IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

R.Sushmitha
31-08-2025

Outline

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

Executive Summary

Summary of methodologies	Summary of all results
1.Data collection from API and WebScraping	1.The best Hyperparameters for Logistic Regression,SVM,Decision Tree,KNN classifier
2.Data Wrangling and Predictive Analysis Classification	2.The methods that performs best using test data
3.Exploratory Data Analysis(EDA) using SQL,Pandas,Matplotlib	
4.Interactive Visual Analytics and Dashboard with Folium and Plotly Dash	

Introduction

- **Project background and context**
- SPACE X is here to compete in the commercial space race. we are making rocket launches relatively inexpensive for everyone
- **Problems you want to find answers**
- SPACE X can save millions in every launches of our eagle rocket because we can reuse its first stage
- In addition, we can determine if the first stage of our competitor will land and determine the cost of the launch by using Data Science and Machine Learning models

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**

  The data was gathered from SpaceX RESTAPI and WebScraping from wiki pages

- **Perform data wrangling**

  The data is collected in the form of json object and HTML tables,after that the data is converted into pandas dataframe for visualisation and analysis

- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**

  Use of machine learning to determine if the first stage of FALCON 9 will land successfully.

Data Collection

- Data collection process involves a combination of API request and SPACEX REST API and WebScraping data from a table in SPACEX's wikipedia entry.
- We have to use both of these data collection methods in order to get complete information about the launches for the more detailed analysis.
- **Data Columns obtained by using SPACEX REST API:**
- Flightno,Date,Boosterversion,PayloadMass,Orbit,LaunchSite,Outcomes,Flights,Gridfins,Reused,Legs,
- LandingPad,Blocked,ReusedCounts,Serial,Longitute,Latitude
- **Data Columns obtained by using wikipedia WebScraping:**
- Flightno.,PayloadMass,Payload,LaunchSite,Orbit,Customer,LaunchOutcome,VersionBoooster,
- BoosterLanding,Date,Time.

Data Collection – SpaceX API

Requesting Rocket Data Launch from SpaceX APi

→ Decoding the response content using .json() and turning it into a Dataframe using .json_normalize()

→ Requesting Needed information about launches from SPaceX API by applying custom function

→ Constructing data we have obtained into a dictionary

→ Creating a Dataframe using dictionary

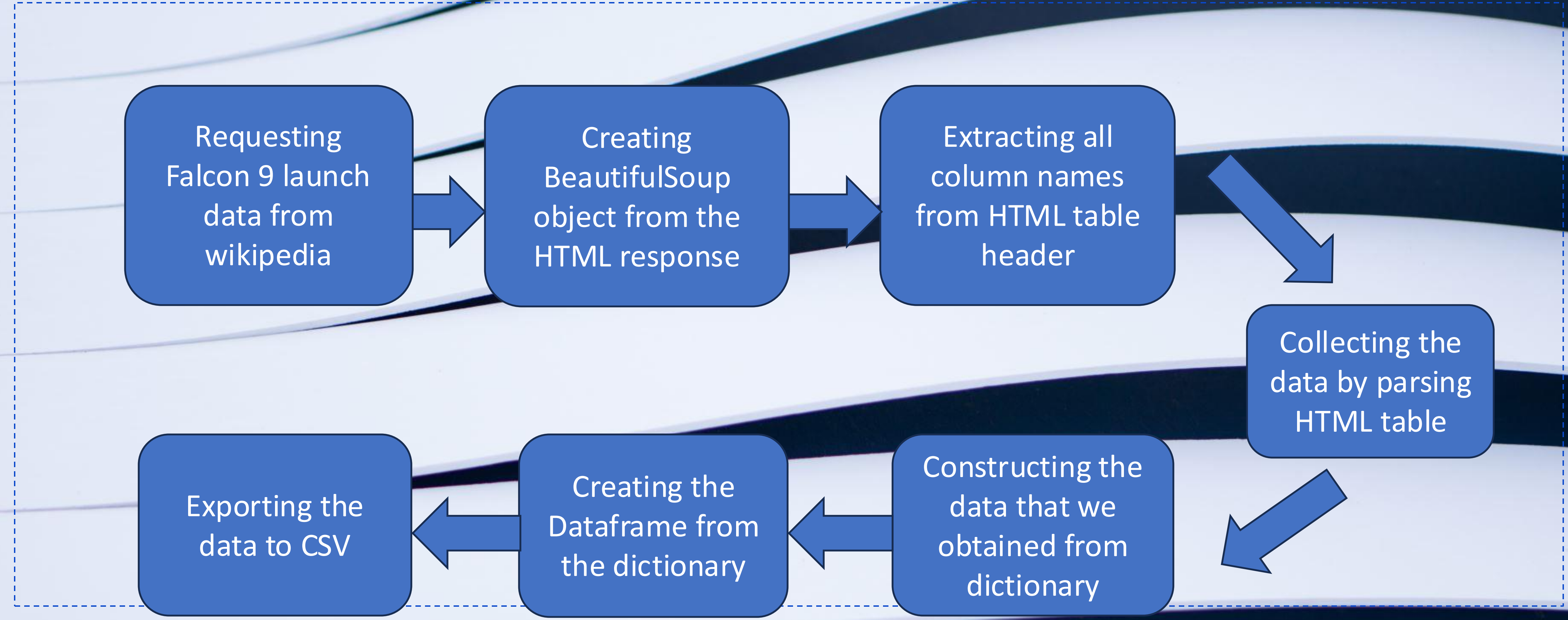
→ Filtering the DataFrame to only include falcon 9 launches

→ Replacing Missing values of PayloadMass columns with calculated .mean() for this column

→ Exporting the data to the CSV

[jupyter-labs-spacex-data-collection-api.ipynb]()

Data Collection - Scraping



jupyter-labs-webscraping.ipynb

Data Wrangling

- In the Data set there are several different cases,where the booster did not land successfully.sometimes the landing was attempted but failed due to an accident,for example,true ocean means the meachine outcome is successfully landed to a specific regionof the ocean while ,False means mission Outcome is unsuccessfully landed to a specific region of ocean.True RTLS means the outcome is sucessfully landed to a ground Pad,False RTLS means the mission outcome is unsuccessfully landed to a ground pad.True ASDS means the mission Outcome is sucessfully landed on a Drone ship.False ASDS means the mission outcome is unsucessfully landed on a Droneship.
- We mainly convert those Outcomes into Traning labels with '1' means booster sucessfully landed.'0' means unsucessfull.

