

# **Análise quanto a moradia para apoiar análises da PRAE (moradia estudantil) e setor imobiliário da grande Florianópolis.**

**Diego Feijó, Gabriel Stedile, João Campos e João Messina**

**Curso de Bacharelado em Sistemas de Informação - Universidade Federal de  
Santa Catarina (UFSC)**

**Prof. José Leomar Todesco**

[diego.feijo13@gmail.com](mailto:diego.feijo13@gmail.com), [gabriel.stedile9@gmail.com](mailto:gabriel.stedile9@gmail.com),  
[joaovictorcampos@gmail.com](mailto:joaovictorcampos@gmail.com), [joao.r.messina@gmail.com](mailto:joao.r.messina@gmail.com)

**Resumo** - Este trabalho tem como objetivo analisar os dados do vestibular da Coperve entre os anos de 2008 a 2012 e a partir deles implementar um Data Mart a fim de possibilitar a investigação através de técnicas de Data Warehouse (DW) e responder alguns questionamentos como: Qual a faixa de poder aquisitivo dos candidatos por cidade, que foram aprovados no vestibular ; Quais candidatos de baixa renda estão aptos a solicitar moradia estudantil para PRAE; Quais candidatos possuem capacidade financeira para alugar ou comprar imóvel na grande Florianópolis; e Existe uma relação de candidatos por ano ou vestibular aptos ao mercado imobiliário ou a PRAE. O processo de Extração, Transformação e Carga (Extract Transform Load - ETL) é uma das técnicas de DW responsável pela extração de dados de várias fontes, sua posterior limpeza, otimização e inserção no DW, onde neste artigo foi demonstrado seu funcionamento através da utilização da ferramenta Kettle. Por fim os dados gerados para serem mostrados aos usuários (front-end) foram apresentados através da ferramenta Power BI.

**Abstract** - This work aims to analyze Coperve's vestibular data between the years 2008 and 2012 and from them to implement a Data Mart to enable the investigation through Data Warehouse (DW) techniques and answer some questions such as: What is the range of purchasing power of candidates per city, which were approved in the entrance exam; Which low-income candidates are eligible to apply for student housing for PRAE; Which candidates have the financial capacity to rent or buy property in the greater Florianópolis; e There is a list of candidates per year or college entrance examination or the PRAE. The Extract Transform Load (ETL) process is one of the DW techniques responsible for the extraction of data from various sources, its subsequent cleaning, optimization and insertion in the DW, where in this article it has been demonstrated its operation through the use the Kettle tool. Finally the data generated to be shown to the users (front-end) were presented through the Power BI tool.

**Palavras-chave:** Data Mart, Data Warehouse, Pentaho, Kettle, Vestibular.

## **1. Introdução**

A universidade Federal de Santa Catarina (UFSC, 2019) tem sua sede principal na cidade de Florianópolis, Santa Catarina. É classificada entre as melhores faculdade do Brasil, com mais de 100 cursos, entre presenciais e a distância (DPGI, 2018). Oferece mais de 6 mil vagas em seu vestibular (DPGI, 2018), se tornando responsável por um grande fluxo de pessoas e potenciais interessados em morar na Grande Florianópolis, principalmente na Trindade, bairro que abriga o centro.

Diferente da tendência geral do setor imobiliário nacional, que passou uma crise nos últimos tempos, a bairro da Trindade passa por uma alta especulação, com crescimento e interesse de imóveis na região com aproximadamente 15% de valorização, sendo os principais motivos o crescimento da população da região e a presença da Universidade que promove uma circulação de estudantes, professores e funcionários (RIBEIRO, 2019).

Entre inúmeras atribuições, sendo uma delas, apoiar a execução de políticas estudantis relativas a moradia estudantil, a PRAE - Pró-Reitoria de Assuntos Estudantis - possui programas de auxílio moradia e disponibilização de moradia estudantil, que tem como objetivo proporcionar melhores condições para permanência na universidade (PRAE, 2019). No entanto, devido a crescente demanda, as políticas da PRAE não conseguem beneficiar a todos os que necessitam. Além disso a própria disponibilidade e valorização imobiliária dos arredores, como exposto acima, dificultam a obtenção de uma moradia por parte dos alunos.

A inovação tecnológica, principalmente através da área de BI (Business Intelligence), vem se tornando um importante aliado para impulsionar os negócios de forma mais precisa e segura, ao melhorar o processo de tomada de decisão e focar na agilidade e análise de dados (DEVMEDIA, 2013).

Nesse contexto, um Data Warehouse é uma ferramenta importante para alcançar o público-alvo de maneira mais eficiente e gerar relacionamentos mais qualificados, como por exemplo potenciais compradores ou inquilinos de imóveis.

