Applied Data Science Capstone

Tysha Batts

November 2022

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications

EXECUTIVE SUMMARY

- Summary of methodologies
  - Data Collection via API, SQL, and Web scraping
  - Data Wrangling and Analysis
  - Interactive Maps with Folium
  - Predictive Analysis for each classification model
- Summary of Results
  - Low weighted payloads perform better than heavy payloads
  - KSC LC 39 A has the most successful launches from all sites
  - Orbit GEO, HEO, SSO, ES L1 has the best success rate
  - Success rate for SpaceX is directly proportional in years

INTRODUCTION

- Predict if the Flacon 9 first stage will land successfully.
- If we can predict if the first stage will land, we can determine the cost of the launch.

- Key Problems?
  - What is the price for each launch?
  - Will SpaceX reuse the first stage of Falcon 9?
  - What information (factors) can help to predict is Space X will reuse the first stage?

METHODOLOGY

- Data Collection
  - SpaceX Rest API
  - Web Scraping
- Data Wrangling
- Exploratory Data Analysis using SQL
- Interactive Visual Analytics with Folium and Ploty Dash
- Predictive Analysis

Data Collection

https://github.com/tyshsan13/testrepo/blob/main/Data%20Collection%20API%20Lab_TB.ipynb

Getting response from API

Converting response to a .json file

Cleaning data

Assign list to dictionary

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

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

```
data = pd.json_normalize(response.json())
```

```
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}
```

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

Data Collection -Scrapping

https://github.com/tyshsan13/testrepo/blob/main/Web%20Scrapping%20Lab_TB.ipynb

1. Getting response from HTML

```
r=requests.get(static_url).text
```

2. Creating Beautiful Soup Object

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

3. Finding tables

```
html_tables = soup.find_all('table')
```

4. Getting column names

```
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if (name != None and len(name) > 0):
        column_names.append(name)
```

5. Creating dictionary

```
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']=[]
launch_dict['Time']=[]
```

6. Converting dictionary to dataframe

```
df=pd.DataFrame(launch_dict)
```

7. Dataframe to .CSV

Data Wrangling

https://github.com/tyshsan13/testrepo/blob/main/Data%20Wrangling%20Capstone_TB.ipynb

