ACADGILD

SESSION: 11 To 15

Assignment

Table of Contents

1. Introduction

This assignment will help you understand the concepts learned in the session.

1. Objective

This assignment will test your skills on the concepts of SQL analytics.

1. Prerequisites

Not applicable.

1. Associated Data Files

Not applicable.

1. Problem Statement

Task 1:

1. Use the given link below and locate the bank marketing dataset. **Data Set Link**

Perform the below operations:

a. Is there any association between Job and default?
b. Is there any significant difference in duration of last call between people having housing loan or not?
c. Is there any association between consumer price index and consumer?
d. Is the employment variation rate consistent across job types?
e. Is the employment variation rate same across education?
f. Which group is more confident?

Task 2:

1. Use the given link: **Data Set.**

Answer the below questions:

a. What are the assumptions of ANOVA, test it out?

b. Why the ANOVA test? Is there any other way to answer the above question?

Task 3:

1. Use the given link: **Data Set.**

Answer the below questions:
a. Visualize the correlation between all variables in a meaningful and clear way of representing. Find out top 3 reasons for having more crime in a city.

b. What is the difference between covariance and correlation? Take an example from this dataset and show the differences if any?

Task 4:

1. Use the below given data set

DataSet

Problem- prediction of the number of comments in the upcoming 24 hours on those blogs, The train data was generated from different base times that may temporally overlap. Therefore, if you simply split the train into disjoint partitions, the underlying time intervals may overlap. Therefore, the you should use the provided, temporally disjoint train and test splits to ensure that the evaluation is fair.

a. Read the dataset and identify the right features

b. Clean dataset, impute missing values and perform exploratory data analysis.

c. Visualize the dataset and make inferences from that

d. Perform any 3 hypothesis tests using columns of your choice, make conclusions

e. Create a linear regression model to predict the number of comments in the next 24 hours (relative to basetime)

f. Fine tune the model and represent important features

g. Interpret the summary of the linear model

h. Report the test accuracy vs. the training accuracy

i. Interpret the final model coefficients

j. Plot the model result and compare it with assumptions of the model

Task 5:

1. Use the below-given data set

DataSet

a. Predict the no of comments in next H hrs

b. Use regression technique

c. Report the training accuracy and test accuracy

1. ## Expected Output

Solution report with commands, explanation of commands, and screenshots of the output should be submitted in .pdf format on GitHub the same GitHub should expected to submit on student dashboard. This assignment contains 700 marks and will be evaluated within 14 days of submission.