**Module 2.7 – PYTHON ASSIGNMENT**

**Chapter 1 – Requirements**

**Problem description:** Antimicrobial resistance (AMR) is a significant public health concern in the European Union (EU), as the levels of resistance to commonly used antibiotics continues to rise. The European Centre for Disease Prevention and Control (ECDC), located in Stockholm, monitors AMR through the European Antimicrobial Resistance Surveillance Network (EARS-Net), which collects data on an annual basis from as many as 37 centres across Europe.

*Klebsiella pneumoniae* is a Gram-negative lactose-fermeting bacterium, which is typically found as part of the normal bacterial population of the intestinal tract, skin and mouth. However, if aspirated into the lungs, it can have serious and destructive effects, including (but not limited to) pneumonia and other upper respiratory tract infections. Infection can also occur in hospital settings (nosocomial infections) and cause urinary tract and wound infections.

New antibiotic-resistant strains of *K. pneumoniae* have been appearing in recent years, giving rise to additional concerns as these strains are showing a (variable) degree of resistance to the, usually, very effective class of antibiotics called carbapenems. The use of this class of antibiotics is normally limited in cases of multidrug-resistant (MDR) bacterial infections, as a “last resort”. As such, until recently, carbapenems were thought to be “immune” to the observed trends of increasing antiobiotic resistance. However, recent data collected and published by the ECDC suggests that *K. pneumoniae* is gradually becoming resistant to carbapenems, generating serious concerns for public health, especially in those countries most affected.

The **objective of this Python program** is to create a line graph showing trends in the resistance, over time, of *K. pneumoniae* to carbapenem antibiotics in Bulgaria. This country is among the most seriously affected (in Europe) by growing antibiotic resistance, including that of *K. pneumoniae*.

**INPUT:** A data file (csv format) with comprehensive resistance data specific to Bulgaria over the period from 2005 to 2023, and obtained (freely and publicly) through the ECDC Surveillance Atlas of Infectious Diseases (<https://atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=4>). The data is expressed as percentage of *K. pneumoniae* isolates resistant to carbapenem antibiotics in Bulgaria, per year (see attached csv file).

This Python program is developed following the structured Waterfall programming model:

1. Problem definition and analysis
2. Design
3. Implementation
4. Testing and verification

**OUTPUT:** A linear graph showing the resistance trends over time in Bulgaria.

**Chapter 2 – Design**

The chosen method of representation for this design is **Pseudocode**.

IMPORT pandas library for data processing and manipulation.

IMPORT matplotlib library for data visualization and generate a line graph.

The Python program runs through the data set.

The data (in csv file format) includes the following key columns:

* **HealthTopic**: Topic of the data (Antimicrobial resistance).
* **Population**: Includes the bacterial species and the specific antibiotics used for the screening, in this case *Klebsiella pneumoniae* culturing in the presence of carbapenem antibiotics.
* **Indicator**: Specifies the data type, in this case the proportion of K. pneumoniae isolates resistant to carbapenem antibiotics.
* **Unit**: Unit of measurement, in this case percentage.
* **Time**: The year of data collection.
* **RegionCode** and **RegionName**: Identify the country.
* **NumValue**: Percentage of isolates resistant to carbapenem antibiotics.​​

Please note that ChatGPT was used as an aid to help identify relevant short scripts for the outputs requested in the assignment. Prompts were inserted in ChatGPT, followed by follow-up prompts, then the proposed scripts were tested directly in Visual Studio Code and the output checked. Development of the full script followed a trial-and-error approach, requiring multiple iterations until a working script was cerated.

This script is designed to draw a line graph showing the percentage of *K. pneumoniae* isolates exhibiting resistance to carbapenem antibiotics over the years from 2005 to 2023, specifically in isolates collected in Bulgaria.

