



PROJETO DE GRADUAÇÃO

**DESENVOLVIMENTO DE UM SISTEMA DA INFORMAÇÃO
MINERANDO DADOS ABERTOS DA RECEITA FEDERAL PARA
GERAÇÃO DE INSIGHTS COMERCIAIS PARA UMA EMPRESA DE
TELECOM**

Por,
TÚLIO DA COSTA TOMÉ

Brasília, 25 de Abril de 2022

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia
Departamento de Engenharia de Produção

PROJETO DE GRADUAÇÃO

**DESENVOLVIMENTO DE UM SISTEMA DA
INFORMAÇÃO MINERANDO DADOS ABERTOS
DA RECEITA FEDERAL PARA GERAÇÃO DE
INSIGHTS COMERCIAIS PARA UMA EMPRESA
DE TELECOM**

POR,

Túlio da Costa Tomé

Relatório submetido como requisito para obtenção da aprovação
Em projeto de Graduação em Engenharia de Produção 2

Banca Examinadora

Prof. Edgard Oliveira, UnB/EPR (Orientador)

Prof. Ari Melo Mariano, UnB/EPR (Convidado)

Brasília, 25 de Abril de 2022

Dedicatória

Gostaria de agradecer esta conquista primeiramente a Deus, pois sem ele jamais teria chegado aqui e a minha mãe, que mesmo não tendo estudado, sempre trabalhou muito para me manter nos estudos e no caminho do trabalho.

RESUMO

Partindo da demanda específica de uma empresa de telecomunicação fictícia, foi desenvolvido um sistema de informação que traz análises que ajudarão a empresa a ter *insights* e a tomar decisões baseadas em números reais, que estão disponíveis no site da Receita Federal. Neste projeto serão abordado vários conhecimentos de *Business Intelligence (BI)* além do levantamento de um processo de *Extract, transform e load (ETL)*, juntamente com a criação de um banco de dados e um *dashboard* em *Power BI*. A demanda específica da empresa partiu de perguntas realizadas para a empresa em questão. Essas perguntas tiveram o objetivo de levantar os principais questionamentos e necessidades da companhia. Com isso, foi possível levantar o real objetivo da empresa, que gira em torno de encontrar oportunidades de vender seus produtos dentro de todo o território brasileiro para pessoas jurídicas (PJ). Partindo de uma mineração de dados dentro da World Wide Web (WEB), foram localizadas bases de dados da Receita Federal do Brasil, que possui informação de todas as empresas que tem CNPJ cadastrado, onde esses dados abertos podem ser baixados por qualquer pessoa ou empresa para serem realizadas análises de mercado ou para fins comerciais. A partir do download das bases foi realizada uma união dos dados utilizando o próprio *prompt* de comando do Windows (CMD) para ser possível subir um único arquivo para o *Integration Service*. Logo em seguida, foi realizado um processo de extração, transformação e carga de dados (ETL) dentro do *Integration Service* que tem como objetivo deixar alinhados os dados da Receita Federal, tirando inconsistências e erros da base de dados para assim ser possível subir esses dados para um banco de dados. Com esse processo realizado, foi possível desenvolver um banco de dados e colocar todos os dados da Receita Federal dentro de um banco de dados com várias tabelas relacionadas entre si. Partindo da etapa finalizada do banco de dados e do processo ETL foi viável conectar o banco de dados a uma ferramenta de visualização de dados chamada Power BI e realizar todas as análises importantes para solucionar e responder todas as perguntas da empresa.

Palavras-chave: Dados, Análise, *Dashboard*, Dados Abertos, Telecom.

ABSTRACT

Starting from the specific demand of a fictitious telecommunication company, an information system will be developed that will have analyzes that will help the company to have insights and make decisions based on real numbers available on the Federal Revenue website. In this project, various Business Intelligence (BI) knowledge will be addressed in addition to the survey of an Extract, transform and load (ETL) process, along with the creation of a database and a dashboard in Power BI. The specific demand of the company came from questions asked to the company in question. These questions aimed to raise the main questions and needs of the company. With this, it was possible to identify the real objective of the company, which revolves around finding opportunities to sell its products within the entire Brazilian territory to legal entities (PJ). Starting from a data mining within the World Wide Web (WEB), databases of the Federal Revenue of Brazil were located, which have information on all companies that have a registered CNPJ, where these open data can be downloaded by any person or company. to carry out market analysis or for commercial purposes. From the download of the bases, a data union was performed using the windows command prompt (CMD) to be able to upload a single file to the Integration Service. Soon after, a process of extracting, transforming and loading data (ETL) was carried out within the Integration Service that aims to align the data from the Federal Revenue, removing inconsistencies and errors from the database so that it is possible to upload this data. to a database. With this process carried out, it was possible to develop a database and place all the data from the Federal Revenue Service within a database with several tables related to each other. Starting from the finalized stage of the database and the ETL process, it was possible to connect the database to a data visualization tool called Power BI and perform all the important analyzes to solve and answer all the company's questions.

Key words: Data, Analysis, Dashboard, Open Data, Telecommunications.

SUMÁRIO

1. INTRODUÇÃO	11
1.1. CONTEXTUALIZAÇÃO	12
1.2. OBJETIVO GERAL.....	13
1.3. OBJETIVOS ESPECÍFICOS	13
1.4. METODOLOGIA.....	14
2. JUSTIFICATIVA E LIMITAÇÕES DA PESQUISA	20
3. REVISÃO DE LITERATURA.....	23
3.1. DADOS E INFORMAÇÃO.....	23
3.2. DADOS ABERTOS	27
3.3. ETL – (EXTRACT, TRANSFORM, LOAD).....	29
3.4. SISTEMAS DE INFORMAÇÃO	31
3.5. BUSINESS INTELLIGENCE (BI)	33
4. APRESENTAÇÃO DA EMPRESA DE TELECOMUNICAÇÃO	36
4.1. APRESENTAÇÃO DA TELECOM	36
4.2. PRINCIPAIS CLIENTES DA TELECOM E PRODUTOS DA TELECOM	36
4.3. INVESTIGAÇÃO DO PROBLEMA QUE A TELECOM QUE SOLUCIONAR	38
4.4. ANÁLISE DE SOLUÇÃO DO PROBLEMA.....	39
5. DESENVOLVIMENTO DO PROJETO	42
5.1. MAPEAMENTO DAS ETAPAS DO PROJETO PARA A CRIAÇÃO DO BUSINESS INTELLIGENCE	43
6. PROPOSTA DE BI PARA UMA TELECOM	53
6.1. VISÃO GERAL DO PROJETO	53
6.2. SOLUCIONANDO OS PROBLEMAS.....	60
6.3. TENDÊNCIAS.....	62
7. CONCLUSÃO	63
8. REFERÊNCIAS	65

LISTA DE FIGURAS

Figura 1 - CRISP-DM	14
Figura 2 - Processo de Extract Transform Load - ETL.....	15
Figura 3 - Gráfico do Quadrante Mágico ETL.....	17
Figura 4 - Gráfico do Quadrante Mágico Banco de Dados.....	18
Figura 5 - Gráfico do Quadrante Mágico BI.....	19
Figura 6 - DBEngines, 2022	26
Figura 7 - DBEngines, 2022	26
Figura 8 - DBEngines, 2022	27
Figura 9 - DBEngines, 2022	27
Figura 10 - Receita Federal, 2022.....	39
Figura 11 - Receita Federal, 2022.....	40
Figura 12 - Receita Federal, 2022.....	41
Figura 13 - Receita Federal, 2022.....	41
Figura 14 - Fluxo das Etapas do Projeto	43
Figura 15 - Tela do Site da Receita Federal.....	45
Figura 16 - CMD: Processo de Concatenação das Segmento de Empresas	45
Figura 17 - CMD: Processo de Concatenação do Segmento de Estabelecimentos..	46
Figura 18 - Tela de download do Visual Studio	46
Figura 19 - Tela de Download do Integration Services.....	47
Figura 20 - Tela de Download Banco de Dados SQL Server	47
Figura 21 - Tela de Download do Management Studio	48
Figura 22 - Tela de Criação do Banco de Dados	48
Figura 23 - Tela de Criação da Tabela Estabelecimentos.....	49
Figura 24 - Tela de Criação da Tabela Empresas.....	49
Figura 25 - Tela de Criação da Tabela CNAE	49
Figura 26 - Tela de Criação da Tabela Natureza Jurídica.....	50
Figura 27 - Tela de Criação da Tabela Empresa Telecom.....	50
Figura 28 - Tela de Criação do Data Flow Task	51
Figura 29 - Tela de Processo de Extração de dados	51
Figura 30 - Tela de Download do Power BI.....	52

Figura 31 - Tela de Conexão do Power BI com o Banco de Dados	52
Figura 32 - Tela de Vinculação dos Bancos de Dados.....	53
Figura 33 - Menu Inicial.....	54
Figura 34 - Indicadores da Receita Federal	55
Figura 35 - Geolocalização da Receita Federal	56
Figura 36 - Dados da Empresa de Telecom.....	57
Figura 37 - Oportunidades de Mercado.....	58
Figura 38 - Informações Gerais.....	59
Figura 39 - Visualização de Oportunidade	60
Figura 40 - Tela de Prospecção de Clientes	61
Figura 41 - Tela de Tendência de Mercado.....	62

LISTA DE TABELAS

Tabela 1 - Tamanho das Bases de Dados	21
Tabela 2 - Principais Ferramentas de ETL	31
Tabela 3 - Valores das Ferramentas de BI	35
Tabela 4 - Portfólio de Produtos no Segmento de Dados	37
Tabela 5 - Requisitos do Cliente	38
Tabela 6 - Requisitos do Produto	43

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i> ou Inteligência de Negócio
ETL	<i>Extract, Transform, Load</i> ou Extração, Transformação e Carga
SQL	<i>Structured Query Language</i> ou Linguagem de Consulta Estruturada
CMD	<i>Prompt of Command</i> ou Prompt de Comando
CNAE	Classificação Nacional de Atividades Econômicas
PJ	Pessoa Jurídica
RF	Receita Federal
VPN	<i>Virtual Private Network</i> ou Rede privada virtual
IOT	<i>Internet of things</i> ou Internet das Coisas
WEB	<i>World Wide Web</i>
TIC	Tecnologias de Comunicação e Informação
DAS	<i>Direct Attached Storage</i>
NAS	<i>Network Attached Storage</i>
Open Source	Software Livre
TI	Tecnologia da Informação

1. INTRODUÇÃO

Atualmente, dentro de várias empresas e órgãos públicos, podem ser analisadas a existência de grandes volumes de dados que muitas vezes não conseguem ser analisados não obstante disso, um segundo ponto de vista em consideração tem vínculo à variedade de dados disponíveis. A vastidão de informações na internet, originadas pelas diferentes fontes de dados, sobrecarrega varios sistemas. Cabe registrar que apenas 1% destes dados é efetivamente analisado (BREITMAN, 2014), dados esses que são gerados por sistemas a cada momento. Com a facilidade de se trabalhar com grandes volumes de dados, empresas começam a tomar suas decisões baseando-se nas informações geradas pelos dados. Assim, pode ser realizada uma análise feita a partir da análise dos dados das instituições e que assim começar a obter um impacto direto na sua lucratividade não obstante disso, as empresas que podem transformar esses dados sobre seus clientes em informações em tempo real ganham uma vantagem competitiva substancial (Sivarajah, Kamal, Irani, & Weerakkody, 2017).

Existem hoje várias vantagens para uma empresa ter uma cultura de dados, pois cada vez mais as instituições estão interessadas em coletar dados importantes para o seu segmento de negócio, com isso, essas companhias utilizam o poder da informação como base de suas estratégias corporativas.

A partir desse grande movimento de utilização de dados, pode ser analisado que as empresas começaram a focar na estruturação de dados para tomada de suas decisões, utilizando ferramentas de *extract, transforme e load – ETL*, para ser possível uma organização de seus dados e um melhor tratamento possível. Esses processos de *ETL* são importantes para fazer a extração, transformação e carga dos dados, pois muitos dados minerados tanto da internet quanto de páginas governamentais e páginas não governamentais, páginas essas tem um grande volume de dados e erros de estruturação, erros esses que precisam ser corrigidos para realmente a empresa conseguir tomar decisões corretas e objetivas acerca dos dados.

Não obstante, empresas começaram a utilizar também ferramentas de visualização de dados para um possível visualização de informações de formas mais simples e objetivas, por meio de gráficos e painéis interativos. Hoje em dia, a maior

ferramenta do mundo de visualização de dados é o *PowerBI* que tem como objetivo proporcionar *insights* na gestão dos dados e facilidade de manuseio desses dados, logo muitas empresas estão utilizando essa ferramenta para possibilitar análises mais rápidas em tempo real.

Em síntese, várias empresas, em busca de uma facilidade de seus processos e análises em tempo real, precisam desenvolver *dashboards* para visualização de dados e que permitam decisões mais rápidas e objetivas por meio de suas informações.

1.1. CONTEXTUALIZAÇÃO

A empresa que foi desenvolvida um sistema da informação é uma Telecom fictícia que atua em todo o Brasil, e tem como objetivo expandir sua carteira de clientes, mas no caso essa empresa desconhece o real tamanho do mercado brasileiro e também não possuem conhecimento a respeito da localização dos maiores nichos de mercado no segmento de Pessoa Jurídica (PJ).

A partir de reuniões feitas com o cliente, foi possível elencar suas reais necessidades e objetivos quem giram em torno de aumentar sua carteira de clientes com foco no segmento PJ. Outro ponto analisado também é o qual o cliente quer ver os dados abertos por mapas georreferenciados, para ter a noção do tamanho de oportunidades que ele tem dentro do Brasil, que é segmentado por estados e assim realizar a filtragem de empresas que tem declarado seu capital social acima de 1 milhão de reais.

A Empresa em questão tem bastante dificuldade de trabalhar com dados abertos e grandes volumes de dados pois a empresa trabalhava apenas com dados vinculados aos seus bancos de dados próprios que giravam em todo de aproximadamente 5 milhões de linhas algo muito pequeno em comparação aos dados abertos da Receita Federal. Com isso, foi desenvolvido um trabalho focado na explicação da importância de se trabalhar com dados abertos e os seus benefícios para a empresa para geração de *inghts* e soluções reais para aumento de carteira e lucratividade.

A partir das solicitações do cliente que podem ser analisadas na Tabela 6, foi necessário estudar onde conseguir adquirir essas informações e onde pegar esses dados abertos de forma gratuita e objetiva.

Com isso, foi analisado que vários sites disponibilizam seus dados de forma aberta para serem baixados. Esses arquivos muitas vezes são baixados para o desenvolvimento de *insights* e análises de dados bem profundas, pois a partir desses *downloads* as instituições tomam suas decisões.

Deste modo, foi analisado que seria necessário baixar os dados da Receita Federal do Brasil, que tem como objetivo liberar os dados de todas as empresas do Brasil pelo o seu segmento e também o capital social, além de outras informações podem ser visualizadas na Tabela 6.

Em conclusão, este trabalho tem como objetivo baixar os dados do site da Receita Federal e desenvolver análises com foco em uma empresa fictícia de Telecom que atua em todo o cenário brasileiro.

Deste modo, para chegar no desenvolvimento dessas análises de dados, foram desenvolvidos processos de extração, transformação e carga de dados além da amplificação de um banco de dados juntamente com a criação de *dashboard*.

1.2. OBJETIVO GERAL

Desenvolver um sistema de informação em *Power BI* capaz de visualizar indicadores matemáticos a partir dos dados abertos da Receita Federal para uma Telecom fictícia.

1.3. OBJETIVOS ESPECÍFICOS

Serão desenvolvidos 3 objetivos específicos, que irão delimitar mais precisamente o escopo deste trabalho. São eles:

Objetivo Específico 1:

Levantar e identificar as necessidades de um cliente da área de Telecom e obter informações para tomada de decisão sobre o nicho de mercado.

Objetivo Específico 2:

Criar um processo Extração, Transformação e Carga (*Extract Transform Load - ETL* no *Integration Services* e um Banco de Dados no *SQL SERVER* com base nos arquivos disponíveis pela Receita Federal.

Objetivo Específico 3:

Criar o sistema da informação em *Business Analytics Tool* utilizando a ferramenta *Power BI* versão 2.97.921.0.

1.4. METODOLOGIA

Foi utilizado a metodologia CRISP-DM que é um modelo de processos hierárquizados e se vincula a um conjuntos de etapas. (CHAPMAN et al., 2000).

O ciclo de desenvolvimento desse projeto envolve todas as 6 fases do CRISP-DM. Neste trabalho a sequência das fases não é sempre a mesma pois há a necessidade de mudanças de uma fase para outra, a cada momento esta sendo feito reuniões com o cliente para melhor entendimento do problema.

