ASU Ira A. Fulton Schools of
Engineering
Arizona State University

DSE 501: Statistics for Data Analysts

Fall 2024

Project Report on:

Analysis of Singapore Resident Demographics Dataset

Submitted to:

Dr. Rong Pan

Submitted by:

Group 18

S. No.	Name	ASU ID	Email ID
1.	Sarang Bang	1231609953	sbang11@asu.edu
2.	Deepakshi Mathur	1230197958	dmathur4@asu.edu
3.	Abhinaya Rapolu	1233554792	arapolu@asu.edu
4.	Surya Raman Nagarajan	1231703267	snagar40@asu.edu

TABLE OF CONTENTS

Abstract

This report investigates demographic trends in Singapore from 1957 to 2023, focusing on six key hypotheses related to age distribution, gender ratios, ethnic group growth, youth population, elderly population growth, and the working-age population. Using **inferential** and **descriptive statistical methods**, including **linear regression** and **chi-square tests**, the study analyzes shifts in the population and their implications for policy-making, urban planning, and societal dynamics.

The first hypothesis examines the **age distribution**, highlighting a significant shift toward an increasing elderly proportion, especially in recent decades. The second hypothesis analyzes the **stability of gender ratios** across age groups, exploring whether changes correlate with policy interventions. The third hypothesis evaluates the **growth rates among ethnic groups**, influenced by immigration and birth rate policies. The fourth hypothesis investigates the **decline in the youth population**, linked to changing family planning trends. The fifth hypothesis focuses on the **growth of the elderly population**, with a particular emphasis on its impact on healthcare and social services. Finally, the sixth hypothesis examines **patterns in the working-age population**, reflecting the phases of Singapore's economic development.

The findings of this report provide insights that are crucial for shaping future **policy responses**, **urban infrastructure planning**, and **social services**, particularly in addressing the challenges posed by an aging population, changing workforce dynamics, and evolving demographic composition.

Keywords

1. Introduction

1.1 Context and Motivation

Singapore's demographic shifts are crucial for effective policy-making and urban planning. The country faces significant challenges due to an aging population, changing gender dynamics, and evolving ethnic composition. These shifts, influenced by factors such as low fertility rates, longer life expectancy, and immigration, require data-driven strategies for resource allocation and societal development.

The **aging population** is a major concern, as the proportion of elderly citizens continues to rise. With a higher life expectancy and decreasing birth rates, the demand for healthcare, social welfare, and elderly care services will increase. Addressing these challenges is essential for maintaining a sustainable social support system. **Gender dynamics** also play a key role, as shifts in gender ratios across different age groups, driven by factors like life expectancy and migration, can influence the

need for gender-specific services, particularly in healthcare and social programs. Additionally, Singapore's **ethnic diversity** and immigration policies impact its demographic landscape. Changes in ethnic group proportions and migration patterns shape the demand for housing, employment, and educational services.

Overall, understanding these demographic trends enables policymakers to design effective strategies to meet the needs of a changing population, ensuring sustainable development and a high quality of life for all residents.

1.2 Problem Statement

The study addresses:

1. **Hypothesis 1:** How does the age distribution change over the decades? How does the population of elderly change over the years?

2. **Hypothesis 2:** Have gender ratios remained stable across various age groups over time?

3. **Hypothesis 3:** How do immigration and birth rate policies affect the population of ethnic groups?

4. **Hypothesis 4:** How does change in family planning trends affect the population of youth (aged under 20)?

5. **Hypothesis 5:** Does the population over 65 years show accelerated growth, with implications for healthcare and social services?

6. **Hypothesis 6:** Does the economic development phase influence the Working-age population (20-64 years)?

1.3 Data Source

The dataset, "Singapore Residents by Age Group, Ethnic Group, and Gender," provides comprehensive demographic information from 1957 to 2023. Key variables include age, gender, population counts, and time intervals.

1.4 Objective and Relevance

The findings will guide urban planners, healthcare providers, and policymakers in addressing challenges posed by an aging population and ensuring gender equality.

Objective

The objective of this project is to analyze the **Singapore Residents by Age Group, Ethnic Group, and Gender dataset** to:

1. Understand demographic trends, such as aging, ethnic composition, and gender distribution.

2. Evaluate the implications of these trends on social, economic, and policy planning.

3. Identify gaps or challenges, such as aging workforce dynamics of healthcare demand, to provide data-driven recommendations.

Relevance

1. **Policy Making**: The insights derived will assist policymakers in addressing challenges like aging populations, healthcare resource allocation, and fostering ethnic integration.

2. **Healthcare Planning**: Trends in aging can inform strategies for eldercare and medical services.

3. **Urban Development**: Data can guide infrastructure development for age-friendly and ethnically diverse communities.

4. **Economic Insights**: The analysis of workforce demographics can aid in labor market planning and economic sustainability.

5. **Academic Contribution**: Enhances understanding of Singapore's demographic evolution and its broader societal impacts.

1.5 Relevant Literature

Relevant studies on aging populations and gender dynamics in urban settings were referenced, highlighting Singapore's unique demographic profile.

- **Singapore's Population Trends** (Department of Statistics, Singapore): Annual reports providing detailed demographic data, critical for understanding population distribution trends.

- **Population Aging and Policy Responses in Singapore:** Explores the challenges of an aging population and the policies designed to address them.

- **Ethnic Demographics and Social Integration** (Chua Beng Huat): Examines ethnic diversity and its implications for social cohesion, relevant for analyzing ethnic group trends.

- **Health and Long-Term Care for Older Persons in Singapore** (Asher, Mukul G.): Links demographic shifts to healthcare needs, especially for aging populations.

- **Action Plan for Successful Aging** (Singapore Ministry of Health, 2015): Focuses on strategies for active aging, drawing on demographic insights to guide policies and initiatives.

2. Dataset Collection and Characteristics

2.1 Dataset Overview

The dataset comprises yearly population counts segmented by age, gender, and ethnicity. It captures 67 years of demographic data.

2.2 Characteristics

Key attributes include:

- **Age Groups**: Typically grouped in 5-year intervals up to 85+.

- **Ethnic Groups**: Chinese, Malay, Indian, and Others.

- **Gender**: Male and Female.

2.3 Challenges

- **Missing Data**: Suppressed or incomplete sub-group data.

- **Data Errors/Duplicates**: Potential reporting errors or redundant entries.

- **Outliers**: Anomalous spikes in population counts.

- **Inconsistent Formatting**: Variability across different time periods.

