

DATA WAREHOUSE: ANÁLISE SOBRE O COVID-19 NA PARAIBA.

*ALCÂNTARA, Cícero Moura de
** MEDEIROS, Fábio Nicácio de

* Graduando em Sistemas para Internet

** Orientador

1. INTRODUÇÃO

Com o passar dos anos e com o crescimento da internet, os grandes volumes de dados ficam espalhados pela rede de forma desorganizada e muitas vezes não significam nada, a implementação de um *Data Warehouse* (DW) e criação de um *dashboard*, visa a organização desses dados e formas de consultas muito mais eficientes e elaboradas. Além do mais, uma característica do DW é a criação de um *dataset* sobre um assunto utilizando fontes de dados distintos, essa característica é fundamental para a elaboração de análises mais detalhadas e a obtenção de informações mais robustas. (ROSSI, 2017)

A implementação do DW pode ser uma etapa complexa a ser executada, pois o número de registros analisados é muito alto e aumenta conforme o DW é atualizado, um segmento muito importante do projeto de modelagem de um DW que consome cerca de 70% do tempo total do projeto, é a fase de extração, tratamento e limpeza (ETL) dos diversos dados coletados. Nessa fase, inicialmente, os dados são coletados das suas fontes, em seguida os dados são padronizados e por último eles são transformados, essa última etapa consiste na exclusão de valores inválidos, chaves repetidas ou registros com tipos de erro variados. (UNO, 2015) (COLAÇO JÚNIOR, 2004)

Dentro dessa lógica, o DW é capaz de armazenar dados que retratam o passado e dados que retratam o presente, isso é um ponto positivo no que diz respeito a comparações entre diferentes linhas temporais que os dados representam e também sendo capaz de auxiliar na descoberta de novas tendências. Diante desses pontos, as organizações que utilizam do DW, para lhe auxiliar na tomada de decisão, se destaca dentro do setor comercial ou científico. (UNO, 2015)

Academicamente, este artigo tem como função complementar estudos extracurriculares, que podem abranger desde modelagem de dados, normalização de

dados, mineração de dados, segurança e até a implementação de *dashboards* interativas para leitura dos dados coletados.

Por causa do cenário de pandemia e de todos os benefícios aqui discutidos sobre a utilização de um DW, é possível levantar o seguinte questionamento: como projetar um DW para auxiliar a entender o avanço do Covid-19 no estado da Paraíba e como criar *dashboards* para leitura desses dados?

De acordo com o problema apresentado, este artigo tem como objetivo geral realizar um projeto de implementação de um DW e criação de um *dashboard* com os dados do Covid-19 no estado da Paraíba.

Com isso, para atender a necessidade do objetivo geral, foi necessário fragmentá-lo nos seguintes objetivos específicos:

- a) Pesquisar e coletar os dados relacionado ao Covid-19 no estado da Paraíba;
- b) Definir as etapas de um projeto de DW;
- c) Implementação de um DW;
- d) Criação e disponibilização do *dashboard*.

A fim de solucionar os objetivos propostos, foram feitas pesquisas dos dados referentes ao estado da Paraíba na internet, a fonte de coleta são *datasets* do governo estadual.

Com isso, o artigo está dividido em 5 capítulos, sendo a introdução o primeiro, e os demais seguindo uma sequência uma lógica de escrita.

O segundo capítulo, é feito a partir da base de estudos sobre a modelagem e implementação de um DW e conceitos de *Business Intelligence* utilizando os dados coletados dos portais de transparência do governo do estado da Paraíba e abordaremos as técnicas utilizadas para criação de *dashboard* interativo.

No terceiro capítulo, é abordado a metodologia, onde será apresentado os métodos utilizados para nortear a criação e solução dos problemas relatados no presente artigo.

No quarto capítulo, serão apresentados os resultados e serão feitas discussões a partir da resolução do problema levantado, com isso serão exibidas as informações coletadas de forma detalhada.

No quinto capítulo, será abordado as considerações finais onde serão apresentados os resultados obtidos durante a pesquisa e criação do artigo.

Por fim, concluímos que o objetivo principal do artigo foi atendido, com base nisso foi observado que a implementação de DW, nos mais diversos ambientes, é de

suma importância para o avanço de pesquisas, tomadas de decisões mais embasadas e de agilidade. Diante do cenário de pandemia se conclui que a implementação de DW pode ajudar a salvar vidas.

2. BUSINESS INTELLIGENCE

Por volta da década de 1990 surge o conceito de *Business Intelligence* (BI), de início foi pautado pelo *Gartner Group*, que tem como finalidade ajudar na gestão de negócios por meio da utilização de aplicações, arquitetura de base de dados mais robustas e ferramentas. A integração dessas tecnologias irá auxiliar na obtenção e na análise dos dados de forma mais otimizada e eficiente, dessa forma a tomada de decisão ficou mais clara e objetiva. (SILVA; SILVA; GOMES, 2016)

Diante disso, o BI tem o propósito de coletar e armazenar os dados de diferentes base de dados, sejam elas homogêneas ou heterogêneas, em seguida, transformar esses dados com o objetivo de que as organizações possam analisar e retirar informações concretas com base em fatos reais, gerando estratégias mais embasadas, auxiliando nos processos de tomadas de decisões, agregando no valor de mercado e na competitividade da organização. (PATRICIO; STÁBILE; TÓFOLI, 2016)

De acordo com essa narrativa podemos listar como principais objetivos do BI: (TURBAN; SHARDA; ARONSON; KING, 2009)

- a) Permitir o acesso interativo aos dados, esse acesso pode ser feito em tempo real ou não;
- b) Proporcionar a manipulação dos dados coletados;
- c) Fornecer aos gestores e analistas a capacidade de realizar uma análise adequada.

2.1 DATA WAREHOUSE

Como o próprio nome diz, *Data Warehouse* é um armazém de dados que integra um grande volume de dados relacionados a um assunto em específico, normalmente os dados que compõe o DW estão espalhados pela rede ou dentro das organizações de forma que não geram nenhuma informação. (SILVA; BRUSCHI; SILVA; TORRES, 2018)

Um dos objetivos de um DW é a capacidade de integrar dados históricos, esses dados retratam o passado do fato analisado, uma vez que os dados históricos estão

integrados no DW é possível analisar os números da organização e gerar indicadores em um espaço de tempo determinado. A análise dos dados de forma temporal é eficaz na capacidade de auxiliar na descoberta de novas tendências e comportamentos dentro de uma organização. (SILVA; BRUSCHI; SILVA; TORRES, 2018)

Com isso podemos listar 4 características básicas de um DW: (INMON, 1997)

- a) Baseado em assuntos: Capacidade de integrar dados sobre assuntos específicos e que são importantes para a organização;
- b) Integrado: Os dados passam de um cenário operacional que tem como base os sistemas da organização para o DW, o processo de carga dos dados ocorre de forma em que as inconstâncias dos dados vindos dos sistemas sejam removidas;
- c) Não-volátil: Em um sistema comum os dados podem ser atualizados normalmente, num ambiente de um DW os dados são carregados em grandes volumes e logo em seguida são acessados, essa técnica é chamada de "*load-and-access*", entretanto a atualização dos dados não ocorre dentro do DW;
- d) Variável em relação ao tempo: O espaço de tempo em que se consegue trabalhar com um DW é maior comparado com um sistema comum. Normalmente os sistemas comuns trabalham com intervalos de tempo de 60 a 90 dias comparado com o intervalo de tempo trabalhado pelo DW que é de 5 a 10 anos. Como os dados do DW não são atualizados caso ocorra uma nova mudança é necessário a criação de uma nova entrada no DW para marcar essa mudança, diferente do sistema comum em que as entradas são atualizadas constantemente e retratam o objeto no momento do acesso ao sistema.

