IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mouhamadou Attinine Soumana
2024-03-01

Outline

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

Executive Summary

- **Summary of methodologies**

❖ Data collection

❖ Data wrangling

❖ EDA with data visualization

❖ EDA with SQL

❖ Building an interactivemap

❖ Building a Dashboard

❖ Classification

- **Summary of all results**

❖ EDA Results

❖ Visualization analysis results

❖ Classification analysis result

Introduction

- **Project background and context**
  - 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.
- **Problems you want to find answers**
  - Can we determine if the first stage will land successufly for a bid against SpaceX for a rocket launch?

Section 1

Methodology

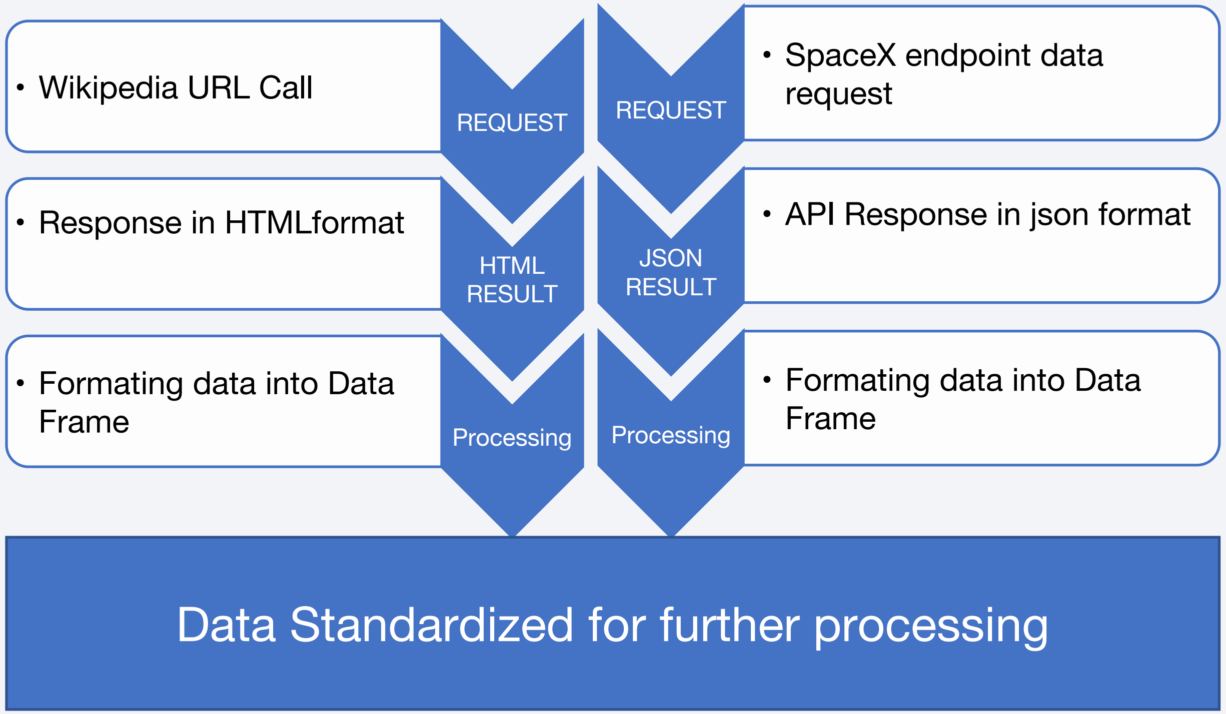
Methodology

Executive Summary

- Data collection methodology:
  - Use of SapaceX Webservice
  - Web Scraping with python from Wikipedia
- Perform data wrangling
  - OneHot Encoding data fiels for Machine Learning and data cleaning
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression,SVM, KNN , Tree models built and evalluated for classifier section

Data Collection

- **Data collection procedure**
  - ❖ Data was collected through SpaceX WebService (REST API)
  - ❖ The request return data about lauches, payload, landingsites, outcomes...
  - ❖ Wthis BeautifulSoup ,one of python’s library for web sracping we wereable parse HTML web pages to extract needed data



Data Collection – SpaceX API

SpaceX Web Service usage flow

- [GitHub source]()

1.Make request

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

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

2.Normalize data

```
# Use json_normalize meethod to convert the js
data = pd.json_normalize(response.json())
```

3.Format data

```
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}
```

4.Create Our DataFrame & exportto SV

```
# Create a data from launch_dict
data = pd.DataFrame.from_dict(launch_dict)
```

data_falcon9.to_csv('dataset_part_1.csv', index=False)

Data Collection - Scraping

1. HTML page request: response

```
# use requests.get() method with the provided static_url
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falc
object = requests.get(static_url)
# assign the response to a object
```

2. Building BeautifulSoup Object

```
soup = BeautifulSoup(object.text)
```

3. Data Extraction

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table'
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number
        if rows.th:
            if rows.th.string:
```

4. Normalization and Formating

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty lis
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
```

