IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nikita Chumakov
Feb 2023

Outline

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

Executive Summary

- The following methodologies were used to complete this project:
  - Data was collected with API & Web Scraping
  - Data was extracted with Load & Transform
  - Data was cleaned so values could be used for the analyses
  - Exploratory Data Analysis (EDA) visualization
  - EDA with SQL
  - Building interactive dashboard with Plotly Dash
  - Predictive Analysis with various ML models

Introduction

- Space Exploration Technologies Corp. (SpaceX) is an American spacecraft manufacturer, launcher, and a satellite communications corporation founded in 2002 by Elon Musk.
- SpaceX stated goal of reducing space transportation costs to enable the colonization of Mars. The company manufactures the Falcon 9, Falcon Heavy, and Starship launch vehicles, several rocket engines, Cargo Dragon and Crew Dragon spacecraft, and Starlink communications satellites.
- The main objective of this project is to predict the probability of successful landing of the booster based on historical data of Falcon 9 rocket launches. The following parameters are included in the model: launch site, orbit, payload, booster version, etc.

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

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

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- The following charts were plotted as part of data visualization:
  - 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

- GitHub URL - https://github.com/NikitaChumakov/Capstone/blob/main/EDA%20dataviz.ipynb

EDA with SQL

- The following SQL queries were executed:
    - All Launch Site Names
    - Launch Site Names Begin with 'CCA'
    - Total Payload Mass
    - Average Payload Mass by F9 v1.1
    - First Successful Ground Landing Date
    - Successful Drone Ship Landing with Payload between 4000 and 6000
    - Total Number of Successful and Failure Mission Outcomes
    - Boosters Carried Maximum Payload
    - 2015 Launch Records
    - Rank Landing Outcomes Between 2010-06-04 and 2017-03-20
- GitHub URL - https://github.com/NikitaChumakov/Capstone/blob/main/EDA%20SQL.ipynb

Build an Interactive Map with Folium

- The following map objects were added to the map:
  - Circles – showing launch sites locations
  - Markers – showing names of launch sites
  - Lines – to illustrate the shortest distance to an objects like city, coastline, etc.
- GitHub URL - https://github.com/NikitaChumakov/Capstone/blob/main/SPACEX%20-%20FOLIUM.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

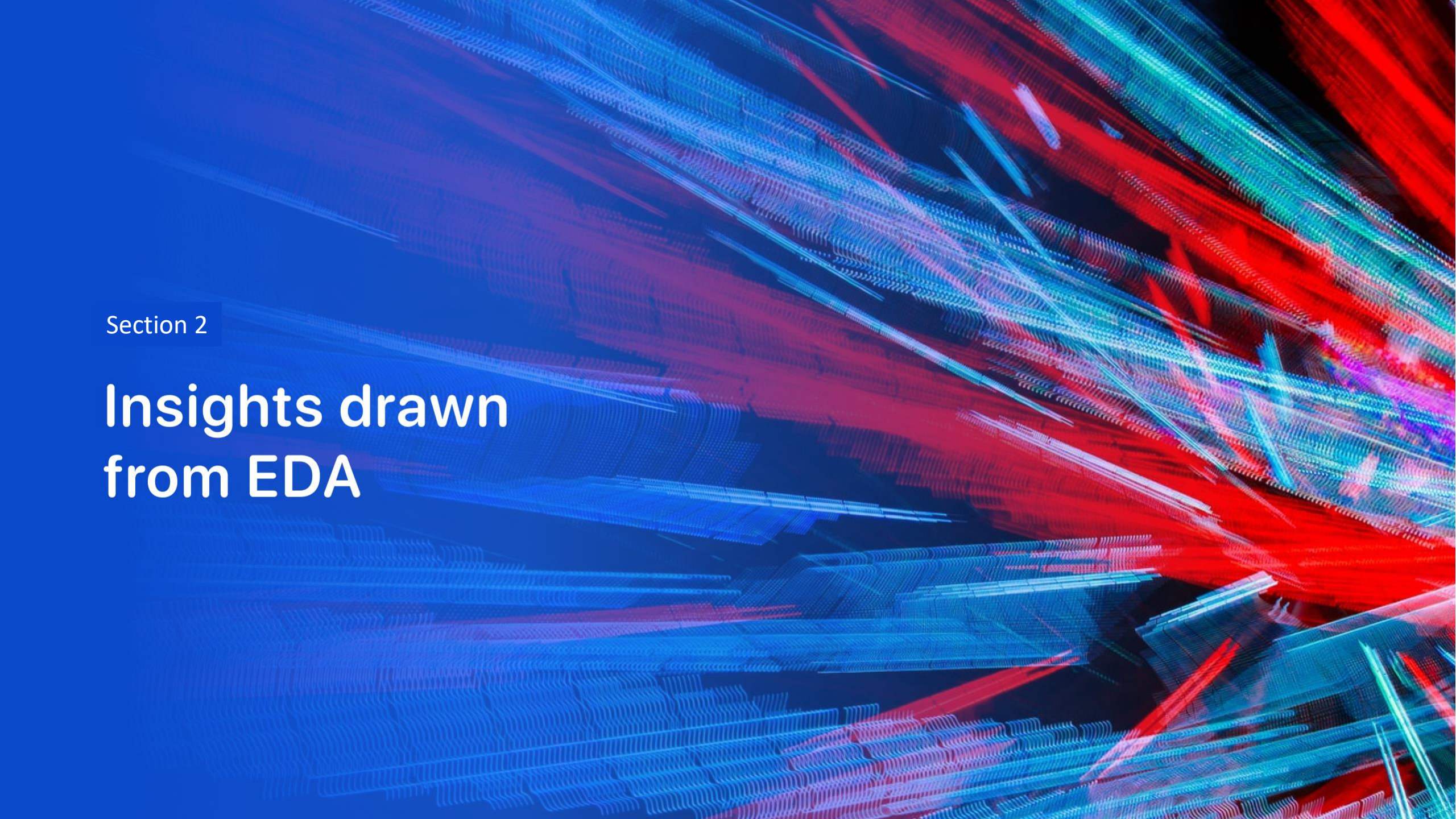
- The dashboard contained the following objects:
  - Dropdown list – enabling the selection of All sites or one of the specific launch sites
  - Pie chart – to represent the share of successful / unsuccessful launches of the selected launch site
  - Slider – to modify / set range of workload for the scatter chart
  - Scatter chart – to illustrate the outcome of the selected load range launches
- GitHub URL - https://github.com/NikitaChumakov/Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

