Assignment

By Jonathan McDonagh - 20074520

BSc (Honours) Software Systems Development

Business Analytics 2

Table of Contents

[Introduction 3](#_Toc34574173)

[Aim 3](#_Toc34574174)

[Preparation 3](#_Toc34574175)

[Data Selection 3](#_Toc34574176)

[Data Cleaning 3](#_Toc34574177)

[Analyse 4](#_Toc34574178)

[Focused Countries 4](#_Toc34574179)

[Ireland and Netherlands Focused Variables 4](#_Toc34574180)

[Ireland and Netherlands Descriptive Statistics 5](#_Toc34574181)

[Countries with an Income between Ireland and Netherlands 7](#_Toc34574182)

[Associations/Correlations 9](#_Toc34574183)

[Relationship between Alcohol Consumption and Suicide Rate 9](#_Toc34574184)

[Conclusion 10](#_Toc34574185)

[References 11](#_Toc34574186)

Table of Figures

[Figure 1: My Focused Countries 4](#_Toc34574333)

[Figure 2: Ireland and Netherlands 4](#_Toc34574334)

[Figure 3: Internet Use Rate and Alcohol Consumption (Bar Charts) 5](#_Toc34574335)

[Figure 4: Descriptive Statistics for all countries 6](#_Toc34574336)

[Figure 5: Descriptive Statistics for Ireland and Netherlands 6](#_Toc34574337)

[Figure 6: Query to restrict observations 7](#_Toc34574338)

[Figure 7: Countries with income between Ireland and Netherlands 7](#_Toc34574339)

[Figure 8: Alcohol Consumption Bar Chart 7](#_Toc34574340)

[Figure 9: Internet Use Rate Bar Chart 8](#_Toc34574341)

[Figure 10: Suicide Per 100,000 Bar Chart 8](#_Toc34574342)

[Figure 11: Alcohol Consumption and Suicide Rate Correlation 9](#_Toc34574343)

# Introduction

For this assignment I will be investigating the Gap Minder dataset. I will be focusing my attention on Ireland and Netherlands comparing these against the following variables Income Per Person, Internet Use Rate, Alcohol Consumption, Life Expectancy and Suicide Per 100 Thousand.

To start of I will identify my research questions which are included in my aim below. After this I will start the data preparation. Once I’ve collected the data needed I will present the data using suitable graphs, tables and other useful ways to represent my findings.

# Aim

In regards to my research questions the goal of this report is to help identify the comparison between two EU countries Ireland and Netherlands. I plan to compare these two countries using the following variables Income Per Person, Internet Use Rate, Alcohol Consumption, Life Expectancy and Suicide Per 100 Thousand. I will also compare the countries with an income rate that is between the income rate of Ireland and Netherlands to see how these other countries with similar incomes compare. I wanted to see if their was any correlation between these variables to see for example if there was a correlation between countries alcohol consumption and suicide rate.

# Preparation

## Data Selection

For data selection I decided to focus on some main areas which was Income Per Person, Internet Use Rate, Alcohol Consumption, Life Expectancy and Suicide Per 100 Thousand.

## Data Cleaning

Data cleaning is a very important part to any data project. A good data analyst should have the skills to prepare the data and limit the issues such as missing data so to do this I replaced empty strings with nan and set the variables to numeric.

# Analyse

## Focused Countries

Firstly I wanted to create the subsets for my focused countries, for this I separated Ireland and Netherlands and had a variable to target them separately. I also created a subset called ‘main\_countries’, which was used to target both Ireland and Netherlands at the same time. As shown in Figure 1 below.

