Term Project Milestone #5: Merging the Data and Storing in a Database/Visualizing Data

Michelle Helfman

Bellevue University

DSC540-T301 Data Preparation (2233-1)

Catherine Williams

June 1, 2023

Merging the Data and Storing in a Database/Visualizing Data

This project has exposed me to things my teammates use that I have wanted to try. This project also identified and corrected misconceptions regarding information obtained programmatically from the internet.

For example, I always thought websites have the same basic structure, comprising various blocks or sections in a cookie-cutter-like fashion. Thinking each website's uniqueness comes from the prose, pictures, tables, and other interactions contributing to the overall message. In my first choice of website for Regions in the United States, I found that while the website, from a visitor's perspective, looked like every other website with a chart or table, this website did not have an embedded table. After attempting unsuccessfully to extract this information for several days, I eventually switched to a combination of two websites with the same information but containing tables that resemble the assignment from the book. APIs, on the other hand, are more standardized on how data is retrieved and parsed. Everyone is issued an API key that controls what information is available, and data is retrieved using topic-based parameters. The API website provides documentation describing the parameters and coding examples. This process repeats itself on other websites using APIs.

Most of this project's information comes from Comma Separated Values (CSV) datasets downloaded from government websites. These datasets were loaded into a SQL Server database using SSMS's Bulk Loader, and a PL/SQL stored procedure was written to combine the columns for this project into one output file. Then, using Python code, the column headers names and data in the output file was standardized, the missing information was filled in from an additional dataset, and outliers were identified. The data from web scraping and the API websites were also standardized, and additional city and state information was included as key columns for

matching between the three datasets. Finally, Python code was written to create tables, upload data, and use the created keys to combine the three separate tables into one final output file.

The three SQL Server tables provide the information to Tableau for visualizations, and the tables were combined using the key columns. For readability, the data was displayed by Region and State. Percentages and averages were used to account for the differences in state populations. In addition, bar graphs, pie charts, and bubble charts convey population, race, income, education, and crime.

There were no ethical implications on where the data came from since this is fact-based information from free public websites, but the weather API has free and subscription-based keys. The free key limits information to 1000 retrievals per day. No other disclaimers or warnings appear on these websites, but the sites were cited in the Reference section below and the Term Project codebook. During the cleansing process, great care was taken to maintain the meaning and purpose of the information, particularly with numbers. The impact of an incorrect decimal place could change profits to losses or the dispersal of resources based on population. Data consistency throughout the process is vital to ensure everything is clear.

This project was challenging but fun. After deciding on my topic, the difficulty is always finding available information in sufficient quantities to satisfy the requirements. However, the exposure to Python, web-scraping, and APIs are invaluable, and I plan to use these new tools outside of class for my employer.

References

Demographics Table – Milestones 1 and 2

DP05: ACS DEMOGRAPHIC AND HOUSING ESTIMATES

U.S. Census Bureau, 2021 American Community Survey 1-Year Estimates Retrieved April 2, 2023

https://data.census.gov/table?q=United+States+demographics&g=010XX00US,$31000M1&tid=ACSDP1Y2021.DP05&tp=true

S1902: MEAN INCOME IN THE PAST 12 MONTHS (IN 2021 INFLATION-ADJUSTED DOLLARS) U.S. Census Bureau, 2021 American Community Survey 1-Year Estimates Retrieved April 2, 2023

https://data.census.gov/table?q=income&tid=ACSST1Y2021.S1902&moe=false&tp=true

S2301: EMPLOYMENT STATUS

U.S. Census Bureau, 2021 American Community Survey 1-Year Estimates Retrieved April 2, 2023

https://data.census.gov/table?q=employment+acs&g=010XX00US$3100000&tid=ACSST1Y2021.S2301&moe=false&tp=true

S2801: TYPES OF COMPUTERS AND INTERNET SUBSCRIPTIONS

U.S. Census Bureau, 2021 American Community Survey 1-Year Estimates Retrieved April 2, 2023

https://data.census.gov/table?q=internet&g=010XX00US$3100000&tid=ACSST1Y2021.S2801&moe=false&tp=true

City and Town Population Totals: 2010-2019

U.S. Census Bureau, 2021 Retrieved March 28, 2023

https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html#ds

Income Tax Rates By State

Tax-Rates.org — The 2022-2023 Tax Resource Retrieved March 31, 2023

https://www.tax-rates.org/taxtables/income-tax-by-state

United States Airports, by Name

Airport Codes Retrieved March 31, 2023

https://www.airportcodes.us/us-airports-by-name.htm

Offenses Known to Law Enforcement by State by City, 2019

FBI - Federal Bureau of Investigation, 2019 Crime in the United States Retrieved April 7, 2023

https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/tables/table-8/table-8.xls/view

Most Educated States

McCann, Adam (February 13, 2023) *2023's Most & Least Educated States in America*

https://wallethub.com/edu/e/most-educated-states/31075

Best & Worst States to Retire

McCann, Adam (January 23, 2023) *2023's Best States to Retire*

https://wallethub.com/edu/best-and-worst-states-to-retire/18592

Cross Reference Table – Joined together by FIPS Codes.

PRINCIPAL CITIES OF METROPOLITAN AND MICROPOLITAN STATISTICAL AREAS, MARCH 2020 – 1270 Rows

U.S. Census Bureau, 2021 Retrieved April 1, 2023

https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html

American National Standards Institute (ANSI) and Federal Information Processing Series (FIPS) Codes

U.S. Census Bureau, Retrieved April 7, 2023

https://www.census.gov/library/reference/code-lists/ansi.html#state

Regions Table – Milestone 3

List Of 50 States And Their Capitals

TheFactFile.Org Retrieved May, 2023

https://thefactfile.org/u-s-states-and-capitals

5 US Regions Map and Facts

Mappr.co Retrieved May, 2023

https://www.mappr.co/political-maps/us-regions-map/

Weather table – Milestone 4

Weather Data & API - Global Forecast & History Data

VisualCrossing Retrieved May 21, 2023

https://www.visualcrossing.com/