**A) Introduction section - explain the problem you are trying to solve**

We are studying the effect that education levels of Canadian employees yielded on the mean number of hours and wages paid to male and females workers in the selected years ranging from 1997 to 2019. The overall aim is to find out time series wage structure, establish the impact of education level (elementary, high school, college and university) on wage and also analyze gender wage disparity across these levels. Further, we also aim at achieving province level data for the regions in the Canadian territory. These evaluative needs are important for formulation of labor market policies, encouraging and supporting educational processes for targeted populations, explaining wage disparities and gender pay gap, and obtaining insights into the economic and social returns to education.

**Why this Dataset:**

With respect to the theoretical framework, we arged that education changes wage trends and the gender wage gap, so we chose the average hourly wage rates of male and female workers in Canada by education level and gender from 1997 to 2019. This period covers important changes in the economy and society, which lettres focus on how education affects wages and the wage gap by gender. The results indicate the breakdown of regions and how education affects the gender wage gap and contribute to designing labour market policies, as well as educational interventions that could narrow wage gaps and reduce persistent labour market inequalities.

**Data source:**

<https://data.ontario.ca/dataset/wages-by-education-level/resource/7b325fa1-e9d6-4329-a501-08cdc22a79df>

**B) Data processing**

**i.** **Segment data sets as tables(4 tables minimum) with relationships**

As per the ERD diagram we have divided the main raw table into 4 different tables using normalisation technique.

* Demographics Table: This table has demographic information with player demographic\_id as unique fields including education level and age group. It is related with wages table through the demographic\_id.
* Geography Table: Geography table creates geography related information with columns such GEOGRAPHY\_ID(Geo-Primary key) And GEOGRAPHY\_NAME. It links to the wages table under the geography\_id.
* Worktypes Table: As shown in the following table, work\_types has unique work\_type\_id for each entry, and attribute work\_type describes the type of the work. It has many-to-one relation with the wages table through the work\_type\_id.
* Wages Table: This table function as the main table and has fields, including wage\_id (primary key), year, geography\_id, work\_type\_id, demographic\_id, wage\_type, and numerical fields of mouth and female. It uses data from demographics, geography and worktypes tables.

**ii. ERD Diagram:**A screenshot of a computer

Description automatically generated

**iii. Data Cleansing:**

To preprocess the dataset, all numerical columns are converted into uniform data types while all categorical columns are stripped of white space and converted to lower case. In the ‘education\_level’ column, mappings are used to transform long descriptions to shortened ones, string matching performed to filter rows with unwanted values and after that rows with such values are removed. Mapping dictionary is changed based on the dataset values so it meets the last dataset after it is checked that all transformations are correctly implemented.

**C) ) Analysis and Insights - Create SQL queries and provide interesting insights about your dataset.**

**1. Rank Regions by Average Male and Female Wages Across All Work Types:**

This query estimates the mean wages of males, females, and a common wage bill (mean of male and female) for each region. It makes use of the RANK() window function to rank all the regions on the basis of total average wage coming down order. The results of each of the regions is, therefore, sorted according to the rank, starting with the region that has the highest rank.

A screenshot of a computer

Description automatically generated

**2. Average Male wage across the Education levels by Years:**

With this query, the average wage of a male employee is computed by year at every level of education. It combines tables using wage and demographics to include the education level and calculate the average male wage by the resulting education level and year. The results are divided by year and education level.

A screenshot of a computer

Description automatically generated

**3. Average Female wage across the Education levels by Years:**

Like the query 2nd one, the result of this query is an average female wage based on the level of education represented by the yellow columns for a specific year represented on the horizontal axis. It connects wages and demographics and computes the average female wage with the result sorted by year and education.

A screenshot of a computer

Description automatically generated

**4. Yearly Wage Gap Between Genders for Each Education Level:**

It also established the gender wage differentials of wages for male and female by education level by year. The wage gap is then obtained by subtracting the average of average female wage from the average of average male wage. The results are also disaggregated by year and education level and arranged in this same manner.

A screenshot of a computer

Description automatically generated

**5. Work Types With the Largest Gender Wage Gap:**

This query defines the work types experiencing highest average gender wage differential. The difference is defined as the crude average of the basic wage rates of males and females in each type of work. Because this study seeks to compare work types based on wage disparity, the results are arranged starting with the work types with the highest disparity, in descending order.

A screenshot of a computer

Description automatically generated

**6. Yearly Comparison of Male and Female Wages Across All Work Types:**

This query computes for the average of male and female wages for specific types of work by year. It gives an average male and female earnings and organizes the results by year and type of work. As for the works he lists major works done by each author which are arranged by work type, year and author.

A screenshot of a computer

Description automatically generated

**7. Wage Trends by Work Type Over Time Using Window Functions:**

This query following employment type wage trends for given periods with the use of the LAG() window function. The average combined wage (male + female) is computed yearly for work type, and then, compared with the previous year using the LAG()function to determine the wage change between consecutive years. The results are grouped by work type and sorted by year.

