IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ekene Emmanuel AJEMBA

December 2022

Outline

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

Executive Summary

- **Summary of methodologies**

  In order to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully, the methods of Data Collection using API with Web Scraping, Data Wrangling, Exploratory Data Analysis (EDA) with SQL and Data visualization, model development, Machine Learning Prediction. model evaluation are applied. Data Collection Methodology describes how data was collected and Data Wrangling describes how data was processed. EDA was performed for visualization, Folium and Plotly Dashboard dependencies was used for interactive visual analytics as the classification model used was the hallmark of the predictive analysis

- **Summary of all results**

  Launch site has a strong relation with success rate. CCAFS SLC-40 and KSC LC-39A lunch sites are most feasible sites with highest success rate and highest amount of launches.

  The ocean & RTLS outcome have succeeded on small and medium mass launches, and there is more chances for improvement. **the best performance model is decided by score evaluation**

Introduction

- **Project background and context**
- SpaceX has gained worldwide attention for a series of historic milestones. It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

- I will take the role of a data scientist working for a new rocket company to determine the price of each launch of SPACE X, Do this by gathering information about Space X and creating dashboards for the team and also determine if SpaceX will reuse the first stage Instead of using rocket science to determine if the first stage will land successfully, I will also train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

- **Problems you want to find answers**
  - Sources of data
  - Data managing
  - Data processing
  - Data analysis
  - Data understanding

Section 1

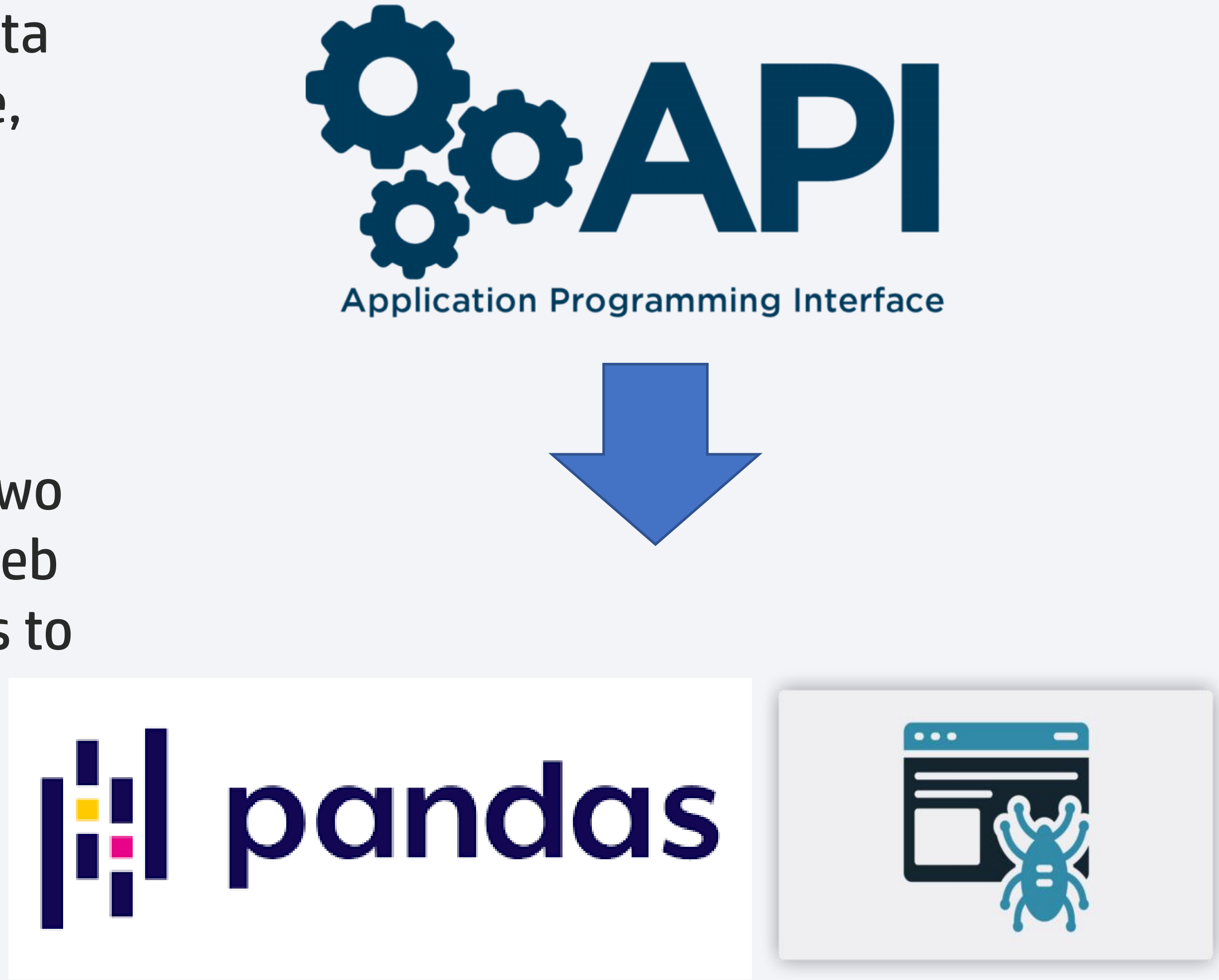
Methodology

Methodology

- Data collection methodology:
  - Describe how data was collected
  - API and web scraping method or mechanism are applied to collect the data.
- Perform data wrangling
  - Preprocess includes filling NaN, discard outlier, generating secondary data, etc.
  - Description on how data was processed
- 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, train and evaluate several classification models, built after exploratory. for example data standardizer, logistic regression training, tuned by the Grid searching method had best performance decided by score evaluation.

Data Collection

- We can say that when there is no data there is no scientific data. Therefore, the fastest way to collect big data is through the Internet.
- API is the acronym for Application Programming Interface, which is a software intermediary that allows two applications to talk to each other. Web scraping is the process of using bots to extract content and data from a website.



Data Collection – SpaceX API

- A get request to the SpaceX REST Application Programming Interface (API) endpoints, and JSON method was used to view and extract data about boosters into flat tables. The goal is to use this data to predict whether SpaceX will attempt to land a rocket or not.

- Python BeautifulSoup package is used to parse or web scrape some HTML tables that contain valuable Falcon 9 launch records.

- further visualization and analysis was applied
- IBM Watson Studio Notebook Link
- GitHub JupyterNotebook Link

Data Collection - Scraping

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Extracted **Falcon 9** launch records HTML table dictionary parsed and converted **to Pandas** Dataframe from Wikipedia

Creation of BeautifulSoup Object, empty dictionary and filling of same with keys from the column names

Extracted columns used for further visualization & analysis of names from the HTML table data about landing pads and boosters

Fill up the dictionary with launch records extracted from table rows

Convert the dictionary into a CSV dataset

- IBM Watson Studio JupyterNotebook Link
- Github Notebook Link

Data Wrangling

- Transforming and mapping data from raw data, which including transforming data to dataframe, discarding the null values, discarding the outliers, normalizing the values, and some basic data exploring.

Use the method `value_counts()` on the column `LaunchSite` to determine the number of launches on each site:

In [7]:
```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

Out[7]:
```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

- An Exploratory Data Analysis (EDA) performance and Training Labels processes was determined and it is given in an overview.

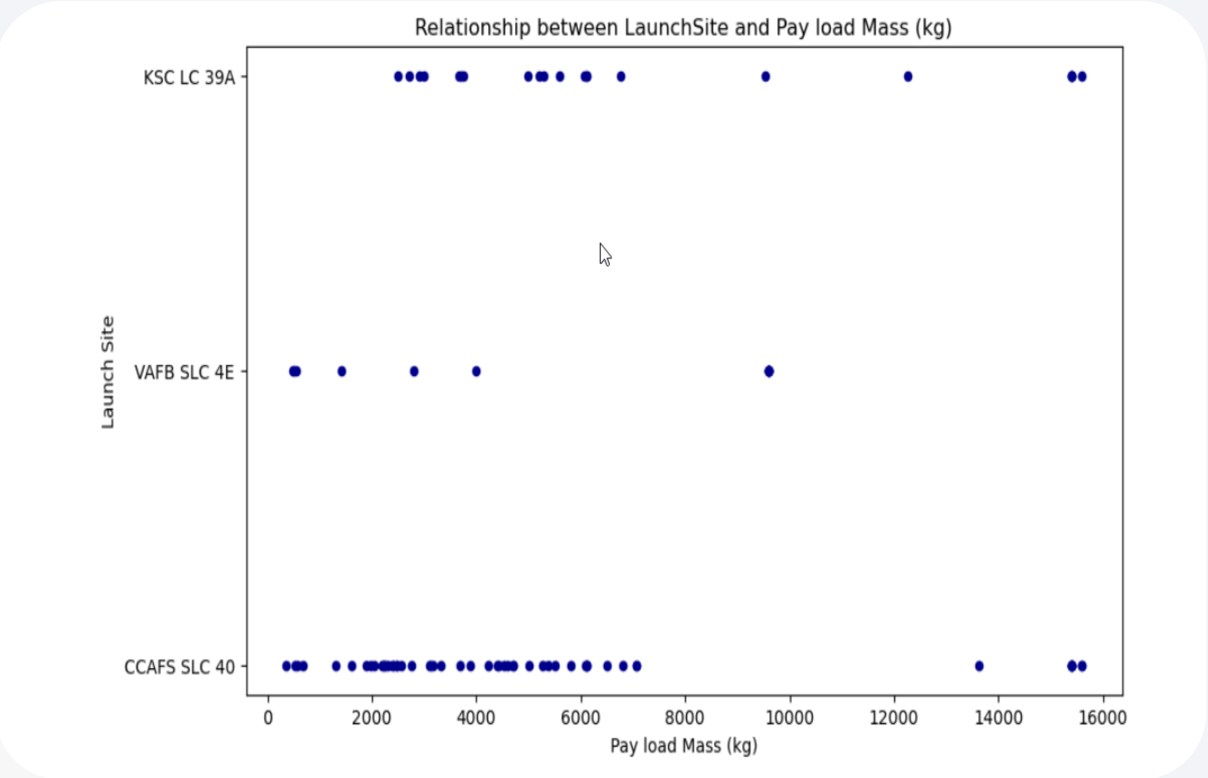
- [IBM Watson Studio JupyterNotebook Link]()
- [Github JupyterNotebook Link]()