Data Wrangling Flowchart

Perform Explorary data analysis and determine Training labels

Calculate the number of launches on each site

Calculate the number and occurrence on each orbit

Calculate the number and occurrence of mission outcome per orbit type

Create a landing outcome label from outcome column

Exporting the data to csv

GITHUB URL :DataWrangling

EDA with Data Visualization

Charts were plotted are:

- FlightNumber vs PayloadMass
- FlightNumber vs LaunchSite
- PayloadMass vs LaunchSite
- Orbit Size vs SucessRate
- FlightNumber vs OrbitType
- PayloadMass vs OrbitType
- Success Rate Yearly Trend

Scatter Plots shows the relationship between variables.if a relationship exists,they could be used in machine learning models.

Bar charts shows comparisons among discrete category.The goal is to show that the relationship between the specific categories being compared and measured value.

Line chart show trends in data over time(time series).

GITHUB URL : EDA with Data Visualization

EDA with SQL

Peformed SQL Queries:

- Displaying the names of the unique launch sites in the space mission.
- Displaying 5 records where launch site begin with the string 'CCA'.
- Displaying total Payloadmass carried by Booster launched by NASA(CRS).
- Displaying average payloadmass by Booster version F9 v1.1.
- Listing the Date when the first successful landing outcome in ground pad was achieved.
- Listing the names of the boosters which has success in droneship and have Payloadmass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes.
- Listing the names of Booster version which have carried the maximum payloadmass.
- Listing the failed tle landing outcomes in droneship,there booster versions and launchsite names for the months in year 2015.
- Ranking the count of landing outcomes(such as failure (droneship)or success(groundpad)) between the date 2010-06-04 and 2017-03-20 in decending order.
- Github URL:EDA with SQL

Build an Interactive Map with Folium

Markers of all Launch sites:

- Added markers with circle,Popup label and Text label of NASA johnson space center using its Latitute and longitutde coordinates as start location.
- Added markers with circles,Popup label and Text label with all Launch Sites using there latitude and longitutde coordinates to show their geographical locations and proximity to equator and coasts.

Colored Markers of the Launch Outcomes for each Launch Site:

- Added colored markers of success(Green) and failed(Red) launches using Marker Cluster to identity which launch sites have relatively high success rate.

Distance between a Launch Sites and to its proximates:

- Added colored line to show the diatance between Launch sites KSC LC-39A(as an example) and its proximities like Railway ,Highway,Coastline and closest city.
- GITHUB URL: Intractive map with Folium

Build a Dashboard with Plotly Dash

Launch Sites Dropdown List:

- Added a Dropdown list to enable Launch site selection.

Pie Chart showing success Launches(All sitess/certain sites):

- Added a pie chart to show total successfull launches count for all the sites and the Success vs Failed counts for all the sites,if a specific launche site was selected.

Slider of PayloadMass range:

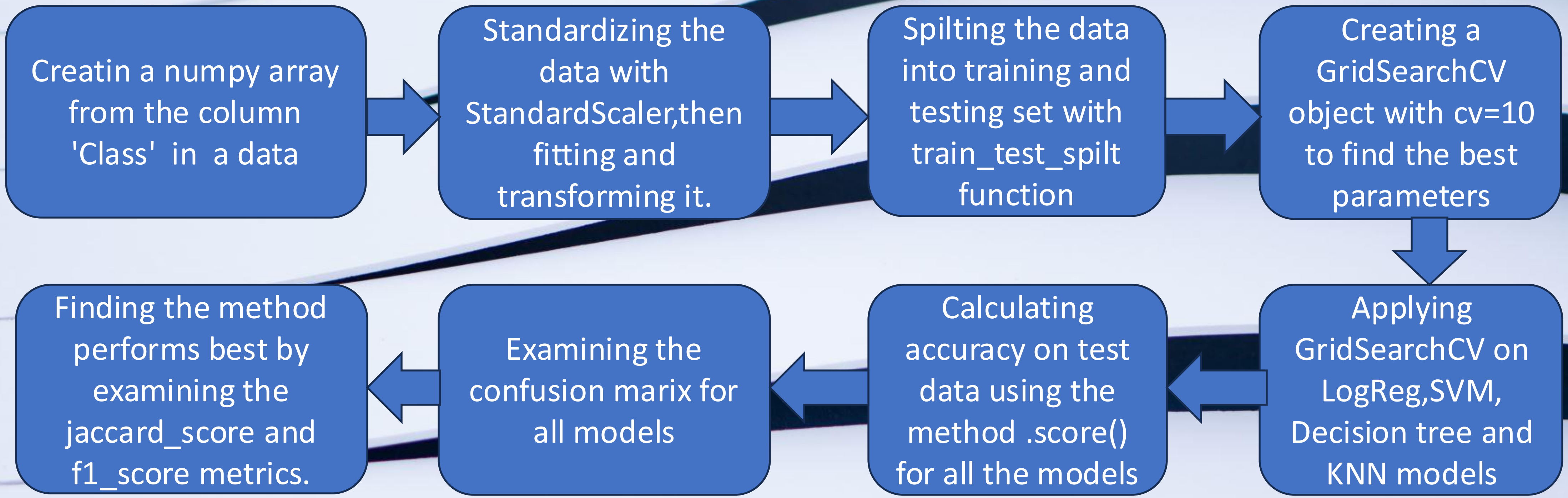
- Added a slider to select payload mass range.

Scatter chart of payload mass vs the Success rate for different Booster Version:

- Added a scatter chart to show the correlations between payload and launch success.

GITHUB URL:Dashboard with Plotly Dash

Predictive Analysis (Classification)



[Predictive Analysis GITHUB URL]

Results

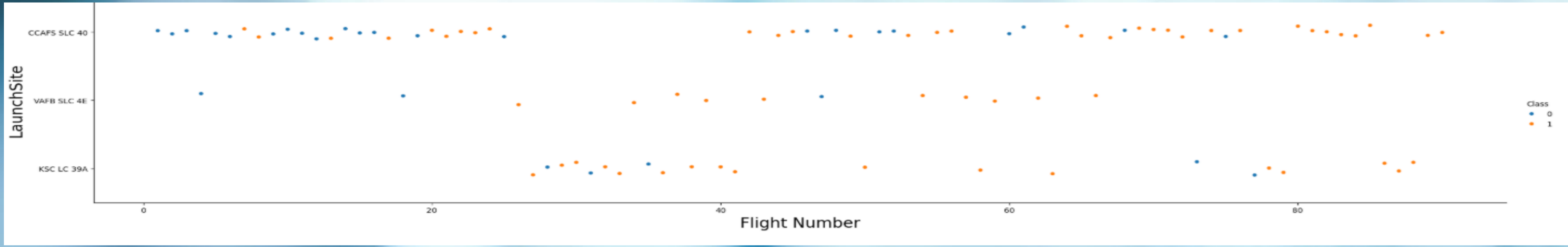
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Flight Number vs. Launch Site

