

---

**VIETNAM NATIONAL UNIVERSITY - HO CHI MINH CITY  
INTERNATIONAL UNIVERSITY**



**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

**DATA SCIENCE & DATA VISUALIZATION**

**PROJECT PROPOSAL**

Course by Dr. Tran Thanh Tung

**Topic: The influence of factors on the life expectancy of each country**

GitHub repository: [link](#)

---

**TEAM MEMBERS**

No.	Name	Student IDs
1	Le Thi Thuy Nga	ITDSIU21024
2	Duong Khanh Trang	ITDSIU21127
3	Tran Quynh Nhu	ITDSIU21105

---

## TABLE OF CONTENT

<b>I. INTRODUCTION</b>	<b>4</b>
A. Abstract	4
B. Purpose	4
C. Goal	4
<b>II. DATA-PULLING PROCEDURES</b>	<b>4</b>
A. Data sources	4
B. Data file	4
C. Aim	4
D. Procedure	4
E. Specific features	5
<b>III. TOOL USED</b>	<b>5</b>
<b>IV. PROJECT TIMELINE</b>	<b>5</b>
<b>V. REFERENCE</b>	<b>6</b>

## I. INTRODUCTION

### A. Abstract

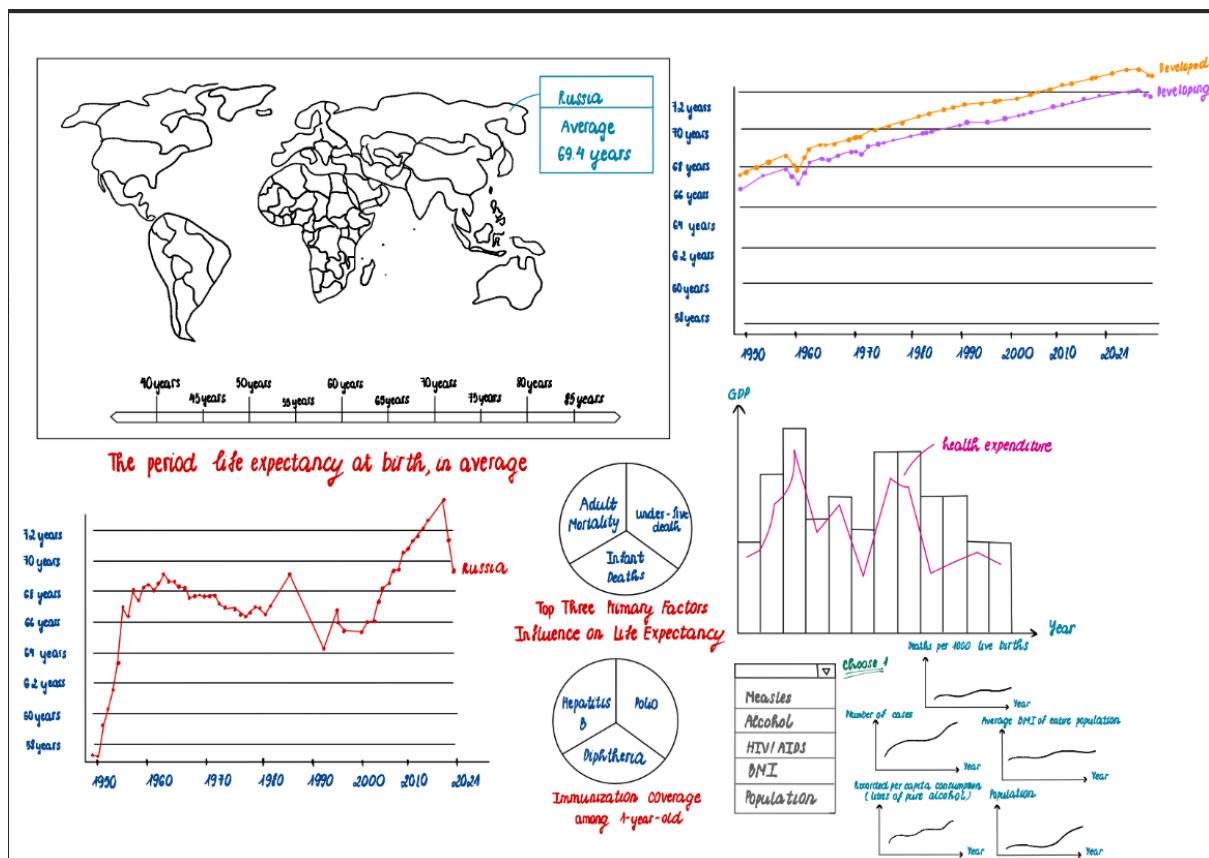
This study explores the multifaceted impact of socio-economic, health, and environmental factors on life expectancy across various nations, utilizing a comprehensive dataset that spans several years. Through quantitative analysis, we aim to discern the relative importance of these variables, acknowledging the complex interplay between lifestyle choices, healthcare access, socio-economic status, and environmental conditions in determining longevity.

