LAB 05 - Thực hành Power BI

Phan Tấn Tài – 22684181 – DHKHDL18A

1. Chọn một bộ dữ liệu bất kỳ từ internet.

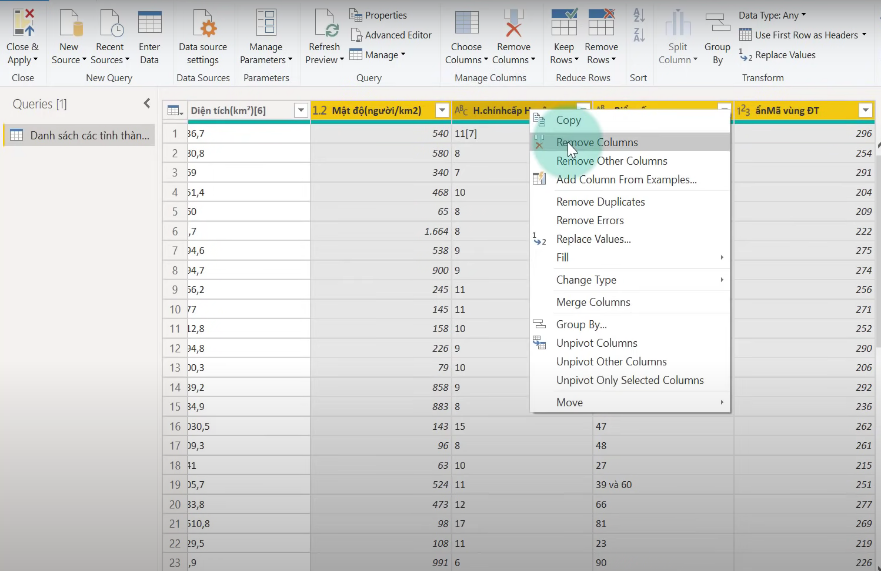
- Mô tả dữ liệu mà bạn có.

- Mô tả mục tiêu nghiên cứu của bạn với bộ dữ liệu

2. Sử dụng Power BI kết nối dữ liệu (từ file)

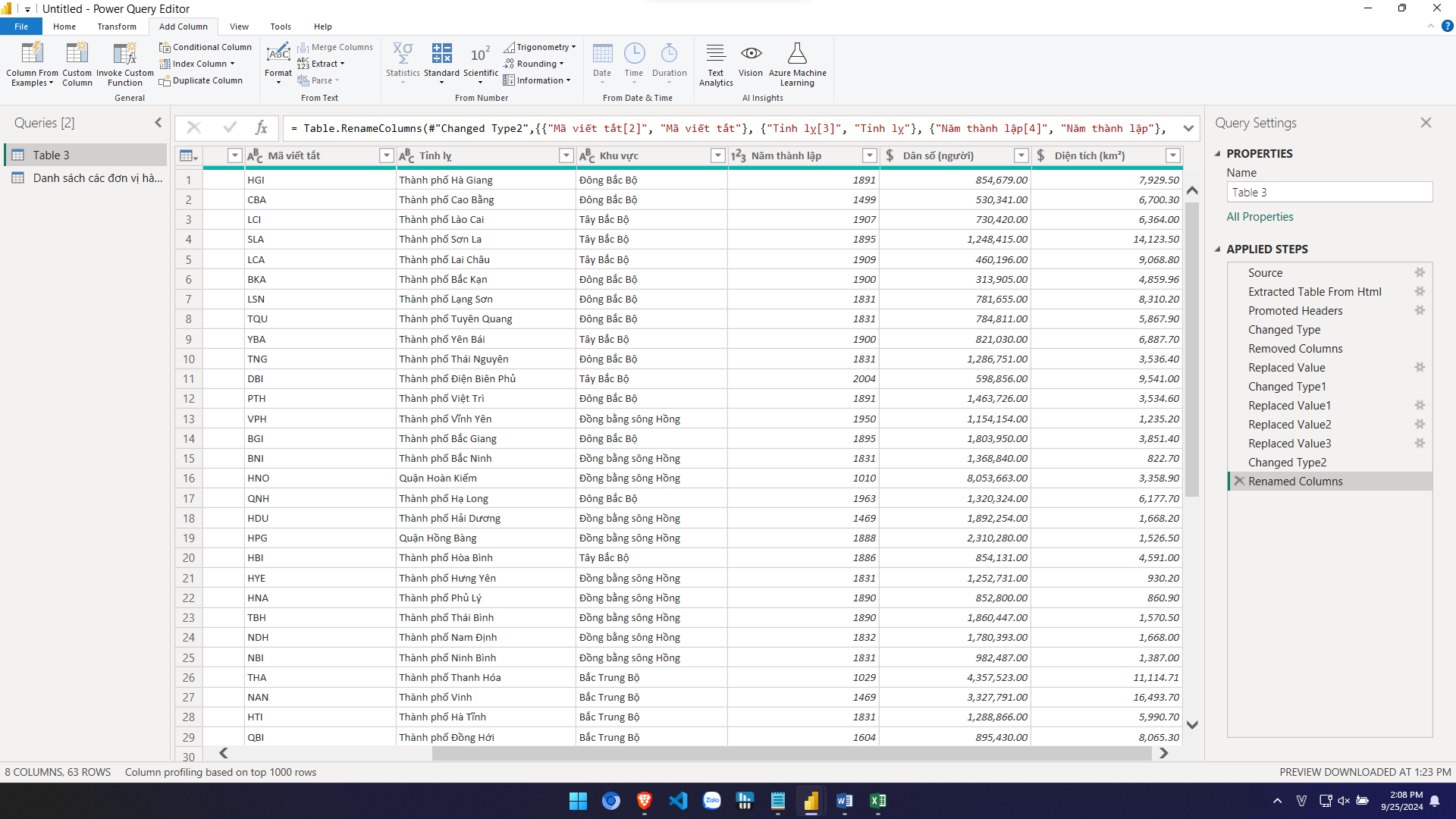
3. Biến đổi hoặc làm sạch dữ liệu với Power BI. Xử lý khi

⦁ Table “Tỉnh\_thành\_Việt\_Nam”

- Tiến hành xoá các cột *“Mật độ(người/km2)”, “H.Chính cấp huyện”, “Biển số xe”, “ẩn Mã vùng ĐT”* vì các cột dữ liệu này không quan trọng, không cần thiết với nhu cầu tạo Dashboard về GDP của các tỉnh thành Việt Nam

