HR DATA ANALYSIS ASSESSMENT

JESSICA ATARI-ATA

Data Acquisition

This Analysis was conducted using a comprehensive dataset provided by **PSYLIQ.**

Project Highlights

A detailed analysis was conducted to identify the following HR parameters:

- Employee Demographics
- Employee Welfare
- Key Performance Indicators
- Other General Employee Data

Data Cleaning Process

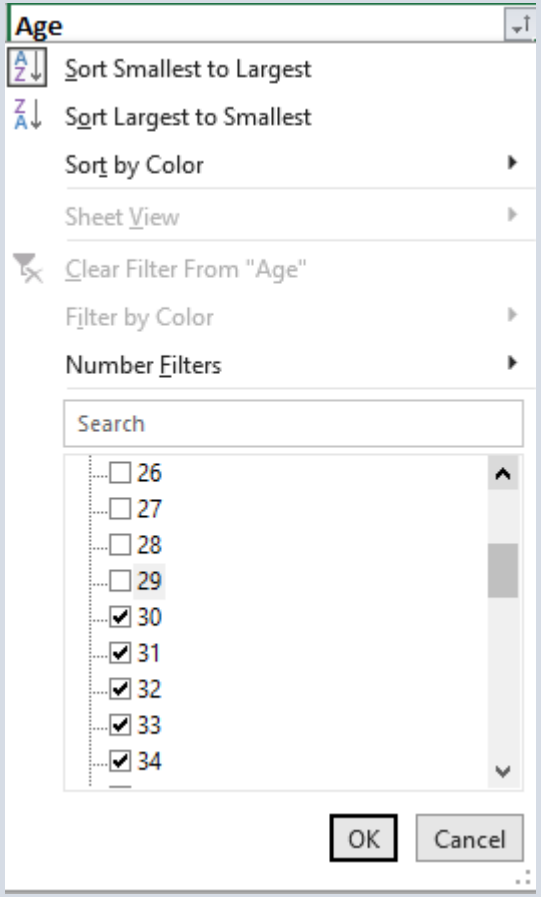
- Sorted the “Age” Column from Smallest to Largest
- Filtered for Duplicates, None was found.
- Corrected some abbreviations e.g. the ‘Y’ to ‘Yes’, and ‘curr’ to ‘Current’ for consistency.
- Filtered out “blanks” from the Dataset.
- Changed the following under the education column to ensure consistency:
- 1 – Below college
- 2 – College
- 3 – Bachelor
- 4 – Masters and
- 5 – Doctor

Project Tool

- Excel & Power BI

1). Using Excel, how would you filter the dataset to only show employees aged 30 and above?

➔ Filter the Age Column and uncheck all other ages below 30 as shown below:



2) Create a pivot table to summarize the average Monthly Income by Job Role.

Job Role	Average Monthly Income
Healthcare Representative	60,983.74
Human Resources	58,528.08
Laboratory Technician	66,314.05
Manager	63,395.88
Manufacturing Director	69,183.72
Research Director	65,473.13
Research Scientist	64,975.68
Sales Executive	65,186.69
Sales Representative	65,370.96
Grand Total	**65,029.31**

3) Apply conditional formatting to highlight employees with Monthly Income above the company's average income.
➔ The Companys' average income is **65,029.31.** The cells highlighted are greater than **65,029.31**

