

Análise BI em um Banco de Dados do ENEM

Ana Paula de Oliveira, José Weverton, Raíssa Pimenta, Alexandre Galvani

Curso de Tecnologia em Banco de Dados (FATEC - BAURU)
Rua Manoel Bento da Cruz, nº30 - Quadra 3 – Centro – 17.015-171 –
Bauru – SP - Brasil

anapauladeoliveira27@gmail.com, jose.pires3@fatec.sp.gov.br,
raissa_pimenta@hotmail.com, alexandre.galvani@fatec.sp.gov.br

***Abstract.** Business Intelligence Analysis (BI) is increasingly popular in companies, doing an analysis of their information and asymmetric and proportions corrections. The objective of this study is to use the BI and other Data Warehouse resources in a database of the National High School Examination (ENEM) and the results obtained from the research for the growth of the number of mentally aged individuals, carry out a comparison of genera and regions of the state of São Paulo during the years 2010 to 2014. It is conclusive that the growth of some indices is relative to the amount of inscriptions and the order in which the candidates will be perceived in these inscriptions, may not all regions who have as correct knowledge about the exam. A comparison between the ages and genders were the most satisfactory, the female and the elderly were the ones that stood out the most over the years and the candidates with difficulties did not reach expectations of increase.*

***Resumo.** A Análise Business Inteligence (BI) está cada vez mais popular nas empresas, fazer uma análise de suas informações e assim trazer falhas e propor correções. O objetivo deste estudo é utilizar a análise BI e outros recursos do Data Warehouse em um banco de dados do Exame Nacional do Ensino Médio (ENEM) e com os resultados obtidos da pesquisa desenvolver relatórios para observar o crescimento de idosos, portadores de deficiência física ou mental, realizar uma comparação de gêneros e das regiões do estado de São Paulo no decorrer dos anos de 2010 até 2014. É conclusivo que o crescimento de alguns tópicos é relativo à quantidade de inscritos e a ordem em que os candidatos foram realizando estas inscrições, pode-se afirmar que não são todas as regiões que tem as informações corretas sobre o exame. A comparação entre as idades e gêneros, foram as mais satisfatórias, o sexo feminino e os idosos são os que mais se destacaram no decorrer dos anos e aos candidatos portadores de deficiência não alcançaram as expectativas de aumento.*

1. Introdução

Segundo Oliveira (2018), o Enem foi concebido originalmente para ser uma avaliação individual dos alunos do ensino médio em 1998. No início dos anos 1990, na gestão do ex-presidente Fernando Henrique Cardoso, a expectativa do então ministro da Educação Paulo Renato Souza era criar um teste nacional diferenciado. A participação inicial em 1998 foi tímida com cerca de 115 mil inscritos, mas superação foi rápida. Em 2001, o exame ultrapassou a marca de 1 milhão de inscritos e em 2014 teve uma marca histórica

de 8,7 milhões de participantes inscritos. Oliveira (2018) ainda ressalta a importância que o exame ganhou em apenas alguns anos de existência, ele descreve que já em 1999 um ano apenas pós criado, o exame já era utilizado como critério de acesso aos cursos de graduação como a PUC-RJ e a Universidade Federal de Ouro Preto (Ufop) que foram as pioneiras em utilizar o sistema do Enem. O autor também ressalta que no ano de 2000 o exame se operacionaliza para atender pessoas com necessidades especiais, passando a oferecer prova em braile, prova ampliada, auxílio para leitura e transcrição e tradutor/intérprete em libras. Em 2001 a política de isenção na inscrição para concluintes do ensino médio é aplicada. Já em 2004 o resultado individual do Enem passa a ser critério de participação dos inscritos a bolsas de estudo integral ou parcial em cursos de graduação de instituições privadas por meio do Programa Universidade para Todos (ProUni). No ano de 2009 foi realizada a criação do sistema de seleção unificada (Sisu) e adoção do novo Enem. Em 2010 o exame começa a ser aplicado para presidiários (Enem PPL). Em 2013 as instituições federais de ensino superior passam a usar o Enem como critério de seleção para novos alunos e concessão de bolsas de estudos do programa Ciência sem Fronteiras. No ano de 2014 foi assinado o primeiro acordo interinstitucional com uma instituição de educação superior portuguesa, a Universidade de Coimbra, para uso das notas do Enem no acesso à vagas a universidade. Desde então o Exame vem sendo melhorado e atualizado a cada ano.

O trabalho descreve como obter informações concretas e válidas para tomadas de decisões que impactem positivamente em futuros exames. Hoje o desafio de grandes empresas é justamente trabalhar com decisões antecipadas que os levem ao êxito. Um exemplo, segundo o blog Think Consulting (2018), a Alemanha campeã da copa de 2014 utilizou uma ferramenta em BI para coletar dados dos jogadores como velocidade em campo, quilômetros percorridos e números de jogadas. Com essas informações a comissão técnica realizou a tomada de decisão de quais jogadores estavam mais aptos fisicamente e taticamente para compor o time titular para a próxima partida, além de saber quais formações poderiam ser otimizadas em jogo, assim como, quais fraquezas deveriam ser debeladas e quais qualidades exploradas. O blog também descreve sobre a empresa Avon utiliza de análise BI para tomadas de decisões de vendas no Brasil, foi descrito que para aumentar a eficiência da Avon Brasil, a empresa realizou a implantação do BI durante 18 meses e a partir da implantação do Data Warehouse e a adoção de métricas de desempenho do negócio, foi possível realizar o monitoramento das estratégias traçadas, identificar oportunidades de novos negócios e realizar uma mudança na cultura da empresa.

