IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

https://github.com/MoSaadcoe/IBM-Applied-Data-Science-

Outline

Executive Summary

In this project I will play the role of a Data Scientist working for a startup intending to compete with SpaceX.

Summary of methodologies:

- Data collection.
- Data wrangling.
- Exploratory data analysis.
- Data visualization.
- Model development.
- Model evaluation..
- Reporting your results to stakeholders.

Summary of all results:

- Exploratory data analysis results.
- Interactive analytics demo in screenshots.
- Predictive analysis result.

Introduction

Project background and context:

This is the final course in the IBM Data Science Professional Certificate as well as the Applied Data Science with Python Specialization. This capstone project course will give you the chance to practice the work that data scientists do in real life when working with datasets. I played the role of role of a Data Scientist working for a startup intending to compete with SpaceX, and in the process follow the Data Science methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and reporting your results to stakeholders.

Finally This project aims to provide stakeholders with Interactive Dashboards and predictive ML models to determine if the first stage of Falcon 9 will land successfully.

Section 1

Methodology

https://github.com/MoSaadcoe/IBM-Applied-Data-Science-

Methodology

Data collection methodology:

- Data used in this project were collected from SpaceX Rest API and from Wikipedia launch table.
- Perform data wrangling
- The wrangling of the collected data included cleaning, preparation for visualization and information extraction
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- SpaceX API
- Scraping

Data Collection – SpaceX API

Request and parse the SpaceX launch data using the GET request

→

Decoding the response using content using and turning it into a data frame and map some rows to readable and meaningful data

→

Filter the data frame to only include `Falcon 9` launches

Output sample:

1	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
2	1	2010-06-04	Falcon 9	6123.547647058824	LEO	CCSFS SLC 40	None None	1	False	False	False
3	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False
4	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False
5	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False
6	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False

Github url: spacex-data-collection-api

Data Collection - Scraping

Perform HTTP get to request Falcon 9 HTML page and
create Beautiful Soup object from
HTML

Create a data frame by parsing the launch HTML tables

Output sample:

1	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster l
2	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	F9 v1.07B0003.18	Failure
3	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.07B0004.18	Failure
4	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.07B0005.18	No atten
5	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success	F9 v1.07B0006.18	No atten
6	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success	F9 v1.07B0007.18	No atten

Github url: webscraping.ipynb

Data Wrangling

Now, the data is collected, but before the visualization and the training it needs to be clean and prepared using NumPy and Pandas as follow:

Keep falcon 9 data only in the dataframe

Dealing with Missing Values by replace the missing payload mass by the mean

Create a landing outcome label from Outcome column. Class 1 for every type of successful landing and 0 otherwise

Github url: [wrangling.ipynb]

EDA with Data Visualization

To understand the data and how different values may effect each other. Multiple figures was made such as scatter plot between multiple columns, bat charts, line charts

explor-with-seeborn.ipynb

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 successful 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.

sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers of all Launch Sites
- Coloured Markers of the launch outcomes for each Launch Site(green for success and red for failed launches)
- Display the distances with lines between a Launch Site to its proximities like Railway, Highway, Coastline and Closest City.

launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Dashboard that contains:

- Launch Sites Dropdown List to allow the user to select.
- Pie Chart showing Success Launches (All Sites/Certain Site).
- Slider of Payload Mass Range to select the range of payloads that the user wont to see the Scatter Chart.
- Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions.

spacex_dash_app.py

Predictive Analysis (Classification)

Create a NumPy array from the column ”class”

Standardize the data

Split the data into training and testing data

Split the data into training and testing data

Train the data using multiple models(will be discussed in detail later)

