**Title: Classification of Income Levels based on US Census Data**

**Abstract:** In this study, we aim to build a model to accurately classify low-income and high-income individuals using US census data. By understanding the attributes that contribute to affluency, we can gain insights into factors that impact income levels and inform policy-making. The study involves three main parts: data cleansing, data exploration, and model building using the K-Nearest Neighbors (KNN) algorithm.

**1. Introduction:**

* Background and significance of the problem
* Objective of the study

**2. Data Cleansing:**

* Description of the dataset used (source, size, features)
* Data cleaning techniques applied
* Handling missing values, duplicates, and outliers
* Encoding categorical variables
* Preparing the dataset for further analysis

**3. Data Exploration:**

* Overview of the dataset (summary statistics, distribution of variables)
* Visualization of income differences based on key attributes:
  + Occupation
  + Education
  + Additional attributes (e.g., marital status, race, gender)
* Discussion of key findings and insights from the visualizations

**4. Model Building and Interpretation:**

* Overview of the KNN algorithm
* Splitting the dataset into training and testing sets
* Feature selection and preparation
* Building multiple KNN models with different combinations of independent variables
* Determining the optimal value of K through hyperparameter tuning or cross-validation
* Evaluation of the models using performance metrics (e.g., accuracy, precision, recall, F1-score)
* Interpretation of the results and key learning points:
  + The importance of different independent variables in predicting income levels
  + Recommendations for policy-making based on the model's accuracy and performance

**5. Conclusion:**

* Summary of the study objectives and methodology
* Key findings and insights from the data exploration and model building process
* Implications for policy-making and improving income equality
* Limitations of the study and suggestions for future research