Foram coletados a base pública do Exame Nacional do Ensino Médio (ENEM) através do Portal Brasileiro de Dados Abertos que é uma ferramenta disponibilizada pelo governo. Utilizou-se para a realização desta pesquisa dados públicos do estado de São Paulo dos anos de 2010 a 2014 para então realizar uma análise BI onde o intuito da pesquisa é analisar o crescimento de alguns tópicos que são considerados importantes no decorrer dos anos, esses tópicos são: regiões do estado de São Paulo, percentual de gênero, idosos e candidatos portadores de algum tipo de deficiência física ou mental.

Portando é possível afirmar que análise de dados possibilita ter uma técnica analítica que busca obtenção de uma grande quantidade de dados sobre um determinado assunto para obter respostas e soluções. Por conta disso pode-se usar a análise de BI para obter informações e percentual de crescimento podendo assim examinar os dados como um todo ou dividi-los em amostras individuais como, por exemplo, o modelo

Estrela que é uma parte importante do Data Warehouse com ele é possível separar os dados em fato principal e dimensões então utiliza-los na análise BI. O que torna possível a aplicação do trabalho com dados do Enem, pois, serão obtidas informações que possibilitarão tomadas de decisões para melhorias do exame.

2. Análise Business Intelligence (BI)

O avanço da tecnologia computacional permite o acesso a um grande volume de dados de diferentes origens e formatos com base em Angeloni e Reis (2006) o conceito Business Intelligence (BI) que traduzido para o português entende-se de Inteligência de Negócios ou Inteligência Empresarial que significa um conjunto de metodologias utilizadas em gestão através de softwares específicos, sua função é propor ganhos nos processos das empresas e da alta administração nas organizações. Angeloni e Reis (2006) também afirmam que o BI tem a capacidade analítica que utilizam ferramentas onde armazenam as informações em um só lugar transformando dados em conhecimento.

Segundo Moraes (2018) BI se resume em adquirir e armazenar os dados do banco de dados para então analisa-los e conseguir obter informações sobre os investimentos avaliando se está trazendo resultados positivos e auxiliando nos objetivos corretos para assim diminuir perdas, evitando sair da linha de foco e reduzindo em custos operacionais. Moraes (2018) ressalta que o BI também consiste em um conjunto de processos que tem por objetivo de entregar informações certas e é dividido em três pilares: estratégia, analítica e operacional.

- a) A estratégia se baseia na coleta de dados onde são analisados o processo de negócio, produtividade, concorrência no mercado, entre outros;
- b) Analítica consiste em organizar e fazer uma análise nos dados para então realizar uma tomada de decisão;
- c) Operacional consta com a parte da ação, o monitoramento dos dados ajuda na escolha para que seus resultados sejam bem-sucedidos.

Moraes (2018) também alega que o BI é muito utilizado em ambientes corporativos que aperfeiçoam os processos de maneira eficaz e direta com a intuição de encontrar falhas e oportunidades de melhorias para assim conseguirem sair do mercado saturado onde existem diversas empresas oferecendo produtos sem seu devido valor. A análise BI pode ser utilizada em diversas maneiras, outra forma é na exploração de dados com a proposta de realizações de pesquisas nas informações.

Bittencourt (2015) afirma que os sistemas Enterprise Resource Planning (ERP) são utilizados para padronizar as informações e centralizar todos os dados de maneira clara e organizada dentro de um único banco de dados. As informações são alimentadas e atualizadas frequentemente por isso auxilia no gerenciamento de suas atividades fazendo assim a diminuição do trabalho manual obtendo um aumento na produtividade. Utilizar as soluções ERP na análise BI possibilitam o cruzamento de dados com o objetivo de emitir relatórios precisos com informações necessárias.

3. Data Warehouse

Siqueira (2014) alega que Data Warehouse é um sistema de dados históricos utilizados no processo de tomada de decisão e por conta disso integra os dados corporativos de

uma empresa em um único local. Também é utilizado na criação de visão única e centralizada das informações contidas dentro do banco de dados e com isso os usuários conseguem fazer consultas, gerar relatórios e realizar análises.

O autor também afirma que para um banco de dados ser considerado um Data Warehouse ele precisa coletar dados de várias fontes, os dados coletados devem ser transformados para que ocorra uma visão única desses dados e os mesmos devem ser usados por aplicativos para obter informações que possam auxiliar na tomada de decisão. Na Figura 1 uma comparação de um banco de dados operacional e um Data Warehouse.

	Banco de Dados Operacional	Data Warehouse
Usuários	Funcionários	Alta administração
Utilização	Tarefas cotidianas	Decisões estratégicas
Padrão de Uso	Previsível	Difícil de prever
Princípio de funcionamento	Com base em transações	Com base em análise de dados
Valores dos dados	Valores atuais e voláteis	Valores históricos e imutáveis
Detalhamento	Alto	Sumarizado
Organização dos dados	Orientado a aplicações	Orientado a assunto

Figura 1. Comparação de um banco de dados operacional e um Data Warehouse.

Fonte. Elaborado pelos autores, 2018.

De acordo com Elias (2014) o objetivo do Data Warehouse é armazenar os dados de forma organizada que permite a análise de um grande volume de informações. Sua base de dados também é utilizada para fazer parte de uma tecnologia do BI.

O autor segue afirmando que essa tecnologia consiste em uma arquitetura dimensional onde pode possuir diversas dimensões e fatos com diversos componentes cujo é chamado de multidimensional. Essas dimensões são ligadas a um ou mais fatos sendo assim elementos presentes e importantes dentro desta estrutura que permite o armazenamento dos dados, desempenho e intuitividade nas análises das informações gerenciadas pelo Data Warehouse.

Elias (2014) também relata que o Data Warehouse trabalha com o modelo Estrela onde sua estrutura é feita de fatos e dimensões que são armazenados em forma de tabela, para um banco ser considerado um modelo Estrela ele precisa ser no mínimo três dimensões. O fato é a tabela principal onde ocorrem as ligações das tabelas dimensões através de chaves estrangeiras de um banco de dados. A tabela fato possui características quantitativas dentro do Data Warehouse onde é possível extrair dela métricas que são cruzadas com os dados das tabelas Dimensões e assim apresentando informações significativas para ocorrer uma análise mais precisa feita pelo usuário em base no assunto pretendido.

