Winning Space Race with Data Science

<Jan Z.>
<13.6.2022>

Outline

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

Executive Summary

- Summary of methodologies:

In order to calculate the success of the mission, data has been obtained using Webscrapping and API interface method.

The final model has been done by 4 different methods: VSM, Decision Tree, Logistics Regression, and KNN in order to predict Success rate based on input data – mainly “PayloadMass”, “Orbit”, “LaunchSite”, “LandingPad” and “Serial” as these based on correlation analyzes having the biggest impact.

- Summary of all results

The best prediction models with 89% success rate was Decision Tree model however not enough data available. With more data the model tends to be more precise.

Introduction

- Project background and context

New Space Y considers flying into space and needs to based on Space X data available find out whether there is any chance for success.

- Problems you want to find answers

Correlation among the values, as well as the model to predict success.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collecting data via API from official spacexdata website as well as with webscprapping from Wikipedia.
- Perform data wrangling
  - With numpy and pandas libraries the data has been analyzed and success (1) / no success (0) has been determinated.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - VSM, Decision Tree, Logistics Regression, and KNN models have been prepared.

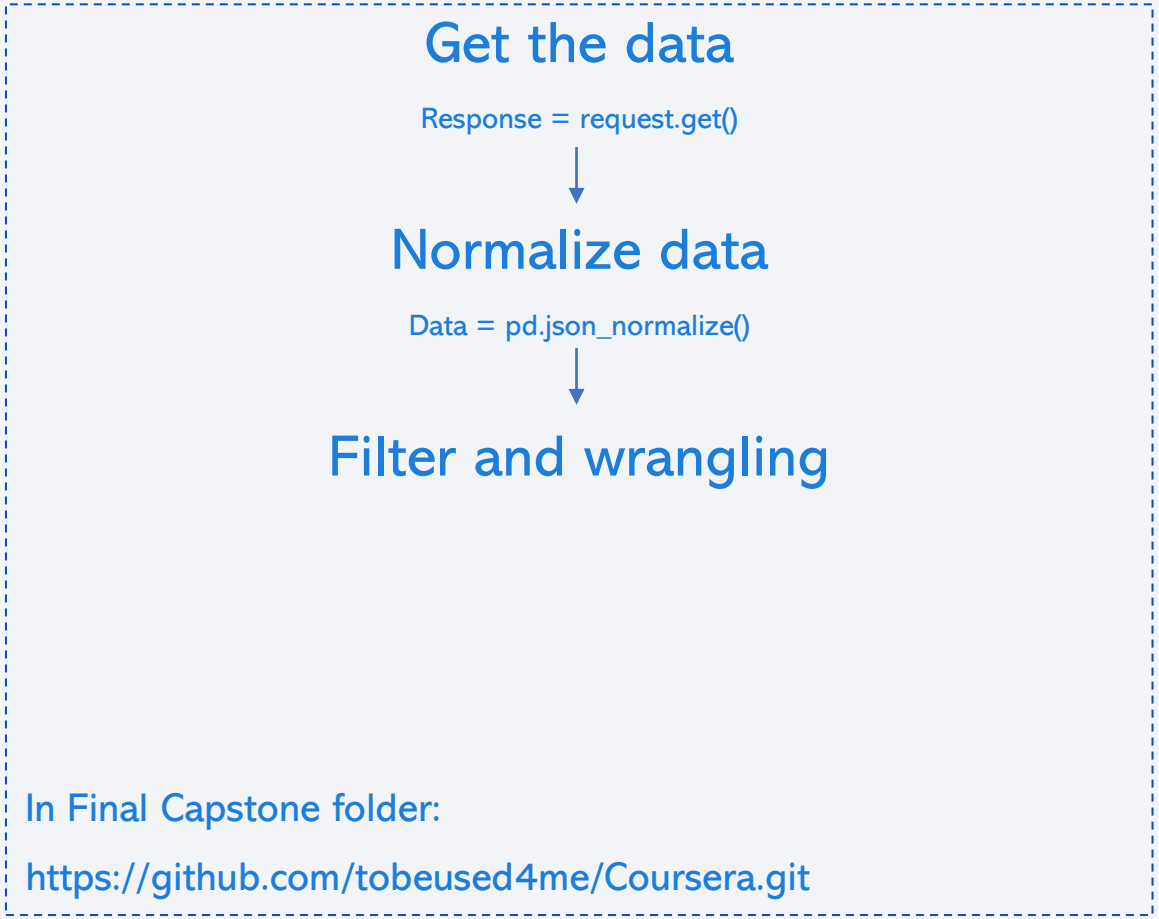
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

