**Exploratory Data Analysis (EDA)**

The **Exploratory Data Analysis (EDA)** phase focuses on understanding the key characteristics of the loan dataset, detecting any patterns, and identifying potential correlations between the variables. The objective is to provide a solid foundation for further analysis by summarizing key statistics, visualizing trends, and uncovering underlying relationships.

1. **Loan portfolio analysis**
2. **Distribution of Loan Status**

The distribution of loan status is crucial for understanding the proportion of loans that are approved or rejected. By analysing loan status, we can gain insights into the performance of the loan portfolio and assess any potential biases in loan approval.

**SQL query:**

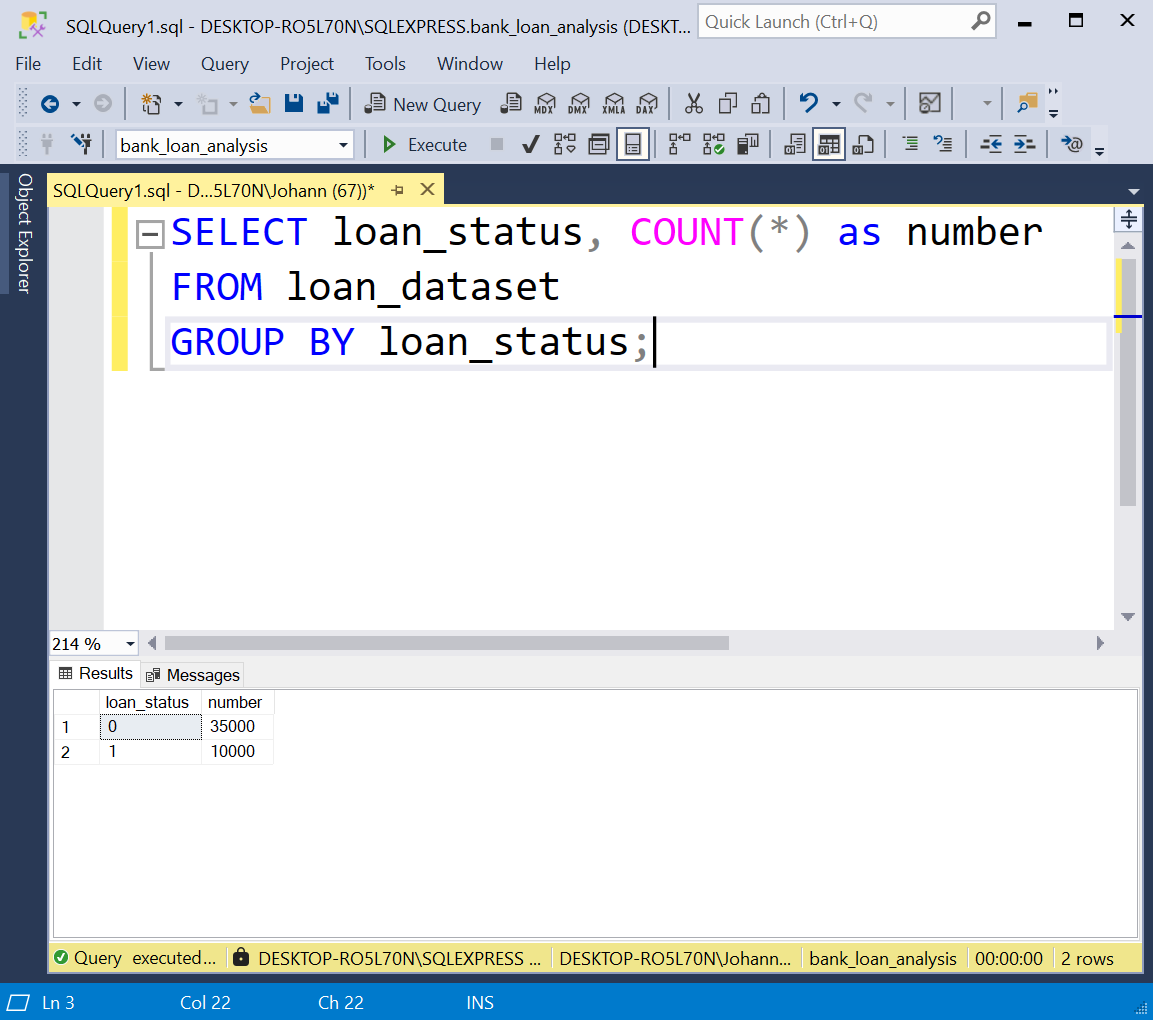


Figure 1 : Distribution of Loan Status

**Result:**

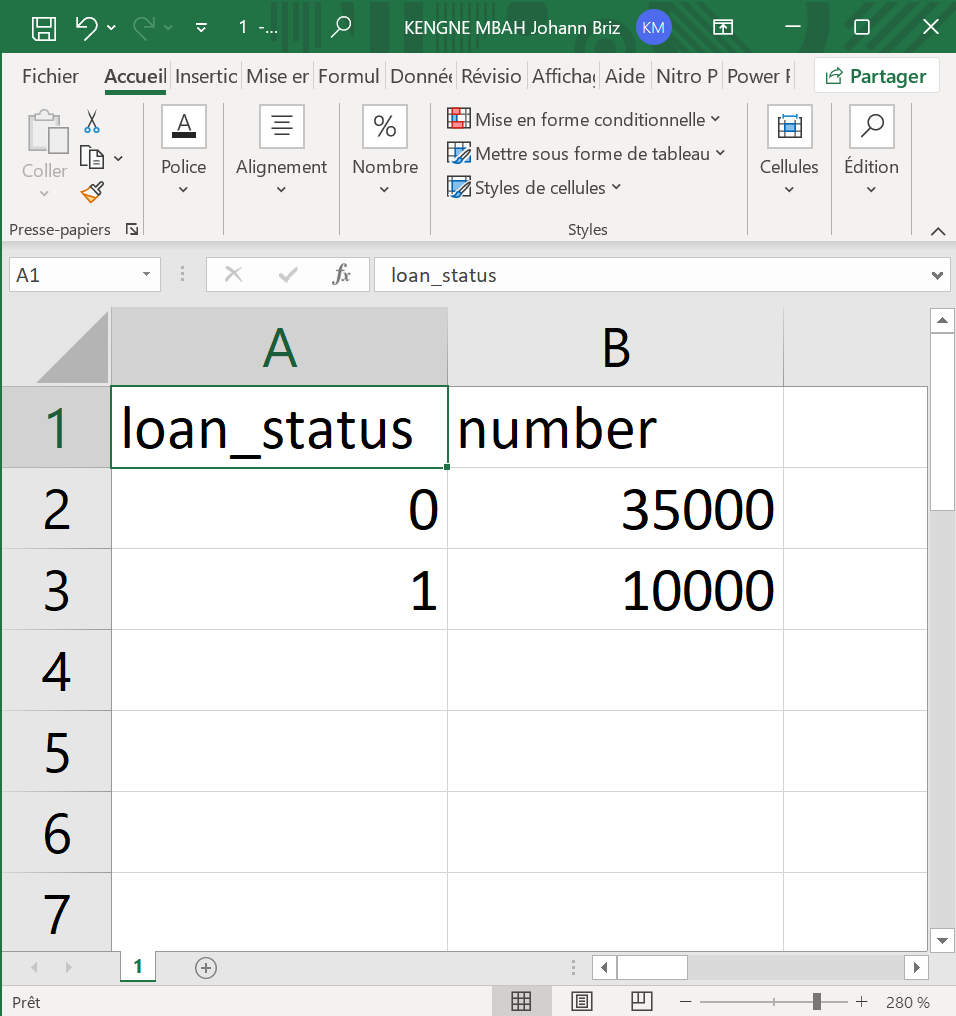


Figure 2 : Distribution of Loan Status Result

**Illustration:**

**Interpretation:**

We now have the number of loans in each status category providing a clear picture of loan performance.

1. **Distribution of Loan Amounts**

The distribution of loan amounts is essential to understand how loans are structured in terms of their size. Understanding the distribution of loan amounts through statistical measures like **mean, median and standard deviation** provides deeper insights into how HSBC structures its loans.

