IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Iwuchukwu Johnpaul

15th May 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection(Web Scraping , Space X public API).
  - Data Wrangling and Analysis.
  - Data exploration.
  - Predictive analysis and Model classification.
- Summary of all results
  - Interactive data visualization.
  - Model Selection and conclusion.

Introduction

- Project background and context
  - With the current commercial space age, companies are making space travel affordable for everyone.
  - Reason SpaceX can do this is the rocket launches are relatively inexpensive, with a cost of $60 million as compared to others $165 million.
  - Hence, we can determine a successful first stage landing of Space X, due to the mission parameters like payload, orbit, and customer.
- Problems you want to find answers
  - To predict a completed lunch and successful landing of rockets.
  - To train a machine learning model and use public information to predict if SpaceX will reuse the first stage

Section 1

Methodology

Iwuchukwu Johnpaul

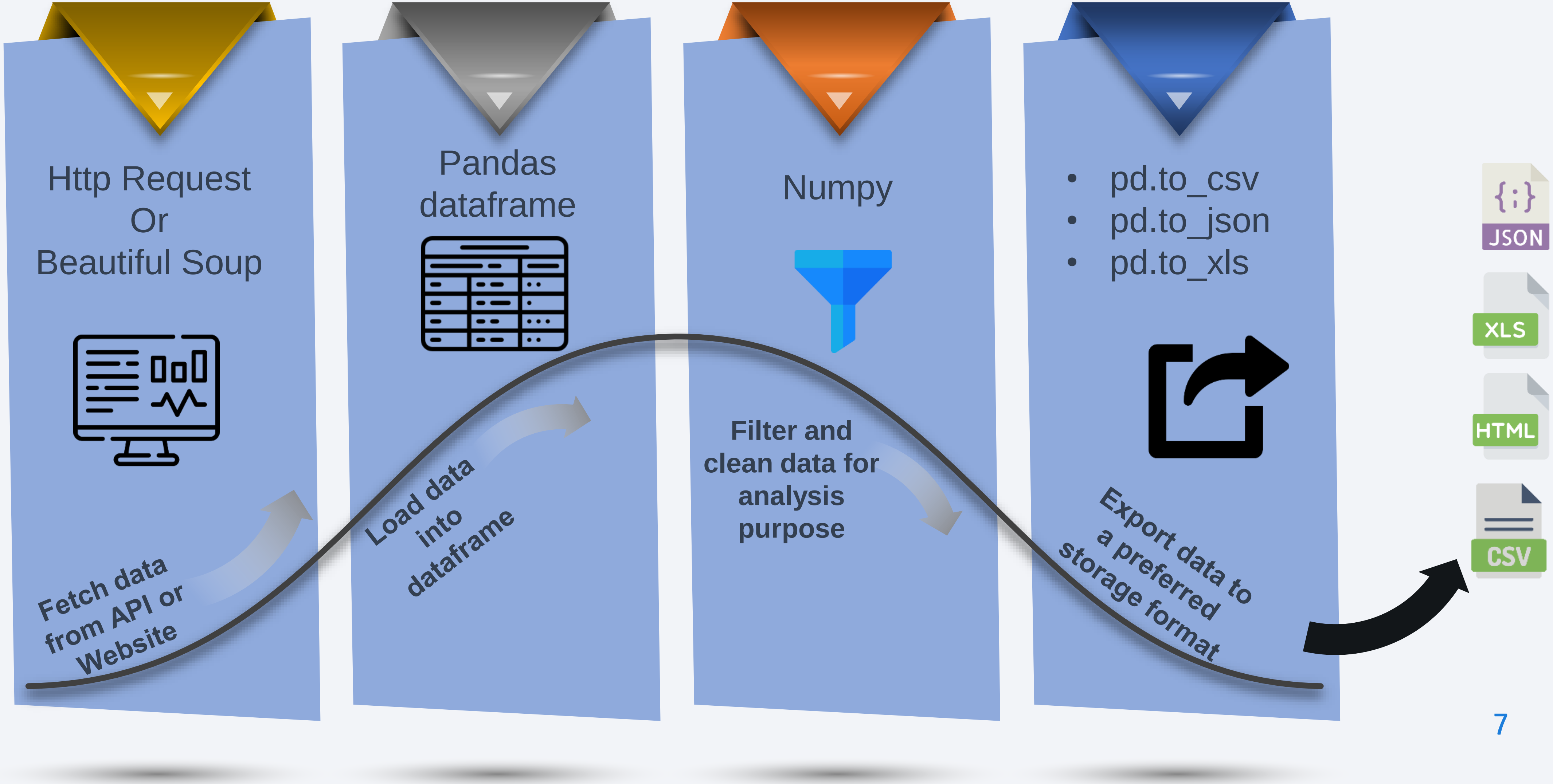
15th May 2022

Methodology

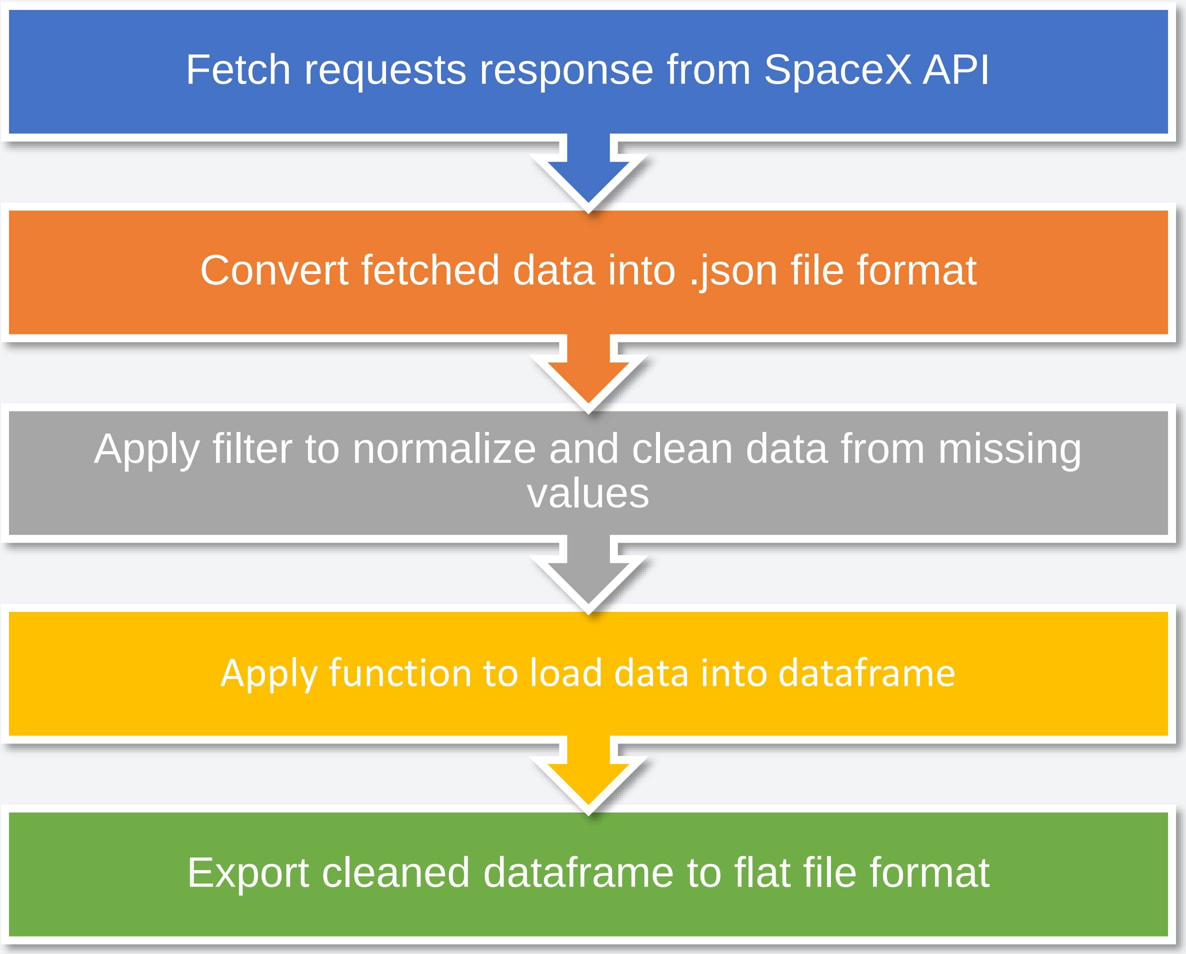
Executive Summary

- Data collection methodology:
  - SpaceX API and Web scraping wikipedia website
- Perform data wrangling
  - Numpy and Pandas
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Scikit-Learn and GridSearchCv

Data Collection



Data Collection – SpaceX API



[Click here for GitHub Link]

Data Collection - Scraping

Fetch requests response from SpaceX HTML

Using the Beautiful soup object to scrap data

Extract column and variables from HTML

Create and load data into the data frame

Export data frame to flat file format

[Click here for GitHub Link]

Data Wrangling

1. Evaluate the total number of lunches at the every site
2. Calculate the number and occurrence of each orbit
3. Calculate the number and occurrence of mission outcome per orbit type
4. Create a landing outcome label from Outcome column
5. Export the data frame to flat file format.

[Click here for GitHub Link]

EDA with Data Visualization

- Dataset visualized explained relationships between variables as listed:
  - FlightNumber vs. PayloadMass (scatter plot)
  - FlightNumber vs LaunchSite (scatter plot)
  - Payload vs Launch Site (scatter plot)
  - relationship between success rate and orbit type (bar plot)
  - FlightNumber and Orbit type (scatter plot)
  - Payload and Orbit type (scatter plot)
