Data science Capstone

Zlatko Dzigal

19.06.2023.

https://github.com/Zlatko-Dz/Capstone

OUTLINE

- Executive Summary – slide 3
- Introduction - slide 4
- Methodology – slide 5
- Results – slide 6 : slide 16
  - Data preparation and analysis
  - Visualization – Charts
  - Dashboard
  - Predictive analysis
- Discussion – slide 17
  - Findings & Implications
- Conclusion – slide 19
- Appendix – slide 20

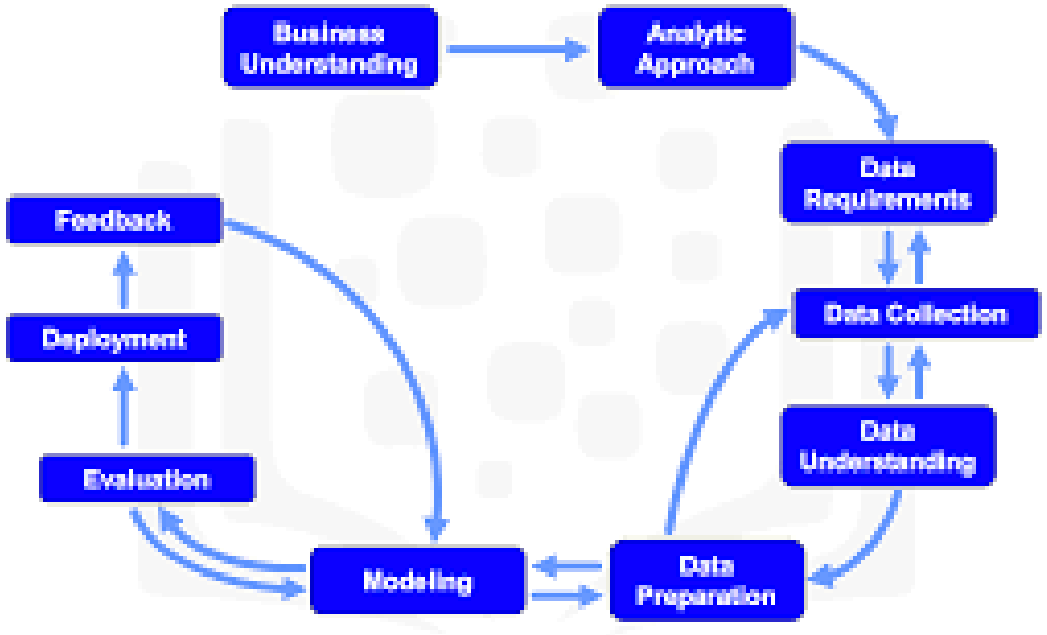
EXECUTIVE SUMMARY

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. Finding the relation between different launch variables (launch sites, orbits, etc.) and launch outcome, we can predict which combination of launch variables is the best in order to achieve the mission's goal - the successful SpaceX first stage landing.
- We are using standard data science methodology - after collecting and understanding the data we prepare it, analyze it, make different visualizations and use machine learning to develop and evaluate the model for prediction.
- After data collecting, wrangling and web scraping, data is prepared, exploratory analysis has been made using SQL and using different data visualizations and we calculated relations between launch variables and succesful landing outcomes.
- Afterwards, using Folium, launch site analysis has been made and Dashboard for SpaceX different Falcon 9 launch sites has been made also, which helped us to identify launch sites with highest success rates, in the terms of successful SpaceX Falcon 9 first stage landing.
- At the end, findings and immplications are presented as conclusions giving information about the probability of mission success.

INTRODUCTION

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Finding the relation between different launch variables (launch sites, orbits, etc.) and launch outcome, we can predict which combination of launch variables is the best in order to achieve the mission's goal - the successful SpaceX first stage landing, helping the Investors approve the project and make positive return on their investment.

METHODOLOGY



- After collecting and understanding the data we prepared it, analyzed it, made different visualizations and used machine learning to develop and evaluate the model.
- As a result we come to the conclusion which launch variables will provide us the highest probability of a mission success – successful landing of SpaceX Falcon 9 first stage successful landing.