Compare the results

Prediction_Part_5.ipynb

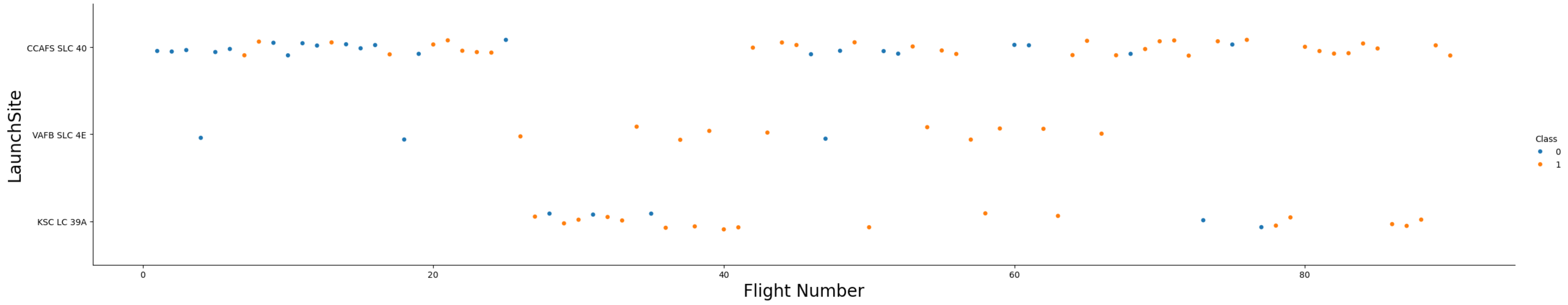
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

