IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

BISA Eric

March 2022

Outline

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

Executive Summary

- Data collection using SpaceX API and Web scraping from Wikipedia and Data wrangling
- Exploratory data analysis and Interactive visual
- Comparison of multiples classification algorithms results
- Predictive analysis using SVM Algorithm
- GEO, ES-L1, SOO and HEO orbits have the best success rates

Introduction

- Space X Falcon 9 rockets have the particularity to have a reusable first stage
- Reuse of the first stage allows to drastically reduce launches prices
- Space Y wants to determine the cost of a rocket launch
- The cost is proportional to the success landing rate of rockets' first stage
- The goal is to predict, using public information, the success of failure landing of a rocket launch

Section 1

Methodology

Methodology

Executive Summary

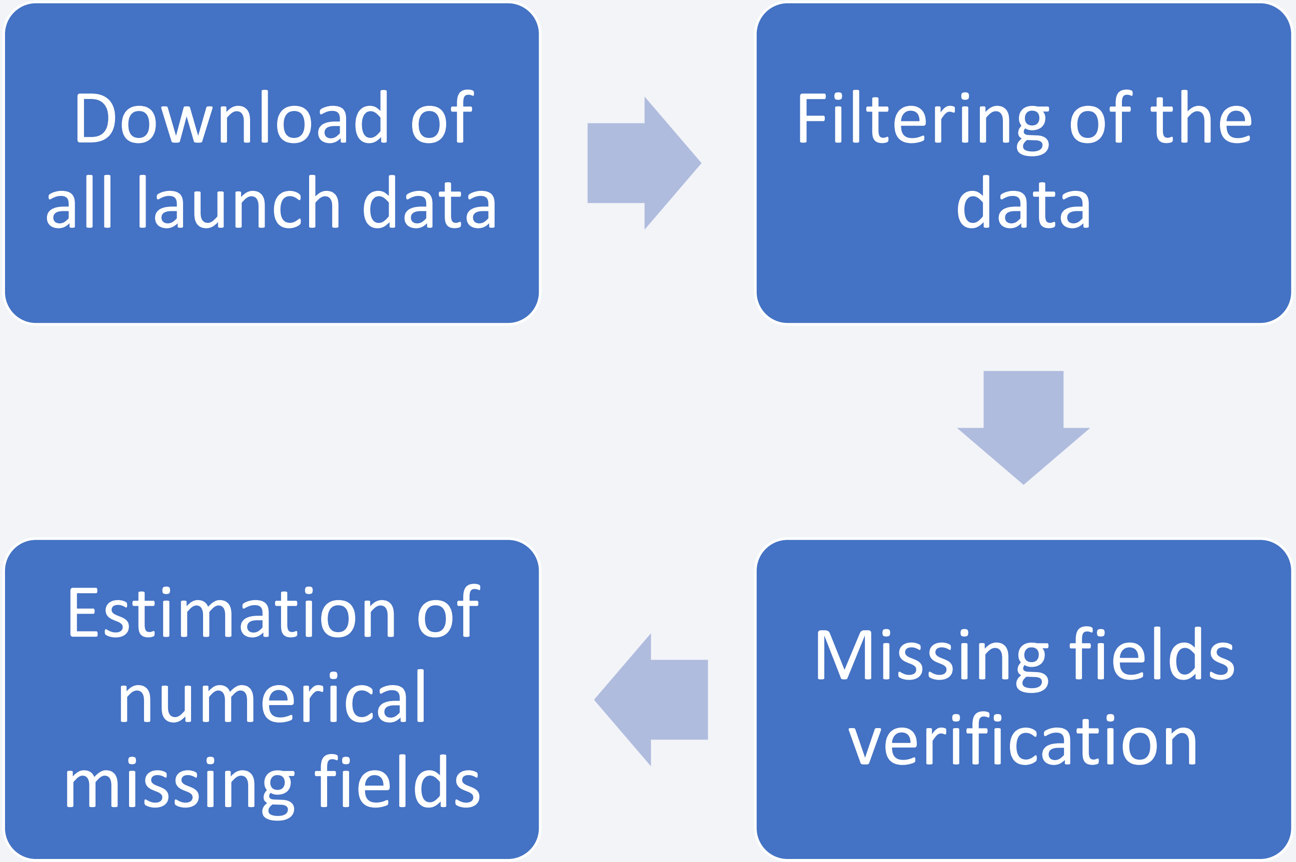
- Data collection methodology:
    - SpaceX API
    - Web scraping from Wikipedia
- Perform data wrangling
    - Manipulation of data using pandas library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Comparison of multiples classification algorithms results

Data Collection

- Data collected in two ways
  - From SpaceX API
  - From Falcon 9 launches Wikipedia page

Data Collection – SpaceX API

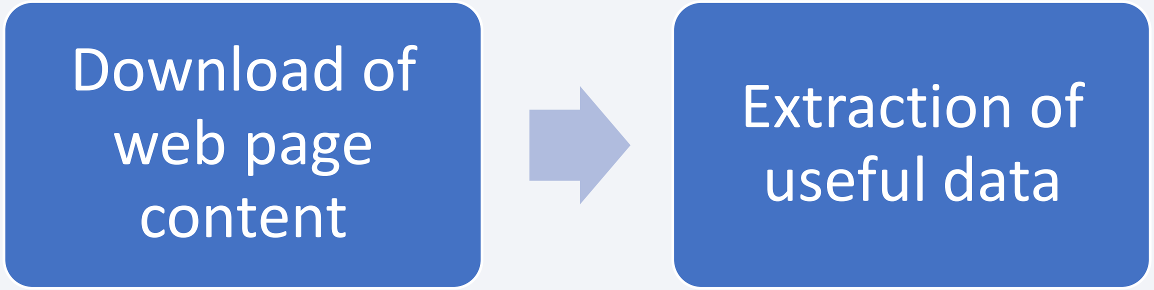
- Data imported from SpaceX API using requests library
- Only keeping the data about Falcon 9 launches
- Replacement of 5 missing payload mass by the mean payload mass



- Notebook on Github

Data Collection - Scraping

- Data imported from Falcon 9 launches Wikipedia page using requests library
- Use of BeautifulSoup library to parse through the data
- Conversion to python dataframe



- Notebook on GitHub

Data Wrangling

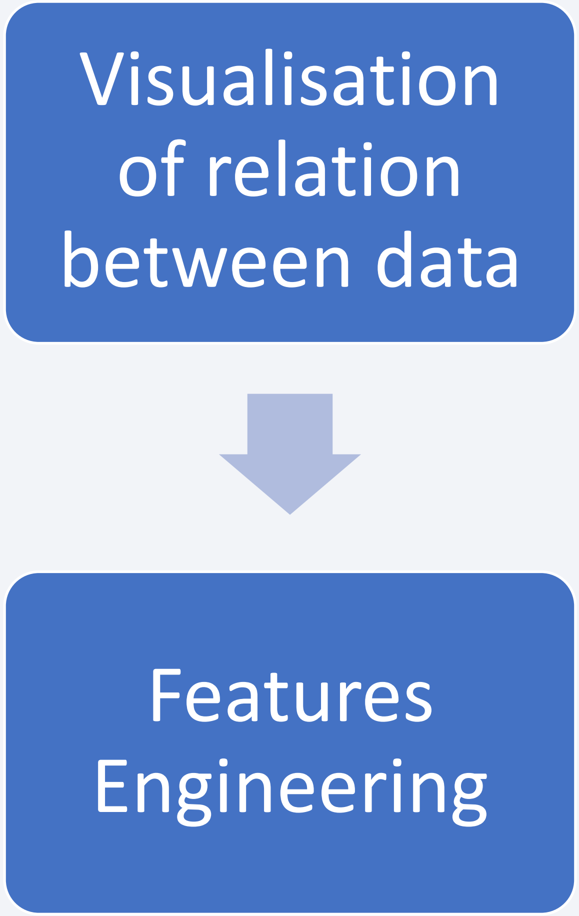
- Data were manipulated using pandas library
- Creation of a label column representing the binary label
- [Notebook on Github]()

