

UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO

VICTOR LUIZ DE OLIVEIRA E SILVA

APLICAÇÃO DE UMA FERRAMENTA DE B.I. PARA ANALISAR O INVESTIMENTO DAS UNIVERSIDADES FEDERAIS

GOIÂNIA

2019

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR
VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE
GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC nº 1204/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG):

Nome completo do autor: *Victor Luiz de Oliveira e Silva*

Título do trabalho: *Apliação de um ferramenta de B.I para análise e investimento das universidades federais*

2. Informações de acesso ao documento:

Concorda com a liberação total do documento ☒ SIM ☐ NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF do TCCG.

Victor Luiz de Oliveira e Silva
(Nome completo do autor)²

Ciente e de acordo:

Thyago Carvalho Marques
(Nome completo do orientador)²

Data: 05 / 03 / 2019

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Versão abril de 2018

² As assinaturas devem ser originais sendo assinadas no próprio documento, imagens coladas não serão aceitas.

VICTOR LUIZ DE OLIVEIRA E SILVA

APLICAÇÃO DE UMA FERRAMENTA DE B.I. PARA ANALISAR O INVESTIMENTO DAS UNIVERSIDADES FEDERAIS

Trabalho de Conclusão apresentado à Coordenação do Curso de Engenharia de computação da Escola de Engenharia Elétrica, Mecânica e de computação da Universidade Federal de Goiás, como requisito para obtenção do título de Bacharel em Engenharia de computação.

Orientador: Prof. Dr. Thyago Carvalho Marques

Goiânia

2019

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Luiz de Oliveira e Silva, Victor
APLICAÇÃO DE UMA FERRAMENTA DE B.I. PARA ANALISAR O
INVESTIMENTO DAS UNIVERSIDADES FEDERAIS [manuscrito] /
Victor Luiz de Oliveira e Silva. - 2019.
xiv, 44 f.: il.

Orientador: Prof. Thyago Carvalho Marques .
Trabalho de Conclusão de Curso (Graduação) - Universidade
Federal de Goiás, Escola de Engenharia Elétrica, Mecânica e de
Computação (EMC), , Cidade de Goiás, 2019.

Bibliografia.

Inclui siglas, abreviaturas, gráfico, tabelas, algoritmos, lista de
figuras.

1. Powerbi. 2. Educação superior. 3. business intelligence. I. ,
Thyago Carvalho Marques, orient. II. Título.

CDU 62:004.3/.4



ATA DE AVALIAÇÃO DE PROJETO FINAL

CURSO

() Eng. Elétrica () Eng. Mecânica (☒) Eng. de Computação
() Projeto Final 1 (☒) Projeto Final II

AVALIAÇÃO DE PROJETO FINAL

Título do projeto: *Aplicação de uma Ferramenta de BI para Analisar o Investimento das Universidades Federais*

BANCA AVALIADORA

Membro 1: *Thyago Carvalho Marques*

Membro 2: *Isabela Brandão Correia*

Membro 3: *Sandraley Ramos Pires*

ESTUDANTES	
Matrícula	Nome
<i>201301344</i>	<i>Victor Luiz de Oliveira e Silva</i>

NOTAS													
Matrícula	Membro 1				Membro 2				Membro 3				Média
	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	NPT	NTE	NAA	NF	
<i>201301344</i>			<i>8,0</i>	<i>8,0</i>			<i>8,0</i>	<i>8,0</i>			<i>7,0</i>	<i>8,0</i>	<i>8,0</i>

NPT – Nota plano de trabalho; NTE – Nota do trabalho escrito; NAA – Nota de apresentação e arguição

Para Eng. Elétrica, Mecânica e PFC2 da Eng. Da Computação: $NF = 0,1 \times NPT + 0,45 \times NTE + 0,45 \times NAA$

Para PFC1 da Eng. Da Computação: $NF = 0,3 \times NPT + 0,7 \times NAA$

Goiânia, 16 de julho de 2019.

Thyago Carvalho Marques

Membro 1

Isabela Brandão Correia

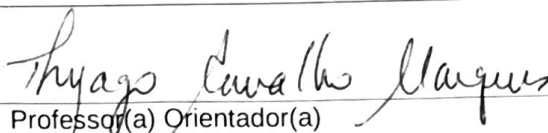
Membro 2

Sandraley Ramos Pires

Membro 3



FREQUÊNCIA - a ser preenchido pelo orientador(a)	
Nome do(a) estudante	Frequência (%)
Victoriz Luiz de Oliveira e Silva	75%


Professor(a) Orientador(a)

ATA DE AVALIAÇÃO DE PROJETO FINAL - Observações

Preencher com modificações solicitadas, caso existam. Em caso de reprovação, informar a justificativa.

- Solicitamos uma alteração no texto conforme as análises da apresentação.
- > Analisar e checar todos os ETL de dados e porar o BD.
- > Elaborar um artigo para CONPEX

AGRADECIMENTOS

Agradeço, antes de tudo à minha família, que sempre me apoia e sempre esteve presente em minha vida. À minha namorada, Amanda, que está sempre ao meu lado nos últimos anos. Aos amigos que sempre tem aquela palavra amiga para nos dar forças nas horas difíceis. Aos professores que tentam nos passar além da disciplina um pouco mais de suas experiências de vida. Ao meu orientador Prof. Dr. Thyago, pela sabedoria com que me guiou nessa trajetória e pela coragem por propor um tema de tamanha importância mas pouco explorado. Enfim, a todos que de uma forma ou de outra contribuíram para a realização deste projeto.

RESUMO

O presente trabalho tem como objetivo estudar os investimentos em educação superior no Brasil. Uma vez que o contexto atual anseia por informações a respeito de decisões passadas e de planejamentos futuros. Nada melhor para obter isso que o Business Intelligence (BI). Este tem um papel importantíssimo para mostrar se o que está sendo feito está mostrando resultado e compará-los para auxiliar no planejamento. Logo temos inteligência organizacional que eleva o nível de eficiência da administração pública e privada. Este trabalho irá apresentar que o ensino é de fato pouco eficiente no Brasil se comparado a países europeus e norte americanos. O projeto foi desenvolvido devido a disponibilidade dos dados em portais online. Foram utilizadas duas ferramentas: o pentaho data integration e o PowerBI para melhor atender as demandas do projeto.

Palavras-chave :Business Intelligence. PowerBI. Educação Superior.

ABSTRACT

This present work has the objective to study government investments in higher education in Brazil. Since the actual context longs to information about past decisions and planning the future. Nothing better than Business Intelligence to get it. This one has a very important job to show if what is being done is having results and compare it to help on planning. We have organization intelligence which brings it to the next level of efficiency of the public and private administrations. This present work was developed due to the availability of data in online portals. Were used duas tools: the pentaho data integration and Microsoft PowerBI to better answer the demands of the Project.

Keywords: Business Intelligence. PowerBI. Tertiary education.

LISTA DE ABREVIATURA E SIGLAS

UFG	Universidade Federal de Goiás
BI	Business Intelligence (Inteligência de negócio)
PDI	Pentaho Data Integration
CSV	Comma Separated Value (Valores separados por vírgula)
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
DW	Data Warehouse (Armazém de dados)
PIB	Produto Interno Bruto
GDP	Gross Domestic Product (Produto interno bruto)
PPP	Purchasing Power Parity (Paridade do poder de compra)
UFRJ	Universidade Federal do Rio de Janeiro
UFESBA	Universidade Federal do sul da Bahia
UFU	Universidade Federal de Uberlândia
IES	Instituição de Ensino Superior

LISTA DE FIGURAS

Figura 1 - Corte de Verba	12
Figura 2 - Quadrante Mágico Gartner	15
Figura 3 - Esquema de Modelo Dimensional Estrela	18
Figura 4 - Etapas de ETL	19
Figura 5 - Transformação no PDI	20
Figura 6 - Job no PDI	21
Figura 7 - Esquema do DW	23
Figura 8 - ETL da tabela fato de docentes	24
Figura 9 - Níveis de Educação OCDE	25
Figura 10 - Alunos de Pós Graduação	26
Figura 11 - Ensino Superior no Brasil	28
Figura 12 - Custo por Universidade Federal	28
Figura 13 - Orçamento por universidade federal	29
Figura 14 - Alunos por Curso	29
Figura 15 - Custo por artigo	30
Figura 16 - Alunos por docente	30
Figura 17 - Correlação entre investimento e artigos	31
Figura 18 - PIB e Artigos por País	31
Figura 19 - Modelagem dos dados no PowerBI	32
Figura 20 – Exemplo de dados da d_Curso	32
Figura 21 - Exemplo de dados da d_tempo	33
Figura 22 - Exemplo de dados da d_universidade	33
Figura 23- Exemplo de dados da d_pais	34
Figura 24-Exemplo de dados da f_%PIB	34
Figura 25-Exemplo de dados da f_artigo	34
Figura 26- Exemplo de dados da f_docentes_mundo	35
Figura 27- Exemplo de dados da f_Matricula	35
Figura 28- Exemplo de dados da f_orcamento	35
Figura 29- Exemplo de dados da f_Artigo_mundo	36
Figura 30- Exemplo de dados de f_docente	36
Figura 31- Exemplo de dados de f_PIB	36
Figura 32 - Comparativo UFG (alunos)	37
Figura 33 - Investimento por país	38

Sumário

1.Introdução	12
2. Referencial Teórico	15
2.1 Quadrante Gartner	15
2.2 Data Warehouse	16
2.3 Modelo Dimensional	17
2.4 Modelo Estrela	17
2.5 Extração, Transformação e Carregamento (ETL)	19
2.6 Ferramentas	19
2.6.1 Pentaho	19
2.6.2 Microsoft PowerBI	21
3. Materiais e Métodos	22
3.1 Criação do DW (Data Warehouse)	22
3.2 Processo de ETL usando o PDI (Pentaho Data Integration)	23
3.3 Visualização no Power BI	24
3.4 ETL no PowerBI	24
4. Resultados	28
4.1 Comparativo	37
5. Discussão	39
6.Conclusão	42
7. Referências	43

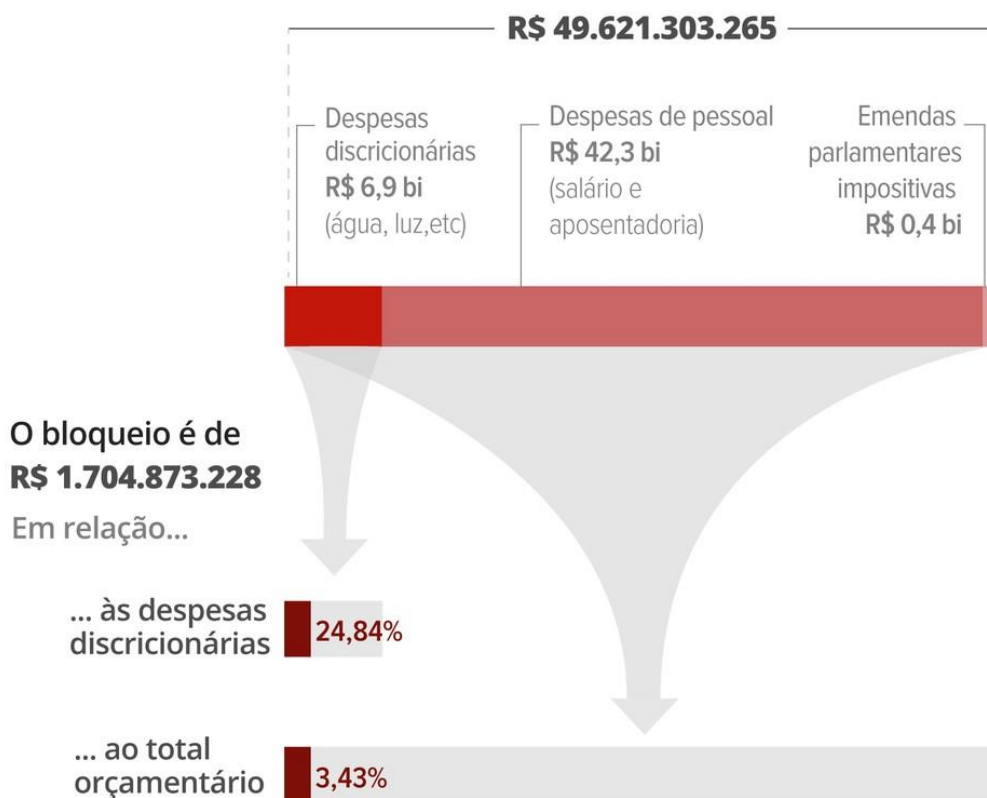
1.Introdução

O cenário atual Brasileiro é de retomada de crescimento da economia, os valores do PIB vêm em promessa de alta. Porém a arrecadação de impostos ainda é baixa aliada a recente troca de governo. O ambiente é propício para questionamentos a respeito de investimentos em áreas como a educação. Principalmente no que diz respeito a eficiência do dinheiro investido.

Não se pode esquecer também do contingenciamento de aproximadamente 3% da verba das universidades federais e que tem causado bastante críticas à gestão.

Orçamento total e o bloqueio

Total orçamentário das universidades em 2019:



Fonte: MEC



Infográfico elaborado em: 14/05/2019

Figura 1 - Corte de Verba

Ao se falar em eficiência, pode-se comparar as diferentes universidades brasileiras em busca de qual a melhor, qual produz mais formandos com menos verba, ou ainda qual a melhor qualificada no ranking de universidades mundiais. Porém existem inúmeras métricas capazes de comparar universidades, este artigo falará apenas no que diz respeito ao valor do investimento, alunos e docentes. Estes fatores foram escolhidos para serem analisados pois são de fácil medição e acompanhamento de resultados ao longo do tempo. Uma vez que medir fatores como a importância do lugar geográfico, investimento de empresas externas e ainda produção de conhecimento, embora sejam fatores importantes para obter a eficiência, são difíceis de medir e comparar.

Outra métrica capaz de analisar a eficiência é o percentual do PIB (Produto Interno Bruto) que é destinado a investimento em educação. Esta medida é importante analisar pois trata os países em tom de igualdade, uma vez que tenham PIB's diferentes, compara-se a parcela deste PIB, ou entende-se o grau de relevância que a educação tem para a economia do país.

