IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chirayu
02/14/2025

Outline

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

Executive Summary

- We analyzed factors affecting success/failure of SPACEX missions in this exercise
- We performed Data procurement, data processing, visualization and predictions
- We identified factors affecting mission success/failure
- We identified factors most important in success/failure prediction of missions

Introduction

- Project background and context
  - This project was to analyze SPACEX launch data to study relationships between several variables such as Rocket Payload, Payload Mass, etc with Success/Failure of mission
- Problems you want to find answers
  - Are there any variables that are better suited to predict whether a lauch will result in success or Failure

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was downloaded from SPACE X API (V4)
- Perform data wrangling
  - Data was processed in several steps
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

https://github.com/chirayug/ds_project.git

Data Collection

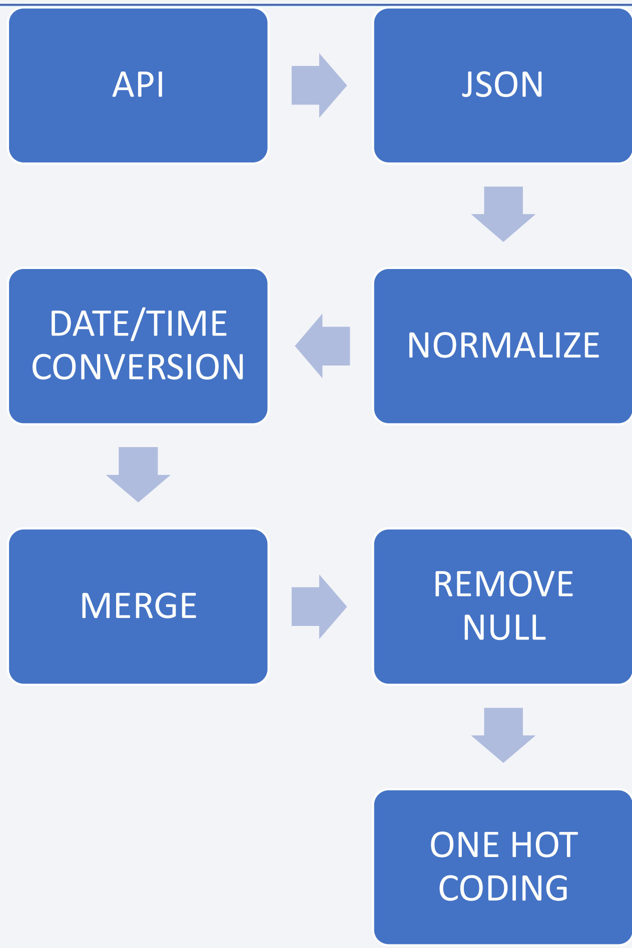
- Describe how data sets were collected.
    - Data was collected from SPACEX API (v4)
- You need to present your data collection process use key phrases and flowcharts
    - Following APIs were used
        - rockets_url = https://api.spacexdata.com/v4/rockets/
        - launchpad_url = https://api.spacexdata.com/v4/launchpads/
        - payloads_url = https://api.spacexdata.com/v4/payloads/
        - core_url = https://api.spacexdata.com/v4/cores/
        - main_url = https://api.spacexdata.com/v4/launches/past
        - static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

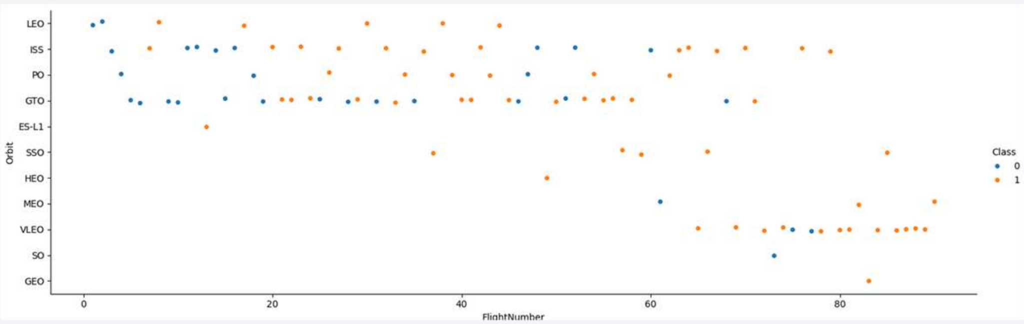
- Repository with code is available at

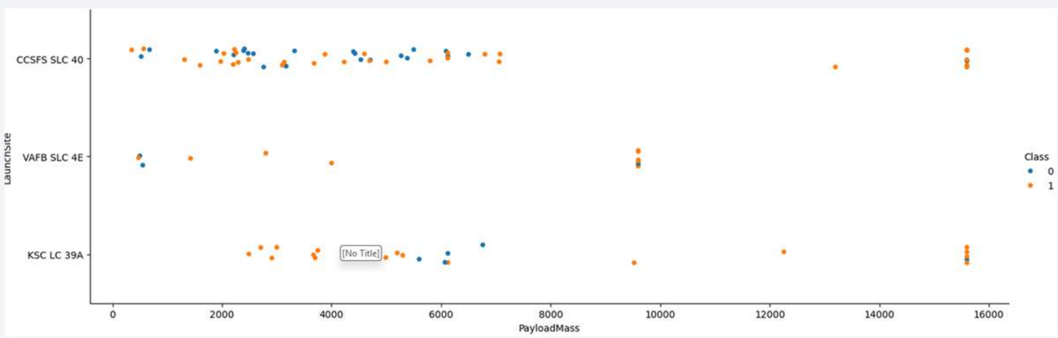
https://github.com/chirayug/ds_project.git

