Winning Space Race with Data Science

Monther Alharbi
5/30/2023

Outline

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

Executive Summary

- Summary of methodologies
    1. Data were collected
    2. ML models were built
    3. Data visualizations were shown
- Summary of all results
    1. The optimal model was acquired
    2. Visualizations were great for decision making

Introduction

- SpaceX can reuse the first stage unlike other providers
- Falcon 9 rocket launches cost 62 million dollars only
- Therefore, if we can predict the first stage will land successfully, we can determine the cost of a launch.

Section 1

Methodology

Methodology

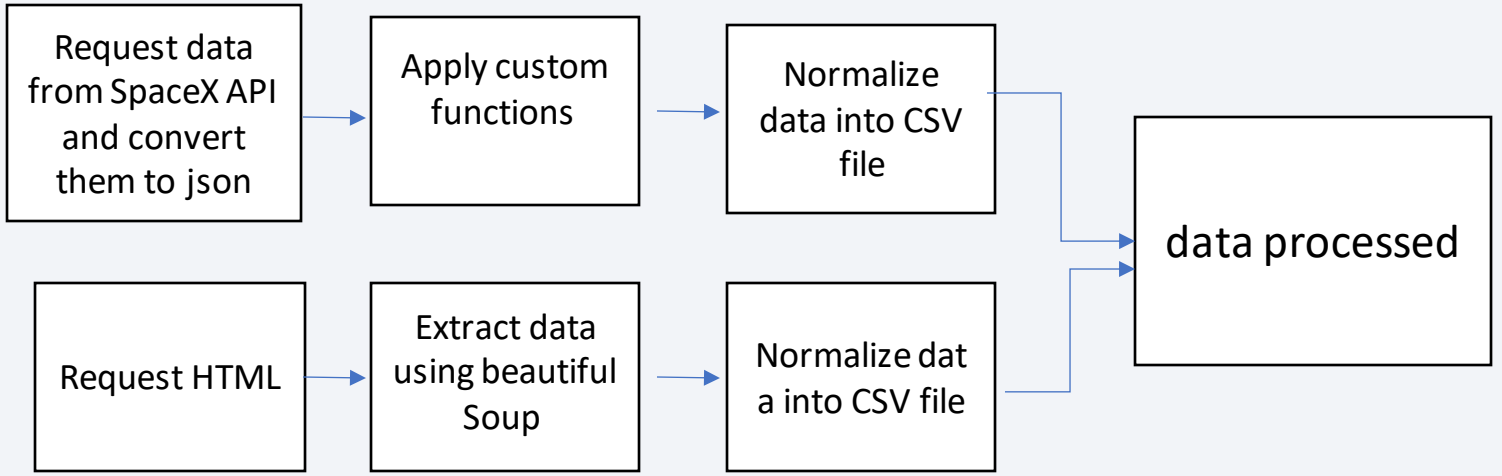
Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scraping from a static page (Wikipedia)
- Perform data wrangling
  - One hot encoding
  - Data cleaning
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, SVM, Decision Tree, KNN models

Data Collection

The sources of processed data are:

- SpaceX REST API
- Wikipedia



Data Collection – SpaceX API

- Data collection with SpaceX REST calls
- https://github.com/Monther-alharbi/testrepo.git

1) Request and convert

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

response = requests.get(spacex_url)

data= pd.json_normalize(response.json())
```

2) Apply functions

```
# Call getLaunchSite
getLaunchSite(data)

# Call getPayloadData
getPayloadData(data)

# Call getCoreData
getCoreData(data)
```

3) Save data into csv file

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection - Scraping

- Data collection with BeautifulSoup
- https://github.com/Monther-alharbi/testrepo.git

1) Request HTML

```
http = requests.get(static_url).text
```

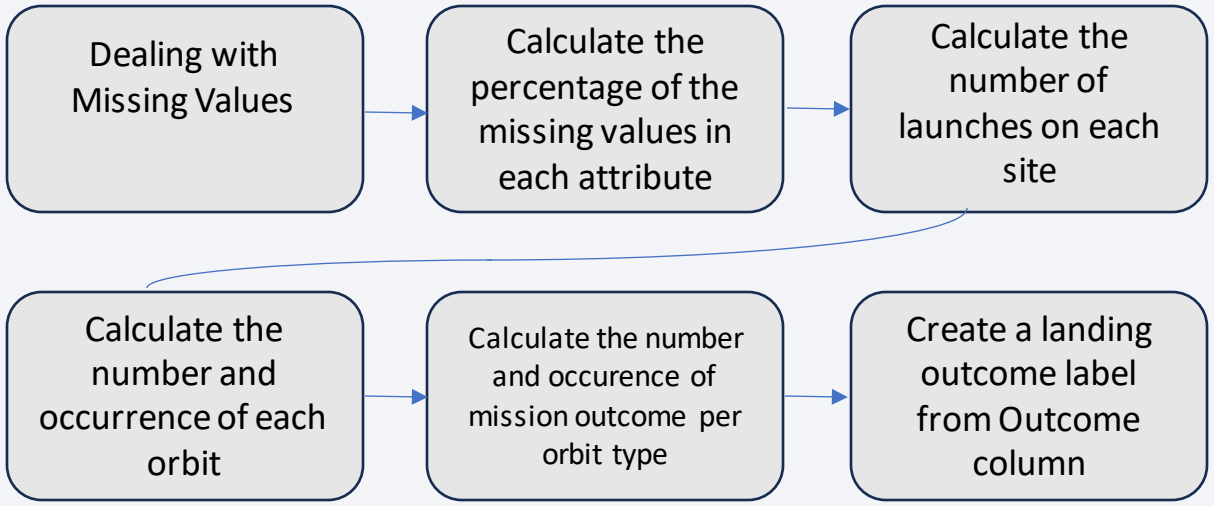
2) Extract data

