Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

To accomplish task set in this project, which is finding correlation between Payload Mass, Launch Site, Booster Version, Orbit and Flight Number, I had to use some visualization tools, like Pie Charts or Scatter Plot, but also some Machine Learning Tools.

Careful analysis brought these conclusions:

- CCAFS LC-40 has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%,
- Success rate increased in 2013 after 20th Flight,
- Highest success rate are for Orbits: ES-L1, GEO, HEO, SSO.

Introduction

Purpose of this project was to find out the best paramteres for a flight to have both success flight and success landing. For analysis I used data from SpaceX flights.

I want to know what is the best Payload Mass, Launch Site, Booster Version and Orbit to reach the goal of having successful flight and landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data sets were collected with a help of SpaceX API and Web Scraping. SpaceX served for downloading csv file with data regarding Flight Number, Orbit, Booster Version etc. For Web Scraping I used BeautifulSoup to Extract a Falcon 9 launch records HTML table from Wikipedia.

Data Collection – SpaceX API

- Data collection with SpaceX REST which provides json text.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

In [8]:
```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

In [9]:
```
response = requests.get(spacex_url)
```

Check the content of the response

In [10]:
```
print(response.content)
```

```
b'[{"fairings":{"reused":false,"recovery_attempt":false,"re
6Ph45r_o.png","large":"https://images2.imgbox.com/5b/02/Qcx
kr":{"small":[],"original":[]},"presskit":null,"webcast":"h
s://www.space.com/2196-spacex-inaugural-falcon-1-rocket-los
c":"2006-03-17T00:00:00.000Z","static_fire_date_unix":11425
```

Data Collection - Scraping

- Scraping data with a help of Requests and BeautifulSoup

- https://github.com/ejzenberg/ibm_ds_project/blob/master/jupyter-labs-webscraping.ipynb

In [20]:
```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

In [21]:
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content)
```

Print the page title to verify if the `BeautifulSoup` object was created properly

In [22]:
```
# Use soup.title attribute
print(soup.title)
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

Data Wrangling

- I performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Charts were plotted to find out if the outcome of landing was successfull or not.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with SQL

- I used WHERE to find Launch Sites started with CCA
- I summed Payload Mass with SUM(payload_mass__kg_)
- I found averaged Payload Mass for „F9 v1.1” with AVG(payload_mass__kg_) from SPACEXTBL WHERE booster_version = 'F9 v1.1’
- I found Booster Version from SPACEXTBL WHERE "Landing _Outcome" = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000
- https://github.com/ejzenberg/ibm_ds_project/blob/master/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Map objects such as markers, circles, lines, etc. I created and added to a folium map served to show from which place most flights occured
- I added those objects to show in an easy way places for most flights and the distance from coast.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/lab_jupyter_launch_site_location%20(1).ipynb

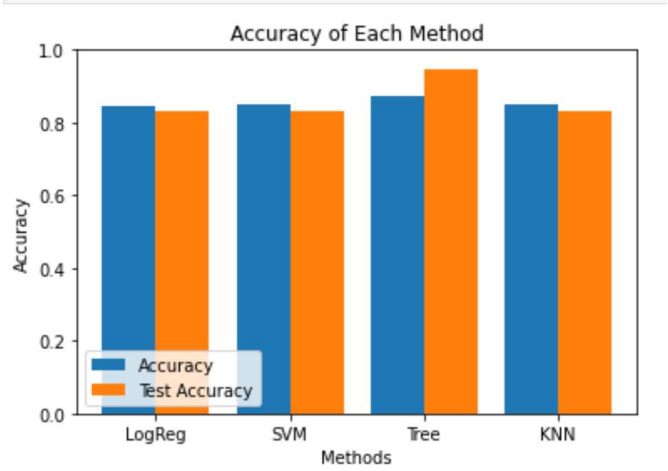
Build a Dashboard with Plotly Dash

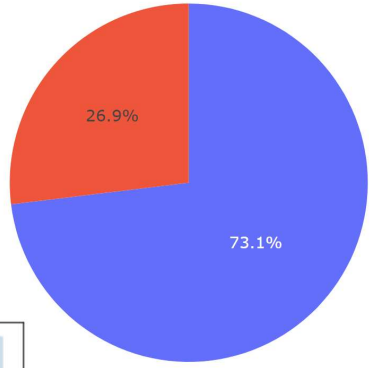
- Plots/graphs and interactions I have added to a dashboard are a great help to understand success rate for different Launch Sites and Payload Mass
- I added those plots and interactions to better explain how Payload Mass and Launch Site affects success rate.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/spacex_dash_app.py

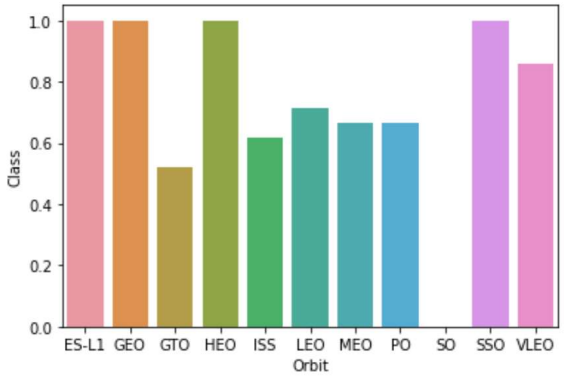
Predictive Analysis (Classification)

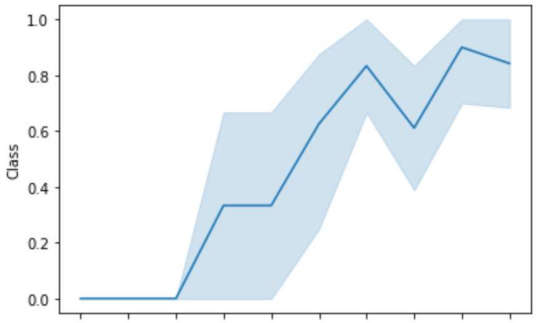
- I built, evaluated, improved, and found the best performing classification model with a help of preprocessing tools like Standard Scaler, splitting data to test and train part, looking for best parameters with GridSearchCV. I created few confusion matrices to find the best performance.
- https://github.com/ejzenberg/ibm_ds_project/blob/master/Machine%20Learning%20Prediction.ipynb