5. Conversion and export to CSV

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
df.to_csv('spacex_web_scraped.csv', index=False)
```

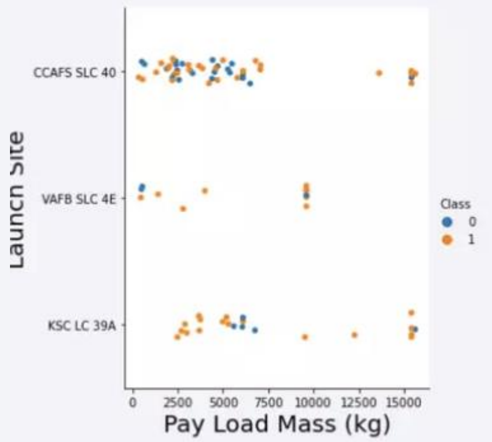
- Github source

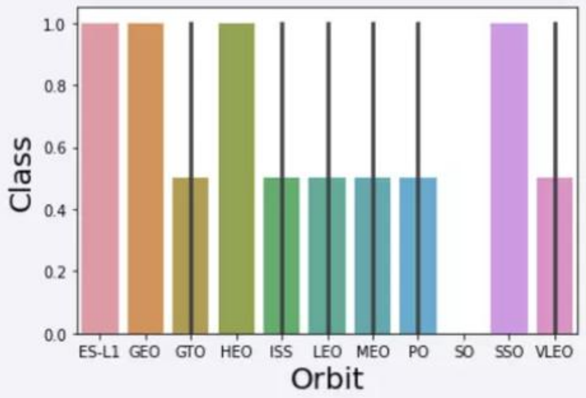
Data Wrangling

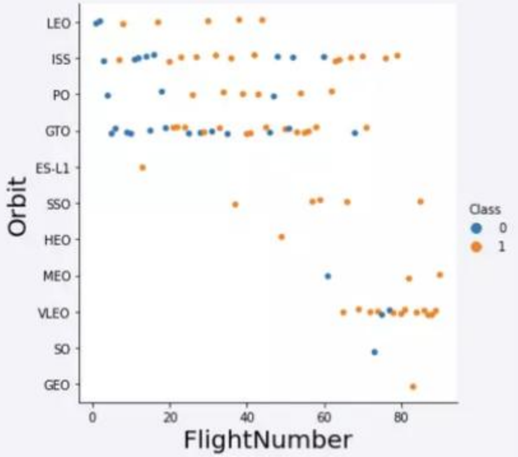
1. Checking Null
2. Compute lauches per site
3. Compute orbit occurrences
4. Evaluate number and occurence of each orbit
5. Evaluate missions outcome per orbit type
6. Landing outcome label and column creation

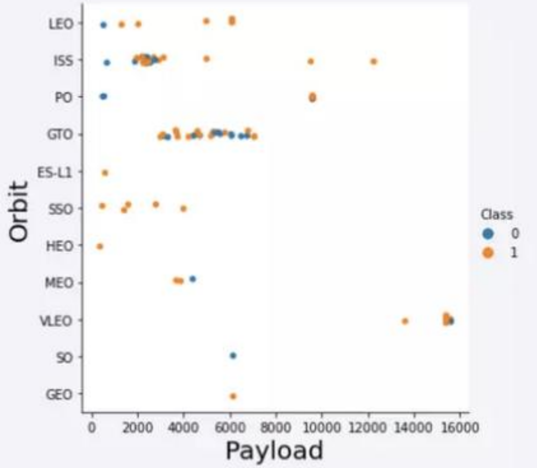
GitHub source

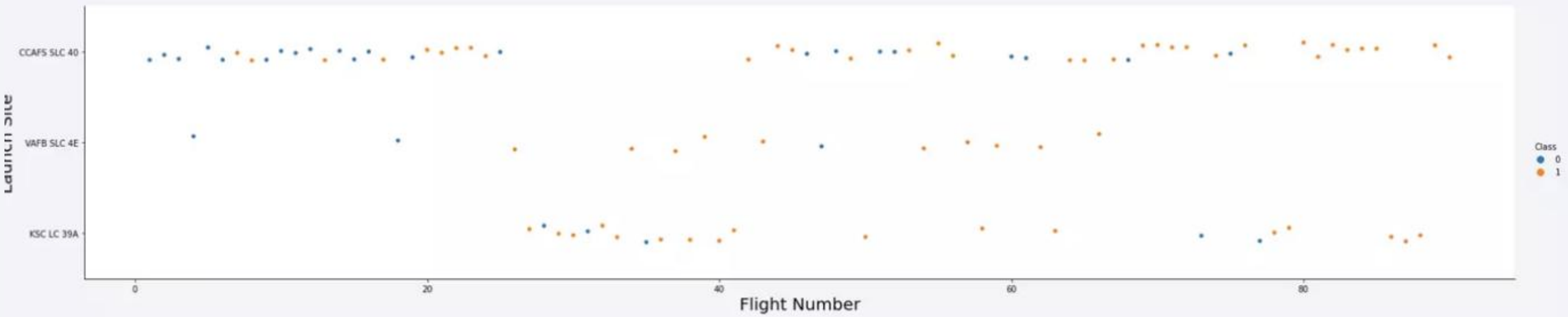
EDA with Data Visualization

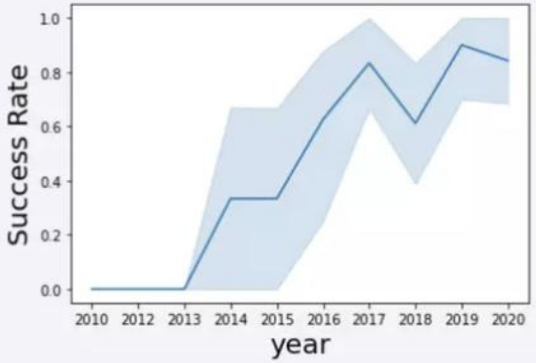










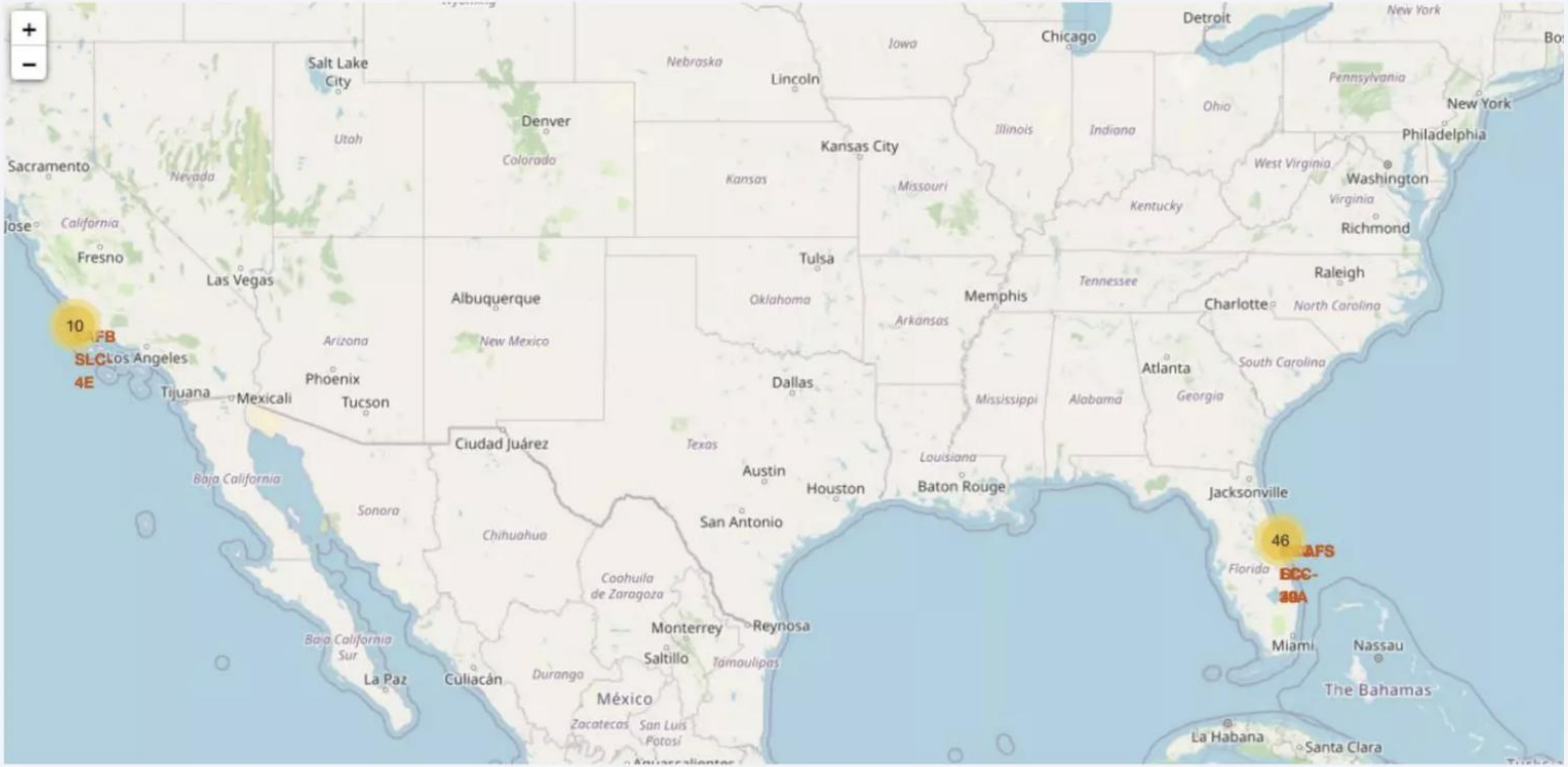


[GitHub source]

EDA with SQL

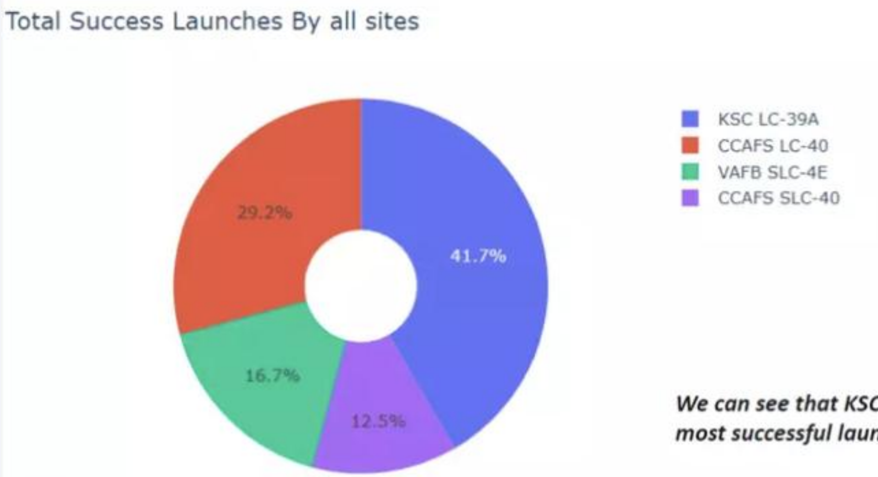
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months 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

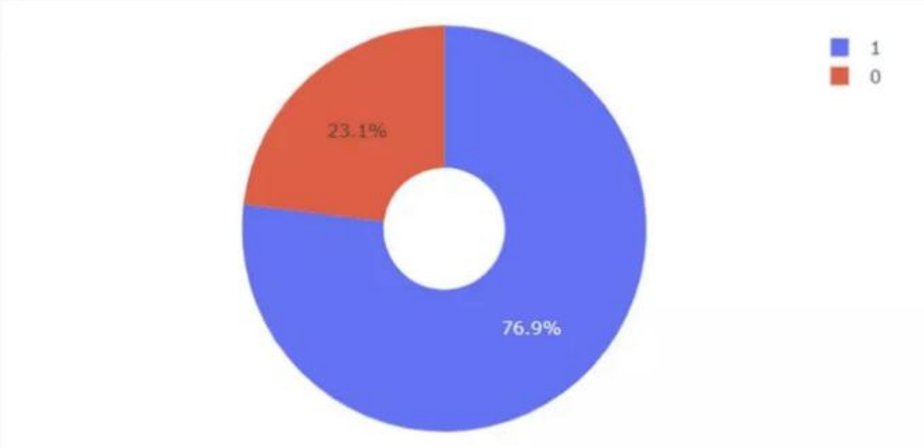


Map was used to identify the best site offering maximum positive outcome

