IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anthony Ford
1/15/2021

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Interactive map with Folium
  - Dashboard with Plotly Dash
- Summary of all results
  - Reviewing the data analysis results
  - Demo the interactive analytics
  - Reviewing the results

Introduction

- Project background and context
  - From the data collected on the SpaceX web site we predicted if the Falcon 9 rocket launches will land successfully. SpaceX advertises that the Falcon 9 rocket can provide launches at a cost of 62 million USD vs other companies with an average cost near 165 million USD. SpaceX states being able to reuse the first stage of the rocket greatly reduces the launch costs. We will be looking into if the first stage lands. The data can help determine the launch cost for an alterative company looking to compete with SpaceX' s launch pricing.
- Open Enquiries
  - What influences an successful landing?
  - Is there an certain variables that effect the rockets landing?
  - What conditions need to be achieved to get the best landing results?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from the SpaceX website with a Rest API
  - The API gave us the launch dates, the rocket used, payload, launch location, etc.
- Perform data wrangling
  - We are able to remove irrelevant data
- 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

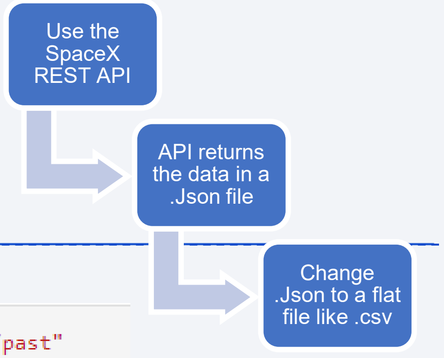
Data Collection

The Collection Process

- Launch data was collected from the SpaceX website with the REST API
- The data was filtered to show launches, rocket type, payload, etc.
- The goal was to use the data to learn if the rocket will land and be reused.

- The data was collected from the SpaceX REST API (api.spacexdata.com/v4/rockets/)
- Data was also collected by web scrapping Wikipedia with BeautifulSoup

Data Collection – SpaceX API



- Here is the process followed to collect the data from the .json file and converted the relevant data to a .cvs file

- The Jupiter notebook located here: [GitHub Data]()

1. Check the connection

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

2. Get the data in .Json File

```
def getBoosterVersion(data):
    for x in data['rocket']:
        response = requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
        BoosterVersion.append(response['name'])
```

3. Normalize the .Json file

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

4. Add a dictionary to the data frame

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

5. Export to a .cvs file

```
data_falcon9.to_csv('dataset_part\_1.csv', index=False)
```

Data Wrangling

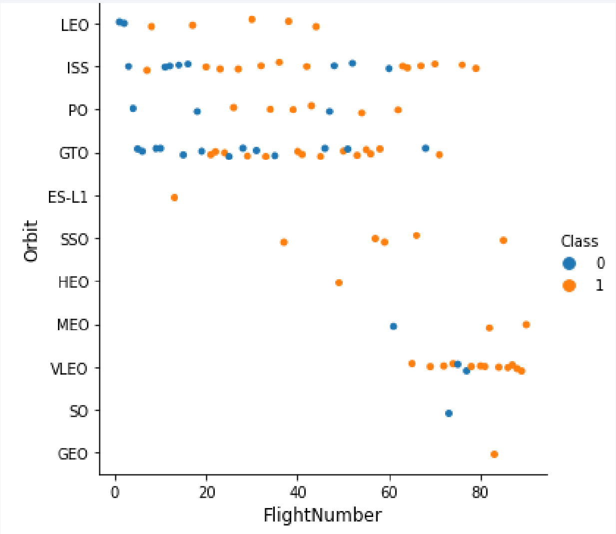
The data on the landings was presented as true or false data if the booster landed. The data was converted then converted to training data with 1 representing the booster landed and 0 for if the booster did not successfully land.

GitHub URL File

EDA with Data Visualization

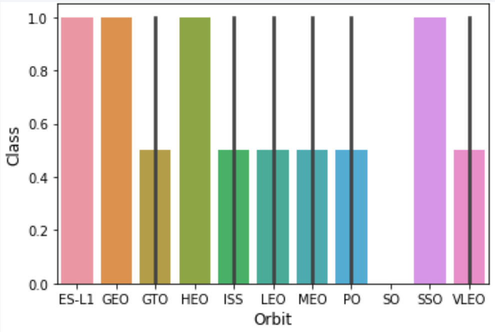
The Scatter Plots Generated:

- Flight Number vs. Payload Mass
- Flight Number vs Launch Site
- Payload and Launch Site
- Flight Number and Orbit type
- Payload and Orbit type



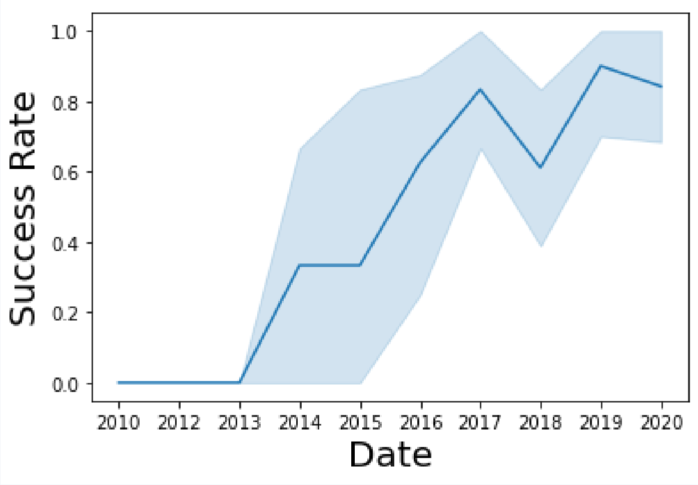
Bar Charts:

- Success Rate of Each Orbit Type



Line Graphs:

Launch Success Yearly Trend



[Get Hub URL]()

EDA with SQL

- Names of the unique launch sites per space mission
- Display 5 records where Launch site begins with CCA
- Total play load mass carried by boosters launched by NANA (CRS)
- Average Payload mass carried by booster F9 v1.1
- Date when the first successful landing was achieved
- Name of boosters which have successful drone ship landing with payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Name of booster with maximum payload mass
- List of failed drone ship landings with booster versions and launch site in 2015
- Rank the landing outcomes between June 4, 2010 to march 20, 2017 in descending order.

