IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Janith Silva
2023/07/13

Outline

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

Executive Summary

Summary of methodologies

The primary objective of this research is to predict if the Falcon 9 first stage will land successfully. The following methodologies were employed in this research.

- Data collection
- Data wrangling
- Exploratory Analysis Using SQL
- Exploratory Analysis Using Pandas and Matplotlib
- Analyzing launch records interactively with Plotly Dash
- Predictive Analysis (Classification)

Summary of all results

In summary, the success rate of launches has shown improvement over time, with certain orbits like ES-L1, GEO, HEO, and SSO achieving a 100% success rate. Test results indicate comparable performance among different models, with the decision tree model showing a slight advantage. Launch sites are strategically located near the equator to benefit from the Earth's rotational speed, resulting in cost savings and efficiency. All launch sites are located close to coastal areas. Among them, KSC LC-39A stands out with the highest success rate.

Introduction

Project background and context

The commercial space age is here, companies are making space travel affordable for everyone. Perhaps the most successful is SpaceX. 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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

Questions to be answered

- How payload mass, launch site, number of flights, and orbits affect first-stage landing?
- How the rate of successful landings behaves over time?
- Which model performs best in identifying a successful landing?

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

- Data was collected using SpaceX REST API and web scraping related wiki pages

Perform data wrangling

- By filtering the data, handling missing values and one hot encoding categorical features

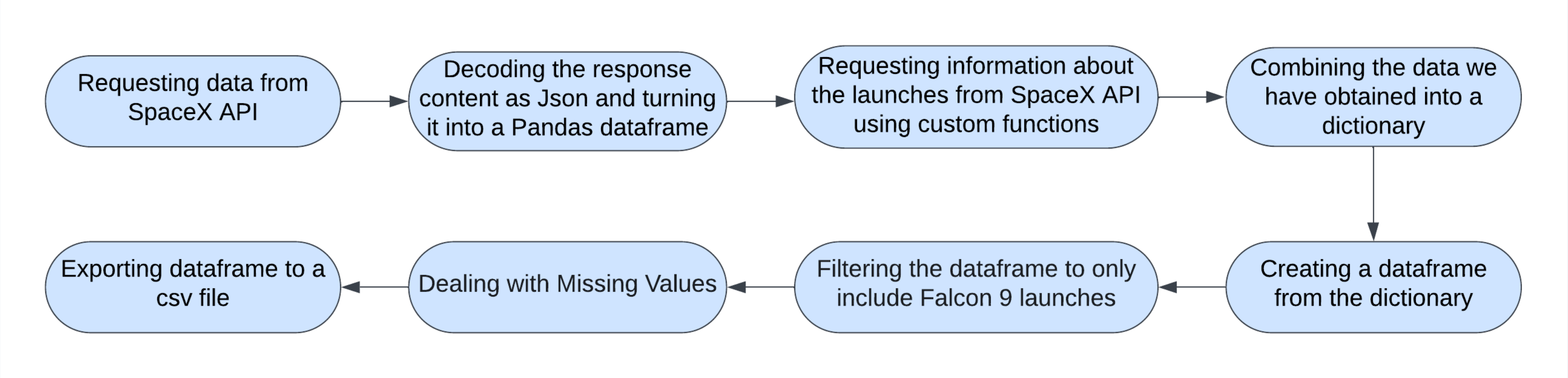
Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

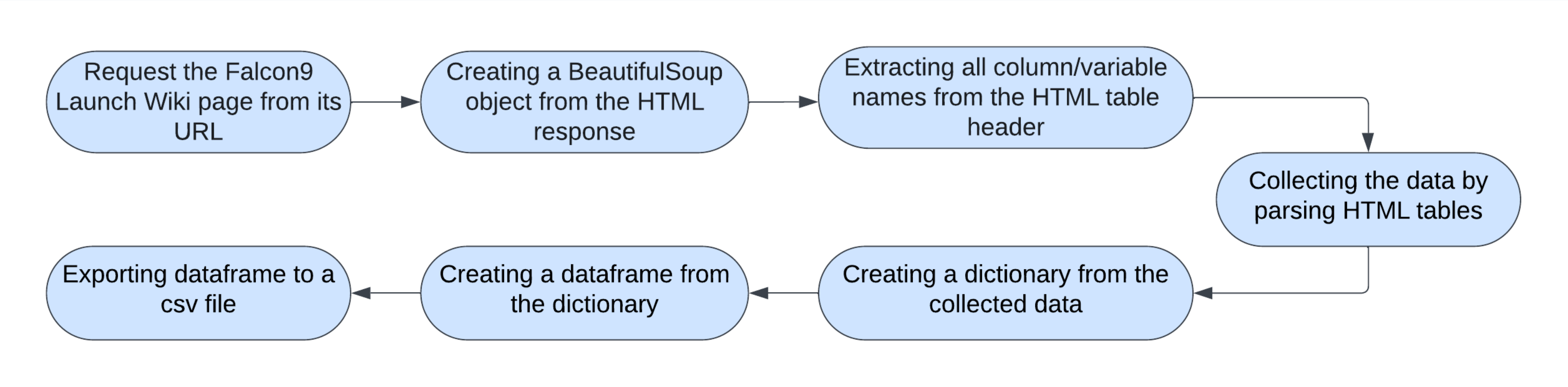
- Building models involves training machine learning classification models using the data set, tuning hyper parameters and evaluating the performance.

Data Collection – SpaceX API



SpaceX API calls notebook

Data Collection - Scraping



[Web scraping notebook]

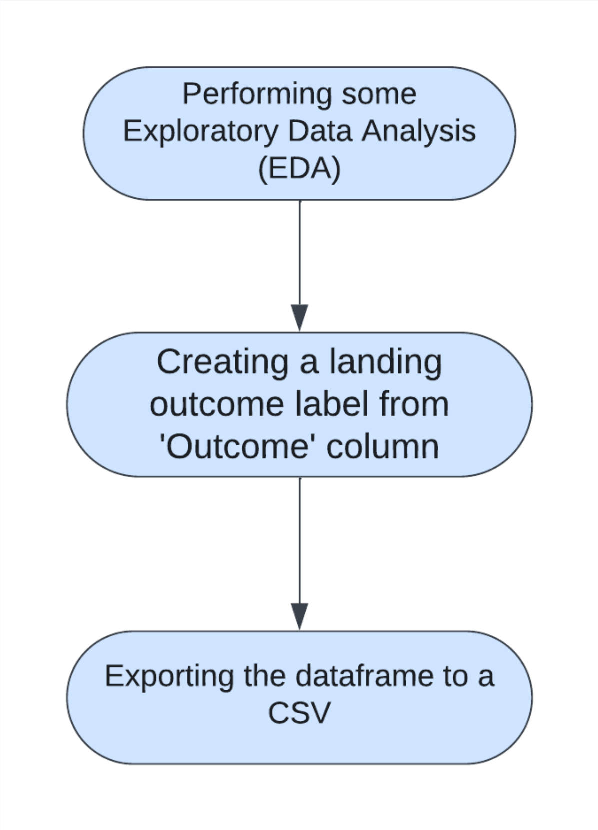
Data Wrangling

Performing Exploratory Data Analysis (EDA) involves

- Calculating the number of launches on each launch site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type

Creating a landing outcome label from Outcome column

Using the column ‘Outcome’ creating a new column ‘landing_class’ where the column element is zero if the corresponding row in ’Outcome’ is in the set bad outcome otherwise, it's one. If the value is zero, the first stage did not land successfully. One means the first stage landed successfully.



