Winning Space Race with Data Science

By :
Vikash Yadav
18/05/2022

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

To understand, analyse and predict rocket launch success for a space company by using data of most successful company in this industry, SpaceX.

Results include success rate based on payload, time, location of launch site and dashboard for clients to analyse and get insights into industry.

Following model can be improved further on availability of more Data in future.

Introduction

- Methodology
- Exploratory Data Analysis and Visualization
- Launch Site Proximities analysis
- Results of success and failed launches
- Building Machine Learning Model

Section 1

Methodology

Methodology

- Executive Summary
- Data collection :
  - Rest API
  - Web Scrapping
- Data wrangling with pandas
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Testing different classifiers using scikit-learn library
  - Calculating accuracy of different models

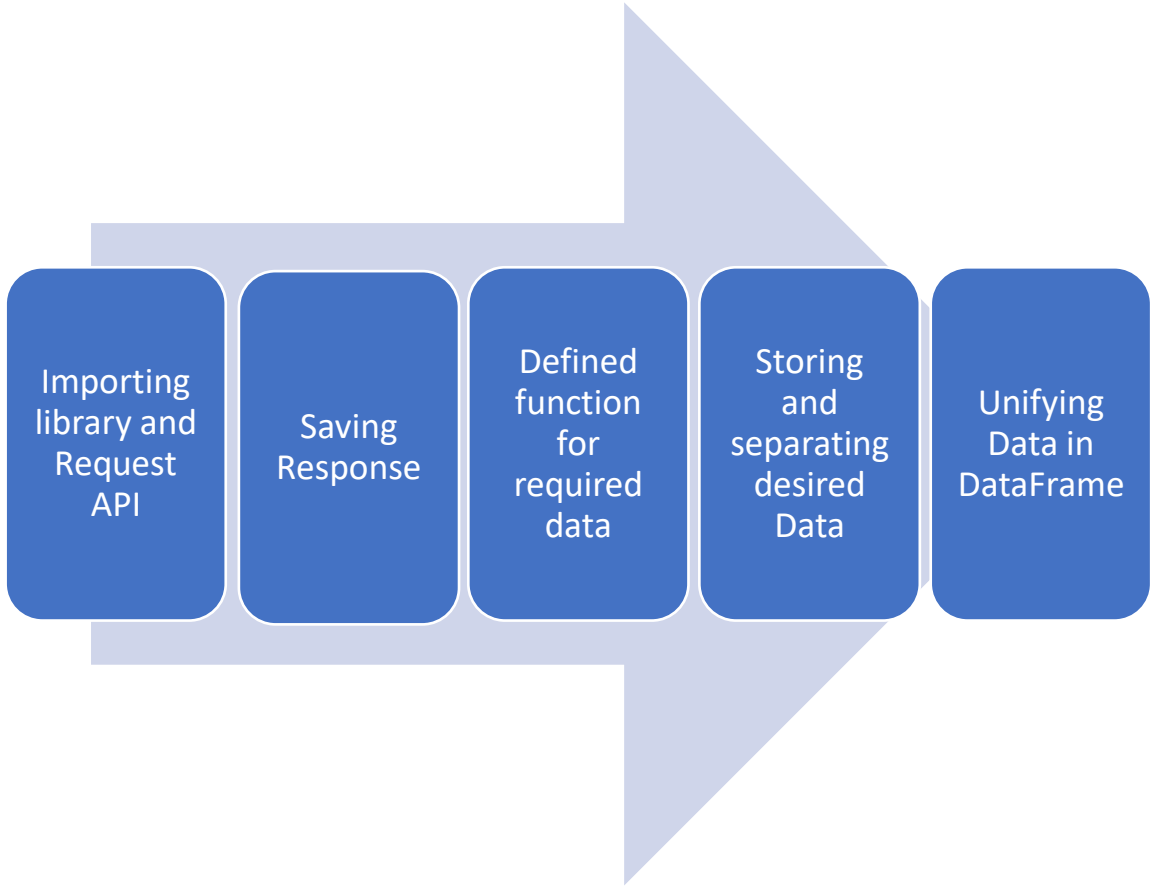
Data Collection

- Master data from SpaceX official API using Python library requests
- Additional Data from Wikipedia page with Web Scrapping using Python Beautifulsoup library

Data Collection – SpaceX API

Master data using requests library on SpaceX official API

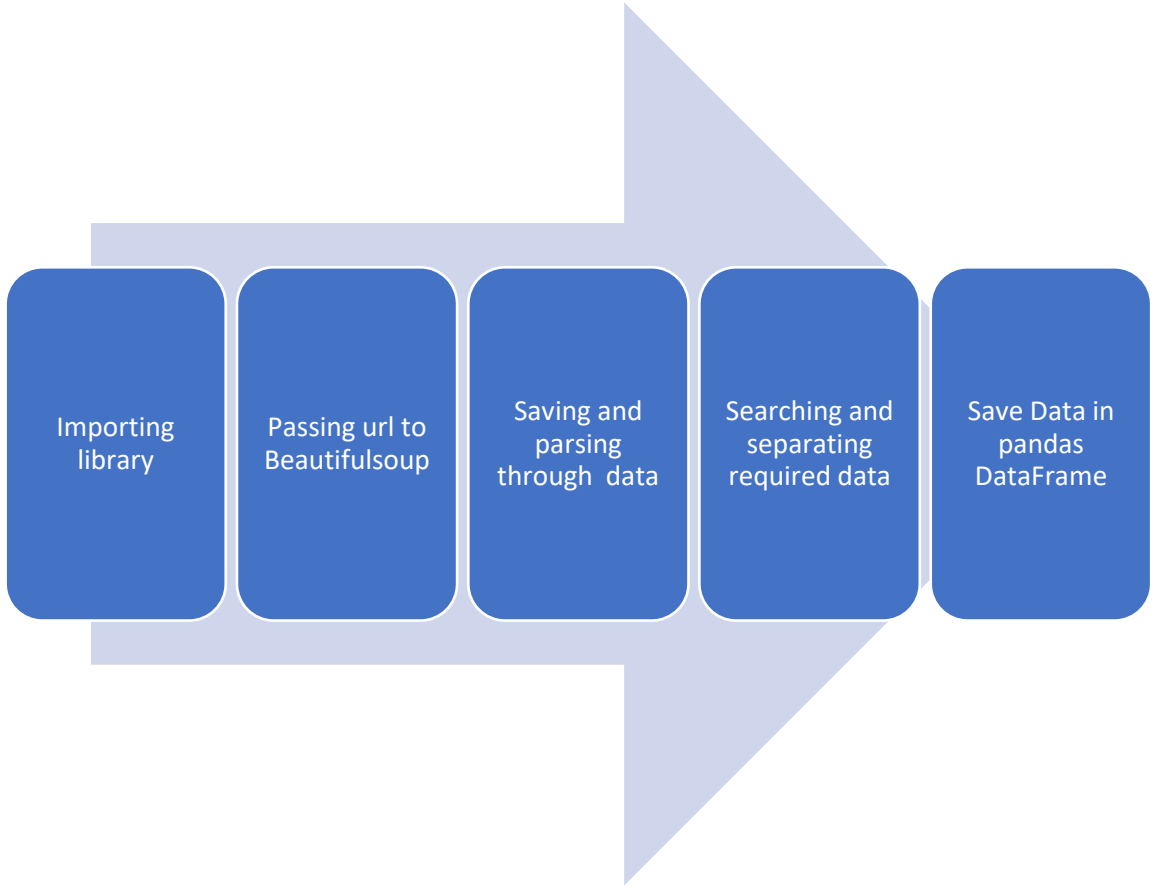
Data Collection Notebook Link



Data Collection – Scraping

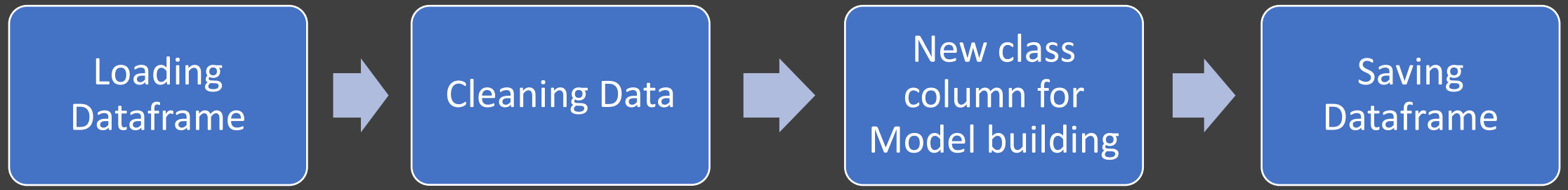
Additional data from Wikipedia using Beautifulsoup library