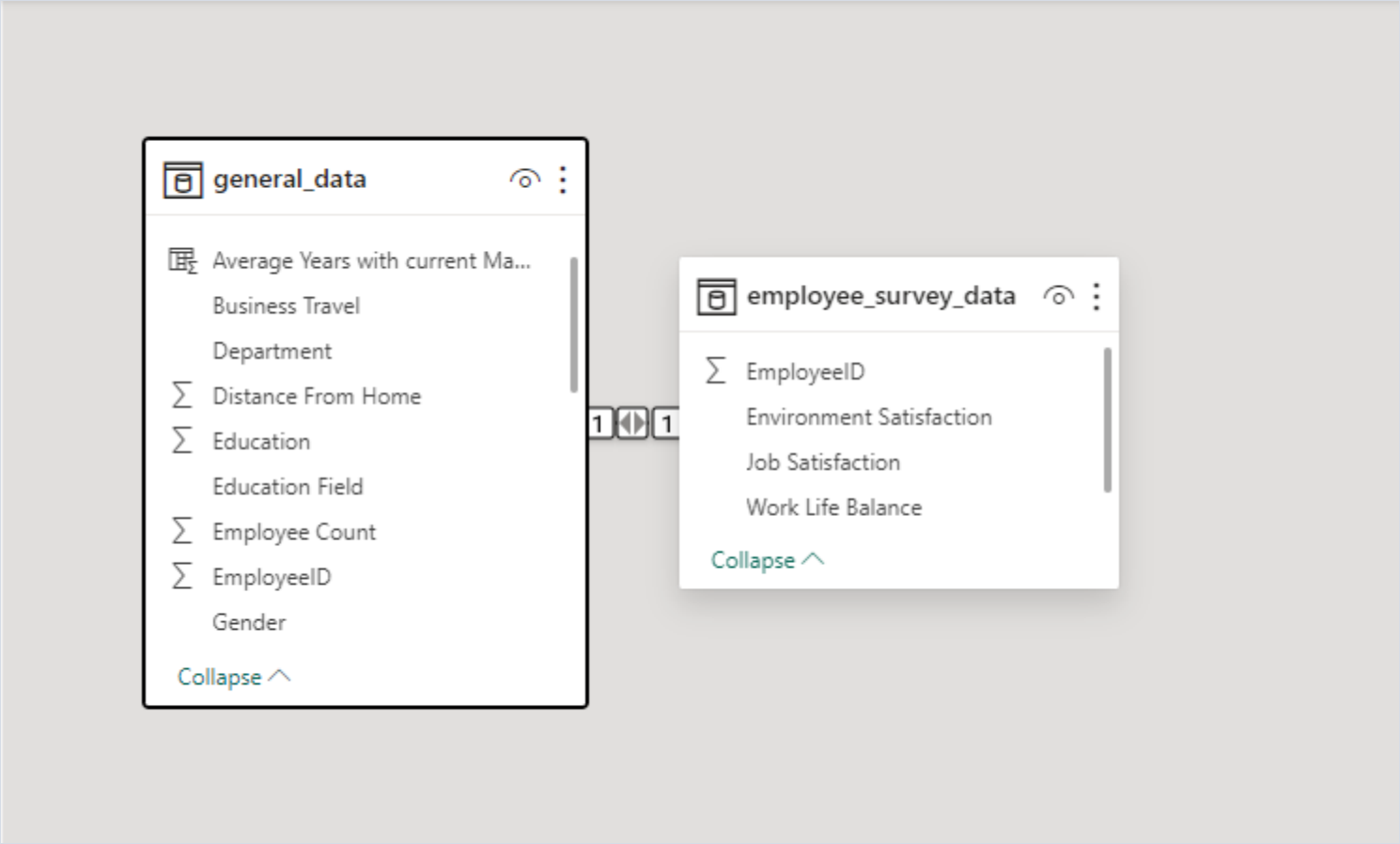
EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	Monthly Income
1016	Male	4	Research Scientist	Married	49,680
1082	Female	2	Laboratory Technician	Single	35,790
1166	Male	2	Sales Executive	Single	25,340
1230	Female	3	Sales Executive	Divorced	108,510
1322	Male	1	Sales Representative	Single	193,310
1341	Male	1	Sales Executive	Married	46,630
1373	Male	1	Laboratory Technician	Married	132,690
1419	Female	4	Healthcare Representative	Divorced	129,360
1596	Male	1	Research Scientist	Married	29,910
1656	Female	2	Laboratory Technician	Single	37,850
1711	Female	2	Laboratory Technician	Married	195,450
1843	Male	1	Sales Representative	Married	11,020
1879	Female	1	Laboratory Technician	Married	124,900
1947	Male	2	Sales Representative	Married	49,070
2073	Male	1	Sales Representative	Married	26,540
2350	Male	3	Sales Executive	Married	97,050
2486	Male	4	Research Scientist	Married	49,680
2552	Female	2	Laboratory Technician	Single	35,790
2636	Male	2	Sales Executive	Single	25,340
2700	Female	3	Sales Executive	Divorced	108,510

4). Create a bar chart in Excel to visualize the distribution of employee ages.

5). Identify and clean any missing or inconsistent data in the "Department" column.

- The Department column was filtered for missing or inconsistent data and none was found.

6). In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.



7). Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.

```
1 Average Years with current Manager = AVERAGE(general_data[Years With Current Manager])
```

	Total Working Years	Training Times Last Year	Years At Company	Years Since Last Promotion	Years With Current Manager	Average Years with current Manager
3	1	2	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
2	1	2	1	0	0	4.12312925170068
3	1	2	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
2	1	2	1	0	0	4.12312925170068
3	1	2	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068
2	1	2	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068
2	1	3	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
2	1	3	1	0	0	4.12312925170068
1	1	3	1	0	0	4.12312925170068
0	1	2	1	0	0	4.12312925170068

