

Centro Universitário Carioca - Unicarioca
Ciência da Computação

Fernanda Cristina da Silva Santos
Israel Pereira de Lucena Junior

**PUBLIC DASH: Uma análise preditiva de crescimento salarial baseada na
escolaridade com Power BI**

Rio de Janeiro
2018

Fernanda Cristina da Silva Santos

Israel Pereira de Lucena Junior

PUBLIC DASH: Uma análise preditiva de crescimento salarial baseada na escolaridade com Power BI

Trabalho de Conclusão de Curso apresentado ao curso de Ciência da Computação, como parte dos requisitos necessários à obtenção do título de Bacharel.

Orientador: Jorge Viana Doria Junior,
M.Sc.

Rio de Janeiro

2018

Fernanda Cristina da Silva Santos

Israel Pereira de Lucena Junior

PUBLIC DASH: Uma análise preditiva de crescimento salarial baseada na escolaridade com Power BI

Trabalho de Conclusão de Curso
apresentado ao curso de Ciência da
Computação, como parte dos requisitos
necessários à obtenção do título de
Bacharel.

Orientador: Jorge Viana Doria Junior,
M.Sc.

Banca Examinadora

Prof. André Luiz Avelino Sobral, M.Sc
Centro Universitário Carioca

Prof. Jorge Viana Doria Junior, M.Sc
Centro Universitário Carioca

Prof. Alessandro Jatobá, D.Sc
Centro Universitário Carioca

Agradecimentos

Israel Pereira de Lucena Junior,

Agradeço a minha família, minha mulher, meus amigos e colegas por todo o apoio que recebi nesta caminhada até o dia de hoje.

Fernanda Cristina da Silva Santos,

Agradeço à minha família e amigos por sempre terem me apoiado durante toda a jornada, por terem me dado suporte nos dias que pensei em desistir e por comemorar todas as vitórias comigo.

Resumo

Atualmente, a sociedade e as organizações têm gerado dados a todo momento. Em função disso, o surgimento da quantidade excessiva de dados desorganizados, com diversas fontes, tipos e formatos gerou a necessidade de organizá-los e transformá-los em informação útil. O conceito de *Big Data* surgiu como o estudo de massa de dados muito grandes e complexos. Nos dias atuais, o termo tem sido utilizado para englobar diversas atividades, tais como: obtenção, análise de dados para gerar indicadores e modelos preditivos com Data Science e Machine Learning. Este trabalho irá utilizar os conceitos citados anteriormente para criar um dashboard no *Power BI*, utilizando dados públicos e consultados no IBGE. O *Power BI* é um serviço de *Business Analytics* baseado em nuvem, que oferece a qualquer profissional a chance de criar visualizações e dashboards de alto nível, a partir da análise de dados e com poucos cliques. Ao final, fala-se sobre os motivos pela escolha da ferramenta *Power BI*, as conclusões do trabalho e, também, serão indicados possíveis modelos preditivos que foram percebidos durante a pesquisa realizada.

Palavras-chave: BI. Power BI. Big Data. dashboards.

Abstract

Today, society and organizations have generated data all the time. Because of this, the appearance of excessive amount of data disorganized, with diverse sources, types and formats generated the need to organize it and turn it into useful information. The concept of Big Data emerged as the study of very large mass and complex data. Nowadays, the term has been used to encompass several activities, such as: obtaining and analyzing data to generate indicators and predictive models with Data Science and Machine Learning. This work will use the concepts previously mentioned to create a dashboard in Power BI, using public data and consulted in IBGE. Power BI is a cloud-based Business Analytics service that gives any trader the chance to create high-level views and dashboards from data analysis and with a few clicks. At the end, we talk about the reasons for choosing the Power BI tool, the conclusions of the work, and also, possible predictive models that were perceived during the research will be indicated.

Keywords: BI. Power BI. Big Data. dashboards.

Sumário

1 Introdução	9
1.1 Contextualização	9
1.2 Motivação	10
1.4 Organização do trabalho	11
2 Conceito de Business Intelligence	11
2.1 Introdução	12
2.2 Data Warehouse	12
2.3 Componente de uma solução de Data Warehouse	13
2.4 Modelagem Dimensional	14
2.5 Fatos e Métricas	14
2.6 Dimensões	16
3 Plataforma de BI Microsoft	16
3.1 Introdução	16
3.2 SQL Server	17
3.3 Integration Services	18
3.4 Analysis Services	22
3.5 Excel, Reporting Services e Power BI	23
3.6 Posicionamento da Microsoft em pesquisas de mercado	29
4 Implementando o projeto de BI do IBGE	31
4.1 Introdução a análise do IBGE	31
4.2 Entendendo as ocorrências salariais e de escolaridades	31
4.3 Elaborando os requisitos da solução para o IBGE	32
4.4 Elaborando o desenho da solução final	34
4.5 Construindo o ETL	35
4.6 Construindo o Data Warehouse	37
4.7 Construindo o modelo semântico	38
4.8 Construindo os Relatórios e Dashboards no Power BI	39
4.9 Compartilhando relatórios no Report Server	41

4.10 Resultados obtidos	43
5 Conclusão	48
5.1 Considerações finais	48
5.2 Limitações encontradas no decorrer do projeto	48
5.3 Possibilidade de melhorias futuras	48
6 Referências	50

1 Introdução

Este capítulo contém a contextualização do assunto no cenário atual, o objetivo e a motivação para a elaboração deste trabalho.

1.1 Contextualização

O aumento do volume de dados exigiu a rápida adequação do mercado a algumas mudanças. Segundo a Gartner, aproximadamente 2,2 milhões de dados são movimentados diariamente. Devido a grande quantidade de dados gerados, o termo *Big Data* tem ficado cada vez mais popular. As empresas têm buscado utilizar essa massa de dados para gerar valor, que pode se tornar uma vantagem competitiva no mercado, uma estratégia para conseguir novos cliente, dentre outros.

Além disso, o uso de análise preditiva no auxílio à tomada de decisão vem crescendo. Baseando-se na extração de dados, geração de relatórios com fundamentos em cálculos estatísticos para visualização de tendências e previsão de comportamentos de públicos, é possível, também, refinar as necessidades do cliente com o intuito de direcionar melhor o atendimento.

Existem ainda outros tipos de análise, como a prescritiva que é utilizada para entender as consequências que podem ocorrer a partir de uma decisão tomada, a descritiva que realiza um mapeamento em tempo real para obter um perfil do cliente e a cognitiva que procura identificar e entender de forma mais específica cada uma das ações, as causas, os impactos e o motivo para terem sido realizadas.

Com essa grande procura por valor a partir da informação, aumentou a busca por serviços de BI de autoatendimento ou *self-service BI*, como também é conhecido. A *Microsoft* desenvolveu uma aplicação chamada *Power BI* que é uma referência em serviços de autoatendimento de *Business Intelligence*. A ferramenta é muito semelhante ao *Microsoft Excel*, o que a torna mais intuitiva que as concorrentes no mercado, facilitando a adaptação. É possível utilizar R, Dax,

PowerQuery e, mais recentemente, *Python* para manipular os dados que podem ser trazidos diretamente de diversas fontes, como requisições rest, banco de dados, *Microsoft Excel* e csv. O *Power BI* será utilizado neste trabalho para a geração dos relatórios.

1.2 Motivação

Com a constante evolução da tecnologia, algumas profissões começaram a ganhar foco, como o cientista de dados. A ciência de dados (*Data Science*) não é uma área nova, mas ganhou foco nos últimos anos por causa do Big Data. Entre 2012 e 2017, nos Estados Unidos, o número de vagas foi multiplicado por aproximadamente 6,5. Já no Brasil, os salários dos profissionais dessa área cresceram 15,25% desde 2017, ficando entre R\$12 mil e R\$22 mil (FOLHA - 2018).

Analisar dados para obter informações relevantes ao negócio tem se tornado cada vez mais importante. A Nubank (banco digital e operadora de cartão de crédito) utiliza *Data Science* em todo o seu processo. A empresa possui um *Squad* de Ciência de Dados totalmente dedicado, com 15 profissionais. Como grandes bancos já existiam e estavam totalmente consolidados, a Nubank, fundada em 2013, resolveu utilizar os dados para criar um diferencial e se tornou uma empresa “*data-driven*”, ou seja, cujo marketing é orientado a dados. Ao trabalhar com a análise de informações geradas pelo aplicativo, é possível entregar uma melhor experiência para o usuário e evitar perdas para a empresa.

Ter profissionais capazes de basear suas decisões em dados é fundamental no mercado que está começando a se desenvolver, afinal, é possível utilizar os dados que a empresa possui para gerar indicadores futuros, o que faz com que as decisões sejam tomadas tendo uma ideia dos possíveis riscos envolvidos naquela decisão.

Existem diversas ferramentas específicas para apoio às soluções de *Business Intelligence* disponíveis no mercado atualmente. A motivação deste trabalho é realizar um estudo de caso utilizando Power BI e explicar o motivo da escolha desta ferramenta, apresentando as vantagens encontradas ao utilizá-la. Ao longo do texto, serão citadas ferramentas já solidificadas no mercado, como o SQL Server e o

Excel. A comparação realizada ao final do trabalho será em relação a esses dois *softwares*.

1.3 Objetivos

O objetivo deste trabalho é implementar e analisar dashboards preditivos utilizando dados públicos disponibilizados pelo IBGE. A versão final do dashboard possuirá visualização mobile e desktop.

1.4 Organização do trabalho

O primeiro capítulo apresenta o objetivo do trabalho e o cenário de grande volume de dados que existe atualmente.

O segundo capítulo trata do Conceito de *Business Intelligence*. Nele serão abordados os tópicos relacionados ao conceito que envolve BI, como *Data Warehouse*, Modelagem Dimensional e Dimensões.

O terceiro capítulo é a análise da Ferramenta de self-service BI da *Microsoft*, o *Power BI*, e irá apresentar também outros sistemas e técnicas disponíveis no mercado.

O quarto capítulo trata da especificação do projeto que utilizará os dados do IBGE e a implementação do modelo de *Data Warehouse*, análise preditiva e implementações relacionadas ao dashboard.

O quinto capítulo é a conclusão, onde serão apresentadas as possíveis melhorias para trabalhos futuros, as limitações encontradas e considerações finais.

2 Conceito de *Business Intelligence*

2.1 Introdução

O conceito de *Business Intelligence* surgiu na década de 90 referindo-se a processos de uma organização que envolvam o tratamento de informações.

Business Intelligence é um processo que envolve métodos, métricas, tecnologias etc. para explicar causas ou eventos. Caso o resultado do evento seja negativo, BI tem a função de descobrir a causa do problema para tratá-lo, já com um resultado positivo, o objetivo torna-se melhorá-los.

A explicação para causas é baseada em dados. O processo reconhece padrões e os utiliza para realizar análise dos dados, garantindo assim os sucessos e evitando fracassos.

O processo de BI deve incluir o analista de BI, que é a pessoa especializada no software que será utilizado para o processo e um analista de negócios que é responsável por extrair os dados para serem utilizados na organização e definir o objetivo do processo.

Este capítulo irá explicar algumas partes do processo de BI, como *Data Warehouse*, a modelagem dimensional, fatos e métricas e dimensões.

2.2 Data Warehouse

O pré-processamento e preparação de dados tem como objetivo melhorar os dados e gerar uma base separada para realizar a análise. Os processos de BI devem ser feitos em bases separadas da base transacional para que não haja nenhum problema nas operações diárias da organização. Desse modo, exemplifica-se o conceito de *Data Warehouse* que é possuir uma base formada por dados de bases transacionais para realizar as análises.

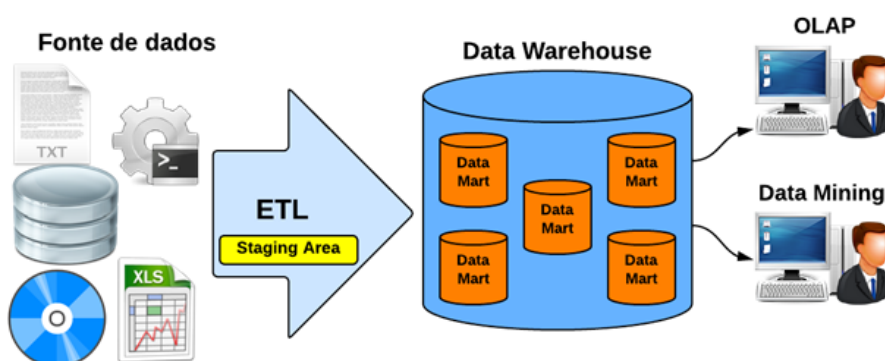
O *Data Warehouse* possui algumas características:

- Orientado por assunto: a partir dos principais assuntos da empresa, a modelagem é feita.
- Integrado: padroniza os dados em uma representação que é transferida para a base de dados do *Data Warehouse*.
- Não volátil: os dados são carregados de forma definitiva, ou seja, não existe alteração de dados após o seu carregamento, apenas inclusão e exclusão.

- Variável com o tempo: as análises a partir do *Data Warehouse* não são em tempo real, isso porque o banco consultado não é o transacional. Existem chaves de tempo que indicam qual o período analisado.

