

Vítor de Godeiro Marques

1. Introdução

O objetivo dessa atividade é desenvolver uma análise dos bairros da cidade de São Paulo para:

1. Estimar o faturamento que uma loja teria em cada um dos bairros;
2. Classificar o potencial de cada bairro como Alto, Médio ou Baixo.

2. Solução

A solução dessa atividade inicialmente consiste em separar os dados da cidade do Rio de Janeiro com os dados da cidade de São Paulo. Os dados do Rio de Janeiro irão servir como base (dados de treinamento) para fazermos a análise dos dados da cidade de São Paulo (estimar o faturamento que uma loja teria em cada um dos bairros e estimar o potencial de cada bairro como alto, médio ou baixo).

Inicialmente fazemos uma análise dos dados da cidade do Rio de Janeiro, nela observamos que para o campo **rendaMedia** está faltando informação para 6 dados da tabela. Para solucionar poderíamos excluir esses dados mas devido a baixa quantidade, (ao total são 160 dados da cidade do Rio de Janeiro), optamos por desenvolver um modelo de *Machine Learning* para estimar esses valores faltantes utilizando como base os 154 dados restantes da cidade do Rio de Janeiro.

A próxima etapa agora é selecionar as melhores *features* para resolver o problema, dentre elas as *features* escolhidas foram: **domiciliosA1, domiciliosA2, domiciliosB1, domiciliosC2, domiciliosD, domiciliosE, rendaMedia e faturamento**. Com as *features* selecionadas desenvolvemos um modelo de *Machine Learning* com o objetivo de estimar valores a partir de uma dada entrada (essas entradas são formadas pelas *features* pré selecionadas) e a partir das informações das cidades de São Paulo encontramos os valores para o faturamento de cada bairro.

Agora que possuímos a informação do faturamento para cada bairro da cidade de São Paulo, iremos partir para o problema de classificar o potencial de cada bairro da mesma como alto, médio ou baixo. Inicialmente iremos utilizar o dado da cidade de Rio de Janeiro como base para realizarmos a análise da cidade de São Paulo.

Inicialmente selecionamos as melhores *features* para resolver o problema, sendo elas: **população, popDe20a24, popDe25a34, popDe35a49, popDe50a59, popMaisDe60, domiciliosA1, domiciliosA2, domiciliosB1, domiciliosB2, domiciliosC1, rendaMedia, faturamento e potencial**. A próxima etapa é verificar que o conjunto de dados do Rio de Janeiro está desbalanceado, ou seja, possui mais dados de um tipo de potencial do que de outro, por tanto iremos solucionar o problema com a criação de dados fictícios para deixar as três classes com a mesma quantidade de dados.

A próxima etapa é desenvolver o modelo de *Machine Learning* com o objetivo de realizar a classificação do potencial de cada dado de entrada, ou seja, as *features* escolhidas para resolver o problema. A partir das informações das cidades de São Paulo encontramos os valores para o potencial de cada bairro.

3. Análise das informações

Com as informações presentes nos dados utilizados para resolver esse problema, podemos observar a distribuição da população da cidade de São Paulo pela sua respectiva faixa etária na Figura 1. Podemos observar que a faixa de idade entre 35 a 49 anos é onde possui a maior parcela da população e as faixas de 25 a 34 anos, de 50 a 59 anos, a cima de 60 anos e até 9 anos possuem quantidades de pessoas semelhantes.

