



Data Professional Survey Project

The Project

This project we are going to analyze the data that come from survey done by Alex Freberg (Alex the Analyst) and published on his GitHub(<https://github.com/AlexTheAnalyst>) and YouTube channel(<https://www.youtube.com/@AlexTheAnalyst>). We will download the data from GitHub, import it into Power BI, transform the data using Power Query and then we will create the visualizations and finalize the dashboard.

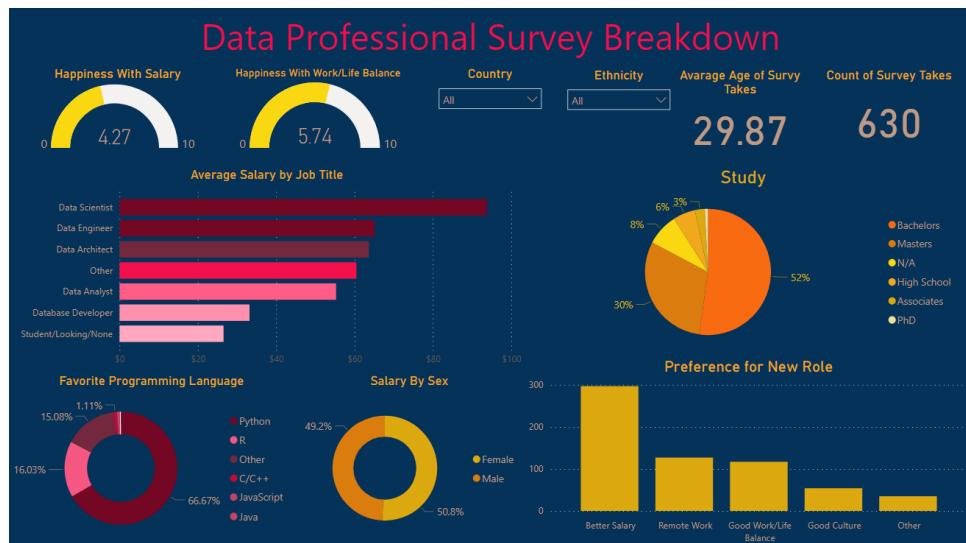


Image 1

About the Data

The data that we are going to be looking at collected from data professionals about August of 2022. Alex posted the survey in LinkedIn, Twitter and others social media, it had 630 who responded the questionnaire.

The File

We are not going to make any changes at all in Excel, just look at file and then move it over to Power BI and start transforming the data. The file has in column A this Unique ID, these all the people that took. The column B have e-mail and column C and D we have the data and hour token.

1	Unique ID	Email	Date Taken [America/New_York]	Time Taken [America/New_York]	Browser	OS	City	Country	Referrer	Time Spent [Q1-Q10]
3	62a33ba1bae91e4b8682e30c	anonymous	6/10/2022	8:40						0:01:30 Det
4	62a33cbdb8f1bf76be1	anonymous	6/10/2022	8:42						0:02:18 Det
5	62a33cd624af2620732219	anonymous	6/10/2022	8:43						0:02:10 Oth
6	62a33d916072d693021e03	anonymous	6/10/2022	8:44						0:01:51 Det
7	62a33dcd5c3394917300177	anonymous	6/10/2022	8:44						0:02:34 Det
8	62a33dd72e54d4000e51	anonymous	6/10/2022	8:44						0:01:15 Det
9	62a33d00b659945a0f8	anonymous	6/10/2022	8:45						0:01:25 Det
10	62a33cc8c22354917300149	anonymous	6/10/2022	8:45						0:02:10 Det
11	62a33cd8b8f1bf76c78c5	anonymous	6/10/2022	8:45						0:01:27 Det
12	62a33c918134dd75ce820	anonymous	6/10/2022	8:45						0:01:29 Det
13	62a33d9077c177f6ba2	anonymous	6/10/2022	8:45						0:02:31 Det
14	62a33d15409be079ed3704	anonymous	6/10/2022	8:46						0:03:20 Det
15	62a33d1eb91e4b8682e107	anonymous	6/10/2022	8:46						0:00:55 Det
16	62a33d1624af2620732244	anonymous	6/10/2022	8:47						0:01:24 Det
17	62a33d5d5c3394915a0f7c	anonymous	6/10/2022	8:47						0:00:47 Det
18	62a33d50fb289945a10a	anonymous	6/10/2022	8:48						0:01:06 Det
19	62a33d6b6f7616468524b	anonymous	6/10/2022	8:48						0:01:04 Stu
20	62a33d6bb0d009f419cd5	anonymous	6/10/2022	8:49						0:01:05 Stu
21	62a33d124af2620732262	anonymous	6/10/2022	8:49						0:01:21 Det
22	62a33d6bb0d009f419ce2b	anonymous	6/10/2022	8:49						0:01:23 Det
23	62a33d6bb0d009f419ced4	anonymous	6/10/2022	8:50						0:01:35 Det
24	62a33e365f72d68d8221d8	anonymous	6/10/2022	8:51						0:01:25 Det
25	62a33e44b2009f419cf9	anonymous	6/10/2022	8:51						0:01:47 Det
26	62a33e257f616e8e8aaef	anonymous	6/10/2022	8:51						0:02:14 Det
27	62a33e30c5491e09499	anonymous	6/10/2022	8:51						0:02:21 Det

Image 2

We have all the questions between column K and AB.