2.3 Componente de uma solução de *Data Warehouse*

Figura 1 - Componentes de um *Data Warehouse*



Fonte:

<https://canaltech.com.br/business-intelligence/conhecendo-a-arquitetura-de-data-warehouse-19266/>

Os componentes da solução são os seguintes:

- Fonte de dados: é um sistema transacional da organização.
- ETL: é o processo que extrai, transforma e carrega os dados até serem armazenados no Data Warehouse.
- Área de *staging*: é uma área intermediária usada para processar os dados durante o processo de ETL, possuindo como objetivo primário aumentar a eficiência do ETL. Os dados gerados nesta etapa estarão limpos, sem duplicidade e transformados para serem utilizados no data warehouse, não serão exibidos para o usuário final.
- *Data Warehouse*: armazena as informações finais. Os dados que se encontram no *Data Warehouse* são os que possuem valor para a organização.
- *Data Mart*: é uma divisão lógica do *Data Warehouse*, dividido por agrupamentos interessantes ao usuário.

- OLAP: *On-line Analytical Processing* é o conjunto de ferramentas que podem analisar diversas perspectivas dos dados armazenados no Data Warehouse.
- *Data mining*: é a mineração de dados. Procura padrões ou tendências a partir de regras. Trabalha com grandes massas de dados, então seria difícil realizar essa mineração manualmente.

2.4 Modelagem Dimensional

A modelagem dimensional, ou multidimensional, é usada nas consultas ao *Data Warehouse* e permite a análise pelas ferramentas OLAP. Possui os modelos *star*, cujas dimensões são desnormalizadas, e *snow flake*, que é parcialmente normalizado.

A estrutura multidimensional possui tabelas desnormalizadas, grande número de redundância e suas atualizações ocorrem em periodicidade menor que a de uma estrutura relacional convencional. A importância de uma modelagem multidimensional bem feita é grande pois permite um bom desempenho, a escalabilidade do *Data Warehouse* e boa intuitividade. Uma boa modelagem garante a geração de um bom armazém de dados, com maior qualidade.

2.5 Fatos e Métricas

A modelagem dimensional possui dois elementos obrigatórios e complementares: as tabelas “Fatos” e as tabelas “Dimensões”.

As tabelas fatos contém as informações qualitativas das descritivas armazenadas nas Dimensões. Elas contém as métricas que são utilizadas para medir os dados desejados pela organização e as *foreign keys*, que realizam a relação com a tabela de Dimensões.

Existem seis tipos de tabelas fato:

- Transacional: é a mais comum. Utiliza métricas que podem ser somadas em todas as dimensões. As duas formas de armazenar informações na fato transacional são: gravar vários itens de uma mesma transação em várias linhas, ou gravar itens com algum campo

contador, como quantidade. Desse modo, haverá menos linhas na tabela.

- Agregada: acelera o desempenho de consultas, juntando apenas os valores desejados pela organização, por exemplo: se uma organização quer saber o tráfego em seu site por data, ao invés de trazer uma consulta contendo as tabelas data e hora, a tabela agregada só traria o campo data e o resto das informações desejadas. Apesar do campo hora conter mais informações a respeito dos acessos, pode não ser necessário esse nível de detalhamento no relatório, então sua geração será mais rápida utilizando a fato agregada.
- Consolidada: é semelhante a fato agregada, que unifica valores, porém a consolidada unifica dois tipos de processos. A criação da tabela ocorre durante o processamento de ETL, ela é gerada juntando duas tabelas com as informações de interesse.
- De *Snapshot* periódico: o *snapshot* é como uma fotografia do momento, ou seja, o periódico é tirado de tempos em tempos.
- De *Snapshot* acumulado: o *snapshot* acumulado é semelhante ao periódico, mas possui o snapshot de vários momentos. Dessa forma, esse *snapshot* visita novamente o fato e insere dados sem deletar registros. Geralmente, o ETL carrega o primeiro campo, depois um job carrega outro pedaço.
- Sem fato: é um fato sem métricas, serve para realizar associação entre as dimensões. É chamada também de fato de associação, é composta apenas por elementos associativos, como *foreign keys*.

As métricas são os valores que a organização exigir. Existem quatro tipos de métricas:

- Aditivas: permitem operações matemáticas, como soma e subtração, por todas as dimensões.
- Derivadas: é calculada a partir de um valor que está no fato. É a criação de uma nova métrica.

- Semi-aditivas: pode ser somada por quase todas as dimensões, menos a de tempo.
- Não-aditivas: não podem ser somadas por nenhuma dimensão, são valores percentuais feitos em tempo de execução.

2.6 Dimensões

As dimensões possuem os dados qualitativos da tabela Fato, ela permite a visualização de informações por perspectivas diferentes.

Existem cinco tipos de dimensões:

- *Degenerate Dimension*: Essa dimensão seria muito pequena para ter uma dimensão e foi inserida como uma coluna da tabela Fato.
- *Slowly changing dimension*: não é uma dimensão e sim uma técnica para atualizar a tabela dimensão.
- *Role-playing dimension*: Quando existem casos em que a dimensão é utilizada por múltiplos lugares, como datas, é criada uma outra dimensão ao invés de colocar a dimensão de tempo várias vezes.
- *Conformed dimension*: é quando a dimensão possui fatos e métricas detalhadas e classificadas do mesmo modo em toda a organização, ela tem dados que diversas fatos precisam.
- *Junk dimension*: É uma solução para quando existem dimensões com cardinalidade baixa ou densidade esparsa. É uma dimensão “lixo”, possui os tipos que não tem como ficar em outro lugar.

3 Plataforma de BI *Microsoft*

3.1 Introdução

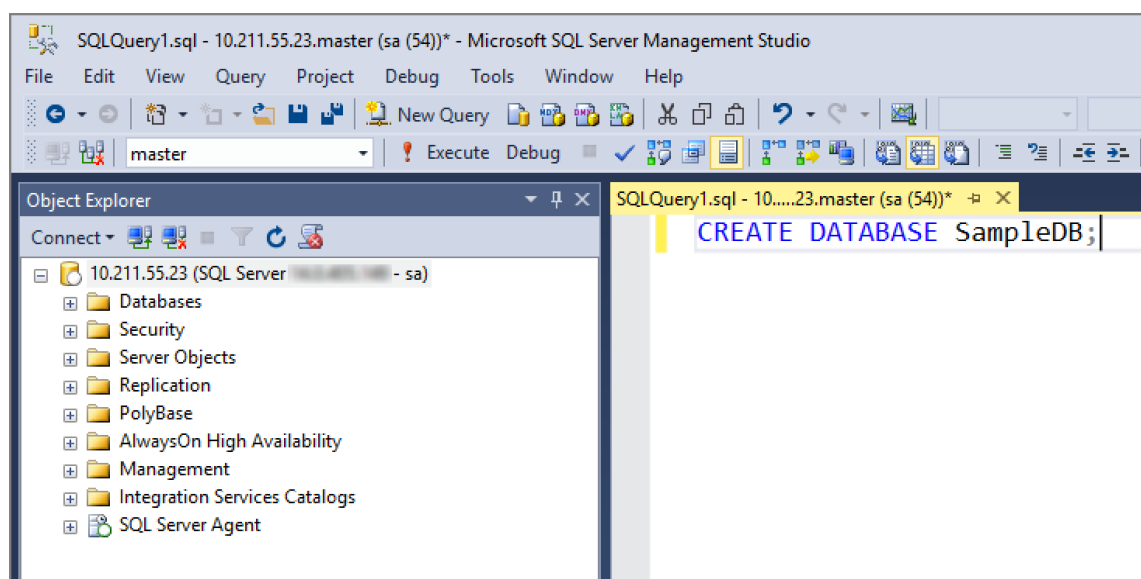
Este capítulo apresenta tecnologias utilizadas pela *Microsoft* para apoiar soluções de *Business Intelligence* e *Data Warehouse*. A introdução ao *SQL Server* está incluída por oferecer suporte às três ferramentas de BI mais conhecidas da *Microsoft*: *Integration Services*, *Analysis Services* e *Reporting Services*.

O *Power BI* atualmente encontra-se entre as plataformas mais bem avaliadas do mercado. Possui fácil interação, pois a ferramenta incorporou em si funcionalidades de outras aplicações existentes, como o *Excel*, *Reporting Services*, dentre outras.

3.2 SQL Server

O *SQL Server* (Figura 2) é um sistema de gerenciamento de banco de dados operacionais e possui a liderança nesse setor. A linguagem utilizada para manipular o *SQL Server*, a partir da versão 2008, é a *Transact-SQL* (T-SQL), que também pode manipular o banco de dados SQL do *Azure*, *SQL Data Warehouse* do *Azure* e o *Parallel Data Warehouse*. A partir da versão 2012 do *SQL Server* a linguagem *Xquery* foi inserida como subconjunto do T-SQL, ela é utilizada para consultar dados do tipo xml. Com o SQL é possível criar banco de dados relacionais para processamento analítico e de transações *online*.

Figura 2 - SQL Server



Fonte: <https://docs.microsoft.com/pt-br/sql/linux/sql-server-linux-run-sql-server-agent-job?view=sql-server-2017>

As tecnologias disponíveis para o *SQL Server* incluem o *SQL Server Management Studio* e o *SQL Server Profiler* que podem ser utilizados para gerenciar

objetos do banco e capturar eventos do servidor, respectivamente. Também é possível realizar a replicação do *SQL Server*, que copia e distribui dados e objetos entre bancos, e os sincroniza para garantir consistência. O *Data Quality Services* é um produto orientado a conhecimento, ele é composto pelo *Data Quality Server* e o *Cliente Data Quality*, e consiste em criar uma base de conhecimento e usá-la para verificar a qualidade, e realizar correções. Além dessas tecnologias, é possível utilizar o *SQL Server* para serviços de *Machine Learning* com as linguagens R e Python.

A *Microsoft* incluiu diversos serviços de *Business Intelligence* e *Data Warehouse* no *SQL Server*, que são o *Integration Services*, o *Reporting Services* e o *Analysis Services*.

3.3 *Integration Services*

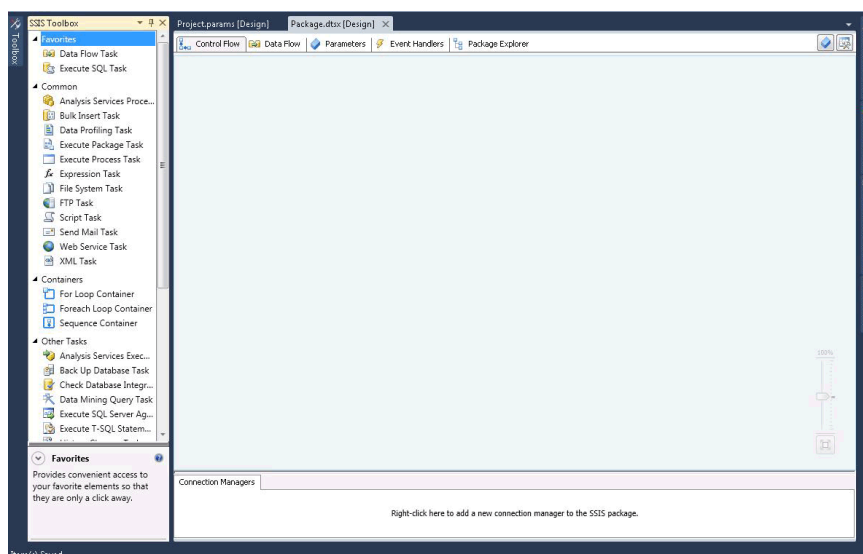
O *Integration Services* é uma plataforma ideal para integrar dados de alto desempenho e transformá-los. Inclui pacotes que permitem utilizar a ferramenta para fazer ETL e armazenar os dados de diversas fontes, como dados relacionais, arquivos simples, dados XML e realizar sua transferência para diversos destinos. É possível instalar um servidor totalmente dedicado a processos de ETL.

O *SQL Server Integration Services Designer* (*SSIS Designer*) é uma interface gráfica (Figura 3) que é utilizada para realizar criações e mantém pacotes do *Integration Services*.

Os componente do *Integration Services* estão divididos da seguinte forma:

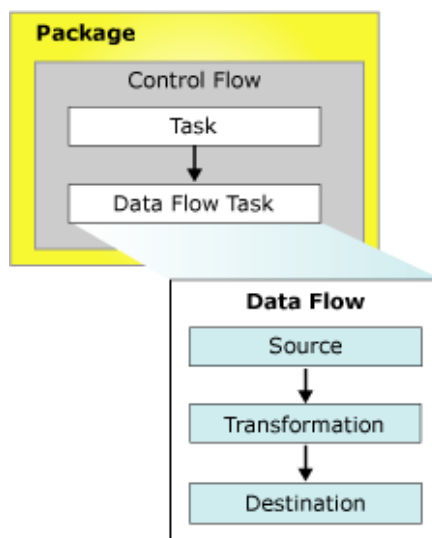
- *Package*: é a unidade recuperada, executada e salva. Sua composição contém conexões, elementos de fluxo de controle e dados, manipuladores de eventos, variáveis, parâmetros e configurações que podem ser agrupadas pela ferramenta de design do SSIS, o *SSIS Designer* ou criadas via programação. O pacote pode ser salvo no armazenamento de packages do SSIS, arquivos de sistemas ou no *SQL Server*. Ao ser criado, o pacote está sem funcionalidade, para adicioná-las é necessário incluir fluxos de controle (*control flow*) e fluxos de dados (*data flow*), opcionalmente (Figura 4).