Web Notebook Link



Data Wrangling

Flowchart :



Data Wrangling Notebook Link

EDA with Data Visualization

Using Python Visualization tools for analyzing relations :

- Flight number vs Launch site
- Payload vs launch site
- Orbit vs success rate
- Flight number vs orbit type
- Payload vs orbit type
- Launch success yearly trend

EDA with Visualization Notebook Link

EDA with SQL

Using SQL queries to explore data :

- Site names
- Payload from specific clients
- Success dates and success landings
- Mission outcomes and more

EDA with SQL Notebook link

Build an Interactive Map with Folium

To get a better understanding behind Launch site locations folium map was used and process is represented by flowchart :

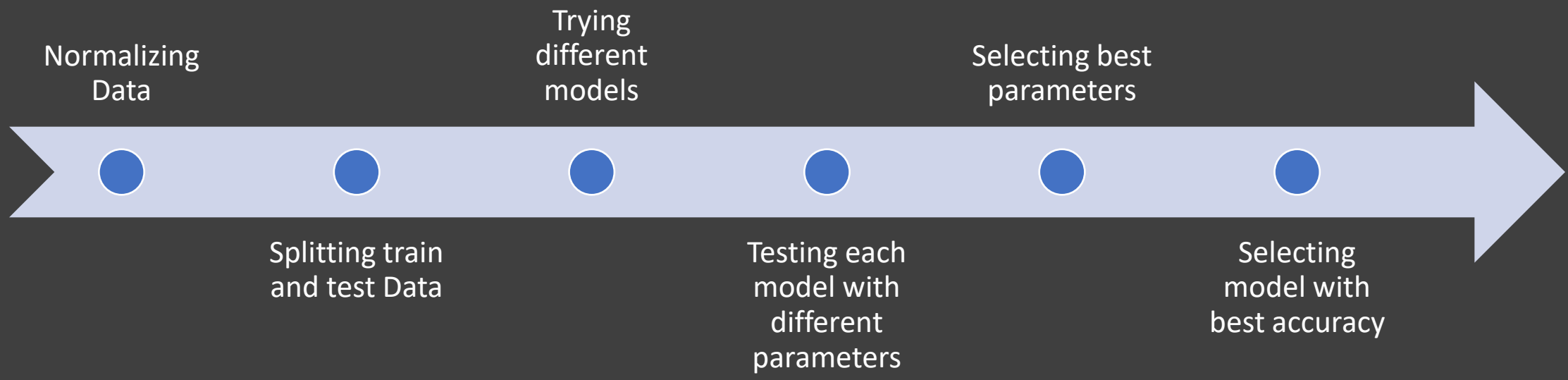
[Launch Site Map Notebook Link]

Build a Dashboard with Plotly Dash

- Dashboard built with pie chart and scatter plot to understand success ratio properly and its relationship with site location and payload.

Dashboard Code Notebook Link

Predictive Analysis (Classification)



Machine Learning Notebook Link

Results

Exploratory data analysis results :

Different plots to analyze relation between payload mass, flight number, orbit etc.

SQL queries to get desired outputs for analysis

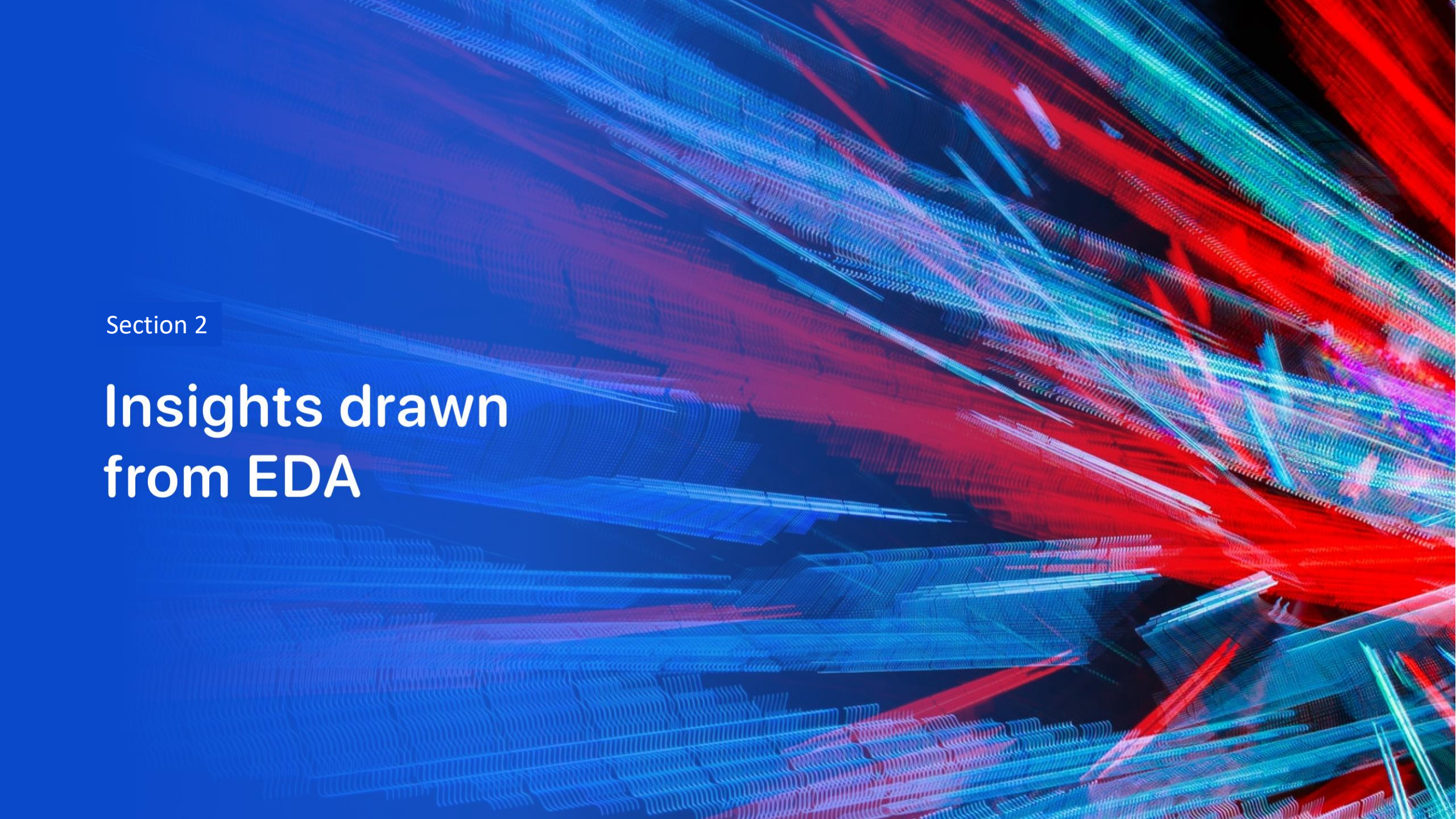
Interactive analytics :

Dashboard with Plotly Dash to check success ratio in relation with Launch Site and Payload Mass.

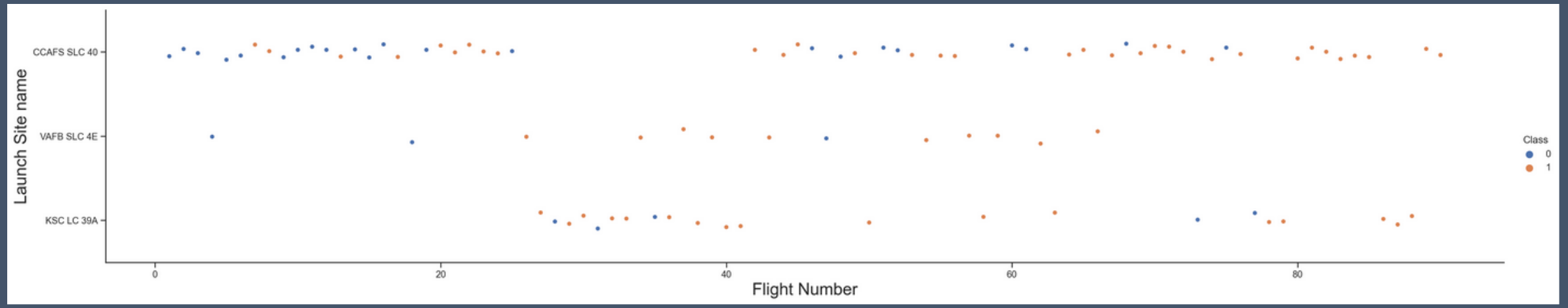
Predictive analysis results :

Tree Map Classification is best.