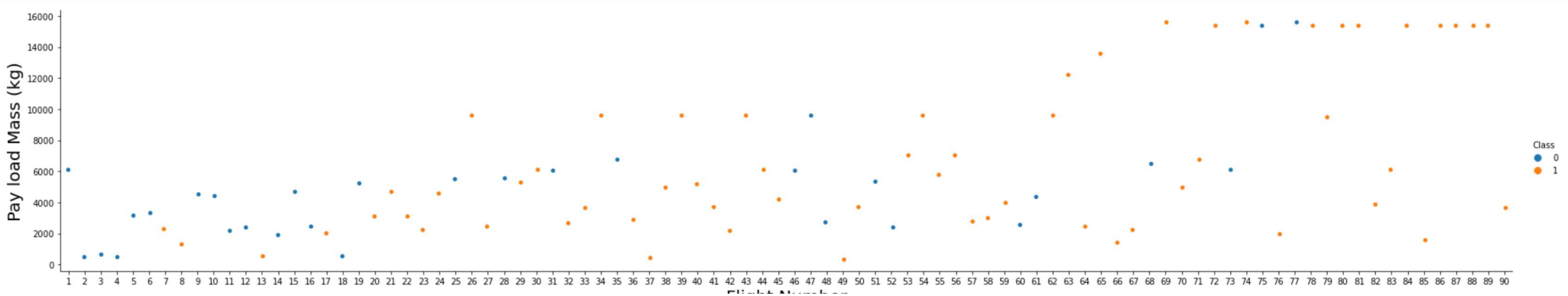
Total Launches for site CCAFS LC-40





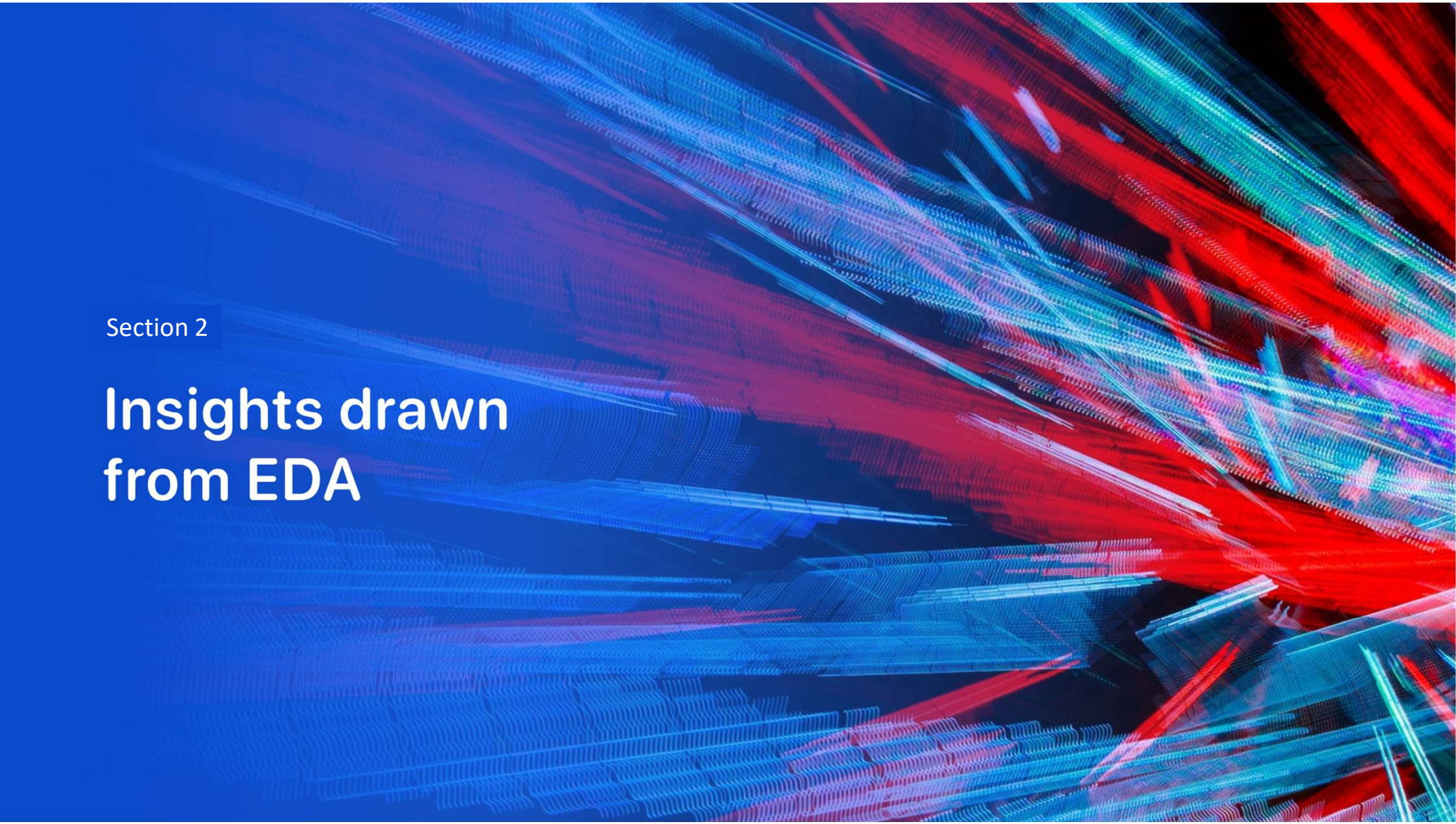




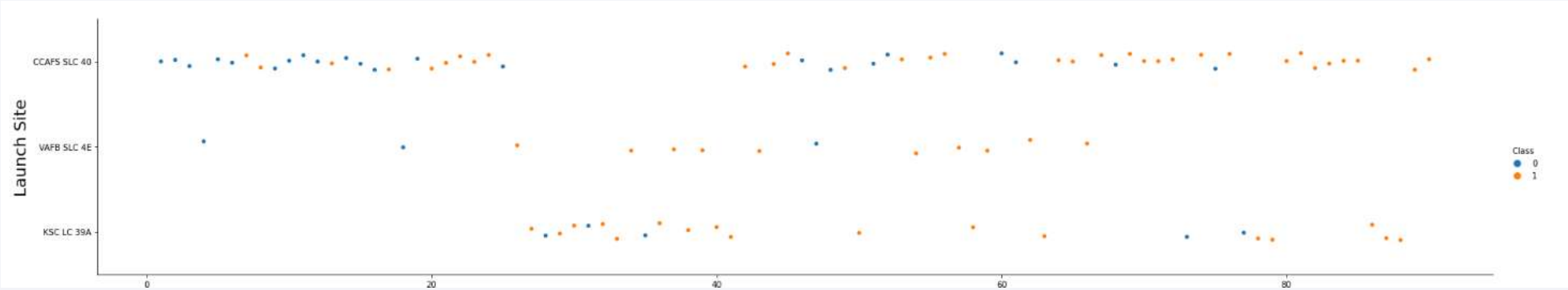


Results

- Careful analysis brought these conclusions:
- CCAFS LC-40 has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%,
- Success rate increased in 2013 after 20th Flight,
- Highest success rate are for Orbits: ES-L1, GEO, HEO, SSO.