Através deste artigo tem-se a intenção de obter informações acerca do ensino superior no Brasil. Saber sobre seus custos, quantidade de alunos, docentes e sua real eficiência de ensino. Com os dados em mãos é possível fazer uma análise a respeito da educação superior pública brasileira e a partir daí compará-la com o ensino em outros países do mundo. Desta forma podemos ter maior conhecimento sobre a eficiência ou não do dinheiro que está sendo gasto nas faculdades públicas.

Temos observado um ataque frontal aos investimentos em educação no Brasil. De um lado o governo que justifica cortes de gastos desnecessários com custo altíssimo, e de outro lado associações de defesa à universidade pública que são contra esses cortes. Nesse contexto o trabalho é levantar e observar essas informações e levantar até que ponto um ou outro está certo e comparar com o contexto de países desenvolvidos para ver o quanto o Brasil se distancia ou não dessa situação.

O objetivo deste trabalho é comparar a educação brasileira com a educação em outros países. Capacidade de comparação, quanto se custa de fato, se está longe ou não. E mostrar os fatos de forma simples e fácil de entender.

O termo Business Intelligence foi primeiramente usado em 1958, quando foi lançado o artigo "Um sistema de Business Intelligence" escrito por H.P. Luhn. É descrito como "um sistema automático desenvolvido para disseminar informação para as várias seções de uma indústria, organização científica ou governamental." [Luhn 1958] (Traduzido do inglês).

Segundo Luhn, "o objetivo do BI é dar suporte com informações adequadas atividades específicas, tais como indivíduos, grupos, departamentos ou até unidades maiores."

Nos dias atuais existem várias ferramentas de BI no mercado utilizadas em todo o mundo por empresas para a tomada de decisão. Podemos mencionar ferramentas como Tableau, Qlikview, Pentaho e Power BI (Microsoft). Todas as ferramentas possuem como objetivo levar informação para a empresa da forma mais eficiente possível.

O conceito de business intelligence vai além de uma simples ferramenta. BI é um processo, um conjunto de técnicas e conceitos. Trata de entregar informação certa para a pessoa certa no tempo certo. Esse processo passa por várias etapas, entre elas temos: a coleta, a organização e análise dos dados, a elaboração de relatórios ou dashboards e todo o acompanhamento e atualização além das adições posteriores e incrementos do projeto. Esses são os fundamentos, a base para entender como tudo funciona. [Rafael Piton]

Uma das formas mais utilizadas para consumir o BI é através do uso de dashboards. “O dashboard é uma ferramenta de visualização de dados que exhibe o status atual das métricas e indicadores, chave de desempenho (KPIs) entre outros dados de uma empresa”. [Rafael Piton Dashboard]

Os dashboards geralmente apresentam as informações que são processadas anteriormente de forma que o usuário possa remover filtros, aplicar novos filtros e ver os dados de diversas formas, e com diferentes perspectivas com poucos cliques. É através de dashboards que iremos apresentar os resultados deste artigo.

A principal ferramenta a ser utilizada neste artigo é o Microsoft PowerBI, atualmente é a líder de mercado, uma vez que permite fazer todo o processo de BI de forma mais rápida que outras e ainda é de fácil aprendizado e utilização. Conta com uma versão grátis que permite que seja feito todo o trabalho de edição de dados, visualização e publicação na nuvem para visualização na web. Além da facilidade para criar dashboards, o PowerBI conta com um módulo de transformação de dados (Power Query) em que pode-se realizar edições de dados com rapidez e facilidade. O segredo para toda a rapidez é o armazenamento de dados em memória, o que deixa as coisas fluírem rapidamente, porém o programa não suporta trabalhar com um volume grande de dados e é necessário utilizar uma outra ferramenta para complementá-lo.

A segunda ferramenta utilizada é o Pentaho, responsável pela etapa de ETL. Foi utilizada por ser uma ferramenta mais robusta e que pôde tratar e resumir um grande volume que foi utilizado para este artigo. Um arquivo CSV obtido pesava mais de 1 gigabyte e contava com mais de 1 milhão de linhas e portanto precisou ser tratado no Pentaho Data Integration (PDI) antes de ser carregado no PowerBI. O Pentaho é uma suíte que conta com vários programas que são capazes de entregar um BI completo para o usuário final. Conta com a versão da comunidade (gratuita) e com a versão paga. A diferença entre elas é a usabilidade, principalmente na confecção de dashboards. Neste artigo usaremos apenas o software de ETL do Pentaho para resumir os grandes arquivos CSVs e conseguir utilizá-los no projeto.

Com o uso das técnicas de Business Intelligence, faremos a extração, transformação e apresentação das informações extraídas de dados públicos a respeito da educação. Estes dados foram retirados de várias fontes na internet, sendo a maior delas o Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), portal da transparência, OECD entre outras. O objetivo é obter insights conclusivos a respeito da educação superior no Brasil. Comparando- a com países desenvolvidos e países com mesmo potencial como por exemplo os BRIC’S.

2. Referencial Teórico

2.1 Quadrante Gartner

A empresa de consultoria Gartner possui sua própria definição de B.I. “Business intelligence é um termo abrangente de aplicações, infraestrutura e ferramentas, e melhores práticas para permitir acesso e analisar as informações e melhorar e otimizar decisões e performance”. Ela produz o “quadrante mágico” que classifica as ferramentas existentes de acordo com sua completude e habilidade de execução.

O quadrante gartner refere-se a uma série publicada pela empresa de consultoria Gartner de pesquisa de mercado. Ele revela que confiar em métodos qualitativos de análise para demonstrar tendências de mercado, como direção, maturidade e participantes daquele nicho. Suas análises são conduzidas por várias indústrias de tecnologia e são atualizadas a cada ano.

Muitas empresas utilizam o quadrante gartner quando pretendem adquirir ou conhecer um pouco mais sobre as opções de software para determinada tarefa e entender qual se encaixa mais com as especificidades da empresa.

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Figura 2 - Quadrante Mágico Gartner

Interpreta-se o quadrante gartner como o primeiro passo para entender os provedores de tecnologia para se considerar para um uso específico de investimento.

Considere que o quadrante de líderes não é sempre a melhor opção. Há bons motivos para se considerar os desafiantes de mercado. E um software de nicho pode ser a melhor opção para o seu negócio. Tudo depende de como se alinha o provedor com seus objetivos de negócio.

- Os líderes de mercado (leaders) como é o caso do PowerBI (Microsoft), Tableau e Qlik se saem bem na visão atual e estão bem posicionados para os desafios de amanhã.
- Os visionários (visionaries) são os que entendem para onde o mercado está se dirigindo ou tem uma visão das mudanças de tendência, mas ainda não executam bem.
- Os de Nicho (niche players) são as empresas focadas no sucesso de um segmento específico, ou não estão focadas em algo. Não inovam e não geralmente não são as melhores opções.
- Os desafiantes (Challengers) executam bem hoje e podem dominar uma grande segmento de mercado, mas não demonstram um entendimento das direções do mercado. [Quadrante Gartner] (Traduzido do inglês)

O Quadrante Gartner ajuda a entender rapidamente sobre as competências tecnológicas e a habilidade de entregar para o usuário final hoje o que eles precisarão no futuro. Entender como o mercado se posiciona e as estratégias que estão sendo usadas para competir por negócios.

Comparar as tecnologias, forças e desafios com suas necessidades especiais. Pode-se perceber que a suíte Pentaho não aparece no Quadrante Gartner deste ano. Isso se deve a ser uma plataforma que, na sua versão gratuita (Community Edition) necessita de um maior conhecimento técnico para entregar as várias funcionalidades que as outras entregam.

Ele necessita de atualizações em sua estrutura para que seja comparado com as consideradas líderes de mercado.

2.2 Data Warehouse

Um dos aspectos mais importantes das organizações são os sistemas de informação. Este tem o propósito de guardar dados e fazer a engrenagem da empresa rodar. Por outro lado, existe o sistema de DW/BI(data warehouse/business intelligence) que é responsável pela saída dos dados e verificar o funcionamento desta engrenagem.

Em computação, um armazém de dados (data warehouse) é um sistema usado para relatórios e análises de dados, e é considerado um componente importante para business intelligence.

D.Ws são o repositório central de uma ou mais fontes diferentes. Guardam informações atuais e históricas em um único lugar que é usado para criar análises e relatórios para os usuários.

[Data Warehouse]

Os dados armazenados no DW são retirados de fontes de dados diversas, como bancos relacionais, planilhas, dados da internet, passando por um processo chamado E.T.L (extração, transformação e carregamento) e após estarem limpos vão para o D.W. Isso garante qualidade dos dados antes de serem analisados.

Alguns dos aspectos que um data warehouse deve apresentar:

- “Permitir acesso a informação de forma fácil”.
- “Informações consistentes”.
- “Ser adaptado a mudanças”.
- “Apresentar informações no momento oportuno”.
- “Ser seguro e proteger as informações.”
- “A comunidade do negócio deve aceitar o sistema de BI.”

(Traduzido do inglês)

Através da adoção de um sistema de DW/BI podemos acabar com alguns problemas frequentes em empresas:

- Obtemos a versão única da verdade, onde as reuniões passam a ser mais produtivas e orientadas a tomada de decisão.
- Diminui-se problemas para grande volume de dados.
- Eliminação de erros e maquiagens nos dados.
- Automação de processos.
- Melhor aproveitamento dos analistas de informação.

2.3 Modelo Dimensional

O modelo dimensional é um modelo de banco de dados que tem como objetivo melhorar performance para consultas. É muito utilizado para sistemas de business intelligence voltado para a criação de relatórios e dashboards.

Modelos dimensionais são bem diferentes de modelos da terceira forma normal (3FN) que visam eliminar redundâncias. No entanto, modelos normalizados são muito complicados para consultas de BI. Os usuários não conseguem entender ou navegar por um sistema tão complexo. [The Data Warehouse Toolkit] (Traduzido do inglês)

Vale lembrar que o modelo dimensional contém a mesma informação que o modelo normalizado, porem guarda essas informações de forma a entregar entendimento, performance e facilidade de mudança. [The Data Warehouse Toolkit] (Traduzido do inglês)

2.4 Modelo Estrela

O modelo dimensional mais usado é o modelo estrela, devido a sua simplicidade. Este contém uma ou mais tabelas fato, e várias dimensões ligadas as fatos. A tabela fato representa a(s) medida(s) (que são as colunas da tabela) que se quer analisar (responde a pergunta “O quê?”). As tabelas de dimensão correspondem a forma como se quer ver a métrica (responde a pergunta “Como?”). Ao projetar um DW, deve-se perguntar ao cliente: ‘qual informação você quer receber?’ A resposta pode ser: ‘Quero saber as vendas, o lucro e despesa por loja, tempo e

produto'. Assim as medidas da tabela são vendas, lucro e despesa. E as dimensões: loja, tempo e produto.

A tabela fato é composta por várias colunas, são elas: uma chave primária, várias chaves estrangeiras (uma para cada dimensão), e as métricas que se tem interesse.

As dimensões são compostas por uma chave primária e os atributos que definem aquela dimensão. Por exemplo na dimensão data, temos atributos de dia, mês, nome do mês, dia da semana, ano, trimestre, nome do trimestre além da chave primária.

Modelo Dimensional (Esquema Estrela)

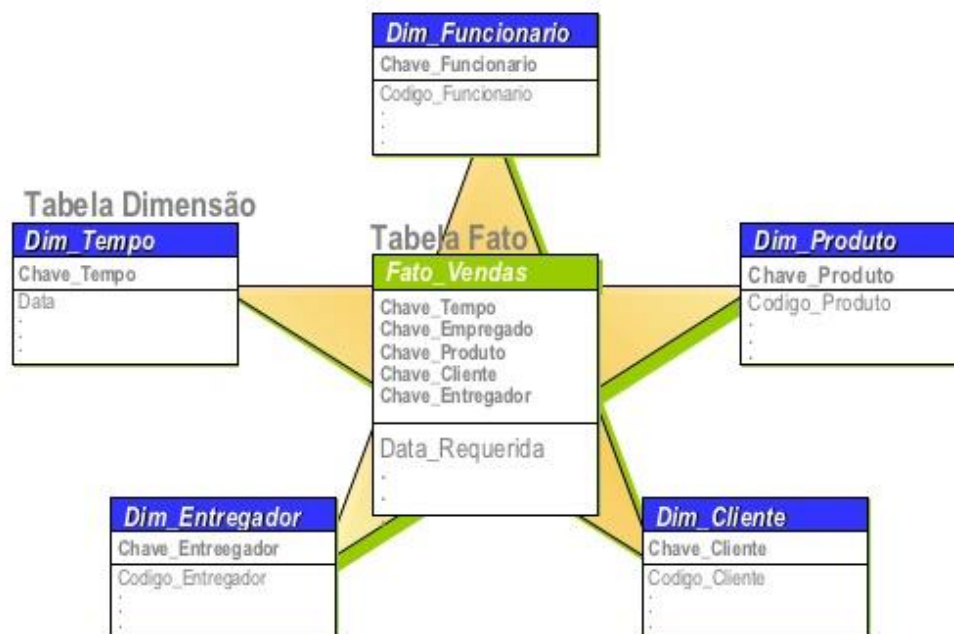


Figura 3 - Esquema de Modelo Dimensional Estrela

Esse modelo, no entanto, possui uma desvantagem. Caso uma dimensão possua mais de um atributo e o atributo seja um subatributo, teremos uma coluna que aparece repetidas vezes.

Deve-se mencionar o conceito de granularidade. Granularidade é o menor nível de detalhamento de uma tabela 'fato'. Seria a informação contida em uma linha da tabela. A definição é definida pelo usuário, porém o analista deve ficar atento para uma granularidade baixa em uma possível expansão do projeto pode levar a ter que fazer outra fato com granularidade maior. Portanto, é

preferível escolher granularidades altas. Por exemplo: em uma tabela fato de vendas, a maior granularidade seria um único item daquela venda.