A estrutura que faz parte do Data Warehouse e que facilita na realização das análises necessárias é o OLAP do português "processamento analítico on-line. Segundo Microsoft (2016) o OLAP consiste em cubos que são recursos do Service Manager e usam a infraestrutura que existe no Data Warehouse para fornecer recursos de BI. Microsoft (2016) afirma que um cubo OLAP é uma estrutura de dados que suporta limitações e proporciona rápida análise dos dados. Os cubos exibem e somam grandes

quantidades de dados enquanto os usuários conseguem fazer pesquisa de qualquer assunto dentro do banco de dados. Estas informações podem ser acumuladas, segmentadas e separadas conforme a necessidade e interesse do usuário.

4. Análise de Processos

Para HEFLO (2016) a análise de processos se baseia em uma ação de conduzir uma revisão e adquirir um entendimento melhor sobre os processos de negócio. Essa revisão envolve componentes de entradas, saídas, processamentos, controles, atores, aplicações, dados, tecnologias e interações que auxiliam nos resultados. HEFLO (2016) segue afirmando que este tipo de análise abrange diversas avaliações como: tempo, custo, capacidade e qualidade destes processos. Referente a coleta de dados, é uma ferramenta de importância para o início e o fim das atividades. Na Figura 2 demonstra as etapas de um processo de análise.

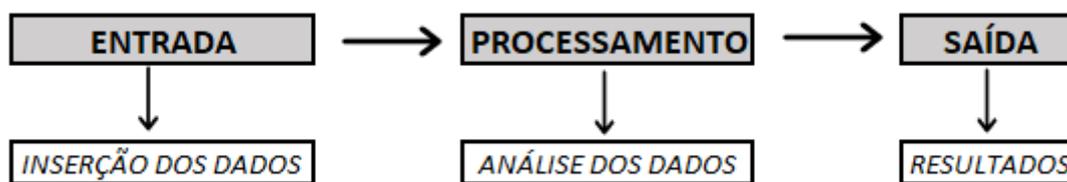


Figura 2. Exemplo de uma análise de processos.

Fonte. Elaborado pelos autores, 2018.

Santos (2016) afirma que a análise dos processos é uma visualização detalhada dos processos existentes que atendem as propostas de um cliente e também afirma que para uma análise obter um resultado positivo ela precisa passar por três tipos de fase. A primeira é composta por uma exploração que é a investigação dos dados onde é possível entender o que pode aprender com esse tipo de análise. A seguinte fase é a de gerar hipóteses sobre as causas se referindo na identificação de possíveis defeitos. E a última é a fase de testes onde utilizam os dados para verificar quais causas contribuem para algum problema.

Scartezini (2009) afirma que o mapeamento de processos é uma ferramenta gerencial analítica que cuja intenção é auxiliar na melhoria dos processos e também é possível utiliza-lo para implantação de estrutura voltada para os processos, esta análise estruturada permite a redução de custos, falhas de integração nos sistemas e melhora o desempenho da organização sendo assim podendo fazer uma análise nos processos de forma ágil.

5. Banco de Dados

Para Meira (s/d) um banco de dados tem como objetivo organizar uma informação, normalmente agrupando informações que serão utilizadas para um mesmo fim, de forma que possam representar coleções de informações que se relacionam de forma que crie um sentido, então podemos dizer que um banco de dados é uma coleção de dados relacionados, sendo assim toda informação que pode ser armazenada e que apresenta algum significado dentro do contexto ao qual ele se aplica.

Segundo Lucidchart (2018) o sistema de um banco de dados é informatizado e acessado por meio de um software mais conhecido como Sistema Gerenciador de Banco de Dados (SGBD). Um SGBD é composto por vários programas que permitem construir e manipular bases de dados.

Lucidchart (2018) segue afirmando que para trabalhar com essas tabelas, algumas restrições precisaram ser impostas para evitar aspectos indesejáveis, como: repetição de informação, incapacidade de representar parte da informação e perda de informação.

CANDIDATO		
ID_CANDIDATO	IDADE	NOTA_MEDIA
12.652	20	569.55
12.653	25	635.78
12.654	19	387.75

LOCAL	
ID_LOCAL	REGIAO
1	NORTE
2	NORDESTE
3	NOROESTE

Figura 3. Exemplo de banco relacional.

Fonte. Elaborado pelos autores, 2018.

Lucidchart (2018) ainda relata que o modelo Relacional é uma teoria matemática criada por Edgar Frank Codd em 1970 para descrever como as bases de dados devem funcionar. É um dos mais utilizados por apresentar uma grande flexibilidade. Esse modelo classifica os dados por tabela, conhecidas como relações, cada uma das quais consiste em colunas e linhas como representado na Figura 3. O modelo também explica os tipos de relações entre essas tabelas, incluindo relações uma para uma, uma para muitas e muitas para muitas.

6. Materiais e Métodos

Para este projeto foi utilizado uma base pública contendo informações sobre o Exame Nacional do Ensino Médio (ENEM) que foi disponibilizado pelo Portal Brasileiro de Dados Abertos. Foi necessária a separação destes dados para apenas obter os candidatos que realizaram as provas no estado de São Paulo dos anos de 2010 até o ano de 2014. A pesquisa consistiu em fazer uma análise BI utilizando o modelo Estrela que permite armazenar uma grande quantidade de informações e assim trazer dados relevantes para a criação de relatório que foram utilizados na conclusão deste projeto. Coletou-se cinco mil dados de cada ano, no total de vinte e cinco mil informações onde analisou-se apenas as tabelas que contribuíram para então obter informações somente dos dados necessários.

