Data Wrangling Report

Introduction

This report describes the implementation of the whole process of data analytics by using actual dataset.It starts from data collection and goes to analysis and visualization.

1. **Data Collection**
   For this project, there are three ways of collecting datasets.One of the dataset of this project has come from twitter's @weratedogs account. The dataset from the twitter account has missed some of the necessary variables. In order to enrich the dataset, data has been collected from api by using python libraries. BeautifulSoup is used to collect the missing data from twitter api. On top of that, another dataset has been read from a txt file too.
2. **Data Assessing**
   In this process, the collected data is assessed both visually and programmatically. Word processors can be used to assess the dataset visually. In order to assess programmatically different methods from the Pandas library is used. We use methods like, info, shape, isNull and so on.
3. **Data Cleaning**
   During the data assessment part, 8 quality and 3 tidness issues are recorded to be cleaned in this process. Before cleaning the dataset, it will be copied to another dataframe. After that, it is cleaned with the define-code-test procedure. First we define what the action that needs to be taken for a specific data issue. The next step is writing the code to solve the issue. Finally, we will test if the correct solution is applied for each issue
4. **Data Storing**
   After cleaning the dataset, It is compiled into one file and saved in a csv file by using Pandas library
5. **Data Analyzing and Visualization**
   Once the dataset is compiled together, it will be analyzed and visualized by using different python libraries. First the file is loaded into a dataframe. Next, we use Pandas library to assess the dataset. Once it is assessed

programmatically, it will be analyzed. Last but not least, the analyzed dataset is visualized using Matplotlib or seaborn libraries.