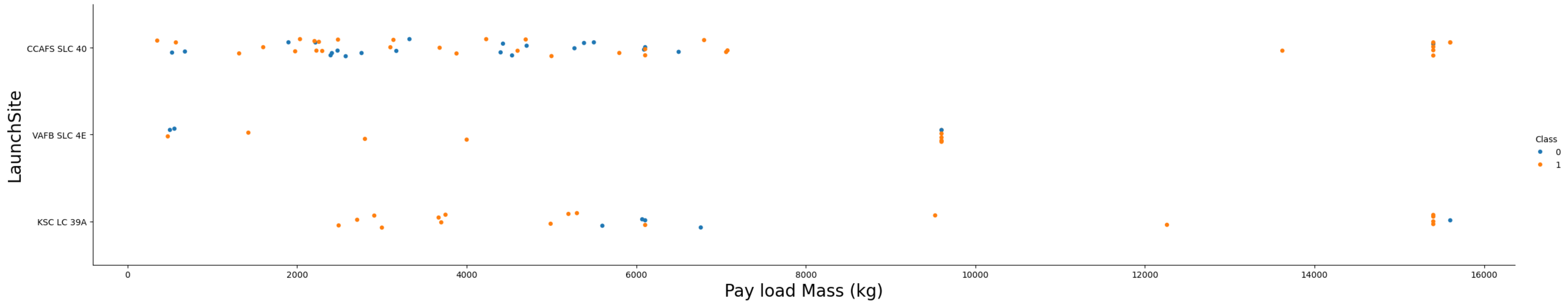
Insights drawn from EDA

Flight Number vs. Launch Site



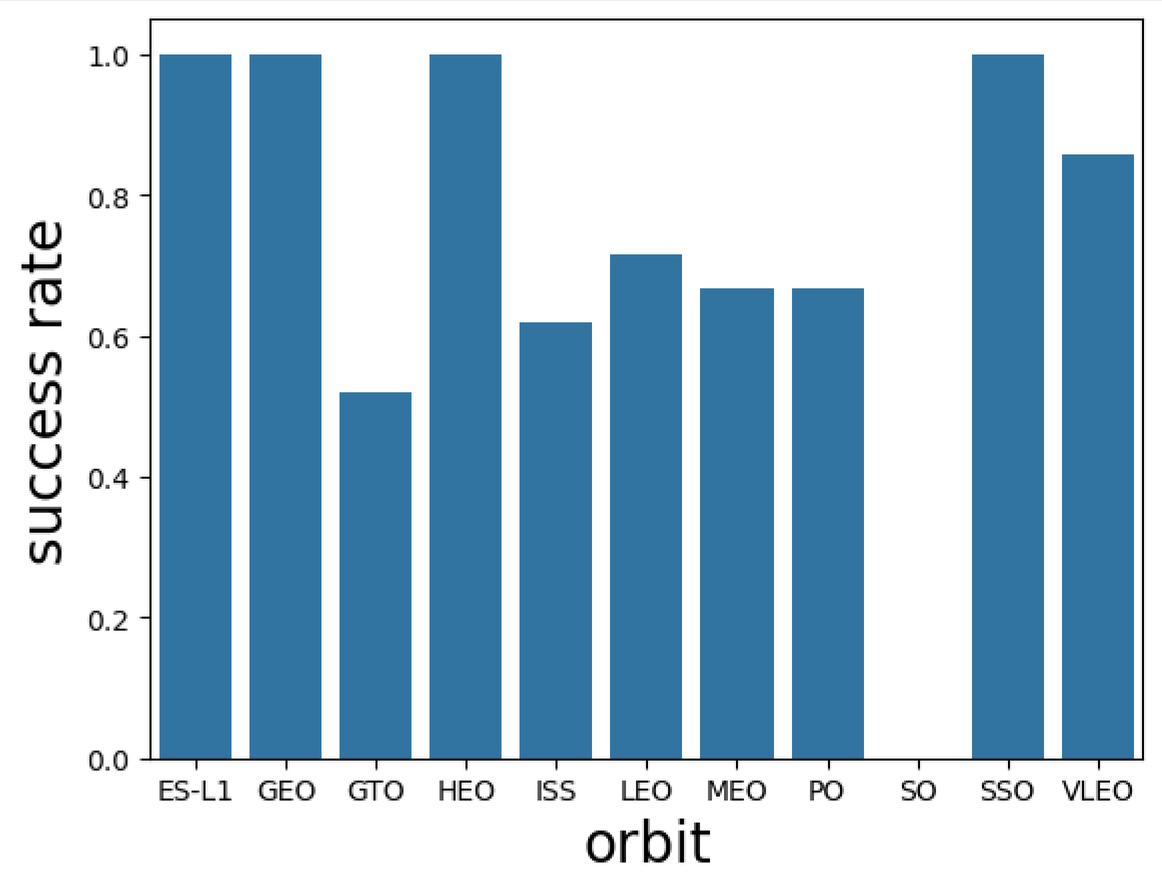
The figures shows that ion general the successful Lunches rate increase as the number of flight increase in all sites.

Payload vs. Launch Site



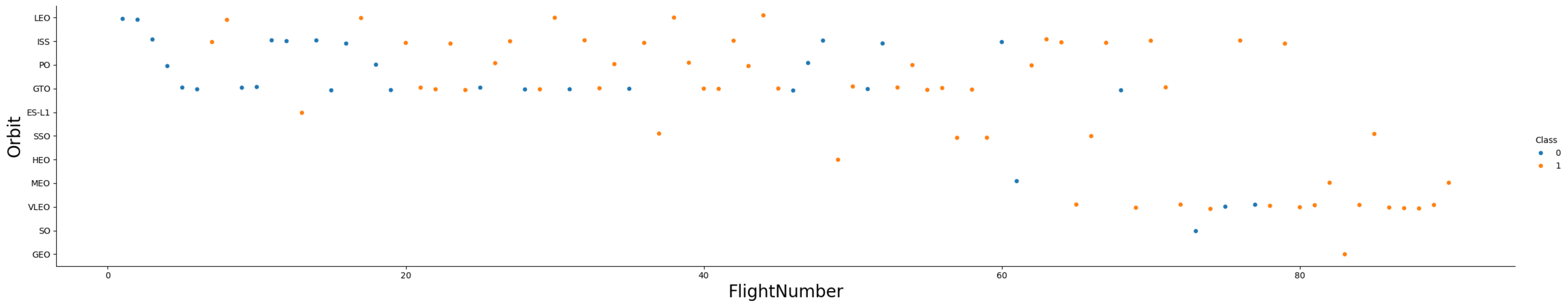
- All lunches from KSC LC 39A site with payload less than 5500Kg are successful
- For lunches from CCAFS SLC 40 site, lunches with payload around 16K kg are successful with high rate.

