

# Statistics Singapore Newsletter



ISSUE 1, 2023

## High Growth Firms in Singapore

by Aldric Han, Eric Ng and Tan Xue Lin

Business Statistics Division

Singapore Department of Statistics

### Introduction

International studies<sup>1</sup> found that High-Growth Firms (HGF) are significant contributors to revenue and employment growths, and they tend to outperform non-HGF. HGF are also more likely to invest in R&D and innovation, generating spillover effects that improve productivity in other firms.

In Singapore, policies have been made to support high-potential firms to grow and become globally competitive, which would contribute to economic growth and job creation. One example is the EnterpriseSG's Scale-Up programme which aims to accelerate the growth of high-potential firms, help them build strong brand recognition and become global champions.

Against this backdrop, DOS conducted a study to examine the characteristics, prevalence, and performance of HGF in Singapore between 2016 and 2019, using an integrated firm-level longitudinal dataset.

### Definition

In this study, HGF are defined as firms with average annualised revenue growth exceeding 10 per cent over a three-year period<sup>2</sup> and with at least 10 employees<sup>3</sup> at the start of the period:

$$\sqrt[3]{\frac{\text{revenue}_{t+3}}{\text{revenue}_t}} - 1 > 0.1$$

where  $\text{revenue}_{t+3}$  denotes revenue at the end of the three-year period and  $\text{revenue}_t$  denotes revenue at the start of the three-year period. Firms are tagged as either HGF or non-HGF at the end of the three-year period. Consequently, firms with at least 10 employees in 2016 and registered more than 10 per cent annualised revenue growth between 2016 and 2019 are deemed as HGF.

### Key Findings

There were about 7,800 HGF, accounting for 19 per cent of firms with at least 10 employees in 2016.

#### Majority of HGF Were Smaller and More Than Five Years Old

A majority of these HGF were smaller firms in terms of revenue or employee size. Nearly half (48 per cent) had revenue exceeding \$1 million but up to \$10 million and 28 per cent had revenue not exceeding \$1 million (Chart 1).

Only 6 per cent had revenue exceeding \$100 million. In terms of employee size, 55 per cent employed between 10 to 25 workers while only 5 per cent employed more than 200 employees.

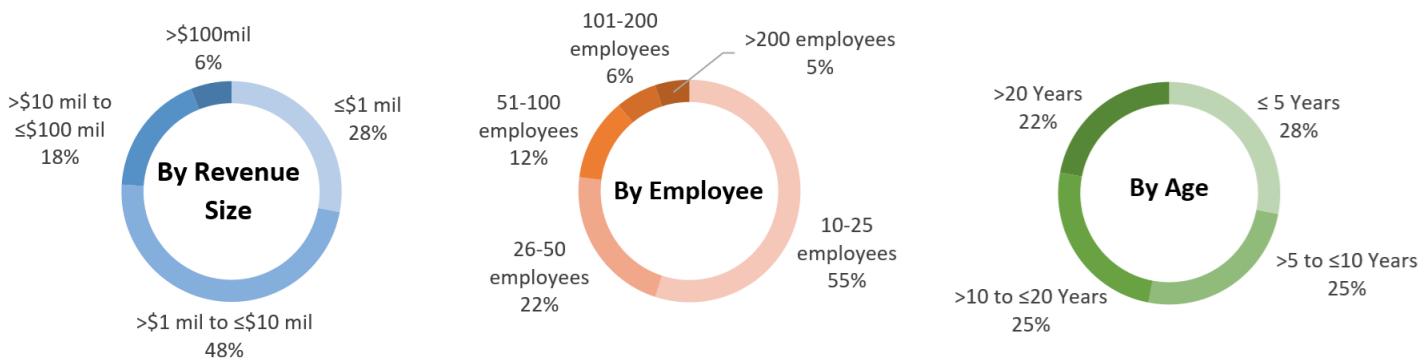
Half of these HGF were aged 5 - 20 years old while 22 per cent were aged over 20 years. About 28 per cent were aged 5 years or below.

1 For more information, refer to the References appended at the end of the article.

2 This is equivalent to at least 33.1 per cent cumulative growth after 3 years. The criterion used is similar to that adopted by Eurostat's and OECD's research studies, although there have been varied definitions used in other studies.

3 The threshold of a minimum of 10 employees is meant to mitigate small firm bias, where a small increase in revenue results in disproportionately high revenue growth due to the small denominator.

**Chart 1**  
**PROFILE OF FIRMS WITH HIGH GROWTH BETWEEN 2016 AND 2019<sup>4</sup>**

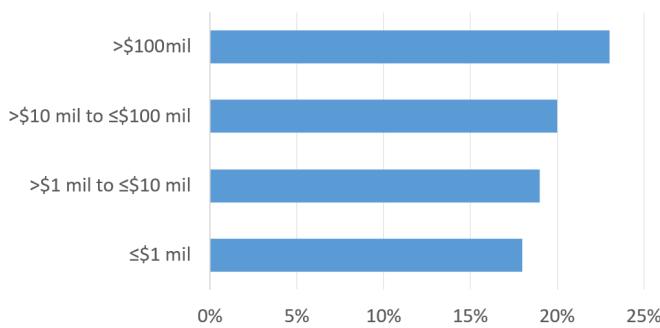


### HGF Were Likely to Appear Across Different Revenue and Employee Size Bands

Across revenue size bands, HGF were most prevalent among firms with revenue exceeding \$100 million (Chart 2).

**Chart 2**

**PREVALENCE OF FIRMS WITH HIGH GROWTH  
BETWEEN 2016 AND 2019,  
BY REVENUE SIZE IN 2016**

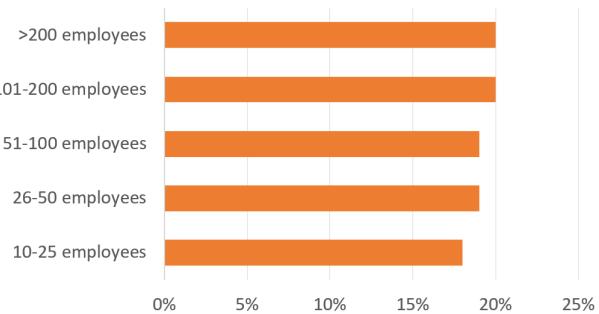


Among firms with revenue not exceeding \$1 million in 2016, 18 per cent registered high growth between 2016 and 2019. The proportion of HGF increased to 20 per cent among firms with revenue over \$10 million and not exceeding \$100 million, and further increased to 23 per cent among larger firms with revenue exceeding \$100 million.

The prevalence of HGF was generally similar across different employee size bands. Among firms with employee size of no more than 25 in 2016, 18 per cent experienced high revenue growth between 2016 and 2019. The proportion of HGF increased marginally to 20 per cent for firms with employee size exceeding 100 (Chart 3).

**Chart 3**

**PREVALENCE OF FIRMS WITH HIGH GROWTH  
BETWEEN 2016 AND 2019,  
BY EMPLOYEE SIZE IN 2016**

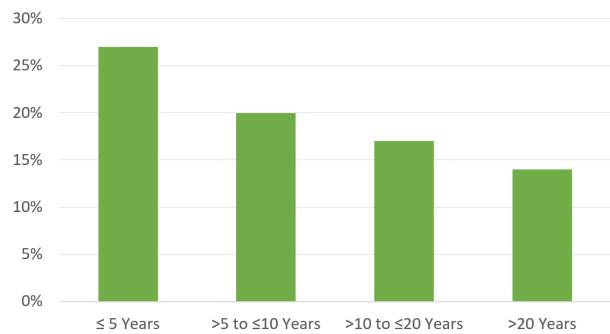


### Younger Firms Were More Likely to Experience High Growth

The prevalence of HGF was higher among younger firms compared to older ones. Among firms that were no more than 5 years old in 2016, 27 per cent registered high growth between 2016 and 2019. This proportion declined progressively with older firms and was 14 per cent among firms over 20 years old (Chart 4).

**Chart 4**

**PREVALENCE OF FIRMS WITH HIGH GROWTH  
BETWEEN 2016 AND 2019, BY AGE IN 2016**



<sup>4</sup> The respective categories refer to the revenue size, employee size and age of firms in 2016, which is the start of the revenue growth period between 2016 and 2019.

## HGF Outperformed Non-HGF

On average, HGF generated higher revenue and employed more workers as compared to firms with lower growth. The average revenue of HGF was more than 5 times the revenue of non-HGF (Table 1).

**Table I**

**AVERAGE REVENUE, EMPLOYEES,  
PRODUCTIVITY AND WAGE PER EMPLOYEE  
OF HGF AND NON-HGF, 2019**

Average Statistics	HGF	Non-HGF
Revenue (\$mil)	257	49
Employees (No.)	90	56
Productivity (\$'000)	233	105
Annual Wage Per Employee (\$'000)	67	52

The average employee size of HGF was 90, which was larger than the non-HGF's average employee size of 56.

