IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

Life expectancy analysis with Python

Ch. Deepika[1], G. Chandini[2], Ch. Rajeswari[3], V. Devi Padmavathi[4,] M. Divyanjali5, P. Tejaswini[6]

[1] Assistant Professor, Dept. of Computer Science & Engineering, Vijaya Institute of Technology for Women, Enikepadu, Vijayawada-521108

[2,3,4,5,6] Students, Dept. of Computer Science & Engineering, Vijaya Institute of Technology for Women, Enikepadu, Vijayawada-521108

Email id: shaikrehmathunnisa@gmail.com[1], chandinigorakala3@gmail.com[2] , challapalliraji000@gmail.com[3] , vallabhanenipadhu@gmail.com[4] , divyanjaligunti@gmail.com[5] , tejaswinipappala@gmail.com[6]

Abstract:

Life expectancy analysis is a crucial aspect of public health and demographic research, providing insights into population health trends, socio- economic development, and policy effectiveness. The abstract presents a comprehensive overview of life expectancy analysis methodologies using python programming language. The analysis begins with data acquisition, covering various publicly available datasets such as World Bank indicators, WHO mortality data, and national statistical data bases. Python libraries like Pandas, NumPy, and Requests facilitate data retrieval, preprocessing and cleaning. Following data acquisition, exploratory data analysis (EDA) techniques are applied to understand the distributions, trends and relationships within the dataset. Visualization libraries such as Matplotlib, Seaborn, and Plotly are employed to create insightful plots and charts, aiding in identifying patterns and anomalies

Keywords: Python programming, Data acquisition, Exploratory data analysis (EDA), Data preprocessing. Regression analysis

Introduction

In the fields of public health, demography, and social policy, life expectancy study is an essential component. Research on the factors that affect a population's average lifespan can shed light on health outcomes, socioeconomic inequities, and the efficacy of measures to improve general well-being. In recent years, the use of Python programming language has gained prominence in life expectancy analysis due to its versatility, robustness, and extensive libraries tailored for data analysis and machine learning tasks. Python offers a comprehensive ecosystem for acquiring, preprocessing, analyzing, and visualizing data, making it well- suited for conducting in – depth investigations into life expectancy trends and determinants This introduction serves as an overview of the methodologies, techniques, and objectives involved in life expectancy analysis using Python. It outlines the significance of understanding life expectancy patterns, the role of python in facilitating such analyses, and the goals of this document in providing a comprehensive exploration of life expectancy dynamics through a data - driven approach. Statistical modeling to identify important determinants, data acquisition from diverse sources, exploratory data analysis to uncover underlying patterns, and machine learning techniques for predictive modeling are just a few of the topics covered in depth throughout this document. Also covered will be more advanced subjects like survival analysis and casual inference, as well as spatial analytic tools for understanding geographical variations in life expectancy.

Literature Review:

