



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ashok Sharma
5/12/2022



Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

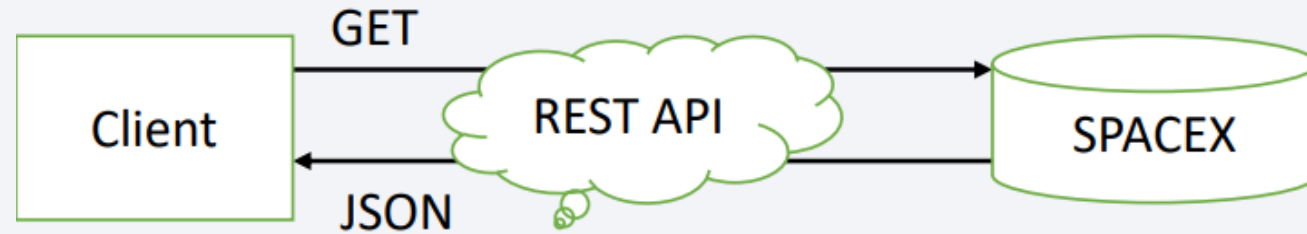
- Data collection methodology:
 - Data collected from SpaceX API which enable us to access some major data.
 - Web scrapping from Wikipedia by using beautiful soup,
- Perform data wrangling
 - Data Cleaning includes removing unnecessary data.
 - Replacing NAN values to the mean of data or simple 0.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - Classification models were created and evaluated by using GridSearchCV

Data Collection

- Data acquisition through use of SpaceX API.([api.spacexdata.com/v4/...](https://api.spacexdata.com/v4/))
- 4 different helper functions were built to request data:
- Booster Version, Launch Site, Payload, and Core Data requests response in the form of .JSON files then get parsed and converted into Data Frames for further analysis.

Data Collection – SpaceX API

- A get requests issued to access past launch data.
- Data obtained is then parsed and filtered to issue additional get requests for our global variables.
- [GITHUB](#)



```
In [14]: #Global variables
BoosterVersion = []
PayloadMass = []
Orbit = []
LaunchSite = []
Outcome = []
Flights = []
GridFins = []
Reused = []
Legs = []
LandingPad = []
Block = []
ReusedCount = []
Serial = []
Longitude = []
Latitude = []
```


Data Collection - Scraping

- Web Scraping was done through the use of HTML GET requests.
- BeautifulSoup library was used to parse HTML tables and convert them into data frames
- GITHUB

```
response = requests.get(static_url).text
soup = BeautifulSoup(response, 'html.parser')
html_tables = soup.find_all("table")
print(html_tables)
```

```
[<table class="multicol" role="presentation" style="border: 1px solid black; width: 100%; border-collapse: collapse;">
<tbody><tr>
<td style="text-align: left; vertical-align: top;">
first_launch_table = html_tables[2]
a = soup.find_all('th')
for x in range(len(a)):
    try:
        name = extract_column_from_header(a[x])
        if (name is not None and len(name) > 0):
            column_names.append(name)
    except:
        pass
```

Data Wrangling

- **Clean NULL values**
- **Determine number of launches per site**
- **Determine occurrence of launches at different orbital distances.**
- **Sort mission outcome per orbit type**
- **Create a label to help associate launch to outcome**

EDA with Data Visualization

As part of the DV several graphs were generated to gain additional insights

- **Launch Site vs Flight Number:** as the number of launches increases so does the chance for a successful landing.
- **Payload vs Launch Site:** higher occurrence for launches below 10000kg.
- **Class vs Orbit:** ES-L1, GEO, HEO and SSO have the highest success rate.
- **Orbit vs Flight Number:** Class vs Orbit needs to be cross referenced with Flight Number to see that SSO and VLEO have the highest success rates backed by a number of flights while some other orbits were only attempted once.
- **Orbit vs Payload:** SSO launches at rather low payloads while VLEO Launch Payloads were at above 12000kg.

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 where the successful landing outcome in drone ship was achieved.
 - Listing the names of the boosters which have success in “drone ship” and have payload between 4000 and 6000kg.
 - Listing the total number of successful and failure mission outcomes
 - Listing the names of the booster versions which have carried the maximum payload mass.

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- An interactive map was generated with Folium Library. The map shows the Markers for the different Launch Sites with Cluster Markers to show the number of launches per site.
 - Distances to coastline and transport infrastructure were added to support benefits for using Launch Sites.

Build a Dashboard with Plotly Dash

An interactive dashboard was generated with Plotly Dash. It offers a dropdown menu to select “ALL” or individual launch sites. A slide bar representing payload values was added afterwards. The dashboard shows dynamically a couple of graphs depending on the values selected for the dropdown menu and the slide bar. The pie chart shows Success Rate while the scatterplot shows the correlation of Success vs Payload.

