IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<LONG AN>
<2023-04-15>

Outline

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

Executive Summary

- Summary of methodologies:
  - Data Collection using SpaceX API and web scraping
  - Exploratory Analysis using SQL, Pandas, Maptplotlib
  - Data Visualization using Ploty Dash and Folium
  - Machine Learning Prediction using Logistic Regression, SVC, Decision Tree Classifier, Kneighbors Classifier, Grid Search CV models.
- Summary of all results:
  - Successfully collected data using API and web scarping. Save the data to csv file after cleaning the data.
  - Determine training labels for supervised models. And the success rate of Falcon 9 first stage is 67%. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%. ES-L1, GEO, HEO, SSO have high success rate. The success rate since 2013 kept increasing till 2020.

Introduction

- Project background and context:
  The goal is to determine the price of each launch for a new rocket company, Space Y, that would like to compete with SpaceX.
- Problems you want to find answers:
  - What data do we need to gather on past SpaceX launches to determine if the first stage was reused or not?
  - What are the key features or variables that may influence whether the Falcon 9 first stage will land successfully?
  - How can we use machine learning algorithms to build a predictive model that can determine the likelihood of first stage landing success based on the available data and features?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Request to the SpaceX API
  - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Find patterns in the data and determine what would be the label for training supervised models.

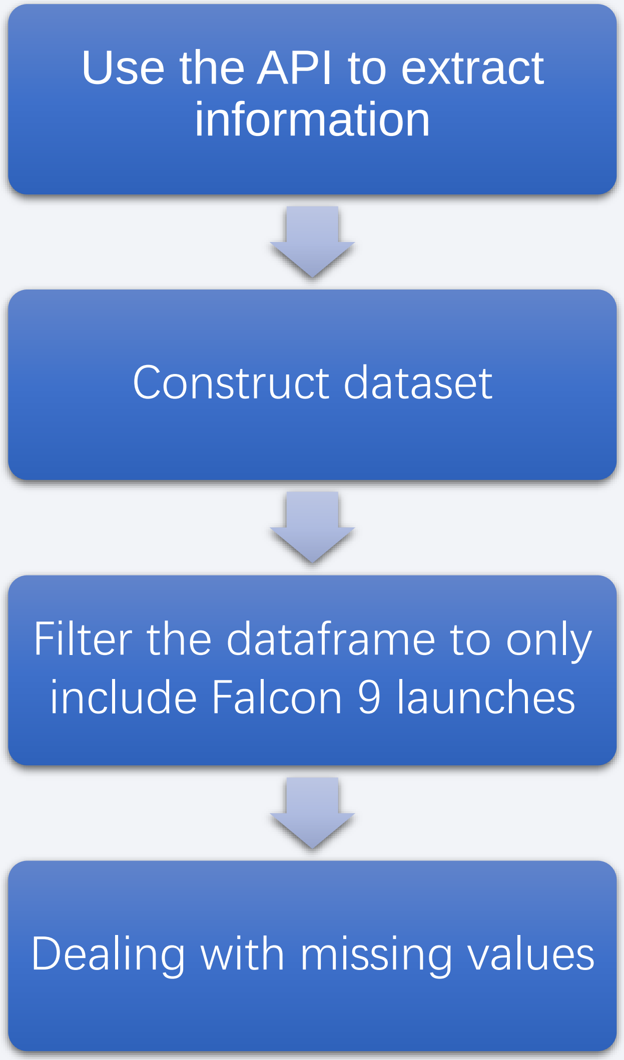
Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Create a NumPy
    - Standardize the data
    - Split the data into training data and test data
    - Create different Model to compare
    - Calculate the accuracy on the test data
    - Find the method performs best

Data Collection – SpaceX API

- Request to the SpaceX API
- Clean the requested data



Source code:
https://github.com/Michaelan171/ibm_ds_capstone/blob/master/Final_Assignment.ipynb

Data Collection - Scraping

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Source code:

https://github.com/Michaelan171/ibm_ds_capstone/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

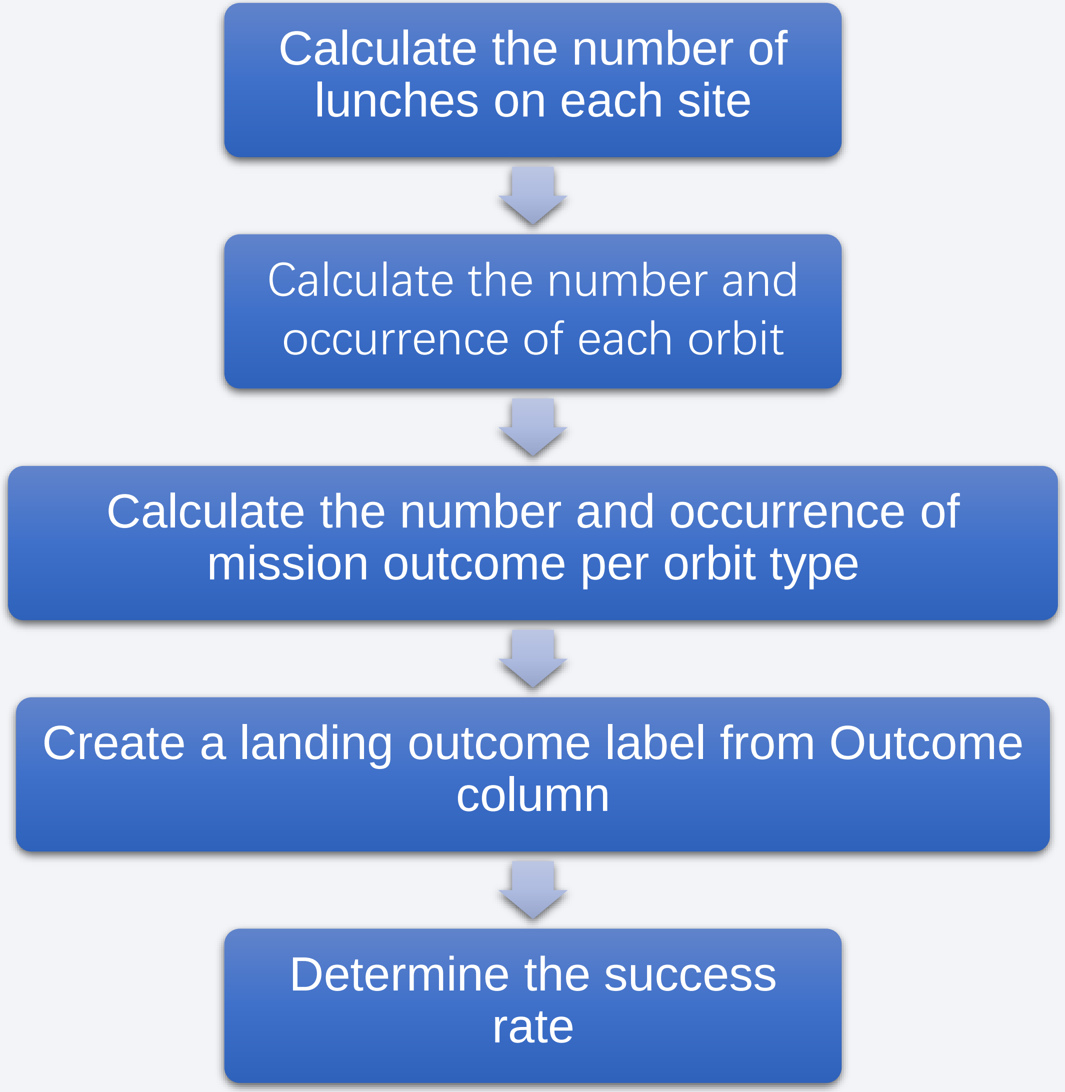
Request the Falcon9 Launch Wiki page from its URL

Extract all column/variable names from the HTML table header

Create a data frame by parsing the launch HTML tables

Data Wrangling

- Exploratory Data Analysis
- Determine Training Labels



Source code:
https://github.com/Michaelan171/ibm_ds_capstone/blob/master/EDA.ipynb

EDA with Data Visualization

- To explore data, scatter chart and bar chart were used to visualize the relationship between pair of features, and line chart to observe success rate trend.