1. Calculate number of launches at each site

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

2. Calculate number and occurrence at each orbit

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

3. Calculate number and occurrence of mission outcome per orbit

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

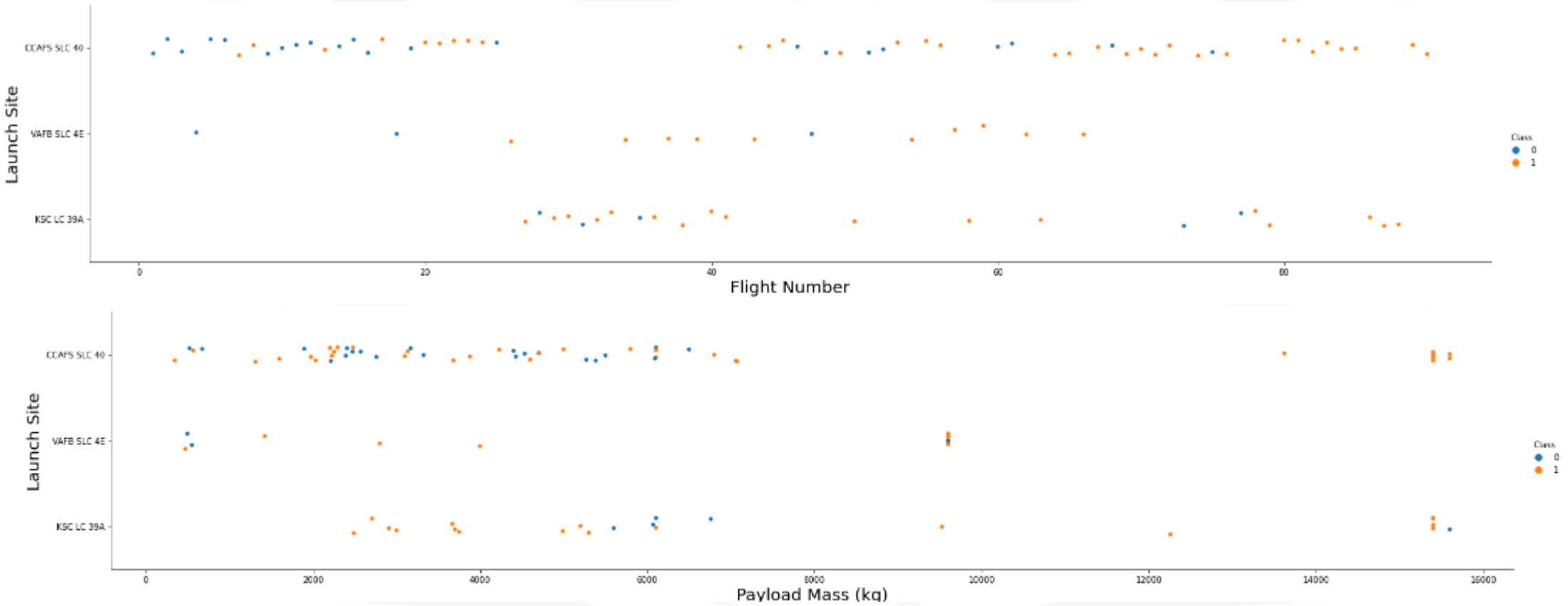
4. Create landing outcome label from Outcome column

5. Export data as .csv

```
df['Class']=landing_class
df[['Class']].head(8)
```

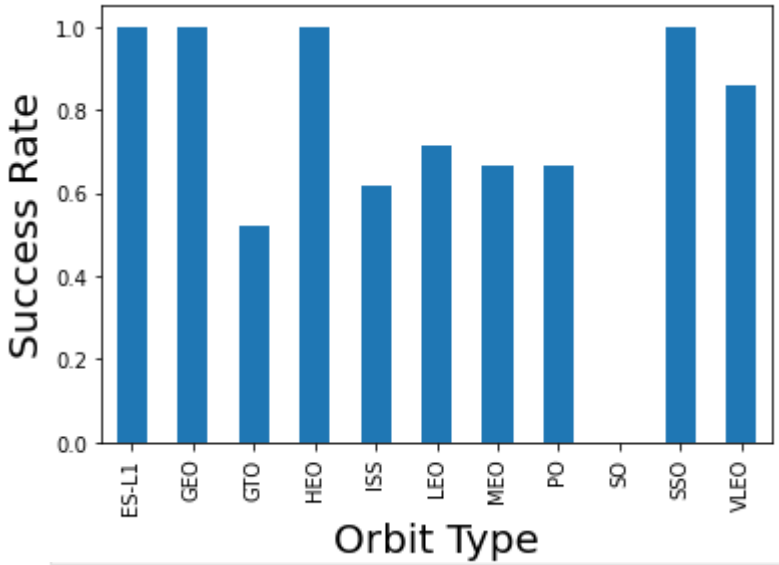
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitu
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561
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	28.561
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	34.632
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	28.561