2.2 MODELAGEM DIMENCIONAL

Com o avanço da tecnologia e a busca por resultados cada vez melhores nas organizações, o conceito de modelagem dimensional surge como uma arquitetura de projeto lógico, muito utilizado em DW, e tem como objetivo melhorar o desempenho dos bancos de dados focados em consulta, tendo como base um conjunto de eventos básicos de medições. (BORGES, 2017)

Nesse tipo de arquitetura, a modelagem dimensional ocorre de forma diferente da que é utilizada no modelo relacional, uma vez que o foco desse tipo de modelagem é auxiliar na busca de resultados rápidos para ajudar na tomada de decisões. (SILVA; SARTORI, 2015)

Nós podemos observar a seguir, na Tabela 1, a comparação entre o modelo dimensional e o modelo relacional.

Tabela 1: Comparação entre o modelo relacional e o modelo dimensional.

Modelo Dimensional	Modelo Relacional
Padrão de estrutura mais fácil e intuitiva	Modelo mais complexo
Anterior ao MER, anos 60	Ênfase em Bancos de Dados Relacionais, anos 70
Tabelas Fato e tabelas Dimensão	Tabelas que representam Dados e Relacionamentos
Tabelas Fato são o núcleo – normalizadas	Todas as tabelas são comumente normalizadas
Tabelas Dimensão são os pontos de entrada	As tabelas são indistintamente acessadas e de filtro inicial
Tabelas Dimensão opcionalmente normalizadas	Todas as tabelas são comumente normalizadas
Facilidade de “joined”	Maior dificuldade de “join” pelo número maior de tabelas
Leitura mais fácil do modelo por usuários não especializados	Maior dificuldade de leitura pelo usuário não especializado

Fonte: (BARBIERI, 2011)

A partir da integração dos dados, na técnica de modelagem dimensional, é formado um modelo em que a informação é encontrada na interseção de várias dimensões ou pontos de um cubo. As dimensões ou pontos do cubo, são os locais onde as medições são encontradas e são nelas onde ocorrem as combinações (Figura 1), como por exemplo, de produto, região e tempo. (BARBIERI, 2011)

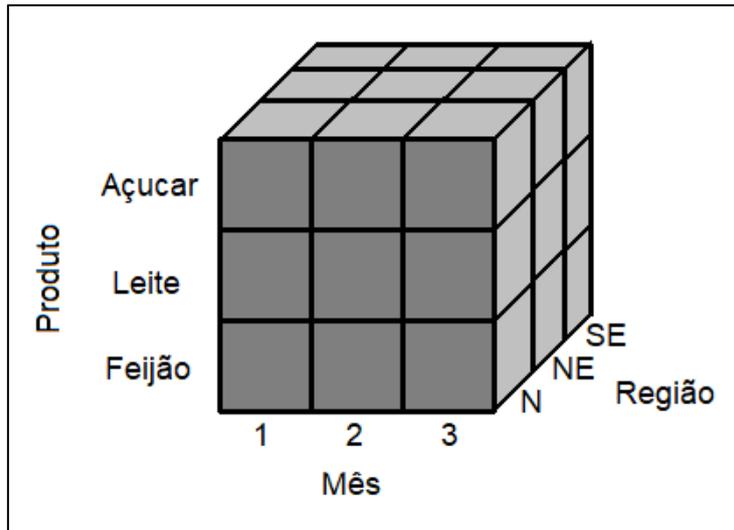


Figura 1: Cubo dimensional

Fonte: (AUTOR, 2020)

O modelo multidimensional é formado com base em três elementos básicos:

- a) Tabela Fato;
- b) Tabela Dimensão;
- c) Medidas.

2.2.1 TABELA FATO

A tabela fato (Figura 2) é a tabela que fica no centro de todo o modelo dimensional, ela tem o papel de integrar os dados das tabelas de dimensão.

Essas tabelas se interligam usando as chaves estrangeiras (FK) das tabelas de dimensão e tem como objetivo associar os dados a um evento que vai ser analisado. Como a finalidade da tabela fato é a análise, muitas vezes essa tabela é composta por atributos numéricos ou métricos das tabelas dimensão. (REIS; TEIXEIRA; ARAÚJO, 2009)

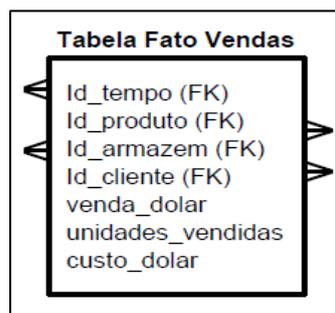


Figura 2: Tabela fato

Fonte: (WAGNER, 2012)

2.2.2 TABELA DIMENSÃO

A tabela dimensão (Figura 3) representa uma parte do fato ou assunto que está sendo analisado, é nela que contém as características do evento abordado. Cada dimensão integra uma ou mais tabelas fato, essa integração acontece a partir de uma chave primária (PK). Esse tipo de tabela oferece uma grande quantidade de registros, auxiliando na análise dos dados. (REIS; TEIXEIRA; ARAÚJO, 2009)



Figura 3: Tabela dimensão

Fonte: (WAGNER, 2012)

Dentro da modelagem dimensional, esse tipo de tabela pode ser implementado de duas formas:

- a) Modelo estrela;
- b) Modelo Floco de Neve.

2.2.3 MODELO ESTRELA

O esquema de estrela (Figura 4) é a estrutura mais básica da modelagem dimensional. Sua composição é dada de forma simples, onde a tabela fato fica ao centro do esquema e as tabelas de dimensões ficam ao redor, formando uma estrela. (REIS, 2018)

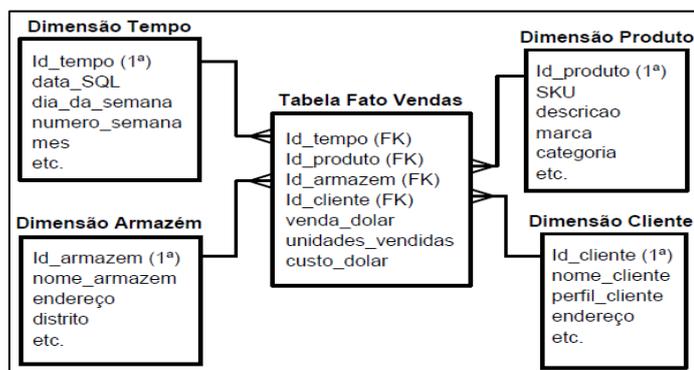


Figura 4: Esquema de estrela

Fonte: (WAGNER, 2012)

A simplicidade do modelo proporciona alguns benefícios em relação ao modelo entidade – relacionamento, como, por exemplo, um desempenho melhor nas consultas, na flexibilidade caso ocorra a necessidade de ser efetuadas mudanças, e a granularidade maior proporciona uma capacidade de dimensionamento maior. (BORGES, 2017)

2.2.4 MODELO FLOCO DE NEVE

Esse tipo de modelo (Figura 5) é eficiente para a remoção de redundâncias na tabela de dimensão, em virtude disso é criada uma tabela de dimensão secundária que passa pelo processo de normalização, essas tabelas são ligadas através de uma chave artificial. (CORREIA, 2017)

