Nota técnica - desafios e potencialidades no desenvolvimento de metodologia de trabalho com dados de alvarás de licenciamento de imóveis residenciais em São Paulo

Por Evandro Luis

Em 27 de abril de 2024

Table of Contents

# 1 Introdução

Este artigo apresenta a metodologia de trabalho desenvolvida pela equipe do Laboratório Arq.Futuro de Cidades do Insper para tratamento dos dados de alvarás de licenciamento de imóveis residenciais disponibilizados pela Prefeitura de São Paulo e faz apontamentos sobre desafios e potencialidades de se discutir políticas urbanas baseadas em evidências a partir deste conjunto de dados públicos. A metodologia foi utilizada para a produção de quatro notas técnicas apresentadas ao Fórum SP 23 que debatem possíveis impactos do PDE de São Paulo a partir dos dados de licenciamento de imóveis e outros dados complementares. O estudo ocorre no âmbito do **Projeto Acesso a Oportunidades no PDE de São Paulo do Laboratório Arq.Futuro de Cidades do Insper**, que tem por objetivo contribuir com os debates de monitoramento e avaliação do atual PDE de São Paulo e suas leis complementares. Após discorrer sobre o percurso metodológico de construção do conjunto de evidências organizado para a elaboração das notas técnicas, trazemos ao fim do artigo uma série de recomendações relacionadas à viabilização do melhor uso deste conjunto de dados como produção de evidência para planejamento e gestão urbana.

# 2 Contexto

## 2.1 Dados imobiliários em São Paulo e opção pelos dados da PMSP/SMUL

Como consequência das perguntas de pesquisa estabelecidas para o desenvolvimento do **Projeto Acesso a Oportunidades no PDE de São Paulo do Laboratório Arq.Futuro de Cidades do Insper**, foi colocado para a equipe a necessidade processar e interpretar dados que informassem sobre a dinâmica imobiliária. Após análise comparativa sobre dados disponíveis e acessíveis sobre a dinâmica imobiliária de São Paulo que cobrissem intervalos de tempo apropriados para a discussão sobre o atual PDE, chegou-se ao seguinte panorama geral de conjuntos de dados elegíveis para a pesquisa e capazes de subsidiar informações a respeito da dinâmica imobiliária de São Paulo no período de interesse:

* EMBRAESP/Secovi/BRAIN - dados de direito privado que trazem registros de lançamentos imobiliários georreferenciados na cidade de São Paulo inclusive registros de preços de venda e de transação;
* DataZap+/Loft/Quinto Andar/Viva Real - dados de direito privado que trazem registros de ofertas de unidades habitacionais em sites de anúncio imobiliário inclusive registros de preço de venda, mas possuem grande ruído estatístico e dificuldades de expansão da amostra para um universo de casos;
* IPTU/ITBI - dados de natureza pública que contém registros importantes sobre o uso do lote e de transações envolvendo o lote, mas que não trazem grandes informações sobre empreendimentos executados neste lote;
* Alvarás PMSP/SMUL - dados de natureza pública que contêm registros das várias etapas do licenciamento imobiliário, mas que não trazem registros do lançamento e informações como preços de venda e transação.

Após análise do panorama geral de alternativas, optou-se por desenvolver a análise da dinâmica imobiliária requerida pelo **projeto priorizando o uso dos dados de alvarás de licenciamento imobiliário da PMSP/SMUL**. Podemos definir como razões principais para esta escolha:

* o fato deste conjunto de dados conter licenciamentos de empreendimentos de pequeno porte e/ou produzidos para faixas de renda mais baixa que não fazem lançamento formal, o que torna este conjunto de dados potencialmente mais representativo do universo de edificações novas na cidade de São Paulo do que as alternativas que têm o lançamento imobiliário como menor unidade de agregação;
* a possibilidade de se construir informação sobre futuros lançamentos imobiliários desde o momento do licenciamento de parcelamento de solo, aprovação de projeto e/ou execução de obra, o que possibilita retratar a melhor a dinâmica imobiliária desde o momento de real tomada de decisão de se incorporar o terreno para uma edificação nova e também torna este conjunto de dados potencialmente mais atualizado perante às alternativas;
* a oportunidade de se ter como um dos produtos da pesquisa o aprimoramento de uma informação de caráter público, de acesso gratuito e de interesse social mais generalizado, o que torna o projeto orientado a contribuir diretamente com a superação de um importante gargalo informacional sobre os processos de licenciamento imobiliário em São Paulo e amplia o seu potencial de impacto social;
* a oportunidade de estabelecer uma cooperação técnica com iniciativa pré-existente da Associação Brasileira de Incorporadores Imobiliários - ABRAINC dedicada ao tratamento deste conjunto de dados;

Essa escolha implicou, porém, em enfrentar o considerável desafio de transformar este conjunto parcialmente estruturado de dados administrativos em um conjunto mais interpretável e mais confiável de informações. Sendo assim, o projeto dedicou recursos para desenvolver um pré-processamento dos dados que fosse detalhado, reprodutível e funcional para processos de monitoramento e avaliação. De modo tal que, além do percurso de análise sobre os dados que é apresentada nas demais notas técnicas submetidas ao Fórum SP 23, tornou-se produto do projeto e tema da presente nota técnica o próprio conjunto tratado de dados de alvarás de licenciamento imobiliário fornecidos pela PMSP/SMUL bem como todo os procedimentos realizados em linguagem de programação para extrair este conjunto dos dados originalmente.

### 2.1.1 Benchmarking

Como forma de entender o estado da arte em relação ao uso dos dados da PMSP/SMUL, foi feita uma análise geral de benchmarking com experiências que tivessem acumulado aprendizado no manejo deste conjunto de dados. A busca resultou em um número escasso de iniciativas que se referem a este conjunto de dados, seja para pesquisa ou para uso em algum sistema/ aplicação em produção. Podendo ser destacadas:

* [Portal de Olho na Obra - por PMSP/SMUL](https://www.prefeitura.sp.gov.br/cidade/secretarias/subprefeituras/sp_mais_facil/de_olho_na_obra/index.php?p=1053/)
* [Plataforma GeoSampa/ Item Sistema Eletrônico de Licenciamento de Construções (SLCe) - por PMSP](https://www.prefeitura.sp.gov.br/cidade/secretarias/licenciamento/noticias/?p=265348)
* [SISSEL - Relatório de Alvarás e documentos emitidos - por PMSP/SMUL](https://www.prefeitura.sp.gov.br/cidade/secretarias/licenciamento/servicos/index.php?p=3334)
* Sistema Urbinet - por ABRAINC
* [Relatório 2021: Monitoramento e Avaliação do 2014-2020 - PMSP/SMUL](https://gestaourbana.prefeitura.sp.gov.br/wp-content/uploads/2021/12/Relat%C3%B3rio-de-Monitoramento-do-PDE-2014-2020.pdf)
* [Dissertação - A Eficácia das ZEIS de imóveis vazios ou subutilizados no Município de São Paulo (2019) - por Rosana Yamaguti](http://lepur.com.br/wp-content/uploads/2019/07/Dissertacao_RosanaYamaguti_revfinal.pdf)
* [Tese - "ZEIS de vazios" em São Paulo 2002-2014: produção habitacional, transformações e permanências do estoque de terra (2018) - por Rodrigo Minoru Hayakawa Tanaka](https://www.teses.usp.br/teses/disponiveis/16/16139/tde-10122018-110131/en.php)

Atribui-se este cenário de pouca pesquisa e desenvolvimento em relação a essa base de dados ao fato de existir um considerável grau de complexidade na implementação do tratamento dos dados de licenciamento imobiliário publicizados pela PMSP/SMUL, sobretudo para abordagens de analíticas do tipo low-code ou no-code. O que configura uma importante barreira à entrada aos pesquisadores, servidores públicos, empreendedores e ativistas que são parte interessada nas informações oferecidas por este conjunto de dados.

Vale mencionar que A PMSP oferece acesso a um conjunto tratado dos dados de alvarás por meio da [seção Dados Abertos da plataforma Monitoramento PDE](https://monitoramentopde.gestaourbana.prefeitura.sp.gov.br/dados-abertos/). Este conjunto, contudo, não foi considerado como tendo sido alvo de tratamento suficiente para responder as perguntas de pesquisa, em particular pela ausência de uma metodologia de agregação das informações de alvarás por empreendimento.

### 2.1.2 Framework

Todo o processo de tratamento e análise exploratória de dados de dinâmica imobiliária em São Paulo foi elaborado em **linguagem R** com intermédio da **IDE RStudio** e da ferramenta **RMarkdown** para produção de relatórios automatizados. Ao longo da seção "Alvarás PMSP/SMUL" do presente documento apresenta-se as decisões de tratamento e suas justificativas.

Além do uso dos algoritmos básicos disponibilizados no R, também foram algoritmos adicionais importados de livrarias/pacotes de repositórios públicos disponibilizados no CRAN – Compreensive R Archive Network que são relatados ao início dos scripts conforme convenção da comunidade de desenvolvedores. E cabe destacar também que para a confecção de mapas o software QGis também foi utilizado como recurso em momentos oportunos e que a produção dos textos que dissertam sobre achados da análise de dados elaborada por meio da ferramenta Microsoft Word/ Microsoft Word Online.

