Winning Space Race with Data Science

Lei Diao
06/08/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Data visualization
  - Exploratory Data Analysis (EDA) with SQL
  - Building interactive map with Folium
  - Building a Dashboard
  - Classification model analysis
- Summary of all results
  - EDA results
  - Predictive analytics

Introduction

- The commercial space age is coming. Companies such as Virgin Galactic, Blue Origin and SpaceX are making space travel affordable for everyone.
- As a startup company, SpaceY would like to compete with SpaceX on Sending spacecraft to the out space.
- By gathering information about SpaceX, machine learning methodologies are used to determine the price of each launch.
- Data are also used to determine if SpaceX will reuse the first stage.

Section 1

Methodology

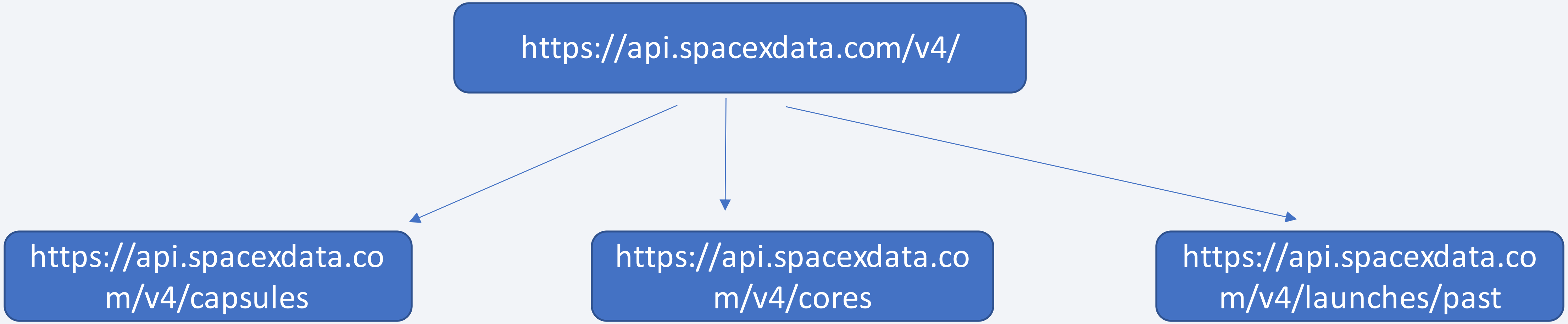
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX launch data is collected from SpaceX REST API
  - Falcon 9 launch data is web scraping related Wiki pages
- Perform data wrangling
  - Landing outcomes in data set is converted to class with value 0 (Failure) or 1 (success)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, SVM, DT, KNN models have been built and evaluated for the best classifier

Data Collection

- SpaceX launch data is collected from SpaceX REST API



- Falcon 9 launch data is web scraping related Wiki page

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Data Collection – SpaceX API

- Data collection with SpaceX REST calls

Getting responds from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
```

Convert .json file to a dataframe

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

Data clean

```
Pmean=data_falcon9['PayloadMass'].mean()
data_falcon9['PayloadMass']=data_falcon9['PayloadMass'].replace(np.nan, Pmean)
```

Data Collection - Scraping

- Web scraping process from Wikipedia

Getting responds from HTML

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

Creating beatifulSoup object and finding tables

```
soup = BeautifulSoup(response.text, "html.parser")
html_tables = soup.find_all('table')
```

Getting column names

Creation of dictionary

Appending data to keys

Converting dictionary to dataframe

```
del launch_dict['Date and time ( )']

