**UNIVERSIDADE FEDERAL DO ABC**

**PROJETO FINAL**

**Campanha Política**

**Banco de Dados de Apoio à Tomada de Decisão**

Diogo Eduardo Lima Alves RA: 21012813

Larissa de Mattos RA: 11087611

Rafael Giannini Nicoletti RA: 21004213

# PERSPECTIVA DE NEGÓCIO

A perspectiva de negócio assumida neste trabalho é a gestão de uma campanha eleitoral. Nestas, estratégias políticas são tomadas a partir da análise do perfil e das necessidades dos eleitores. Esta análise denota o percurso que a campanha eleitoral deverá seguir a fim de captar seus eleitores. Desta forma, pesquisas são o canal de comunicação que candidatos e sua equipe de campanha possuem para identificar qual o perfil e necessidades dos cidadãos.

O planejamento estratégico do candidato e de sua equipe envolve a compreensão do eleitorado e suas demandas, identificando os temas fundamentais para a campanha. Deste modo, é possível definir-se as estratégias e diretrizes essenciais para definição de uma campanha que cative os eleitores e obtenha seus votos.

# ESTUDO DOS DADOS DE ORIGEM

Escolhemos dados gerais utilizados em Índices de Desenvolvimento Humano e condições de moradia e saneamento básico para poder formular estratégias para uma hipotética campanha política, de modo que seja possível formular políticas públicas razoáveis por região e para maximizar o número de eleitores atingidos.

# DATA WAREHOUSE

Adotou-se para o Data Warehouse a estratégia de modelagem Data Vault. Sua aplicação possibilita um Data Warehouse flexível possibilitando um crescimento gradual do Data Warehouse e simplificando seu desenvolvimento e gestão. Sua metodologia ágil de desenvolvimento é capaz de absorver mudanças mais prontamente e responder satisfatoriamente ao surgimento de novas origens de dados. Características essenciais em um ambiente de negócios em constante mudança como o gerenciamento de uma campanha eleitoral. Desta forma definiu-se as seguintes tabelas, representadas através da figura 1:

* **Hubs:** *h\_domicilios, h \_microrregiao*;
* **Satélite:** *s\_domicilios;*
* **Link:** *l\_esta.*

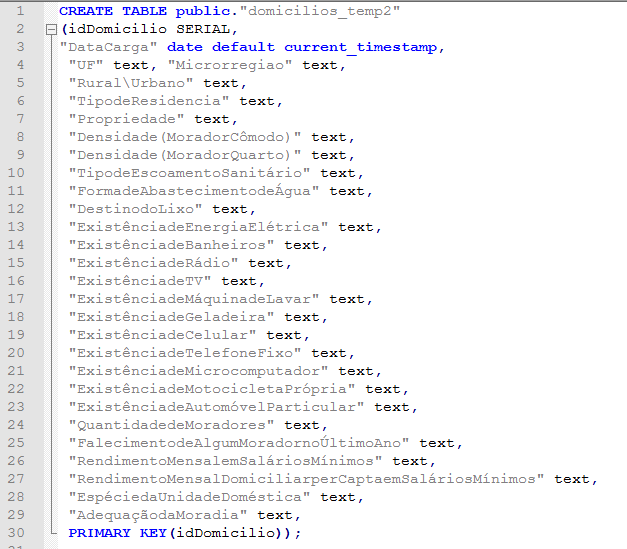
**Figura 1**. Modelagem do Data Warehouse: relacionamentos



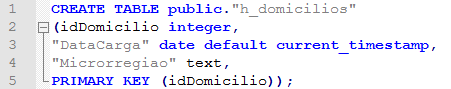
Tendo em vista que a base de dados de domicílios do IBGE não dispõe de chave primária de identificação de suas tuplas, foi necessário aplicar-se a estratégia de *staging area*. Desta forma, utilizou-se uma tabela temporária, *domicilios\_temp2*, de modo que após o processo de transformação a mesma dividiu-se sendo carregada nas tabelas do DW *h\_domicilios* e *s\_domicilios*, garantindo que as chaves referenciam os mesmos objetos em ambas as tabelas. Também foi adicionado o “2010” ao final de cada chave, pois como trata-se de um censo, esta é a chave temporal, caso sejam carregados novos dados futuramente.

As figuras 2 à 6 apresentam o código *PostgreSQL* implementado para criação das tabelas através da ferramenta PgAdmin 4.