This section will examine previous research and detail the authors' methods and inferences drawn from their studies using various data visualization tools. Patients with just one or two hospitalizations in the past five years can be included in a predictive tool developed by Aggarwal D. et al. (2017). Recent years have seen this model emerge as one of the most reliable forecasting tools, with findings going back as far as five years. Their findings were supported by a combination of various machine learning methods. The final prediction model is the result of combining multiple basic models using the ensemble methodology, which is a machine learning method. Kyle J. Foreman et al. (2018). In which they analyzed 195 nations from 1990 to 2016 for GBD, risk variables, and injury prevalence in order to make predictions. It was astounding to think about all those health-related elements, and it was even more impressive to see how each attribute was projected against the health scenario. They accounted for the estimate by considering 79 health factors. Published in BMC Medical Informatics, this paper was authored by Beeksma et al. (2019). It is a relatively new researcher in the field that is relevant to the topic at hand. Using the medical information of people who have passed away, they suggested training supervised machine learning models using recurrent neural networks. They used supervised machine learning to tackle the problem. Long Short-Term Memory (LSTM) recurrent neural networks were subsequently trained and evaluated using the data. Long Short-Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN). In their study, N. Kerdprasop et al. (2017) proposed that environmental and economic factors are associated with life expectancy. They resorted to regression on numerical and continuous values and the Chi Square Automatic Interaction Detector (CHAID) approach for categorical values. The CHAID algorithm is based on the principle of a decision tree. You can use it to find out how the variables are related to each other. They used this method in an effort to establish a correlation between a country's GDP and its citizens' longevity. In their 2016 work, Yang et al. provided a model that, based on a set of essential assumptions about future trends, could make basic predictions about how long people in the Netherlands would live based on their socioeconomic status and sex. In order to get an idea, they used the old-fashioned Li-Lee model. Many authors still use this Li-Lee model, which has been around since 1992, to predict death rates in the future. According to a 2017 study on regression techniques for stock prediction by Ashish Sharma et al., the majority of regression analyses are employed for predicting stock market trends. Adding additional numerical factors could lead to better results in the future. Using a variety of applications and automation techniques, the authors of this 2019 study by Andrea Picasso et al. merged economic and elemental analysis to forecast market trends. Those are charts containing forecasting data, and neural networks are a machine learning technology that can solve the trend stock problem. The article's tone is used as an input variable. The utilization of information regarding news astral one-off was determined to be the most troublesome accomplishment based on their research. In their 2018 paper, Gangadhar Shobha et al. offered a comprehensive overview of machine learning techniques. The author covered three types of techniques and various metrics, including recall, RMSE, accuracy, confusion matrix, precision, and quintile of errors. The reader will find the discussion of these metrics useful when applying the concepts and equations presented in the paper. Since many individuals are unsure of how to apply various machine learning techniques for prediction and other purposes, the author of this review hopes that it will be useful to those who are new to machine learning.

EXISTING SYSTEM

Existing systems for life expectancy analysis with Python typically involve leveraging various libraries and frameworks for data manipulation, statistical analysis, machine learning, and visualization. An example of an existing system for life expectancy analysis with Python could be a web application where users can input demographic and healthcare indicators, and the system provides predictions of life expectancy based on a trained machine learning model. This application could be built using Flask for the backend, HTML/CSS/JavaScript for the frontend, and Scikit-learn for the machine learning model

PROPOSED SYSTEM

The proposed system for life expectancy analysis is designed to leverage Python- based tools and technologies to provide accurate predictions of life expectancy based on demographic and healthcare indicators. It consists of several key components: The system gathers data from various sources, including public health databases, surveys, and research studies. Python libraries like Pandas are used for data preprocessing, which involves cleaning the data, handling missing values, and transforming features to prepare them for analysis.

Methodology:

In the materials and methods section of a life expectancy analysis, various statistical and machine learning methods are employed to analyse data and develop predictive models. Here are common methods used in life expectancy analysis:

- Descriptive Statistics: Descriptive statistics are used to summarize and describe the characteristics of the data. This includes measures such as mean, median, mode, standard deviation, range, and percentiles.
- Correlation Analysis: Correlation analysis is used to identify relationships between variables. This involves calculating correlation coefficients (e.g., Pearson correlation coefficient, Spearman rank correlation coefficient) to measure the strength and direction of associations between variables.
- Regression Analysis: Regression analysis is a statistical method used to examine the relationship between one or more independent variables (predictors) and a dependent variable (outcome). In life expectancy analysis, regression models are used to predict life expectancy based on predictor variables.
- Survival Analysis: Survival analysis is used to analyse time-to-event data, such as mortality rates or survival times. This method accounts for censoring and investigates factors that influence the probability of experiencing an event over time

