Winning Space Race with Data Science

Uthayan Muthukrishnan
14-Dec-2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis
  - Interactive Dashboard
  - Predictive Analysis using Classification
- Summary of all results
  - Success rate increase for higher orbit
  - Low success rate for booster version v1.0, v1.1, higher success rate for FT, B4 and B5
  - Higher success rate for higher payload Mass
  - Higher success rate for Kennedy space center

Introduction

- Project background and context
  - Space X advertises low-cost Falcon 9 rocket launches compared to competitors
  - This success is because of reusability of first stage rockets
- Problems you want to find answers
  - If we can determine first stage will be success, then we can determine the overall cost for the launch.
  - Above result can also be used by competitors to determine the success rate of SpaceX and bid against them

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX-API
  - Web scraping of SpaceX Wikipedia page
- Perform data wrangling
  - Missing values in Payload Mass column is replaced with Mean value.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Analyze outcome by orbit type, Payload mas and booster version
  - Visual analysis with charts by payload mass, time, orbit type and launch Site
- Perform interactive visual analytics using Folium and Plotly Dash
  - Visual Analysis with map by site
  - Interactive Dashboard using plotly : Analysis by site
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, Decision Tree, KNN
  - Parameter Tuning with Grid search

Data Collection

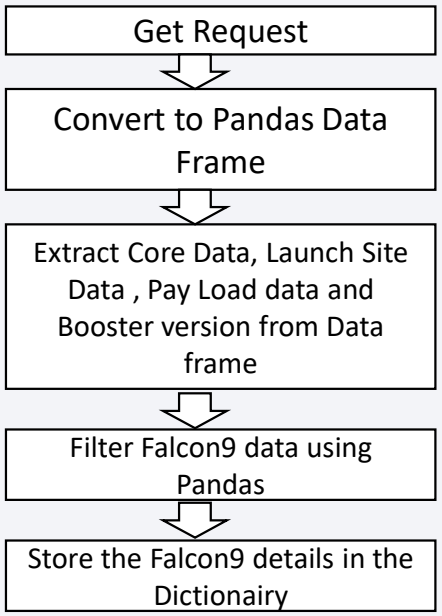
- Data were collected using Web scraping of SpaceX Wikipedia Page
  - Using Python Html requests package
  - Using Beautiful Soup package for web scrapping

- Data Collected also using SpaceX Rest API
  - Restful Interface
  - Get Coredata
  - Get Booster Vesrion
  - Get Payload Mass
  - Get Launch Site data

Data Collection Lab Notebook

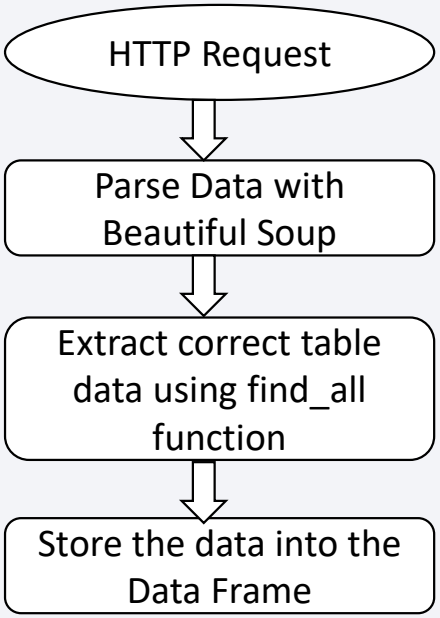
Data Collection – SpaceX API

- Send Get request to SpaceX rest API interface website
- Parse data into dataframe
- Extract Core data, Launch site data and Pay Load data and Booster version
- Filter Falcon9 data alone from the above and store it separately.
- SpaceX API calls notebook



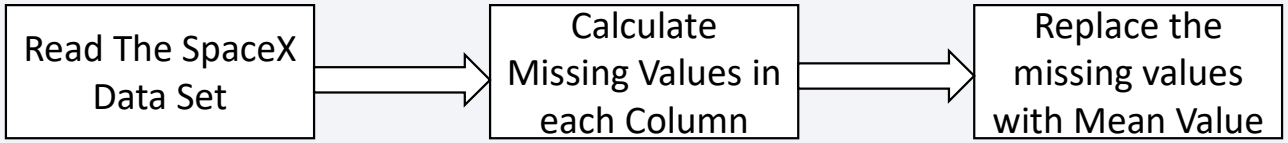
Data Collection - Scraping

- Send Http request to SpaceX Wikipedia Page
- Parse the data and find relevant table using beautiful soup
- Extract the data and store it in Pandas Data frame.
- Data Collection Notebook



Data Wrangling

- Checked the Missing values in Payload mass Column
- Replaced the Missing values with the mean value of that column
- Data Wranging



EDA with Data Visualization

- Used Charts for Data Visualization of Success rate Vs other Parameters.
  - Payload Mass vs Flight number vs Success rate using Seaborn Catplot: which showed success rate with respect to Payload mass and Flight number
  - Launch Site vs Flight Number vs Succes rate : Success rate of each launch site is debicted using seaborn catplot.
  - Launch Site Vs Payload Mass vs Success rate using seaborn catplot.
  - Orbit type vs Success Rate : Barplot is used
  - Orbit type vs Flight number vs Success Rate: Catplot is used
  - Orbit type vs Payload Mass vs Success Rate: Catplot is used
  - Success Rate vs Year: line plot.
- Data Exploration Lab

EDA with SQL

- SQL queries performed
  - Extract a list of all launch sites
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which carried the maximum payload mass
  - List the failed landing_outcomesin drone ship, their booster versions, and launch site names for in year 2015
  - 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
- SQL Query Lab

Build an Interactive Map with Folium

- Map Objects
  - Edged Circles (radius 1000m): Space launch sites
  - Markers: for labeling all objects
  - MarkerCluster: for creating a bunch of markers around space launch sites to indicate success (green) or failure (red) of the landing of the rocket's first stage
  - Lines: Measure the distince between the launch site and the next coast or next city
- Maps With Folium Lab