Tendo em vista o exposto acima, o objetivo deste projeto é desenvolver um Data Mart (DM) e uma camada de apresentação dos dados gerados que em conjuntos, constitua uma plataforma de inteligência de mercado imobiliário,

## **2. Materiais**

Como fonte dos dados do trabalho, foi utilizado a base de dados da COPERVE - Comissão Permanente do Vestibular da UFSC. Foi analisado o banco de dados dos vestibulares da UFSC no intervalo de 2008 a 2012, a fim de obter as informações necessárias para executar este projeto.

## **3. Métodos**

Data Warehouse, conhecido também como armazém de dados ou ainda depósito de dados tem como objetivo agregar informações de uma ou mais fontes de dados, estas geralmente tendo um caráter heterogêneo, para logo após tratar, formatar e consolidar em uma única estrutura de dados (FERREIRA, 2010). Tem como principais características ser um repositório de dados não volátil, variável através do tempo, buscando a integração e contribuindo para análise e tomada de decisão. Estruturalmente um DW é a união de um ou mais DataMarts, estes sendo repositórios de dados que consolidam apenas as informações de uma determinada área.

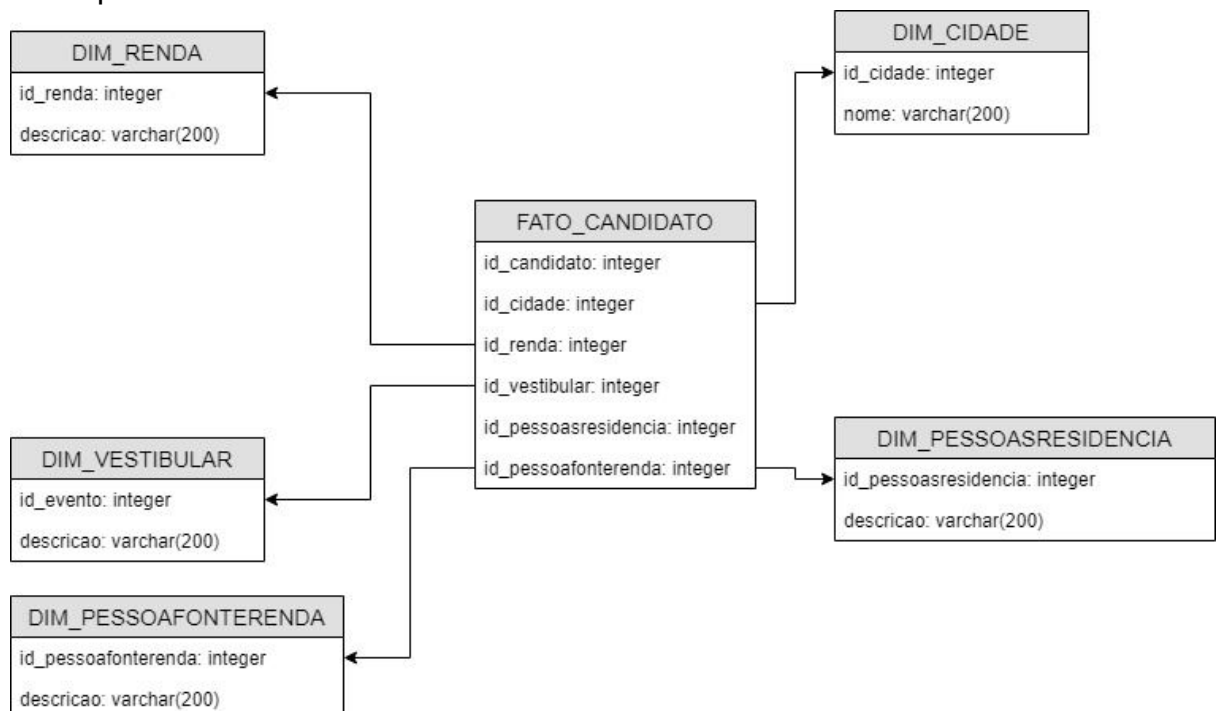
Para realização deste projeto foi utilizado uma estratégia de construção do DW de forma Bottom-up, onde primeiro os Data Marts são desenvolvidos até gerar um DW final. Como este trabalho possui um escopo pequeno e específico, foi criado um Data Mart que resolva o que foi proposto e se economiza custos e tempo para o desenvolvimento de um DW completo, caso fosse necessário.