Success Rate vs. Orbit Type



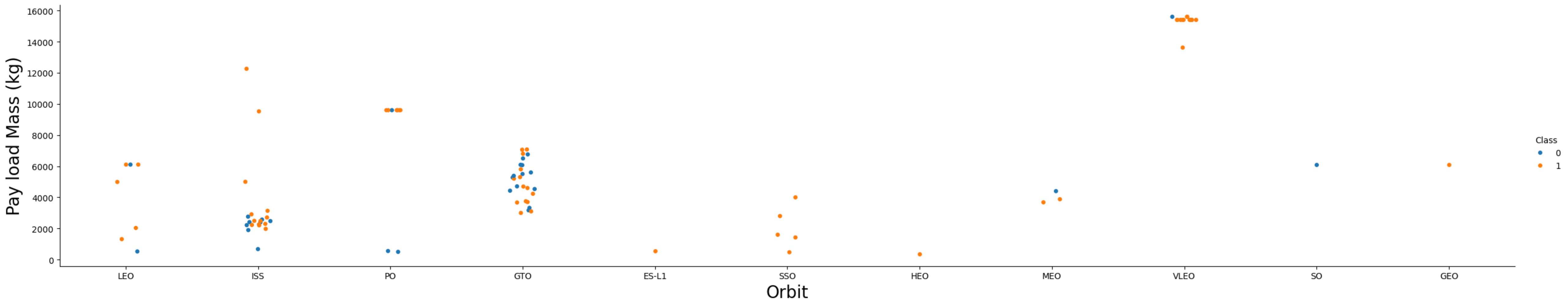
- Lunches for ES-L1 , GEO, HEO, and SSO are allows successful.
- GTO has the lowest successful rate.

Flight Number vs. Orbit Type



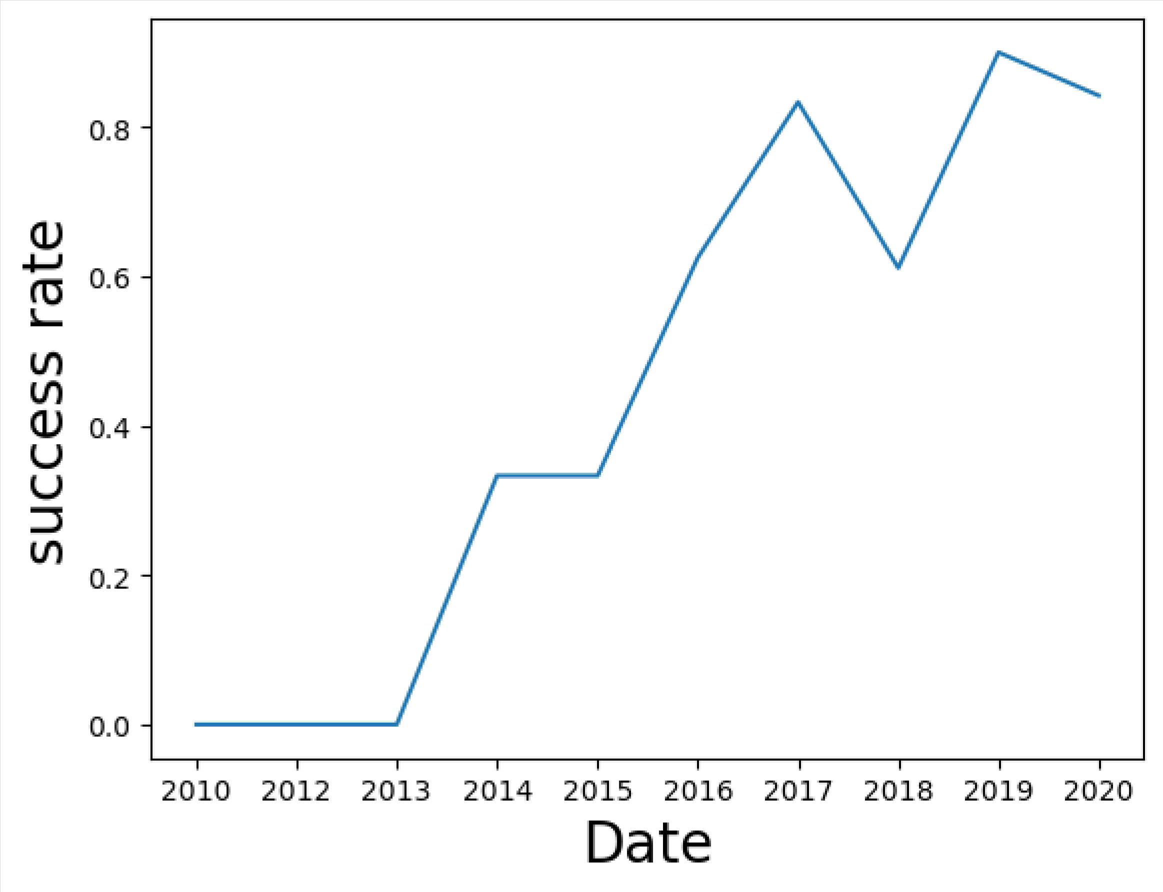
Lunches 0-20 have low successful rate as expected. And all lunches higher than 80 were successful.

