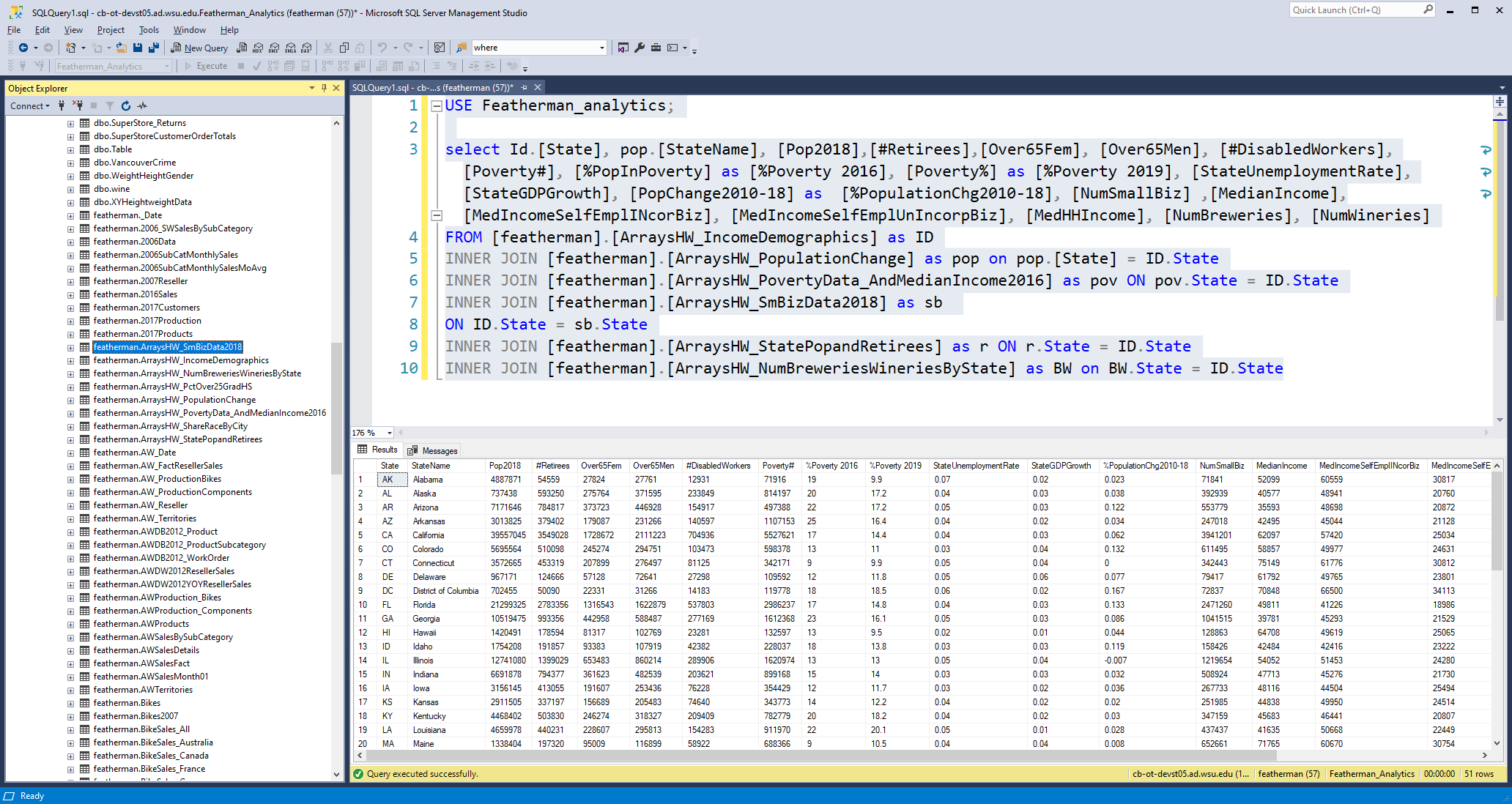
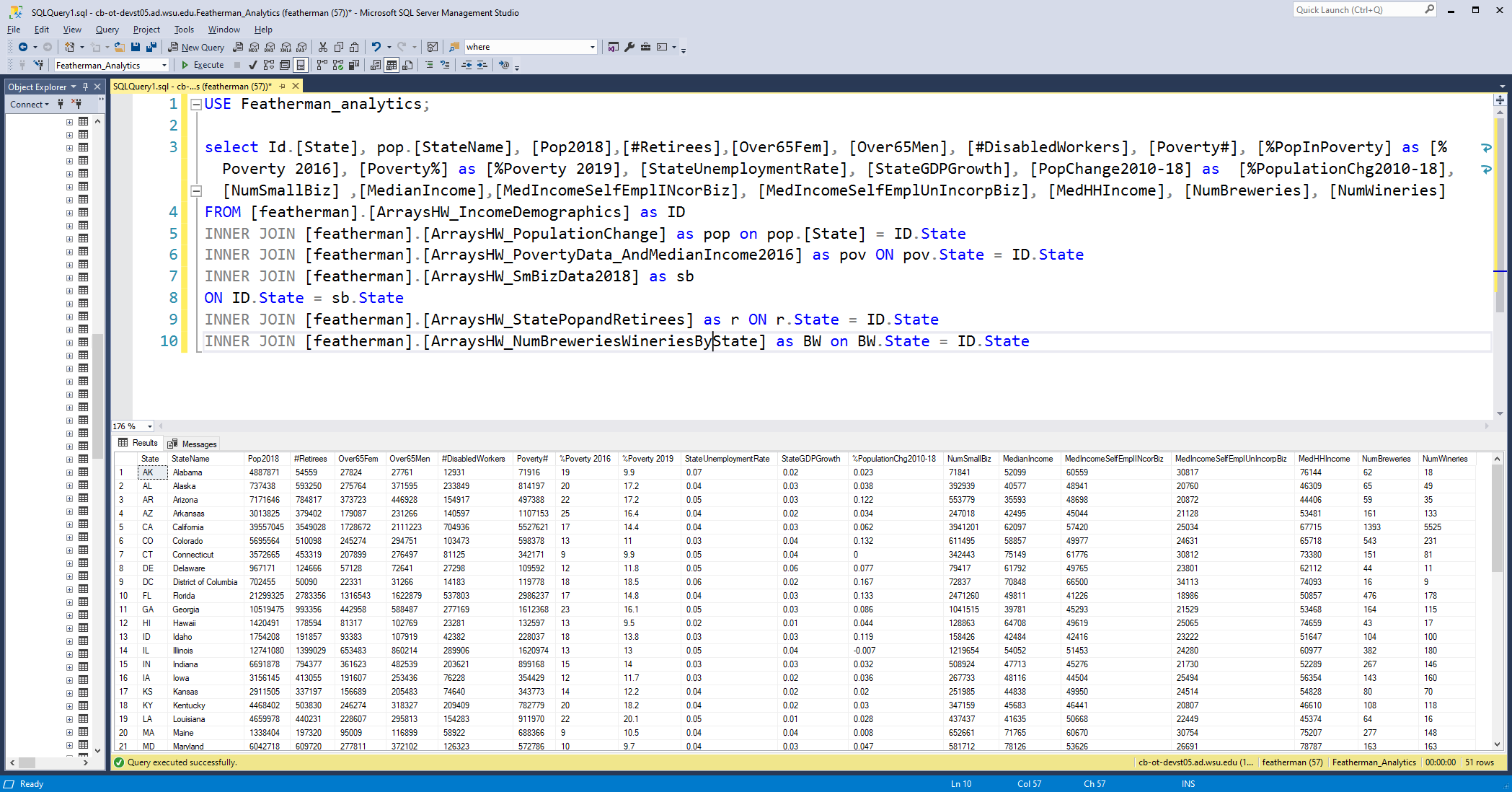
**Featherman’s Analytics Adventures© – Array Table Assignment**