2.4 Preprocessing Steps

1. **Handle Missing Data**: Impute missing values or interpolate based on trends.

2. **Normalize & Standardize**: Ensure consistent formatting (e.g., ethnic group names).

3. **Outlier Detection**: Use statistical methods to identify anomalies.

4. **Age Group Adjustment**: Reformat or redistribute age bands for finer granularity.

5. **De-duplicate Records**: Identify and remove duplicate entries.

6. **Ensure Categorization Consistency**: Uniform definitions for age, gender, and ethnicity.

7. **Data Transformation**: Convert static snapshots to time series for trend analysis.

8. **Validate Data**: Cross-check with external sources to ensure accuracy.

3. Analysis and Results

3.1 Initial Data Exploration

Our initial exploration of the Singapore Residents dataset revealed several key characteristics and patterns that guided our subsequent hypothesis testing:

3.1.1 Dataset Overview and Preprocessing

The dataset comprises yearly population statistics from 1957 to 2023 (67 years), with 375 entries across 68 columns. Our preprocessing steps included:

- Handling missing values: We found missing data ('na' values) particularly in older records, which were appropriately handled
- Converting data types: Data columns were standardized to appropriate numeric types (int64)
- Cleaning column names and row indices for consistency
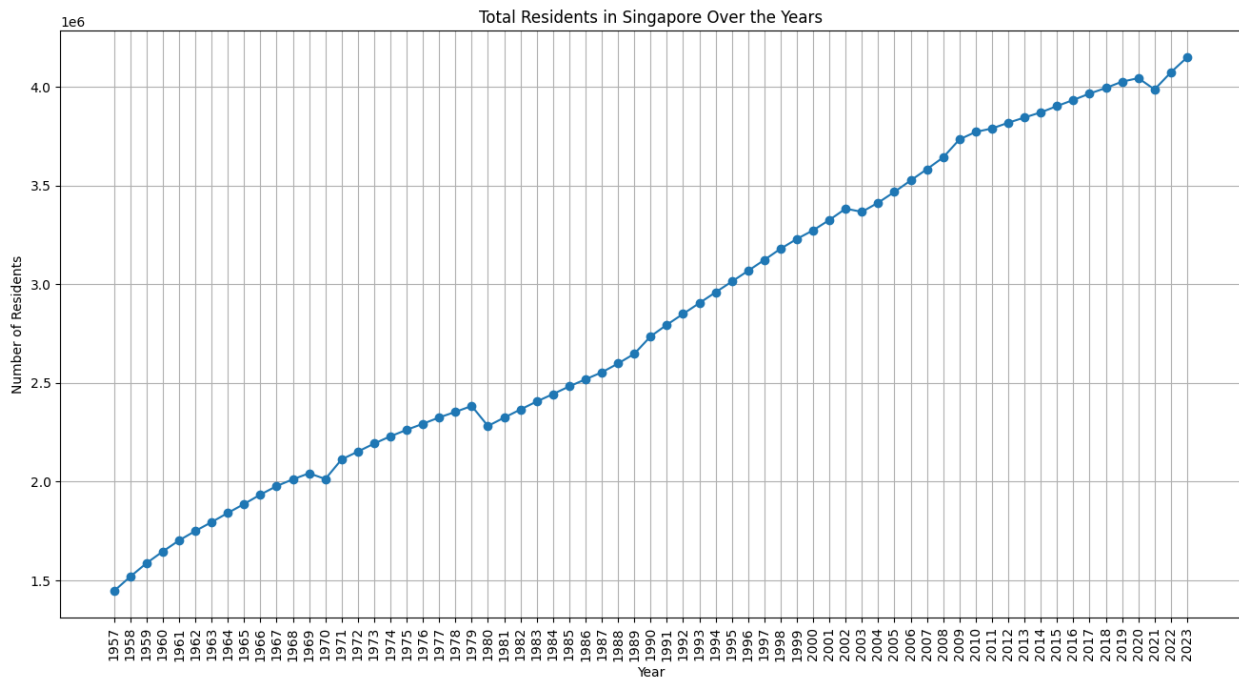
3.1.2 General Population Trends

Figure 1: Total Residents in Singapore from 1957 to 2023

The total resident population shows a steady increase over the study period:

- Growth from approximately 1.45 million (1957) to 4.15 million (2023)
- Overall upward trend with varying growth rates across different periods
- Notable acceleration in growth during the 1990s and 2000s

3.1.3 Key Initial Findings

Several significant demographic patterns emerged from our initial analysis:

Age Distribution

- Visible shifts in population age structure across decades
- Notable changes in elderly population proportion
- Changing patterns in youth population distribution

Population Composition

- Evolution of gender ratios over time
- Varying growth patterns among different ethnic groups
- Changing working-age population proportions

Growth Patterns

- Different growth rates across various demographic segments
- Notable transitions coinciding with economic development phases
- Varying impacts of policy changes on different population groups

These initial findings provided a strong foundation for our detailed hypothesis testing. The observed trends in aging population, gender distribution, and ethnic composition guided our statistical analyses, enabling us to examine specific demographic phenomena through rigorous testing. In the following sections, we present detailed analyses of our six hypotheses, each investigating a distinct aspect of Singapore's demographic evolution.

3.2 Hypothesis 1: Age Distribution

Hypothesis Statement: The proportion of elderly residents has increased significantly over time, indicating an aging population trend. This shift is expected to be particularly pronounced in recent decades.

