ETL Project proposal

2020 appears to have far exceeded voting rates in recent US presidential elections. According to estimates on Wikipedia, 2020 voting among eligible population was between 66.4% and 72.1%. If correct, this would be the highest turnout % since the election of 1900 (re-election of William McKinley). At-least 22 million more votes were cast in 2020 vs. 2016 with an eligible population increase of only approximately 9 million.

My project collects election related data from various sources to build a database that will allow analysis, by state of changes in voter Eligible age population and voter turnout %.

2020 data will be collected via Kaggle or electionproject.com

2016 data will be collected via electionproject.com and census data tables.

If time allows- data for the 2018 midterm may be included.

Data will be extracted by csv/xlsx download or table web-scraping.

Data will be cleaned for each election, and a summary DataFrame with variances will be created with Python.

A final Database will be created with PostgreSQL