IBM Developer
SKILLS NETWORK

SpaceX Falcon 9 Landing Prediction

Tariq Elsheikh

April 11th, 2023

Outline

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

Executive Summary

- ***Summary of methodologies:***

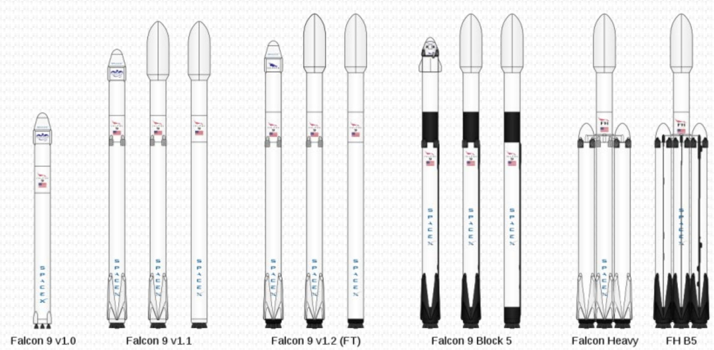
1. Data collection of SpaceX Falcon 9 using Spacex API.
2. Web scraping of SpaceX Falcon 9 records with BeautifulSoup.
3. Data wrangling of SpaceX Falcon 9.
4. Explore and prepare data of SpaceX Falcon 9 using Pandas and Matplotlib.
5. Analysis of SpaceX Falcon 9 launch sites locations with Folium & Poly Dash.
6. Machine Learning Prediction of SpaceX Falcon 9.

- ***Summary of results:***

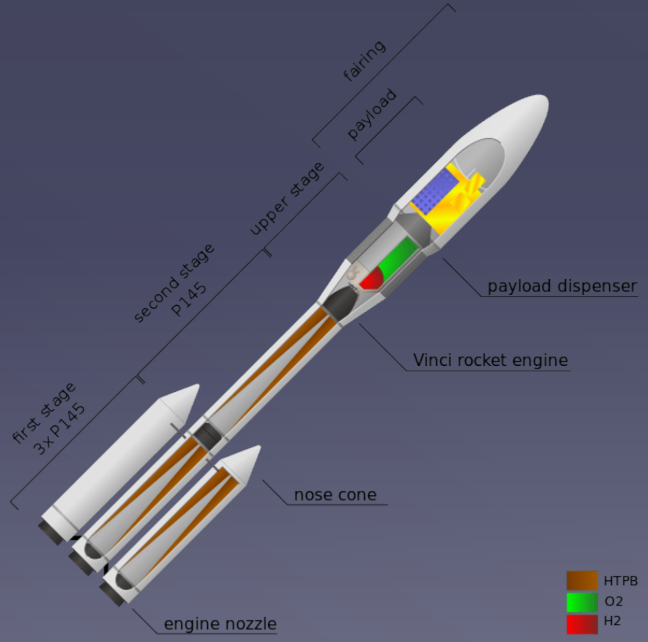
1. EDA with visualization & SQL results
2. Interactive map with Folium results
3. Ploty Dash dashboard results
4. Predictive analysis (classification) results

Introduction

- SpaceX Falcon 9 is a reusable rocket designed for commercial and government missions. The starting price of each launch is $62 million. By improving the efficiency of the manufacturing processes & research, the cost of launch & materials can be reduced.
- For safe and designated landing areas, First Stage Landing Prediction tools are applied.



Methodology



Methodology

- Data Collection
- Data Wrangling
- Exploratory Data Analysis (EDA) by using visualization and SQL.
- Interactive visual analytics by using Folium and Plotly Dash
- Predictive analysis by using classification models.

Data Collection Description

- The SpaceX API get request was used for Data collection.
- The response was decoded as a Json using .json() function call and turn it into a pandas dataframe using .json_normalize().
- The data was cleaned and checked for missing values to fill where necessary.
- Web scraping was performed from Wikipedia for Falcon 9 launch records with BeautifulSoup.
- The launch records was extracted as HTML table, then the table was parsed and converted to a pandas dataframe.

Data Collection – SpaceX API

- The SpaceX API was used to collect data, then cleaning, wrangling and formatting were done based on the provided data.
- The GitHub URL of the notebook: Data Collection - GitHub

Now let's start requesting rocket launch data from SpaceX API with the following URL:

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

```
[7]: response = requests.get(spacex_url)
```

Check the content of the response

```
[8]: print(response.content)
```


```
b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"shi
om/94/f2/NN6Ph45r_o.png","large":"https://images2.imgbox.com/5b/02/QcxHUb5V_o.
```

Data Collection – Web Scraping

- Web Scraping was performed to collect Falcon 9 Historical launch records with BeatifulSoup.
- Data frame was created by parsing the HTML table from the Wikipedia page.
- The GitHub URL of the notebook: [Web Scraping - GitHub]

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
[5]: # use requests.get() method with the provided static_url
     # assign the response to a object
     response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
[6]: # Use BeautifulSoup() to create a BeautifulSoup object from a response text content
     soup = BeautifulSoup(response.content, 'html.parser')
```

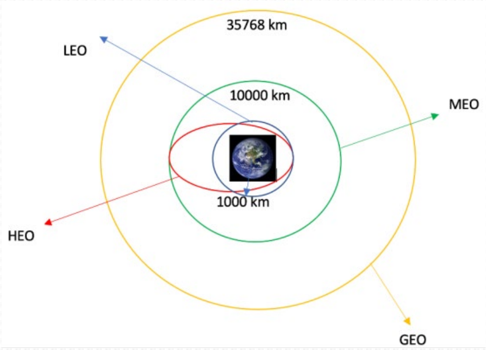
Print the page title to verify if the `BeautifulSoup` object was created properly

```
[7]: # Use soup.title attribute
     soup.title

[7]: <title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

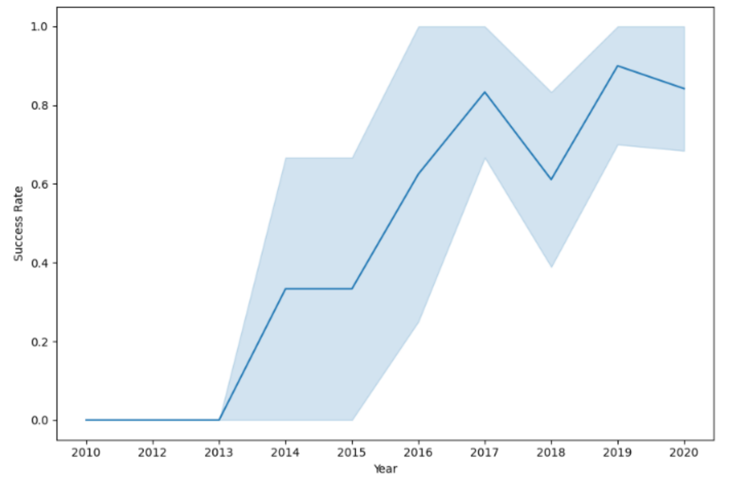
Data Wrangling

- Exploratory data analysis (EDA) was performed to determine the training labels through patterns.
- The number of launches at each site and the number of occurence of each orbit were calculated
- Landing outcome label was created from outcome column and the results were exported to csv.
- The GitHub URL of the notebook: [Data Wrangling - GitHub]



EDA – Data Visualization

- By exploring the data, it's possible to visualize the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly trend.
- The GitHub URL of the notebook: [Data Visualization - GitHub]



EDA - SQL

SQL was used to get insights from the data such as:

1. Display the names of the unique launch sites in the space mission
2. Display the total payload mass carried by boosters launched by NASA (CRS)
3. Display average payload mass carried by booster version F9 v1.1
4. List the total number of successful and failure mission outcomes
5. List the names of the booster versions which have carried the maximum payload mass
6. Display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the requested date
7. Rank the count of successful landing outcomes between certain dates.