K	L	M	N	O	Q5 - Favorite Programming Languages	Q6 - How Happy are you
1	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 Languages	Q6 - How Happy are you
3	Data Analyst	No	41k-65k	Finance	R	
4	Data Engineer	No	0-40k	Other (Please Specify)/Clean Energy	Python	
5	Other (Please Specify)/Analytics Consultant	Yes	150k-225k	Finance	R	
6	Data Analyst	Yes	41k-65k	Healthcare	R	
7	Data Analyst	Yes	0-40k	Other (Please Specify)/Coworking space	Python	
8	Data Scientist	Yes	0-40k	Finance	Python	
9	Data Engineer	Yes	125k-150k	Other (Please Specify)/Retail	Other/SQL	
10	Data Analyst	Yes	85k-105k	Healthcare	R	
11	Data Analyst	Yes	41k-65k	Telecommunication	Python	
12	Data Analyst	Yes	65k-95k	Other (Please Specify)/Logistics	Python	
13	Data Analyst	Yes	0-40k	Other (Please Specify)/Gaming	Python	
14	Data Analyst	Yes	0-40k	Tech	R	
15	Data Scientist	Yes	0-40k	Education	Python	
16	Data Analyst	No	41k-65k	Construction	R	
17	Data Analyst	Yes	41k-65k	Finance	Python	
18	Data Analyst	Yes	0-40k	Tech	Python	
19	Student/Looking/None	Yes	0-40k	Other (Please Specify)/Retail	Python	
20	Student/Looking/None	No	41k-65k	Tech	R	
21	Data Analyst	No	0-40k	Other (Please Specify)/Aviation	Python	
22	Data Analyst	Yes	41k-65k	Finance	Python	
23	Data Analyst	Yes	100k-125k	Tech	Other/Mostly use sql but that's not prog	
24	Data Analyst	Yes	0-40k	Healthcare	Python	
25	Data Analyst	No	0-40k	Finance	Python	
26	Data Analyst	No	0-40k	Tech	Python	
27	Data Analyst	No	0-40k	Tech	Python	

Image 3

Importing Data

Now we are going to import the file to Power BI. First **open the Power BI**.

1. Click in import data from Excel or Excel workbook (Image 4).

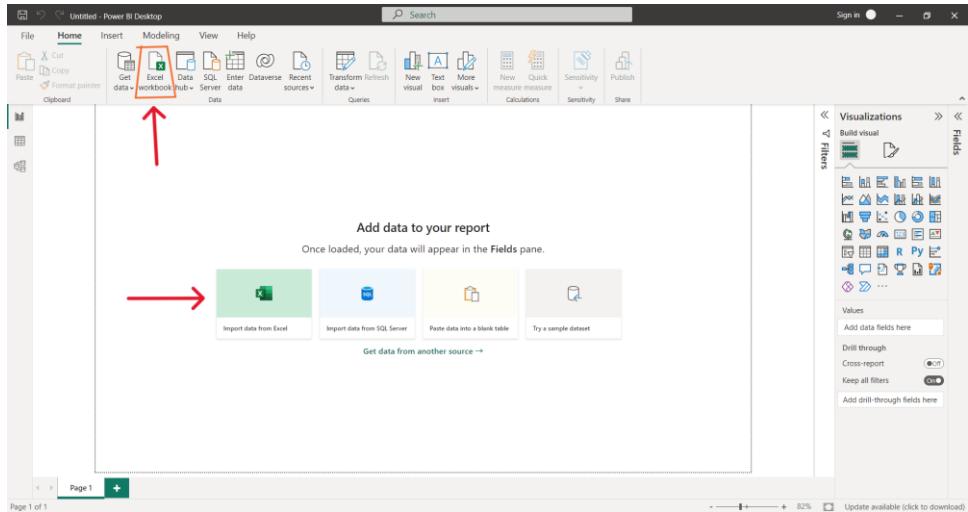


Image 4

2. Select the archive you have downloaded.

3. Click on transform data (Image 5).

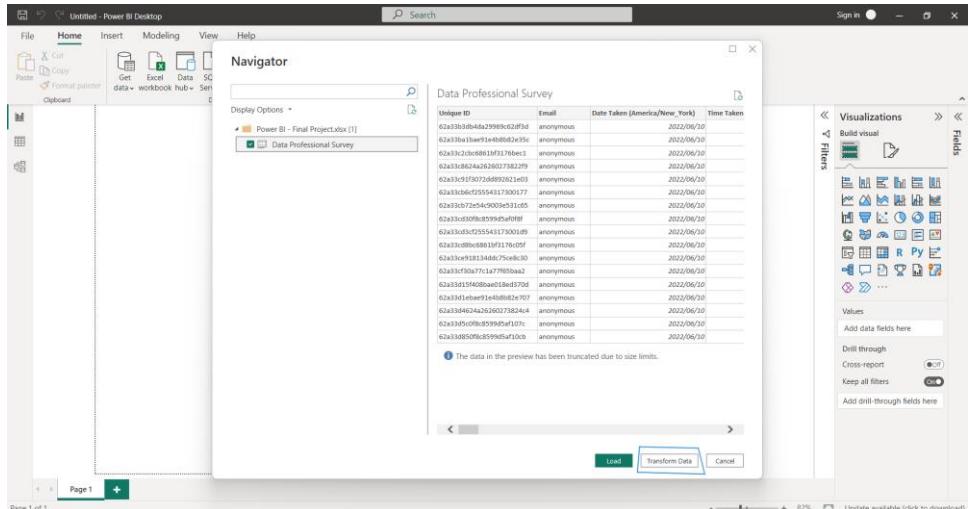


Image 5

Transforming Data

After clicked on **transform data**. The fill will be open in Power Query, right off we can already tell you that there are **columns** that we can just **delete**.

We are **left click on browser column** -> **hit shift and click on the reference column** -> **delete column**.

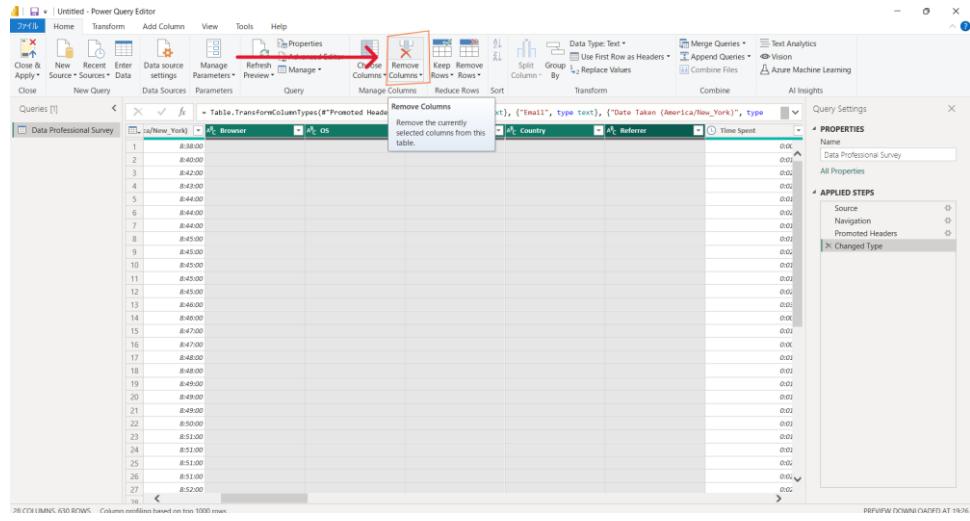


Image 6

We can see in the *image 7* that some columns are filled with the option “*others*”. These is king of the issue; it is do the data not very clean.

