A DATA ANALYSIS PROJECT

BASED ON A SURVEY OF DATA PROFESSIONALS ACROSS COUNTRIES

PREPARED BY: CHINEDU ELEKWA. P

BRIEF INFORMATION

A survey was given to several data professionals across different countries to create a data with which analyst can help explain the work life among them. The data analysis would answer certain puzzling questions that gives insights to Business Owners and Human Resource Managers. The survey data was collected and stored in an excel spreadsheet.

- DATA CLEANING TOOL: Power Query
- DATA VISUALIZATION TOOL: PowerBi
- NUMBER OF COLUMNS: 27
- NUMBER OF ROWS: 631

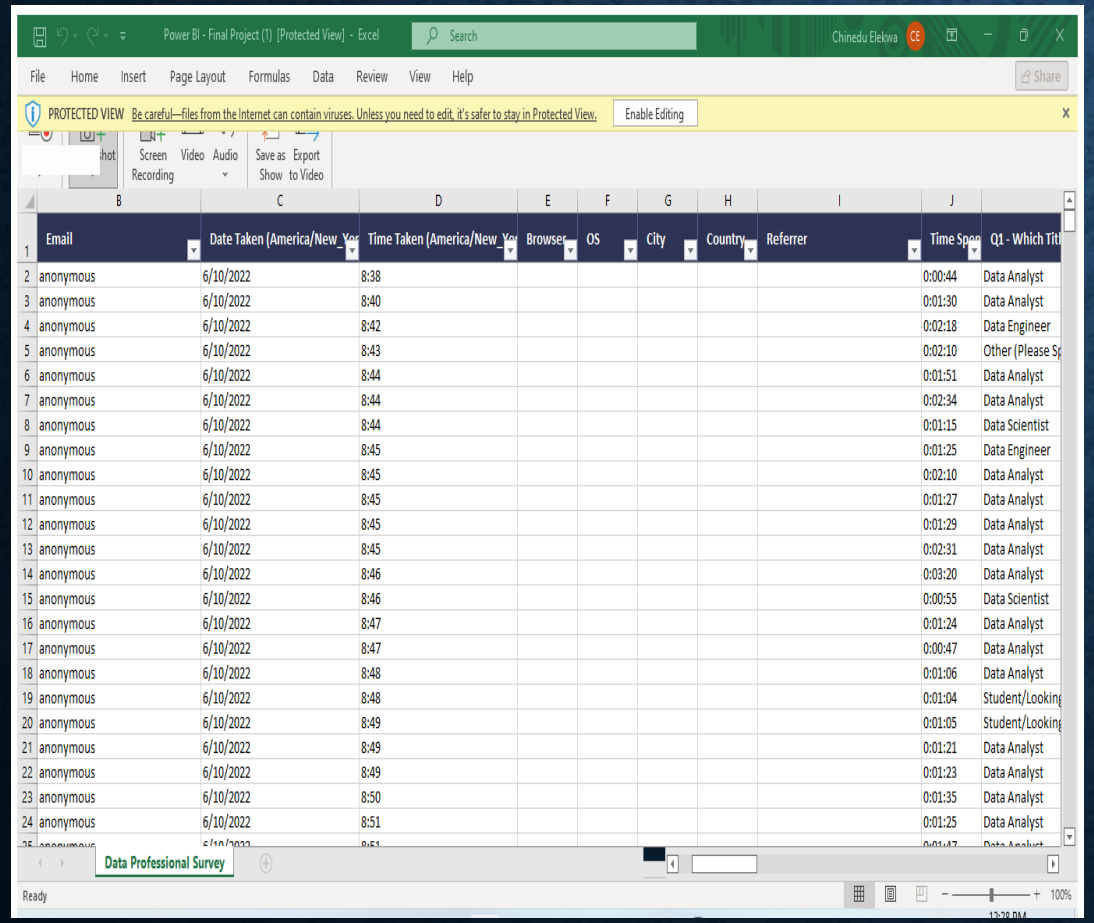
PROBLEM STATEMENT

1. Favorite programming language of data professionals
2. Average Salary of data professionals
3. Average of data professionals happy learning through work
4. Average of data professionals happy with salary