```
for head in first_launch_table.find_all('th'):
    column_names.append(extract_column_from_header(head))
```

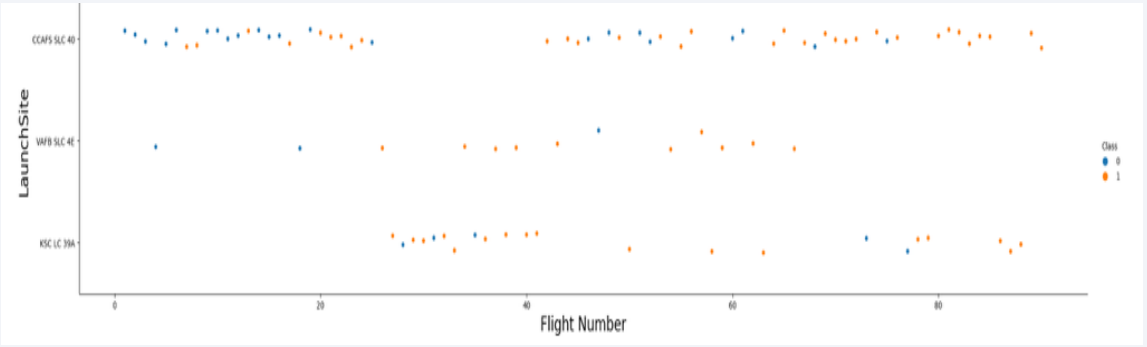
3) Save into CSV

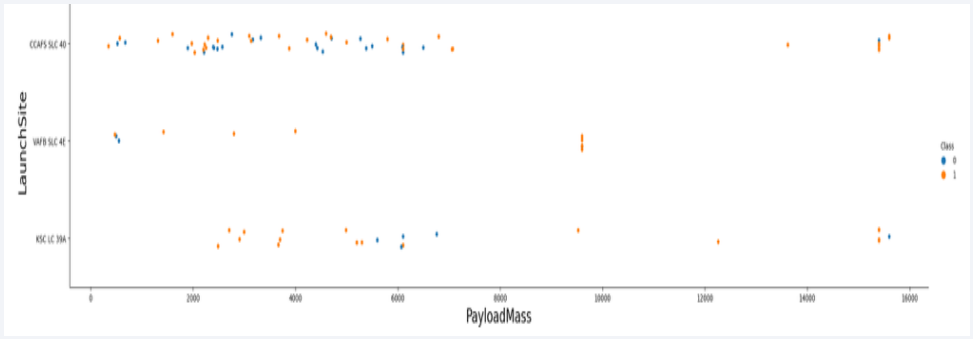
```
df.to_csv('spacex_web_scraped.csv', index=False)
```

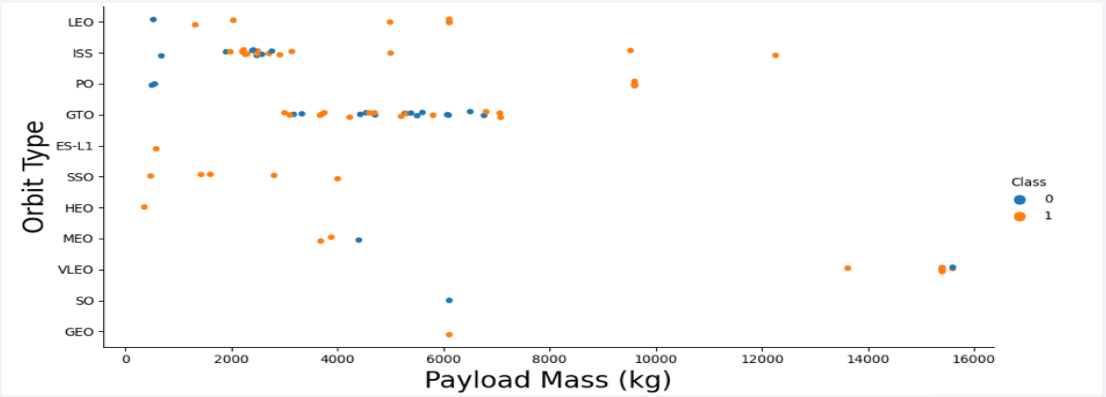
Data Wrangling

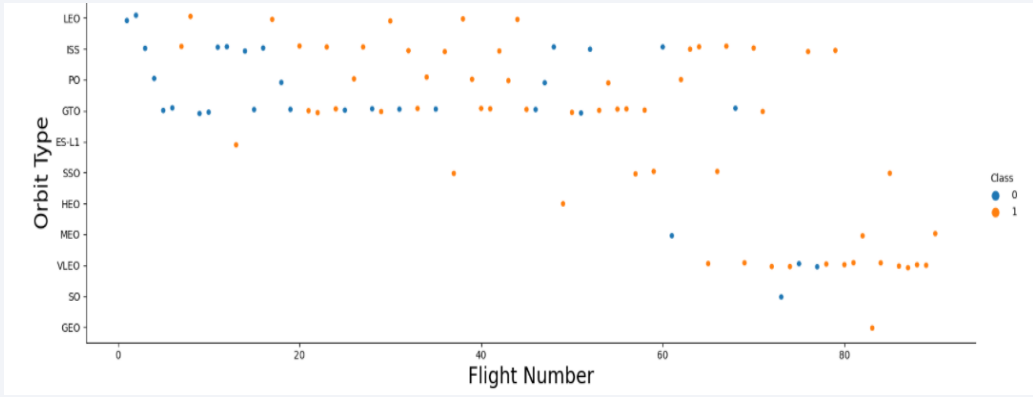


EDA with Data Visualization

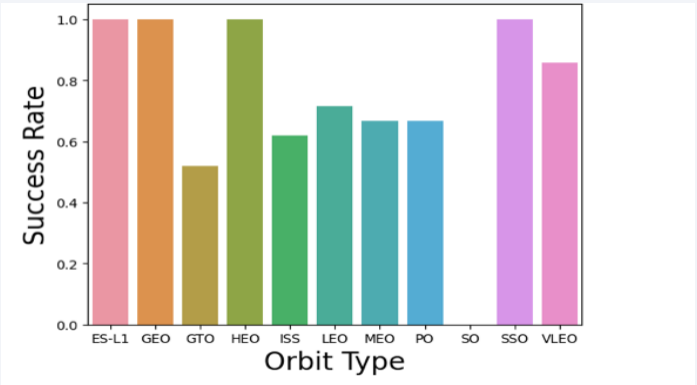


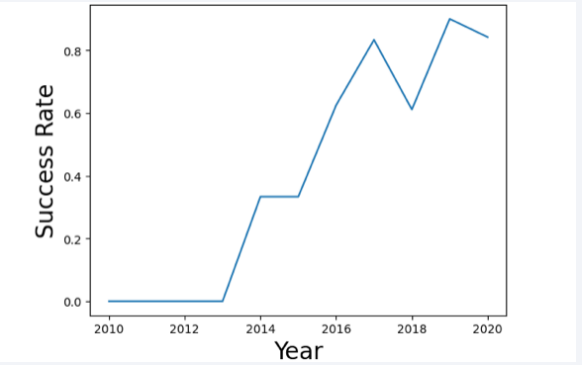