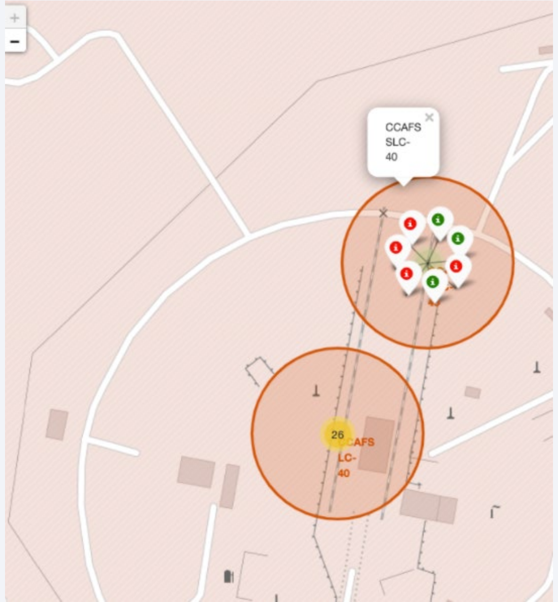
Interactive Map with Folium

In analyzing the Launch Data an interactive map was created. Using the longitude and Latitude coordinates of each launch site a circle marker and label was added to each launch site.

Using the launch data we assigned markers with MarkerCluster() to show the **Successful (1)** and **Failure (0)** outcomes with **Green** or **Red** Markers.

We also calculated the distance between landmarks to see if there was a correlation to landings

Examples: railroads, highways, coastlines, and cities



GitHub Data

Predictive Analysis (Classification)

Building the Model

1. The data was loaded into Pandas and NumPy
2. Data was transformed
3. Data was split into a training set and test data set
4. Check the number of test samples were created
5. Determined the most effective machine learning logarithm for the dataset
6. Set parameters and algorithms
7. Fitted our datasets and trained the data

Validating the Model

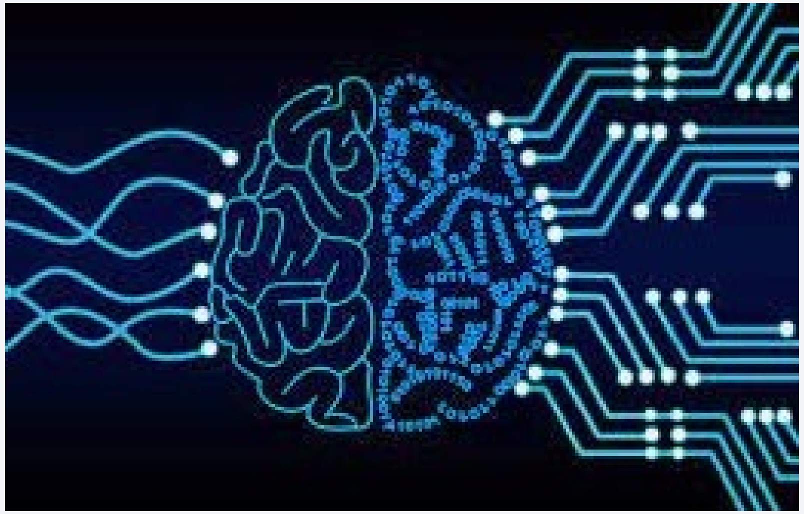
1. Checked the accuracy
2. Input the hyperparameters for each algorithm
3. Plotted the Confusion Matrix

Improving the Model

1. Tuning

Best Preforming Model

1. The best preforming model will have the greater accuracy score.



[GitHub File]

Results

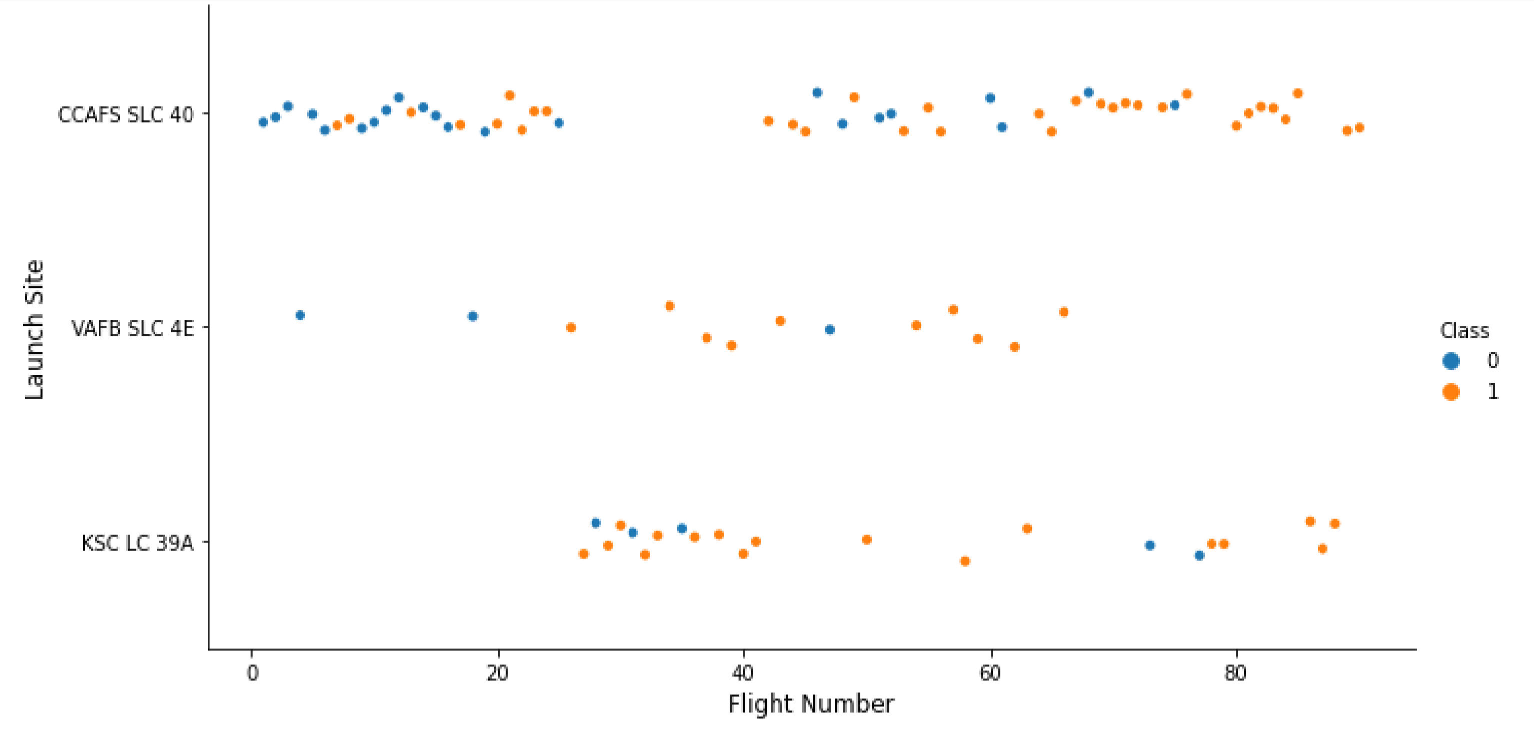
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