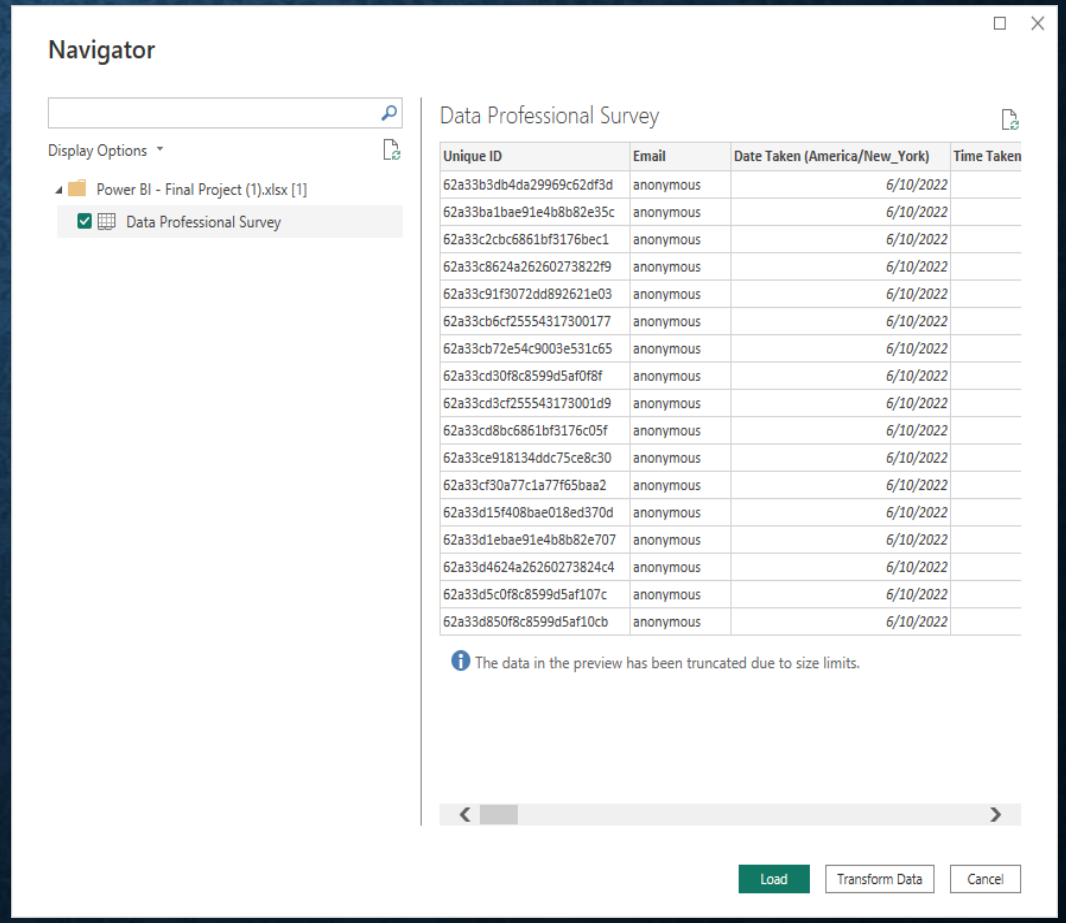
COLUMN HEADINGS OF THE DATA

- Unique ID, Email Date Taken (America/New_York), Time Taken (America/New_York), Browser, OS, City, Country, Referrer, Time Spent, Q1 - Which Title Best Fits your Current Role?, Q2 - Did you switch careers into Data?, Q3 - Current Yearly Salary (in USD), Q4 - What Industry do you work in?, Q5 - Favorite Programming Language, Q6 - How Happy are you in your Current Position with the following? (Salary), Q6 - How Happy are you in your Current Position with the following? (Work/Life Balance), Q6 - How Happy are you in your Current Position with the following? (Coworkers), Q6 - How Happy are you in your Current Position with the following? (Management), Q6 - How Happy are you in your Current Position with the following? (Upward Mobility), Q6 - How Happy are you in your Current Position with the following? (Learning New Things), Q7 - How difficult was it for you to break into Data?, Q8 - If you were to look for a new job today, what would be the most important thing to you?, Q9 - Male/Female?, Q10 - Current Age, Q11 - Which Country do you live in?, Q12 - Highest Level of Education, Q13 - Ethnicity

I imported the excel file into PowerBi and clicked the transform button so as to clean the data using Power Query

