IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eric Spencer
3-27-24

Outline

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

Executive Summary

- Utilizing Python and SQL we collected a vast amount of data regarding SpaceX launches
- Used that data to make predictions
- Used data to create a dashboard app for ease of finding information

Introduction

- This project was undertaken as a final capstone to the IBM Data Science Certificate
- Interested in finding correlations of launch sites to success rates

Section 1

Methodology

Methodology

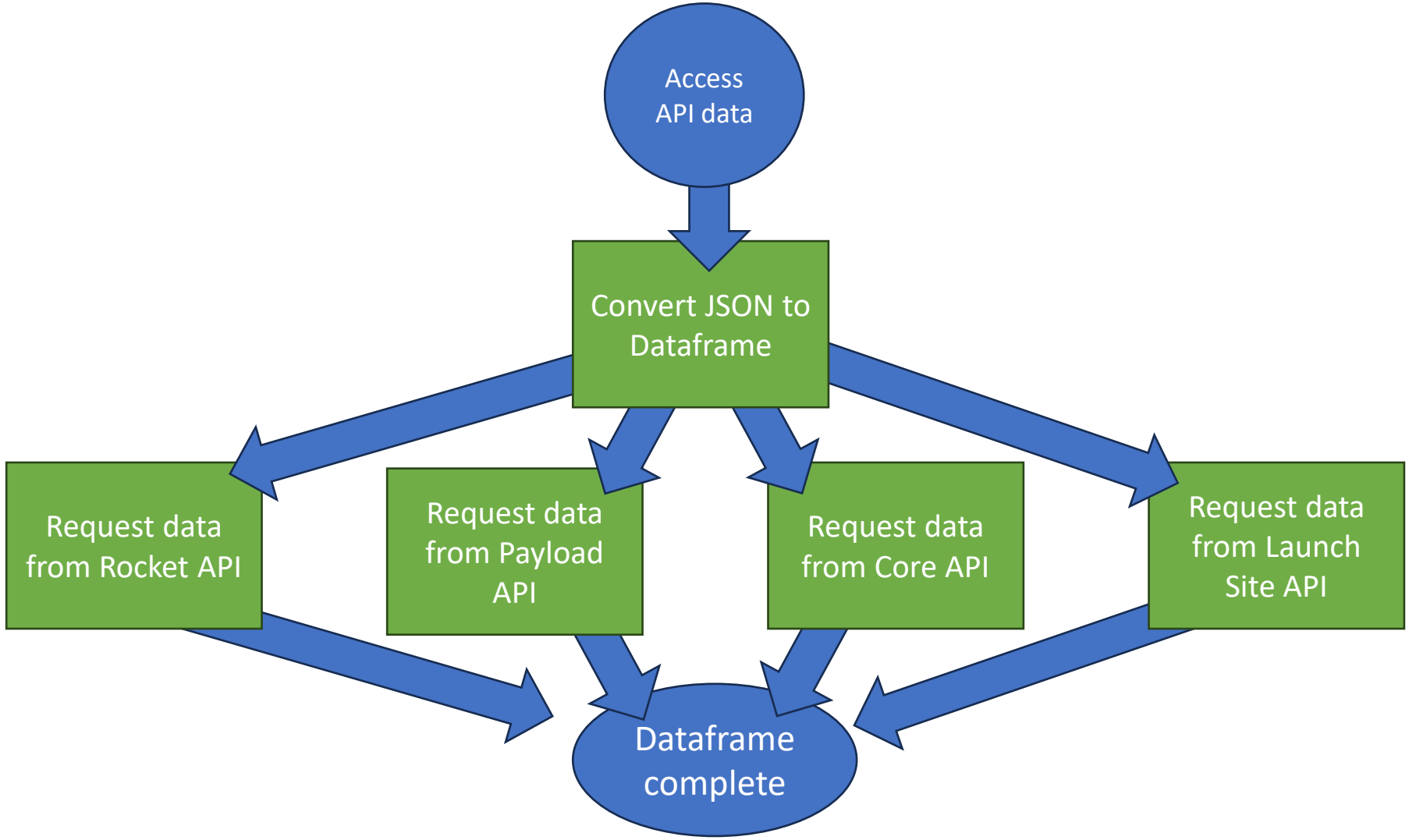
Executive Summary

- Data collection methodology:
  - Called SpaceX API for data and used python to build dataframe column by column
- Perform data wrangling
  - Used SQL to organize data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built four different classification models and tested their accuracy

Data Collection

- Utilized python to access SpaceX API
- Through the API, requested data set of past launches
- Conducted data cleanup
- Data set ready for additional analysis

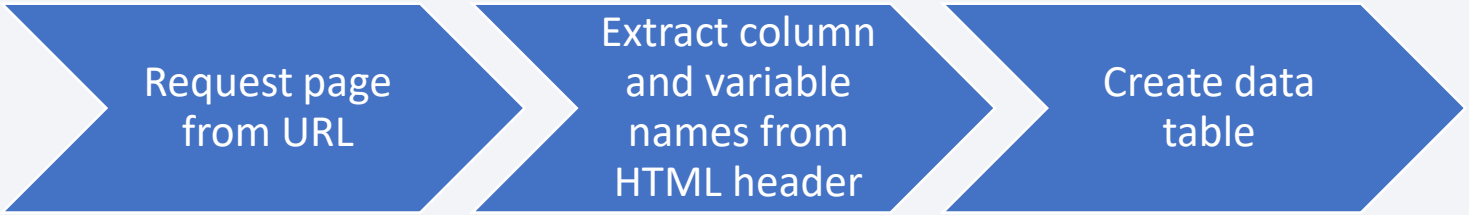
Data Collection – SpaceX API



https://github.com/spencerej/CapstoneProject/blob/70372e5052fd9e6c6dc9d5b8c2d0a3dcab503f2d/spacex-data-collection-api.ipynb

Data Collection - Scraping

- Webscraped data from Wikipedia and constructed data table column by column



- [CapstoneProject/webscraping.ipynb at main · spencerej/CapstoneProject (github.com)]

Data Wrangling

- Calculated numbers of launches and different orbit types. Bucketed bad outcomes and good outcomes separately and added that data to table.

- [CapstoneProject/Week1 labs-jupyter-spacex-Data wrangling.ipynb at main · spencerej/CapstoneProject (github.com)]

EDA with Data Visualization

- Created multiple plots comparing different characteristics of flights.
- Used scatter plots comparing Flight Number, Payload, Orbit, Launch Site, etc. and visualized success of flights with different color plot points.
- Visualized flight success rate using bar chart comparing different Launch Sites. This easily shows the most successful sites vs. the least successful
- Changed data parameter to only show year and created a line chart to track the success rate trend over the years. Chart clearly shows that over time successful flights are more common.
- https://github.com/spencerej/CapstoneProject/blob/main/Week2%20-%20edadataviz.ipynb

EDA with SQL

- Utilizing SQL, performed queries to understand the generalizations of the data set.
- Explored total payload mass of launches performed by Nasa
- Explored average mass of launches by booster type
- Explored which boosters saw successful flights limited by payload mass
- https://github.com/spencerej/CapstoneProject/blob/main/eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

