POWER BI PROJECT

**DATA CLEANING (FIRST PHASE)**

The data source is completely row. We are not going to make changes in data source ie the Excel file.

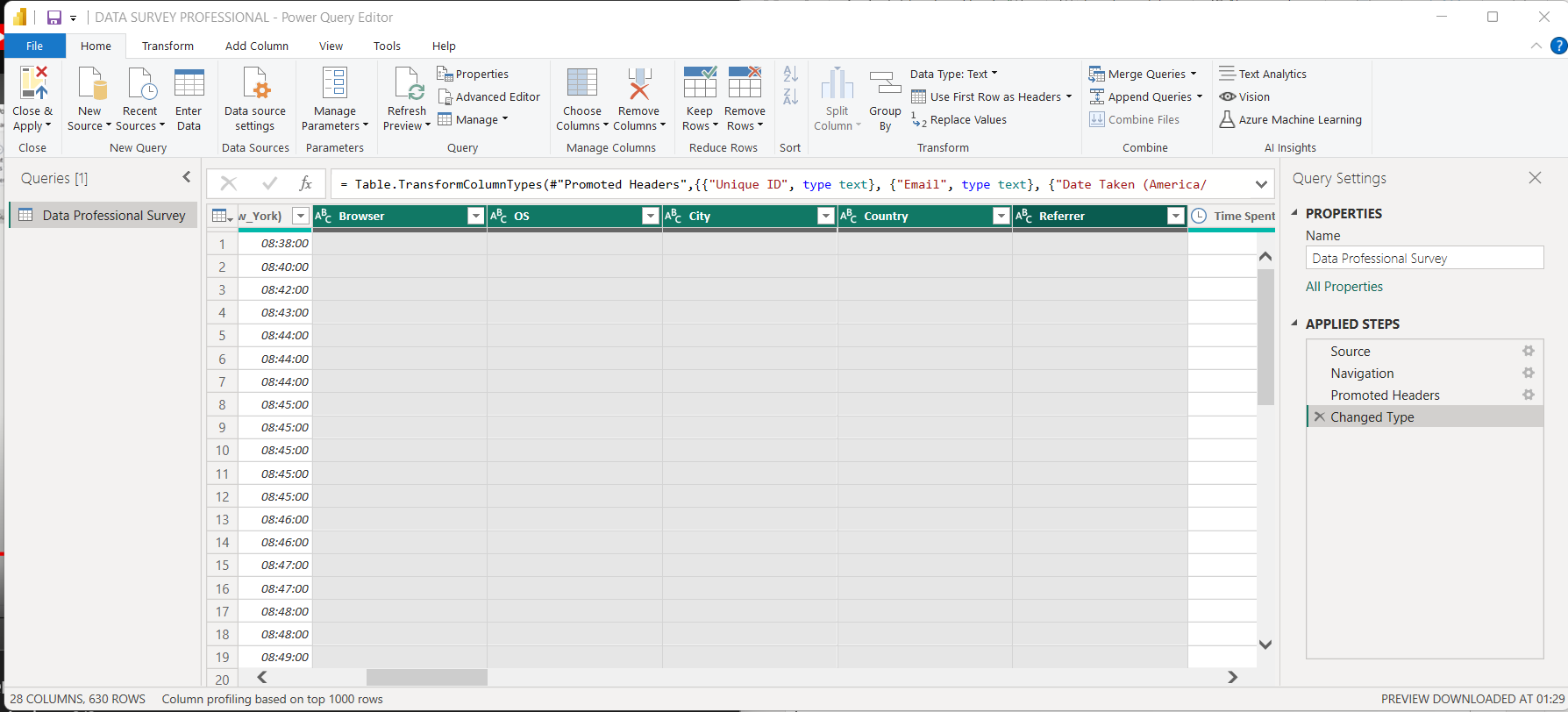
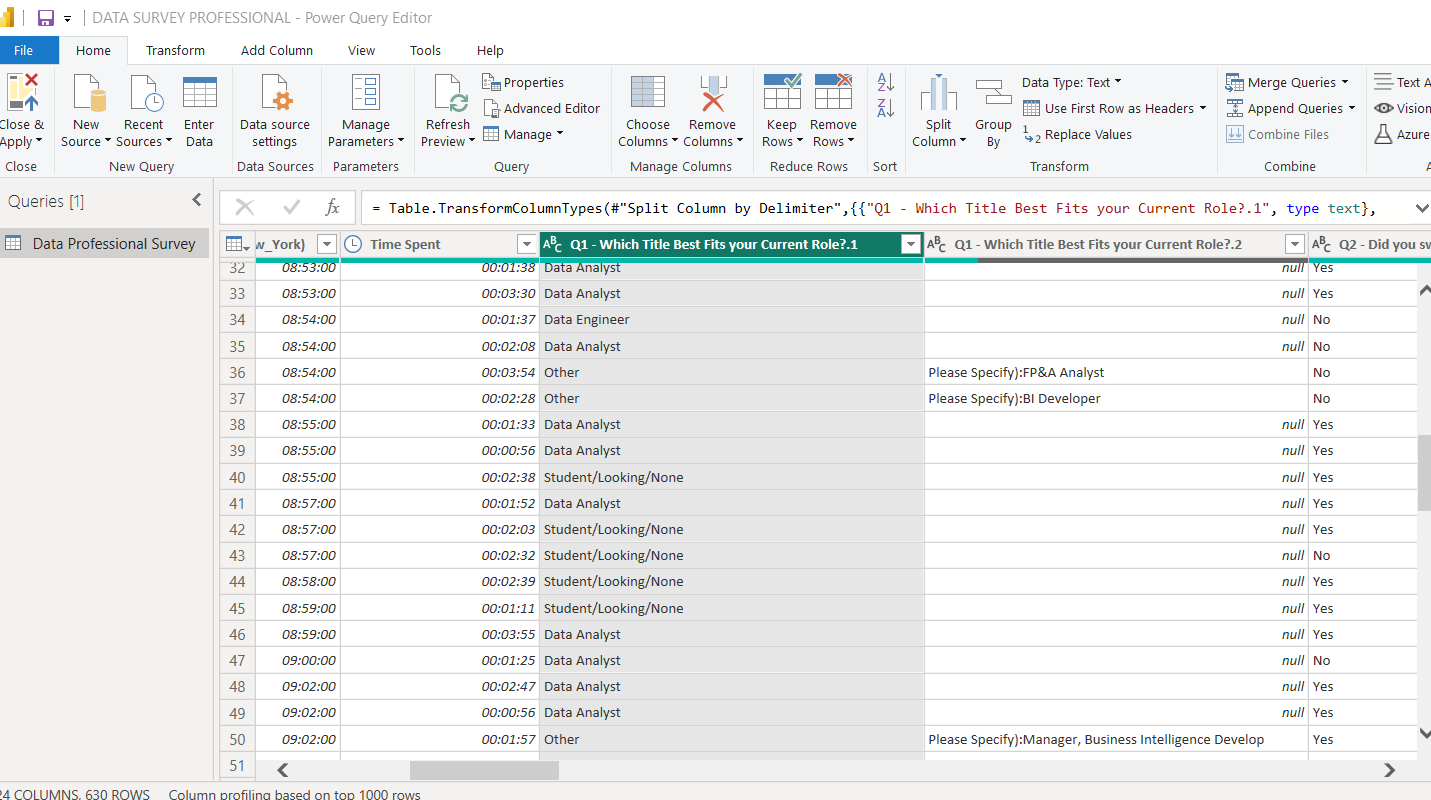