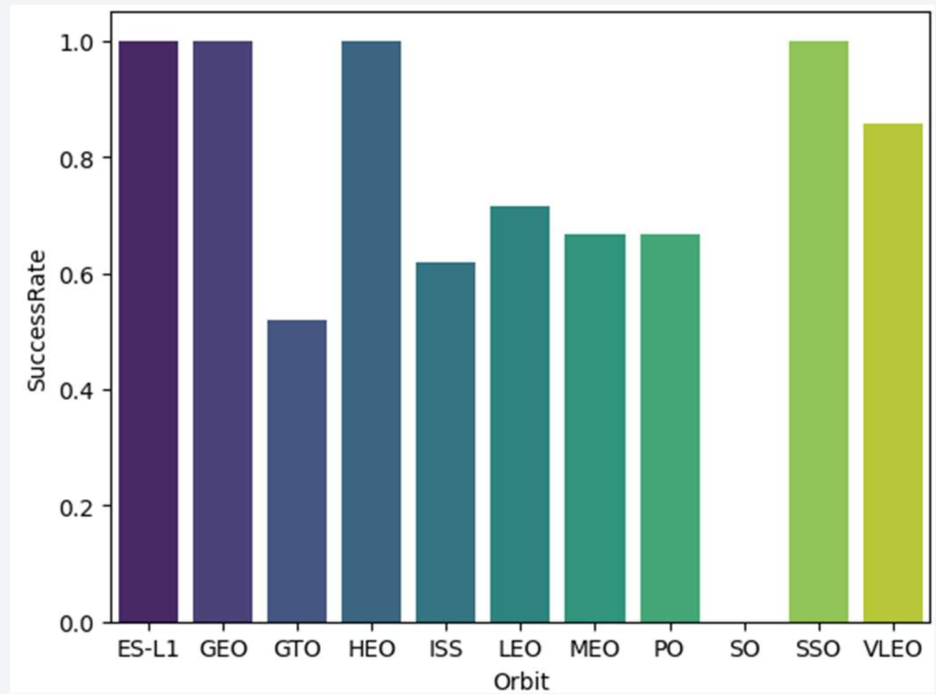


https://github.com/chirayug/ds_project.git

EDA with Data Visualization







https://github.com/chirayug/ds_project.git

EDA with SQL

- Following queries were performed for analysis
  - %sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
  - %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer = 'NASA (CRS)’
  - %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1’
  - %sql select min(Date) from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)’
  - %sql select distinct(Booster_Version) from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

https://github.com/chirayug/ds_project.git

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- I used markers, circles and lines for launc sites, success/failure clusters, and nearest landmarks to calculate distances

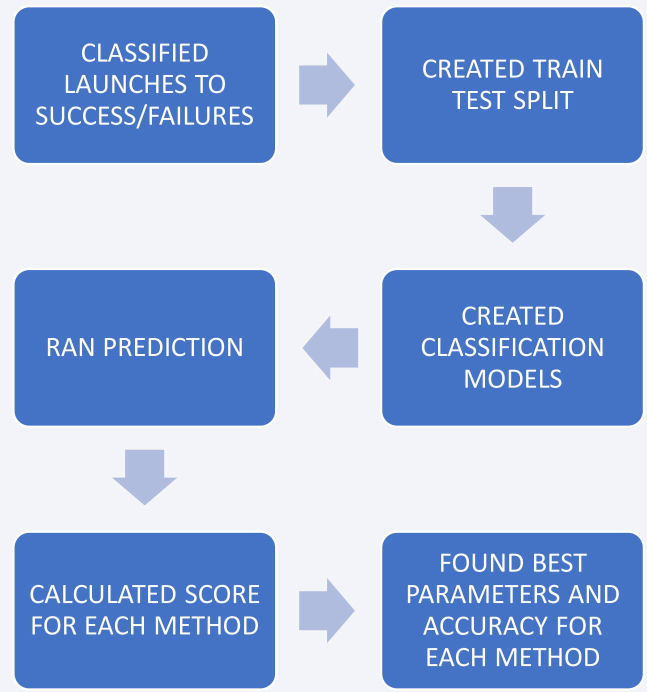
https://github.com/chirayug/ds_project.git

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- I added percent launch success/failures, by site, also filterable by payload mass to the plot. This helps us visualize total success/failures, success/failures by launch site, and further for each launch site, by payload mass.

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model



Results

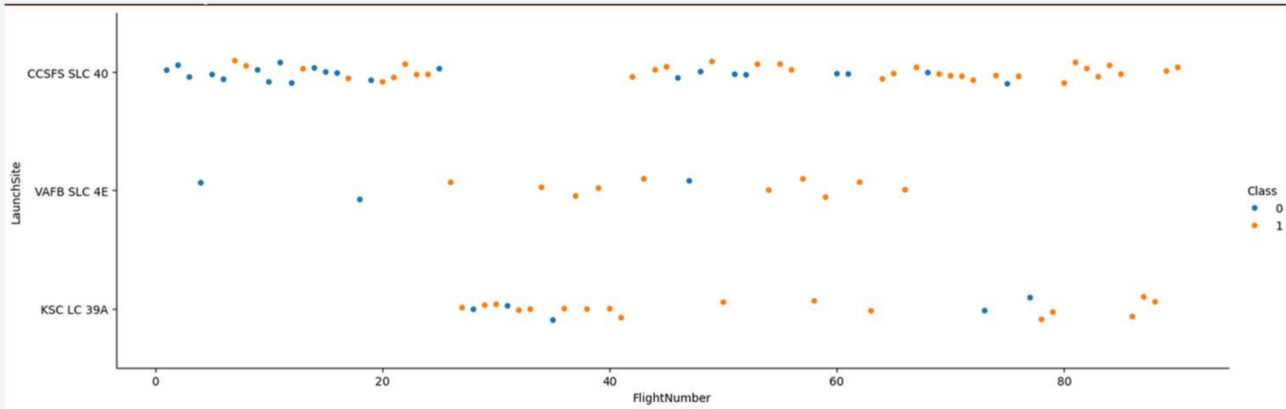
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

