09-09-2025

WEB SCRAPPING:

- We started by creating a new report in Power BI and named it as web scraping.
- In google, I browsed for World Happiness Report
- Now load the contents of url select table -17 and load it to the Power BI.
- Rename the table as WHR.
- Go to table view and double-check the data types of the columns change the data category of country column to country/region.

COLUMNS AND THEIR DESCRIPTION:

1. Overall Rank

- **Definition**: Position of a country in the report based on its **Happiness Score** (descending order).
- **Formula**:

  Rank(i)=Position of country i when countries are sorted by Score (highest → lowest)

2. Country or Region

- **Definition**: Name of the country or region.
- **No formula** – just a label.

3. Score (Happiness Score / Ladder Score)

- **Definition**: Average response to the **Cantril Ladder question**:

"Imagine a ladder with steps from 0 (worst possible life) to 10 (best possible life). On which step do you stand?"

- **Formula**:

  Score=N∑(Responses) / N
  where N = number of survey respondents.

4. Log GDP per Capita

- **Definition**: Natural log of GDP per capita (PPP, constant international $).
- Used instead of raw GDP to reduce skewness.
- **Formula**:

  Log GDP per capita=ln(GDP per capita, PPP)

5. Social Support

- **Definition**: Measure of perceived social support.
- Based on the Gallup World Poll question:

"If you were in trouble, do you have relatives or friends you can count on to help you when needed?"

- **Formula**:

  Social Support=Yes responses/N

6. Healthy Life Expectancy

- **Definition**: Expected years of life in good health.
- Taken from WHO/World Bank data.
- **Formula (as provided by WHO)**:

Healthy Life Expectancy=Life Expectancy at Birth−Years in Ill Health

7. Freedom to Make Life Choices

- **Definition**: Measure of perceived freedom.
- Based on Gallup World Poll question:

"Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"

- **Formula**:

  Freedom Score=Satisfied responses/N

8. Generosity

- **Definition**: Propensity to donate and share with others.
- Based on Gallup World Poll question:

"Have you donated money to a charity in the past month?"

- Adjusted for income using regression (so it reflects generosity independent of GDP).
- **Formula (simplified)**:

Generosity=Residual from (Charitable donations vs GDP model)

9. Perceptions of Corruption

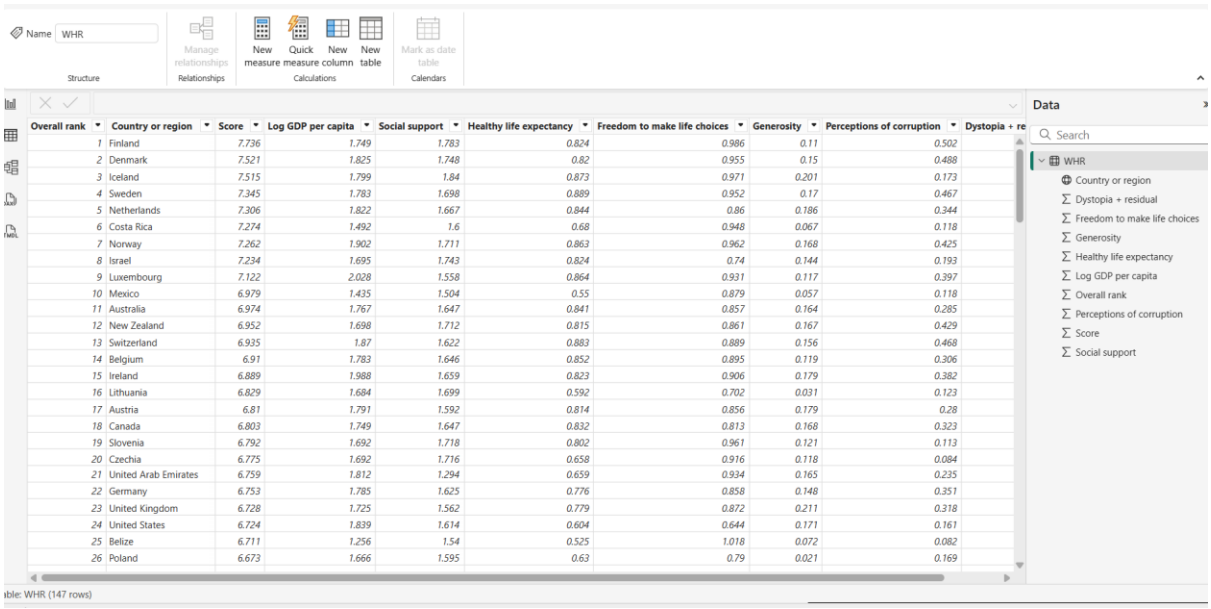
- **Definition**: Trust in government and business.
- Based on Gallup World Poll questions:
    - "Is corruption widespread throughout the government?"
    - "Is corruption widespread within businesses?"
- **Formula**:

Perceptions of Corruption=1−(Yes responses (avg of both)/N)

10. Dystopia + Residual

- **Definition**: A baseline reference plus the unexplained part of happiness.
    - *Dystopia* = a hypothetical "lowest possible country" with worst values for all factors.
    - *Residual* = difference between actual Score and the sum of contributions from all measured factors.
- **Formula**:

  Dystopia + Residual=Score−(β1(lnGDP)+β2(Social Support)+β3 (Healthy Life)+β4(Freedom)+β5(Generosity)+β6(Corruption))
- This ensures that no country has negative contributions and that all Scores can be reconstructed.

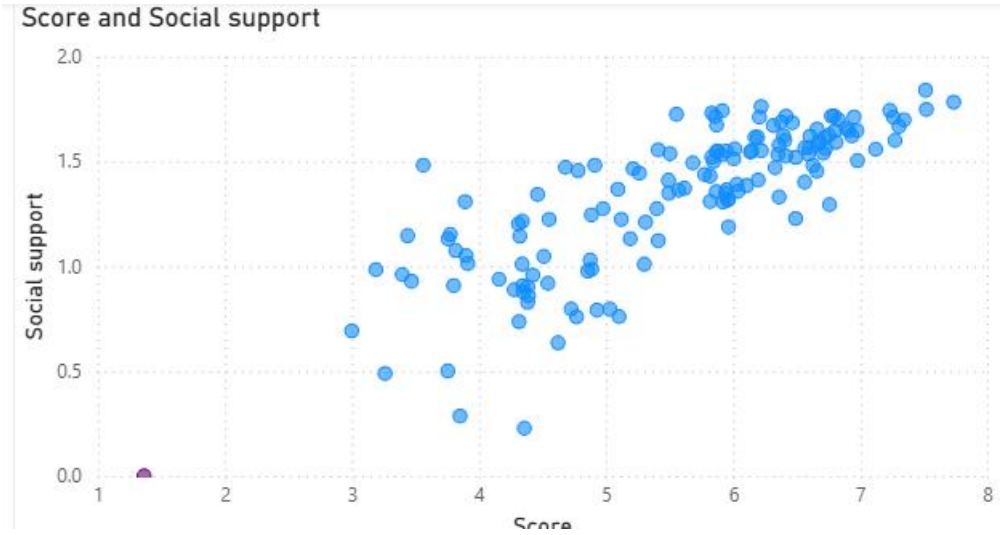


Overall rank	Country or region	Score	Log GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Dystopia + re
1	Finland	7.736	1.749	1.783	0.824	0.986	0.11	0.502	
2	Denmark	7.521	1.825	1.748	0.82	0.955	0.15	0.488	
3	Iceland	7.515	1.799	1.84	0.873	0.971	0.201	0.173	
4	Sweden	7.345	1.783	1.698	0.889	0.952	0.17	0.467	
5	Netherlands	7.306	1.822	1.667	0.844	0.86	0.186	0.344	
6	Costa Rica	7.274	1.492	1.6	0.68	0.948	0.067	0.118	
7	Norway	7.262	1.902	1.711	0.863	0.962	0.168	0.425	
8	Israel	7.234	1.695	1.743	0.824	0.74	0.144	0.193	
9	Luxembourg	7.122	2.028	1.558	0.864	0.931	0.117	0.397	
10	Mexico	6.979	1.435	1.504	0.55	0.879	0.057	0.118	