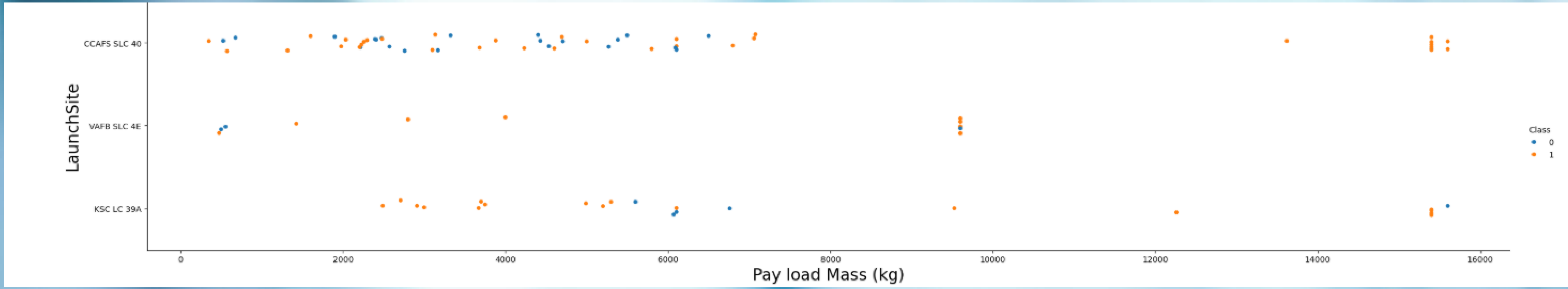


Explanation:

- The earliest flights all failed and the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have a higher success rate.
- It can be assumed that each new launches has the higher success rate.

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

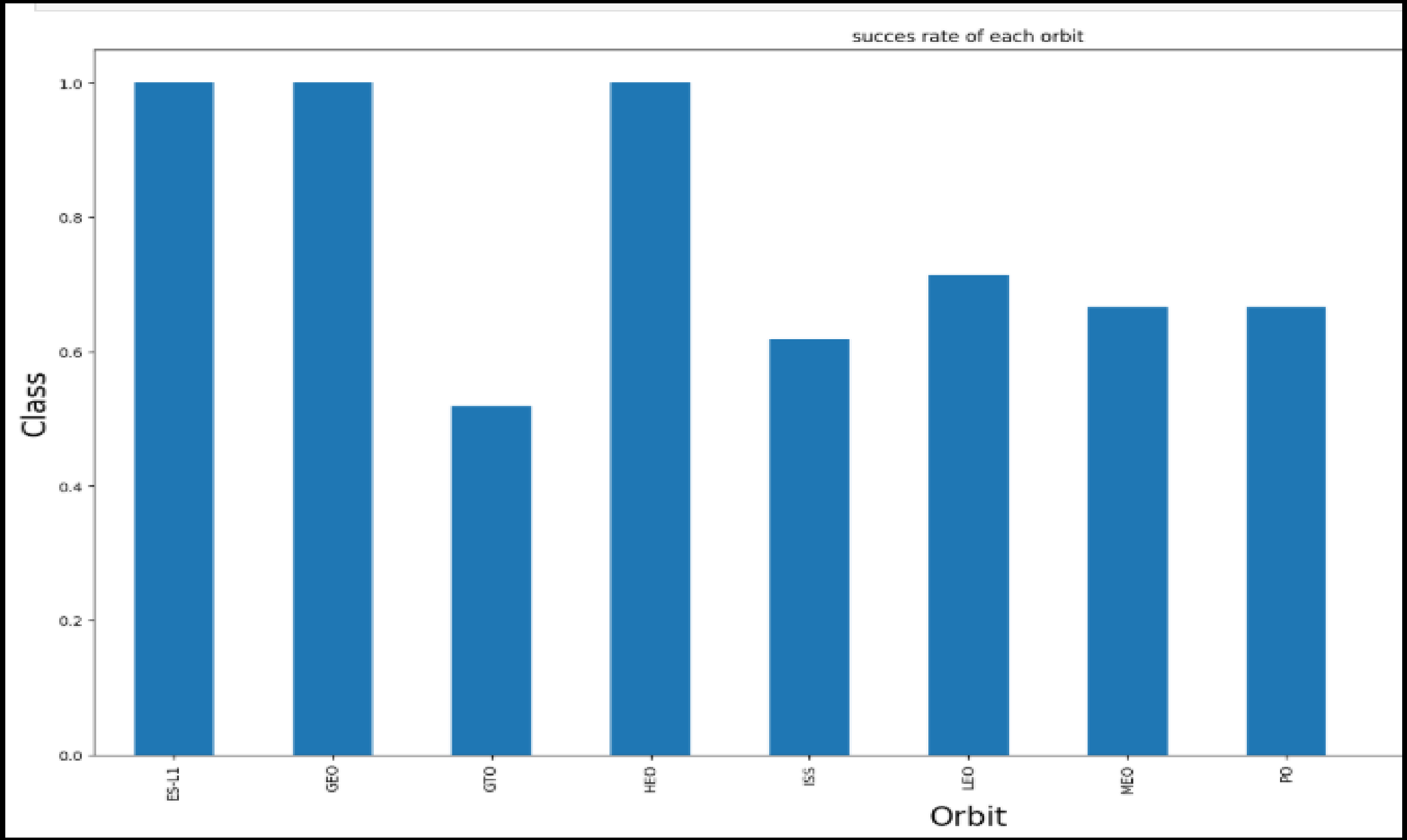


Explanation:

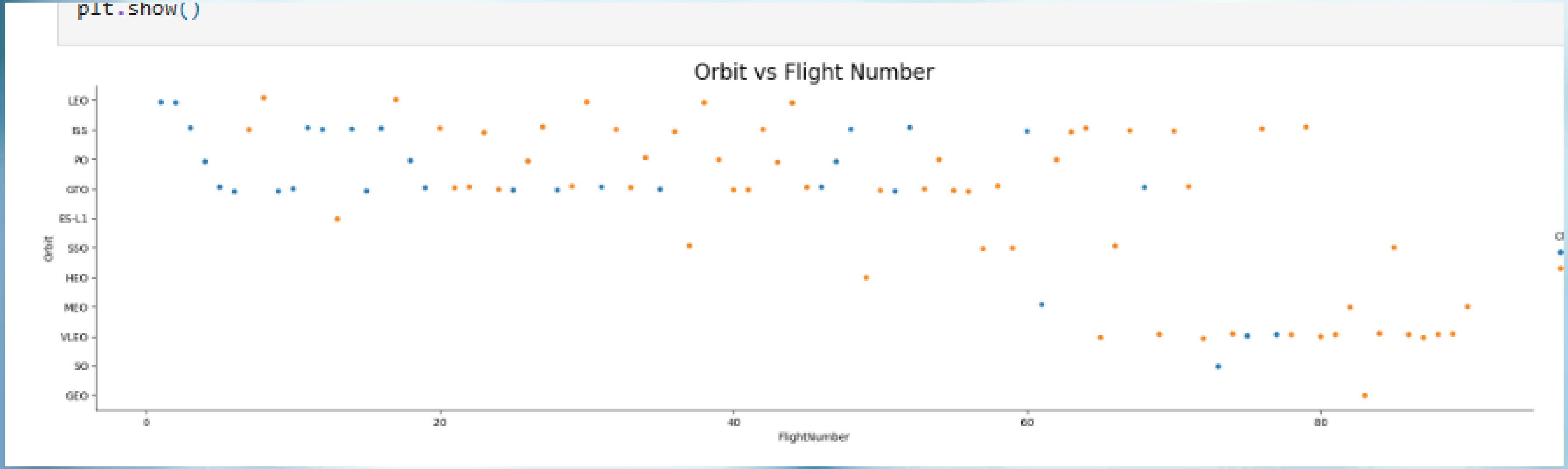
- For every Launch site the higher the payload mass,the higher the success rate.
- Most of the Launches with payload mass over 7000kg was successful.
- KSC LC 39A has a hundred percent success rate for payload mass under 5500kg too.

Success Rate vs. Orbit Type

- **Explanation:**
- Orbit with 100% success rate:
- - ES-L1,GEO,HEO,SSO
- Orbit with 0% success rate:
- - SO
- Orbits with Success rate between 50% and 85%:
- - GTO,ISS,LEO,MEO,PO



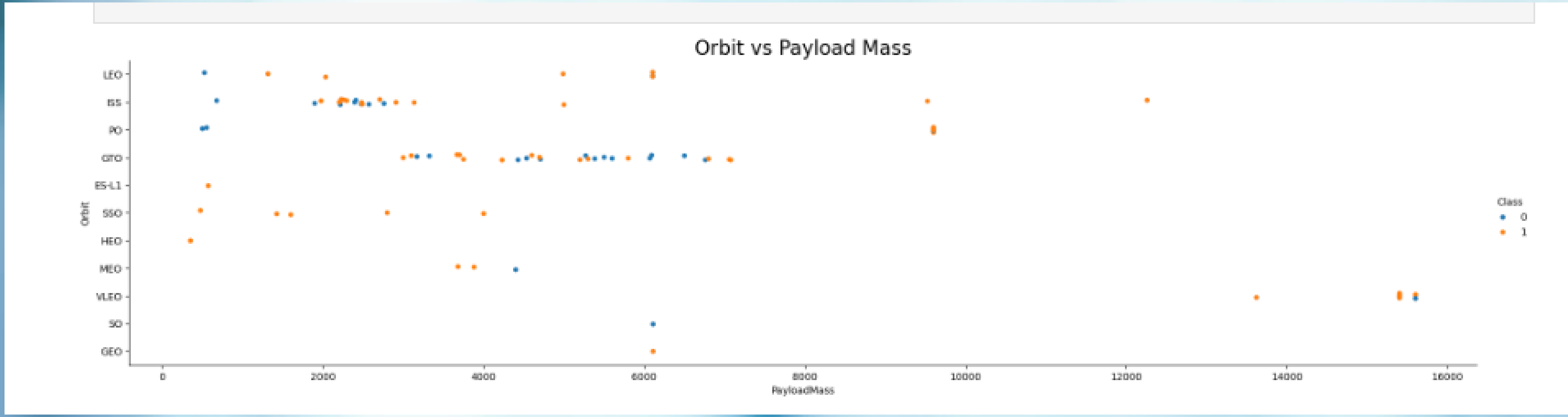
Flight Number vs. Orbit Type



Explanations:

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



Explanation:

Heavy Payload have a negative influence on GTO orbits and positive on GTO and polar LEO(ISS) orbits.

Launch Success Yearly Trend

Explanation:

The success rate since 2013,kept increasing till 2020.

All Launch Site Names

In [12]:
```
%sql select distinct Launch_Site from spacextable
```

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

Out[12]:

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

Explanation:

Displaying the names of the unique launch sites in the space mission.

Launch Site Names Begin with 'CCA'

Task 2

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

[13]:
```
%sql select * from spacextable where Launch_Site like 'CCA%' limit 5
```

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

[13]:

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

Explanation:

Displaying 5 records where launch site begins with the strings 'CCA'.

Total Payload Mass

Task 3

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

In [21]:
```
%sql select sum(PAYLOAD_MASS__KG_) as SUM from spacextable where customer like 'NASA (CRS)'
```

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

Out[21]:

SUM
45596

Explanation:

Displaying the Total Payload Mass carried by boosters launched by NASA(CRS).

Average Payload Mass by F9 v1.1

Task 4

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

[22]:
```
%sql select avg(PAYLOAD_MASS__KG_) as AVG from spacextable where Booster_Version like 'F9 v1.1'
```

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

[22]:

AVG
2928.4

Explanation:

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

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

In [25]:
```
%sql select min(date) as Date from spacextable where Landing_Outcome like '%ground pad%'
```

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

Out[25]:

Date
2015-12-22

Explanation:

Listing the date of the first successful landing outcome on ground pad was achieved.

