Project:

Hypothesis: Employee characteristics may be correlated to the number of tickets (issues) and their resolution and that insights may be drawn from these.

Task:

Build a logistic regression model by following the steps below:

1. Plot a descriptive statistics of the data. Unique categories in each column, total counts of each categories in each column, mean and median Tenure_Months
2. Take Issue_Severity as a dependent variable. Level 1 is most severe and level 4 least severe. This is a multinomial (issue_severity) classification task.
3. Predictor Variables: Issue_Category, Support_Channel, Building, Position, Tenure_Months, Career_Desc. Among these all of them are category variables except Tenure_Months. For the variable – **Building**, take MOBILE = Category 1, HOME OFFICE = Category 2, and the rest as Category 3. For all other category variables consider each type as separate categories.
4. Run a logistic regression using **StatsModel** and using **scikit learn**. Explain the coefficient significance using the **stats model** approach using the p-values and standard errors obtained. Using the **scikit learn** approach, divide the data into train and test and find the accuracy, ROC curve and confusion matrix. Display the confusion matrix, precision, recall, F1 and support
5. Use **Random Forest Classifier** to classify Issue_Severity. Use One-Hot Encoding and find feature importance. Plot ROC curve and compute classification accuracy and other metrics and compare with Logistic Regression.
6. Choose a **clustering** algorithm (FAMD or UMAP) from the following link for mixed data (numerical + categorical) and run an algorithm and plot the clusters after recoding numerical and categorical variables.

   https://medium.com/analytics-vidhya/the-ultimate-guide-for-clustering-mixed-data-1eefa0b4743b

7. Please extract support (justification) for the business insights and recommendations and create any visualizations (if needed) to support your insights and recommendations.
8. Please provide a specification laying out the roadmap for more and better actionable insights into tickets (issues) created by employees.