IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

J. Hidalgo Sánchez
December 2023

# Outline

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

# Executive Summary

- In this capstone, we predicted if the Falcon 9 first stage will land successfully by implementing several machine learning models.
- Models developed were Logistic Regression, Support Vector Machine, Decision Tree and KNN.
- Data was collected form SPACE X using its own API and from Wikipedia.
- Several data visualizations were developed in order to perform exploratory data analysis.
- Best models reach an accuracy of 83%.
- Amn interactive dashboard and terrain maps were also generated.

# Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Furthermore, an interactive dashboard and several data visualizations were generated with the purpose to gain a better data understanding.

Section 1

# Methodology

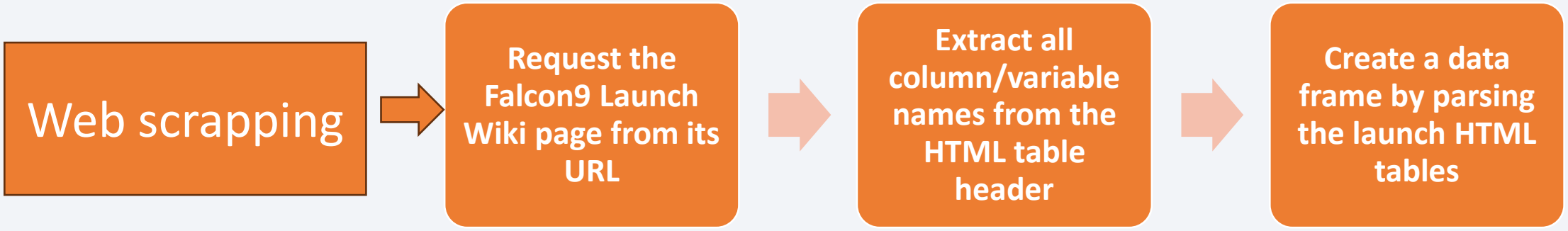
# Methodology

## Executive Summary

- Data collection methodology:
  - Data was collected from Space X and Wikipedia.
- Perform data wrangling
  - Data was cleaned, exploratory Data Analysis performed, and Training Labels defined.
- 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

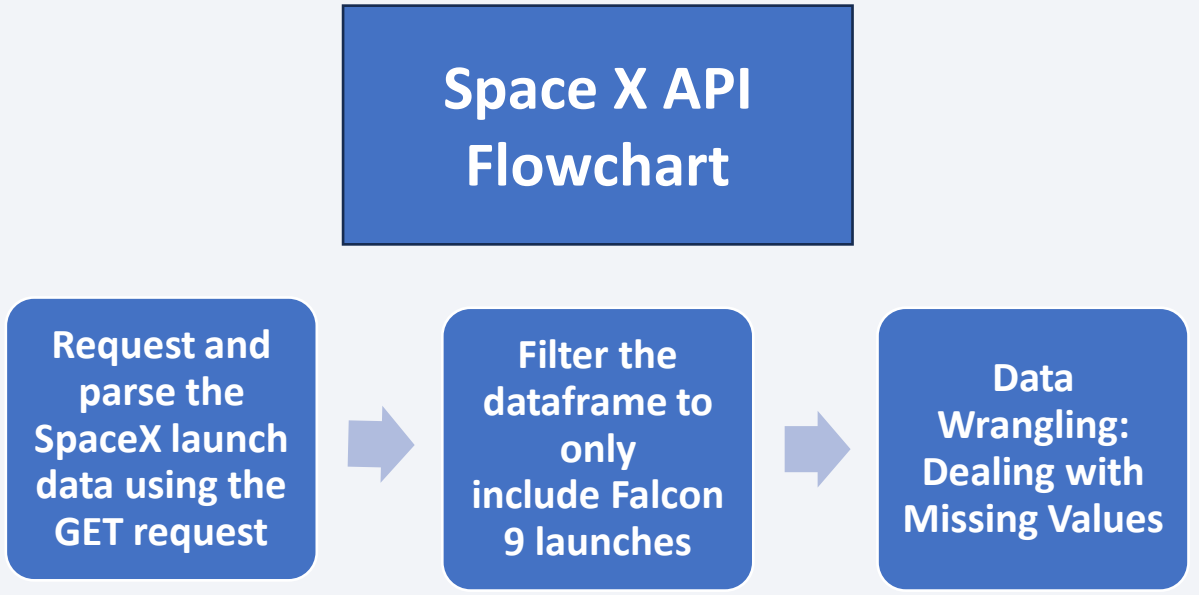
- Raw data was collected directly from Space X by using its own API.
- Furthermore, Falcon 9 launch records were extracted from an HTML table from Wikipedia.



# Data Collection – SpaceX API

- GitHub URL Link:

[IBM-Applied-Data-Science-Capstone/01.-jupyter-labs-spacex-data-collection-api.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)]

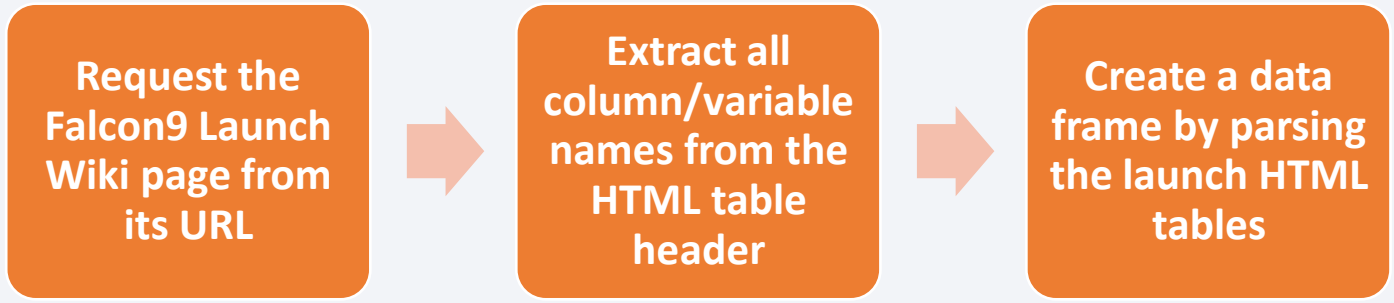


# Data Collection - Scraping

- GitHub URL Link:

[IBM-Applied-Data-Science-Capstone/02.-jupyter-labs-webscraping.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)]

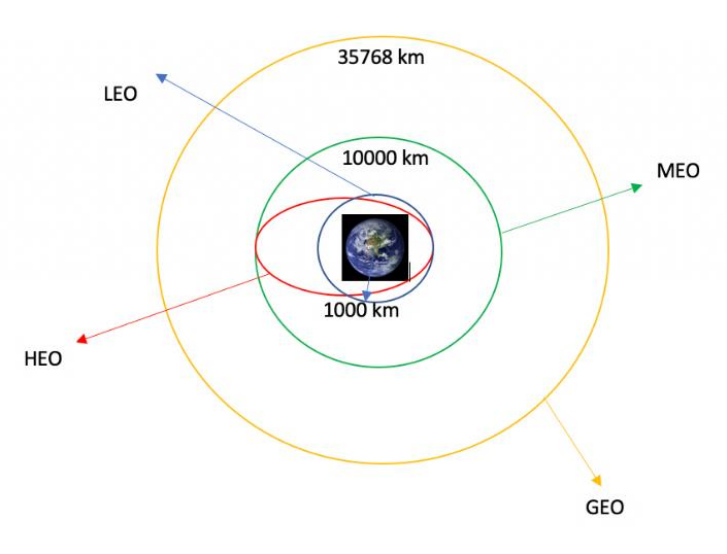
Wikipedia Web scrapping Flowchart



# Data Wrangling

- GitHub URL Link:

IBM-Applied-Data-Science-Capstone/03.-labs-jupyter-spacex-Data wrangling.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)



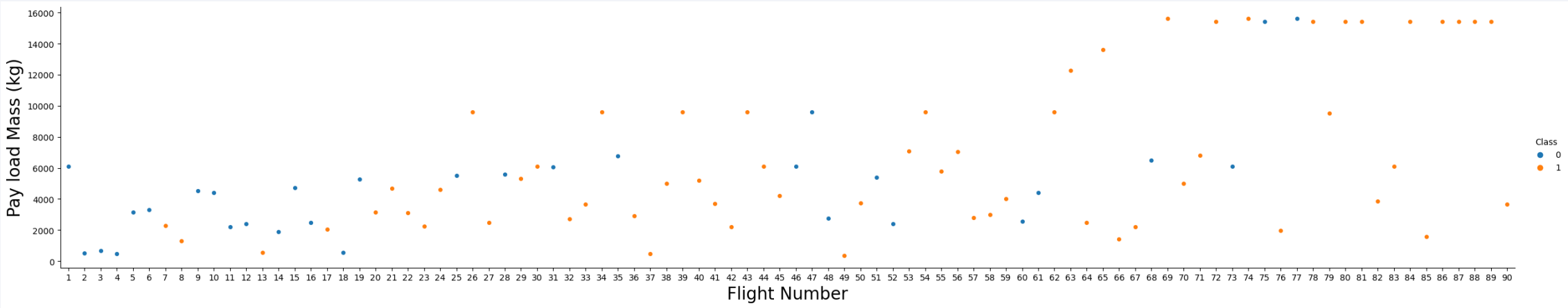

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome of the orbits
- Create a landing outcome label from Outcome column

# EDA with Data Visualization

- GitHub URL Link:

[IBM-Applied-Data-Science-Capstone/05.-jupyter-labs-eda-dataviz.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)]

| Visualizations |
| --- |
| Relationship between Flight Number and Launch Site |
| Relationship between Payload and Launch Site |
| Relationship between success rate of each orbit type |
| Relationship between Flight Number and Orbit type |
| Relationship between Payload and Orbit type |
| Launch success yearly trend |
| Features Engineering |