No major difference between Models because of small data set.


Section 2
Insights drawn from EDA

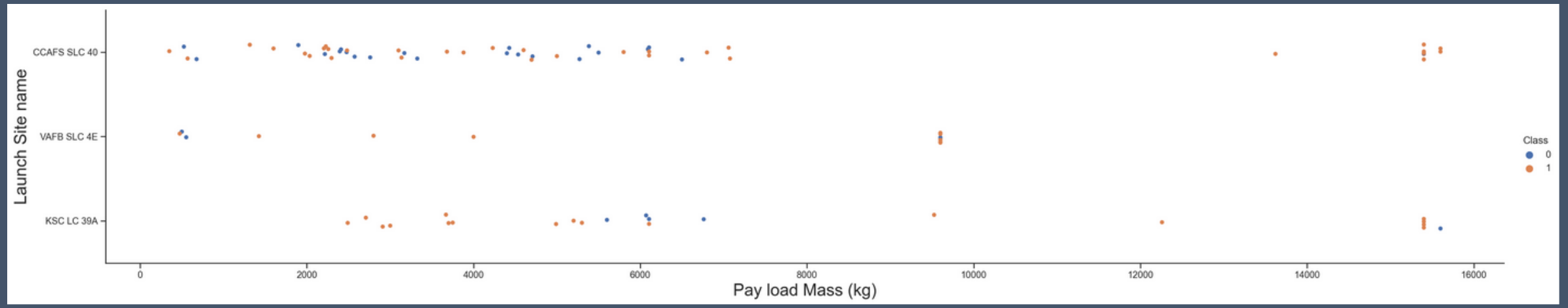
Flight Number vs. Launch Site



Interesting information from this scatter includes :

- CCAFS SLC-40 is most used site, specially at beginning
- Low success rate for each site in beginning and then it improved.

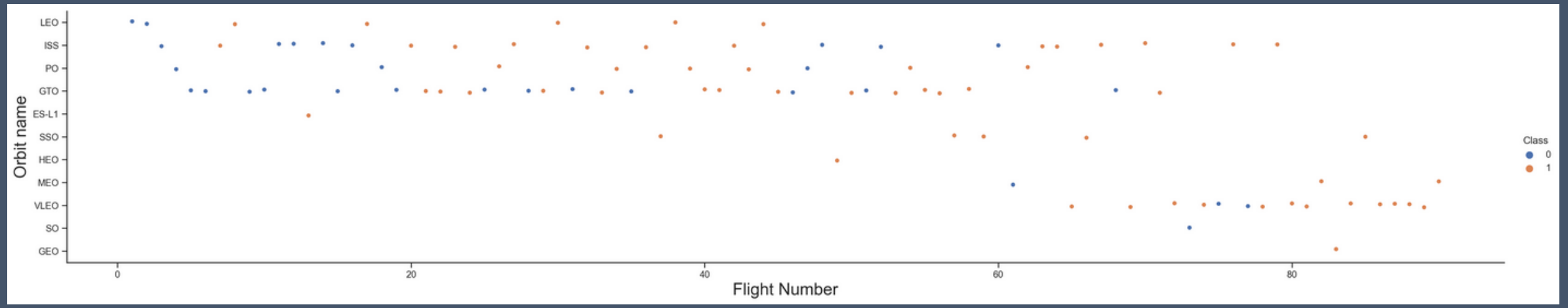
Payload vs. Launch Site



Interesting information from this scatter includes :

- CCAFS SLC-40 and KSC LC 39A are used for very high payload
- KSC LC 39A have highest success ratio

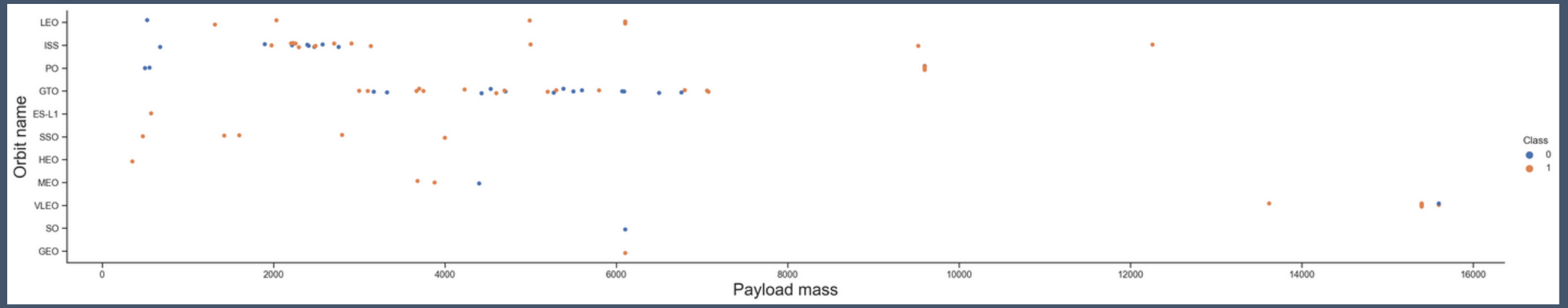
Flight Number vs. Orbit Type



Interesting information from this scatter includes :

- SpaceX started with LEO market and then moved further
- Initially suffered a lot of failed mission outcome