**SQL query:**

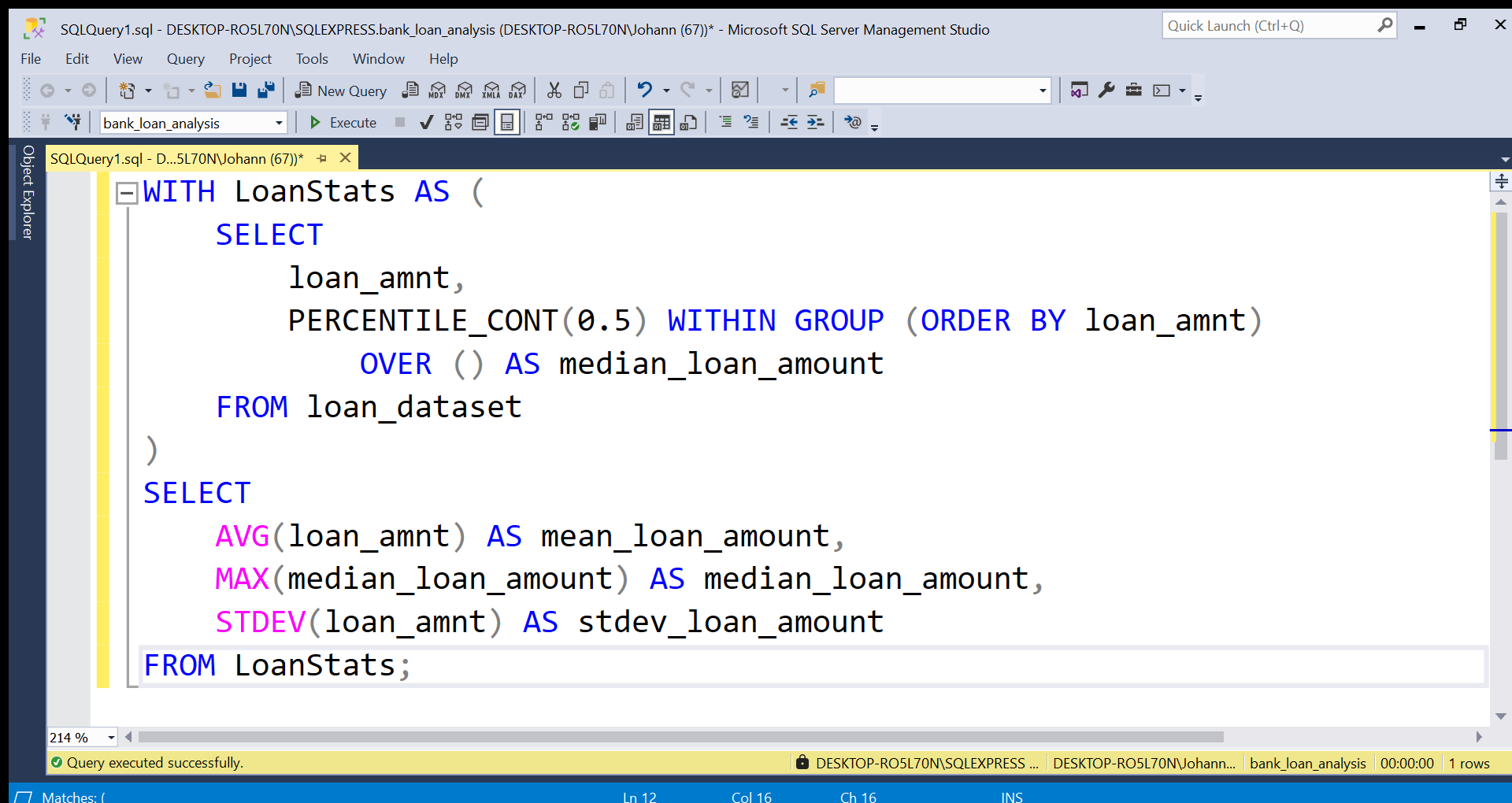


Figure 3 : Distribution of Loan Amounts

**Result:**

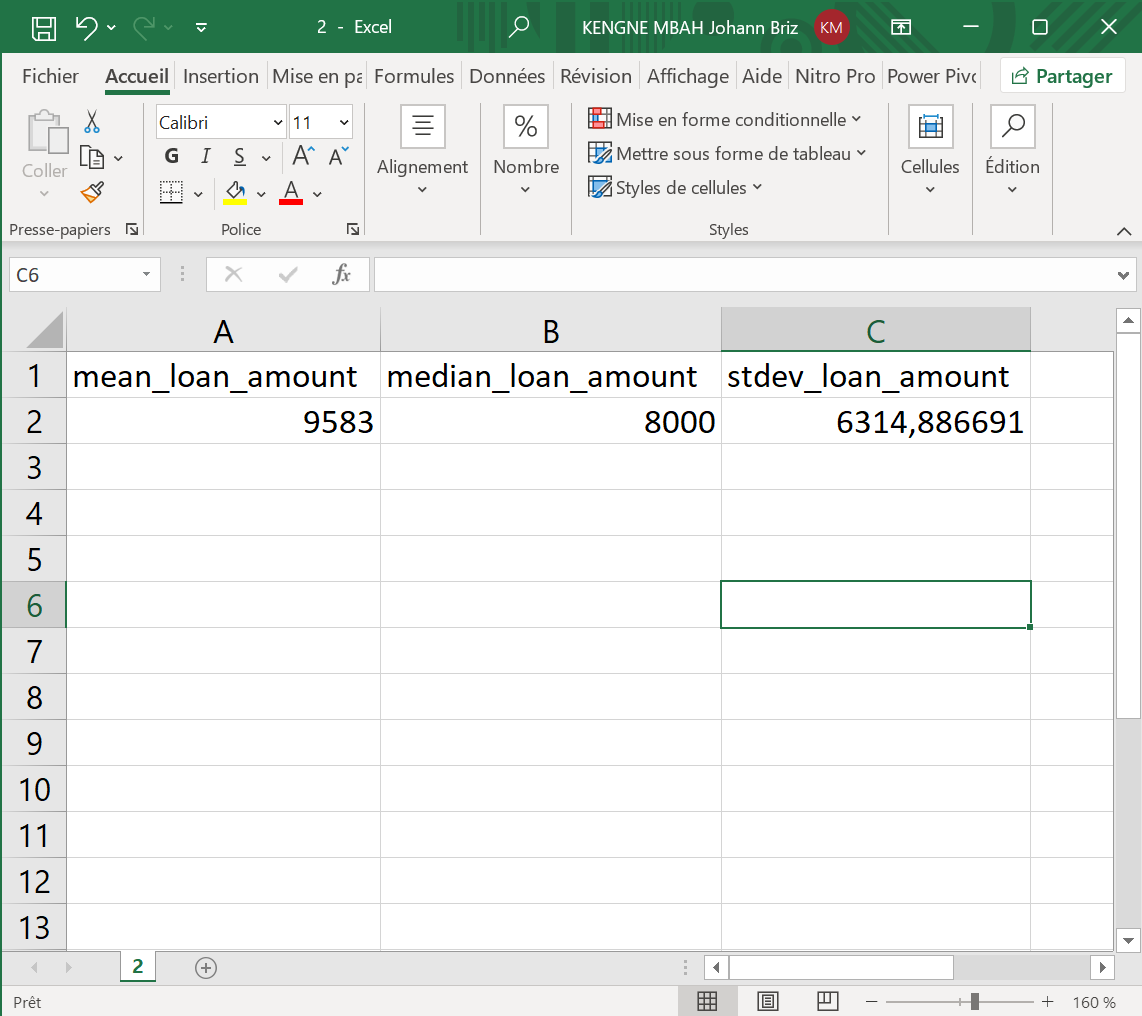


Figure 4 : Distribution of Loan Amounts Result

**Illustration:**

**Interpretation:**

We now have the **mean, median, and standard deviation** of loan amounts, providing a statistical overview of loan distribution. This helps us understand typical loan sizes, detect potential outliers, and optimize lending strategies to mitigate risk.

1. **Average Interest Rate Applied**

The **average interest rate** is crucial for HSBC, as it helps in:

* Evaluating the **profitability of the loan portfolio**.
* Identifying trends in **interest rate policies**.
* Assessing the **impact of interest rates on defaults and loan demand**.