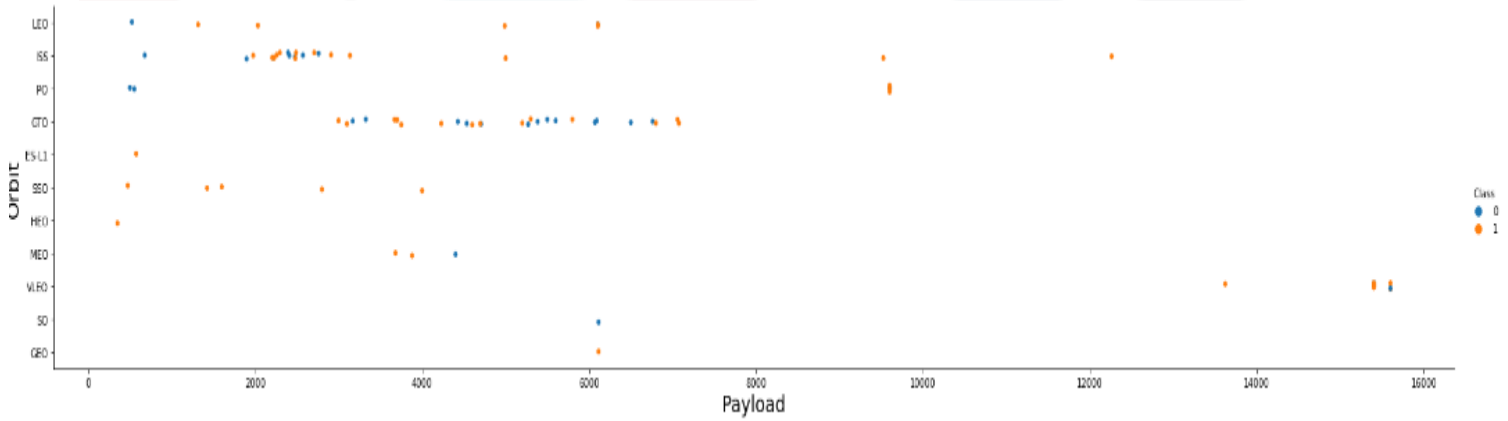
EDA with Data Visualization

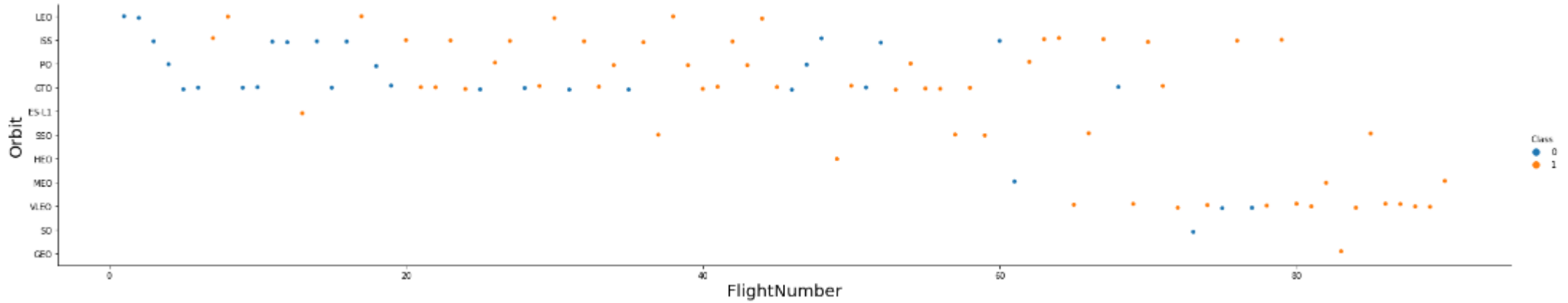


https://github.com/tyshsan13/testrepo/blob/main/EDA%20with%20Visualization_TB.ipynb

EDA with Data Visualization







EDA with SQL

https://github.com/tyshsan13/testrepo/blob/main/EDA%20with%20SQL.ipynb

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites being with ‘CCA”

```
%%sql
SELECT DISTINCT LAUNCH_SITE
FROM SPACEXTBL;
```

- Display the total payload mass carried by boosters

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Customer = 'NASA (CRS)';
```

- Display the average payload

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.0%';
```

EDA with SQL

- List the date when the first successful outcome was achieved

```
%%sql
SELECT MIN(Date)
FROM SPACEXTBL
WHERE "Landing__Outcome" = "Success (ground pad)";
```

