NikaYz /
DataVisualization_Assignment3 🔒

<> **Code**   ⊙ Issues   ⑂ Pull requests   ▶ Actions   ⊞ Projects   ⊘ Security   ⌁ Insigh

⑂ main ▾   **DataVisualization_Assignment3** / **README.md** ⎘   ···

NikaYz Update README.md   1bca02e · now   🕘

85 lines (54 loc) · 3.3 KB

Preview   Code   Blame   Raw ⎘ ⬇   ✎ ▾   ☰

Visual Analytics- DAS732 - A3

This repository contains analytics visualization for dataset : Higher Education Attrition Rate of Australia for 2005 - 2013, as done in Assignment 1. Please refer to Assignment 1 for the basic inference as this is the extension of that.

Folder Structure

- **/Story1_DifferentPlots**: Contains code in python to generate various other plots.
- **/Story2_OtherInferences**: Contains code in python to generate plots for combining attributes based on ATAR Score and Broad Field Of Education (BFOE).
- **/Story3_GeographicalAnalysis**: Contains code in python to generate new data based on clustering and used Tableau to generate plots.
- **/Images**: Contains images used in assignment.
- **/Assignment3.twb**: Contains Tableau plots to generate LGA and SA3 heatmaps and scatter plots based on clustering.
- **/Assignment1**: Contains Assignment 1 works.
- **/Database**: Contains dataset for the Higher Education Attrition Rate of Australia.
- **/A3**: Contains report generated for this assignment.
- **/Extra**: Contains extra work that was not used in the assignment found irrevalent.

Note

Please note that since the dataset is small, it was not possible to run any models to derive various other inferences as outlined in Assignment 1. I did my best to generate visual analytics using the available data. Additionally, I experimented with several different plots that were not used in Assignment 1.

Workflow diagram

Please refer to Images folder, one can observe files named as workflow{number} named as workflow used for Story or task(number+1).

Dataset

Please refer to Assignment 1, as same dataset was used. In addition to this for Story 3, cluster data was added to datasets of LGA and SA3 files using K-Means Clustering in python.

Getting Started

Prerequisites

1. **Python Installation**: Make sure Python 3.12.0 or later is installed.

   ```
   python3 --version
   ```

   If not installed, download from [Python Downloads](#).

2. **pip Installation**: Ensure pip 24.3.1 or later is installed.

   ```
   pip3 --version
   ```

   If not installed, follow the [pip Installation Guide](#).

Installation of Required Packages

Navigate to the main project directory and run:

```
pip install pandas numpy matplotlib seaborn scikit-learn plotly
```

Note

Please keep the project structure the same as provided. If any files or folders are moved, you will need to update file paths in the code to reflect the new structure (if changed, one could change file path as mentioned in step 4 of running the code using .py scripts section).

Running the Code

Running `.py` scripts.

Run `.py` Scripts

1. **Navigate to the Folder**: Move to the folder containing the code for the desired visualization (e.g., `Story1_DiferentPlot/`).

2. **Run the Python Script**: Execute the main Python script (e.g., `sunburst.py`) to generate plots, which will be saved in the same folder.

   ```
   python3 sunburst.py
   ```

   Optional: Uncomment `plt.show()` in the script to display plots as they are generated.

   Modify the file path based on your system structure.

Conclusion

This repository provides a comprehensive suite of visualizations aimed at visual analytics.