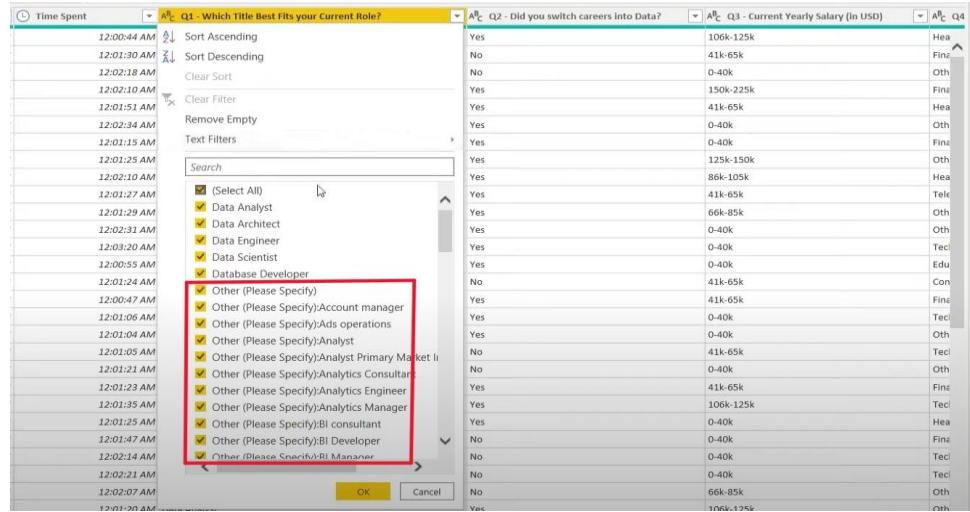


Image 7

For the propose of this study, we are not going to take every single one of these options. If we separated these out would be different options when we created our visualizations, and we don't want that. So, for this **we will split this column(s) to the left by the delimiter “(“**because we only want what comes before the “(“. Starting for the *column Q1*.

The screenshot shows the Microsoft Power BI Data Editor interface. A context menu is open over a column named 'Time Spent'. The 'Split Column' option is selected, which opens a dialog box titled 'Split Column by Delimiter'. In the dialog, the 'Delimited' tab is selected. The 'Delimited' dropdown shows 'By Delimiter' is chosen. Below it, a checkbox labeled 'Split value in the selected column based on the specified delimiter' is checked. The 'APPLIED STEPS' pane on the right lists 'Removed Columns'.

Image 8

The screenshot shows the Microsoft Power BI Data Editor interface. A context menu is open over a column named 'Time Spent'. The 'Split Column' option is selected, which opens a dialog box titled 'Split Column by Delimiter'. In the dialog, the 'Custom' dropdown is selected under 'Select or enter delimiter'. The 'Split at' field contains a dot ('.') and has a red arrow pointing to it. The 'OK' button is visible at the bottom right of the dialog. The 'APPLIED STEPS' pane shows the step 'Removed Columns'.

Image 9

After clicked in button ok, it should create another column. The original column will have “.1” at the end of its name and the new column will have “.2” at the end of its name. The new column (with the final .2) does not interest us, so we will simply delete it.

Table.TransformColumnTypes("#Split Column by Delimiter", {"Q1 - Which Title Best Fits your Current Role?":1}, {"Q1 - Which Title Best Fits your Current Role?":2}, {"Q2 - Did you switch careers into Data?":3})

	Time Spent	A% Q1 - Which Title Best Fits your Current Role?1	A% Q1 - Which Title Best Fits your Current Role?2	A% Q2 - Did you switch careers into Data?3
1	12:00:44 AM	Data Analyst	null Yes	
2	12:01:30 AM	Data Analyst	null No	
3	12:02:18 AM	Data Engineer	null No	
4	12:02:10 AM	Other	Yes	
5	12:01:51 AM	Data Analyst	null Yes	
6	12:02:34 AM	Data Analyst	null Yes	
7	12:01:19 AM	Data Scientist	null Yes	
8	12:01:25 AM	Data Engineer	null Yes	
9	12:02:10 AM	Data Analyst	null Yes	
10	12:01:27 AM	Data Analyst	null Yes	
11	12:01:29 AM	Data Analyst	null Yes	
12	12:02:31 AM	Data Analyst	null Yes	
13	12:03:20 AM	Data Analyst	null Yes	
14	12:00:55 AM	Data Scientist	null Yes	
15	12:01:24 AM	Data Analyst	null No	
16	12:00:47 AM	Data Analyst	null Yes	
17	12:01:06 AM	Data Analyst	null Yes	
18	12:01:04 AM	Student/Looking/None	null Yes	
19	12:01:05 AM	Student/Looking/None	null No	
20	12:01:21 AM	Data Analyst	null No	
21	12:01:23 AM	Data Analyst	null Yes	
22	12:01:35 AM	Data Analyst	null Yes	
23	12:01:25 AM	Data Analyst	null Yes	
24	12:01:47 AM	Data Analyst	null No	
25	12:02:14 AM	Data Analyst	null No	

Image 10

We are going to do exact the same thing to column Q4. In the column Q5 we have the same problem, but in this case, we will **split the column by the delimiter colon**.

- Other:Power bi
- Other:Qlik sense script
- Other:SAS
- Other:SAS SQL
- Other:Sql
- Other:sql
- Other:SQL
- Other:SQL
- Other:Sql & plsql
- Other:SQL because that is all I know really well
- Other:SQL Postgres
- Other:Stata
- Other:unknown
- Other:Vba
- Other:VBA
- Python

Image 11

Because it is the colon that separates what we want from what we don't want. After clicked in “*Split Column by Delimiter*” we select the option **colon -> left-most delimiter -> ok**.

