IBM Developer SKILLS NETWORK

Applied DS Capstone Project

Data Driven Space Exploration:

Predicting Falcon9 Landings

Presented By:
Harini Sampat
21st December, 2024

OUTLINE


Executive Summary
Introduction
Methodology
Results
Conclusion
Appendix

EXECUTIVE SUMMARY

Methodology Adopted:

- Data Collection
- Data Wrangling
- Exploratory Data Analysis with:
  - Data Visualization
  - SQL
- Interactive Map with Folium
- Dashboard with Plotly Dash
- Predictive Analysis - Classification

Results

- Insights drawn from EDA
- Launch Sites Proximities Analysis
- Dashboard with Plotly Dash
- Predictive Analysis

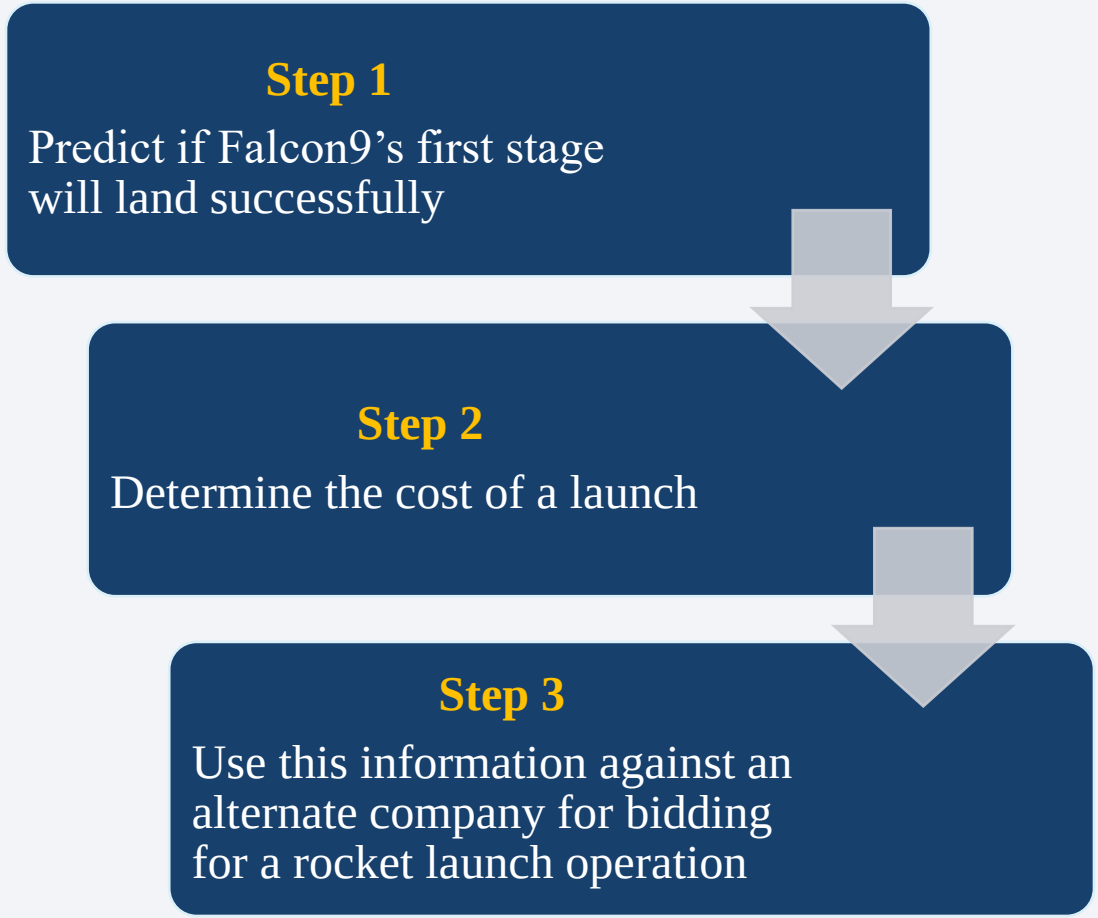
INTRODUCTION

Space Industry Today:

- Lead by SpaceX's Falcon9 Rocket
- Due to its cost-effective launches priced at $62 million
- 62% lesser than others

Reason for the cost efficiency?

- Reusable rocket stages



INTRODUCTION

PROBLEM STATEMENT

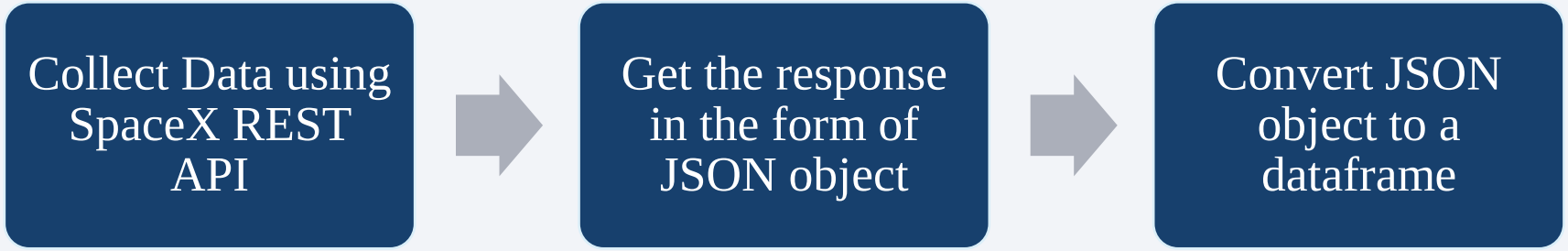
- ✓ For a given set of features about Falcon9 rocket launch (i.e. Payload Mass, Orbit Type, Launch specifications, Landing specifications, Landing Outcome etc.),
  Will the first stage of the rocket land successfully?

- ✓ What is the best algorithm that can be used for classification tasks of this dataset?

METHODOLOGY

DATA COLLECTION

Method 1: SpaceX API



https://api.spacexdata.com/v4/launches/past

Method 2: Scraping

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

DATA COLLECTION – SpaceX API

Define helper functions to extract information about booster name, payload details etc.

Request and parse the SpaceX launch data using GET request

Create a Pandas Dataframe to organize the data

Filter the Dataframe to only include “Falcon9” launches

Reset the “FlightNumber” column

➢ **Output:** Dataframe with 90 rows x 17 columns

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093

Figure: *Dataframe with data extracted using SpaceX API*

GitHub Link: *Lab1_DataCollectionwithAPI*

DATA COLLECTION – Scraping

Define helper functions to extract information about booster name, payload details etc.

Request Falcon9 Wikipedia Launch page from its URL

Use BeautifulSoup Object to extract the column names from the table

Create a Dataframe by parsing the launch HTTML tables

Reset the "FlightNumber" column

➢ **Output:** Dataframe with 121 rows x 11 columns

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
1	1	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.07B0003.18	Failure	4 June 2010	18:45
2	2	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.07B0004.18	No attempt\n	8 December 2010	15:43
3	3	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.07B0005.18	No attempt	22 May 2012	07:44
4	4	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.07B0006.18	No attempt\n	8 October 2012	00:35

Figure: *Dataframe with data extracted using Webscraping*

GitHub Link: *Lab2_DataCollection_WebScraping*

DATA WRANGLING

Check for Missing Values

PayloadMass = 5

LandingPad = 26

Use replace function

Replace NaN values in PayloadMass with its mean

Classify Landing Outcomes

Good = successful landing

Bad = unsuccessful landing

Create a column "Class"

0 = Bad landing outcome

1 = Good landing outcome

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Figure: *Dataframe post data wrangling*

GitHub Link: *Lab3_DataWrangling*

EDA WITH DATA VISUALIZATION

Libraries Used

1. Matplotlib

- A plotting library
- Used for creating interactive visualizations in Python

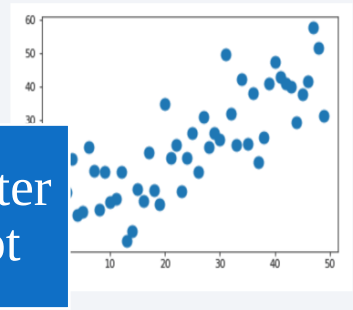
2. Seaborn

- A data visualization library based on Matplotlib
- Used to plot statistical graphs

GitHub Link: *Lab4_EDAwithVisualization*

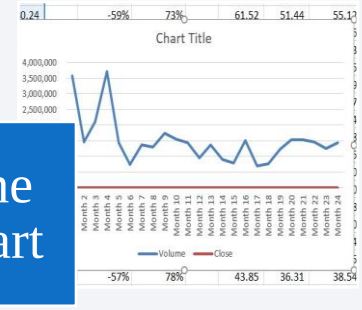
Types of Charts used:

Scatter Plot



Bar Chart

Line Chart



EDA WITH DATA VISUALIZATION

Following charts were plotted:

- Scatter Plot between Flight Number and Launch Site
- Scatter Plot between Payload Mass and Launch Site
- Bar Chart for the success rate of each Orbit
- Scatter Plot between Flight Number and Orbit type
- Scatter Plot between Payload Mass and Orbit type
- Line Chart for the launch success yearly trend