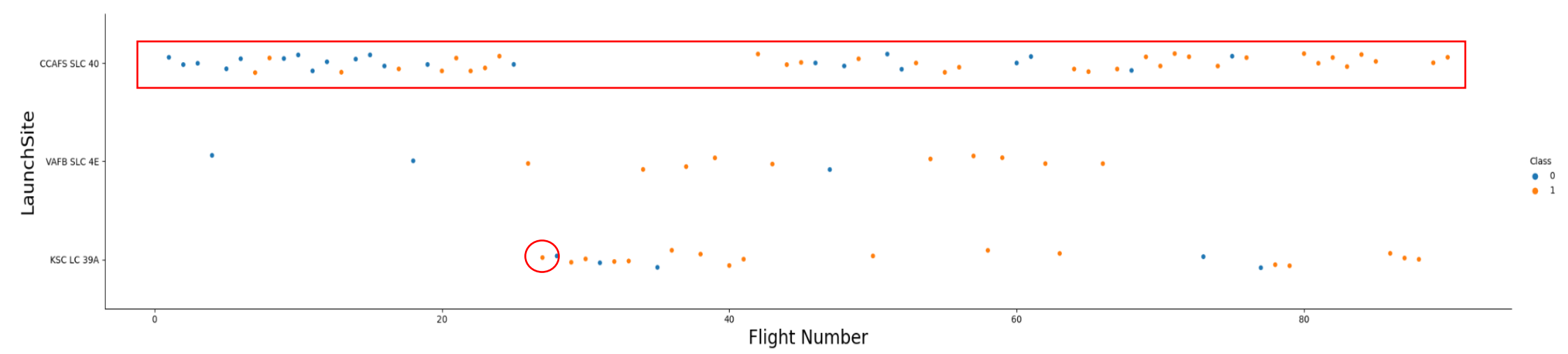
- Predictive analysis part contained the following steps:
  - Standardization of the data
  - Splitting into training data and test data
  - Finding best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Finding the method performs best using test data
- GitHub URL - https://github.com/NikitaChumakov/Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_ML.py

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

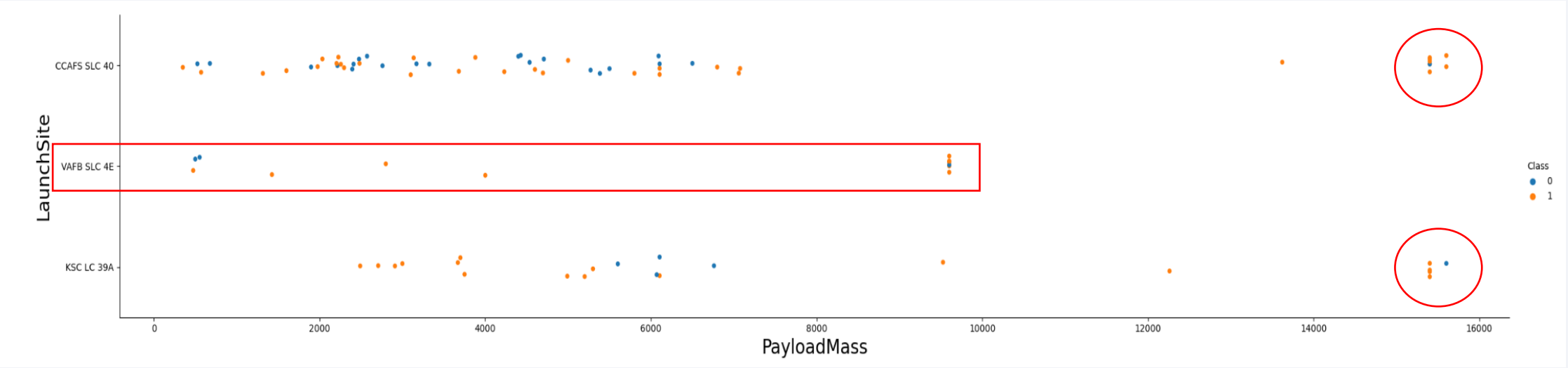

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



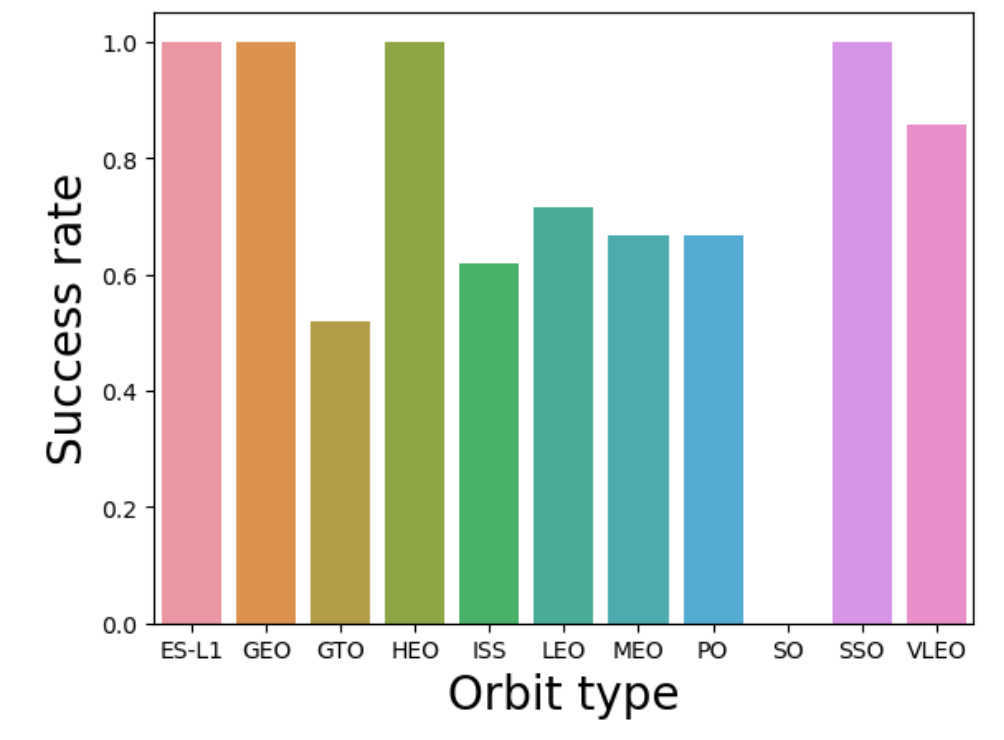
- CCAFS SLC 40 site was used for the first launch and remained the main launch site for the first 26 launches
- It was 27th launch when KSC LC 39A site was used for the first time
- VAFB SLC 4E site was not utilized lately (for the last 24 launches)