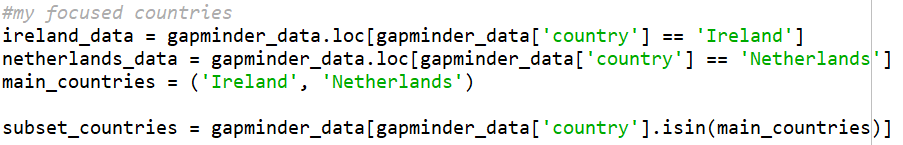


Figure 1: My Focused Countries

## Ireland and Netherlands Focused Variables

As stated in my aim above I wanted to focused on the following variables Income Per Person, Internet Use Rate, Alcohol Consumption, Life Expectancy and Suicide Per 100 Thousand. As shown below in Figure 2 I have displayed the outcome for these variables within Ireland and Netherlands.

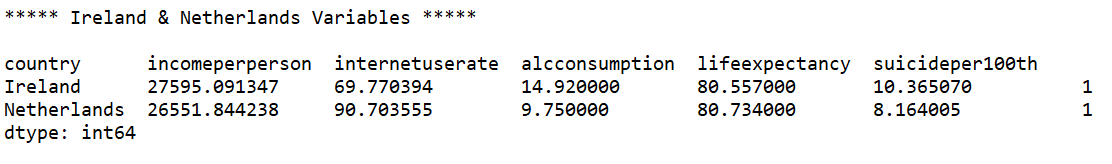


Figure 2: Ireland and Netherlands

As we can see from the output above, Ireland and Netherlands have very similar stats when it comes to Income, Life Expectancy and Suicide Per 100 thousand. But when it comes to Internet Use Rate and Alcohol Consumption we see a bit of a difference, we can see Netherlands has a higher Internet Use Rate but a lower Alcohol Consumption when compared to Ireland.

According to the World Health Organisations annual report into alcohol consumption Dutch people ranked near the bottom of the list in regards to drinking, however this still isn’t the lowest in Europe with Italy and Norway taking the credit for that. (Iamexpat.nl, 2020).

As we can see with the bar charts in Figure 3 below for Ireland and Netherlands Internet Use Rate and Alcohol Consumption to get a visual comparison of the differences between both countries in terms of these two variables as the Income Per Person, Life Expectancy and Suicide Per 100,000 is very similar.

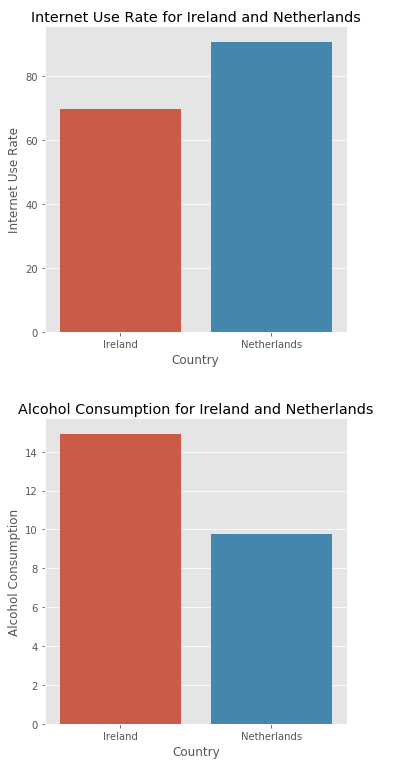


Figure 3: Internet Use Rate and Alcohol Consumption (Bar Charts)

### Ireland and Netherlands Descriptive Statistics

In order to compare the descriptive statistics between these two countries and all countries I decided to do a descriptive statistics for Ireland and Netherlands with my focused variables and compare it to the descriptive statistics for all countries with my focused variables. As shown in Figure 4 below you can see the Descriptive Statistics for all the countries and in Figure 5 we can see the Descriptive Statistics for Ireland and Netherlands combined.