2.5 Extração, Transformação e Carregamento (ETL)

A sigla E.T.L (extração, transformação e carga) se refere ao processo de extração, limpeza, transformação de dados e carga de um sistema fonte para serem um DW.

A extração consiste na leitura dos dados, visualizações entre outras formas de extrair. A transformação está muito ligada as regras de negócio, transformar os dados para que tenham utilidade na visualização, de forma que se ganhem a forma do Data Warehouse. A carga é o armazenamento dos dados no D.W.

“A extração, transformação e carregamento consomem a maior parte do tempo requerido para construir um ambiente de BI” [Kimball e Ross 2013].

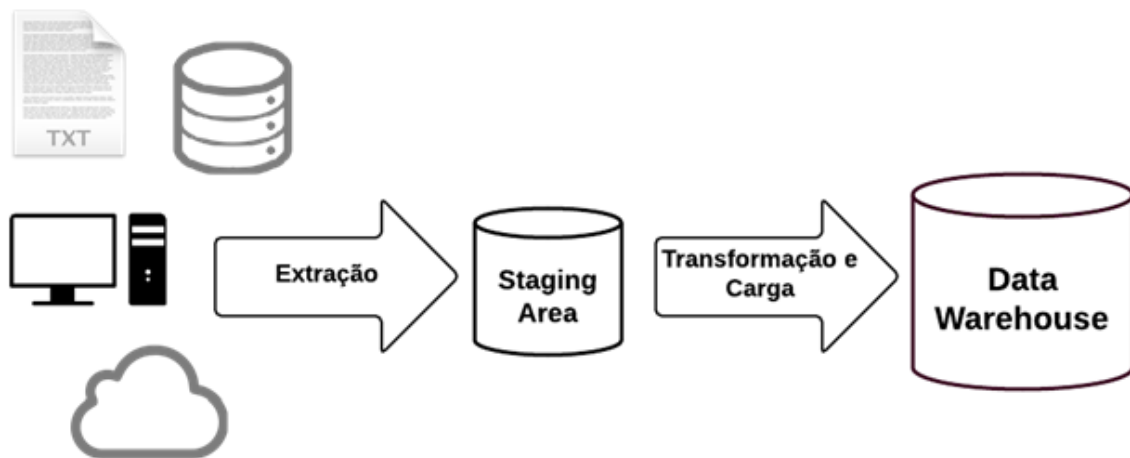


Figura 4 - Etapas de ETL

Na figura podemos ver o processo de E.T.L Os dados são extraídos para uma área de ‘staging’, lá são limpos, transformados, combinados e/ou ainda normalizados. E então finalmente podem ser carregados em um DW(Data Warehouse) . O encarregado a criar todas essas rotinas são os analistas, sendo assim, usuários não tem acesso a este nível do BI.

2.6 Ferramentas

2.6.1 Pentaho

O Pentaho é uma suíte completa de Business Analytics. Este possui ferramentas de ETL, criação de cubos OLAP, servidor com possibilidade de gerenciamento de usuários, visualização dos cubos

OLAP, criação de dashboards além da criação de relatórios personalizados. Porém, isso tudo tem um custo, que está relacionado com a visualização dos resultados. A criação de dashboards interativos com filtros é um processo que toma muito tempo, além de precisar ter conhecimento em linguagem MDX. É um processo custoso.

Possui duas versões: a Enterprise Edition que é a versão comercial, e a Community Edition que é a versão gratuita. A diferença entre as duas é a usabilidade. A versão Community possui maiores dificuldades e exige maior conhecimento técnico, principalmente para se fazer dashboards.

A vantagem de escolher trabalhar com o Pentaho é trabalhar com o Pentaho Data Integration que será apresentado a seguir.

2.6.1.1 Data Integration

O Pentaho Data Integration(PDI) é uma excelente ferramenta para se fazer ETL. Nela utiliza-se os componentes para fazer alterações nos metadados. Cada componente tem sua função, e através da montagem dos componentes e conexão entre eles(hops) montamos a transformação.

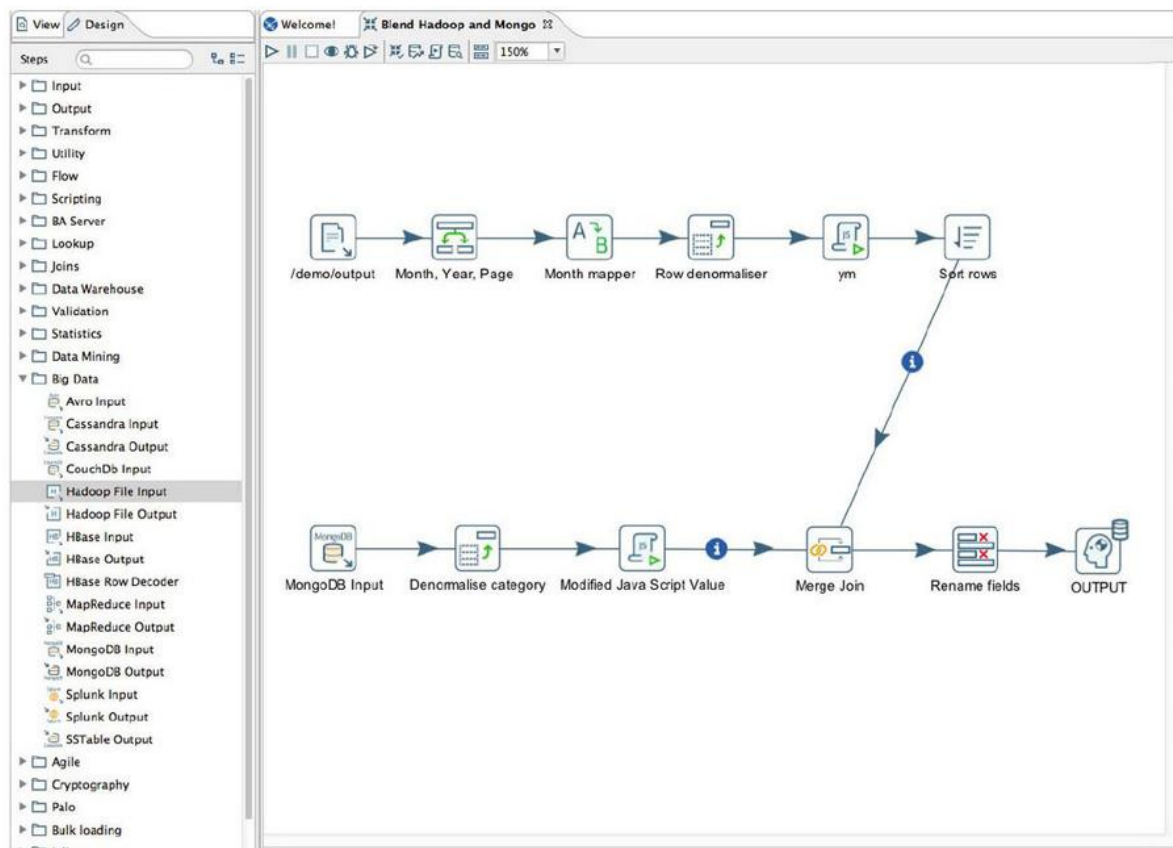


Figura 5 - Transformação no PDI

Ele utiliza o conteúdo de metadados, de forma que cada componente faça alguma modificação nos metadados para o componente seguinte. As modificações vão desde um 'de ... para', uma

filtragem de valores, um 'join' ou ainda 'group by' entre outros. Vale a pena lembrar que o ETL é sempre feito voltado para as regras de negócio que foram definidas inicialmente na criação do DW.

As transformações do PDI são o processo de carga que se realiza para uma determinada tabela do DW, seja ela fato ou dimensão.

Pode-se criar o 'Job', que é o mecanismo que o PDI utiliza para executar as transformações ou ainda outros Jobs. Ele pode ser agendado, mandar email em caso de falha na transformação, gravar o log entre outras funcionalidades.

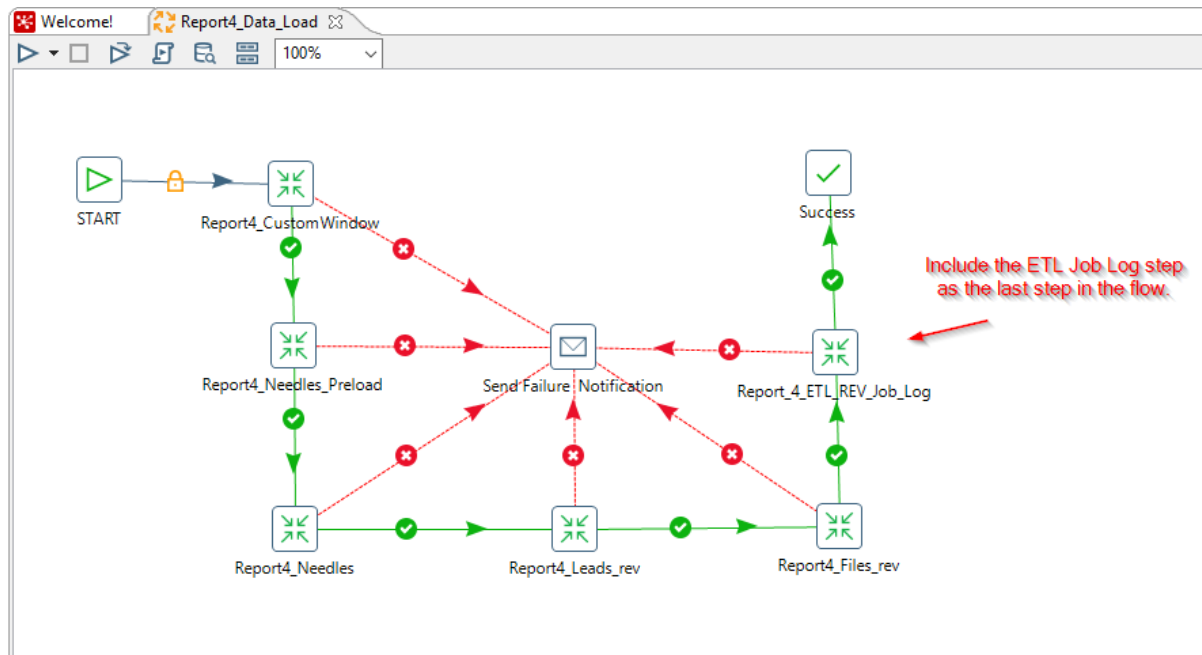


Figura 6 - Job no PDI

2.6.2 Microsoft PowerBI

O PowerBI, uma ferramenta que permite criação de consultas, conexões a diversos tipos de dados e criação de relatórios que podem ser facilmente compartilhados com outras pessoas. Com o Power BI Desktop (versão gratuita), é possível conectar dados, transformar dados e criar modelos de relatórios que permitem a visualização e o compartilhamento com outras pessoas.

Uma grande vantagem em usar essa ferramenta está na dinamicidade e praticidade que temos para gerar relatórios. Uma vez criado o dashboard não é necessário recriá-lo. A cada vez que novos dados forem inseridos este atualizará automaticamente. Para isso basta clicar no botão atualizar ou configurar o gateway on premises.

Vale lembrar que os dados só serão atualizados se estiverem na mesma estrutura de quando foram inseridos, caso haja alguma mudança poderá ocorrer erros no momento da atualização. Ao selecionar a fonte de dados e clicar em Editar, abrimos o Power Query Editor, onde podemos tratar os dados.

No Power Query é onde se realiza o ETL. Ele utiliza a linguagem M para realizar modificações nos dados, porém muitas das modificações desejadas podem ser feitas apenas com “cliques”. Por trás de cada “clique” existe um código em linguagem M, que pode ser visualizado tanto na barra de fórmula, quanto no Editor Avançado.

A linguagem M é uma linguagem com uma sintaxe diferente das usuais e portanto não usaremos neste projeto, haja vista que todo o necessário para conclusão do projeto pode ser feito com a interface gráfica do Power Query.

3. Materiais e Métodos

Os dados encontrados na internet vem em várias formas diferentes (planilhas, arquivos CSVs, arquivos zipados) cada um vem em um formato único. Para se obter a informação, é necessário tratar este dado, limpá-lo e transformá-lo. Esta transformação tem o nome de ETL(do inglês extração, transformação e carga). São transformados em um modelo que seja de fácil análise, único, e que comporte grandes volumes de dados sem causar maiores problemas. Aí que entra a fase de ETL do BI. A sigla significa extração, transformação e carregamento (do inglês “Load”), em que estes dados passam por um área que antecede sua fase final e são processados até entrar em um banco de dados ou onde de fato serão armazenados.

Os dados de despesas e orçamentos do governo federal foram encontrados no portal da transparência (<http://www.transparencia.gov.br/download-de-dados/despesas-execucao>). Os dados relacionados a matrículas de universidades foram encontrados no site do Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) (<http://portal.inep.gov.br/web/guest/microdados>). E os dados de artigos publicados foram encontrados no portal capes (<https://dadosabertos.capes.gov.br/dataset/projetos-dos-programas-de-pos-graduacao-stricto-sensu-no-brasil-2017>). A partir daí foi feito um DW baseado nas informações desejadas para analisar estes dados.

3.1 Criação do DW (Data Warehouse)

Obtidos os arquivos, que eram do tipo separado por ponto e vírgula (csv), foi analisado o conteúdo dos dados para poder criar o DW. Vale mencionar que os dados do portal da transparência disponíveis eram de 2014 até 2017. Portanto os outros dados foram coletados também do mesmo período para se obter resultados em todos os aspectos.

A partir daí, a criação de um armazém de dados voltado para consultas, foi baseada no modelo estrela.