- In conclusion Classiffication Tree is the Best ML model to predic results on given data (Date and Class) with test accuracy 94%.

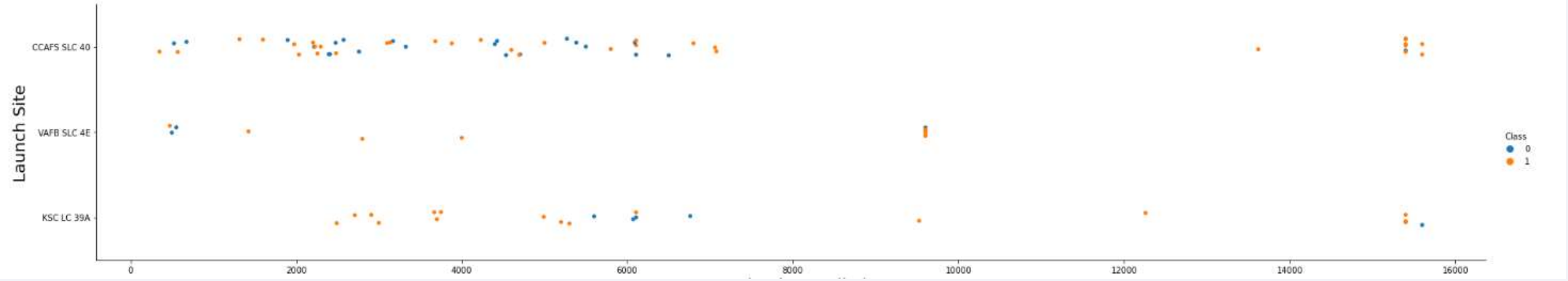

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site



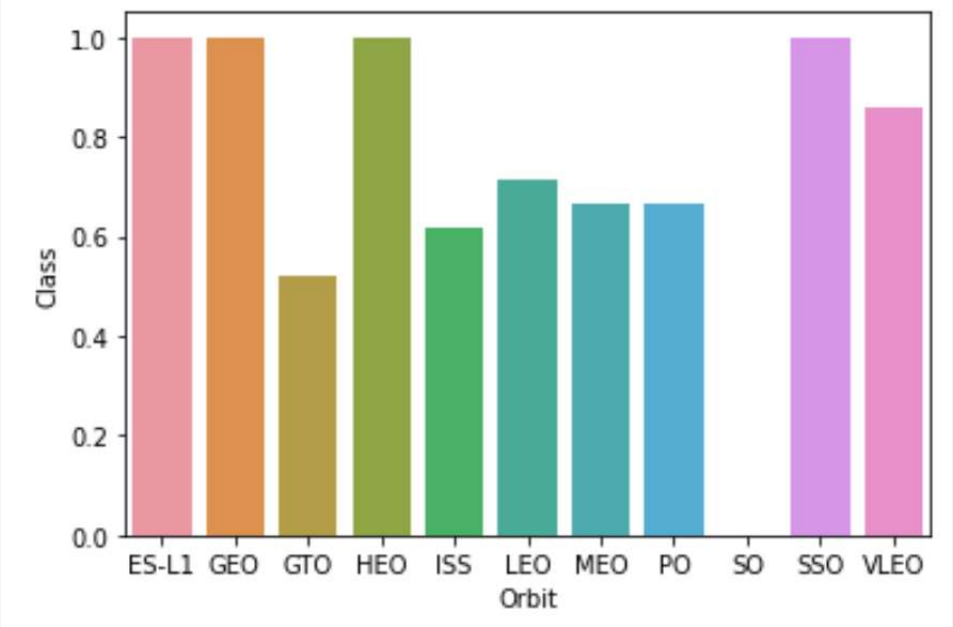
- On the screenshot we can see that after 20th flight success rate increased.

Payload vs. Launch Site



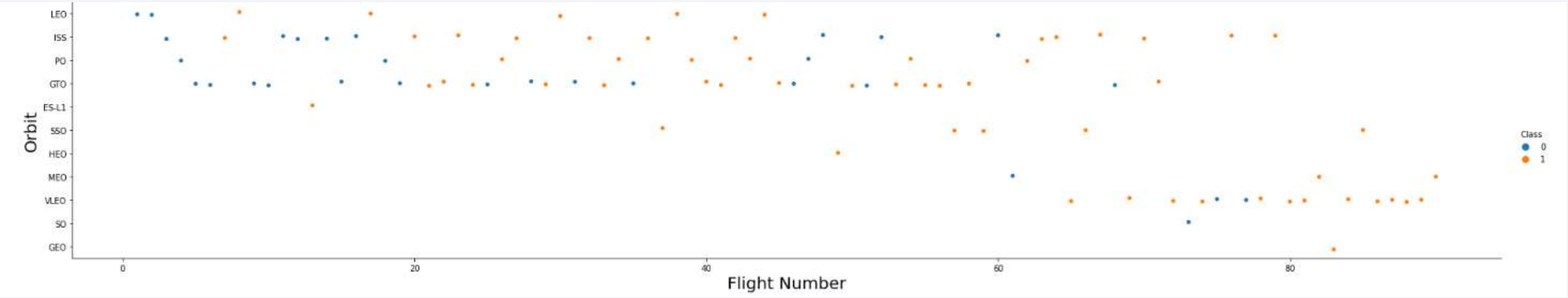
- VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