O primeiro passo para se incorporar os dados no DW é conhecido como processo ETL, que resumidamente se traduz em extração, transformação e inclusão dos dados no DW. Neste trabalho esse aspecto é realizado pela ferramenta pentaho (Kettle), por apresentar uma interface amigável, gráfica e com boa usabilidade (PENTAHO, 2017).

Já a ferramenta de front-end adotada foi a Power Bi, desenvolvida pela Microsoft, que gera dashboards, gráficos e relatórios em poucos segundos, de forma simples e intuitiva.

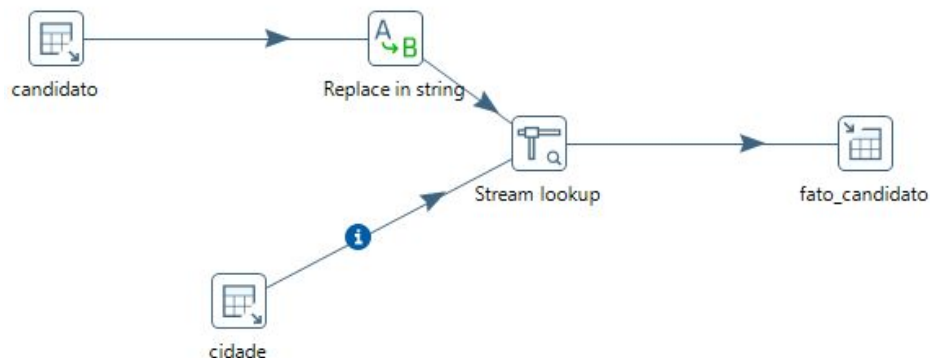
#### 4. Metodologia

Foi elaborado um esquema estrela - modelagem dimensional - para possibilitar o posterior desenvolvimento da base de dados analítica:



Para poder popular a dimensão cidade, foi necessário transformar os dados extraídos do banco de dados original, da tabela candidato. A informação de cidade do candidato era apresentada somente como uma propriedade do modelo, e não um objeto forte, logo foi necessário tratar os nomes quanto a erros de digitação e duplicidade. Portanto antes de popular o *fato\_candidato*, foram tratadas as strings. Para tal, a tabela dimensional *dim\_cidade*, já populada, foi carregada com um *StreamLookup* no *Spoon*, que faz com que os dados estejam acessíveis em memória antes de popular o fato. Então foram substituídos os nomes das cidades pelas suas chaves correspondentes.

Abaixo segue uma ilustração do programa Spoon:



Foi encontrado um pequeno empecilho à análise, pois o dado dos candidatos quanto a sua renda já vieram de forma qualitativa, impedindo uma medição mais direta.

Para analisar quais candidatos estariam aptos neste caso, partimos do princípio que a norma da PRAE prevê o auxílio moradia aos alunos matriculados cuja renda familiar seja de até 1,5 salário mínimo.

Depois, sabendo que cada candidato está em apenas 1 **Faixa de Renda Bruta** (a soma da renda de sua família), sendo elas:

- 1) Até 1 salário mínimo;
- 2) Acima de 1 até 3 sal. mín.;
- 3) Acima de 3 até 5 sal. mín.;
- 4) Acima de 5 até 7 sal. mín.;
- 5) Acima de 7 até 10 sal. mín.;
- 6) Entre 10 e 20 sal. mín.;
- 7) Entre 20 e 30 sal. mín.;
- 8) Acima de 30 sal. mín.

Cada candidato também possui 1 valor de **Pessoas na Residência** (familiares cuja renda seria distribuída), sendo as opções:

- 1 pessoa;
- 2 pessoas;
- 3 pessoas;
- 4 pessoas;
- 5 pessoas;
- Acima de 5 pessoas.

Com estes dados, podemos inferir, por exemplo, que qualquer candidato na primeira faixa de renda (até 1 salário mínimo) estará apto ao auxílio, independente de quantas pessoas morem em sua residência.

Por outro lado, candidatos na terceira faixa de renda (Acima de 3 até 5 salários mínimos) com certeza não estarão aptos se possuírem apenas 1 pessoa na residência (3 salários / 1 pessoa = 3 salários); mas candidatos nesta faixa, com mais pessoas, potencialmente estarão aptos (3 salários / 2 pessoas = 1,5 salário).

Observamos ainda outro exemplo, onde um candidato está na terceira faixa de renda, com 3 pessoas na residência; este candidato seria avaliado como potencialmente apto, pois sua renda bruta poderia ser exatamente 3,1 salários mínimos, o que o deixaria com renda per capita de 1,03 salários. Porém, ainda seria necessária análise detalhada de seu cadastro sócio-econômico, pela possibilidade de sua renda bruta ser 4,9 salários mínimos, o que torna sua renda per capita de 1,63 salários, reprovando-o para o auxílio.

