IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shawn Cormier
September 7, 2024

Outline

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

Executive Summary

- Methodologies
  - Data was accessed and collected from SpaceX dataset
  - Data was cleaned and formatted for model accuracy and accuracy of the visual data presented
- Results
  - Success rates improve over time
  - Several machine learning models were used and all yielded similar accuracy
  - Analysis determined that several boosters had a high success rate when combined with varying ranges of payload mass

Introduction

- SpaceX has gained worldwide attention for a series of historic milestones.
- It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars wheras other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

- Object is to determine the success rate of the launches with their respective payloads which then it would be possible to draw insights on what would be a reasonable cost to per launch as well as predicting the best rate of success to select which would be the optimal rocket to use.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected firstly from SpaceX using the REST api
  - Additional supporting data was also collected by webscraping Wikipedia
- Perform data wrangling
  - Data was converted from it's collected JSON format to a Pandas dataframe for analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Hyperparameters were searched using GridSearchCV
  - Training data was then applied in the next step

Methodology

- Parameters were validated using the score() method
- Confusion matrix is used to visualize the results

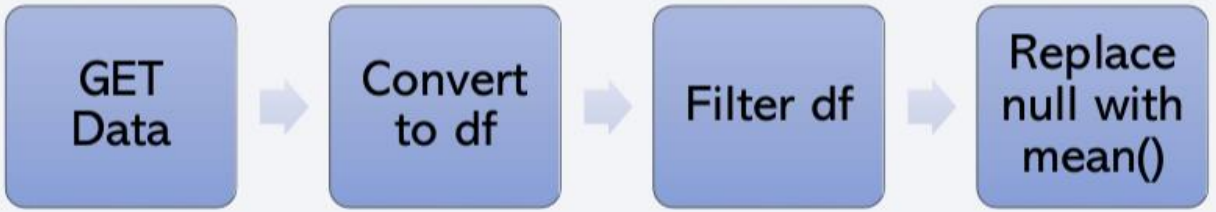
Data Collection

- Describe how data sets were collected. in the next step
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

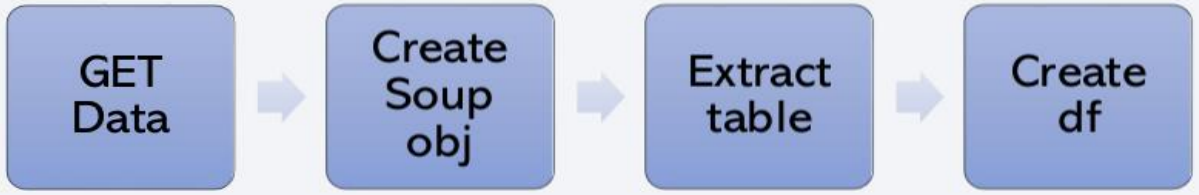
- Flow chart of SpaceX API data collection

- https://github.com/scormier113/spacex-analysis/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Flowchart of webscrape using BeautifulSoup
- https://github.com/scormier113/spacex-analysis/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Data Wrangling was used to clean the source data and organize the data to determine success and failures
- https://github.com/scormier113/spacex-analysis/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- The following charts were used to determine success and failure rates of the varying types of launches
  - Scatter plot of flight number vs. Payload mass
  - Scatter plot of flight number vs. Launch site
  - Scatter plot of payload mass vs. Launch site
  - Bar chart of the relationship between orbit type and success rate
  - Scatter plot of the relationship between flight number and orbit type
  - Scatter plot of the relationship between payload and orbit type
  - Line chart of the yearly trend of launch successes
- https://github.com/scormier113/spacex-analysis/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- SQL queries performed:
  - Distinct Launch Sites
  - Launch sites that begin with the string of 'CCA'
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 1.1
  - Date of the first successful landing outcome for ground pad
  - Names of boosters that have success with landing drone ship and payload mass between 4000kg and 6000kg
  - Total number of successful and failure mission outcomes
  - List names of boosters that carried the maximum payload mass
  - List records in 2015 where failure outcome was landing drone ship
  - Rank landing outcomes count between the dates of 2016-06-04 and 2017-03-20
- https://github.com/scormier113/spacex-analysis/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium Map objects used:
  - Circles: To find the sites easier
  - Map labels: List names of Launch sites to visually show the names
  - Popup labels: Add interacivity to the map
  - Launch markers: Visually display the launches
  - Lines between launch site and other map elements: Measure distances from launch site between other map objects
- https://github.com/scormier113/spacex-analysis/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Items added to Plotly Dashboard:
  - Drop down menu to act as a slicer for launch site locations
  - Pie chart to show the difference between Successful and failure launch outcomes
  - Slider bar to act as a slicer for payload mass values
  - Scatter plot to show payload launch site success at different payload masses as well as their corresponding booster version

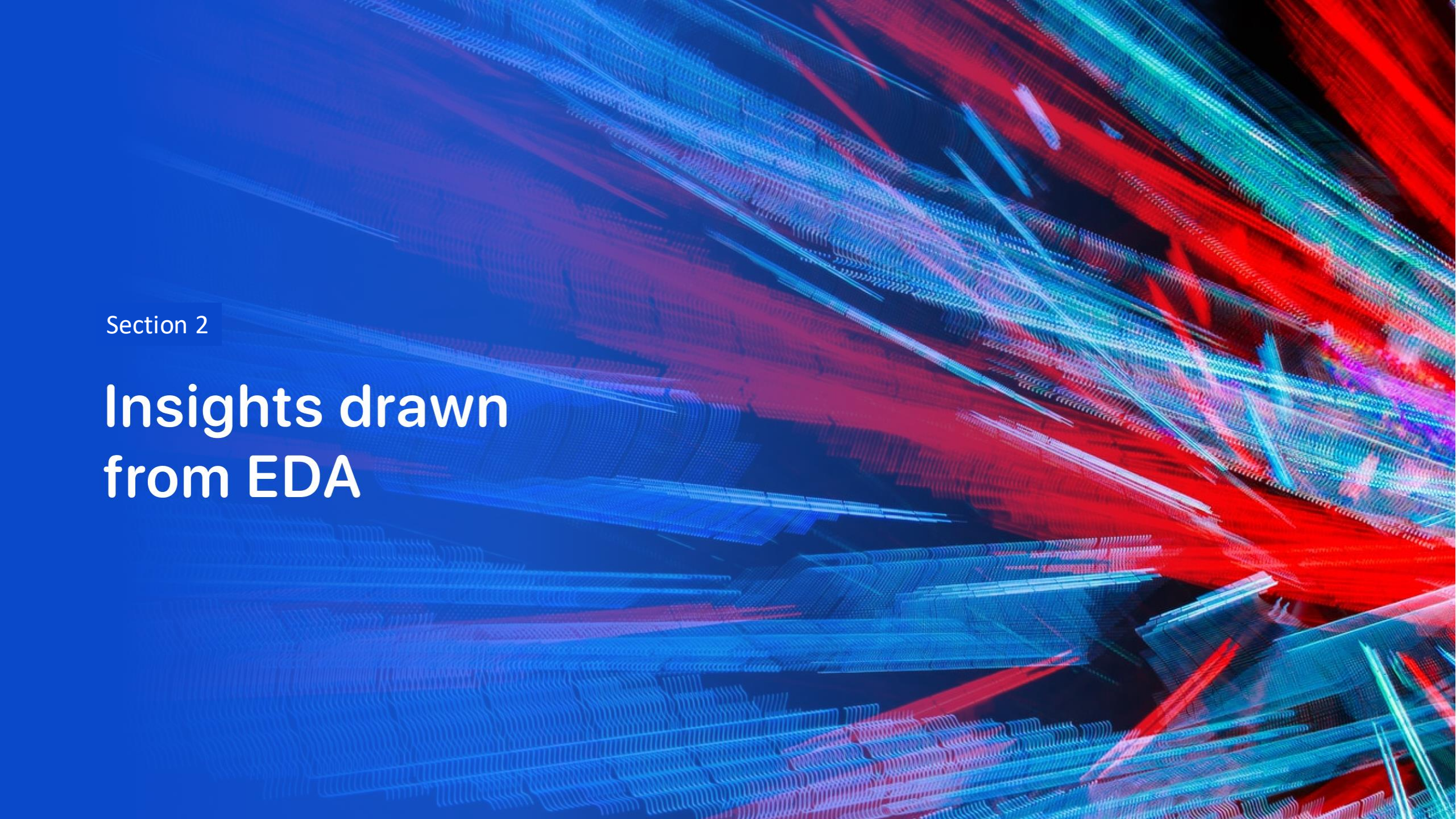
- https://github.com/scormier113/spacex-analysis/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