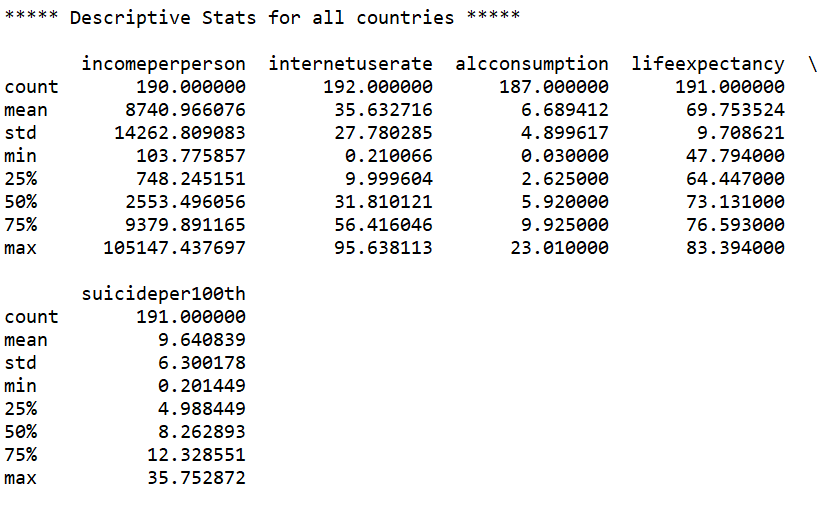


Figure : Descriptive Statistics for all countries

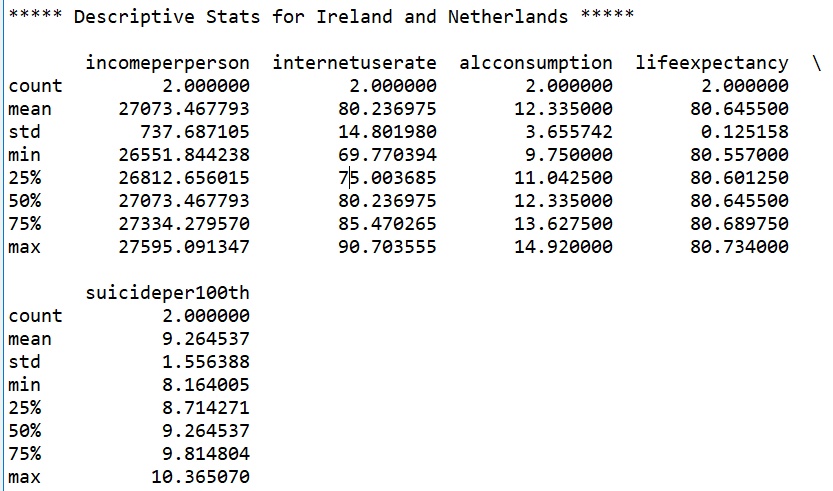


Figure : Descriptive Statistics for Ireland and Netherlands

**Income Per Person:** As we can see above for all countries the income per person mean is a lot lower than the mean of Ireland and Netherlands combined.

**Internet Use Rate:** For Internet Use Rate we can see that Netherlands has a high Internet Use Rate seeing as the max out of all the countries is 95.63 out of 100 while Netherlands has 90.70 out of 100 showing us Ireland as a considerably low Internet Use Rate.

**Alcohol Consumption:** Here Ireland has 14.92 while the max out of all the countries is 23 with Netherlands having a low Alcohol Consumption rate with only 9.75.

**Life Expectancy:** Here we see Ireland and Netherlands have almost the same life expectancy, while compared to the mean of all countries is only 69.75. In 2017 life expectancy at birth in the European Union (EU) was estimated to be 80.9 years old. (Ec.europa.eu, 2020)

**Suicide Per 100,000:** Here both Ireland and Netherlands have a similar Suicide Rate Per 100,000 but Netherlands has less at 8.1 compared to 10.3. While the mean with all the countries is 9.6 showing Netherlands is just a little bit below average.

## Countries with an Income between Ireland and Netherlands

As stated in my aim I also wanted to compare the countries with an Income that was between Ireland and Netherlands, in order to do this I had to do a query as shown in Figure 6 below.



Figure : Query to restrict observations

This was done to restrict the observations to be between 26550 and 27596.

I found that that the countries with an income between Ireland and Netherlands was Austria and Finland, another two Europe countries which I was expecting to see.

