ETL Project

Team Members:

Alexandra Taft
Anthony English
Arlette Varela
Nathan Bolt

Extract:Transform:Load - Automating Jobs Away

August 2020

OVERVIEW/OUTLINE

The datasets we will be using for our project combines automation probability data with a breakdown of the number of jobs by salary in each occupation by state (within the US). For the purposes of our project we will filtering down to the detailed level of class titles and assigning them an occupational group and then region based on US census categorizations. We will be loading into PgAdmin as a SQL database.

DATASET(S):

Kaggle Datasets: Occupation, Salary, and the Likelihood of Automationhttps://www.kaggle.com/andrewmvd/occupation-salary-and-likelihood-of-automation?select=automation_data_by_state.csv

https://www.kaggle.com/andrewmvd/occupation-salary-and-likelihood-of-automation?select=automation_data_by_state.csv

 Occupation, Salary and Likelihood of Automation • kaggle.com

---

TASK BREAKDOWN:

- [x] ~~Set up Github Repository @Arlette V~~
  - [x] ~~Each member create a branch~~
- [x] ~~Search for data @doc @~~
- [x] ~~Set up notebooks (Jupyter, VSC, etc.) @doc~~
- [x] ~~Create ERD~~
  - [x] ~~Export ERD code into PgAdmin~~
- [x] ~~Bring CSV files into Jupyter~~
  - [x] ~~Organize and clean dataframes, this will include:~~
  - [x] ~~automation~~
  - [x] ~~occupation_broad~~
  - [x] ~~occupation_detail~~
  - [x] ~~occupation_major~~
  - [x] ~~occupation_minor~~
- [x] ~~Set up PgAdmin~~
  - [x] ~~Create tables/schemas~~
  - [x] ~~Make sure tables are good in Jupyter~~