**LMIC test Junior Data Scientist position**

**Dataset Analysis:**

1. The ‘naics’ column values of the provided dataset contain alphanumeric, special characters. It doesn’t seem relevant, at least for the given tasks; only the alpha-characters seem valuable as they represent different sectors.

* I cleaned up ‘naics’ values by keeping only alpha characters.
* Removed all spaces and converted each to uppercase to ensure all data are included in the analysis, ignored case-sensitivity or additional spaces.

1. The given ‘Month’ column values are not in date format. Also they were not ordered based on month’s sequence.

* I processed ‘month’ column values according to the tasks. For 1st task, I considered only the year to perform group-by operation. For 2nd task, I converted those into standard date time month format for sorting, as the changes of values over months should be sequential.

1. When plotted data for 3 sectors of each province, I found missing values for “British Columbia” for “Apr”, whereas all the other provinces had data for months.

* I had to programmatically determine how many months of data are available of each province for the given range. However, this missing data has indeed an impact on the general conclusion while comparing number of employments across provinces.

**Data Visualization (Task 2)**

* The general pattern across provinces is the number of employment in “W**holesale and retail trade”** are always higher than “**Accommodation and food services**” and **“Other services** throughout the time range for each month.
* The total number of employment combining the give 3 sectors shows similar trend across provinces during the given months. The total number of employments was higher at the beginning of year; lowest in Mar-Apr; but gradually increased after that. The only exception is Quebec province; in “Apr” it had negative number of employments.

|  |  |
| --- | --- |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Alberta.png** | **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_British Columbia.png** |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Saskatchewan.png** | **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Manitoba.png** |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_New Brunswick.png** | **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Newfoundland and Labrador.png** |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Nova Scotia.png** | **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Ontario.png** |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Prince Edward Island.png** | **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Quebec.png** |
| **E:\All Aplication\Interview-Task\LIMC\Data_visualization\Visualization_Canada.png** | |