As we can see in Figure 7 below, we can see that of course all the countries have a similar income but we can see that Netherlands has the lowest alcohol consumption rate, highest internet use rate and the lowest suicide per 100,000.

We can see that the four countries have a similar Income Per Person and Life Expectancy so I won’t focus on these areas too much. The four countries Alcohol Consumption varies with Netherlands on the lower end at under 10. The Internet Use Rate for the four countries varies a lot with Ireland having the lowest out of these four countries and Netherlands having the highest. With Suicide Per 100,000 Finland has almost double the amount of suicides per 100,000 when compared to the Netherlands.

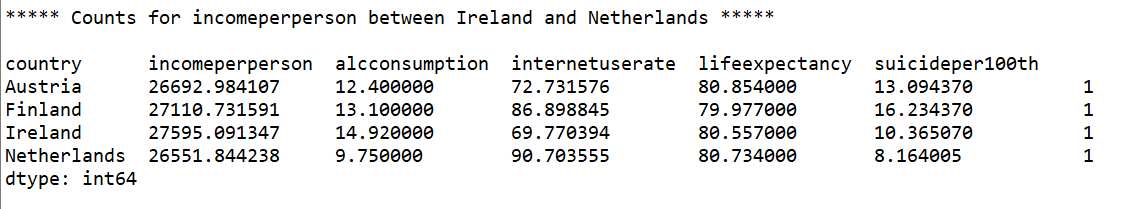


Figure : Countries with income between Ireland and Netherlands

As we can see below are the bar charts for Alcohol Consumption, Internet Use Rate and Suicide Per 100,000. I decided to not include the Income Per Person and Life Expectancy bar charts as they were all very similar. Below in Figure 8, 9 and 10 we can see a visual representation for these variables in these countries.

