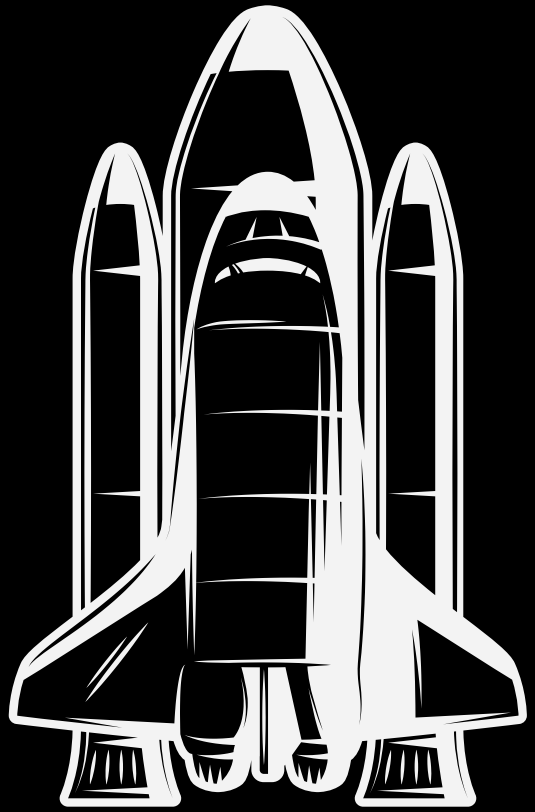
0001



PREDICTING A SUCCESSFUL LANDING W/ DATA SCIENCE

DR. LISA MARIE LEE

OUTLINE

EXECUTIVE SUMMARY

SUMMARY OF METHODOLOGIES

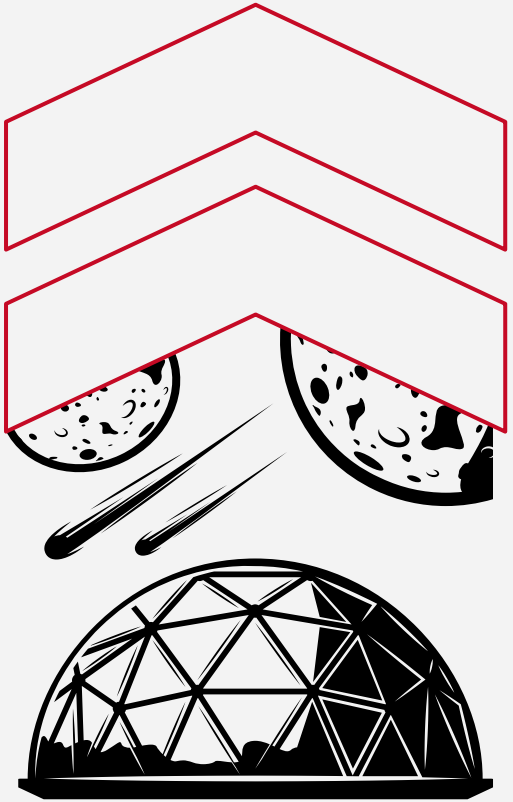
- DATA COLLECTION WITH API
- DATA COLLECTION WITH SCRAPPING
- DATA WRANGLING
- EXPLORATORY DATA ANALYSIS W/DATA VISUALIZATION
- EXPLORATORY DATA ANALYSIS W/ SQL
- INTERACTIVE VISUAL W/FOLIUM & PLOTLY DASH
- PREDICTIVE ANALYSIS

MARKET & COMPETITION

- EXPLORATORY DATA ANALYSIS RESULTS
- INTERACTIVE DATA ANALYSIS RESULTS
- PREDICTIVE ANALYSIS RESULTS

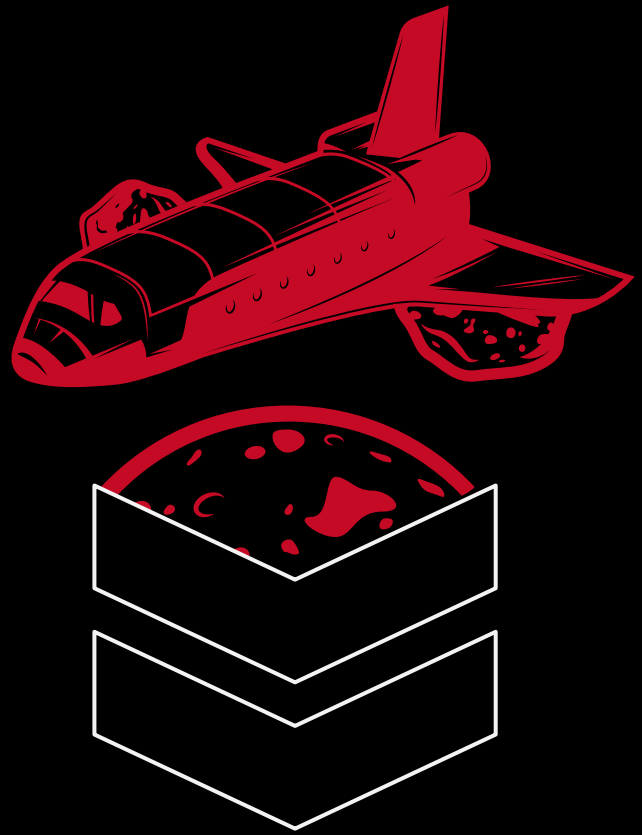
INTRODUCTION

The goal of the entire project is to make use of Data Science and build a machine learning model that can predict the first stage will land successfully.