Source code:
https://github.com/Michaelan171/ibm_ds_capstone/blob/master/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- The 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 achieved.
- 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.
- 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

Build an Interactive Map with Folium

- Markers, circles, lines and marker cluster were used with Folium Maps
- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site
- Lines are used to indicate distances between two coordinates.

Source code:
https://github.com/Michaelan171/ibm_ds_capstone/blob/master/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Scatter and Pie Charts were added to the dashboard
  - Pie Chart is to show the percentage of launches by site
  - Scatter Chart is to show the Correlation between payload and success for selected payload range and site

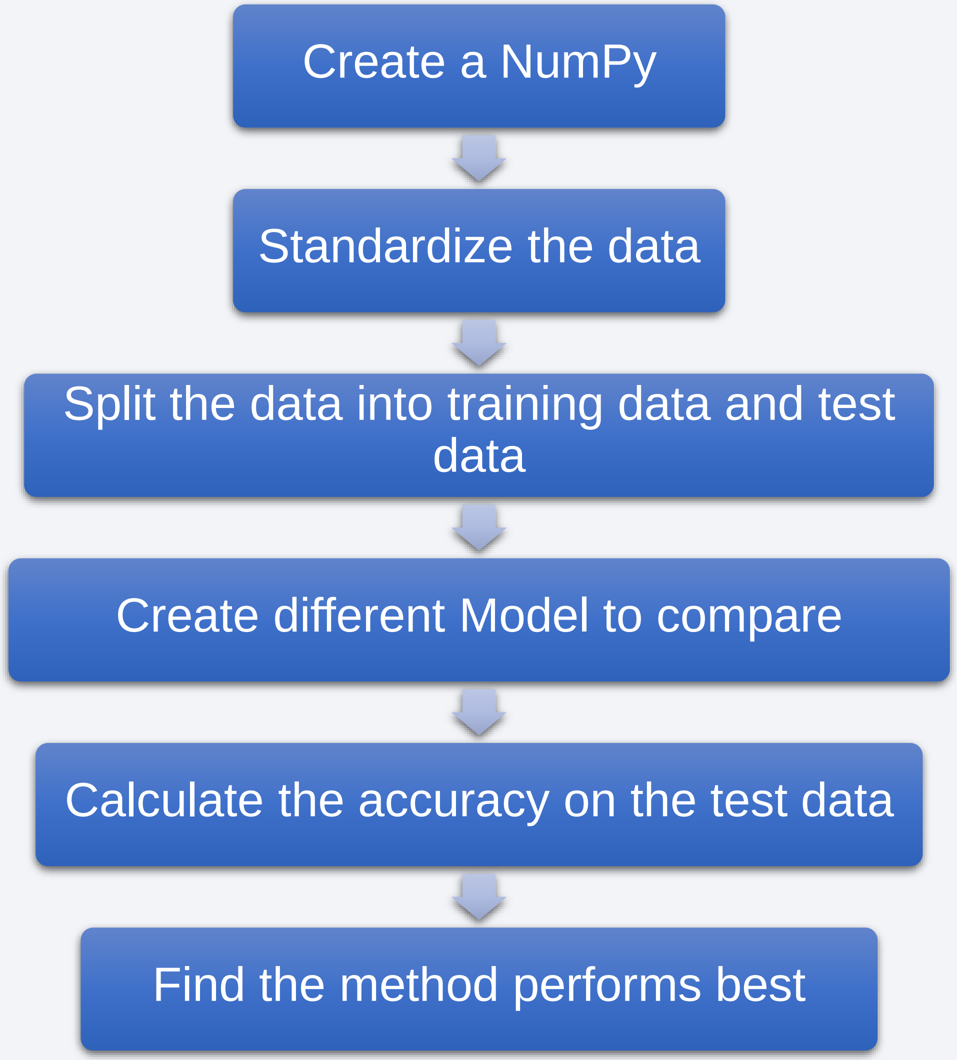
Source code:
https://github.com/Michaelan171/ibm_ds_capstone/blob/master/Space_X_dash.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

Source code:

https://github.com/Michaelan171/ibm_ds_capstone/blob/master/Machine%20Learning%20Prediction.ipynb



Results

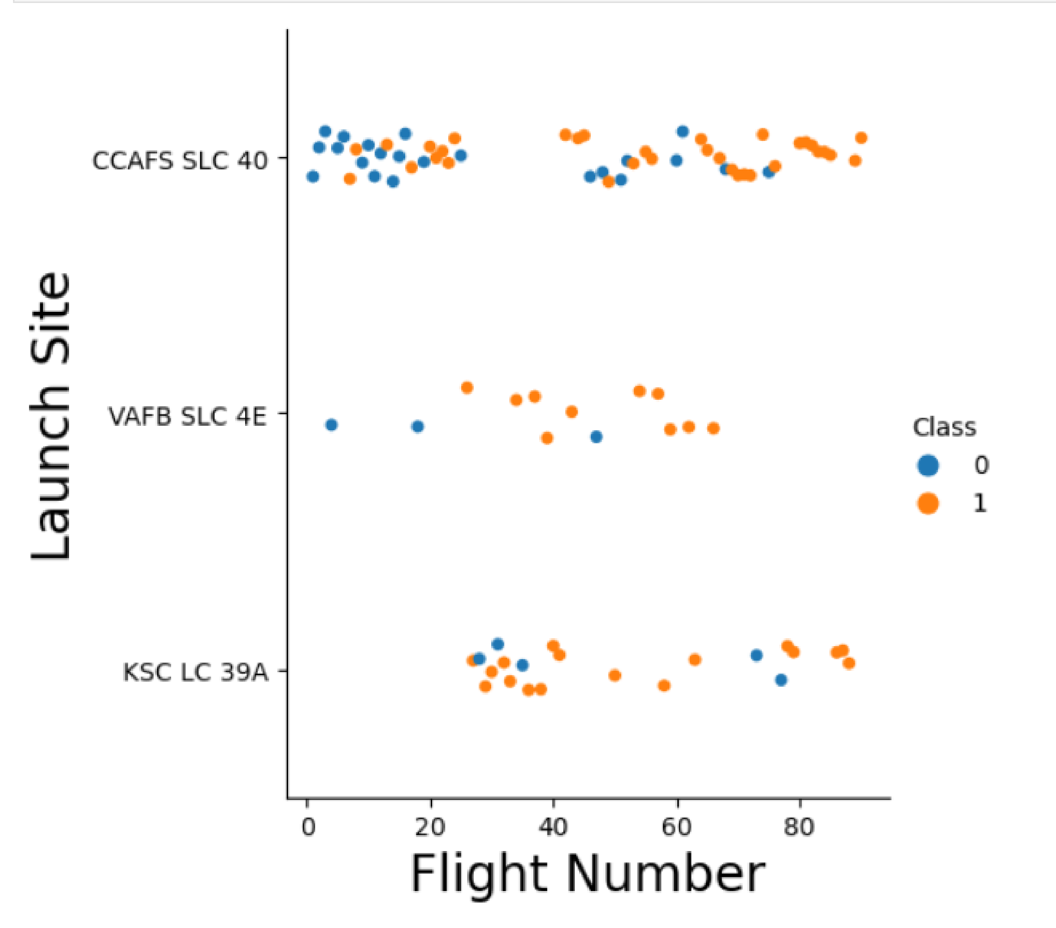
- At the KSC LC 39A launch site, there are no rockets launched for flight numbers under 20.
- The VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)
- ES-L1, GEO, HEO, SSO orbits have high success rate.
- There seems to be no relationship between flight number when in GTO orbit
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- The success rate since 2013 kept increasing till 2020Most launches happens at east cost launch sites
- Decision Tree Classifier is the best model to predict successful landings

