THE SPACE WITH DATA SCIENCE

BRENO BARTHOLOMEU PITTA

JANUARY 2023

IBM Developer

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

OUTLINE

- **Summary Methodologys**

  Data collection with API
  Data collection with WebScraping
  Data Wrangling
  Data Analysis with SQL
  Data Visualization
  Data Analytic with Folium
  Machine Learning Methodos

- **Summary Results**

  Data Exploratory
  Interactive DashBoard
  Predictive Analysis

EXECUTIVE SUMMARY

• Context

The objective of this endeavor is to devise a machine learning pipeline that can predict the success of the first stage landing for SpaceX's Falcon 9 rocket launches. As advertised on the company's website, the cost for a launch is significantly lower than that of other providers, with a cost of $62 million as opposed to $165 million or more. This cost savings is largely attributed to SpaceX's ability to reuse the first stage. Thus, by determining the likelihood of a successful first stage landing, other companies may be able to use this information to compete with SpaceX in bidding for rocket launch services.

• What we want

- Why do some rockets take off and others don't?
- What it takes to have a successful landing
- Factors that alter the landing success rate

INTRODUCTION

METHODOLOGY

Executive Summary

- **Data Collection -** SpaceX API and WebScraping from Wikipedia
- **Data Wrangling -** One Hot Encode to Categorical Features
- **Analisys with SQL**
- **Analisys with Folium**
- **Interactive Dashboard**
- **Predictive analisys with Classification Models -** Build, tune and evaluation

- First we use get request to get the data SpaceX from the API. We use some functions to decode .json and normalize the data with the pandas library. After this, the data after that, the empty data values were removed, replaced (by the average) or just ignored.

- We use the webscraping method from beautifulsoup to get the launch data from wikipedia. The data obtained by wikipedia were transferred and modeled to a dataframe for later analysis.

DATA COLLECTION

- The get request was used to collect data from the SpaceX API. Some data cleaning and formatting was done

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%201/Lab's/Data%20Collection%20API.ipynb

DATA COLLECTION- API

- Webscraping with BeautifulSoup in WikiPedia to get the Falcon 9 launches

- All data has been converted to a dataframe

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%201/Lab's/Data%20Collection%20with%20Web%20Scraping.ipynb

DATA COLLECTION- WEB SCRAPING

- From the analysis of the data, we define the data of independent variables

- Trends were calculated for each orbit and launch pad

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%201/Lab's/Data%20wrangling.ipynb

- At this stage, graphs were created showing the relationships between flight number and launch site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly trend.

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%203/Lab's/Data%20Visualization.ipynb

- The connection between the IBM SQL server and the jupyter environment was made

- Some insights were analyzed with SQL

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%202/Lab's/SQL%20Notebook%20for%20Peer%20Assignment.ipynb

- We add circles and markers from each release with Folium, creating an interactive map
- We define a binary selection for the dependent variable "Class", with success equal to 1 and failure equal to 0
- With the marker cluster argument, we checked which launches had high success percentages
- Calculated the distance between some instances, such as a city and a coast
- **Link to Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%203/Lab's/Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

DATA FOLIUM

- An interactive dashboard was built using plotly and dash

- Graphs were built, like a pie chart showing the relationship between releases and certain sites. A scatter plot was constructed, relating Outcome and payload mass (kg), for different boosters

- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%203/Lab's/spacex_dash_app.py

DASHBOARD INTERACTIVE

- We load the data, transform it and split it into training and testing groups

- From GridSearchCV, we build several predictive models, always looking for the best parameters

- From accuracy metrics, we found the best data prediction model

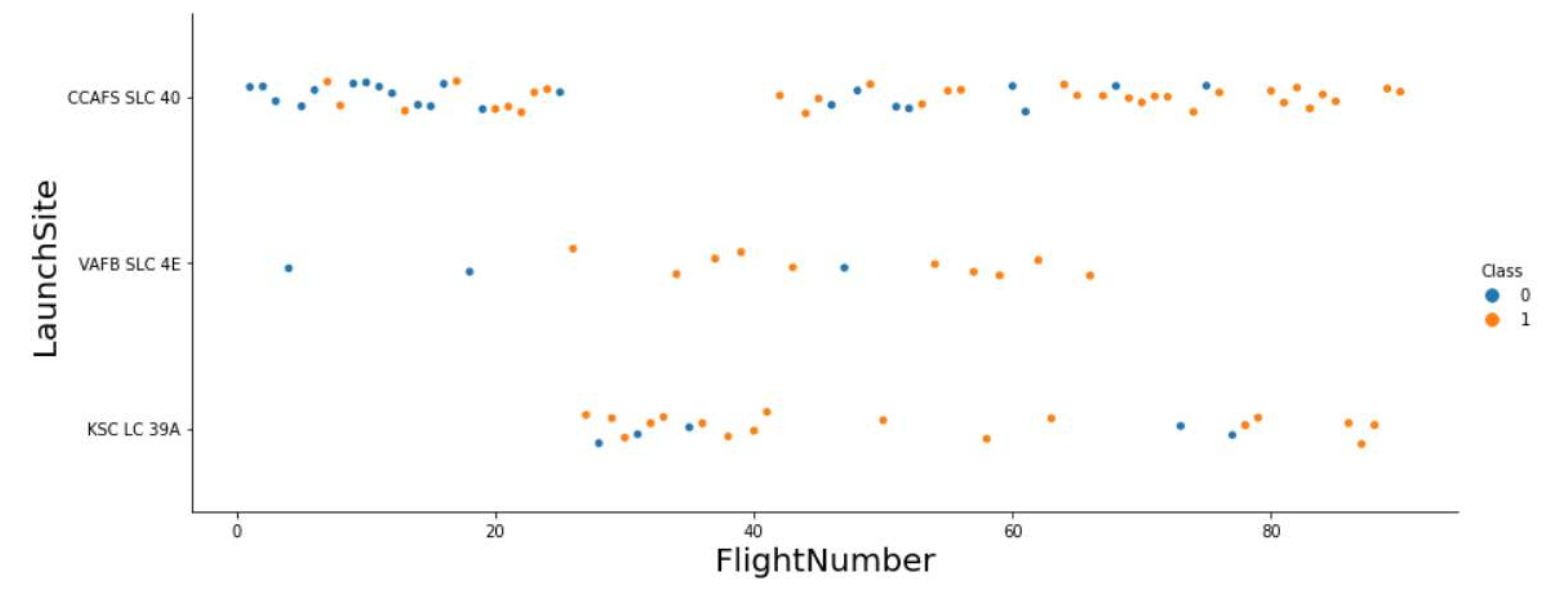
- **Link to the Notebook:** https://github.com/elpitta/MyRepository/blob/main/IBM%20Professional%20Data%20Science/10.%20Applied%20Data%20Science%20Capstone/Week%204/Lab's/_Falcon%209%20First%20Stage%20Landing%20Prediction.ipynb