Para fins de compromisso com a reprodutibilidade e a replicabilidade da pesquisa planeja-se tornar público em momentos subsequentes à finalização do Projeto Acesso a Oportunidades no PDE, para além do que é exposto nesta nota técnica sobre as decisões de tratamento, o conjunto de scripts e procedimentos de tratamento dos dados desenvolvidos em linguagem R em formato open source.

### 2.1.3 Parcerias

A implementação do tratamento dos dados de licenciamento imobiliário de São Paulo exposto neste texto foi fruto de uma cooperação técnica estabelecida entre o Laboratório Arq.Futuro de Cidades do Insper e a Associação Brasileira de Incorporadores Imobiliários - ABRAINC. E é importante destacar que grande parte das decisões de tratamento foram inspiradas ou mesmo herdadas de iniciativa predecessora implementada do Sistema Urbinet - que é desenvolvido e mantido pela ABRAINC.

As rotinas de tratamento dos dados da PMSP/SMUL pelo Urbinet foram compartilhadas com a equipe do projeto Acesso a Oportunidades no PDE de São Paulo e traduzidas da linguagem Java para a linguagem R. Exercício este entendido também como oportunidade de não apenas traduzir literalmente decisões de tratamento, mas de fazer aperfeiçoamentos e de se gerar mais chances de produção de informação relevante do que se tinha no ponto de partida.

O processo de validação do tratamento de dados também contou com o apoio de equipe técnica da ABRAINC e com conhecimentos específicos da instituição acumulados com a experiência prática de licenciar imóveis em São Paulo. Além da disponibilização de horas técnicas para verificação da consistência do novo conjunto de dados, a ABRAINC também se dispôs a contribuir com reuniões periódicas de avaliação parcial de produtos e com 8 reuniões de alinhamento após a implementação da nova rotina de tratamento exclusivamente dedicadas ao tema.

# 3 Alvarás de licenciamento imobiliário da PMSP/SMUL

Ao longo desta seção descreve-se em maiores detalhes como se deram os processos de obtenção, tratamento e produção de informação a partir dos dados de alvarás de licenciamento imobiliário da PMSP/SMUL e outros dados complementares utilizados para contextualização dos dados originais.

## 3.1 Dados brutos

O conjunto original de dados de licenciamentos imobiliários em São Paulo é constituído de tipos diversos de alvarás de licenciamento imobiliário residencial e comercial registrados na cidade entre 2000 e 2022. A série histórica procede de registros cadastrais da Prefeitura de São Paulo gerados pelo Sistema de Administração do Código de Obras e Edificações (SISACOE) e tornados acessíveis pelo intermédio do Sistema SEL (SISSEL). Os registros do SISSEL são disponibilizados para o público geral em formato de planilhas (.xlsx) no site oficial da Prefeitura através do acesso ao link: < <https://www.prefeitura.sp.gov.br/cidade/secretarias/licenciamento/servicos/index.php?p=3334> >.

Não foram encontradas muitas explicações acerca do funcionamento do SISACOE e do SISSEL e nem de como a relação entre os dois sistemas produz os relatórios disponibilizados ao público. Segundo informação oferecida em uma notícia no sítio oficial da prefeitura, podemos inferir que o SISSEL é um aplicativo que visa facilitar a interface do SISACOE com o usuário (o servidor público responsável por realizar licenciamento) e, ao mesmo tempo, gerar relatórios automatizados integrados a sistemas de registros de processo administrativo como SIMPROC e SEI:

[…] o sistema (SISSEL) agiliza o licenciamento de obras, uma vez que reúne de maneira prática todas as informações necessárias para a emissão de um alvará. Tabelas que antes eram feitas manualmente agora são produzidas de modo automatizado e padronizado, o que evita erros e torna a análise mais rápida. Interligado com o SISACOE e SIMPROC, sistemas da Prefeitura que atualmente são utilizados na aprovação de processos, e com uma interface simples e prática, o SISSEL oferece mais autonomia aos técnicos da secretaria, já que eles mesmos poderão produzir e emitir um alvará. O tempo médio de emissão do documento passa de cerca de seis horas para apenas doze minutos. O aplicativo também permite que os próprios técnicos possam tramitar processos e documentos para seus supervisores, sem que o processo precise passar pelo expediente, o que antes gerava gargalos que dificultavam o trabalho1.

De acordo com o catálogo de dados da PMSP, os dados do SISACOE são de responsabilidade da SMUL, mais especificamente da Coordenadoria de Cadastro, Análise de Dados e Sistema Eletrônico de Licenciamento - CASE e da Supervisão de Licenciamento Eletrônico e Análise de Dados - STEL. E segundo o decreto nº 60.061/2021 (decreto de competências da SMUL) é atribuição da SMUL/CASE/STEL a gestão do SISACOE além da produção de relatórios e indicativos relacionados à política municipal de licenciamento:

Art. 64. A Supervisão de Licenciamento Eletrônico e Análise de Dados - STEL tem as seguintes atribuições:   
  
I - implantar, monitorar e o Sistema Eletrônico de Licenciamento;   
  
II - armazenar e analisar os dados de licenciamento produzidos no âmbito da SMUL;   
  
III - produzir relatórios técnicos e estatísticos para subsídio da gestão da Política Municipal de Licenciamento;   
  
IV - elaborar indicadores relativos aos processos de licenciamento;   
  
V - atuar no atendimento e suporte técnico aos servidores e munícipes usuários do Sistema.

## 3.2 Decisões de tratamento

Para a construção de um conjunto de dados mais coeso e mais confiável como base de evidências, foi necessário aplicar aos compilados anuais de licenciamentos disponibilizados no sítio oficial da prefeitura uma série de decisões de seleção, visualização, georreferenciamento e engenharia de atributos. Além disso, há importantes inconsistências no registro de licenciamento disponibilizado pela PMSPM/SMUL via SISSEL/SISACOE que demandam tratamento muito específico, tais como duplicatas, erros de digitação e, principalmente, a computação de variáveis numéricas que estão em formato de texto.

Podemos listar como decisões críticas de tratamento:

* **Automatização da importação de dados** - importação de dados brutos mediante raspagem de dados (web scrapping) do sítio da PMSP/SMUL na web;
* **Construção da série histórica com atributos padronizados** - estruturação da série histórica a partir dos dados originais separados anualmente, com a redefinição de alguns nomes de atributos;
* **Normalização dos registros de blocos, pavimentos e unidades** - contabilização dos valores de Blocos, Pavimentos e Unidades que estavam combinados em uma única variável, momento que exigiu o uso de expressões regulares combinadas com ferramentas auxiliares para operações vetoriais/matriciais;
* **Definição dos tipos de alvarás relevantes** - Definição dos tipos de alvarás de licenciamento imobiliário relevantes para a monitoramento e avaliação de licenciamentos de imóveis residenciais em São Paulo;
* **Agrupamento de alvarás por empreendimento** - Agrupamento dos alvarás por lote/empreendimento e construção do conjunto de dados alvaras\_por\_lote.parquet a partir da série histórica de alvarás, que agrega as informações de alvarás por código Setor-Quadra-Lote ou código INCRA para imóveis rurais;
* **Georreferenciamento dos empreendimentos** - georreferenciamento pela base de lotes oficial na plataforma GeoSampa (cerca de 1/3 dos lotes de interesse) e geolocalização dos endereços via API's (cerca de 2/3);
* **Ajustes em observações com parcelamento de solo** - identificação e tratamento de casos de parcelamento de terreno erroneamente atribuídos a um único empreendimento e, por consequência, subnotificando algumas contagens de atributos.
* **Seleção de alvarás e empreendimentos de interesse** - seleção de alvarás e de empreendimentos/lotes de interesse para a pesquisa;
* **Engenharia de novos atributos** - engenharia de novos atributos para construção de indicadores de dinâmica imobiliária, mediante análises qualitativas, operações vetoriais/matriciais e estatística descritiva.

A seguir trazemos comentários adicionais sobre cada uma destas etapas, contextualizando a tomada de decisão nos limites e nas potencialidades identificados neste conjunto de dados.

#### 3.2.0.1 Automatização da importação de dados

A rotina de tratamento desenvolvida faz a raspagem (web scrapping) dos dados via script e faz outras etapas de pré-processamento de dados também via scripts. O formato de publicização dos dados via web é o de download de planilhas .xls/.xlsx pelo usuário via navegador. Foi entendimento da equipe do projeto, porém, que o uso da compilação de código implicaria em uso de melhor prática do que o de ferramentas do estilo point-to-click, oferecendo uma forma semiautomatizada de executar incorporação dos dados organizadas que reduz chances de erro por consequência de manipulação imprópria de planilhas via Microsoft Excel ou software similar.

#### 3.2.0.2 Construção da série histórica com atributos padronizados

A série histórica não está plenamente construída nos dados originais, que são separados em planilhas referentes a um ano de operação que é agregação de registros mensais. Para estruturar a série histórica foi necessário redefinir alguns nomes de atributos, identificando de forma quase qualitativa mudanças de nomeação para um mesmo atributo em anos diferentes.