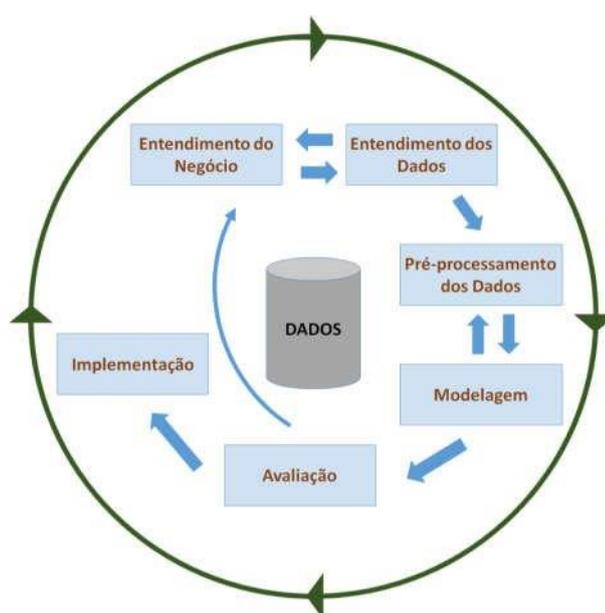


Figura 1 - CRISP-DM

Com o objetivo de fazer a caracterização da demanda do cliente, o desenvolvimento de um processo ETL e concepção de um banco de dados, além do desenvolvimento um *dashboard* utilizando *business Analytics Tools* o trabalho tende a seguir as seguintes etapas:

A primeira parte teve como propósito levantar o objetivo do cliente, que é de suma importância para o desenvolvimento deste projeto, pois sem ele não seria possível conseguir mirar nas reais demandas que o cliente queria solucionar, por meio dos dados exportados da Receita Federal.

A segunda parte consistiu no *download* dos dados da Receita Federal, que se encontra no próprio website da Receita Federal do Brasil, e que pode ser baixado no link:

“<https://www.gov.br/receitafederal/ptbr/assuntos/orientacaotributaria/cadastros/consultas/da-dos-publicos-cnpj>”. A partir desses *downloads*, foi desenvolvido um processo chamado Extração, Transformação e Carga (*Extract Transform Load - ETL*), que teve como objetivo primordial obter os dados baixados da Receita Federal e colocá-los dentro do Banco de Dados *SQL Server 18.9.2*.

A terceira parte consistiu no desenvolvimento de um *dashboard*, que teve como objetivo mostrar de forma intuitiva os dados e as análises desenvolvidas para solucionar as questões da empresa de Telecom.

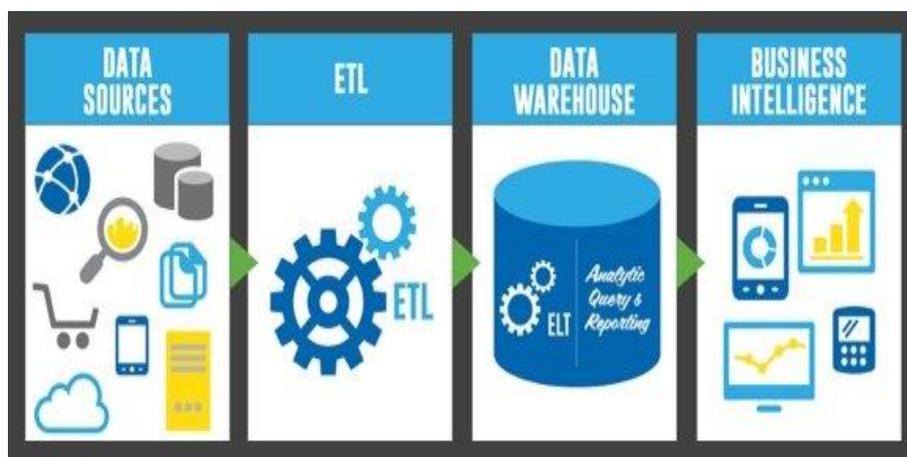


Figura 2 - Processo de Extract Transform Load - ETL

Para alcançar o objetivo específico 1, foi necessário:

Foi feita uma reunião com o cliente, com o objetivo de levantar todos os questionamentos que o cliente possui em relação ao seu negócio, com isso foram feitas várias perguntas para saber qual seria o escopo do projeto. Todas as perguntas podem ser vistas na Tabela 6.

A partir das perguntas levantadas, foi possível abrir um panorama das reais dificuldades do cliente, bem como propor a resolução dos problemas.

Para alcançar o objetivo específico 2, foi necessário:

Procurar alguma fonte de dados que poderia levantar as informações solicitadas, e com isso, foi encontrada uma base de dados aberta no site da Receita Federal, que poderia ser usada para solucionar esse problema do cliente e resolver todos as questões do cliente para esse escopo.

Com isso, foi necessário fazer toda a extração da base de dados da Receita Federal que se encontra no *link* para *download* no site da Receita Federal do Brasil: <https://www.gov.br/receitafederal/ptbr/assuntos/orientacaotributaria/cadastros/consultas/da-dos-publicos-cnpj> que pode ser visto também na figura 15.

A partir da base baixada foram realizados os processos de unificação de bases, pelo o próprio *CMD* do Windows, que possibilitou colocar todos os dados de uma só vez no processo de *ETL*, que pode ser observado também nas figuras 16 e 17.

Para levantar a melhor ferramenta de mercado foi utilizado o gráfico de *Gartner*, conhecido também como o Gráfico do Quadrante Mágico, que tem como objetivo auxiliar em quais são as melhores ferramentas dentro do seu segmento. Desta forma, consegue-se analisar várias ferramentas líderes e foi decidido utilizar a ferramenta *Integration Services* por ser uma ferramenta de fácil *download* e excelente para desenvolver processos de *ETL*.



Source: Gartner (August 2021)

Figura 3 - Gráfico do Quadrante Mágico ETL

O *Integration Services*, que atualmente é muito utilizado para realizar extração, transformação e carga de grandes volumes de dados, foi utilizado nesse trabalho para desenvolver todo o processo de *ETL* e colocar os dados dentro do Banco de Dados.

Em seguida, foi necessário utilizar o Gráfico do Quadrante Mágico novamente, para escolher qual banco de dados utilizar, pois a ferramenta de *ETL* não possibilita a visualização exata dos dados.



Figura 4 - Gráfico do Quadrante Mágico Banco de Dados

Como foi utilizado para fazer o processo de *ETL* uma ferramenta da *Microsoft*, foi decidido continuar utilizando ferramentas dessa própria empresa, sendo que essa empresa é uma ferramenta líder de mercado. Deste modo então, foi escolhido utilizar o *SQL Service*.

Desta forma, foi possível colocar os dados que se encontram no *Integration Services* em um Banco de Dados, que teve como objetivo nos possibilitar a visualização completa dos dados, além de possibilitar desenvolver consultas e análise dentro da própria ferramenta.

Para alcançar o objetivo específico 3, foi necessário:

Para desenvolver o *dashboard* do cliente foi necessário realizar uma análise de mercado de qual ferramenta utilizar. Para levantar a melhor ferramenta de mercado

2. JUSTIFICATIVA E LIMITAÇÕES DA PESQUISA

Este trabalho se justifica pela demanda solicitada de uma empresa de Telecom fictícia que tem como objetivo aumentar sua carteira de clientes. Contudo, para desenvolver esse projeto foi necessário criar todo um sistema da informação com etapas de *ETL*, criação de banco de dados e desenvolvimento de um *dashboard* para visualização das informações solicitadas pelo o cliente que podem ser vista na Tabela 6.

Pode ser justificado também o motivo de utilizar os dados abertos da Receita Federal para gerar as informações desejadas pelo o cliente. Segundo o site Dados.gov, “dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura.”

Também citado no site Dados.gov, existem hoje cinco motivos para a abertura dos dados, os quais são:

- 1 - Transparência na gestão pública;
- 2 - Contribuição da sociedade com serviços inovadores ao cidadão;
- 3 - Aprimoramento na qualidade dos dados governamentais;
- 4 - Viabilização de novos negócios;
- 5 - Obrigatoriedade por lei.

Este trabalho tem como objetivo gerar *insights* com base nas análises de dados que foram feitas a partir da coleta de dados abertos, que muitas vezes não são utilizados pelas empresas por falta de informação, ou por falta de conhecimento técnico para manusear os dados para alcançar o objetivo de visualização de dados.

A grande limitação dessa pesquisa se vincula ao fato da dificuldade de trabalhar em um computador pessoal com um grande volume de dados, que, segundo Boyd e Crawford (2011) citado por Pimenta (2013), nada mais é do que um grande volume de dados, mas mais do que seu volume, sua articulação em rede, sua velocidade e diversidade possibilitam a produção de mais dados, a partir dos dados já existentes,

sobre indivíduos, grupos ou sobre a própria informação, quaisquer que sejam ela, disponível.

Não obstante, ainda nas limitações do projeto de graduação em questão, temos todo o desenvolvimento do projeto, a exportação de dados e o armazenamento desses dados em uma máquina pessoal que pertence ao autor desse projeto: Túlio da Costa Tomé, estudante de Engenharia de Produção da Universidade de Brasília. Este computador em questão tem 8 gigas de memória, processador Intel 7 e um pequeno SSD de 240 gigas.

Logo, pode-se concluir que foi necessário fazer um estudo de quais dados realmente precisam ser baixados e quais processos realmente precisam ser feitos, para não sobrecarregar a máquina em questão, pois fazer bastante *downloads* de base de dados acaba deixando o computador muito pesado e sem força para processar grandes volumes de dados, e isso gera o questionamento tendo como premissa a pergunta que Valdyanathan (2011) fez:

Estaremos nos afogando em dados, incapazes de distinguir o que é bom do que é ruim e de diferenciar o verdadeiro do falso? (...) Que ferramentas nos ajudam a lidar com essa abundância? (VAIDHYANATHAN, 2011, p. 191)

Analisando o projeto em questão, pode ser observado que ele possui um grande volume de dados, que acaba sendo um grande problema, pois isso acarreta em perda de performance no banco de dados, o que gera uma dificuldade para fazer o processo da cópia dos dados que foram baixados da Receita Federal em CSV para o banco de dados em questão.

Para maior entendimento, foi desenvolvido na (Tabela 1), uma melhor visualização do tamanho das bases de dados.

Tabela 1 - Tamanho das Bases de Dados

Arquivos CSV da Receita Federal	Tamanho
Estabelecimentos	10,4 GB (11.227.792.031 bytes)
Empresas	3,35 GB (3.604.951.630 bytes)

CNAE	86,1 KB (88.215 bytes)
------	------------------------

Desta forma, pode-se visualizar de forma bastante clara, que as bases de dados da Receita Federal são bastante pesadas e difíceis para manter em um banco de dados e vincular essas informações às bases de dados da empresa e seus sistemas.

3. REVISÃO DE LITERATURA

A revisão de literatura é uma parte vital do processo de investigação de trabalhos desenvolvidos, artigos e projetos. Como nos informam Cardoso et al (2010) “cada investigador analisa minuciosamente os trabalhos dos investigadores”. Deste modo, a seguir, estão elencadas as fontes para este estudo.

3.1. DADOS E INFORMAÇÃO

Definições acerca de dados e informações:

Atualmente, é observado que os dados são um conjunto de caracteres, observações documentadas ou resultados de alguma medição, logo é possível definir o termo dado como uma sequência de símbolos quantificados ou quantificáveis. (SEMIDÃO, 2014)

Outro termo bastante em alta quando o nome dados ou informação é citado é o conceito: “*big data* ou grandes volumes de dados que dá início a uma verdadeira corrida, na qual, quem dispor de pessoal habilitado e tiver os instrumentos para decifrar mais precisamente as informações por trás de uma miríada de dados, ficará com o ouro.” (GLIKAS, 2018). Essa afirmação gerou uma grande força porque praticamente tudo o que é feito hoje na internet formam dados para alguma empresa tomar decisões.

Contudo, os dados são somente a matéria-prima da informação, eles não dizem nada sobre uma causa ou efeito, pois dados são um conjunto de registros qualitativos ou quantitativos conhecidos que organizados, agrupados, categorizados e padronizados adequadamente transforma-se em informação”. (MIRANDA, 1999). Para eles se tornarem realmente valiosos precisamos tratá-los e fazer esses dados se tornarem informação. Assim, segundo Setzer não é possível processar informação diretamente em um computador. Para isso é necessário reduzi-la a dados.” (SETZER, 2015). Esse aprimoramento de dados é importante, pois a partir disso pode ser feita a limpeza das bases de dados e focar nas informações pertinentes para descobrir a resposta para uma eventual pergunta.

Não obstante, é importante também definir o que é informação, que de acordo com Yves François Le Coadic, autor do livro “A Ciência da Informação”, “A informação é um registro, em suporte físico ou intangível, disponível à assimilação crítica para produção de conhecimento.” (LE COADIC, 1996).

Dessa forma, podemos inferir que a informação nada mais é do que os dados tratados e analisados para produzir um eventual conhecimento desejado.

A partir da definição de dados e informação é possível analisar um contexto bastante utilizado também na análise de dados que é o conceito de mineração de dados.

Mineração de Dados é um ramo da computação que teve início nos anos 80, quando os profissionais das empresas e organizações começaram a se preocupar com os grandes volumes de dados informáticos estocados e inutilizados dentro da empresa. Nesta época, Data Mining consistia essencialmente em extrair informação de gigantescas bases de dados da maneira mais automatizada possível. Atualmente, Data Mining consiste sobretudo na análise dos dados após a extração. (AMO, 2003).

A partir do conhecimento citado acima, podemos traçar a ideia de mineração de dados, que se iniciou nos anos 80 e teve como início o grande volume de dados encontrados dentro das organizações e que não eram utilizados para o levantamento de análises para os seus negócios. Desta forma, podemos definir mineração de dados como sendo a técnica de extrair grandes volumes de dados de fontes de dados que se encontra em um mesmo local ou locais diferentes. (AMO, 2003).

Aplicações do uso de dados e informações:

Atualmente as dados e a informação são usados em praticamente todas as empresas em seus mais variados segmentos. Fora do Brasil, a pesquisa em bases de dados é parte integrante em qualquer busca de informações para negócios. Com o uso da internet e o aumento do seu acesso, possibilitou as empresas pegarem informações de seus clientes em vários segmentos como compras online, cadastros de clientes em sites e outros.

Não obstante, várias empresas começaram a inserir seus dados em banco de dados e em sistemas, o que possibilitou a elas a redução de papeladas de documentos físicos que as empresas utilizavam.

Muitas das bases de dados de informação para negócios ou de notícias são atualizadas diariamente ou, mesmo, a cada minuto. Hoje em dia, alguns tipos de informação podem até mesmo estar disponíveis apenas em forma eletrônica. Soma-se a essas vantagens o maior poder de recuperação de informação em uma busca informatizada. (Cendón, 2014).

Deste modo, podem ser analisadas as facilidades que os dados trazem nos dias de hoje, além redução de papel que ajuda bastante na sustentabilidade das empresas. Nos últimos anos, com a utilização da tecnologia e comunicação, o papel físico enfrenta a substituição de algumas de suas utilizações por meios eletrônicos e digitais de comunicação. (MORAES, 2011).

Tecnologias atualmente usadas para dados e informações:

As maiores tecnologias que armazenam dados são os *databases* ou banco de dados. Essas tecnologias são abertas em duas categorias de banco de dados: os bancos de dados relacionais e os bancos de dados não relacionais.

O banco de dados relacional possibilita a criação de um relacionamento lógico entre as informações por meio de tabelas que se conectam por chaves primárias e secundárias, isso evita a necessidade da repetição de informações e agiliza as consultas feitas às fontes de dados.

Segundo o *DB-Engines Ranking* que tem como objetivo classificar os sistemas de gerenciamento de banco de dados de acordo com sua popularidade, classificou as principais tecnologias no segmento banco de dados relacional. Segue a lista abaixo.

Rank			DBMS	Database Model	Score		
Feb 2022	Jan 2022	Feb 2021			Feb 2022	Jan 2022	Feb 2021
1.	1.	1.	Oracle +	Relational, Multi-model	1256.83	-10.05	-59.84
2.	2.	2.	MySQL +	Relational, Multi-model	1214.68	+8.63	-28.69
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	949.05	+4.24	-73.88
4.	4.	4.	PostgreSQL +	Relational, Multi-model	609.38	+2.83	+58.42
5.	5.	5.	IBM Db2	Relational, Multi-model	162.88	-1.32	+5.26
6.	6.	↑7.	Microsoft Access	Relational	131.26	+2.31	+17.09
7.	7.	↓6.	SQLite +	Relational	128.37	+0.94	+5.20
8.	8.	8.	MariaDB +	Relational, Multi-model	107.11	+0.69	+13.22
9.	9.	↑10.	Microsoft Azure SQL Database	Relational, Multi-model	84.95	-1.37	+13.67
10.	↑11.	↑22.	Snowflake +	Relational	83.18	+6.36	+64.96

Figura 6 - DBEngines, 2022

Os Bancos de Dados não relacionais ou banco de dados *NoSQL* (*Not Only SQL*) tem o objetivo de resolver a demanda de armazenagem de dados, apresentando uma abordagem diferente da persistência de dados.

Os bancos de dados *NoSQL* são classificados em bases de esquema Chave/Valor (*Key/Value Store*), Bancos de dados orientados à documentos, Bancos de dados de Colunas de chave valor e Bancos de dados de Grafos. (Lakshman, 2008) citado por Samuel Oliveira, (2014). Segue a lista abaixo segundo o site *Engines Ranking* dos principais banco de dados *NoSQL*.

