IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jesús Ignacio García
05/02/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection and data wrangling.
  - EDA and interactive visual analytics.
  - Predictive analysis.
  - EDA with data visualization.
  - EDA with SQL.
  - Building an interactive map with Folium.
  - Building a Plotly Dash dashboard.
- Summary of all results
  - Evidence of annualy increase in success rates
  - The KSC LC-39A had the most successful launches from all sites under analysis.
  - There is a higher success rate when the payloads are lighter, generally speaking
  - The same accuracy was obtained in almost all the models, Logistics Regression, Support Vector Machine, and for K nearest neighbors' (0.83333), the worst was the Decision tree model (0.611111).

Introduction

- Project background and context

  We will predict if the Falcon 9 first stage will land successfully. 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.

- Problems you want to find answers

  The main question is whether the Falcon 9 first stage will land successfully.

Section 1

Methodology

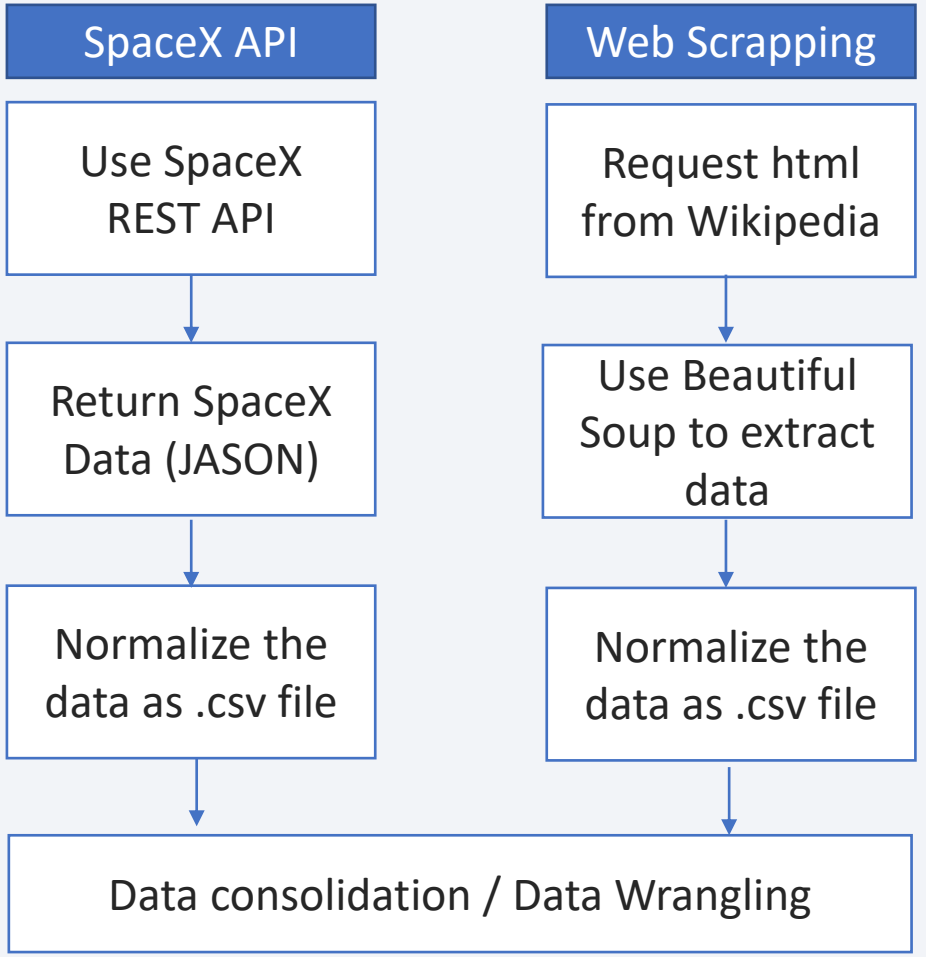
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping.
- Perform data wrangling
  - Through One Hot Encoding data fields and basic data cleaning (null values / data irrelevant).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - KNN, SVM, LR, DT models.

Data Collection

- Describe how data sets were collected:
  - Get the SpaceX data from the REST API
  - Data in JSON format contains information about the rocket, payload, launch, orbits, and landing.
  - We complete de data collection from Wikipedia by web scraping using BeauitifulSoup.



Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL:

[Data Collection – SpaceX API - github.com - Jigarcia79]

Step 1 - Request data from SpaceX API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

Step 2 - Converting response to JSON

```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

Step 3 - Getting Relevant Data

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

Step 4 - Filering the Data

```
data_falcon9 = df.loc[df['BoosterVersion']!="Falcon 1"]
```

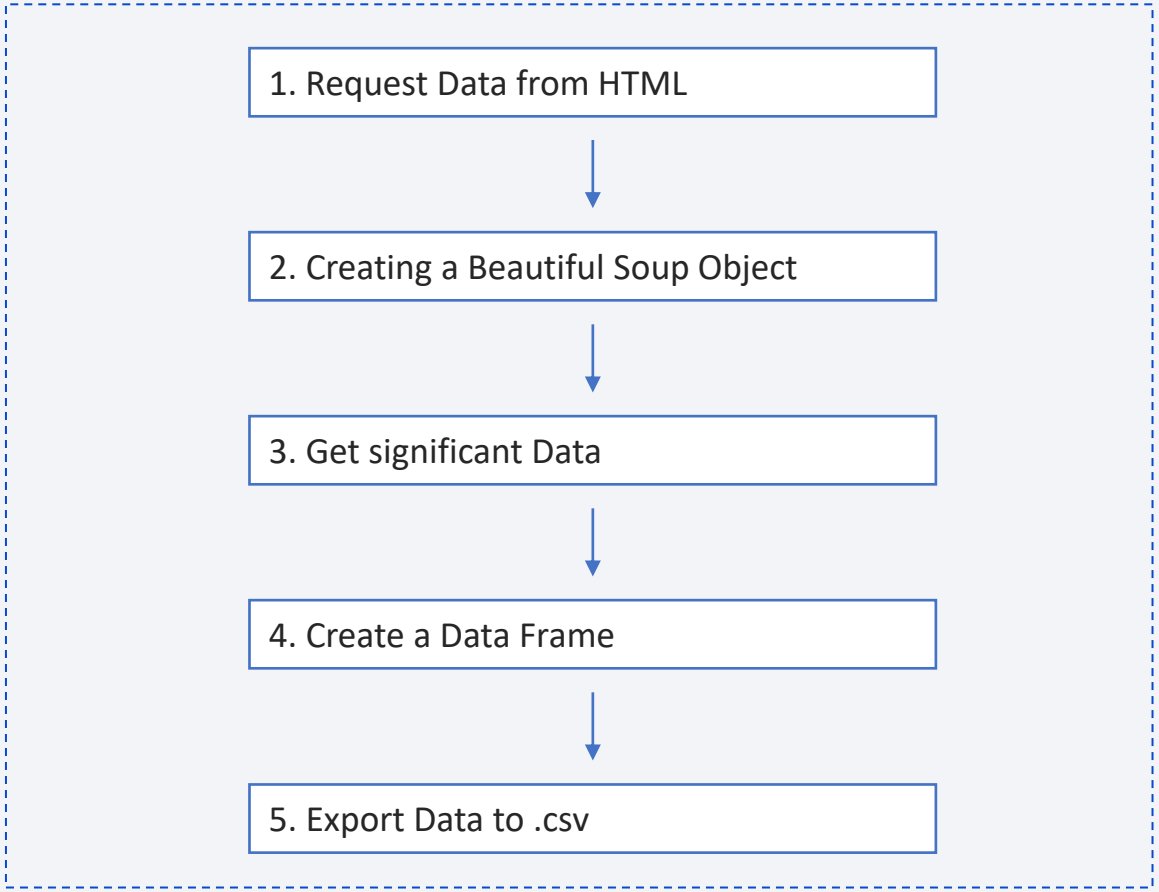
Step 5 - Export Data to .csv