HGF were more productive than non-HGF, with the average productivity<sup>5</sup> of HGF more than double of the non-HGF. The average annual wage per employee<sup>6</sup> of HGF was also higher than non-HGF.

## Conclusion

The findings showed that close to one in five firms with at least 10 employees in 2016 were HGF in 2019. While HGF can be found among firms of different sizes and ages, a majority of HGF were smaller in terms of revenue or employee size. Younger firms were also more likely to become HGF.

HGF tended to outperform non-HGF with higher productivity, generate higher revenue and employ more workers. This suggests that there are economic benefits to supporting firms with high-growth potential.

Further research may be undertaken to study if these HGF are able to sustain their growth potential over a longer period of time.

## References

Australian Department of Industry, Innovation and Science (2018)  
*What Drives High-Growth? Characteristics of Australian Firms*

EuroStat (2020)  
*1 in 10 EU Enterprises Classified as High Growth*

Organisation for Economic Co-operation and Development (2021)  
*Understanding Firm Growth: Helping SMEs Scale Up*

Statistics Canada (2020)  
*High-Growth Firm Characteristics in Canada*

World Bank (2019)  
*High-Growth Firms: Facts, Fiction and Policy Options for Emerging Economies*



<sup>5</sup> Productivity is measured as nominal value-added per worker.

<sup>6</sup> Wage per employee is derived as total wages divided by total number of employees.

# Experimental Uses of Machine Learning and New Data Sources in Updating the Statistical Business Register

by Chee Rong Can and Peh Li Lin  
 Business Statistics Division  
 Singapore Department of Statistics

## Introduction

The Singapore Department of Statistics (DOS) manages the Statistical Business Register (SBR), which serves as the foundational statistical database for the compilation of business and economic statistics.

DOS's SBR provides a comprehensive coverage of the economic units for survey frame production used in sample selections for business surveys, facilitates the compilation of business indicators and supports in-depth analysis for policy insights.

In response to increasingly complex data demands, DOS has undertaken initiatives to transform and acquire innovative and new capabilities across the data value chain. This article shares DOS's experiences in leveraging machine learning (ML) and artificial intelligence (AI) techniques to enhance data availability and update the SBR.

## Coverage of the SBR

The SBR covers all entities registered in Singapore. These comprise companies and businesses (including sole proprietorships and partnerships) registered with the Accounting and Corporate Regulatory Authority

(ACRA) of Singapore, charities and societies registered with the Ministry of Culture, Community and Youth (MCCY) and the Registry of Societies (ROS) respectively, and other entities registered with their respective registration authorities<sup>1</sup>.

Basic identification (e.g., Unique Entity Number (UEN), enterprise name) and enterprise characteristics (e.g., registration date, status) are readily available from the various administrative registration authorities.

Enterprises in the SBR are uniquely identified by their UENs, which are issued to them upon successful registration in Singapore. Enterprises use UENs in their interactions with the Government, such as the application of business licenses and permits, as well as filing of tax returns. It follows that the UEN enables DOS to process and integrate enterprise-level data from various sources efficiently and accurately.

## Data Sources and Maintenance of the SBR

Administrative data are the primary sources of the SBR because of their comprehensive coverage of Singapore-registered enterprises. Table 1 presents examples of administrative data that are regularly used to update the SBR.

**TABLE I**  
**EXAMPLES OF ADMINISTRATIVE DATA USED TO UPDATE THE SBR**

Source Agency	Administrative Data
Accounting and Corporate Regulatory Authority (ACRA) of Singapore	<ul style="list-style-type: none"> <li>■ Identification and Basic Enterprise Information (e.g., UEN, Enterprise Name, Registration Date, Shareholder Information, Industrial Classification)</li> <li>■ Financial Information (e.g., Revenue, Profit)</li> </ul>
Inland Revenue Authority of Singapore (IRAS)	Financial Information
Ministry of Manpower (MOM) Central Provident Fund Board (CPF)	Employment and Wages
Singapore Customs (CUSTOMS) Enterprise Singapore (ESG)	Merchandise Trade (i.e., Imports, Exports)

<sup>1</sup> More information on the UEN registration authorities can be found [here](#).

## Challenges with Using Administrative Data

DOS has been collaborating with data source agencies on the use of administrative data for statistical and analytical purposes as well as the coordination and streamlining of operational processes:

1. *Changes in administrative systems:* For example, agencies would notify DOS in advance of forthcoming data changes such as revisions in administrative filing requirements. This would allow DOS time to assess the potential impact of the data changes and implement necessary measures to minimise disruption to statistical production.
2. *Data consistency and quality:* DOS also provides regular feedback and suggestions on data quality improvement and assists data source agencies in data capability building via knowledge sharing. Through such communication and interactions, DOS and data partner agencies develop a shared understanding and mutual trust to optimise the use of administrative data for statistical and analytical purposes.
3. *Timeliness and coverage:* While administrative data are the primary data sources for the SBR, some may not be timely (e.g., corporate tax filings are only available 1-2 years after an enterprise's financial year ending), readily available in machine readable format or are unavailable. Therefore, data collected through surveys conducted by DOS or other government agencies are also used to supplement the updating of the SBR. Nonetheless, in view of respondent burden, it is not feasible to continually leverage surveys to collect data that are not available from administrative sources.

## Experimental Uses of Machine Learning and New Data Sources

In response to these challenges and the need to meet increasingly complex data demands, DOS is building its capabilities to tap on new opportunities including big data and advances in technology such as AI and ML.

Big data offer potential advantages of higher data frequency, greater granularity as well as lower data collection cost. Big data also provide additional

information not available in the existing administrative data and surveys.

Using AI and ML, DOS can extract and process such new data sources efficiently. The next section showcases two pilot projects that used AI and ML techniques to improve data availability in the SBR.

### (I) Web-Based Data Sources to Profile Enterprises with Internet Presence

Over the last few decades, the internet provided growth opportunities for many enterprises. However, information on whether enterprises have an internet presence is not available from administrative sources.

Hence, DOS undertook a pilot project to text mine web-based data and use supervised ML to study how enterprises in Singapore utilise their corporate websites. In this project, enterprises are broadly classified into three major categories according to their internet presence and corresponding usage (Table 2).

**Table 2**  
CATEGORISATION OF ENTERPRISES  
ACCORDING TO THEIR INTERNET PRESENCE

Internet Category	Definition
A	Enterprises without websites
B	Enterprises with websites/ online presence but do not generate income directly from their websites <u>Example</u> Websites with information on products/ services
C	Enterprises which generate income directly from their websites <u>Example</u> Online retail stores where customers can place orders directly

The target population was first identified through the enterprise information available in the SBR. The Uniform Resource Locators (URLs) or the website addresses of these enterprises were then gathered from various sources such as surveys, administrative data, online directory, and the Singapore Network Information Centre (SGNIC) which is the domain registry of website addresses ending with ".sg".

These information on enterprises' URLs were subsequently merged with the target population to generate a web crawling list. Web scraping technique was then applied to extract selected features from the website addresses on the crawling list.

To minimise burden on the websites, DOS only scrapped each domain once a year and included idle time when scrapping the pages within each domain.

Some examples of the extracted features include whether the website displays information on products and services or any online shopping facility. With the extracted features, a supervised ML classifier algorithm was applied to classify the enterprises into different categories of internet presence (Figure 1).

The indicator on enterprises' internet presence can be integrated with enterprise characteristics (e.g., economic activity, enterprise's age) available in the SBR to derive new insights for further analysis.

This pilot project demonstrated the feasibility of text mining web-based data and using ML to derive new indicators, which can be used to enhance the information in the SBR.

Compared to the traditional way of conducting a survey to obtain the data, this approach costs less and does not impose burden on survey respondents.

## (2) Leveraging AI for Data Extraction of Unstructured Data from Financial Statements

While DOS has been relying on structured data to update the SBR and for statistical compilation, unstructured data from enterprises' financial statements are also a rich source of financial information and new insights.