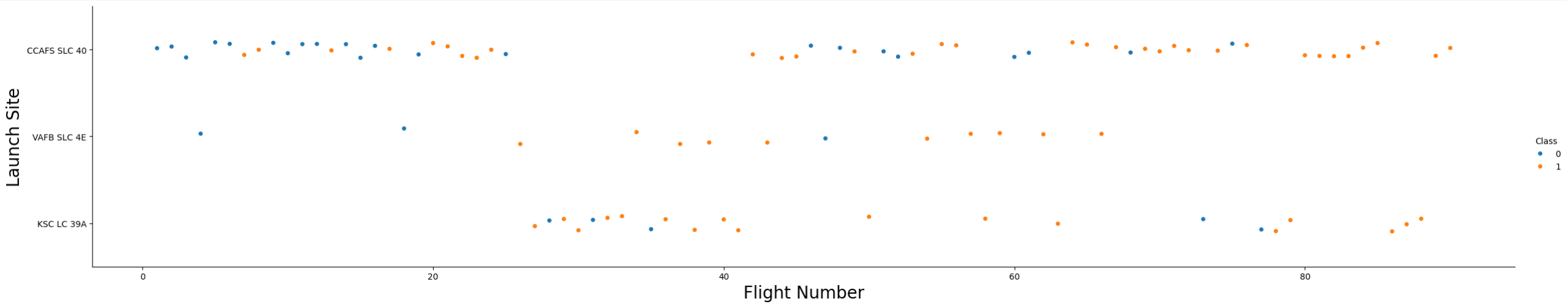
- Classification model was built using the following algorithms:
  - Logistic Regression
  - SVM
  - Decision Tree
  - KNN
- https://github.com/scormier113/spacex-analysis/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results:
  - Results of data analysis are in the following slides
- Predictive analysis results:
  - Models tested were evaluated with a result of 83% across all the models
  - 18 launch sites were in the testing set with 12 successful predictions of successful launches while 3 failure launches were accurately predicted. Of the results there were 3 false positives

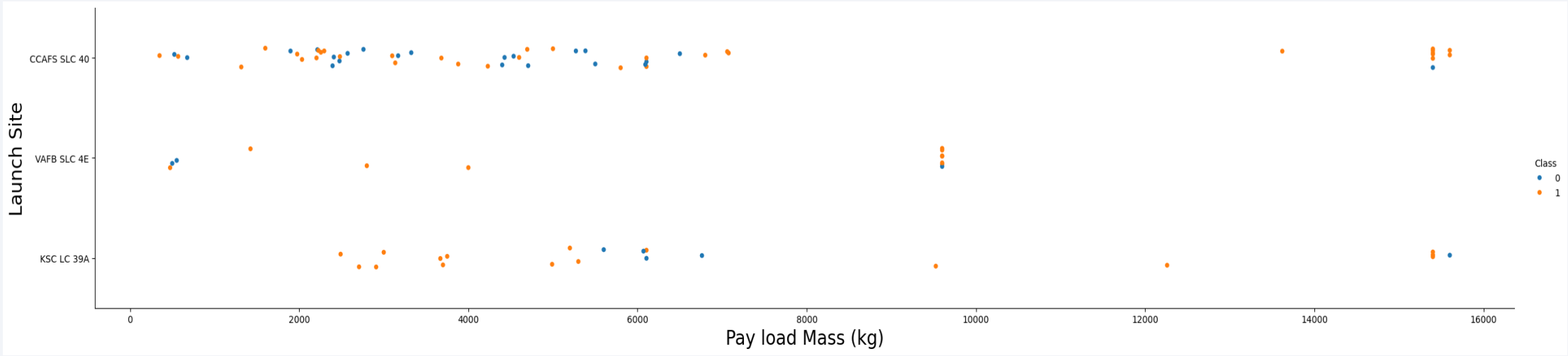

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- The majority of the first launches were from the site CCAFS SLC-40
- VAFB SLC-4E has no launches after Flight Number 66 and launches are more spread apart then the other sites
- From flight number 60 and on, the majority of the launches were from CCAFS SLC-40

