IBM Data Science Capstone Project

Jimmy Vásquez Ochoa

31 March 2023

Executive Summary

- Introduction
- Data Collection
- Data Wrangling
- EDA with Visualization
- EDA with SQL
- Launch Sites Locations Analysis with Folium
- Build a Dashboard with PLOTLY
- Machine Learning Prediction
- Conclusion

Introduction

In this capstone, we will predict if the Falcon 9 first stage will land successfully. 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.

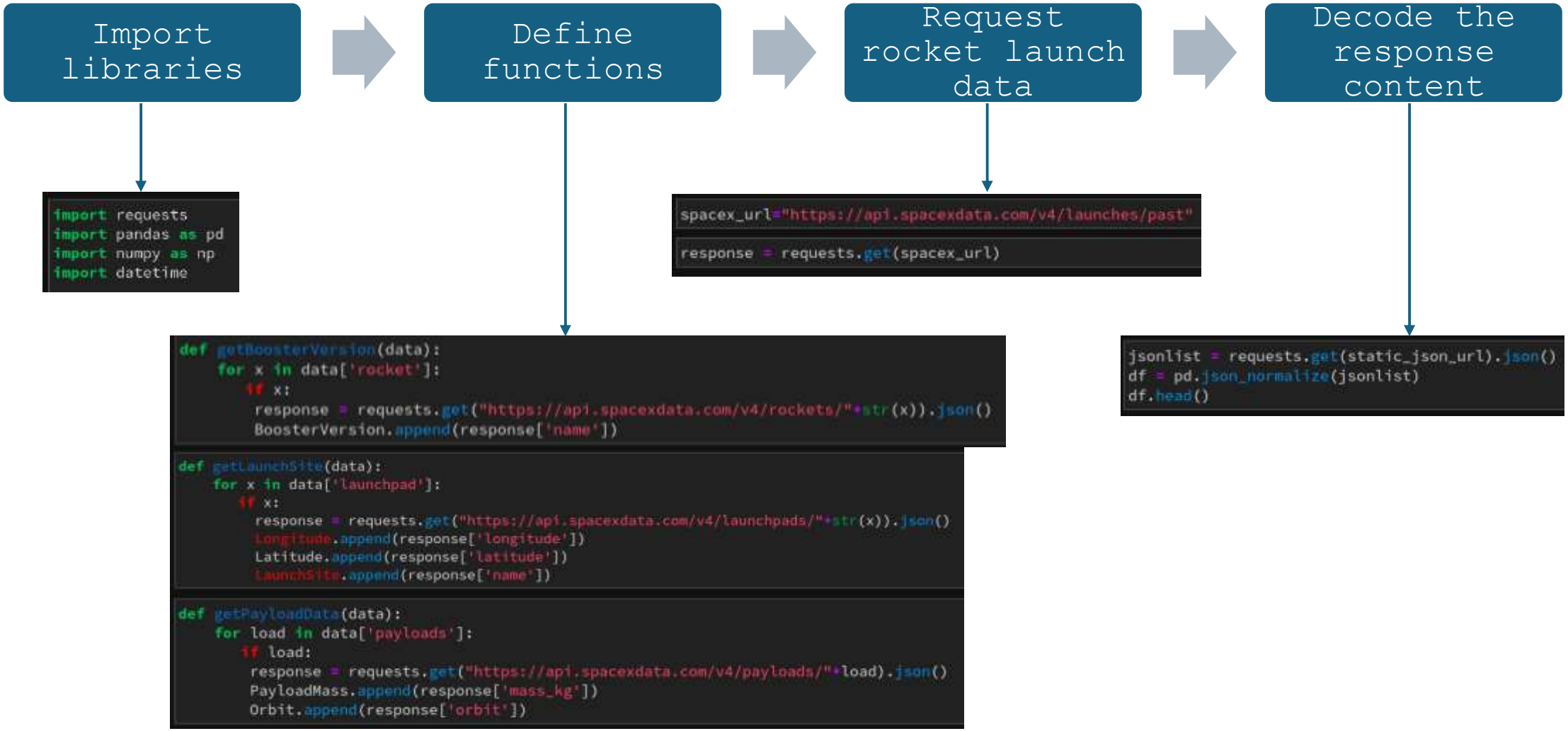
Data Collection

Objectives

In this lab, you will make a get request to the SpaceX API. You will also do some basic data wrangling and formating.

- Request to the SpaceX API
- Clean the requested data

Data Collection SpaceX API



Data Collection SpaceX API

Construct dataset

```
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}
```

Remove Falcon 1

```
data_falcon9.drop(data_falcon9[data_falcon9['BoosterVersion']!='Falcon 9'].index, inplace = True)
```

Final dataframe

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Collection Web Scrapping

Scrape the data

Create a BeautifulSoup object

Find all tables

Extract columns name

Create dictionary

Create dataframe

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
data  = requests.get(static_url).text
```

```
soup = BeautifulSoup(data, 'html5lib')
```

```
html_tables=soup.find_all("table")
first_launch_table = html_tables[2]
```

```
ths = first_launch_table.find_all('th')
for th in ths:
    name = extract_column_from_header(th)
    if name is not None and len(name) > 0:
        column_names.append(name)
```

```
launch_dict= dict.fromkeys(column_names)
```

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success	F9 v1.0B0007.1	No attempt	1 March 2013	15:10

Data Wrangling

Objectives

Perform exploratory Data Analysis and determine Training Labels

- Exploratory Data Analysis
- Determine Training Labels

Data Wrangling

Calculate number of launches