**Figura 2**. Modelagem do Data Warehouse: criação da tabela temporária



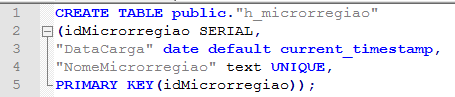
**Figura 3**. Modelagem do Data Warehouse: criação da hub h\_domicílios



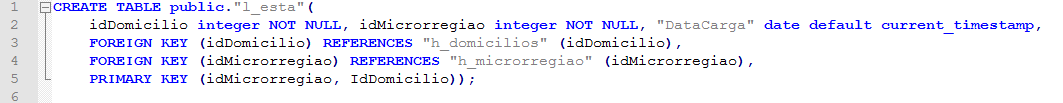
**Figura 4.**. Modelagem do Data Warehouse: criação do satélite s\_ domicílios



**Figura 5**. Modelagem do Data Warehouse: criação da hub h\_microrregião



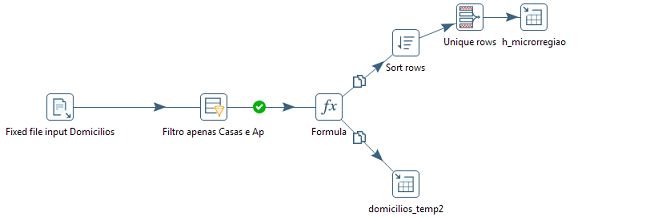
**Figura 6**. Modelagem do Data Warehouse: criação do link l\_esta



# PROCESSO DE ETL

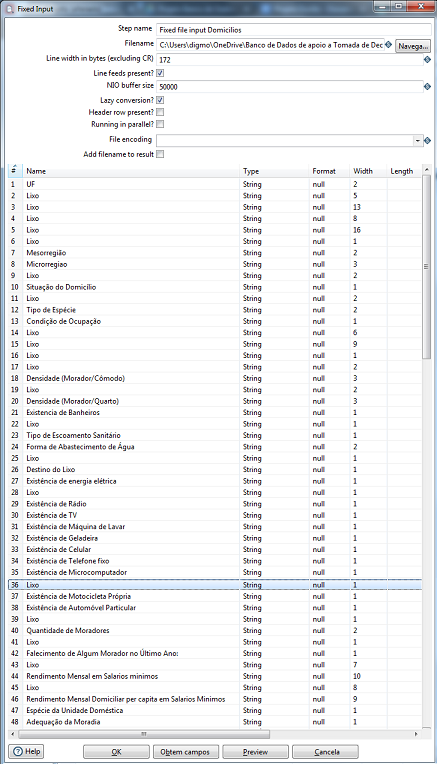
A ferramenta de ETL open source Pentaho possibilitou o processo de Extração, Transformação e Carga do projeto, no qual foi efetuado o tratamento e limpeza dos dados extraídos da base de dados de domicílios do IBGE, para a inserção dos dados transformados em um Data Warehouse modelado para a perspectiva de negócio definida, uma campanha política.

**Figura 7.** Processo de ETL



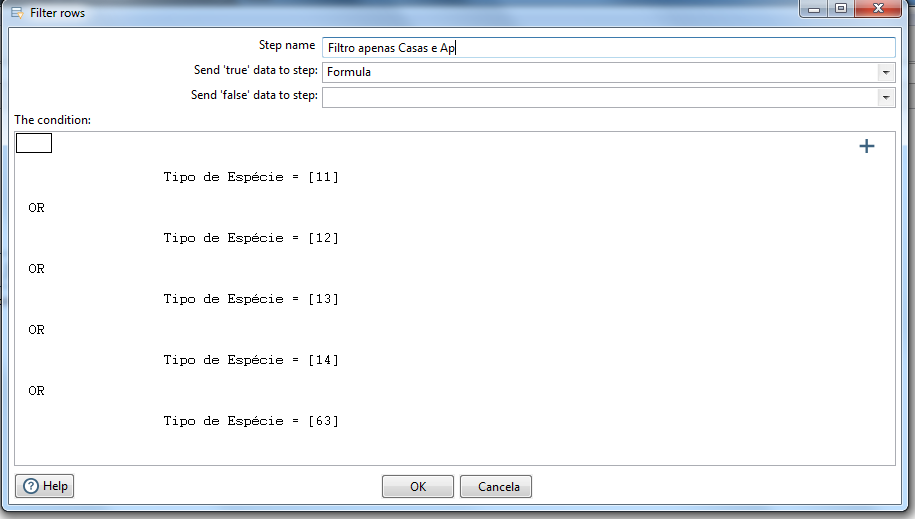
A figura 7 apresenta o processo de ETL aplicado na base de dados de origem. A tabela *domicilios\_temp2* é uma tabela temporária parte do processo de *staging area* e será posteriormente dividida em *h\_domicilios* e *s\_domicilios*. Após a extração dos dados da base de domicílios, efetuou-se a seleção dos campos relevantes à proposta, e os campos desconsiderados foram representados através da palavra chave “lixo”, conforme figura 8.

