**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**

**PÓS GRADUAÇÃO EM ESPECIALIZAÇÃO EM BUSINESS INTELLIGENCE**

Rodrigo Tofani Simões de Brito

COMPARATIVO DOS CENSOS DE 2000 E 2010 COM POWER BI

Belo Horizonte

2019

Rodrigo Tofani Simões de Brito

COMPARATIVO DOS CENSOS DE 2000 E 2010 COM POWER BI

Relatório Técnico apresentado ao programa de Especialização da Pontifícia Universidade Católica de Minas Gerais – IEC/PUC, no curso Business Intelligence, como requisito para a obtenção do certificado de Especialista em Business Intelligence.

Orientador: Prof. Rodrigo Vitorino Moravia

Belo Horizonte

2019

RESUMO

Este trabalho tem como objetivo explorar alguns dados dos Censos de 2000 e 2010, mais especificamente, sexo, raça e concentração populacional no campo e urbana utilizando as técnicas de Business Intelligence (BI), Extração, Carga e Transformação (ETL) e criação de um *dashboard* para facilitar a visualização, interpretação e comparação dos dados dos Censos de 2000 2010.

**Palavras-chave:** *Business Intelligence*, Censo, Dashboard, ETL, *OLAP*, *Power BI*.

**ABSTRACT**

This work aims to explore some data from the 2000 and 2010 Censuses, specifically, gender, race and population concentration in the urban and rural areas using the techniques of Business Intelligence (BI), Extraction, Loading and Transformation (ETL) and creation of a dashboard to facilitate the visualization, interpretation and comparison of the data of the Censuses of 2000 and 2010.

**Keywords:** *Business Intelligence*, Censo, Dashboard, ETL, *OLAP*, *Power BI*.

**LISTA DE ILUSTRAÇÕES**

**Figura 1:** Arquitetura de BI

**Figura 2:** Modelo Multidimensional

**Figura 3:** Crescimento Populacional entre 2000 e 2010

**Figura 4:** População por Raça em 2000 e 2010

**Figura 5**: População por Sexo em 2000 e 2010

**Figura 6:** População Rural x Urbana

**LISTAS DE ABREVEATURAS E SIGLAS**

BI Business Intelligence

DW Data Warehouse

ETL Extraction, Transformation and Loading

IBGE Instituto Brasileiro de Geografia e Estatística

OLAP On-line Analytical Processing

SIDRA Sistema IBGE de Recuperação Automática

**SUMÁRIO**

[1. INTRODUÇÃO 9](#_Toc4349199)

[2. CENSO 9](#_Toc4349200)

[3. BUSINESS INTELLIGENCE 10](#_Toc4349201)

[6. FERRAMENTAS 13](#_Toc4349202)

[7. METODOLOGIA 13](#_Toc4349203)

[8. VISUALIZAÇÕES NO POWER BI 14](#_Toc4349204)

[8.1. CRESCIMENTO POPULACIONAL 14](#_Toc4349205)

[8.2. POPULAÇÃO POR RAÇA 15](#_Toc4349206)

[8.3. POPULAÇÃO POR SEXO 16](#_Toc4349207)

[8.4. POPULAÇÃO URBANA x RURAL 17](#_Toc4349210)

[9. CONCLUSÃO 18](#_Toc4349213)

1. INTRODUÇÃO

Atualmente as grandes empresas estão passando por um grande problema, os dados gerados pelos seus sistemas de informação estão crescendo cada vez mais e isso torna difícil a extração de informações que possam ser usadas para tomada de decisão.

Para resolver este problema e os dados virarem informações, as empresas estão cada vez mais utilizando Business Intelligence (BI), um conjunto de conceitos e metodologias que visam a apoiar a tomada de decisões nos negócios, a partir da transformação do dado em informação e da informação em conhecimento.