**SQL query:**

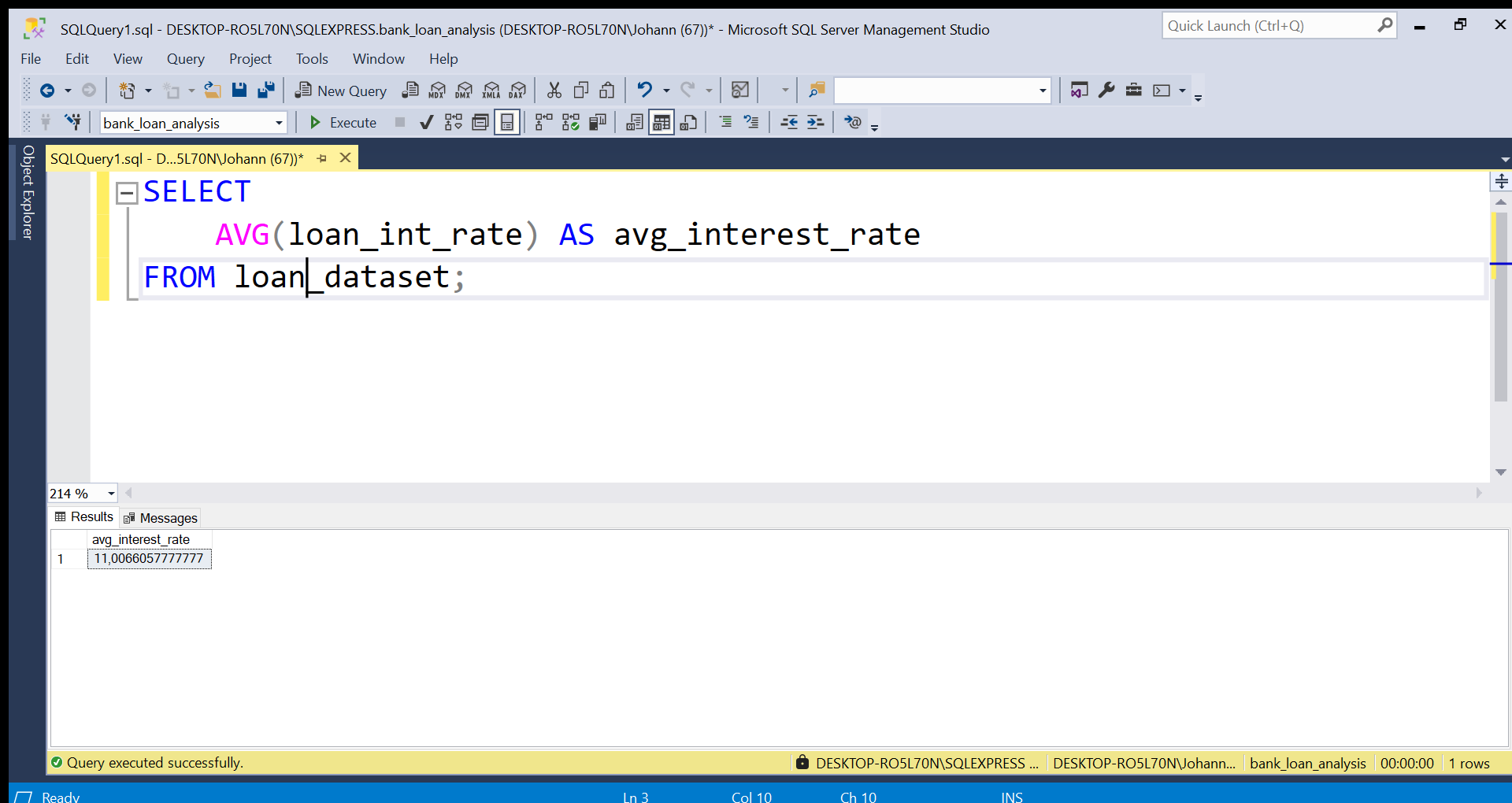


Figure 5 : Average Interest Rate Applied

**Result:**

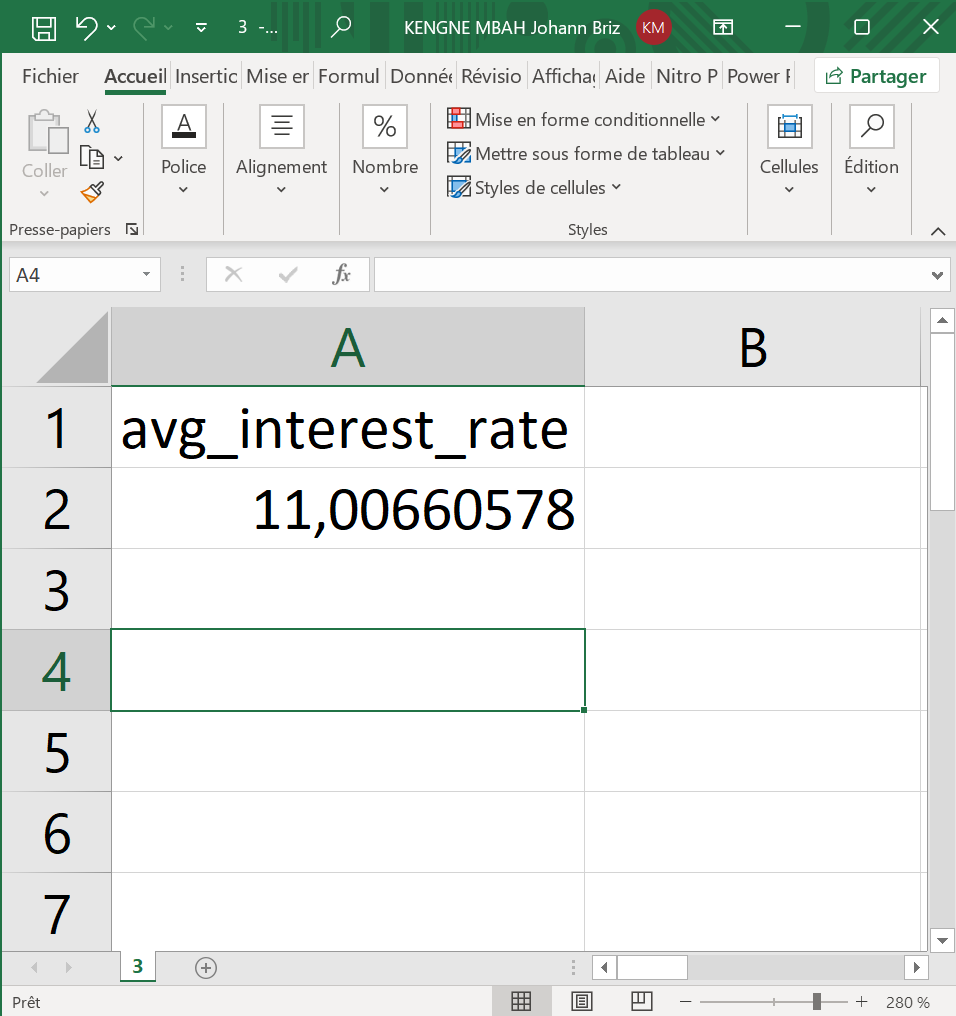


Figure 6 : Average Interest Rate Applied Result

**Interpretation:**

This result indicates that the **average interest rate applied across all loans is 11%**. HSBC can use this information to compare its rates with competitors, adjust pricing strategies, and optimize risk management policies.

1. **Borrower Profile**

### ****Loan Distribution by Borrower Education****

Understanding the **distribution of loans by borrower education** helps us:

* Identify the **most common borrower profiles**.
* Assess **risk levels** associated with different borrower types.
* Optimize **loan policies** and **targeted marketing strategies** for different customer segments.

**SQL query:**

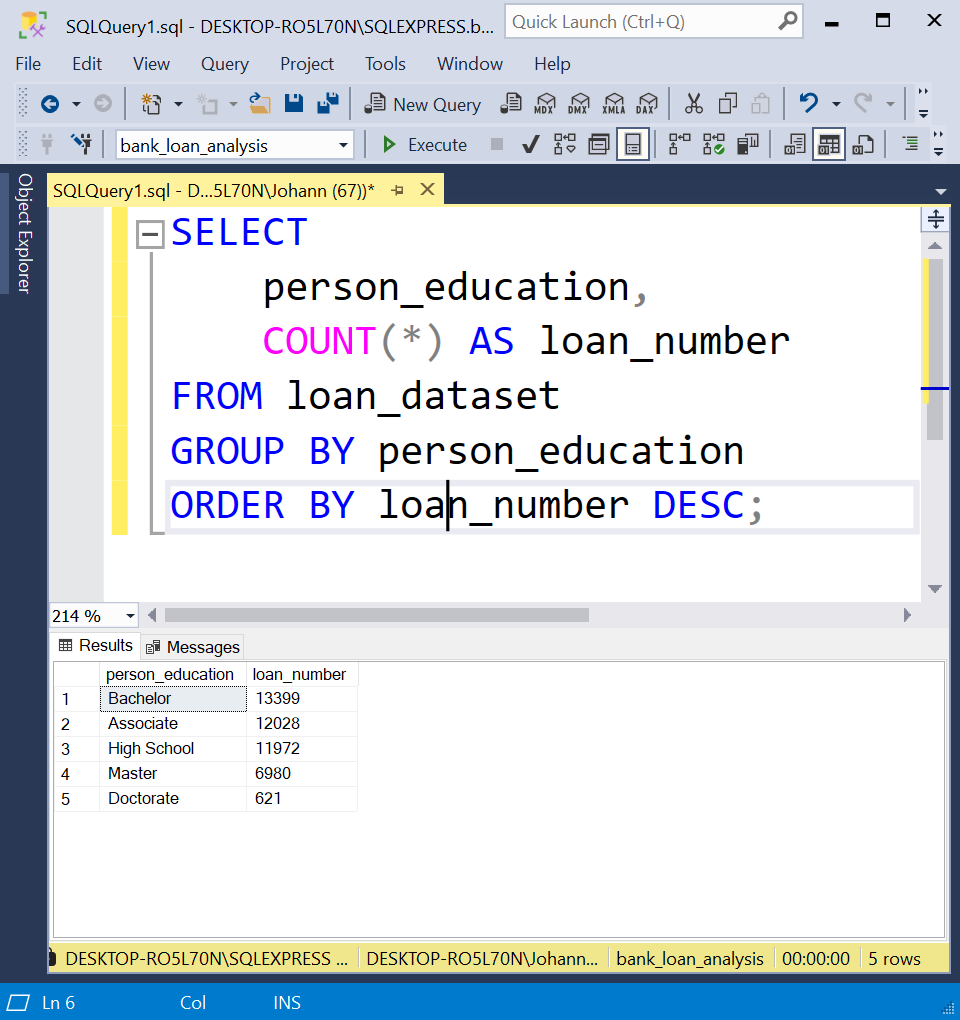


Figure 7 : Loan Distribution by Borrower Education

**Result :**

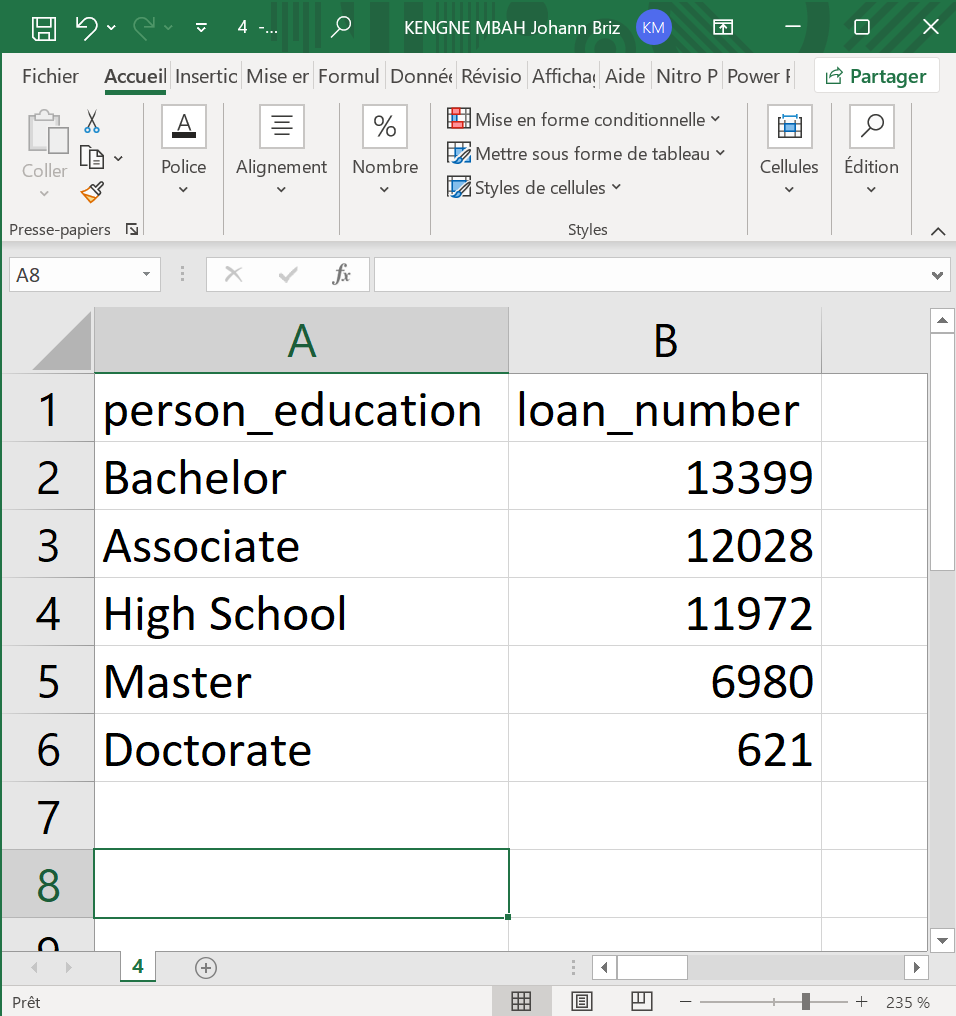


Figure 8 : Loan Distribution by Borrower Education Result

**Illustration:**

**Interpretation:**

This insight helps HSBC **understand which borrower segments are most active**, assess risk associated with different education levels, and **fine-tune lending criteria** accordingly.