  - launch success yearly trend (line plot)

Click here for GitHub Link

EDA with SQL

Here we used the IBM Db2 storage to store the extracted dataset, while exploring the database with SQL queries on cloud jupyter notebook.
Summarizing the SQL queries performer:

- Load dataset into SQL database and create SQL connections with the IBM Db2 database.
- Evaluate the names of the unique launch sites in the space mission,
- Evaluate dataset Records (5) where launch sites begin with the string 'CCA'.
- Evaluate total payload mass carried by boosters launched by NASA (CRS).
- Evaluate average payload mass carried by booster version F9 v1.1,
- Find date when the first successful landing outcome in ground pad was achieved.
- Find the boosters name which have success in drone ship and have payload mass greater than 4000 but less than 6000,
- Evaluate the total number of successful and failure mission outcomes,
- List the names of the booster versions which have carried the maximum payload mass,
- List 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

Click here for GitHub Link

Build an Interactive Map with Folium

- Dataset visualized maps out different sites launch locations.
- Elements added to map:
    - circle and marker for each launch site, added as point of attention at map,
    - marker cluster – easy way to show number of launches at each site,
    - mouse position- to get coordinates,
    - poly lines (lines between site and coastline)

Click here for GitHub link

Build a Dashboard with Plotly Dash

- The purpose of adding interactive dashboard was to give the users opportunity to decide which information one would like to see.
- Features add to dashboard:
  - dropdown menu was used to let one select launch site and information related
  - intractable bar for payload mass
  - scatter plot that show correlation between payload mass and landing result
  - Pie chart was used to show proportion between successful landing and unsuccessful.

Click here for GitHub Link

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Click here for GitHub Link

Create a NumPy array from Class, make a Pandas series Y.

Standardize the data, then reassign it to the variable using the transform method.

Split datasets , Set the parameter test size to 0.2 and random state to 2

Create the model (logistic regression, SVM, decision tree, KNN) with the GridsearchCv object.

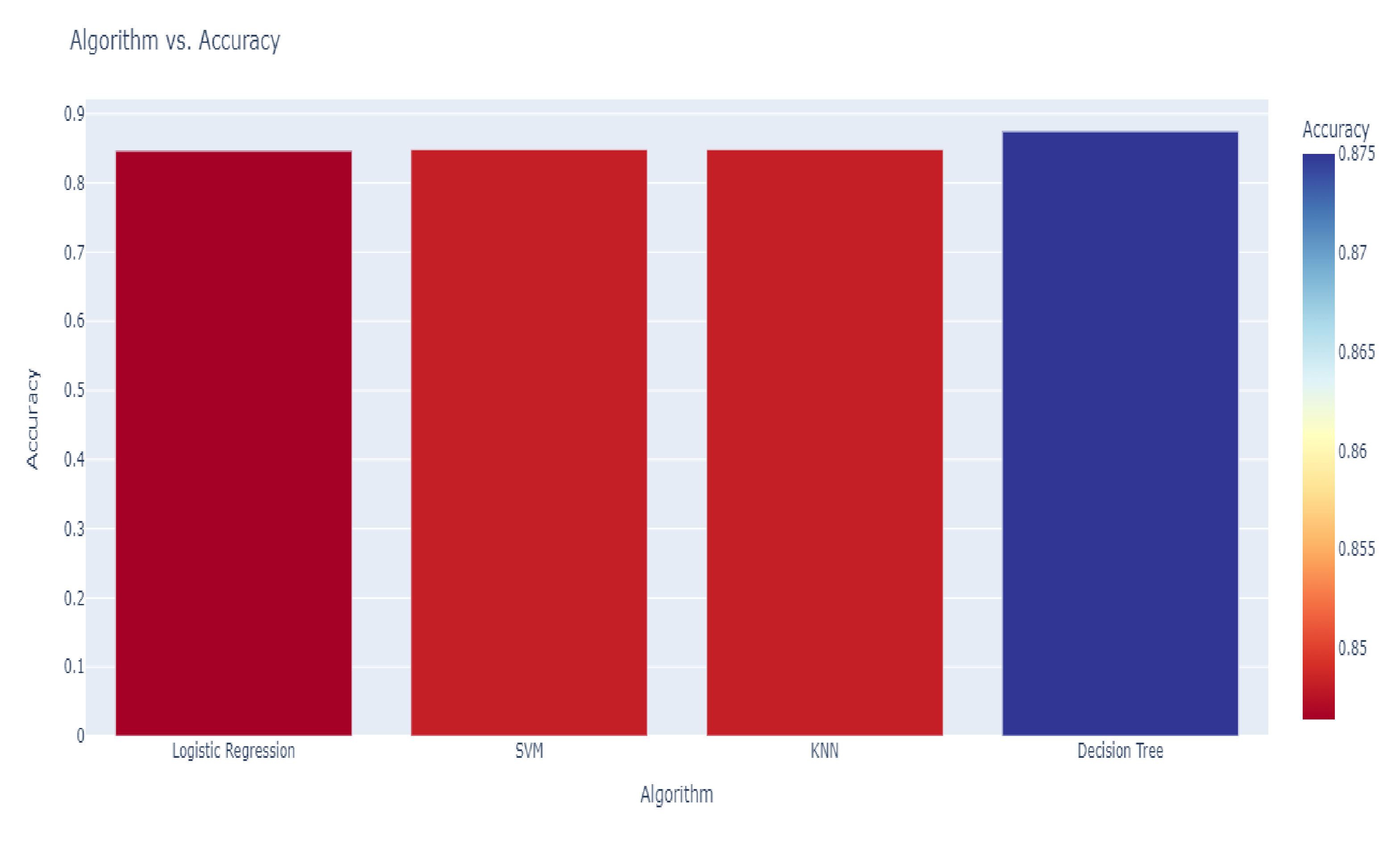
Calculate the accuracy score for each model