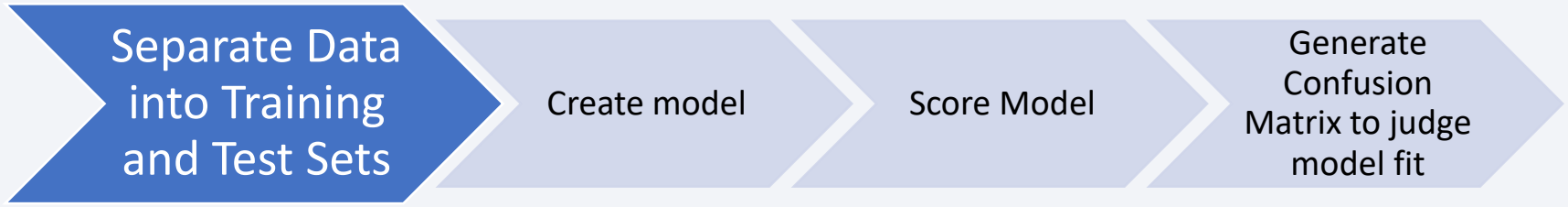
- Using Folium map created circles to designate launch sites, markers to track success/failure, and lines to track distances.
- Markers made tracking success/failure easily because color could be used. Red for failure, Green for success. Lines created made it easy to show distances between sites and other landmarks.
- https://github.com/spencerej/CapstoneProject/blob/main/week3-launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- In dashboard creation used pie charts and line charts to visualize launch data.
- Pie charts can at a glance show common generalizations of the data.
- Line charts shows trends over time.
- https://github.com/spencerej/CapstoneProject/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

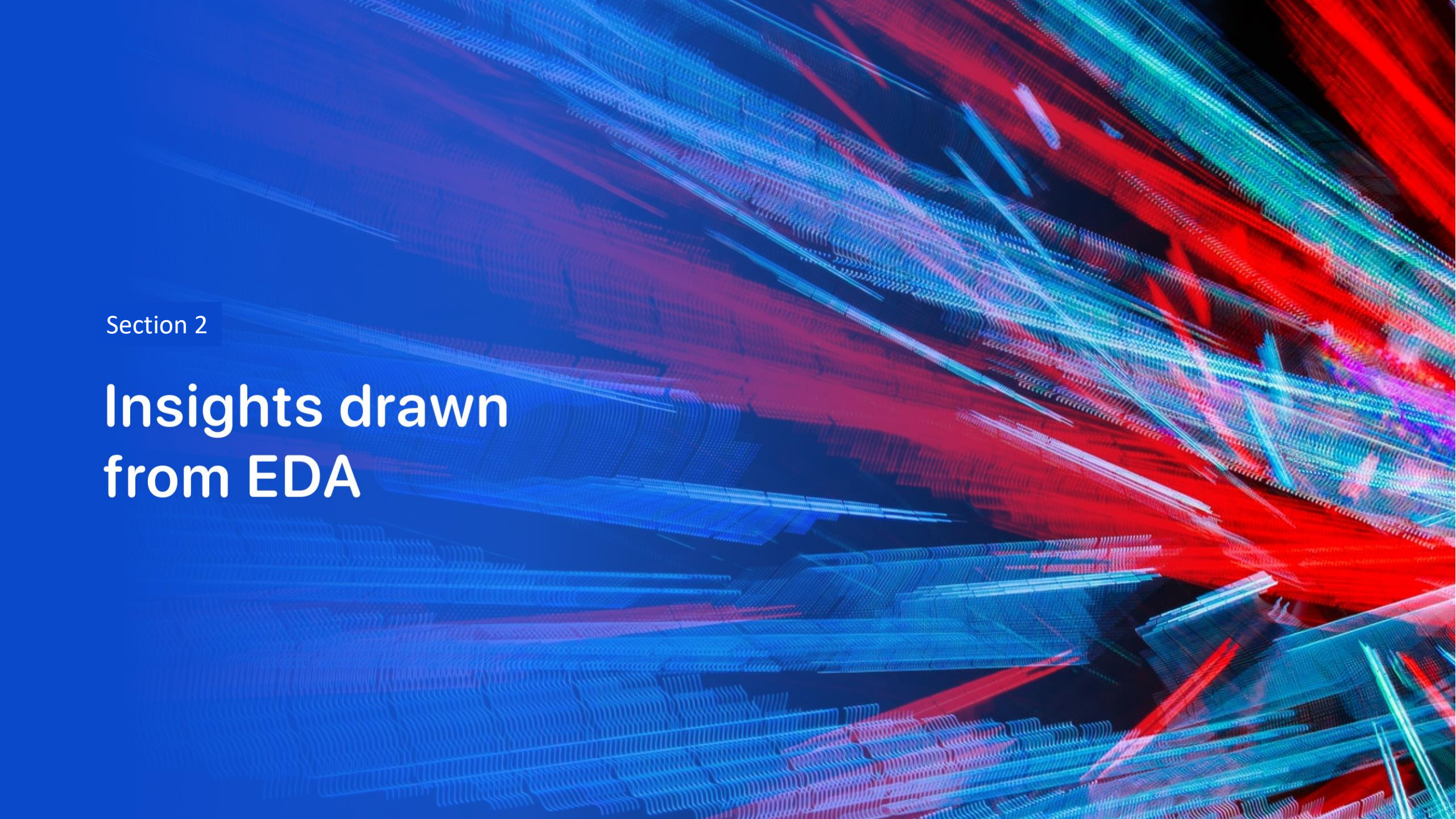
- Created 4 separate models for classification and prediction (Log Regression, KNN, Decision Tree, Support Vector)
- For each model type executed process as below:



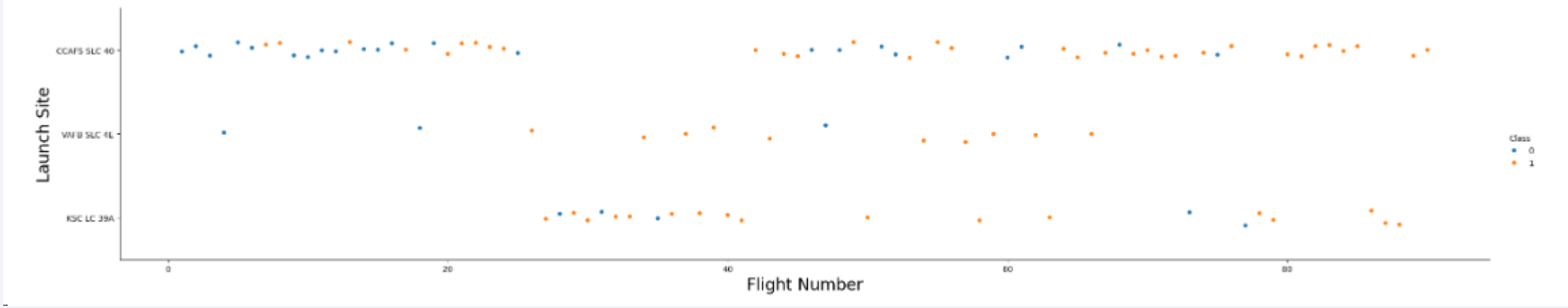
- https://github.com/spencerej/CapstoneProject/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Success rate of flight seems to be related to Launch Site
- As more launches are performed, success rate increases
- Models generally perform well and are able to predict future flight success rates