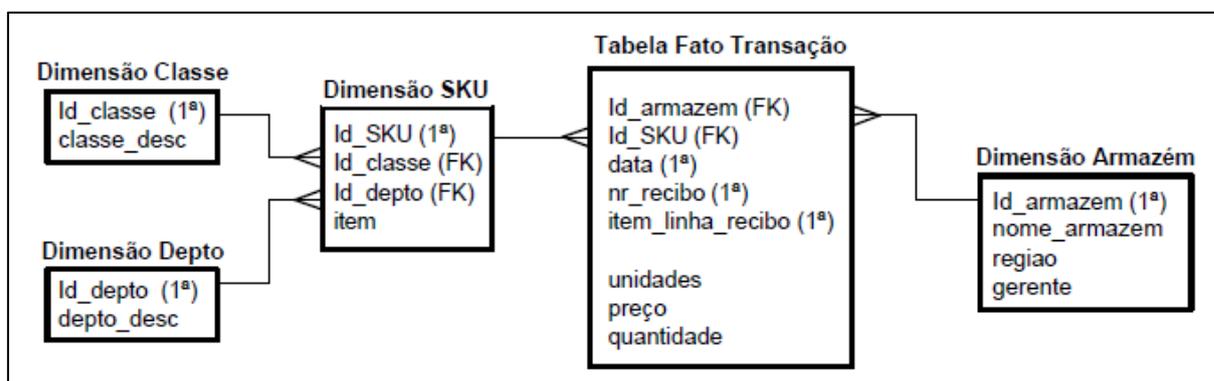


Figura 5: Modelo floco de neve

Fonte: (WAGNER, 2012)

Esse tipo de modelo não é recomendado para ser utilizado em DW por ser uma modelagem mais complexa e que se aproxima muito de uma modelagem entidade - relacionamento, onde a quantidade de tabelas é maior e conseqüentemente aumentará o número de *JOINS* nas consultas, tornando o processo de análise mais lento. (BORGES, 2017)

2.2.5 MEDIDAS

Os atributos numéricos que representam um fato são chamados de medida ou métrica, são exemplos dessas medidas: o valor em reais de uma venda, o número de unidades vendidas, estoque, custo de venda, custo de fabricação, imposto, desconto, entre outros. As medidas são criadas a partir da combinação das tabelas de dimensão

e estão integradas na tabela fato como atributos. As medidas podem ser classificadas em dois grupos: (REIS, 2018)

- a) Valores aditivos: São medidas que podem ser usadas em operações de subtração, soma e média.
- b) Valores não aditivos: São os valores que não podem ser manipulados, como, por exemplo, valores percentuais.

2.3 ETL

O processo de extração, transformação e carga (ETL) é uma etapa importante dentro de um projeto de DW e consome em média 70% do tempo do projeto. ETL tem como objetivo extrair os dados de qualquer fonte, transformar os dados de acordo com o banco de dados e carregar os dados limpos no DW (Figura 6), esses dados podem ser oriundos de qualquer fonte de dados, como por exemplo, aplicações de *mainframe*, arquivos de texto, planilha do Excel, entre outros. (TURBAN; SHARDA; ARONSON; KING, 2009)

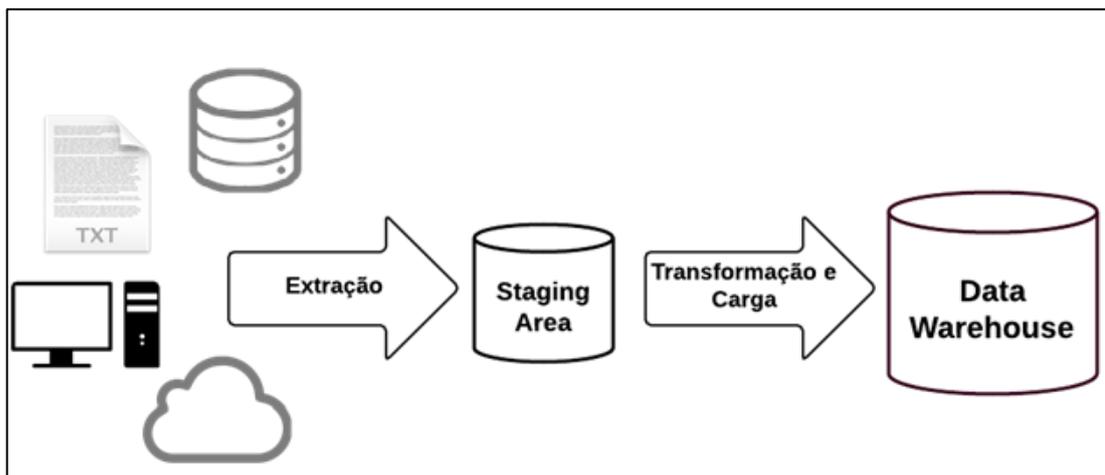


Figura 6: Etapas do processo de ETL

Fonte: (LYRA, 2016)

O processo de ETL consiste na execução de 3 etapas básicas: (LYRA, 2016)

- a) Extração: É a primeira etapa do processo de ETL, é onde ocorre a coleta dos dados dos locais de origem, esses dados são colocados na área de preparação (*staging area*).
- b) Transformação: A segunda etapa ocorre na área de preparação, onde ocorre a transformação e limpeza dos dados com o intuito de melhorar a qualidade do dado.

- c) Carga: A última etapa é onde ocorre a carga do DW com os dados que vieram da área de preparação. Dependendo da estratégia de BI a carga desses dados podem variar, os dados podem ser atualizados semanalmente ou a cada meia hora.

3. METODOLOGIA

O procedimento metodológico adotado para o desenvolvimento deste artigo científico consistiu nos seguintes itens: tipo de pesquisa; instrumento da pesquisa; análise dos dados; e descrição dos resultados obtidos. A pesquisa, através de métodos científicos, busca encontrar soluções para o problema proposto: como projetar um DW para auxiliar a entender o avanço do Covid-19 no estado da Paraíba e como criar *dashboards* para leitura desses dados. Desse modo, se fez necessário uma pesquisa exploratória. Para tal, a investigação teórica foi realizada a partir de consultas em artigos de periódicos nacionais, sendo complementada pela busca em livros e teses.

O desenvolvimento da proposta de DW se deu no desenvolvimento de quatro etapas:

- a) Pesquisar as fontes de dados: Os dados foram consultados em sites do governo do estado da Paraíba;
- b) Modelagem das tabelas: Nesta etapa são criadas as tabelas de dimensão e a tabela fato;
- c) Inserção dos dados nas tabelas: Com as tabelas de dimensão e fato criadas a próxima etapa é a inserção dos registros dentro das tabelas;
- d) Análise dos dados: Por fim serão feitas as análises a partir do desenvolvimento de um *dashboard*, que será manipulado com os dados extraídos do DW.

4. DESENVOLVIMENTO

Neste capítulo, será detalhado as etapas de criação do projeto de DW e a criação do *dashboard* para análise, será apresentado também as fontes dos dados, os códigos em *Transact-SQL* (T-SQL), as tabelas de dimensão, tabela fato e apresentação do *dashboard*.

4.1 FONTES DE DADOS

Os dados que foram selecionados para compor o DW são oriundos do site do governo do estado da Paraíba (Figura 7), a princípio, os dados estão armazenados na web para consulta através de arquivos com extensão CSV.