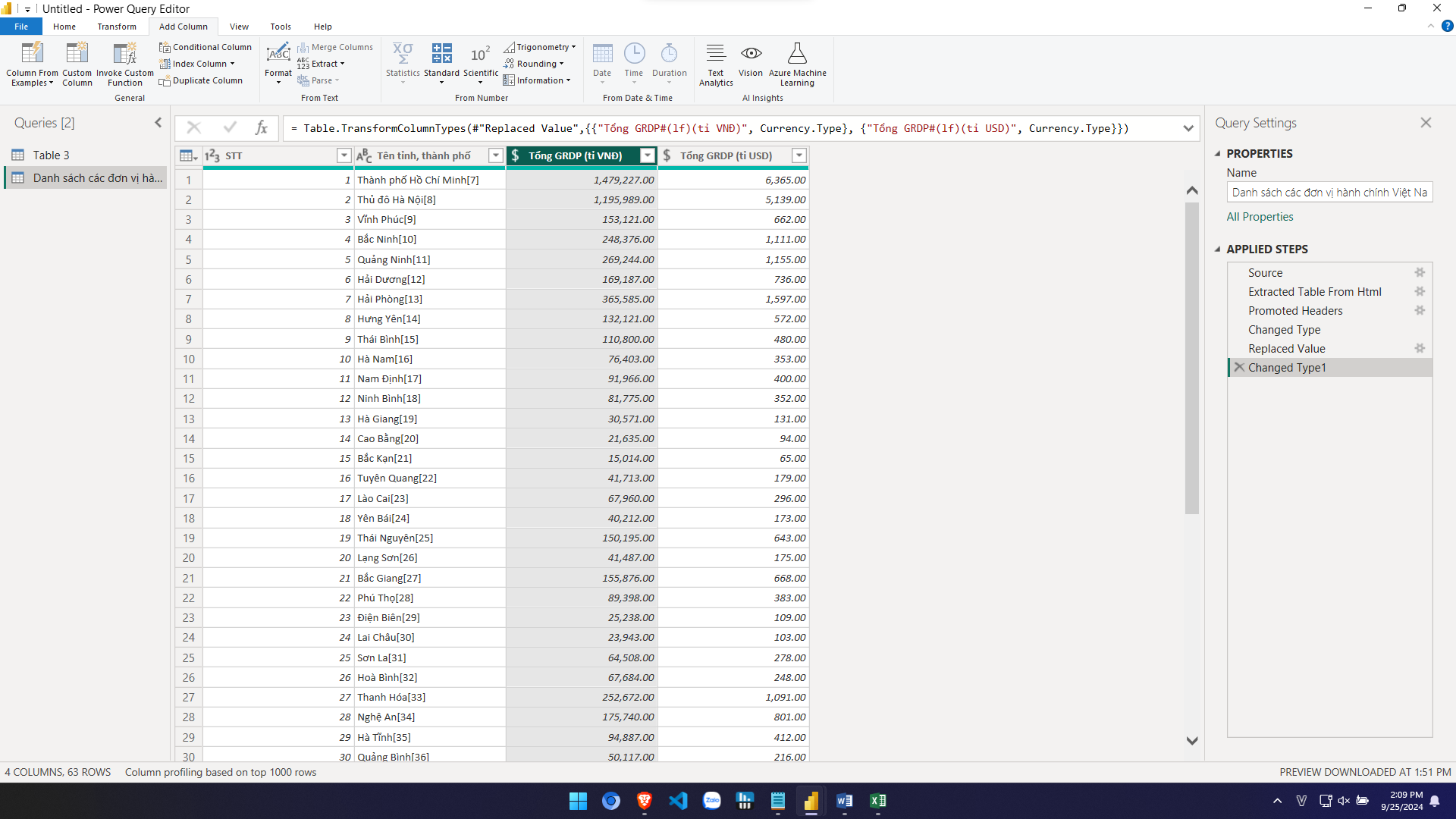
- Để tránh gây sai sót trong quá trình tạo biểu đổ và Dashboard ta cần chuyển dấu *“ . ”* Trong dữ liệu cột *“Dân số(Người)[5]”*  và cột *“Diện tích (km²)[6]”* thành dấu *“ , ”* bởi vì PowerBI phân biệt hàng đơn vị số bằng dấu *“ , ”* chứ không là dấu *“ . ”*.

- Chuyển các định dạng dữ liệu của 2 cột *“Dân số(Người)[5]”*  và cột *“Diện tích (km²)[6]”* từ *text* về *fixed demical number*



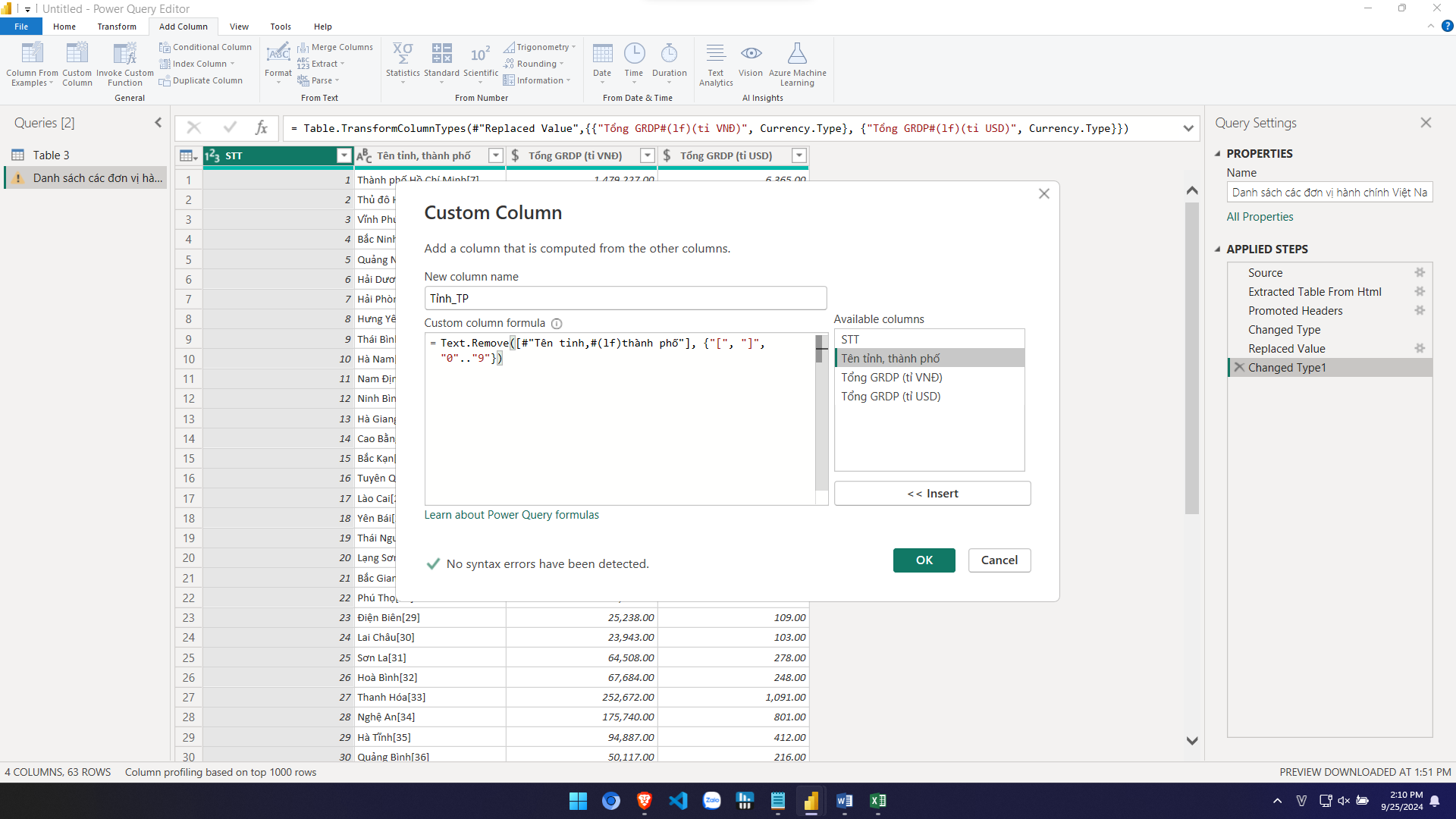
⦁ Table “Danh\_sách\_đơn\_vị\_hành\_chính\_Việt\_Nam\_theo\_GRDP”