PREDICTIVE ANALYSIS

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

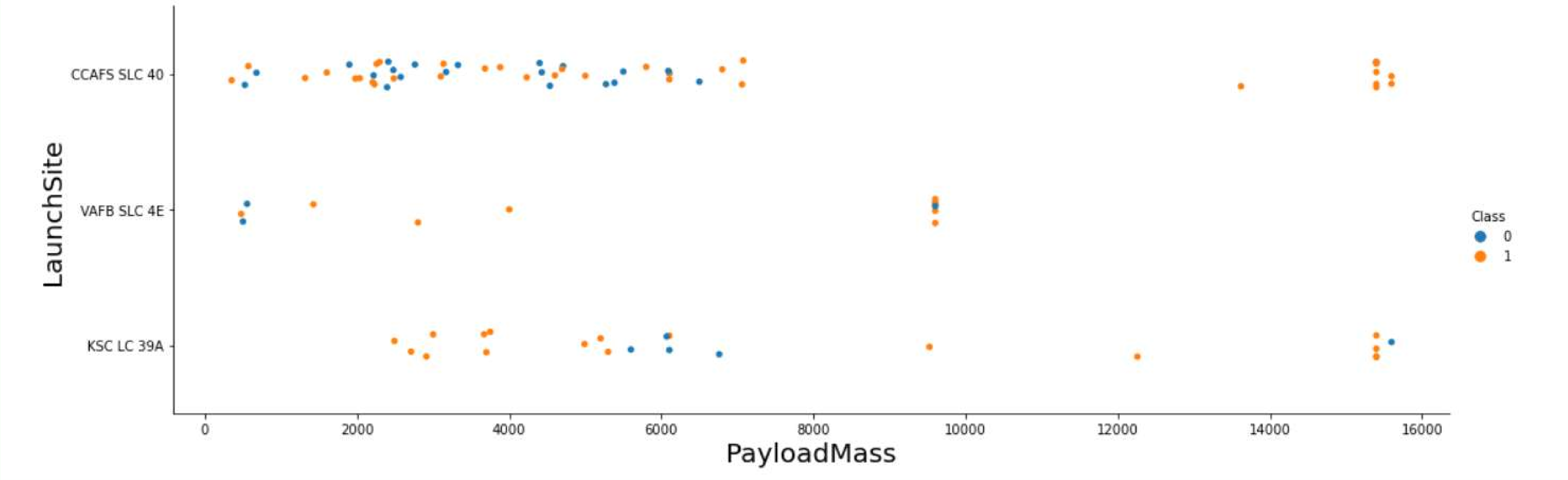
RESULTS

INSIGHTS DATA



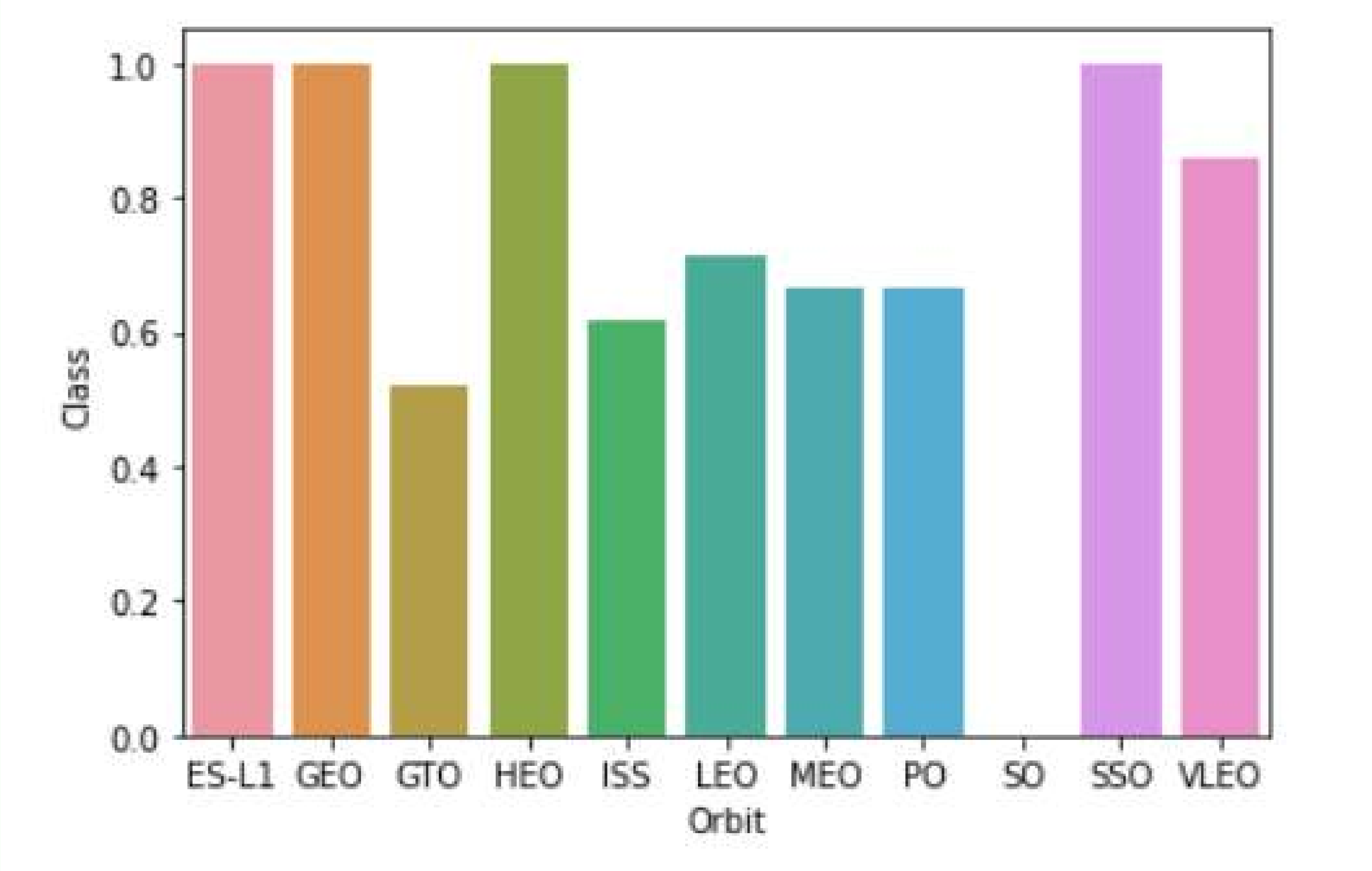
From the chart, we see that the higher the flight number, the higher the success rate for the launch site.