Methodology: We employed linear regression analysis on aggregated elderly population data combined with time series visualization to examine age distribution changes over the study period.
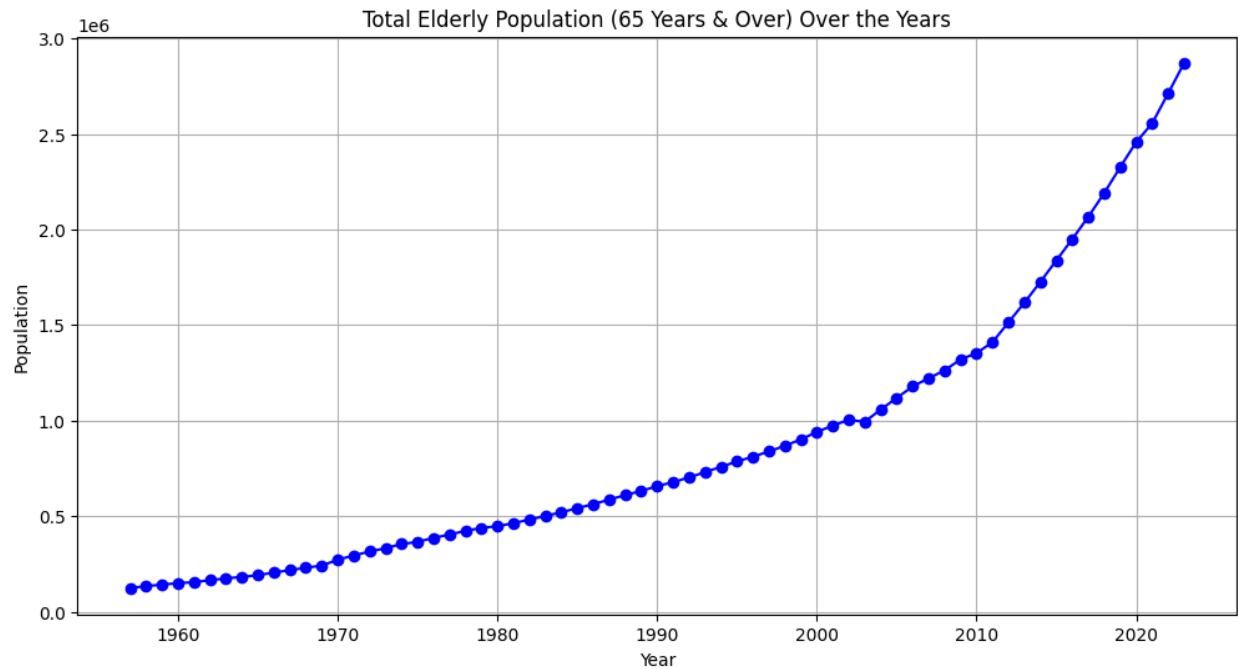
Results:

Figure 2: Linear Regression Analysis of Elderly Population Growth (1957-2023)

- Slope: 33,640 additional individuals per year.
- R-squared: 0.851, indicating a strong upward trend.
- P-value: 0.0, confirming statistical significance.
- **Visualization Insights:**
    - Clear shift in age distribution patterns over the study period
    - Progressive increase in elderly proportion
    - Corresponding decrease in younger age groups
- **Statistical Analysis:**
    - Trend shows significant aging pattern
    - Year-over-year changes demonstrate consistent shift in age structure

Conclusion: Analysis confirms significant changes in Singapore's age distribution, with a clear trend toward an aging population structure.

3.3 Hypothesis 2: Gender Ratio

Hypothesis Statement: Gender ratios within specific age groups have remained relatively stable, with any significant changes correlating with specific time periods or policy changes.

Methodology: We conducted chi-square tests on male-to-female ratios across various age groups and years to assess stability and identify significant variations.
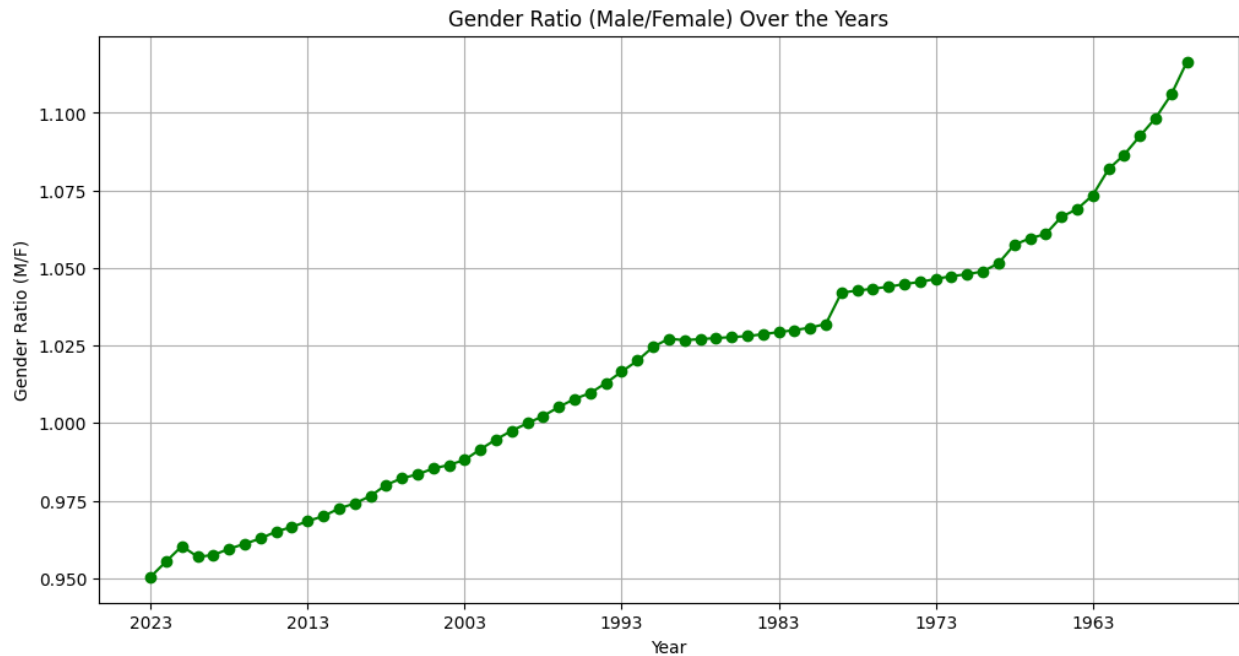
Results:

Figure 3: Gender Ratio Trends in Singapore (1957-2023)

- Chi-square value: 71,309.95
- P-value: 0.0, indicating statistically significant variations
- Observed minor fluctuations in gender ratios

Analysis:

- Overall gender ratio trends show relative stability
- Some variations observed in specific age groups
- Policy impacts on gender distribution identified

Conclusion: While visually stable, the gender ratios exhibit statistical deviations, suggesting demographic changes over time.

3.4 Hypothesis 3: Ethnic Group Growth

Hypothesis Statement: Population growth rates vary significantly among different ethnic groups, influenced by immigration and birth rate policies.

