UNIVERSITY OFWINDSOR

Masters in Applied Computing (MAC)

COMP 8157-Unit1

Advanced Database Topics – Part II

ASSIGNEMENT - I

Submitted by:

Keerthi Supriya Ravi

(1051418425)

Analysis of World’s Educational Attainment for Population Aged 25 and over

By

Keerthi Supriya Ravi

INTRODUCTON:

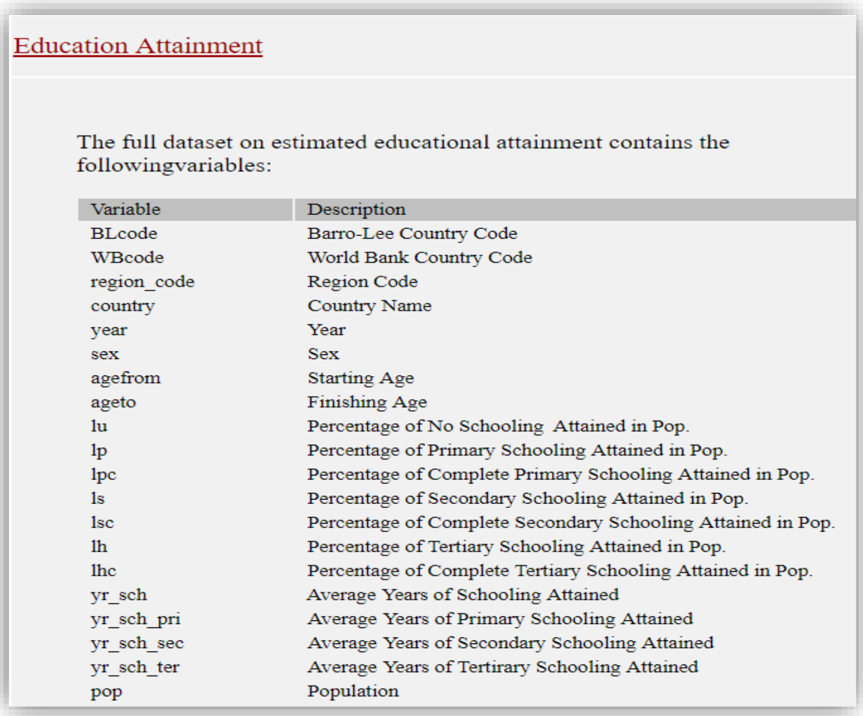
With the increase in population, there are many individuals being educated across the world. Furthermore, the people studying can be comprised of different Age groups such as 15-60 years, 0-16 years etc. The problem is to identify the following questions among these people across the world:

- how many of the people in the current population are educated or studying?
- What age groups does these people belong to?

The dataset I have selected will provide answers to such questions easily on analyzing it by using various visualization tools like RapidMiner, Tableau etc. I have selected RapidMiner to analyze and perform visualizations to extract meaningful insights from the data.

ABOUT THE DATASET:

The Educational Attainment Dataset used in this report contains 2582 rows and 20 columns. This can be downloaded from the link (**http://www.barrolee.com/Lee_Lee_LRdata_dn.htm**) and the filename is “**OUP_long_MF2564_v1.csv**”.The following picture shows the detailed column names used in the dataset.



Education Attainment

The full dataset on estimated educational attainment contains the followingvariables:

Variable	Description
BLcode	Barro-Lee Country Code
WBcode	World Bank Country Code
region_code	Region Code
country	Country Name
year	Year
sex	Sex
agefrom	Starting Age
ageto	Finishing Age
lu	Percentage of No Schooling Attained in Pop.
lp	Percentage of Primary Schooling Attained in Pop.
lpc	Percentage of Complete Primary Schooling Attained in Pop.
ls	Percentage of Secondary Schooling Attained in Pop.
lsc	Percentage of Complete Secondary Schooling Attained in Pop.
lh	Percentage of Tertiary Schooling Attained in Pop.
lhc	Percentage of Complete Tertiary Schooling Attained in Pop.
yr_sch	Average Years of Schooling Attained
yr_sch_pri	Average Years of Primary Schooling Attained
yr_sch_sec	Average Years of Secondary Schooling Attained
yr_sch_ter	Average Years of Tertirary Schooling Attained
pop	Population

PROCESS STEPS TO ANALYZE THE DATA USING RAPIDMINER TOOL

- ✓ A Repository is first created with two sub-folders namely Data and Resources.
- ✓ The Dataset is imported into the workspace using the “Import Data” option.
- ✓ In the Design View, the operators like Retrieve, Store and Filter are used to load and remove the redundant and missing values.
- ✓ The connections to the input and output are made to retrieve the results**.**
- ✓ In the “Turbo Prep” view, we can select load data to retrieve the dataset and click on pivot to select the columns with or without aggregators along with the ‘GroupBy’ clause.
- ✓ After selecting the desired columns for analysis, click on “Commit Pivot” and the table with the values corresponding to our query is generated.
- ✓ Click on “Charts” to perform visualizations using various plots.