A screenshot of a computer

Description automatically generated

**8. Identify Regions Where Female Wages Exceed Male Wages:**

This query defines areas where the salary earned by women is higher than the salary earned by man by different types of working. It tells us the percentage of records where the female wage rate is more than the male wage rate and then categorizes them according to geographical locations of an organization and types of work performed. All the findings are arranged based on the percentage of cases in which female wages are higher than the male wages, in a descending manner.

A screenshot of a computer

Description automatically generated

**9. Yearly Wage Disparities Between Work Types:**

The degree of the wage differential for each pairwise combination of work types in each year is calculated by this query. The ratio quantifies the absolute change in average wage for distinct work types and categorizes the results into years. This facilitates a comparison with the results obtained in the previous sections where the outcomes are arranged by year and the absolute measure of wage gap in descending order.

A screenshot of a computer

Description automatically generated

**10. Top Regions for Each Work Type by Wage:**

This query assigns the highest region for each type of work in respect to the average wage. It also computes an average of wages for every region, work type, and then using window function RANK() ranks the regions in terms of work types. After that, the query sorts the works again by rank which states the highest rank is 1 for each work type and orders the result based on descending order of average wage.

A screenshot of a computer

Description automatically generated

**D) Visualization:**

**1. Rank Regions by Average Male and Female Wages Across All Work Types:**

This data compares the average hourly wages for men and women in Canadian provinces:

Ontario: $15.70 overall, $16.37 for men, $15.04 for women

Alberta: $14.97 overall, $15.06 for men, $14.87 for women

Quebec: $14.85 overall, $15.31 for men, $14.38 for women

Other provinces follow a similar pattern.

Key points:

Men generally earn more than women in each province.

The wage gap varies by province.

There are significant regional differences in wages, with Ontario and Alberta being higher and Prince Edward Island and Newfoundland and Labrador being lower.

A graph of a number of wage

Description automatically generated with medium confidence

**2. Average Male wage across the Education levels by Years:**

The information illustrates the changes in the highest average male wages by education throughout 1997 to 2019. For the workers with education level of 0-8 years, the highest wage was $7.70 in 2019, which also increased year by year. All salaries, however, increased across the board with college graduates seeing the highest rise and stand at $16.58 in 2019, which showed that there was high demand for educated labor. Diploma holders had a peak wage of about $16. 45 for those who graduated from high school in 2019 and like the other levels, there was an increase as well. Those with university degrees had the highest overall wages at $20.82 in 2019 showing that despite gender wage disparities at all educational attainment levels, there was high returns to education investment.

**3. Average Female wage across the Education levels by Years:**

The data shows the average wages of women by educational attainment from 1997 to 2019. The highest average wage for low literacy employees with 0-8 years of education was $ 5.39 in 2019, which also increases year by year. Employees with college degree earned about $15.59 per head in the fiscal year 2019 due to the rise in the requirement of intellectuals. Again, the average wages of high school graduates increased to dollars 15.48 in year 2019 showing some improvement too. University degree holders earned generally the highest wages in this regard and increased to as high as $21.61 in 2019, nonetheless, this underlines the importance of earnings through education with a microcosm of tallied gender wage discriminations among employees with varied educational attainment.

**4. Yearly Wage Gap Between Genders for Each Education Level:**

The maximum gender wage gap as well as the gender wage gap by education level at the given points of time 1997, 2003, 2009, 2014 & 2019 is provided in the data. For those with 0 to 8 years of education the wage gap was at its highest at 2.32 in 2019. On the other hand, the low earners with college education displayed a negative wage gap in several years; women low earners outearned men low earners by 0.99 in 2019. High school graduateds had the highest disparity of roughly 1:1.21 in 2003. On the other hand, university registration maintained a negative wage differentials throughout the period, and more pronounced at -0.251 in 2019 showing that women at thislevel of education earned less than their male counterpart.

**5. Yearly Wage Disparities Between Work Types**

The figure below establishes the fact that there is a differential wage rate between the full time and part time employees from 1997 to 2019. The gap widened from 8.90 in 1997 to 15.95 in 2019 with fluctuating figures slightly diluting after 2002, 2005, and rising steadily from 2008 till the end of the period under analysis.

A graph showing the growth of the company's sales

Description automatically generated

**E) Summary:**

The trends and gender pay gap over education levels-average hourly wages by gender for 1997-2019 are discussed in the article. Education is the biggest determinant of wages, with employees holding a university degree earning the highest wages of $20.82 for a man and $21.61 for woman in 2019. However, gender wage discrimination still remains, and it is even more pronounced at all educational achievements in particular, in the group of employees with 0-8 years of education, when the indicator was 2.32 in 2019. On the other hand, college-educated women earned 1.99 times more than college degree educated men in 2019 to give the message that there are changes in wage expectations. In total, this research highlights the need for education as a way of also increasing wages, albeit highlighting that there are still issues pertaining to gender pay which needs an efficient policy approach towards wage differentials.

**Contribution:**

|  |  |
| --- | --- |
| **Pratik** | **SQL queries** |
| **Mitul & Irfan** | **DATA Cleaning and Normalization** |
| **Jasbir** | **SQL and Visualization** |
|  |  |