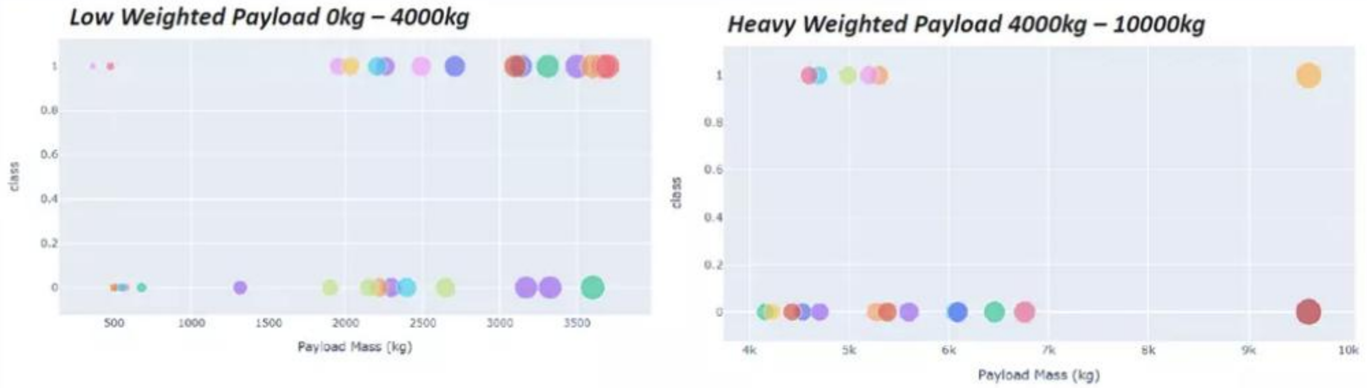
Build a Dashboard with Plotly Dash



We can see that KSC LC-39A had the most successful launches from all the sites



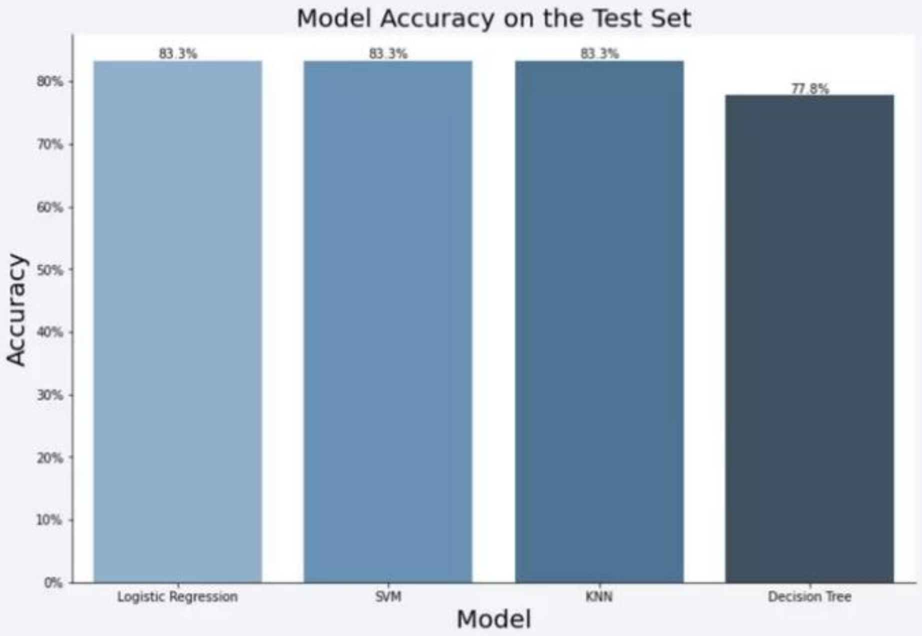
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

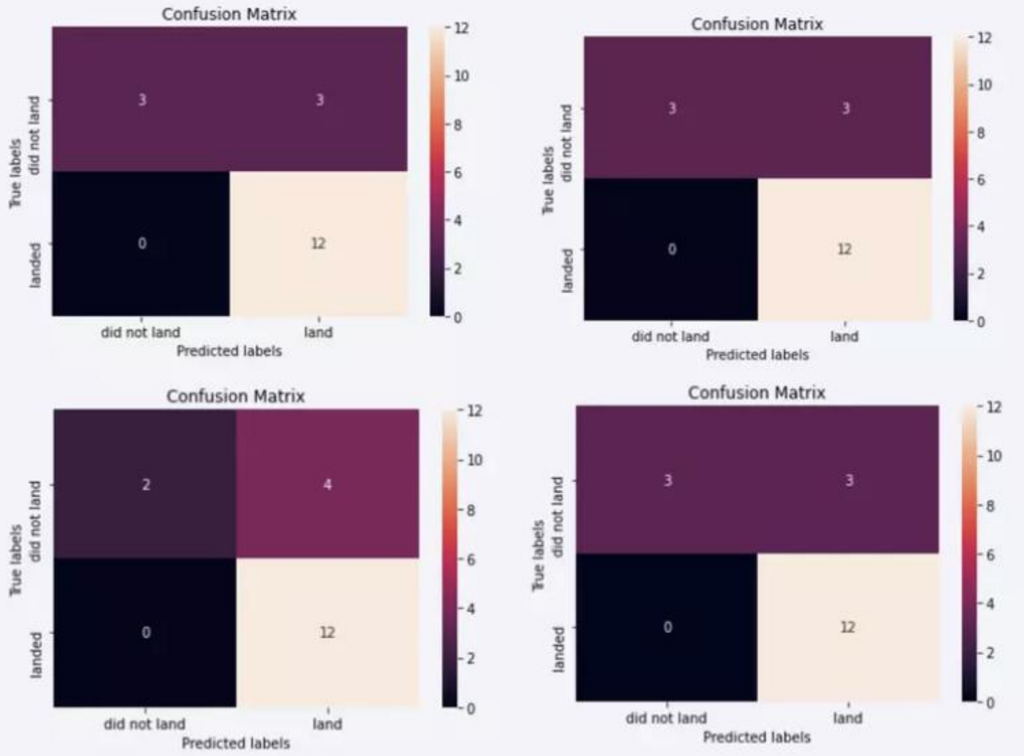


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.





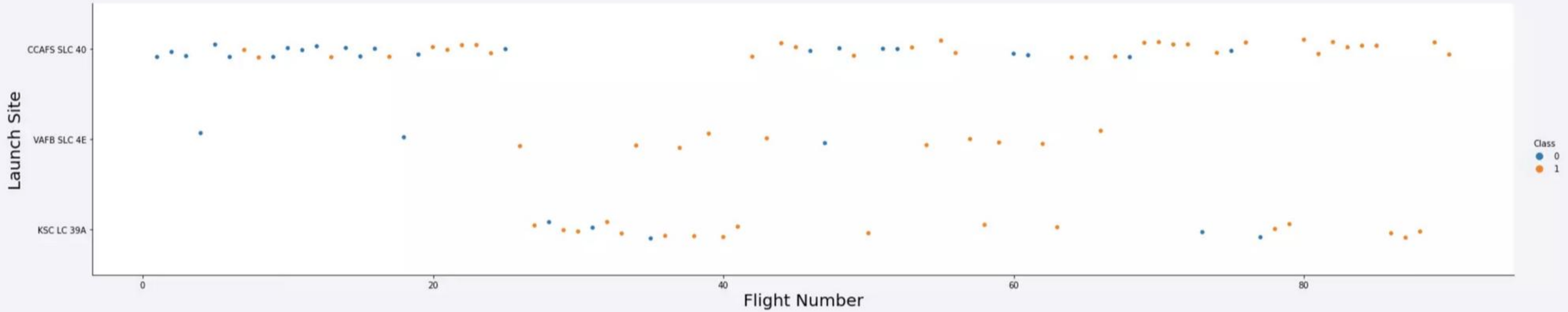
Results

- SVM,KNN and logistic Regression models happen togive best accuracy for this dataset
- Outcomes show low weight payload have better perfomace
- Succes rate tends get better over the years.
- KSC LC 39A have better launch score
- GEO,HEO,SSO,ES L1 show higher success rate

Section 2

Insights drawn from EDA

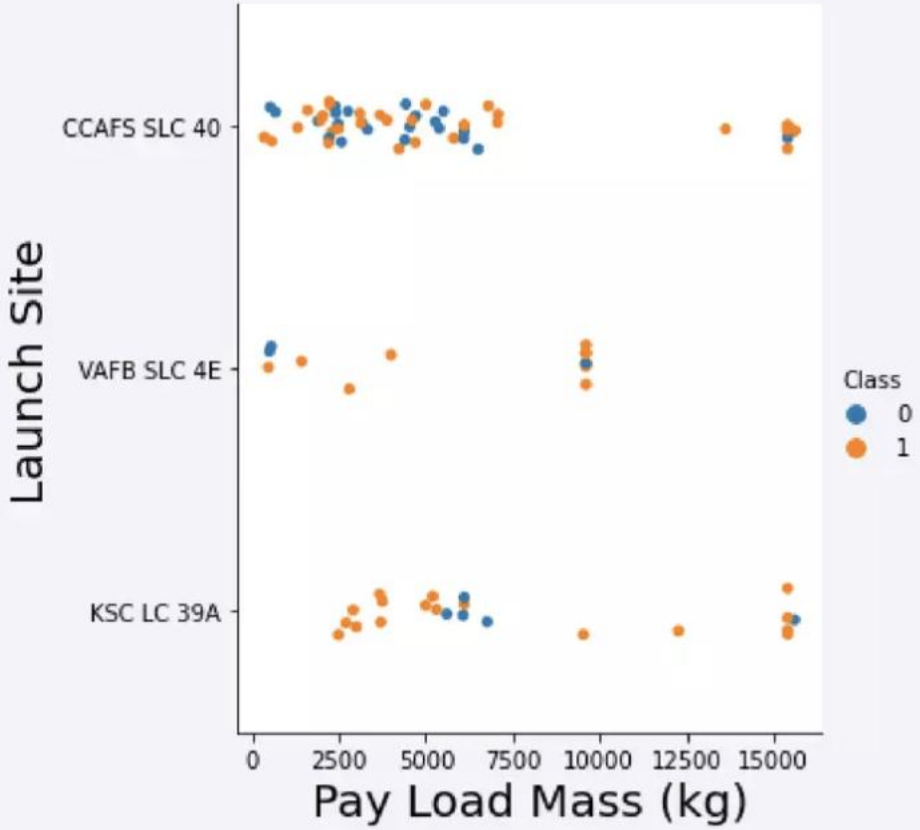
Flight Number vs. Launch Site



CCAFS SCL 40 site shows more launches than other site

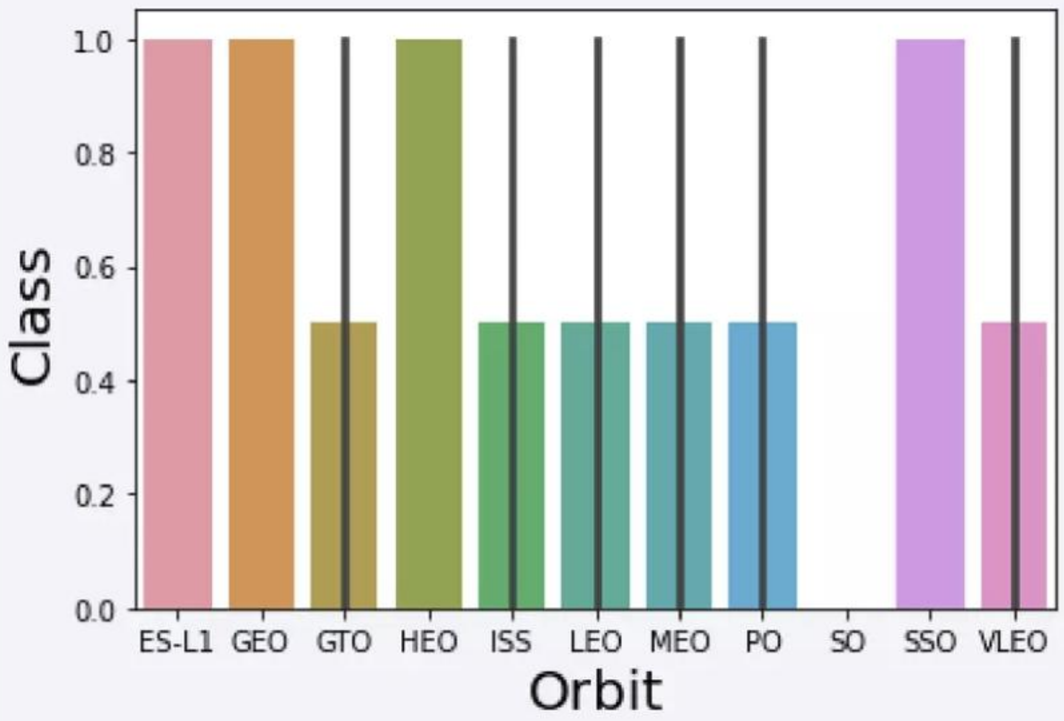
Payload vs. Launch Site

- Most of Launches done from CCAFS SCL 40 were low mass.



