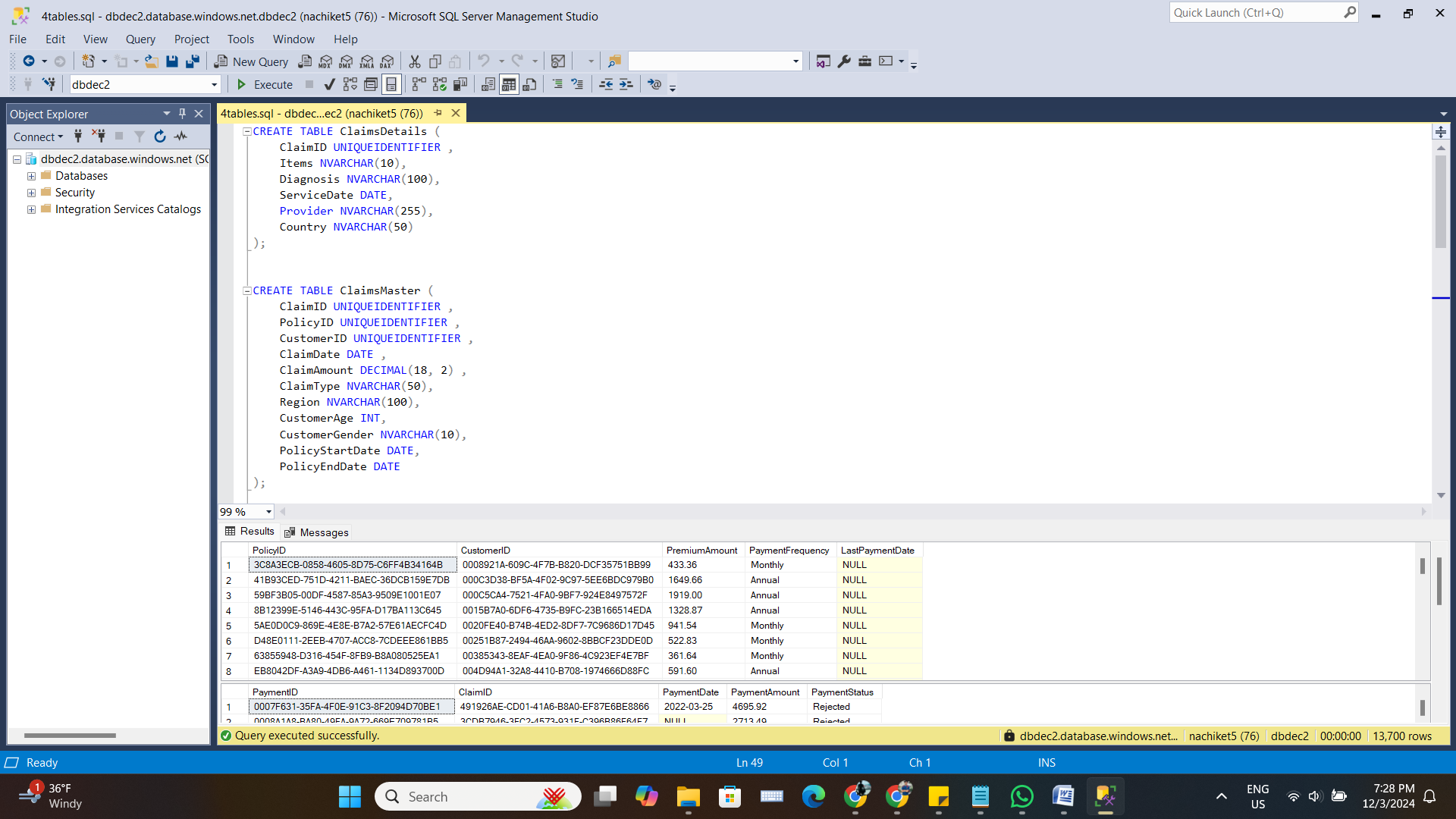
**Clean Duplicate Data and Null Records from SQL Database Table**

**Objective**  
The goal of this project is to clean a given SQL database table by identifying and separating duplicate and null records into distinct categories: good records and bad records. The cleaned data will then be exported to two separate CSV files.

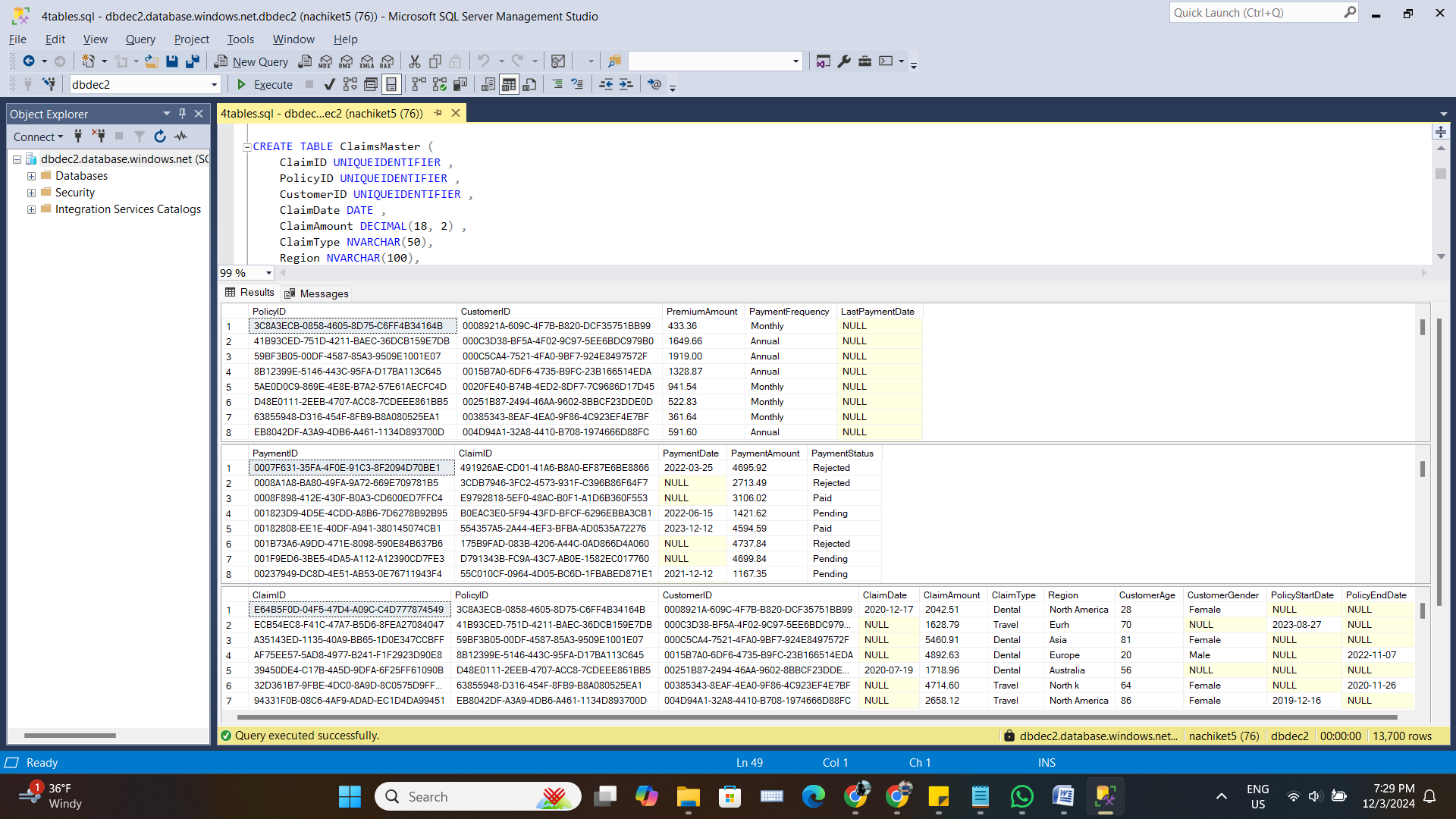
#### **Instructions**

1. **Analyze the SQL Database Table**
   * Inspect the given table and identify columns with duplicate and null values.