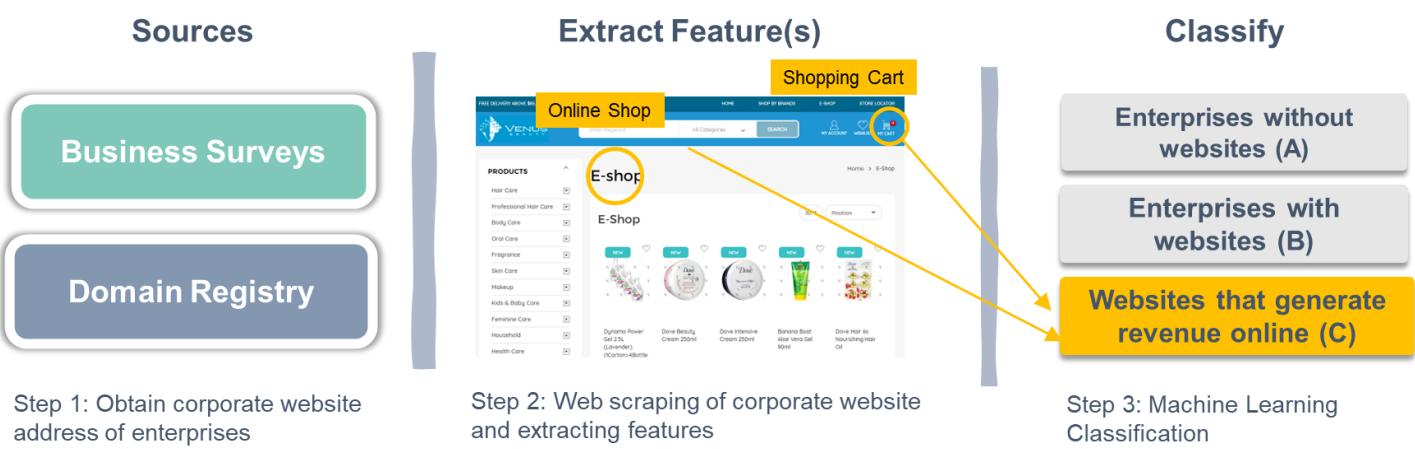
However, considerable manual effort is required to read, analyse and extract the relevant information. This curtails the number of financial statements and data points that can be captured and processed for statistical use.

To overcome this limitation, DOS is developing AI capabilities for data extraction and processing of unstructured data from enterprises' financial statements.

Advanced semantic and reasoning algorithms are used to automatically identify, extract, cleanse and validate the required information from financial statements. The AI model is developed based on training datasets (i.e., a small set of financial statements) and deployed for data extraction from a large volume of financial statements.

A Proof-of-Concept (PoC) of the AI solution, co-developed with a commercial AI solution provider,

**Figure 1**  
WEB SCRAPING AND ML  
FOR CLASSIFICATION OF ENTERPRISES WITH INTERNET PRESENCE



was conducted by DOS to extract data items such as the type of fixed assets, names of overseas subsidiaries and ultimate shareholders from financial statements. The AI solution successfully extracted the required data from the unstructured information in the financial statements with reasonable accuracy.

The information extracted by AI can be used to supplement existing financial information in the SBR for compilation of economic and business indicators. Two examples are presented below:

- Detailed assets information extracted from the notes of the financial statements can be used to support in-depth analysis on enterprises' asset structure and investment (Figure 2).

**Figure 2**  
**DETAILED ASSETS INFORMATION<sup>2</sup>**

Non-current assets	Note	Group		Company	
		2020 \$'000	2019 \$'000	2020 \$'000	2019 \$'000
Property, plant and equipment	4	1,929	2,165	247	213
Intangible assets	5	769	990	637	773
Investment properties	6	2,730	2,981	—	—
Subsidiaries	7	—	—	86,663	86,163
Other investments	9	18,819	25,096	54	14
Loans, advances, hire purchase and leasing receivables	10	82,332	83,092	75,837	69,368
Deferred tax assets	12	3,692	3,856	—	—
Right-of-use assets	38	2,525	2,839	1,834	2,020
		112,796	121,019	165,272	158,551

Extract value '247000' based on interpretation of column names (i.e., year 2020 and units ('000) and row name (i.e., Property, plant and equipment)

- More detailed shareholding information supplement existing machine-readable data available in the SBR, for ownership analysis (Figure 3).

**Figure 3**  
**SHAREHOLDING INFORMATION**

As at 31 December 2019, the Company's immediate holding company is AB Limited, a company incorporated in the Republic of Singapore. The Company's intermediate holding company is ABO, a company incorporated in Denmark, and the ultimate holding company is AB Foundation, an enterprise foundation registered in Denmark.

Extract name and country of the immediate, intermediate and ultimate companies (highlighted)

The experience and knowledge gained in the PoC helped DOS plan and scale up the actual implementation of the AI solution that will be rolled out in production by end 2023. The new AI capability enables DOS to improve operational

processes in data collection, statistical editing, imputing and estimation, and ensure that more detailed data are available for analyses.

## Conclusion

The SBR plays an important role in the data collection, compilation and analysis of economic and business statistics.

While administrative data remain the primary sources for the SBR, new data sources such as big data and unstructured data are becoming increasingly important. Leveraging AI and ML opens up newer data sources for DOS to enhance the information in the SBR.

## References

**EuroStat (2021)**  
*European Business Statistics Methodological Manual for Statistical Business Registers*

**Singapore Department of Statistics (2022)**  
*Using Big Data to Profile Singapore's Internet Economy*

**Statistics Netherlands (2020)**  
*Measuring the Internet Economy with Big Data*

**United Nations (2020)**  
*United Nations Guidelines on Statistical Business Registers*

**United Nations Economic Commission for Europe (2015)**  
*Guidelines on Statistical Business Registers*

**United Nations Economic Commission for Europe (2021)**  
*Machine Learning for Official Statistics*



<sup>2</sup> Information is sourced from the annual report made available on an enterprise's corporate website.

# The Retirement and Health Study - Management of a Large-Scale Longitudinal Study in Singapore

by Jocelyn Ow, Neo Ming Yuan and Koh Phuay Leng  
Central Provident Fund Board

## Introduction

With one of the most rapidly ageing populations globally, it is imperative to ensure Singapore's social, healthcare, financial, and other systems evolve to cater to the needs of changing demographics.

Timely and pertinent research on retirement planning, health status, income, and living arrangements helps the government and relevant organisations understand the needs and challenges faced by an ageing population, and informs the development of suitable programmes to support the well-being of seniors.

In 2014, Singapore's first large-scale integrated longitudinal study on residents' retirement and healthcare needs, the Retirement and Health Study (RHS), was launched to facilitate evidence-based decision-making.

The RHS is jointly conducted by the Central Provident Fund Board (CPFB), the Housing and Development Board (HDB), the Ministry of Finance (MOF), the Ministry of Health (MOH) and the Ministry of Manpower (MOM), with the Singapore Department of Statistics (DOS) providing technical advisory.

Modelled after other similar studies overseas (e.g., the Health and Retirement Study in the United States, the English Longitudinal Study of Ageing in the United Kingdom), the multi-disciplinary RHS examines information on various aspects of ageing in Singapore to gain a deeper understanding of the related needs of seniors.

## Survey Coverage

The survey covers a sample of Singapore residents aged 45 years and over who are interviewed once every two years.

The sample is selected based on a stratified design, with the sampling frame divided into different groups (or strata) according to age group, sex and race. Systematic sampling by broad dwelling type is then performed within each stratum. Younger cohorts are added periodically to ensure that the RHS sample continues to represent the Singapore resident population aged 45 years and over.

The main unit of analysis is the Main Respondent who goes through a 1.5 hour interview. The Main Respondent's spouse or partner is also invited to participate in a separate 30 minute interview on selected topics from the Main RHS Questionnaire (Table 1).

**TABLE I**  
**TOPICS COVERED**  
**IN THE LATEST RHS QUESTIONNAIRES**

### Topics in Main RHS Questionnaire

Demographics	}	Topics in a separate questionnaire for Spouse or Partner
Health Status		
Employment and Retirement		
Assets and Liabilities		
Financial Transfers and Other Income		
Financial Planning Behaviour		
Household Expenditure (Non-healthcare)		
Healthcare Insurance Plans and Schemes		
Healthcare Utilisation and Expenditure		
Health Risk Factors, Social Connectedness and Lifestyle		
Cognitive Status		

To date, four waves of the RHS have been completed, and Wave 5 is in progress. Around 11,700 to 15,100 Main Respondent interviews were completed in each of the past four waves. The longitudinal all-waves response rate for Main Respondent is 60 per cent in Wave 4, with 8,200 Main Respondents who participated in all four waves (Table 2). The response rate for Spouse/ Partner is around 70 per cent from Waves 1 to 4 (Table 3).

The next section provides an overview of the strategies adopted by the Retirement and Health Study Office (RHSO) in the areas of questionnaire design, panel management and data quality assurance, to achieve a robust longitudinal panel with quality data across waves.

