Faculty of Computer Science and Engineering - Skopje

Web based systems

Analysis of a dataset on mortality in Macedonia

Professors:

Dimitar Trajanovik
Milos Jovanovikj

Prepared by:

Magdalena Petrushevska

April 2023

Description of the data set

The data is taken from the World Health Organization (WHO) and focuses on the death rate in the Republic of Macedonia for the period of 2018-2021. The data set we will consider has a size of 912 x 10. The categories by which the data are distributed are by age group, year, gender and cause of death. There were 3 different categories of death cause and these are:

- Communicable, maternal, perinatal and nutritional conditions:
  - Infectious and parasitic diseases, Respiratory infections, Maternal conditions and Perinatal conditions
- Noncommunicable diseases
  - Malignant and other neoplasms, Diabetes mellitus and endocrine disorders, Neuropsychiatric conditions, Cardiovascular diseases, Respiratory diseases, Digestive diseases, Skin diseases, Musculoskeletal diseases
- Injuries

Deaths from infectious diseases are grouped together with deaths of women and newborns that were related to pregnancy and birth and deaths from insufficient nutrient intake.

Description overview

	RegionCode	RegionName	CountryCode	CountryName	Year	Sex	AgeGroupCode	AgeGroup	Causes	Number
0	EU	Europe	MKD	North Macedonia	2021	All	Age00	[0]	All	85
1	EU	Europe	MKD	North Macedonia	2021	All	Age01_04	[1-4]	All	14
2	EU	Europe	MKD	North Macedonia	2021	All	Age05_09	[5-9]	All	12
3	EU	Europe	MKD	North Macedonia	2021	All	Age10_14	[10-14]	All	12
4	EU	Europe	MKD	North Macedonia	2021	All	Age15_19	[15-19]	All	32
...	...	...	...	...	...	...	...	...	...	...
922	EU	Europe	MKD	North Macedonia	2018	Female	Age65_69	[65-69]	Injuries	12
923	EU	Europe	MKD	North Macedonia	2018	Female	Age70_74	[70-74]	Injuries	6
924	EU	Europe	MKD	North Macedonia	2018	Female	Age75_79	[75-79]	Injuries	12
925	EU	Europe	MKD	North Macedonia	2018	Female	Age80_84	[80-84]	Injuries	14
926	EU	Europe	MKD	North Macedonia	2018	Female	Age85_over	[85+]	Injuries	20

912 rows × 10 columns

Dataset origin

The data used and the entire code can be found on the following link.

In order to have bigger dataset to work wth, I aggregate data related to mortality in Macedonia for multiple years (2018, 2019, 2020, 2021). And also based on 3 different death causes ('Injuries', 'Communicable, maternal, perinatal and nutritional conditions', 'Noncommunicable diseases').

Dataset analysis

For the purpose of this project, first of all data preparation took place, in order to get rid of useless data and to get to work with clean data. Records containing ambiguous data like ‘Unknown age’ were removes from the dataset.

Later I did a check no missing values and also check duplicate records. Based on these check it was concluded that the dataset is free from missing values and duplicate records.

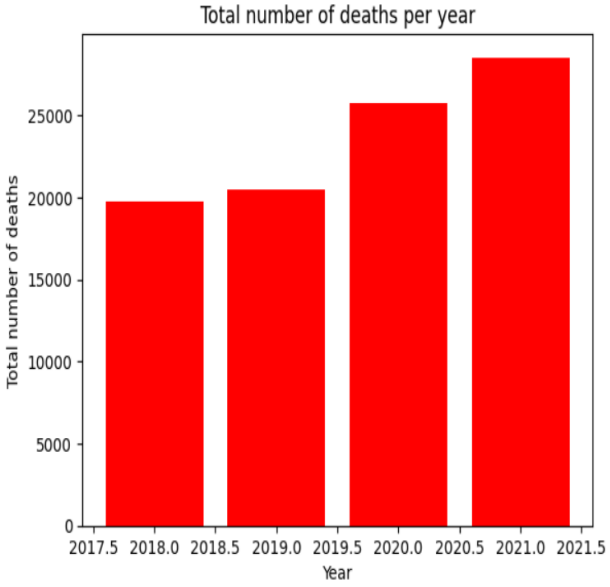
Further I got some continuous variable description for ‘Number’ column, which is the main star in these project containg the number of deaths per certain conditions/categories.