### ****Average Income of Borrowers****

Understanding the **average income of borrowers** is crucial for us to:

* Assess **borrowers' repayment capacity**.
* Identify **high-income vs. low-income segments** and their loan behaviours.
* Optimize **loan approval policies** and risk assessment models.

**SQL query:**

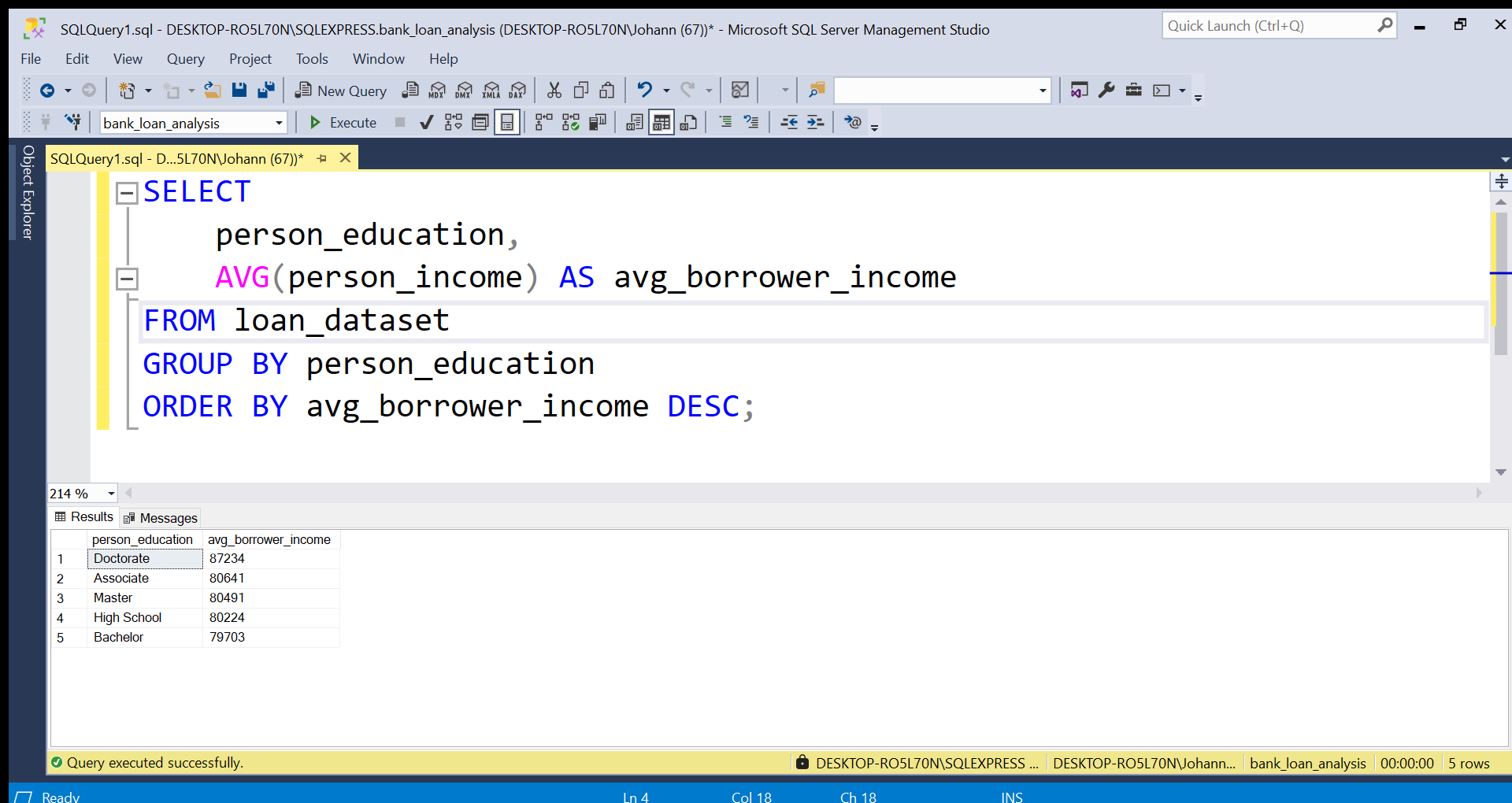


Figure 9 : Average Income of Borrowers

**Result**:

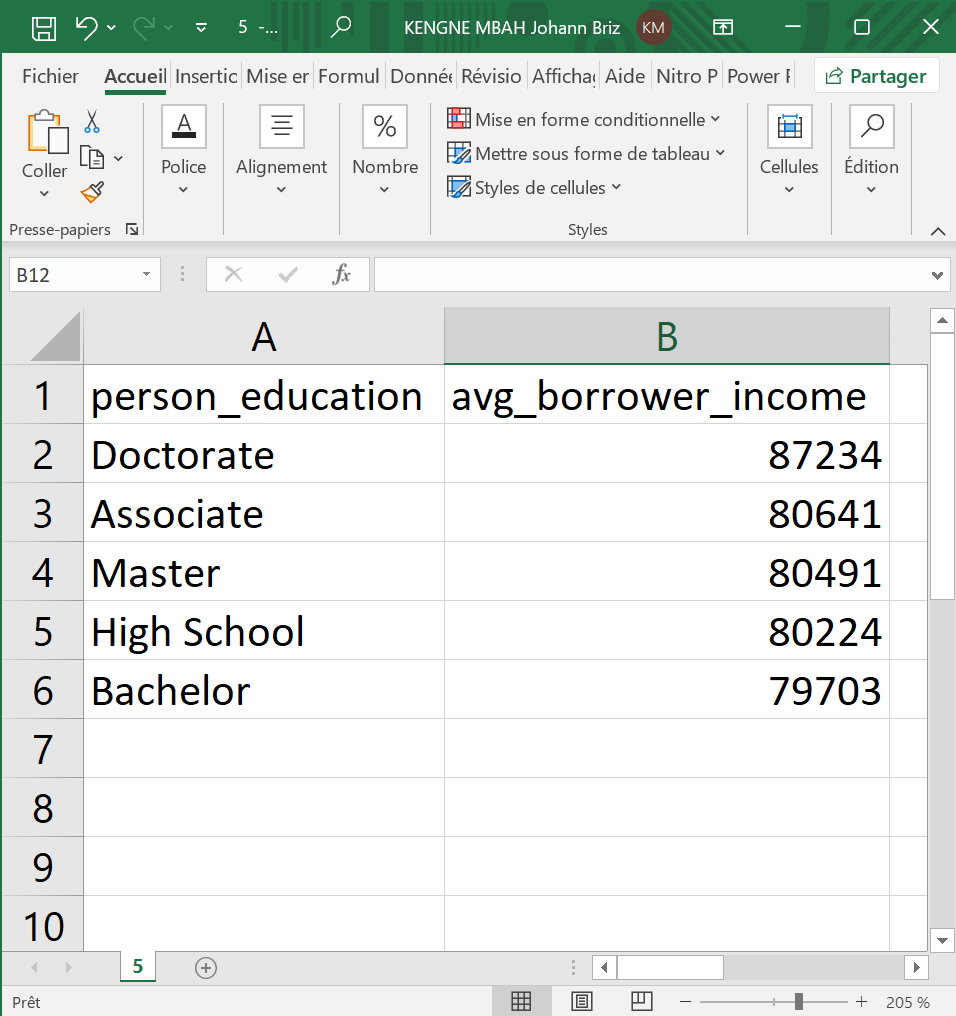


Figure 10 : Average Income of Borrowers Result

**Illustration:**

**Interpretation:**

This result indicates that the **average borrower income is approximatively £80,000**. HSBC can use this data to develop **income-based loan offerings** and ensure that loans are granted to borrowers with sufficient repayment capacity.

### ****Correlation Between Income and Loan Amount****

Understanding the **relationship between borrower income and loan amount** helps us to:

* Determine if higher-income individuals tend to borrow more.
* Assess whether income is a strong predictor of loan size.
* Improve credit risk models and loan approval strategies.