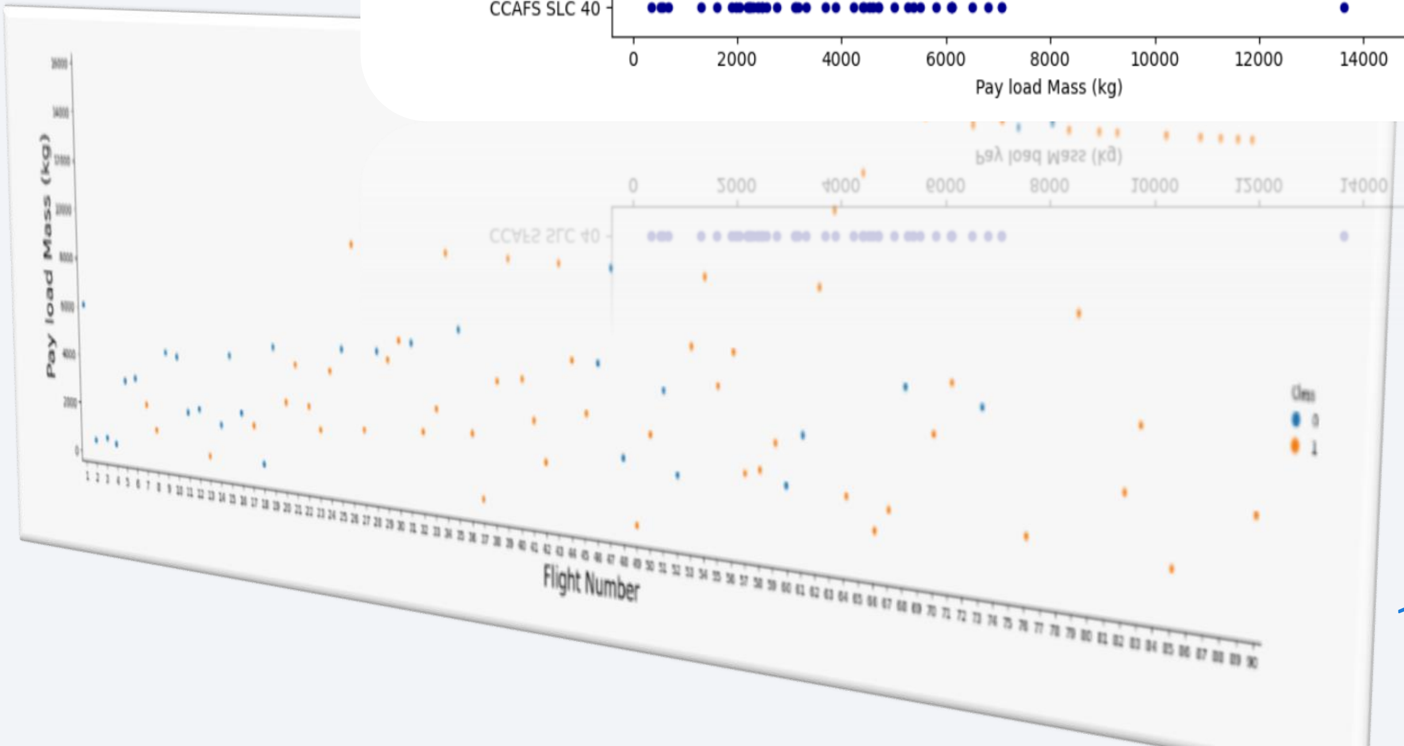
- Check the LINK for completed IBM Watson Studio JupyterNotebook LAB analysis and GitHub notebook Repository

EDA with Data Visualization

- **Types of Charts Used :**
- **Scatter plot** - Variable relationship and easy to handle of Flight Number vs Payload Mass plot, Flight Number vs Launch Sites , Payload and Launch Sites , Flight Number and Orbit Type , Payload and Orbit Type
- • **Bar chart** – Success rate of each orbit with categorical relationship & impactful Visualization.
- • **Line plot** – success rate has time series tendency and Date
- EDA with Data Visualization notebook link is given below

- **Scatter plot, Bar plot, and Line plot are used to visualize the data.**





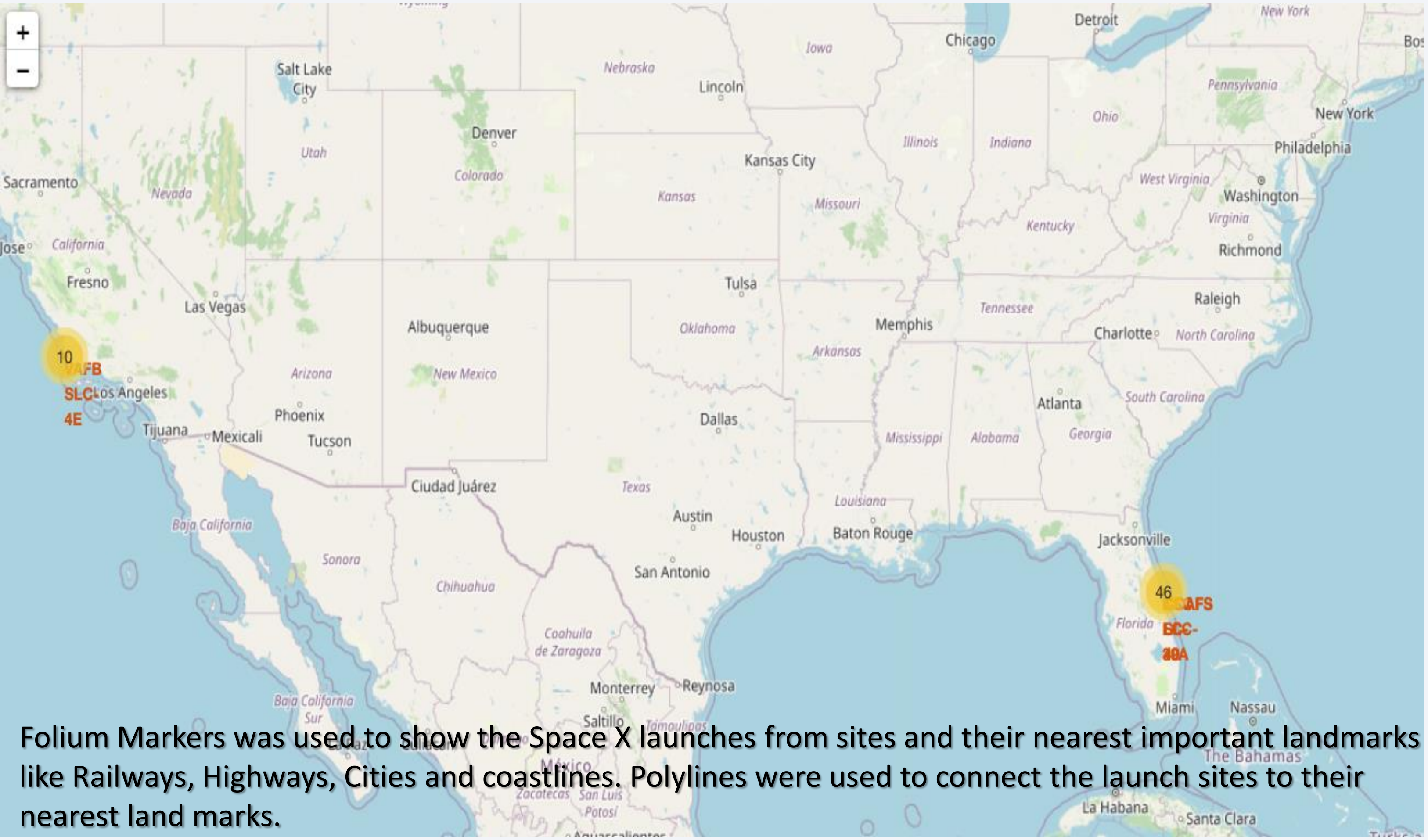
- [IBM Watson Studio Link]
- [GitHub Jupyter Notebook Link]

EDA with SQL

- [IBM Watson Studio JupyterNotebook Link]
- [Github JupyterNotebook Link]

- %load_ext sql
  - Load the sql and build a connection
- %%sql
  - Magic command for sql in python
- Where: extract or filter conditions
- Like: character string match detection
- Group by: summary rows
- Between: select values within a range
- Order: sort the result in order

Build an Interactive Map with Folium



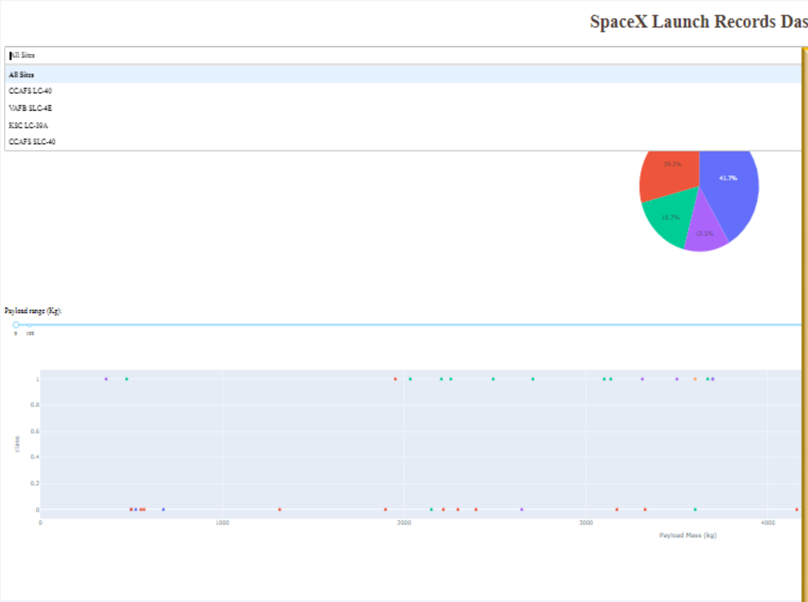
Folium Markers was used to show the Space X launches from sites and their nearest important landmarks like Railways, Highways, Cities and coastlines. Polylines were used to connect the launch sites to their nearest land marks.