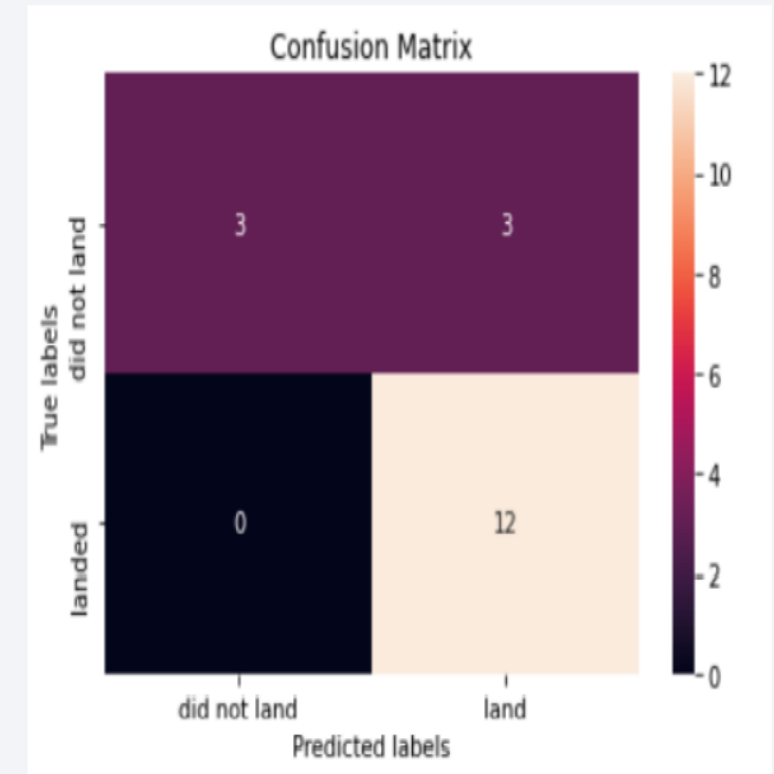
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- For the predictive analysis, the data set was preprocessed by normalizing the data.
- `train_test_split` method was used to create a training set and a testing set.
 - `GridSearchCV` was then used to optimize and find the best hyperparameters for all models.
- Confusion Matrices and Accuracy scores were used to compare the different models.

Results

- For this particular dataset all models perform the same at 83.3% accuracy.
- Low payload launches perform better than heavy ones.
- KSC LC 39A had the highest success rate while also presenting launches at load and high loads.
- Orbits ISS has the better associated performance followed by VLEO considering the frequency of success landings.

	Score
Model	
knn_cv	0.833333
tree_cv	0.833333
svm_cv	0.833333
logreg_cv	0.833333

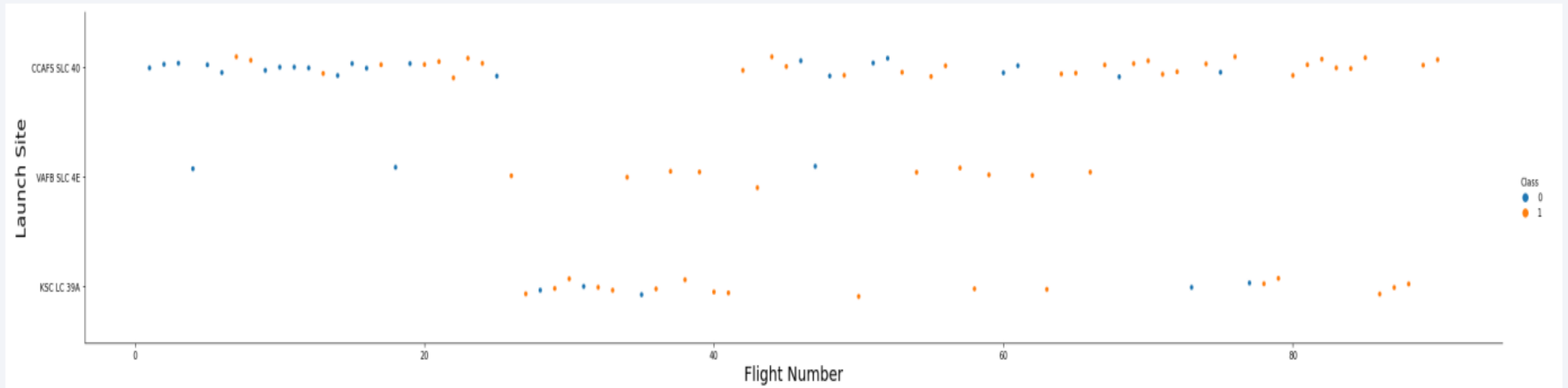


The background of the slide is an abstract composition. It features a dark blue field on the left side, which transitions into a complex pattern of diagonal streaks in shades of blue, red, and teal on the right. These streaks have a textured, almost woven appearance. Overlaid on this pattern is a faint, light blue grid that recedes into the distance, creating a sense of depth and perspective.

