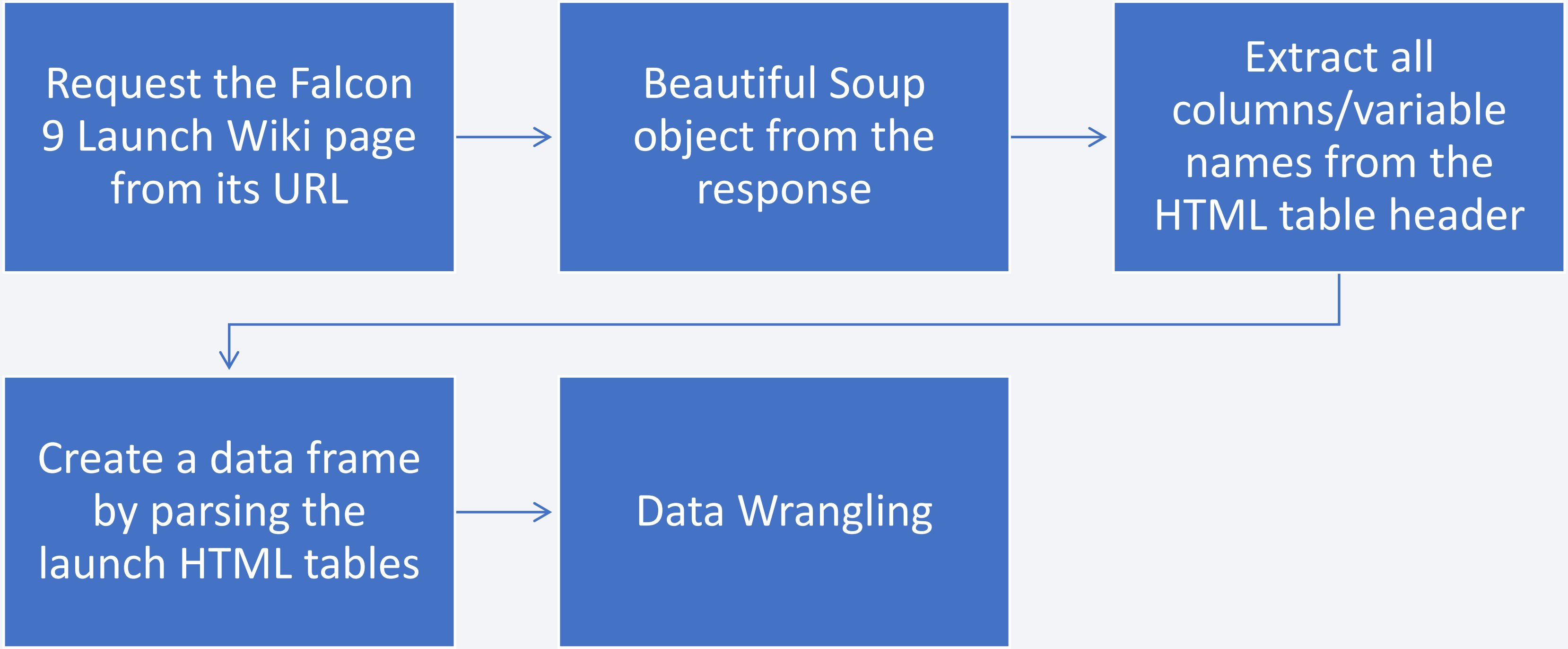
Collect and make sure the data is in the correct format from an API



https://github.com/pinnaclewealth/capstone-space-Y.git

Data Collection - Scraping

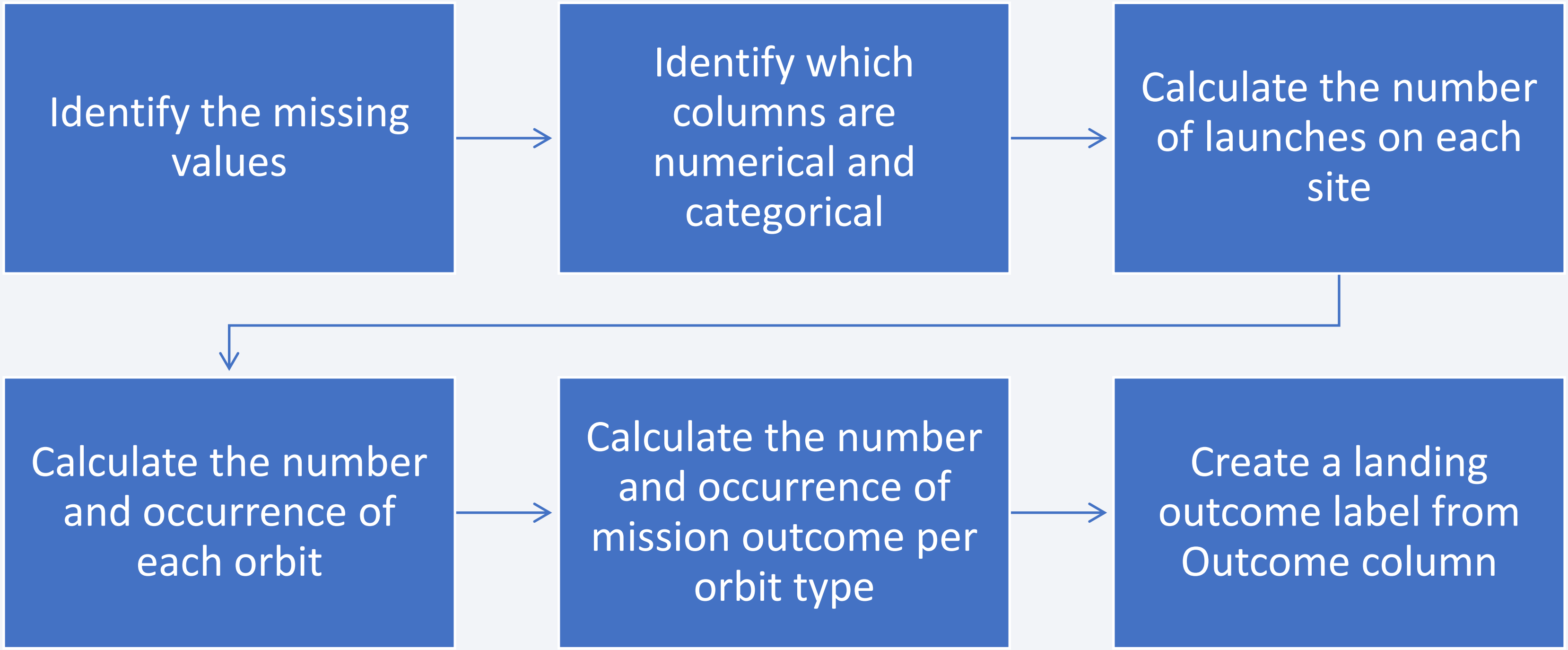
Perform web scraping to collect Falcon 9 historical launch records from Wikipedia page.



https://github.com/pinnaclewealth/capstone-space-Y.git

Data Wrangling

Perform data wrangling to obtain data from web scraping and API connections.



https://github.com/pinnaclewealth/capstone-space-Y.git

EDA with Data Visualization

Perform Exploratory Data Analysis (EDA) to find patterns in the data and determine what would be the label for supervised models.