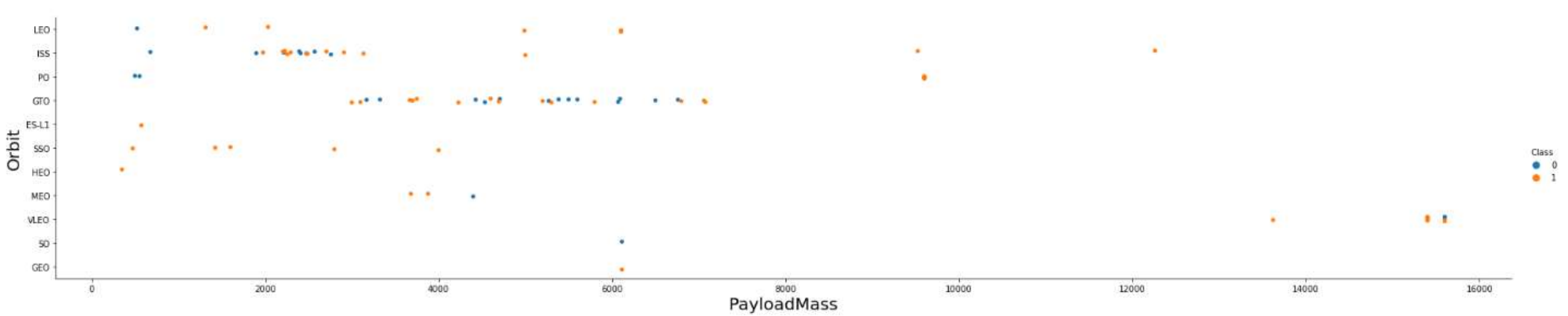
- Highest success rate for ES-L1, GEO, HEO, SSO

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.

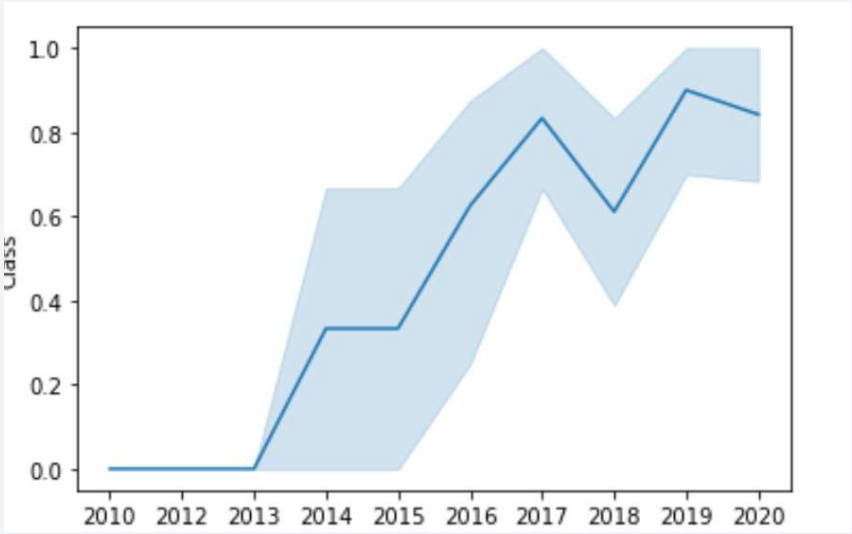
Payload vs. Orbit Type



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

- 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



- You can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

In [5]:
```
%sql select DISTINCT(launch_site) from SPACEXTBL
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

Out[5]:

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

- Result gives us unique names of Launch Site.

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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

- 5 records where launch sites begin with `CCA`

Total Payload Mass

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select SUM(payload_mass__kg_) from SPACEXTBL WHERE customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

1
45596

- Calculate the total payload carried by boosters from NASA, it is 45596 kg.

Average Payload Mass by F9 v1.1

```
%sql select AVG(payload_mass__kg_) from SPACEXTBL WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

1
2928

- Calculate the average payload mass carried by booster version F9 v1.1, which is 2928 kg.

First Successful Ground Landing Date

```
%sql select MIN(DATE) from SPACEXTBL WHERE "Landing _Outcome" = 'Success (ground pad)'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
        1
2015-12-22
```

- Find the date of the first successful landing outcome on ground pad. It was on 2015-12-22.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql select booster_version from SPACEXTBL WHERE "Landing _Outcome" = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql select COUNT(*) from SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success%'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

1

```
%sql select COUNT(*) from SPACEXTBL WHERE "Landing _Outcome" LIKE 'Failure%'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

1

10

- Calculate the total number of successful and failure mission outcomes, which is 61 and 10.

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select booster_version from SPACEXTBL WHERE payload_mass__kg_ = (select MAX(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

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

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

2015 Launch Records

List 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 from SPACEXTBL WHERE DATE LIKE '2015%' and "Landing _Outcome" LIKE 'Failure (drone ship)'
```

```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015, it gives 2 results.

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 "Landing _Outcome", COUNT("Landing _Outcome") from SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing _Outcome"
```