O modelo Estrela é composto por um fato principal que é rodeado por outras tabelas de dimensões, ficando parecido com uma estrela. Esse modelo ocupa mais espaço em disco porém é mais fácil de implementar e acabou sendo utilizado porque além de hoje em dia o espaço ser muito barato, ele permite entregar projetos por pedaços.

Após a coleta dos dados realizou-se a estruturação das tabelas no banco de dados de acordo com a elaboração do modelo Estrela representado na Figura 4. A criação das tabelas e a inserção dos dados obtidos na base pública foram realizadas no banco de dados Oracle Server que foi virtualizada no Oracle Virtual Box, software de virtualização de sistemas.

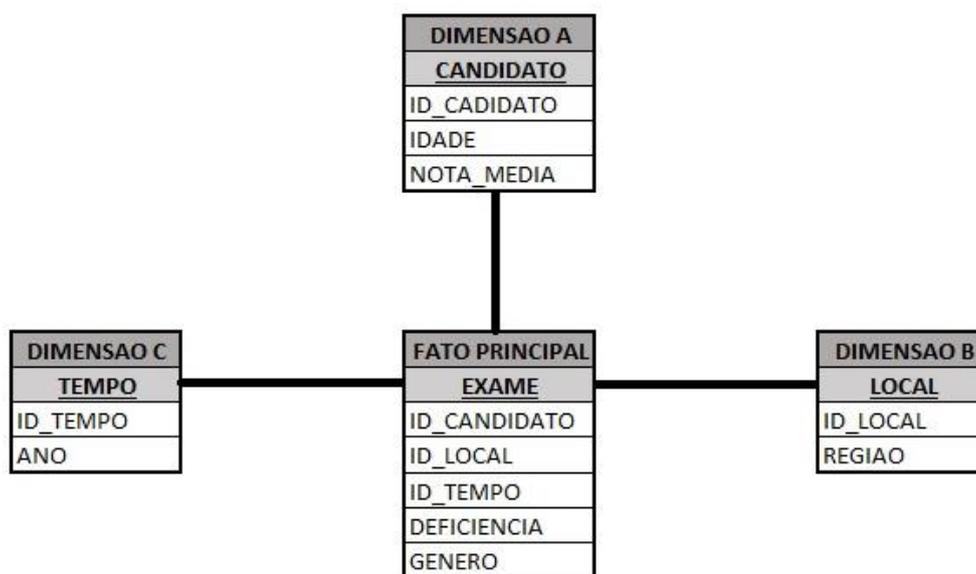


Figura 4. Exemplo de um modelo Estrela.

Fonte. Elaborado pelos autores, 2018.

Este projeto foi desenvolvido em três tabelas dimensões, são elas: Candidato, armazenando a idade e a média final de cada um deles, seguindo com a tabela Local armazenando as regiões do estado de São Paulo, a tabela Tempo para obtermos o ano em que foi realizado o ENEM, seguindo assim com o fato principal sendo a tabela Exame e fazendo a ligação com as tabelas dimensões através de chaves estrangeiras

ainda na tabela do fato principal foram acrescentadas mais duas informações: Gênero, se o candidato é masculino ou feminino e Deficiência caso o candidato tenha ou não alguma.

Todas as dimensões e o fato principal foram alimentados com dados importados de suas respectivas tabelas, que estavam armazenadas em arquivos no formato CSV. Foram utilizadas as importações dos arquivos por dois motivos. Primeiro, é um formato compatível com banco de dados do Oracle, ou seja, evita a necessidade de realizar criação *PL/SQL* e *Procedures*, o que consome tempo para a análise de verificação da forma que os dados estão ordenados. Segundo, possibilita visualização das informações antes da importação, como por exemplo, no Excel, ferramenta que foi utilizada para filtrar informações desnecessárias antes da importação para a base de dados.

A análise BI foi desenvolvida no Power BI Desktop que é um software disponibilizado pela Microsoft, por sua vez essa tecnologia traz uma solução de análise de negócios que permite que você visualize seus dados e compartilhe soluções para a organização. Power BI permite conectar seus dados de onde quer esteja e faça a exploração de seus dados de forma interativa, através dele realizou-se uma conexão com o banco de dados que teve as informações do ENEM em um modelo Estrela, o programa realizou a análise necessária de todos os dados solicitados para então ao fim utilizar a análise de processo desta base de dados pública e disponibilizar as informações através de uma interface intuitiva e dinâmica para cruzamento de dados, que provê visualização gráfica, permitindo utilizar sua percepção visual para melhor analisar, compreender as informações e gerar um relatório final com a comparação dos anos, todas as tabelas dimensões possuem “id”.

Foi feito o armazenamento dos dados em forma de linhas e colunas, inseridos no Power BI onde foram separados os dados que foram utilizados e foram criados as dimensões para gerar os gráficos de forma que o programa mostrasse de forma de fácil visualização e entendimento para analisar possíveis erros e corrigi-los de forma mais efetiva e eficaz..

7. Resultados

Para o prosseguimento do desenvolvimento desta pesquisa foi necessário à separação dos dados no Excel para então atribuir as dimensões no fato principal baseado no modelo Estrela que foi citado na Seção 6, foram desenvolvidos *scripts* em SQL para gerar a criação das tabelas dentro do bando de dados Oracle Server. As Figuras a seguir descrevem a criação cada uma delas, todas as dimensões com padrão de possuir “id” para realizar a conexão com a tabela Exame por chaves estrangeiras. A Figura 5 representa a criação da tabela Candidato.

```
CREATE TABLE CANDIDATO(  
    ID_CANDIDATO NUMBER PRIMARY KEY,  
    IDADE NUMBER(3) NOT NULL,  
    NOTA_MEDIA NUMBER (5,2) NOT NULL  
);
```

Figura 5. Representação da Dimensão A referente à tabela Candidato.