**Figura 8.** Campos de Input



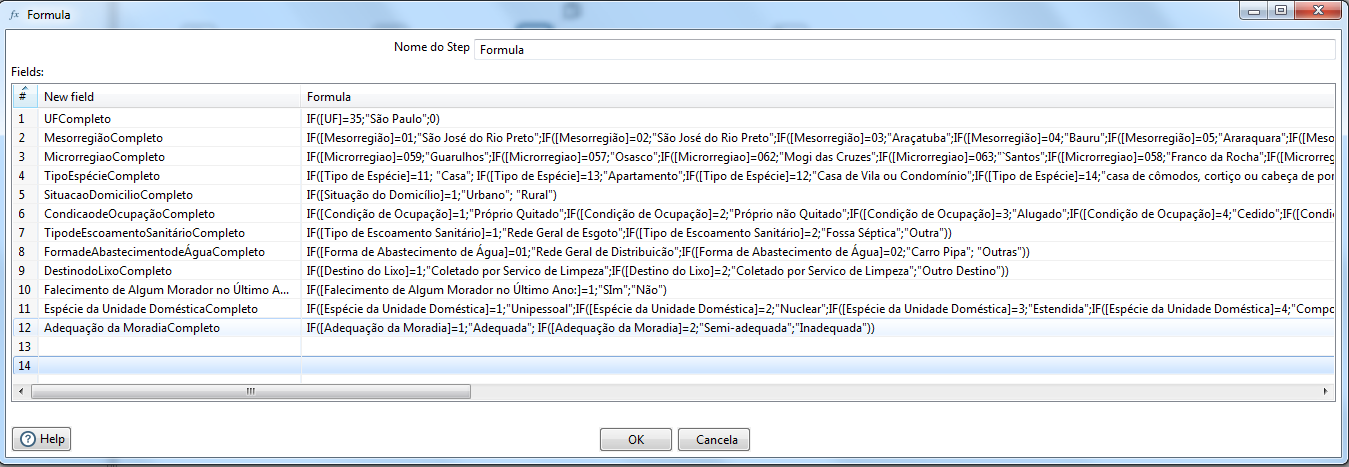
Em seguida, por meio de um filtro de linhas, restringiu-se os domicílios a casas e apartamentos, conforme figura 9.

**Figura 9.** Filtro das linhas



Com os dados filtrados, os campos nos quais códigos denotavam informações relevantes à visualização final foram traduzidos através da aplicação de fórmulas, conforme Figura 10.