****This assignment is designed to give you insight about the United States of America and practice using INSERT INTO SELECT FROM SQL queries, and also UPDATE SET() and CASE() processing. Data manipulation is made far easier when the data is loaded into an array. This assignment gives you the opportunity to gain experience manipulating data in arrays. The UPDATE SET() command is very useful in this context to build new columns of metrics.  
  
The scenario of this assignment is that you are considering moving to take advantage of a hot IT job market. But what states are livable according to your interests and needs? You decided that your town is great but because the real estate is too overpriced, that you feel that you want to start your career in a less expensive state. You decide to return to the Pacific Northwest only after building considerable wealth (five year’s equity in a duplex that can be leveraged to buy a starter home, and a year’s worth of expenses in the rainy-day fund.)

But how to make the decision of what state(s) are acceptable to move to? You recall your analytics professor suggested to use both system 1 (emotional) and system 2 (rational) and so you decide to pull data together to analyze different states and regions. Due to the importance of the decision, you do not want to rely on emotion and bias alone. As a result, you decided to look for datasets that are indicators of quality of life, and criteria that help decide if the state is going to be a good place to live. While many more and different datasets could be captured, (e.g. food production and prices, level of commerce, trade in goods and services, global connectedness, agriculture production, level of property and retail taxes, etc.), you feel the eight tables in the above picture are a good start to assess the quality of life in the state. You realize these tables and any analysis derived from them is just a first version of continued research, but you also surmise that starting with eight datasets is a good enough start for this week. Here is a partial dataset that can be created using these tables. You may find and merge more datasets as long as they are grouped by state.

The tables above were imported into the Featherman\_analytics database on cb-ot-devst05.ad.wsu.edu (or devst06). A listing of the columns of these tables and their column meanings is in the appendix. You can pull other perhaps more relevant data into your array from data.gov ([many datasets use a state level of granularity](https://catalog.data.gov/dataset?q=state+level&organization_type=State+Government&sort=score+desc%2C+name+asc)).