Rank			DBMS	Database Model	Score		
Feb 2022	Jan 2022	Feb 2021			Feb 2022	Jan 2022	Feb 2021
1.	1.	1.	MongoDB +	Document, Multi-model	488.64	+0.07	+29.69
2.	2.	2.	Amazon DynamoDB +	Multi-model	80.36	+0.50	+11.21
3.	3.	3.	Microsoft Azure Cosmos DB +	Multi-model	39.95	-0.09	+8.29
4.	4.	4.	Couchbase +	Document, Multi-model	30.07	+1.21	-0.59
5.	5.	5.	Firebase Realtime Database	Document	19.15	-0.21	+3.15
6.	6.	6.	CouchDB	Document, Multi-model	17.45	+0.81	+1.84
7.	↑8.	7.	MarkLogic	Multi-model	9.46	+0.27	+0.00
8.	↓7.	8.	Realm +	Document	9.42	-0.17	+0.27
9.	9.	9.	Google Cloud Firestore	Document	9.06	+0.14	+1.01
10.	↑12.	↑11.	ArangoDB +	Multi-model	5.40	+0.67	+0.33

Figura 7 - DBEngines, 2022

Rank			DBMS	Database Model	Score		
Feb 2022	Jan 2022	Feb 2021			Feb 2022	Jan 2022	Feb 2021
1.	1.	1.	Redis +	Key-value, Multi-model ⓘ	175.80	-2.18	+23.23
2.	2.	2.	Amazon DynamoDB +	Multi-model ⓘ	80.36	+0.50	+11.21
3.	3.	3.	Microsoft Azure Cosmos DB +	Multi-model ⓘ	39.95	-0.09	+8.29
4.	4.	4.	Memcached	Key-value	25.77	+0.43	+0.10
5.	5.	5.	etcd	Key-value	11.71	+0.16	+3.07
6.	6.	6.	Hazelcast	Key-value, Multi-model ⓘ	9.50	-0.03	+1.08
7.	7.	↑ 11.	Riak KV	Key-value	6.92	+0.83	+1.93
8.	↑ 9.	↑ 12.	Ignite	Multi-model ⓘ	6.50	+0.48	+1.77
9.	↓ 8.	↓ 7.	Ehcache	Key-value	6.38	+0.29	-0.50
10.	10.	↓ 8.	Aerospike +	Key-value, Multi-model ⓘ	5.61	-0.33	-0.06

Figura 8 - DBEngines, 2022

Rank			DBMS	Database Model	Score		
Feb 2022	Jan 2022	Feb 2021			Feb 2022	Jan 2022	Feb 2021
1.	1.	1.	Neo4j +	Graph	58.25	+0.21	+6.08
2.	2.	2.	Microsoft Azure Cosmos DB +	Multi-model ⓘ	39.95	-0.09	+8.29
3.	↑ 4.	↑ 4.	ArangoDB +	Multi-model ⓘ	5.40	+0.67	+0.33
4.	↓ 3.	↑ 6.	Virtuoso +	Multi-model ⓘ	5.39	+0.02	+3.02
5.	5.	↓ 3.	OrientDB	Multi-model ⓘ	5.03	+0.47	-0.10
6.	↑ 7.	↑ 8.	Amazon Neptune	Multi-model ⓘ	2.99	+0.36	+0.92
7.	↓ 6.	7.	GraphDB +	Multi-model ⓘ	2.93	+0.07	+0.79
8.	8.	↓ 5.	JanusGraph	Graph	2.36	-0.03	-0.17
9.	9.	↑ 12.	TigerGraph +	Graph	2.24	+0.22	+0.91
10.	10.	10.	Stardog +	Multi-model ⓘ	1.98	+0.09	+0.52

Figura 9 - DBEngines, 2022

3.2. DADOS ABERTOS

Definição geral de dados abertos dentro e fora do Brasil:

Os dados abertos giram em torno do conceito de qualquer pessoa poder acessar e baixar informações relativas a uma base de dados governamental ou não.

Segundo o maior site do Brasil de dados abertos, Dados.gov, a definição de dados abertos é quando qualquer indivíduo pode facilmente acessar, utilizar, modificar e compartilhar para qualquer fim suas informações.

A Lei de Acesso à Informação é a obrigatoriedade de divulgação por meio da Internet e por meio de formatos abertos e legíveis por máquinas. Destaca, também, que isso permite que desenvolvedores possam validar a consistência dos dados e publicá-los de forma mais criativa e acessível à sociedade por meio de aplicativos. Zaidan (2012) apud Dutra; Lopes (2013).

A abertura de dados abertos surge em grande escala pelos órgãos públicos, pois em 2011 ocorreu a declaração do governo aberto, que tem como objetivo reconhecer que as pessoas de todo o mundo possam exigir mais transparências de seus governos. (DUTRA; LOPES, 2013)

Deste modo, segundo Araújo e Souza (2011) “Atualmente, muitos dados governamentais estão disponíveis na Web”, assim, pode ser verificado que várias instituições e órgãos governamentais disponibilizam seus dados para análise porque é importante uma democratização da informação e transparência na gestão governamental.

Alguns sites disponibilizam seus dados e suas informações de forma legal para estudos e análises de mercado, um desses sites é o site da Receita Federal (<https://www.gov.br/receitafederal/pt-br>) que é bastante utilizado para levantar as análises referentes às empresas de todo o Brasil e será utilizado neste trabalho.

Aplicações mais usuais dos dados abertos:

Os dados abertos podem ser aplicados para qualquer fim que o indivíduo ou as corporações tenham em mente, focando basicamente nas seguintes aplicações a seguir:

- **Transparência:** promover melhor acesso aos dados.
- **Análise de Mercado:** promover *insights* e galgar novos mercados.
- **Análises governamentais:** analisar gastos do governo e buscar mais transparência dos valores investidos ou gastos no país, além de dados dos nossos políticos.
- **Reputação:** verificar a reputação das empresas ou do governo.
- **Reclamação:** analisar reclamações feitas para empresas ou governos.

Deste modo, com as aplicações de técnicas é possível obter novas ideias e informações úteis além de melhores formas de visualizar os dados públicos. (MOREIRA, 2020).

Tecnologias mais usadas para acessar e fazer uso de dados abertos:

Existem atualmente duas formas de armazenar os dados sendo elas chamadas de *Direct Attached Storage (DAS)* ou armazenamento ligado diretamente e o *Network Attached Storage (NAS)*.

O *Direct Attached Storage* inclui todos os armazenamentos físicos entre os dados e o *hardware*, sendo eles: discos rígidos (HD), *pendrives* e as unidades de estado sólido (SSD'S). Já o *Attached Storage* permite que vários computadores tenham acesso as informações remotamente.

Contudo, existem também as principais ferramentas para guardar os dados abertos, que são os bancos de dados relacionais e não relacionais. Não obstante, podem ser utilizadas também ferramentas como csv, excel e o próprio txt.

3.3. ETL – (EXTRACT, TRANSFORM, LOAD)

Definições da extração, transformação e carga dos dados:

O processo de extração, transformação e carga (*ETL*) é um processo que tem como objetivo extrair dados de um banco de dados ou de algum arquivo existente que contém alguma informação, e jogar esses dados em um banco de dados muitas vezes chamado de *data warehouse*. (Ferreira et al, 2010).

O processo de *ETL* é muito utilizado para pegar dados de várias fontes de dados distintas ou até iguais, onde necessitam fazer alguma tratativa ou ser alocadas para alguma outra fonte de dados.

Dentro de um processo de *ETL* o principal objetivo é capturar as fontes de dados primárias e transformar esses dados nos seus próprios objetivos, fazendo

cálculos, excluindo informações e acrescentando ideias, além de trazer a carga das informações em um banco de dados, de preferência dentro de um *data warehouse*. (AGUENA, 2017).

Aplicações do ETL:

O *ETL* teve fama nos anos 70, quando as empresas começaram a ter vários bancos de dados para armazenar diferentes tipos de informações de negócios. Com isso, surgiu a necessidade de relacionar os dados de vários bancos de dados diferentes. Deste modo, o *ETL* virou o método mais usado de coleta dados, transformação e carga.

Logo podem ser citadas as seguintes aplicações para o *ETL*, segundo Rodrigues (2020):

- Junção ou consolidação de várias fontes de dados.
- Transformação e desenvolvimento de cálculos matemáticos.
- Carregamento de bases de dados.

Tecnologias mais usadas de ETL no mundo:

Atualmente existem várias tecnologias de *ETL* com suas mais variadas possibilidades, qualidades e pontos de melhoria. Segundo (FERREIRA, 2010) é muito importante tomar cuidado na decisão de qual ferramenta utilizar pois dificilmente ela poderá ser mudada ao longo do tempo.

Na tabela 2 denominada como Gráfico do Quadrante Mágico *ETL* que foi tirada do gráfico de *Gartner*, presente no site Time Smart, em 2021 é possível analisar as ferramentas mais utilizadas no momento e acrescentar o *Pentaho* que é uma ferramenta de *software* livre *open source*.

Tabela 2 - Principais Ferramentas de ETL

Revendedoras	Ferramentas
Informática	<i>PowerCenter</i>
<i>IBM</i>	<i>IBM Data Stage</i>
<i>SAP</i>	<i>Data Integrator & Data Services</i>
<i>Oracle</i>	<i>Oracle Data Integrator</i>
<i>Talent</i>	<i>Talent ETL</i>
<i>Microsoft</i>	<i>Microsft Integration Services</i>
<i>Denodo</i>	<i>Data Visualization and ETL</i>
<i>Hitachivantara</i>	<i>Pentaho Data Integrato</i>

3.4. SISTEMAS DE INFORMAÇÃO

Definição base do sistema da informação:

Sistema de informação (SI) é o conjunto de desafios gerenciais, tecnologias da informação, conceitos fundamentais, processos de desenvolvimento e aplicações de negócios. (O'BRIEN; MARAKAS, 2012).

Em outras palavras, um sistema de informação (SI), pode ser definido como um amontoado de componentes relacionados que coletam, processam, armazenam e distribuem informação com o objetivo de ajudar na tomada de decisão. (LAUDON, 2004).

Pode ser analisado que com o avanço da tecnologia as corporações se transformaram em instituições mais complexas, hierarquizadas, além de especializadas, que demandam supervisão e gerência a todo momento. Deste modo, a inquietação das empresas passou a ser com autoridade, responsabilidade, planejamento, controle, coordenação e relações no trabalho (MOTTA, 1998).

As tecnologias da informação, incluindo sistemas de informação com base na internet, têm hoje papel vital e crescente na administração. A tecnologia da informação é capaz de auxiliar todos os tipos de negócios a aprimorar a produtividade e a eficácia de seus processos administrativos, a tomada de decisão gerencial e a colaboração de grupos de trabalho. (O'Brien, 2013).

Uma definição final que expressa muito bem o que seria o papel de um sistema da informação é a de entrada de dados, processamento dos dados e de saída de dados. (LAUDON, 2004).

Aplicações do uso dos sistemas da informação:

As aplicações dos sistemas de informação tem como objetivo investigar, analisar, projetar, implementar e manter suas informações(O'BRIEN; MARAKAS, 2012).

O SI dentro de uma instituição está a cada momento capturando dados, processando e transformando esses dados em informações importantes que ajudarão a empresa a tomar decisões. (MARCONDES, 2020).

As principais aplicações de um sistema da informação dentro de uma empresa seriam: o aumento de eficiência operacional da empresa, processamento de transações básicas, coleta e transmissão de informações gerenciais, monitoramento e registro do desempenho dos empregados e setores, manutenção do registro de status e mudanças nas funções centrais do negócio da empresa. (GURBAXANI, WHANG, 1991).

De acordo com O'Brien e Marakas (2012) os sistemas de informação podem ser divididos em quatro tipos:

- Sistemas de informações transacionais.
- Sistemas de informações gerenciais.
- Sistemas de apoio a decisão.
- Sistemas de informações executivas.

Os sistemas da informação tem como objetivo primordial nas suas aplicações a entrega da possibilidade das decisões tomadas terem grande assertividade. A pessoa que tem a decisão final precisa ter a compreensão das consequências ou resultados do seu parecer, deste modo, deverá escolher a melhor entre as alternativas propostas. (STEPHEN, COULTER, 1996).

Tecnologias mais usadas de S.I:

Tanto quanto existem diferentes tipos de aplicações para sistemas da informação, existem também diferentes tipos de tecnologias relacionadas a SI. (Mesquita, 2018).

Os tipos de tecnologias de sistemas da informação são:

- ERP: *Software* de planejamento de recursos.
- CRM: Gestão de relacionamento com os clientes.
- SCM: Gestão de cadeia de fornecimento.
- SIG: Sistema de informação gerencial.
- Relatórios programados ou de exceção: Relatórios com informações específicas muitas vezes vem em formato txt, csv ou xlsx.
- Planilhas: muitas vezes feitas em formato csv ou xlsx.

O mais importante na parte da escolha da tecnologia é que o sistema que for escolhido deverá oferecer a possibilidade da companhia tomar decisões sob condições de incertezas, pois cada SI citado acima tem como objetivo um devido fim e aplicação. (BAZZOTTI, GARCIA, 2000).

3.5. BUSINESS INTELLIGENCE (BI)

Definições sobre o business intelligence:

O termo *Business Intelligence (BI)*, que foi citado inicialmente pela empresa *Gartner Group*, foi vinculado ao método de alcance, organização de informações, análise, compartilhamento e monitoramento de informações. (GARCIA, 2022).

Os sistemas de BI retém dados de diferentes fontes de dados e ofertam uma apresentação de conhecimentos complexos que podem ajudar as empresas a tomarem as decisões corretas. (NEGASH, 2004).

Esses sistemas, ganharam espaço no mercado com o advento da era da informação, em meados da década de 90, onde o diferencial competitivo passou a ser a capacidade analítica da empresa (Barbieri, 2011).

Portanto, é observado que cada detalhe e análise de dados tem grande importância no aumento do lucro ou prejuízo de uma empresa. Tomar decisões com base na emoção atualmente é praticamente banalizada em relação as ferramentas de BI que tem como objetivo mostrar os dados muitas vezes em tempo real para a empresa tomar a decisão.

Aplicações do uso de B.I:

As três principais aplicações do BI são “consultar, relatar e analisar” (TURBAN & VOLONIMO, 2013, p. 330).

Atualmente as principais aplicações de BI são encontradas nas áreas de finanças, marketing e atendimento ao cliente, maciçamente concentradas em bancos, seguradoras e empresas de comércio. (DETONI JUNIOR, 2018).

Dentre os vários proveitos do avanço da tecnologia na TI nas últimas décadas, podemos elencar as aplicações de superação de barreiras espaciais e temporais que anteriormente dificultavam a comunicação. (GOMES, 2011). Hoje, com o BI, todas essas aplicações de análises temporais podem ser realizadas para as tomadas de decisões.

Tecnologias mais usadas de B.I:

Segundo o site *knowsolution*, podemos analisar as principais ferramentas de *Business Intelligence*, que tem como objetivo auxiliar o desenvolvedor de BI a criar *dashboards* ou painéis interativos para os colaboradores da companhia em questão, e analisar suas informações em tempo real e levar a tomada de decisão.

Nem sempre é fácil escolher uma ferramenta de BI para atender às necessidades de sua empresa, segundo (SAYER, 2018).

Analisando o mercado de ferramentas de BI é possível entender porque nem sempre é fácil escolher uma ferramenta de BI, pois cada ferramenta mesmo tendo as mesmas definições e características possuem valores de compra totalmente diferentes.

Segue abaixo uma tabela onde foram analisado as 3 ferramentas líderes de mercado. Essa análise vai confirmar a citação do parágrafo acima.

Tabela 3 - Valores das Ferramentas de BI

Ferramentas	Valores de Mercado
<i>Power BI PRO</i>	57,10 Reais
<i>Tableau Creator</i>	339,49 Reais
<i>Qlik Sense Business</i>	141,48 Reais

Fica entendido que essas 3 ferramentas são bem parecidas e entregam os mesmos resultados em análise de dados. Porém essas ferramentas tem visuais diferentes e possibilidades de apresentação de dados totalmente diferentes umas das outras.

Em síntese, podemos entender que praticamente todas as ferramentas líderes de mercado no segmento de BI entregam os mesmo resultados no que diz respeito a análise dados, mas é importante analisar seus valores e as funções de visualização de dados.

4. APRESENTAÇÃO DA EMPRESA DE TELECOMUNICAÇÃO

4.1. APRESENTAÇÃO DA TELECOM

A empresa em questão é uma Telecom fictícia, que atua em todos os estados brasileiros e possui uma carteira no segmento de pessoas jurídicas (PJ) de 49.999 clientes e um faturamento mensal de R\$ 19.954.502 com um ticket médio de 399 reais.

Esta Telecom possui dois principais combos de produtos, sendo eles o combo avançado de dados e telefonia e o combo de dados básicos e telefonia, tendo um faturamento mensal de R\$ 15.167.110 e R\$ 4.786.952, respectivamente.

A empresa foi fundada em 2010 e ela possui apenas 21 anos desde sua criação, seus valores mais profundos são o de pessoas ajudando pessoas, o que diz muito sobre a cultura da empresa de sempre estar dando bastante atenção para seus clientes e suporte para os seus funcionários.

4.2. PRINCIPAIS CLIENTES DA TELECOM E PRODUTOS DA TELECOM

Os principais clientes dessa empresa de telecomunicação são as empresas PJ que precisam de dados de telefonia. A empresa quer focar em empresas que tem um faturamento acima de 1 milhão de reais por mês e que esteja disposta a investir em dados ou na sua área comercial com *softwares* ou *IOT* que possibilitem fortalecer sua área comercial a partir da compra de sistemas de telefonia ou *softwares* de segurança.

A empresa possui vários serviços e produtos. Segue a abaixo os seguintes produtos na Tabela 4.

