CSC343 - Project Phase 1

Xiaochun Tong, Farzaneh Tabandeh
Sept 2021

1.Project Domain

This project will investigate the relation between the Gross Domestic Product (GDP), manufacturing enterprise size and employment rate by age and education of world's countries. We will try to find how the countries' employment rate and enterprise size affects the GDP of the countries. We also explore how different age and education level groups constitute the total employment rate of the counties. The utilized datasets will consist of the data points of a certain time period between 2016-2019.

2. Datasets

We will use the following data sets:

Countries Gross Domestic Product (GDP)
Countries employment rate
Employment based on age group
Employment based on education level
Number of enterprises of specific size

In order to work with these datasets in a meaningful way, we should learn about the precise definitions of the indicators that they measure. In addition, we will need to remove some non-relevant attributes from these datasets. Some of these datasets will need to be merged into one dataset, this is because we want to consider a certain time interval of years for the data. For example the data about the employment by the age group provides one dataset per age group, which we will merge them to all be in one single dataset.

3. Questions

We will investigate the following questions over the years 2016-2019:

1) How does age affect the employment rate? (i.e. Which country has the greatest proportion of old people in its employment rate?)

2) How does education affect employment and in turn the countries' GDP? (i.e. Do more educated people comprise the greatest proportion of employed people?)

3) Does the country with the greatest number of big sized enterprises also have a greater number of employment with higher education? (This question is influenced by thinking that more educated people can form more successful bigger sized companies, but does the data show that this thinking is true?)

4. Schema

4.1. Relational schema

We will have 5 relations as follows:

Countries(countryCode, countryName, year, gdp, gdpPerCapita)
Employment(countryCode, year, employmentRate)
EmploymentByAge(countryCode, year, ageGroup, employmentByAgeRate)
EmploymentByEducation(countryCode, year, educationLevel, employmentByEducation)
EmploymentByEnterprise(countryCode, year, enterpriseSize, numEnterprises)

Employment[countryCode] ⊆ Countries[countryCode]
EmploymentByAge[countryCode] ⊆ Countries[countryCode]
EmploymentByEducation[countryCode] ⊆ Countries[countryCode]
EmploymentByEnterprise[countryCode] ⊆ Countries[countryCode]

4.2. Data dictionary

Countries

Attribute	Description	Type	Required	Default
countryCode	The code of the country.	TEXT	yes	
countryName	The name of the country.	TEXT	yes	
year	The year in which the gdp and gdpPerCapita rates were measured.	TEXT	yes	
gdp	Yearly income obtained through the production of goods and services in a country.	INT	yes	
gdpPerCapita	gdp divided by the country's population.	DOUBLE	yes	

Employment

Attribute	Description	Type	Required	Default
countryCode	The code of the country.	TEXT	yes	
year	The year in which the employment rate was measured.	TEXT	yes	
employmentRate	The ratio of the number of people who are employed and have a job with respect to the total number of available labour resources in a country.	DOUBLE	yes	

EmploymentByAge

Attribute	Description	Type	Required	Default
countryCode	The code of the country.	TEXT	yes	
year	The year in which the employment by age rate was measured.	TEXT	yes	
ageGroup	The age group that this rate falls into. The age groups will consist of the people aged "**15_24"** (those just entering the labour market following education); people aged **"25_54"** (those in their prime working lives); people aged **"55_64"** (those passing the peak of their career and approaching retirement).	TEXT	yes	
employmentByAgeRate	The measure of the number of employed people of a given age as a percentage of the total number of people in that same age group.	DOUBLE	yes	

EmploymentByEducation

Attribute	Description	Type	Required	Default
countryCode	The code of the country.	TEXT	yes	
year	The year in which the employment by education rate was measured.	TEXT	yes	
educationLevel	The education level for which the employment rate has been calculated, which can be any of the **“below upper secondary”**, **“upper secondary non-tertiary”**, or **“tertiary”**.	TEXT	yes	
employmentByEducation	The percentage of the number of people with a certain education level in the total population in working age.	DOUBLE	yes	

EmploymentByEnterprise

Attribute	Description	Type	Required	Default
countryCode	The code of the country.	TEXT	yes	
year	The year in which the number of enterprises with different sizes was measured.	TEXT	yes	
enterpriseSize	The size of an enterprise which can be any of the following: Enterprises with fewer than 10 employees which is shown by **“1_9_EMPLOYED”**. Enterprises with 10 to 49 employees which is shown by **“10_49_EMPLOYED”**. Enterprises with 50 to 249 employees which is shown by **“50_249_EMPLOYED”**. Enterprises with more than 250 employees which is shown by **“250More_EMPLOYED”**.	TEXT	yes	

numEnterprises	The number of enterprises with a specific size in this country.	INT	yes	

4.3. Justification of design

We decided to have different relations that look into different aspects of employment and its effects over the years 2016-2019. The “Countries” relation will hold the info about one the most direct outcome of the employment rate which is the gdp and gdpPerCapita. The “EmploymentByAge” and “EmploymentByeducation” will have the data about how employment consists of different age and education groups and how those rates will affect the total employment in the “Employment” retaliation. Lastly, “EmploymentByEnterprise” will enable us to see the potential effects of education and age on the number of enterprises with different sizes, which in turn also influences the gdp.