BACKGROUND

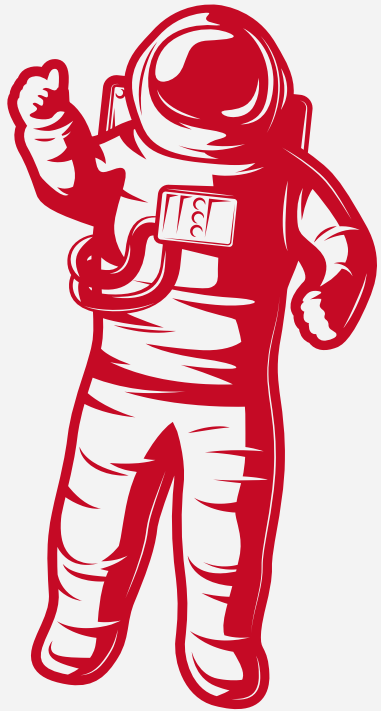
SpaceX's Falcon 9 mission sends rockets to space and the ISS at a cost of $62 million, significantly lower than the $165 million of other rockets, due to their ability to reuse the first stage. By determining if the first stage will land, we can estimate the launch cost. This information is valuable for alternate companies looking to bid against SpaceX for rocket launches.



PROBLEMS TO BE SOLVED

- What factors determine the rocket will land successfully?
- The interaction amongst the various features that determine the successful landing?
- What are the different conditions for successful landing of the rocket?

METHODOLOGY



SUMMARY OF METHODOLOGIES

- DATA COLLECTION WITH SPACEX API
- DATA WRANGLING FROM WIKIPEDIA
- EXPLORATORY DATA ANALYSIS W/DATA VISUALIZATION
- EXPLORATORY DATA ANALYSIS W/ SQL
- INTERACTIVE VISUAL W/FOLIUM & PLOTLY DASH
- PREDICTIVE ANALYSIS

DATA COLLECTION SUMMARY

- REQUESTED & PARSED THE DATA USING GET REQUEST FROM SPACEX API
- DECODED THE RESPONSE AS JSON USING json() FUNCTION & TURNED INTO PANDAS DATAFRAME USING json_normalize()
- PERFORMED WEB SCRAPING FROM WIKIPEDIA FOR FALCON 9 LAUNCH RECORDS W/ BEAUTIFULSOUP
- PARSED THE LAUNCH HTML TABLES INTO PANDAS DATA FRAME

LINKS TO ASSIGNMENTS:

#1 SpaceX Falcon 9 First Stage Landing Prediction - Lab 1 Data Collection

#2 SpaceX Falcon 9 First Stage Landing Prediction - Lab 1 Web scraping

DATA WRANGLING SUMMARY

- PERFORMED EXPLORATORY ANALYSIS & DETERMINED TRAINING LABELS
- CALCULATED THE NUMBER OF LAUNCHES OF EACH SITE & THE NUMBER OF OCCURRENCES OF EACH ORBITS
- CALCULATED THE SUCCESS RATE OF LANDING BY FINDING THE MEAN
- EXPORTED OUTPUT TO CSV

LINKS TO ASSIGNMENTS:
#3 Space X Falcon 9 First Stage Landing Prediction - Lab 2 Data Wrangling

EDA W/ DATA VISUALIZATION

- EXPLORED THE DATA BY VISUALIZING THE RELATIONSHIP BETWEEN
  - FLIGHT NUMBER & LAUNCH SITE
  - PAYLOAD & LAUNCH SITE
  - SUCCESS RATE OF EACH ORBIT
  - FLIGHT NUMBER & ORBIT TYPE
  - PAYLOAD & ORBIT TYPE
  - ORBIT TYPE & SUCCESS RATE YEARLY TREND

LINKS TO ASSIGNMENTS:
#5 - Space X Falcon 9 First Stage Landing Prediction - EDA with Pandas & Matplotlib

EDA W/ SQL SUMMARY

- LOADED DATA SET INTO JUPYTER NOTEBOOKS
- APPLIED EDA WITH SQL TO GET INSIGHTS. APPLIED QUERIES TO FIND OUT:
  - NAMES OF UNIQUE LAUNCH SITES IN SPACE MISSION
  - DISPLAY TOP 5 RECORDS WHERE LAUNCH SITE NAME BEGINNING WITH PARTICULAR STRING.
  - TOTAL AND AVERAGE PAYLOAD MASS CARRIED BY BOOSTERS
  - DATE OF 1ST SUCCESSFUL LANDING IN GROUND
  - TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSIONS AND MANY MORE.

LINKS TO ASSIGNMENTS:

#4 Space X Falcon 9 First Stage Landing Prediction - EDA with SQL

INTERACTIVE MAP W/ FOLIUM SUMMARY

- MARKED ALL THE LAUNCH SITE ON WORLD MAP
- ADDED CIRCLES TO FIND THE SUCCESS AND FAILURE LAUNCHES FROM EACH SITES. FOR THAT WE USE COLOR MARKER. GREEN COLOR SHOWS THE SUCCESSFUL LANDING OUTCOME WHEREAS RED COLOR SHOWS THE FAILURE IN THE MISSION.
- SHOWED THE DISTANCE FROM THE LAUNCH SITES TO ITS CLOSE LANDMARKS LIKE CITY, HIGHWAY, RAILWAY STATION.
- OBSERVED THAT ALL THE SITES ARE NEAR THE COASTS AND AWAY FROM POPULATED AREAS

LINKS TO ASSIGNMENTS:
#6- Space X Falcon 9 First Stage Landing Prediction - Launch Sites Location Analysis with Folium

DASHBOARD W/ PLOTLY DASH SUMMARY

- BUILT AN INTERACTIVE DASHBOARD WITH PLOTLY DASH
- PLOTTED PIE CHARTS SHOWING THE TOTAL LAUNCHES BY CERTAIN SITES
- PLOTTED THE SCATTER GRAPH SHOWING THE RELATIONSHIP WITH OUTCOME AND PAYLOAD MASS FOR DIFFERENT BOOSTER VERSION.