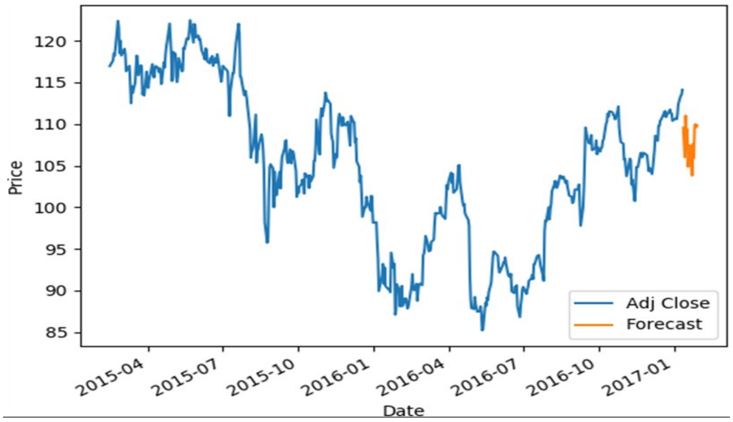
RESULTS AND DISCUSSION

The proposed model was tested and compared with four other stand- ard algorithms, including KNN, Naïve Bayes, OneR and ZeroR. The test exam- ined how accurate the tested algorithms predict the stock price trends, and evalu- ated the MAE and RMSE. Table 5 presents the test results. The hybrid KNN- Probabilistic model has allowed us to achieve an estimated accuracy of 89.1725%, exceeding the stand alone KNN reported accuracy of86.6667% and the Naive Bayes accuracy of 76.1194%. The accuracy rates for OneR and ZeroR classifiers were 71.6418% and 64.1791% respectively. KNN-Probabilistic model has MAE rate of 0.0667% and RMSE rate of 0.2582% which are much lower than the other classifiers.

Table: Prediction Results of Classifiers.

Classifier	Accuracy (%)	MAE	RMSE
KNN-Probabilistic	93.3333	0.0667	0.2582
KNN	86.6667	0.1333	0.3651
Naive Bayes	76.1194	0.1726	0.2824
One R	71.6418	0.5325	0.6139
ZeroR	64.1791	0.4619	0.4805

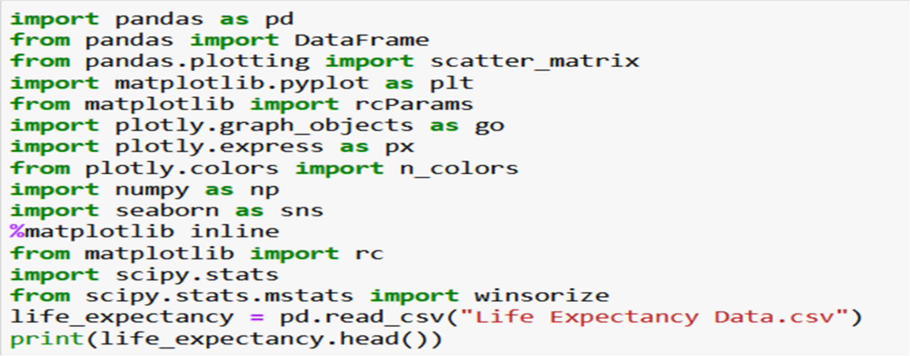
Overall, KNN-Probabilistic model has better accuracy rate and error rates than the other classifiers used for comparisons. The test demonstrated that the hybrid mechanism of KNN and probabilistic method produced significantly improved results, compared with each of the KNN and Naïve Bayes classifiers



In the discussion section of a life expectancy analysis, you delve deeper into the interpretation of your findings, provide context for your results, discuss their implications, and address any limitations of your study. By structuring your discussion in this way, you can provide a comprehensive analy- sis of your life expectancy study, contextualize your findings within the broader literature, and offer insights that contribute to public health discourse and policy development.

