**INTRODUÇÃO**

Retirado do Censo de 2022 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), este conjunto de dados reúne informações sobre a situação de vida da população em cada um dos municípios e localidades do país, incluindo informações como sexo, idade e distribuição de renda, por exemplo.

Após uma análise de todos os arquivos disponibilizados, optou-se por avaliar a desigualdade entre as regiões e Unidades da Federação (UFs) brasileiras, com foco principalmente na Renda Média e sua correlação com outros indicadores socioeconômicos, como: Educação, posse de bens, acesso à esgoto e abastecimento de água, além da condição de moradia.

LEMBRETE: Em todos os trabalhos de Excel e PowerBI, caso deseje avaliar o trabalho feito no processo de ETL do PowerQuery, deve-se substituir o caminho presente na etapa ‘Fonte’ pelo caminho onde seus arquivos foram salvos.

**DESENVOLVIMENTO**

Primeiramente, foi realizado um backup dos dados originais e, em seguida, definido quais planilhas/tabelas seriam utilizadas no projeto. As planilhas utilizadas foram copiadas e dispostas em outra pasta chamada ‘Dados Trabalhados’. Como a análise tinha o foco nas diferenças entre regiões, foram retiradas do projeto planilhas que avaliavam os dados de outra forma: Gênero, Cor, Posição, etc.

Não foram desenvolvidas análises temporais devido à inconsistência nos dados do IBGE. Por exemplo, algumas planilhas continham dados de vários anos disponíveis (2016, 2017, etc.), enquanto outras só estavam preenchidas a partir de 2021 ou apenas continham o ano de 2022. O usuário deste projeto pode navegar por todas as planilhas disponíveis na pasta de trabalho chamada ‘Dados Originais’. Por esse motivo, foram utilizados dados somente do ano de 2022, com o intuito de realizar uma avaliação mais atualizada do perfil socioeconômico brasileiro.

O objetivo deste dashboard é ser informativo e interativo, por isso alguns dados foram simplificados para facilitar a experiência de usuários que conhecem apenas as funções básicas da ferramenta.

ETL e MODELAGEM

As etapas de ETL e Modelagem foram desenvolvidas simultaneamente. Enquanto se realizava o ETL, também foi definida a criação manual de uma tabela no próprio Excel contendo os estados e regiões. Esta seria a tabela responsável por iniciar e facilitar o relacionamento entre as outras através da mescla de consultas, visando manter a correta cardinalidade.

Após a padronização das tabelas, foram removidas algumas linhas e colunas, alterados tipos de colunas e elaboradas fórmulas na Linguagem M para transformar valores em porcentagem.

VISUALIZAÇÃO

O primeiro Dashboard avalia o perfil socioeconômico por região, contendo um botão para navegar entre as regiões, facilitando a extração de insights.

O segundo dashboard faz uma relação entre a renda média e todos os outros indicadores, porém desta vez por Unidade Federativa, focando nos 5 melhores e 5 piores estados.

Para a elaboração de dashboards que facilitassem a compreensão do usuário e não fossem poluídos com excesso de informações, os dados a seguir foram agrupados:

Nível de Instrução: Levado em consideração uma formação completa na escola, ou seja, Ensino Médio Completo como Alto Nível de Instrução devido ao seu impacto no desenvolvimento pessoal, além de ser requisito mínimo para diversos cargos no mercado de trabalho.

Acesso à rede de esgoto e abastecimento de água: O critério utilizado para definir se os acessos são adequados ou não foi retirado do PLANSAB (Plano Nacional de Saneamento Básico), que indica que deve ser considerado acesso adequado à rede de esgoto como população servida de Rede Geral ou Pluvial, Fossa Séptica Ligada à Rede e Fossa Séptica Não Ligada à Rede. Para o abastecimento de água, o PLANSAB indica que está adequado à população que possui acesso via Rede Geral de Distribuição, Poço Profundo ou Artesiano e Poço Raso, Poço Freático ou Cacimba.

Condição de Moradia (Dashboard de Indicadores): As porcentagens utilizadas para este dashboard são de pessoas que possuem casa própria, excluindo as cedidas ou por aluguel. Como se está fazendo um comparativo com a Renda Média, seria válido separar a população que já tem a casa própria de quem paga aluguel ou ainda tem a casa cedida de alguma forma.

Bens em Domicílio (Dashboard de Indicadores): Bens em Domicílio não foi colocado nesse dashboard devido ao tamanho da abrangência das definições de "Bens Essenciais" e "Luxos" ou "Bens de Alto Valor". Essa abrangência muda dependendo do estado, maneira de aquisição, forma de trabalho e, como neste dashboard o objetivo não era colocar todos os bens por estado devido à poluição visual que ocasionaria, foi optado por deixar de fora este indicador pela impossibilidade de resumi-lo.