As planilhas ano-a-ano estão organizadas de forma pouco regular, por vezes com registros em abas diferentes e colunas vazias, exigindo tratamento caso a caso. E possuem indicativos de registros corrigidos manualmente em momento posterior à sua primeira elaboração, o que não é considerado boa prática e que pode afetar a qualidade e a confiabilidade dos registros.

Vale destacar que os nomes dos atributos nos diferentes anos (o conteúdo dos cabeçalhos das planilhas) foram utilizados como referência para estabelecer a relação dos atributos entre os anos - diferentemente, por exemplo, da versão do Sistema Urbinet da ABRAINC que foi compartilhada com a equipe do projeto, que fazia esta relação pela posição do atributo na planilha.​

Além disso, a série histórica organizada pela equipe do Laboratório Arq.Futuro de Cidades não incluiu os alvarás emitidos entre 2000 e 2003, sendo iniciada a partir de 2004. Esta decisão se deu em função da ausência de registros de quantidades de blocos, pavimentos e unidades para o ano de 2003.

Importante mencionar também que, durante o período de desenvolvimento do trabalho com os dados, o mesmo tipo de problema de ausência de dados apresentado em 2003 surgiu novamente 18 anos depois. A atualização das planilhas do primeiro semestre de 2021 não trouxe consigo registro de unidades, por razões que não ficam claras a partir das explicações da prefeitura. Esta ausência teve um efeito negativo na pesquisa, prejudicando conclusões que envolviam o ano de 2021, mas foi ainda sim mantida na série histórica.

#### 3.2.0.3 Normalização dos registros de blocos, pavimentos e unidades

Predomina na série histórica original a apresentação dos atributos "blocos", "unidade" e "pavimentos" combinados em um único atributo separado por hífen "blocos-pavimentos-unidades". Mais especificamente, este atributo apresenta-se geralmente em padrão similar ao seguinte: "B - 01 - P - 02 - U - 04", exemplo em que teríamos aqui um empreendimento com 1 bloco, 2 pavimentos por bloco e 4 unidades por bloco.

Além de dificultar a contabilização de valores numéricos, este formato apresenta padrão irregular ao longo da série histórica. As observações/linhas por vezes contém mais de um registro de "B-P-U" por linha, exigindo, portanto, a agregação de dados do atributo naquela observação. E também são declarados em determinados períodos na forma invertida "U-P-B".

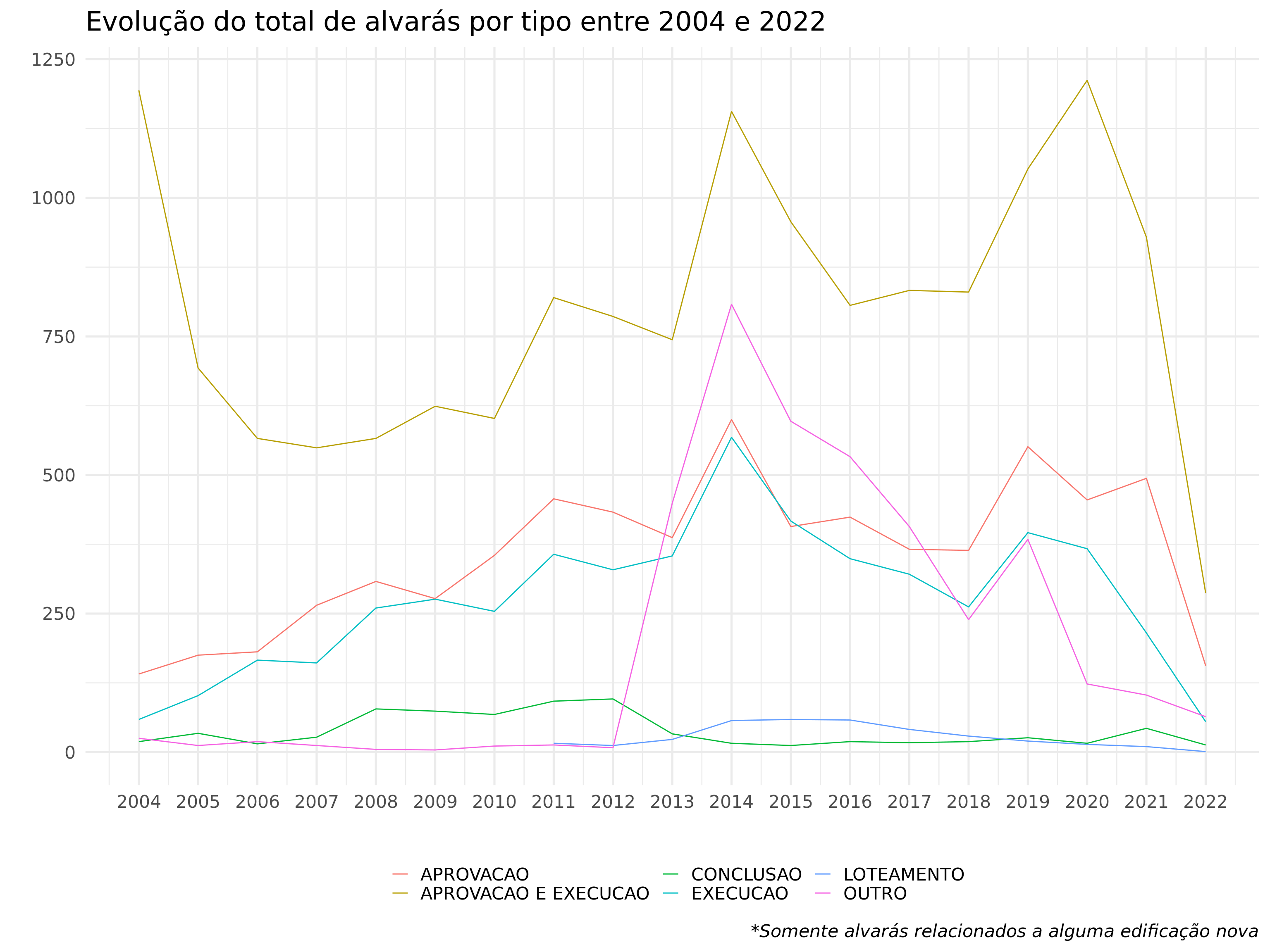
O tratamento deste atributo exigiu, portanto, atenção especial, tendo sido realizado mediante o uso de expressões regulares (regex) antes da transformação das informações em registros numéricos. Seguido de uma combinação dos valores numéricos, feita com uso de ferramentas auxiliares para operações vetoriais/matriciais oferecidas pela linguagem R e pacotes desenvolvidos pela comunidade de desenvolvedores.

#### 3.2.0.4 Definição dos tipos de alvarás relevantes

Os dados brutos possuem vários tipos de alvarás relacionados ao processo de execução de obras em São Paulo, tanto residenciais quanto comerciais, sendo cada alvará referente a uma determinada etapa de um processo de licenciamento de um empreendimento em determinado/s lote/s. De um modo tal que se torna necessário, como um dos primeiros passos da extração de informações a partir dos dados originais, definir quais desses diversos alvarás de licenciamento imobiliário são relevantes para a compreensão da dinâmica de licenciamento de imóveis residenciais em São Paulo e para a agregação posterior dos dados por lote/empreendimento.

Após análise qualitativa e rodadas de validação com especialistas da ABRAINC com experiência acumulada em processos de licenciamento imobiliário em São Paulo, foram identificados como tipos relevantes para análise e inferência os alvarás de Edificação Nova do tipo Aprovação, Execução, Aprovação e Execução além de alvarás de loteamento/desmembramento de gleba e de conclusão de obra.





#### 3.2.0.5 Agrupamento de alvarás

A abordagem aos dados disponibilizados pela Prefeitura deve atentar-se ao fato de que uma série de alvarás de tipos distintos são a unidade de observação básica do conjunto de dados - ou seja, as linhas das tabelas - e que várias observações listadas podem se referir a um mesmo empreendimento imobiliário. Sendo assim, torna-se pressuposto básico para uma análise de dinâmica imobiliária tal qual a que fizemos para as notas técnicas a realização de um agrupamento de alvarás por lote/empreendimento imobiliário.

As operações de agregação dos atributos numéricos alternam-se entre a soma e a média dos atributos dos alvarás individualizados que têm contabilização condicionada ao fato de serem ou não tipificados como relevantes. Além disso, a agregação pressupõe escolher um dos vários registros de datas de alvarás como ano de registro do empreendimento. Para os estudos referentes aos alvarás de licenciamento imobiliário no âmbito do Projeto Acesso a Oportunidades no PDE de São Paulo, foi atribuído como ano do empreendimento o ano do seu último alvará de execução[[1]](#footnote-44).

#### 3.2.0.6 Ajustes em observações com parcelamento de solo

Após rodadas de reuniões de cooperação técnica para validação dos produtos intermediários, os especialistas da ABRAINC chamaram a atenção da equipe do projeto ao fato de que havia alguns casos de parcelamento de terreno erroneamente atribuídos a um único empreendimento. A análise detalhada do problema revelou que, apesar de representar um número muito pequeno de observações, a base de fato traz uma limitação no registro original e que existe impacto de subnotificação em algumas contagens de atributos.