Payload vs. Orbit Type

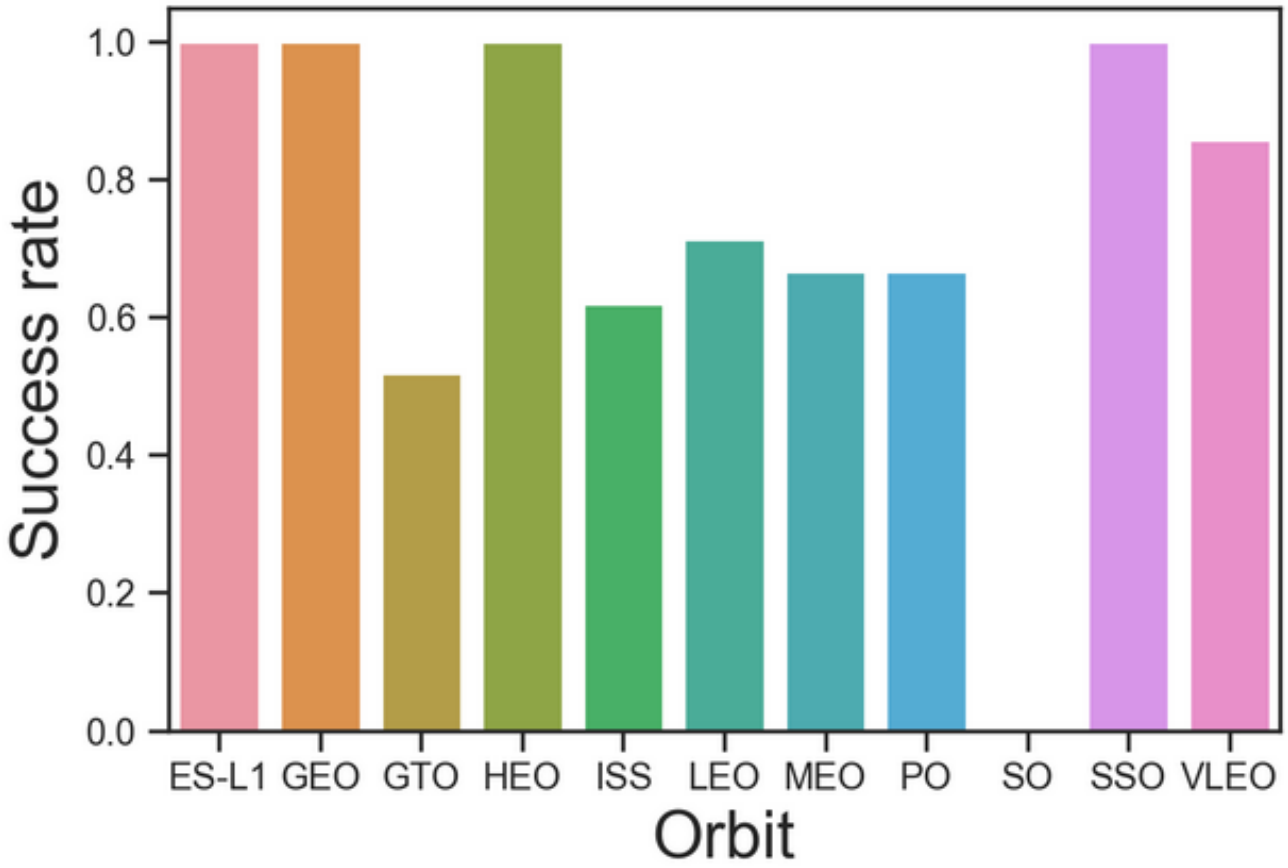


Interesting information from this scatter includes :

- GTO still have a very low success ratio with high payload

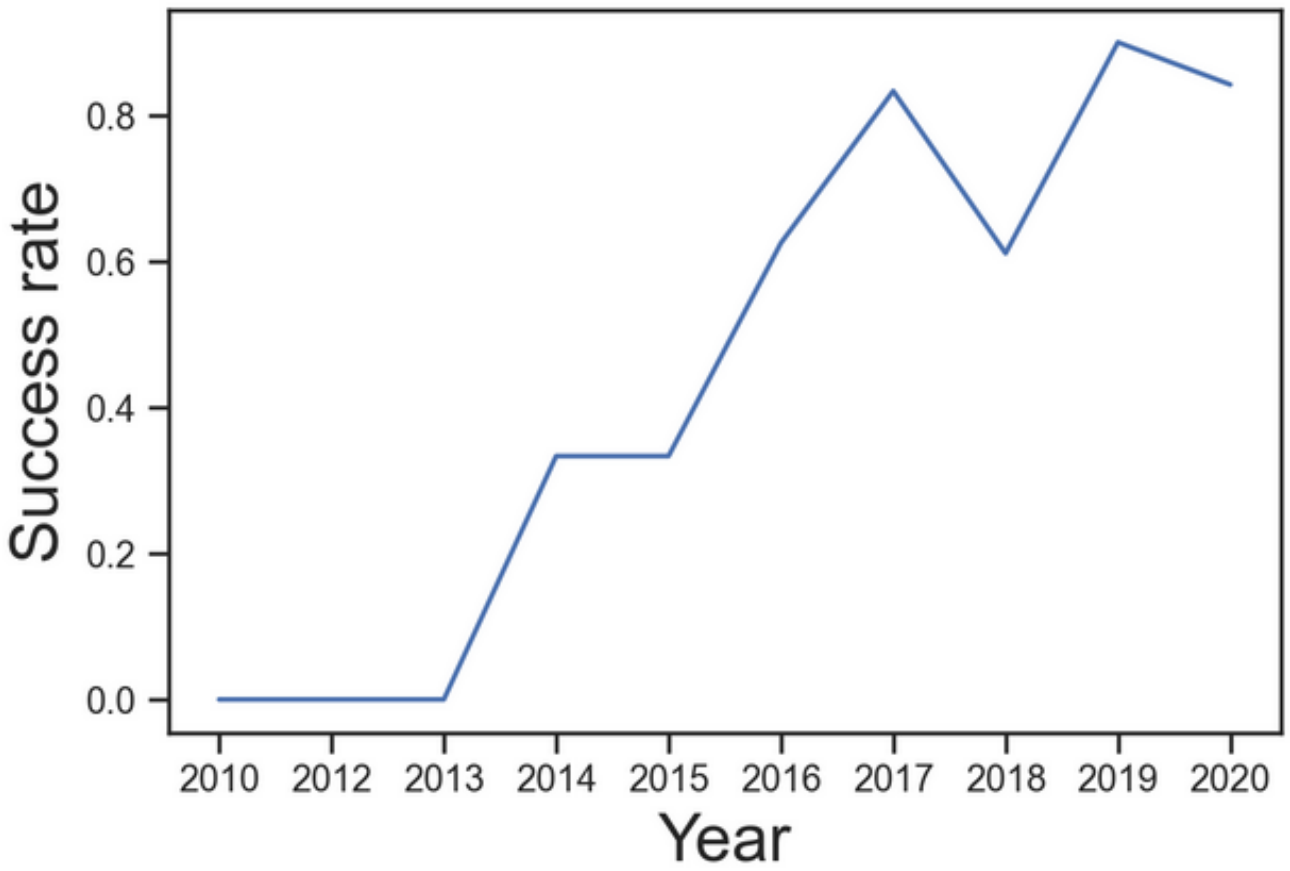
Success Rate vs. Orbit Type

Success rate for orbits vary. It is 100% for GEO, HEO and some other orbits while being around 50% for GTO and 70% for LEO.



Launch Success Yearly Trend

Success rate of SpaceX launches has been improving since 2013. Now it is above 80% . Initially company did struggle to get success for three years.



All Launch Site Names

SQL query and result to find name of All launch sites

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

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[10]:

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

Launch Site Names Begin with 'CCA'

SQL query and result for launches with site name ‘CCA’

[14]:
```
%%sql
select * from SPACEXTBL
where launch_site LIKE 'CCA%'
limit 5 ;
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[14]:

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

Total Payload Mass

SQL query and result for Total payload mass send by NASA in SpaceX rocket

In [13]:
```
%%sql
select  sum(payload_mass__kg_) from SPACEXTBL
where LTRIM(RTRIM(customer)) =  'NASA (CRS)'
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

Out[13]:

1
45596

Average Payload Mass by F9 v1.1

SQL query and result for Average Payload Mass launched from F9 v1.1 booster

[56]:
```
%%sql
select AVG(payload_mass__kg_) from SPACEXTBL
where booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[56]:

1
2534

First Successful Ground Landing Date

SQL query and result for date of first successful launch

[52]:
```
%%sql
select min(date) from SPACEXTBL
where landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[52]:

1
2015-12-22

Successful Drone Ship Landing with Payload(4000-6000Kg)

SQL query and result for list of booster versions of successful launches with payload in range(4000-6000Kg)

[65]:
```
%%sql
select booster_version from  SPACEXTBL
where landing__outcome = 'Success (drone ship)' and  payload_mass__kg_  > 4000 and payload_mass__kg_ < 6000 ;
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[65]:

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 query and result for count of mission outcomes(success/failure)

[71]:
```
%%sql
SELECT mission_outcome, count(mission_outcome) AS CountOf
FROM spacextbl
GROUP BY mission_outcome
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[71]:

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

Boosters Carried Maximum Payload

SQL query and result for list of booster versions which carried maximum payload

[80]:
```
%%sql
select booster_version  from spacextbl
where  *payload_mass__kg_ →*  = (select max( payload_mass__kg_) from spacextbl)
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[80]:

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 Failed Launch Records

SQL query and result for list of booster versions failed (drone ship) landing outcomes in year 2015

[82]:
```
%%sql
select booster_version, launch_site from spacextbl
where date like '2015%' and landing__outcome = 'Failure (drone ship)'
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

[82]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

SQL query and result for list of count of different landing outcomes between desired dates