Summary Of charts that were plotted:

- Catplot to visualize the relationship between Flight Number and Payload
- Catplot to visualize the relationship between Flight Number and Launch
- Catplot to visualize the relationship between Payload and Launch Site
- Barchart to visualize the relationship between success rate and Orbit type
- Catplot to visualize the relationship between Flight Number and Orbit type
- Catplot to visualize th
- Line chart to visualize the launch success yearly trend

EDA with SQL

SQL queries performed:

- Display the names of the unique launch sites in the space mission:
  - SELECT DISTINCT(launch_site) FROM SPACEXTBL:
- Display 5 records where launch sites begin with the string ‘CCA’:
  - SELECT * FROM SPACEXTBL WHERE launch_site LIKE ‘CCA%’ LIMIT 5;
- Display the total payload mass carried by boosters launched by NASA (CRS):
  - SELECT SUM(payload_mass_kg) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE customer = ‘NASA (CRS)’;
- Display average payload mass carried by booster version F9 v1.1:
  - SELECT AVG(payload_mass_kg) AS AVG_PAYLOAD_MASS FROM SPACEXTBL WHERE booster_version = ‘F9 v1.1’;
- List the data when the first successful landing outcome in ground pad was achieved:
  - SELECT MIN(DATA) AS first_successful_landing FROM SPACEXTBL WHERE (landing_outcome) = ‘Success(ground pad)’;

Build an Interactive Map with Folium

Summary of map objects that were created and added to the Folium map

- folium.Circle and folium.Marker to add a highlighted circle area with a text label on a specific coordinate for each launch site on the site map.
- MarkerCluster object to simplify a map containing many markers having the same coordinate
- MousePosition on the map to get coordinate for a mouse over a point on the map.
- Folium.PolyLine to draw a line between a launch site to its closest city, railway and highway

Build a Dashboard with Plotly Dash

This dashboard application contains input components sucha s a dropdown list and range slider to interact to perform interactive visual analytics on SpaceX launch data in real-time.

- A launch Site Drop-down Input Component with 4 different launch sites and a dropdown menu to select different launch sites.
- Callback function to render success-pie-chart based on the selected dropdown.
- Range slider to select payload for visual patterns

Predictive Analysis (Classification)

Predictive Analysis allows for building, evaluating, improving, and identifying the best performing classification model.

- Creation of a NumPy array from the column Class in data.
- Data standardization
- Use of function train_test_split to split the data X and Y into training and test data
- Search for hyperparameters for logistic regression, SVM, Decision Tree, and KNN classifiers
- Searching for the method that performs best using data