Foi desenhado, portanto, um procedimento de ajuste para identificar casos de provável parcelamento ainda no momento de agregação dos dados por lote/empreendimento, de forma a minimizar o impacto de alguns casos com este tipo de problema de registro. Sendo esses casos tratados como dois ou mais empreendimentos em vez de um só.

#### 3.2.0.7 Georreferenciamento de lotes/empreendimentos

Os dados originais não possuem de antemão georreferenciamento em algum Sistema de Informações Geográficas (SIG). Tendo sido selecionados a partir da base agregada por lotes/empreendimentos para georreferenciamento e tratamento em maiores detalhes registros de 2013 a 2021. Este grupo consiste num total de 11095lotes/empreendimentos agrupados a partir da análise e tratamento de um total de 0 de alváras de licenciamento atribuídos a esses mesmos lotes ao longo da série histórica de 2004 e 2022.

O georreferenciamento foi feito no sistema SIRGAS 2000 mediante a identificação do dígito SQL na base de lotes oficial disponibilizada na plataforma GeoSampa (cerca de 1/3 dos lotes de interesse) e geolocalização dos endereços via API’s de georreferenciamento (cerca de 2/3).

#### 3.2.0.8 Seleção de alvarás e empreendimentos de interesse de pesquisa

Após as várias etapas de qualificação e agregação dos dados originais, o produto intermediário do tratamento foi reservado como tal para comparação e novo tratamento foi aplicado a este produto intermediário com a finalidade mais específica de atender às questões de pesquisa do projeto Acesso a Oportunidades no PDE de São Paulo. Sendo questão primeira selecionar alvarás e empreendimentos de interesse.

Foram, então, excluídos da fase final de refinamento dos dados:

* empreendimentos de cunho exclusivamente comercial ou não diretamente relacionados a um alvará de edificação nova;
* empreendimentos que não possuíam tanto o alvará de aprovação de projeto quanto o de execução da obra

#### 3.2.0.9 Engenharia de novos atributos

De modo geral, o tratamento buscou refletir diretamente o sentido dos atributos originais, ainda que com algumas alterações importantes de nomenclatura. Por exemplo, optou-se por tratar o número de unidades como "n\_unidades\_por\_bloco" e não como "unidades", em função do novo nome refletir melhor a métrica efetivamente apresentada no conjunto de dados original.

No primeiro momento de qualificação e agregação dos dados brutos por lote, evitou-se a criação de novos atributos que estivessem muito além da decomposição do atributo blocos-pavimentos-unidades, visando manter o máximo de fidelidade ao formato original trazido pelo dado bruto. Para o momento final, porém, foram combinados e agregados por operações topológicas/espaciais vários atributos novos para responder as questões de pesquisa.

## 3.3 Integração com dados complementares

A transformação de dados de registros administrativos de licenciamento imobiliário em evidências sobre incidência ou não-incidência do PDE e seus processos correlatos implica também na necessidade de integrar os dados que versam sobre a dinâmica de licenciamento stricto sensu a dados complementares que versam sobre aspectos socioespaciais do lugar em que os empreendimentos estão inseridos e sobre a legislação urbana aplicável aos empreendimentos por consequência das diversas territorializações definidas para a cidade de São Paulo.

A PMSP conta com um notável projeto de sistema de informações georreferenciadas, a **Plataforma GeoSampa**, que, via de regra, foi a fonte das informações complementares à análise de licenciamentos imobiliários. Em momentos pontuais em que algum dado complementar necessário não pôde ser coberto pelo catálogo de dados da Plataforma GeoSampa, fez-se também o uso da **Plataforma Gestão Urbana da PMSP/SMUL**, que é explicitamente dedicada a compartilhar conteúdos de monitoramento da legislação urbana de São Paulo e oferece um sistema de informações menos estruturadas.

### 3.3.1 Geografia

A camada base dos mapas apresentados na análise segue o CRS SIRGAS 2000 e é oriunda da plataforma GeoSampa. As subdivisões territoriais de São Paulo são referenciadas ou a) na divisão administrativa em subprefeituras e distritos; ou b) na divisão do território conforme legislação aplicável e regramentos urbanos, em particular a divisão em Macroáreas e em Eixos de Estruturação da Transformação Urbana (EETU's), tais como previstos no PDE 2014 e na Lei de Uso e Ocupação do Solo de 2016. Em alguns momentos também foi adicionada à camada geográfica base uma projeção em linhas das redes de infraestutura de transporte que originaram a classificação dos EETU's: linhas de metrô, linhas de Trem e corredores de ônibus.

#### 3.3.1.1 EETU’s

A atribuição de infraestruturas de transporte aos Eixos de Transformação Urbana demarcados ou previstos fez uso do banco de dados da plataforma GeoSampa, tendo sido implementada mediante a combinação das shapefiles referentes às quadras EETU's assim definidas no PDE 2014 com aquelas referentes aos corredores de ônibus e às estações de metrô e trem da cidade.

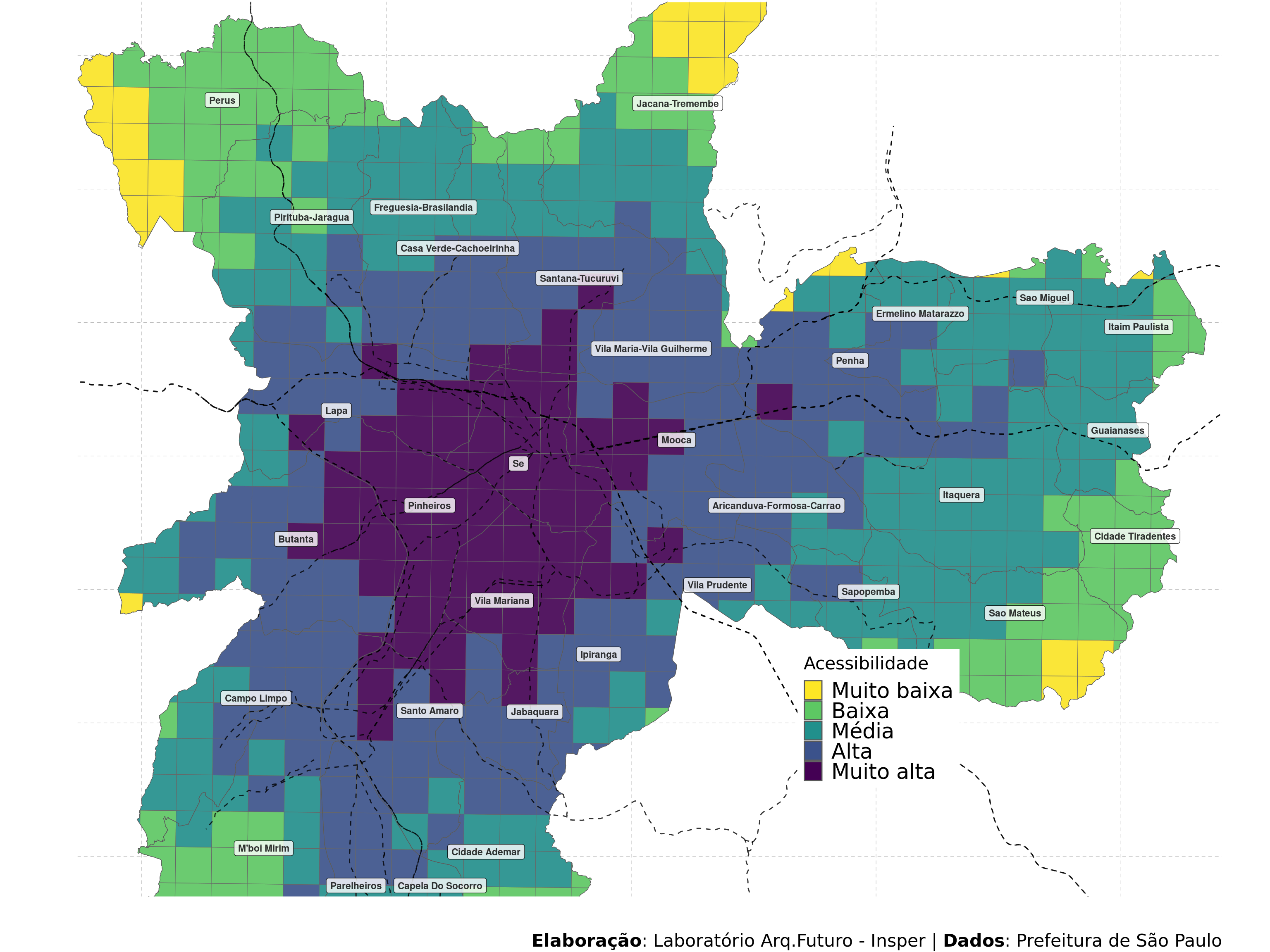
Para identificar qual seria a infraestrutura que "ativa" o eixo, considerou-se a distância euclidiana entre o centroide da quadra-eixo e o centroide da infraestrutura mais próxima além de um critério qualitativo de hierarquização do nível de atratividade de cada tipo de infraestrutura. De modo que o tipo metrô prevaleceu sobre o tipo trem que, por sua vez, prevaleceu sobre o tipo corredor de ônibus.

