Task 2 : Employment Data Analysis

Introduction

Hello Everyone! Welcome to my Blog. In this blog, I will be demonstrating the second task of my virtual internship given by PsyliQ. I will be Analyzing Employee Dataset.

Datasets

They have provided me with four datasets :

1. Employee_data: It contains information about the Employee demographics.
2. Employee_engagement_survey_data: It contains information related to the survey date.
3. Recruitment_Data: It contains the recruitment details of each employee.
4. Training_and_Development_data: It contains the Training date and details of the Training program of each employee.

Let’s begin the Analysis

Task Questions

1. Can you create a pivot table to summarize the total number of employees in each department?

Select the Employee_data by using **Ctrl + Shift + End > Go to Insert Tab > Click on Pivot table > Drag the Department Column in the Rows field > Employee_ID Column in the Values field**. Change the name of the header.

Department	Total_Employees
Admin Offices	80
Executive Office	24
IT/IS	430
Production	2020
Sales	331
Software Engineering	115
Grand Total	**3000**

2. Apply conditional formatting to highlight employees with a "Performance Score" below 3 in red.

Copy the First Name, Last Name, and Current Employee Rating from the Employee_Data into the new sheet. After that, I concatenated the First and last name of the employee into one Column called **Employee_name**. I have used **=CONCAT(B2," ",C2)** function for concatenation. Then **Select the Current Employee Rating Column > Go to the home tab> Conditional Formatting > Highlight Cells Rules > Less than > type 3 and Click OK.**

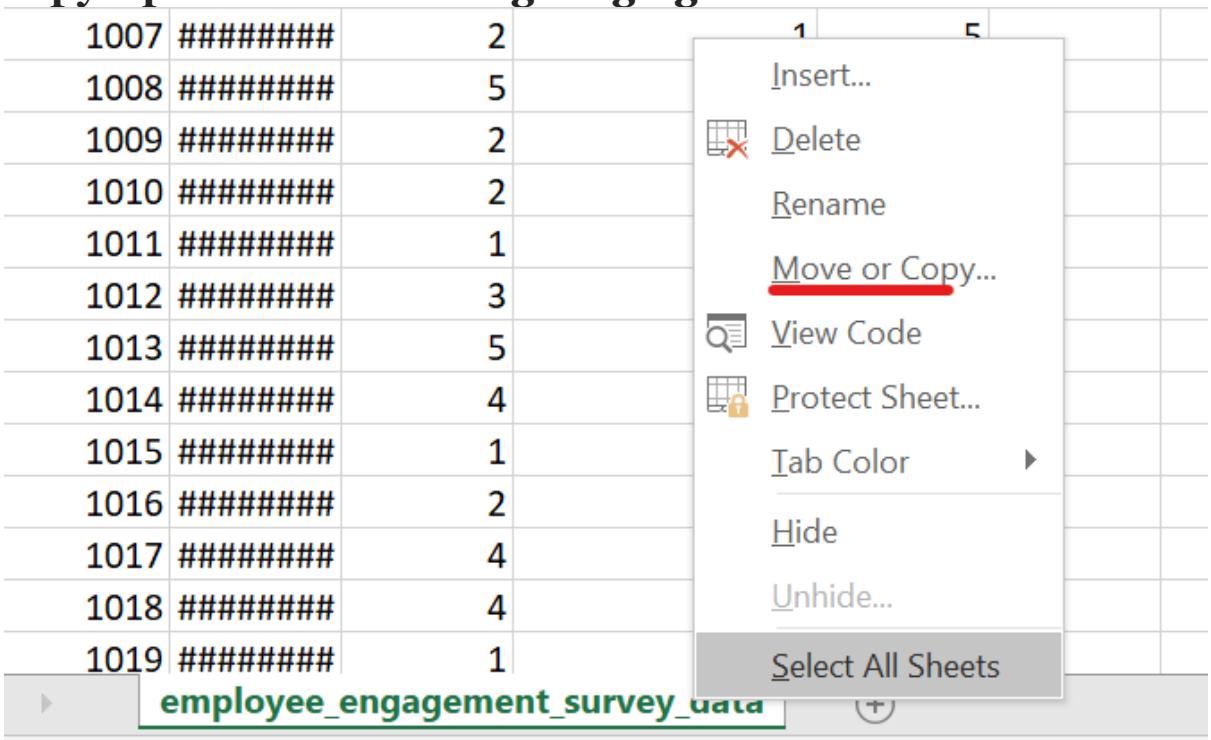
Employee Name	Current Employee Rating
Zoie Rowland	3
Zoie Mercado	3
Zoie Logan	1
Zoey Spence	3
Zoey Page	3
Zoe Colon	3
Zechariah Strong	1
Zayne Mccullough	3
Zayne Hunter	3
Zavier Reese	2
Zariah Ritter	3
Zariah Castaneda	5
Zariah Black	2
Zaria Mcintosh	4
Zaria Kidd	1
Zaria Benton	3
Zara Werner	5
Zara Pitts	3
Zaniyah Roman	3
Zaniyah Parks	4