Utilizando 3 tabelas fatos (identificadas por um prefixo ‘f_’) e 3 dimensões (identificadas por um prefixo ‘d_’). Foi criado um banco MySQL para receber os dados finais após o processo de ETL. O banco foi projetado da seguinte forma:

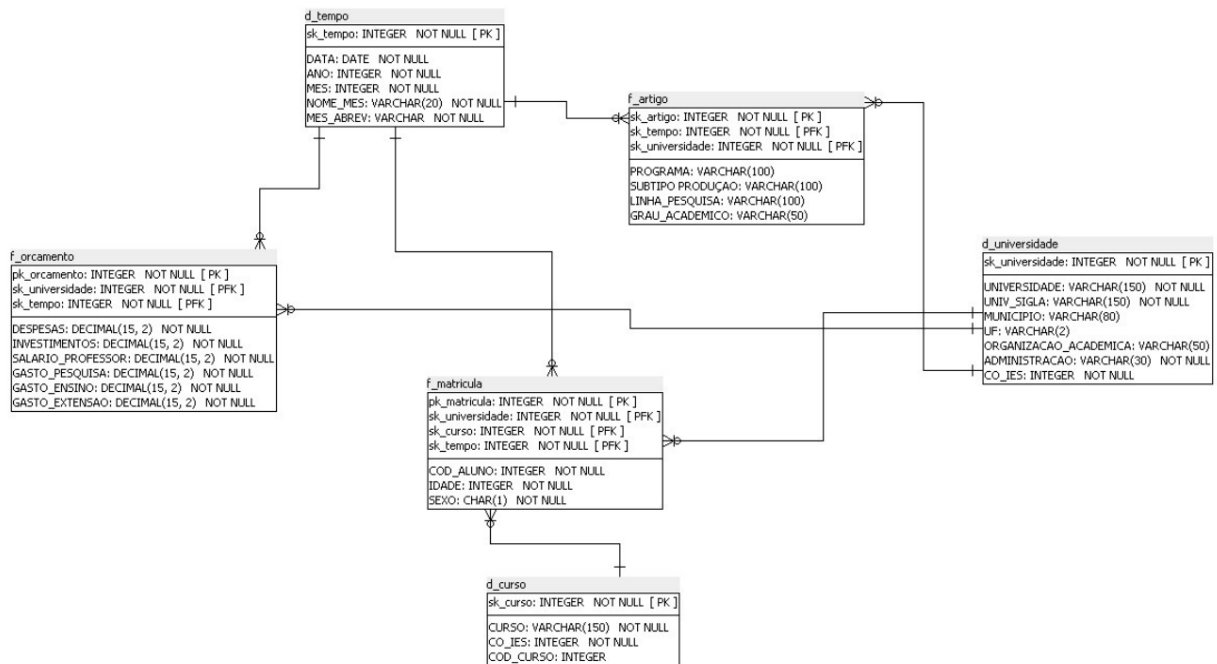


Figura 7 - Esquema do DW

3.2 Processo de ETL usando o PDI (Pentaho Data Integration)

Este artigo não tem a intenção de detalhar como foram feitas as cargas de dados nas tabelas de dimensão e fato. Porém caso seja feito algum filtro de dado ou agrupamento de linhas que cause alguma alteração, será detalhada a forma como foi feita.

Com o banco de dados criado e rodando, fomos a transformação dos dados. Inicialmente é feita a carga de dados das tabelas de dimensões:

Para a dimensão tempo foram necessários dados de mês e ano. Os dados foram gerados no próprio ETL e foram incluídos dados dos anos de 2010 a 2019;

Na dimensão universidade carregamos a partir dos micro dados do INEP todas as universidades de todos os anos analisados;

Na dimensão curso foram carregados, assim como na universidade, todos os cursos de acordo com os micro dados do INEP, não permitindo a inclusão de cursos com o mesmo código.

Com os dados de dimensões incluídos partimos agora para o carregamento das tabelas de fatos:

A granularidade da fato matrícula é por situação, curso, universidade e tempo. Foi escolhido assim devido a um grande número de registros que se fossem adicionados individualmente causaria perda de performance e pouco acrescentaria ao “dashboard”. Segue-se na fato matrícula os alunos matriculados agrupados pelo ano da matrícula, pela universidade, por curso e pela situação atual (cursando, trancado, formado, transferido, desvinculado ou falecido). Foi utilizada essa metodologia pois não tinha a exigência de granularidade de aluno. O tamanho da tabela fato, considerando que existe em média 11 milhões de alunos por ano (11 milhões de linhas) foi diminuído para uma média de 130 mil linhas por ano.

A seguir foi adicionada a fato de orçamento. Para conseguir carregar totalmente esta fato foram feitas modificações em algumas linhas da dimensão universidade. Isso se deve ao fato de a mesma universidade terem nomes diferentes no sistema do Inep e no portal da transparência (Ex. 'UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL' para 'UNIVERSIDADE FEDERAL DE MATO G. SUL'), além de algumas universidades no portal transparência apresentarem o nome "Fundação" antes do nome da universidade, o que levou a um maior tratamento a nível de ETL.

Adicionou-se, por fim, a fato de artigos, em que também continha alguns nomes de universidades que contavam com a palavra 'FUNDAÇÃO' antes da universidade e foram tratados a nível de ETL para que ao fazer o lookup em outra tabela, consiga trazer a chave estrangeira correta.

3.3 Visualização no Power BI

Inicialmente foi criada a conexão do Power BI com o Data Warehouse para obter os dados a serem carregados para os dashboards . Foi necessário a criação de um novo atributo na dimensão universidade. Este atributo é o resultado da concatenação da cidade com a unidade federativa. Esta concatenação é necessária para a pesquisa de local utilizando o elemento de mapa.

Além disso foram criadas novas métricas nas tabelas de fato:

Fato Artigos: Quantidade de Artigos; Artigos Doutorado; Artigos Mestrado; Artigos Mestrado Profissional; Custo Artigo.

Fato Matrícula: Alunos; Alunos Cursando; Alunos Evasão; Alunos Formados; Alunos Total; Taxa Evasão; Taxa Formatura.

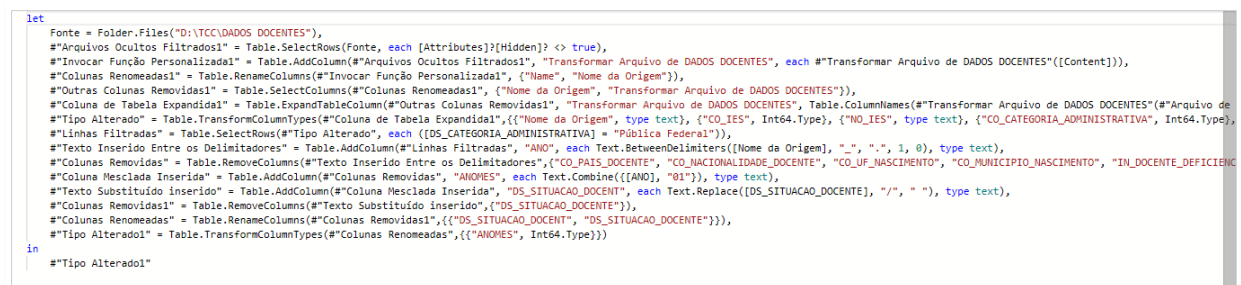
Fato Orçamento: Custo Aluno; Custo Mensal Aluno; Total Pago.

A primeira página do dashboard foi feita para trazer uma visão geral do panorama brasileiro. No elemento de mapa foi mostrado o custo do aluno por universidade e onde a mesma se encontra geograficamente no país. Isso pode nos mostrar onde está o aluno mais caro. Alguns cartões com dados foram inseridos para situar melhor, como por exemplo o Orçamento do ensino superior e a quantidade de alunos. E dados relacionados às medias, como de custo por aluno por mês de faculdade, taxa de evasão e taxa de alunos formandos por ano.

3.4 ETL no PowerBI

Foi feita a inserção dos dados de Docentes para se obter as informações de alunos por docente nas faculdades brasileiras. Como as dimensões já estavam criadas bastou inserir os dados e relaciona-los com as dimensões de tempo e universidade. Obtivemos assim os dados de docentes por tempo e por universidade.

fDocentes



```
let
    Fonte = Folder.Files("D:\TCC\DAOS DOCENTES"),
    #"Arquivos Ocultos Filtrados1" = Table.SelectRows(Fonte, each [Attributes][Hidden] <> true),
    #"Invocar Função Personalizada1" = Table.AddColumn(#"Arquivos Ocultos Filtrados1", "Transformar Arquivo de DAOS DOCENTES", each #"Transformar Arquivo de DAOS DOCENTES"([Content])),
    #"Colunas Renomeadas1" = Table.RenameColumns(#"Invocar Função Personalizada1", {"Name", "Nome da Origem"}),
    #"Outras Colunas Removidas1" = Table.SelectColumns(#"Colunas Renomeadas1", {"Nome da Origem", "Transformar Arquivo de DAOS DOCENTES"}),
    #"Coluna de Tabela Expandida1" = Table.ExpandTableColumn(#"Outras Colunas Removidas1", "Transformar Arquivo de DAOS DOCENTES", Table.ColumnNames(#"Transformar Arquivo de DAOS DOCENTES"("#Arquivo de
    #"Tipo Alterado" = Table.TransformColumnTypes(#"Coluna de Tabela Expandida1",{{("Nome da Origem", type text), ("CO_IES", Int64.Type), ("NO_IES", type text), ("CO_CATEGORIA_ADMINISTRATIVA", Int64.Type),
    #"Linhas Filtradas" = Table.SelectRows(#"Tipo Alterado", each ([DS_CATEGORIA_ADMINISTRATIVA] = "Pública Federal")),
    #"Texto Inserido Entre os Delimitadores" = Table.AddColumn(#"Linhas Filtradas", "ANO", each Text.BetweenDelimiters([Nome da Origem], "_", ".", 1, 0), type text),
    #"Colunas Removidas" = Table.RemoveColumns(#"Texto Inserido Entre os Delimitadores",{"CO_PAIS_DOCENTE", "CO_NACIONALIDADE_DOCENTE", "CO_UF_NASCIMENTO", "CO_MUNICIPIO_NASCIMENTO", "IN_DOCENTE_DEFICIENC
    #"Coluna Mesclada Inserida" = Table.AddColumn(#"Colunas Removidas", "ANOMES", each Text.Combine([ANO], "01")), type text),
    #"Texto Substituído inserido" = Table.AddColumn(#"Coluna Mesclada Inserida", "DS_SITUACAO_DOCENTE", each Text.Replace([DS_SITUACAO_DOCENTE], "/", " "), type text),
    #"Colunas Removidas1" = Table.RemoveColumns(#"Texto Substituído inserido",{"DS_SITUACAO_DOCENTE"}),
    #"Colunas Renomeadas" = Table.RenameColumns(#"Colunas Removidas1",{"DS_SITUACAO_DOCENTE", "DS_SITUACAO_DOCENTE"}),
    #"Tipo Alterado1" = Table.TransformColumnTypes(#"Colunas Renomeadas",{{("ANOMES", Int64.Type)}})
in
    #"Tipo Alterado1"
```

Figura 8 - ETL da tabela fato de docentes

Após a inserção de docentes foram feitas comparações a nível mundial dos dados encontrados do Brasil com o restante do mundo.

Foi feita uma busca no site da OCDE (Organização para a Cooperação e Desenvolvimento Econômico) para o custo por aluno em universidade federal no mundo (<https://data.oecd.org/eduresource/education-spending.htm>). É encontrado no filtro de “public spending on education”. Este dado vem em formato de % do PIB do país e nível terciário da educação. A OCDE entende o nível terciário como terciário de ciclo curto, graduação, mestrado e doutorado (ISCED 5 a 8). O ideal seria obter o custo aluno em dólares por aluno para comparar com o custo aluno do Brasil que obtivemos, porém este dado que existe no site da OCDE leva em conta universidades públicas e privadas e portanto não serve como comparativo.

A seguir buscamos dados relacionados a quantidade de professores por aluno. Estes dados possuem um filtro diferente do anterior, leva-se em conta também os professores e alunos de graduação com situação “em exercício” e “cursando” respectivamente, além de serem apenas mestrado e doutorado. (ISCED 6 a 8) A fonte destes dados está no seguinte endereço: https://stats.oecd.org/Index.aspx?DataSetCode=EAG_PERS_RATIO.

Terms used in this publication	ISCED classification
Early childhood education Refers to early childhood programmes that have an intentional education component and aim to develop cognitive, physical and socio-emotional skills necessary for participation in school and society. Programmes at this level are often differentiated by age.	ISCED 0 (sub-categories: 01 for early childhood educational development and 02 for pre-primary education)
Primary education Designed to provide a sound basic education in reading, writing and mathematics and a basic understanding of some other subjects. Entry age: between 5 and 7. Typical duration: 6 years.	ISCED 1
Lower secondary education Completes provision of basic education, usually in a more subject-oriented way with more specialist teachers. Programmes may differ by orientation, general or vocational, though this is less common than at upper secondary level. Entry follows completion of primary education and typical duration is 3 years. In some countries, the end of this level marks the end of compulsory education.	ISCED 2
Upper secondary education Stronger specialisation than at lower secondary level. Programmes offered are differentiated by orientation: general or vocational. Typical duration is 3 years.	ISCED 3
Post-secondary non-tertiary education Serves to broaden rather than deepen the knowledge, skills and competencies gained in upper secondary level. Programmes may be designed to increase options for participants in the labour market, for further studies at tertiary level, or both. Usually, programmes at this level are vocationally oriented.	ISCED 4
Short-cycle tertiary education Serves to deepen the knowledge developed at previous levels by imparting new techniques, concepts and ideas not generally covered in upper secondary education.	ISCED 5
Bachelor's or equivalent level Designed to provide participants with intermediate academic and/or professional knowledge, skills and competencies, leading to a first degree or equivalent qualification. Typical duration: 3-4 years full-time study.	ISCED 6
Master's or equivalent level Stronger specialisation and more complex content than bachelor's level. Designed to provide participants with advanced academic and/or professional knowledge. May have a substantial research component.	ISCED 7
Doctoral or equivalent level Designed to lead to an advanced research qualification. Programmes at this level are devoted to advanced study and original research, and exist in both academic and professional fields.	ISCED 8

Figura 9 - Níveis de Educação OCDE

Para comparar o resultado obtido previamente no nível Brasil, em que temos dados de professores e alunos disponibilizados pelo Inep, iremos incluir o valor total dos alunos de mestrado e doutorado, segue a tabela com os valores retirados do MCTIC: (https://www.mctic.gov.br/mctic/opencms/indicadores/detalhe/Recursos_Humanos/RH_3.5.1.html):