Tabela 4 - Portfólio de Produtos no Segmento de Dados

Produtos	Descrição
VPN	<p>O VPN permite a criação de uma rede privativa entre matriz e filiais, capaz de possibilitar o tráfego de aplicações com diferentes requisitos fazendo o uso de QoS.</p> <p>O recurso de QoS (<i>Quality of Service</i>) permite que a rede classifique, diferencie, priorize e encaminhe, de forma eficiente, o tráfego, com garantia dos recursos de rede para atendimento das necessidades de aplicações mais sensíveis à latência, congestionamento e perda de pacotes e que exigem elevado <i>uptime</i> de rede, integrando todas unidades de negócio do cliente.</p>
LAN TO LAN	<p>LAN to LAN tem como objetivo interconectar unidades de negócio do cliente através de um circuito de comunicação <i>Ethernet</i> (rede estatística) baseado em transporte de dados, utilizando camada (L2), com topologias ponto a ponto ou vários ponto a ponto, com base no padrão <i>Metro Ethernet Fórum (MEF)</i>.</p>
WAVELENGTH	<p><i>Wavelength</i> é uma solução que conecta unidades de negócio do cliente, através de circuito determinístico ponta a ponta de forma dedicada e transparente a protocolos de camada (L2). Além disso, esse serviço possui em suas características, alta escalabilidade, qualidade de serviço e baixa latência.</p>
IOT	<p><i>SIMCARD</i> com Atuação em 5 operadoras Algar, Vivo, TIM, Oi e Claro.</p>
Celular	<p>Telefone celular é um aparelho de comunicação por ondas eletromagnéticas que permite a transmissão bidirecional de voz.</p>
Telefone Fixo	<p>Telefone fixo é um sistema de telecomunicação capaz de transmitir som através de sinais elétricos nas redes de telefone.</p>

4.3. INVESTIGAÇÃO DO PROBLEMA QUE A TELECOM QUE SOLUCIONAR

A investigação foi feita para entender o objetivo da empresa em querer desenvolver um sistema de informação focado no levantamento de indicadores, para isso, foram desenvolvidas perguntas para o cliente, onde conseguimos levantar sua principal indagação, que seria aumentar sua carteira de clientes PJ que possuem capital social acima de 1 milhão de reais em todos os estados brasileiros Tabela 5.

Na investigação do problema não constam as perguntas feitas para a empresa para tentar descobrir quais problemas precisamos resolver com o *dashboard* que será feito. Nessa etapa constam apenas as perguntas feitas de forma geral para entender de forma objetiva o que o cliente quer saber.

Tabela 5 - Requisitos do Cliente

Perguntas	Sim	Não
O sistema da informação precisa ser de Fácil Manipulação	X	
É necessário ter Disponibilidade de outras pessoas visualizarem o sistema da informação?	X	
É preciso ter a existencia de filtros	X	
O sistema da informação precisa ser rapido?		X
O sistema da informação precisa estar atualizado a cada 3 meses?	X	

A partir do levantamento das perguntas feitas e dos questionamentos que o cliente possui, foi analisado que para solucionar esses problemas seria necessário baixar os dados abertos do site da Receita Federal, no qual pode ser encontrado todos os dados relativos as empresas do Brasil.

4.4. ANÁLISE DE SOLUÇÃO DO PROBLEMA

Para desenvolver esta análise foram baixados todos os dados da Receita Federal referentes as empresas e aos estabelecimentos, além dos dados de CNAE e natureza jurídica como pode ser observado na figura 15.

Com isso, os dados passaram por um processo *ETL* que possibilitou a visualização de todos os campos em um banco de dados. As informações contidas nas bases de dados são:

EMPRESAS	
Campo	Descrição
CNPJ BÁSICO	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ).
RAZÃO SOCIAL / NOME EMPRESARIAL	NOME EMPRESARIAL DA PESSOA JURÍDICA
NATUREZA JURÍDICA	CÓDIGO DA NATUREZA JURÍDICA
QUALIFICAÇÃO DO RESPONSÁVEL	QUALIFICAÇÃO DA PESSOA FÍSICA RESPONSÁVEL PELA EMPRESA
CAPITAL SOCIAL DA EMPRESA	CAPITAL SOCIAL DA EMPRESA
PORTE DA EMPRESA	CÓDIGO DO PORTE DA EMPRESA: 1 - NÃO INFORMADO 2 - MICRO EMPRESA 03 - EMPRESA DE PEQUENO PORTE 05 - DEMAIS
ENTE FEDERATIVO RESPONSÁVEL	O ENTE FEDERATIVO RESPONSÁVEL É PREENCHIDO PARA OS CASOS DE ÓRGÃOS E ENTIDADES DO GRUPO DE NATUREZA JURÍDICA 1XXX. PARA AS DEMAIS NATUREZAS, ESTE ATRIBUTO FICA EM BRANCO.

Figura 10 - Receita Federal, 2022

ESTABELECEMENTOS	
Campo	Descrição
CNPJ BÁSICO	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ).
CNPJ ORDEM	NÚMERO DO ESTABELECEMENTO DE INSCRIÇÃO NO CNPJ (DO NONO ATÉ O DÉCIMO SEGUNDO DÍGITO DO CNPJ).
CNPJ DV	DÍGITO VERIFICADOR DO NÚMERO DE INSCRIÇÃO NO CNPJ (DOIS ÚLTIMOS DÍGITOS DO CNPJ).
IDENTIFICADOR MATRIZ/FILIAL	CÓDIGO DO IDENTIFICADOR MATRIZ/FILIAL: 1 - MATRIZ 2 - FILIAL

NOME FANTASIA	CORRESPONDE AO NOME FANTASIA
SITUAÇÃO CADASTRAL	CÓDIGO DA SITUAÇÃO CADASTRAL: 01 - NULA 2 - ATIVA 3 - SUSPENSA 4 - INAPTA 08 - BAIXADA
DATA SITUAÇÃO CADASTRAL	DATA DO EVENTO DA SITUAÇÃO CADASTRAL
MOTIVO SITUAÇÃO CADASTRAL	CÓDIGO DO MOTIVO DA SITUAÇÃO CADASTRAL
NOME DA CIDADE NO EXTERIOR	NOME DA CIDADE NO EXTERIOR
PAIS	CÓDIGO DO PAIS
DATA DE INÍCIO ATIVIDADE	DATA DE INÍCIO DA ATIVIDADE
CNAE FISCAL PRINCIPAL	CÓDIGO DA ATIVIDADE ECONÔMICA PRINCIPAL DO ESTABELECEMENTO
CNAE FISCAL SECUNDÁRIA	CÓDIGO DA(S) ATIVIDADE(S) ECONÔMICA(S) SECUNDÁRIA(S) DO ESTABELECEMENTO
TIPO DE LOGRADOURO	DESCRIÇÃO DO TIPO DE LOGRADOURO
LOGRADOURO	NOME DO LOGRADOURO ONDE SE LOCALIZA O ESTABELECEMENTO.
NÚMERO	NÚMERO ONDE SE LOCALIZA O ESTABELECEMENTO. QUANDO NÃO HOUVER PREENCHIMENTO DO NÚMERO HAVERÁ 'S/N'.
COMPLEMENTO	COMPLEMENTO PARA O ENDEREÇO DE LOCALIZAÇÃO DO ESTABELECEMENTO
BAIRRO	BAIRRO ONDE SE LOCALIZA O ESTABELECEMENTO.
CEP	CÓDIGO DE ENDEREÇAMENTO POSTAL REFERENTE AO LOGRADOURO NO QUAL O ESTABELECEMENTO ESTA LOCALIZADO
UF	SIGLA DA UNIDADE DA FEDERAÇÃO EM QUE SE ENCONTRA O ESTABELECEMENTO
MUNICÍPIO	CÓDIGO DO MUNICÍPIO DE JURISDIÇÃO ONDE SE ENCONTRA O ESTABELECEMENTO

DDD 1	CONTÉM O DDD 1
TELEFONE 1	CONTÉM O NÚMERO DO TELEFONE 1
DDD 2	CONTÉM O DDD 2
TELEFONE 2	CONTÉM O NÚMERO DO TELEFONE 2
DDD DO FAX	CONTÉM O DDD DO FAX
FAX	CONTÉM O NÚMERO DO FAX
CORREIO ELETRÔNICO	CONTÉM O E-MAIL DO CONTRIBUINTE
SITUAÇÃO ESPECIAL	SITUAÇÃO ESPECIAL DA EMPRESA
DATA DA SITUAÇÃO ESPECIAL	DATA EM QUE A EMPRESA ENTROU EM SITUAÇÃO ESPECIAL

Figura 11 - Receita Federal, 2022

CNAEs	
Campo	Descrição
CÓDIGO	CÓDIGO DA ATIVIDADE ECONÔMICA
DESCRIÇÃO	NOME DA ATIVIDADE ECONÔMICA

Figura 12 - Receita Federal, 2022

NATUREZAS JURÍDICAS	
Campo	Descrição
CÓDIGO	CÓDIGO DA NATUREZA JURÍDICA
DESCRIÇÃO	NOME DA NATUREZA JURÍDICA

Figura 13 - Receita Federal, 2022

A partir dessas colunas segmentadas dentro do banco de dados, por meio de tabelas específicas, serão desenvolvidas toda a resolução de problemas alegados pelo cliente. Pois, como pôde ser visto nas figuras 10, 11 e 12, 13 temos todos os campos que ajudarão o cliente a solucionar suas indagações, como por exemplo, o campos CNPJ, empresa, capital social e etc.

5. DESENVOLVIMENTO DO PROJETO

O projeto tem como objetivo primordial desenvolver um sistema da informação minerando dados abertos da receita federal para desenvolver insights comerciais para uma empresa de telecom.

Desta forma, podemos definir mineração de dados como sendo a técnica de extrair grandes volumes de dados de fontes de dados que se encontra em um mesmo local ou locais diferentes.

Não obstante disso, pode ser verificado que várias instituições e órgãos governamentais disponibilizam seus dados para análise porque é necessária uma democratização da informação e transparência na gestão governamental.

Alguns sites disponibilizam seus dados e suas informações de forma legal para estudos e análises de mercado, um desses sites é o site da receita federal (<https://www.gov.br/receitafederal/pt-br>) que será usado para levantar as análises referentes às empresas de todo o Brasil.

A Partir desses dados localizados e baixados pode ser desenvolvido um processo chamado Extração, Transformação e Carga (*Extract Transform Load - ETL*) que tem como objetivo pegar essas bases de dados da sua fonte primária, transformar esses dados para os seus próprios objetivos, fazendo cálculos, excluindo informações e acrescentando ideias além de fazer a carga das informações em um banco de dados de preferência dentro de um *data warehouse*.

Dessa forma, será desenvolvido bancos de dados que são conjuntos de informações com relacionamentos entre si, que tem como objetivo ser nossa fonte primaria de informação dentro do projeto, pois nele irá conter os dados baixados do site da receita federal.

A partir do desenvolvido dos bancos de dados será criado informações em Power BI que serão vistas geralmente em formas de gráficos e painéis interativos tem como objetivo por meio dos dados gerar ideias e informações sobre os dados mostrados, que possibilitará a tomada de decisão de uma empresa de Telecom fictícia.

5.1. Mapeamento das etapas do projeto para a criação do business intelligence

Abaixo segue o mapeamento feito de todas as etapas necessarias para o desenvolvimento do projeto.



Figura 14 - Fluxo das Etapas do Projeto

A imagem acima tem como objetivo exemplificar todas as etapas do desenvolvimento do projeto. Abaixo poderão ser visualizadas as aberturas desse mapeado acima.

Para alcançar o objetivo específico 1, foi necessário:

Desenvolver perguntas para saber qual era o real objetivo do cliente e quais questões ele queria responder por meio do estudo levantado.

Tabela 6 - Requisitos do Produto

Perguntas	Sim	Não
Deseja encontrar oportunidades de mercado como endereço, telefone e localidade dos futuros <i>prospects</i> ?	X	
Deseja saber o capital social das empresas que podem vim a ser novos clientes?	X	
Deseja visualizar dados de geolocalização das empresas?	X	
Será importante fazer distinção de clientes ativos da base e novos clientes?		X
O sistema da informação precisa mostrar a quantidade de clientes por estado?	X	

O sistema da informação precisa mostrar a quantidade de empresas matrizes e filiais?	X	
Deseja visualizar os dados de empresas abertas, fechadas e etc?	X	
Deseja visualizar faturamento das empresas?		X
Deseja ver a segmentação das empresas por pequenas empresas, médias e grandes empresas?	X	
Tem objetivo de ver os dados de segmentos por CNAE?	X	
Deseja visualizar as empresas por natureza jurídica?	X	
Deseja ver cruzamento de dados de natureza jurídica por capital social?	X	
Tem como objetivo aumentar a sua carteira de clientes	X	
Deseja tentar realizar venda para empresas com capital social abaixo de 1 milhão de reais cadastrado na Receita Federal?		X
Deseja conseguir rastrear empresas em todo o Brasil?	X	
Deseja descobrir o endereço e o telefone dessas empresas?	X	
Deseja ter um painel de fácil visualização das informações acerca da empresa	X	
Deseja poder localizar o porte da empresa	X	
Deseja visualizar gráficos e indicadores matemáticos	X	

Para alcançar o objetivo específico 2, foi necessário:

Acessar o site da Receita Federal, que se encontra no link a seguir: <https://www.gov.br/receitafederal/ptbr/assuntos/orientacaotributaria/cadastros/consultas/dados-publicos-cnpj>, e fazer os downloads das bases de dados no computador.

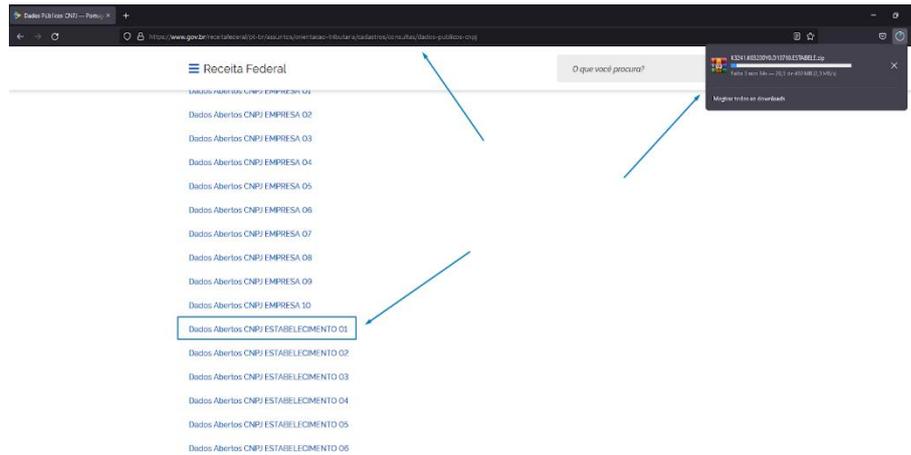


Figura 15 - Tela do Site da Receita Federal

Após o *download* das bases de dados foi necessário consolidar as bases de empresas e estabelecimentos, pois elas possuem um grande volume de dados. Foi usado um comando para consolidação no CMD - *Prompt* de Comando: `for %f in (*.csv) do type "%f" >> "pasta para qual vai o arquivo\"nome do arquivo que deseja.CSV"`. Essa consolidação deixa todos os arquivos em apenas um arquivo.

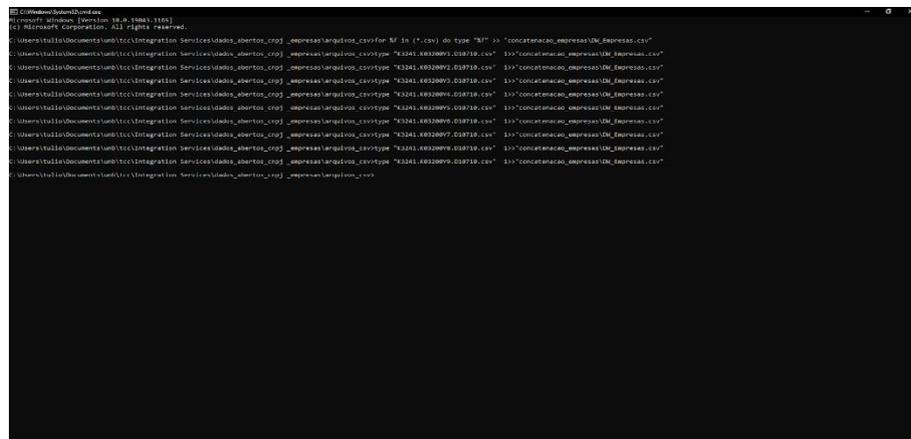


Figura 16 - CMD: Processo de Concatenação das Segmento de Empresas

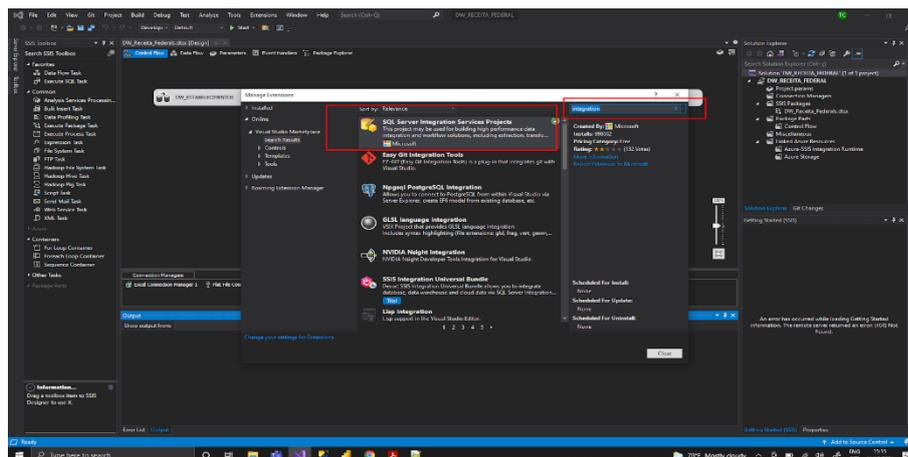


Figura 19 - Tela de Download do Integration Services

Após a instalação do *software* de *ETL*, é importante fazer a instalação do banco de dados *SQL Server* e criar as tabelas de dados que vão receber os dados da Receita Federal.

