CHAPTER 4

INITIAL RESULTS

4.1 Introduction

This chapter 4 will discuss regarding the presents the findings and analyzation of data obtained from the datasets. All the findings will be presented, tabulated, and summarized to obtain the results of the study. This chapter begins with the introduction of exploratory data analysis as well as data visualization and descriptive statistics. The contents also provide initial insights gained from the data visualization findings by the researcher. Lastly, the feature engineering that involves during the exploratory data analysis process will be emphasized in this chapter.

4.2 Exploratory Data Analysis (EDA)

In the data science life cycle, the preliminary analysis phase known as "Exploratory Data Analysis" should be made. The purpose of this phase that needs to be done is to help the researchers to gain insight into the conditions of the data itself. This phase involves gaining an understanding of the response and its influencing factors before building the actual model. The graphical examination of data distribution within unique variables of a character utilizes bar graphs, while correlations among various elements are illustrated through visual methods such as scatter plots and heat maps (GeeksforGeeks, 2023).

In this section will presents the initial findings that contain data analysis and suitable visualization to help the reader to understand and be able to interpret the findings of the data. The interpretation and insights will be elaborated after the visualization made to gain clear understanding of the correlation between the variables. Furthermore, the feature engineering will be present as the researcher made some feature engineering involved during data cleaning process.

4.2.1 Data Visualization & Descriptive Statistics

This section will present several initial analyses that can be done with dataset collected. The findings includes the distribution of population by age group, gender based population analysis and state-wise population comparison, as the dataset contains the following variables such as "population", "age" , "state", "ethnicity" and "date".

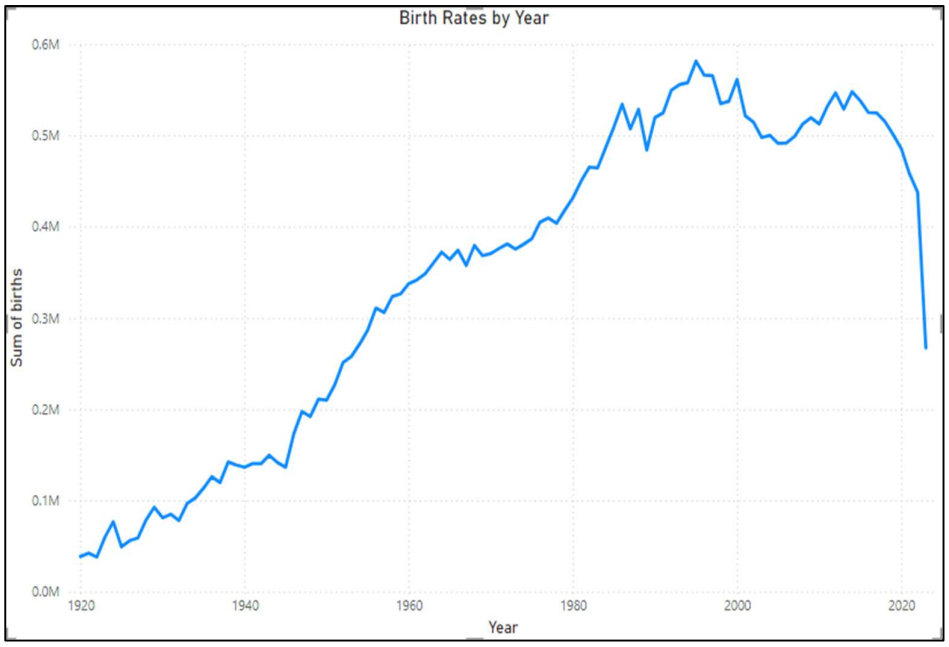

Figure 4.1 shows the line chart of birth trends from 1920 to 2023.

Figure 4.1 shows the line chart illustrates the trend in birth rates over time, spanning from 1920 to 2020. Birth rates exhibit a consistent upward trend from 1920 to the 1960s. Birth rates reached a peak around the late 20th century (1980s-2000s), suggesting a period of demographic expansion and stability. From the mid-2000s onward, there is a noticeable decline in birth rates. The sharp drop near 2020 may be linked to extraordinary circumstances, such as the COVID-19 pandemic, which disrupted healthcare systems, economic stability, and family planning decisions.