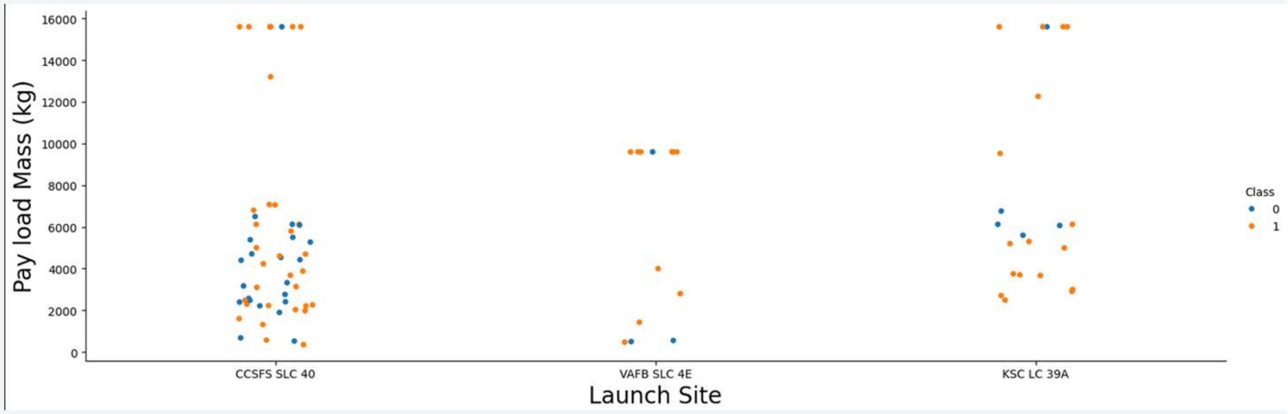


- This plot shows success/failures of consecutive flights. Initial flights had a mix, where as recent flights have been all successful irrespective of launch site

Payload vs. Launch Site

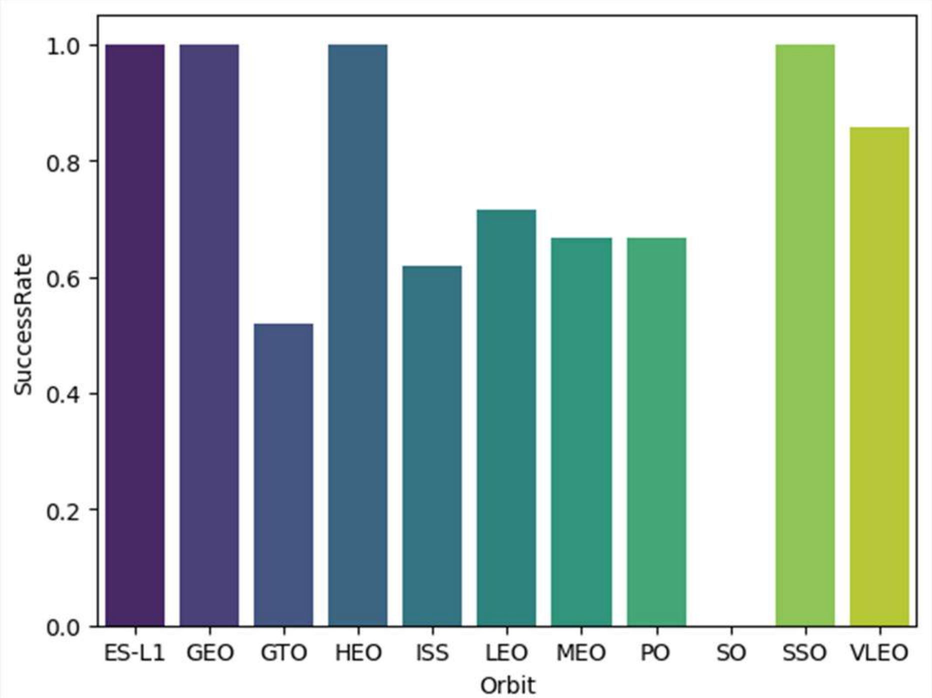
- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

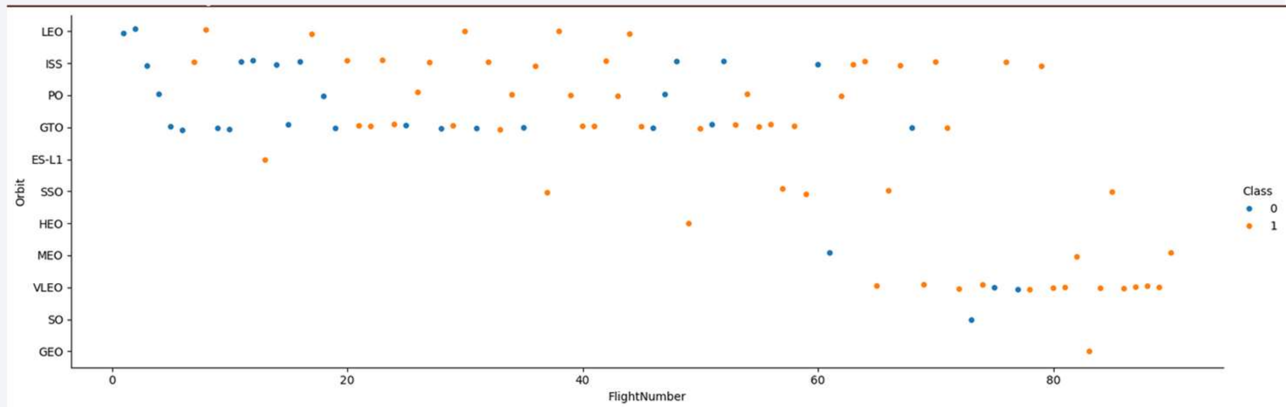
- Show a bar chart for the success rate of each orbit type
- This graph shows that some orbits have a 100% success rate (e.g., ES-L1, GEO) whereas some (e.g., GTO) have almost half failure rate



