**To prepare the dataset for the next step in Tableau using SSMS, here are the detailed steps you might have taken:**

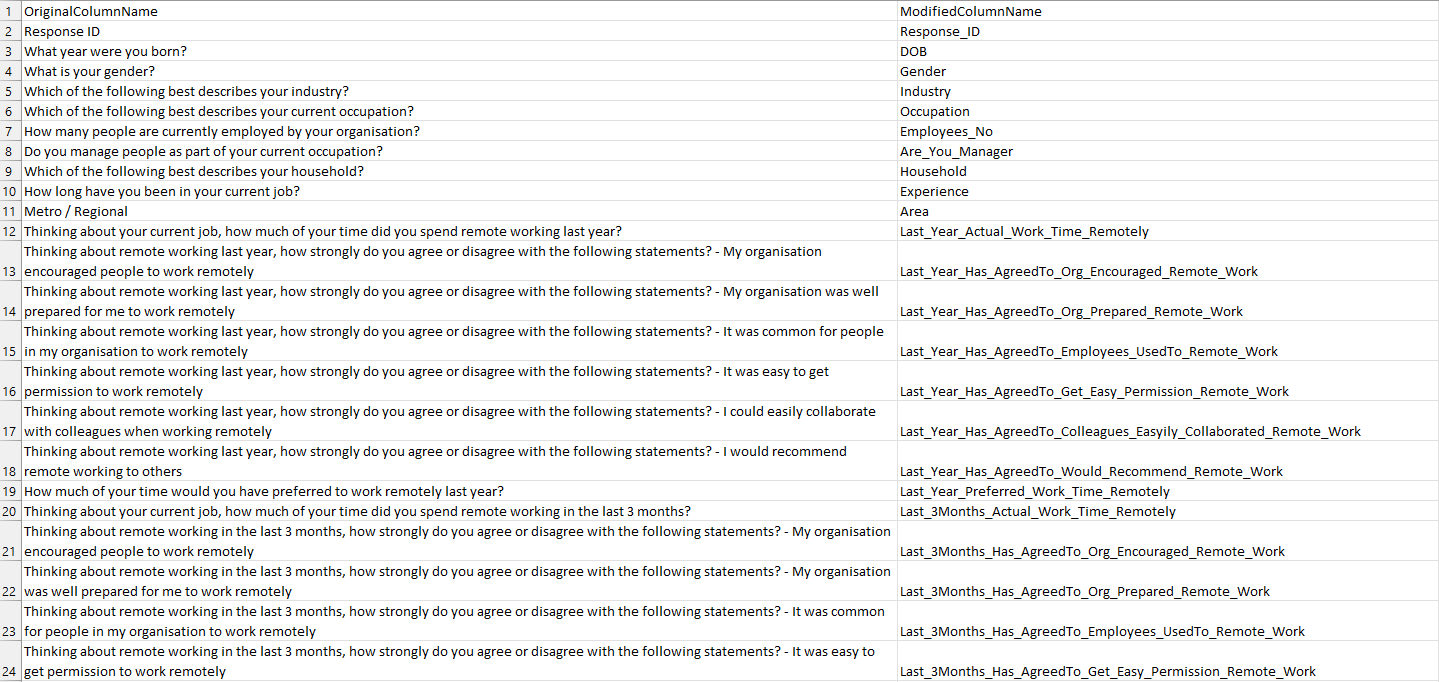
1. **Data Import and Validation**

* **Pre-Importing Data:**

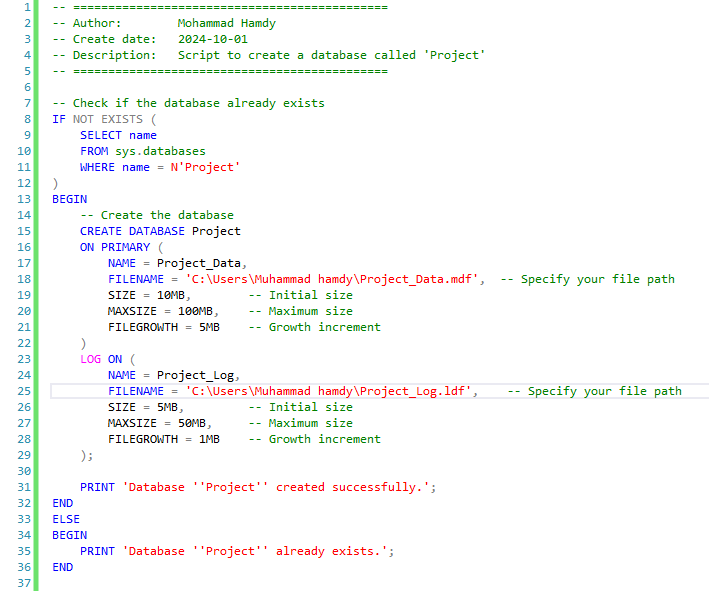
1. **SQL Column Nomenclature: With Excel I handled columns names as it was too long to fit as a column name in SQL DB, and not unified, so I copy names of columns, which were the questions of the survey, and used transpose feature and put the column names in a new sheet and make new names for columns, referring to the old ones names.**

