Task 4 - Inclusion Dependency Discovery

1. Reading the CSV data and storing each table
2. Creating a flattened column data frame (key, value)
   - The key column contains the attribute values of each table
   - The value column contains the column name of the attribute
   - This means that each row contains: (attribute value, column name)
3. Group by the keys (attribute values) and aggregate the columns to a set
4. Then use the explode function on the column set
   - This will add a row for each entry of each column set
   - So each column is associated with the rest of the column set
5. Use groupByKey and mapGroup to create the inclusion list
   - The grouByKey is done on the exploded values
   - Map each column name to the inclusion list if one exists
     - Intersect all sets of the grouped column

Task 4 - Inclusion Dependency Discovery

6. Filter out the rows that contain null values
7. Sort the results
8. Print them in the terminal