RESEARCH QUESTIONS:

After examining the dataset, I have framed some questions and performed the analysis with the help of the tool to identify the facts in the dataset. Some of the questions I have identified are discussed below:

QUESTION 1:

To analyze the Average percentage of complete tertiary schooling attained in the Population?

SOLUTION:

In this question, I tried to find out the mean percentage of the population who showed interest in completing their tertiary schooling education. I have selected this question because from the results we are able to determine the following:

- The values of people’s interest belonging to this age group motivated to pursue tertiary education.
- The countries that has reported the highest and lowest values of enrolment.

By Examining the plot below, I found out that the USA has reported the highest value as 10.413% while Mozambique being the lowest with 0.0517%. This explains that majority of the people living in the USA are interested in continuing their studies after their secondary schooling education within the age groups 25-64 years. Although the plot shows many countries and their mean percentage values, we can clearly notice that majority of the countries showed a value of 3% and above. This clearly tells us that the there is an importance given for completing the tertiary education in many countries.

The scatter plot represents country on the X-Axis and the Average percentage values of educational attainment on the Y-Axis.

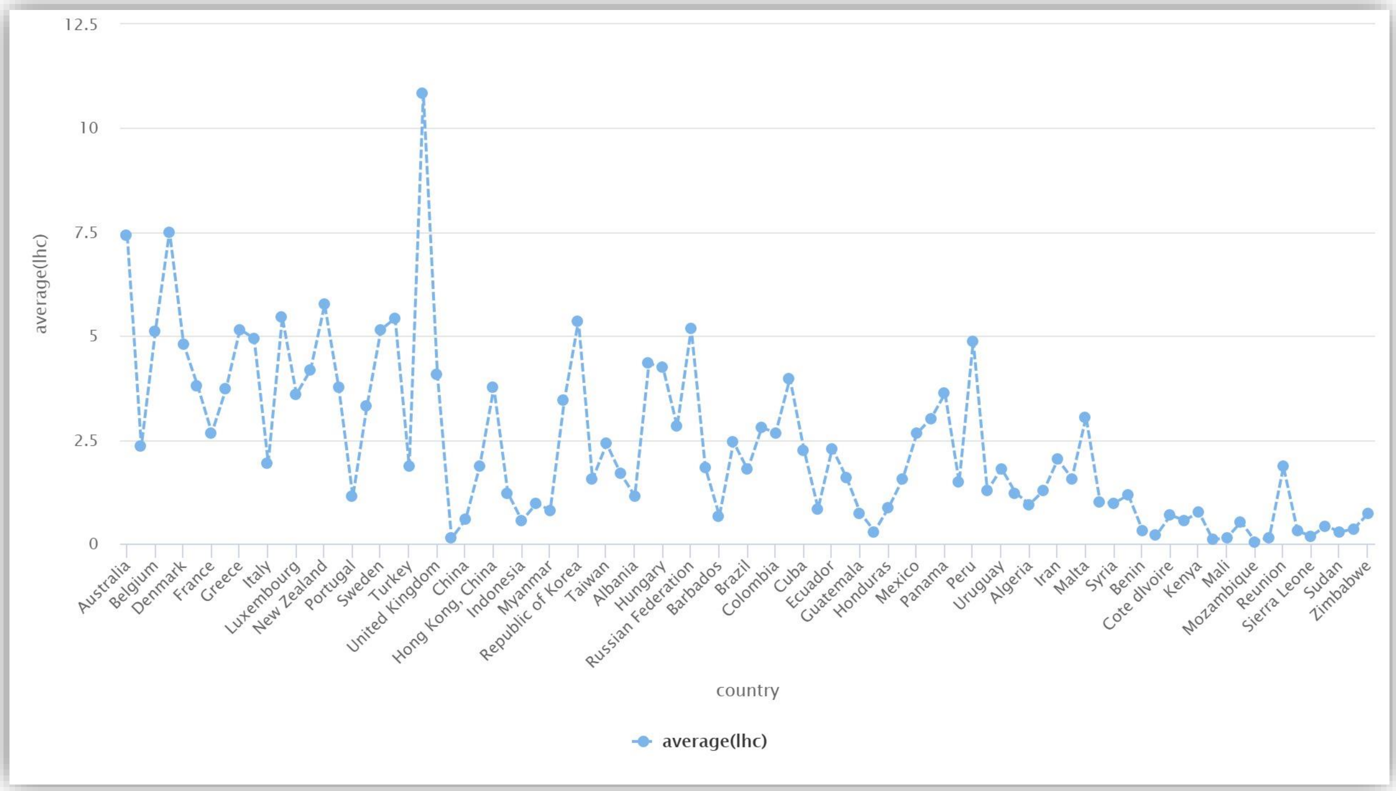


Figure 1: A scatter plot on Average percentage of population attained complete tertiary schooling

QUESTION 2:

To analyze the total population attained education in a year categorized by a region?