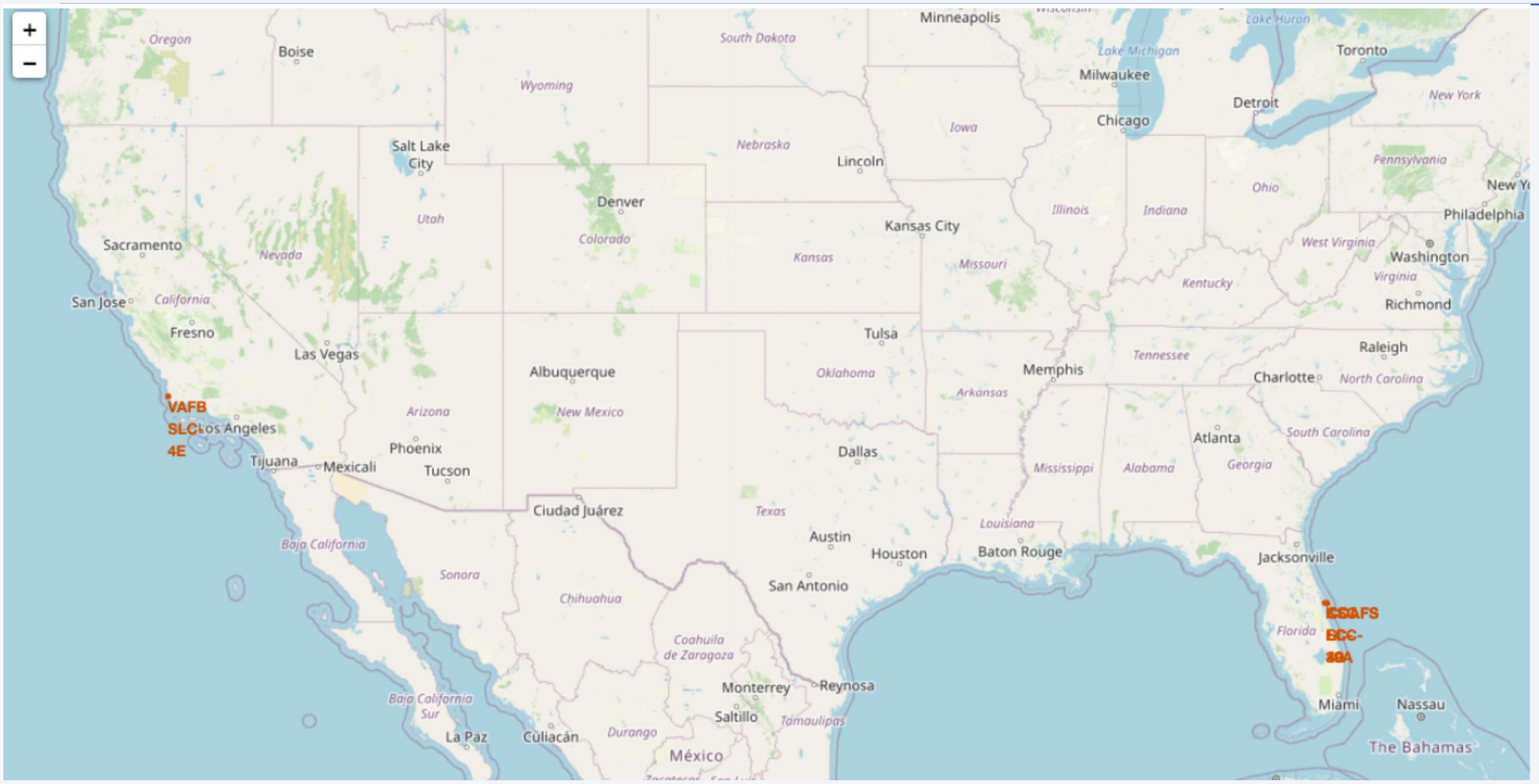
```
 * ibm_db_sa://vbp61812:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/BLUDB
Done.
```

Landing _Outcome	2
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

- 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.