Build a Dashboard with Plotly Dash

- Input Elements:
  - Dropdown list for the launch site (with option to select all)
  - RangeSlider for selecting the payload mass
- Output Elements:
  - PieChart: for showing the success rate of each launch site, or (if all sites are selected) showing the number of successful landing outcomess
  - Scatterplot: Show success/failure by payload and booster version
- Interactive Plotly Dashboard

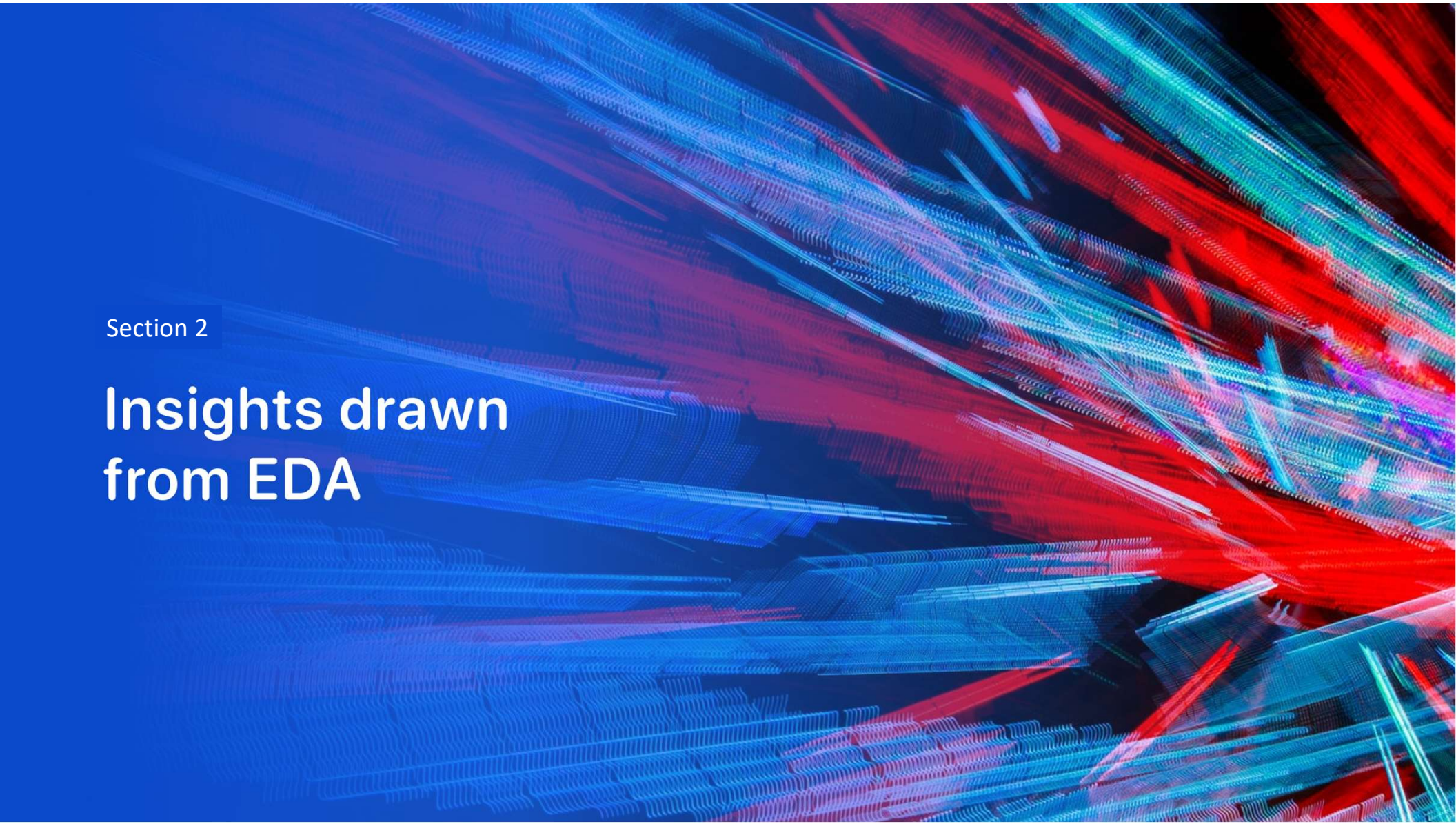
Predictive Analysis (Classification)

- Preprocessing
  - One-Hot-Encoding for Categorical Features
  - Split data into dependent/independent variables and train/test data
  - Scale Data with StandardScaler
- Model Building for each Method
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbor
- Optimization
  - Use Gridsearch for optimizing the models based on their hyperparameters
- Evaluation
  - Use Accuracy of Gridsearch for selecting the best parameter
  - Use Score to compare each classification method

Machine Learning Prediction Lab Link

Results

- Exploratory data analysis results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Higher success rate for higher payload mass
  - Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
  - Higher success rate for Kennedy Space center and recent starts at Cape Canaveral
- Predictive analysis results
  - Best prediction results with Logistic Regression and Support Vector Machine