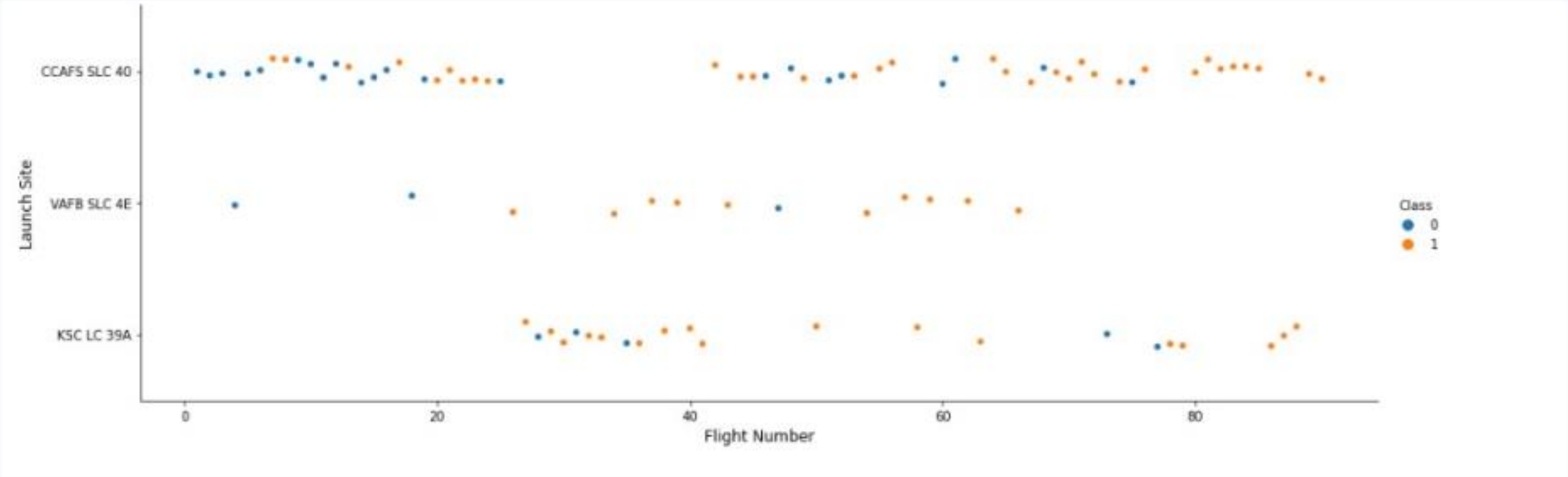
Results

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

Section 2

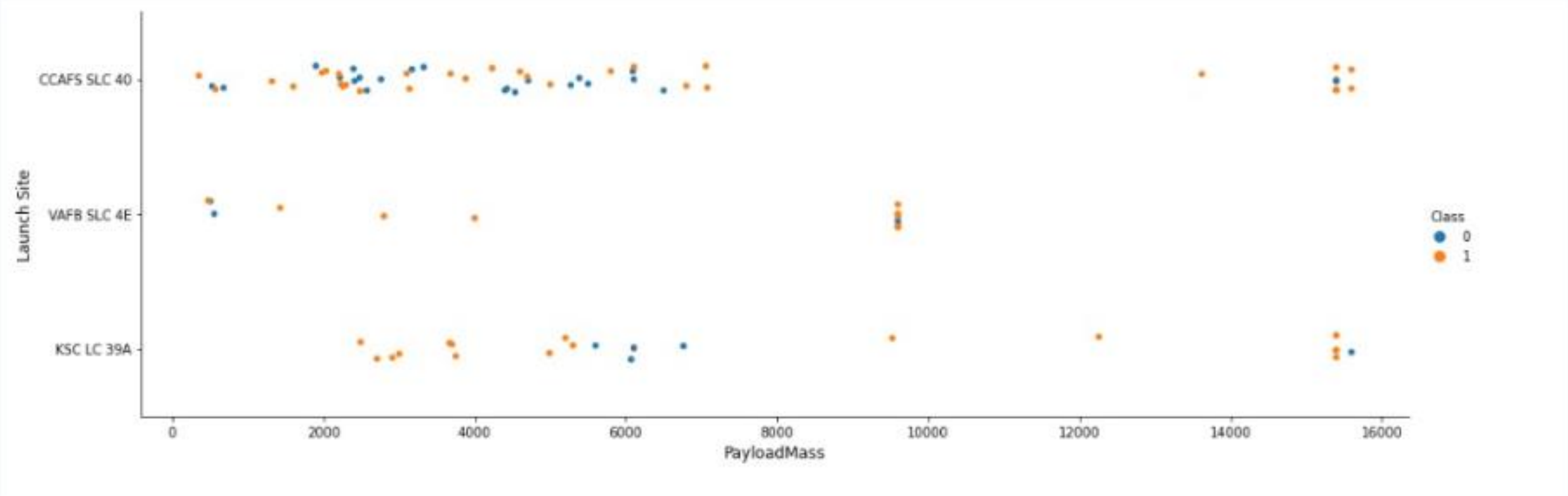
Insights drawn from EDA

Flight Number vs. Launch Site



- With time the successful rate has increased for every Launch Site, especially for CCAFS SLC 40, where are concentrated the majority of the launches.
- VAFB SLC 4E and KSC LC 39A has a higher successful rate but represents one third of the total launches.

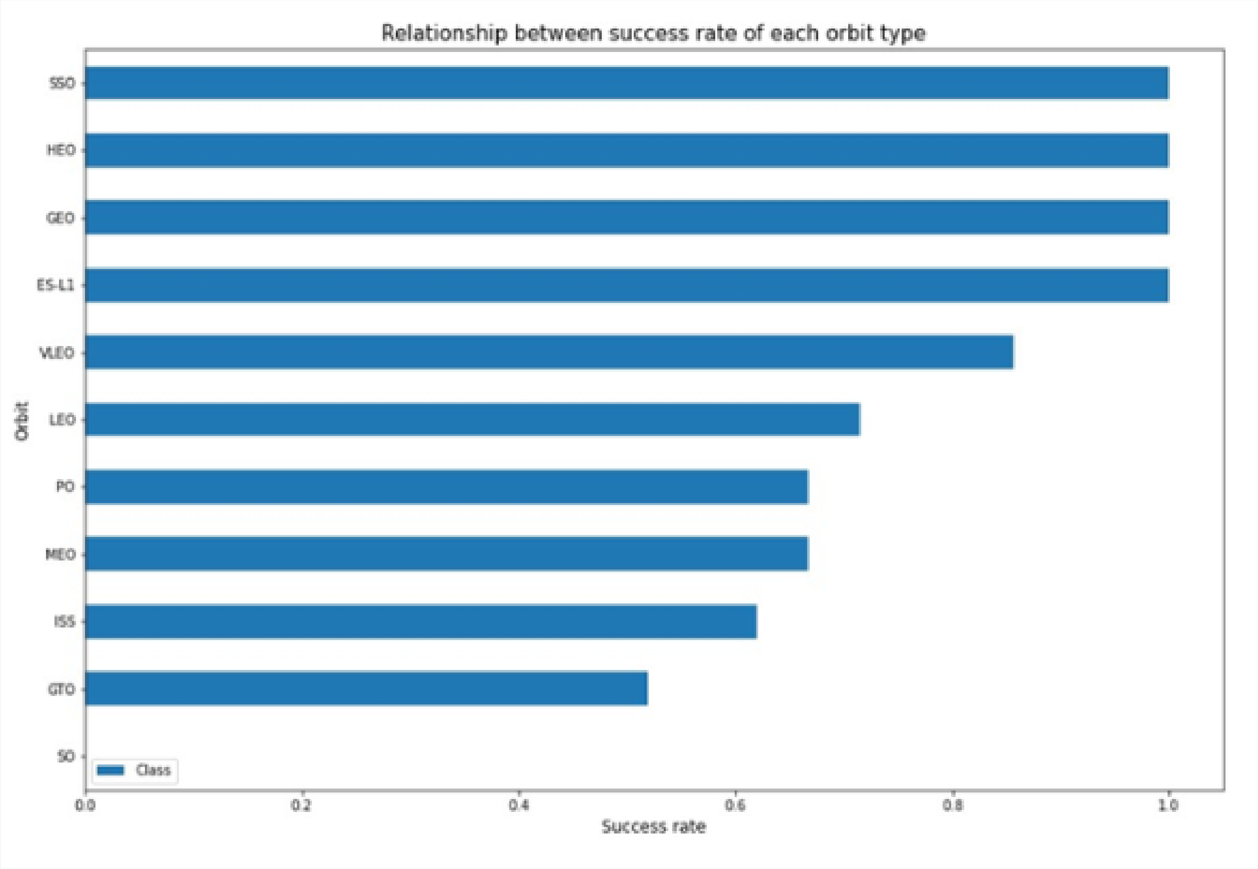
Payload vs. Launch Site



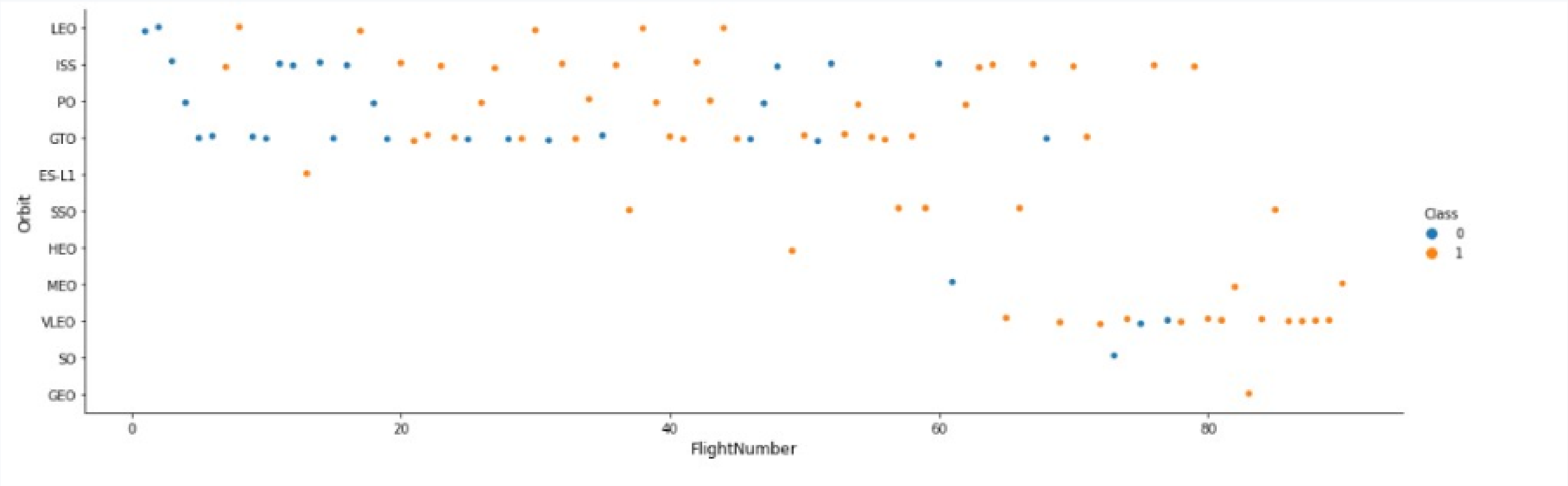
- In VAFB-SLC launch site there are no rockets launched for heavy payloadmass (greater than 10000 kg).
- In KSC LC launch site there are no rockets launched for lower payloadmass (less than 2500kg).
- CCAFS SLC has launched rockets less than 7500kg and more than 13000kg payloadmass but not in between.