Data do Óbito	Sexo	Idade	Método	Início Sintomas	Município de Residência	Doenças preexistentes
2020-05-06	Masculino	50	Teste Rápido	2020-04-20	João Pessoa	Diabetes Mellitus, Doença Hepática
2020-06-12	Masculino	39	Swab	2020-06-03	Pedras de Fogo	Sem comorbidades
2020-06-23	Masculino	46	Teste Rápido	2020-06-21	Alhandra	Diabetes Mellitus, Doença Renal, Doença Hepática
2020-05-29	Masculino	68	Teste Rápido	2020-05-16	João Pessoa	Hipertensão, Diabetes Mellitus, Doença Neurológica
2020-06-04	Feminino	78	Teste Rápido	2020-05-25	Itapororoca	Cardiopatia, Obesidade
2020-05-19	Feminino	82	Swab	2020-05-15	Pitimbu	Hipertensão, Diabetes Mellitus
2020-05-17	Masculino	86	Swab	2020-05-05	Campina Grande	Sem comorbidades
2020-05-13	Feminino	42	Swab	2020-05-04	Areal	Diabetes Mellitus, Doença Respiratória
2020-04-13	Masculino	42	Swab	2020-04-07	João Pessoa	Imunossupressão
2020-06-26	Masculino	42	Teste Rápido	2020-06-08	Campina Grande	Obesidade
2020-05-08	Masculino	24	Swab	2020-05-22	João Pessoa	Imunossupressão
2020-06-15	Masculino	79	Swab	2020-06-07	Cachoeira dos Índios	Cardiopatia

Figura 7: Site do governo do estado

Fonte: (TARGET, 2020)

A coleta dos dados ocorreu em dois links, são eles:

a) Dados dos óbitos que ocorreram no estado da Paraíba (Figura 8):

https://superset.plataformatarget.com.br/superset/explore_json/?form_data=%7B%22slice_id%22%3A1549%7D&csv=true

	A	B	C	D	E	F	G
1	Data do Óbito	Sexo	Idade	Método	Início Sintomas	Município de Residência	Doenças preexistentes
2	06/05/2020	Masculino	50	Teste Rápido	20/04/2020	João Pessoa	Diabetes Mellitus, Doença Hepática
3	12/06/2020	Masculino	39	Swab	03/06/2020	Pedras de Fogo	Sem comorbidades
4	23/06/2020	Masculino	46	Teste Rápido	21/06/2020	Alhandra	Diabetes Mellitus, Doença Renal, Doença Hepática
5	29/05/2020	Masculino	68	Teste Rápido	16/05/2020	João Pessoa	Hipertensão, Diabetes Mellitus, Doença Neurológica
6	04/06/2020	Feminino	78	Teste Rápido	25/05/2020	Itapororoca	Cardiopatia, Obesidade
7	19/05/2020	Feminino	82	Swab	15/05/2020	Pitimbu	Hipertensão, Diabetes Mellitus
8	17/05/2020	Masculino	86	Swab	05/05/2020	Campina Grande	Sem comorbidades
9	13/05/2020	Feminino	42	Swab	04/05/2020	Areal	Diabetes Mellitus, Doença Respiratória
10	13/04/2020	Masculino	42	Swab	07/04/2020	João Pessoa	Imunossupressão
11	26/06/2020	Masculino	42	Teste Rápido	08/06/2020	Campina Grande	Obesidade
12	08/06/2020	Masculino	24	Swab	22/05/2020	João Pessoa	Imunossupressão
13	15/06/2020	Masculino	79	Swab	07/06/2020	Cachoeira dos Índios	Cardiopatia

Figura 8: Arquivo CSV com os óbitos que ocorreram na Paraíba

Fonte: (AUTOR, 2020)

De início, para o arquivo CSV de óbitos na Paraíba (Figura 8), foi feita a transformação dos valores para o padrão Unicode UTF-8, esse tipo de problema ocorre porque o Excel não consegue ler o arquivo CSV no padrão UTF-8 automaticamente. Nessa base de dados pode se observar a data do óbito, o gênero, idade, tipo do teste que foi feito no paciente, início dos sintomas, o município de residência e as doenças preexistentes do paciente.

b) Dados referentes aos boletins diários do estado da Paraíba (Figura 9):

https://superset.plataformatarget.com.br/superset/explore_json/?form_data=%7B%22slice_id%22%3A1550%7D&csv=true

	A	B	C	D	E	F	G	H
1	data	casosAcumulados	casosNovos	descartados	recuperados	obitosAcumulados	obitosNovos	Letalidade
2	17/10/2020	128416	384	176960	104493	2985	8	0.0232
3	16/10/2020	128032	576	176543	104015	2977	7	0.0233
4	15/10/2020	127456	665	175570	103487	2970	12	0.0233
5	14/10/2020	126791	718	174661	102755	2958	14	0.0233
6	13/10/2020	126073	140	173432	101794	2944	14	0.0234
7	12/10/2020	125933	56	172947	101388	2930	8	0.0233
8	11/10/2020	125877	164	172947	101212	2922	8	0.0232
9	10/10/2020	125713	428	172874	101073	2914	10	0.0232
10	09/10/2020	125285	536	172333	100970	2904	10	0.0232
11	08/10/2020	124749	434	171611	100334	2894	10	0.0232

Figura 9: Arquivo CSV com os acumulados de óbitos e recuperados

Fonte: (AUTOR, 2020)

Nessa base de dados (Figura 9) não foi necessário fazer alguma transformação nos valores e pode ser observado os dados por dia, desde o início da coleta dos dados da pandemia, são detalhados dados referentes ao número de casos, número de óbitos, casos descartados, casos recuperados e letalidade.

4.2 MODELAGEM DAS TABELAS

Nesta etapa foram modeladas as tabelas de dimensões e tabela fato. A princípio foi modelado a tabela 'dbo.obitosParaiba' (Figura 10) com a função de detalhar cada óbito que ocorreu no estado da Paraíba e tem os seguintes campos: id_obitosParaiba (*Primary Key* ou PK), dataObito, gênero, idade, tipoTeste, inicioSintomas, municipioResidencia e doencasPreexistentes.

```
3 create table dbo.obitosParaiba (  
4     id_obitosParaiba int not null identity PRIMARY KEY,  
5     dataObito date,  
6     genero varchar(50),  
7     idade int,  
8     tipoTeste varchar(50),  
9     inicioSintomas date,  
10    municipioResidencia varchar(100),  
11    doencasPreexistentes varchar(300)  
12 );
```

Figura 10: Script de criação da tabela 'dbo.obitosParaiba'

Fonte: (AUTOR, 2020)

Em seguida foi modelado a tabela 'dbo.boletinsDiarioParaiba' (Figura 11) detalhando o total de casos por dia na Paraíba e possui os seguintes campos: dataBoletim (PK), casosAcumulados, casosNovos, descartados, recuperados, obitosAcumulados, obitosNovos e letalidade.

```
3 create table dbo.boletimDiarioParaiba (  
4     data smalldatetime PRIMARY KEY,  
5     casosAcumulados int,  
6     casosNovos int,  
7     descartados int,  
8     recuperados int,  
9     obitosAcumulados int,  
10    obitosNovos int,  
11    letalidade varchar(50)  
12 );
```

Figura 11: Script de criação da tabela 'dbo.boletimDiarioParaiba'

Fonte: (AUTOR, 2020)

Para fazer relação com o tempo foi necessário criar a dimensão de tempo (Figura 12) 'dbo.dim_tempo' com os seguintes campos: Data (PK), Ano, Dia, MesNome e MesNumero.

```
3 CREATE TABLE [dim_tempo](  
4     [Data] [smalldatetime] NOT NULL Primary Key,  
5     [Ano] [smallint] NOT NULL,  
6     [Dia] [smallint] NOT NULL,  
7     [MesNome] [varchar](10) NOT NULL,  
8     [MesNumero] [smallint] NOT NULL,  
9 )
```