Section 2

Insights drawn from EDA

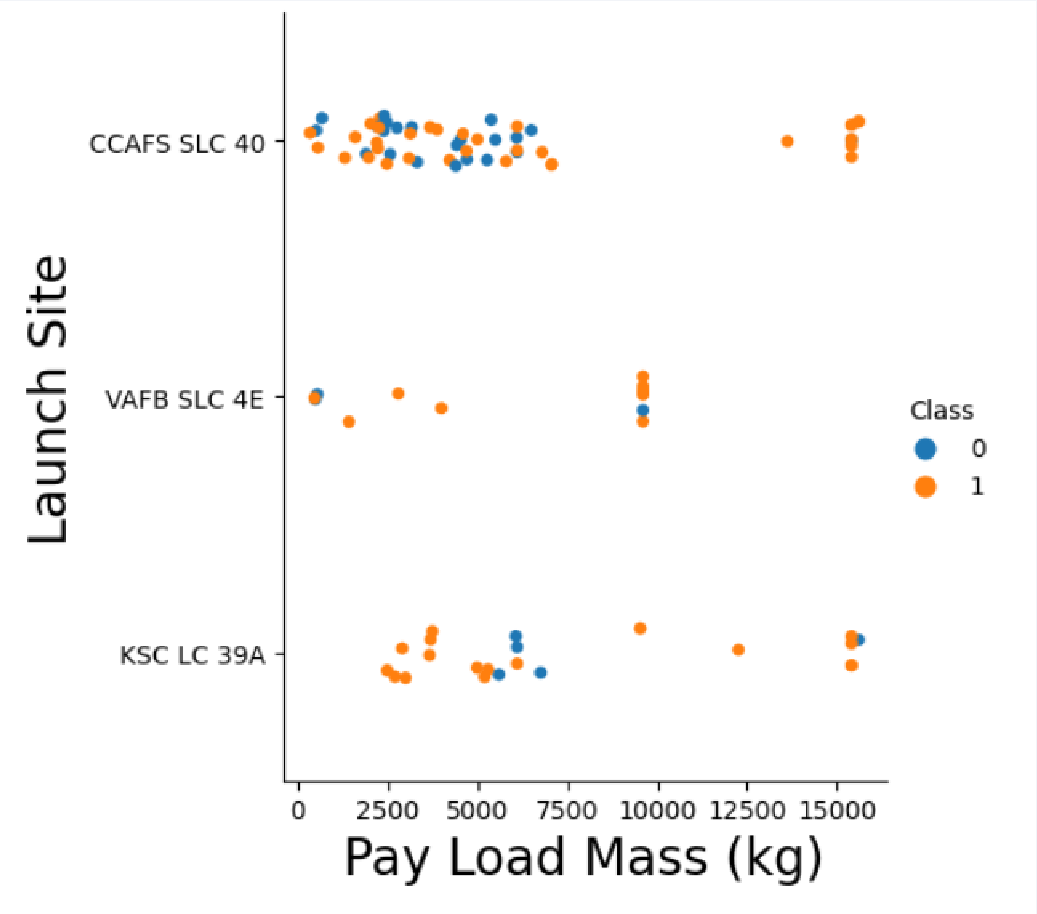
Flight Number vs. Launch Site

- At the KSC LC 39A launch site, there are no rockets launched for flight numbers under 20.



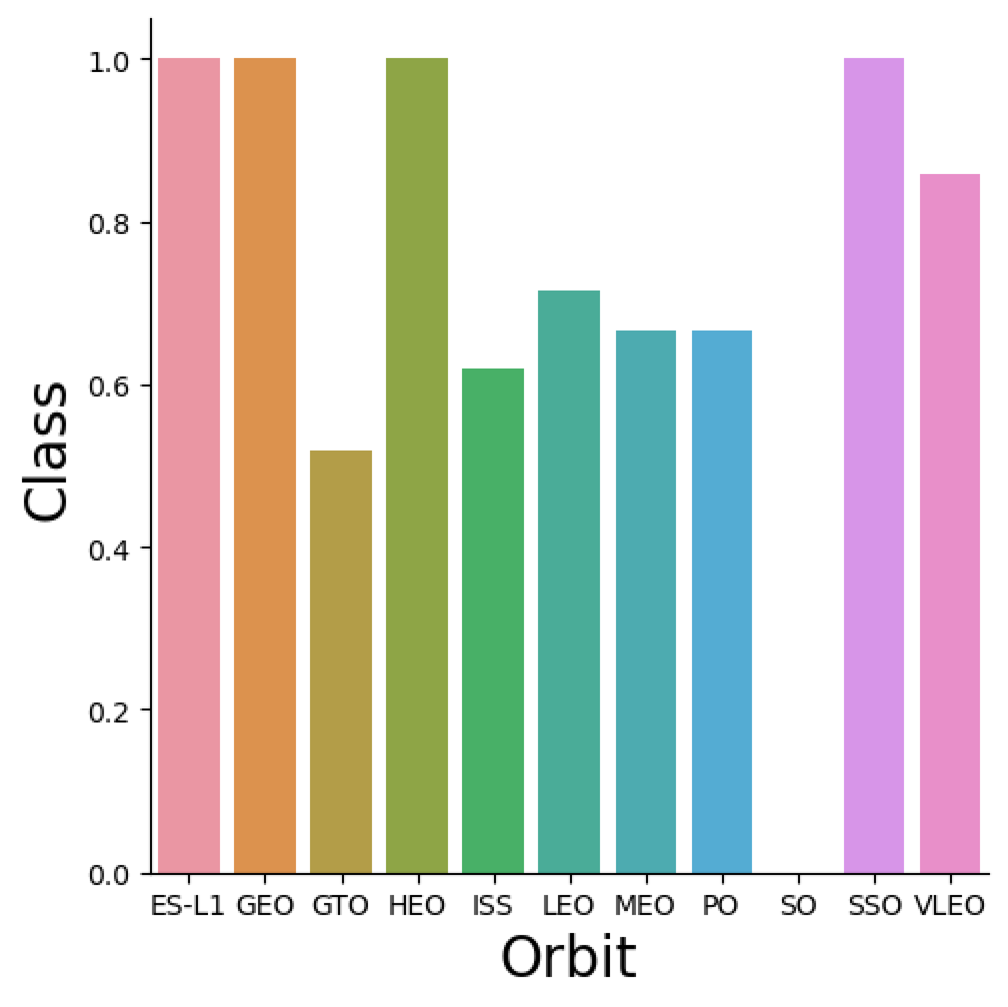
Payload vs. Launch Site

- The VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)