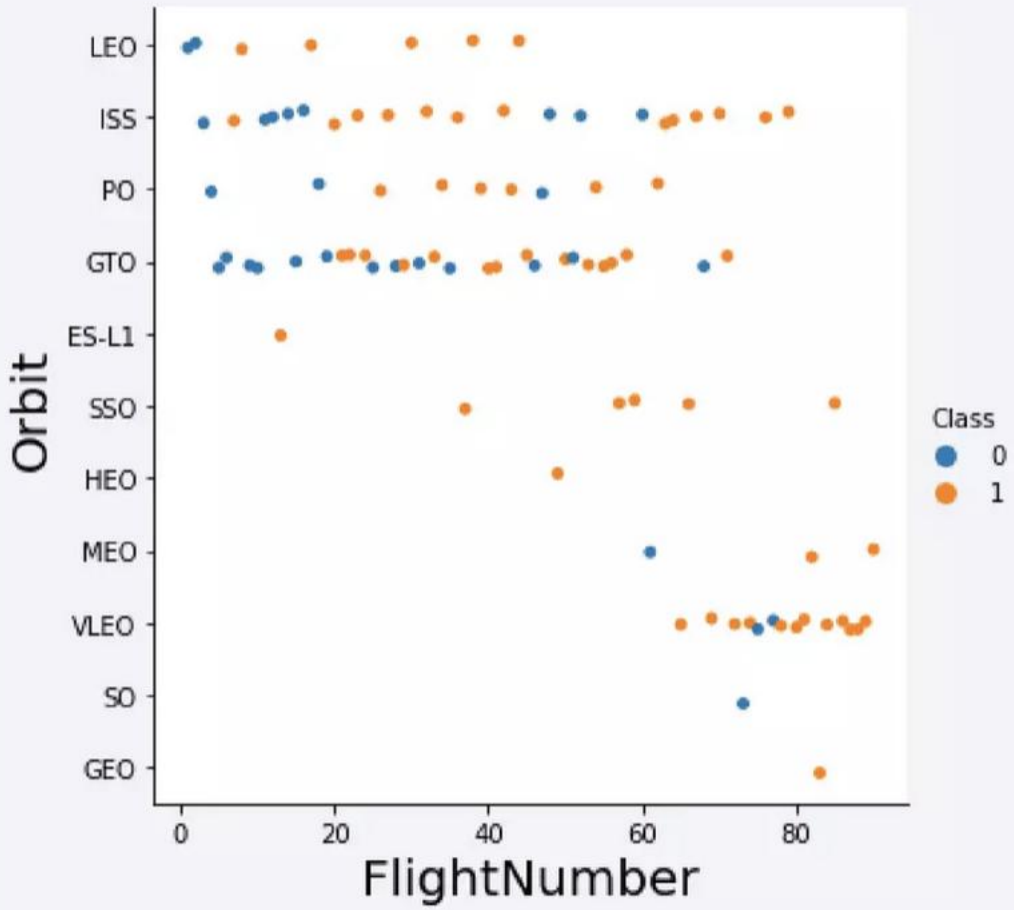
Success Rate vs. Orbit Type

- ES–L1, GEO,HEO,SSO shows the highest success rate



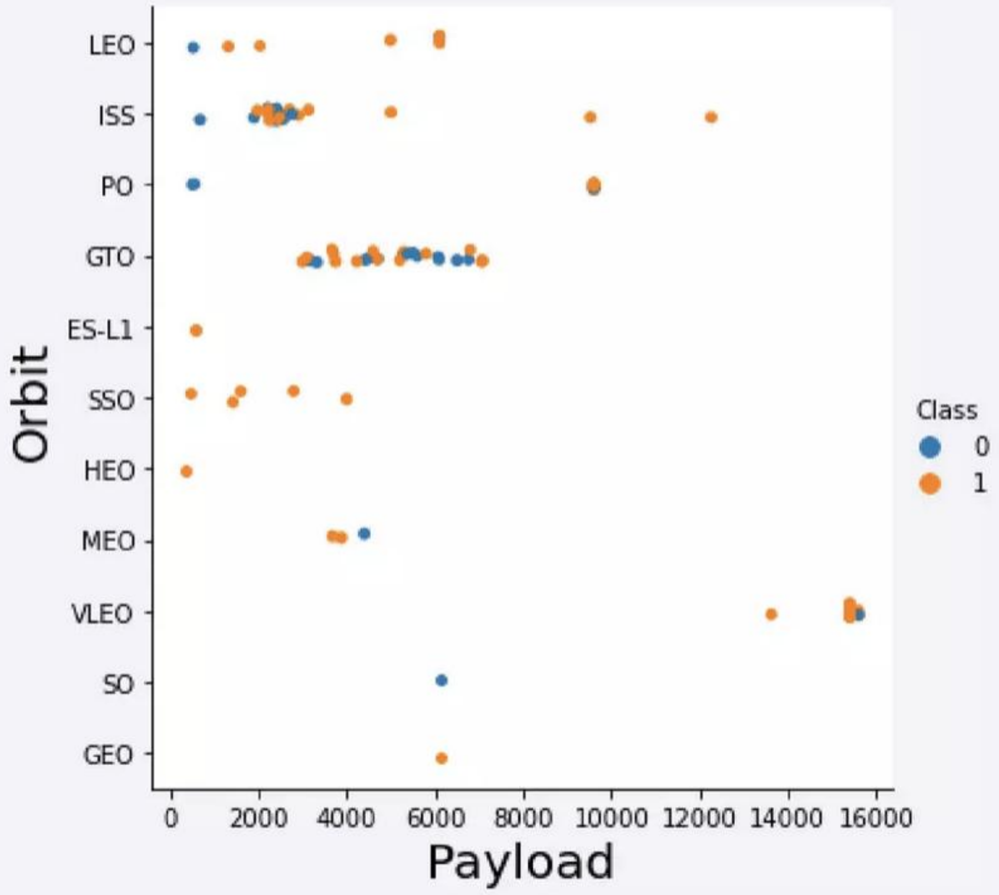
Flight Number vs. Orbit Type

- We notice a trend drawing whith VLEO launches



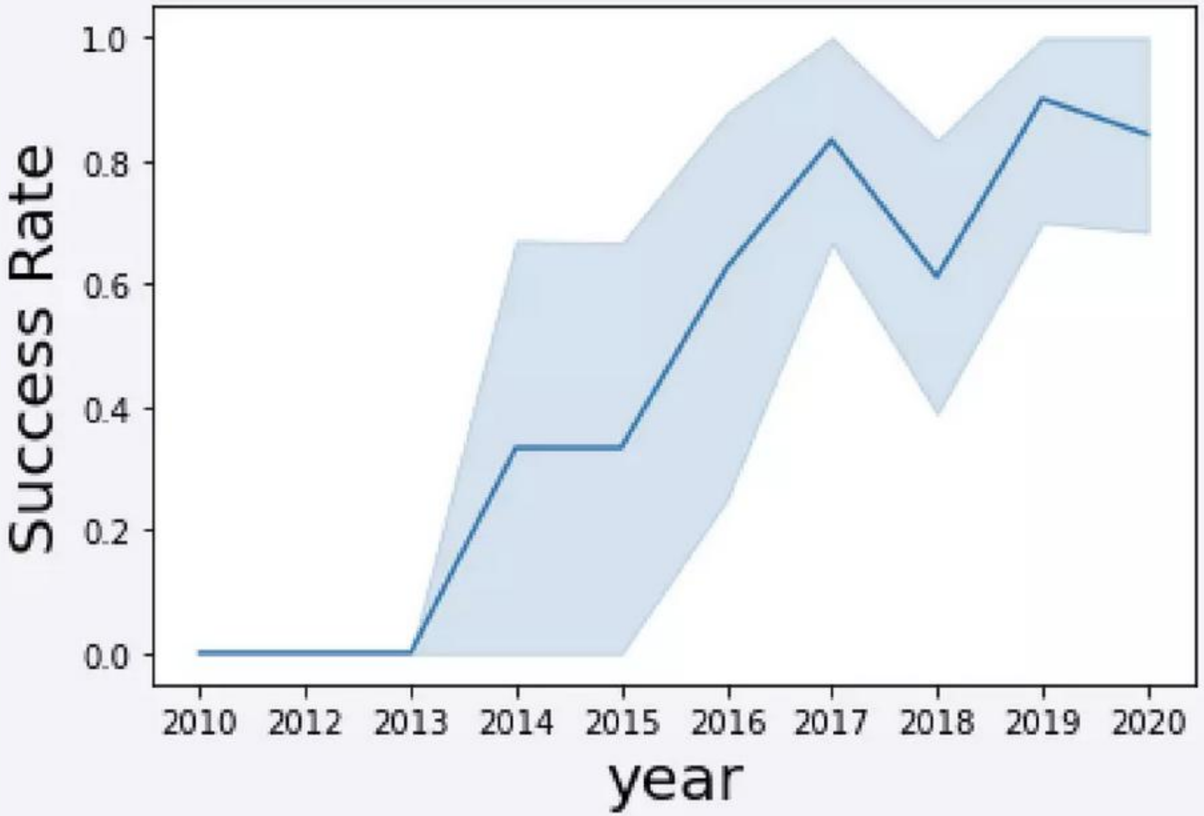
Payload vs. Orbit Type

- ISS and payload seems to be tightly correalted as well as GTO and the range of 4000–8000



Launch Success Yearly Trend

- We observe constant success rate increase from 2013



All Launch Site Names

- Query: “Select disctinct(launche_site)from SPACEXTBL”
- We select the column named launch_site with distinct

keyword which will elinimatethe occurences.

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

Launch Site Names Begin with 'CCA'

- SQL: “Select * from SPACEXTBL where launch_site like 'CCA%' limit 5”
- We select all data where launch site name starts with CCA and set a limit to the result returned

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

- SQL: “select sum(PAYLOAD_MASS_KG) from SPACEXTBL where customer='NASA (CRS)'”
- Sum keyword sums all masses in entries corresponding to the speciefied customer in this case NASA

Result : 45596

Average Payload Mass by F9 v1.1

- SQL: “select avg(PAYLOAD_MASS_KG) from SPACEXTBL where booster_vesrsion='F9 v1.1'”
- Avg keyword calculates the averagel payload masse for entries corresponding to the speciefied booster version in this case F9 v1.1

Result PAYLOAD_MASS_KG average : 2928.400000

First Successful Ground Landing Date

- SQL: “select min(DATE) from SPACEXTBL where Landing_Outcome='Success (ground pad)'”
- Min keyword evaluates the date for entries corresponding to the speciefied conditions in this case landing outcome success and output the first date

Result : 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

%sql select BOOSTER_VERSION from SPACEXTBL 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

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

100

Boosters Carried Maximum Payload