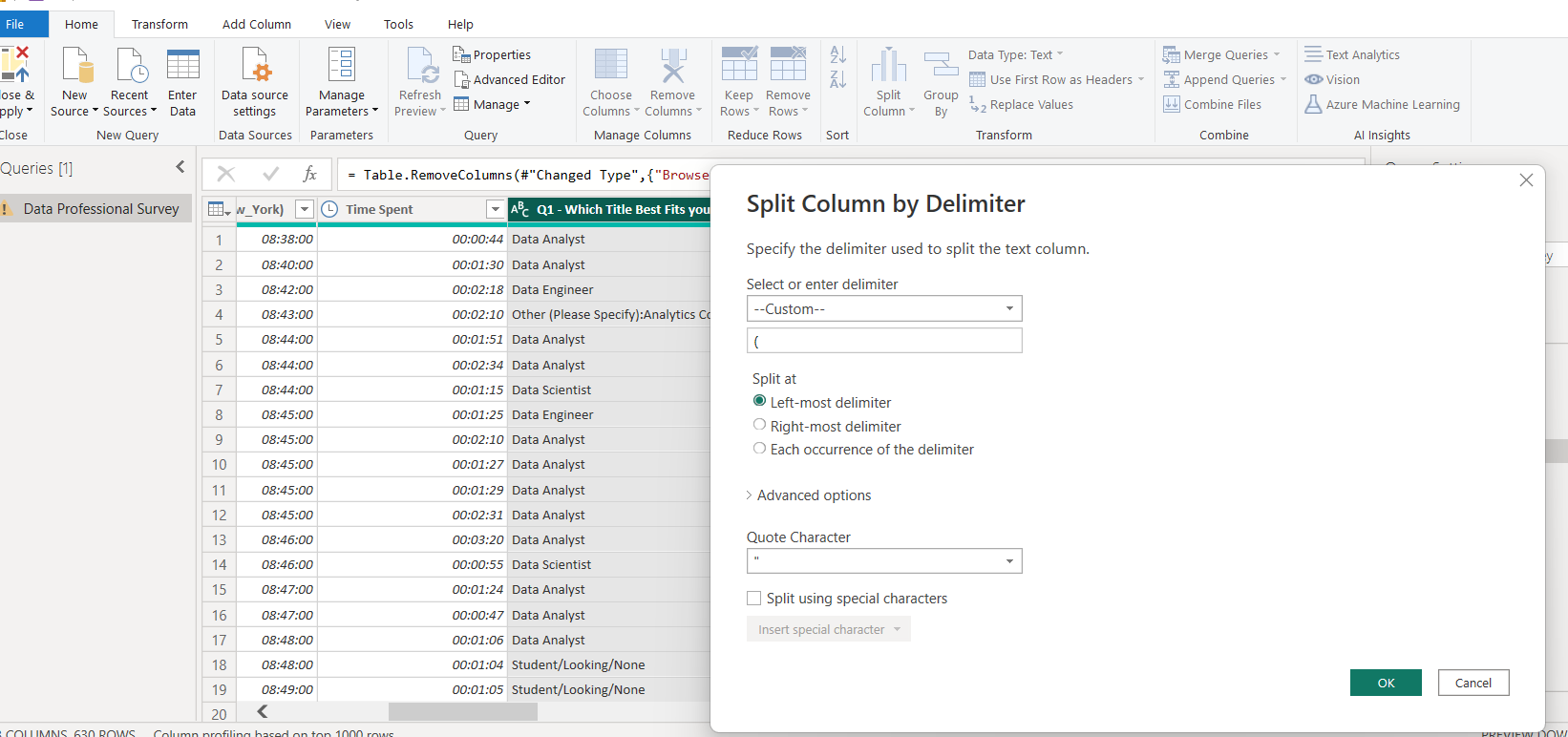
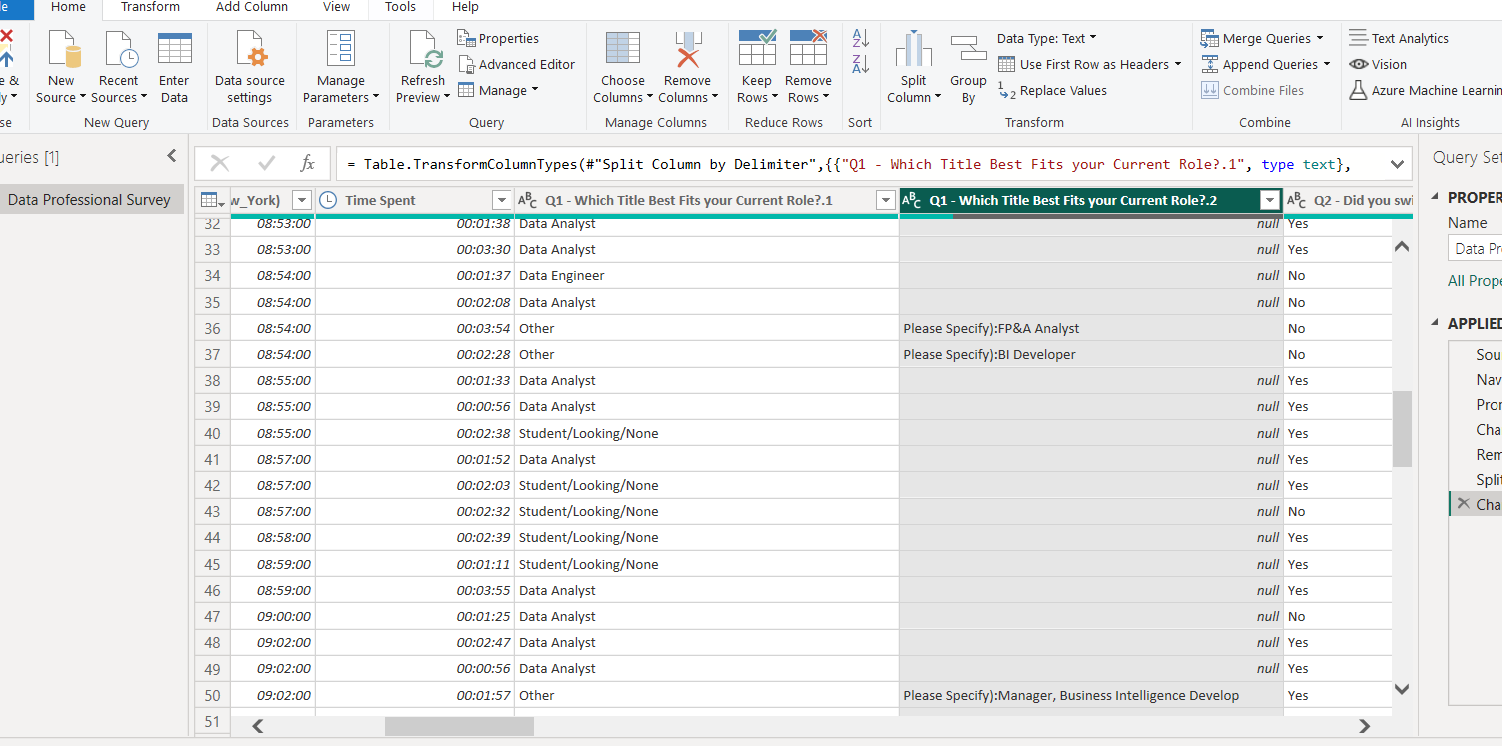
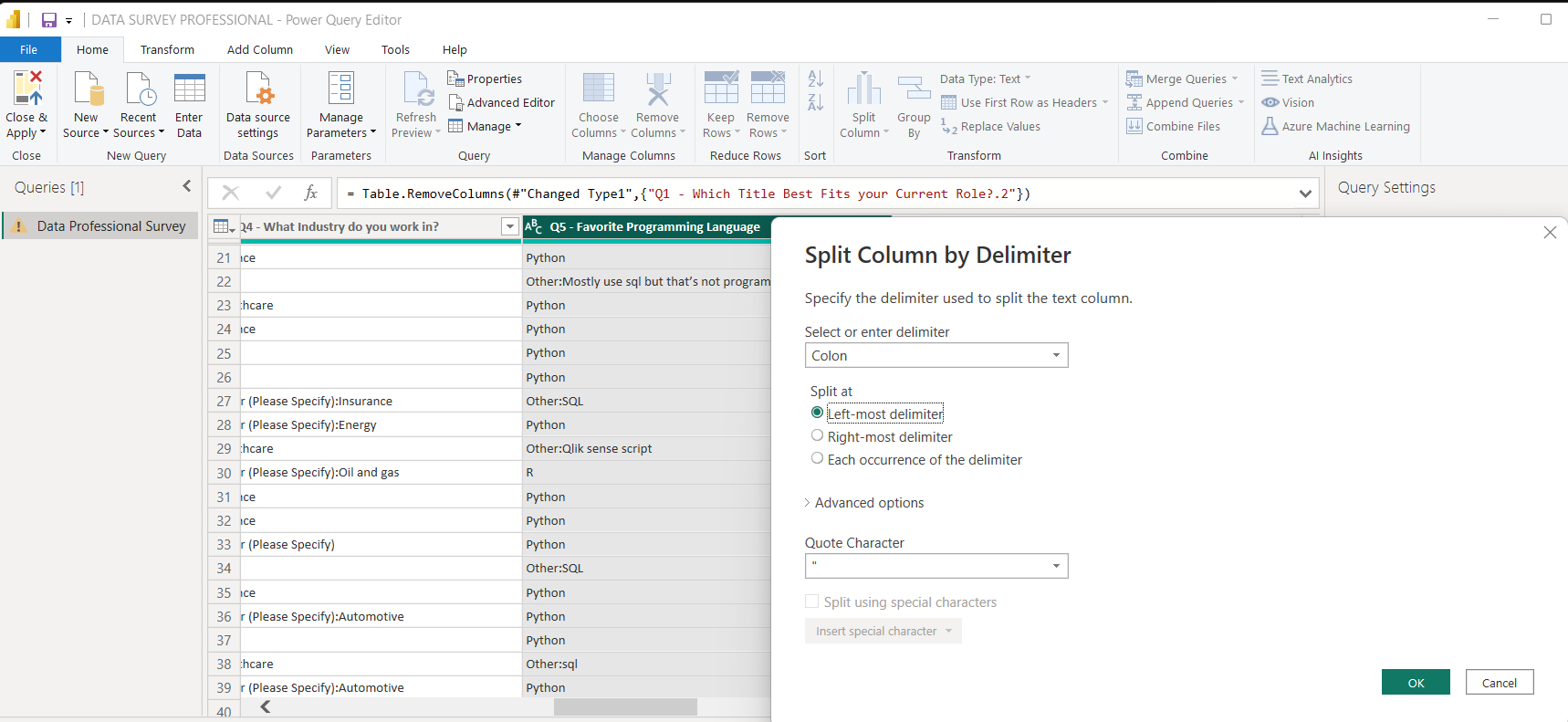
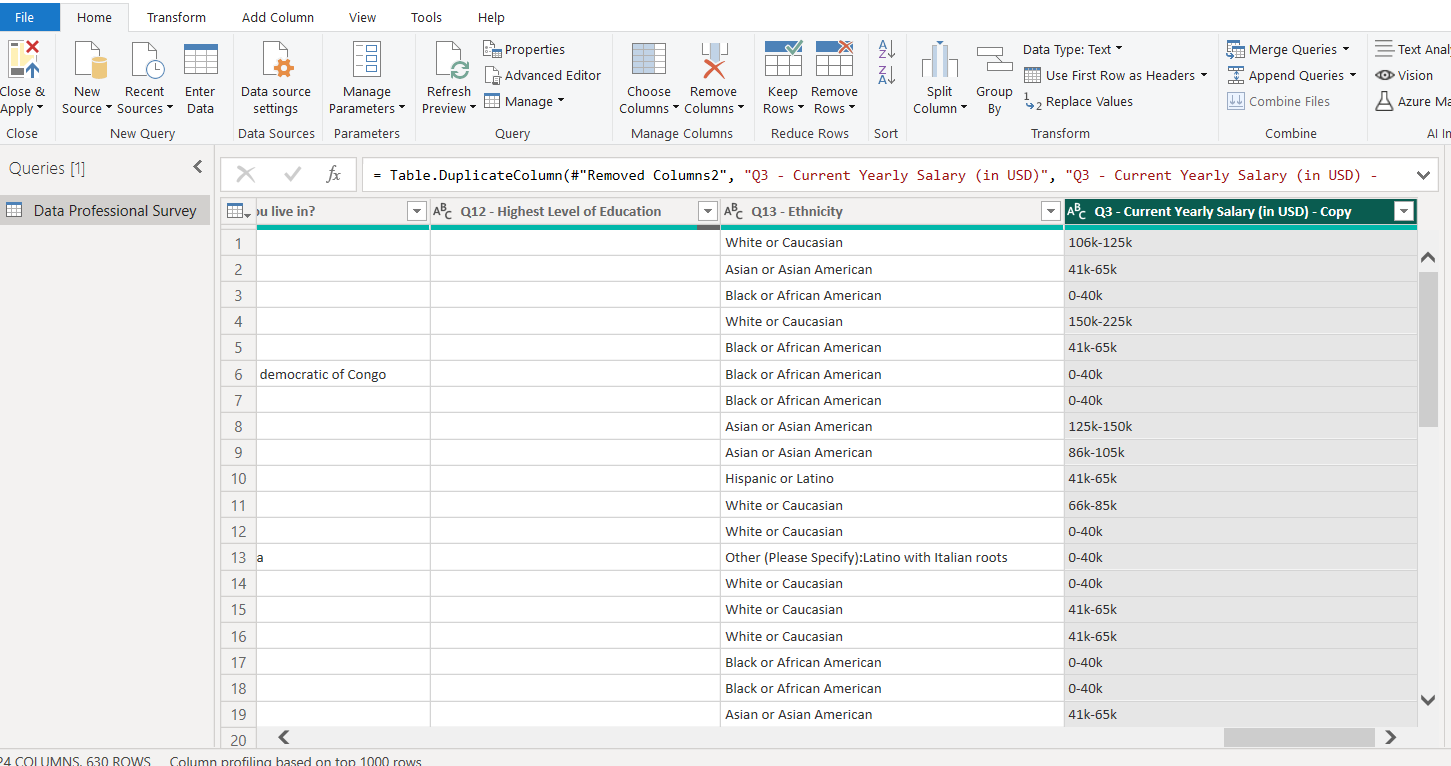
The transformation will be done in Power BI Transform not in in Excel.