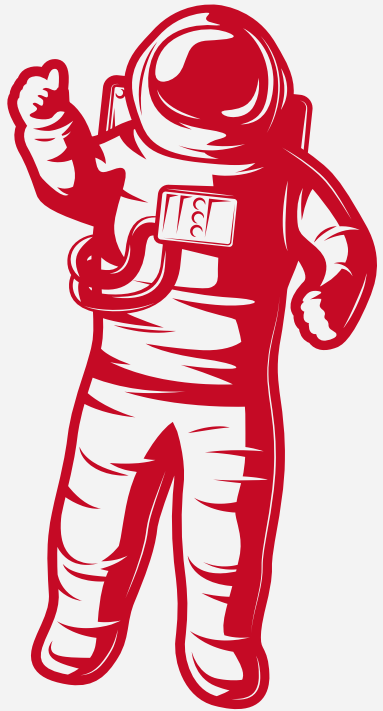
LINKS TO ASSIGNMENTS:
#7 SapceX- Hands-on Lab Build an Interactive Dashboard with Ploty Dash

PREDICTIVE ANALYSIS (CLASSIFICATION) SUMMARY

- LOADED THE DATA USING PANDAS AND NUMPY, AND SPLIT THE DATA IN TRAINING AND TESTING SETS
- BUILT DIFFERENT MACHINE LEARNING MODELS
- USED ACCURACY AS THE METRIC FOR OUR MODEL AND IMPROVED THE MODEL USING FEATURE ENGINEERING
- FOUND THE BEST PERFORMING CLASSIFICATION MODEL. WE SEE THE DECISION TREE MODEL HAS THE HIGHEST TRAINING ACCURACY

LINKS TO ASSIGNMENTS:
#8 Space X Falcon 9 First Stage Landing Prediction

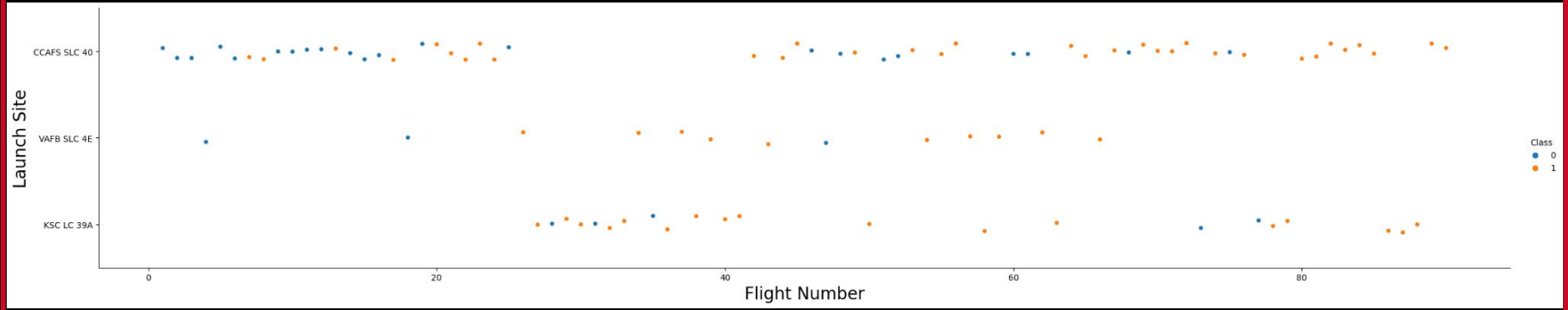
RESULTS



EDA VISUALIZATION

FLIGHT NUMBER VS. LAUNCH SITE

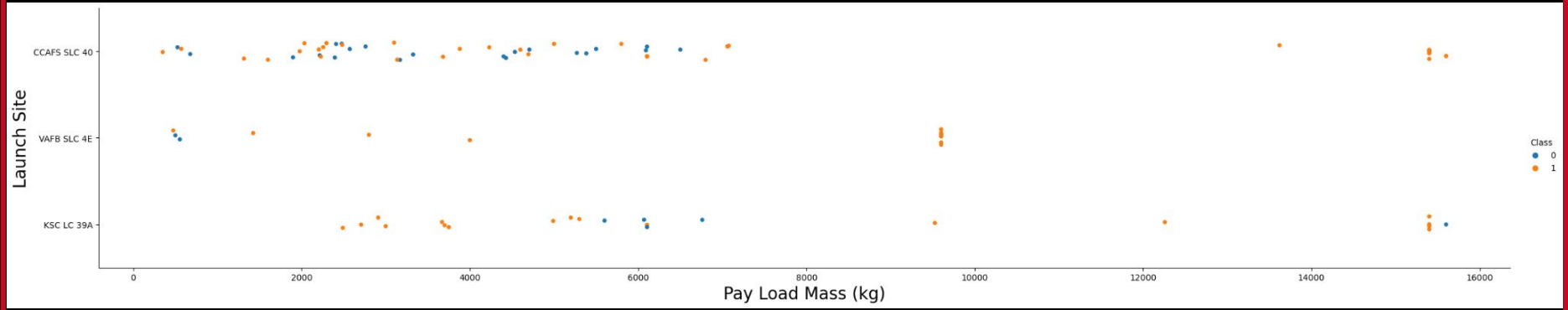
FROM THE PLOT, WE FIND THAT THE LARGER THE AMOUNT OF FLIGHTS, AT A PARTICULAR LAUNCH SITE, THE GREATER IS THE SUCCESS RATE. HERE CCAFS-SLC 40 HAS THE MAXIMUM SUCCESS RATE.



EDA VISUALIZATION

PAYLOAD VS. LAUNCH SITE

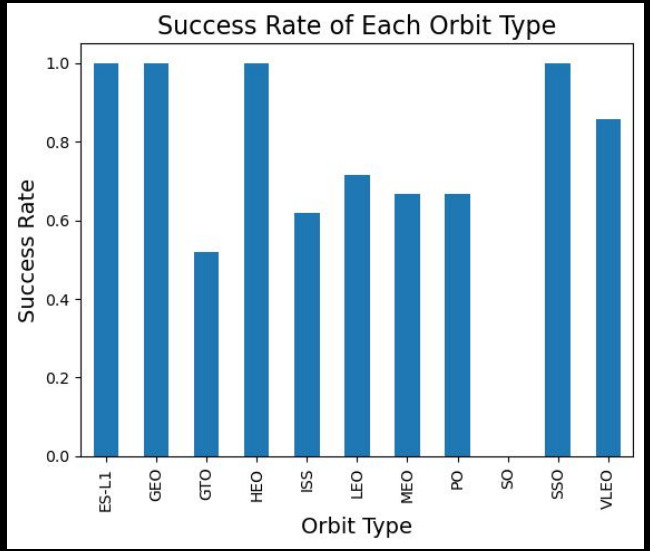
- FROM THE PLOT, WE FIND THAT VAFB SLC 4E LAUNCH SITE, THERE RE NO ROCKETS OF HEAVY PAYLOAD MASS GREATER THAN 10000.
- ALSO CCAFS SLC 40 HAS HIGH SUCCESS FOR HIGH PAYLOAD MASS AROUND 15000



EDA VISUALIZATION

SUCCESSION RATE VS. ORBIT TYPE

- FROM THE BAR PLOT, WE CAN CLEARLY SEE THAT ES-L1, GEO, HEO AND SSO HAVE 100% SUCCESS RATE FOLLOWED BY VLEO.
- SO HAS 0 SUCCESS RATE.