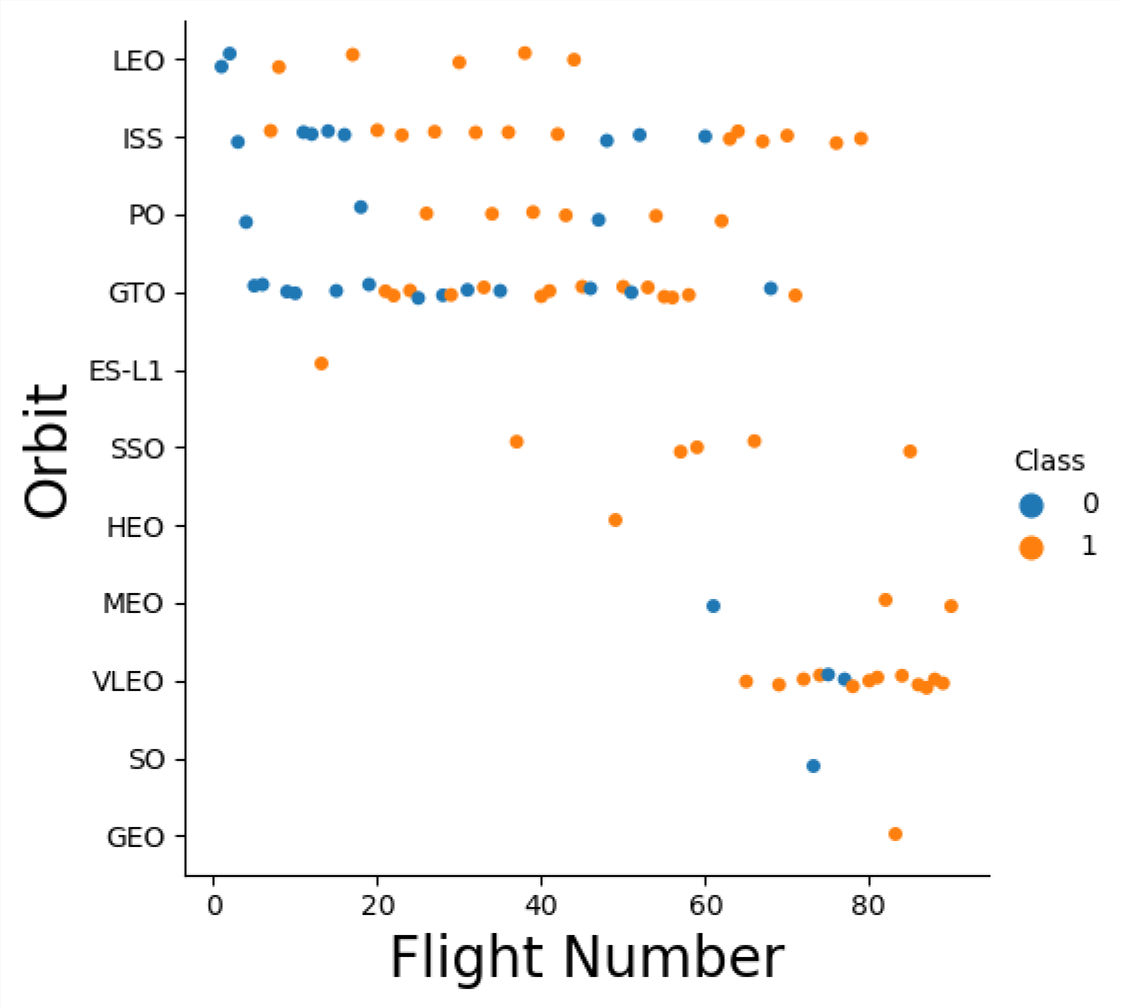
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO orbits have high success rate.