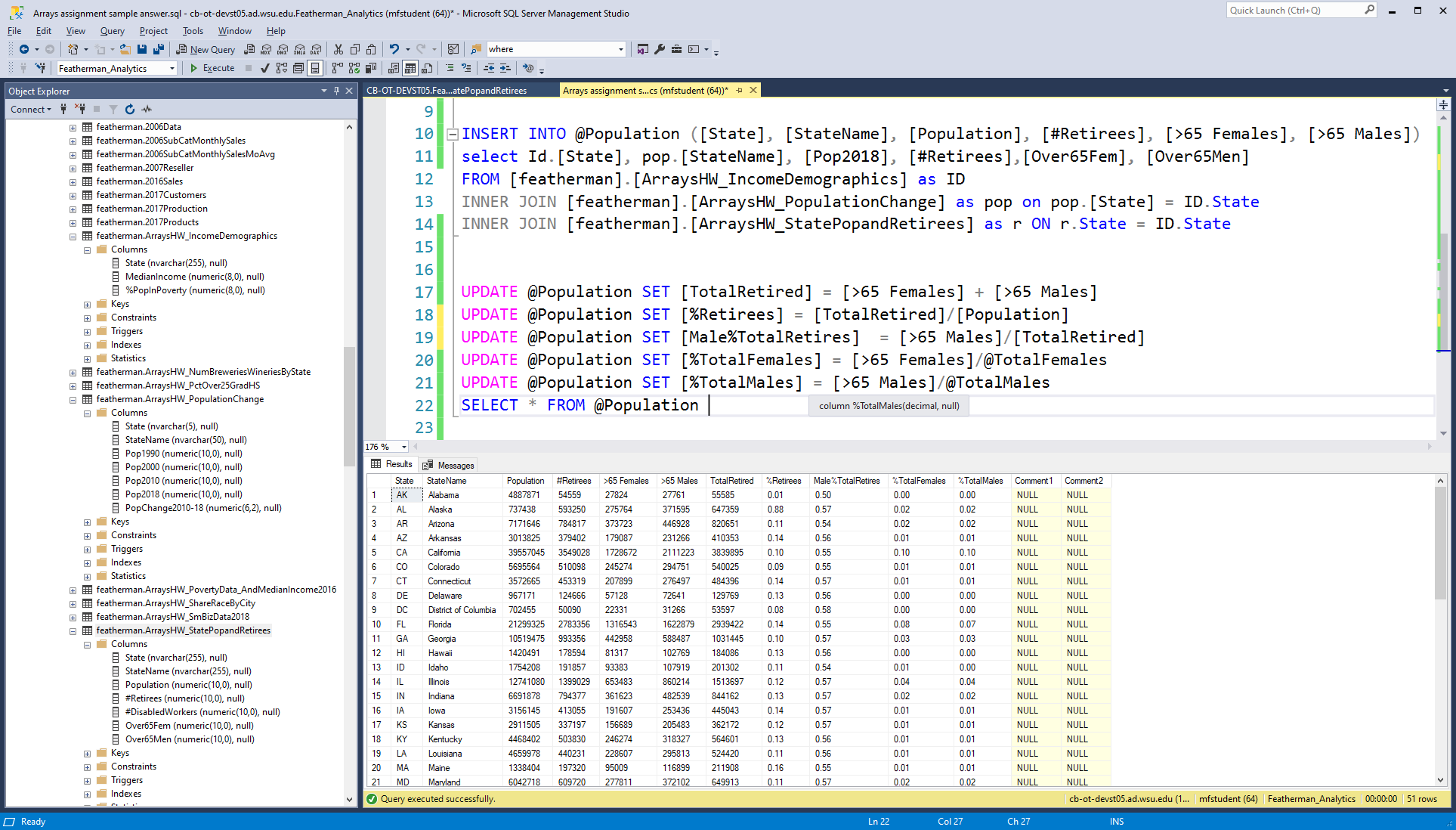
**Background**

You decide to do an analysis of the states for potential employment and domicile. You decide to build your analysis on several decision criteria, therefore you break the analysis into different sections. In the end you will rank the top 5 states using some the decision criteria in the next section. A future extension could introduce more data such as housing prices, housing affordability (as a percentage of salary), salary levels by job category, and other ratings based on intended lifestyle, etc. Again, you realize this analysis is just a start, version one, so 8 tables of demographics is a good start.

