109550198 卜銳凱

HW2

Report

This report describes the analysis of the dataset from the CSV file "2023 June Unemployment Rate by County (Percent).csv." The basic goal of the code is to load the data, do preliminary data exploration, and illustrate the presence of missing values.

**Code Summary:**

The provided code accomplishes the following tasks:

Data Loading: The code reads the dataset from the CSV file and saves it in a Pandas Data Frame with the name dataset.

Data Splitting:

It divides the dataset into two arrays, x and y, where x contains all except the last column and y contains the last column. This is commonly done to differentiate between features (independent variables) and the desired variable (dependent variable).

A screenshot of a computer

Description automatically generated

Data Exploration:

It shows the first ten rows of the Data Frame, providing an overview of the data's structure and content.

It computes and displays summary statistics, providing insights into the central patterns and spreads of numeric columns.

It discovers and counts missing values in the dataset and visualizes them using a heatmap.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Usage: The code can be used to get a sense of the structure and quality of the provided dataset. It lets users to inspect the data, look for missing values, and depict the distribution of such values.

**Discuss possible problems you plan to investigate for future studies**.

For future studies, I plan to investigate:

Data Cleaning: t is critical to handle missing values effectively. Consider imputation or removal of rows or columns with missing values, depending on the amount of the missing data.

Exploratory Data Analysis (EDA): To acquire deeper insights, try undertaking more complete EDA, such as investigating connections between variables, displaying distributions, and detecting outliers.

Data Visualization: Extend the visualization capabilities to incorporate different sorts of plots, such as histograms, box plots, scatter plots, or time series plots, to expose new insights in the data.

The provided code is intended to be used as a first step in evaluating the "2023 June Unemployment Rate by County" dataset. It provides data loading, basic exploration, and missing value visualization. To extract useful information and enable informed decision-making, more data cleaning and in-depth exploratory analysis are recommended.