Figura 3 - SSIS Designer



Fonte: <https://docs.microsoft.com/pt-br/sql/integration-services/ssis-designer?view=sql-server-2017>

Figura 4 - Package

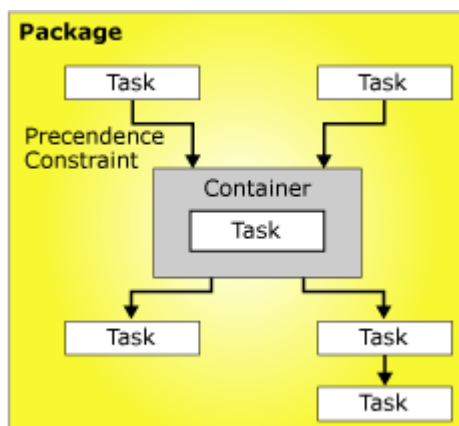


Fonte: <https://docs.microsoft.com/pt-br/sql/integration-services/integration-services-ssis-packages?view=sql-server-2017>

- **Control Flow:** é composto por tarefas, responsáveis pela funcionalidade, além de contêineres responsáveis por fornecer a estrutura e restrições de precedência, realizando a conexão desses contêineres e das tarefas de forma ordenada. Eles são executados junto com o pacote, é possível agrupá-los e executá-los de forma

repetida. A Figura 5 mostra um contêiner conectado a seis tarefas no qual cinco estão diretamente ligadas ao contêiner e a sexta tarefa depende outra realizar sua execução.

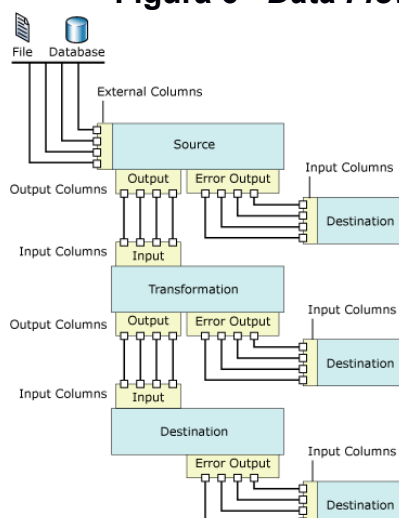
Figura 5 - Exemplo de Control Flow



Fonte: <https://docs.microsoft.com/pt-br/sql/integration-services/control-flow/control-flow?view=sql-server-2017>

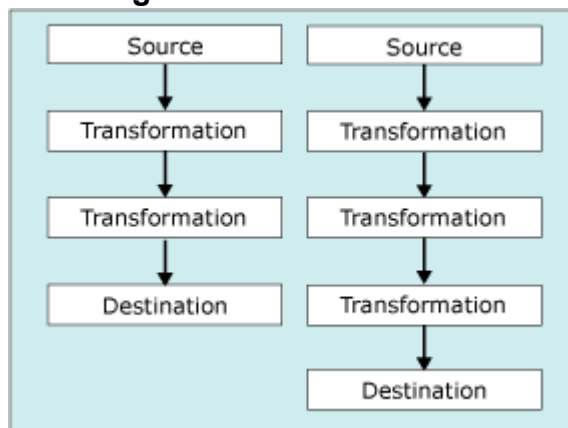
- **Data Flow:** existem três tipos de componente de *data flow* (Figura 6), as fontes, que tem o papel de extrair dados armazenados, as transformações, que modificam, resumem e limpam dados, e os destinos que carregam os dados na memória ou em repositórios. A conexão de componentes de *data flow* é feita entre a saída de fontes e destinos e a entrada de transformações e destinos. Existe ainda o *Data Flow Task* (Figura 7), que encapsula o mecanismo de *data flow*. Sua adição permite que o pacote realize o ETL.

Figura 6 - Data Flow



Fonte: <https://docs.microsoft.com/pt-br/sql/integration-services/data-flow/data-flow?view=sql-server-2017>

Figura 7 - Data Flow Task



Fonte: <https://docs.microsoft.com/pt-br/sql/integration-services/control-flow/data-flow-task?view=sql-server-2017>

- Gerenciadores de conexão: é o *link* entre a fonte de dados e o pacote, que possui pelo menos um gerenciador de conexão. O SSIS inclui conexão para dados do tipo texto, XML, bancos relacionais e projetos do SSAS (SQL Server Analysis Service).

3.4 Analysis Services

Analysis Services é uma ferramenta de OLAP e *data mining* do Microsoft SQL Server para aplicativos de *Business Intelligence*. O SSAS (SQL Server Analysis Service) é utilizado para analisar informações, que podem estar em bancos de dados, tabelas, ou arquivos diferentes.

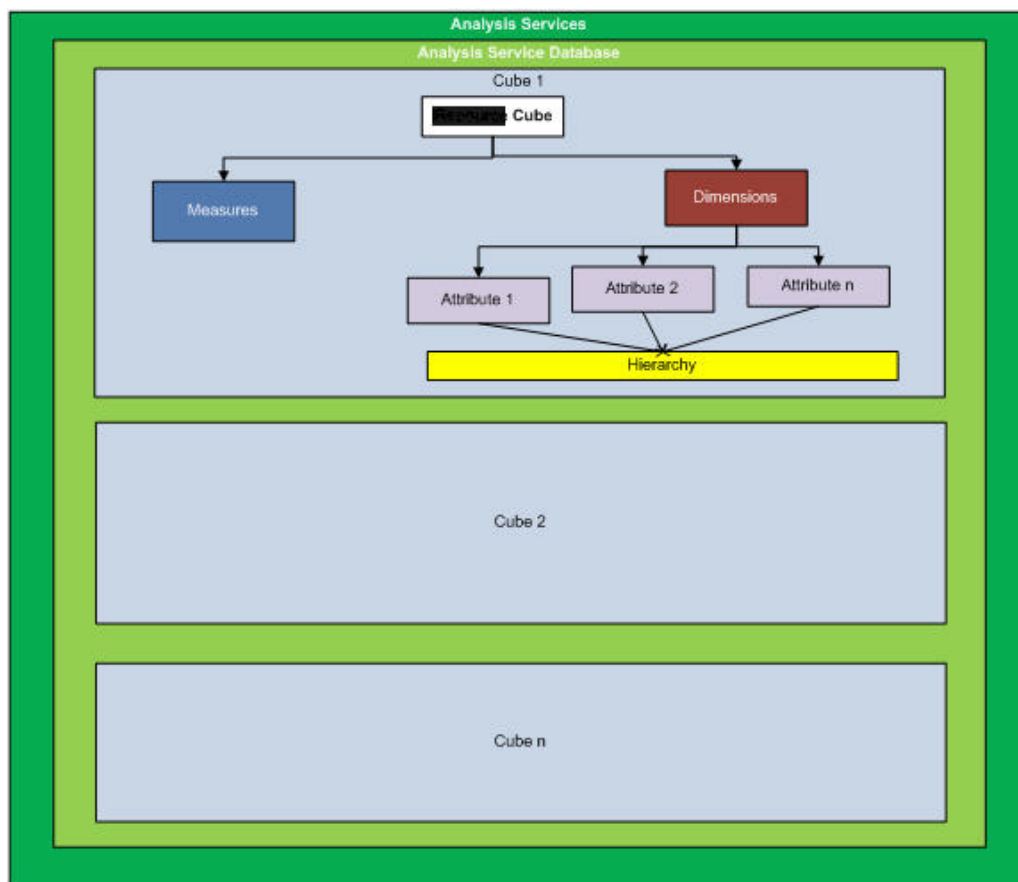
O *Analysis Services* auxilia no projeto, criação e visualização de modelos de *data mining*, ele constrói cubos de dados que podem ser acessados pelo Microsoft Excel, que atua como uma excelente ferramenta de exibição dos dados coletados.

Um cubo do *Analysis Services* é composto por: medida, que são valores numéricos, agregados, calculados em tempo real, as dimensões, que são semelhantes a tabelas no modelo relacional, atributos, que seria equivalente a colunas no modelo relacional (Figura 8).

Dentre as vantagens de uso estão a arquitetura escalonável para lidar com OLAP, o suporte para aplicativos personalizados, flexibilidade na criação e

gerenciamento de dados, a integração com fonte de dados, segurança, relatórios etc.

Figura 8 - Analysis Services



Fonte: <https://bennyaustin.files.wordpress.com/2009/06/cube-model.jpg>

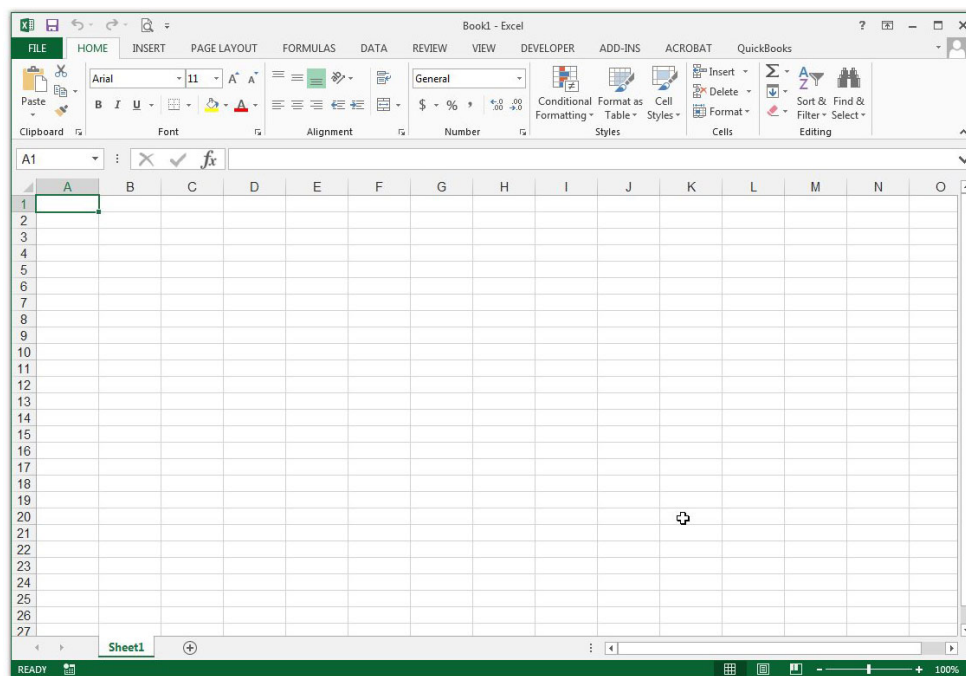
3.5 Excel, Reporting Services e Power BI

O *Microsoft Excel* é um programa de planilhas que realiza armazenamentos e operações com dados (Figura 9). Sua estrutura é semelhante a uma planilha de papel, possui linhas e colunas, o encontro entre linha e coluna forma a célula.

Ao referenciar uma célula, o correto é unir o nome de sua coluna com o de sua linha, a representada na figura 9 é A1, onde “A” é o nome da coluna e “1” é o número da linha. Os tipos de dados mais simples que podem existir em uma célula são data e hora, texto e número. Existem os tipos booleano, também conhecido como valor lógico, que possui o retorno como true ou false (0 ou 1), e fórmula, que pode ser uma simples adição, ou uma expressão mais complexa. É comum que o

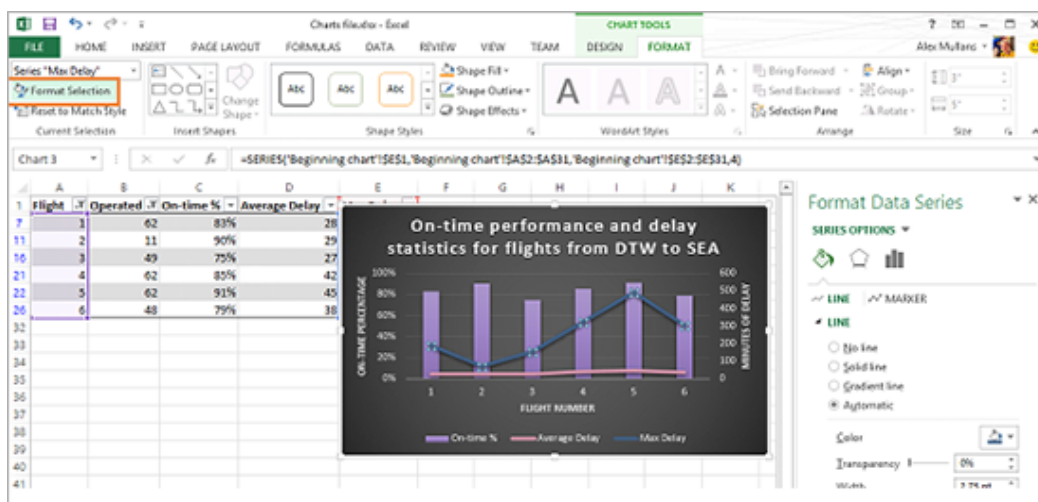
Microsoft Excel seja utilizado para realizar cálculos financeiros, mas possui outras utilidades, como importar dados de bancos para análise e para gerar dashboards de visualização (Figura 10).

Figura 9 - Planilha *Microsoft Excel*



Fonte: <https://www.solveyourtech.com/version-microsoft-excel-using/>

Figura 10 - Relatório com *Microsoft Excel*



Fonte: <https://www.microsoft.com/en-us/microsoft-365/blog/2013/10/02/telling-a-story-with-charts-in-excel-2013/>

O *Reporting Service* (Figura 11), mais conhecido como SSRS (*SQL Server Reporting Services*), é uma ferramenta local para criar, publicar e gerenciar relatórios, é possível enviá-los por e-mail, exibir em sites e aplicativos.