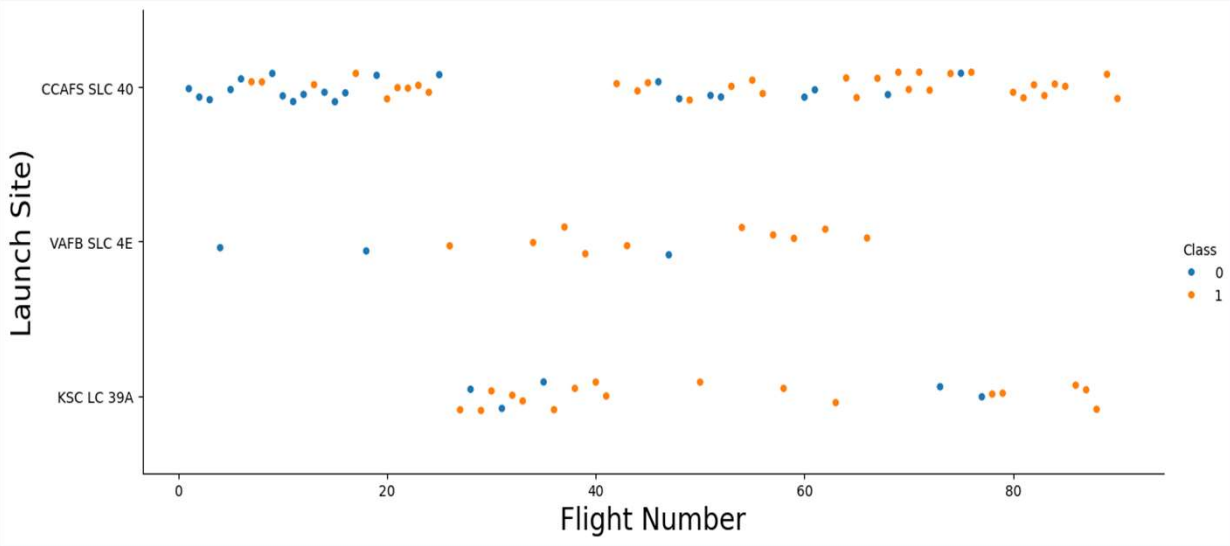
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

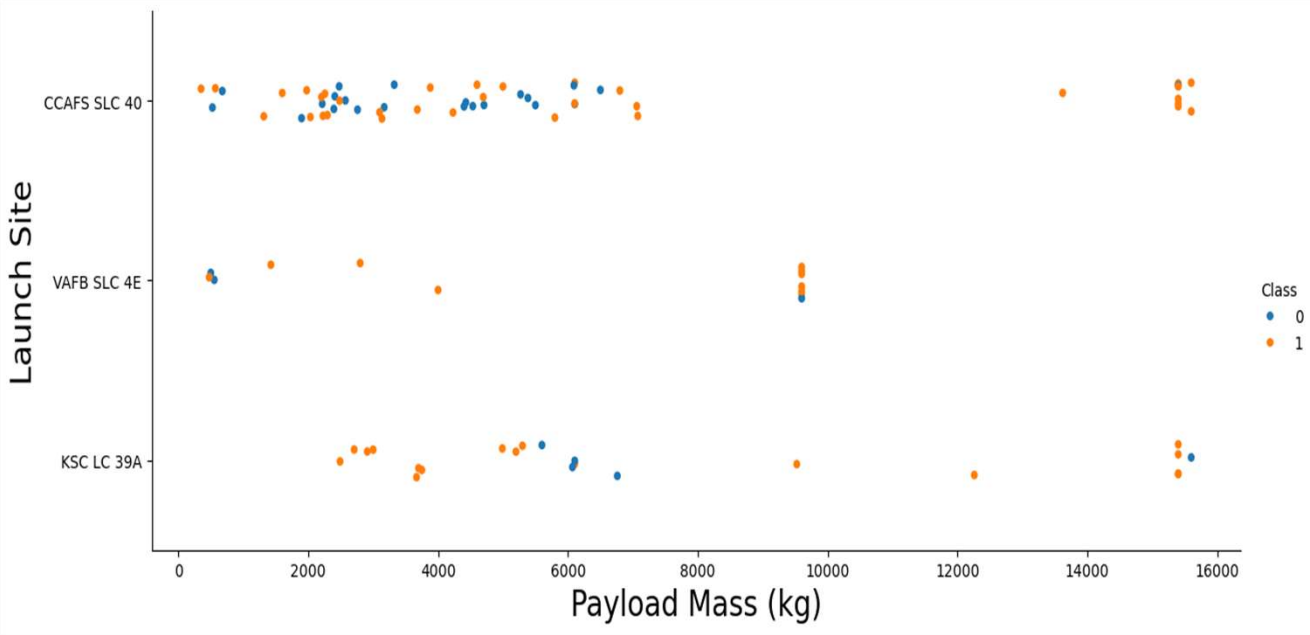
Detailed Launch Records of Each launch sites is visualized using Scatter Plots.

From the chart it is clear that number of Launch happened in CCAFS SLC 40 launch sites higher compared to other Launch sites.



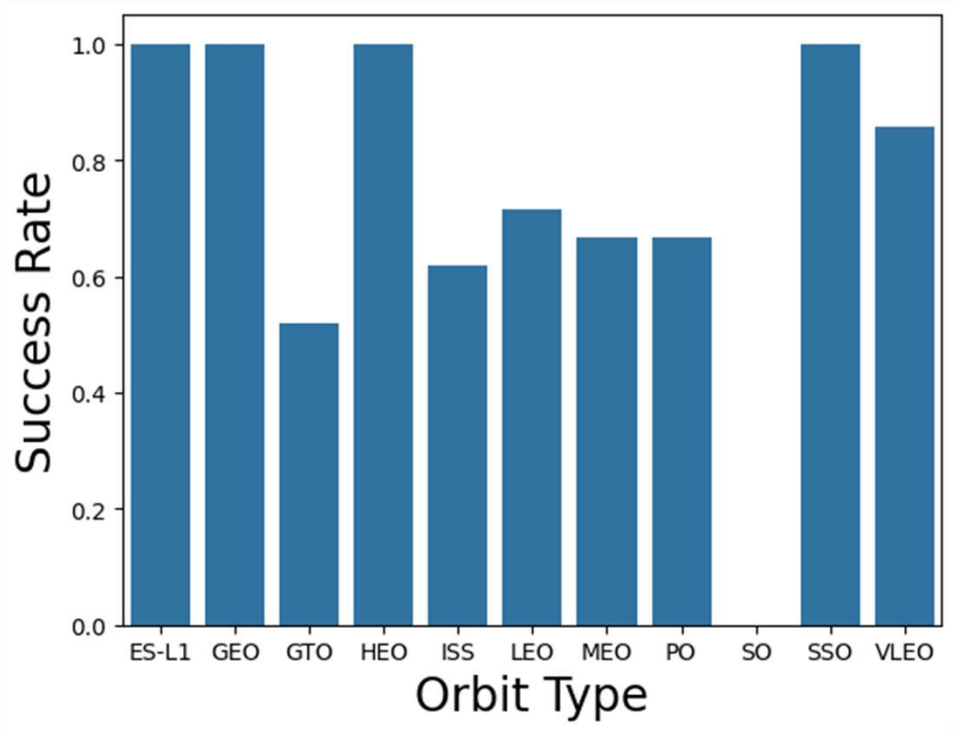
Payload vs. Launch Site

- Relation between Launch Site and Payload mass is depicted in the scatter plot chart.
- From the chart we can infer that in launch site VAFB SLC 4E there are no rockets launched for heavy payload mass greater than 10000



