About the dataset

Job patterns for minorities and women in private industry (EEO-1: 2018)

As part of mandate under Civil Rights Act of 1964, the Equal Employment Opportunity Commission requires periodic reports from public/private employers, and unions and labor organizations which indicate the composition of their work forces by gender/race/ethnicity.

As a result, data is collected annually from private employers with 100 or more employees or federal contractors with 50 or more employees. This dataset represents 73,400 employers representing 56.1 million employees whose identities are masked by aggregation.

Below are the important columns that were used to perform this analysis –

1. Region - indicating regional composition of workforces – Midwest, NorthEast, South, West
2. States – indication state-wise composition of workforces – All 50 states
3. North American Industry Classification System (NAICS) – including 20 different types of profession

For the below use cases, we performed the analysis in comparison to -

1. Race – White, Asian, Black or African American, Biracial, Native Hawaiian or other pacific Islander, Hispanic or American Indian or Alaska Native
2. Gender – Male, Female
3. Job Profile – Labors, Service, Clerical, Tecnicians, Sales Workers, Mid Off and Managers, Operatives, professionals, Senior Off and Managers, Craft

According to each use case, different subset of data had to be considered to create different views. There was a sub-routine written to create a map and automate this process to smartly pick the right columns as per the requirement.

Inferences –

1. The data is largely collected for Health Care and Social Assistance, Retail Trade, Manufacturing and Professionals (in that order) so it’s a good metric to get a consolidated view of the situation of diversity in 2018.
2. In terms of variety in industries, United States is one of the most diverse countries in the world with every form of occupation contributing towards the count of workforce.
3. Race –
4. Whites have consistently dominated in each industry, region and state. This should come as no surprise as white American constitute x% of the population in 2018.

Structure –

1. Introduction
2. About the Dataset
3. Tech Stack
4. What have I done?
5. Insights
6. Future Scope