

Design and Implementation of Chinese Population Data Analysis and Visualization System

Jin Wang^{1, a}, Taizhi Lv^{2, b}

¹College of Information Technology, Jiangsu Maritime Institute, Jiangsu Nanjing, 211170, China;

²Nanjing Longyuan Microelectronic Company Limited, Jiangsu Nanjing 211106, China.

^a 1497163039@qq.com, ^b 404050569@qq.com

Abstract

Based on the importance of population analysis, this paper designs and implements a population data analysis system. Based on the Python language, the system uses crawler technology to obtain population data from the census, analyzes the population status, analyzes China's male to female ratio, the total population, as well as the birth rate and death rate, and the distribution of population age groups. Make predictions about future population development based on statistics. The crawler module uses Request to obtain the source code of the webpage, analyzes and locates tags through XPath, and calls Openpyxl to store the data in Excel. The data analysis module uses Pandas and Sklearn to implement data statistics and prediction functions. The visualization module uses Pyecharts to complete the drawing.

Keywords

China population, Python, data analysis, data visualization.

1. Introduction

What is the most important problem in today's society? It is undoubtedly the population problem. The population problem has a pivotal impact on my country's economy, society, politics and other aspects. China is a populous country with a population of 1.3 billion. The population has been growing, and it is expected to reach about 1.4 billion in 2035 [1-2]. At present, China's population is facing many problems. The structure shows a three-up and three-down trend, that is, the size of women of childbearing age, the size of births, and the size of the labor force have declined, and the size of the elderly population, the size of family households, and the dependency ratio have risen. The size of China's aging population is also increasing. In the long run, it will affect the country's resource allocation and the country will enter a state of overload. Therefore, the country needs to continuously introduce new population policies to adapt to the ever-evolving population structure and achieve reasonable resources. Configuration helps to promote the rationalized development of my country's economy and society. This article mainly analyzes China's population problems through data collection, data cleaning, data analysis, data visualization and other processes, and makes predictions about the future.

2. Data crawl

This article mainly crawls the population data of the National Resources Network in recent decades, including the total population, birth rate, death rate, male to female ratio, and the population distribution of each age group, and then according to these data, according to data

collection, data cleaning, and data analysis. The steps of data visualization realize the system flow.

The data crawling module first imports the relevant libraries that need to be used in the project, such as the Pyecharts configuration method, the chart drawing method in the module, and the global theme customization method, and import Openpyxl to process the relevant data in excel. Initiate a request to the target site through the HTTP library. The request can include additional headers and other information. If the server can respond normally, it will get a response, which is the content that needs to be obtained.

Use the lxml parsing library to extract the data from the label analysis of the source data. The project uses MySQL to store the crawled data by SQLAlchemy[3-4]. Use mysqlclient in the Scrapy pipeline to obtain the connection and execute the MySQL insert operation by self.client.population.china.insert(item). Among them, population is the name of the database and china is the name of the table. After the execution, the data is inserted into it.

Assessment methods are flexible and appropriate to ensure the consistency between students' On-campus learning and actual work, combine on-campus performance assessment with enterprise practice assessment, and explore the integration of classroom and practice location.

3. Data exploration

3.1. Data exploration

First, master the basic situation of the data through basic exploration of the data, including the following explorations

- (1) The number of births and deaths in each year, these data can reflect the trend of population changes.
- (2) The proportion of male population and female population. These data can reflect the structure of the population.
- (3) The proportion of the population of each age group. The data can reflect the number of laborers and the demographic dividend.

3.2. Data cleaning

Generally, dirty data is divided into three categories: incomplete data, wrong data, duplicate data [5]. Through data cleaning, the problems of missing value, out of bounds value, inconsistent code, duplicate data and so on are solved. Browse the data of each field in the database, determine the data type, analyze the fields that need to be cleaned such as birth population, death population, male population, female population, etc., remove empty data in these fields, and useless data such as values to be determined. Remove Useless characters.

4. Data visualization

Pyecharts is a powerful data visualization tool combining Python and Echarts [6]. This paper realizes the development of charts based on Pyecharts.

4.1. Population trends graph

The visualization of population trends is completed by the following steps:

Step 1: Obtain the population data in the MySQL database by SQLAlchemy.

Step 2: Create multiple lists to store time, total population, number of births, number of deaths, number of males, and females, respectively, and write them into the corresponding lists.

Step 3: Traverse all data elements, extract target data items, and return 6 lists.

Step 4: Input the parameters and draw the trend graph of China's population change over time.

Step 5: Finally, save it as an html file.

The trend chart of China’s population change over time is shown in Figure 1. The total population of China from 1989 to 2019 and the change law of my country’s total population over the past few decades are obvious. In the past few decades, Chinese population is increasing always, but the rate of increase is getting slower and slower. This shows that under the implementation of family planning and other population policies in China. The total population of China has been controlled to a certain extent. Although the trend of population growth has slowed down, China population is still in the process of growth due to its large base.