https://github.com/Monther-alharbi/testrepo.git





EDA with SQL

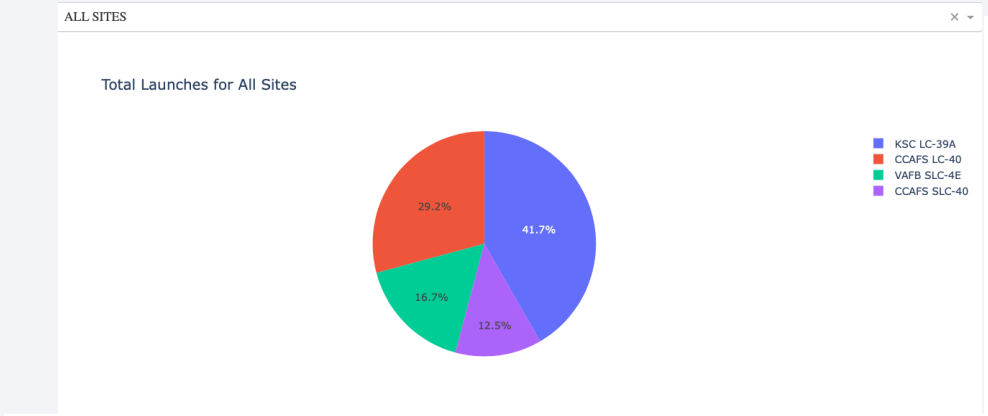
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1¶
- List the date when the first succesful 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.¶

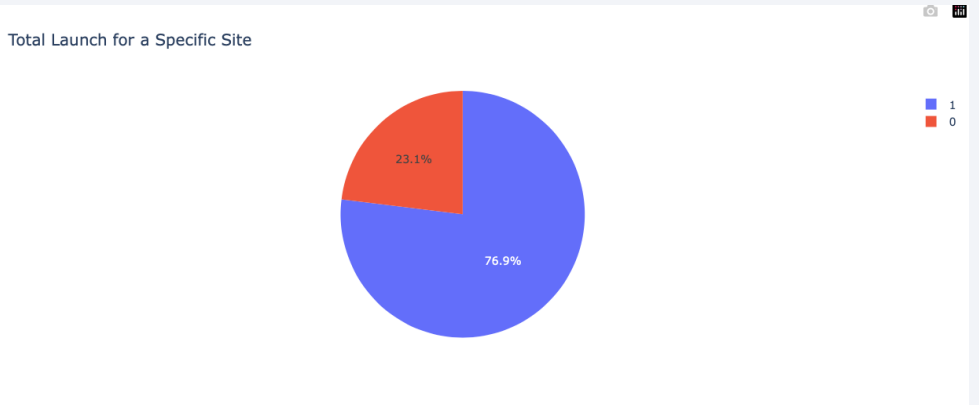
https://github.com/Monther-alharbi/testrepo.git

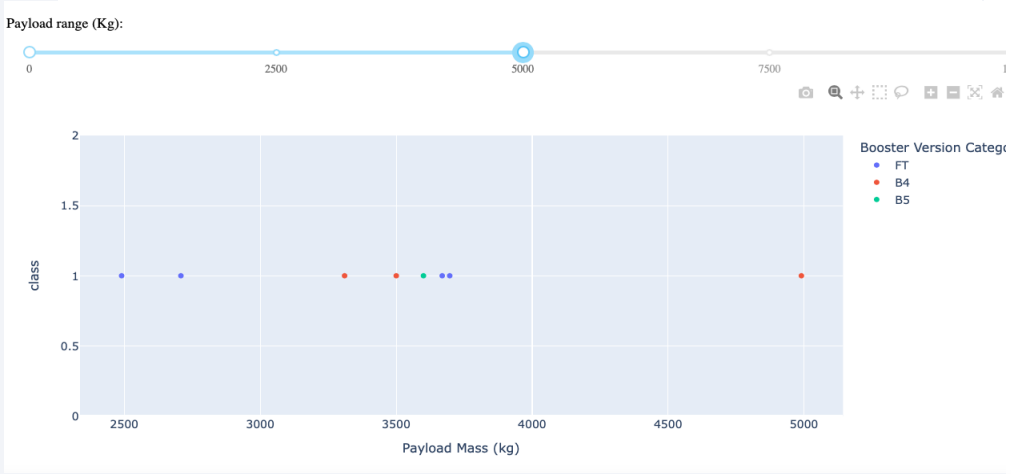
Build an Interactive Map with Folium

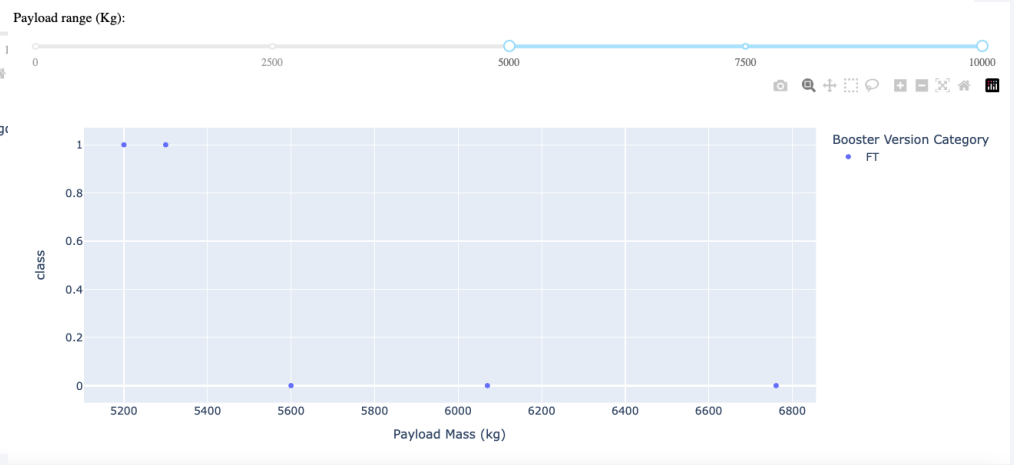
https://github.com/Monther-alharbi/testrepo.git