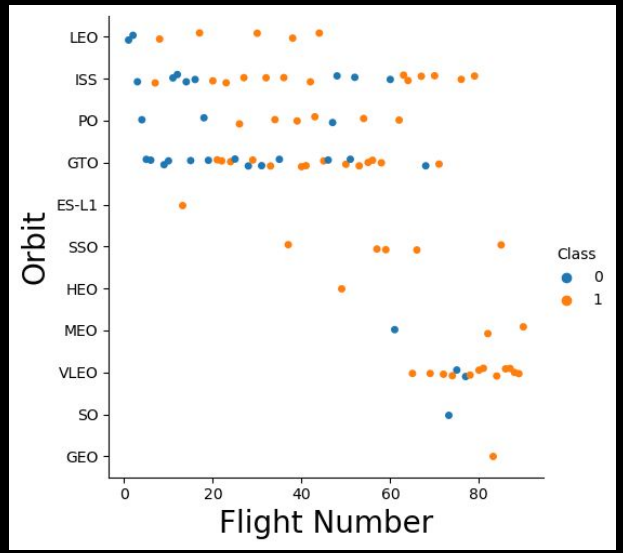


EDA VISUALIZATION

FLIGHT NUMBER VS. ORBIT TYPE

2020

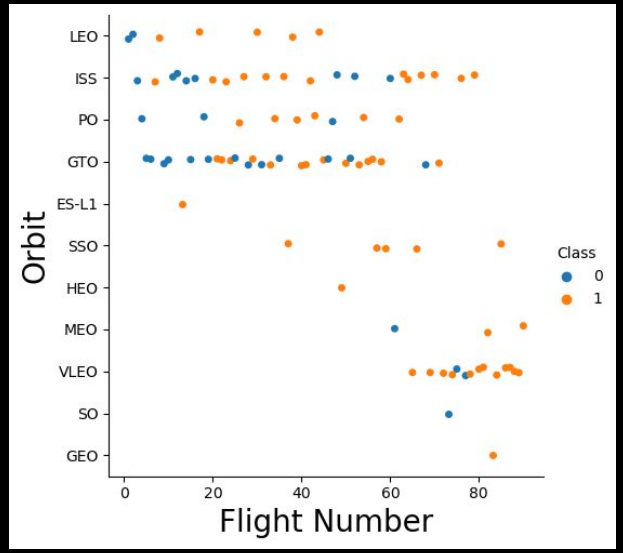
- FROM THE PLOT, WE SEE THAT THERE ARE VERY FEW FLIGHTS FLOWN FROM LEO ORBIT. HOWEVER ALL THE FLIGHTS ARE SUCCESSFUL. ON THE OTHER HAND, THERE SEEMS TO BE NO RELATIONSHIP BETWEEN FLIGHT NUMBER WHEN IN GTO ORBIT.



EDA VISUALIZATION

PAYLOAD VS. ORBIT TYPE

- FROM THE PLOT, WE SEE WITH HEAVY PAYLOADS THE SUCCESSFUL LANDING RATES ARE MORE FOR POLAR, LEO AND ISS.
- HOWEVER FOR GTO, WE CANNOT DISTINGUISH THIS AS BOTH POSITIVE AND NEGATIVE LANDING ARE THERE HERE.

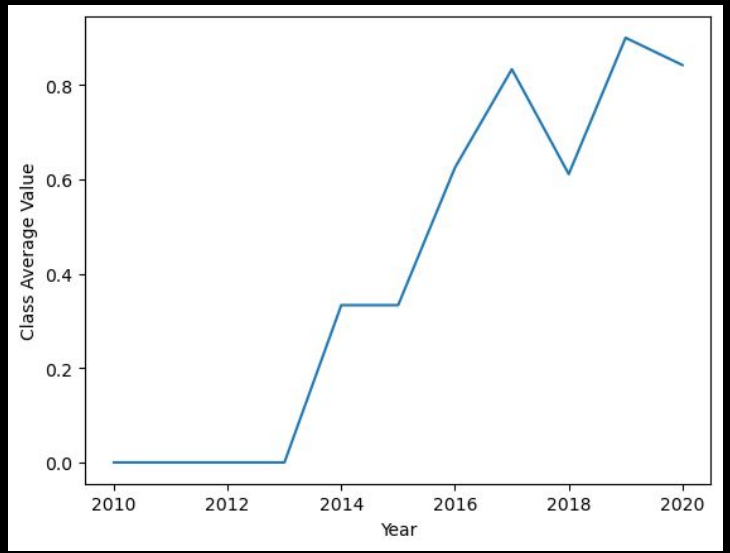


EDA VISUALIZATION

LAUNCH SUCCESS YEARLY TREND

2022

- FROM THE PLOT, WE CAN SEE THE SUCCESS RATE IS INCREASED IN 2013 AND IT IS CONSTANT TILL 2015, AFTER THAT IT BOOSTED UP TILL 2017.
- IN MID OF 2018 WE CAN SEE SUDDEN DECREASE WHICH AGAIN INCREASED 2019.



EDA SQL

UNIQUE LAUNCH SITES

- USED THE KEYWORD 'DISTINCT' TO FIND OUT UNIQUE LAUNCH SITES FOR THE MISSION.