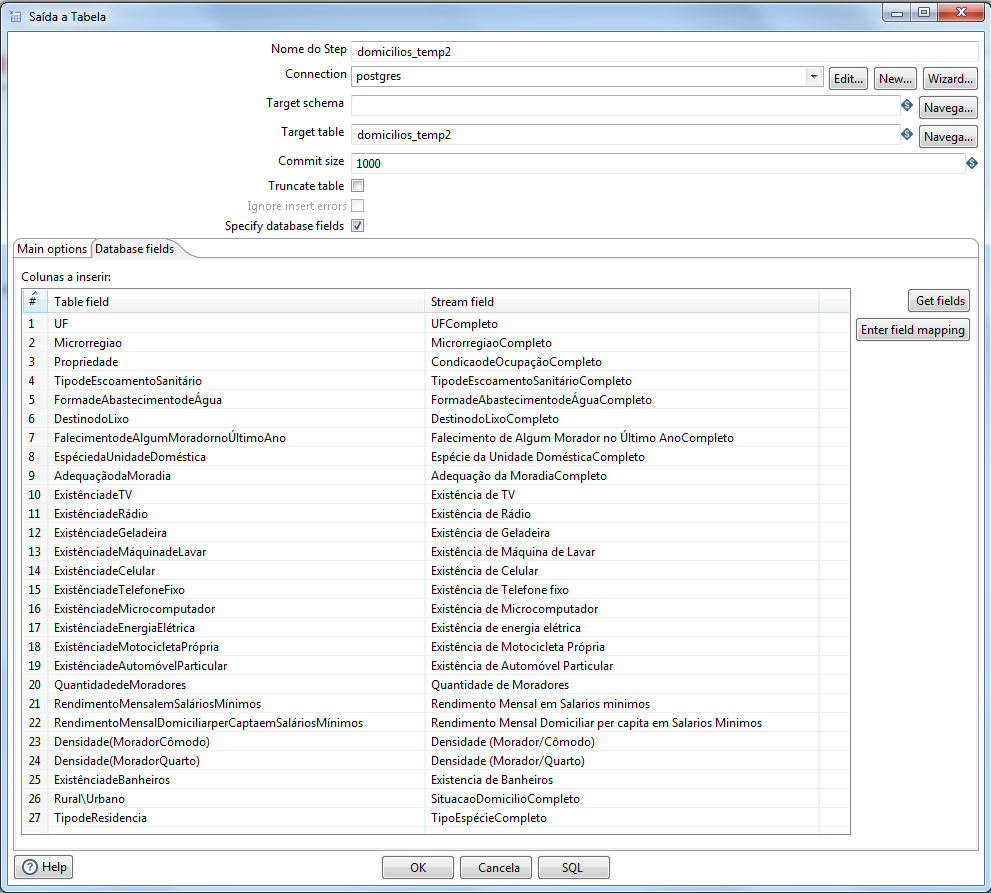
**Figura 10.** Fórmulas



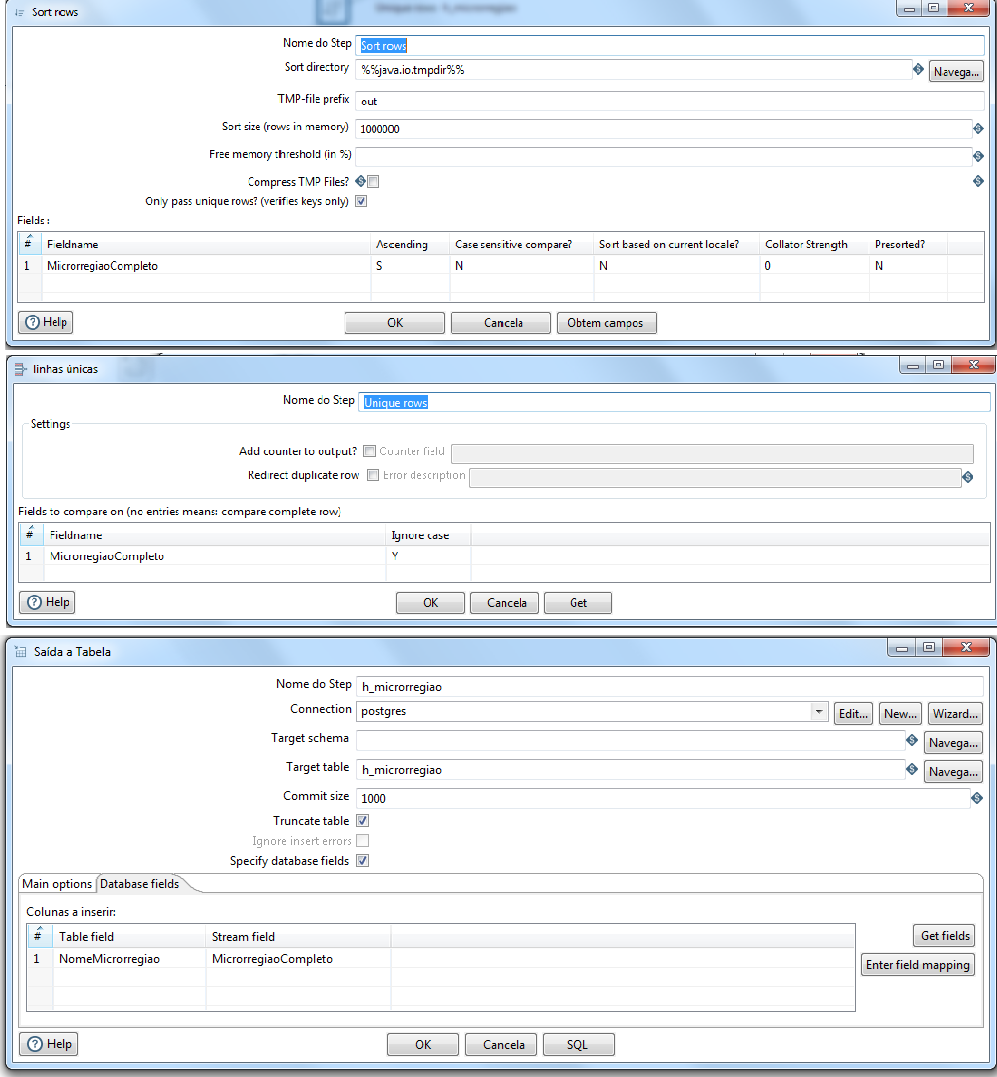
Em seguida, duas tabelas são populadas: *h\_microrregião e domicilios\_temp2*. A figura 11 apresenta os campos inseridos na tabela *domicilios\_temp2.*

Para popular a tabela *h\_microrregião,* primeiramente efetuou-se a ordenação ascendente do campo *MicrorregiaoCompleto,* ea seguir removeu-se as linhas duplicadas de forma a serem carregadas na tabela apenas o campo microrregião, conforme figura 12.

**Figura 11.** Saída a tabela *domicilios\_temp2*

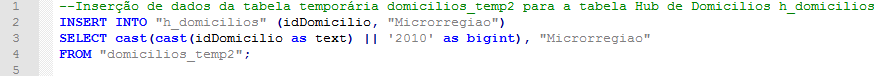


**Figura 12.** Processo de carregamento da tabela *h\_microrregião*



Após o processo de carga na tabela temporária *domicilios\_temp2* e na hub *h\_microrregião*, *inserts* foram executados por meio da ferramenta PgAdmin 4 a fim de carregar as tabelas h\_domicilios e s\_domicilios, conforme figuras 13 à 15.