Red represents rocket launch failures

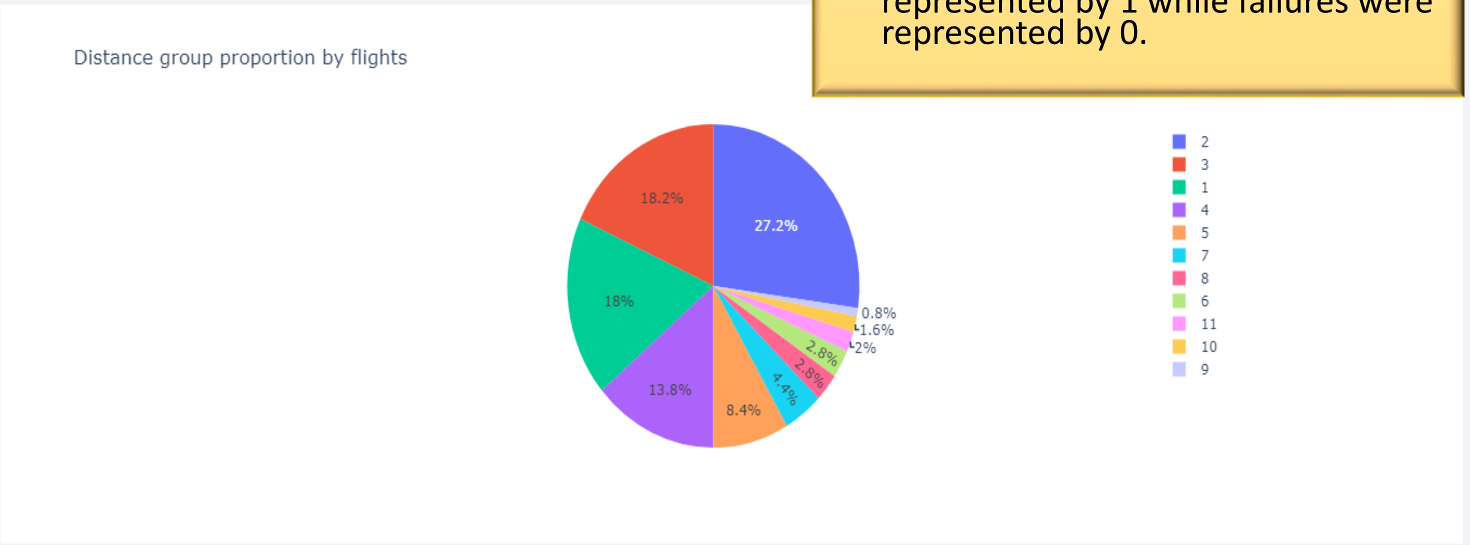
Green represents the successes

- Summary of map objects:
  - Markers: show a geo location from **lat** and **lon** data
  - Cluster: show a group of markers
  - Circles: show a single location
  - Lines: show distance between two markers
- [IBM Watson Studio JupyterNotebook Link]()
- [GitHub Jupyter Notebook Link]()

Build a Dashboard with Plotly Dash

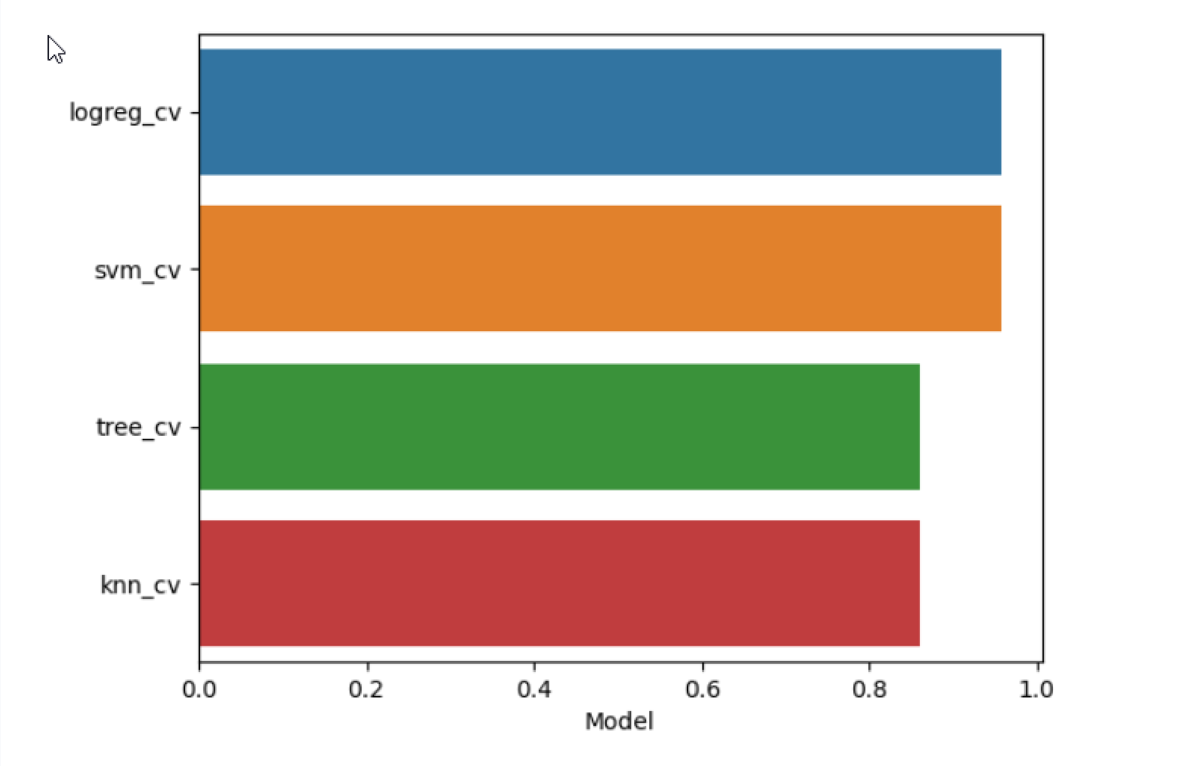


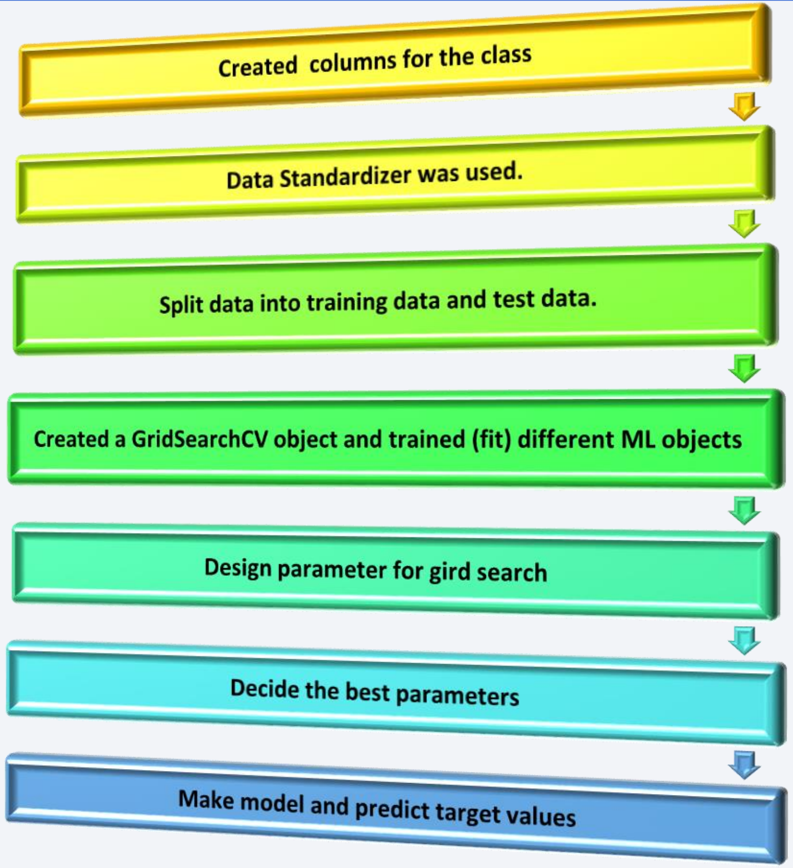
- Summary of plots:
- Pie charts and scatter charts were used to visualize the launch records of Space X.
- • These charts displayed the rocket launch success rate per launch site. We are were able to get
- an understanding of the factors that may have been influencing the success rate at each site.
- Such as the payload mass and booster versions.
- • Successful launches were represented by 1 while failures were represented by 0.



- **Built Dashboard App Development with Python Library Plotly Dash and Pandas Dash**
- Built Dashboard app for users to perform interactive visual analytics on SpaceX launch data in real-time
- App contains input **drop-down** components for launch site and also **range slider** for payload selection
- An operational **callback** function for rendering success Pie chart and payload scatter chart scatter plot.
- This Dashboard was completed and lunched on a dedicated computer on a cloud IDE exclusively.
- Output to be the graph with success pie chart
- IBM Watson Studio Jupyter Notebook Link
- GitHub Jupyter Notebook Link

Predictive Analysis (Classification)





- The models are trained by the spited data sets, the optimal parameters are decided by applying grid searching method, and evaluated by scores. Scikit-learn is Machine Learning library that was used for predictive analysis. • Created a machine learning pipeline to predict if the first stage will land given the data. KSC LC-39A has the highest successful rate.
- • The CCAFS SLC-40
- Key findings: The most successful launch sites are CCAFS SLC-40 and KSC LC-39A

- IBM Watson Studio JupyterNotebook Link
- GitHub URL link

Results

Exploratory data analysis results

- Both API and web scraping are capable to collect Space X data
- successful landing outcomes are somewhat correlated with flight number and also with significant increase since 2015

Interactive analytics demo in screenshots

- EDA with sql is effective for data filtering
- EDA with interactive visualization provides informative information
- Plotly Dash is powerful to show instant data change

Predictive analysis results

- The logistical regression model has the best accuracy of predicting.

