**HR data Analysis Project**

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

Answer: A dataset is organised into a central fact table encircled by dimension tables that are linked together via one-to-many relationships in order to create a star schema. Let's discuss how to build a star schema for a dataset that contains employee details such as job descriptions, performance metrics, and demographics.   
  
Determine the Fact Table: In this case, the fact table might be a table with performance metrics or other quantifiable employee-related business event, including productivity, sales, or performance evaluations.

Determine Dimension Tables: Demographics Dimension: Information like Employee ID, Name, Gender, Age, Marital Status, Education Level, and so forth would be included in this table.   
Job Dimension: Information about a job's department, role, level, employment status, hire date, and other details are included in this table.

Build Relationships: Using common keys, create one-to-many relationships between each Dimension Table and the Fact Table. For instance, the Employee ID may function as the Fact Table's foreign key and the Demographics Dimension's main key.  
Advantages of Employing a Star Schema

Simpleness: Star schemas are straightforward and simple to comprehend. Their core fact table, encircled by dimension tables, facilitates user-friendly navigation and data analysis.

Performance: Generally speaking, star schemas result in quicker query performance. Queries can run more quickly because dimension tables are denormalized and just have descriptive characteristics.

Flexibility: Star schemas provide an analytically flexible solution. By slicing and dicing data along multiple dimensions, users can obtain insights from diverse angles.  
  
Query Optimisation: Star schemas simplify queries by reducing their complexity through fewer joins than other schema types like snowflake schemas. This makes query optimisation activities easier to do.

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

Answer: Comparing the actual data with the anticipated structure and format specified by the schema allows one to determine whether the data follows a predefined schema. The processes to carry out this verification are as follows, along with what to do if discrepancies are discovered:

Examine the preset schema: Examine the established schema to learn about the restrictions, data types, and intended structure for every field and attribute.

Examine the Information: Examine a sample of the data to make sure the predetermined schema is being followed. Examine the data for any inconsistencies, unusual values, improper data types, and missing values.

Data Profiling: Apply data profiling methods to learn more about the quality, distribution, and uniqueness of the data. Determine trends, outliers, and possible problems with the data.  
Compare the Data with the Schema: Employ scripts or automated tools.

Examine the Information: Examine a sample of the data to make sure the predetermined schema is being followed. Examine the data for any inconsistencies, unusual values, improper data types, and missing values.

Data Profiling: Apply data profiling methods to learn more about the quality, distribution, and uniqueness of the data. Determine trends, outliers, and possible problems with the data.  
Compare the Data with the Schema: Employ scripts or automated tools