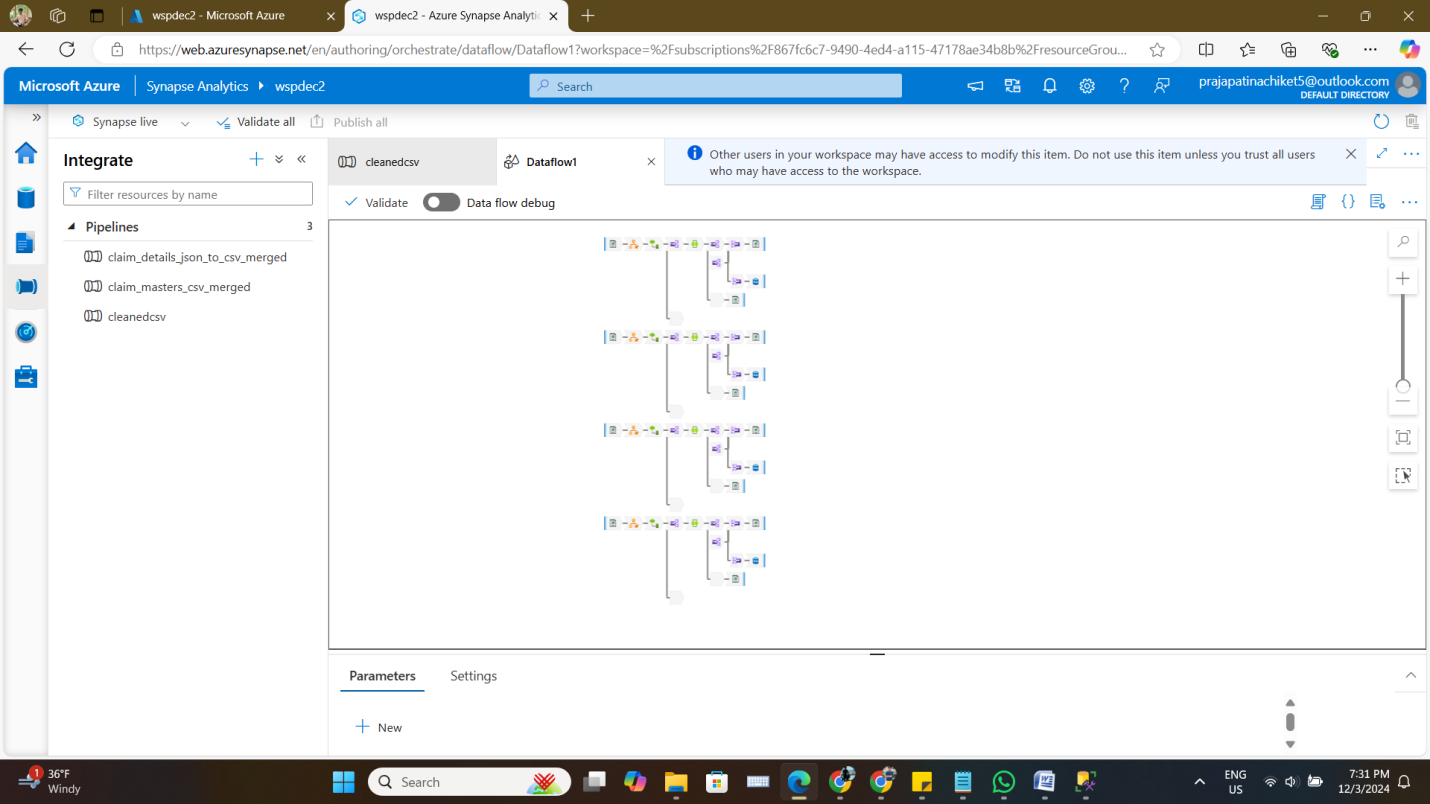


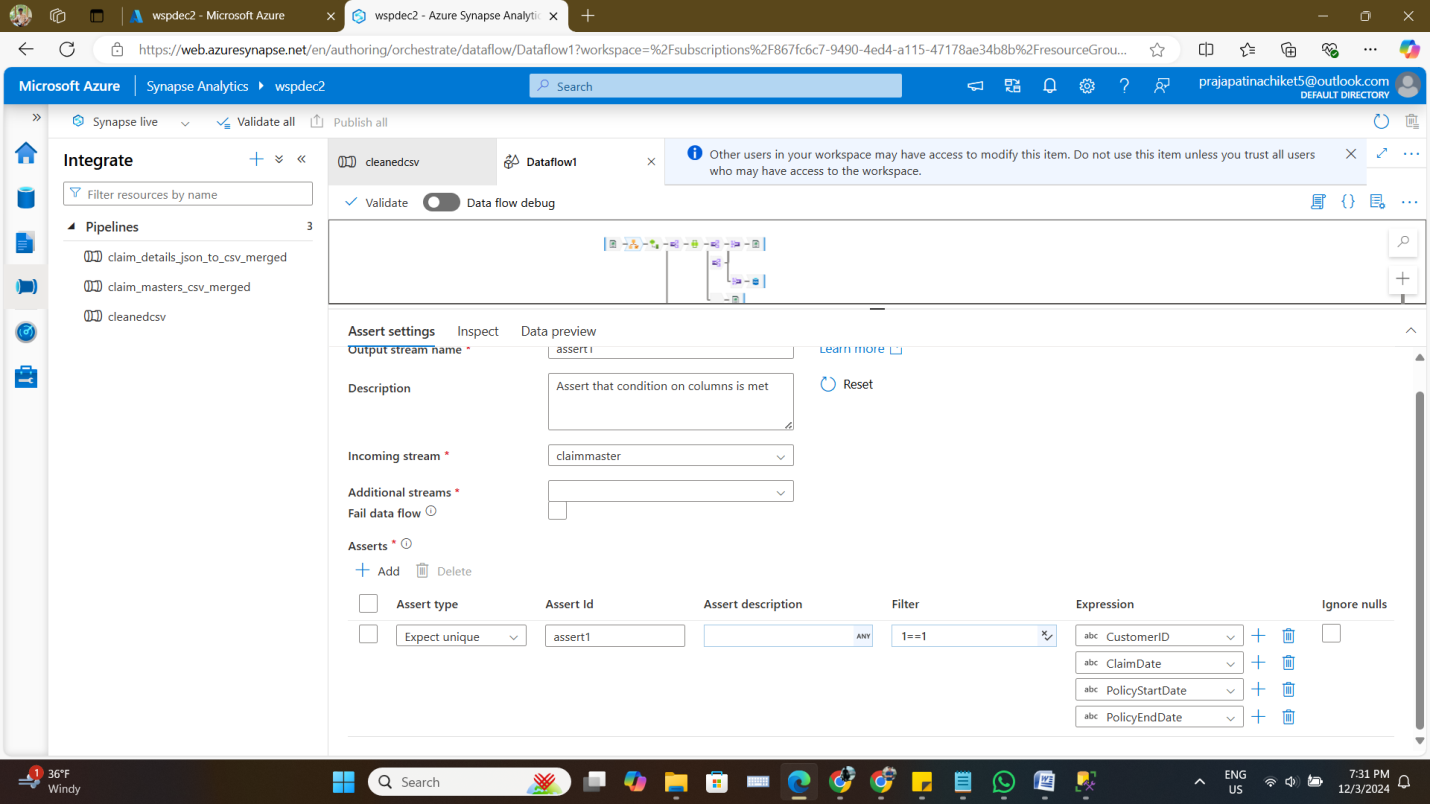
* + Understand the schema and determine which columns are critical for identifying good records.