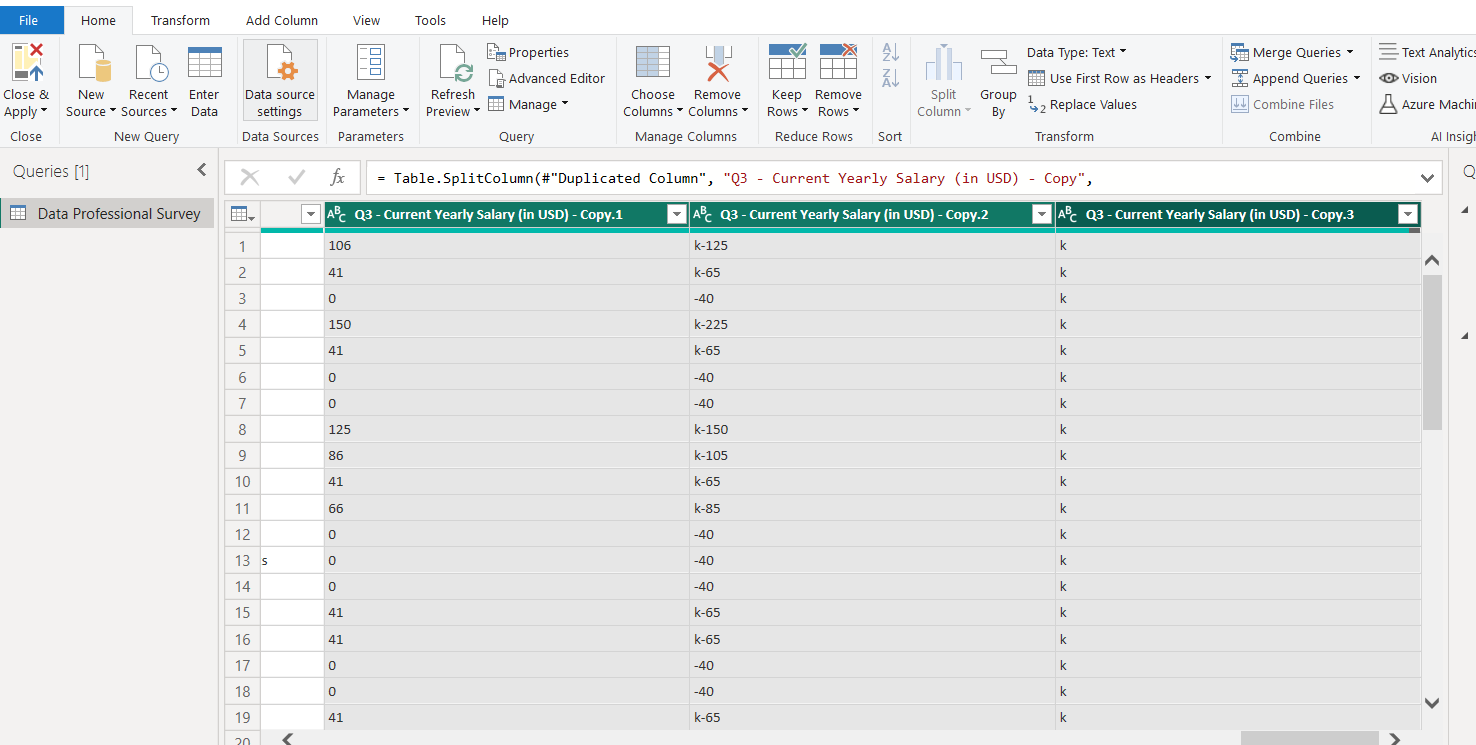
The column Q1 has Other attribute value has lot of data…that needs to be cleaning

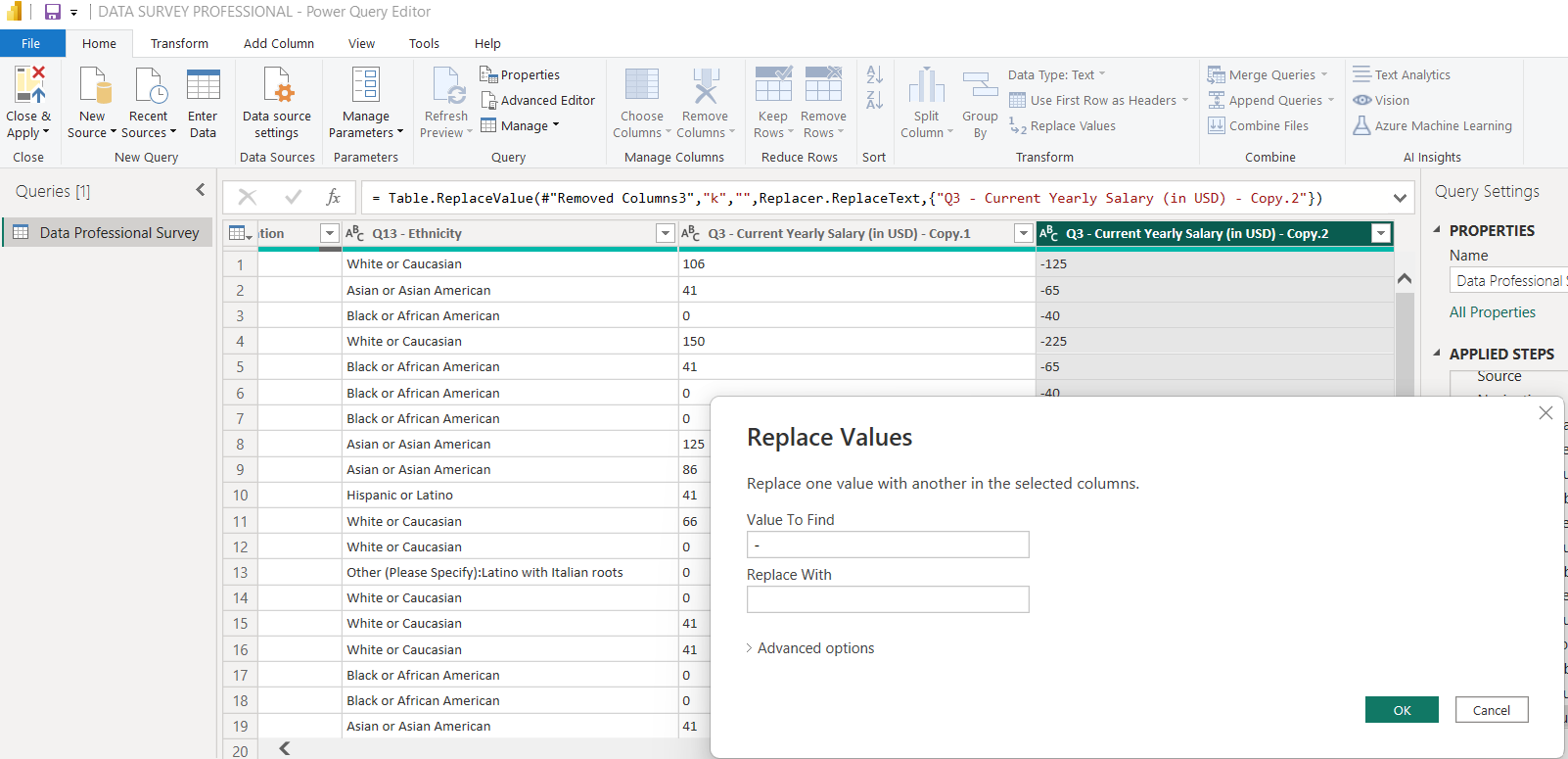
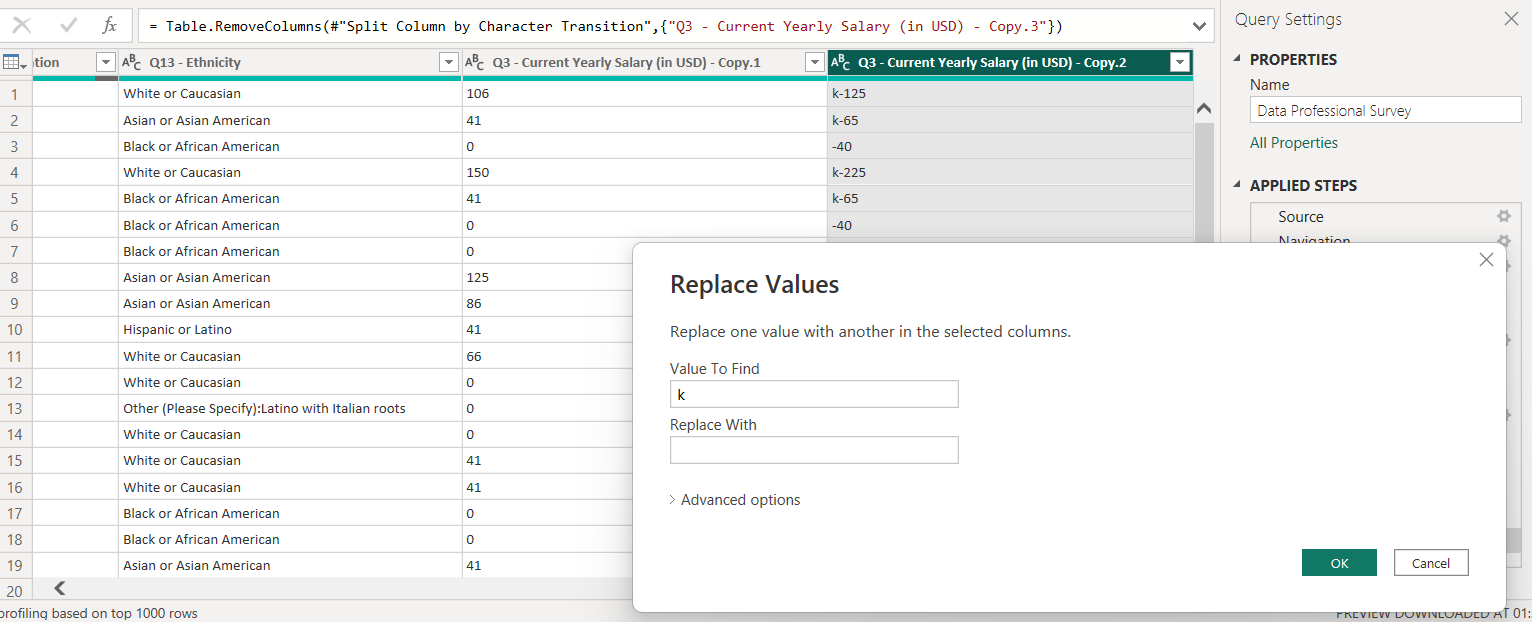
The column Q2 has yearly salary in ranges. So that’s needs to be cleaned.