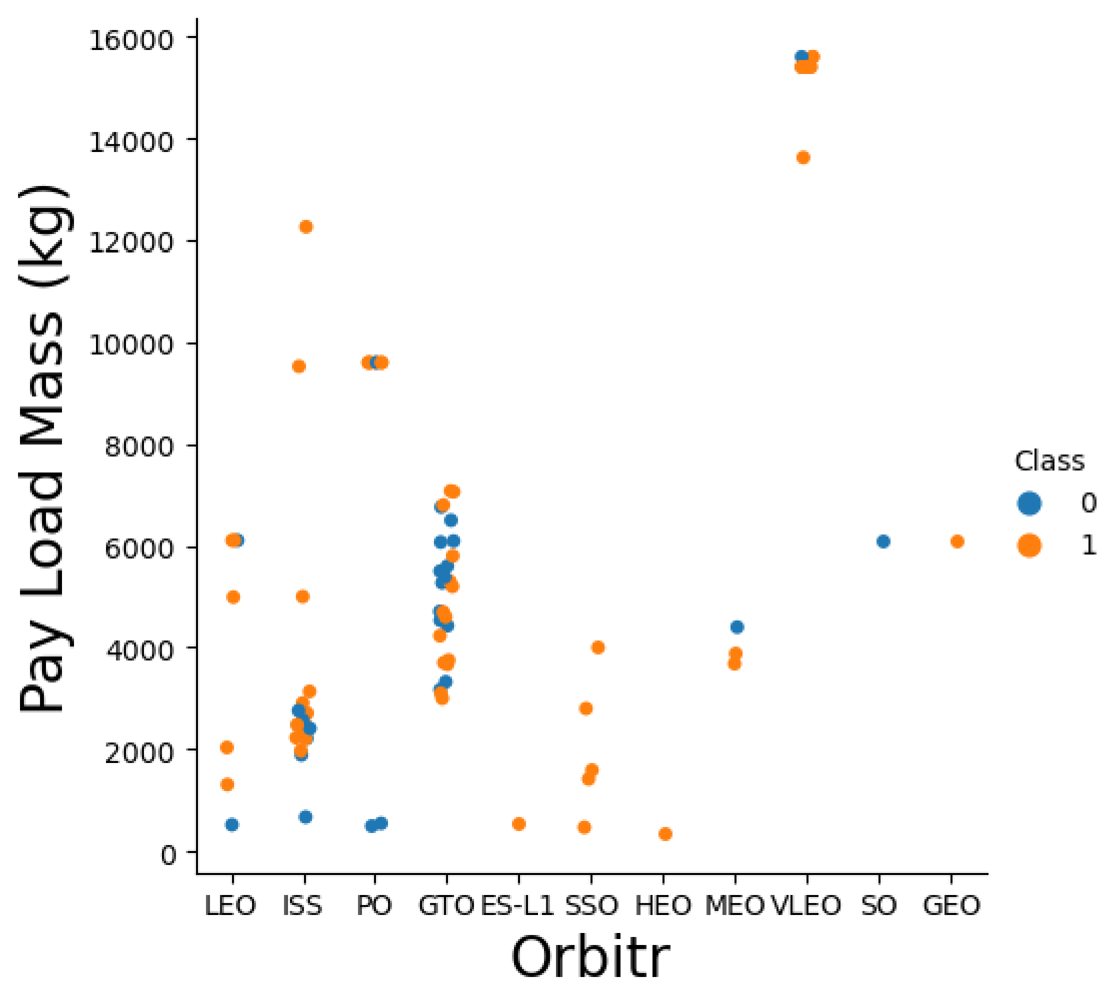
Flight Number vs. Orbit Type

- There seems to be no relationship between flight number when in GTO orbit



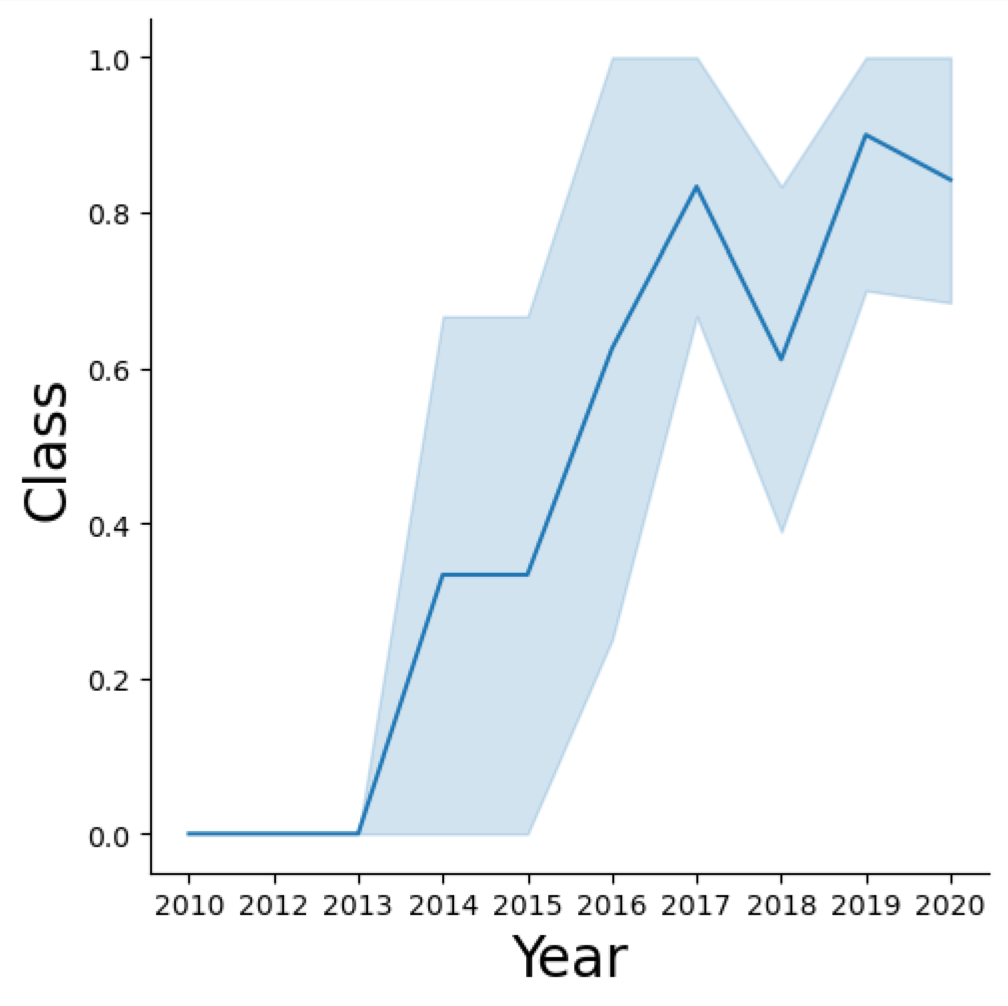
Payload vs. Orbit Type

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



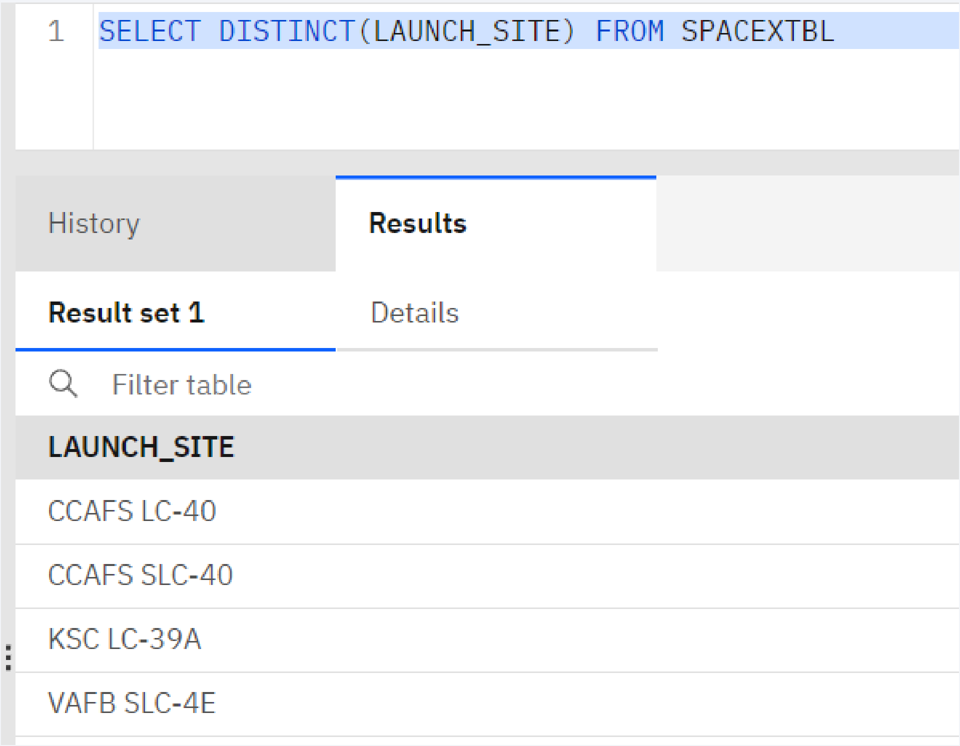
Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- There are four launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E



Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
1  SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