#### 3.3.1.2 Possíveis eixos na MEM

A identificação de possíveis eixos na Macroárea de Estruturação Urbana - MEM é utilizada na Nota Técnica dedicada à análise dos EETU's. Para construir a tipificação, foram utilizados shapefiles de quadras filtradas pelo perímetro da MEM estabelecido na shapefile referente às macroáreas e combinadas às shapefiles referentes às infraestruturas de transporte - todas também disponíveis na plataforma GeoSampa. Cabe destacar que quadras inseridas em áreas de Operação Urbana não foram consideradas como possível EETU.

#### 3.3.1.3 Acessibilidade

Em determinado momento é também utilizado o indicador de acessibilidade a empregos, como indicador de acesso a oportunidades no entorno do empreendimento. Esta camada complementar de dados faz uso de Pontos de Interesse (POI’s) combinados às informações disponibilizadas na plataforma GeoSampa. Os cálculos de acessibilidade a empregos foram retirados de COSTA, Adriano B.; RAMOS, Camila; ZHENG, Siqi. *Subway expansion, job accessibility improvements, and home value appreciation in four global cities*: Considering both local and network effects. The Journal of Transport and Land Use. v. 15 n. 1, pp. 1-21, 2022. <http://dx.doi.org/10.5198/jtlu.2021.2146.>



## 3.4 Dados tratados

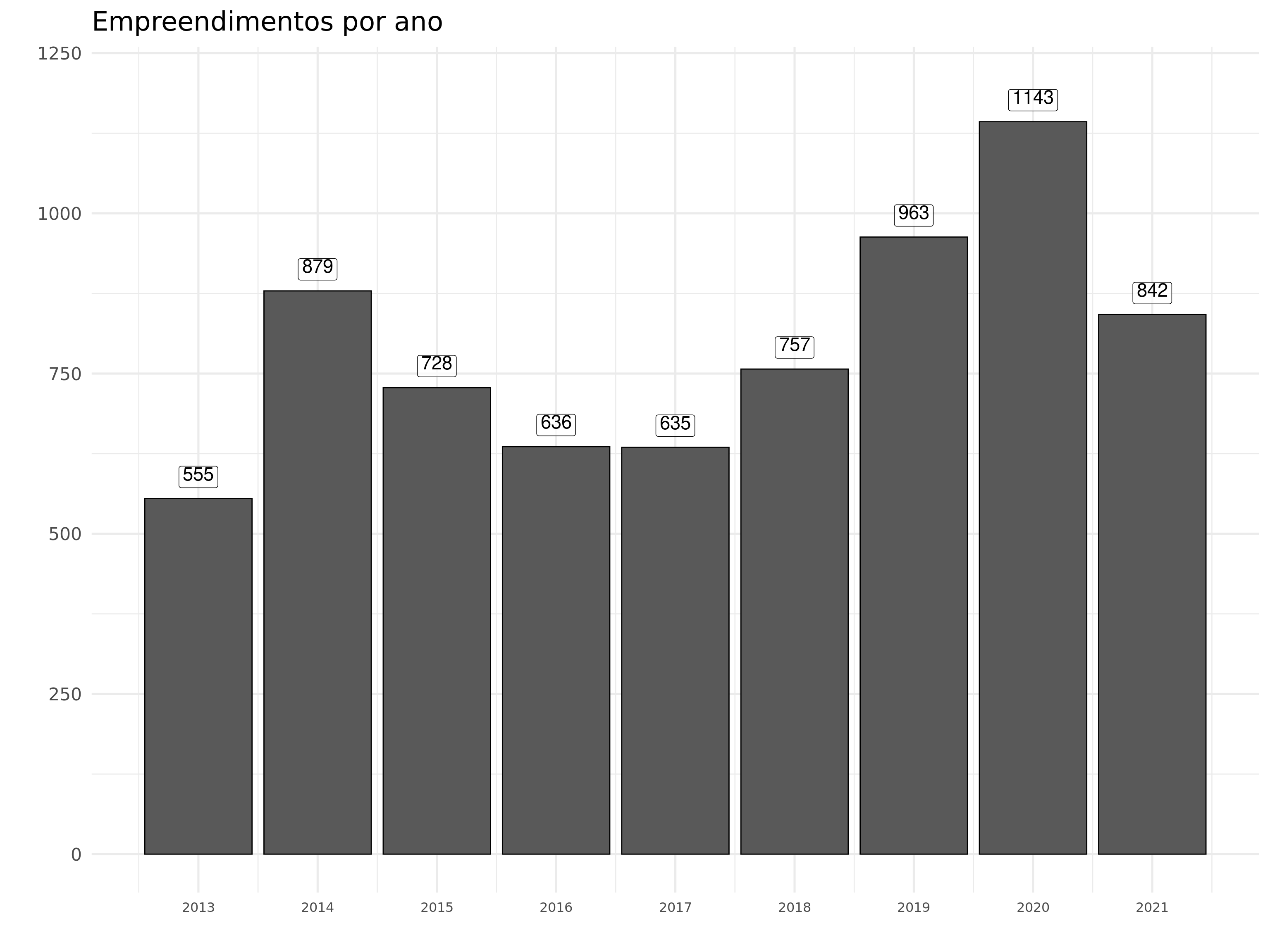
O conjunto de dados georreferenciados **alvaras\_por\_lote.parquet** é a resultante da série de tratamentos aplicados aos dados brutos do SISACOE e é o grande ponto de referência para as análises debatidas nas notas técnicas. Tendo sido identificado após todos os procedimentos de extração, transformação e carregamento de dados um total de 7138 empreendimentos residenciais licenciados para execução de edificação nova em São Paulo entre 2013 e 2021.



### 3.4.1 Distribuição

Em termos da distribuição dos lotes/empreendimentos ao longo da série histórica, podemos observar valores mais altos em anos recentes.

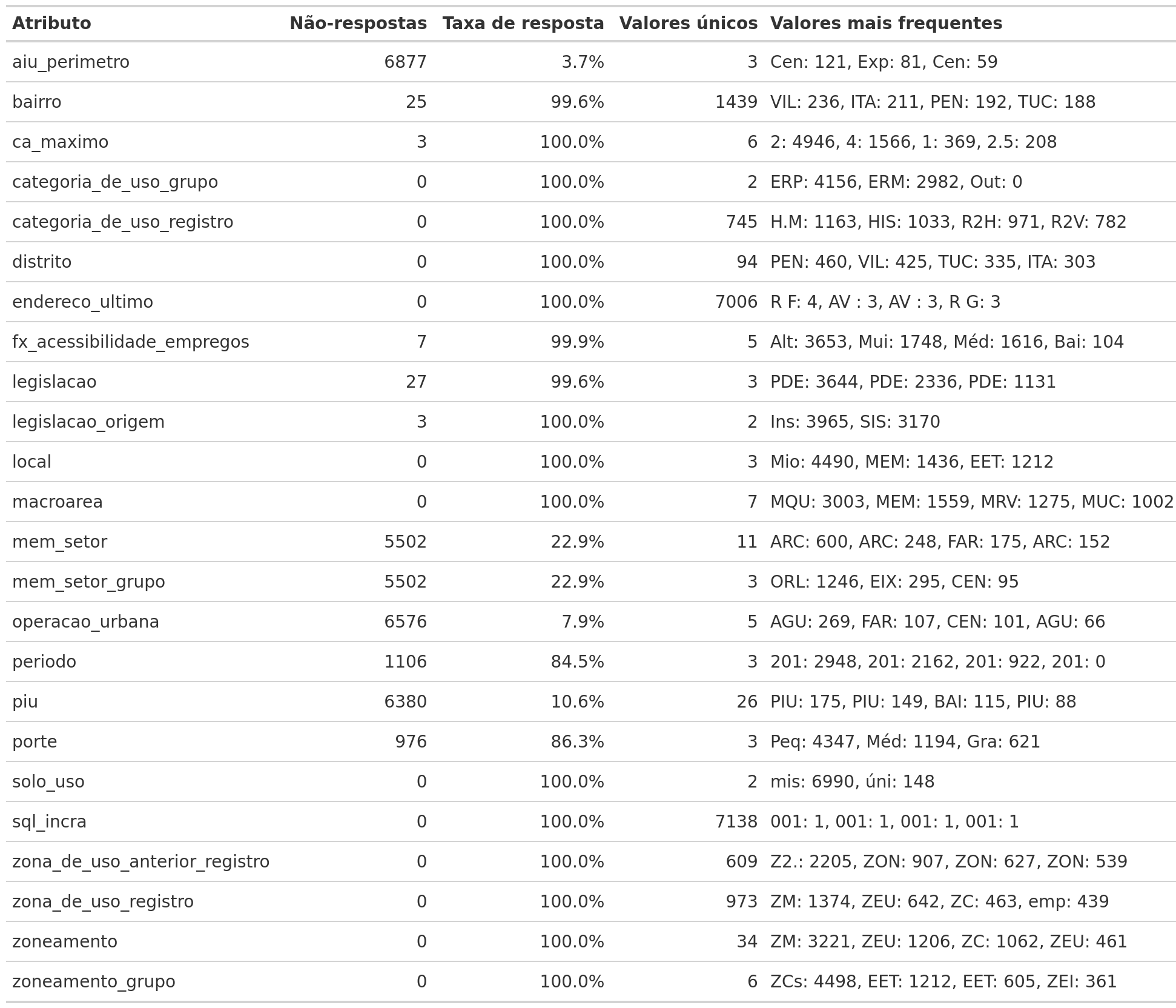
#### 3.4.1.1 Distribuição ao longo da série histórica