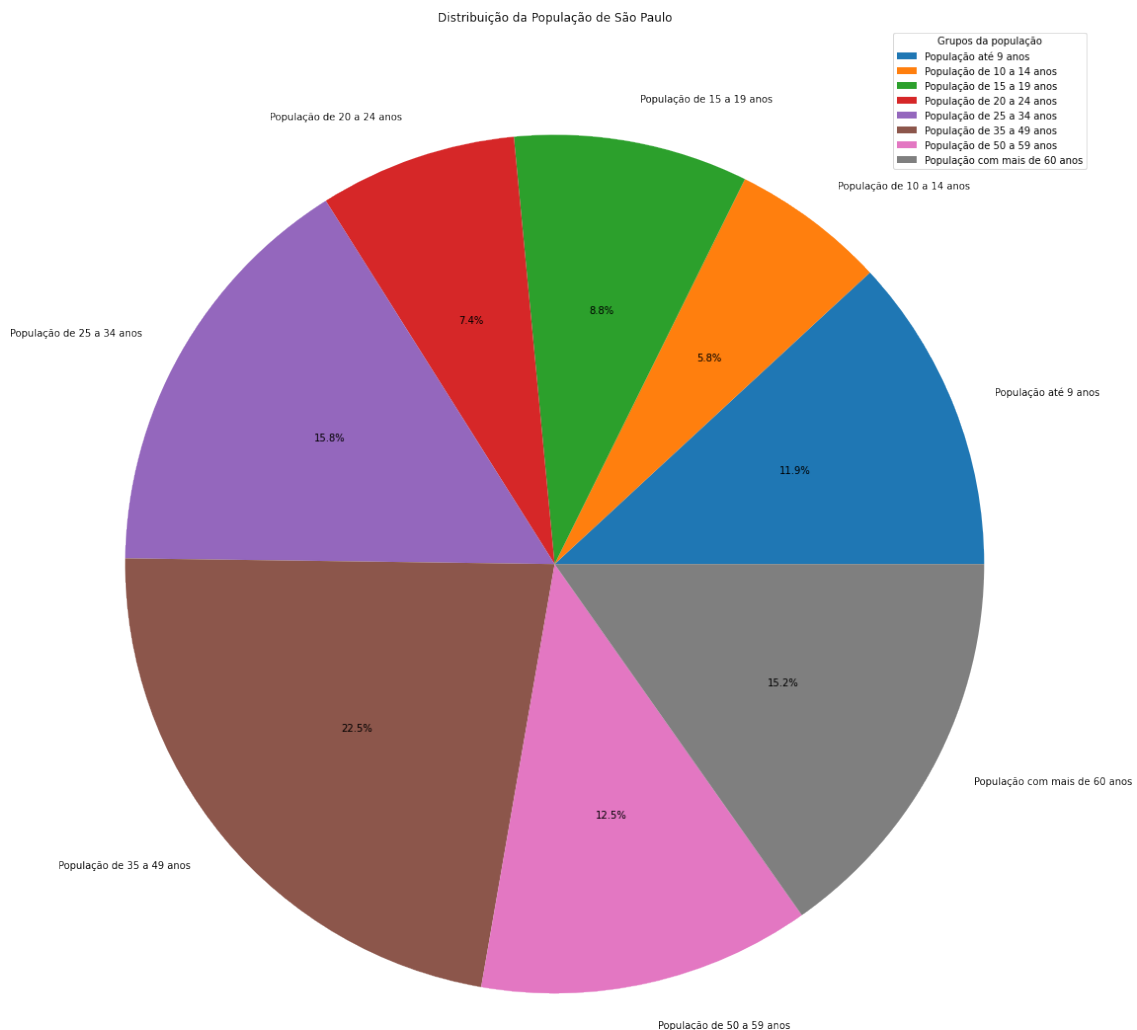


Figura 1. Distribuição da população de São Paulo por faixa etária.

Na Figura 2 observamos que a maioria dos domicílios da cidade de São Paulo são da **Classe C**, e as classes de interesse do problema **Classes A e B** juntas representam 37.9% da população da cidade de São Paulo.

Na Figura 3 observamos que a maioria dos bairros da cidade de São Paulo são de **alto potencial**, os bairros de potencial alto e médio representam cerca de 82% dos bairros da cidade, indicando ser uma excelente cidade para a abertura das filiais.

Na Figura 4 observamos o mapa da cidade de São Paulo com a divisão por bairros dos potenciais de cada um deles. Na figura o azul escuro representa a que possui um

