**UNIVERSIDADE VIRTUAL DO ESTADO DE SÃO PAULO**

Flávio Ribeiro Máximo

Gabriella Wandeur de Freitas

Grazielle Aparecida dos Santos

Leiza Santos Brito Menezes

Mileide Maria da Silva Barbosa

Miriam Rose de Castro Meneses

Rafael Paoleschi Iurovschi

Sandra Camargo

**AQ-CADE Arujá: Manual e Análise de Dados no sistema CartCID**

Arujá - SP

2025

**SUMÁRIO**

[1 INTRODUÇÃO 3](#_Toc1)

[2 GERAR RELATÓRIOS 4](#_Toc2)

[2.1 Acesse o Google Colab 4](#_Toc3)

[2.2 Importe o arquivo carteirinha.csv 4](#_Toc4)

[2.3 Execute as células do notebook 5](#_Toc5)

[2.4 Valide as análises 5](#_Toc6)

[2.3 Exportar o arquivo 5](#_Toc7)

[3 ANÁLISE EXEMPLO 6](#_Toc8)

[3.1 Análise de Idade 6](#_Toc9)

[3.2 Análise de Gênero 8](#_Toc10)

[3.3 Análise por tipo Sanguíneo 10](#_Toc11)

[3.4 Análise por Bairro 12](#_Toc12)

## 1 **INTRODU**ÇÃO

Este manual tem como objetivo orientar os analistas de dados da Prefeitura da Cidade de Arujá na realização de análises com base nos dados do sistema CartCID. O documento apresenta um passo a passo para importar os dados extraídos do sistema e executar os scripts desenvolvidos para gerar visualizações e interpretações úteis para campanhas de acessibilidade e inclusão social.

As instruções aqui contidas foram elaboradas a partir de um ambiente de testes, utilizando dados simulados e anonimizados,sem relação com pessoas reais. Essa abordagem garante conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018), ao mesmo tempo em que demonstra o potencial analítico do sistema. O manual inclui também detalhes sobre os ambientes necessários, formas de execução local ou em nuvem (Google Colab), e a estrutura básica dos arquivos utilizados na análise.

O documento também apresenta demonstrações práticas de exploração de dados, com gráficos gerados utilizando a linguagem de programação Python. As análises foram desenvolvidas no ambiente Jupyter Notebook e exemplificam como as informações podem ser segmentadas por critérios como bairro, gênero e faixa etária. Esses recursos visuais facilitam a interpretação dos dados e apoiam a criação de campanhas públicas mais eficazes, especialmente voltadas à acessibilidade e inclusão social.

## 2 GERAR RELATÓRIOS

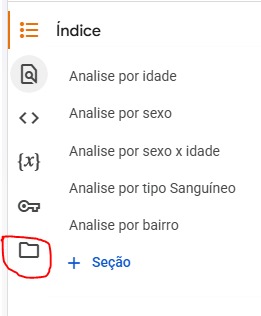
#### 2.1 Acesse o Google Colab

Vá até [Google Colab](https://colab.research.google.com/drive/1NvHKMCYp-iY_6V5cjC5VJrGite1oQpX3?usp=sharing) e faça login com sua conta Google, se necessário.

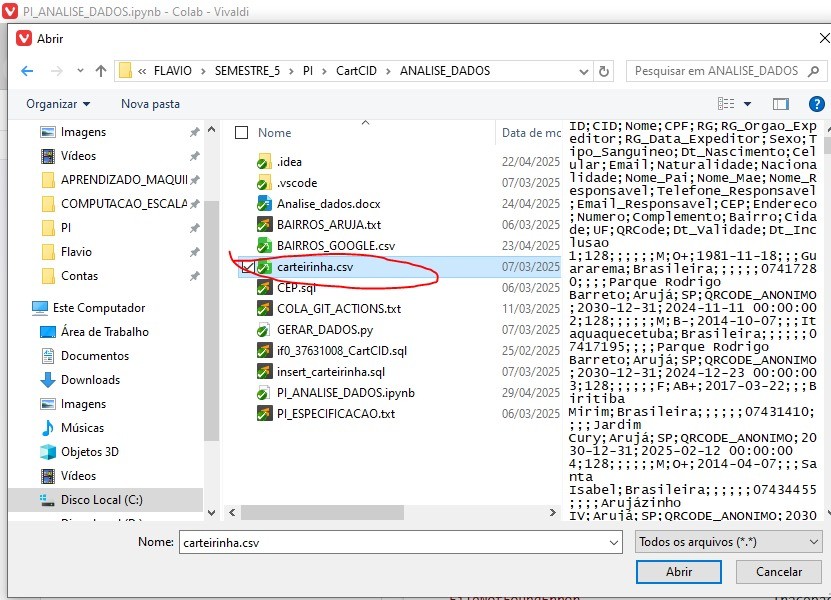
#### 2.2 Importe o arquivo carteirinha.csv