Após esta análise das variáveis, definimos como critério para o candidato estar "apto" (potencialmente) ao auxílio, se ele atender a um dos seguintes pares de condições de Renda/Pessoas:

<b>Faixa de Renda Bruta</b>	<b>Pessoas na Residência</b>
1) Até 1 salário mínimo.	Quaisquer pessoas na Residência.
2) Acima de 1 até 3 sal. mín.	Quaisquer pessoas na Residência.
3) Acima de 3 até 5 sal. mín.	2 pessoas; 3 pessoas; 4 pessoas; 5 pessoas; 6 pessoas.
4) Acima de 5 até 7 sal. mín.	4 pessoas; 5 pessoas; 6 pessoas.
5) Acima de 7 até 10 sal. mín.	5 pessoas; 6 pessoas.
6) Entre 10 e 20 sal. mín.	6 pessoas.
7) Entre 20 e 30 sal. mín.	6 pessoas.
8) Acima de 30 sal. mín.	6 pessoas.

Alternativamente, definimos que um candidato provavelmente tem capacidade financeira para aluguel ou compra de imóvel em Florianópolis se atender a um dos seguintes pares de condições:

<b>Faixa de Renda Bruta</b>	<b>Pessoas na Residência</b>
3) Acima de 3 até 5 sal. mín.	1 pessoa.
4) Acima de 5 até 7 sal. mín.	1 pessoa; 2 pessoas; 3 pessoas.
5) Acima de 7 até 10 sal. mín.	1 pessoa; 2 pessoas; 3 pessoas; 4 pessoas.

6) Entre 10 e 20 sal. mín.	1 pessoa; 2 pessoas; 3 pessoas; 4 pessoas; 5 pessoas; 6 pessoas.
7) Entre 20 e 30 sal. mín.	1 pessoa; 2 pessoas; 3 pessoas; 4 pessoas; 5 pessoas; 6 pessoas.
8) Acima de 30 sal. mín.	1 pessoa; 2 pessoas; 3 pessoas; 4 pessoas; 5 pessoas; 6 pessoas.

A partir do modelo acima foi gerado um banco de dados MySQL e realizada uma série de transações com o apoio da ferramenta Kettle da suíte Pentaho para carga do Data Mart. Para o processo de ETL, foi realizado a seguinte sequência geral de passos:

- 1 - Extração dos dados das tabelas de respostas dos questionários do vestibular
- 2 - Realização de limpeza de alguns dados, principalmente nos relacionados a cidades.
- 3 - Carga de Dimensões.
- 4 - Carga dos Fatos.
- 5 - Desenvolvimento de jobs para automação no Kettle.

Após a carga completa dos dados, foi utilizado a ferramenta Power BI para gerar e observar as análises relacionadas ao domínio do problema proposto, através de gráficos e relatórios que dão suporte a tomada de conclusões.

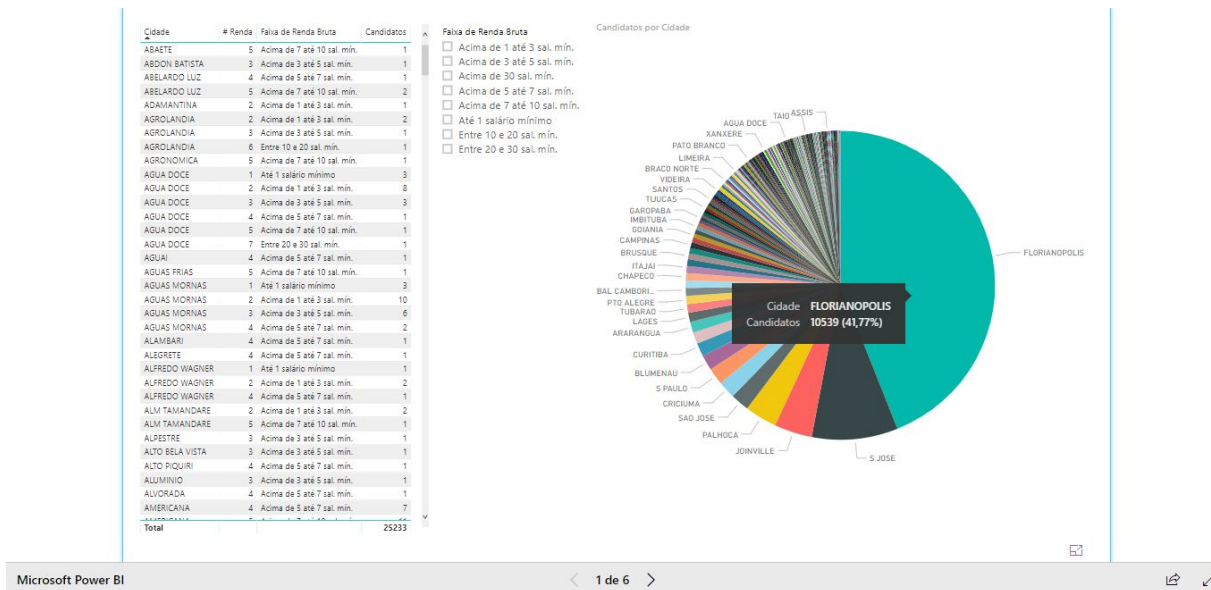
Pelo Power Bi, para auxiliar visualmente a análise dos dados, organizamos os dashboards e gráficos em 6 telas:

- Candidatos por Cidade;
- Candidatos por Ano e Faixa de Renda;
- Candidatos Aptos (consolidado);
- Candidatos Aptos (listagem);
- Candidatos com Capacidade Financeira (consolidado);
- Candidatos com Capacidade Financeira (listagem).

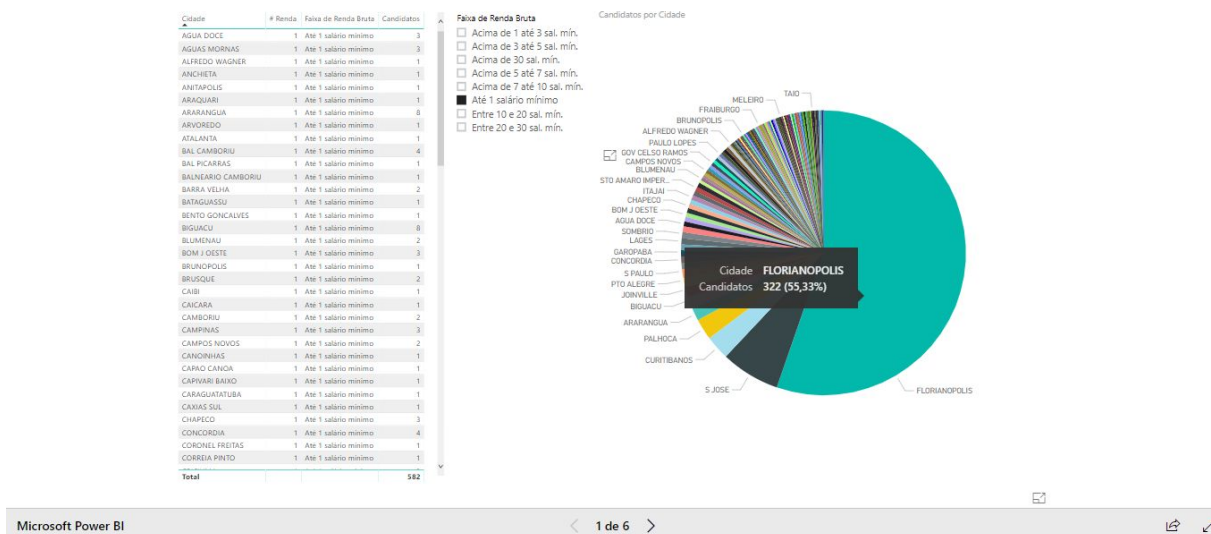
Finalmente, buscamos as respostas às perguntas propostas inicialmente, a partir do ponto de vista do usuário da aplicação.

## 5. Resultados

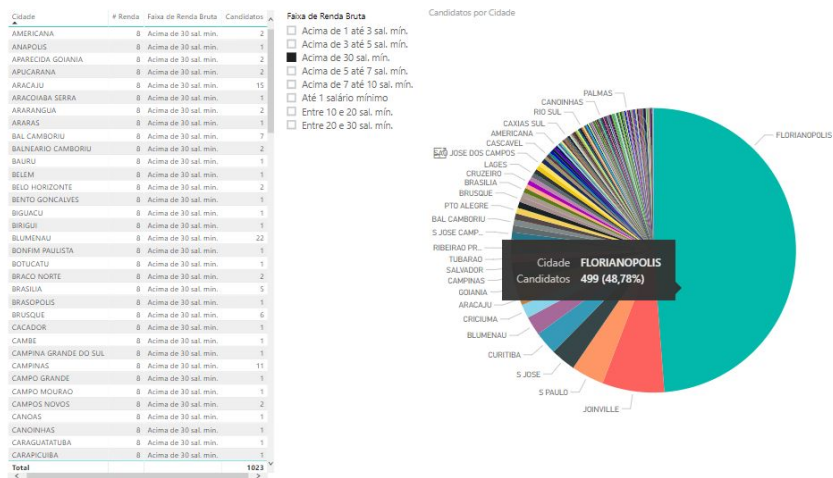
Na primeira tela podemos ver a quantidade de candidatos para cada faixa de renda em cada cidade. Ao lado, um gráfico pizza mostrando o total de candidatos por cidade. Nele, podemos ver que dos 25233 candidatos aprovados nos 5 vestibulares amostrados, a cidade com o maior número de candidatos é **Florianópolis com 41,77%** dos candidatos, seguida de **São José com 8,54%**.