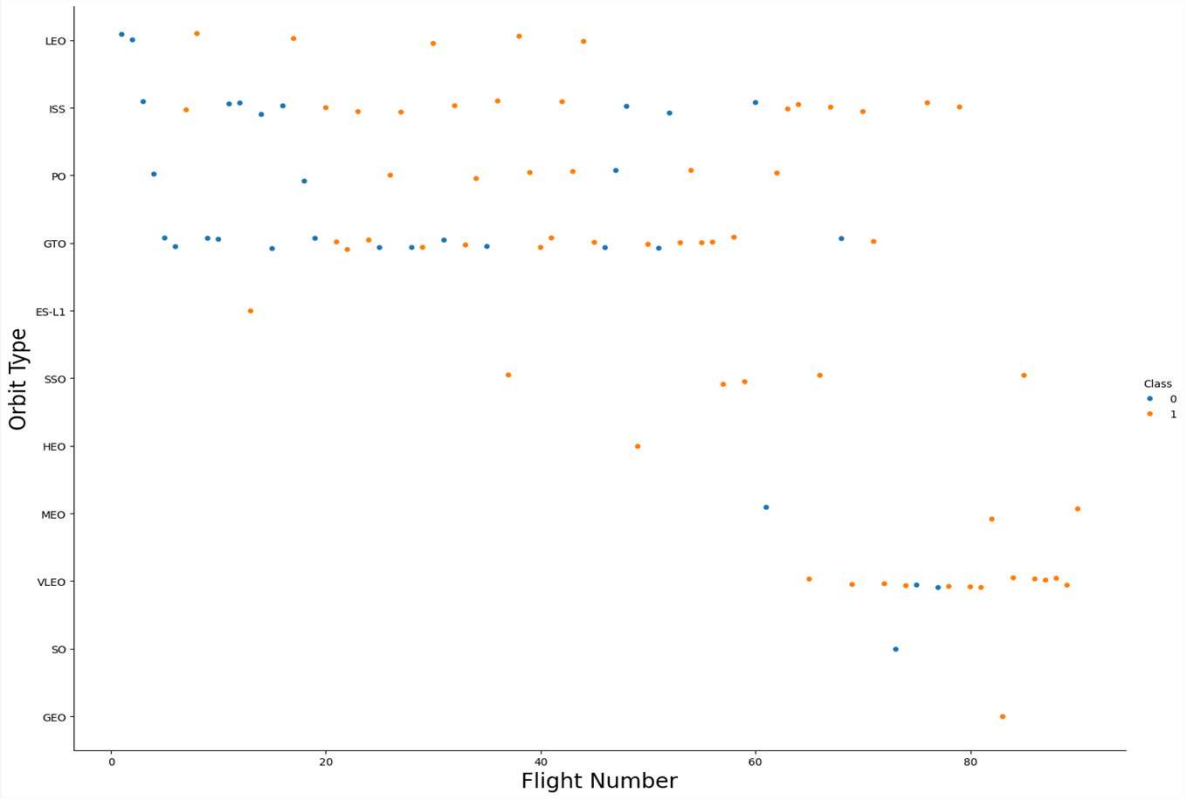
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations



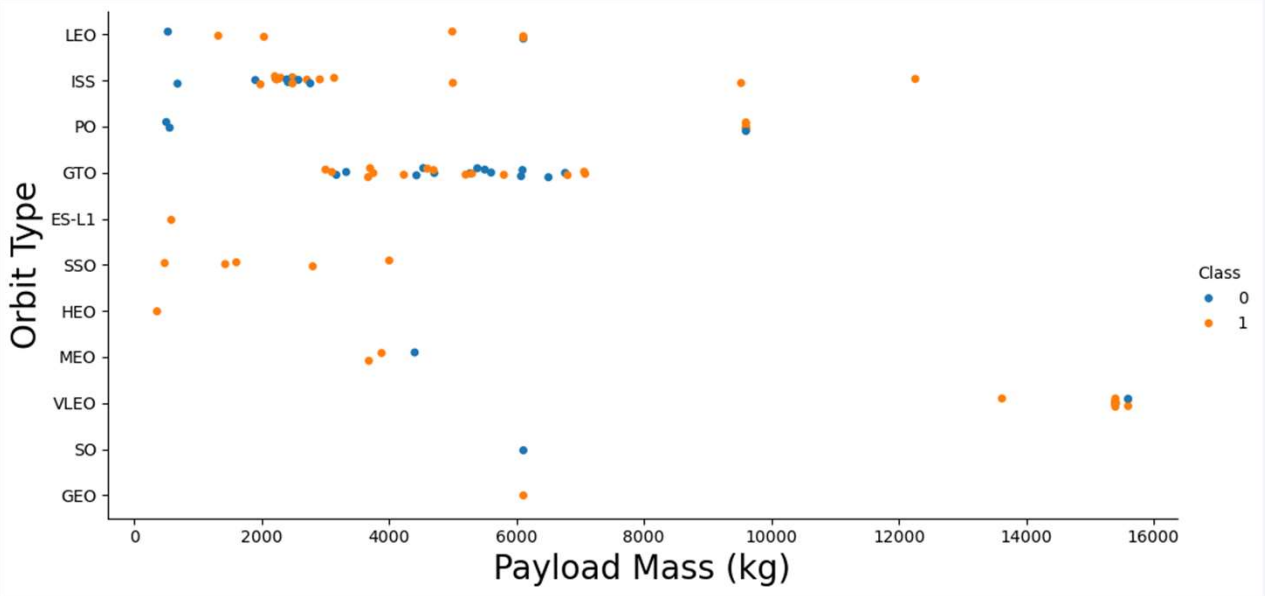
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



