IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shibil PM
27-03-2023

Outline

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

Executive Summary

- Summary of methodologies

  Collecting data through API

  Collecting data through web scraping

  Data wrangling

  Exploratory analysis using SQL

  Exploratory analysis using visualization

  Interactive visual analytics with folium and dash

  Machine learning prediction

- Summary of all results

  Exploratory analysis result

  Predictive analytics result

Introduction

- Project background and context

  To evaluate Space X, for Space Y to compete.

- Problems you want to find answers

  What factors determine if the rocket will land successfully?

  The best way to estimate the total cost for launches?

  Where is the best place to make lauches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Scrapping (Wikipedia)
- Perform data wrangling
  - Collected data and cleaned properly through One-hot-encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data normalized and split to training and test sets. Using these sets evaluated accuracy and score for four different models

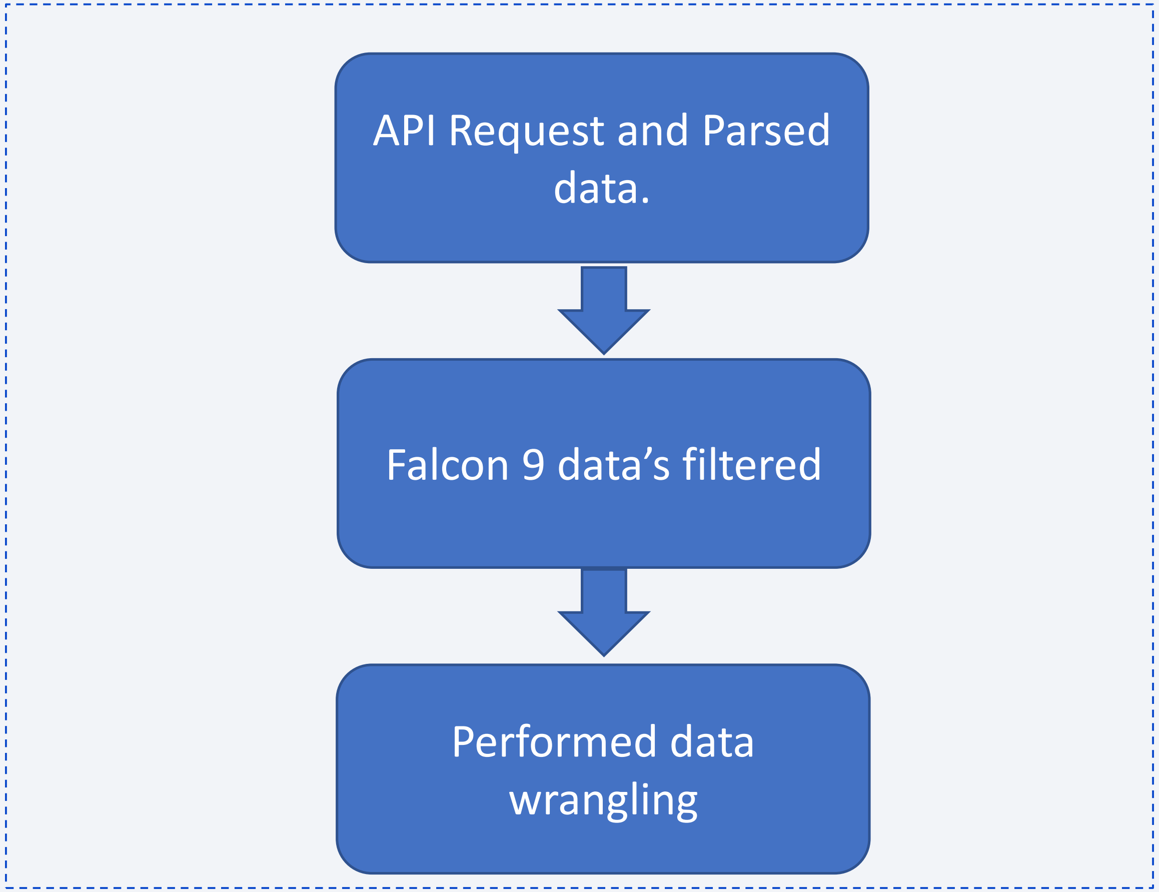
Data Collection

- Data collected from Space X API .
- Data collected using Web Scrapping Wikipedia.

Data Collection - SpaceX API

- Data collected through get request and it was in json format after it stored in to a pandas data frame using .json_normalize. And performed data wrangling.

- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb



API Request and Parsed data.

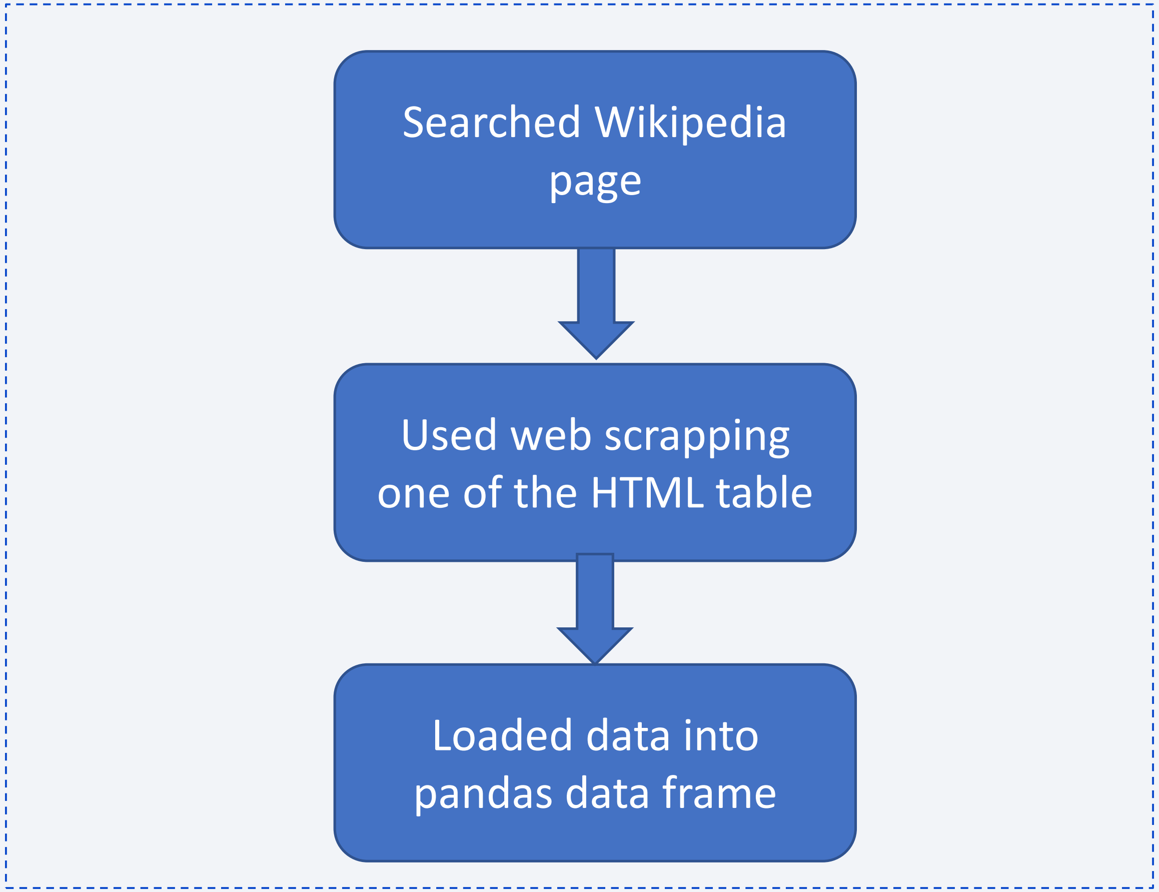
Falcon 9 data's filtered

Performed data wrangling

Data Collection - Scraping

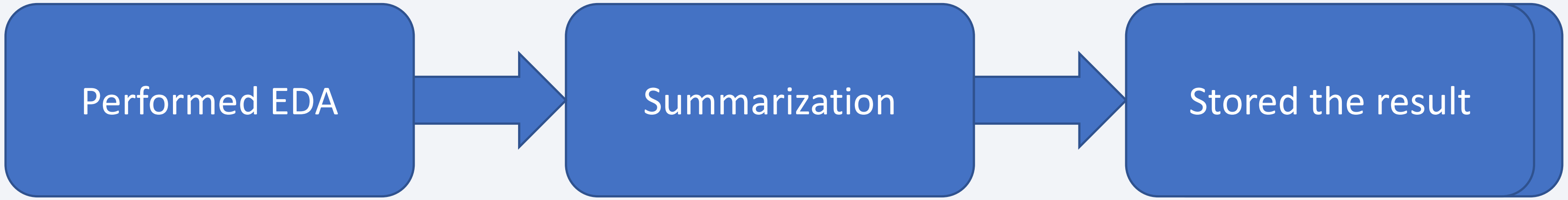
- Data collected through web scrapping. Utilized one of the Wikipedia pages and data scrapped from one of the html table.

- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb


Searched Wikipedia page
Used web scrapping one of the HTML table
Loaded data into pandas data frame

Data Wrangling

- Performed exploratory data analysis and determined the training labels.
- Calculated the number of launches at each site, and the number and occurrence of each orbits
- Created landing outcome label from outcome column and exported the results to csv.
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/spacex-Data-20wrangling.ipynb