Para fazer a Instalação do *Software SQL SERVER* é necessário acessar o site: <https://www.microsoft.com/pt-br/sql-server/sql-server-downloads>.

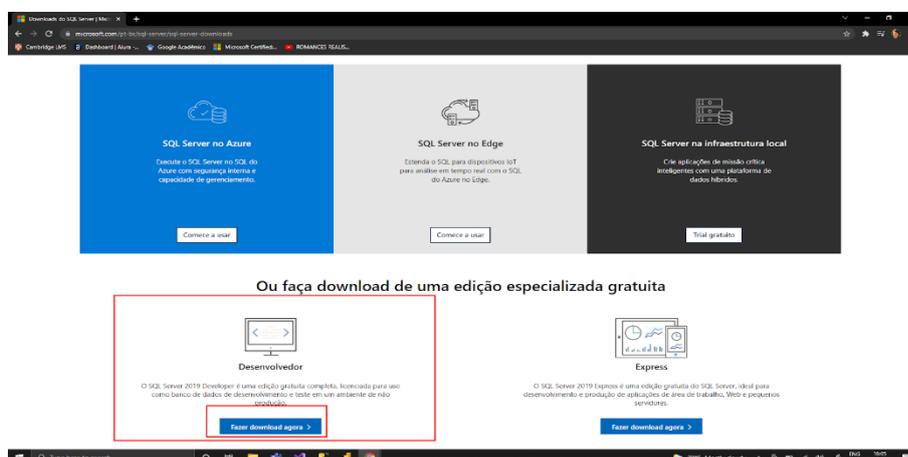


Figura 20 - Tela de Download Banco de Dados SQL Server

Após a instalação do banco de dados é importante instalar o *Microsoft SQL Server Management Studio*, para acessar o banco de dados e criar os bancos de dados.

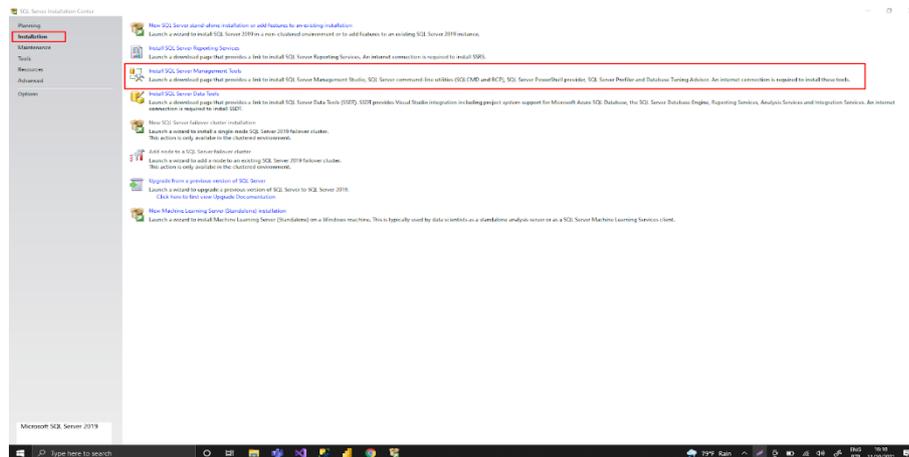


Figura 21 - Tela de Download do Management Studio

Logo em seguida, foi criado o banco de dados, que terá o nome de [dw_receita_federal] e foi desenvolvido usando o seguinte comando: *create database [dw_receita_federal]*, onde pode ser visto facilmente na imagem a baixo:

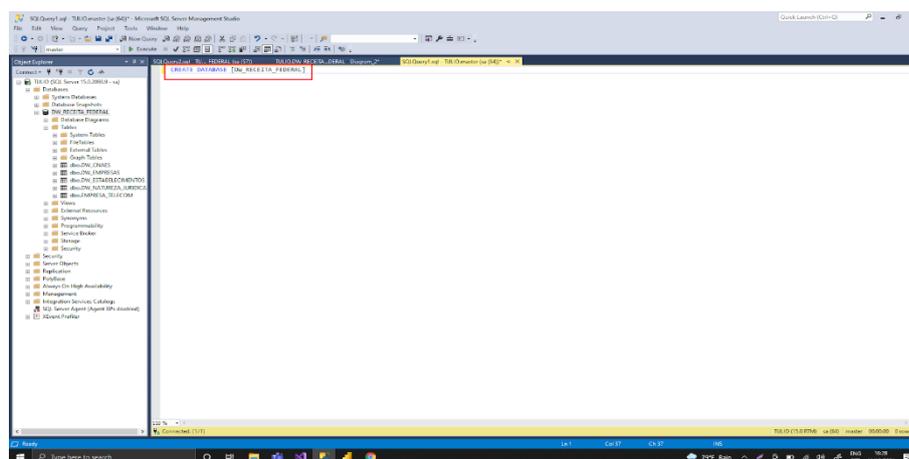


Figura 22 - Tela de Criação do Banco de Dados

Não obstante do processo de criação do banco de dados, foram criadas 5 tabelas dentro da base de dados. Elas têm o objetivo de pegar os dados da Receita Federal e a base de dados da empresa de telecom.

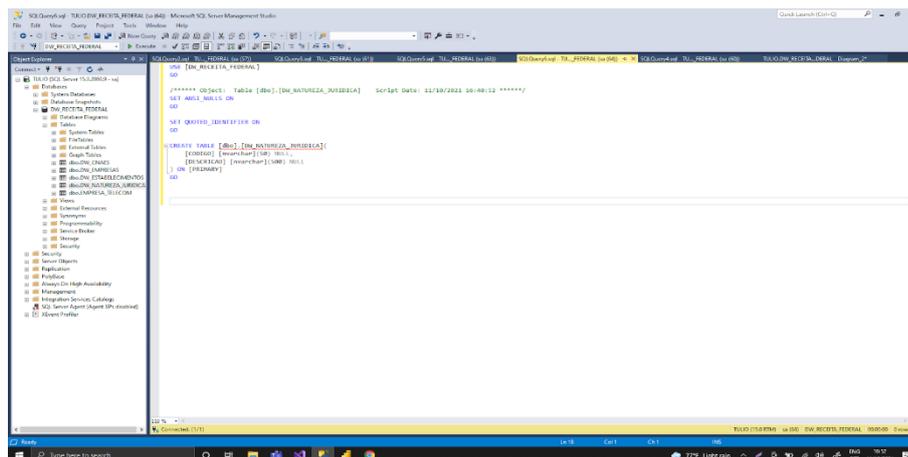


Figura 26 - Tela de Criação da Tabela Natureza Jurídica

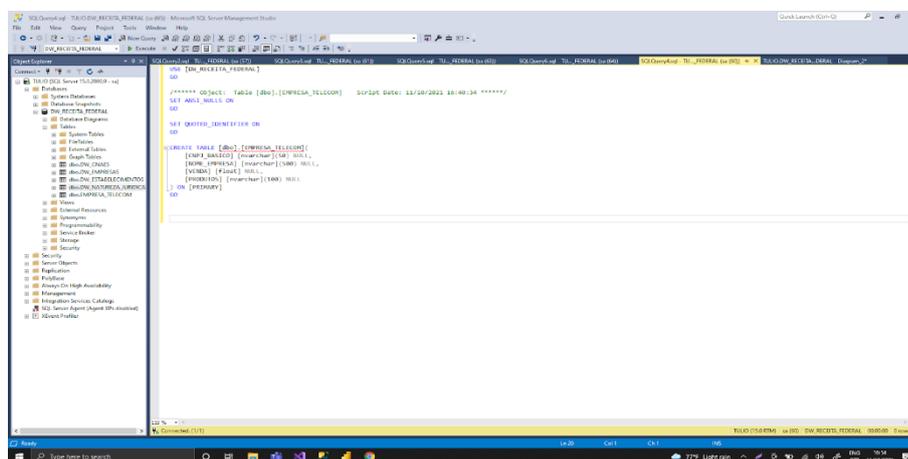


Figura 27 - Tela de Criação da Tabela Empresa Telecom

A partir da criação das bases de dados, foi desenvolvido o processo de *ETL* utilizando a ferramenta do *Integration Services* que se encontra dentro do *Visual Studio*. Esse processo *ETL*, tem como objetivo primordial pegar os dados baixados da Receita Federal e importar esses dados para as tabelas criadas no banco de dados.

Usando os comandos de *Data flow task* para criar as áreas restritas para fazer o processo de *ETL* e utilizando processos simples como *flat file sources*, *data conversion* e conectando esses arquivos aos bancos de dados pode ser criado o processo *ETL*, onde vamos pegar os dados que se encontram nos arquivos em formato *CSV* e colocados no banco de dados.

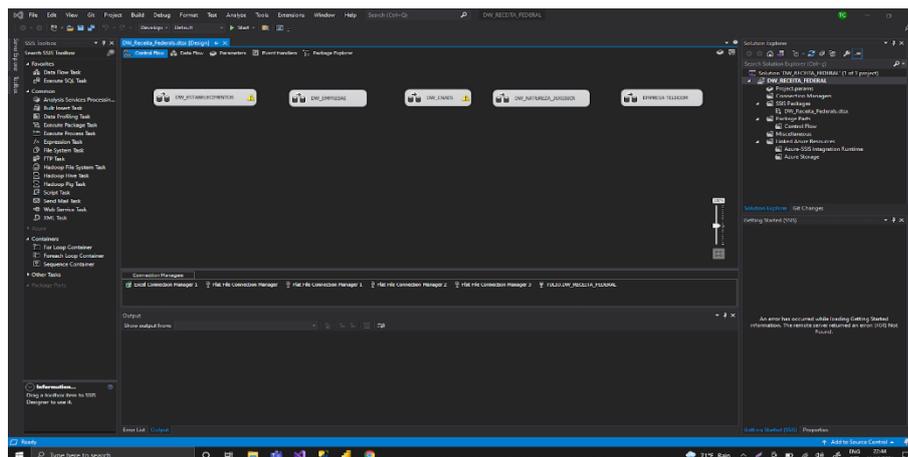


Figura 28 - Tela de Criação do Data Flow Task

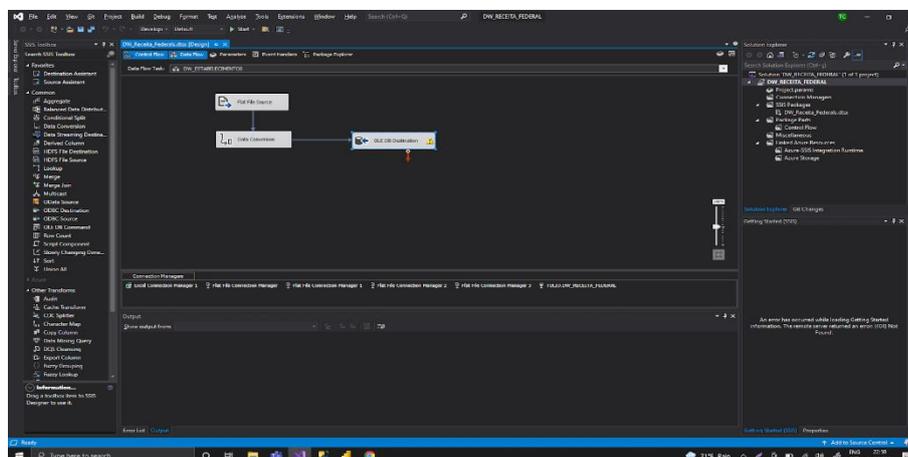


Figura 29 - Tela de Processo de Extração de dados

A partir do levantamento das necessidades do cliente são localizadas as bases de dados da RF. Inicia-se o processo de *ETL*, que tem como objetivo pegar todos os dados da Receita Federal e colocar esses dados dentro de um banco de dados como podem ser visto nas figuras 22 e 23.

Logo após ao processo de *ETL* podem ser analisado os bancos de dados que tem como objetivo mostrar todos os dados baixados da Receita Federal, para desenvolvimento do *dashboard*.

Para alcançar o objetivo específico 3, foi necessário:

Fazer a instalação do software *Power BI* que pode ser instalado no site:
<https://powerbi.microsoft.com/pt-br/desktop/>

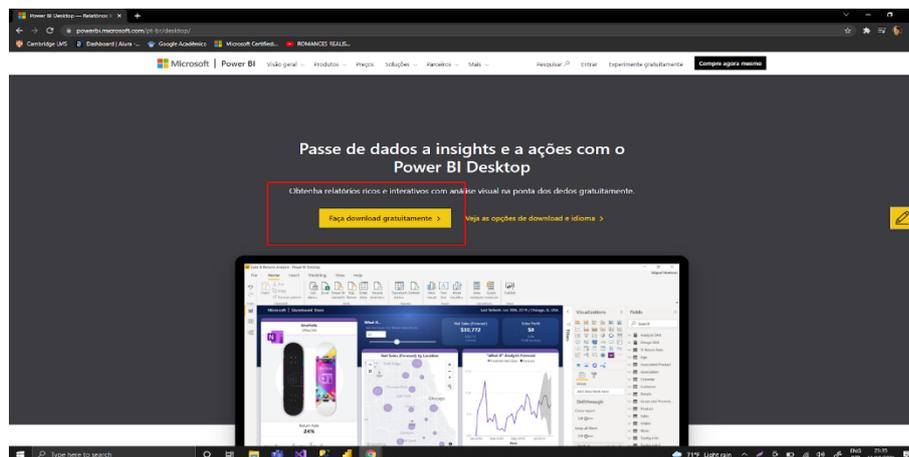


Figura 30 - Tela de Download do Power BI

Logo após o *download* do *software* foram importados todos os dados para o *Power BI*, os dados que foram importados para a ferramenta são os dados exportados do site da Receita Federal que foram inseridos dentro no banco de dados criado chamado DW_RECEITA_FEDERAL pelo processo do *Integration Services*.

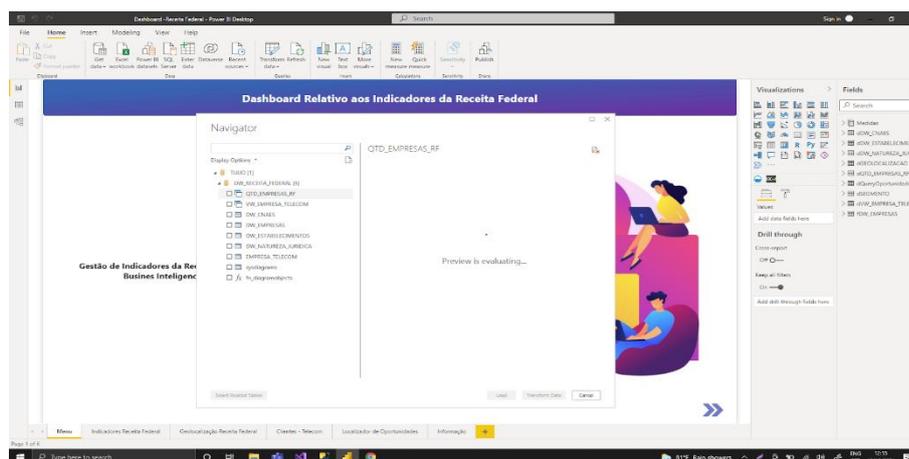


Figura 31 - Tela de Conexão do Power BI com o Banco de Dados

Com os dados dentro no *Power BI*, foram desenvolvidas todas as conexões e os vínculos entre as tabelas para que todas as tabelas se comuniquem.

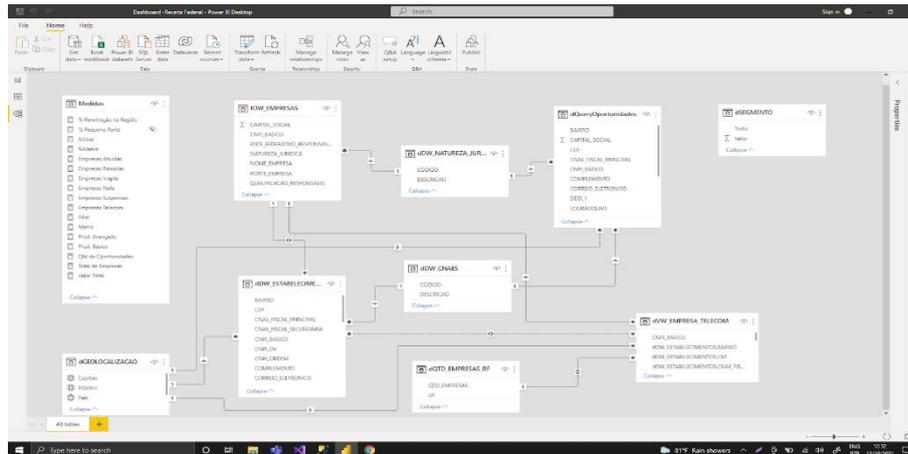


Figura 32 - Tela de Vinculação dos Bancos de Dados

6. PROPOSTA DE BI PARA UMA TELECOM

Saindo do pressuposto que *Business Intelligence* ou, inteligência de negócio, tem como objetivo a coleta, organização e toda a análise de dados, além da liberação das informações e o monitoramento dessas informações.

Deste modo, o *BI* é detentor de um conjunto de ferramentas e conhecimentos necessários para desenvolver um projeto.

Com este contexto foi utilizado uma ferramenta de *ETL* chamada *Integration Services* e um banco de dados o *SQL server* para desenvolver todo o *back-end*, e para o desenvolvimento do *front-end* será utilizado uma ferramenta chamada *Power BI*.

