Machine Learning Project 1

Jonathan Chacko Pattasseril

Marine Data Analysis

Reflection

# 1. Reflect on the analysis process, discussing any challenges encountered and how they were overcome.

The analysis of the Marine Data involved several stages & plenty of challenges were there in each stage.

If I were to break down according to the Stages in the Project it would be =

## Data Selection:

* That was the Easiest as I have always wanted to Analyse the Data the Company Provided me & I created a Restricted API with the Information needed for this Project.

## Data Cleaning, Standardisation & Pre-Processing:

* Understanding the Data & Possible Analysis was First Challenge
* Another Challenge was Understanding how to clean it & remove error’s WHILE not changing the Integrity of the Data.

## Exploratory Data Analysis (EDA):

* This Process Involved Creating Pivot Tables, Heatmap’s, etc…
* The main Challenge I faced was that the Original Data Set I had no proper Correlation between Columns (which I could anyway identify) so I had to get Creative & figure out Alternative’s which ended up me dividing the Table’s into Sub-groups.

## Statistical Inference:

* I found this Challenging as I’m still learning Chi2 & the other Statistic’s Formula’s
* Hence, I needed to do a lot of research on how this is done & why?

## Visualisation and Presentation:

* This was my Favorite Part the Only Challenge was Trying to Figure out how to Further Subdivide &or Categorise the Data and then Plot it onto the Same Graph the way I wanted.

## Other:

* Limited Time as I had a lot more Plan’s on how to Update this (& I will) More than I already Have.

# 2. Propose further questions for exploration or additional analyses that could be performed with more advanced techniques.

## Predictive Analysis/Forecasting:

The Whole Reason I wanted to work on this Data-set is because one of my End Goals for doing this Course is that I will be able to Help the Company Create Forecast’s & Optimise their Inventory, which should reduce Losses.

## Machine Learning for Pattern Recognition:

This would be a Solution for another Problem the Company Faces which is that User’s add a lot of Incorrect Data during their Manual Entry. I want a ML Program to identify Outliers &or Errors in the Data to Flag/Notify to the Management and Users.

## Updates to the API:

Merging Data from Other Tables which will support the Analysis.

## Updates to the Cleaning Process:

Have Planned Selenium Web Scrappers which should help fill up missing Values in Few Particular Columns.

## Advanced Data Visualisation:

Create an Interactive Dashboard for Company Management to help them make Decisions.