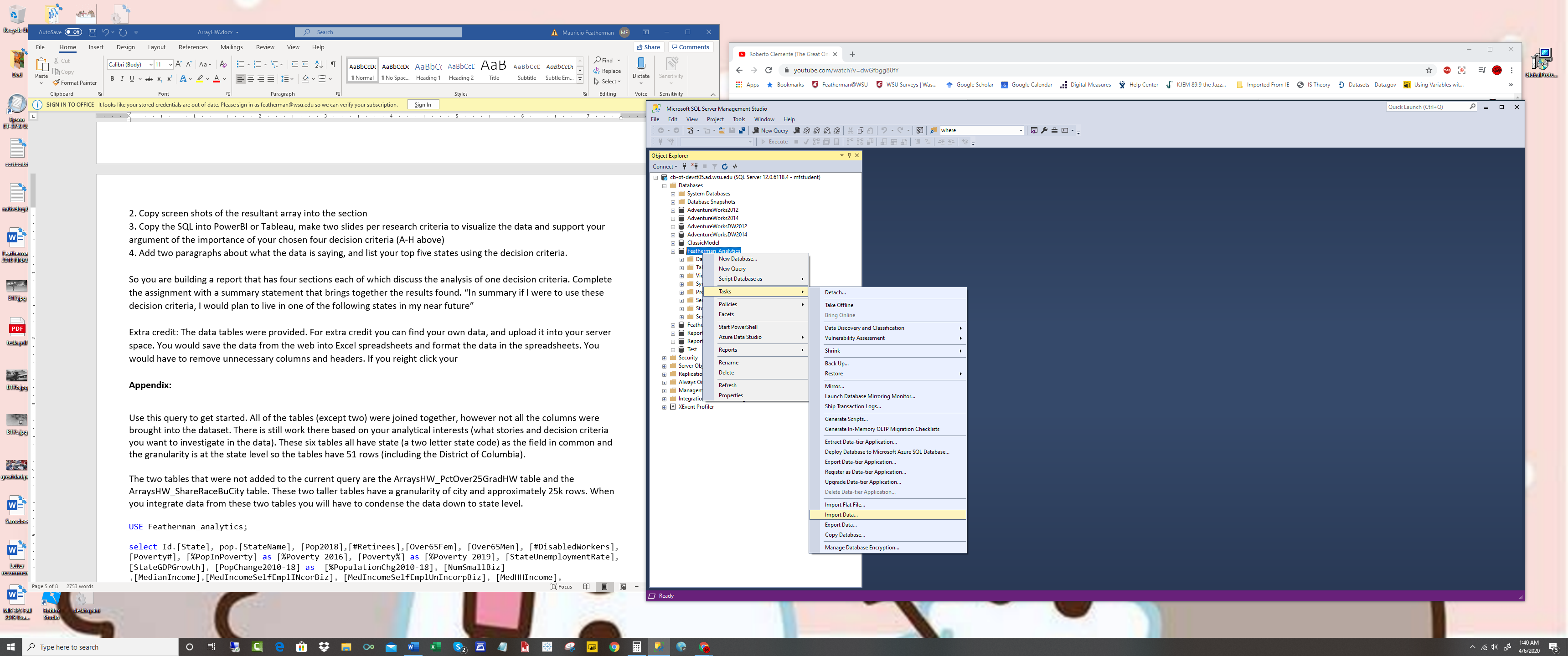
**Suggested Decision Criteria**

Below is a list of decision criteria believed to be evidenced in the first set of data tables. Were this an actual decision you would create your own decision criteria and then go find the data and merge it together. In this assignment some introductory has already been provided to speed up the process.   
  
a) population and level of population change (you want to live in a growing state, not declining as you think a state with increasing population correlated to higher real estate appreciation).  
b) level of poverty, high school graduation rate, and unemployment in the state and whether that is improving (you do not want to live in a state where prosperity is hard to come by, as you think poverty levels are correlated to higher crime).  
c) the level of retires in the state (you figure states with higher levels of retirees must have strong medical and abundant services for senior citizens, therefore the state can help to provide a higher standard of living in retirement).  
d) the GDP growth of the state (as a measure of prosperity and growth) and number of new jobs  
e) small business growth and importance of small business as a engine of economic growth in the state. The number of small businesses can be analyzed and percentage of employees in several industries that are employed by small businesses (% of exporters, % employment by small business, % manufacturing, %mining and %professional services employed by small businesses in the state). You know that the strength of America is a vibrant small business sector so you decide to factor this into your decision of what state to live in.  
f) You are interested in median income as this is a measure pf prosperity and you want to live in a state that is prospering and growing and becoming more wealthy, not in decline.  
g) you are interested in beer and wine and want to make sure there are breweries and wineries available to visit.  
h) Finally you are interested in the racial makeup of the state (black, white, native America, asian, hispanic).

Not all the fields to perform the analysis mentioned above were added to the query in the appendix. For example, many columns useful to analyze the importance and growth of small businesses were not included. Using the base query provided and changes you want to make, bring any columns of data into arrays for analysis.