FLIGHT NUMBER X LAUNCH SITE



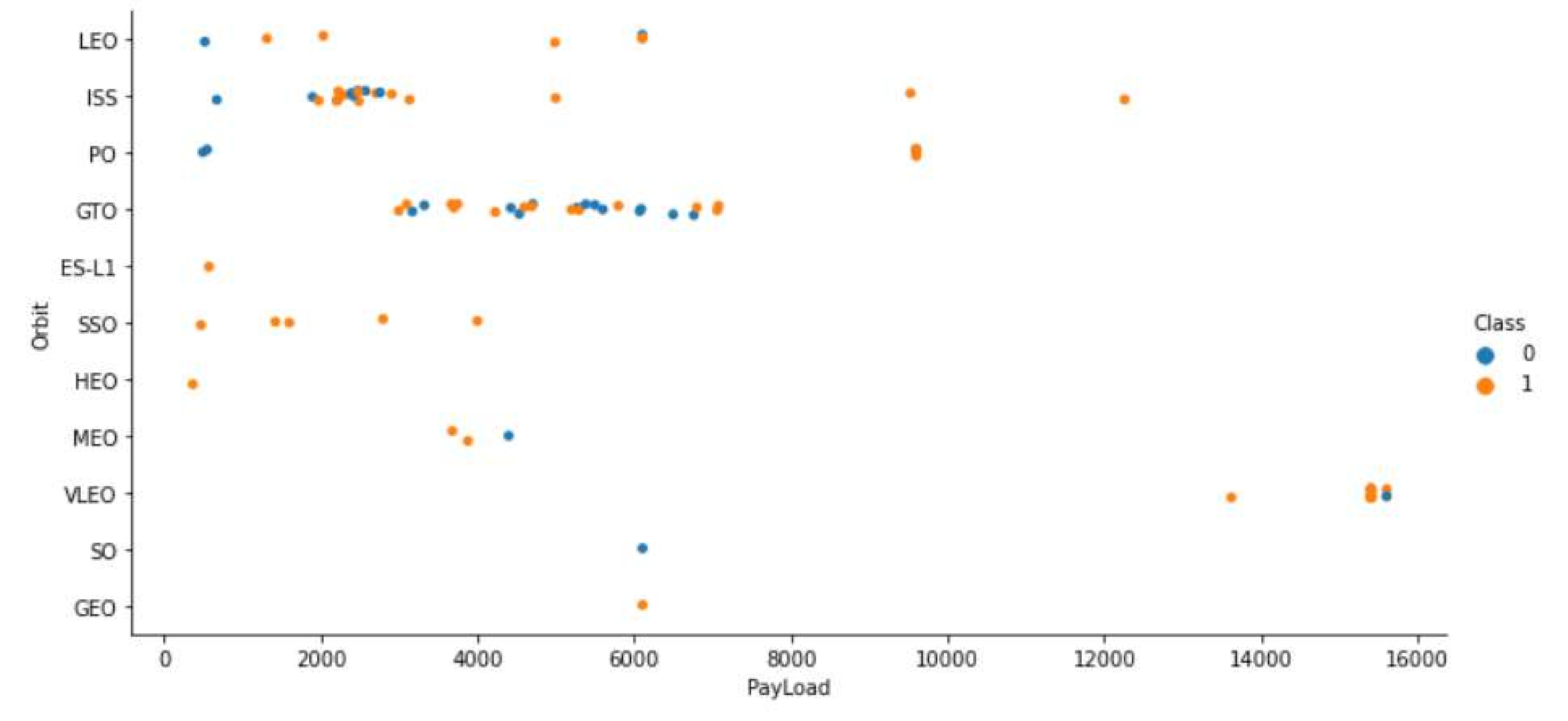
The higher the payloadmass for the SLC-40, the greater the chance of success

PAYLOAD MASS X LAUNCH SITE



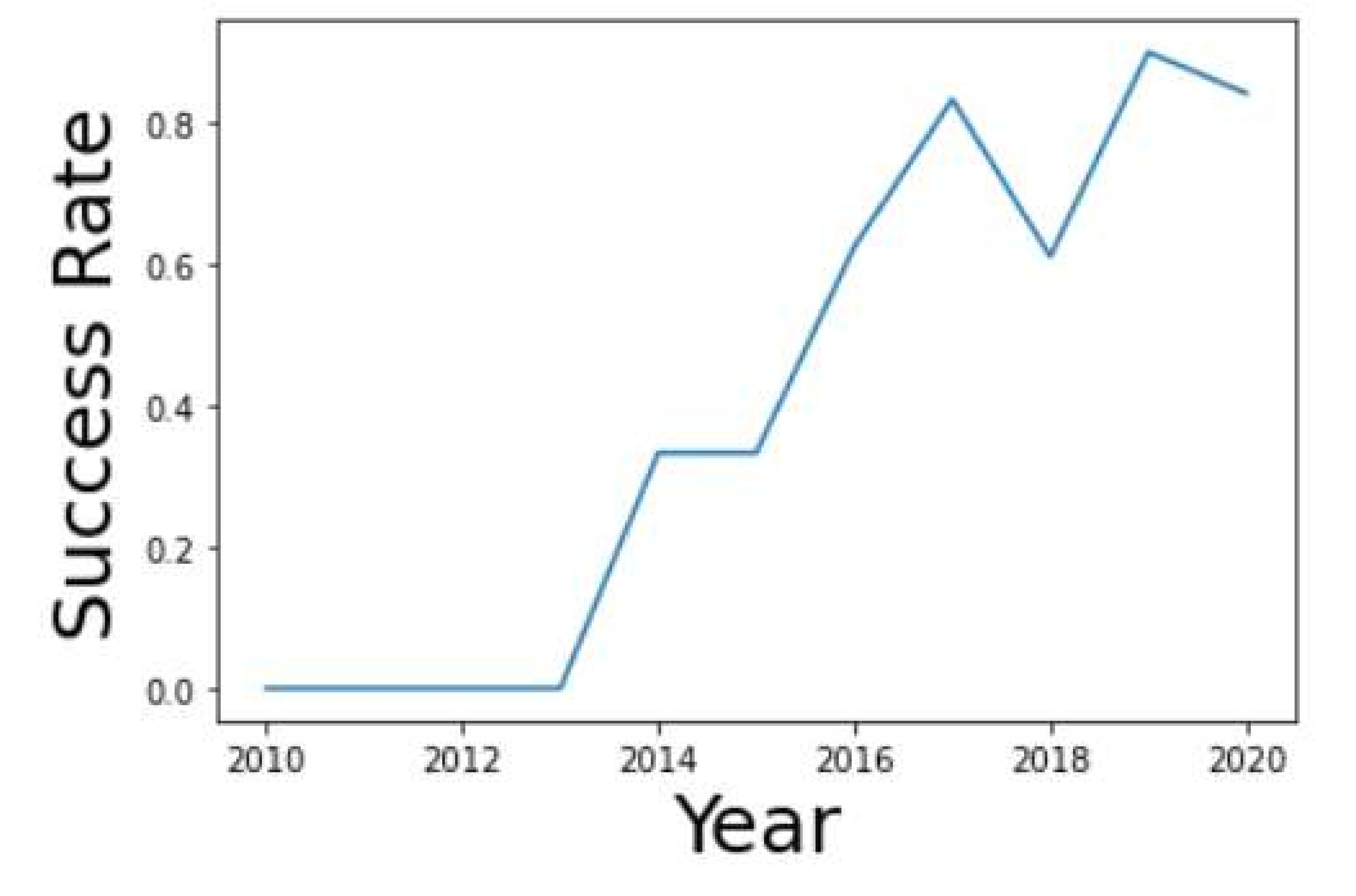
The best chances of success are for ES-L1, GEO, HEO, SSO and VLEO orbits

SUCCESS RATE X ORBIT



Types of payloadmass influence the type of orbit chosen. for example, mass of 10000 kg we only have success with PO, mass of 1000 kg we only have failures.