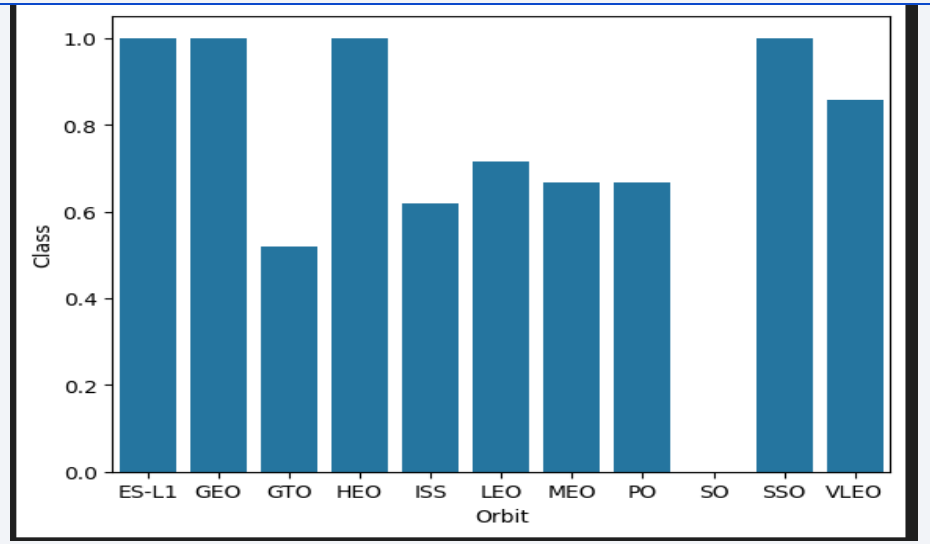
Payload vs. Launch Site

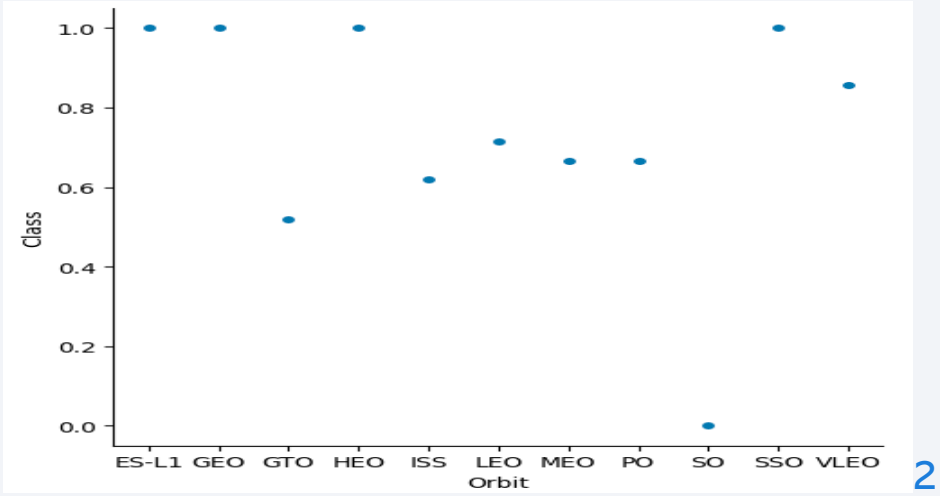


- The majority of the launches across all sites occur with a payload mass under 7000kg.
- There are no Launches for VAFB SLC-E greater than 1000kg
- CCAFS SLC-40 has the greatest number of launches below 7000kg and above 10000kg.

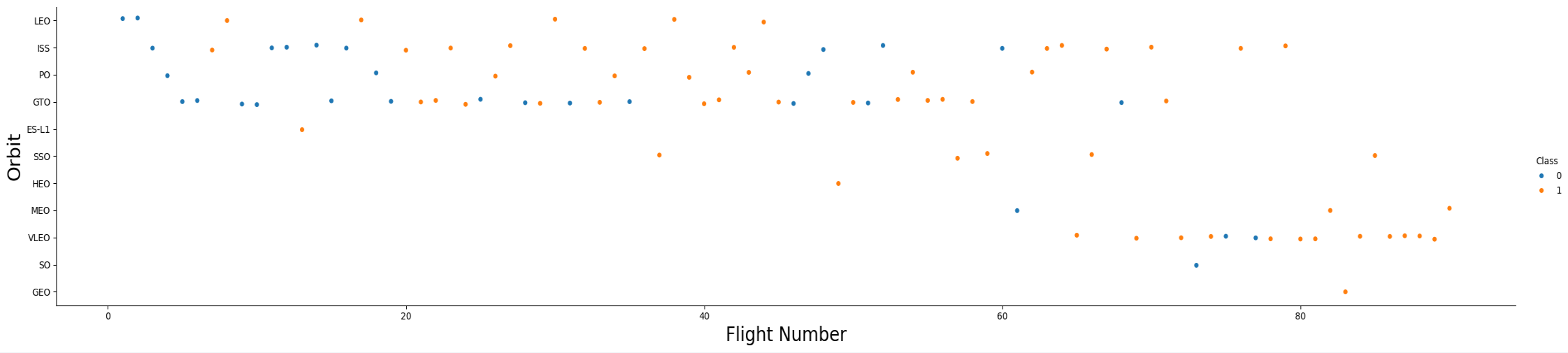
Success Rate vs. Orbit Type

- Orbit types that had the highest success rate:
  - ES-L1
  - GEO
  - HEO
  - And SSO
- Orbit GTO had the lowest success rate





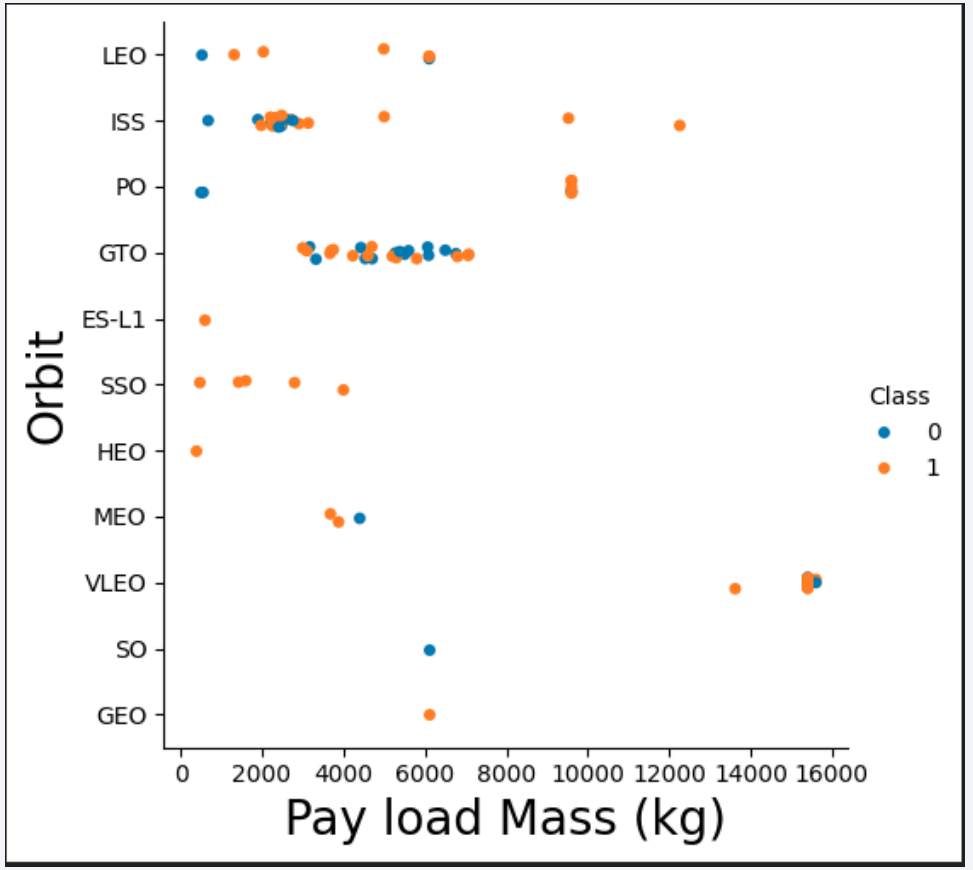
Flight Number vs. Orbit Type



- Orbits LEO, ISS, PS, and GFO are more frequent in earlier flights under 60
- VLEO Orbit type is the most frequently used with later Flight Numbers in the series past 60
- GEO orbit type was only used once