Payload vs. Orbit Type



Lunches for ISS orbit with payload higher than 4k kg are always successful.

Launch Success Yearly Trend



In general, with the time successful rate increases, but there is a drop at 2018 need to be searched about.

All Launch Site Names

The querie:
%sql select distinct Launch_Site from SPACEXTBL ;

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

Launch Site Names Begin with 'CCA'

The querei:
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit(5);
(like 'CCA%') search for Lunch sit values start 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	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

Total Payload Mass

%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL
where Customer == "NASA (CRS)";
(SUM()) function display the sumof all rows in the givin column.

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like "F9 v1.1%";
```

```
 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

```
%sql select min(Date)  from SPACEXTBL where  Landing_Outcome == 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct Booster_Version from
SPACEXTBL where Landing_Outcome ==
'Success (drone ship)' and 4000 <
PAYLOAD_MASS__KG_ < 6000;
```

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome , count(Mission_Outcome) from SPACEXTBL group by (Mission_Outcome);
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	count(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%sql select Booster_Version , PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_ == (select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

%sql SELECT booster_version,launch_site , substr(Date, 6,2) , Landing_Outcome FROM SPACEXTBL WHERE substr(Date, 0, 5) = '2015' and Landing_Outcome == 'Failure (drone ship)' ;

Booster_Version	Launch_Site	substr(Date, 6,2)	Landing_Outcome
F9 v1.1 B1012	CCAFS LC-40	01	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	04	Failure (drone ship)

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

```
%sql select Landing_Outcome , count(Landing_Outcome) from SPACEXTBL group by Landing_Outcome order by count(Landing_Outcome) desc ;
```