OBS: Além das macros, ambas as planilhas possuem botões. No Dashboard de Perfil Socioeconômico, você pode navegar entre as regiões, assim como no Dashboard de Indicadores, você pode escolher qual indicador, além do fixo que é a Renda Média, deseja realizar a comparação. Também existe a possibilidade de exibir as tabelas para análises exploratórias mais profundas e avaliação das fórmulas utilizadas na planilha de ‘Fonte Dashboards’. Nela estão dispostas tabelas dinâmicas, cálculos e fórmulas (ÍNDICE, CORRESP, PROCX), onde estão concentrados todos os dados usados para elaboração dos dashboards interativos.

**CONCLUSÕES**

\*\* Dashboard Perfil Socioeconômico por Região: \*\*

As regiões com maior Renda Média são as que possuem maior nível de instrução, além de infraestrutura básica. Alguns motivos que podem estar ligados a estes fatores são que regiões com maior renda média tendem a ter mais recursos disponíveis para investir em infraestrutura, incluindo sistemas de abastecimento de água e redes de esgoto. Isso pode resultar em melhores condições de vida e acesso a serviços básicos.

Regiões com maior renda média também oferecem mais oportunidades de emprego e salários mais altos. Isso pode motivar as pessoas a buscar níveis mais elevados de instrução ou atrair pessoas de níveis mais elevados de instrução de outras regiões, na tentativa de garantir empregos melhores e mais estáveis, o que contribui para um maior nível educacional nessas áreas.

A maior surpresa deste dashboard é a de que existe uma disparidade entre as regiões com maior renda e a posse de imóveis. Os motivos que podem levar a esse fator são o custo de vida e habitação que são maiores nessas regiões. Outro motivo que poderia levar a essa disparidade é a presença de indivíduos que estão transitando para estudar ou iniciar a carreira, preferindo por alugar imóveis ao invés de comprá-los. As políticas habitacionais em áreas de menor rendimento também podem afetar esses indicadores, pois os habitantes de regiões de renda inferiores podem ter maior facilidade na aquisição de imóveis.

Em Bens em Domicílio, o que se nota é que nas regiões Norte e Nordeste, a população possui uma porcentagem significativamente menor de automóveis, microcomputadores, acesso à internet e máquinas de lavar roupa, coincidindo com o fato de terem uma menor renda média e, consequentemente, um menor poder de compra. Vale notar que o inverso ocorre para a posse de motocicletas, que é um meio de locomoção mais barato.

\*\* Dashboard Indicadores: \*\*

Destaca-se o Distrito Federal, a unidade federativa de maior renda média, com aproximadamente 35% a mais que São Paulo, o segundo colocado neste indicador. Ele é responsável por elevar a média da região, tornando-a a maior neste quesito, com R$ 3.195,75. Além disso, é a única UF onde mais de 70% da população possui um alto nível de instrução.

Essa informação complementa os insights que tivemos no outro dashboard e traz à luz novos motivos para a correlação renda média x nível de instrução, pois para o Distrito Federal, as causas podem ser de uma maior concentração de riqueza e também a concentração de serviços públicos e instituições governamentais.

Em contrapartida, o Distrito Federal possui a população com a menor taxa de casas próprias do país.

Já do outro lado, e por motivos já explicados após avaliar o primeiro Dashboard, temos o Maranhão, com a 2ª menor renda média do país, apresentando a 2ª melhor condição de moradia, onde mais de 83% da população tem casa própria.

**INTRODUCTION**

Taken from the 2022 Census conducted by the Brazilian Institute of Geography and Statistics (IBGE), this dataset brings together information on the living conditions of the population in each of the country's municipalities and localities, including information such as gender, age, and income distribution, for example.

After an analysis of all the files made available, it was decided to evaluate the inequality between the Brazilian regions and Federative Units (UFs), focusing primarily on Average Income and its correlation with other socioeconomic indicators, such as: Education, asset ownership, access to sewage and water supply, as well as housing conditions.

REMINDER: In all Excel and PowerBI works, if you wish to evaluate the work done in the PowerQuery ETL process, you must replace the path present in the 'Source' step with the path where your files were saved.

**DEVELOPMENT**

First, a backup of the original data was made, and then it was defined which spreadsheets/tables would be used in the project. The spreadsheets used were copied and placed in another folder called ‘Processed Data’. Since the analysis focused on the differences between regions, spreadsheets that evaluated data in other ways were removed from the project: Gender, Color, Position, etc.

No temporal analyses were developed due to inconsistencies in the IBGE data. For example, some spreadsheets contained data from several years available (2016, 2017, etc.), while others were only filled in from 2021 onwards or only contained the year 2022. The user of this project can browse all the spreadsheets available in the work folder called ‘Original Data’. For this reason, only data from the year 2022 was used, with the aim of conducting a more up-to-date assessment of the Brazilian socioeconomic profile.