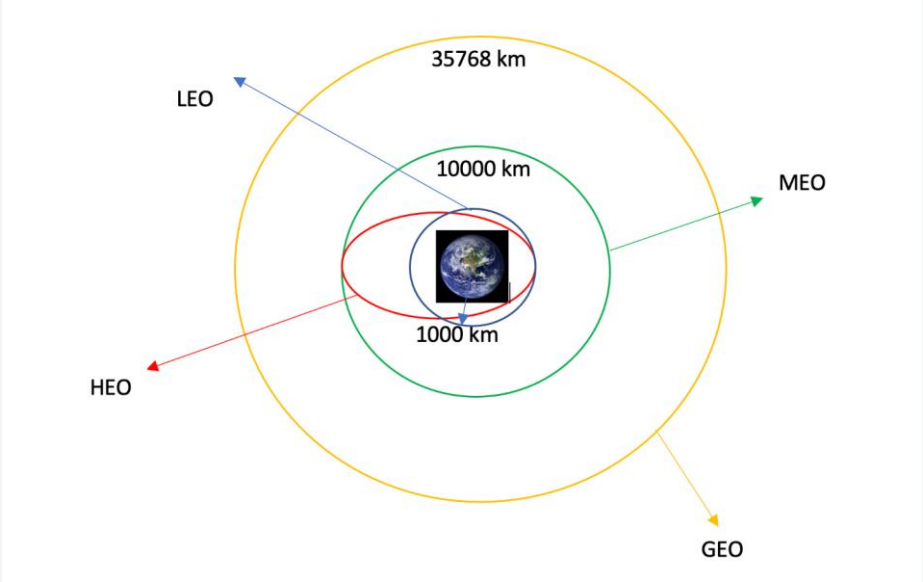
Payload vs. Launch Site



- Maximum payload of Falcon 9 launched from VAFB SLC 4E site was <10000 kgs
- Both CCAFS SLC 40 and KSC LC 39A sites were used to launch Falcon 9s with >15000 kgs payload
- CCAFS SLC 40 has the largest number of launches with >15000 kgs payload

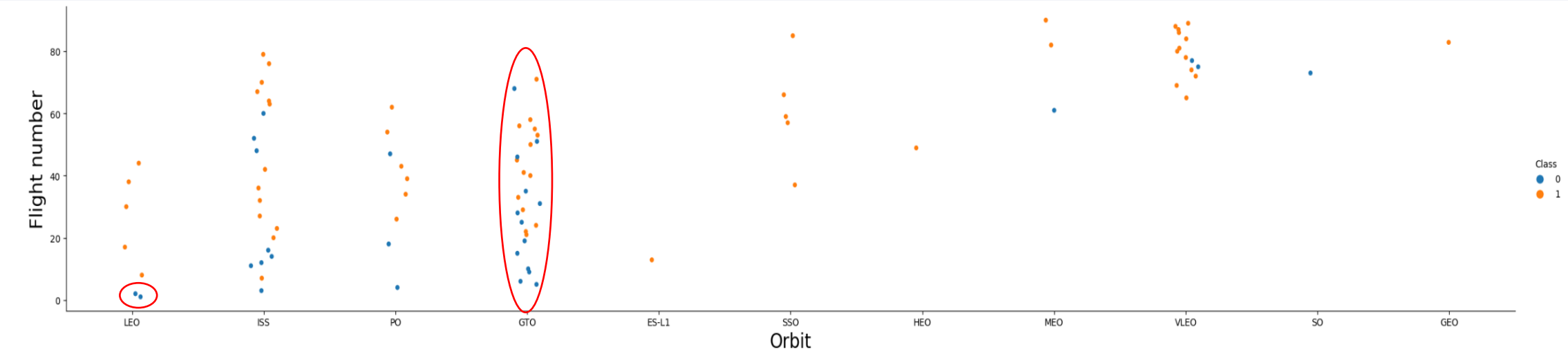
Success Rate vs. Orbit Type





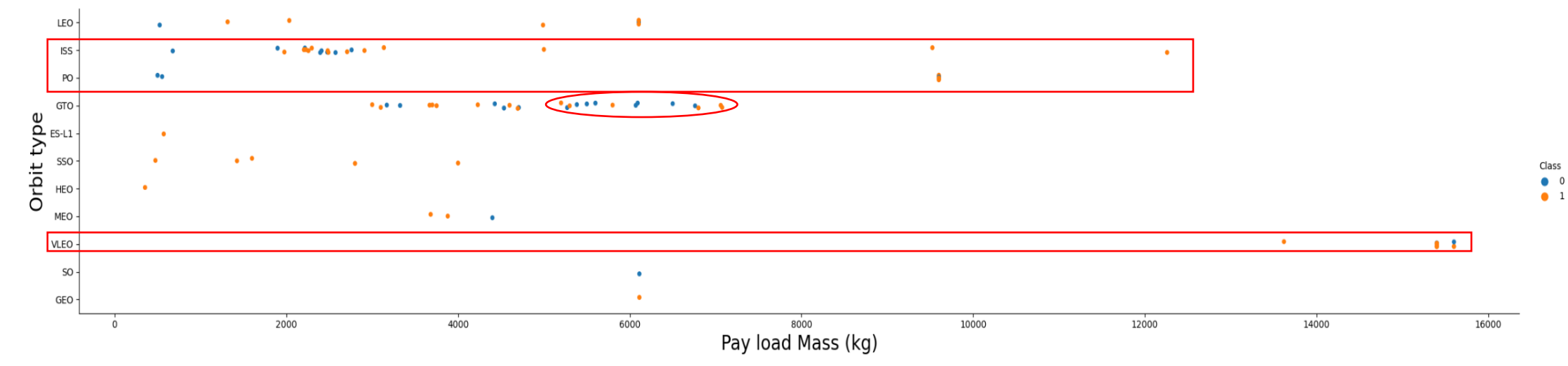
- Orbits ES-SL1, GEO, HEO and SSO have the highest launch success rate
- So far orbit GTO had the lowest success rate

Flight Number vs. Orbit Type



- Launch success for LEO orbit changed with the number of flights, first two launches failed
- GTO orbit seems not to have a relationship between flight number and launch success
- Orbits ES-L1, HEO, SO and GEO only one launch each
- SSO orbit is the only one which had 100% success rate for more than 1 launch