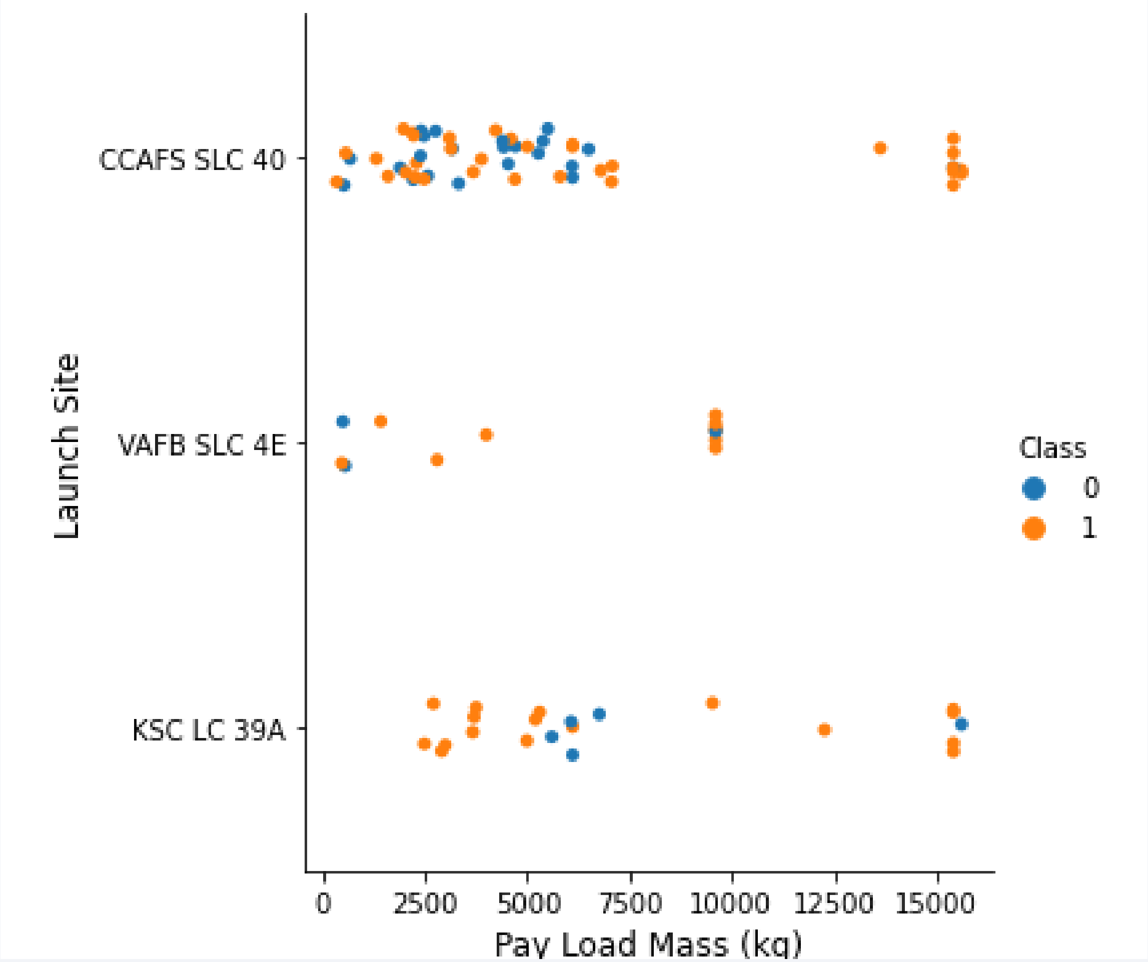
Flight Number vs. Launch Site

- We can see the success rate has increased over time at the launch sites
- CCAFS SLC 40 looks to be where most test flights launched.