Missing field verification → Data type analysis → Definition of data labels

EDA with Data Visualization

- Data visualized using matplotlib.pyplot and seaborn libraries
- Plot of relation between multiple duos of data and launch outcome
- Conversion of categorical data into numerical data One hot encoding

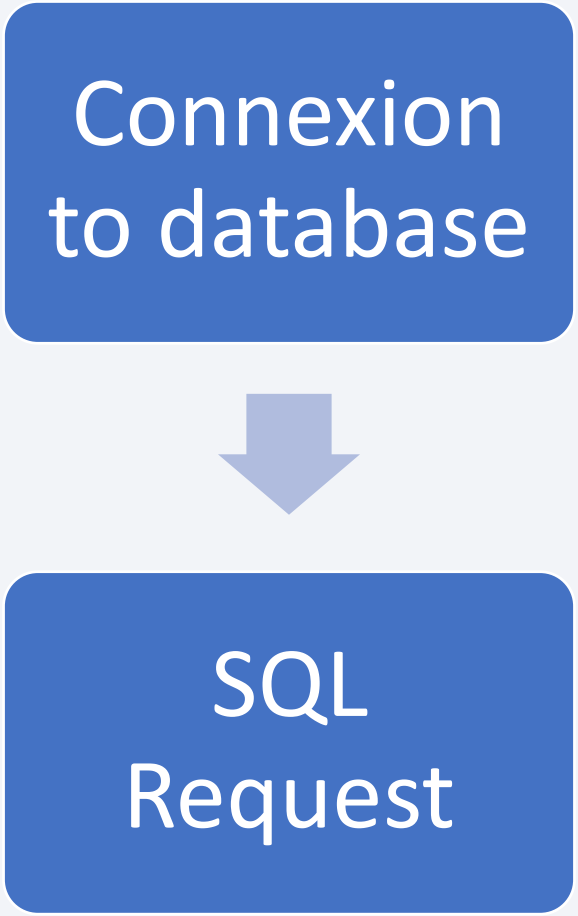
- [Notebook on Github]



EDA with SQL

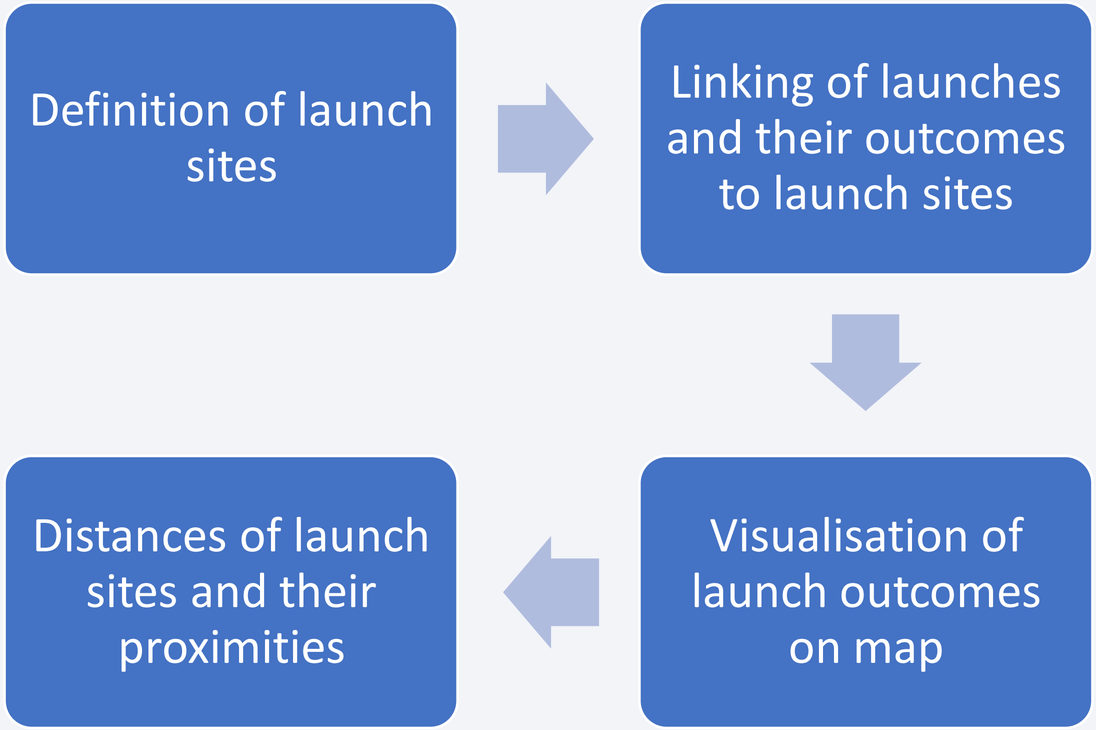
- Connexion and request to the dabased were made with in SQL with Jupyter Magics

- Notebook on Github



Build an Interactive Map with Folium

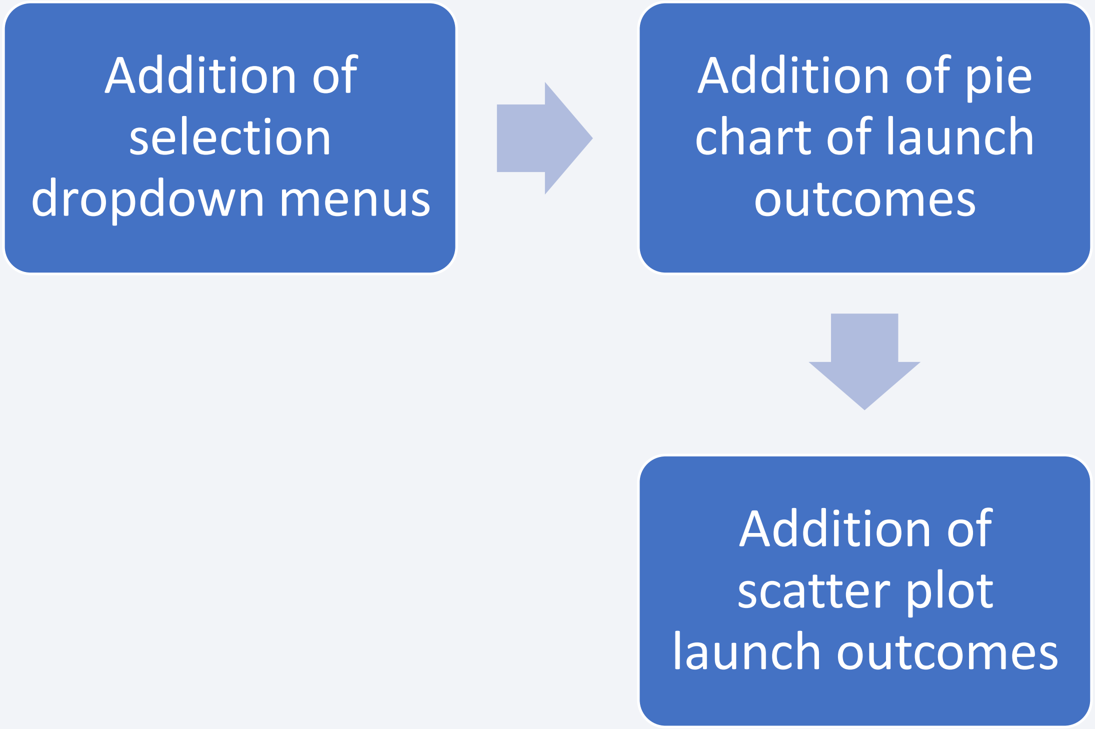
- Launch sites added to a Folium map using their geospatial coordinates
- Visual representation of launch outcomes on the map
- Computation of distance of launch sites and railway, highway, cities