```
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
```

Calculate the number of each orbit

```
df['Orbit'].value_counts()
```

```
GTO     27
ISS     21
VLEO    14
PO       9
LEO      7
SSO      5
MEO      3
ES-L1    1
HEO      1
SO       1
GEO      1
```

Mission outcome of the orbts

```
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

Create a landing outcome

```
df['Class'] = df['Outcome'].apply(lambda landing_class: 0 if landing_class in bad_outcomes else 1)
df[['Class']].head()
```

Create a dataframe

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Clas
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	

EDA with Visualization

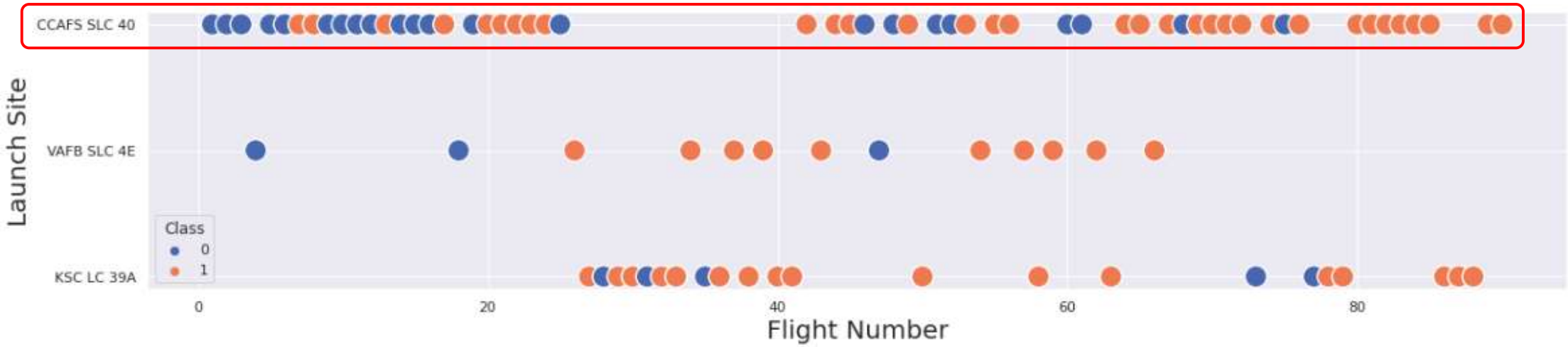
Objectives

Perform exploratory data analysis and feature engineering using pandas and matplotlib

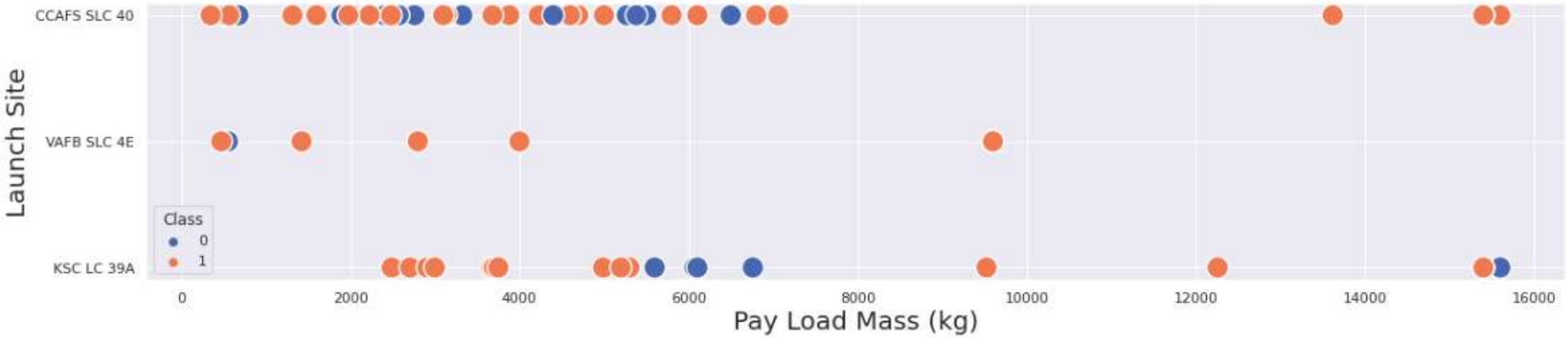
- Exploratory data analysis
- Feature engineering