Figura 11 - Reporting Service



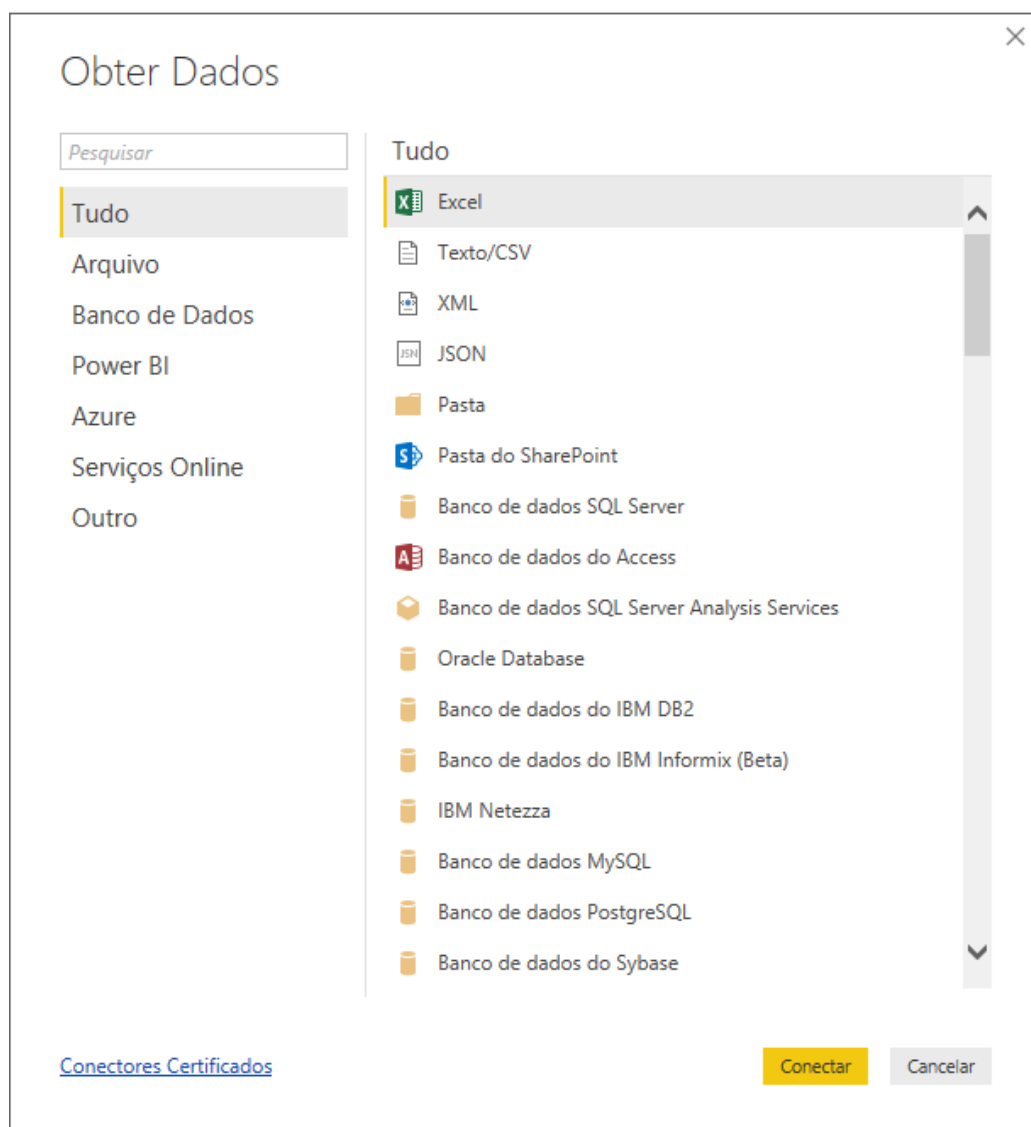
Fonte: <https://docs.microsoft.com/pt-br/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-2017>

Qualquer banco de dados que possua um componente de conexão com o *Reporting Service*, pode gerar relatórios a partir dele, além disso, a ferramenta possui suporte a *APIs RESTful*. A partir da versão 2017 foi possível introduzir comentários e anexos comentados nos relatórios, o DAX nativo foi inserido para modelos do SSAS que possuam suporte. O SSRS gera visualizações paginadas, móveis responsivas e possui um portal compatível com a maioria dos navegadores, que é totalmente personalizável e cria *KPIs* (*Key Performance Indicator* - Indicadores-Chave de desempenho) direto na *web*. O conteúdo é separado por tipo: *KPIs*, *dashboards mobile* e paginados, pastas de trabalho do *Microsoft Excel*, dados e fontes para serem utilizados no relatório.

O *Power BI* é uma ferramenta de *Business Intelligence* utilizada para gerar painéis e relatórios de forma rápida e intuitiva. É possível conectar a aplicação a

fontes de dados locais, na nuvem e em fontes de *Big Data*, como o *Hadoop* e o *Spark*. A conexão de dados pode ser selecionada em uma lista, como exemplificado na Figura 12.

Figura 12 - Fontes de dados do Power BI

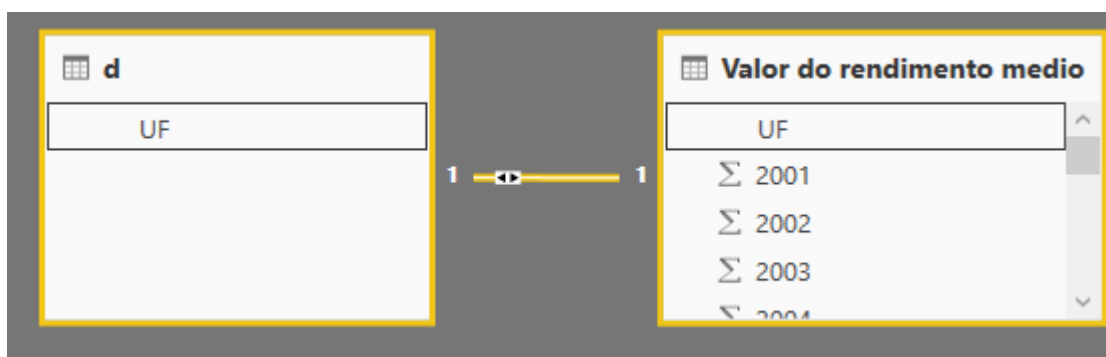


Fonte: Própria(2018)

O *Power BI* oferece recursos para modelagem de dados, ao detectar identificadores iguais nos dados importados, ele realiza a associação automaticamente (Figura 13). Para realizar toda a etapa de obtenção de dados, transformação, combinação e compartilhamento, é utilizado o *Power Query*, que já foi incorporado no *Microsoft Excel* na versão 2010. A linguagem utilizada pelo *Power*

Query é o M, todas as ações realizadas durante o processo de edição são registradas no editor avançado (Figura 14). Outras linguagens utilizadas pelo *Power BI* são DAX (Figura 15) e R, que necessita de instalação prévia na máquina.

Figura 13 -Modelagem com Power BI



Fonte: Própria(2018)

Figura 14 -Tabelas



Fonte: Própria(2018)

Figura 15 - DAX



Fonte: Própria(2018)

O *Power BI* possui as seguintes versões atualmente:

- *Desktop*: versão gratuita que permite grande processamento de dados e acesso a nuvem, desde que possua um e-mail de estudante ou corporativo. Também é possível utilizar a função “Publicar na *web*”, mas o relatório pode ser acessado por qualquer um que possua o *link*.
- *Pro*: é um serviço na nuvem que custa atualmente U\$9,99 por mês, por usuário. A grande diferença desta versão para a *Desktop* é a possibilidade de controlar o acesso aos relatórios, além disso, é possível programar *Gateways* para automatizar atualizações. É necessário que quem compartilha e quem consome o relatório possua a conta *PRO*.
- *On-Premises (local)*: pode ser adquirida a partir de U\$ 3.700,00, essa versão é empresarial mas local. Caso a empresa não esteja preparada para utilizar serviços na nuvem, esta é a versão indicada. Possui integração com o *SQL Server Reporting Services*, o que viabiliza a utilização do *SQL Server Enterprise* para realizar a publicação dos relatórios apenas para a organização. Para utilizar estas funcionalidades é necessário adquirir o *SQL Server Enterprises* ou a versão *Premium* do *Power BI*.
- *Premium (cloud e local)*: o plano se encontra disponível a partir de U\$4.999,00 por mês, esse pacote é vendido de acordo com a capacidade de utilização. É possível trabalhar com dados da nuvem e

com dados locais. As licenças do *Power BI Embedded*, que permite anexar o relatório em aplicações e *SQL Server Reporting Service Power BI Edition* estão incluídas na versão *Premium*. Diferente da versão PRO, não é necessário que o consumidor do relatório possua uma conta PRO, somente autores e editores de conteúdo.

Para a execução do caso de uso deste trabalho, foram utilizadas queries em R e *Power Query*, para persistência no *data warehouse* e realização do ETL, respectivamente. O R já é conhecido por ser utilizado em outras ferramentas, como o Visual Studio. O Power Query é mais popularmente conhecido por ser utilizado no *Excel*. Sua sintaxe é ampla e permite realizar desde operações matemáticas básicas, a grande saneamento de dados. Neste projeto, algumas funções foram amplamente utilizadas, como:

- **Let:** a expressão se encontra no início de todas as queries utilizadas. Sua utilização indica computação de grupos de valores, com armazenamento em nomes, que são utilizados em expressões que serão executadas após sua criação.
- **Excel.Workbook:** é a função que abre conexão com o *Excel*, que foi a fonte de dados utilizada neste trabalho.
- **File.Contents:** função que recebe o arquivo.
- **Table.TransformColumnTypes:** realiza a transformação dos valores de uma coluna.
- **Table.PromoteHeaders:** foi utilizado para promover a primeira linha a cabeçalho, que ao ser importado, vem como sendo uma linha de arquivo comum.

Já a sintaxe utilizada pelo R, era composta por:

- **library:** função que escolhe qual biblioteca será utilizada.
- **odbcDriverConnect:** abre a conexão com o *driver* desejado, nesse caso, SQL Server.
- **odbcClearError:** utilizada para capturar exceções.
- **sqlSave:** salva os dados no banco de dados.
- **close:** fecha a conexão com o banco.

Todas as queries utilizadas no projeto se encontram no *GitHub*¹, que é um repositório *online* de códigos.

O *Power BI* uniu o que havia de melhor no *Microsoft Excel* e no *Reporting Services*, gerando uma ferramenta robusta e de fácil interação.

3.6 Posicionamento da *Microsoft* em pesquisas de mercado

A *Microsoft* tem sido líder no ranking da Gartner, o “*magic quadrant for Business Intelligence and analytics platforms*”, nos últimos 11 anos de pesquisa. Os motivos para essa grande aceitação no mercado são:

- Custo: o *Power BI* é uma das soluções com menor preço do mercado atualmente e oferece o serviço de *desktop* de forma gratuita.
- Facilidade de uso: esse foi o ponto mais apontado por consumidores de conteúdo, que consideram que a ferramenta possui baixo nível de complexidade de utilização. Entretanto, o cliente final acredita que ainda possa haver melhoria na parte administrativa, facilitando a parte de criação quando os dados vêm de fontes mais complexas.
- Visão: o produto tem ficado bem colocado no eixo *Completeness of Vision*, isso tem ocorrido pois a *Microsoft* se preocupa com as pesquisas de qualidade, e tenta sempre melhorar a experiência do usuário final. Os lançamentos de atualizações quase mensais, sua integração com recursos de *Machine de Learning* e com outros produtos *Microsoft*, também tem se mostrado um atrativo para o público.
- Comunidade ativa: a comunidade ativa com diversos vídeos, tutoriais e fóruns, também tem se destacado como um grande diferencial para o usuário final.

Apesar de estar bem posicionada nos últimos anos, alguns pontos ainda são levantados como negativos, como a maturidade do produto. A *Microsoft* conseguiu lidar com diversos cenários complexos, como o problema de nuvem híbrida, mas os usuários não estão satisfeitos com funcionalidades básicas, como a geração de

¹ <https://github.com/FernandaCSantos/publicdash>

tabelas dinâmicas com subtotais. A comunidade é um ponto positivo mas o suporte prestado pela empresa tem aparecido como um ponto negativo na avaliação dos usuários.

Figura 16 - Ranking da Gartner



Fonte: <https://www.atscale.com/blog/gartner-magic-quadrant-for-business-intelligence-bi-2018-the-good-the-bad-the-ugly>

4 Implementando o projeto de BI do IBGE

4.1 Introdução a análise do IBGE

Para realizar a implementação do projeto com *Power BI*, os componentes do grupo procuraram um tema para explorar a ferramenta da *Microsoft*. O assunto escolhido foi educação e como ela está relacionada com a média salarial dos estados brasileiros e verificar o relacionamento existente entre a pobreza e a evasão escolar. Após analisar os dados, será possível deduzir qual será o *ranking* de estados com a maior média salarial nos próximos anos. O site Séries Estatísticas e Séries Históricas², pertencente ao IBGE, disponibiliza algumas séries estatísticas, fornecidas em formato csv ou xls, dependendo da busca realizada. Para a análise que será realizada o período considerado será entre 2007 e 2015, com exceção da evasão escolar, que será analisada até 2016, para realizar a análise preditiva dos próximos anos.