11	Australia	6.974	1.767	1.647	0.841	0.857	0.164	0.285	
12	New Zealand	6.952	1.698	1.712	0.815	0.861	0.167	0.429	
13	Switzerland	6.935	1.87	1.622	0.883	0.889	0.156	0.468	
14	Belgium	6.91	1.783	1.646	0.852	0.895	0.119	0.306	
15	Ireland	6.889	1.988	1.659	0.823	0.906	0.179	0.382	
16	Lithuania	6.829	1.684	1.699	0.592	0.702	0.031	0.123	
17	Austria	6.81	1.791	1.592	0.814	0.856	0.179	0.28	
18	Canada	6.803	1.749	1.647	0.832	0.813	0.168	0.323	
19	Slovenia	6.792	1.692	1.718	0.802	0.961	0.121	0.113	
20	Czechia	6.775	1.692	1.716	0.658	0.916	0.118	0.084	
21	United Arab Emirates	6.759	1.812	1.294	0.659	0.934	0.165	0.235	
22	Germany	6.753	1.785	1.625	0.776	0.858	0.148	0.351	
23	United Kingdom	6.728	1.725	1.562	0.779	0.872	0.211	0.318	
24	United States	6.724	1.839	1.614	0.604	0.644	0.171	0.161	
25	Belize	6.711	1.256	1.54	0.525	1.018	0.072	0.082	
26	Poland	6.673	1.666	1.595	0.63	0.79	0.021	0.169	

CORRELATION:

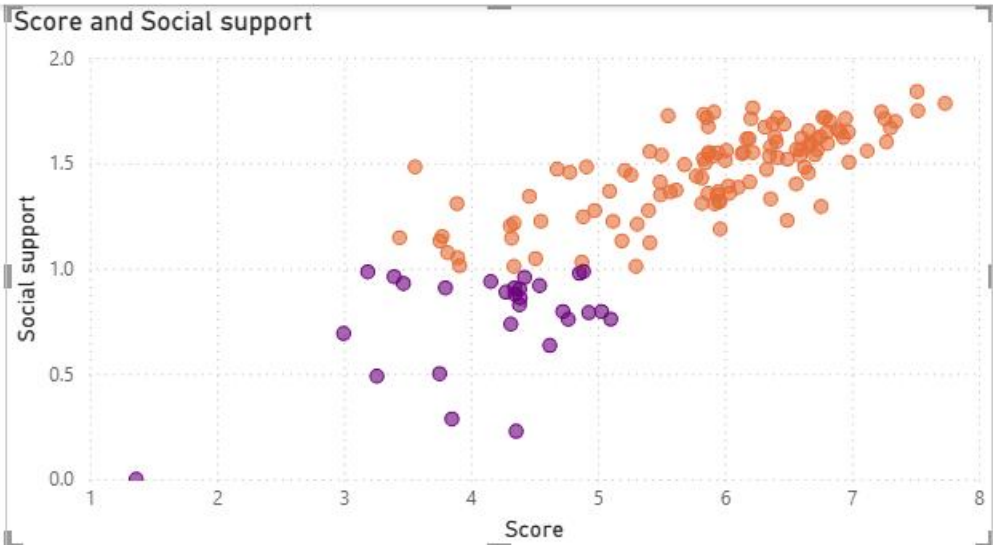
- When we have more numerical data and only one categorical data then we go with the concept of correlation.
- **positive correlation**: two numerical values which are directly proportional to each other.
- **no correlation**: numerical values without any relation.
- **negative correlation**: numericals which are indirectly proportional to each other.
- **charts used**: Scatter plot