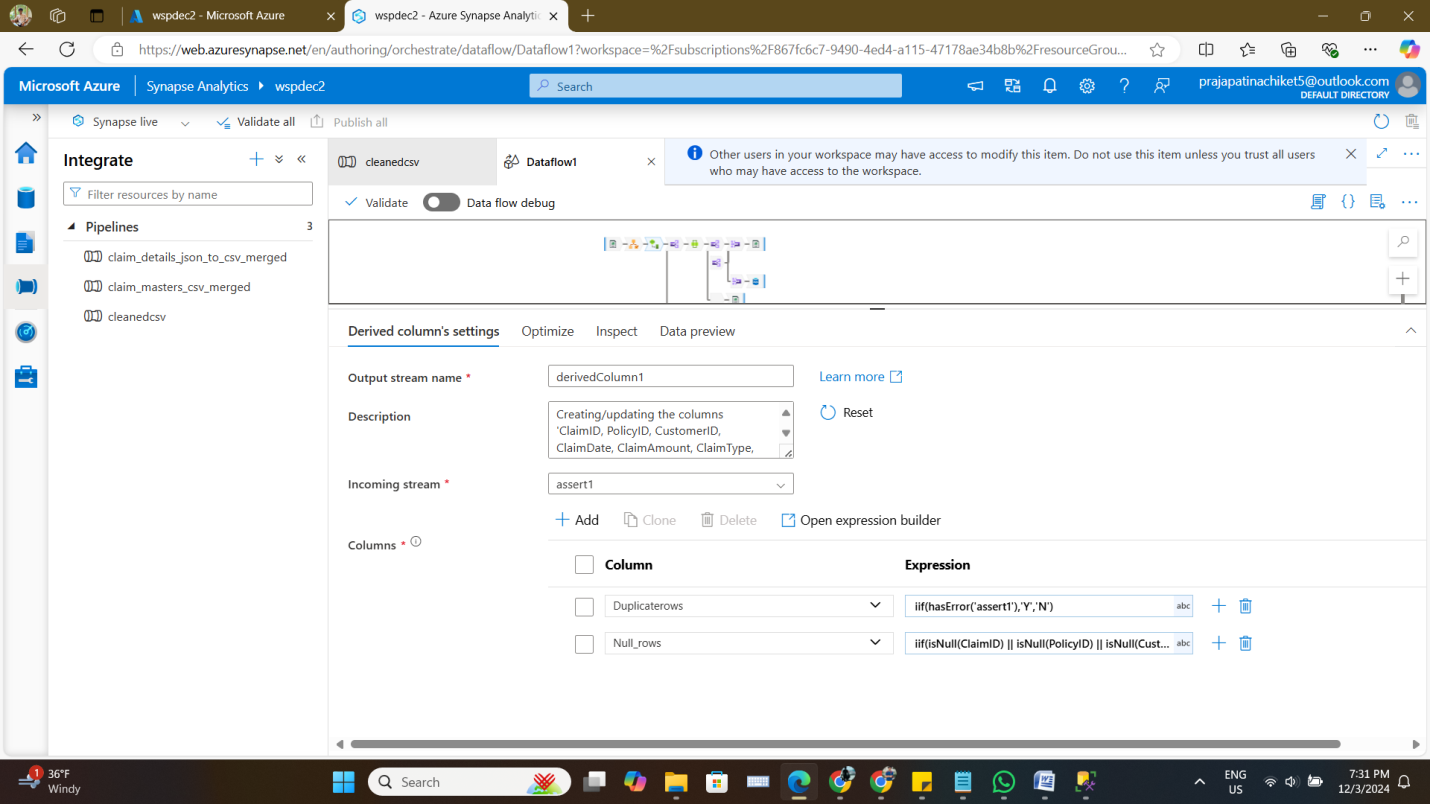
For analysis dates are important for my final business growth insights so i should not use validdate column to remove nulls from date columns.



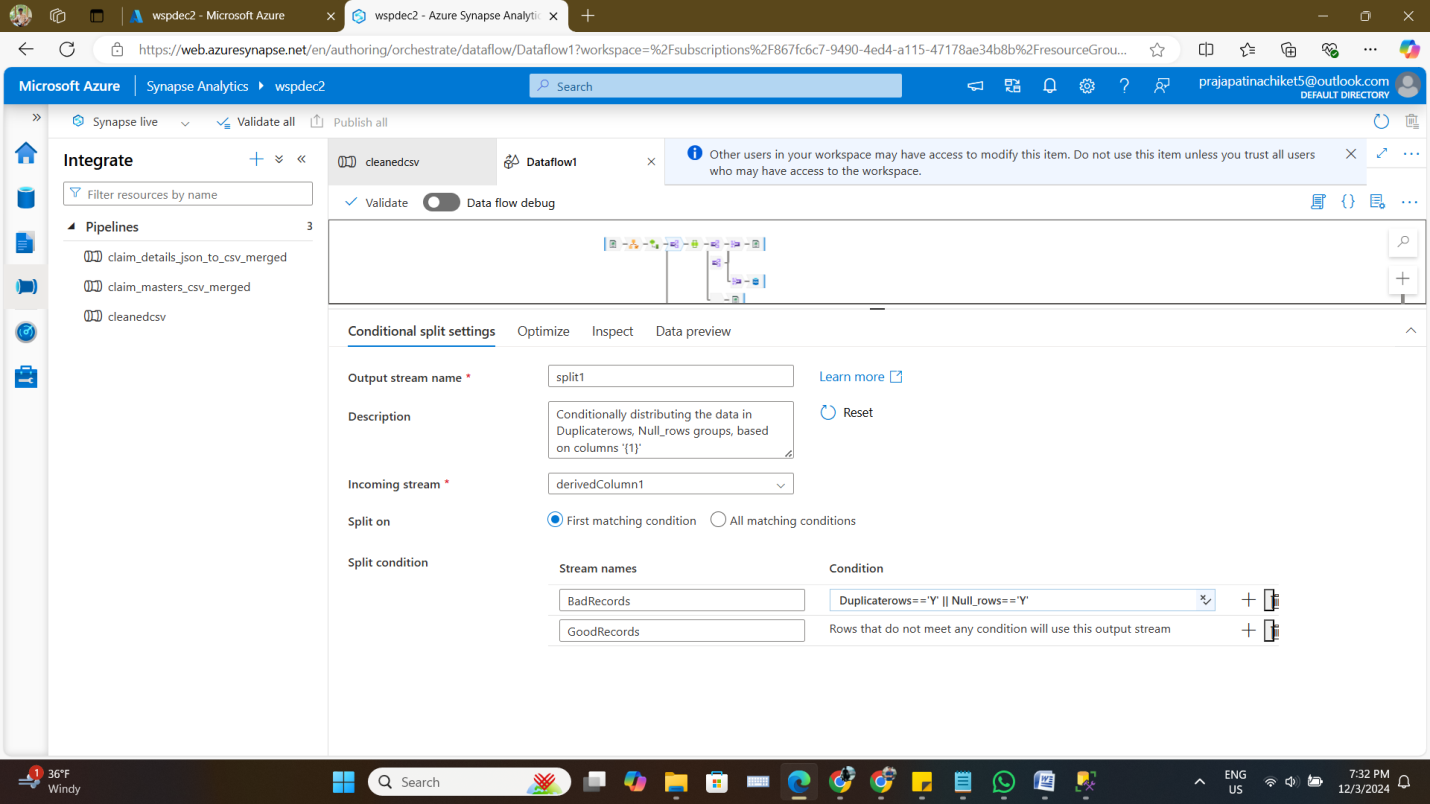
1. **Perform Data Cleaning**
   * Remove duplicate records from the dataset based on specified columns.