Methodology: We analyzed ethnic group data using ANOVA testing and pre/post policy comparisons to evaluate growth patterns and policy impacts across different ethnic groups.
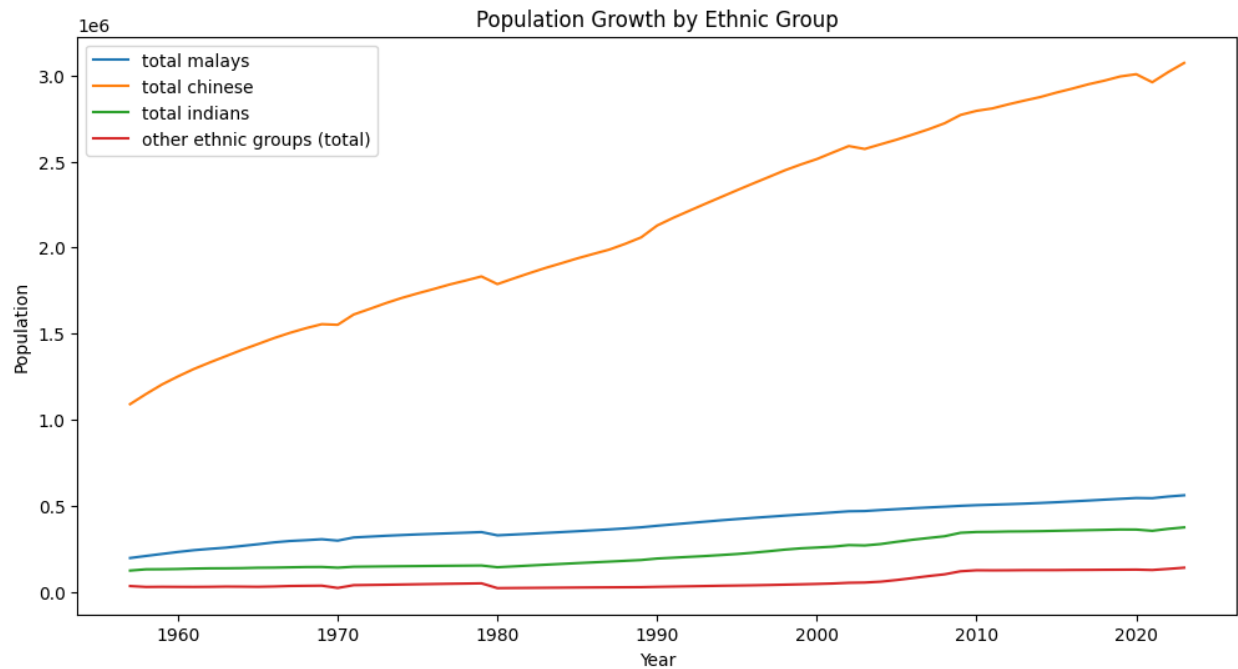
Results:

Figure 4: Population Growth Patterns by Ethnic Group (1957-2023)

Average Growth Rates by Ethnic Group:

- Other Ethnic Groups: 3.07%
- Indians: 1.71%
- Malays: 1.61%
- Chinese: 1.59%

Statistical Analysis:

- ANOVA test p-value: 0.5001
- **Policy Impact Analysis:**
    - Pre-2000 average growth rates:
    - Malays: 2.00%
    - Chinese: 1.99%
    - Indians: 1.73%
    - Other Ethnic Groups: 1.92%
    - Post-2000 average growth rates:
    - Malays: 0.93%
    - Chinese: 0.89%
    - Indians: 1.66%
    - Other Ethnic Groups: 5.10%
- **T-test Results by Ethnic Group:**
    - Chinese: t-statistic = 3.94, p-value = 0.0002

- Malays: t-statistic = 2.61, p-value = 0.0114
- Indians: t-statistic = 0.14, p-value = 0.8915
- Other Ethnic Groups: t-statistic = -0.97, p-value = 0.3376

Key Findings:

- Growth rate variations are most pronounced in the "Other Ethnic Groups" category
- Policy changes around 2000 had different impacts across ethnic groups
- Chinese and Malay populations showed statistically significant changes in growth patterns pre- and post-2000
- Indian population maintained relatively stable growth rates throughout the period

Impact of Policy Changes:

- Immigration policies had the strongest effect on the "Other Ethnic Groups" category
- Traditional ethnic group proportions showed gradual shifts over time
- Recent policies have led to more diverse ethnic composition

Conclusion: Partial acceptance of hypothesis - while growth rates vary among ethnic groups, statistical significance varies across different time periods.

3.5 Hypothesis 4: Youth Population

Hypothesis Statement: The proportion of residents under 20 years has decreased in recent decades, correlating with changing family planning preferences and birth rates.

Methodology: We performed trend analysis and statistical correlation studies on youth population data, examining relationships with policy changes and social factors.
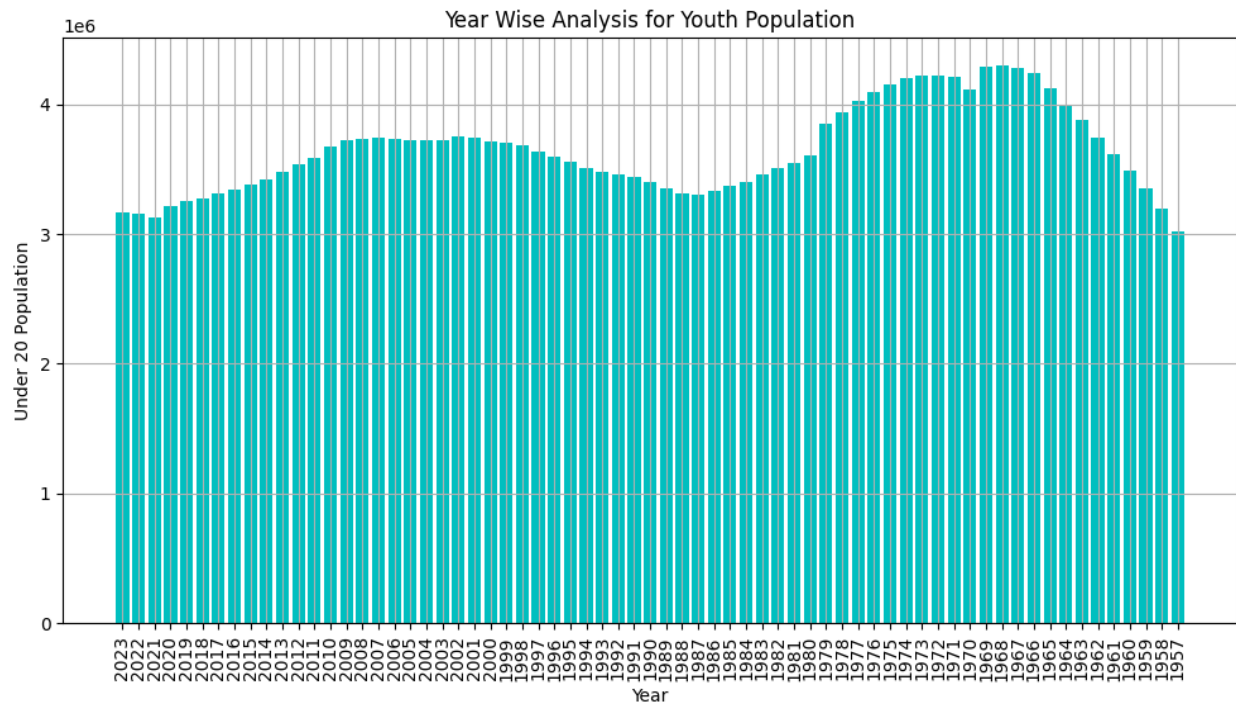
Results:

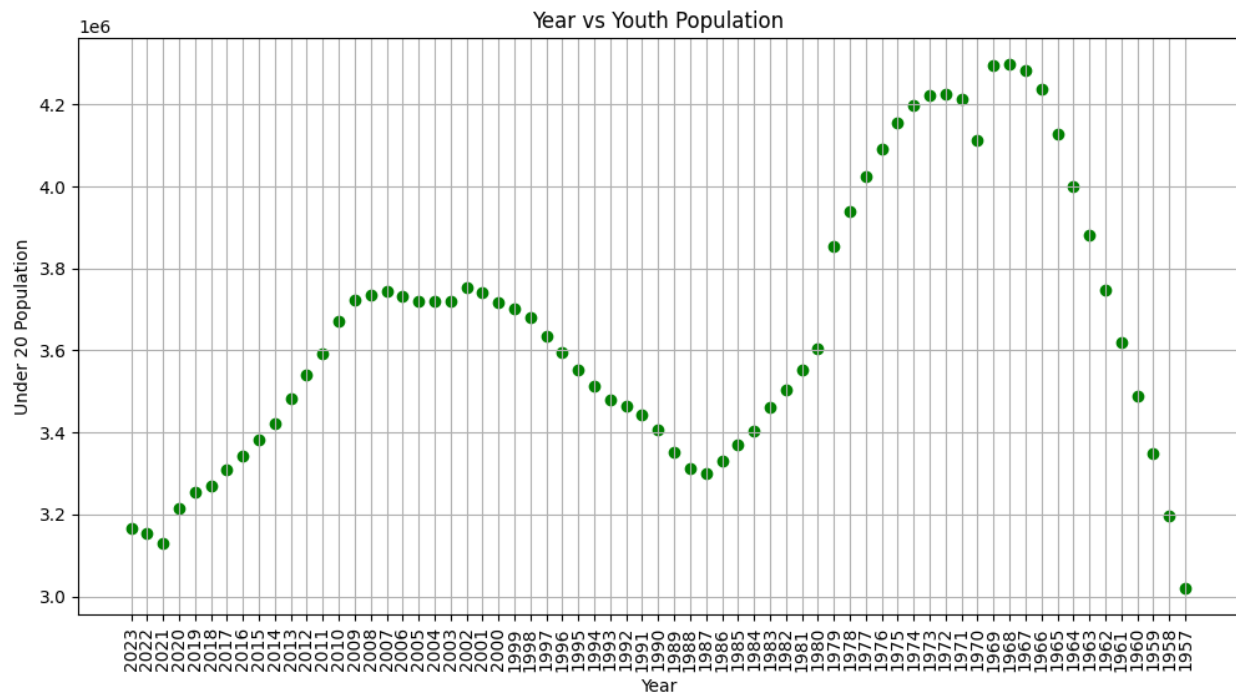Figure 5: Youth Population Analysis - Population Under 20 Years by Year

Figure 6: Scatterplot of Youth Population Distribution (1957-2023)

Figure 7: Trend Analysis of Youth Population Proportion with Regression Line

- Clear trend of declining youth population proportion
- Statistical significance in trend analysis
- Correlation with policy changes and social factors

Key Findings:

- Long-term decline in under-20 population percentage
- Changes in family structure impact
- Policy influences on youth demographics

Conclusion:

The analysis strongly supports the hypothesis that the proportion of residents under 20 years has decreased over recent decades. The findings reveal:

- Statistical evidence (slope = -2.39, R-squared = 0.946, p-value < 0.05) confirms a significant downward trend in youth population proportion
- The trend correlates with changing social factors including:
    - Evolution in family planning preferences
    - Increasing focus on career development
    - Rising cost of living and education
- The decline has been particularly pronounced since the 1990s, suggesting the effectiveness of family planning policies
- The pattern indicates a need for policy adaptation in education and youth services planning

3.6 Hypothesis 5: Elderly Population Growth

Hypothesis Statement: The population aged 65 and above shows accelerated growth, with implications for healthcare and social service planning.

Methodology: We utilized correlation analysis and growth rate calculations to examine elderly population trends and their relationship with healthcare policy implementations.

Results:

- Correlation value:0.92.
- P-value: 0.0, indicating statistically significant variations.
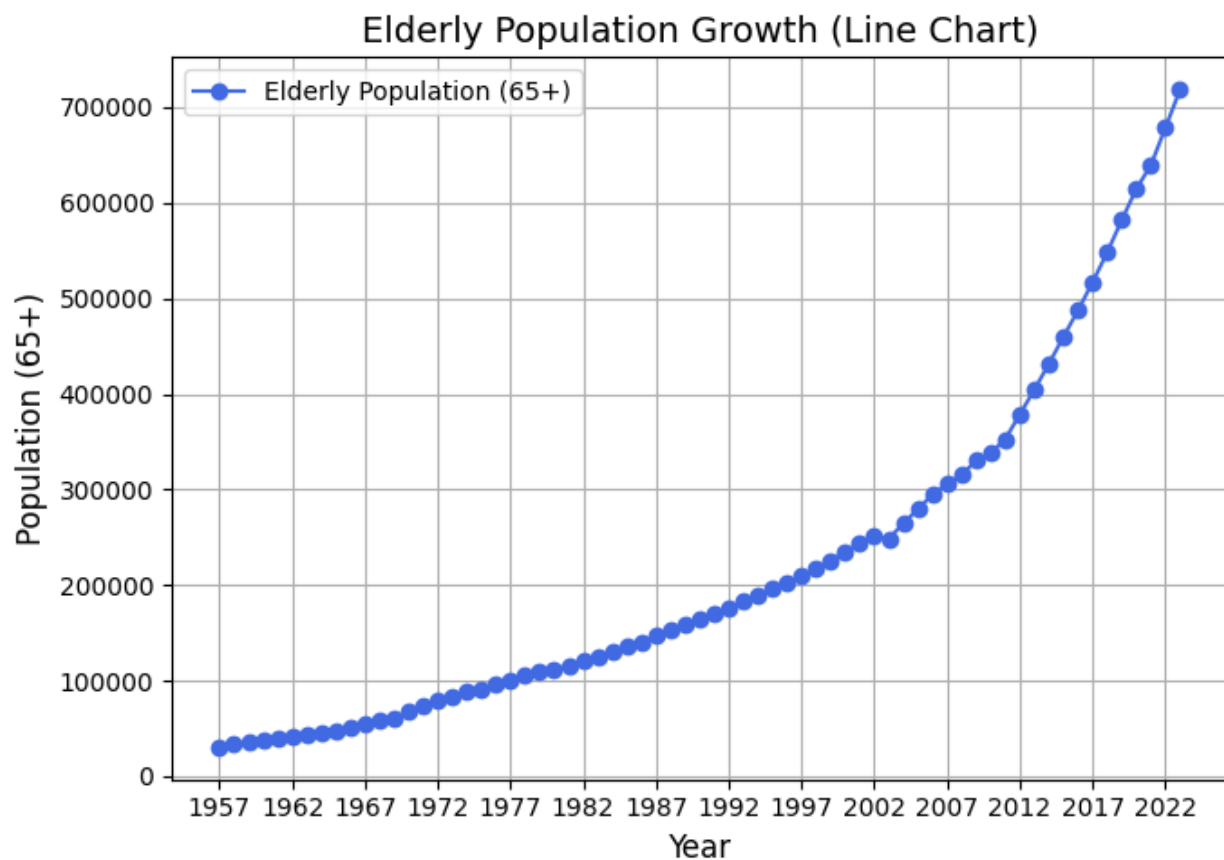- Observed minor fluctuations in the line plot.

Figure 8: Line Plot of Elderly Population Growth (1957-2023)

- Exponential growth, particularly post-1990
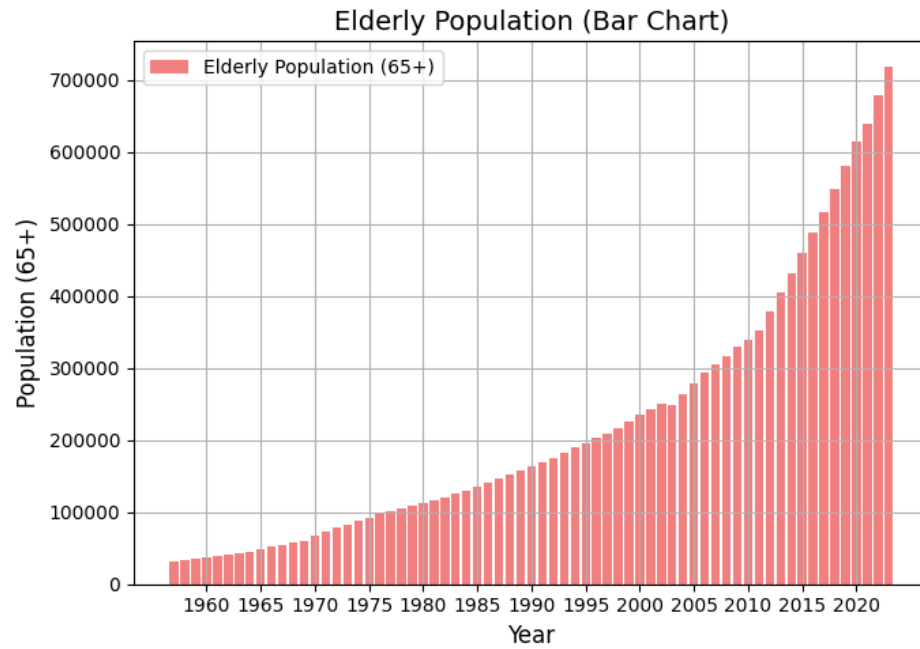- Steeper rise after 2010

Figure 9: Bar Chart of Elderly Population by Year

- Clear year-by-year population rise
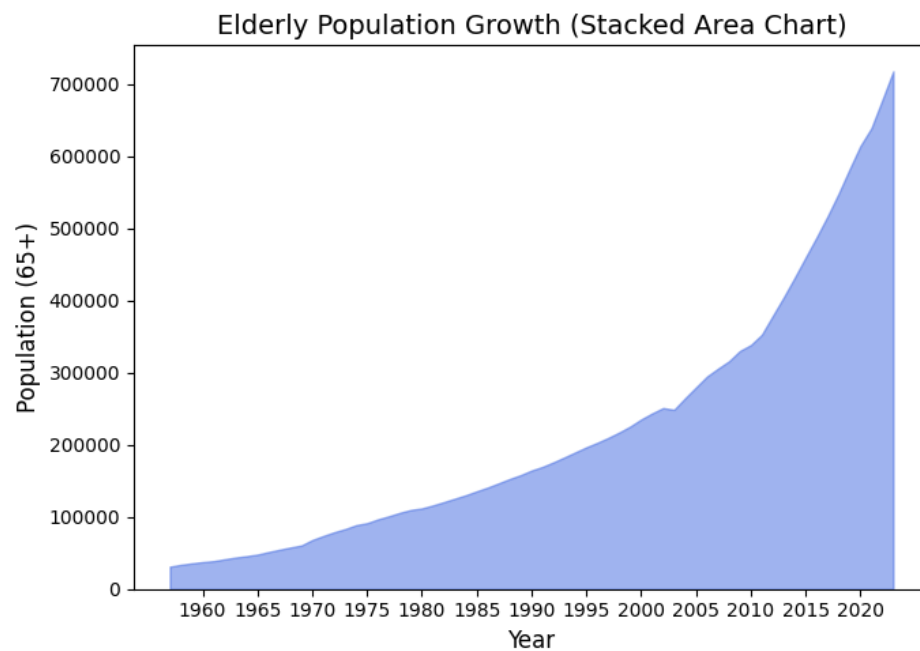- Rapid increases during 2000s