Build a Dashboard with Plotly Dash



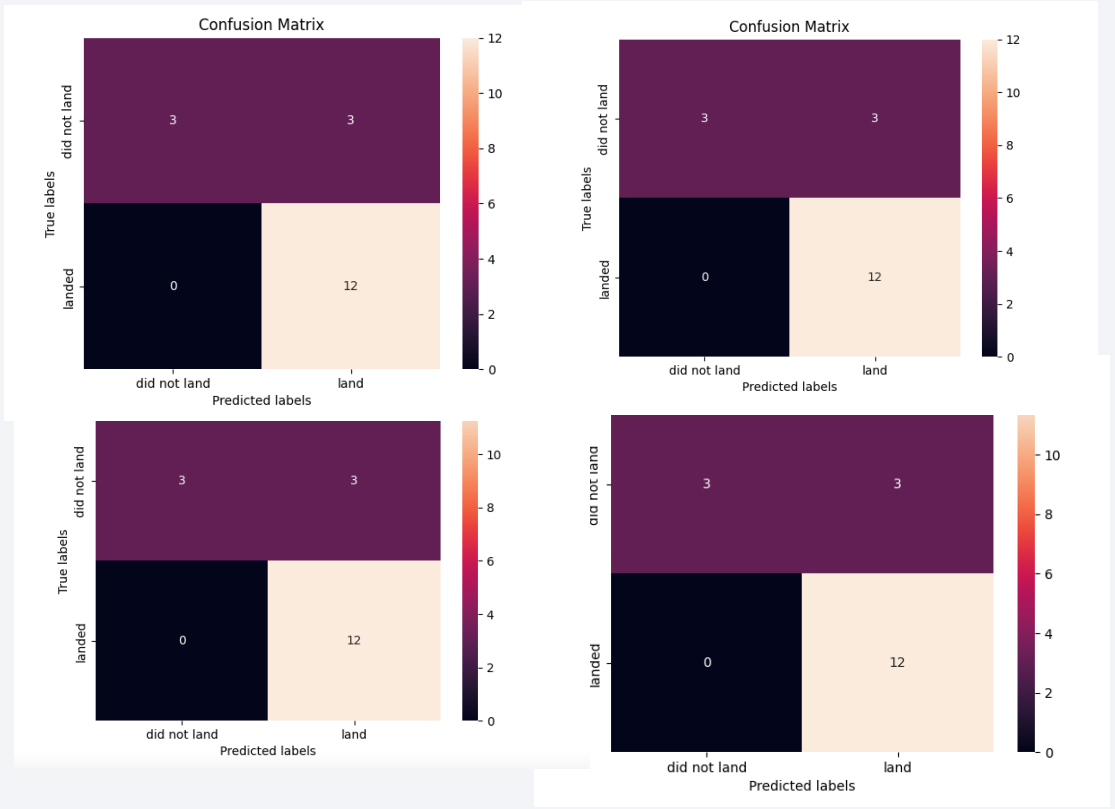






https://github.com/Monther-alharbi/testrepo.git

Predictive Analysis (Classification)



	Algorithm	Accuracy
0	Logistic Regression	0.846429
1	SVM	0.848214
2	KNN	0.848214
3	Decision Tree	0.901786

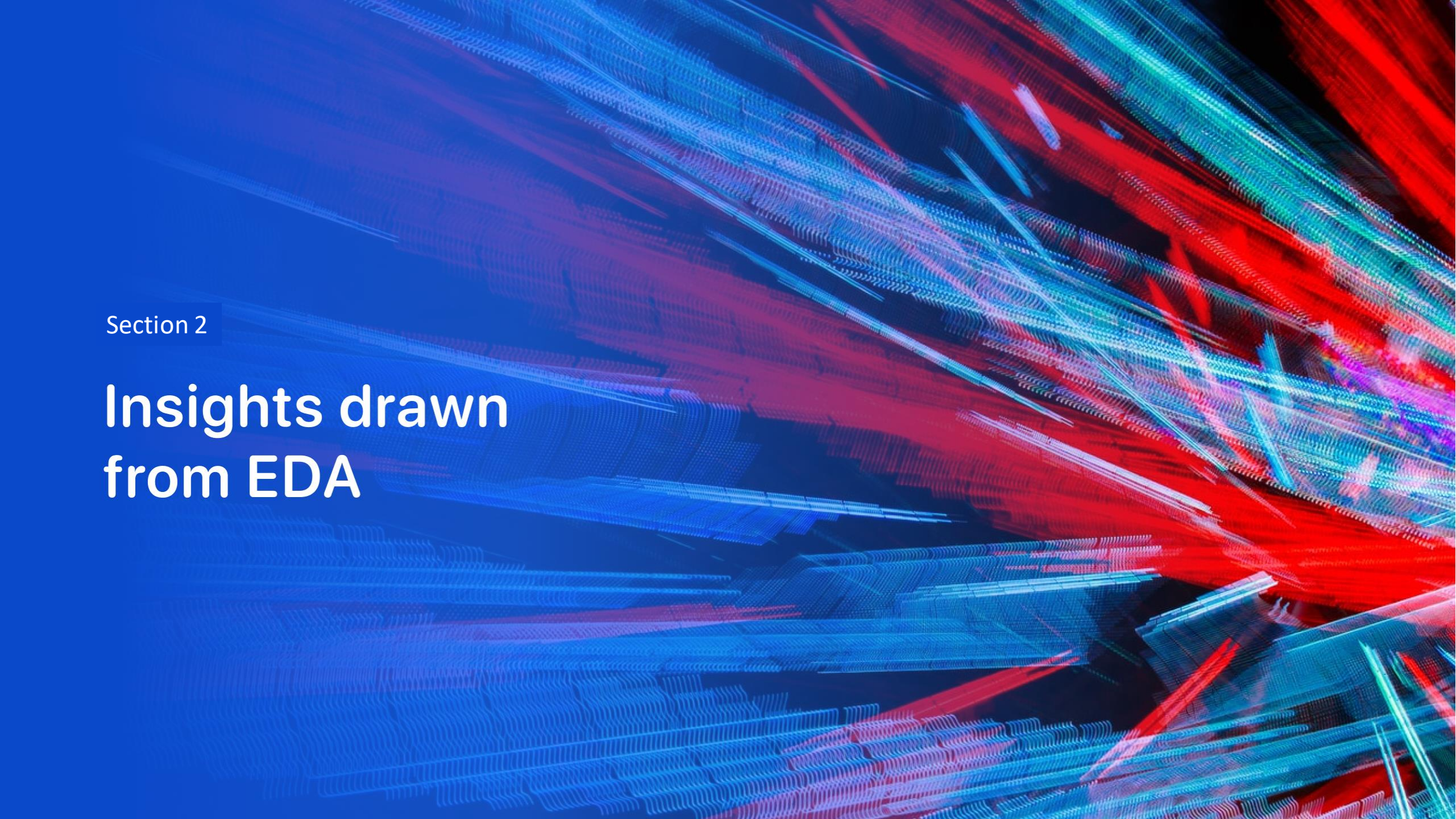
https://github.com/Monther-alharbi/testrepo.git

Results

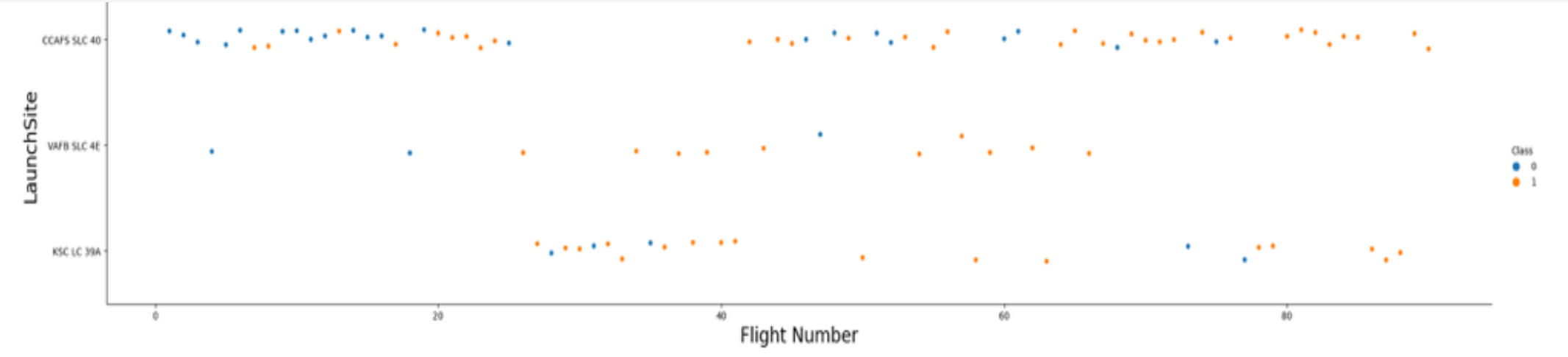
- KSC LC-39A has the highest score site
- Payload of 0kg to 5000kg was more diverse