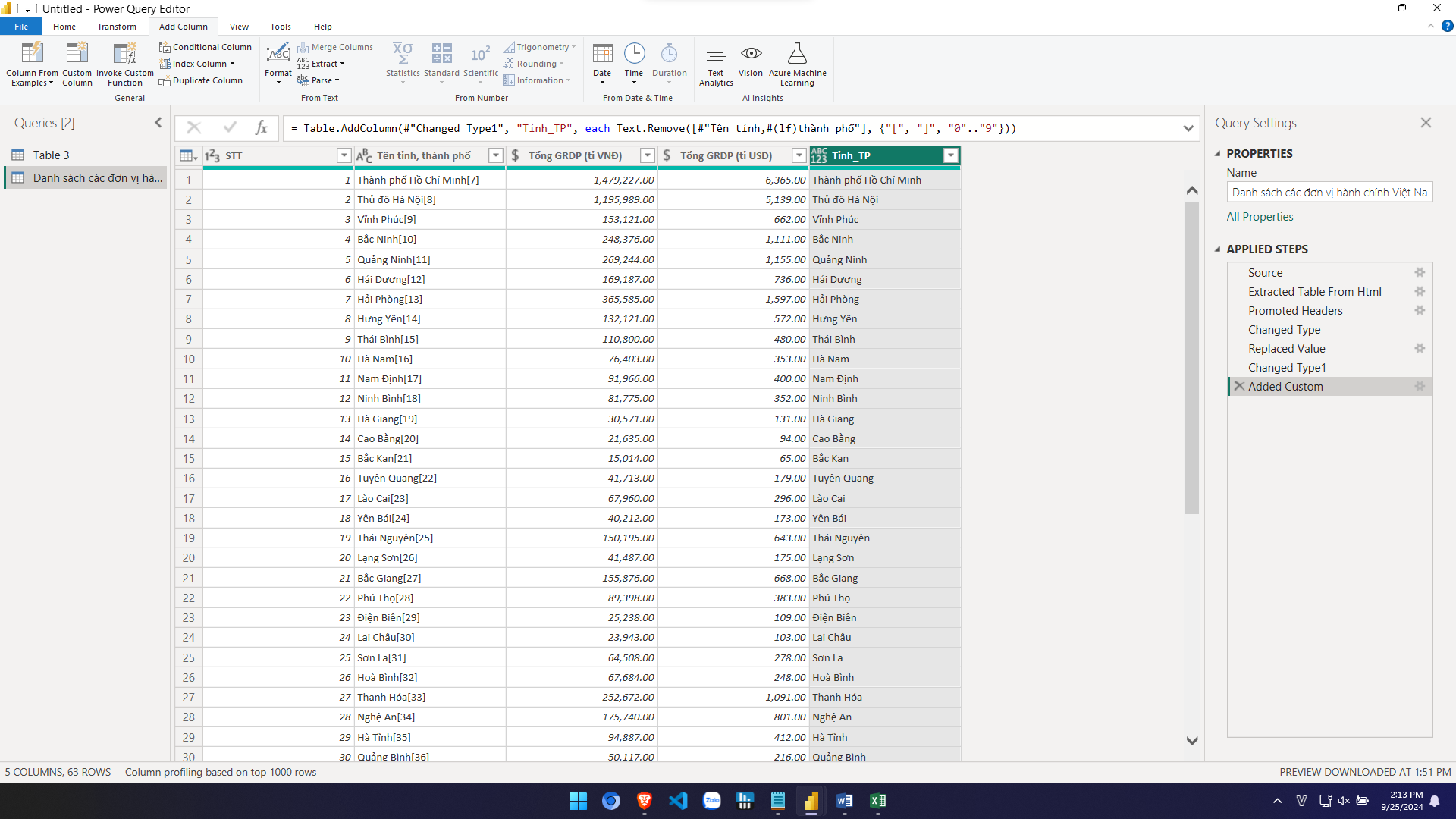
- Tương tự như tệp dữ liệu “Tỉnh\_thành\_Việt\_Nam” ta cũng chuyển đổi *“ . ”* qua thành *“ , ”* và đưa các định dạng dữ liệu từ *text* về *fixed demical number*



- Ở cột *“Tên tỉnh, thành phố”* dữ liệu thường có các chú thích *[number]*. Ta thực hiện tạo 1 *custom column* và đặt tên là *“Tỉnh\_TP”*. Sau đó viết 1 query để xử lý các chú thích

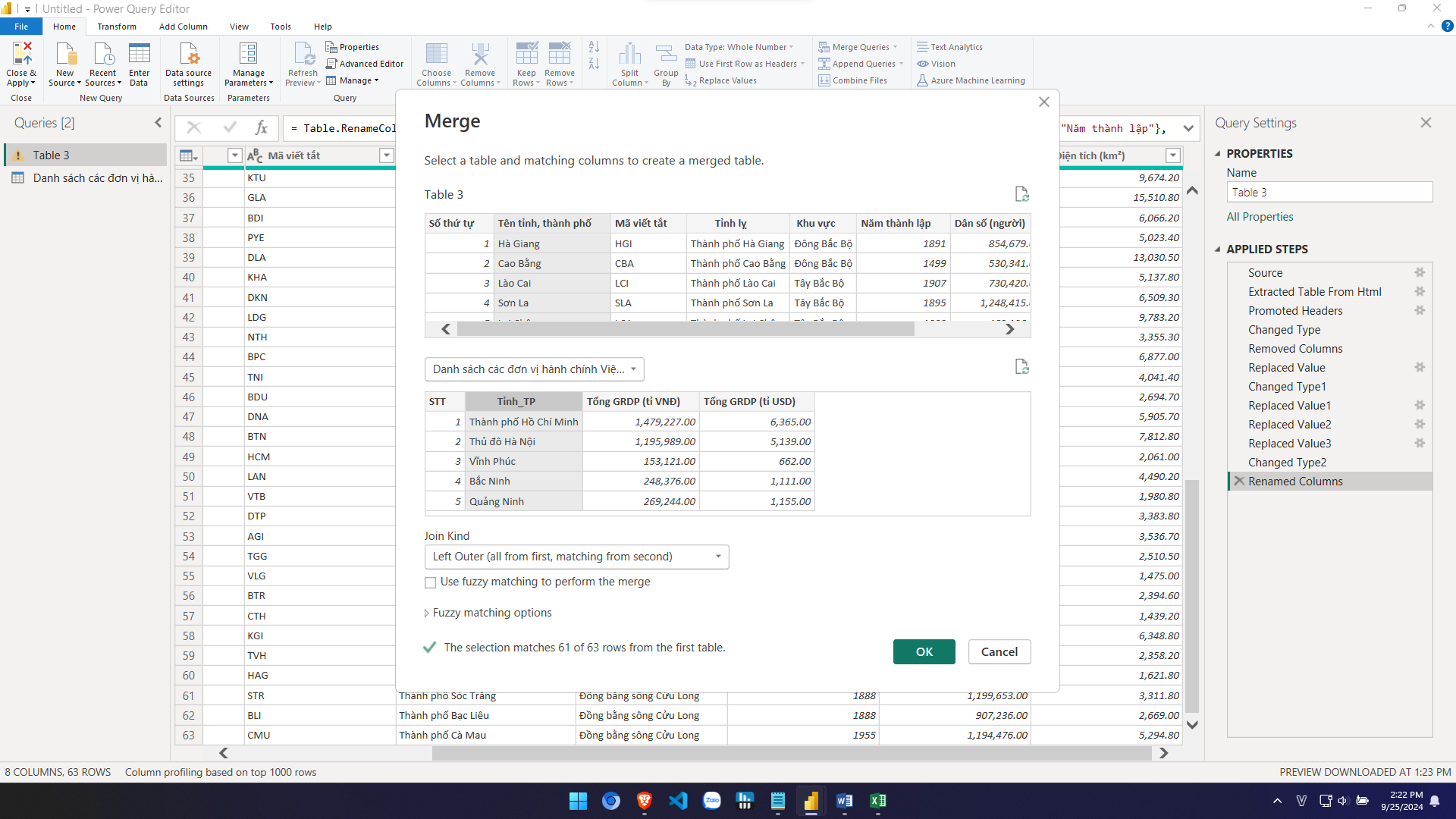
*“Text.Remove([#"Tên tỉnh,#(lf)thành phố"], {"[", "]", "0".."9"})”*