PAYLOAD MASS X ORBIT



Over the years, the success rate has increased, with some drops between certain periods

YEARLY X SUCCESS RATE

```
%%sql

SELECT DISTINCT launch_site FROM SPACE ;
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84
Done.
```

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

Disitinct names of launch sites for space missions were displayed

LAUNCH SPACE NAMES

```
%%sql

SELECT * FROM SPACE WHERE (launch_site) LIKE 'CCA%' LIMIT 5 ;
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/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

Was displayed records where launch begins with string "CCA"

RECORD NAMES

```
%%sql

SELECT sum(payload_mass__kg_) as "Total mass by CRS" FROM SPACE where customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Total mass by CRS
45596

We calculated the total payload carried by boosters launched by NASA (CRS)

TOTAL PAYLOAD NASA

```
pd = %sql SELECT avg(payload_mass__kg_) as "Average mass F9 v1.1" FROM SPACE WHERE Booster_version = 'F9 v1.1'

pd
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Average mass F9 v1.1
2928

Average payload mass carried by booster version F9 v1.1 was calculated

AVERAGE PAYLOAD MASS

```
%%sql

SELECT min(DATE) FROM SPACE WHERE landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

1
2015-12-22

Was calculated on the date when the first successful landing outcome in ground pad was achieved.

DATE SUCCESSFUL LANDING

```
%%sql

SELECT distinct booster_version FROM SPACE WHERE landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

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

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

SUCCESS BOOSTERS

```
%%sql

select count(*) as "Total", mission_outcome from SPACE group by mission_outcome
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

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

The total number of successful and failure mission outcomes was calculated

SUCCESS AND FAILURE

```
%%sql

SELECT distinct booster_version as "Booster Version" FROM SPACE where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACE)
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

Booster Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

The names of the booster versions which have carried the maximum payload mass were calculated

BOOSTER VERSIONS

```
%%sql

SELECT distinct Landing__Outcome, Booster_Version, Launch_Site from SPACE where Landing__Outcome='Failure (drone ship)' and DATE LIKE '2015%'
```

```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 were calculated

FAILED LANDING OUTCOMES

```
%%sql

SELECT Landing__Outcome, count(*) as "Total" from SPACE where DATE between '2010-06-04' and '2017-03-20'group by  Landing__Outcome order by 2 desc
```

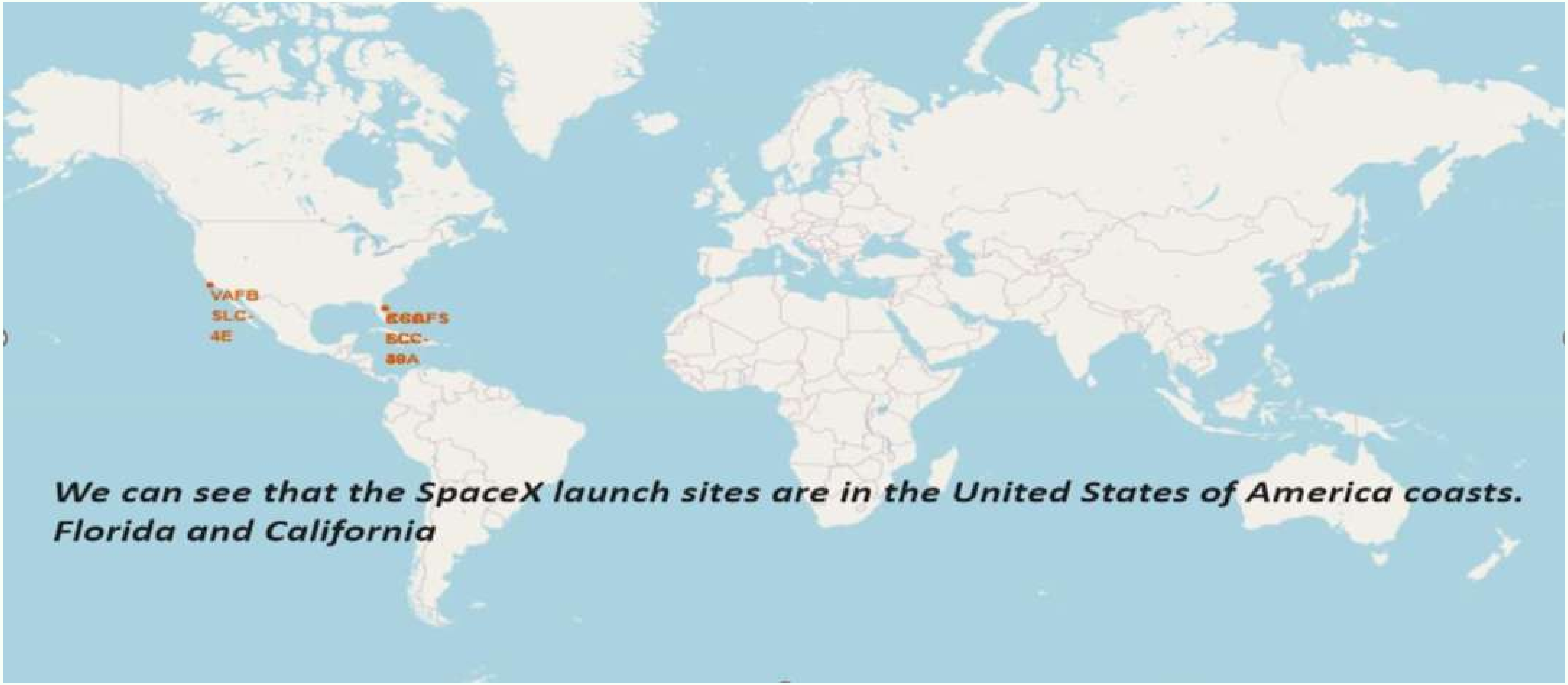
```
 * ibm_db_sa://hxr06276:***@h1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/bludb
Done.
```

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

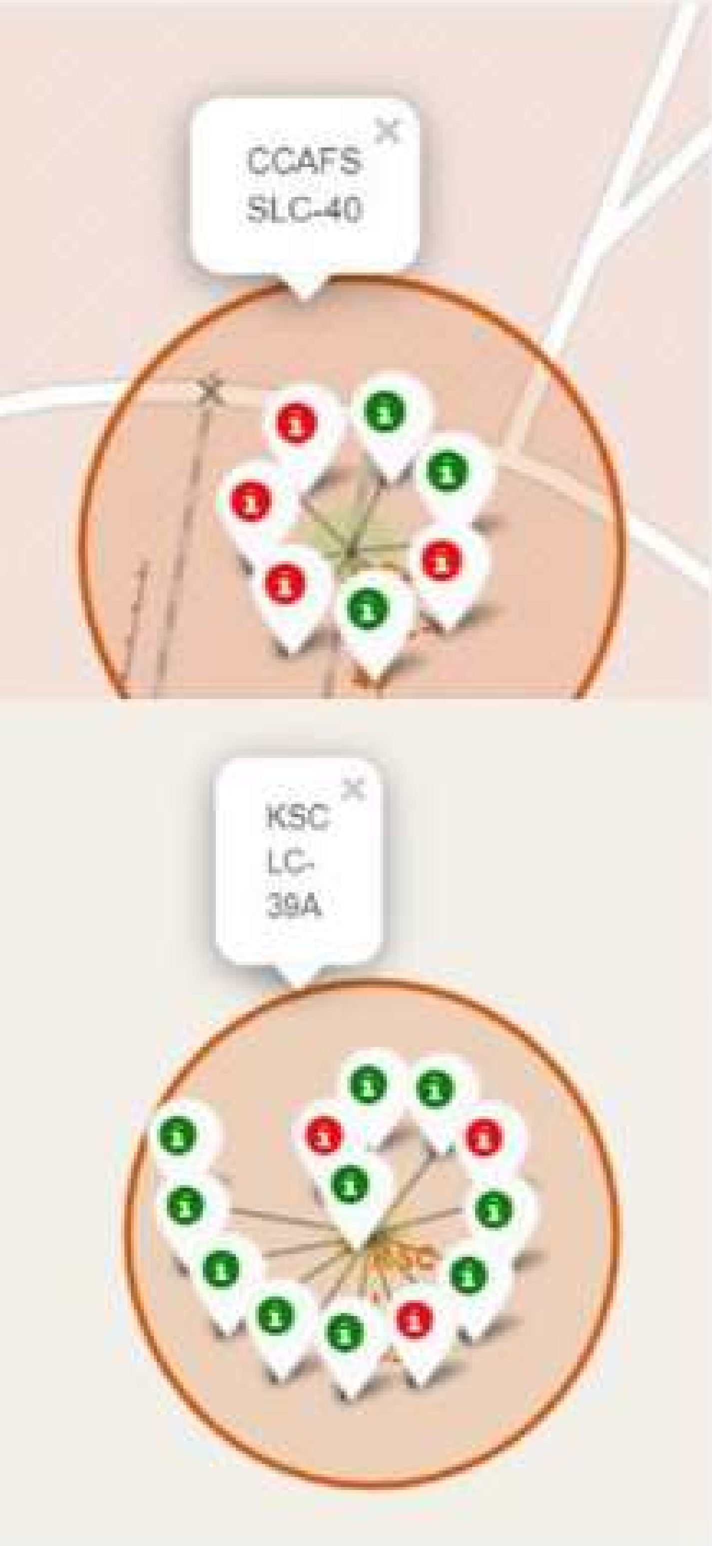
Was ranked the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