GitHub Link: *Lab4_EDAwithVisualization*

EDA WITH SQL

Following SQL queries were executed:

- Names of unique launch sites used
- Records with launch sites beginning with the string ‘CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date of first successful landing in ground pad
- Names of boosters which have success in drone ship and payload mass greater than 4000 and lesser than 6000
- Total number of success and failure outcomes
- Names of boosters carrying the maximum payload mass
- Ranking of landing outcomes

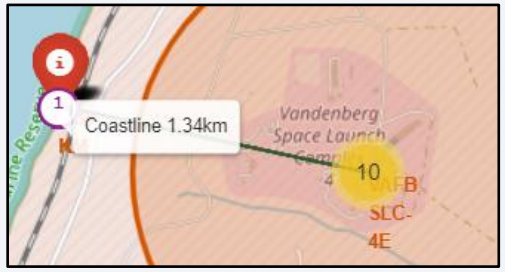
GitHub Link: *Lab5_EDAwithSQL*

INTERACTIVE MAPS WITH FOLIUM

➢ **Folium Library**: Used to visualize the data through interactive maps

Marker

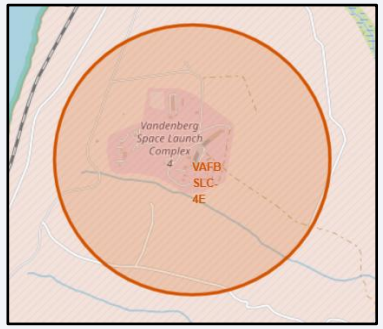
- Used to mark the location of the launch sites



Polyline

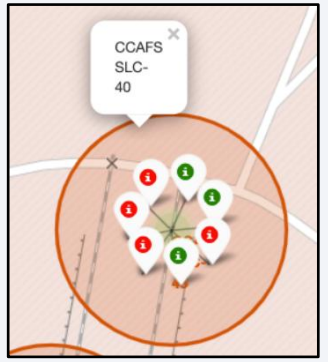
- Used to connect the launch site to other points of interest (e.g.: coastline)

Map Objects Used



Circle

- Used to highlight the region around the marker



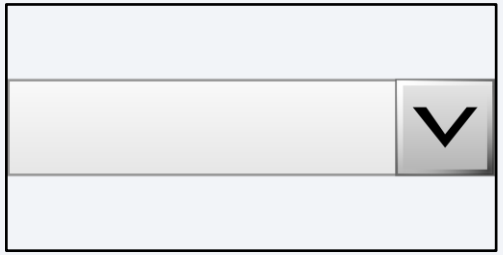
Marker Cluster

- Group nearby markers for enhanced readability

GitHub Link: *Lab6_VisualAnalyticswithFolium*

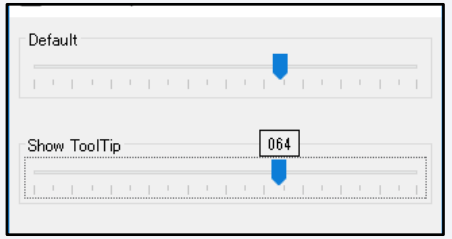
DASHBOARD WITH PLOTLY DASH

➢ **Plotly Dash**: A Python framework used to create interactive dashboards



Dropdown

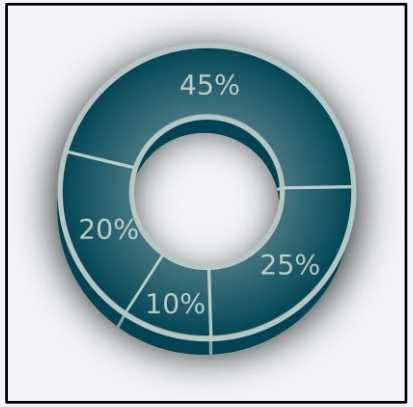
- Used to filter the data to select a particular launch site



Range Slider

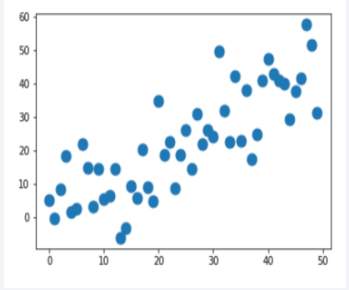
- Used to adjust the range of the values to selectively view data

Elements used



Pie Chart

- Used to display the success rate of the launches

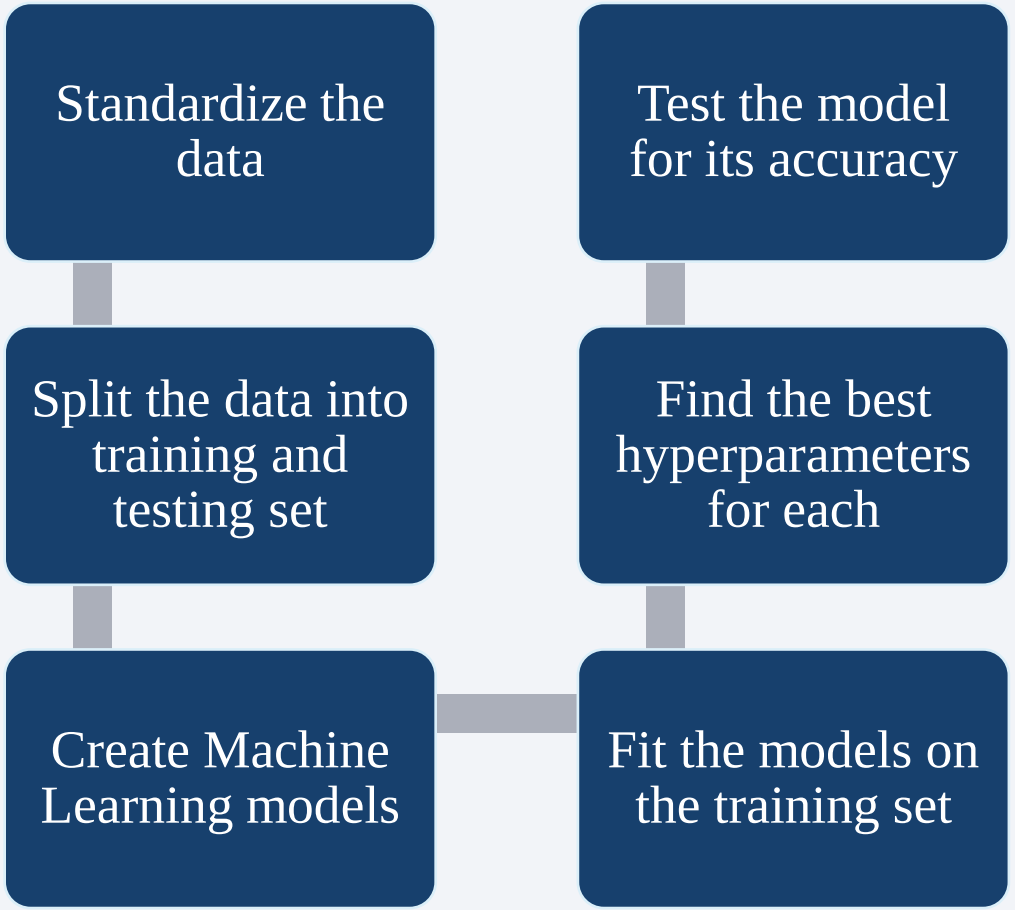


Scatter Plot

- Used to show the correlation between two variables

GitHub Link: *Lab7_spacex_dash_app*

PREDICTIVE ANALYSIS – CLASSIFICATION



GitHub Link: *Lab8_MachineLearningPrediction*

Machine Learning models used:

1. Logistic Regression
2. Support Vector Machines (SVM)
3. Decision Tree
4. K – Nearest Neighbors (KNN)

RESULTS

SECTION 1

INSIGHTS DRAWN FROM EDA

FLIGHT NUMBER v/s LAUNCH SITE

Observations:

As the flight number increases, the landings become more successful (represented by the orange dots)

- At Site “CCAFS SLC 40”, beyond Flight Number = 60, most of the landings are successful