Section 2

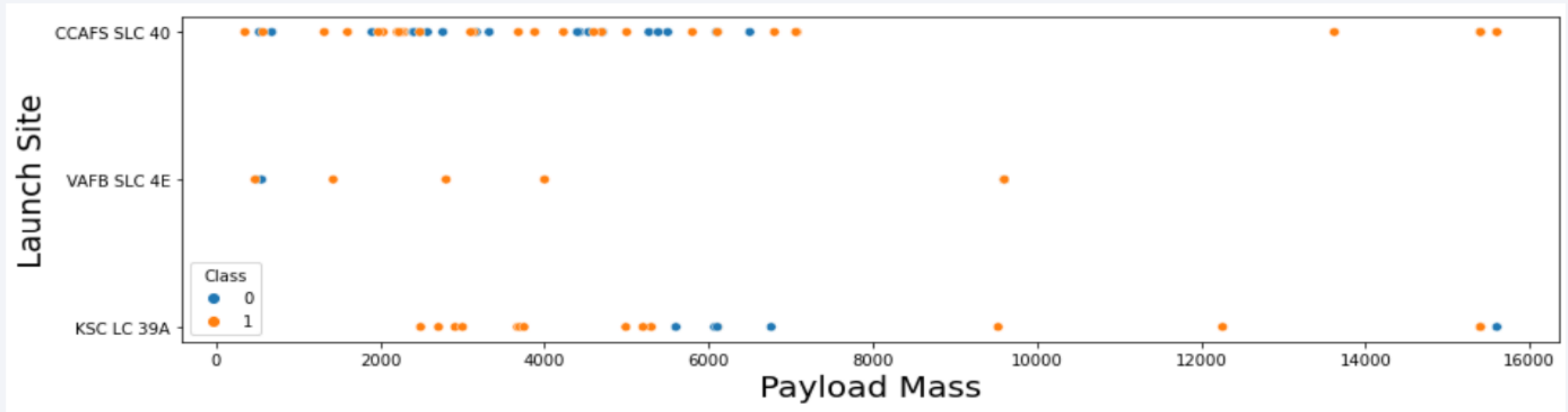
Insights drawn from EDA

Flight Number vs. Launch Site



Site CCAFS SLC 40 was used for the initial launches showing a high occurrence of failed landings.