SOLUTION:

In this question, I tried to study the sum of the populace who attained education belonging to a specific region over 4 decades in the dataset. I have selected to answer this question as from the results we are able to determine the following:

- The trends in education completion levels for comparing various regions
- The total population values belonging to a specific region for very decade

On observing the trends depicted in the graph below, I can conclude that there is a dramatical increase in the total population (approximately 1592k) who showed interest towards achieving a complete schooling certificate in Asia and the Pacific region from the year 1940 to 2010. Another interesting pattern to be noticed is that the regions “Middle East North America” and “Sub-Sharan Africa” has shown very similar values. The Latin America and Caribbean has interchanged their pattern trends from the year 1960. Overall, it is observed that the first highest value is reported in the Asia and Pacific, the second maximum value is found at Advanced Economies (approximately 525k) while the remaining regions showing similar patterns except the above-mentioned trend.

The graph shows years on X-Axis, Sum of population values on Y- axis categorized by region code

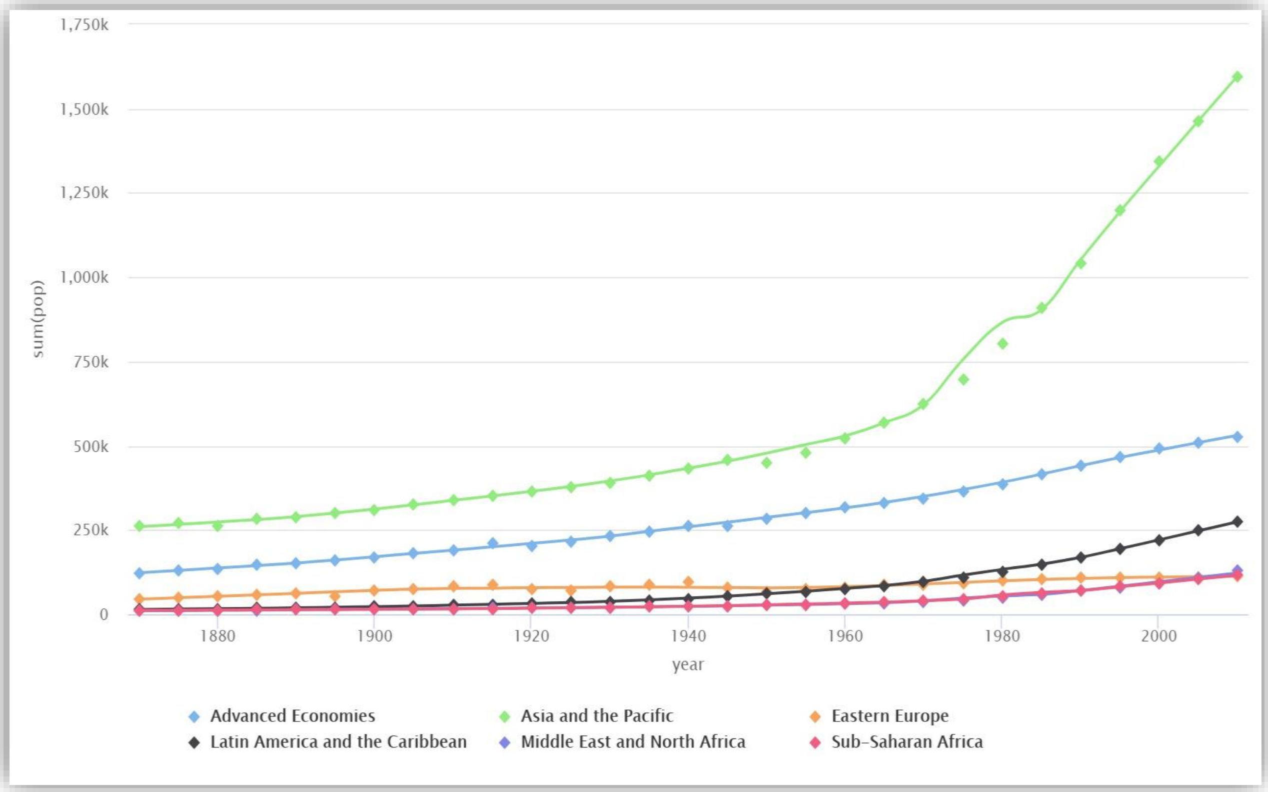


Figure 2: A line graph on total population attained education in a year categorized by a region

QUESTION 3:

To analyze the median of percentage of people not attending schooling by world bank country code (WBCode)?