Data wrangling notebook

EDA with SQL

SQL queries performed

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was acheived.
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass. Use a subquery
- Listing the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- Ranking 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 notebook

Build an Interactive Map with Folium

Marked all launch sites on a map

- Created a blue circle at NASA Johnson Space Center's coordinate with a popup label showing its name using its latitude and longitude coordinates
- Added a red circle object for each launch site, based on its coordinate (Lat, Long) values. In addition, added launch site name as a popup label.

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

- Added colored markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

Calculated the distances between a launch site to its proximities

- Added colored lines to show distances between the launch site CCAFS SLC-40 (as an example) and its proximities like railway, highway, coastline and closest city.

Interactive Map with Folium notebook

Build a Dashboard with Plotly Dash

Launch sites dropdown list

- Added a dropdown list to enable Launch Site selection

Pie chart showing success launches in launch sites

- Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected

Slider of payload mass range

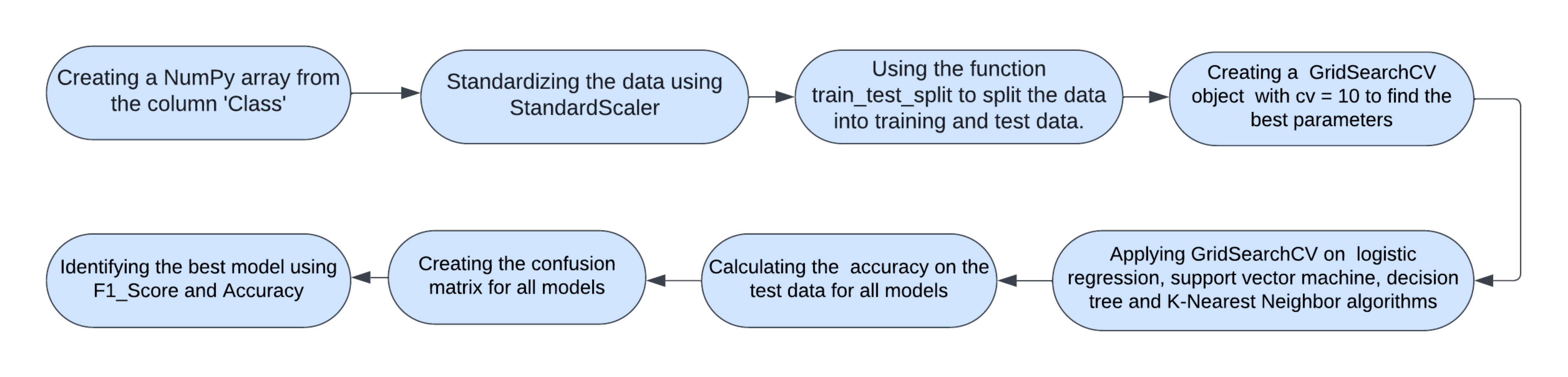
- Added a slider to select Payload range

Scatter chart of payload mass vs. success rate for the different booster versions

- Added a scatter chart to show the correlation between Payload and Launch Success

Dashboard with Plotly Dash

Predictive Analysis (Classification)



