

FRONTIER TECHNOLOGY INSTITUTE

DATA SCIENCE CERTIFICATION

MODULE -II: EXPLORATOTY DATA ANALYSIS

HOME WORK-1

You are given a Sales Dataset . Create a Python Notebook and perform the Exploratory Data Analysis . You should do the following.

- 1. Perform basic data quality assessment including missing value analysis and incorrect value analysis. Replace all incorrect values with NaN and then remove all such records with NaN values. You can impute missing values using MCT where ever applicable
- 2. To Perform univariate analysis , plot the most relevant chart for each attribute (You may leave ID etc) . You may Plot Histograms for categorical attributes and Boxplots or line charts for all numeric attributes .
- 3. Submit A python notebook and a report of your analysis and findings (10-20 Pages)

Additional Resources:

- 1. https://www.kaggle.com/residentmario/univariate-plotting-with-pandas
- 2. http://walkerke.github.io/geog30323/slides/eda-1/#/21
- 3. https://towardsdatascience.com/build-the-story-around-data-using-exploratory-data-analysis-and-pandas-c85bf3beff87

Deadline: Submit through Email by Next Sunday November 17, 2019