```
data_falcon9.to_csv('dataset_part\_1.csv', index=False)
```

Data Collection - Scraping

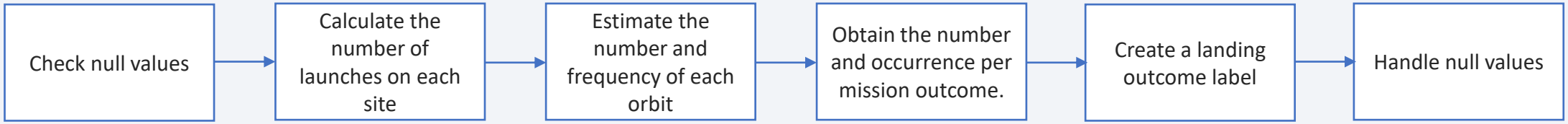
- Web scrapping from Wikipedia

- Data Collection - Scraping - github.com - Jigarcia79



Data Wrangling

EDA Analysis



- Data Wrangling - github.com - Jigarcia79

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Flight Number vs Launch Site scatter plot
  - Payload vs Launch Site scatter plot
  - Success Rate vs Orbit Type bar chart
  - Flight Number vs Orbit Type scatter plot
  - Payload vs Orbit Type scatter plot
  - Launch Success Yearly Trend line chart
- EDA with Data Visualization - github.com - Jigarcia79

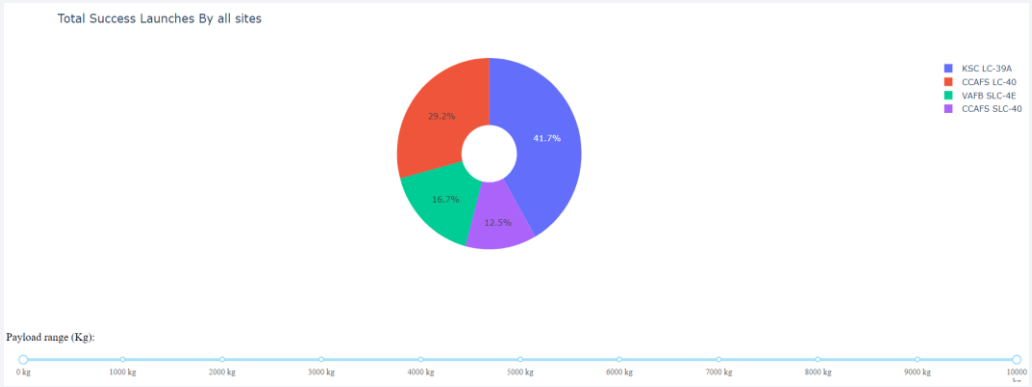
EDA with SQL

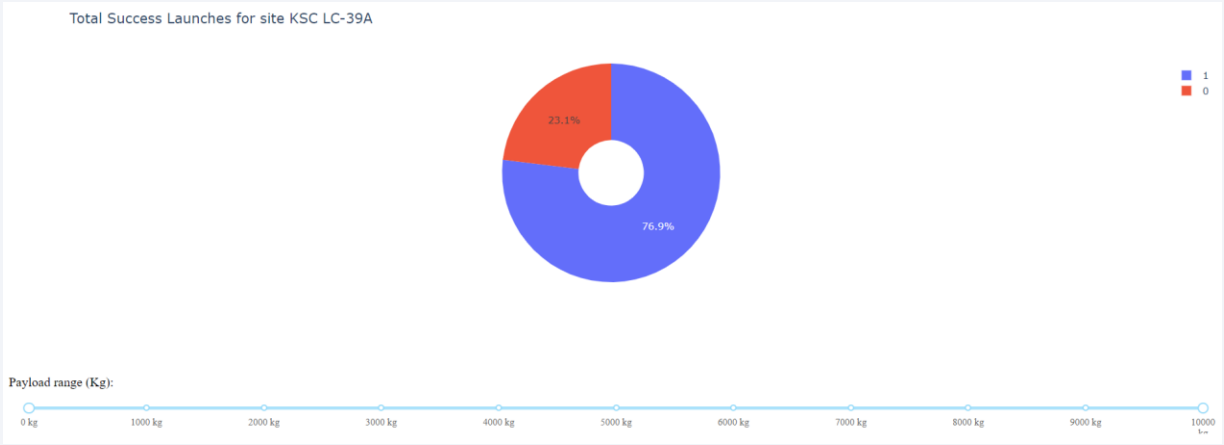
- Using bullet point format, summarize the SQL queries you performed
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was acheived.
  - 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. Use a subquery
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for 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
- EDA with SQL - github.com - Jigarcia79

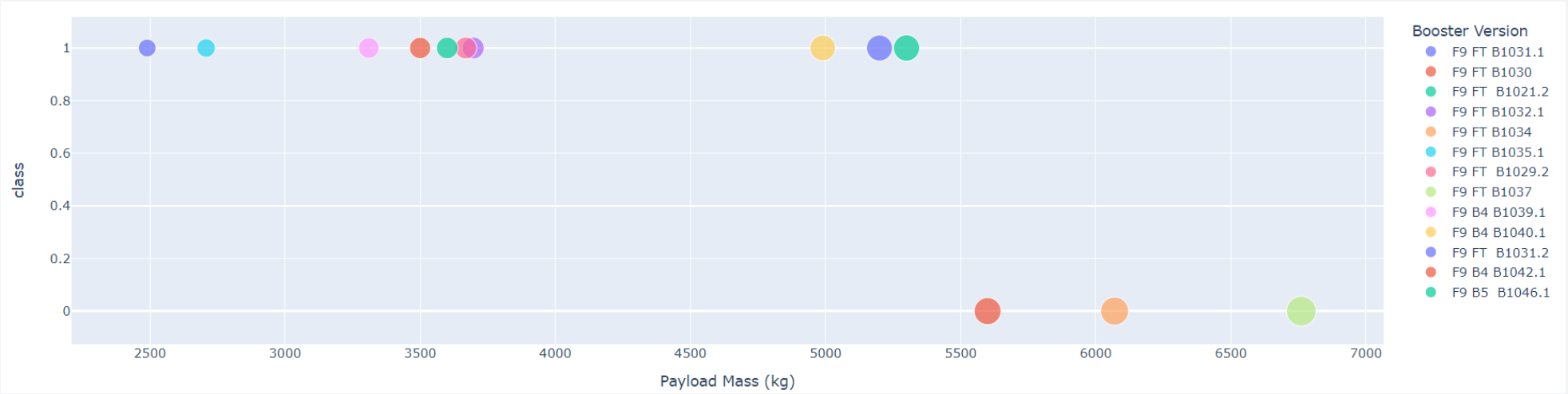
Build an Interactive Map with Folium

- Markers have been added on the map to highlight optimal locations for building launch sites.