3. Calculate the average "Satisfaction Score" for male and female employees separately using a pivot table.

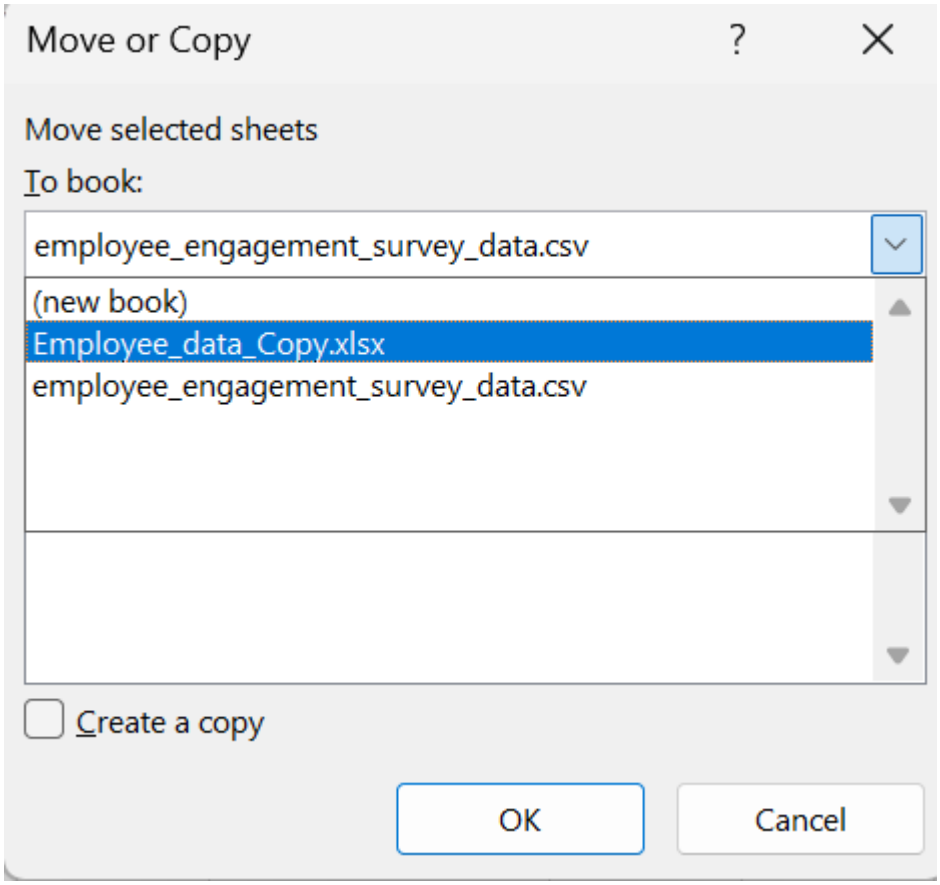
The Satisfaction Score column is in the **Employee_engagement_survey_data workbook** and the Gender Column is in the **Employee_data Workbook**.

So To create a **pivot table**. First I have to merge these two workbooks into one. So, that I can summarize them in a table.

