WeRateDogs Data Wrangling Process

Introduction

This research was conducted on behalf of the Twitter account on rating dogs, known as "WeRateDogs". In this analysis, the implementation of data wrangling has been used with its three main features: gathering, assessing, and cleaning the given datasets. All these parts were clearly represented and described in this report.

Gathering data

Initially, I collected the given data by using Python's Pandas and Requests libraries. Thus, directly downloaded the records from the archive. An API Request has been utilized by dint of the Tweepy library. As a consequence of these findings, whole three tables were set up.

Assessing data

In the second phase, data assessments were conducted visually and programmatically in depth. Overall, eight quality and three tidiness issues were deeply dissected. In regard to quality problems, missing values, wrong data types, and invalid data cases were found. Coming to tidiness-related ones, extra columns and separate tables that can be merged were identified.

Cleaning data

The third stage was about cleaning the detected affairs in the previous step. For each issue, the individual approach was carried out. This includes dropping unnecessary columns, editing the wrong information entries, and merging tables. All the solutions were defined, coded, and tested thoroughly.

Storing data

Hence, all the sorted and cleaned records were stored in one complete master dataset.

Analyzing and Visualizing data

Using the polished data source, in total, four different types of charts were depicted. With the help of these discoveries, four insights were obtained.

Conclusion

To summarize, I acquired the core methods for wrangling data by implementing the well-described concepts. Mainly, I broadened my horizons to analyze any dataset regardless of its difficulty levels. Using the advanced Python libraries, I managed to gather complex datasets and assess them in various ways. Most importantly, I developed my data cleaning skills to polish high-level data collection. Doubtlessly, these acquired skills will aid me to prosper in my future career in the data science field.