CS306 Group Project – Step 1

Group Number 16

Berke Ceylan	27895
İsmail Çakmak	29496
Nisa Erdal	28943
Salih Kaya	27890
Yunus Topçu	28880

Can be accessed at https://github.com/Salihmaya/CS306_Group16

Tobacco Consumption: Causes and Effects

Introduction

The Project proposes to create a database of tobacco consumption related data and to obtain meaningful correlations between its causes and effects, of which will be filtered and visualised accordingly. Selection of different datasets are conducted with the purpose of finding relevant attributes, which are all ordered by location and time entries. In further steps of the project, countries and their given data will be filtered by commonly shared time periods in order to provide consistent analysis. Although it is yet to be decided, these findings and interpretations will be presented in a user-friendly website to showcase the end user experience of a Database Management System, which will be implemented via MySQL. Through the analysis of these interpretations, the project aims to distinguish statistically accurate causes of changes in tobacco consumption and to establish methods for reducing its negative effects.

Methods

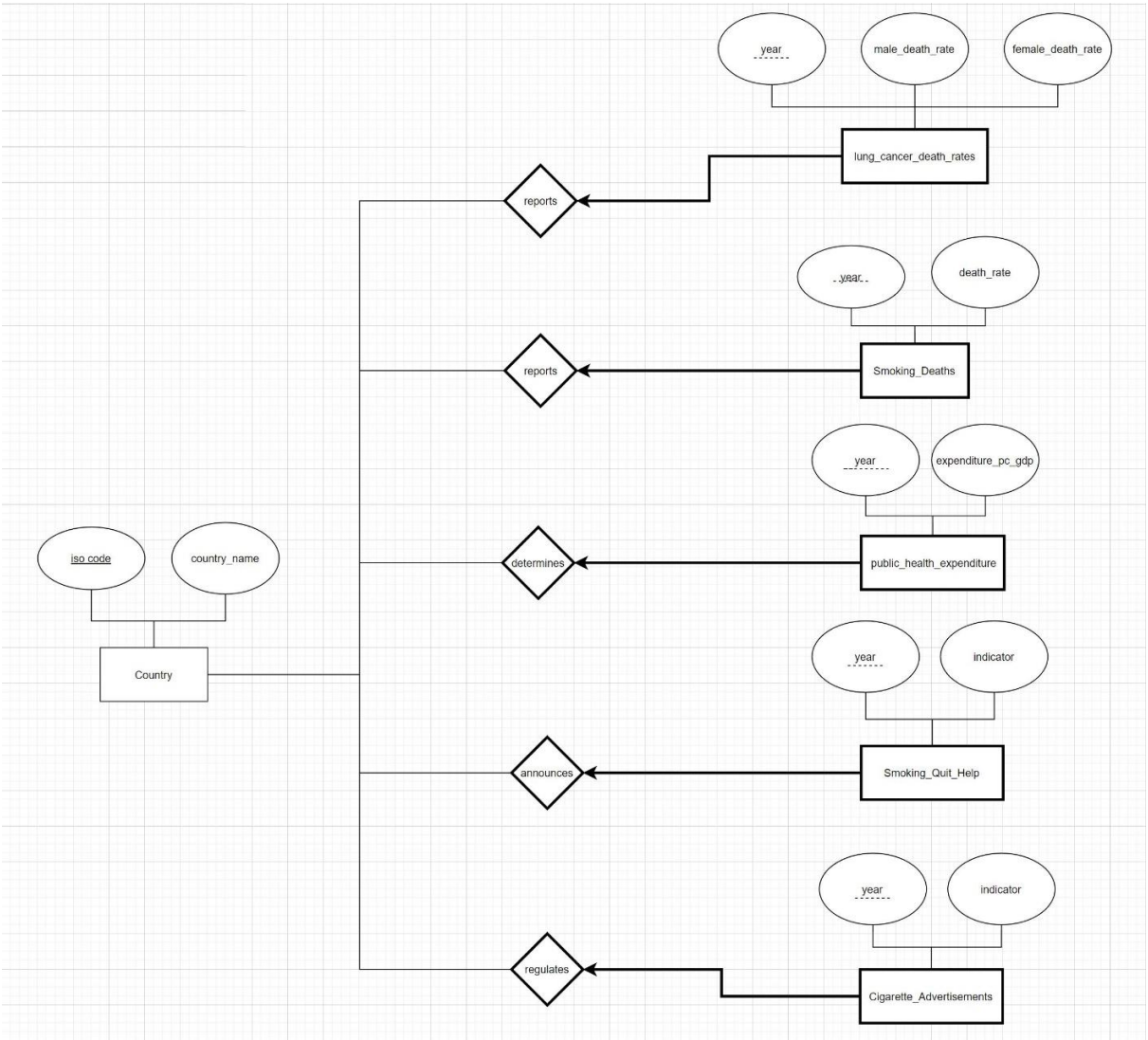
There are currently five different tables for each dataset in the project as depicted both in the ER diagram and as separate ".csv" files in the GitHub folder, besides the location dataset which contains country names and their respective ISO codes. ISO codes and years are all shared between the datasets, and they are named as the following:

- Lung cancer death rates
- Share of deaths that are attributed to smoking
- Government health expenditure as a share of GDP
- Support to help quit smoking
- Enforcement of bans on tobacco advertising

From the raw data provided by “ourworldindata.org”, each dataset was transferred from text to columns (to comply with the excel format) and was checked for duplicates. Since the datasets were selected individually, no separation of columns was conducted. Power Query was used to test and compare data points for deciding on whether the dataset in question would be relevant or not.

Results

Processed “.csv” files corresponding to the below ER Diagram can also be accessed from the GitHub folder.



References

Our World in Data. (n.d.). *Enforcement of bans on tobacco advertising*. Retrieved March 15, 2023, from https://ourworldindata.org/grapher/enforcement-of-bans-on-tobacco-advertising

Our World in Data. (n.d.). *Share of deaths from smoking*. Retrieved March 15, 2023, from https://ourworldindata.org/grapher/share-deaths-smoking

Our World in Data. (n.d.). *Lung cancer deaths per 100,000 by sex, 1950-2002*. Retrieved March 15, 2023, from https://ourworldindata.org/grapher/lung-cancer-deaths-per-100000-by-sex-1950-2002?country=ESP-1~ESP~ESP~USA-1~USA~USA

Our World in Data. (n.d.). *Support to help quit tobacco use*. Retrieved March 15, 2023, from https://ourworldindata.org/grapher/support-to-help-to-quit-tobacco-use

Our World in Data. (n.d.). *Public health expenditure (% of GDP).* Retrieved March 15, 2023, from https://ourworldindata.org/grapher/public-health-expenditure-share-gdp-owid