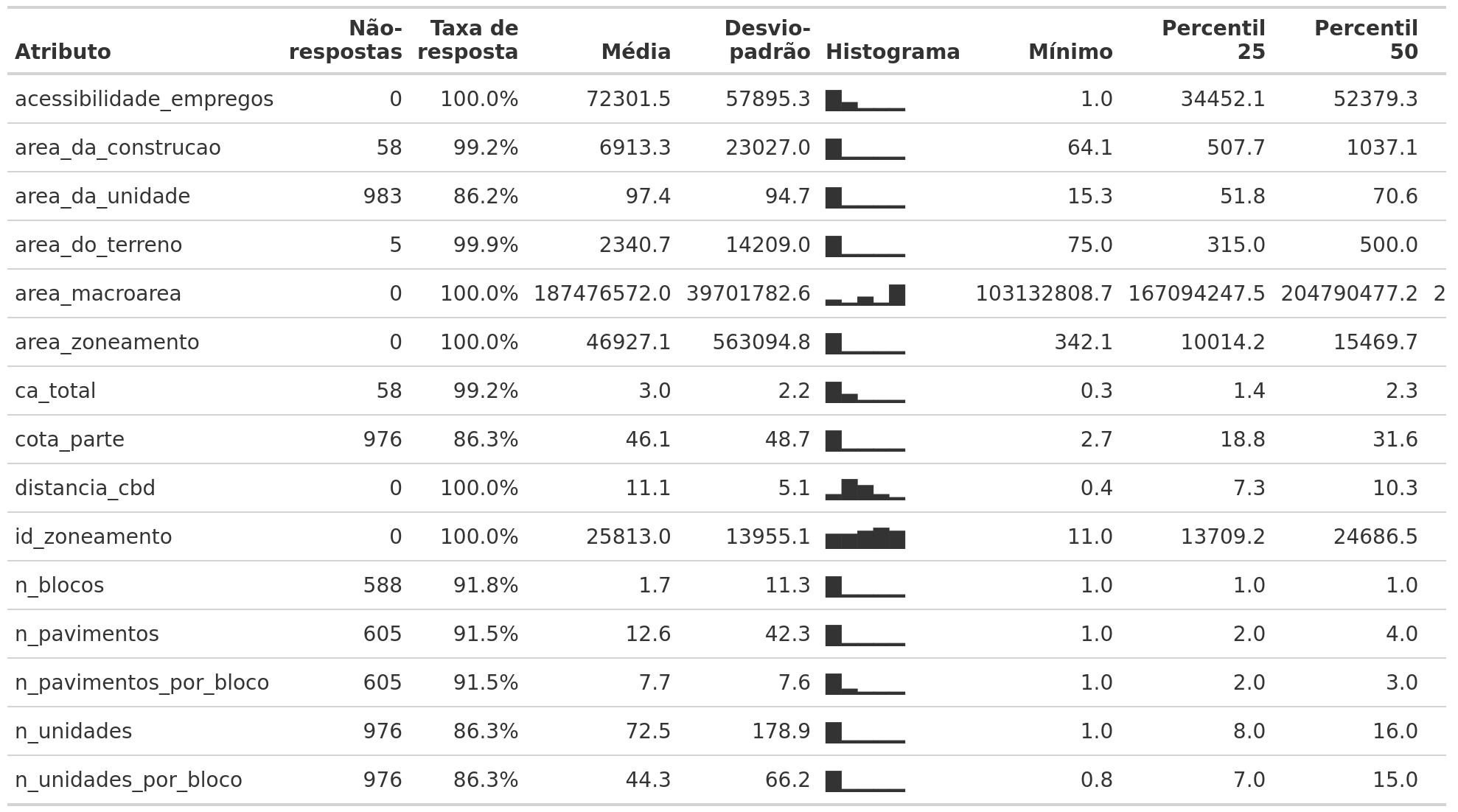
### 3.4.2 Atributos

Os atributos dos lotes/empreendimentos nos informam questões como ano de execução, categoria de uso, zoneamento, total de unidades habitacionais, total de área de construção, acesso a oportunidades de emprego no entorno do empreendimento, cota parte, etc. Os gráficos a seguir trazem breve resumo estatístico dos atributos do conjunto de dados alvaras\_por\_lote:

#### 3.4.2.1 Atributos do tipo categórico



#### 3.4.2.2 Atributos do tipo numérico



#### 3.4.2.3 Atributos do tipo lógico



#### 3.4.2.4 Correlações



### 3.4.3 Tipologias de agrupamento (análises qualitativas)

A definição sobre como agregar alguns dos atributos exigiu maior nível de análise qualitativa da equipe do projeto e merece ter uma explicação detalhada.

#### 3.4.3.1 Legislação aplicável (PDE e LPUOS)

O conjunto de dados original não traz consigo um registro que é em tese bastante simples, mas muito importante para a contextualização dos dados: a regulação urbana que incide sobre aquele determinado alvará. É razoável crer que a data de autuação do processo administrativo para pedido de emissão do alvará tem forte relação com a legislação aplicável naquele momento, porém não há como sabermos de antemão se a legislação que incide no momento é aquela em que o empreendimento relacionado àquele alvará está sendo construído. tal prática se constitui para o que é conhecido como direito de protocolo, situação em que o empreendedor pode optar entre a legislação do momento de aprovação do alvará e a anterior (quando processo foi autuado). Mas também pela relação entre os vários tipos de alvará de licenciamento imobiliário - sendo alguns mais determinantes do que outros para esta incidência - e pela existência de um intervalo de tempo entre a autuação de um alvará e sua aprovação na prefeitura.

Para contornar este problema, a equipe do projeto desenvolveu alguns critérios determinantes da legislação e atribuiu estes critérios ao conjunto tratado de dados - ou seja, não ao nível mais desagregado do alvará, mas ao nível dos lotes/empreendimentos. Em tese, este problema poderia ser resolvido com o estabelecimento do vínculo do conjunto de dados tratado pelo projeto com o conjunto de dados tratados pela PMSP/SMUL, uma vez que os dados tratados que foram disponibilizados na plataforma Monitoramento PDE trazem consigo uma atribuição de legislação aos alvarás. Contudo, há um grande universo de alvarás sem esta atribuição na base tratada pela prefeitura, sendo necessário pensar em critérios adicionais para termos uma visão satisfatória sobre a regulação incidente.

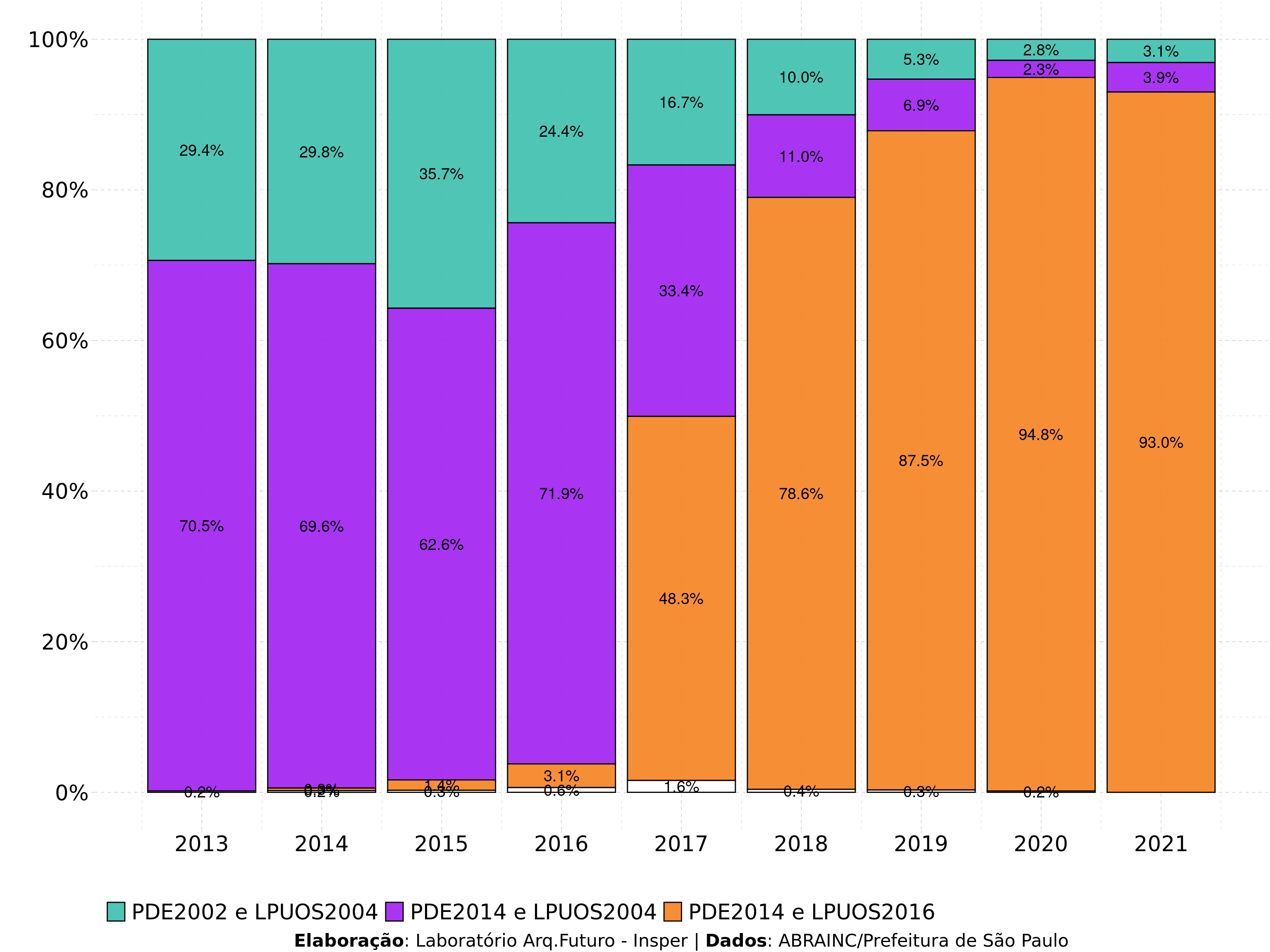
Para os alvarás sem informações, o primeiro critério de classificação utilizou as informações da coluna “zoneamento”. Neste universo, foi inicialmente utilizada a nomenclatura específica a cada Lei de Parcelamento, Uso e Ocupação do Solo. Desta forma, zonas que se claramente se referem à Lei 13.885/04, foram classificadas como tal. O mesmo vale para Lei 16.402/16 (ver listagem de nomes de zonas passíveis de aplicação direta) e PDE 2014 (quando denominado EETU). ​Em seguida, para as zonas que possuem nomenclatura similar nas legislações, foram utilizados critérios cruzados entre nomenclatura e informações da coluna zoneamento antigo3. E por fim, foram analisados caso a caso quando denominado por tipos de ZEIS que não apresentavam clareza dos critérios acima elencados. Para estes, foi consultado o sítio eletrônico “De olho na obra” para verificação do alvará e em outros casos o SIMPROC, para leitura do despacho no diário oficial.