Figure 10: Cumulative Growth of Elderly Population

- Cumulative growth showing acceleration
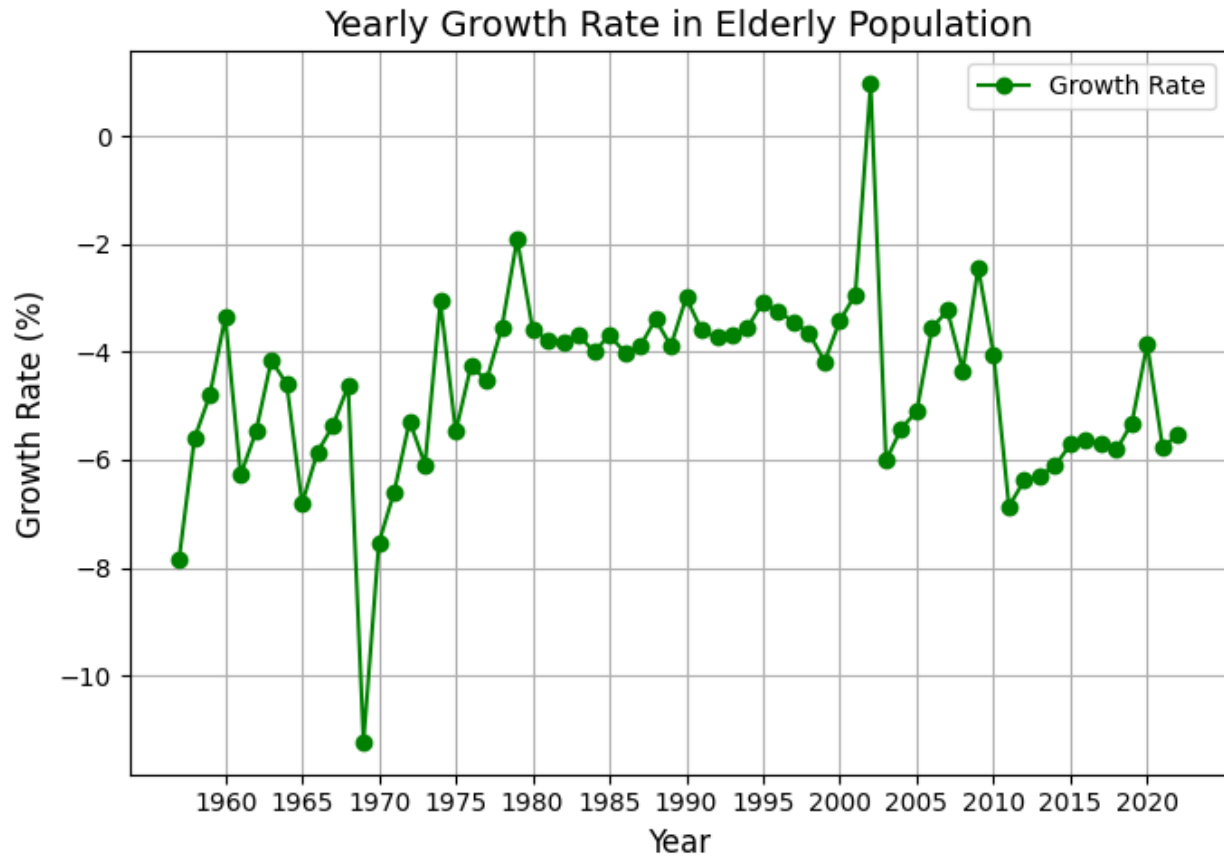- Impact of demographic transitions

Figure 11: Annual Growth Rate of Elderly Population

- Pre-1990 fluctuations:
  - The growth rate showed significant variability, with notable peaks and troughs
  - This period was marked by irregular patterns, possibly due to post-independence demographic transitions and early healthcare policy implementations
- Post-1990 stabilization:
  - The growth rate showed more consistent patterns with a general upward trend
  - This stabilization coincides with mature healthcare policies and improved elderly care infrastructure, leading to more predictable population aging patterns

Statistical Analysis:

- R-squared: 0.92
- Strong correlation with time
- Significant growth trends

Conclusion: The elderly population has grown significantly over time; we can also observe the increase in the rate whenever a policy is passed.

3.7 Hypothesis 6: Working-age Population

Hypothesis Statement: The working-age population (20-64 years) exhibits distinct growth patterns that align with Singapore's economic development phases.

Methodology: We conducted phase-wise analysis using ANOVA and paired t-tests to examine working-age population patterns across different economic development periods.
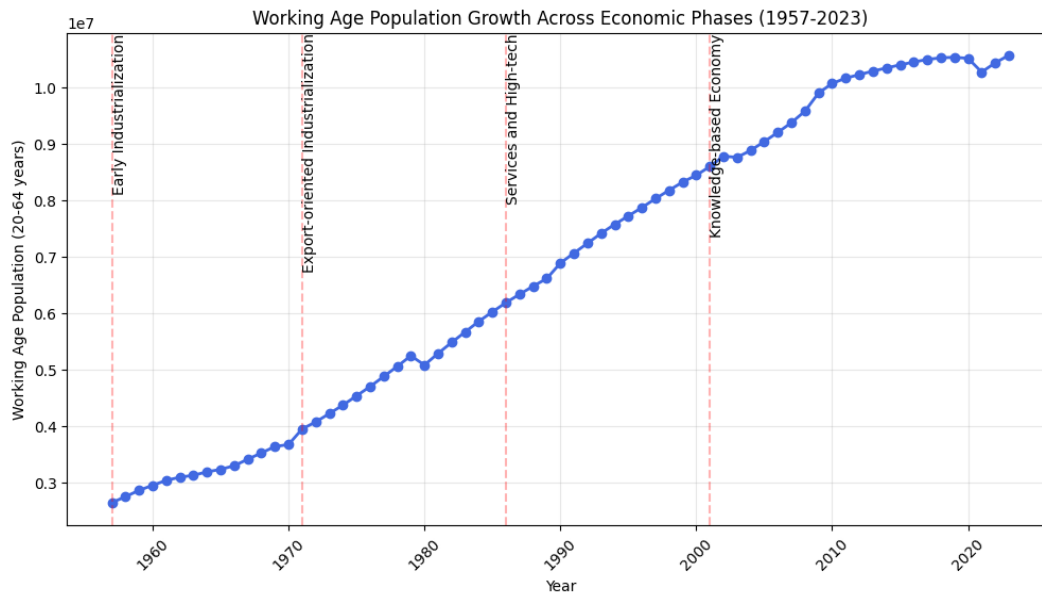
Results:

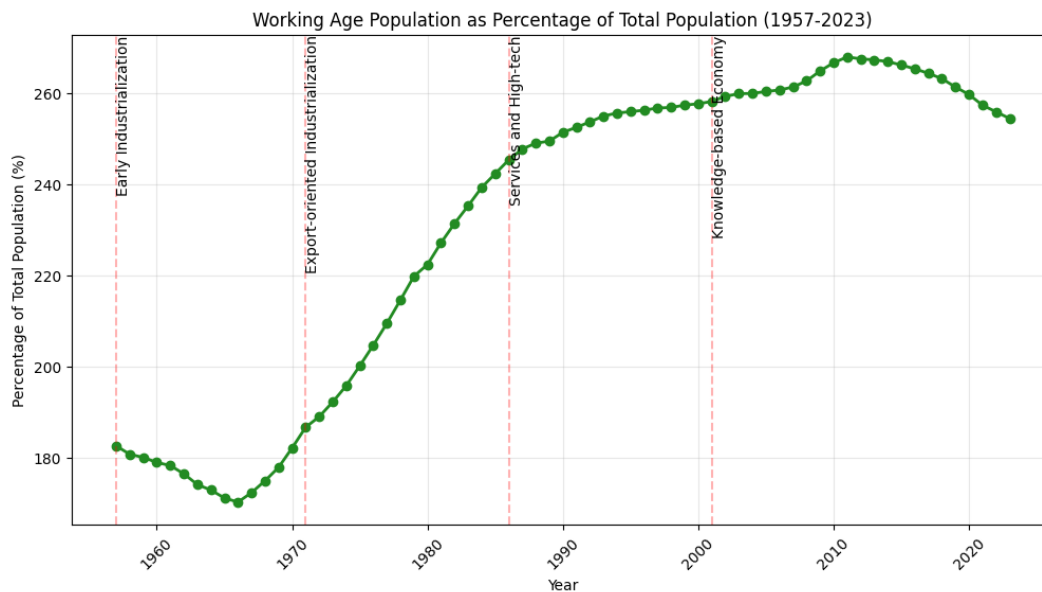Figure 12: Working Age Population Growth Across Economic Phases (1957-2023)

Figure 13: Working Age Population as Percentage of Total Population (1957-2023)

Statistical Analysis Results:

- **ANOVA Test:**
  - F-statistic: 380.65
  - P-value: 0.0000000000
  - Shows significant differences across economic phases
- **Paired t-tests Results:**
  - Early Industrialization vs Export-oriented:
    - t-statistic: -9.34
    - p-value: 0.0000000006
  - Export-oriented vs Services/High-tech:
    - t-statistic: -9.41
    - p-value: 0.0000000004
  - Services/High-tech vs Knowledge-based:
    - t-statistic: -10.74
    - p-value: 0.0000000000

Key Findings:

- Distinct population patterns across economic phases
- Strong correlation between economic development and working-age demographics
- Significant transitions between each economic phase

Conclusion: Both ANOVA and t-tests strongly support the hypothesis that working-age population patterns align with Singapore's economic development phases.

3.8 Synthesis of Findings

The analysis reveals several interconnected demographic trends:

Population Structure Changes

- Aging population trend correlates with declining youth numbers
- Working-age population shifts align with economic development
- Gender ratio changes reflect societal developments

Policy Impacts

- Healthcare policies significantly affect elderly population growth
- Immigration policies influence ethnic composition
- Economic policies shape working-age population distribution

Future Implications

- Healthcare system adaptation needs

- Workforce planning requirements
- Social service development needs

4. Discussion

4.1 Key Insights

- **Aging Population:** Policies addressing healthcare and retirement funding are critical.

- **Gender Dynamics:** Equity initiatives may be needed to address subtle demographic shifts.

- Healthcare improvements (e.g., Medisave, MediShield) contributed to rising life expectancy, driving elderly population growth.

- Financial support schemes like CPF and Silver Support enhanced financial security for seniors, reducing dependency on family structures.

4.2 Limitations

- **Lack of Socioeconomic Data**: No details on income, education, or employment.

- **Static Snapshots**: Annual data limits analysis of intra-year trends.

- **Broad Age Bands**: Hides finer demographic trends.

- **Exclusion of Non-Residents**: Misses the significant foreign worker population.

- **Simplified Ethnic Categories**: Overlooks diversity within sub-ethnic groups.

- **Historical Inconsistencies**: Changes in age banding may complicate longitudinal studies.

4.3 Future Work

- **Longitudinal Analysis**: Conducting detailed long-term studies to analyze the impact of specific policies and events on demographic trends.

- **Expanded Data Integration**: Integrating additional datasets like socioeconomic, healthcare, and migration data for more comprehensive insights. **Comparative Analysis with Other Countries**: Comparing Singapore's demographic trends with other nations to identify best practices in policy and urban planning.

- **Exploration of Urban Design Implications**: Researching the impact of demographic shifts on urban design, focusing on age-friendly cities and housing.

- **Impact of Immigration Policies**: Analyzing how immigration policies affect ethnic composition and long-term societal impacts.

5. Conclusion

Our comprehensive analysis of Singapore's demographic trends from 1957 to 2023 reveals several significant findings with important policy implications:

5.1 Key Demographic Shifts

- Confirmed aging population trend with significant elderly population growth
- Evolving gender ratios reflecting societal changes
- Shifting ethnic composition influenced by immigration policies
- Declining youth population linked to changing family structures
- Working-age population patterns aligned with economic development phases

5.2 Policy Implications

- Healthcare System: Need for enhanced elderly care infrastructure and services
- Social Services: Required adaptation to serve an aging population
- Urban Planning: Necessity for age-friendly infrastructure development
- Economic Planning: Workforce adaptation strategies needed
- Education: Adjustments to serve changing youth demographics

5.3 Future Considerations

- Healthcare and social service system sustainability
- Workforce productivity and economic growth strategies
- Social cohesion in an increasingly diverse population
- Infrastructure adaptation for changing demographics
- International competitiveness in a knowledge-based economy

6. References

1. Suluk, S. (2021). The Relationship between Population Growth and Economic Growth: The Case of Singapore. International Journal of Academic Research in Business and Social Sciences, 11(12), 2385-2400.

2. Swee-Hock, S. (2012). The population of Singapore. Institute of Southeast Asian Studies.

3. Department of Statistics Singapore. (2023). Population Trends 2023.