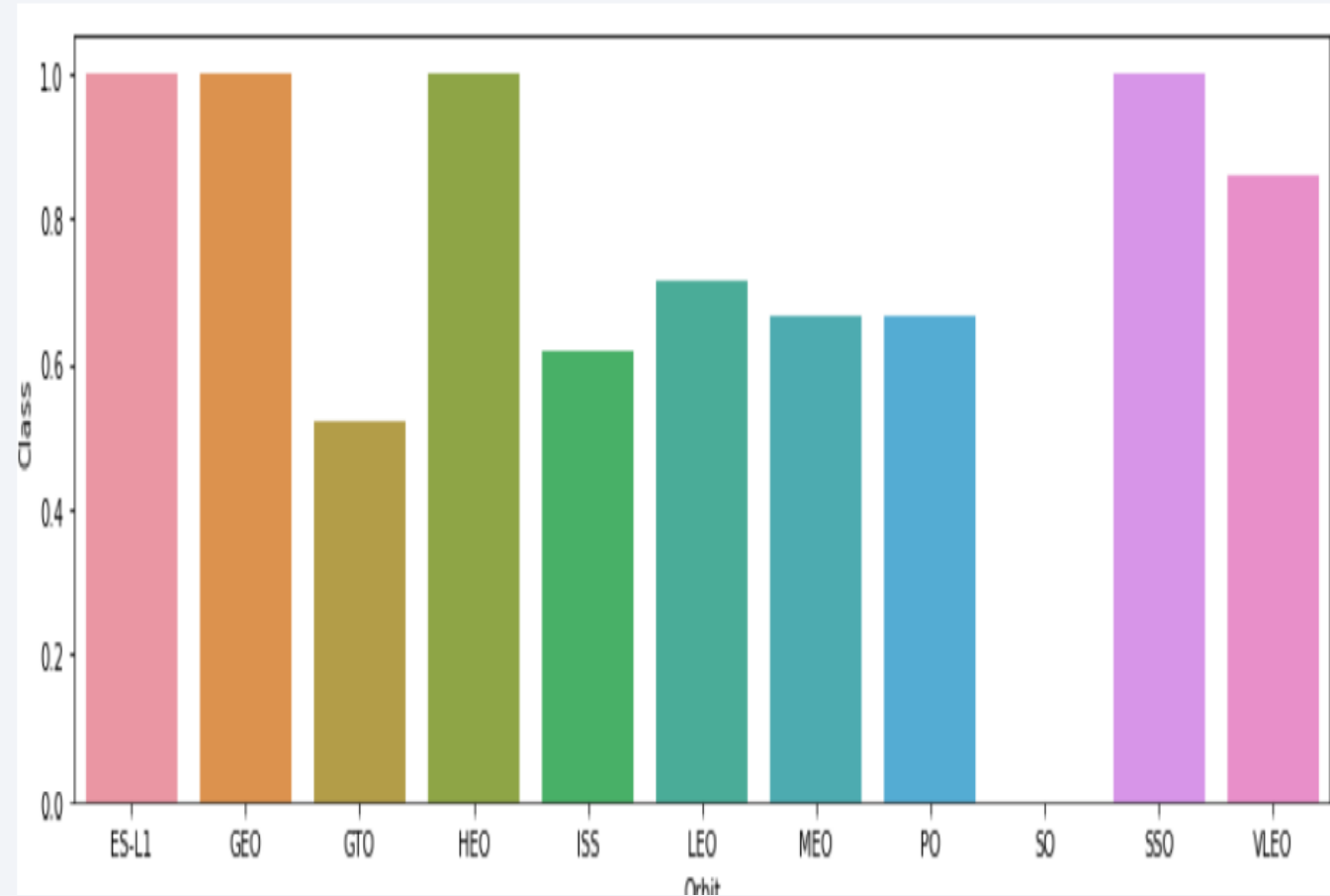
Payload vs. Launch Site



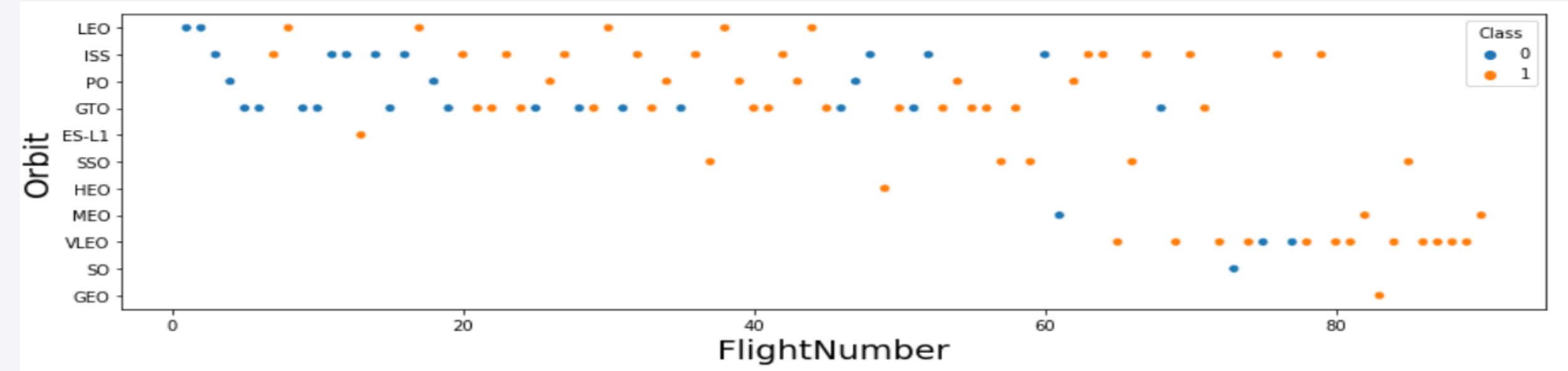
- On this graph it is easier to see the correspondence Success Rate to Low Payload

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have a mean success rate of 100%.

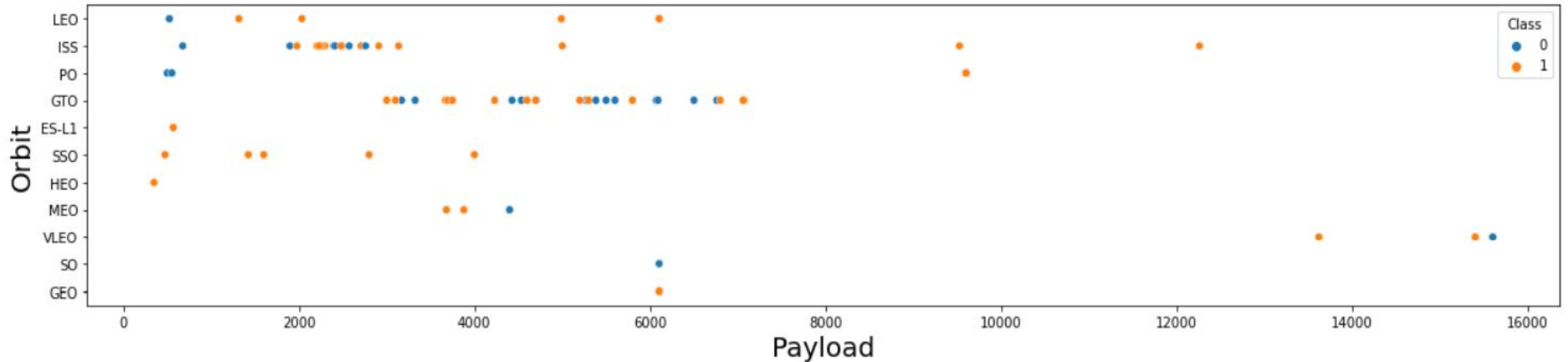


Flight Number vs. Orbit Type



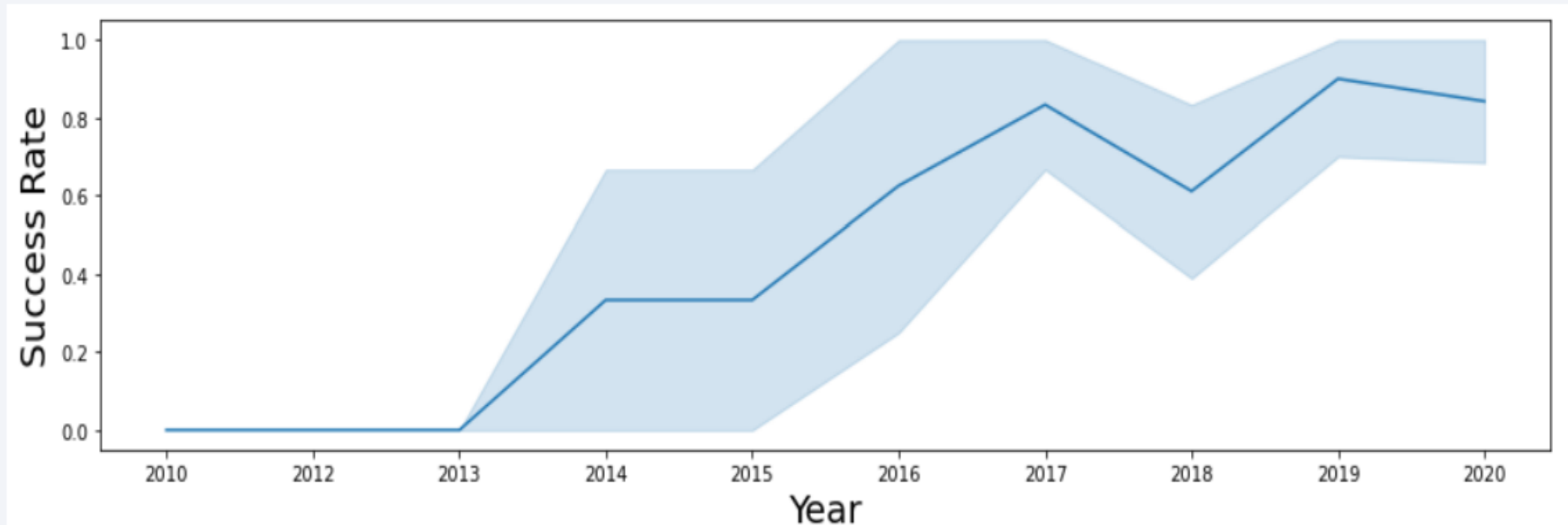
This graph shows how while some orbits have a 100% success rate, they have only a few associated flights (ES-L1, GEO, HEO). At the same time it is observed that SSO has several associated flights and all were successful landings.