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

```
%sql select DISTINCT LAUNCH_SITE from SPACEXTBL
```

Running query in 'sqlite:///my_data1.db'

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

EDA SQL

LAUNCH SITES BEGINNING W/ 'CCA'

- USED THE 'LIKE' KEYWORD TO FIND THE LAUNCH SITE NAMES STARTING WITH 'CCA' AND 'LIMIT' TO FIND 1ST 5 SITES.

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

```
%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

Running query in 'sqlite:///my_data1.db'

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

EDA SQL

TOTAL PAYLOAD MASS BY NASA CRS

- USED THE ALIAS ‘AS’ KEYWORD TO CHANGE THE COLUMN NAME OF PAYLOAD MASS.
- USED ‘WHERE’ KEYWORD TO FILTER OUT THE CUSTOMERS.
- USED ‘SUM’ FOR ADDING ALL THE PAYLOAD MASS VALUES.

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

```
]: %sql SELECT sum(PAYLOAD_MASS__KG_) as TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE CUSTOMER LIKE '%NASA (CRS)%';
```

Running query in 'sqlite:///my_data1.db'

]:

TOTAL_PAYLOAD_MASS
48213

EDA SQL

AVERAGE PAYLOAD MASS BY F9 v1.1

- USED ‘LIKE’ HERE AGAIN AS THERE ARE MANY BOOSTERS STARTING WITH F9 V1.1 AND ‘AVG’ TO CALCULATE THE AVERAGE OF THE PAYLOAD MASS.

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

```
%sql SELECT AVG (PAYLOAD_MASS__KG_) as TOTAL_AVERAGE_PAYLOAD_MASS FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1
```

Running query in 'sqlite:///my_data1.db'

TOTAL_AVERAGE_PAYLOAD_MASS
2928.4

EDA SQL

FIRST SUCCESSFUL GROUND LANDING DATE

- USED ‘LIKE’ HERE AGAIN AS THERE ARE MANY BOOSTERS STARTING WITH F9 V1.1 AND ‘AVG’ TO CALCULATE THE AVERAGE OF THE PAYLOAD MASS.

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

```
%sql SELECT AVG (PAYLOAD_MASS__KG_) as TOTAL_AVERAGE_PAYLOAD_MASS FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1
```

Running query in 'sqlite:///my_data1.db'

TOTAL_AVERAGE_PAYLOAD_MASS
2928.4

EDA SQL

FIRST SUCCESSFUL LANDING OUTCOME IN GROUND PAD

- USED THE 'MIN' KEYWORD TO FIND THE FIRST SUCCESSFUL MISSION OF 2015

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

Hint:Use min function

```
%sql SELECT MIN(DATE) as FIRST_SUCCESSFUL_LANDING_OUTCOME_IN_GROUND_PAD FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Succ
```

Running query in 'sqlite:///my_data1.db'

FIRST_SUCCESSFUL_LANDING_OUTCOME_IN_GROUND_PAD
2015-12-22

EDA SQL

SUCCESSFUL DRONE SHIP LANDING W/ PAYLOAD 4000 & 6000

- USED THE 'BETWEEN' KEYWORD TO FILTER OUT THE PAYLOAD MASS 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

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING_OUTCOME ='Success (drone ship)' AND PAYLOAD_MASS__KG_> 4000
```

Running query in 'sqlite:///my_data1.db'

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

EDA SQL

TOTAL NUMBER OF SUCCESSFUL & FAILURE MISSION OUTCOMES

- USED THE 'COUNT' KEYWORD TO COUNT THE TOTAL SUCCESS AND FAILURE MISSION.
- USED 'GROUP BY' KEYWORD TO GROUP ALL THE MISSION OUTCOMES TOGETHER.

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS TOTAL FROM SPACEXTBL WHERE MISSION_OUTCOME IN ('Success', 'Failure (in flig
```

Running query in 'sqlite:///my_data1.db'

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

EDA SQL

BOOSTERS CARRIED MAXIMUM PAYLOAD

- USED SUB QUERY TO FIND MAXIMUM PAYLOAD.

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

Running query in 'sqlite:///my_data1.db'

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

Truncated to displaylimit of 10.

EDA SQL

2015 LAUNCH RECORDS

- USED SUBSTRING TO EXTRACT REQUIRED DATA.

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING_OUTCOME LIKE 'FAIL%' AND DATE LIKE '2015
```

Running query in 'sqlite:///my_data1.db'

Date	Booster_Version	Launch_Site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