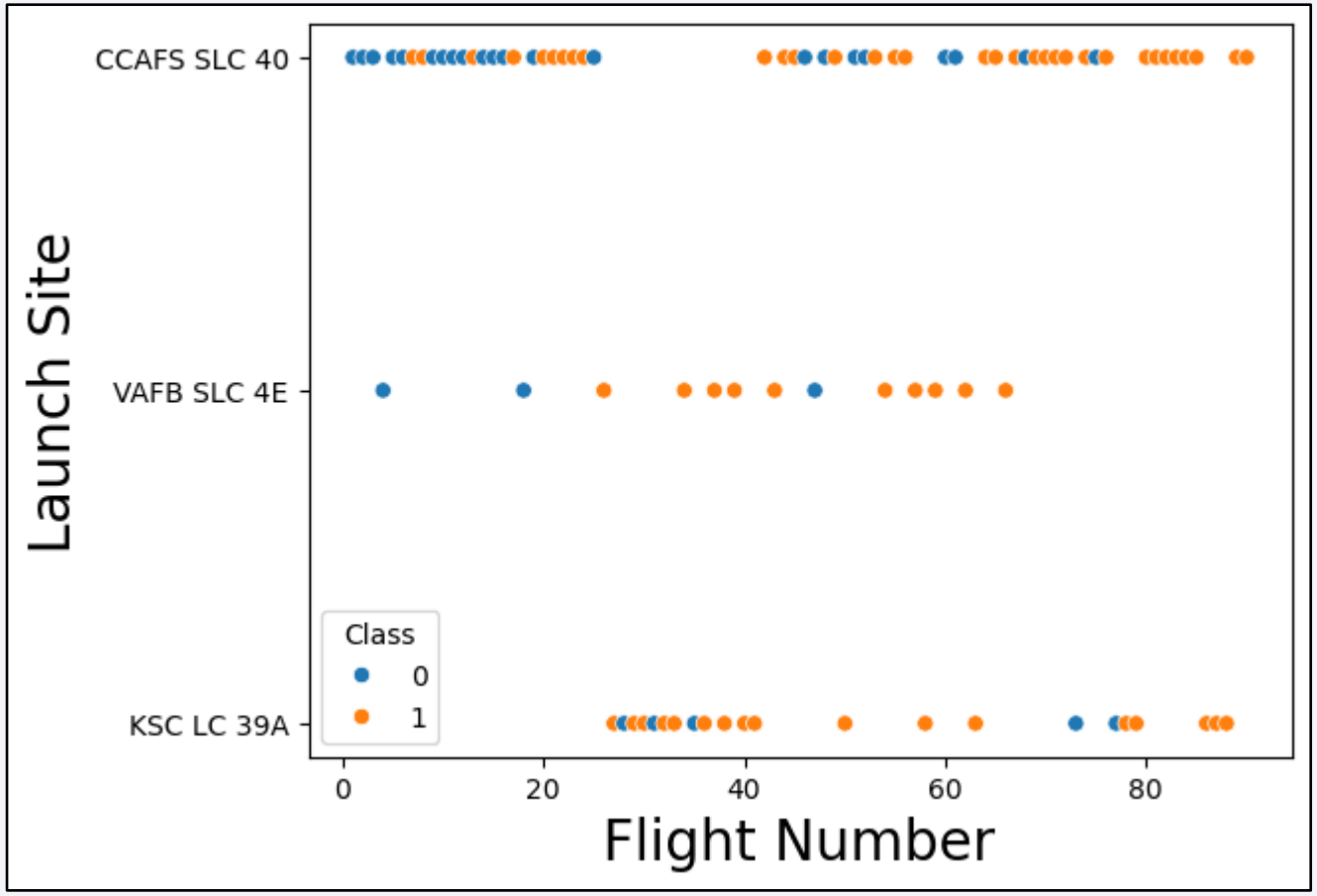


Chart: *Scatter Plot of Flight Number v/s Launch Site*

PAYLOAD v/s LAUNCH SITE

Observations:

- For the launch site "VAFB SLC 4E", there are no rockets launched for heavy payload mass (i.e., greater than 10,000 kg)

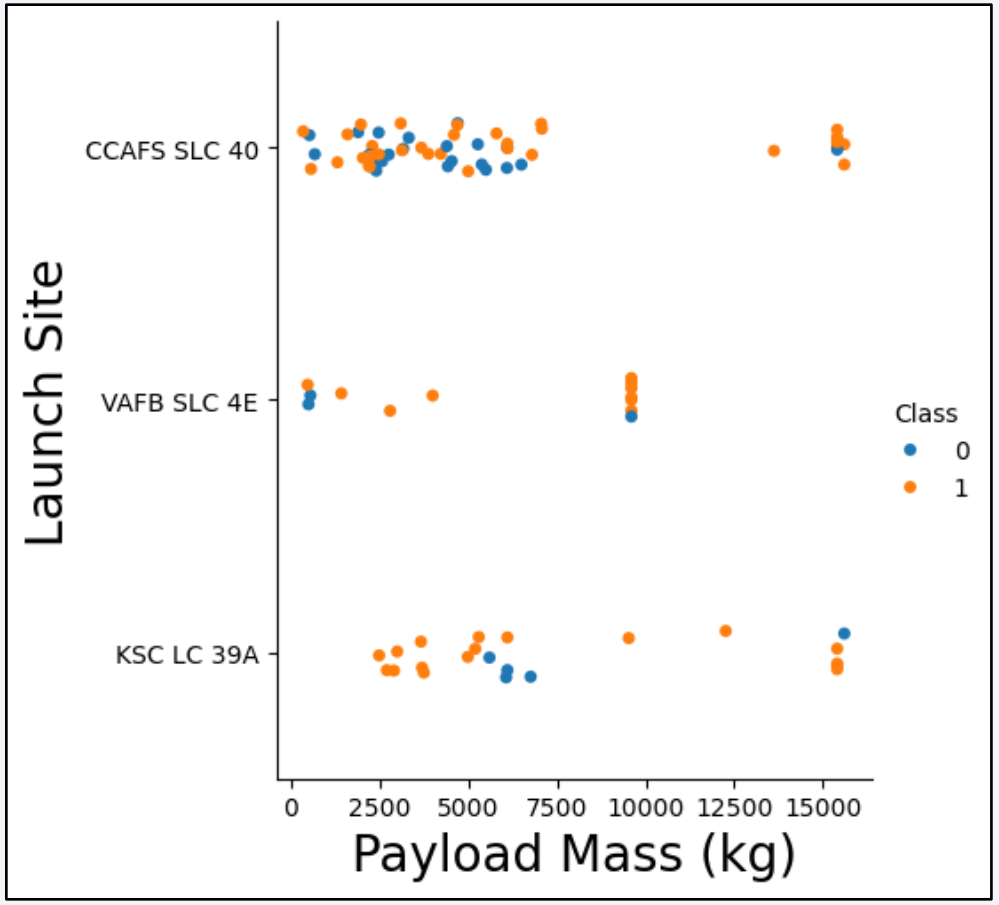


Chart: *Scatter Plot of Payload Mass v/s Launch Site*

ORBIT TYPE v/s SUCCESS RATE

Observations:

- The following orbits have a 100% Success Rate (i.e. all rockets in this orbit have successfully landed):
  1. ES – L1
  2. SSO
  3. HEO
  4. GEO
- Orbit with 0% Success Rate: SO

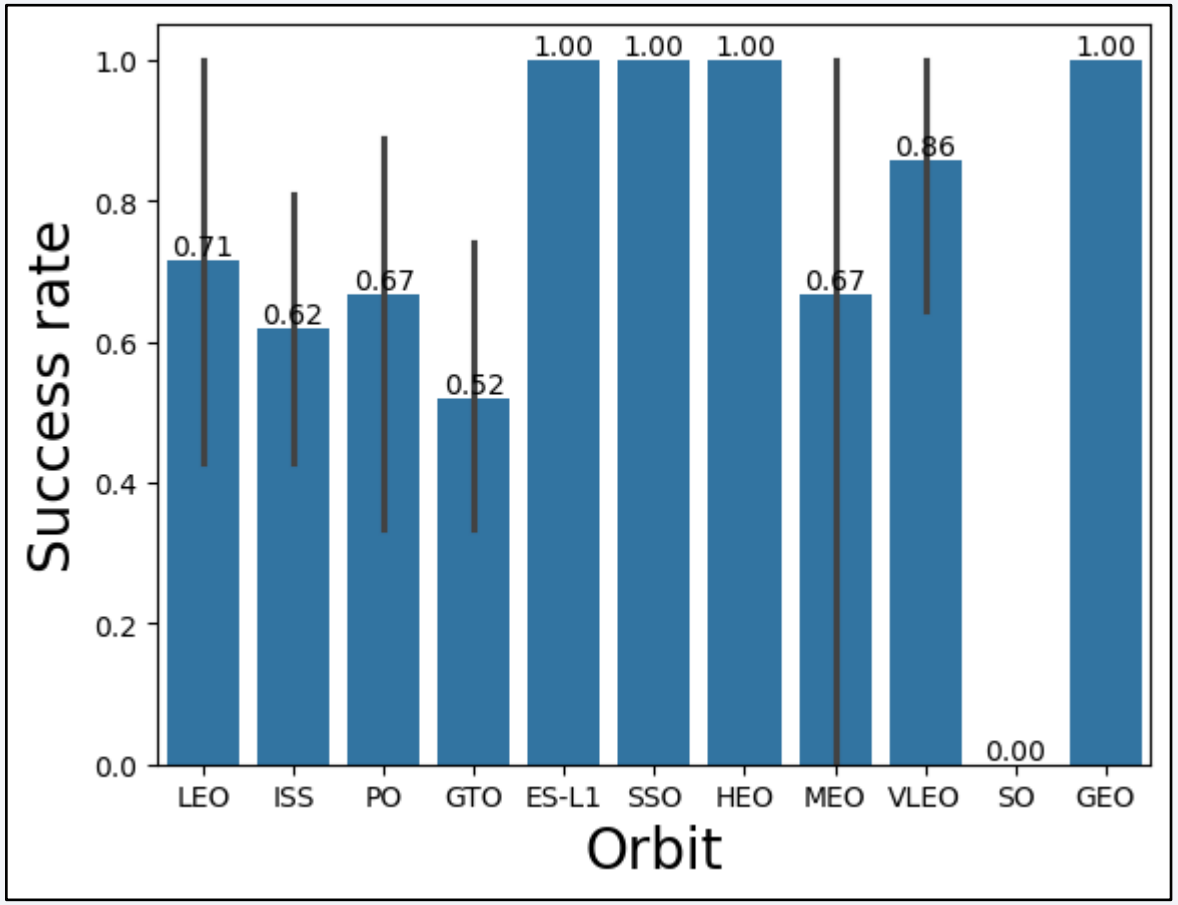


Chart: *Bar Plot of Orbit Type v/s Success Rate*

FLIGHT NUMBER v/s ORBIT TYPE

Observations:

- In Orbit “Leo”, success has a positive relationship with Flight number
- In Orbit “GTO”, there is no relationship of success with Flight number

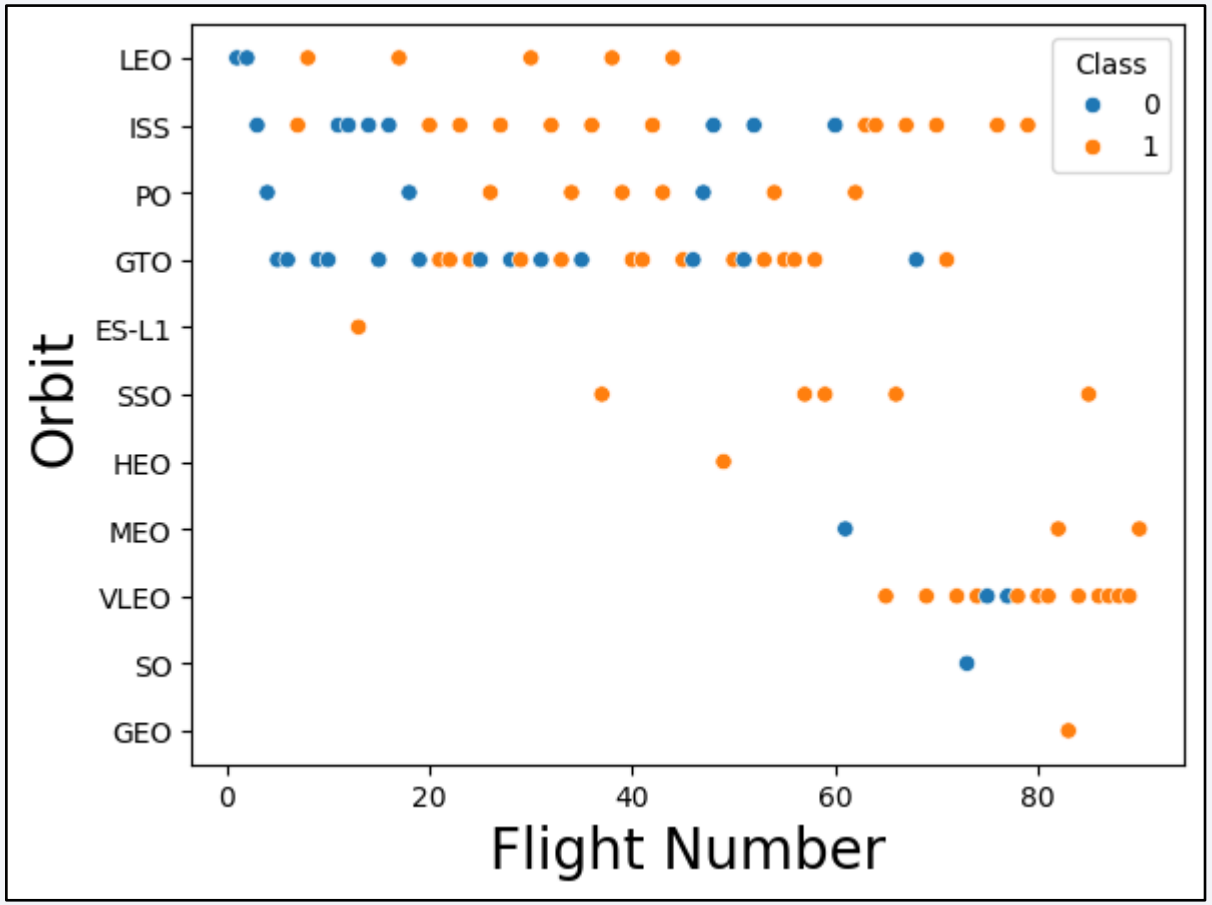


Chart: *Scatter Plot of Flight Number v/s Orbit Type*

PAYLOAD v/s ORBIT TYPE

Observations:

- With heavy payload masses, the successful landings are more for the following orbits:

1. PO
2. LEO
3. ISS

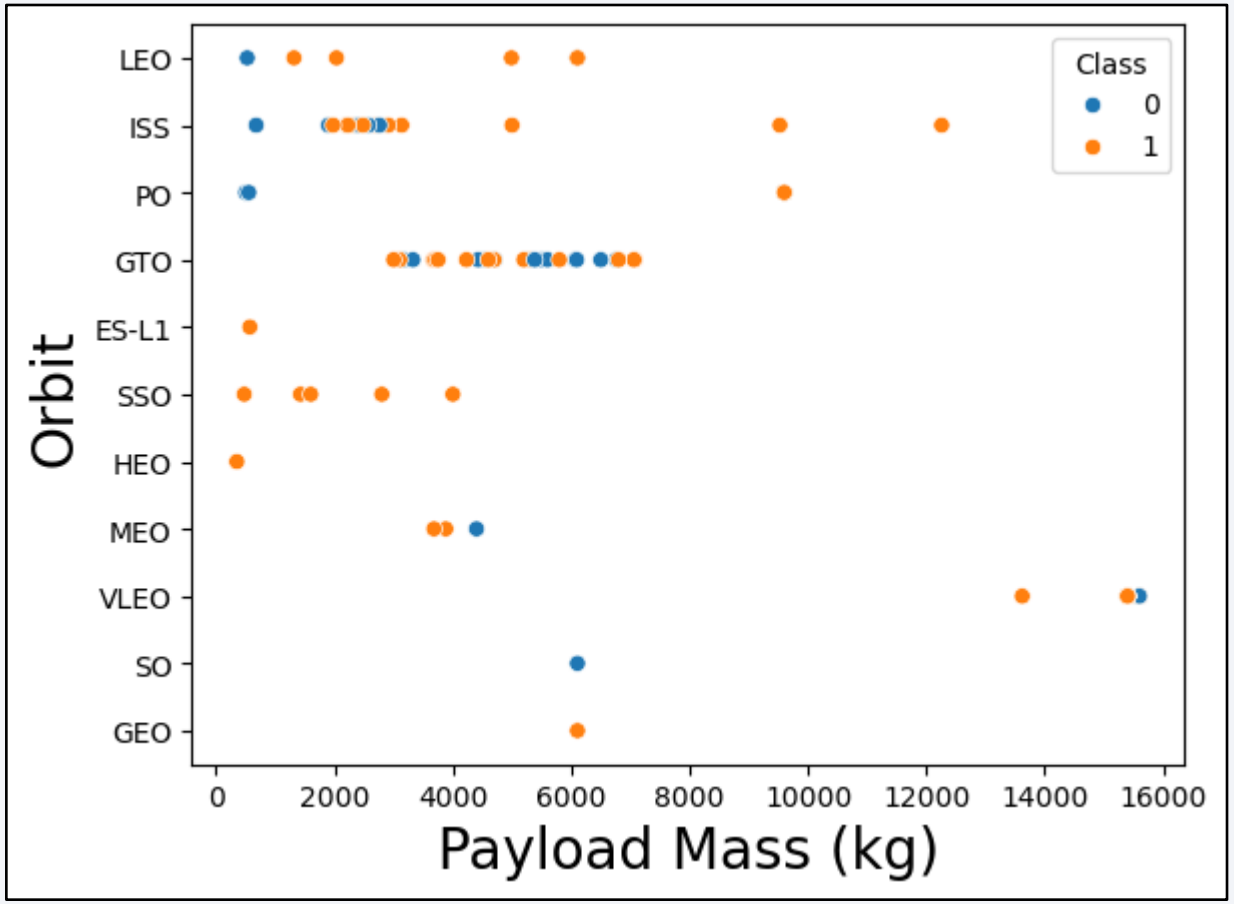


Chart: *Scatter Plot of Payload Mass v/s Orbit Type*

LAUNCH SUCCESS YEARLY TREND

Observations:

- Since 2013, there has been an upward increasing trend in the success rate of the landings till 2019.