Fonte. Elaborado pelos autores, 2019.

A Figura 6 representa a criação da tabela Local para receber as sete regiões do estado de São Paulo.

```
CREATE TABLE LOCAL (  
    ID_LOCAL NUMBER PRIMARY KEY,  
    REGIAO VARCHAR2 (120) NOT NULL  
);
```

Figura 6. Representação da Dimensão B referente à tabela Local.

Fonte. Elaborado pelos autores, 2019.

Em seguida na Figura 7 é referente tabela Tempo para receber os cinco anos da pesquisa.

```
CREATE TABLE TEMPO (  
    ID_TEMPO NUMBER PRIMARY KEY,  
    ANO NUMBER (4) NOT NULL  
);
```

Figura 7. Representação da Dimensão C referente à tabela Tempo.

Fonte. Elaborado pelos autores, 2019.

A Figura 8 demonstra a tabela Exame além de fazer as ligações com as dimensões também colhe gênero e deficiência de cada candidato.

```
CREATE TABLE EXAME (  
    ID_CANDIDATO NUMBER NOT NULL,  
    ID_LOCAL NUMBER NOT NULL,  
    ID_TEMPO NUMBER NOT NULL,  
    GENERO CHAR (1) NOT NULL,  
    DEFICIENCIA CHAR (1) NOT NULL,  
    CONSTRAINT CDT_ENEM FOREIGN KEY (ID_CANDIDATO) REFERENCES CANDIDATO (ID_CANDIDATO),  
    CONSTRAINT LCL_ENEM FOREIGN KEY (ID_LOCAL) REFERENCES LOCAL (ID_LOCAL),  
    CONSTRAINT TMP_ENEM FOREIGN KEY (ID_TEMPO) REFERENCES TEMPO (ID_TEMPO),  
    CONSTRAINT CK_GENERO CHECK (GENERO IN('M','F')),  
    CONSTRAINT CK_DEFICIENCIA CHECK (DEFICIENCIA IN('S','N'))  
);
```

Figura 8. Representação do Fato Principal referente à tabela Exame.

Fonte. Elaborado pelos autores, 2019.

Todas as informações foram descritas como *not null*, ou seja, o campo não pode estar vazio, é de extrema importância popular todos os campos para que não ocorra erros quando os dados necessários forem inseridos no banco. Após a criação das tabelas foi inserido as informações no banco de dados nas dimensões local e tempo por terem poucas informações foram possíveis inseri-las através de scripts sem a necessidade de utilizar o Excel, as Figuras 9 e 10 demonstram como foi desenvolvido.

```

INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (1, 'NORTE');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (2, 'NORDESTE');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (3, 'NOROESTE');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (4, 'SUL');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (5, 'SUDESTE');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (6, 'OESTE');
INSERT INTO LOCAL (ID_LOCAL, REGIAO) VALUES (7, 'CENTRAL');

```

Figura 9. Representação do *script* de inserção da tabela Local.

Fonte. Elaborado pelos autores, 2019.

```

INSERT INTO TEMPO (ID_TEMPO, ANO) VALUES (1, 2010);
INSERT INTO TEMPO (ID_TEMPO, ANO) VALUES (2, 2011);
INSERT INTO TEMPO (ID_TEMPO, ANO) VALUES (3, 2012);
INSERT INTO TEMPO (ID_TEMPO, ANO) VALUES (4, 2013);
INSERT INTO TEMPO (ID_TEMPO, ANO) VALUES (5, 2014);

```

Figura 10. Representação do *script* de inserção da tabela Tempo.

Fonte. Elaborado pelos autores, 2019.

Para a inserção das tabelas candidato e exame é necessário o uso do Excel, pois, o mesmo permite grande volume de dados e facilita no preenchimento das tabelas, as tabelas foram separadas em linhas e colunas, divididas na tabelas Exame e Candidatos, ambas foram salvas em formato CSV que é o formato aceito pelo banco de dados da Oracle e assim importadas no banco de dados, após cada *create table* e *insert* foi executado o comando *commit* para efetivar os dados.

Prosseguindo ainda no Oracle Server foi realizada a criação de scripts de *select* utilizando a função *group by cube* através dele podemos demonstrar apenas a quantidade das informações desejadas, o cubo foi desenvolvido conforme os tópicos citados anteriormente que são: avaliar quantidade de idosos, portadores de deficiência, fazer comparação dos candidatos por gêneros e por regiões. A Figura 11 mostra como foi desenvolvido o script com base em obter informações sobre pessoas com mais de sessenta anos no ano de 2010, para avaliação dos outros anos e idades apenas é necessário trocar o “id” do tempo e o valor da idade desejada e isso será mantido no restante deste projeto.

```

SELECT C.IDADE, T.ANO, COUNT(E.ID_CANDIDATO)
FROM EXAME E
INNER JOIN CANDIDATO C
ON C.ID_CANDIDATO = E.ID_CANDIDATO
INNER JOIN TEMPO T
ON T.ID_TEMPO = E.ID_TEMPO
WHERE C.IDADE BETWEEN 60 AND 100 AND T.ID_TEMPO = 1
GROUP BY CUBE (C.IDADE, T.ANO);

```

Figura 11. Representação do *script group by cube* referente a candidatos com mais de sessenta anos no ano de 2010 representado pelo “id” 1.

Fonte. Elaborado pelos autores, 2019.

A Figura 12 mostra o desenvolvimento do *script* para avaliarmos candidatos portadores de deficiência (seja ela física ou mental) representados pela letra 'S' no ano de 2011.

```
SELECT C.DEFICIENCIA, T.ANO, COUNT(E.ID_CANDIDATO)
FROM EXAME E
INNER JOIN CANDIDATO C
ON C.ID_CANDIDATO = E.ID_CANDIDATO
INNER JOIN TEMPO T
ON T.ID_TEMPO = E.ID_TEMPO
WHERE T.ID_TEMPO = 2 AND E.DEFICIENCIA = 'S'
GROUP BY CUBE (C.DEFICIENCIA, T.ANO);
```

Figura 12. Representação do *script group by cube* referente a candidatos com deficiência no ano de 2011 representado pelo "id" 2.