Predictive analysis notebook

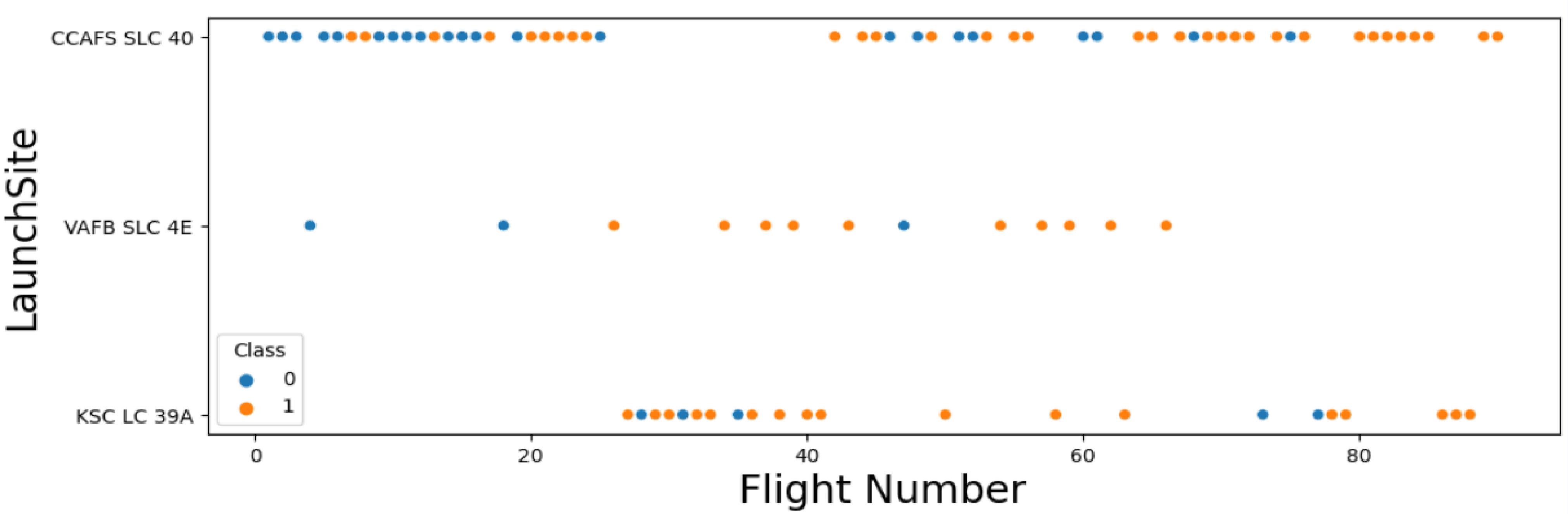
Results

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

Section 2

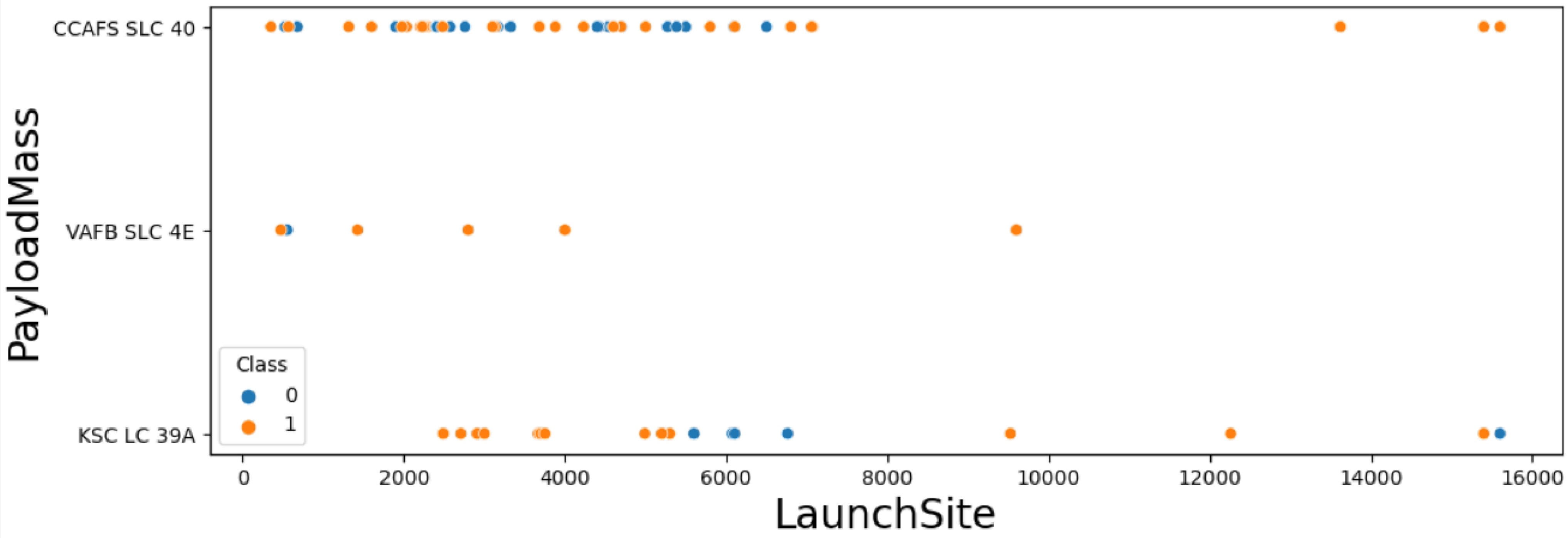
Insights drawn from EDA

Flight Number vs. Launch Site



- Around half of launches were from CCAFS SLC 40 launch site
- VAFB SLC 4E and KSC LC 39A have comparatively higher success rates
- Earlier flights had a lower success rate
- The success rate have increased over time