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']=[]
```

Data Wrangling

1 Calculate the number of launches on each site

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

2 Calculate the number of occurrence of each orbit

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

3 Calculate the number of occurrence of mission outcome per orbit type

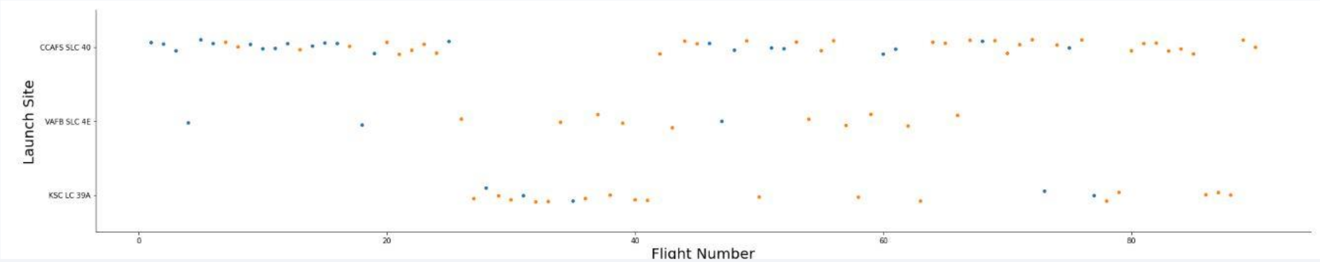
```
landing_outcomes=df['Outcome'].value_counts()
```

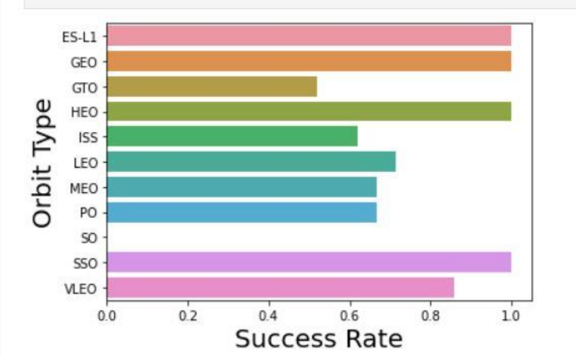
4 Create a landing outcome label from Outcome column

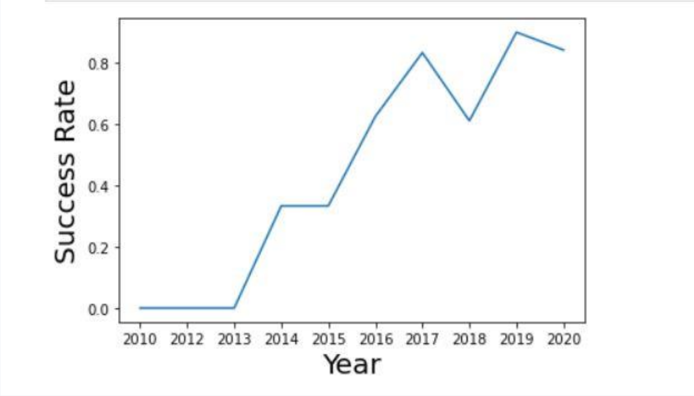
```