Payload vs. Orbit Type



With this graph we can further associate the success rate per orbit type to the relatively low payload used for the highest performance (ES-L1, GEO, HEO, SSO)

Launch Success Yearly Trend



With this graph we see the start of the positive trend during 2013.

All Launch Site Names

List all site names from SPACEX

```
%%sql
```

```
select DISTINCT(LAUNCH_SITE) from SPACEX
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb  
Done.
```

```
']:
```

```
launch_site
```

```
CCAFS LC-40
```

```
CCAFS SLC-40
```

```
KSC LC-39A
```

```
VAFB SLC-4E
```

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%%sql
select LAUNCH_SITE from SPACEX
where LAUNCH_SITE LIKE 'CCA%'
limit 5
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

```
]:
```

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

```
%%sql
select SUM(PAYLOAD_MASS__KG_) from SPACEX
where LAUNCH_SITE LIKE 'CCA%'
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

```
]: 1
321400
```


Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

Dates of the first successful landing outcome on ground pad

```
%%sql
select min(DATE) from SPACEX
where LANDING__OUTCOME LIKE 'Success%'
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

```
1]
```

1

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List of the of boosters names which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
select BOOSTER_VERSION, PAYLOAD_MASS_KG_ from SPACEX
where (LANDING__OUTCOME LIKE 'Success (drone ship)') AND (PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000)
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

```
!]:
```

booster_version	payload_mass_kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes

```
%%sql
select COUNT(MISSION_OUTCOME) FROM SPACEX

* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46
Done.

1
202
```

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum payload mass

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

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0r
Done.
```

```
] :
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

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

```
%%sql
SELECT BOOSTER_VERSION, LAUNCH_SITE, DATE, LANDING__OUTCOME FROM SPACEX
WHERE LANDING__OUTCOME LIKE 'Failure (drone ship)' AND YEAR(DATE)=2015

* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n4
Done.
```

```
4]:
```

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

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

Ranking of 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 landing__outcome, COUNT(*) AS count FROM SPACEX
WHERE DATE BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY landing__outcome
ORDER BY COUNT(landing__outcome) DESC
```