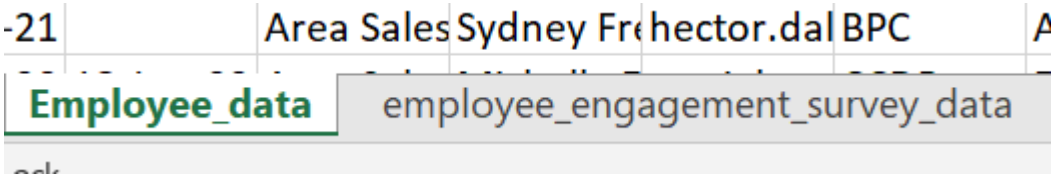
First of all, **Open Employee_engagement_survey_data workbook > right-click on the sheet > Click on Move or copy Option as shown in giving figure**



Then **Select the Employee_Data_Copy sheet** or Sheet where you want to move > **Click OK.**



Now you can see, you have successfully moved the sheet from one workbook to another.



Now Moving on to answer the question :

We get the required sheet. Now I want the Satisfaction Score column on Employee_data sheets for that I have used the **VLOOKUP** Function.

Vlookup function is used when you need to find things in a table or a range by row.

In its simplest form, the VLOOKUP function says:

=VLOOKUP(What you want to look up, where you want to look for it, the column number in the range containing the value to return, return an Approximate or Exact match — indicated as 1/TRUE, or 0/FALSE).

So to bring the Satisfaction Score I have written a flowing formula.

=VLOOKUP([EmployeeID],employee_engagement_survey_data!A2:E3001,4,FALSE)

I have successfully extracted the Column.

Satisfaction Score
5
5
5
5
4
2
1
2
5
4
2
5
4
4
1
3
1
1
3
1

Now for ease, I have converted both sheets into a Table

Why is that?

Converting the Range into the Table has lots of benefits :

1. When you convert your Range of Cells into the table. You will get an extra tab in the Excel ribbon called Table Design. It highlights alternate rows of the table and it looks pretty good! There is a bunch of designs present in Excel which you can choose from.
2. Inside the Table Design tab, you can get multiple features. One of the features is it has a filter button on the header so you don't have to go into a Data Tab to apply a filter.
3. You can get the Total Row, and You can even calculate the Aggregation function individually on numerical values. So do not have to write a formula for that.
4. You can also refer one table to a different table with a table name or a column name.
5. The Table is Dynamic. You can easily add a new row or column in the table and the formatting of the table will apply to a new row. And if you have done any aggregation by using the formula. It will also consider the value of the new row.

6. You can also add slicer. A slicer is more like a filter where you can filter your data according to the values you have selected.

7. When you have millions of rows in the Dataset. In range, you have to freeze your header to see the header. But In the table, the freezing feature is already applied.

Here is the summarization of the Satisfaction Score of Male and Female Gender.

Gender	Avg of Satisfaction Score
Female	3.01
Male	3.04

4. Create a chart to visualize the distribution of "Work-Life Balance Score" for different job functions

The work-life balance Score is on a different sheet. To bring this column in Employee_data. I have used the Vlookup Function.

Here Employee ID is a lookup value, Employee_Engagement is my table range and 2 is my Column index Number and I want Exact Match.

=VLOOKUP([Employee ID],Employee_Engagement,2,FALSE)