**I followed these Naming Guidelines:**

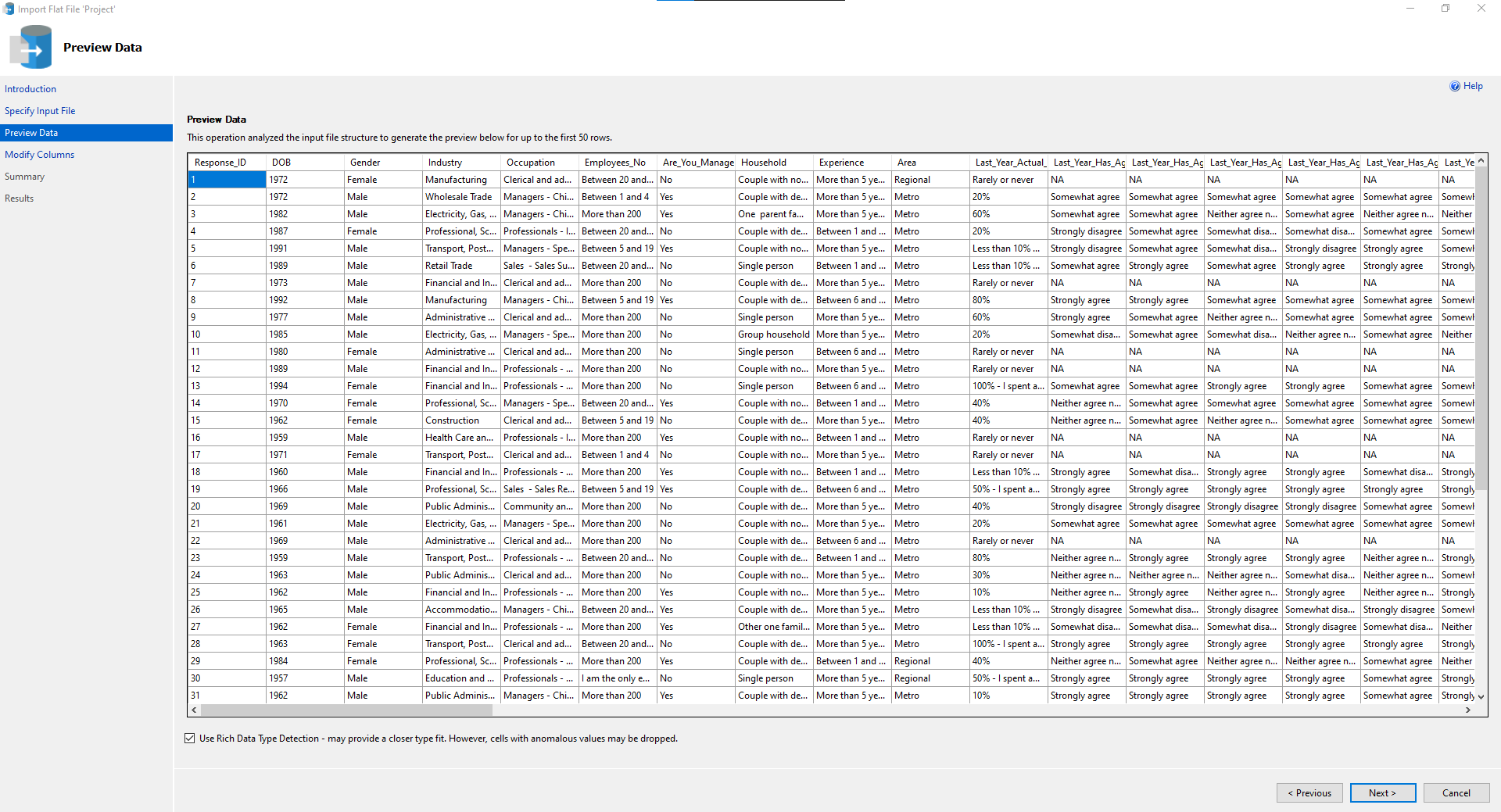
* **Meaningful Names**
* **Consistency**
* **Avoided Reserved Keywords**
* **Used Snake Case Style**
* **Prefixing and Suffixing like Data Type Prefixes**
* **Abbreviations and Acronyms**
* **Avoided Uncommon Abbreviations and Used Standard Acronyms**
* **Naming Conventions for Specific Data Types like Boolean Columns and Numeric Columns**
* **Documentation**
* **Commenting**