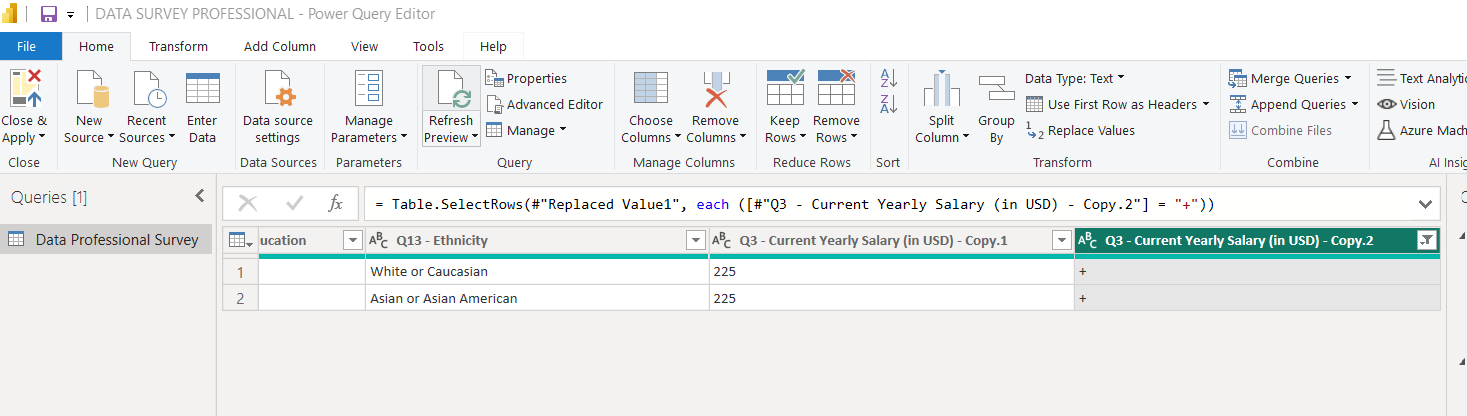
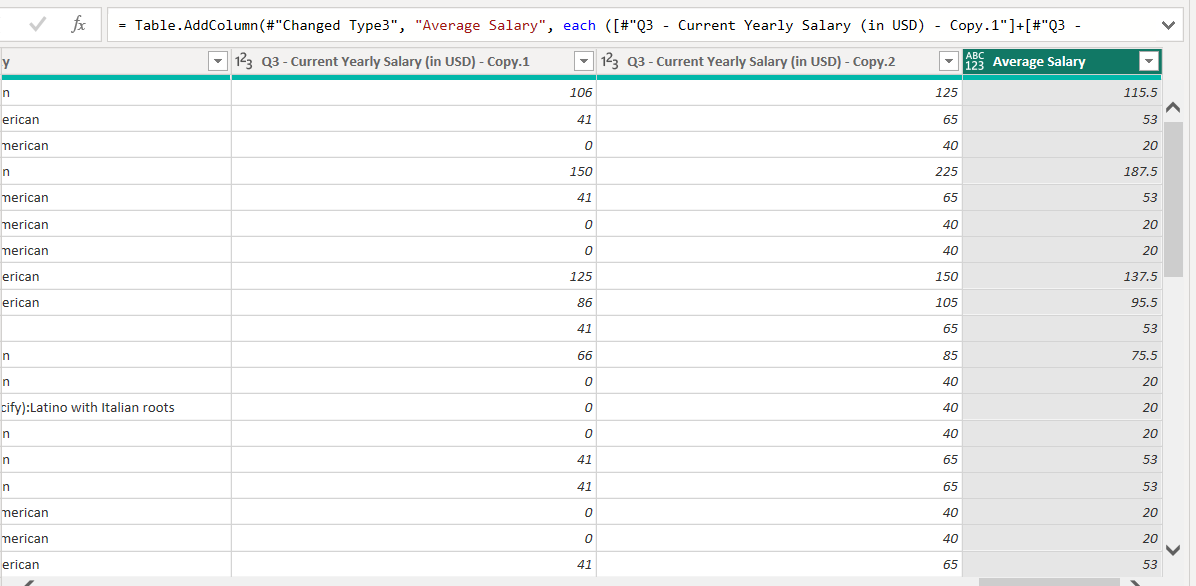
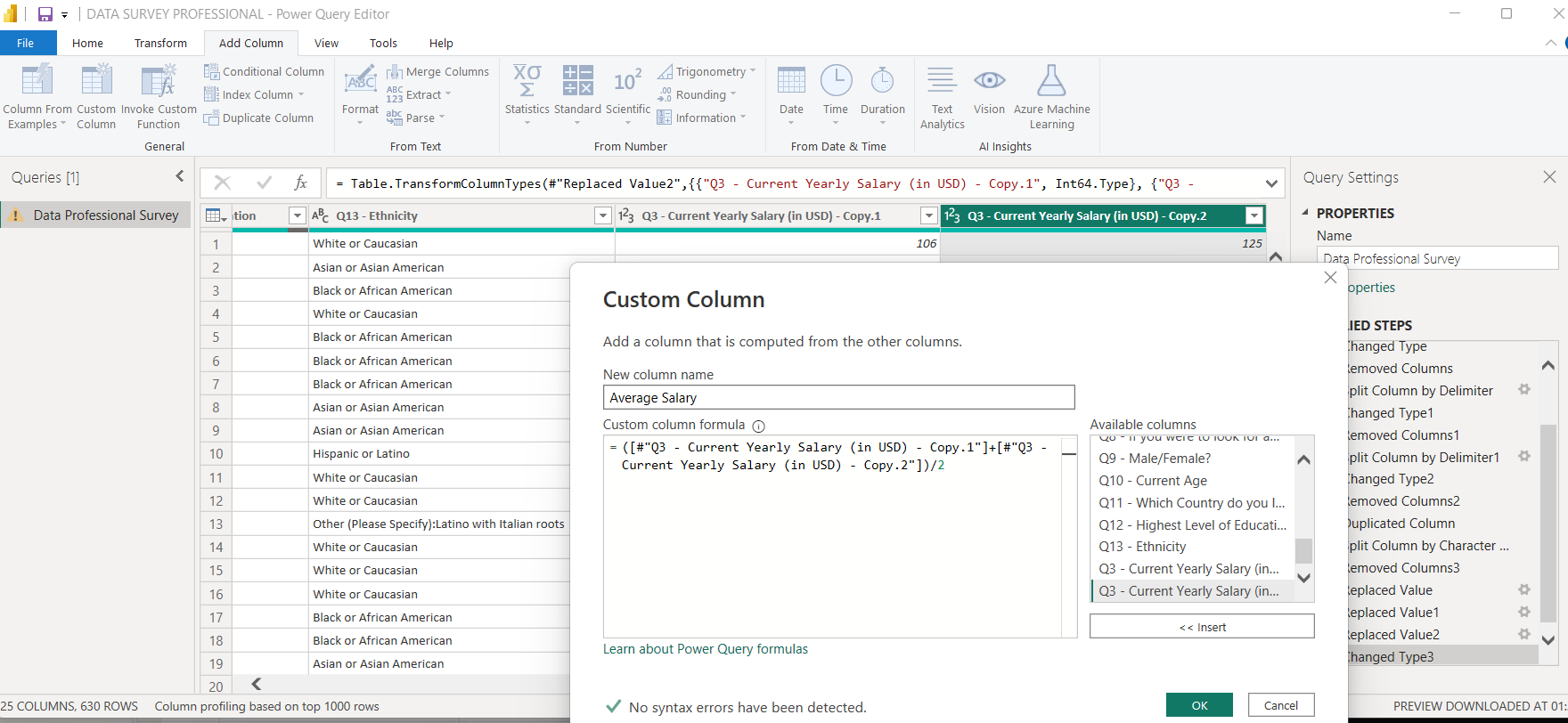
Lets get started.

**First Transform data.**

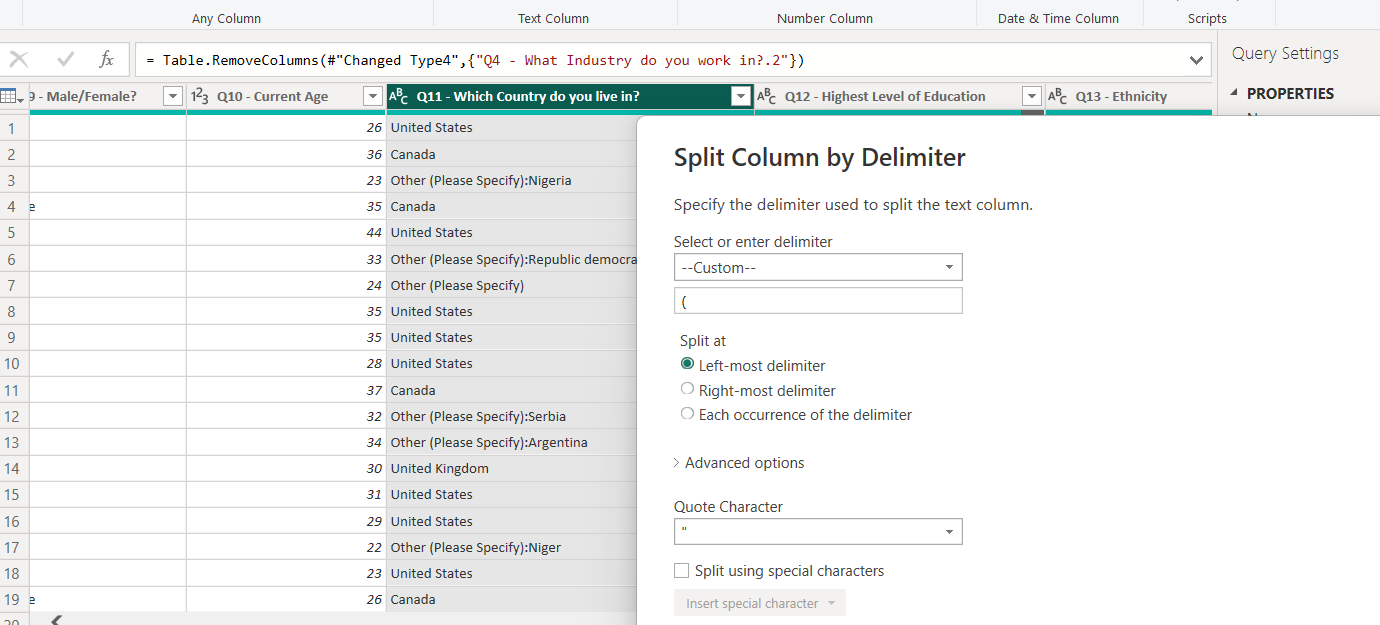
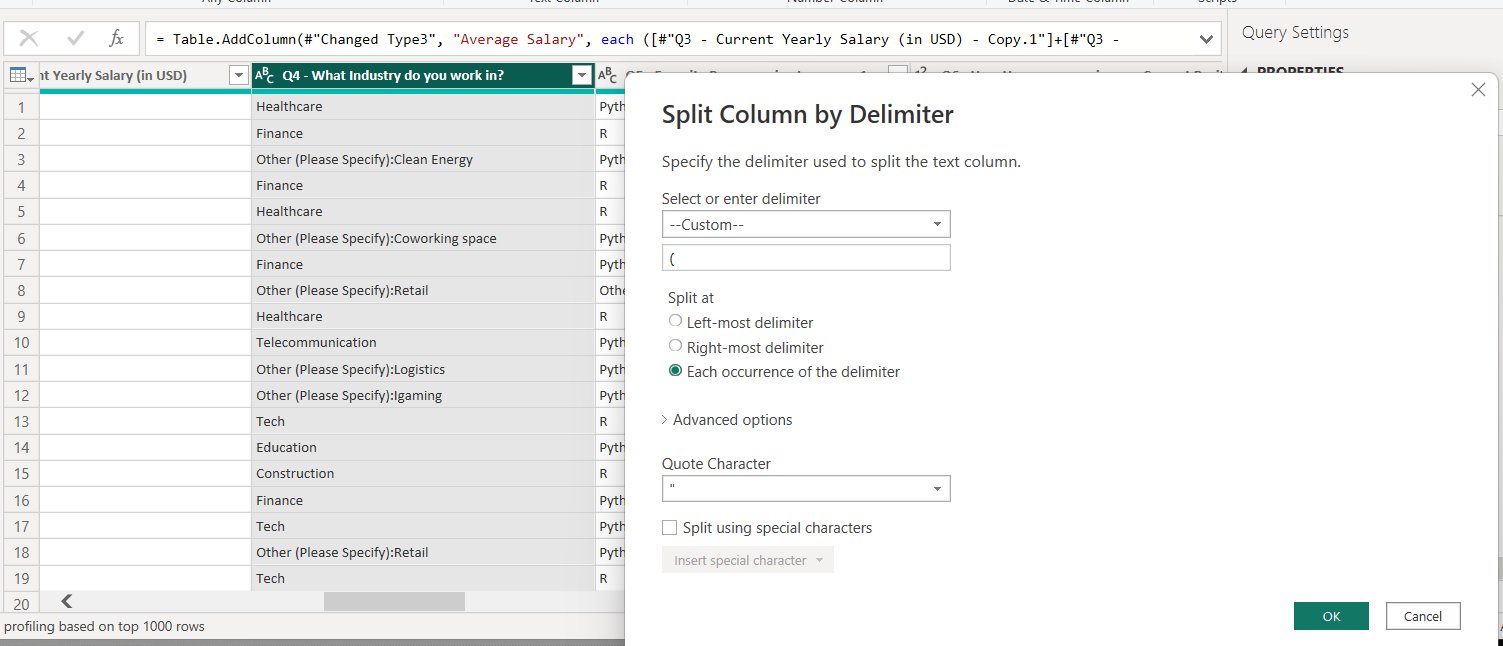
1. We delete the columns as shown below. 
2. The Q1 columns. We need to remove those extra values attached to value=Other. So we create split column based on delimiter ‘(‘ . 
3. We delete the newly splited column. 
4. The above two steps is applied for column Q5 Favourite Programming Language. We split the others from original column. And delete the new column. Note the delimiter (colon). 
5. Now we need clean the Q3 salary. We create duplicate copy of that column which appears at end. And we split that copied column by Digit to Non-Digit.