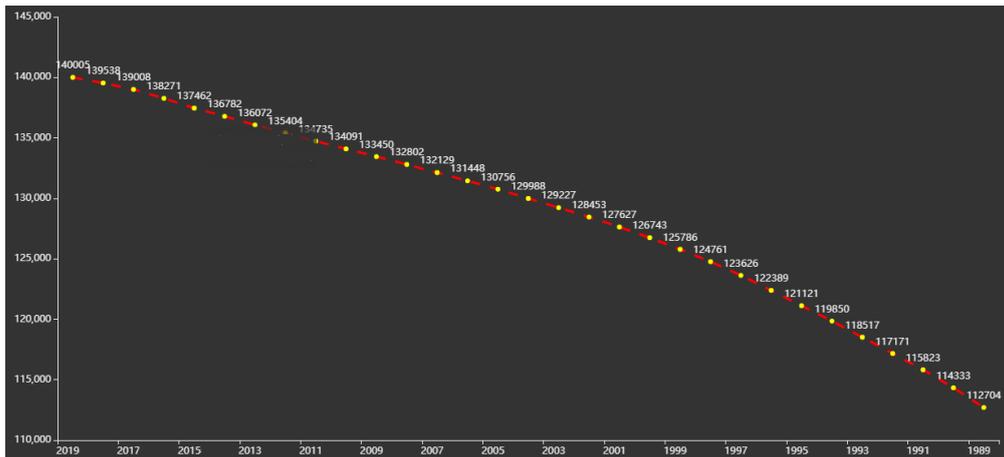


Figure 1. population trends graph

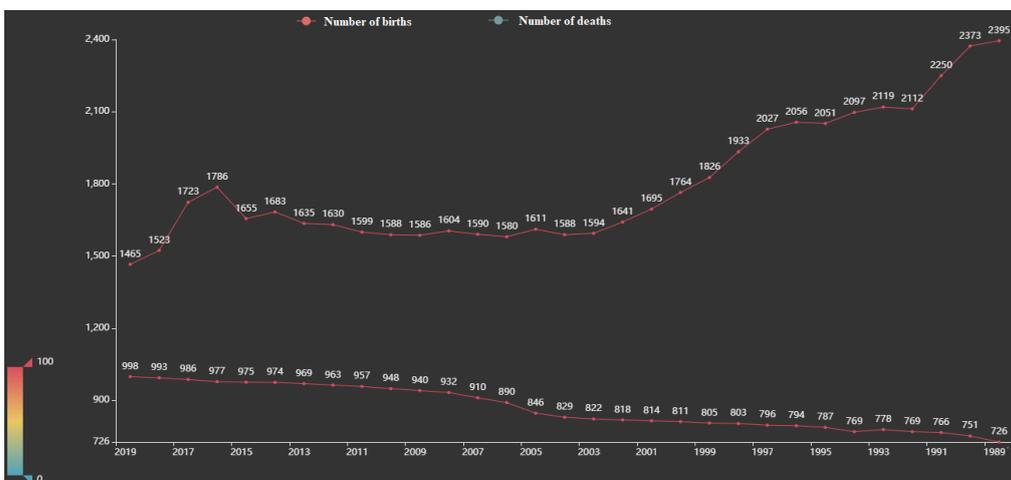


Figure 2. Birth rate and death rate comparison chart

4.2. Birth rate and death rate comparison chart

The development process of the birth rate and death rate comparison chart is basically similar to the visualization of population trends. The three parameters of time, number of births, and number of deaths are used as formal parameters. It can be seen from Figure 2 that my country’s birth rate has been declining slowly since 1989, mainly due to the introduction of the national family planning policy. However, it can be seen from the graph that there was a fluctuation in 2016. Thanks to the country's reopening of the second-child policy, the number of births has increased again, and then it has started to decline. The main reason is that with the development of society, people's views on raising children have also changed. Thanks to the rapid development of the national medical level, people's welfare has improved, so people's life span has been extended. The death rate curve tends to be flat and declining. This is due to the continuous improvement of the national economy, medical care, and social welfare. People live

and work in peace and contentment. They are not ravaged by wars as before. Therefore, people's deaths and injuries have been reduced, and the birth rate has increased. This will also lead to an increase in the total population.

4.3. Age distribution chart

The age distribution chart is realized through initialization, data reading, graphic drawing, saving as HTML file and other steps, as shown in Figure 3.