Flight Number vs. Orbit Type

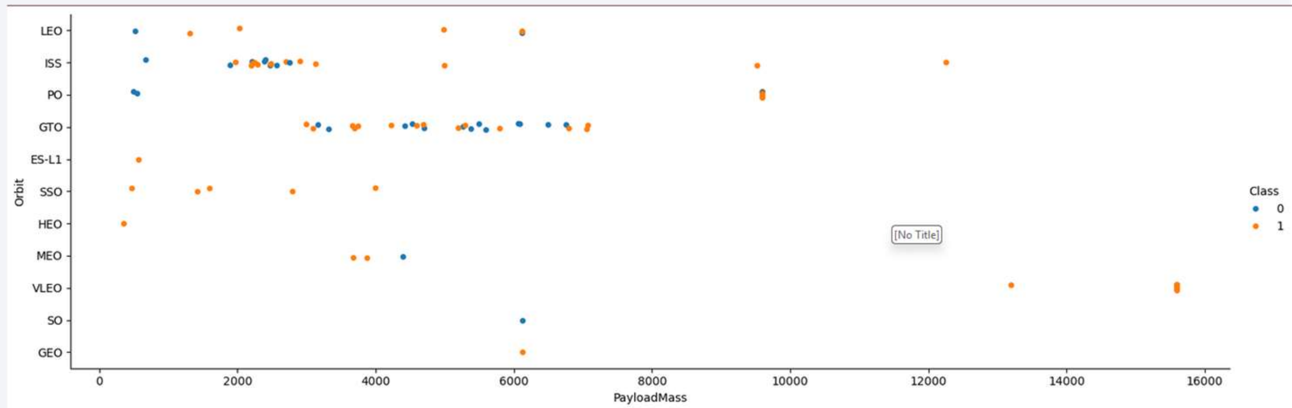
- Show a scatter point of Flight number vs. Orbit type

- This plot shows a pattern change in orbit selection between old and recent flights with most recent ones being VLEO with high success rate.



Payload vs. Orbit Type

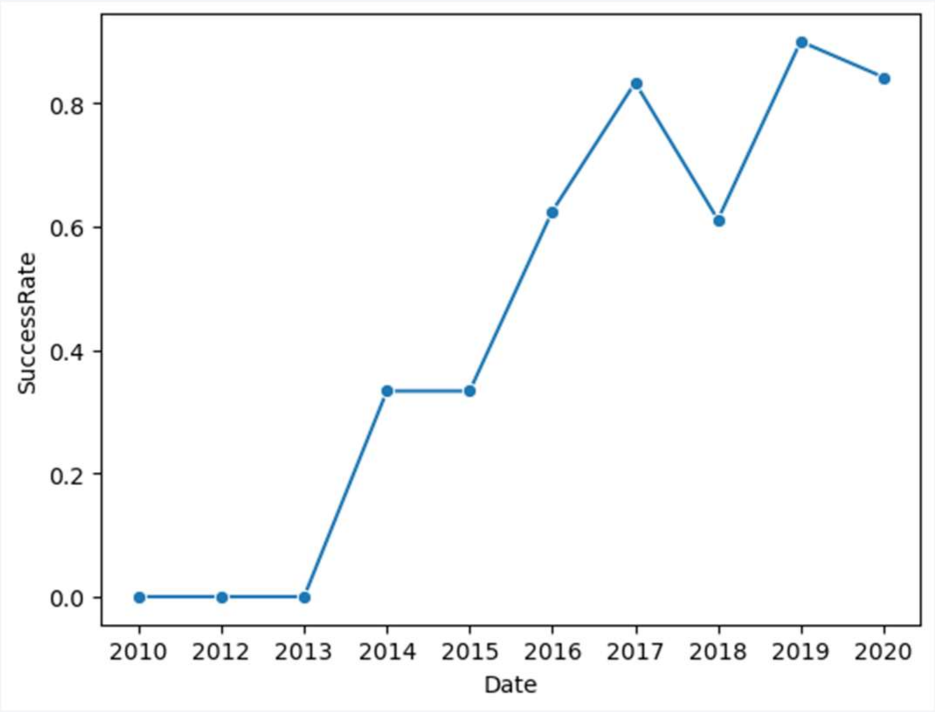
- Show a scatter point of payload vs. orbit type
- Higher payloads seem to have better success rates in any specific orbit types
- Some orbits such as ISS and GTO have mixed success rates.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- This graph shows increase in success rate with passing of time.