The goal of this dashboard is to be informative and interactive, which is why some data has been simplified to facilitate the experience for users who only know the basic functions of the tool.

ETL AND MODELING

The ETL and Modeling stages were developed simultaneously. While performing the ETL, it was also decided to manually create a table in Excel itself containing the states and regions. This would be the table responsible for starting and facilitating the relationship between the others through the merging of queries, aiming to maintain the correct cardinality.

After the tables were standardized, some rows and columns were removed, column types were changed, and formulas were developed in M Language to transform values into percentages.

VISUALIZATION

The first Dashboard evaluates the socioeconomic profile by region, containing a button to navigate between regions, facilitating the extraction of insights.

The second dashboard makes a relationship between average income and all other indicators, but this time by Federative Unit, focusing on the 5 best and 5 worst states.

To create dashboards that facilitated user comprehension and were not cluttered with excess information, the following data was grouped:

Education Level: A complete school education was considered, that is, complete High School as a High Level of Education due to its impact on personal development, as well as being a minimum requirement for several job positions in the job market.

Access to sewage and water supply: The criteria used to define whether access is adequate or not was taken from the PLANSAB (National Basic Sanitation Plan), which indicates that adequate access to the sewage network should be considered as a population served by General or Pluvial Network, Septic Tank Connected to the Network, and Septic Tank Not Connected to the Network. For water supply, PLANSAB indicates that it is adequate for the population that has access via the General Distribution Network, Deep or Artesian Well, and Shallow Well, Freatic Well or Cistern.

Housing Condition (Indicator Dashboard): The percentages used for this dashboard are of people who own their own home, excluding those that are provided or rented. As a comparison is being made with Average Income, it would be valid to separate the population that already owns their own home from those who pay rent or still have their home provided in some way.

Assets in the Home (Indicator Dashboard): Assets in the Home were not included in this dashboard due to the size of the scope of the definitions of "Essential Assets" and "Luxuries" or "High-Value Assets". This scope changes depending on the state, method of acquisition, form of work, and, as the objective of this dashboard was not to include all assets by state due to the visual clutter it would cause, it was decided to leave this indicator out due to the impossibility of summarizing it.

NOTE: In addition to macros, both spreadsheets have buttons. In the Socioeconomic Profile Dashboard, you can navigate between regions, just as in the Indicator Dashboard, you can choose which indicator, in addition to the fixed one which is Average Income, you want to compare. There is also the possibility of displaying the tables for deeper exploratory analyses and evaluation of the formulas used in the ‘Dashboard Source’ spreadsheet. It contains dynamic tables, calculations, and formulas (INDEX, MATCH, VLOOKUP), where all the data used to create the interactive dashboards is concentrated.

**CONCLUSIONS**

\*\* Socioeconomic Profile Dashboard by Region: \*\*

The regions with the highest Average Income are those that have the highest level of education, as well as basic infrastructure. Some reasons that may be linked to these factors are that regions with higher average incomes tend to have more resources available to invest in infrastructure, including water supply systems and sewage networks. This can result in better living conditions and access to basic services.

Regions with higher average incomes also offer more job opportunities and higher salaries. This can motivate people to seek higher levels of education or attract people with higher levels of education from other regions, in an attempt to secure better and more stable jobs, which contributes to a higher level of education in these areas.

The biggest surprise of this dashboard is that there is a disparity between the regions with higher incomes and home ownership. The reasons that may lead to this factor are the cost of living and housing, which are higher in these regions. Another reason that could lead to this disparity is the presence of individuals who are transitioning to study or start their careers, preferring to rent properties instead of buying them. Housing policies in areas with lower incomes can also affect these indicators, as residents of lower-income regions may have greater ease in acquiring property.

In Assets in the Home, what is observed is that in the North and Northeast regions, the population has a significantly lower percentage of cars, computers, internet access, and washing machines, coinciding with the fact that they have a lower average income and consequently a lower purchasing power. It is worth noting that the opposite occurs for motorcycle ownership, which is a cheaper means of transportation.

\*\* Indicator Dashboard: \*\*

The Federal District stands out as the federative unit with the highest average income, approximately 35% higher than São Paulo, the second-placed in this indicator. It is responsible for raising the region's average, making it the highest in this category, at R$ 3,195.75. In addition, it is the only UF where more than 70% of the population has a high level of education.

This information complements the insights we had in the other dashboard and sheds light on new reasons for the correlation between average income and level of education, because for the Federal District, the causes may be a higher concentration of wealth and also the concentration of public services and government institutions.

Conversely, the Federal District has the population with the lowest home ownership rate in the country.

On the other hand, and for reasons already explained after evaluating the first Dashboard, we have Maranhão, with the 2nd lowest average income in the country, presenting the 2nd best housing condition, where more than 83% of the population owns their home.