My Walkthrough steps following Alex The Analyst, PowerBI Final project: Made by Luke Yeo

Note:

* Dataset: <https://github.com/AlexTheAnalyst/Power-BI/blob/main/Power%20BI%20-%20Final%20Project.xlsx>
* Just for this project, there will be no data cleaning in Excel for this project, will do it in PowerBI
* Data cleaning is not the Main focus, the cleaning in this project is an extremely rough cleaning, more in depth cleaning can be done on own time
* Designing a dashboard is not the Main/important aspect to me for this project (Last part)

Screening from the Data In Excel for a few seconds:

1. Lots of Blank column, may delete
2. Email is just ‘anonymous’, may delete
3. Unique\_ID is mix of numbers and letters, need to check if there is any duplication (if it is even allowed)
4. Date is different though it goes month/day/year I think, not day/month/year
5. Salary has a range
6. Some columns have a lot of options like the “Job” column
7. 2 different types of time, 1 is time taken to do the survey, other is when they do the survey

Loading the data to PowerBI:

* Going to transform data after importing the data, realised that there is an error in the Date column for the dataset, it is showing “error”, this is due to PowerBI reading it as day/month/year, and not month/day/year (where the number is higher than 12 for the month column)

Formatting Dates in PowerBI:

(Can only use if it’s just formatting directly or if the date starts off with “ABC” format , not “Error” immediately) :

Solution 1:

1. Split the date using “Split column” -> Delimiter, ‘/’
2. Rename each column according to day, month, year … then put them in that order ( month, day, year -> day, month, year )
3. Select all 3 columns, merge column then custom, ‘/’

Solution 2:

* Changing regional setting (File -> Options -> Region)

Solution 3:

* Change the date format using Local -> Region ‘English-US’, change the format to ‘Date’

**Sadly nothing worked for me so I will actually do this in Excel instead.**

* Change the Date of the first of each day, then ‘CTRL’ drag the right bottom corner till the next day

Data Cleaning:

* Delete Empty columns, ’Browser’ to ‘Referrer’
* For Q1, there are too many title that have the same job title but due to specific, everything is classified as its own individual. (Normally we will use excel to handle stuff like that but we are strictly trying to only use PowerBI so we will do it the “easy” way)

**Best method** is to spend A LOT more time to change/replace the value (aus, austr, australia, ect -> all in to value ‘Australia’.

Solution: Split Column by delimiter -> Custom -> ‘(’, left most. Then remove the additional column

* For Q4, do the same as Q1 (This can be done for the other columns too if want)
* For Q5, change the thing to ‘Colon’ rather than Custom, ‘(‘
* For Q3, since the salary is a range, we will use the median/average of the range (0-30K), we will put 15K), IT IS NOT A PERFECT SOLUTION but at least it makes it usable – an numeric value

Solution 1:

* Step 1: Duplicate the original cell. Split them by ‘digit to non-digit’, delete the cell that only has ‘k’.
* Step 2: Right click the 2nd column -> replace values ‘k’ with ‘ ‘, ‘-‘ with ‘ ‘ , ‘+’ with ‘225‘.
* Step3: Change the 2 columns to numerical first (‘Whole numbers’), then create a custom column call “Average Salary” with the formula ‘([#"Q3 - Current Yearly Salary (in USD) - Copy.1"] + [#"Q3 - Current Yearly Salary (in USD) - Copy.2"]) / 2 ‘.
* Step 4: Finally delete the split columns and move the Average salary to beside the original Q3
* Step 5: Change the Data type of Average Salary from a ‘Text’ to a ‘Decimal’

Data Visualization (Creating the graphs):

Note: For each X, Y or anything being used, do rename the column and title so it fits well.

1. Create a text box as our title ‘Data survey breakdown’, change the effect -> background
2. Create a Card -> Unique\_ID -> In the fields, change it to ‘Distinct Count’, change name to ‘Count of Survey Takers’
3. Create a Card -> Current Age, Q10 -> In the fields, change it to ‘Average’, change name to ‘Average Age of Survey Taker’
4. Create a Stacked Bar Chart -> X: Average Salary (by default it will be SUM, change it to Average), Y & Legend: Job title, Change the title to ‘Avg pay by job’ in general -> title
5. Create a Stack Column Chart -> X: Fav programming language, Y: Unique\_ID. Change the title to ‘Fav programming languages’. Add ‘Job title’ to the legend
6. Create a Tree Map (for other projects, try the ‘Filled Map’) -> Country data in both category and value
7. Create a Guage -> Work life balance (for value, minimum value, maximum value), change each of them (avg, mini, max). Change the title to ‘Work life balance rating’
8. Create a Guage -> Salary satisfactory (for value, minimum value, maximum value), change each of them (avg, mini, max). Change the title to ‘Salary Satisfactory rating’
9. Create Pie Chart or Donut Chart -> Gender (legend), Values can be whatever, in this case , I pick ‘Average salary’ or ‘Difficulty into data science’

Data Visualization (Designing – colour, shapes, ect ):

* For Donuts and Pie chart, change the colour at ‘Slices’
* At the top, go to View, then apply the filter, I picked the dark looking ones (the colors are able to be customize too)