- [Notebook on Github]

Build a Dashboard with Plotly Dash

- Addition of two dropdown menus to select :
  - Launch site
  - Launch Year
- Addition of a pie chart of launch outcomes
- Addition of a scatter plot of launch outcomes based on the payload mass
- [Notebook on Github]()



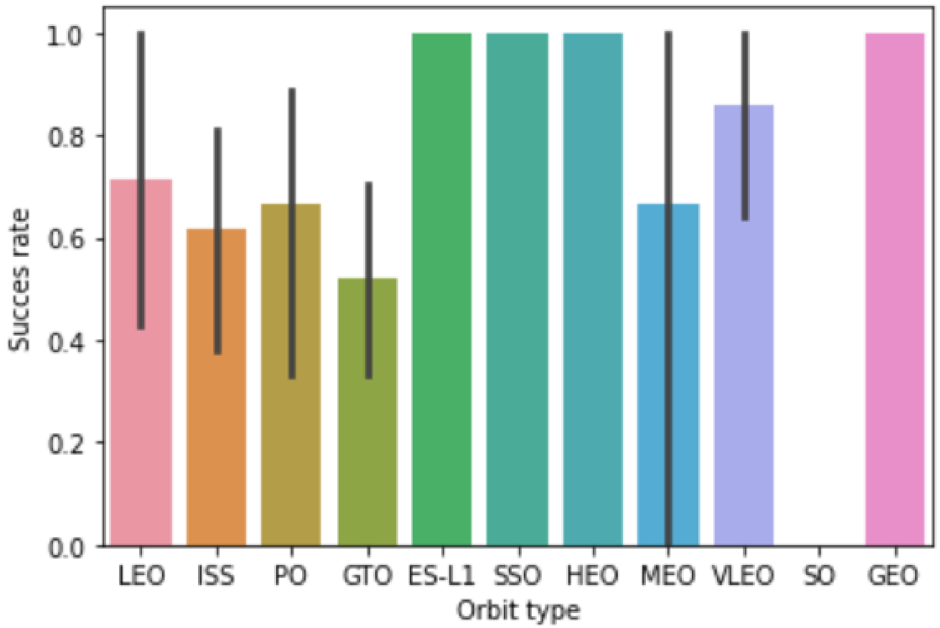
Predictive Analysis (Classification)

- The predictive analysis was made using sklearn libraries
- 80% of data were used for training and 20% for testing
- The classification algorithms tested were :
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - k-nearest neighbors
- Notebook on Github

- Data standardisation
- Separation of data into train and test sets
- Test of different classification algorithms
- Hyperparameter tuning with grid search
- Selection of the best ML model

Results

```
# HINT use groupby method on Orbit column and get
sns.barplot(x='Orbit', y='Class', data=df)
plt.xlabel("Orbit type")
plt.ylabel("Succes rate")
plt.show()
```

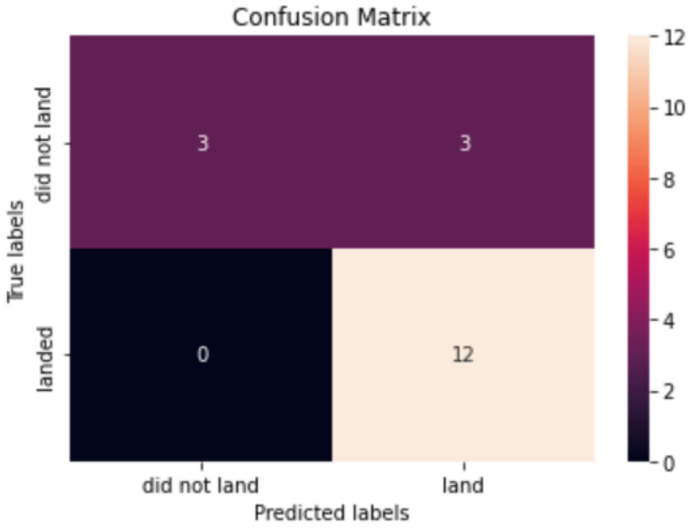


[13]:
```
Y_predict=logreg_cv.predict(X_test)
logreg_cv.score(X_test,Y_test)
```

[13]:

Lets look at the confusion matrix:

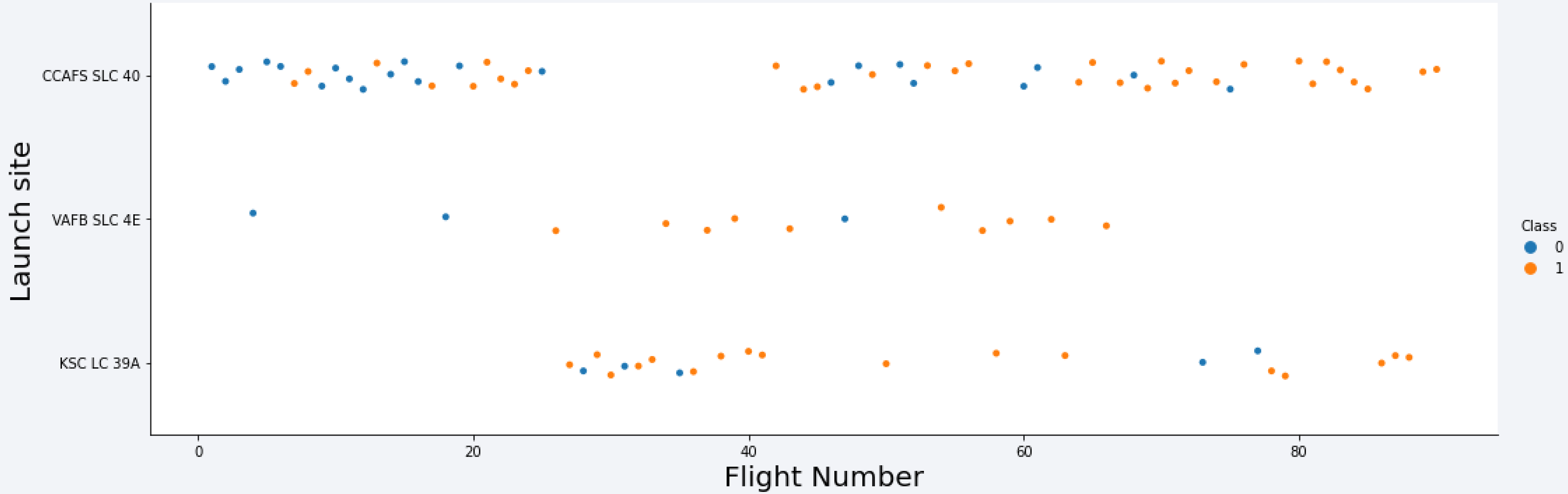
[14]:
```
yhat=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Section 2

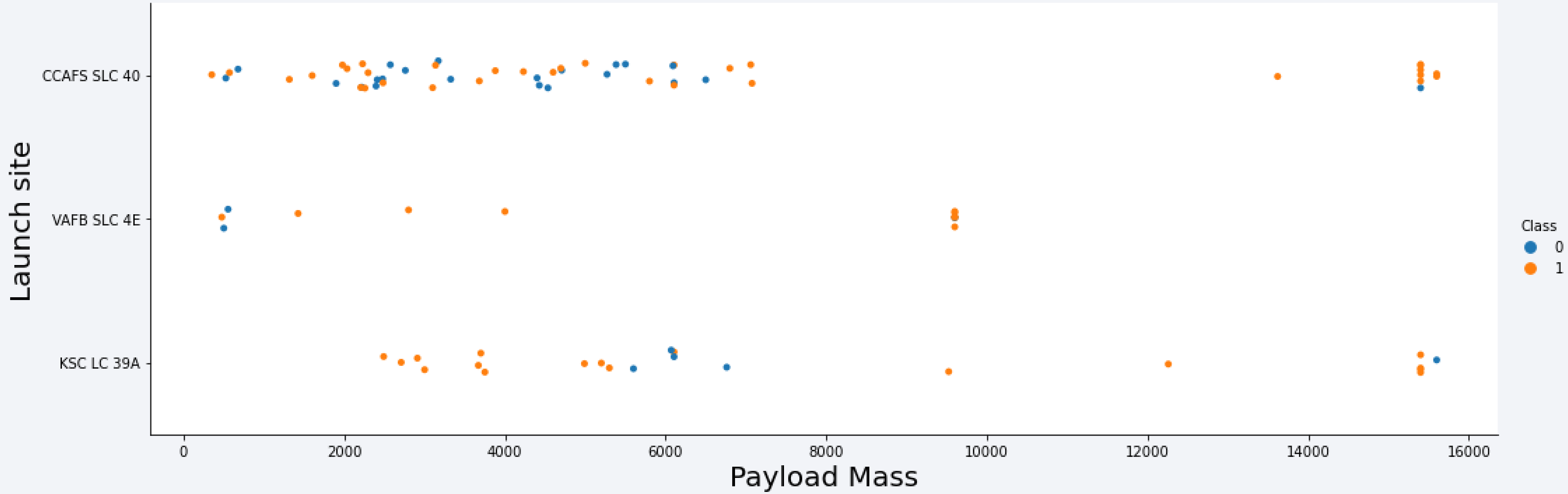
Insights drawn from EDA

Flight Number vs. Launch Site



- The failure rate is much more important in the Launch Site CCAFS LC-40 than in the other 2 sites

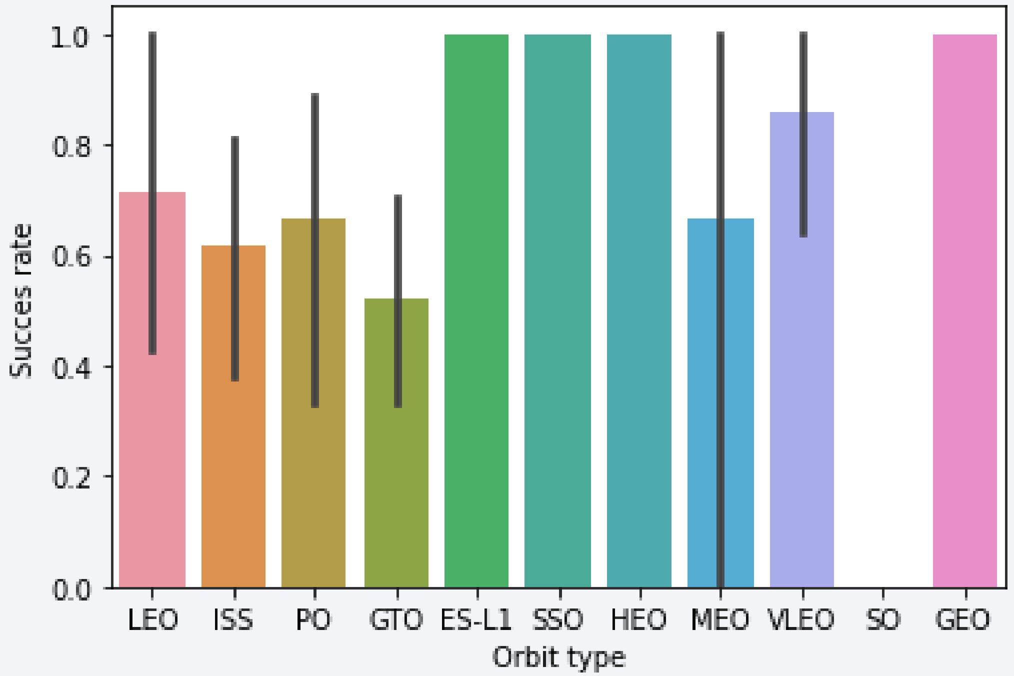
Payload vs. Launch Site



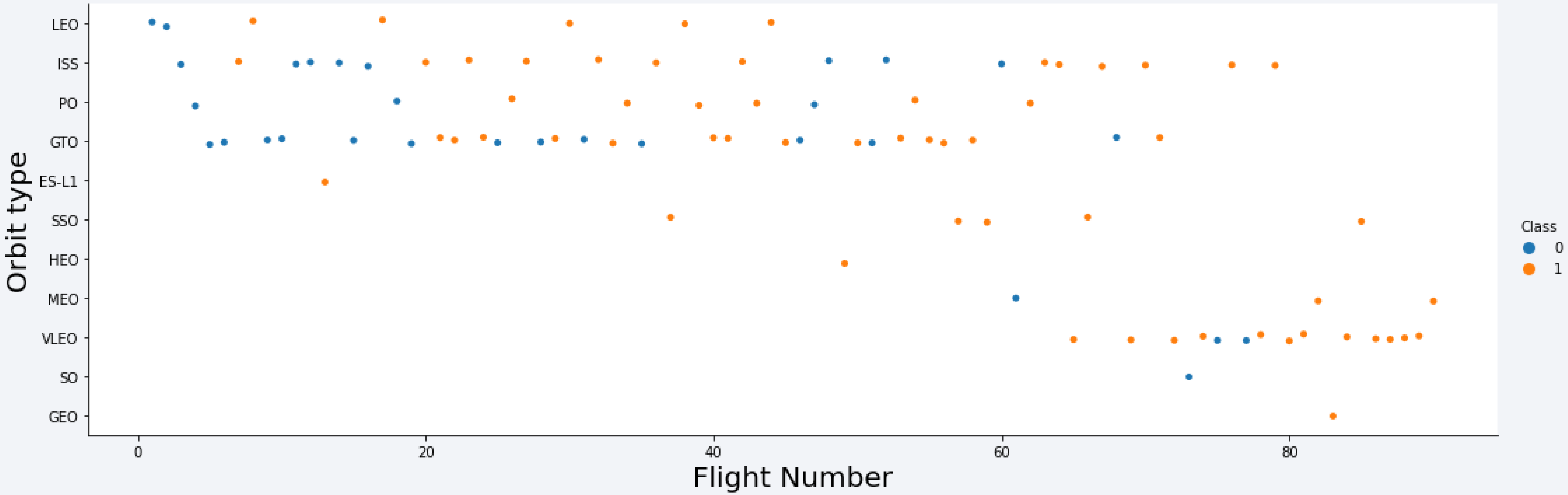
- It seems like there is no relationship between payload mass and launch site

Success Rate vs. Orbit Type

- The GEO, ES-L1, SOO and HEO orbits have the best success rates

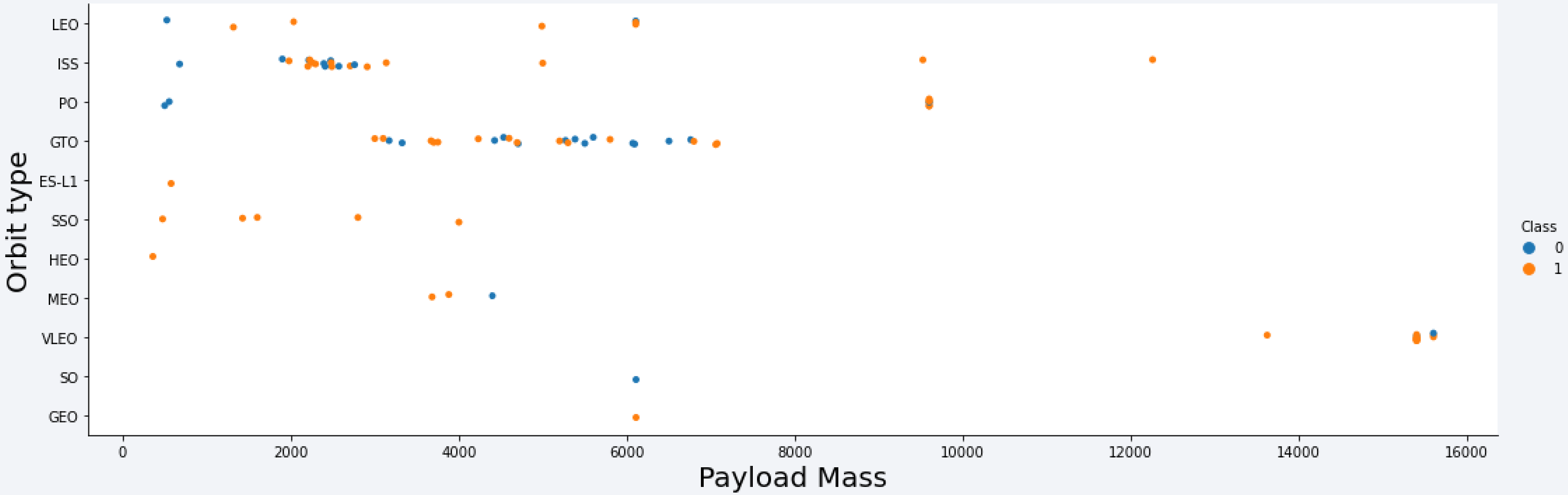


Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit

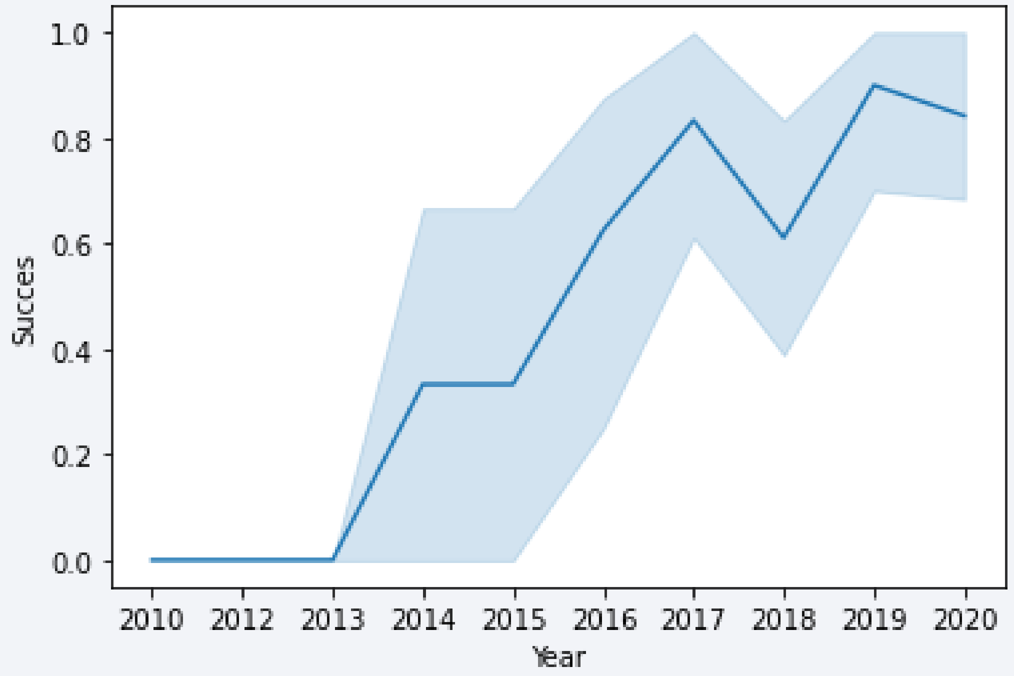
Payload vs. Orbit Type



- With heavy payloads the successful landing rate are more for Polar,LEO and ISS