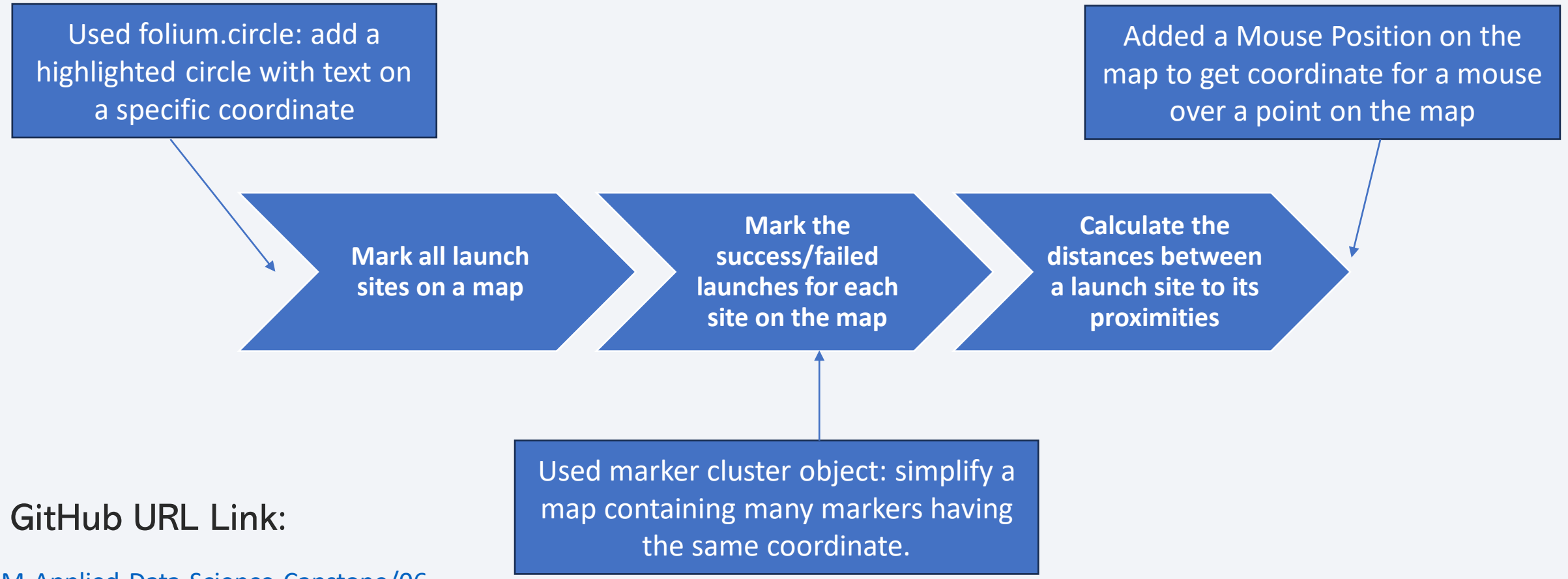
# EDA with SQL

- GitHub URL Link:

IBM-Applied-Data-Science-Capstone/04.-jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)

| SQL Queries |
| --- |
| 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 successful landing outcome in ground pad was achieved. |
| 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order. |

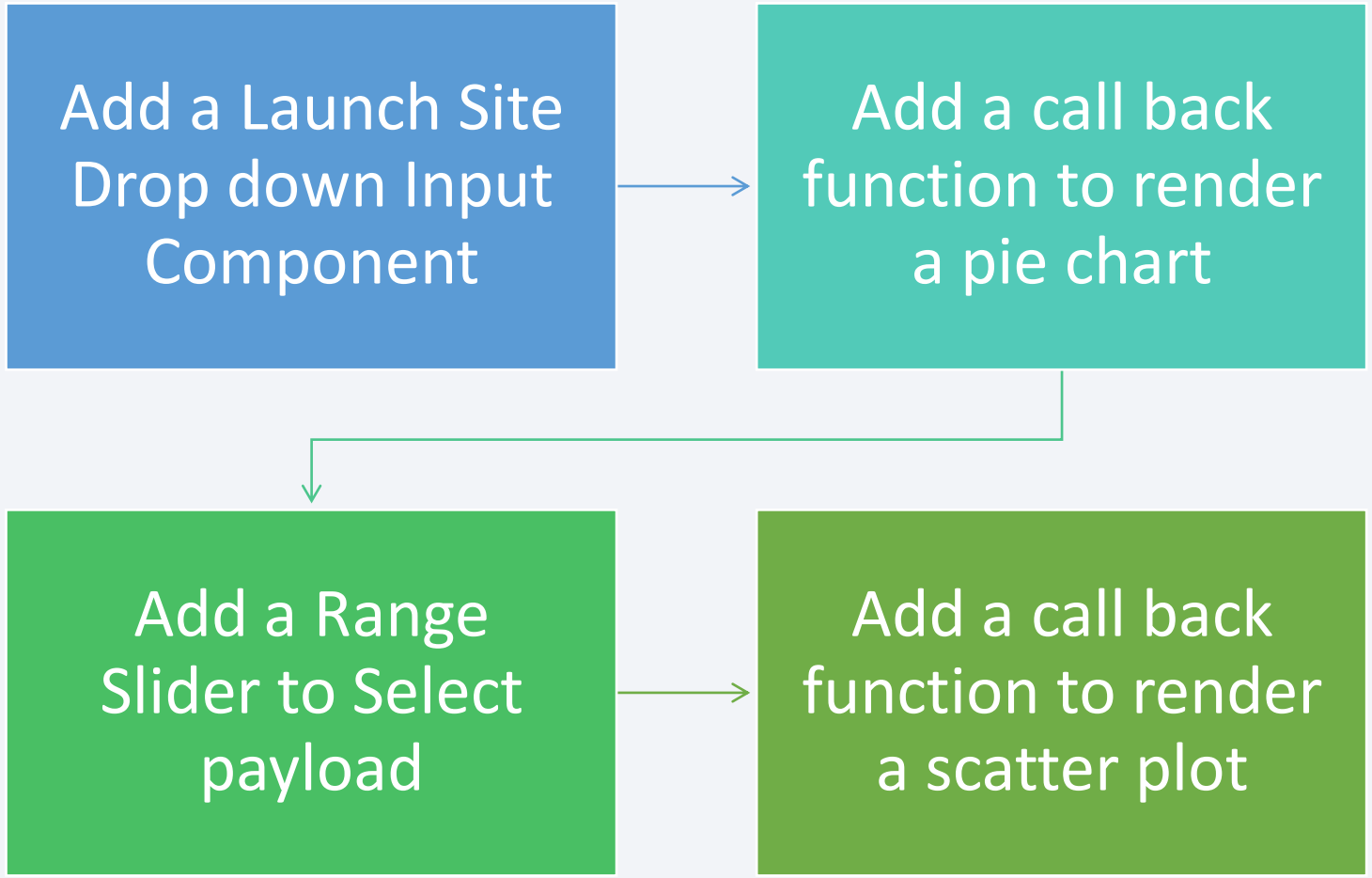
# Build an Interactive Map with Folium




- GitHub URL Link:

IBM-Applied-Data-Science-Capstone/06.-lab_jupyter_launch_site_location.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)

# Build a Dashboard with Plotly Dash

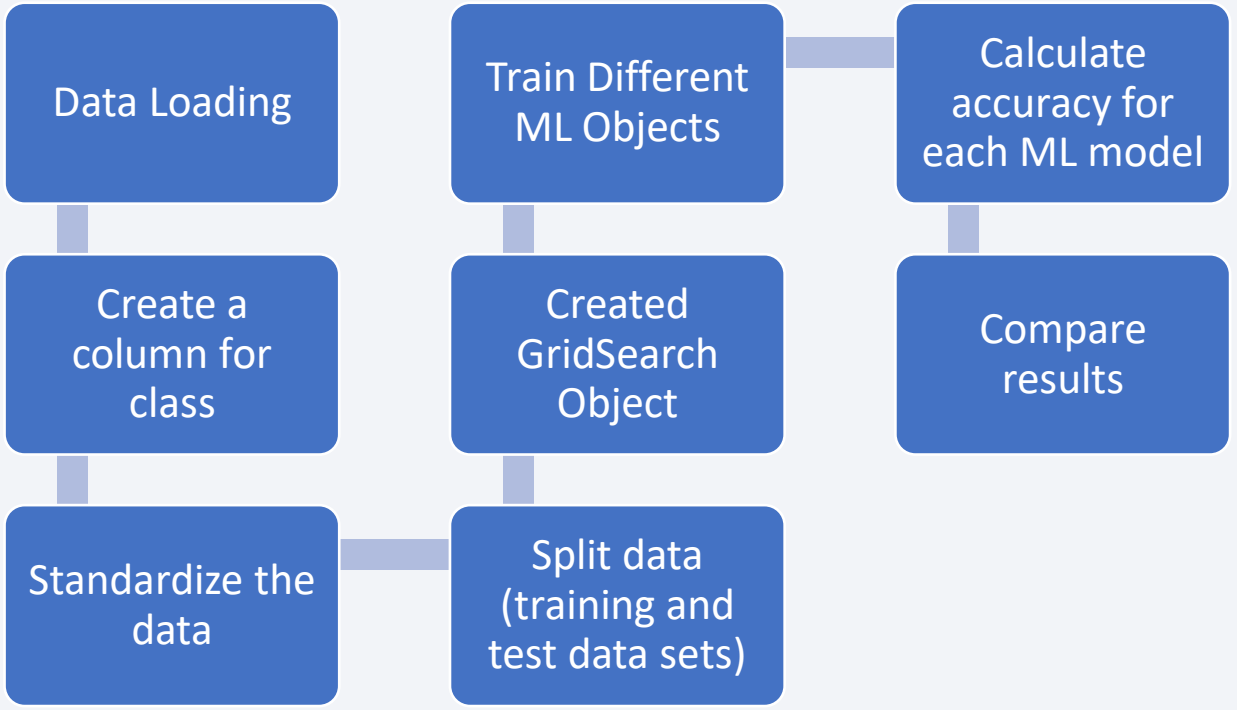



- GitHub URL Link:

IBM-Applied-Data-Science-Capstone/07.-spacex_dash_app.py at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)

# Predictive Analysis (Classification)

- Scikit learn library was used for predictive analysis.
- GridSearchCV to identify best parameters for models.