```
* ibm_db_sa://bqn92294:***@b0aebb68-94fa-46ec-a1fc-1c96
Done.
```

]:

landing__outcome	COUNT
No attempt	20
Failure (drone ship)	10
Success (drone ship)	10
Controlled (ocean)	6
Success (ground pad)	6
Failure (parachute)	4
Uncontrolled (ocean)	4
Precluded (drone ship)	2

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The image is a composite of a solid blue background on the left and a satellite photograph of Earth on the right. The Earth's surface is dark blue, with numerous bright yellow and orange lights representing cities and urban areas. The horizon line of the Earth is visible, separating the dark surface from the blackness of space.

Section 3

Launch Sites Proximities Analysis

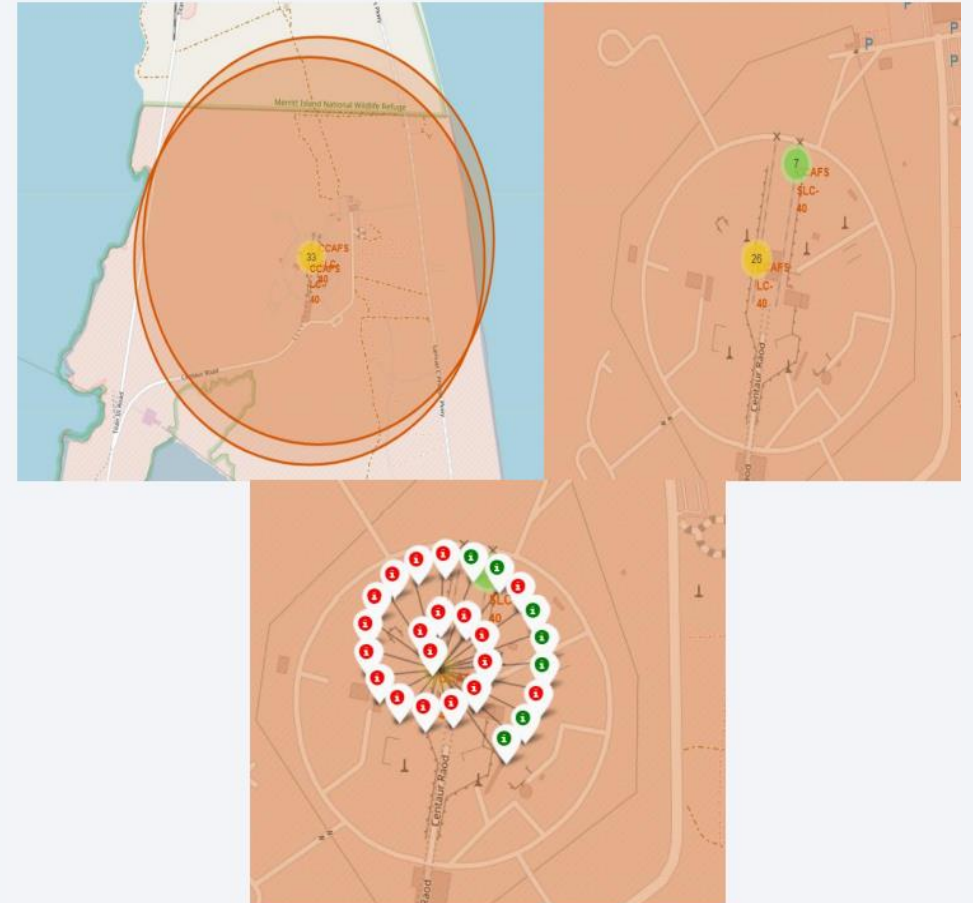
Interactive Map

This is a
screenshot of
the Folim map
showing the
different sites
across the US



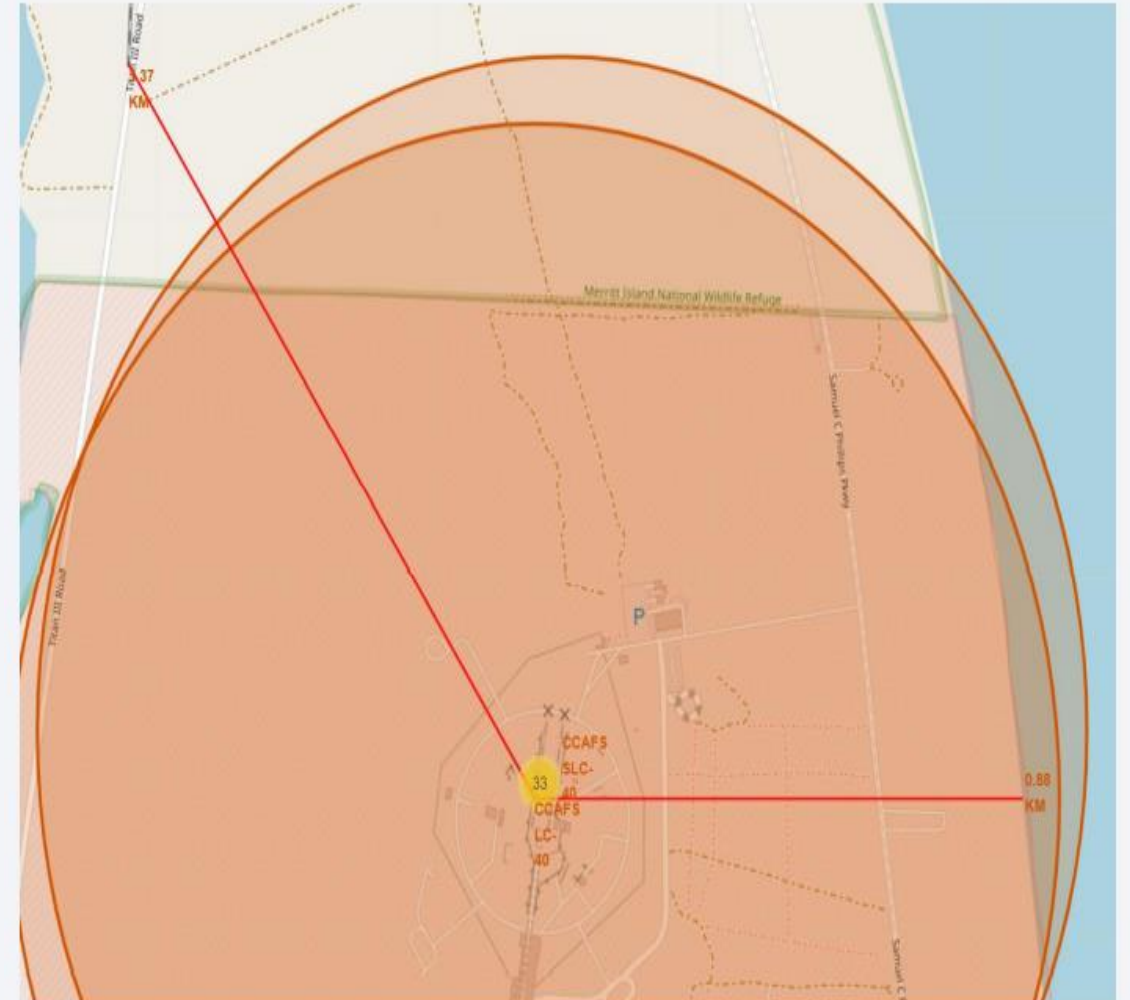
Interactive Markers

- The screenshots show the added Circle markers as well as the MarkerCluster which show the number of launches carried out at a site and if clicked it shows the outcome of the the individual launches.



Surrounding Site Elements

- Screenshot showing distances from the CCAFS LC-40 to the coastline (0.88 km) to the east and a railway (1.37 km) to the northwest.





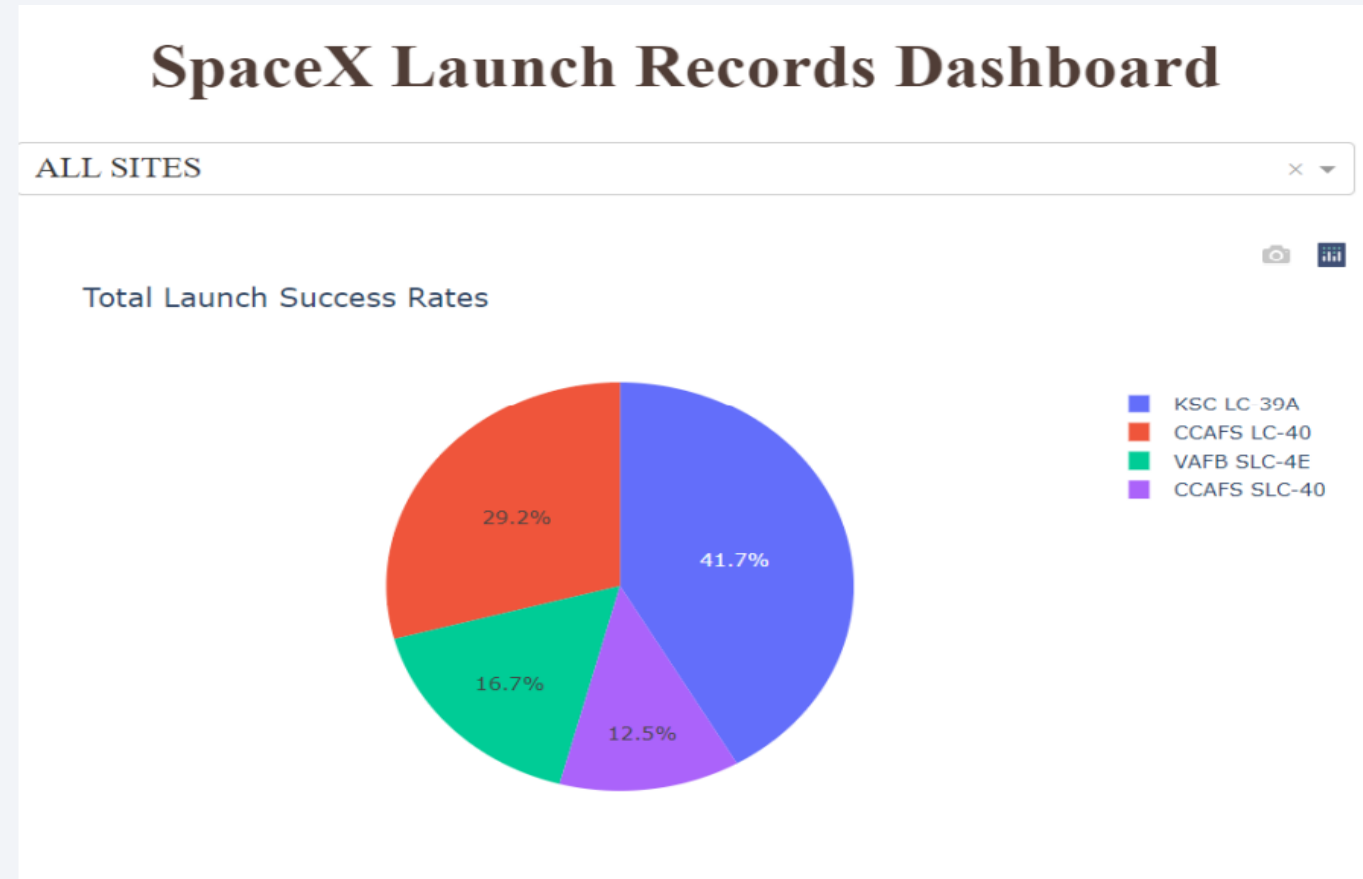
Section 4

Build a Dashboard with Plotly Dash