**SQL query:**

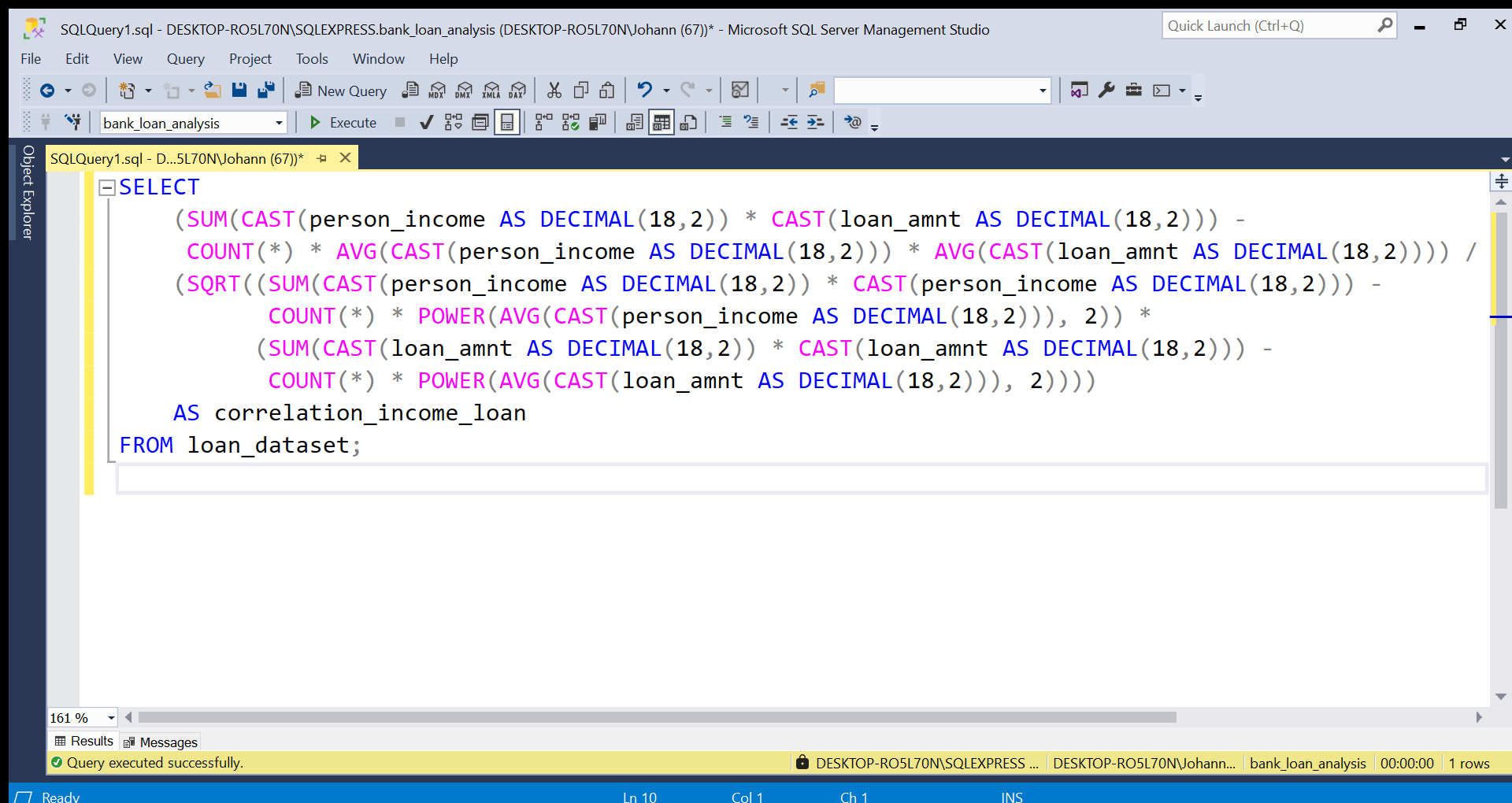


Figure 11 : Correlation Between Income and Loan Amount

**Result**:

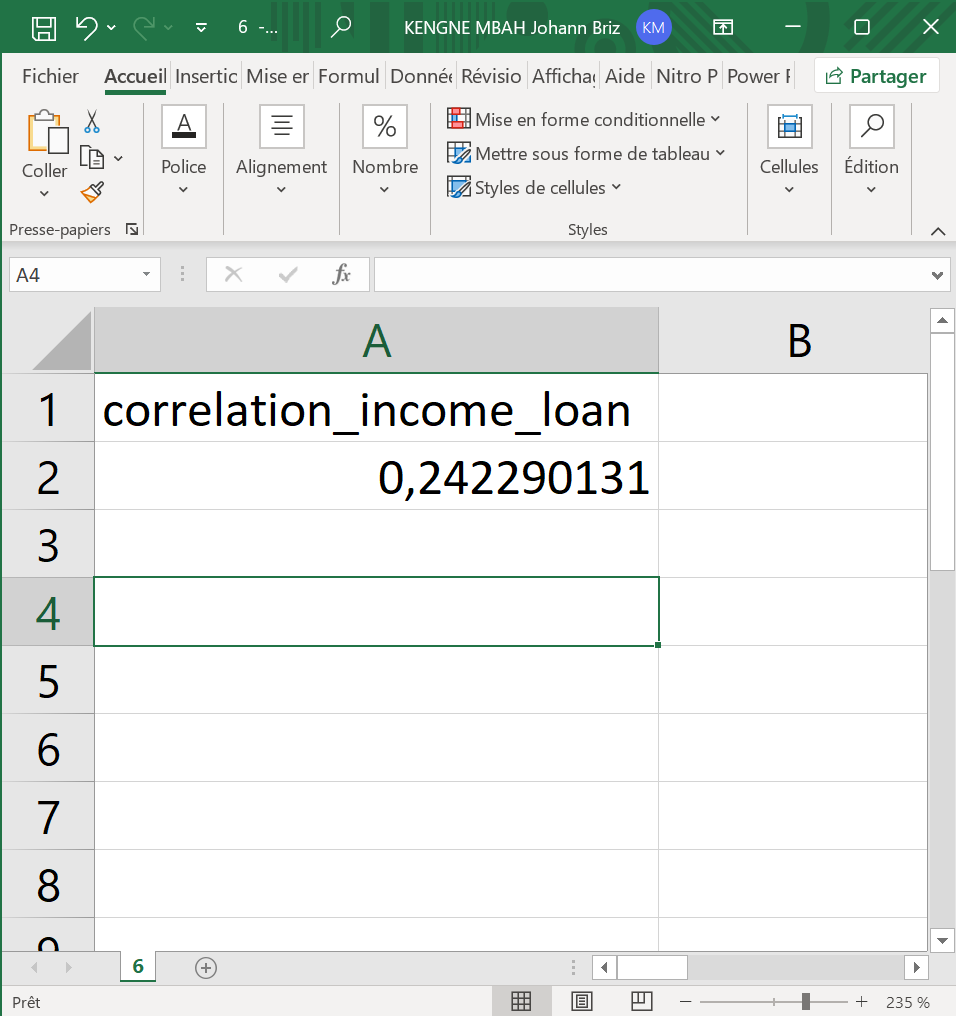


Figure 12 : Correlation Between Income and Loan Amount Result

**Interpretation:**

The correlation coefficient is 0.242 so there is a weak positive correlation between income and loan amount. This indicates that there is a slight tendency for higher-income individuals to borrow more, but the relationship is **not strong**.

### ****Distribution of Socio-Economic Profiles Among Borrowers****

The distribution of socio-economic profiles based on age, profession, and location is crucial for understanding the key characteristics of borrowers in HSBC loan portfolio. By analysing these profiles, we can gain insights into the most common borrower groups and potentially target high-performing segments for future loan offerings.