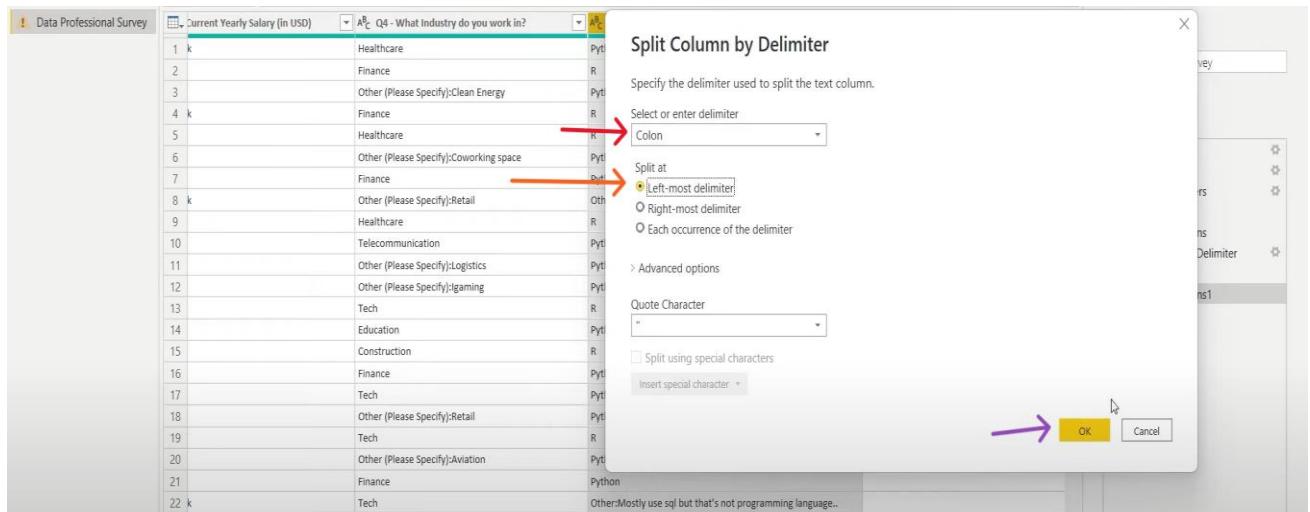


Image 12

Here we will delete the new column again.

The *column Q3* we want to do is break up these **numbers 106 - 125** and then take the **average** of those numbers. We will use some docs in there so for this create in two separate columns. Then we will create a third column that will give us the average of those two numbers.

We will start by copying column Q3. **Right-click on the Q3 column -> Duplicate Column.**

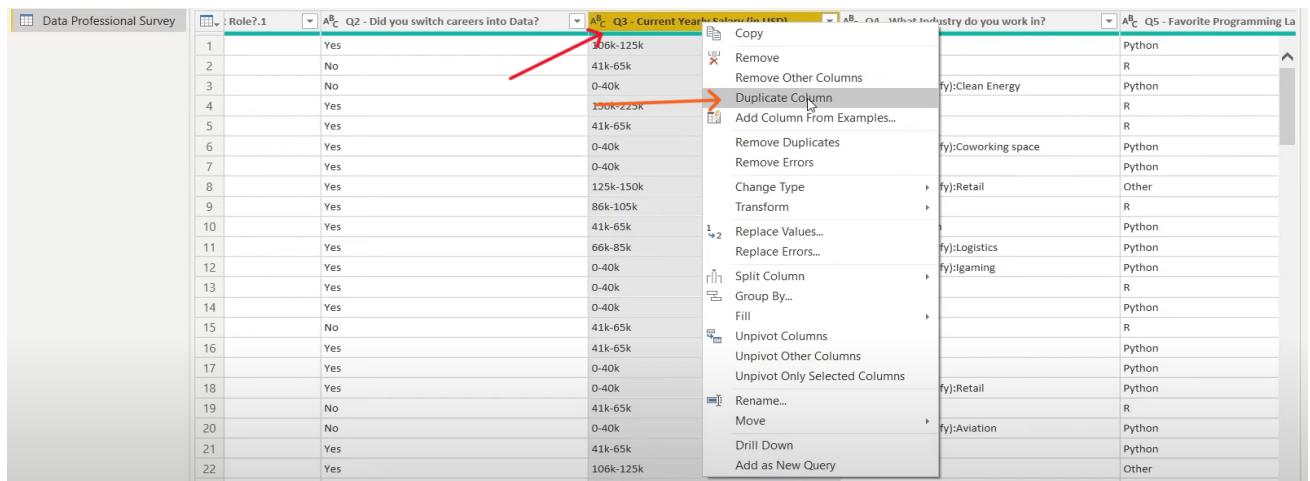


Image 13

Now we have this copy in the end and we can use this one. Next step is split this one up we are to **Left Click on the Column Header -> Split Column -> By Digit to Non-Digit.**

23 COLUMNS: 630 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 19:26

Image 14

The **first column** we have just **numeric values**, the **second** we have **k dash numeric** or just **dash** and this can be easily cleaned.

23 COLUMNS: 630 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 19:26

Image 15

The **third one** we can just get rid of, because it is **only the value k**. So, we will remove. (Image 5)

Now we will **take off** the value **k** and **dash** from the **column Q3 copy2**. Right Click on the **Column Header** -> **Replace Values** -> Value to Find: **k**, Replace With: **not fill**. We want to **delete** the **K** value, for this we are going to **replace the value for nothing**.

A screenshot of the Power BI Advanced Editor interface. The main area shows a table with three columns: "Highest Level of Education", "Q13 - Ethnicity", and "Q3 - Current Yearly Salary (in USD) - Copy.1". The "Q3 - Current Yearly Salary (in USD) - Copy.1" column contains values like 106, 41, 0, 150, etc., with some entries starting with 'k-' followed by a number. A context menu is open over the last few rows of the salary column, with a red arrow pointing to the "Replace Values..." option under the "Transform" section.

Image 16

A screenshot of the Power BI Advanced Editor interface. The main area shows the same table as in Image 16. A "Replace Values" dialog box is open in the foreground. It has two main input fields: "Value To Find" with the value "k" and "Replace With" with an empty input field. A red arrow points to each of these input fields. At the bottom right of the dialog box are "OK" and "Cancel" buttons.

Image 17

Images 15 and 16 show us how we **eliminated the K**, now we are going to do exactly the same to **eliminate the dash**. **Column Header -> Replace Values -> Value to Find: “-” (dash), Replace With: “”**.

We have a plus too, let's get rid of that because that is when some people had 250 or plus. **click on the header arrow -> select + -> ok**.