Proximities of launch sites and distance

- All launch sites are located near the coast line. Perhaps, a hallmark to test rocket landings in the water.
- Sites are also located near highways and railways. This may facilitate transportation of equipment and research material.
- **The machine learning used were able to predict the landing success of rockets with an accuracy score of 83.33%.**

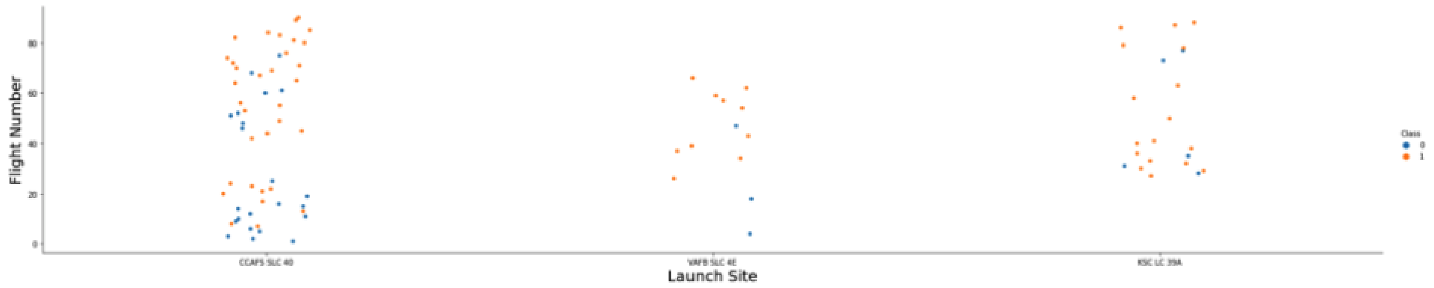
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- [screenshot] Shows a scatter piont chart of Flight Number vs. Launch Site

- It appears that there were more successful landings as the flight numbers increased. launch site **CCAFS SLC 40** had the most number of landing.



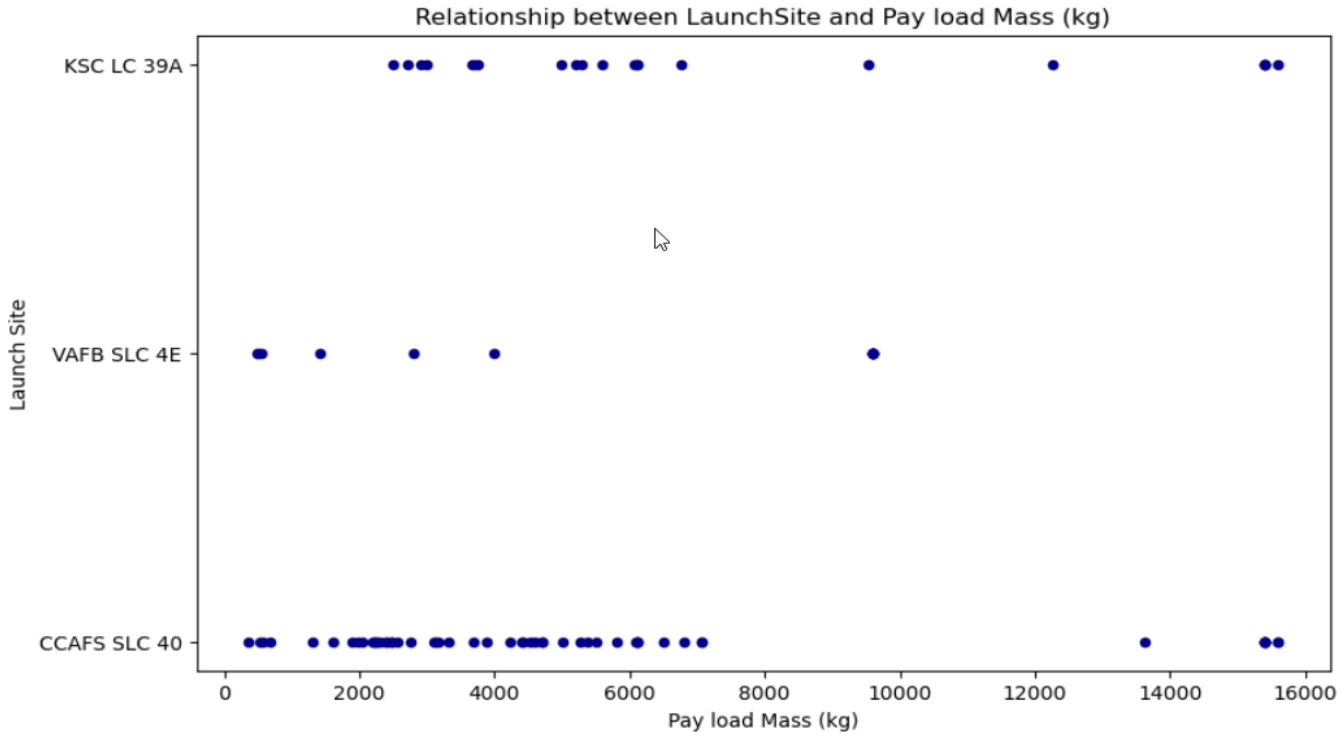
Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

- CCAFS SLC 40 seems to be the oldest launch platform and also lasted for a long time to support the new launches. It is the main launch site, and the failure rate decrease gradually.
- Launches at WAFB SLC 40 failed at the beginning.
- KSC has a low failure rate.

Payload vs. Launch Site

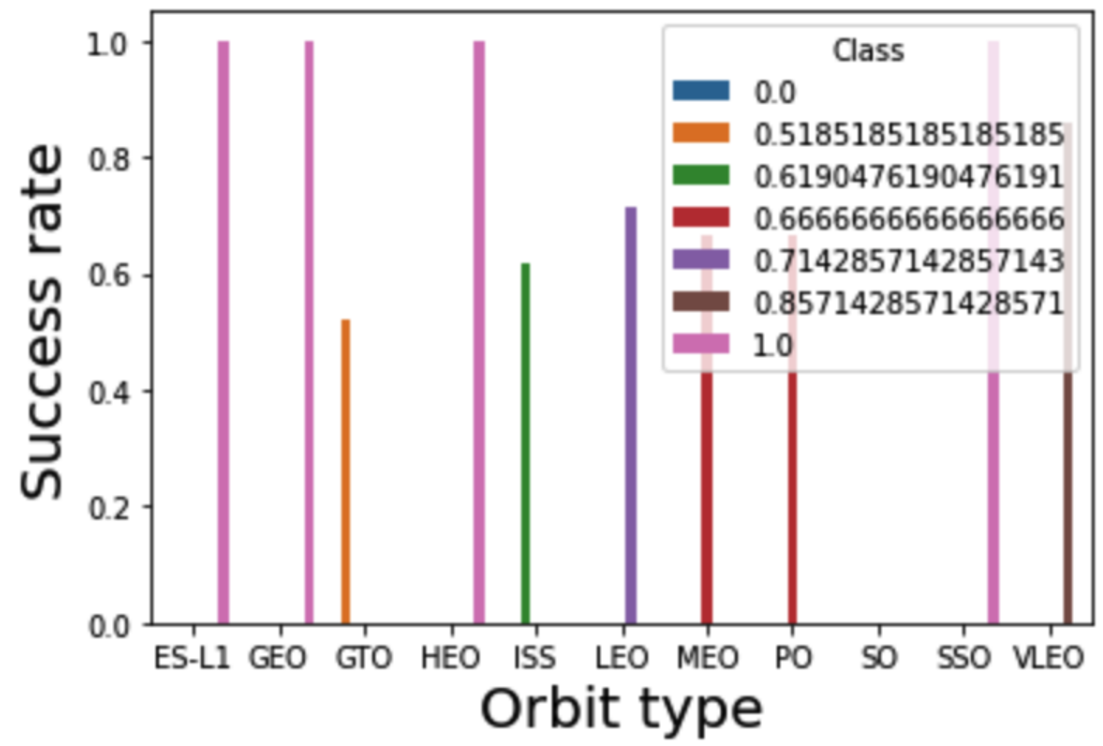
- [Shows a scatter point chart] of Payload vs. Launch Site

- Now if you observe the scatter point chart, you will find out that for the VAFB-SLC launch site, there are no rockets launched for heavy payload mass(greater than 10000).



- CCAFS SLC 40 seems to be a predominant launch site, most small-medium mass level load mass launches are located here.
- WAFB SLC 40 seem to ba small mass launch platform, but quite important as it provides medium mass launches, the failure rate here is relatively low.

Success Rate vs. Orbit Type



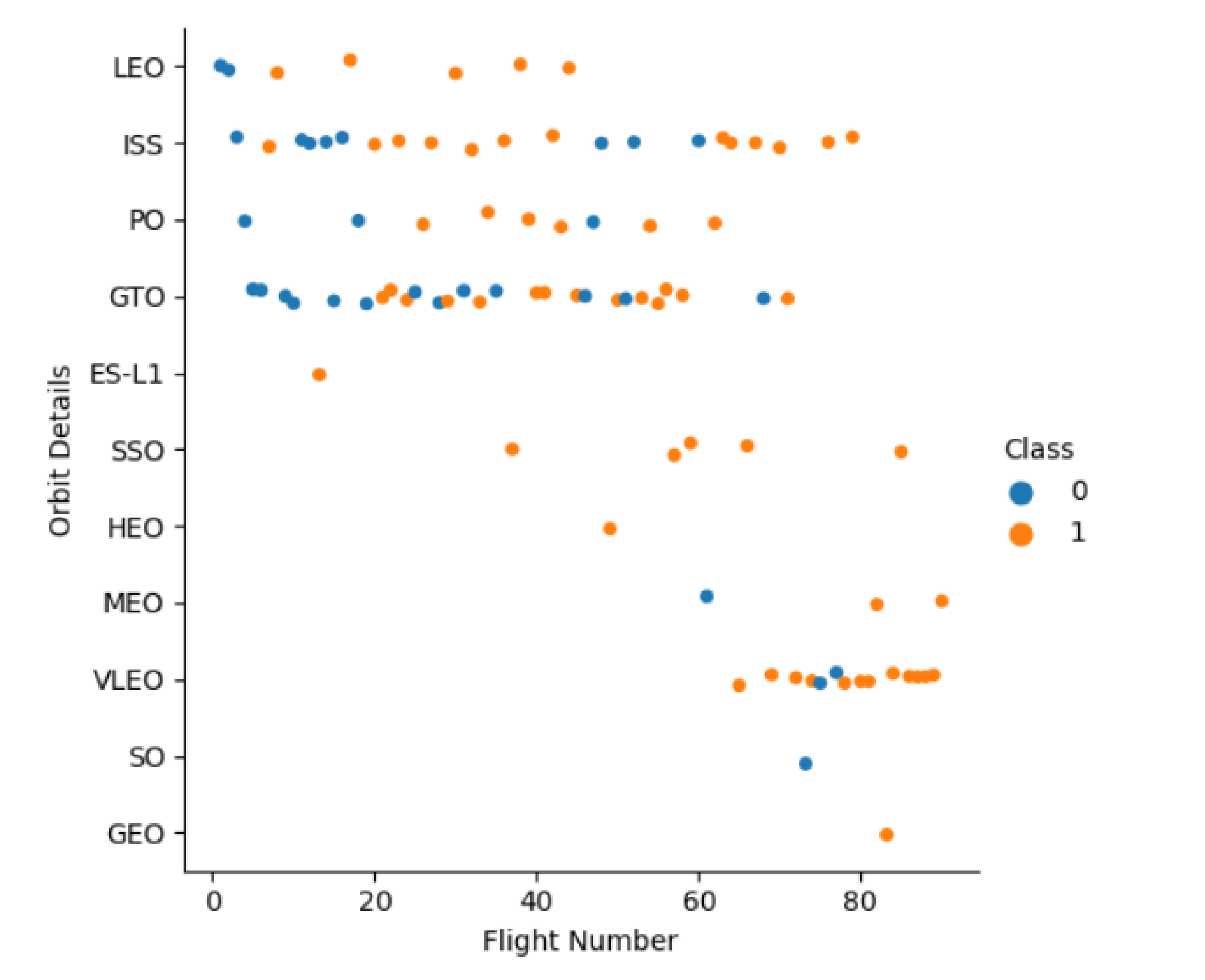
Analyze the ploted bar chart try to find which orbits have high sucess rate.