****

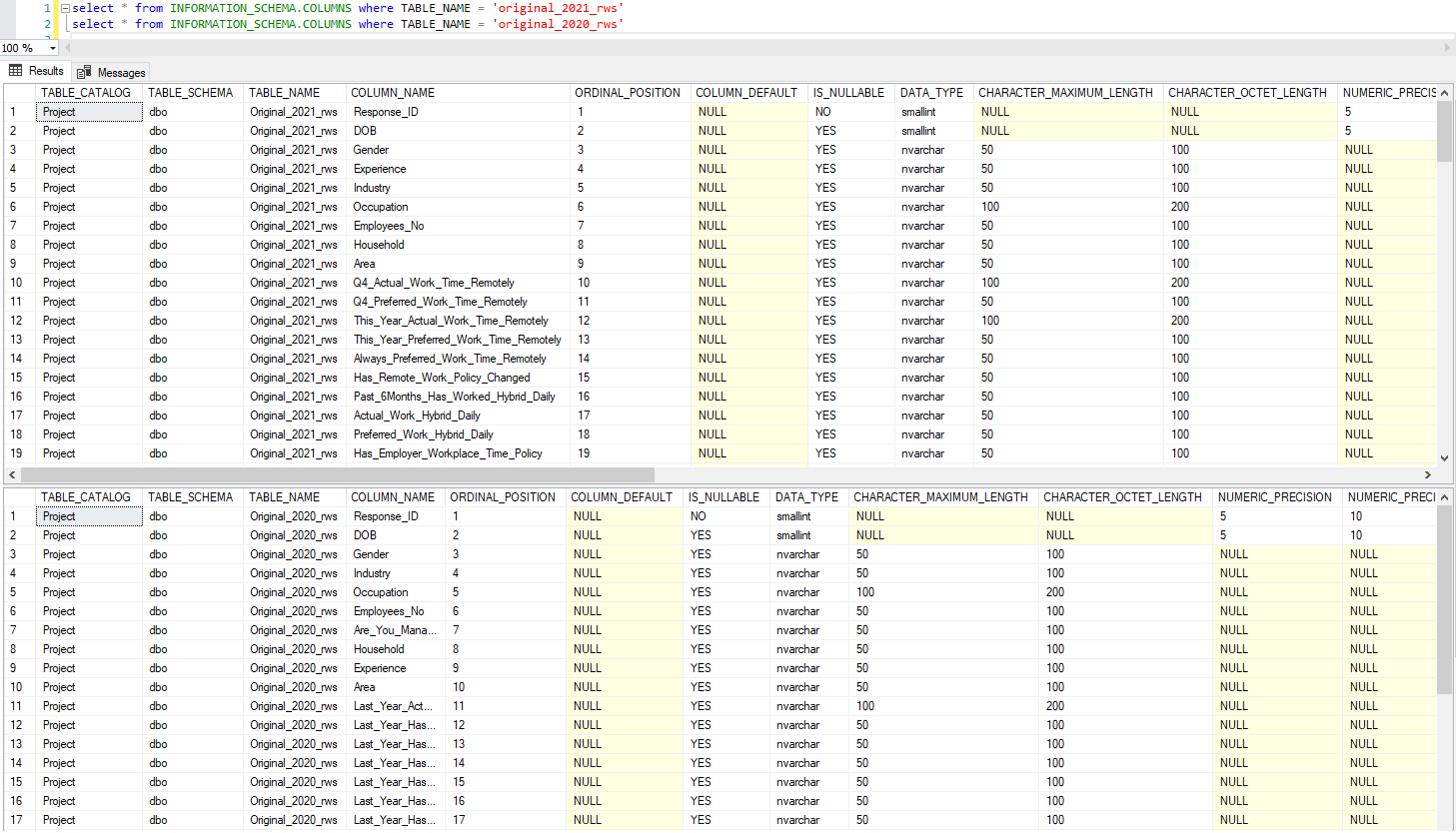
1. **Creating Database:**

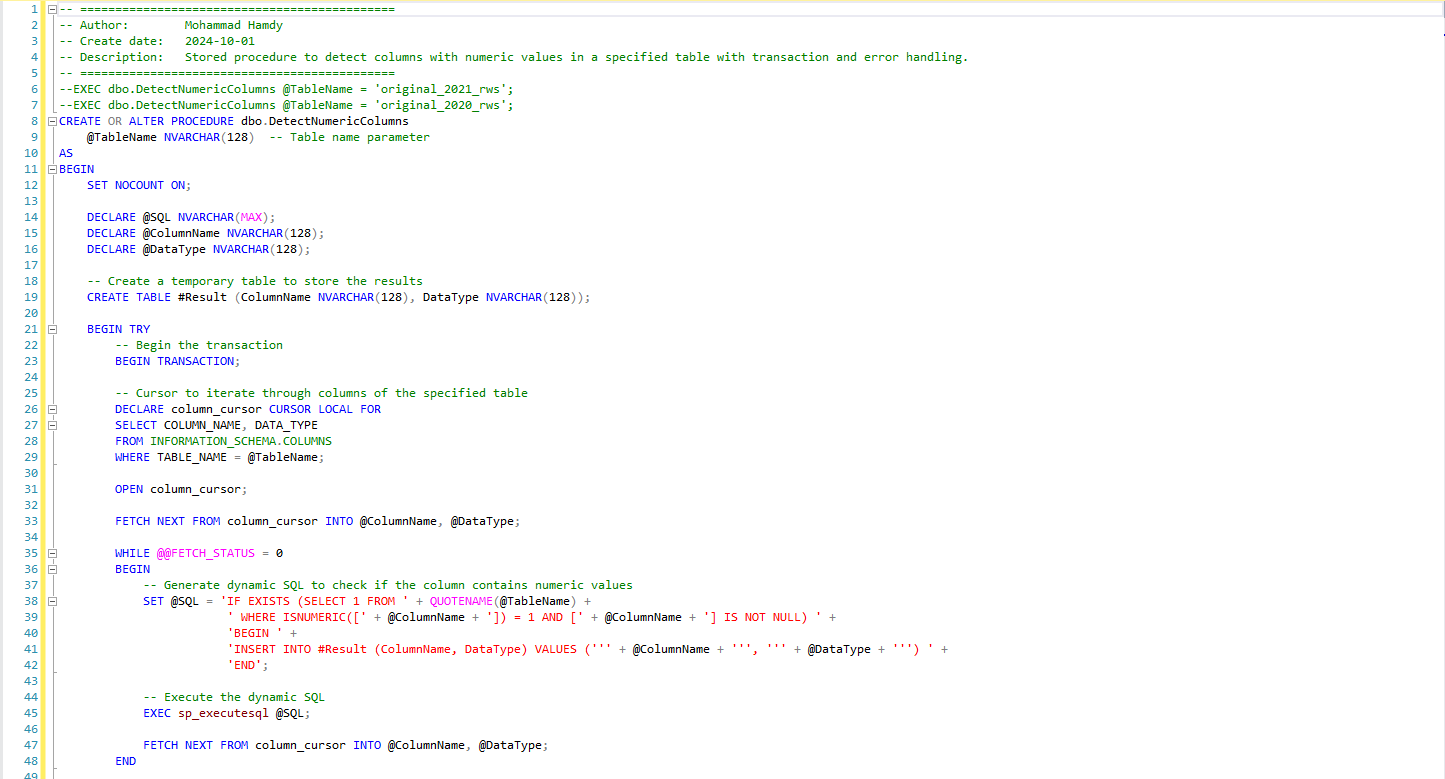
****

* **Import Data: Imported the dataset into SQL Server via SSMS, either by directly using SQL scripts or through the Import/Export wizard, ensuring that the data was loaded into the correct tables.**

****

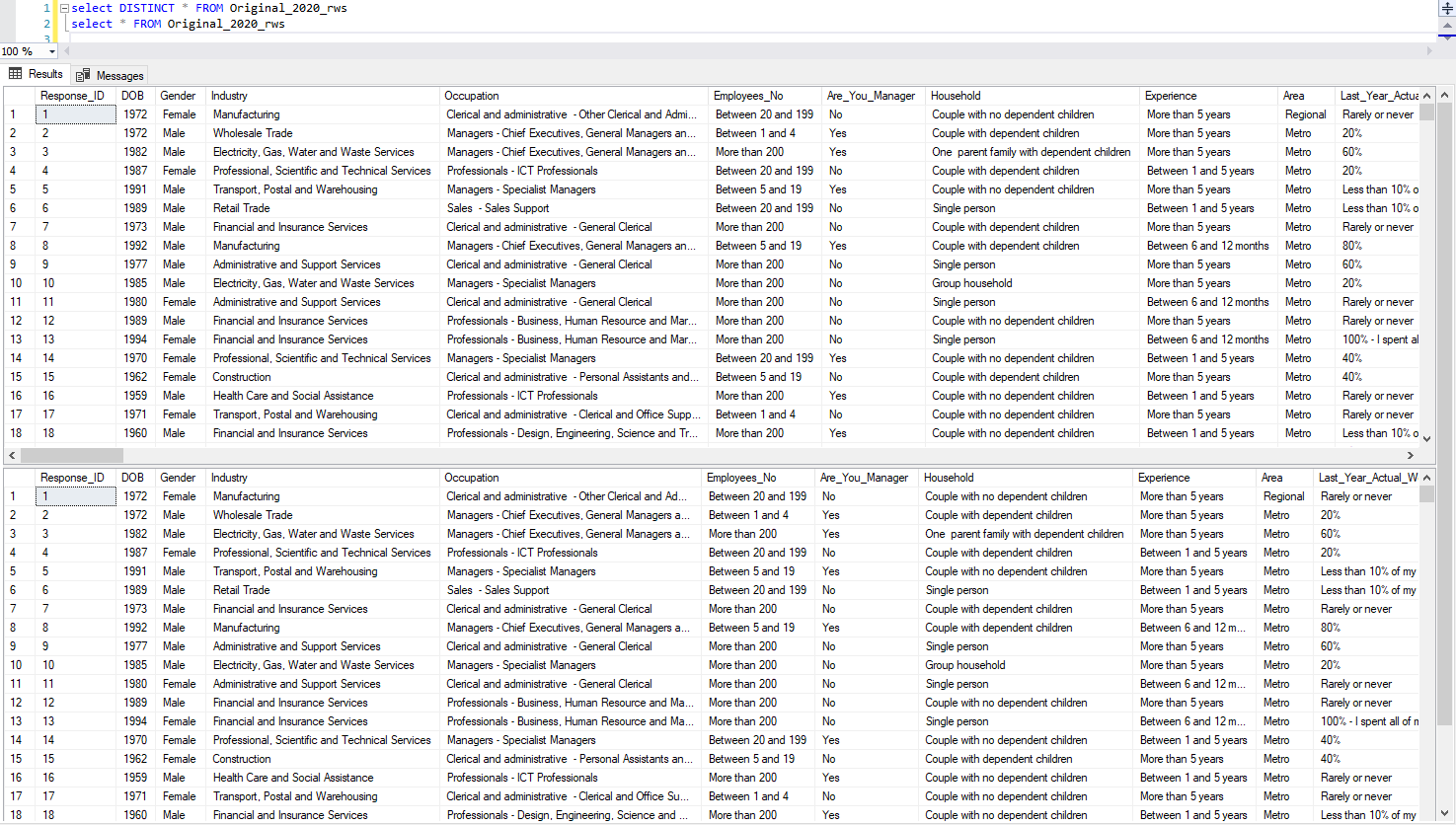
* **Data Validation: Checked if the data loaded correctly by running SELECT queries to verify the row counts, data types, and content consistency.**