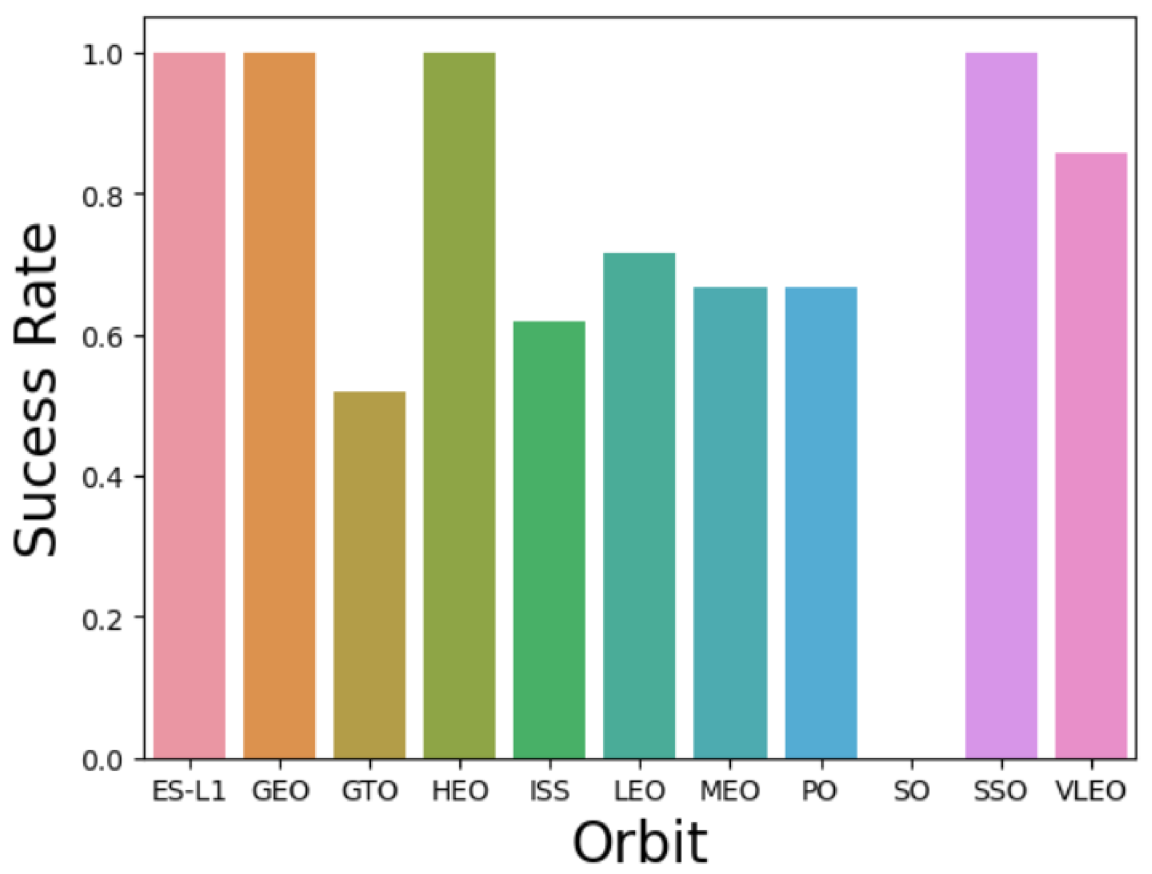
Payload vs. Launch Site



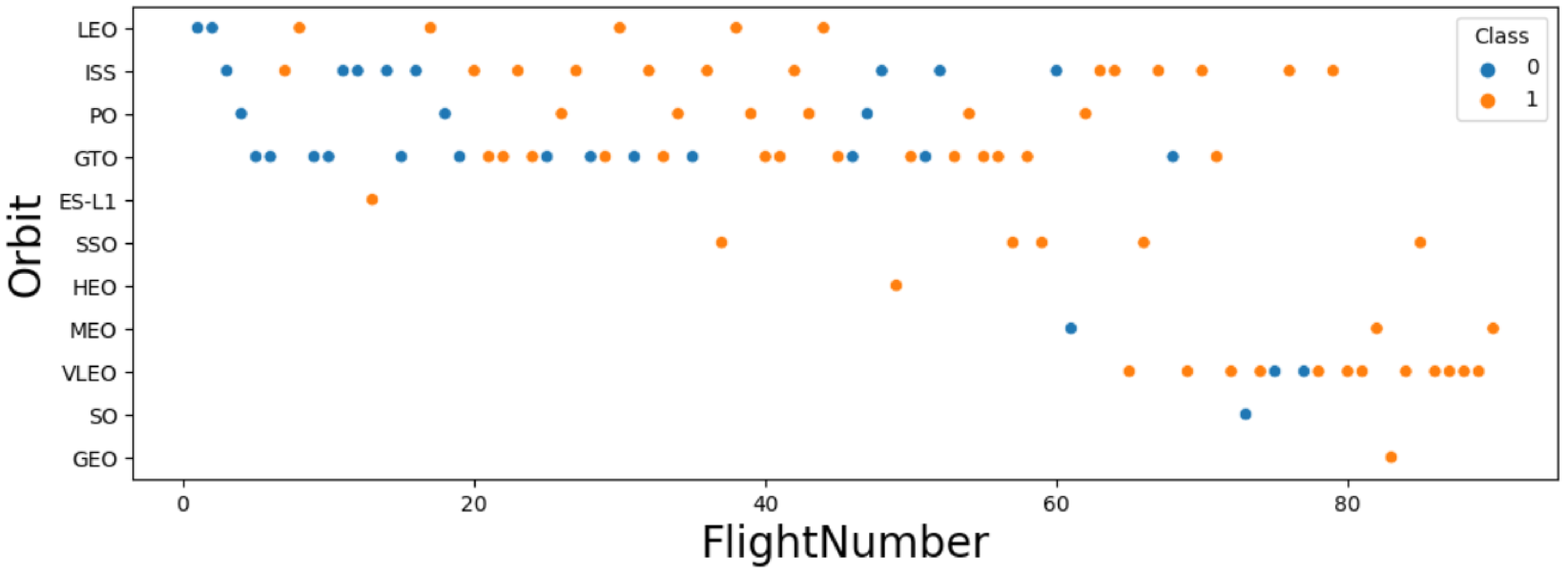
- We can see that higher the payload mass (kg), the higher the success rate
- VAFB-SLC launch site has not launched rockets for heavy payload mass (greater than 10000)
- Most launces with a payload greater than 7,000 kg were successful
- KSC LC 39A has a 100% success rate for launches less than 5000kg

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO orbits have 100% success rate
- GTO, ISS, LEO, MEO, PO orbits have success rate between 50% and 85%
- SO orbit have 0% success rate

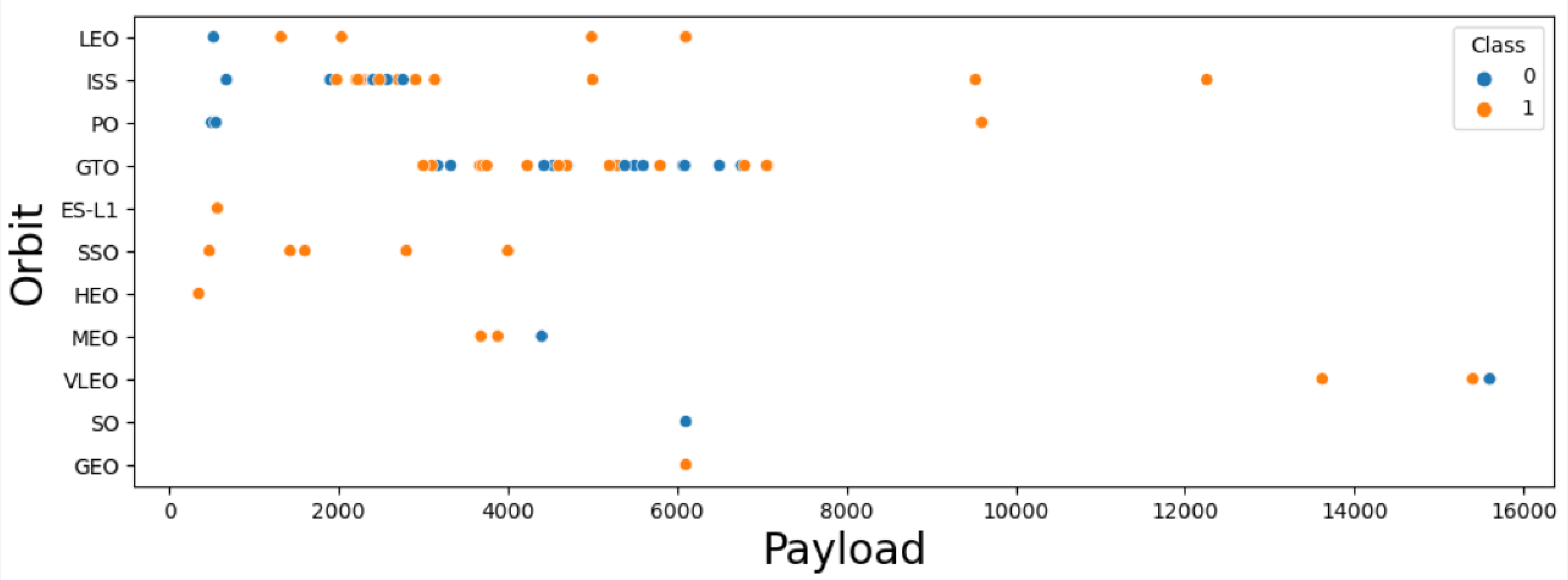


Flight Number vs. Orbit Type



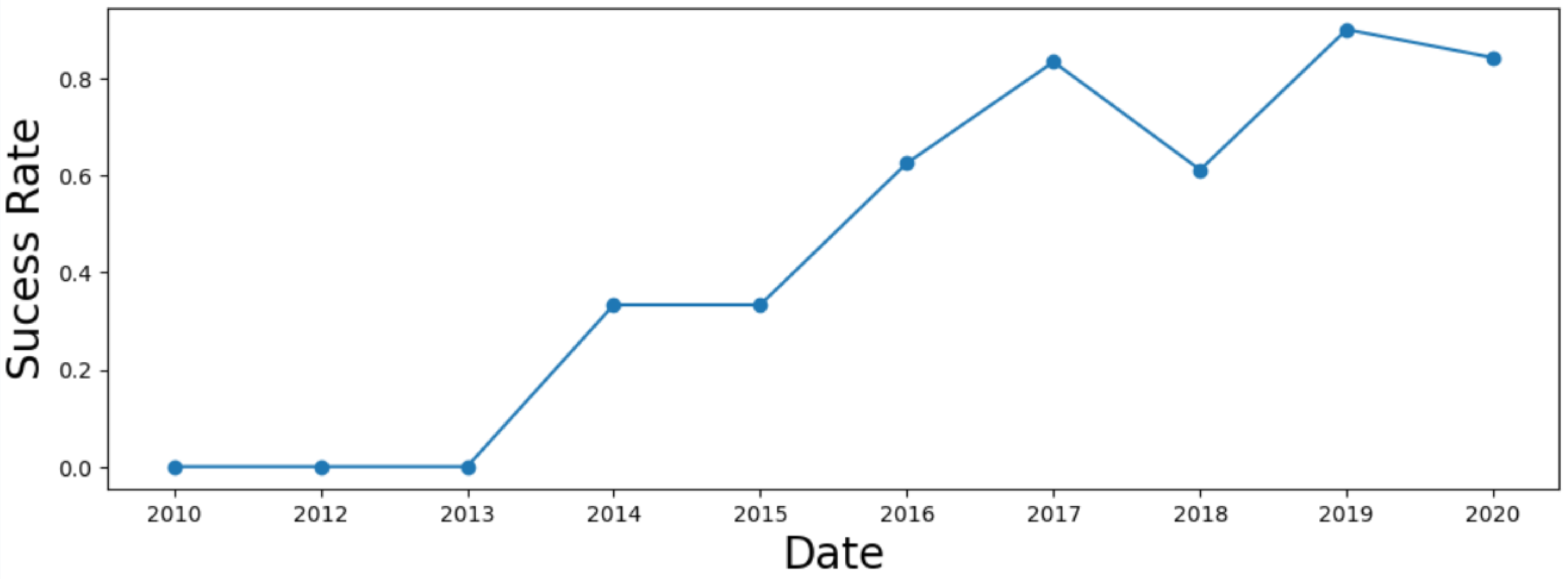
LEO orbit the Success appears related to the number of flights. on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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