Para nosso exemplo estamos utilizando uma massa de dados para teste, mas o analista da prefeitura de arujá deverá utilizar os dados importados do sistema.   
É importante realizar a extração e renomea-la para carteirinha.csv

No menu lateral esquerdo, clique no ícone de pasta 🗂 para abrir o navegador de arquivos. Clique em **"Upload"**



Selecion e envie o arquivo **carteirinha.csv**, que deve conter os dados exportados do sistema.

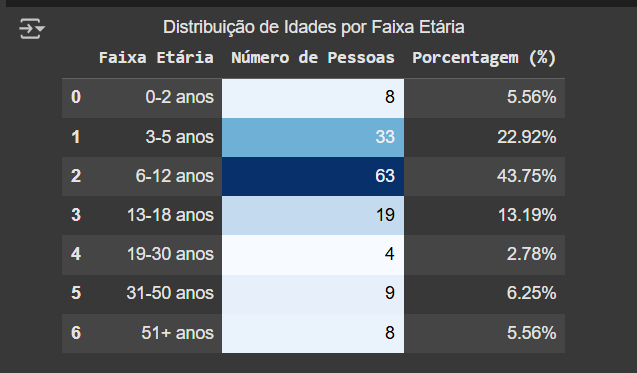


#### **2.3 Execute as células do notebook**

Após abrir o notebook, clique em **"Executar tudo"** (ícone ▶️ no topo da tela ou Runtime > Run all). Isso rodará todas as análises e gráficos automaticamente com base no arquivo de dados importado.

#### **2.4 Valide as análises**

O notebook exibirá gráficos e tabelas com base nas colunas *idade, gênero, bairro e Tipo sanguíneo.* Gerando insights que podem ser usados para decisões públicas e campanhas de acessibilidade.



#### **2.3 Exportar o arquivo**

Se quiser salvar uma cópia em PDF ou HTML:

1. Vá em **"Arquivo" > "Fazer download"** e selecione o formato desejado.
2. Ou use a opção de imprimir o notebook como PDF com Ctrl + P.

## 3 **ANÁLISE EXEMPLO**

As análises aqui exemplificadas foram elaboradas a partir de um ambiente de testes, utilizando dados simulados e anonimizados, sem relação com pessoas reais. Essa abordagem garante conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018), ao mesmo tempo em que demonstra o potencial analítico do sistema.

Com o objetivo de ilustrar como os dados do sistema CartCID podem ser utilizados de forma estratégica, foram realizadas análises exploratórias com base em dados simulados. As informações selecionadas para análise foram: faixa etária, gênero, tipo sanguíneo e bairro de residência.

As variáveis utilizadas para as análises, como faixa etária, gênero, tipo sanguíneo e bairro de residência, foram fornecidas pelos próprios colaboradores da prefeitura durante uma pesquisa realizada. Ao utilizar as variáveis sugeridas pelos colaboradores, buscamos garantir que as análises refletissem as necessidades reais da população, assegurando que as decisões tomadas fossem baseadas em dados pertinentes e relevantes para o público-alvo.