EDA SQL

RANK LANDING OUTCOMES BETWEEN 2010-06-04 & 2017-03-20

- USED THE 'GROUP BY' AND 'ORDER BY' CLAUSE TO GROUP ALL THE EVEN RECORDS AND SORT THEM. FURTHER WE USE 'DESC' TO SORT IN DESCENDING ORDER.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP
```

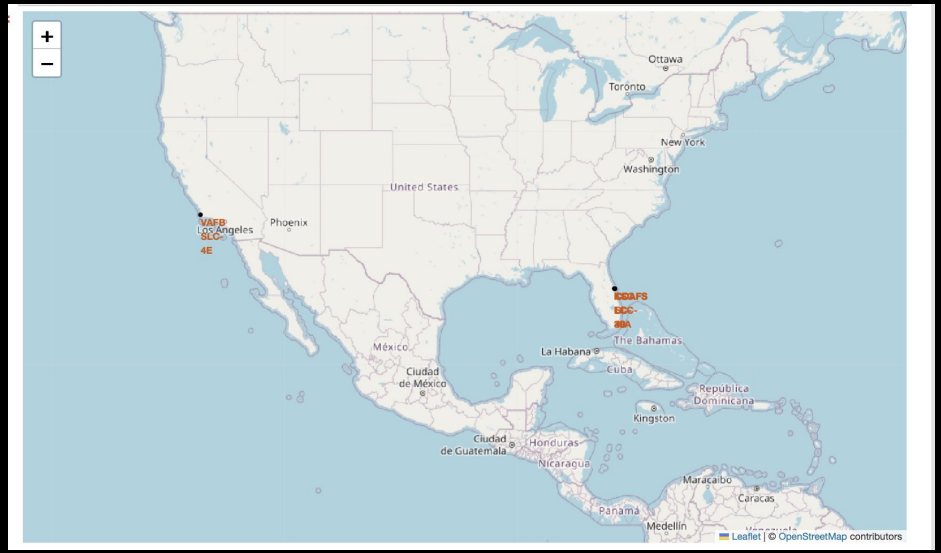
Running query in 'sqlite:///my_data1.db'

Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

INTERACTIVE MAP W/ FOLIUM

2015 LAUNCH RECORDS

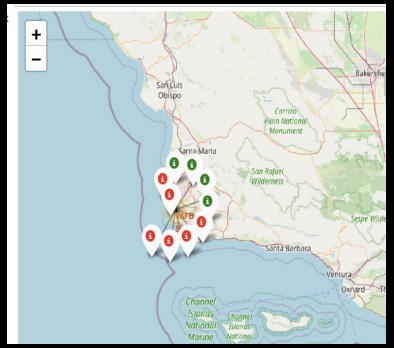
- WE CAN SEE THE SPACEX LAUNCH SITES ARE IN USA COASTS IN FLORIDA AND CALIFORNIA

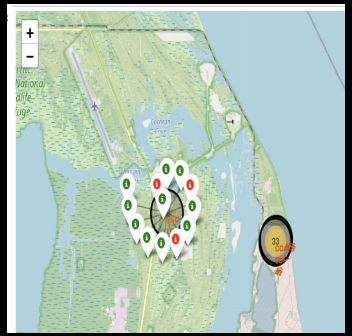


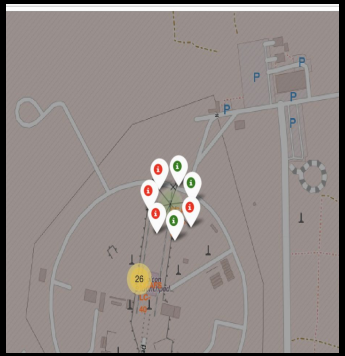
INTERACTIVE MAP W/ FOLIUM

SUCCESSFUL & FAILED LAUNCHES FOR EACH SITE

- GREEN MARKER SHOWS SUCCESSFUL LAUNCHES AND RED MARKER SHOWS FAILURES



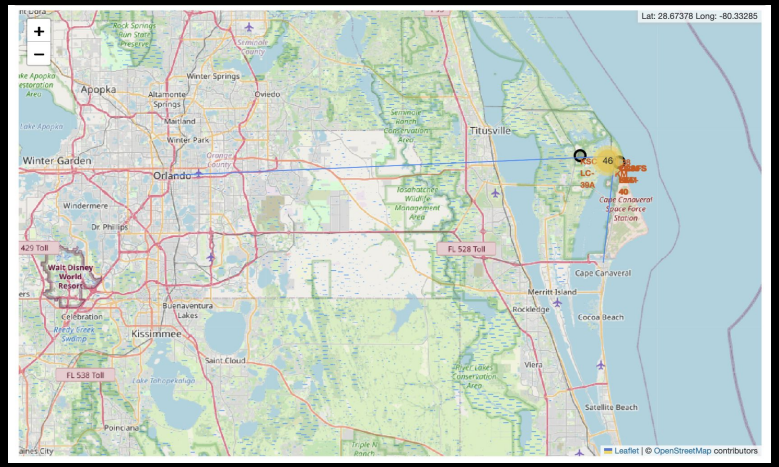


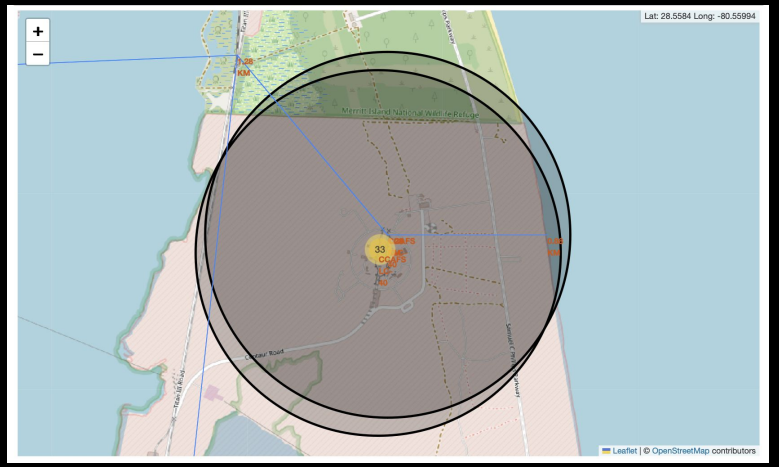


INTERACTIVE MAP W/ FOLIUM

LAUNCH SITE DISTANCE FROM LANDMARKS

- THE LAUNCH SITES ARE NEAR THE COASTS.
- THE NEAREST CITY WHICH IS ORLANDO, IS FAR AWAY FROM THE 2 LAUNCH SITES ON THE EASTERN COAST.
- ALSO HIGHWAYS AND RAILWAY LINES ARE FAR AWAY FROM THE SITES.

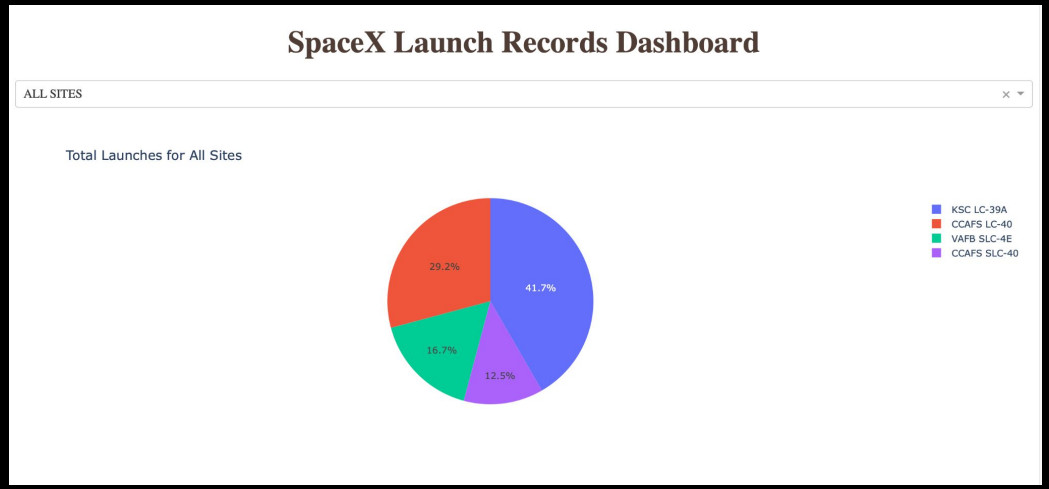




