**Data Analysis Internship**

**HR Data Analysis**

Q11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

Ans. A star schema is a type of data model used to organize data in a database for easy understanding and analysis. It’s the simplest style of data mart and is widely used to develop data warehouses and dimensional data marts.

In the given dataset, I have made a Star Schema with the following connections –

1. **employee\_survey\_data <-----> general\_data : EmployeeID - One to One Relationship**
2. **manager\_survey\_data <-----> general\_data : EmployeeID - One to One Relationship**
3. **in\_time <-----> general\_data : EmployeeID - Many to One Relationship**
4. **out\_time <-----> general\_data : EmployeeID - Many to One Relationship**

The benefits of using a star schema include:

* **Simplicity**: The star schema is easy to understand and query due to its simple structure.
* **Performance**: The simple structure allows for fast query response times and efficient use of database resources.
* **Flexibility**: The star schema can be easily extended by adding new dimension tables or measures to the fact table, making it a scalable and flexible solution for data warehousing.

Q16. 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.

Ans. We can perform a What -if analysis in Excel to understand the impact of a 10% increase in Percent Salary Hike on Monthly Income using the “Data Table” feature.

1. **Set up the base model**: First, we need to set up a base model in Excel. This should include a formula that calculates the Monthly Income based on the Percent Salary Hike. For example, if the current Monthly Income is in cell B2 and the Percent Salary Hike is in cell B3, we will have a formula in cell B4 like =B2\*(1+B3).
2. **Set up the data table**: Next, we need to set up a data table. This is a range of cells where Excel will substitute different values for the Percent Salary Hike and recalculate the formula. In cell C3, enter the value 0.10 (representing a 10% increase). In cell C4, reference the formula cell (=B4).
3. **Perform the What-If analysis**: Now we are ready to perform the What-If analysis. Go to the Data tab on the Ribbon, then click What-If Analysis > Data Table. In the Row input cell field, enter the cell reference for the Percent Salary Hike (B3). Leave the Column input cell field blank, then click OK.

Excel will now recalculate the formula in cell B4 for a 10% increase in Percent Salary Hike and display the result in the data table.

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

Verifying if data adheres to a predefined schema and handling inconsistencies in Excel involves several steps:

1. **Understanding the Schema**: Understand the structure of the data and the predefined schema. The schema may include specific data types, range of values, or specific formats that each column should adhere to.
2. **Manual Review**: Manually review the data to identify any obvious inconsistencies. This could include checking for misspelled entries, incorrect data types, or values outside the expected range.
3. **Use Excel Features**: Excel provides several features to help identify inconsistencies:
   * **Filters**: Apply filters to specific columns to view unique values and quickly spot inconsistent data.
   * **Conditional Formatting**: Use formatting rules to highlight cells that meet specific criteria or mismatched data between columns.
   * **Data Validation**: Use data validation to ensure that the data entered into a cell meets specific criteria. For example, you can use data validation to make sure a value is a number, a date, or to present a dropdown menu with predefined choices to a user.
4. **Advanced Excel Functions**: Utilize functions such as MATCH and VLOOKUP to compare data sets and identify discrepancies between them.

If inconsistencies are found, here are some actions to take:

* **Correct the Data Manually**: For small datasets or a few errors, you might correct the data manually.
* **Use Find and Replace**: For consistent errors (like a misspelled word), you can use the Find and Replace feature.
* **Use Formulas or Functions**: For more complex inconsistencies, you might use formulas or functions to correct the data. For example, you could use the TRIM function to remove extra spaces.
* **Delete the Data**: If the data is not necessary or too inconsistent, it might be easier to delete it.