6.1. Visão Geral do Projeto

Na última etapa do projeto foi desenvolvido um projeto em *Power BI* que teve como objetivo solucionar os questionamentos dos clientes, como pode ser visto na Tabela 5 e 6 foram feitas perguntas para o cliente que possibilitou criar os requisitos do cliente e do produto respectivamente.

Com isso, foi desenvolvido um menu que tem como objetivo fazer uma apresentação do *dashboard* interativa, que a partir do clique nas setas na parte inferior ao lado direito pode transportar a visualização para a página de indicadores da Receita Federal.

Este menu também foi inspirado nas cores da própria Receita Federal trazendo uma melhor harmonização da tela e uma melhor acessibilidade de navegação.

O objetivo real dessa visualização inicial é criar uma imagem personalizada do dashboard com intuito de o indivíduo começar a mexer na ferramenta clicando nas setas no canto inferior a direita.

Dashboard Relativo aos Indicadores da Receita Federal

Gestão de Indicadores da Receita Federal
Business Intelligence



Figura 33 - Menu Inicial

Logo após o clique feito na aba home nas setas no canto inferior direito, o dashboard se modifica e pode ser visto uma aba feita com inumeros indicadores matemáticos.

Ao acessar a tela de indicadores da Receita Federal, é possível visualizar todas as informações relativas da Receita Federal, sendo elas informações de empresas ativas, matrizes, filiais, CNAE, natureza jurídica e quantidade de empresas por Estado.

No dashboard pode ser analisado indicadores gerais da receita federal, como o total de empresas tanto ativas quanto canceladas além de ser visto também dados de empresas matrizes e filiais e os seus determinados segmentos e as quantidades de empresas por estados.

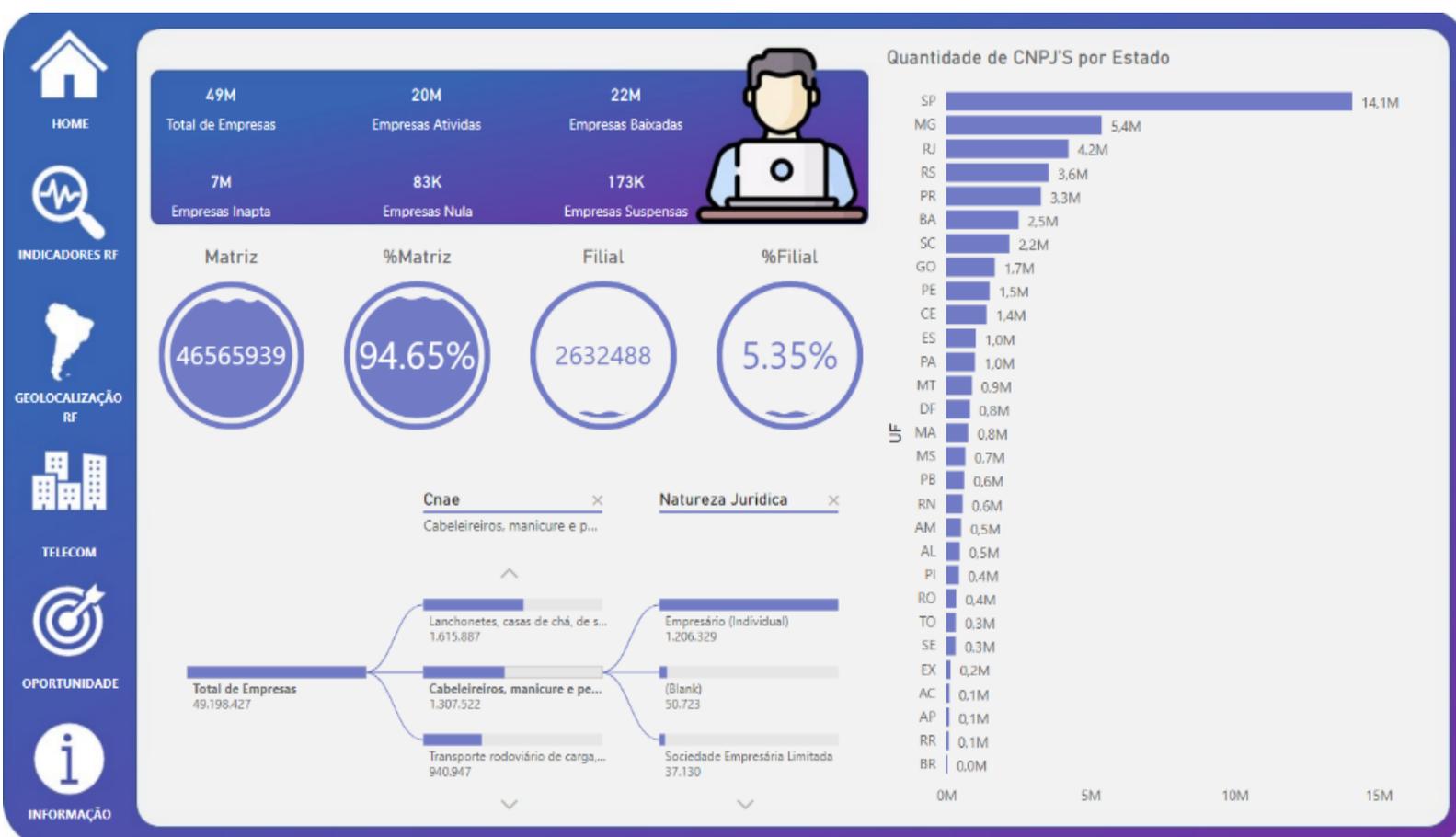


Figura 34 - Indicadores da Receita Federal

Clicando no ícone de mapa onde consta o nome geolocalização RF, será ativado outra aba do *dashboard* na qual consta os indicadores de geolocalização, mostrando os dados das empresas dentro do Brasil e a sua natureza jurídica em cruzamento com o capital social além de todo o capital social dessas empresas.

Outro ponto desse painel de indicadores é que podem ser filtradas as informações por CNAE, situação cadastral, natureza jurídica e porte da empresa.

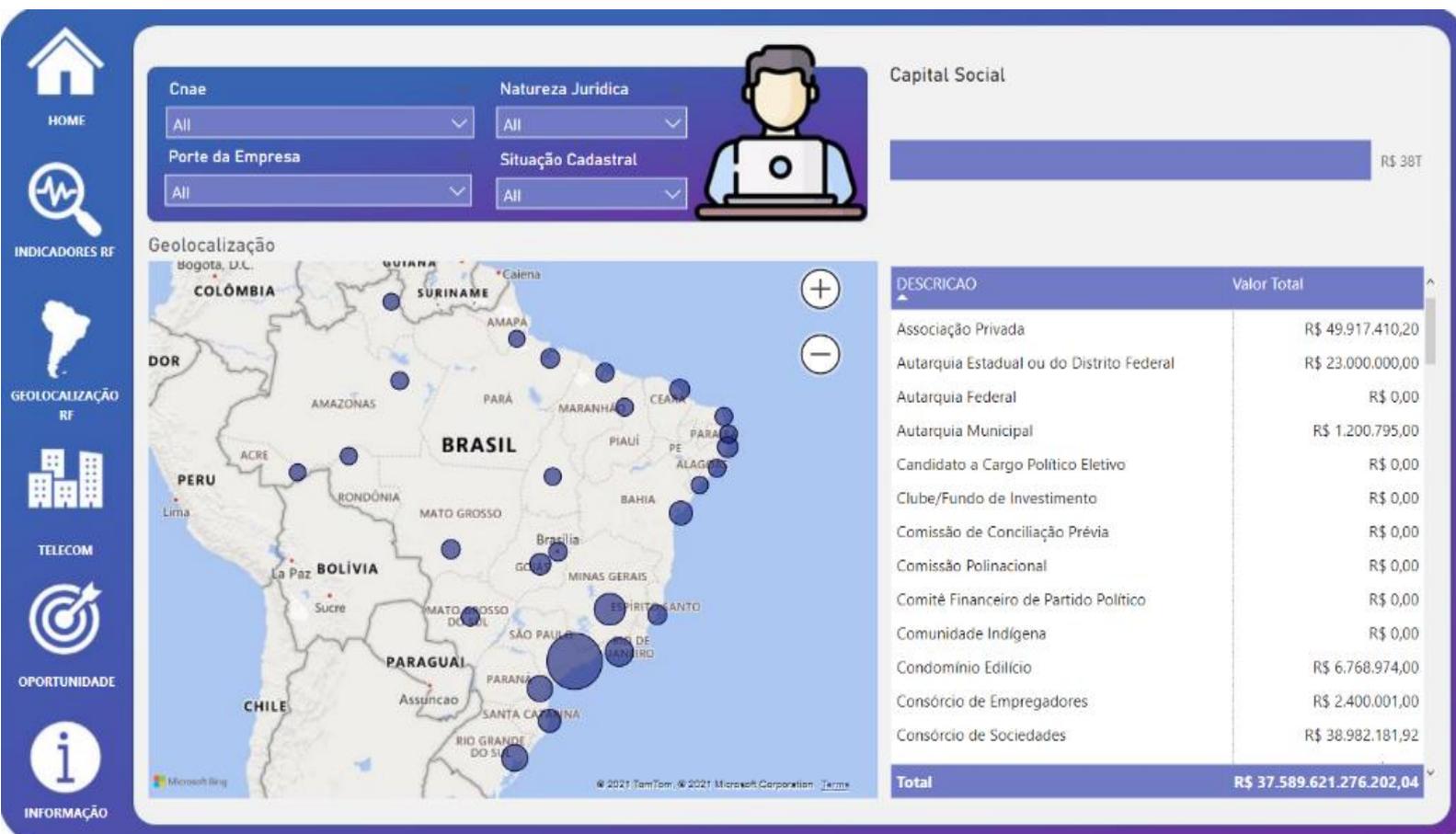


Figura 35 - Geolocalização da Receita Federal

Ao clicar no ícone de prédios onde consta o nome Telecom, podem ser analisados os dados georreferenciados da empresa, além do seu valor de faturamento respectivamente.

Esses dados são importantes para termos uma notação total do tamanho da empresa que estamos querendo encontrar as oportunidades e visualizar onde existe seu maior numero de clientes.

Não obstante disso, pode ser visualizado os dados relativos a quantidade vendida de seus produtos.

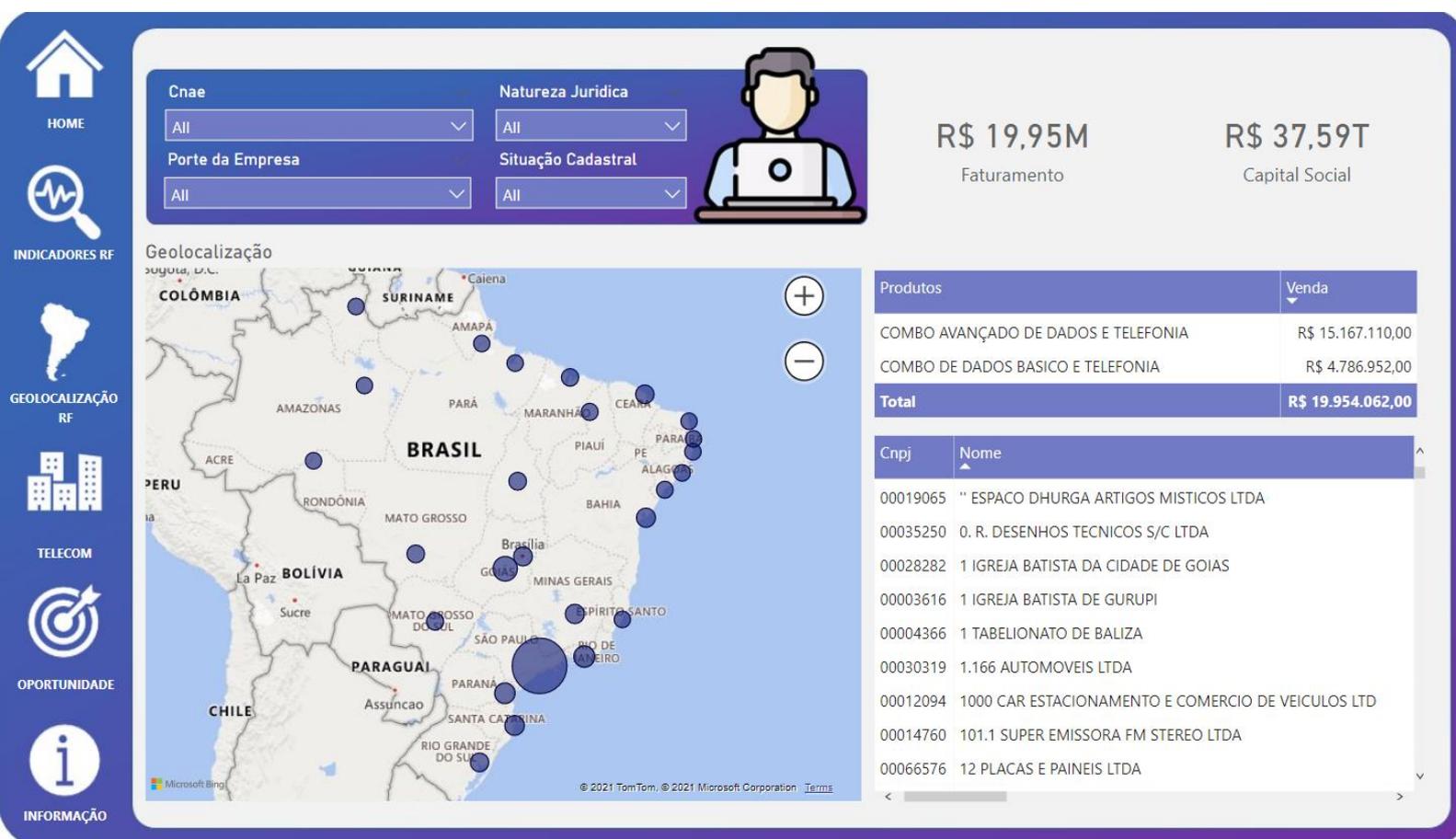


Figura 36 - Dados da Empresa de Telecom

No painel de oportunidades podem ser encontradas empresas que possuem faturamento acima de 1 milhão de reais, seus telefones e endereços, além de uma série de dados que serão utilizadas para a empresa de Telecom fazer prospecção.

Esses dados tem como objetivo responder as questões da empresa como um todo. Pois a empresa sempre quis saber quais eram as oportunidades dentro do Brasil tendo acesso a seu endereço, telefone e email.

Quantidade de Oportunidades

586.901

NOME_EMPRESA	CNPJ_BASIC
AGRA VERBIER INCORPORADORA LTDA - EM RECUPERACAO	09580789
AGROPECUARIA COMIN LTDA	19351784
AROABA EMPREENDIMENTOS IMOBILIARIOS LTDA - EM REC	07311447
AROABA EMPREENDIMENTOS IMOBILIARIOS LTDA - EM REC	07311447
BLUE BEACH HOLDING LTDA	20865365
- COMPANHIA DE TECNOLOGIA DA INFORMACAO DE CACHOEI	31720485
ELLEN MIMI LOCACOES IMOBILIARIAS LTDA	20431413
EVCOSTA ADMINISTRADORA DE IMOVEIS LTDA	23265974
IDEA ZARVOS PLANEJAMENTO IMOBILIARIO LTDA - SCP A	28265019
IDEA ZARVOS PLANEJAMENTO IMOBILIARIO LTDA - SCP F	28263946
JMC STOPPIGLIA ADMINISTRACAO DE BENS PROPRIOS LTD	03214985
" LABORATORIO OSLER DO BRASIL LTDA."	05020272
" LABORATORIO OSLER DO BRASIL LTDA."	05020272
LASER FAST DEBILITACAO LTDA SCP	36763371

Figura 37 - Oportunidades de Mercado

No último ícone de informações podem ser encontradas a descrição de todas as colunas encontradas no *dashboard*.

Nessas informações tem como objetivo geral mostrar o que cada tabela e coluna significa dentro do dashboard.

 HOME
 INDICADORES RF
 GEOLOCALIZAÇÃO RF
 TELECOM
 OPORTUNIDADE
 INFORMAÇÃO

Seja Bem-Vindo(a) ao Painel de Informações



Informações das Tabelas

EMPRESAS	Descrição
Campo	Descrição
CNPJ BASICO	NUMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ)
RAZÃO SOCIAL / NOME EMPRESARIAL	NOME EMPRESARIAL DA PESSOA JURÍDICA
NATUREZA JURÍDICA	CODIGO DA NATUREZA JURÍDICA
QUALIFICAÇÃO DO RESPONSÁVEL	QUALIFICAÇÃO DA PESSOA FÍSICA RESPONSÁVEL PELA EMPRESA
CAPITAL SOCIAL DA EMPRESA	CAPITAL SOCIAL DA EMPRESA
PORTE DA EMPRESA	CODIGO DO PORTE DA EMPRESA: 1 - NÃO INFORMADO 2 - MICRO EMPRESA 03 - EMPRESA DE PEQUENO PORTE 05 - DEMAIS
ENTE FEDERATIVO RESPONSÁVEL	O ENTE FEDERATIVO RESPONSÁVEL É PREENCHIDO PARA OS CASOS DE GRUPOS E ENTIDADES DO GRUPO DE NATUREZA JURÍDICA 1.000. PARA AS DEMAIS NATUREZAS, ESTE ATRIBUTO FICA EM BRANCO.