[88]:
```
%%sql
SELECT landing__outcome, count(landing__outcome) AS CountOf
FROM spacextbl
where date between '2010-06-04' and '2017-03-20'
GROUP BY landing__outcome
```

```
 * ibm_db_sa://mct06771:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

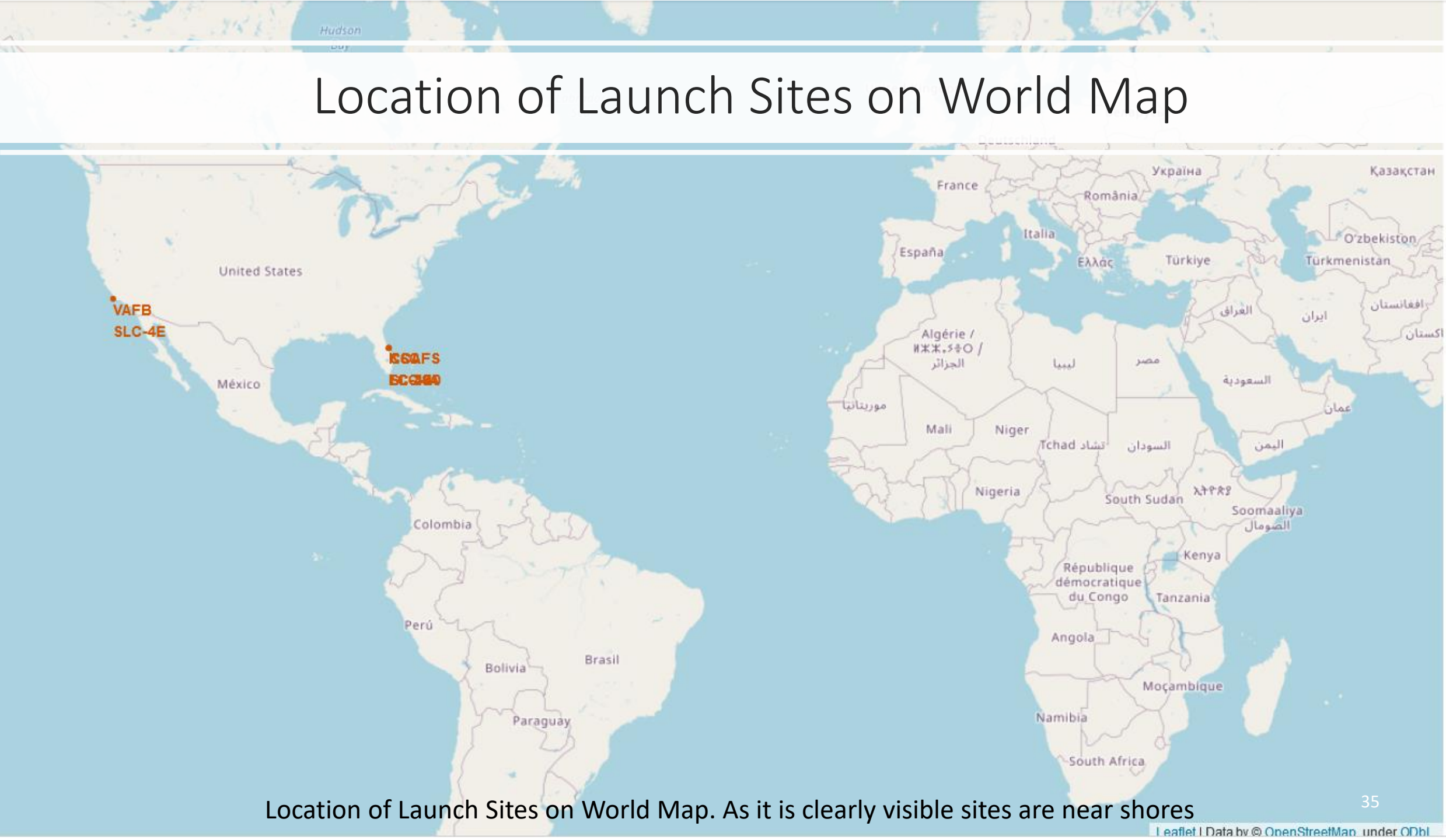
[88]:

landing__outcome	countof
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 3

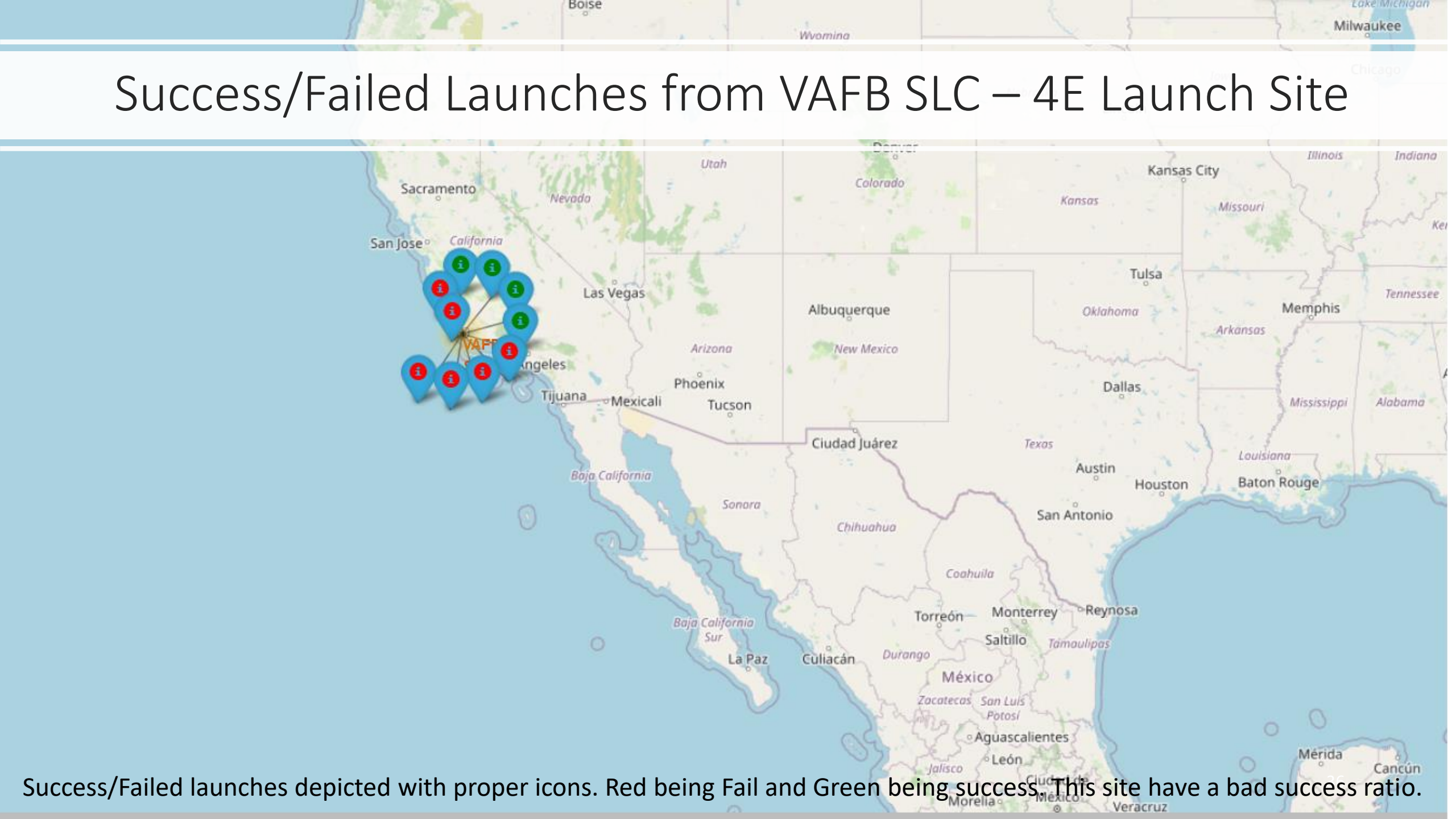
Launch Sites Proximities Analysis

Location of Launch Sites on World Map



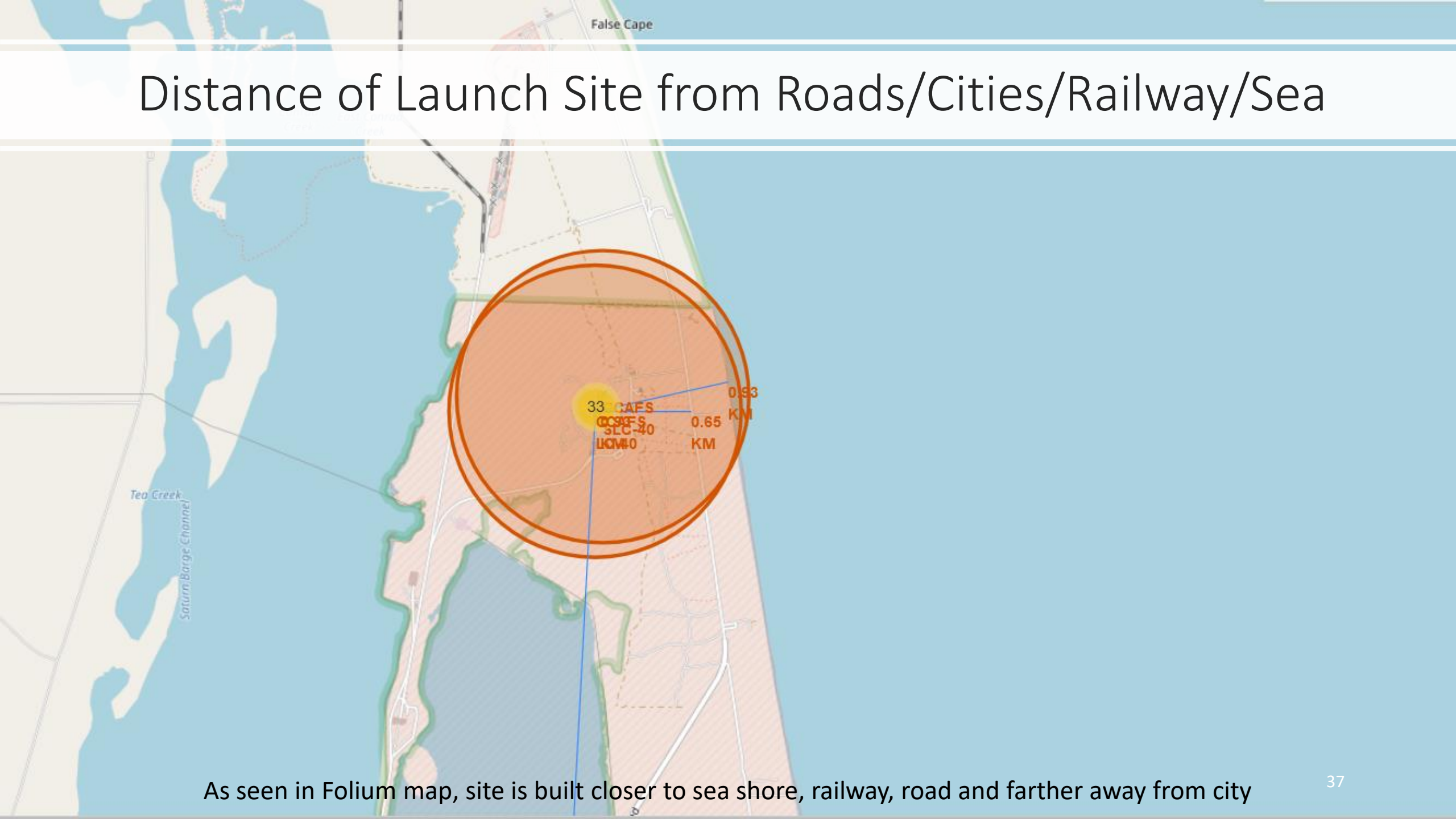
Location of Launch Sites on World Map. As it is clearly visible sites are near shores

Success/Failed Launches from VAFB SLC – 4E Launch Site

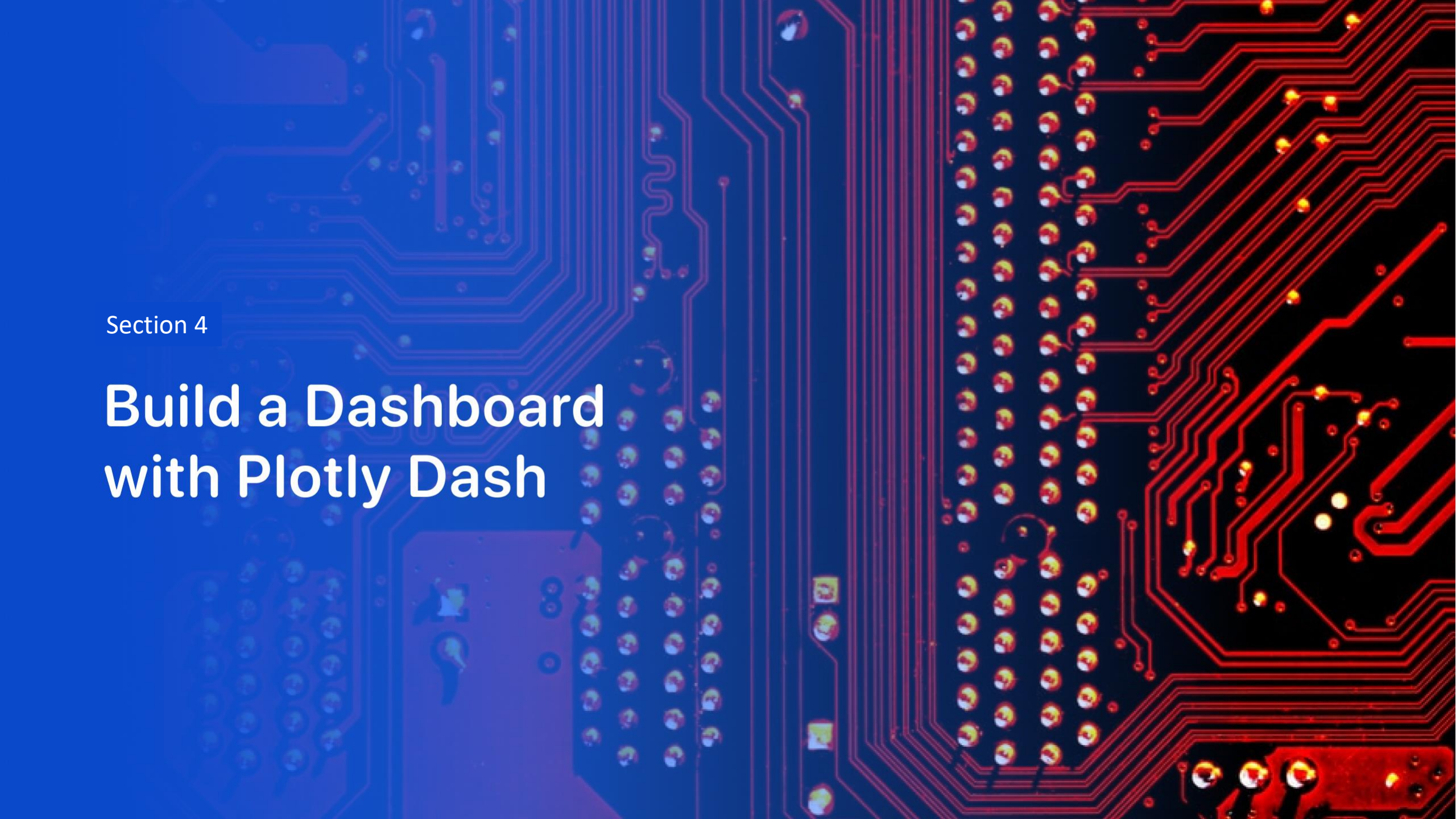