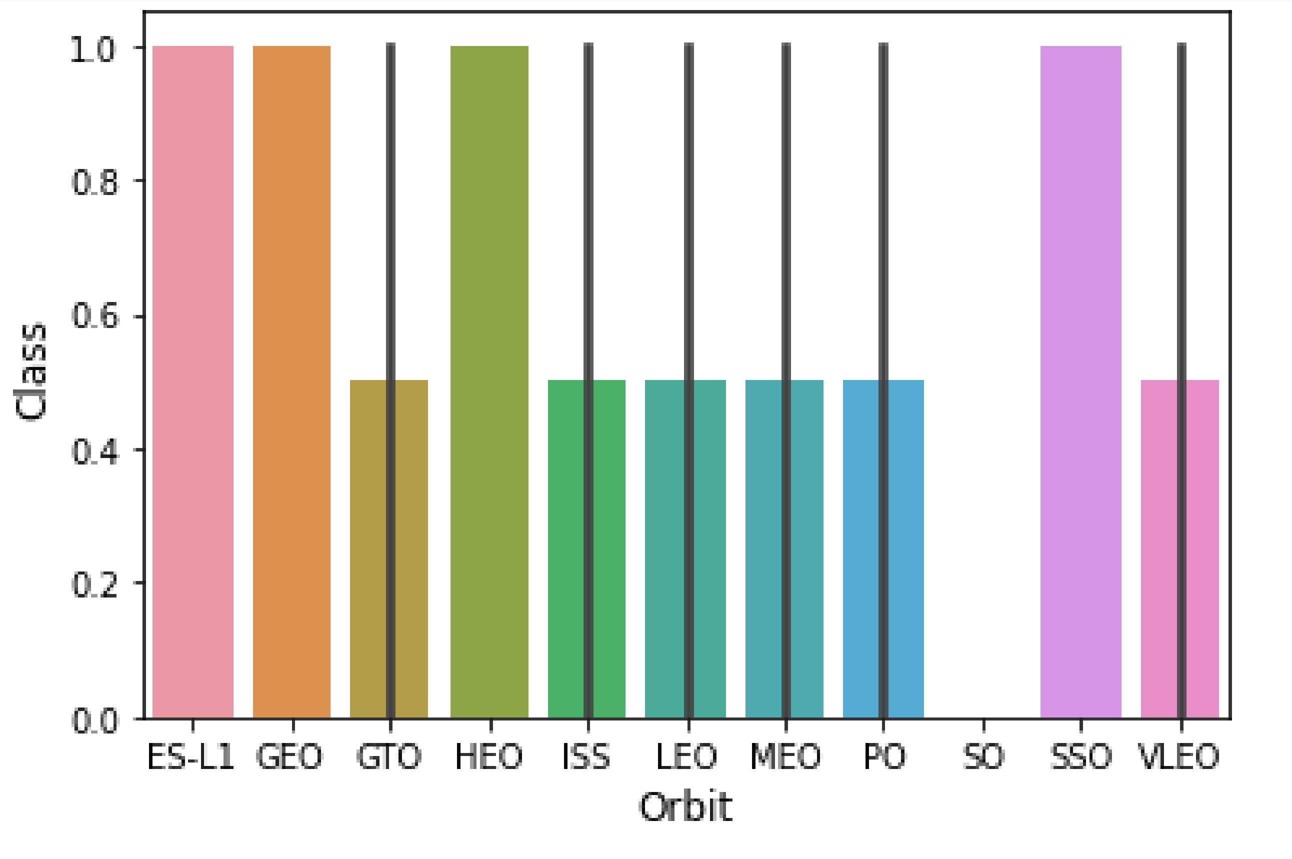
Payload vs. Launch Site

- The success rate seems to increase with payloads greater than 10,000 kg



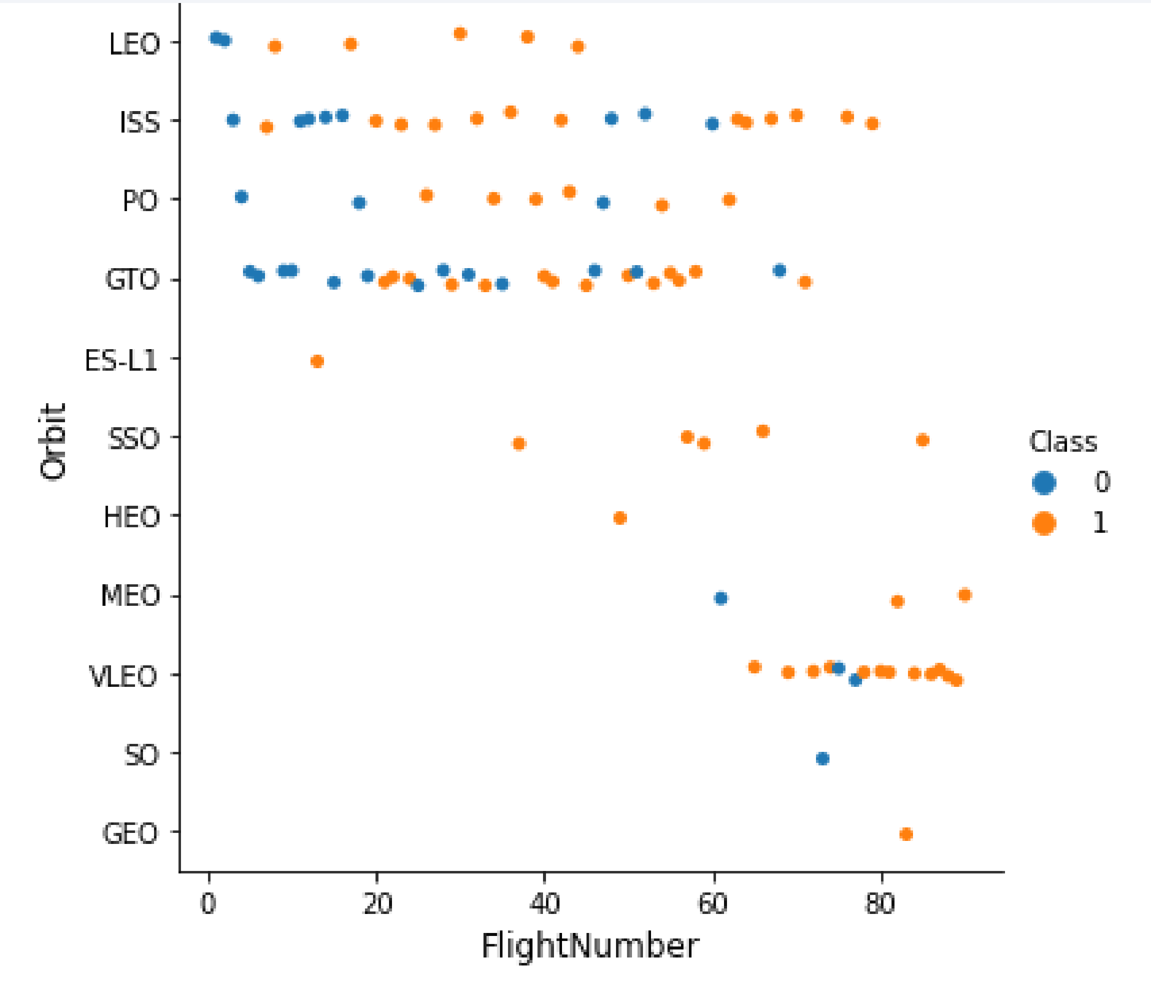
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO were the orbits with the highest success rate.