Create a confusion matrix for each model

Analyse the accuracy score and Select the best performing model

Results

- Predictive analysis results : all models have similar accuracy 0.86 in approximate, except for Decision tree which has the highest of 0.87



Section 2

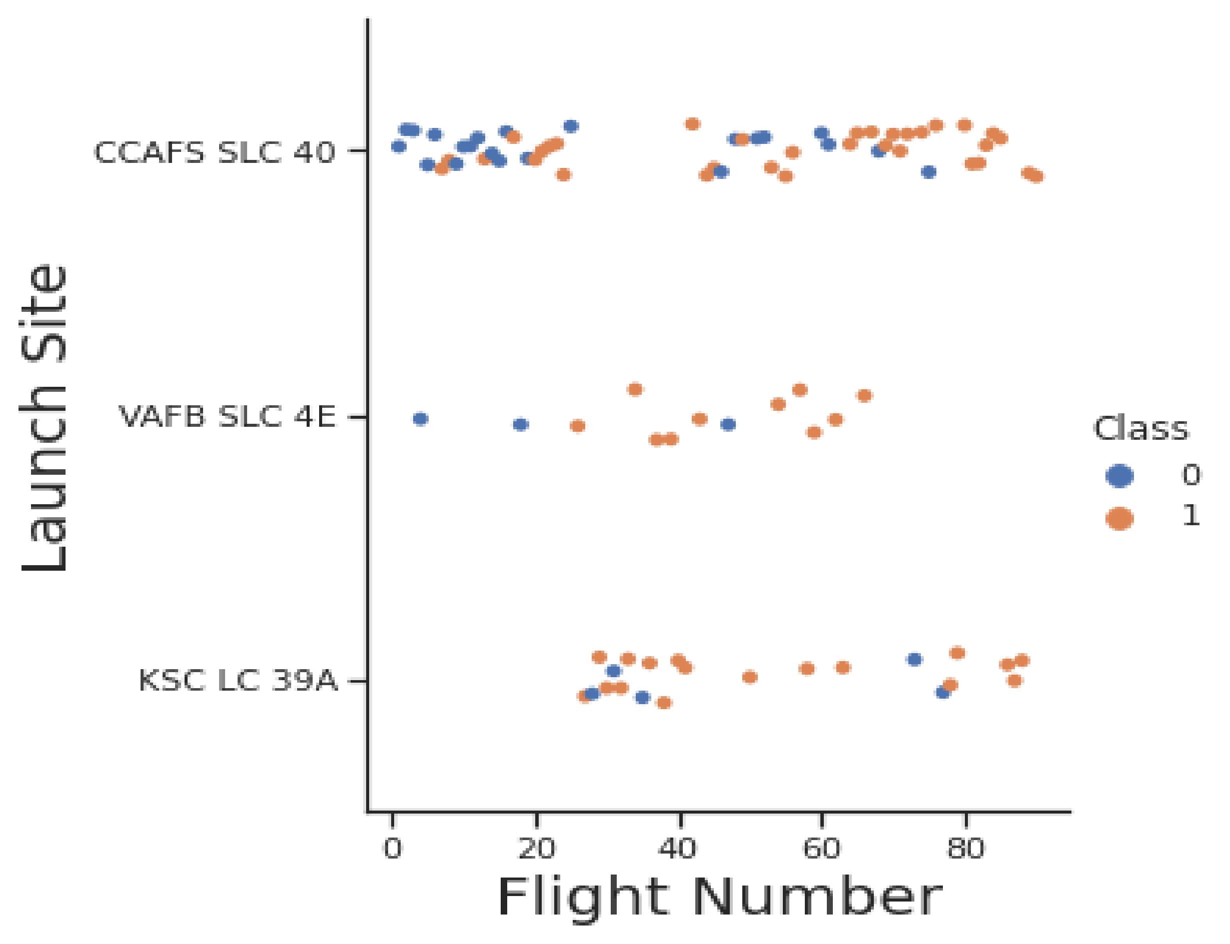
Insights drawn from EDA

Iwuchukwu Johnpaul

15th May 2022

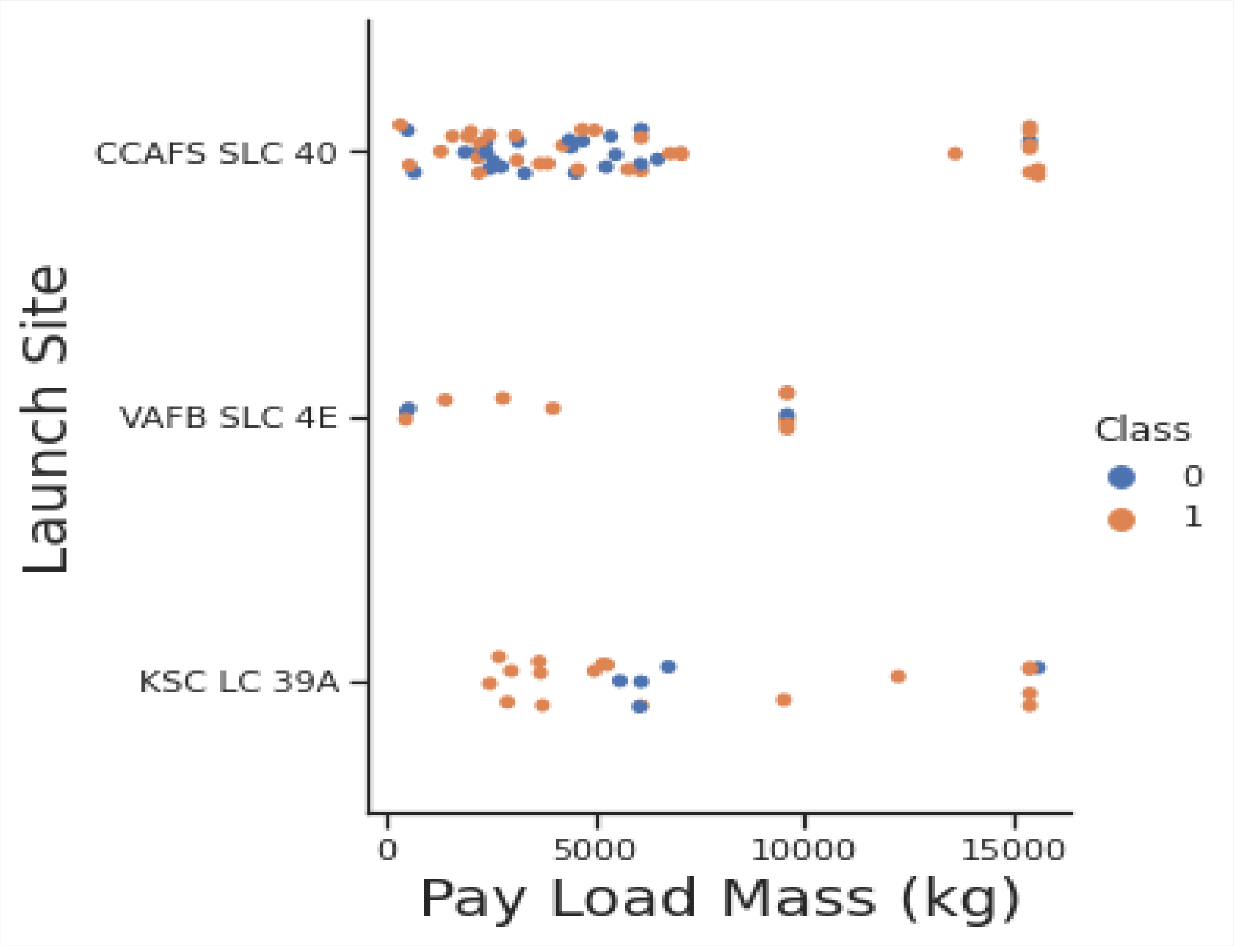