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

All Launch Site Names

```
[21]: %sql select DISTINCT Launch_Site from SPACEXTBL;
 * sqlite:///my_data1.db
Done.
```

[21]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

- Displaying the names of the unique launch sites in the space mission

Launch Site Names Begin with 'CCA'

```
[9]: %%sql
SELECT * FROM SPACEXTBL
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

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

[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

- Displaying 5 records where launch sites begin with the string 'CCA'

Total Payload Mass

```
[10]: %%sql
SELECT SUM(PAYLOAD_MASS__KG_)as 'total payload mass' FROM SPACEXTBL
WHERE Customer='NASA (CRS)';

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

[10]:

total payload mass
45596.0

- Displaying the total payload mass carried by boosters launched by NASA (CRS).

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)as 'average payload mass' FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

average payload mass
2534.6666666666665

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

First Successful Ground Landing Date

```
[12]: %%sql
SELECT min(substr(Date, 7, 4 ) || '-' || substr(Date, 4, 2 ) || '-' || substr(Date, 1, 2 )) AS Date FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';

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

[12]:

Date
2015-12-22

- Listing the date when the first successful landing outcome in ground pad was achieved

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[13]: %%sql
SELECT DISTINCT Booster_Version as 'Booster' FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

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

[13]:

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

- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000kg but less than 6000kg

Total Number of Successful and Failure Mission Outcomes

[18]:
```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) as 'total number' FROM SPACEXTBL GROUP BY Mission_Outcome ;
```

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

[18]:

Mission_Outcome	total number
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Listing the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
[15]: %%sql
SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

[15]:

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

- Displaying the names of the booster versions which have carried the maximum payload mass

2015 Launch Records

```
[16]: %%sql
SELECT substr(Date, 4, 2) AS month, date, booster_version, launch_site, Landing_Outcome FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date,7,4)='2015';
```

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

[16]:

month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	01/10/2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14/04/2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Displaying the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

[17]:

```
%%sql
SELECT  Landing_Outcome, count(*) AS 'Count' FROM SPACEXTBL
WHERE Date BETWEEN '04/06/2010'AND '20/03/2017'
GROUP BY Landing_Outcome
ORDER BY COUNT DESC;
```

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

[17]:

Landing_Outcome	Count
Success	20
No attempt	9
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

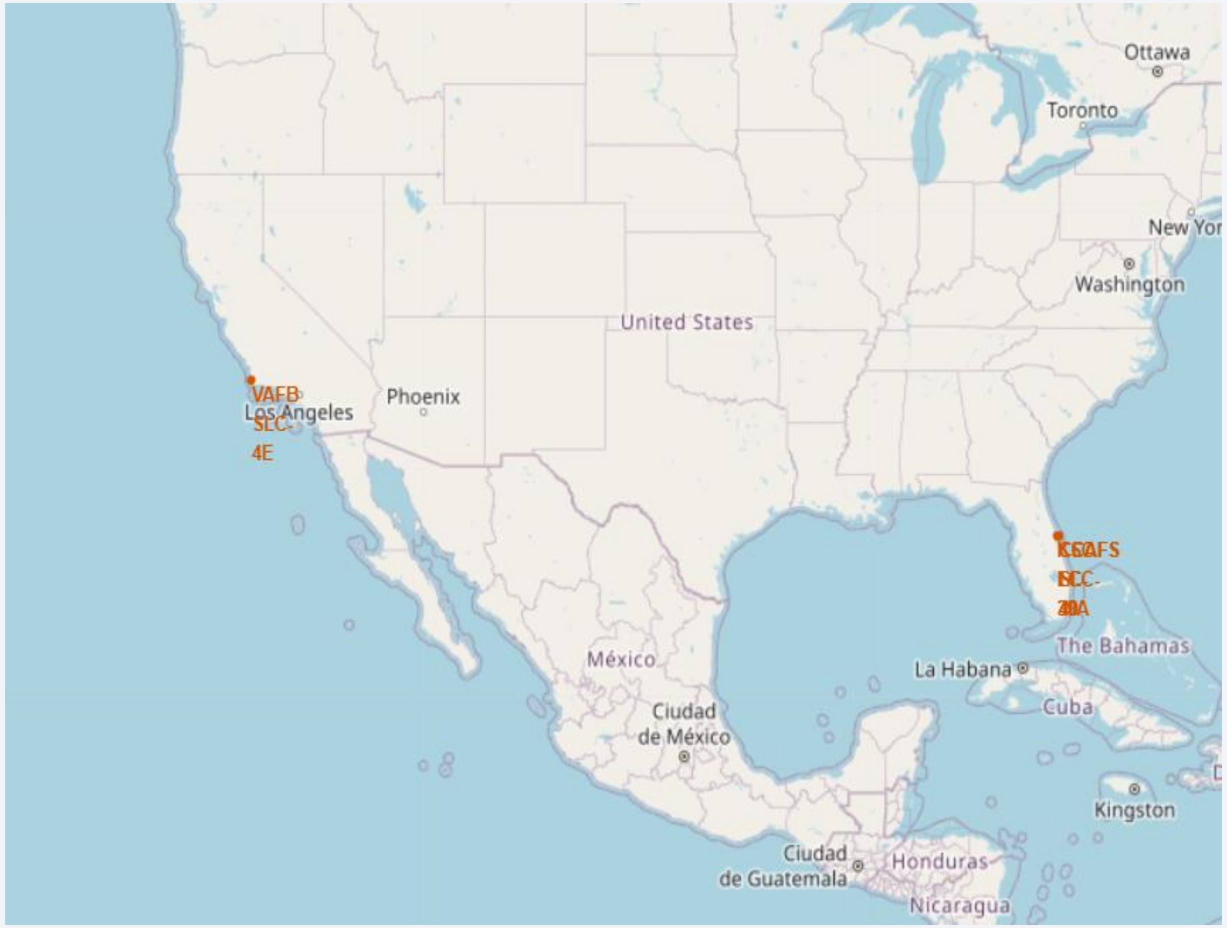
Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Section 3

Launch Sites Proximities Analysis

All launch sites’ location markers on the global map

Most of Launch sites are in proximity to the Equator line. Launch sites near the Equator provide certain advantages for space launches due to the Earth's rotation. The rotational speed of the Earth is highest at the Equator, which can provide an additional velocity boost to rockets during launch. This is particularly beneficial for missions requiring a significant payload or missions aiming to reach geostationary orbits. All launch sites are in very close proximity to the coast, Launching rockets over open water reduces the risk to populated areas in case of an accident or failure during launch. In the event of an emergency, it is safer to have the rocket and its debris fall into the ocean rather than over land.