DASHBOARD W/ PLOTLY DASH

PIE CHART SHOWING THE SUCCESS PERCENTAGE ACHIEVED BY EACH LAUNCH SITE

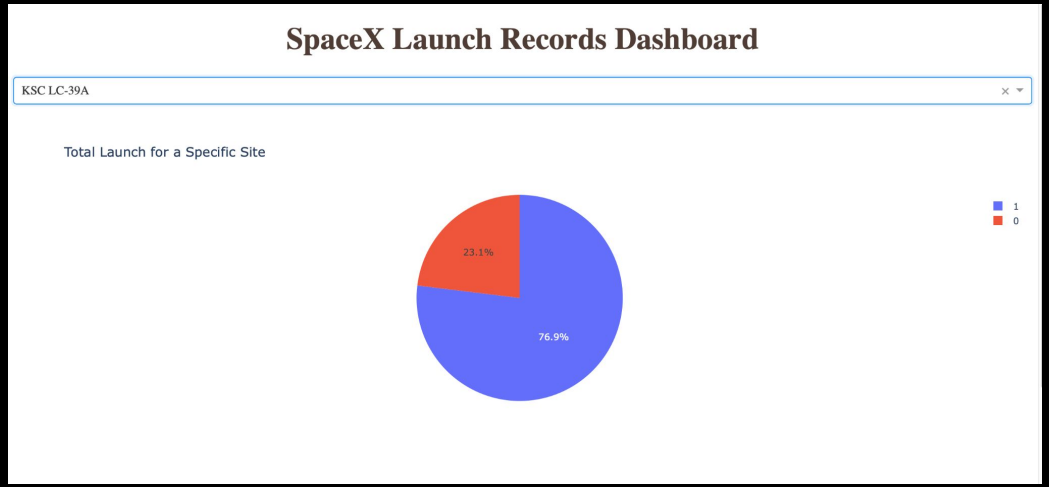
- KSC LC-39A HAD THE MOST SUCCESSFUL LAUNCHES FROM ALL THE SITES
- KSC LC -39A HAS THE MOST SUCCESSFUL LAUNCHES COMPARED TO ALL OTHER SITE



DASHBOARD W/ PLOTLY DASH

PIE CHART SHOWING THE LAUNCH SITE W/ THE HIGHEST LAUNCH SUCCESS RATE

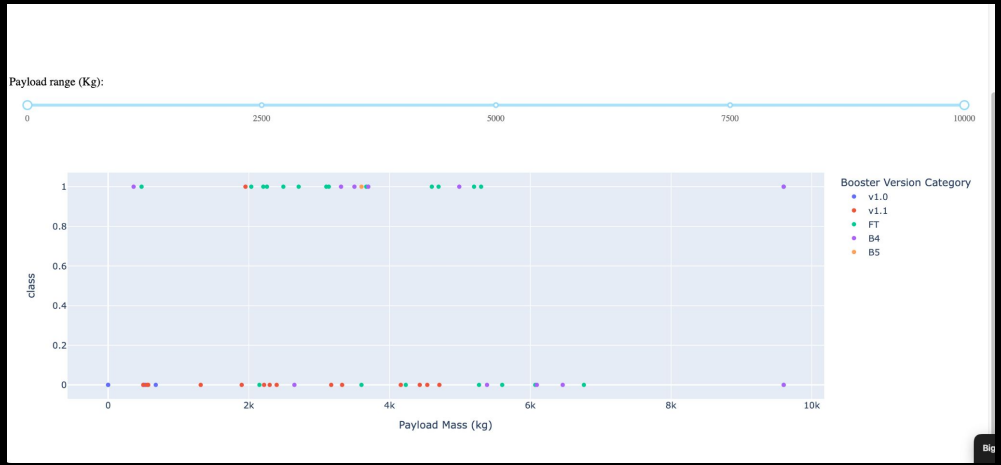
- KSC LC-39A HAD 76.0% SUCCESSFUL RATE, WHILE GETTING A 23.1% FAILURE RATE



DASHBOARD W/ PLOTLY DASH

SCATTER PLOT OF PAYLOAD VS LAUNCH OUTCOME FOR ALL SITES

- SUCCESS RATES OF LOW WEIGHTED PAYLOADS IS HIGHER THAN THE HEAVY WEIGHTED PAYLOADS



Questions

Which site has the largest successful launches? VAFB-SLC-4E

Which site has the highest launch success rate? VAFB-SLC-4E

Which payload range(s) has the highest launch success rate? 0 - 4000

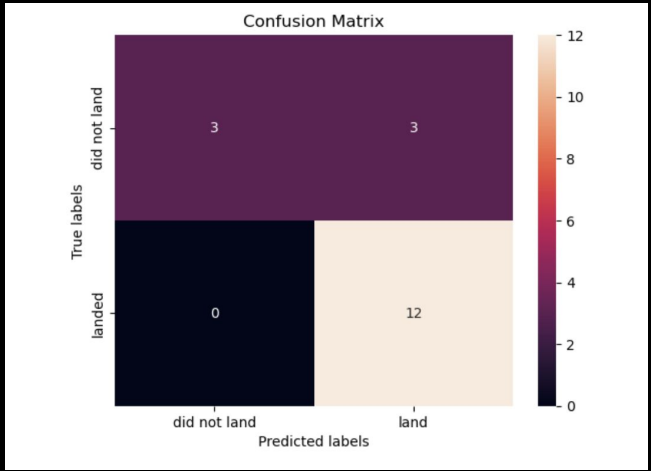
Which payload range(s) has the lowest launch success rate? 4000 - 10000

Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? FT

PREDICTIVE ANALYSIS

CONFUSION MATRIX

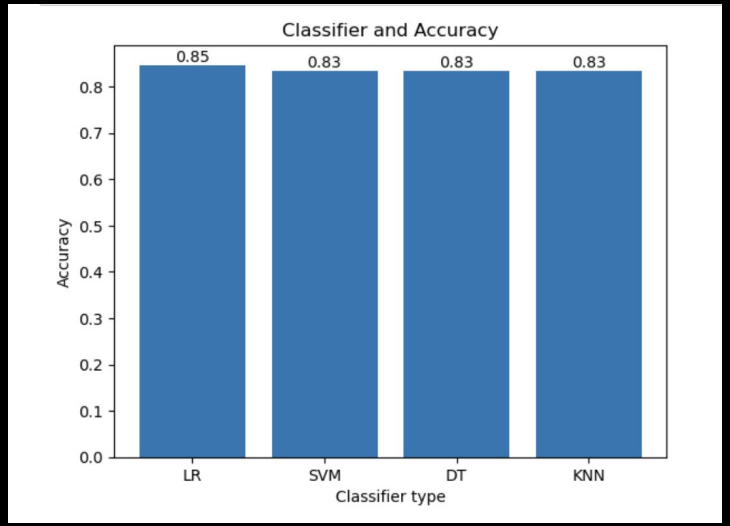