Figura 12: Script de criação da tabela 'dbo.dim_tempo'

Fonte: (HENRIQUE; VAN BILJON, 2016)

E por fim foi modelado a tabela fato 'dbo.fatoCovidParaiba' (Figura 13) com os seguintes campos: Data (Foreign Key- FK) fazendo referência a tabela 'dbo.dim_tempo', dataBoletim (FK) fazendo referência a tabela 'dbo.boletimDiarioParaiba', id_obitosParaiba (FK) fazendo referência a tabela 'dbo.obitosParaiba', obitosNovos, casosNovos, letalidade e idade.

```

1 create table dbo.fatoCovidParaiba (
2     Data smalldatetime references dbo.dim_tempo(Data),
3     dataBoletin smalldatetime references dbo.boletinDiarioParaiba(dataBoletin),
4     id_obitosParaiba int references dbo.obitosParaiba(id_obitosParaiba),
5     obitosNovos int,
6     casosNovos int,
7     letalidade varchar(50),
8     idade int
9 );

```

Figura 13: Script de criação da tabela 'dbo.fatoCovidParaiba'

Fonte: (AUTOR, 2020)

Com a criação das tabelas de dimensões e tabela fato, foi gerado, no SQL Server, o diagrama de entidades (Figura 14) e se observou que o modelo dimensional que foi criado se assemelha com o modelo de estrela, esse possui uma tabela fato no centro e as tabelas de dimensão ao redor da tabela fato.

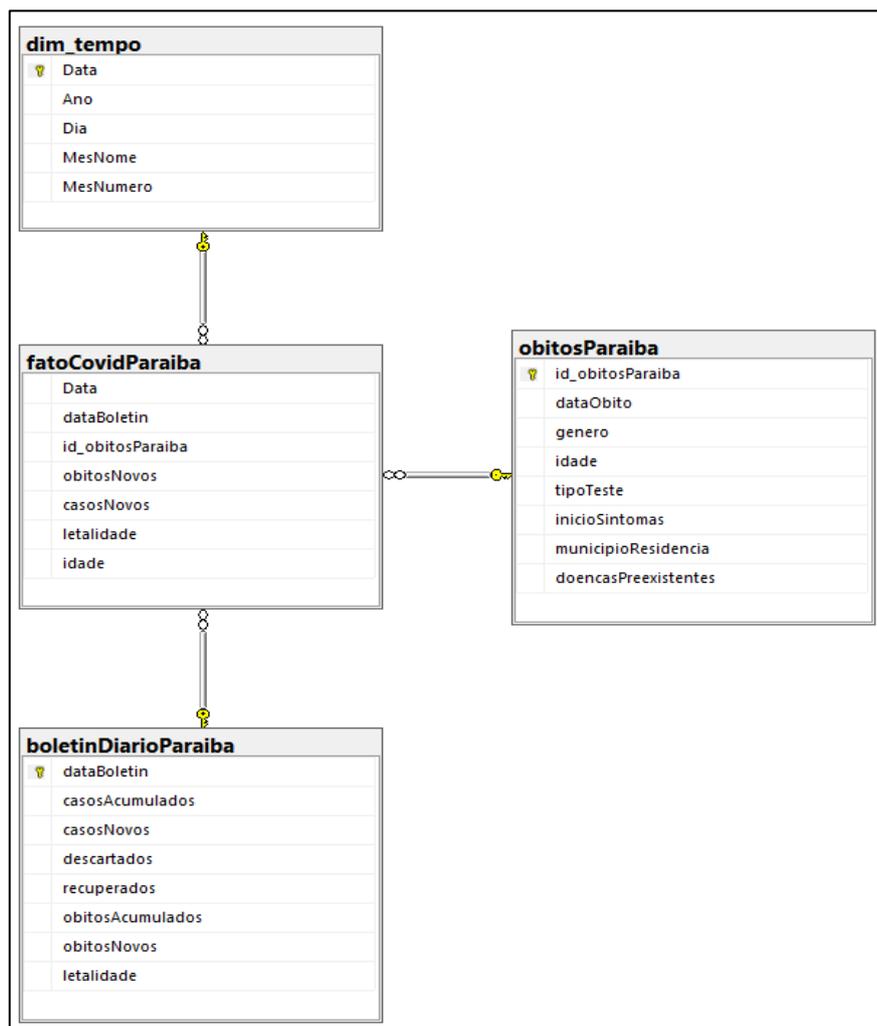


Figura 14: Diagrama de entidades

Fonte: (AUTOR, 2020)

4.3 INSERÇÃO DOS DADOS NAS TABELAS

A inserção dos dados nas tabelas 'dbo.obitosParaiba' e 'dbo.boletinDiarioParaiba' se deu com o auxílio do assistente de importação de dados do SQL-Server, a princípio é selecionado o arquivo do Excel com os dados que se quer importar e a tabela que será inserido os dados, feito isso todos os registros do arquivo CSV serão adicionados as tabelas.

Diferente da tabela 'dbo.dim_tempo' em que os registros foram inseridos a partir de um código em T-SQL (Figura 15), é nesse momento em que é escolhido o intervalo de tempo em que os dados serão analisados.

```
11 SET LANGUAGE Brazilian
12
13 DECLARE @Ano smallint
14 DECLARE @Data smalldatetime
15 DECLARE @DataIni smalldatetime
16 DECLARE @DataFim smalldatetime
17 DECLARE @Dia smallint
18 DECLARE @MesNome varchar(10)
19 DECLARE @MesNumero smallint
20
21 SET @DataIni='16/03/2020'
22 SET @DataFim='17/10/2020'
23
24 SET @Data=@DataIni
25 While @Data<=@DataFim
26 Begin
27     Set @Dia = day(@Data)
28     Set @MesNome = datename(month,@Data)
29     Set @MesNumero = month(@Data)
30     Set @Ano = YEAR(@Data)
31     INSERT INTO dim_tempo values(@Data, @Ano, @Dia, @MesNome, @MesNumero)
32     Set @Data=dateadd(day,1,@Data)
33 End
```

Figura 15: Script de inserção dos dados na tabela 'dbo.dim_tempo'

Fonte: (HENRIQUE; VAN BILJON, 2016)

Analisando o código da figura 15, na linha 11 foi atribuído a linguagem dos dados de datas, nas linhas 21 e 22 foi atribuído o intervalo de tempo em que os dados serão analisados de acordo com dados que retratam o passado, no caso o intervalo de março a outubro, e por fim a partir da linha 24 as datas foram inseridas.

```

11 insert dbo.fatoCovidParaiba(Data, dataBoletim, id_obitosParaiba, obitosNovos, casosNovos, letalidade, idade)
12 select dbo.dim_tempo.Data, dbo.boletimDiarioParaiba.dataBoletim, dbo.obitosParaiba.id_obitosParaiba,
13        dbo.boletimDiarioParaiba.obitosNovos, dbo.boletimDiarioParaiba.casosNovos,
14        dbo.boletimDiarioParaiba.letalidade, dbo.obitosParaiba.idade
15 from dbo.dim_tempo
16 inner join dbo.boletimDiarioParaiba
17 on dbo.boletimDiarioParaiba.dataBoletim = dbo.dim_tempo.Data
18 inner join dbo.obitosParaiba
19 on dbo.obitosParaiba.dataObito = dbo.boletimDiarioParaiba.dataBoletim
20 group by Data, dataBoletim, id_obitosParaiba, obitosNovos, casosNovos, letalidade, idade;
21

```

Figura 16: Script inserção dados na tabela 'dbo.fatoCovidParaiba'

Fonte: (AUTOR, 2020)

E por fim foi populado a tabela de fato através do seguinte código T-SQL (Figura 16).

