1. Log-GDP: Log transformation of GDP per capita in current U.S dollars. GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.
2. NB: Multi-Categorical feature representing the number of new limited liability corporations registered in the calendar year. The country is categorized as ‘high’ if the number of new business registered larger than 20000, ‘medium’ if between 1000 and 20000, ‘low’ if smaller than 1000.
3. Cost: Multi-Categorical feature representing the cost to register a business normalized by presenting it as a percentage of gross national income (GNI) per capita. The country is categorized as ‘High’ if the cost in percentage larger than 75, ‘Medium High’ if between 25 and 75, ‘medium low’ if between 25 and 10, ‘low’ if smaller than 10.
4. LowerS: Ratio of people entering last grade of lower secondary education to the actual age population entering last grade.
5. percent\_urban\_population: Urban population ratio with whole population.
6. Procedures: Start-up procedures are those required to start a business, including interactions to obtain necessary permits and licenses and to complete all inscriptions, verifications, and notifications to start operations.
7. Sexratio: Male to female ratio.
8. Time - Number of days required to start.