- Low success rate: GTO, ISS, MEO, PO and LEO
- High success rate: ES-L1, GEO, HEO, SSO

- Show a bar chart for the success rate of each orbit type

- The highest success rate ORBITS are
- **ES-L1**
- **GEO**
- **SSO**
- **HEO**

Flight Number vs. Orbit Type

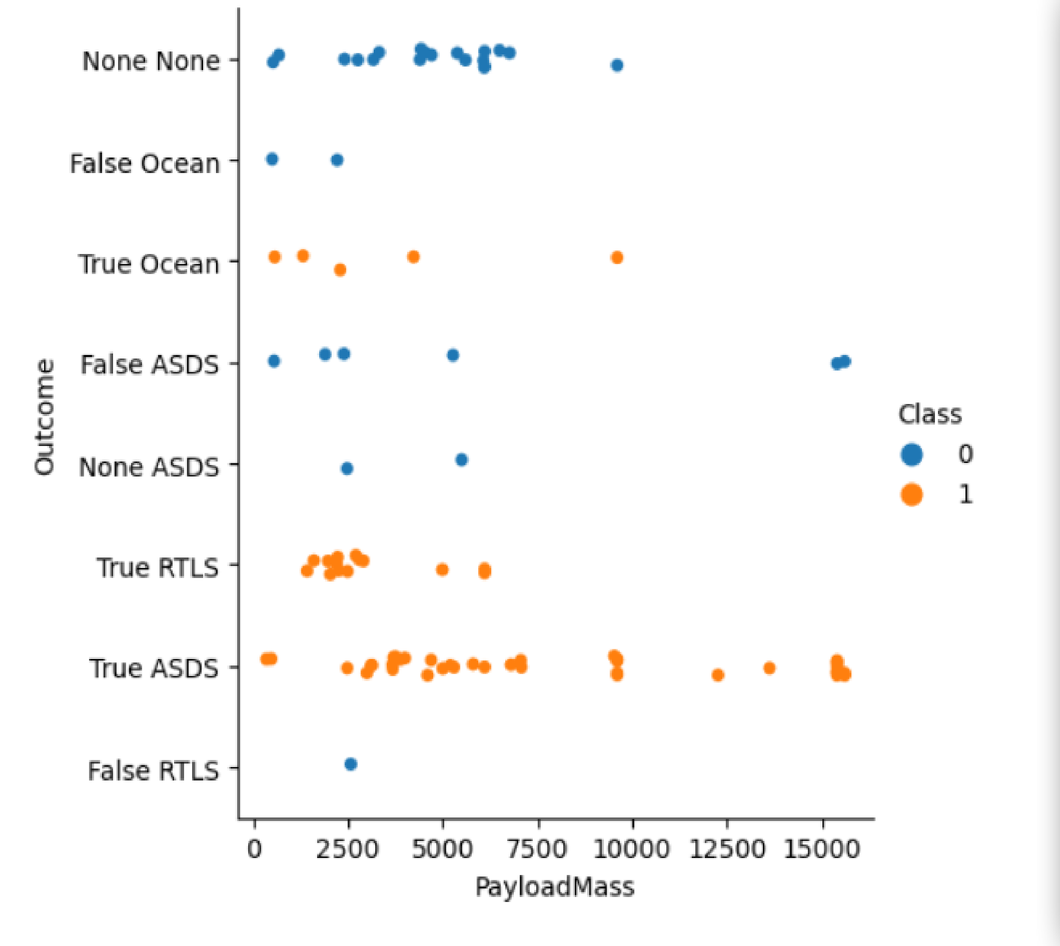


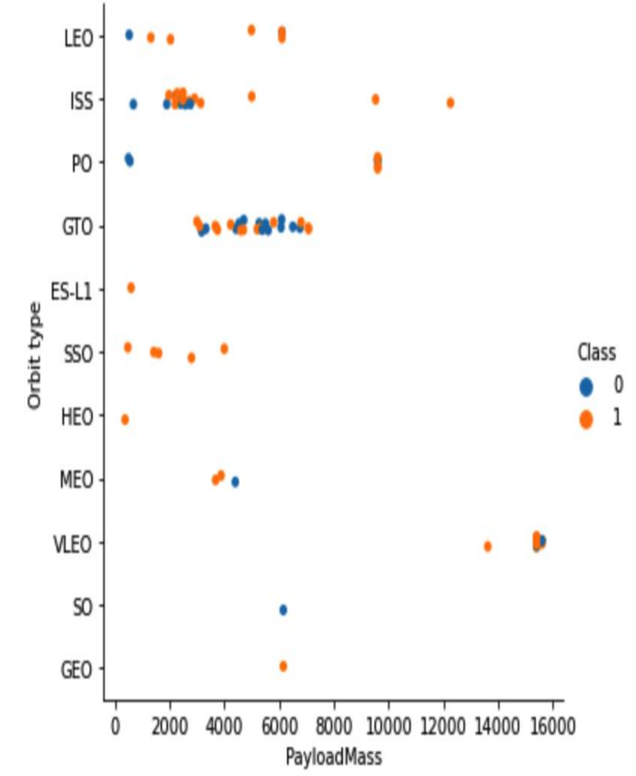
- [Scatter Point Chart] Shows a scatter point of Flight number vs. Orbit type
- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

- Shows a scatter point chart of payload vs. orbit type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here

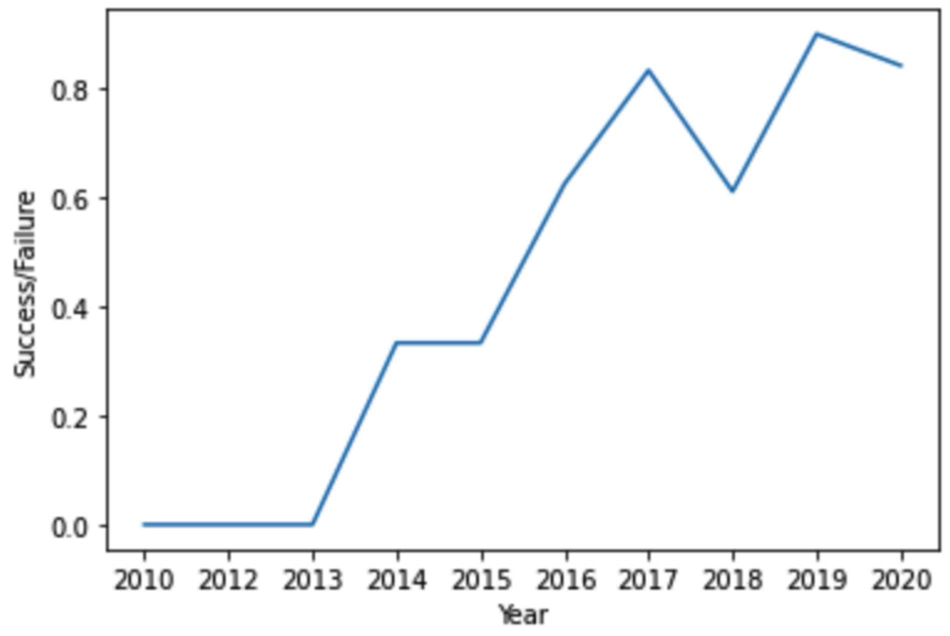




You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- *Plotted a line chart with success or Failure axis and Year axis to be the extracted year for average success rate*
- *The aim was* to get the average launch success trend.



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- The names of the unique launch sites
  - There are four launch sites, CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E.

In [8]:
```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL
```

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

Out[8]:

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

Launch Site Names Begin with 'CCA'

- Displays 5 records where launch sites begin with the string `CCA`
- My short query result analysis is that:
  - They are all CCAFS LC-40.
  - Space X CRS successes attempted no parachute landing outcomes or any landing attempts

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

In [9]:
```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE "CCA%" LIMIT 5
```

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

Out[9]:

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

Total Payload Mass

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

```
In [10]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = "NASA (CRS)"
 * sqlite:///my_data1.db
Done.
```

Out[10]:

SUM(PAYLOAD_MASS__KG_)
45596

- **THE TOTAL PAYLOAD CARRIED BY BOOSTERS FROM NASA:**
- The information in the picture displays the total payload mass carried by boosters launched by NASA
  - Well, um… that's the idea …

Average Payload Mass by F9 v1.1

```
In [11]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE "F9 v1.1%"
 * sqlite:///my_data1.db
Done.
Out[11]:
```

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

- Calculates the average payload mass carried by booster version F9 v1.1
- The average payload mass is 2534.6 KG
  - An average of about 3000 kg, quite small mass.