History | **Results**

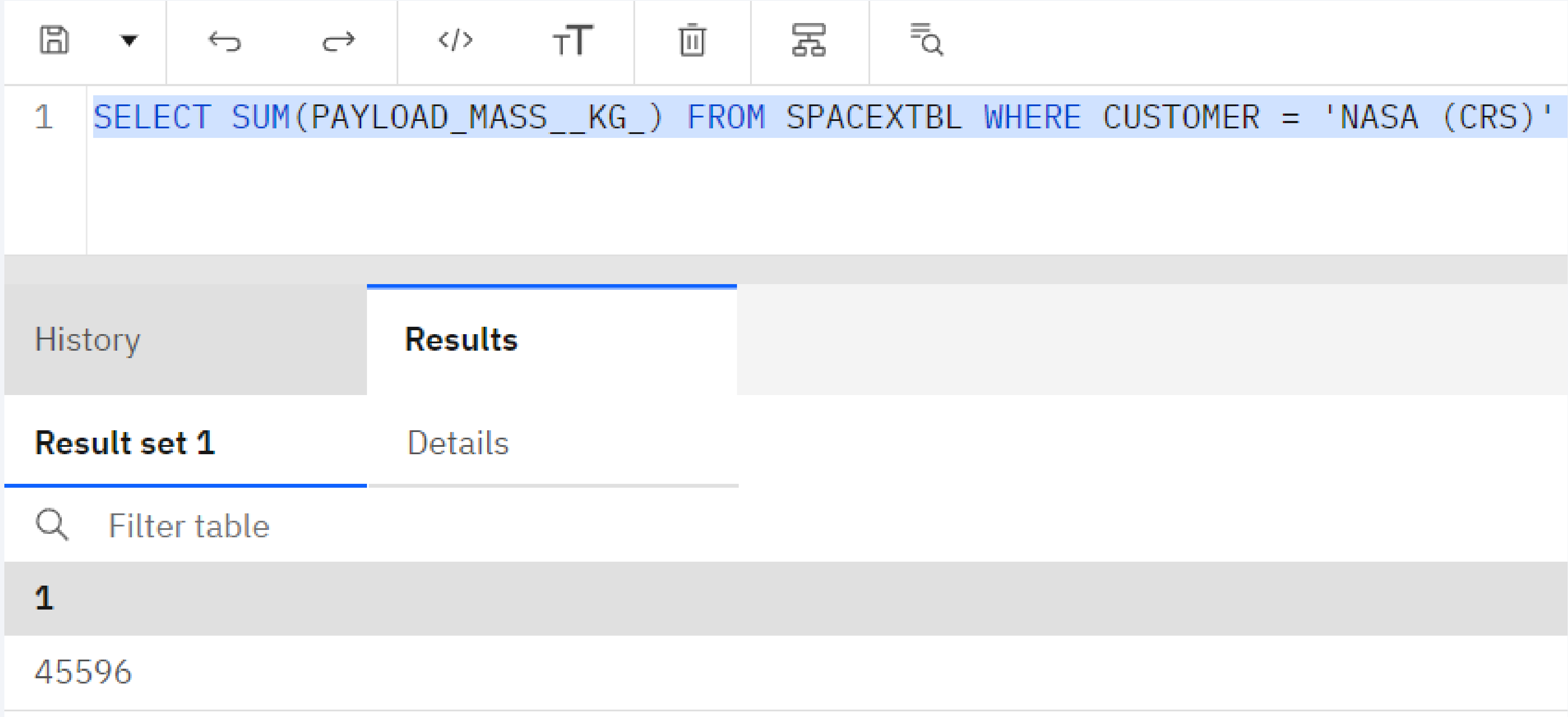
Result set 1 | Details

Filter table

DATE	TIME__UTC_	BOOSTER_VERSION	LAUNCH_SITE	PAYLOAD
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2

Total Payload Mass

- Total payload carried by boosters from NASA



Average Payload Mass by F9 v1.1

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

```sql
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

History | **Results**

Result set 1 | Details

Filter table

1
2928

First Successful Ground Landing Date

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

```
1  SELECT min(DATE) FROM SPACEXTBL WHERE Landing__Outcome = 'Success (ground pad)'
```

History | **Results**

Result set 1 | Details

Filter table

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
SELECT BOOSTER_VERSION
  FROM SPACEXTBL
  WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

History | **Results**

Result set 1 | Details

Filter table

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

```
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL FROM SPACEXTBL GROUP BY MISSION_OUTCOME
```

History | **Results** | SPACEXTBL

Result set 1 | Details

Filter table

MISSION_OUTCOME	TOTAL
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

```sql
SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

History | Results | SPACEXTBL

Result set 1 | Details

Filter table

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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```sql
SELECT MONTH(DATE) AS MONTH, LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
  FROM SPACEXTBL
  WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND YEAR(DATE) = 2015
```

History | **Results** | SPACEXTBL

Result set 1 | Details

Filter table Total:2

MONTH	LANDING__OUTCOME	BOOSTER_VERSION	LAUNCH_SITE
1	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
4	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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