- List the name of the boosters greater than 4000 but less than 6000

```
%%sql
SELECT "BOOSTER_VERSION"
FROM SPACEXTBL
WHERE "LANDING__OUTCOME" = "Success (drone ship)"
    AND 4000 < PAYLOAD_MASS__KG_ < 6000;
```

- List the total number of successful and failure mission

```
%%sql
SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS TOTAL_NUMBER
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME;
```

EDA with SQL

- List the name of the booster_versions which have carried maximum pavload

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

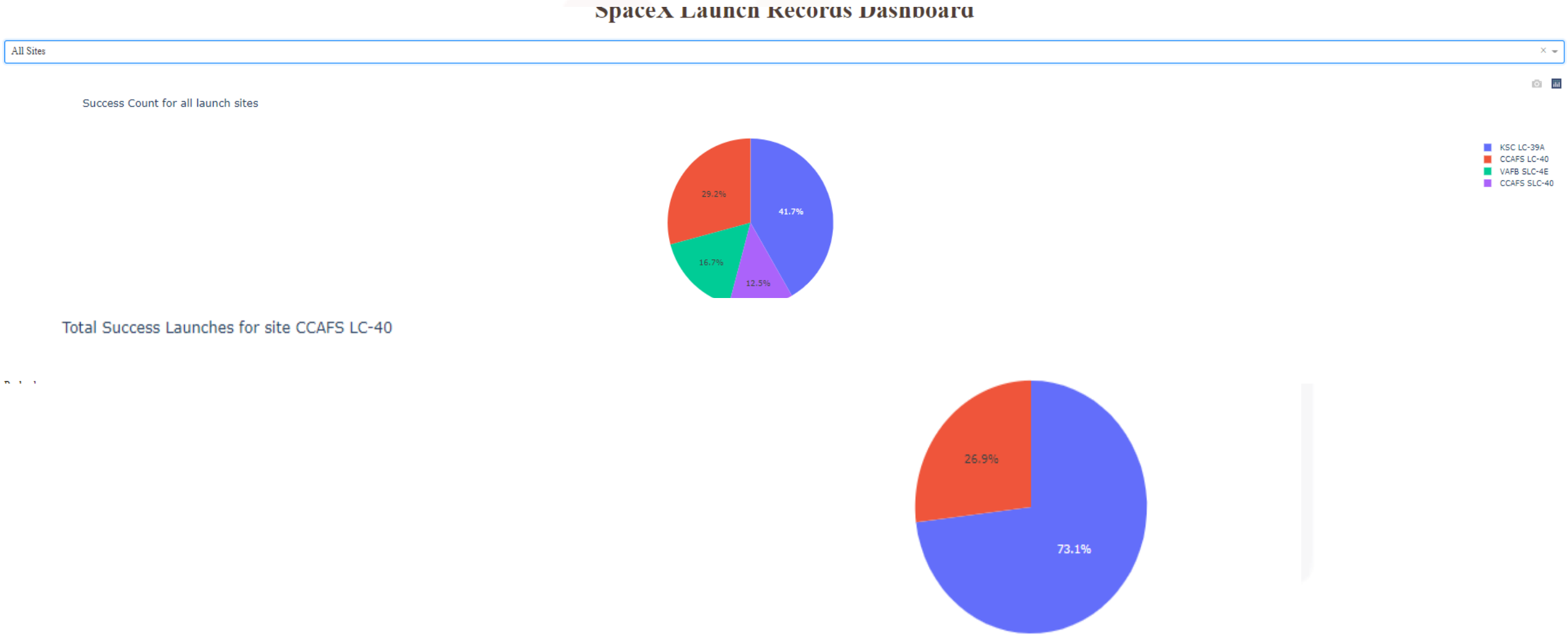
- List the records for months in year 2015

```
%%sql
SELECT "LANDING__OUTCOME", "BOOSTER_VERSION", "LAUNCH_SITE"
FROM SPACEXTBL
WHERE "Landing__Outcome" = "Failure (drone ship)"
    AND "substr(DATE)" = '2015';
```

