**09-09-2025**

**WEB SCRAPPING:**

* We started by creating a new report in power bi and names it as web scrapping.
* In google browsed for world happiness report and navigated to the table section of 2025 and copied its url.
* In the power bi menu bar there will be an option called get data in that navigate to web,
* Click on web and keep the specification as basic and paste the url
* 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**:

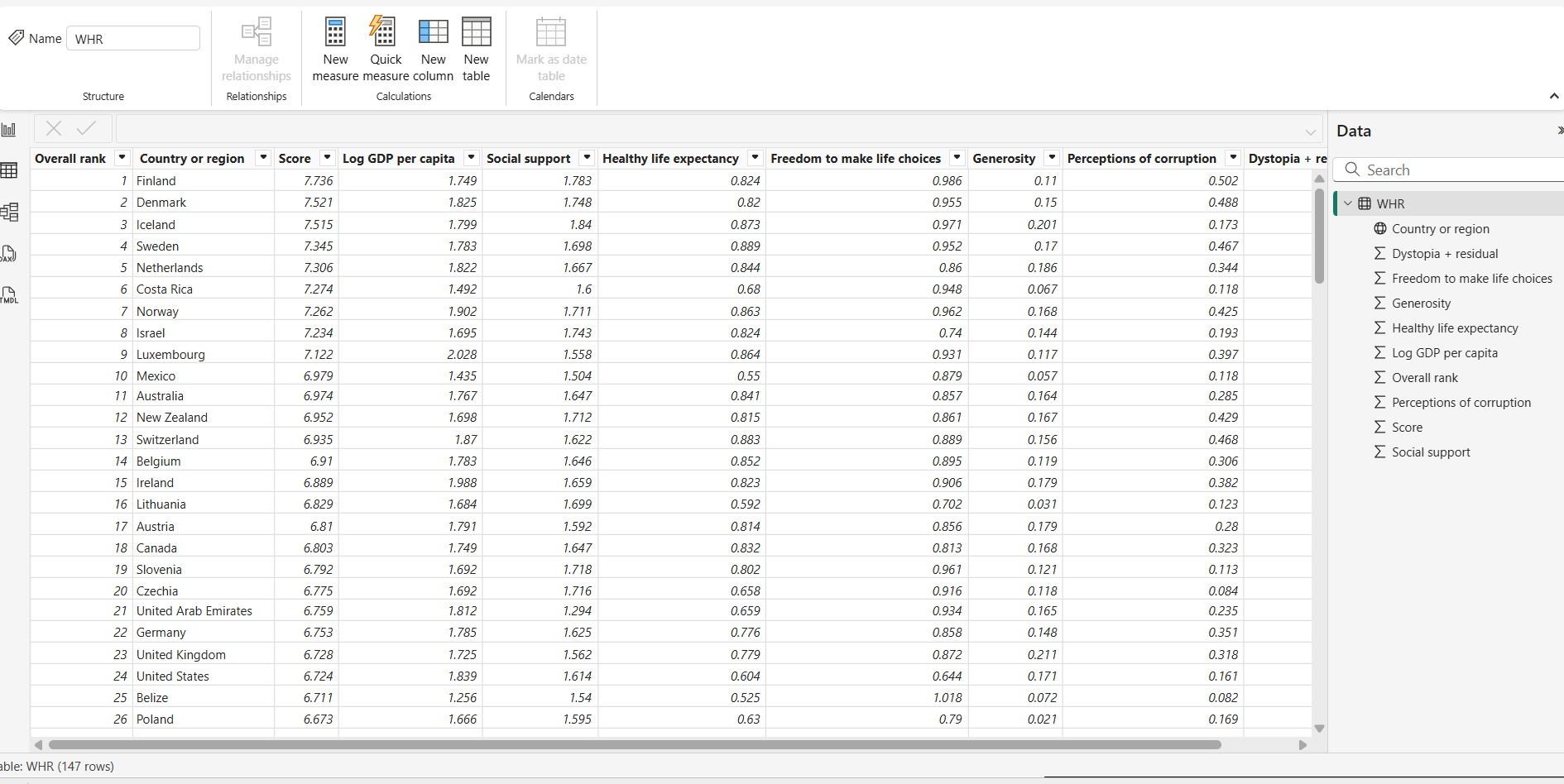
Perceptions of Corruption=1−(Yes responses (avg of both)​/N)

**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))

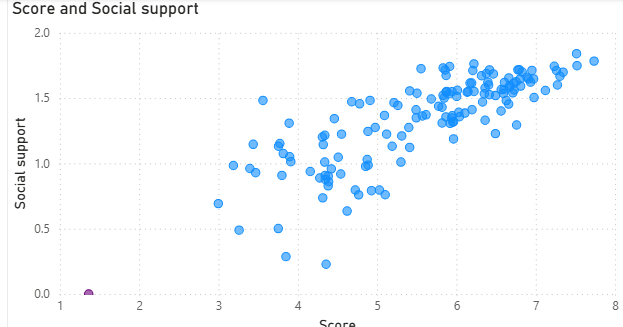
* This ensures that no country has negative contributions and that all Scores can be reconstructed.