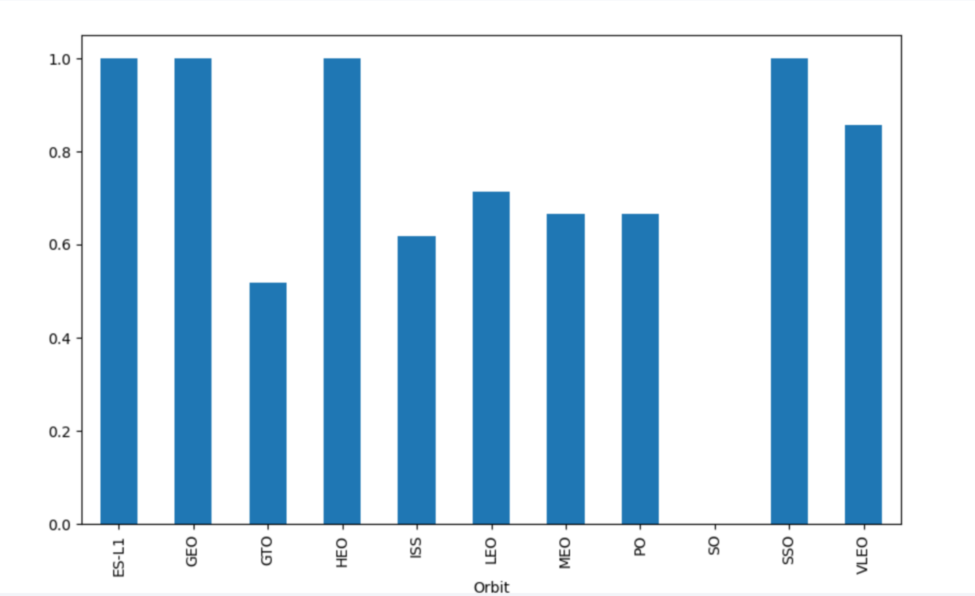
Performed EDA

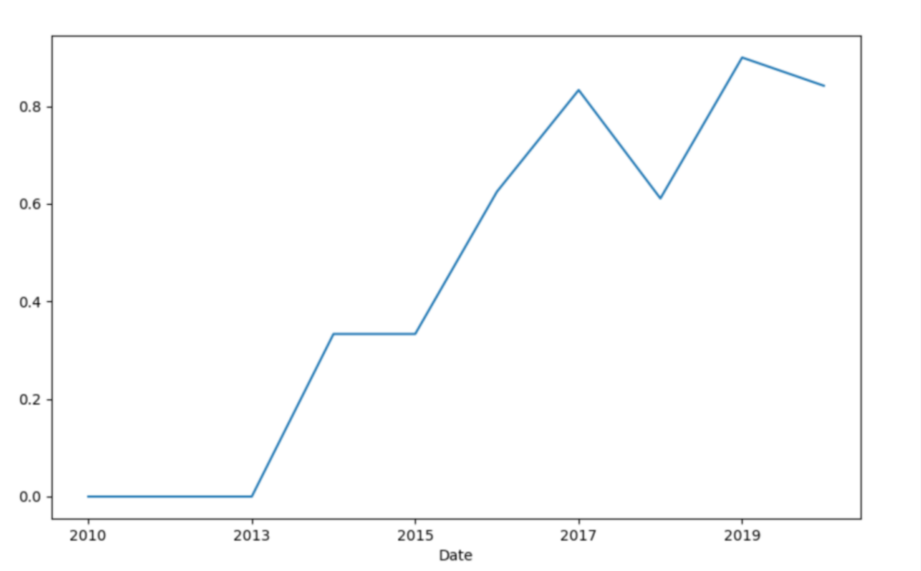
Summarization

Stored the result

EDA with Data Visualization

- Explored the data by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly Add the
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.ipynb





EDA with SQL

- Names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- The total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first successful landing outcome in ground pad was acheived.
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster_versions which have carried the maximum payload mass. Use a subquery
- The failed landing_outcomes in drone ship, their booster versions, and launch site names for 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
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/sql-coursera%20(1).ipynb

Build an Interactive Map with Folium

- Marked all launch sites, and added map objects such as markers, circles, lines to mark the success or failure of launches for each site on the folium map.
- Assigned the feature launch outcomes (failure or success) to class 0 and 1.
- Using marker clusters, we identified which launch sites have relatively high success rate.
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Graphs and plots used:

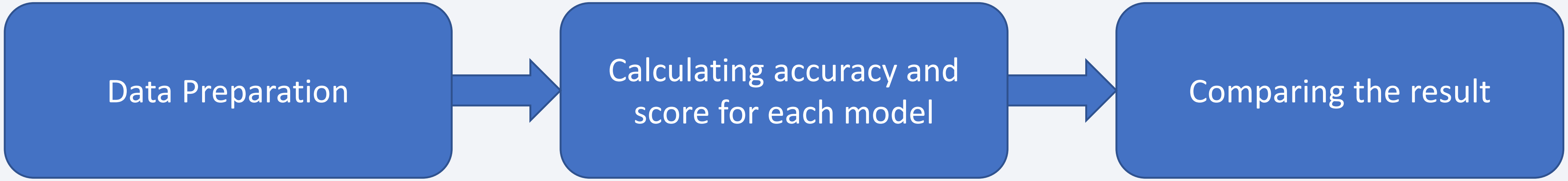
  Percentage of launches by site (Pie Chart)

  Payload range(scatter plot)

- This combination allowed to quickly analyze the relation between payload and launch sites
- It helped to to find where is the best place to launch according to payloads
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/dashboard.py

Predictive Analysis (Classification)

- Loaded the data using numpy and pandas, transformed the data, split our data into training and testing.
- Built different machine learning models and tune different hyperparameters using GridSearchCV.
- Found the best performing classification model.
- Github: https://github.com/ShibilPM/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.ipynb



Data Preparation

Calculating accuracy and score for each model

Comparing the result

Results

- Exploratory data analysis results:

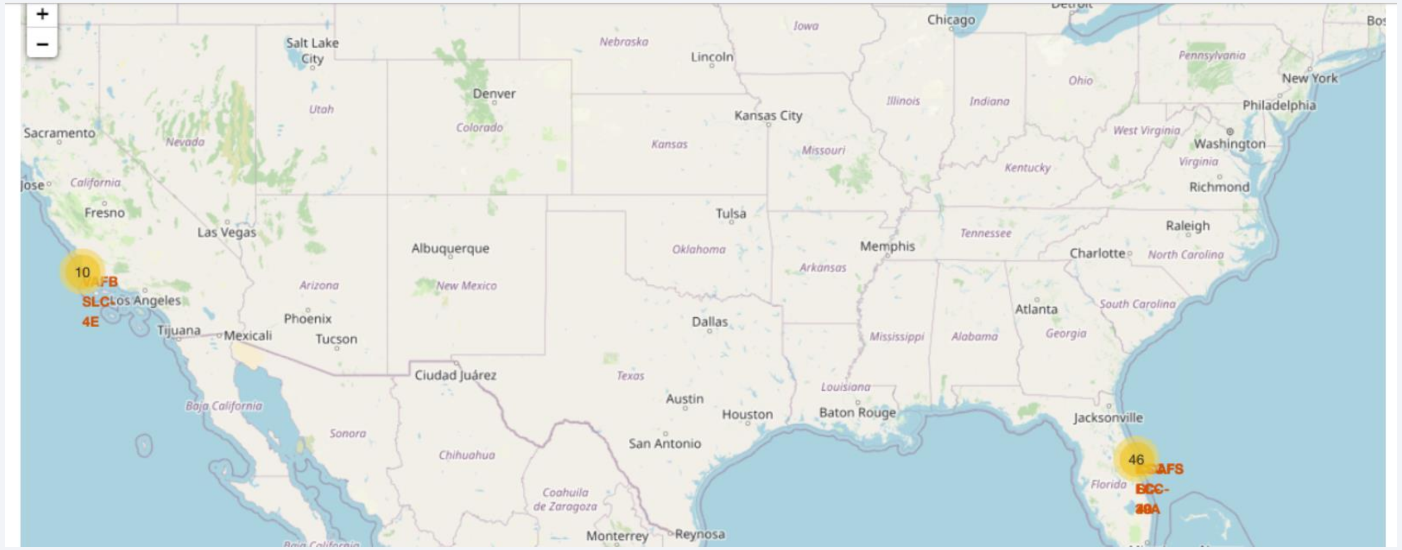
  -SpaceX uses 4 different launch sites.

  -Two booster versions failed at landing drone ships in 2015.

  -Many Falcon 9 booster version were successful at landing in drone ships having payload above average.

  -The number of landing outcomes became as better as years passed.

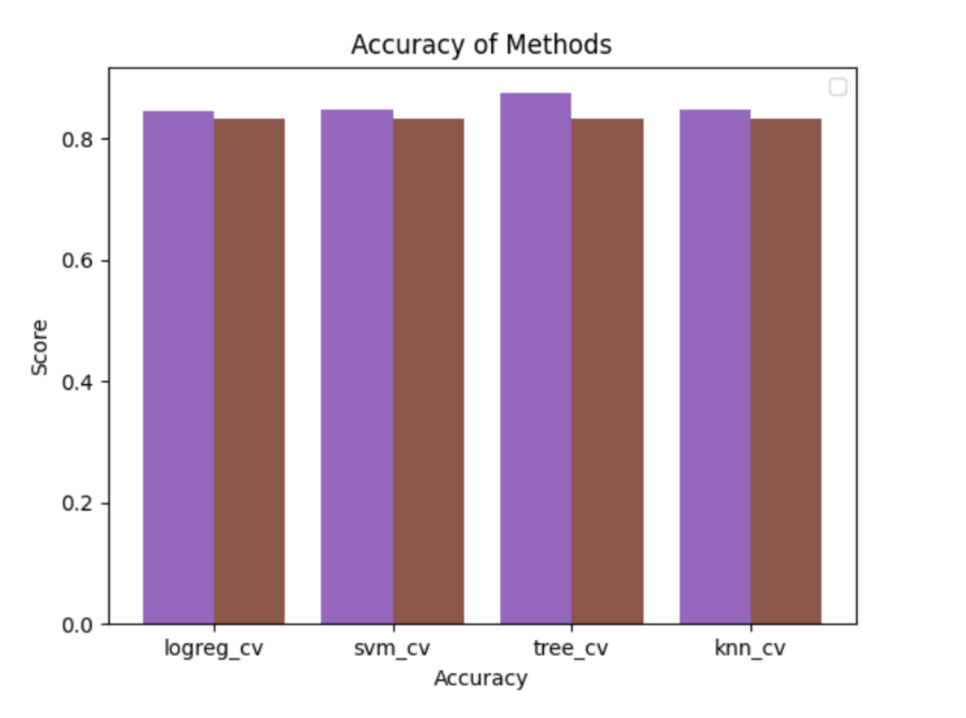
Results



- Most launches happend at east cost launch sites.

Results

- Decision Tree Classifier is the best model having accuracy over 87.5%.