**SQL query:**

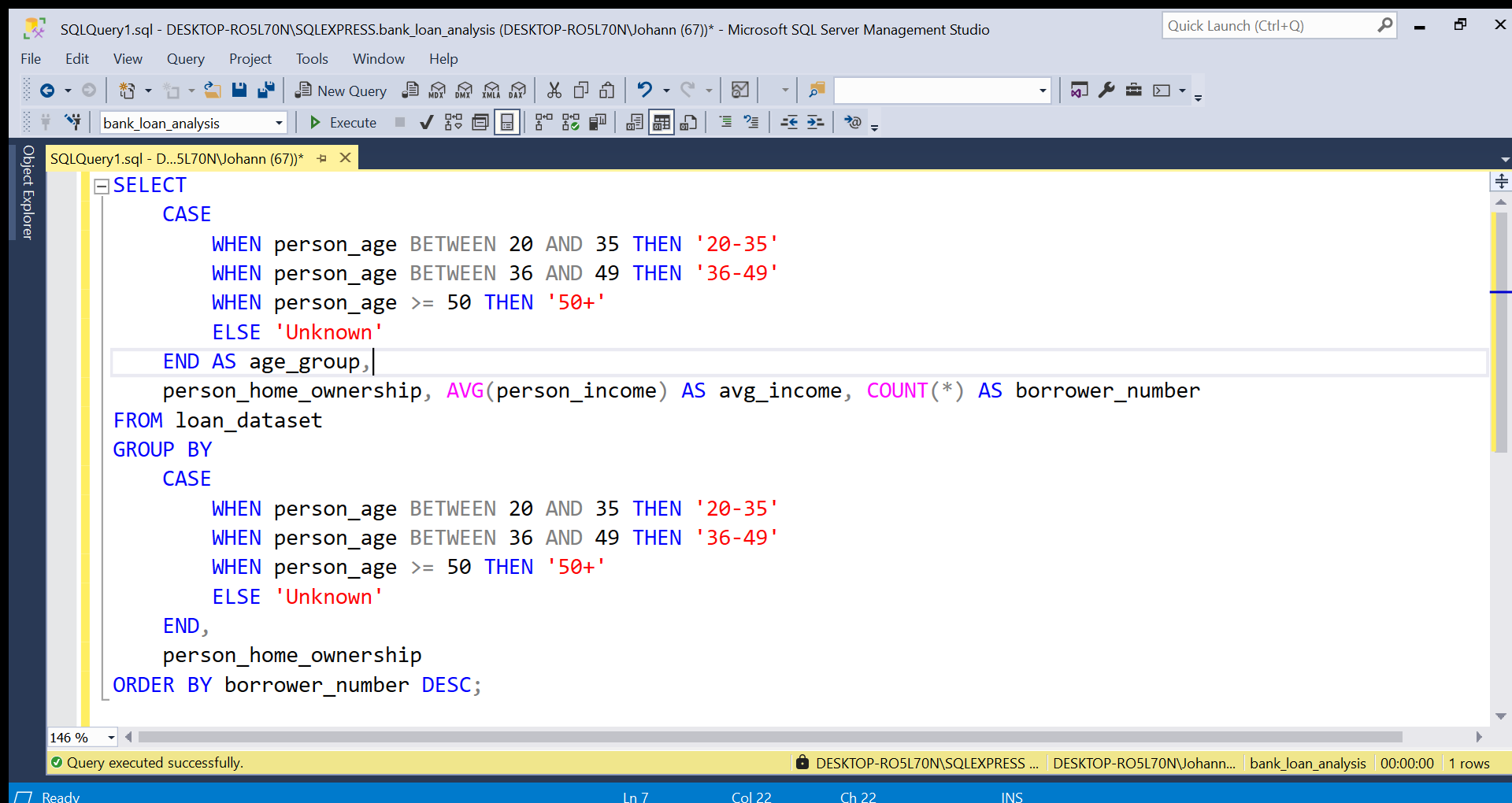


Figure 13 : Distribution of Socio-Economic Profiles Among Borrowers

**Result**:

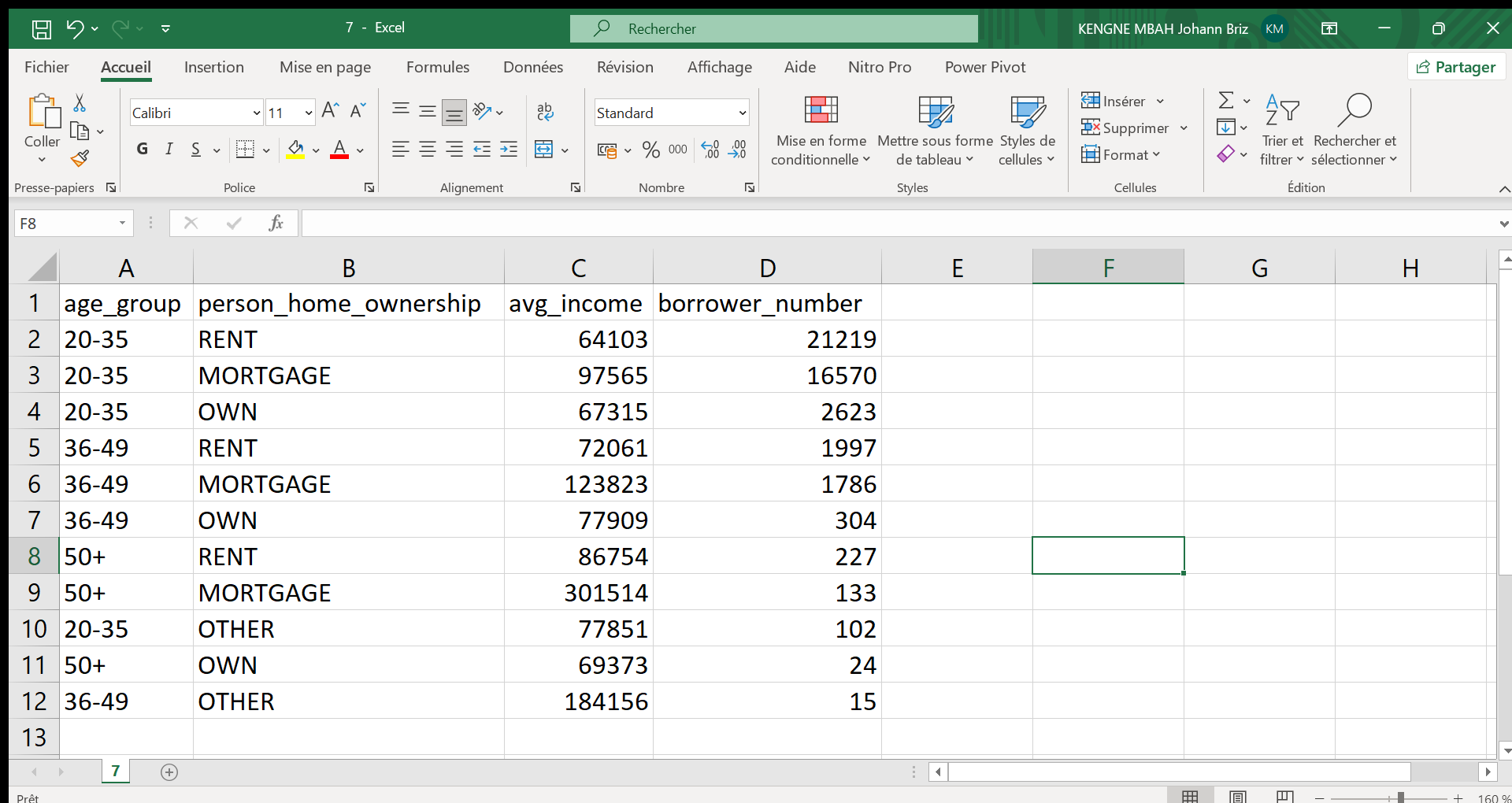


Figure 14 : Distribution of Socio-Economic Profiles Among Borrowers Result

**Illustration:**

**Interpretation:**

* **Young Professionals** have moderate income across all property categories, but their mortgage holders have the highest average income, which is expected as mortgages often signify more financial stability.
* **Middle-Aged Professionals** tend to have higher incomes overall, especially among those with mortgages. They have a good balance of homeownership and mortgage holders, showing financial stability and career maturity.
* **Experienced Borrowers** have the highest incomes, especially those with mortgages. This could reflect long career experience, accumulated wealth, or investments in higher-value properties. Renters in this group also show relatively high-income levels, suggesting financial flexibility.

Understanding these profiles allows HSBC to tailor their loan offerings, risk assessments, and marketing strategies to the most common borrower segments. It also provides insight into the socio-economic diversity of HSBC loan portfolio.

1. **Loan performance**
2. **Types of Loans with the Best Interest Rates**

Understanding which types of loans carry the best interest rates is crucial for HSBC to evaluate its pricing strategy. By identifying the loan types with the lowest interest rates, the bank can ensure that its pricing is competitive, while also optimizing profitability for different types of borrowers.