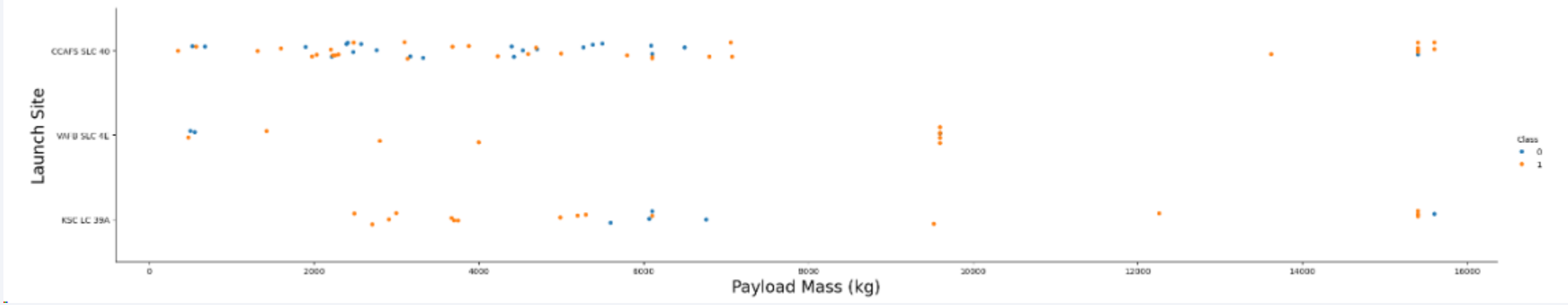

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- With more flights comes better success
- CCAPS SLC 40 is the primary launch site

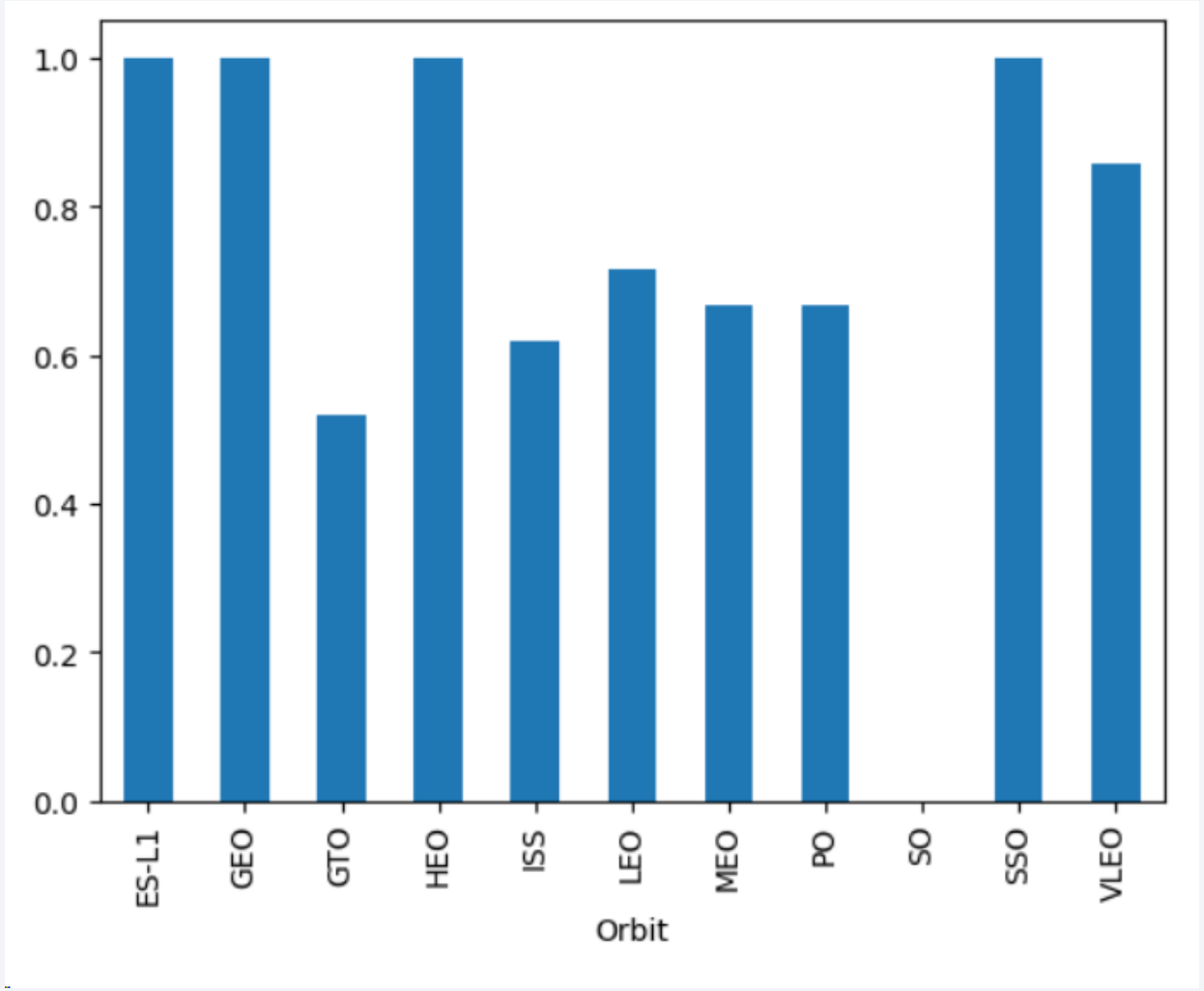
Payload vs. Launch Site



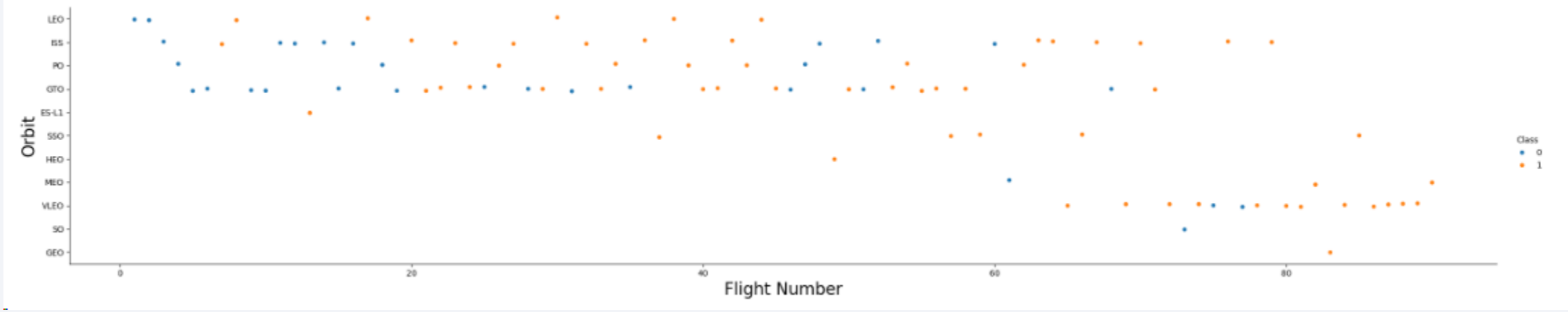
- VAFB SLC launch site does not launch heavy rockets (over 10k kg

Success Rate vs. Orbit Type

- SO Orbit has not had a successful flight
- ES-L1, GEO, HEO, SSO Orbits have perfect success rates for flights