O objetivo deste relatório técnico é mostrar como a utilização de um dashboard pode disponibilizar informações que facilitem a interpretação dos dados obtidos de forma bruta em uma maneira simples, utilizando como exemplo uma parte dos dados dos censos de 2000 e 2010 obtidos no website do Instituto Brasileiro de Geografia e Estatística (IBGE).

1. CENSO

O censo ou recenseamento demográfico é um estudo estatístico referente a uma população que possibilita o recolhimento de várias informações, tais como o número de homens, mulheres, crianças e idosos, onde e como vivem as pessoas. Esse estudo é realizado, normalmente, de dez em dez anos, na maioria dos países. (Wikipédia, 2011).

Além de proporcionar conhecimento do aumento demográfico brasileiro, com os dados obtidos das condições de vida da população o governo pode traçar planos de políticas públicas com o intuito de melhorar as condições de vida.

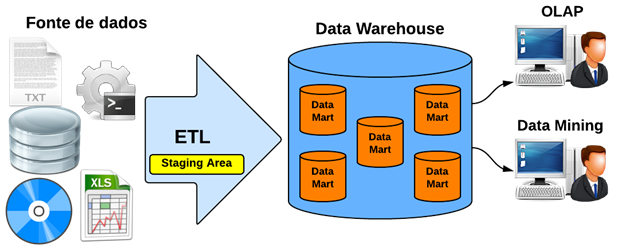
1. BUSINESS INTELLIGENCE

O conceito de Business Intelligence (BI), de forma mais ampla, pode ser entendido como a utilização de variadas fontes de informação para definir estratégias de competitividade nos negócios da empresa (BARBIERI, 2011).

As arquiteturas tradicionais de BI utilizam vários elementos e técnicas para transformação (processamento) de dados em informação (SILVA, 2011). De maneira mais ampla, pode‑se dividir a arquitetura de BI em três principais componentes:

* ETL (Extraction, Transformation and Loading), processo responsável por extrair os dados das bases operacionais (transacionais) da organização, efetuar transformações a fim de gerar informações válidas para a análise e apoio ao processo decisório e, por último, armazená‑las em um repositório que facilite o acesso às informações.
* Repositório de dados analíticos: são representados pelos Data Warehouses (DW), repositórios de dados que utilizam modelagens (geralmente modelagem dimensional), as quais podem dispor os dados de maneira mais natural para a análise e o processo de decisão.
* O último componente é a camada de apresentação. Essa camada pode‑se utilizar de uma série de técnicas e / ou ferramentas para auxiliar o consumo e apresentação das informações armazenadas pelo DW.

Figura 1 – Arquitetura de BI



Fonte: ELIAS, 2014

Fonte de dados: abrange todos os dados de origem que irão compor as informações do DW. Compreende os sistemas OLTP, arquivos em diversos formatos (XLS, TXT, etc), sistemas de CRM, ERP, entre vários outros.

**ETL**: o ETL, do inglês Extract, Transform and Load, é o principal processo de condução dos dados até o armazenamento definitivo no DW. É responsável por todas as tarefas de extração, tratamento e limpeza dos dados, e inserção na base do DW.

**Staging Area**: a Staging Area é uma área de armazenamento intermediário situada dentro do processo de ETL. Auxilia a transição dos dados das origens para o destino final no DW.

**Data Warehouse**: essa é a estrutura propriamente dita de armazenamento das informações decisivas. Apenas os dados com valor para a gestão corporativa estarão reunidos no DW.

**Data Mart**: o Data Mart é uma estrutura similar ao do DW, porém com uma proporção menor de informações. Trata-se de um subconjunto de informações do DW que podem ser identificados por assuntos ou departamentos específicos. O conjunto de Data Marts em conformidade dentro da organização compõe o DW.

**OLAP**: o OLAP, do inglês On-line Analytical Processing, na arquitetura de um DW se refere as ferramentas com capacidade de análise em múltiplas perspectivas das informações armazenadas.