RESULTS / Data collecting & wrangling

- First we collected the data from link provided below:
  - ✓ `spacex_url="https://api.spacexdata.com/v4/launches/past"`
- Then we made some some basic data wrangling and formating in order to clean the data and prepare it for further analysis.
- After normalising the data, we created dataframe and filtered it so we can use only Falcon 9 launches.
- At the end we dealt with missing values and stored it in the 'csv' file.
- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

RESULTS / Data exploratory analysis

- We started with EDA – exploratory data analysis and:
  - ✓ Calculated the number of launches on each site
  - ✓ Calculated the number and occurrence of each orbit
  - ✓ Calculate the number and occurence of mission outcome per orbit type
  - ✓ Determined the launch success rate

- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

RESULTS / Data web scraping

- We used Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
  - ✓ `static_url = https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922`
- We extracted variables / column names and we created data frame by parsing the launch HTML tables and filtered it so we can use only Falcon 9 launches.
- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/jupyter-labs-webscraping.ipynb

RESULTS / EDA with SQL

- We used SQL for data exploratory analysis to find relations between:

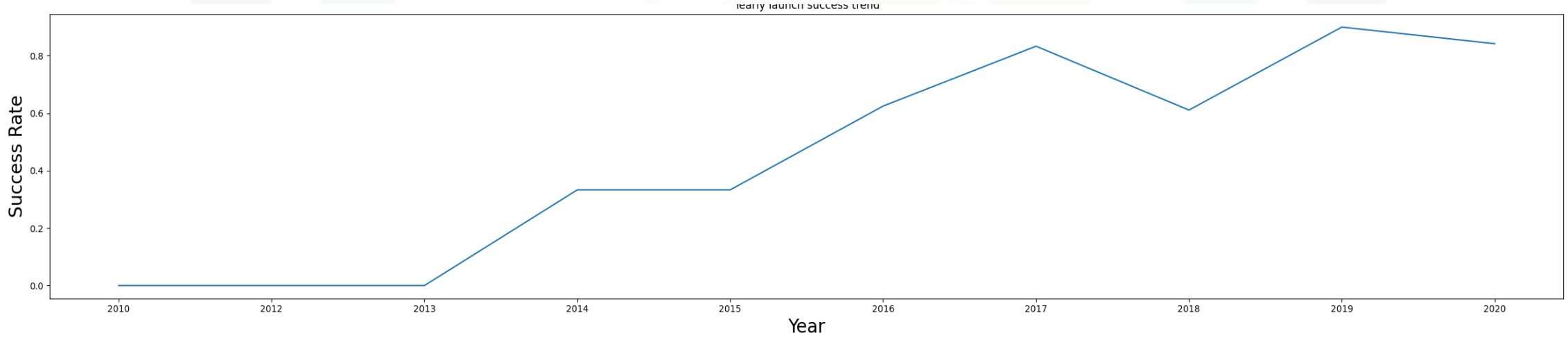
  - ✓ Launch Site
  - ✓ Payload Mass
  - ✓ Booster Versions
  - ✓ Landing Outcome

- As the result we got information about number of successful landing outcomes.

- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

RESULTS / EDA with Data Visualization

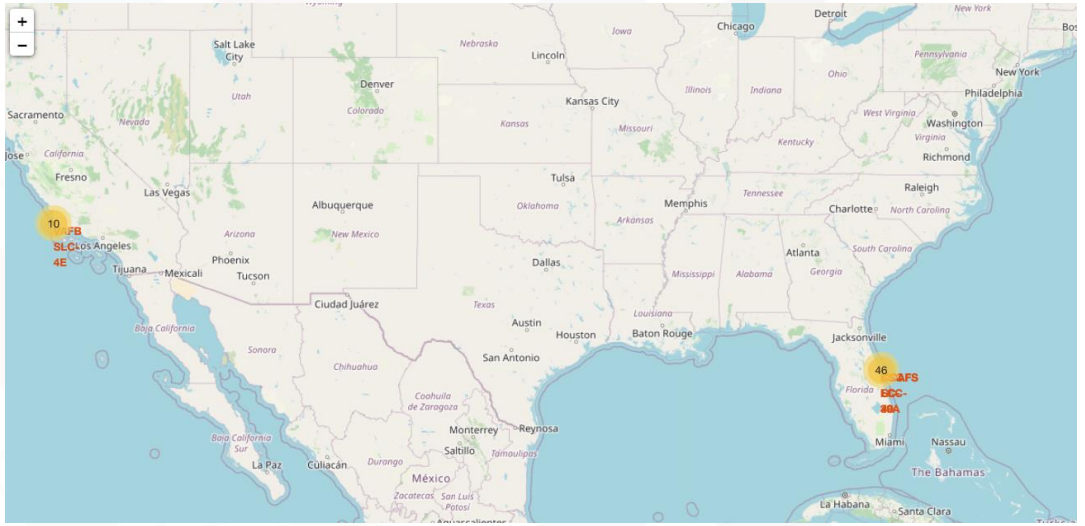
- We used Pandas and Matplotlib for exploratory data analysis (EDA).
- We found which combinations of launch variables have high success rates.
- Determined the trend of successsful launch outcomes from 2010 until 2020.
- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb



FOLIUM / Launch Sites Locations Analysis

In this part we analyzed the launch success rate dependancy on the location and proximities of a launch site, i.e., the initial position of rocket trajectories.