We remove the column that has k’s.

Then we look at Q3..Copy 2. Here we need to remove k , then remove ‘-‘ and some have only + which needs to be by corresponding value of that row. 

1. As you see two rows have + sign so replace with the value from column Copy1. 
2. Convert the type of Copy1 and Copy2 column into Whole Numbers. Then we take average of both in new column Average Salary. And delete the Copy1 and Copy2 columns. 

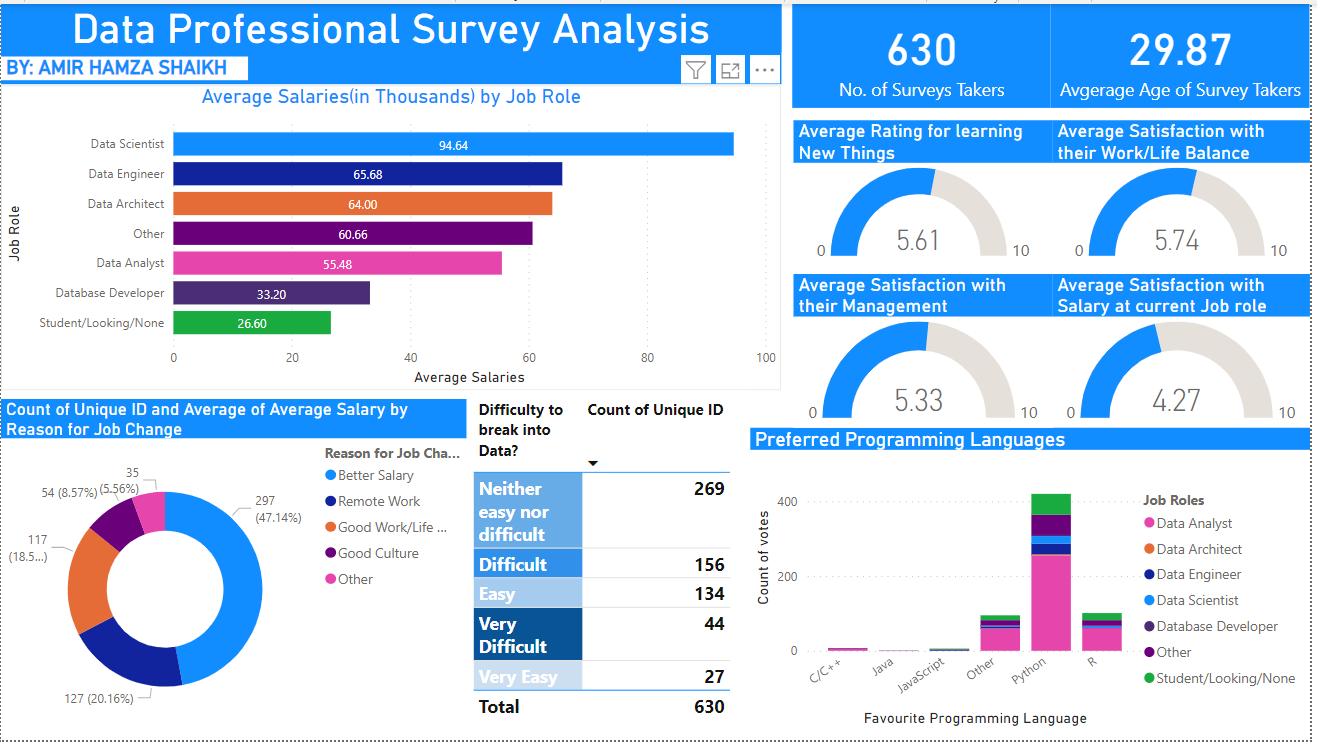
Make sure that Avgerage Salary is in Whole Number type.

1. Similary try to remove the Others from The column Q11 and Q4. 

This enough for Data Cleaning. The way taking insights from Salary range might not be so accurate in taking averages. But we this method/ way as simple.

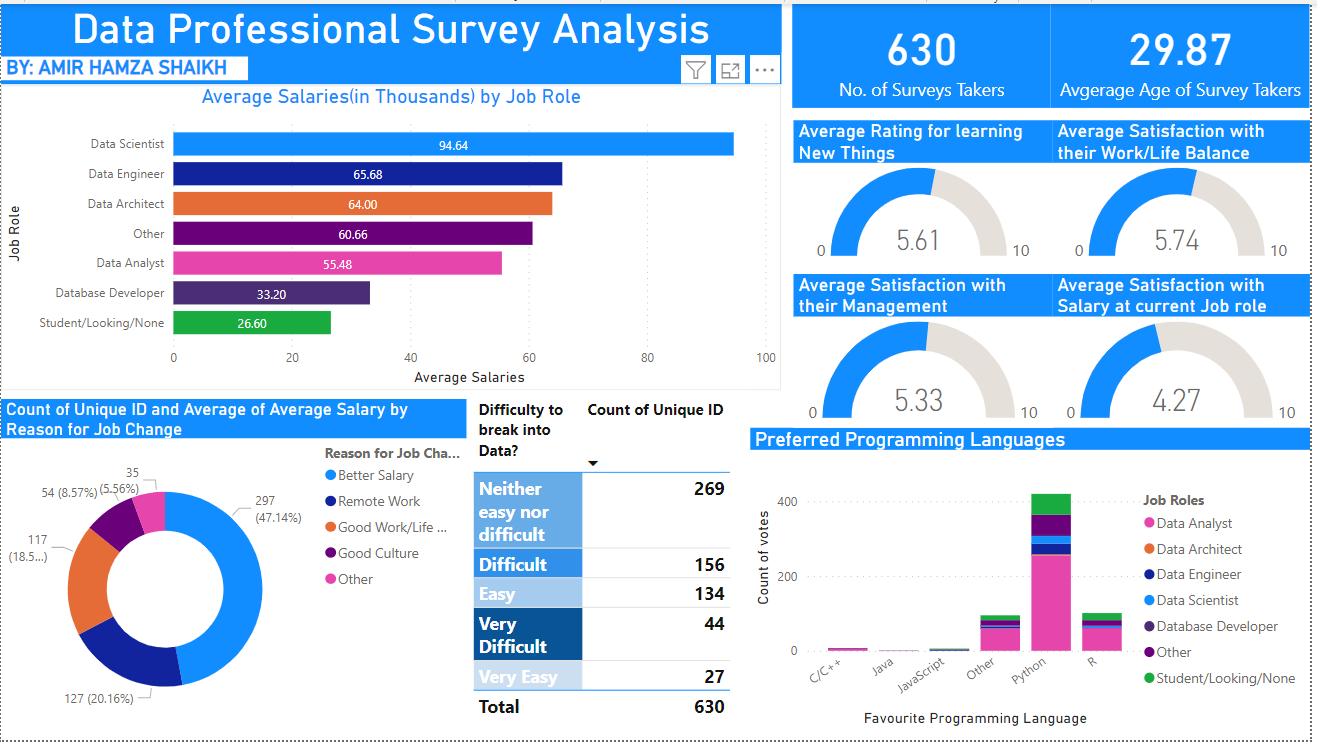
Then click on Close&Apply button.