All Launch Site Names

- Find the names of the unique launch sites
  - %sql select distinct Launch_Site from SPACEXTABLE
- Present your query result with a short explanation here
- These are the 4 Launch sites in the data

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here
  - %sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer = 'NASA (CRS)'

TOTAL PAYLOAD MASS

45596

- Present your query result with a short explanation here
  - They key is to use aggregation function 'sum' in the query, and limiting the Customer to NASA (CRS)
  - Total Payload Mass for NASA CRS Rockets was 45596 KG

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here
  - %sql select avg(PAYLOAD_MASS__KG_) as `F9v1.1 PAYLOAD` from SPACEXTABLE where Booster_Version = 'F9 v1.1'
  - Average Payload for F9 v1.1 booster was 2534 KG

F9v1.1 AVG PAYLOAD
2534.6666666666665

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
    - %sql select min(Date) as `OLDEST LAUNCH DATE` from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)'
- Present your query result with a short explanation here
    - This query also uses the aggregation ‘min’ function to find oldest data of successful launch

OLDEST LAUNCH DATE

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - %sql select distinct(Booster_Version) as `BOOSTER VERSION` from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
- Present your query result with a short explanation here

BOOSTER VERSION

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - %sql select count(*) as `SUCCESSFUL MISSIONS` from SPACEXTABLE where Mission_Outcome = 'Success'
  - %sql select count(*) as `FAILED MISSIONS` from SPACEXTABLE where Mission_Outcome <> 'Success'
- Present your query result with a short explanation here

SUCCESSFUL MISSIONS	FAILED MISSIONS
98	3

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here
  - %sql select distinct(Booster_Version) as `MAX PAYLOAD BOOSTERS` from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
  - This query uses function distinct and a subquery.
  - Seems like F9 B5 version boosters have carried max palyload mass

MAX PAYLOAD BOOSTERS
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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here
  - %sql select substr(Date, 6,2) as 'Month', substr(Date,0,5) as 'Year' from SPACEXTABLE where substr(Date,0,5) = '2015' and Landing_Outcome like 'Failure%' limit 5
  - Results show in year 2015, only months of Jan and Apr had Failures

Month	Year
01	2015
04	2015

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

- 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
- Present your query result with a short explanation here
  - The results show that No Attempt was the major Landing Outcome of missions between 2010-06-04 and 2017-03-20, followed by Drone Ship Success. Least frequent outcome was Precluded (Drone Ship)

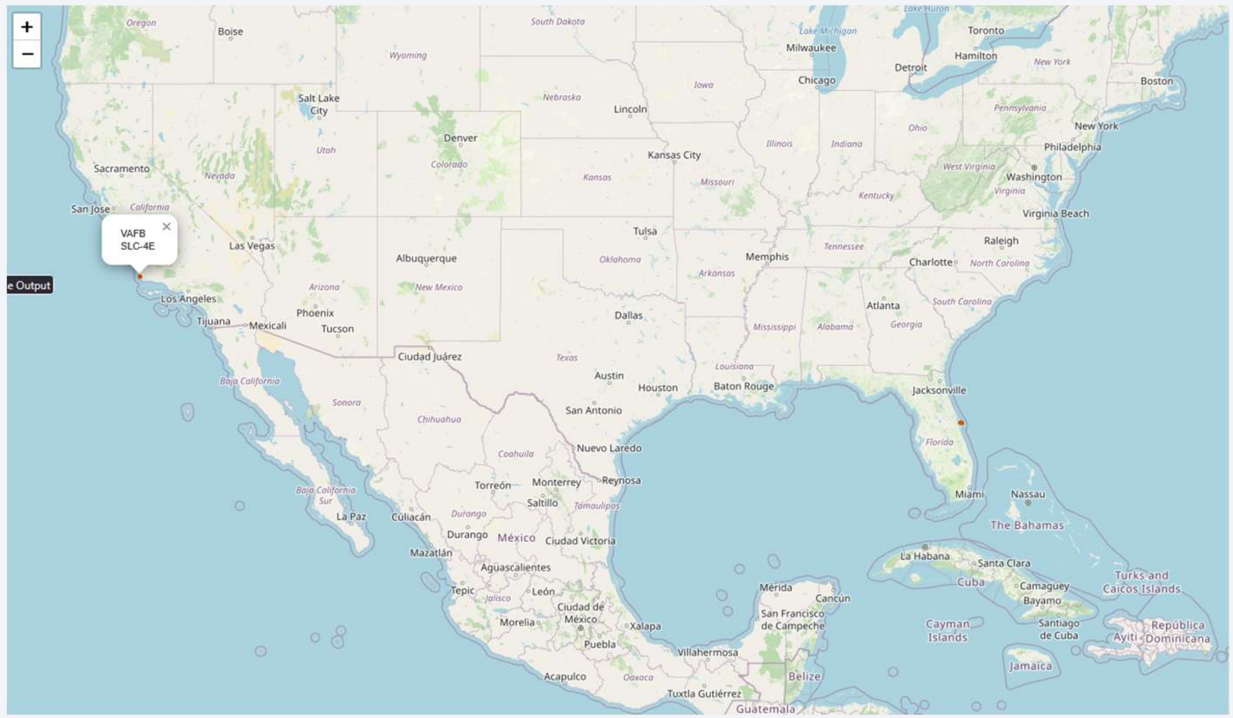
Landing_Outcome	OUTCOME
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