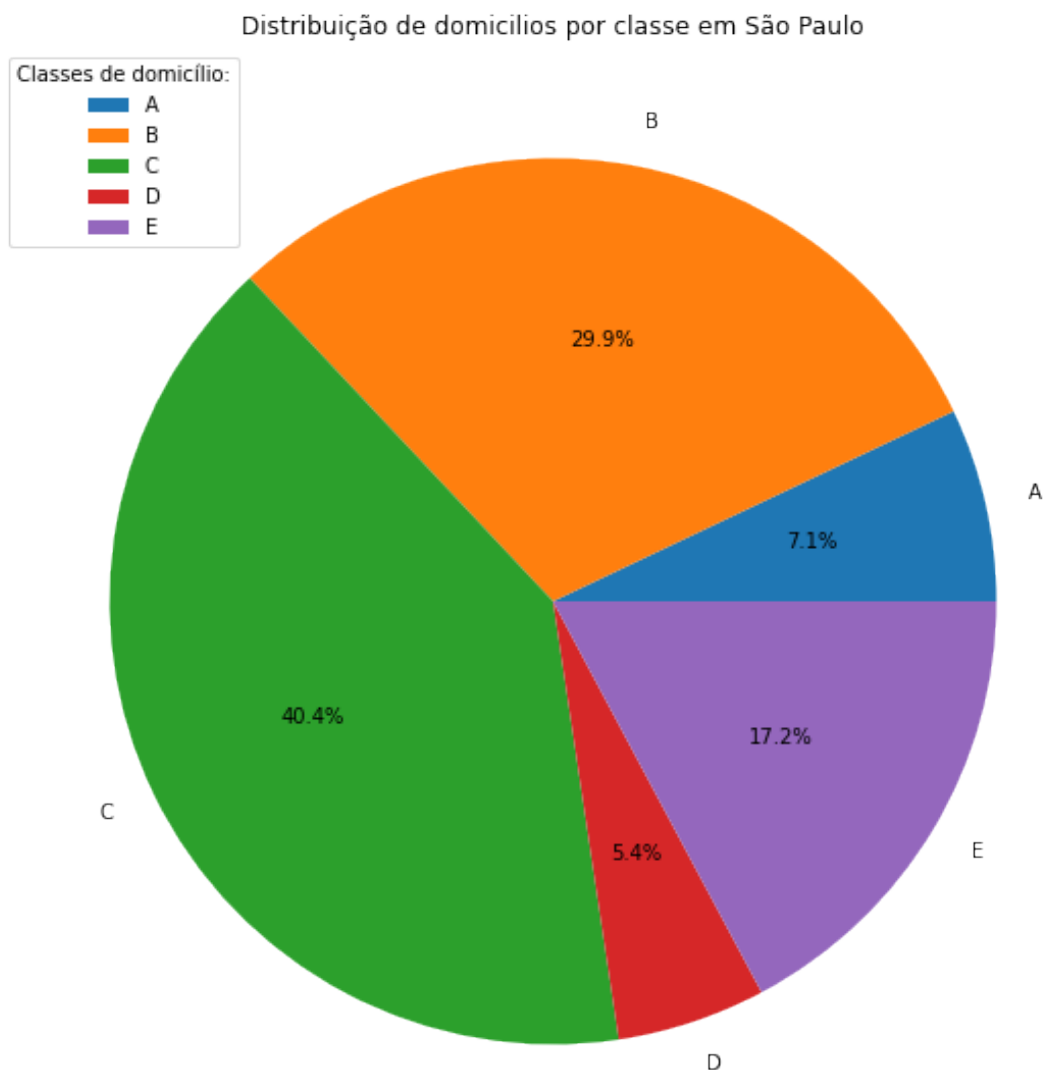


Figura 2. Distribuição dos bairros do município de São Paulo por classe de renda.

alto potencial, o marrom de potencial médio e o azul claro de baixo potencial. Como já sabíamos que a maioria dos bairros são de alto potenciais, com a ajuda do mapa podemos observar que existe diversos bairros que são rodeados por outros bairros com alto potencial, indicando assim serem bairros mais *atrativos* para a abertura de filiais. Como por exemplo os bairros: **Vila Formosa, Jardim Paulista, Moema**.

Na Figura 5 observamos o mapa da cidade de São Paulo com a segmentação por bairros de acordo com o faturamento de cada um deles. As cores podem ser interpretadas da forma que quanto mais puxado para o roxo menor será o faturamento daquele bairro e quanto mais puxado para o amarelo maior o faturamento daquele bairro. Portanto com a análise desse mapa e o mapa da Figura 3, observamos que o bairro da **Moema** está rodeado de bairros com alto potencial e de alto faturamento para esse tipo de filial, indicando ser um bairro altamente recomendado.