**Data Mining**: Data Mining ou Mineração de Dados, se refere as ferramentas com capacidade de descoberta de conhecimento relevante dentro do DW. Encontram correlações e padrões dentro dos dados armazenados.

1. Modelo Multidimensional

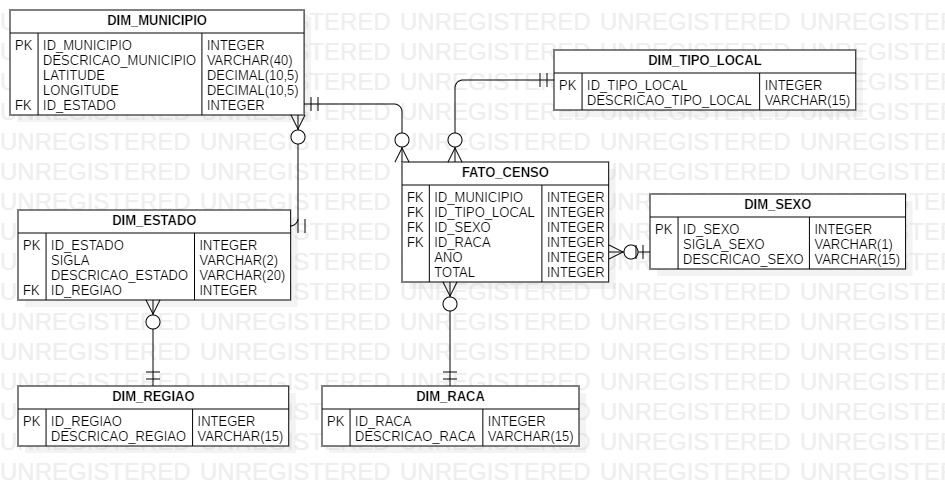
Geralmente, os dados que compõem um Data Warehouse são organizados numa estrutura chamada Multidimensional. Isto porque há uma estrutura principal de dados (fatos) e estruturas auxiliares (dimensões). (Loh,2014).

As tabelas fato contêm muitos registros de valores e medidas relacionadas ao negócio. Os valores são armazenados de acordo com a interseção de todas as dimensões.

As dimensões contêm descrições textuais das dimensões do negócio, e elas geralmente são utilizadas como variáveis de corte ou filtros.

A figura 2 mostra a modelagem dimensional que será usada neste trabalho.

**Figura 2: Modelagem Dimensional**



1. DASHBOARDS

Dashboards são painéis que mostram métricas e indicadores importantes para alcançar objetivos e metas traçadas de forma visual, facilitando a compreensão das informações geradas.

O objetivo dos Dashboards, é possibilitar o monitoramento dos resultados de uma empresa distribuídos em diversos indicadores. Para chegar a esse patamar é preciso responder perguntas essenciais para ter as respostas desejadas. Se você não souber o que perguntar, não será possível construir um Dashboard.

1. FERRAMENTAS

As ferramentas utilizadas neste trabalho foram, o banco de dados SQL Server e o Power BI, ambas da Microsoft.

O SQL Server foi usado para armazenar os dados que foram importadas das planilhas disponibilizadas pelo IBGE.

O Power BI é um serviço de análise de negócios da Microsoft. É um conjunto de ferramentas que tem como objetivo, fornecer visualizações interativas e recursos de BI com uma interface simples o suficiente para que os usuários finais criem seus próprios relatórios e painéis. O Power BI possibilita o usuário monitorar seu negócio através de diferentes fontes de dados, tudo consolidado em uma única dashboard.

1. METODOLOGIA

Trata-se de um relatório técnico contendo toda a metodologia de BI que foi utilizada para disponibilizar de maneira mais organizada e intuitiva os dados da população dos censos de 2000 e 2010 desde a importação, passando pela transformação e carga, até ser disponibilizado para o usuário.