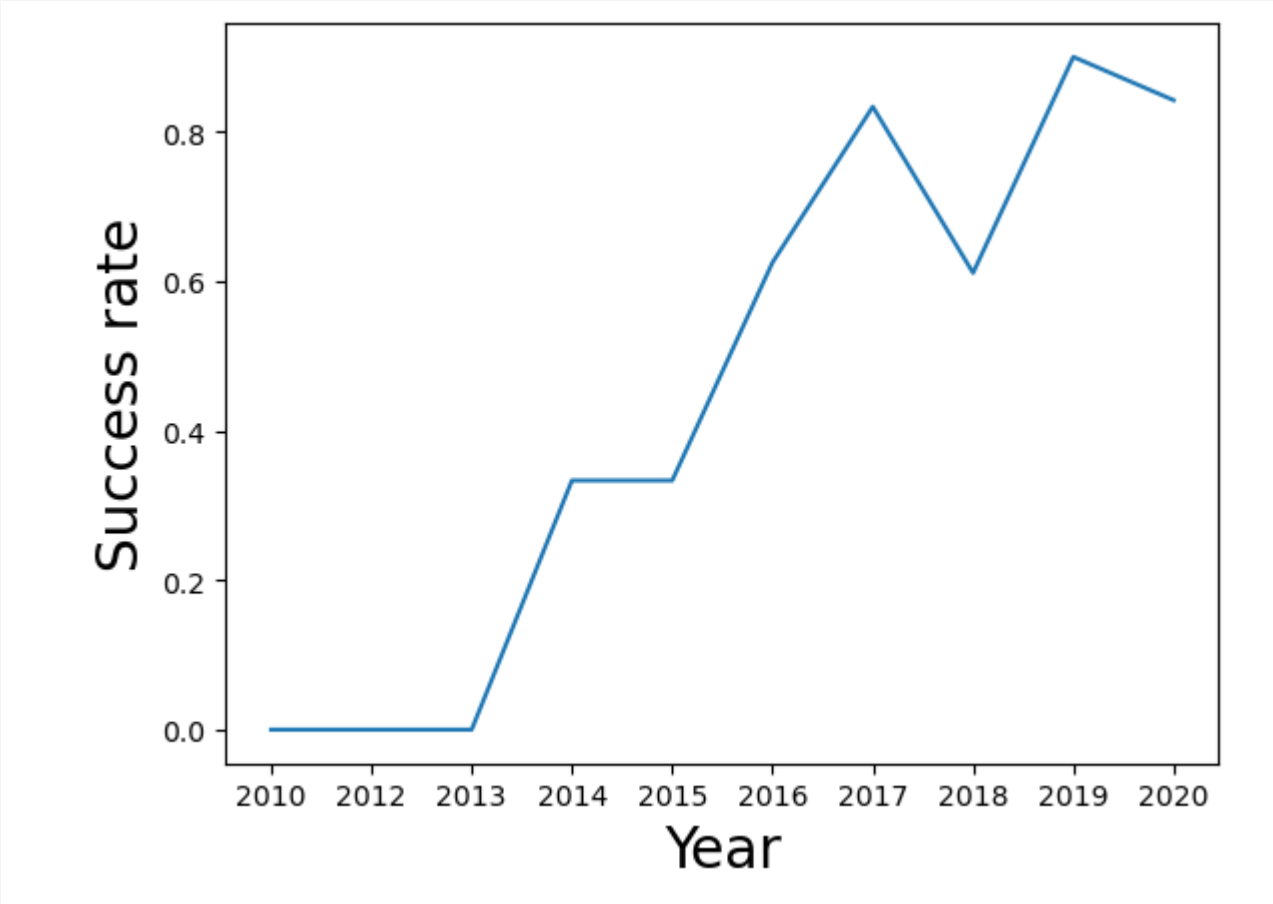
Payload vs. Orbit Type



- PO, VLEO and ISS orbits have high successful landing rate for heavy payloads launches
- GTO orbit has a mixed history of positive and negative landings for heavy payloads launches

Launch Success Yearly Trend

- The launch success rate increased from 0% in 2010-2013 up to ~80% in 2020
- The highest launch success rate was demonstrated in 2019, it reached >80%
- All launches during first three years were unsuccessful



All Launch Site Names

- The unique names launch sites were as following:

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

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX
```

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA` are:

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

```
%sql SELECT * FROM SPACEX WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

Total Payload Mass

- Total payload carried by boosters from NASA:

launch_site	2
CCAFS LC-40	67363
CCAFS SLC-40	254037
KSC LC-39A	208837
VAFB SLC-4E	89730

%sql SELECT launch_site, SUM(payload_mass__kg_) FROM SPACEX GROUP BY launch_site

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2534 kg


```
%sql SELECT avg(payload_mass__kg_) FROM SPACEX WHERE booster_version LIKE 'F9 v1.1%'
```

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad is 2015-12-22

```
%sql SELECT MIN(DATE) FROM SPACEX WHERE landing__outcome='Success (ground pad)'
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Out[11]:

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

```
%sql SELECT booster_version FROM SPACEX WHERE landing__outcome='Success (drone ship)'
AND PAYLOAD_MASS__KG_>4000 AND PAYLOAD_MASS__KG_<6000
```

Total Number of Successful and Failure Mission Outcomes

- Total number of successful outcomes is 100 (99 success and 1 success outcome with unclear payload status) and total number of failures is 1

Out[13]:

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

%sql SELECT Mission_Outcome, count(Mission_Outcome) FROM SPACEX GROUP BY Mission_Outcome

Boosters Carried Maximum Payload

- There were 12 boosters versions which carried maximum workload of 15600 kgs:

Out[31]:

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