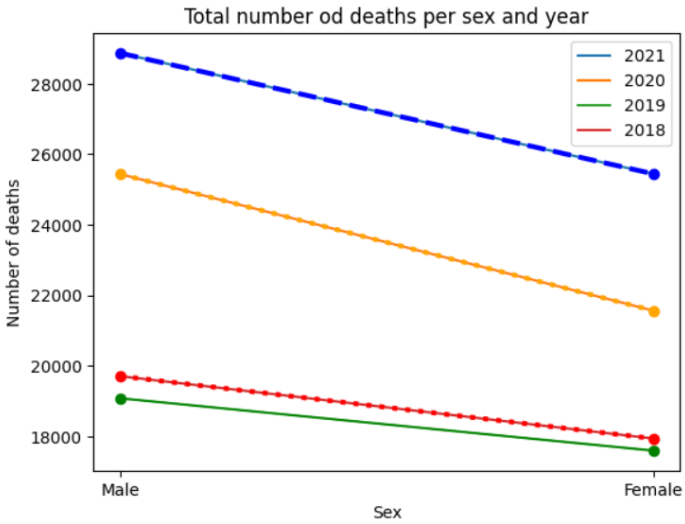
```
df.Number.describe()
```

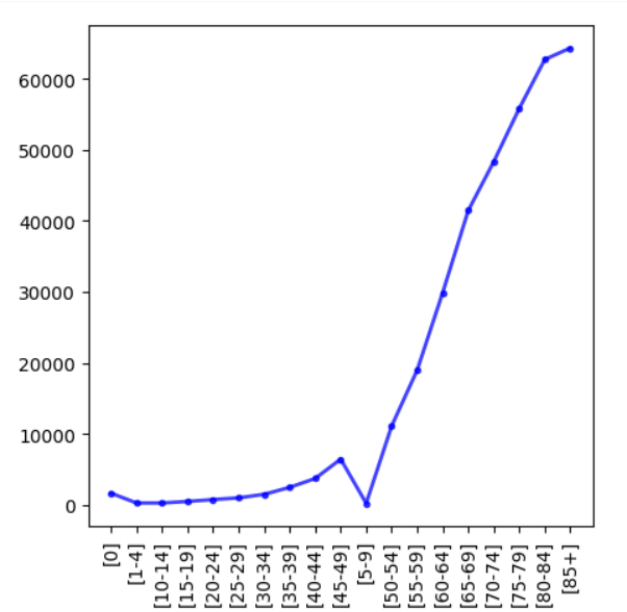
```
count     912.000000
mean      385.166667
std       812.449750
min         0.000000
25%         7.000000
50%        26.000000
75%       243.750000
max      5177.000000
```

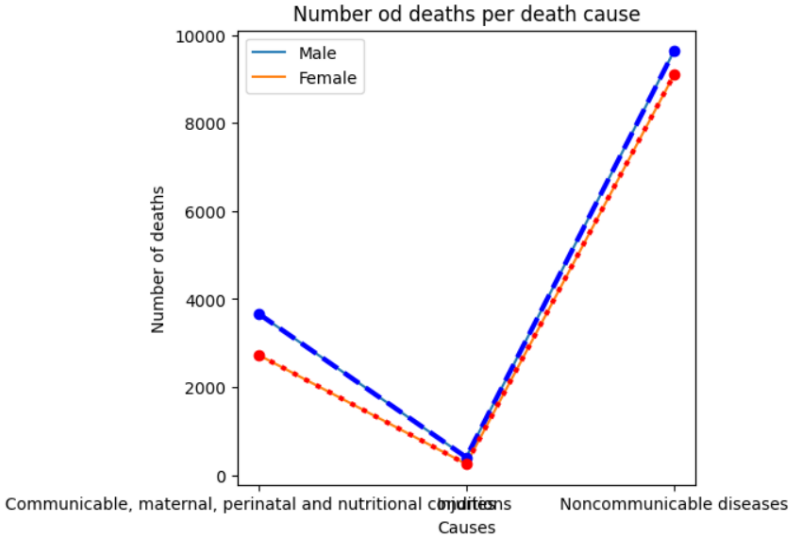
Also I did multiple dataset visual analysis, by using different types of charts. In general the overall conclusion was that every consequent year in a period of 2018-2021, an increasing number of deaths is reported. Also the largest number of deaths is in the age limit over 75 years. However, a significant number of deaths also began to be recorded from age of 50.
The biggest cause of deaths are noncommunicable deseases with percentange of 85%.
The number of death men and women does not differ much. Although statistically there are still more dead men than women.
It is also worth mentioning that the as older people get, the biggest probability is to died from noncommunicable type of deseases.