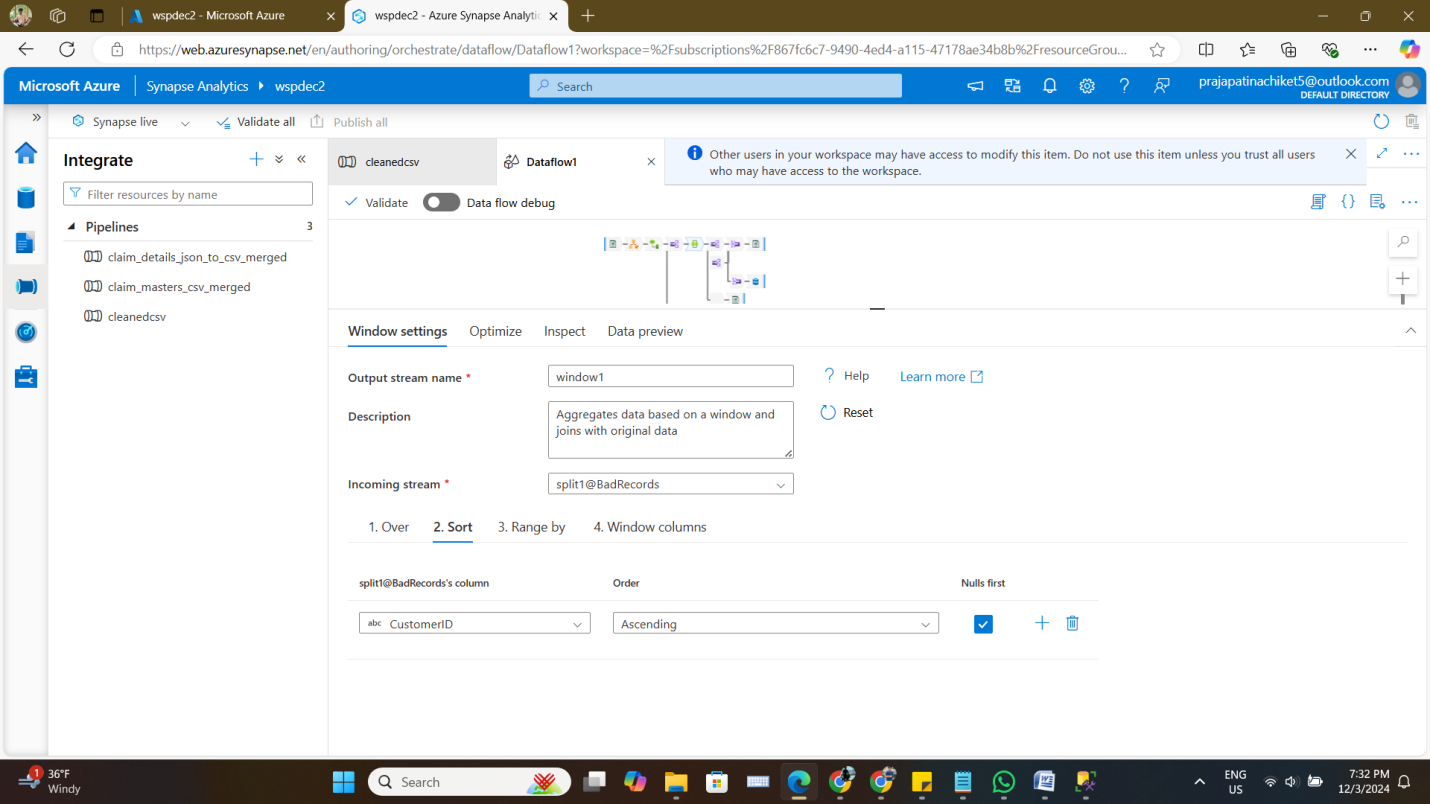


* + Identify and segregate records containing NULL values in any of the critical columns.