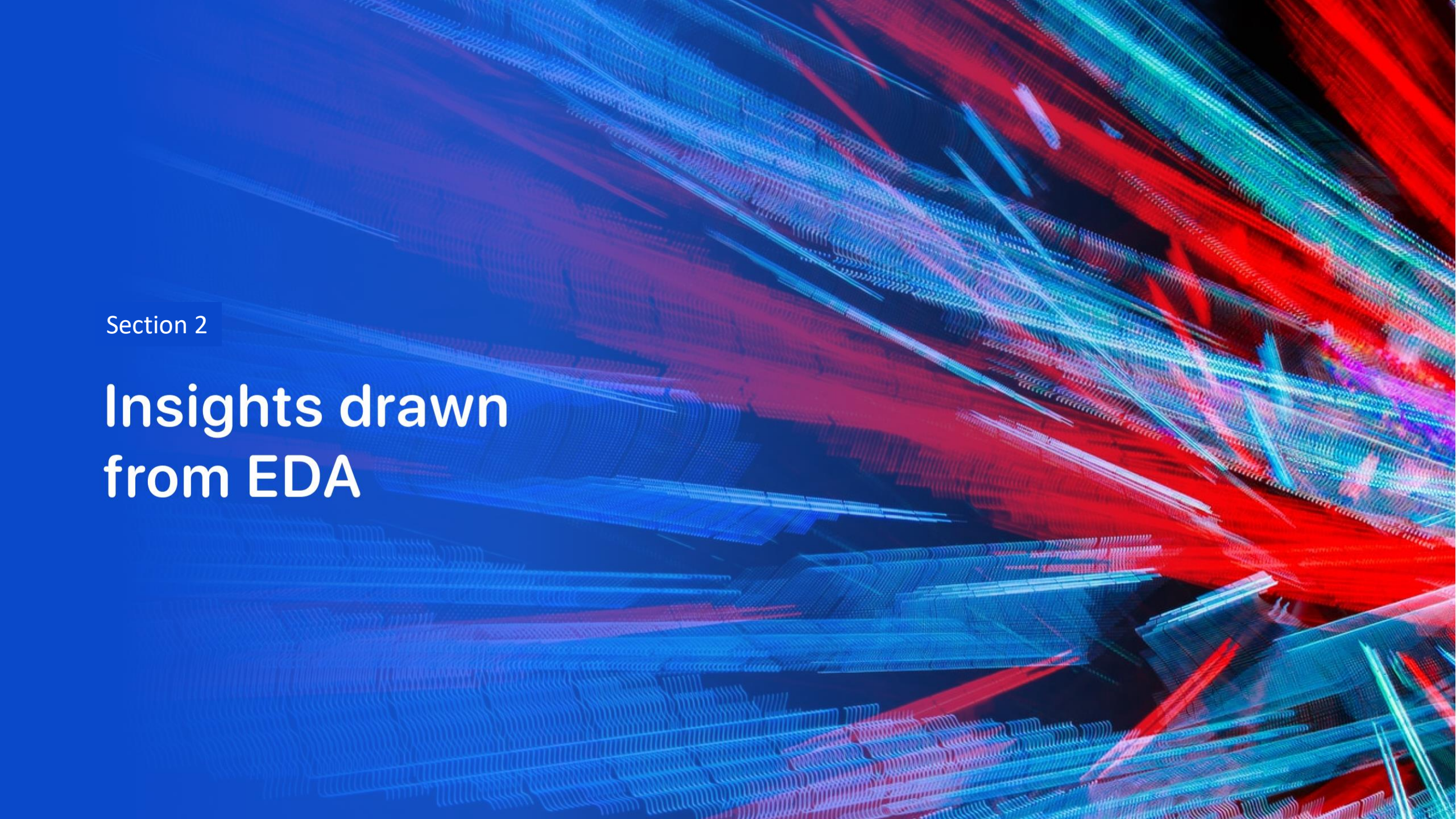



- GitHub URL Link:

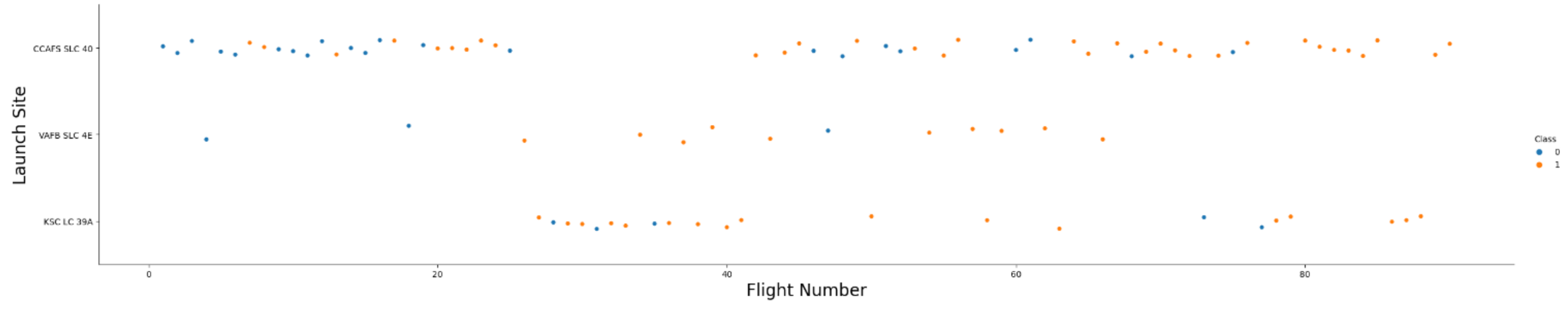
IBM-Applied-Data-Science-Capstone/08.-SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb at main · jesushidalgosanchez/IBM-Applied-Data-Science-Capstone (github.com)

# Results

- Machine learning models were able to predict landing success of rockets with an accuracy score of 83 %.
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.
- The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

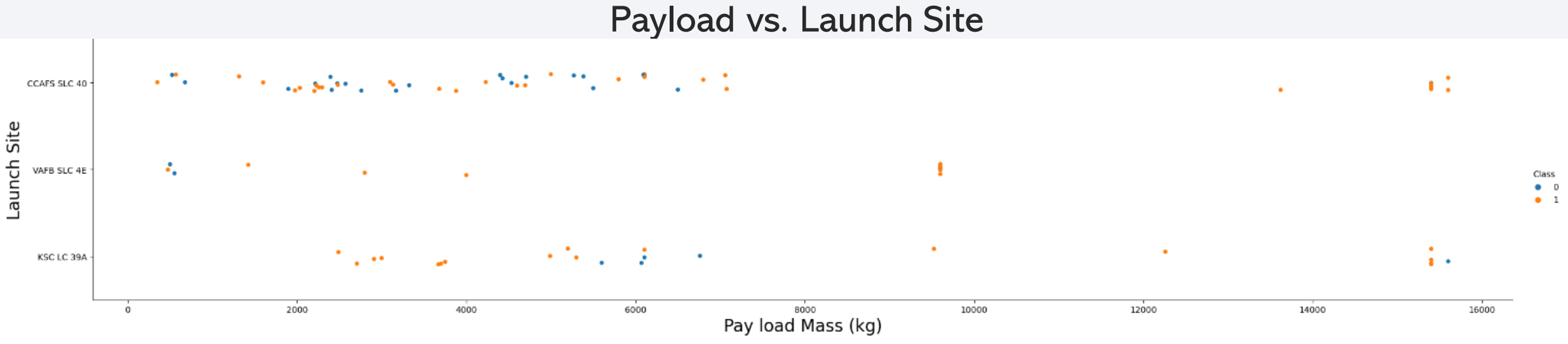

Section 2
Insights drawn from EDA

# Flight Number vs. Launch Site




- As flight number increases successful landings also increase.
- Most used Launch site is CCAFS SLC 40.

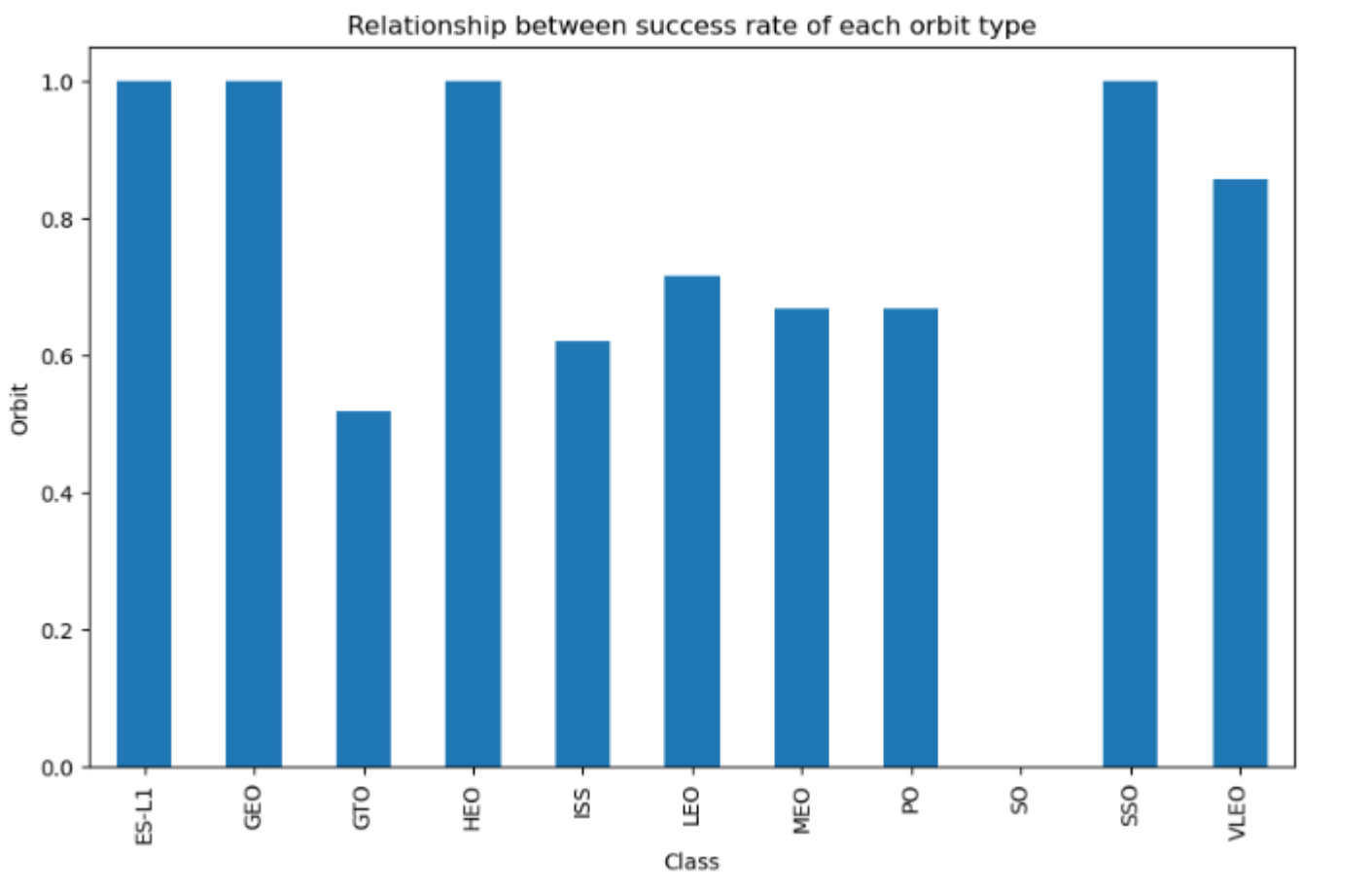
# Payload vs. Launch Site




- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- Most of heavy payload launches are successful.

