GitHub repository for the project: https://github.com/ihagverdi/Database-Systems-Project

Group 2:

Hagverdi Ibrahimli 30014
Amil Kazımoğlu 27891
Raman Afravi 30061
Ayçelen Kaptan 28826

Project title and description:

Our project title is “Smoking kills” and the main idea behind the project is to examine the relation between smoking and health. The data we shall utilize in this database application covers the following topics: Affordability of cigarettes, consumption per smoker per day, cigarette sales per day, share of cancer deaths due to tobacco, share of deaths due to smoking, share of lung cancer deaths. The real world problems which this database application addresses narrow down to uncovering the negative relationship between the smoking rate of the individuals and their health.

Updated CSV files:

For the second phase of the project, we spotted some problems in the previously edited CSV files from step one. For that reason, we further cleaned-up our CSV files and uploaded the edited files to our repository. Furthermore, due to a minor change in our ER diagram which we will explain after this section, we merged some CSV files to have a flawless conversion from our conceptual model to the relational one. Below is the summary of the changes made to the CSV files:

- We ran a detailed cleaning-up process on all the CSV files since there were errors in some rows and blank spaces throughout the datasets.
- We edited the two CSV files “share_deaths_smoking.csv” and “share_of_cancer_deaths_due_to_tobacco.csv” such that we combined their datasets together to make a new CSV file “deaths_due_to_tobacco_usage.csv”, the reason for which we will explain in the next section.

Updated ER (Entity relationship diagram):