```sql
SELECT *
FROM
  SPACEXTBL
WHERE
  LANDING__OUTCOME LIKE 'Success%' AND (DATE BETWEEN '2010-06-04' AND '2017-03-20')
ORDER BY DATE DESC
```

History | **Results** | SPACEXTBL

Result set 1 | Details

Filter table — Total:8

DATE	TIME__UTC_	BOOSTER_VERSION	LAUNCH_SITE	PAYLOAD	PAYLOAD_I
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034

Section 3

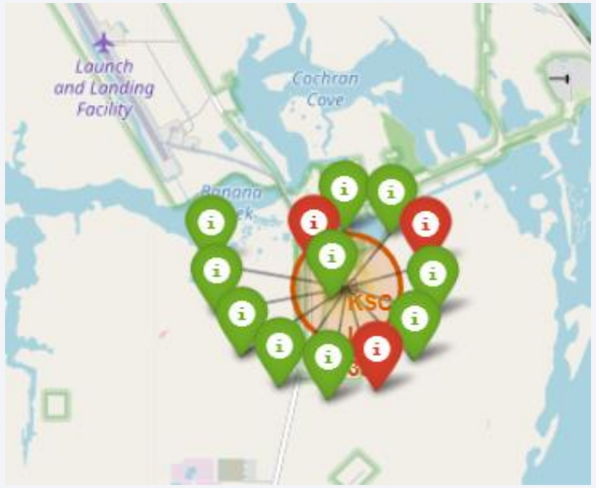
Launch Sites Proximities Analysis

All launch sites

- Launch sites are near sea.

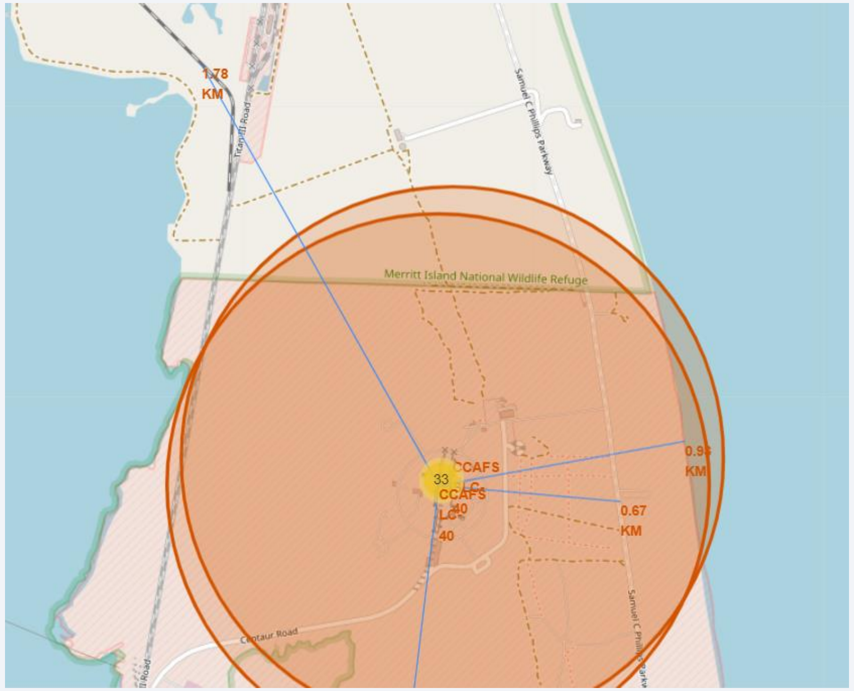
Success/Failed launches for each site

- The KSC LC-39A have the highest success rate.



Distances between a launch site to its proximities

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road

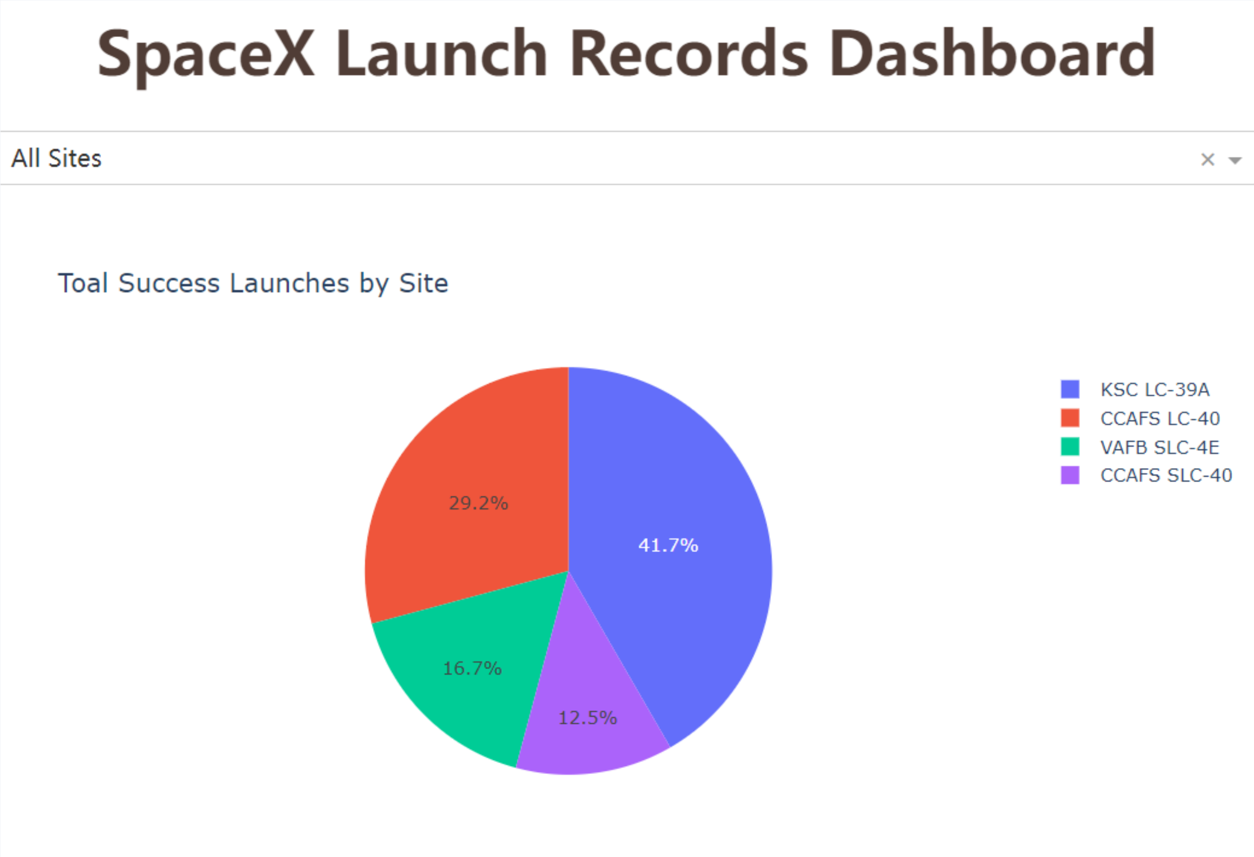


Section 4

Build a Dashboard with Plotly Dash

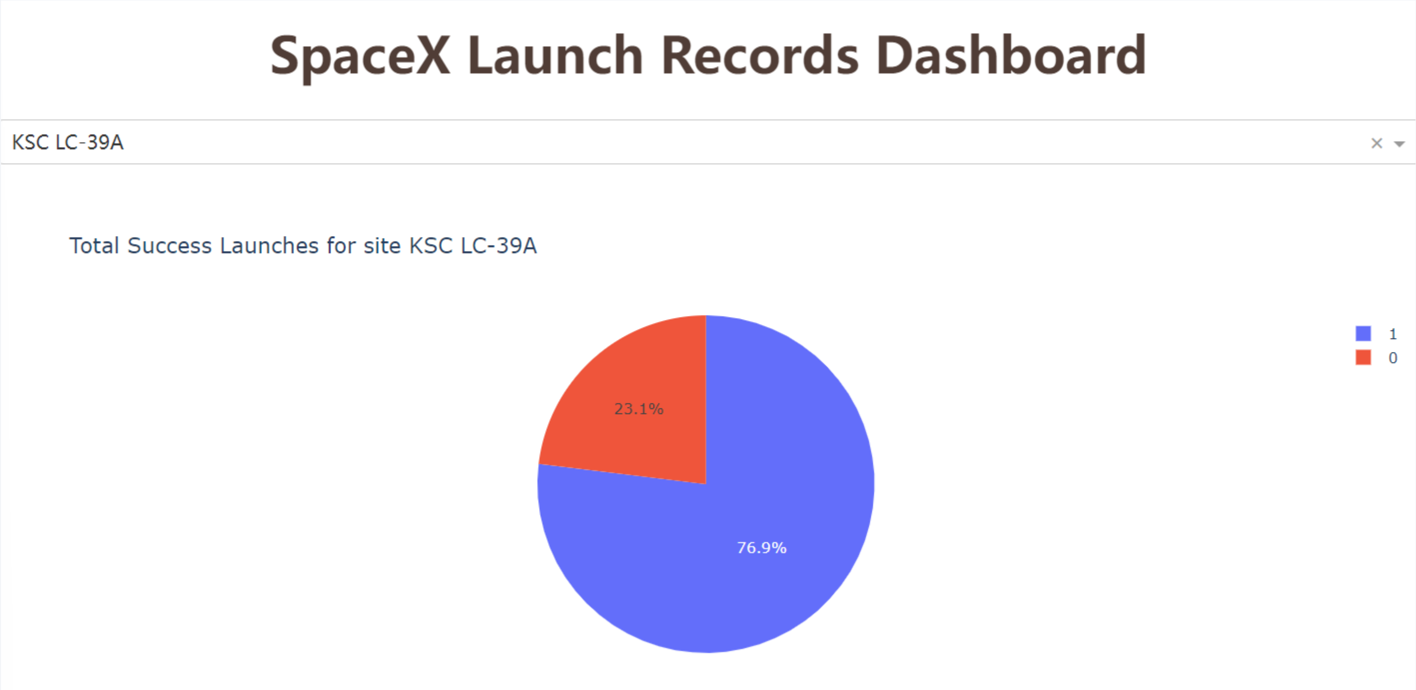
Successful Launches by Site

