

# Term 2 Project - Factors contributing to a Nation's Happiness

*Hugo TING, Tunay TOKMAK, Seng Moon Ja*

## Research Questions

The main goal of the project is to find out the state of happiness of 30 European countries in 2021. The term “happiness” of a country is a combination of positive feelings that provide a general sense of satisfaction. To explore this combination, we look into the questions as follows:

- Which European country has the highest / lowest happiness level?
- How does each of the social aspects affect people's happiness?
- Is there any correlation between the education level, GDP per capita, press freedom, population density and corruption within the country?

## Data Source

To explore these questions we acquire data from the following links:

[Happiness Data](#)

[World Press index](#)

[Country Code](#)

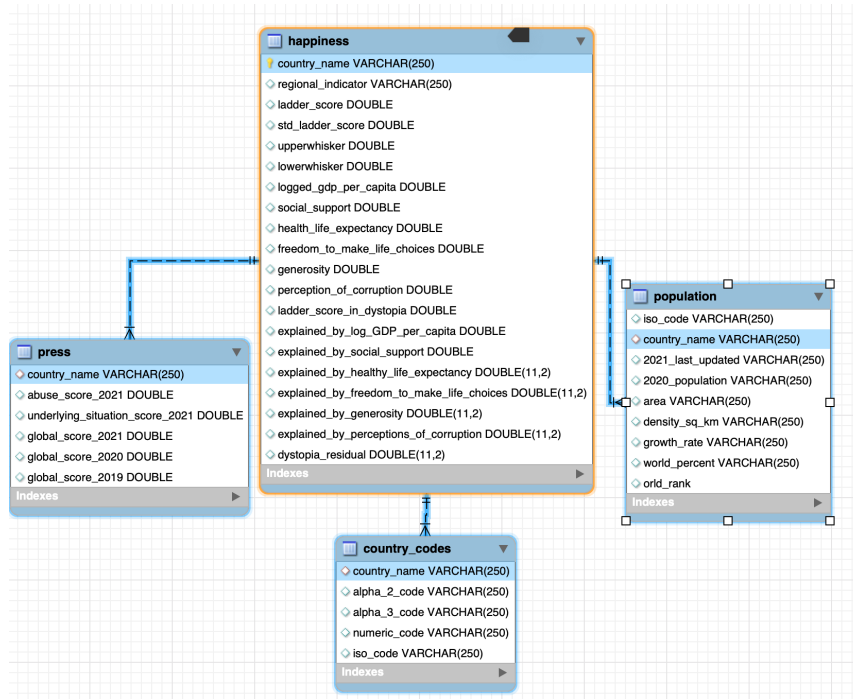
[Population Stat](#)

Education level from Eurostat API (JSON)

## Set Up

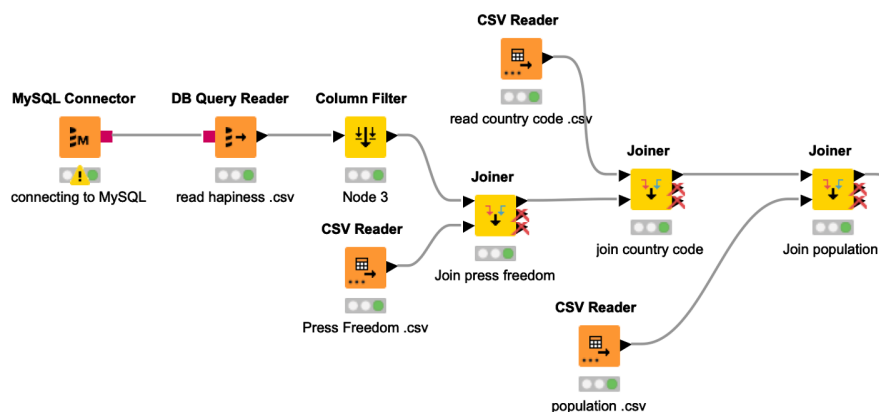
### MySQL

For this project happinessdb database was created and in that database 4 tables are built namely “*happiness*”, “*press*”, “*population*” and “*country code*”. Within the tables the interconnection are made through foreign key from each table. Each table's data is inserted from 4 different .csv files. The ER diagram which shows the relationship among the tables. The “*press*”, “*country\_codes*” and “*population*” tables are connected with their foreign key to the “*happiness*” primary key, which is the country\_name.