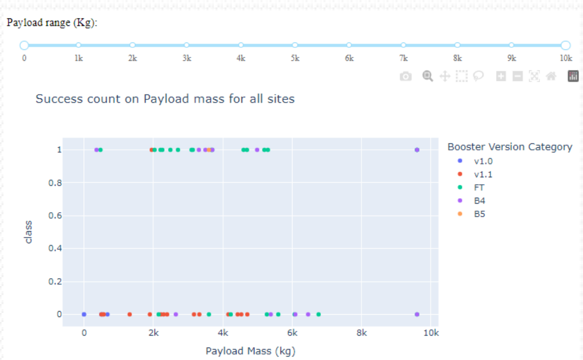
The GitHub URL of the notebook: SQL - GitHub

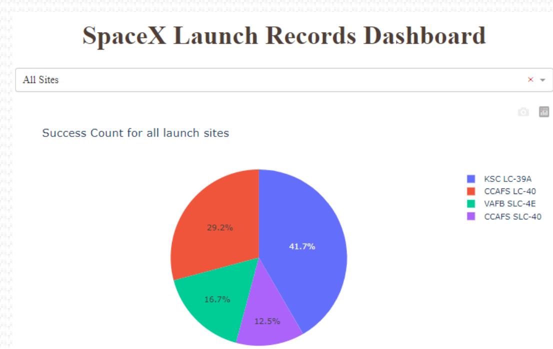
Interactive Map with Folium

- Folium map was created to mark all the launch sites and map objects such as circles, markers and lines to find out the success or failure rate of launches.
- Set outcomes for the launch are success=1, failure=0.
- The GitHub URL of the notebook: [Interactive Map - GitHub]

Dashboard with Plolty Dash

- an interactive dashboard with Plotly dash was built.
- Pie charts that show the total launches by a certain sites were plotted.
- Scatter graph that show the relationship with Outcome and Payload Mass (Kg) for the different booster version were plotted.
- The GitHub URL of the notebook: [Dashboard - GitHub]

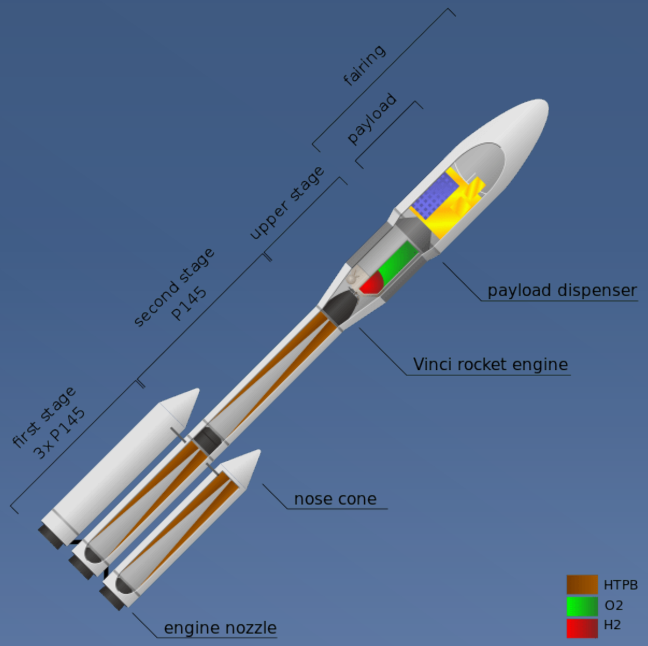




Predictive Analysis (Classification)

- The data was loaded using numpy and pandas, then transformed and split our data into training and testing.
- Different machine learning models were built and expolored different hyperparameters using GridSearchCV.
- Accuracy was assigned as the metric for the model, the model was improved using feature engineering and algorithm tuning.
- The best performing classification model was found.
- The GitHub URL of the notebook: Predictive Analysis - GitHub

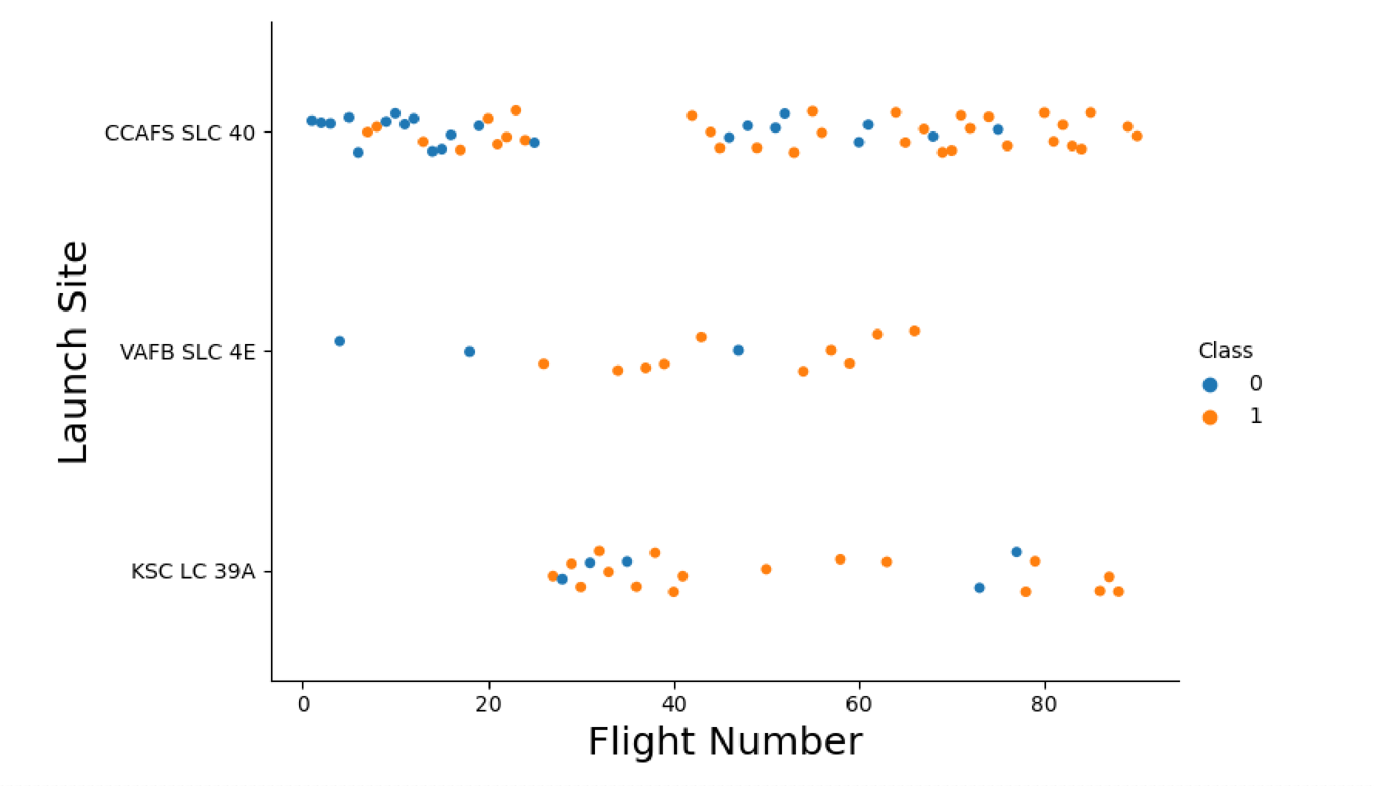
Results


fairing
payload
upper stage
second stage
P145
first stage
3x P145
payload dispenser
Vinci rocket engine
nose cone
engine nozzle
HTPB
O2
H2