```
%sql SELECT BOOSTER_VERSION FROM SPACEX WHERE PAYLOAD_MASS__KG_ =
(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX)
```

2015 Launch Records

- Below is the list of 2 failed landing outcomes in drone ship including their booster versions and launch site names for 2015:

Out[32]:

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

%sql SELECT booster_version, launch_site FROM SPACEX WHERE landing__outcome = 'Failure (drone ship)' AND year(Date)='2015'

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

- Table below contains counts of landing outcomes between 2010-06-04 and 2017-03-20, sorted in descending order:

Out[53]:

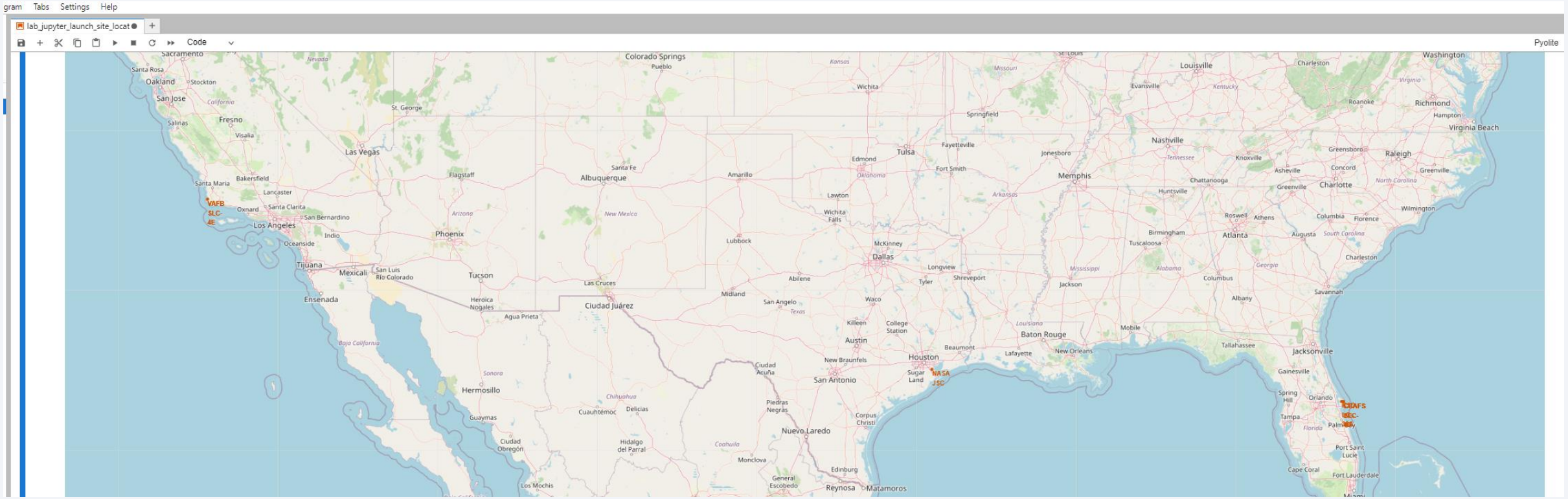
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