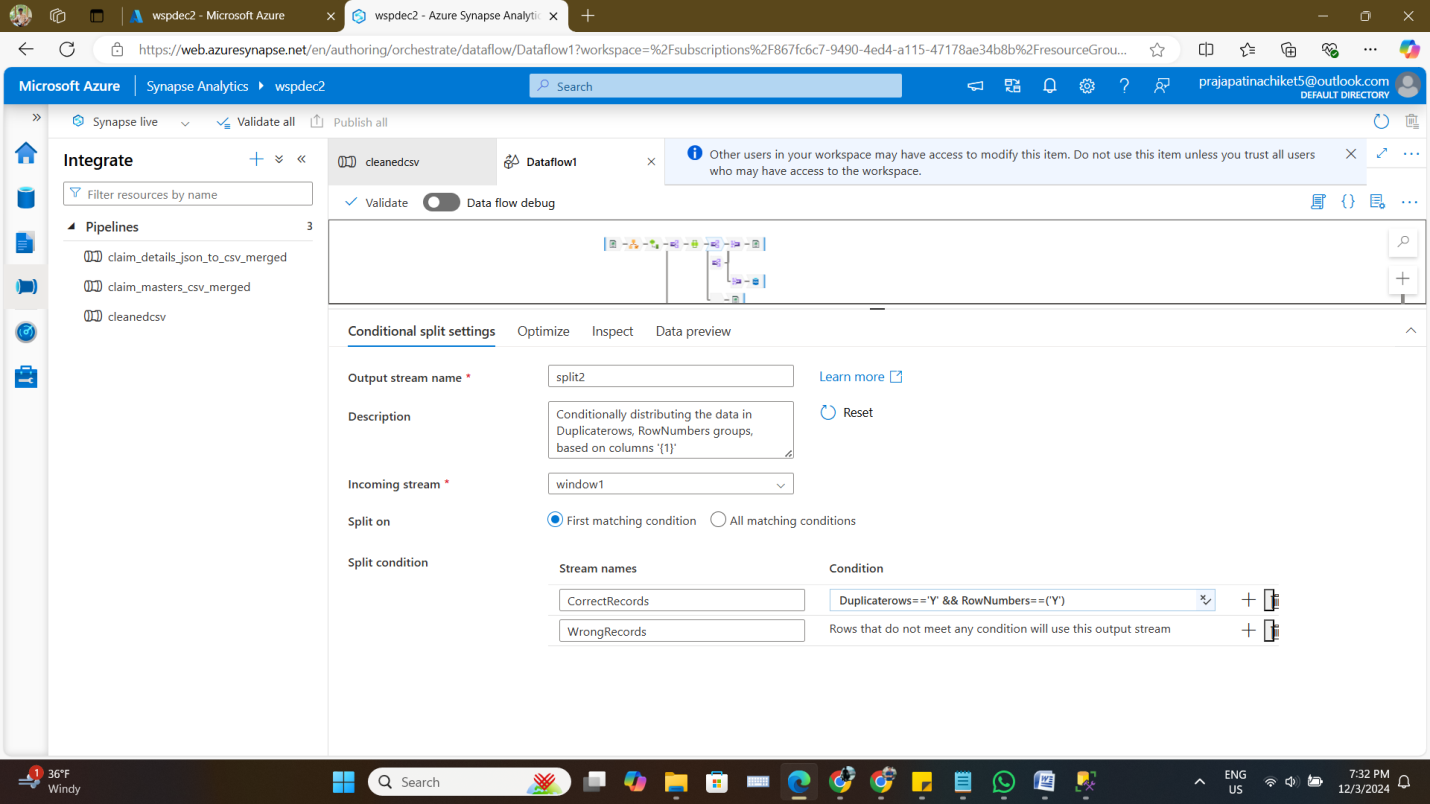
****

**Window**

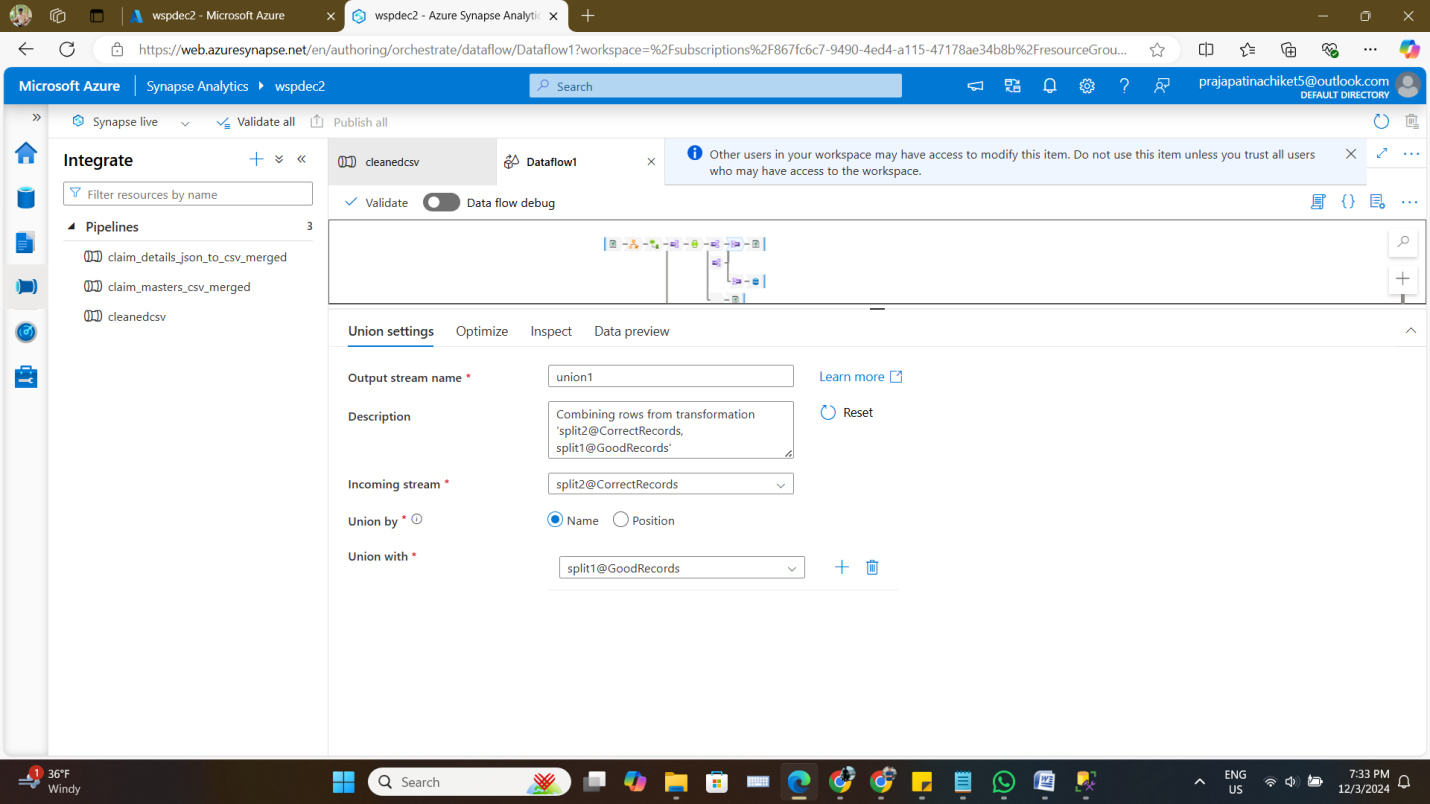
****

****

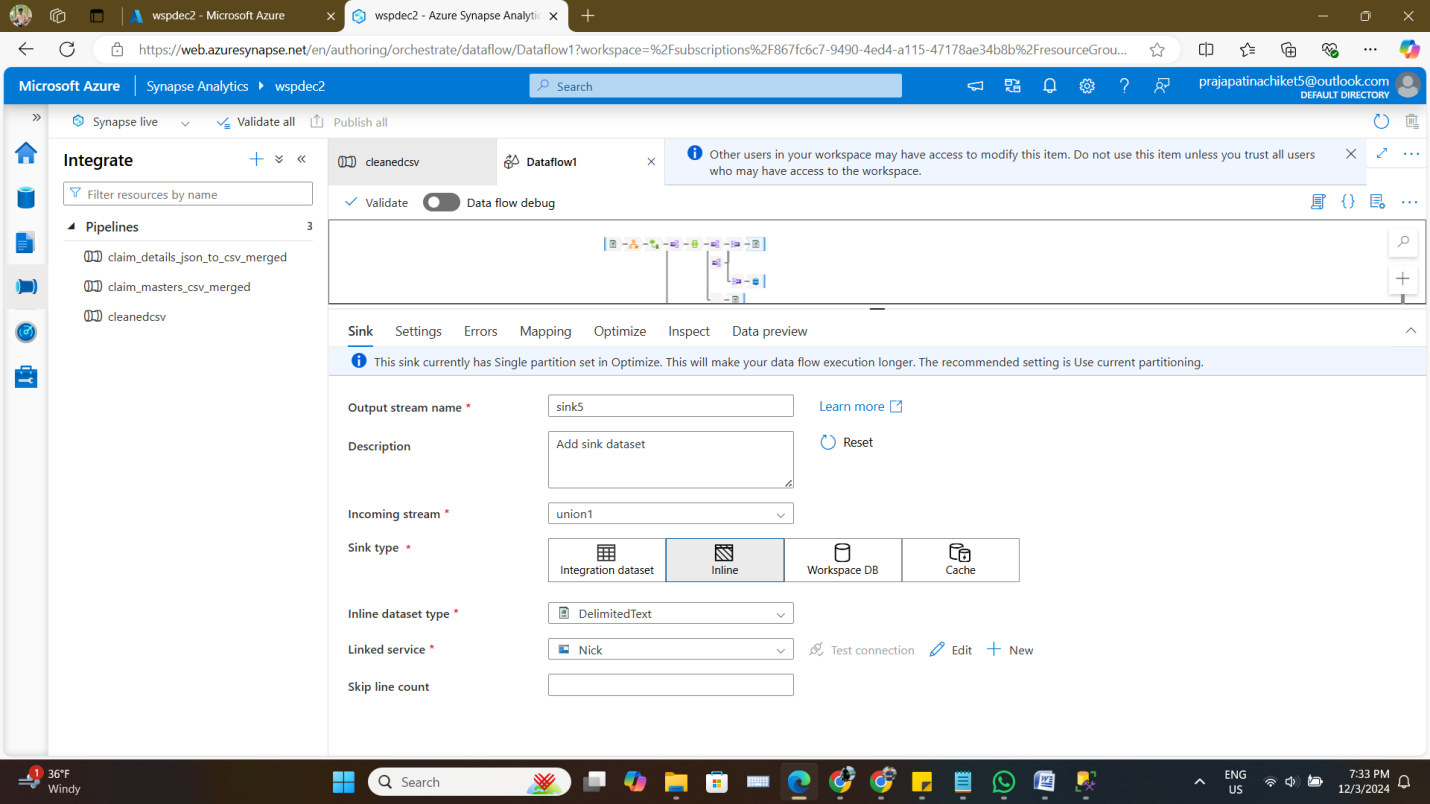
****

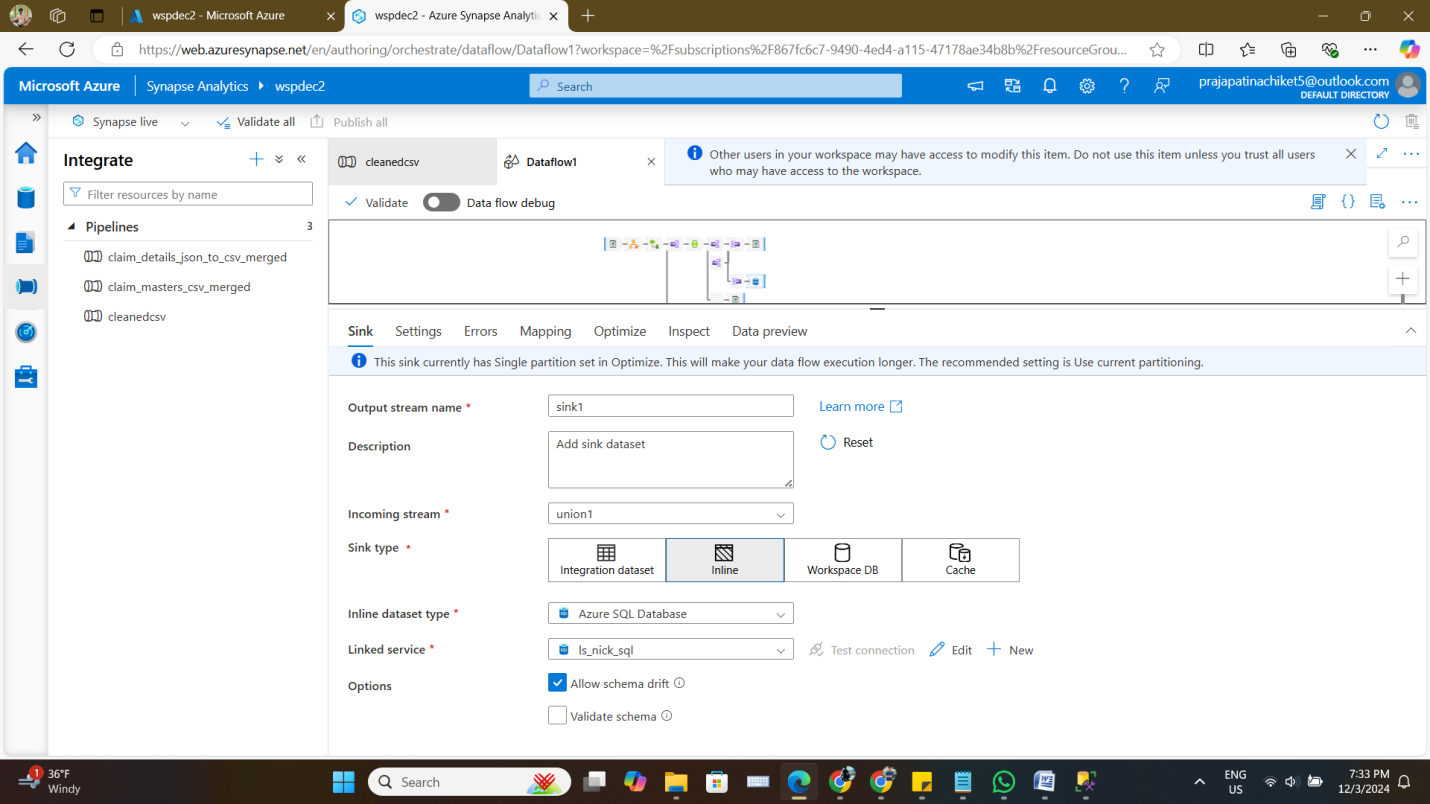