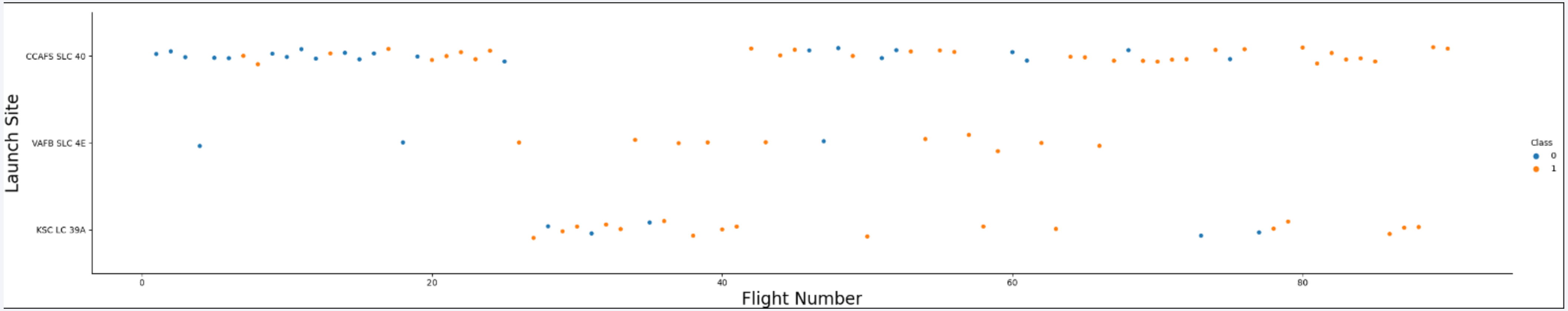
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Flight Number vs. Launch Site

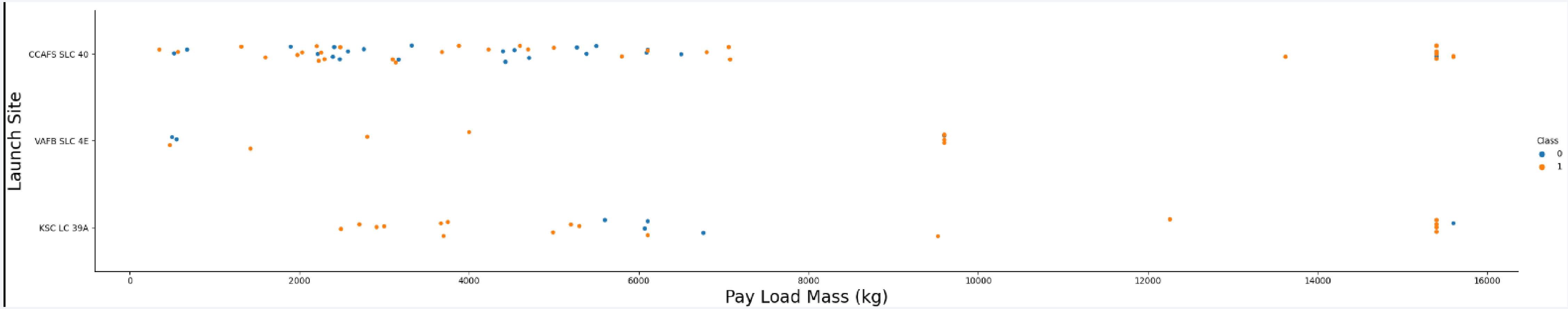
-Found that the larger the flight amount at a launch site, the greater the success rate at a launch site.



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

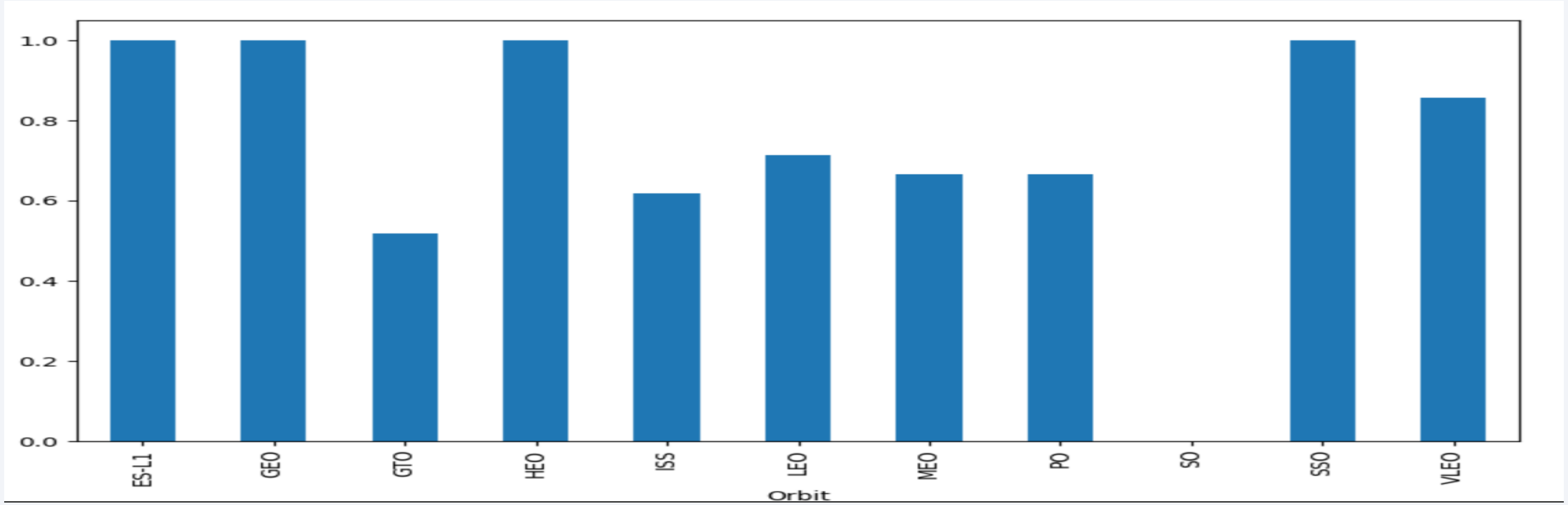
  -Payload over 9000kg have excellent success rate.



Success Rate vs. Orbit Type

- success rate of each orbit type

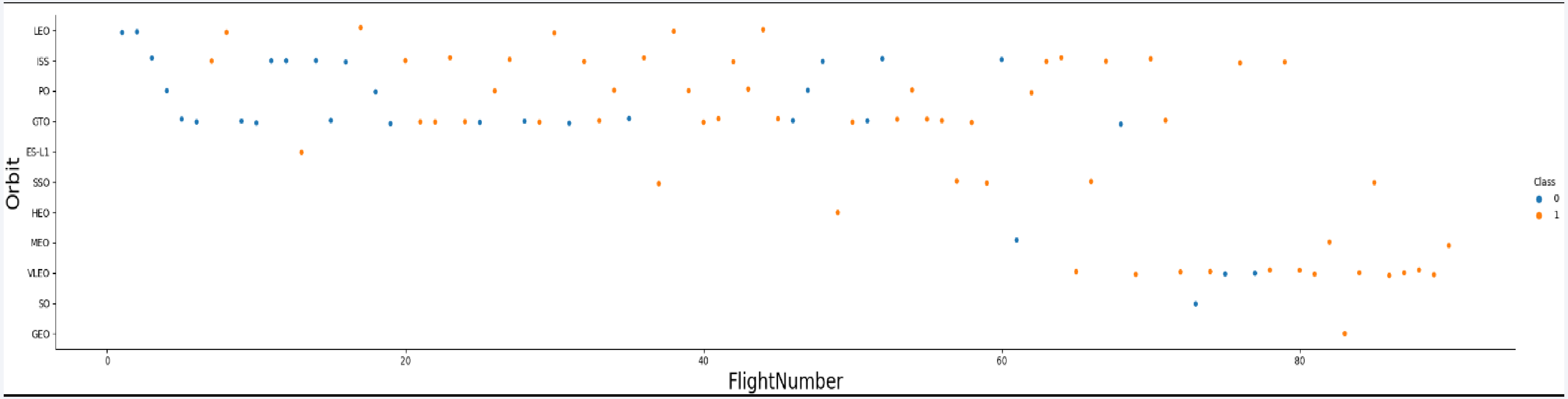
    -ES-L1, GEO, SSO and HEO have most success rates.



Flight Number vs. Orbit Type

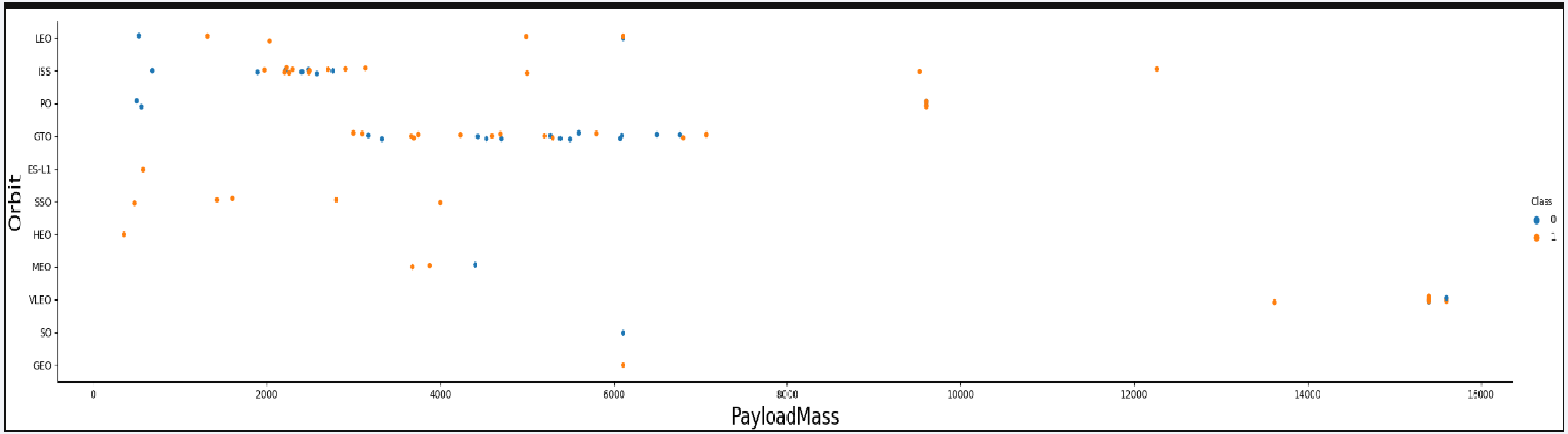
- Flight number vs. Orbit type

  -Success rate improves over time for all orbits.