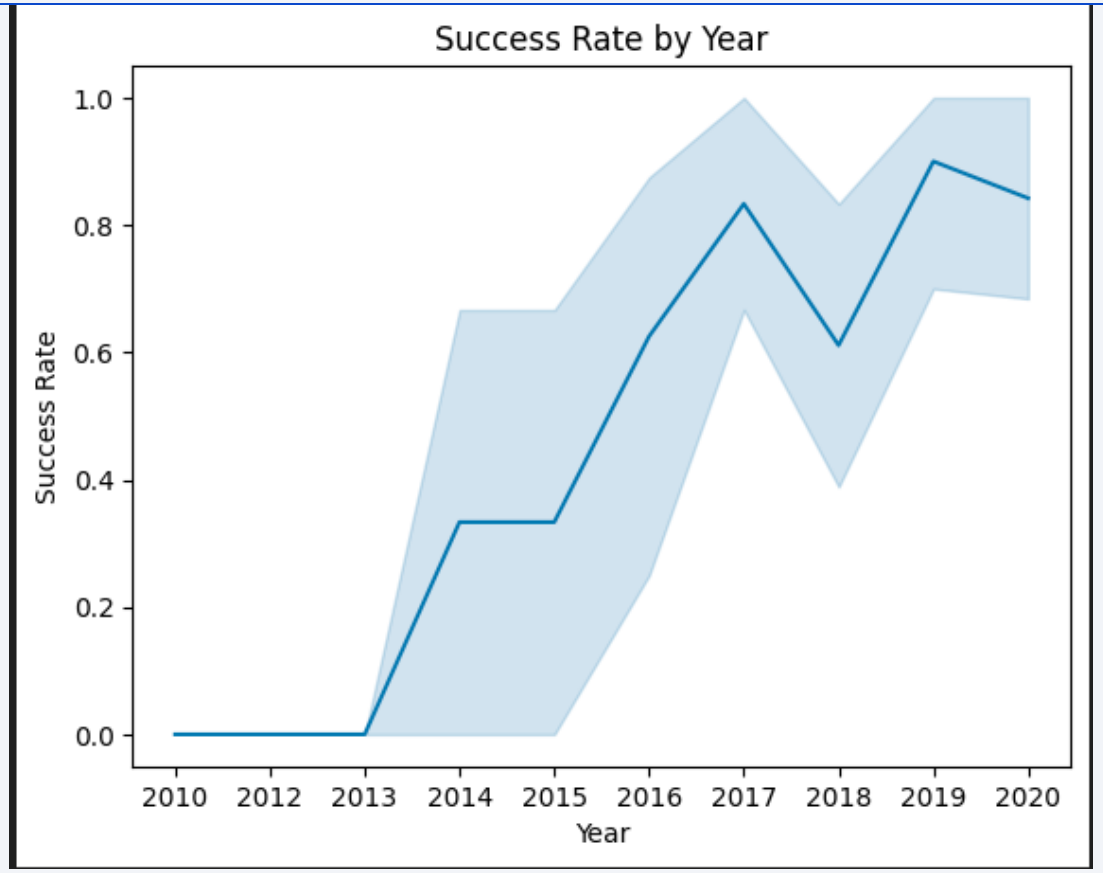
Payload vs. Orbit Type

- Highest success rate of the GTO orbit were in between Payload mass 3000kg and 8000kg
- VLEO orbit has the highest success rate for payloads greater that 15000kg



Launch Success Yearly Trend

- Success rate starts to increase starting at 2013
- Success rate is the highest in 2019



All Launch Site Names

- Queried the table for unique names in the Launch_Site column using the following query:

```
%sql select distinct Launch_Site from SPACEXTABLE

✓ 0.0s
```

- The results were the following four launch sites:

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

Launch Site Names Begin with 'CCA'

Used the following query to get the first 5 results from launch sites names that begin with 'CCA'

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

The results of the query are listed below in chronological order from earliest to latest launch dates

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

Total Payload Mass

```
%sql select SUM(PAYLOAD_MASS__KG_) as TotalPayloadMass from SPACEXTABLE where Customer like 'NASA (CRS)'
```

```
Done.
...  TotalPayloadMass
     45596
```

- Total payload mass was 45596kg

Average Payload Mass by F9 v1.1

```
%sql select AVG(PAYLOAD_MASS__KG_) as AverageMassF9 from SPACEXTABLE where Booster_Version like 'F9 v1.1%'
```

AverageMassF9
2534.6666666666665

- Average Payload Mass for the F9 v1.1 Booster was 2534.667kg

First Successful Ground Landing Date

```
%sql select MIN(Date) as FirstSuccessDate from SPACEXTABLE where Landing_Outcome like 'Success%'
```

FirstSuccessDate
2015-12-22

- Date of the First Success Ground Landing was December 22nd, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

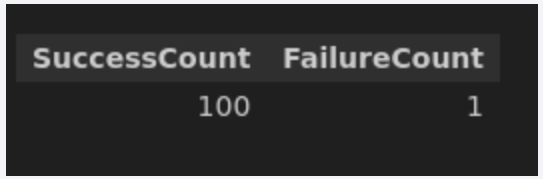
```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome like 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

- The following booster versions had a successful drone ship landing with a payload mass between 4000kg and 6000kg
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 B1031.2

Total Number of Successful and Failure Mission Outcomes

- **%sql select COUNT(Mission_Outcome) as SuccessCount, (select COUNT(Mission_Outcome) from SPACEXTABLE where Mission_Outcome like '%Failure%') as FailureCount from SPACEXTABLE where Mission_Outcome like '%Success%'**



SuccessCount	FailureCount
100	1

- Success count was 100 and Failure count was 1

Boosters Carried Maximum Payload

```
%sql select Booster_Version, PAYLOAD_MASS__KG_ as MaxAmount FROM SPACEXTABLE where (PAYLOAD_MASS__KG_) in (select MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE) ORDER by MaxAmount
```

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

- The query results for the booster versions which carried the maximum payload of 15600kg

2015 Launch Records

```
%sql select Booster_Version, Launch_Site, case substr(Date,6,2) when '01' then 'January' when '02' then 'February' when '03' then 'March' when '04' then 'April' when '05' then 'May' when
'06' then 'June' when '07' then 'July' when '08' then 'August' when '09' then 'September' when '10' then 'October' when '11' then 'November' when '12' then 'December' end as monthname
from SPACEXTABLE where Landing_Outcome like 'Failure (drone ship)' and substr(Date,0,5)='2015'
```

Booster_Version	Launch_Site	monthname
F9 v1.1 B1012	CCAFS LC-40	January
F9 v1.1 B1015	CCAFS LC-40	April

- There were only two results queried for landing outcomes that were failures on the drone ship in 2015

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