df['Class'] = df['Outcome'].apply(lambda x: 0 if bad_outcomes.count(x)>0 else 1)
```

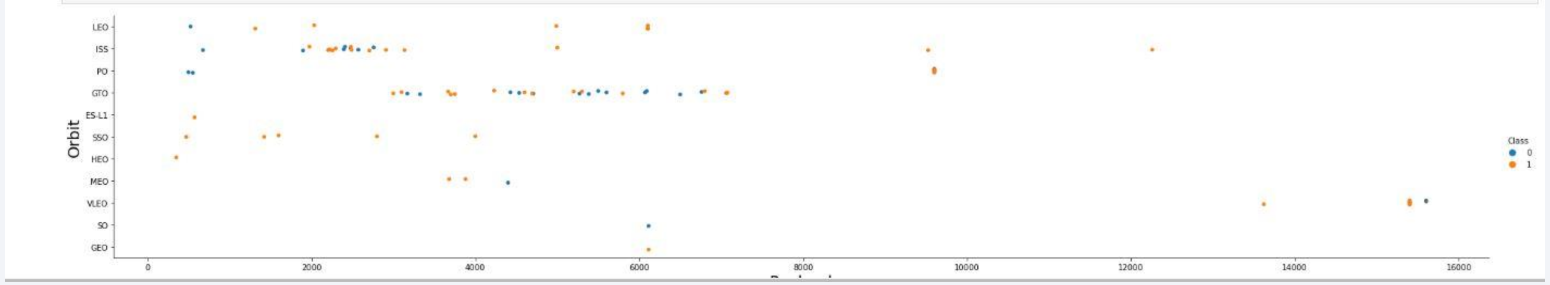
5 Save to .csv

EDA with Data Visualization





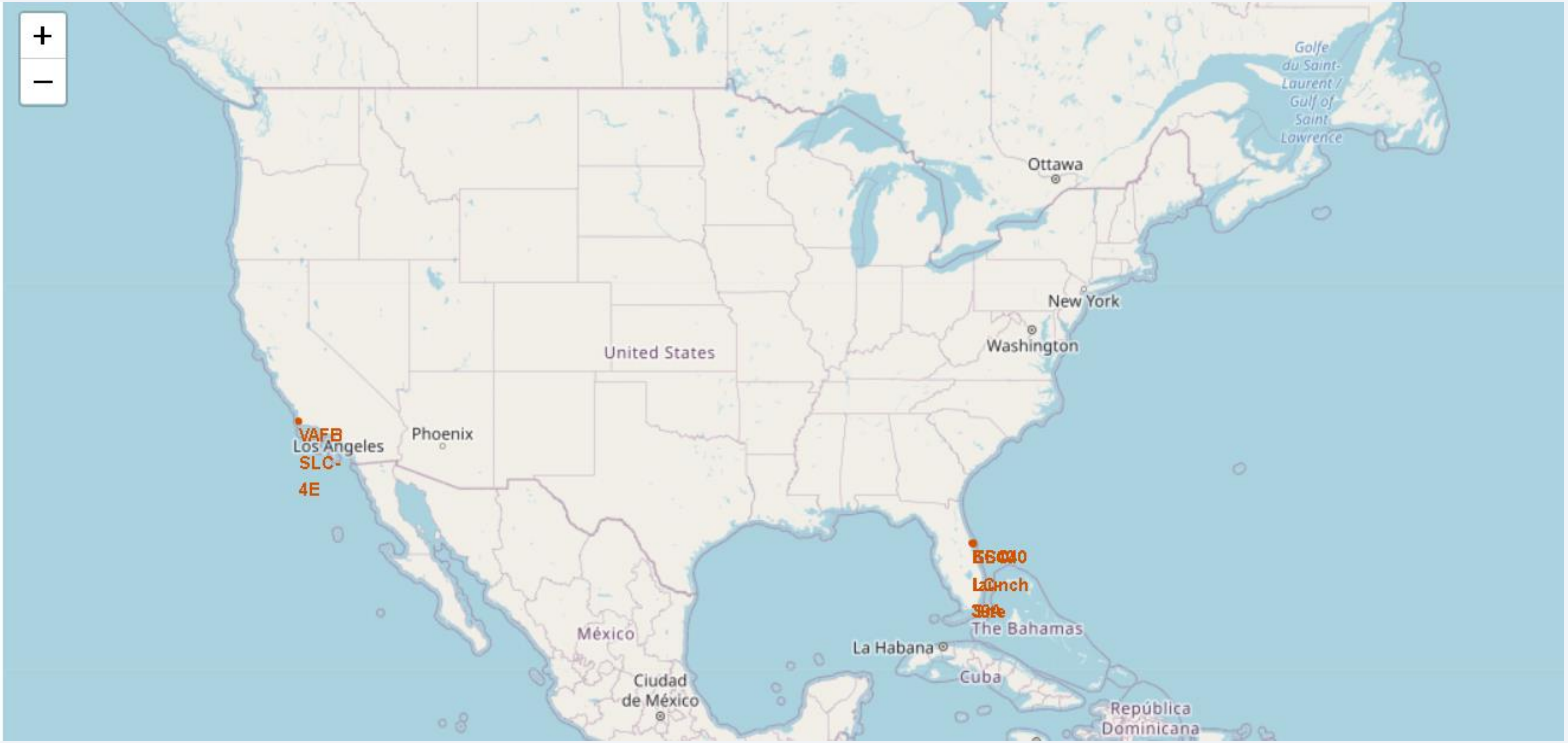




EDA with SQL

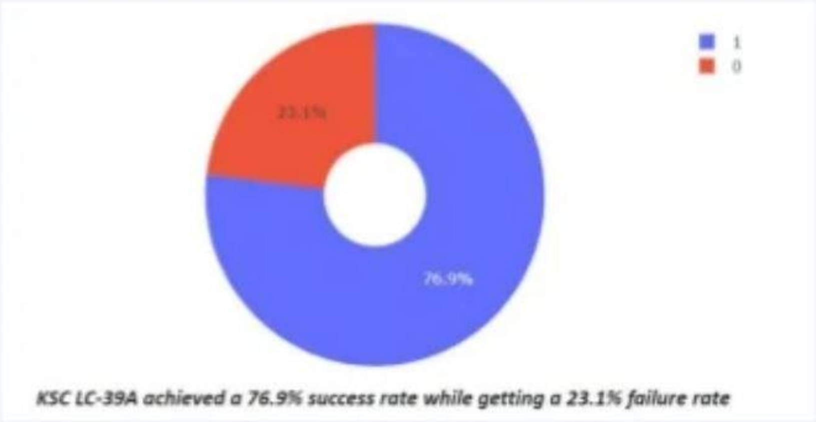
- Using SQL queries to perform:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved
  - 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 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

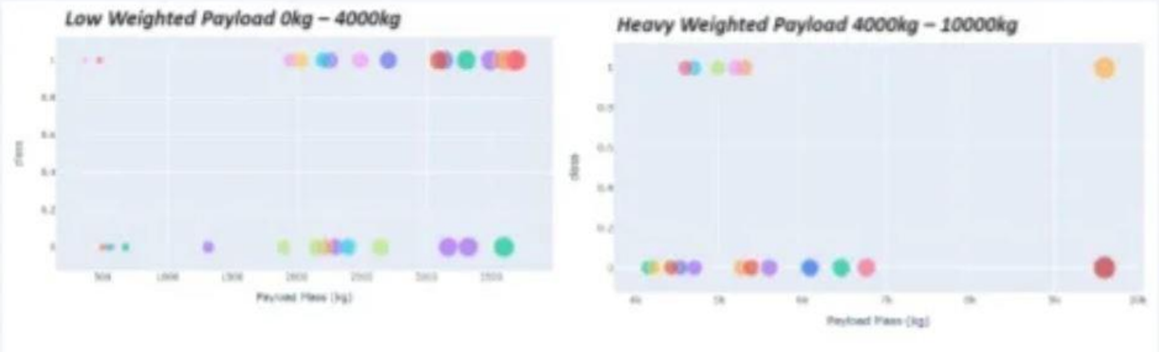
Build an Interactive Map with Folium



- Map markers have been added to the map to find the best location for building a launch site.