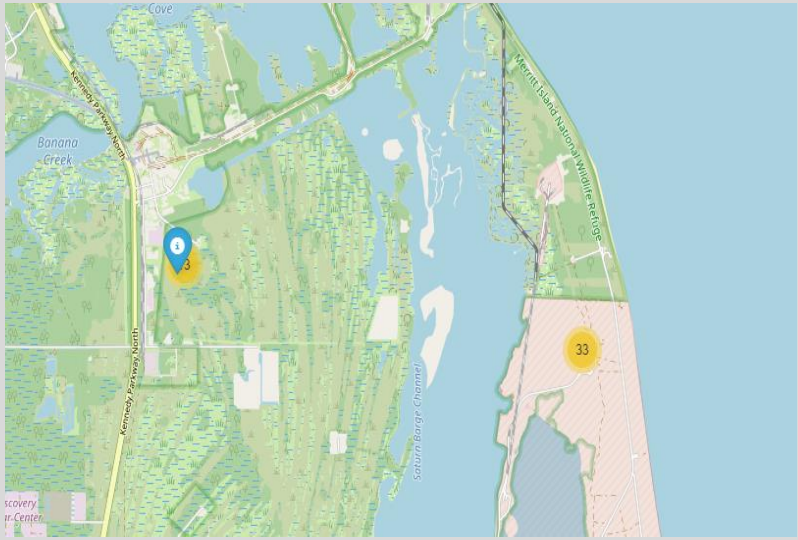
Landing_Outcome	count(Landing_Outcome)
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Section 3

Launch Sites Proximities Analysis

USA Launch Sites in California and Florida

This map shows Lunches sites in California and Florida with the number of lunches in each site.



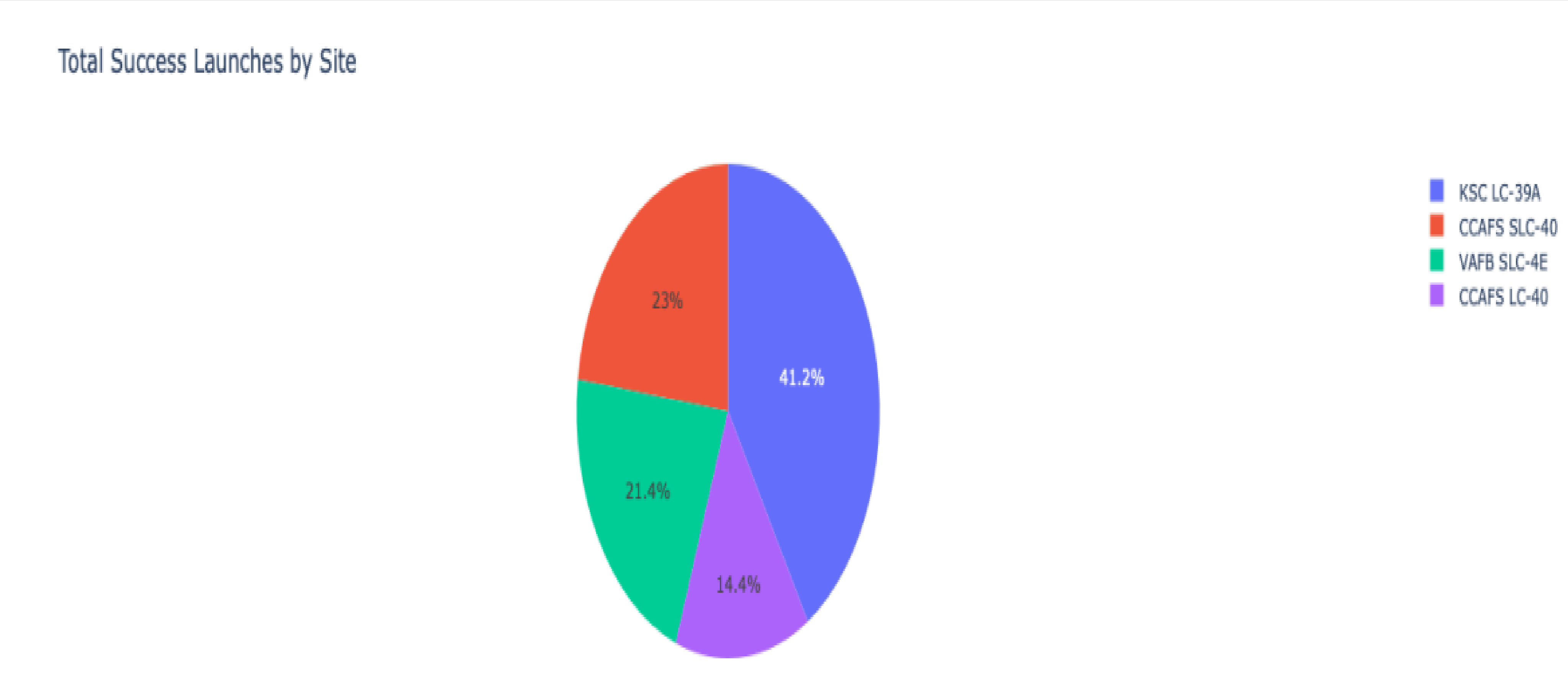
Successful and failed lunches.