Visualize the relationship between Flight Number vs Launch Site

The greater number of flight are from CCAFS SLC 40

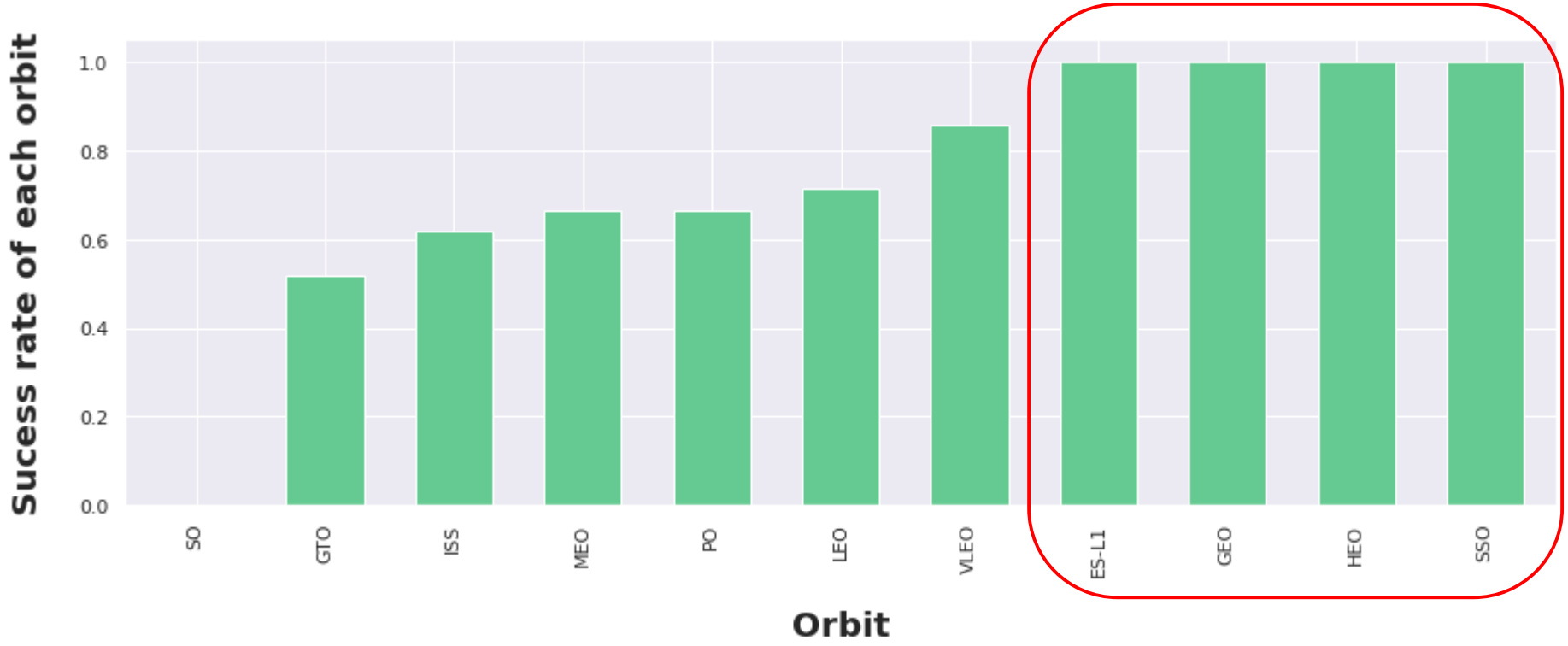


Visualize the relationship between Payload Mass vs Launch Site



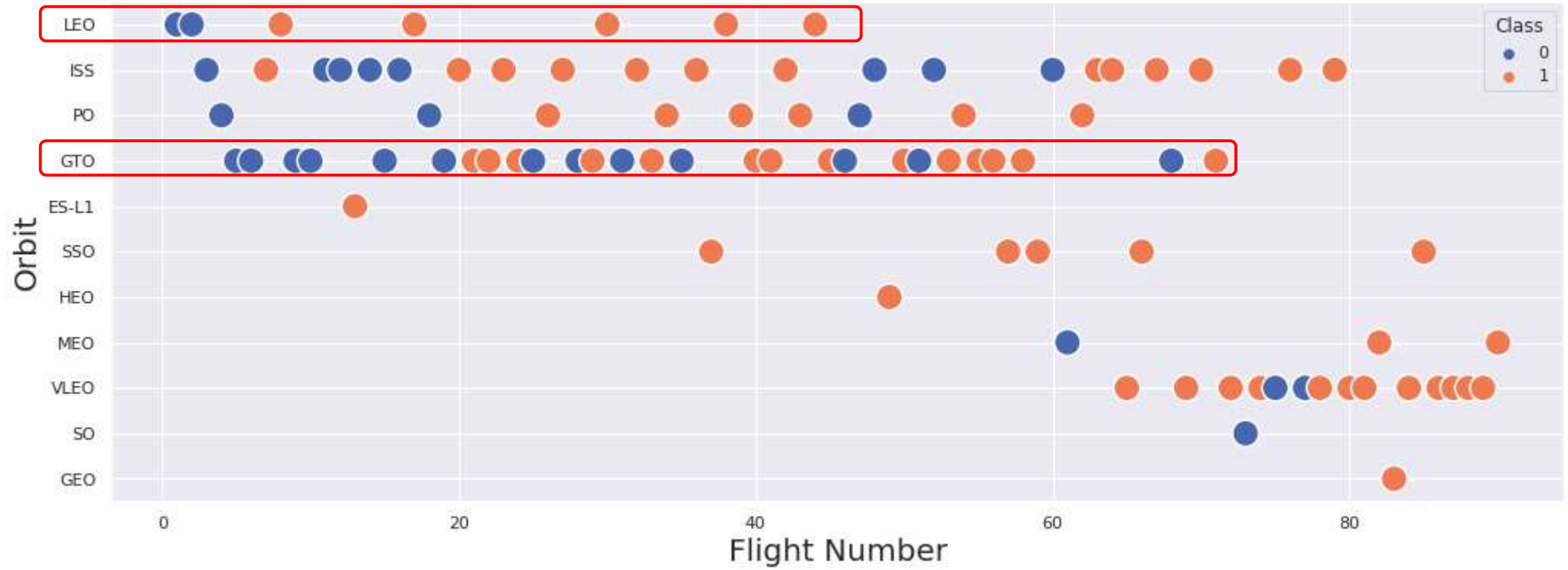
Success Rate vs Orbit

The orbits of ES-L1, GEO, HEO and SSO has the highest rate of success