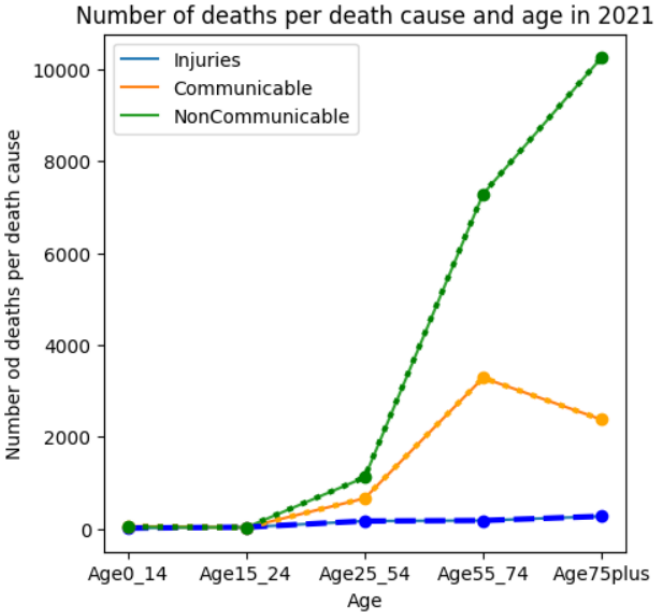
The details results from the analysis can be seen from the provided screenshot below.

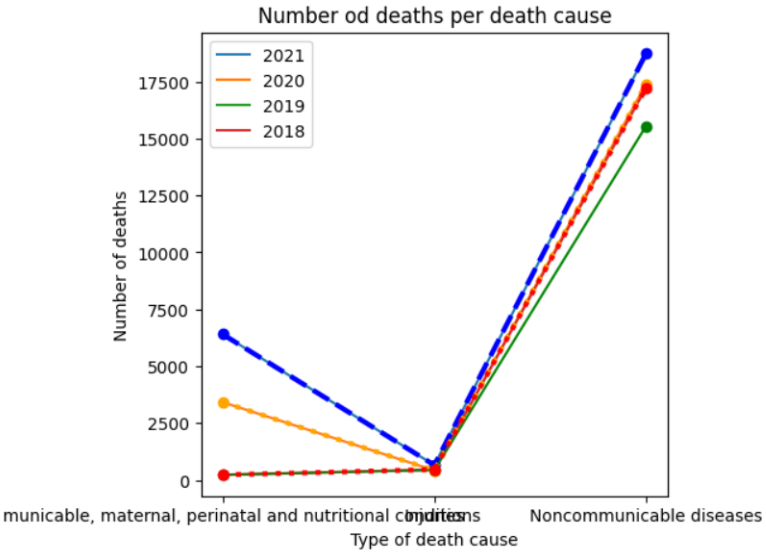


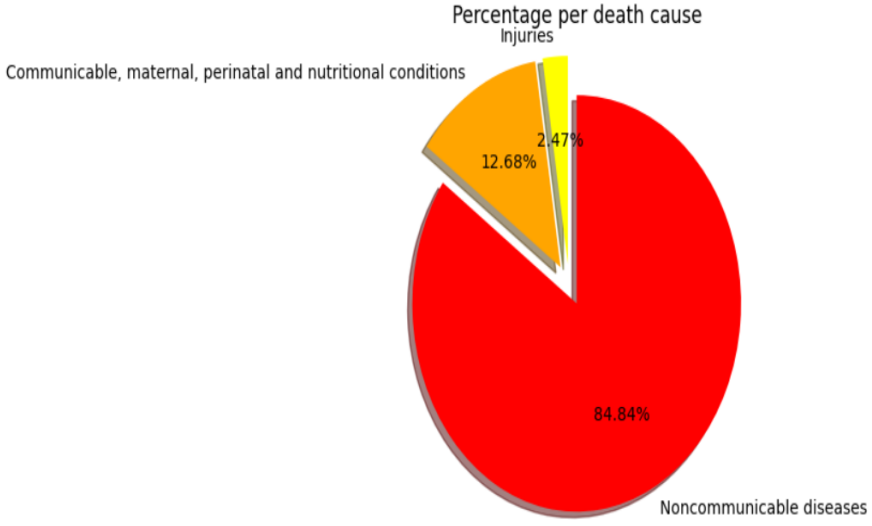



60000
50000
40000
30000
20000
10000
0
[0]
[1-4]
[10-14]
[15-19]
[20-24]
[25-29]
[30-34]
[35-39]
[40-44]
[45-49]
[5-9]
[50-54]
[55-59]
[60-64]
[65-69]
[70-74]
[75-79]
[80-84]
[85+]


Number od deaths per death cause
Male
Female
Number of deaths
10000
8000
6000
4000
2000
0
Communicable, maternal, perinatal and nutritional conditions
Injuries
Noncommunicable diseases
Causes


Number of deaths per death cause and age in 2021
Injuries
Communicable
NonCommunicable
Number od deaths per death cause
10000
8000
6000
4000
2000
0
Age0_14
Age15_24
Age25_54
Age55_74
Age75plus
Age


Number od deaths per death cause
2021
2020
2019
2018
Number of deaths
17500
15000
12500
10000
7500
5000
2500
0
municable, maternal, perinatal and nutritional conditions
Injuries
Noncommunicable diseases
Type of death cause


Percentage per death cause
Injuries
Communicable, maternal, perinatal and nutritional conditions
2.47%
12.68%
84.84%
Noncommunicable diseases