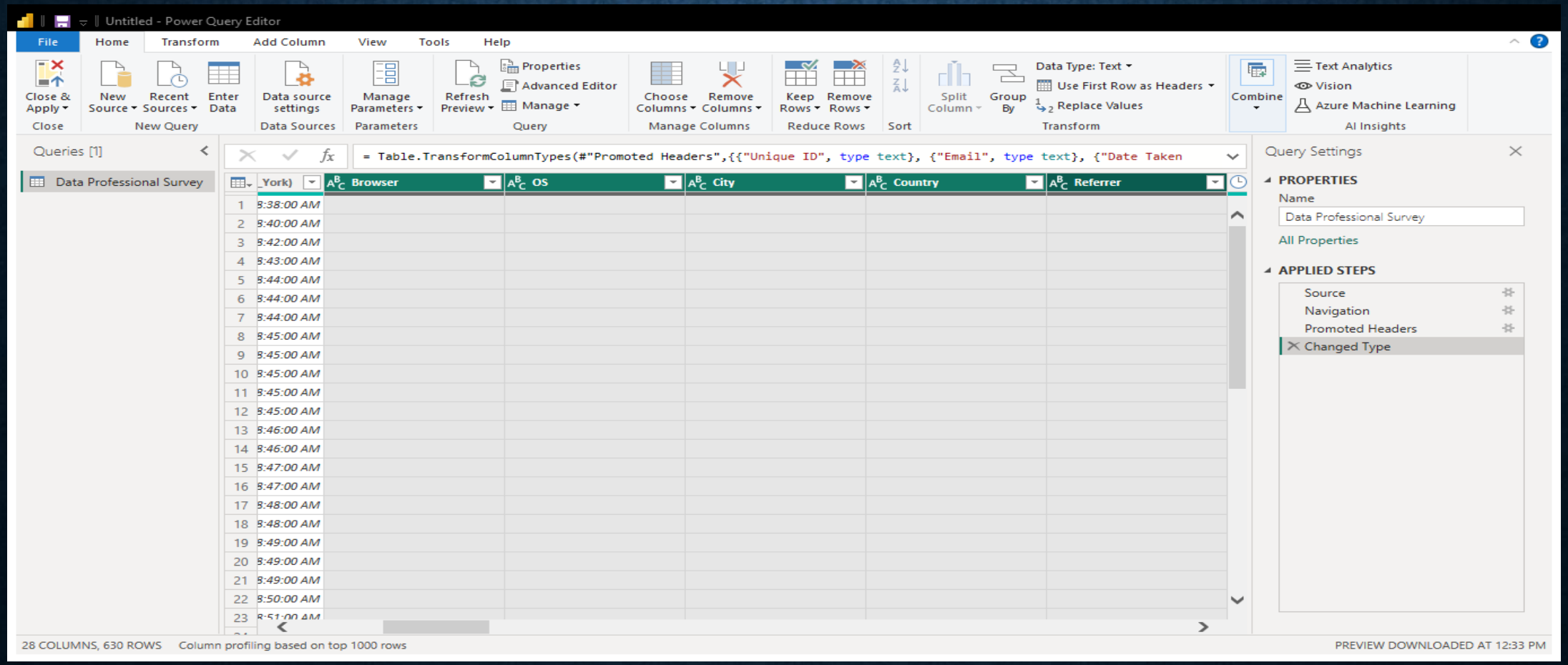


	Email	Date Taken (America/New_Yor	Time Taken (America/New_Yo	Browser	OS	City	Country	Referrer	Time Spen	Q1 - Which Titl
2	anonymous	6/10/2022	8:38						0:00:44	Data Analyst
3	anonymous	6/10/2022	8:40						0:01:30	Data Analyst
4	anonymous	6/10/2022	8:42						0:02:18	Data Engineer
5	anonymous	6/10/2022	8:43						0:02:10	Other (Please Sp
6	anonymous	6/10/2022	8:44						0:01:51	Data Analyst
7	anonymous	6/10/2022	8:44						0:02:34	Data Analyst
8	anonymous	6/10/2022	8:44						0:01:15	Data Scientist
9	anonymous	6/10/2022	8:45						0:01:25	Data Engineer
10	anonymous	6/10/2022	8:45						0:02:10	Data Analyst
11	anonymous	6/10/2022	8:45						0:01:27	Data Analyst
12	anonymous	6/10/2022	8:45						0:01:29	Data Analyst
13	anonymous	6/10/2022	8:45						0:02:31	Data Analyst
14	anonymous	6/10/2022	8:46						0:03:20	Data Analyst
15	anonymous	6/10/2022	8:46						0:00:55	Data Scientist
16	anonymous	6/10/2022	8:47						0:01:24	Data Analyst
17	anonymous	6/10/2022	8:47						0:00:47	Data Analyst
18	anonymous	6/10/2022	8:48						0:01:06	Data Analyst
19	anonymous	6/10/2022	8:48						0:01:04	Student/Looking
20	anonymous	6/10/2022	8:49						0:01:05	Student/Looking
21	anonymous	6/10/2022	8:49						0:01:21	Data Analyst
22	anonymous	6/10/2022	8:49						0:01:23	Data Analyst
23	anonymous	6/10/2022	8:50						0:01:35	Data Analyst
24	anonymous	6/10/2022	8:51						0:01:25	Data Analyst