The screenshot shows the Power BI Data Editor with a table titled "Table.ReplaceValue(#"Replaced Value","-","",Replacer.ReplaceText,{"Q3 - Current Yearly Salary (in USD) - Copy.2"})". The table has three columns: "Highest Level of Education", "Q13 - Ethnicity", and "Q3 - Current Yearly Salary (in USD) - Copy.2". The "Q3 - Current Yearly Salary (in USD) - Copy.2" column contains values like 41, 86, 66, etc. A context menu is open on the header of this column, showing options like Sort Ascending, Sort Descending, Clear Sort, Clear Filter, Remove Empty, and Text Filters. A search bar is present. A modal dialog box titled "Replace Values" is displayed, showing a list of values: 105, 125, 150, 225, 40, 65, 85. The "Select All" checkbox is checked. A red arrow points to this checkbox, and a purple arrow points to the "OK" button.

Image 18

We have just two values plus. For our average we will replace this value by 225, which is the highest we have.

The screenshot shows the Power BI Data Editor with a table titled "Table.SelectRows(#"Replaced Value1", each ([#"Q3 - Current Yearly Salary (in USD) - Copy.2"] = "+"))". The table has three columns: "Highest Level of Education", "Q13 - Ethnicity", and "Q3 - Current Yearly Salary (in USD) - Copy.1". The "Q3 - Current Yearly Salary (in USD) - Copy.1" column contains values 225 and 225. A context menu is open on the header of this column, showing options like Copy, Remove, Remove Other Columns, Duplicate Column, Add Column From Examples..., Remove Duplicates, Remove Errors, Change Type, Transform, Replace Values..., Replace Errors..., Split Column, Group By..., Fill, Unpivot Columns, Unpivot Other Columns, Unpivot Only Selected Columns, Rename..., Move, Drill Down, and Add as New Query. The "Replace Values..." option is highlighted with a red arrow.

Image 19

Replace the value of + with 225.

The screenshot shows the Power BI desktop interface with the 'Replace Values' dialog box open. The dialog box contains two input fields: 'Value To Find' with the value '+' and 'Replace With' with the value '225'. A red arrow points to the 'Value To Find' field, and an orange arrow points to the 'Replace With' field. A pink arrow points down from the dialog box towards the main Power BI interface.

Image 20

To unfelted these, select all and ok.

The screenshot shows the Power BI desktop interface with the 'Replace Values' dialog box open. The dialog box contains two input fields: 'Value To Find' with the value '+' and 'Replace With' with the value '225'. A red arrow points to the 'Value To Find' field, and an orange arrow points to the 'Replace With' field. A pink arrow points down from the dialog box towards the main Power BI interface. A red arrow also points to the 'OK' button in the dialog box. The main interface shows a table with columns 'Q13 - Ethnicity' and 'Q3 - Current Yearly Salary (in USD) - Copy.1'.

Image 21

Now we will construct the average column. Go to Add Column -> Custom Column.

The screenshot shows the Microsoft Power BI Data Editor interface. The top navigation bar has tabs: File, Home, Transform, Add Column, View, Tools, and Help. The 'Add Column' tab is currently selected. Below the ribbon, there are several icons for different data transformation operations like 'Conditional Column', 'Index Column', 'Duplicate Column', etc. A red arrow points from the 'Custom Column' icon on the ribbon to the 'Custom Column' option in the dropdown menu under 'Add Column'. The main workspace displays a table titled 'Data Professional Survey' with columns: 'Highest Level of Education', 'Q13 - Ethnicity', 'Q3 - Current Yearly Salary (in USD) - Copy.1', and 'Q3 - Current Yearly Salary (in USD) - Copy.2'. The table contains data rows for various ethnicities and their corresponding salaries.

Image 22

We will name the column for average salary. In the Custom column formula, we will write the average formula, $Q3 - \text{copy 1} + Q3 - \text{copy 2} / 2$. For the write the formula, open parenthesis and click on the **Q3 - copy 1** -> click on **insert** -> type "+" -> click on the **Q3 - copy 2** -> click on **insert** -> close parenthesis and **divide by 2**.

The screenshot shows the Microsoft Power BI Data Editor interface with the 'Custom Column' dialog box open. The dialog box title is 'Custom Column' and it says 'Add a column that is computed from the other columns.' It has a 'New column name' field containing 'Average Salary' and a 'Custom column formula' field containing the formula $= ([#\"Q3 - Current Yearly Salary (in USD) - Copy.1"] + [#\"Q3 - Current Yearly Salary (in USD) - Copy.2"]) / 2$. To the right of the formula is a list of 'Available columns' with several items highlighted in yellow. Arrows point from the formula text area to the 'Available columns' list, indicating the selection of columns for the formula. The background shows the 'Data Professional Survey' table with columns: 'Highest Level of Education', 'Q13 - Ethnicity', 'Q3 - Current Yearly Salary (in USD) - Copy.1', and 'Q3 - Current Yearly Salary (in USD) - Copy.2'. The 'Query Settings' pane on the right shows the 'Properties' section with 'Name' set to 'Data Professional Survey' and the 'Applied Steps' section listing various data processing steps.

Image 23

Next steps, we are broken the Q4, Q11, Q13 columns, as the same way than column Q1 (images 7,8 and 9).

The column *Q11 – Which country do you live* we are going to break because they have issues, there are countries with the wrong grammar (Argentina and Argentine means the same country but will be two different values). So, we cannot normalize those values.

Dashboard

Let us divide the dashboard into 8 blocks as shown in the image below (image 24).

