**42047: Data Processing with Python**

**Assignment (Part C)**

**Report: Data Analysis and Visualization**

**Student Name and ID:** [Add names and student ID]

**Date:** [Add date]

**Table of Contents**

[**Abstract (1pts)** 3](#_30j0zll)

[**1. Introduction and Background (4Pts)** 4](#_1fob9te)

[**2 Overview of the Data Analysis Pipeline (20pts)** 5](#_3znysh7)

[**3 Discussion and Conclusions (5pts)** 6](#_2et92p0)

[**4 References** 6](#_tyjcwt)

**Abstract (1pts)**

<Add a short abstract>

**1. Introduction and Background (4Pts)**

[Provide a short introduction of the Data analysis task and the dataset]

**1.1 The problem you tried to solve (1pts)**

[Here you describe the problem *in your own words*!]

**1.2 Business Question (1Pts)**

[What is the business question you want to answer as a result of the data exploratory analysis.]

**1.3 Dataset (2Pts)**

[Describe briefly the dataset you have used, the total number of samples, attributes, reference (link, publication, etc.) and add some screenshot from the dataset if appropriate. List the attributes you would like to explore to answer your business question].

**2 Overview of the Data Analysis Pipeline (20pts)**

[Describe the different parts of your exploratory data analysis pipeline in details. Add a flow diagram/flowchart of your system (if appropriate), and briefly explain each of the steps]

**2.1 Flow Diagram/Flowchart/Work Flow** [Remove what is not applicable]

[Briefly describe the workflow and add a flow diagram/flowchart]

**2.2 Data Preparation (5Pts)**

[Briefly describe briefly what are the techniques you have used for data preparation, use of head, tail, statistical methods (such mean, median, mode, etc.) for finding dataset information etc. and their interpretation, appropriate plot for checking data distribution. You can add screenshots for the plots.]

**2.3 Missing value exploration (5Pts)**

[Briefly describe with Screenshots and explain the how you have explored whether the dataset has missing values, use appropriate plot/visualization techniques to find missing values, and how you have handled them, etc.]

**2.4 Outlier identification (5Pts)**

[Use of Appropriate visualization techniques to identify whether there are outliers in the dataset, and take appropriate action to removed/handle them. You can include screenshots of the plots, etc.]

**2.5 Data Visualization (5Pts)**

[Use of Appropriate visualization techniques according attribute types, plots such as distribution, pair, bar, pie, box, etc. Provide some interpretation with reference to the plots.]

**3 Discussion and Conclusions (5pts)**

[Summarize your data exploration experiment, answer your business question, etc. Add your final thoughts.]

**4 References**

[References should be cited (i.e. actually be referred to) at the appropriate place in your text, they should visibly influence your document, and they should convey as much information as possible to the reader.]

Example:

[1] Shaoqing Ren, Kaiming He, Ross B. Girshick and Jian Sun, *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, <http://arxiv.org/abs/1506.01497>, 2015