SCATTER PLOTS:

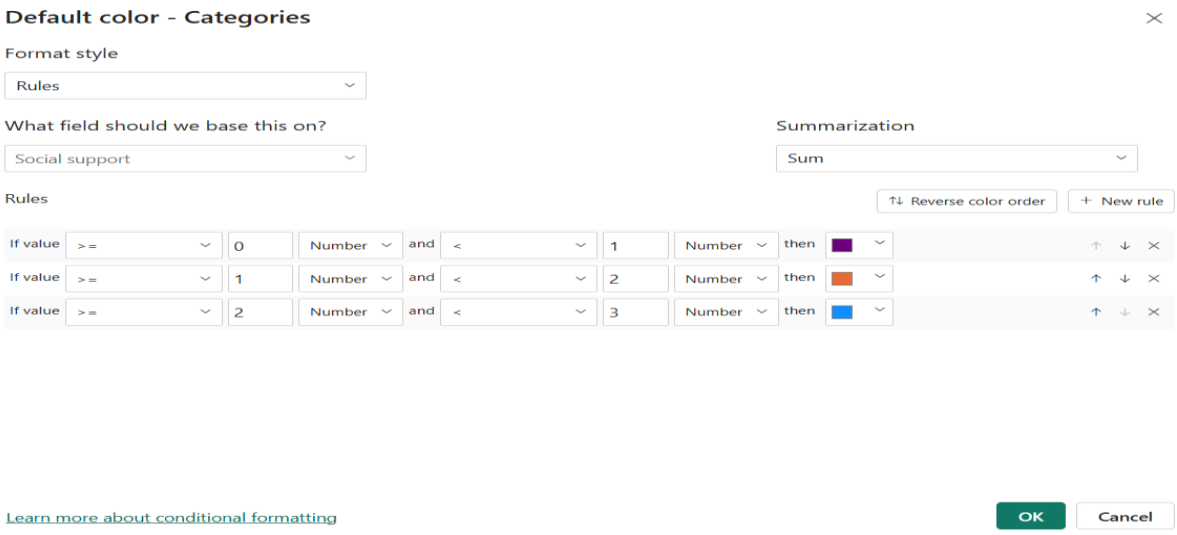


STEPS:

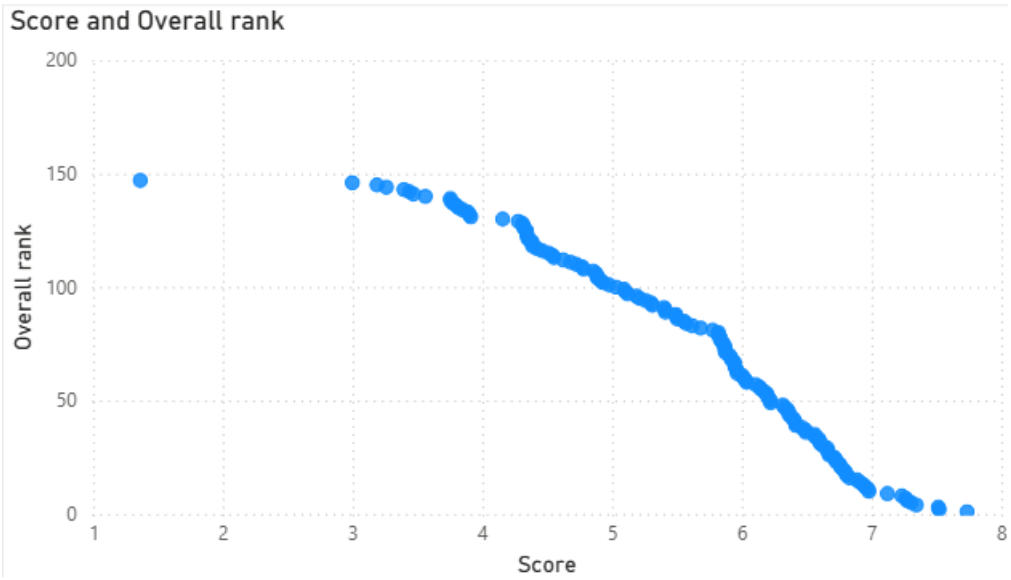
- From visualization pane select a scatter plot.
- Drag and drop score to the x-axis and social support to the y-axis.
- Default the columns will be summarized, we don't want them to be summarized so click on down arrow in column name and select don't summarize.
- Now we can see the scatter plot.
- From the visual we can interpret that score and social support are weak positively correlated.