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

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts

Load the data → Analysis → Creating Labels

- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

In Final Capstone folder:

https://github.com/tobeused4me/Coursera.git

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

Several different charts have been plotted but mainly scatter plot and bar plot to better understand the relationship among the data.

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

In Final Capstone folder: https://github.com/tobeused4me/Coursera.git

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

- %sql select Unique(LAUNCH_SITE) from SPACEXDATA;
- %sql SELECT LAUNCH_SITE from SPACEXDATA where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
- %sql select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXDATA;
- %sql select avg(PAYLOAD_MASS__KG_) as payloadmass from SPACEXDATA;
- %sql select min(DATE) from SPACEXDATA;
- %sql select BOOSTER_VERSION from SPACEXDATA where LANDING__OUTCOME='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
- %sql select count(MISSION_OUTCOME) as missionoutcomes from SPACEXDATA GROUP BY MISSION_OUTCOME;
- %sql select BOOSTER_VERSION as boosterversion from SPACEXDATA where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXDATA)
- %sql SELECT MONTH(DATE),MISSION_OUTCOME,BOOSTER_VERSION,LAUNCH_SITE FROM SPACEXDATA where EXTRACT(YEAR FROM DATE)='2015'
- %sql SELECT LANDING__OUTCOME FROM SPACEXDATA WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

In Final Capstone folder: https://github.com/tobeused4me/Coursera.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

- Explain why you added those objects

Several visualizations to see spatial position among several places of start for further analysis

- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

In Final Capstone folder: https://github.com/tobeused4me/Coursera.git

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard

I did not succeed to create the dashboard.

- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

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

4 different models have been done: KNN, VSM, Logistics regression and Decision tree with use of GridSearchCV – several different parameters to be tested during one session.

- You need present your model development process using key phrases and flowchart

Obtaining data → scale the data → split into Train/Test → training → prediction → matrices

- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

In Final Capstone folder: https://github.com/tobeused4me/Coursera.git

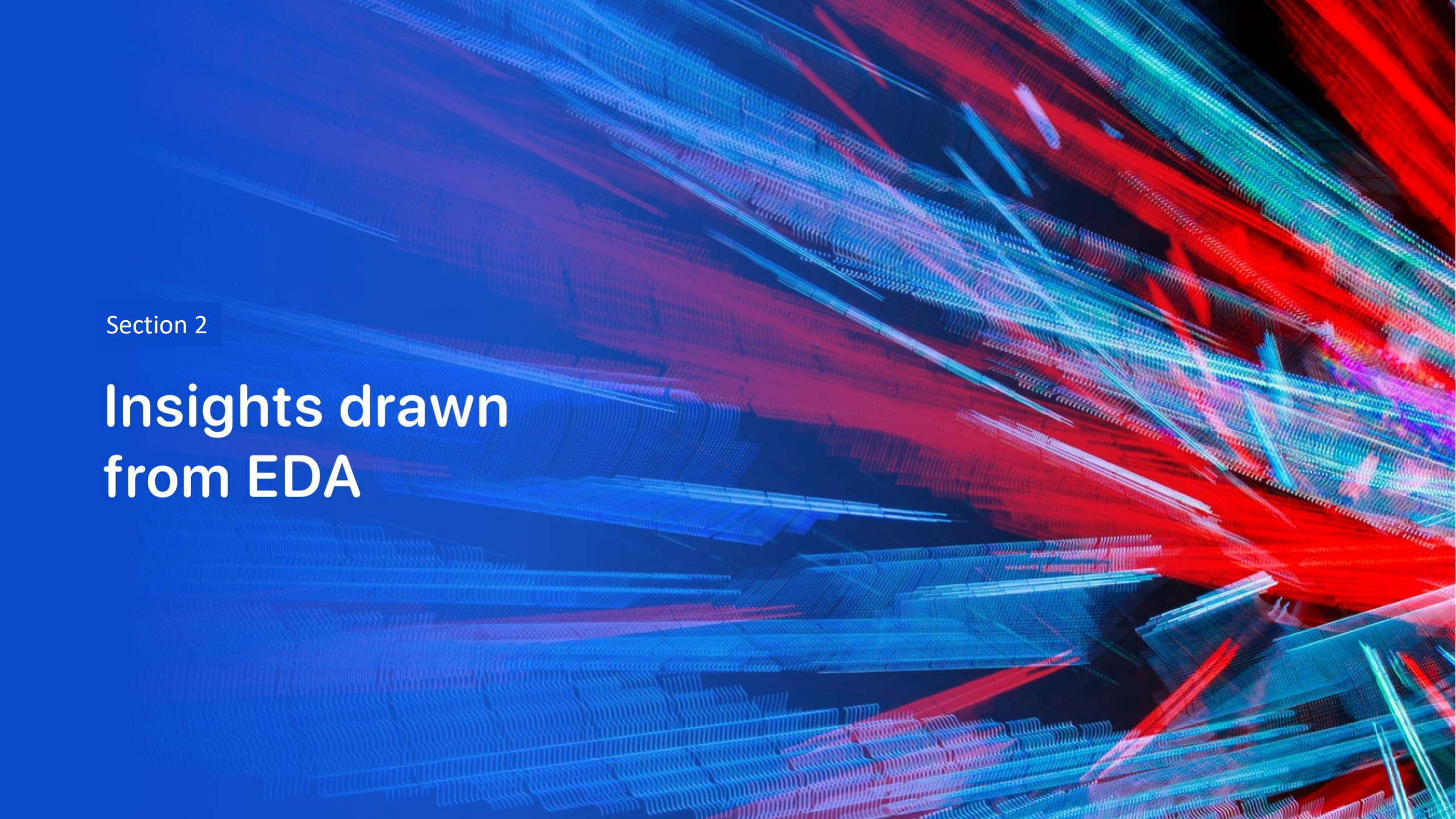
Results

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

Logistic regression and SVM = 88,9%

Tree decision = 83,3%

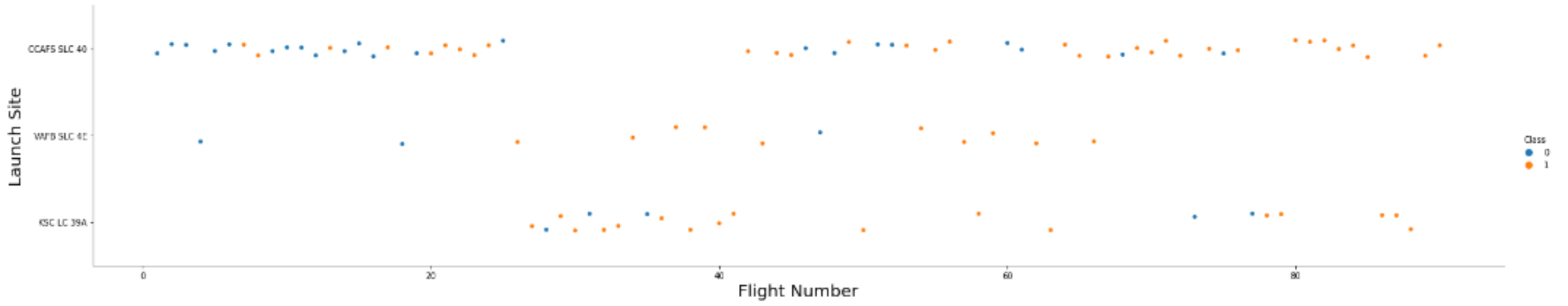
KNN = 0,77%