It can be seen from the figure that China is still dominated by young and middle-aged people, accounting for a large proportion. The population aged 0-14 is 253.38 million, accounting for 17.95%; The population aged 15-59 is 894.38 million, accounting for 63.35%; The population aged 60 and above is 264.02 million, accounting for 18.70% (among them, the population aged 65 and above is 190.64 million, accounting for 13.50%). Compared with 2010, the proportion of people aged 0-14, 15-59, 60 and above increased by 1.35 percentage points, decreased by 6.79 percentage points and increased by 5.44 percentage points respectively. China's juvenile population has increased, and the adjustment of fertility policy has achieved remarkable results, but the number of elderly populations has also increased, which means that China's aging degree will be more and more serious.

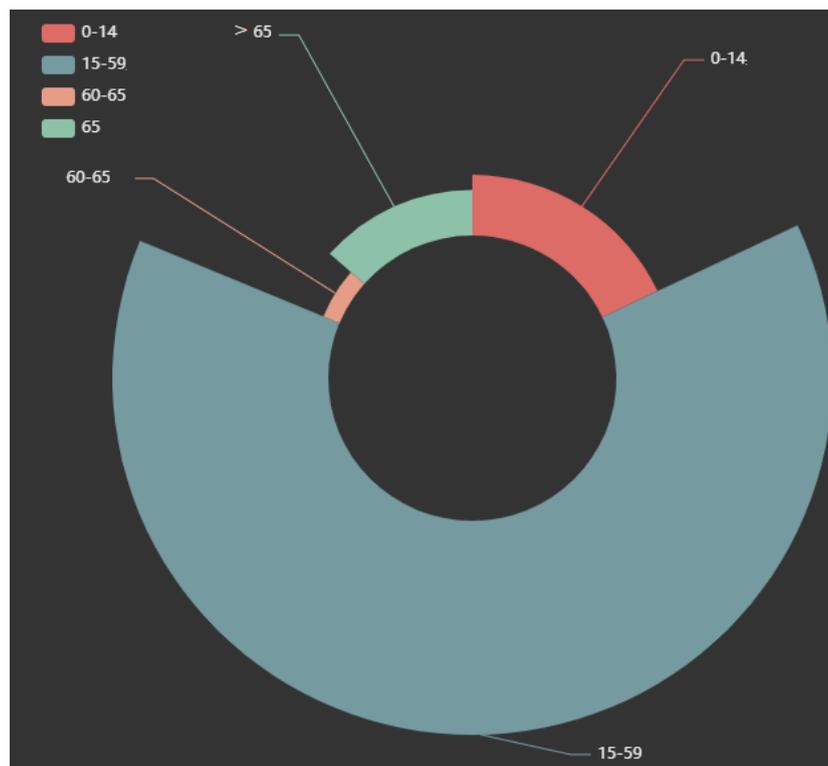


Figure 3. Age distribution chart

5. Conclusion

Although China has implemented various population plans in recent years, it still cannot change the phenomenon of China's sustained population growth. Moreover, with the continuous growth of population, the continuous improvement of medical level and people's living standards, China's population aging will be more serious. The purpose of this paper is to make statistics on China's population, then make charts, reasonably analyze the population and predict the future population development, so as to facilitate the introduction of reasonable population policies and make China's population present a reasonable structure, rather than favor one aspect, such as serious population aging, insufficient labor force and excessive

proportion of men and women, The population is too large, the population is too large, and the resources are evenly distributed. This paper collects population data based on crawler technology, stores the crawled data through MySQL database, realizes data statistics and analysis based on pandas, and realizes data visualization through Pyecharts.

Acknowledgements

This work was financially supported by the funding the Qianfan science and technology team of Jiangsu Maritime Institute(Big data analysis and application research team), the young academic leaders of Jiangsu colleges and universities QingLan project, the excellent teaching team of Jiangsu colleges and universities QingLan project (Innovative teaching team of software technology specialty), and the big data collaborative innovation center of Jiangsu Maritime Institute.

References

- [1] Liu,Y.J. S. Tang, and K.Wu, et al. " Logical framework of China's population research from the perspective of economic geography." *Economic Geography* 41,10(2021):97-105.
- [2] Cai, H. B., and J. R. Han. " Artificial Intelligence Relieves the Pressure of Population Aging:Mechanism and Path." *Expanding Horizons* 2021.6(2021):20-26.
- [3] Relan, Kunal. "Database Modeling in Flask." *Building REST APIs with Flask*. Apress, Berkeley, CA, 2019. 27-58.
- [4] Wang, Yangchen, et al. "Character Relationship Management of Play Script System Based On SaaS." *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*. IEEE, 2018:709-713.
- [5] Gudivada, Venkat, Amy A, and Junhua D. "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations." *International Journal on Advances in Software* 10.1 (2017): 1-20.
- [6] Liu, Juan, and Xidong Guan. "Job information data analysis and visualization system implementation based on Python crawler." *Softw Eng Appl* 9.4 (2020): 317-325.