- We have different scatter plot formattings, one of them is conditional formatting
- In this scenario I had gave different colours for the markers based on rules.

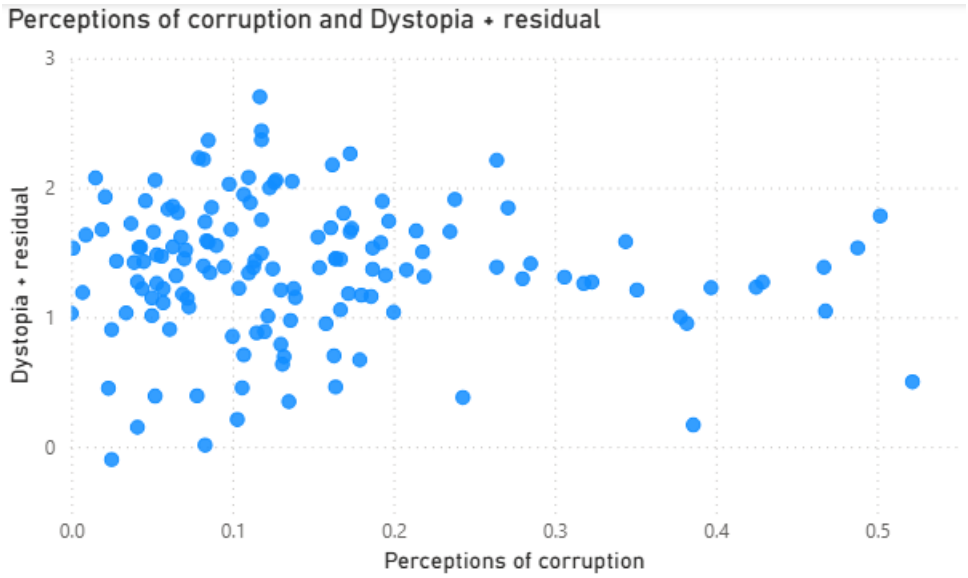


- Go to format visual and navigate to markers section, then scroll to colour and we find fx button for conditional formatting.
- Select format style as rules,field to be based as social support
- Create rules and specify different colours to the different rules and click on ok.



STEPS:

- From visualization pane select a scatter plot.
- Drag and drop score to the x-axis and overall rank to the y-axis.
- Default the columns will be summarized, we don’t want them to be summarized so click on down arrow in column name and select don’t summarize.
- Now we can see the scatter plot.
- From the visual we can interpret that score and overall rank are strongly negative correlated.
- But logically the interpretation will be mis leaded as the rank is in descending order so while dealing with ranking in the plot be aware of the order of ranking.
- We can change the order of ranking in the table view.



STEPS:

- From visualization pane select a scatter plot.
- Drag and drop perceptions of corruption to the x-axis and Dystopia + residual to the y-axis.
- Default the columns will be summarized, we don't want them to be summarized so click on down arrow in column name and select don't summarize.
- Now we can see the scatter plot.
- From the visual we can interpret that the perceptions of corruption and Dystopia + residual have no correlation.