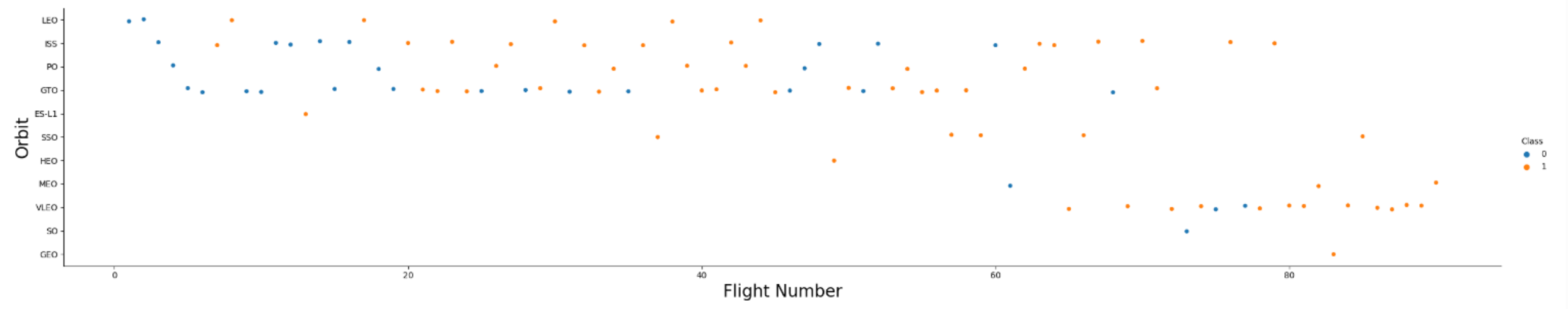
# Success Rate vs. Orbit Type

Success rate of each orbit type



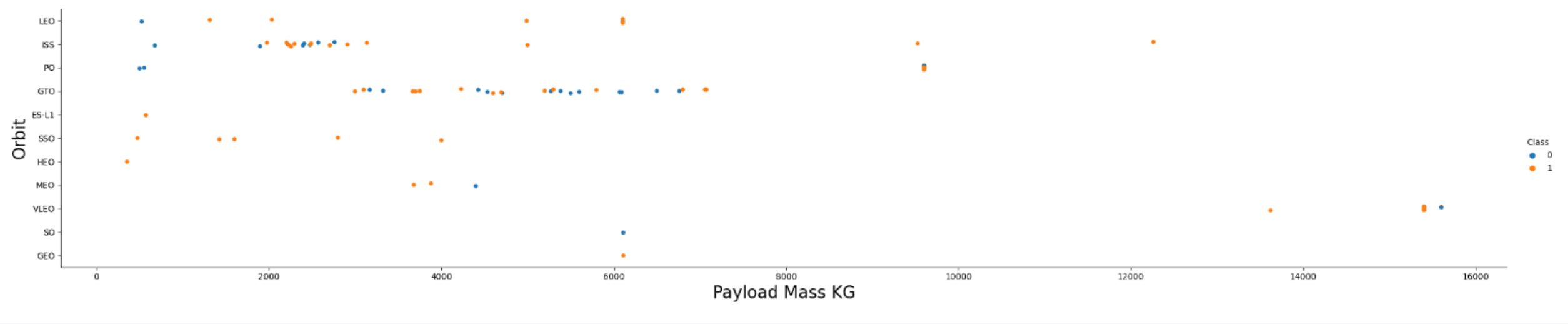

ES-L1, GEO, HEO and SSO orbits have a success rate of 100%

# Flight Number vs. Orbit Type




- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

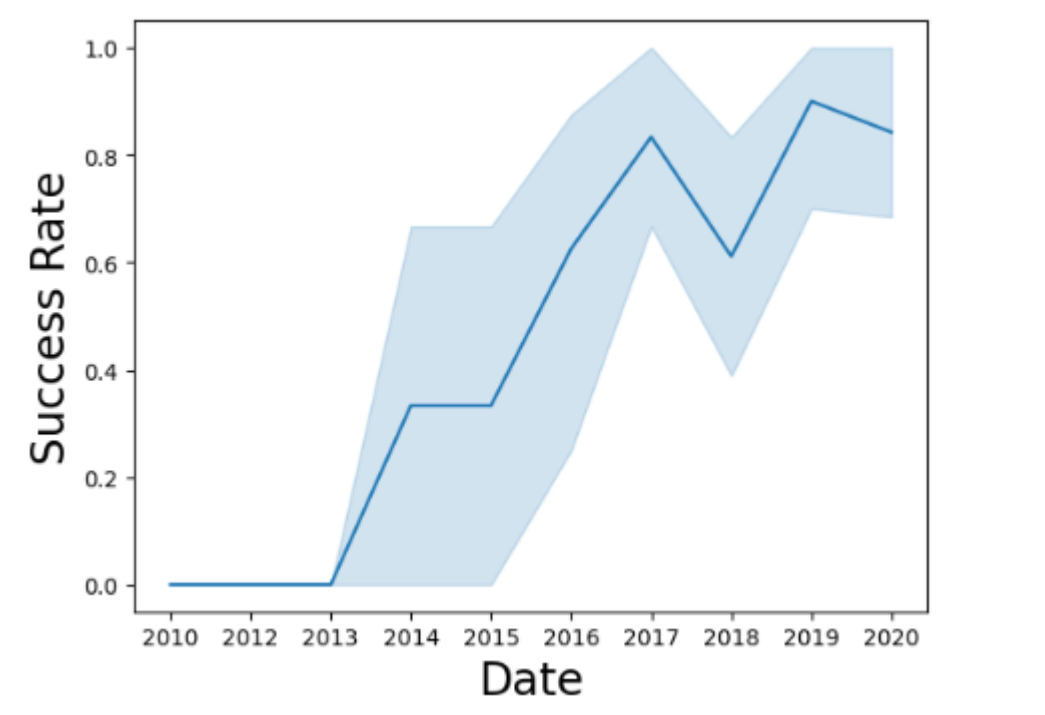
# Payload vs. Orbit Type




- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are both there here.

# Launch Success Yearly Trend

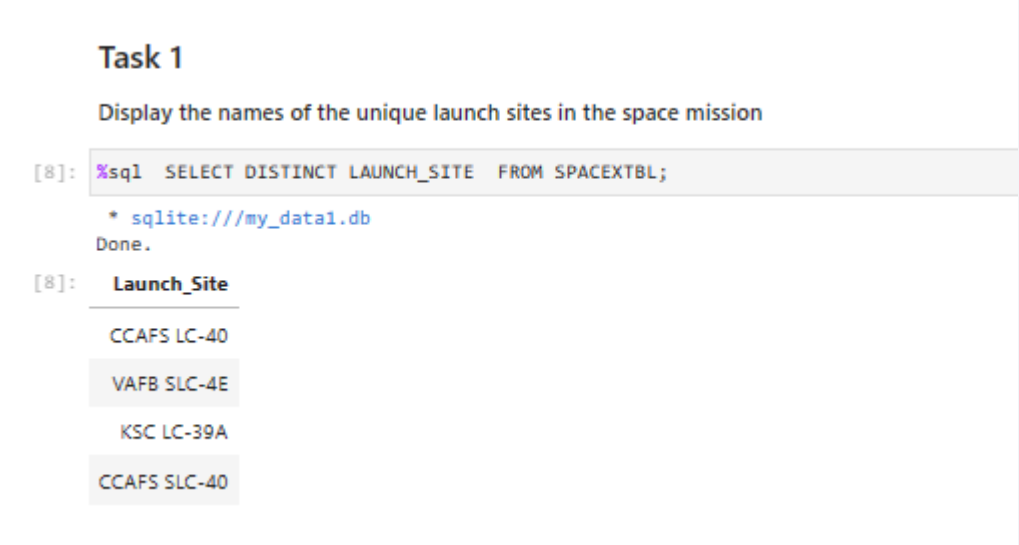
Line chart of yearly average success rate




Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

# All Launch Site Names

- Query result: DISTINCT operator was used.



Task 1

Display the names of the unique launch sites in the space mission

```
[8]: %sql  SELECT DISTINCT LAUNCH_SITE  FROM SPACEXTBL;
```

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

[8]:

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

# Launch Site Names Begin with 'CCA'

- Query result:

### Task 2

Display 5 records where launch sites begin with the string '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 |

# Total Payload Mass

- Query result:45596 kg

**Task 3**

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[13]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

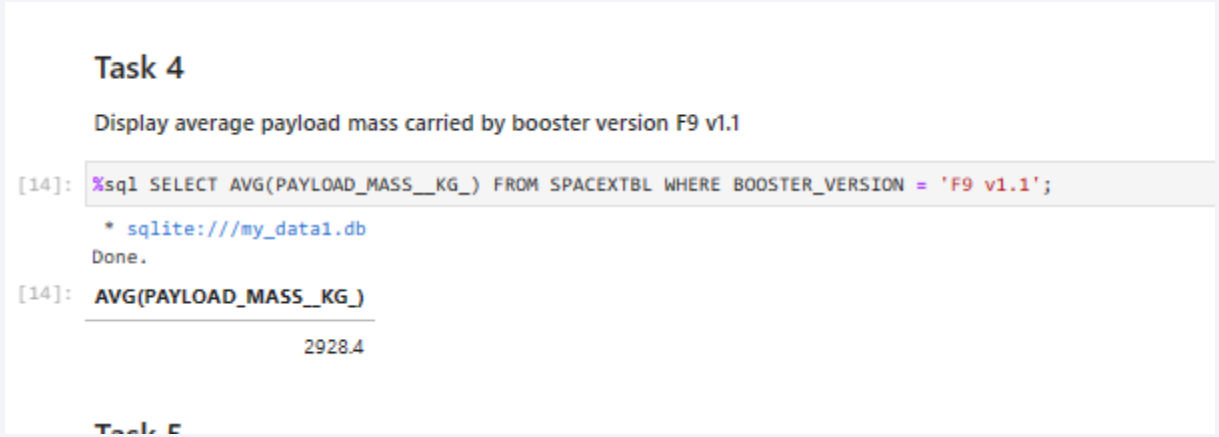
```
 * sqlite:///my_data1.db
Done.
```

[13]:

| SUM(PAYLOAD_MASS__KG_) |
|---|
| 45596 |

# Average Payload Mass by F9 v1.1

- Query result: average is 2928,4 Kg



Task 4

Display average payload mass carried by booster version F9 v1.1

```
[14]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

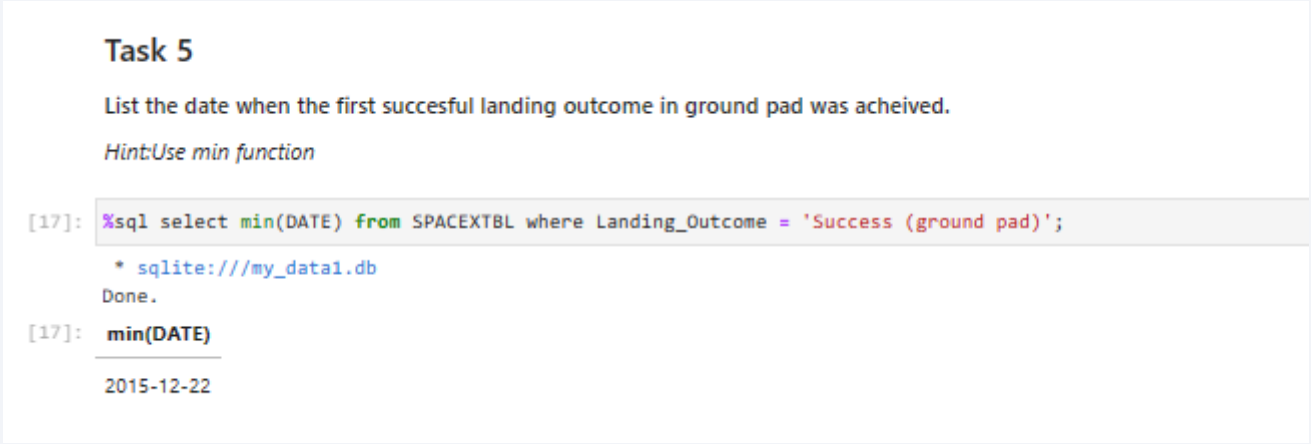
[14]:

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

Task 5

# First Successful Ground Landing Date

- Query result: 22nd December 2015



Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

*Hint:Use min function*

```
[17]: %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

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

[17]:

| min(DATE) |
|---|
| 2015-12-22 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- Query result:

### Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[19]:
```
%sql SELECT Booster_Version from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ < 6000;
```

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

[19]:

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

# Total Number of Successful and Failure Mission Outcomes

- Query result: 99 missions

### Task 7

List the total number of successful and failure mission outcomes

```
[20]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)';
```

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

[20]:

| count(MISSION_OUTCOME) |
|---|
| 99 |

# Boosters Carried Maximum Payload

- Query result: a subquery was used.

### Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
[21]: %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

[21]:

| 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

- Query result: 5 in total

### Task 9

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.

**Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.**

[24]:
```
%sql SELECT BOOSTER_VERSION,LAUNCH_SITE,LANDING_OUTCOME, DATE FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Failure (drone ship)' and DATE('2015');
```

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

[24]:

| Booster_Version | Launch_Site | Landing_Outcome | Date |
|---|---|---|---|
| F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) | 2015-01-10 |
| F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) | 2015-04-14 |
| F9 v1.1 B1017 | VAFB SLC-4E | Failure (drone ship) | 2016-01-17 |
| F9 FT B1020 | CCAFS LC-40 | Failure (drone ship) | 2016-03-04 |
| F9 FT B1024 | CCAFS LC-40 | Failure (drone ship) | 2016-06-15 |

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

- Query result:

**Task 10**

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
[33]: %sql select * from SPACEXTBL where Landing_Outcome = 'Success (ground pad)' or DATE between '2010-06-04' and '2017-03-20' order by date DESC;
```

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

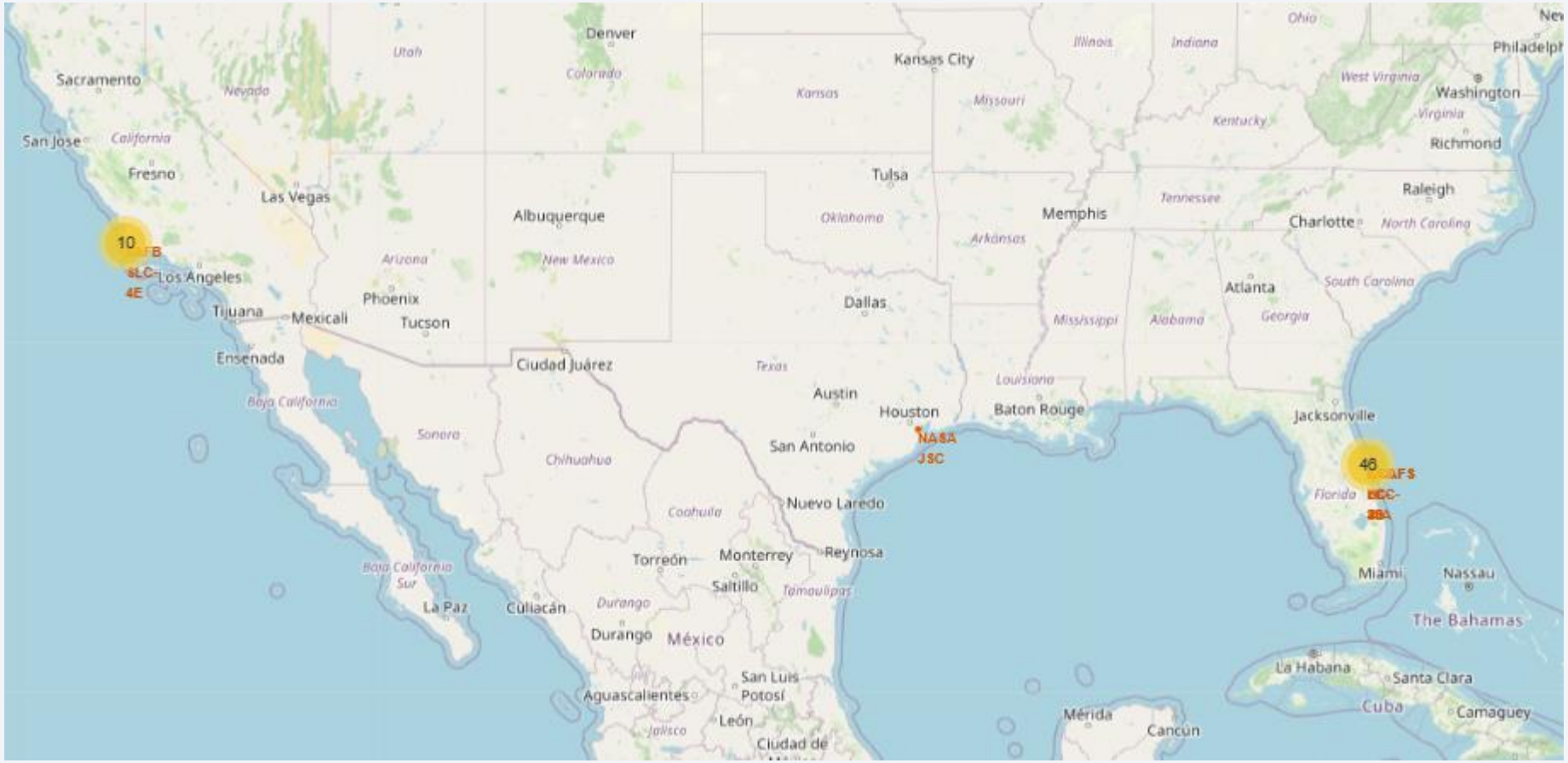
[33]:

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 2018-01-08 | 1:00:00 | F9 B4 B1043.1 | CCAFS SLC-40 | Zuma | 5000 | LEO | Northrop Grumman | Success (payload status unclear) | Success (ground pad) |
| 2017-12-15 | 15:36:00 | F9 FT B1035.2 | CCAFS SLC-40 | SpaceX CRS-13 | 2205 | LEO (ISS) | NASA (CRS) | Success | Success (ground pad) |
| 2017-09-07 | 14:00:00 | F9 B4 B1040.1 | KSC LC-39A | Boeing X-37B OTV-5 | 4990 | LEO | U.S. Air Force | Success | Success (ground pad) |
| 2017-08-14 | 16:31:00 | F9 B4 B1039.1 | KSC LC-39A | SpaceX CRS-12 | 3310 | LEO (ISS) | NASA (CRS) | Success | Success (ground pad) |
| 2017-06-03 | 21:07:00 | F9 FT B1035.1 | KSC LC-39A | SpaceX CRS-11 | 2708 | LEO (ISS) | NASA (CRS) | Success | Success (ground pad) |
| 2017-05-01 | 11:15:00 | F9 FT B1032.1 | KSC LC-39A | NROL-76 | 5300 | LEO | NRO | Success | Success (ground pad) |
| 2017-03-16 | 6:00:00 | F9 FT B1030 | KSC LC-39A | EchoStar 23 | 5600 | GTO | EchoStar | Success | No attempt |
| 2017-02-19 | 14:39:00 | F9 FT B1031.1 | KSC LC-39A | SpaceX CRS-10 | 2490 | LEO (ISS) | NASA (CRS) | Success | Success (ground pad) |
| 2017-01-14 | 17:54:00 | F9 FT B1029.1 | VAFB SLC-4E | Iridium NEXT 1 | 9600 | Polar LEO | Iridium Communications | Success | Success (drone ship) |
| 2016-08-14 | 5:26:00 | F9 FT B1026 | CCAFS LC-40 | JCSAT-16 | 4600 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 2016-07-18 | 4:45:00 | F9 FT B1025.1 | CCAFS LC-40 | SpaceX CRS-9 | 2257 | LEO (ISS) | NASA (CRS) | Success | Success (ground pad) |
| 2016-06-15 | 14:29:00 | F9 FT B1024 | CCAFS LC-40 | ABS-2A Eutelsat 117 West B | 3600 | GTO | ABS Eutelsat | Success | Failure (drone ship) |
| 2016-05-27 | 21:39:00 | F9 FT B1023.1 | CCAFS LC-40 | Thaicom 8 | 3100 | GTO | Thaicom | Success | Success (drone ship) |
