**Insights and Q&A**

This file would act as a tracking book of the Insights and challenges I acquired/ faced while building the **career-counsellor-ml** project. Along with the Insights, I would also keep the record of a few Questions and Answers which I think are important, and which really forced me to think about the concept deeply.

**I – Insights**

**Q – Questions**

**A – Answers**

**C – Challenges**

**I1 -** In the describe function, include = "all" parameter makes sure to include all the columns while drawing the statistics table.

**Q1 – Which type of graphs are used for which type of data?**Ans) Categorical variables 🡪 Bar plot (Matplotlib), Count plot (Seaborn)  
 Numerical variables 🡪 Histogram, Boxplot  
 Time-series 🡪 Line plot  
 Multiple Numeric variables 🡪 Heatmap (Just like in the project)

**Q2 – How to detect and handle missing values?**Ans) Missing values can be detected by “df.isnull( ).sum( )”. There are multiple ways to handle missing values – Dropping rows/ columns, Imputation methods (Replacing with Mean, Median and Mode, Interpolation methods, model based imputation), Missingness as a feature.

*Important: Write about the value error faced in step 6 soon after writing train\_test\_split function -* ***ValueError****: The least populated class in y has only 1 member, which is too few. The minimum number of groups for any class cannot be less than 2.*

*This is due to stratify = y not having enough members in the class to distribute evenly*