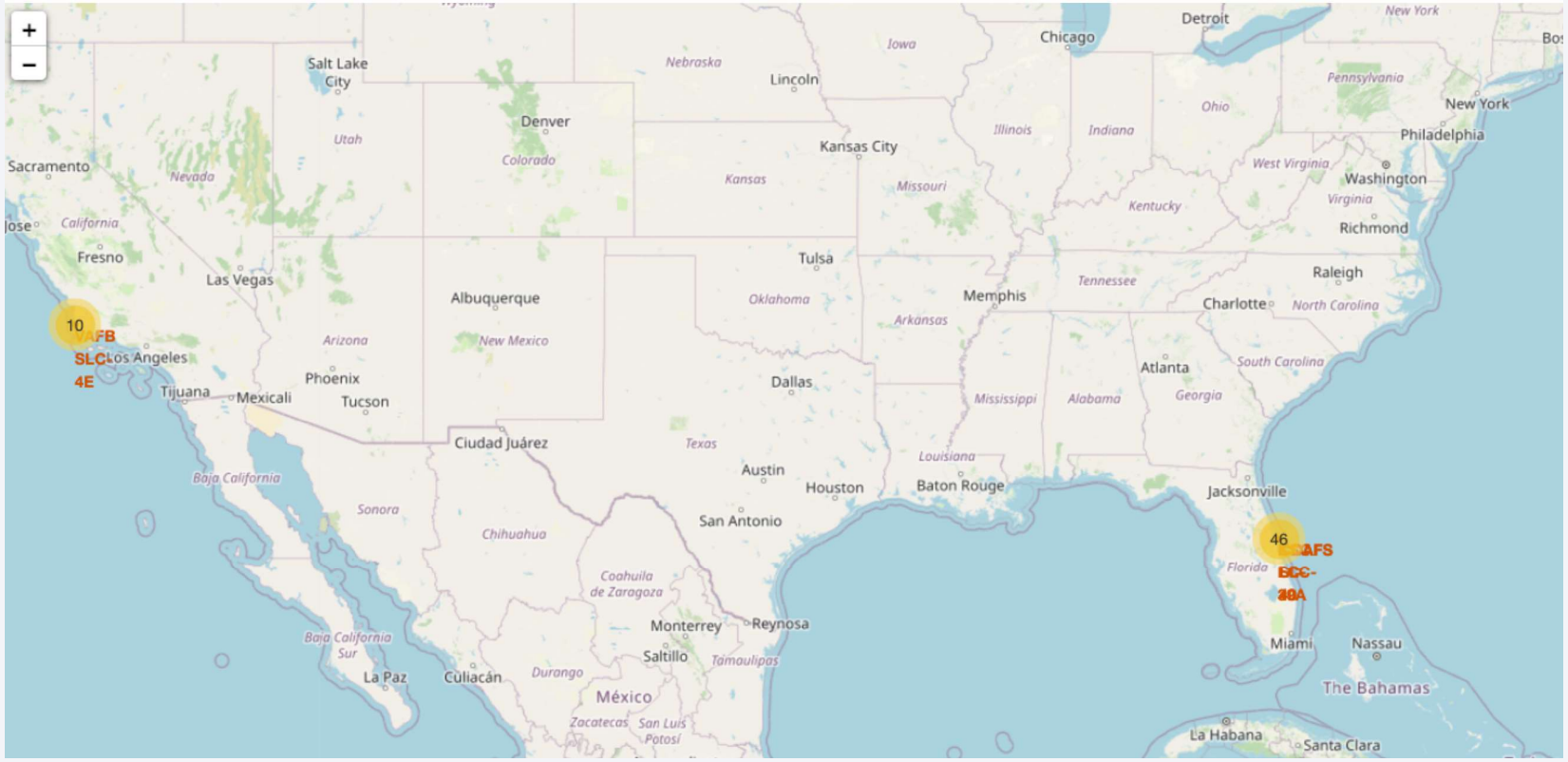

Section 3
Launch Sites
Proximities Analysis

Launch Sites Locations Analysis with Folium



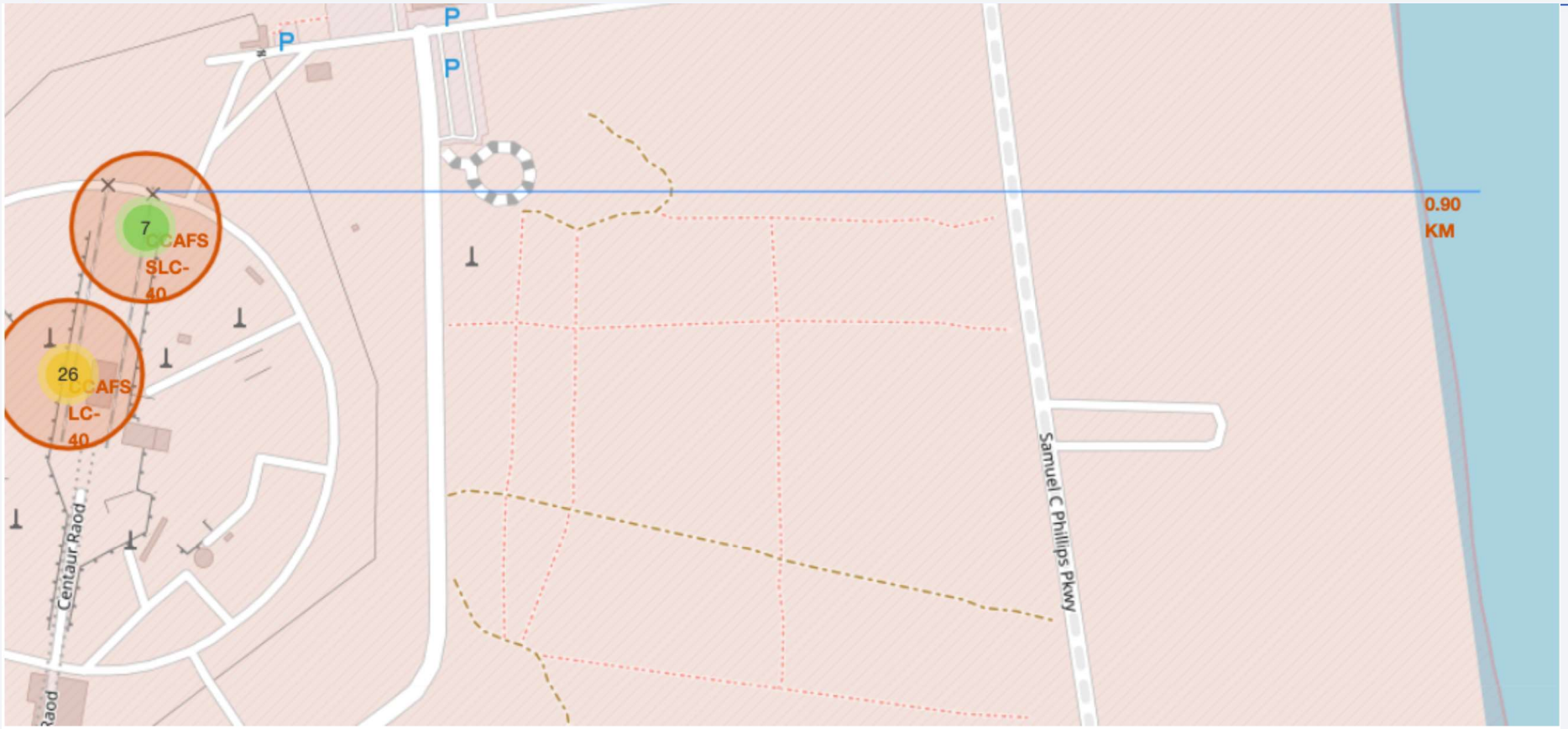
- We can see Launch Sites are located near coasts.