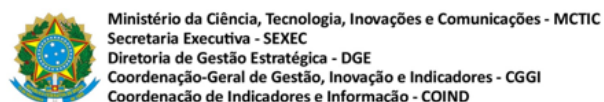


Tabela 3.5.1

Brasil: Alunos matriculados e titulados nos cursos de mestrado e doutorado, ao final do ano, 1998-2017

Ano	Doutorado matriculados	Doutorado titulados	Mestrado acadêmico matriculados	Mestrado acadêmico titulados	Mestrado profissional matriculados	Mestrado profissional titulados	Total Matriculados
1998	26.697	3.915	49.387	12.351	-	-	
1999	29.895	4.831	54.792	14.938	589	43	85.276
2000	32.900	5.318	60.425	17.611	1.131	210	94.456
2001	35.134	6.040	62.353	19.651	2.956	362	100.443
2002	37.728	6.894	63.990	23.457	4.350	987	106.068
2003	40.213	8.094	66.951	25.997	5.065	1.652	112.229
2004	41.261	8.093	69.190	24.755	5.809	1.903	116.260
2005	43.942	8.989	73.805	28.605	6.301	2.029	124.048
2006	46.572	9.366	79.050	29.742	6.798	2.519	132.420
2007	49.667	9.915	84.356	30.559	7.638	2.331	141.661
2008	52.750	10.711	88.295	33.360	9.073	2.654	150.118
2009	57.917	11.638	93.016	35.686	10.135	3.102	161.068
2010	64.588	11.314	98.611	36.247	10.213	3.343	173.412
2011	71.890	12.321	105.240	39.544	12.505	3.689	189.635
2012	79.478	13.912	109.515	42.878	14.724	4.260	203.717
2013	88.337	15.650	109.720	45.490	20.810	6.045	218.867
2014	95.383	17.286	114.341	46.245	25.326	6.998	235.050
2015	102.207	18.996	120.050	47.644	28.384	9.023	250.641
2016	107.640	20.603	126.436	49.002	32.742	10.612	266.818
2017	112.004	21.591	129.220	50.306	37.568	10.841	278.792

Figura 10 - Alunos de Pós Graduação

Estes dados da OCDE vêm de forma resumida, portanto foi necessário pouca limpeza e transformação dos mesmos. Esta transformação foi feita toda na própria ferramenta do Power BI (Power Query) o que gerou maior rapidez.

Além destes, buscou-se dados a respeito da publicação de artigos por país. Os dados são do data world bank (https://data.worldbank.org/indicator/ip.jrn.artc.sc?most_recent_value_desc=true). Serão usados para encontrar uma correlação entre o percentual investido do PIB com o total de artigos publicados em um determinado ano.

Além destes dados, foi procurado também o valor do PIB dos países para se ter uma real noção de quanto foi investido por país no setor terciário. Os dados são da OCDE (<https://data.oecd.org/gdp/gross-domestic-product-gdp.htm>) e não contam com filtros específicos, vale lembrar que os valores estão em milhões de dólares e não são valores reais do PIB, mas um PIB nominal.

O Produto interno Bruto é a medida padrão criada para a produção de valores e serviços em um país durante um período de tempo. Mede também o que foi recebido por aquela produção, ou o total gasto nos bens e serviços (menos importação).

Em quanto o PIB é o indicador mais importante que captura a atividade econômica, falha em relação ao medida de bem estar de pessoas para o qual outros indicadores seriam mais apropriados. Este indicador é baseado no PIB nominal (também chamado de PIB para preços atuais ou PIB em valor) e é disponibilizado em milhões de dólares. Os dados são compilados de acordo com o 'System Nacional Accounts(SNA) de 2008. Esse indicador é menos utilizado para comparações relacionadas ao tempo, devido à não ser apenas utilizado o crescimento real, mas sim a variações em preços e PPPs (paridade do poder de compra).

[Traduzido do inglês]

Além destes dados foi inserida a dimensão País para poder comparar os fatos das diferentes fontes por país. Esta dimensão foi populada com os dados de um arquivo CSV baixado com os dados de artigos mundiais.

As informações obtidas são mostradas através de dashboards interativos que serão mostrados a seguir:

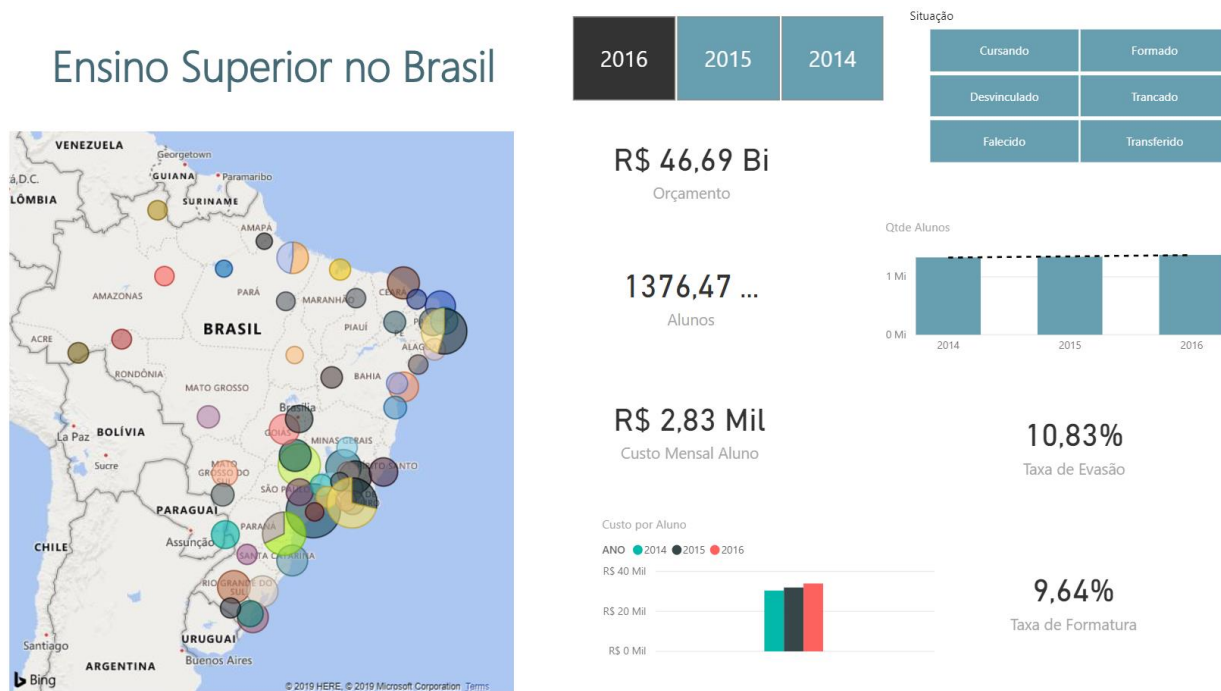


Figura 11 - Ensino Superior no Brasil

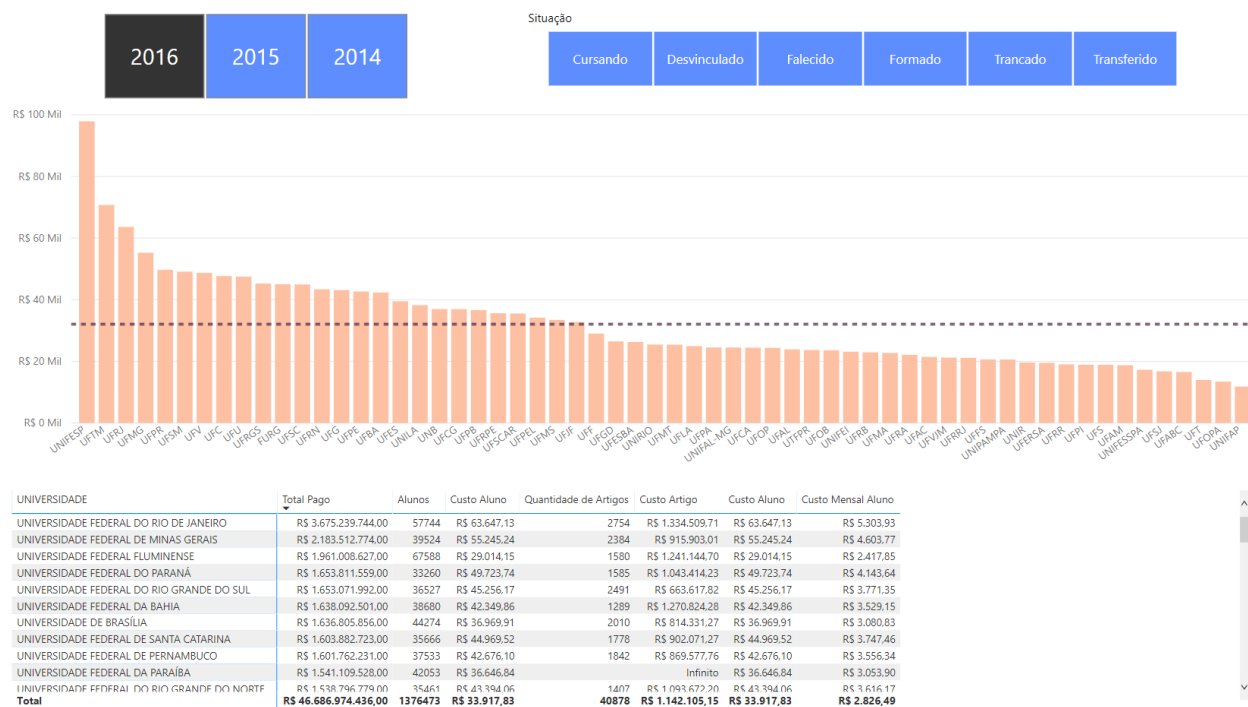


Figura 12 - Custo por Universidade Federal

Orçamento por Universidade

2016

2015

2014

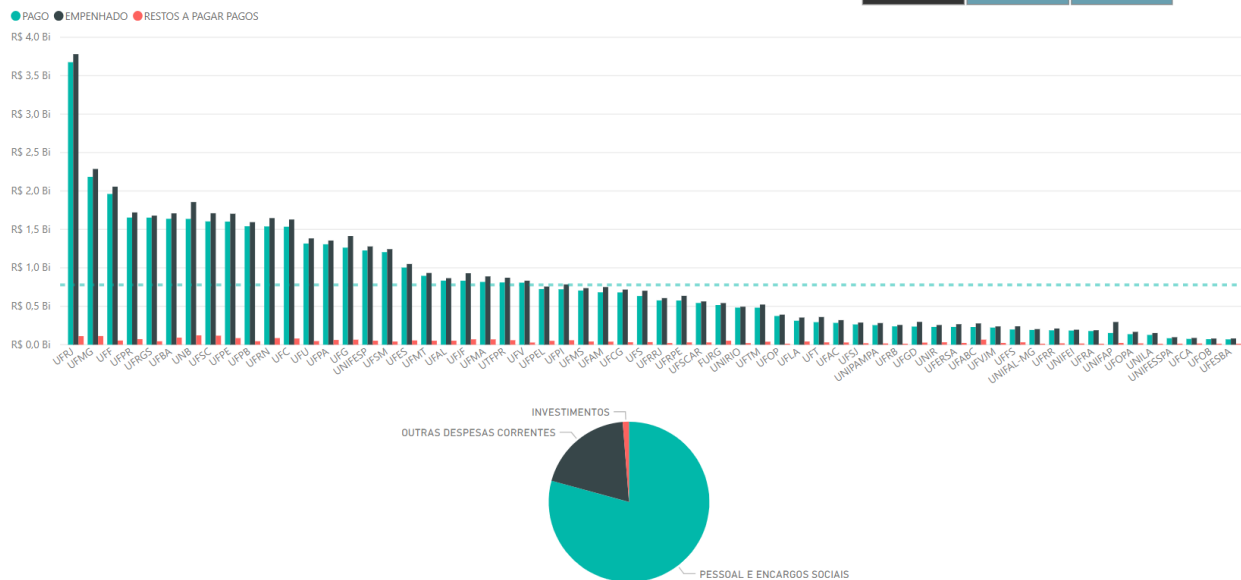


Figura 13 - Orçamento por universidade federal

2016

2015

2014

SITUAÇÃO

Cursando

Desvinculado

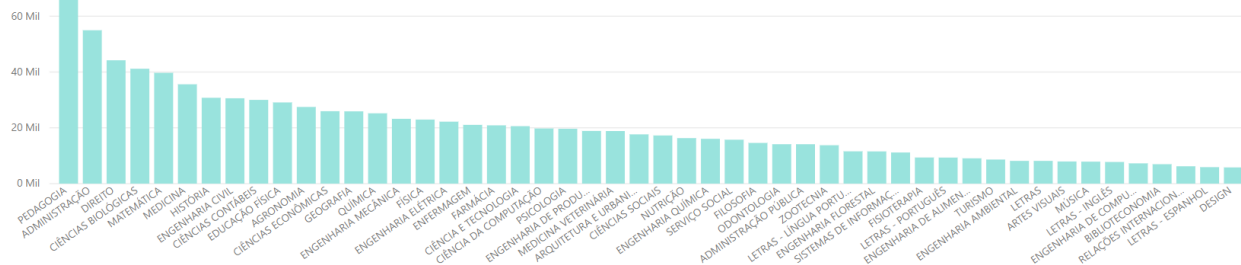
Falecido

Formado

Trancado

Transferido

Alunos por CURSO



CURSO	Alunos	Alunos Cursando	IDADE	Alunos de Evasão	Taxa Evasão	Alunos Formados	Taxa Formatura
PEDAGOGIA	66438	43432	30.41	5820	8.76%	9618	14.48%
ADMINISTRAÇÃO	44214	33707	26.16	2364	5.35%	5777	13.07%
DIREITO	54972	37527	26.31	5566	10.13%	5205	9.47%
CIÊNCIAS BIOLÓGICAS	41177	25935	24.93	4832	11.73%	4474	10.87%
MEDICINA	35599	29734	23.59	817	2.30%	4260	11.97%
ENGENHARIA CIVIL	30583	23773	23.52	1699	5.56%	3406	11.14%
EDUCAÇÃO FÍSICA	29096	20020	26.54	2858	9.82%	3389	11.65%
CIÊNCIAS CONTÁBEIS	30008	21203	26.78	2983	9.94%	3070	10.23%
HISTÓRIA	30766	19829	27.96	4054	13.18%	2925	9.51%
GEOGRAFIA	25907	16261	28.54	3464	13.37%	2907	11.22%
AGRONOMIA	27482	20332	24.13	2518	9.16%	2888	10.51%
ENFERMAGEM	21065	14992	24.62	2044	9.70%	2798	13.28%
MATEMÁTICA	39709	22496	27.36	6820	17.17%	2564	6.46%
Total	1376473	950510	26,20	149110	10,83%	132669	9,64%