- Decision tree was the best model with accuracy of 0.90

Section 2

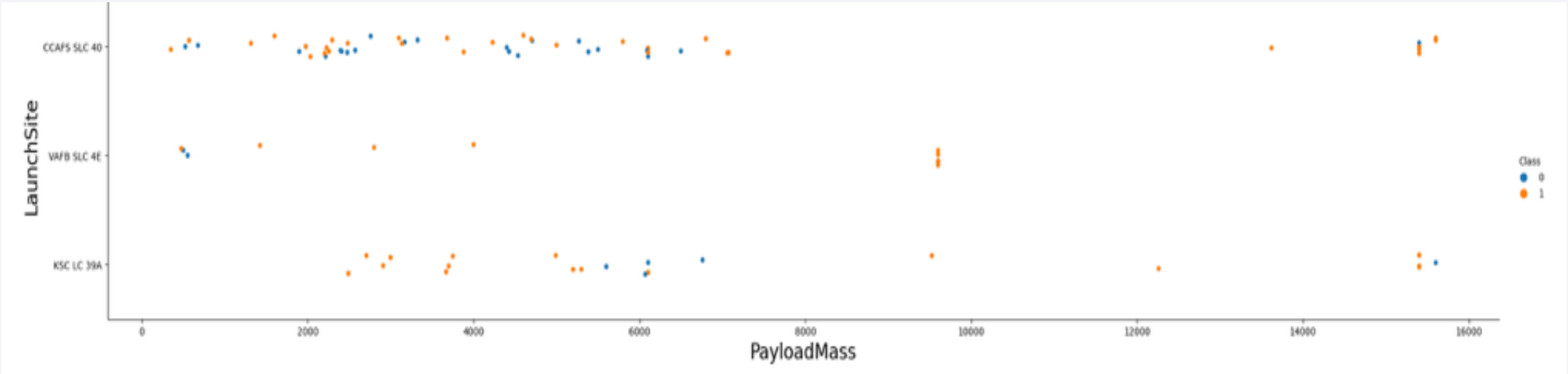
Insights drawn from EDA



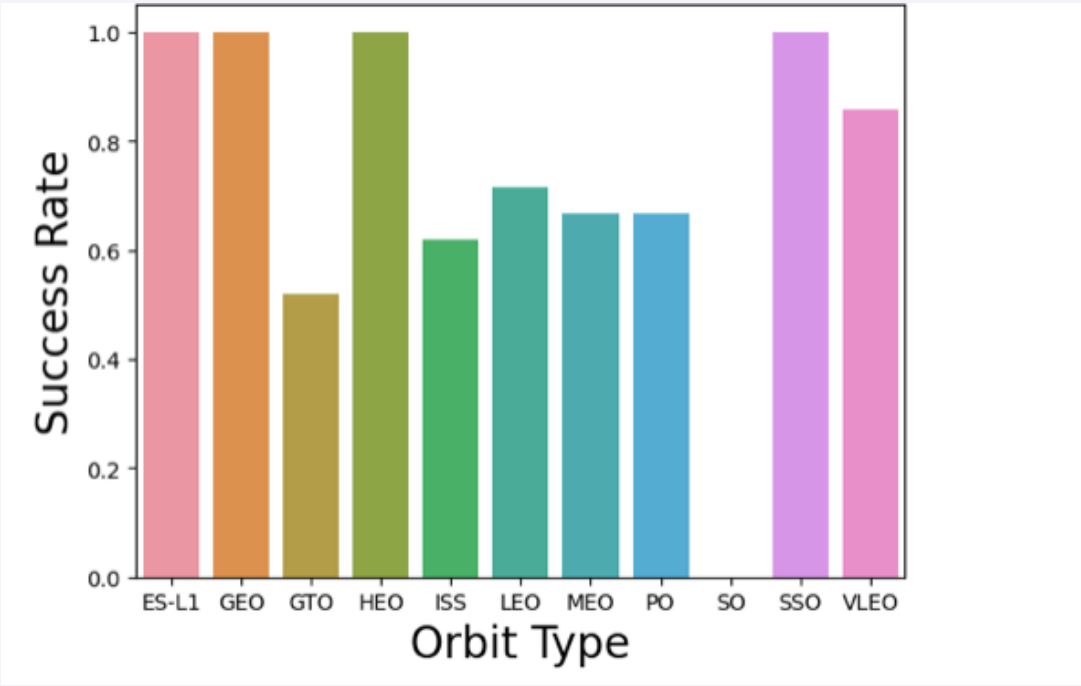
Flight Number vs. Launch Site



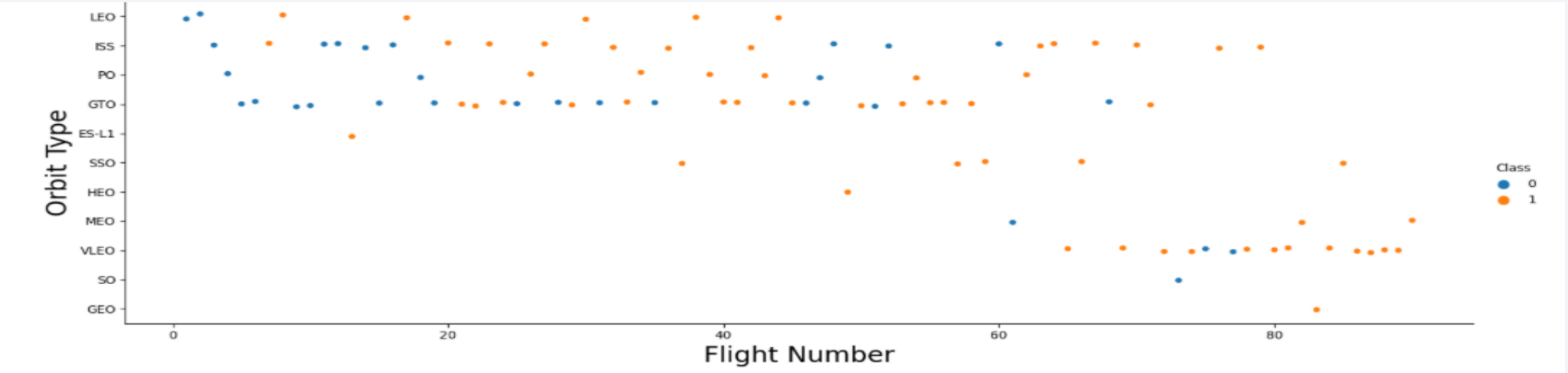
Payload vs. Launch Site



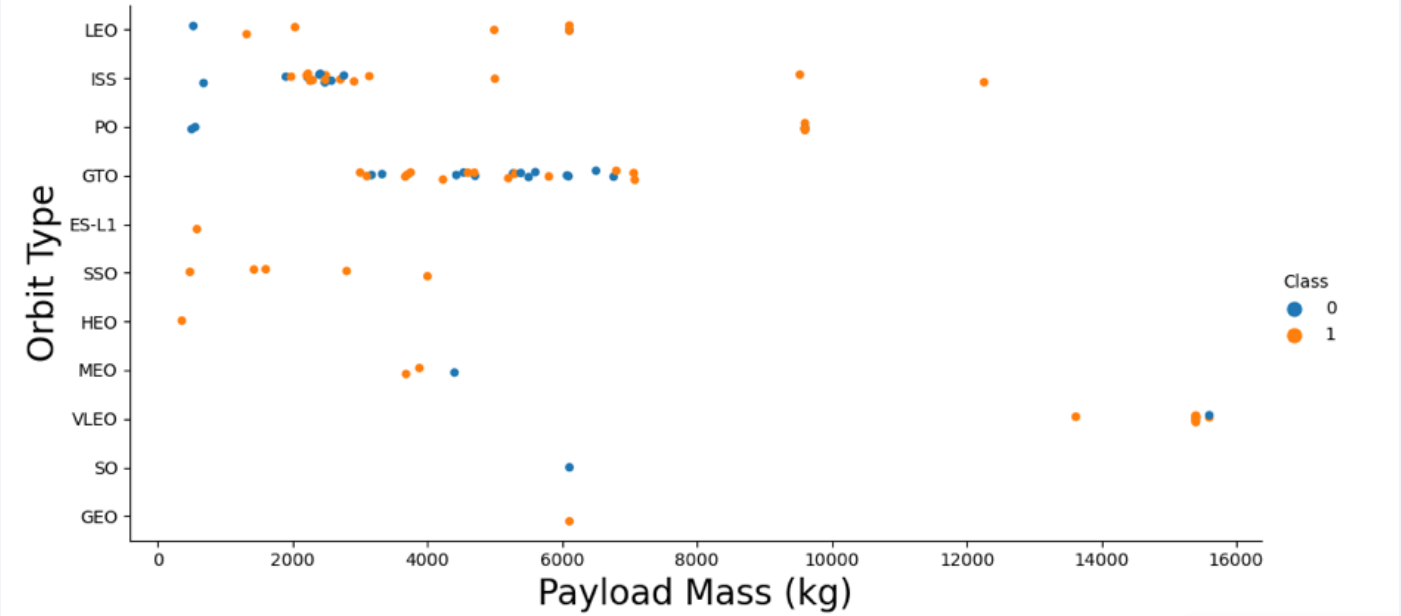
Success Rate vs. Orbit Type



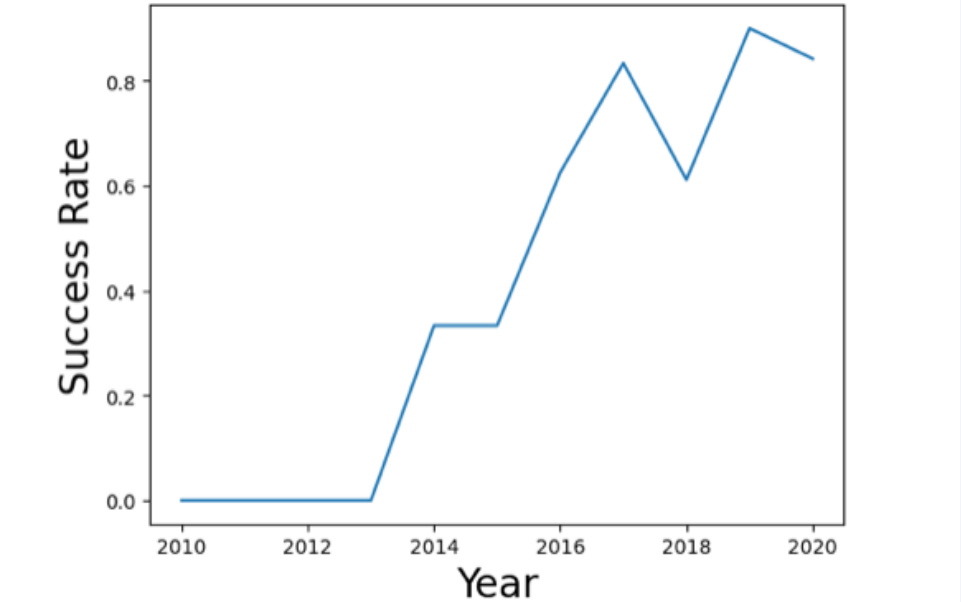
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
[11]: %sql select distinct Launch_Site from SPACEXTBL
```