4.2 Entendendo as ocorrências salariais e de escolaridades

O salário mínimo é o menor valor que um funcionário pode receber, resguardado pela CLT (Convenção das Leis Trabalhistas). Essa quantia sofre reajustes anuais com a intenção de fazer com que a pessoa não perca o padrão de vida, mesmo com o avanço da inflação. Alguns estados possuem piso salarial, maior que o salário mínimo, diferente para algumas profissões.

A educação básica é obrigatória dos quatro aos dezessete anos, segundo o inciso I do artigo 208 da Constituição Federal de 1988, mas apesar disso, ainda existe um número expressivo de crianças e adolescentes que evadem o sistema escolar dentro dessa idade.

² <https://seriesestatisticas.ibge.gov.br>

4.3 Elaborando os requisitos da solução para o IBGE

O projeto surgiu do interesse do grupo em verificar como o ensino pode interferir na economia de um estado, como as informações necessárias para realizar esta análise estão em arquivos diferentes, sua centralização será realizada no Power BI.

Ao finalizar a interpretação de dados, deve ser possível responder às seguintes perguntas:

- Qual o estado com a maior média salarial?

Verificar, dentro do período analisado, qual estado apresentou a maior média salarial. Vale ressaltar que o Distrito Federal possui salários mais elevados por sediar órgãos políticos de grande porte, como o Congresso Nacional.

- Qual a porcentagem da população daquele estado se encontra nessa situação salarial?

Dado o local, responder qual a porcentagem de pessoas possuem de fato uma realidade dentro do valor relatado, ou com valor inferior.

- Qual a porcentagem de evasão escolar no ensino médio, nos estados de maior e menor renda verificadas?

Verificar qual a taxa de evasão escolar nesses estados e verificar se há alguma relação entre a renda identificada com a evasão escolar.

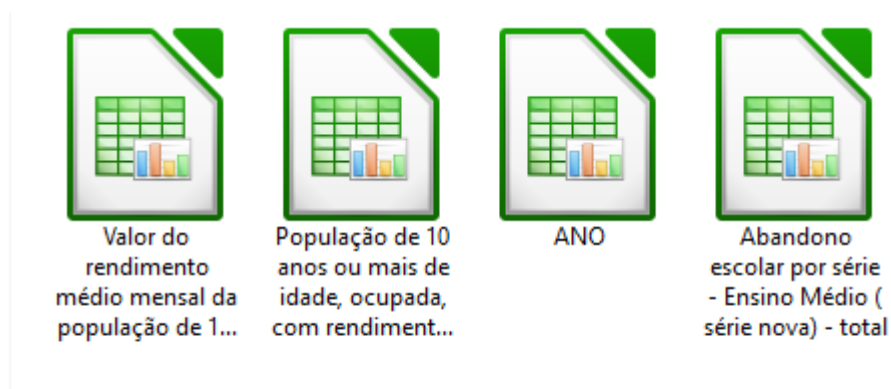
- Qual a previsão dos três estados com a maior média salarial?

Demonstrar segundo os relatórios, qual serão os três primeiros estados do *ranking*.

A Figura 17 mostra os arquivos encontrados no site do IBGE, com exceção do arquivo “ANO” que foi gerado para suprir uma necessidade do projeto.

O *Data Warehouse* está sendo utilizado para centralizar as informações. Seu objetivo no projeto é facilitar o acesso à informação e permitir um acesso fácil aos dados.

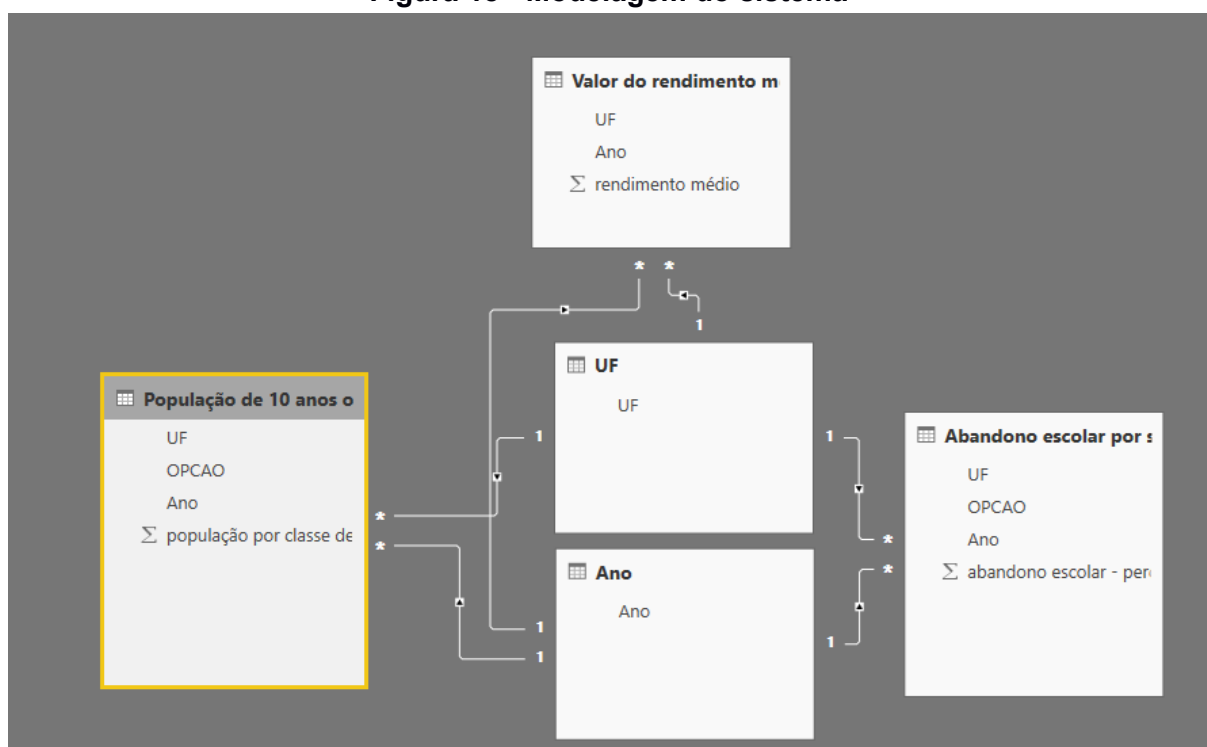
Figura 17 - Arquivos utilizados



Fonte: Própria(2018)

O modelo lógico foi elaborado com a todas as tabelas concentradas em torno de duas tabelas auxiliares criadas para facilitar a integração entre elas. O esquema pode ser verificado na Figura 18.

Figura 18 - Modelagem do sistema



Fonte:Própria(2018)

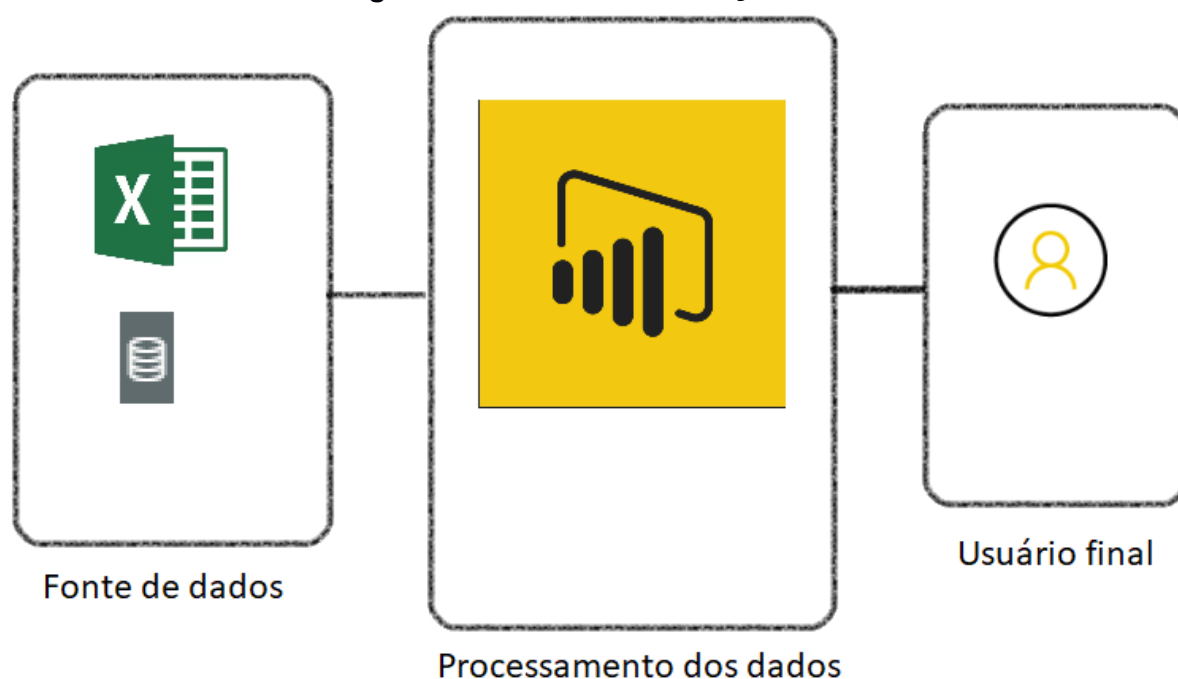
4.4 Elaborando o desenho da solução final

Para a elaboração do projeto, o *Power BI* foi utilizado em todas as camadas.

- O ETL foi realizado utilizando o *Power Query*.
- A carga para o *Data Warehouse* foi feita utilizando a linguagem R com o *Power BI* para realizar as inserções no banco de dados *SQL Server*.

A quantidade de dados utilizados no projeto não justificou a utilização de uma aplicação mais robusta para popular o *Data Warehouse*. O projeto foi centralizado em uma máquina com um processador i5, 8GB de memória RAM e com sistema operacional *Windows 10* com 200GB de memória. A imagem abaixo ilustra o desenho da solução final.

Figura 19 - Desenho da solução final

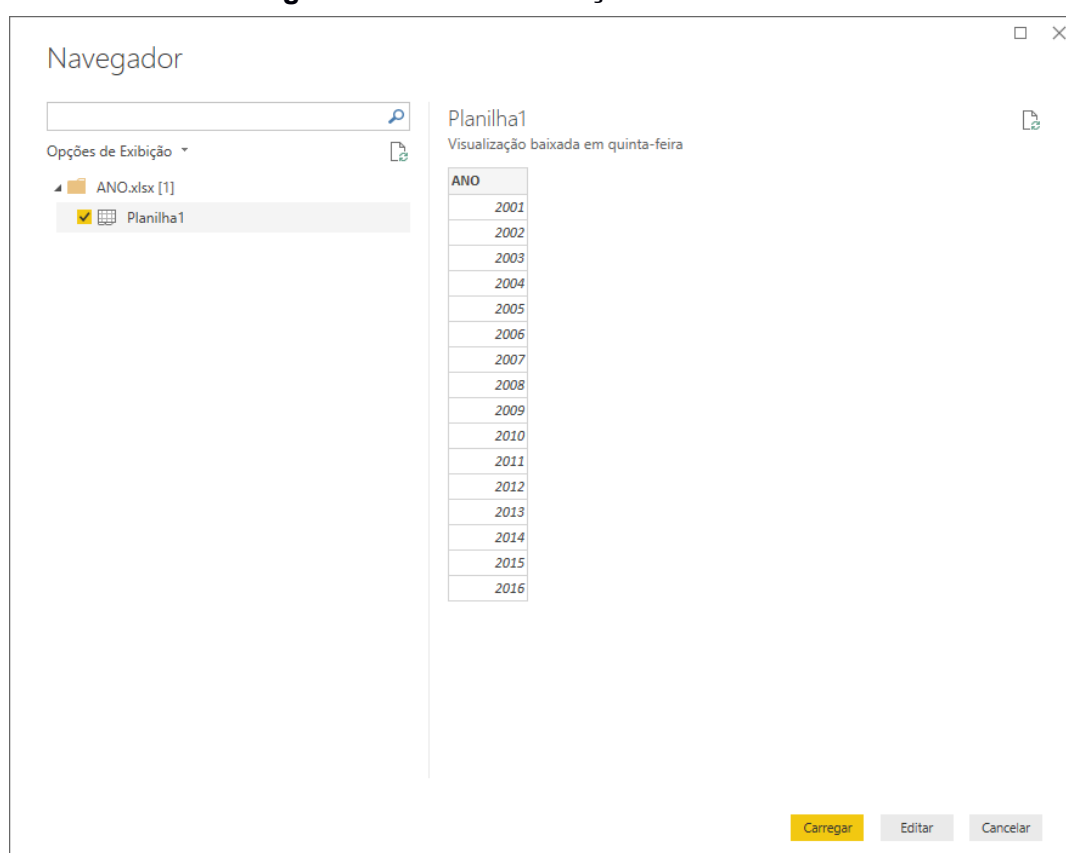


Fonte: Própria utilizando ícones da *Microsoft* (2018)