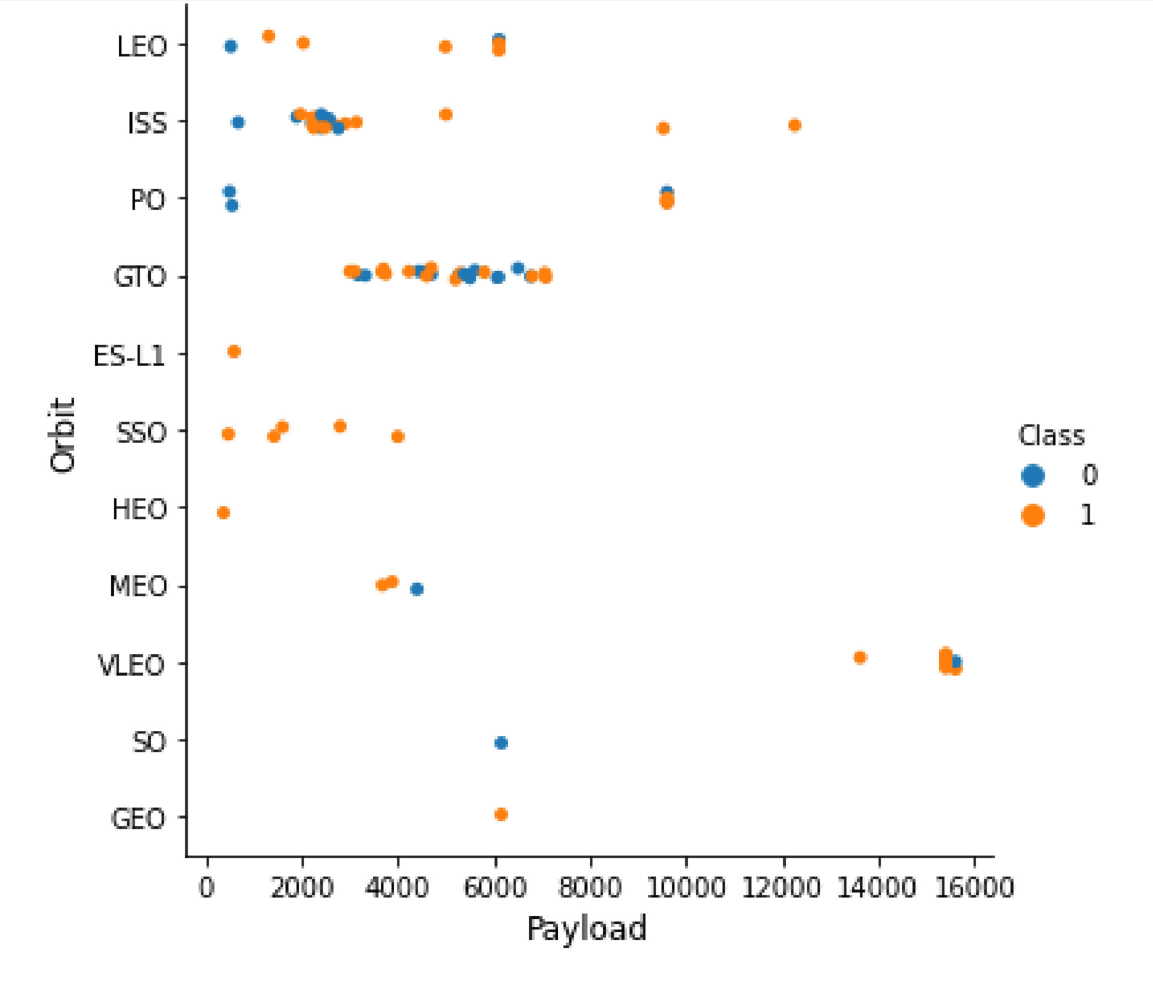
Flight Number vs. Orbit Type

- Most of the flights seem to be to ISS, and GTO obits
- In the later flights there seems to be an increase on VLEO obit launches.



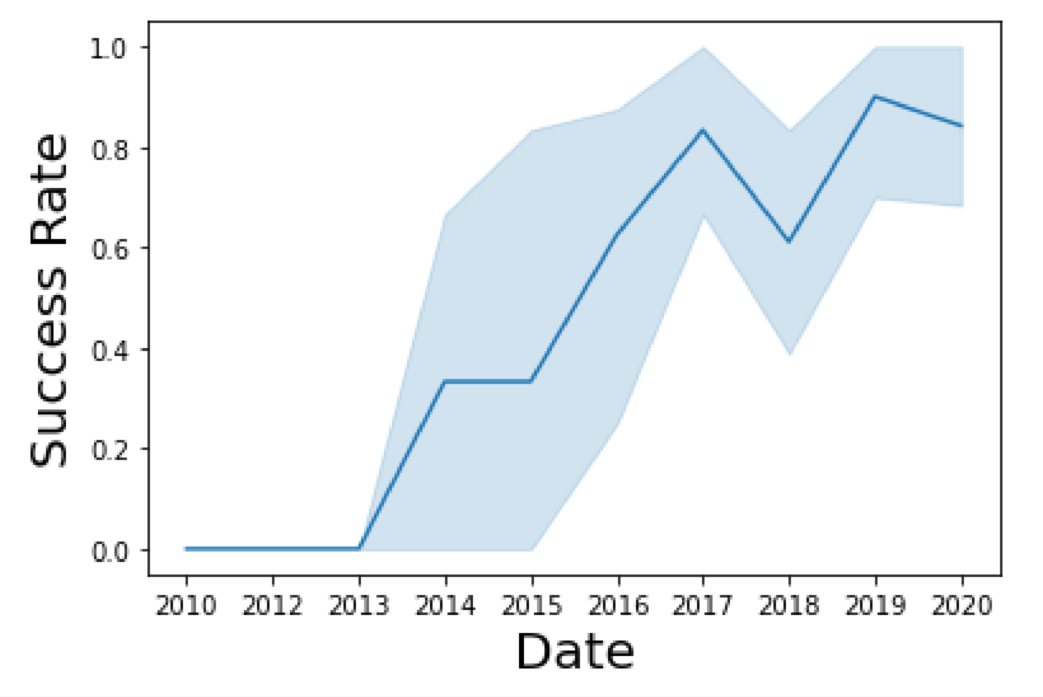
Payload vs. Orbit Type

- Large payloads look to be launched to VLEO obit
- Payloads less than 8,000 kg are mostly sent to GTO and ISS obits



Launch Success Yearly Trend

- The Success Rate has be steadily increasing over time.
- There was a notable decrease in Success Rate in 2018.



All Launch Site Names

Query:

Using the word DISTINCT in the query means that it will only show Unique values in the Launch_Site column from tblSpaceX

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

Launch Site Names Begin with 'CCA'

Query:

Using the word TOP 5 in the query means that it will only show5 records from tblSpaceX and LIKE keyword has a wild card with the words ‘CCA%’ the percentage in the end suggests that the Launch_Site name must start with CCA.