****

****

**2. Data Cleansing**

* **Check for Duplicates to Remove Them: Used the DISTINCT keyword with a Common Table Expression (CTE) to identify and remove duplicate records.**

****

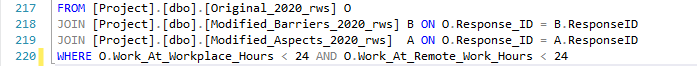
* **Handle Null Values: Checked for NULL values using ISNULL() and COALESCE() functions and replaced them with default values where necessary (NA, average using AVG() function, or zeros).**

****

* **Standardize Formats: Standardized data formats such as numbers, and text using functions like TRY\_CAST() for cast string data types that are only contain numerical fields to float data type.**

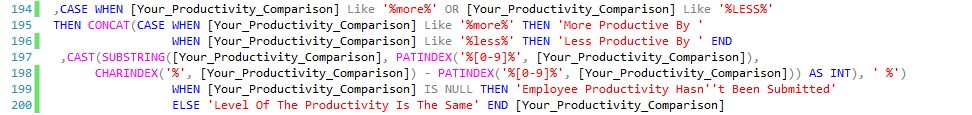
****

* **Remove Unlogic Outliers: Like in questions asking about work hours per day, using where to remove hours more than 24 Hours/Day**

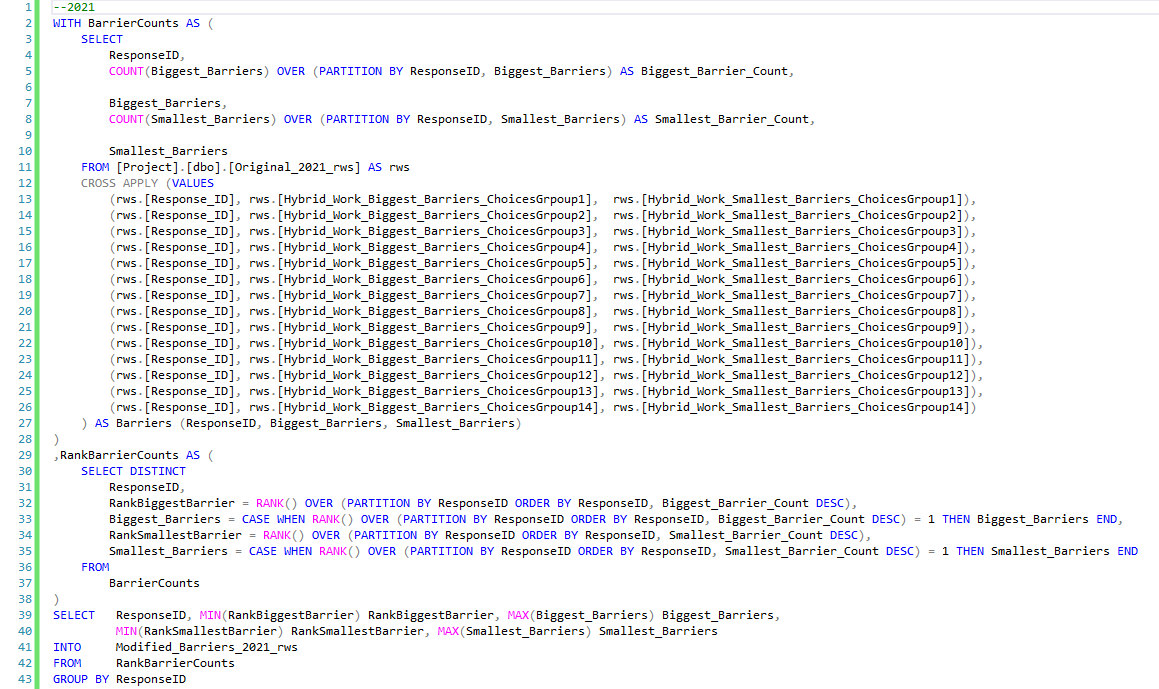
****

**3. Data Transformation**

* **Create New Columns/Derived Values: Added new columns using CASE WHEN statements or functions like or CONCAT() for text combinations, to unify data in both 2020 and 2021 tables like concatenation Occupation and Occupation\_Detailed in one column as Occupation, and to handle ordering data for better and correct usage in the visualization stage by Tableau.**