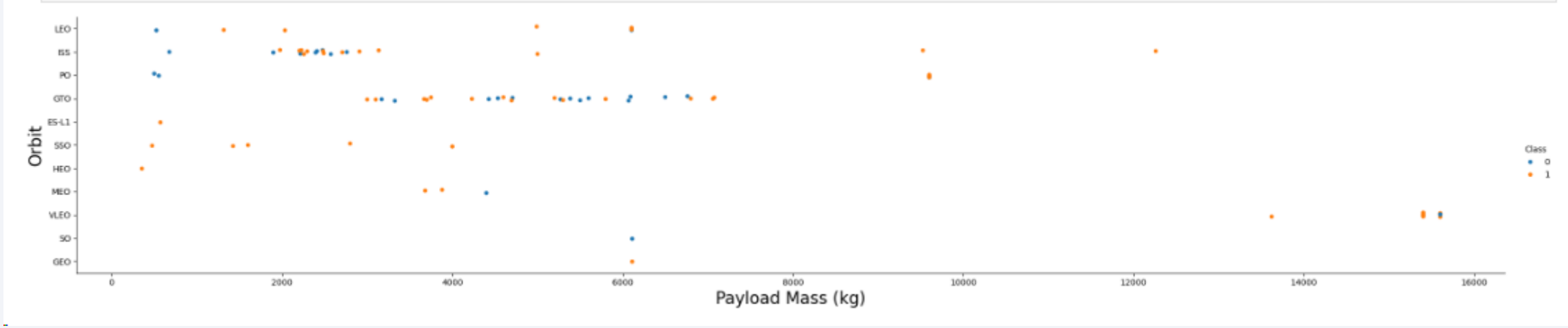


Flight Number vs. Orbit Type



- As time goes on Orbit types move from LEO, ISS, PO, GFO to more VLEO
- Only 1 GEO orbit attempt

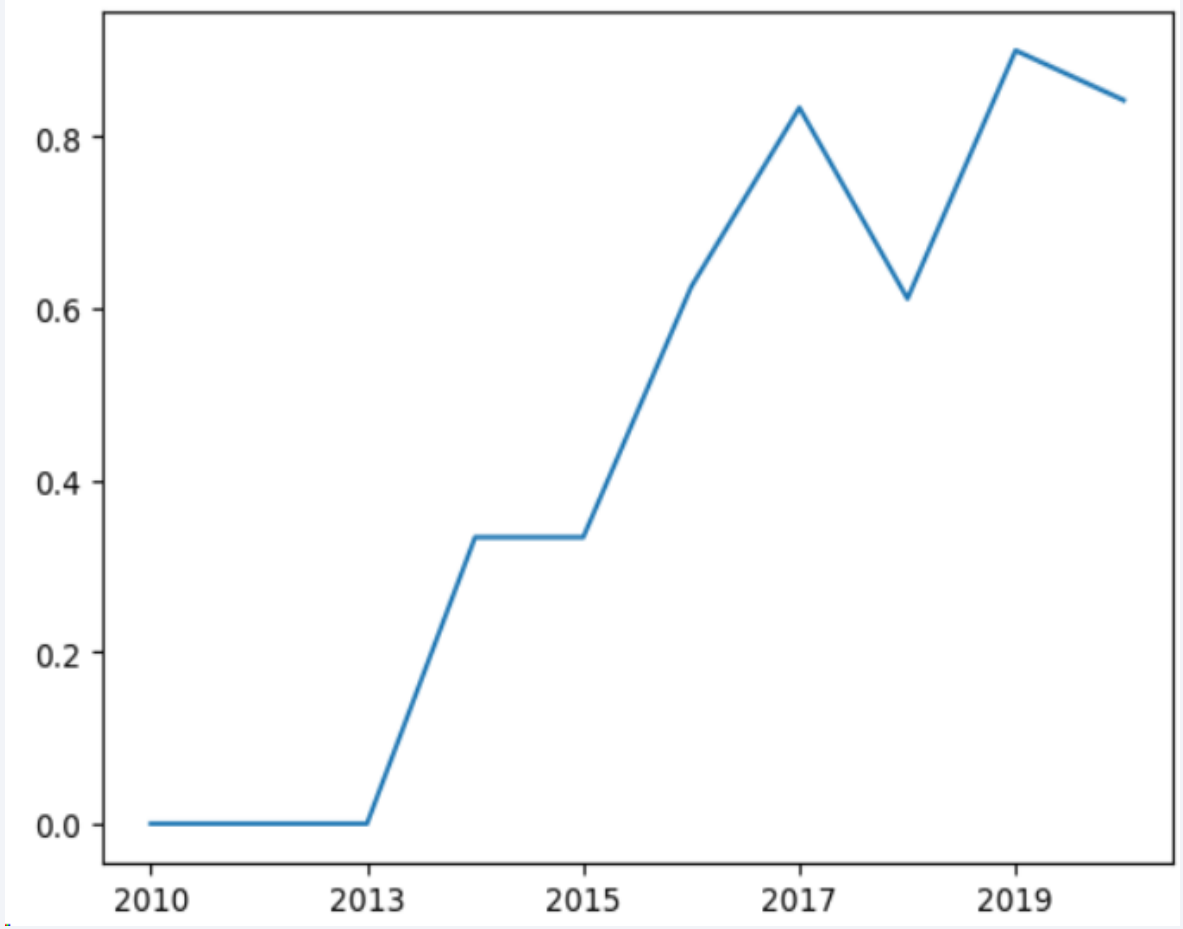
Payload vs. Orbit Type



- Heavy payloads have higher success rates in Polar, LEO, and ISS orbits
- GTO orbit inconclusive as success rate is mediocre

Launch Success Yearly Trend

- Since 2010, Launch Success steadily increases over time.



All Launch Site Names

- Utilizing SQL able to obtain each unique launch site

In [113…

```
#%sql select * from spacextbl where Launch_Site like 'CCA%' limit 5;
%sql select distinct(Launch_Site) from spacextbl
```

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

Out[113…

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

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
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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Manipulating data to get a feel for the data set

Total Payload Mass

```
In [14]: %sql select sum(PAYLOAD_MASS__KG_) from spacextbl where Customer like 'NASA (COTS)%';
 * sqlite:///my_data1.db
Done.
Out[14]:
```

sum(PAYLOAD_MASS__KG_)
525

- Total Payload in KG for NASA is 525

Average Payload Mass by F9 v1.1

```
In [18]: %sql select avg(PAYLOAD_MASS__KG_) from spacextbl where Booster_Version like 'F9 v1.1%'
 * sqlite:///my_data1.db
Done.
Out[18]: avg(PAYLOAD_MASS__KG_)
         2534.6666666666665
```

- Average mass carried by F9 v1.1 rocket is ~2535 KG

First Successful Ground Landing Date

```
In [23]: %sql select min(Date) from spacextbl where Landing_Outcome like 'Success (ground pad)';
 * sqlite:///my_data1.db
Done.
Out[23]: min(Date)
         2015-12-22
```

- First successful ground landing date is 12-22-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

In [35]:
```
%sql select distinct(Booster_Version) from spacextbl where Landing_Outcome like 'Success (drone ship)' and PAYLOAD_MASS__I
```

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

Out[35]:

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

- The above rockets had successful drone ship landings. These payloads were between 4000 and 6000.

Total Number of Successful and Failure Mission Outcomes

In [46]:
```
%sql select Mission_Outcome, Count(*) from spacextbl group by Mission_Outcome order by 2 desc;
```

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

Out[46]:

Mission_Outcome	Count(*)
Success	98
Success (payload status unclear)	1
Success	1
Failure (in flight)	1

- Mission success rate is 99%

Boosters Carried Maximum Payload

```
In [76]: %sql select Booster_Version from spacextbl where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextbl);
 * sqlite:///my_data1.db
Done.
```

Out[76]:

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

- The above boosters are able to carry max payloads.