```
%sql SELECT landing__outcome, count(*) as COUNT FROM SPACEX WHERE DATE>='2010-06-04' and DATE <= '2017-03-20' GROUP BY landing__outcome ORDER BY COUNT DESC
```

Section 3

Launch Sites Proximities Analysis

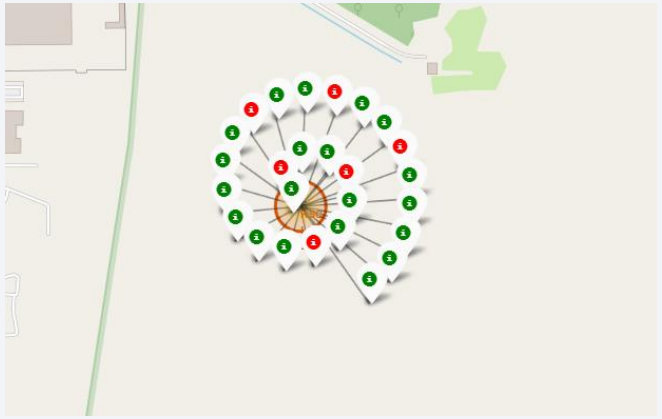
Location of launch sites



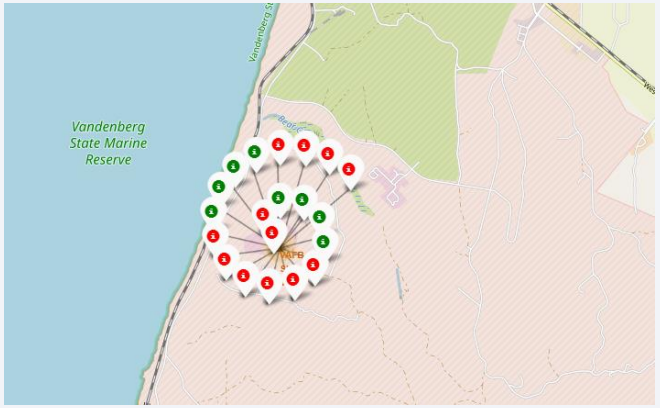
- One launch site is on the West Coast - VAFB SLC-4E
- Three launch sites are located on an East Coast – KSC LC-39A, CCAFS LC-40 and CCAFS SLC-40
- CCAFS LC-40 launch site is closest to the equator

Launch outcomes by launch sites

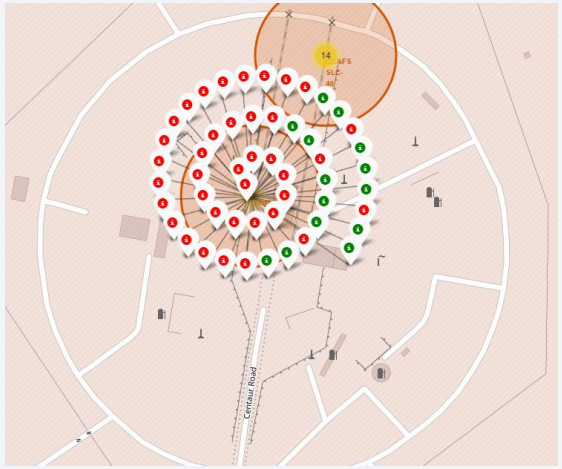
KSC LC-39A



VAFB SLC-4E



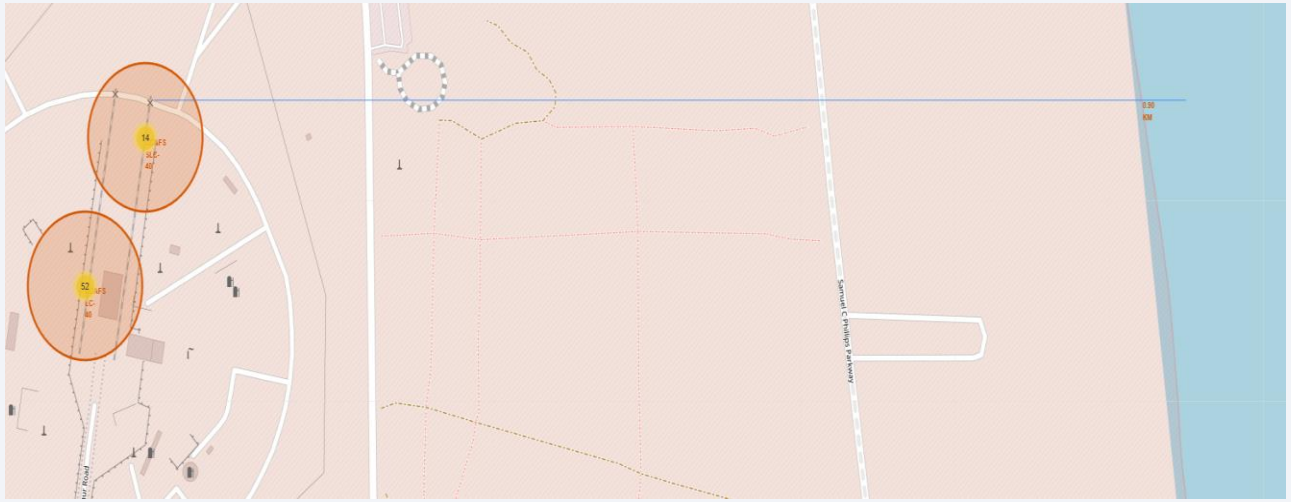
CCAFS LC-40



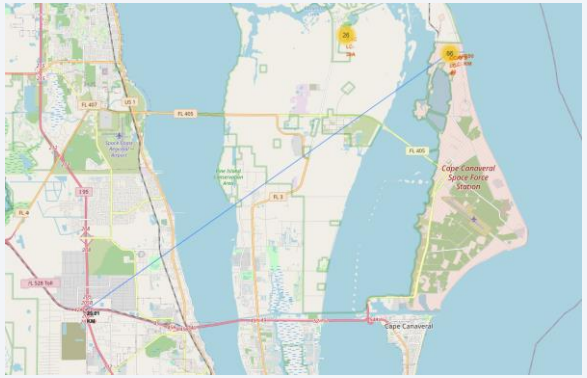
CCAFS SLC-40

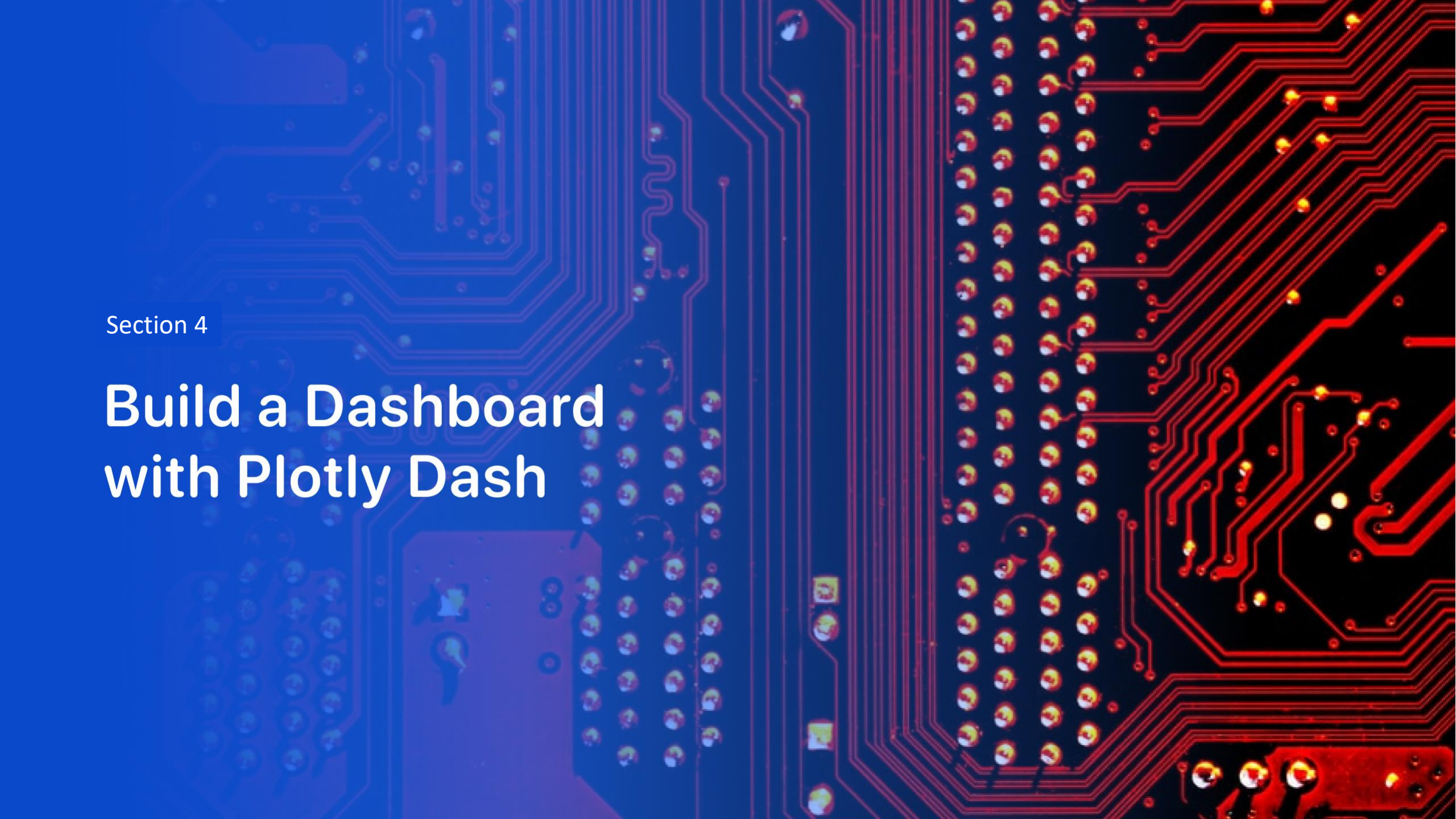
Proximity of CCAFS SLC-40 to a city and coastline