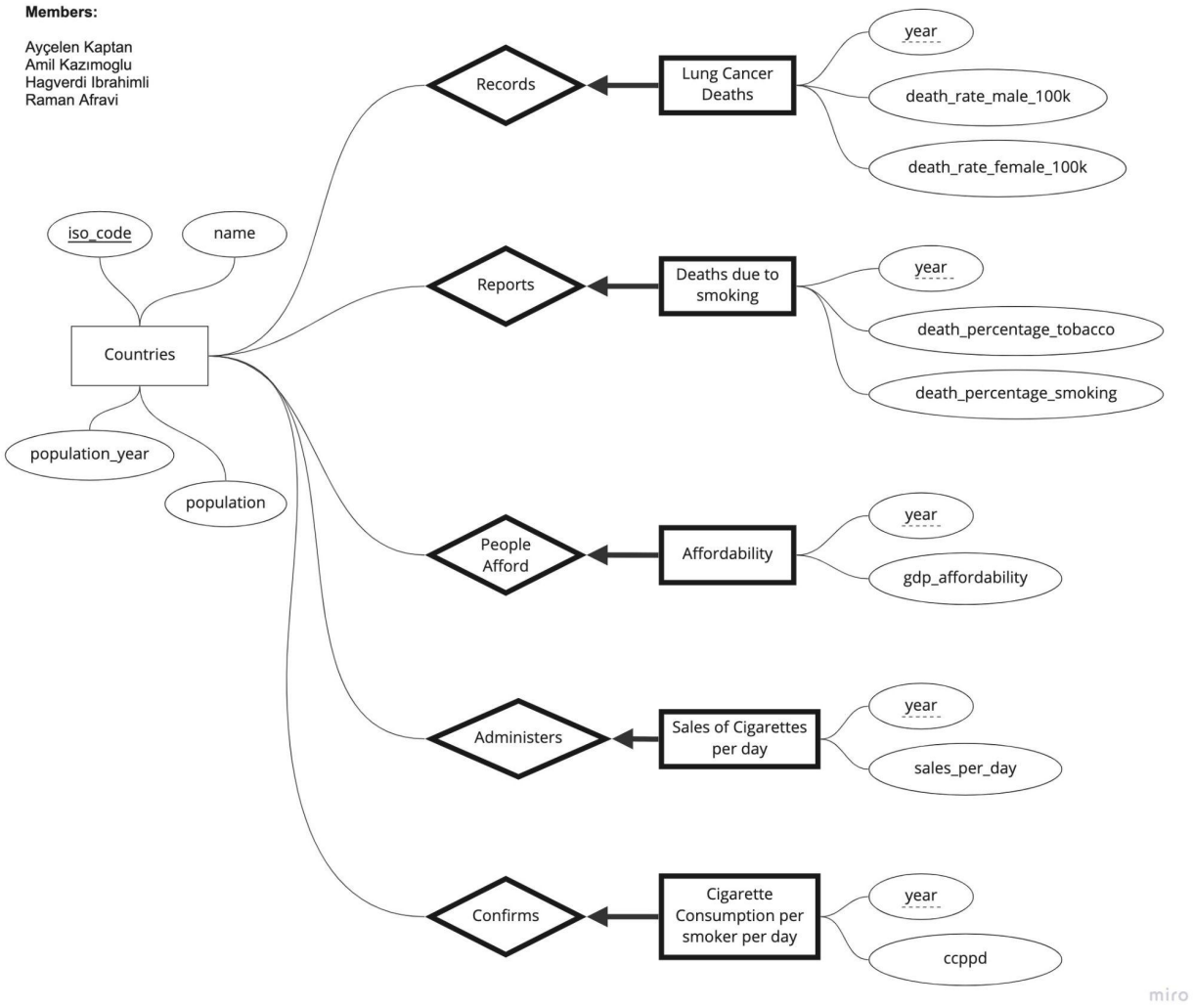

Figure 1. Updated ER diagram

The major revision in our ER diagram from step one of the project is the removal of the ISA hierarchy. The main reason behind this was that the “Lung Cancer Deaths” weak entity set did not meet the requirements of the ISA hierarchy since it contained information about the general deaths related to lung cancer, not necessarily related to smoking. For that reason, we separated the “Lung Cancer Deaths” weak entity set and made a new relationship “Records” with “Countries” entity set. Furthermore, we decided to merge the entity sets “Share of cancer deaths attributed to tobacco” and “Death due to smoking” into a general “Deaths due to smoking” entity set which contains their attributes combined. The idea behind this merge was that their datasets

contained the same range of date values for the similar type of information which is the percentage of deaths in this case.

SQL statements for creating the tables:

It is important to note that we have 1 entity set, 5 weak entity sets, and 5 relation sets as for our current ER diagram. As it is also evident in the SQL statements below, we do not need to create separate tables for weak entity sets as we only need to create the corresponding tables of the relation sets. However, since “Countries” is not a weak entity set, we create a separate table for it.

- “Countries” entity set:

```
create table countries(
iso_code Varchar(5),
population INT,
population_year INT,
country_name Varchar(100),
primary key(iso_code, population_year)
);
```

- The command corresponds to the following SQL file: “project smoking_countries.sql”-

- “People Afford” relation set:

```
create table affordability(
iso_code varchar(5),
gdp_affordability Float,
data_year INT,
primary key(iso_code,data_year),
foreign key(iso_code) references countries(iso_code) on delete cascade);
```

-The command corresponds to the following SQL file:”project smoking_affordability.sql”-

- “Administers” relation set:

```
create table cig_sales_daily(
iso_code varchar(5),
data_year INT,
sales_daily float,
primary key(iso_code,data_year),
```

foreign key(iso_code) references countries(iso_code) on delete cascade);
- The command corresponds to the following SQL file:
"project smoking_cig_sales_daily.sql"-

- "Confirms" relation set:
create table cig_consumption_daily(
iso_code varchar(5),
data_year INT,
consumption_daily float,
primary key(iso_code,data_year),
foreign key(iso_code) references countries(iso_code) on delete cascade);
- The command corresponds to the following SQL file:
"projectsmoking_cig_consumption_per_smoker_daily.sql"-

- "Reports" relation set (we merged two of our csv including tobacco_deaths and smoking_deaths files into one):
create table deaths_tobacco_usage(
iso_code varchar(5),
data_year INT,
smoking_deaths float,
tobacco_deaths float,
primary key(iso_code,data_year),
foreign key(iso_code) references countries(iso_code) on delete cascade);
- The command corresponds to the following SQL file:
"project smoking_deaths_tobacco_usage.sql"-

- "Records" relation set:
create table deaths_lung_cancer(
iso_code varchar(5),
data_year INT,
death_rate_male Float,
death_rate_female Float,
primary key(iso_code,data_year),
foreign key(iso_code) references countries(iso_code) on delete cascade);

- The command corresponds to the following SQL file:

  “project smoking_deaths_lung_cancer.sql”-