Adicionalmente, ao filtrar pela faixa de renda "**Até 1 salário mínimo**", vemos que Florianópolis compõe 55,33% dos candidatos desta faixa de renda, seguida de São José com 6,7%.

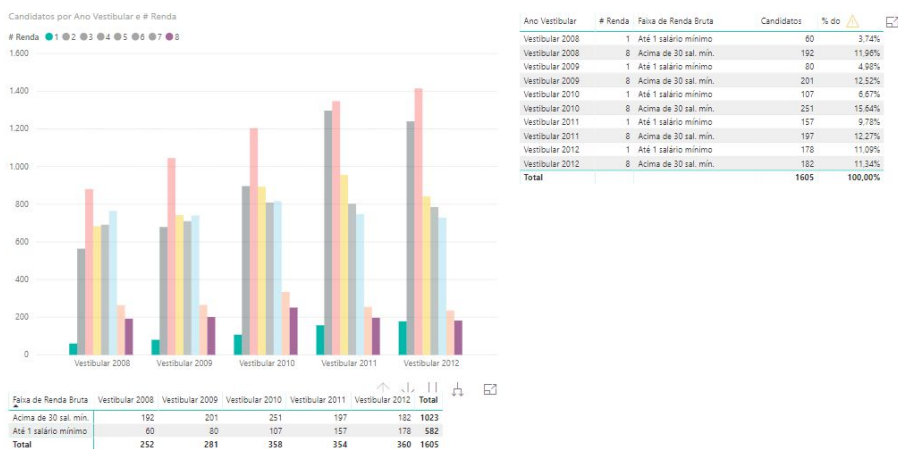


Enquanto para a última faixa de renda "**Acima de 30 sal. mín.**", Florianópolis compõe 48,78% dos candidatos, seguida de Joinville com 7,04%, São Paulo com 3,71% e São José agora com 2,83%.



Com isso podemos inferir que candidatos de baixa renda têm mais tendência a já morar nas proximidades, enquanto candidatos oriundos de outras metrópoles tendem a ter uma renda mais alta.

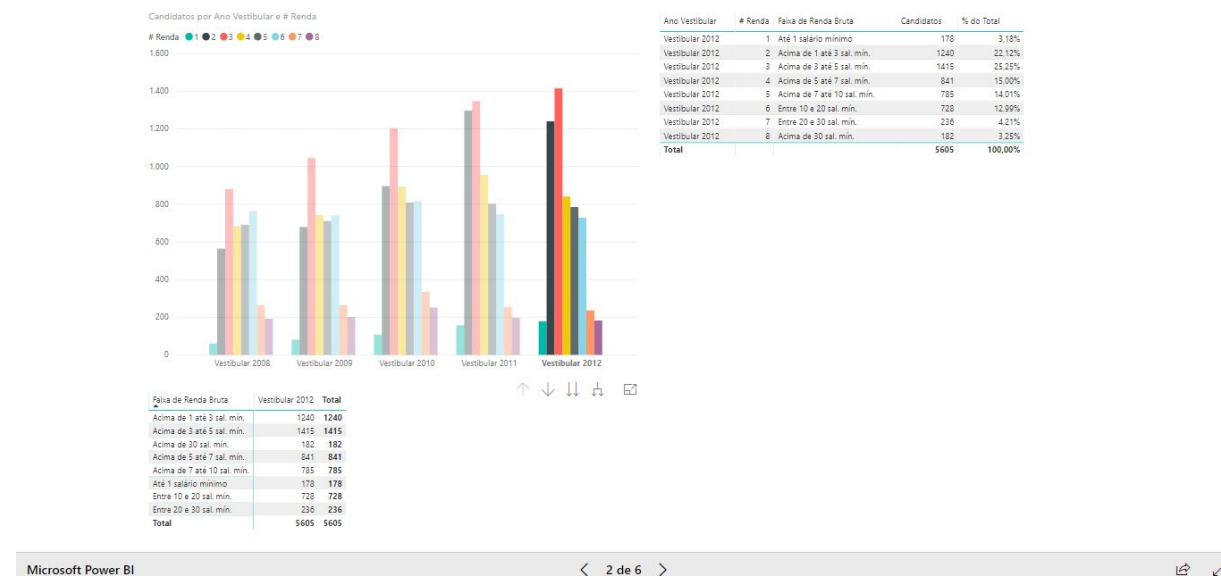
Na segunda tela propomos uma visão dos candidatos por ano do vestibular e por renda, possibilitando também uma análise *drill down* dos vestibulares. Ao selecionar a primeira e última faixa de renda, vemos pelo gráfico de barras e pela matriz abaixo que a quantidade de candidatos com potencialmente a menor renda teve uma ascensão nos 5 anos, enquanto que a quantidade de candidatos com potencialmente a maior renda ficou relativamente constante.



