SAS Programing Project
Texas County Economic Indicators Compilation

This project involves the compilation of a group of county level economic indicators for Texas. The ultimate objective is to compile a panel style dataset of monthly unemployment rates, per capita sales tax collections, and per capita alcohol sales for the period January 2017 – August 2022 for the ten largest Texas counties by population.

You are provided the following input datasets:

Texas_County_Unemployment.csv

A CSV (comma separated values) formatted file with monthly unemployment rates over the period January 1990 – August 2022. Variables and layout appear as follows:

	geoname	date	unemployment_rate
2	Anderson County, TX	1/1/1990	7.2
3	Anderson County, TX	2/1/1990	7.2
4	Anderson County, TX	3/1/1990	6.6
5	Anderson County, TX	4/1/1990	6.4
6	Anderson County, TX	5/1/1990	8.2

Texas_County_SalesTax_Receipts.csv

A CSV (comma separated values) formatted file with monthly sales tax receipts over the period January 2012 – October 2022. Variables and layout appear as follows:

	County	date	sales_tax_receipts
2	Anderson	1/1/2012	466,440.23
3	Anderson	2/1/2012	694,920.67
4	Anderson	3/1/2012	436,622.94
5	Anderson	4/1/2012	475,761.17
6	Anderson	5/1/2012	647,752.86

mix_beverage_receipts.sas7bdat

Establishment level data on mixed beverage alcohol sales in SAS dataset format. See the file "Mixed_Beverage_Layout.pdf" for variable names. You will need to aggregate these data to a county-month level. The key variables you will need are as follows:

Location_County: A numeric county code variable (note that this variable is different than the U.S. Census Bureau geographic FIPs codes). See the description of the "County Codes" Excel file below.

Obligation_End_Date: Numeric SAS date variable formatted as MM/DD/YYYY. This variable reflects the last day of each month for which receipts are reported. For example, 07/31/2019 reflects receipts for the month of July 2019.

Total_Receipts: Total dollar receipts reported for each month for each reporting establishment.

County Codes.xlsx
Excel workbook format. Provides a bridge to the "Location_County" variable in the mix_beverages_receipts dataset and county names. Variables and layout as follows:

	County	Code
2	ANDERSON	1
3	ANDREWS	2
4	ANGELINA	3
5	ARANSAS	4
6	ARCHER	5

Texas_County_Population.csv

A CSV (comma separated values) formatted file with annual county population estimates over the period 2014 –2021. Variables and layout appear as follows:

	ctyname	year	TOT_POP
2	Anderson County	2014	57829
3	Anderson County	2015	57639
4	Anderson County	2016	57526
5	Anderson County	2017	58175

Us cpi.xlsx

Excel workbook with the monthly U.S. Consumer Price Index (2020=100) over the period January 2012 – September 2022. Variables and layout appear as follows:

	A	B	C	D
1	Series_Id	Year	Period	CPI_2020
2	CUUR0000SA0	2012	M01	87.579
3	CUUR0000SA0	2012	M02	87.965
4	CUUR0000SA0	2012	M03	88.633
5	CUUR0000SA0	2012	M04	88.901
6	CUUR0000SA0	2012	M05	88.796
7	CUUR0000SA0	2012	M06	88.666
8	CUUR0000SA0	2012	M07	88.522
9	CUUR0000SA0	2012	M08	89.014

As noted above, the ultimate objective is to compile of panel style dataset of monthly unemployment rates, per capita sales tax collections, and per capita alcohol sales for the period January 2017 – August 2022 for the ten largest Texas counties by population. The ten counties of interest (by order of 2021 population) are as follows:

Harris	Dallas	Tarrant	Bexar	Travis
Collin	Denton	Hidalgo	El Paso	Fort Bend

You are asked to convert the Sales Tax and Alcohol to a real per capita basis using the annual population data and U.S. CPI data to adjust for inflation. For a 2022 population estimate, you are asked to calculate each county's Average Compound Annual Growth Rate in population over the 2014 – 2021 period and apply that growth rate to 2021 population levels.

$$ACGR = \left(\frac{Pop_{2021}}{Pop_{2014}}\right)^{\left(\frac{1}{7}\right)}$$

$$Pop_{2022} = ACGR \times Pop_{2021}$$

Note: Use the integer portion of the second calculation for POP_{2020}.

For auditing purposes, for any intermediate steps that require merging datasets using more than a single matching variable (more than one BY variable) you are asked to create a single combined matching variable.

Your final combined dataset should reflect monthly data for the 10 counties noted above for the 68 months covering January 2017 – August 2022 (a total of 680 observations). The dataset should include the following variables:
Date
County Name
Unemployment Rate
Real Per Capita Sales Tax Receipts
Real Per Capita Alcohol Sales

All your analysis and data compilation should be completed in SAS. Turn in a copy of your final SAS program in Canvas by midnight on Friday, October 28th.

Note: To read the CSV files into SAS, I suggest you use PROC IMPORT as discussed in the Lecture 7 slide deck so that you leave an "audit trail."
To read the Excel workbook data you may either use a LIBNAME link to the workbooks or use PROC IMPORT.