#### 3.1 Análise de Idade

Analisando os dados pelo ponto de vista da idade nota-se a maior concentração de pessoas está na faixa de 6 a 12 anos, totalizando 63 indivíduos, o que representa 43,75% da amostra. As faixas etárias de 0 a 2 anos e 3 a 5 anos somam 9 e 32 indivíduos, respectivamente, o que corresponde a 6,25% e 22,22%. Somadas, as faixas etárias até 12 anos representam 72,22% do total, indicando que a amostra é majoritariamente composta por jovens. As demais faixas apresentam uma frequência significativamente menor, com destaque para a faixa de 13 a 18 anos, que representa 13,19% da amostra. Faixas adultas, como 19 a 30 anos (2,78%), 31 a 50 anos (6,25%) e 51 anos ou mais (5,56%), compõem juntas apenas 14,59% do total.

Essa distribuição reforça o caráter infantil juvenil da população estudada, o que pode estar diretamente relacionado a melhorias aplicadas na metodologia de diagnóstico.

Foram imputados 144 registros onde:

* A **idade média** é de aproximadamente **12,13 anos**.
* O **desvio padrão** é 13,26, o que indica que há bastante variação nas idades.
* A **idade mínima** é 0 (o usuário não completou 1 ano ainda).
* A maioria das idades está entre **4 e 13 anos**, pois esses são os quartis.
* A **mediana** (valor central) é 8 anos, ou seja, metade dos registros tem idade menor que isso.
* O valor **máximo** registrado é 58 anos.

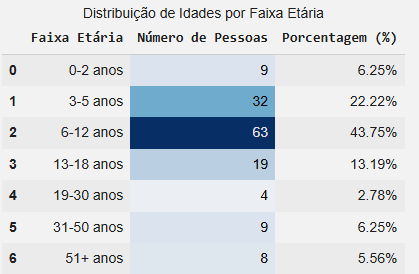


Tabela 1 – Distribuição de Idades por Faixa Etária

Gráfico 1 – Distribuição de Idades por Faixa Etária

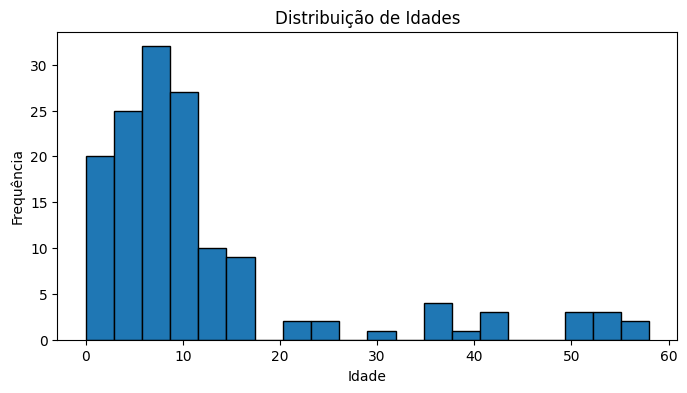
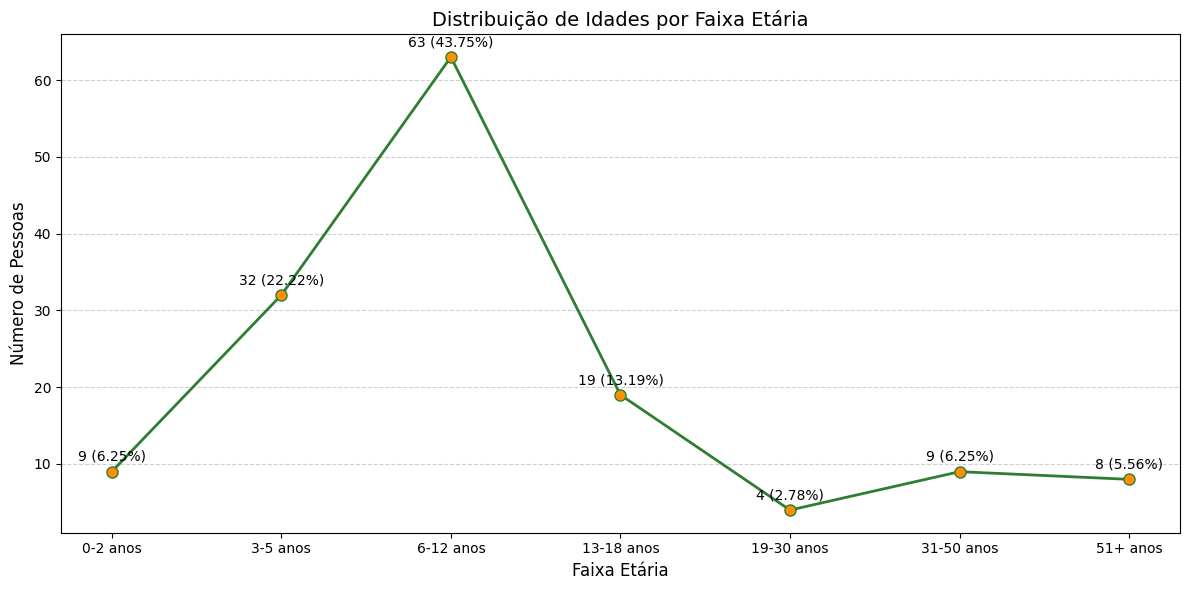