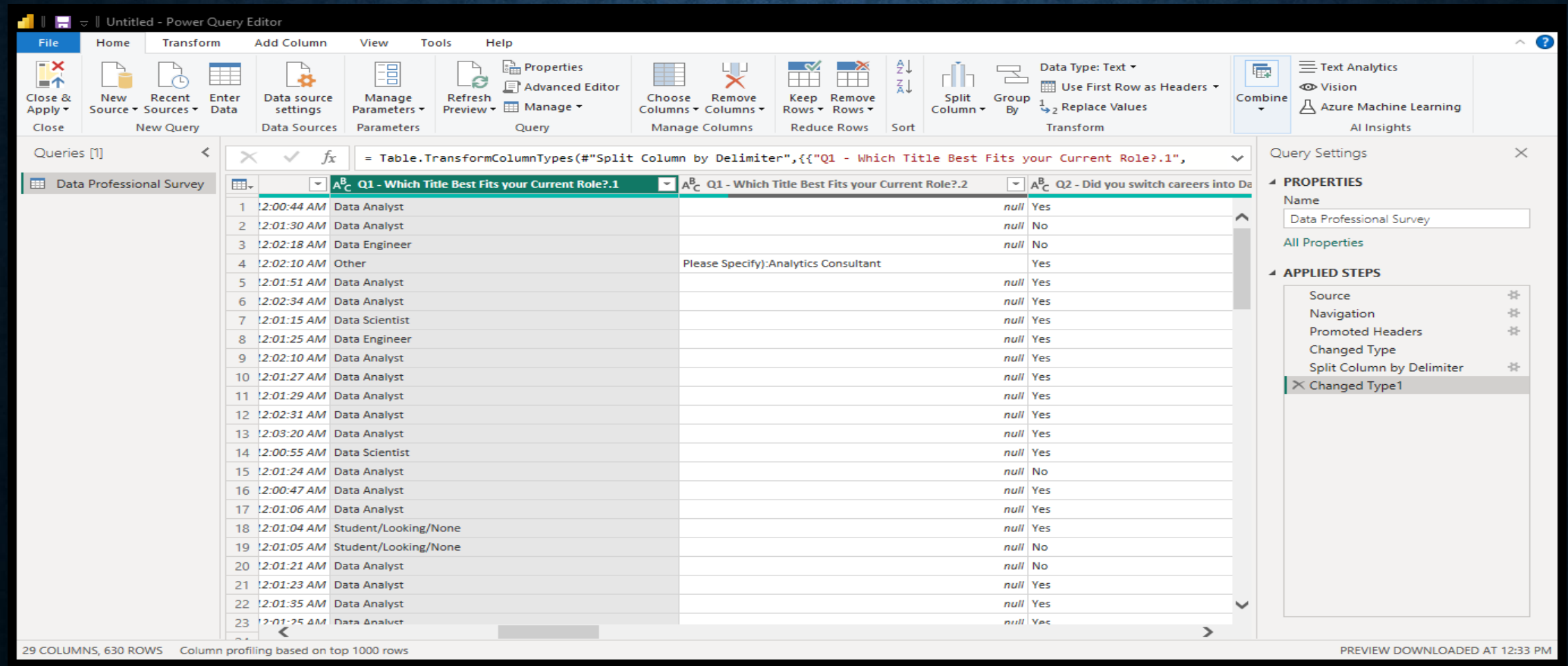


Unique ID	Email	Date Taken (America/New_York)	Time Taken
62a33b3db4da29969c62df3d	anonymous	6/10/2022	
62a33ba1bae91e4b8b82e35c	anonymous	6/10/2022	
62a33c2cbc6861bf3176bec1	anonymous	6/10/2022	
62a33c8624a26260273822f9	anonymous	6/10/2022	
62a33c91f3072dd892621e03	anonymous	6/10/2022	
62a33cb6cf25554317300177	anonymous	6/10/2022	
62a33cb72e54c9003e531c65	anonymous	6/10/2022	
62a33cd30f8c8599d5af0f8f	anonymous	6/10/2022	
62a33cd3cf255543173001d9	anonymous	6/10/2022	
62a33cd8bc6861bf3176c05f	anonymous	6/10/2022	
62a33ce918134ddc75ce8c30	anonymous	6/10/2022	
62a33cf30a77c1a77f65baa2	anonymous	6/10/2022	
62a33d15f408bae018ed370d	anonymous	6/10/2022	
62a33d1ebae91e4b8b82e707	anonymous	6/10/2022	
62a33d4624a26260273824c4	anonymous	6/10/2022	
62a33d5c0f8c8599d5af107c	anonymous	6/10/2022	
62a33d850f8c8599d5af10cb	anonymous	6/10/2022	