Payload vs. Orbit Type

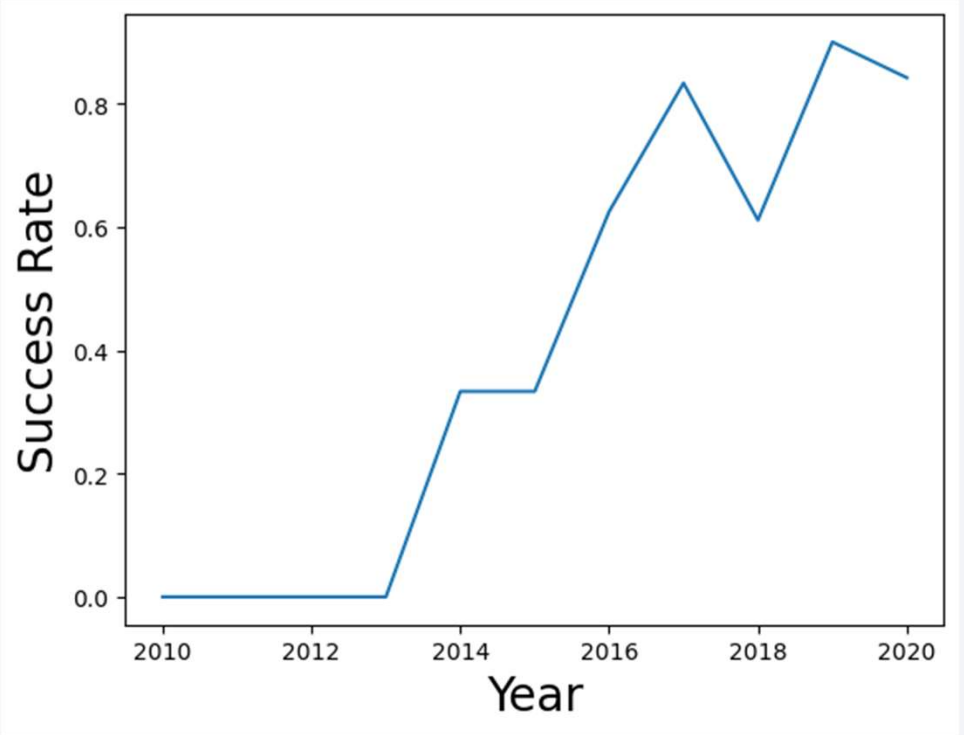
- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- KSC: Kennedy Space Center
- CCA: Cape Canaveral Launch Center
- VAFB: Vandenburg Air Force Base

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

Launch Site Names Begin with 'CCA'

- Some List of launch records in launch sites CCA

```
features.head()
```

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Clas
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003	
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005	
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007	
3	4	500.000000	PO	VAFB SLC 4E	1	False	False	False	NaN	1.0	0	B1003	
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004	

Total Payload Mass

- Total payload carried by boosters from NASA

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad

min(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

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

count(MISSION_OUTCOME)
99

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

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

Landing__Outcome	Booster_Version	Launch_Site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1017	VAFB SLC-4E
Failure (drone ship)	F9 FT B1020	CCAFS LC-40
Failure (drone ship)	F9 FT B1024	CCAFS LC-40

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

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

Section 3

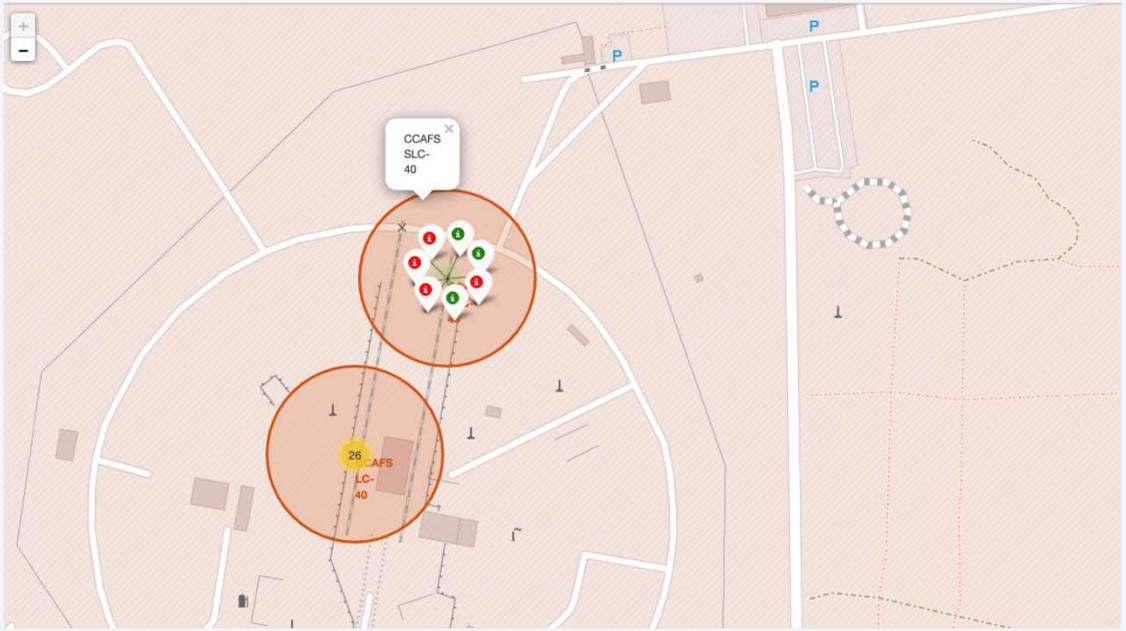
Launch Sites Proximities Analysis

Launch Site Location

- Launch Sites are at the East and West coast, near the southernmost U.S mainland area, Which is Florida and California

