



Act Locally, Impact Globally

FRONTIER TECHNOLOGY INSTITUTE

DATA SCIENCE CERTIFICATION

MODULE -II : EXPLORATORY DATA ANALYSIS

HOME WORK -3

You are given a Census Dataset . Create a Python Notebook and perform the following tasks:

1. Using Chi Square statistic find out if there is any attribute for which Null Hypothesis (H_0 : There is no association or dependency between two variables) is True . You have to consider all the categorical attributes for this . Use $\alpha=0.05$
2. Discretize the age dimension so that it is converted to a categorical type.
3. Change the class variable to a quantitative attribute by converting $\geq 50k$ as 1 and $< 50k$ as 0. Now compute the Pearson correlation between class variable and all other continuous variables.
4. Apply manual encoding to education , one hot encoding to gender and race and label encoding to all other remaining categorical attributes. Do not add additional columns. Make changes to the original column.
5. Apply min-max normalization to fnlwgt , decimal scaling to education_num and z-score normalization to all other remaining continuous attributes
6. Make a report in Microsoft Word and submit along with Python Notebook.

Deadline : Submit through Email by Sunday December 01 , 2019