### B. Purpose

- Implementing a dashboard with clear visualization and charts to show differences in factors that affect the life expectancy of each country.
- Providing people knowledge about how factors can affect longevity.
- Providing fundamental insights to answer the question.

### C. Goal

- Preparing data to determine qualities needed for visualization.
- Providing several charts and dashboard layouts to assist customers get information.
- Updating and changing charts with D3.js libraries to better accommodate future changes.
- Using charts to communicate and to display important data from the dashboard.



---

## II. DATA-PULLING PROCEDURES

### A. Data sources

Kaggle stands out as a renowned platform for data science, offering researchers a vast selection of datasets. These datasets, ideal for data visualization, come from reliable sources and maintain a high level of quality.

### B. Data file

The chosen dataset is in csv format, which is a common and easy to use format for data visualization. The dataset contains 22 features, including 2 string features and 20 numerical features. However, just some of these features are used to visualize because of their meaning in this study.

### C. Aim

The aim of this project is to obtain meaningful insights into impacts of lifestyle choices, healthcare access, socio-economic status, and environmental conditions on life expectancy. The data-set aims to answer the following key questions:

- Do various predicting factors which have been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting life expectancy?
- Should a country having a lower life expectancy value(<65) increase its healthcare expenditure in order to improve its average lifespan?
- How do Infant and Adult mortality rates affect life expectancy?
- Does Life Expectancy have a positive or negative relationship with drinking alcohol?
- Do densely populated countries tend to have lower life expectancy?
- What is the impact of Immunization coverage on life Expectancy?

### D. Procedure

The following steps will be taken to achieve the aim of the project:

- Data processing: Remove some specific features and values to obtain clean data for better understanding.
- Data visualization: Several data visualization methods, including JavaScripts libraries, will be employed to delve into the dataset and uncover patterns and trends. This approach may encompass the use of charts, graphs, and maps.
- Interpretation: The results of the data visualization will be interpreted to draw meaningful insights into the impacts of factors on life expectancy of each country.

### E. Specific features

This project will have some features such as zoom in and zoom out of world map, sort by year to express factors affected on life expectancy.

## III. TOOL USED

The following tools will be used to complete the project:

- GitHub: control the version of the system that will be used to manage the code and data for the project. [Link](#)
- Visual Studio Code: build the web pages for the data visualization.

- HTML/CSS/JS: build the web pages for the data visualization.
- Python: process the data, create demo data visualizations, and test further features

#### IV. PROJECT TIMELINE

STAGE	TASK	MEMBER	WEEK
<b>PLANNING</b>	Build the planning	Nga	1
	Topic research	All	
	Establish the analysis's objectives and purpose	All	
	Research for science-based information on the topic	All	
<b>DATA PROCESSING</b>	Collect dataset	Trang	2-3
	Data cleaning	Nhu	
<b>c</b> <b>DATA VISUALIZATION DESIGN</b>	Design UI/UX	Nga	4-7
	Select a theme and charts for visualization	All	
	Sketch the diagrams and user flow	Trang	
	Sketch activity diagram	Trang	
	Review the final design	All	
<b>WEBSITE DESIGN</b>	Design the website's layout	Nga + Trang	8-13
	Coding the backend and frontend	All	
	Review the final web coding	All	
<b>PRESENTATION</b>	Write final report	All	14
	Do presentation slide	All	

#### V. REFERENCE

1. Dataset:  
[https://www.kaggle.com/code/varunsaikanuri/life-expectancy-visualization?fbclid=IwAR3lERVqXxcH5RMTFWKgt77kQ8j2iAHonEqUokNZl69dU5R\\_\\_w4lqyijl4](https://www.kaggle.com/code/varunsaikanuri/life-expectancy-visualization?fbclid=IwAR3lERVqXxcH5RMTFWKgt77kQ8j2iAHonEqUokNZl69dU5R__w4lqyijl4)
2. D3. (n.d.). Retrieved from <https://d3js.org/>