**SQL query:**

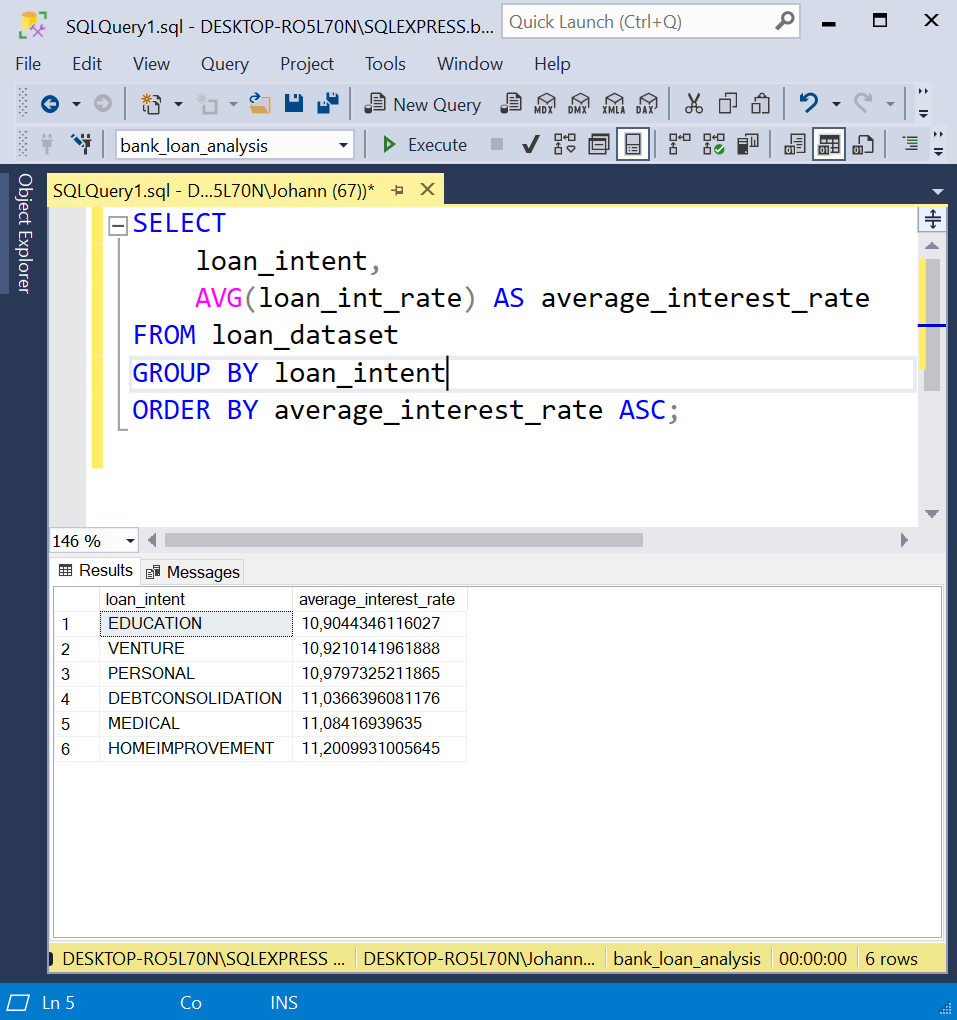


Figure 15 : Types of Loans with the Best Interest Rates

**Result**:

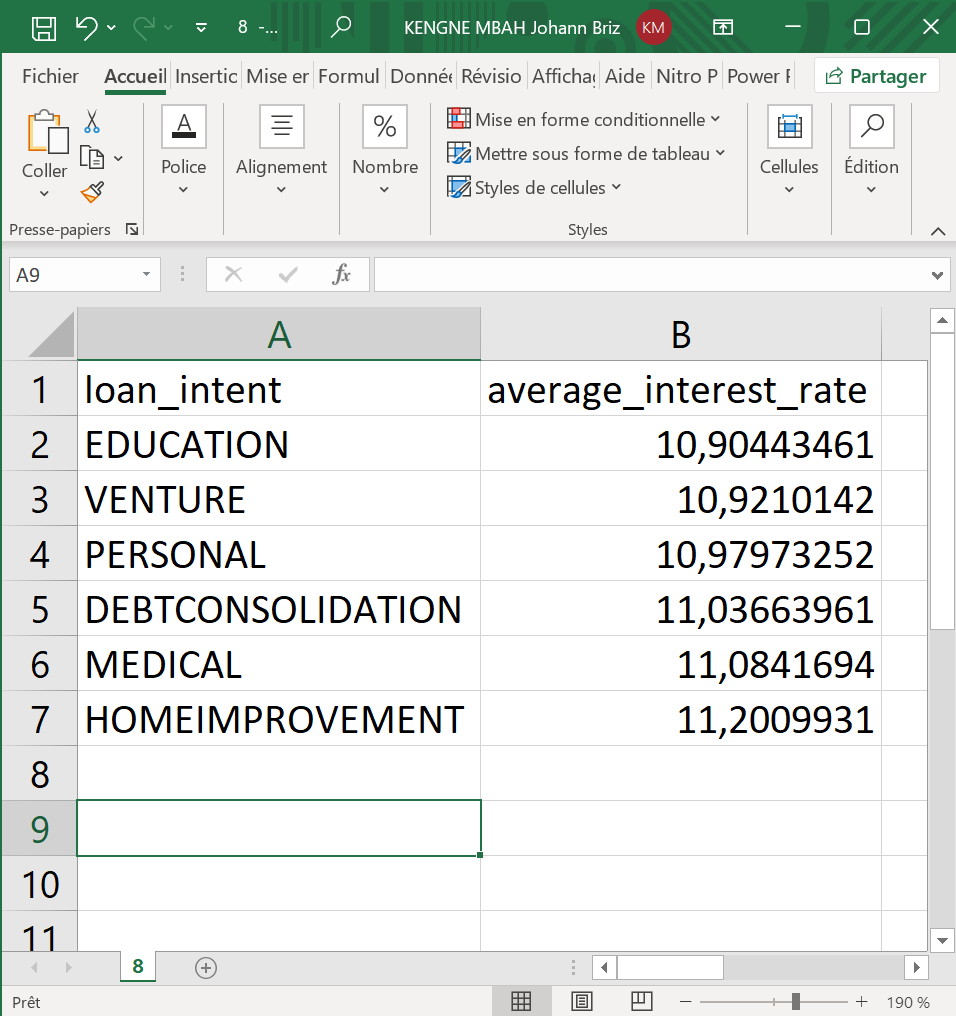


Figure 16 : Types of Loans with the Best Interest Rates Result

**Illustration:**

**Interpretation:**

By analysing the results, we can see that the average interest rate is approximatively the same for the different types of loans.

1. **Evolution of Interest Rates Based on Loan Amounts**

Understanding how interest rates evolve based on the amount borrowed can help HSBC optimize its lending strategy. By analysing the relationship between loan amounts and interest rates, the bank can better assess whether larger loans are associated with higher or lower interest rates, and adjust its pricing strategies accordingly to attract specific borrower profiles.

**SQL query:**

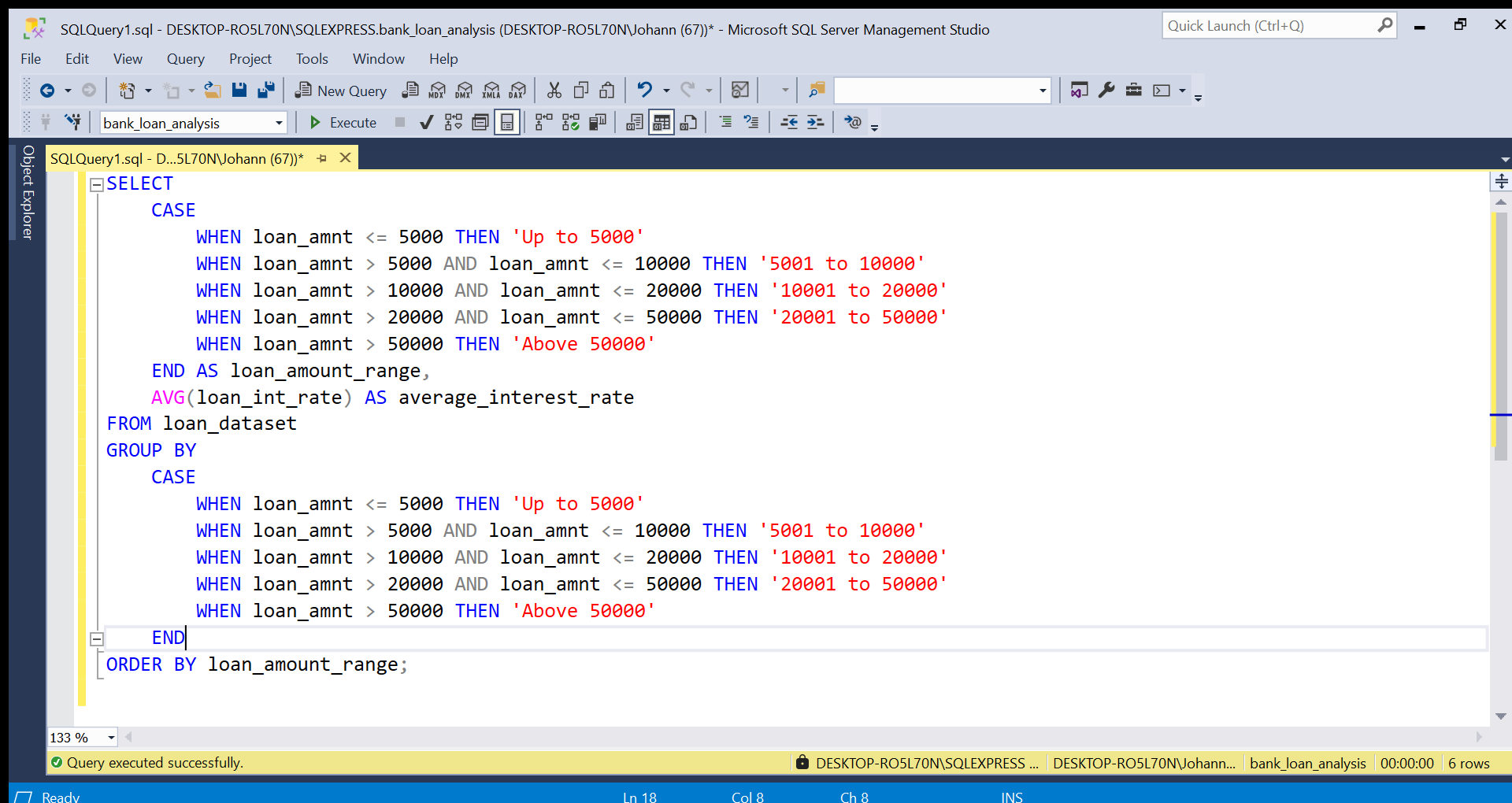


Figure 17 : Evolution of Interest Rates Based on Loan Amounts

**Result :**

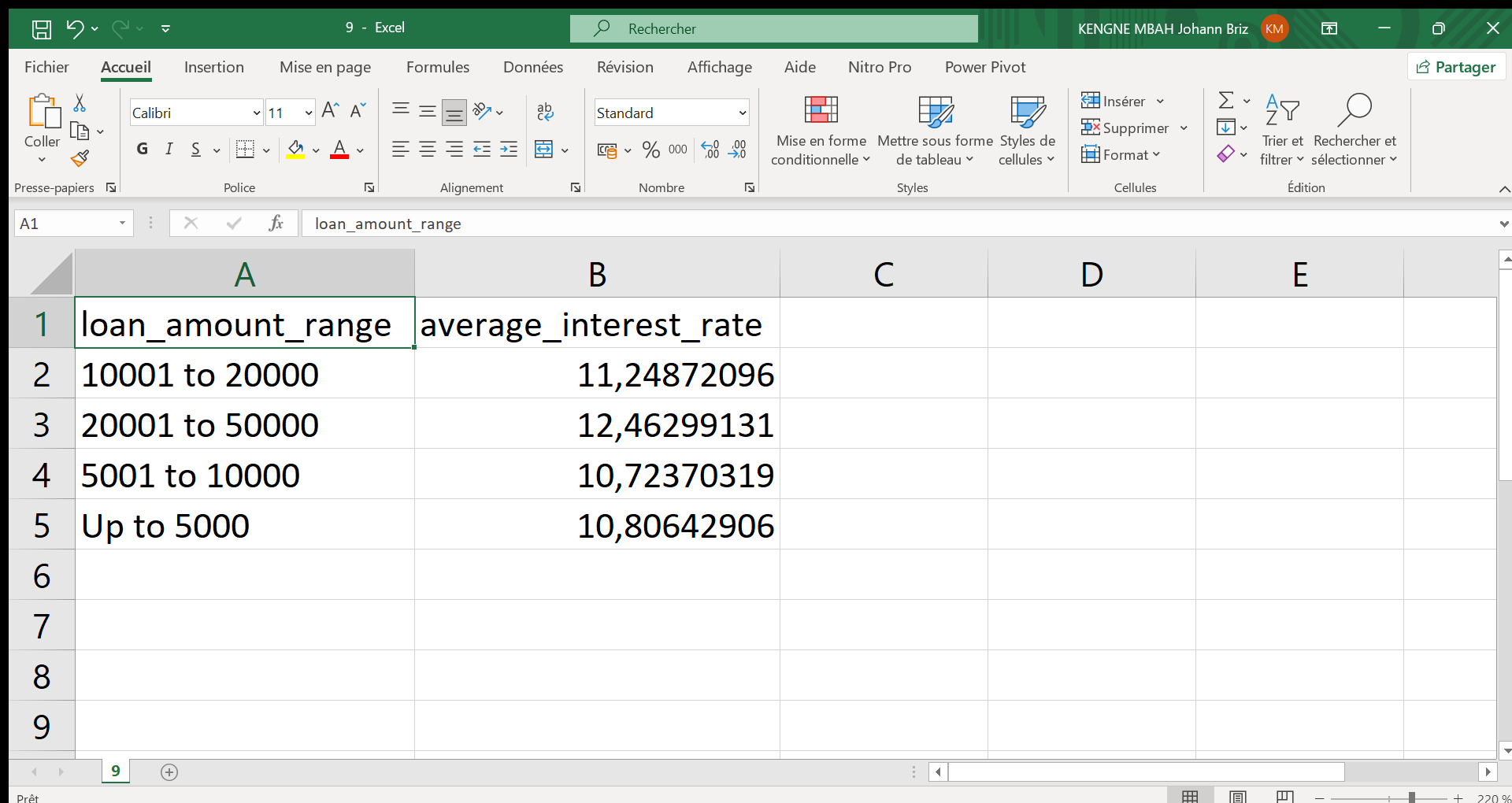


Figure 18 : Evolution of Interest Rates Based on Loan Amounts Result

**Illustration:**

**Interpretation:**

From the results, we can observe:

* Relatively stable rates for small loans: Interest rates are fairly similar for loans up to $10,000.
* Significant increase for larger loans: There is a notable increase in the interest rate for loans over $10,000, and even more so for those over $20,000.