Para a inserção dos dados na tabela fato foram utilizados *joins* entre todas as tabelas do modelo, gerando assim a população de 2.985 registros referentes aos óbitos no estado da Paraíba.

4.4 ANÁLISE DOS DADOS

Nessa seção será apresentado uma análise dinâmica dos dados que foram coletados, para facilitar o entendimento foram criados *dashboards* no *Qlikdview* como podemos observar a seguir.

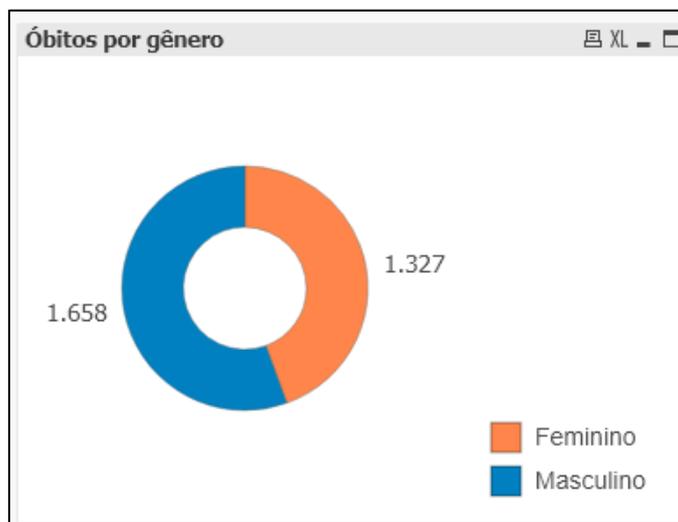


Figura 17: Óbitos por gênero

Fonte: (AUTOR, 2020)

Na Figura 17 contém os dados do número de óbitos por gênero, masculino e feminino, e se observa que o maior índice de óbitos se dá em indivíduos do gênero masculino com uma diferença de aproximadamente 300 óbitos.

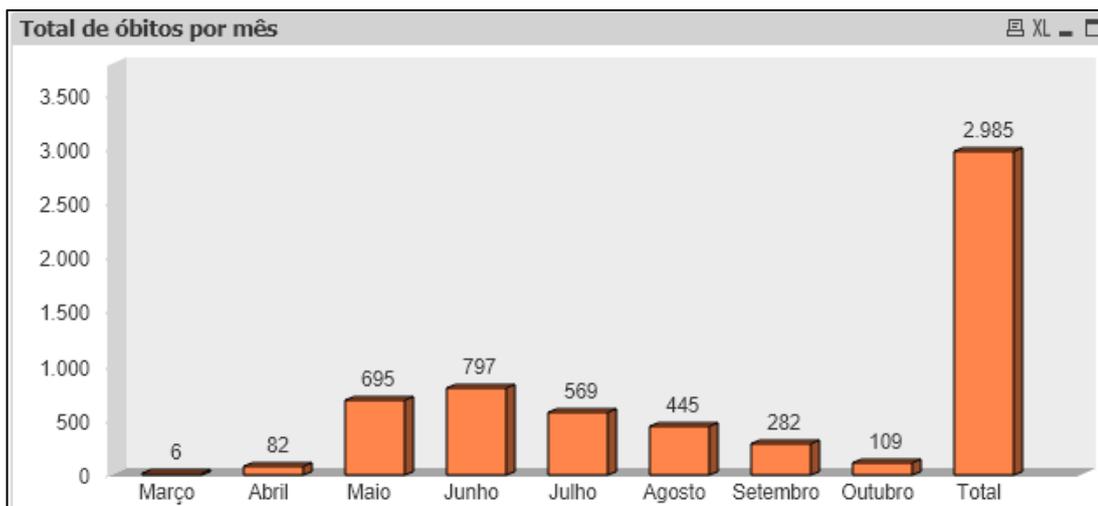


Figura 18: Total de óbitos por mês.

Fonte: (AUTOR, 2020)

A Figura 18 representa o total de óbitos por mês e o seu total que é a soma de todos os meses analisados, pode se observar que o pico de óbitos se deu no mês de junho com 797 óbitos, a partir desse mês os óbitos tendem a cair, uma vez que foram adotadas medidas de contenção do vírus, chegando a 282 óbitos no mês de setembro.

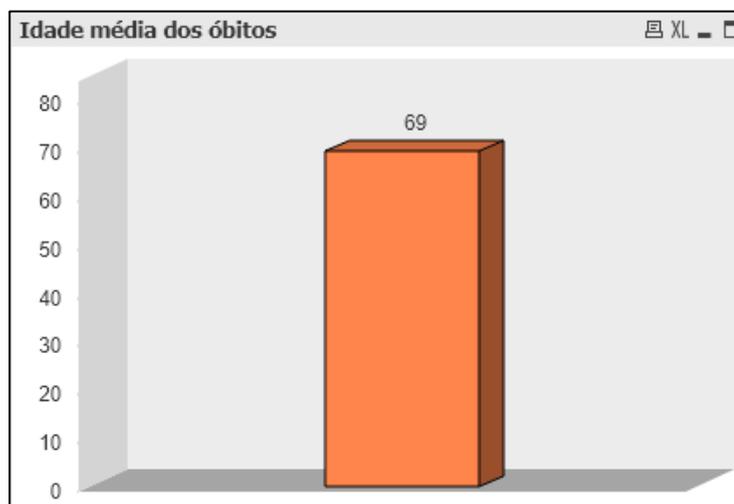


Figura 19: Idade média dos óbitos.

Fonte: (AUTOR, 2020)

A figura 19 representa a idade média dos óbitos e se observa que as pessoas mais idosas que estão mais propensas a morrer em decorrência do Covid-19, essa média pode variar de cidade para cidade e de mês para mês, algo em torno de 2 ou 3 anos, mas num todo a média gira na casa dos 69 anos.

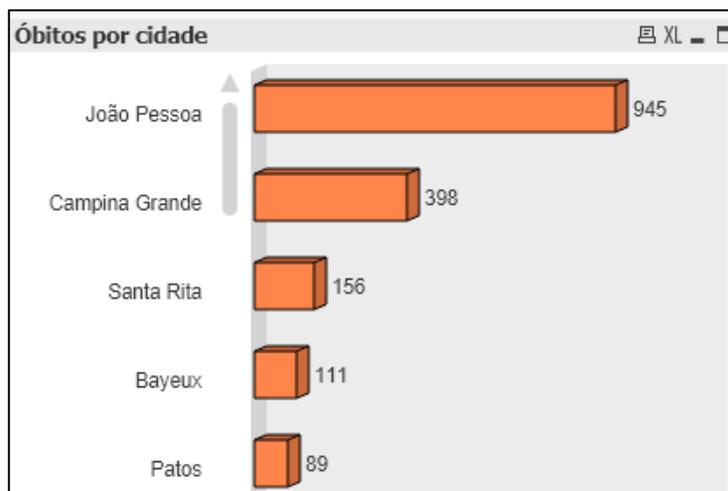


Figura 21: Óbitos por cidade

Fonte: (AUTOR, 2020)

A Figura 20 mostra os óbitos que ocorreram nas cidades da Paraíba, o maior índice de óbitos se dá nas cidades com maior número de habitantes, são elas João Pessoa e Campina Grande, logo em seguida pode observar as cidades do interior com números menores de óbitos. Seleccionando a cidade é possível analisar os números da cidade, como óbitos por gênero, idade média etc.