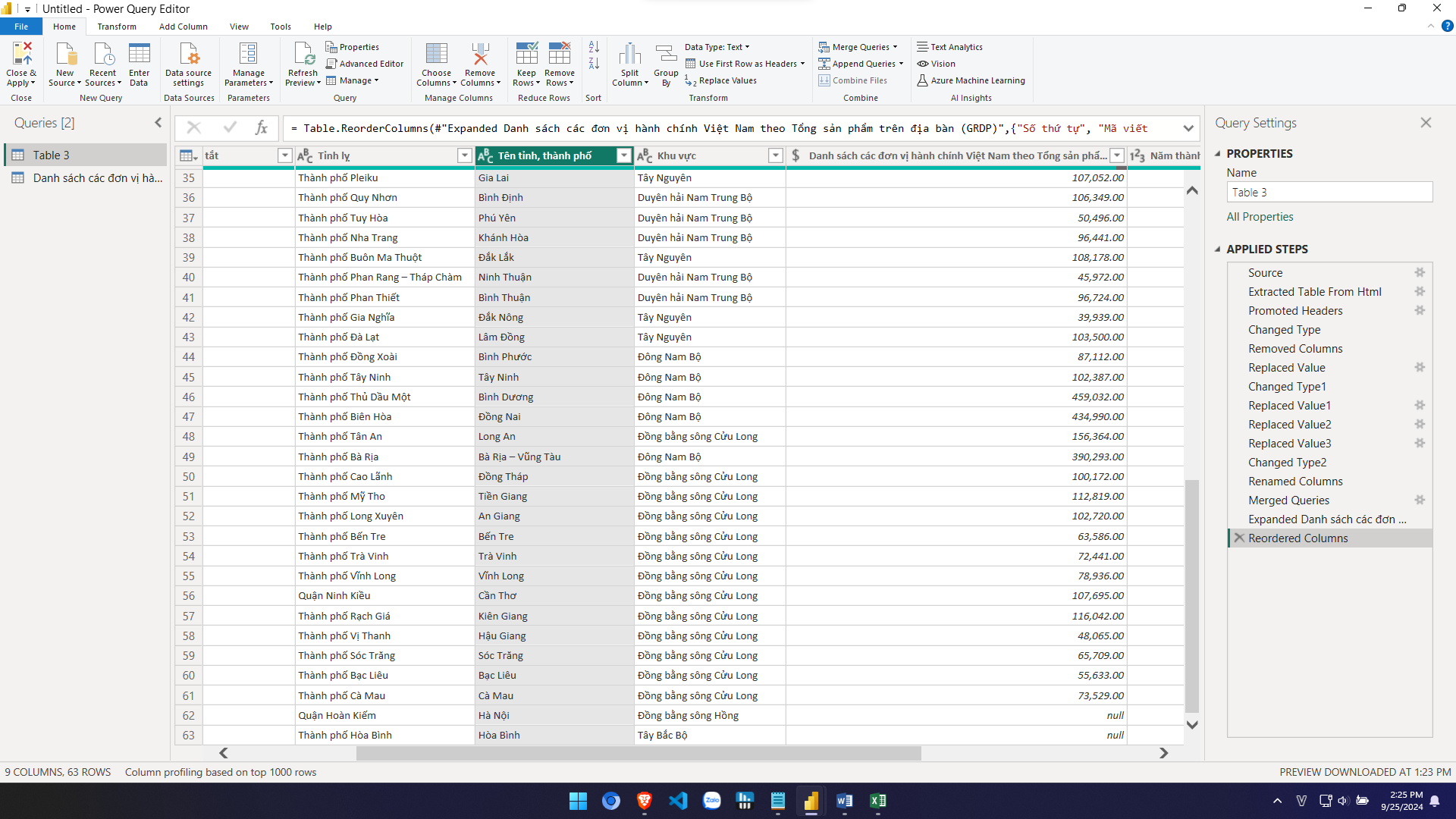




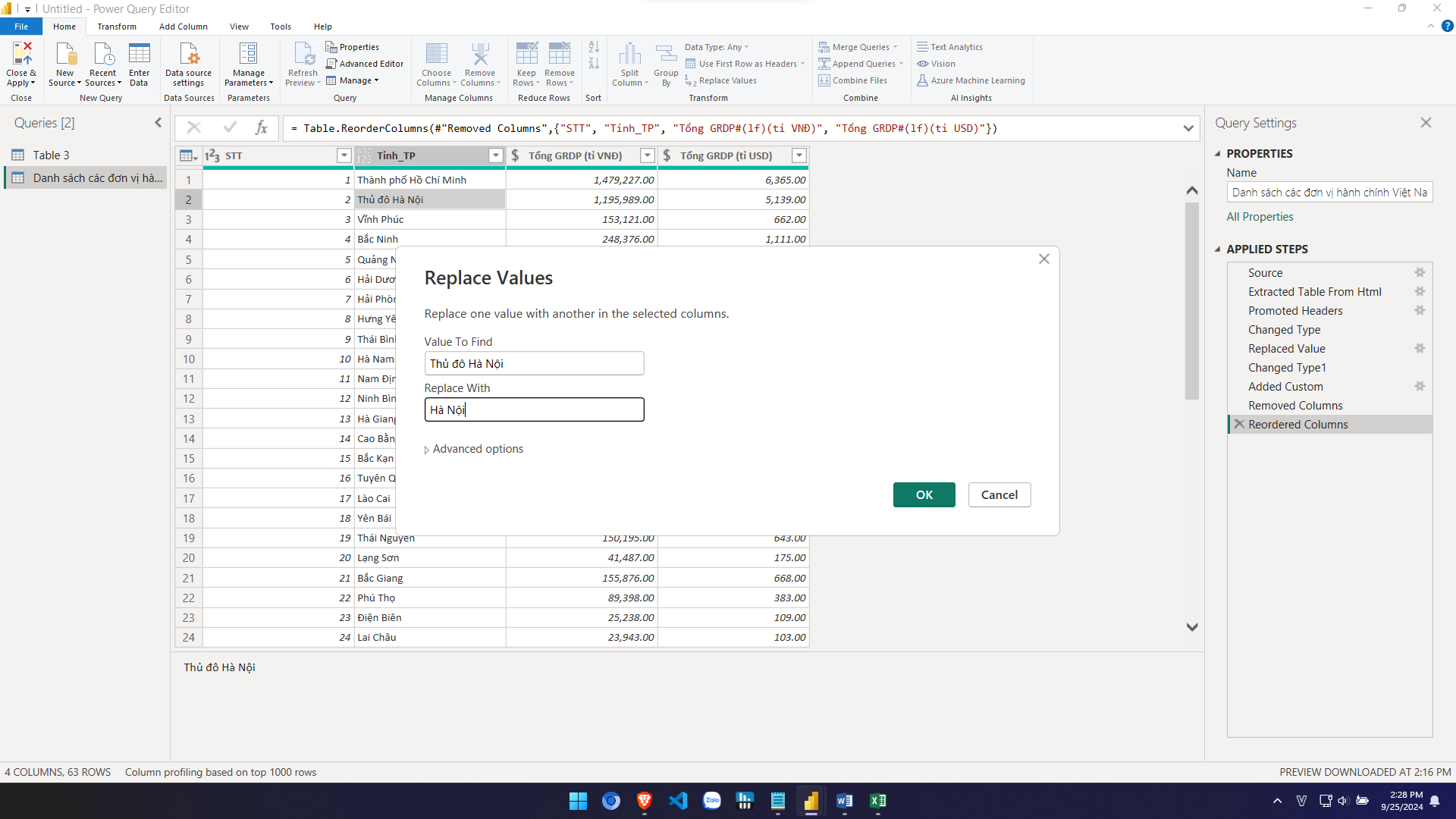
- Sau khi tiền xử lý 2 table “Tỉnh\_thành\_Việt\_Nam” và “Danh\_sách\_đơn\_vị\_hành\_chính\_