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

[11]:

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

Launch Site Names Begin with 'CCA'

```
12]: %sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
 * sqlite:///my_data1.db
Done.
```

12]:

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

Total Payload Mass

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer='NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version='F9 v1.1'
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTBL where Landing_Outcome='Success (ground pad)'
```

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

min(Date)
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct Booster_Version from SPACEXTBL where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select substr(Mission_Outcome,1,7) as Mission_Outcome, count(*) from SPACEXTBL  group by 1
```

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

Mission_Outcome	count(*)
None	898
Failure	1
Success	100

Boosters Carried Maximum Payload

```
%sql select distinct Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

2015 Launch Records

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome FROM SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and
```

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

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)

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

```
%sql SELECT Landing_Outcome, count(*) as count_outcomes FROM SPACEXTBL WHERE DATE between '04-06-2010' and '20-03-2017' group by Landing_Outcome order by
```

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

Landing_Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

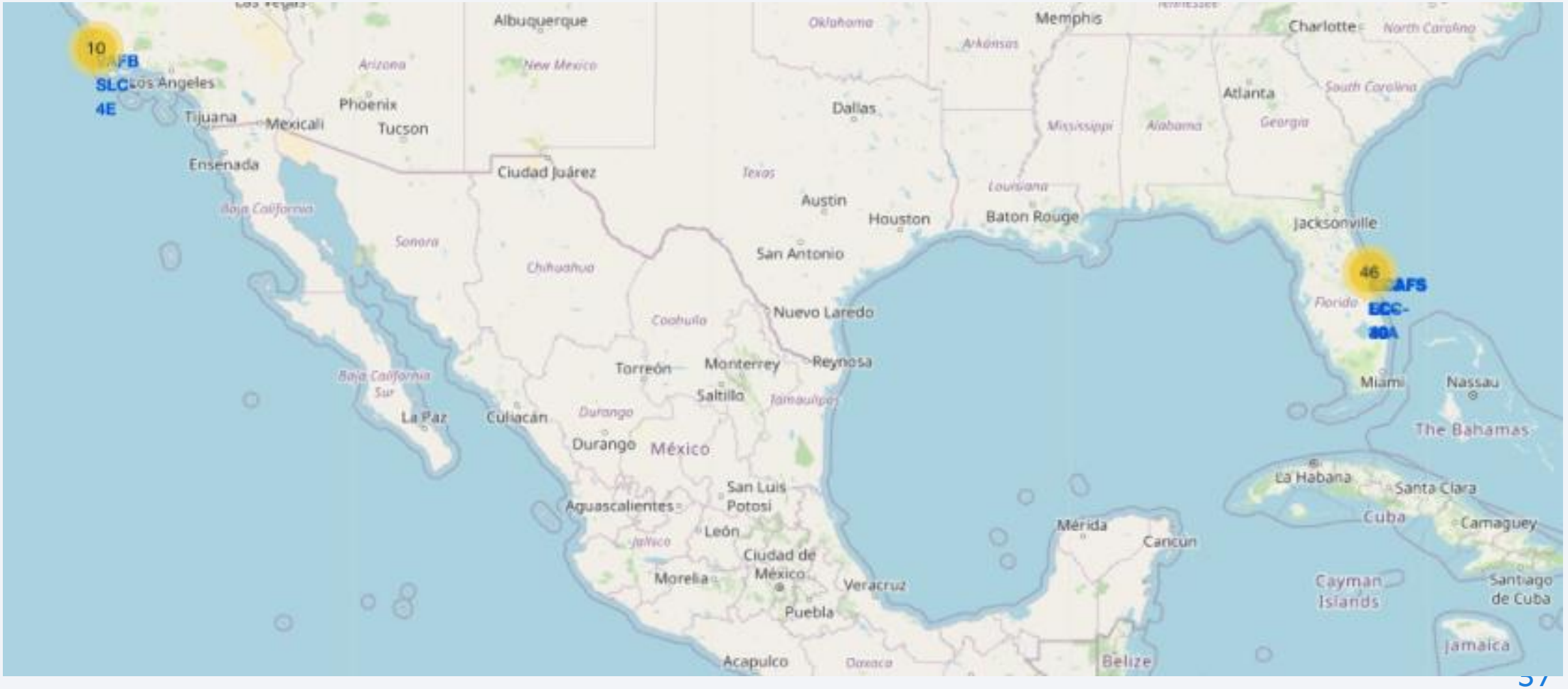

Section 3
Launch Sites
Proximities Analysis

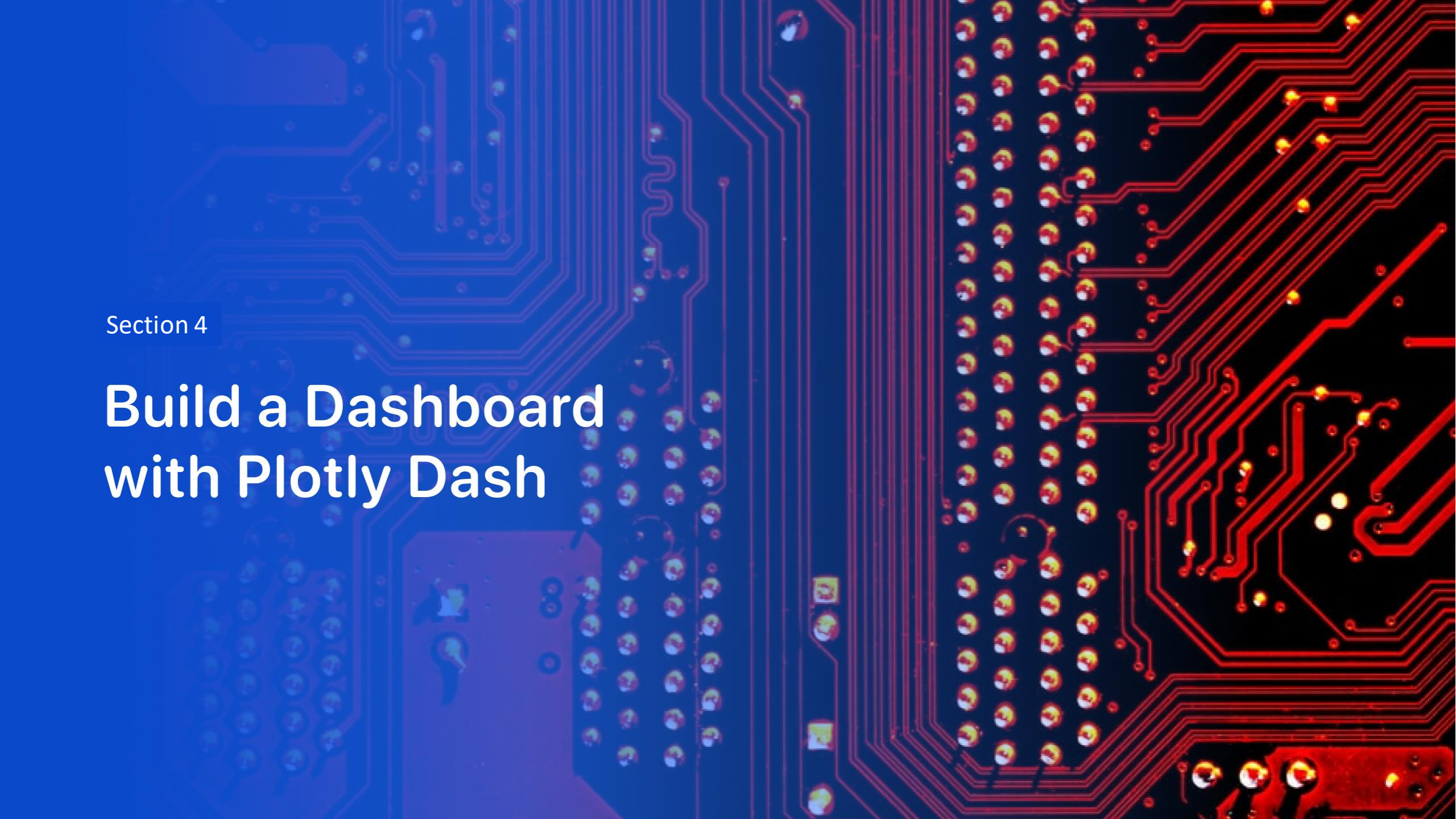
All launch sites marked on map

success/failed launches for each site on the map

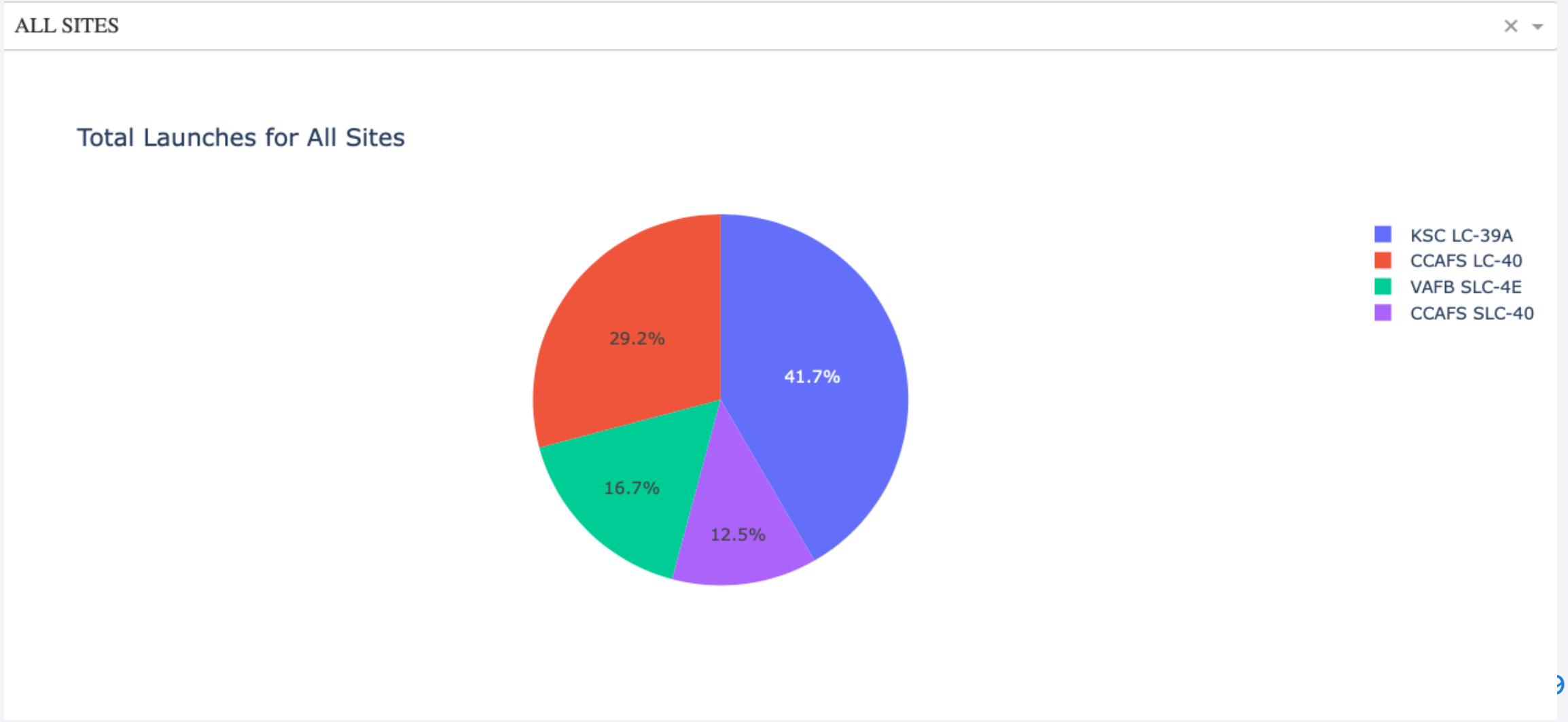
- Green means success
- Red means failed