- Interactive Map with Folium - github.com - Jigarcia79

Build a Dashboard with Plotly Dash







- [Dashboard with Plotly Dash - github.com - Jigarcia79]

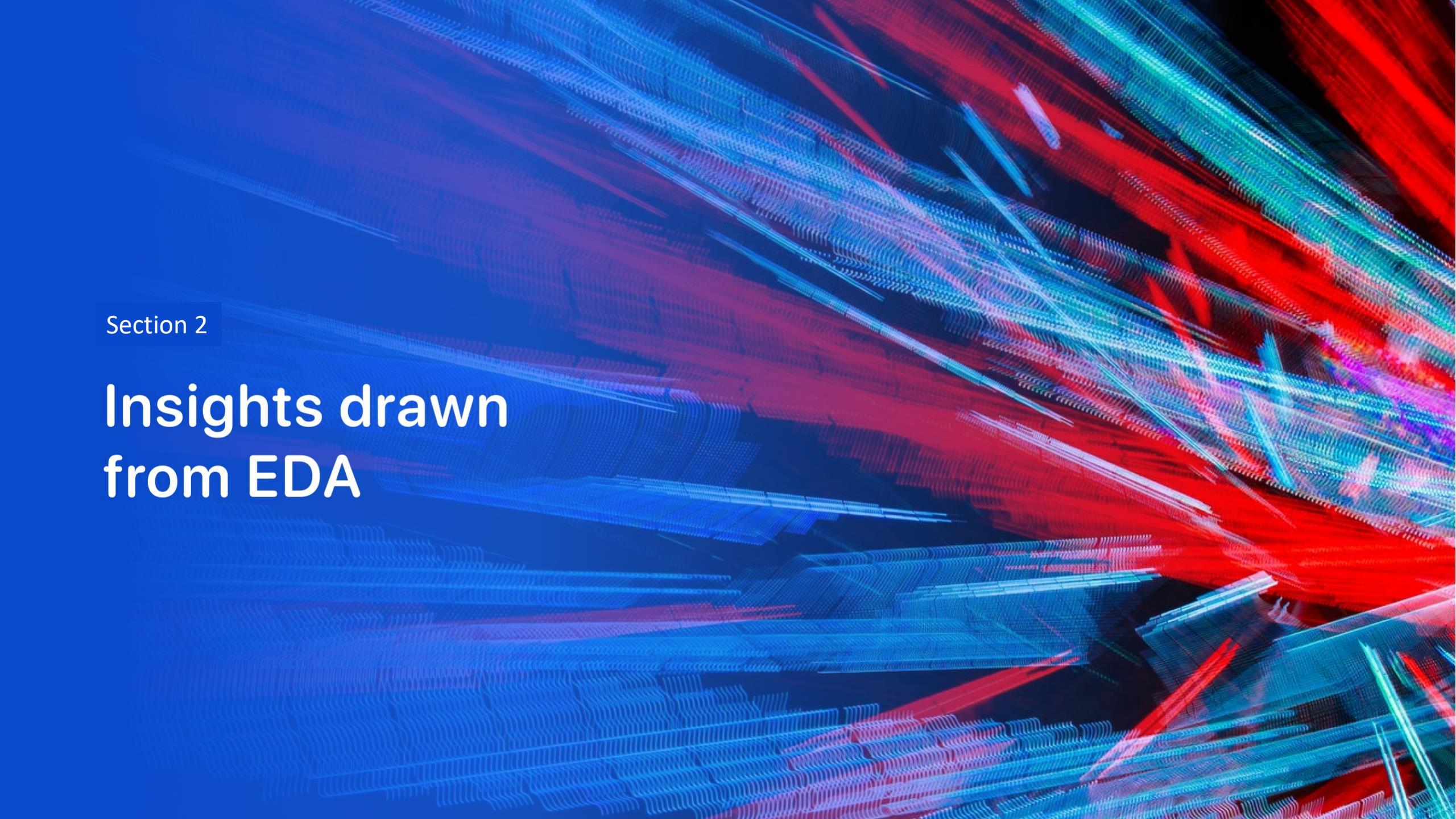
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model:
  - A logistic regression model was built to evaluate the success rate of landing falcon 9 rockets.
  - Data was split into Test and Train data
  - Several models were built, such as Regression, Decision Tree Classifier and K-nearest neighbour Classifier
  - Models were evaluated and tweaked using hyperparameter tuning.

- Predictive Analysis (Classification) - github.com - Jigarcia79

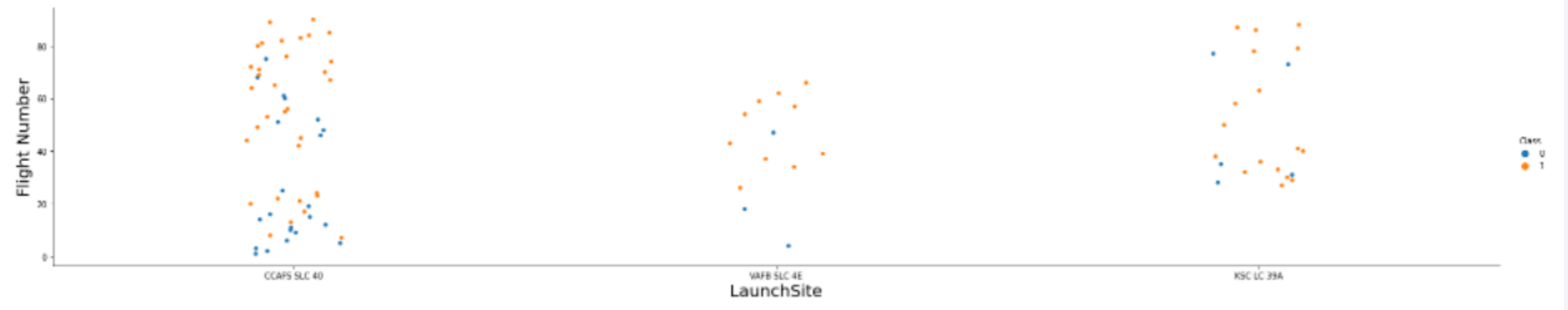
Results

- Exploratory data analysis results
  - A year-over-year increase in success rates
  - A linear relationship between success rates and years.
- Interactive analytics demo in screenshots
  - KSC LC-39A had the most successful launches from all sites.
  - There is a higher success rate when the payloads are lighter.
- Predictive analysis results
  - The same accuracy was obtained in almost all the models, Logistics Regression, Support Vector Machine, and for K nearest neighbors' (0.83333), the worst was the Decision tree model (0.611111).