## Extract, Transform and Load pipeline in KNIME

### Extract



**MySQL Connector:** Connecting KNIME with the happinessdb in MySQL

**DB Query Reader:** Read happiness table from happinessdb

**Column Filter:** Filter out duplicate columns from “happiness” table

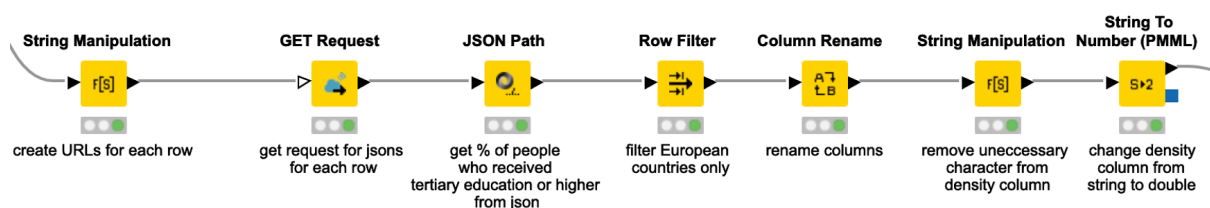
**Joiner 1:** Join the node between happiness table and Press Freedom.csv with “country name” from each data. It joins the left table which is happiness and the inner join method is used and merges joined columns.

**Joiner 2 :** To the Joiner 1, the country code.csv is joined with “country name” by inner join method. This table is joined to match the country codes in the JSONs in later stages.

**Joiner 3 :** To the Joiner 2, the population.csv is joined with iso\_code to Alpha-3 code and inner join method is used.

In order to include the data of “the percentage education level” from European countries into our KNIME, we use [Eurostat](#) data with API requests. The data is retrieved in JSONs format with API request. In order to be able to access it, we use three steps such as String Manipulation, GET Request, JSONPath.

## Transform



**String Manipulation :** Country code from “country code” and “country code” from Euro-Stat data are matched and created a new column to store the requested URL from Eurostat for each row.

**Get Request:** In each row of matched country code, a JSONS path that contains the data from Euro stat, specifically the data of “the percentage of people with tertiary education” was inserted.

**JSON Path:** The definite JSON path is converted to the selected KNIME type.

**Row Filter:** Filtering only European countries i.e deselect the data that is N/A.

**Column Rename :** Rename the column to be able to easily read and remember.

**String Manipulation:** Since there is unnecessary character in the population density per km square from population data, therefore those characters are removed.

**String to Number:** In order to be able to calculate, the population density per km square column is converted to double from string format.

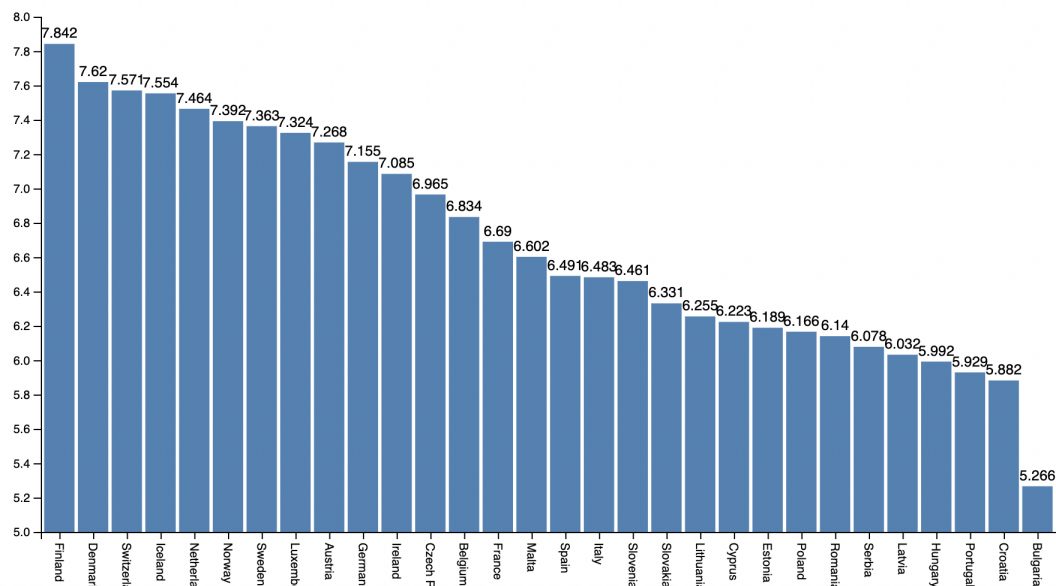
## Load

**Column Filter:** Filter columns which are only necessary for visualisation or analysis, saved into a data warehouse.

**Maths Formula:** The Press Freedom Score was in a reverse scale (the lower the value, the more free it is). In order to show the correlation correctly in visualisations, we subtract the press freedom score from 100 ie. (100-Press Freedom Score 2021\$).

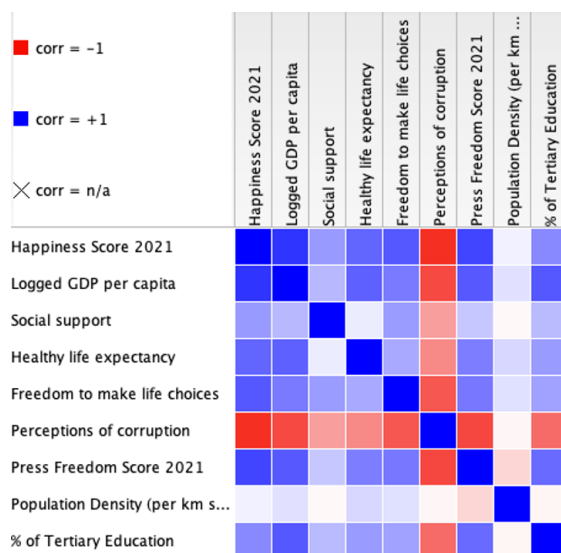
# Findings

## The Happiness of European countries



From the above bar chart, we can see the happiness level ranking of 30 European countries, Finland is the country with the highest happiness level, while Bulgaria is the country with the least happiness score.