Success/Failed launches depicted with proper icons. Red being Fail and Green being success. This site have a bad success ratio.

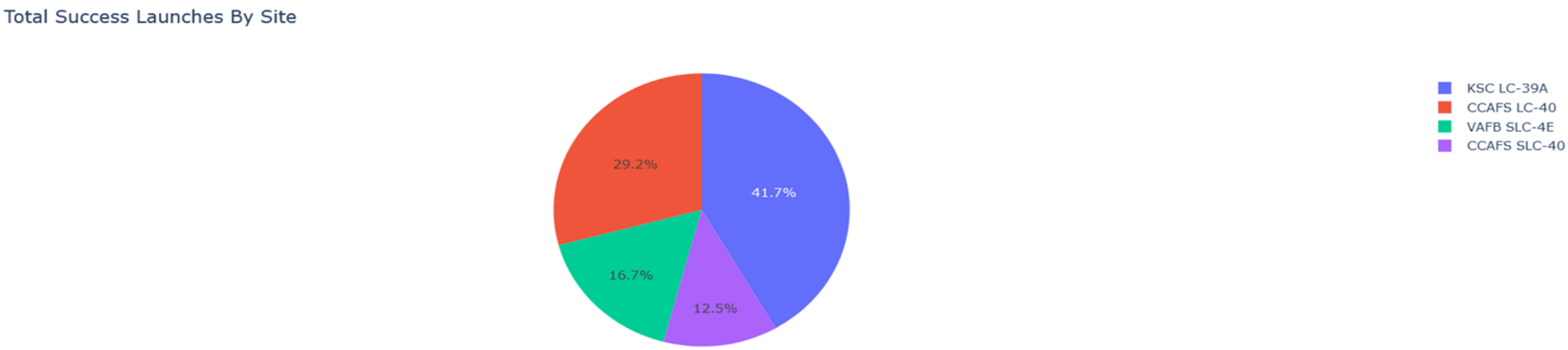
Distance of Launch Site from Roads/Cities/Railway/Sea



As seen in Folium map, site is built closer to sea shore, railway, road and farther away from city

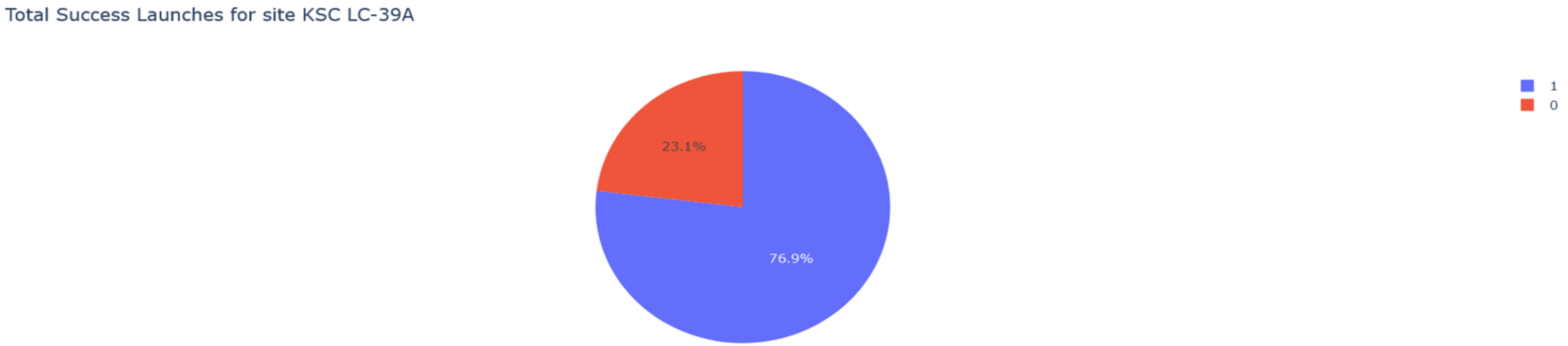

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches By Site



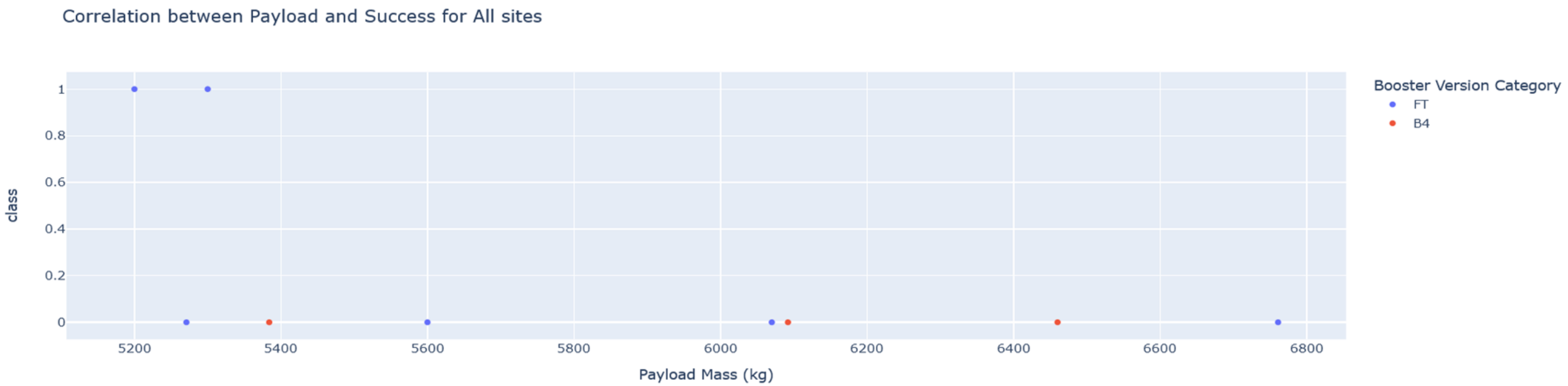
KSC LC-39A launches comprise largest portion of successful launches. But it does not necessary mean it is most successful site as it might have more launches than other sites. To find site with best success ratio we have to check each site.

Launch Profile of Most Successfull Site