Build a Dashboard with Plotly Dash

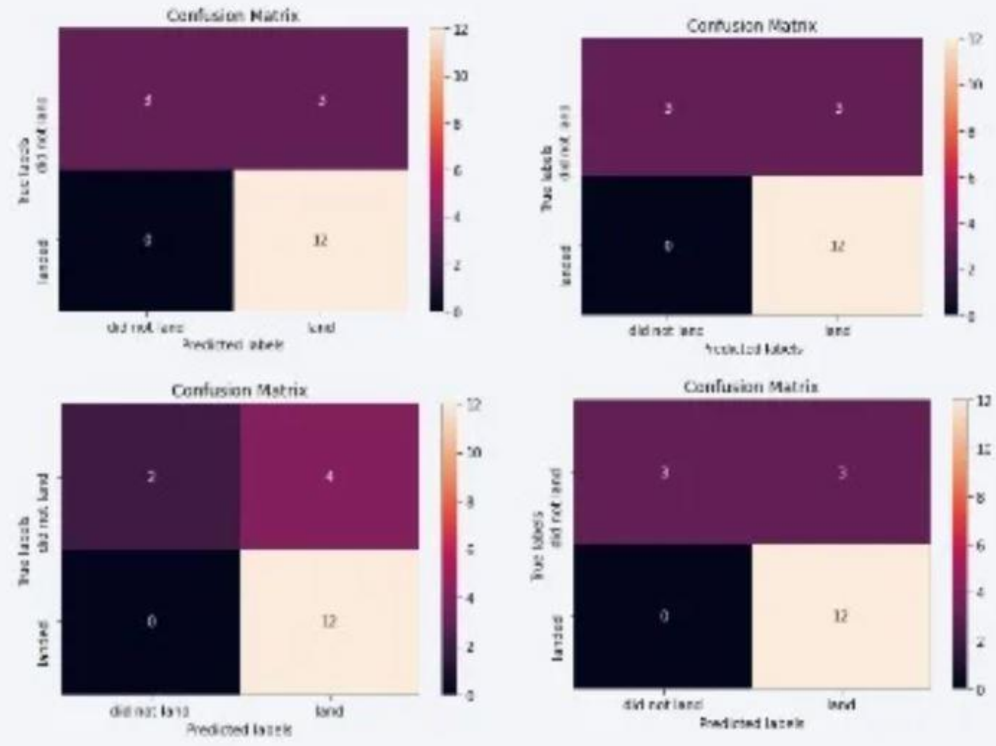


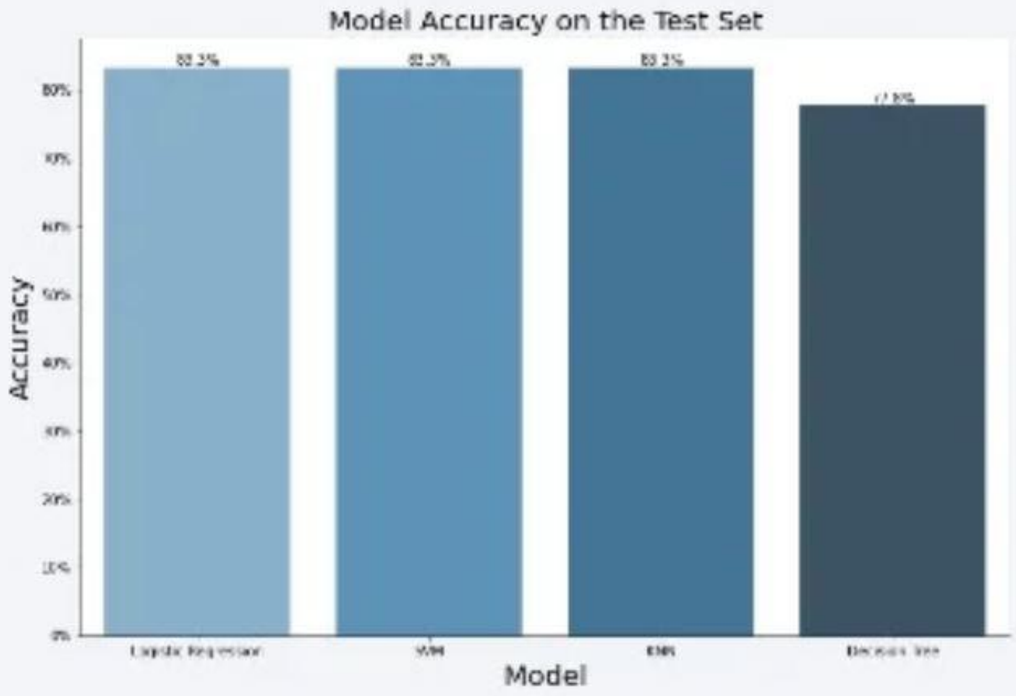


- Pie charts and scattering charts were added to the dashboard with dropdown interactions to show success rates for all sits or individual site.

Predictive Analysis (Classification)

- The SVM, KNN, and the logistic Regression model perform best with the highest accuracy at 83.3%





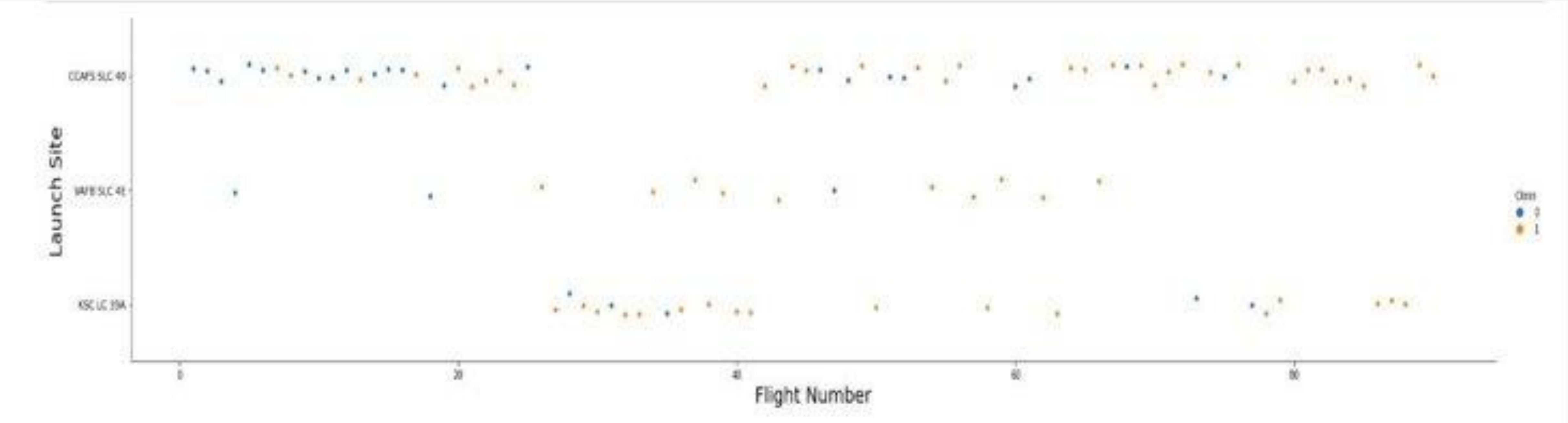
Results

- Launches with low payload mass performs better than that with high payload mass
- The SpaceX launch success rates is getting higher with time in years