Work-Life Balance Score
5
3
2
3
5
1
5
2
1
2
3
4
3
1
3
2
3
4
5
2
4
4
3

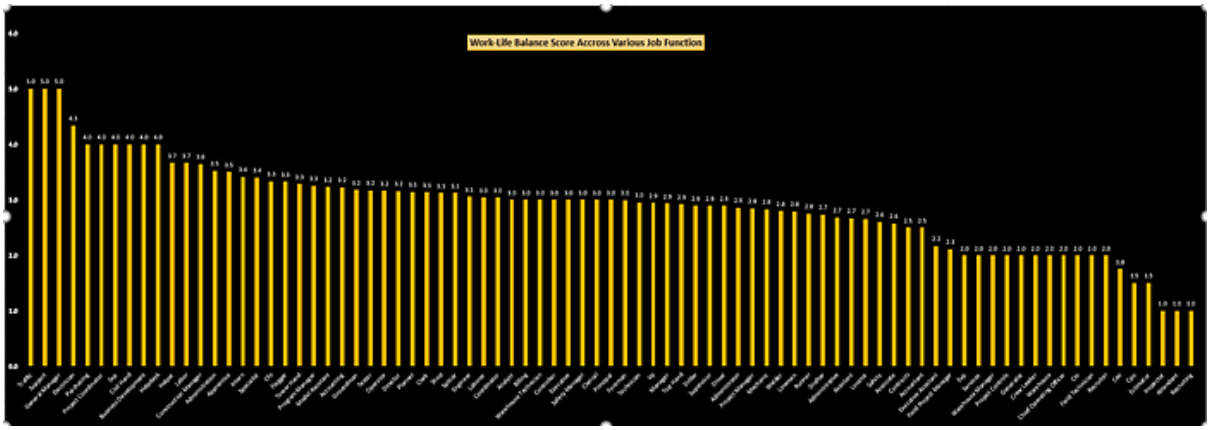
Now I have already made a pivot table for the previous question. So, I do not have to make another one by going back to the employee table and doing the the process again. So, I just have to update a pivot table to reflect the changes I have made in the employee Table.

How can we update the pivot table?

Select any cell of the pivot table then Click on the Analyze tab in the Ribbon and then Click Refresh.

Then you will see the work-life balance score Column get updated. Now you can make a pivot table by dragging the Work-life balance score column in the values field and Job functions in Rows Filed. Then select the entire pivot table. After that Go to the Insert tab and then Recommended Charts.

Here is the Work-life balance Across Various Functions.



5. Filter the data to display only terminated employees and find out the most common “Termination Type.”

Drag the Termination type column in the Row field and then again drag the Termination type column in the values field to count the number of termination types.

Unk is the most common termination type

Termination Type	Total Termination
Unk	1467
Involuntary	388
Voluntary	388
Resignation	380
Retirement	377
Grand Total	**3000**

6. Calculate the average "Engagement Score" for each department using a pivot table.

With the help of the Vlookup function. Bring the Engagement Score Column from the Engagement table (Employee_engagement_surevey_data) into the Employee Table(Employee_data).

=VLOOKUP([Employee ID],Employee_Engagement,5,FALSE)

Refresh the pivot table field list to see the Engagement Score. After that Drag the Departments and Engagement Score column into rows and values fields and sort the Average in descending order.

Departments	Average of Engagement Score
Executive Office	3.38
IT/IS	3.03
Sales	2.99
Software Engineering	2.97
Admin Offices	2.93
Production	2.91
Grand Total	**2.94**

7. Use VLOOKUP to find the supervisor's email address for a specific employee.

The supervisor's email address is in the recruitment workbook. So I move the workbook into the employee_data workbook. After adding a new sheet, I inserted a few Employee_IDs and then used the Vlookup function to search for the supervisor's Email ID.

Employee_ID is the lookup value, recruitment_data is the table array, 8 is the column index number, and False means I want an Exact match.

=VLOOKUP([Employee ID],recruitment_data,8,FALSE)

Supervisor Email Address

Employee ID	Supervisor Email
3005	rhenry@example.org
1089	smiller@example.com
2576	marshallnicholas@example.org

8. Can you identify the department with the highest average "Employee Rating?"

Select the Employee_data and make a pivot table. Drag the Department Type and Current Employee Rating Column in the Rows and Value field. After that change the Number format of the Average Rating Column and Sort the Column from highest to lowest, Change the headers name and Highlight the Highest Avg Rating with green color.

Departments	Average Rating
Admin Offices	3.03
Production	2.98
IT/IS	2.97
Sales	2.91
Software Engineering	2.90
Executive Office	2.79

9. Create a scatter plot to explore the relationship between "Training Duration (Days)" and "Training Cost."

Select Training Development data Insert a Pivot table. Drag the Total Duration (Days) Column in the Rows field and Total Cost in the Values field. Select the pivot table, Go to the Insert tab, and then Charts.