2015 Launch Records

In [101…

```
%sql select substr(Date, 6, 2) as month, Booster_Version, Launch_Site, Landing_Outcome from spacextbl where Landing_Outcor
```

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

Out[101…

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

- Above shows the failures for drone ship. They were carried by B1015 and B1012 boosters at the CCAFS LC-40 launch site

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

```
In [112…
%sql select Landing_Outcome, count(*) as Countof from spacextbl where Date between '2010-06-04' and '2017-03-20' group by
 * sqlite:///my_data1.db
Done.
```

Out[112…

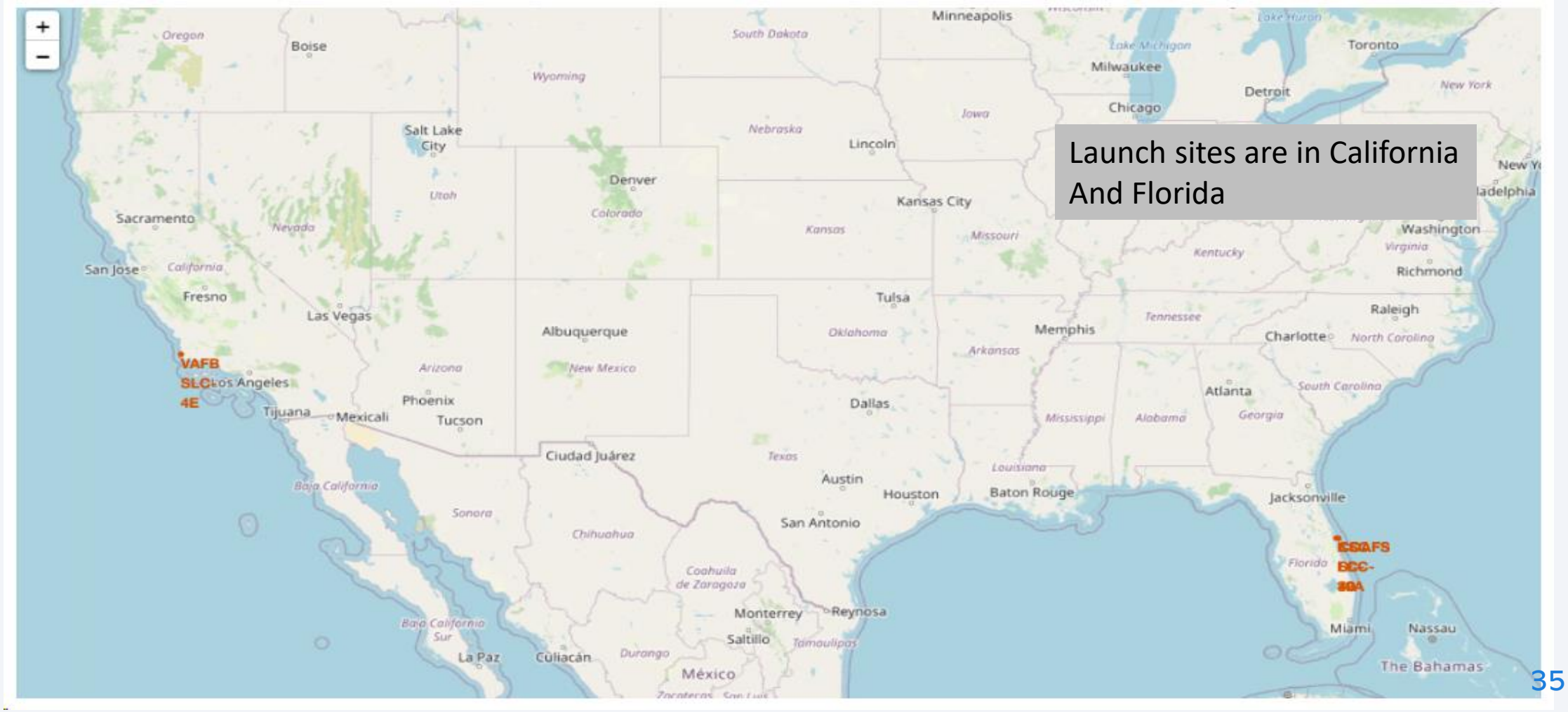
Landing_Outcome	Countof
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- There were 31 launches between 6-4-2010 and 3-20-2017
- Above are the outcomes of those launches