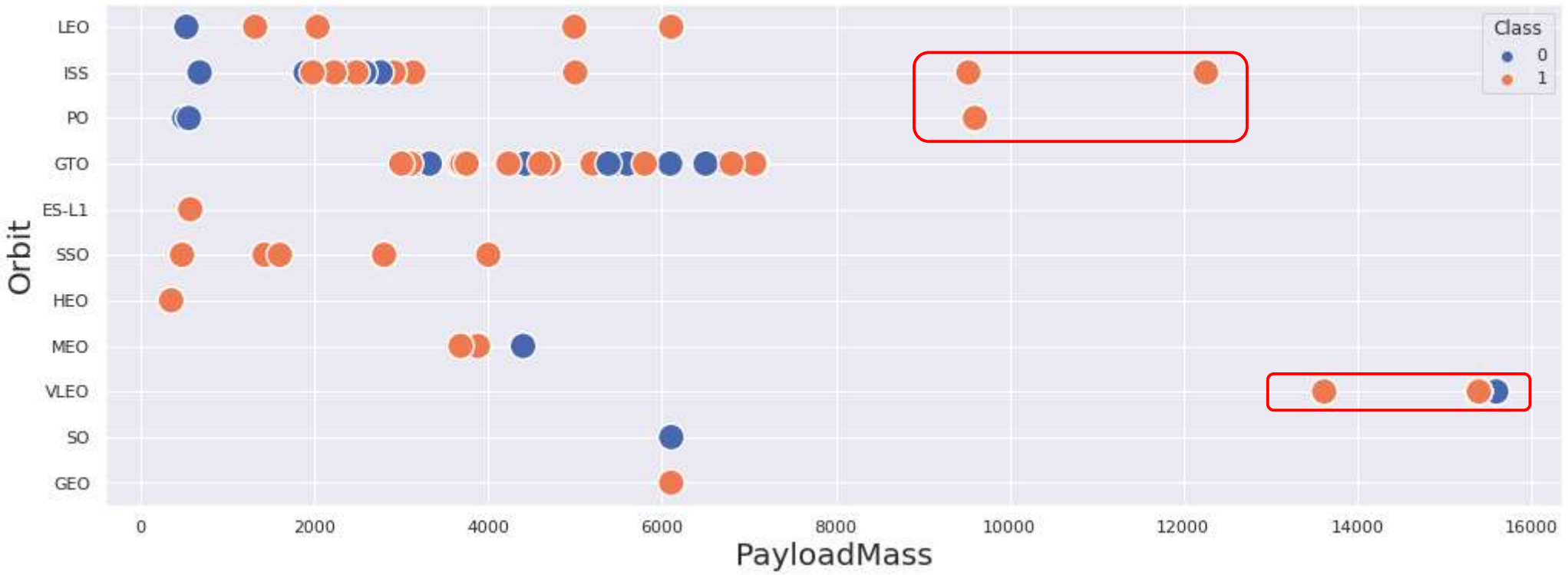
Visualize the relationship bewteen Flight Number vs Orbit Type

We see that in the LEO orbit the success appears related to the numbers os flights, on the other hand, there seems to be no relationship between flight number when in GTO orbit



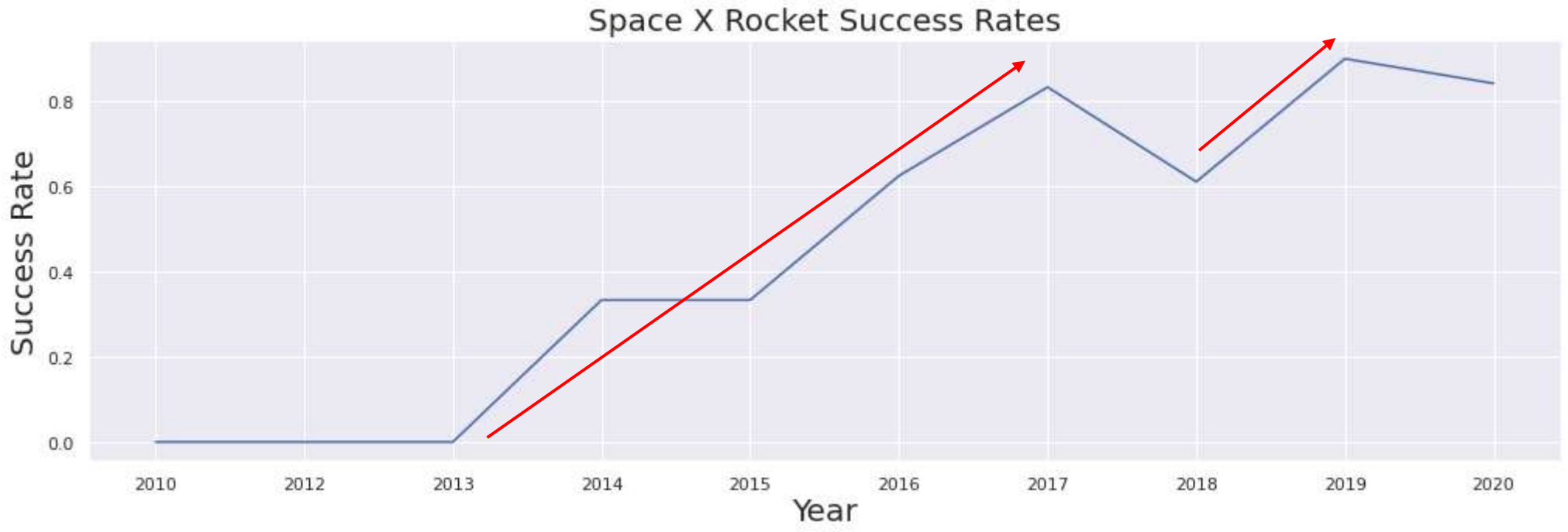
Visualize he relationship between Payload Mass vs Orbit type