- For GTO the landing outcomes does not seem to be correlated with the payload mass

Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- There are 4 different launch sites

Launch sites
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin 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

- 5 launches from launch sites with a name beginning with “CCA”

Total Payload Mass

- More than 45 tons of NASA (CRS) payload have left-off with Falcon 9 boosters

Total payload mass for NASA (CRS) [kg]
45596

Average Payload Mass by F9 v1.1

- Average payload mass is about 2.5 tons

Average payload mass by booster F9 v1.1
2534

First Successful Ground Landing Date

- The first successful landing occurred the 22/12/2015

Date of the first successful landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Those boosters are of type B4, B5, B5B and FT

booster_version
F9 B4 B1040.1
F9 B4 B1043.1
F9 B5 B1046.2
F9 B5 B1046.3
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5 B1058.2
F9 B5B1060.1
F9 B5B1062.1
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026
F9 FT B1032.1

Total Number of Successful and Failure Mission Outcomes

- Even though not all ground landing were successful, almost all the mission were

Mission outcome	Number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Those boosters have caried a 15,6 tons

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- There were only 2 failed landing in 2015
- Both failures happened with boosters V1.1

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- There were mainly non attempts of landing
- There was 5 drone ship successes and 3 ground successes
- Most of the landing outcomes were controlled crashes in the ocean and failures

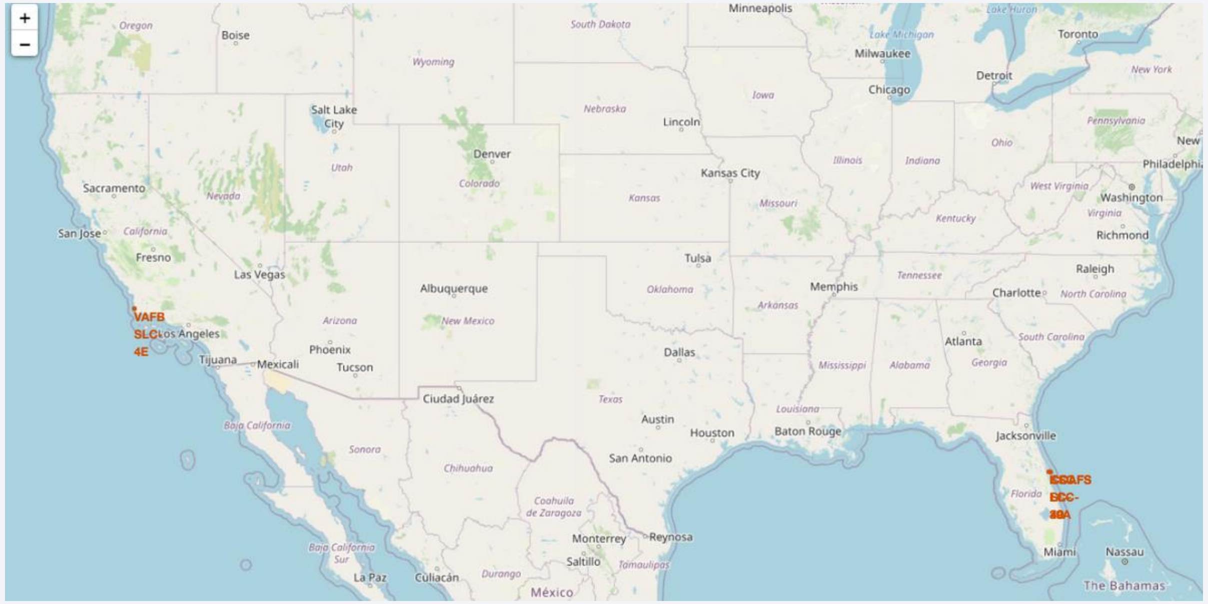
Landing outcome	**Number**