- Launch sites are mainly located on East or West Coast

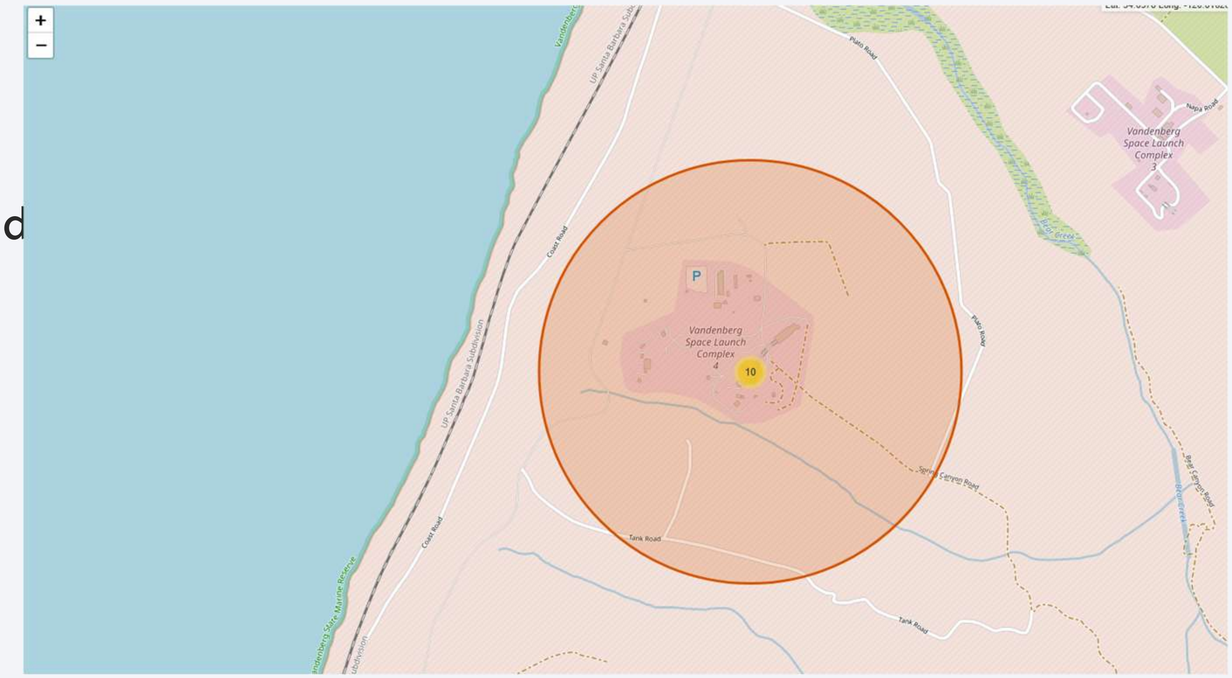


Success/Failure for Launc Sites

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Map shows Success (green) and Failure(red) outcomes of launches from California site

Landmarks near California Launch Site

- Launch Site VAFB SLC-4E Proximity to Landmarks
- Distance between launcsite and nearest coastline is 1.390 Km