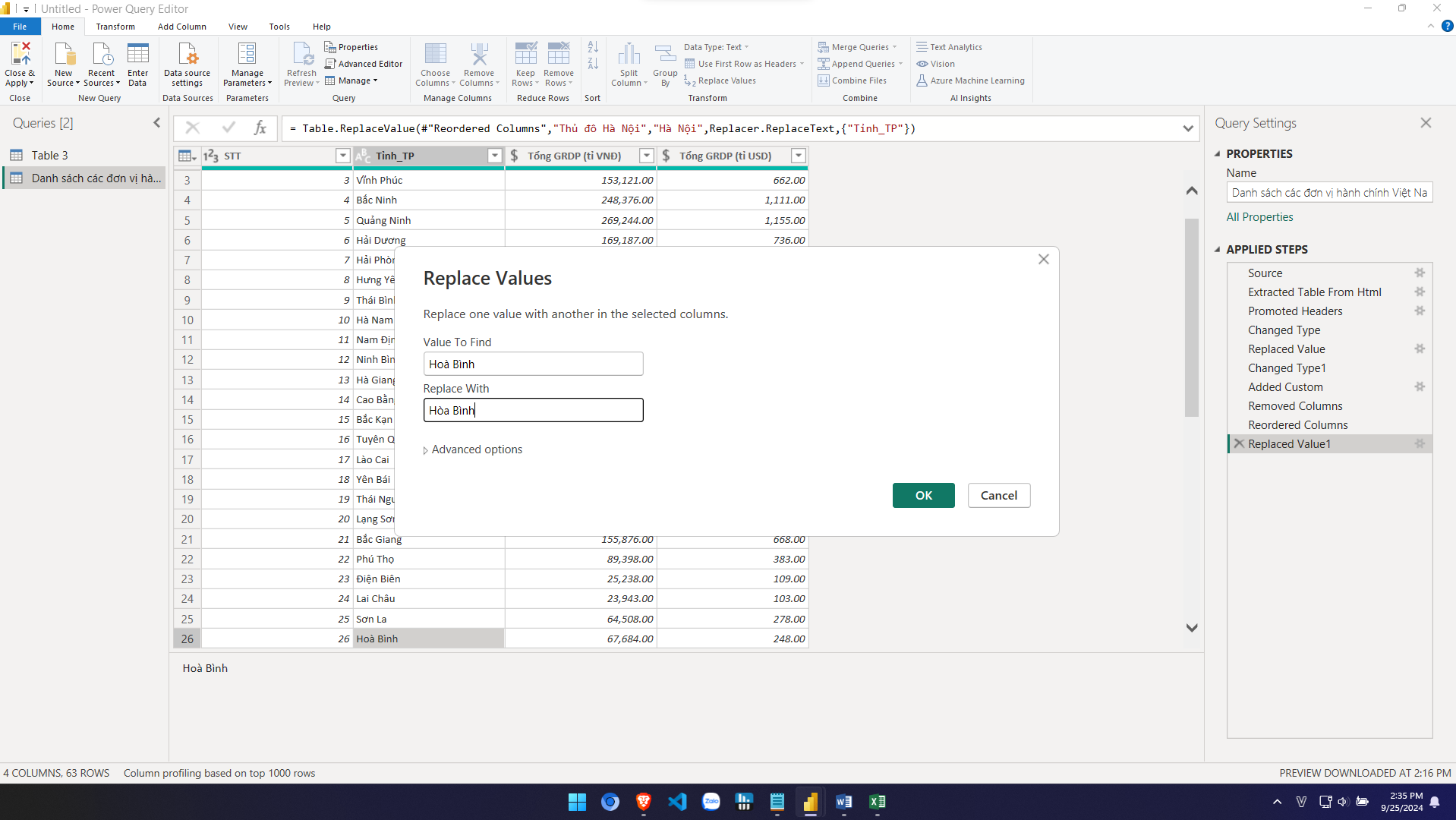
Việt\_Nam\_theo\_GRDP” ta tiến thành *Merge* 2 table lại.



- Sau khi *Merge* ta thấy có 2 giá trị *null* trong table là “Hà Nội” và “Hoà Bình”  


- Ta tiến hành đồng nhất tên của 2 dữ liệu “Hà Nội” và “Hoà Bình” ở 2 tệp dữ liệu

