

# UGN1 – UGN1 TASK 2: LOGISTIC REGRESSION ANALYSIS

STATISTICAL DATA MINING – D600

PRFA – UGN1

Preparation

Task Overview

Submissions

Evaluation Report

## EVALUATION REPORT – ATTEMPT 1 – REVISION NEEDED

### Overall Evaluator Comments

#### EVALUATOR COMMENTS

The submission logically develops a research problem seeking to predict IsLuxury classification based on readily available attributes. Please review the rubric's comments for information about aspects requiring revisions.

### A. GITLAB REPOSITORY

**Approaching Competence** The subgroup and project are created in GitLab, but 1 or more of the given actions are not completed, or they are completed incorrectly.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission logically seeks to determine if a luxury classification can be predicted based on independent attributes within the housing dataset. The response is incomplete because the GitLab link provided in the Word document is broken.

### B1. PROPOSAL OF QUESTION

**Competent** The submission proposes 1 question that is relevant to a real-world organizational situation and can be answered using logistic analysis.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission logically seeks to determine if it is feasible to forecast a luxury status based on home prices, crime rate, property tax, and previous sale price.

## B2. DEFINED GOAL

**Competent** The submission defines 1 reasonable goal for data analysis that is within the scope of the scenario and is represented in the available data.

There are no comments for this aspect.

## C1. VARIABLE IDENTIFICATION

**Competent** The submission logically identifies the dependent variable and at least 2 independent variables that are required to answer the research question. The submission also logically justifies the selection of these variables.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appropriately defines lsLuxury as the dependent variable and price, crime rate, property tax rate, and previous sales prices as the independent variables.

## C2. DESCRIPTIVE STATISTICS

**Approaching Competence** The submission provides screenshots of the descriptive statistics for some of the variables identified in part C1.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to reference an R file used for descriptive statistics. This will be assessed when revisions to previous task prompts are in place.

## C3. VISUALIZATIONS

**Approaching Competence** Univariate and bivariate visualizations of the distributions of some variables from part C1 are provided, or the dependent variable is not included in some bivariate visualizations.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission indicates that the data plots are found in an R-file. The response is incomplete because the GitLab

link is broken.

## D1. SPLITTING THE DATA

**Approaching Competence** The submission provides training and test datasets, but the split is not reasonably proportioned.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission indicates a 70/30 data split was applied to train and evaluate the logistic regression model. This will be assessed when revisions to previous task prompts are in place.

## D2. MODEL OPTIMIZATION

**Not Evident** The submission does not demonstrate model optimization.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not appear to indicate an explanation of the optimization technique, nor does it include a screenshot or access to the code.

## D3. CONFUSION MATRIX AND ACCURACY

**Approaching Competence** The confusion matrix and accuracy of a model is provided prior to model optimization.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to provide a training confusion matrix along with predictive accuracy. This will be assessed when revisions to previous task prompts are in place.

## D4. PREDICTION

**Approaching Competence** A screenshot of the prediction results is provided, but the optimized model was not used.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to provide a test confusion matrix along with predictive accuracy. This will be assessed when revisions to previous task prompts are in place.

### E1. PACKAGES OR LIBRARIES LIST

**Not Evident** The submission does not list the packages or libraries chosen for Python or R.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not list any of the packages or libraries used in the analysis.

### E2. OPTIMIZATION METHOD

**Not Evident** The submission does not discuss the method used to optimize the model.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not appear to discuss the applied optimization method.

### E3. METHOD JUSTIFICATION

**Not Evident** The submission does not justify the approach used to optimize the model.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not appear to discuss rationale for the method used to optimize the model.

### E4. ASSUMPTION SUMMARY

**Approaching Competence** The submission summarizes less than four assumptions of logistic regression.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to vaguely verify the assumptions without any text explanation. This will be assessed when revisions to previous task prompts are in place.

## E5. VERIFICATION OF ASSUMPTIONS

**Approaching Competence** A code snippet or a screenshot is submitted to verify *some* of the assumptions discussed in part E4.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to vaguely verify the assumptions without any text explanation. This will be assessed when revisions to previous task prompts are in place.

## E6. EQUATION

**Approaching Competence** The regression equation is not fully provided or the discussion of the coefficient estimates is incorrect.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to provide the logistic regression equation without the intercept or coefficients. This will be assessed when revisions to previous task prompts are in place.

## E7. MODEL METRICS

**Approaching Competence** The discussion of the model metrics does not address *each* of the identified parts.

### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to provide limited performance evaluation metrics for logistic regression. This will be assessed when revisions to previous task prompts are in place.

## E8. RESULTS AND IMPLICATIONS

**Not Evident** The submission does not discuss both the results and implications of the prediction analysis.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not appear to discuss any results or implications for the conducted analysis.

### E9. COURSE OF ACTION

**Not Evident** The submission does not recommend a course of action for the real-world organizational situation from part B1.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission does not appear to discuss any recommended course of action for the conducted analysis.

### F. PANOPTO RECORDING

**Approaching Competence** The Panopto video recording provided is missing either a screen share of the presenter demonstrating the functionality of the code used or a discussion commenting on the programming environment. Or either the demonstration or the summary is inaccurate.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to include a 7:38-minute Panopto video describing the R-script used. This will be assessed when revisions to previous task prompts are in place.

### G. SOURCES

**Approaching Competence** The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations or reference list is incomplete or inaccurate.

#### EVALUATOR COMMENTS: ATTEMPT 1

The submission appears to indicate that no sources were used for this assignment. This will be assessed when revisions to previous task prompts are in place.

### H. PROFESSIONAL COMMUNICATION

**Approaching Competence** This submission includes substantial errors in professional communication related to grammar, sentence fluency, contextual spelling, or punctuation. Specific errors have been identified by Grammarly for Education under the Correctness category.

EVALUATOR COMMENTS: ATTEMPT 1

The submission is very limited in detail and lacks clear explanations; however, it appears to be mostly free from grammatical and spelling errors, as reflected by the Grammarly score of 95/100.