SOLUTION:

For the above question, I have calculated the average of total percentage of people in a specific region listed as per the WBCode in the dataset. Additionally, it shows the population values who have not attained education at all by the age of 64. I have opted this question to answer the following listed question:

- To find out the countries that reported more non-education attainment percentage values across the world and make relevant comparisons

By Examining the pie chart below, it is interesting to know that most of that countries have shown similar median percentage values. The codes in the picture below are arranged in a descending order of the values with 99.3 for KHM (Cambodia) and 1.1 for KHM (Finland). It is observed that most of the values lie beyond the mean value. There are 15 countries showing median values greater than 90 and 9 countries above 80 and so on.
The pie chart represents the median proportion values of uneducated people grouped by World Bank country code.

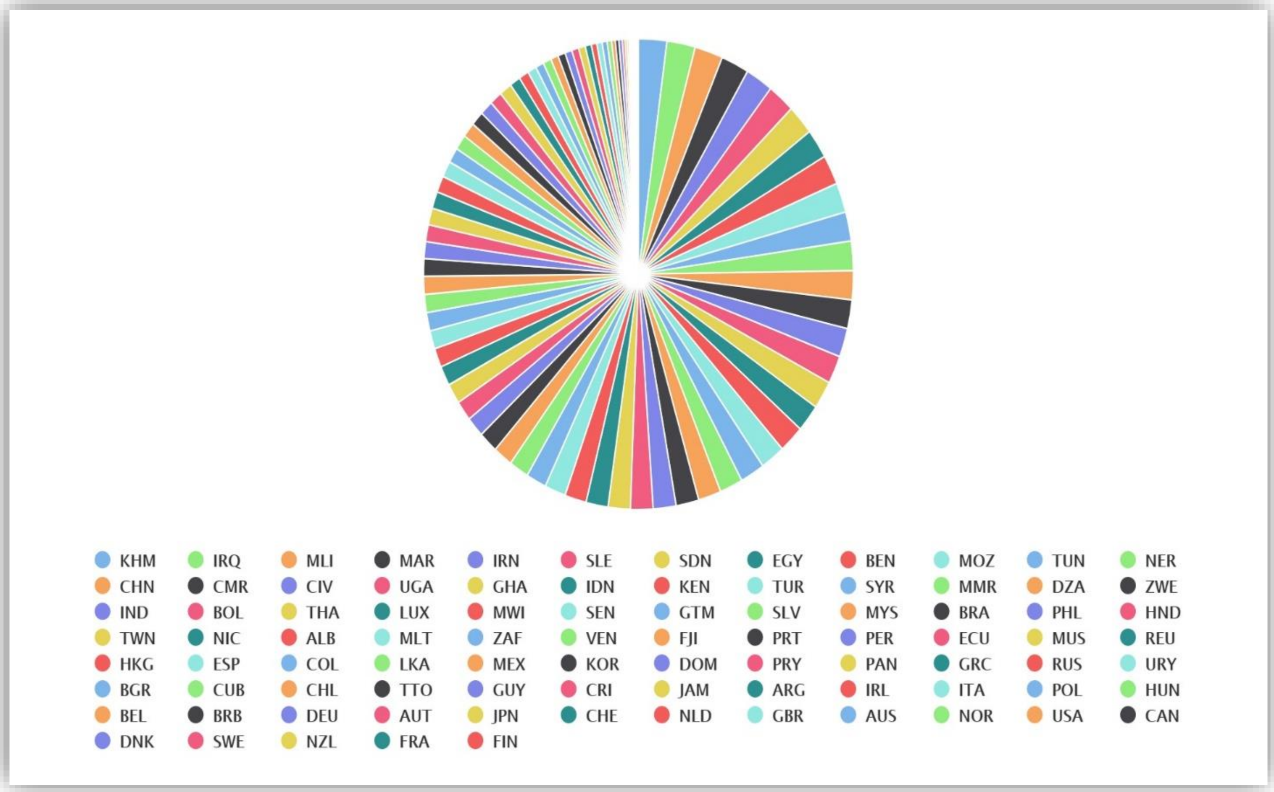


Figure 3: A Pie Chart representing median percentage of people not attending schooling by region

CONCLUSION:

Each year, there is a lot of amount of data generated which must be stored and maintained to perform some predictions in the future. The education sector especially has a lot of records to be used for determining the tuition rate across the globe. The education attainment dataset used in this report was based on the information for the years 1870-2010. By performing the analysis, the values are analyzed in a visual format easily with the help of RapidMiner tool. The insights found by answering the questions are beneficial for future analysis like prediction for the next decade etc. In addition, it helps in comparing the future trends when compared to the past observed results so that the accuracy of the prediction values can be easily determined.