With heavy payloads the successful landing or positive landing rate are more for Polar, VLEO and ISS



Visualize the Launch Success Yearly Trend

The success rate since 2013 kept increasing until 2017 and then decrease and recover in 2019



EDA with SQL

Display the names of the unique Launch Sites in the Space Mission and 5 records where Launch Sites begin with the string CCA

Select Launch Sites

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTABLE;
```

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

Select Launch Sites with CCA

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

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

Display the Payload Mass carried by boosters launched by NASA (CRS) and Average Payload Mass carried by boosters versión F9 v1.1

Total Payload Mass

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

sum(PAYLOAD_MASS__KG_)
45596

Carried by booster F9

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

avg(PAYLOAD_MASS__KG_)
2928.4

Successful Ground Landing Date and Drone Ship Landing with Payload between 4000 and 6000

First Successful Landing

```
%sql select min(DATE)  AS "First Succesful Landing Outcome in Ground Pad" from SPACEXTABLE
where Landing_Outcome = 'Success (ground pad)'
```

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

Payload Mass >4000

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTABLE WHERE LANDING_OUTCOME = 'Success (drone ship)'
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

Sucessful Mission

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Successful Mission" FROM SPACEXTABLE WHERE MISSION_OUTCOME
LIKE 'Success%';
```

Successful Mission
100

Failure Mission

```
%sql SELECT COUNT(MISSION_OUTCOME) AS "Failure Mission" FROM SPACEXTABLE WHERE MISSION_OUTCOME
LIKE 'Failure%';
```

Failure Mission
1

Booster Version wich carried the Maximum Payload Mass

Maximum Payload Mass

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

Booster Versions which carried the Maximum Payload Mass
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

Records wich will display the months in 2015 and the Landing Outcome between 2010-2017