**TABLE 2**  
RESPONSE RATES FOR MAIN RESPONDENTS BY RHS WAVE

RHS Wave	Year Respondents Introduced	Fieldwork Period	No. of Main Respondents	Cross-Sectional Unconditional Response Rate <sup>1</sup> (%)	Cross-Sectional Conditional Response Rate <sup>2</sup> (%)	Longitudinal Response Rate <sup>3</sup> (%)	Longitudinal All-Waves Response Rate <sup>4</sup> (%)
I	2014	Jul 2014 – Jun 2015	15,100	61	-	-	-
2		Jun 2016 – Apr 2017	12,900	69	81	80	80
3		Jun 2018 – Mar 2019	11,700	72	83	76	67
4		Sep 2020 – Aug 2021	11,200	83	90	76	60
4 <sup>5</sup>	2020		3,500	54	-	-	-

1 Cross-sectional unconditional response rate = (No. of respondents who responded in Wave X)/ (No. of eligible sample units released in Wave X), where X = 1, 2, 3 or 4

2 Cross-sectional conditional response rate = (No. of respondents who responded in Wave Y)/ (No. of respondents who responded in Wave (Y-1) and eligible in Wave Y), where Y = 2, 3 or 4

3 Longitudinal response rate = (No. of respondents who responded in Waves 1 & Y)/ (No. of respondents who responded in Wave 1 and eligible in Wave Y)

4 Longitudinal all-waves response rate = (No. of respondents who responded in all waves)/ (No. of respondents who responded in Wave 1 and eligible in Wave Y)

5 A new panel of individuals aged 45-50 years old was added in Wave 4

**TABLE 3**  
RESPONSE RATES FOR SPOUSES OR PARTNERS OF MAIN RESPONDENTS BY RHS WAVE

RHS Wave	Fieldwork Period	No. of Spouses or Partners of Main Respondents	Response Rate <sup>1</sup> (%)
I	Jul 2014 – Jun 2015	7,700	69
2	Jun 2016 – Apr 2017	6,500	70
3	Jun 2018 – Mar 2019	6,100	74
4 <sup>1</sup>	Sep 2020 – Aug 2021	7,400	70

1 Response rate = (No. of spouses or partners who responded in Wave X)/ (No. of main respondents who are married or living with a partner in Wave X), where X = 1, 2, 3 or 4

fieldwork. Lastly, data collected from respondents are checked and processed during and after fieldwork.

To provide researchers with robust data despite the twin constraints of limited time and resources, the RSHO adopts multiple strategies at each stage to:

- i. increase the survey response rate
- ii. enhance process efficiency
- iii. ensure data quality and user-friendly datasets

#### Pre-Fieldwork: Reviewing Questionnaire Design

Given the longitudinal nature of the study, the questionnaire has to be broadly similar across waves to effectively measure changes in the respondents' situations and needs. Hence, a well-designed questionnaire is crucial to preserve longitudinal data quality.

With the respective domain agencies, the RSHO conducts a review of the questionnaire before the start of each wave. This ensures that the information collected remains meaningful and relevant for policymaking, keeps the respondent burden manageable, and improves data quality (Table 4).

## Strategies

There are three broad stages in the collection of RHS data. Firstly, preparatory actions, such as designing the questionnaire and panel, are conducted prior to fieldwork. Secondly, interviews are conducted with respondents during

**TABLE 4**  
**OBJECTIVES AND STRATEGIES ADOPTED DURING QUESTIONNAIRE REVIEW**

<b>Objective</b>	<b>Strategies Adopted</b>
1 Review relevance of questions to current or new issues	a) Seek inputs from domain agencies on new or emerging issues b) Monitor ageing-related policies for assessment on the application to RHS
2 Reduce respondent burden	a) Re-evaluate value of existing questions, especially for repeat respondents b) Pre-fill responses using past wave survey data to aid respondents' recall c) Tap on administrative data sources to populate survey-equivalent responses
3 Enhance quality of data collected	Identify gaps in question clarity based on quality of data collected in past waves, as well as feedback from respondents and interviewers

### Pre-Fieldwork: Maintaining a Robust Panel

Maintaining a robust panel with high response rates is essential to ensure representativeness of the RHS sample. Before initiation of each wave, RHO reviews the panel management strategies to be implemented, such as assessing if differentiated approaches are needed for different groups.

These strategies seek to retain as many repeat respondents as possible and boost response rates from groups that are relatively harder to reach (e.g., private housing dwellers and respondents from the younger age group).

### Fieldwork: Ensuring Operational Efficiency

For effective allocation of fieldwork resources, the RHO uses administrative information to identify individuals who are out-of-scope due to long-term institutionalisation, changes in residency status and vital status.

Contact information from administrative database and data collected in previous waves are also utilised to aid interviewers in contacting respondents for the survey. To assure respondents on the value and authenticity of the survey, the RHO sends out invitation letters and emails to respondents. Following which, interviewers contact respondents to schedule appointments via calls and/or house visits.

To enhance efficiency in fieldwork operations, the RHO has developed a machine learning model to identify optimal time slots for interviewers to call on respondents without prior appointments. This model uses various information (e.g., respondents' socio-demographic characteristics) to achieve a higher rate of survey completion with fewer house visits.

### Fieldwork: Enhancing Participation through Digitalisation

Besides conventional ways of making appointments via the RHS hotline and email, respondents can make appointments via an online appointment form (available in the four official languages viz. English, Chinese, Malay and Tamil). Respondents can indicate their preferred appointment day, time and mode of the interview on the form. A substantial proportion of appointments have been made via this channel since it was introduced in Wave 4.

During the COVID-19 pandemic, the RHS was one of the first government surveys to pilot the use of Video Call Interviews (VCI) as an alternative to face-to-face interviews for respondents who had concerns with in-person meetings.

VCI has enabled the RHO to reach out to more tech-savvy respondents and those overseas during the survey period. Despite being a novel interview mode at that time, the interviewing process remains effective, with minimal impact on data quality. VCI has also enhanced the resilience of the RHS project to better overcome unforeseen disruptions to fieldwork.

In view of concerns over potential scams, the RHO has introduced an online link via CheckFirst for respondents to easily verify interviewers' identity. Many respondents have found this useful as they can verify any time instead of calling or emailing the RHO.

Finally, the RHO provides updated information for respondents through the RHS webpage, to facilitate their participation and reinforce the authenticity of the study. For Wave 5, an RHS chatbot

named Rosy was deployed to address Frequently Asked Questions from respondents instantly.

## Fieldwork and Post-Fieldwork: Engaging Respondents

As a token of appreciation to respondents who committed their time to complete the RHS interview, shopping vouchers are given to them. To encourage continued participation, a loyalty incentive was introduced since Wave 2 for repeat respondents.

The RHSO sends greetings to respondents during festive seasons, as a way of maintaining contact in-between waves. An RHS newsletter is also published regularly. The newsletter provides information on upcoming waves and assurance on data confidentiality, as well as presents useful findings from the RHS data.

## Data Management: Ensuring Data Quality

The RHSO implements various measures to ensure that complete, valid and consistent data are collected and compiled so that robust findings can be generated.

To ensure the collected data meet the intent of the questionnaire, interviewer training is conducted before the launch of each wave. Fieldwork audits are conducted regularly, and feedback is provided to interviewers with one-on-one coaching conducted to reinforce learning.

Front-end validation rules to trigger the verification of anomalous responses with respondents are also incorporated as part of the survey program to ensure accuracy of the collected data. Further back-end validation checks and editing are carried out by the RHSO to ensure cross-sectional consistency across related data domains, as well as longitudinal consistency for relatively static data items.

Administrative data are integrated with the survey data for respondents who gave their consent. With over 90 per cent of respondents consenting to linkage since Wave 1, the resultant integrated datasets are more complete and accurate. Key derived variables (e.g., non-work income), supplementary guides and tools (e.g., a search engine which enables discoverability of RHS variables) are also developed to improve data users' experience.

Finally, automation and machine learning techniques are used to improve the RHSO's efficiency in data processing.

A point-and-click interface program was created using Python programming to automate common data processing tasks. In terms of machine learning techniques, the RHSO makes use of natural language processing and a linear support vector machine model to process unstructured job and company-related text responses into the Singapore Standard Occupational and the Singapore Standard Industrial Classifications codes. Another automated valuation model is used to estimate the value of respondents' residential properties and improve the measure of respondents' housing wealth.

## Conclusion

Over the past decade, the RHSO has implemented efficient fieldwork processes and robust data management methods, developing the RHS into an authoritative source of longitudinal data for ageing-related policy research in Singapore.

The RHS data have been used by government agencies and academics to undertake studies which have contributed to policy design and thinking in diverse fields such as ageing and retirement, healthcare, private financial wealth, and social relationships. Findings based on the RHS had also been cited in local media and published in international research journals.

The RHSO is committed to promoting greater use of the RHS data to maximise their value, while continuing to safeguard respondents' confidentiality.