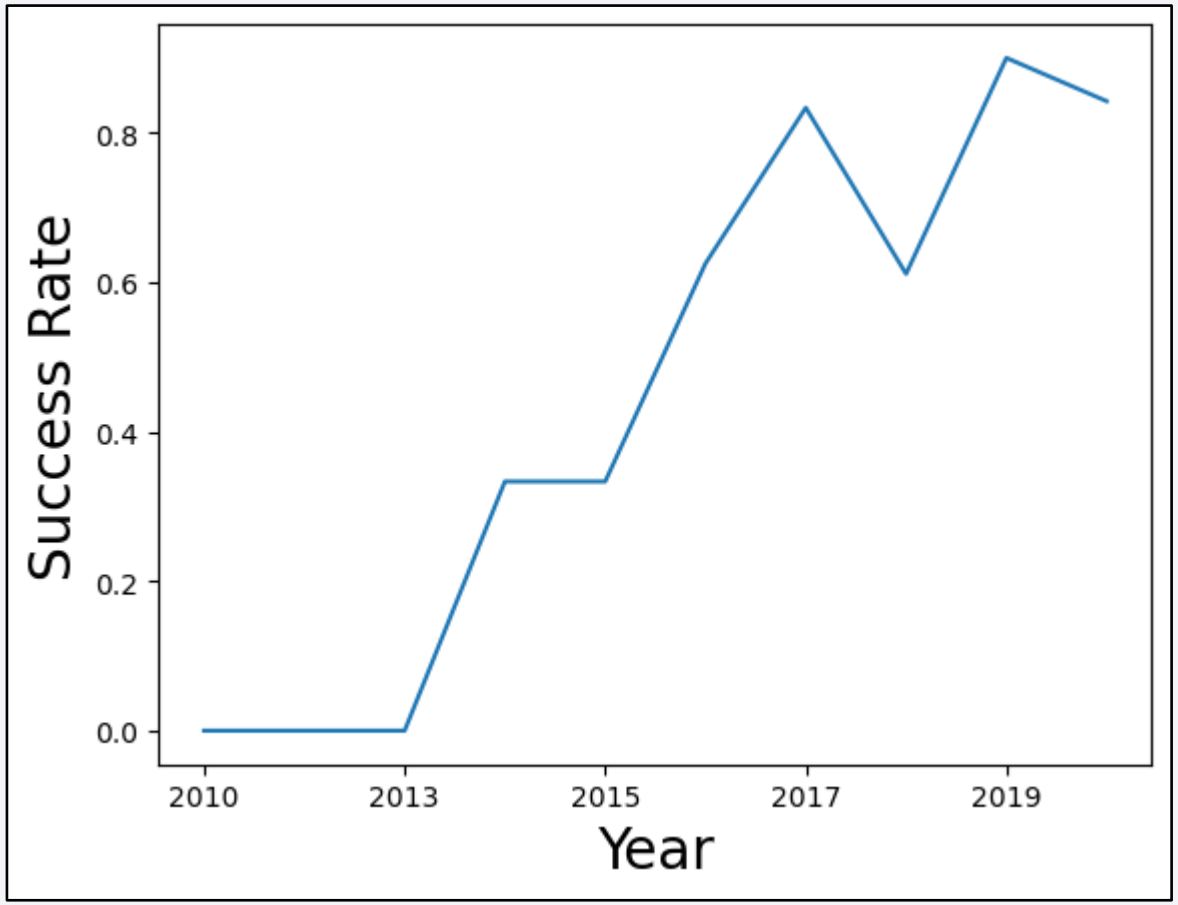
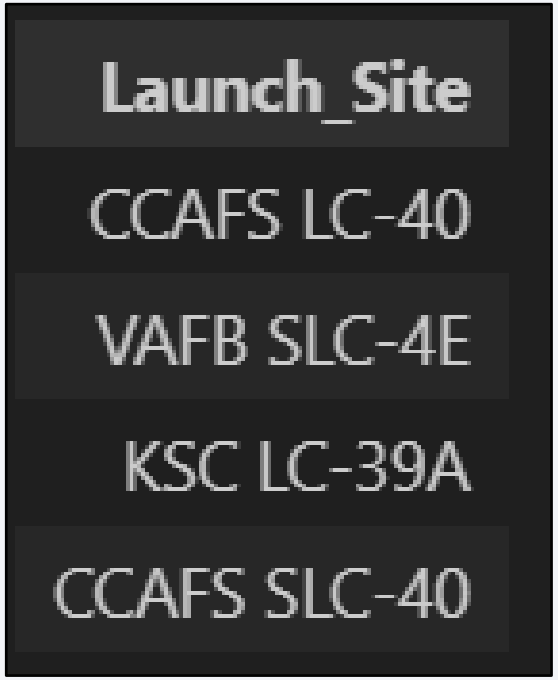


Chart: *Line Plot of Year v/s Success Rate*

ALL LAUNCH SITE NAMES

There are 4 unique Launch Sites extracted using the **“DISTINCT”** keyword in SQL:

1. CCAFS LC – 40
2. VAFB SLC – 4E
3. KSC LC – 39A
4. CCAFS SLC - 40



Output: *All Launch Site Names*

LAUNCH SITE NAMES STARTING WITH 'CCA'

The first 5 records with Launch Sites beginning with "CCA" are:

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

Output: *Dataframe records with launch site names starting with 'CCA'*

PAYLOAD MASS DETAILS

Total Payload Mass carried by boosters launched by NASA (CRS) is **45596 kg**.

sum("PAYLOAD_MASS__KG_")
45596

Output: *Total payload mass carried by NASA (CRS) boosters*

The average payload mass carried by booster version F9 v1.1 is **2928.4 kg**

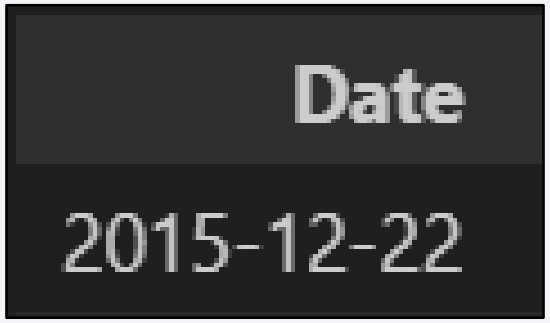
avg("PAYLOAD_MASS__KG_")
2928.4

Output: *Average payload mass carried by booster F9 v1.1*

FIRST SUCCESSFUL GROUND LANDING DATE

The first successful landing outcome on ground pad was recorded on

“22nd December 2015”



Output: *Date of first successful ground pad landing*

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD RANGE

There are 4 boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

1. F9 FT B1022
2. F9 FT B1026
3. F9 FT B1021.2
4. F9 FT B1031.2

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

Output: *Names of boosters with successful drone ship landing within payload range of 4000-6000kg*

TOTAL SUCCESSFUL & FAILURE MISSION OUTCOMES

Based on the output table:

- Total number of successful mission outcomes = 100
- Total number of failure mission outcomes = 1

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

Output: *Total Successful and Failure mission outcomes*

BOOSTERS CARRYING MAXIMUM PAYLOAD

- Maximum Payload Mass = 15600 kg
- Number of boosters carrying maximum payload mass = 12

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

Output: *Boosters carrying maximum payload*

LAUNCH RECORDS OF 2015

There are 2 records with failure outcome in landing in the drone ship in the year 2015:

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Output: *Records with failure landing outcomes in drone ship in 2015*

RANK OF LANDING OUTCOMES

Ranking of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order is:

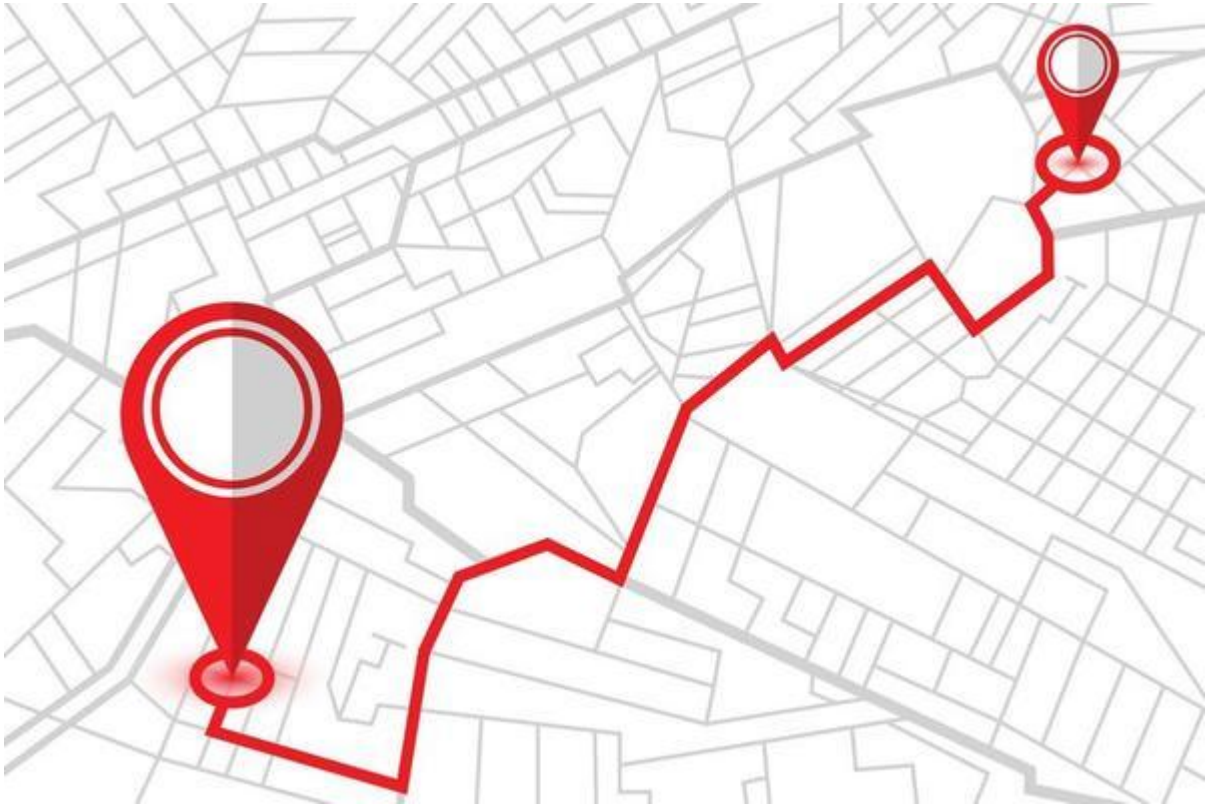
- 10 rockets made no landing attempts
- 8 rockets successfully landed on drone ship and ground pad collectively
- 7 rockets failed to land on drone ship and ground pad collectively