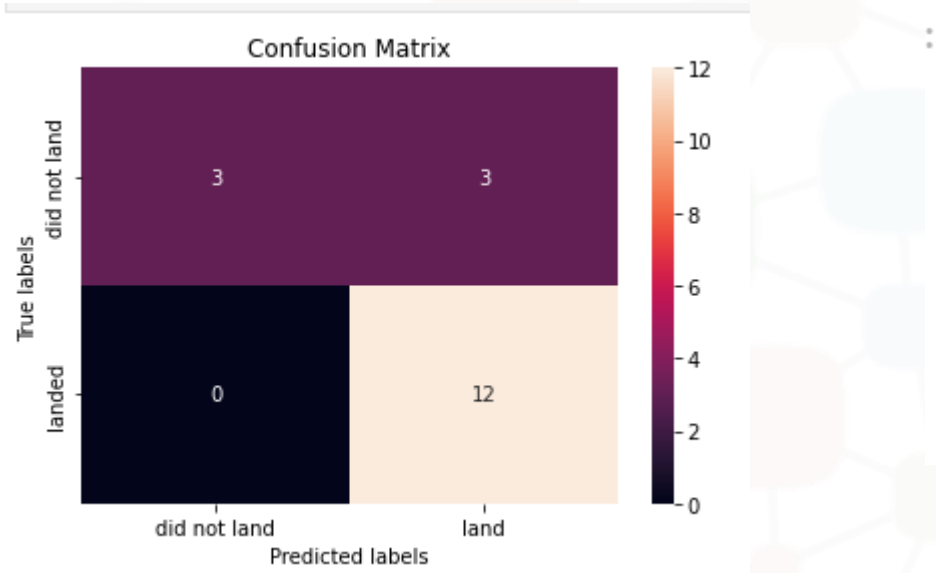
- Rank the count of successful landing_outcomes

```
%%sql
SELECT "LANDING__OUTCOME", "COUNT(LANDING__OUTCOME)" AS "TOTAL_NUMBER"
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "LANDING__OUTCOME"
ORDER BY "TOTAL_NUMBER" DESC
```

Build a Dashboard with Plotly



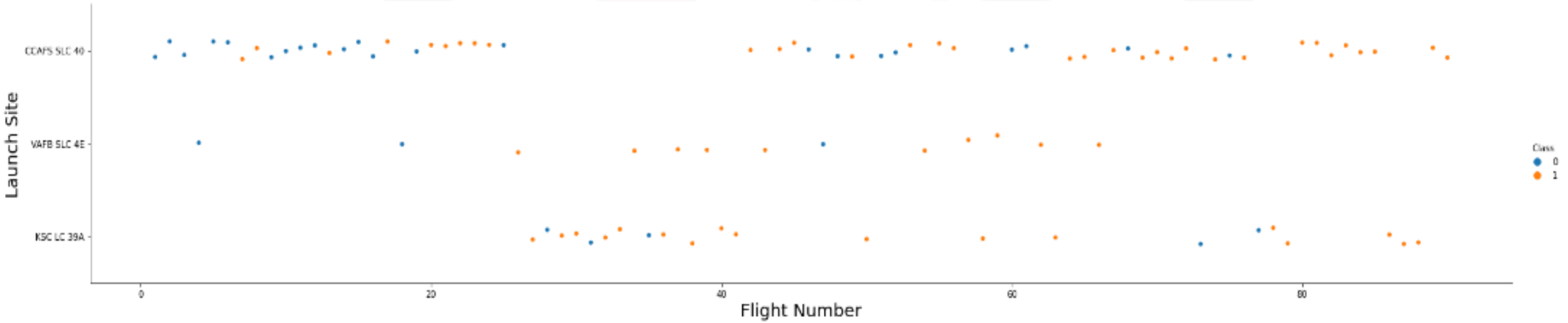
Predictive Analysis



	Accuracy
KNN	0.848214
Decision Tree	0.875000
Logistic Regression	0.846429
SVM	0.848214