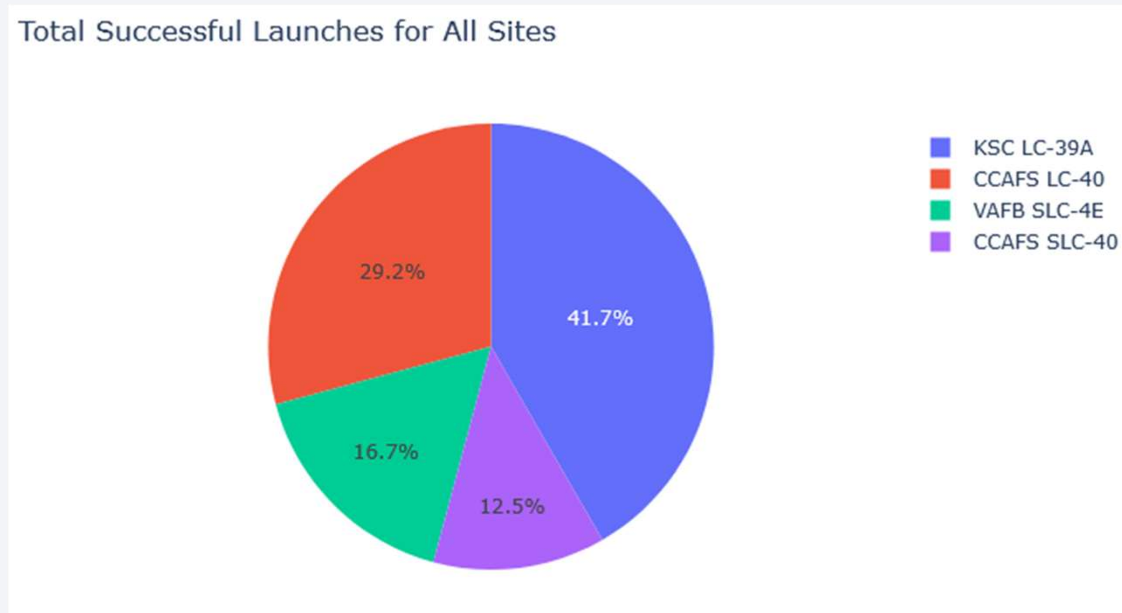
Section 4

Build a Dashboard with Plotly Dash

Success/Failure per site

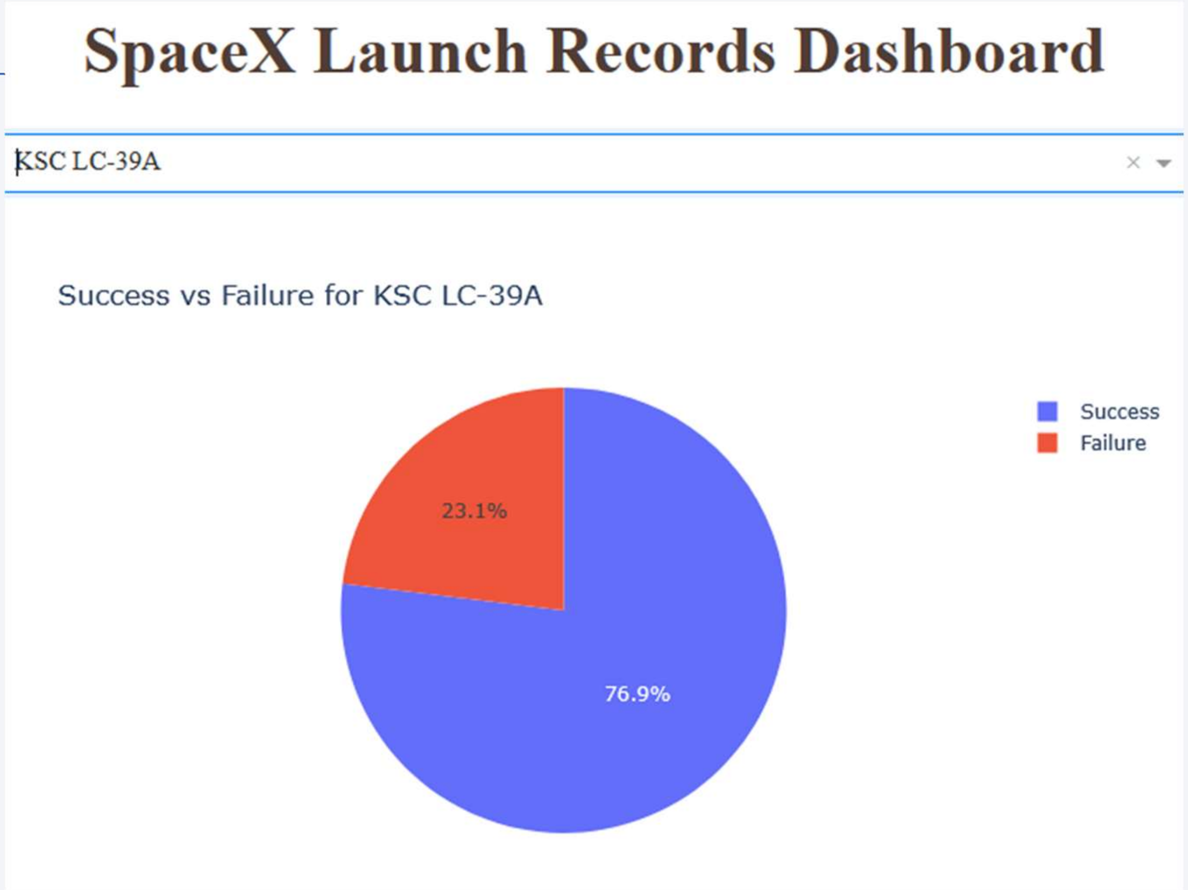
- The piechart shows success rate for different sites

- Site KSC has highest success rate which site CCAFS has least.



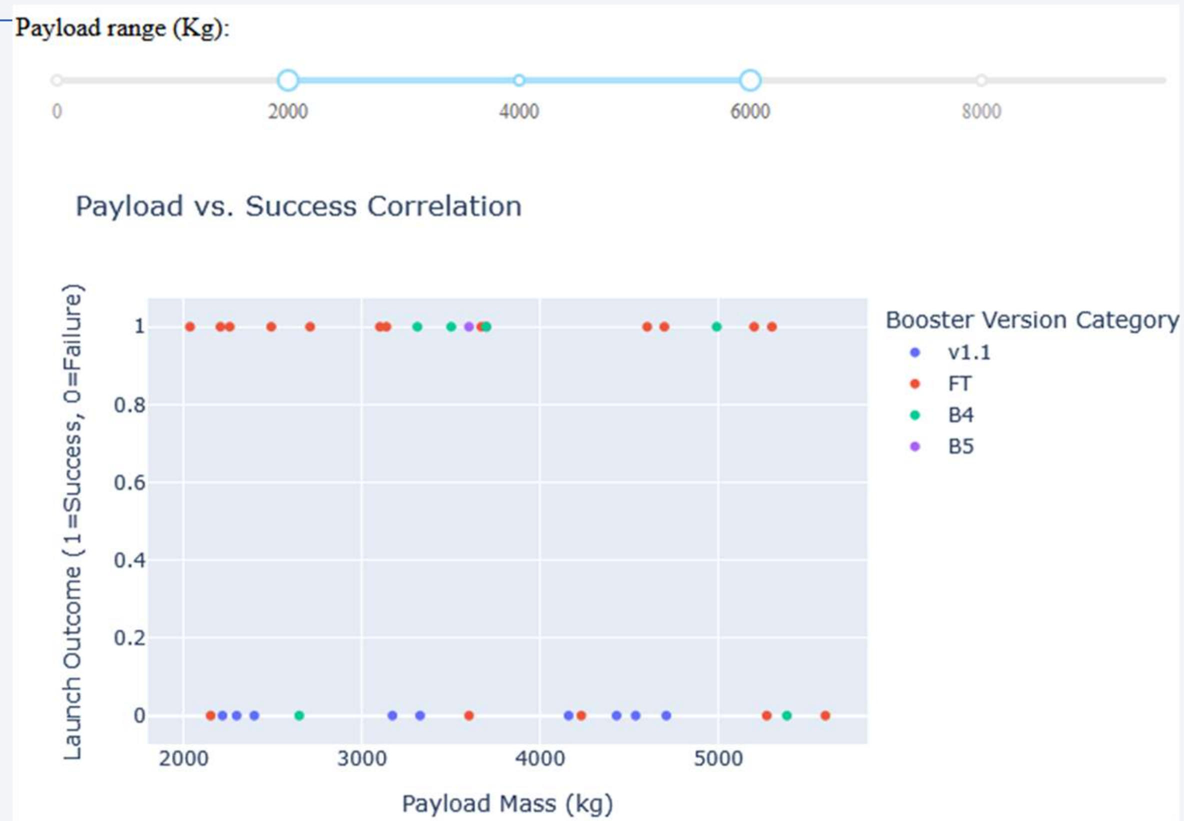
Highest success site

Site KSC LC-39A has 77% Success rate



Payload vs Success Correlation

- Figure displays plot for all launch sites for payload between 2000 and 6000 KGs