```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

%sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
[illegible]	[illegible]	CCAFS LC-	[illegible]	[illegible]	[illegible]	SKY Perfect JSAT	[illegible]	[illegible]

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

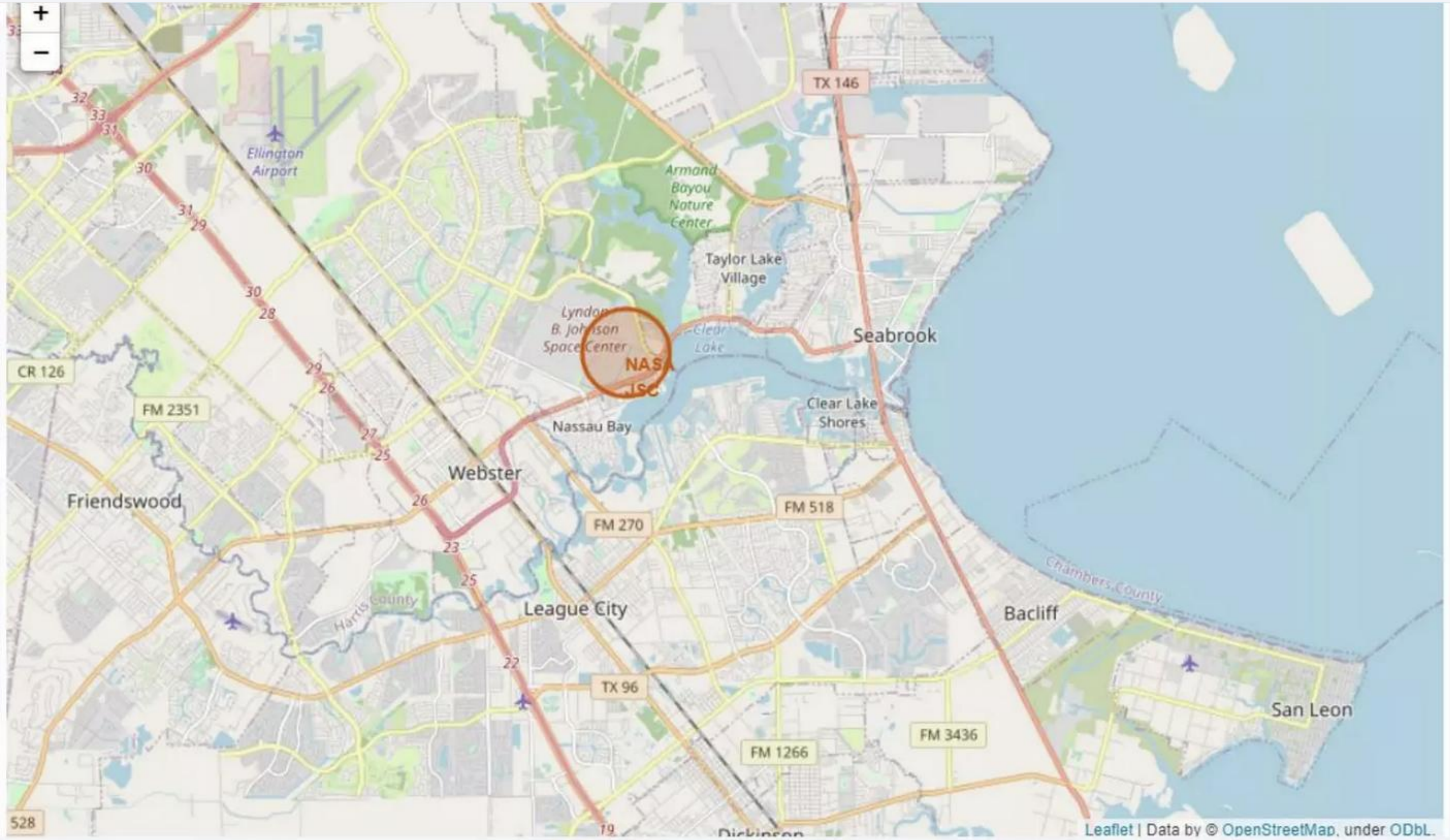
- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