This map shows the successful and failed lunches in each site.

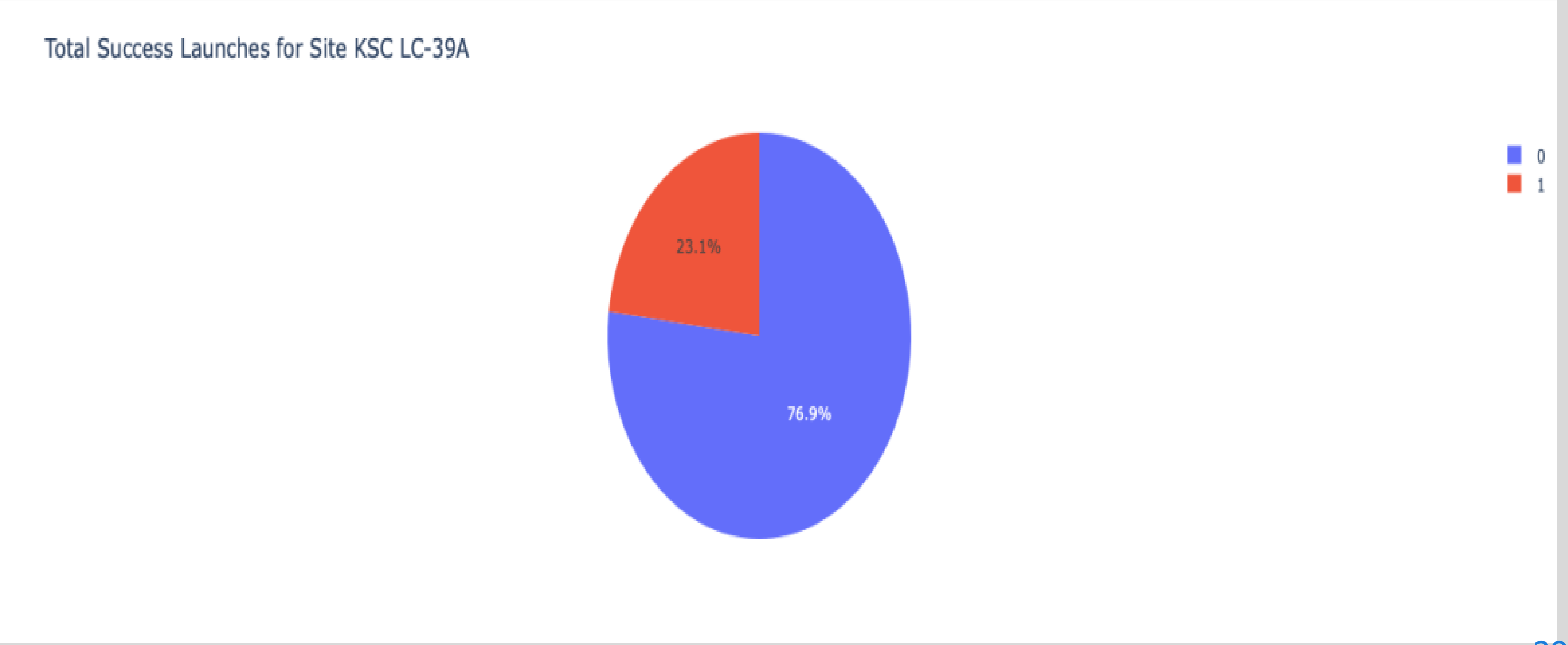
Section 4

Build a Dashboard with Plotly Dash

Launch success count for all sites



Launch site with highest launch success ratio



Payload Mass vs. Launch Outcome for all sites

Section 5

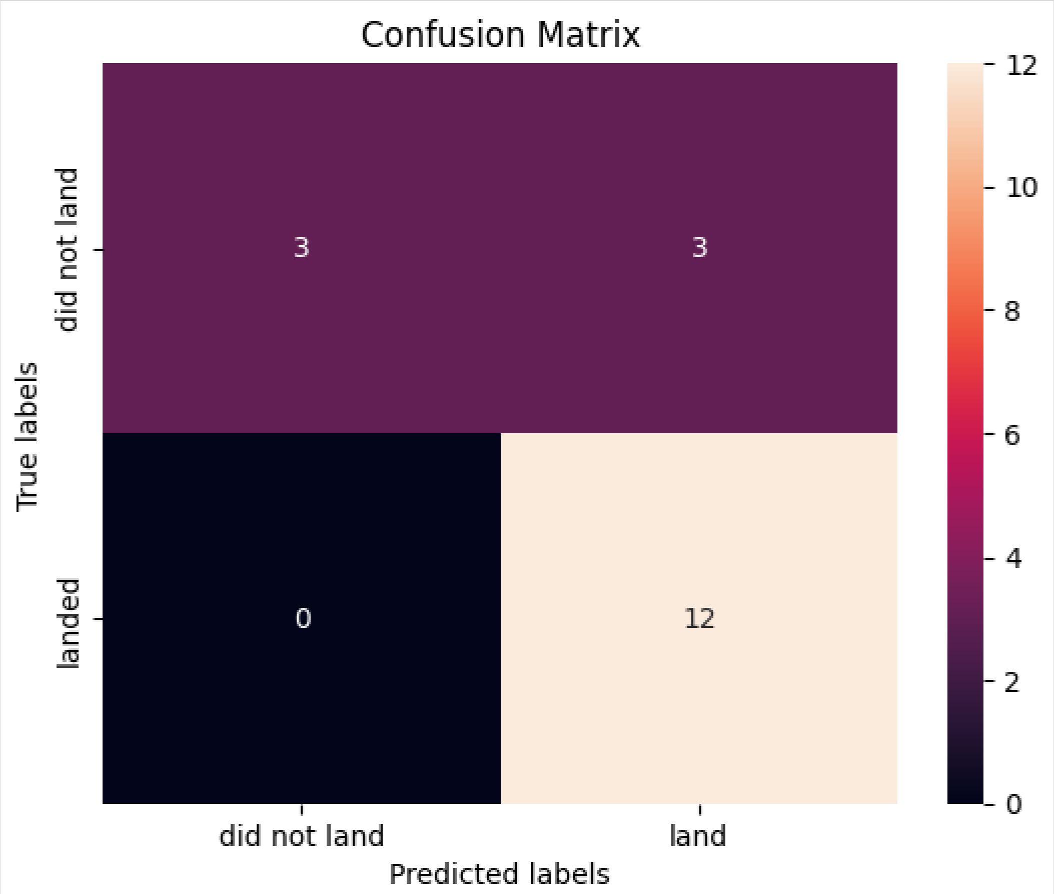
Predictive Analysis (Classification)

Classification Accuracy

- Accuracy for Logistics Regression method: 0.8333333333333334
- Accuracy for Support Vector Machine method: 0.8333333333333334
- Accuracy for Decision tree method: 0.8333333333333334
- Accuracy for K nearsdt neighbors method: 0.8333333333333334

Confusion Matrix

The Confusion Matrix shows that from 18 lunches, 6 did not land and 3 only was expected to land On the other hand, 12 lunches landed and the 15 was expected so.



Conclusions

- For a reason, all deferent classification methods has similar results.
- Some orbits has very Haigh successful rate.
- As time goes, the successful rate increases.

Thank you!