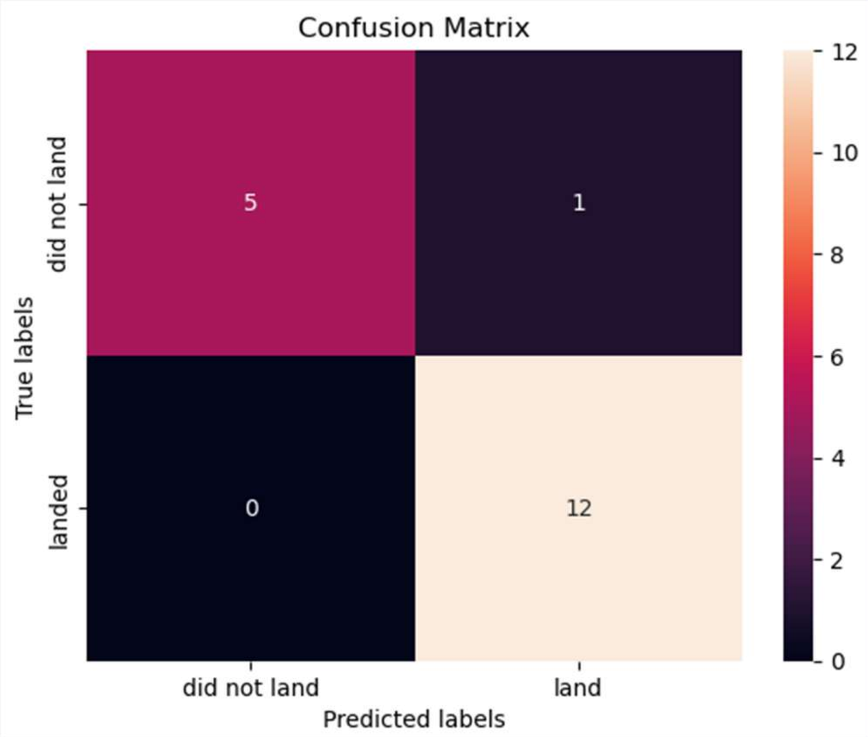


Section 5

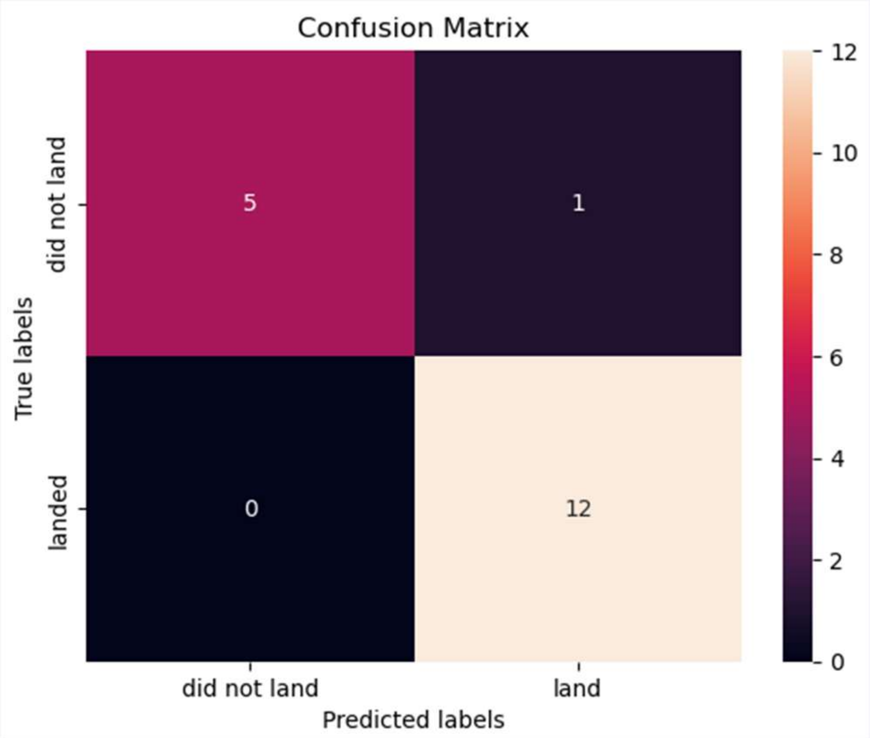
Predictive Analysis (Classification)

Classification Accuracy

- Models Assesses
  - Logistic Regression
  - SVM
  - Decision Tree
  - K-Nearest Neighbor
- Model with highest accuracy was SVM with a 94.44% accuracy



Confusion Matrix



Conclusions

- Failure Rate depends on factors such as Orbit and payload mass
- It was noticed though that recent launches have a success rate independent of payload mass
- Some sites were noticed to have a higher success rate compared to others
- Certain classifiers are better suited for predicting Success/Failure rate compared to others
- In the model, SVM performed as best classifier for predicting success/failure rate

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


Thank you!