

Simple Analysis of World Suicide Rate with Python

Zheng Ma, Baorong Zhong

School of Computer Science, Yangtze University, Jingzhou, 434020, China

Abstract

With the advent of the era of artificial intelligence, data analysis has attracted wide attention from all walks of life. The computer language programming method is used to complete the data analysis work, which is not only highly efficient, but also has high reliability of the data analysis results, and has multiple uses. In the data-driven mode, whether it is product design and development or project management, strong data analysis is required to provide support. The data source of this project is from the World Health Organization, which includes the number of suicides and total number of people in 141 countries from 1979 to 2016 in all ages. Panel data, etc. Analyze the correlation between univariate and multivariate for several of the parameter variables, and mine the associated factors, so as to analyze the suicide rate of countries around the world.

Keywords

Suicide rate; Python data analysis.

1. Python Data Analysis Technology

Python language is an interpreted high-level programming language that supports object-oriented programming and design, and can handle dynamic data types. At present, third-party program databases and practical tools have been developed. Since the Python language was publicly released in the 1990s, it has become a popular scripting language because of its concise syntax and rich class libraries, which are suitable for rapid development activities. When programming in Python, programmers do not need to consider low-level details such as memory footprint, which can effectively improve programming efficiency. And the Python language has good consistency and can be ported on multiple platforms. It belongs to the open source code, and can be run directly from the source code during the process of program interpretation and execution, without compiling. These characteristics have laid the foundation for the widespread application of Python language. And Python has powerful data analysis functions, which can be applied to many fields such as website development, image processing, data statistics, and visual expression. In May 2017, the Python language successfully surpassed the Java language and ranked first in the programming language rankings. It is currently one of the most important programming languages and has become the programming technology of choice for data analysis.

2. Data Analysis Process

The use of Python to carry out data analysis work specifically includes the following basic processes. In the process of data analysis, it is necessary to first clarify data usage requirements, such as financial data analysis, process analysis, etc., and according to the characteristics of target data Use appropriate data analysis methods to dig deep into data characteristics and potential utilization values. Data acquisition. After clarifying the data analysis requirements, the target data should be collected as comprehensively as possible. Specifically, local acquisition and web crawler acquisition methods can be used. Among them, the method for obtaining web crawlers is to use Python programming to legally obtain relevant data in the

network to meet the data acquisition requirements in big data analysis. Data pre-processing. Before the formal data analysis, the data pre-processing must be completed first, specifically by merging, cleaning, transforming and standardizing the target data to meet the needs of subsequent modeling and analysis. In the process, the quality of the data can be improved, and the efficiency of data analysis can be improved. Modeling and optimization. In the process of data analysis, modeling is an important link. Specifically, the target data can be processed by establishing cluster models, association rules, and intelligent algorithms. After the modeling is completed, its performance should be evaluated and optimized to ensure that it is suitable for actual data analysis conditions. Result generation. After the data analysis is completed, the results must be displayed and output in a visual form, which is convenient for users to use the data results.

3. Description of the Experiment Process

3.1. Data Cleaning

Because of the time span of the data and the large amount of national design, some of the data are missing or complete statistics, which need to be processed for missing values. Read the data, browse the data format and type as a whole.

3.2. Data Regularization

Data regularization includes slicing, aggregation, grouping, and reconstruction to analyze data from different dimensions. It is necessary to slice, aggregate, group, and reconstruct data from different variables, so that we can visualize the data. First, starting from the overall data, we use country and time as keywords to aggregate and sum the data. Perform a slicing operation on the data, extract all the data of a single country, perform a slicing operation on the data, and then extract the data of all countries in a certain year, pivot the age column 'length' into 'width', and then merge the country and year columns into One column, restructure the data table, pivot the age column 'length' to 'width'.

4. Data Visualization Analysis

4.1. Observe the Data from A Holistic Perspective

Observing the data from an overall perspective, we can find that the annual suicide rate fluctuated little from 1979 to 2016, and after 2000, the world suicide rate decreased year by year; from the total suicide rate of all countries from 1979 to 2016, it can be found that the total suicide rate from 1979 to 2016 is large. The suicide rate in some countries is below 0.02%, a small number of countries are between 0.02% and 0.03%, and only a few countries have a high suicide rate.

4.2. Special Case Analysis

For countries with a high suicide rate (for example, Belize), we take out the data separately for analysis. It can be found that the suicide rate of men in this country is generally higher than that of women, and the suicide rate in certain years (such as 1994, 2000, etc.) has shown a significant explosive growth. In these years, the suicide rate of 75+ elderly people Very high, 55-74 people also have a high suicide rate, and the suicide rate shows a positive correlation at the age level. For the reasons, we take one year's male data set (1994 and 2000) as a reference, observe the suicide rate of each age level in various countries of the world in the year, and analyze whether it is a special phenomenon in Belize individual countries or a common phenomenon.

4.3. Analyze from A Certain Angle

It can be seen from the figure that as the age increases, there is a positive correlation between suicide rate, which is a common phenomenon. The data of 5-14 is the most dense, indicating that the suicide mortality rate of children in most countries in the world is less than 1 in 10,000; 15-24, 25-34, 35-54, 55-74, 75+ five age levels are gradually sparse, showing that as the age increases, the probability of suicide rate gradually increases; The 75+ mortality rate clearly shows that the degree of aggregation is very low, showing a uniform distribution, which indicates that the suicide rate of the elderly is generally high in countries around the world and is significantly higher than other age levels. Belize has a very high overall suicide rate, but the suicide rate in 1995 and 2000 was at an intermediate level, indicating that its average annual mortality rate is high.

5. Summary

From the above analysis, we draw some conclusions. The suicide rate in most countries of the world is at a normal level, and it has been decreasing year after year since 2000. The average annual suicide rate in individual countries is high, and the suicide rate is generally higher in men than in women. The suicide rate of children is generally low. As the age increases, the suicide rate is on the rise, and the suicide rate of 75+ elderly people is generally higher. There is no obvious correlation between the suicide rates of all ages.

References

- [1] TorresMendoza Yaritbel et al. Community Assessment for Mental and Physical Health Effects After Hurricane Irma - Florida Keys, May 2019.[J]. MMWR. Morbidity and mortality weekly report, 2021, 70(26): 937-941.
- [2] Liao Juan and Ruan Yunfei. Research on APP Intelligent Promotion Decision Aiding System Based on Python Data Analysis and AARRR Model[J]. Journal of Physics: Conference Series, 2021, 1856(1)
- [3] Tian Zichun and Wang Bin. Use Python Data Analysis to Gain Insights from Airbnb Hosts[J]. Advances in Mathematical Physics, 2021, 2021.
- [4] Groh Micah et al. PandAna: A Python Analysis Framework for Scalable High Performance Computing in High Energy Physics[J]. EPJ Web of Conferences, 2021, 251.
- [5] Cao Shengjia et al. Research on Python Data Visualization Technology[J]. Journal of Physics: Conference Series, 2021, 1757(1): 012122.