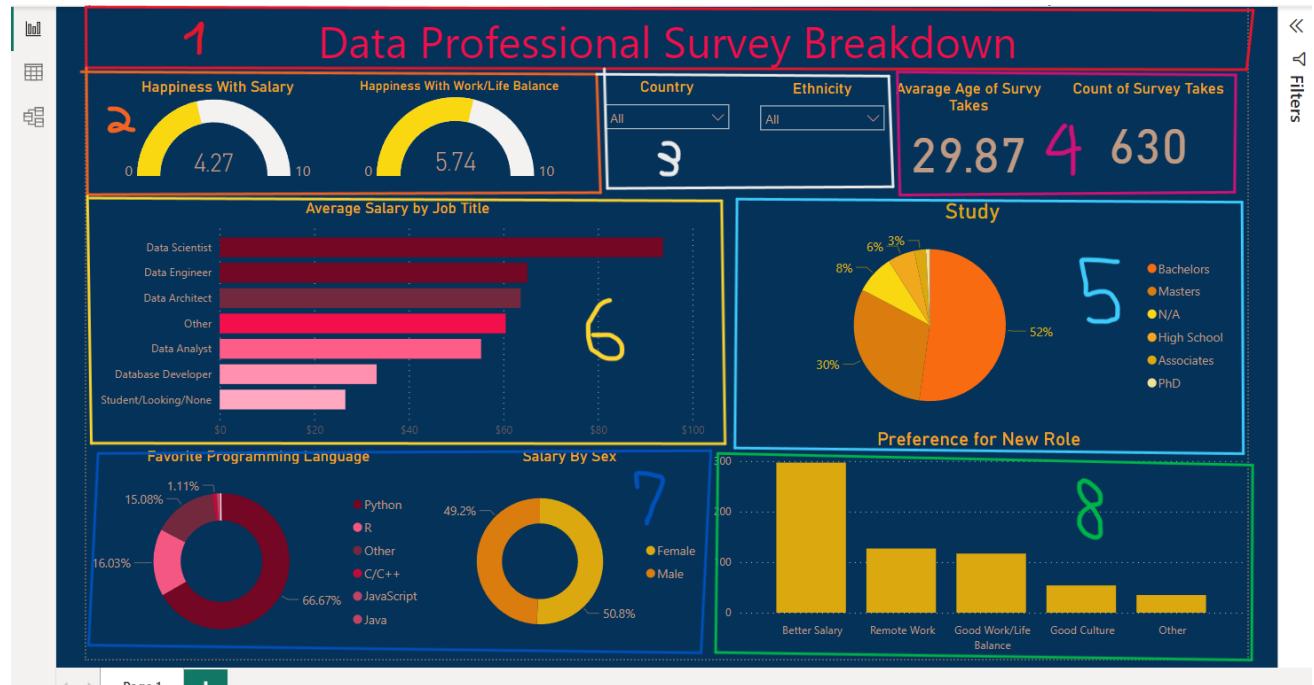
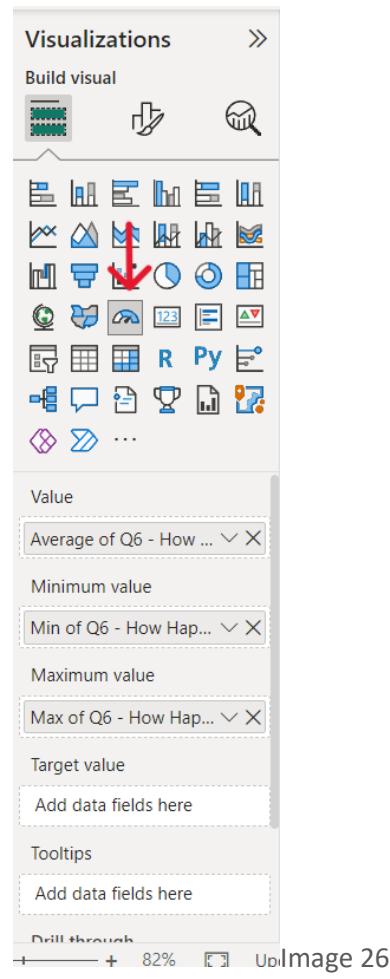
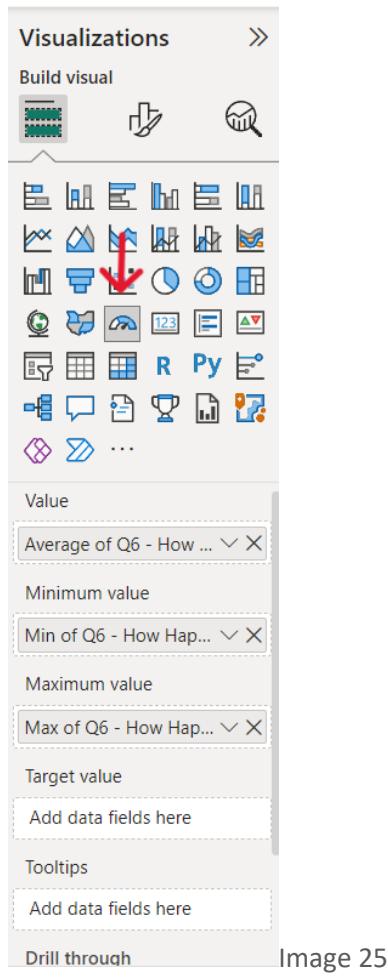


Image 24

Block 1: On the home page click **Text Box** -> write the text -> scrolls to the top of the dashboard.

Block 2: **Visualizations -> Gauge.**

- **Happiness With Salary parameters** (Image 25): *Variable = Q6 - How Happy are you in your Current Position with the following? (Salary)*
 - **Value = Average of variable.**
 - **Minimum Value = Min of variable.**
 - **Maximum Value = Max of variable.**
- **Happiness With Work/Life Balance** (Image 26): *Variable = Q6 - How Happy are you in your Current Position with the following? (Work/Life Balance)*
 - **Value = Average of Q6.**
 - **Minimum Value = Min of Q6.**
 - **Maximum Value = Max of Q6.**



Block 3: Visualizations -> Slicer -> Visual -> Slicer Setting -> Option -> Style -> Dropdown.