10. Build a pivot table that shows the count of employees by "RaceDesc" and "GenderCode."

Drag the Race Column in the Rows field, Gender Column in the Column field, and Employee ID in the values field.

Race	Gender		
	Female	Male	Grand Total
Asian	346	283	629
Black	346	272	618
Hispanic	325	247	572
Other	318	264	582
White	347	252	599
Grand Total	**1682**	**1318**	**3000**

11. Use INDEX and MATCH functions to find the "Training Program Name" for an employee with a specific ID.

Training is the table array, and L6 is the look-up value. Here Lookup values are the Employee ID, Lookup values are going to match with the Employee ID of the training table, and 0 means I want an exact match, and I want column no 3 to display.

=INDEX(training,MATCH(L6,training[Employee ID],0),3)

Employee ID	Training Programe
2678	Project Management
1290	Leadership Development
3000	Technical Skills
1678	Leadership Development
1020	Technical Skills

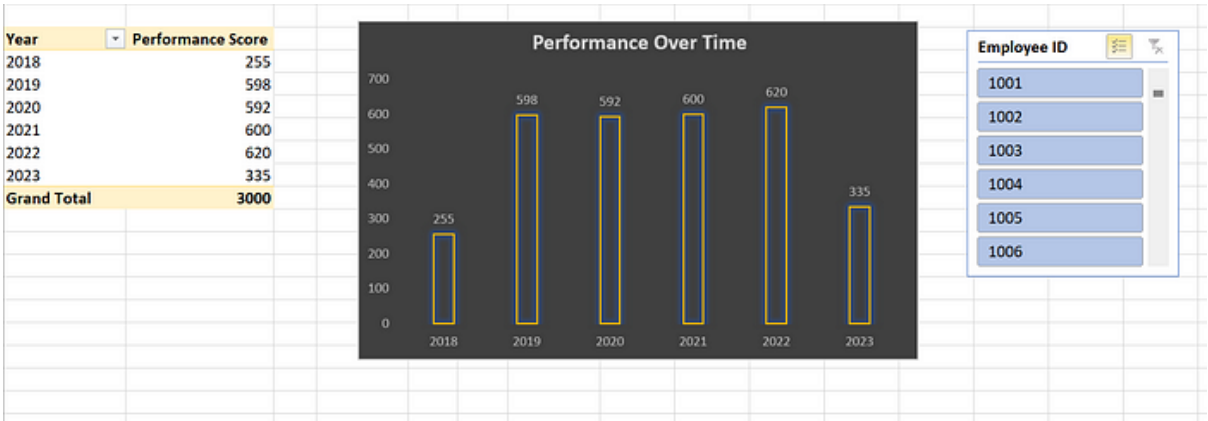
12. Create a multi-level pivot table to analyze the "Performance Score" by "BusinessUnit" and "JobFunctionDescription."

Drag the Business Unit column in the row field, the Job function description column below the Business Unit Column in the row field, and the Performance score column in the values field.

Business Unit	Performance Score
⊞BPC	303
⊞CCDR	300
⊞EW	302
⊞MSC	296
⊞NEL	304
⊞PL	301
⊞PYZ	299
⊞SVG	304
⊞TNS	297
⊞WBL	294
Grand Total	3000

13. Design a dynamic chart that allows users to select and visualize the performance of any employee over time.

Drag the Year column in the row field, and Performace Score in the values field. Select the pivot table, Insert a Column Chart, and then Select any cell of the pivot table after that go to the Analyze tab in the ribbon and then Insert slicer.



14. Calculate the total training cost for each “Training Program Name” and display it in a bar chart.

Drag the Training Program name in the rows field and Training cost in the values field. Select the pivot table, Insert a Bar Chart, and do some formatting.