| 2016-05-06 | 5:21:00 | F9 FT B1022 | CCAFS LC-40 | JCSAT-14 | 4696 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 2016-04-08 | 20:43:00 | F9 FT B1021.1 | CCAFS LC-40 | SpaceX CRS-8 | 3136 | LEO (ISS) | NASA (CRS) | Success | Success (drone ship) |
| 2016-03-04 | 23:35:00 | F9 FT B1020 | CCAFS LC-40 | SES-9 | 5271 | GTO | SES | Success | Failure (drone ship) |
| 2016-01-17 | 18:42:00 | F9 v1.1 B1017 | VAFB SLC-4E | Jason-3 | 553 | LEO | NASA (LSP) NOAA CNES | Success | Failure (drone ship) |
| 2015-12-22 | 1:29:00 | F9 FT B1019 | CCAFS LC-40 | OG2 Mission 2 11 Orbcomm-OG2 satellites | 2034 | LEO | Orbcomm | Success | Success (ground pad) |
| 2015-06-28 | 14:21:00 | F9 v1.1 B1018 | CCAFS LC-40 | SpaceX CRS-7 | 1952 | LEO (ISS) | NASA (CRS) | Failure (in flight) | Precluded (drone ship) |
| 2015-04-27 | 23:03:00 | F9 v1.1 B1016 | CCAFS LC-40 | Turkmen 52 / MonacoSAT | 4707 | GTO | Turkmenistan National Space Agency | Success | No attempt |
| 2015-04-14 | 20:10:00 | F9 v1.1 B1015 | CCAFS LC-40 | SpaceX CRS-6 | 1898 | LEO (ISS) | NASA (CRS) | Success | Failure (drone ship) |
| 2015-03-02 | 3:50:00 | F9 v1.1 B1014 | CCAFS LC-40 | ABS-3A Eutelsat 115 West B | 4159 | GTO | ABS Eutelsat | Success | No attempt |
| 2015-02-11 | 23:03:00 | F9 v1.1 B1013 | CCAFS LC-40 | DSCOVR | 570 | HEO | U.S. Air Force NASA NOAA | Success | Controlled (ocean) |
| 2015-01-10 | 9:47:00 | F9 v1.1 B1012 | CCAFS LC-40 | SpaceX CRS-5 | 2395 | LEO (ISS) | NASA (CRS) | Success | Failure (drone ship) |
| 2014-09-21 | 5:52:00 | F9 v1.1 B1010 | CCAFS LC-40 | SpaceX CRS-4 | 2216 | LEO (ISS) | NASA (CRS) | Success | Uncontrolled (ocean) |
| 2014-09-07 | 5:00:00 | F9 v1.1 B1011 | CCAFS LC-40 | AsiaSat 6 | 4428 | GTO | AsiaSat | Success | No attempt |
| 2014-08-05 | 8:00:00 | F9 v1.1 | CCAFS LC-40 | AsiaSat 8 | 4535 | GTO | AsiaSat | Success | No attempt |
| 2014-07-14 | 15:15:00 | F9 v1.1 | CCAFS LC-40 | OG2 Mission 1 6 Orbcomm-OG2 satellites | 1316 | LEO | Orbcomm | Success | Controlled (ocean) |


Section 3
Launch Sites
Proximities Analysis

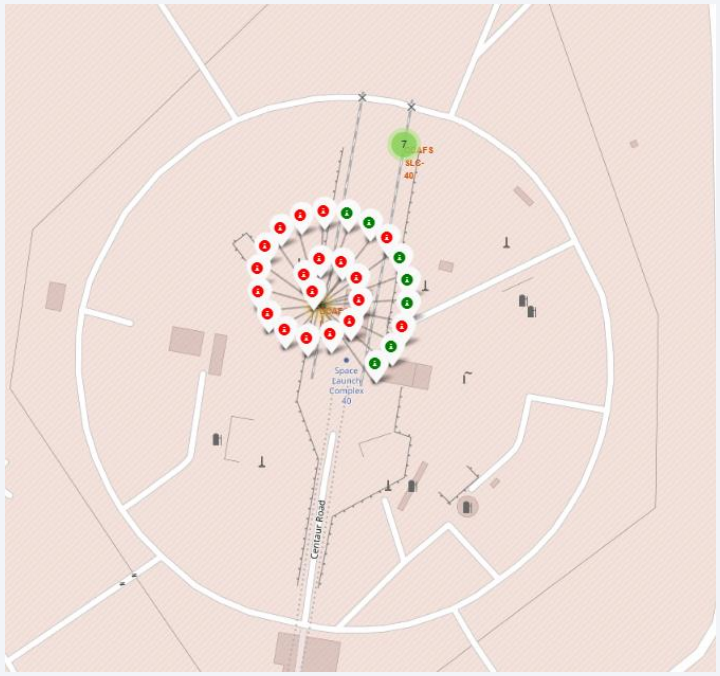