| **Árvore de decisões - Atribuição da legislação** |
| --- |
| Decisão |
| **Critério** |
| Zoneamento |
| ZM-X/00X | ZM-X | ZM-X; OU |
| ZCP-A | ZCP-B |
| ZER-X/XXXX |
| ZEIS-1/XXXX |
| ZEPAM/XXXX |
| ZPI/XXXX |
| ZM-P |
|  |
| ZEIS-5 |
|  |
| ZPR |
| QA |
| ZM | ZMA |
| ZEM | ZEMP |
| ZC-ZEIS |
| ZDE-X |
| ZC | ZCA |
| ZPI-1 | ZPI-2 |
| ZMIS | ZMISA |
| Zoneamento + Zoneamento Anterior + Categoria de Uso |
| Categoria de uso + Zoneamento Anterior |
| LPUOS1972 |
|  |
| Zoneamento + Zoneamento Anterior |
|  |
| ZER |
| Data de autuação + Zoneamento Anterior |
|  |
| LPUOS2004 |
|  |
| LPUOS2016 |

Como parte do tratamento realizado nos dados do SISSEL/SISACOE, identificamos então a legislação incidente em cada alvará a partir de três categorias:

* PDE 2002 & LPUOS 2004
* PDE 2014 & LPUOS 2004
* PDE 2014 & LPUOS 2016

Na primeira categoria encontram-se os empreendimentos protocolados e regulados pela legislação urbana anterior, a última representa os empreendimentos licenciados sob a atual legislação. Já a segunda categoria refere-se a empreendimentos protocolados entre 2014 e 2016 e que seguem regimentos híbridos, com certos regramentos já alterados pelo PDE-2014, mas ainda tendo por base a LPUOS-2004.



Os dados de alvarás de construção desde 2004 estão disponíveis a partir do tratamento realizado. No entanto, nossa análise temporal se restringe ao período mais recente. Optamos por analisar o período entre 2013 e 2021 em função da possibilidade de identificarmos três períodos distintos em termos de legislação urbanística incidente nos alvarás. Em função do direito de protocolo, a regulação urbanística aprovada em 2014 no PDE e complementada em 2016 na LPUOS incide apenas em parte dos alvarás de construção emitidos a partir desta data.

A partir da incidência da regulação urbanística aplicada em cada alvará, conforme ilustrado no gráfico anterior, podemos identificar a formação de padrão de três períodos, que foram utilizados em momento oportuno durante a análise:

* Entre 2013 e 2015, em que a legislação anterior predomina;
* Entre 2016 e 2018 é um período de transição, em que há empreendimentos seguindo o regramento anterior em função do direito de protocolo, empreendimentos seguindo o regramento do híbrido (PDE2014&LPUOS2004), mas também empreendimentos que já seguem o regramento atual;
* Entre 2019 e 2021, em que o regramento atual já predomina, ainda que, de forma residual, haja alvarás que sigam o regramento anterior devido ao direto de protocolo.

Ainda que as discussões públicas para a revisão do PDE tenham se iniciado em abril 2013 e que o primeiro período inclua momentos logo após a aprovação do atual PDE, podemos considerar que o efeito expectativa do novo regramento era restrito, pois era incerto o que seria validado na lei de zoneamento, além do fato de que o ciclo da produção habitacional é longo. Temos que levar em conta que estes períodos também marcam diferentes fases do mercado imobiliário. O período de transição também foi um período de recessão econômica e de retração do mercado por fatores macroeconômicos.

#### 3.4.3.2 Grupos de categoria de uso

A atribuição de grupos/classes de categoria de uso foi amplamente utilizada na pesquisa, a partir de categorias de uso do empreendimento prevista na legislação e registradas no SISACOE. De tal maneira, categorias como HIS, HMP, R2V e R2H foram agrupadas por duas tipologias básicas:

* **Empreendimentos Residenciais de Mercado (ERM)**
* Voltado à população em geral que tenha condições de arcar com o preço de mercado de uma moradia
* Preço e condições de financiamento definidas pelo mercado
* Características construtivas e custo da unidade habitacional definidos de acordo com o mercado-consumidor alvo
* Qualquer família pode adquirir um imóvel desde que tenha condições econômicas para sua aquisição
* **Empreendimento Residencial Popular (ERP)**
* Destinada ao atendimento habitacional das famílias de baixa renda, podendo ser de promoção pública ou privada
* Preço e condições de financiamento reguladas pelo governo
* Características construtivas e incentivos do poder públicos para tornar seu custo menor
* A categoria ERP engloba as subcategorias Habitação de Interesse Social (HIS), para famílias que recebem até seis salários mínimos e Habitação de Mercado Popular (HMP), para famílias que recebem entre seis e dez salários mínimos mensais.

É importante ressaltar que poderia haver várias categorias de uso atribuídas ao mesmo lote/empreendimento, uma vez que alvarás individualizados podem se referir a usos diferentes do terreno e/ou compor um tipo de uso de diferentes padrões de produto imobiliário dentro do mesmo lote. Para alguns casos em que havia um grau de pertinência do lote/empreendimento tanto ao grupo ERP quanto ao grupo ERM tomou-se a decisão de definir como prevalecente o grupo ERP. Além disso, em momento julgado oportuno, foram usadas subcategorias, com destaque particular ao desmembramento da categoria ERP para análise de HIS, HMP ou HIS junto de HMP.

#### 3.4.3.3 Grupos de zoneamento

A definição de zoneamento utilizada na pesquisa é atribuída diretamente ao lote pelo georreferenciamento, tratando-se de tomar como referência não o zoneamento do momento do registro, mas o zoneamento atribuído ao lote do empreendimento na legislação atual. Para tornar este atributo mais apto à produção de informações relevantes, utilizou-se também uma agregação desses zoneamentos em grupos de zoneamento desenvolvida a partir de análise qualitativa e resumida no quadro abaixo:



### 3.4.4 Limitações do processo de ETL

Espera-se publicizar todos os detalhes do tratamento após a conclusão da pesquisa, o que exige, além da simples disponibilização de scripts e bases de dados em repositórios públicos, algumas revisões e aportes ao processo de tratamento.

Implementado o tratamento dos dados de licenciamento imobiliário, foram listadas pela equipe do Projeto Acesso a Oportunidades alguns objetivos estratégicos no que tange ao desenvolvimento do processo de Extração, Transformação e Carregamento dos dados - ETL:

* Compreender as razões para taxas de não-resposta em determinados atributos (sobretudo, total de unidades/ média de unidades por pavimento) e delinear estratégias de minimização de impactos em análises sobre os dados;
* Avaliar a acurácia do georreferenciamento dos empreendimentos e pensar formas de aprimoramento;
* Publicizar os scripts de tratamento em repositório público via GitHub;
* Disponibilizar as bases tratadas com ferramentas de fácil acesso ao usuário via Portal de Dados Urbanos do Insper (projeto complementar, ainda em elaboração);
* Elaborar recomendações para o desenvolvimento de uma tecnologia de monitoramento e avaliação dos licenciamentos em São Paulo, maximizando o aproveitamento de resíduos informacionais decorrentes da prestação de serviços públicos e o acompanhamento de dados cadastrais.