Fonte. Elaborado pelos autores, 2019.

Na Figura 13 está a demonstração da quantidade de candidatos do gênero feminino representado pela letra 'F' no ano de 2012 representado pelo "id" 2.

```
SELECT E.GENERO, T.ANO, COUNT(E.ID_CANDIDATO)
FROM EXAME E
INNER JOIN CANDIDATO C
ON C.ID_CANDIDATO = E.ID_CANDIDATO
INNER JOIN TEMPO T
ON T.ID_TEMPO = E.ID_TEMPO
WHERE T.ID_TEMPO = 3 AND E.GENERO = 'F'
GROUP BY CUBE (C.GENERO, T.ANO);
```

Figura 13. Representação do *script group by cube* referente a candidatos do gênero feminino no ano de 2012 representado pelo "id" 3.

Fonte. Elaborado pelos autores, 2019.

Figura 14 demonstra a quantidade de candidatos do gênero masculino representado pela letra 'M' no ano de 2013.

```
SELECT E.GENERO, T.ANO, COUNT(E.ID_CANDIDATO)
FROM EXAME E
INNER JOIN CANDIDATO C
ON C.ID_CANDIDATO = E.ID_CANDIDATO
INNER JOIN TEMPO T
ON T.ID_TEMPO = E.ID_TEMPO
WHERE T.ID_TEMPO = 4 AND E.GENERO = 'M'
GROUP BY CUBE (C.GENERO, T.ANO);
```

Figura 14. Representação do *script group by cube* referente a candidatos do gênero masculino no ano de 2013 representado pelo "id" 4.

Fonte. Elaborado pelos autores, 2019.

```

SELECT L.LOCAL, T.ANO, COUNT(E.ID_CANDIDATO)
FROM EXAME E
INNER JOIN CANDIDATO C
ON C.ID_CANDIDATO = E.ID_CANDIDATO
INNER JOIN TEMPO T
ON T.ID_TEMPO = E.ID_TEMPO
INNER JOIN LOCAL
ON L.ID_LOCAL = E.ID_LOCAL
WHERE T.ID_TEMPO = 5 AND L.ID_LOCAL = 4
GROUP BY CUBE (L.ID_LOCAL, T.ANO);

```

Figura 15. Representação do *script group by cube* referente a região Sul do estado de São Paulo referente ao ano de 2014 representado pelo “id” 5.

Fonte. Elaborado pelos autores, 2019.

Após a avaliação da quantidade dos tópicos desejados no banco de dados Oracle, foi necessário o uso da ferramenta Power BI Desktop, os gráficos CSV que foram salvos através do Excel foram importados para esta ferramenta e então foi possível gerar os gráficos intuitivos e realizar as comparações dos tópicos acima. As Figuras a seguir representam os gráficos baseados no cubo desenvolvido anteriormente.

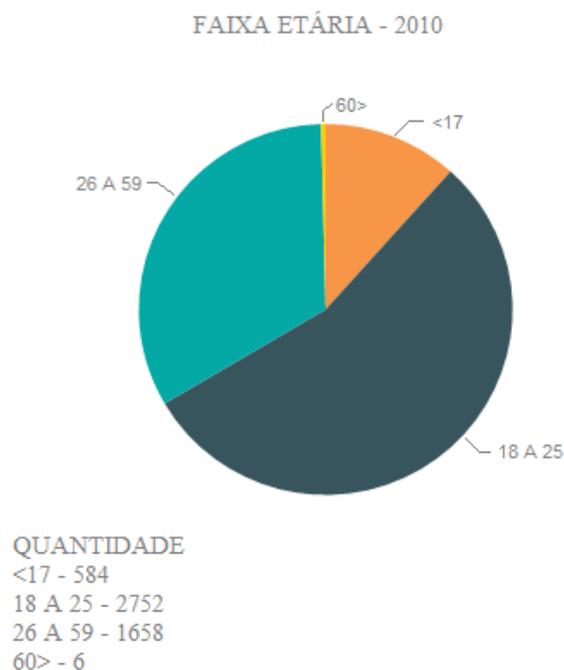
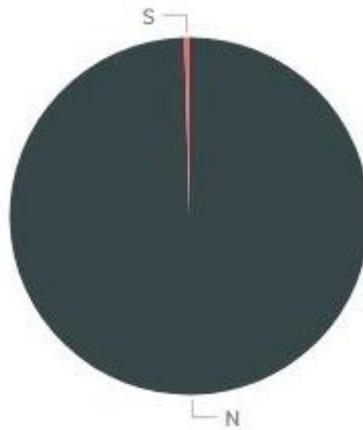


Figura 16. Representação do gráfico de idade separada por faixa etária do ano de 2010.

Fonte. Elaborado pelos autores, 2019.

DEFICIÊNCIA - 2010



QUANTIDADE

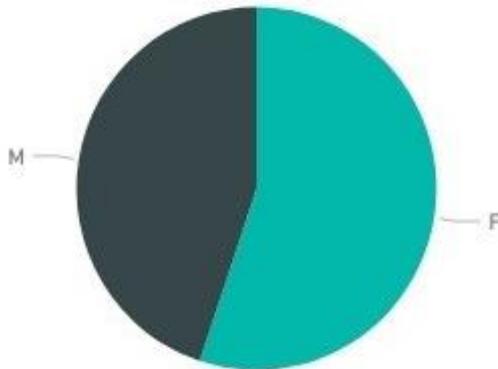
S - 23

N - 4977

Figura 17. Representação do gráfico de deficiência do ano de 2010.