Success Rate vs. Orbit Type

- The first 4 Orbit types has the best successful rate. But how many attempts are per orbit type?
- The bar chart must be interpreted with the number of launches per orbit type.

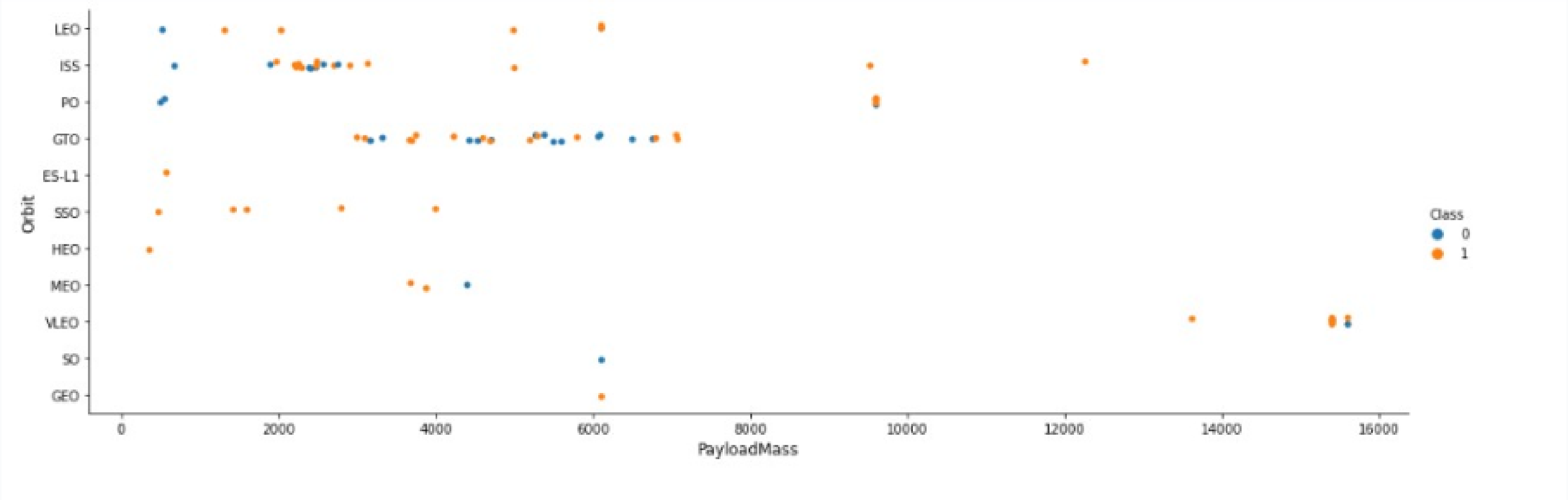


Flight Number vs. Orbit Type



- As expected, there are more failures at the beginning of the series of launches, but, after the first 40 launches, the ratio improves by reducing the 50 percent of unsuccessful landings.
- GTO and ISS orbits has the higher concentration of launches with the lowest ratio of successful landings.
- The orbits with higher successful rate, has one or just a few number of launches.

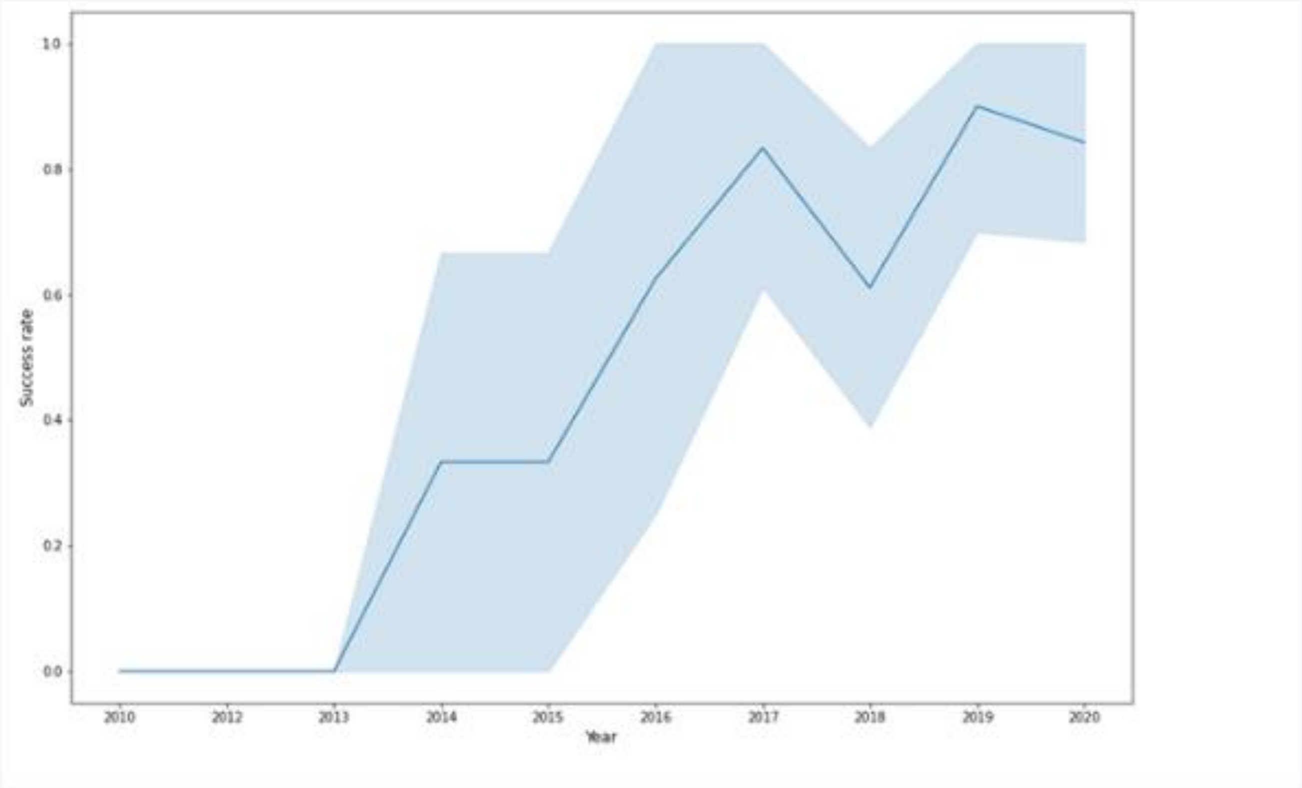
Payload vs. Orbit Type



- Exists a visible limit of Payload around 7600 kg. Less than 10 launches exceed that limit.
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- However for GTO, we cannot distinguish this well as both, positive landing rate and negative landing are both there here.