4.5 Construindo o ETL

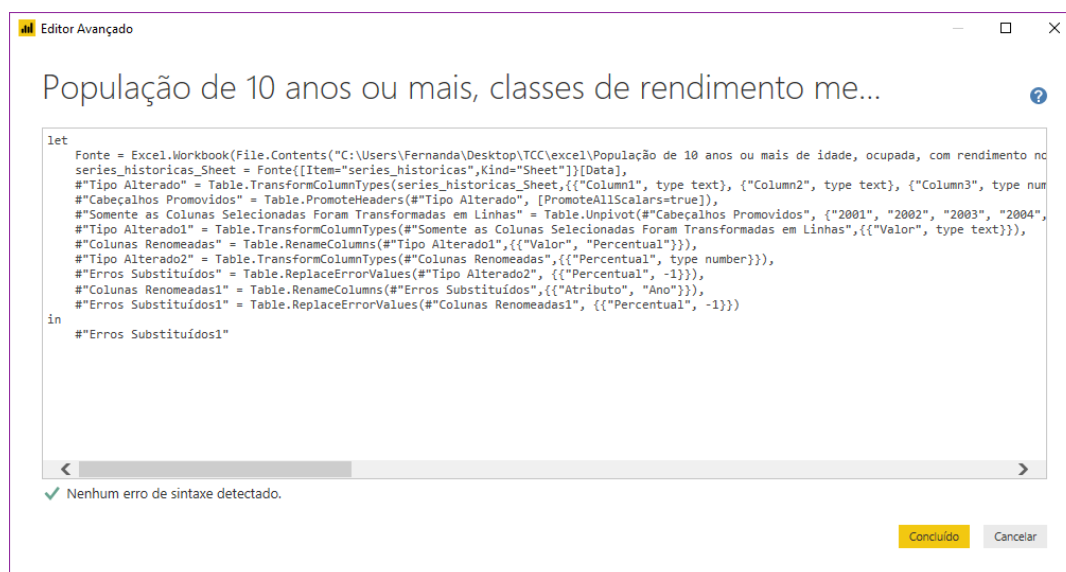
O processo de ETL foi realizado pelo *Power BI*, através do *Power Query*. A extração dos dados pode ser feita através das opções “Obter dados” ou “Fontes recentes” (Figura 12). Depois de selecionar a fonte de dados, será exibida uma pré visualização dos dados (Figura 20). Após a extração, a transformação é realizada através do editor de consultas, que armazena todas as alterações feitas nos dados importados. Desse modo, é possível garantir que os novos registros que serão incluídos, se estiverem formatados do mesmo modo, serão transformados sem a necessidade de um novo saneamento de dados. A Figura 21 exemplifica o código de uma transformação de arquivo do tipo *Excel* que foi usado neste trabalho. Foi possível alterar o tipo dos dados e deixá-los padronizados para a carga no *Data Warehouse* e utilizados em relatórios.

Figura 20 - Pré visualização dos dados



Fonte: Própria(2018)

Figura 21 - Power Query utilizado para fazer o ETL dos dados



Fonte: Própria(2018)

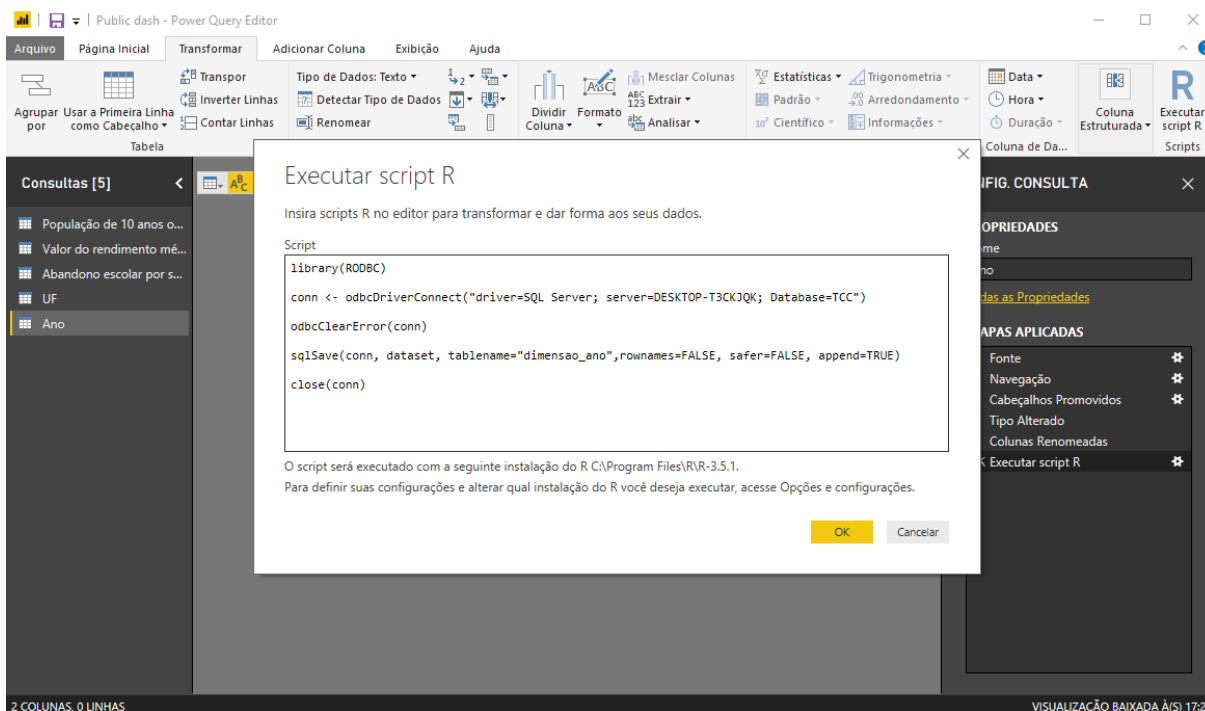
4.6 Construindo o Data Warehouse

Com a consulta no *Power BI* é possível popular o *Data Warehouse*. Foi implementada uma solução usando a linguagem de programação R, como mostra a Figura 22. Uma conexão com o banco de dados, previamente criado no SQL Server, foi estabelecida, e executado um script em R para criar e preencher as tabelas utilizando o ETL do *Power BI*.

Foram criados dois tipos de tabelas:

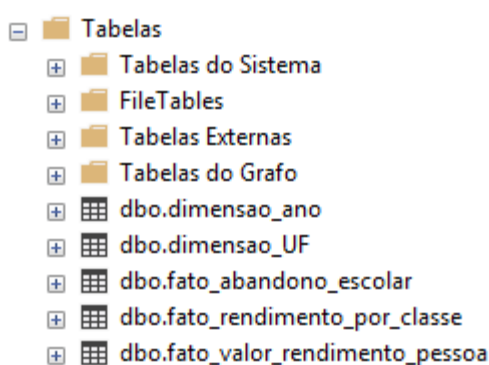
- **Dimensão:** contém dados textuais associados a um evento/fato. Eles descrevem, nesse caso, quando e onde ocorre o fato. As tabelas de dimensões do projeto são: Ano e UF.
- **Fato:** armazena as métricas do negócio. Corresponde ao maior volume de dados de um *Data Warehouse*, tendo em vista que possui granularidade que corresponde ao fato que é observado. As tabelas de fatos são: fato_valor_rendimento_pessoa, fato_rendimento_por_classe e fato_abandono_escolar. Podemos ver a construção tanto da tabela fato quanto da tabela dimensão na figura 23.

Figura 22 - Script na linguagem de programação R



Fonte:Própria(2018)

Figura 23 - Hierarquia do projeto no SQL Server

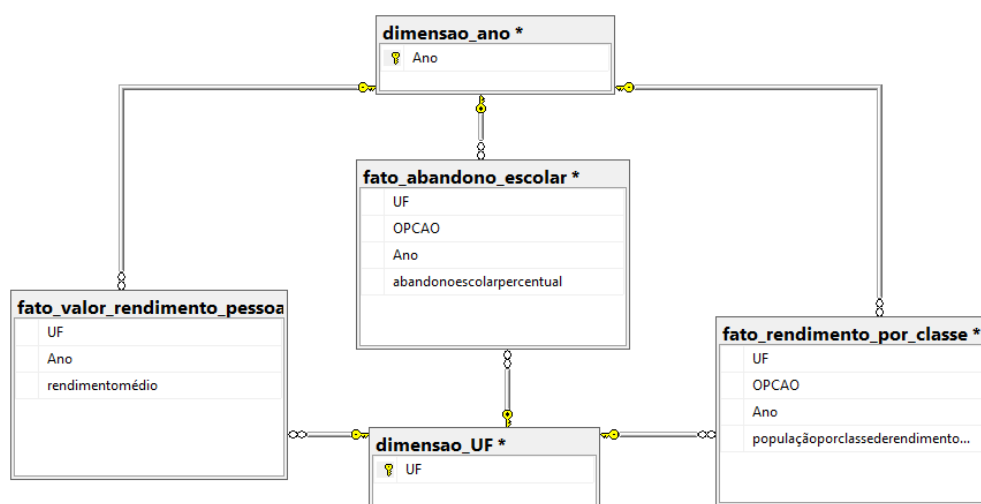


Fonte:Própria(2018)

4.7 Construindo o modelo semântico

A figura 24 mostra o seguinte modelo das tabelas gerado pela criação do *Data Warehouse*. As tabelas dimensão tem relacionamento com as tabelas de fatos, que usam seus dados nas informações de Ano e UF.

Figura 24 - Modelo das tabelas geradas



Fonte:Própria(2018)

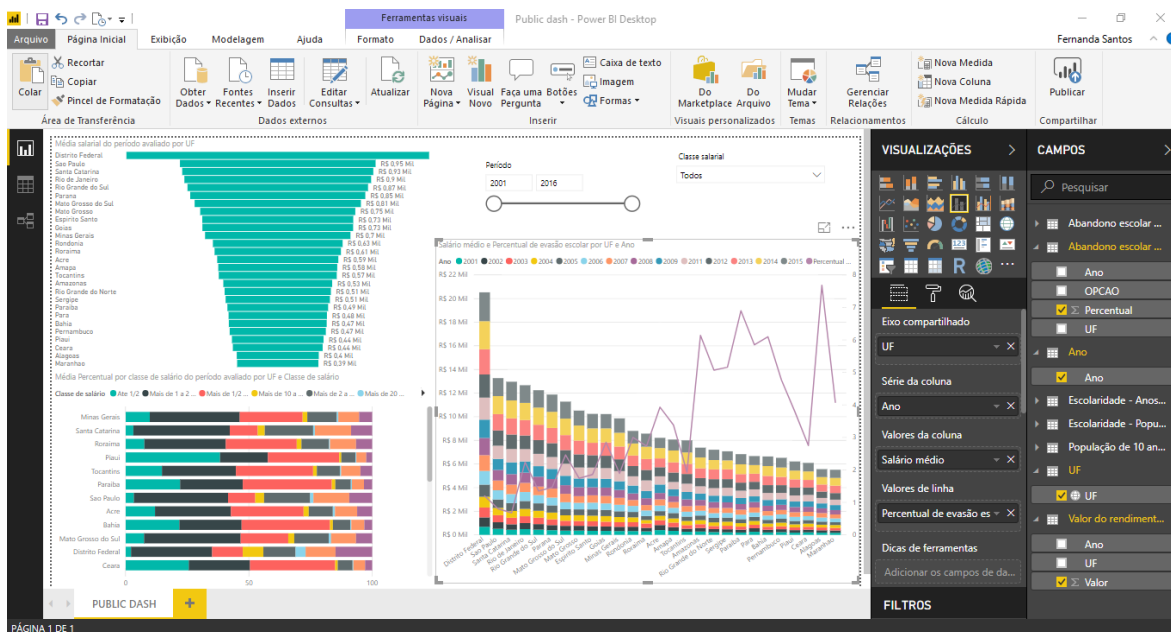
4.8 Construindo os Relatórios e Dashboards no *Power BI*

O *Power BI* foi utilizado para implementar a camada de visualização. Após realizar a importação dos dados, é necessário escolher o tipo de gráfico desejado na aba “Visualizações” e, então, selecionar os valores na aba “Campos” (Figura 25). As tabelas “Ano” e “UF” foram utilizadas para realizar associação entre as outras.

No menu “Exibição”, há a opção “Layout do Telefone” onde é possível organizar os gráficos para dispositivos mobile (Figura 26), que após a sua geração, podem ser publicados no *Report Server*, que é o serviço do *Power BI* na nuvem. Para realizar o acesso, é necessário possuir uma conta no *Office 365*. Os recursos estarão disponíveis de acordo com o plano adquirido.

Para ter acesso aos relatórios na plataforma *mobile*, basta fazer o download do aplicativo do *Power BI*. Caso não possua uma visualização mobile, a exibição será a visualização padrão do *desktop*.

