## The Discovery of Handwashing – A Python Recreation

**Description**

The classical saying 'waste not, want not' has never been more relevant in the current world situation complete with the heightened global consciousness and concern for the environment as matter of the day. By his finding the evolution of the sanitary regulations in medical care started and still used today. Through the process of this project, we are in quest for copies of Semmelweis's discovery where Python programming being the major language in data analysis and visualization will be used.

The project relies on the historical data that reflects various groups within a hospital setting, this groups having ranges in mortality rates during different periods. Python will be used to thoroughly go through the data with a fine-toothed comb so we can identify patterns and trends. This is going to be of great importance even as we establish the interactions among the variables and unravel the impact of hand washing on death rate. Thanks to Python statistics to data manipulation and analysis of , we will conduct calculations like prior and after the implementation of hand washing protocols.

1. Data Loading and Exploration: The first thing to do is load that historical dataset into your Python environment and then proceed with the preceding data analysis to understand the structure and content of the database.
2. Relationship Analysis: By utilizing Python's data analysis tools we will go even deeper and try to find out how the time period, deaths rate, and handwashing techniques are connected to each other.
3. Calculations and Comparisons: With regard to the implementation of the Python numerical computing element of the study, mortality rates will be computed before and after implementing hand washing techniques. The aim of this stage is to measure how the hand washing campaign affect on reducing disease death rates.
4. Visualization: A crucial tool for the delivery of insights through visual representation of data is thus to be highlighted. There will be Python visualization libraries to be employed in order to generate graphs and charts that will be easy to interpret and will effectively show the effect of hand washing on the mortality rates over time.

**Analysis of the relationship between variables over different periods**

This project is intended to make an analysis of some factors which are time-related in order to enable us to comprehend the role of hand washing process in the death rates. Through it, I explored the professional career of Dr. Ignaz Semmelweis. For this, we will first begin the collection of some data which includes the mortality rates separated by the time periods of a hospital.

Our work will also include assessment of common elements, like if they were targeted at specific periods of time or health understanding, hand washing or nursing training. Python is a data analysis tool that we will be using to identify the patterns in the given data. The research will be done to check if there will be major changes in mortality indexes shortly after the handwashing protocol imposition.

To do a numerical assessment of the effectiveness of the hand washing campaign, we will carry out mathematical calculations of the shift in mortality rate after the introduction of hand hygiene measures. This stage is going to be dedicated to the use of statistical analysis to determine any tangible differences in mortality rate between the periods when hand washing was provided and the period when hand washing was not provided.

In addition to visualizing our findings, the visual form of representation will also be highly appropriate to our presentations. We maan will display heat maps while there are relationships between the variables and period that are generated through Python's visualization libraries. These visual graphics including charts will be introduced to present in a clear format what steps people can take to help to decrease global mortality rate through hand washing.

The purpose of this research is to reveal the far-reaching effect of personal hygiene on health. Additionally, it draws a parallel between the modern world and the past when data monitoring, analysis, and statistics played a vital role in discovering medical breakthroughs.

**Now Let’s Look at the Code Snippets**

**First We will import all the necessary module libraries :**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

Now to read the yearly dataset that is provided to us

yearly\_df = pd.read\_csv("yearly\_deaths\_by\_clinic.csv")

yearly\_df

yearly\_df.shape

This code provides us with the details of the dataset for example number of columns and number of rows present :

yearly\_df.info()

This gives all the information of the dataset such as datatype of rows and , NULL or NON NULL

yearly\_df.groupby("clinic") ["deaths"].sum()

Grouping the data by the "clinic" column, and then calculates the sum of the "deaths" column for each group.

yearly\_df["Proportion of Deaths"] = yearly\_df["deaths"] / yearly\_df["births"]

yearly\_df

calculating the proportion of deaths for each entry and adding the a new column in the dataframe with a name of Proportions of death .

clinic\_1 = yearly\_df[yearly\_df["clinic"] == "clinic 1"]

clinic\_2 = yearly\_df[yearly\_df["clinic"] == "clinic 2"]

Seperating the database for clinic1 and clinic 2

fig,ax = plt.subplots(figsize = (10,4))

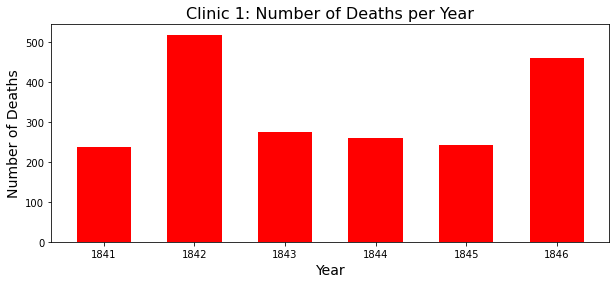
plt.bar(clinic\_1.year, clinic\_1.deaths, width= 0.6, color= "red")

plt.title("Clinic 1: Number of Deaths per Year", fontsize=16)

plt.xlabel("Year", fontsize=14)

plt.ylabel("Number of Deaths", fontsize=14)

Displaying the number of deaths per year for clinic1 using matplotlib



fig,ax = plt.subplots(figsize = (10,4))