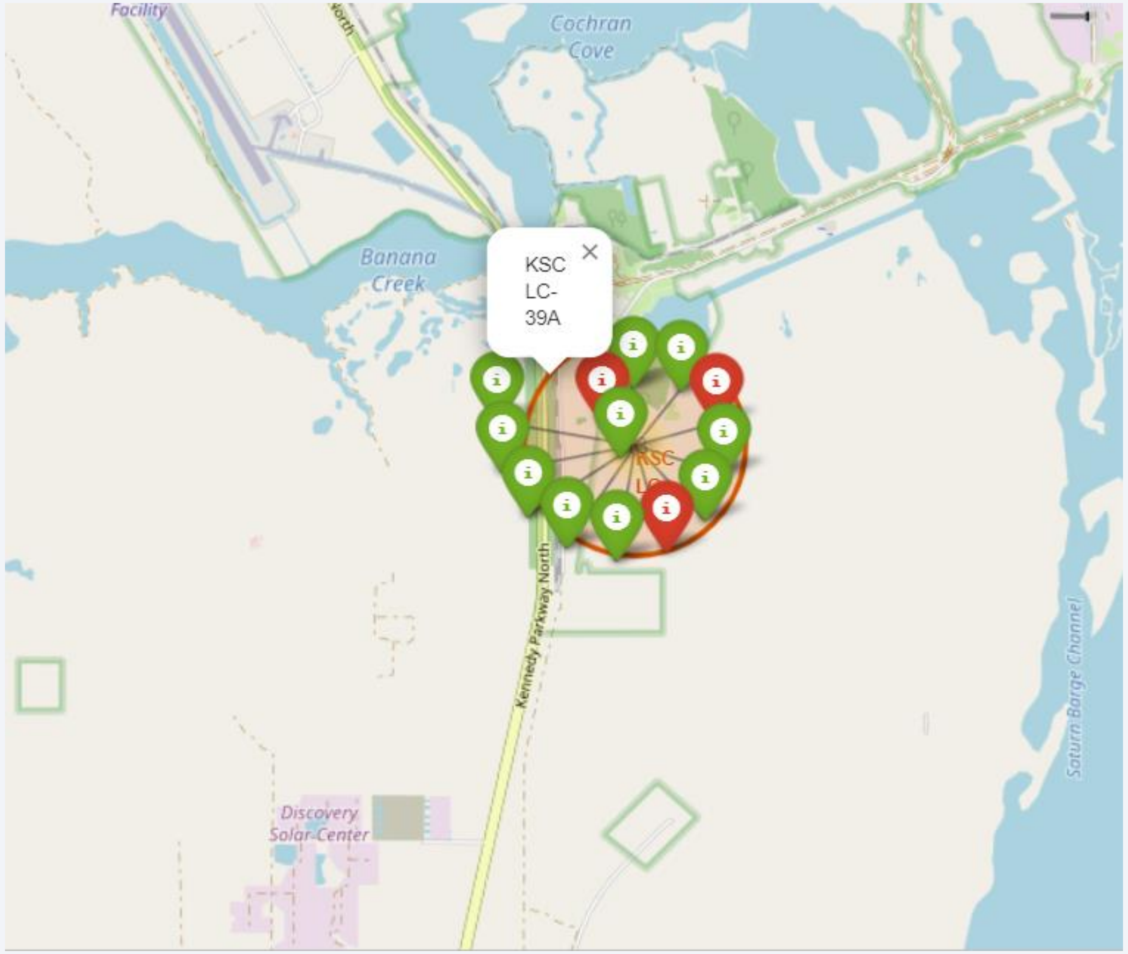


Color labeled launch outcomes on the map

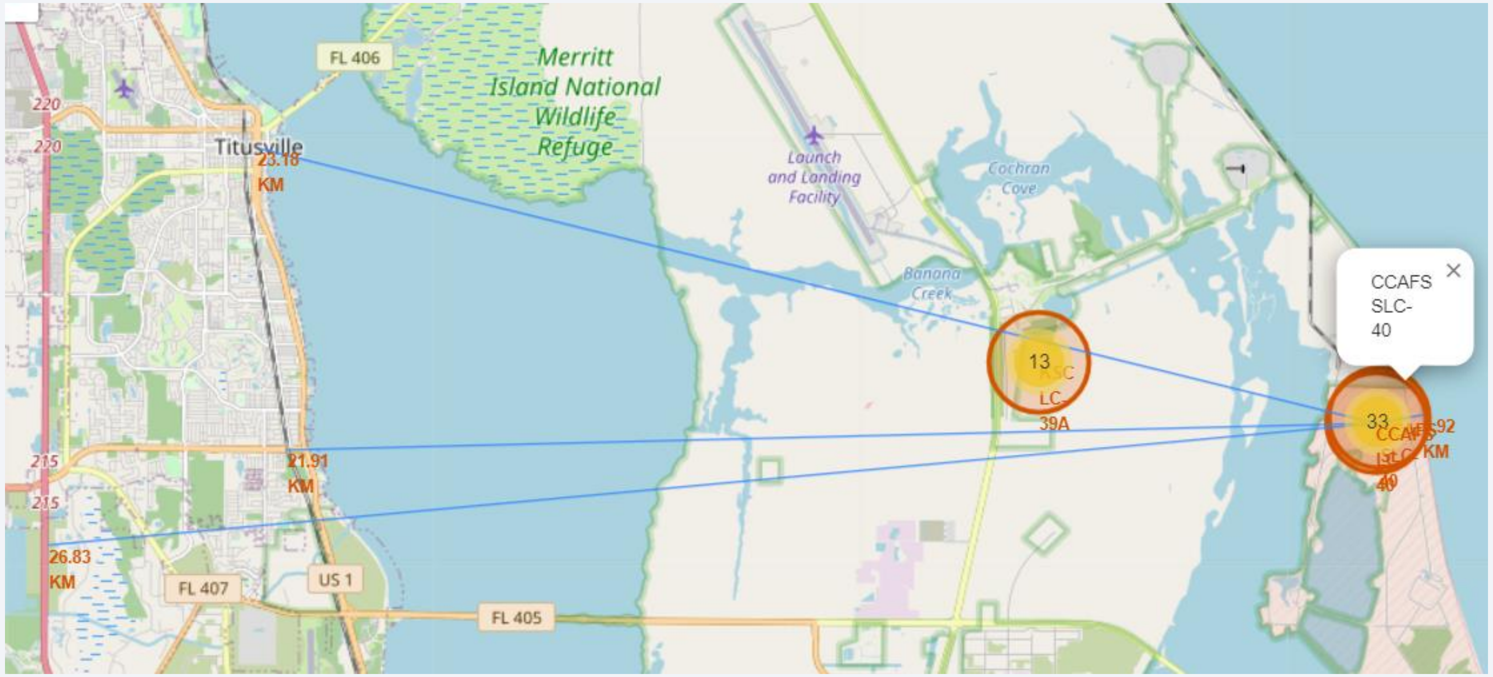
From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

- Green marker indicates a successful launch
- Red marker indicates a failed launch

Launch site KSC LC-39A has a very high success rate.



Distance from the launch site CCAFS SLC-40 to its proximities

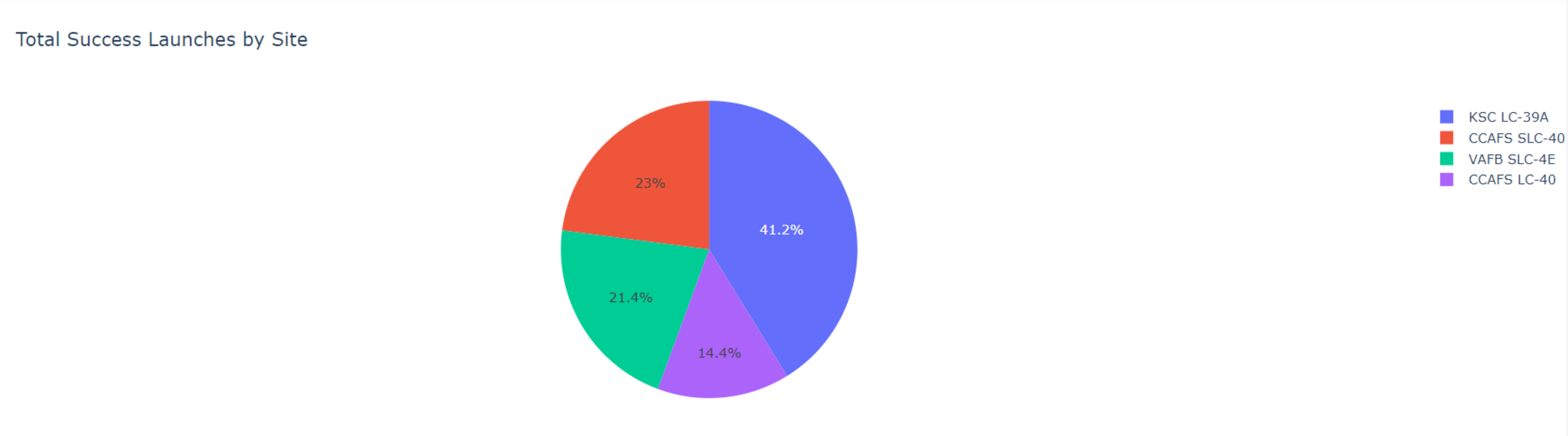


We can clearly see that In close proximity to the launch site, there are several key features including a city, a highway and a railway. Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to these populated areas.

Section 4

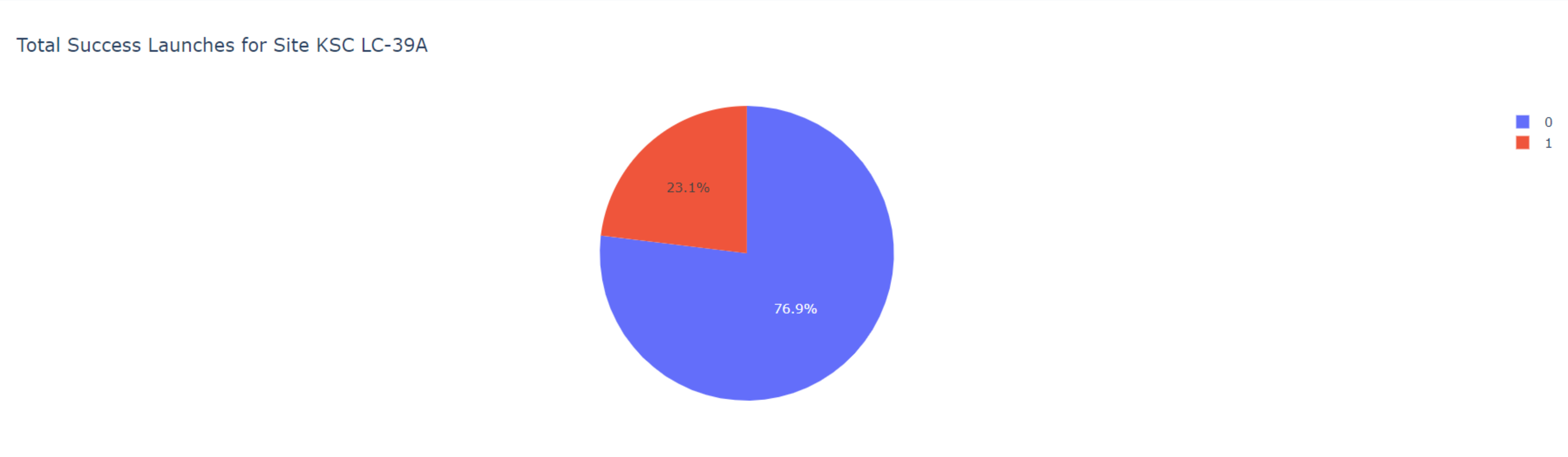
Build a Dashboard with Plotly Dash

Launch success count for all sites



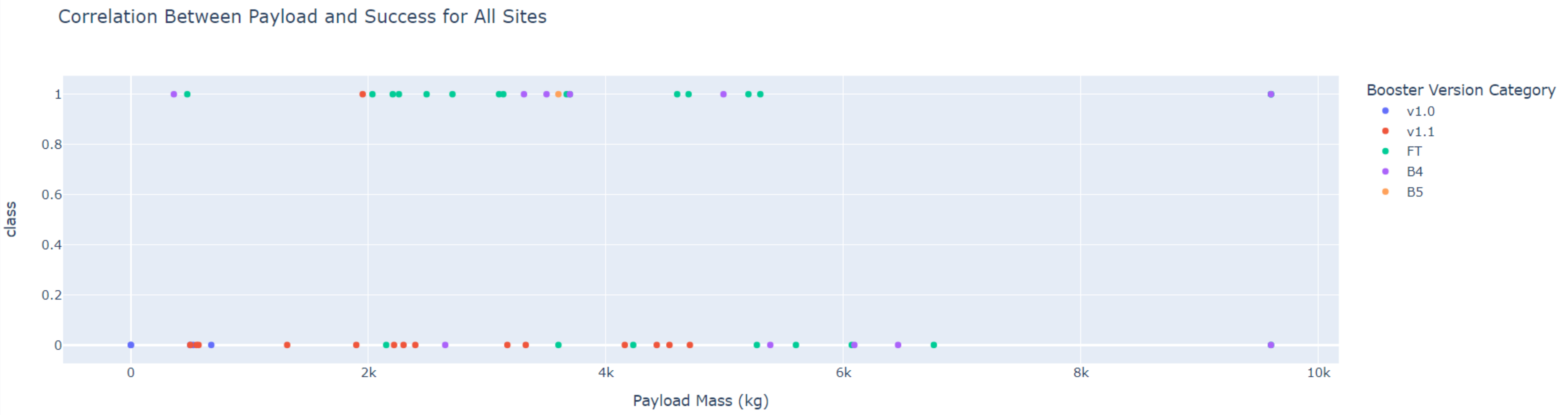
We can clearly see that from all the sites, KSC LC-39A has the most successful launches.

Launch site with highest launch success ratio



- KSC LC-39A has the highest launch success rate which is 76.9%

Payload mass vs. Launch Outcome scatter plot for all sites



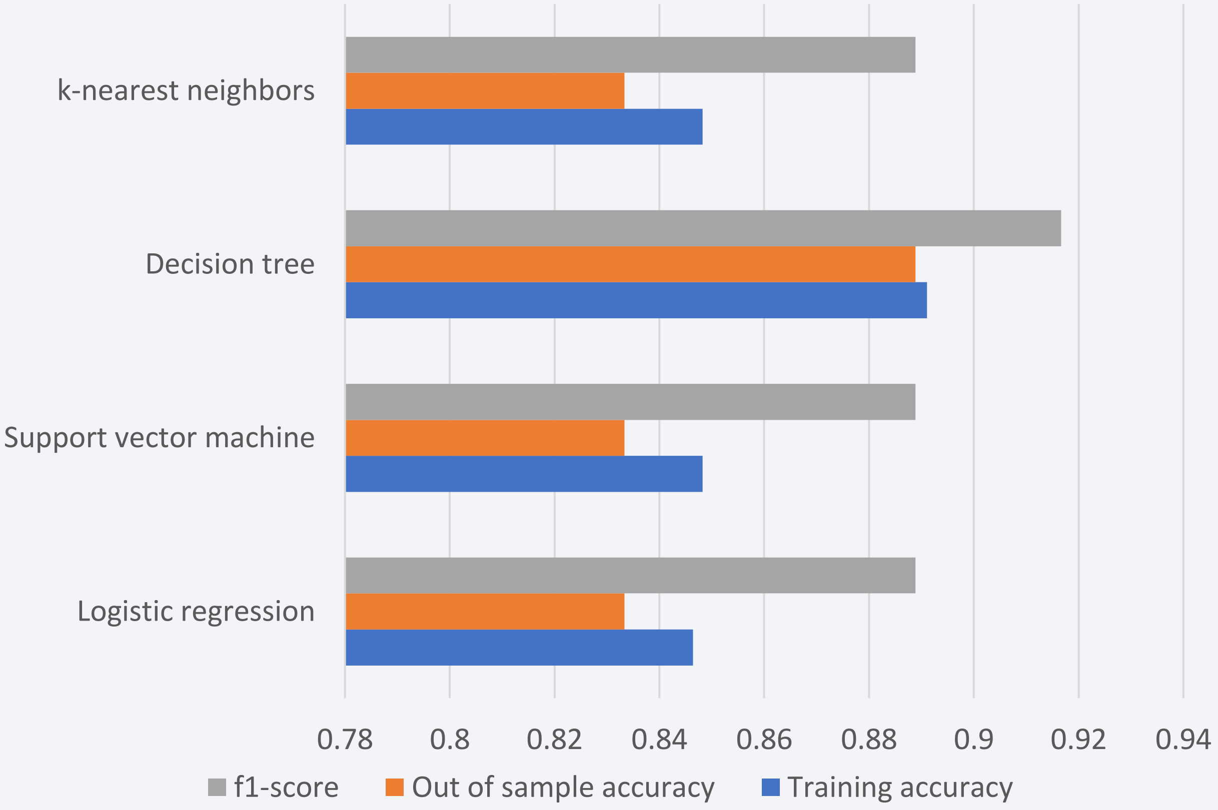