Landing_Outcome	count(Landing_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

Output: *Ranking of landing outcomes between 4th June 2010 and 20th March 2017*

SECTION 2

LAUNCH SITES' PROXIMITIES ANALYSIS



LOCATION MARKERS FOR LAUNCH SITES

4 Markers locating the launch sites:

- VAFB SLC – 4E: West Coast
- KSC LC – 39A: East Coast
- CCAFS LC – 40: East Coast
- CCAFS SLC – 40: East Coast

➢ All launch sites are near the Equator and Coastlines

Figure: *Map showing location markers of all 4 launch sites*

Note: *On the east coast, 3 launch site markers are overlapping*

COLOR CODED LAUNCH OUTCOMES

Colored Markers are used to analyze the success rates of launch sites:

- **Red Marker** = Unsuccessful Launch
- **Green Marker** = Successful Launch

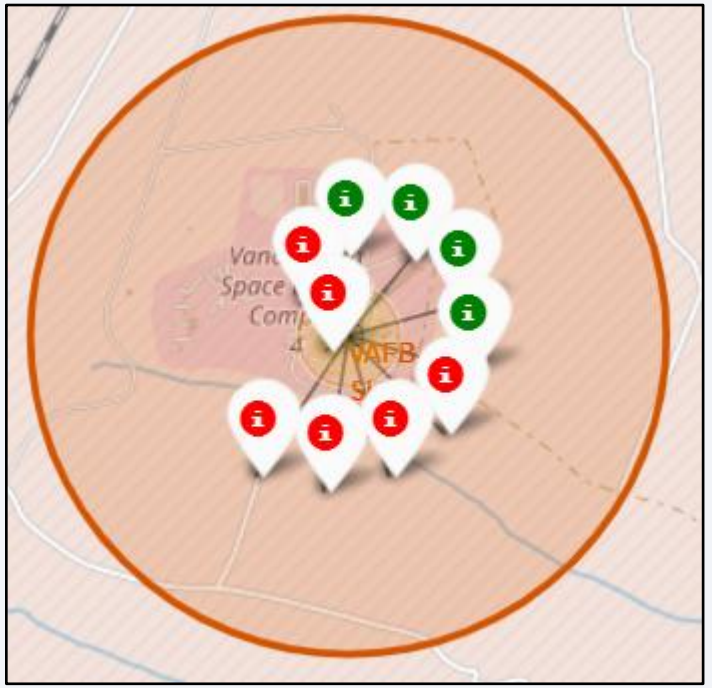


Figure: *Successful and Unsuccessful launch outcomes for a particular launch site – represented by Markers*

DISTANCE FROM KSC LC – 39A TO ITS PROXIMITIES

Launch Site: **KSC LC – 39A**

- ➢Distance to Railway: 15.32 km
- ➢Distance to Highway: 20.28 km
- ➢Distance to Coastline: 14.99 km
- ➢Distance to Titusville (closest city): 16.32 km
  - ✓Well positioned site with optimal safety and logistical efficiency

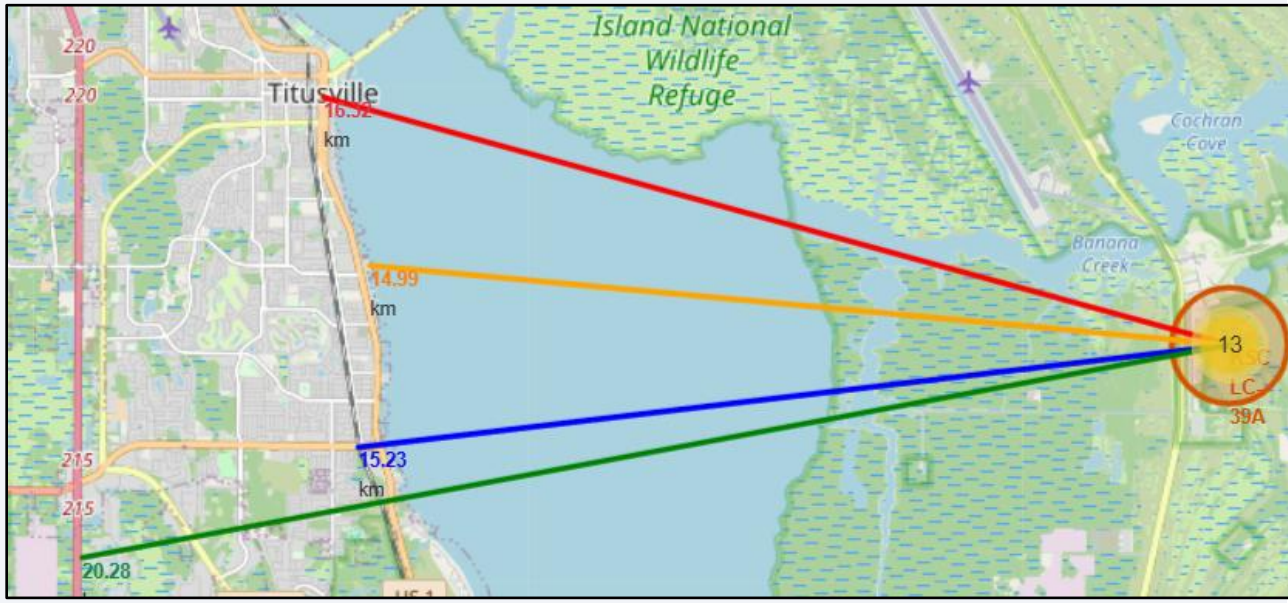
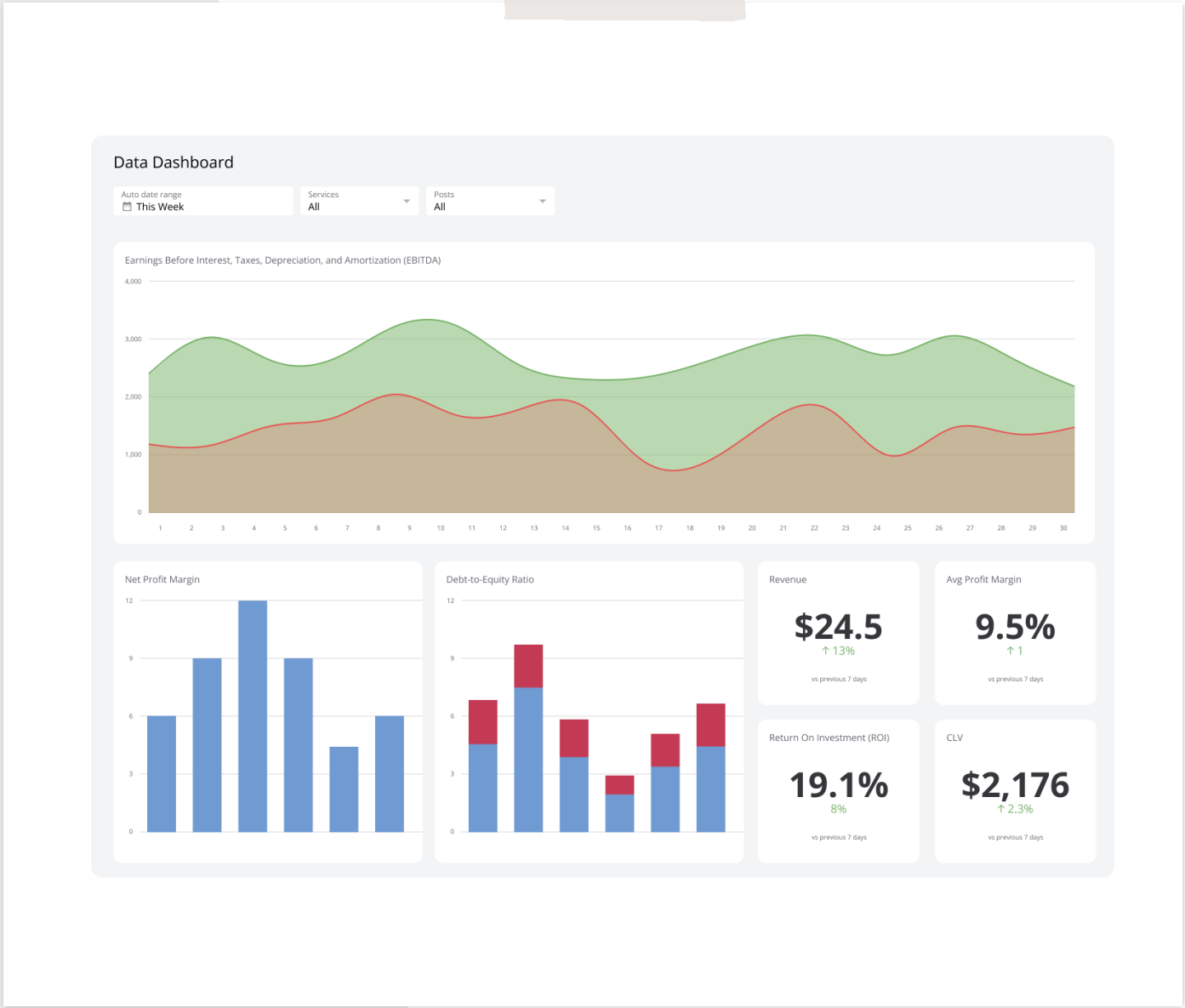