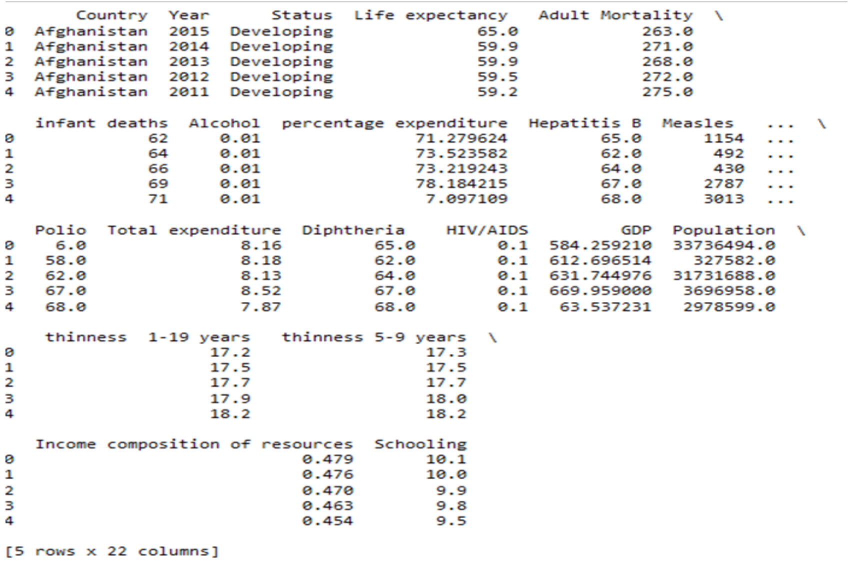
IMPLEMENTATION

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external. Our data mainly consists of the previous year stock prices. Initially, we will be analysing the live dataset and according to the accuracy, we will be using the model with the data to analyse the predictions accurately



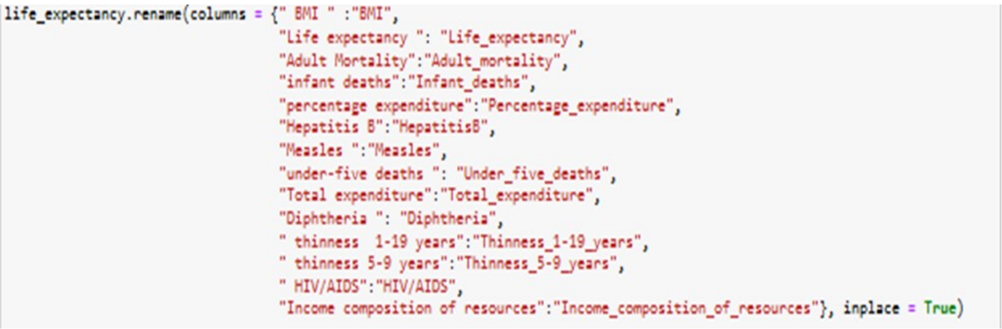
```
import pandas as pd
from pandas import DataFrame
from pandas.plotting import scatter_matrix
import matplotlib.pyplot as plt
from matplotlib import rcParams
import plotly.graph_objects as go
import plotly.express as px
from plotly.colors import n_colors
import numpy as np
import seaborn as sns
%matplotlib inline
from matplotlib import rc
import scipy.stats
from scipy.stats.mstats import winsorize
life_expectancy = pd.read_csv("Life Expectancy Data.csv")
print(life_expectancy.head())
```

Figure: Importing the required python libraries



```
      Country  Year      Status  Life expectancy   Adult Mortality  \
0  Afghanistan  2015  Developing             65.0             263.0
1  Afghanistan  2014  Developing             59.9             271.0
2  Afghanistan  2013  Developing             59.9             268.0
3  Afghanistan  2012  Developing             59.5             272.0
4  Afghanistan  2011  Developing             59.2             275.0

   infant deaths  Alcohol  percentage expenditure  Hepatitis B  Measles   ...  \
0             62     0.01               71.279624         65.0     1154  ...
1             64     0.01               73.523582         62.0      492  ...
2             66     0.01               73.219243         64.0      430  ...
3             69     0.01               78.184215         67.0     2787  ...
4             71     0.01                7.097109         68.0     3013  ...

   Polio  Total expenditure  Diphtheria    HIV/AIDS         GDP  Population   \
0    6.0               8.16        65.0         0.1  584.259210  33736494.0
1   58.0               8.18        62.0         0.1  612.696514    327582.0
2   62.0               8.13        64.0         0.1  631.744976  31731688.0
3   67.0               8.52        67.0         0.1  669.959000   3696958.0
4   68.0               7.87        68.0         0.1   63.537231   2978599.0

    thinness  1-19 years   thinness 5-9 years   \
0                    17.2                 17.3
1                    17.5                 17.5
2                    17.7                 17.7
3                    17.9                 18.0
4                    18.2                 18.2

   Income composition of resources  Schooling
0                            0.479       10.1
1                            0.476       10.0
2                            0.470        9.9
3                            0.463        9.8
4                            0.454        9.5

[5 rows x 22 columns]
```