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

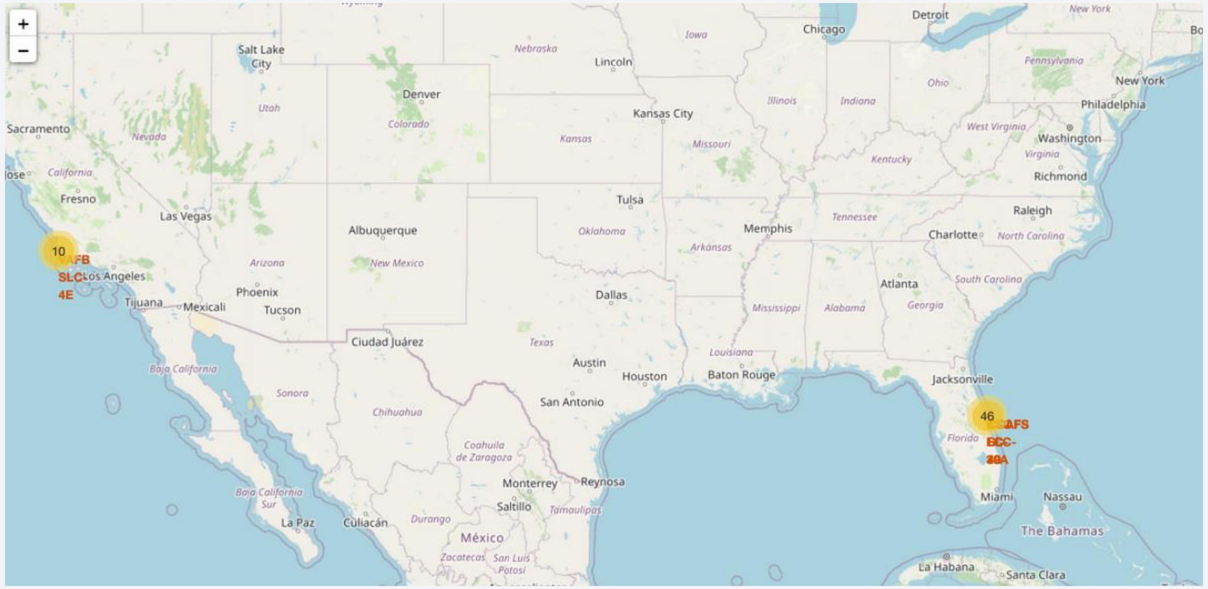
Map of launch sites

- Launch sites are in proximity of the equator line
- Launch sites are in very close proximity to the coast



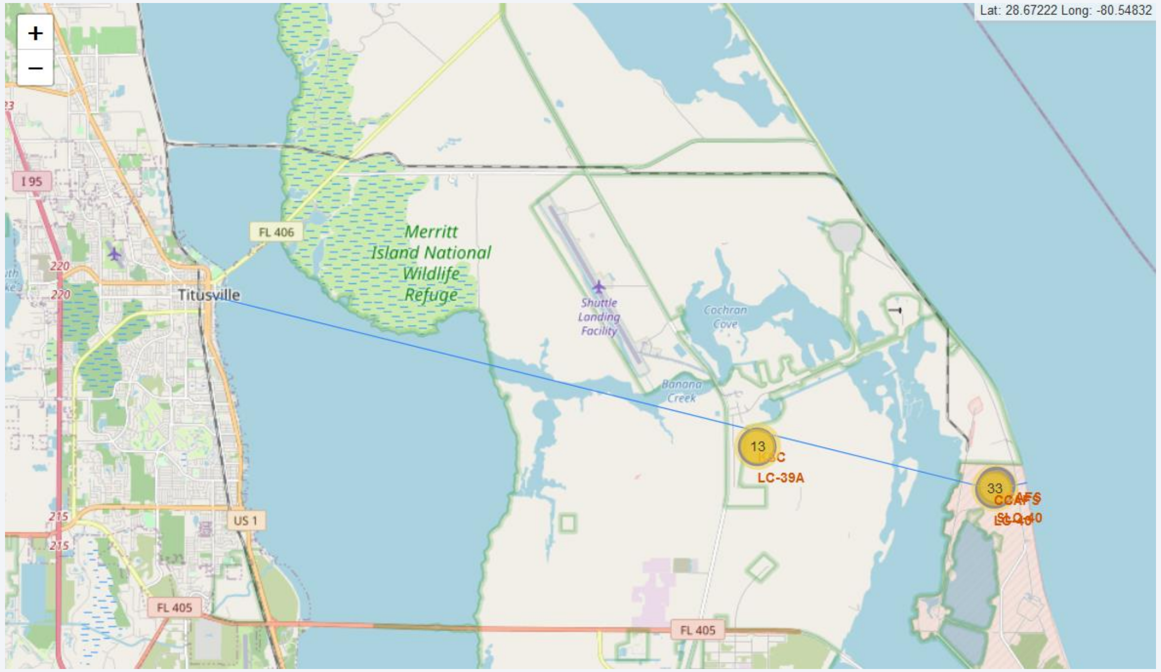
Map of success/failed landing

- Legend:
  - Green: Successful landing
  - Red: Failed landing
- Landing sites in yellow (green + red) from far



Launch sites and their proximities

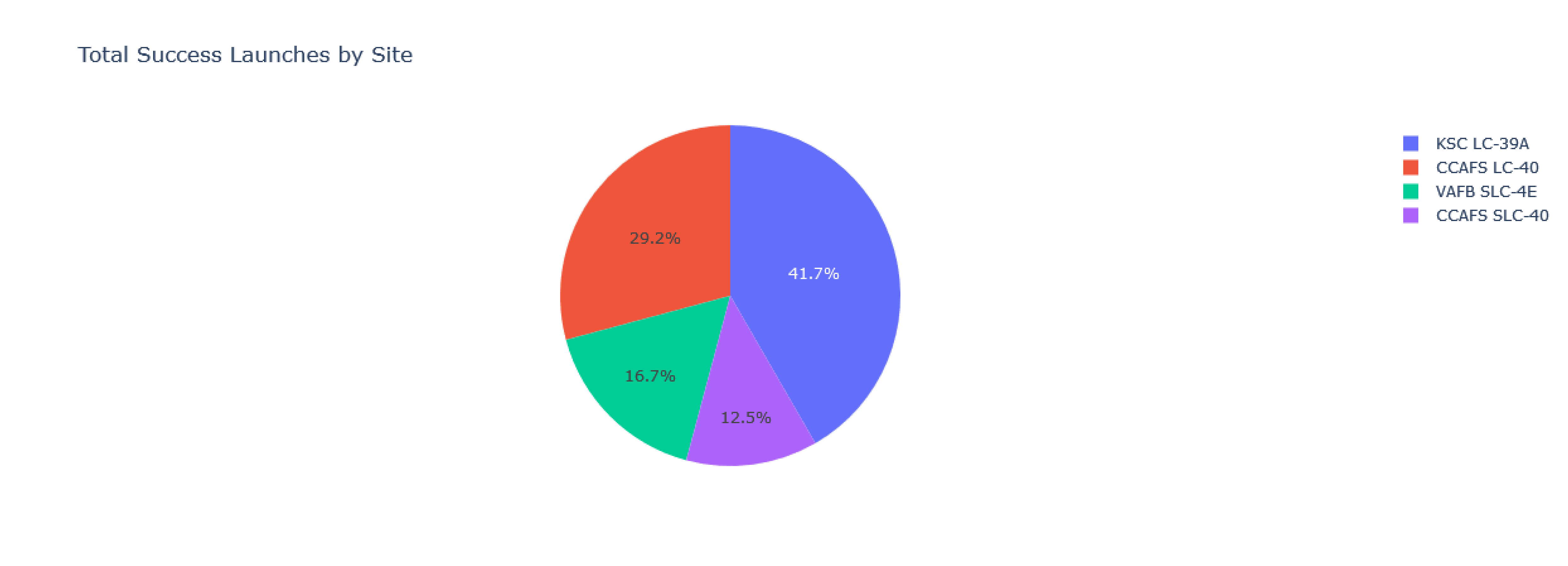
- Launch sites are close to coastlines, highways and railways
- Launch sites are far from cities



Section 4

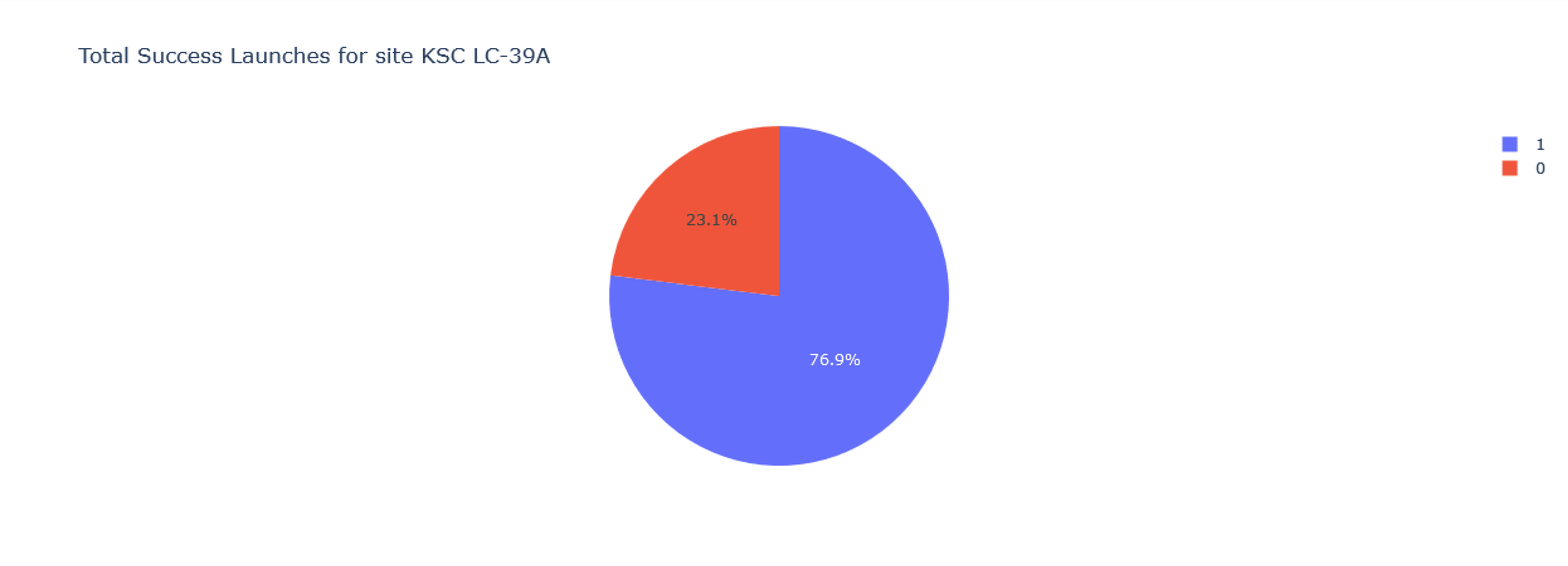
Build a Dashboard with Plotly Dash

Success launch by sites



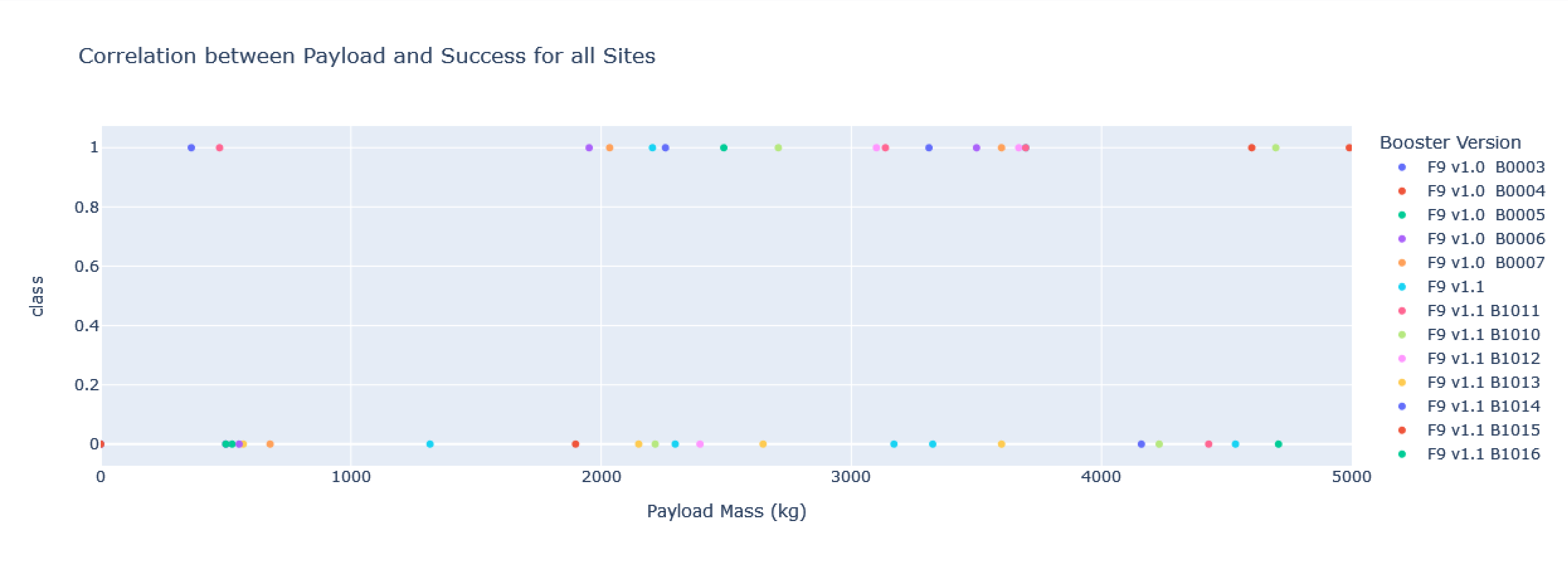
Most of the successful launches happened on the site KSC LC-39A

KSC LC-39A launch success rate



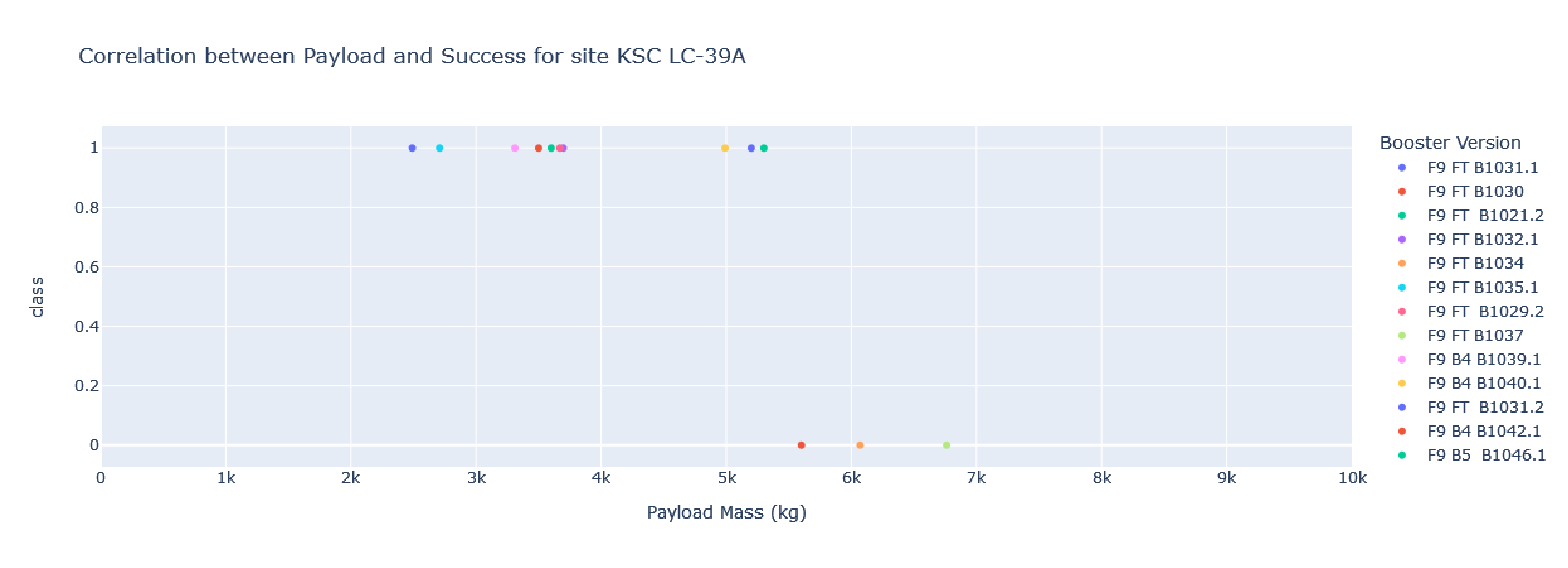