Gráfico 2 – Distribuição de Idades por Faixa Etária

#### 3.2 Análise de Gênero

Analisando os dados pelo ponto de vista do gênero (Sexo) nota-se que a uma ligeira predominância do sexo feminino (52,08%), com 75 registros, contra 47,92% do sexo masculino (69 registros). A diferença absoluta é de 6 pessoas entre os grupos, indicando uma distribuição equilibrada, sem viés significativo em relação ao gênero. Essa proporção (~52% mulheres e ~48% homens) está alinhada com a distribuição demográfica brasileira, onde as mulheres representam cerca de 51,1% da população.

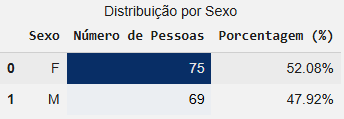


Tabela 2 – Distribuição por sexo.

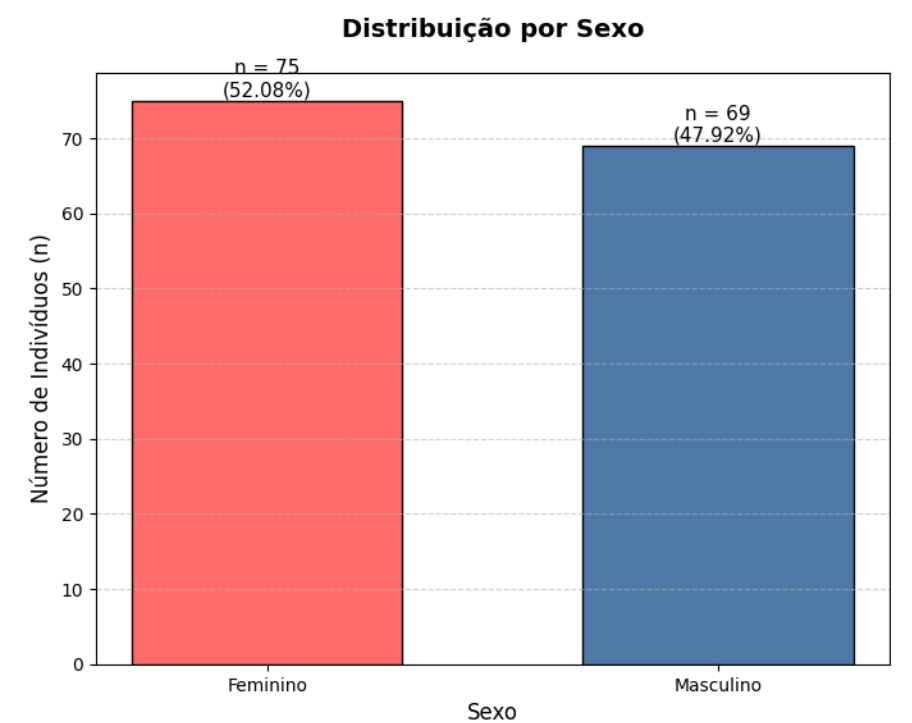


Gráfico 3 – Distribuição por sexo

**Análise Descritiva:**

1. **Crianças (0-12 anos)**:
   * Representam **55.56% da amostra** (80/144), com leve predominância feminina (52.5%).
2. **Adolescentes (13-18 anos)**:
   * **22.92% da amostra**, mantendo proporção similar entre sexos (~54% F vs ~45% M).
3. **Adultos (31+ anos)**:
   * A partir dos 31 anos, há inversão: homens passam a ser maioria (54.55% M na faixa 31-50 anos, 60% M em 51+).
4. **Consistência geral**:
   * A proporção total (52.08% F / 47.92% M) mantém-se estável nas faixas mais jovens, com variações significativas apenas em grupos menores (idosos).