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)

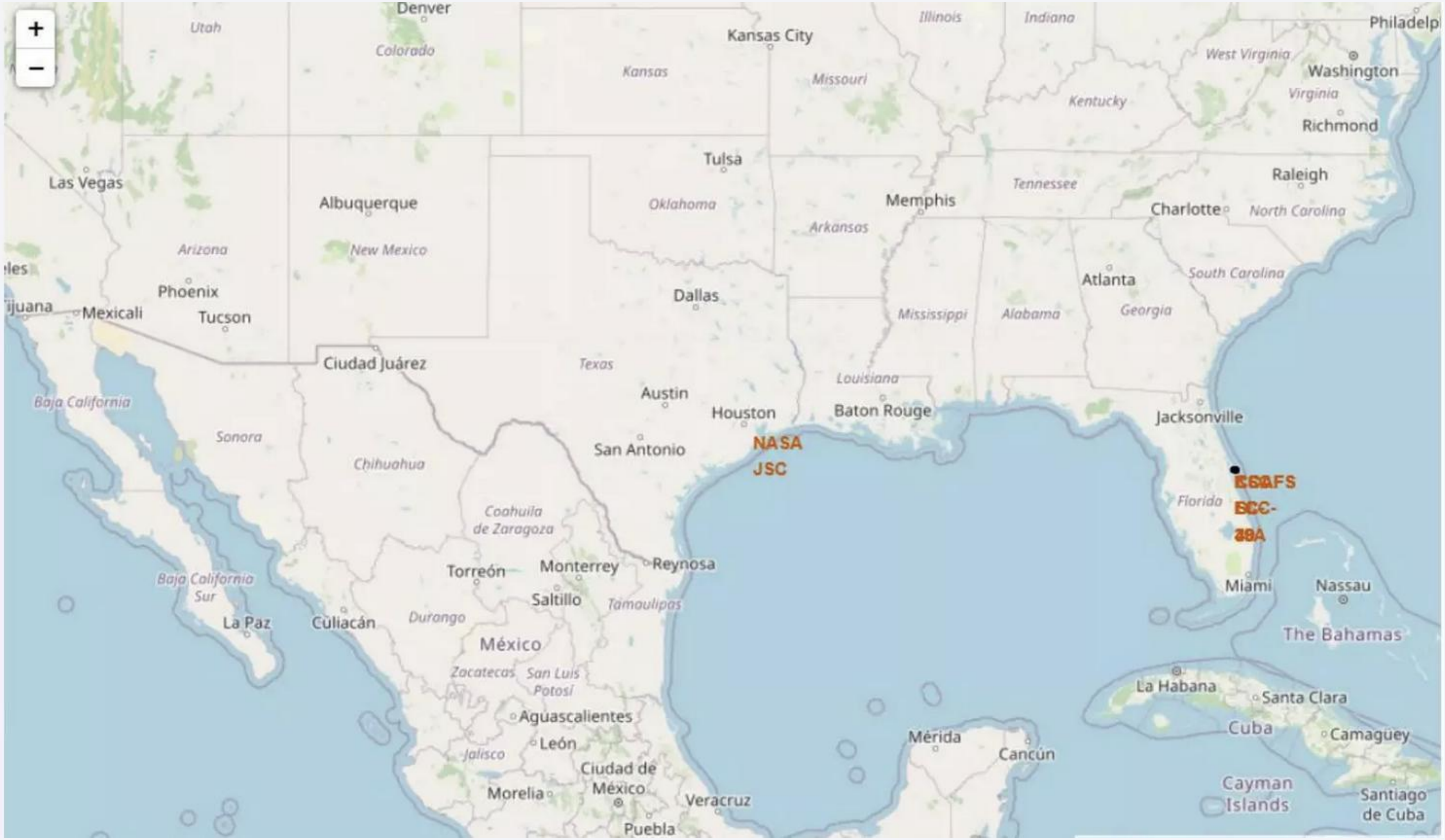
Section 3

Launch Sites Proximities Analysis

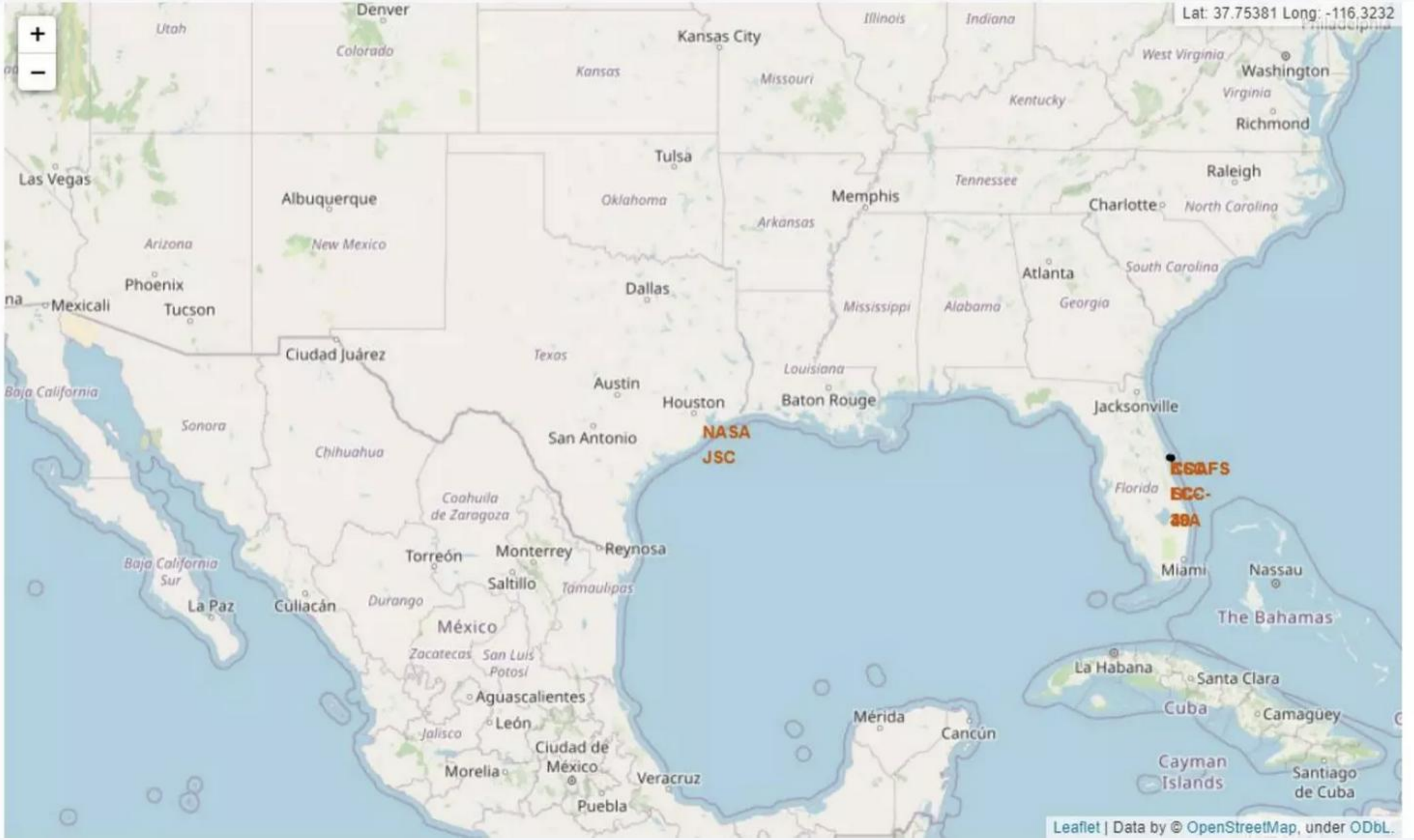
Launch site marked on a map



Launches with Sucess/fail marked on the map



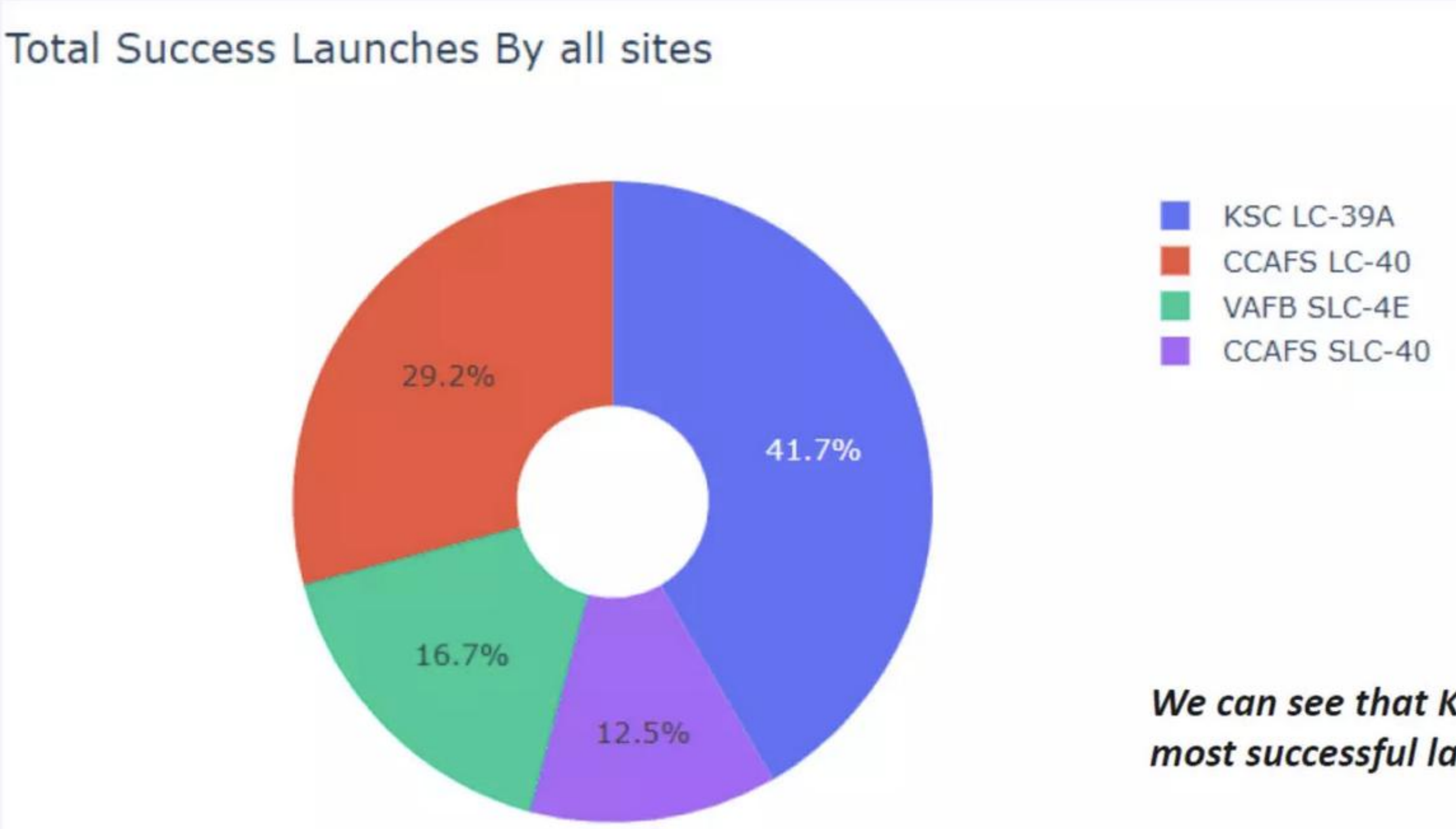
Distances between launche site to its proximities



Section 4

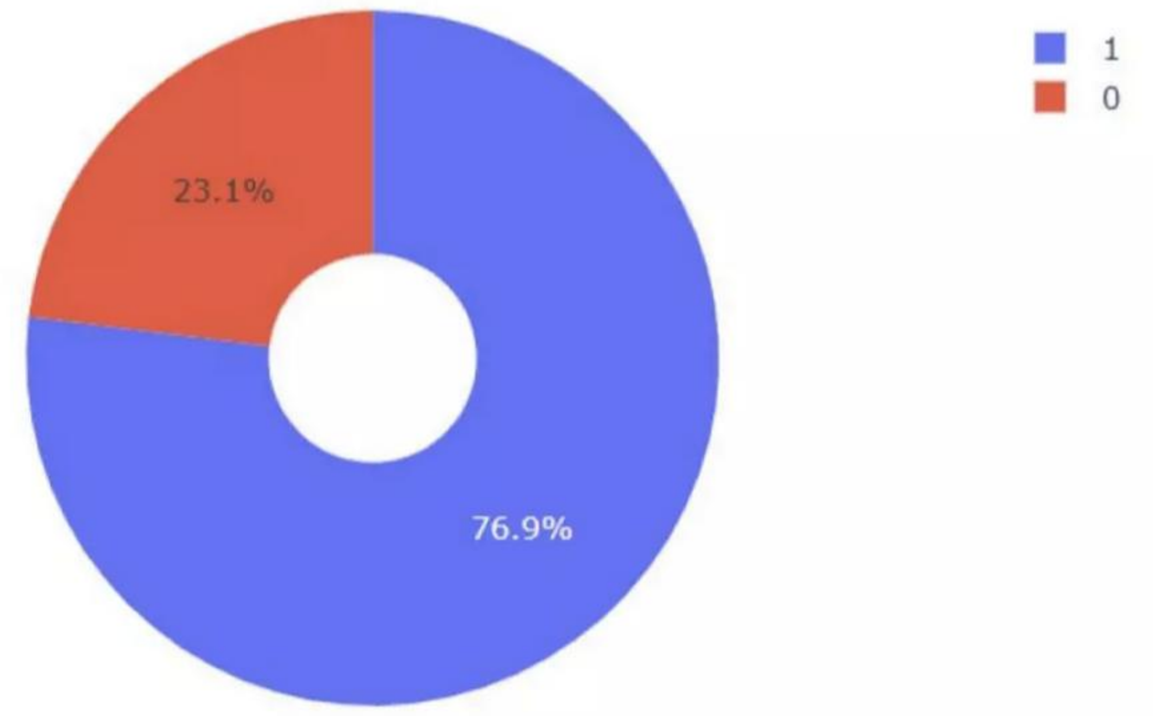
Build a Dashboard with Plotly Dash

Sucess lauches by Site



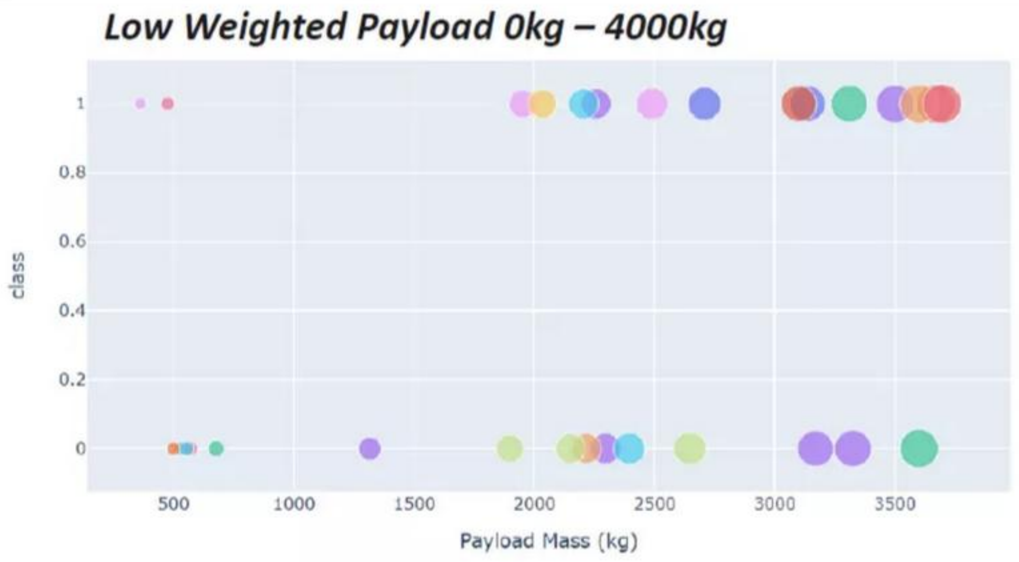
We can see that KSC LC-39A had the most successful launches from all the sites