Success/failed launches for each site on the map

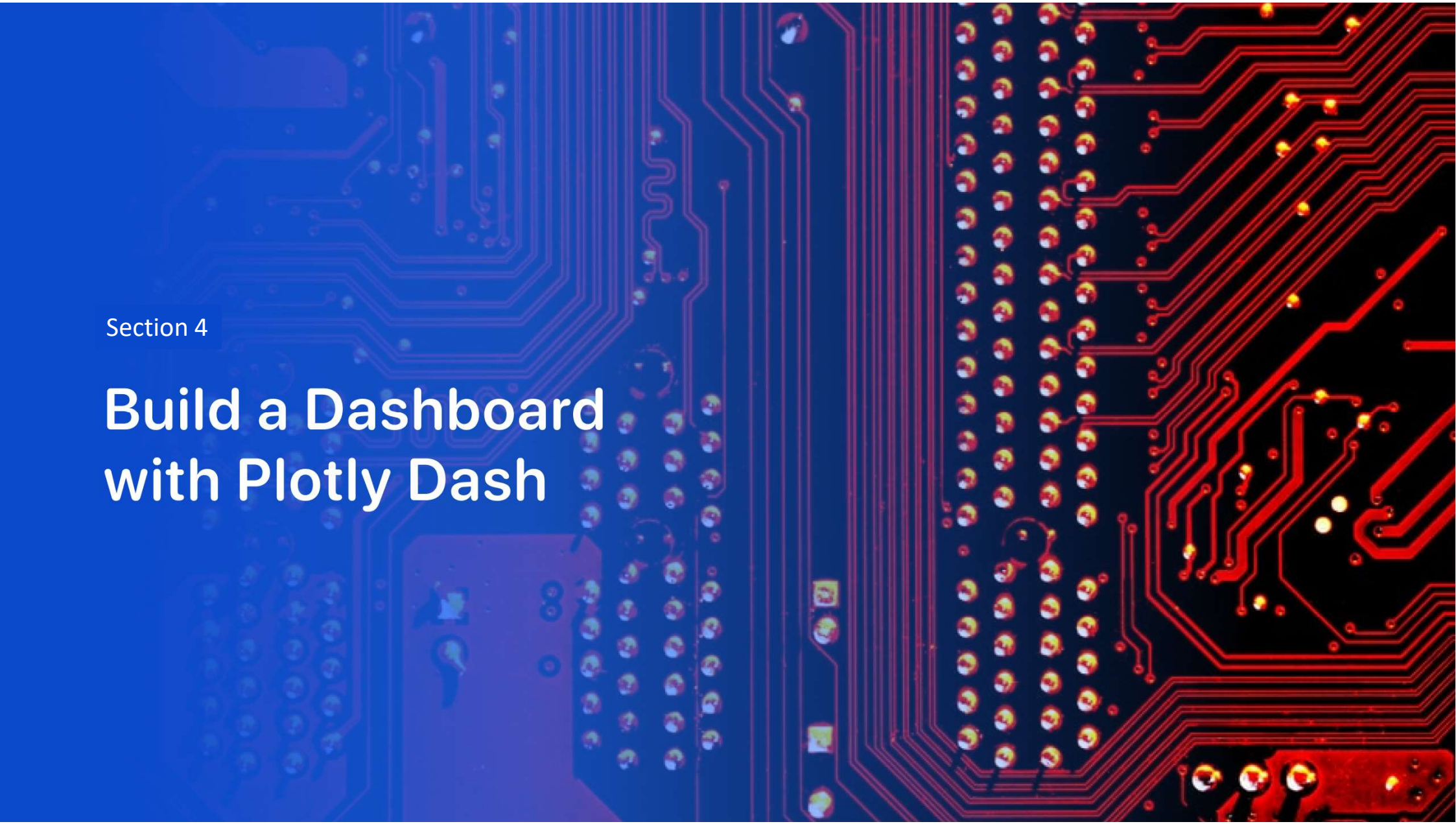


- Most flights were launched from east coast.

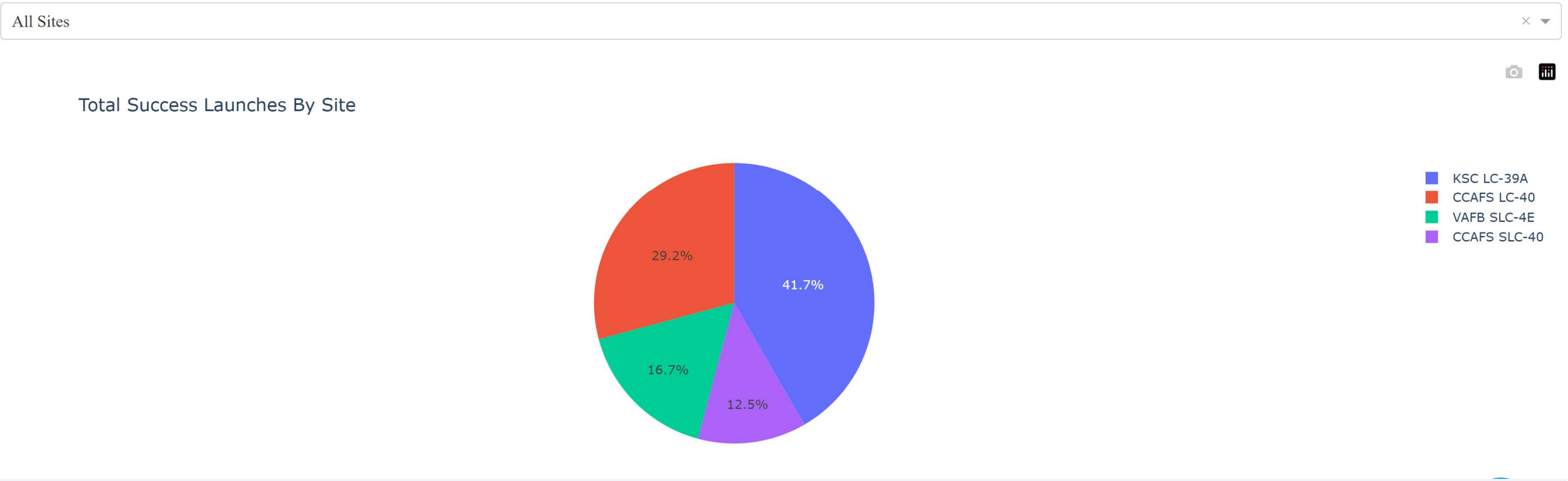
<Folium Map Screenshot 3>



- Flights launched 0,90 km away from a coast.