Fonte: "Dados processados pelo autor (2025)".



Tabela 1 – Distribuição por sexo e Faixa Etária .

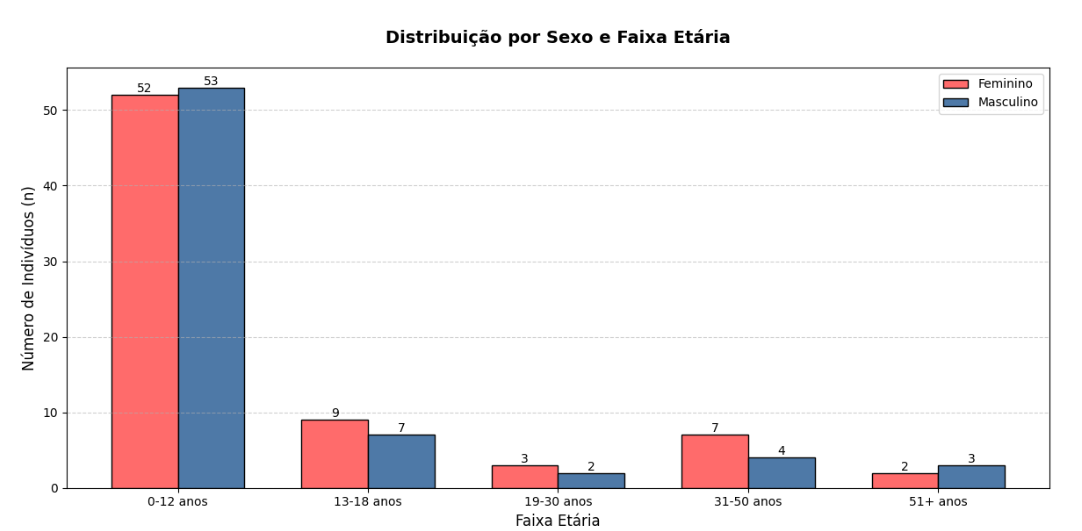


Gráfico 1 – Distribuição por sexo e Faixa Etária .

#### 3.3 Análise por tipo Sanguíneo

A distribuição de tipos sanguíneos em uma população é um tema de grande relevância para a saúde pública, planejamento médico e gestão de recursos em emergências. Conhecer a prevalência de cada tipo sanguíneo (A, B, AB e O, combinados com os fatores Rh+ e Rh-) permite otimizar estoques em bancos de sangue, orientar campanhas de doação e antecipar desafios em procedimentos críticos, como transfusões e transplantes. Em contextos locais, como o município de **Arujá/SP**, essa análise ganha ainda mais importância, pois reflete particularidades demográficas e necessidades específicas da comunidade.

No Brasil, segundo a Fundação Pró-Sangue, a distribuição aproximada é:

|  |  |
| --- | --- |
| Tipo sanguíneo | Porcentagem % |
| **O+** | 36 |
| A+ | 34 |
| B+ | 9 |
| AB+ | 3 |
| O- | 6 |
| A- | 2 |
| B- | 2 |
| AB- | 1 |

Esses dados nacionais, porém, podem variar regionalmente devido a fatores genéticos, migratórios e étnicos. Por isso, análises locais, como a realizada neste estudo, são fundamentais para embasar políticas públicas precisas.