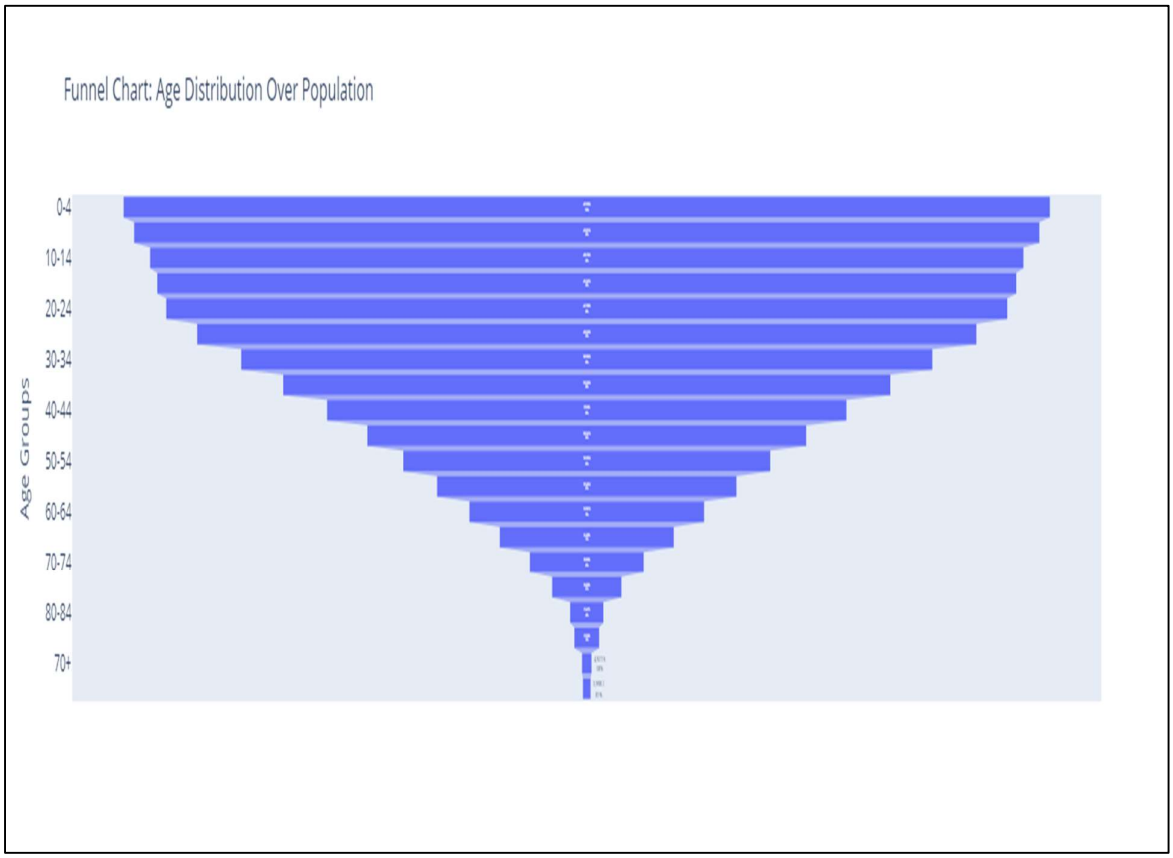

Figure 4.2 show funnel bar chart of age distribution of Malaysian population.

Figure 4.1 shows the funnel bar chart indicates the age distribution of Malaysian population. The bar chart is divided into age groups showing a breakdown of the population for each group. The funnel shape indicates a gradually declining population from younger to older groups. This is consistent with typical age distributions in many societies where younger generations are more numerous due to higher birth rates.

