FINAL PROJECT OVERVIEW

This project is based on the data set called ERIMdata.xlsx that includes about 3,000 households in two midwestern cities in the United States. The data contain demographic information such as household incomes, number of household members, education levels of the heads of households as well as information on the purchases of several retail products such as frozen dinners and yogurt. The data were collected between 1985 and 1988 by a marketing research firm, AC Nielsen.

Your assignment is first to propose a business analytics plan based on the CRISP-DM framework and identify and complete the appropriate tasks for each of the six CRISP-DM phases. The project deliverables include a final written report and an oral presentation that should follow the outline shown below.

1. **Business understanding:** Describe the business opportunities that the data present and formulate relevant business questions.

2. **Data understanding**: Explore the data set with descriptive analytics tools and provide relevant information. Examine the possibility of supervised and unsupervised analysis techniques and identify possible variables for further analysis. Keep in mind the business opportunities and questions formulated in the first phase. The following criteria may also be considered as a guide.

   - Does a target variable exist?
   - Does the data set contain historical values of the target variables?
   - Does the data set have a sufficient number of observations to support data partitioning that may be required to answer the business question(s)?

3. **Data preparation:** Determine and perform the necessary data wrangling and preparation tasks based on the decision made during the business and data understanding phases. Explain the rationale for these tasks and document the changes that you have made to the data set.

4. **Modeling**: Consider the strengths and weaknesses of different modeling techniques. Implement the appropriate techniques, explain the rationale for your selection, and present relevant analysis results and interpretation. For the supervised techniques, determine whether to use classification or prediction models and explain your decision. Use appropriate data partitioning and performance measures to evaluate the competing models implemented in the modeling phase. Identify the best model(s).

5. **Evaluation:** Refocus on the business objectives of the project. Review the steps executed to construct the model to ensure no key business issues were overlooked. Evaluate whether the models have properly achieved the business objectives outlined during the business understanding phase. Formulate actionable recommendations based on the findings.

6. **Deployment:** Communicate the findings and relevant business insights with a written report and oral presentation that incorporate appropriate statistical information and visuals. The main focus should be placed on providing actionable business recommendations for a managerial and non-technical audience.

DATA DICTIONARY

HHID   The household's identification number

ResType   Types of residence: 1 for Apartment, 2 for Condo, 3 for Single Family, 4 for Multiple Family, 5 for Mobile, and 6 for Other.

ResStatus   Residence status: 1 for owned home, 2 for rented, and 3 for other.

HHlnc   The average annual income of a household; there are 14 categories for this variable.

HHNbr   The number of members in the household.

MWrkHrs   The average hours worked each week by the male head of household.

MEdu   Education level of the male head of household: values less than 9 imply varying education levels prior to a college degree, 9 for graduated from college, 10 for attended graduate school, and 11 for post-graduate degree.

FWrkHrs   The average hours worked each week by the female head of household.

FEdu   Education level of the female head of household. See MEdu for detail.

FBirth   The birth year of the female head of household.

F Rel   Relationship within the household: 1 for female head of household, 2 for male head of household, 3 for daughter, 4 for son, and 5 for other.

MBirth   The birth year of the male head of household.

M Rel   Relationship within the household: 1 for female head of household, 2 for male head of household, 3 for daughter, 4 for son, and 5 for other.

Cable   Whether or not the household has cable; 1 if yes, 0 otherwise.

Cats   Whether or not the household has cats; 1 if yes, 0 otherwise.

Dogs   Whether or not the household has dogs; 1 if yes, 0 otherwise.

YogExp   A household's yogurt expenditures (in$)

DinExp   A household's frozen dinner expenditures (in$)