Section 3

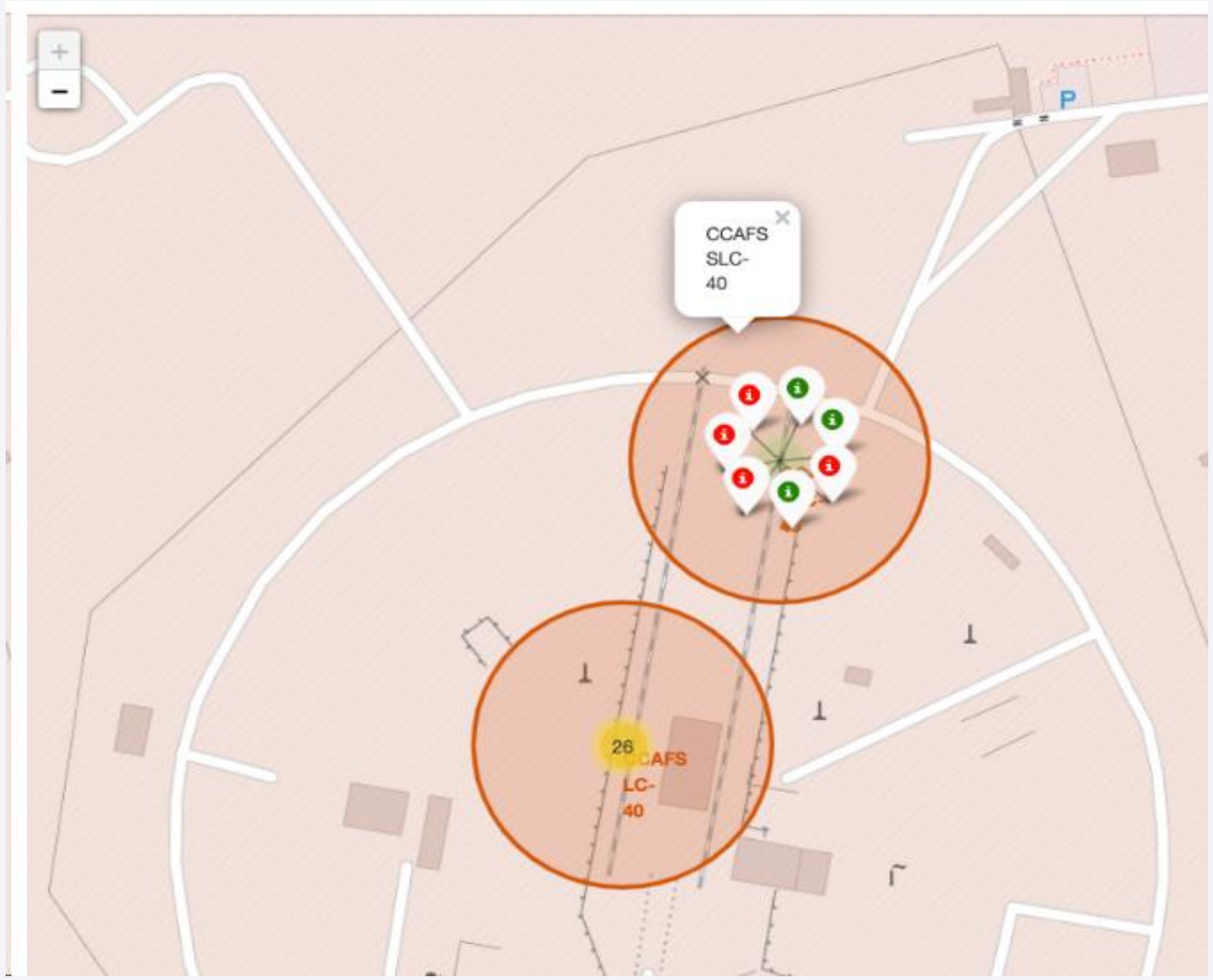
Launch Sites Proximities Analysis

Launch Sites on the Map

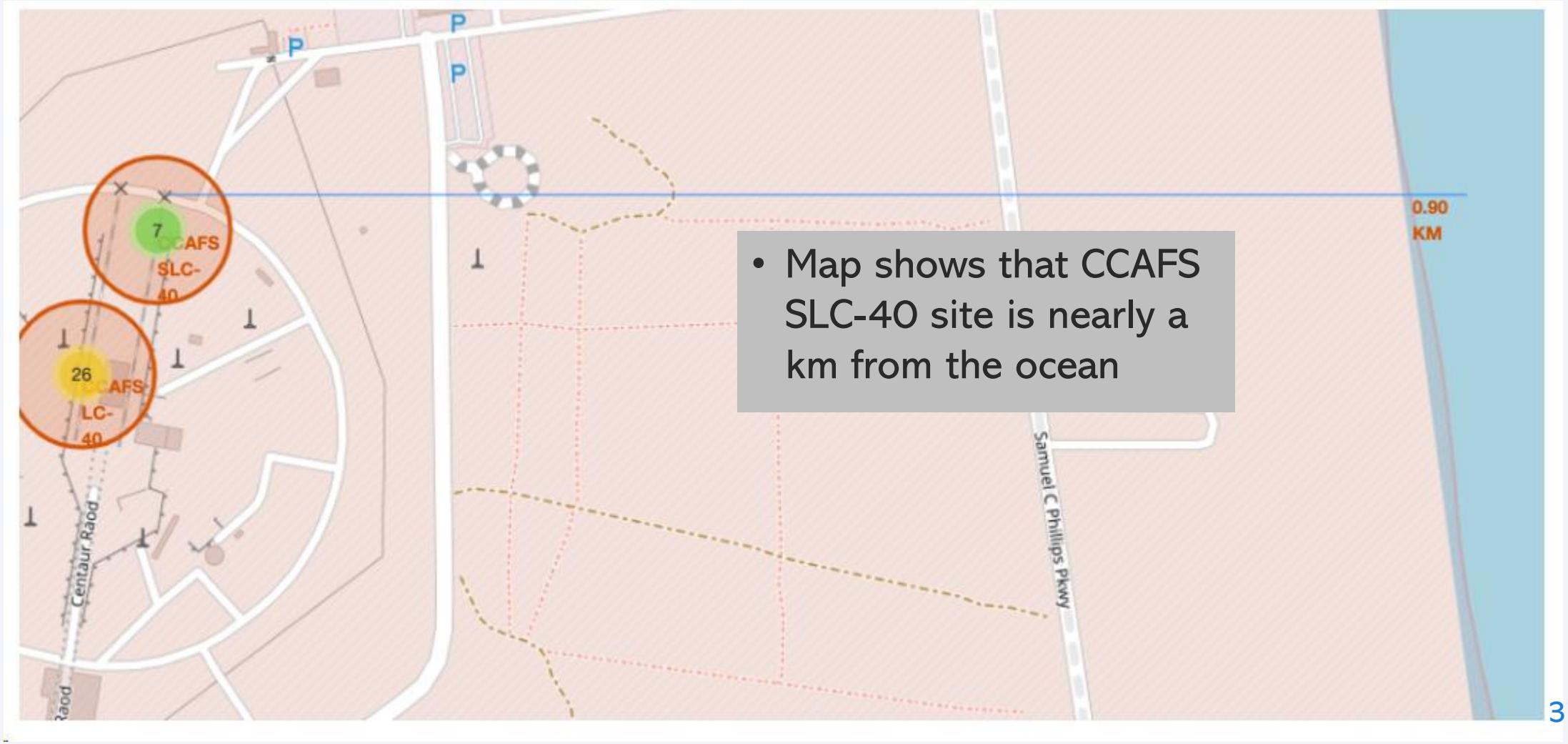


Tracking Launch Outcomes

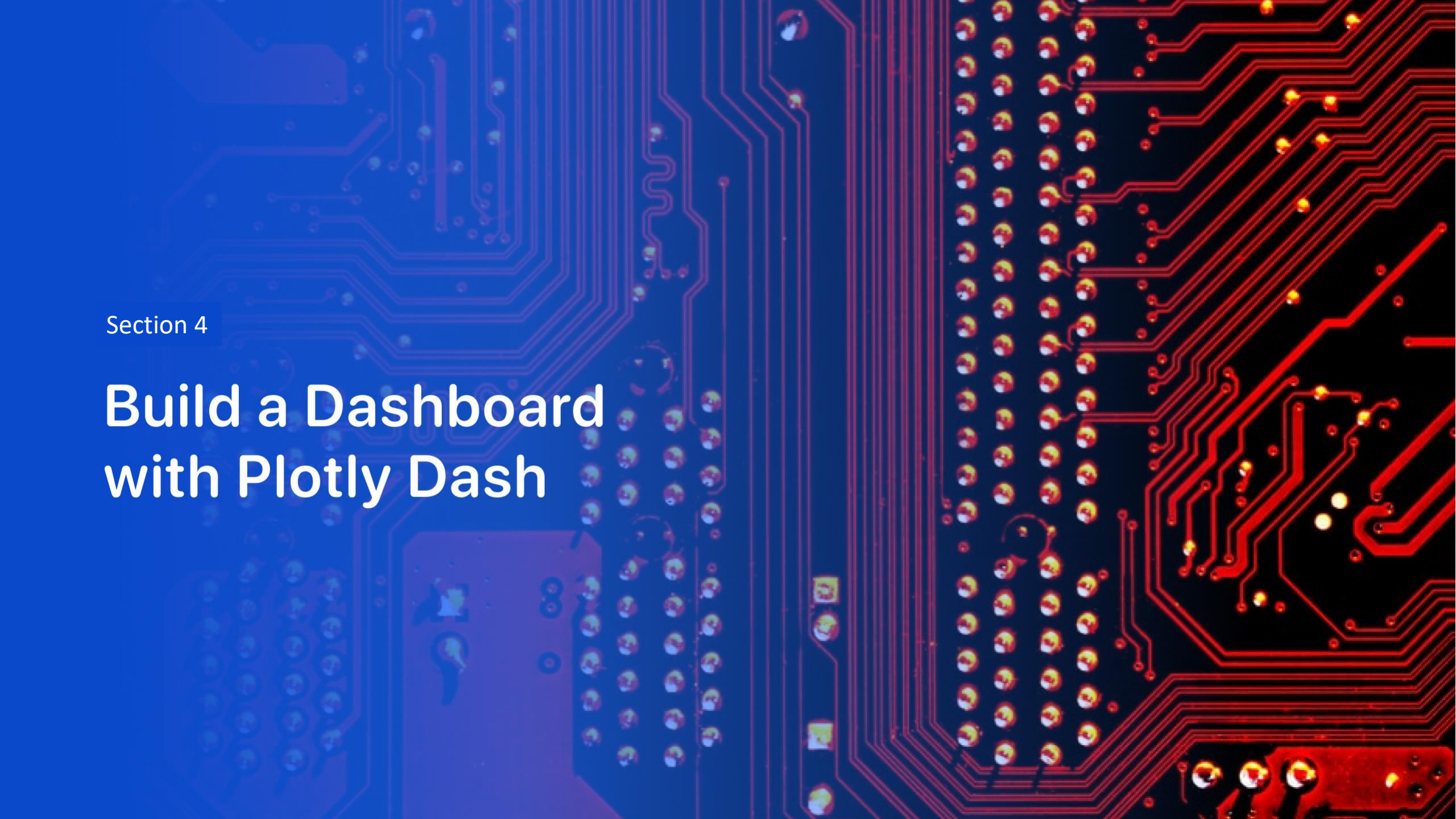
- Utilizing Folium we are able to easily show successful or failed launches by site
- CCAFS LC-40 had 26 other launches