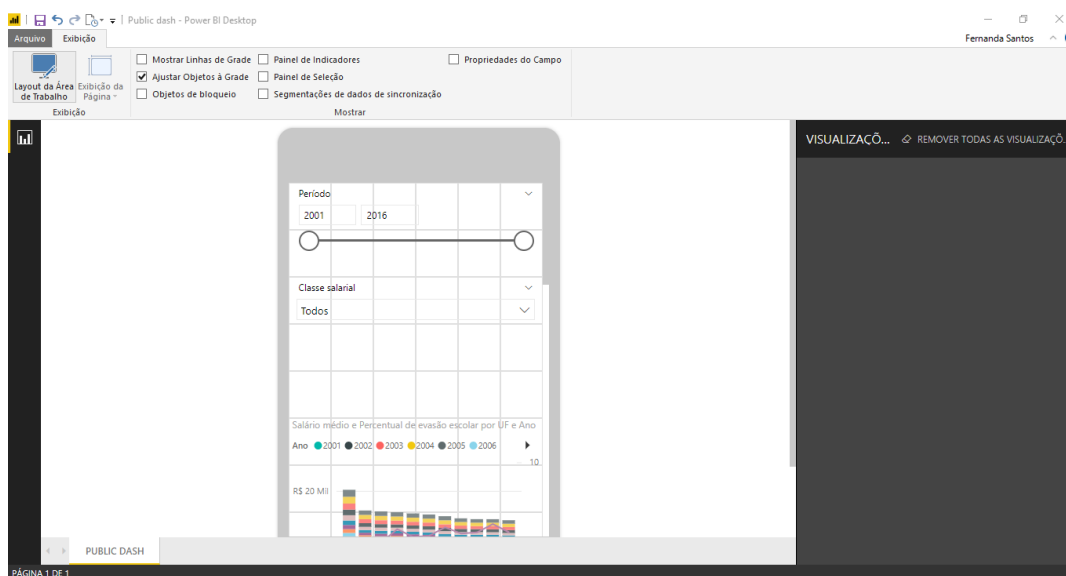
Figura 25 - Visualizações no Power BI



Fonte:Própria(2018)

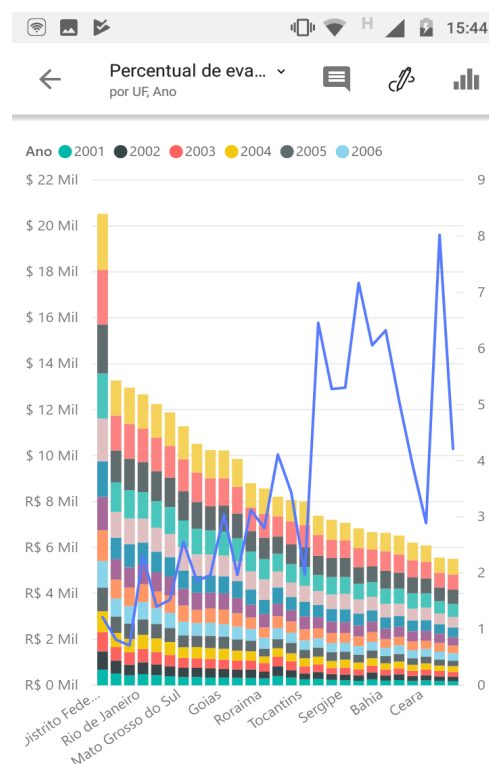
A construção do painel de visualização pode ser feita no navegador após a publicação do relatório. Basta selecionar o gráfico desejado e clicar em “Fixar visual”. É possível montar *dashboard* para *desktop* e *mobile* e realizar buscas no painel, escrevendo a informação desejada em “Faça uma pergunta sobre seus dados” (Figura 27).

Figura 26 - Construção da visualização mobile



Fonte:Própria(2018)

Figura 27 - Painel online na versão *mobile*



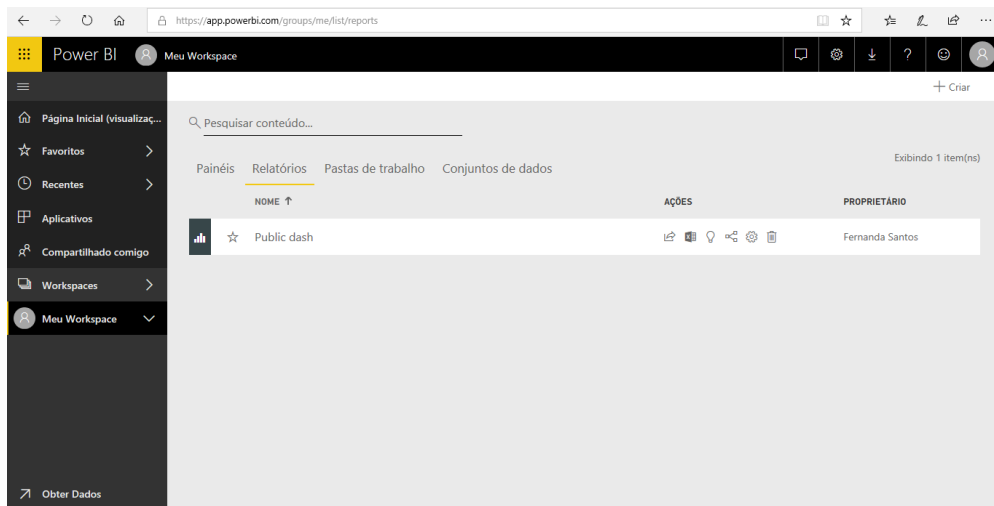
Fonte: Própria(2018)

4.9 Compartilhando relatórios no *Report Server*

É possível compartilhar na nuvem as visualizações geradas na versão *Desktop*. Para publicar o relatório, o usuário deve realizar login com sua conta do *Office 365* e selecionar a opção “Publicar” no canto superior direito. Será necessário informar o espaço de trabalho online, para submeter a publicação. É possível visualizar os relatórios publicados (Figura 28) no site do *Power BI*, em “Meu *Workspace*”, no item “Relatórios”. Os itens podem ser compartilhados e exportados para análise no *Excel*, além disso, também é possível gerar Insights automaticamente (Figura 29). A visualização *online* é muito semelhante a versão *desktop* e permite algumas edições.

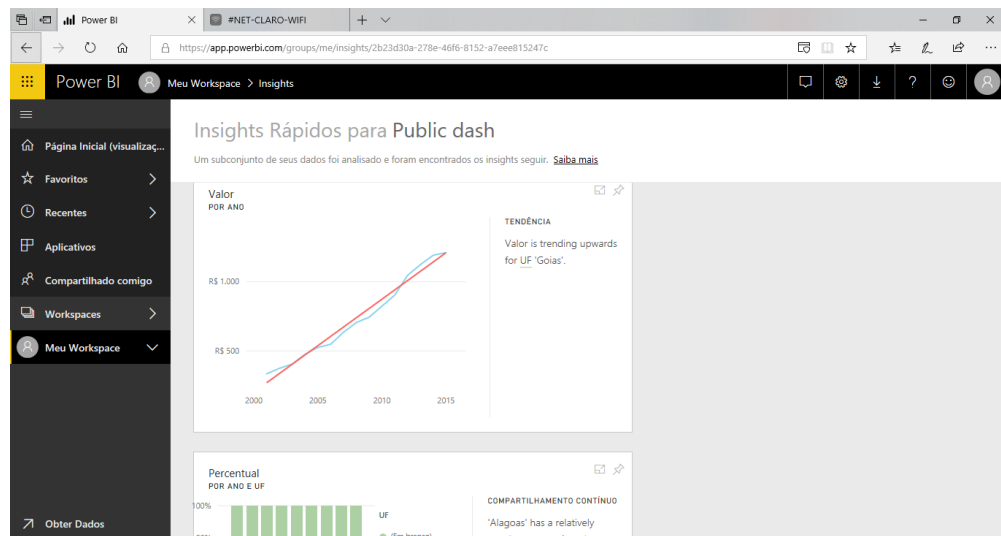
O tipo de conta utilizada para a realização deste trabalho, foi a *Desktop*, sendo assim, o compartilhamento do arquivo exemplificado na Figura 30 não possui restrição de acesso ou compartilhamento.

Figura 28 - Relatórios online



Fonte: Própria(2018)

Figura 29 - Insights



Fonte: Própria(2018)

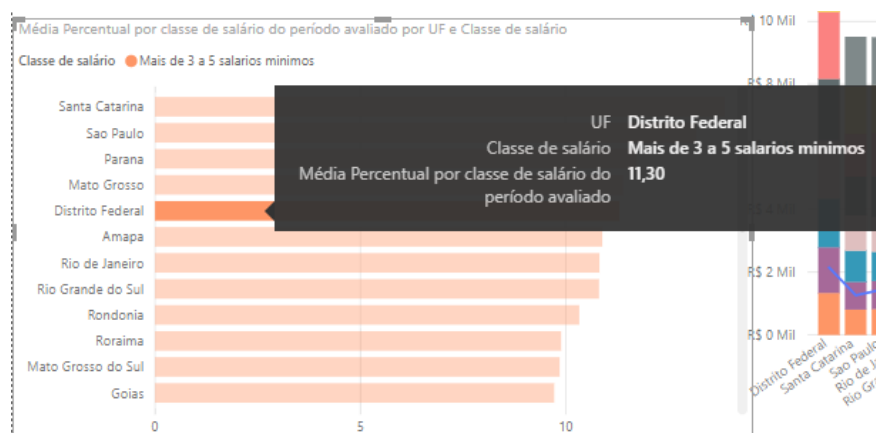
Santa Catarina: a média salarial é de 1,88 salários e 41,30% das pessoas recebia entre um e dois salários (Figura 33). Aproximadamente 13,7% da população recebia menos que essa faixa.

Figura 31 - Maior média salarial



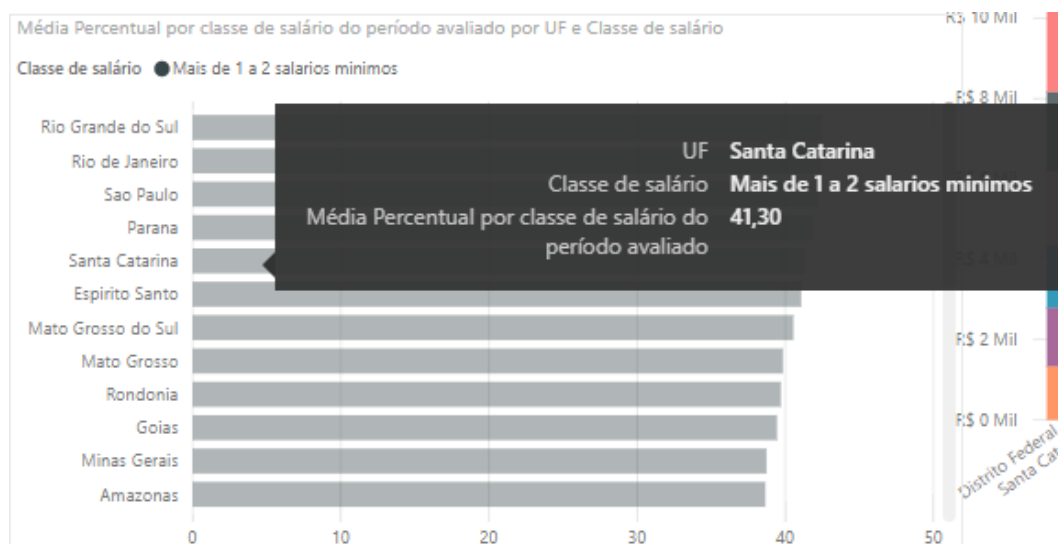
Fonte: Própria(2018)

Figura 32 - Porcentagem de pessoas que ganharam entre três e cinco salários



Fonte: Própria(2018)

Figura 33 - Porcentagem de pessoas que ganharam entre dois e três salários



- Qual a porcentagem de evasão escolar no ensino médio, nos estados de maior e menor renda verificadas?

O estado com a menor renda é Alagoas, com uma média de aproximadamente R\$512 (Figura 34), e possui 8,02% de evasão escolar. O Distrito Federal possui uma taxa de evasão muito menor, 1,21%. Avaliando o gráfico da Figura 35, nota-se que apesar de algumas oscilações, há uma relação entre o salário e o abandono escolar.