Records in 2015

```
%sql SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] \
FROM SPACEXTABLE where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

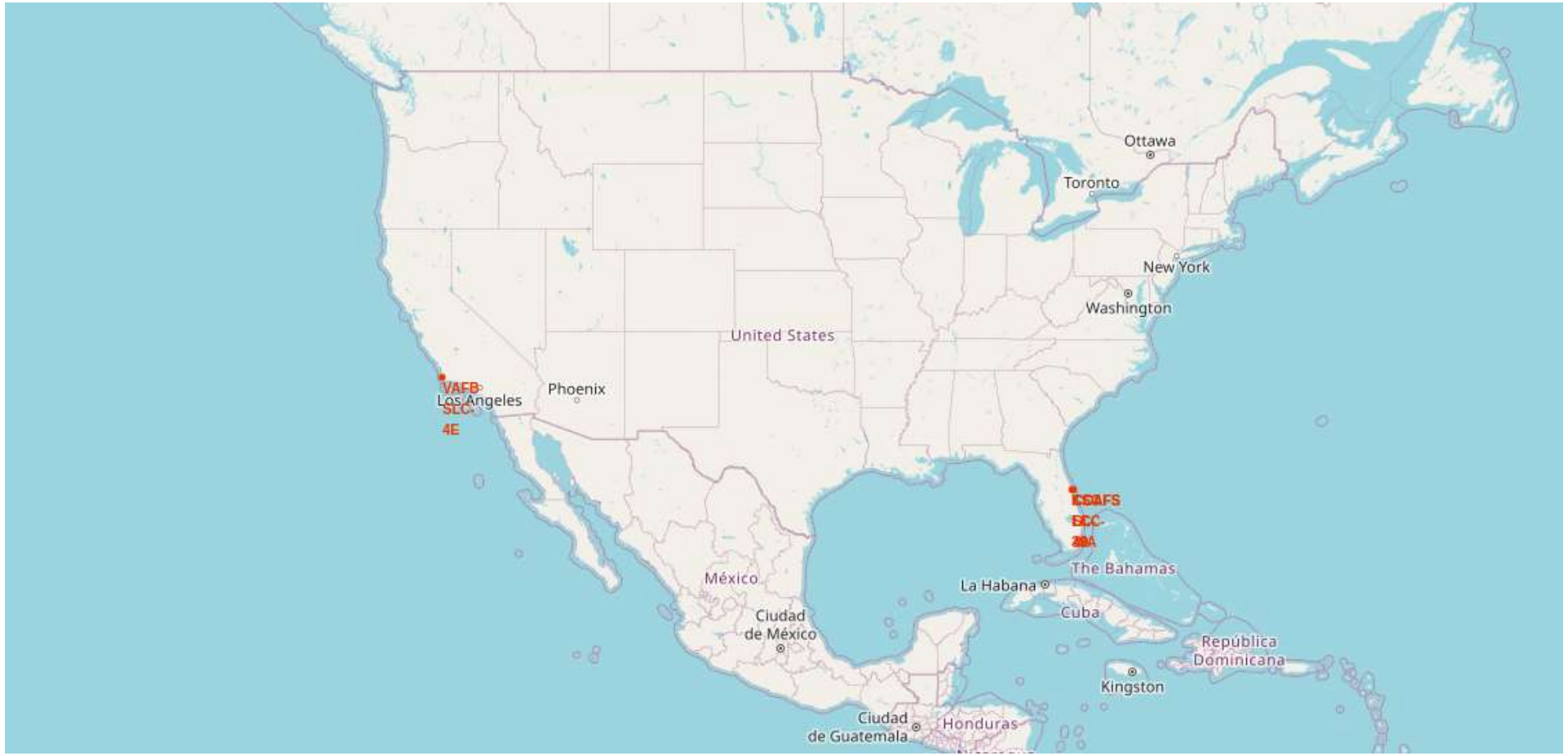
Landing Outcome 2010-2017

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

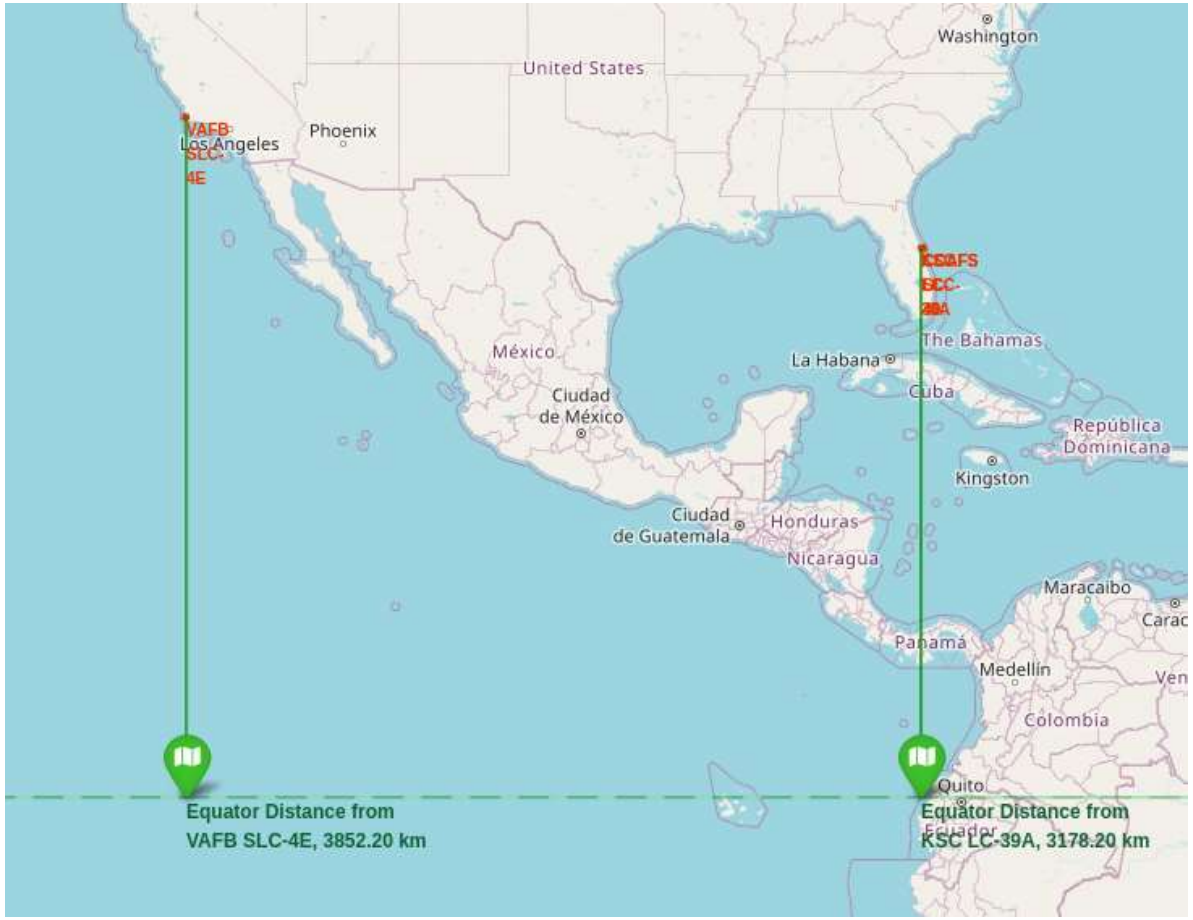
Landing Outcome	Total Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Launch Sites Locations Analysis with Folium

All Launch Sites on a Map

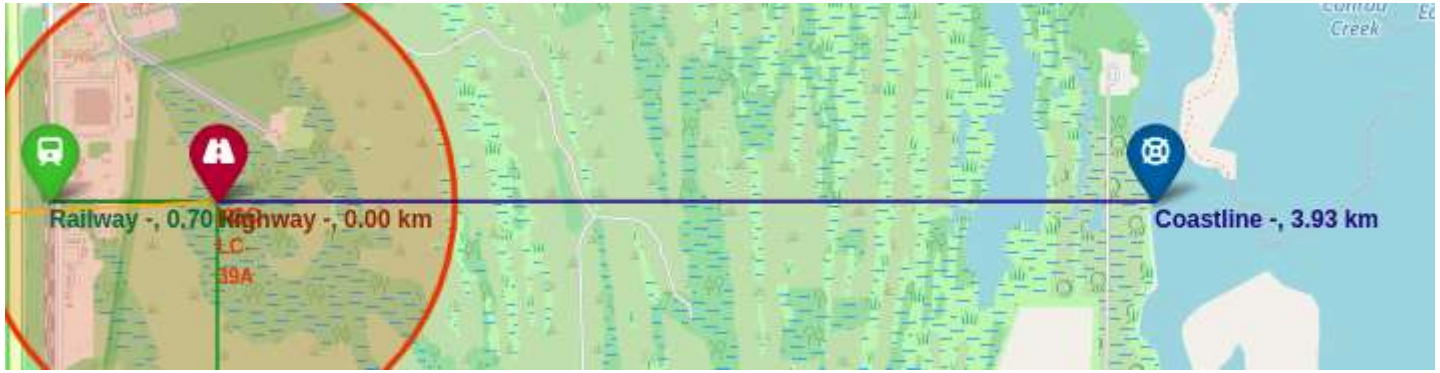


Distances between Launch Sites to its proximities



Distances between Launch Sites to Equator is more than 3000 km but less than 4000 km

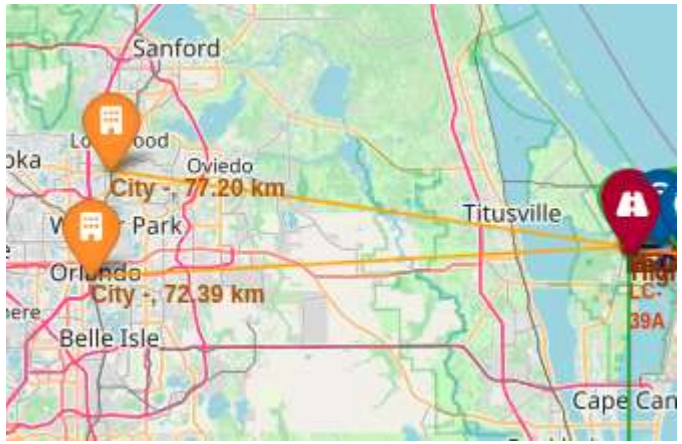
Distances between Launch Sites to its proximities



Distances between Launch Sites to its proximities

- Distances between Launch Sites to Railways is more than 0.7 km but less than 1.5 km
- Distances between Launch Sites to Coast Line is more than 0.8km but less than 4 km