Payload vs. Orbit Type

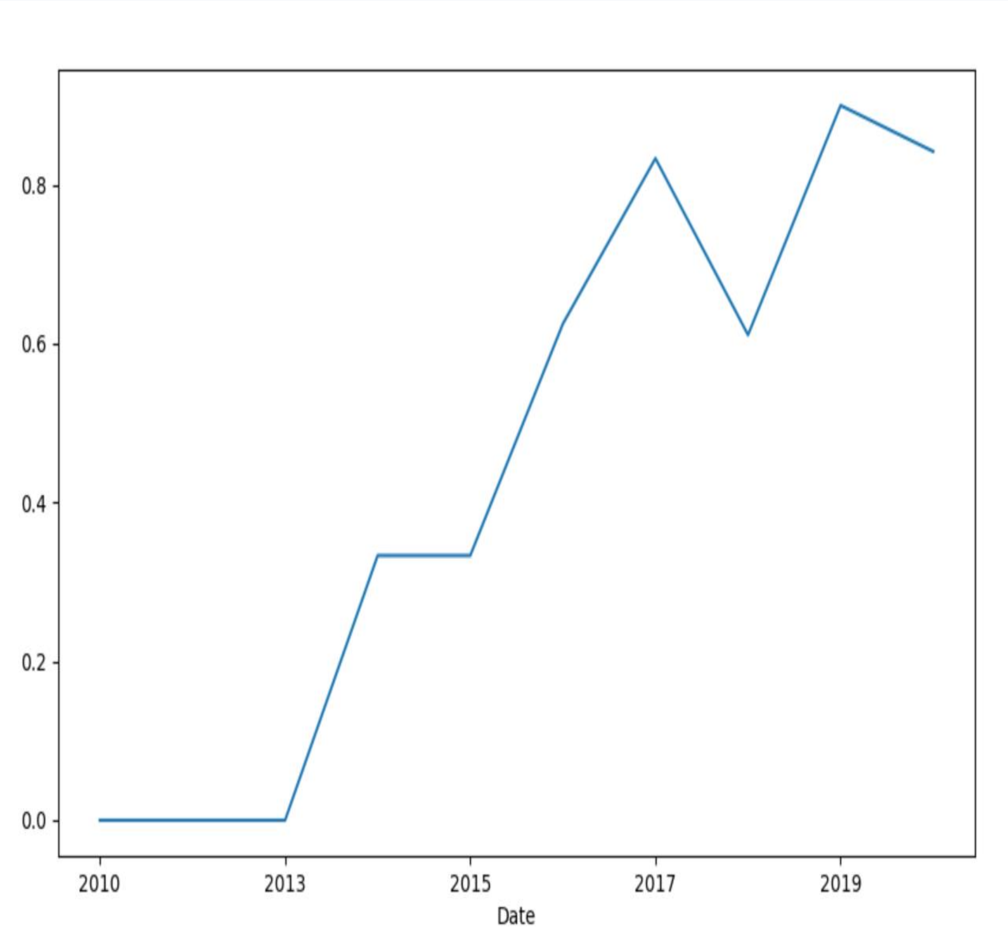
- payload vs. orbit type



Launch Success Yearly Trend

- yearly average success rate

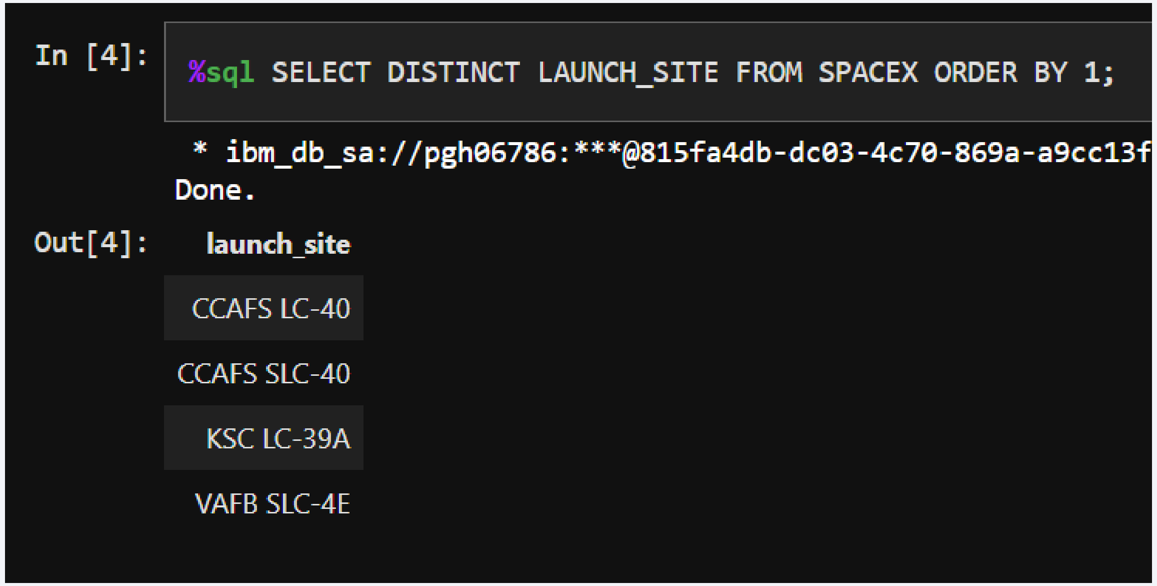
-Success rate since 2013 kept on increasing till 2020.



All Launch Site Names

- Find the names of the unique launch sites

Found the launch site names without any duplicates using Distinct

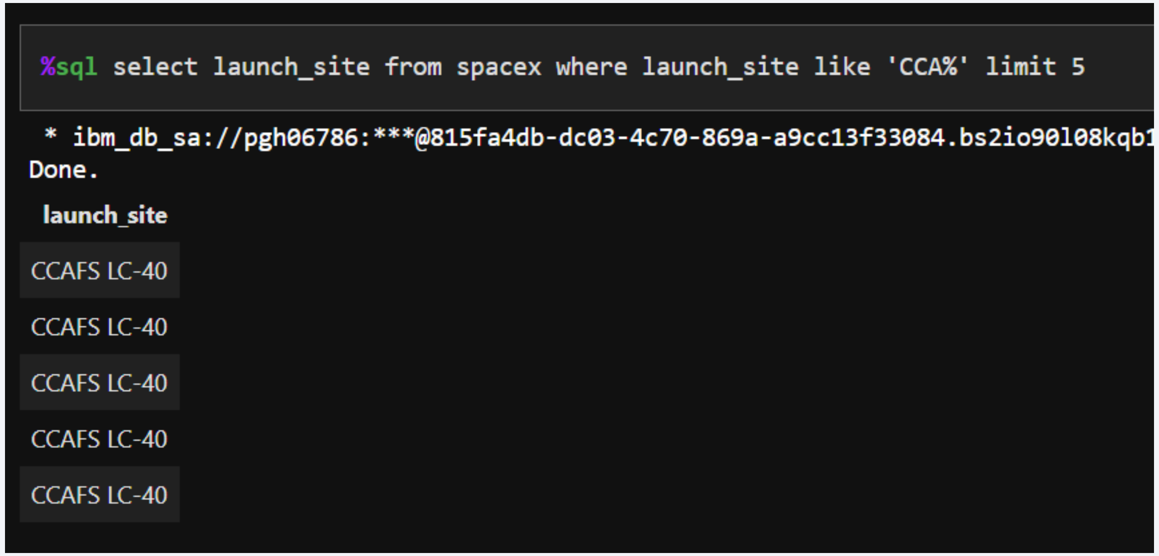


```
In [4]: %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX ORDER BY 1;

         * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f
        Done.

Out[4]:   launch_site
          CCAFS LC-40
         CCAFS SLC-40
           KSC LC-39A
          VAFB SLC-4E
```

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

  - Fetched data using like keyword



```
%sql select launch_site from spacex where launch_site like 'CCA%' limit 5
```

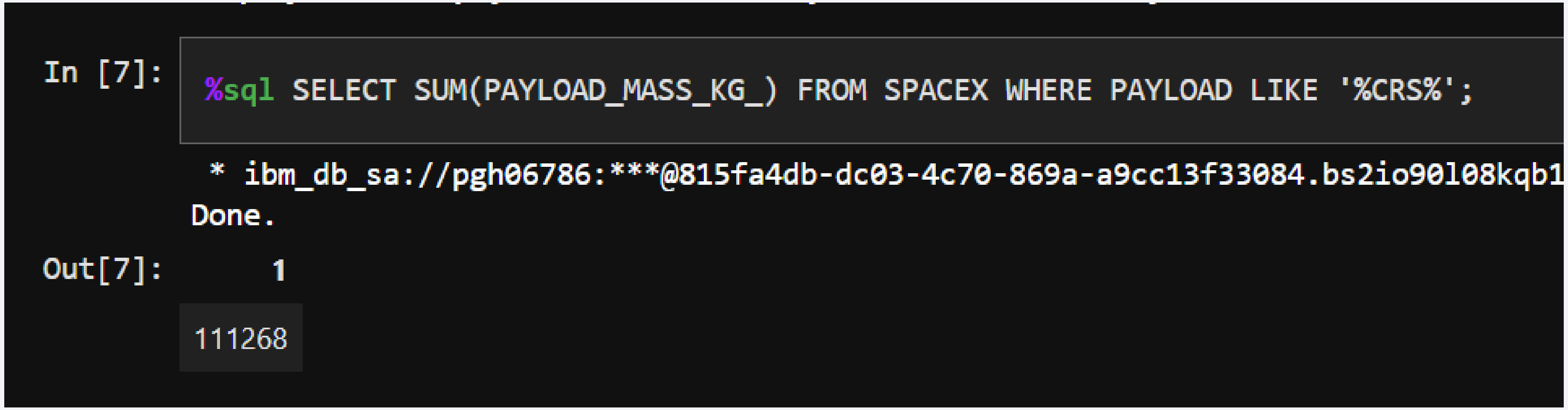
```
 * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1
Done.
```

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA.

  -Found the total using sum() function

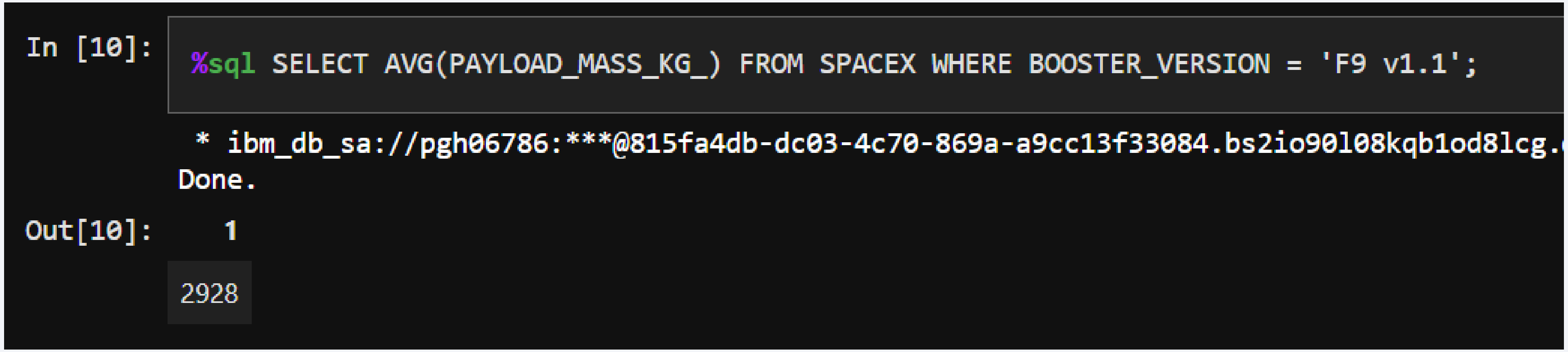


```
In [7]:  %sql SELECT SUM(PAYLOAD_MASS_KG_) FROM SPACEX WHERE PAYLOAD LIKE '%CRS%';

          * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1
         Done.
Out[7]:      1
         111268
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

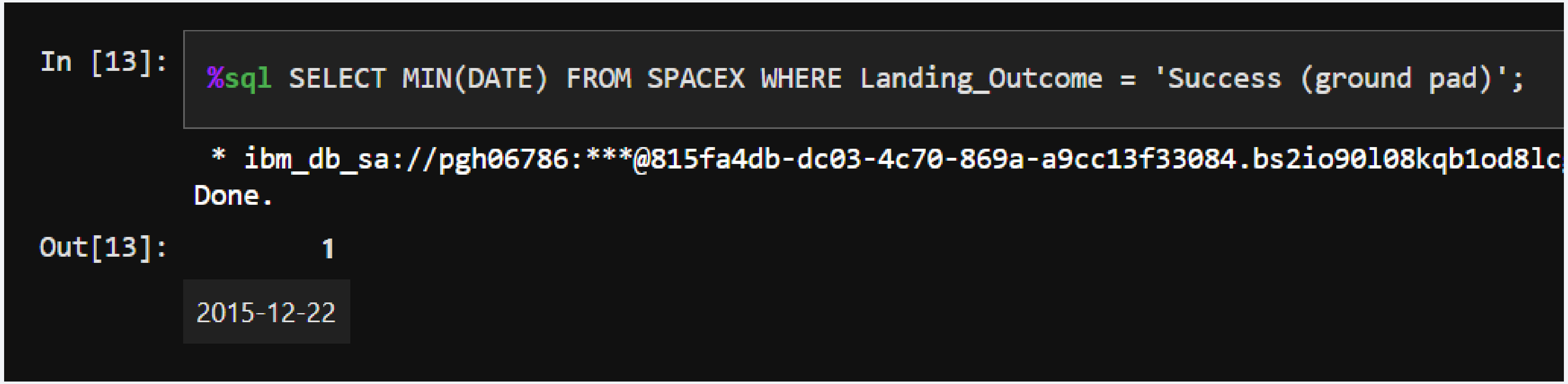
  - Found the average using AVG() function.