**Second INSIGHTS data visualization .(SECOND PHASE)**



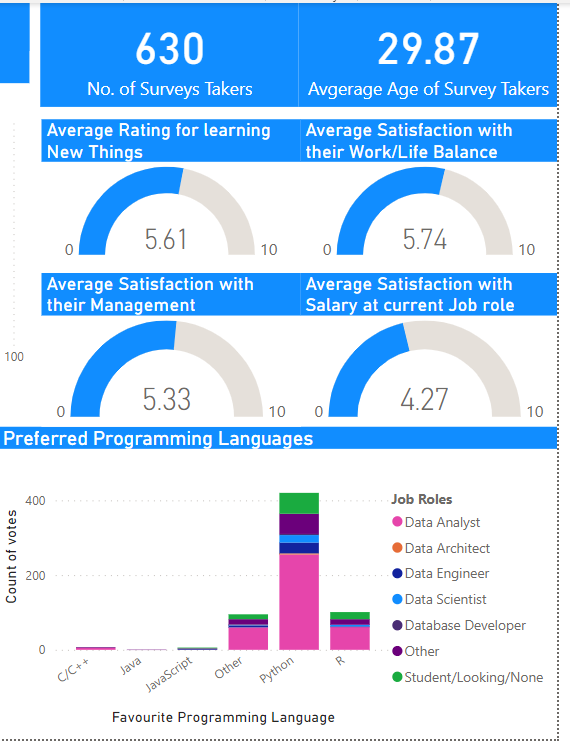
**Observations:**

**1)**



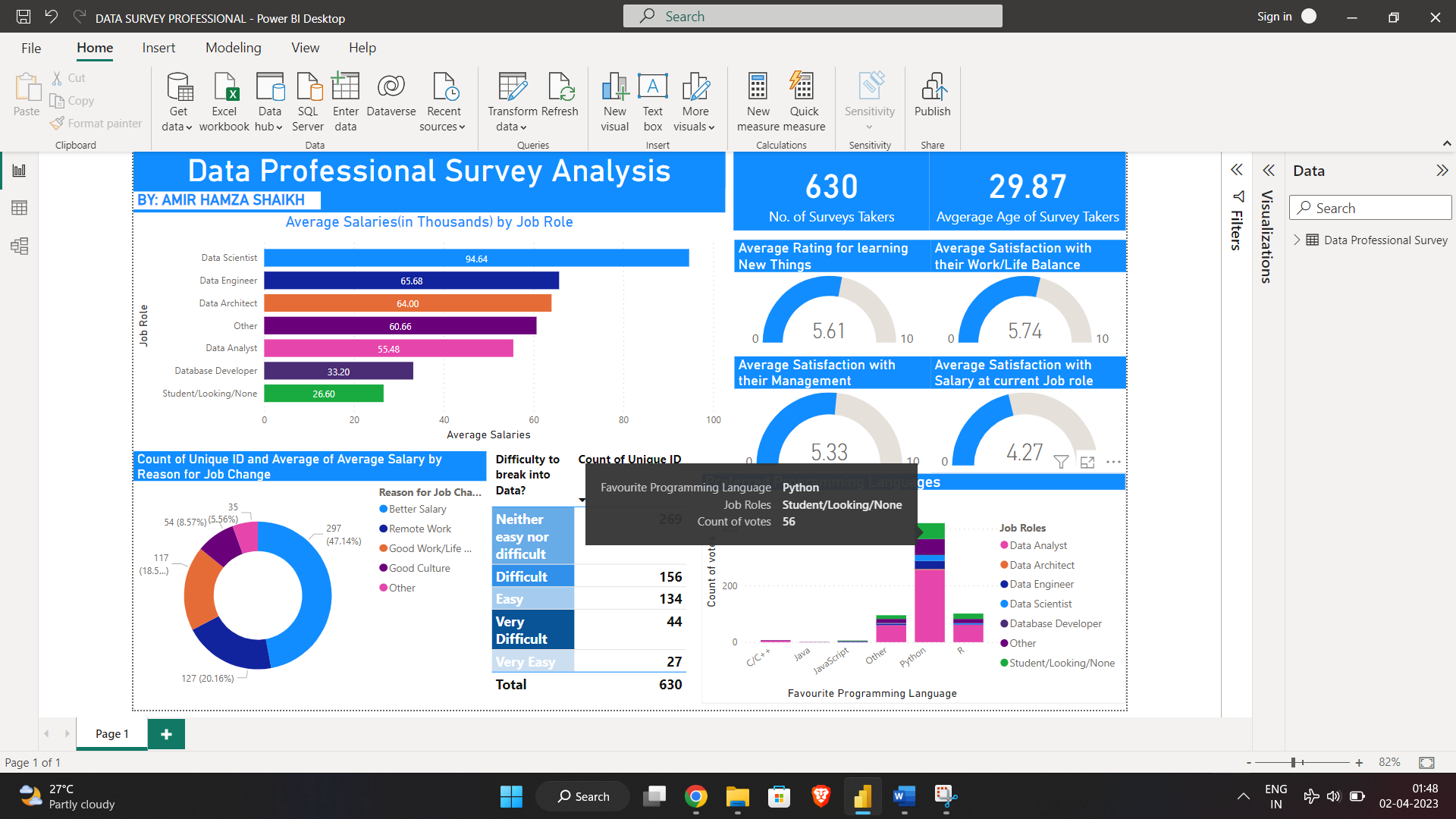
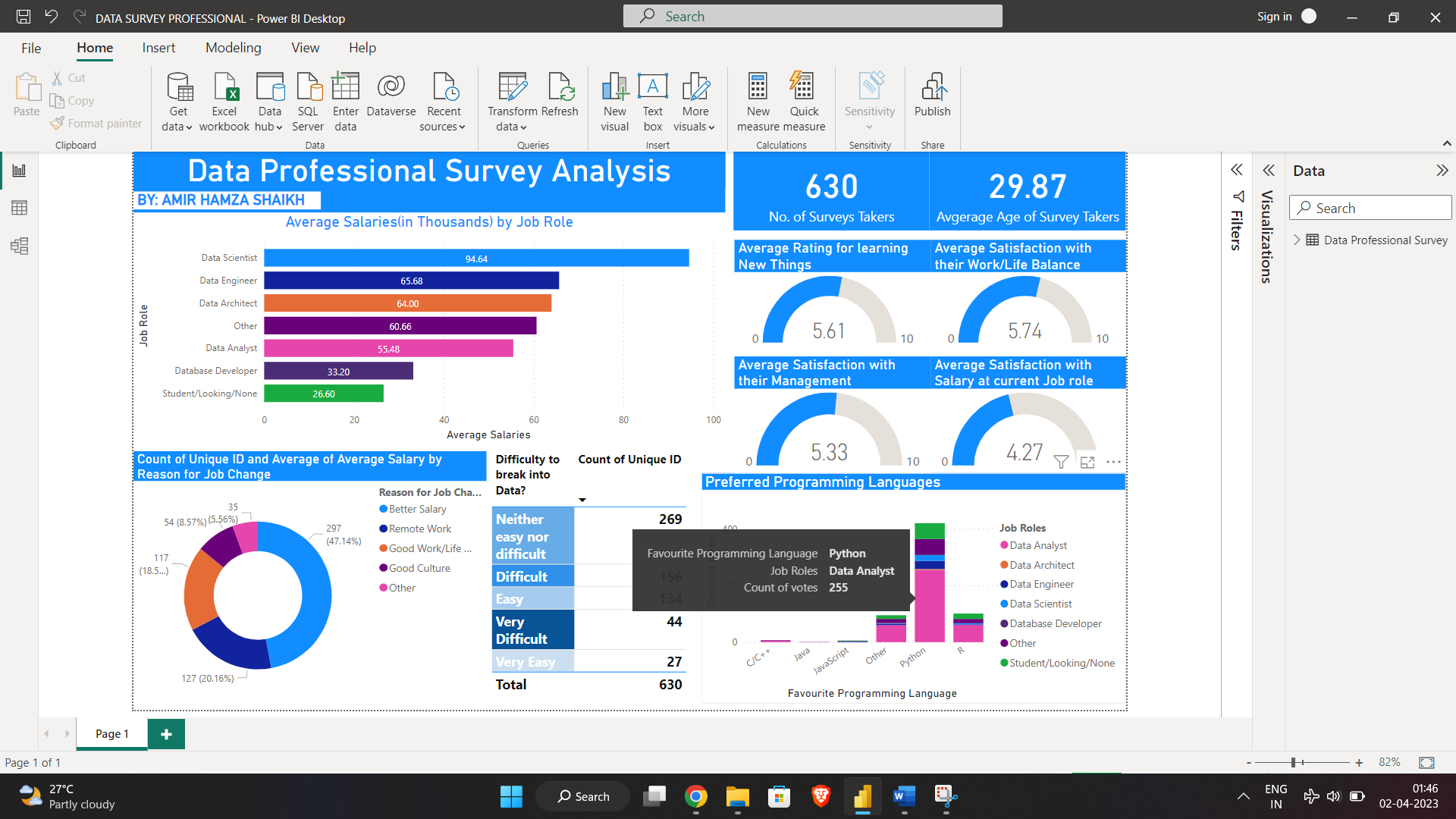
Here we see that we have highest Average Salary for Data Scientist and Database developer as lowest while ignoring the Student/Looking/None. The Data Analyst earn 55.48k on an average.

2)