Payloads between 2,000 kg and 5,000 kg have the highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

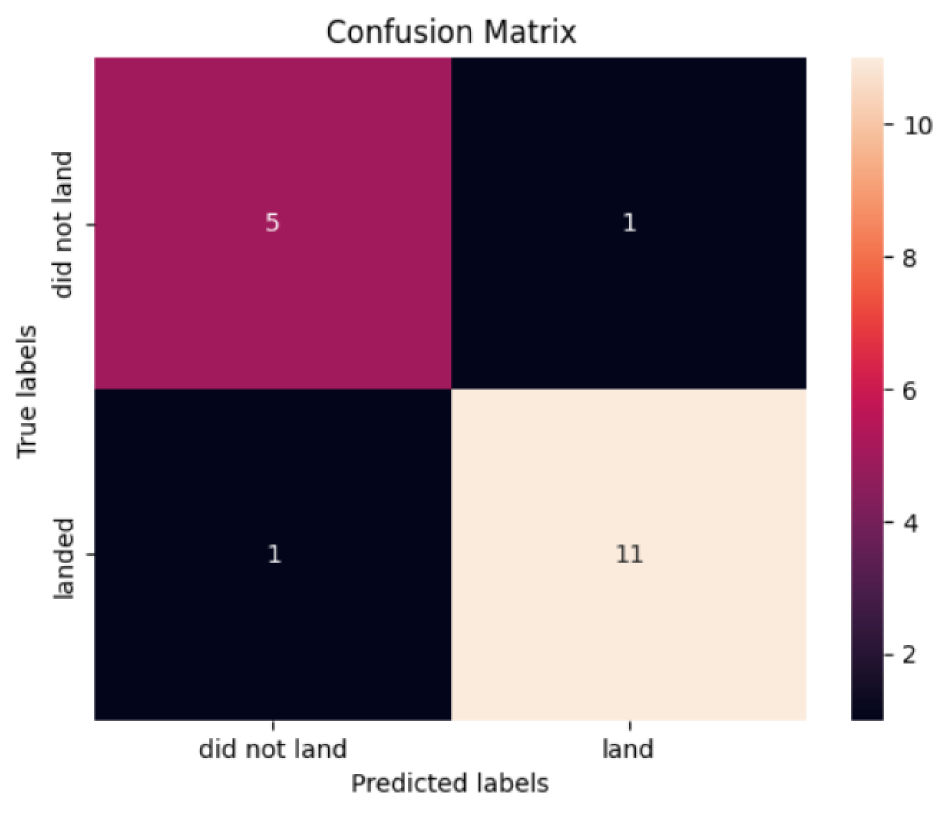
All the models exhibited similar performance, with comparable scores and accuracy. However, the Decision Tree model demonstrated a slight superiority, surpassing the others in terms of performance.



Confusion Matrix

Confusion Matrix Outputs

- 11 True positive predictions
- 5 True negative predictions
- 1 False positive predictions
- 1 False Negative predictions



Conclusions

- The success rate of launches have increased over the years.
- ES-L1, GEO, HEO and SSO orbits have 100% success rate.
- The test results indicated comparable performance between the models, with the decision tree model exhibiting a slight advantage over the others.
- Launch sites are strategically located near the equator to take advantage of an inherent advantage provided by the Earth's rotational speed. This natural boost translates into significant cost savings by reducing the need for additional fuel and boosters.
- All of the launch sites are situated in close proximity to coastal areas.
- KSC LC-39A has the highest success rate among launch sites.

Thank you!