Neste trabalho, utilizamos uma base de dados de **144 registros anonimizados** de carteirinhas de Arujá/SP para explorar:

1. **Prevalência dos tipos sanguíneos** na amostra.
2. **Comparação com padrões nacionais e internacionais**.
3. **Implicações práticas** para gestão de bancos de sangue e saúde coletiva.

Os resultados revelaram, por exemplo, que **O+** é o tipo mais frequente (24,31%), seguido por **A+** (17,36%), enquanto **AB-** e **AB+** são os mais raros (6,94% e 5,56%, respectivamente). Essa distribuição reforça a necessidade de estratégias direcionadas para garantir estoques adequados de tipos raros, além de campanhas educativas para sensibilizar doadores.

Além disso, a análise por tipo sanguíneo pode ser cruzada com outras variáveis (como idade, sexo e bairro) para identificar padrões ocultos. Por exemplo:

* **Crianças** tendem a ter maior prevalência de tipos O+ e A+?
* Há **bairros** com concentração incomum de tipos raros?

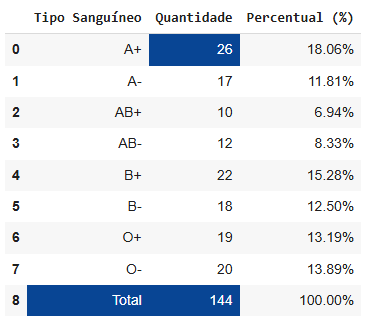


Tabela 4 – Distribuição por tipo Sanguíneo

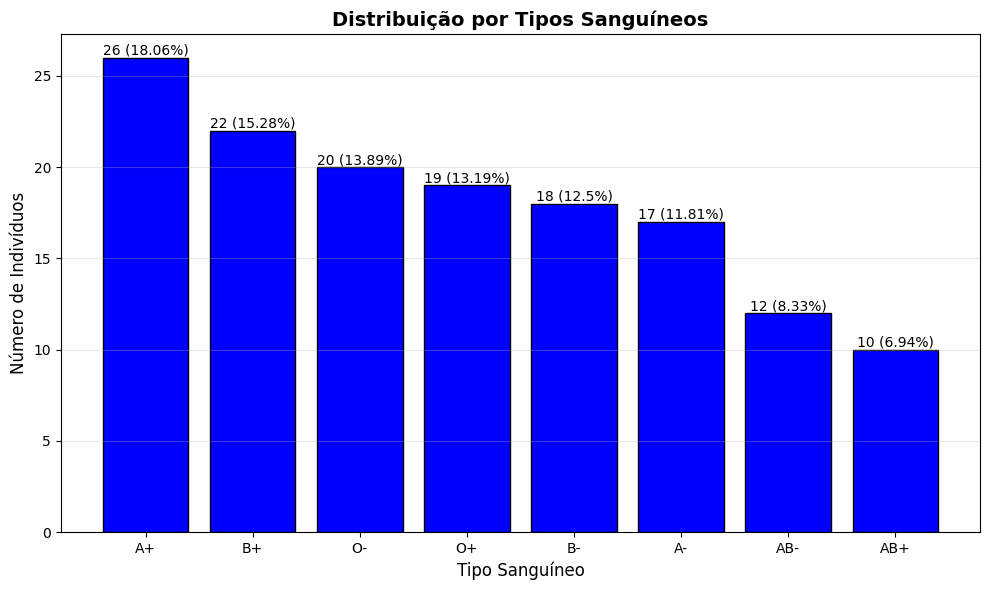


Gráfico 5—Distribuição por tipo Sanguíneo

#### 3.4 Análise por Bairro

A distribuição geográfica de dados populacionais é um instrumento fundamental para a gestão pública, permitindo identificar desigualdades, alocar recursos de forma eficiente e direcionar políticas sociais com base em evidências. No contexto deste estudo, a análise por bairros revela padrões espaciais críticos que podem refletir desde desigualdades no acesso a serviços até necessidades específicas de comunidades locais.