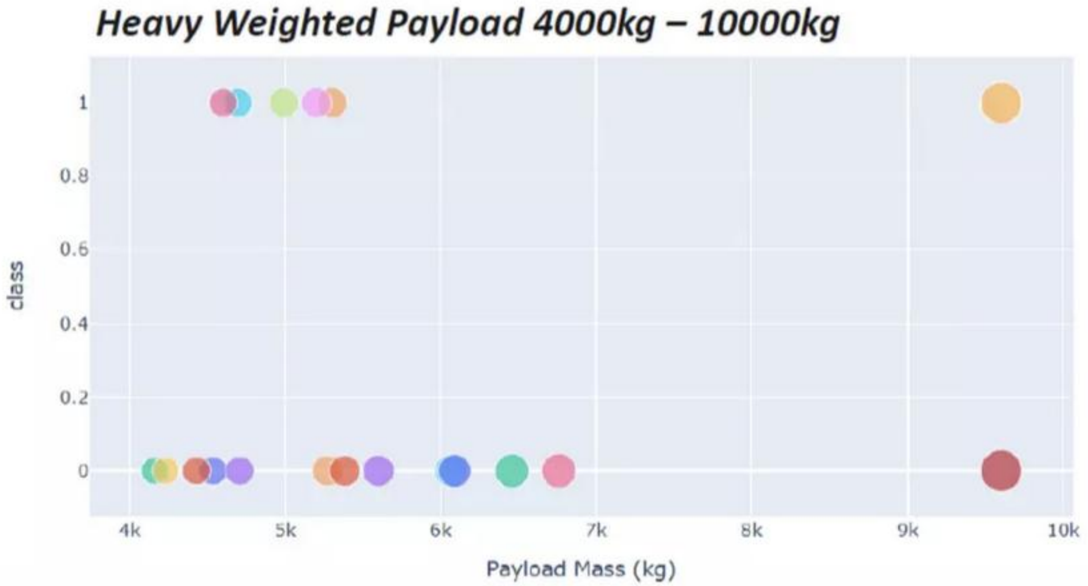
Success rate byt site



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs launch outcome



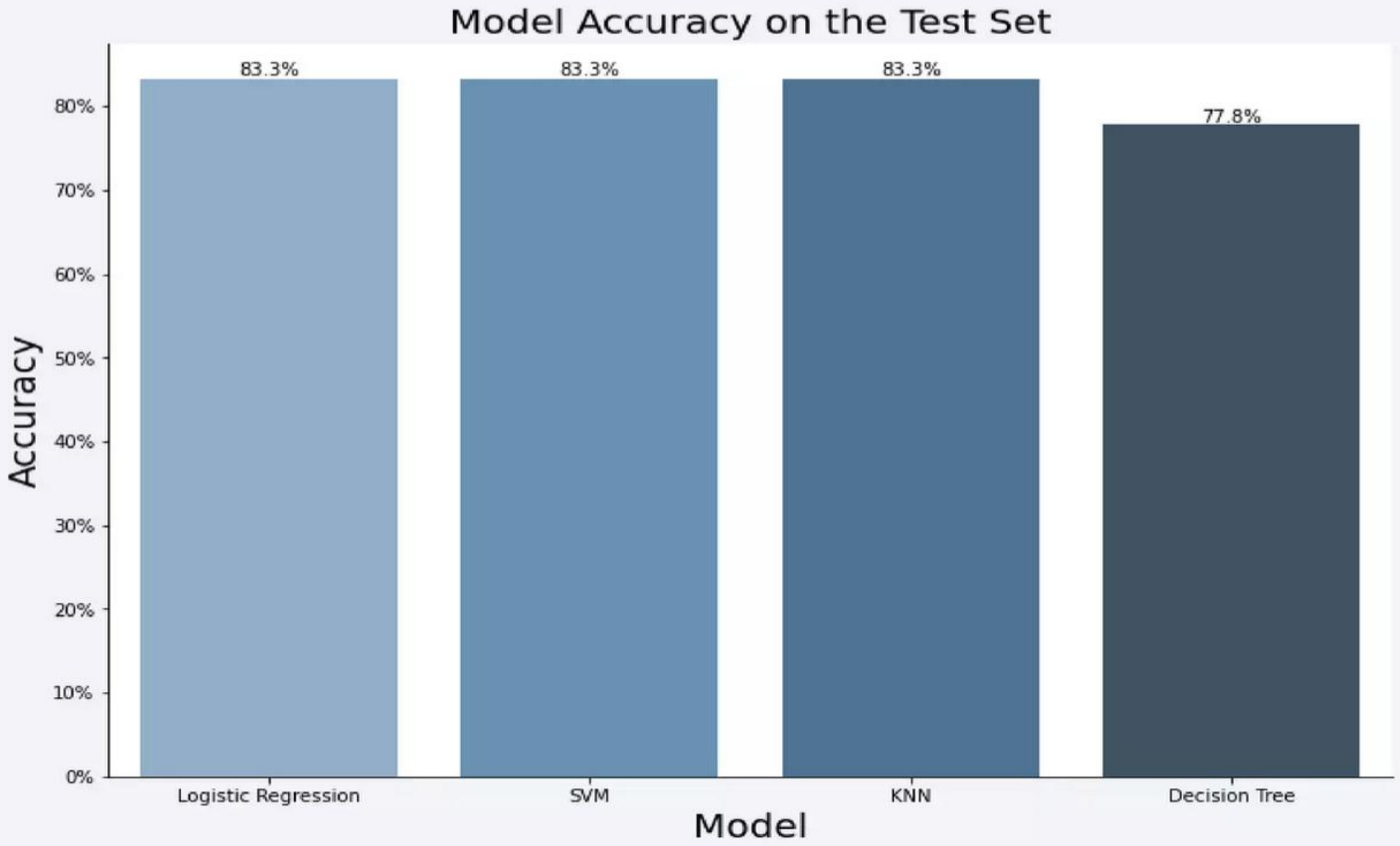


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

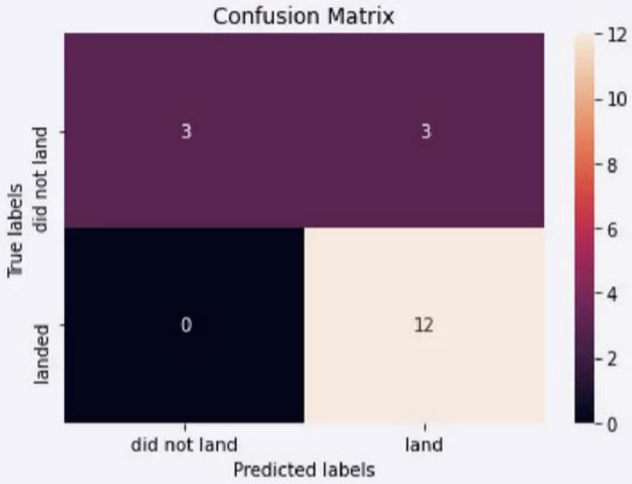
Section 5

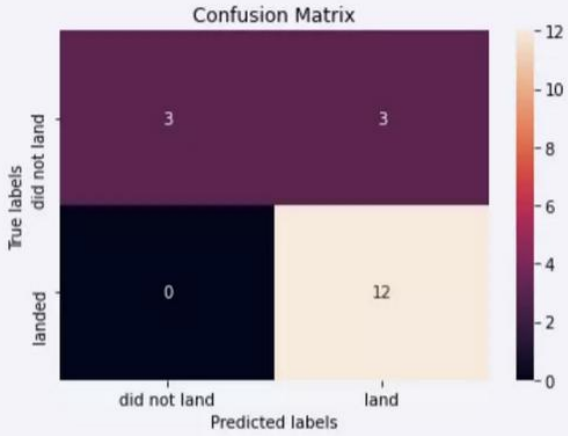
Predictive Analysis (Classification)

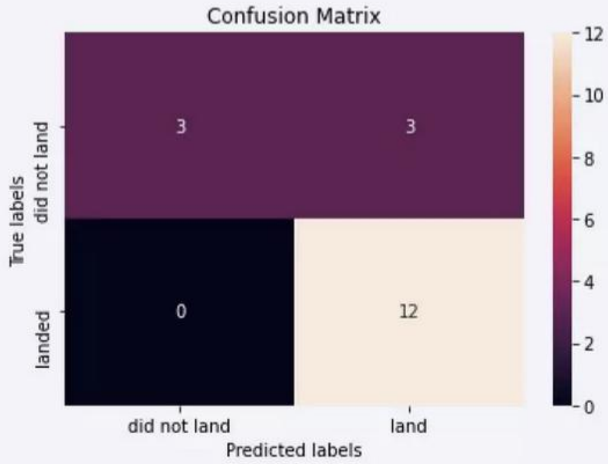
Classification Accuracy



Confusion Matrix







Conclusions

- SVM,KNN and logistic Regression models happen togive best accuracy for this dataset
- Outcomes show low weight payload have better perfomace
- Succes rate tends get better over the years.
- KSC LC 39A have better launch score
- GEO,HEO,SSO,ES L1 show higher success rate

Thank you!