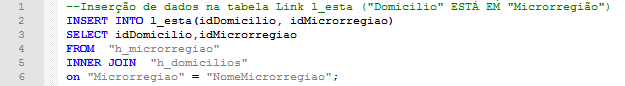
**Figura 13.** Processo de carregamento da tabela *h\_domicilios*



**Figura 14.** Processo de carregamento da tabela *s\_domicilios*



**Figura 15.** Processo de carregamento da tabela *l\_esta*

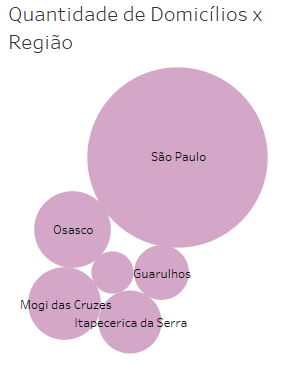


# VISUALIZAÇÕES

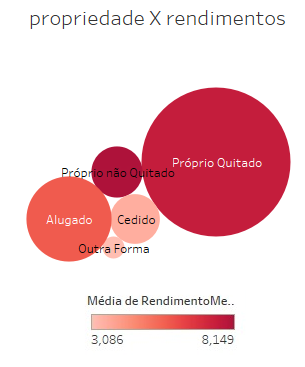
Com o objetivo de elaborar uma campanha eleitoral para o governo do estado de São Paulo, analisamos as condições de moradias e saneamento básico de algumas regiões. A princípio analisamos as informações como um todo, para obtermos uma visão mais ampla destas informações em relação ao estado. Como ponto inicial, observamos a quantidade de domicílios por região e a relação entre os rendimentos familiares e o tipo de moradia.

Observando a quantidade de domicílios por região (figura 16), observamos que grande parte dos entrevistados pelo CENSO pertence a microrregião de São Paulo. Podemos observar, também, que a grande parte dos tipos de moradias (figura 17) são próprios quitados e alugados. Os imóveis próprios não quitados aparecem em menor quantidade, mas seus moradores estão entre aqueles que possuem as maiores médias salariais. As menores rendas estão atreladas as moradias que foram cedidos ou outros tipos de moradia. Mas, ao analisarmos as informações sobre a adequação das moradias (figura 18), percebemos que estas moradias, mesmo quando próprias quitadas, não são totalmente adequadas. Menos de 70% das moradias são consideradas adequadas para a habitação.

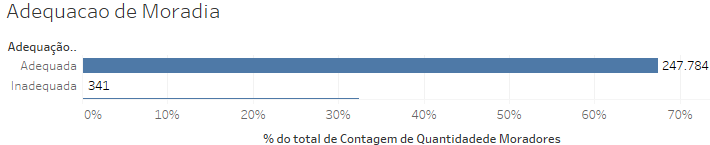
**Figura 16**. Quantidade de domicílios por região



**Figura 17.** Domicílios por rendimento mensal em salários mínimos



**Figura 18.**  Adequação de moradia

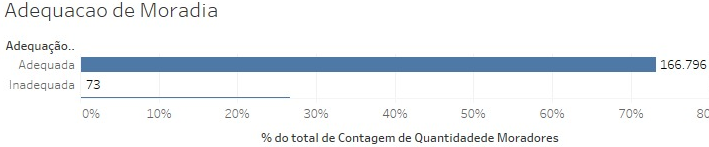


Quando observamos cada região isoladamente, fica fácil a identificação dos problemas de moradia existentes relativos a região observada. A figura 19 mostra a microrregião de São Paulo, onde grande parte das moradias são próprias quitadas ou alugada. As pessoas que possuem maiores rendimentos ainda não possuem moradia própria (estão em processo de quitação). As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. Também podemos observar, através da figura 20, que pouco mais de 70% das moradias são consideradas adequadas para habitação.

**Figura 19.** Domicílios por rendimento mensal em salários mínimos em São Paulo

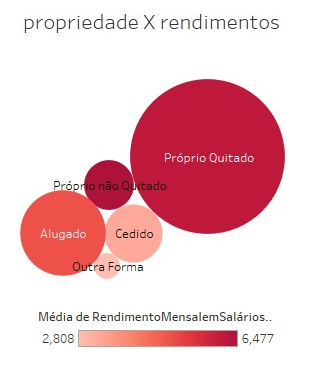


**Figura 20.** Adequação de moradia

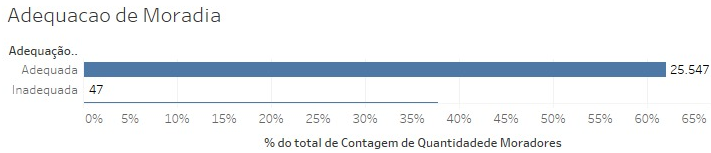


A figura 21 mostra a microrregião de Osasco, onde grande parte das moradias são próprias quitadas ou alugada. As pessoas que possuem maiores rendimentos ainda não possuem moradia própria (estão em processo de quitação). As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. Esta microrregião se assemelha muito a de São Paulo. Também podemos observar, através da figura 22, que pouco mais de 60% das moradias são consideradas adequadas para habitação.

**Figura 21.** Domicílios por rendimento mensal em salários mínimos em Osasco



**Figura 22.** Adequação de moradia

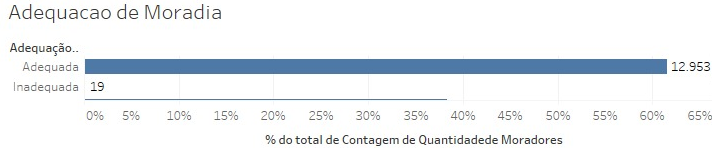


A figura 23 mostra a microrregião de Guarulhos, onde grande parte das moradias são próprias quitadas ou alugada. As pessoas que possuem maiores rendimentos ainda não possuem moradia própria (estão em processo de quitação). As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. Esta microrregião se assemelha muito a de São Paulo. Também podemos observar, através da figura 24, que pouco mais de 60% das moradias são consideradas adequadas para habitação.