**Requirements**Using the tables shown pull columns into arrays to analyze the state data. Choose your decision criteria from the list above or scrutinize the tables of data and decide which measures and categorization are useful. Build one array of columns to analyze the states for your chosen decision criteria. Your job is to categorize and rate the state using each of decision criteria that you choose. In the array create calculated columns of metrics that pull data together from the 8 tables to analyze the states on the decision criteria. You can for example add a textual column to the array that puts the state into categories based on GDP growth (e.g., stagnant state, slow-growth state, moderate growth state, fast-growing state).  
  
You are required to make a report with at least 4 sections (each analyzing one decision criteria). In addition to pulling the relevant columns of data together for analysis, your query needs to have three new metrics of your own design (numeric) and two textual metrics (using CASE() statements). Your metrics are graded for usefulness to analyze the decision criteria. Use the functionality shown in the provided arrays documentation and training videos. Add two paragraphs of textual interpretation and analysis for each of the decision criteria. Mention the top 5 states in desirability in each section.   
  
You will be creating the array, selecting values into it, maybe creating local variables (for example creating a percent of total calculation and others) and loading them with values, and adding columns of metrics to the array to analyze . Use an UPDATE SET command (and perhaps UPDATE JOIN) to calculate values for the new columns of data. You are actually using only a small percentage of the functionality presented in the arrays module as most of the data is at the same level of granularity, and already organized for you.  
  
Most of the tables are at the state level of granularity but two of the tables are at the level of city (the ArraysHW\_PctOver25GradHW, and ArraysHW\_ShareRaceBuCity tables are at the city level). When you bring in data from these two tables (required to use at least one of them) you will have to change the granularity of the data to match the state level. For two of your decision criteria incorporate the city data.

Before starting this assignment, be sure to set your foundational knowledge and experience with using arrays. Please run the queries from the arrays module that are needed to complete this assignment. Set your knowledge before you begin by running the related examples so that the assignment completion process is orderly. Do not jump into an assignment without creating your own arrays for practice.   
  
The arrays module code that you should practice with include:  
  
a) the DECLARE ing of the array and understanding of the data types needed (e.g. any column that is to display a percent should be declared DECIMAL(5,2) to allow display of the values, or nvarchar(6) to display percent signs.  
b) The INSERT INTO @array columns SELECT FROM queries to load an initial set of columns from the arrays  
c) The UPDATE SET commands to calculate values for columns in your array.  
d) PRINT functions to see calculated values   
e) CASE statements to write textual analytics  
f) the Tableau and PowerBI videos in modules 8 and 9  
g) other techniques as envisioned and performed

**Turn-in**Build a MS-Word document that   
a) Lists your chosen decision criteria for making the decision which state to move to. Write two sentences about the importance or value of each.  
b) Choose your decision criteria and build metrics and reports/charts for each. Use headings to organize your responses such as Decision Criteria 1: “title of decision criteria” Have at least 4 decision criteria so four sub-sections are needed.  
  
***For each of the 4 sub-sections:***  
1. Create the query section.  
2. Copy screen shots of the resultant array into the section  
3. Copy the SQL into PowerBI or Tableau, make two visualizations per research criteria to present the data and support your argument of the importance of your chosen decision criteria.  
4. Add two paragraphs about what the data is saying, and list your top five states using the decision criteria.  
  
So you are building a report that has four sections each of which discuss the analysis of one decision criteria. Complete the assignment with a summary statement that brings together the results found. “In summary if I were to use these decision criteria, I would plan to live in this short list of states.”   
  
Extra credit: The data tables were provided. For extra credit you can find your own data, and upload it into your server space. You would save the data from the web into Excel spreadsheets and format the data in the spreadsheets. You would have to remove unnecessary columns and headers. If you right-click your database you can select Tasks| Import data. You can watch a YouTube video on importing excel data in to SSMS if you get stuck.

