Submitted By:
Ashutosh Verma

Expenditure Analysis Report

Contents of the file:

- One CSV contains the records in which the categories of the expenditure are generalised and grouped into 8 categories.
- Result CSV containing the following information:
  Name of user
  The amount spent on the specific category
  The month and year of the expenditure

Methodology:

The data has been cleaned by the following tasks:

- Removing whitespaces from the values
- Converting the time date to standard formats(Pandas date time)
- Removing NaN records.
- Removing rows with mismatched columns.

We have initially counted the unique number of categories present in the dataset, then we parsed the categories and the keywords present in each category were analyzed and processed to generate **8 generic categories of budget expenses namely: entertainment General, shopping, travel, cab, food, book.**
On the basis of these generated categories, the expenditure of each user in each month and also keeping track of the year, the total expense was calculated and the result CSV was generated in the required format.

Modules Used:
Python3 Pandas

Tool : Jupyter Notebook

Description of contents:
The folder contains 3 CSV's, given dataset(CleanStatements),
New_Data_State.CSV - New dataset formed after categorizing tasks.
Category_wise_expense_data- The resulting CSV with all the information for a user for a particular month and year.

OneBanc_assignment_Code.ipynb - Code for processing and generating the results(python code).

To Run the file open the ipynb file in any python environment like jupyter, colab etc and run the code.