On further checking we find that KSC LC-39A is indeed most successful site with success ratio of 76.9%.

Correlation between Payload and Success for All sites



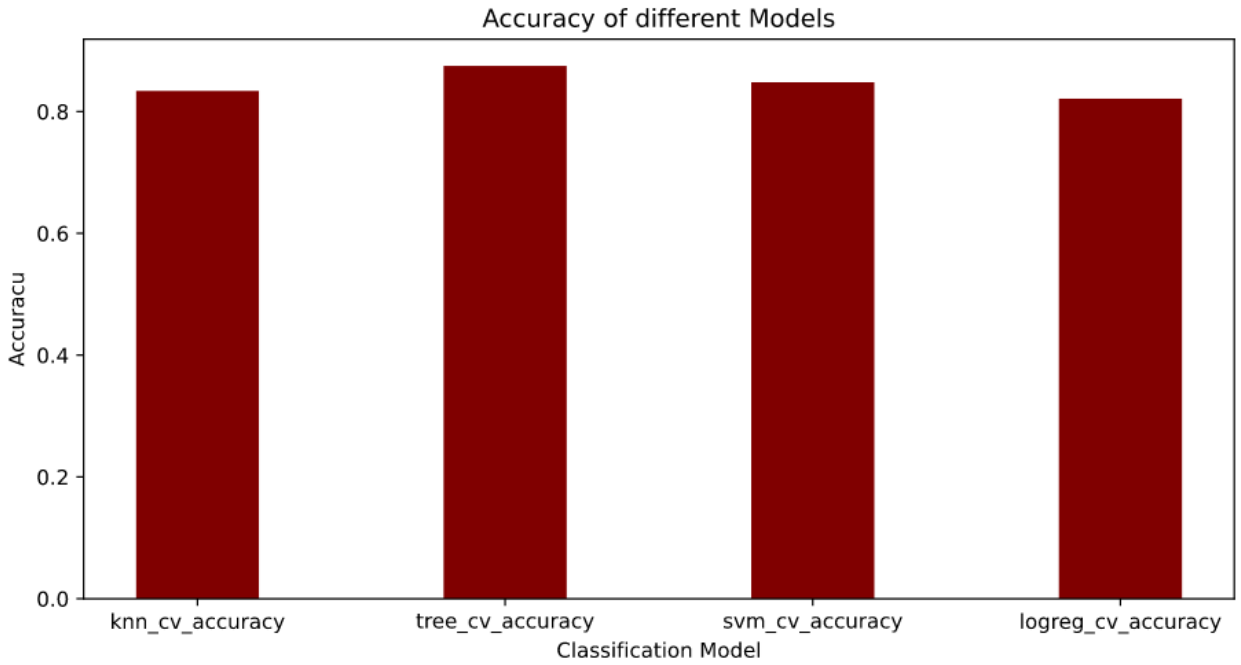
Success Ratio for Booster Versions in Payload Range (5000 - 7000 Kg). In this range there are more failed launches. Only FT booster version have successful launches.

Section 5

Predictive Analysis (Classification)

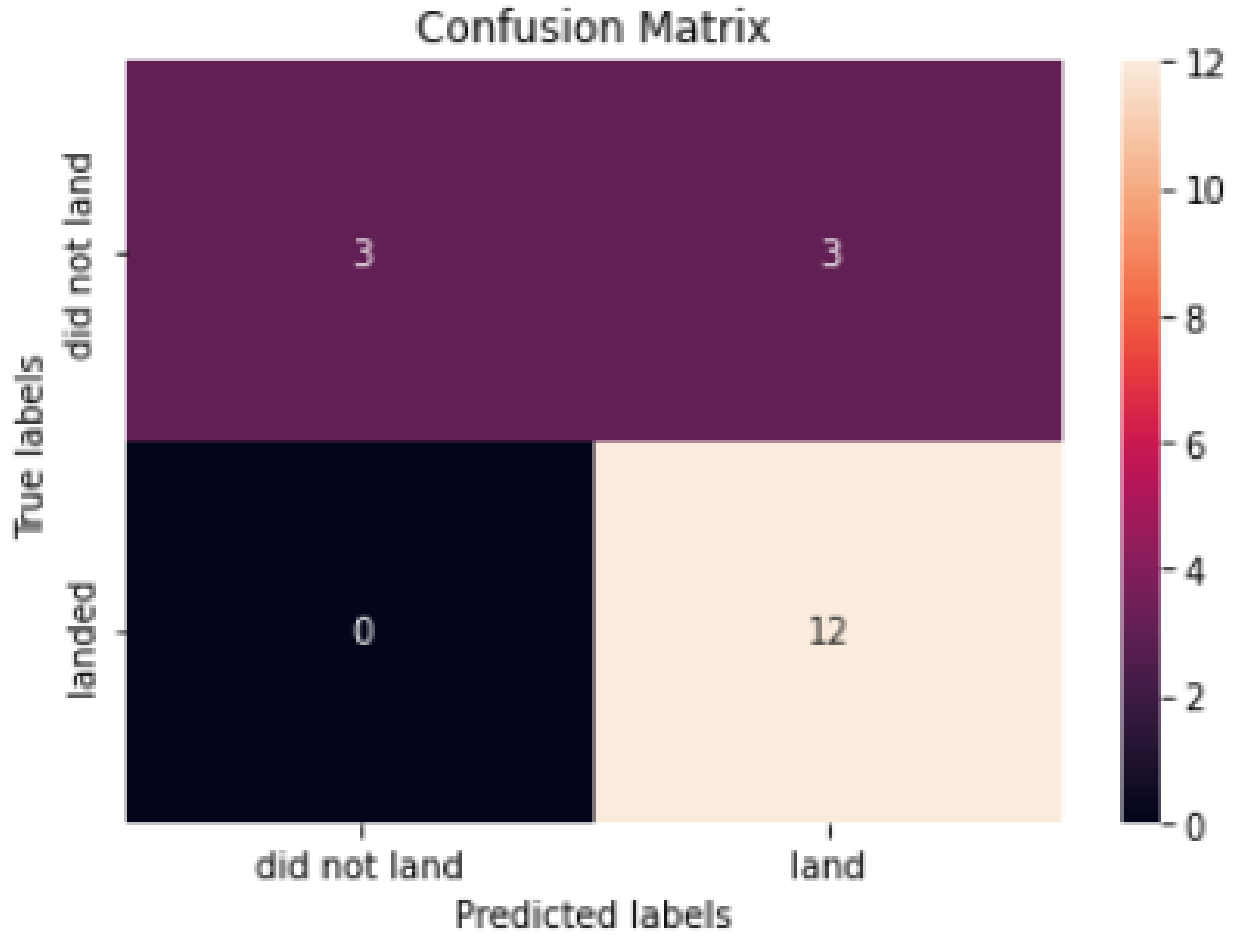
Accuracy of Different Classification Models

- Tree map have slightly better accuracy.
- Accuracy is almost same for each model.
- This is due to small amount of data available to build model.



Confusion Matrix for Tree-Map Model

- Predicted Correctly
  - Successful(TP) = 12