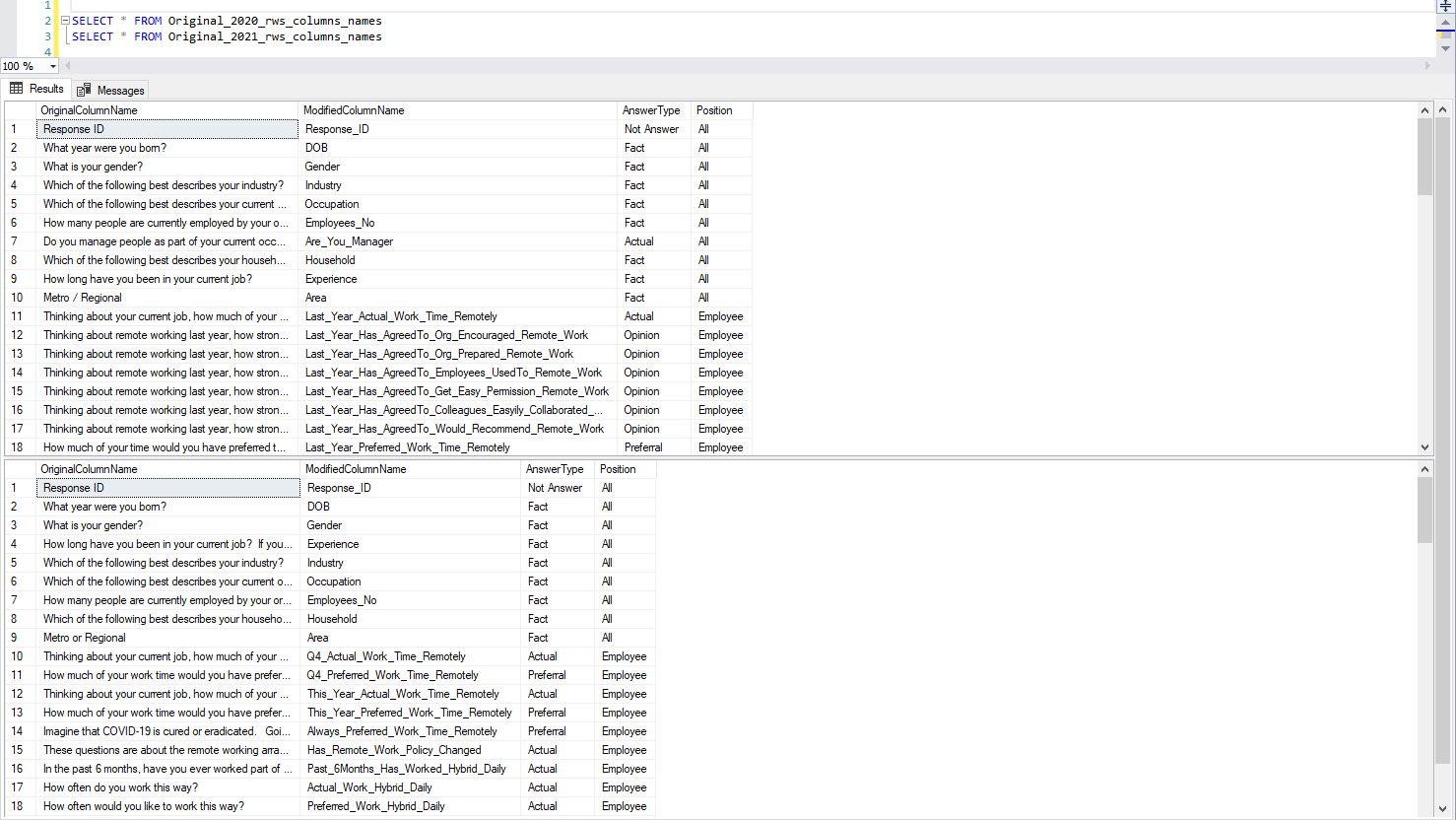
****

* **Normalize Data: Checked for columns if it needed to be split into separate fields if necessary for normalization.**
* **Joins: Joined data across multiple tables using INNER JOIN, or CROSS APPLY to combine related data from different sources.**
* **Data Aggregation: Created summary tables using aggregate functions such as COUNT(), MIN() and MAX() based on business needs.**
* **Data arrangement: Used window functions like ROW\_NUMBER(), RANK()**

****

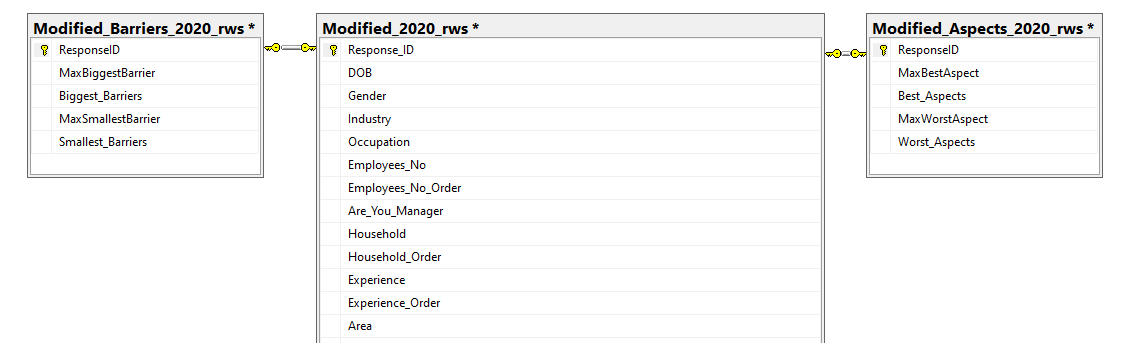
**4. Data Enrichment**

* **Add Lookup Tables: Linked data to lookup tables to enrich it with additional information, so I made a table contains original survey question column, modified column name, survey question types, and survey takers types to make it easy to get insights.**

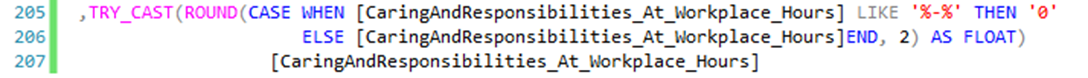
****

**5. Data Integrity Checks**

* **Foreign Key and Referential Integrity: Ensured that all relationships were consistent by verifying foreign key constraints.**

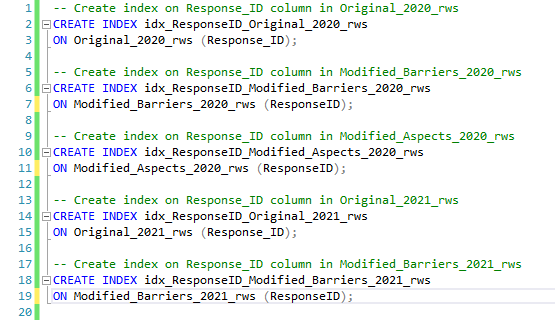
****

* **Data Validation Rules: Implemented and checked any business logic validation rules (e.g., ensuring no negative values for certain fields).**