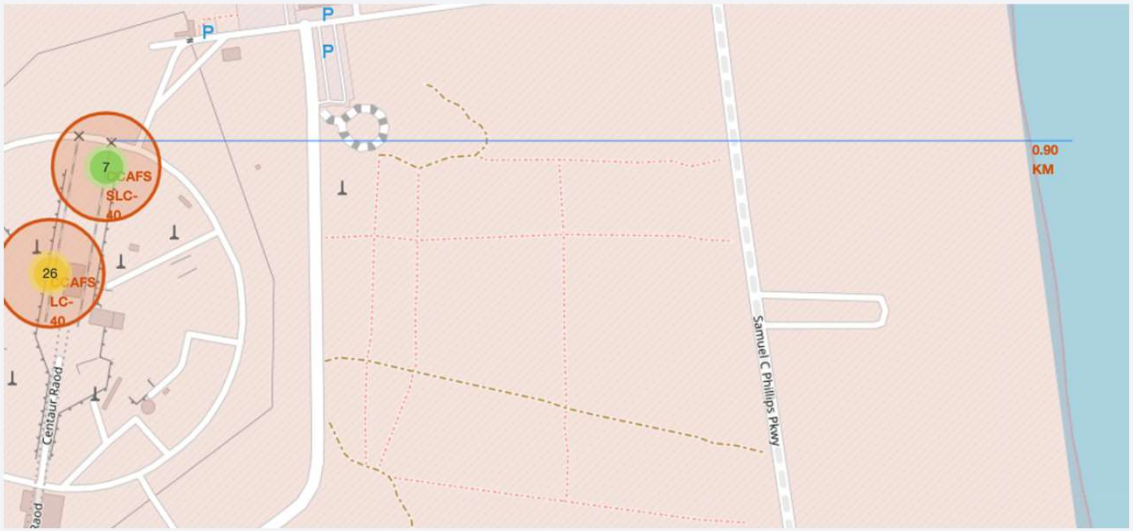
Launch Site with Success Rate

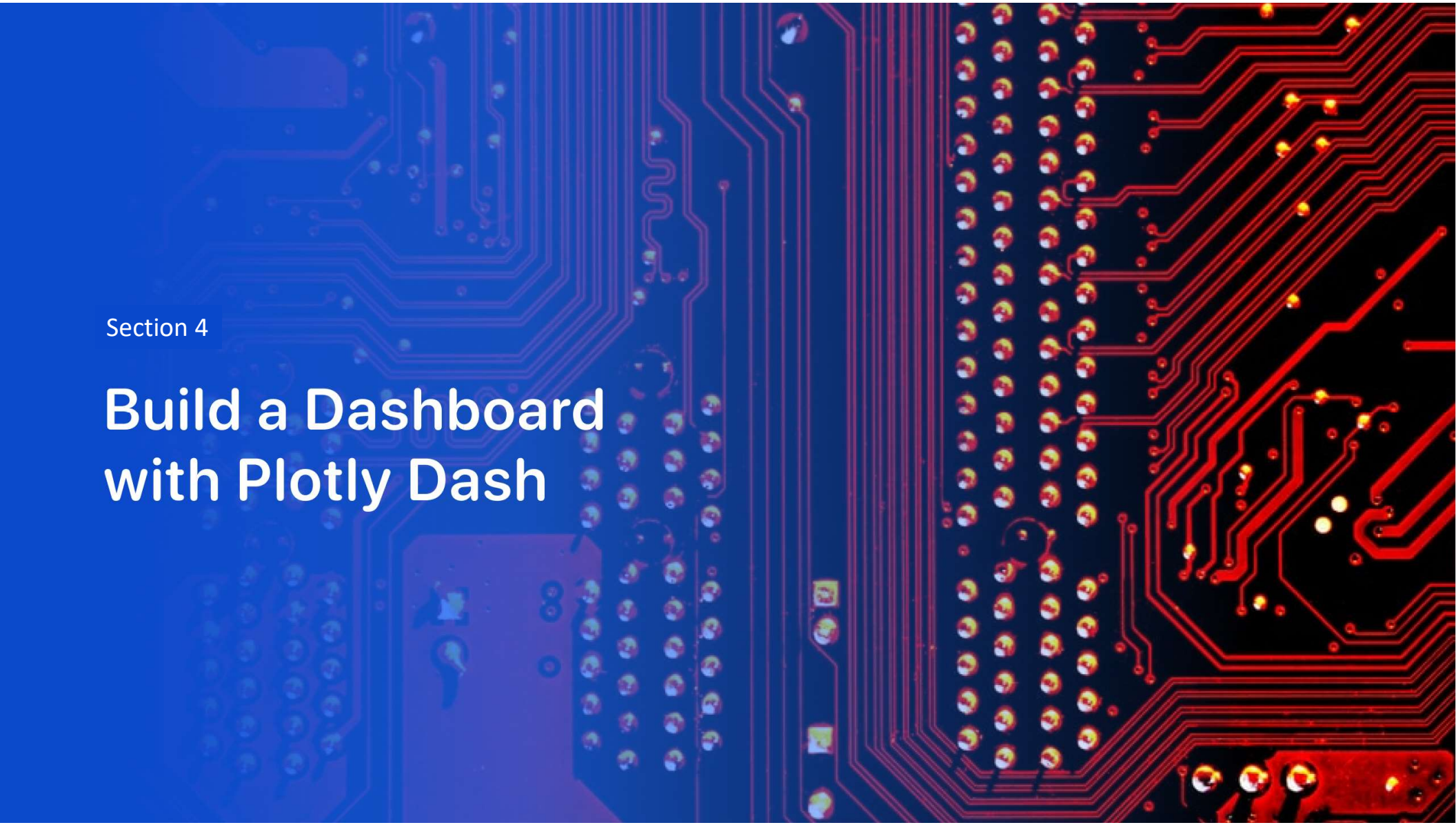
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map