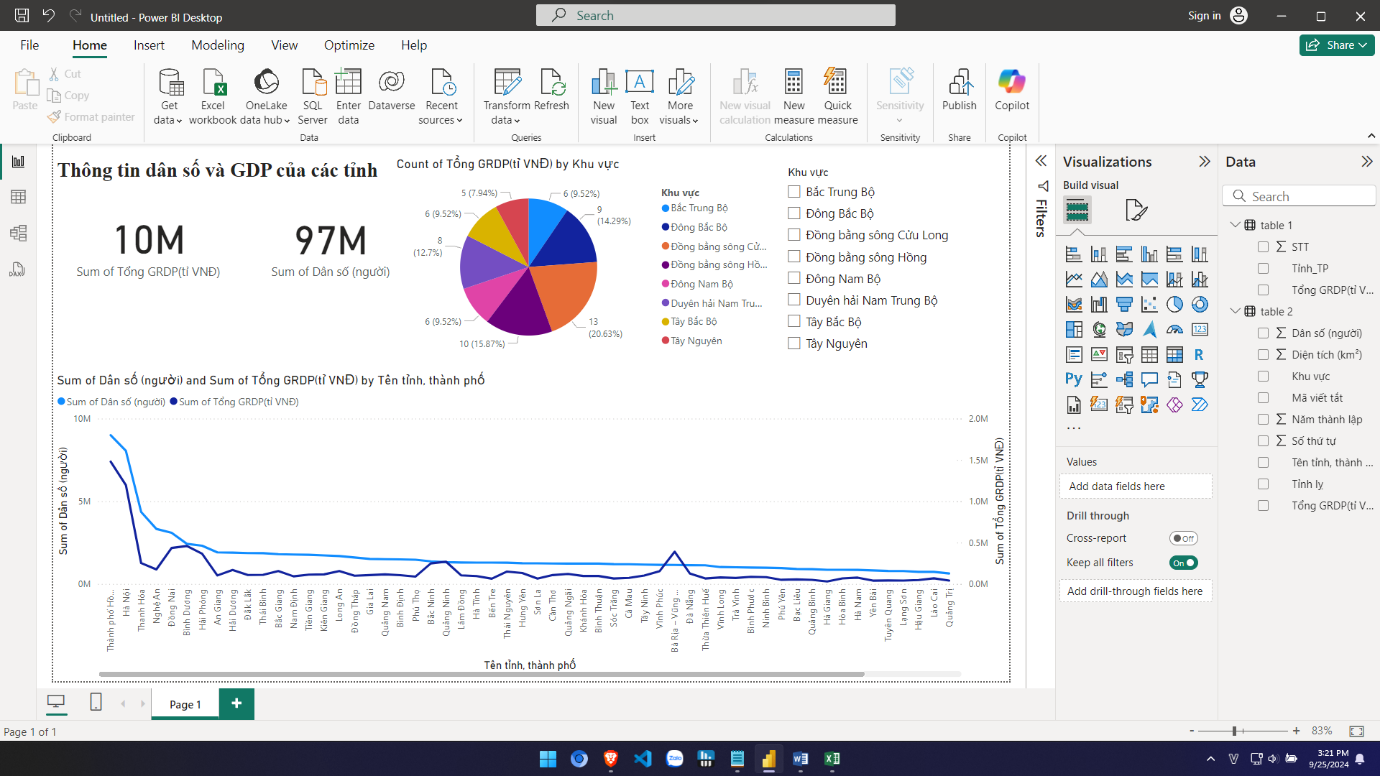




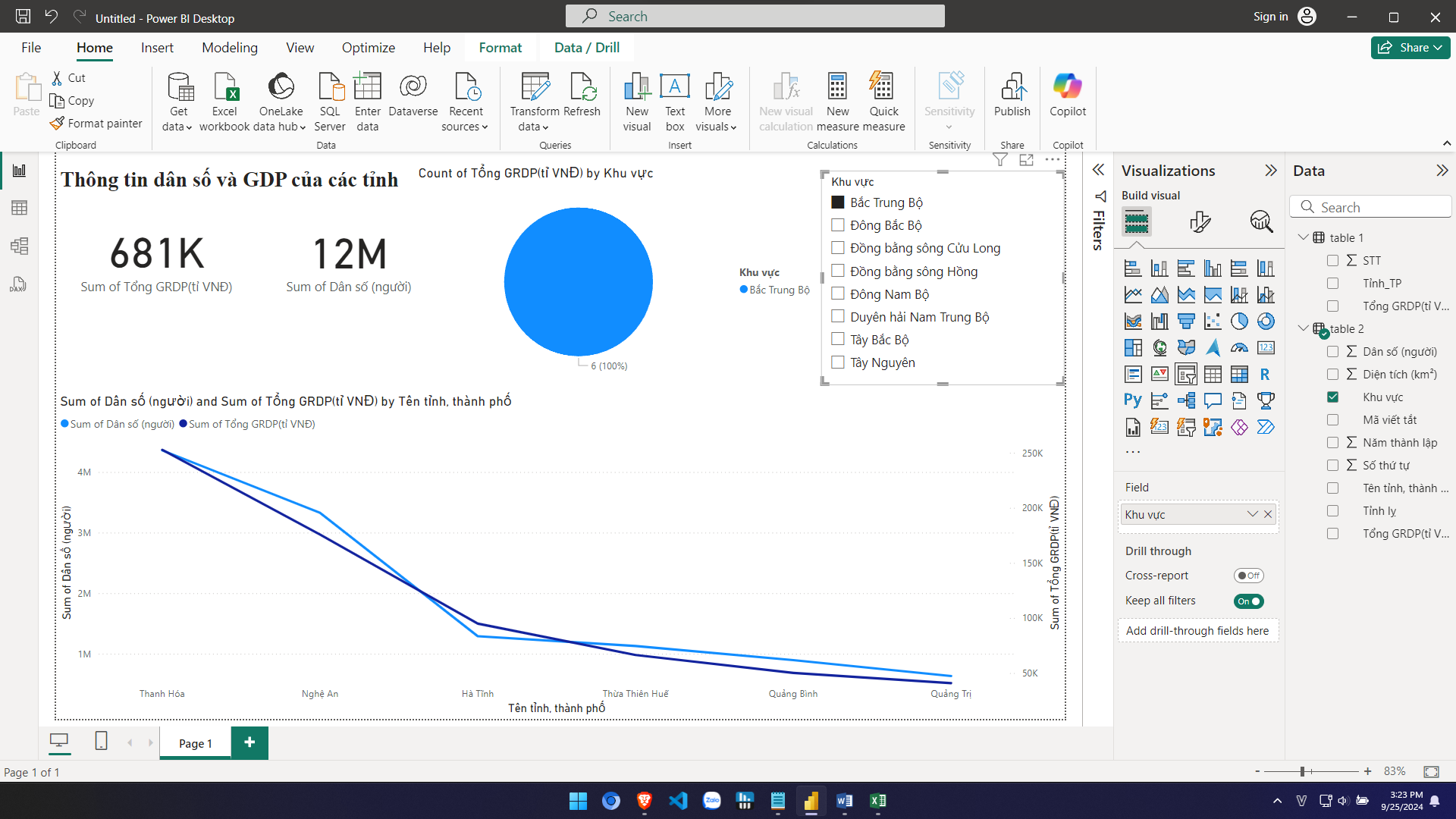
**Kết quả là không còn tồn tại giá trị null trong table dữ liệu**

4. Vẽ biểu đồ và tạo một Dashboard với Power BI

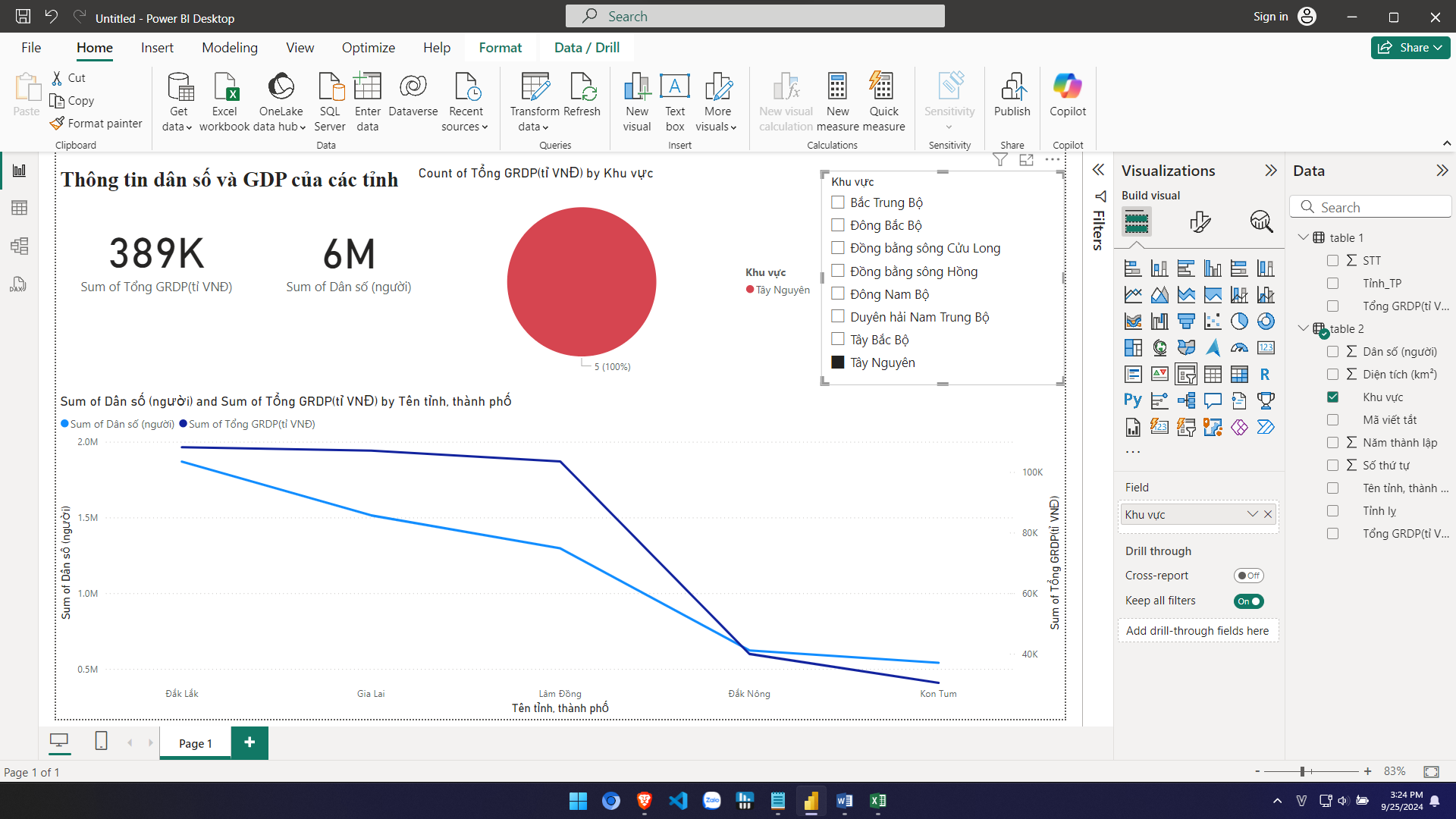
**Dashboard tổng quát báo cáo về dân số và GDP của các tỉnh thành Việt Nam**