Fonte. Elaborado pelos autores, 2019.

GÊNERO - 2010



QUANTIDADE

F - 2756

M - 2244

Figura 18. Representação do gráfico de gêneros do ano de 2010.

Fonte. Elaborado pelos autores, 2019.



Figura 19. Representação do gráfico de local do ano de 2010.

Fonte. Elaborado pelos autores, 2019.

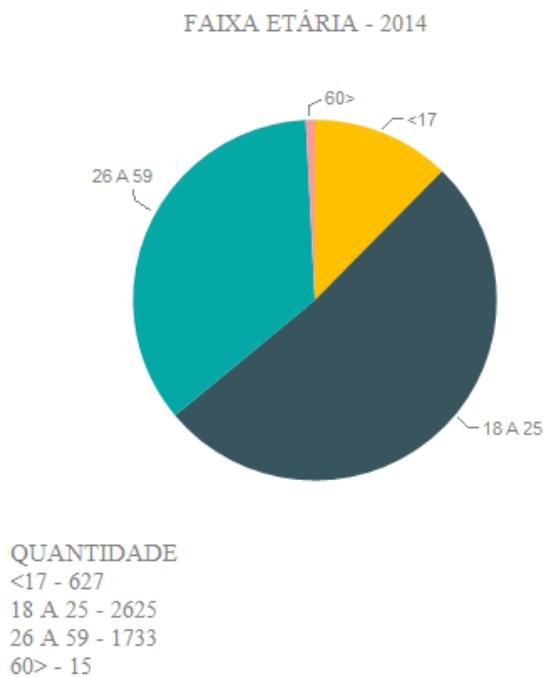


Figura 20. Representação do gráfico de idade separada por faixa etária do ano de 2014.

Fonte. Elaborado pelos autores, 2019.

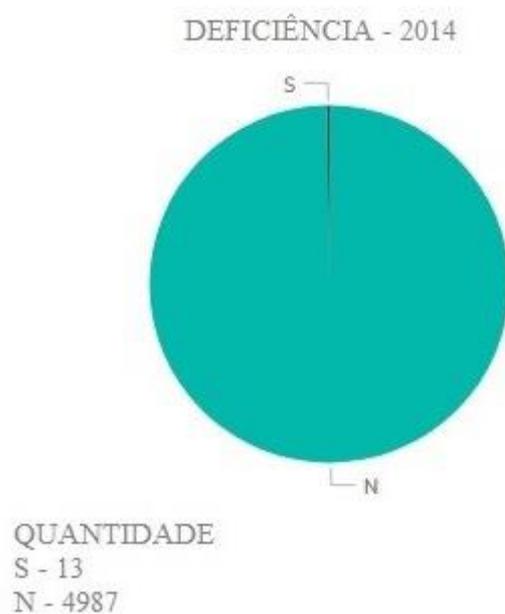


Figura 21. Representação do gráfico de deficiência do ano de 2014.

Fonte. Elaborado pelos autores, 2019.

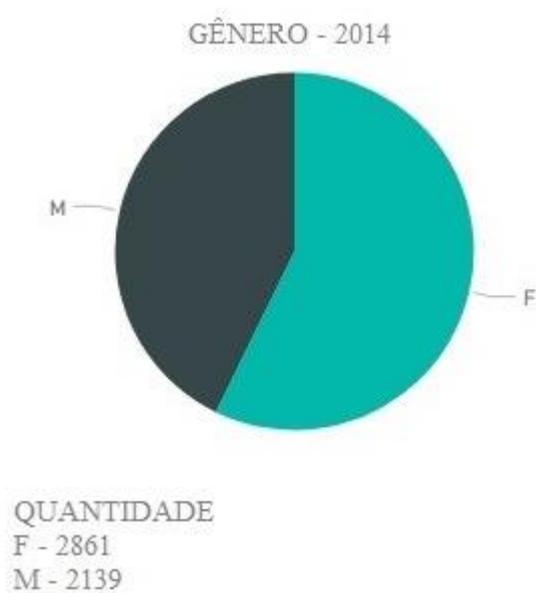


Figura 22. Representação do gráfico de gêneros do ano de 2014.

Fonte. Elaborado pelos autores, 2019.

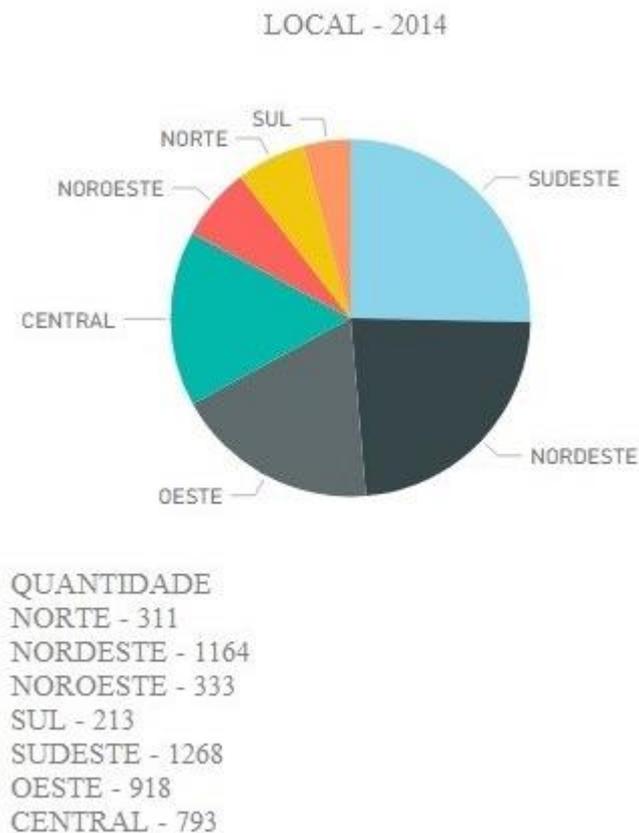


Figura 23. Representação do gráfico de local do ano de 2014.