To further analyze the data, the following functions were included in the script:

* **Array** to store and highlight the years where the percentage exceeds the set threshold (30%).
* **IF function**: if the value exceeds the set threshold, the relevant years are highlighted by a vertical red, segmented line.
* **Function to calculate the year-to-year change in values** and then find the year with the highest change. This is followed by a print function to print the results of year with highest change.
* **IF-ELSE function**: if the percentage value exceeds 30%, the relevant data point is marked by a yellow “X”. If the value does not exceed 30%, the relevant data point is marked by a blue “o”.
* The plot is customized to include a background grid and a tight layout, with integers on the x axis (Year) and Percentage on the y axis.
* The resulting plot/figure is saved as “resistance\_trends.png”

**Chapter 3 – Implementation**

See attached python script in text file, output and line graph below:

CSV file loaded successfully.

Line graph created.

**Yearly Change (percentage):**

Time NumValue YearlyChange

0 2005 0.000000 NaN

1 2006 0.000000 0.000000

2 2007 0.000000 0.000000

3 2008 0.000000 0.000000

4 2009 0.000000 0.000000

5 2010 0.000000 0.000000

6 2011 0.000000 0.000000

7 2012 1.851852 1.851852

8 2013 0.000000 -1.851852

9 2014 7.194245 7.194245

10 2015 3.157895 -4.036350

11 2016 4.402516 1.244621

12 2017 12.426035 8.023520

13 2018 21.243523 8.817488