Distances between a launch site to its proximities

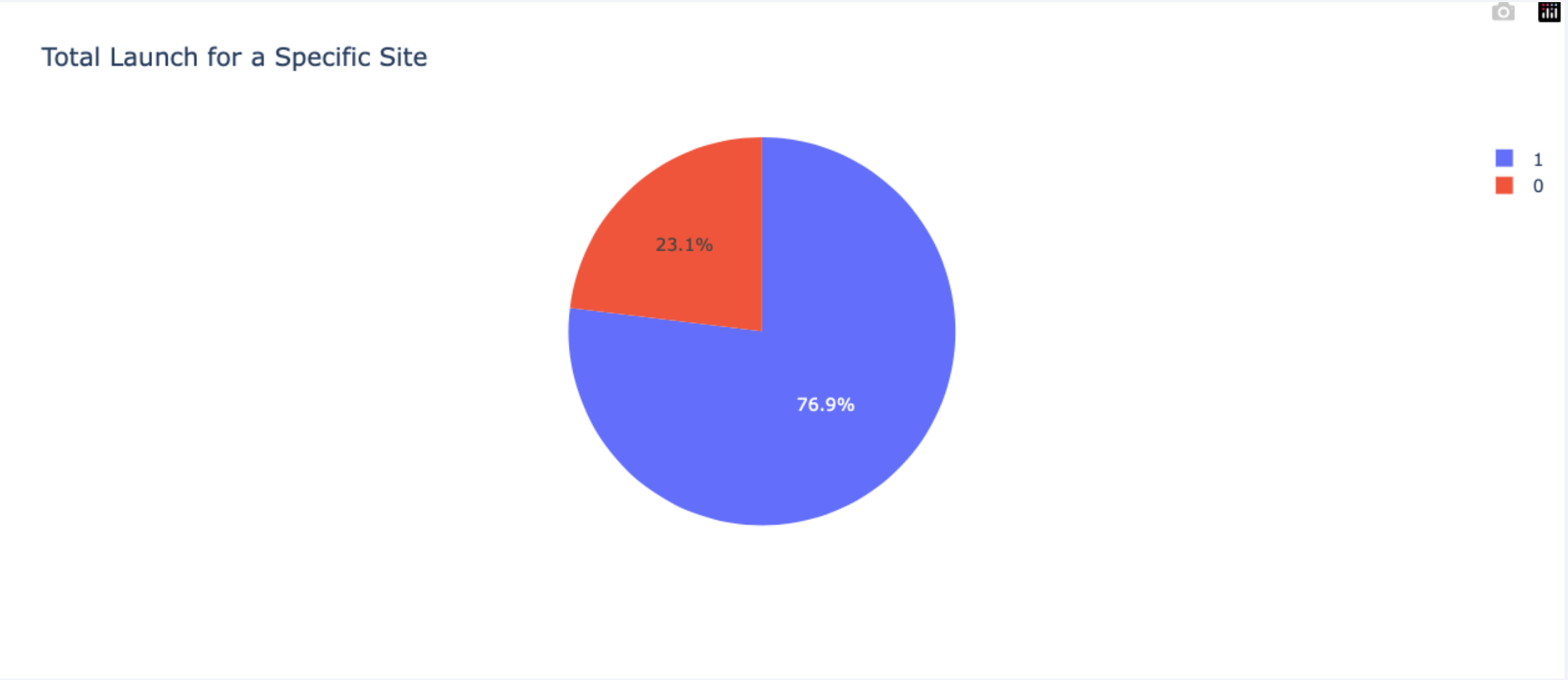



Section 4
Build a Dashboard with Plotly Dash

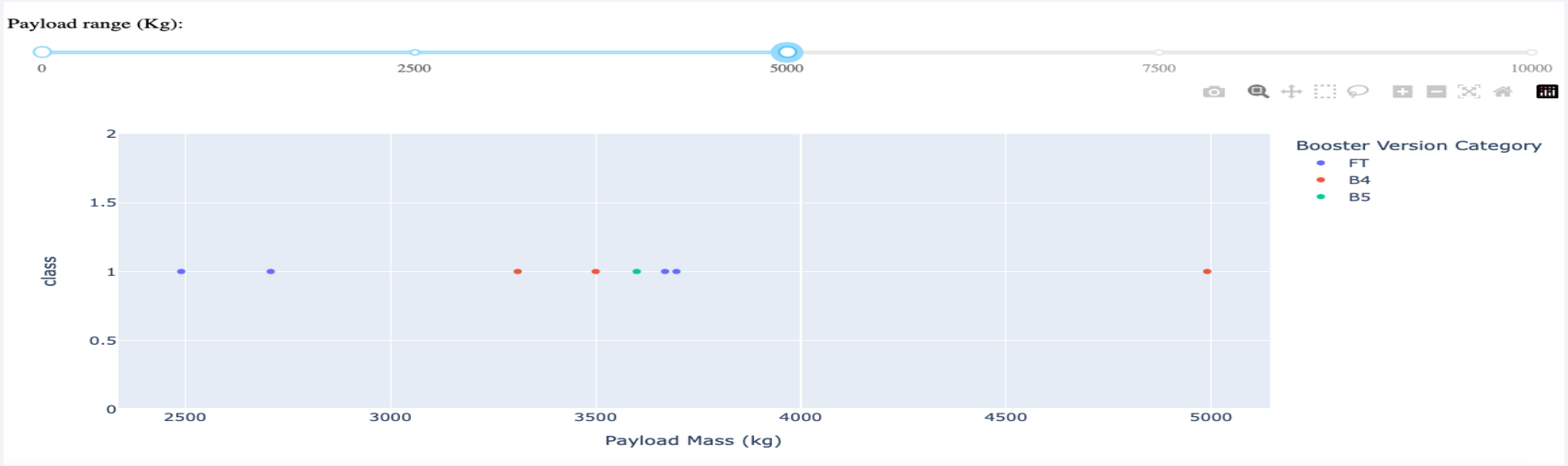
Total success lanunches by all sites

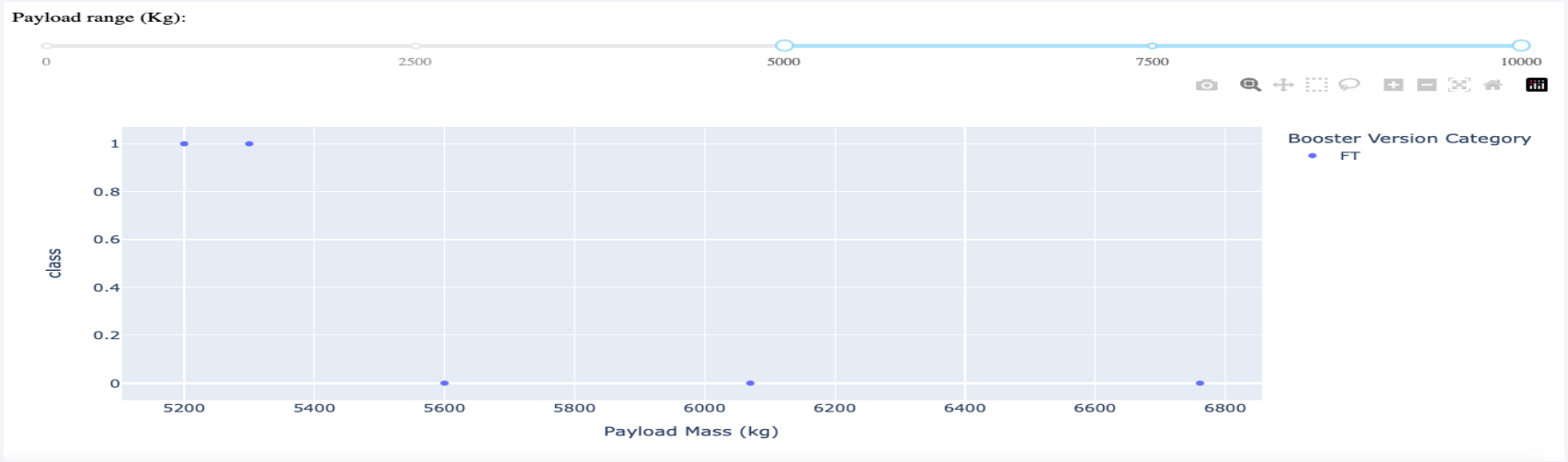


Success rate by site



<Payload vs launch outcome>





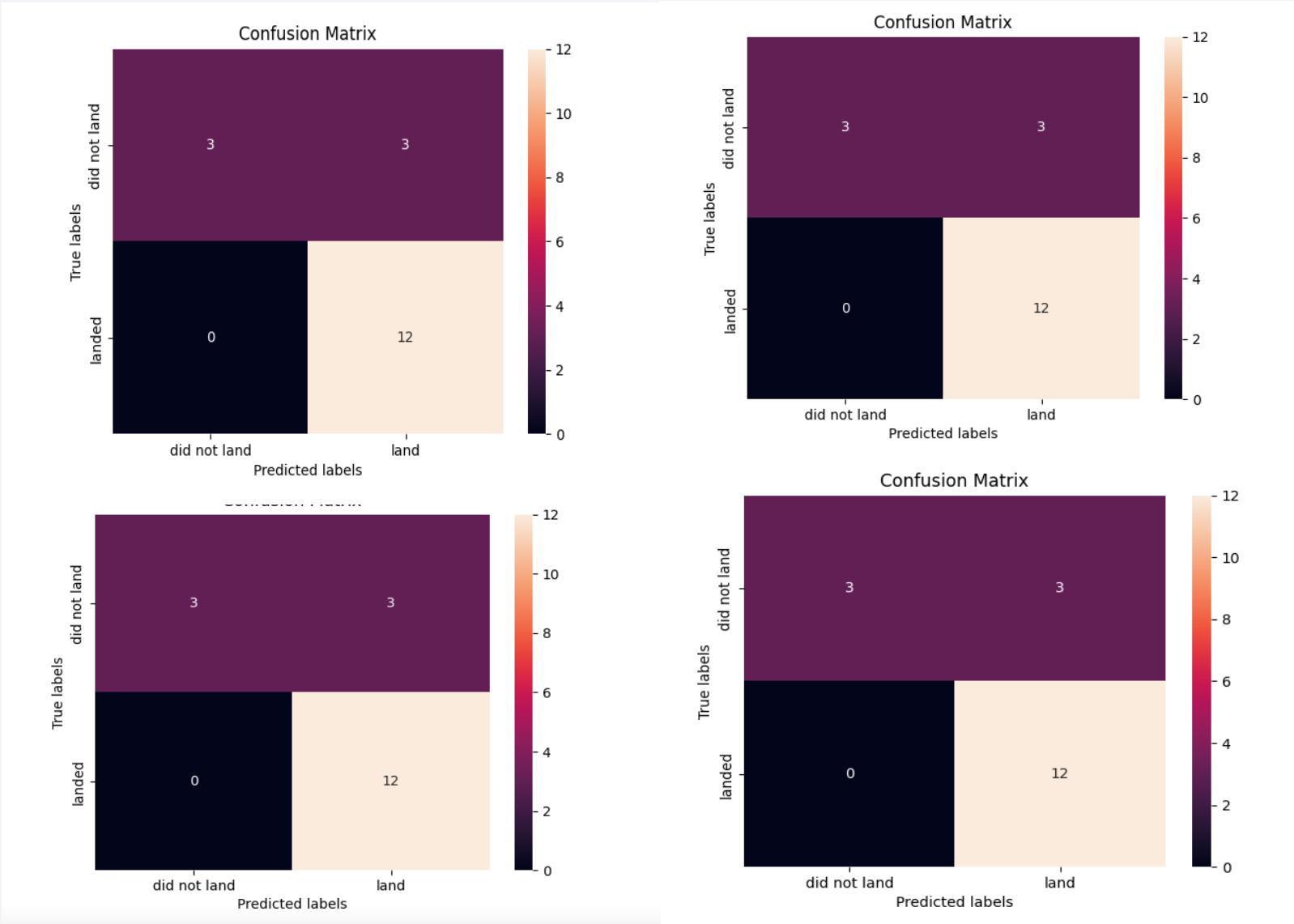
Section 5

Predictive Analysis (Classification)

Classification Accuracy

	Algorithm	Accuracy
0	Logistic Regression	0.846429
1	SVM	0.848214
2	KNN	0.848214
3	Decision Tree	0.901786

Confusion Matrix



Conclusions

- We know the launch sites distance to its proximities
- KSC LC-39A has the highest score site
- Payload of 0kg to 5000kg was more diverse
- Decision tree was the best model with accuracy of 0.90

Appendix

- https://github.com/Monther-alharbi/testrepo.git

Thank you!