```
In [10]:  %sql SELECT AVG(PAYLOAD_MASS_KG_) FROM SPACEX WHERE BOOSTER_VERSION = 'F9 v1.1';

           * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.
          Done.
Out[10]:    1
          2928
```

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

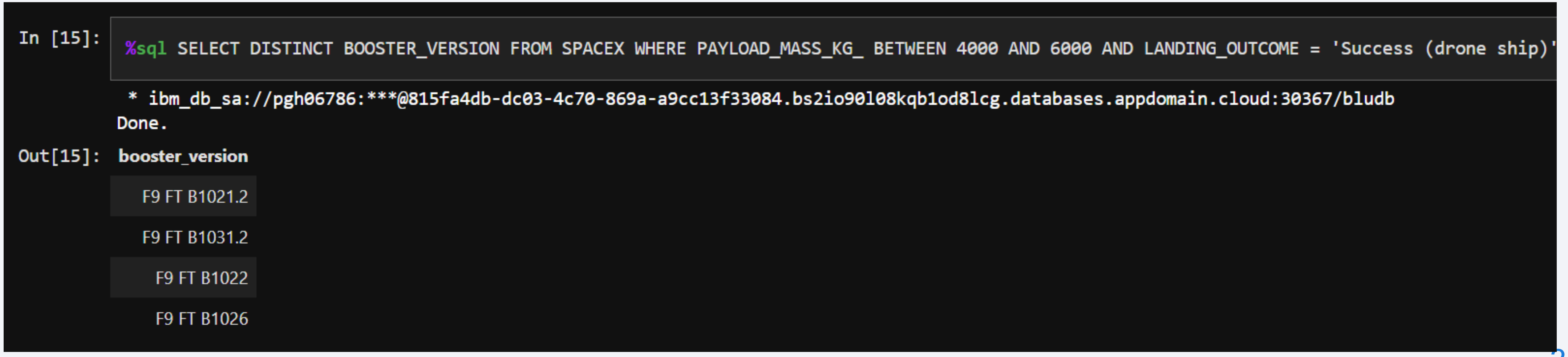


```
In [13]:  %sql SELECT MIN(DATE) FROM SPACEX WHERE Landing_Outcome = 'Success (ground pad)';

           * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lc
          Done.

Out[13]:          1
          2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

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

  - Used between keyword to select the range.



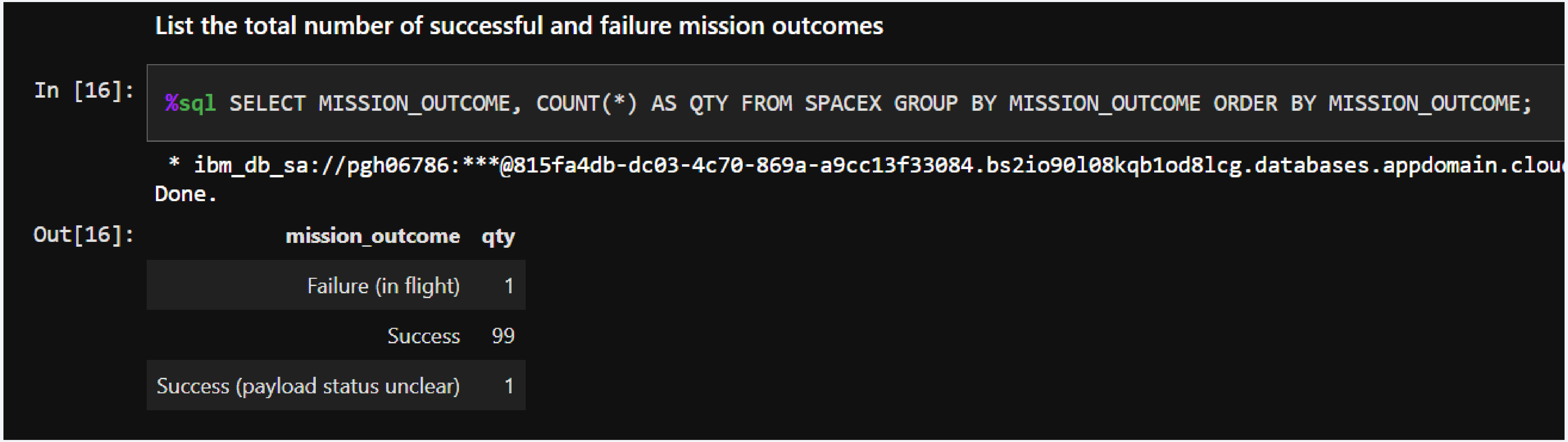
```
In [15]:  %sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEX WHERE PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000 AND LANDING_OUTCOME = 'Success (drone ship)'

           * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30367/bludb
          Done.
```

Out[15]:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcome.



List the total number of successful and failure mission outcomes

```
In [16]: %sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEX GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

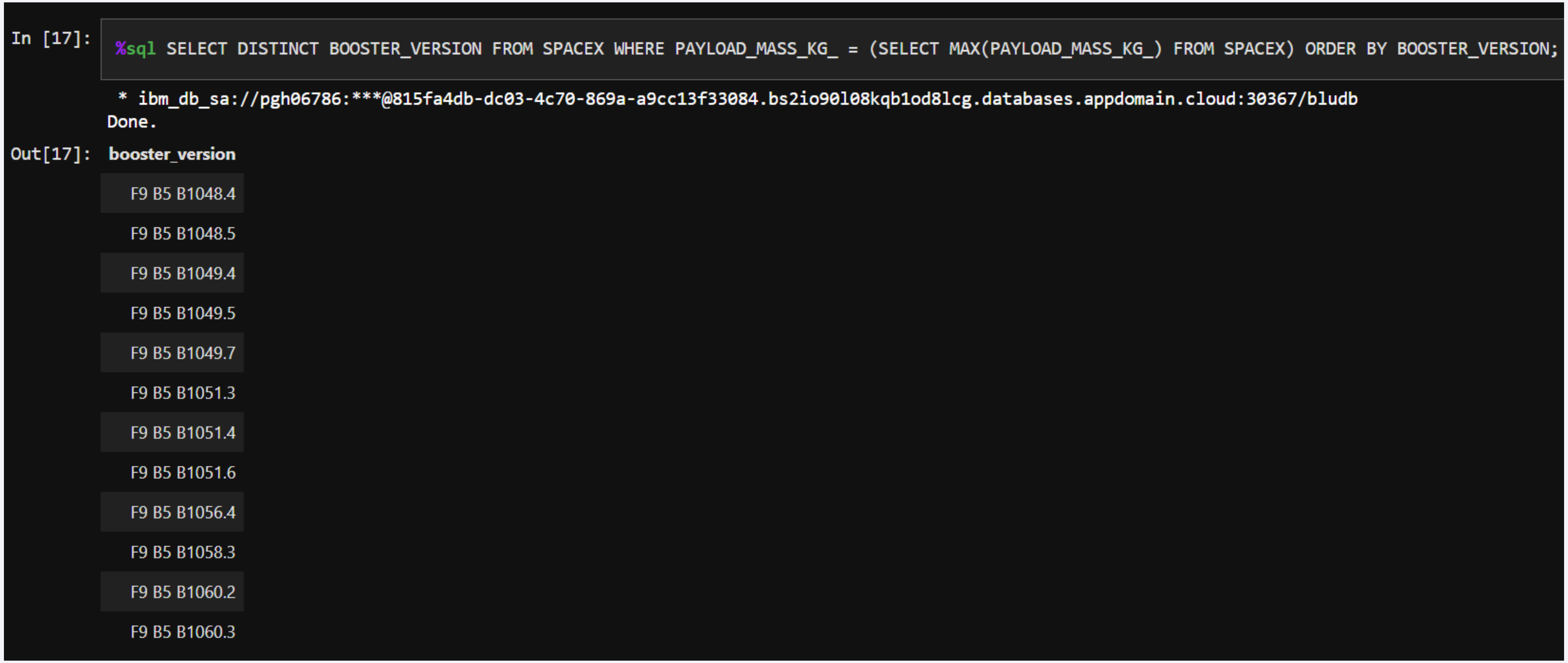
```
 * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.clou
Done.
```

Out[16]:

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

Boosters Carried Maximum Payload

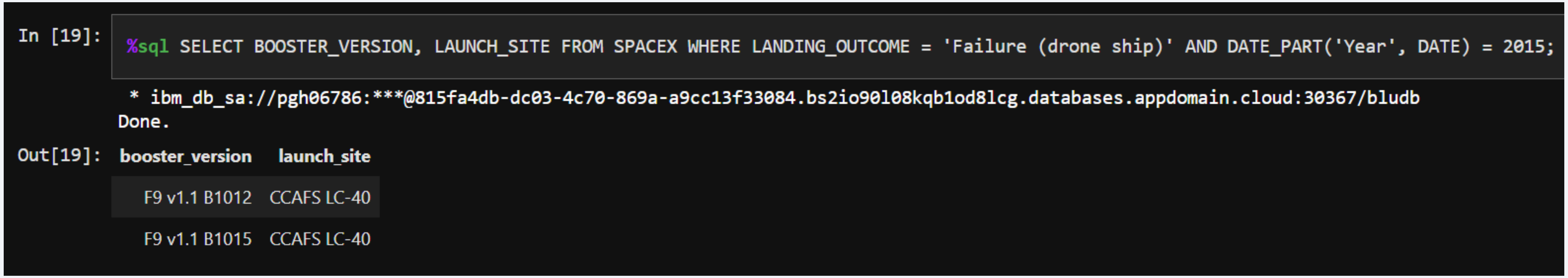
- List the names of the booster which have carried the maximum payload mass