NATUREZAS JURÍDICAS	Descrição
Campo	Descrição
CODIGO	CODIGO DA NATUREZA JURÍDICA
DESCRIÇÃO	NOME DA NATUREZA JURÍDICA

CNAEs	Descrição
Campo	Descrição
CODIGO	CODIGO DA ATIVIDADE ECONÔMICA
DESCRIÇÃO	NOME DA ATIVIDADE ECONÔMICA

ESTABELECIMENTOS	Descrição
Campo	Descrição
CNPJ RADICAL	ASSINADO PARA O REGISTRO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ)
CNPJ ESTAB	NUMERO DO ESTABELECIMENTO DE INSCRIÇÃO NO CNPJ (DOIS ÚLTIMOS DÍGITOS DO CNPJ)
CNPJ DV	CODIGO VERIFICADOR DO NUMERO DE INSCRIÇÃO NO CNPJ (DOIS ÚLTIMOS DÍGITOS DO CNPJ)
IDENTIFICADOR INSTITUCIONAL	CODIGO DO IDENTIFICADOR INSTITUCIONAL: 1 - INTERJUR 2 - PRAZ

SITUAÇÃO CADASTRAL	Descrição
Campo	Descrição
DATA SITUAÇÃO CADASTRAL	DATA DO EVENTO DA SITUAÇÃO CADASTRAL
ATIVIDADE SITUAÇÃO CADASTRAL	CODIGO DO EVENTO DA SITUAÇÃO CADASTRAL
NOME DA CIDADE NO EXTERIOR	NOME DA CIDADE NO EXTERIOR
PAIS	CODIGO DO PAIS
DATA DE INÍCIO ATIVIDADE	DATA DE INÍCIO DA ATIVIDADE
CNAE PRINCIPAL ATIVIDADE	CODIGO DA ATIVIDADE ECONÔMICA PRIMÁRIA DO ESTABELECIMENTO
CNAE SECUNDARIA	CODIGO DAS ATIVIDADES ECONÔMICAS SECUNDARIAS DO ESTABELECIMENTO
TIPO DE LOGRADOURO	DESCRIÇÃO DO TIPO DE LOGRADOURO
LOGRADOURO	NOME DO LOGRADOURO ONDE SE LOCALIZA O ESTABELECIMENTO
NUMERO	NUMERO ONDE SE LOCALIZA O ESTABELECIMENTO, QUANDO NÃO EXISTIR, PREENCHER O NUMERO NULAS, 999
COMPLEMENTO	COMPLEMENTO PARA O ENDEREÇO DE LOCALIZAÇÃO DO ESTABELECIMENTO
BARRIO	BARRIO ONDE SE LOCALIZA O ESTABELECIMENTO
CAP	CODIGO DE PADRÃO/CONTADOR REFERENTE AO LOGRADOURO NO QUAL O ESTABELECIMENTO ESTÁ LOCALIZADO
UF	SIGLA DA UNIDADE DA FEDERAÇÃO EM QUE SE ENCONTRA O ESTABELECIMENTO
MUNICIPIO	CODIGO DO MUNICIPIO DE JURISDIÇÃO ONDE SE ENCONTRA O ESTABELECIMENTO

DDD 1	CONTEN O DDD 1
TELEFONE 1	CONTEN O NUMERO DO TELEFONE 1
DDD 2	CONTEN O DDD 2
TELEFONE 2	CONTEN O NUMERO DO TELEFONE 2
DDD DO FAX	CONTEN O DDD DO FAX
FAX	CONTEN O NUMERO DO FAX
COMER ELETRONICO	CONTEN O E-MAIL DO CONTRIBUANTE
SITUAÇÃO ESPECIAL	SITUAÇÃO ESPECIAL DA EMPRESA
DATA DA SITUAÇÃO ESPECIAL	DATA EM QUE A EMPRESA ENTROU EM SITUAÇÃO ESPECIAL

Figura 38 - Informações Gerais

6.2. Solucionando os problemas

Saindo das necessidades do cliente citadas na Tabela 5, o *dashboard* foi desenvolvido para solucionar todos os seus questionamentos em apenas uma tela, onde ele pode inserir todas as informações desejadas, como pode ser visto na imagem abaixo na figura 40.

Nessa visão chamada de oportunidades dentro do dashboard o cliente pode facilmente escolher um segmento de CNAE específico para filtrar as suas oportunidades dentro de um determinado segmento, além de ser possível filtros por estado, situação cadastral e natureza jurídica.

Outro ponto que deve ser constado aqui é que todas as empresas encontradas como oportunidade tem um capital social acima de 1 milhão de reais.

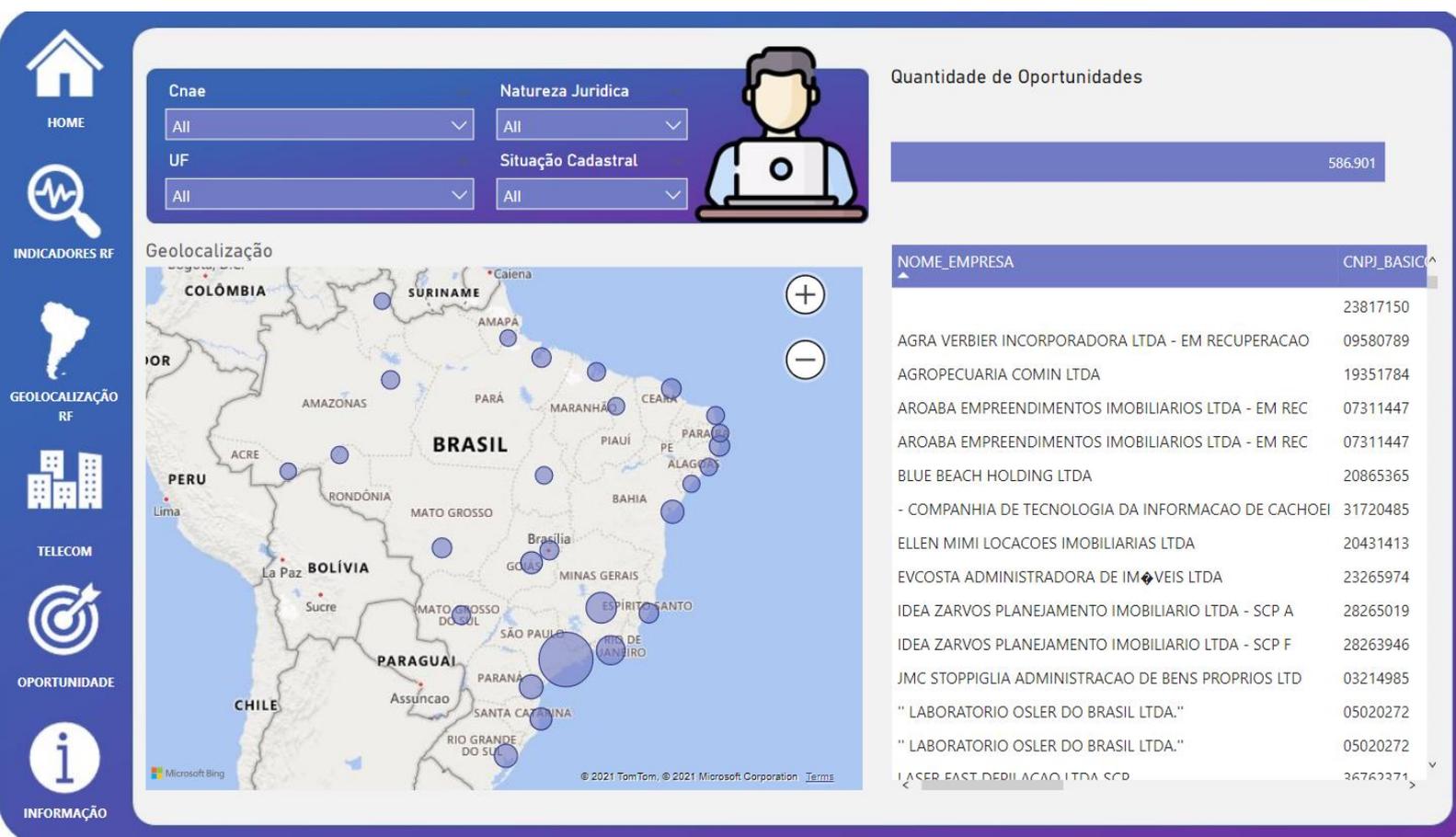


Figura 39 - Visualização de Oportunidade

Expandindo a tabela encontrada na aba oportunidade que se localiza no canto inferior direito pode ser localizado todos os dados solicitados para o cliente, sendo eles:

- NOME DA EMPRESA
- CNPJ
- UF
- BAIRRO
- ENDEREÇO
- TELEFONE

NOME_EMPRESA	CNPJ_BASICO	UF	CEP	BAIRRO	TIPO_LOGRADOURO	LOGRADOURO	COMPLEMENTO	DDD_1	TELEFONE_1
AGRA VERBIER INCORPORADORA LTDA - EM RECUPERACAO	23817150	PR	87160000	CENTRO	AVENIDA	MUNHOZ DA ROCHA		44	33051631
AGROPECUARIA COMIN LTDA	09580789	SP	04571010	CIDADE MONCOES	AVENIDA	ENGENHEIRO LUIZ CARLOS BERRINI	ANDAR: 11; SALA: AGRA VERBIER LTDA;	11	43737000
AGROPECUARIA COMIN LTDA	19351784	MS	79041903	BAIRRO TIRADENTES	RUA	JOAQUIM MURTINHO		67	21073535
AROABA EMPREENDIMENTOS IMOBILIARIOS LTDA - EM REC	07311447	SP	04548005	VILA OLIMPIA	AVENIDA	DOUTOR CARDOSO DE MELO	ANDAR 10 CONJ 101	11	32673210
AROABA EMPREENDIMENTOS IMOBILIARIOS LTDA - EM REC	07311447	SP	05315030	VILA NOVA LEOPOLDINA	RUA	NAGEL		11	32960100
BLUE BEACH HOLDING LTDA	20865365	SP	02336000	AGUA FRIA	RUA	JOSE DE ALBUQUERQUE MEDEIROS	APT 07	11	22068425
- COMPANHIA DE TECNOLOGIA DA INFORMACAO DE CACHOEI	31720485	ES	29300100	SEDE	RUA	VINTE E CINCO DE MARCO			
ELLEN MIMI LOCACOES IMOBILIARIAS LTDA	20431413	SP	13092133	NOVA CAMPINAS	AVENIDA	DOUTOR HERMAS BRAGA		11	55381039
EVCOSTA ADMINISTRADORA DE IMOVEIS LTDA	23265974	PR	86050730	GUANABARA PARQUE BOULEVARD	RUA	DOUTOR DIMAS DE BARROS	APT 801	43	33235102
IDEA ZARVOS PLANEJAMENTO IMOBILIARIO LTDA - SCP A	28265019	SP	05443000	VILA MADALENA	RUA	NATINGUI	CONJ 8	11	35501849
IDEA ZARVOS PLANEJAMENTO IMOBILIARIO LTDA - SCP F	28265946	SP	05443000	VILA MADALENA	RUA	NATINGUI	CONJ 8	11	35501849
JMC STOPPIGLIA ADMINISTRACAO DE BENS PROPRIOS LTD	03214985	SP	13278403	VALE DO ITAMARACA	RUA	DOUTOR JOAO MENDES JUNIOR		19	38719500
" LABORATORIO OSLER DO BRASIL LTDA."	05020272	SP	01451917	JARDIM PAULISTANO	AVENIDA	BRIGADEIRO FARIA LIMA	CONJ 721	11	30319484
" LABORATORIO OSLER DO BRASIL LTDA."	05020272	SP	07775240	EMPRESARIAL PAINEIRA (JORDANESIA)	AVENIDA	RIBEIRAO DOS CRISTAIS (G PRETO)	GALPAO2 PARTE A	11	30979991
LASER FAST DEPILACAO LTDA SCP	36762371	SP	15015100	CENTRO	RUA	ANTONIO DE GODOY		17	41413262
LASER FAST DEPILACAO LTDA SCP ARACATUBA V	36967250	SP	15015100	CENTRO	RUA	ANTONIO DE GODOY		17	41413262
LASER FAST DEPILACAO LTDA SCP AVARE IV	40977539	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BALNEARIO CAMBORIU	40616261	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BALNEARIO CAMBORIU	40619567	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BAURU SHOPPING II	40812485	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BLUMENAU	40574298	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BLUMENAU IV	40574474	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BLUMENAU V	40574506	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BRAGANCA PAULISTA I	40683654	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262
LASER FAST DEPILACAO LTDA SCP BRASILIA VIII	36970495	SP	15015100	CENTRO	RUA	ANTONIO DE GODOY		17	41413262
LASER FAST DEPILACAO LTDA SCP COTIA V	40517717	SP	15090190	CHACARA MUNICIPAL	AVENIDA	FRANCISCO DAS CHAGAS OLIVEIRA	ANEXO XVIII	17	41413262

Figura 40 - Tela de Prospecção de Clientes

6.3. Tendências

O cliente atualmente atende 49.999 clientes PJ que correspondem em um faturamento R\$ 19.954.502 de reais aproximadamente. Foi identificado, como pode ser visto na figura 39, que o cliente possui oportunidade de 586.901 clientes para serem prospectados. Caso ele venda para 5% desta base de prospecção a carteira de clientes da empresa subirá para 79.344 clientes. Saindo dessas informações e do valor do ticket médio da empresa que está em torno de R\$ 399 reais, podemos analisar que o cliente aumentaria sua receita em 159%.

Atualmente				
Faturamento	Clientes	Ticket Medio		
R\$ 19.954.502	49.999	R\$	399,10	

Caso ele Aumente sua Carteira em:				
Oportunidade	586.901			
Aumento	Quantidade de Clientes	Faturamento	Crescimento	
1%	55.868	R\$ 22.296.812,28	112%	
2%	61.737	R\$ 24.639.122,56	123%	
3%	67.606	R\$ 26.981.432,85	135%	
4%	73.475	R\$ 29.323.743,13	147%	
5%	79.344	R\$ 31.666.053,41	159%	
6%	85.213	R\$ 34.008.363,69	170%	
7%	91.082	R\$ 36.350.673,97	182%	
8%	96.951	R\$ 38.692.984,25	194%	
9%	102.820	R\$ 41.035.294,54	206%	
10%	108.689	R\$ 43.377.604,82	217%	

Figura 41 - Tela de Tendência de Mercado

7. CONCLUSÃO

O trabalho tinha como objetivo resolver um objetivo geral e três objetivos específicos sendo eles respectivamente:

- Desenvolver um sistema de informação em *Power BI* capaz de visualizar indicadores matemáticos a partir dos dados abertos da Receita Federal para uma Telecom fictícia.
- Levantar e identificar as necessidades de um cliente da área de Telecom por informações para tomada de decisão sobre nicho de mercado.
- Criar um processo Extração, Transformação e Carga (*ETL*) no *Integration Services* e um banco de dados no *SQL SERVER* com base nos arquivos da Receita Federal.
- Criar o sistema da informação em *business analytics tool* utilizando a ferramenta *Power BI*

Todos esses objetivos foram realizados. A conclusão dos objetivos gerou um resultado de mais de 500 mil oportunidades em todo o Brasil.

Não obstante disso este trabalho teve como objetivo mostrar todo um processo de análise de dados focando na exportação da base de dados da Receita Federal além da criação de processos de *ETL*, criação de banco de dados e desenvolvimento de *dashboards*. Essa base de dados da Receita Federal abriu várias possibilidades e visões que no decorrer do projeto realizado veio a solucionar as demandas do cliente.

Um ponto que pode ser levado em consideração é que baixar os dados da Receita Federal por si só não resolve o problema, pois esses dados vem em formatos não usuais e com seus dados não organizados. Por isso é a importância de fazer sempre um processo de *ETL*.

Não obstante, pode ser verificado que todas as informações da Receita Federal consolidadas geram dados bastantes robustos que precisam ser jogados em um banco de dados para só assim ser feito um *dashboard*. Deste modo, podemos concluir que a importância do *Business Intelligence* para a análise traz bastante benefícios para a resolução de problemas reais dentro das instituições.

Pode ser elencado também que foi de suma importância entender as reais necessidades do cliente, pois só assim foi possível realmente buscar ferramentas e dados para levantar informações que vieram a ser aproveitadas pela empresa.

Em síntese, todas as demandas da empresa fictícia de Telecom foram solucionadas por meio de um *dashboard* em *Power BI* que responde todas as perguntas feitas para a empresa e mostra os dados de forma simples e objetiva para o cliente tomar as melhores decisões com as informações oferecidas.