**CORRELATION:**

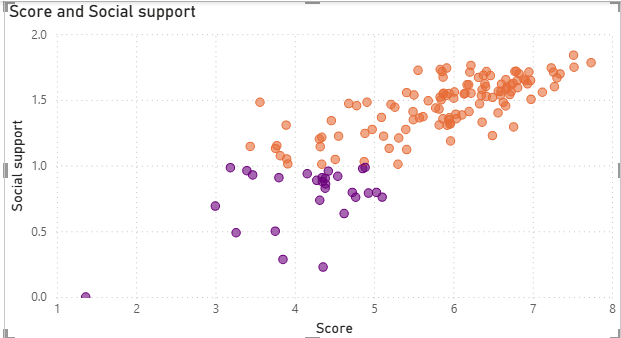
* When we have more numerical data and only one categorical data then we go with the concept of correlation.
* **positive correlation**: two numerical values which are directly proportional to each other.
* **no correlation**: numerical values without any relation.
* **negative correlation**: numericals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

**SCATTER PLOTS:**

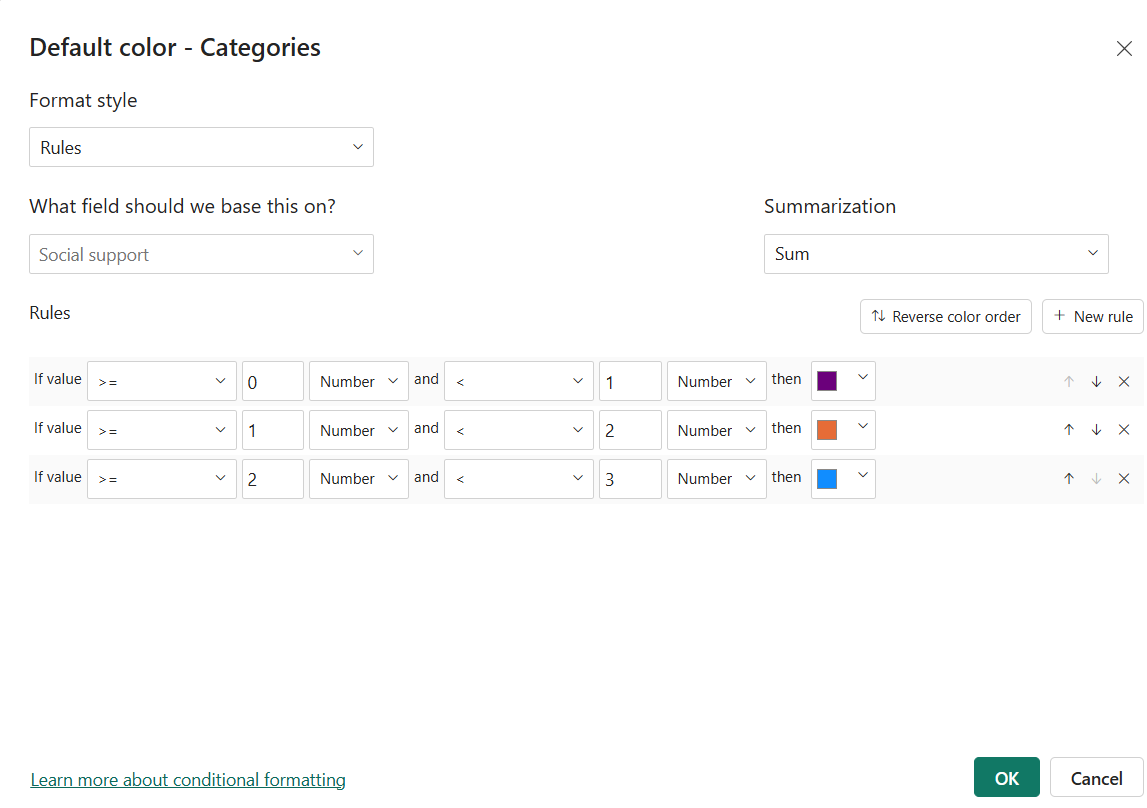


STEPS:

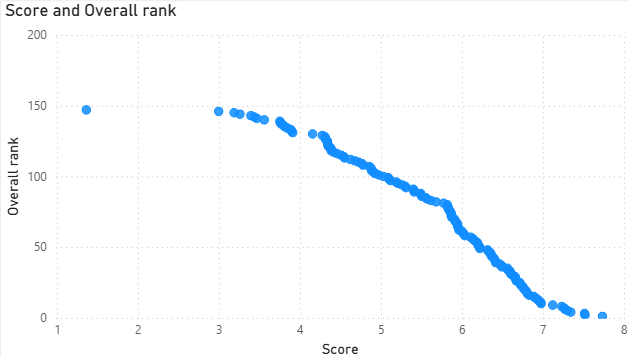
* From visualization pane select a scatter plot.
* Drag and drop score to the x-axis and social support to the y-axis.
* Default the columns will be summarized, we don’t want them to be summarized so click on down arrow in column name and select don’t summarize.
* Now we can see the scatter plot.
* From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

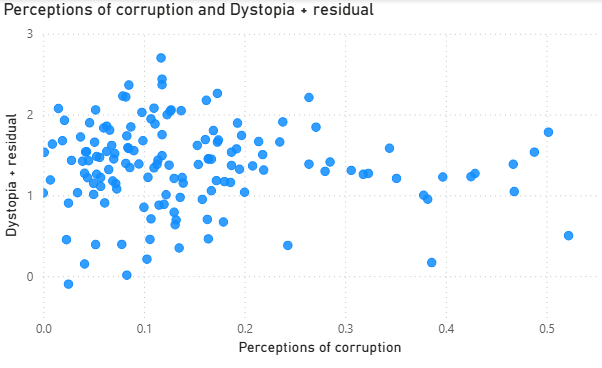


* Go to format visual and navigate to markers section, then scroll to colour and we find fx button for conditional formatting.
* Select format style as rules,field to be based as social support
* Create rules and specify different colours to the different rules and click on ok.



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.