Figura 14 - Alunos por Curso

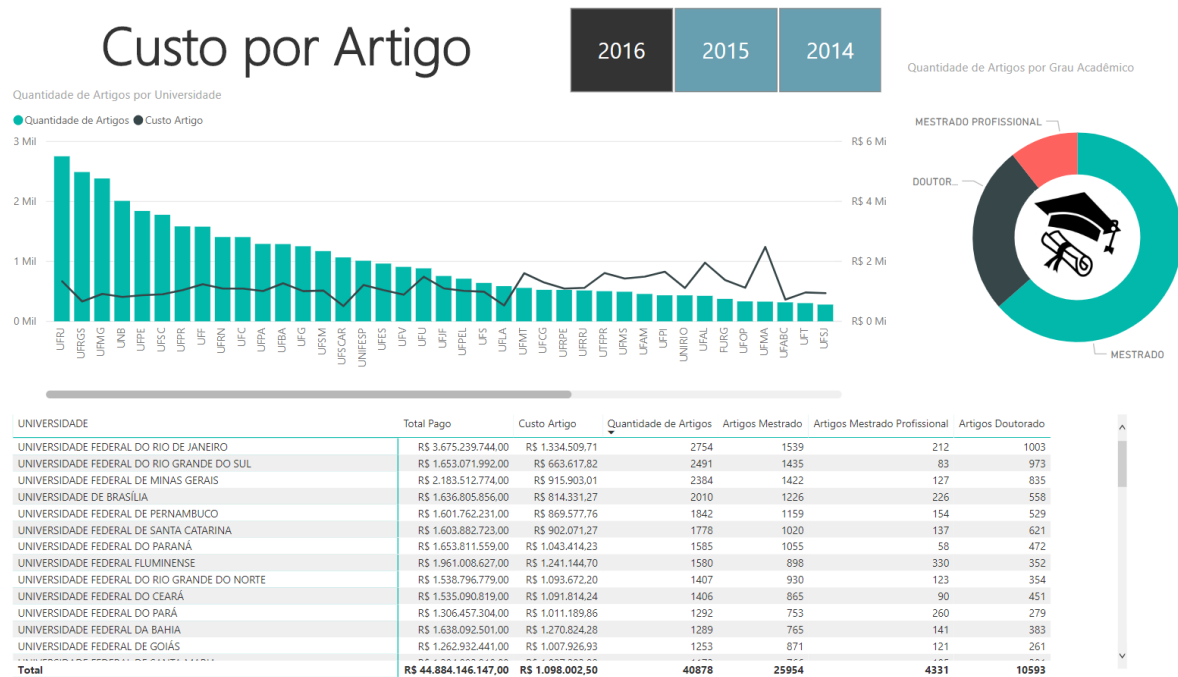
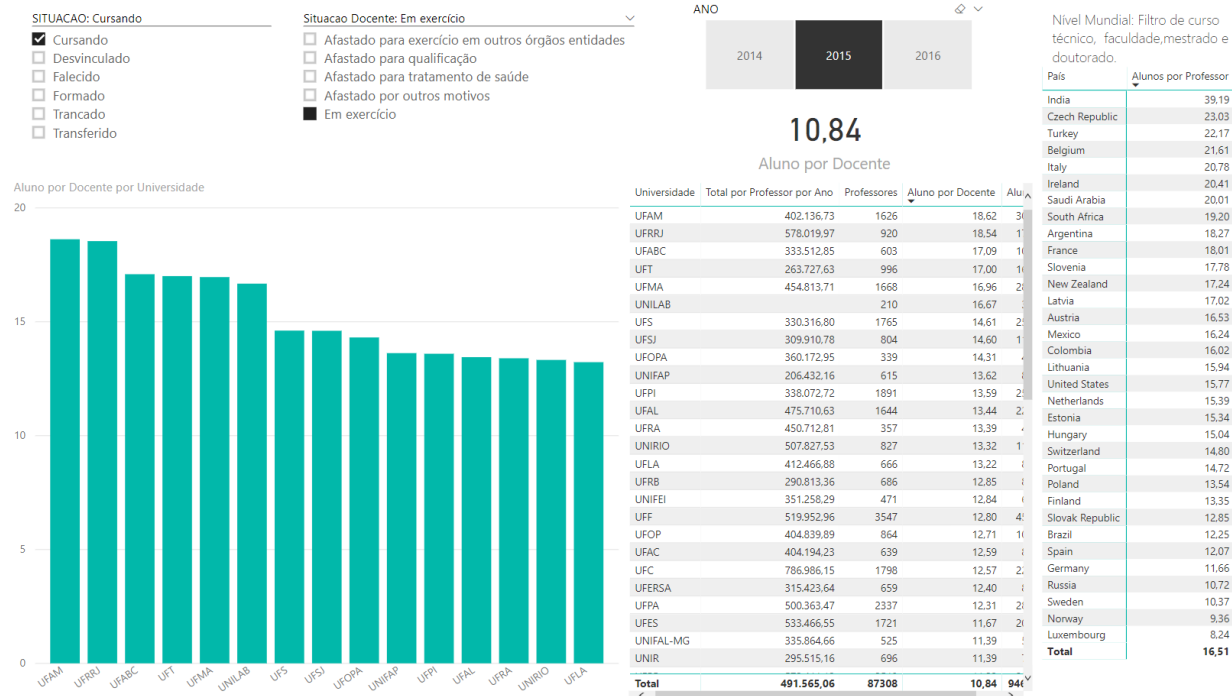


Figura 15 - Custo por artigo

Alunos por docente no Brasil e no mundo.



Aluno por Docente por Universidade

Universidade	Total por Professor por Ano	Professores	Aluno por Docente
UFAM	402.136,73	1626	18,62
UFRRJ	578.019,97	920	18,54
UFABC	333.512,85	603	17,09
UFT	263.727,63	996	17,00
UFMA	454.813,71	1668	16,96
UNILAB		210	16,67
UFS	330.316,80	1765	14,61
UFSJ	309.910,78	804	14,60
UFOPA	360.172,95	339	14,31
UNIFAP	206.432,16	615	13,62
UFPI	338.072,72	1891	13,59
UFAL	475.710,63	1644	13,44
UFRA	450.712,81	357	13,39
UNIRIO	507.827,53	827	13,32
UFLA	412.466,88	666	13,22
UFRB	290.813,36	686	12,85
UNIFEI	351.258,29	471	12,84
UFF	519.952,96	3547	12,80
UFOP	404.839,89	864	12,71
UFAC	404.194,23	639	12,59
UFC	786.986,15	1798	12,57
UFERSA	315.423,64	659	12,40
UFPA	500.363,47	2337	12,31
UFES	533.466,55	1721	11,67
UNIFAL-MG	335.864,66	525	11,39
UNIR	295.515,16	696	11,39
Total	491.565,06	87308	10,84

Figura 16 - Alunos por docente

OBS.: Vale lembrar que a diferença entre a relação aluno docente no Brasil e a relação Aluno docente no mundo para o Brasil fica por conta dos alunos de pós graduação (ISCED 7 e 8) que no ano de 2015 foi de 250.000 alunos de acordo com dados da Capes.



Figura 17 - Correlação entre investimento e artigos

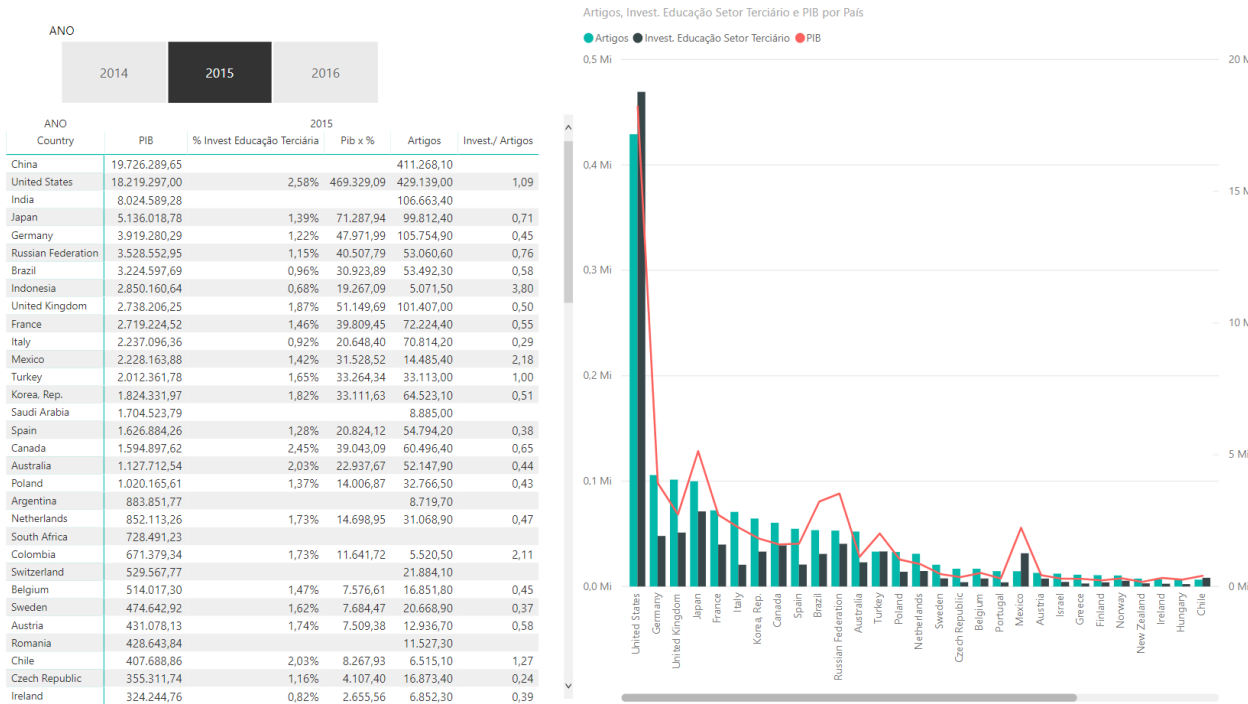


Figura 18 - PIB e Artigos por País

sk_tempo	DATA	ANO	MES	NOME_MES	MES_ABREV
201610	quinta-feira, 1 de outubro de 2020	2016	10	Outubro	Out
201611	quinta-feira, 1 de outubro de 2020	2016	11	Novembro	Nov
201612	quinta-feira, 1 de outubro de 2020	2016	12	Dezembro	Dez
201701	quinta-feira, 1 de outubro de 2020	2017	1	Janeiro	Jan
201702	quinta-feira, 1 de outubro de 2020	2017	2	Fevereiro	Fev
201703	quinta-feira, 1 de outubro de 2020	2017	3	Março	Mar
201704	quinta-feira, 1 de outubro de 2020	2017	4	Abril	Abr
201705	quinta-feira, 1 de outubro de 2020	2017	5	Mai	Mai
201706	quinta-feira, 1 de outubro de 2020	2017	6	Junho	Jun
201707	quinta-feira, 1 de outubro de 2020	2017	7	Julho	Jul
201708	quinta-feira, 1 de outubro de 2020	2017	8	Agosto	Ago
201709	quinta-feira, 1 de outubro de 2020	2017	9	Setembro	Set
201710	quinta-feira, 1 de outubro de 2020	2017	10	Outubro	Out
201711	quinta-feira, 1 de outubro de 2020	2017	11	Novembro	Nov
201712	quinta-feira, 1 de outubro de 2020	2017	12	Dezembro	Dez
201801	quinta-feira, 1 de outubro de 2020	2018	1	Janeiro	Jan
201802	quinta-feira, 1 de outubro de 2020	2018	2	Fevereiro	Fev
201803	quinta-feira, 1 de outubro de 2020	2018	3	Março	Mar

Figura 21 - Exemplo de dados da d_tempo

sk_universidade	UNIVERSIDADE	UNIV_SIGLA	MUNICIPIO	UF	ORGANIZACAO_ACADEMICA	ADMINISTRACAO	version	date_from	date_to	CO_JES	Municipio, UF
249	FACULDADES ESEFAP		Tupã	SP	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	270	Tupã, SP
992	FACULDADE AMÉRICA LATINA		Caxias do Sul	RS	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	3596	Caxias do Sul, RS
1674	Faculdades Integradas do Rio Grande do Norte - FANORTES		Natal	RN	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	3643	Natal, RN
1865	FACULDADE DE TECNOLOGIA CÉSAR LATTES		Itu	SP	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	4209	Itu, SP
2414	FACULDADE UNINASSAU MANAUS		Manaus	AM	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	domingo, 9 de dezembro de 2018	4135	Manaus, AM
2416	UNAMA FACULDADE DA AMAZÔNIA DE ANANINDEUA		Ananindeua	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	35642	Ananindeua, PA
2417	FACULDADE UNINASSAU BELÉM		Belém	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	domingo, 9 de dezembro de 2018	13982	Belém, PA
2418	Faculdade Brasil Inteligente		Belém	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	18093	Belém, PA
2419	Faculdade Cosmopolita		Belém	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	18077	Belém, PA
2424	FACULDADE MIRIENDE		IGARAPE-MIRI	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	18516	IGARAPE-MIRI, PA
2426	FACULDADE DOS CARAJÁS		Marabá	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	domingo, 9 de dezembro de 2018	13728	Marabá, PA
2428	Faculdade Metropolitana de Paragominas		Paragominas	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	17666	Paragominas, PA
2429	Faculdade Metropolitana de Parauapebas		Parauapebas	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	terça-feira, 31 de dezembro de 2199	17667	Parauapebas, PA
2431	Unama Faculdade da Amazônia de Santarém		Santarém	PA	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	domingo, 9 de dezembro de 2018	757	Santarém, PA
5489	ESFAP PAPE - ADEMIÁ		Marabá	AD	Faculdade	Privada com fins lucrativos	1	segunda-feira, 1 de janeiro de 1900	domingo, 9 de dezembro de 2018	11609	Marabá, AD