- In all orbits, GEO, SSO, HEO and ES L1 has the highest success rates
- In all launch sites, KSC LC-39A has the highest success rates
- The SVM, KNN, and the logistic Regression model perform best with highest accuracy

Section 2

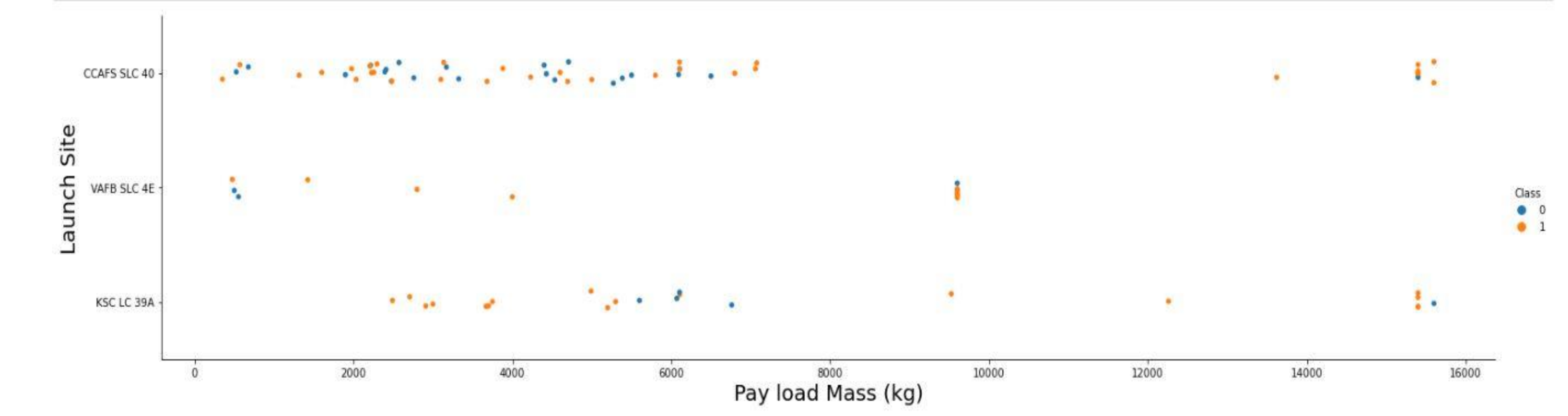
Insights drawn from EDA

Flight Number vs. Launch Site



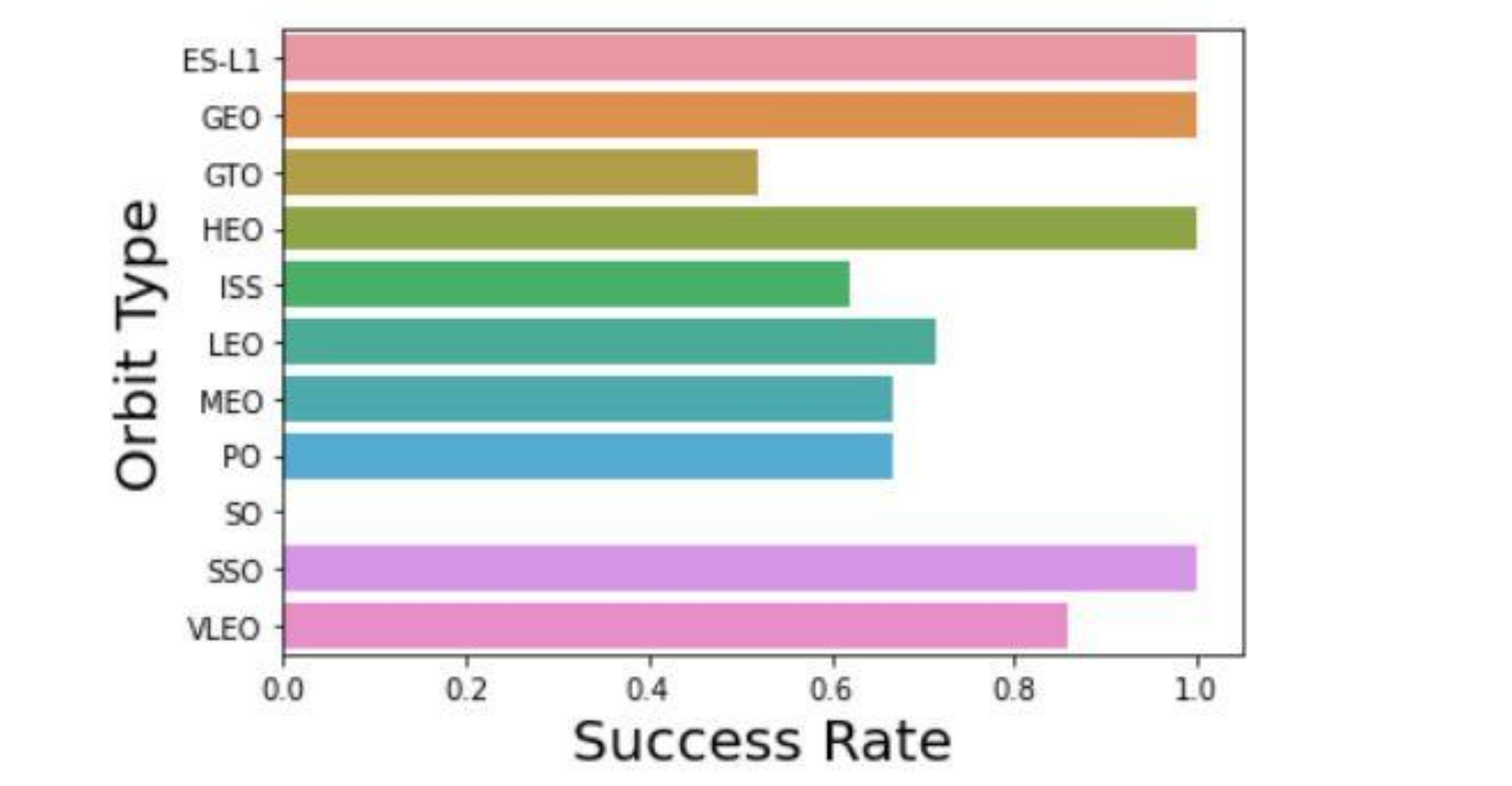
- Launch site CCAFS SLC-40 has the most number of launches

Payload vs. Launch Site



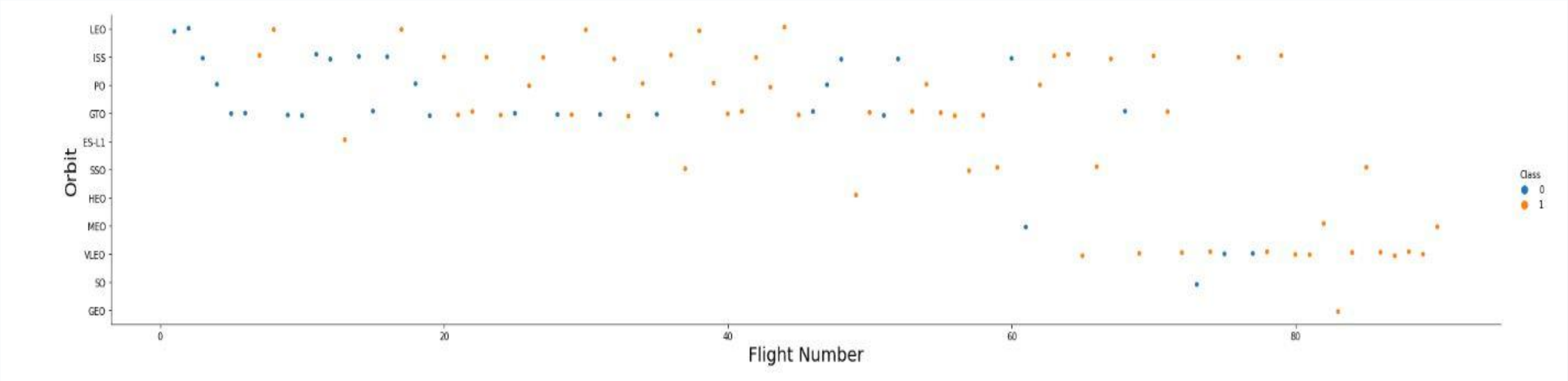
- For the VAFB SLC-4E launch site there are no rockets launched for heavy payload mass(greater than 10000)