Distances between Launch Sites to its proximities

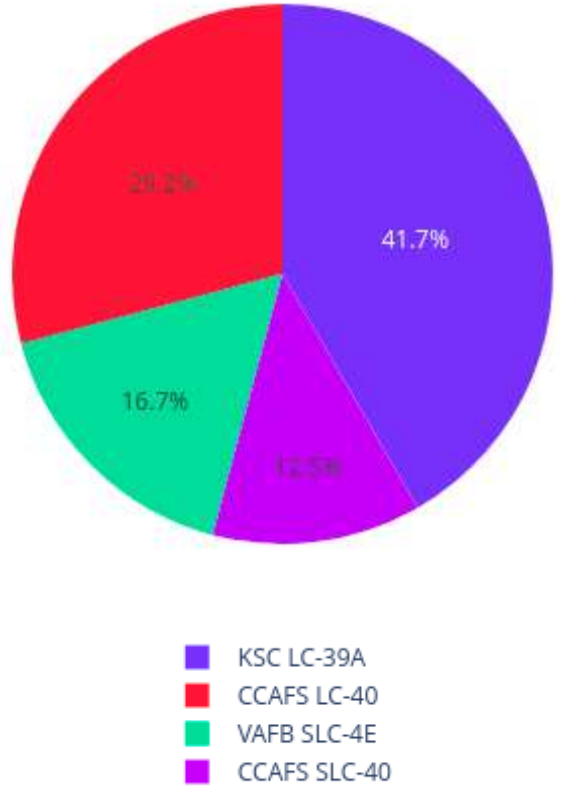


Distances between Launch Sites to Cities is more than 14km but less than 80 km

Build a Dashboard with PLOTLY

Launch Success Count for All Sites

Total Success Launches for All Sites

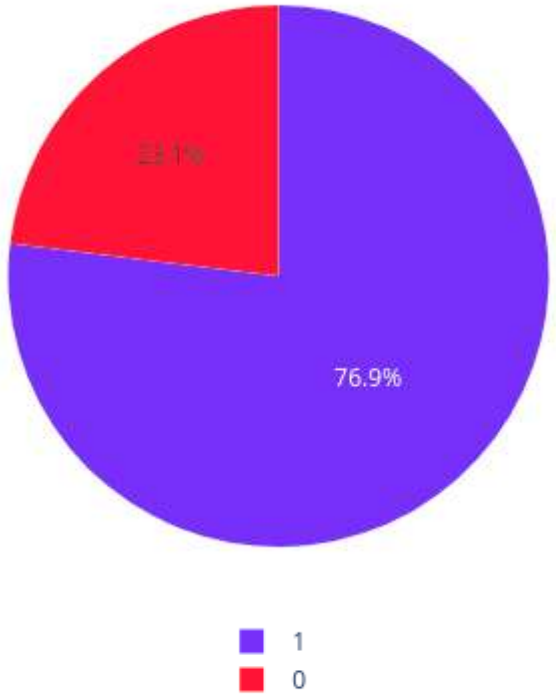


The smaller success launch is
CCAFS SLC-40

The bigger success launch is
KSC LC-39A

Launch Site with highest Success Ratio

Total Launch for KSC LC-39A



KSC LC-39A has the bigger success rate 76.9% with a 23.1% rate of failure

FINDING INSIGHTS VISUALLY

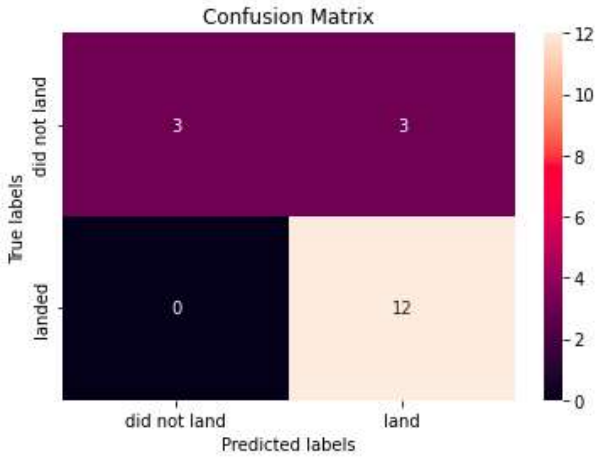
1.Wich site has the largest succesfull launches?
KSC LC-39ª

2.Wich site has the highest launch succes rate?
KSC LC-39A

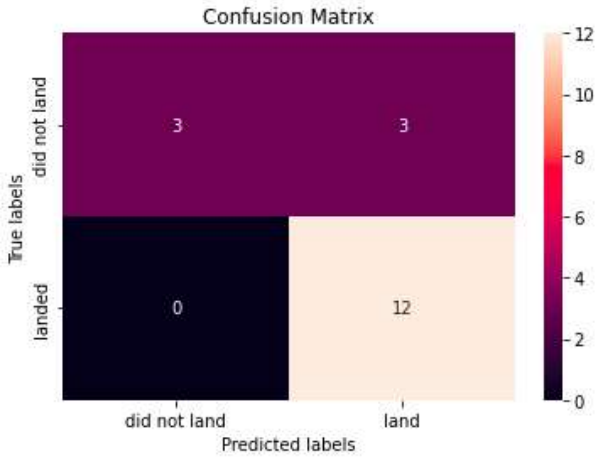
Machine Learning Prediction

Confusion Matrix

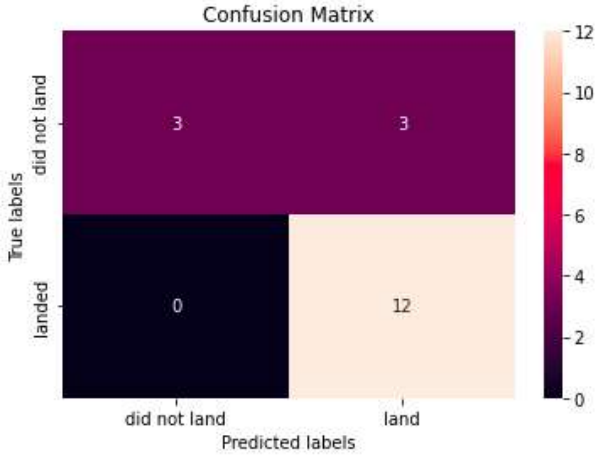
Logistic Regression



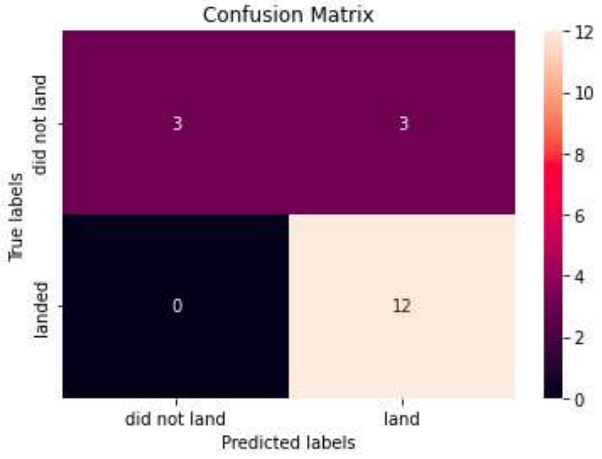
Decision Tree



SVM



KNN



Classification Accuracy

Algorithm	Accuracy	Tuned Hyperparameters
Logistic Regression	0,821428571428571 4	{'C':1, 'penalty' : 'l2', 'solver': 'lbfgs'}
SVM	0,848214285714285 8	{'C':1, 'gamma' : ',0,03162277', 'kernel': 'sigmoid'}
Decision Tree	0,848214285714285 8	{'criterion': 'entropy' , 'max Depth': 16, 'max feature':'sqrt' , 'min samples left': 4 , 'min samples Split': 2, 'splitter': 'best'}
KNN	0,833928571428571 4	{'algorithm': 'auto', 'n neighbors':6 , 'p': 1}

Conclusion

Conclusion

- The greater number of flight are from CCAFS SLC 40
- The orbits of ES-L1, GEO, HEO and SSO has the highest rate of success
- With heavy payloads the successful landing or positive landing rate are more for Polar, VLEO and ISS
- The success rate since 2013 kept increasing until 2017 and then decrease and recover in 2019
- The smaller success launch is CCAFS SLC-40 and the bigger success launch is KSC LC-39A
- Decision Tree and SVM algorithms are the best Machine Learning models