Launch Success Yearly Trend

- The success rate since 2013 kept increasing until 2020.



All Launch Site Names

- The four unique launch sites in the space mission.
- I have used "DISTINCT" statement to find the unique values in the launch site column.

```
%sql SELECT DISTINCT(launch_site) FROM SPACEXTBL;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa3
Done.
```

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with the string 'CCA'. The query uses WHERE, LIKE and LIMIT.

```
%sql SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 total payload mass carried by boosters launched by NASA (CRS) using SUM function and WHERE clause.

```
%sql SELECT SUM(payload_mass__kg_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE customer='NASA (CRS)';
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdom
Done.
```

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 using AVG() function.

```
%sql SELECT AVG(payload_mass__kg_) AS AVG_PAYLOAD_MASS FROM SPACEXTBL WHERE booster_version='F9 v1.1';
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdoma
Done.
```

avg_payload_mass
2928

First Successful Ground Landing Date

- The date when the first successful landing outcome in ground pad was achieved using MIN function.

```
%sql SELECT MIN(DATE) AS first_successful_landing FROM SPACEXTBL WHERE (landing_outcome)='Success (ground pad)';
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3
Done.
```

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000, combining WHERE clause with AND operator.

```
%sql SELECT booster_version, payload_mass__kg_, landing_outcome FROM SPACEXTBL \
      WHERE landing_outcome='Success (drone ship)' AND (payload_mass__kg_ BETWEEN 4000 AND 6000) ;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdc
Done.
```

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)

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes. The query uses a combination of COUNT function with GROUP BY statement.

```
%sql SELECT mission_outcome, COUNT(mission_outcome) AS TOTAL FROM SPACEXTBL GROUP BY mission_outcome;
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdom
Done.
```

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster_versions which have carried the maximum payload mass. Using a subquery.

```
%sql SELECT DISTINCT(booster_version), (SELECT MAX(payload_mass__kg_) AS "maximum_payload_mass" FROM SPACEXTBL) FROM SPACEXTBL
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

booster_version	maximum_payload_mass
F9 B4 B1039.2	15600
F9 B4 B1040.2	15600
F9 B4 B1041.2	15600
F9 B4 B1043.2	15600
F9 B4 B1039.1	15600

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT landing_outcome, booster_version, launch_site, DATE FROM SPACEXTBL WHERE landing_outcome LIKE '%Failure (drone ship)%' 
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