  - Failed(TN) = 3
- Predicted Wrong
  - Successful(FP) = 0
  - Failed(FN) = 3
- Accuracy
  - (Total True/Total)*100
  - = 83.3334%

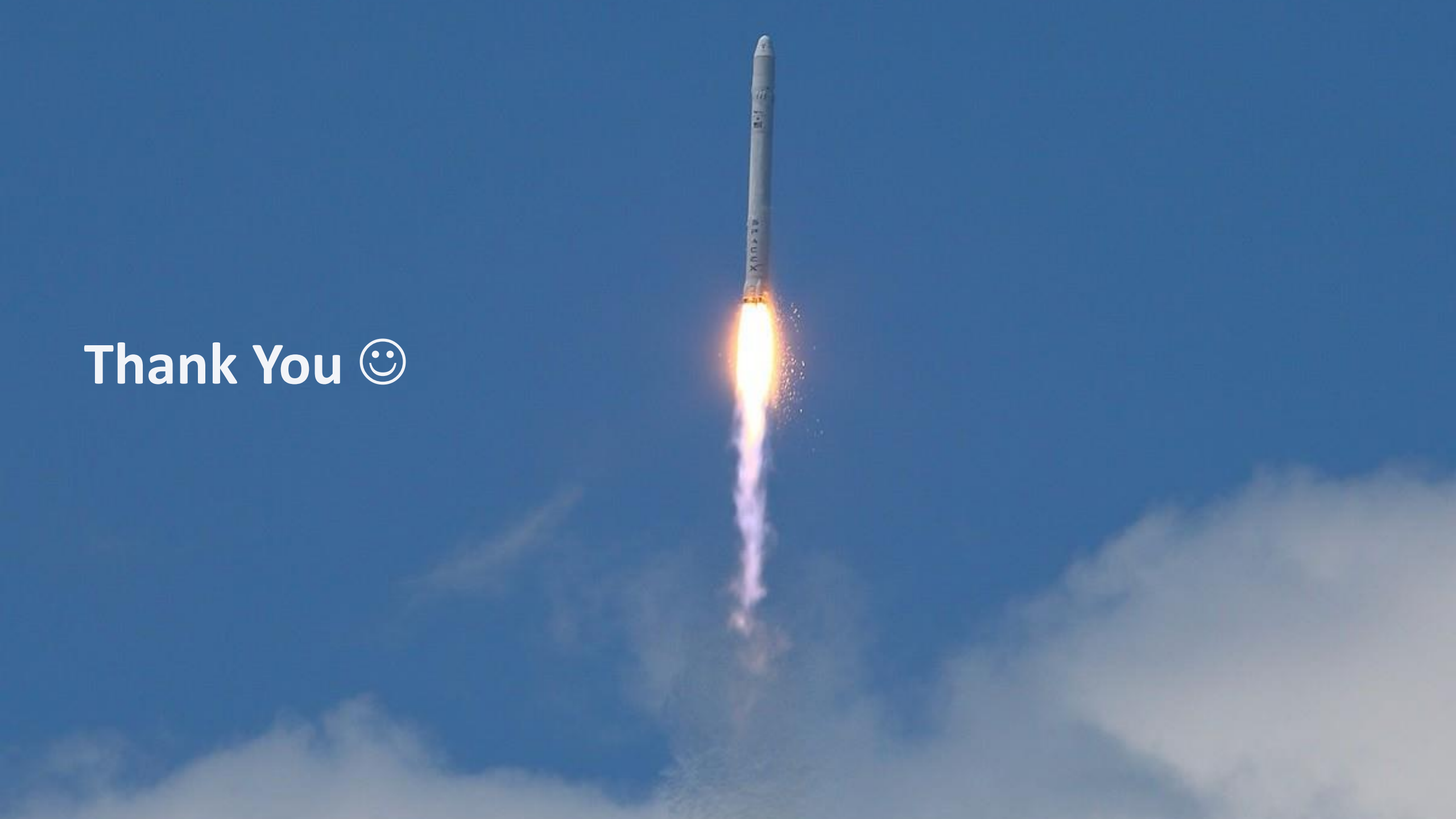


Conclusions

- SpaceX
  - Improvement in success rate
- Launch sites locations
  - Near ocean, railway, road in order
  - Farther from city
- Accuracy is almost same for all models in this case because of small dataset.
- Data Science gives deep insight into the selected field and can be used to build blueprint for decision making.

Appendix

- Python version 3.2
- Data files collected and created include :
  - dataset_part_1.csv
  - spacex_web_scraped.csv
  - dataset_part_2.csv
  - dataset_part_3.csv
  - spacex_launchdash.csv
- Master GitHub repository link


Thank You ☺