KSC LC-39A has the best launch success rate of all sites

Failures for low payload flights



For low payloads, failures mainly happened with V1.1 boosters

Payload VS Success launch for site KSC LC-39A



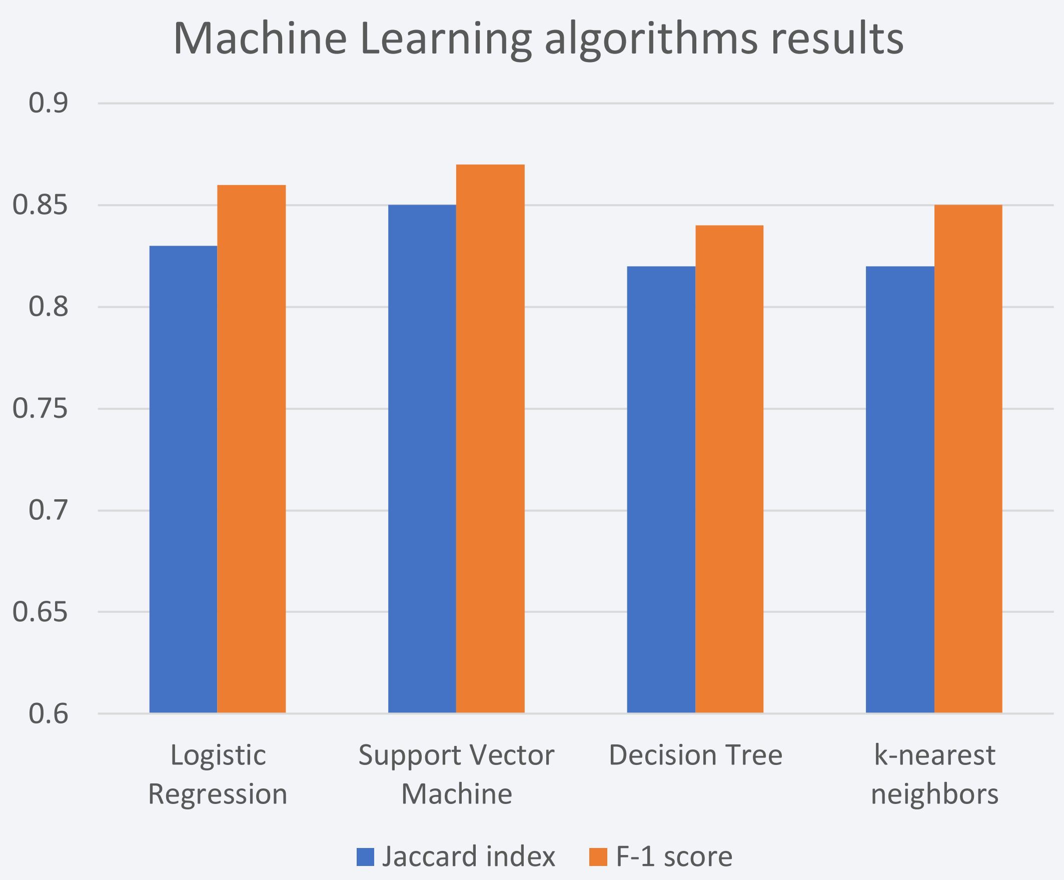
Launches from KSC LC-39A used only B4, B5 and FT boosters

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- SVM is the best prediction algorithm



SVM Confusion Matrix

- SVM can distinguish between the different classes
- The biggest problem with the model is false positives

Conclusions

- FT, B4 and B5 booster have the best landing success rate
- KSC LC-39A launch site has the best success rate
- The GEO, ES-L1, SOO and HEO orbits have the best success rates
- Future launches success can be predicted using an SVM algorithm

Appendix

Logistical regression confusion matrix

Decision tree confusion matrix

Thank you!