landing_outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

- 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. The query uses COUNT, WHERE, BETWEEN and GROUP BY.

```
%sql SELECT landing_outcome, COUNT(landing_outcome) AS "total" FROM SPACEXTBL WHERE (DATE BETWEEN '2010-06-04' AND '2017-03-20')
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

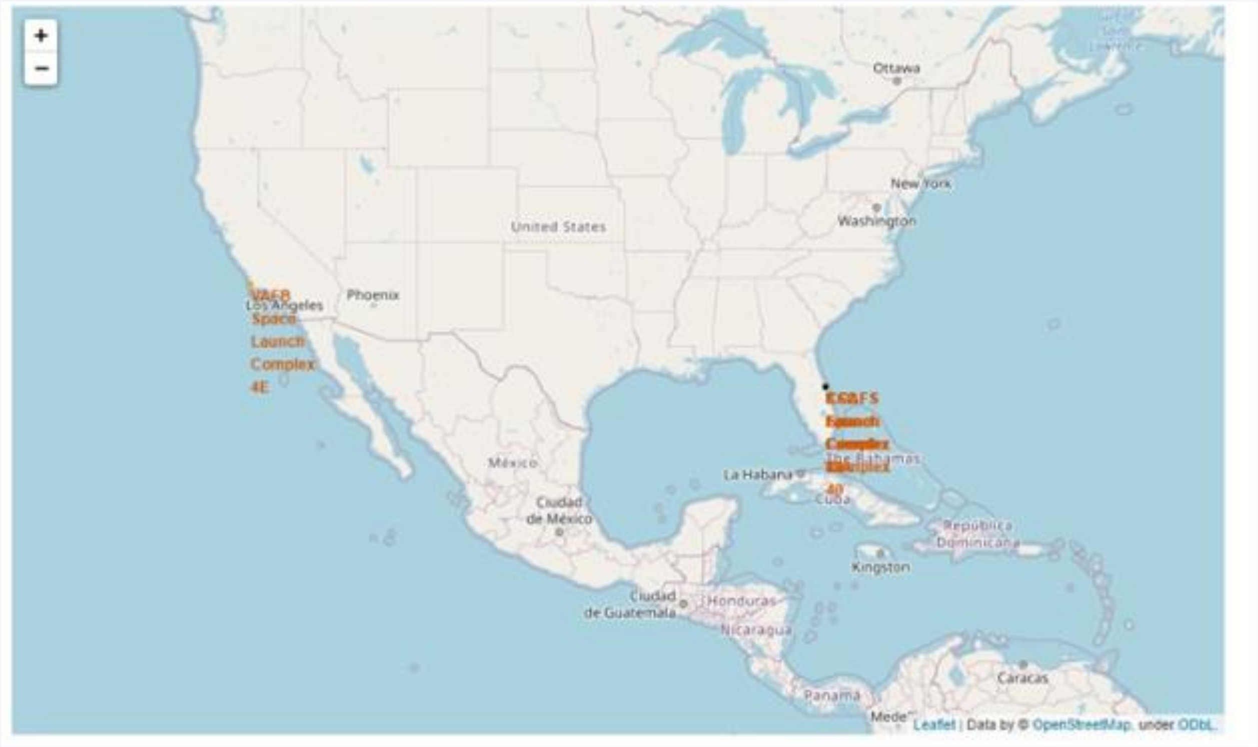
landing_outcome	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

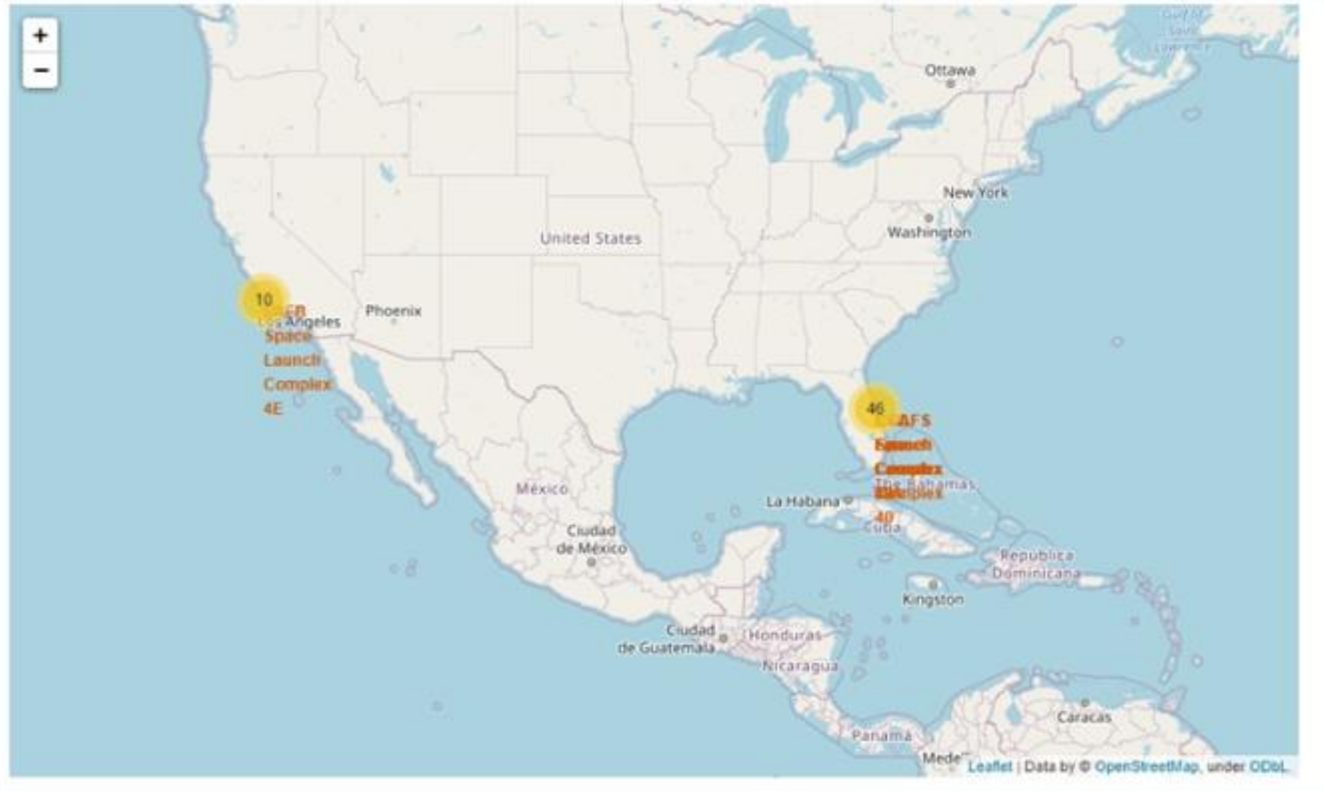
All Launch Sites

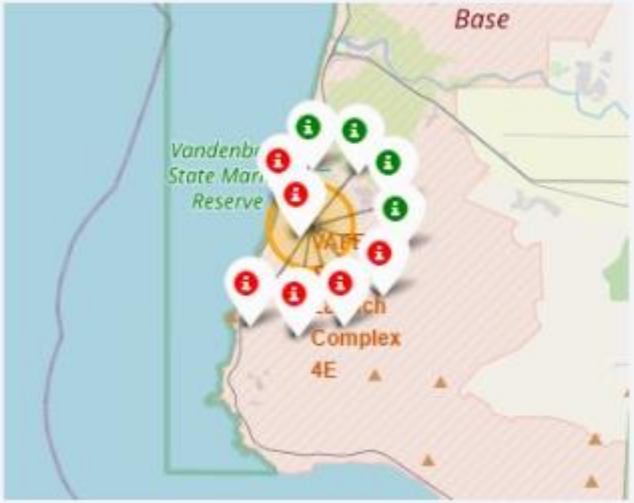
All launch sites are in very close proximity to the coast and in restricted areas.



Success/Failed Launches for Each Site