14 2019 26.966292 5.722769

15 2020 28.112450 1.146158

**16 2021 46.280992 18.168542**

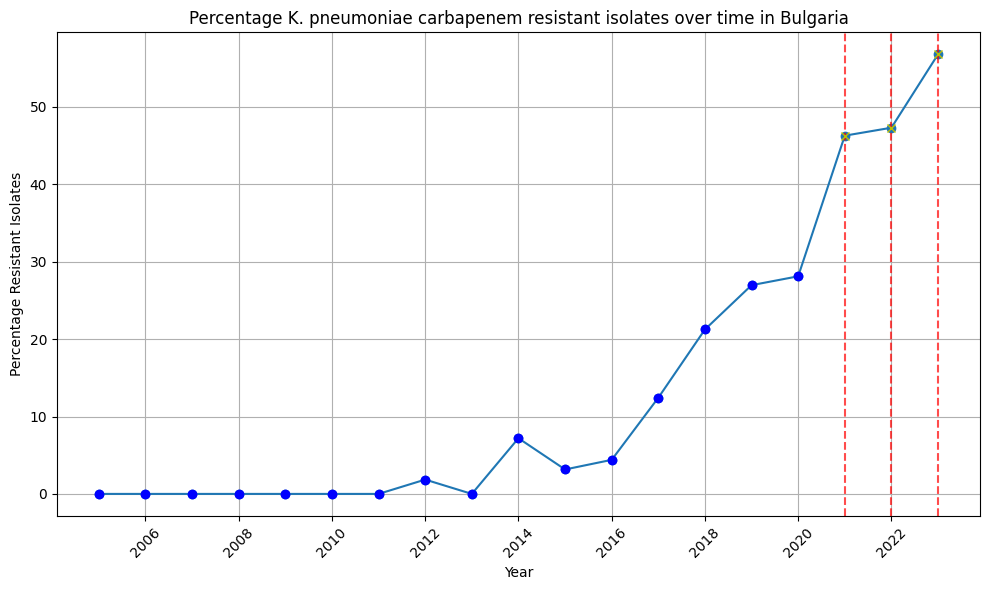
17 2022 47.307692 1.026701

18 2023 56.790123 9.482431

**Year with the highest change in percentage resistant isolates:**

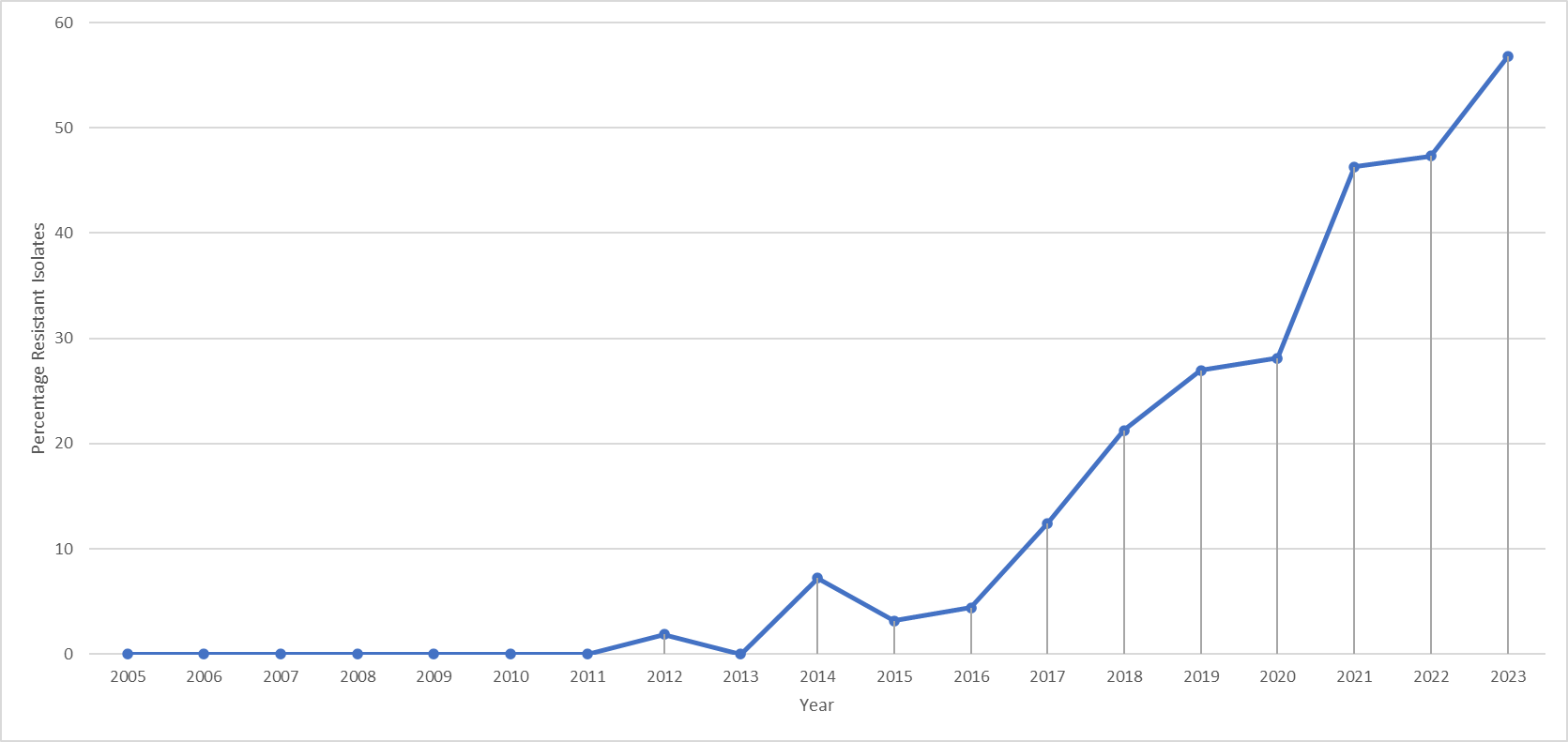
Year: 2021

Change: 18.17



**Chapter 4 – Validation**

The output of the Python script, in particular the line graph, has been validated by reviewing the output graph and checking the axis labels and titles, as well as by visually inspecting the graph to make sure that the data is represented correctly. In addition, an Excel line graph has been created with the same data (here below), demonstrating that the output graph from Python is indeed correctly representing the data. A simple calculation in Excel also demonstrates that the calculation of change in percentage of resistant isolates year-on-year is also correct (please see below).



EXCEL validation :



**Chapter 5 – REFERENCES**

European Centre for Disease Prevention and Control (ECDC)- Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report 2023 – 18 Nov 2024 (<https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-eueea-ears-net-annual-epidemiological-report-2023#:~:text=The%20estimated%20total%20EU%20incidence%20of%20carbapenem%2Dresistant%20Klebsiella%20pneumoniae,2.39%20per%20100%20000%20population>)

OpenAI. (2025). ChatGPT (Version 4) [Large language model]. https://chat.openai.com

* END OF THE DOCUMENT -