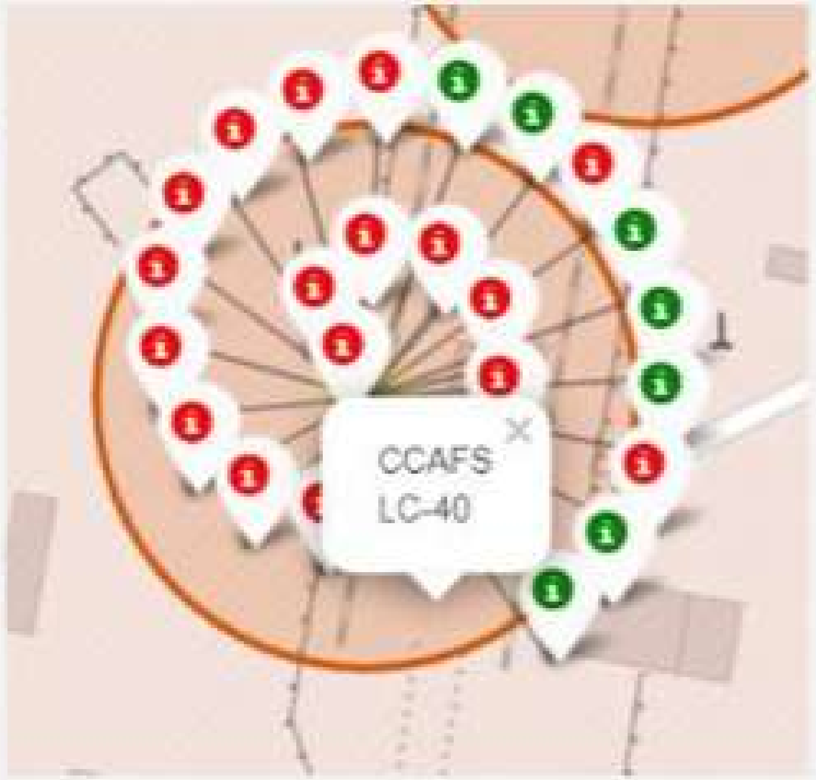
LANDING OUTCOMES

LAUNCH SITES WITH FOLIUM



LAUNCH SITES IN GLOBAL MAP



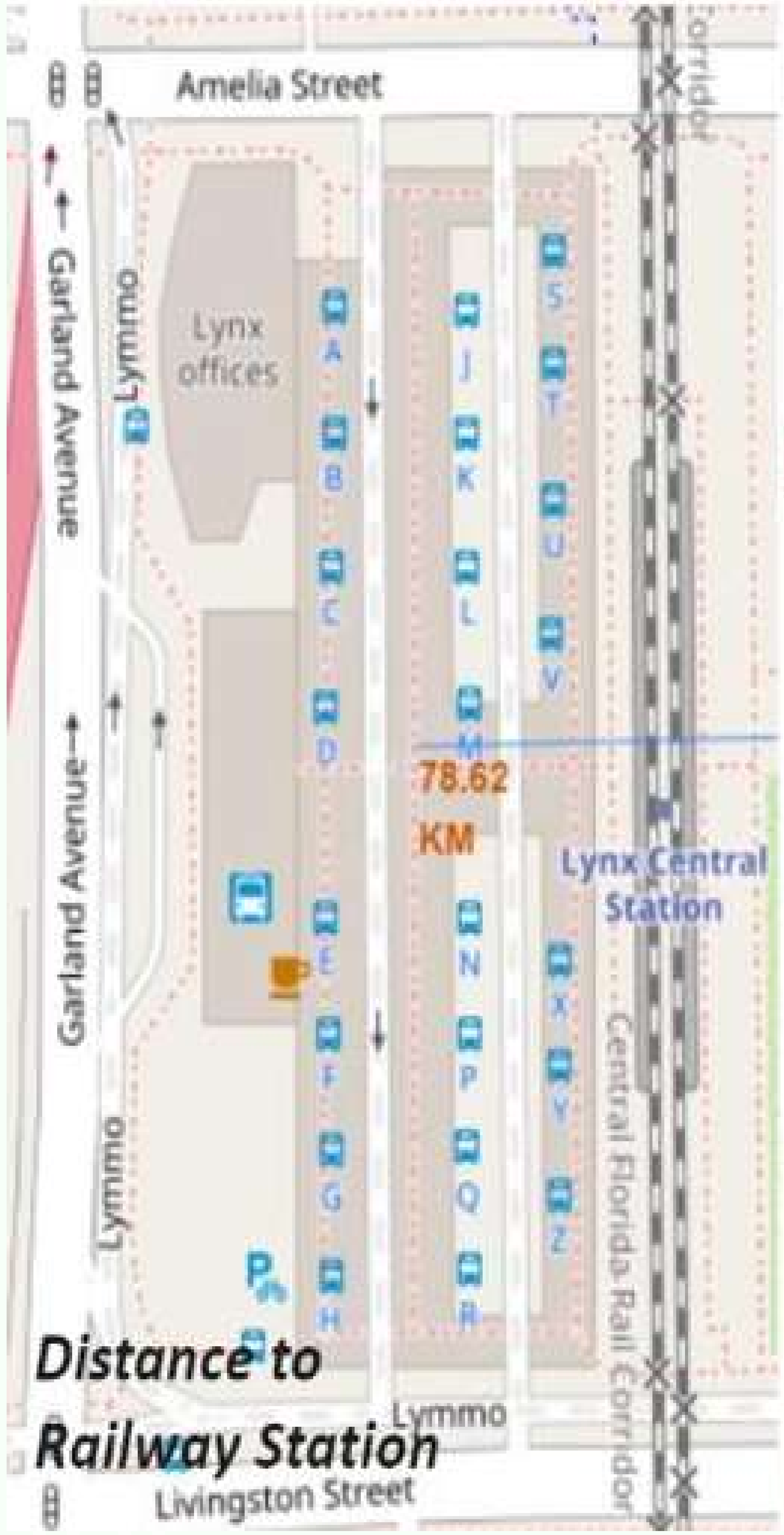


Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

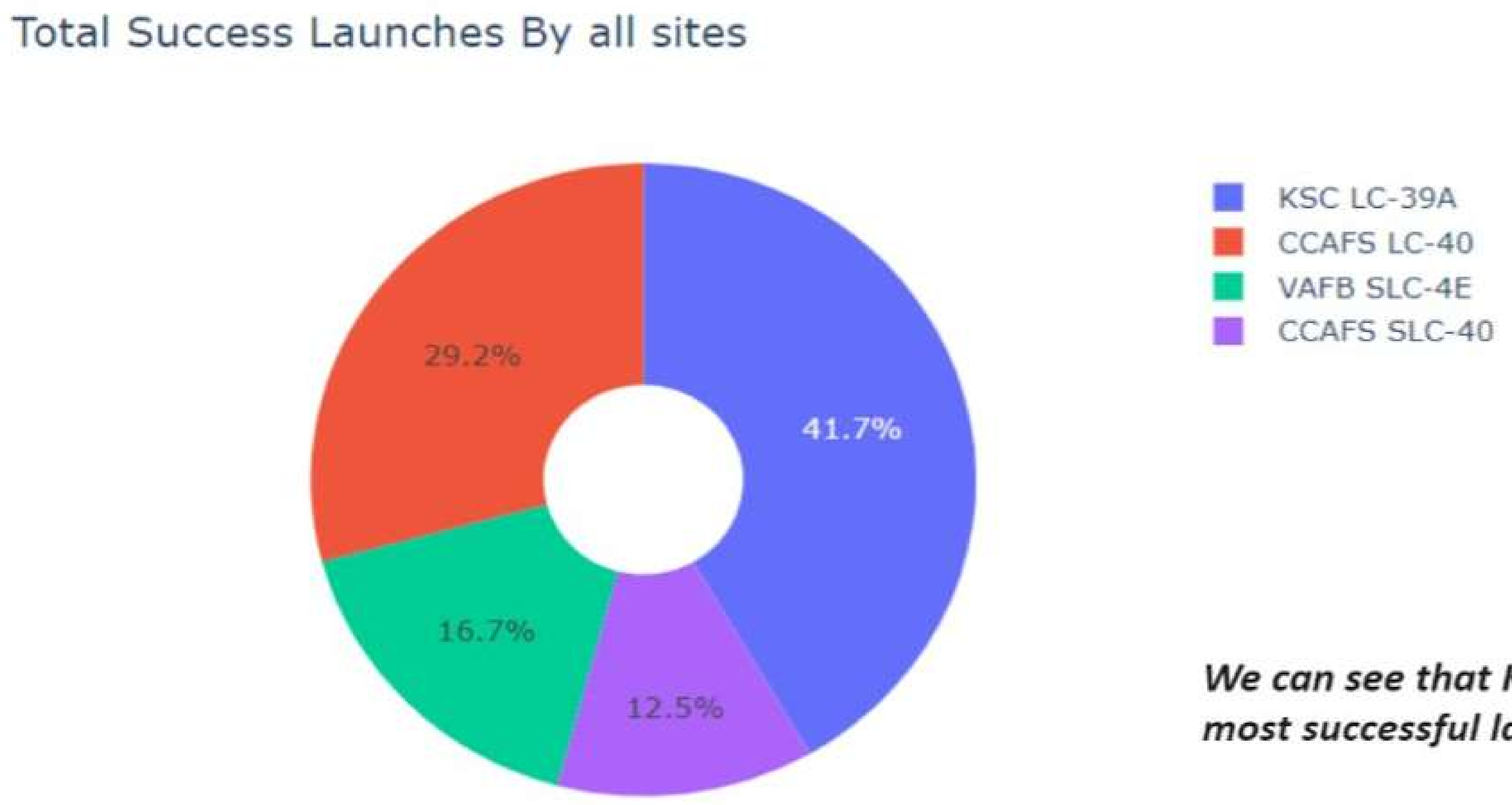
MARKERS WITH COLORS LAUNCH SITES



- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

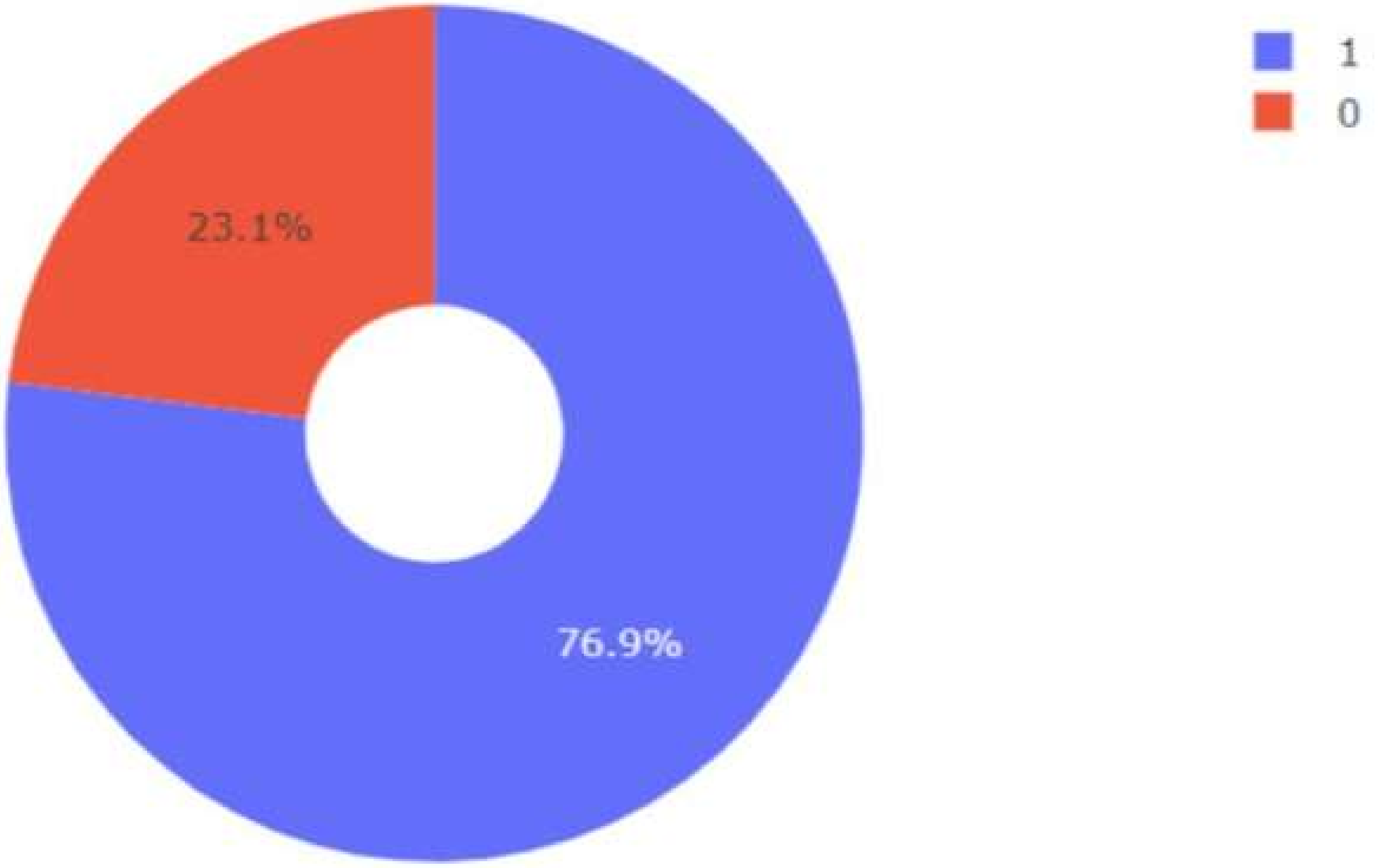