```
In [17]:  %sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEX WHERE PAYLOAD_MASS_KG_ = (SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEX) ORDER BY BOOSTER_VERSION;

           * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30367/bludb
          Done.
Out[17]:  booster_version
             F9 B5 B1048.4
             F9 B5 B1048.5
             F9 B5 B1049.4
             F9 B5 B1049.5
             F9 B5 B1049.7
             F9 B5 B1051.3
             F9 B5 B1051.4
             F9 B5 B1051.6
             F9 B5 B1056.4
             F9 B5 B1058.3
             F9 B5 B1060.2
             F9 B5 B1060.3
```

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015



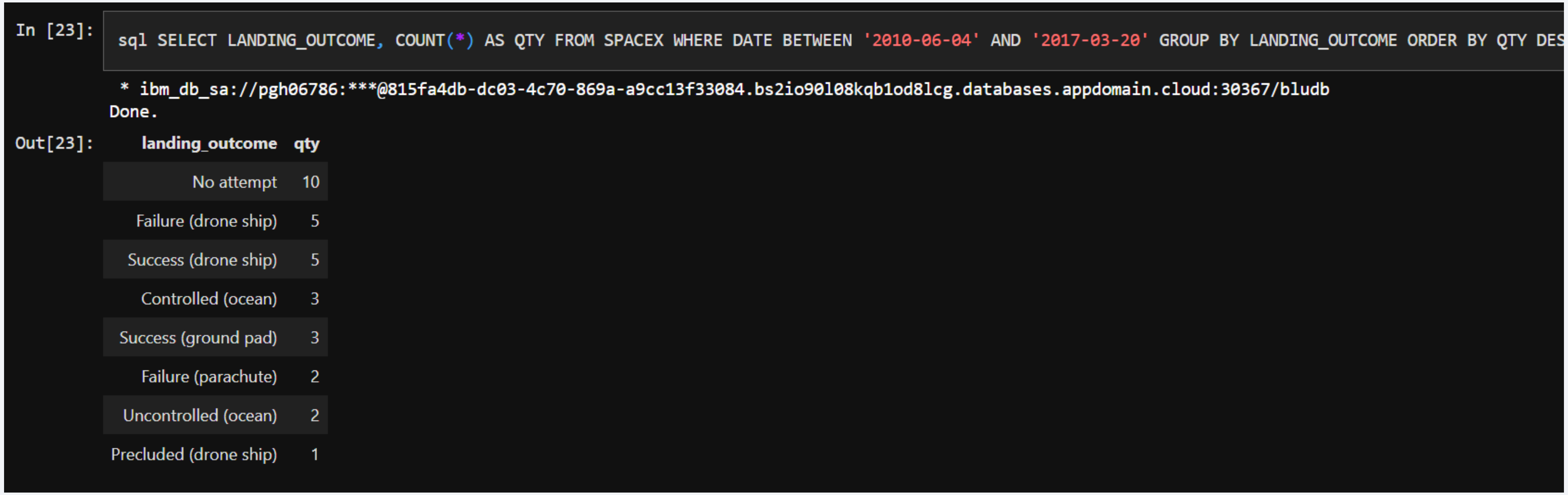
```
In [19]: %sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEX WHERE LANDING_OUTCOME = 'Failure (drone ship)' AND DATE_PART('Year', DATE) = 2015;

          * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30367/bludb
         Done.
```

Out[19]:

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

- 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



```
In [23]: sql SELECT LANDING_OUTCOME, COUNT(*) AS QTY FROM SPACEX WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY QTY DES
```

