**TASK: LINEAR REGRESSION ANALYSIS SCENARIO**

A major healthcare system wants to reduce job-related stress among its employees. In particular, the nursing staff endures a significant amount of pressure due to 12-hour rotating shifts, required overtime, and challenging job assignments. A few years ago, the hospital's human resources (HR) department discovered that its nurse turnover rate was much higher than any other professional position at the hospital. Those nurses leaving employment cited work conditions and emotional stress as their main concerns.

To curb this trend, the hospital proactively developed an employee well-being program. For over three years, a program manager has sponsored monthly activities specifically designed to improve employee morale and reduce stress. Participation in the program has been strictly voluntary, with a $50 cash bonus awarded to all employees who consistently complete the various well-being activities every quarter. Nurse participation is specifically being tracked to see if the program has successfully improved job satisfaction and reduced job turnover.

Enrollment in the program has generally increased over time. Sponsored activities are currently developed by a full-time athletic trainer, a massage therapist, and a yoga instructor. These professionals also provide individualized services to participants upon request. A full-time program coordinator is tasked with administering the program's day-to-day operations, including the responsibility of expanding employee participation through periodic marketing emails that tout or attempt to see the program benefits and upcoming activities. Of course, there are significant internal costs associated with subsidizing this program. Nevertheless, it is widely believed that these expenses are minuscule or small when compared to the substantial costs associated with replacing employees—nurses in particular*. Accordingly, the program costs and predicted savings are not being considered as part of this initial analysis*.

*The purpose of this study is to determine if there is a significant relationship between the monthly rate of nurse participation and the nurse attrition rate over the span of 36 months*. Since attrition can be considered a lagging indicator of job satisfaction, the attrition rate is retroactively tied back to each nurse’s enrollment month. Categorized data for program participation and nurse attrition was extracted and assembled from the program enrollment database and the HR employee database, with the data for analysis shown below.

The hospital executive council is reviewing the program’s efficacy in reducing nurse attrition through program participation as part of its routine funding plan for the next five years. However, the initial meeting will primarily focus on program efficacy and predictions for future enrollment growth and nurse retention.

**INSTRUCTIONS**

A. Describe a business question that could be answered by applying linear regression analysis and is derived from the scenario stated above.

B. Describe the data provided in the attached “Linear Regression Analysis Resources” by doing the following:

1. Describe the relevant data characteristics for your linear regression analysis, including each of the following:

* *the independent variable(s)*
* *the dependent variable*
* *type of data*
* *quantity of data*

2. Create a graphical display of the data using a scatter plot or line chart, including each of the following:

* *chart title*
* *legend*
* *axis titles*
* *data intervals*

***Note: This display should be a summary or representation of the data provided, not raw data***.

C. Report how you analyzed the data using linear regression by doing the following:

1. Provide the output and calculations of the linear regression analysis you performed.

***Note: You may submit the analysis output and calculations using a separate spreadsheet attachment***

***Note: The output should include the output from the software you used to perform the analysis which in this case Microsoft Excel****.*

2. Justify why linear regression is the appropriate analysis technique for predicting the dependent variable, including relevant details from the scenario to support your justification.

D. Describe the implications of your data analysis from the scenario by doing the following:

1. State the null hypothesis for this linear regression analysis.

2. Interpret the results of the data analysis by doing the following:

a. Discuss the *goodness of fit* with the supporting test statistic from your linear regression analysis output.

b. Discuss the *significance of the independent variable(s)* with support from your linear regression analysis results.

c. Create the *linear equation* and explain its purpose using your analysis results.

3. Discuss a limitation of the research that could affect a recommended course of action.

4. Recommend a course of action that aligns with your linear regression analysis results.

***Note: Your recommendation should focus on the results of your analytic technique output from part C1.***

E. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased or summarized.