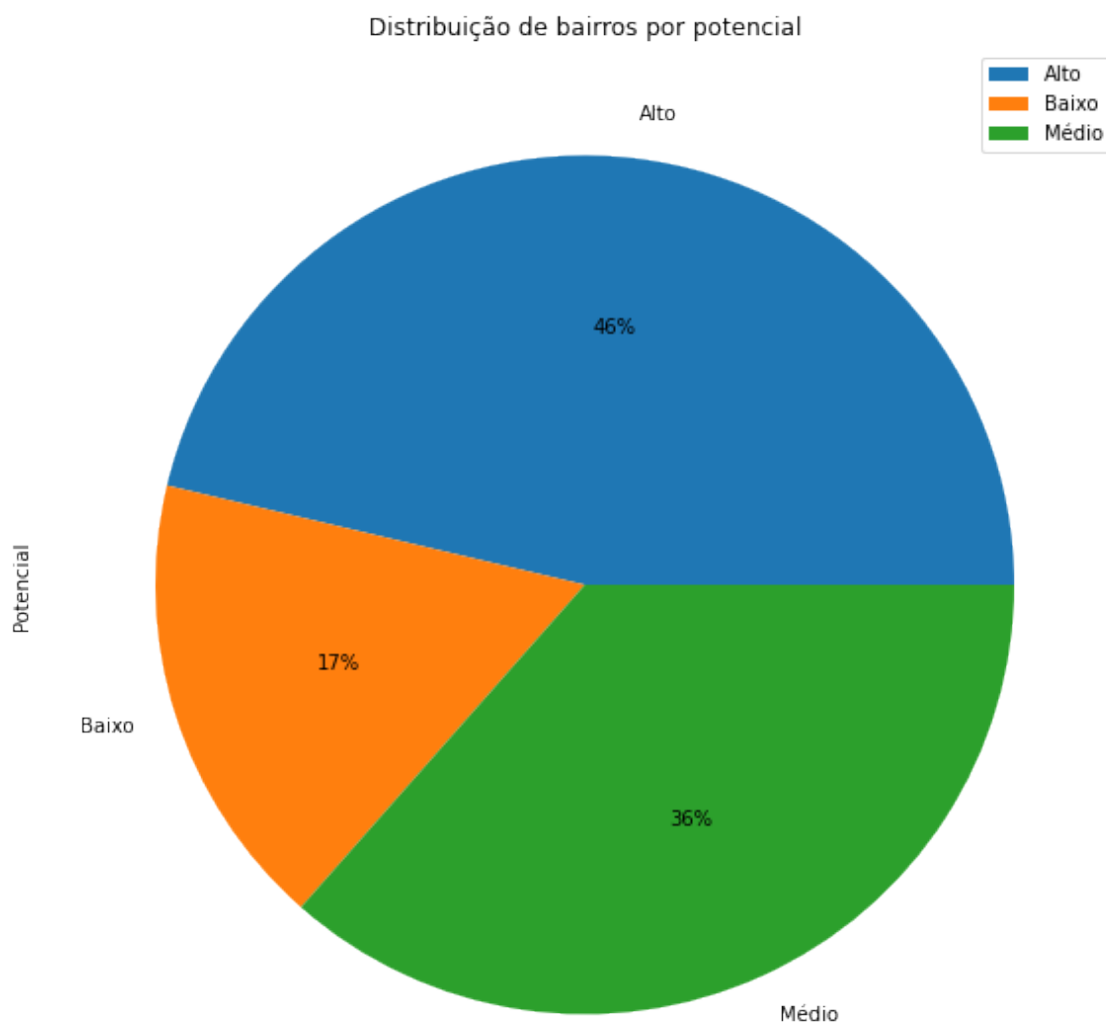


Figura 3. Distribuição dos bairros do município de São Paulo por potencial.

Adicionalmente podemos destacar os 15 bairros com maior faturamento estimado na cidade de São Paulo, sendo eles em ordem decrescente: Moema, Perdizes, Triunfo, Chácara Itaim, Vila Andrade, Vila Mariana, Saúde, Paraíso, Tatuapé, Pamplona, Brooklin, Marechal Deodoro, Jardim Marajoara, Alfredo, Pujol e Viera De Moraes. Todos eles foram estimados com alto potencial, portanto são excelentes bairros para a abertura de uma filial, em especial o bairro de **Moema** por ser o de maior faturamento estimado e perto de diversos outros como o Brooklin, Viera De Moraes e Chácara Itaim, estando esses no top 15 bairros com maior potencial de faturamento.