LAUNCH SITE DISTANCES

INTERACTIVE DASHBOARD



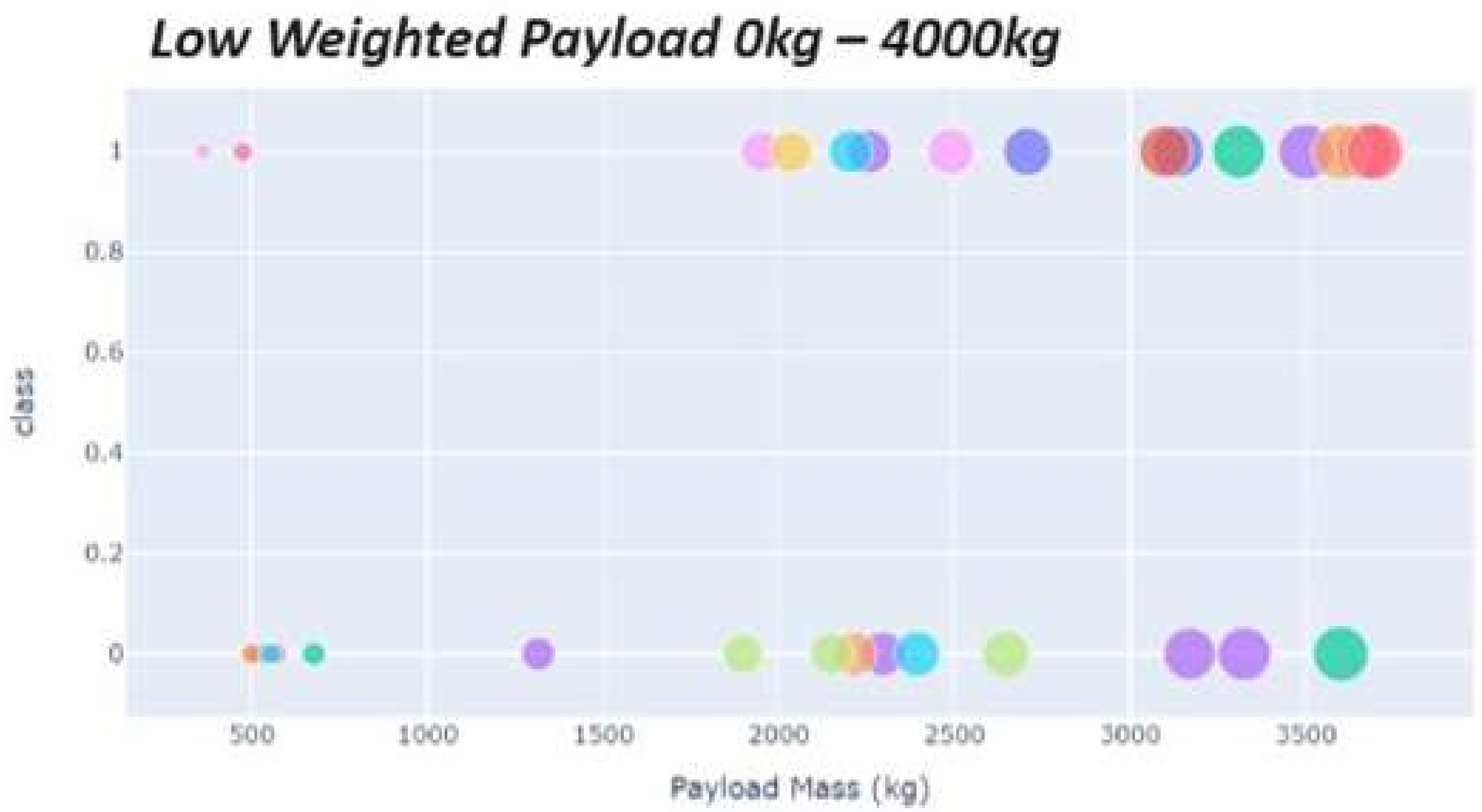
We can see that KSC LC-39A had the most successful launches from all the sites