F. Demonstrate professional communication in the content and presentation of your submission.

**RUBRIC**

**A1:DATA-DRIVEN DECISION**

* ***NOT EVIDENT****: A description of a business question is not provided*.
* ***APPROACHING COMPETENCE****: The business question described is not derived from the scenario or is not consistent with the application of linear regression analysis*.
* ***COMPETENT****: The business question described is derived from the scenario and is consistent with the application of linear regression analysis.*

**B1: DESCRIPTION OF RELEVANT DATA**

* ***NOT EVIDENT****: A description of data characteristics is not provided.*
* ***APPROACHING COMPETENCE****: The description does not accurately address the relevant data characteristics of the linear regression analysis, or 1 or more of the given elements are not included.*
* ***COMPETENT****: The description accurately addresses the relevant data characteristics of the linear regression analysis, including each of the given elements*.

**B2: GRAPHICAL DISPLAY**

* ***NOT EVIDENT****: A graphical display of the data is not provided.*
* ***APPROACHING COMPETENCE****: The graphical display of the data does not use a scatter plot or line chart, does not accurately represent the data, or does not include 1 or more of the given elements.*
* ***COMPETENT****: The graphical display of the data uses a scatter plot or line chart to accurately represent the data, including each of the given elements***.**

**C1: ANALYSIS OUTPUT AND CALCULATIONS**

* ***NOT EVIDENT****: A linear regression analysis output and calculations are not provided.*
* ***APPROACHING COMPETENCE****: The linear regression analysis output or calculations provided contain 1 or more inaccuracies.*
* ***COMPETENT****: The linear regression analysis output and calculations provided are accurate.*

**C2: ANALYSIS TECHNIQUE JUSTIFICATION**

* ***NOT EVIDENT****: A justification of linear regression analysis is not provided.*
* ***APPROACHING COMPETENCE****: The justification of linear regression analysis does not logically explain why it is the appropriate technique to predict the dependent variable, or the justification is not supported with relevant details from the scenario.*
* ***COMPETENT****: The justification of linear regression analysis logically explains why it is the appropriate technique to predict the dependent variable and is supported with relevant details from the scenario.*

**D1: NULL HYPOTHESIS**

* ***NOT EVIDENT****: A null hypothesis is not provided.*
* ***APPROACHING COMPETENCE****: The null hypothesis provided is not accurately stated for the linear regression analysis or is not relevant to the scenario.*
* ***COMPETENT****: The null hypothesis provided is accurately stated for the linear regression analysis and is relevant to the scenario.*

**D2A: GOODNESS OF FIT**

* ***NOT EVIDENT****: A discussion of the goodness of fit is not provided.*
* ***APPROACHING COMPETENCE****: The discussion of goodness of fit is not supported with the appropriate test statistic from the linear regression analysis output, or the interpretation of the test statistic is inaccurate.*
* ***COMPETENT****: The discussion of goodness of fit is supported with the appropriate test statistic from the linear regression analysis output, and the interpretation of the test statistic is accurate***.**

**D2B: INDEPENDENT VARIABLE(S) SIGNIFICANCE**

* ***NOT EVIDENT***: A discussion of the significance of the independent variable(s) is not provided.
* ***APPROACHING COMPETENCE***: The discussion of the significance of the independent variable(s) is not logical or is not supported by the linear regression analysis results, or the associated interpretation of the comparison to the significance level is inaccurate.
* ***COMPETENT***: The discussion logically addresses the significance of the independent variable(s) with support from the linear regression analysis results, and the associated interpretation of the comparison to the significance level is accurate.

**D2C: LINEAR EQUATION**

* ***NOT EVIDENT***: A linear equation and an explanation of the equation are not provided.
* ***APPROACHING COMPETENCE***: The linear equation provided is not accurate, or the explanation does not logically address the purpose of the linear equation using analysis results.
* ***COMPETENT***: The linear equation provided is accurate, and the explanation logically addresses the purpose of the linear equation using analysis results.