Một vài hình ảnh trực quan hoá dữ liệu về các khu vực Bắc Trung Bộ:



Một vài hình ảnh trực quan hoá dữ liệu về các khu vực Tây Nguyên:



5. Xây dựng một model đơn giản với Power BI (linear regression...)

Để có thể tạo ra một model đơn giản là Linear regression với PowerBI. Em đã thực hiện các bước sau để giải và tìm ra phương trình tuyến tính của tập dữ liệu:

Bước 1: Đi tìm tổng cộng luỹ kế (Running Total) với mục đích:

+ Làm mượt dữ liệu

+ Xử lý các dữ liệu không tuyến tính hoặc biến động mạnh

+ Chuẩn bị dữ liệu phù hợp cho mô hình hồi quy

**Code Dax trong PowerBI:**

*FrameGRDP =*

*CALCULATE(*

*SUM(dsGRDP[GRDP]),*

*FILTER(*

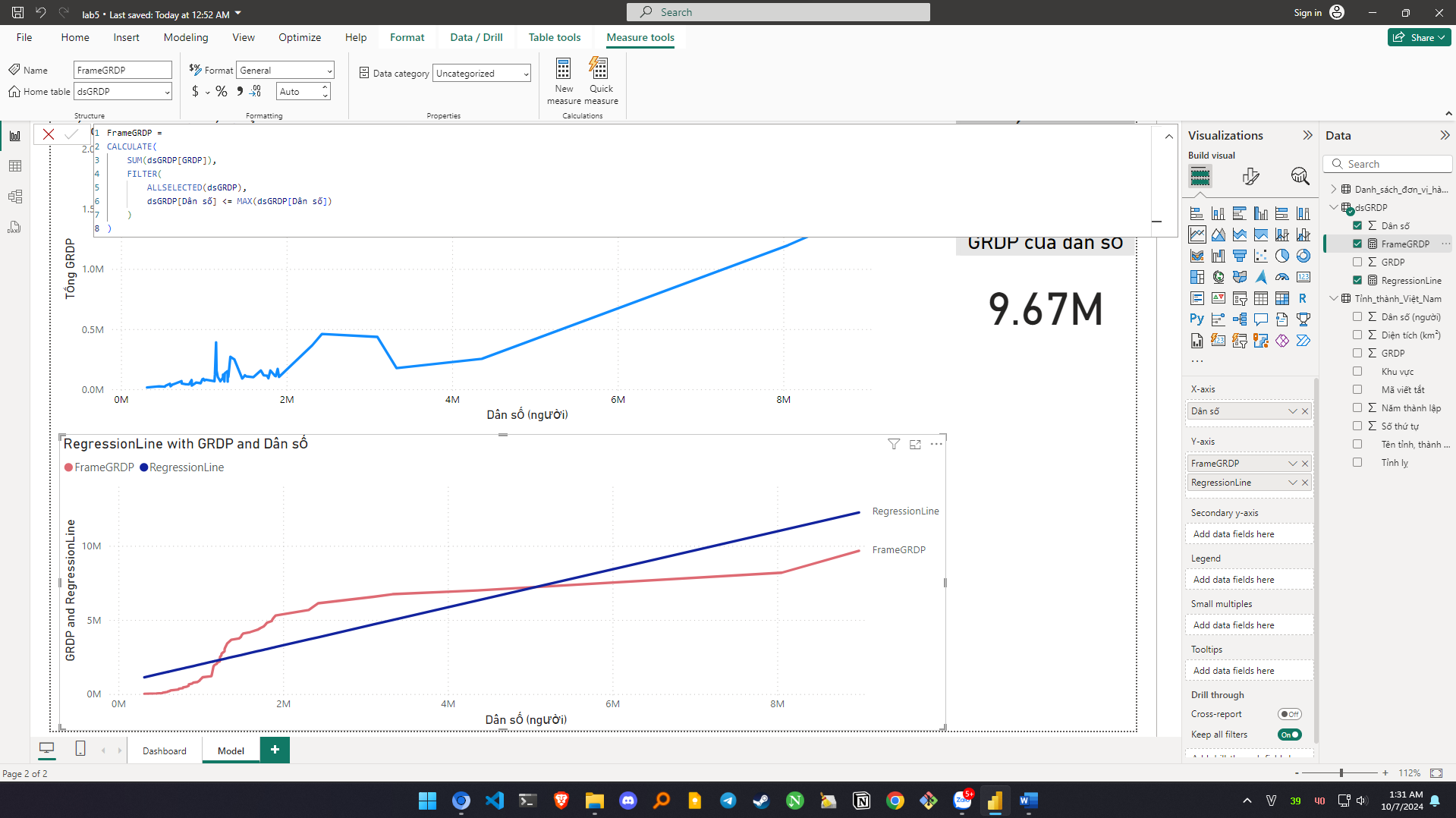
*ALLSELECTED(dsGRDP),*

*dsGRDP[Dân số] <= MAX(dsGRDP[Dân số])*

*)*

*)*

**Hình ảnh code**:



Bước 2: Thiết lập phương trình hồi quy tuyến tính (Linear Regression)

**Code Dax trong PowerBI:**

*RegressionLine =*

*-- Tạo biến dữ liệu cần thiết*

*VAR data =*

*FILTER(*

*SELECTCOLUMNS(*

*ALLSELECTED(dsGRDP),*

*"x\_values", dsGRDP[Dân số],*

*"y\_values", [FrameGRDP]*

*),*

*NOT(ISBLANK([x\_values])) &&*

*NOT(ISBLANK([y\_values]))*

*)*

*-- Tính toán các biến cần thiết cho phương trình hồi quy*

*VAR y\_sum = SUMX(data, [y\_values])*

*VAR x\_sum = SUMX(data, [x\_values])*

*VAR x2 = SUMX(data, [x\_values] ^ 2)*

*VAR xy = SUMX(data, [x\_values] \* [y\_values])*

*VAR row\_count = COUNTROWS(data)*

*-- Giải quyết phương trình để tìm intercept và slope*

*VAR Intercept = (y\_sum \* x2 - x\_sum \* xy) / (row\_count \* x2 - (x\_sum) ^ 2)*

*VAR Slope = (row\_count \* xy - x\_sum \* y\_sum) / (row\_count \* x2 - (x\_sum) ^ 2)*