Results

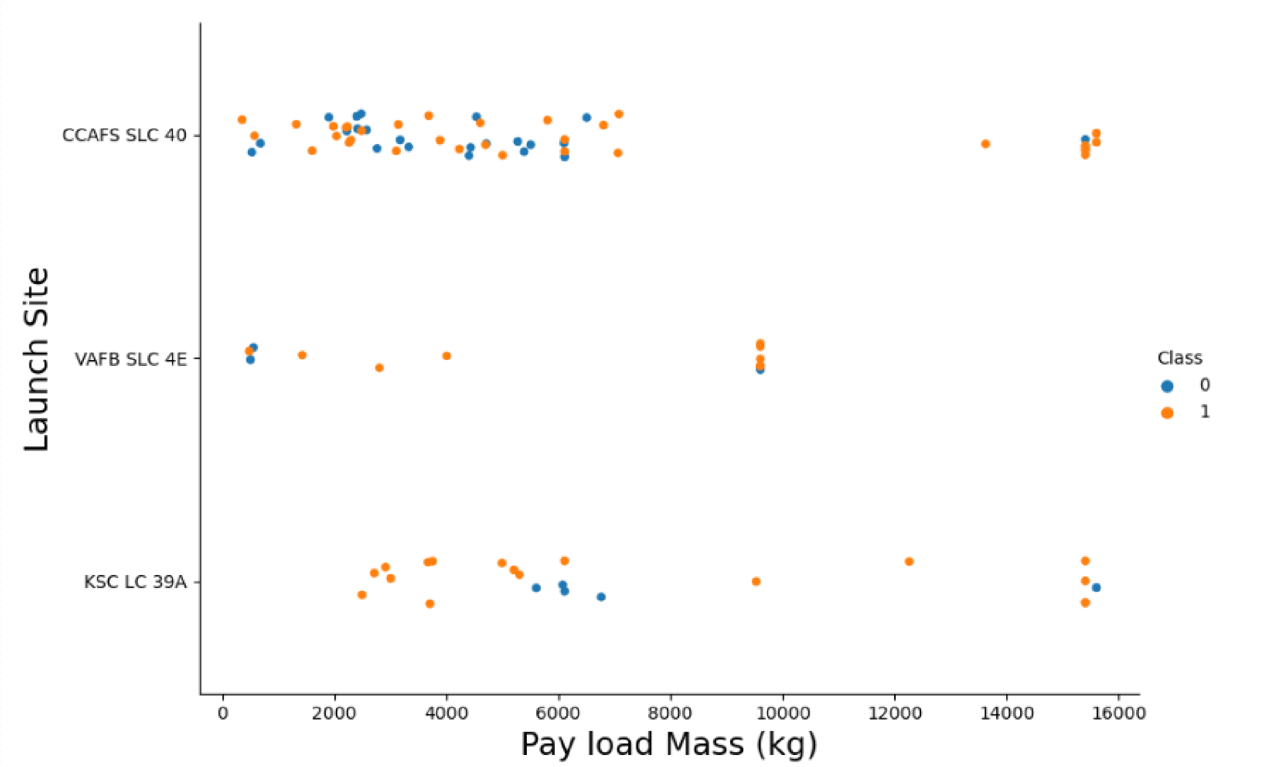
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Flight Number vs. Launch Site



From plot, the larger the flight amount at a launch site, the greater the success rate at a launch site.

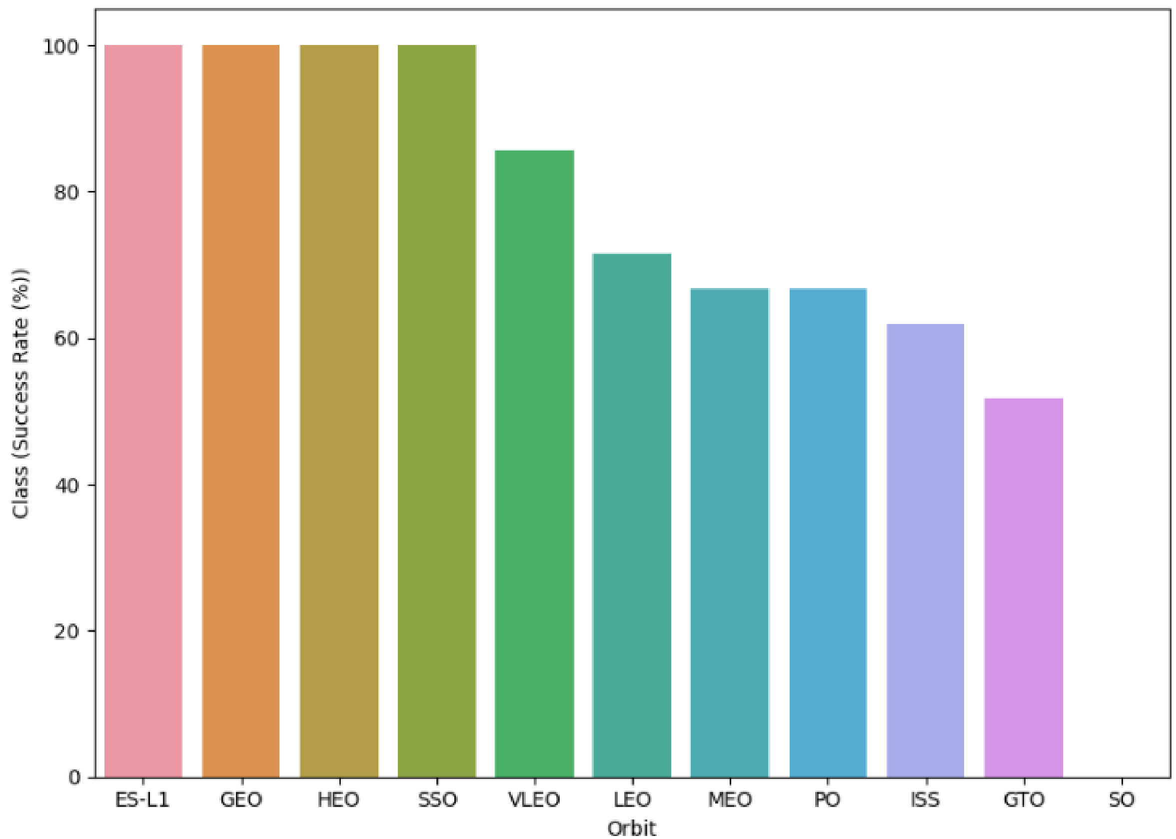
Payload vs. Launch Site



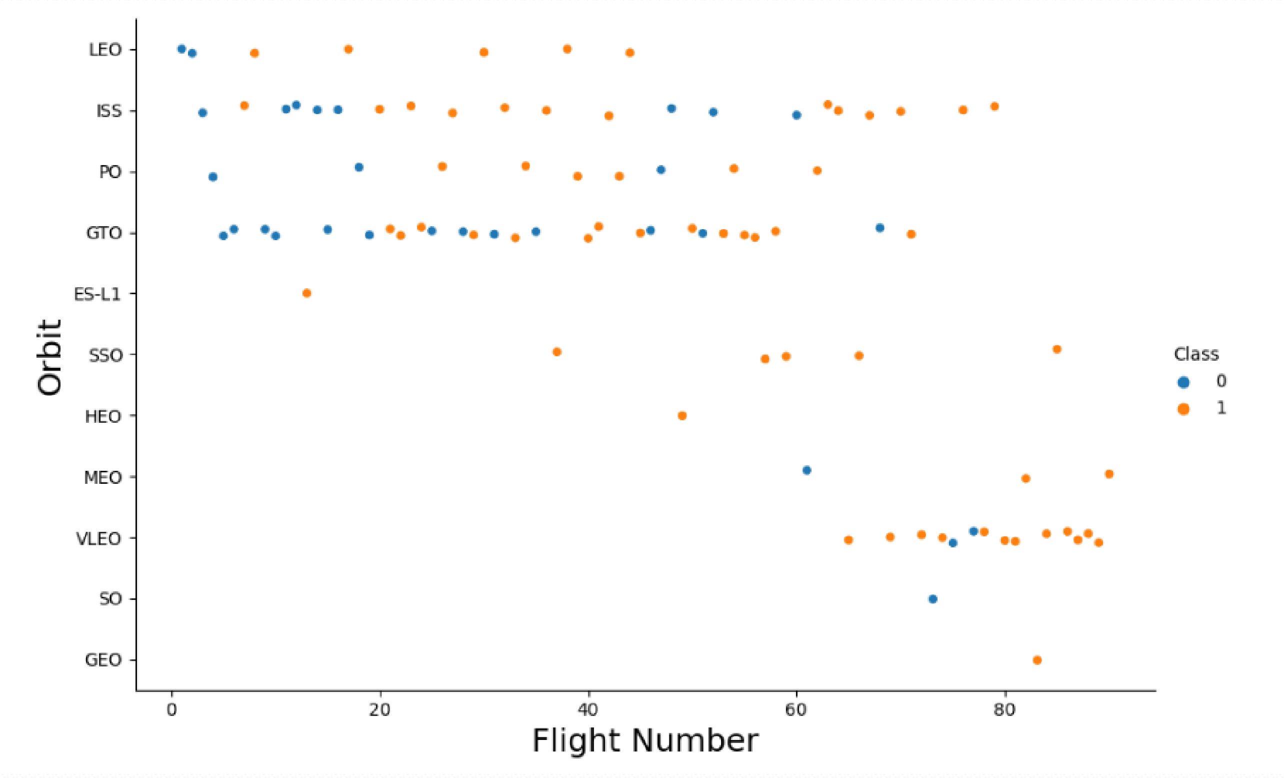
From plot, the greater the payload mass for launch site CCAFS SLC 40, the higher the success rate for the rocket. For VAFB SLC 4E, there are no launches for Pay Load Mass greater than 10,000 kg.

