README - Data Analysis and Combination Probability Tool

Overview

This Python program is designed to process CSV data, compute various probabilities, and export the results. It allows for analyzing combinations of data columns, calculating probabilities, entropies, chi-squared values, and more.

Features

- Load and process data from CSV files.
- Calculate column-wise probabilities.
- Generate and analyze combinations of data columns.
- Calculate joint probabilities, information measures, and chi-squared values.
- Export calculated data to CSV files.

Setup and Requirements

- Python 3.x
- Pandas Library
- Math Library
- CSV File located in the directory [.\dist\ ]

Installation

Ensure that Python 3 and Pandas are installed on your system. If Pandas is not installed, you can install it via pip:

```
pip install pandas
```

Main Functions

CSVPreprocessor: Loads and processes CSV files. *ColumnProbability*: Calculates probabilities for each column in the data. *CombinationProbability*: Analyzes combinations of columns for joint probabilities and other statistical measures. *Export to CSV*: Exports the results of the calculations to a CSV file.

Usage

Load Data: Start by loading your CSV file. You will be prompted to enter the file path and the desired data range (e.g., 'A1:D10'). ***Calculate Probabilities***: The program automatically calculates individual column probabilities and stores them. ***Explore Combinations***: Use the 'perms' command to explore header combinations. You can specify the order (e.g., '5' for up to order 5, '15' for only order 5). ***View Data***: The 'viewdata' command allows you to inspect the currently loaded DataFrame, probability results, header combinations, and combination probabilities. ***Export Data***: Use the 'export' command to export the combination probabilities to CSV files. You will be prompted to enter specific headers and the maximum order for export. Only rows with AssociationValidity = True will be exported. ***Additional Commands***: Use 'show', 'table', and 'header' commands to display specific column probabilities, the entire data table, and the list of headers, respectively.

Command List

show: Display probabilities for a specific column. **table**: Show the entire data table. **header**: List all headers/column names. **viewdata**: View detailed data stored in the program. **perms**: Explore header combinations. **export**: Export combination probabilities to CSV files. **quit**: Exit the program.

CSV File Processing

The program can process specific ranges within a CSV file. Specify the range in the format 'A1:D10' when prompted.

Probability and Statistical Calculations

Calculates column-wise probabilities. Analyzes combinations of columns for joint probabilities. Computes information measures, chi-squared values, and significance. Special calculation for combinations of order 2 and higher.

Examples

- Load CSV: Enter the path to your CSV file: data.csv
- View Data: Enter a command: viewdata
- Export Data:

```
Enter the path to your CSV file: path/to/your/file.csv
Enter the desired data range (e.g., 'A1:D10'): A1:D10
Enter the support threshold (default 0): 0.05
Command: ('show', 'table', 'header', 'viewdata', 'perms', 'export', 'quit')
Enter a command: export
Enter headers to export (separated by commas): Header1, Header2
Enter the maximum order to export (2-6): 3
```

Notes

Ensure the CSV file path and data range are correct. The export functionality allows for selecting specific headers and the maximum order of combinations to be included in the export. Ensure your CSV file is correctly formatted and accessible at the provided path. The tool handles a range of data analysis tasks, but it's primarily focused on association patterns and their significance.