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

In [26]:
```
%sql select Booster_Version as name from spacextable where landing_outcome like '%drone ship%' and PAYLOAD_MASS__KG_ > 400
```

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

Out[26]:

name
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Explanation:

Listing the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [31]:
```
%sql select count(*) from spacextable where mission_outcome like '%Failure%'
```

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

Out[31]:

count(*)
1

In [35]:
```
%sql select count(*) from spacextable where mission_outcome like '%success%'
```

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

Out[35]:

count(*)
100

Explanation:

Listing the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
 * sqlite:///my_data1.db
Done.
Out[37]:
```

Maximum_payload_mass
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

Explanation:

Listing the names of the booster which have carried the maximum payload mass.

2015 Launch Records

```
#AND substr("Date", 1, 4) = '2015'

cur.execute(query)

# Fetch all results
failed_landings_2015 = cur.fetchall()

# Print the results
print("Month Name | Landing Outcome | Booster Version | Launch Site")
print("------------------------------------------------------------")
for record in failed_landings_2015:
    print(f"{record[0]} | {record[1]} | {record[2]} | {record[3]}")
```

```
Month Name | Landing Outcome | Booster Version | Launch Site
------------------------------------------------------------
January | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40
April | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40
```

Explanation:

Listing the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

Task 10

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.

In [47]:
```
%sql select landing_outcome, count(landing_outcome) as COUNT from spacextable group by landing_outcome order by count(landi
```

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

Out[47]:

Landing_Outcome	COUNT
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

Explanation:

- Ranking 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

Section 3

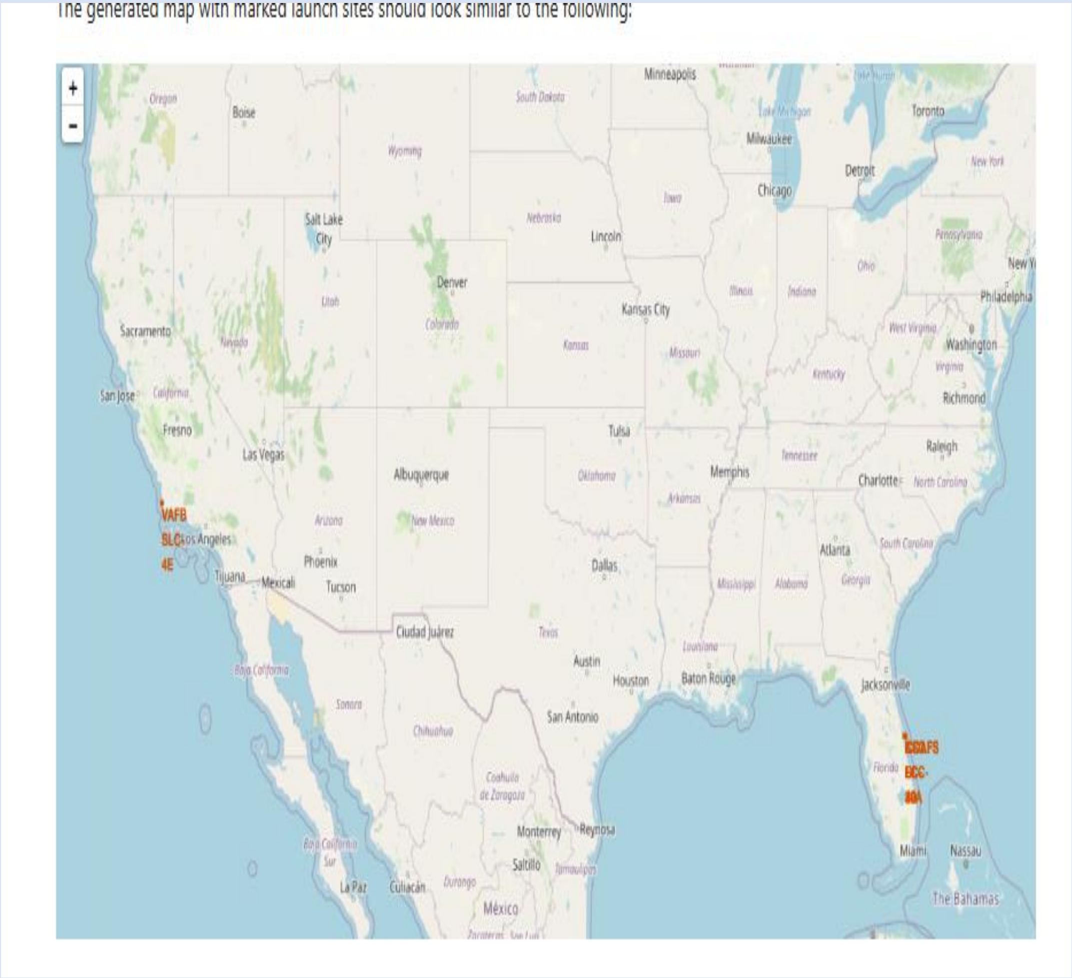
Launch Sites Proximities Analysis

All Launch Site Location Markers on a global map

Explanation:

- Most of the Launch sites are in proximity to the equator line.the land is moving faster at the equator than any other place on the surface of the earth.Anything on the surface of the earth at the equator is already moving at 1670 km/hr.if a ship is launched from a equator it goes up into space,it is also moving around the earth at the same speed it was moving before launching.This is because of inertia.This speed will help the spacecraft keep up a good enough speed to stay in orbit.

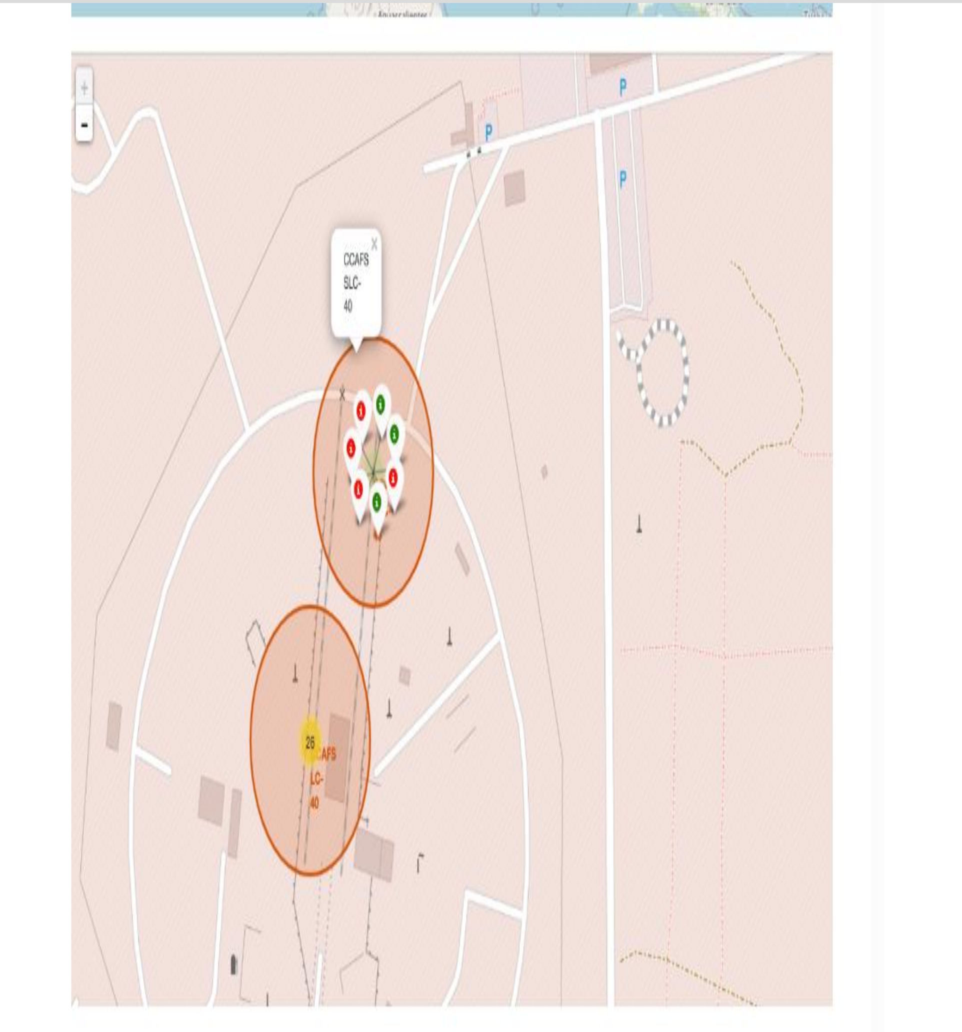
- All Launch site are in very close proximity to the coast, while launching rockets towards the ocean it minimises the risk of having an debris dropping or exploding near people,



Colour-Labeled Launch Records on the Map

Explanation:

- From the colour-Labeled markers we should be able to easily identify that which launch sites have relatively high success rate.
- Green Marker-Sucessful Launch
- Red Marker- Failed Launch
- Launch site KSC LC – 39A has high success rate.