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

Total Payload Mass

Total payload carried by boosters from NASA

Query:

Using the function SUM summates the total in the column PAYLOAD_MASS_KG_

The WHERE clause filters the dataset to only perform calculations on Customer NASA (CRS)

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
45596

Average Payload Mass by F9 v1.1

Query:

Using the function AVG works out the average in the column PAYLOAD_MASS_KG_

The WHERE clause filters the dataset to only perform calculations on Booster_version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
2928.400000

First Successful Ground Landing Date

Query:

Using the function MIN works out the minimum date in the column Date

The WHERE clause filters the dataset to only perform

calculations on Landing_Outcome Success (drone ship)

```
%sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Query

Selecting only Booster_Version

The WHERE clause filters the dataset to Landing_Outcome = Success (drone ship)

The AND clause specifies additional filter conditions Payload_MASS_KG 4000 AND Payload_MASS_KG 6000

```
%sql select BOOSTER_VERSION from SPACEXTBL where Landing__Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS_
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
100

Boosters Carried Maximum Payload

List the 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)
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

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

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

```
%sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
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	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
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	05:21:00	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)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

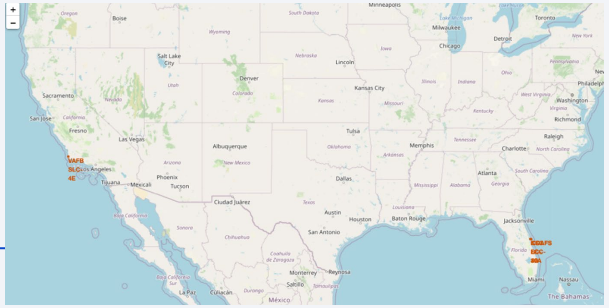
Section 4

Launch Sites Proximities Analysis

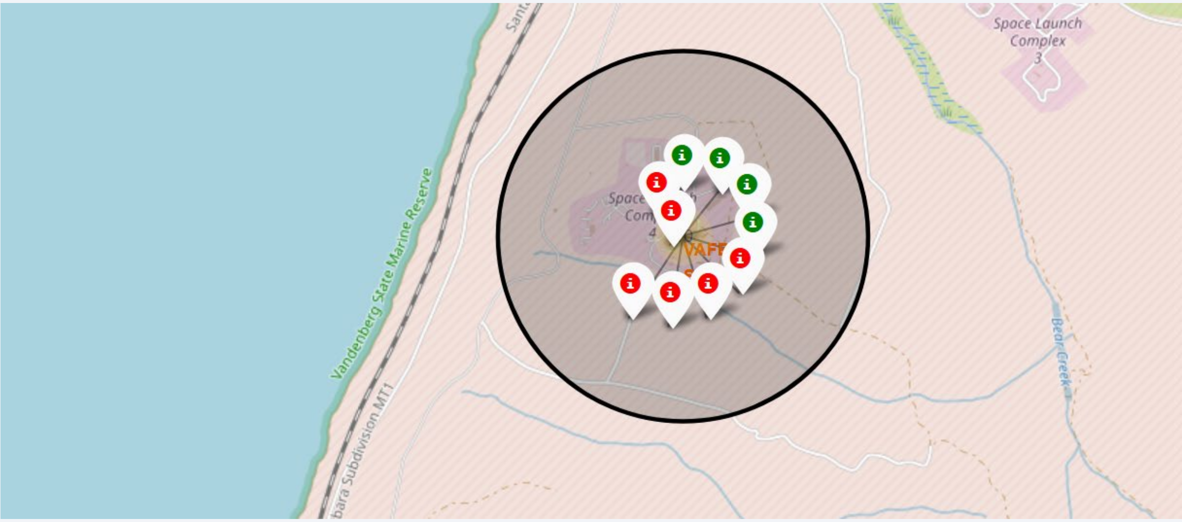
All Launch Sites

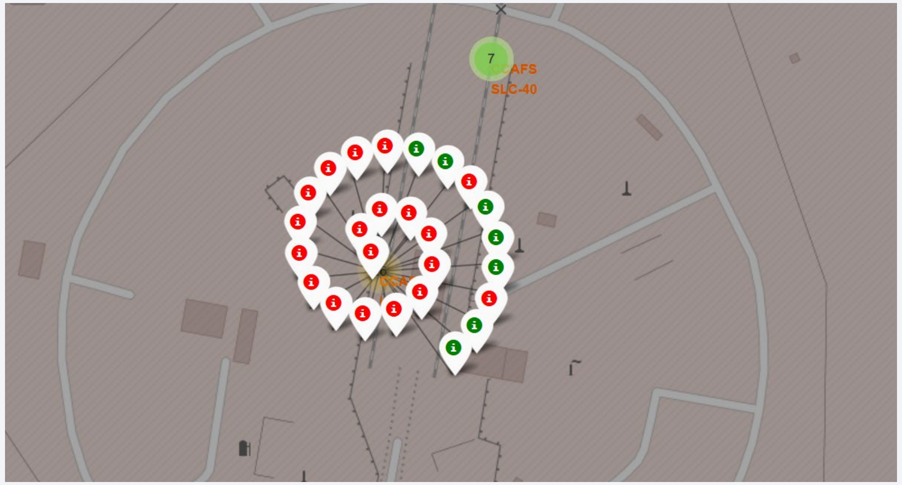
We can see that all launches are focused in CA and FL.

Launch Labeled Markers

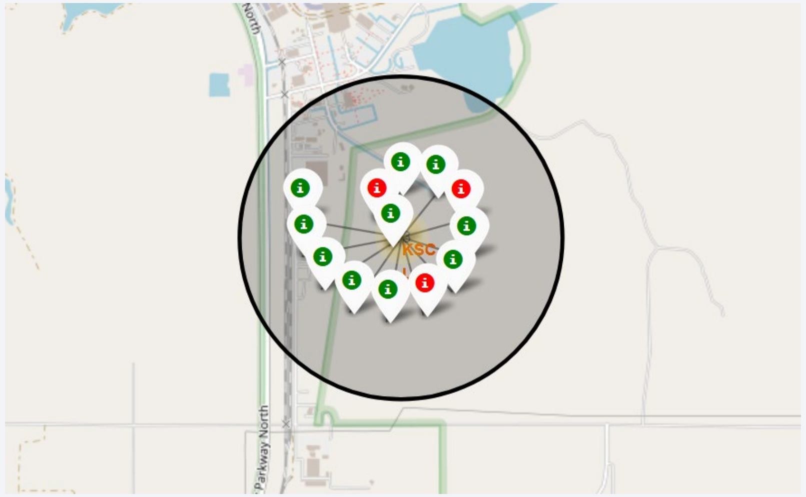


Exploring the map the **successful** landings are marked with a **Green** marker and **Red** for the **unsuccessful** landings

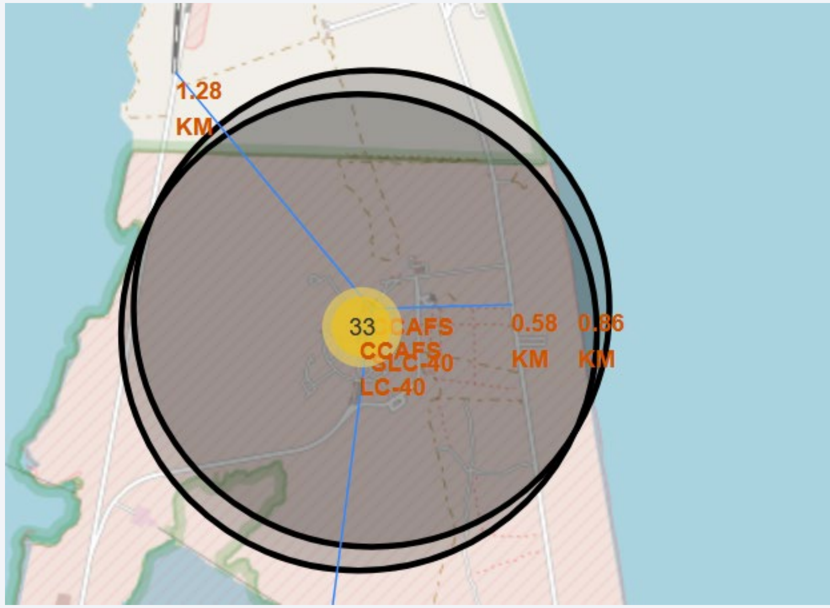




It was found that KSC LC-39A had more successful landings with only 3 failures.