Success Rate vs. Orbit Type

From plot, orbit ES-L1, GEO, HEO & SSO have the highest success rate at 100%, while SO success rate is 0%.

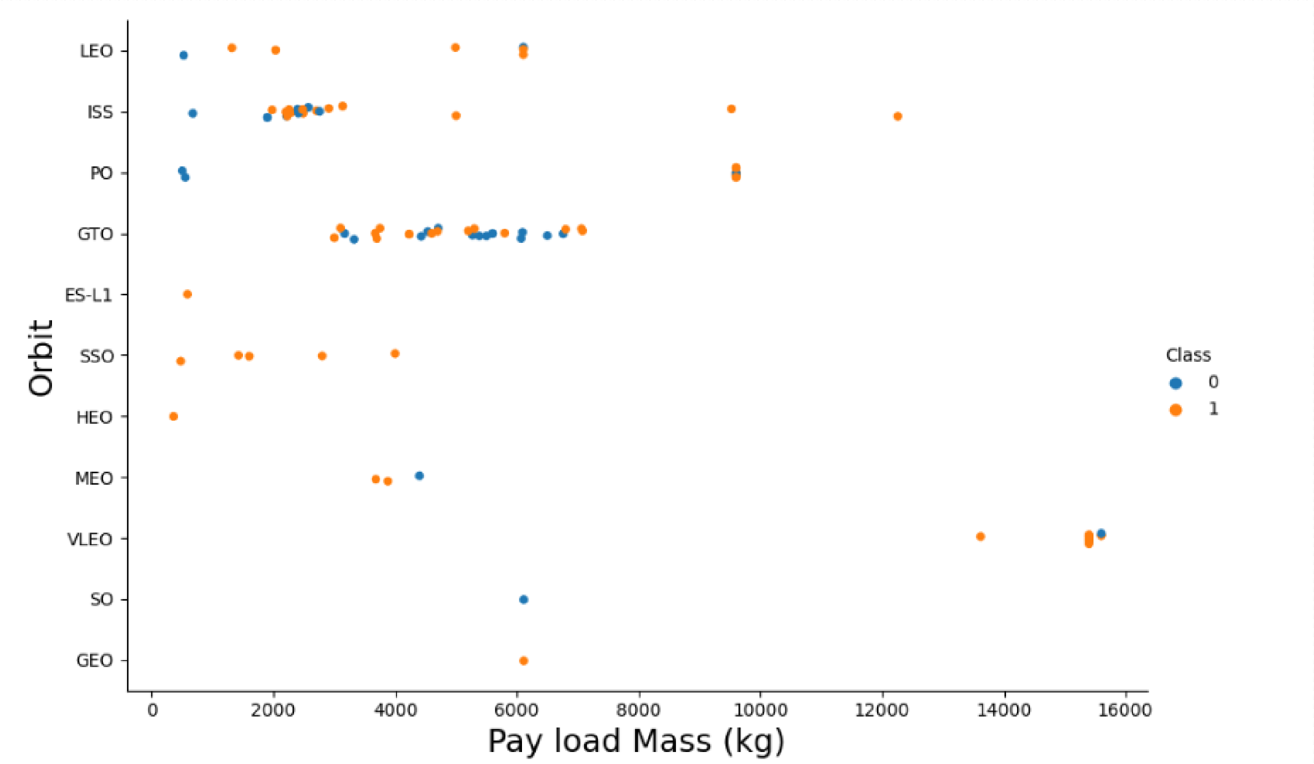


Flight Number vs. Orbit Type



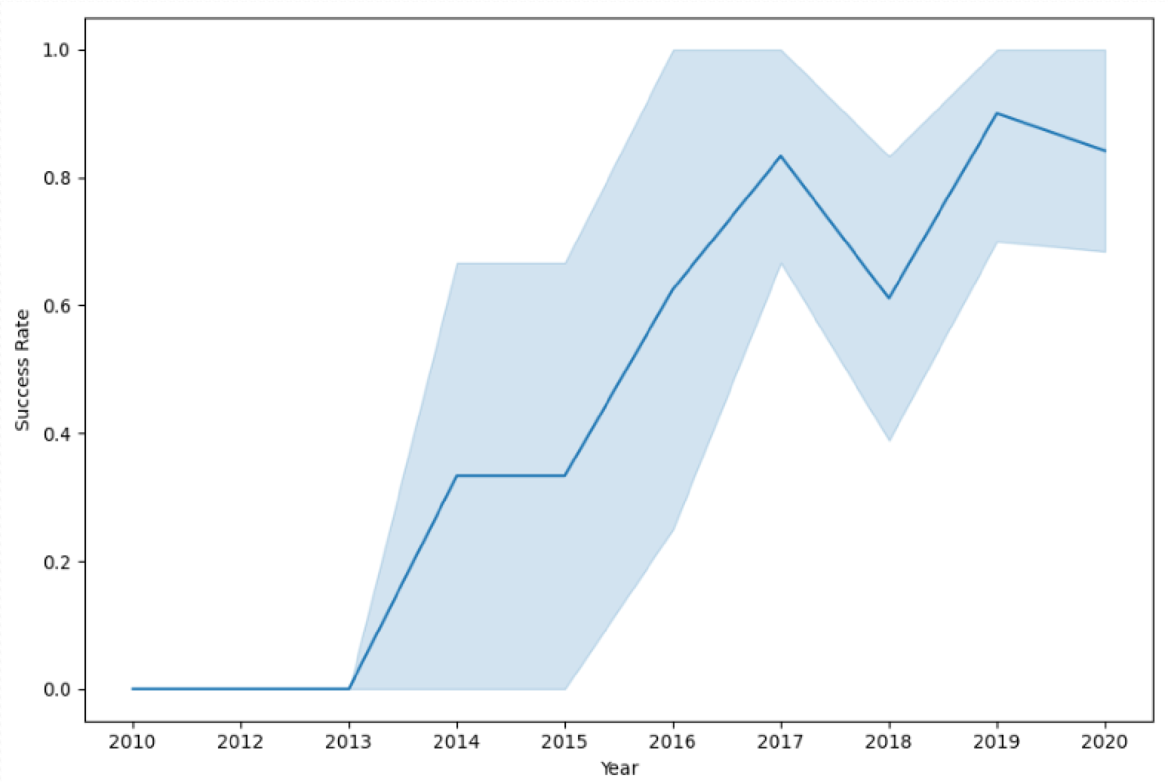
From plot, LEO orbit success rate is related to the number of flights. While for GTO orbit, there is no relationship between flight number and the orbit.

Payload vs. Orbit Type



From plot, LEO, ISS & PO orbits success rate is related to the pay load mass. While for GTO orbit, there is no relationship between the pay load mass and the orbit.

Launch Success Yearly Trend



From plot, the success rate increases from 2013 to 2020, the rate of increase vary from year to year as shown.

All Launch Site Names

Display the names of the unique launch sites in the space mission

```
[7]: %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

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

[7]:

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

DISTINCT keyword was used to show only unique launch sites from the SpaceX data.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
[8]: %sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

The shown command was used to view 5 records where launch sites begin with ‘CCA’

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

[9]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
```

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

[9]:

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

The total payload carried by boosters from NASA is 45596 kg using the shown command.

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
[10]: %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

[10]:

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

The average payload mass carried by booster version F9 v1.1 is 2534.666 using the shown command.

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[11]: %sql SELECT MIN(DATE) FROM 'SPACEXTBL' WHERE "Landing _Outcome" = "Success (ground pad)";
```

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

[11]:

MIN(DATE)
01-05-2017

The first successful ground landing date is 01-05-2017 using the shown command.

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[12]: %sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG
```

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

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

The boosters that have successfully Drone Ship Landing were filtered to determine successful landing with payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

The number of successful and failure missions outcomes are represented using the shown code.

Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FR
```

List of boosters that have carried maximum payload mass is shown which is 15600 kg.

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

```
%sql SELECT substr(Date,7,4), substr(Date, 4, 2),"Booster_Version", "Launch_Site", Payload, "PAYLOAD_MASS__KG_", "Mission_Outcome", "Land
```

substr(Date,7,4)	substr(Date, 4, 2)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing _Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

Failed landing outcomes in drone ship, their booster versions, and launch site names for year 2015 were measured by using the shown code.

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

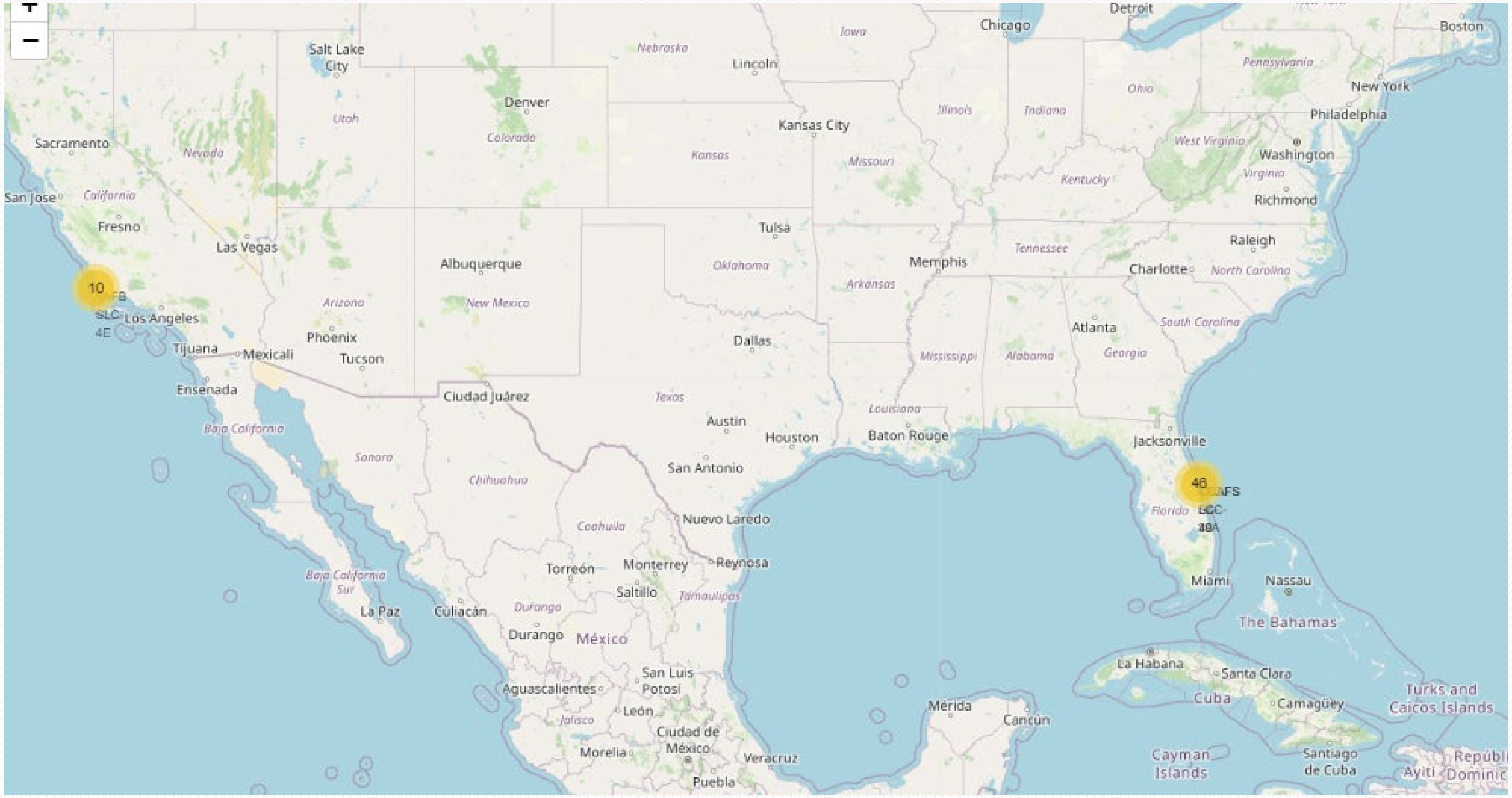
Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%sql SELECT * FROM SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success%' AND (Date BETWEEN '04-06-2010' AND '20-03-2017') ORDER BY Date DESC
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)
17-12-2019	00:10:00	F9 B5 B1056.3	CCAFS SLC-40	JCSat-18 / Kacific 1, Starlink 2 v1.0	6956	GTO	Sky Perfect JSAT, Kacific 1	Success	Success