- **Country**
 - *Field = Q11.*
- **Ethnicity**
 - *Field = Q13.*

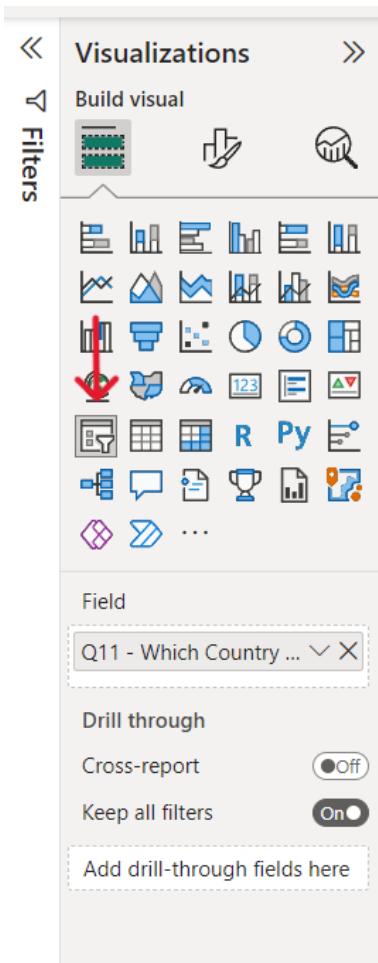


Image 27

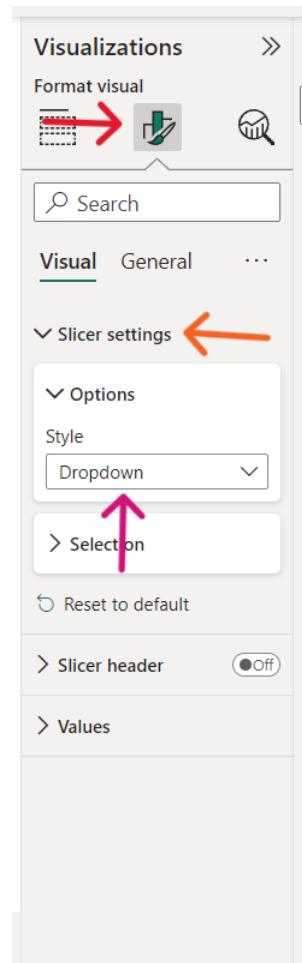


Image 28

Block 4: Visualizations -> Slicer.

- **Average Age of Survey Takes**
 - *Field: Q10* (purple arrow, image 29).
- **Count of Survey Takes**
 - *Field: Uniq ID* (blue arrow, image 29).

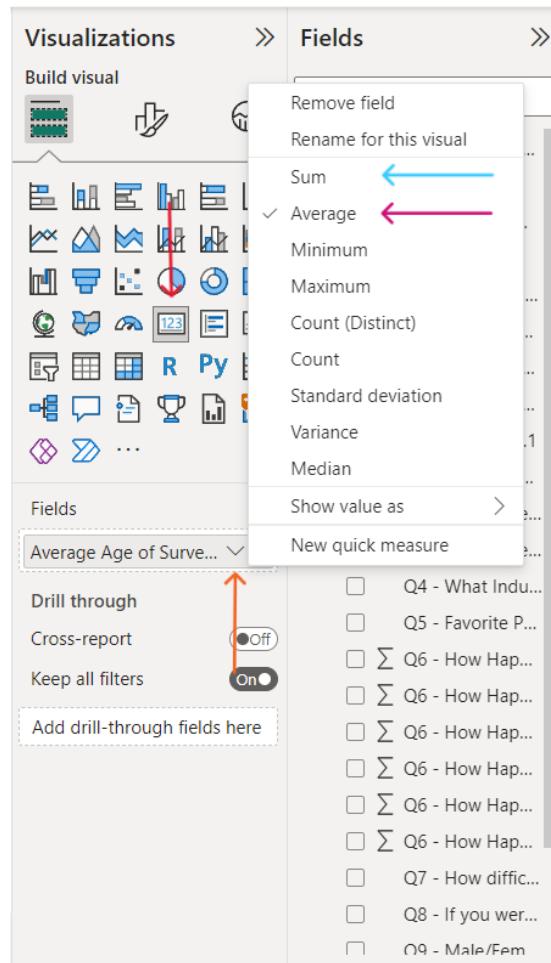


Image 29

Block 5: Visualizations -> Pie Chart.

- **Average Salary by Job Title**
- **Y-axis: Q1**
- **X-axis: Average Salary**
- **Legend: Job Title**

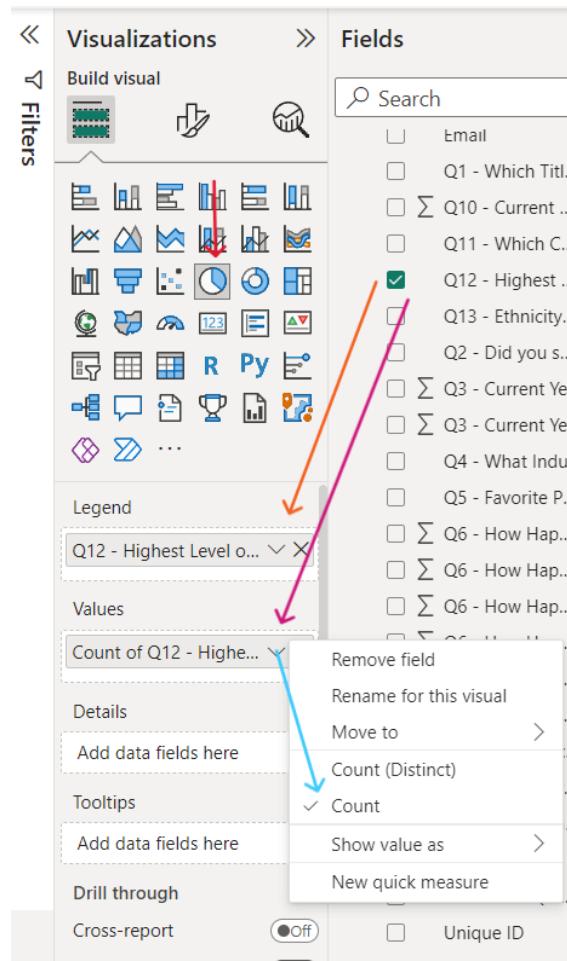


Image 30

Block 6: Visualizations -> Stacked Bar Chart.

- **Study:**
 - *Legend: Q12.*
 - *Values: Q12(Count).*

The screenshot shows the Power BI interface with the 'Visualizations' pane open. The 'Build visual' section is selected. The 'Y-axis' dropdown is set to 'Job Title'. The 'X-axis' dropdown is set to 'Average Salary'. The 'Fields' pane on the right lists various data items, with several checkboxes checked. Three red arrows point from the Y-axis and X-axis dropdowns to specific items in the Fields pane: 'Q1 - Which Tit...', 'Q6 - How Hap...', and 'Q7 - How diffic...'. Other visible items include 'Email', 'Q10 - Current ...', 'Q11 - Which C...', 'Q12 - Highest ...', 'Q13 - Ethnicity.1', 'Q2 - Did you s...', 'Q3 - Current Ye...', 'Q3 - Current Ye...', 'Q4 - What Indu...', 'Q5 - Favorite P...', 'Q6 - How Hap...', 'Q8 - If you wer...', 'Q9 - Male/Fem...', 'Time Spent', 'Time Taken (A...', and 'Unique ID'.