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations

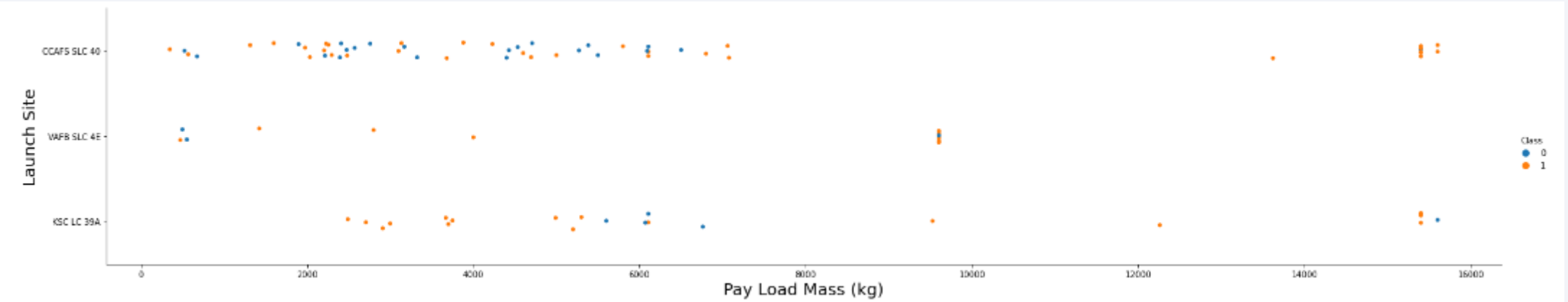
- No correlation between Flight Number vs. Launch Site and success of mission



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

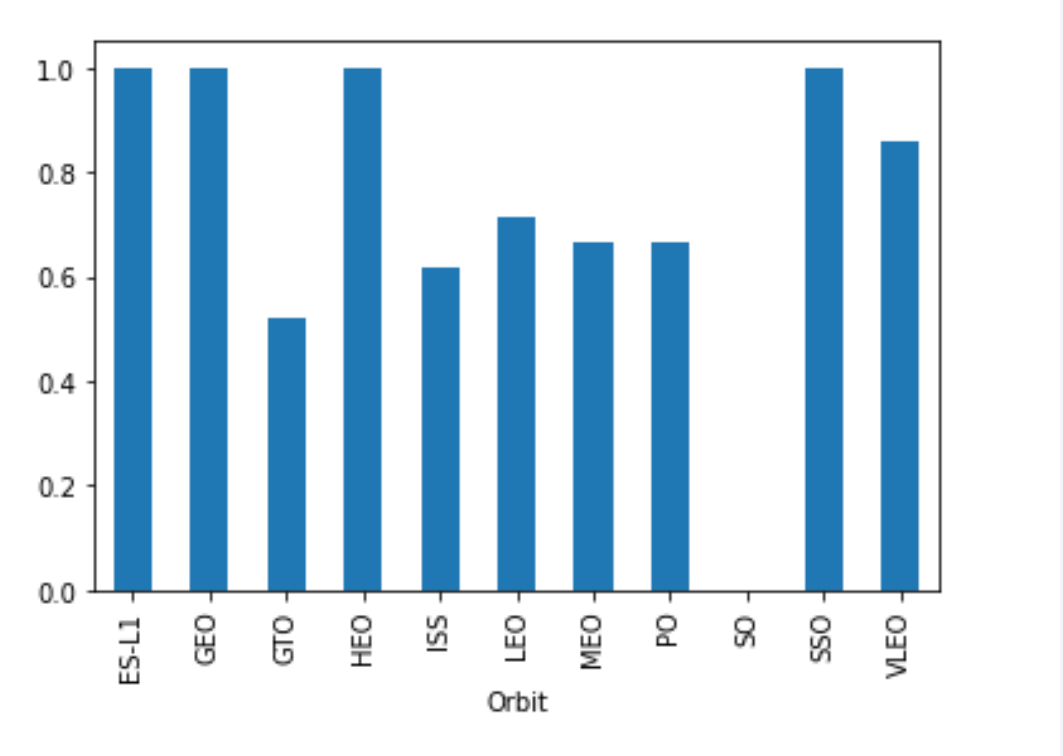
- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