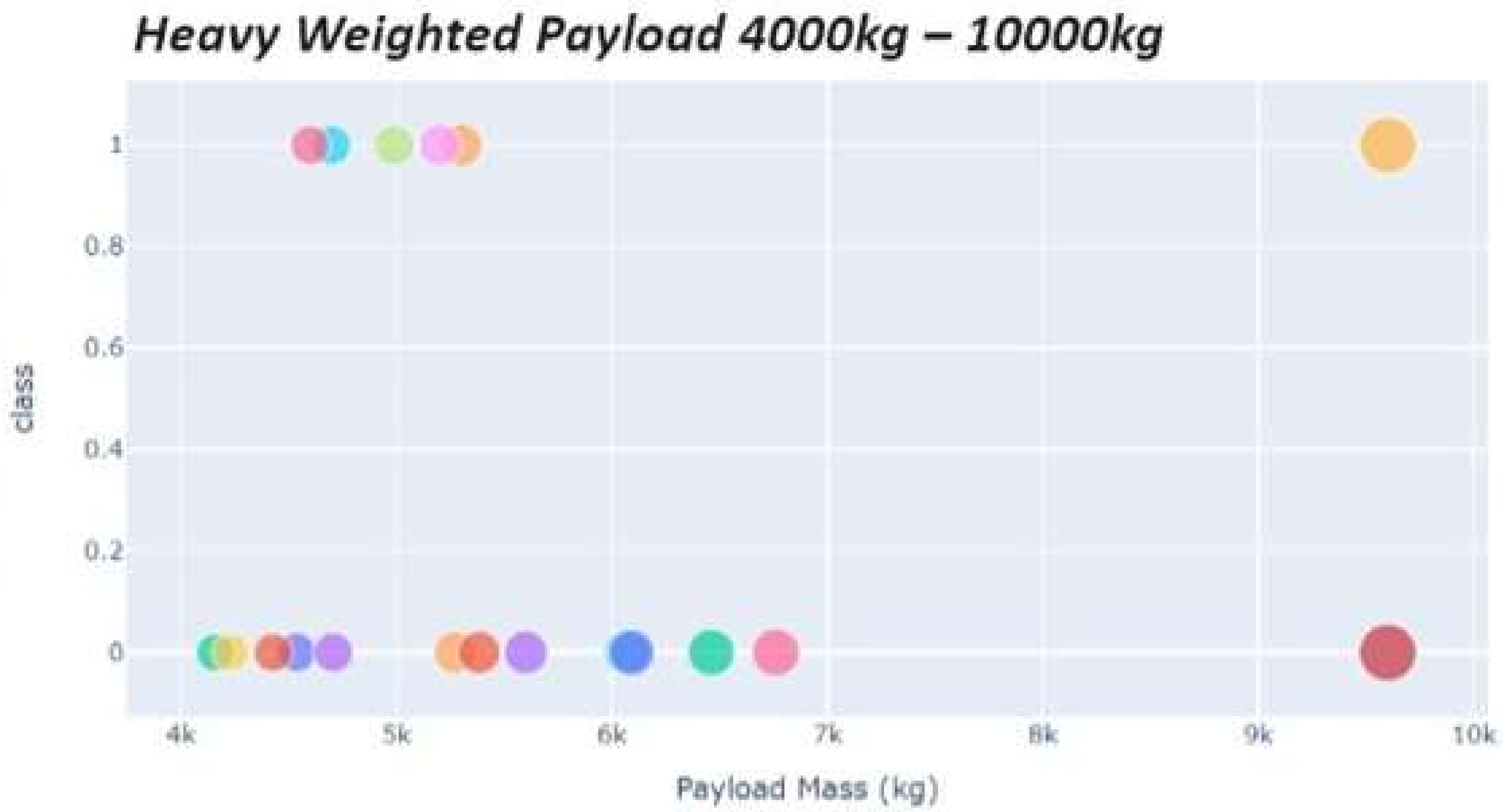
SUCCESS RATE BY LAUNCH SITE



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

LAUNCH SITES IN GLOBAL MAP





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

PREDICTIVE ANALYSIS

Create a decision tree classifier object then create a `GridSearchCV` object `tree_cv` with cv = 10. Fit the object to find the best parameters from the dictionary `parameters`.

```
parameters = {'criterion': ['gini', 'entropy'],
     'splitter': ['best', 'random'],
     'max_depth': [2*n for n in range(1,10)],
     'max_features': ['auto', 'sqrt'],
     'min_samples_leaf': [1, 2, 4],
     'min_samples_split': [2, 5, 10]}

tree = DecisionTreeClassifier()
```

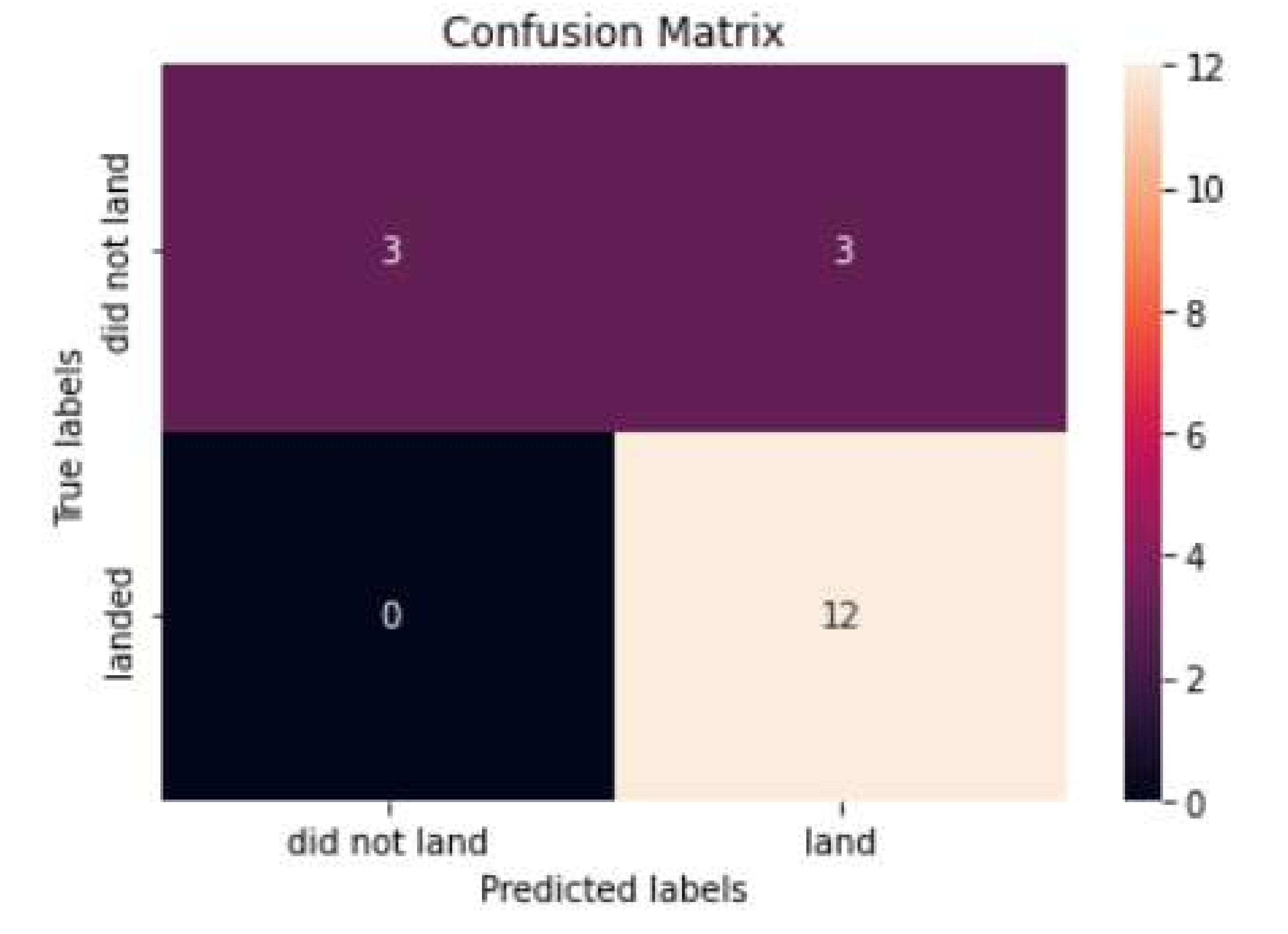
```
grid_search = GridSearchCV(tree, parameters, cv=10)
tree_cv = grid_search.fit(X_train, Y_train)
```

```
print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
print("accuracy :",tree_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
accuracy : 0.8892857142857142
```

The decision tree was the predictive model with the highest accuracy rate, with 88,92%

CLASSIFICATION ACCURACY



The confusion matrix got a good result. The only problem is the high amount of false positives.

CONFUSION MATRIX

- The more flights that take place at a launch site, the higher the likelihood of success
- From 2013 to 2020, there was a steady rise in the success rate of launches.
- Certain orbits, such as ES-L1, GEO, HEO, SSO, and VLEO, had particularly high success rates.
- The Kennedy Space Center's LC-39A launch site had the most successful launches of any site.
- A decision tree classifier is the best type of machine learning algorithm for analyzing this data.

CONCLUSIONS

- **COMPLETE NOTEBOOK LINK:**

  HTTPS://GITHUB.COM/ELPITTA/MYREPOSITORY/BLOB/MAIN/IBM%20PROFESSIONAL%20DATA%20SCIENCE/10.%20APPLIED%20DATA%20SCIENCE%20CAPSTONE/FINAL%20PROJECT/PROJECT%20SPACEX.IPYNB

- **GITHUB LINK:**

  HTTPS://GITHUB.COM/ELPITTA/MYREPOSITORY/TREE/MAIN

MORE INFORMATIONS

THANK YOU!!