Image 31

Block 7: Visualizations -> Donut Chart.

- **Favorite Programming Language**
 - *Legend = Q5*
 - *Values = Uniq ID (Count)*
- **Salary by Sex**
 - *Legend = Q9*
 - *Values = Average Salary (Average)*

The screenshot shows the 'Build visual' interface in Microsoft Power BI. The left pane displays various visualization icons, with the 'Clustered Column Chart' icon highlighted by a red arrow. The right pane lists data fields under the heading 'Fields'. The field 'Q5 - Favorite P...' is checked, while others like 'Q8 - If you wer...' are unchecked.

Field	Status
Average Salary	unchecked
Date Taken (A...)	unchecked
Email	unchecked
Q1 - Which Titl...	unchecked
Q10 - Current ...	unchecked
Q11 - Which C...	unchecked
Q12 - Highest ...	unchecked
Q13 - Ethnicity.1	unchecked
Q2 - Did you s...	unchecked
Q3 - Current Ye...	unchecked
Q3 - Current Ye...	unchecked
Q4 - What Indu...	unchecked
Q5 - Favorite P...	checked
Q6 - How Hap...	unchecked
Q7 - How diffic...	unchecked
Q8 - If you wer...	unchecked
Q9 - Male/Fem	unchecked

Image 32

Block 8: Visualizations -> Clustered Column Chart.

- Preference for New Role
 - X-axis = Q8
 - Y-axis = Q8 (Count)

Save

After finishing the dashboard design, click on menu -> file -> save as (images 33{red arrow}and 34), or just click on the floppy disk button(image 33{orange arrow}).

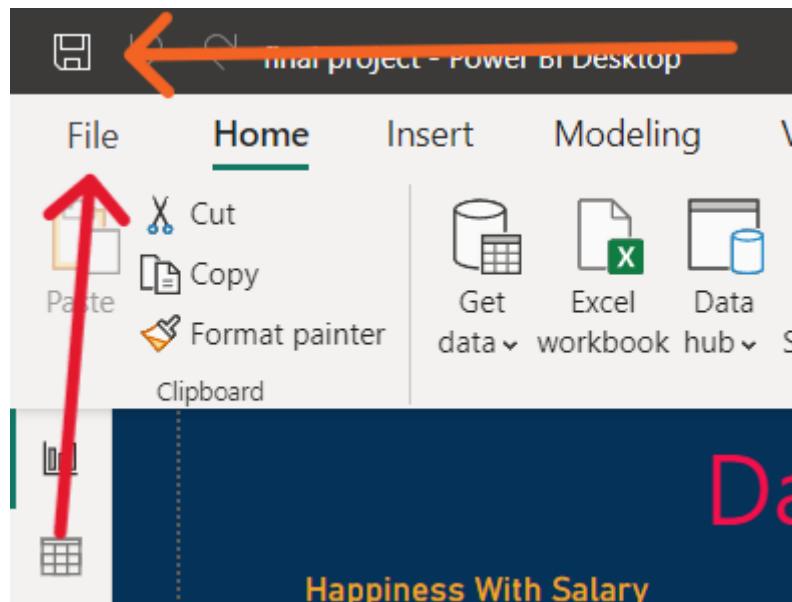


Image 33

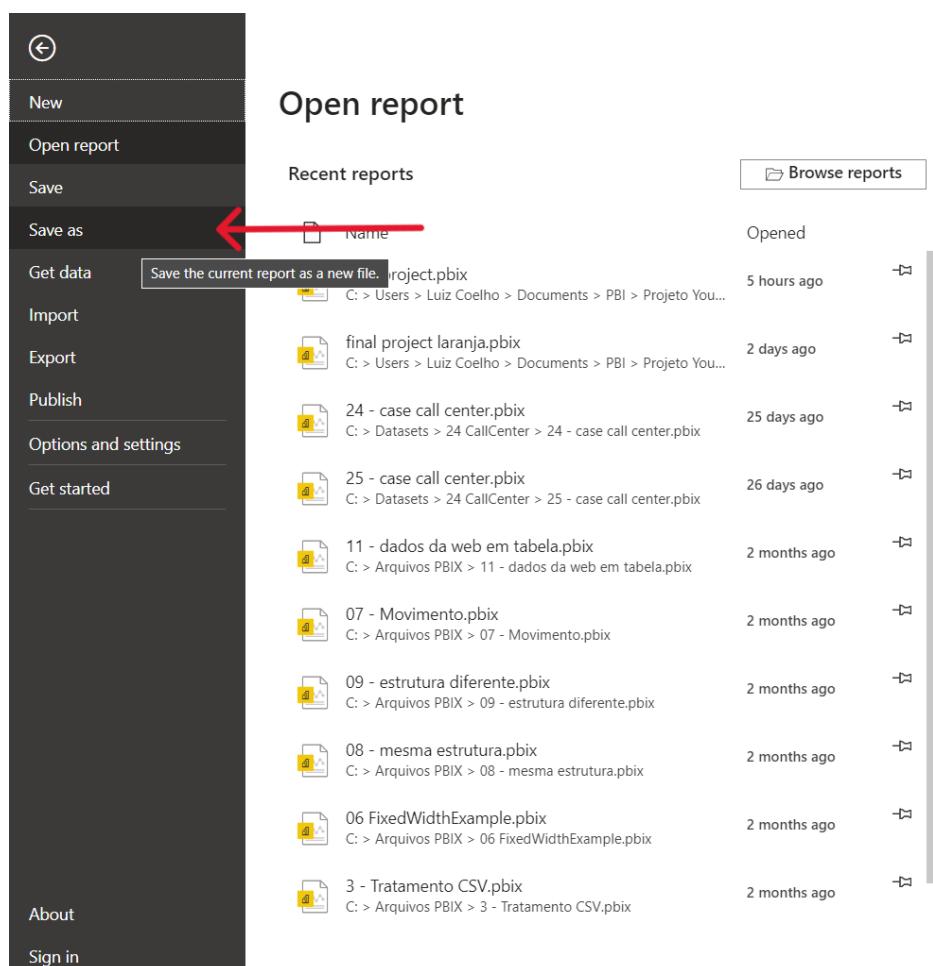


Image 34