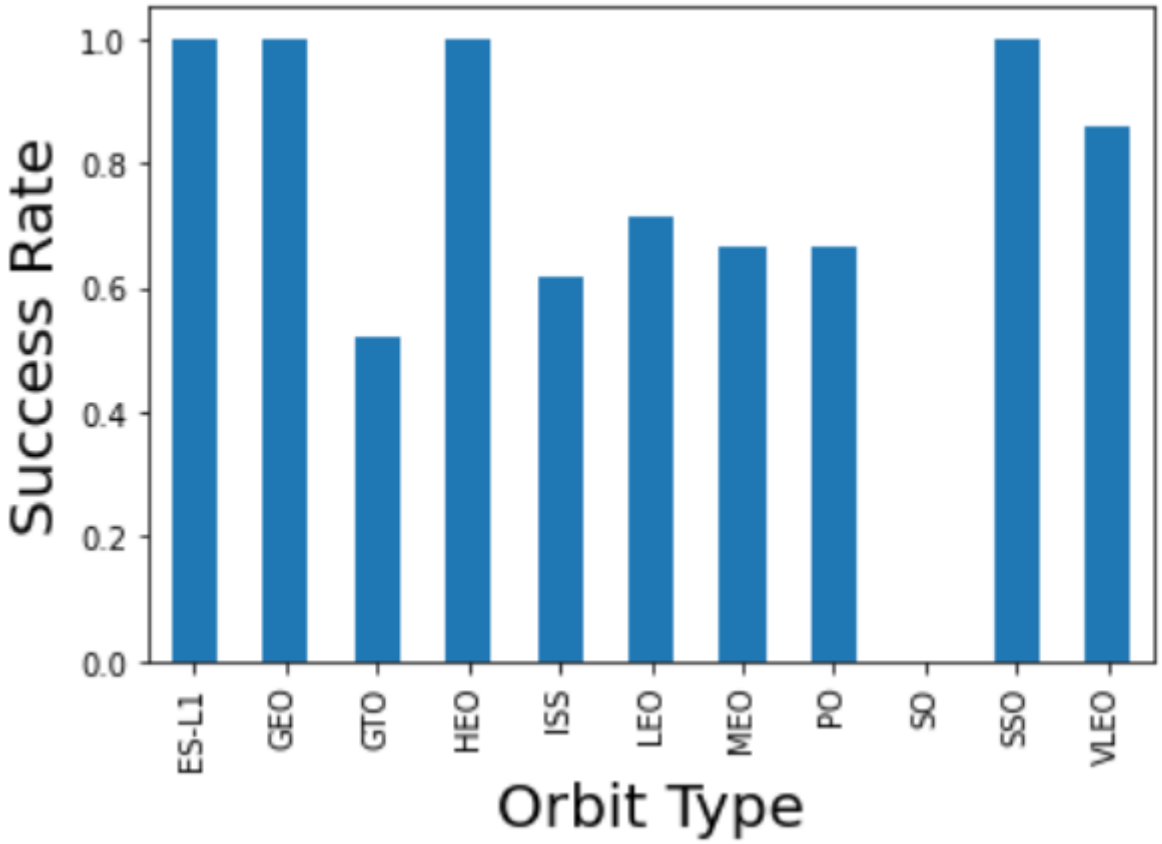
https://github.com/tyshsan13/testrepo/blob/main/Machine%20Learning%20Prediction%20Lab_TB.ipynb