# How DOS's Data Tools Are Helping Businesses Make Data-Driven Decisions

by Joel Tan  
Business Statistics Division  
Singapore Department of Statistics

## Introduction

Service journeys conducted with businesses revealed that while businesses recognise the importance of official statistics, they face challenges in consuming such data.

Businesses have provided feedback that data are not contextualised, making them difficult to understand and use. As a case in point, a business owner wanted to understand the customer demographics in an area but was unaware that the information can be found within an existing statistical report.

The Singapore Department of Statistics (DOS) stands ready to help businesses make better use of official statistics relevant to them. To address the feedback received, DOS has launched the Data for Businesses webpage which provides data tools containing curated and contextualised data from different agencies.

These tools aim to empower businesses to make data-driven decisions with insights covering three broad themes, viz. (1) customers, (2) Industry, and (3) business performance. They were developed in collaboration with key partner agencies and enhanced based on businesses' feedback during pilot testing.

## Theme I: Know My Customer

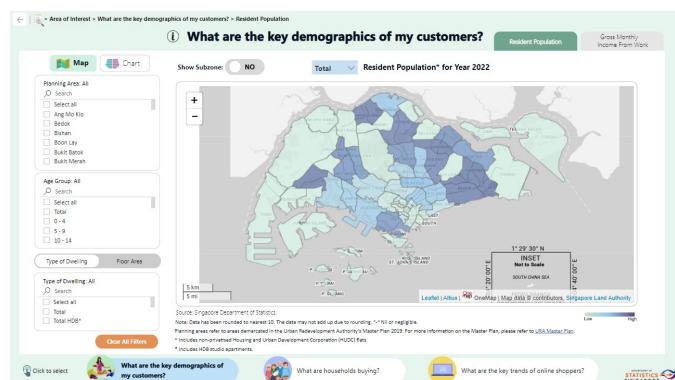
Businesses are in constant pursuit of better customer-insights to elevate customer experiences. The data tools under this theme provide information and insights on customers' demographics, household consumption patterns and trends of online shoppers. By supplementing their own customer data with information from these data tools, businesses can gain enhanced insights on their customers for their business planning purposes.

### Customer Demographics

Detailed information on consumer populations by geographic area could be useful to businesses. By fusing population and geospatial data, the data tool enables businesses to zoom in on specific geographic areas of interest and tap into the rich demographic information (e.g., resident

population with breakdown by type of dwelling, age group and sex) (Figure 1).

**Figure 1**  
**KNOW MY CUSTOMER**  
**— KEY DEMOGRAPHICS OF CUSTOMERS**

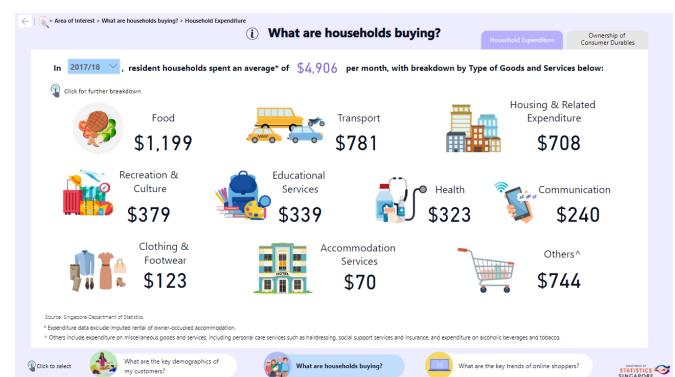


The data tool also presents resident population data on a map, which makes it easier for businesses to visualise and discover pertinent information about the demographics of their customers residing in an area of interest.

### Consumption Patterns

Understanding customers' expenditure patterns facilitates businesses' market analyses. The data tool allows businesses to readily discover customers' expenditure patterns based on data from DOS's Household Expenditure Surveys. Using the data tool, businesses can further delve into the details of what households are spending on and their average expenses by type of goods and services (Figure 2).

**Figure 2**  
**KNOW MY CUSTOMER**  
**— HOUSEHOLD SPENDING**



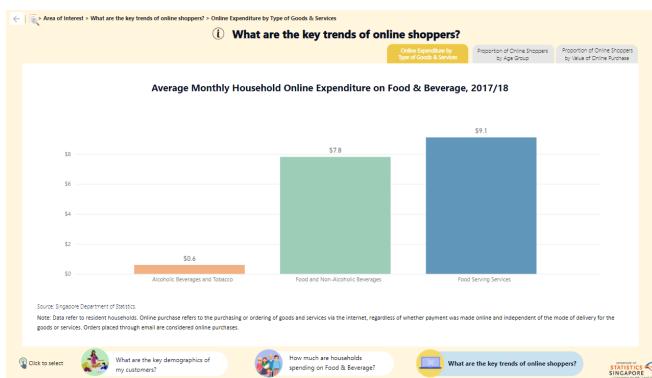
## Trends of Online Shoppers

Online spending is growing in prevalence. The data tool also presents data from DOS's Household Expenditure Survey and the Infocomm Media Development Authority's (IMDA) online spending data in user-friendly charts.

Businesses can leverage such information to size the potential market for online sales from insights on online expenditure by type of goods and services and the profile of online shoppers (Figure 3).

**Figure 3**

### KNOW MY CUSTOMER — ONLINE SPENDING



## Theme 2: Know My Industry

Having a strong understanding of the industry helps power business strategies. Businesses can obtain industry-specific information on business performance, costs, and labour market situation from the data tool. These insights can help augment businesses' own market research to provide a more holistic view of the industry.

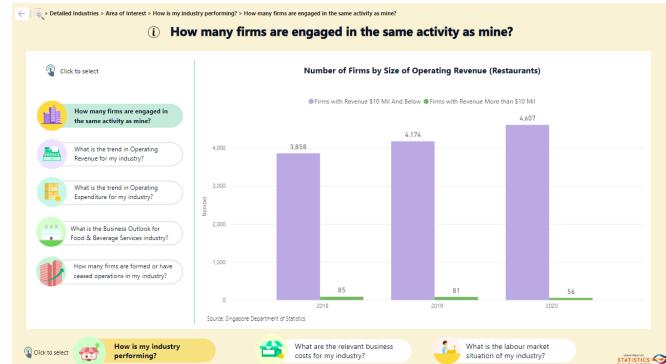
### Industry Performance

A full suite of data on the industry, ranging from the number of companies engaged in the same activity to the business performance and outlook of the industry is available in the data tool. Data from various surveys are presented in charts for businesses to perform market analyses.

Besides quantitative data, the data tool also presents information on industry-specific business outlook gathered from DOS's Business Expectations Survey. Businesses can leverage the collective sentiments of fellow industry players on the projected business outlook for their industry in their planning and risk management (Figures 4 and 5).

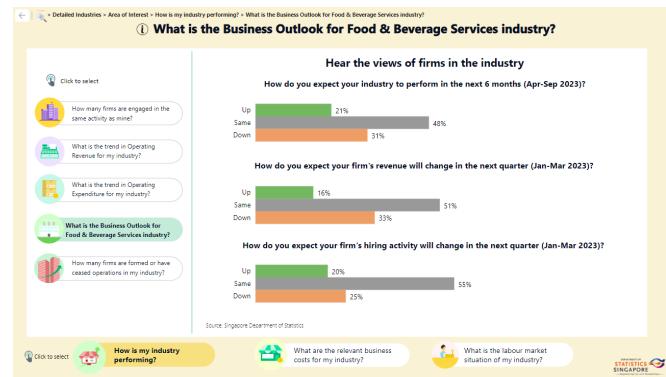
**Figure 4**

### KNOW MY INDUSTRY — NUMBER OF SIMILAR FIRMS



**Figure 5**

### KNOW MY INDUSTRY — BUSINESS OUTLOOK

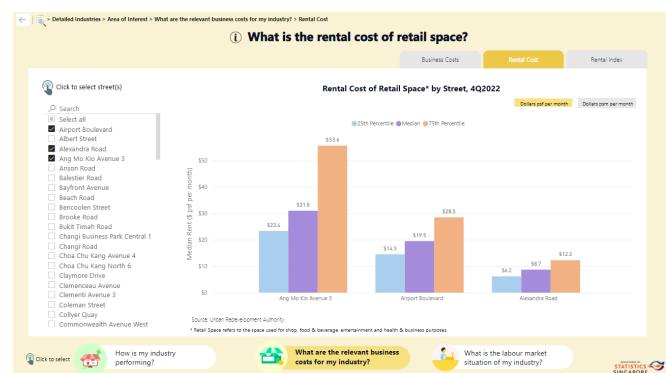


### Business Costs

Business cost is a crucial component of businesses' bottom line, which businesses seek to optimise. With the data tool, businesses can find out the top 3 business costs facing the industry and obtain information on rental costs by street (Figure 6). Businesses can also use the latter to make an informed decision before renting any premises.