First Successful Ground Landing Date

List the date when the first succesful landing outcome

Hint:Use min function

In [12]:
```
%%sql
SELECT MIN("Date")
FROM SPACEXTBL
WHERE
    "Landing _Outcome" = 'Success'
```

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

Out[12]:

MIN("Date")
02-03-2019

FIRST SUCCESSFUL LANDING DATE
(DRONE SHIP)

In [37]:
```
%%sql
select min(DATE),
from SPACEXTBL
where landing__outcome like 'Success (g%';
```

```
 * ibm_db_sa://zgr48431:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[37]:

FIRST SUCCESSFUL GROUND LANDING DATE
(GROUND PAD)

1
2015-12-22

- Found the date(s) of the first successful landing outcome on ground pad EVEN drone ship from an SQL queries.

The first successful ground landing date was 2015-12-22.

- First successful landing after 5 years attempts, LOL!!

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [21]: %sql SELECT Booster_Version,Customer FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
 * sqlite:///my_data1.db
Done.
```

Out[21]:

Booster_Version	Customer
F9 v1.1	AsiaSat
F9 v1.1 B1011	AsiaSat
F9 v1.1 B1014	ABS Eutelsat
F9 v1.1 B1016	Turkmenistan National Space Agency
F9 FT B1020	SES
F9 FT B1022	SKY Perfect JSAT Group
F9 FT B1026	SKY Perfect JSAT Group
F9 FT B1030	EchoStar
F9 FT B1021.2	SES
F9 FT B1032.1	NRO
F9 B4 B1040.1	U.S. Air Force
F9 FT B1031.2	SES EchoStar
F9 B4 B1043.1	Northrop Grumman
F9 FT B1032.2	SES
F9 B4 B1040.2	SES
F9 B5 B1046.2	Telkom Indonesia
F9 B5 B1047.2	Es hailSat
F9 B5 B1046.3	Spaceflight Industries
F9 B5B1054	USAF
F9 B5 B1048.3	PSN, SpaceIL / IAI
F9 B5 B1051.2	Canadian Space Agency (CSA)
F9 B5B1060.1	U.S. Space Force

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- There are 24 names of the boosters but different versions which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - The version code seem to be continued

Total Number of Successful and Failure Mission Outcomes

Double check for missing Successful Mission Outcomes, Success payload status unclear

```
In [18]: %sql SELECT DISTINCT(COUNT(Mission_Outcome)) FROM SPACEXTBL WHERE Mission_Outcome = "Success (payload status unclear)"
 * sqlite:///my_data1.db
Done.
Out[18]:
```

(COUNT(Mission_Outcome))
1

Double Check for successful misssion outcomes

```
In [19]: %sql SELECT COUNT(Mission_Outcome) FROM SPACEXTBL WHERE Mission_Outcome = "Success"
 * sqlite:///my_data1.db
Done.
Out[19]:
```

COUNT(Mission_Outcome)
98

Double Check for failure mission outcomes

```
In [20]: %sql SELECT COUNT(Mission_Outcome) FROM SPACEXTBL WHERE Mission_Outcome Like "Fail%"
 * sqlite:///my_data1.db
Done.
Out[20]:
```

COUNT(Mission_Outcome)
1

```
In [49]:
%%sql
select mission_outcome, count(mission_outcome)
from SPACEXTBL
group by mission_outcome;

 * ibm_db_sa://zgr48431:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[49]:
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- Mission outcomes are preforming well
- The Above screenshot show the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
In [23]: %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) from spacextbl)

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

Out[23]:

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

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here
  - The version code seem to be continued

2015 Launch Records

- **The records which displays the month names as substr(Date,4,2), with integer value, Failure landing_outcomes in drone ship, Booster versions, Launch_Site name for the months in year 2015.**

- **However, there is no record of F9 v1.1 B1017 'Landing Outcome' or any of its booster being lunched in any unique Launch sites.**
  - **Maybe drone ship is problem before improvement**

In [24]:
```
%%sql
SELECT
    "Landing _Outcome",
    Booster_Version,
    Launch_Site,
    substr(Date,4,2)
FROM SPACEXTBL
WHERE
    substr(Date,7,4) = '2015'
```

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

Out[24]:

Landing _Outcome	Booster_Version	Launch_Site	substr(Date,4,2)
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	01
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40	02
No attempt	F9 v1.1 B1014	CCAFS LC-40	03
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04
No attempt	F9 v1.1 B1016	CCAFS LC-40	04
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40	06
Success (ground pad)	F9 FT B1019	CCAFS LC-40	12

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

- Success outcome sits on top of the blockchain as 'No attempt' outcomes follows. The droneship and ground pad success landing outcomes are runners up while failure (parachute) and controlled ocean landing sits on the bottom on the ranks between the date 2010-06-04 and 2017-03-20, in descending order
  - Most failure cases are on drone ship recently.

In [25]:

```
%%sql
SELECT
    "Landing _Outcome",
    COUNT("Landing _Outcome"),
    RANK() OVER(ORDER BY COUNT("Landing _Outcome") DESC) as ranking
FROM SPACEXTBL
WHERE
    Date BETWEEN '04-06-2010' AND '20-03-2017'
GROUP BY
    1
ORDER BY
    2 DESC
```

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

Out[25]:

Landing _Outcome	COUNT("Landing _Outcome")	ranking
Success	20	1
No attempt	10	2
Success (drone ship)	8	3
Success (ground pad)	6	4
Failure (drone ship)	4	5
Failure	3	6
Controlled (ocean)	3	6
Failure (parachute)	2	8
No attempt	1	9

Section 4

Launch Sites Proximities Analysis

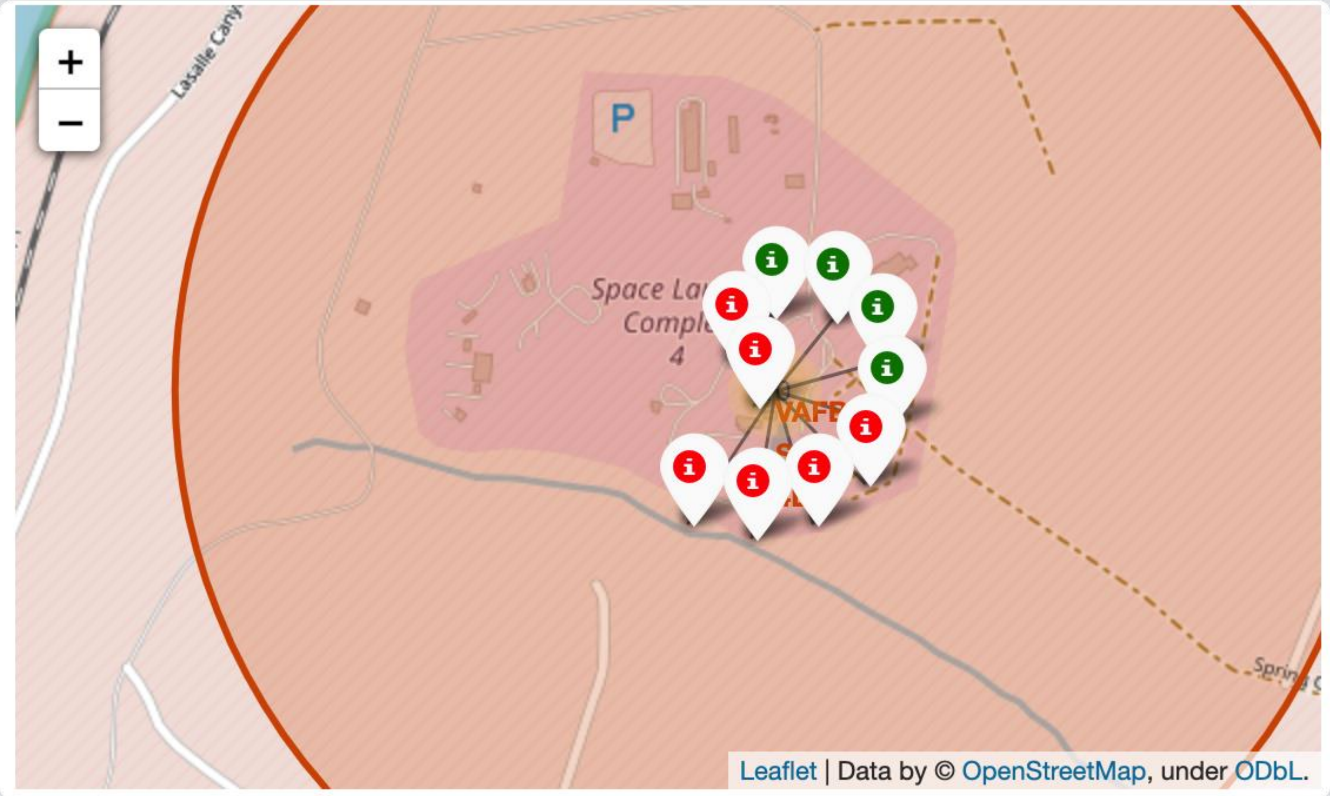
Best launch sites

Both ground and sea surface sites are necessary, also, south areas maybe a proper areas.

- The transportation base chosen maybe important
- All launch sites are in very close proximity to the coast and they are also a couple thousand kilometers away from the equator line.

Red vs. Green

- Markers was used to show the Space X launches from sites and their nearest important landmarks like Railways, Highways, Cities and coastlines. Polylines were used to connect the launch sites to their nearest land marks.
- Red represents rocket launch failures
- Green represents the successes
    - Color icons is impressive way to show the rate of success, in VABF, failure rate is high.



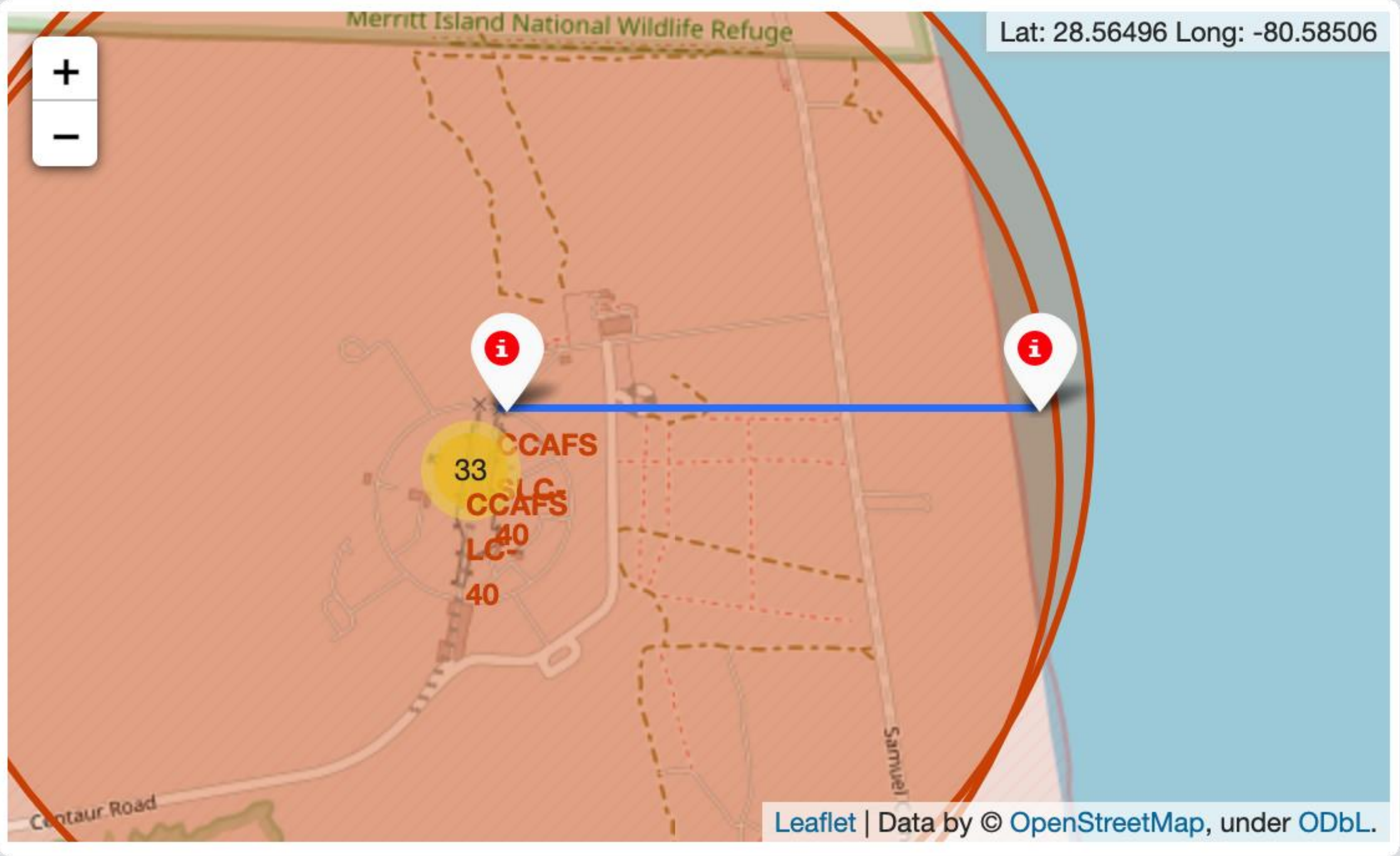
Transportation

Visualization shows importance of transportation, for example the railway, coast or port.

It appears that launch sites are usually set up at least 18 km away from cities. This may be because of the desire to prevent any crashes near populated areas.

- It is also apparent that launch sites are in very close proximity to railways and highways. Perhaps, due to the necessary transportation requirements for rocket parts.

- The sites are close the coast line. This is evident with the many rocket landing tests on water bodies like the ocean.

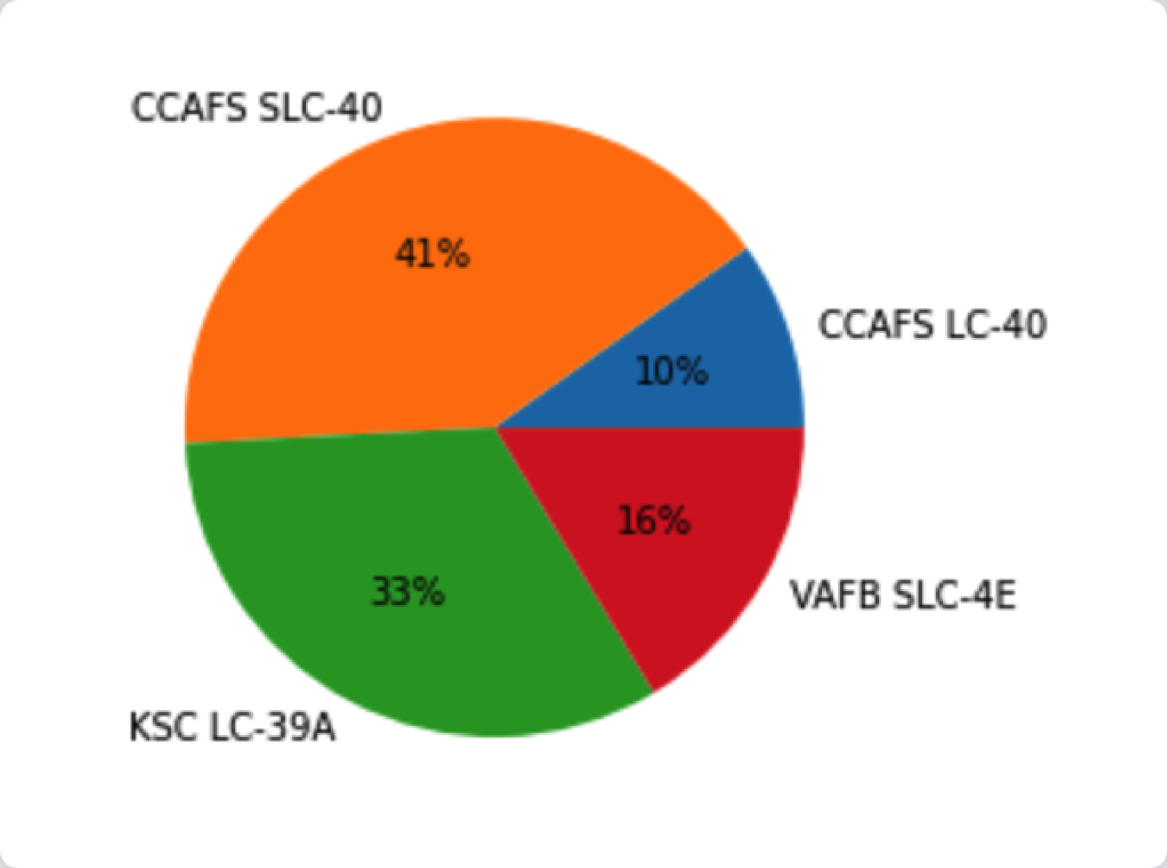


Section 5

Build a Dashboard with Plotly Dash

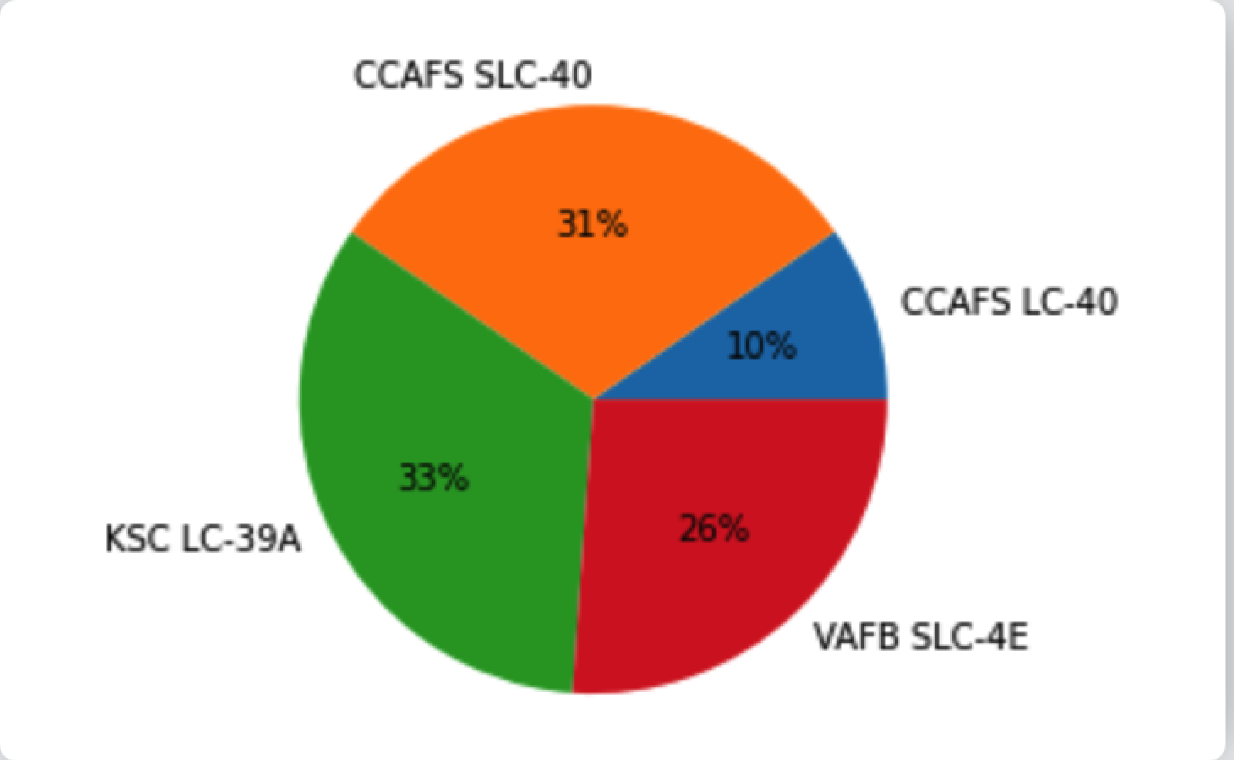
The most successful launch site