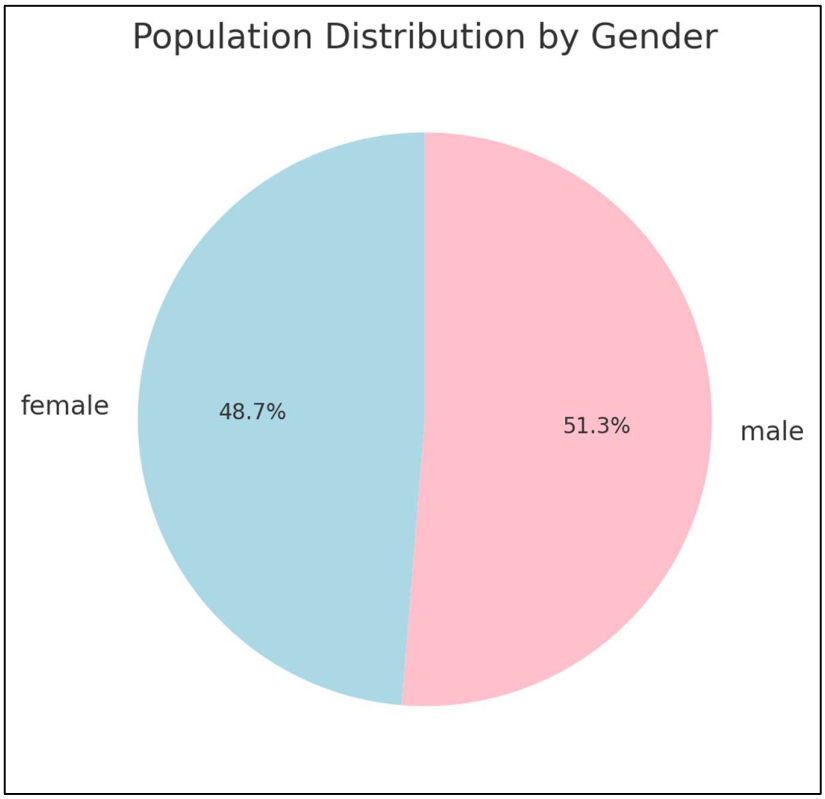

Figure 4.3 shows a pie chart of population distribution by gender.

Figure 4.3 shows the bar chart that illustrates the percentage of the total population by state. The chart shows that 51.3% of the population is male, slightly more than half of the total population. Females make up 48.7% of the population, a slightly smaller proportion compared to males. The distribution suggests a near balance between the genders, with males being marginally more prevalent. Such small differences may reflect natural variations or cultural and social factors affecting the gender ratio. This level of gender parity is common in many populations and may suggest an equitable distribution in terms of natural birth rates and survival rates across genders.

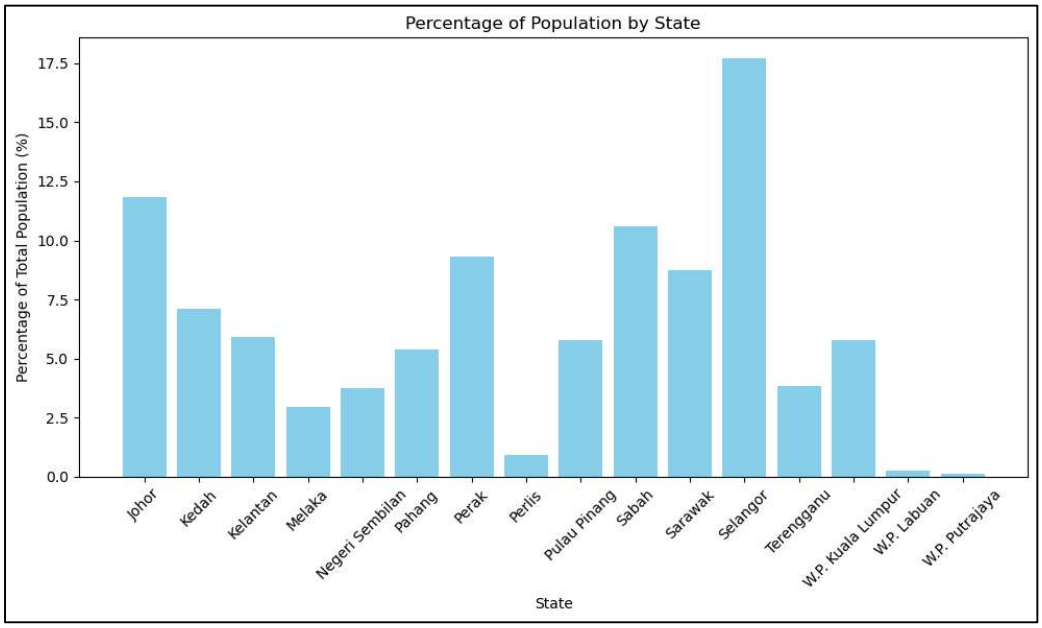

Figure 4.4 shows bar chart of the percentage of the total population by state in Malaysia.

Figure 4.4 shows the bar chart that illustrates the percentage of the total population by state. Selangor has the highest percentage of the population, exceeding 17%, making it the most populous state. This could be due to urbanization, industrial development, and better job opportunities. Johor has the second-highest population percentage, slightly above 12%. This reflects its strategic location and economic importance. States like Perlis**,** Wilayah Persekutuan Putrajaya, Labuan**,** and Wilayah Persekutuan Putrajaya have the smallest population percentages, likely due to their smaller geographical sizes or administrative roles.