Flight Number vs. Launch Site

- From the plotted graph, the amount of successful landing increase as flight numbers increases and “CCAFS SLC 40” appears to be the most often used platform.



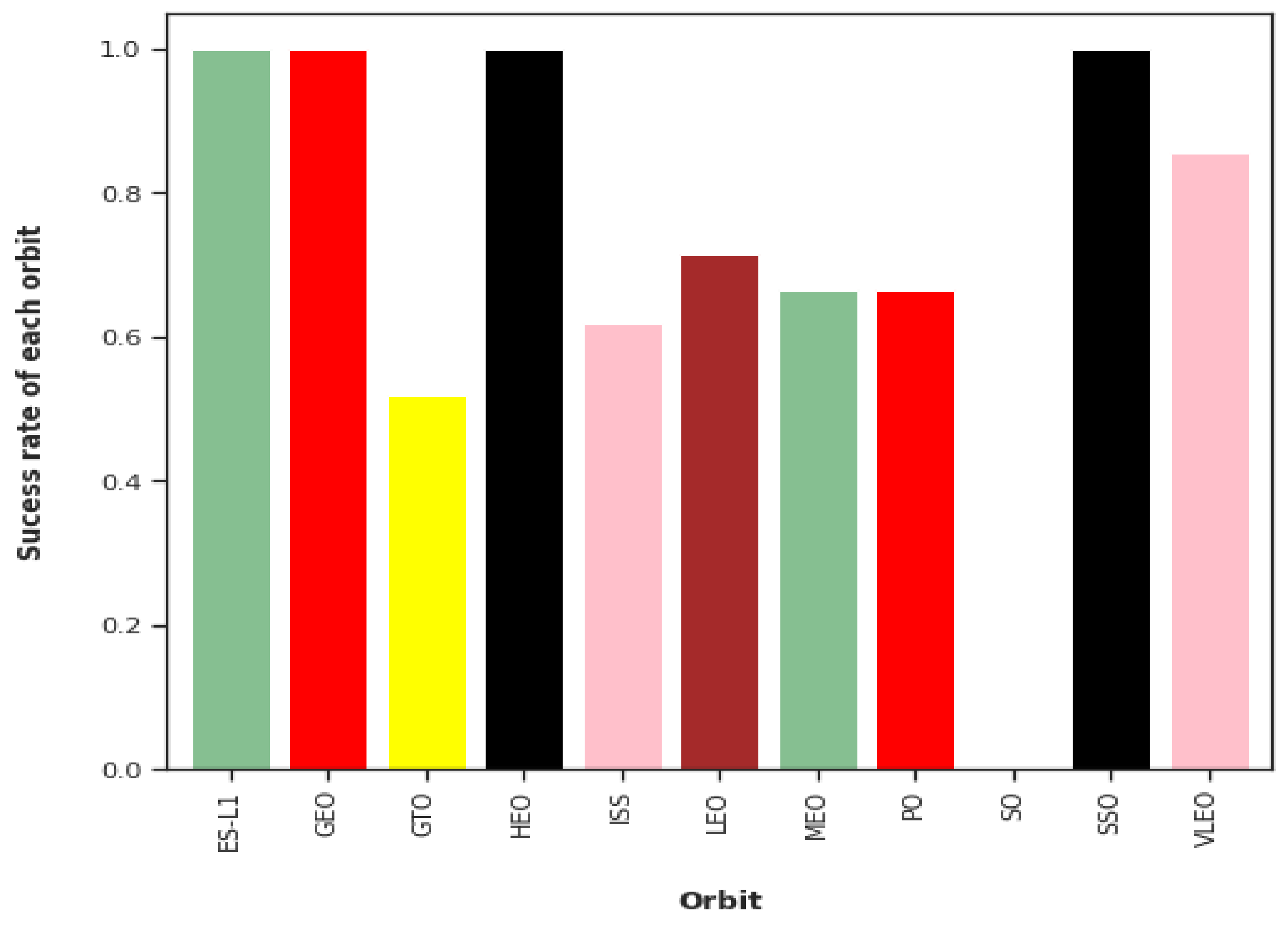
Payload vs. Launch Site

- Most of the launches have payload below 8000kg
- Most of the launches with payload over 8000kg were successful