****

**Union**

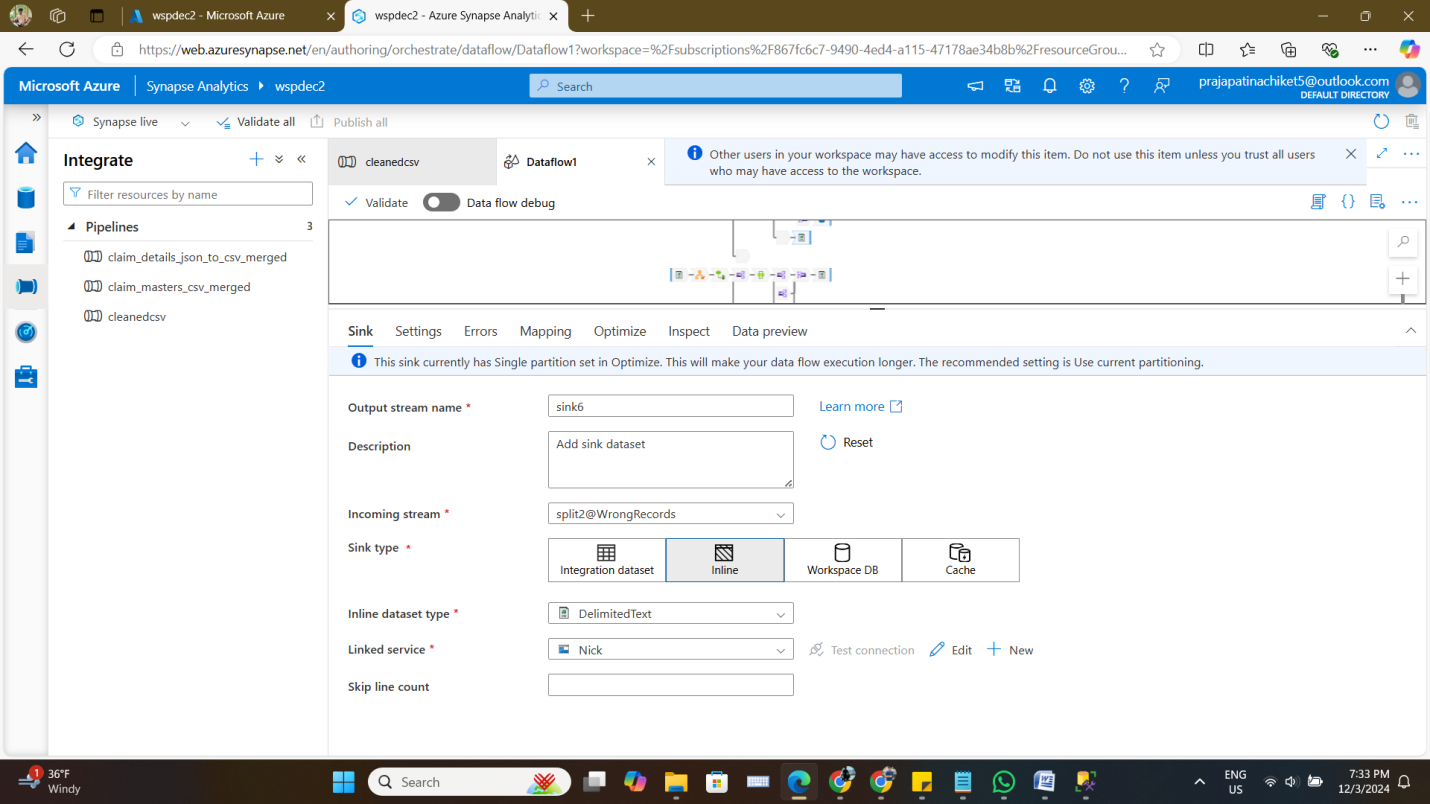
****

1. **Export Cleaned Data**
   * Export all good records (cleaned and valid) into a good\_records.csv file.