Figura 22 - Exemplo de dados da d_universidade

Country Code ▾	Region ▾	IncomeGroup ▾	SpecialNotes ▾	TableName ▾	Column1 ▾	País ▾
ABW	Latin America & Caribbean	High income		Aruba		
AFG	South Asia	Low income		Afghanistan		
AGO	Sub-Saharan Africa	Lower middle income		Angola		
ALB	Europe & Central Asia	Upper middle income		Albania		
AND	Europe & Central Asia	High income		Andorra		
ARE	Middle East & North Africa	High income		United Arab Emirates		
ARM	Europe & Central Asia	Upper middle income		Armenia		
ASM	East Asia & Pacific	Upper middle income		American Samoa		
ATG	Latin America & Caribbean	High income		Antigua and Barbuda		
AZE	Europe & Central Asia	Upper middle income		Azerbaijan		

Figura 23- Exemplo de dados da d_pais

LOCATION ▾	INDICATOR ▾	SUBJECT ▾	MEASURE ▾	FREQUENCY ▾	TIME ▾	Value ▾	Flag Codes ▾	País ▾	ANOMES ▾
AUS	EDUEXP	TRY	PC_GDP	A	2000	1,50%		Austrália	200001
AUS	EDUEXP	TRY	PC_GDP	A	2005	1,52%		Austrália	200501
AUS	EDUEXP	TRY	PC_GDP	A	2008	1,54%		Austrália	200801
AUS	EDUEXP	TRY	PC_GDP	A	2009	1,62%		Austrália	200901
AUS	EDUEXP	TRY	PC_GDP	A	2010	1,68%		Austrália	201001
AUS	EDUEXP	TRY	PC_GDP	A	2011	1,63%		Austrália	201101
AUS	EDUEXP	TRY	PC_GDP	A	2012	1,60%		Austrália	201201
AUS	EDUEXP	TRY	PC_GDP	A	2013	1,70%		Austrália	201301
AUS	EDUEXP	TRY	PC_GDP	A	2014	1,85%		Austrália	201401
AUS	EDUEXP	TRY	PC_GDP	A	2015	2,03%		Austrália	201501
AUT	FDIEXP	TRY	PC_GDP	A	2012	1,75%		Austria	201201

Figura 24-Exemplo de dados da f_%PIB

sk_artigo ▾	sk_tempo ▾	sk_universidade ▾	PROGRAMA ▾	SUBTIPO PRODUÇÃO ▾	AREA ▾	LINHA_PESQUISA ▾	GRAU_ACADEMICO ▾	version ▾	date_from ▾	date_to ▾
240072	201301	2278	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
241789	201301	495	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
242053	201301	495	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246782	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246790	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246848	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246862	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246873	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246879	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246886	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246887	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246889	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246893	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
246896	201301	4	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262783	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262868	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262871	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262881	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262912	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			
262916	201301	2263	LETRAS	DISSERTAÇÃO	NÃO INFORMADO		MESTRADO			

Figura 25-Exemplo de dados da f_artigo

COUNTRY	Country.1	ISC11_LEVEL_CAT	Level of education	REF_SECTOR	Reference sector	INDICATOR	Indicator.1	Value	ANOMES
AUT	Austria	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	16,884	201401
AUT	Austria	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	16,527	201501
AUT	Austria	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	16,36	201601
BEL	Belgium	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	21,196	201401
BEL	Belgium	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	21,606	201501
BEL	Belgium	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	20,288	201601
CZE	Czech Republic	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	22,353	201401
CZE	Czech Republic	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	23,033	201501
CZE	Czech Republic	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	18,888	201601
DNK	Denmark	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	14,256	201401
FIN	Finland	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	13,245	201401
FIN	Finland	L5T8	Total tertiary education (ISCED2011 levels 5 to 8)	INST_PUB	Public educational institutions	PERS_RATIO_INST	Ratio of students to teaching staff by type of institutions	13,347	201501

Figura 26- Exemplo de dados da f_docentes_mundo

pk_matricula	sk_tempo	COD_CURSO	IDADE	SEXO	version	date_from	date_to	COD_IES	N_ALUNOS	SITUACAO
202	201701	141	20					2	1	Transferido
431	201701	100358	20					1952	1	Transferido
859	201701	58524	20					29	1	Transferido
1359	201701	1279825	20					11645	1	Transferido
2173	201701	1127925	20					1808	1	Transferido
2762	201701	1206306	20					17828	1	Transferido
2787	201701	107090	20					641	1	Transferido
3093	201701	108378	20					1690	1	Transferido
3650	201701	1323694	20					1497	1	Transferido
3821	201701	71261	20					2616	1	Transferido
4033	201701	1192415	20					2	1	Transferido
4140	201701	20290	20					170	1	Transferido
4373	201701	303167	20					56	1	Transferido
4609	201701	1182592	20					1149	1	Transferido

Figura 27- Exemplo de dados da f_Matricula

pk_orcamento	sk_universidade	sk_tempo	DES_EMPENHADO	DES_LIQUIDADO	DES_PAGO	DES_RESTOS_PAGOS	version	date_from	date_to	NOME_DESPESA	ELEMENTO_DESPESA
80783	2075	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
80791	2278	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81375	2075	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81393	2075	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81413	2278	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81464	2	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81478	2075	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
81513	2	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82045	2161	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82059	3	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82119	1306	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82130	2278	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82143	2161	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82194	136	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82757	2	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO
82779	136	201612	R\$ 0,00	R\$ 0,00	R\$ 0,00	R\$ 0,00				OUTRAS DESPESAS CORRENTES	MATERIAL DE CONSUMO

Figura 28- Exemplo de dados da f_orcamento

Country Code	Country Name	Atributo	Artigos	ANOMES
AFG	Afghanistan	2013	27,3	201301
AGO	Angola	2013	24,1	201301
ALB	Albania	2013	359,5	201301
AND	Andorra	2013	5,9	201301
ARE	United Arab Emirates	2013	1775,6	201301
ARG	Argentina	2013	8268,5	201301
ARM	Armenia	2013	552	201301
ATG	Antigua and Barbuda	2013	1,9	201301
AUS	Australia	2013	49921,9	201301
AUT	Austria	2013	12557,6	201301
AZF	Azerbaijan	2013	481,1	201301

Figura 29- Exemplo de dados da f_Artigo_mundo

Country Code	Country Name	Atributo	Artigos	ANOMES
AFG	Afghanistan	2013	27,3	201301
AGO	Angola	2013	24,1	201301
ALB	Albania	2013	359,5	201301
AND	Andorra	2013	5,9	201301
ARE	United Arab Emirates	2013	1775,6	201301
ARG	Argentina	2013	8268,5	201301
ARM	Armenia	2013	552	201301
ATG	Antigua and Barbuda	2013	1,9	201301
AUS	Australia	2013	49921,9	201301
AUT	Austria	2013	12557,6	201301
AZF	Azerbaijan	2013	481,1	201301

Figura 30- Exemplo de dados de f_docente

LOCATION	INDICATOR	SUBJECT	MEASURE	Value	ANOMES
AUS	GDP	TOT	MLN_USD	1118350,267	201401
AUS	GDP	TOT	MLN_USD	1127712,5423	201501
AUS	GDP	TOT	MLN_USD	1215897,6861	201601
AUT	GDP	TOT	MLN_USD	417059,5187	201401
AUT	GDP	TOT	MLN_USD	431078,1342	201501
AUT	GDP	TOT	MLN_USD	451297,2461	201601
BEL	GDP	TOT	MLN_USD	499974,7211	201401
BEL	GDP	TOT	MLN_USD	514017,2959	201501
BEL	GDP	TOT	MLN_USD	535003,4891	201601
CAN	GDP	TOT	MLN_USD	1621391,8322	201401
CAN	GDP	TOT	MLN_USD	1594897,6212	201501
CAN	GDP	TOT	MLN_USD	1628880,1751	201601
CZE	GDP	TOT	MLN_USD	339582,5988	201401

Figura 31- Exemplo de dados de f_PIB

4.1 Comparativo

Nesta etapa iremos mostrar como está a situação da UFG comparando-a com outras faculdades do mesmo padrão e ainda comparar as mais eficientes. Se compararmos baseado na quantidade de alunos, podemos notar que possui um orçamento adequado, e produz uma quantidade de artigos razoável, tem um custo por artigo mais baixo se comparadas as faculdades com quantidade próxima de alunos. Porém percebe-se que é a que tem a menor média de alunos por docente. Se Comparada com a média brasileira a UFG tem o custo por aluno mais caro que a média, tem o custo por artigo mais barato e tem uma quantidade de alunos por docente mais baixa que a média nacional.

UNIVERSIDADE	Total Pago	Alunos	Custo Aluno	Quantidade de Artigos	Custo Artigo	Custo Aluno	Custo Mensal Aluno	Aluno por Docente
UNIVERSIDADE FEDERAL DE ALAGOAS	R\$ 833.393,047,00	34875	R\$ 23.896,57	426	R\$ 1.956.321,71	R\$ 23.896,57	R\$ 1.991,38	20,29
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ	R\$ 810.373.421,00	34180	R\$ 23.708,99	503	R\$ 1.611.080,36	R\$ 23.708,99	R\$ 1.975,75	11,75
UNIVERSIDADE FEDERAL DE SERGIPE	R\$ 633.317.656,00	33525	R\$ 18.890,91	641	R\$ 988.015,06	R\$ 18.890,91	R\$ 1.574,24	18,52
UNIVERSIDADE FEDERAL DO PARANÁ	R\$ 1.653.811.559,00	33260	R\$ 49.723,74	1585	R\$ 1.043.414,23	R\$ 49.723,74	R\$ 4.143,64	13,80
UNIVERSIDADE FEDERAL DO CEARÁ	R\$ 1.535.090.819,00	32189	R\$ 47.689,92	1406	R\$ 1.091.814,24	R\$ 47.689,92	R\$ 3.974,16	16,64
UNIVERSIDADE FEDERAL DE GOIÁS	R\$ 1.262.932.441,00	29283	R\$ 43.128,52	1253	R\$ 1.007.926,93	R\$ 43.128,52	R\$ 3.594,04	9,95
UNIVERSIDADE FEDERAL DE UBERLÂNDIA	R\$ 1.316.211.993,00	27709	R\$ 47.501,24	884	R\$ 1.488.927,59	R\$ 47.501,24	R\$ 3.958,44	13,91
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO	R\$ 576.414.515,00	27245	R\$ 21.156,71	515	R\$ 1.119.251,49	R\$ 21.156,71	R\$ 1.763,06	29,65
UNIVERSIDADE FEDERAL DE JUIZ DE FORA	R\$ 832.587.048,00	25421	R\$ 32.751,94	757	R\$ 1.099.850,79	R\$ 32.751,94	R\$ 2.729,33	14,68
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	R\$ 1.003.520.731,00	25404	R\$ 39.502,47	964	R\$ 1.040.996,61	R\$ 39.502,47	R\$ 3.291,87	13,40
UNIVERSIDADE FEDERAL DE SANTA MARIA	R\$ 1.204.093.910,00	24509	R\$ 49.128,64	1172	R\$ 1.027.383,88	R\$ 49.128,64	R\$ 4.094,05	11,52
Total	R\$ 46.686.974.436,00	1376473	R\$ 33.917,83	40878	R\$ 1.142.105,15	R\$ 33.917,83	R\$ 2.826,49	14,59

Figura 32 - Comparativo UFG (alunos)

A universidade brasileira que mais gastou em 2016 foi a UFRJ com R\$ 3.675.239.744 e a que menos gastou foi a UFESBA (Universidade Federal do sul da Bahia) com R\$ 69.781.822 em 2016.

A universidade com menor custo por aluno foi a Universidade Federal do Amapá com R\$ 11.821 por aluno por ano, já a Universidade Federal de São Paulo com R\$ 97.922 por aluno por ano foi a que mais gastou em 2016.

A universidade que teve o maior custo por artigo foi a Universidade Federal do triângulo mineiro com R\$ 2.957.641 por artigo, já a Universidade Federal de São Carlos foi a que teve o menor custo por artigo de 2016 com R\$ 508.801.

A universidade que teve a maior média de alunos por docente é a UFRJ com 19,98 alunos e a que teve menor média foi a Universidade Federal de São Paulo com 5,91.

Ao comparar o Brasil com o restante do mundo podemos notar que o investimento é baixo se comparado a países menos populosos que investem bem mais, como por exemplo o Canada, Turquia, Coréia do sul, México, França entre outros.