**Figura 23.** Domicílios por rendimento mensal em salários mínimos em Guarulhos



**Figura 24.** Adequação de moradia

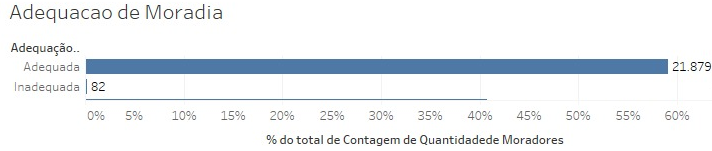


A figura 25 mostra a microrregião de Mogi das Cruzes, onde grande parte das moradias são próprias quitadas. Nesta microrregião, as pessoas que possuem maiores rendimentos estão entre as pessoas que possuem moradia própria quitada e as que possuem moradia própria não quitada. As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. O número de moradores em moradias próprias não quitadas se aproxima muito dos que moram em aluguel. Também podemos observar, através da figura 26, que menos de 60% das moradias são consideradas adequadas para habitação.

**Figura 25.** Domicílios por rendimento mensal em salários mínimos em Mogi das Cruzes

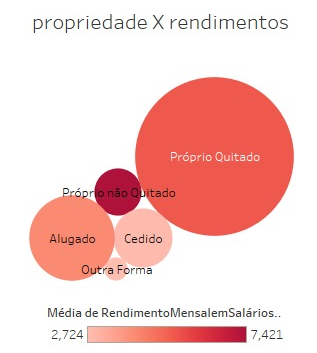


**Figura 26.** Adequação de moradia

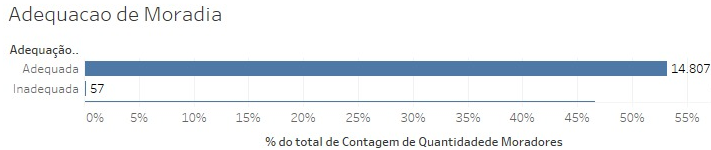


A figura 27 mostra a microrregião de Itapecerica da Serra, onde grande parte das moradias são próprias quitadas ou alugadas. Nesta microrregião, as pessoas que possuem maiores rendimentos estão entre as pessoas que possuem moradia própria quitada não quitada. As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. Também podemos observar, através da figura 28, que apenas metade (aproximadamente) das moradias são consideradas adequadas para habitação.

**Figura 27.** Domicílios por rendimento mensal em salários mínimos em Itapecerica da Serra

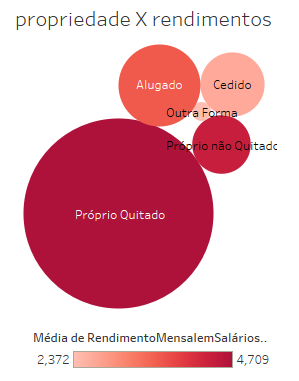


**Figura 28.**  Adequação de moradia

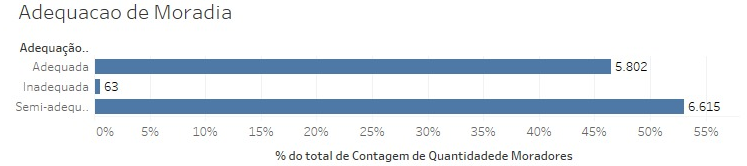


A figura 29 mostra a microrregião de Franco da Rocha, onde grande parte das moradias são próprias quitadas. Nesta microrregião, as pessoas que possuem maiores rendimentos estão entre as pessoas que possuem moradia própria quitada e as que possuem moradia própria não quitada. As que possuem os menores rendimentos médios possuem moradias cedidas ou outras formas de moradias. O número de moradores em moradias próprias não quitadas se aproxima muito dos que moram em aluguel. Também podemos observar, através da figura 30, que apenas metade (aproximadamente) das moradias são consideradas adequadas para habitação.

**Figura 29.** Domicílios por rendimento mensal em salários mínimos em Franco da Rocha



**Figura 30.** Adequação de moradia

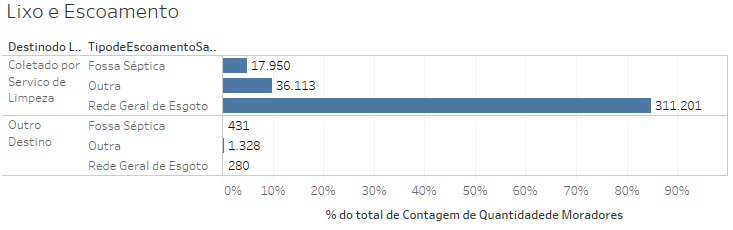


Observando estas informações, podemos elaborar algumas metas de campanha que podem ser utilizadas nas maiorias das microrregiões. As principais metas que podem ser obtidas através destas informações são:

* A criação de moradias populares em todas as regiões (ou a conclusão das obra inacabadas, caso existam obras em andamento), com foco nas pessoas que moram em situação de aluguel e/ou para as pessoas que habitam em moradias impróprias (demonstrados através das outras formas nas figuras acima);
* Planos de crédito imobiliário mais baratos e/ou revisão das taxas de financiamento, para facilitar a compra de moradias mais adequadas e/ou facilitar a quitação destes imóveis ou dos que já estão em processo de quitação.