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

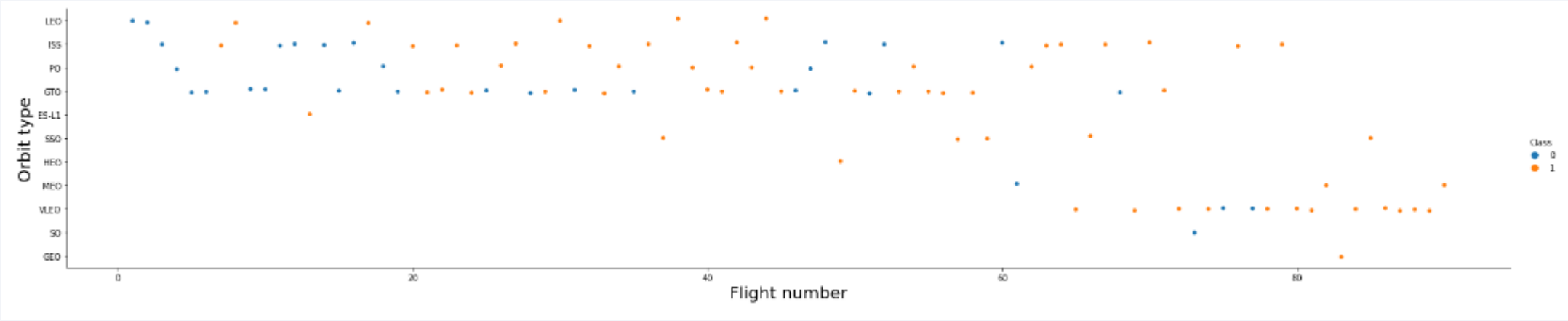
No success at SO at all.



Flight Number vs. Orbit Type

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

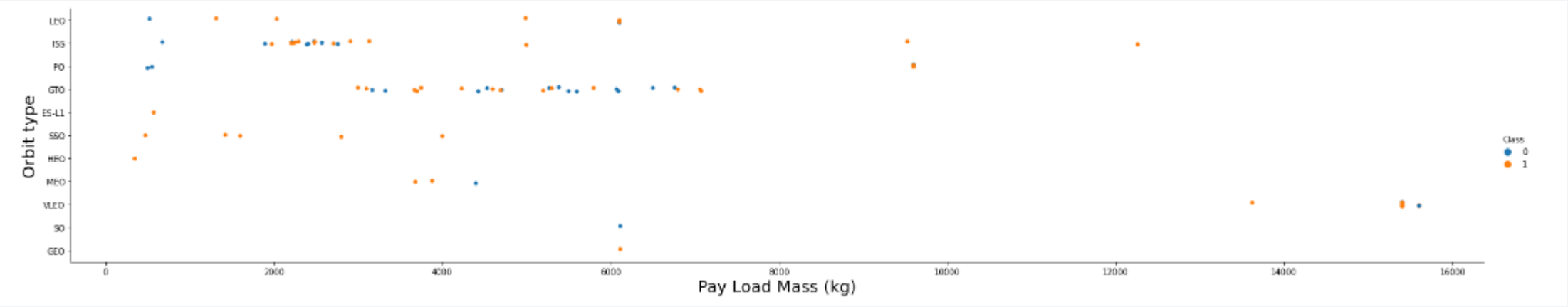
- Show the screenshot of the scatter plot with explanations



Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

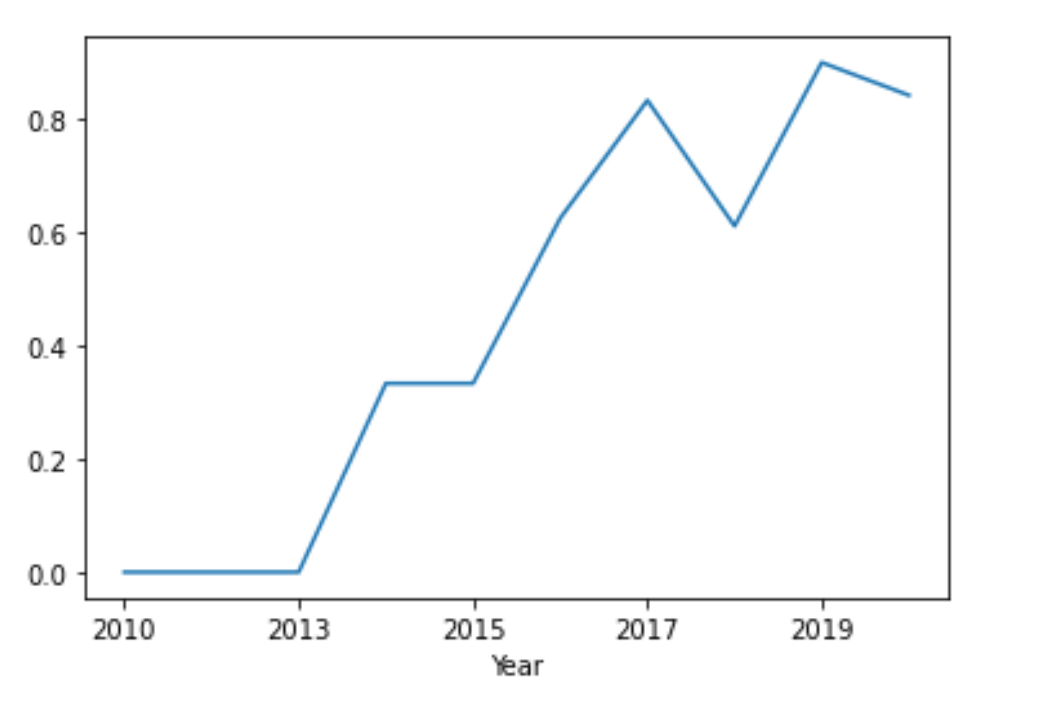


Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

Positive trend since 2013.



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

payloadmass
619967

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

payloadmass
6138

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

1
2010-06-04

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

- 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
- Present your query result with a short explanation here

missionoutcomes
1
99
1

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

boosterversion
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

1	mission_outcome	booster_version	launch_site
1	Success	F9 v1.1 B1012	CCAFS LC-40
2	Success	F9 v1.1 B1013	CCAFS LC-40
3	Success	F9 v1.1 B1014	CCAFS LC-40
4	Success	F9 v1.1 B1015	CCAFS LC-40
4	Success	F9 v1.1 B1016	CCAFS LC-40
6	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
12	Success	F9 FT B1019	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

- Present your query result with a short explanation here

landing_outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)
No attempt
Controlled (ocean)
Failure (drone ship)
Uncontrolled (ocean)
No attempt
No attempt
Controlled (ocean)
Controlled (ocean)
No attempt
No attempt
Uncontrolled (ocean)
No attempt
No attempt
No attempt
Failure (parachute)
Failure (parachute)