The analysis indicates that **larger loans** tend to have **higher interest rates**, which may be due to the perceived **higher risk** associated with lending larger amounts. Borrowers requesting larger loans may be seen as riskier, or the bank may charge higher rates to offset the potential risk of non-repayment.

1. **Most Profitable Customer Segments**

Identifying the most profitable customer segments is essential for HSBC to refine its marketing and lending strategies. By understanding which borrower profiles bring in the most revenue, the bank can tailor its offerings and optimize its efforts to target high-value customers, while maintaining a balanced risk profile.

**SQL query:**

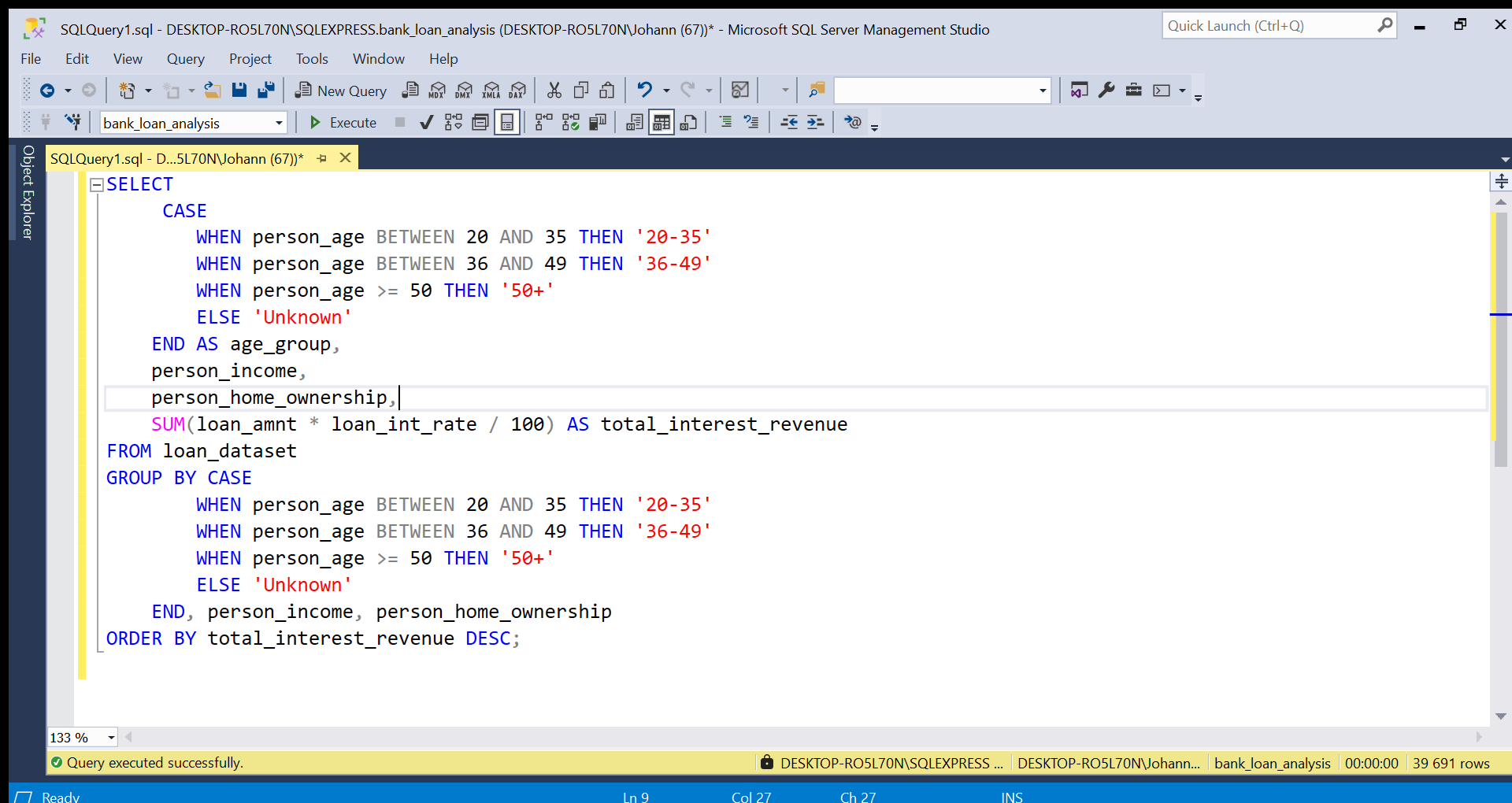


Figure 19 : Most Profitable Customer Segments

**Result :**

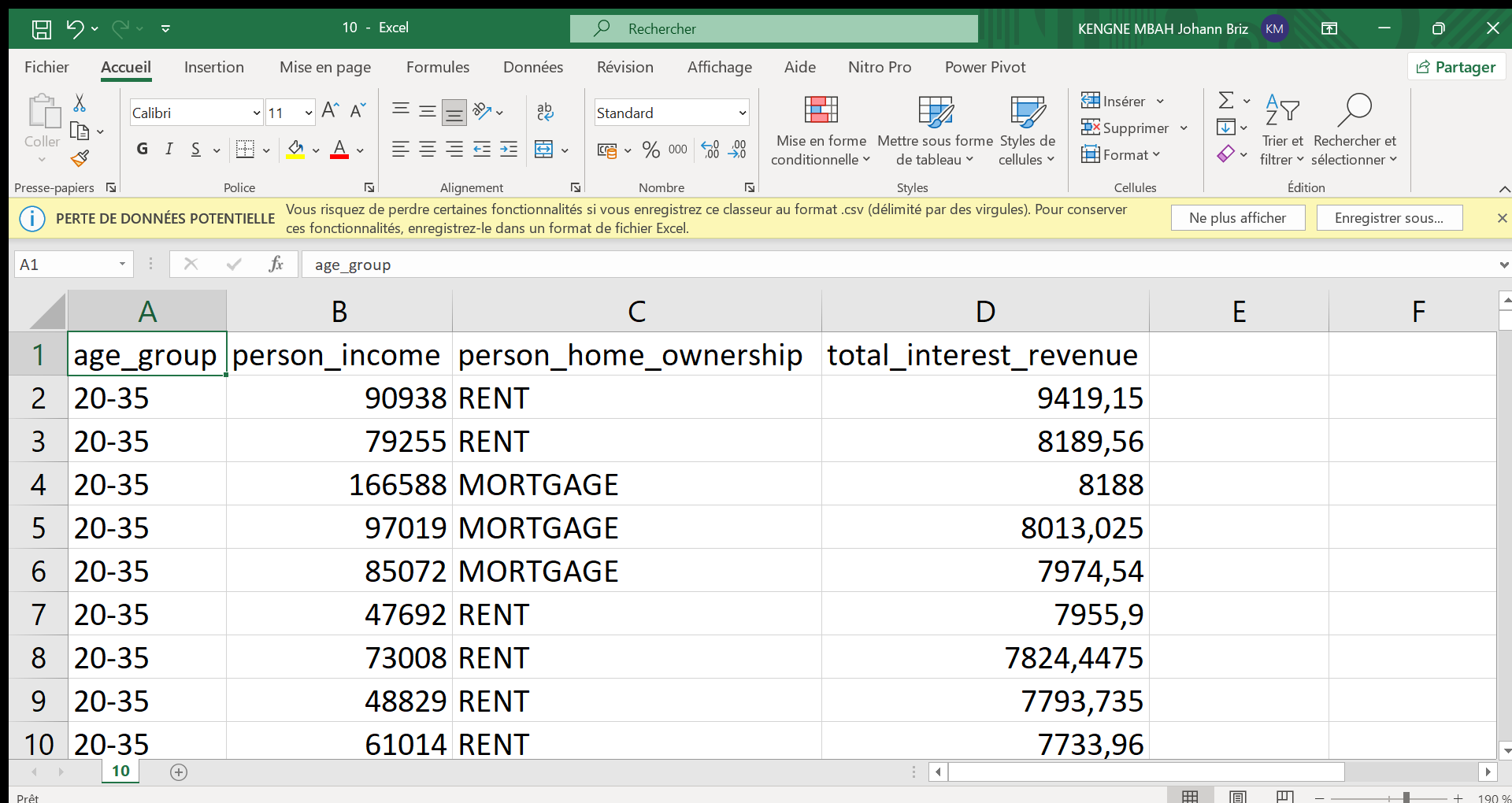


Figure 20 : Most Profitable Customer Segments Result

**Illustration:**

**Interpretation:**

From the results, we can see that the most profitable customer segments are typically **young professionals (20-35 years old)** customers with **rent and mortgage ownership**.

For HSBC, targeting young professionals’ borrowers with rent or mortgage ownership could lead to higher profitability. Offering competitive rates and targeted marketing to these segments can be a strategic approach to maximize revenue while maintaining a balanced risk profile.

**CONCLUSION**

The study highlights key trends in borrower behaviour and loan structuring at HSBC. The main challenge lies in optimizing credit criteria and revising pricing policies to maximize profitability while maintaining controlled risk. A more detailed analysis of risk factors would make banking offerings more competitive.