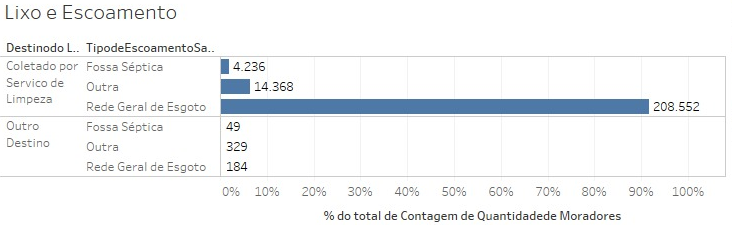
Em relação a saneamento básico, observamos que a grande maioria dos entrevistados já está conectada à rede geral de esgotos da sua região, mas existe uma grande parcela da população que ainda despeja seus resíduos de forma inadequada (cerca de 15%).

**Figura 31.** Lixo e escoamento referente a coleta de água e esgoto



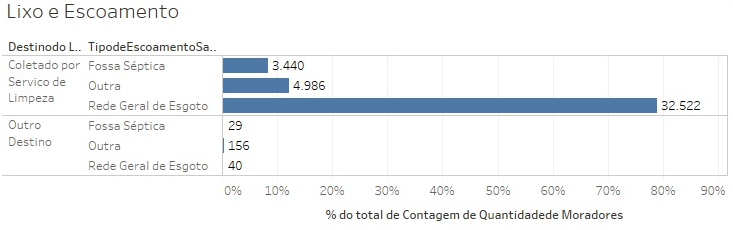
Observando a microrregião de São Paulo (figura 32), vemos que esta região é a que possui melhores índices de coleta de esgoto, superior a 90%, mas boa parte do que não é coletado é simplesmente descartado de maneira totalmente irregular (em córregos e rios, por exemplo).

**Figura 32.** Lixo e escoamento referente a coleta de água e esgoto de São Paulo



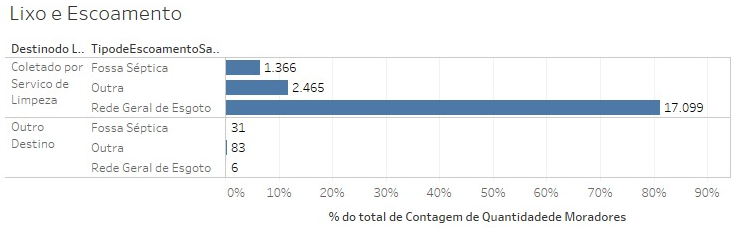
Observando a microrregião de Osasco (figura 33), vemos que, nesta região, menos de 80% das moradias estão conectadas à rede geral de esgoto e boa parte do que não é coletado é simplesmente descartado de maneira totalmente irregular.

**Figura 33.** Lixo e escoamento referente a coleta de água e esgoto de Osasco



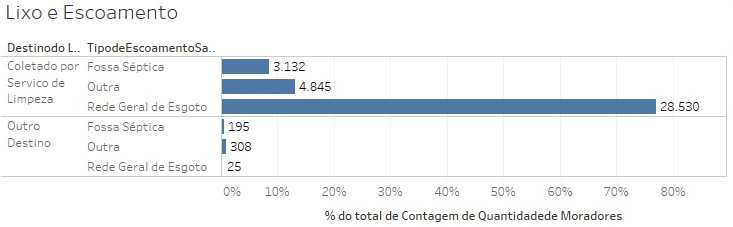
Observando a microrregião de Guarulhos (figura 34), vemos que, nesta região, menos de 80% das moradias estão conectadas à rede geral de esgoto e boa parte do que não é coletado é simplesmente descartado de maneira totalmente irregular.

**Figura 34.** Lixo e escoamento referente a coleta de água e esgoto de Guarulhos



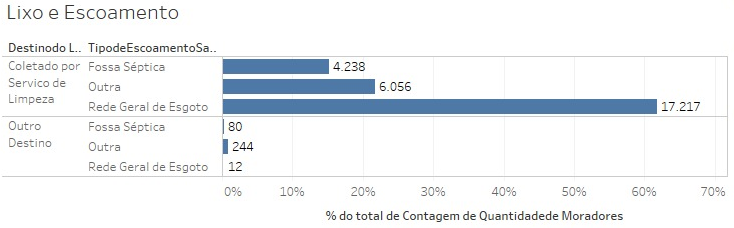
Observando a microrregião de Mogi das Cruzes (figura 35), vemos que, nesta região, menos de 80% das moradias estão conectadas à rede geral de esgoto e boa parte do que não é coletado é simplesmente descartado de maneira totalmente irregular.