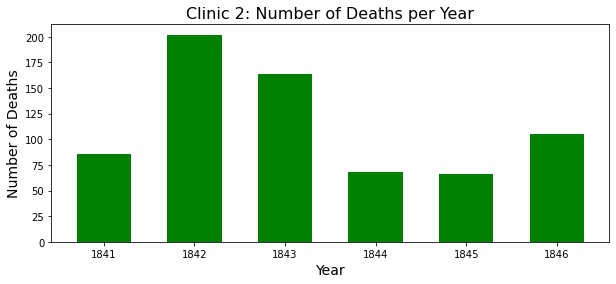
plt.bar(clinic\_2.year, clinic\_2.deaths, width= 0.6, color= "green")

plt.title("Clinic 2: Number of Deaths per Year", fontsize=16)

plt.xlabel("Year", fontsize=14)

plt.ylabel("Number of Deaths", fontsize=14)

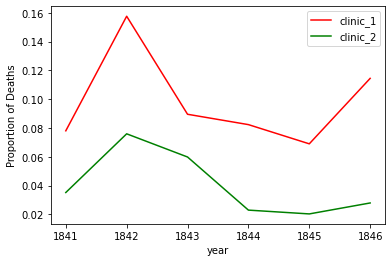
Displaying the number of deaths per year for clinic2 using matplotlib



ax= clinic\_1.plot(x= "year", y= "Proportion of Deaths", label= "clinic\_1", color="red")

clinic\_2.plot(x= "year", y= "Proportion of Deaths", label= "clinic\_2", ax=ax, ylabel= "Proportion of Deaths", color="green")

Creating a line plot using pandas' built-in plotting functionality, with the proportions of deaths plotted over the years for Clinic 1 and Clinic 2.



monthly\_df = pd.read\_csv("monthly\_deaths.csv")

monthly\_df.head(5)

Reading the Monthly Dataset

monthly\_df.info()

Showing the Information of the dataset

monthly\_df["Proportion of Deaths"]= monthly\_df["deaths"] / monthly\_df["births"]

monthly\_df.head(5)

calculating the proportion of deaths for each entry and adding the a new column in the dataframe with a name of Proportions of death .

monthly\_df.dtypes

monthly\_df['date'] =  pd.to\_datetime(monthly\_df['date'])

converts the "date" column to datetime format using the pd.to\_datetime() function. This function converts the "date" column from its current data type

start\_handwashing = pd.to\_datetime('1847-06-01')

before\_washing = monthly\_df[monthly\_df["date"] < start\_handwashing]

after\_washing = monthly\_df[monthly\_df["date"] >= start\_handwashing]

split the DataFrame monthly\_df into two separate DataFrames based on a condition related to the "date" column.

fig,ax = plt.subplots(figsize = (10,4))

x= before\_washing["date"]

y= before\_washing["Proportion of Deaths"]

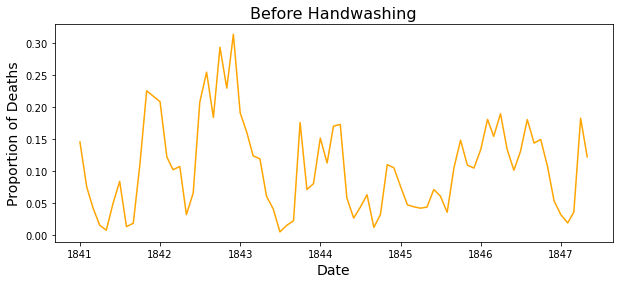
plt.plot(x, y, color= "orange")

plt.title("Before Handwashing", fontsize=16)

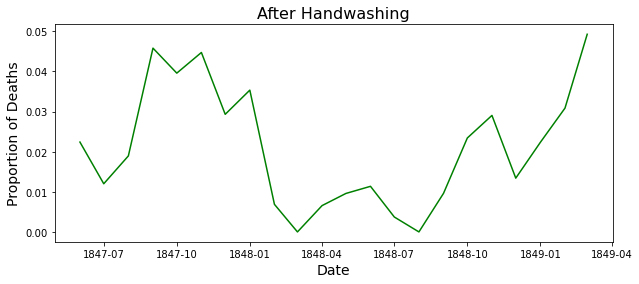
plt.xlabel("Date", fontsize=14)

plt.ylabel("Proportion of Deaths", fontsize=14)

creates a line plot using matplotlib, showing the proportion of deaths over time before the implementation of handwashing.



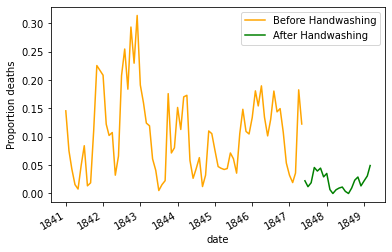
Creating a line plot using matplotlib, showing the proportion of deaths over time after the implementation of handwashing.



ax= before\_washing.plot(x= "date", y= "Proportion of Deaths", label= "Before Handwashing", color="orange")

after\_washing.plot(x= "date", y= "Proportion of Deaths", label= "After Handwashing", ax=ax, ylabel= "Proportion deaths", color="green")

creating a line plot, displaying the proportion of deaths over time before and after the implementation of handwashing on the same graph.



before\_proportion = before\_washing["Proportion of Deaths"]

after\_proportion = after\_washing["Proportion of Deaths"]

before\_proportion.mean()

calculating exactly how much did handwashing decreased the proportion of deaths on average.

after\_proportion.mean()

mean\_diff = after\_proportion.mean() - before\_proportion.mean()

mean\_diff

O/p : -0.0839566075118334

Calculating the difference between both proportions

And here minus sign in the output indicates that there is a disease .