Customer Churn Prediction Report Summary

1. **Data Loading & Exploration:**
   Loaded Telco Customer Churn dataset and performed basic exploration (shape, data types,
   Missing values, statistics, and unique values).
2. **Preprocessing & EDA:**
   - Converted 'TotalCharges' to numeric and imputed missing values.
   - Dropped 'customerID' as it's non-informative.
   - Visualized churn distribution, distributions of numerical and categorical features.
   - Boxplots and churn rate analysis were done to assess feature behavior.
3. **Feature Engineering:**
   - Created 'TenureGroup' and 'AvgMonthlySpend' features.
   - Derived binary indicators for services and a new feature 'NumAdditionalServices'.
   - Replaced infinite values and cleaned the data.
4. **Data Preparation:**
   - Split data into training and testing sets.
   - Identified numerical and categorical features.
   - Built preprocessing pipeline using imputation, scaling, and one-hot encoding.
5. **Model Building & Evaluation:**
   - Trained and evaluated Logistic Regression, Random Forest, Gradient Boosting, and Neural Network models using pipelines.
   - Performed hyperparameter tuning with GridSearchCV.
   - Used metrics like accuracy, precision, recall, F1, ROC AUC, and plotted ROC curves and confusion matrices.
6. **Best Model:**
   - Random Forest provided the best performance among the models evaluated.
7. **Model Interpretation:**
   - Used SHAP and feature importance from Random Forest to interpret model predictions