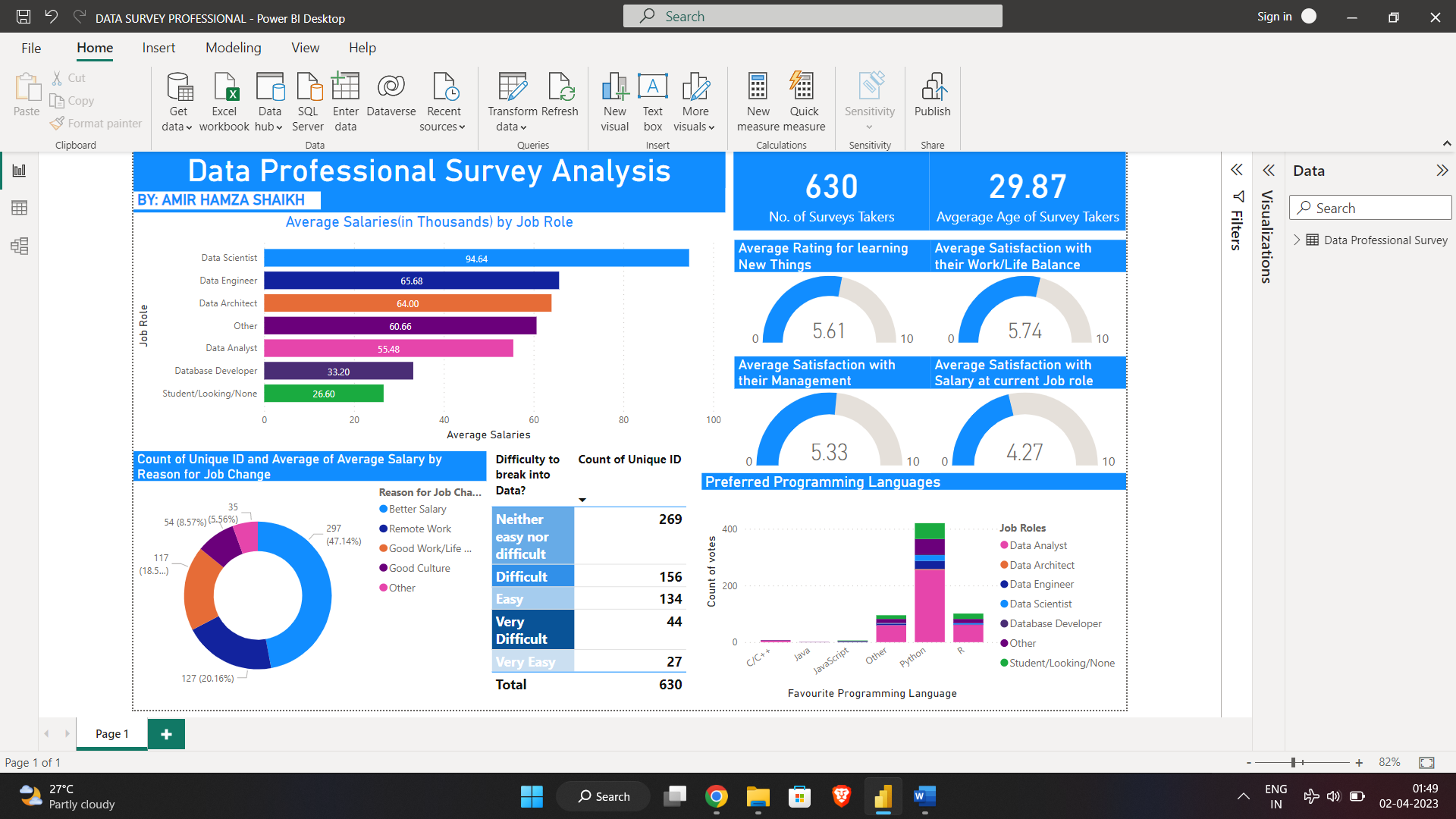


We have four Gauge Charts which says the following for example The people learning new things have average rating of 5.6/10 irrespective of their the Job role. One important thing to know that most of people are not happy with their Job work/satisfaction and with their current role and salary.

The Most preferred language preferred for various Roles is Python. If I just hover it. On python pink section. Python is mainly used by Data Analyst as compared to R. For the students/freshers they highly prefer Python as their Main Language.



3)



We have two observations for this.

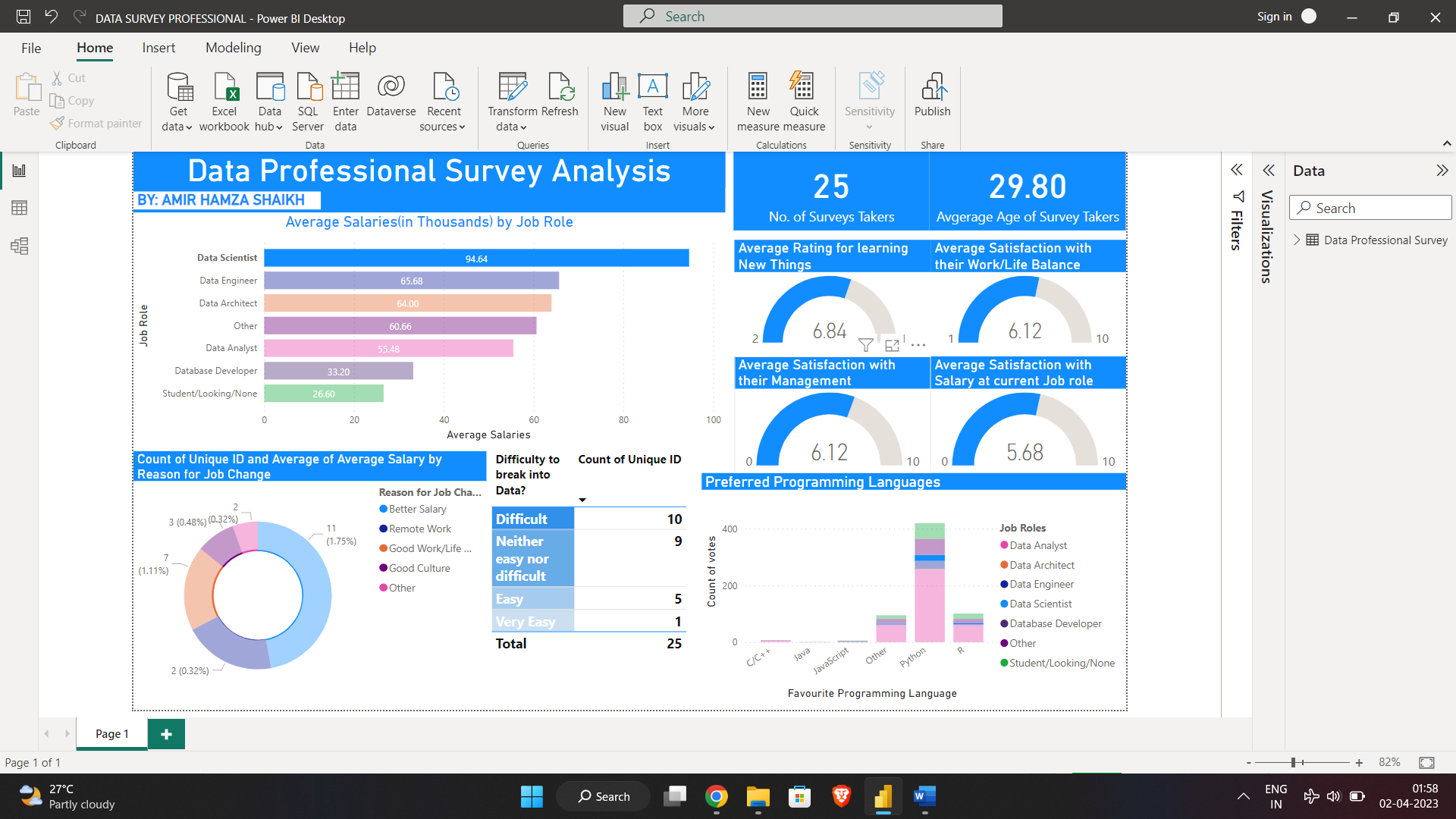
On right side is self explainable. Most of people find that it is average/medium level to break into Data related Jobs. Only 134+27 found that getting Job is Esay!

On Left side we found that people switch to the next Job fot two Main reasons:

* To have better Salary as comapred to current job position and salary
* Most people prefer to work at home instead of working in office. As they have become adaptable working at home during pandemic.

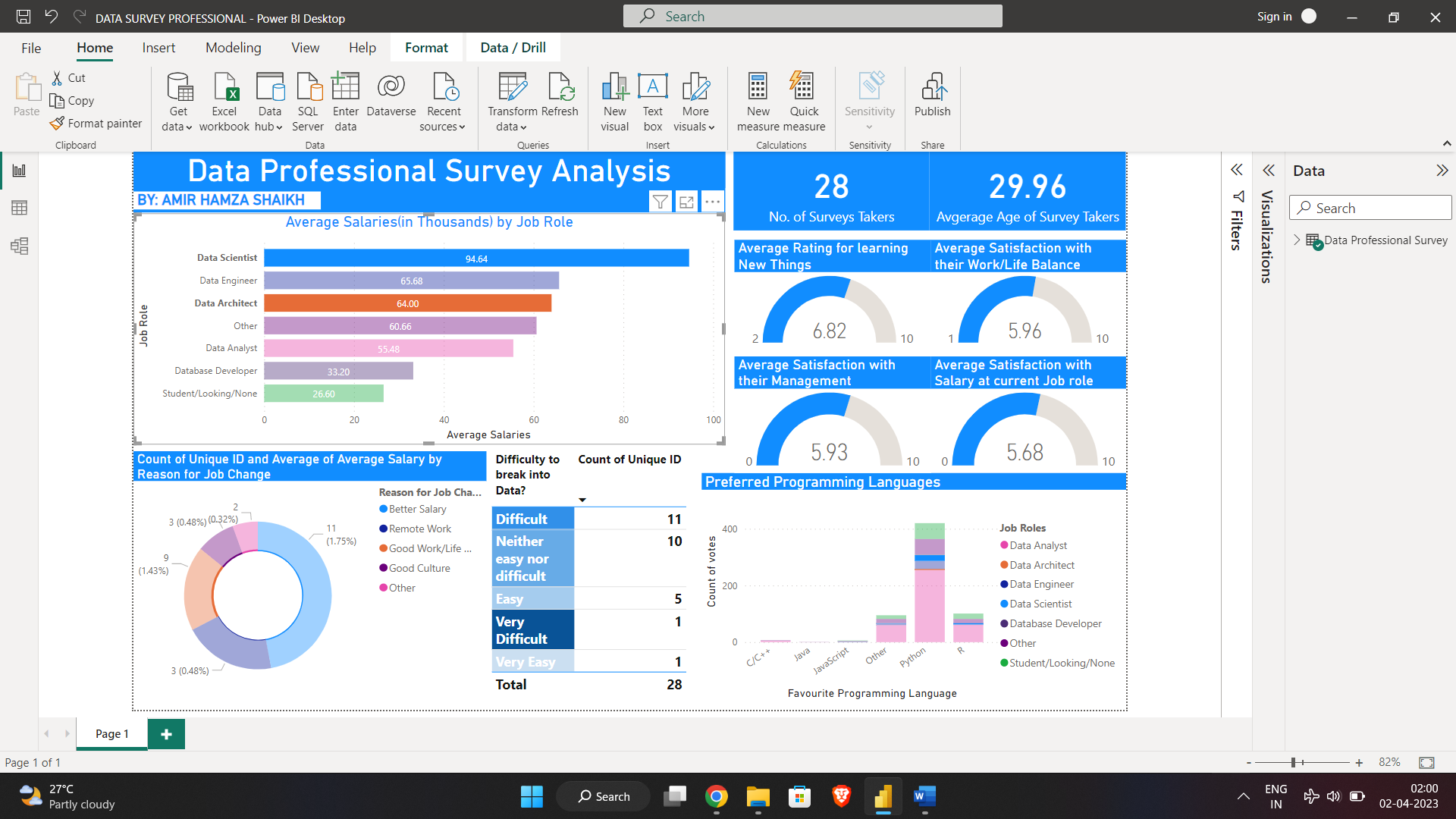
Deeper Observation:

1)