**Figura 35.** Lixo e escoamento referente a coleta de água e esgoto de Mogi das Cruzes



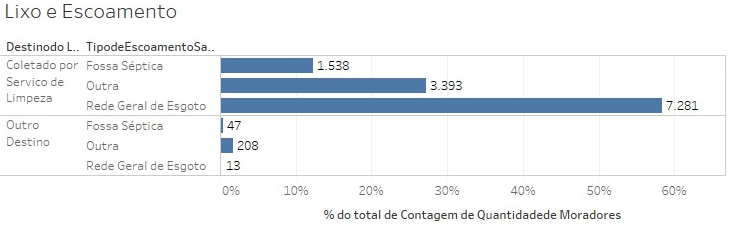
Observando a microrregião de Itapecerica da Serra (figura 36), vemos que, nesta região, apenas cerca de 62% das moradias estão conectadas à rede geral de esgoto e mais de 20% do que não é coletado é simplesmente descartado de maneira totalmente irregular.

**Figura 36.** Lixo e escoamento referente a coleta de água e esgoto de Itapecerica da Serra



Observando a microrregião de Franco da Rocha (figura 37), vemos que, nesta região, menos de 60% das moradias estão conectadas à rede geral de esgoto e mais de 25% do que não é coletado é simplesmente descartado de maneira totalmente irregular.

**Figura 37.** Lixo e escoamento referente a coleta de água e esgoto de Franco da Rocha

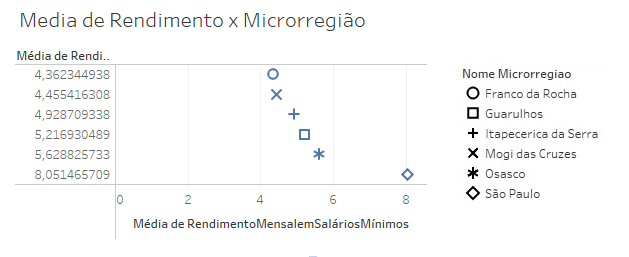


Com essas informações, percebemos que grande parte do problema de saneamento está localizado nas regiões mais carentes, onde os moradores possuem menos poder aquisitivo e vivem em moradias mais precárias (mesmo sabendo que a grande maioria vive em moradias próprias quitadas, isto não informa qual é o tipo de moradia na qual elas estão habitando). Neste caso, podemos elaborar as seguintes metas para o plano de governo:

* Para a microrregião de São Paulo, podemos elaborar uma meta de coletar 100% do esgoto e a criação de novas estações de tratamento, com o foco de tratar todo o esgoto coletado;
* Para as microrregiões de Osasco, Guarulhos e Mogi das Cruzes, podemos elaborar uma meta de coletar 100% do esgoto e a criação de novas estações da tratamento, com o foco de tratar pelo menos metade do esgoto coletado;
* Para as microrregiões de Itapecerica da Serra e Franco da Rocha, podemos elaborar uma meta de coletar todo o esgoto. Como há muito a ser feito nestas duas regiões, não é aconselhável elaborar uma meta mais ousada, mas não há a limitação de parar neste ponto quando todo o esgoto for coletado. Neste momento poderemos estudar a construção de estações de tratamento de esgoto.

Observando a figura 38, podemos perceber que a microrregião de São Paulo é a que possui maior média salarial entre as microrregiões analisadas. Com esta informação podemos elaborar metas para a criação de mais escolas técnicas e universidades públicas nas outras microrregiões e incentivos para que grandes empresas venham a se instalar nestas regiões, oferecendo, assim, mais empregos com melhores condições de trabalho e salários melhores. Para suprir a nova demanda de mão de obra especializada que irá surgir, podemos elaborar metas para a construção de novas escolas técnicas e universidades públicas (tanto estaduais quanto federais - estas necessitando de apoio do governo federal, não sendo recomendada a divulgação antes de acordos já estabelecidos para a construção das mesmas).

**Figura 38.** Média de rendimento, em salários mínimos, por microrregião



# CONSIDERAÇÕES FINAIS

O trabalho executado teve como objetivo o desenvolvimento de um ambiente de Data Warehouse voltado ao gerenciamento de uma campanha eleitoral, de modo que as visualizações expostas neste relatório visam o fornecimento de subsídios para a tomada de decisão dentro do contexto proposto.

A maior dificuldade encontrada ocorreu na etapa de transformação e tratamento dos dados durante o processo de ETL, no qual fez-se necessário a aplicação de diversas conversões de tipos de dados para possibilitar o carregamento dos mesmos no Data Warehouse modelado. Logo, percebe-se a complexidade envolvida no processo de extração dos dados das diversas origens, e a dificuldade de relacioná-los de modo rápido e seguro, a fim de transformá-los em informações substanciais na tomada de decisão da corporação.

Desta forma, acredita-se que o objetivo deste trabalho foi amplamente atingido, no qual os integrantes da equipe desenvolveram com sucesso as atividades de planejamento, desenvolvimento e gerenciamento de um Data WareHouse.