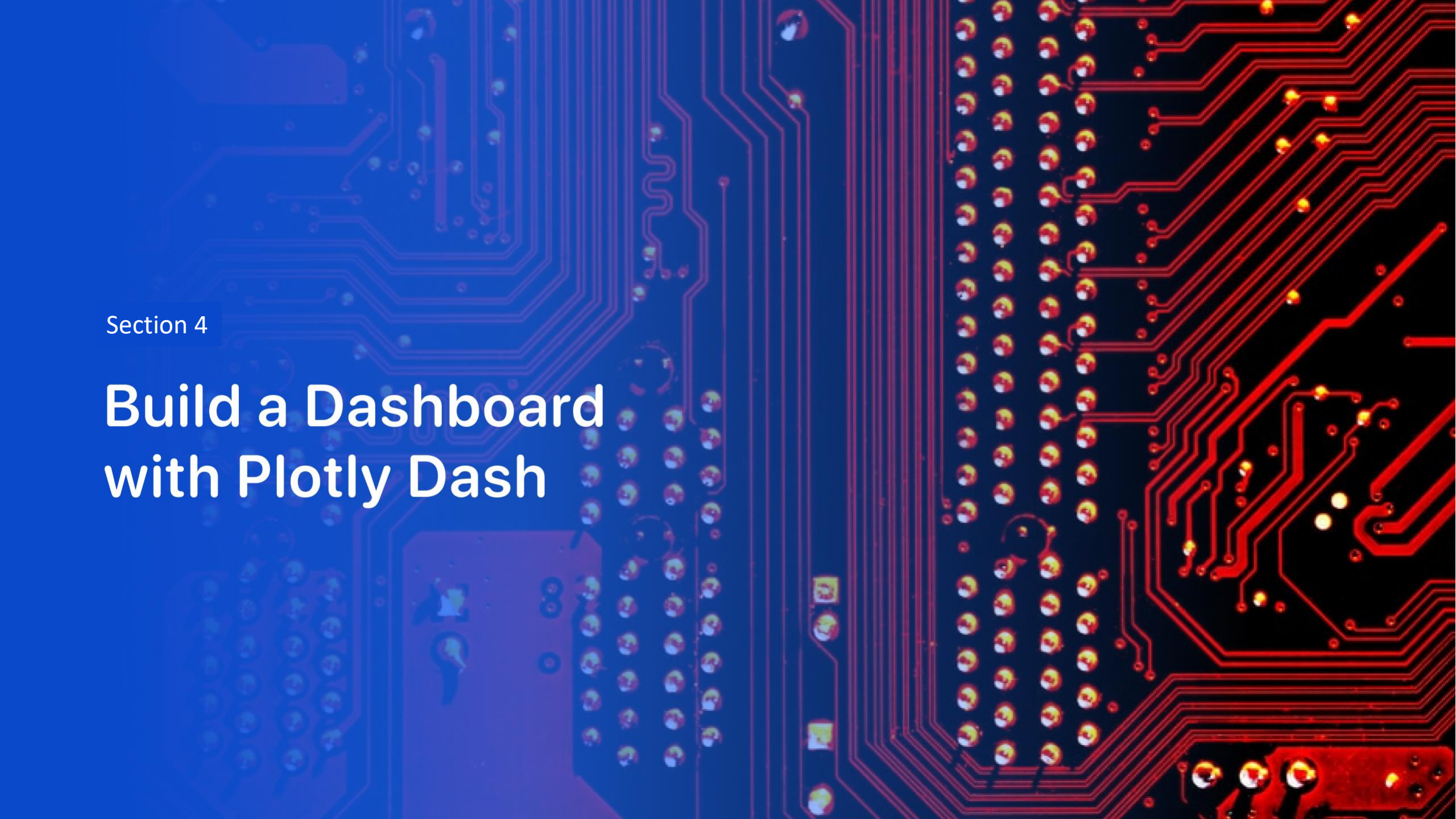
Section 3

Launch Sites Proximities Analysis

Launch Sites on Map

Color labeled launch outcomes

Proximity to key locations


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

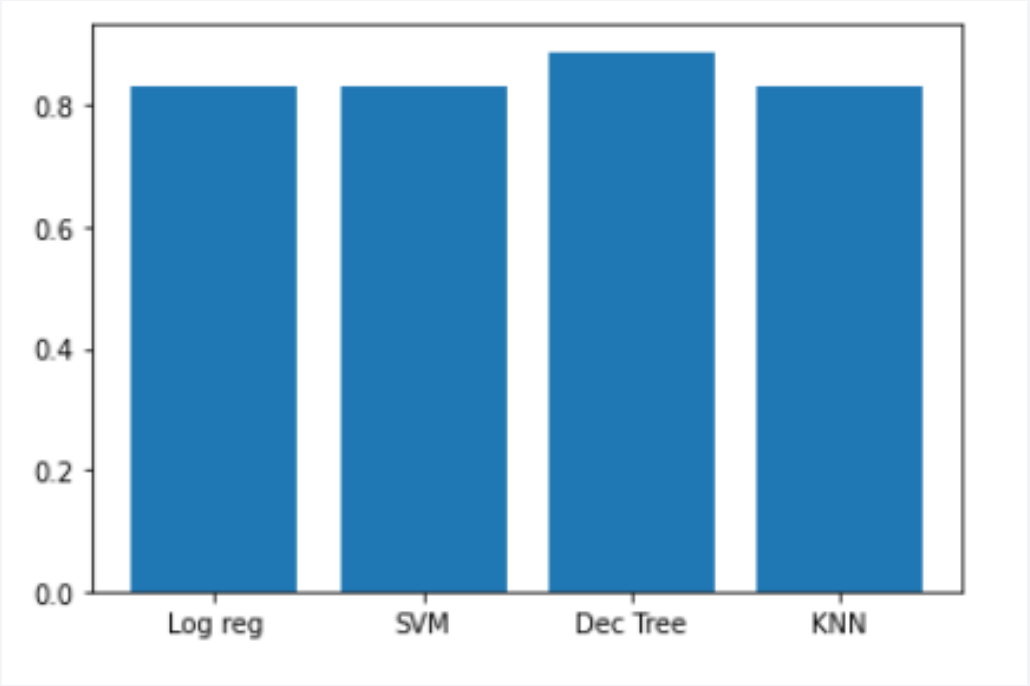
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