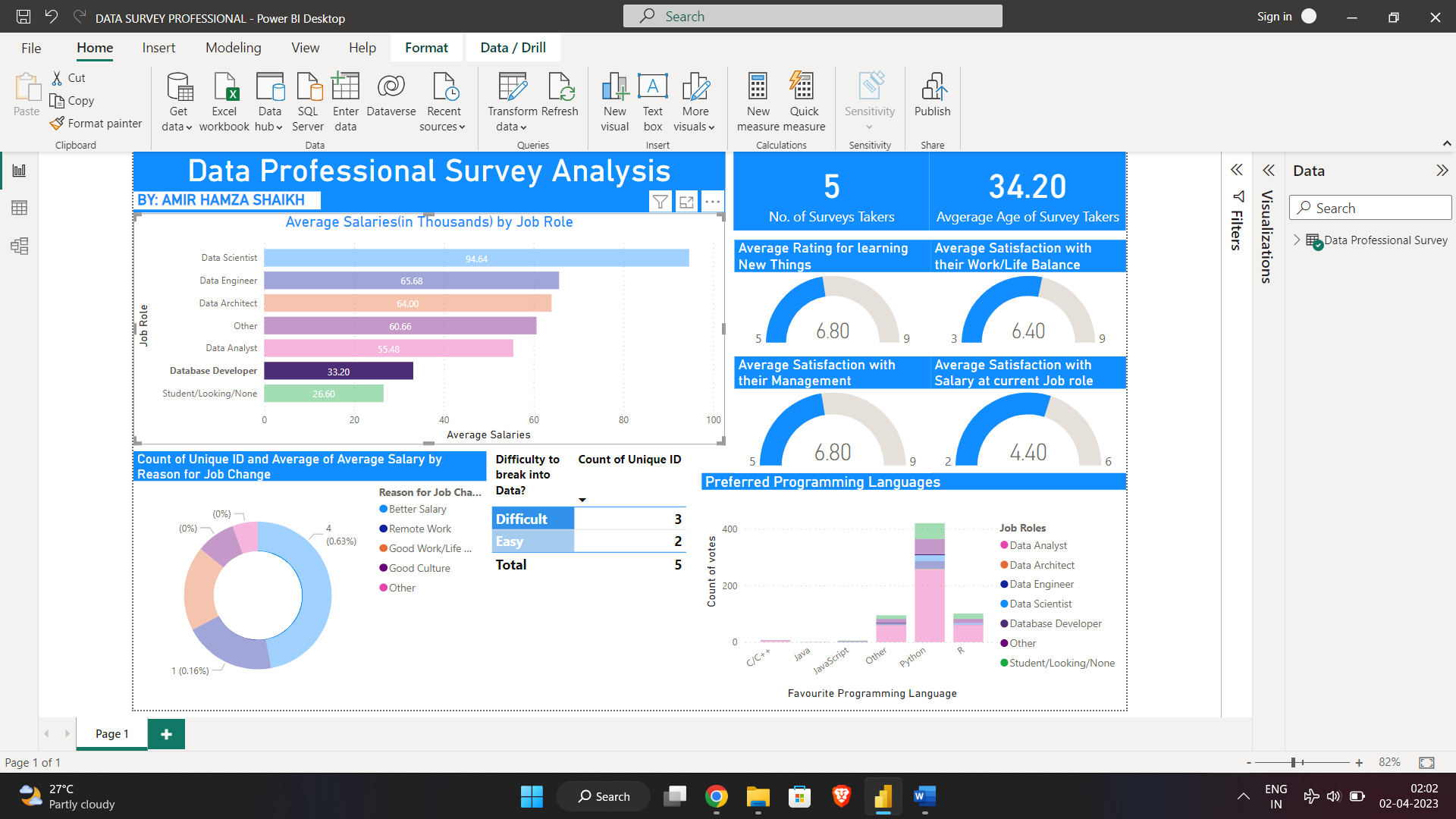
The Data scientists have the highest rating for learning New things.

2)



Both Data Scientists and Data architects have highest Job Satisfaction with current role and salary.

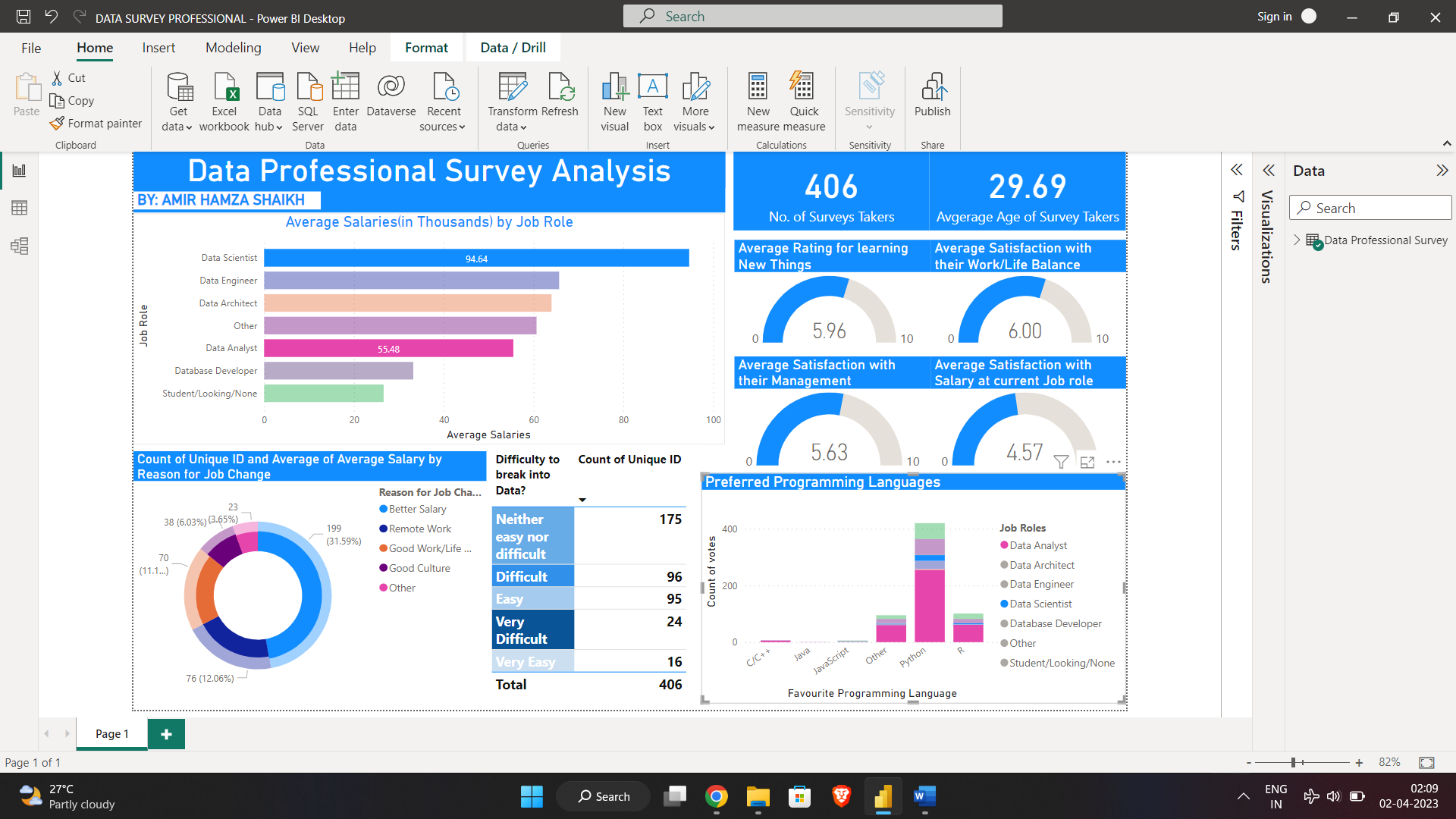
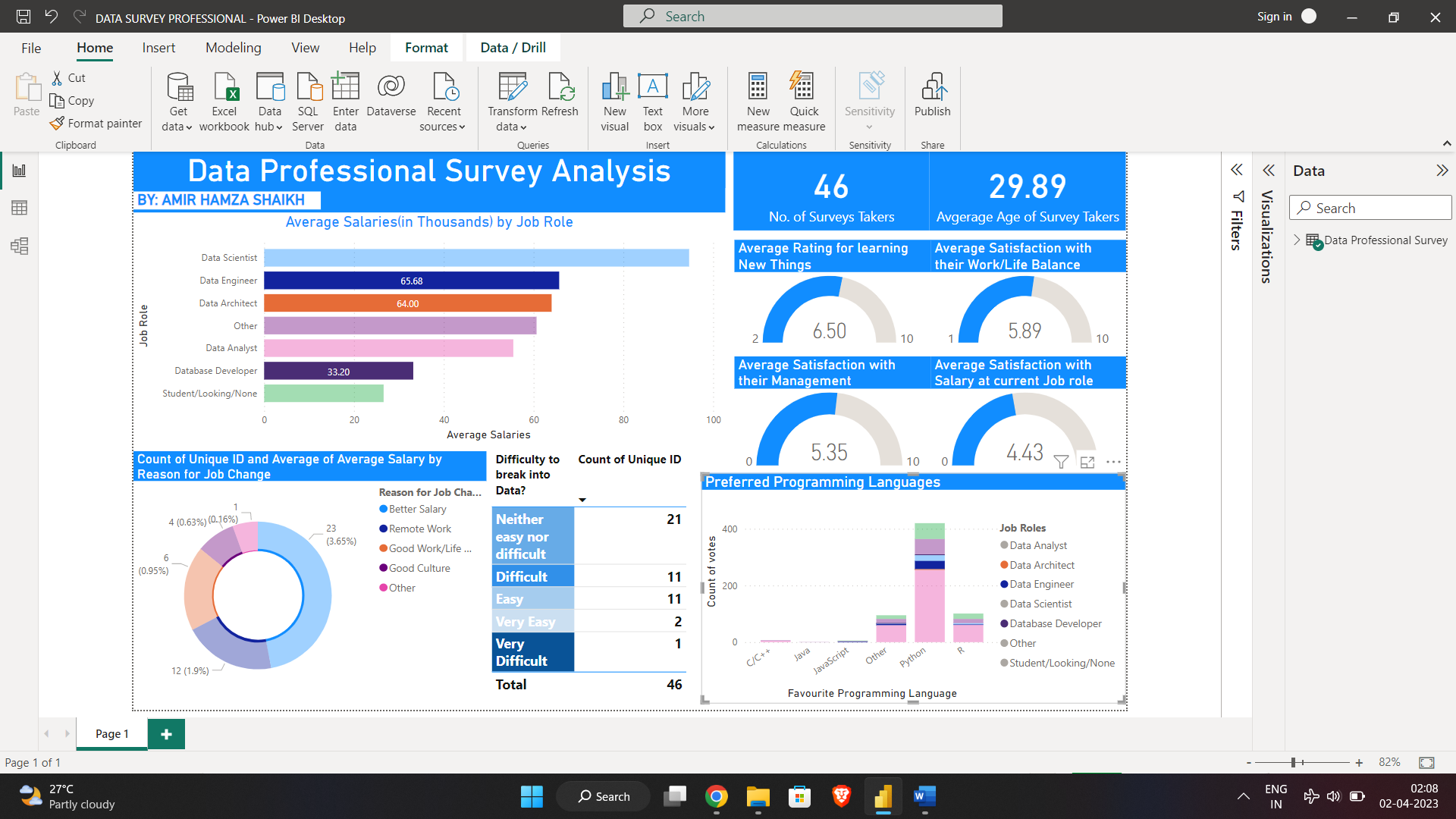
3)



The Database developer have no tension with their Management but they not happy with their current role and income.

4)

4)



The people working with their current role as Data analyst or scientist are open to get better company as they gain experiences. They are highly demanding jobs.