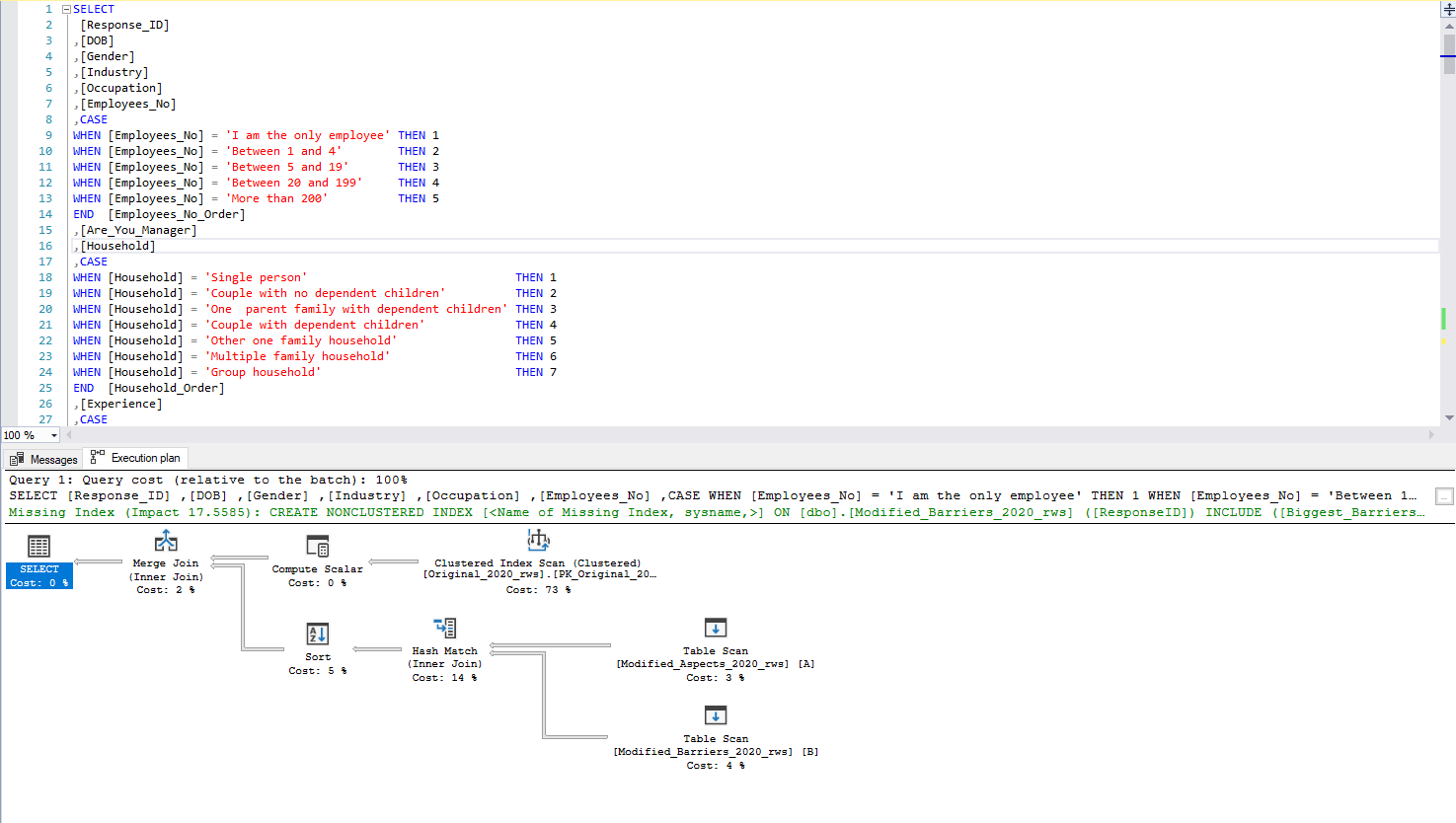
****

**6. Performance Optimization**

* **Indexes: Added indexes on key columns to improve query performance.**

****

* **Query Optimization: Optimized SQL queries by analyzing execution plans and removing any unnecessary subqueries or joins that could cause performance degradation.**

****

**7. Create Views/Stored Procedures Contains Transactions & Error Handling**

* **Views: Created views to simplify complex queries and provide pre-aggregated data for Tableau users, allowing them to access ready-to-use data.**

**Union the 2020 and 2021 remote work survey datasets into one table to conduct a comparative analysis across both years. However, the phrasing differences could introduce inconsistencies in the interpretation of the data, as "last 3 months" vs. "last 6 months" or "last year" vs. "this year" cover different timeframes.**

**To address this:**

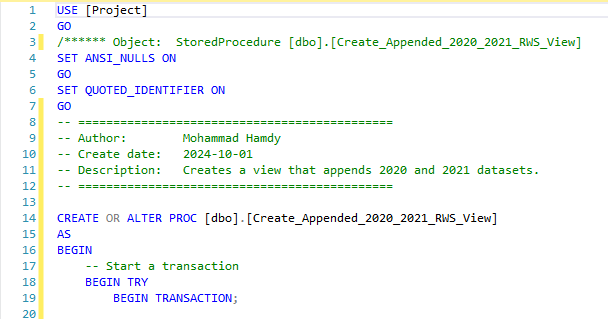
1. **I Added a Year Column: Include a column to identify the survey year (2020 or 2021) in the unified table.**
2. **I Standardized the Timeframes: You could either:**
   * **I Normalized the responses by recalculating metrics to reflect consistent time periods (e.g., converting 6 months to 3 months or yearly responses).**
   * **I Segmented the analysis based on year and timeframe (analyzing data from different periods separately).**
3. **I Annotated Question Variants: Track question differences in your documentation, so you can clearly explain any discrepancies in your analysis.**

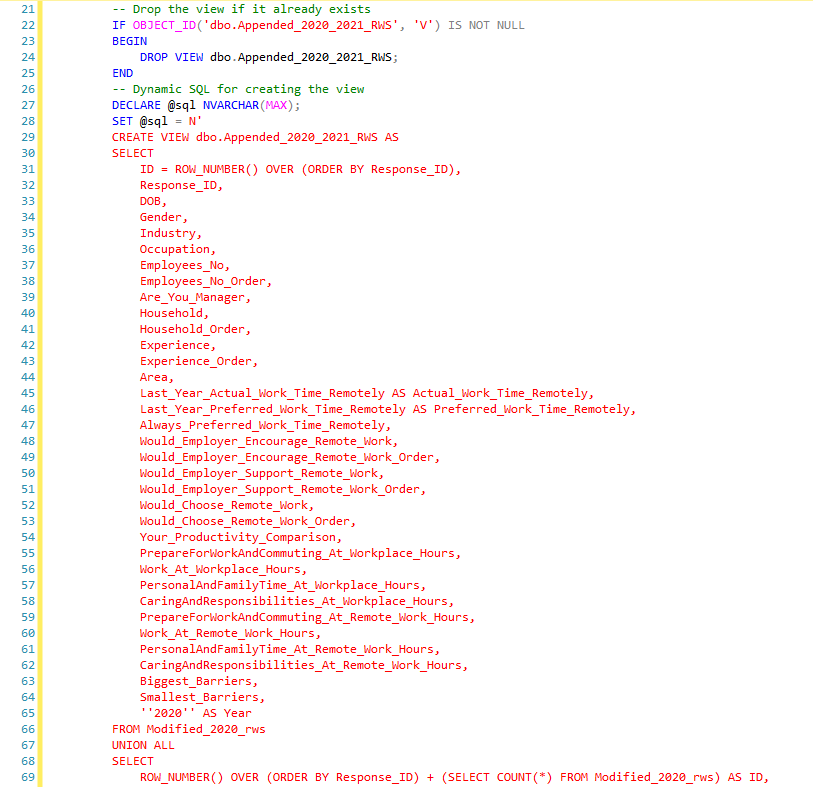
**This approach allowed me to union the datasets while keeping potential biases in check.**