```
 * ibm_db_sa://pgh06786:***@815fa4db-dc03-4c70-869a-a9cc13f33084.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30367/bludb
Done.
```

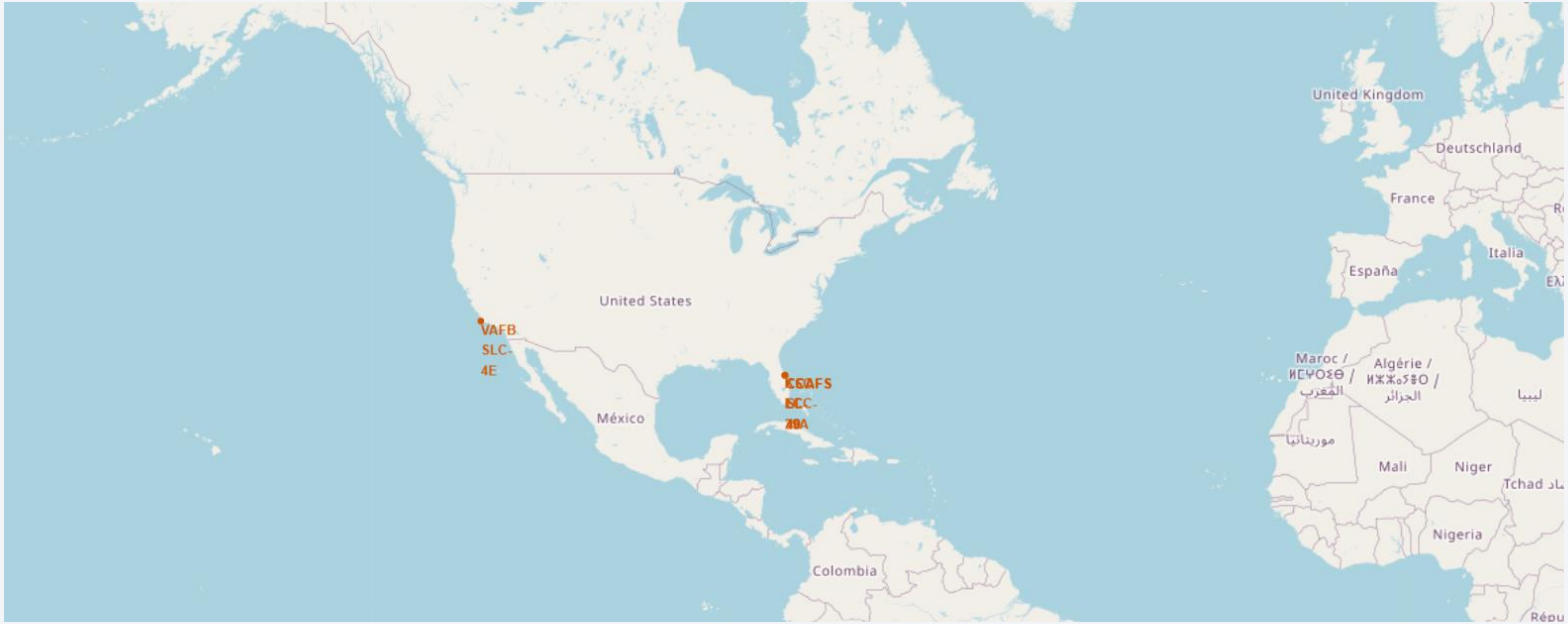
Out[23]:

landing_outcome	qty
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

Section 3

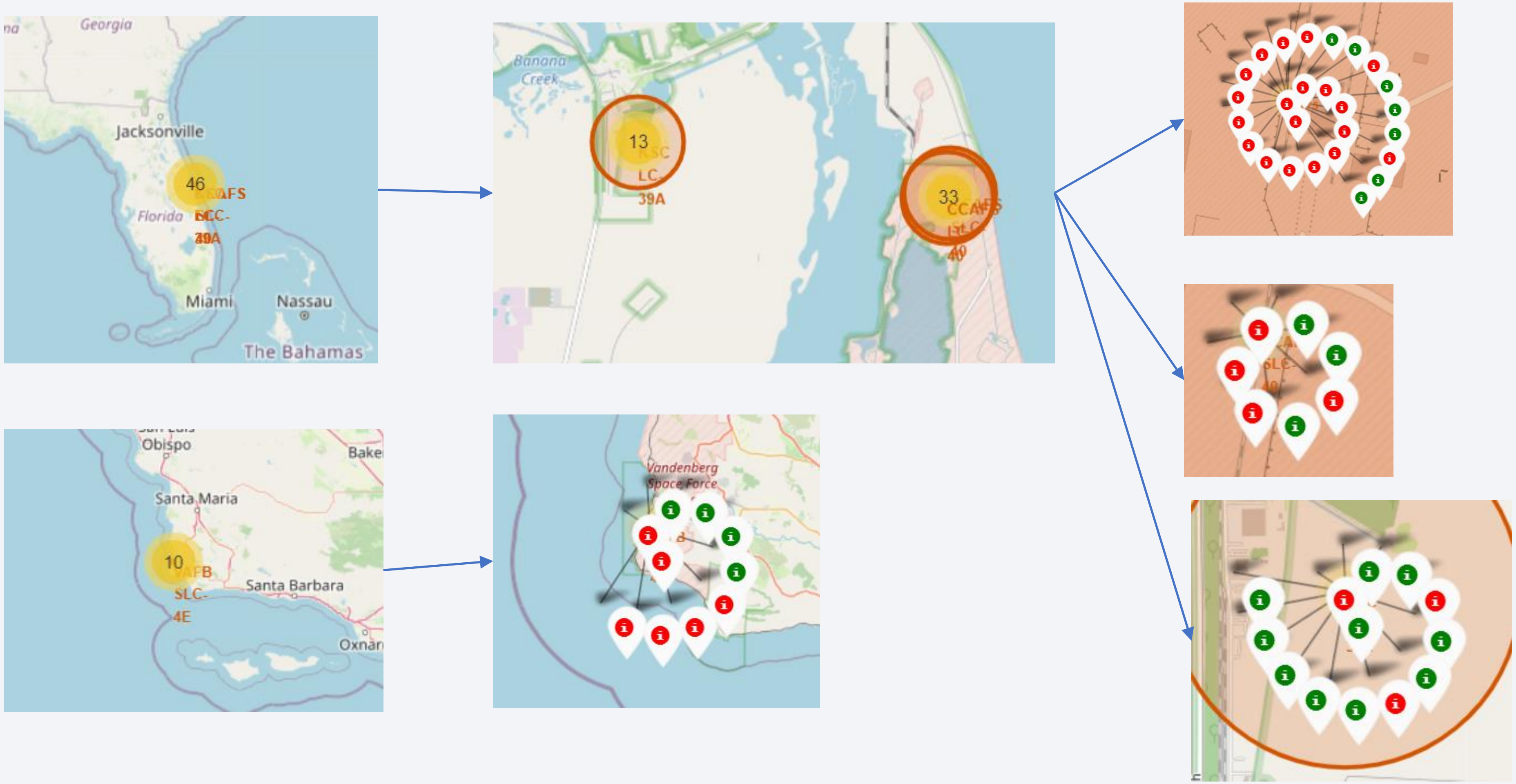
Launch Sites Proximities Analysis

Launch Sites



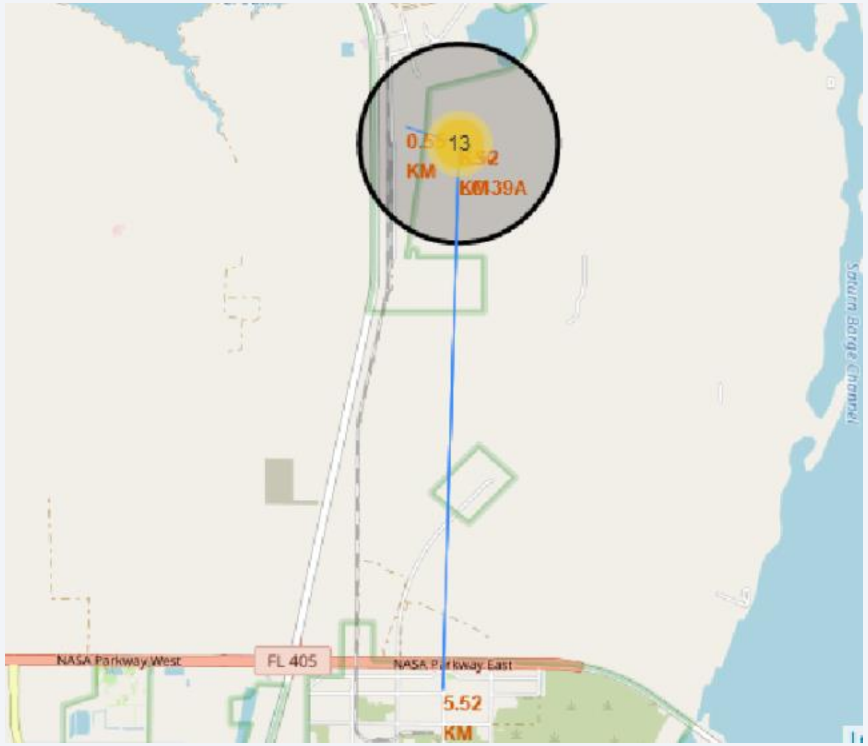
SpaceX launch sites are in united states ameria coasts.

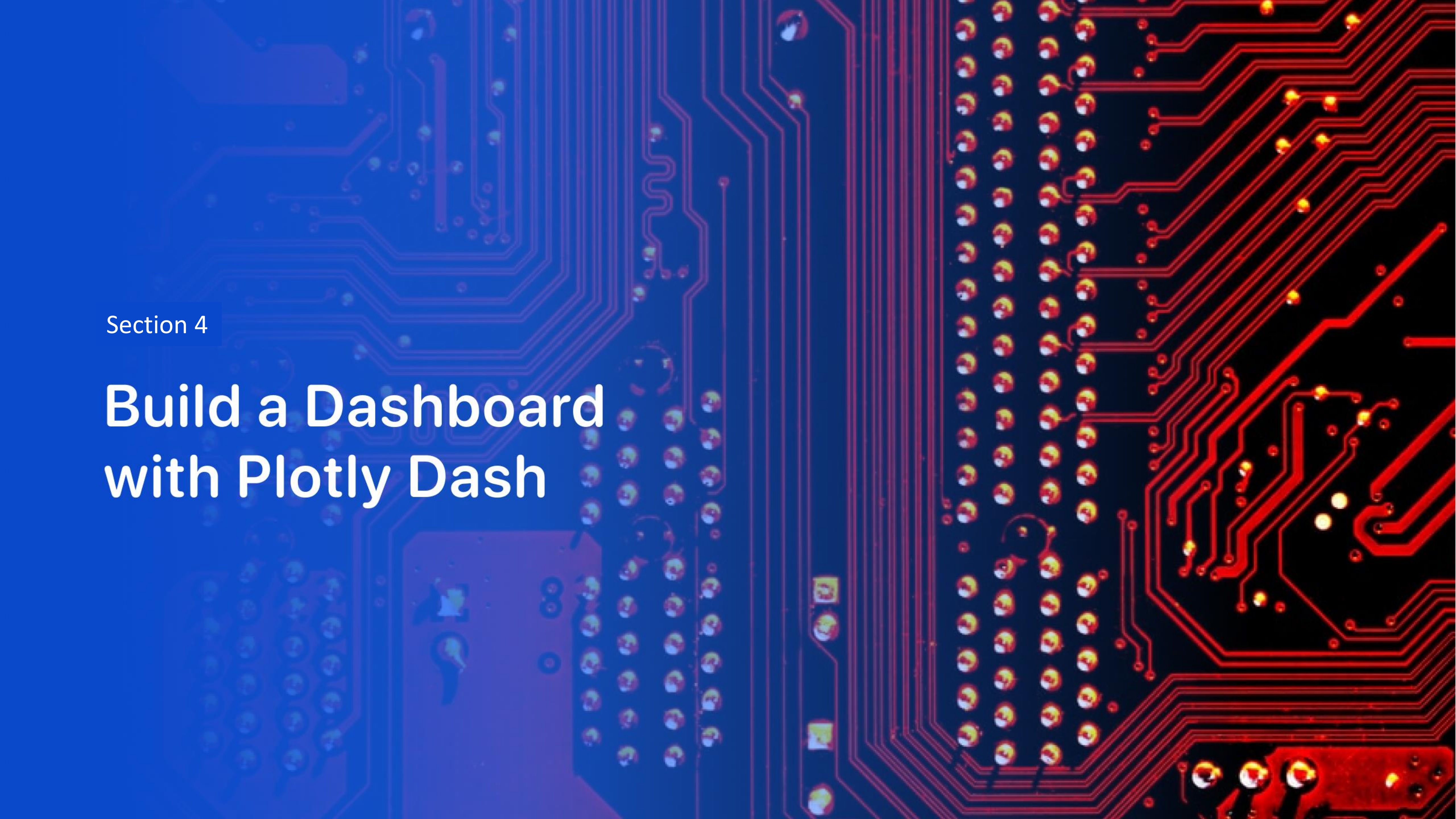
Outcomes by sites



- Green is success and red is failure

Distance to landmarks

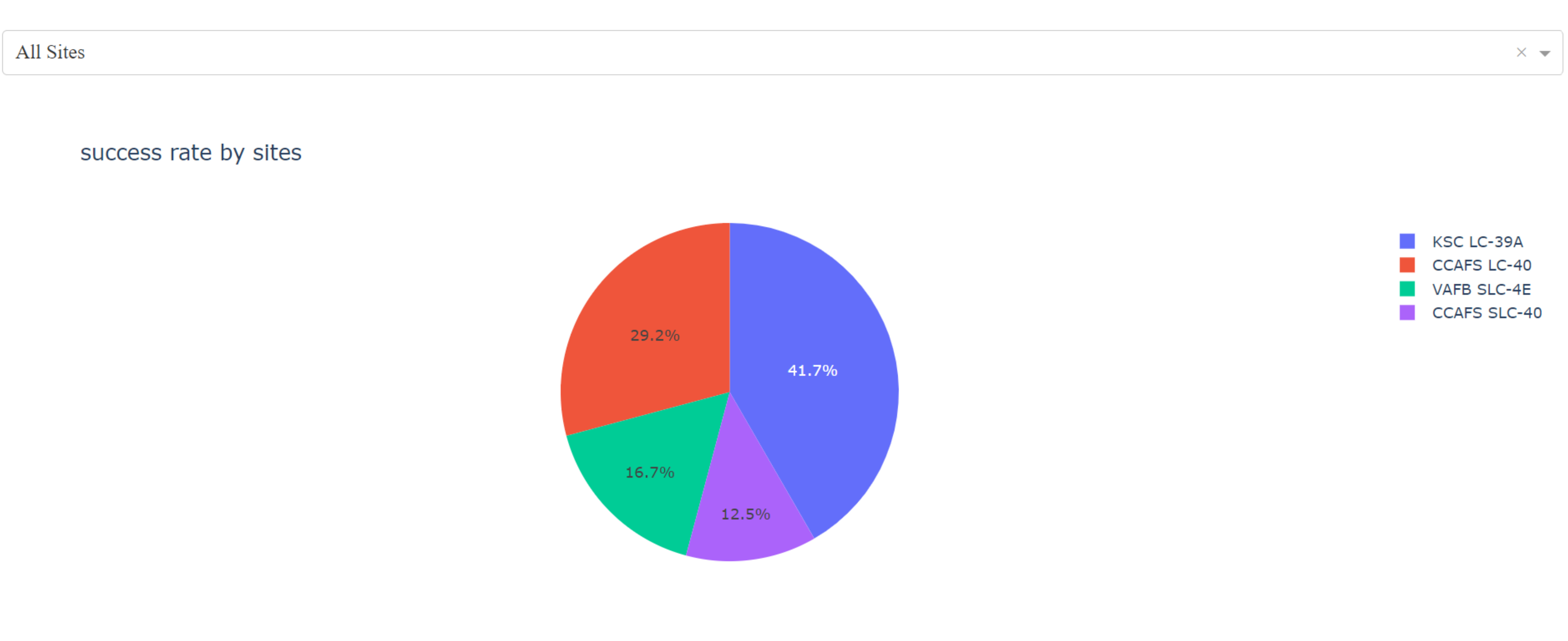




Section 4

Build a Dashboard with Plotly Dash

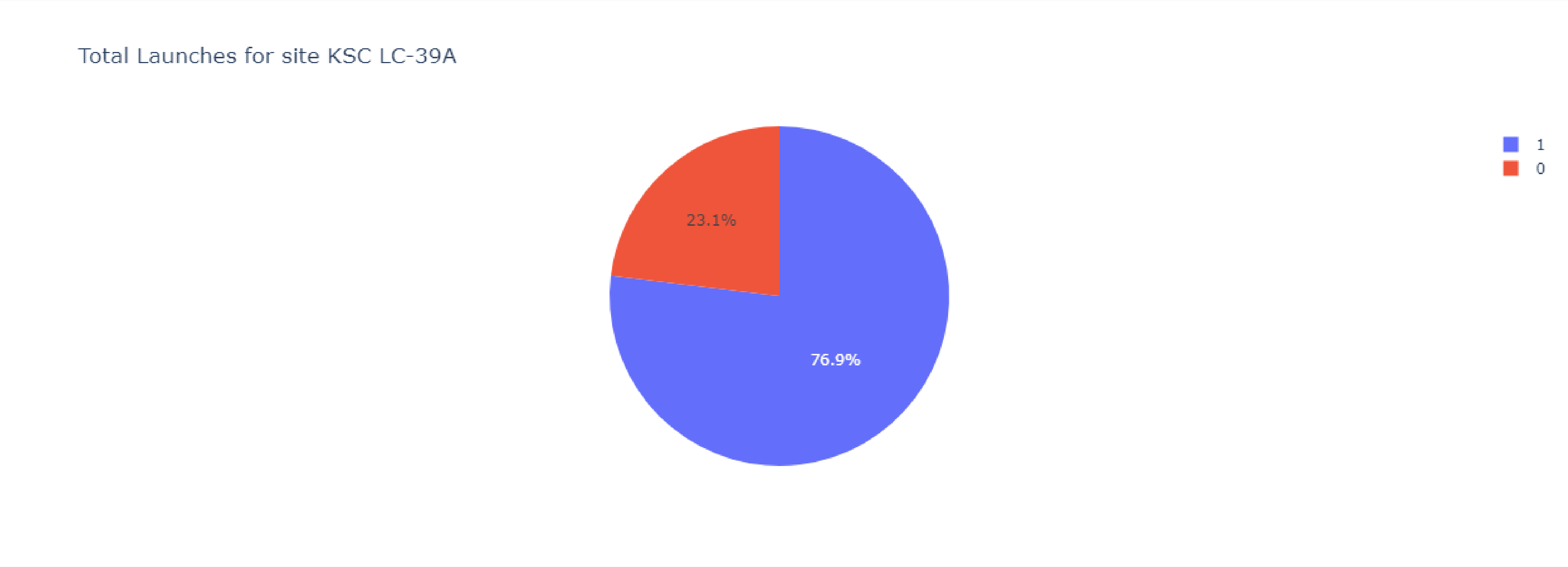