Comparando o primeiro e o último ano, vemos que em 2008 houveram 1,46% dos candidatos declarados na faixa de renda bruta "Até 1 salário mínimo", contra



4,69% dos candidatos na faixa "Acima de 30 sal. min"; já em 2012 essas faixas ficaram mais equilibradas, tendo 3,18% de candidatos "Até 1 salário mínimo" e 3,25% de candidatos "Acima de 30 sal. min."



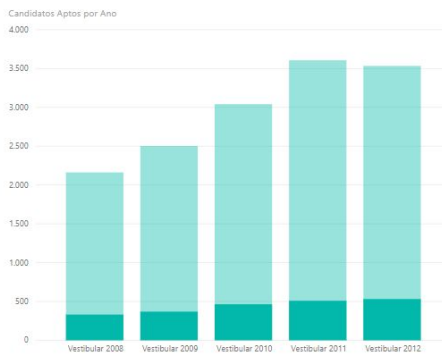
Na terceira tela avaliamos os candidatos que potencialmente estão aptos a receberem o auxílio moradia, baseando em sua faixa de renda bruta e quantidade de pessoas em sua residência.

Primeiro são listados, para cada ano, os cenários possíveis para o candidato ser apto, e quantos candidatos compõem esta categoria, totalizando 14848 candidatos potencialmente aptos.

O gráfico de barras ao lado indica o total de candidatos "aptos" por ano. Nele podemos ver uma progressão de candidatos aptos, atingindo o ápice em 2011 com 3607 candidatos "aptos".

Abaixo, a matriz de faixa de renda bruta por pessoas na residência mostra que os cenários mais comuns são a faixa "Acima de 3 até 5 sal. min.", com 5719 candidatos no total, a quantia de 4 pessoas na residência, com 5265 candidatos no total, e a intersecção destes grupos com 2204 candidatos, grupo que também vemos que teve um aumento constante durante os 5 anos.

Ano	# Renda	Faixa de Renda Bruta	Pessoas na Residência	Candidatos
Vestibular 2008	3	Acima de 3 até 5 sal. mín.	4 pessoas	331
Vestibular 2009	3	Acima de 3 até 5 sal. mín.	4 pessoas	369
Vestibular 2010	3	Acima de 3 até 5 sal. mín.	4 pessoas	454
Vestibular 2011	3	Acima de 3 até 5 sal. mín.	4 pessoas	509
Vestibular 2012	3	Acima de 3 até 5 sal. mín.	4 pessoas	531
Total				2204



Faixa de Renda Bruta	1 pessoa	2 pessoas	3 pessoas	4 pessoas	5 pessoas	Acima de 5 pessoas	Total
Acima de 1 até 3 sal. mín.	502	950	1295	1217	509	199	4676
Acima de 3 até 5 sal. mín.	745	1419	2204	979		912	5719
Acima de 5 até 7 sal. mín.					60	60	120
Acima de 7 até 10 sal. mín.			1757	669		231	2657
Acima de 10 até 15 sal. mín.			725			193	918
Até 1 salário mínimo	230	119	93	87	33	20	582
Entre 10 e 20 sal. mín.						169	169
Entre 20 e 30 sal. mín.						67	67
Total	732	1814	2871	3265	2915	1251	14848

Na quarta tela apenas são listados exatamente quais candidatos formam o grupo dos 14848 candidatos potencialmente aptos, ordenados por ano do vestibular, faixa de renda bruta e quantidade de pessoas na residência.