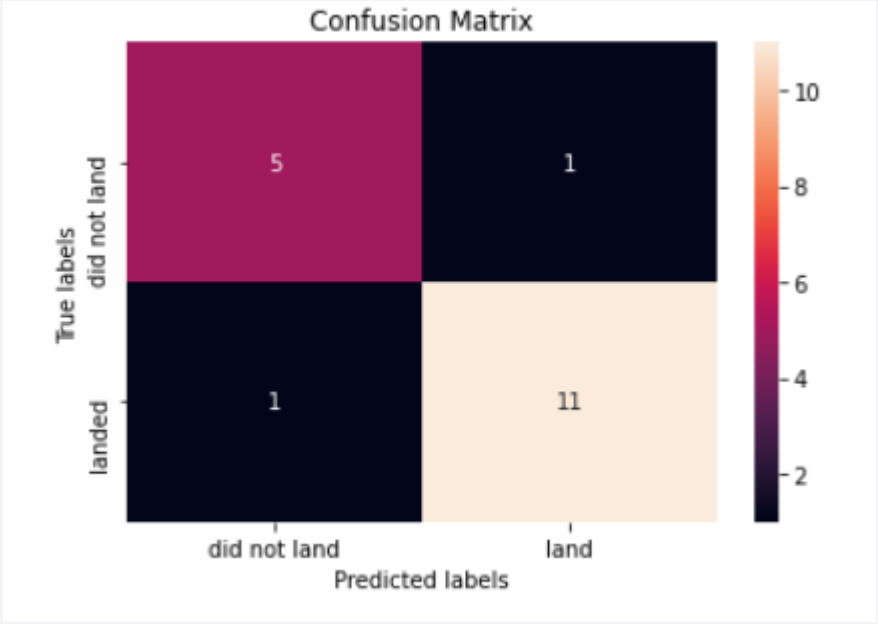
The highest accuracy has Decision Tree model.



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

100% accuracy with land vs. landed



Conclusions

The Decision Tree model has the best results in terms success prediction.

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!