Proposal 8 : California employment and wage analysis in the past decade

Problem:

Analyze and visualize the employment and wage in the last decade (2011-2021) in California.

Dataset:

Quarterly Census of Employment and Wages (QCEW)

(https://catalog.data.gov/dataset/quarterly-census-of-employment-and-wages-qcew)

The original dataset is formatted in a CSV file, which contains information about California geographic areas, timestamp, ownership, industry codes, employment times, wages, etc. from 2004 to 2021 (17 years.

Reference: Industry Code (https://www.census.gov/naics/?58967?yearbck=2022)

Proposed Solution and Real World Application :

Our proposed solution is to visualize the several key attributes of employment and build statistics models to analyze their relations with each other.

This solution would have a real world application, because it would generate insight to the change of employment over time, and would also identify the correlation between geometrics, industry, wages, and etc.

Statistical Modeling proposed:

The kind of statistical relationships we want to explore include correlations of employment trends between different industries and different sectors, comparison of the employment trends of an industry between different years etc. We also propose to use the time series data of some industries with adequate data points (for example: industry with NAICS 92212) for employment numbers along the years for time series forecast prediction using a simple RNN, LSTM or a statistical model.

Project steps

Step	Estimated completion time	Person(s) in charge (among the group of 5)

1. Extracting and cleaning up data	One week	Zhiyan
2. Data visualization / graphics code	One week	Yusuf, Siddharth
3. Python Models - i.e. time series	Two weeks	Xi Yang, Siddharth
4. Code reviews / organization	One week	Yusuf, Xi Yang, Siddharth, Zhiyan, Yuhao
5. Presentation/report	One week	Yusuf, Xi Yang, Siddharth, Zhiyan, Yuhao