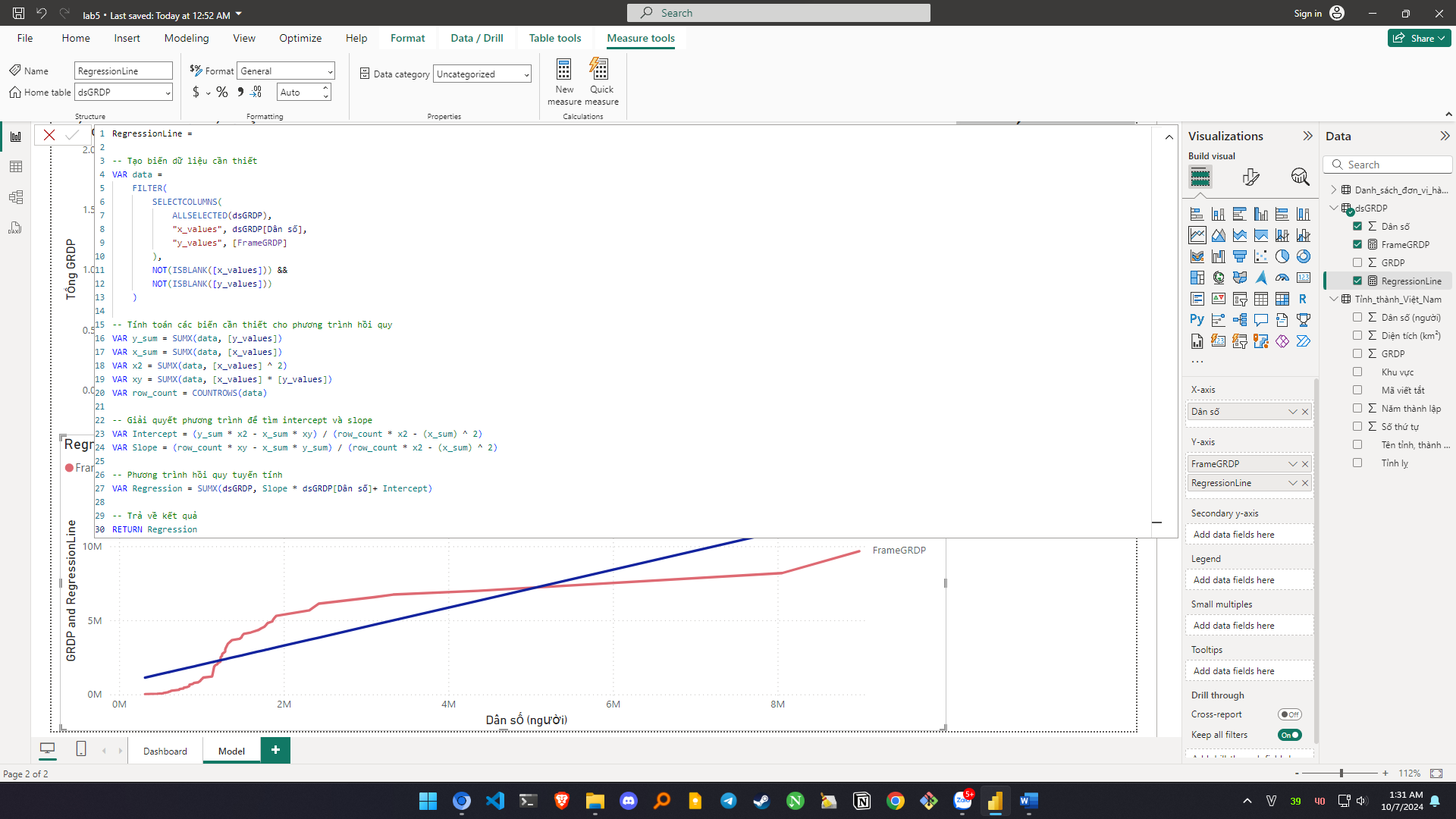
*-- Phương trình hồi quy tuyến tính*

*VAR Regression = SUMX(dsGRDP, Slope \* dsGRDP[Dân số]+ Intercept)*

*-- Trả về kết quả*

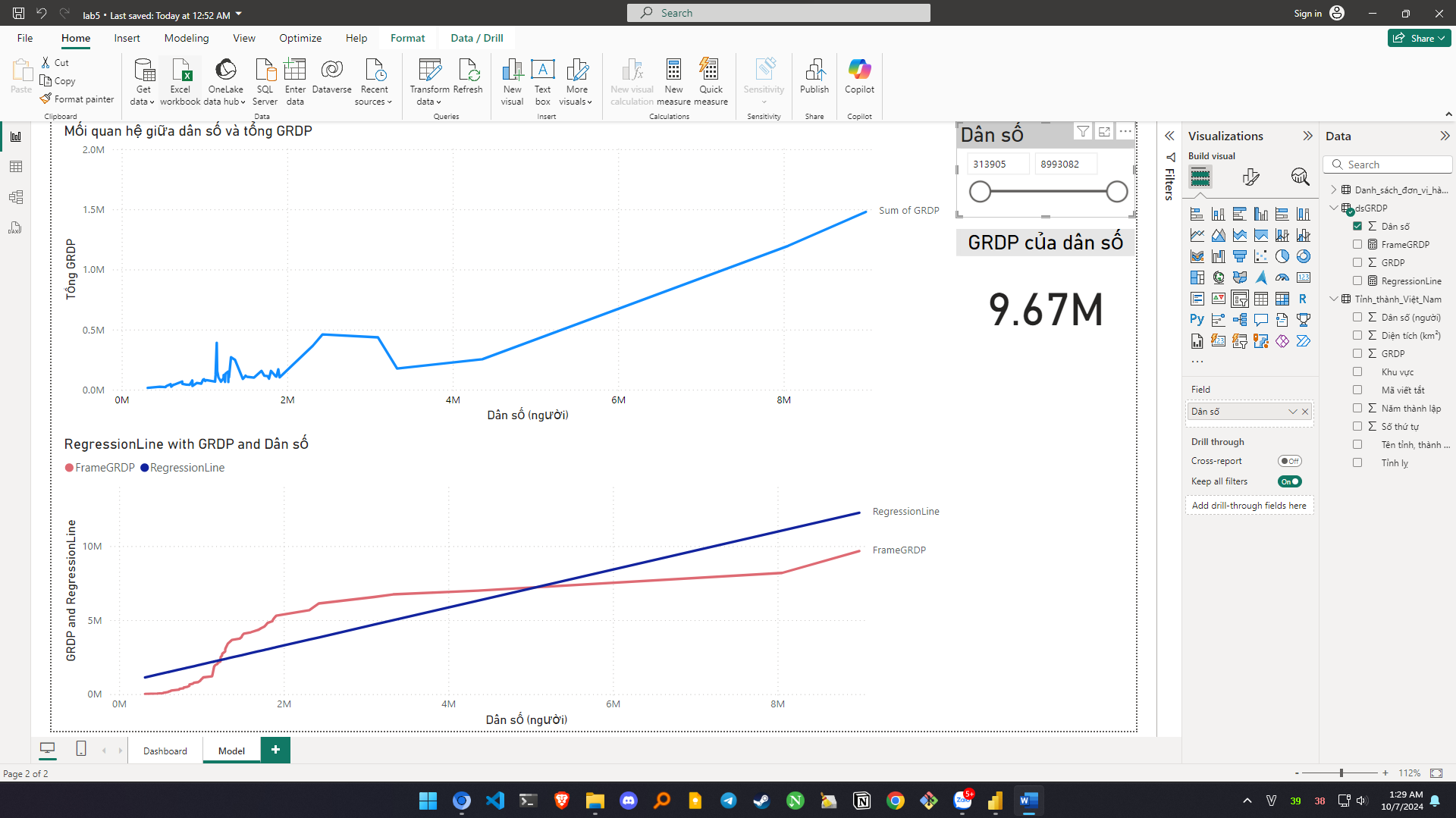
*RETURN Regression*

**Hình ảnh code:**



Bước 3: Trực quan hoá với dữ liệu và dự đoán với mô hình

**Hình ảnh visual về mô hình hồi quy:**



5. Chia sẻ report / Dashboard kết quả bằng Power BI services

6. Thử bỏ file dữ liệu của bạn vào một Hệ quản trị CSDL bất kỳ (sqlite, MS.SQL Server, PosgreSQL...)

----------------------------------------------------------------------------------

NỘP BÀI

- Đại diện một sv trong nhóm nộp bài

- Lưu kết quả vào file word. Cú pháp đặt tên: <stt>\_<ten-project>.docx

- Trong báo cáo chứa kết quả và link đến services của bạn.

----------------------------------------------------------------------------------

Tham khảo

- Get data

https://learn.microsoft.com/en-us/training/modules/get-data-power-bi/

- Clean, transform, and load data in Power BI

https://learn.microsoft.com/en-us/training/modules/clean-data-power-bi/

- Choose a model

https://learn.microsoft.com/en-us/training/modules/choose-power-bi-model-framework/

- Report

https://learn.microsoft.com/en-us/training/modules/power-bi-effective-reports/

- Dashboard

https://learn.microsoft.com/en-us/training/modules/create-dashboards-power-bi/

- Publish & Share

https://learn.microsoft.com/en-us/training/modules/publish-share-power-bi/

- Others

https://learn.microsoft.com/en-us/training/browse/?products=power-bi&roles=data-analyst