Os dados dos censos dos anos de 2000 e 2010 foram baixados através do Sistema IBGE de Recuperação Automática (SIDRA), em uma planilha Excel e importado para o banco de dados SQL Server. Após a importação do arquivo para o banco de dados, foram executados scripts que fizeram a carga na tabela de fato e das tabelas de dimesão. O arquivo sidra.xlsx, o script de criação do banco e demais scripts, estão disponíveis em xxxxxxxxxxx

Como resultado deste relatório vamos disponibilizar um dashboard que exibirá informações da população brasileira baseada nos dados coletados nos censos dos anos 2000 e 2010.

1. VISUALIZAÇÕES NO POWER BI
   1. CRESCIMENTO POPULACIONAL

Como podemos ver na figura 3, atualmente o Brasil possui 190.755.799 habitantes. Esse é o resultado do Censo 2010 divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Em dez anos, o aumento da população foi de 11%, em números absolutos isso significa 20.882.945 pessoas.

A Região Sudeste ainda é a mais populosa do Brasil, com 80.364.410 pessoas. São Paulo é o estado mais populoso, com 41.262.199 pessoas. Já Roraima é o estado menos populoso, com 450.479 pessoas.

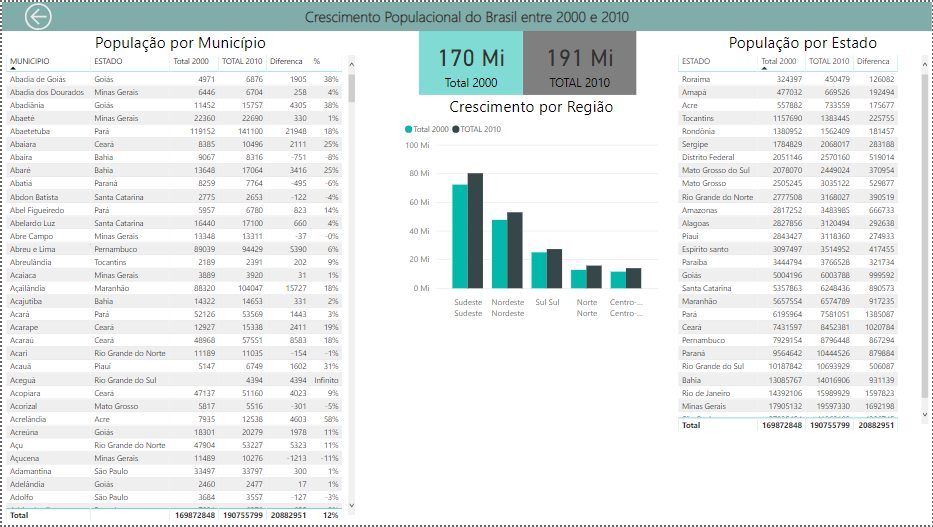
Houve mudanças no ranking dos dez municípios mais populosos do país, com Brasília, que pulou do 6º para 4º lugar e Manaus, que saiu do 9º para 7º, ganhando posições. De acordo com o IBGE, Brasília tem hoje 2.570.160 habitantes e Manaus, 1.802.014.

Belo Horizonte, Curitiba e Recife perderam posições na comparação com o censo realizado em 2000. No levantamento antigo, essas capitais ocupavam as posições de número 4, 7 e 8, respectivamente. Em 2010, passaram para as posições 6, 8 e 9.

Na relação dos municípios que tiveram maior crescimento no número de habitantes, aparecem ainda Rio das Ostras (190%), no Rio de Janeiro; Balbinos (182%), em São Paulo e Pedra Branca do Amapari (169%), no Amapá .

Outros 1.861 municípios apresentaram queda no número de habitantes. Os cinco que tiveram maior redução foram: Maetinga (BA), Itaúba (MT), Brejo de Areia (MA), Severiano Melo (RN) e Ribeirão do Largo (BA) .