Total Success Rate

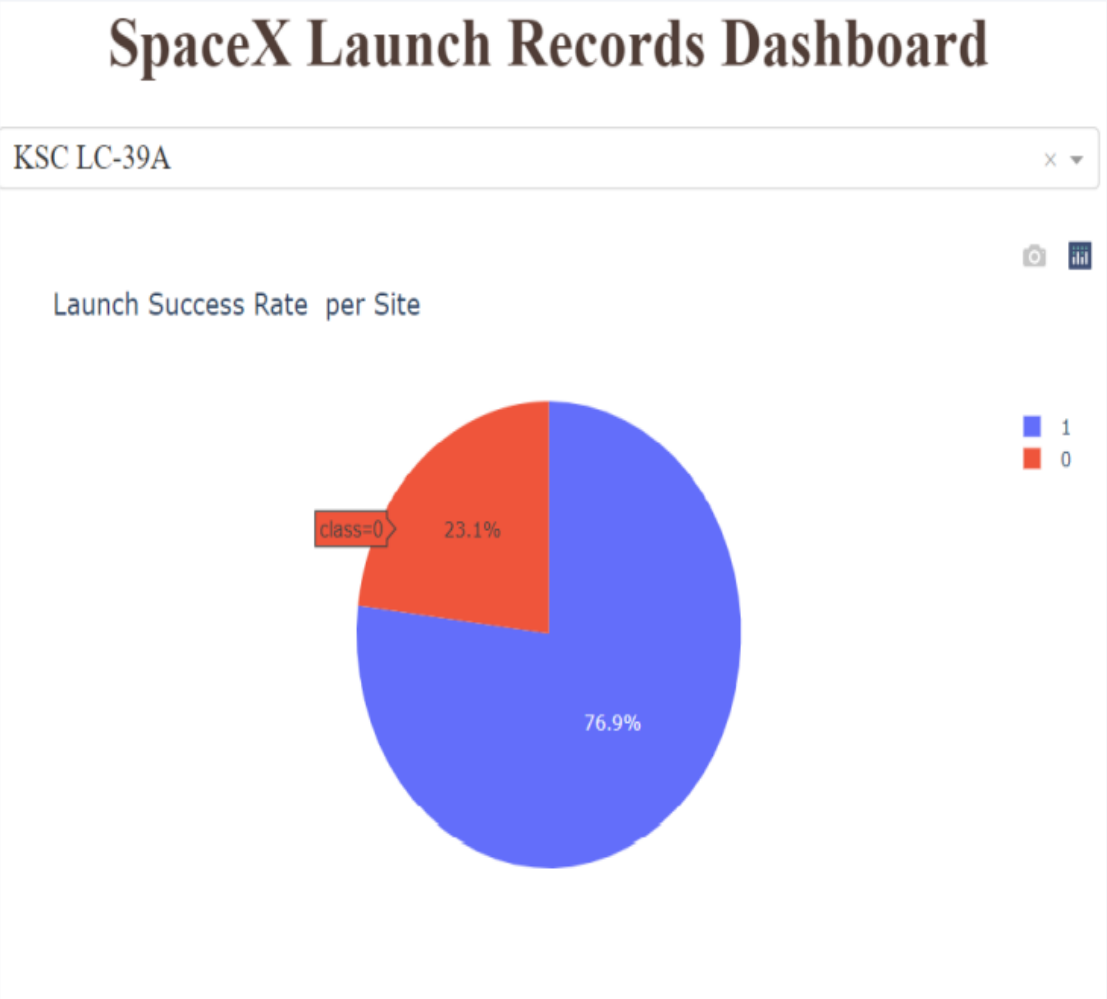
This is a screen shot of the interactive dashboard.

- The pie-chart shows the success rate for all sites while the color key separates the different sites.



Success Rate per Site

Dashboard showing individual success rate for site KSC LC-39A which is the one with the highest success rate with 76.9%



Payload vs Success rate correlation



This screenshot shows how there is a higher frequency for success landings at low payload values vs high payload values.



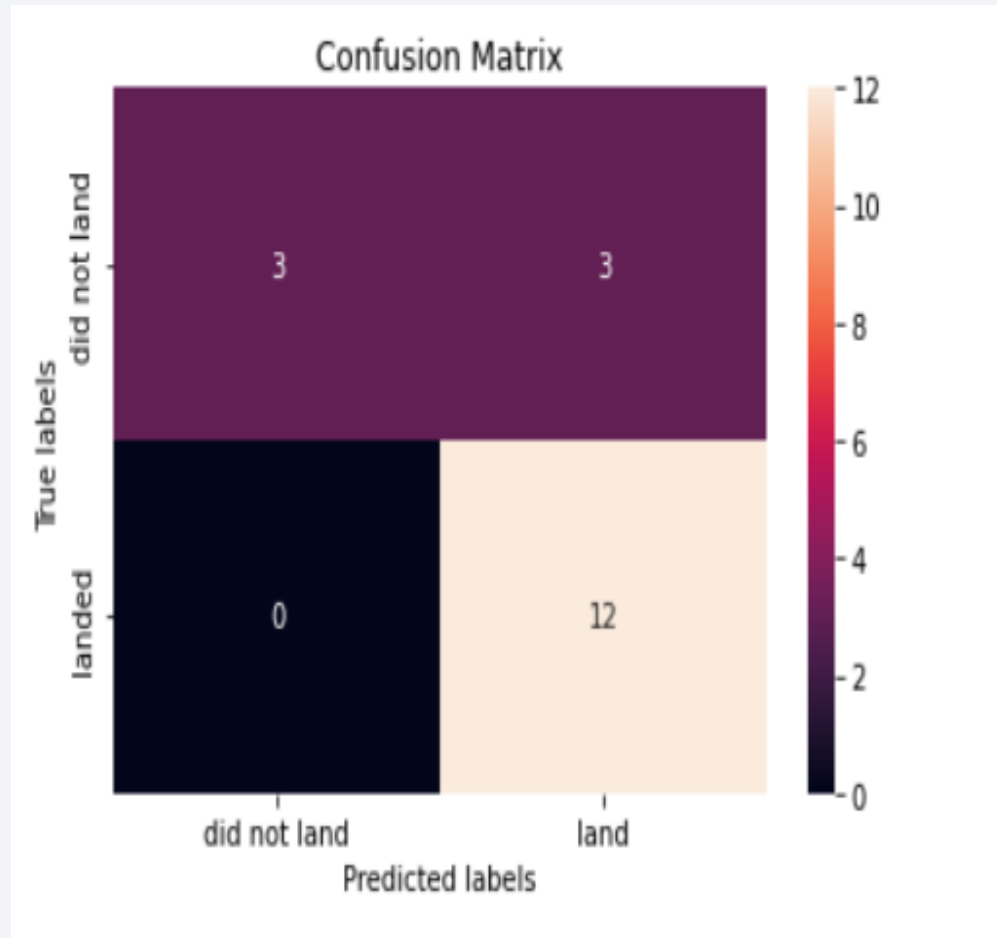
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- ...

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!