Distance From Launch Sites KSC LC -39A to its Proximities

Explanation:

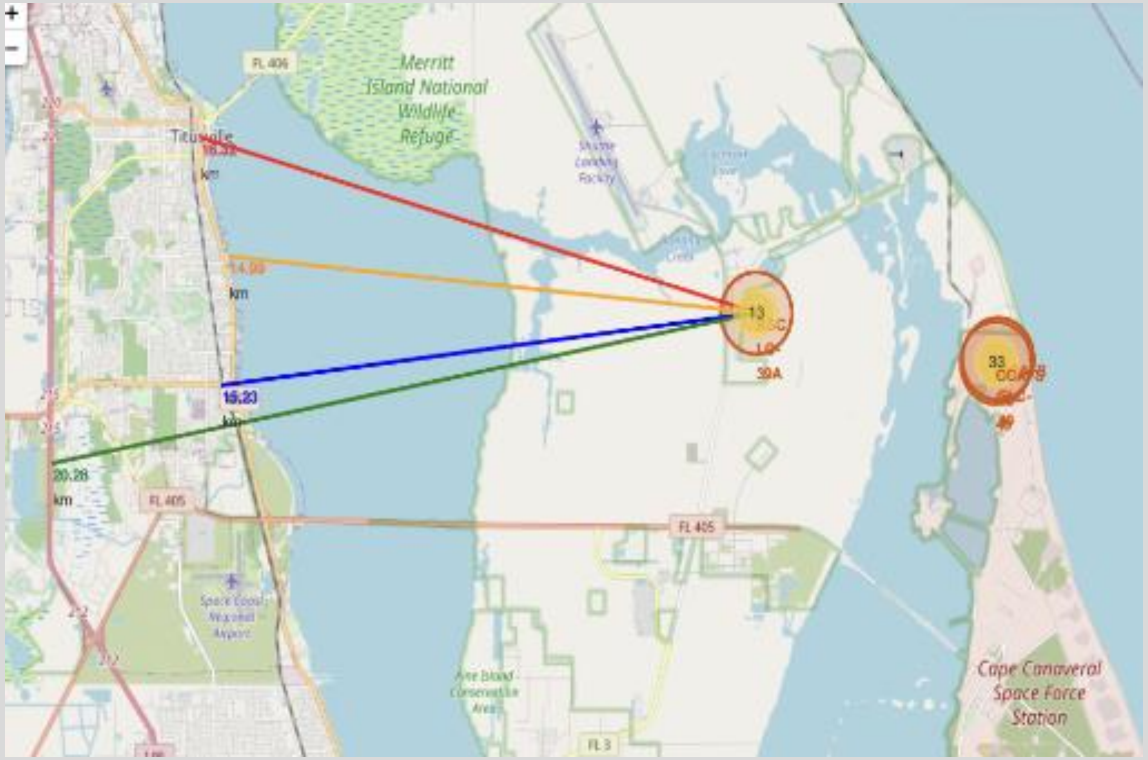
- From the visual analysis of the launch site KSC LC 39A we can clearly see that:

  -Relative close to railway(15.23km)

  -Relative close to highway(20.28km)

  -Relative close to coastline(14.99km)

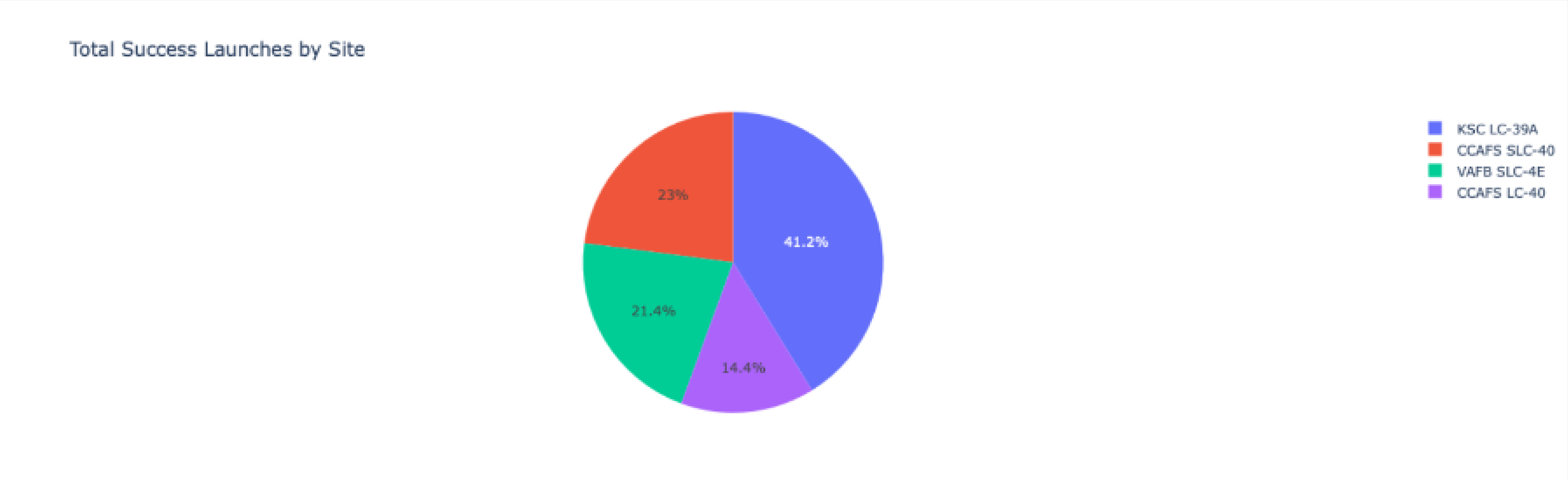
- Also the Launch site KSC LC-39A is relatively close to the closest city Titusvilli(16.32km)
- Failed rocket with its high speed can cover distance like 15-20km in few seconds.it could be potentially dangerous to the populated area.



Section 4

Build a Dashboard with Plotly Dash

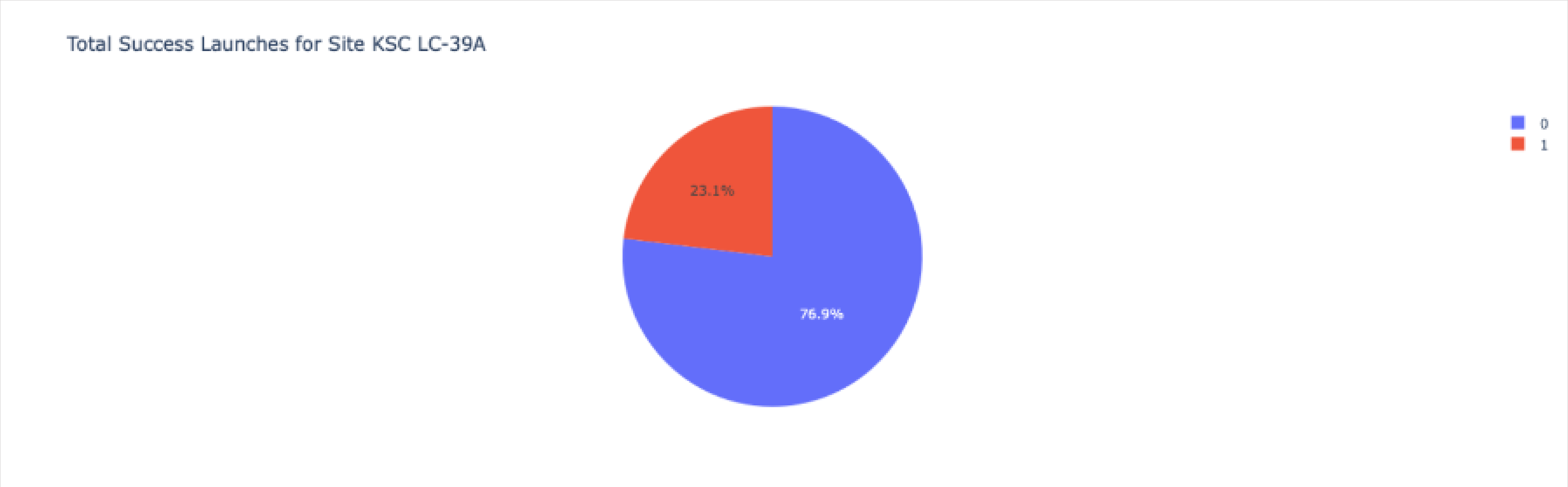
Launch Success count for all sites



Explanation:

The chart clearly shows that from all the sites KSC LC-39A has the most successful launches.

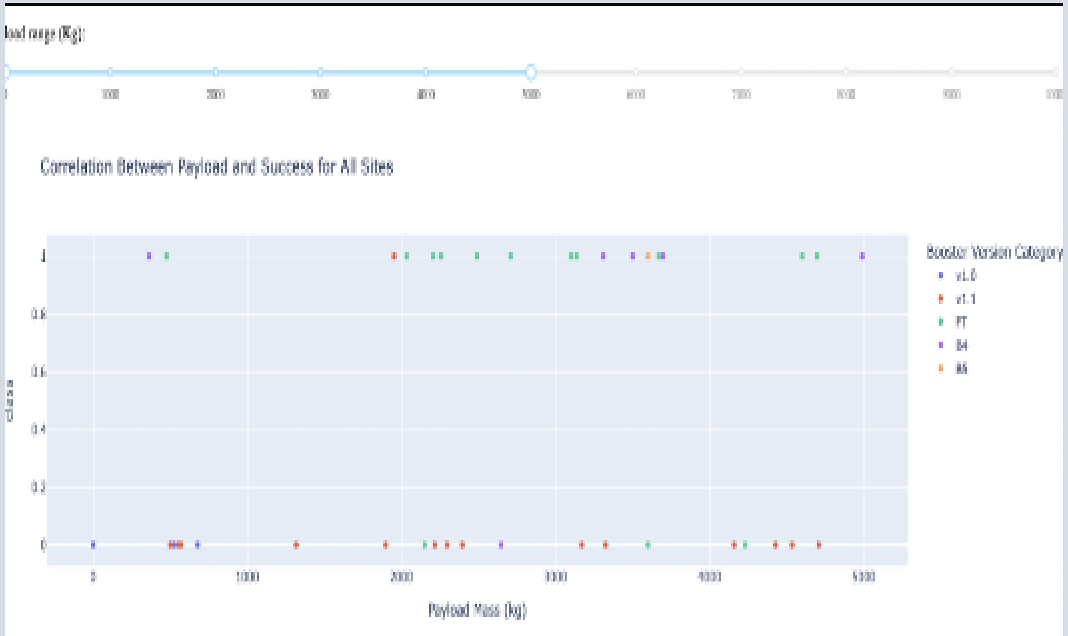
Launch site with highest launch success ratio

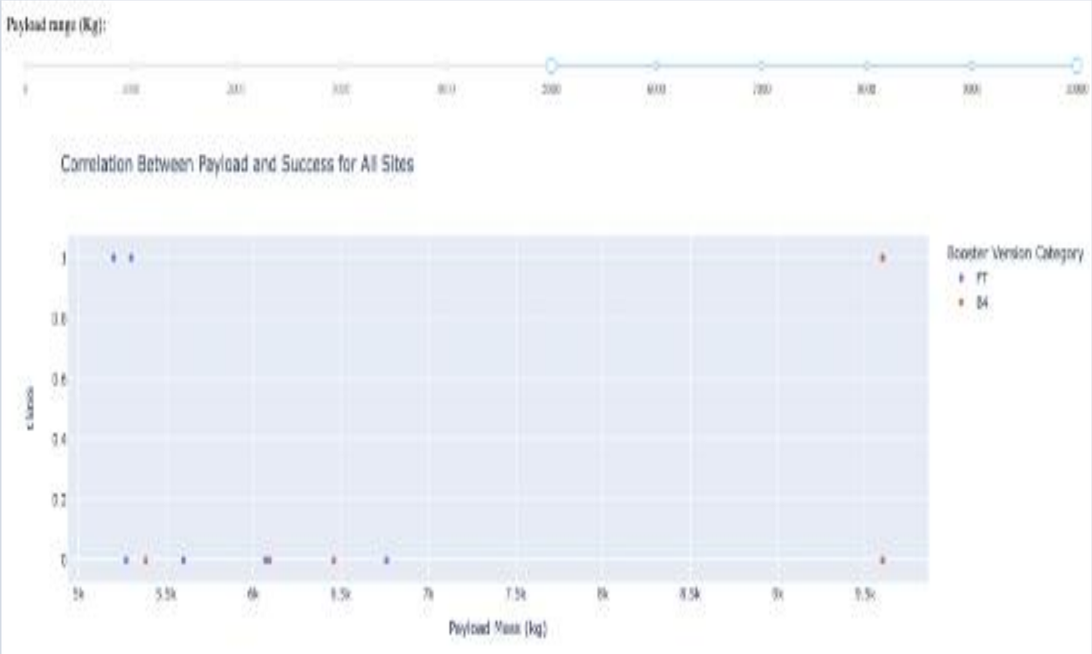


Explanation:

KSC LC 39A has the high launch success rate(76.9%) with 10 successful the 3 failed landing.

Payload Mass vs Launch Outcome for all sites





Explanation:

The chart shows that payload between 2000 and 5500 kg have the highest success rate.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Explanation:

- Based on the scores of the test set,we cannot confirm that which performs best.
- Same test set scores may be due to the small test sample size(18 samples).therefore we tested all methods based on the whole dataset.
- The score of the whole dataset confirm that the best model is Decision Tree Model.This Model has not only higher scores,but also has highest accuracy.

In [16]:

```
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

TASK 5

Calculate the accuracy on the test data using the method `score` :

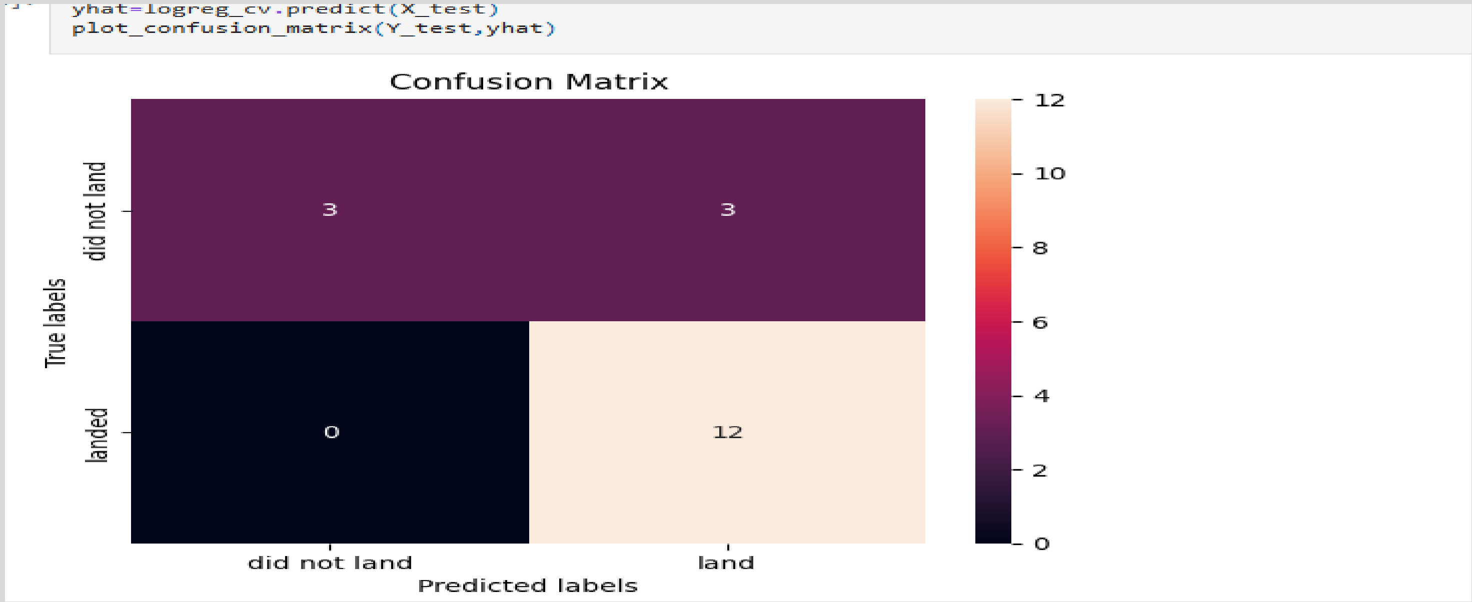
In [21]:

```
test_accuracy=logreg_cv.score(X_test,Y_test)
print("accuracy :" ,test_accuracy)
```

```
accuracy : 0.8333333333333334
```

Lets look at the confusion matrix:

Confusion Matrix



Explanation:

Examining the confusion matrix,we see that Logistic Regression can distinguish between different classes.We see that the major problem is false positive.

Conclusions

- Decision Tree algorithm is the best model for this dataset.
- Launches with a low payload mass shows better results than launches with more payload mass.
- Most of the launche site is in proximity to the equator line and all the sites are in very close proximity to the coast.
- The success rate of Launches increases over the year.
- KSC LC 39A has the highest success rate of launches from all the sites.
- Orbits ES-L1,GEO,HEO and SSO have 100% success rate.

Appendix

Special Thanks to :

[Instructor]

[Coursera]

[IBM]

Thank you!