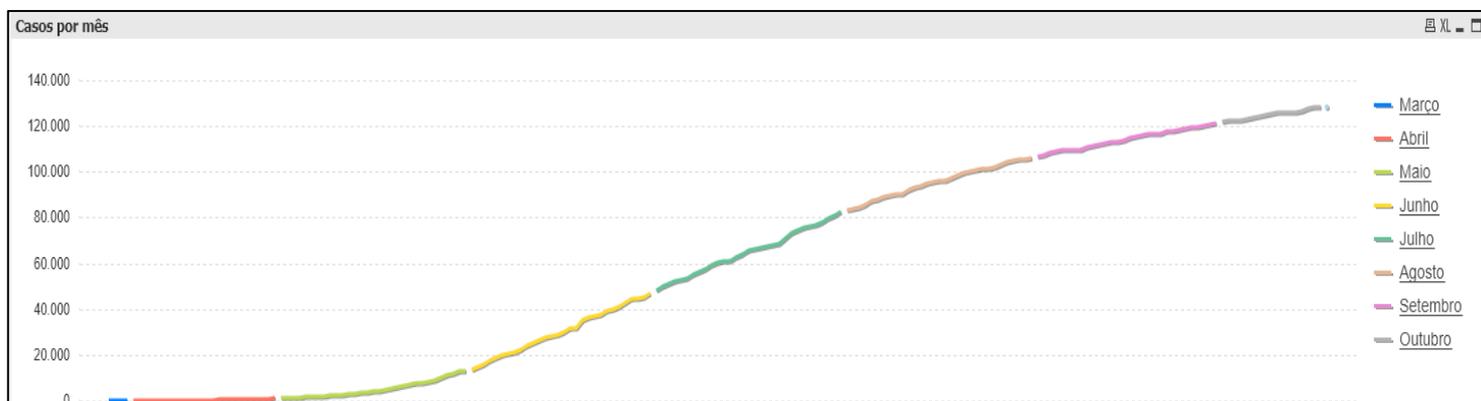


Figura 20: Casos por mês.

Fonte: (AUTOR, 2020)

Na Figura 21 é mostrado o acúmulo de casos por mês, e se observa que a curva de casos vem diminuindo em relação aos meses iniciais da pandemia, os meses de maio, junho e julho contabilizam o maior número de casos confirmados na Paraíba chegando a 82.794 casos.

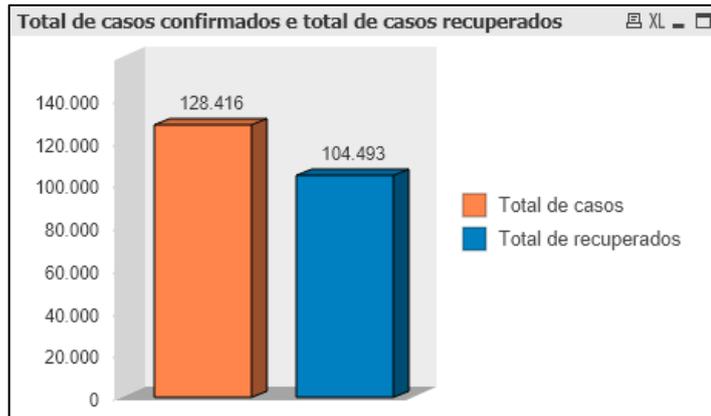


Figura 23: Total de casos confirmados e total de casos recuperados.

Fonte: (AUTOR, 2020)

No gráfico da Figura 22 contém o número total de casos confirmados em comparação com o total de casos recuperados, com isso podemos observar que existem cerca de 20.000 casos em observação tirando o número de óbitos.

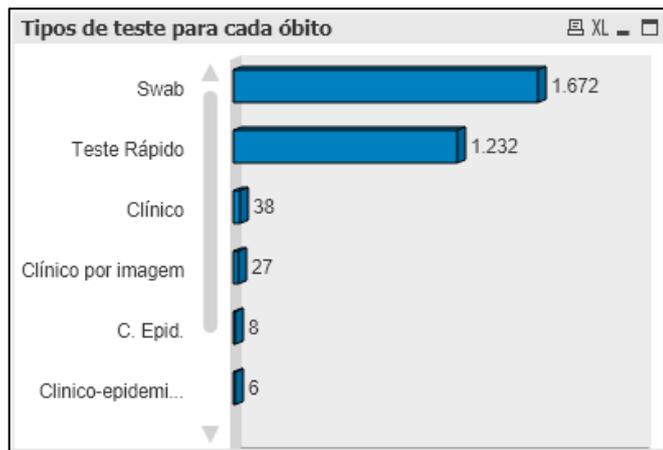


Figura 22: Tipos de teste para cada óbito

Fonte: (AUTOR, 2020)

A Figura 23 retrata os testes que foram realizados nas pessoas que vieram a óbito e os testes que foram usados por mais vezes foram os testes de *Swab*, que são testes feitos com cotonetes longos que são introduzidos no nariz do paciente, e testes rápidos que podem ser feitos em farmácias, esses não são testes tão precisos em relação ao *Swab*.

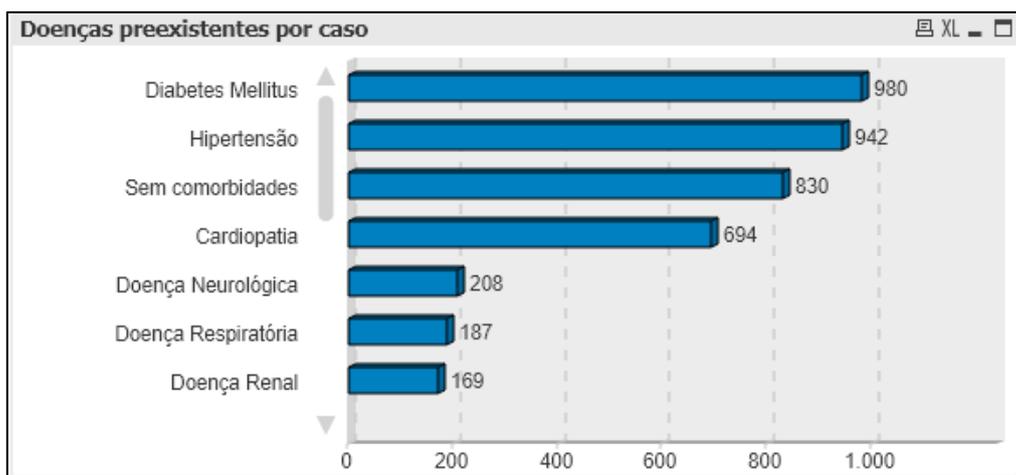


Figura 24: Doenças preexistentes por óbito.

Fonte: (AUTOR, 2020)

A Figura 24 mostra as doenças que os paciente já possuíam, muitos dos pacientes portavam mais de uma doença, com isso é observado um número total de doenças maior que o número de óbitos, a diabetes e hipertensão lideram o gráfico de doenças, esses tipos de doenças são mais comuns em pessoas com idade mais avançada, isso ajuda o vírus a se tornar mais mortal em pessoas idosas.

5. CONSIDERAÇÕES FINAIS

O desenvolvimento de um projeto de DW por si só é um grande desafio, foram abordados conceitos em que a academia e a população possam ler, compreender e utilizar de pesquisa, foram citados conceitos de BI, de DW e modelagem de dados, com isso se observou que a implementação de um projeto de DW pode ajudar na interpretação de base de dados de qualquer lugar e de qualquer tamanho.