```
%sql select Landing_Outcome, count(Landing_Outcome) as outcomecount from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by outcomecount desc
```

Landing_Outcome	outcomecount
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

- Landing outcomes and their count are ranked in from highest to lowest

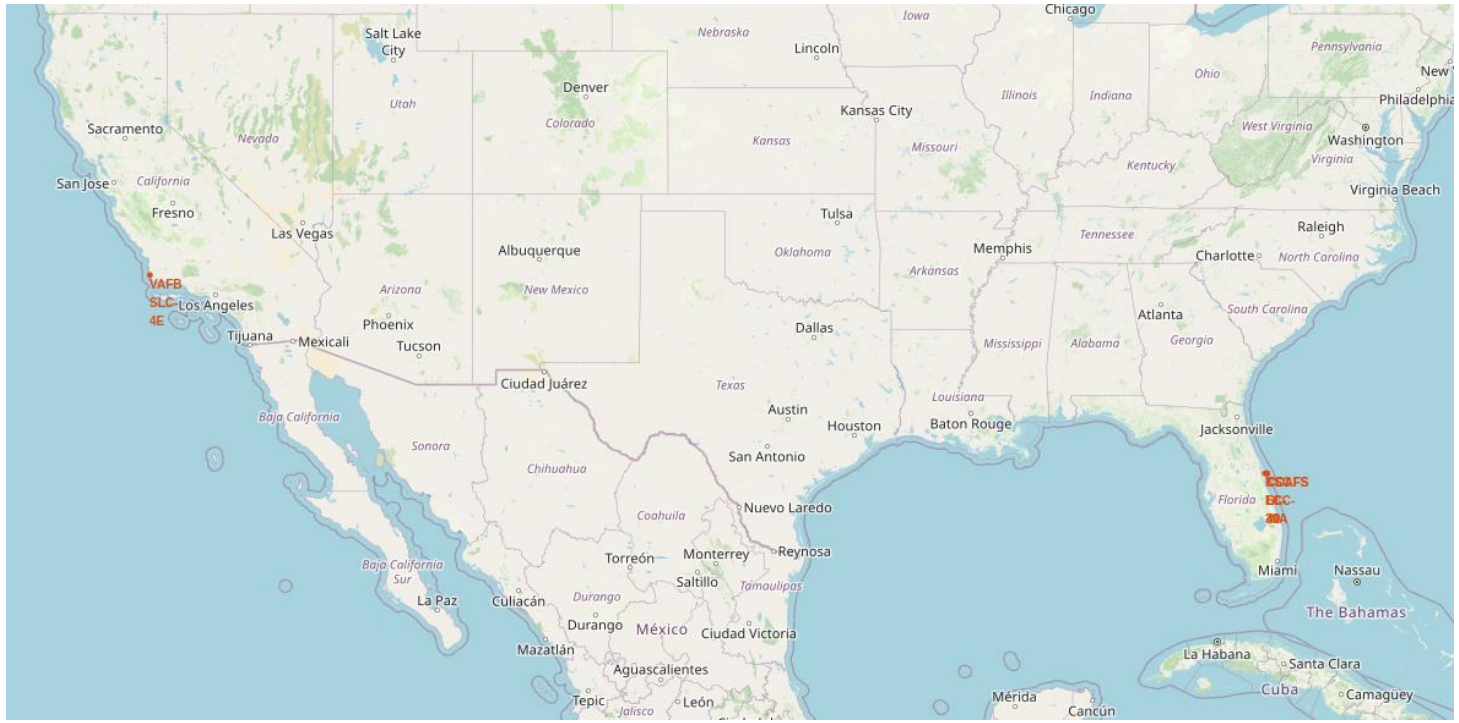
Section 3

Launch Sites Proximities Analysis

Launch Site Locations Map

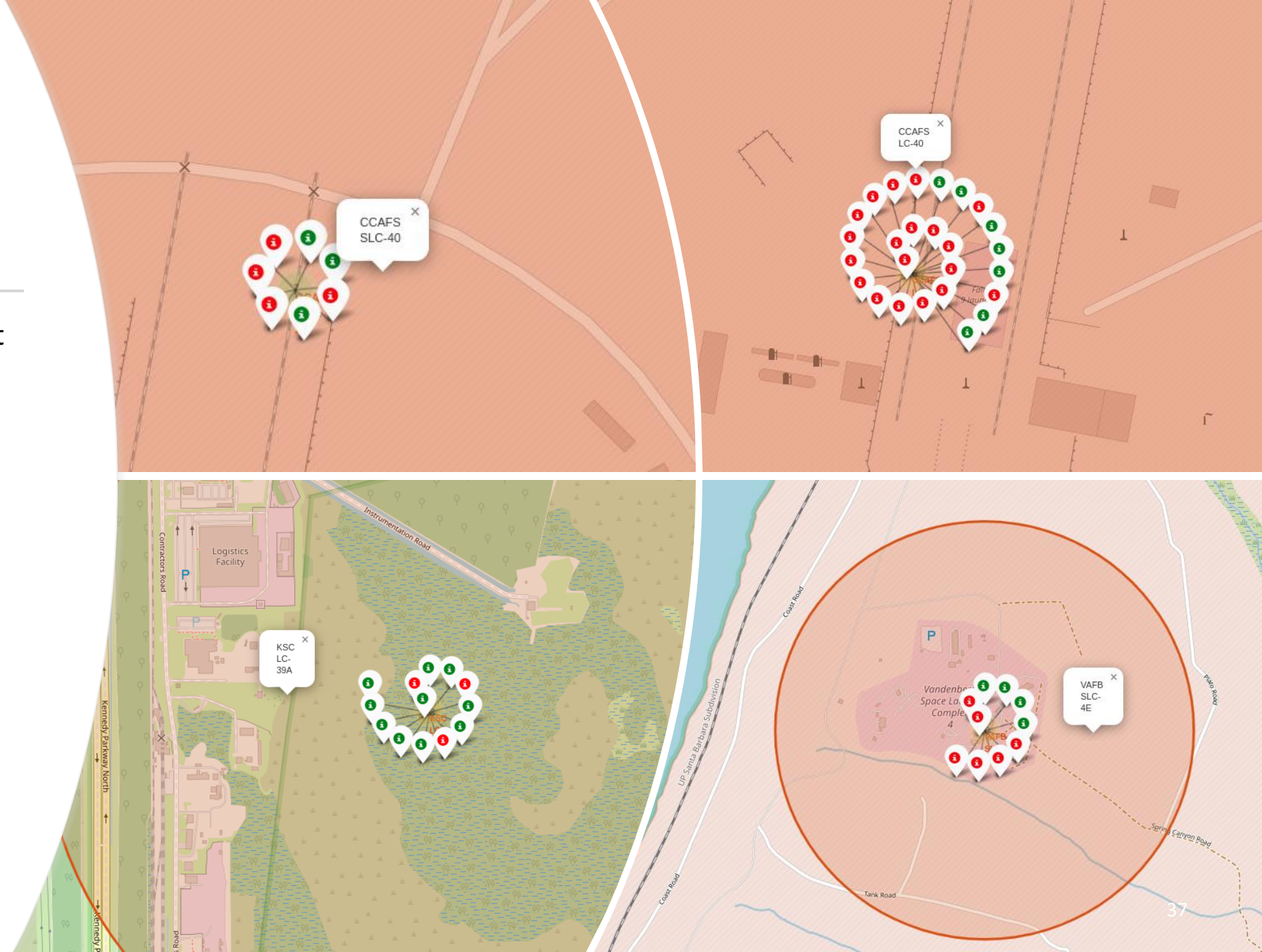
*Only 1 Site is located on the West Coast of USA.

- All sites are in close proximity to an ocean coastline.
- All sites are near the eqator
- The other three sites are located in close proximity to Cape Canaveral Space Force Station

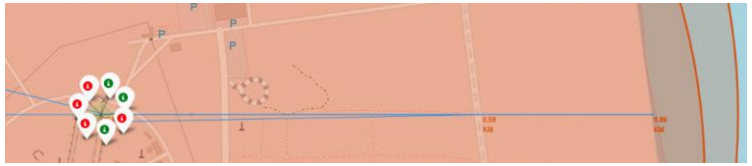


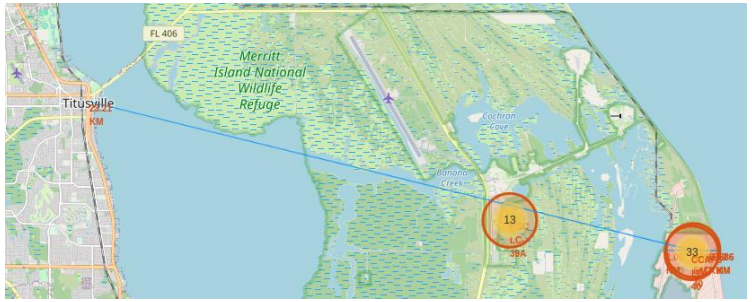