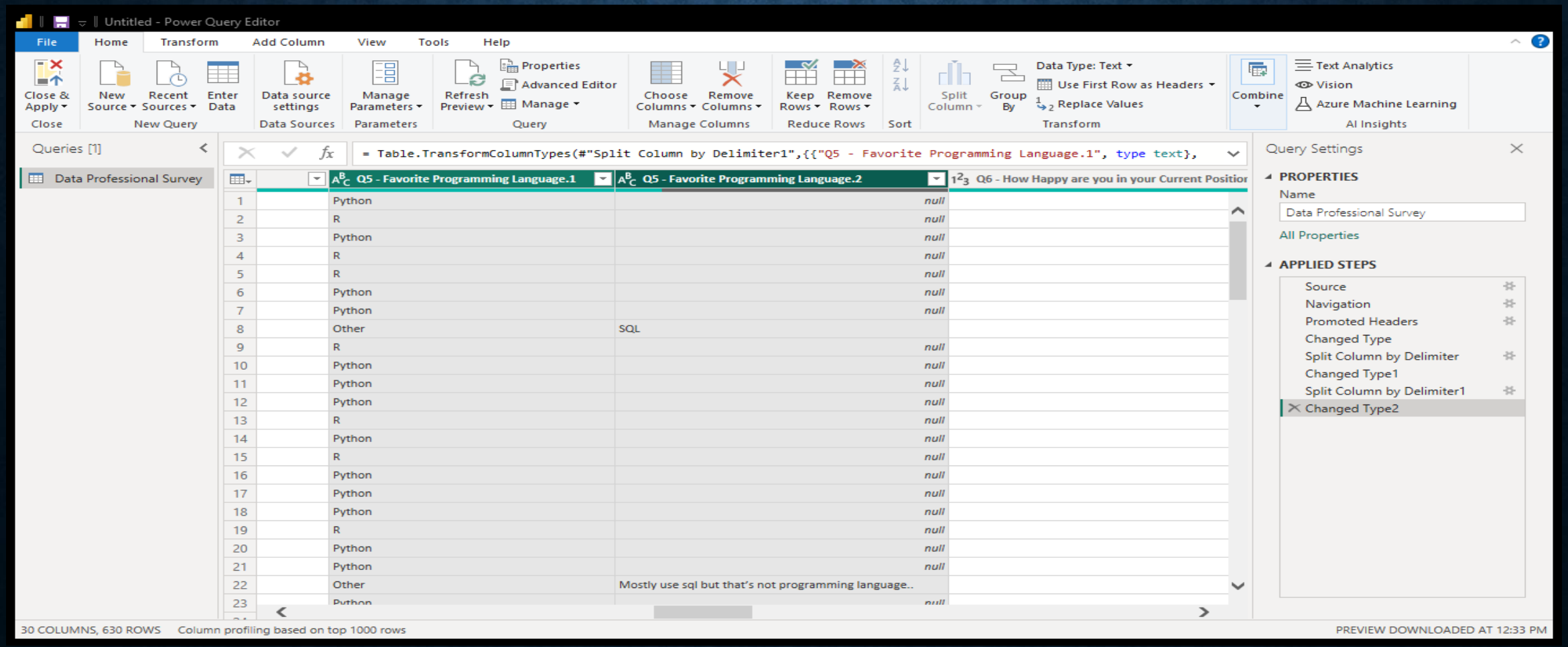
I deleted irrelevant columns (Browser – Referrer) by clicking on the “browser” header, held “shift”, then “referrer” and finally removed the columns.



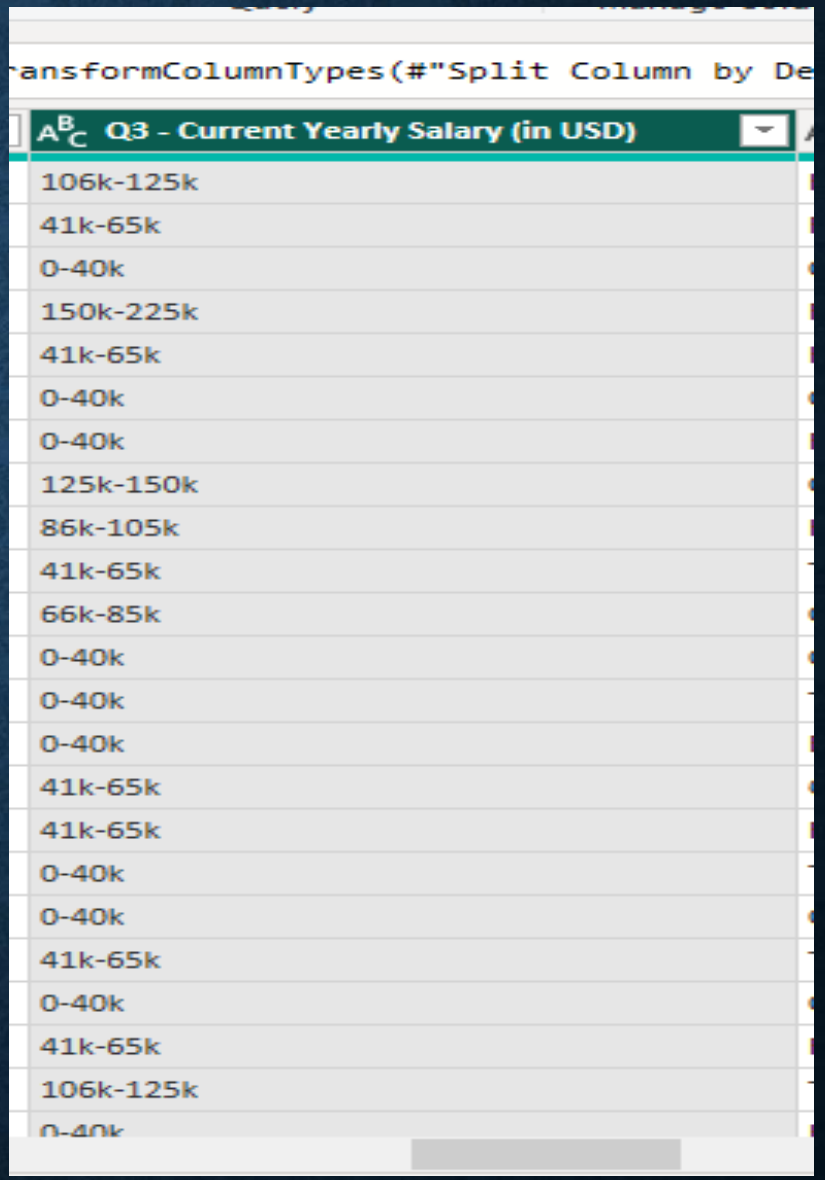
Q1 includes texts in brackets and after it that are not relevant, to deal with this, I split Q1 column (using Split column). "Split column By Delimeter" using the "c" delimeter, "Split at left-most delimeter". This creates a new column – Q1. 2 of text within () and proceeding it.. I deleted the new column. The same for column Q4.