Success Rate vs. Orbit Type



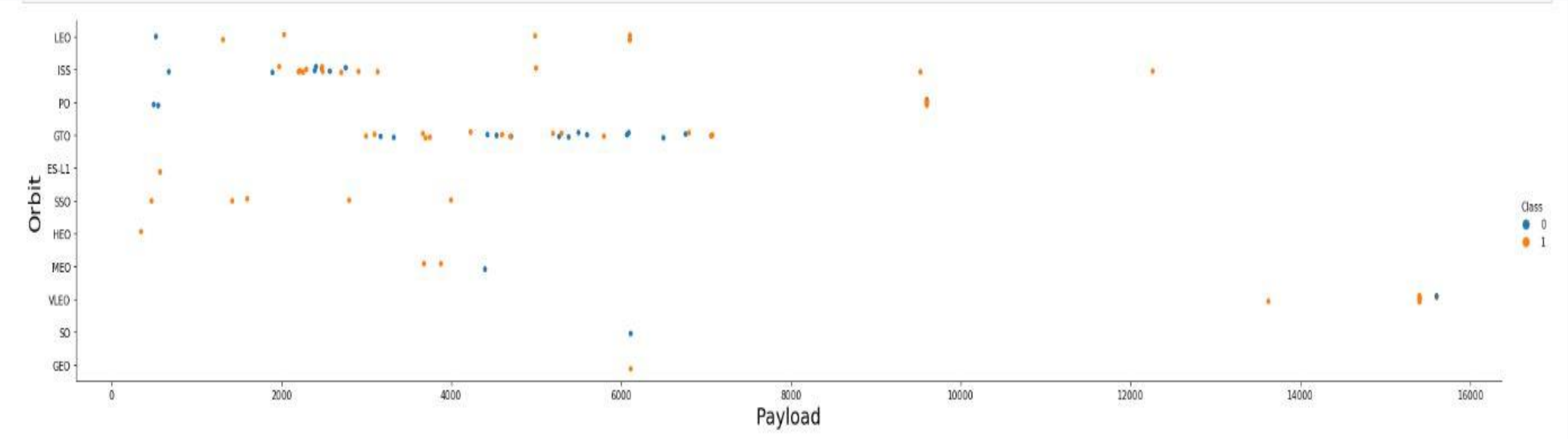
- The orbit of ES-L1, GEO, HEO and SSO have the highest success rates. The orbit SO has zero success rate

Flight Number vs. Orbit Type



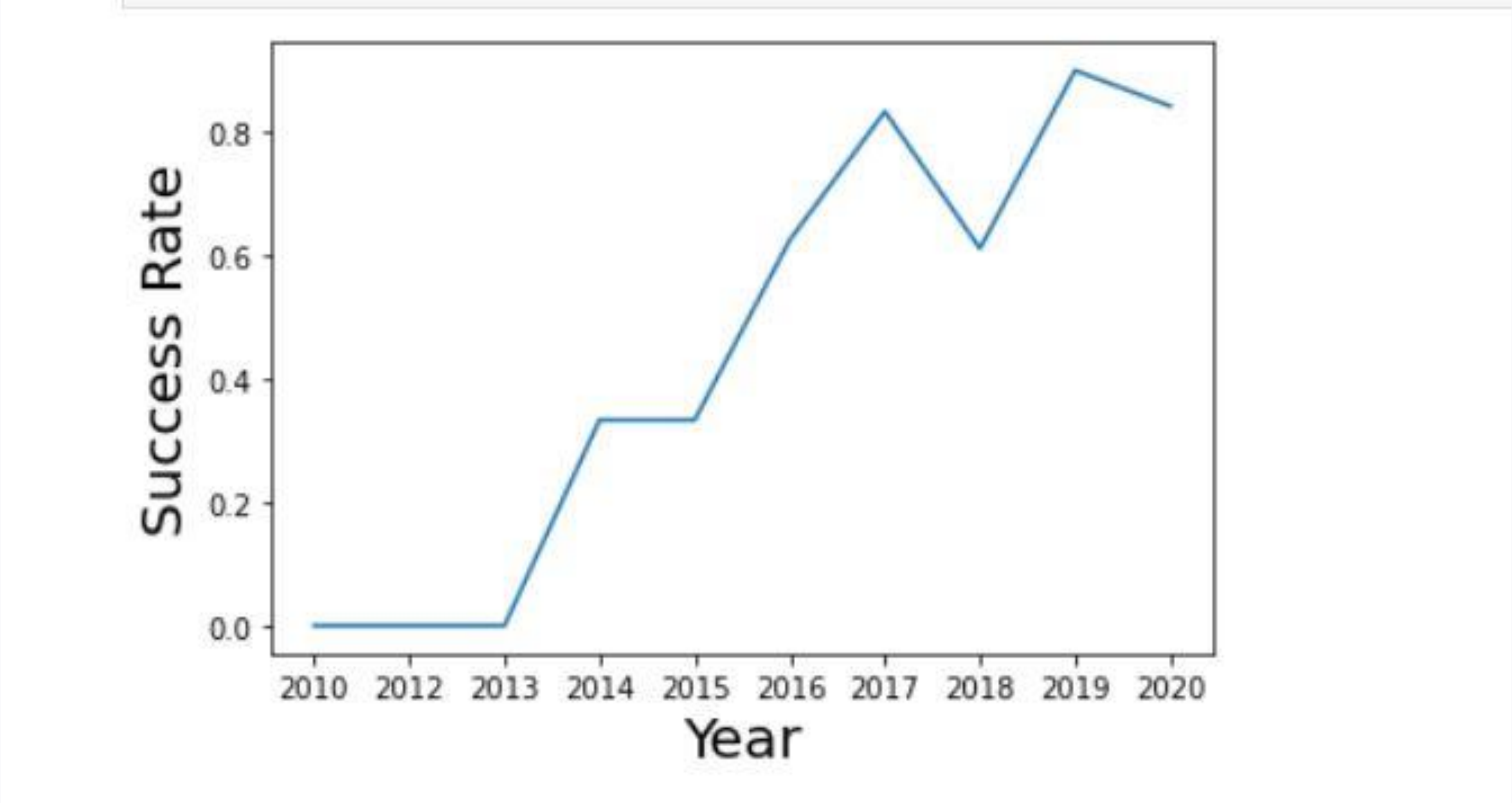
- In the LEO orbit the success rate appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



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

Launch Success Yearly Trend



- The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
5]: %sql select distinct launch_site from SPACEXTBL

 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b
Done.
```