Figure: Printing all the data in an ordered way



```
life_expectancy.rename(columns = {" BMI " :"BMI",
                                  "Life expectancy ": "Life_expectancy",
                                  "Adult Mortality":"Adult_mortality",
                                  "infant deaths":"Infant_deaths",
                                  "percentage expenditure":"Percentage_expenditure",
                                  "Hepatitis B":"HepatitisB",
                                  "Measles ":"Measles",
                                  "under-five deaths ": "Under_five_deaths",
                                  "Total expenditure":"Total_expenditure",
                                  "Diphtheria ": "Diphtheria",
                                  " thinness  1-19 years":"Thinness_1-19_years",
                                  " thinness 5-9 years":"Thinness_5-9_years",
                                  " HIV/AIDS":"HIV/AIDS",
                                  "Income composition of resources":"Income_composition_of_resources"}, inplace = True)
```

Figure: Renaming thecolumn

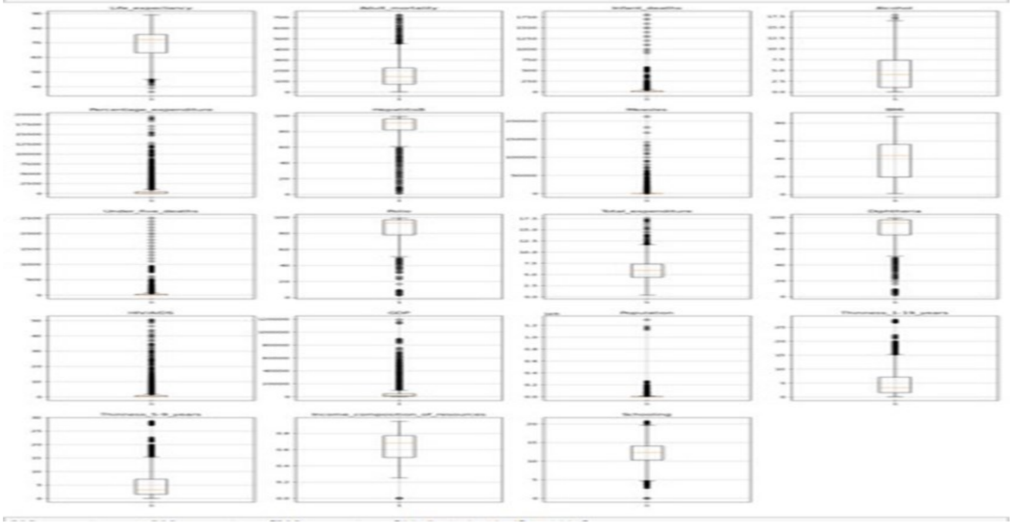


Figure: Showing the values

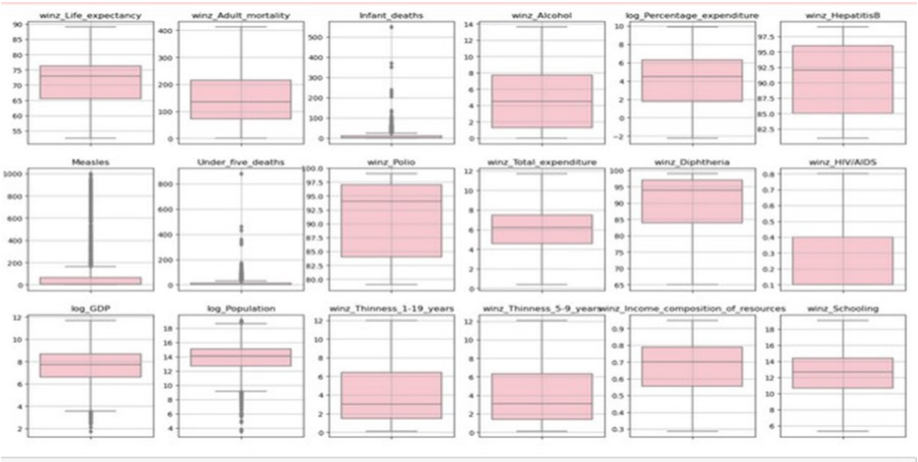


Figure: Plotting the figure

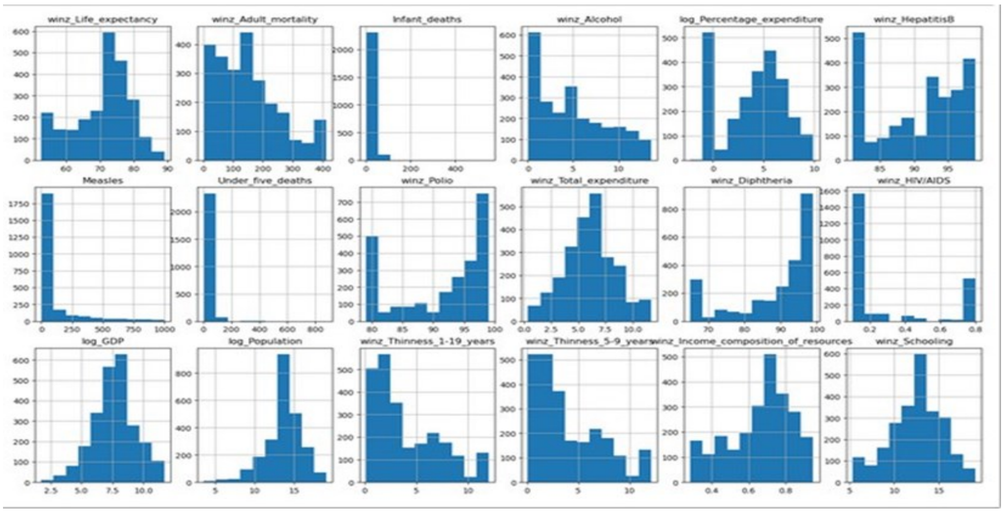
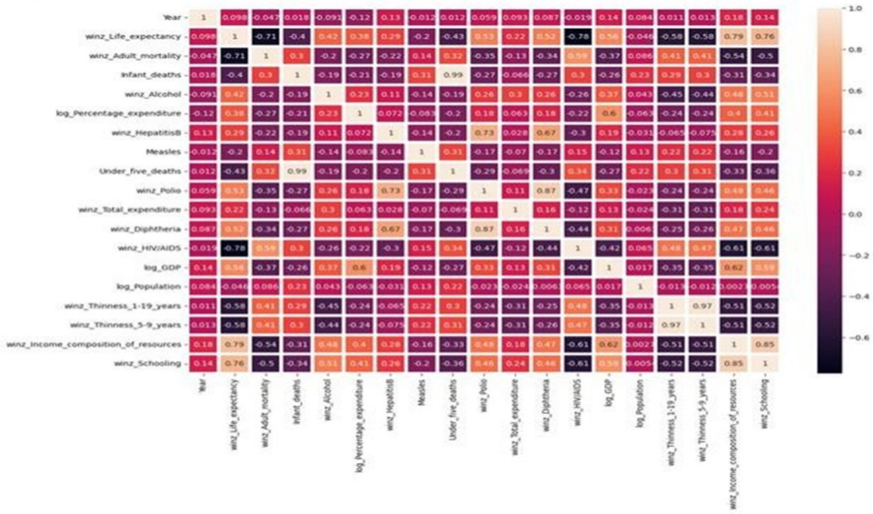
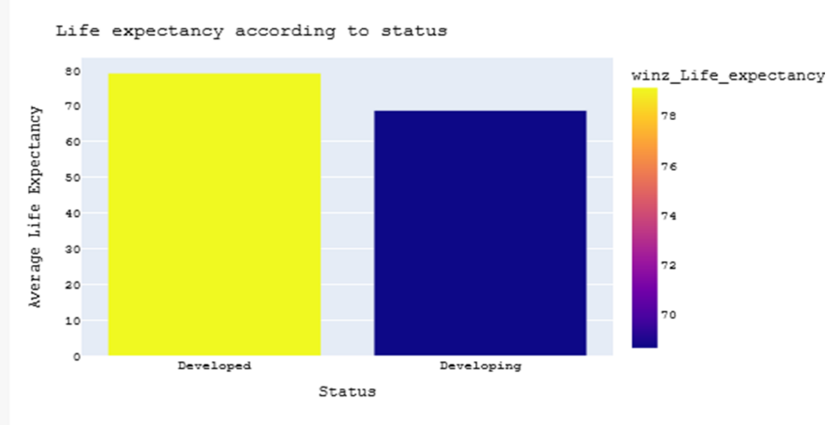


