**Hi team, #TeamLogic**

Below is a brief outline of what I have found and how I think we can build/modify etc on it. Please note, it is slightly different from the initial proposal because I came across rich data that we can play around with in many ways.

1. **.csv Datasets found (By state)**

***For now the csv’s are in this google drive***

1. SAINC1 Personal Income Summary Personal Income, Population, Per Capita Personal Income (1929 to 2019)
2. Personal consumption expenditures per capita (1997 to 2019)

Source: <https://www.bea.gov/>

1. **API’s,**
2. City Month average weather information

Source: <https://openweathermap.org/api/statistics-api#month>

1. **Other possible datasets to consider (not yet found)**
2. Crime rates per state
3. Car crashes per state vs historical weather conditions
4. Demographics vs real estate ownership per state
5. Number of universities in a state
6. Alcohol abuse incidents per state
7. The list can be endless

**Methodology/Questions (based on 1 & 2)**

1. Compare two states and analyze which state has a better economy, better income per capita, and lowest expenditure per capita in 2019 using a bar chart or alternative
2. Compare income and expenditure for each state in the last 2 decades using a line graph
3. Use statistical methods to find the outlier years of economic activity since 1997. (I am curious if using the boxplot will pick out the dotcom era, and the 2008 recession, and who knows other outstanding years?)
4. Determine the correlation of Clothing and footwear expenditure vs Food and beverages purchased for off-premises consumption using regression methods and scatter plotting.
5. Use a line plot to show the average monthly temperatures of 2 cities using APIs to create dataset.

NB: Once our code is running smoothly, we can investigate a number of combinations of variables and different state pairs to tell an interesting story in our presentation.

**Technical areas that will be displayed**

1. Cleaning datasets and creating dataframes for a period of interest
2. Merging the datasets on state
3. Possible grouping/cutting if we choose to analyze our data per certain periods e.g. per decade or west coast vs east coast states etc
4. Statistical methods including regression
5. Plotting i.e. boxplot, bar plot, line plot
6. If we have time as a bonus we can code for user input for 2 states and 2 cities in a way that does not break our code if the user inputs a wrong state or city etc.

My advise is that we do the basics first, enough for a good submission, then build on our code thereafter as we have fun.