**Figure 6**

### KNOW MY INDUSTRY — RENTAL COSTS

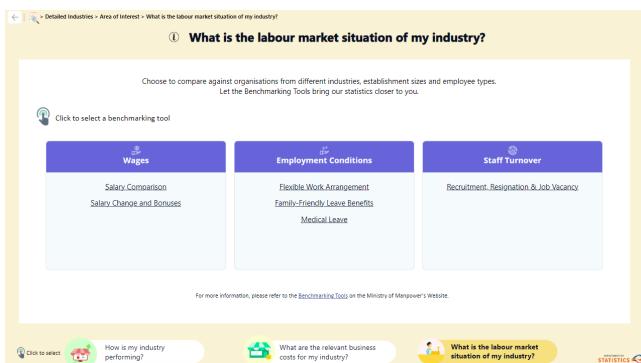


## Labour Market Situation

Labour market analysis is an integral part of a successful business. The data tool provides links to the Ministry of Manpower's data tools that allow businesses to compare salaries and bonuses as well as employment conditions against industry norms. Businesses can also analyse staff turnover rates against industry norms to review their recruitment or retention strategies (Figure 7).

**Figure 7**

### KNOW MY CUSTOMER — LABOUR MARKET SITUATION



## Theme 3: Benchmark My Performance

The data used in this benchmarking tool are contextualised and personalised, providing businesses with an improved experience in effortlessly benchmarking their performances against their contemporaries.

For example, businesses can compare their financial ratios with the industry average. Commonly used ratios such as Return on Assets, Return on Equity, Current Ratio and Equity Ratio are also represented on the data tool (Figure 8).

**Figure 8**

### BENCHMARK MY PERFORMANCE — SELECTED FINANCIAL RATIOS



Deep dives and benchmarking exercises can be performed on the firm's profit margin, profit per worker and revenue per worker, to possibly identify areas done well and areas for improvement. Expenditure related ratios can be compared against fellow industry players to identify relative inefficiencies in spending (Figure 9).

**Figure 9**

### BENCHMARK MY PERFORMANCE — BUSINESS PERFORMANCE



## Conclusion

These data tools, by providing businesses with more insights on their customers, industry and business performance, empowers them to make data-driven decisions.



### Hear from Businesses Their Feedback on the Data Tools!

The interactive electronic publication of data on SingStat's new website is a useful tool for our business. Through features such as the visualisation dashboards, we are able to easily access relevant statistics which help us gain a better understanding of our industry, how we are doing relative to other players, and potentially identify new growth opportunities through a data-driven approach.

**Mr Pang Fu Wei,  
Group Managing Director, Mothercare**

The business tools gave concise insights to the F&B industry statistics and helped me gain a better understanding on the changing customer demographics.

**Mr Mervin Lee,  
Vice President, Fei Siong Group**

 **Start your data journey with us today!**

Visit [go.gov.sg/dfb-ssn](http://go.gov.sg/dfb-ssn) to use our business tools (available for Retail and F&B industries).

# Which Fields of Study Have Higher Proportion of Males/ Females?

by Erica Ngan  
Household, Income and Population Division  
Singapore Department of Statistics

## Introduction

Over the years, residents in Singapore have made significant progress in education. In 2020, the proportion of residents aged 45 years and over with a polytechnic diploma or university qualification was 30.0 per cent for males and 18.9 per cent for females, up from 9.3 per cent and 3.1 per cent in 2000 respectively. Among those aged 25-34 years in 2020, the proportion was closer at 70.1 per cent for males and 73.8 percent for females.

With more residents obtaining polytechnic diplomas and university qualifications, do the fields of study with a larger proportion of males or females in 2000 continue to be so in 2020? How has the distribution by sex within the fields of study changed over the same period?

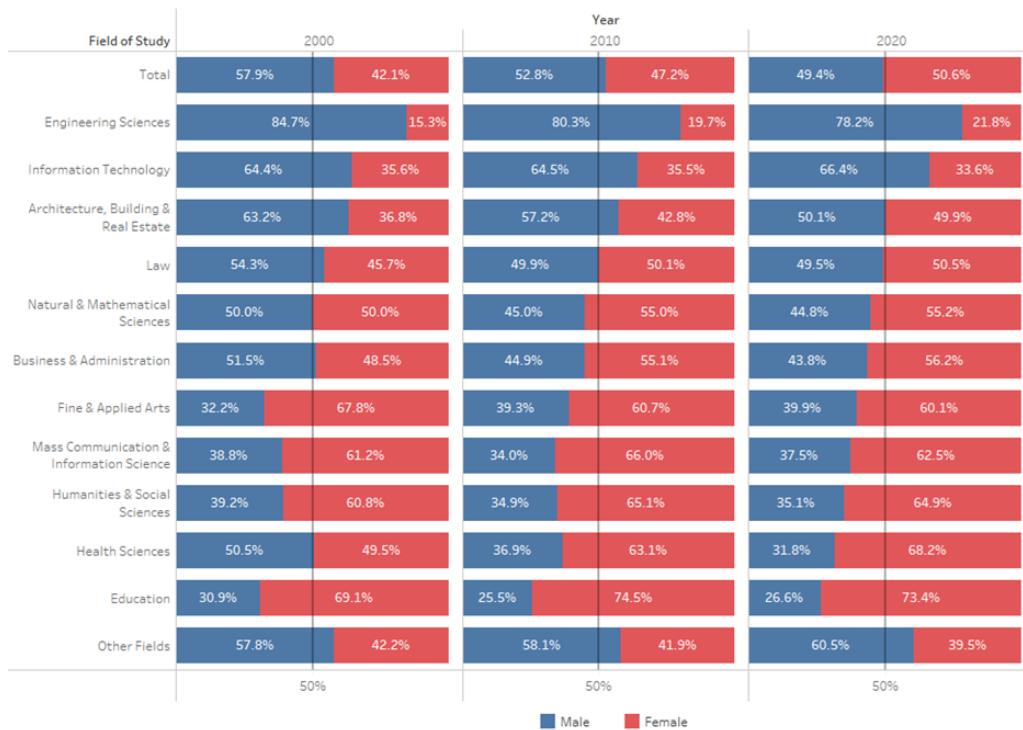
This article examines these questions using data on residents aged 25 years and over with a polytechnic diploma or university qualification<sup>1</sup> from the Censuses of Population 2000, 2010 and 2020. Residents comprise both Singapore citizens and permanent residents.

## Sex Distribution by Field of Study

**Engineering Sciences and Education Remained the Top Fields of Study with More Male and More Female University Graduates Respectively**

In 2020, there were as many female university graduates as their male counterparts. This was unlike two decades ago when males outnumbered females (Chart 1). The distributions have also become

**CHART I**  
**RESIDENT UNIVERSITY GRADUATES AGED 25 YEARS AND OVER**  
**BY FIELD OF STUDY AND SEX, 2000, 2010 AND 2020**



<sup>1</sup> Data exclude those who were attending educational institutions as full-time students. Those who were upgrading their qualifications through part-time courses are included.

less equal within more fields of study over the years, with females outnumbering males in more than half of the fields of study since 2010.

Education has consistently been the field of study with the largest proportion of female graduates. From 69.1 per cent in 2000, the share of female graduates in this field of study increased to close to 75 per cent in 2010 and 2020. Health Sciences, which had an equal proportion of males and females in 2000, saw a sizeable increase for females over the last two decades, with close to 7 in 10 being females in 2020.

Fine & Applied Arts, Mass Communication & Information Science and Humanities & Social Sciences continued to have a larger share of female graduates since 2000. While there continued to be more females in Fine & Applied Arts, the proportion had declined from 67.8 per cent in 2000 to 60.1 per cent in 2020.

Engineering Sciences remained the field of study with the largest proportion of male university graduates. In 2020, close to 8 in 10 university graduates in Engineering Sciences were males. Despite the increase in the proportions of females for most fields of study, the share of male university graduates in Information Technology continued to increase from 64.4 per cent

in 2000 to 66.4 per cent in 2020. This field of study had the second highest proportion of male graduates.

Architecture, Building & Real Estate as well as Law saw an equal split between the sexes among university graduates in 2020. For the former, less than 4 in 10 of the university graduates in this field of study were females two decades ago.

### Health Sciences and Education Had Highest Share of Females While Engineering Sciences Had Highest Share of Males among Polytechnic Diploma Graduates Since 2000

Similar to university graduates, the proportion of females among polytechnic diploma graduates also rose compared to 2000. However, unlike university graduates, the proportion of males remained higher than that of females, although it had declined from 65.9 per cent in 2000 to 59.4 per cent in 2020 (Chart 2).