- The location from which launches are conducted appears to be a crucial factor in the success of space missions.



The Launch Success Ratio for KSC LC-39A

- This site has a success rate of 76.9% for launches.



Payload vs. Launch Outcome

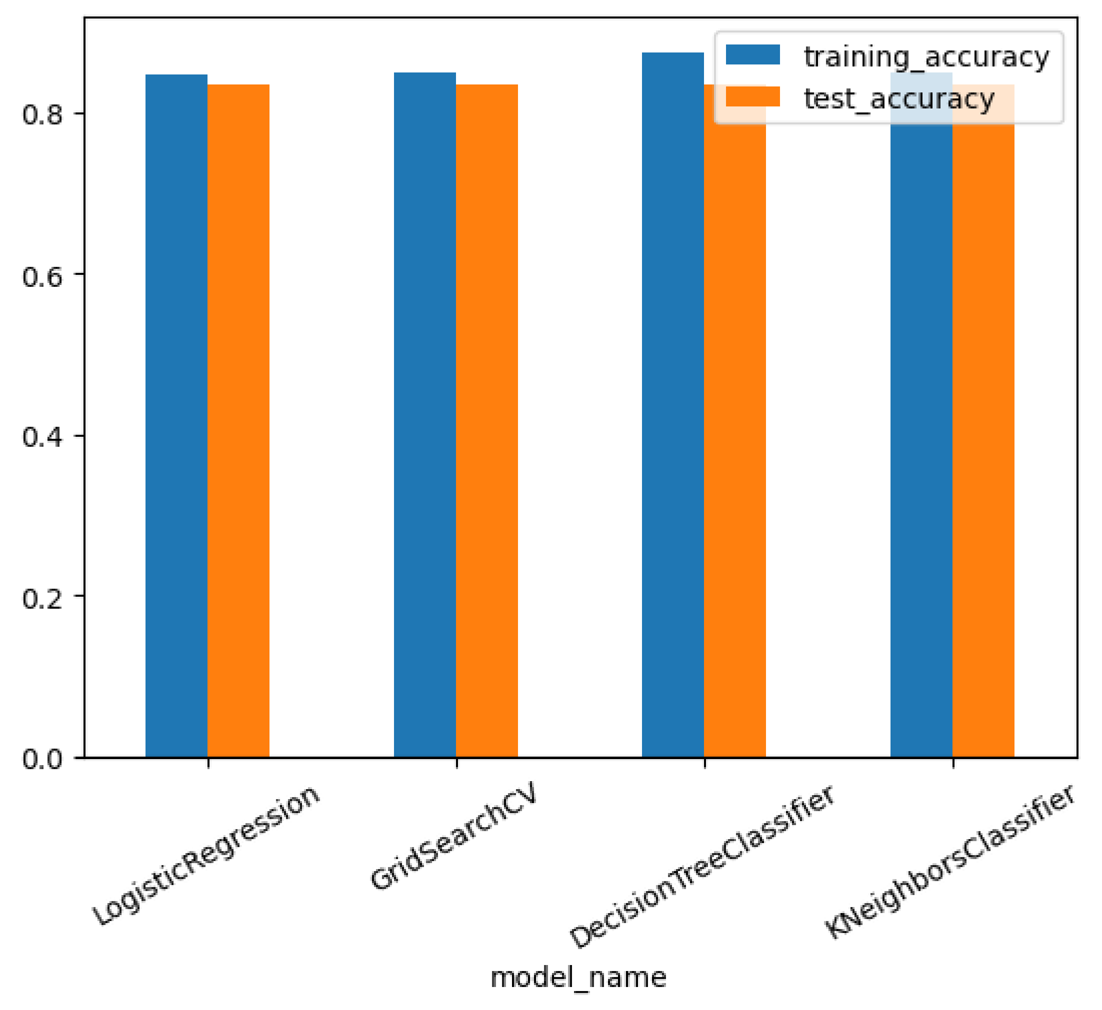
- The payload between 1500kg and 5500kg and FT boosters are the most successful combination.

Section 5

Predictive Analysis (Classification)

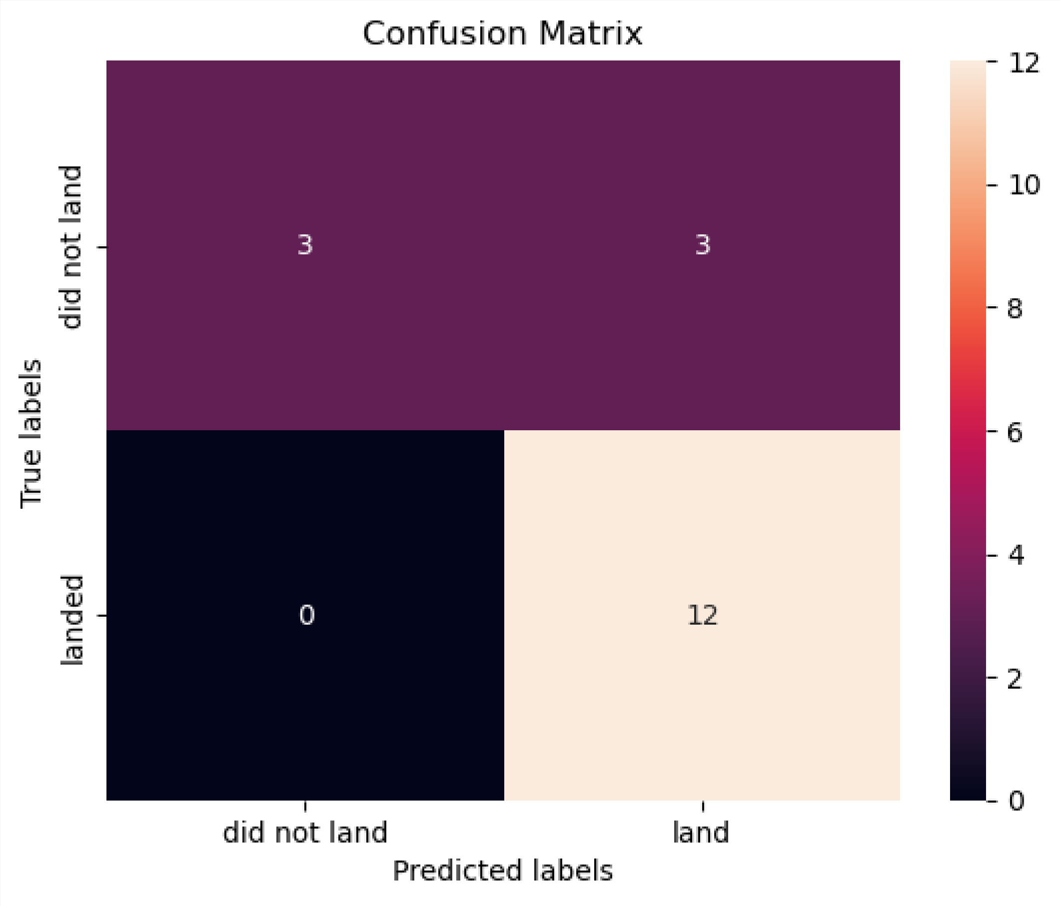
Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier.



Confusion Matrix

- The confusion matrix of the Decision Tree Classifier demonstrates its accuracy by displaying a higher number of true positives and true negatives compared to false positives and false negatives.



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve overtime, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings

Appendix

- Can’t connect notebook with IBM db2. So I finished SQL part on the cloud.

Thank you!