**Group 5 Proposal: Global life expectancy analysis**

**Problem:**

Global life expectancy analysis

**Dataset:**

The dataset will be acquired from WHO dataset ([http://www.who.int](http://www.who.int/)) and Kaggle open dataset ([https://www.kaggle.com](https://www.kaggle.com/)), integrating health conditions of 193 countries in the near 15 years. Many factors related to those conditions are listed in the table, such as Hepatitis B, measles, HIV, infant death and adult mortality etc. Also, we will discovering other related information such as GDP, population, climate data which might be related to life expectancy.

**Proposed Statement:**

Our proposed solution is to analyze global life expectancy from different aspects. We intend to compare country’s life expectancy geographically and chronologically, showing the health condition comparison visually. We aim to derive some key factors that influence the life expectancy of different continents and even some specific countries. Plus, we are going to analyze geographically, drawing a world map with different climates, economics, diseases, regions, cultures. As for chronological analysis, we will focus on the exact country and curve the fluctuation of life expectancy in the past 15 years. This analysis may show us some details and regular patterns, which can be used to provide some recommendations for different countries or regions, such as specific health-care, population control etc. We may also be able to make future predictions by seeing the tendency in terms of dataset from the past 15 years.

**Implication:**

One of the real-world applications of this solution is to raise awareness about factors effecting life span. If we visualize it clearly, we could provide guidelines to different countries, about what they should focus on for longer life spans. The other application is to predict the life expectancy tendency of countries or regions. We could draw more attention to countries with lower life expectancy as we would have concrete evidence and analysis.

**Project steps**

|  |  |  |
| --- | --- | --- |
| Steps | Estimated completion time | Person(s) involved |
| 1. Extracting & integrating data | Half week | Arda Bati, Weinan Li |
| 2. Cleaning data | half week | Qingyuan Jin, Wei Zhang |
| 3. Data analysis (comparison and prediction) | One week | All group |
| 4. Visualization | One week | All group |
| 5.Recommendation(application) | Half week | All group |