Health Sciences and Education remained the fields of study with the highest proportions of females among polytechnic diploma graduates with more than 7 in 10 being females, despite slight declines in the proportions over the past two decades.

**CHART 2**  
RESIDENT POLYTECHNIC DIPLOMA GRADUATES AGED 25 YEARS AND OVER  
BY FIELD OF STUDY AND SEX, 2000, 2010 AND 2020



■ Male ■ Female

Note: Figures may not add up to 100 per cent due to rounding.

Similarly for Business & Administration, which traditionally had more females, the share of females declined from 68.8 per cent in 2000 to 63.2 per cent in 2020.

Approximately 8 in 10 polytechnic diploma graduates in Engineering Sciences were males in 2020, making it the field of study with the highest proportion of males. This stayed unchanged since 2000. Male polytechnic graduates continued to outnumber females in fields such as Services, Information Technology, as well as Architecture, Building & Real Estate in 2020. While the proportion of males in Architecture, Building & Real Estate had declined slightly over the two decades, the other two fields of study observed increases.

## Sex Distribution by Field of Study and Age Group

**Among University Graduates in 2020, Education Had Higher Proportion of Females among Younger Age Groups, Whereas Engineering Sciences Had Higher Proportion of Males among Older Age Groups**

For university graduates, fields of study with disproportionately more females tend to have higher proportions of females among the younger age groups relative to the older age groups (Chart 3). This is

apparent for Education, which had the highest proportion of females in 2020. Among university graduates aged 25-34 years in this field, 80.6 per cent were females, higher than those in the age groups of 35-44 years (73.7 per cent) and 45 years and over (69.5 per cent).

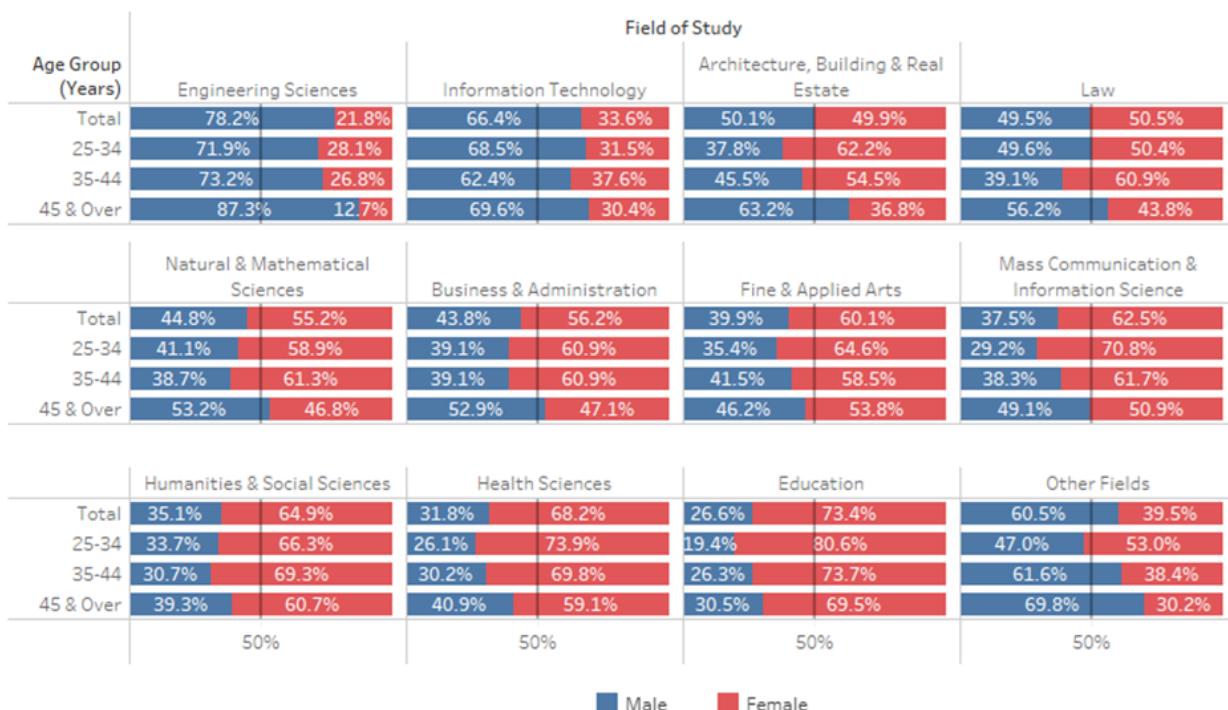
Notably, Mass Communication & Information Science university graduates aged 25-34 years and 35-44 years also had a larger proportion of females at 70.8 per cent and 61.7 per cent respectively. Those aged 45 years and over had almost equal proportions of males and females.

On the other hand, Engineering Sciences had the highest proportion of males, with a considerably higher proportion of males among the older age groups. Among graduates aged 45 years and over, 87.3 per cent were males compared to 73.2 per cent and 71.9 per cent for those aged 35-44 years and 25-34 years, respectively.

Despite an overall equal split between both sexes in Architecture, Building & Real Estate, the younger age groups of 25-34 years and 35-44 years had higher proportions of females, unlike those aged 45 years and over. In contrast, Law saw an equal share of male and female graduates among those aged 25-34 years. Among those aged 35-44 years, females formed a larger share, while the opposite was true for those aged 45 years and over.

**CHART 3**

RESIDENT UNIVERSITY GRADUATES AGED 25 YEARS AND OVER BY AGE GROUP, FIELD OF STUDY AND SEX, 2020



Note: Figures may not add up to 100 per cent due to rounding.

## Education Had Higher Share of Females in Younger Age Groups While Services Had Higher Proportion of Males in Older Age Groups Among Polytechnic Graduates in 2020

Among polytechnic diploma graduates, the trend by age group differs slightly for the two fields of study with the highest proportion of females. In the case of Health Sciences, the proportions of females ranged similarly between 74.3 per cent and 78.9 per cent across age groups (Chart 4).

In comparison, for Education, the proportion of females was higher at 83.6 per cent among those aged 25-34 years, compared to 76.3 per cent and 67.9 per cent for the other two older age groups. Additionally, while Mass Communication & Information Science had higher proportions of females among the younger age groups of 25-34 and 35-44 years, the proportion of females among those aged 45 years and over was lower at 40.0 per cent.

In Engineering Sciences, both proportions of males among those aged 25-34 years and 45 years and over exceeded 80 per cent while the proportion of males among those aged 35-44 years was slightly lower at 72.9 per cent. While in Services, the proportion of males was relatively higher for those in the older age

groups, ranging from 78.2 per cent for those aged 45 years and over to 58.8 per cent for those aged 25-34 years.

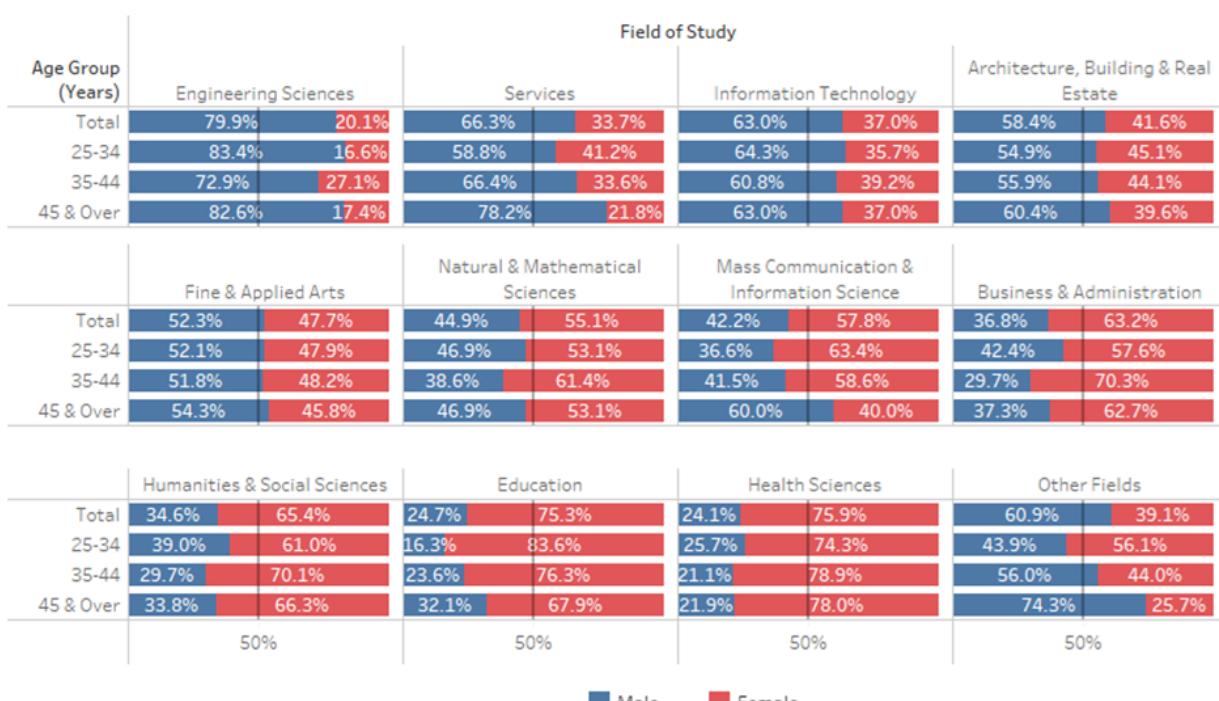
## Conclusion<sup>2</sup>

While the proportions of polytechnic diploma and university graduates among residents have increased over the last two decades, the fields of study with higher proportions of males or females were generally unchanged over the same period.