Proximity to coastline – 0.9 km

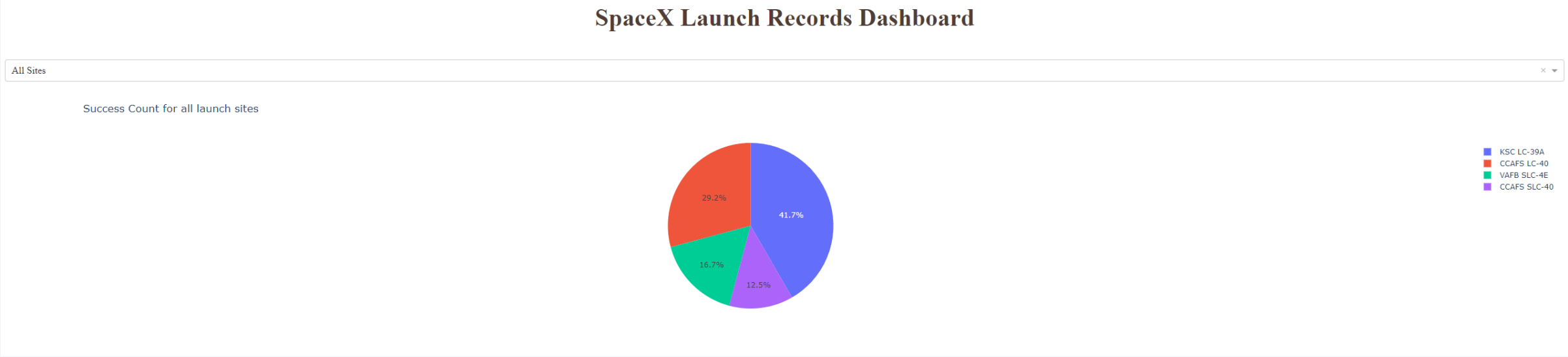


Proximity to a city of Canaveral – 29.2 km



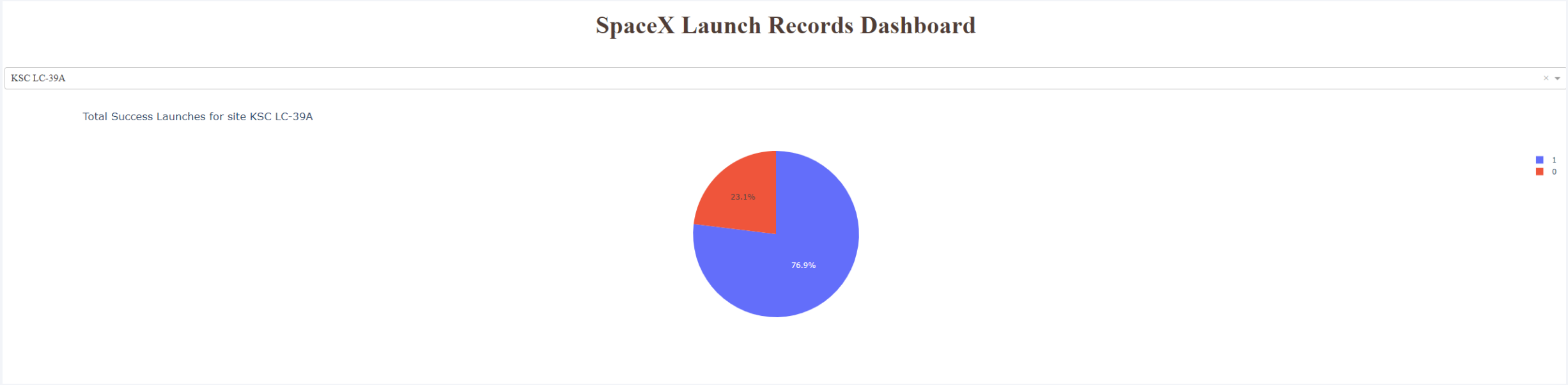

Section 4
Build a Dashboard with Plotly Dash

Success count for all launch sites



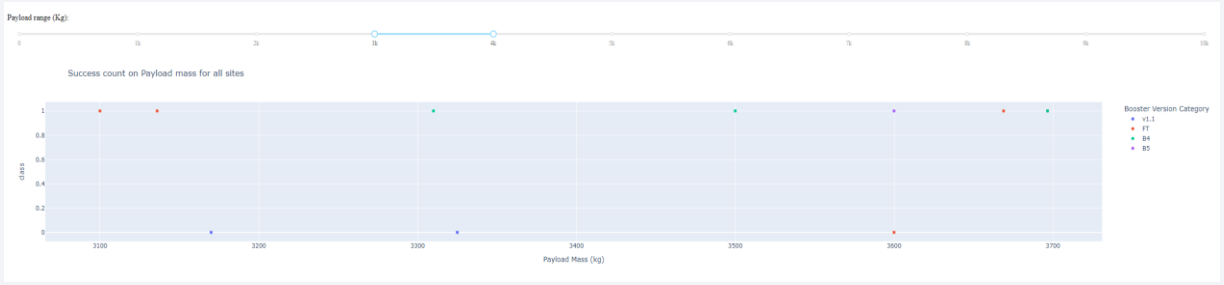
- The KSC LC-39A site had 10 successful launches which constitutes 41.7% of total number
- CCAFS SLC-40 site had only 3 successful launches which contributed 12.5% to the total number of successful launches

Total success launches for site KSC LC-39A



- The KSC LC-39A site had 10 successful and 3 unsuccessful launches

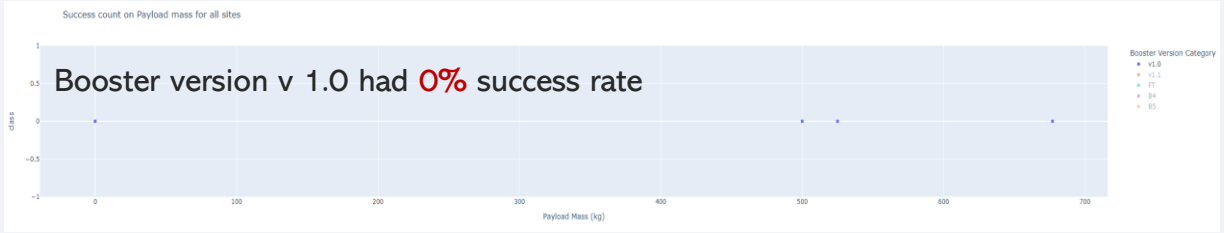
Success rates per payload range and booster version

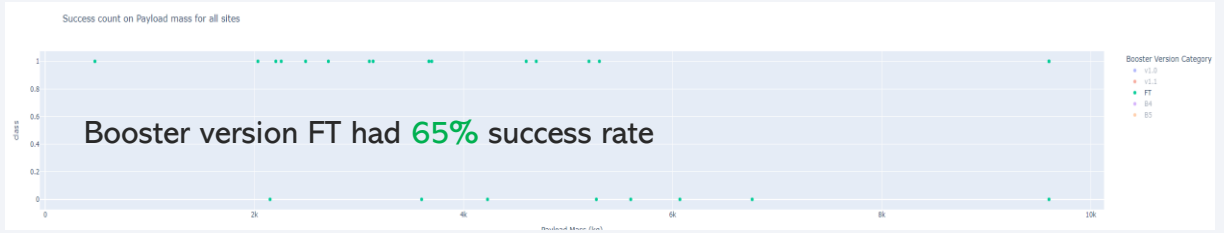


- Range between 3,000 and 4,000 kgs had 70% success

- Range between 2,000 and 3,000 kgs had 50% success