# Folium Map: Launch Site Locations

- Launch sites are located at both east and west coasts.
- Nasa Launch site is located at the gulf of Mexico



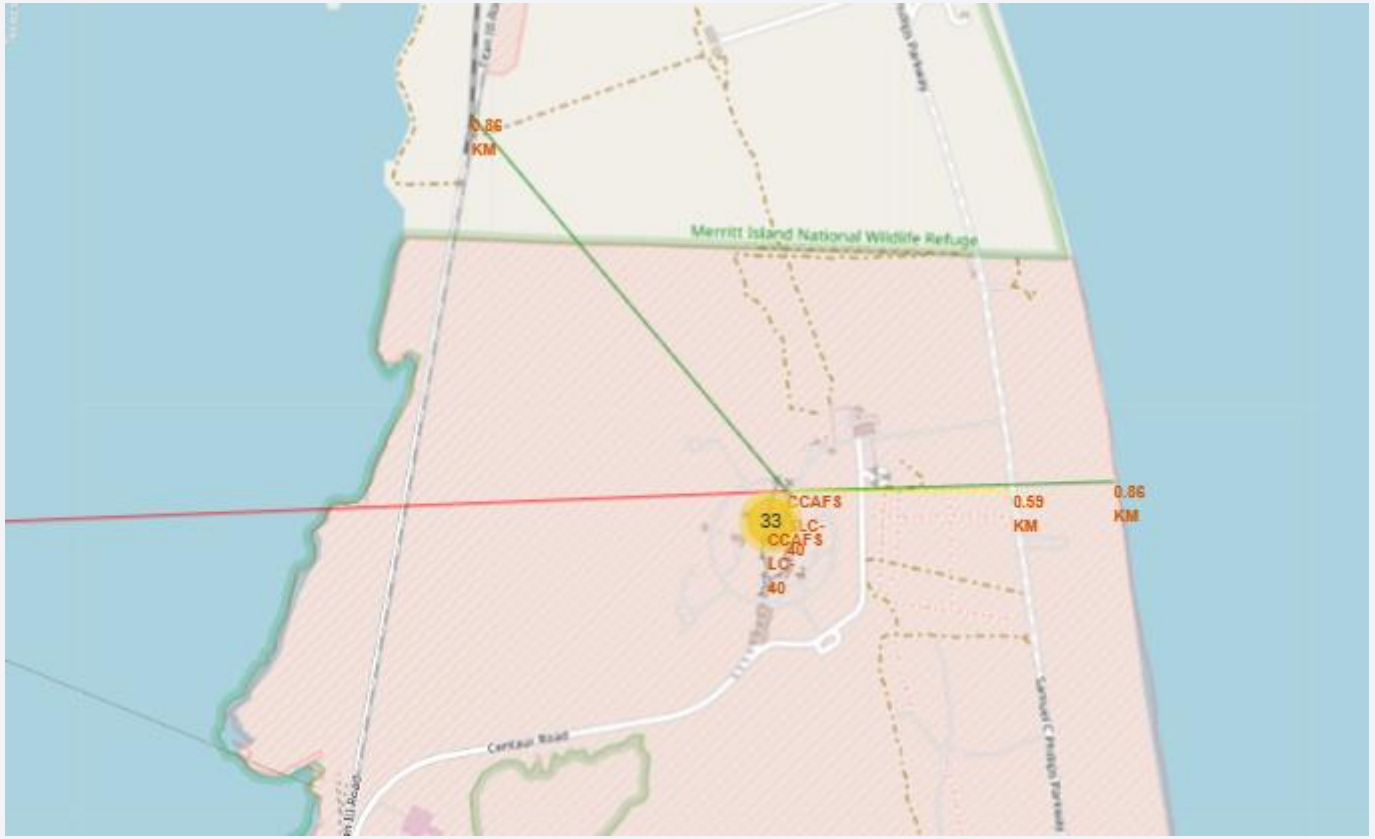
# Folium Map: Success Rate of Rocket Launches

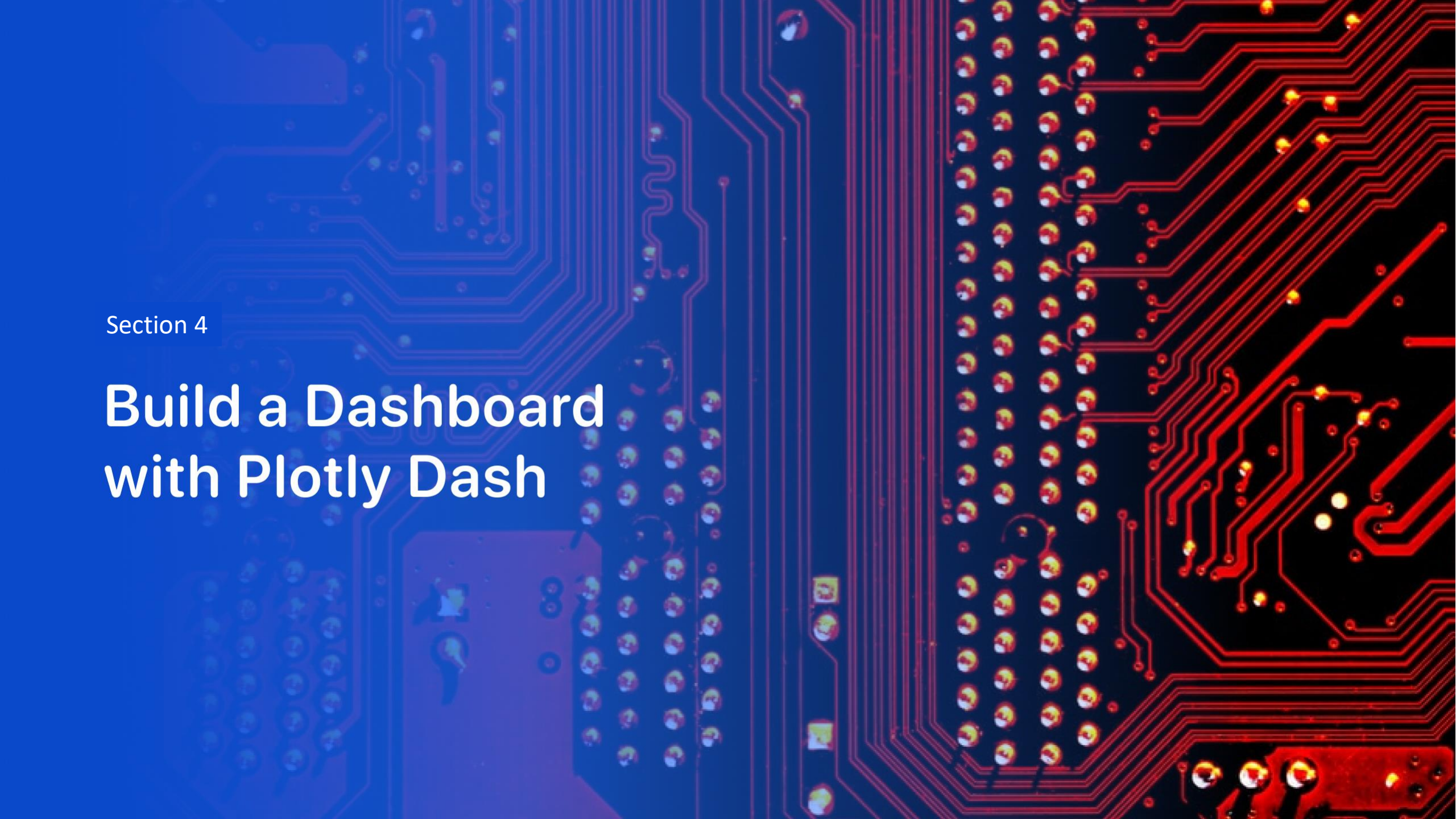
- Green markers represent the successful launches and red markers represents failed launches.

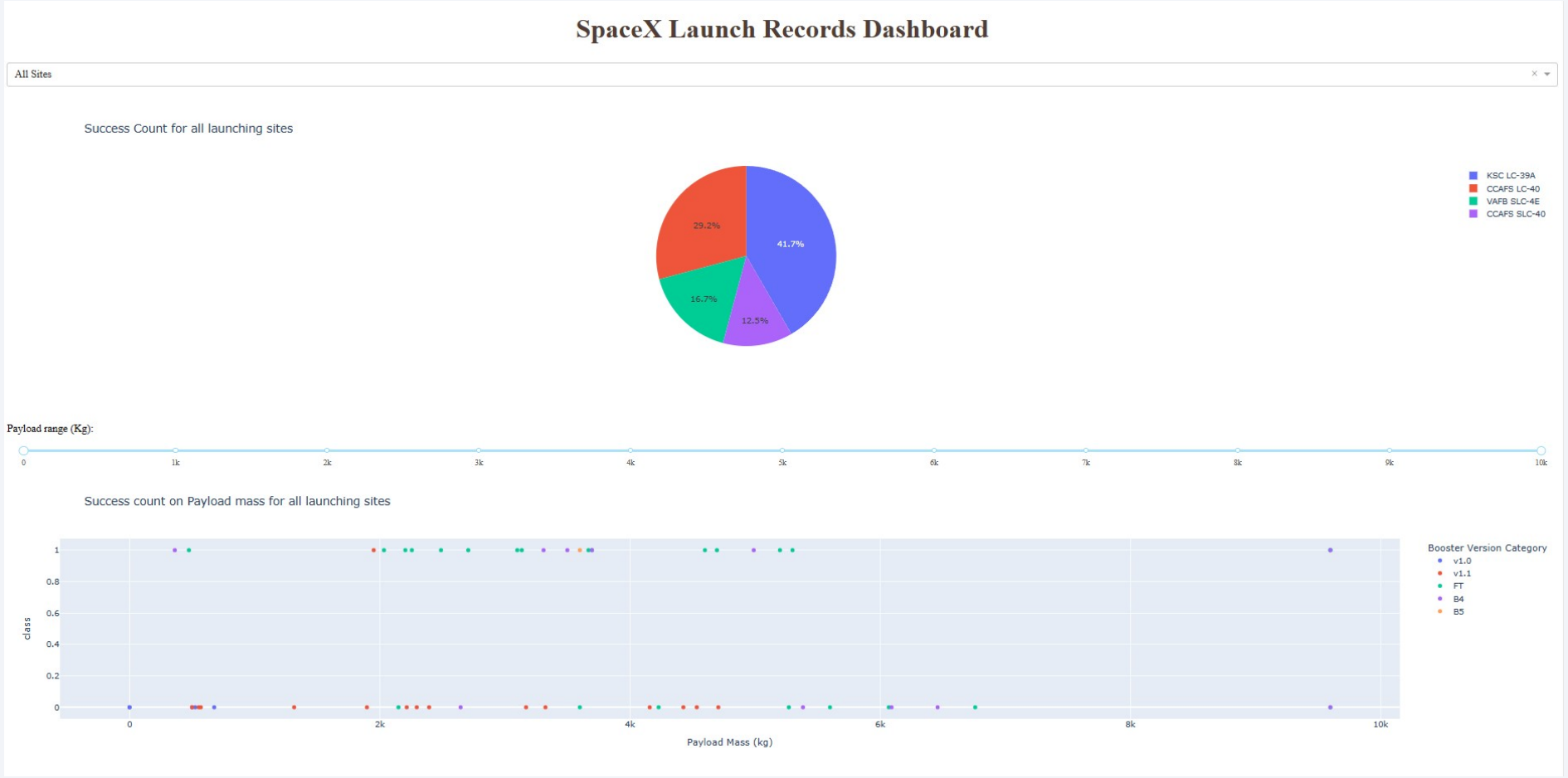

# Folium Map: Surrounding Landmarks

- Several distances were calculated CCAFS LC 40 launch site.



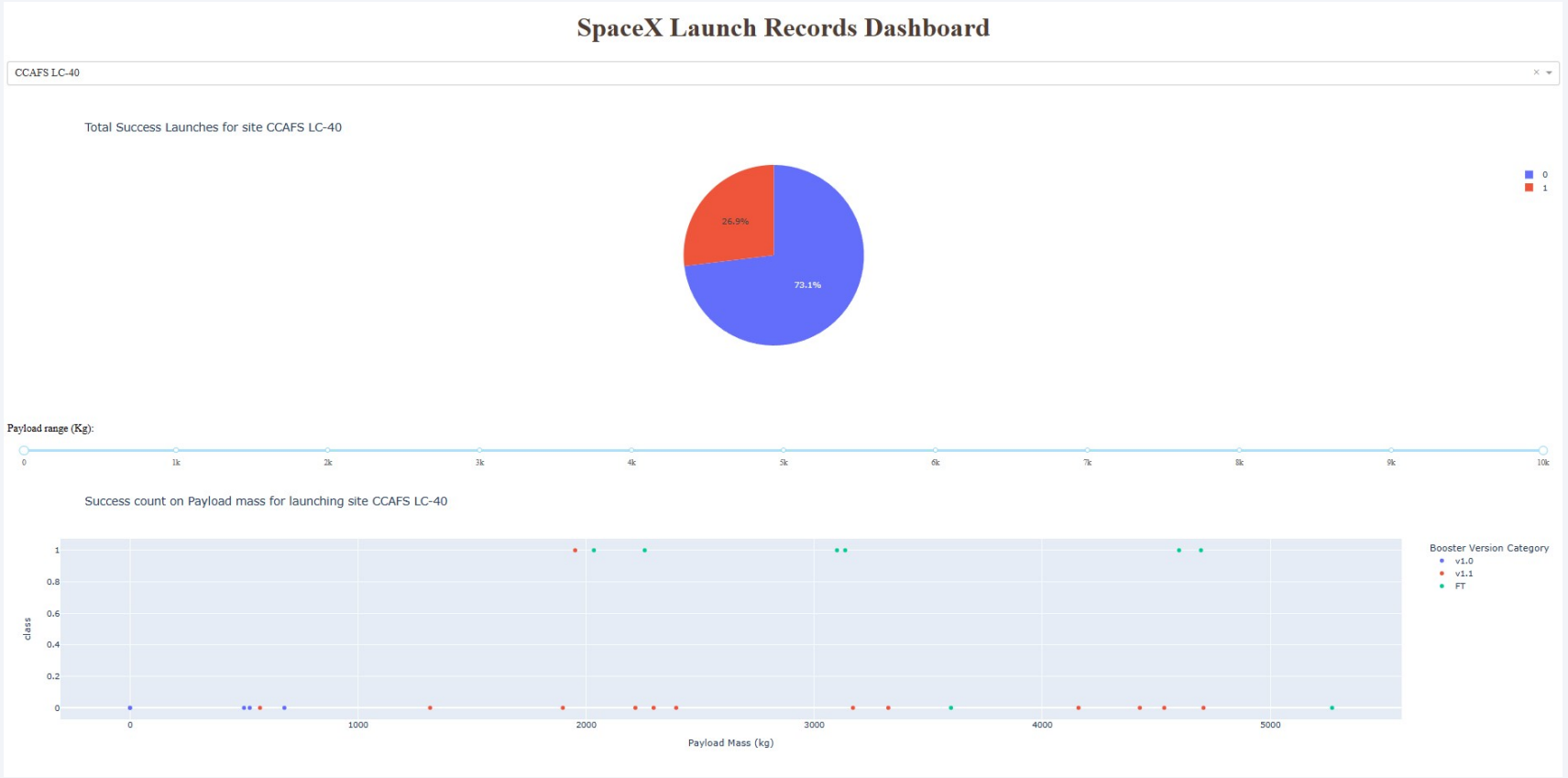

Section 4
Build a Dashboard with Plotly Dash

# Full Dashboard




- Information for all launch sites is showed.

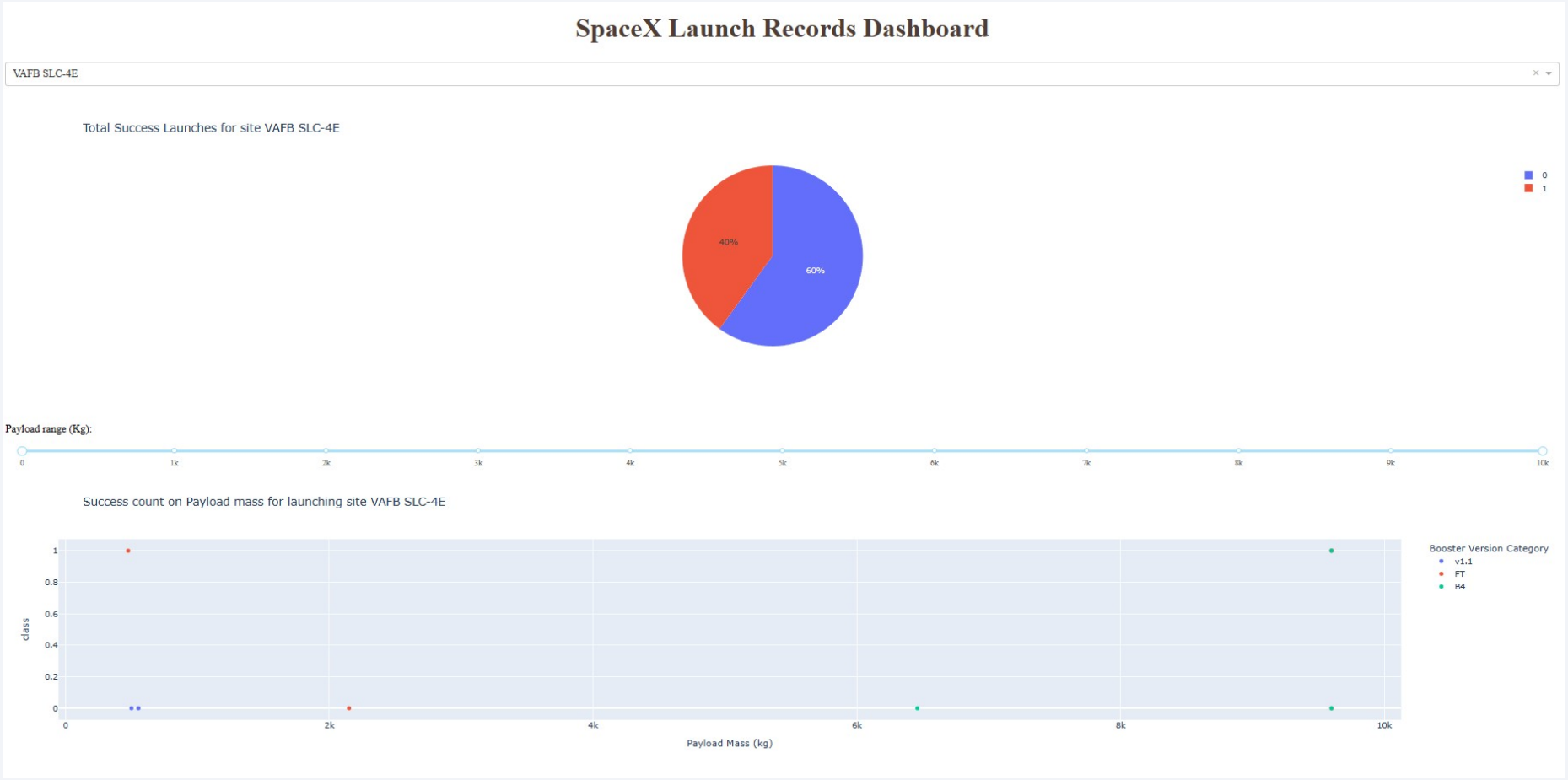
# CCAFS LC-40




- CCAFS LC 40 Pie chart (successful and unsuccessful launches) and payload scatter plot

# VAFB SLC-4E

- VAFB SLC-4E launch site, payload mass is also filtered



Section 5

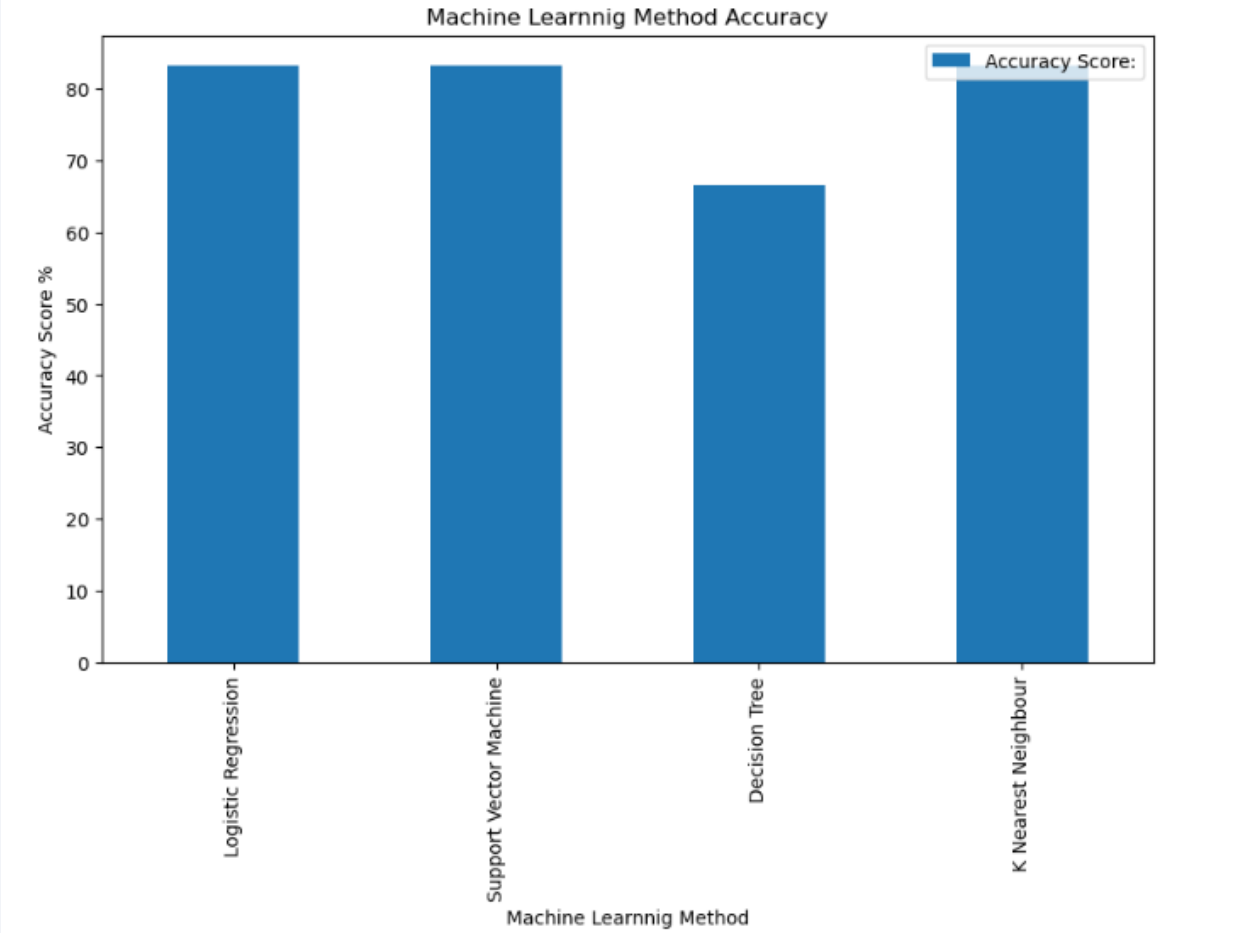
# Predictive Analysis (Classification)

# Classification Accuracy

- Logistic Regression, Support Vector Machine and KNN models have the biggest accuracy (83%).

[35]:

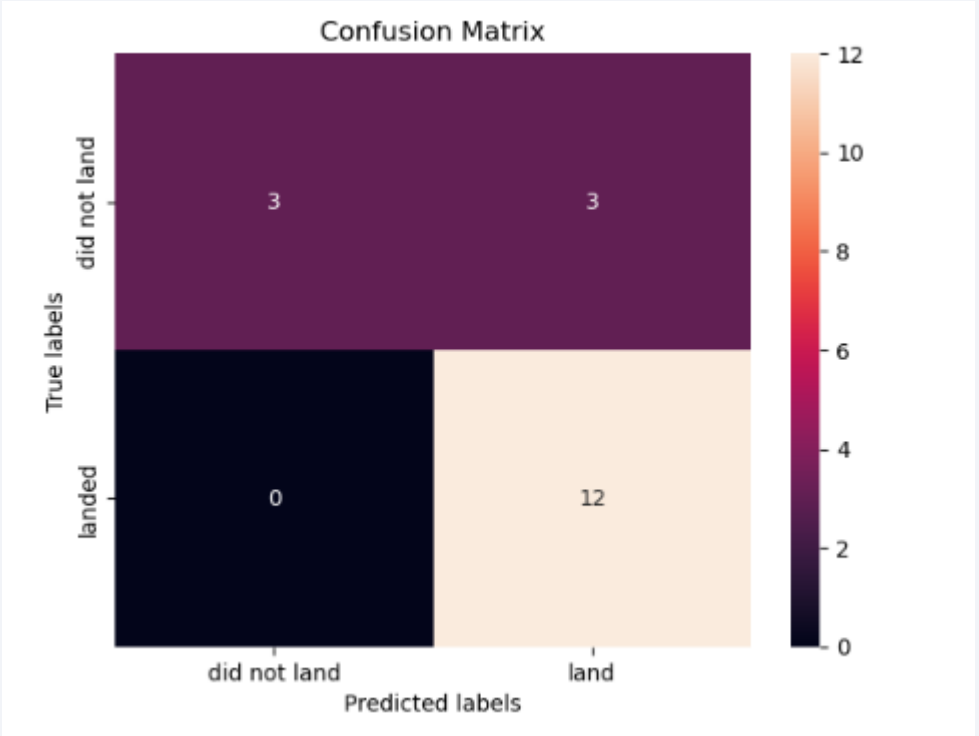
| Machine Learnnig Method | Accuracy Score: |
|---|---|
| Logistic Regression | 83.333333 |
| Support Vector Machine | 83.333333 |
| Decision Tree | 66.666667 |
| K Nearest Neighbour | 83.333333 |



# Confusion Matrix

- Confusion matrix of the best performing model corresponds to Logistic Regression, Support Vector Machine and KNN models .

- Al landed rockets were properly classified (12) while for the rockets that did not land 3 were properly classified and 3 failed in the classification.



# Conclusions

- As flight number increases successful landings also increase.
- Most used Launch site is CCAFS SLC 40.
- ES-L1, GEO, HEO and SSO orbits have a success rate of 100%
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- Most of heavy payload launches are successful.
- Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.
- Launch sites are located at both east and west coasts.
- Logistic Regression, Support Vector Machine and KNN models have the biggest accuracy (83%).

# Thank you!