- THE CONFUSION MATRIX FOR DECISION TREE SHOWS THAT THE CLASSIFIER CAN DISTINGUISH BETWEEN DIFFERENT CLASSES.
- THE FALSE POSITIVE IS A PROBLEM I.E THE CHANCE OF SUCCESSFUL LANDING MARKED AS SUCCESSFUL BY THE CLASSIFIER.



PREDICTIVE ANALYSIS

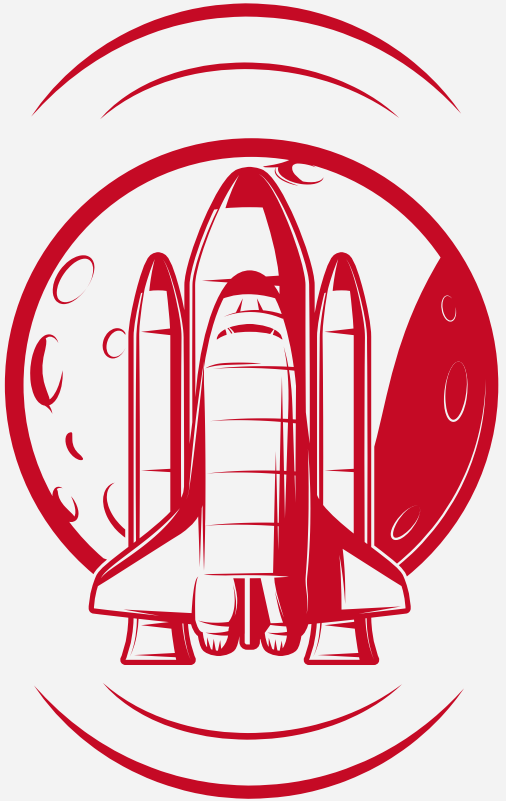
CLASSIFICATION ACCURACY

- THE BARPLOT CLASSIFIES THE ACCURACY OF TRAINING SETS OF DIFFERENT CLASSIFIERS.
- OF ALL THE MODELS IS ALMOST THE SAME AROUND 83%, HOWEVER THE ACCURACY OF LOGISTIC REGRESSION MODEL IS HIGHER THAT OTHERS, SO IN THIS CASE LR IS MOST ACCURATE CLASSIFIER.



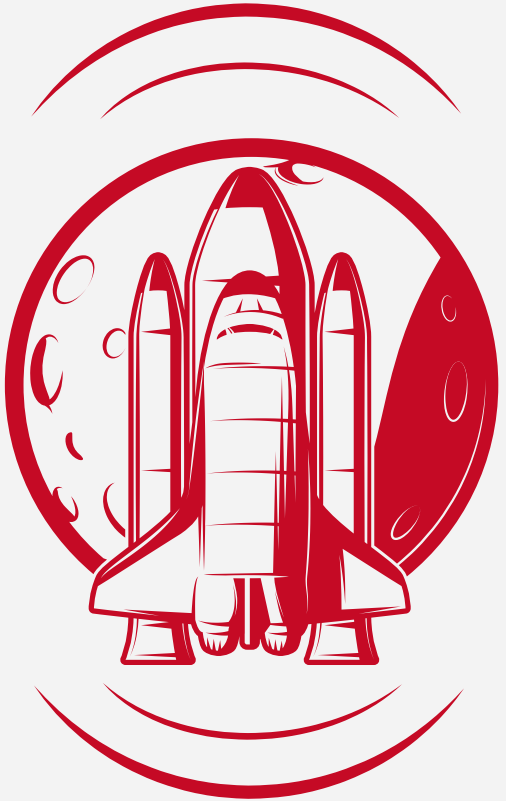
CONCLUSIONS

- LOGISTIC MODEL IS THE BEST ALGORITHM FOR THIS DATASET.
- THE LARGER THE FLIGHT AMOUNT AT THE LAUNCH SITE, THE GREATER IS THE SUCCESS RATE AT A LAUNCH SITE.
- LAUNCH SUCCESS FOR LOW PAYLOAD MASS IS BETTER FOR ALL THE ORBITS.
- LAUNCH SUCCESS RATE STARTED TO INCREASE AS THE YEARS WENT PAST.
- ORBITS ES L1, GEO, HEO, SO, VLEO HAD THE MOST SUCCESS RATE.
- KSC LC-39A HAD THE MOST SUCCESSFUL LAUNCHES OF ANY SITES.
- LAUNCH SITES ARE PLACED NEAR COASTAL AREAS AND AWAIT FROM POPULATED AREAS.



APPENDIX

- Special Thanks to –
- All the instructors throughout the course
- Coursera
- IBM



0043

THANKS!