First map shows clusters for each launch site while the second shows a green marker for successful launches and a red marker for failed launches.





Launch Site and its Proximities

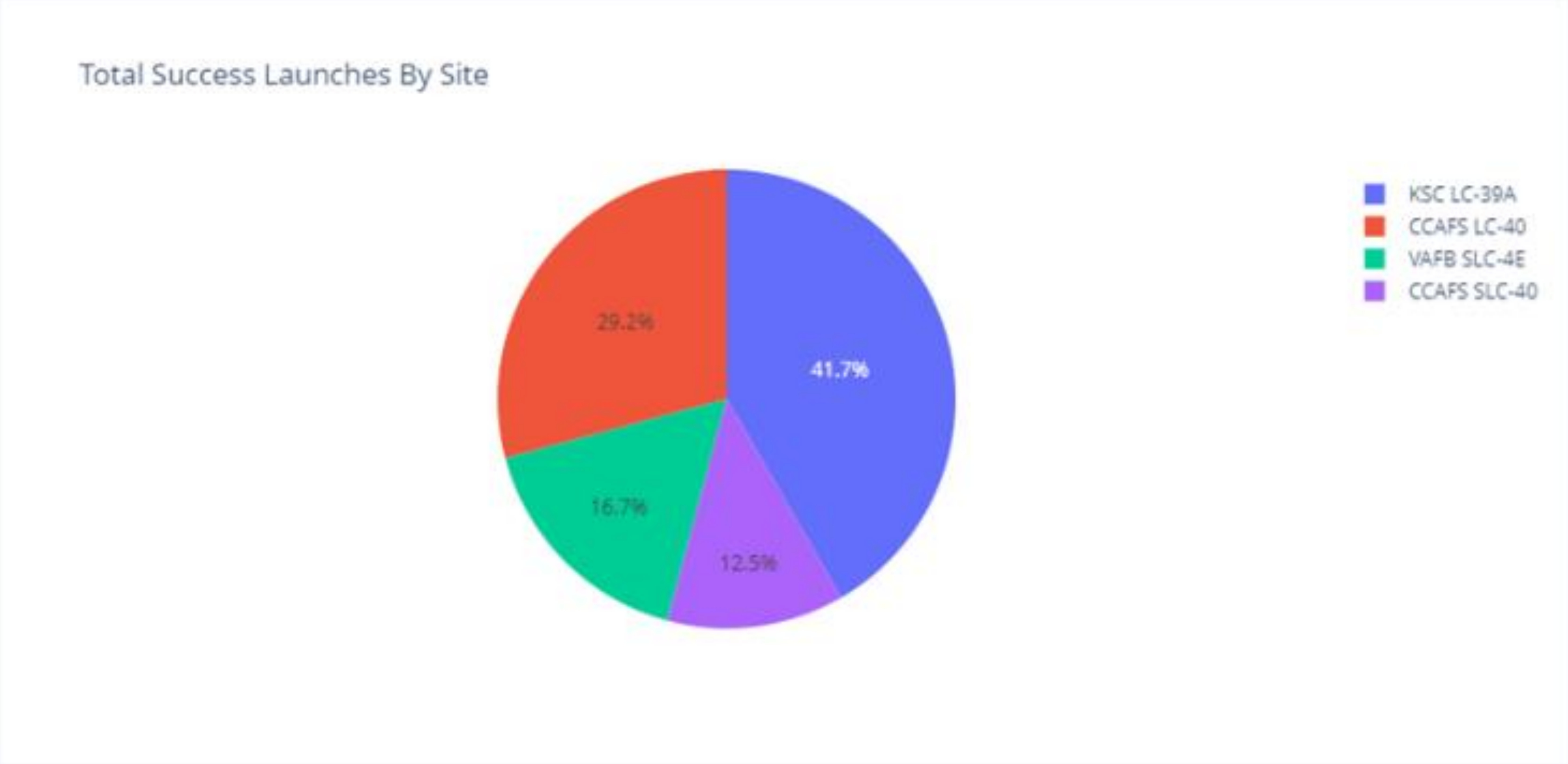
Launch sites are near railways, roads, highways and coastlines. Maintaining a safe distance to cities is crucial for success.

Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site

KSC LC-39A is the site with the higher success launches followed by CCAFS LC-40.



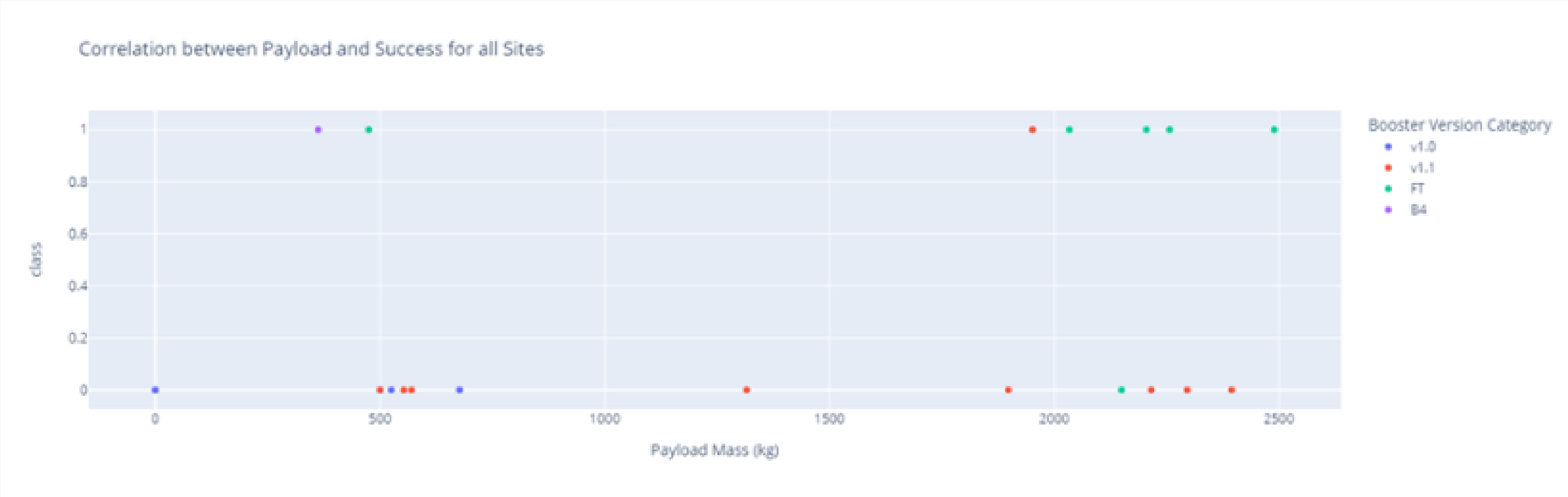
KSC LC-39A

The piechart for the launch site KSC LC-39A shows the site with highest launch success ratio.

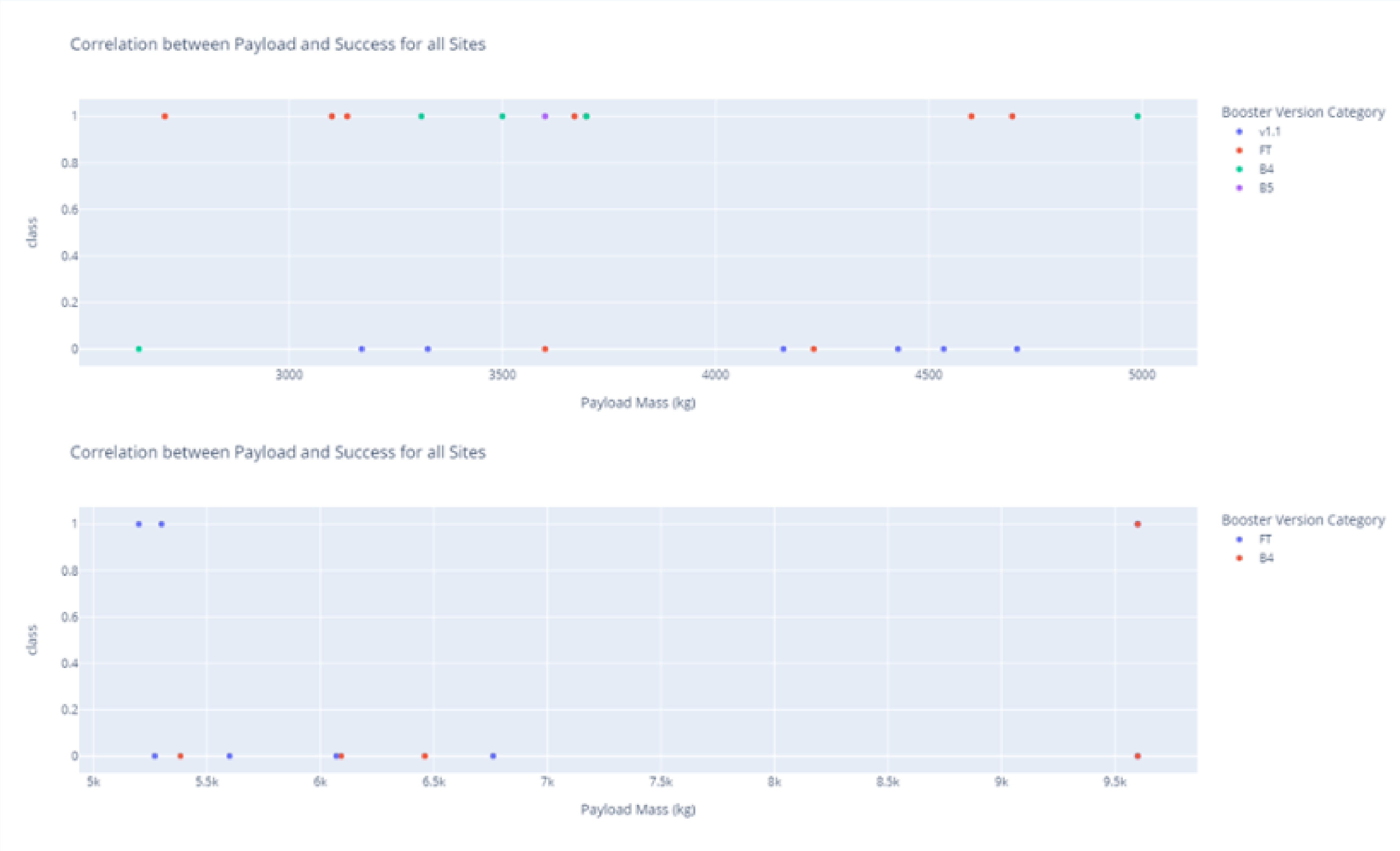
Payload vs. Launch Outcome

Scatter plot for all sites with 2500(kg), 5000(kg) and 10000(kg) payload ranges.

