Assignment 10 - Comprehensive assignment

INSTRUCTIONAL DETAILS

Please refer to the previous 9 assignments for syntax examples and method specific instructional details.

The BrightSpace course shell also contains additional [Python syntax examples for each of the methods taught](#).

BACKGROUND

The comprehensive assignment for Wentworth's Applied Business Analytics course involves analyzing real-world data to solve a series of business problems using various statistical and data analysis techniques. You will need to complete the following steps:

Problem definition: Identifying a specific business problem after reading a set of research question. Defining the objective of the analysis.

Data collection: Review, clean, and import the relevant data.

Exploratory Data Analysis (EDA): Conducting an initial analysis of the data to understand its patterns and relationships, using visualizations and summary statistics.

Model building: Selecting appropriate statistical models, such as regression, decision trees, etc., and fitting them to the data to make predictions or inferences.

Validation: Evaluating the model's performance and accuracy using metrics like R-squared, confusion matrix, etc.

Model improvement: Iteratively improving the model by considering additional variables, transforming variables, etc.

Result interpretation and conclusion: Interpreting the results of the analysis, drawing conclusions, and providing recommendations based on the insights obtained.

The assignment requires the use of Python and a presentation of the results in a professional report.

RESEARCH QUESTION

Descriptive research questions - -

How does age impact job satisfaction scores among employees?

Is there a significant difference in job satisfaction scores between male and female employees?

How does ethnicity impact job satisfaction scores among employees?

How does country of origin influence job satisfaction scores among employees?

Is there an interaction effect between age and gender on job satisfaction scores?

Is there an interaction effect between ethnicity and country on job satisfaction scores?

Do certain age groups have higher job satisfaction scores than others based on gender, ethnicity, and country?

Do certain ethnic groups have higher job satisfaction scores than others based on age, gender, and country?

Do certain countries have higher job satisfaction scores than others based on age, gender, and ethnicity?

Is there a difference in job satisfaction scores among employees from different countries within specific age and ethnicity groups?

Inferential research questions - -

Is there a statistically significant difference in job satisfaction scores between employees of different ages?

Is there a statistically significant difference in job satisfaction scores between male and female employees?

Is there a statistically significant difference in job satisfaction scores between employees of different ethnicities?

Is there a statistically significant difference in job satisfaction scores between employees from different countries?

Is there a statistically significant interaction effect between age and gender on job satisfaction scores?

Is there a statistically significant interaction effect between ethnicity and country on job satisfaction scores?

Is there a statistically significant difference in job satisfaction scores between different age and gender groups?

Is there a statistically significant difference in job satisfaction scores between different ethnic and country groups?

Is there a statistically significant difference in job satisfaction scores between employees of different ages, ethnicities, and countries?

Can a logistic regression model be developed to predict job satisfaction scores based on age, gender, ethnicity, and country?

What is the significance of each predictor in the logistic regression model for job satisfaction scores?

Is there a significant relationship between age and job satisfaction scores?

Is there a significant relationship between gender and job satisfaction scores?

Is there a significant relationship between ethnicity and job satisfaction scores?

Is there a significant relationship between country and job satisfaction scores?

What is the strength and direction of the relationship between age and job satisfaction scores?

What is the strength and direction of the relationship between gender and job satisfaction scores?

What is the strength and direction of the relationship between ethnicity and job satisfaction scores?

What is the strength and direction of the relationship between country and job satisfaction scores?

REQUIREMENTS FOR SUBMISSION

A successful assignment will include the following - -

Data handling and cleaning: Ability to import, manipulate, and clean data from various sources using libraries such as Pandas.

Exploratory Data Analysis (EDA): Knowledge of various visualization techniques and statistical methods for data exploration, such as histograms, box plots, and summary statistics.

Model building: Knowledge of machine learning algorithms, such as regression, decision trees, etc., and the ability to implement them using libraries such as scikit-learn.

Validation: Ability to evaluate the performance of models using metrics such as R-squared, confusion matrix, etc.

Result interpretation: Ability to interpret the results of the analysis and draw meaningful insights.

Reporting: Ability to present the results of the analysis in a clear and professional manner using tools such as Matplotlib and Seaborn for visualizations and Jupyter notebooks for report writing.

The following Python libraries and tools should be familiar to complete such an assignment:

Pandas for data manipulation and cleaning
Numpy for numerical operations
Matplotlib and Seaborn for data visualization
Scikit-learn for machine learning algorithms

In addition to the technical skills, the assignment should also require clear and concise written communication, critical thinking, and attention to detail.

FORMATTING
See "Assignment 1 - Descriptives" for a detailed list of assignment formatting guidelines. Also, assignment formatting guidelines can be found in the course document cache.

DATASET DETAILS (all data sets can be found here or here)
job_satisfaction.csv

DATASET FIELDS
Satisfaction - satisfaction score for the employee
important_recognitionAchievements - employee score for importance of recognition of achievements
satisfied_careerAdvance - employee satisfaction with career advancement
satisfied_recognitionAchievements - employee satisfaction with recognition of achievements
age - age of employee
gender - gender of employee
ethnicity - ethnicity of employee
country - country of residence of employee