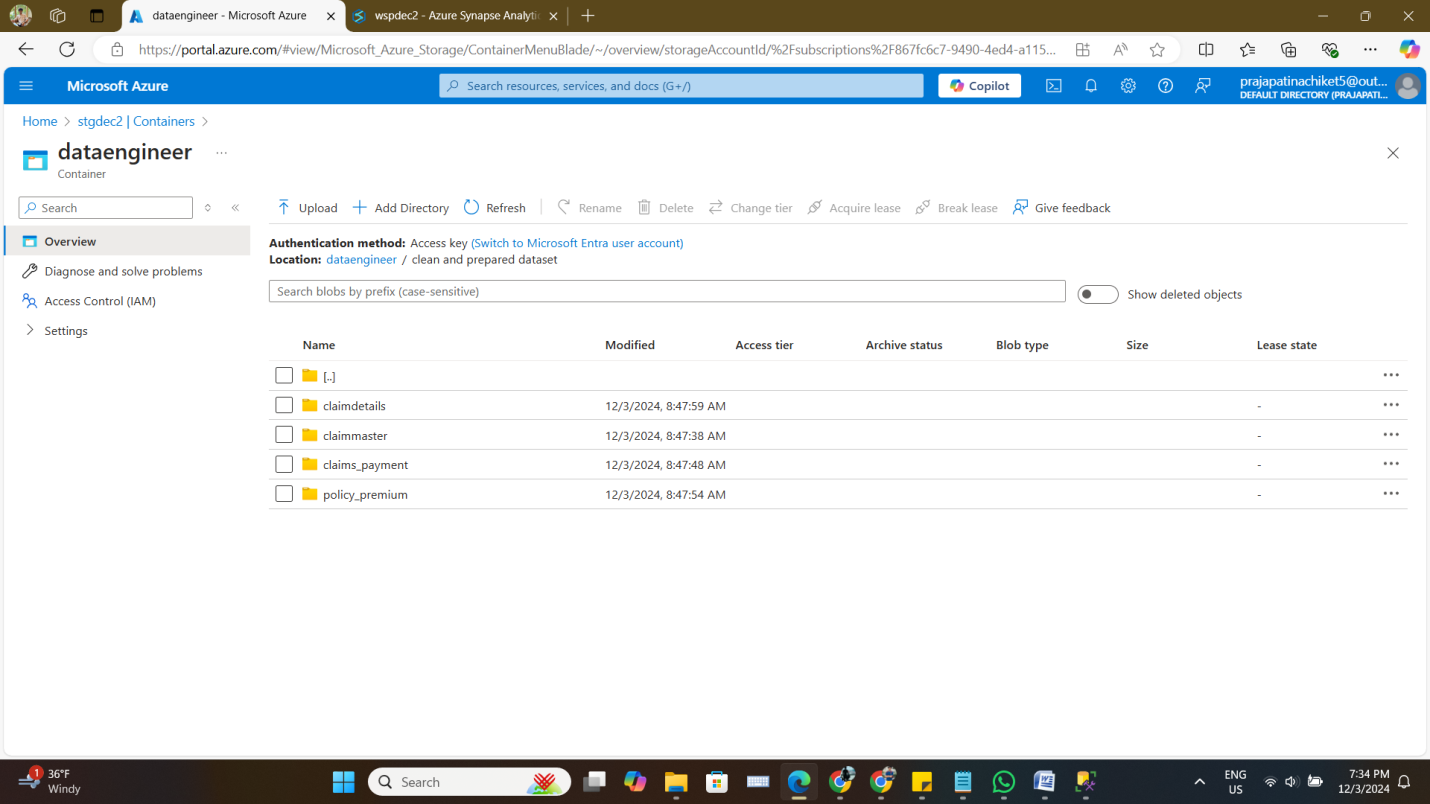


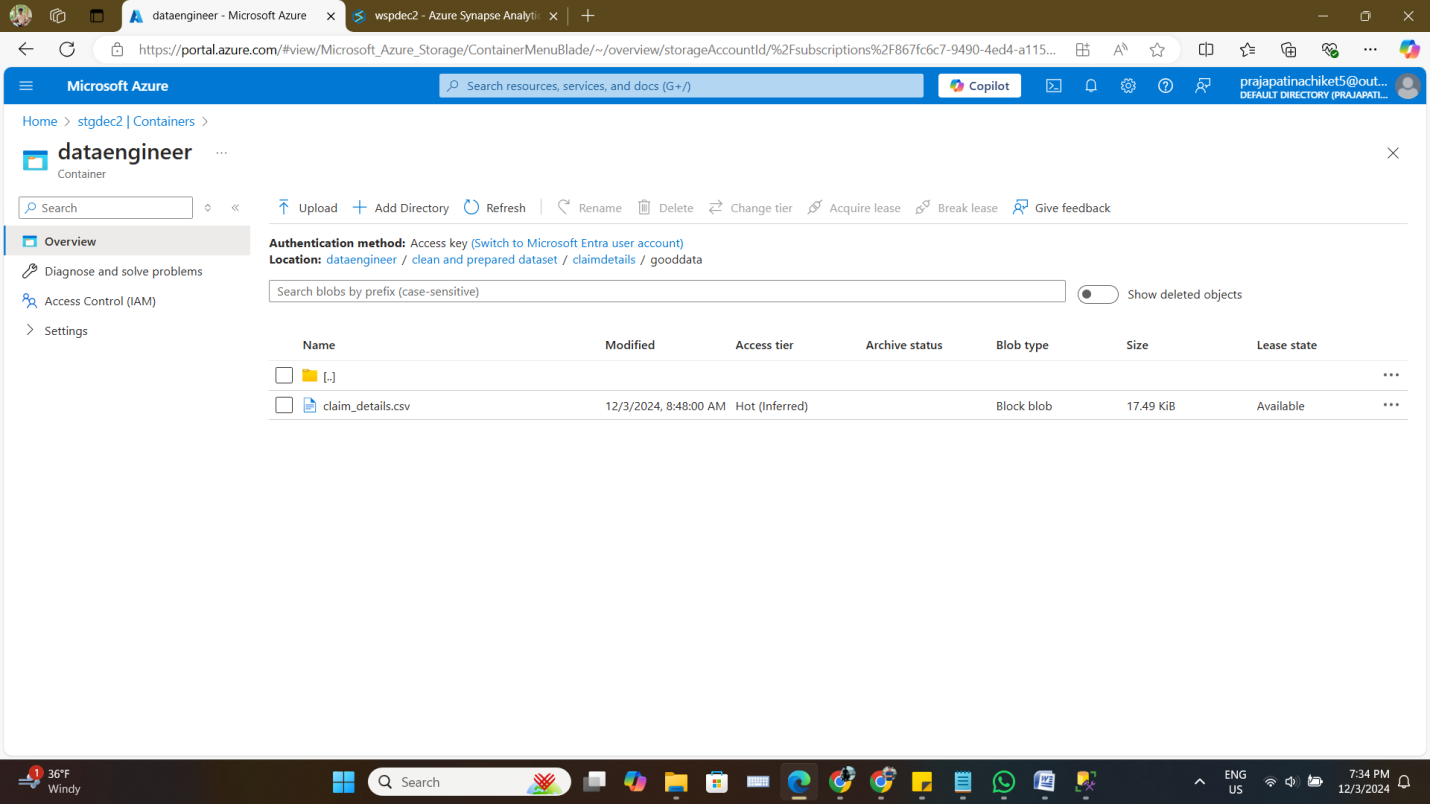


* + Export all bad records (duplicates and records with null values) into a bad\_records.csv file.

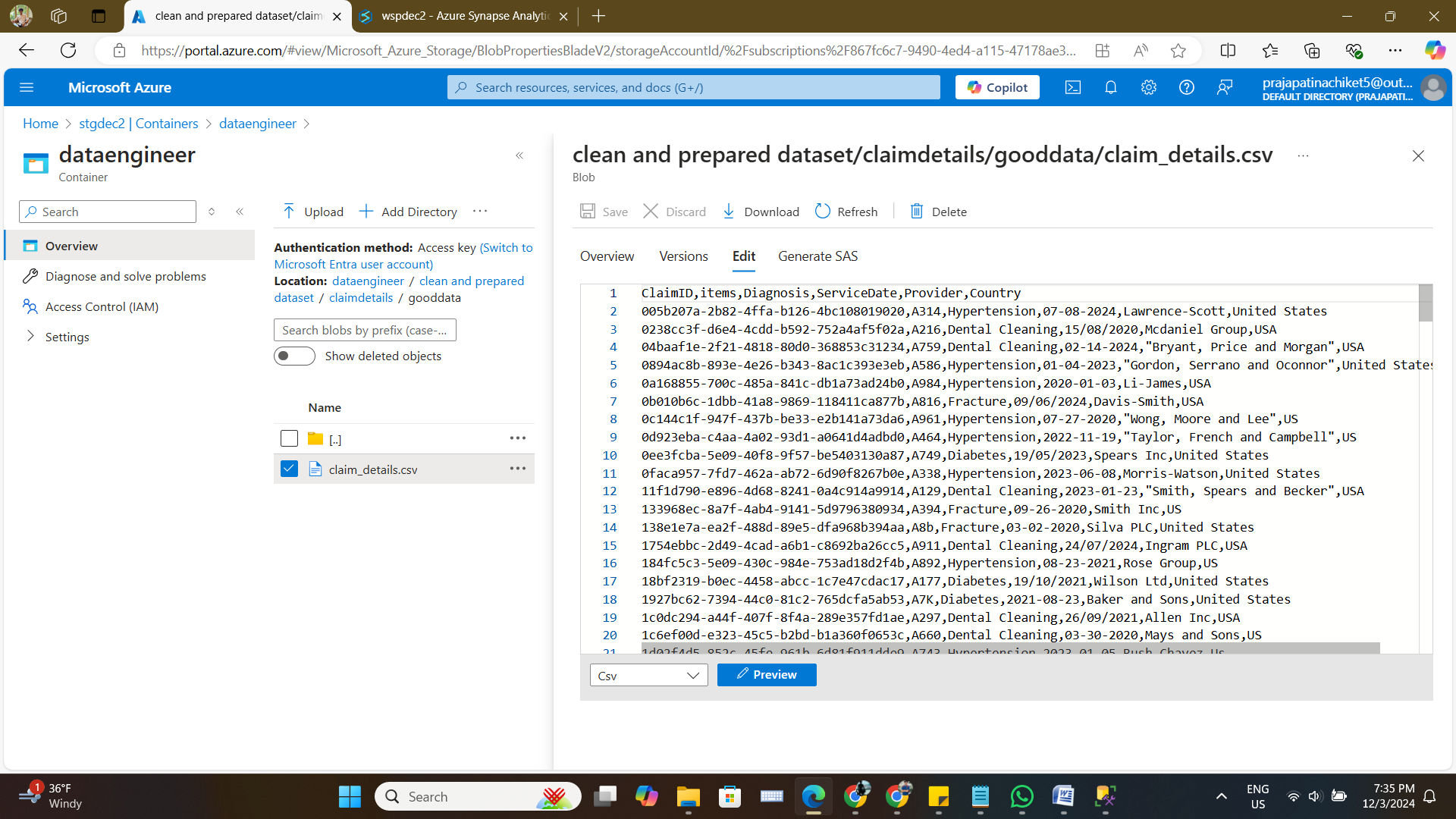


1. **Verify Output**
   * Validate the CSV files to ensure correct segregation and cleaning of records.