Q5 includes texts after colon which are not relevant. To deal with this, I repeated the same process for Q5. I splitted the columns by delimeter, using the "colon" as delimeter, "Split at left-most delimeter". This creates a new column Q5. 2. I then deleted the new column. The same for column Q13



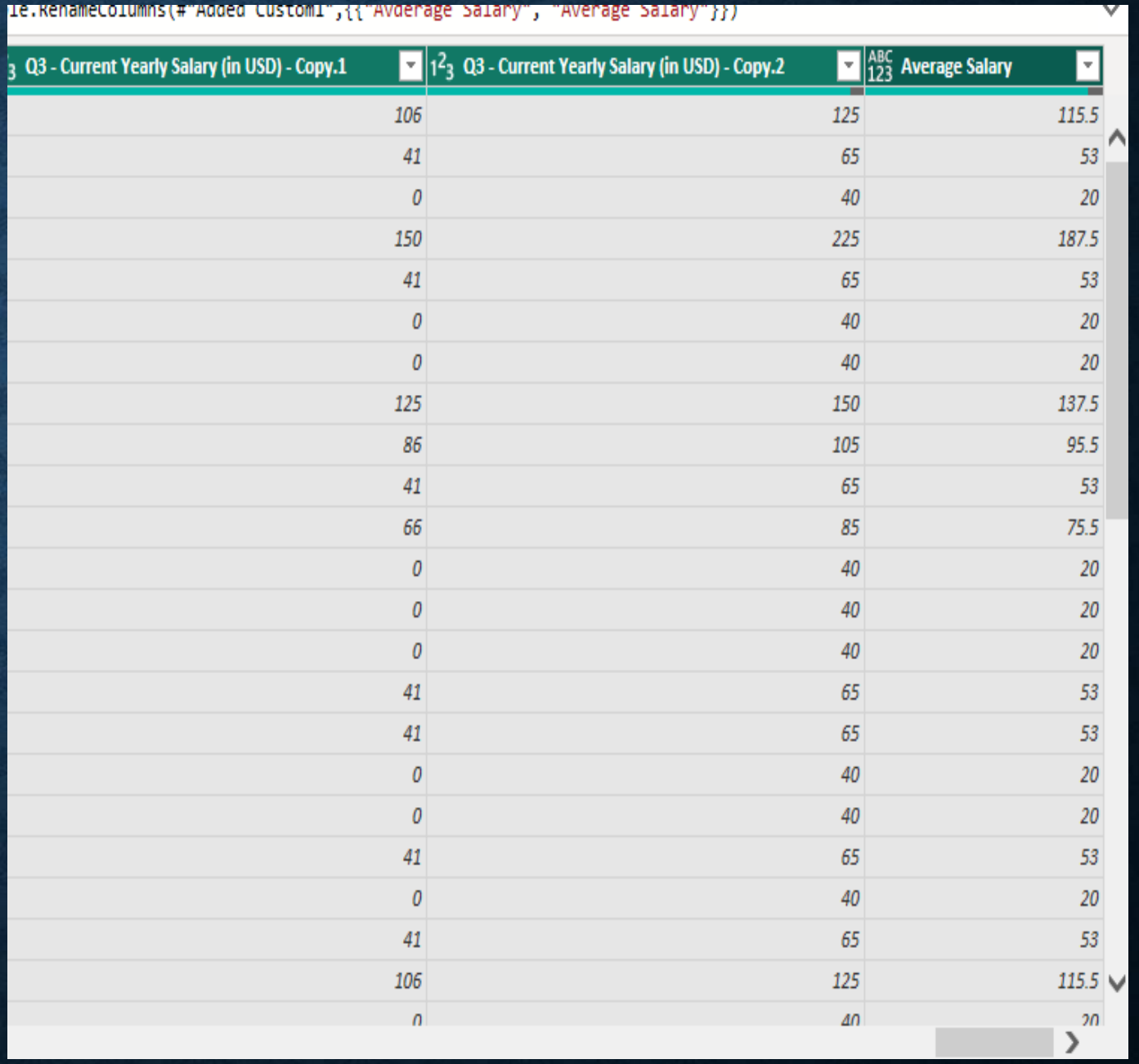
Q3 is in ranges which is not good enough. It should have definite values. First, I duplicated the column, splitted the duplicate (By digits to non-digits). I deleted the column with “k” values (copy 3) Then for copy 2, using “Find and Replace”, I replaced “+” with 225, “-” with nothing and “k” with nothing.