**Try to have some fun with this assignment!**

**Appendix:**

The six tables all have state (a two letter state code) as the field in common and the granularity is at the state level so the tables have 51 rows (including the District of Columbia). Two tables (ArraysHW\_PctOver25GradHW table and the ArraysHW\_ShareRaceBuCity) are taller tables (approximately 25k rows) with a granularity of city. When you integrate data from these two tables you will have to condense the data down to state level.

|  |  |  |
| --- | --- | --- |
| Table Name | Column Name | Column name meaning |
| SmBizData2018 | State | State code 2 letters |
|  | StateName | Name of State |
|  | NumSmallBiz | Number small businesses in state |
|  | PercentEmpbySmBiz | Percent Employees in state working in small businesses |
|  | MedIncomeSelfEmplINcorBiz | Median income of self-employed individuals in *Incorporated* businesses |
|  | MedIncomeSelfEmplUnIncorpBiz | Median income of self-employed individuals in *UNIncorporated* businesses |
|  | NumNewBiz2018 | Number new businesses created in 2018 |
|  | NumNewJobsFromNewBiz2018 | Number new jobs from new businesses in 2018 |
|  | PctOfExportersThatAreSmBiz | Percent of Exporters that are small businesses |
|  | StateGDPGrowth | State GDP growth |
|  | StateUnemploymentRate | State Unemployment Rate |
|  | PctMfrEmpFromSmBiz | Percent of *manufacturing employees* are that in small businesses |
|  | PctProfessionalSvcEmpfromSMBiz | Percent of *professional service* employees are that in small businesses |
|  | PctMiningEmpfromSMBiz | Percent of *mining* employees are that in small businesses |
|  |  |  |
| Table Name | Column Name | Column name meaning |
| Income Demographics | State | State code 2 letters |
|  | MedianIncome | Median Income for residents of the state |
|  | %PopInPoverty | Percentage of the population of the state that is in poverty |
|  |  |  |
| Table Name | Column Name | Column name meaning |
| NumBreweries WineriesByState | State | State code 2 letters |
|  | NumBreweries | Number of breweries in the state |
|  | NumWineries | Number of wineries in the state |
|  |  |  |
| Table Name | Column Name | Column name meaning |
| PercentOver25 Completed HighSchool XL. This dataset is by city so you will have to group the data by state | State | State code 2 letters |
| City | City |
| PercentCompetedHS | Percent of state’s population that completed high school |
| Table Name | Column Name | Column name meaning |
| State – Poverty and Median Income 2016 | State | State code 2 letters |
|  | StateName | Name of the state |
|  | Poverty# | Number of people in state living in poverty |
|  | Poverty% | Percent of people in state living in poverty |
|  | MedHHIncome | Median household income |
|  |  |  |
| Table Name | Column Name | Column name meaning |
| Population Change | State | State code 2 letters |
|  | StateName | Name of the state |
|  | Pop1990 | State population in 1990 |
|  | Pop2000 | State population in 2000 |
|  | Pop2010 | State population in 2010 |
|  | Pop2018 | State population in 2018 |
|  | PopChange2010-18 | Percentage population changed from 2010 to 2018 |
|  |  |  |
| Table Name | Column Name | Column name meaning |
| Share Race By City  The data is by city so you have to aggregate it by state | State | State code 2 letters |
| City | City |
| %white | Percent of population that is white |
| %black | Percent of population that is black |
| %native\_american | Percent of population that is Native Am. |
| %asian | Percent of population that is asian |
| %hispanic | Percent of population that is hispanic |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Table Name | Column Name | Column name meaning |
| StatePop andRetireees | State | State code 2 letters |
|  | StateName | Name of State |
|  | Population | State population |
|  | #Retirees | Number of retired people in state |
|  | #DisabledWorkers | Number of people collecting disability in state |
|  | Over65Fem | Number females in state that are over 65 |
|  | Over65Men | Number males in state that are over 65 |
|  |  |  |
|  |  |  |