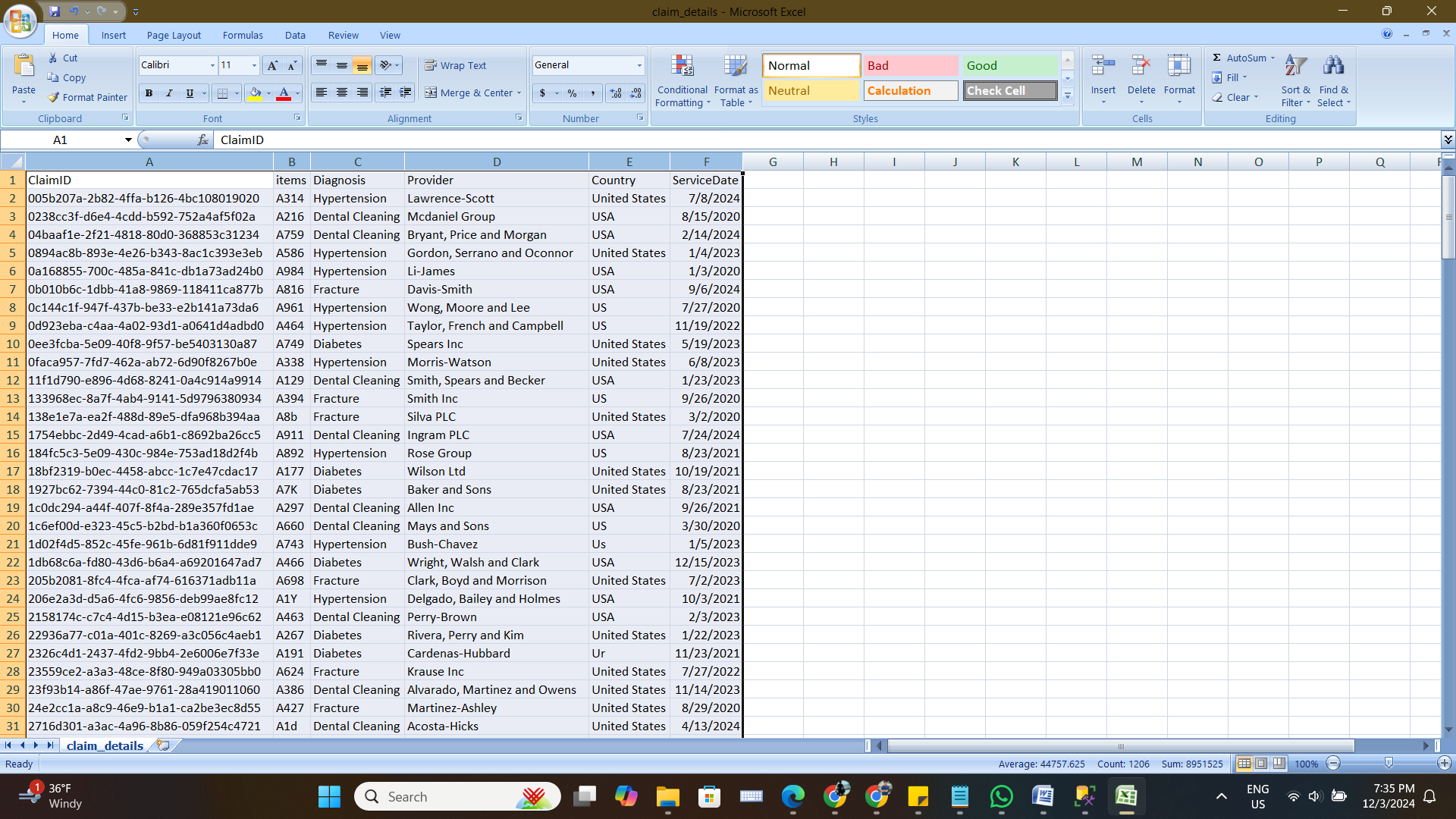


* + Ensure the good records meet all the criteria for completeness and uniqueness.

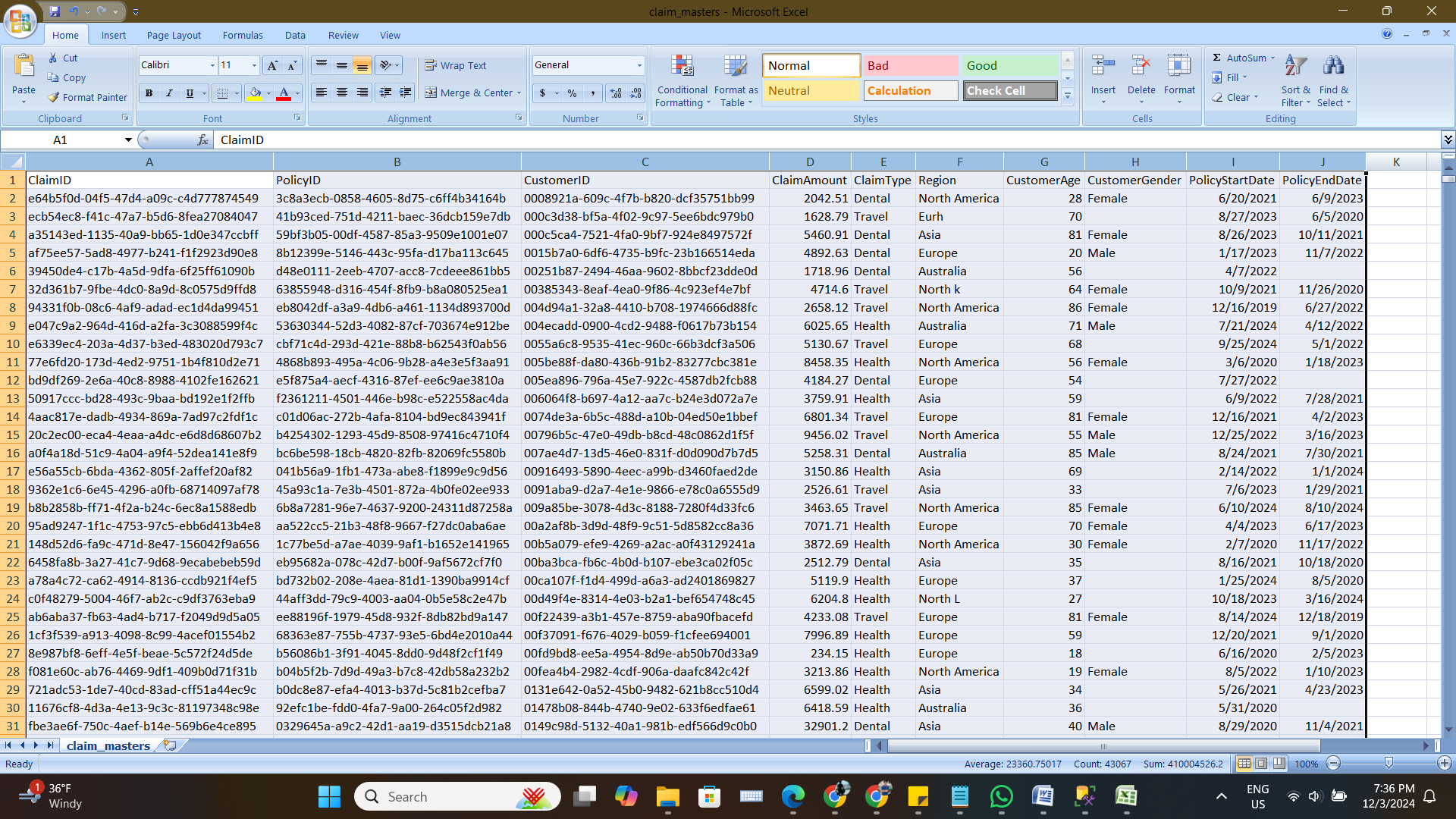


#### **Expected Output**

1. **good\_records.csv**
   * Contains all cleaned records without duplicates or null values.



* + Each record should represent valid and unique data.



1. **bad\_records.csv**
   * Contains all records flagged as duplicates or containing null values.



* + Each record should indicate why it was classified as a bad record (e.g., Duplicate, Null Value).

It is not showing because of validdate column, in the end i will need that date column for analyzing five objectives for business ingsights