**Figura 3: Crescimento Populacional entre 2000 e 2010**

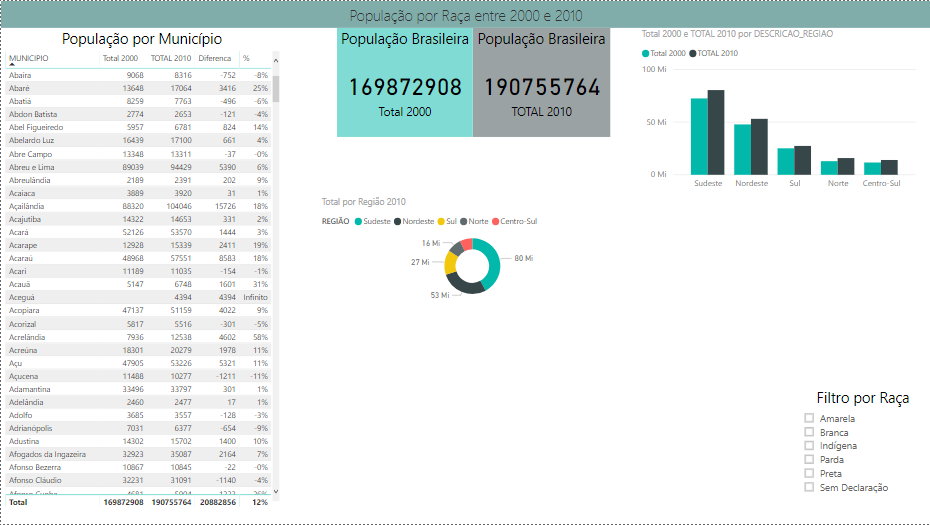


* 1. POPULAÇÃO POR RAÇA

Pela primeira vez na História do Censo, a população do Brasil deixa de ser predominantemente branca. Pelos dados de 2010, as pessoas que se declararam brancas são 47,50% da população, enquanto em 2000 eram 53,74%. Nos outros Censos, até agora, os brancos sempre tinham sido mais que 50%, superando a soma de pretos, pardos, amarelos e indígenas. A população branca foi, assim, a única que diminuiu. Passou de 91298094 em 2000 para 90.621.280 em 2010.

A população negra aumentou, indo de 10.554.323 em 2000 para 14.351.130. Já a parda, foi de 65.318.104 para 82.820.427. A parcela de indígenas cresceu de 734.155 para 821.493, e a amarela, de 761.578 para 2.105.376.

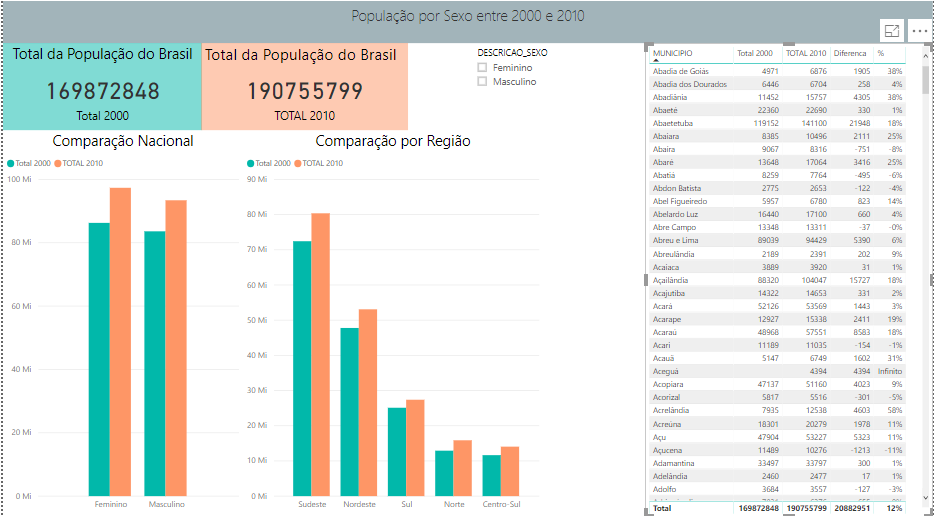
**Figura 4: População por Raça**



* 1. POPULAÇÃO POR SEXO

A população brasileira é formada de 97.34.809 mulheres e 93.406.990 homens. Entre os municípios, o que tinha maior percentual de homens era Balbinos (SP), com 82,2%. Já o que tinha maior percentual de mulheres era Santos (SP), com 54,25%.