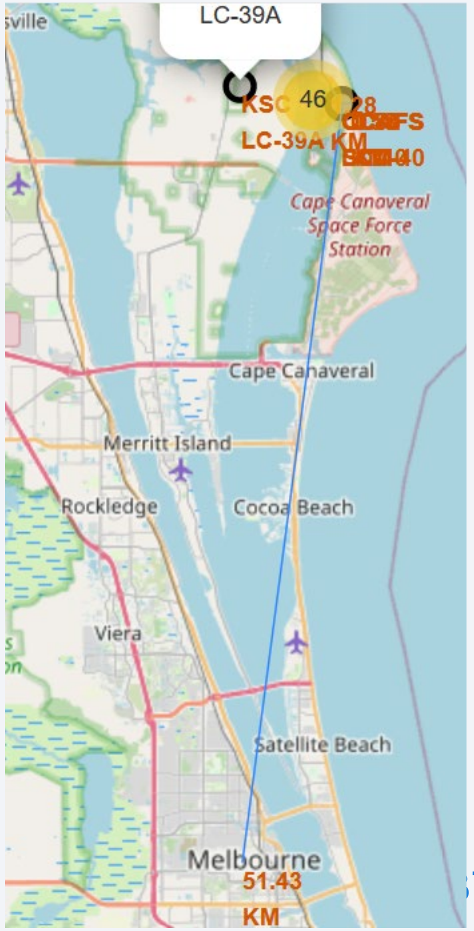


Distances from the Launch Site



Landing site distances:

- 0.86 KM from the ocean
- 0.58 KM from the nearest road
- 1.28 KM from the Railroad
- 51.43 KM from the nearest International Airport



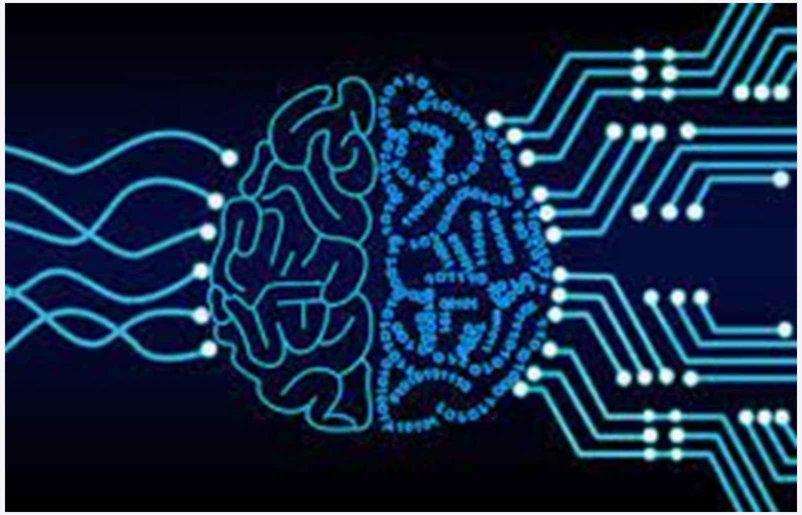
Section 6

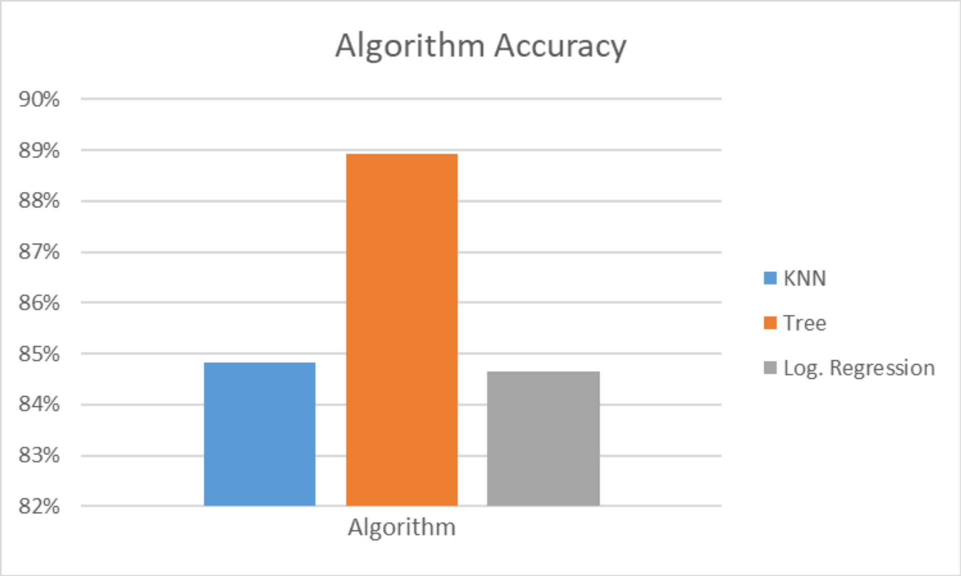
Predictive Analysis (Classification)

Classification Accuracy

- It was found that the Tree Algorithm was the most accurate to predict this data set.

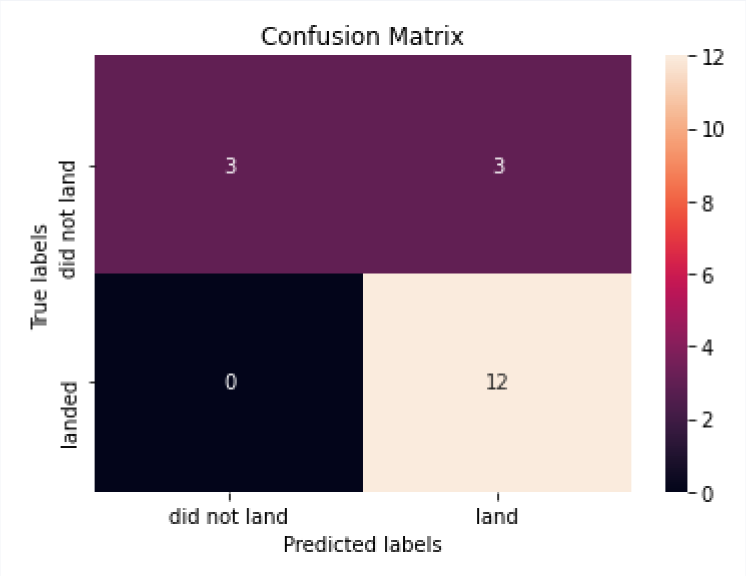
Accuracy	Algorithm
0.848214	KNN
0.889285	Tree
0.846428	Log. Regression

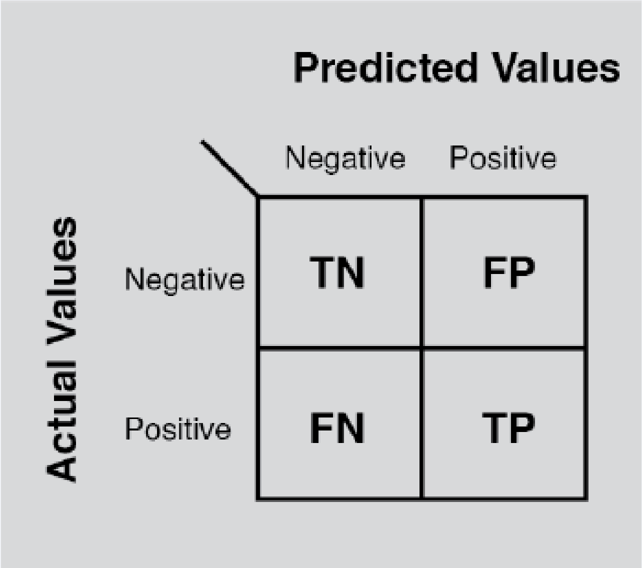




Confusion Matrix

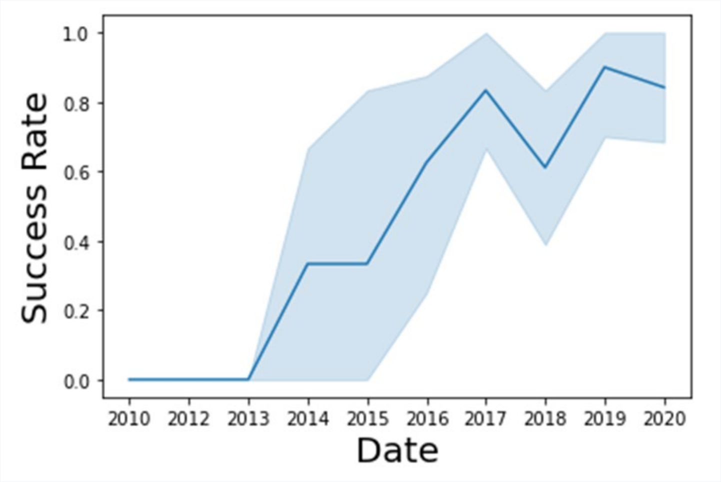
- From the confusion matrix we can see tree algorithm does predict some false positives in the data set.

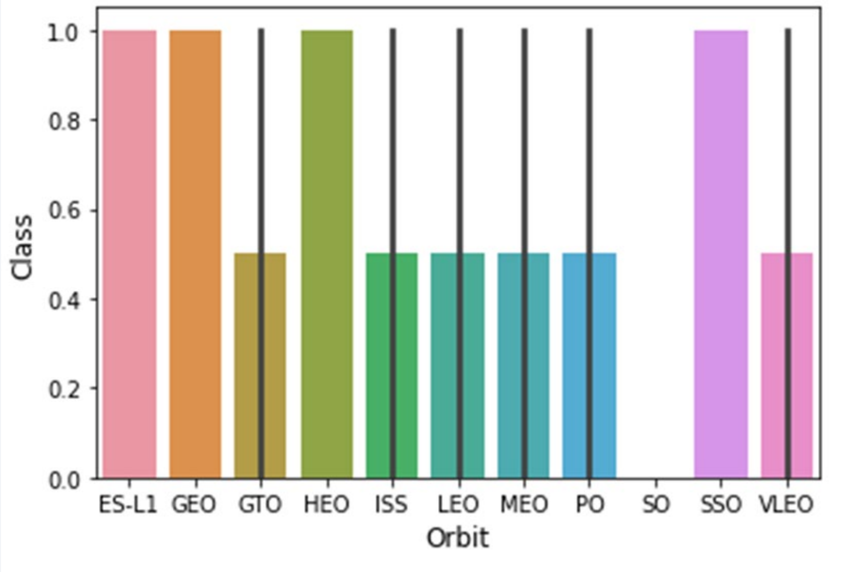


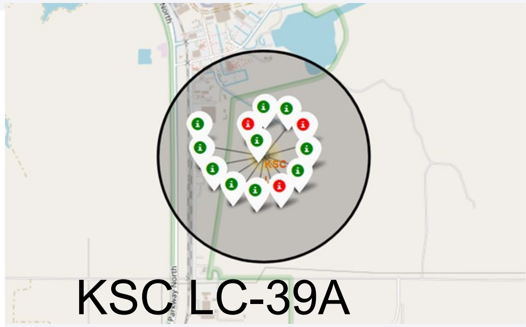


Conclusions

- The best Classifier Algorithm was the Tree Classifier with 0.889285 accuracy.
- KSC LC-39A had more successful landings.
- The Success Rate has increased over time.
- ES-L1, GEO, HEO, and SSO were the orbits with the highest success rate.







Appendix

IBM Cloud

Link to all the GetHub Files

Thank you!