Figure: *Map showing polylines indicating the distances between launch site and nearby proximities*

SECTION 3

DASHBOARD WITH PLOTLY DASH



LAUNCH SUCCESS PIE CHART FOR ALL SITES

Observation:

➢KSC LC – 39A (represented by blue slice): Max launch success = 41.7%

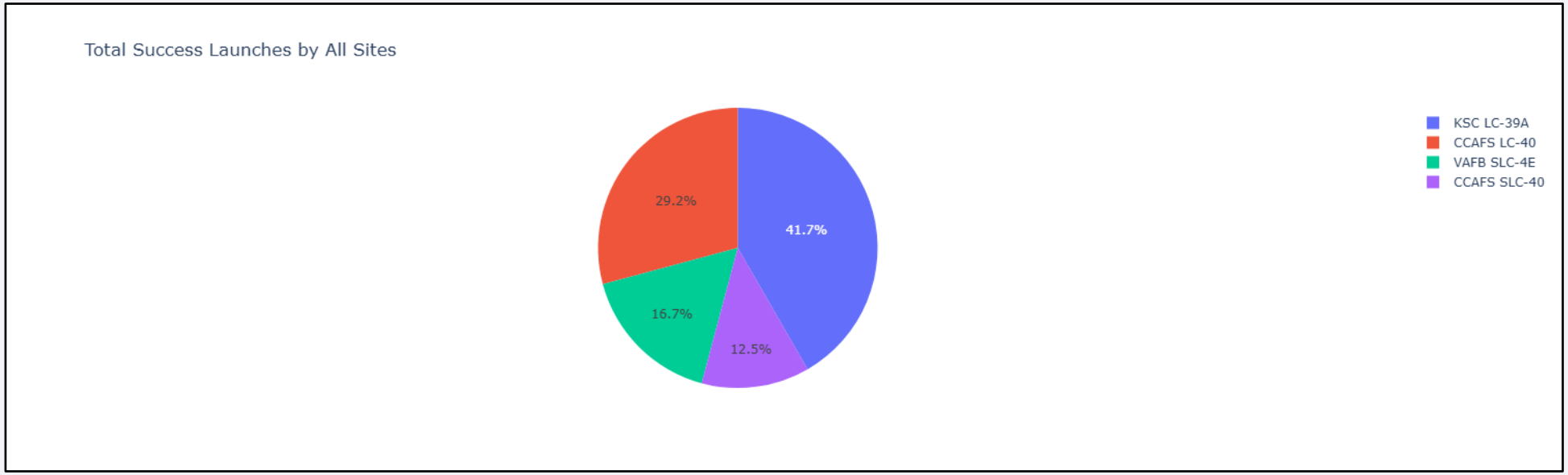


Chart: *Pie Chart showing launch success percentage for all sites*

SUCCESS RATE OF KSC LC – 39A

- Launch site with highest launch success ratio: KSC LC – 39A
- Success Rate = 76.9%

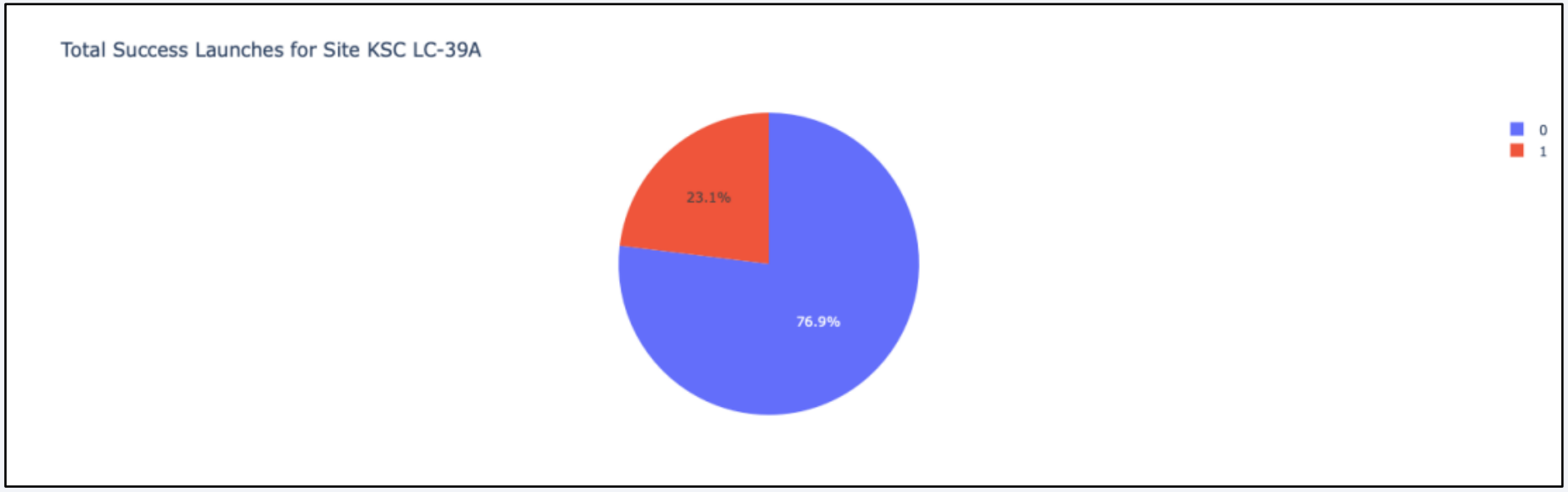


Chart: *Pie Chart showing launch success percentage for KSC LC – 39A*

SCATTER PLOT FOR CORRELATION

Launch Site with highest success rate: KSC LC – 39A

- **Observation:** Booster Version = FT has the highest success for payload range between 0 to 10000kg

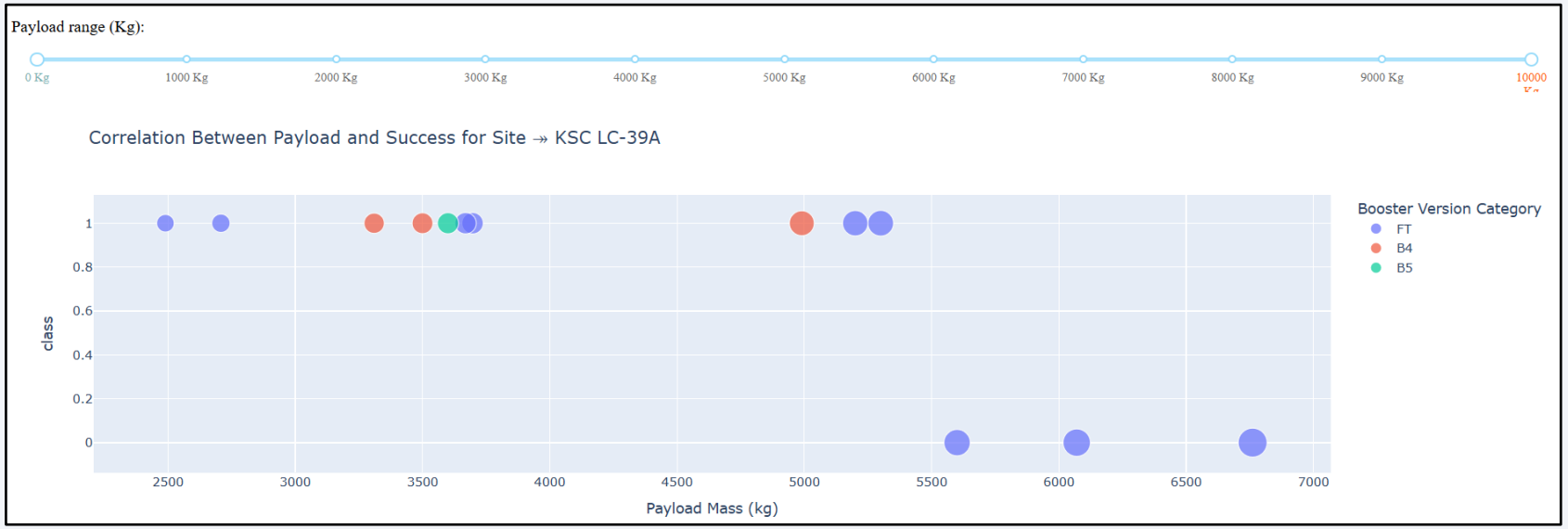


Chart: *Scatter Plot showing correlation between payload mass and launch class for KSC LC – 39A*