8. REFERÊNCIAS

- AGUENA, Carine Calixto. **ETL4NoSQL**: um framework programável para extração, transformação e carga de banco de dados NoSQL. 2017. 94 f. Dissertação (Mestrado) – Universidade Federal de Pernambuco. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/29771/1/DISSERTA%c3%87%c3%83O%20Carine%20Calixto%20Aguena.pdf>. Acesso em: 28 de fevereiro de 2022.
- AMO, Sandra de. **Técnicas de mineração de dados**, 2003. Disponível em: https://www.researchgate.net/profile/Sandra-Amo/publication/260300816_Tecnicas_de_Mineracao_de_Dados/links/54230bd80cf290c9e3ae25e3/Tecnicas-de-Mineracao-de-Dados.pdf. Acesso em: 28 de fevereiro de 2022.
- ARAÚJO, Lucas de Ramos; SOUZA, Jairo Francisco de. Aumentando a transparência do governo por meio da transformação de dados governamentais abertos em dados ligados. **Revista Eletrônica de Sistemas de Informação**, [S.l.], v. 10, n. 1, ago. 2011. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/880>. Acesso em: 28 de fevereiro de 2022.
- BARBIERI, C. **BI2 - Business Intelligence**: modelagem e qualidade. Rio de Janeiro: Elsevier, 2011, 392 p.
- BAZZOTTI, Cristiane; GARCIA, Elias, 2000. A importância do sistema de informação gerencial na gestão empresarial para tomada de decisões, **Ciências Sociais Aplicadas em Revista**, Cascavel, Vol. 6, n 11, 2006. Disponível em: <https://e-revista.unioeste.br/index.php/csaemrevista/article/view/368/279>. Acesso em: 26 de fevereiro de 2022.
- BECKER, João Luiz. 2015. **Estatística básica**: transformando dados em informação. Porto Alegre: Bookman editora, 2015. 502 p.
- BOTELHO, Rigo Fernando; RAZZOLINI FILHO, Edelvino. Conceituando o termo business intelligence: origem e principais objetivos. **Revista Iberoamericana de Sistemas, Cibernética e Informática**, v.1, n.1, 2014. Disponível em: <http://www.iiisci.org/Journal/pdv/risci/pdfs/CB793JN14.pdf>. Acesso em: 26 de fevereiro de 2022.
- CARIBÉ, Rita de Cássia do Vale. Sistema de indicadores: uma introdução. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 7, n. 1, p. 1–23, 2009. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1986>. Acesso em: 4 de agosto de 2021.
- CENDÓN, Beatriz Valadares. Bases de dados de informação para negócios no Brasil. **Ciência Da Informação**, Brasília, v. 32, n. 2, 2003.. Disponível em: <http://revista.ibict.br/ciinf/article/view/1002/1057>. Acesso em: 26 de fevereiro de 2022.

DADOS.GOV. **O que são dados abertos**. Disponível em:
<https://dados.gov.br/pagina/dados-abertos>. Acesso em: 26 de fevereiro de 2022.

DB –ENGINES. **DB- Engines Ranking of Documents Stores**, 2022. Disponível em:
<https://db-engines.com/en/ranking/document+store>. Acesso em: 22 de abril de 2022.

DB –ENGINES. **DB- Engines Ranking of Graph DBMS**, 2022. Disponível em:
<https://db-engines.com/en/ranking/graph+dbms>. Acesso em: 22 de abril de 2022.

DB –ENGINES. **DB- Engines Ranking of Key-values Stores**, 2022. Disponível em:
<https://db-engines.com/en/ranking/key-value+store>. Acesso em: 22 de abril de 2022.

DB –ENGINES. **DB-Engines Ranking of Relation DBMS**, 2022. Disponível em:
<https://db-engines.com/en/ranking/relational+dbms>. Acesso em: 22 de abril de 2022.

DFREIRE COMUNICAÇÃO E NEGÓCIOS. **InterSystems é reconhecida Líder pelo Quadrante Mágico Gartner 2019 para Sistemas de Gerenciamento de Banco de Dados Operacional**, 2019. Disponível em:
<https://paranashop.com.br/2019/12/intersystems-e-reconhecida-lider-pelo-quadrante-magico-gartner-2019-para-sistemas-de-gerenciamento-de-banco-de-dados-operacional/>. Acesso em: 22 de abril de 2022.

DE OLIVEIRA, Jones Dhyemison Quito; RIBEIRO, Leonardo Andrade. **Avaliação de Processos ETL para Análise de Dados usando SGBD Orientado a Grafos**. In: **ESCOLA REGIONAL DE INFORMÁTICA DE GOIÁS (ERI-GO)**, 7, 2019, Goiânia. Porto Alegre: Sociedade Brasileira de Computação, 2019, p. 61-74. Disponível em:
<https://sol.sbc.org.br/index.php/erigo/article/view/9086/8988>. Acesso em: 26 de fevereiro de 2022.

DE TEFFÉ, Chiara Spadaccini; VIOLA, Mario. Tratamento de dados pessoais na LGPD: estudo sobre as bases legais. **Civilística**, Rio de Janeiro, ano 9, n. 1, 2020. Disponível em:
https://www.academia.edu/42993125/Tratamento_de_dados_pessoais_na_LGPD_e_studo_sobre_as_bases_legais. Acesso em 4 de outubro de 2021.

DETONI JUNIOR, Waldecir. Business intelligence: uma revisão de literatura. **Revista Científica Semana Acadêmica**, Fortaleza, v.1, ed. 132, 2018. Disponível em: <https://semanaacademica.org.br/artigo/business-intelligence-uma-revisao-de-literatura>. Acesso em: 28 de fevereiro de 2022.

DUTRA, Claudio Crossetti; LOPES, Karen Maria Gross. **Dados abertos**: uma forma inovadora de transparência. In: IV Congresso CONSAD de Gestão Pública. Brasília, 2013. Disponível em: <http://www.sgc.goias.gov.br/upload/arquivos/2014-09/dados-abertos---uma-forma-inovadora-de-transparEncia.pdf>. Acesso em: 26 de fevereiro de 2022.

FRAGA, Bruna; ERPEN, Julio Graeff; VARVAKIS, Gregório; Santos, Neri dos. Business Intelligence: métodos e técnicas de gestão do conhecimento e as

tendências para avanços do capital intelectual. **Navus**, Florianópolis, v. 7, n. 1, p. 43-56, 2017. Disponível em:

https://www.researchgate.net/publication/312245492_Business_Intelligence_metodos_e_tecnicas_de_gestao_do_conhecimento_e_as_tendencias_para_avancos_do_capital_intelectual. Acesso em: 26 de fevereiro de 2022.

FERRARI, Alberto; RUSSO, Marco. **Introducing Microsoft Power BI**. Germany: Microsoft Press, 2016. 306 p.

FERREIRA, João Tiago Antunes et al. **O processo ETL em sistemas data warehouse**. *In*: INForum, 2010, II Simpósio de Informática, 9-10 de setembro. p. 757-765. Disponível em:

https://www.researchgate.net/publication/265195317_O_Processo_ETL_em_Sistemas_Data_Warehouse Acesso em: 15 de outubro de 2021.

FORTULAN, Marcos Roberto; GONÇALVES FILHO, Eduardo Vila. Uma proposta de aplicação de business intelligence no chão-de-fábrica. **Gestão & Produção**, v. 12, n. 1, 2005 pp. 55-66. Disponível em:

<https://www.scielo.br/j/gp/a/ydtVGxxBtD65zcx4VmJDJGw/?lang=pt#>. Acesso em: 28 de fevereiro de 2022.

GARCIA, Marco. **Data analytics, big data, data science**, 2022. Disponível em: <https://www.cetax.com.br/blog/business-intelligence-tudo-que-voce-precisa-saber/>. Acesso em: 1 de março de 2022.

GLIKAS, Alexandre. **Indústria 4.0**: empresas são desafiadas a aproveitar “novo petróleo”, 2018. Disponível em: <https://inforchannel.com.br/2018/12/20/industria-4-0-empresas-sao-desafiadas-a-aproveitar-novo-petroleo/>. Acesso em: 28 de fevereiro de 2022.

GOMES, Luiz Flavio Autran Monteiro et al. Uma abordagem multicritério para a seleção de ferramentas de business intelligence. **Eletrônica de Sistemas de Informação**, v. 10, n. 2, 2011. Disponível em:

<http://www.periodicosibepes.org.br/index.php/reinfo/article/view/950>. Acesso em: 28 de fevereiro de 2022.

GUIMARÃES, Eliane Marina Palhares; ÉVORA, Yolanda Dora Martinez. Sistema de informação: instrumento para tomada de decisão no exercício da gerência. **Ciência Da Informação**, Brasília, n. 33, 2004. Disponível em:

<https://www.scielo.br/j/ci/a/XyD5tMC7NK7YFCsPgTvQCWM/?lang=pt>. Acesso em: 28 de fevereiro de 2022.

GURBAXANI, V.; WHANG, S. J. **O impacto dos sistemas de informações em organizações e mercados**. Nova York: ACM, 1991.

INFORMATICA. **A Informatica é novamente a líder no Quadrante Mágico do Gartner de para Ferramentas de Integração de Dados**, 2022. Disponível em:

<https://www.informatica.com/pt/data-integration-magic-quadrant.html>. Acesso em: 22 de abril de 2022.

KAUFELD, J. **Access 95 para Windows para leigos**: Um manual para novos usuários. LUDEMIR, J. São Paulo: Berkeley Brasil, 1996. 352 p.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados**: um guia sobre visualização de dados para profissionais de negócios. 2. ed. Rio de Janeiro: Alta Books, 2019. 256 p.

KNOWSOLUTION. **O que é Business Intelligence (BI)?** Disponível em: <https://www.knowsolution.com.br/o-que-e-business-intelligence-bi/>. Acesso em: 1 de março de 2022.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação gerenciais**: administrando a empresa digital. 5. ed. São Paulo: Prentice Hall, 2004.

LE COADIC, Yves François. **A Ciência da Informação**, tradução de Maria Yêda F. S. de Filgueiras Gomes. Brasília: Briquet de Lemos, 1996.

MARCONDES, José Sérgio. **Sistema de informação**: o que é, conceitos e definições, 2020. Disponível em: <https://gestaodesegurancaprivada.com.br/sistema-de-informacao-o-que-e-conceitos/>. Acesso em: 28 de julho de 2020.

MATHEUS, Renato F; PARREIRAS, Fernando F. Inteligência Empresarial Versus *Business Intelligence*: abordagens complementares para apoio à tomada de decisão no Brasil. *In: Congresso Anual da Sociedade Brasileira de Gestão do Conhecimento*, São Paulo, n. 3, 2004. Disponível em: https://zenodo.org/record/12310/files/ie_bi.pdf. Acesso em: 28 de fevereiro de 2022.

MESQUITA, Renato. **O que é sistema de informação e quais são as suas características**, 2018. Disponível em: <https://rockcontent.com/br/blog/sistema-de-informacao/>. Acesso em: 28 de fevereiro de 2022.

MICROSOFT. **É assim que você faz o software**, 2021. Disponível em: <https://visualstudio.microsoft.com/pt-br>. Acesso em: 22 de abril de 2022.

MICROSOFT. **Experimente o SQL Server na infraestrutura local ou na nuvem**, 2022. Disponível em: <https://www.microsoft.com/pt-br/sql-server/sql-server-downloads>. Acesso em: 22 de abril de 2022.

MICROSOFT. **Fazer download de ferramentas e aplicativos do Power BI**, 2022. Disponível em: <https://powerbi.microsoft.com/pt-br/downloads/>. Acesso em: 22 de abril de 2022.

MIRANDA, Roberto Campos da Rocha. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da Informação**, v. 28, n. 3, 1999, pp. 286-292. Disponível em: <<https://doi.org/10.1590/S0100-19651999000300006>>. Acesso em: 28 de fevereiro de 2022.

MORAES, Gustavo H. S. M.; CAPPELLOZZA, Alexandre; MEIRELLES, Fernando S. Será o fim do papel? Os avanços tecnológicos e seus possíveis impactos no

consumo de papel. **Internext – Revista Eletrônica de Negócios Internacionais da ESPM**, São Paulo, v. 6, n. 2, p. 48-65, jul./dez. 2011. Disponível em: https://pesquisa-eaesp.fgv.br/sites/gvpesquisa.fgv.br/files/arquivos/meirelles_-_sera_o_fim_do_papel_os_avancos_tecnologicos_e_seus.pdf. Acesso em: 28 de fevereiro de 2022.

MOREIRA, Fábio Mosso et al. A qualidade na recuperação de dados governamentais: um estudo sobre dados de políticas públicas na internet. **Perspectivas em Ciência da Informação**, v. 25, n. 02, 2020, pp. 103-132. Disponível em: <https://doi.org/10.1590/1981-5344/3994>. Acesso em: 28 de fevereiro de 2022.

MORESI, Eduardo Amadeu Dutra. Delineando o valor do sistema de informação, **Ci. Inf.**, Brasília, v. 29, n. 1, p. 14-24, jan./abr. 2000. Disponível em: <https://www.scielo.br/j/ci/a/pzj7MLqJc6jX5zHLxH5PFwq/?format=pdf&lang=pt>. Acesso em: 26 de fevereiro de 2022.

MOTTA, Paulo Roberto. **Transformação organizacional: a teoria e a prática de inovar**. Rio de Janeiro: Qualitymark, 1998.

NEGASH, Solomon. BUSINESS INTELLIGENCE, **Communications of the Association for Information Systems**, v.13, p. 177-195, 2004. Disponível em: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais>. Acesso em: 26 de fevereiro de 2022.

O'BRIEN, James A; MARAKAS, George M. **Administração de sistemas de informação**. 15 ed. São Paulo: AMGH Editora, 2012.

OLIVEIRA, Lairson; LÓSCIO, Bernadette. Uma abordagem para captura de informações sobre aplicações que fazem uso de dados abertos. **Revista Brasileira de Administração Científica**. Aquidabã, v.5, n.2, out 2014. Disponível em: https://www.researchgate.net/publication/291099159_Uma_abordagem_para_captura_de_informacoes_sobre_aplicacoes_que_fazem_uso_de_dados_abertos. Acesso em: 28 de fevereiro de 2022.

OLIVEIRA, Samuel Silva de. Bancos de dados não-relacionais: um novo paradigma para armazenamento de dados em sistemas de ensino colaborativo, **REVE@P: Revista Eletrônica da Escola de Administração Pública do Amapá, Macapá**, v.2 n. 1, p. 184–194, ago.- dez. 2014. Disponível em: <https://www2.unifap.br/oliveira/files/2016/02/35-124-1-PB.pdf>. Acesso em: 26 de fevereiro de 2022.

PETER, Sayer. **12 principais ferramentas de Business Intelligence em 2019**. EUA, 2018. Disponível em: <https://itforum.com.br/noticias/12-principais-ferramentas-de-business-intelligence-em-2019/>. Acesso em: 26 de fevereiro de 2022.

PIMENTA, R. M. Big data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 2, 2013. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119514>. Acesso em: 28 de fevereiro de 2022.

PINHEIRO, Patricia Peck. **Proteção de Dados Pessoais**: vomentários à Lei n. 13.709/2018-LGPD. 23. ed. São Paulo: Saraiva Jur, 2020. 152 p.

RECEITA FEDERAL. **Dados Públicos CNPJ**, 2018. Disponível em: <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/dados-publicos-cnpj>. Acesso em: 22 de abril de 2022.

RECEITA FEDERAL. **Novo layout dos dados abertos do CNPJ**, 2021. Disponível em: <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/arquivos/NOVOLAYOUTDOSDADOSABERTOSDOCNPJ.pdf>. Acesso em: 22 de abril de 2022.

RODRIGUES FILHO, José; LUDMER, Gilson. Sistema de informação: que ciência é essa? **JISTEM - Journal of Information Systems and Technology Management**, v. 2, n. 2, p. 151-166, 2005.. Disponível em: <https://www.scielo.br/j/jistm/a/V4mV4rpjqkYJTFK7BBfghbc/?format=pdf&lang=pt>. Acesso em: 28 de fevereiro de 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação Informação e Inovação em Saúde**, Rio de Janeiro, v. 8, n. 2, 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>. Acesso em: 26 de fevereiro de 2022.

SEMIDÃO, Rafael Aparecido Moron. **Dados, informação e conhecimento enquanto elementos de compreensão do universo conceitual da ciência da informação**: contribuições teóricas. 2014. 198 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências de Marília, 2014. Disponível em: <http://hdl.handle.net/11449/110783>. Acesso em: 28 de fevereiro de 2022.

SETZER, de Valdemar W. **Dado, Informação, Conhecimento e Competência**. Disponível em: <https://www.ime.usp.br/~vwsetzer/dado-info.html>. Acesso em: 28 de fevereiro de 2022.

STEPHEN P. Robbins; COULTER, Mary. **Administração**. 5. ed.. Prentice: Hall Interamericana, 1996.

SILVA, Rafaela Alexandre da; Silva, Fernando Cesar Almeida; Gomes, Carlos Francisco Simões. O uso do business intelligence (bi) em sistema de apoio à tomada de decisão estratégica. **Revista geintec-gestao inovacao e tecnologias**, v. 6, n.1, 2016. Disponível em: <https://www.revistageintec.net/index.php/revista/article/view/726/630>. Acesso em: 28 de fevereiro de 2022.

SILVA FILHO, Demóstenes Ferreira da et al. Banco de dados relacional para cadastro, avaliação e manejo da arborização em vias públicas. **Revista Árvore**, v. 26, n. 5, pp. 629-642, 2002. Disponível em: <<https://doi.org/10.1590/S0100-67622002000500014>>. Acesso em: 26 de fevereiro de 2022.

TIME SMART. **Quadrante Mágico Gartner Group 2021 BI e Analytics**, 2021. Disponível em: <https://blog.smartconsulting.com.br/quadrante-magico-gartner-2021-melhores-solucoes-bi-e-analytics/>. Acesso em: 28 de fevereiro de 2022.

TURBAN, E., & VOLONIMO, L. **Business Intelligence e Suporte à Decisão**. Porto Alegre: Bookman, 2013.

VAIDHYANATHAN, Siva. **A Googlelização de Tudo (e por que devemos nos preocupar)**: a ameaça do controle total da informação por meio da maior e mais bem-sucedida empresa do mundo virtual. São Paulo: Cultrix, 2011.

WOEBCKEN, Cayo. **Quadrante mágico Gartner**: o que é e qual a aplicabilidade? Disponível em: <https://rockcontent.com/br/blog/quadrante-magico-gartner/>. Acesso em: 26 de fevereiro de 2022.