RESULTS- Flight Number vs. Launch Sites



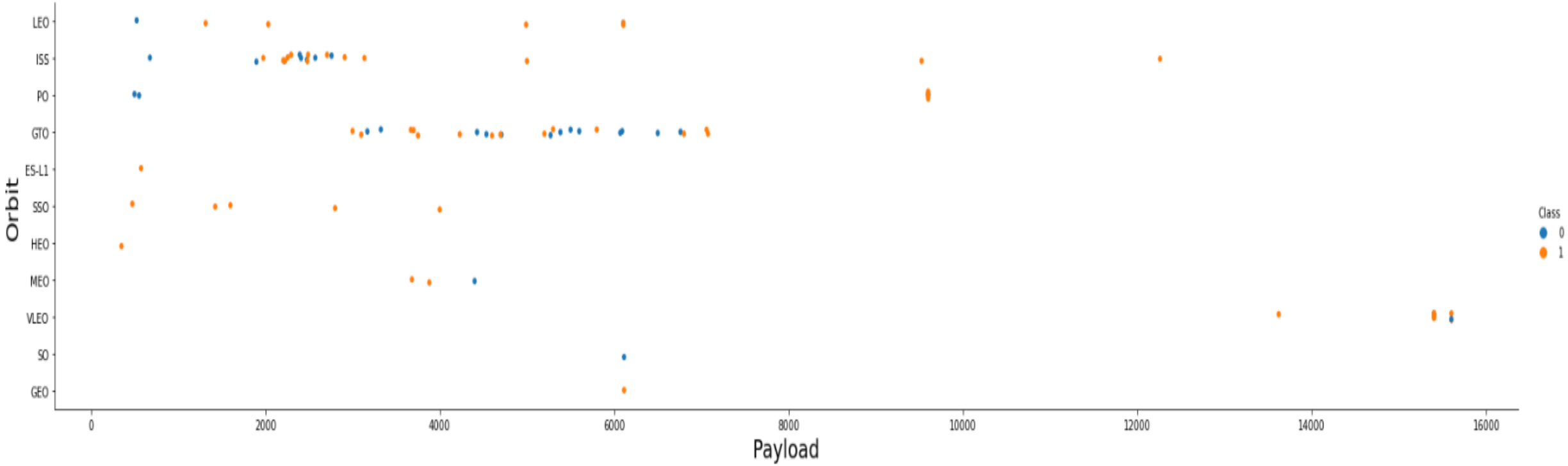
We see that flight numbers greater than 30 have a better success rate.

Results – Success Rate Vs. Orbit



You see that ES-L1, GEO, HEO, and SSO have the highest success rates.

Results – Payload vs. Orbit Type



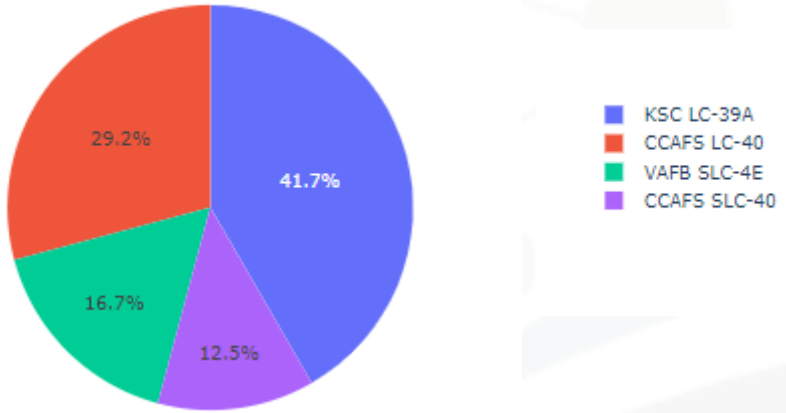
We see that Polar, LEO, and ISS have more of a success rate with heavy payloads.

Discussion- Findings and Implications

	Accuracy
KNN	0.848214
Decision Tree	0.875000
Logistic Regression	0.846429
SVM	0.848214

The Decision Tree is best in terms of the SVM, KNN, and Logistic models.

The KSC LC 39 A had the most successful launches from all sites.



Discussion- Findings and Implications

Low weighted payloads have a better success rate than have payloads.