The fields of study with higher proportions of female university graduates than males generally had a higher proportion of females in the younger age groups relative to the older age groups. The reverse was true for Engineering Sciences, the field with the highest proportion of male university graduates, with higher proportion of males in the older age group relative to younger age groups.

Overall, Health Sciences and Education had the highest proportions of female polytechnic diploma graduates. The proportions of females in Health Sciences were relatively similar across all age groups while Education saw a larger share of females among the younger age groups. In contrast, the proportion of male polytechnic diploma graduates in Services was higher in the older age groups compared to those in the younger age groups.

**CHART 4**  
RESIDENT POLYTECHNIC DIPLOMA GRADUATES AGED 25 YEARS AND OVER  
BY AGE GROUP, FIELD OF STUDY AND SEX, 2020



Note: Figures may not add up to 100 per cent due to rounding.

2 Read more in other related SSN article: SSN Issue 2, 2022. "[Gender Differentials in Educational Profile and Field of Study among Residents](#)"

# SINGAPORE 2022



## ECONOMY

Real Gross Domestic Product Growth was **3.6%**



Overall Unemployment Rate was **2.1%**



Inflation Rate was **6.1%**

Why does the reported inflation differ from my personal experience?  
Infographic

Gross National Savings reached **S\$265.5 bil**



International Visitor Arrivals was **6.3 mil**



Air Cargo Handled was **1.9 mil tonnes**



Per Capita Gross National Income was **S\$95,787**



Nominal Value Added for the Manufacturing Sector was **S\$134.5 bil**



Sea Cargo Handled reached **578.2 mil tonnes**



Labour Force Participation Rate was **70.0%**



Investment Commitments in Manufacturing and Services (Fixed Assets Investments) amounted to **S\$22.5 bil**



Total Merchandise Trade was **S\$1,365.4 bil**



## SOCIO-DEMOGRAPHY

Total Population was **5.6 mil** as at end-Jun 2022



Population Trends Dashboard

Mean Years of Schooling was



**12.1 years**



**11.3 years**

for residents aged 25 years and older

Proportion of Owner-Occupied Resident Households\* was **89.3%**



Total Marriages was **29,338<sup>P</sup>**



Mobile Population Penetration Rate reached **165.9%**



Residential Wired Broadband Household Penetration Rate was **92.7%**



Resident Total Fertility Rate was **1.05<sup>P</sup> per female**



Key Household Income Trends Dashboard

Median Monthly Household Income from Work Per Household Member was **S\$3,287**

Physical Crimes Rate was **358** per 100,000 population

Scams & Cybercrimes Rate was **597**

P: Preliminary

\* Refers to the proportion of households where the household reference person or any other member owns the house, and is as reported by respondents.



Find out more:  
[Singapore Economy Infographic](#)



Find out more:  
[Singapore Population Infographic](#)

Find us at:

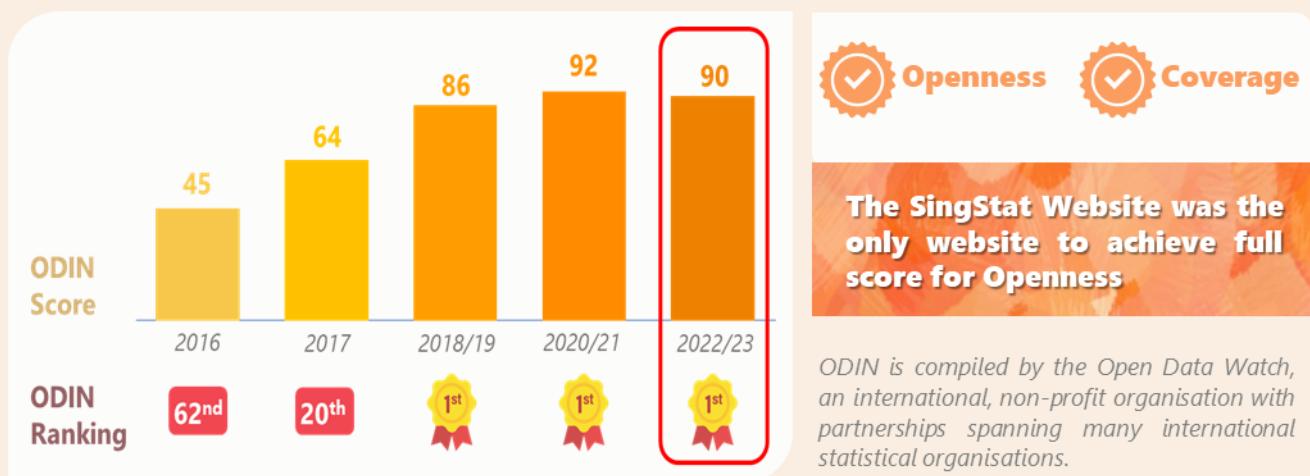


DEPARTMENT OF  
**STATISTICS SINGAPORE**  
Empowering You with Trusted Data



## The SingStat Website Retains Top Spot in the Global Open Data Inventory (ODIN) 2022/23 Ranking

The ODIN 2022/23 assessed the websites of 192 National Statistical Offices (NSOs) on the Openness (i.e., accessibility) and the Coverage (i.e., availability) of their data. The SingStat Website has been topping the ODIN ranking since 2018/19.



The indicators assessed are based on recommendations from international agencies and country practices, including those required to monitor the Sustainable Development Goals. They cover key domains of:



### Economic Statistics

(e.g., National Accounts, Price Indices, Money and Banking, International Trade)



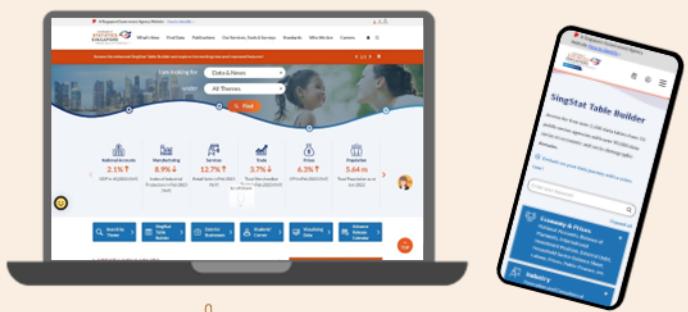
### Social Statistics

(e.g., Population and Vital Statistics, Education Outcomes, Health Outcomes)



### Environmental Statistics

(e.g., Agriculture and Land Use, Resource Use, Energy)



The SingStat Website & the SingStat Table Builder

To improve Coverage, i.e., data availability, the Singapore Department of Statistics continues to work closely with public sector agencies to facilitate convenient and user-friendly access to a wide range of statistics via the [SingStat Website](#) and the [SingStat Table Builder](#).

# CONTENTS

- 1 High Growth Firms in Singapore**
- 4 Experimental Uses of Machine Learning and New Data Sources in Updating the Statistical Business Register**
- 8 The Retirement and Health Study — Management of a Large-Scale Longitudinal Study in Singapore**
- 12 How DOS's Data Tools Are Helping Businesses Make Data-Driven Decisions**
- 15 Which Fields of Study Have Higher Proportion of Males/ Females?**
- 19 Singapore 2022**
- 20 The SingStat Website Retains Top Spot in the Global Open Data Inventory (ODIN) 2022/23 Ranking**



## Statistics Singapore Newsletter Issue 1, 2023

The Statistics Singapore Newsletter is issued twice a year by the Singapore Department of Statistics.

It aims to inform readers on recent statistical findings as well as latest information on statistical methodologies, processes, products and services.

[www.singstat.gov.sg](http://www.singstat.gov.sg)  
[info@singstat.gov.sg](mailto:info@singstat.gov.sg)

Find us on