Os dados coletados demonstram uma distribuição heterogênea da amostra entre os bairros do município, com concentrações significativas em regiões como o Parque Rodrigo Barreto (17,36% dos registros) e o Jardim Fazenda Rincão (12,5%), enquanto outras áreas aparecem com representação mínima. Essa disparidade pode ser influenciada por fatores como:

* Acesso a serviços públicos (ex.: proximidade de postos de saúde, escolas ou centros de cadastro);
* Densidade populacional (bairros mais habitados tendem a aparecer com maior frequência);
* Barreiras socioeconômicas (populações em áreas periféricas podem ter menor acesso a registros formais).

Além disso, a análise por bairros é essencial para:

* Priorizar investimentos em infraestrutura, saúde e educação onde há maior demanda;
* Identificar lacunas na cobertura de serviços públicos;
* Promover equidade, garantindo que políticas públicas atendam tanto a áreas centrais quanto a regiões menos assistidas.

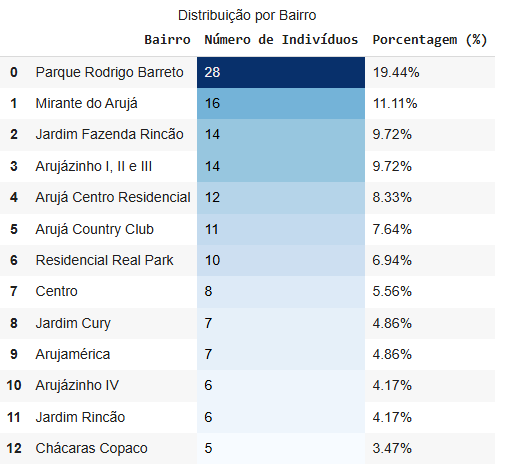


Tabela 5 – Distribuição por Bairro

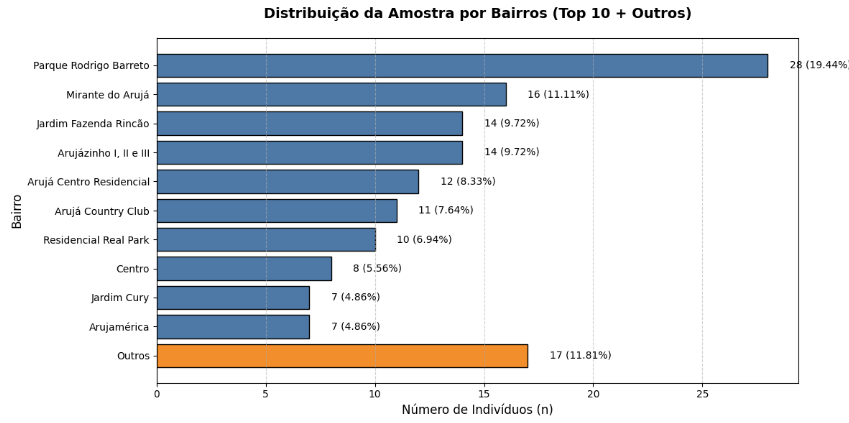


Gráfico 6 – Distribuição da Amostra por Bairros (Top 10 + Outros)