Choosing the optimal location for launching the SpaceX Falcon may increase the probability of a mission success.



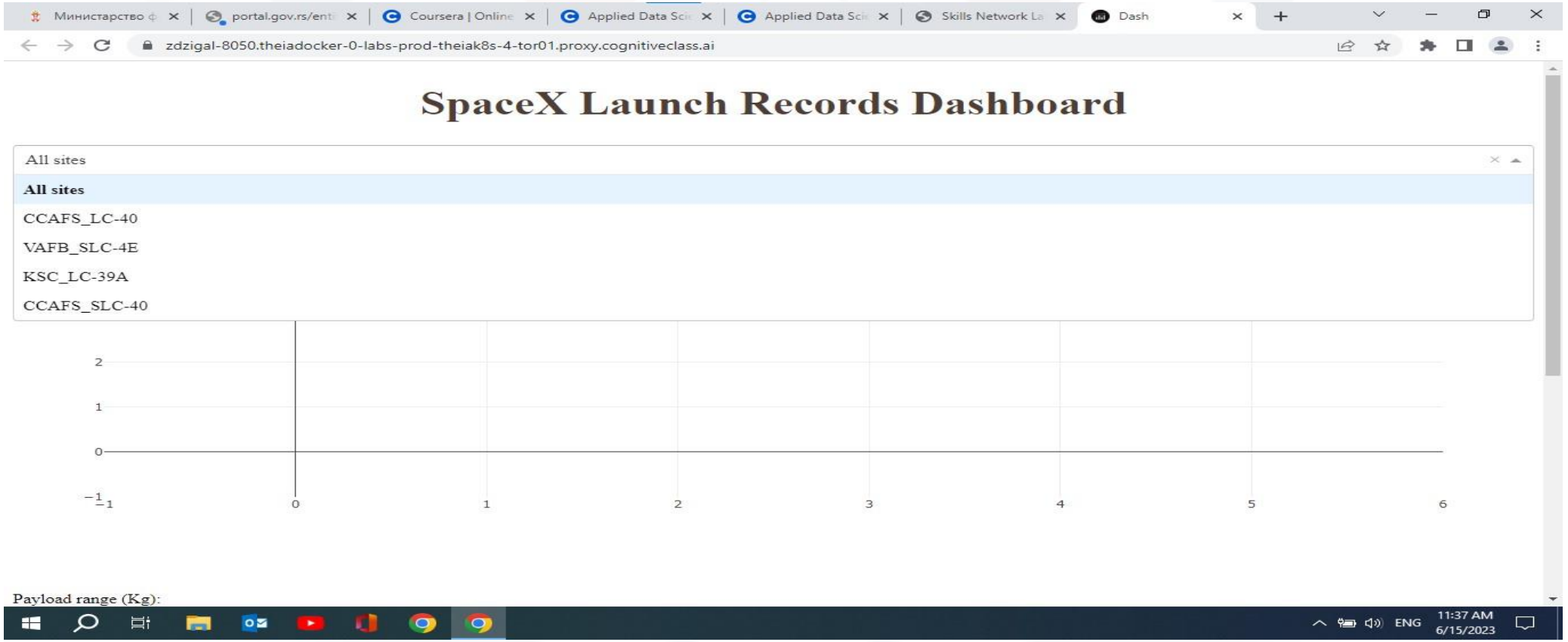
https://github.com/Zlatko-Dz/Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