Fonte. Elaborado pelos autores, 2019.

8. Conclusão

A pesquisa visou mostrar uma técnica que pode ser usada para a análise de dados sobre um determinado assunto, foram feitas exploração de dados onde utilizamos a investigação dos processos fazendo que fossem gerados os gráficos e relatórios sobre o assunto apresentado no trabalho. O processo que foi adquirido para chegar até os gráficos e os relatórios foram a utilização de alguns métodos como análise BI para armazenar os dados, foi implementado Data Warehouse para verificar com dados ser analisados e qual seria o processo de tomada de decisão e houve a análise de processos para conduzir uma revisão das informações.

Após fazer todos os processos descritos acima foi possível começar a análise dos gráficos, concluímos que o crescimento de alguns tópicos é relativo à quantidade de inscritos e a ordem em que os candidatos foram realizando as inscrições, ressaltando que foram analisados os cinco mil primeiros candidatos que compareceram a prova. Em comparação dos anos de 2010 e 2014 podemos concluir que houve uma diminuição de candidatos na faixa etária de 18 a 25 anos e um aumento de candidatos menores de 17 anos e maiores de 60 com isso, relatou-se que alunos do ensino médio estão valorizando os estudos e desejando iniciar um ensino superior, em relação aos candidatos idosos é possível concluir que os mesmos estão querendo voltar a estudar e adquirir conhecimento técnico e prático de alguns assuntos.

Um ponto importante que foi avaliado é em relação aos portadores de deficiência, como a tecnologia está cada vez mais avançadas, nossas expectativas em relação a estes candidatos eram outras, com base na acessibilidade que vemos hoje em dia, acreditávamos que eles seriam os candidatos que mais trariam aumento, mas o resultado foi ao contrário, ocorreu uma queda. Essa informação em relação aos deficientes abre uma questão, porque há poucos e por que diminui? Vale também ressaltar que é possível sim ter um aumento destes candidatos, porém, os mesmos não foram os primeiros a realizar as inscrições.

Em relação ao gênero, o grupo feminino cresceu gradativamente enquanto o masculino fez o mesmo de forma inversamente, lembrando que quando um gênero aumenta outro diminui na mesma proporção. O grupo feminino mesmo sendo a maior população do país, vem se descartando cada vez mais nas universidades e hoje em dia as mulheres não dependem mais de seus maridos e tentam cada vez serem independentes. Entre as localidades das execuções das provas são notórios que nos anos de 2010 e 2014 as regiões Nordeste e Sudeste foram as que mais obtiveram candidatos presentes e as regiões Sul e Norte as que mais implicaram com a diminuição de candidatos inscritos no decorrer dos anos. É possível também afirmar que na questão de local é conclusivo que estas regiões podem não ter tanta facilidade ao acesso as informações sobre o ENEM e por conta disto não tiveram tantos inscritos.

Referências

- Angeloni, M. T. e Reis, E. S. (2006). “Business Intelligence como Tecnologia de Suporte a Definição de Estratégias para a Melhoria da Qualidade do Ensino”. Salvador, Bahia, Brasil, Agosto.
- Bittencourt, M. R. (2015). “Sistemas ERP X Ferramentas de B.I”. <http://brasaosistemas.com.br/blog/sistemas-erp-x-ferramentas-de-b-i>, Agosto.
- Elias, D (2014). “Dimensões e Fatos no contexto do Business Intelligence”. <https://canaltech.com.br/business-intelligence/dimensoes-e-fatos-no-contexto-do-business-intelligence-bi-18710>, Setembro.
- HEFLO, (2016). “Análise de Processos”. <https://www.heflo.com/pt-br/definicoes/analise-de-processos>, Outubro.
- Lucidchart, (2018). “O que é um modelo de banco de dados?”. <https://www.lucidchart.com/pages/pt/o-que-%C3%A9-um-modelo-de-banco-de-dados>, Novembro.
- Meira, R. (s/d). ”Banco de Dados.” <http://www.regilan.com.br/wp-content/uploads/2013/10/Apostila-Banco-de-Dados.pdf>, Novembro.
- Microsoft, (2016). "Visão geral de cubos OLAP do Service Manager para análise avançada". <https://docs.microsoft.com/pt-br/system-center/scsm/olap-cubes-overview?view=sc-sm-1807>, Setembro.
- Moraes, D. (2018). “Business Intelligence: o que é e como fazer análise de dados de inteligência empresarial?”. <https://inteligencia.rockcontent.com/business-intelligence>, Agosto.
- Oliveira, Junia Oliveira (2018). “20 anos do Enem: confira a história do exame e o que mudou nesse período”. https://www.em.com.br/app/noticia/especiais/educacao/2018/11/11/internas_educacao,1004738/20-anos-do-enem-confira-a-historia-do-exame-e-o-que-mudou-nesse-perio.shtml, Junho.
- Santos. V. F. M. (2016). “O que é Análise de Dados? Como estruturar a sua?”. <https://www.fm2s.com.br/analise-de-dados-como-estruturar>, Outubro.
- Scartezini, L. M. B. (2009). “Análise e Melhoria de Processos”. Goiânia, Goiás, Brasil, Agosto.
- Siqueira, F. (2014). “Data Warehouse”. http://www.inf.ufsc.br/~frank.siqueira/bd_univag/DataWarehouse-Peq.pdf, Setembro.
- Think Consulting (2018). “3 empresas que utilizam business intelligence e obtiveram sucesso”. <https://blog.thinkconsulting.com.br/3-empresas-que-utilizam-business-intelligence-e-obtiveram-sucesso/>, Junho.