* **Stored Procedures: Set up stored procedures for specific reporting purposes, allowing for automation and ease of access to specific data subsets, using Transactions and Error Handling in them, as they are crucial for ensuring data integrity and reliability in database operations. Transactions allow multiple operations to be treated as a single unit, ensuring that either all changes are applied or none, preventing partial updates in case of failures.**

**Error handling ensures that unexpected issues are caught and managed gracefully, allowing the system to roll back changes and maintain consistency.**

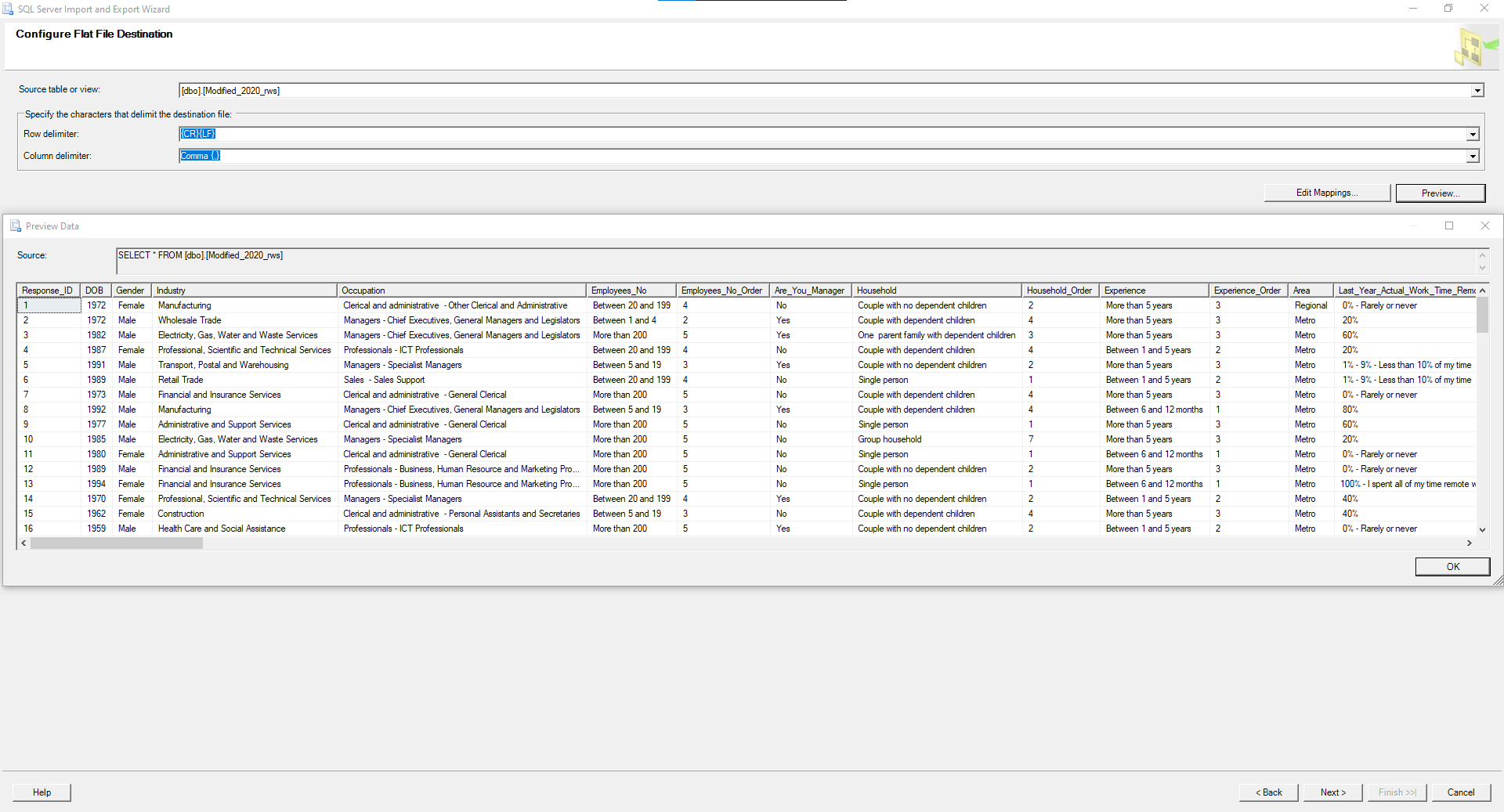
**Together, they enhance robustness, reduce the risk of data corruption, and improve system reliability.**

****

****

**8. Export Dataset for Tableau**

* **Export as CSV: Exported the prepared dataset to CSV files that can be easily imported into Tableau.**