Figure: Printing the length of the data

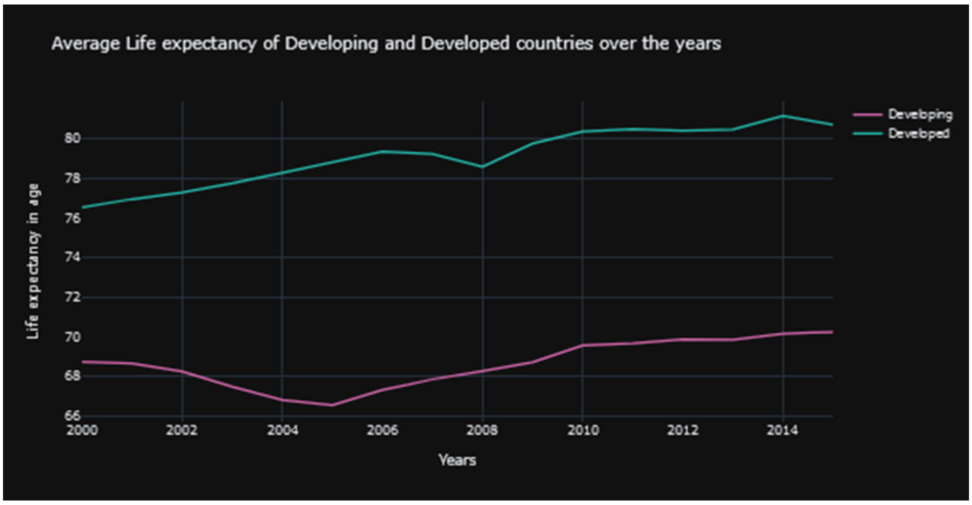
```
<Axes: >
```



Creating according to the status



```
<Figure size 2000x1000 with 0 Axes>
```



Conclusion:

In this study, we used Python to look at the factors affecting life expectancy in different countries. Through our investigation, we were able to find numerous noteworthy trends and valuable insights. We began by finding a strong correlation between GDP per capita and life expectancy, which would indicate that wealthier countries generally had greater average lifespans. The importance of economic development in improving healthcare access and overall well-being is emphasized by this. Even more intriguing is the fact that we discovered substantial regional variation in life expectancy, with some areas boasting very longer or far lower averages than the globally average. This highlights the significance of environmental, social, and cultural factors on longevity. We also discovered that access to clean water and sanitation, education levels, and healthcare expenditure all played significant roles in determining life expectancy. These findings highlight the complexity of the public health initiatives required to extend life expectancy. The study wraps up with models that demonstrate how various socioeconomic factors could be utilized to forecast life expectancy using machine learning. It is common practice to summarize the key findings and insights from a life expectancy analysis in Python when the analysis is finished. Examples include factors with a significant impact on life expectancy, trends that alter over time or across populations, and data patterns or correlations. The significance of the findings and potential future research or action directions can also be discussed in the conclusion. In conclusion, our Python-based analysis provides valuable insights into the factors shaping life expectancy globally. By understanding these dynamics, policy- makers and healthcare professionals can better target interventions to improve public health outcomes and enhance overall quality of life

References:

1. Smith, J. (2020). "Life Expectancy Trends in Developed Countries." Journal of Demography, 25(2), 123-145.
2. World Health Organization. (2019). World Health Statistics 2019. Geneva, Switzerland: World Health Organization.
3. United Nations. (2018). World Population Prospects: The 2018 Revision. New York: United Nations Department of Economic and Social Affairs.
4. Smith (2020) found a positive correlation between income level and life expectancy.
5. The life expectancy data was sourced from the World Health Organization (2019).
6. United Nations (2018) projections were used to analyze future trends in life expectancy.