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

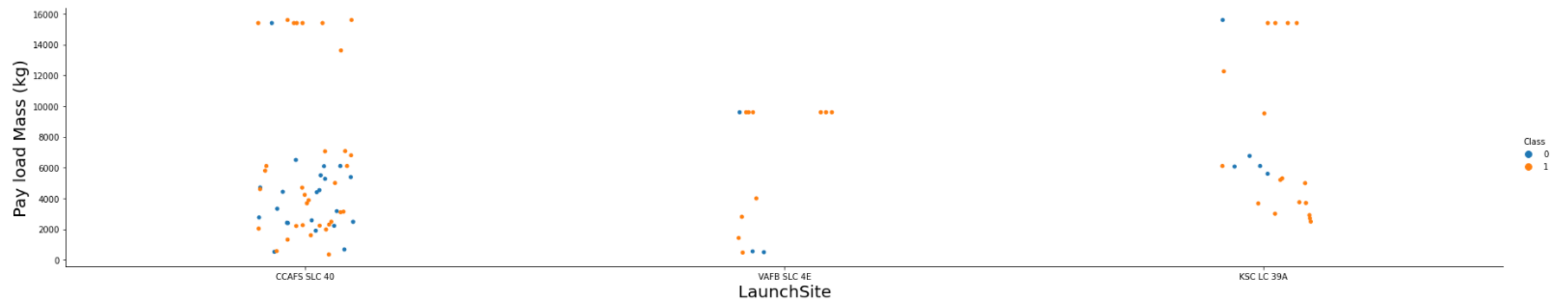
- Scatter plot of Flight Number vs. Launch Site



- Launches from the site of CCAFS SLC 40 are significantly higher than the other two sites under analysis.
- Apparently, the success rate in the launch site VAFB SLC 4E is higher than the others.

Payload vs. Launch Site

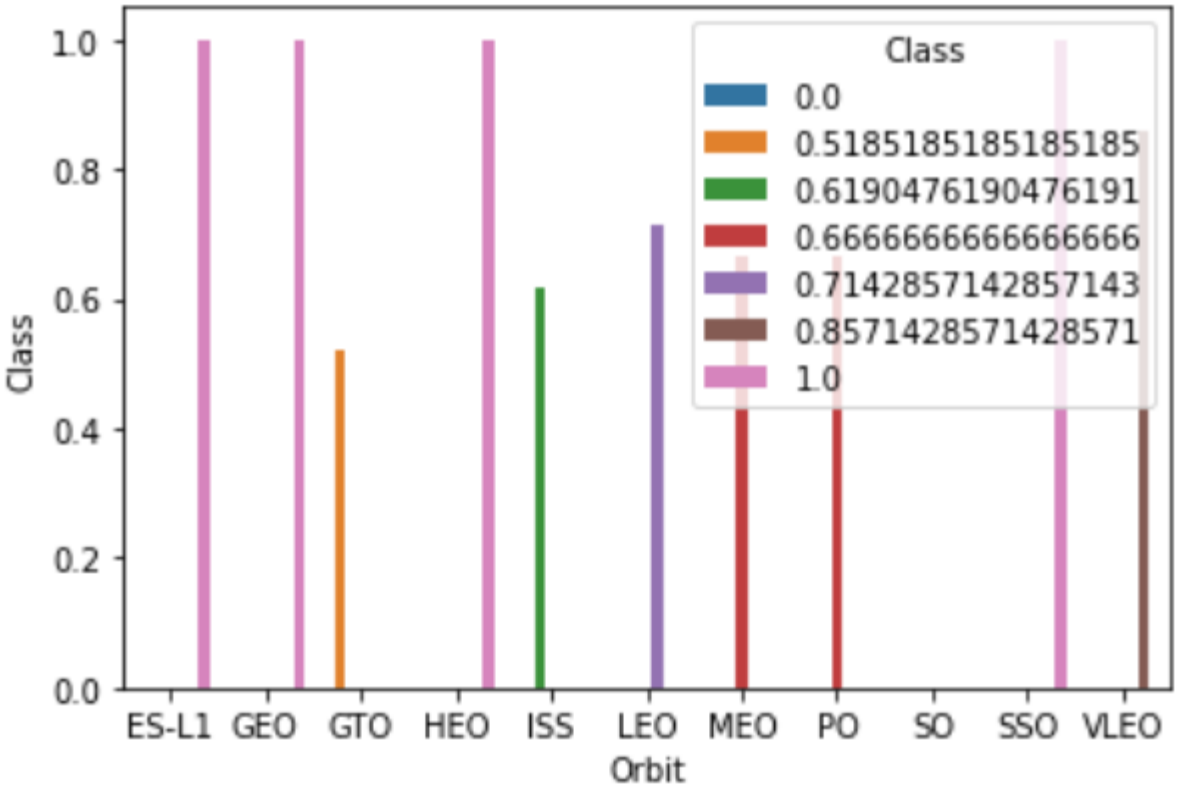
- Scatter plot of Payload vs. Launch Site



- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

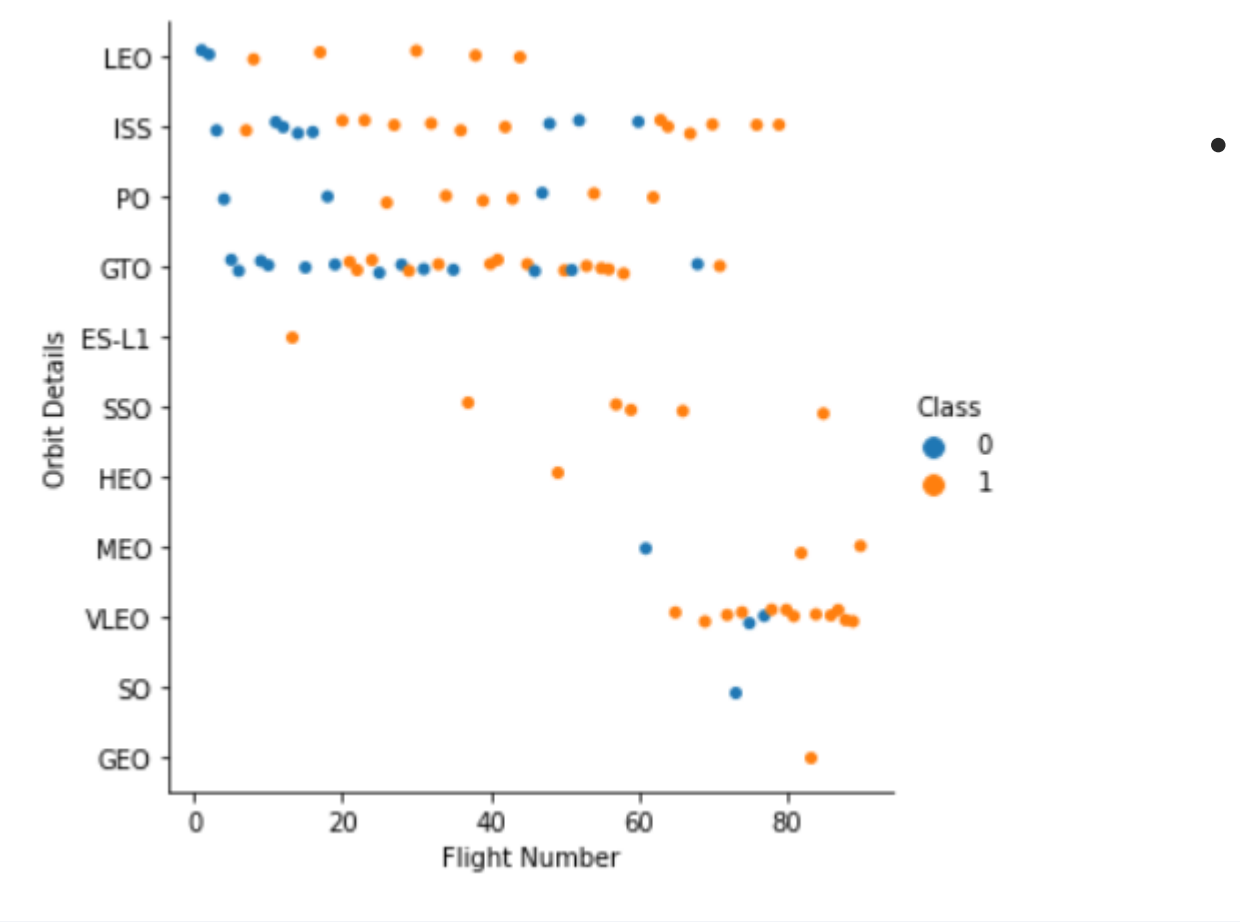
- Bar chart between the success rate of each orbit type.



- The orbit types ES-L1, GEO, HEO and SSO are among the highest success rate.

Flight Number vs. Orbit Type

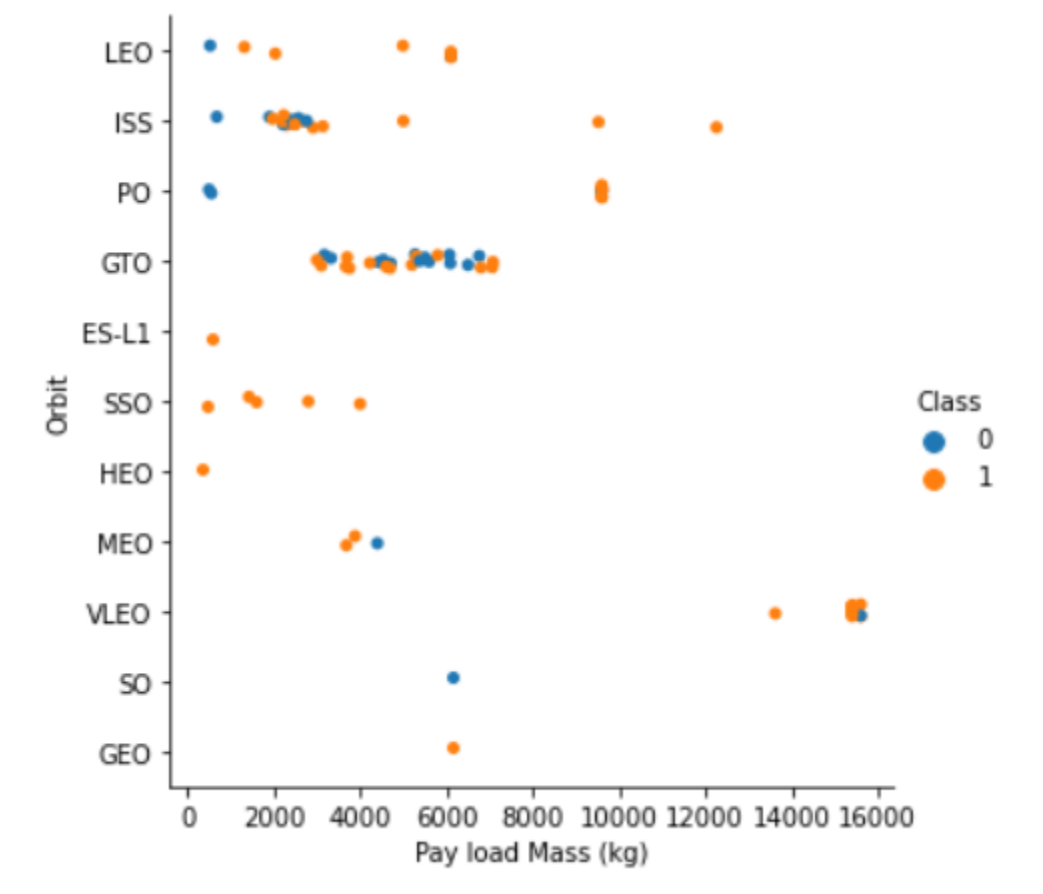
- Show a scatter point of Flight number vs. Orbit type



- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

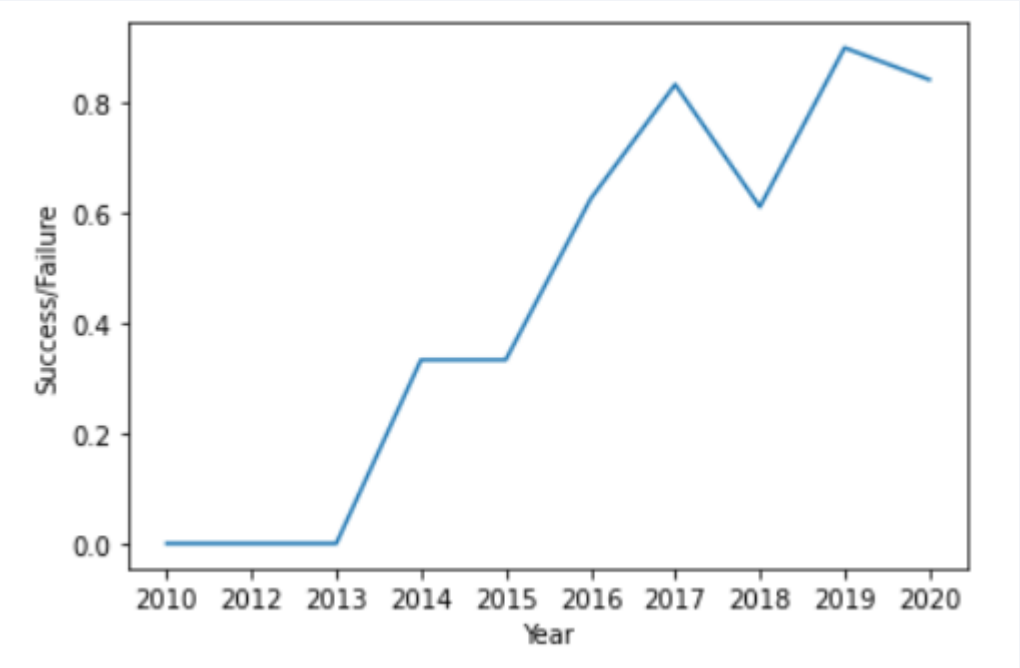
- Show a scatter point of payload vs. orbit type



- With heavy payloads the successful landing or positive 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(unsuccessful mission) are both there here.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- We can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites

```
%sql SELECT Distinct LAUNCH_SITE FROM SPACEX
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

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

- The dataset shows that there are 4 unique launch sites

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%sql SELECT * FROM SPACEX WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEX WHERE CUSTOMER='NASA (CRS)'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

1
22007

- The total payload (sum of all payloads) carried by boosters launched by NSA (CRS) is 22,007kg

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEX WHERE BOOSTER_VERSION='F9 v1.1'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

1
3676

- The average payload mass carried by booster version F9 v1.1 is 3,676kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql SELECT min(DATE) FROM SPACEX WHERE LANDING__OUTCOME='Success (ground pad)'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

1
2017-01-05

- The date when the first successful landing outcome in ground pad was achieved occurred on 2017-01-05

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 SPACEX WHERE PAYLOAD_MASS__KG_ between 4000 and 6000 AND LANDING__OUTCOME='Success (drone shi
p)'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

booster_version
F9 FT B1022
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FROM SPACEX WHERE MISSION_OUTCOME LIKE '%Success%' OR MISSION_OUTCOME LIKE '%Failure%'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

1
45

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql SELECT BOOSTER_VERSION FROM SPACEX WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX)
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

2015 Launch Records

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

```
%sql SELECT TO_CHAR(TO_DATE(MONTH("DATE"), 'MM'), 'MONTH') AS MONTH_NAME, \
    LANDING__OUTCOME AS LANDING__OUTCOME, \
    BOOSTER_VERSION AS BOOSTER_VERSION, \
    LAUNCH_SITE AS LAUNCH_SITE \
    FROM SPACEX WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND "DATE" LIKE '%2015%'
```

```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

month_name	landing__outcome	booster_version	launch_site
OCTOBER	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%sql SELECT "DATE", COUNT(LANDING__OUTCOME) as COUNT FROM SPACEX \
    WHERE "DATE" BETWEEN '2010-06-04' and '2017-03-20' AND LANDING__OUTCOME LIKE '%Success%' \
    GROUP BY "DATE" \
    ORDER BY COUNT(LANDING__OUTCOME) DESC
```

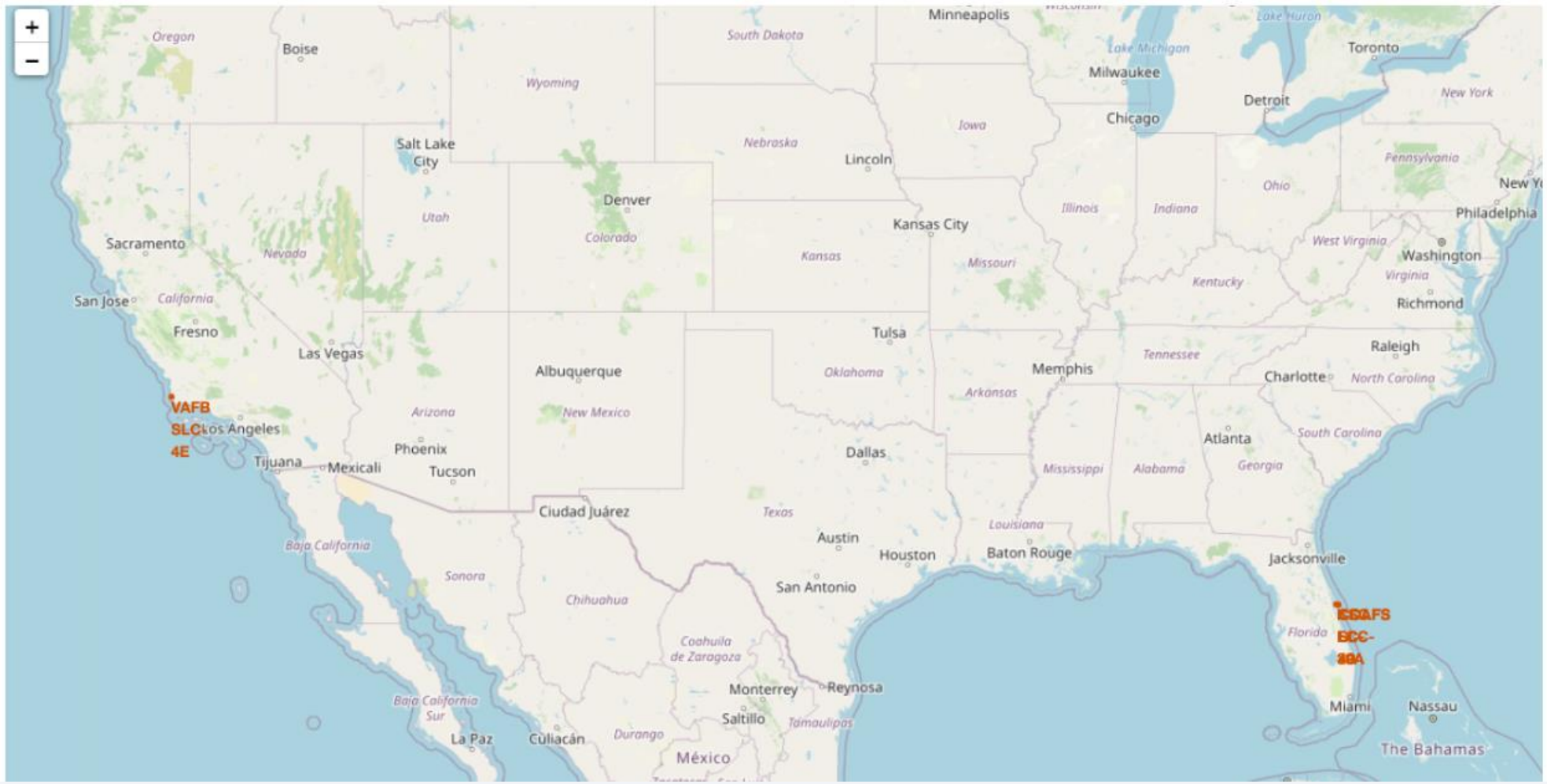
```
 * ibm_db_sa://srb41116:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/bludb
Done.
```