**Figura 5: População por Sexo**

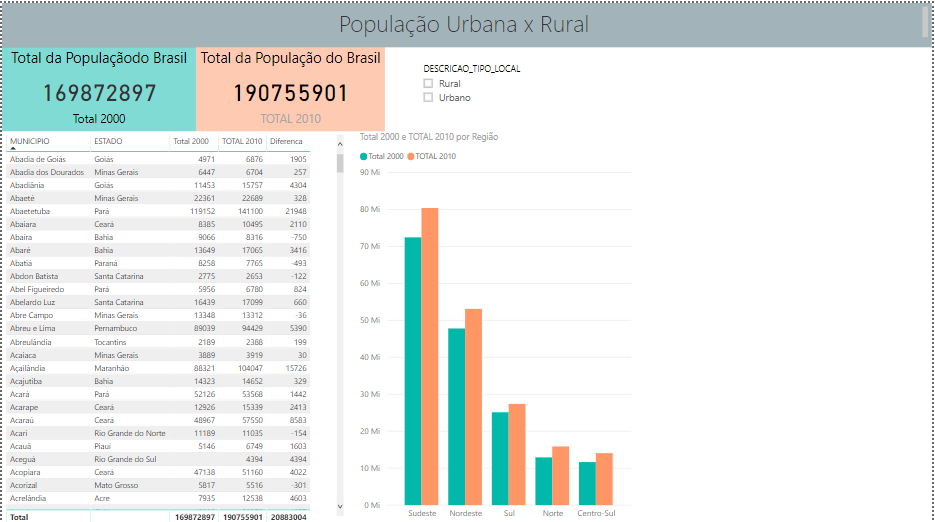


* 1. POPULAÇÃO URBANA x RURAL

A população está mais urbanizada que há 10 anos. Em 2000, 81% dos brasileiros, ou 137.925.255, viviam em áreas urbanas, agora são 84%, que representam 160.934.709.

Em 2010, entre os municípios, 69 possuem 100% de sua população vivendo em situação urbana. Em 2000, entre os municípios, 60 tinham 100% de sua população vivendo em situação urbana e o único município do país a ter 100% de sua população em situação rural era Nova Ramada (RS).

**Figura 6: População Rural x Urbana**



1. CONCLUSÃO

Este relatório técnico demonstrou um exemplo do uso de Business Intelligence, apresentando dados da população brasileira levantados nos censos de 2000 e 2010 utilizando tabelas, gráficos e dashboards na ferramenta Power BI.

Como podemos ver, exploramos apenas as características de raça, sexo e urbanização e verificamos que já houve mudanças significativas na população.

Para trabalhos futuros, podem ser acrescentados vários outros aspectos como renda, escolaridade e religião dentre outros e logo os dados do próximo censo, o de 2020.

**REFERÊNCIAS**

BARBIERI, Carlos. BI2 - Business Intelligence: Modelagem e Qualidade. Rio de Janeiro: Elseiver, 2011.

ELIAS, Diego. Conhecendo a arquitetura de Data Warehouse, 2014. Disponível em < <https://canaltech.com.br/business-intelligence/conhecendo-a-arquitetura-de-data-warehouse-19266/>>. Acessado em 03 Mar. 2019.

LOH, Stanley. BI na Era do Big Data para Cientistas de Dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões. Porto Alegre, 2014.

SILVA, Dhiogo Cardoso da. Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural. 2011. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2011.

WIKIPÉDIA. Censo Demográfico. 2011. Disponível em: <http://pt.wikipedia.org/wiki/Censo\_demográfico>. Acessado em: 15 Mar. 2019.