Folium Map With Distance

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed



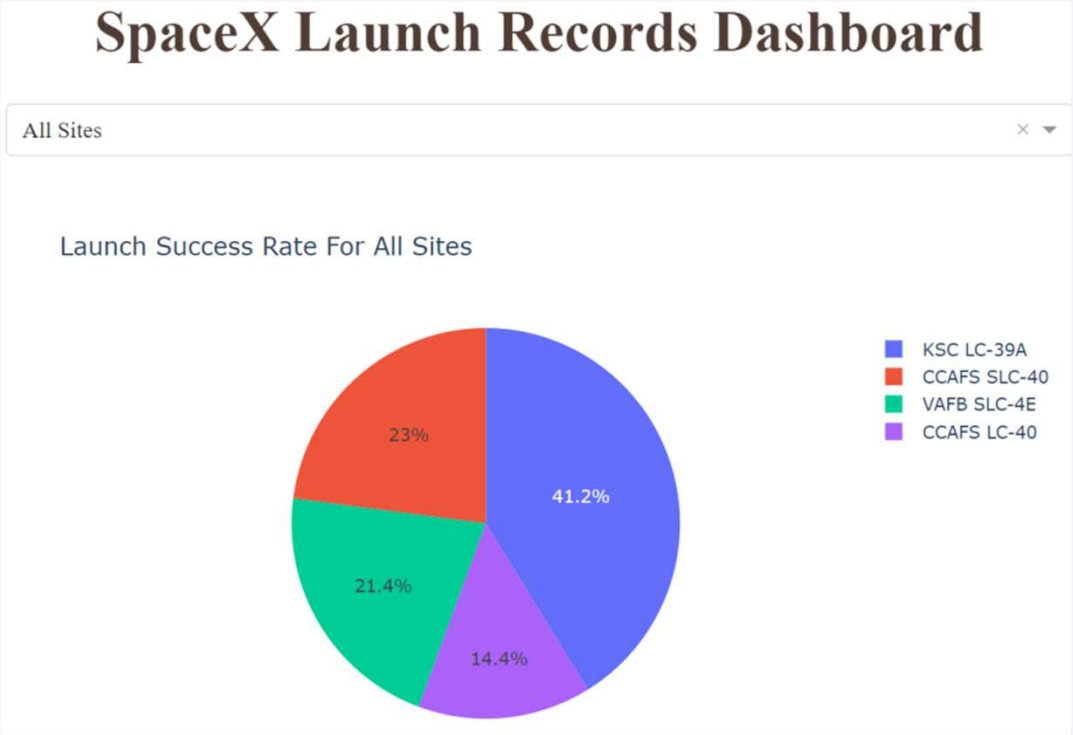


Section 4

Build a Dashboard with Plotly Dash

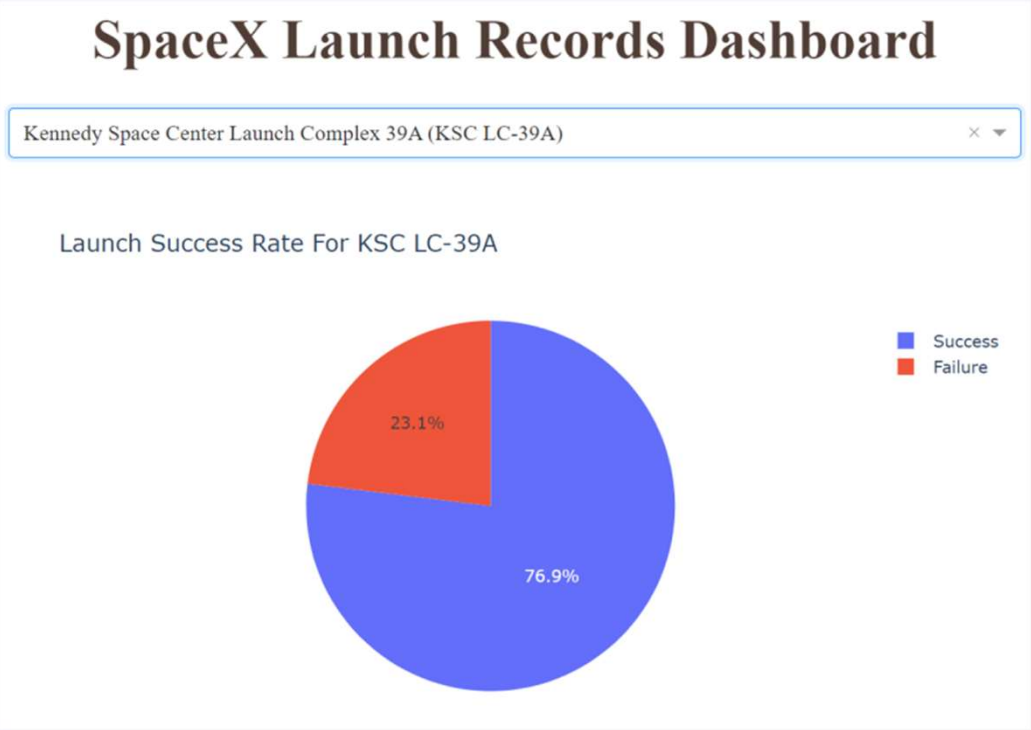
Dashboard: Launch Success Count For all Sites

- Kennedy Space Center (KSC LC-39A) hast the most successful stage-1 landings
- Vandenberg Air Force Base (VAFB SLC-4E) has the least number of successful stage-1 landings