Training Program Name	Total Training Cost
Communication Skills	₹ 3,65,023.24
Project Management	₹ 3,43,313.17
Leadership Development	₹ 3,23,902.03
Technical Skills	₹ 3,23,072.61
Customer Service	₹ 3,20,575.04
Grand Total	**₹ 16,75,886.09**

15. Apply advanced conditional formatting to highlight the top 10% and bottom 10% of employees based on “Current Employee Rating.”

Select the Current Employee Rating Column, then Go to the Home tab then click Conditional Formatting. Then Choose the Top/Bottom Rules option and Click Top 10 % for the first condition and Bottom 10% for the second condition.

Current Employee Rating
3
3
1
3
3
3
1
3
3
2
3
5
2
4
1
3
5
3
3
4
1
3
3

16. Use a calculated field in a pivot table to determine the average “Engagement Score” per year.

Bring the Engagement Score Column from the Engagement table into the Employee Table using the Vlookup function and then select the Employee table to insert a pivot table. After that Drag Year, and Engagement Score into the rows and values field. Then change the value setting to Average.

Year	Average of Engagement Score
2018	2.86
2019	3.03
2020	2.94
2021	2.87
2022	2.91
2023	2.84
Grand Total	**2.92**

17. Can you build a macro that automates the process of updating and refreshing all pivot tables in the workbook?

Yes, I can build a Macro that automates the process of updating and refreshing the pivot table.

Here I am summarizing the Total training cost of training programs.

Training program Name	Total Training Cost
Communication Skills	3,65,023.24
Customer Service	3,20,575.04
Leadership Development	3,23,902.03
Project Management	3,43,313.17
Technical Skills	3,23,072.61
Grand Total	**16,75,886.09**

Now, I will create a macro that will refresh the table. First of all, I will turn on the developer option and then I will record a macro and refresh the pivot table by clicking any cell in the pivot table and then I will stop recording a macro. I will insert a button for refreshing the table. After that, From the source table, I will delete the

communication training program. I will come back to the pivot table and click on the Refresh button.

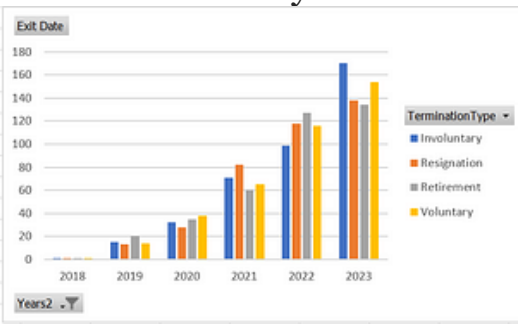
Training program Name	Total Training Cost
Customer Service	3,20,575.04
Leadership Development	3,23,902.03
Project Management	3,43,313.17
Technical Skills	3,23,072.61
(blank)	3,65,023.24
Grand Total	**16,75,886.09**

Refresh

18. Create a histogram to understand the distribution of “ExitDate” for terminated employees

Drag Exit date, Termination Type , and EmployeeID in the rows field, Column field, and Values field. Select the pivot table and insert the column chart. As the histogram is not available in my Excel.

Exit Date	Termination Type: Involuntary	Resignation	Retirement	Voluntary	Grand Total
2018	1	1	1	1	4
2019	15	13	20	14	62
2020	32	28	35	38	133
2021	71	82	60	65	278
2022	99	118	127	116	460
2023	170	138	134	154	596
Grand Total	388	380	377	388	1,533



19. Utilize the SUMPRODUCT function to calculate the total training cost for employees in a specific location.

The sumproduct function is used to multiply the corresponding values in the ranges and then sum up those products.

Here I have used the sum product to calculate the training cost of the Bruceshire location. Training8 is the table name, Location is the column name.

=SUMPRODUCT((training8[Location] = “Bruceshire”)* training8[Training Cost])

Bruceshire	145.99

20. Develop a dashboard that provides an overview of key HR metrics, including headcount, performance, and training costs, using charts and pivot tables.

Thank you for reading. Feel free to share your thoughts, ask questions, and let’s continue this journey together.

Additionally, if you’d like to connect with me, you can find my LinkedIn ID and GitHub profile below.

LinkedIn
GitHub

Happy learning!