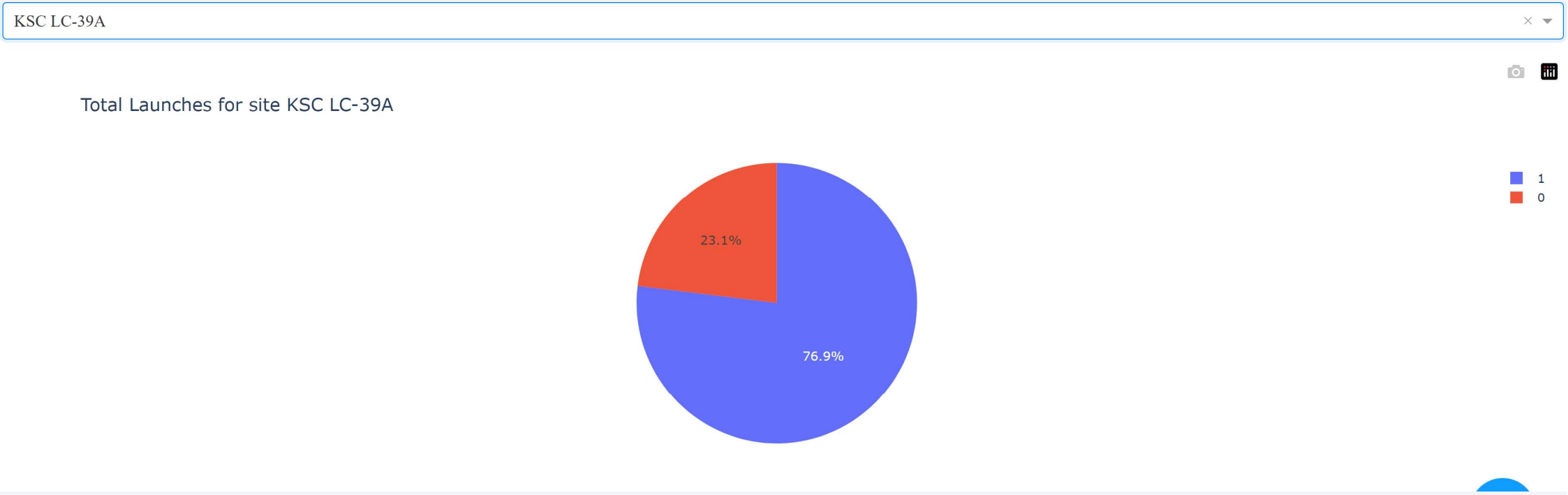

Section 4
Build a Dashboard with Plotly Dash

Total Success Lanuches for All Sites



- Most launches for KSC

Total Launches for Site KSC LC-39A



- Screenshot of the piechart for the launch site with highest launch success ratio

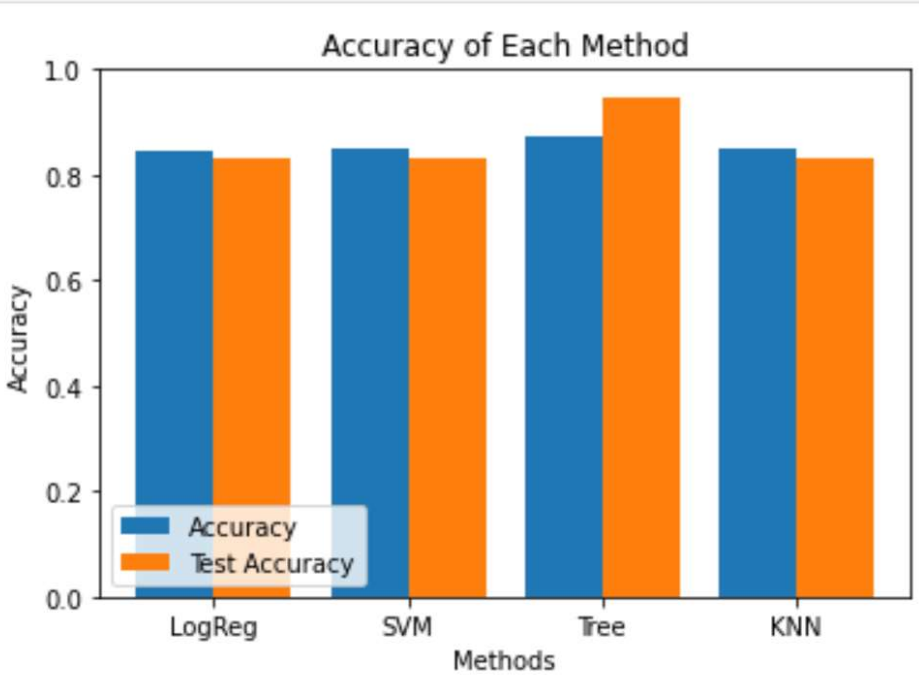
Payload vs. Launch Outcome scatter plot for all sites

- Best success rate for Payload between 2k and 6k.

Section 5

Predictive Analysis (Classification)

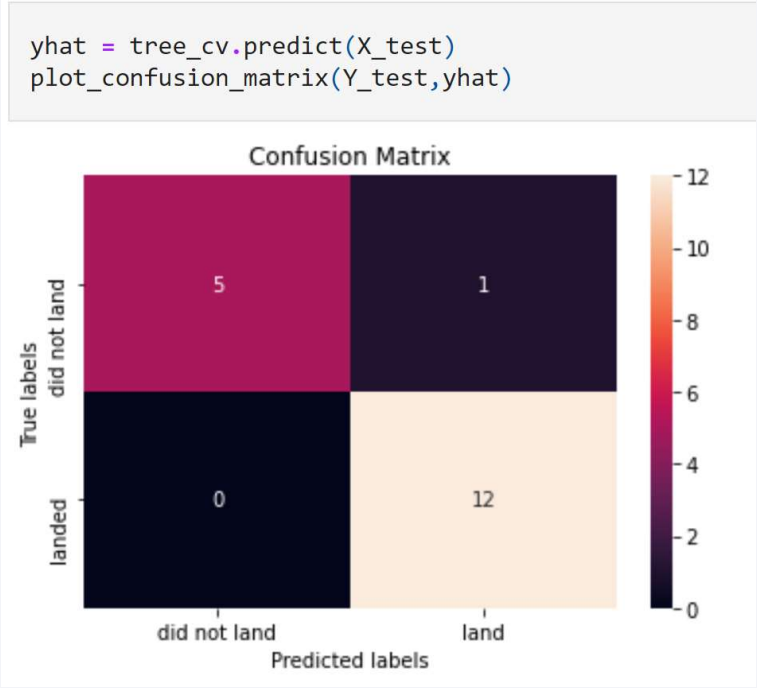
Classification Accuracy



- Visualize the built model accuracy for all built classification models, in a bar chart
- The highest classification accuracy has Classiffication Tree model.

Confusion Matrix

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



- Confusion matrix of the best performing mode. The most true positive cases and true negative cases.

Conclusions

- CCAFS LC-40 has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%,
- Success rate increased in 2013 after 20th Flight,
- Highest success rate are for Orbits: ES-L1, GEO, HEO, SSO.
- Best for classification is Classiffication Tree method.

Appendix

- My github:
- https://github.com/ejzenberg/ibm_ds_project/tree/master

Thank you!