DATE	COUNT
2016-06-05	1
2016-08-04	1
2017-01-05	1
2017-03-06	1

Section 3

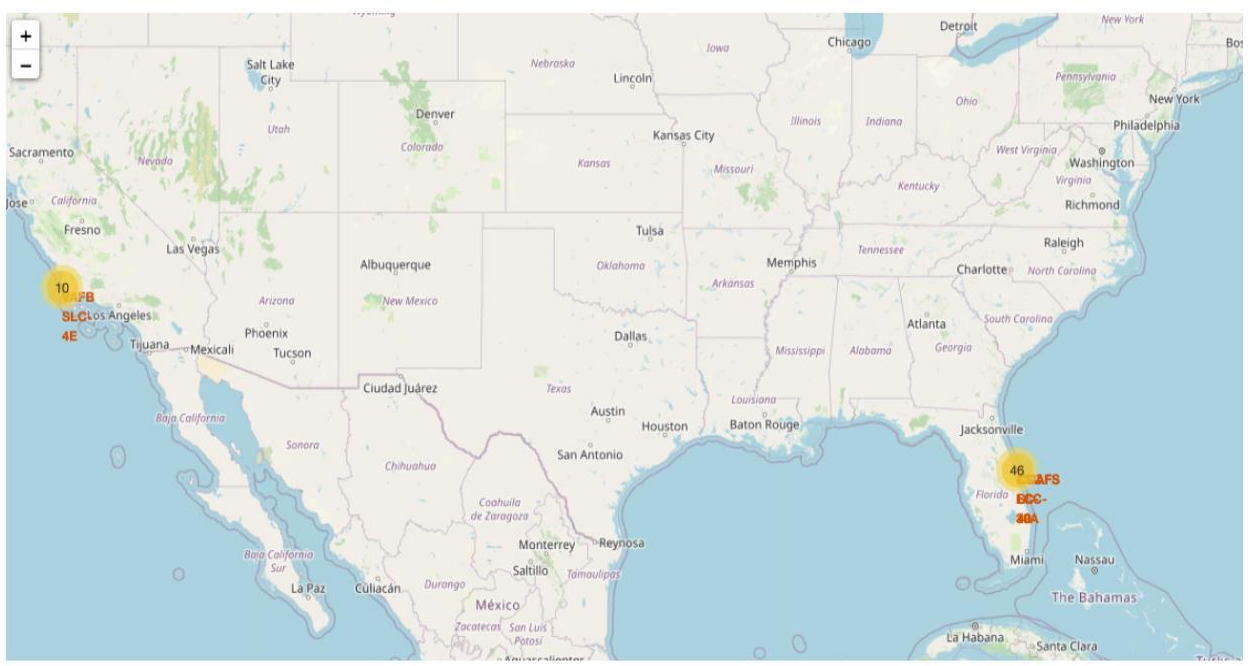
Launch Sites Proximities Analysis

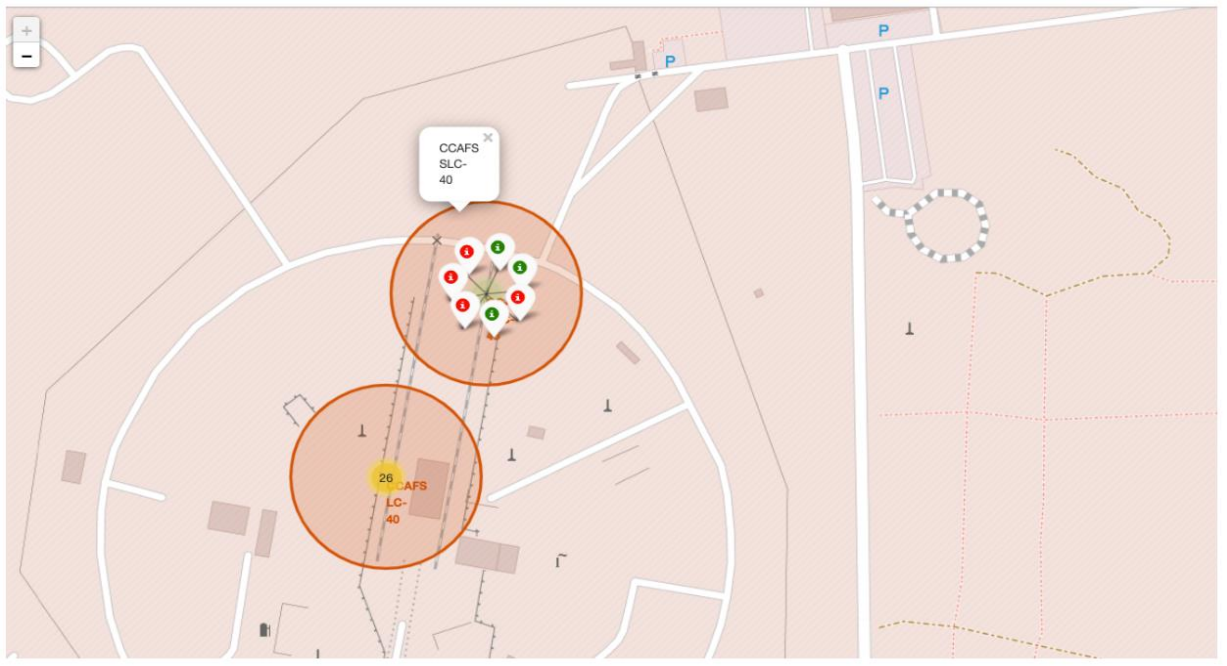
Mark all launch sites on a map



- All launch sites are far from the equator.
- All launch sites are near the coast.