I converted the two duplicate columns data type to “whole numbers” . Furthermore, to create the needed salary column, I clicked on “Add Column”, selected “Custom Column” and named it “Average Salary”. In the Custom Column formula, I inserted “Copy 1” + “Copy 2”/2

([#”Q3-current yearly salary (in USD)-copy.1”]+[#”Q3-current yearly salary (in USD)-copy.2)]/2

Then I deleted copy1 and copy2, left with column Average Salary.

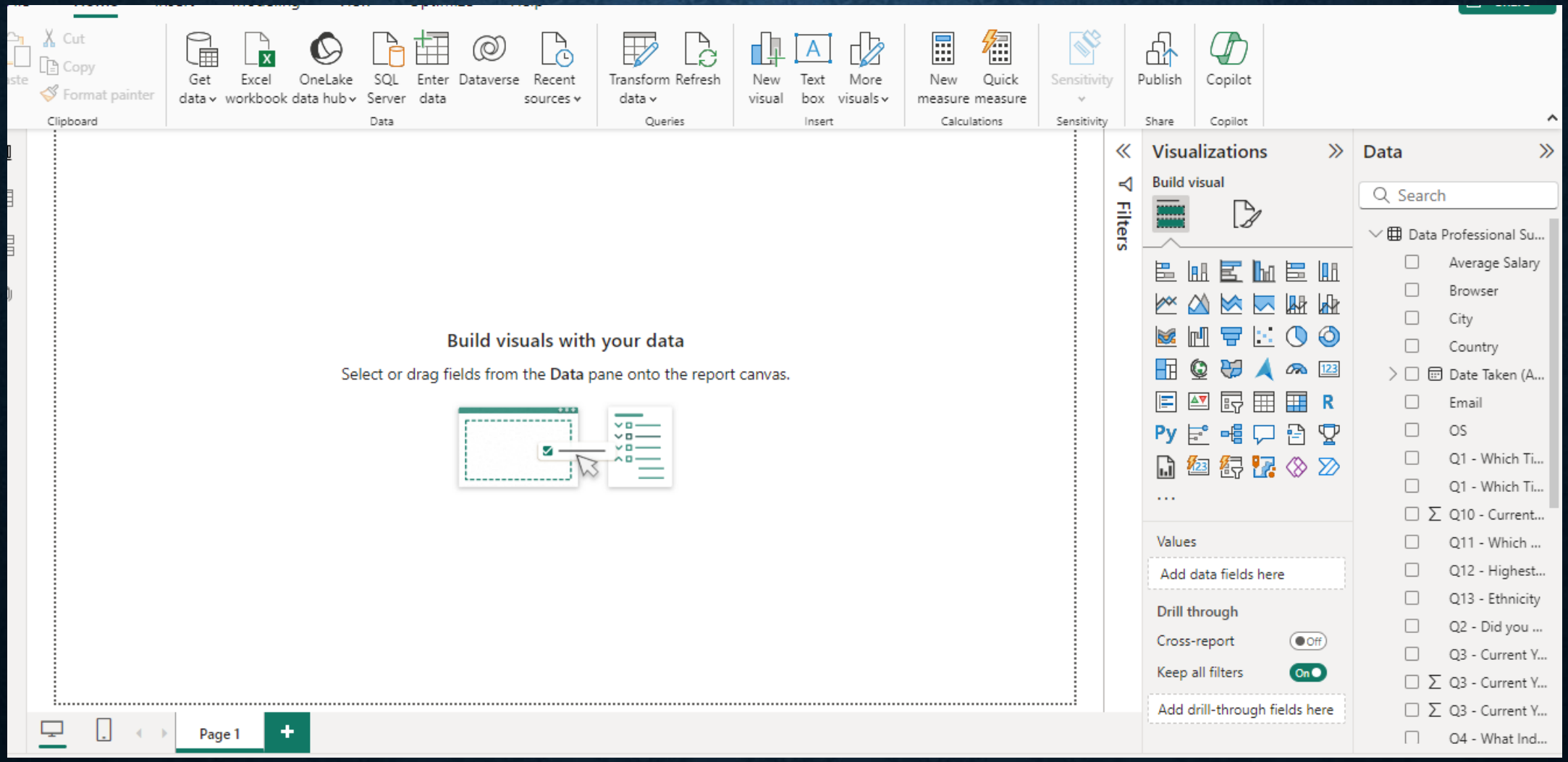


```
le.RenameColumns(#"Added Custom1",{{"Average Salary", "Average Salary"}})
```

Q3 - Current Yearly Salary (in USD) - Copy.1	Q3 - Current Yearly Salary (in USD) - Copy.2	Average Salary
106	125	115.5
41	65	53
0	40	20
150	225	187.5
41	65	53
0	40	20
0	40	20
125	150	137.5
86	105	95.5
41	65	53
66	85	75.5
0	40	20
0	40	20
0	40	20
41	65	53
41	65	53
0	40	20
0	40	20
41	65	53
0	40	20
41	65	53
106	125	115.5
0	40	20

With the cleaning done so far, the data is finally ready for the analysis.