16-11-2020	00:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500	LEO (ISS)	NASA (CCP)	Success	Success
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
15-11-2018	20:46:00	F9 B5 B1047.2	KSC LC-39A	Es hail 2	5300	GTO	Es hailSat	Success	Success
14-08-2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
13-06-2020	09:21:00	F9 B5 B1059.3	CCAFS SLC-40	Starlink 8 v1.0, SkySats-16, -17, -18, GPS III-03	15410	LEO	SpaceX, Planet Labs	Success	Success

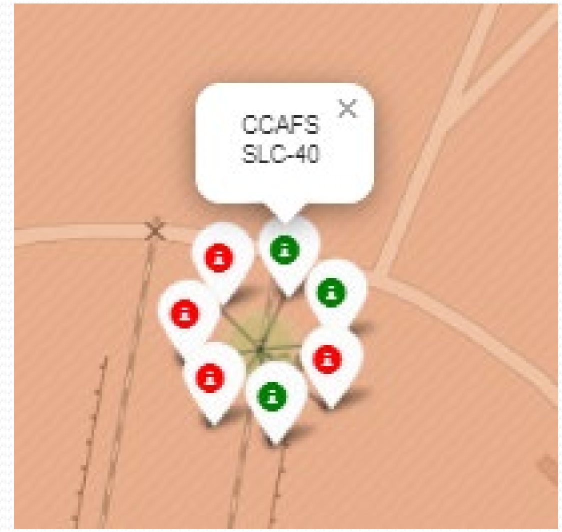
The count of landing outcomes was ranked between 2010-06-04 and 2017-03-20.

Markers of all launch sites on US



It’s noticed all launch sites are close to the US coast in Florida & California.

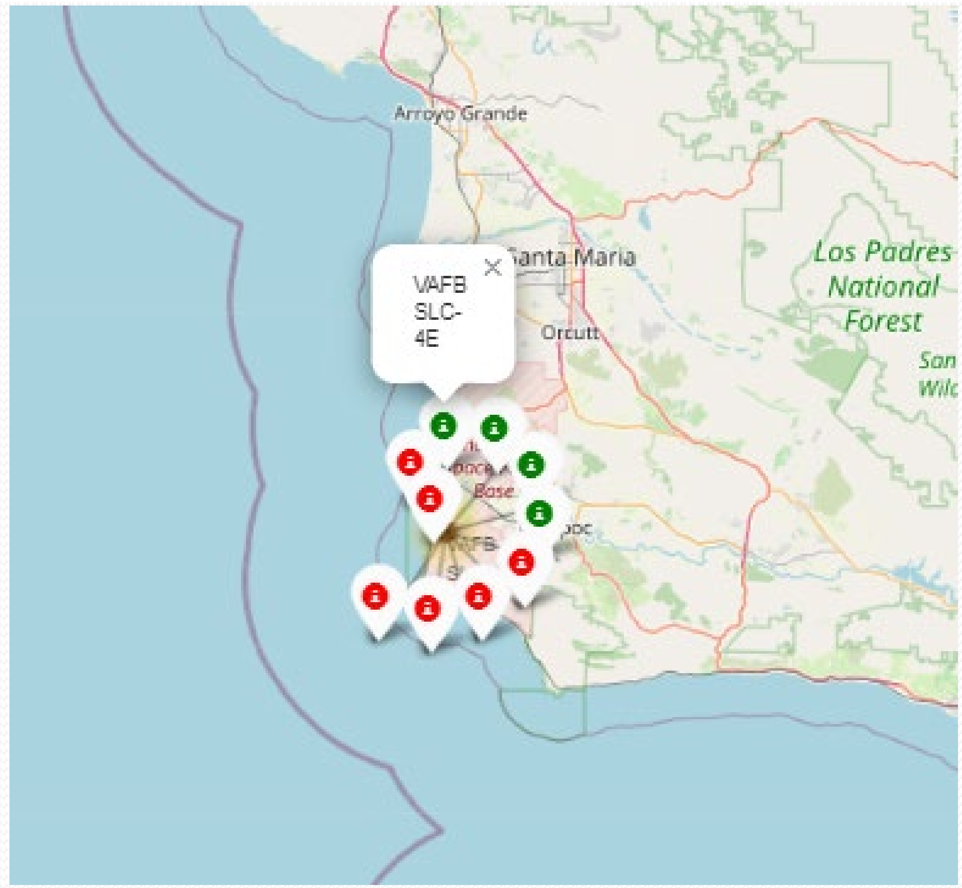
Launch outcomes on map with Color Markers



- Florida Sites: Launch Site KSC LC-39A has higher success rate compared to the other two sites.

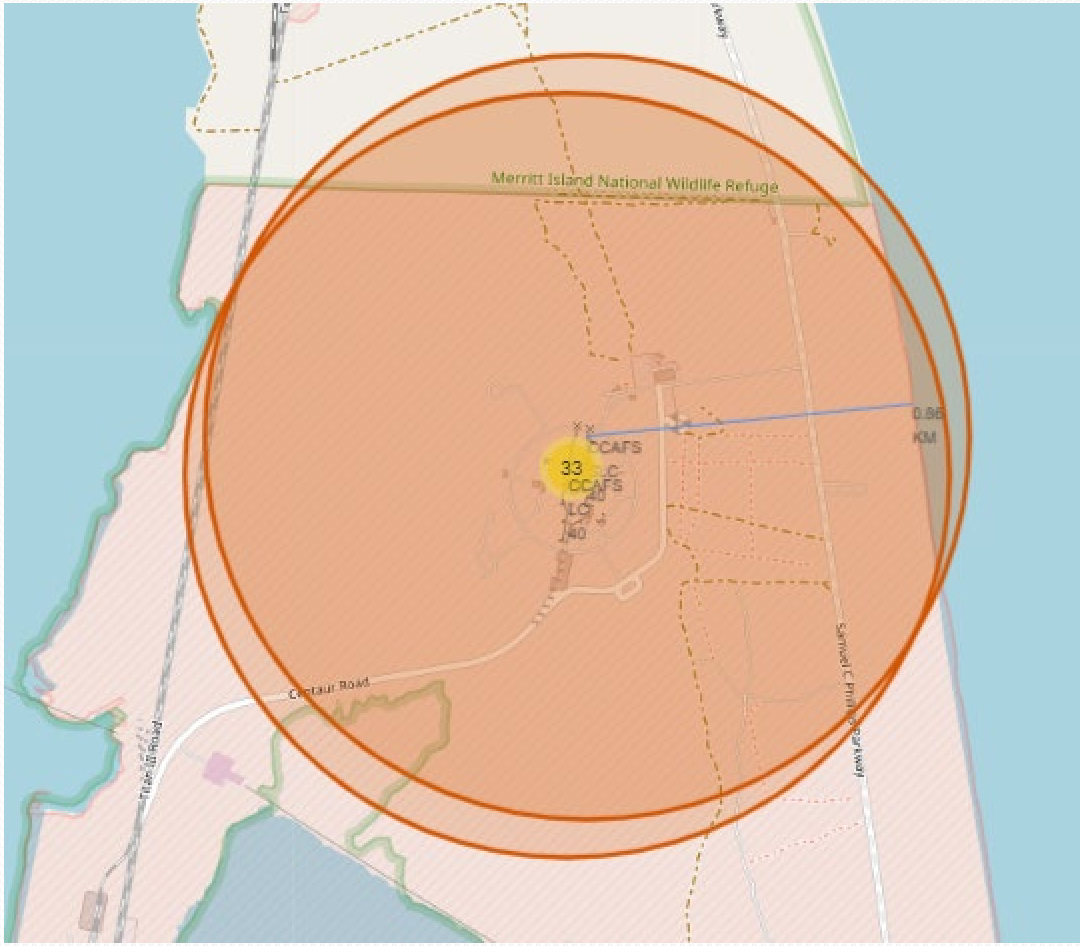
Launch outcomes on map with Color Markers

- California Sites: Launch Site VAFB SLC-4E has lower success rate compared to KSC LC-39A launch site in Florida

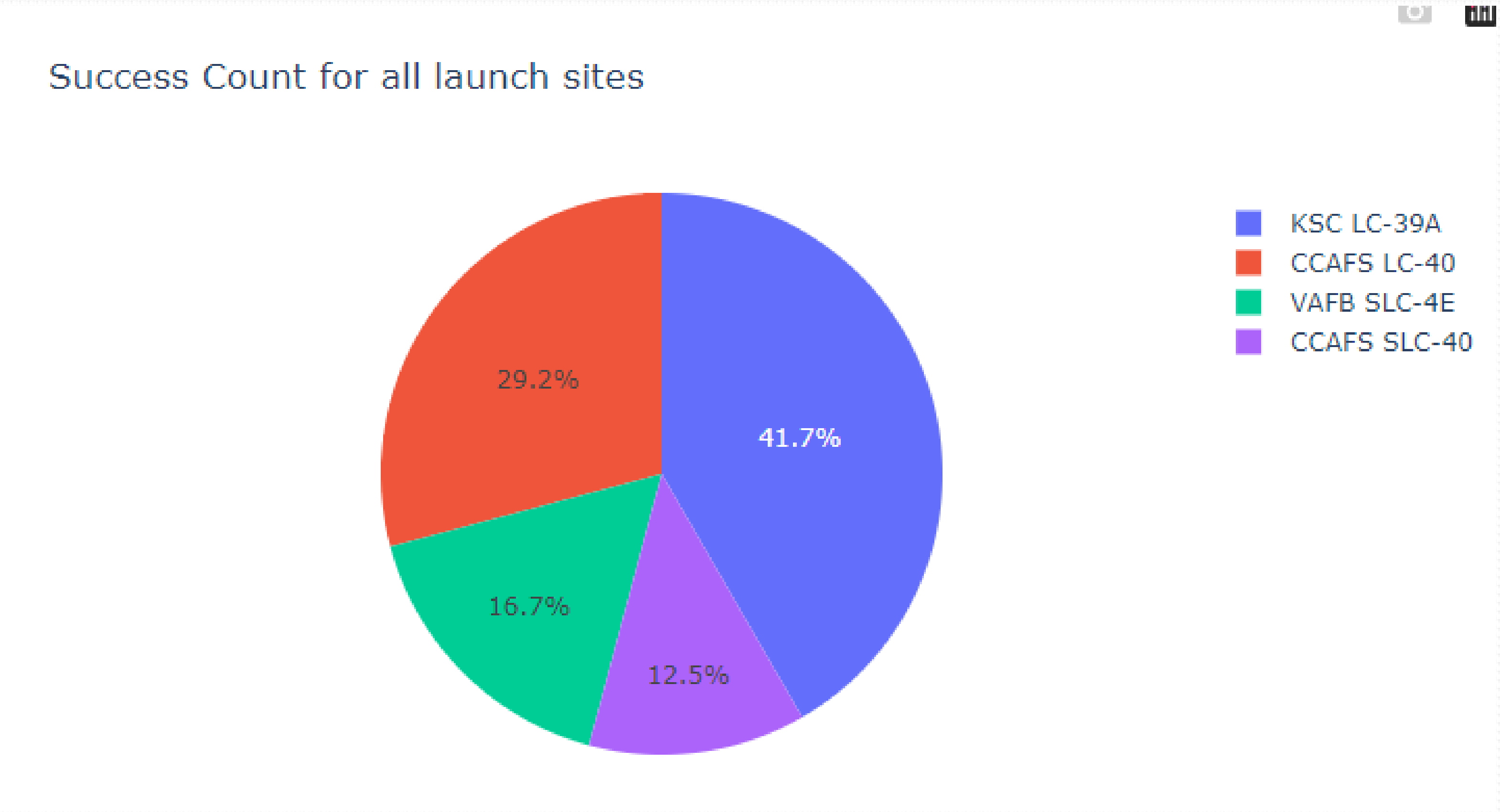


Distance between a launch site to its proximities