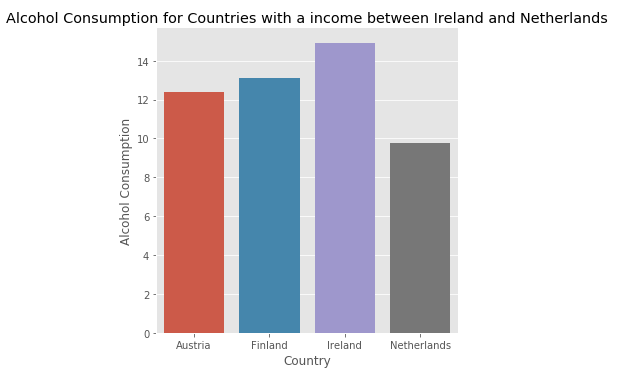


Figure 8: Alcohol Consumption Bar Chart

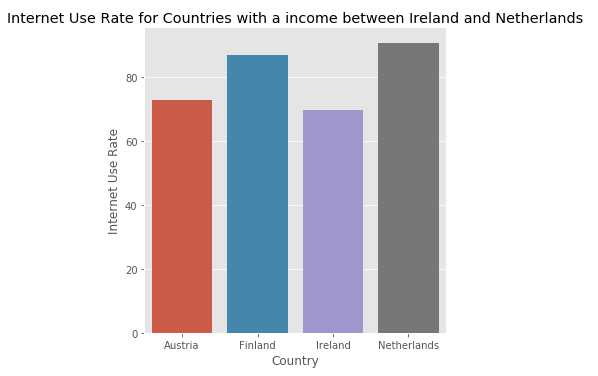


Figure 9: Internet Use Rate Bar Chart

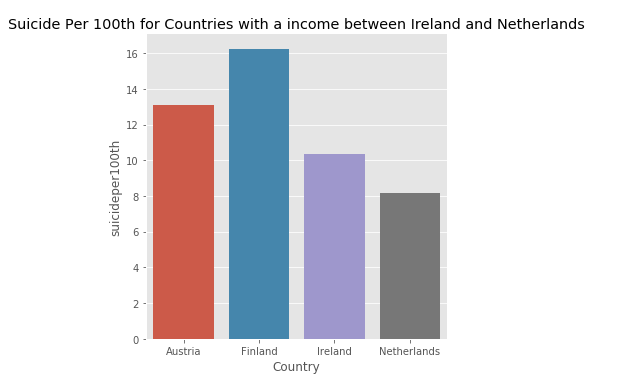


Figure 10: Suicide Per 100,000 Bar Chart

A key takeaway that I would take from this would be to see if there is any correlation between Alcohol Consumption and Suicide Rate as Netherlands as a lower Alcohol Consumption and Suicide Rate when compared to the other three countries. I will investigate below using the a scatterplot and all of the Gap Minder Data.

## Associations/Correlations

### Relationship between Alcohol Consumption and Suicide Rate

After my findings above I wanted to investigate if there was any correlation between Alcohol Consumption and Suicide Rate after my findings with Netherlands but in order to do this and to get a more accurate finding I had to target the entire Gap Minder file including all countries. As we can see below in Figure 7 there is a moderate correlation between Alcohol Consumption and Suicide Per 100,000.

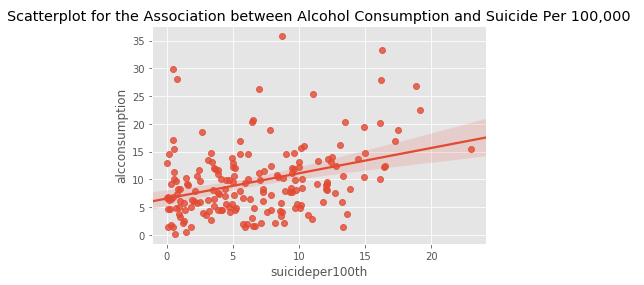


Figure : Alcohol Consumption and Suicide Rate Correlation

After doing some further research I this I found that people with alcoholism are up to 120 times more likely to commit suicide than those who are not dependent on alcohol and 29% of suicide victims in America were found with alcohol in their system. (Smith, 2020).

# Conclusion

I really enjoyed completing this assignment I did find it difficult at times with it all being done through python code but the more I worked through it the easier it became. I got to continue improving my data analytics skills from last semester. I tried to do some further research on any findings I found to back them up for example with the Alcohol Consumption and Suicide Per 100,000 having a correlation and with Netherlands having a low Alcohol Consumption.

In regards to my research questions, I feel like I met my research questions. I wanted to try identify the comparison between Ireland and Netherlands. I wanted to compare countries with a income between Ireland and Netherlands. I wanted to see if their was any correlation between variables such as alcohol consumption and suicide rate.

Some things that did not make it into my report from my python script:

* The temp.html file that gets created once the file is ran which includes Ireland and Netherlands with all the data for these two rows.
* Displaying Ireland and Netherlands and the focused variables individually.
* Mode and Median for the focused variables.
* Two new variables – Alcohol Consumption in number of pints (500ml) and Life Expectancy in seconds.
* Hypothesis Testing.

Overall I really enjoyed the process of this assignment even though some areas were a bit more difficult than others.

# References

Mullally, B. (2020). Tutors. [online] Tutors-design.netlify.com. Available at: https://tutors-design.netlify.com/talks/business-analytics-2-2020.netlify.com [Accessed 1 Mar. 2020].

Iamexpat.nl. (2020). Netherlands has one of the lowest rates of alcohol consumption in Europe. [online] Available at: https://www.iamexpat.nl/expat-info/dutch-expat-news/netherlands-has-one-lowest-rates-alcohol-consumption-europe [Accessed 2 Mar. 2020].

Ec.europa.eu. (2020). Life expectancy at birth in the EU: men vs. women. [online] Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190725-1 [Accessed 3 Mar. 2020].

Smith, C. (2020). The Link Between Alcohol Use and Suicide - Alcohol Rehab Guide. [online] Alcohol Rehab Guide. Available at: https://www.alcoholrehabguide.org/resources/dual-diagnosis/alcohol-and-suicide/ [Accessed 4 Mar. 2020].