- Key findings: The most successful launch sites are CCAFS SLC-40 and KSC LC-39A



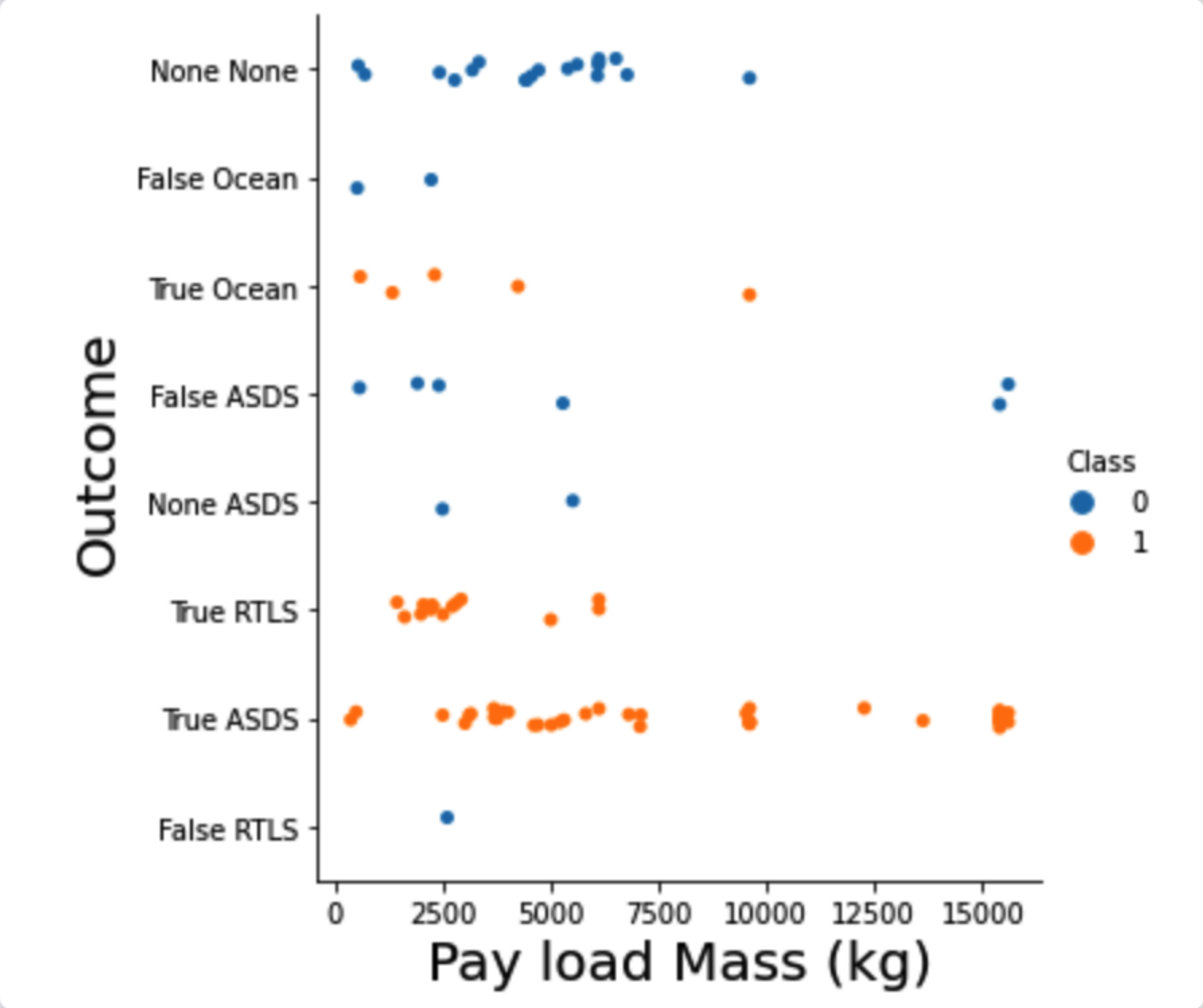
The most feasible launch site

- KSC LC-39A has the highest successful rate.
    - The CCAFS SLC-40 follows.



Promising launch sites

- One key point:
  - The ocean & RTLS outcome has high successful rate for the small and medium mass launch



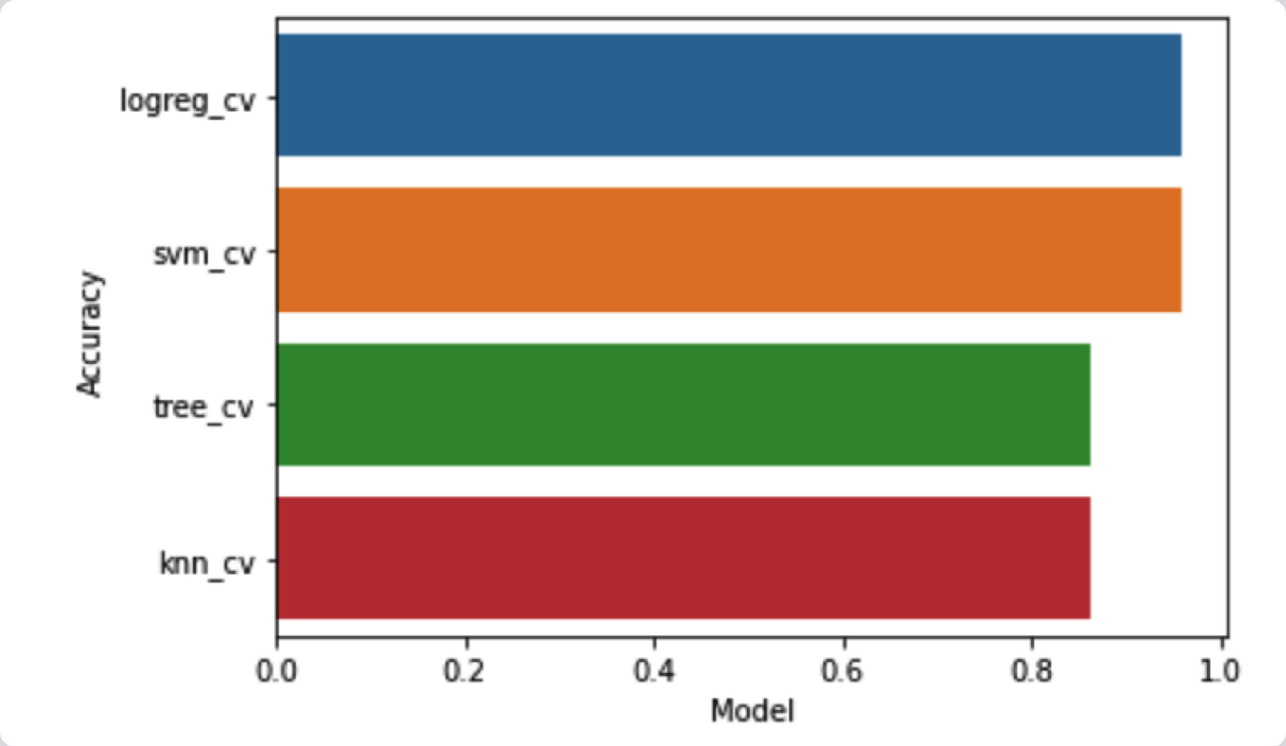
Section 6

Predictive Analysis (Classification)

Classification Accuracy

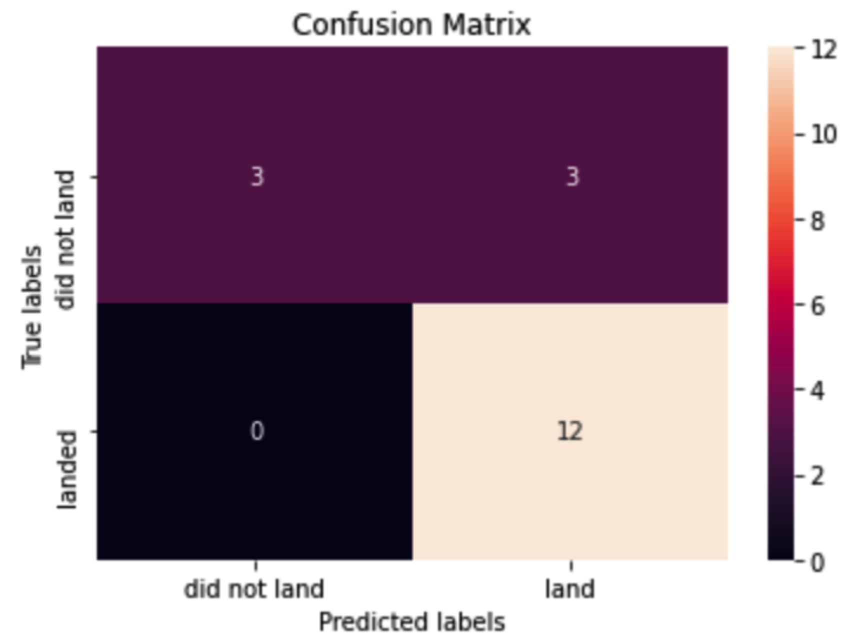
- I visualized the built model accuracy for all built classification models, in a bar chart and from observation, Simple Vector Machine (svm_cv) computer vision and Logistic Regression performed better in classification modelling.

    - Logistic Regression model performed best and has earned the highest classification accuracy because it can distinguish among other models, especially classes.



Confusion Matrix

- Shows the confusion matrix of the best performing model with predicted labels and True Lables
- The chart shows the confusion matrix of the Logistic Regression model that was chosen.
- • However, the model only failed to accurately predict 3 labels.



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

Conclusions

- Point 1: Launch site has a strong relation with success rate.
- Point 2: CCAFS SLC-40 and KSC LC-39A are most feasible sites with highest success rate and amount of launch.
- Point 3: The ocean & RTLS outcome have succussed on small and medium mass launches, and there is more chances for improvement.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

- **SpaceX Falcon 9 first stage Landing Prediction- Data Collection**
- **Space X Falcon 9 First Stage Landing Prediction – Web Scraping**
- **Complete the EDA with SQL lab – Database Snippets and SQL queries**
- **Complete the EDA with Visualization lab - Prepared Data Feature Engineering**
- **Launch Sites Locations Analysis with Folium – Lunch site distances & proximities**
- **Space X Falcon 9 First Stage Landing Prediction – ML best classifier**

Thank you!