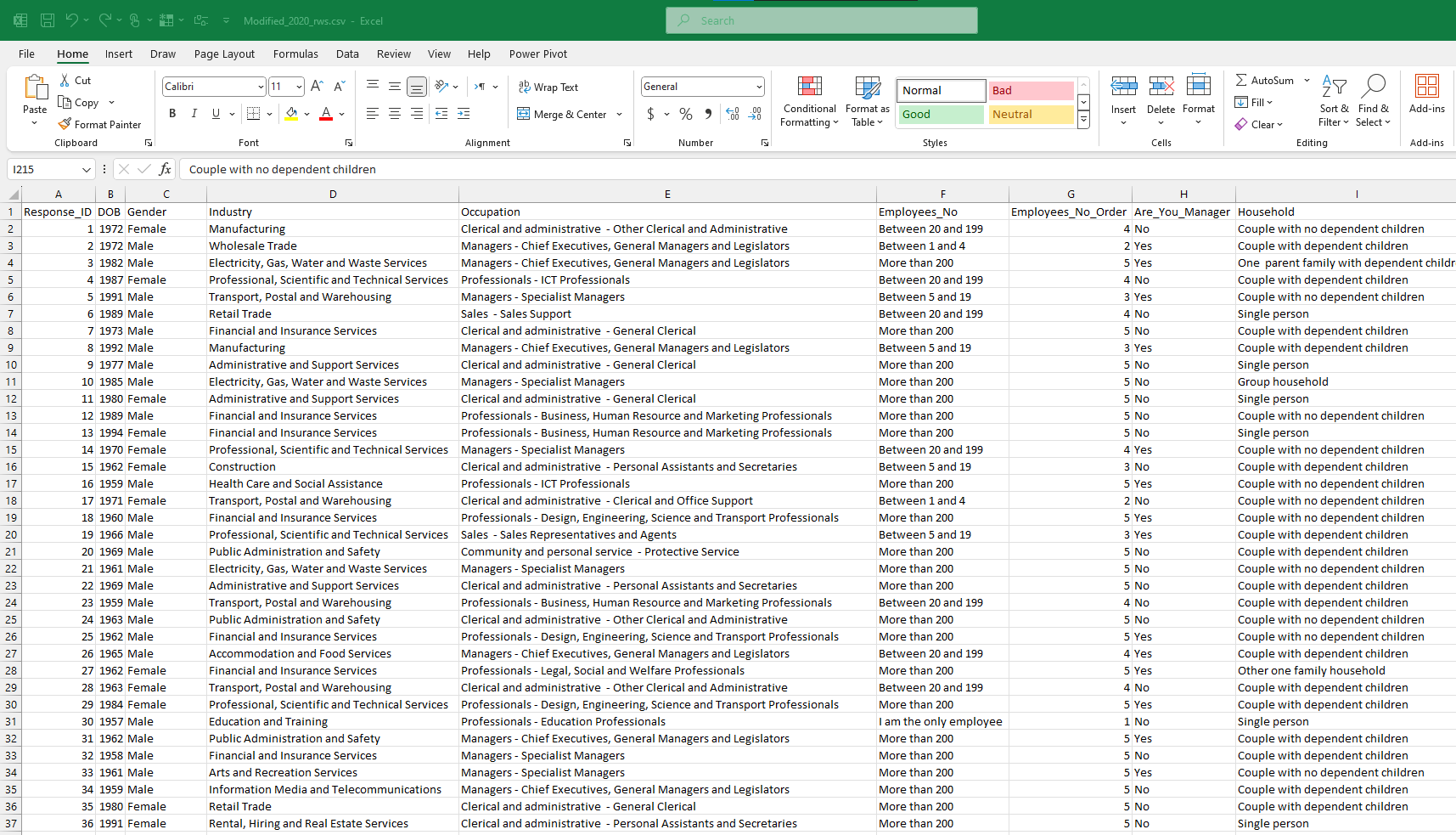
****

* **Direct Connection Setup: Alternatively, set up a direct connection to SQL Server for Tableau, allowing colleagues to use live queries. So, I made a database backup to give it to my teammates to work on.**

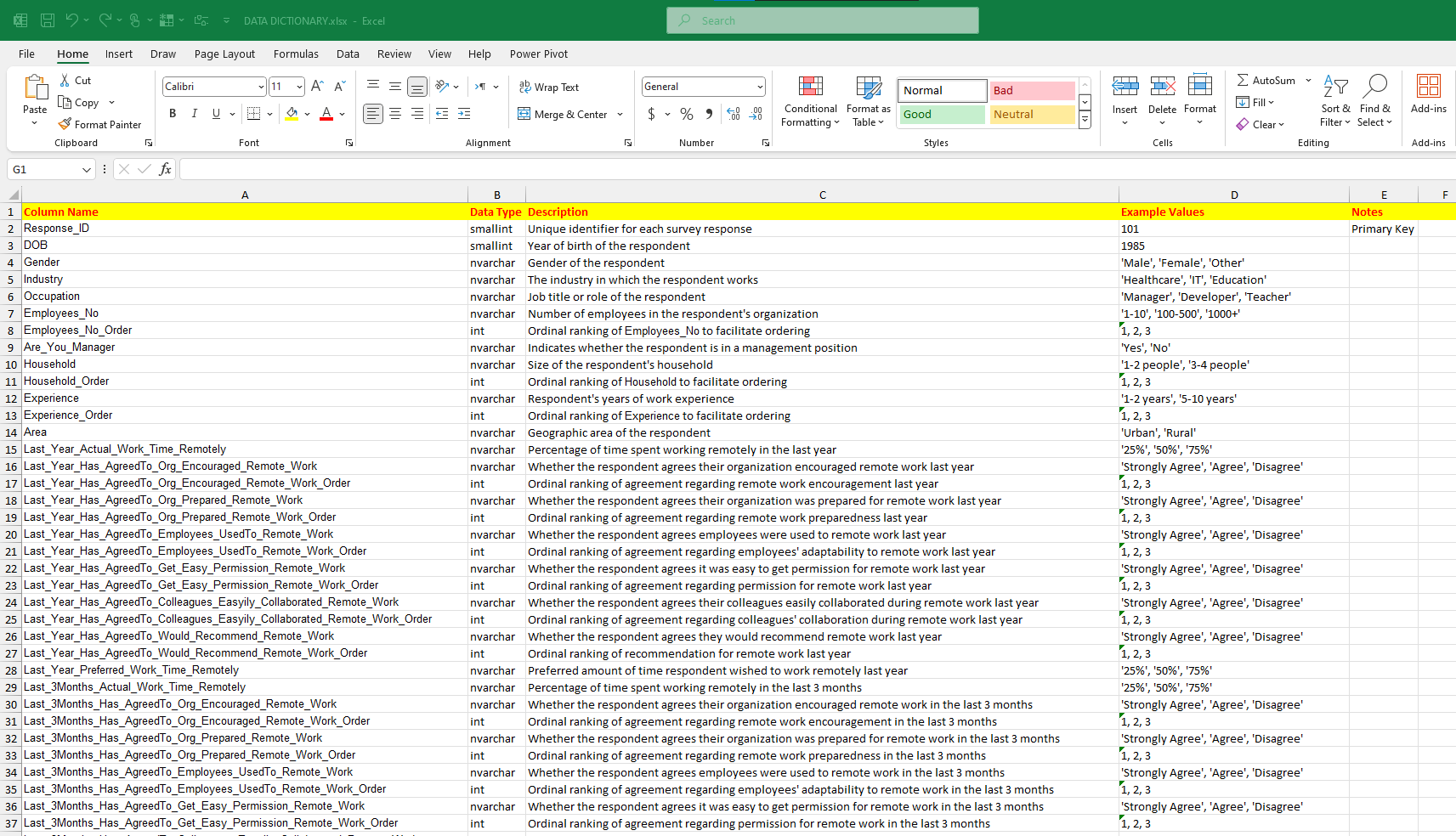
****

**9. Documentation**

* **Document Data Model: Provided detailed documentation outlining the data sources, transformations, and key fields in the dataset to ensure colleagues understand the structure when using it in Tableau.**

****

* **Data Dictionary: Created a data dictionary to explain the columns, relationships, and data definitions.**

****

**By completing these steps, my datasets are now clean, organized, and ready for analysis in Tableau.**