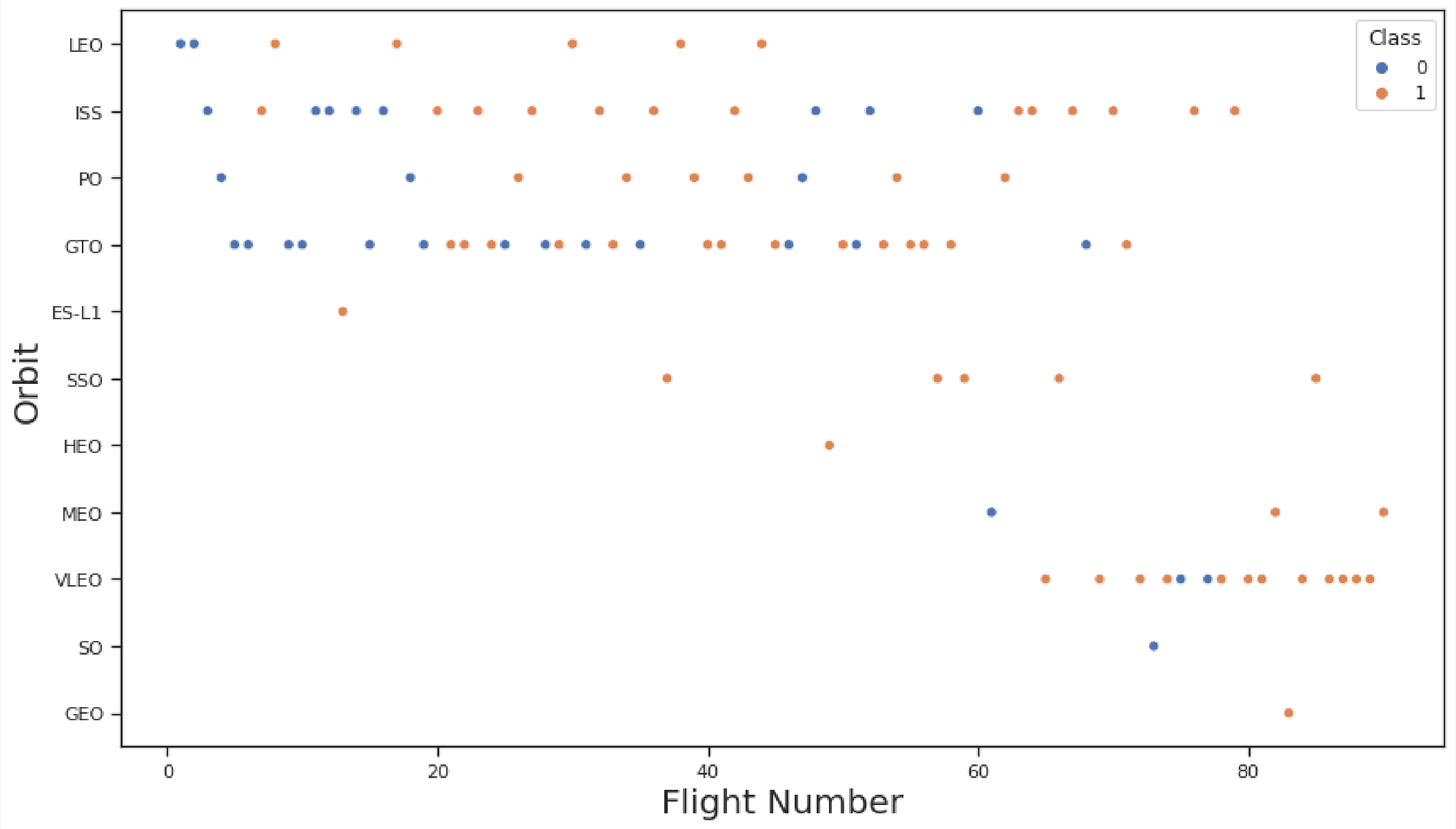
Success Rate vs. Orbit Type

- All other orbits have success rate over 50%, except for SO orbit which do not experience any launching.



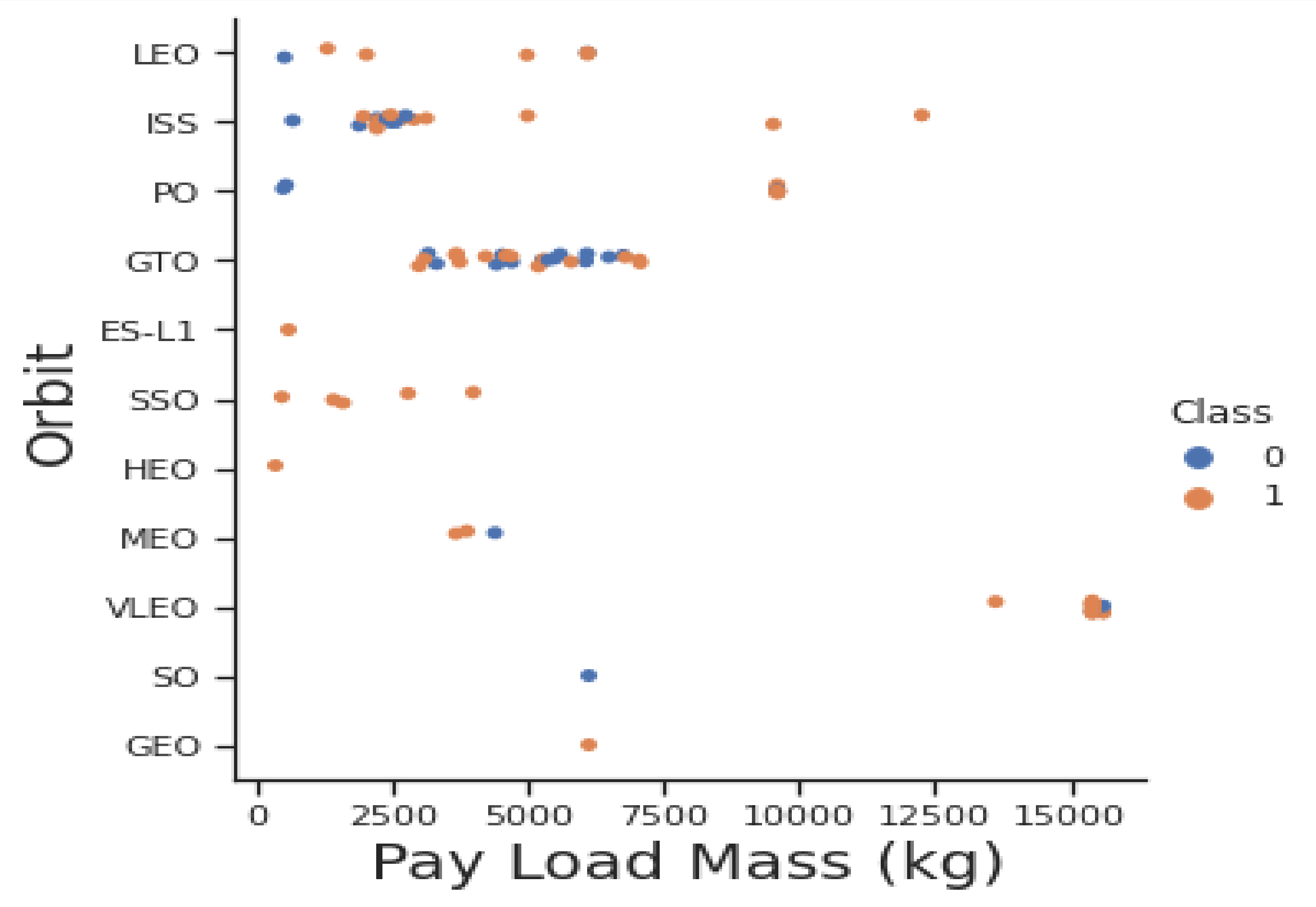
Flight Number vs. Orbit Type

- There was an orbit type changes with flight number increase.