Figura 34 - Estado com menor média salarial

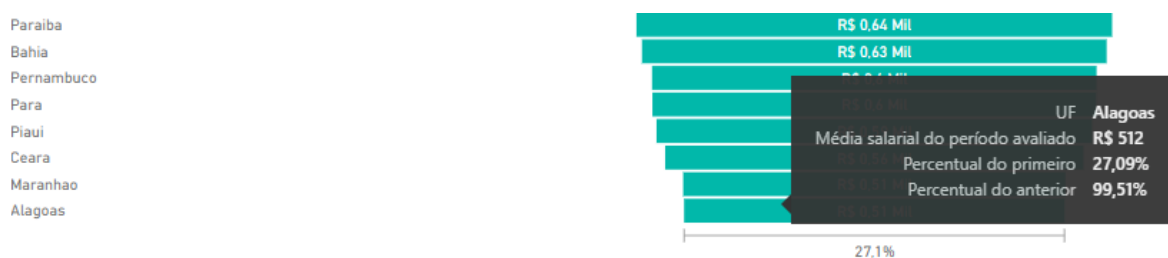
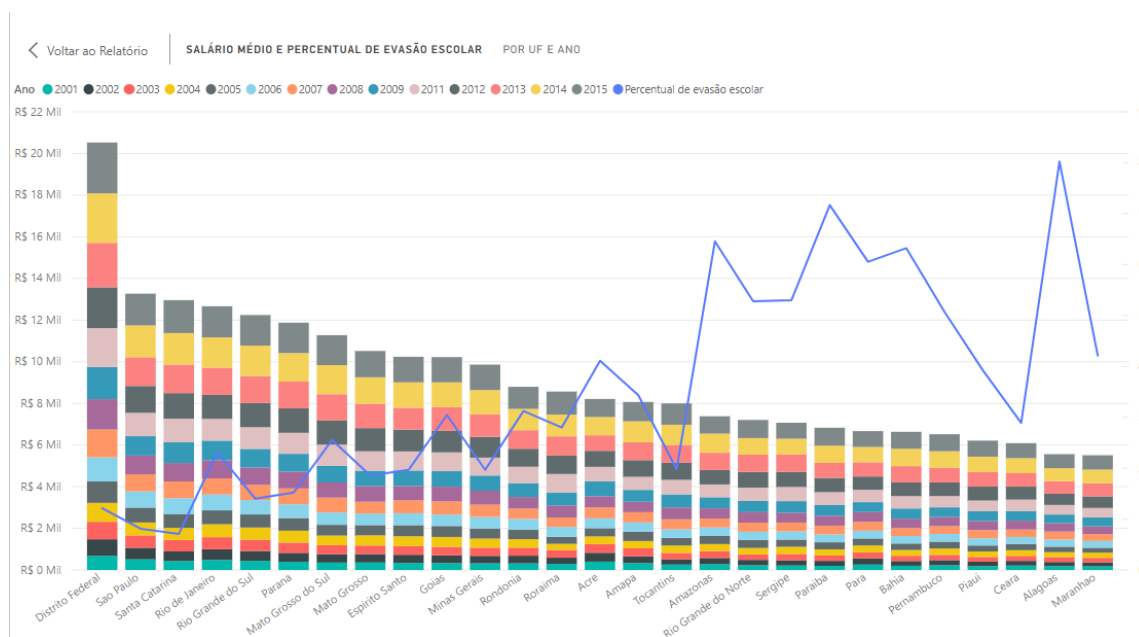


Figura 35 - Salário médio x Percentual de abandono escolar

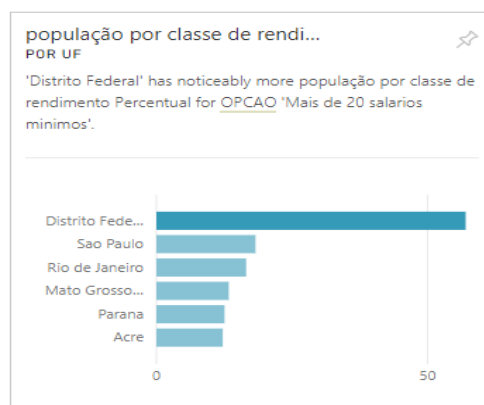


- Qual a previsão dos três estados com a maior média salarial?

Ao gerar os insights a partir do relatório, foi verificado que o Distrito Federal possui o maior salário devido a quantidade da população que recebe mais de vinte salários, o que explica a disparidade salarial (Figura 36). Além disso, o insight exibiu o aumento de rendimento constante com poucas oscilações e a diminuição do abandono escolar (Figura 37).

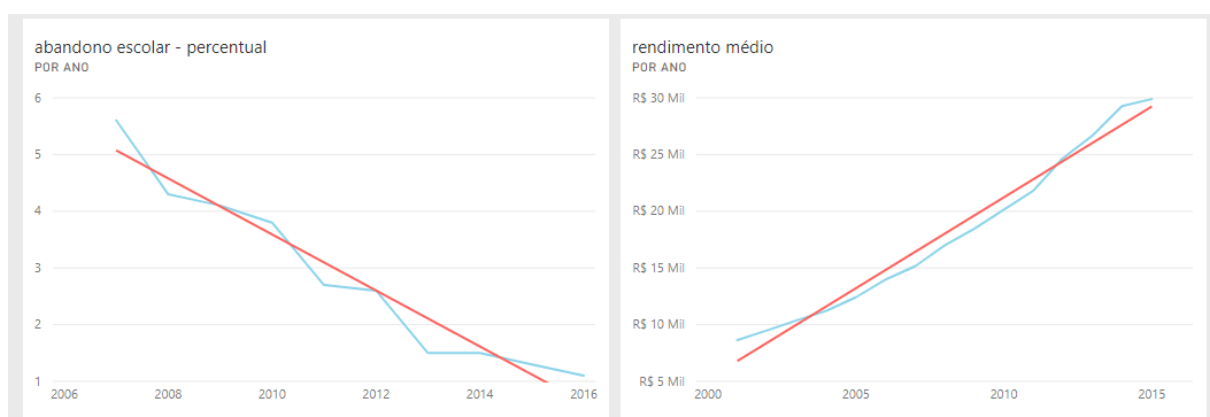
Pernambuco foi o único estado que apresentou mudanças significativas no insight (Figura 38). O aumento salarial entre 2011 e 2015 aconteceu durante o período de queda do abandono escolar, sua posição no relatório de rendimento mensal foi de 24^a para 20^a. Em 2011 o estado apresentava uma taxa de 4,10% de evasão escolar, e em 2015 possuía 2,20%. Avaliando os gráficos já disponibilizados no trabalho, é possível deduzir que o ranking não sofrerá grandes alterações, tendo em suas três primeiras posições: Distrito Federal, São Paulo e Santa Catarina, os dois últimos alternando entre si. Observa-se que, exceto o Distrito Federal, os primeiros colocados no *ranking* são grandes polos econômicos.

Figura 36 - Insight relativo ao percentual de rendimento por classe



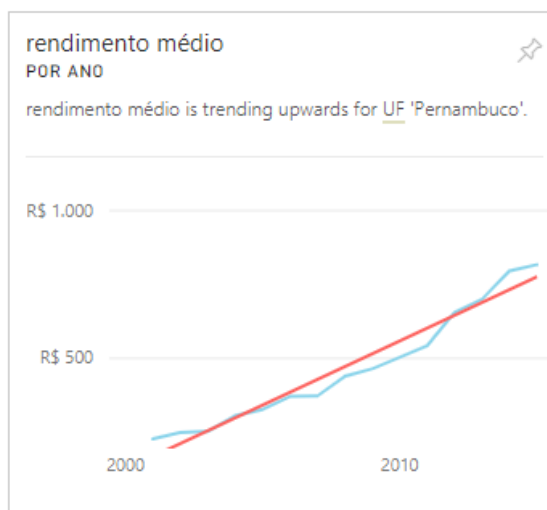
Fonte: Própria (2018)

Figura 37 - Insights de abandono escolar e rendimento médio



Fonte: Própria (2018)

Figura 38 - Crescimento no rendimento médio de Pernambuco



Fonte: Própria (2018)

5 Conclusão

5.1 Considerações finais

Este projeto de graduação buscou introduzir conceitos de *business intelligence* e a ferramenta *Power BI* da *Microsoft* para realizar análise em dados disponibilizados pelo IBGE.

Após a avaliação dos resultados, verificou-se que os estados que ocupam as primeiras posições do ranking, são estados em que a evasão escolar é muito baixa. Deve ser considerado o fato de que o Distrito Federal possui renda mais alta por sediar diversos órgãos políticos de grande porte. São Paulo e Santa Catarina são grandes polos econômicos, o que já elevaria a sua média salarial. Pernambuco foi o único estado, no período avaliado, que se destacou no *insight*. Seu abandono escolar diminuiu e sua remuneração média cresceu. Também foi possível verificar que o estado com a pior renda, é o que possui a maior taxa de evasão escolar. Apesar de ser possível implementar os requisitos e avaliá-los, a falta de dados impediu uma análise mais detalhada. A quantidade da população que nunca foi à escola, a taxa de gravidez precoce, e outros dados podem afetar de forma significativa a renda e não estariam contemplados no cenário mapeado.

O *SQL Server*, apesar de ser uma ferramenta mais robusta e consolidada no mercado, apresenta um nível de complexidade mais alto, o que não era necessário para o porte do estudo de caso. O *Excel* possui algumas similaridades com o *Power BI*, como toda a parte de “Editor de Consultas”, que é igual. O *Excel* possui velocidade maior quando comparado ao *Power BI*, mas a geração de *insights* de maneira automática, o processo de ETL e compartilhamento de relatórios de forma prática e a construção de visualizações mobile, foram fatores que contribuíram para a preferência do *Power BI*.

5.2 Limitações encontradas no decorrer do projeto

Por não possuir um período maior e mais atual, e por não terem sido levados em consideração dados mais minimalistas, a análise não foi capaz de fornecer a estimativa de valor.

5.3 Possibilidade de melhorias futuras

Para trabalhos futuros, é possível estender o escopo da pesquisa e tentar chegar a um valor aproximado de média salarial utilizando cálculos que levem em conta a inflação. A implementação das linguagens Python e R para a análise e geração de visualizações no relatório também pode ser uma evolução para o trabalho. Outro ponto que poderia ser avaliado, é a relação do salário com o nível de ensino da população. A utilização de fontes que possuam dados mais atuais e detalhados é recomendada para realizar análise mais assertiva.

6 Referências

ATKINSON, Malcolm Phillip. **Database access and integration services on the Grid**. The University of Edinburgh, [S.L], fev. 2002. Disponível em: <https://www.researchgate.net/profile/Malcolm_Atkinson/publication/228882346_Database_access_and_integration_services_on_the_Grid/links/5664a90408ae418a786d6c9f.pdf>. Acesso em: 01 nov. 2018.

AUSLAND (2015). **Conceito de business intelligence**. Disponível em: <<http://ausland.com.br/blog/conceito-de-business-intelligence/>>. Acesso em: 13 out. 2018.

CANAL TECH (2014). **Entendendo a modelagem multidimensional**. Disponível em: <<https://canaltech.com.br/business-intelligence/entendendo-a-modelagem-multidimensional-19988/>>. Acesso em: 17 out. 2018.

DEV MEDIA (2009). **O que é data warehouse? arquitetura de um data warehouse**. Disponível em: <<https://www.devmedia.com.br/data-warehouse/12609>>. Acesso em: 13 out. 2018.

EXAME (2018). **Por que o nubank sempre busca cientistas de dados e paga até R\$ 25 mil**. Disponível em: <<https://exame.abril.com.br/carreira/por-que-o-nubank-sempre-busca-cientistas-de-dados-e-paga-ate-r-25-mil/>>. Acesso em: 29 set. 2018.

EXTRA (2018). **Veja cinco profissões que estarão em alta em 2019**. Disponível em: <<https://extra.globo.com/noticias/educacao/vida-de-calouro/veja-cinco-profissoes-que-estao-em-alta-em-2019-22831370.html>>. Acesso em: 29 set. 2018.

FOLHA (2018). **Oportunidades aumentam para quem estuda ciência de dados**. Disponível em: <<https://www1.folha.uol.com.br/sobretudo/carreiras/2018/06/1971998-oportunidades-aumentam-para-quem-estuda-ciencia-de-dados.shtml>>. Acesso em: 29 set. 2018.

KIMBALL, R. e Ross, M. (2013), **The Data Warehouse Toolkit**, 3th edition.

LOH, Stanley. **Bi na era do big data para cientistas de dados**: indo além de cubos e dashboards na busca pelos porquês, explicações e padrões. 1a edição ed. Porto Alegre: [s.n.], 2014. 158 p.

GUIA TRABALHISTA. **Salário mínimo - tabela de valores**. Disponível em: <http://www.guiatrabalhista.com.br/guia/salario_minimo.htm>. Acesso em: 10 nov. 2018.

MALCOLM, G. e O'Neill, C. (2014), **Implementing a SQL Data Warehouse**.

MICROSOFT POWER BI (2018). **Gartner recognizes microsoft as a leader in analytics and bi platforms for 11 consecutive years**. Disponível em: <<https://powerbi.microsoft.com/en-us/blog/gartner-recognizes-microsoft-as-a-leader-in-analytics-and-bi-platforms-for-11-consecutive-years/>>. Acesso em: 26 out. 2018.

MICROSOFT (2018). **Power bi free vs pro**. Disponível em: <<https://docs.microsoft.com/en-us/power-bi/service-features-license-type>>. Acesso em: 10 nov. 2018.

NITEO. **Self-service bi. o poder na mão do usuário..** Disponível em: <<http://www.niteo.com.br/self-service-bi-o-poder-na-mao-do-usuario/>>. Acesso em: 28 set. 2018.

OFICINA DA NET (2016). **O que é data science?**. Disponível em: <<https://www.oficinadanet.com.br/post/16919-o-que-e-data-science>>. Acesso em: 28 set. 2018.

RAFAEL PITON. **Data warehouse - tipos de métricas**. Disponível em: <<https://rafaelpiton.com.br/blog/data-warehouse-tipos-metricas/>>. Acesso em: 22 out. 2018.

RAFAEL PITON. **Data warehouse – o que é tabela fato (e seus tipos)**. Disponível em: <<https://rafaelpiton.com.br/blog/data-warehouse-tipos-fatos/>>. Acesso em: 19 out. 2018.

RAFAEL PITON. **Tabela dimensão: os 5 tipos que você deve conhecer.**

Disponível em:

<<https://rafaelpiton.com.br/blog/data-warehouse-tipos-dimensoes/>>. Acesso em: 22 out. 2018.

Stephens, M. (2011), **A importância de Business Intelligence.**

Disponível em

<<http://www.blog.msbrasil.com.br/a-importancia-do-business-intelligence>>

Acessado em 10 de outubro de 2017.

TRANSFORMAÇÃO DIGITAL (2018). **O que é análise preditiva e como ela ajuda na tomada de decisões?**. Disponível em:

<<https://transformacaodigital.com/o-que-e-analise-preditiva-e-como-ela-ajuda-na-tomada-de-decisoes/>>. Acesso em: 28 set. 2018.