Foi discutido no capítulo 4 na seção 4.4, que existe uma crescente no número de casos no estado da Paraíba, esses números tendem a diminuir e achatar a curva do gráfico de casos nos meses de setembro e outubro, desses casos contabilizados existem 2.985 óbitos, fica evidente um número maior de óbitos entre as pessoas do gênero masculino com média de idade de 69 anos e que possuem doenças preexistentes, como por exemplo, hipertensão, cardiopatia e diabetes.

Entre as pessoas do gênero feminino é possível observar que existe um número um pouco menor de óbitos, em relação ao gênero masculino, e a idade média dos óbitos é de 71 anos, um pouco acima em comparação ao gênero masculino. As doenças preexistentes se assemelham muito com o quadro masculino, sendo predominante óbitos em pessoas com diabetes, hipertensão e cardiopatia.

É importante deixar claro que pessoas do gênero masculino e feminino que não tenham doenças preexistentes também podem chegar a óbito, de acordo com os dados, 506 pessoas do gênero masculino e 324 pessoas do gênero feminino vieram a óbito e não apresentavam nenhum caso clínico de doenças preexistentes.

As projeções para o futuro são boas caso a população adote as medidas impostas pelo governo do estado, os casos por dia vem caindo de forma constante desde o início das medidas de contenção do vírus até o dia da coleta dos dados, caso nada seja feito é perigoso pois se trata de uma vírus altamente contagioso, que afeta principalmente pessoas mais idosas e com a saúde debilitada, é importante analisar os dados de forma constante com o intuito de blindar esse grupo de pessoas para que não fiquem expostas ao vírus e criar, se necessário, políticas públicas que possam diminuir o impacto de tais ações.

REFERÊNCIAS

BARBIERI, Carlos. **Business Intelligence: modelagem e qualidade**. Rio de Janeiro: Axcelbooks, 2011. 416 p.

BORGES, Hercules Alex. **Data Warehouse e ETL aplicados ao setor agrícola para obtenção de informações estratégicas**. 2017. 51 f. TCC (Graduação) - Curso de Curso de Especialização em Banco de Dados, Universidade Tecnológica Federal do Paraná, Pato Branco, 2017. Disponível em: <http://repositorio.roca.utfpr.edu.br/jspui/handle/1/9830>. Acesso em: 21 set. 2020.

CORREIA, Rafael Luciano Gonçalves. **Técnicas de aquisição de conhecimento na coleta de requisitos na modelagem de um banco data warehouse**. 2017. 38 f. TCC (Graduação) - Curso de Análise e Desenvolvimento de Sistemas, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2017. Disponível em: http://repositorio.roca.utfpr.edu.br/jspui/bitstream/1/8350/1/PG_COADS_2017_2_12.pdf. Acesso em: 26 set. 2020.

COLAÇO JÚNIOR, Methanias. **Projetando sistemas de apoio à decisão baseados em data warehouse**. Rio de Janeiro: Axcel Books do Brasil, 2004. 214 p. Editor Responsável: Ricardo Reinprecht.

HENRIQUE, Ozimar; VAN BILJON, Edward. **Dimensão Tempo**. 2016. Disponível em: <https://social.technet.microsoft.com/wiki/pt-br/contents/articles/12578.dimensao-tempo.aspx>. Acesso em: 20 out. 2020.

INMON, William H. **Como Construir O Data Warehouse**. 2. ed. Rio de Janeiro: Campus, 1997. 404 p. Tradução de Ana Maria Netto Guz.

LYRA, Antônio Luiz Bonna de. **Uso de um processo de ETL em um modelo Data Warehouse para a geração de Dashboards de indicadores de redes de telefonia celular**. 2016. 106 f. TCC (Graduação) - Curso de Engenharia Eletrônica e de Computação., Departamento de Engenharia Eletrônica e de Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

PATRICIO, Thiago Seti; STÁBILE, Samuel; TÓFOLI, Eduardo Teraoka. **A importância do Business Intelligence na tomada de decisões em gerenciamento de projetos**. 2016. 15 f. Monografia (Especialização) - Curso de Mba em Gerenciamento de Projetos, Centro Universitário de Lins, Lins, 2016.

REIS, E., TEIXEIRA, F. e ARAÚJO, M. A. (2009). **“Implementando uma solução de Business Intelligence com o Microsoft SQL Server 2005 – Parte 1”**. SQL Magazine, Rio de Janeiro, n. 59, p. 52-66.

REIS, Cláudio. **Uma arquitetura de BI: portal de acesso à informação e transparência do TJSC - gestão de pessoas**. 2018. 79 f. TCC (Doutorado) - Curso de Sistemas de Informação, Universidade do Sul de Santa Catarina, Florianópolis, 2018. Disponível em:

<https://www.riuni.unisul.br/bitstream/handle/12345/5984/TCC.pdf?sequence=1&isAlloved=y>. Acesso em: 26 set. 2020.

ROSSI, Rafael Germano. **Análise de Componentes Principais em Data Warehouse**. 2017. 81 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017. Disponível em: <https://www.teses.usp.br/teses/disponiveis/45/45134/tde-07012018-182730/publico/mestrado.pdf>. Acesso em: 10 set. 2020

SILVA, Luís Alexandre da; BRUSCHI, Gustavo Cesar; SILVA, Franciny Quinelato da; TORRES, Juliano Rodrigues. P.97 **Proposta de Data Warehouse para ser consumido por gestores de produção**. Revista Fatec Sebrae em Debate Gestão, Tecnologias e Negócios, São Paulo, v. 2, n. 2, p. 1-190, 18 abr. 2018. Disponível em: <http://revista.fatecsebrae.edu.br/index.php/em-debate/article/view/18>. Acesso em: 18 set. 2020.

SILVA, Marcio; SARTORI, Marco. **Data Warehouse a vantagem da modelagem dimensional de dados**. In: Encontro Científico Cultural Interinstitucional, 13., 2015, Cascavel. Anais [...]. Cascavel: Fag, 2015. p. 1-10.

SILVA, R. A.; SILVA, F. C. A.; GOMES, C. F. S. **O uso do Business Intelligence (BI) em sistema de apoio a tomada de decisão estratégica**. Revista Gestão Inovação e Tecnologias, [S.L.], v. 6, n. 1, p. 2780-2798, 27 mar. 2016. Associação Acadêmica de Propriedade Intelectual. Disponível em: <http://dx.doi.org/10.7198/s2237-0722201600010005>. Acesso em: 10 set. 2020.

TARGET. **Microdados: óbitos na paraíba**. 2020. Disponível em: <https://superset.plataformatarget.com.br/superset/dashboard/microdados/>. Acesso em: 13 nov. 2020.

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay E.; KING, David. **Business Intelligence: Um enfoque gerencial para a inteligência do negócio** para a. [S.L]: Bookman, 2009. 249 p. Tradução de: Fabiano Bruno Gonçalves.

UNO, Nairemilia Kuchauski Alves. **A Importância do Data Warehouse no Business Intelligence**. In: Encontro Toledo de Iniciação Científica, 11., 2015, Presidente Prudente. Anais [...]. Presidente Prudente: Etic, 2015. v. 11, p. 1-50. Disponível em: <http://intertemas.toledoprudente.edu.br/index.php/ETIC/article/view/4758>. Acesso em: 10 set. 2020.

WAGNER, Cláudio. **Data Warehouse (DW)**. 2012. Elaborado por: Cláudio Wagner.
Disponível em: <http://cacau-indicou.blogspot.com/2012/02/data-warehouse-dw.html>.
Acesso em: 1 fev. 2012.