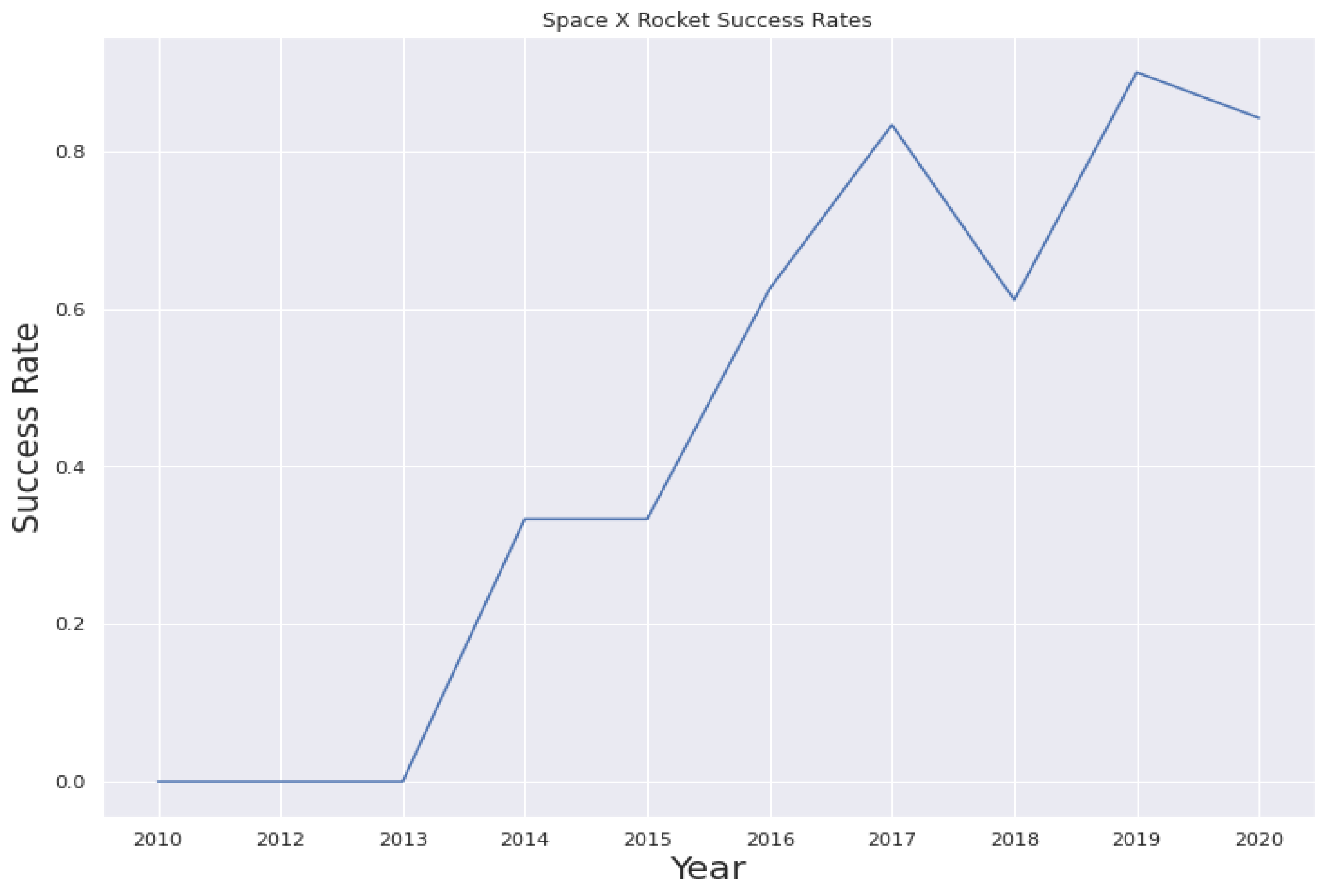
Payload vs. Orbit Type

- There is the correlation between the payload and orbit chosen.



Launch Success Yearly Trend

- Success rate increased in all the years except in 2018 which dropped to 0.61
- Launches between 2010 and 2013 were completed unsuccessfully



All Launch Site Names

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

In [4]:
```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEX;
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[4]:

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

- In dataset there are 4 unique launch sites names

Launch Site Names Begin with 'CCA'

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

In [39]:

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

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[39]:

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

- 5 records where launch sites begin with `CCA`

Total Payload Mass

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

```
[6]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total payload mass by NASA(CRS)" FROM SPACEX WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

t[6]:

Total payload mass by NASA(CRS)
45596

- The total payload carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

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

[7]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS "Average payload mass by version F9 v1.1" FROM SPACEX WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

[7]:

Average payload mass by version F9 v1.1
2928

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

First Successful Ground Landing Date

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

Hint:Use min function

[9]:

```
#Make sure to correct the error in the column name from 'Landing _Outcome' to 'Landing_Outcome' before loading into the database
#sql will not read the column name with a space
%sql SELECT MIN(DATE) AS "First Succesful Landing Outcome in Ground Pad" FROM SPACEX WHERE LANDING_OUTCOME = 'Success (ground pad)';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

[9]:

First Succesful Landing Outcome in Ground Pad
2015-12-22

- The first successful landing outcome on ground pad was on 22nd of December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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 SPACEX WHERE LANDING_OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

t[19]:

booster_version
F9 FT B1022

- There are four boosters version with payload mass between 4000 and 6000 kg that have successfully landed on drone ship

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

[12]:
```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Successful Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Success%';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

[12]:

Successful Mission
100

[13]:
```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Failure Mission" FROM SPACEX WHERE MISSION_OUTCOME LIKE 'Failure%';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

[13]:

Failure Mission
1

- There were 100 successful mission and only 1 failure.

Boosters Carried Maximum Payload

[15]:

```
%sql SELECT DISTINCT BOOSTER_VERSION AS "Booster Versions which carried the Maximum Payload Mass" FROM SPACEX \
WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX);
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

t[15]:

Booster Versions which carried the Maximum Payload Mass
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

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

2015 Launch Records

[21]:
```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEX WHERE DATE LIKE '2015-%' AND \
LANDING_OUTCOME = 'Failure (drone ship)';
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

t[21]:

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

- The failed landing outcomes in drone ship in 2015

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

[27]:

```
%sql SELECT LANDING_OUTCOME as "Landing Outcome", COUNT(LANDING_OUTCOME) AS "Total Count" FROM SPACEX \
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY  LANDING_OUTCOME \
ORDER BY COUNT(LANDING_OUTCOME) DESC;
```

```
 * ibm_db_sa://tbp10921:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

t[27]:

Landing Outcome	Total Count
No attempt	9
Success (ground pad)	6
Failure (drone ship)	5
Controlled (ocean)	3
Success (drone ship)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- 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