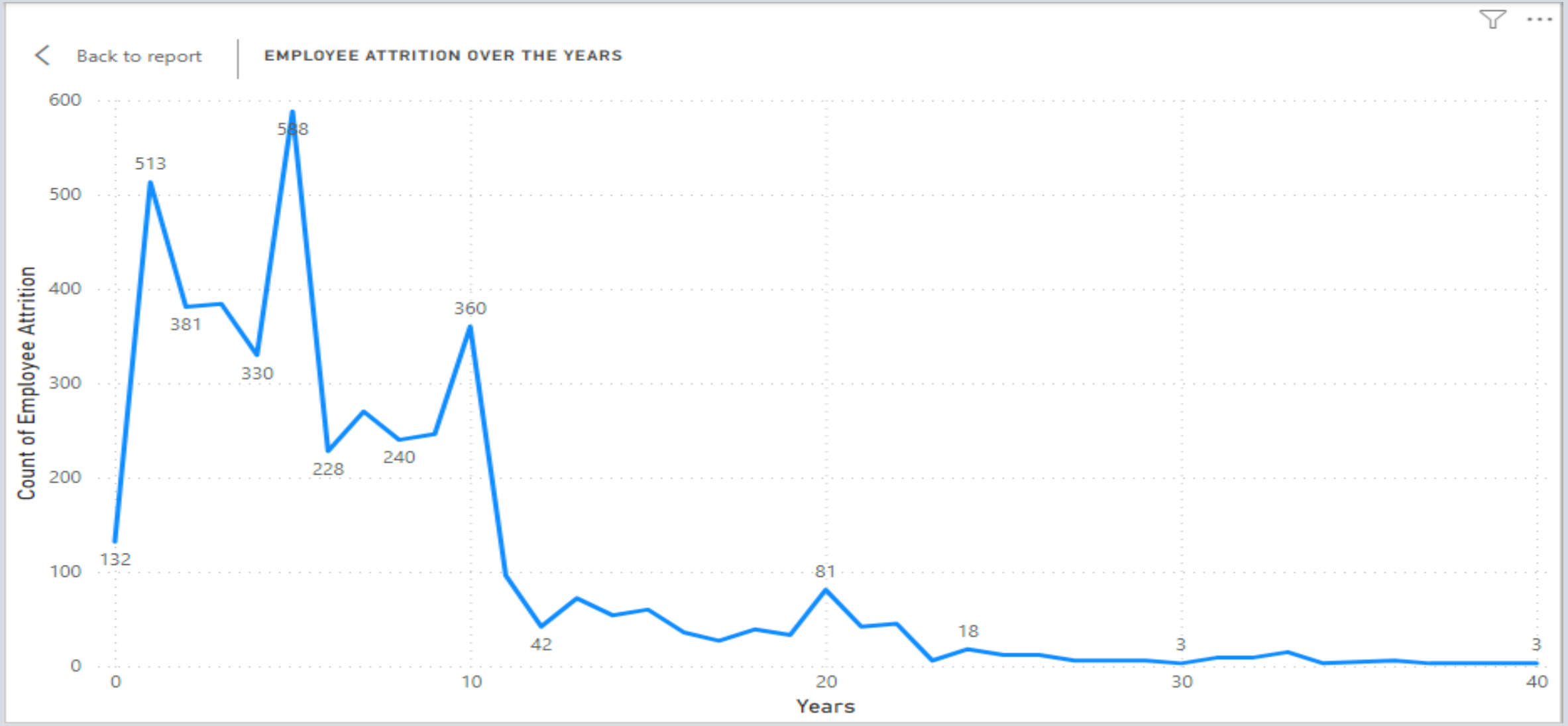
8). Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.

Department	Divorced	Married	Single	Grand Total
Human Resources	21	96	72	**189**
Research & Development	621	1350	912	**2883**
Sales	339	573	426	**1338**
Grand Total	**981**	**2019**	**1410**	**4410**

9). Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.

Gender	Job Level	Job Role	Marital Status	Monthly Income	Job Satisfaction
Male	1	Sales Executive	Single	109,650	4
Male	2	Manager	Single	186,060	4
Male	1	Research Director	Single	76,440	3
Male	1	Sales Executive	Single	191,970	4
Male	1	Sales Executive	Single	115,570	4
Male	2	Manager	Single	152,020	4
Male	1	Research Director	Single	160,640	4
Male	3	Sales Representative	Single	191,970	4
Male	1	Sales Executive	Single	98,240	3
Male	1	Research Scientist	Single	115,570	4
Male	1	Research Director	Single	191,970	3
Male	3	Sales Representative	Single	98,240	3
Male	1	Sales Executive	Single	115,570	4
Male	2	Research Scientist	Single	152,020	4
Male	3	Healthcare Representative	Single	93,960	4
Male	3	Laboratory Technician	Single	83,920	3
Male	1	Manager	Single	84,120	4

10) In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



11). Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

These are the steps to create a star schema for this dataset:

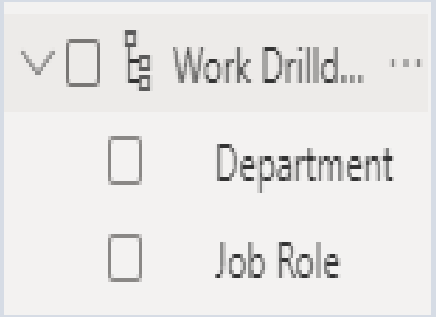
- First step is to identify the business process for analysis; in this case, HR analysis to help define the scope and purpose of the star schema.
- Second step is to identify the Measures or facts that make up the core of the schema. The fact table will contain the measures that are relevant to the business requirement such as; Employee ID, Department, Attrition, Performance Rating, Job Satisfaction and so on.
- Third step is to identify the dimensions which will contain the descriptive attributes that characterize the facts such as; Job level, Job title, Education level, Years at company and so on; making sure that the dimension tables created have primary keys that match the foreign keys in the fact table.
- **Benefits**
- Creating a schema for this dataset will help simplify analysis because the organization of data will make it easier to identify trends and patterns that might otherwise be missed.
- It can also help to handle large amounts of data that needs to be combined in dynamic ways and analyzed quickly.
- It is easy to maintain and update the star schema because of its simple design, while reducing the risks of errors.
- Queries will run faster on a star schema model enabling analysts get answers quickly and efficiently.

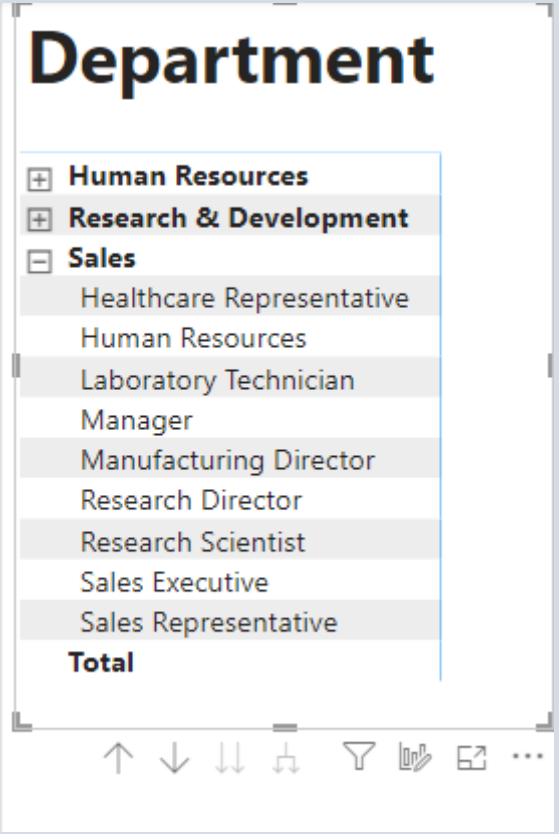
12). Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

```
1 Rolling Average monthly income = AVERAGEX(FILTER( ALLSELECTED('Date'),'Date'[Date]<= MAX('Date'[Date])), [Monthly income])
```

Times Last Year	Years At Company	Years Since Last Promotion	Years With Current Manager	Average Years with current Manager	Rolling Average monthly income
2	1	0	0	4.12312925170068	152020
3	1	0	0	4.12312925170068	160640
2	1	0	0	4.12312925170068	76440
2	1	0	0	4.12312925170068	152020
2	1	0	0	4.12312925170068	191970
2	1	0	0	4.12312925170068	191970
3	1	0	0	4.12312925170068	160640
2	1	0	0	4.12312925170068	76440

13). Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.





14). How can you set up parameterized queries in Power BI to allow users to filter data based on the Distance from Home column?

- Select **Home ➔ Transform data** to open the Power Query Editor.
- In the Power Query Editor, under **Manage Parameters,** select **New Parameters.**
- In the **Manage Parameters** window, fill out the information about the parameter.

15). In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.

Row Labels	Sum of MonthlyIncome
Human Resources	**3,259,140**
3	1,648,500
4	754,800
5	855,840
Research & Development	**53,502,900**
3	28,117,740
4	15,277,290
5	10,107,870
Sales	**22,974,330**
3	11,792,400
4	8,753,070
5	2,428,860
Grand Total	**79,736,370**

16). Explain how to perform a What-If analysis in Excel to understand the impact of a 10% increase in Percent Salary Hike on Monthly Income.

- Go to the **Data** tab, click **What-If Analysis** and choose **Data Table**.
- In **Row input cell**, enter the reference to the cell with Percent Salary Hike and click ok.

	41890	131160
5%	43984.5	137718
10%	46079	144276
15%	48173.5	150834

17). Verify if the data adheres to a predefined schema. What actions would you take if you find inconsistencies?

If inconsistencies are found in my predefined schema, I would do the following:

- Change the primary keys for a table.
- Change the relationship between tables.
- If it is a validation error, review your coding to identify the missing tag and input it.

Thank you