Success rate for each launching site



All Sites

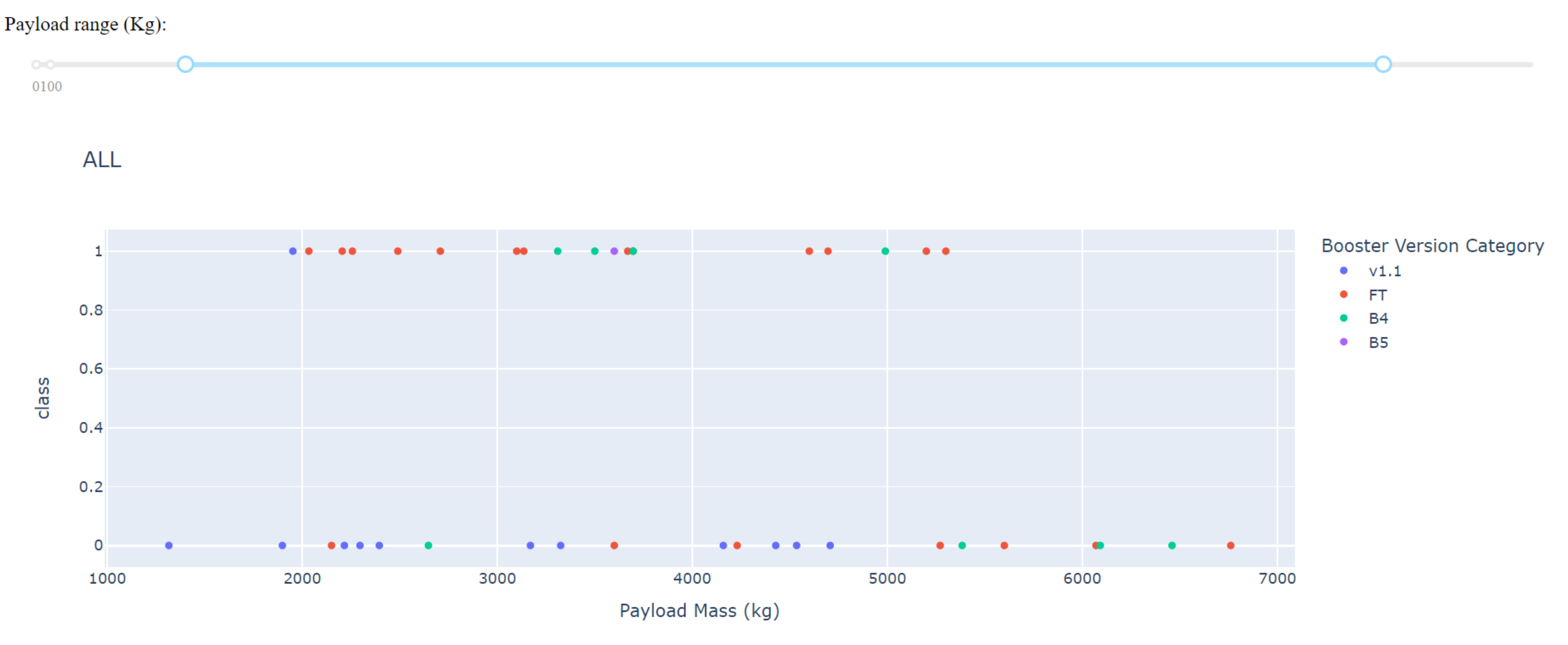
- KSC LC-39A have most success rate.

Launch site with highest success ratio



76.9% Success and 23.1% Failure.

Payload vs class



Payload range (Kg):

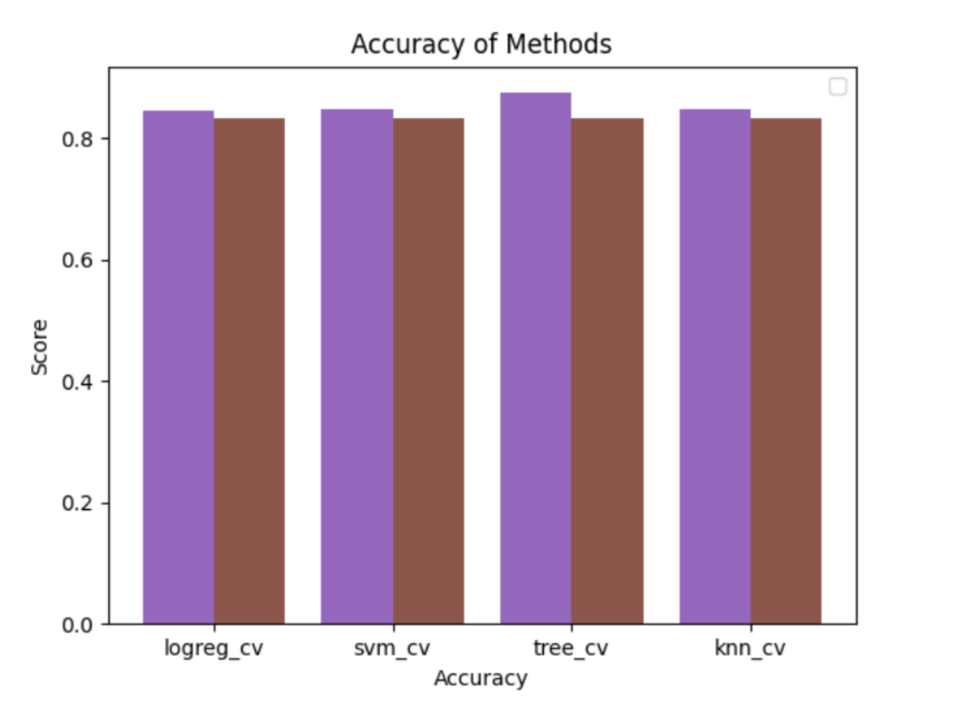
- FT booster is most successful at this range

Section 5

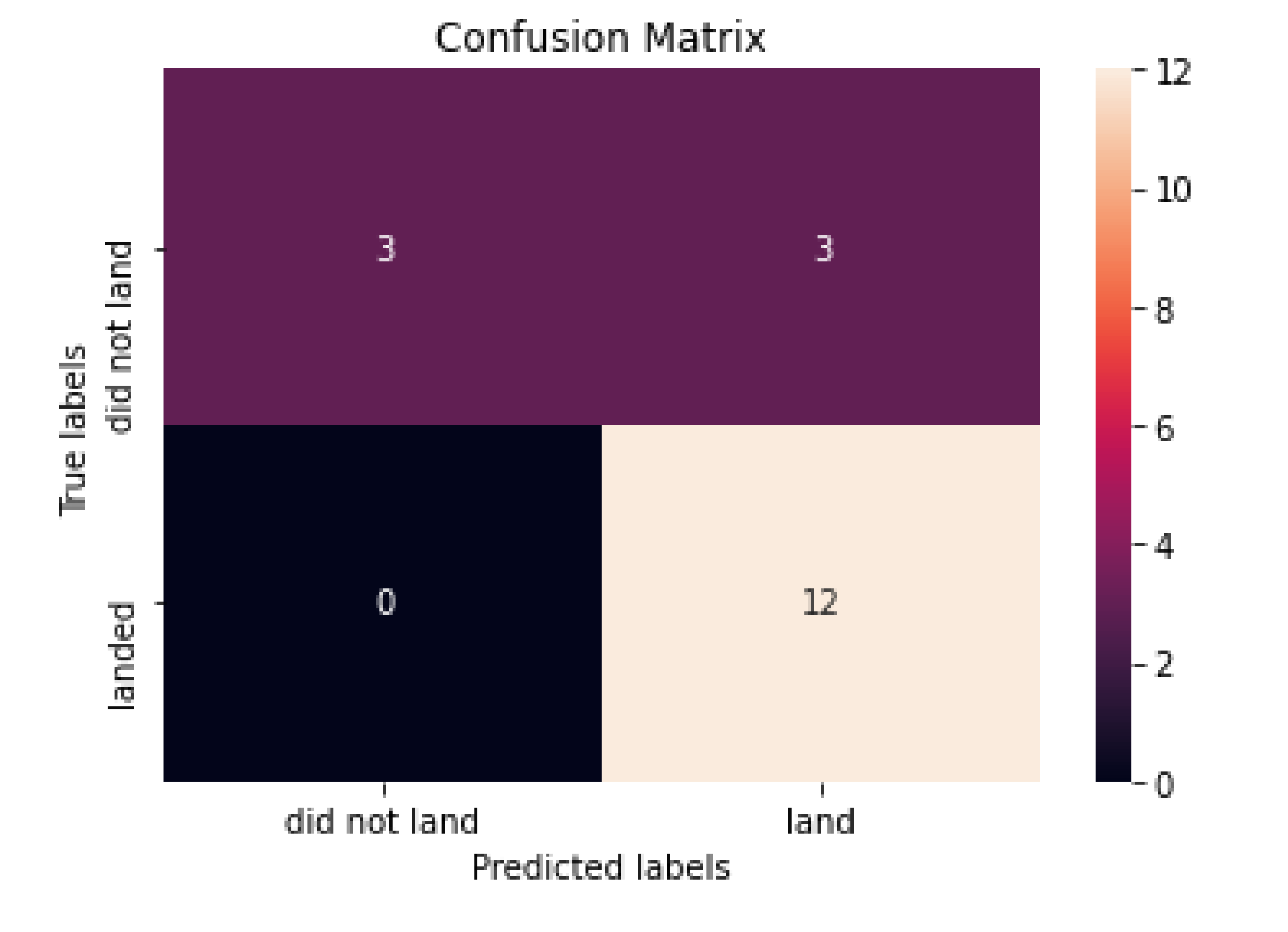
Predictive Analysis (Classification)

Classification Accuracy

- Decision Tree Classifier is the best model having accuracy over 87.5%.



Confusion Matrix



Conclusions

- Launch success rate started to increase in 2013 till 2020.
- Launches above 7000kg are less risky
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

Thank you!