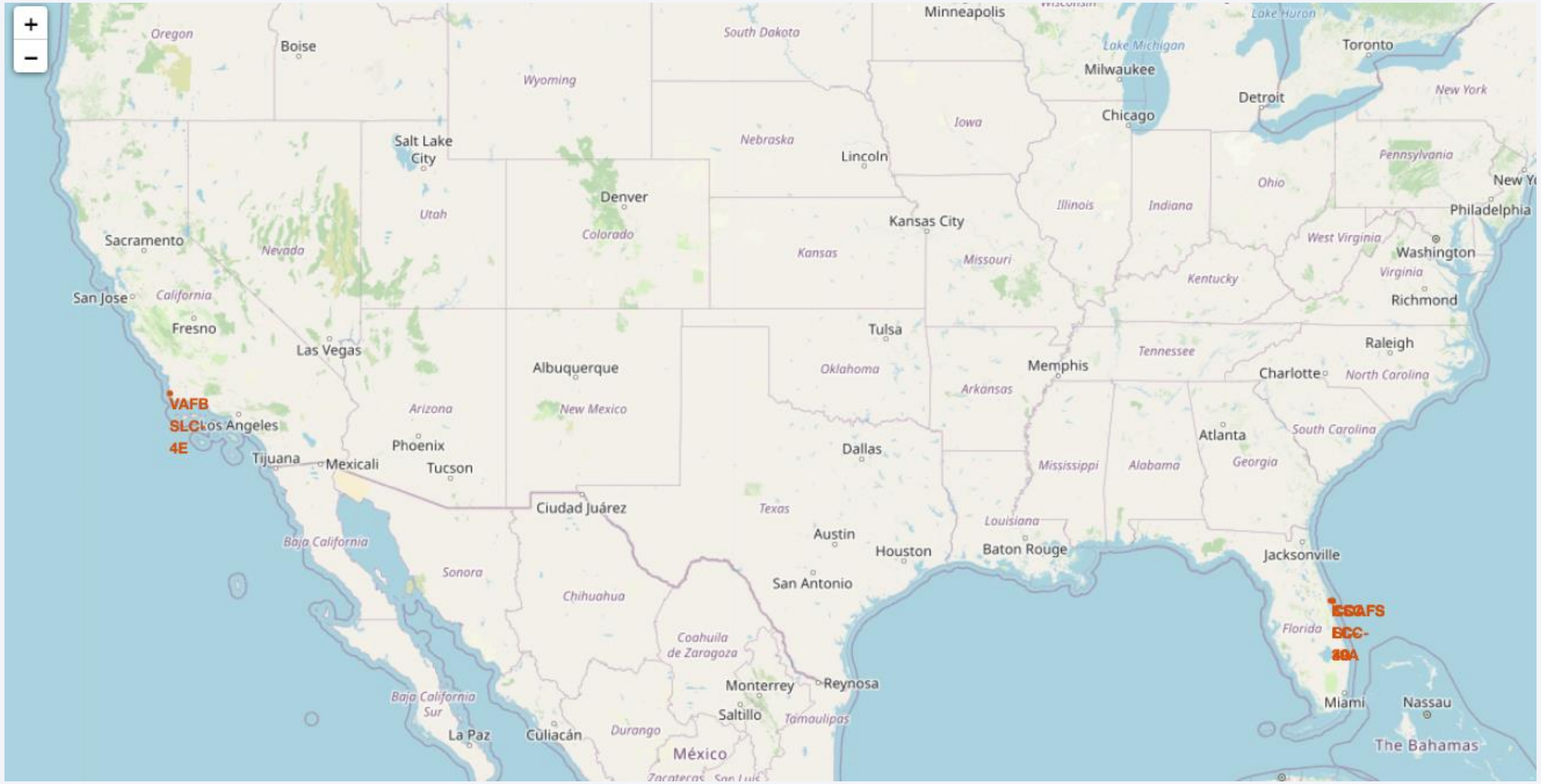
Section 3

Launch Sites Proximities Analysis

Iwuchukwu Johnpaul

15th May 2022

Launch Site Locations



- All launch sites are located at coastline

Landing outcome

- Coordinate landing outcome of 3 successful outcomes and 4 unsuccessful

Coastline distance – Folium Map

- Showing the coastline distance from launch site

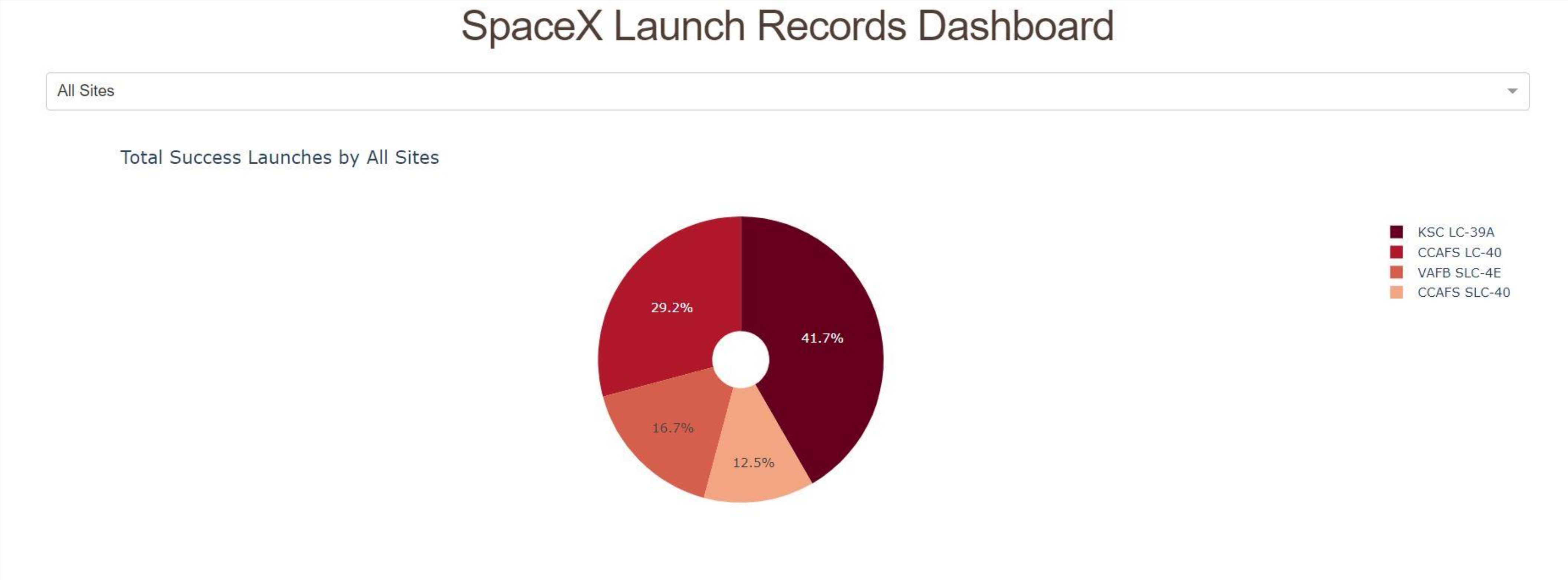
Section 4

Build a Dashboard with Plotly Dash

Iwuchukwu Johnpaul

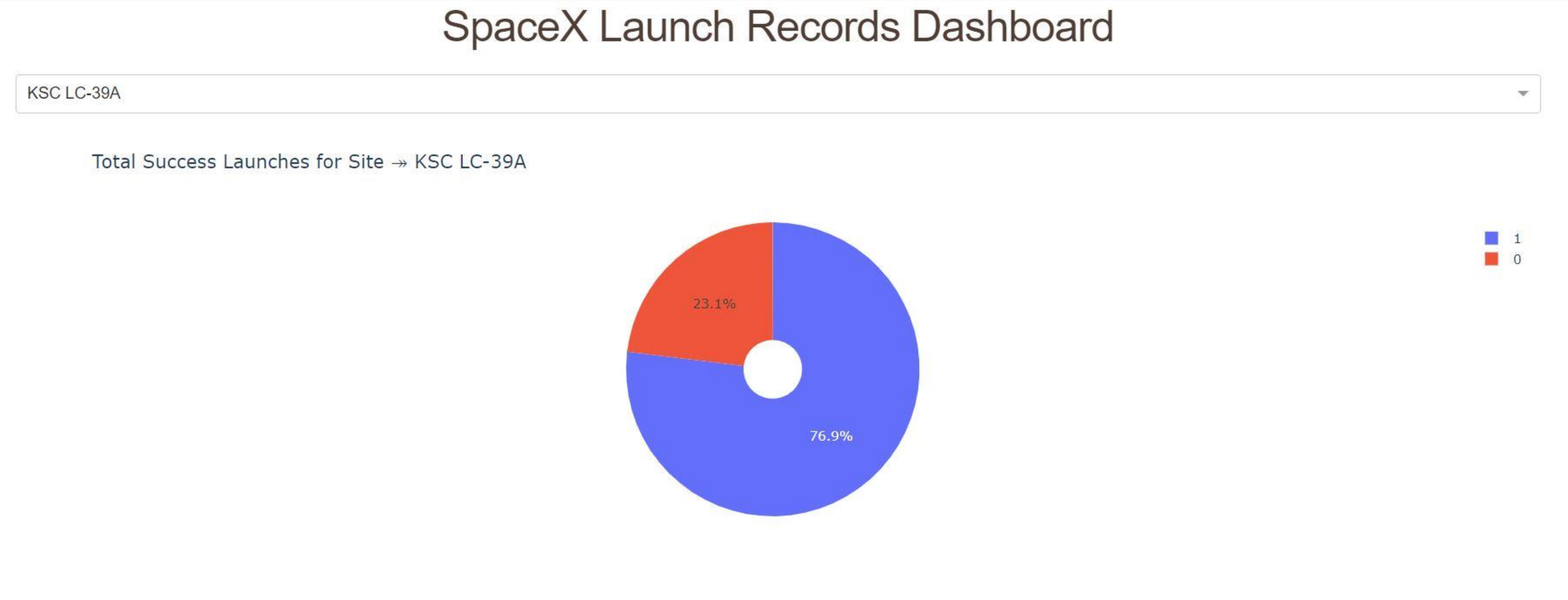
15th May 2022

All successful Launch sites



KSC LC-39A is the site with the highest number of successful launches ~42%

Highest Launch Ratio



Approximately 77% of launches at KSC LC-39A was completed successfully.

Payload mass relationship with success rate

In payload range between 2000 – 5500 kg, the biggest number of successful launches was for FT booster.

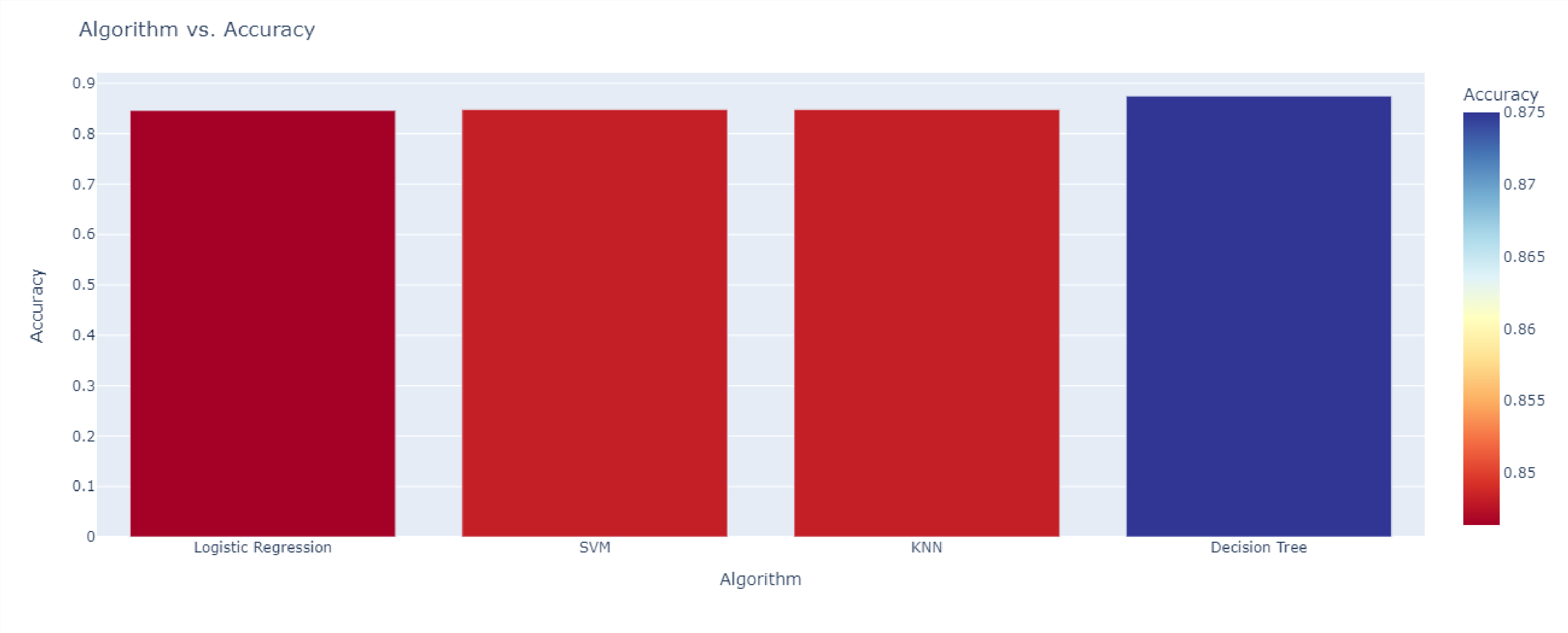
Section 5

Predictive Analysis (Classification)

Iwuchukwu Johnpaul

15th May 2022

Classification Accuracy



- Predictive analysis results : all models have similar accuracy 0.86 in approximate, except for Decision tree which has the highest of 0.87

Confusion Matrix

- Decision Tree - confusion matrix
- 3 unsuccessful landing was predicted correctly,
- 3 successful landing was predicted wrongly,
- 11 successful landings was predicted correctly

Conclusions

- The purpose of this project was to predict if the Stage 1 of the Falcon rocket land successfully or unsuccessfully.
- These information are required by Space Y (competitor of Space X) that want to send rocket cheaper than Space X.
- Launches data was received from Wikipedia and Space X API
- The best model predicted the outcome of landings with accuracy about 0.87

Appendix

- GitHub: Click Here
- SpaceX Data: Click Here
- Wikipedia: Click Here
- NASA: Click Here
- Coursera: Click Here

Thank you!

For you time