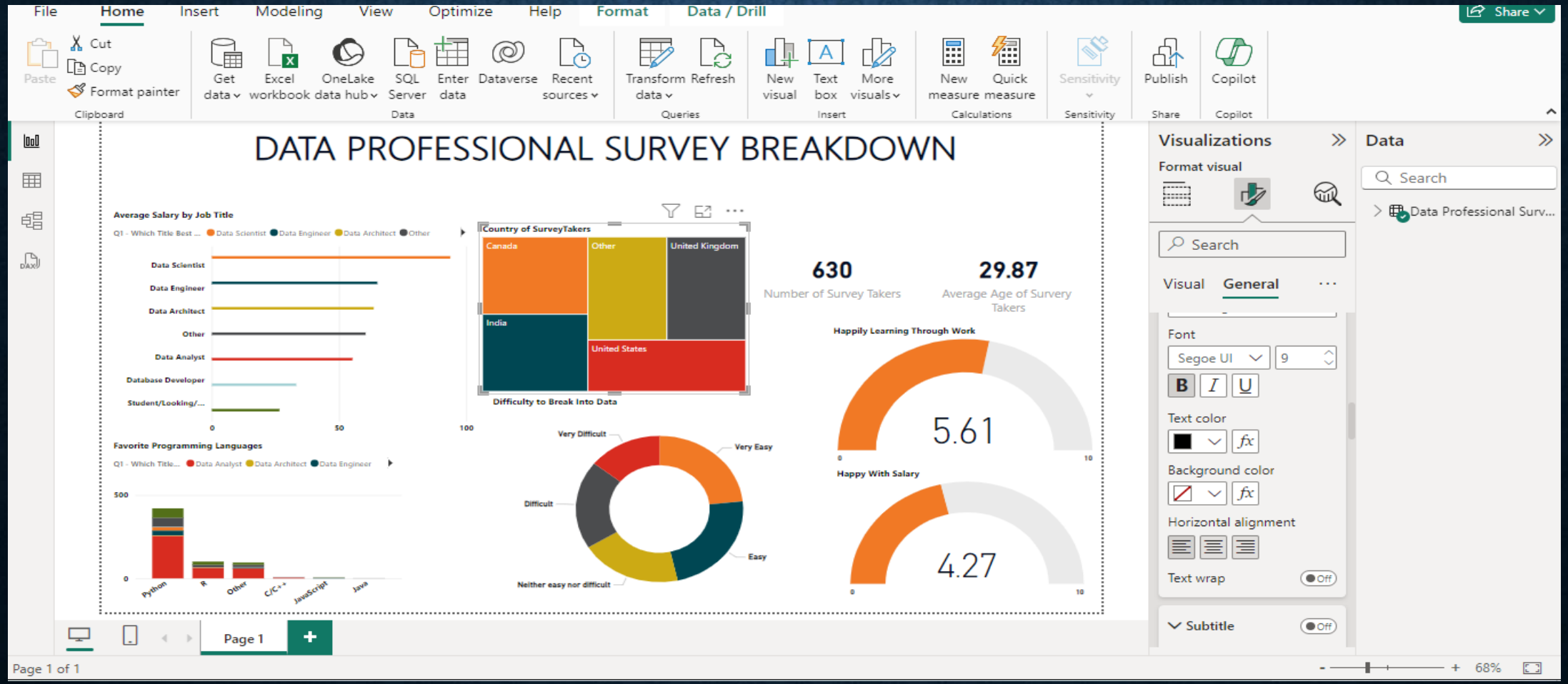
I then clicked "CLOSE AND APPLY to load the ready cleaned data into the visualization pane chart creations



RESULTS

- The total number of survey takers is 630
- Country of Survey Takers are: Canada, India, United Kingdom, United States and Others
- Average Age of survey Takers is 30 years
- Favorite programming language is Python
- Difficulty to break into Data: Easy
- There are more data professionals happy learning through work
- There are less data professionals happy with their salary
- Average Salary by job roles (in USD): DATA SCIENTIST – $93.78K, DATA ENGINEER – $65.09K, DATA ARCHITECT – $63.69K, OTHER – $60.49K, DATA ANALYST – $55.30K, DATABASE DEVELOPER – $33.20K

SCREENSHOT OF THE DASHBOARD



LINKS

- PORTFOLIO LINK: https://www.datascienceportfol.io/ElekwaChinedu
- GITHUB LINK: https://github.com/Nedupelekwa
- LINKEDIN LINK: https://www.linkedin.com/in/chinedu-elekwa-7656a91b0/

- All documents for this project can be downloaded and viewed from my github account specifically at: https://github.com/Nedupelekwa/DATA-PROFESSIONALS-SURVEY.git