Na Figura 6 observamos o mapa da cidade de São Paulo com a segmentação por bairros de acordo com a quantidade de domicílios de interesse (A e B), na Figura 7 observamos a segmentação de acordo com a população entre 25 e 49 anos, na Figura 8 observamos a segmentação de acordo com a renda média. As cores podem ser interpretadas da forma que quanto mais puxado para o roxo menor será o valor naquele bairro e

quanto mais puxado para o amarelo maior o valor naquele bairro. Essas informações seriam interessantes para auxiliar na estratégia, investimento e tomada de decisões para o marketing dessas novas filiais, tornando possível especificar o tipo de marketing que será realizado em cada bairro para atender um público alvo em específico.

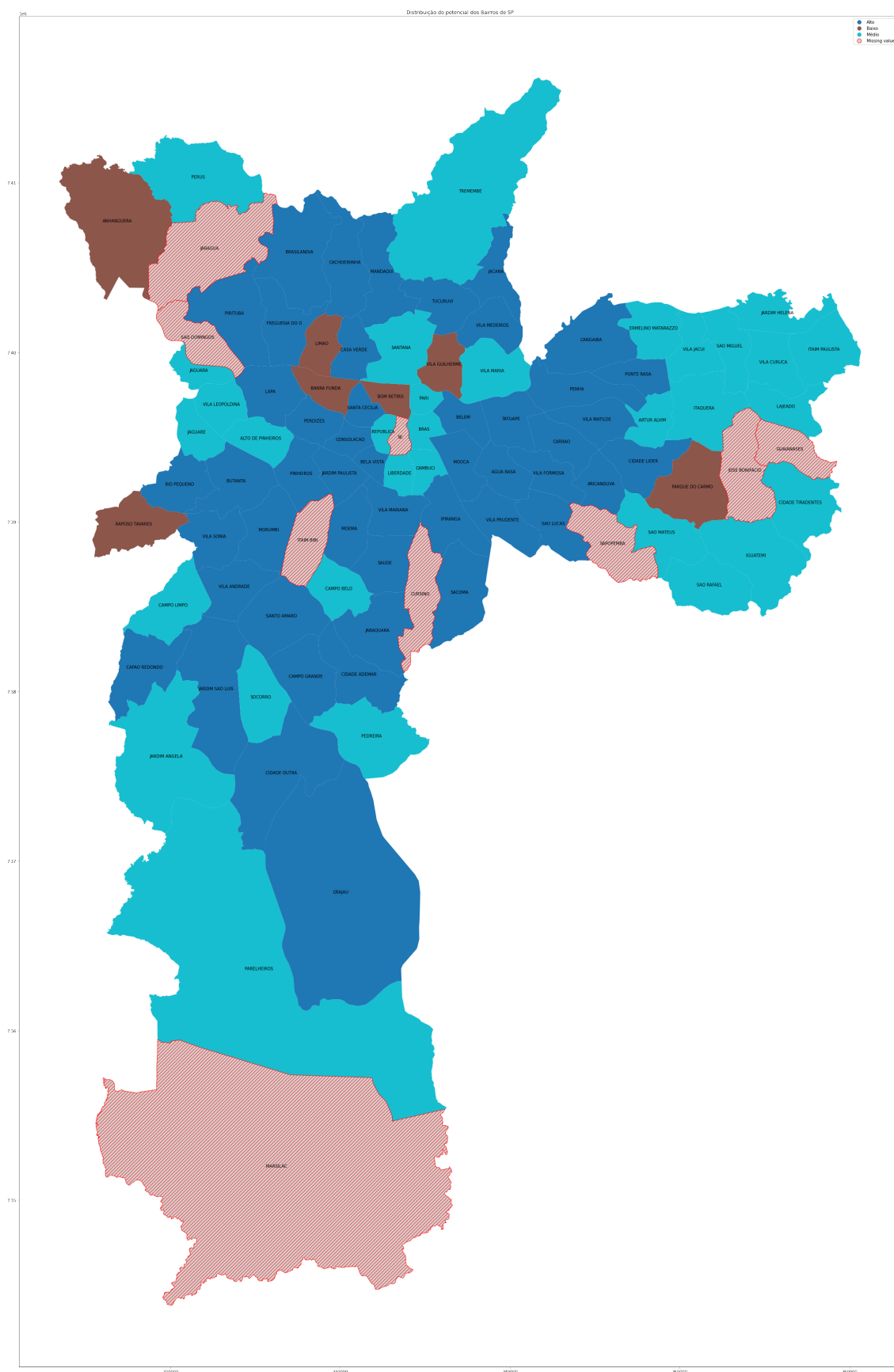


Figura 4. Mapa com a segmentação dos bairros do município de São Paulo por potencial.

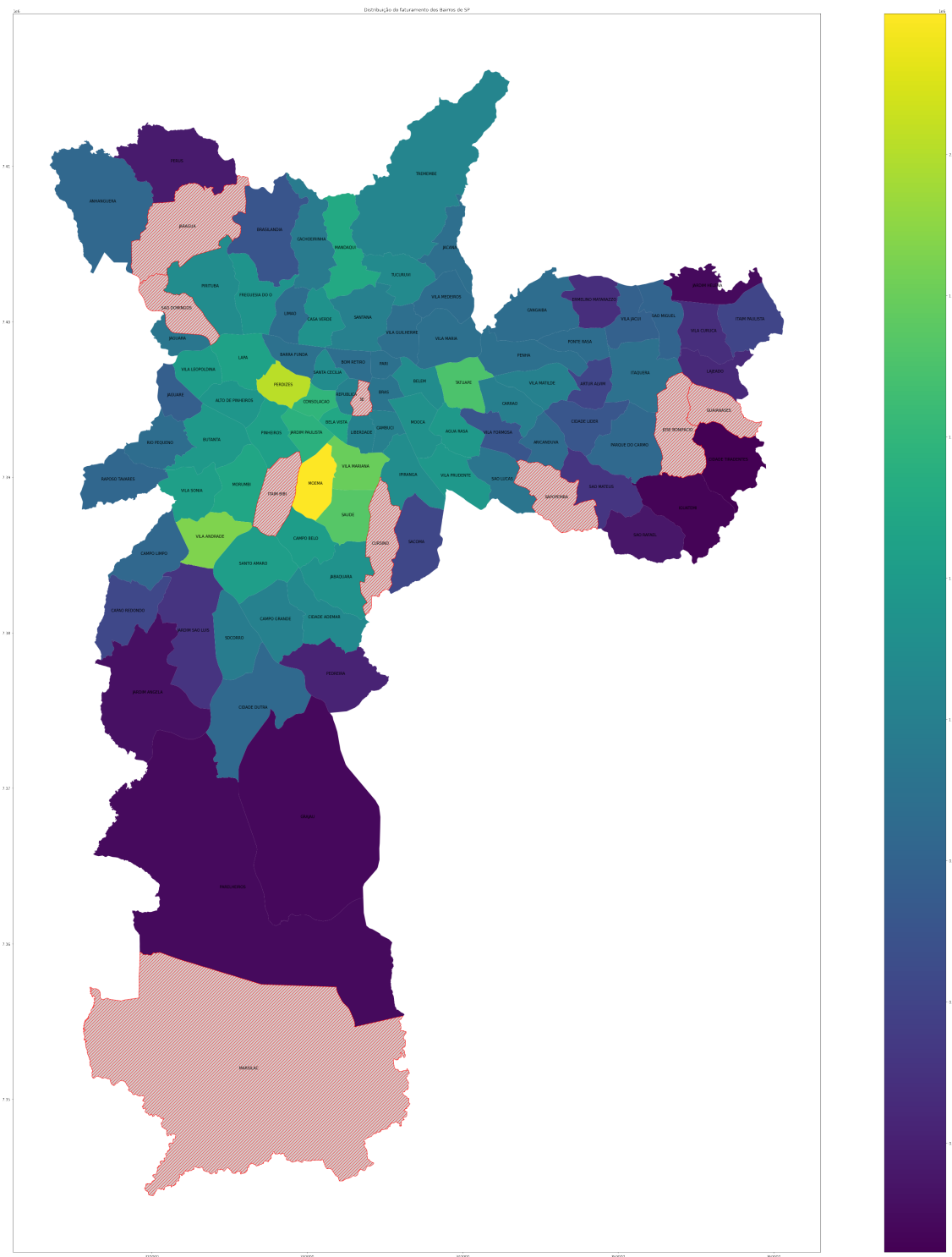


Figura 5. Mapa com a segmentação dos bairros do município de São Paulo por faturamento.