Country	PIB (MM)	% Invest Educação Terciária	Pib x %	Artigos	Invest./ Artigos
China	19.726.289,65			411.268,10	
United States	18.219.297,00	2,58%	469.329,09	429.139,00	1,09
India	8.024.589,28			106.663,40	
Japan	5.136.018,78	1,39%	71.287,94	99.812,40	0,71
Germany	3.919.280,29	1,22%	47.971,99	105.754,90	0,45
Russian Federation	3.528.552,95	1,15%	40.507,79	53.060,60	0,76
Brazil	3.224.597,69	0,96%	30.923,89	53.492,30	0,58
Indonesia	2.850.160,64	0,68%	19.267,09	5.071,50	3,80
United Kingdom	2.738.206,25	1,87%	51.149,69	101.407,00	0,50
France	2.719.224,52	1,46%	39.809,45	72.224,40	0,55
Italy	2.237.096,36	0,92%	20.648,40	70.814,20	0,29
Mexico	2.228.163,88	1,42%	31.528,52	14.485,40	2,18
Turkey	2.012.361,78	1,65%	33.264,34	33.113,00	1,00
Korea, Rep.	1.824.331,97	1,82%	33.111,63	64.523,10	0,51
Saudi Arabia	1.704.523,79			8.885,00	
Spain	1.626.884,26	1,28%	20.824,12	54.794,20	0,38

Figura 33 - Investimento por país

5. Discussão

Inicialmente, olhando as informações de custo a nível de Brasil percebemos que temos um valor bem alto, principalmente se comparado à universidades privadas. O custo médio de 2.800 reais de mensalidade, como pode ser observado no ano de 2016, caso fosse repassado ao estudante este recorreria ao ensino privado que na maioria dos cursos seria mais barato. Pode-se observar também que a taxa de formatura é baixa, se considerarmos que um formando leva em média 5 anos para formar, uma taxa de formatura ideal seria de 20% ao ano. Ou ainda que demore 6 anos em média essa taxa vai para 16,6% ainda bem longe do valor observado. O que pode explicar essa discrepância entre valores reais e ideais além, da alta taxa de evasão é a desmotivação do aluno, falta de informação a respeito do custo real da sociedade ou ainda um foco maior para o mercado de trabalho e empreendedorismo.

Com base nessas informações é importante mostrar para o aluno, o principal fator das universidades, o quanto ele custa para os cofres públicos para que haja conscientização por parte deste. Este investimento que não traz retorno (formaturas) não é interessante para o país que paga um alto valor pela educação, alta mensalidade por um bom tempo, e forma poucos alunos.

Olhando as informações providas pelos dados obtidos podemos ver não só a taxa de formatura como também o custo anual que tem o aluno de cada universidade federal do Brasil. Observamos que há uma grande diferença de orçamento e custo por aluno nas universidades (divisão do orçamento total pela quantidade de alunos matriculados), essa diferença mostra que no Brasil não há uma relação direta entre dinheiro investido e número de formandos por universidade, podemos perceber que em quanto um aluno da Universidade Federal de São Paulo custa quase R\$100.000 no ano, um aluno da Universidade Federal do Amapá custa apenas R\$11.000, informações do ano de 2016. Através dessas discrepâncias podemos perceber que não há uma proporcionalidade, se aumentarmos o orçamento para educação em 10% não se pode dizer precisamente que irá melhorar proporcionalmente a taxa de formatura. Comparando com a UFG podemos ver que estamos mais próximos da média nacional de custo por aluno que é de aproximadamente R\$34.000, em quanto na UFG é de aproximadamente R\$ 43.000 .

Através desta análise podemos perceber que esta somente não é suficiente para obter um diagnóstico da educação superior no Brasil, uma vez que podemos perceber o tamanho das divergências e ineficiências de umas e eficiências de outras, a partir daí iremos analisar também dados de artigos publicados por IES (Instituição de Ensino Superior).

Ao analisar a quantidade de artigos publicados podemos comparar melhor as IES, constatou-se que a Universidade que mais produz artigos é a UFRJ, com 2754 artigos publicados em 2016, porém não diz muito ao percebermos que é a segunda maior IES pública do Brasil em quantidade de alunos. Essa informação isolada não diz respeito com relação à eficiência, porém ao analisar o custo por aluno notamos que este é bem alto, aproximadamente R\$ 63.000. E ao comparar com uma universidade que produz menos artigos como a Universidade Federal do Oeste da Bahia que produziu em 2016 apenas 14 artigos publicados e com um custo por aluno de aproximadamente R\$ 23.500.

Além dessa análise, fizemos também a análise de artigos dividido pela quantidade de alunos da IES, assim conseguimos obter informações a respeito de produção acadêmica da universidade sem entrar em mérito de quantidade de alunos. Assim obtivemos o seguinte resultado: a melhor taxa de artigos por aluno foi na UNIFESP, com 0,081 artigo por aluno ou 8,1%, em quanto que as piores foram na Universidade Federal do

Oeste do Pará e na Universidade Federal do Amapá, ambas com uma taxa de 0,004 artigos por aluno ou 0,4%, não distante a Universidade Federal do Oeste da Bahia apareceu com 0,5%.

Outra métrica utilizada para medir a eficiência nas universidades federais foi a quantidade de artigos publicados dividido pelo orçamento da universidade (custo por artigo). O custo artigo tem também grandes divergências, filtrando apenas as universidades com mais de 100 artigos publicados no ano, a Universidade Federal de São Carlos tem o custo aproximado de R\$ 508.000 por artigo em quanto a Universidade Federal do Triângulo Mineiro tem o custo aproximado de R\$ 2.957.000 por artigo no ano de 2016. Na UFG o custo por artigo foi de aproximadamente R\$ 1.008.000.

A partir destas informações fica difícil dizer se uma Universidade é eficiente ou não, pois percebe-se que quando o custo por aluno é pequeno, a quantidade de artigos publicados também é pequena, apesar de ter uma eficiência para formação de alunos a um baixo custo, por outro lado para grandes custos por aluno percebe-se que existe uma maior taxa de publicação de artigos acadêmicos.

Uma informação interessante é a quantidade de alunos formandos por curso. Percebe-se que os cursos que mais formam alunos são pedagogia, administração e direito. Isso se deve principalmente à grande quantidade de alunos que entram para esses cursos e a baixa taxa de evasão observada para tais cursos. Na UFG os cursos com maior número de formandos em 2016 são direito, pedagogia, educação física e agronomia.

Fazendo uma comparação pela quantidade e alunos pode-se perceber a realidade das universidades federais: em um intervalo de 25.000 a 35.000 alunos, em que a UFG tem 29.283 alunos matriculados em 2016, pode-se observar que a quantidade de artigos que são produzidos na federal do Ceará é maior que na UFG, tem um orçamento pouco maior que o da UFG. Ao analisar a quantidade de docentes percebe-se que a UFG tem muito mais docentes e uma relação de alunos por docente consideravelmente menor (16,6 alunos por docente contra 9,95 da UFG). Percebe-se também que a quantidade de artigos que cada docente faz por ano é de cerca de 0,73 em média enquanto na UFG a média é de 0,43 artigo por docente. Devido à grande quantidade de docentes percebe-se que a UFG está inchada, ocorrendo indícios de uma má gestão principalmente se comparado a outras universidades similares.

Ao comparar a UFG a Universidade Federal e Uberlândia podemos perceber que a situação é parecida com a federal do Ceará: A quantidade de alunos por docente na UFU é de 13,91. O orçamento da UFU é bem parecido com a UFG porém tem 1000 alunos a menos que a UFG. Esse inchaço ganha ainda mais força se olharmos pelo lado da localização geográfica.

Feitas as comparações internas no Brasil, partimos para a comparação a nível mundial. Ao comparar o Brasil com países de PIB semelhante, porém pouco menor, notamos que é o que tem menor investimento percentual do PIB em educação pública. Neste mesmo grupo de países, o Brasil investiu 0,96% do PIB. Multiplicando pelos respectivos PIBs para obtermos valores reais, percebemos que investimos menos que no Reino Unido, França, México, Turquia e Coréia, ficando à frente apenas de Itália e Indonésia. Neste grupo de 8, ficamos em antepenúltimo.

Vale mencionar a grande diferença de investimento se comparada à Rússia, um país pertencente aos BRICS (agrupamento de países de mercado emergente em relação ao desenvolvimento econômico) que tem o PIB aproximadamente 10% maior que o do Brasil, porém conta com um investimento em educação pública cerca de 30% maior que o investimento brasileiro. E ainda conta com uma população de 144

milhões de pessoas, bem menor que a do Brasil que está na casa dos 200 milhões de pessoas. O que mostra que o que é investido proporcionalmente é maior.

Além disso ao comparar o Brasil com o mundo, observamos que ainda é um país que tem muito a investir. Investe-se menos de 1% do PIB em educação superior, enquanto países mais desenvolvidos (europeus e norte americanos) como por exemplo países como Estados Unidos, Canadá, Reino Unido, Coreia investiram mais de 1,8% a 2,5% do PIB no setor terciário no ano de 2015.

6. Conclusão

Este trabalho teve o intuito de estudar sobre o modelo de business intelligence voltado para os órgãos públicos federais para determinarmos se os investimentos feitos em educação são suficientes e ainda saber sua posição em relação ao mundo.

Através do projeto de business intelligence é possível mostrar informações reais e atuais a respeito do ensino público superior no Brasil. O fim das falácias a respeito do ensino, aliada a possibilidade de tomada de decisão baseada em fatos são o ponto forte deste projeto.

Através desta análise é possível perceber que para melhorar o nível da educação superior no Brasil, deve-se conscientizar não apenas os alunos, porém toda a cadeia que participa do processo de formação terciária. É preciso conscientizar professores, coordenadores de curso, reitores, secretários de educação entre outros para obter um melhor resultado a nível de Brasil.

Diante de toda essa busca por transparência e governança eficiente, nota-se o a pouca eficiência que ainda temos no ensino superior do Brasil. O alto custo de um aluno para a sociedade e a baixa taxa de formação, juntamente com a alta taxa de evasão são a prova disso.

Se comparado ao ensino de países desenvolvidos o Brasil ainda tem muito o que crescer. Tanto em investimentos como em eficiência na gestão das universidades. É importante tomar conhecimento dessa realidade o quanto antes para que se possa tomar as medidas certas o mais rápido possível.

Pode-se dizer que o Brasil encontra-se com a educação em desenvolvimento, se comparado a países desenvolvidos. Mas para se tornar um país desenvolvido o caminho principal é através da educação.

Vale ressaltar a importância de um projeto de BI para conhecer melhor a máquina pública e realmente saber se o recurso público está sendo usado de forma correta. A partir de agora pode-se avaliar as políticas por meio da eficiência ou da falta dela no ensino superior.

Somente com a conscientização através da informação consegue-se medir e progredir.

7. Referências

[Student-teacher ratio and average class size]

Disponível em: <https://stats.oecd.org/viewhtml.aspx?datasetcode=EAG_PERS_RATIO&lang=en> .

Acessado em: 29/06/2019.

Public spending on education – tertiary

Disponível em: < <https://data.oecd.org/eduresource/public-spending-on-education.htm#indicator-chart>>. Acessado em 29/06/2019.

Gross domestic product (GDP) total

Disponível em: < <https://data.oecd.org/gdp/gross-domestic-product-gdp.htm>>. Acessado em 02/07/2019.

Treinamento em Power BI

Disponível em: <<https://www.passeidireto.com/arquivo/50763368/apostila-treinamento-power-bi-v2>>.

Acessado em: 22/11/2018.

KIMBALL, R.; ROSS, M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd. ed. [S.l.]: Wiley Publishing, 2013. ISBN 1118530802, 9781118530801.

[Luhn 1958]LUHN, H. P. A business intelligence system. IBM J. Res. Dev., IBM Corp., Riverton, NJ, USA, v. 2, n. 4, p. 314–319, out. 1958. ISSN 0018-8646. Disponível em: <http://altaplana.com/ibmrd0204H.pdf>

Gartner Business Intelligence - Glossário IT.

Disponível em: <<https://www.gartner.com/it-glossary/business-intelligencebi/>>. Acessado em: 25/11/2018.

[Rafael Piton] O que é BI – Business Intelligence?

Disponível em: <<https://rafaelpiton.com.br/blog/o-que-e-bi-businessintelligence/>> . Acessado em 25/11/2018.

[Rafael Piton] O que é Dashboard?

Disponível em: <<https://rafaelpiton.com.br/blog/o-que-e-dashboard/>>. Acessado em 25/11/2018.

[Power BI Visualização] Como os cross filtros se afetam no Power BI.

Disponível em: <<https://docs.microsoft.com/enus/power-bi/consumer/end-user-interactions>>.

Acessado em 26/11/2018.

Entenda o corte de verba das universidades federais e saiba como são os orçamentos das 10 maiores

Disponível em: <<https://g1.globo.com/educacao/noticia/2019/05/15/entenda-o-corte-de-verba-das-universidades-federais-e-saiba-como-sao-os-orcamentos-das-10-maiores.ghtml>>. Acessado em 27/06/2019

[Figura Quadrante Gartner] [https://www.qlik.com/us/-](https://www.qlik.com/us/-/media/B0DCF50C816448EEAC8981B0238781B1.ashx)

[/media/B0DCF50C816448EEAC8981B0238781B1.ashx](https://www.qlik.com/us/-/media/B0DCF50C816448EEAC8981B0238781B1.ashx) [Acessado em 27/06/2019].

[Data Warehouse] Data Warehouse. https://en.wikipedia.org/wiki/Data_warehouse. [Acessado em 26/11/2018]

[Data Warehouse Toolkit]

<http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf> . Acessado em 26/11/2018.

[Figura Modelo Estrela] Disponível em: <<https://www.dirceuresende.com/wp-content/uploads/2017/02/Star-schema.jpg>> . Acessado em 28/06/2019.

[Figura Etapas de ETL] Disponível em: <http://bufallos.com.br/bg_br/wp-content/uploads/2017/10/fluxoETL-1.png>. Acessado em 28/06/2019.

[Figura Transformação Pentaho Data Integration] Disponível em:

<<https://www.infointelligence.com.br/wp-content/uploads/2018/06/data-integration-advantage.jpg>> Acessado em 26/06/2019.

[Figura Job] Disponível em: <https://assignittous.com/wp-content/uploads/2018/01/Pentaho_Automation_1.png> Acessado em 28/06/2019.

[Figura Níveis de Educação] Disponível em: <https://www.oecd-ilibrary.org/sites/eag-2018-en/1/1/4/index.html?itemId=/content/publication/eag-2018-en&_csp_=f25abc58c25319ce4c8c74d26c3f40dd&itemIGO=oecd&itemContentType=book> [Acessado em 02/07/2019.]

[Figura Alunos de Pós Graduação] Disponível em:

<https://www.mctic.gov.br/mctic/opencms/indicadores/detalhe/Recursos_Humanos/RH_3.5.1.html> Acessado em 04/07/2019.