Dashboard: Success Rate Kennedy Space center

- More than 3 of 4 landings have been successful at Kennedy Space Center



Dashboard: Booster Version V1.0, V1.1

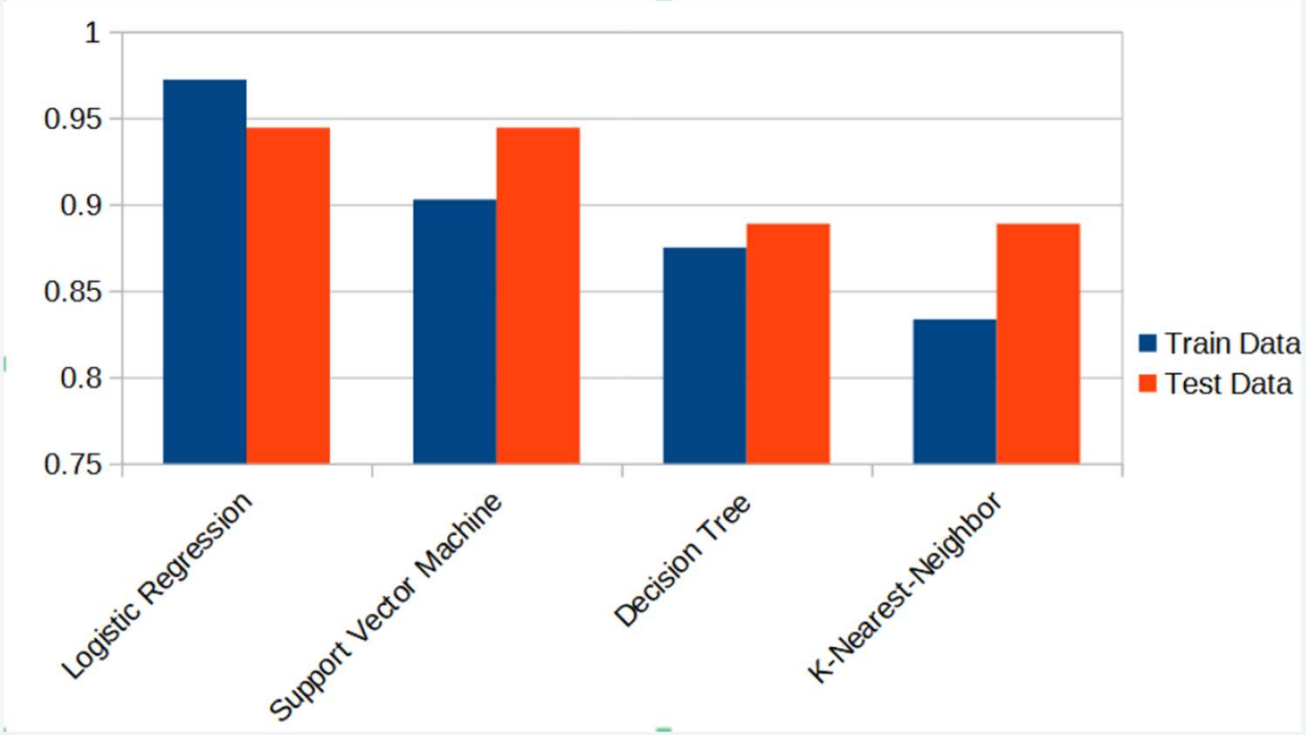
- Success rate forBooster versions v1.0 and v1.1 is quite small in the payload range to 10000kg

Section 5

Predictive Analysis (Classification)

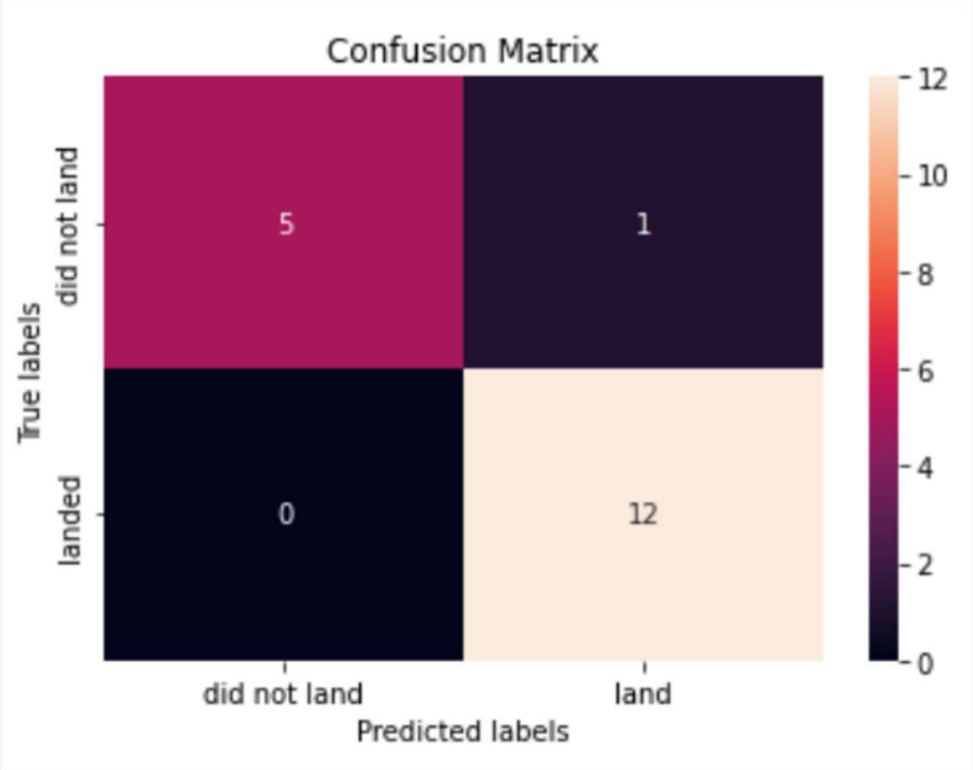
Classification Accuracy

- Logistic Regression has the best result for train data
- Logistic Regression and Support Vector Machines have the best results on test data



Confusion Matrix

- True Positives: 12
- True Negatives: 5
- False Positives: 1
- False Negatives: 0



Conclusions

- Prediction with Logistic Regression is quite accurate
- Support Vector Machine also provide a good result for predicting the landing outcome
- None of the models had false negatives
- All models had at least one false positive

Appendix

- All Code will be downloaded from the following Github Links
- uthayan-muthukrishnan/testrepo: Coursera Project (github.com)

Thank you!