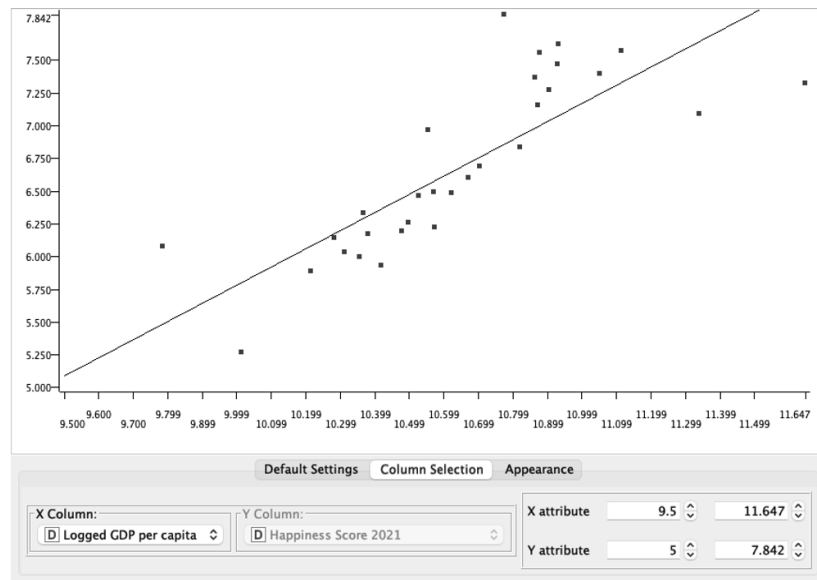
## Correlation between indicators



The figure above shows the correlation between indicators we are looking into. The results show that there is a strong positive correlation between press freedom and happiness score (corr. coeff = 0.75). There is a negative correlation between perceptions of corruption and happiness score, as the lower the perceptions of corruption in the country, the country will be happier (corr. coeff = -0.85). There is no correlation between population Density and happiness score (corr. coeff = 0.06) though we assume that higher population density will lead to least happiness. There is a mild positive correlation

between % of people with tertiary education(corr. coeff = 0.50) however, it is not as strong as other factors.

## Relationship between Socio-Economic status and happiness of people



The scatter and regression plot above demonstrates the distribution of countries according to their GDP per capita and happiness level. In this project, the GDP per capita of a country is the socio-economic status indicator. In the scatter plot, Luxembourg has the highest GDP per capita in Europe (the rightmost dot). However, it is not the happiest country in Europe. Though Finland's GDP per capita is not the highest, its scores of happiness are the highest among European countries (the topmost dot).

In general, the regression curve is upward sloping, which means that if the country performs better socio-economically, the higher the probability of people being happier. On the left of the regression line, countries are happier than the “expected happiness”, probably they have other factors other than socio-economic status which make people happier. While countries on the right of the regression line are not as happy as expected, as their country did not perform well in other aspects which are positively correlated to the level of happiness.

## Project Summary

Answering our analytics questions, from the visualised results of the data warehouse, Finland is the happiest country in 2021, while Bulgaria is the least happy country.

| Positively Correlated Aspects          | Negatively Correlated Aspect |
|--|------------------------------|
| Socio-Economic<br>(+0.8)               | Corruptions<br>(-0.85)       |
| Freedom of Press<br>(+0.75)            |                              |
| Freedom to make life choice<br>(+0.68) |                              |
| Health<br>(+0.64)                      |                              |
| Education<br>(+0.50)                   |                              |
| Social Support<br>(+0.43)              |                              |
| Population Density<br>(+0.06)          |                              |

From the table above, the correlation between different social aspects and happiness level of people can be summarised. Most of the aspects are positively correlated as the focus of our project is to find out the aspects which possibly increase the happiness level of people.

The correlation between “population density” and happiness level is out of our expectation. We have expected a negative correlation as people always complain about crowded public areas, noticeably on public transports. However, from the results, population density is likely to have no correlation with people’s happiness level.

Even though there is a saying “money can’t buy happiness”, GDP per capita has a strong relationship with a nation's happiness.

## Division of Labor

Every team member contributed, collaborated and workload were distributed equally throughout the whole project. The main contributor of each parts are as follows:

MySQL database creation: Tunay

KNIME Workflow: Hugo

Project Report: Seng Moon and Hugo

Presentation Slides: Hugo and Seng Moon