Ano	# Renda	Faixa de Renda Bruta	Pessoas na Residência	Candidato
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2017933
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2066602
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2057166
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2106515
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2115840
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2121930
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2136465
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2153572
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2172208
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2185113
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2234564
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2263750
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2265117
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2316137
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2328879
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2302999
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2805529
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2808030
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2809338
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2815010
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2815842
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2816229
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2819855
Vestibular 2008	1	Até 1 salário mínimo	1 pessoa	2823284
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2081011
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2129337
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2145294
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2174529
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2205130
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2800590
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2803307
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2818345
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2819821
Vestibular 2008	1	Até 1 salário mínimo	2 pessoas	2819953
Vestibular 2008	1	Até 1 salário mínimo	3 pessoas	2049362
Vestibular 2008	1	Até 1 salário mínimo	3 pessoas	2113155

Na quinta tela são mostrados dados no mesmos formatos que a terceira tela dos candidatos aptos, porém agora para os candidatos que provavelmente possuem condição financeira para aluguel ou compra de imóvel em Florianópolis.

Pelo gráfico vemos que os candidatos com capacidade financeira mantêm uma quantidade semelhante ao longo dos anos, diferente do visto com os candidatos aptos. Estes candidatos provavelmente com condições financeiras para

aluguel/compra de imóveis começam com um mínimo de 1991 candidatos em 2008, atingem o ápice em 2010 com 2340 e nos anos seguintes decresce continuamente.

A matriz também mostra que a faixa de renda bruta mais comum entre não-aptos é "Entre 10 e 20 sal. min" com 3795 candidatos, 4 pessoas na residência também é o cenário mais comum com 4654 candidatos, enquanto a intersecção de ambas variáveis compõem 1830 candidatos.



Finalmente na sexta tela são listados estes candidatos que provavelmente possuem condições financeiras para aluguel/compra de imóvel em Florianópolis, no mesmo formato da quarta tela dos candidatos potencialmente aptos.

## 6. Conclusões e Trabalhos Futuros

A partir da análise dos resultados obtidos, principalmente comparando a evolução dos números de candidatos potencialmente aptos, ao passo que os candidatos provavelmente com condições financeiras não apresentam uma evolução drástica, podemos concluir que nos últimos anos os candidatos com menor renda passaram a ingressar com mais frequência na UFSC, o que possivelmente é um indício consequente às Políticas de Ações Afirmativas implementadas pela PRAE, que visam dar uma maior oportunidade de ingresso na UFSC a grupos pouco favorecidos.

No geral, com este trabalho também foi possível vislumbrar e colocar em prática conhecimentos no contexto do Data Warehouse e Business Intelligence, principalmente no que tange o processo de ETL. Tal processo é essencial para as organizações conseguirem sistematizar o tratamento e a limpeza da grande quantidade de dados oriundos das mais diversas formas, principalmente de OLTPs (Online Transaction Processing).

## 7. Bibliografia

FERREIRA, João; MIRANDA, Miguel; ABELHA, António; MACHADO, José. O Processo ETL em Sistemas Data Warehouse.2010. Disponível em: < [https://www.researchgate.net/publication/265195317\\_O\\_Processo\\_ETL\\_em\\_Sistemas\\_Data\\_Warehouse](https://www.researchgate.net/publication/265195317_O_Processo_ETL_em_Sistemas_Data_Warehouse) >. Acessado em 12 de Junho de 2019.

RIBEIRO, Adriano. Universitários movimentam mercado imobiliário na região da Trindade. Informe Floripa. Janeiro, 2019. Disponível em: < <http://www.informefloripa.com/2019/01/11/universitarios-movimentam-mercado-imobiliario-na-regiao-da-trindade/> >. Acessado em 18 de Junho de 2019.

PRAE. Atribuições. Disponível em: < <http://prae.ufsc.br/atribuicoes/> >. Acessado em 12 de Junho de 2019.

UFSC. A UFSC. 2019. Disponível em: < <http://estrutura.ufsc.br/> >. Acessado em 12 de Junho de 2019.

DPGI.UFSC em Números.2018. Disponível em: < <http://dpgi.seplan.ufsc.br/ufsc-em-numeros/> >. Acessado em 12 de Junho de 2019.

DEVMEDIA. Porque tomar decisões utilizando Business Intelligence.2013. Disponível em: < <https://www.devmedia.com.br/porque-tomar-decisoes-utilizando-business-intelligence/28341> >. Acessado em 11 de Junho de 2019.

COPERVE. Vestibulares Anteriores. 2018. Disponível em: < <http://coperve.ufsc.br/vestibulares-anteriores/> >. Acessado em 12 de Junho de 2019.

PENTAHO. Data Integration - Kettle.2017. Disponível em: < <https://community.hitachivantara.com/docs/DOC-1009855> > . Acessado em 10 de Junho de 2019.