

README for the Reproducibility Package for the Prosperity Insight Note: Entrepreneurship Database

Table of Contents

Overview.....	1
Data Availability Statement	2
Instructions for Replicators	2
List of Exhibits	2
Software Requirements	3
Code Description & Folder Structure	3

Overview

The Entrepreneurship Database is a global cross-country dataset on the entry and exit of formal private enterprises. The 8th Edition of the Entrepreneurship Database contains annual data on the number of limited liability companies covering 188 economies and spanning the period 2006-2024.

The aim of this reproducibility package is to generate the figures present in the Prosperity Insight Note: Entrepreneurship Database. To do so, we calculate several metrics including the business density rates for each country and year, which are published on the Entrepreneurship Database website (<https://www.worldbank.org/en/programs/entrepreneurship>). These include the business density rates which are based on the data collected from business registries on the flow and stock of limited liability companies (LLCs), which include the number of new, and closed, LLCs during each calendar year and total LLCs at the end of each calendar year. The business density rates that we publish are defined as:

- New business density, defined as the number of newly registered companies with limited liability per 1,000 working-age adults (ages 15–64) per calendar year.
- Closed business density, defined as the number of closed companies with limited liability per 1,000 working-age adults (ages 15–64) per calendar year.
- Total business density, defined as the total number of registered firms with limited liability per 1,000 working-age people (ages 15-64) at the end of each calendar year.

Data on the number of LLCs per country and year were obtained from business registries. Data on the population of working-age adults (ages 15-64) were compiled from the World Bank's World Development Indicators (WDI) (<https://databank.worldbank.org/source/world-development-indicators>).

We also include sex disaggregated data across three categories:

- LLC business owners, defined as the individuals who hold at least one share in a newly registered limited liability company during the calendar year.

- LLC business directors, defined as the individuals responsible for managing the affairs of a newly registered limited liability company.
- Sole proprietors, defined as single individuals who own and manage a business that is not legally separate from the owner and who bear personal liability.

For each category, counts are disaggregated by gender, and we calculate the shares that represent the proportion of female or male participants out of the total in that category, by dividing the number of female (or male) counts by the total.

Data Availability Statement

Dataset Name	File Name(s)	Link/Access Instructions	Access Date	Data Access
Entrepreneurship Database	Firm Statistics - New Firms	https://www.worldbank.org/en/programs/entrepreneurs-hip	September 2025	Open
Entrepreneurship Database	Firm Statistics - Total Number of Firms	https://www.worldbank.org/en/programs/entrepreneurs-hip	September 2025	Open
Entrepreneurship Database	Firm Statistics - Closed Firms	https://www.worldbank.org/en/programs/entrepreneurs-hip	September 2025	Open
Entrepreneurship Database	Gender Statistics – Business Owners	https://www.worldbank.org/en/programs/entrepreneurs-hip/gender	September 2025	Open
Entrepreneurship Database	Gender Statistics – Business Directors	https://www.worldbank.org/en/programs/entrepreneurs-hip/gender	September 2025	Open
Entrepreneurship Database	Gender Statistics – Sole Proprietors	https://www.worldbank.org/en/programs/entrepreneurs-hip/gender	September 2025	Open
World Development Indicators (WDI)	Population ages 15-64, total	https://databank.worldbank.org/source/world-development-indicators	September 2025	Open

Instructions for Replicators

To reproduce the figures included in the Prosperity Insight Note: Entrepreneurship Database, the replicator should open the folder named “Reproducibility Package”, then:

- 1- Open the master do file named “main.do”
- 2- Replace our directory with their own directory in Line 11
- 3- Run the code

All Figures can be replicated as they appear in the Prosperity Insight Note: Entrepreneurship Database

List of Exhibits

Figures in the Prosperity Insight Note: Entrepreneurship Database. These are:

- Figure 1: Average new business density over time from 2006 to 2024 per income group
- Figure 2: New business density in Cabo Verde from 2008 to 2024
- Figure 3: Average share of female LLC owners and directors, and sole proprietors in 2024
- Figure 4: Evolution of female share of sole-proprietors for economies with 50% or more females in 2024

Software Requirements

Stata

Excel

Code Description & Folder Structure

All the material needed for reproducing the figures are included in the folder "Reproducibility Package" which contains three subfolders: Data, Programs, and Results.

Data Folder

The data folder contains an excel file named "EntrepreneurshipDatabase_8thEdition.xlsx" which contains two sheets. The first sheet, named Firms data contains data on the number of new, closed, and total LLCs for each country and year from 2006-2024, gathered from the business registries. We have added to this excel sheet data on the population of working adults aged 15-64 gathered from the WDI as well as the country classification. This excel sheet was then converted in stata to a dta file named "EntrepreneurshipDatabase_Firms data.dta". The second sheet contains data on the female, male, and total number of LLC business owners, LLC business directors, and sole proprietors. We have added to this excel sheet data on the country classification gathered from the WDI. This excel sheet was then converted in stata to a dta file named "EntrepreneurshipDatabase_Sex disaggregated data.dta".

This data on the number of firms and sex disaggregated data is available on the Entrepreneurship Database website (<https://www.worldbank.org/en/programs/entrepreneurship>) under the Firm number tab and Gender statistics tab respectively (

Programs Folder

This folder contains the do files that produce the four figures available in the Prosperity Insight Note: Entrepreneurship Database. To run the files, just replace the code on line 11 in the main.do file with the appropriate directory. Then run the main file and the ProsperityNote-Graphs.do file will automatically run and the figures will be populated in the Results folder.

Results Folder

Running the do files in the programs folder will generate a gph and png file for each figure.