DASHBOARD / SpaceX Dash for different SpaceX Falcon 9 launch sites

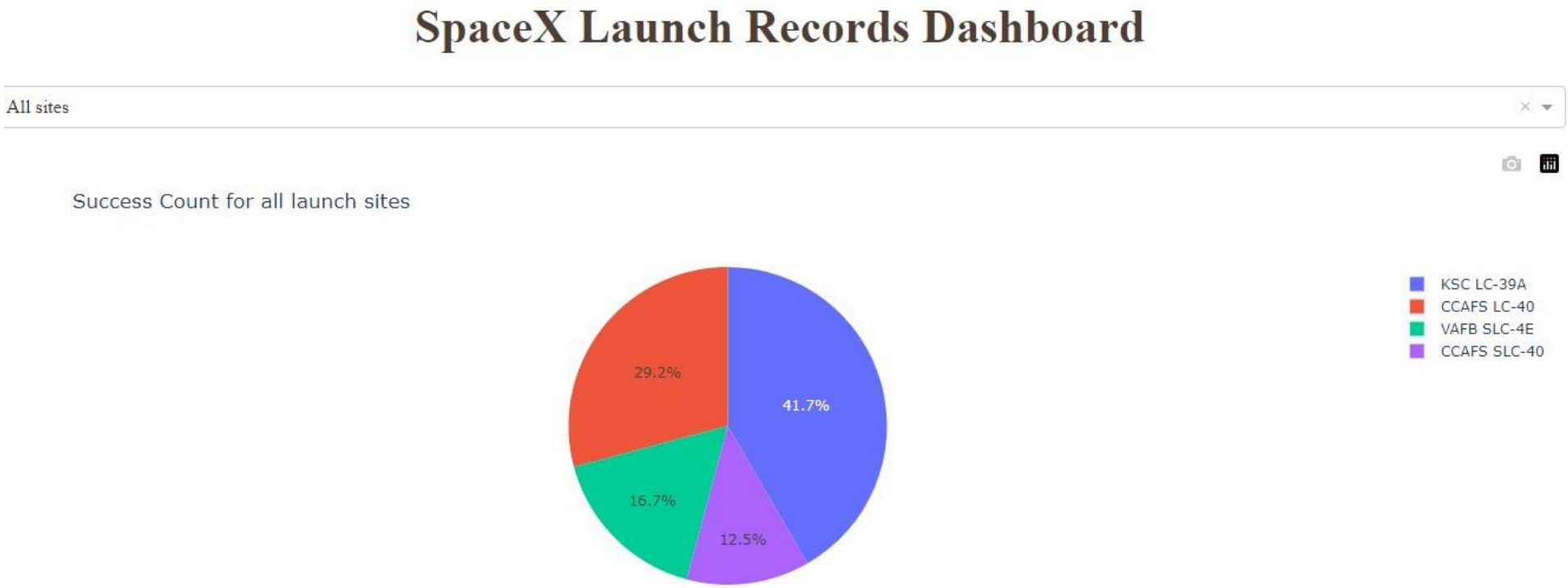
In this part we created dashboard to analyze and visualize the launch information for various launch sites.

https://github.com/Zlatko-Dz/Capstone/blob/main/spacex_dash_app%20(3).py

DASHBOARD TAB 1 / Dropdown menu for different SpaceX launch sites



DASHBOARD TAB 2 / @app callback with success count for all launch sites



DASHBOARD TAB 3 / Payload range and success count on Payload mass for all sites

RESULTS / Maschine learning prediction

- After standardizing the data we split the data into training and testing set.
- Using the Logistic Regression and GridSearchCV, SVM, Decision Tree and KNN we trained the model and examined the accuracy of the prediction in our test data set.
- Our accuracy score was over 0,83 which represents good result for prediction and confusion matrix gave us results in order to determine the case with the least "false positive result" - the min of "false positives" means that model's prediction of successful landing of SpaceX Falcon9 first stage is minimal when you check the true landing outcome, in this case unsuccessful landing outcome, from the test data set.
- As a final result we found Decision Tree as the best in prediction the launch outcome.
- The link of the Jupyter notebook is provided below:
  - ✓ https://github.com/Zlatko-Dz/Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite%20(1).ipynb

DISCUSSION

OVERALL FINDINGS & IMPLICATIONS

Findings

- The most successful Launch Site is KSC LC-39 A, followed by CCAFS SLC-40.
- The most successful Booster Version is FT followed by B4.
- The most succesfull landing is on drone ship followed by ground pad.

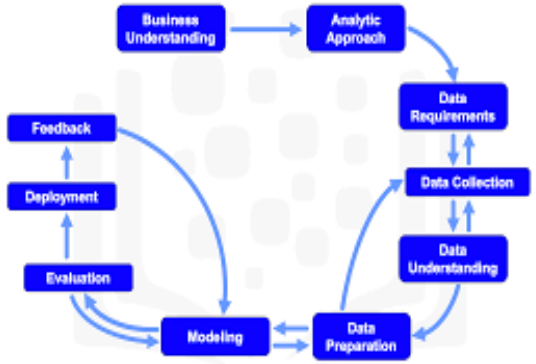
Implications

- The highest probability for SpaceX Falcon9 first stage to land successfully is to use the KSC LC-39A launching site.
- Booster Version FT returns the best probability for mission success.
- Selecting scenario with landing on a drone ship gives more chance for successful landing.

CONCLUSION

- The highest probability for successful project – successful landing of the first stage of SpaceX Falcon9 we would use:
- KSC LC – 39A launch site
- Booster Version FT
- Provide landing on a drone ship
- Decision Tree should be used depending on different launch variables in order to determine the scenario with the highest mission success rate.

APPENDIX



- Downloaded from:

https://www.researchgate.net/profile/Gerard-Schouten/publication/345261491/figure/fig2/AS:953987410976768@1604459731012/IBM-Data-Science-Methodology-9.jpg