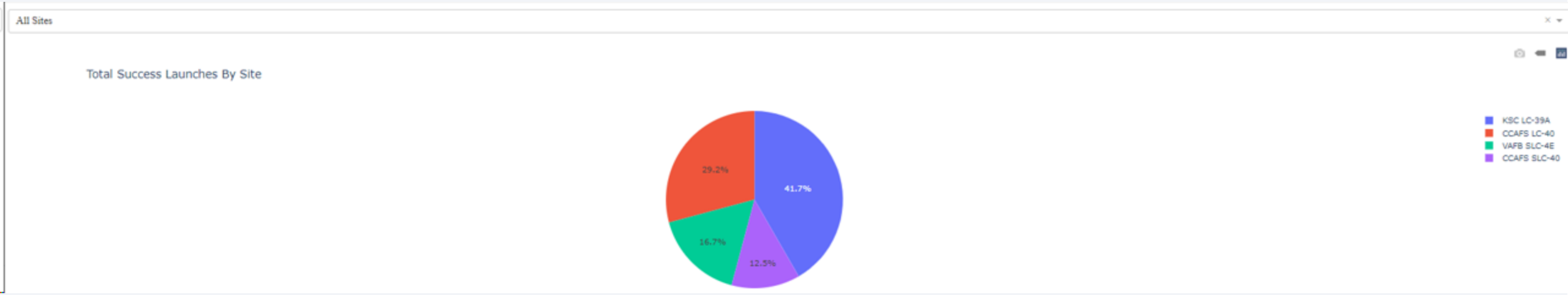
Proximity to Landmarks



- Map shows that CCAFS SLC-40 site is nearly a km from the ocean

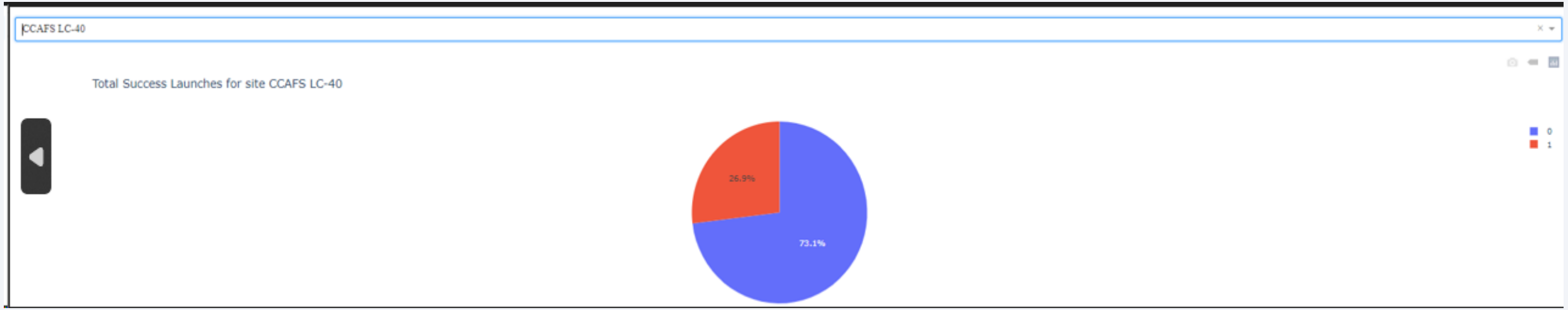

Section 4
Build a Dashboard with Plotly Dash

Dashboard Pie Chart



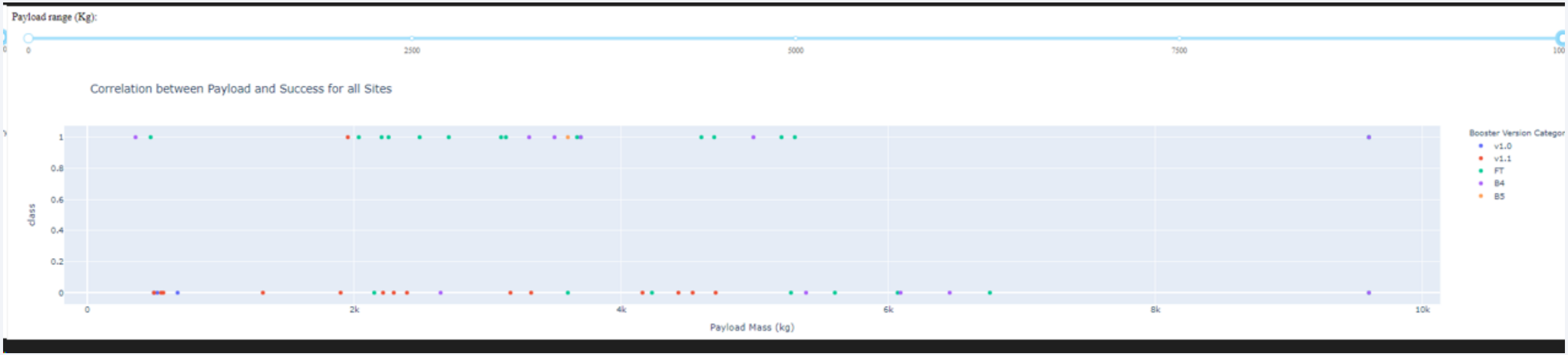
- Successful launches by site

Most Successful Site



- CCAFS LC-40 is the most successful site with a success rate of 73%

More info about success rates



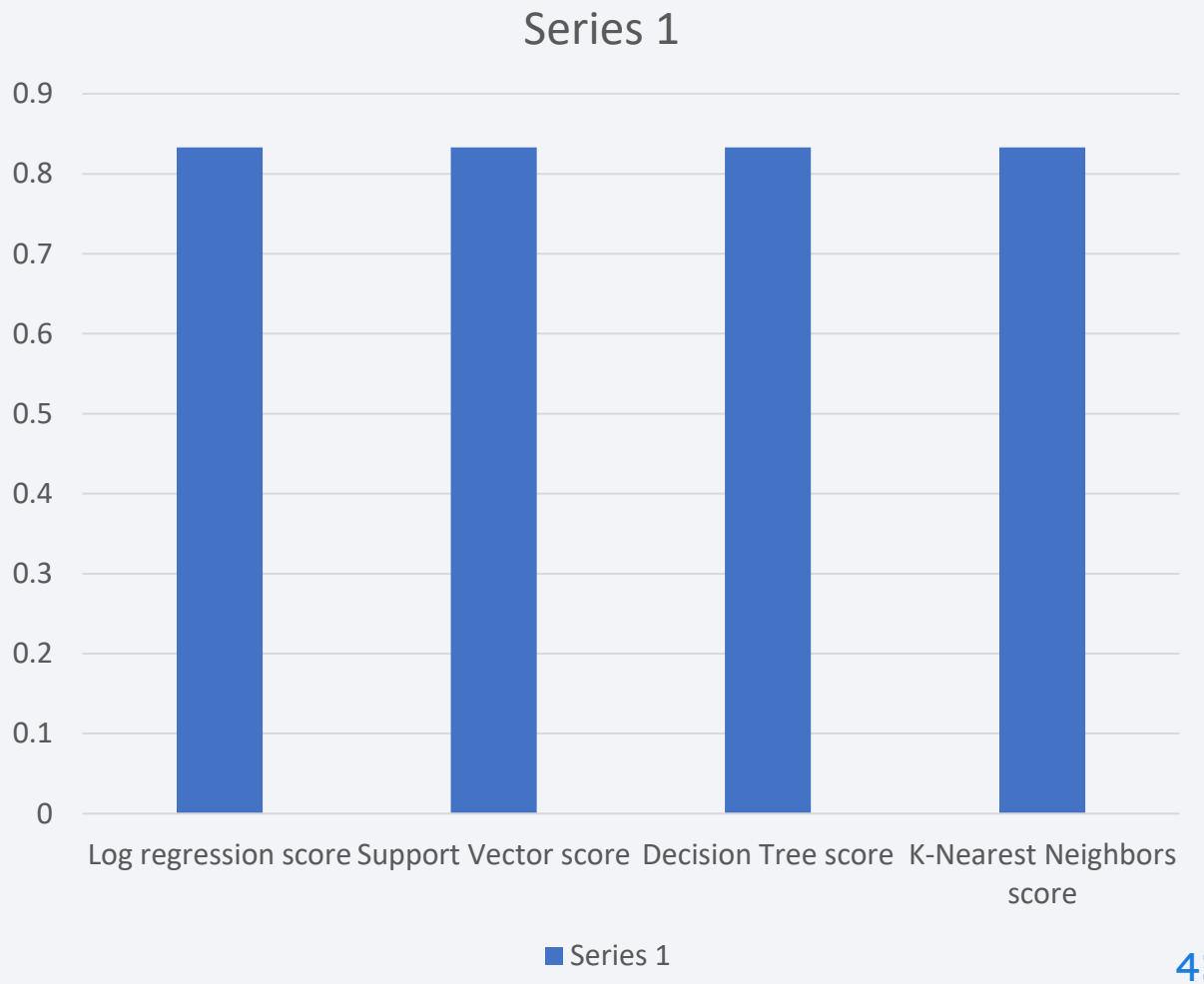
- Success rates by site and payload

Section 5

Predictive Analysis (Classification)

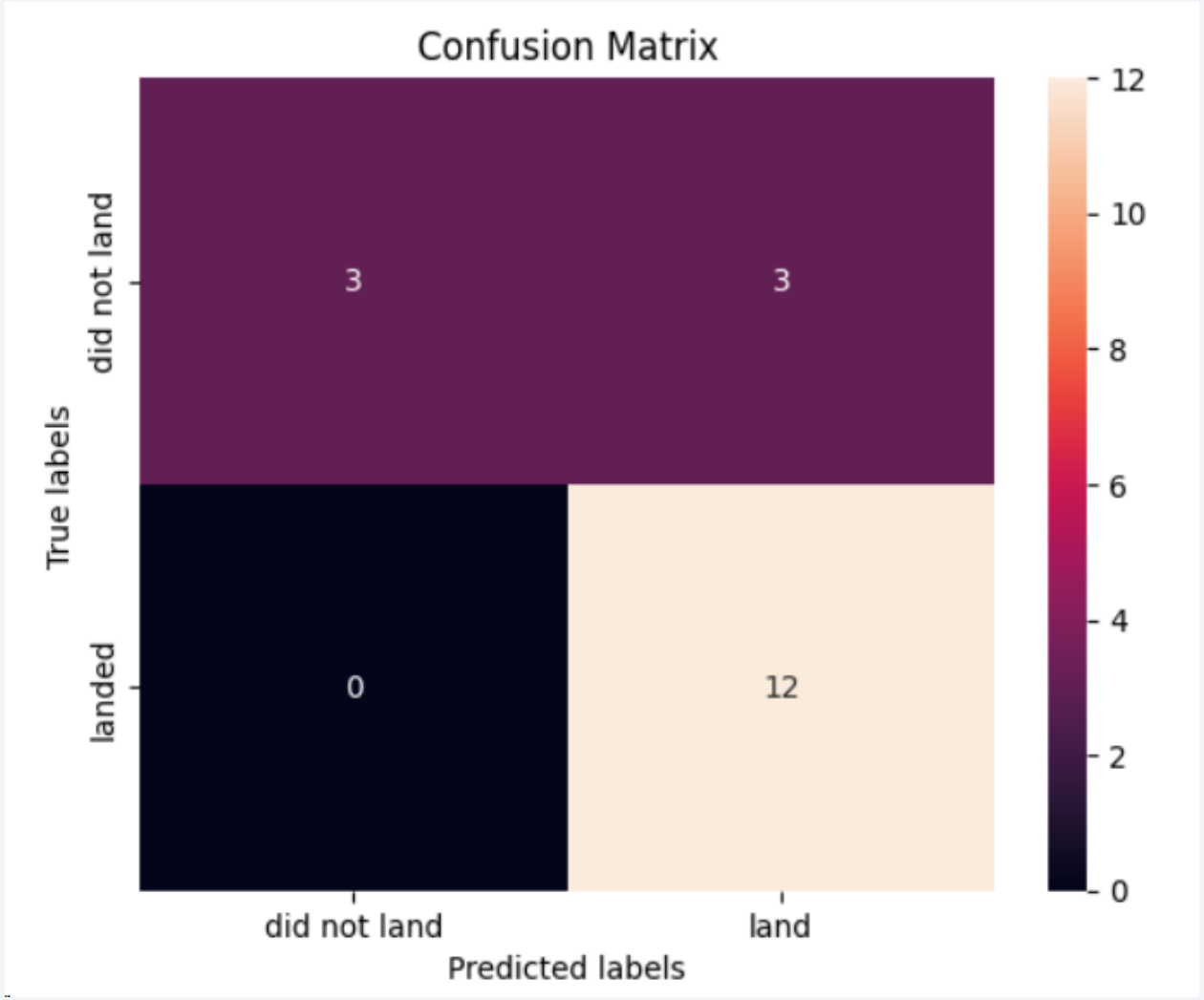
Classification Accuracy

- All models were equally accurate in prediction



Confusion Matrix

- Confusion matrix shows accuracy of predictions
- All models had nearly identical accuracies in predicting landing or not landing



Conclusions

- There is a great amount of correlating data that shows that SpaceX is improving launch successes with time.
- We can use dashboard app to quickly gain information about flight data
- Data Analysis shows that some orbits are more successful than others

Appendix

```
# Select relevant sub-columns: `Launch Site`, `Lat(Latitude)`, `Long(Longitude)`, `class`
spacex_df = spacex_df[['Launch Site', 'Lat', 'Long', 'class']]
launch_sites_df = spacex_df.groupby(['Launch Site'], as_index=False).first()
launch_sites_df = launch_sites_df[['Launch Site', 'Lat', 'Long']]
launch_sites_df
```

```
# Initial the map
site_map = folium.Map(location=nasa_coordinate, zoom_start=5)
# For each launch site, add a Circle object based on its coordinate (Lat, Long) values. In addition, add Launch site name
```

More code snippets

Thank you!