Launch Site Outcomes Map

- KS CLC-39A has the highest rate of successful launch outcomes
- VAFB SLC-4E has almost all successful launches on the eastern side of the launch site
- CCAFS LC-40 has the greatest amount of attempts and launch failures
- CCAFS SLC-49 has the fewest launch attempts

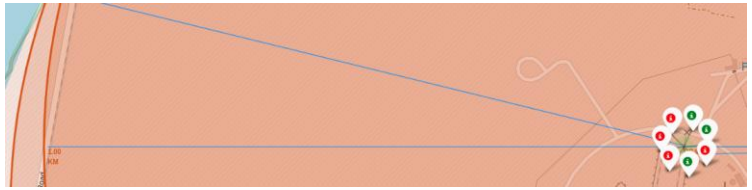


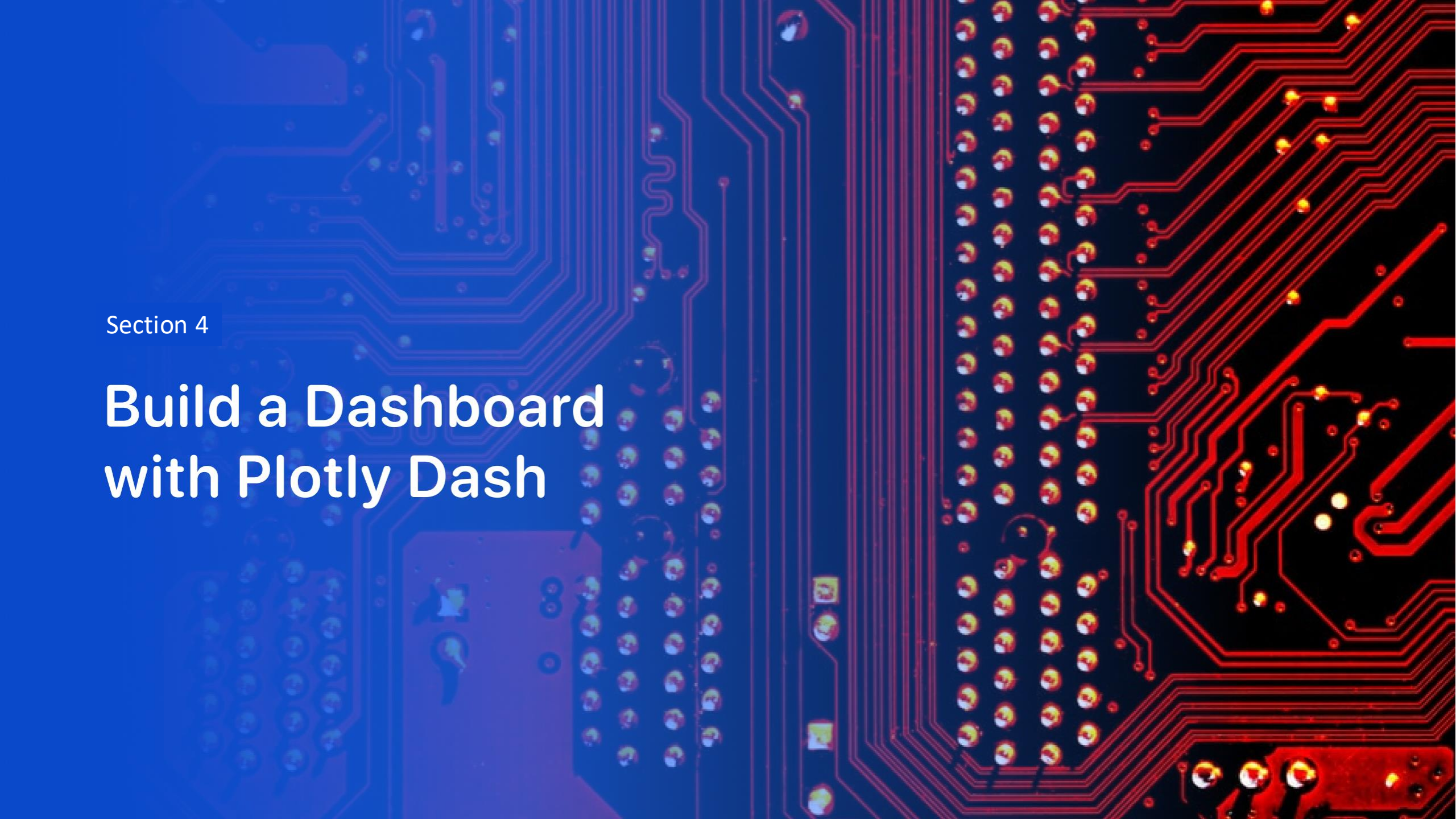
CCAFS SLC-40 Proximities to City, Hwy, and Rail



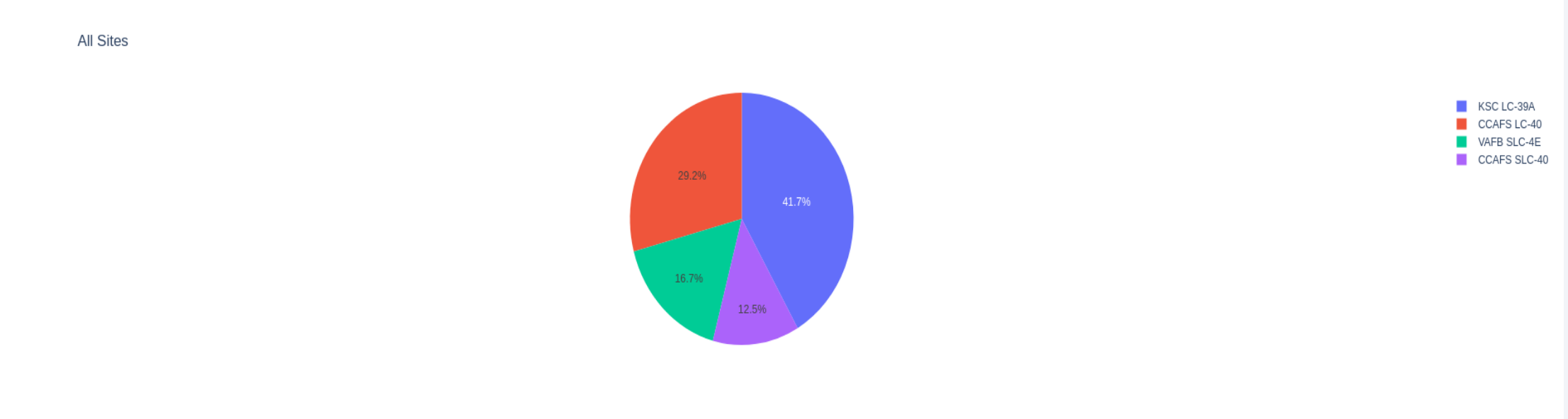


- CCAFS SLC-40 is relatively close to coastline, rail and major highway all within 1km of the launch site.
- Closest City to the launch Site it Titusville which is 23km away.
- The Launch sites are gererally at least 20km away from major cities and would be near the coastline as well as Hwy and Rail for access to supplies and equipment.