- Range between 6,000 and 8,000 kgs had 0% success

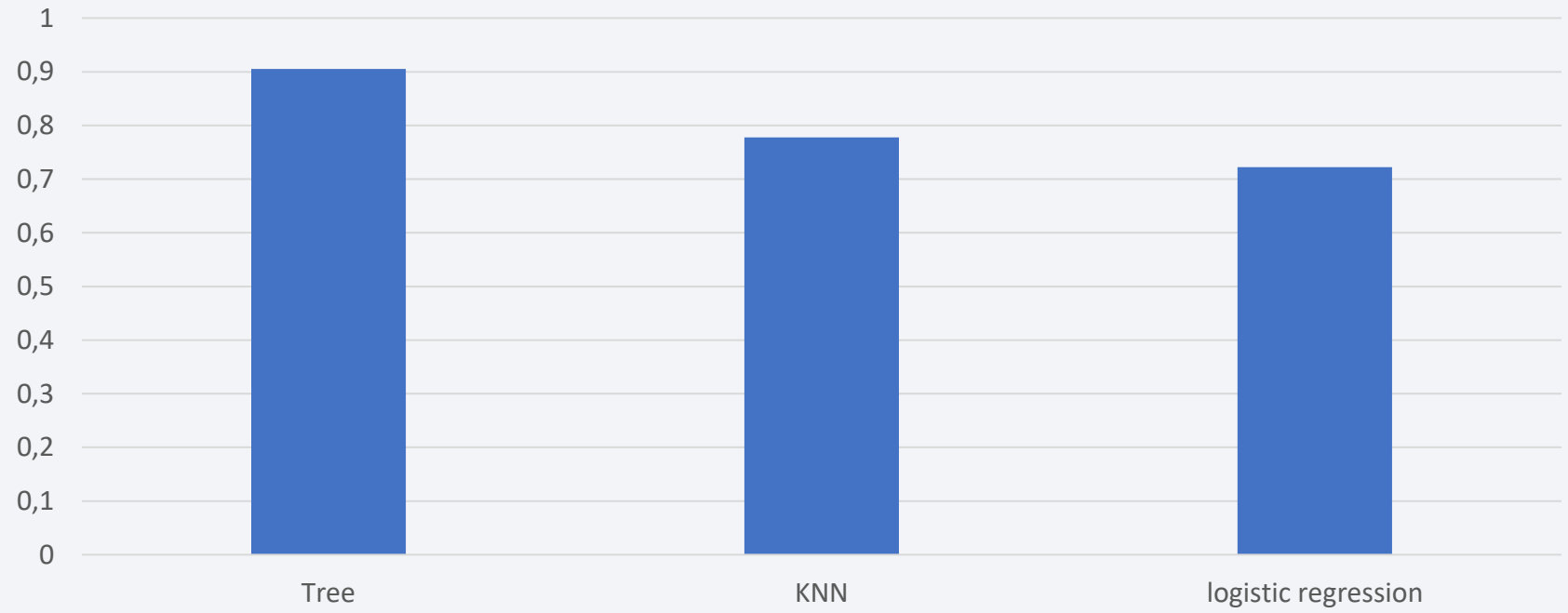




Section 5

Predictive Analysis (Classification)

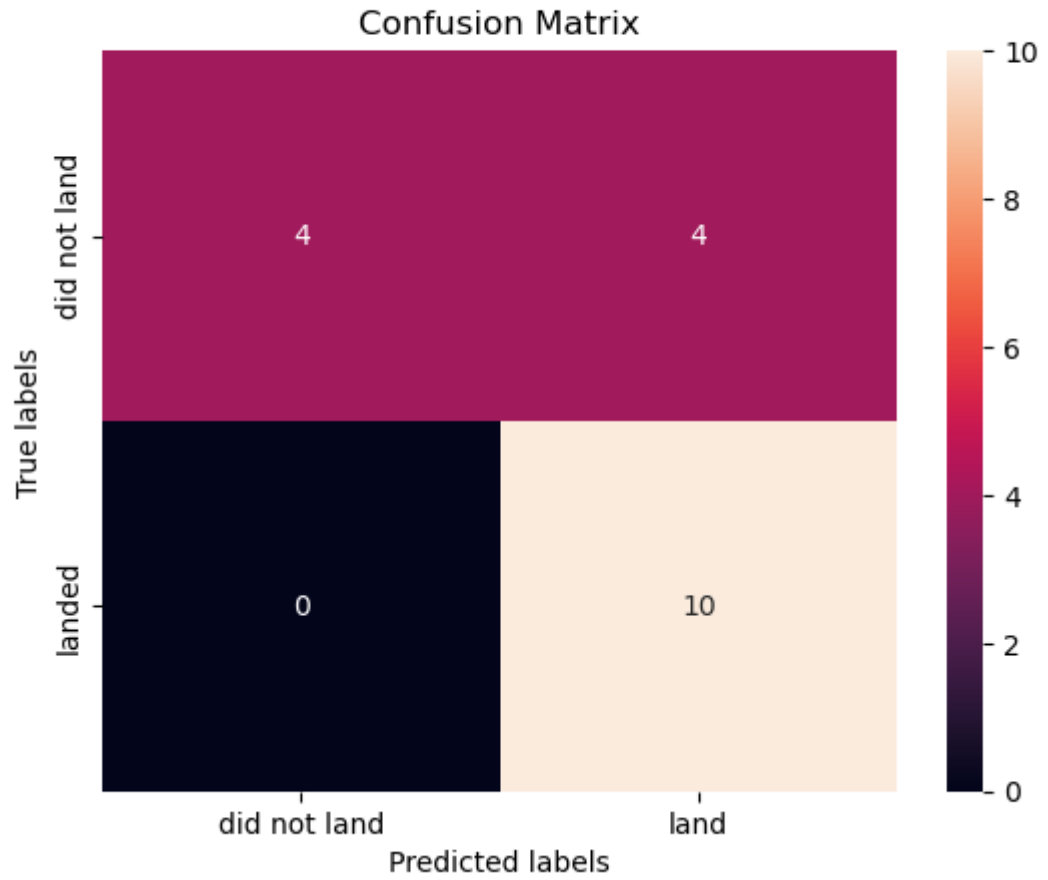
Classification Accuracy



- Best Algorithm is Tree with a score of 0.9053571428571429

Confusion Matrix

```
In [65]: yhat = svm_cv.predict(X_test)
         plot_confusion_matrix(y_test,yhat)
```



Conclusions

- Orbits ES-SL1, GEO, HEO and SSO have the highest launch success rate
- The launch success rate increased from 0% in 2010-2013 up to ~80% in 2020
- There is a correlation between the payload and launch outcome, the heavier the payload, the higher risk of unsuccessful launch:
  - Range between 3,000 and 4,000 kgs had 70% success
  - Range between 6,000 and 8,000 kgs had 0% success

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!