The 2500-5000(kg) range concentrate the majority of the successfully launches, the 0-2500(kg) range has most failed launches but all three are similar.



Payload vs. Launch Outcome

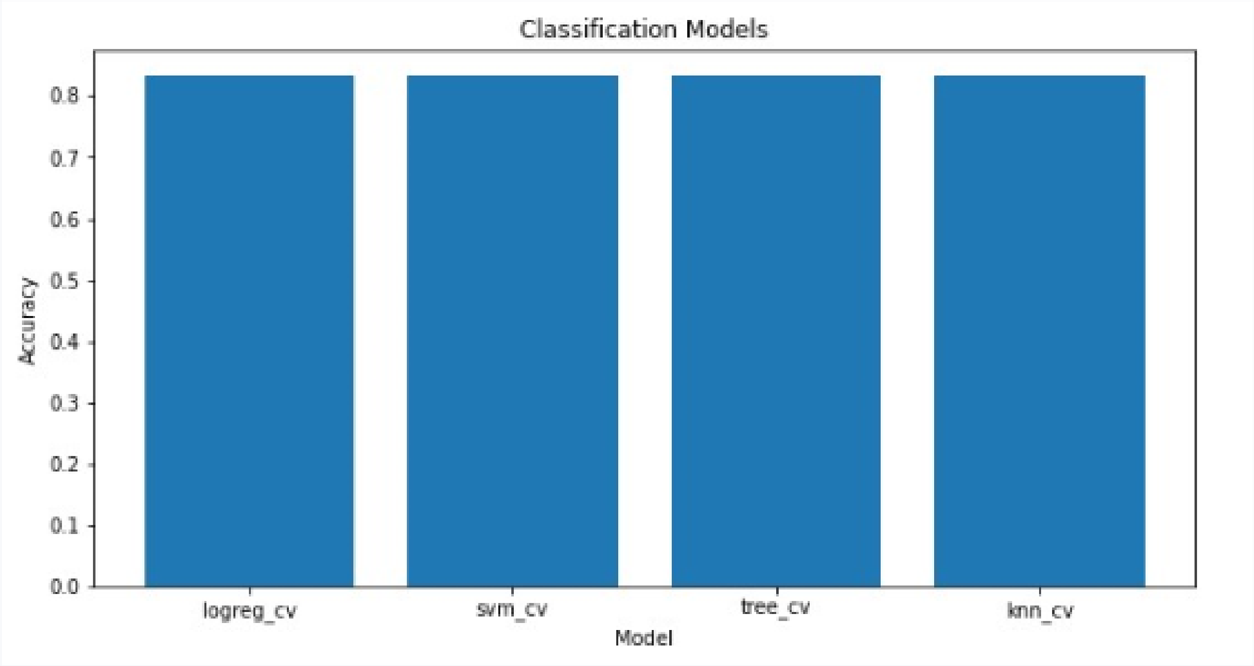


Section 5

Predictive Analysis (Classification)

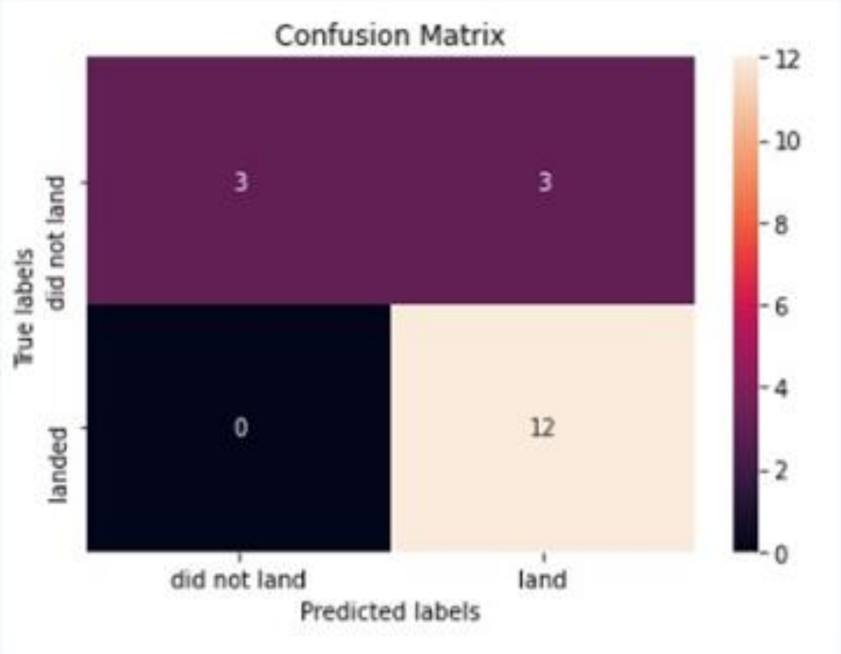
Classification Accuracy

The accuracy is the same for all models.



Confusion Matrix

The confusión matrix is the same for all models.



Conclusions

- As all the algorithms are giving the same accuracy, they all perform practically the same.
- By using our machine learning model, we can predict if the first stage of our competitor will land and determine the cost of a launch.

Appendix

- For all related files please view GitHub link:
  - [Data Science Capstone]

Thank you!