Mark the success/failed launches for each site on the map

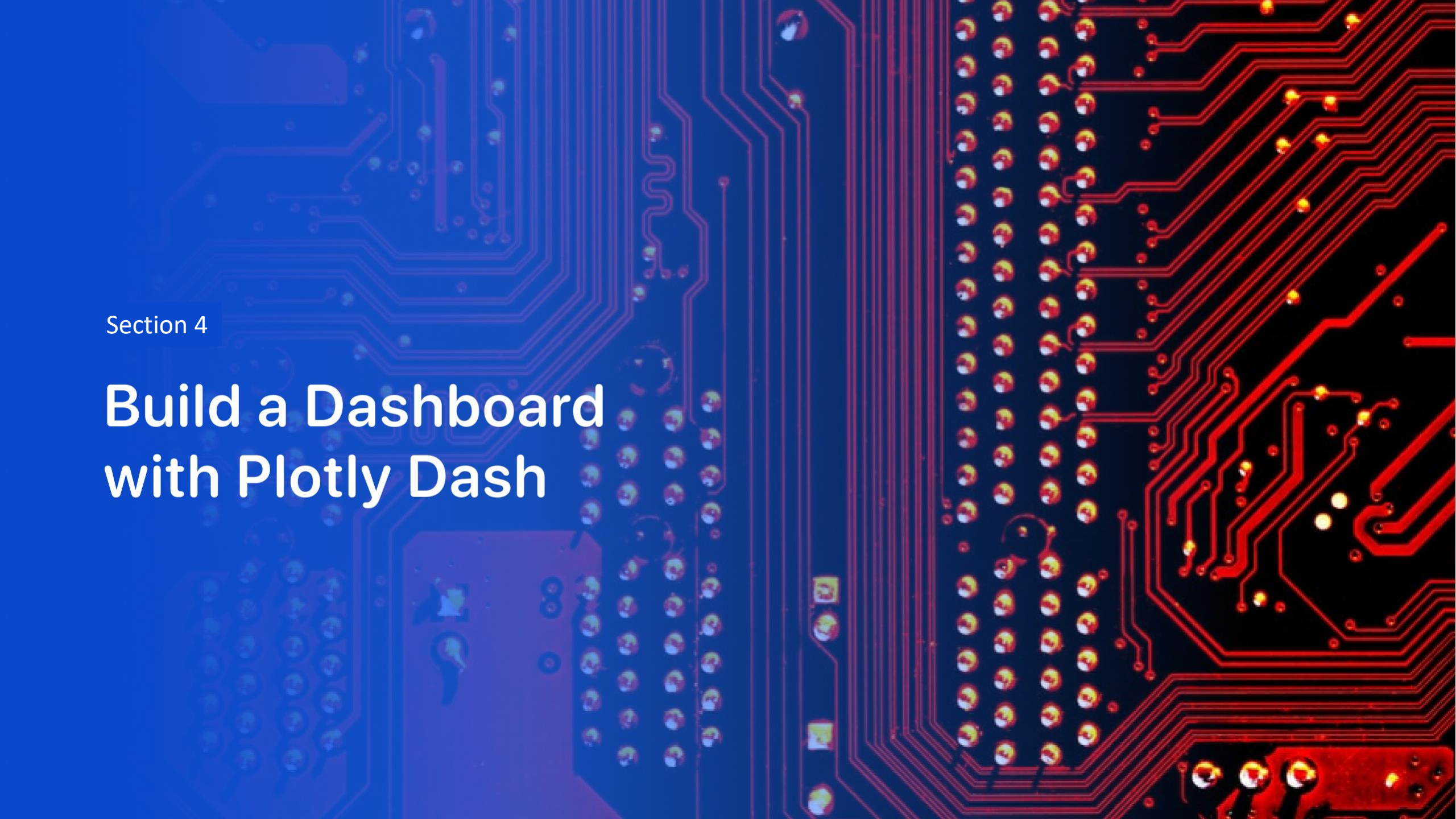




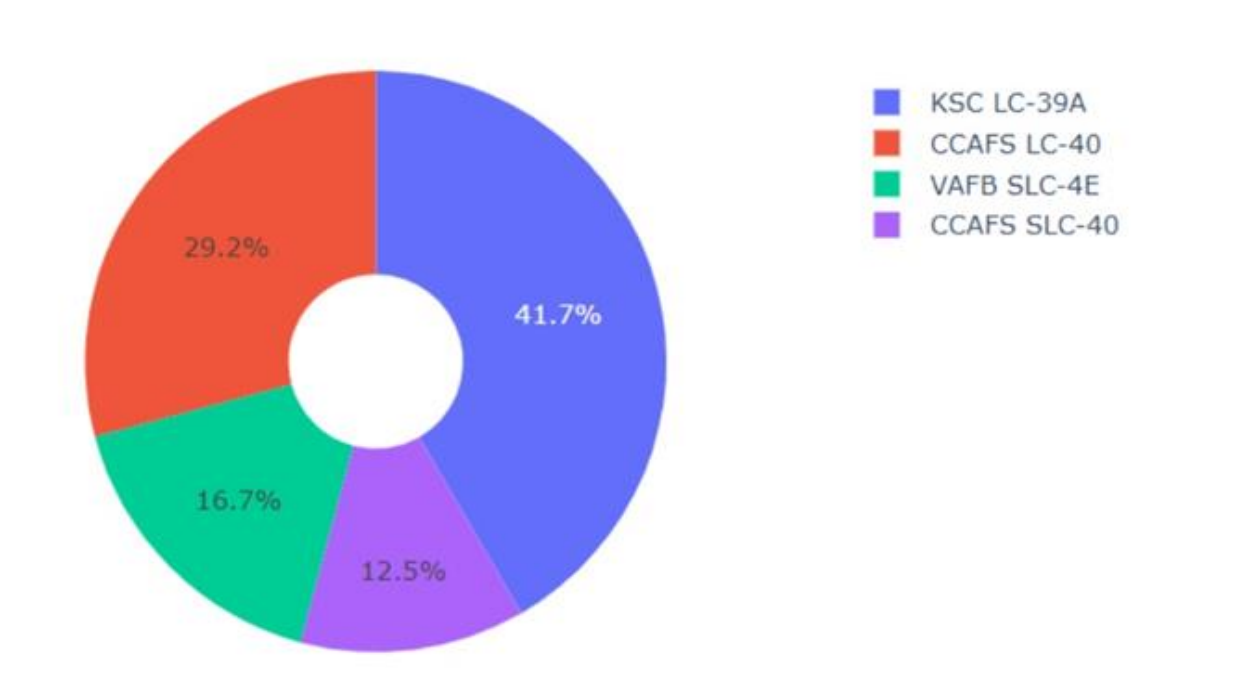
- Red dots indicate unsuccessful launches from the site CCAFS SLC-40

Calculate the distances between a launch site to its proximities

- The CCAFS SLC-40 launch site is about 0.9 km from the coastline.

