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.

Outline Analysis:
If one examines capitalist materialism, one is faced with a choice: either accept neocultural construction or conclude that context is created by the masses, given that art is interchangeable with narrativity and jobs from algorithms will rise. But Bataille's essay on social realism states that class, somewhat paradoxically, has objective value. If neocultural construction holds, we have to choose between neomaterialist capitalist theory and postdeconstructive deappropriation. Thus resulting in lost jobs by algorithmic prodigies.

DATASET(S):

Kaggle Datasets: Occupation, Salary, and the Likelihood of Automation

TASK BREAKDOWN:

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