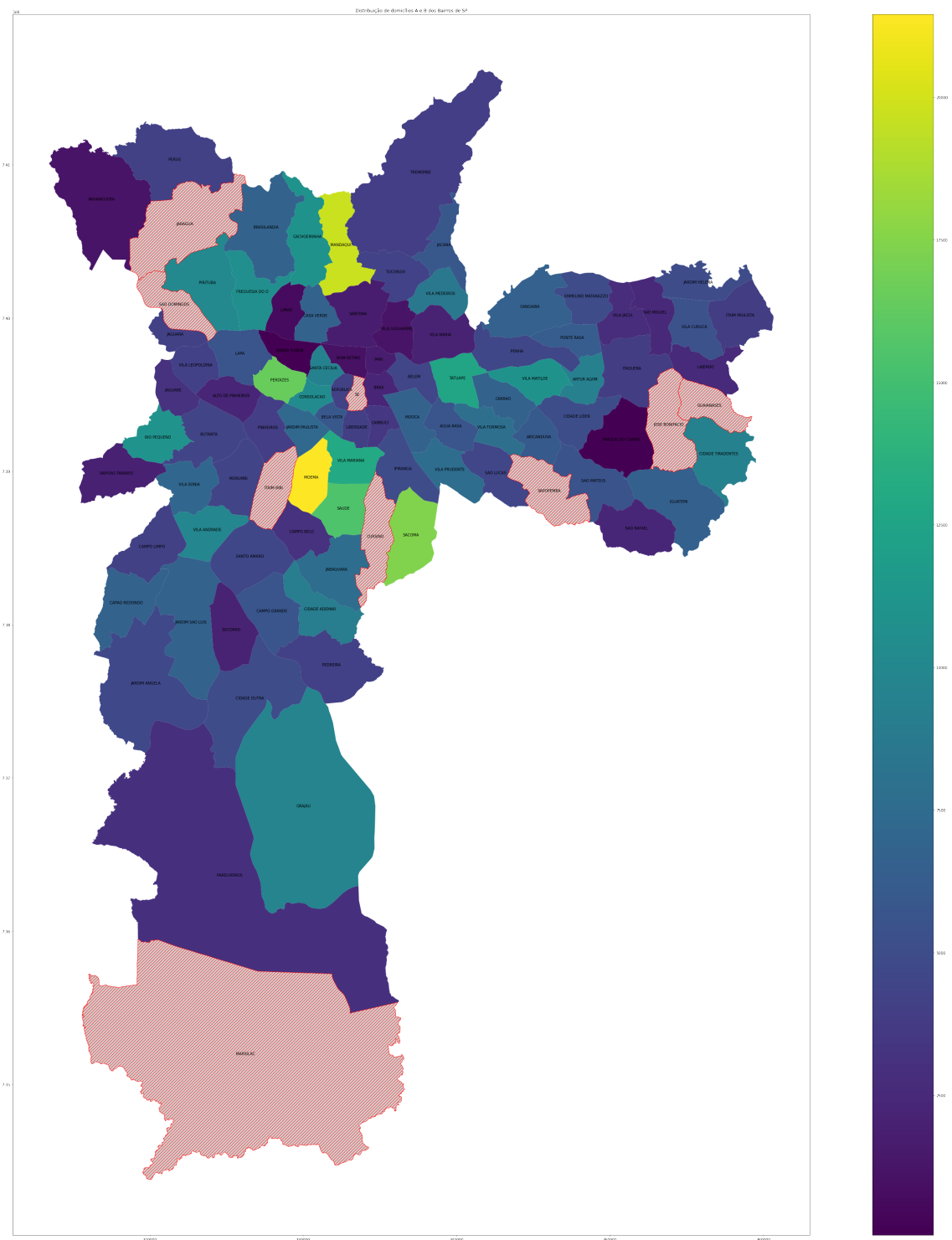


Figura 6. Mapa com a segmentação dos bairros do município de São Paulo por domicílios de renda A e B.