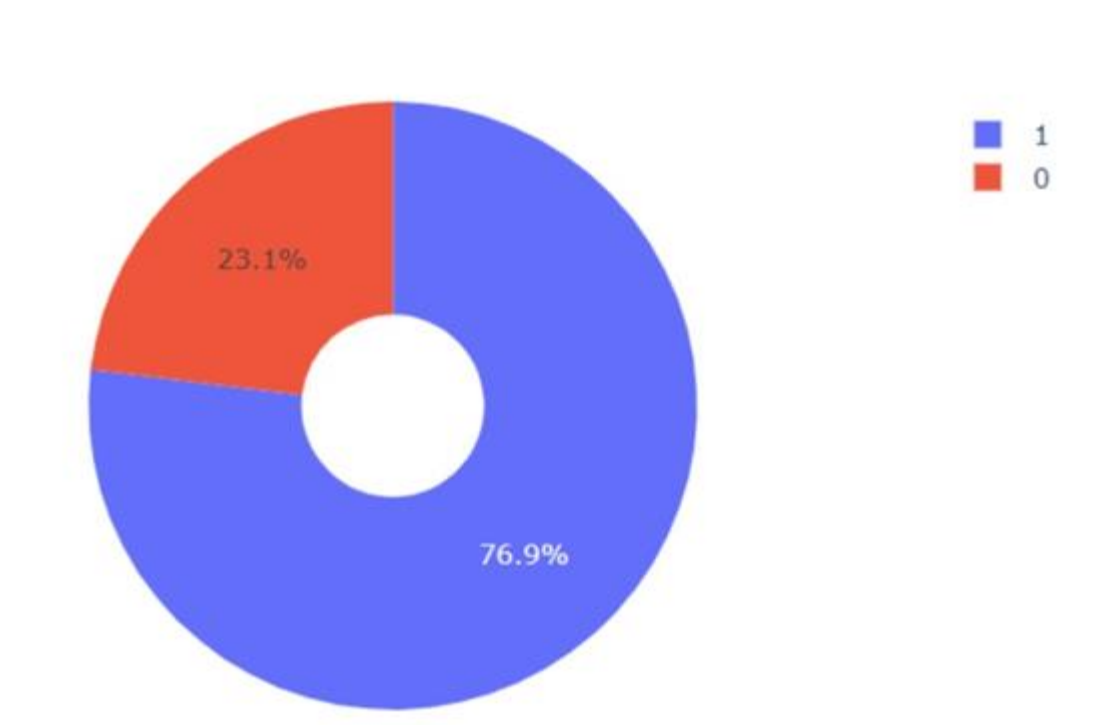

Section 4
Build a Dashboard with Plotly Dash

Total success launches by all sites



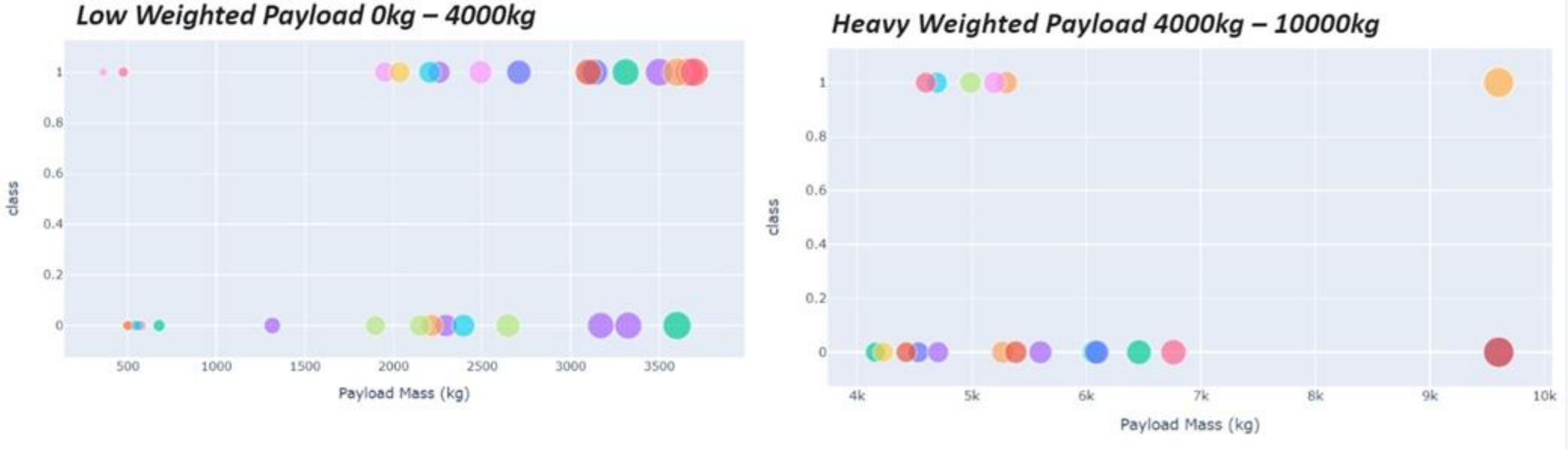
- The KSC LC – 39A launch site has the highest success rate (41.7%).
- The CCAFS SLC-40 launch site has the lowest success rate (12.5%)

Success rate by site



- The success rate of the KSC LC - 39A launch site is equal to 76.9%.

Payload vs launch outcome

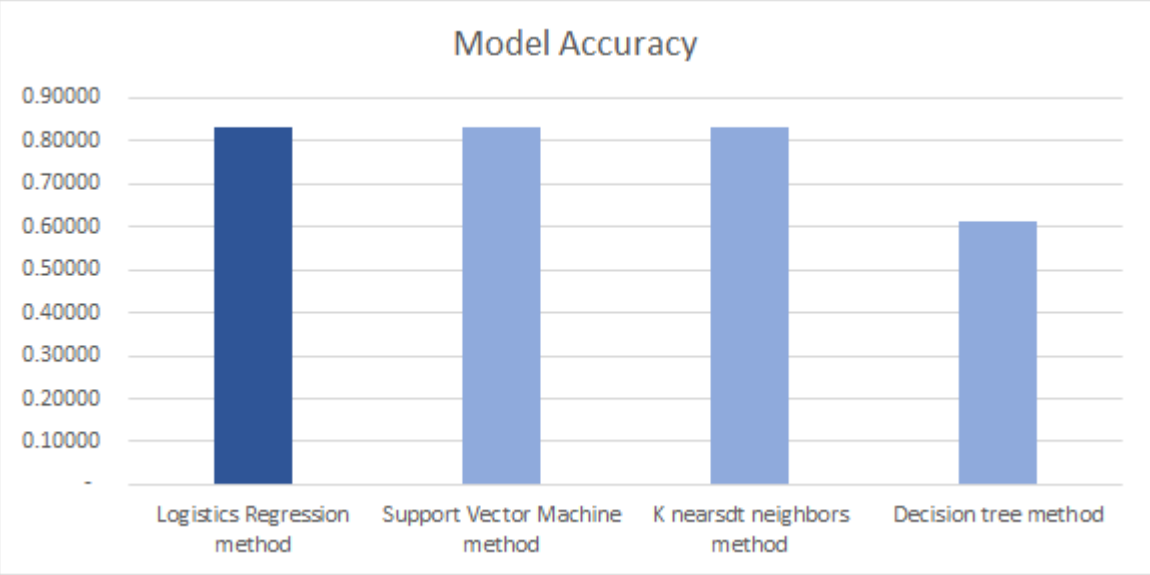


- According to the established ranges, it appears that the lower the weight of the load, the higher the launch success rate.

Section 5

Predictive Analysis (Classification)

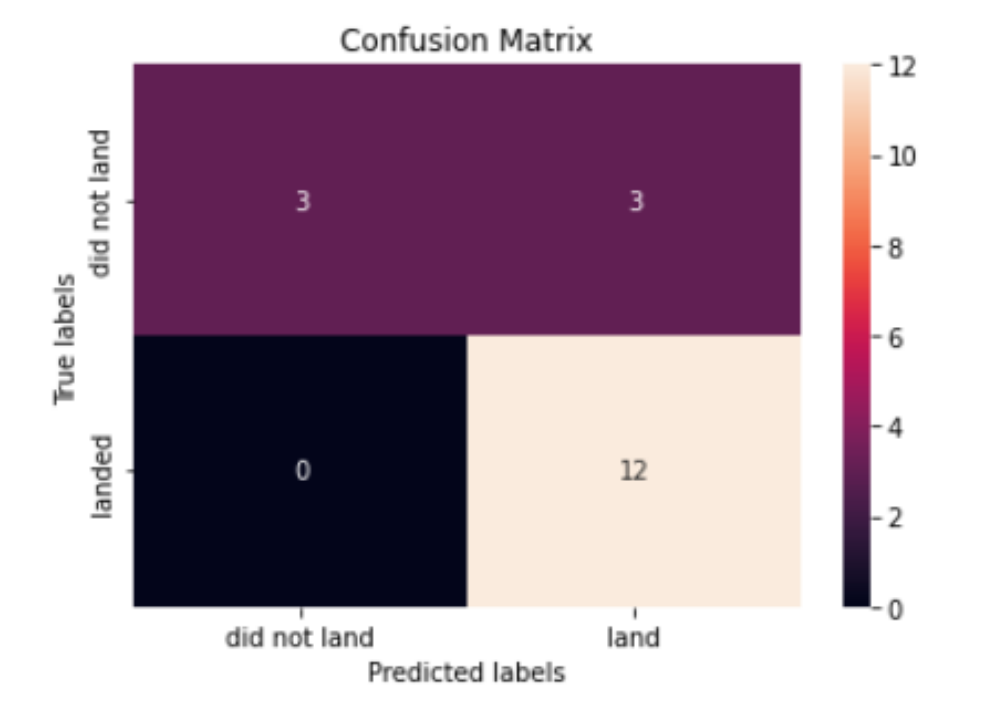
Classification Accuracy



	Model	Accuracy
1	Logistics Regression method	0.83333
2	Support Vector Machine method	0.83333
3	K nearsdt neighbors method	0.83333
4	Decision tree method	0.61111

The same accuracy was obtained in almost all the models, Logistics Regression, Support Vector Machine, and for K nearest neighbors' (0.83333), the worst was the Decision tree model (0.611111).

Confusion Matrix - Logistic Regression



- The confusion matrix for the logistic regression model, show:
  - The logistic regression model predicts 0 False Negatives and 3 False Positive.
  - In the other hand, we have 3 True Positives and 12 True Negatives.

Conclusions

- The Logistic Regression, Support Vector Machine, and K nearest neighbors' models has the best performance for predict.
- There is evidence that the lower the weight of the load, the higher the success rate of the launch.
- The launch site KSC LC-39A had the highest success rate of the four sites.

Thank you!