A próxima sessão traz considerações mais detalhadas a respeito do último objetivo listado.

# 4 Pontos para o desenvolvimento de Monitoramento e Avaliação

Trazemos, por fim, uma série recomendações para o desenvolvimento de uma tecnologia de monitoramento e avaliação dos licenciamentos imobiliários em São Paulo, maximizando o aproveitamento de resíduos informacionais decorrentes da prestação de serviços públicos e aprimorando o acompanhamento de dados cadastrais.

* Abrir os dados de alvarás de licenciamento imobiliário foi um grande avanço da PMSP/SMUL que merece ser celebrado como tal. Porém, é necessário dizer que uma abertura sem explicações técnicas detalhadas para as partes interessadas e sem a devida contextualização desses dados em informações geoespaciais complementares cria barreiras para se extrair informação de qualidade dos dados. Deixamos, então, como sugestão à PMSP/SMUL a incorporação do processo de ETL desenvolvido em código aberto pela Equipe do Projeto Acesso a Oportunidades no PDE como forma de explorar visões de quais seriam os passos de tratamento e contextualização necessários para a elaboração de análises, monitoramentos e avaliações de impacto a partir dos dados de alvarás de licenciamento imobiliário;
* Para fins de tornar os processos de licenciamento imobiliário em São Paulo mais próximos aos princípios de Governo Aberto, seria também interessante, a nosso ver, que a PMSP/SMUL delimitasse a estratégia de abertura dos dados na forma de um [Plano de Dados Abertos - PDA](https://www.gov.br/governodigital/pt-br/sisp/guia-do-gestor/pda#:~:text=O%20que%20%C3%A9%3F,dados%20abertos%20nas%20organiza%C3%A7%C3%B5es%20p%C3%BAblicas.) de caráter público e participativo. O uso deste instrumento, que é amplamente reconhecido como uma boa prática de Governo Aberto e controle social, representaria grande avanço em termos de desenvolvimento institucional em relação ao atual contexto de publicização dos dados de alvarás de licenciamento e promoveria melhores condições para que esses dados sejam parte de um real instrumento de gestão urbana.
* Diante da nossa experiência com o consumo dos dados disponibilizados pela PMSP/SMUL, nos parece interessante a materialização de um cenário em que o PDA estabeleceria um marco de mudança no versionamento do conjunto de dados até então disponibilizado, pensando duas linhas de ação estratégica complementares: a) o aprimoramento da abertura de dados estabelecida até o presente momento, incorporando aprendizados da Equipe do Projeto Acesso a Oportunidades no PDE e outras iniciativas de desenvolvimento de projetos acadêmicos e/ou práticos com os dados da PMSP/SMUL; b) pensar diretrizes de uma nova versão da base de dados para o futuro em médio e longo prazo, tendo em perspectiva outros critérios de coleta, processamento e comunicação que pudessem contornar falhas já conhecidas no conjunto de dados;
* A partir do PDA, seria fundamental pensar em estratégias de engajamento de partes interessadas na sociedade de civil na elaboração e na implementação do plano. Este é um passo essencial para que a abertura de dados atinja melhores critérios de transparência e para a facilitação a inovação social relacionadas à discussão dos licenciamentos imobiliários - coisas que tem como beneficiário, ao fim e ao cabo, o próprio poder público. Para tanto, devem ser prioridade levar o PDA de Licenciamentos Imobiliários tanto para canais institucionais mais formais - tais como conselhos e conferências - quanto para espaços de menor formalidade e burocratização - tal como, por exemplo, as [iniciativas de Café Hacker desenvolvidas pela PMSP/CGM](http://wiki.govit.prefeitura.sp.gov.br/index.php?title=Caf%C3%A9_Hacker);
* Como parte de qualquer tecnologia de ciência de dados e inteligência de negócio, a condição básica de se desenvolver procedimentos que fazem com que os ciclos de produção e consumo de dados possam solucionar problemas reais é o entendimento do "negócio" - ou seja, compreender a fundo quais são os problemas sociais que motivam a existência destes dados e quais são os mecanismos associados a esses problemas. Portanto, no caso dos dados que discutimos aqui, é necessário que as partes interessadas estudem cuidadosamente quais são os desafios urbanos, políticos e econômicos que precisam ser respondidos pelo aprimoramento tecnológico do processo de licenciamento imobiliário de São Paulo. Entende-se que seria interessante que o processo de discussão de um PDA para os alvarás de licenciamento imobiliário fosse acompanhado de discussões sobre o processo de licenciamento como problema urbano em si, em encontros mais ou menos formais de caráter amplo e orientados a produção de resultados de desenvolvimento institucional sistema atualmente implementado.
* Além de um alinhamento à discussão geral pautado pela política de licenciamento imobiliário de São Paulo, seria interessante que o debate sobre o ciclo de produção e consumo dos dados de alvarás no PDA fosse alinhado à discussão geral sobre política urbana. Se alinhando, aos debates mais gerais sobre o atual PDE e as diretrizes de desenvolvimento urbano de São Paulo e também buscasse se alinhar a discussões de caráter nacional sobre urbanismo e licenciamento imobiliário. Cabendo destaque particular às discussões que se iniciaram em torno do [programa "Construa Brasil"](https://www.gov.br/produtividade-e-comercio-exterior/pt-br/ambiente-de-negocios/competitividade-industrial/construa-brasil) e do "[Guia Orientativo de Boas Práticas para Obtenção de Alvarás de Construção](https://www.gov.br/produtividade-e-comercio-exterior/pt-br/ambiente-de-negocios/competitividade-industrial/construa-brasil/produtos/GuiaOrientativodeBoasPrticasparaObtenodeAlvarsdeConstruocompressed.pdf)" promovidos pelo Governo Federal, que dão maior centralidade ao processo de licenciamento como um problema urbano nacional e sugerem metas para o desenvolvimento desta atividade em particular. Além de se atentar aos novos contornos que esta e outras diretrizes nacionais de política urbanas devem ganhar no novo ciclo de governo.
* É também essencial mapear e promover fluxos de integração do PDA de Licenciamentos Imobiliários com iniciativas que já estão em curso. Priorizar a integração com o que já existe - fora e principalmente dentro da própria prefeitura - é etapa primordial para que o planejamento alcance condições mais realistas de implementação e também para a promoção da inovação social;
* De acordo com o [catálogo de dados da PMPSP](https://www.prefeitura.sp.gov.br/cidade/secretarias/controladoria_geral/coordenadoria_de_promocao_da_integridade/index.php?p=225080), existem quatro bases de dados mantidas pela PMSP/SMUL diretamente associadas ao processo de licenciamento imobiliário: SLCe, Sisacoe, Portal do Licenciamento - Aprova Digital, Portal de Licenciamento - Anistia Imobiliário. Portanto, a nosso ver, fica subentendido como um passo número zero da integração de dados de licenciamento imobiliário a integração de processos dentro da própria PMSP/SMUL, unificando as bases de dados referentes a uma finalidade fortemente similar;
* Dentro do escopo da integração de processos dentro da prefeitura, nos parece uma meta interessante para o PDA a de se ter o processo de abertura dos dados do SISACOE/SLCe se integrando à [vitrine de API's da PMSP/SMIT](https://apilib.prefeitura.sp.gov.br/store/) e, por consequência, avançando ao melhor estágio no que tange ao [percurso das 5 estrelas dos dados abertos](https://5stardata.info/pt-BR/). Com os dados disponibilizados em formato de API o processo de inovação social relacionado ao dado é fortemente facilitado, dando condições mais simples para o desenvolvimento de aplicativos, sistemas e estudos acadêmicos - a exemplo das várias iniciativas que se sucederam após à [abertura de dados da SPTRans, caso que pode ser tomado pela PMSP/SMUL](https://g1.globo.com/sao-paulo/noticia/2013/10/sptrans-abre-dados-do-transporte-para-desenvolvimento-de-aplicativos.html) como uma importante referência de boa prática;
* A integração com dados de política urbana produzidos na SMUL e em outras instâncias de governo também deveria ser tomada como uma atividade prioritária. A agregação de dados espaciais da Plataforma GeoSampa amplia fortemente a contextualização dos licenciamentos na cidade, como bem sugere o estudo de caso promovido pela equipe do Projeto Acesso a Oportunidades no PDE e publicado nas notas técnicas apresentadas ao Fórum SP23. Outras bases de dados também poderiam ser agregadas trazendo várias outras hipóteses de pesquisa e desenvolvimento, como dados do IPTU, de ITBI, de Lançamentos Imobiliários, etc. Como uma meta mais concreta ao PDA, poderia ser pensada uma forma de integrar a base de alvarás à base de Lotes da Prefeitura - incluindo memória de desmembramentos de lotes - o que facilitaria amplamente a comparação dos licenciamentos imobiliários aos demais dados cadastrais de imóveis em São Paulo.

1. Tendo em vista que mais de um alvará execução de edificação nova pode estar atribuído ao mesmo lote/empreendimento, em razão de especifidades do processo de licenciamento. [↑](#footnote-ref-44)