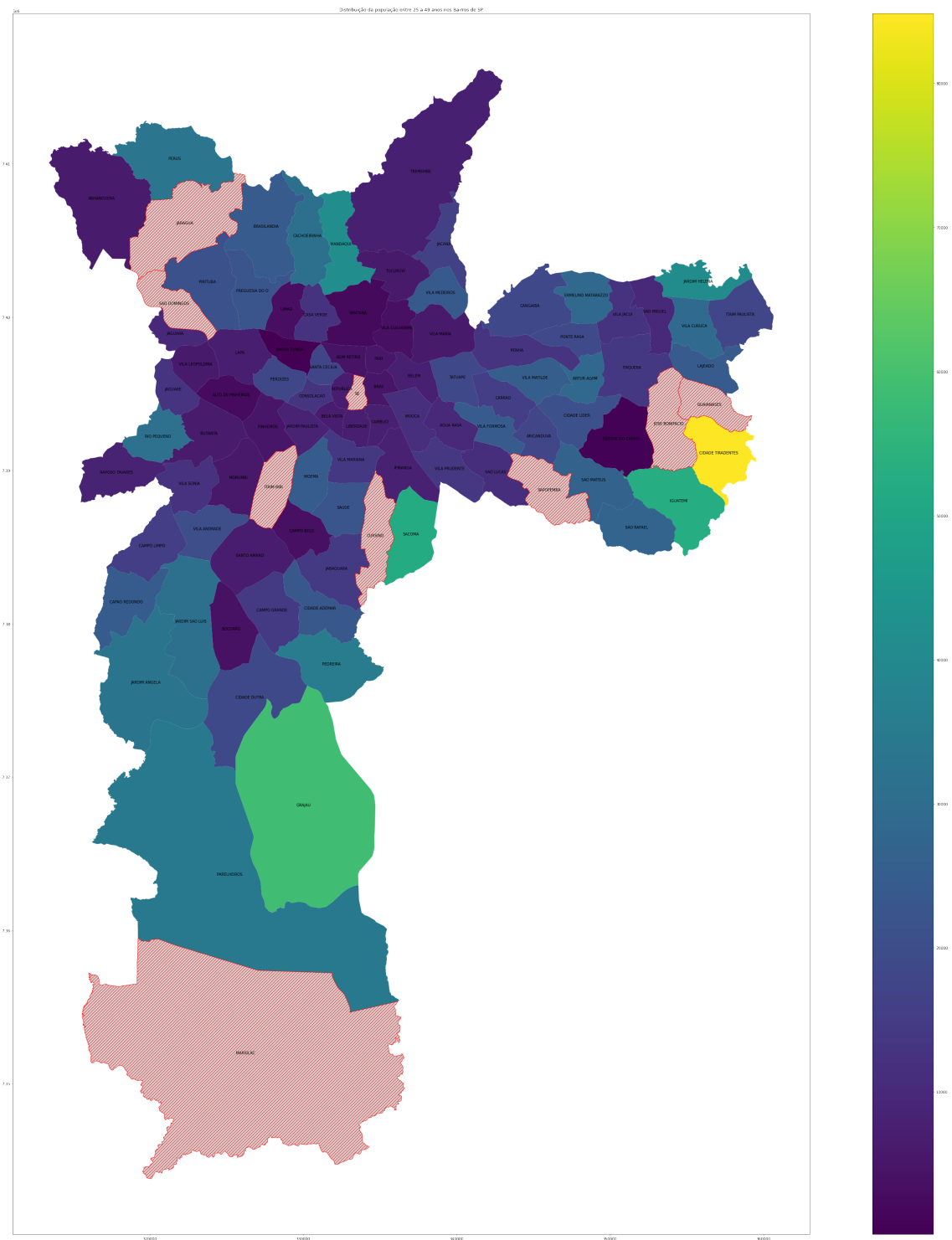


Figura 7. Mapa com a segmentação dos bairros do município de São Paulo por tamanho da população entre 25 e 49 anos.