- California Sites: Launch Site CCAFS SLC-40 proximity to coastline is 0.86 km.



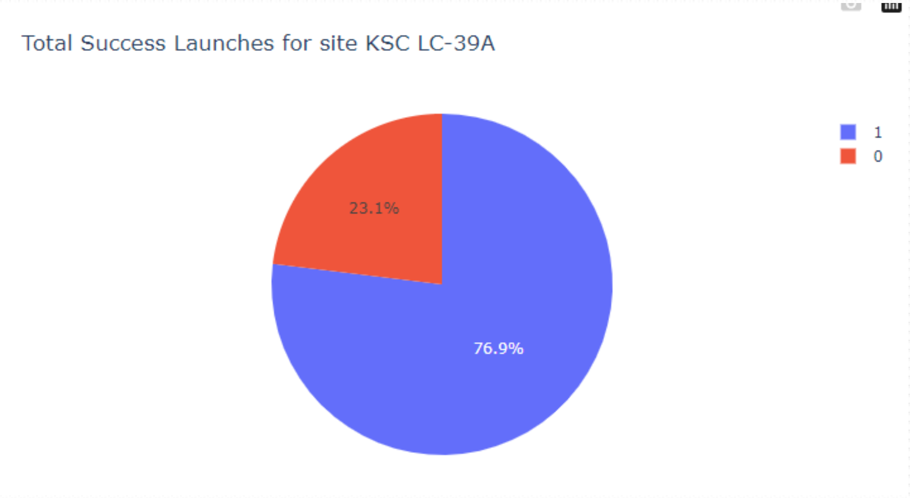
Pie Chart



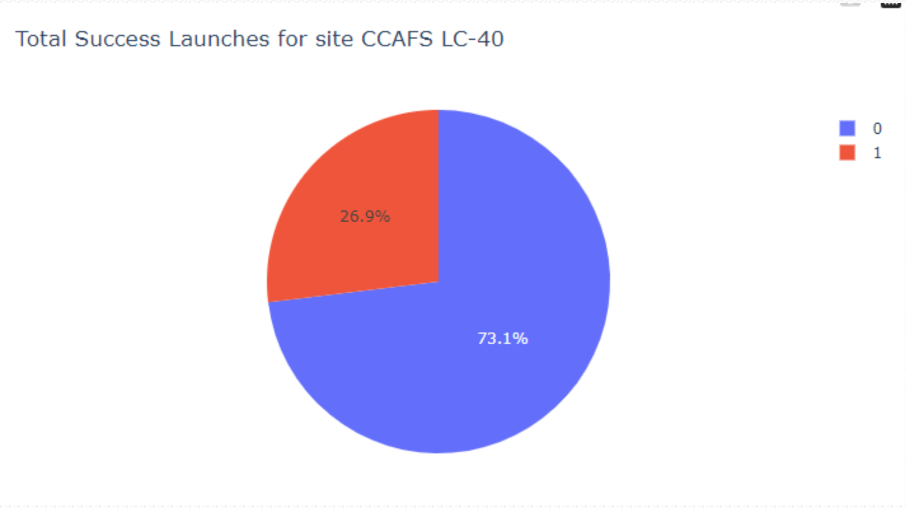
Pie Chart for launch success count for all sites

Pie Chart

1. KSC LC-39A
2. 1st Highest Success Launches
3. Success rate is 76.9% & Failure rate is 23.1%

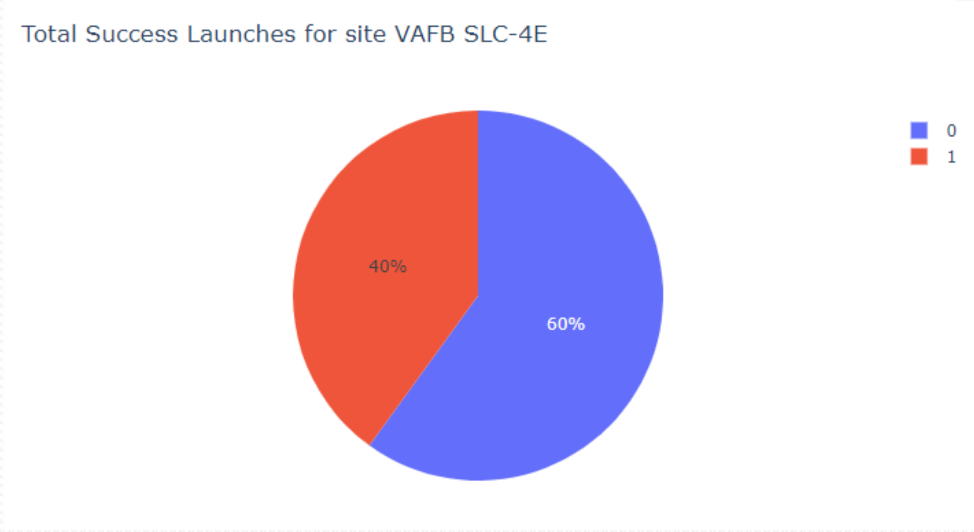


1. CCAFS LC-40
2. 2nd Highest Success Launches
3. Success rate is 73.1% & Failure rate is 26.9%

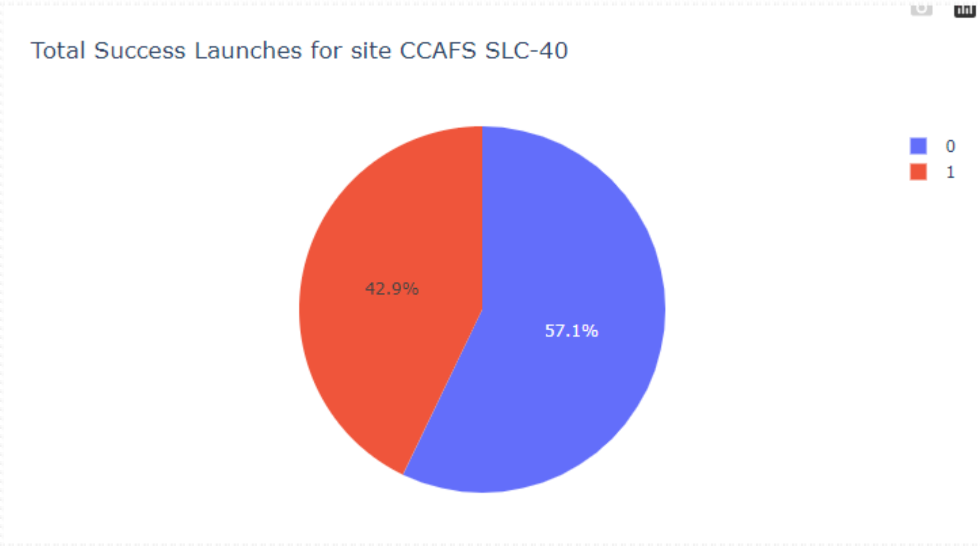


Pie Chart

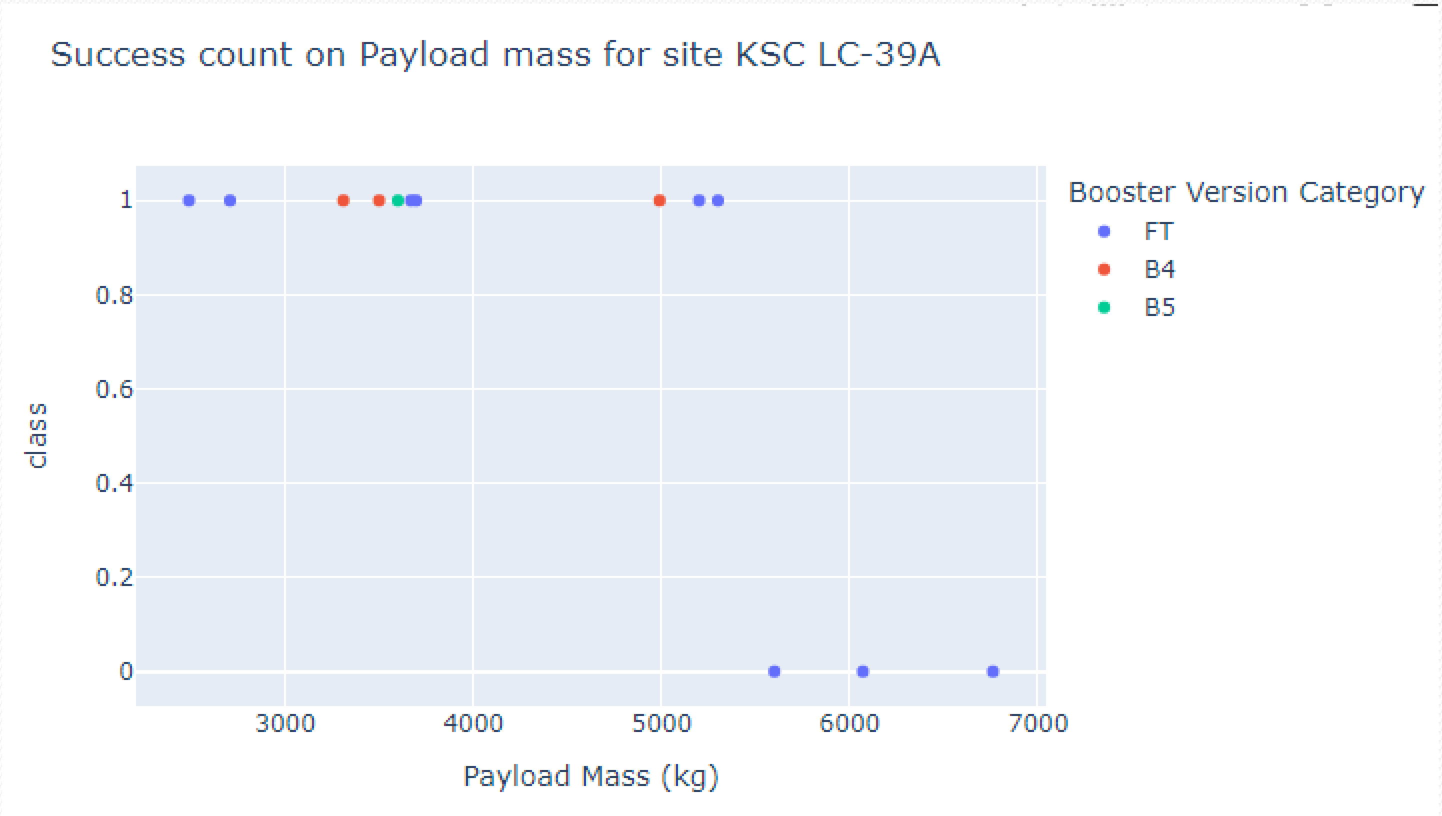
1. VAFB SLC-4E
2. 3rd Highest Success Launches
3. Success rate is 60% & Failure rate is 40%



1. CCAFS SLC-40
2. 4th Highest Success Launches
3. Success rate is 57.1% & Failure rate is 42.9%



Payload vs. Launch Outcome scatter plot for all sites



The success rate increases when the payload mass decreases.

Classification Models Accuracy

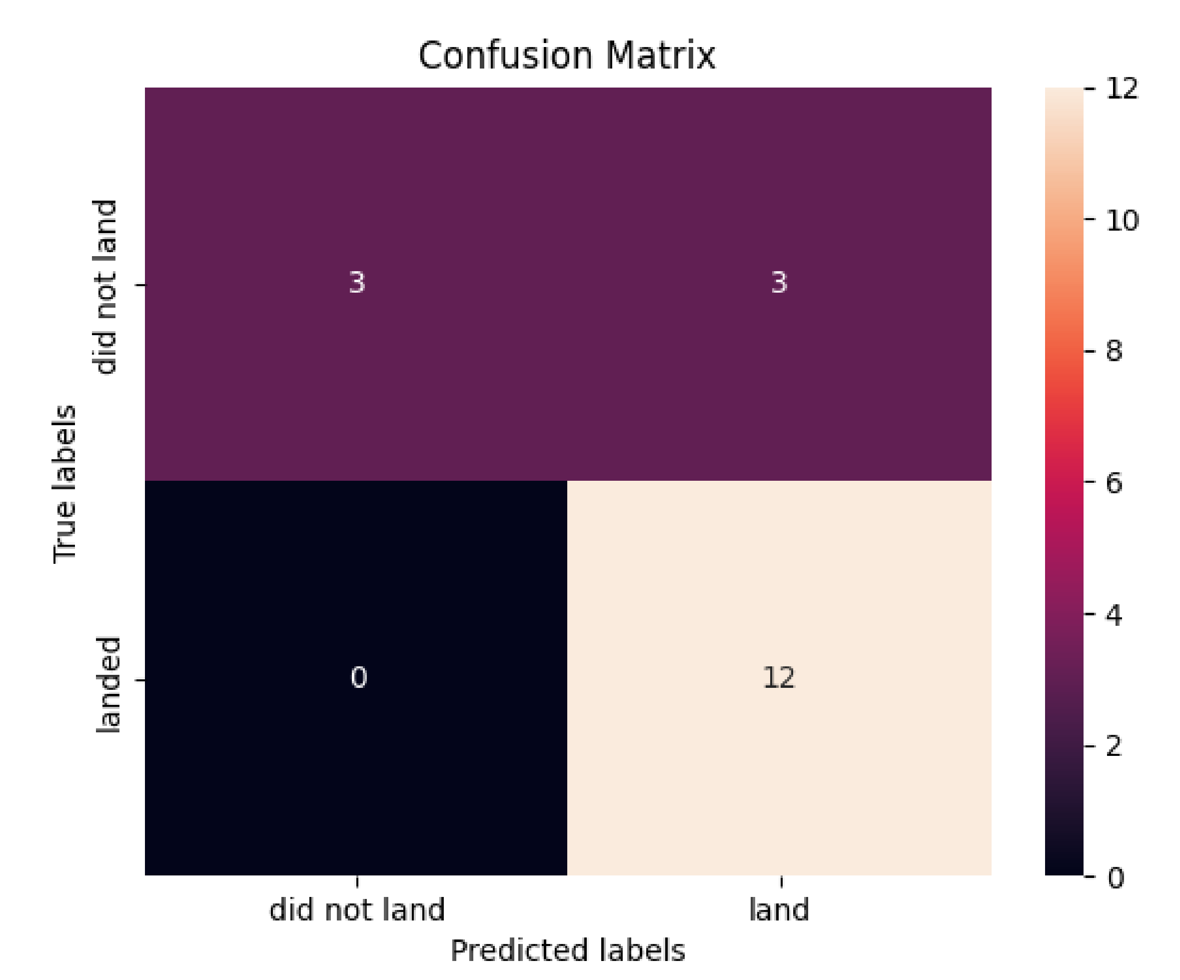
All 4 methods have the same accuracy of 0.8333

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

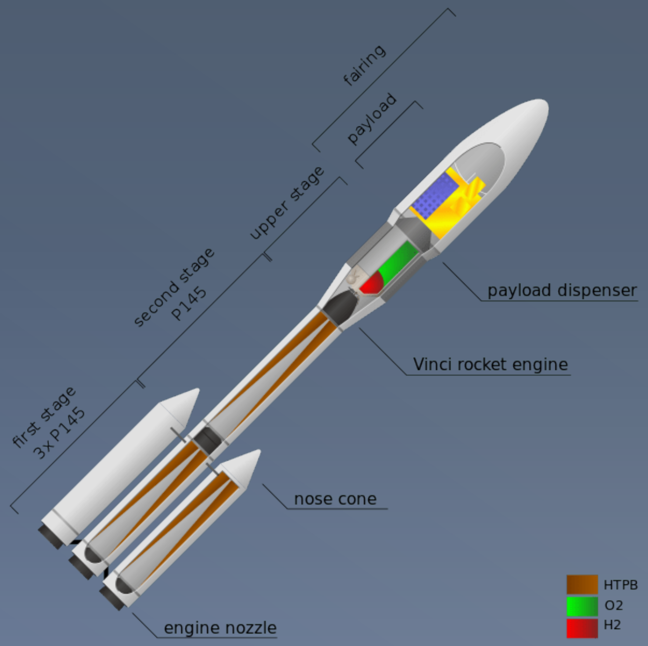
Confusion Matrix

- All 4 classification models have the same confusion matrixes and can distinguish between the different classes.

- One of the major problems is the false positives such as unsuccessful landing marked as successful landing by the classifier.



Conclusion



fairing
payload
upper stage
second stage
P145
first stage
3x P145
payload dispenser
Vinci rocket engine
nose cone
engine nozzle
HTPB
O2
H2

Conclusion

- Various launch sites have different success rates. For example, KSC LC-39A has the highest success launch rate with 76.9% while CCAFS SLC-40 success launch rate is 57.1%

- Launch success rate started to increase from 2013 to 2020.

- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.

- The Decision tree classifier is the best machine learning algorithm for this task.

Appendix – GitHub Links

- Main GitHub: [Capstone Project - GitHub]
- Data Collection: Data Collection - GitHub
- Web Scraping: Web Scraping - GitHub
- Data Wrangling: Data Wrangling - GitHub
- Data Visualization: Data Visualization - GitHub
- SQL: SQL - GitHub
- Interactive Map with Folium: Interactive Map - GitHub
- Dashboard with Plolty Dash: Dashboard - GitHub
- Predictive Analysis (Classification): Predictive Analysis - GitHub


Thanks
IBM Developer
SKILLS NETWORK