SECTION 3

PREDICTIVE ANALYSIS (CLASSIFICATION)

CLASSIFICATION ACCURACY

Model	Accuracy
Logistic Regression	84.64%
SVM	84.82%
Decision Tree	**86.25%**
KNN	84.82%

➢ Best Model: Decision Tree

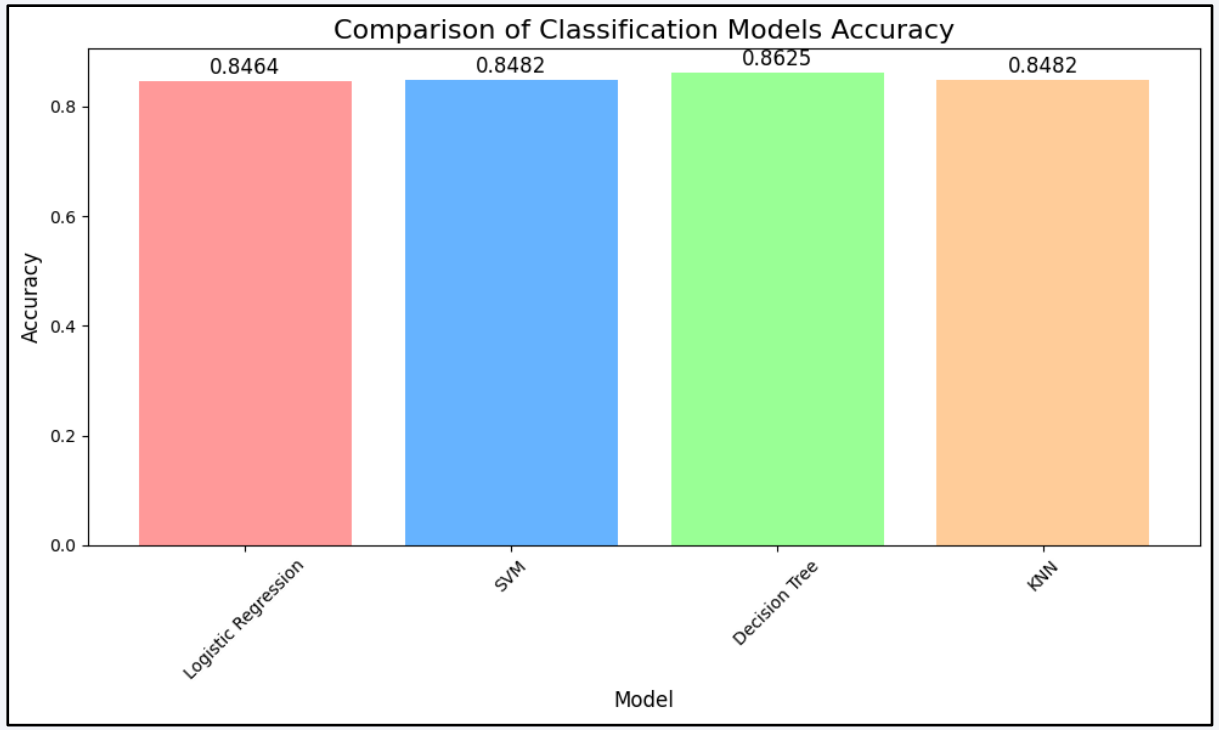
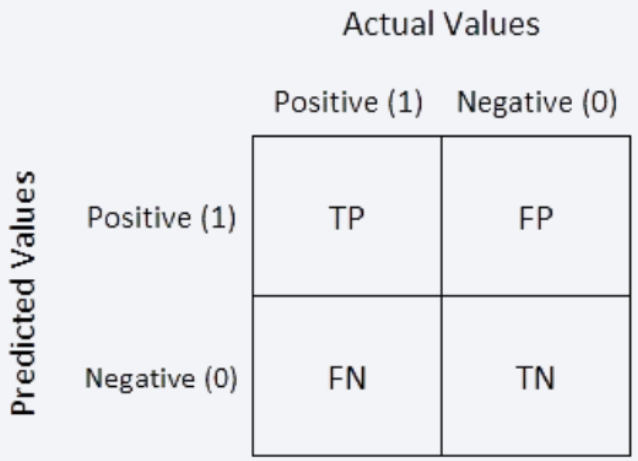


Chart: *Bar Chart showing accuracy of classification models*

CONFUSION MATRIX

- Correctly classified 14 out of 18 records
- Model Accuracy = 78%



Ideal Confusion Matrix

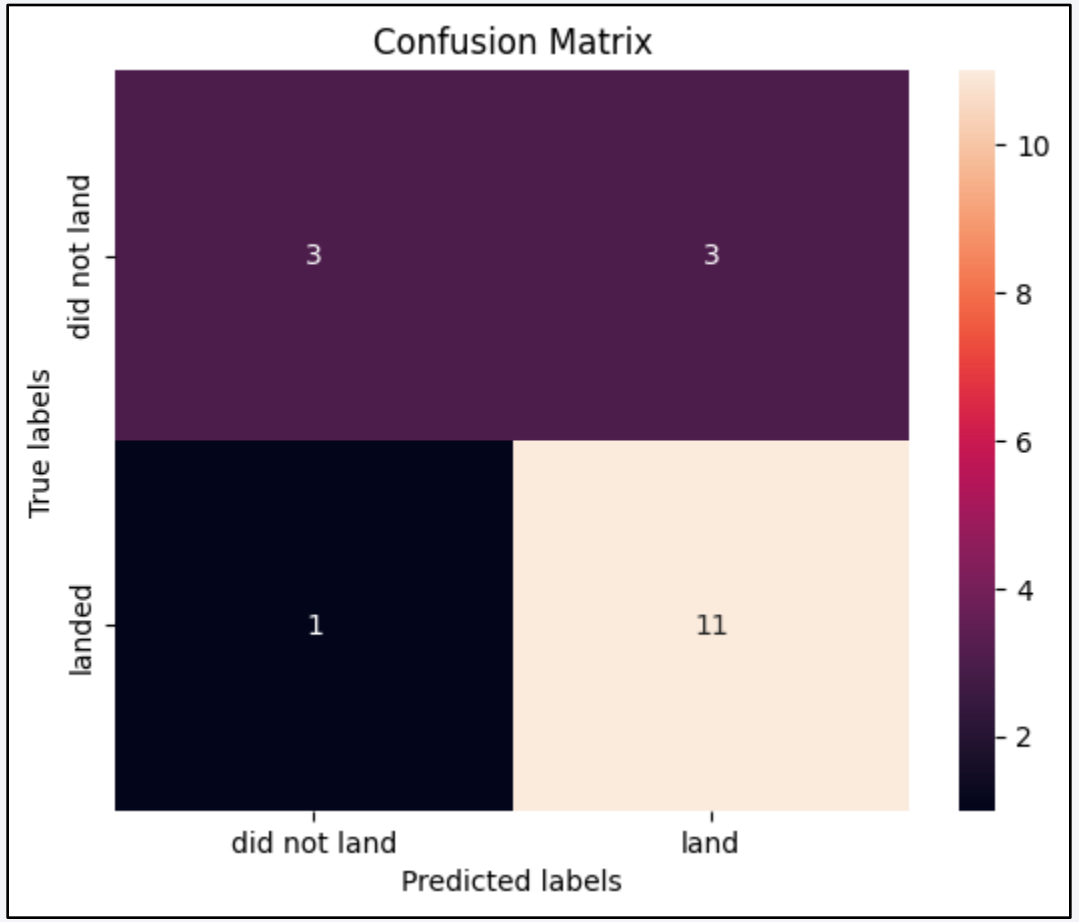
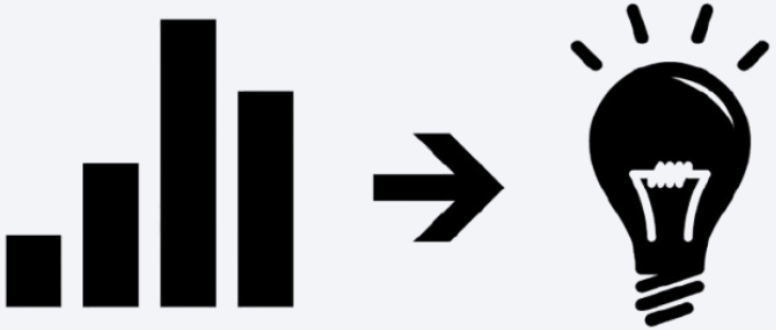


Chart: *Confusion Matrix for Decision Tree Model*

CONCLUSION

- Orbits ES – L1, GEO, HEO and SSO have 100% success rates.
- The success rate of launches increases over the years
- Total number of successful mission outcomes = 100
- KSC LC – 39A has the highest success rate of all launch sites.
- Decision Tree Model is highest performing model for classification tasks in this dataset



APPENDIX

GitHub Link:
Coursera_Applied_DS_Capstone_Project.git

THANK YOU!