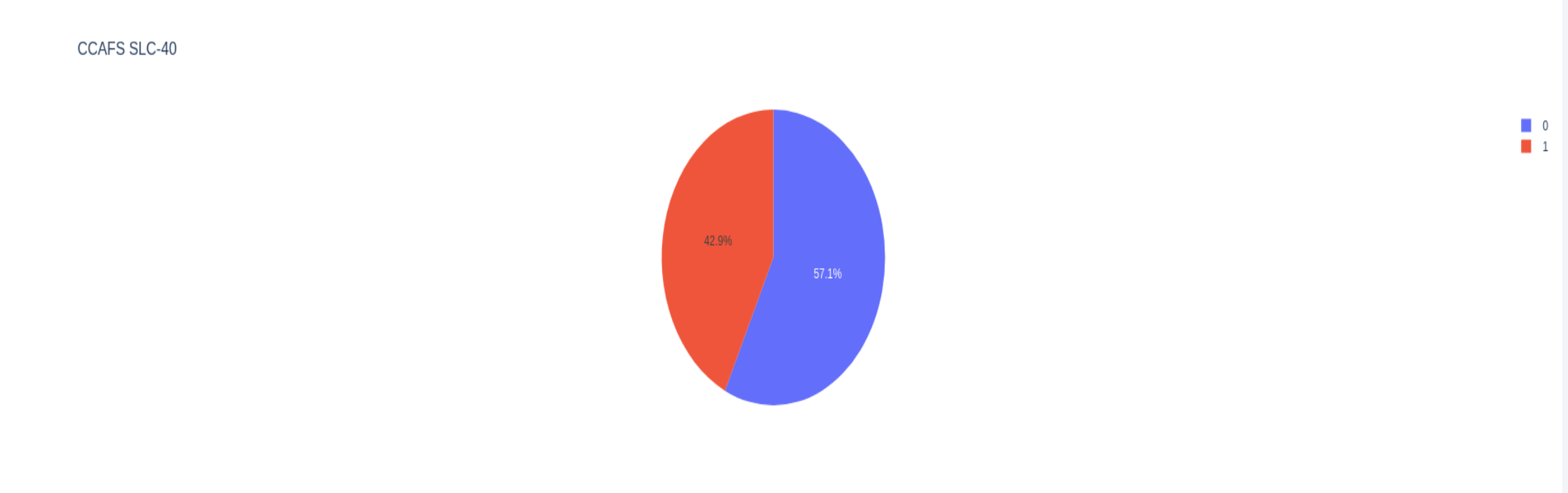

Section 4
Build a Dashboard with Plotly Dash

All Sites Launch Success



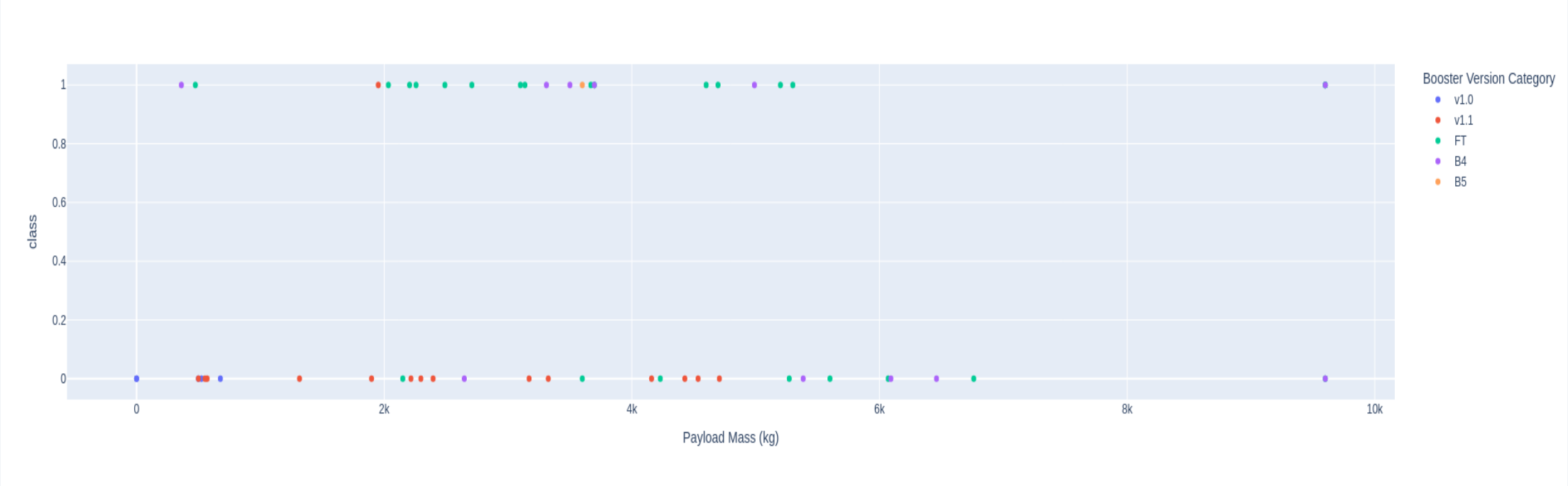
KSC LC- 39A Has the highest launch success while CCAFS LC-40 is the second highest. The last two sites share similar Launch success rates.

Highest Launch Success



- CCAFS SLC–40 Has a success rate of 42.9% while the failure rate for the launches is 57.1%

Successful Launches By Payload Mass

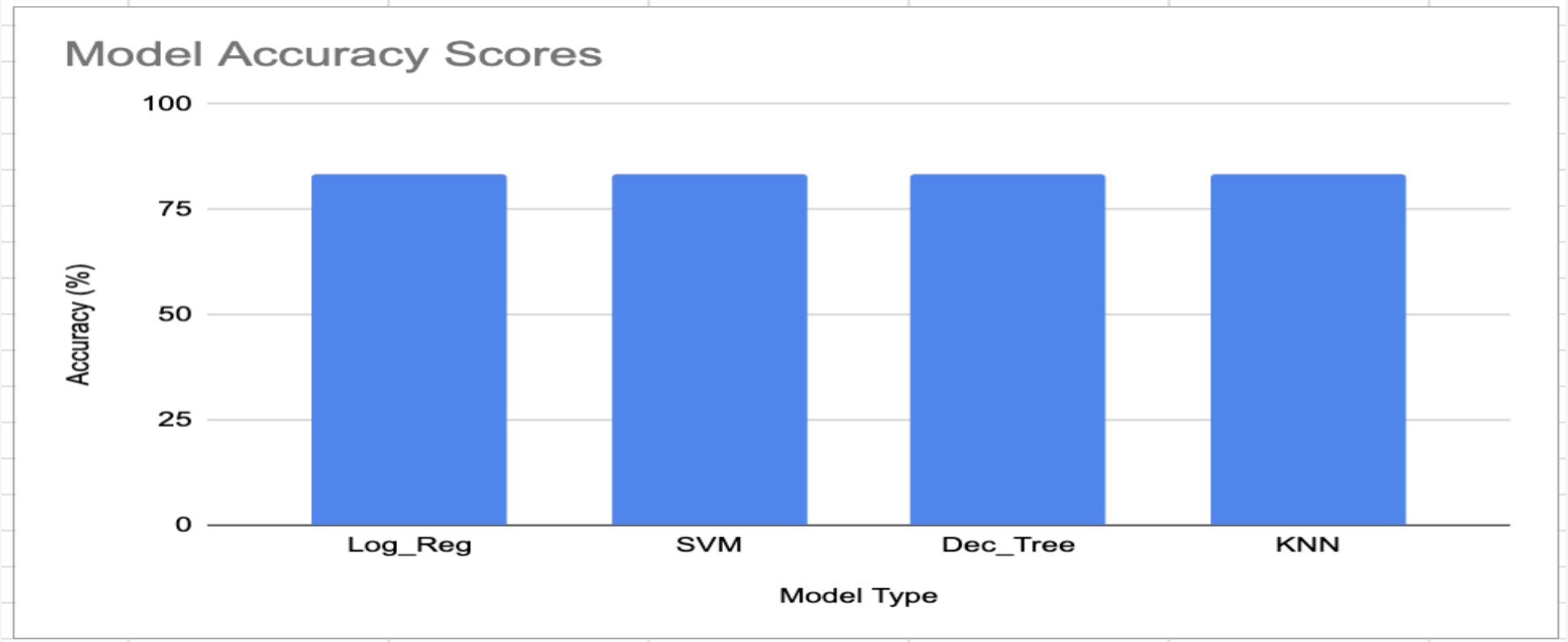


- Booster Version FT has the highest amount of successful launches between 2k to 3k and also between 4k to 6k
- Booster version B4 has the highest payload successful launch

Section 5

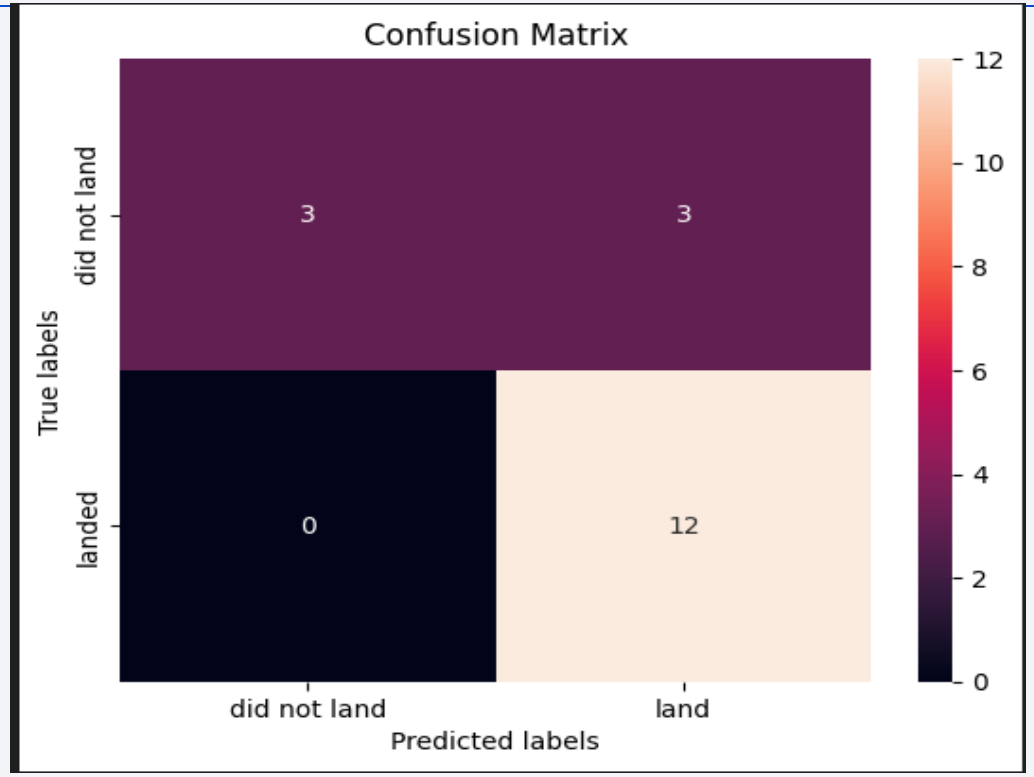
Predictive Analysis (Classification)

Classification Accuracy



- All Models had similar results which were 83%

Confusion Matrix



- 12 Successful Launches were successfully predicted
- 3 Accurate predictions were observed for failures and 3 false positives were predicted as well

Conclusions

- Most successful launches were in 2019
- Machine learning is able to predict landing outcome with a relative level of success, however the sample set would be more accurate if it had included the data for more launches
- Success among different payloads can be achieved by utilizing the booster version as well as orbit type.
- As more launches occur in the future their data can be added and these models can be re-evaluated for a greater degree of accuracy

Thank you!