5]:

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

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

```
[6]: %sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where customer = 'NASA (CRS)'

 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39
Done.
```

[6]:

1
45596

Average Payload Mass by F9 v1.1

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Versi
```

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n4
Done.
```

1
2928

First Successful Ground Landing Date

```
%sql select MIN(date) from SPACEXTBL where Landing__Outcome ='Success (ground pad)'
```

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0ngnrk39u98g.
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version, payload_mass__kg_ from SPACEXTBL where
Landing__Outcome ='Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.
Done.
```

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(*) as count from SPACEXTBL group by mission_outcome
```

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.dat
Done.
```

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286
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

2015 Launch Records

```
%sql select substr(Date, 6, 2) as month,booster_version, launch_site from SPACEXTBL
where Landing__Outcome = 'Failure (drone ship)' and substr(Date,1,4)='2015'
```

```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98
Done.
```

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

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

```
%sql select landing__outcome, count(*) as count from SPACEXTBL where date between
'2010-06-04' and '2017-03-20' group by landing__outcome order by count DESC
```

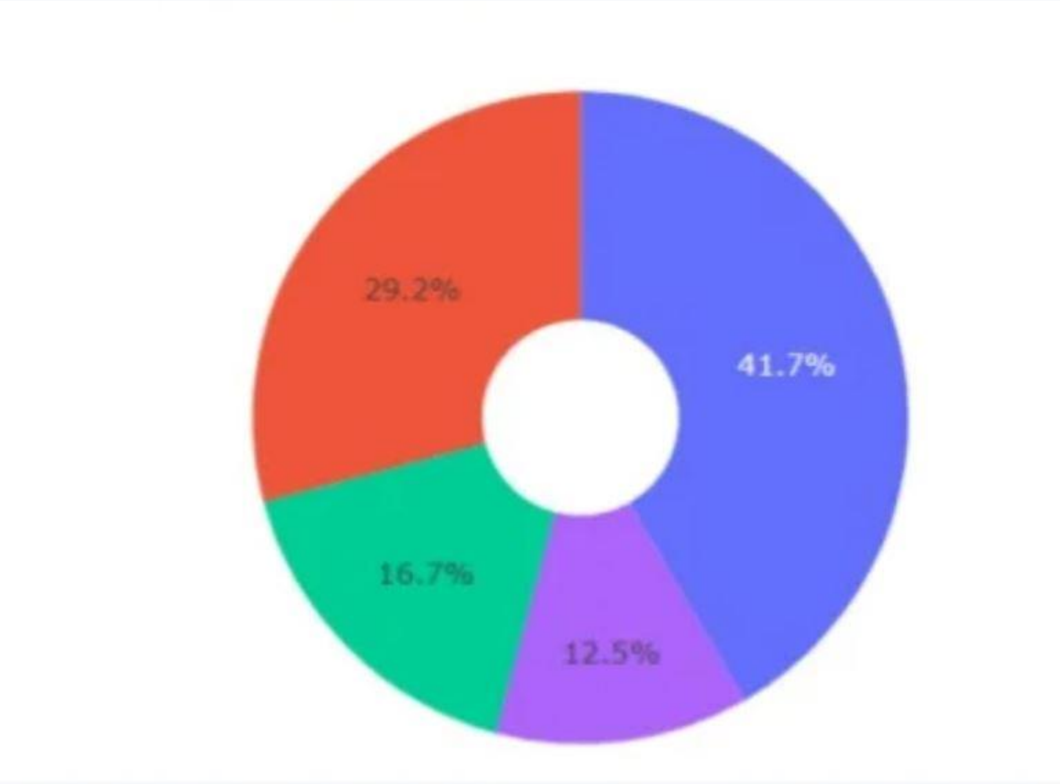
```
 * ibm_db_sa://dtq88624:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u
Done.
```

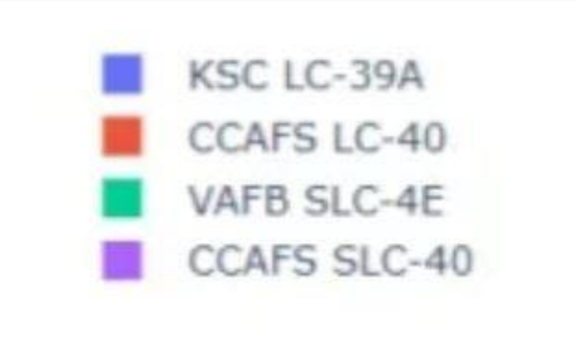
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Build a Dashboard with Plotly Dash

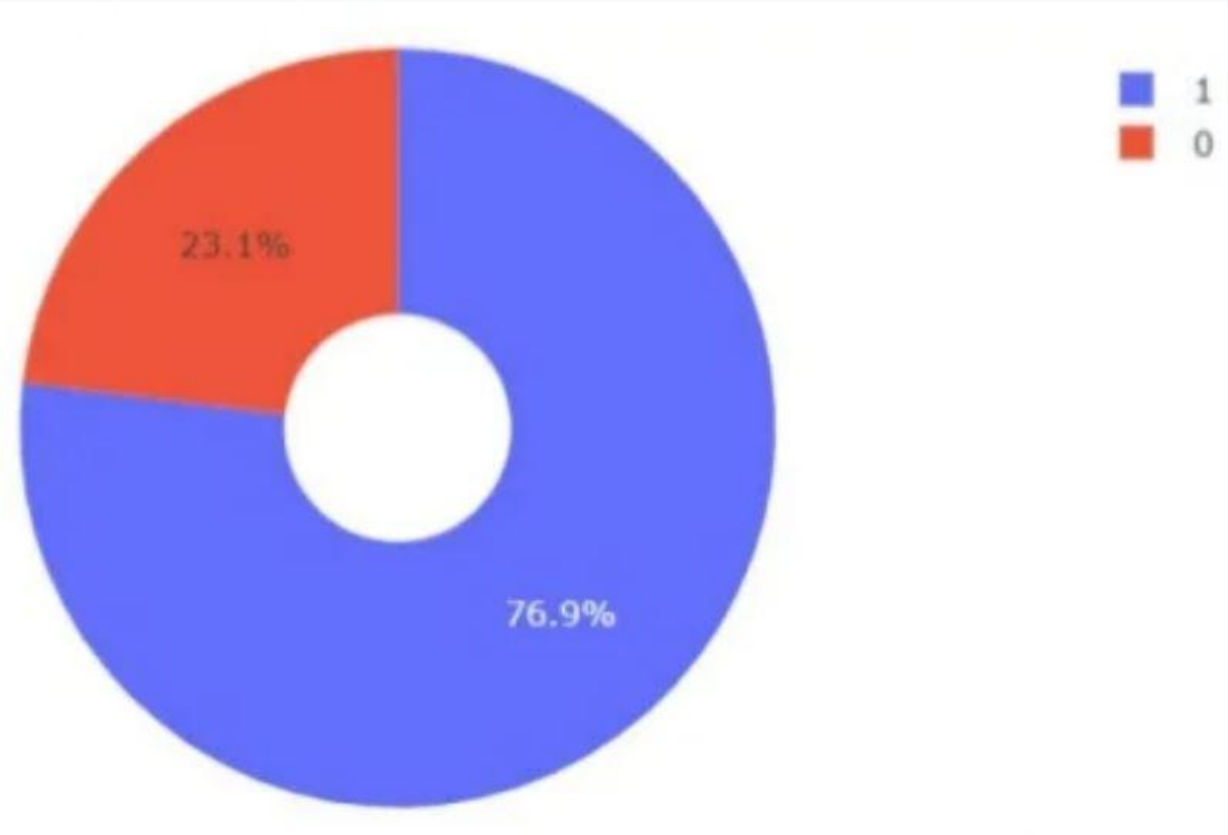
Total success launches for all sites





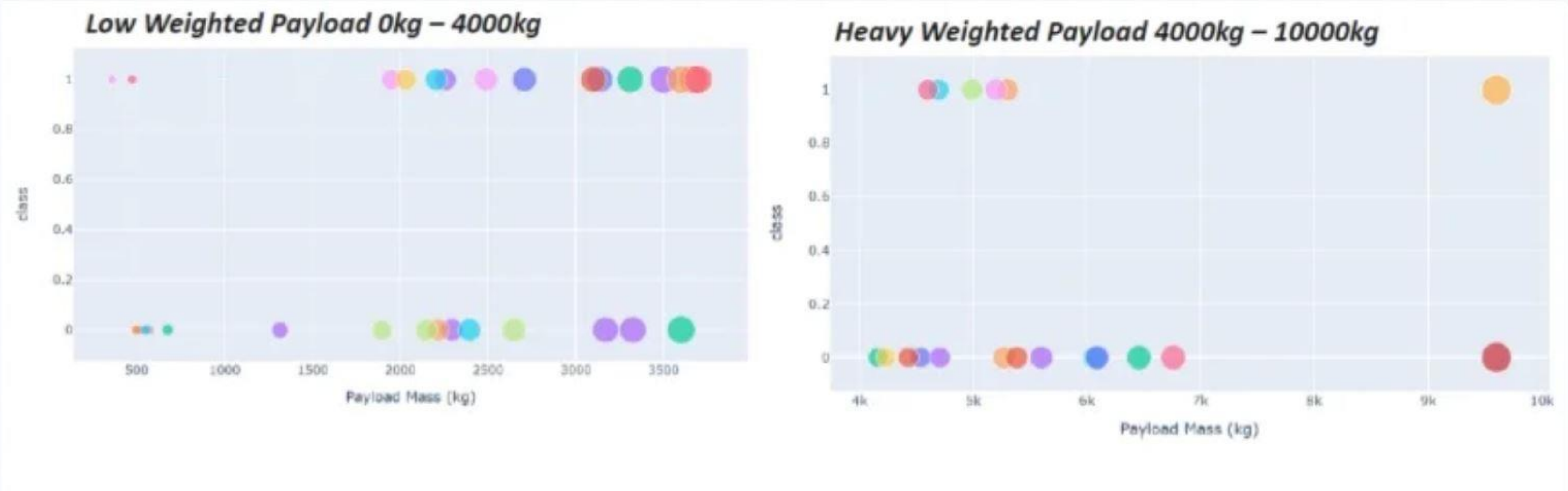
- Site KSC LC-39A has the most successful launches among all sites

Success rate by sites



- LSC LC-39A has a 76.9% success rate

Payload vs launch outcome

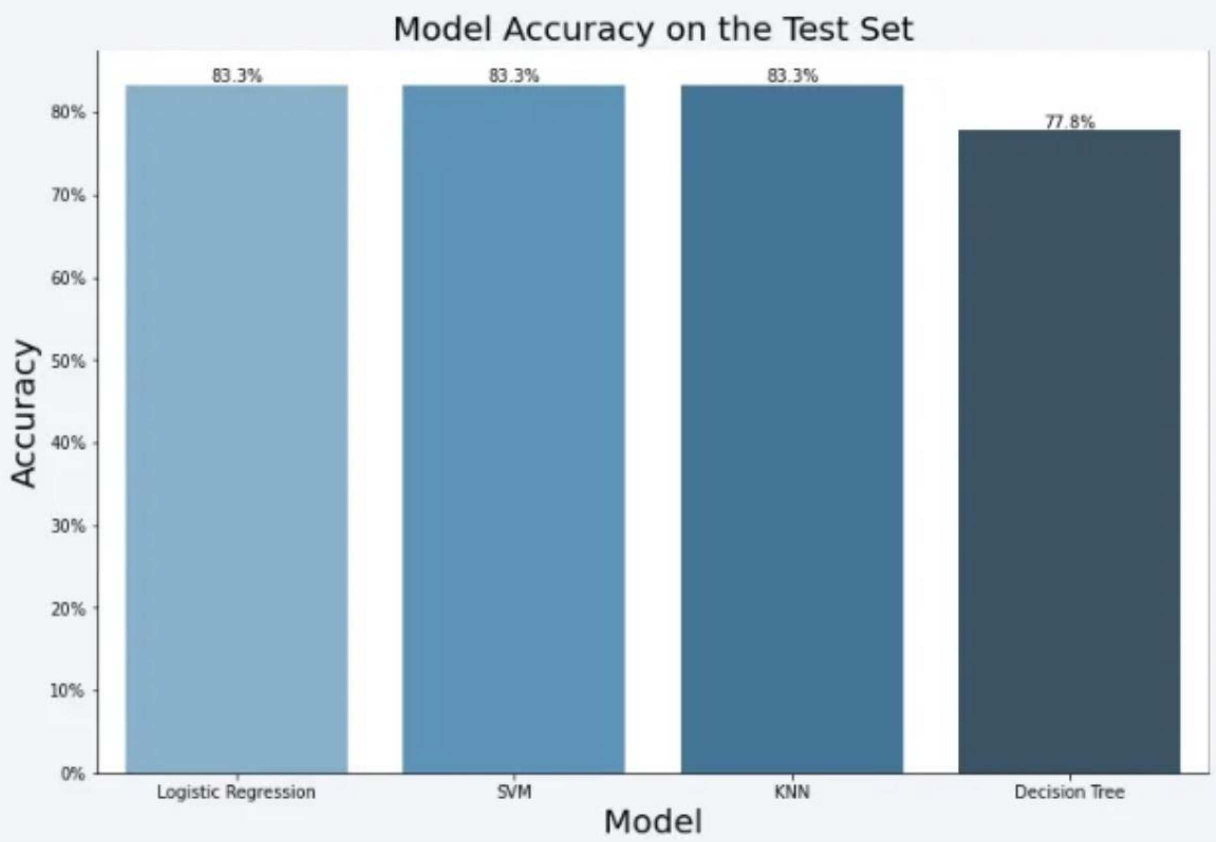


- Low weight payloads have higher success rates than heavy payloads

Section 5

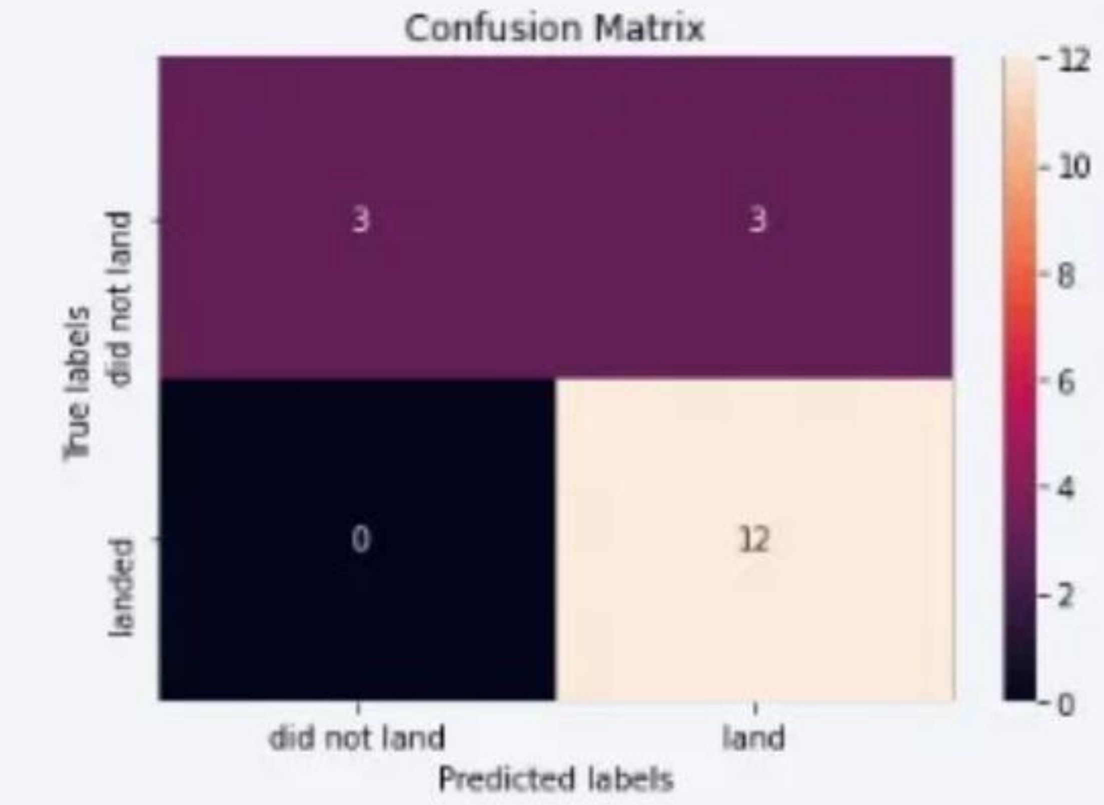
Predictive Analysis (Classification)

Classification Accuracy



- Logistic Regression, SVM and KNN have highest accuracy of 83.3%

Confusion Matrix



- Logistic Regression, SVM and KNN models have confusion matrix. The models labeled 15 out of 18 samples correctly with the accuracy 15/18=83.3%

Conclusions

- Launches with low payload mass perform better than that with high payload mass
- The success rates for SpaceX lunches are proportional to time in years since they are getting better over time
- Orbit GEO, SSO, HEO, ES L1 have the highest success rate
- KSC LC-39A has the most successful launches in all sites
- The Logistic Regression, SVM and KNN models have highest accuracy of 83.3%, in terms of prediction accuracy

Thank you!