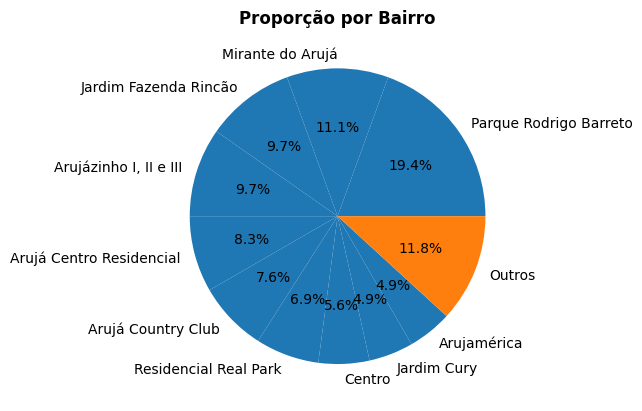
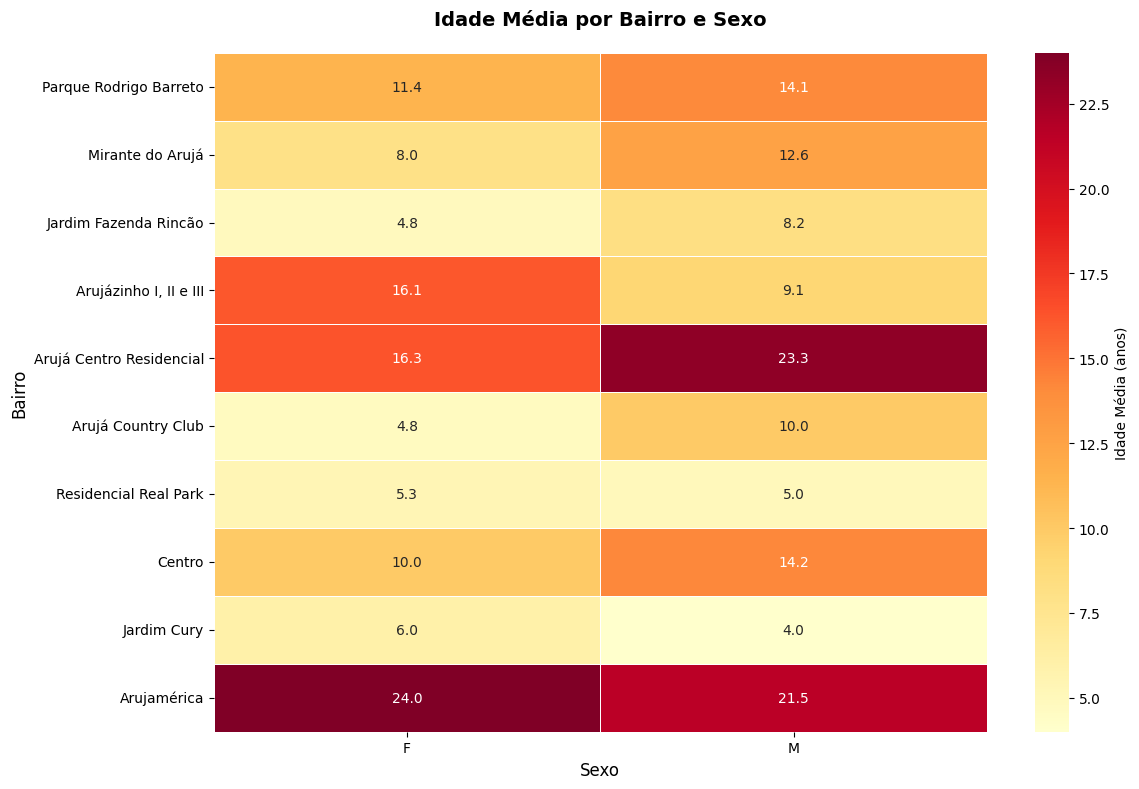


Gráfico 7 – Proporção por bairro

Gráfico 8 – Idade Média por Bairro e Sexo



3.4.1 Mapa georreferenciado  
 O mapa georreferenciado (<https://www.google.com/maps/d/viewer?mid=16e9rmzIG-UX9QjpSrtZQ4e0OzMVU4JU&ll=-23.397215394932733%2C-46.300880564263906&z=15>) revela a distribuição territorial dos indivíduos cadastrados no município de Arujá/SP, permitindo identificar clusters de concentração e áreas sub-representadas. Os dados espaciais destacam três padrões críticos:

1. Concentração em Áreas Centrais

Regiões como o Centro e o Parque Rodrigo Barreto apresentam maior densidade de registros, refletindo:

* Facilidade de acesso a serviços públicos (ex.: posto de saúde onde os dados foram coletados).
* Maior densidade populacional típica de zonas urbanizadas.

1. Dispersão em Periferias

Bairros como Chácaras Copaco e Arujazinho (pontos verdes isolados) têm poucos registros, sugerindo:

* Barreiras geográficas (distância até locais de cadastro).
* Carência de infraestrutura (ex.: transporte público para chegar aos pontos de atendimento).

1. Corredores de Exclusão

Áreas ao longo da Rodovia Presidente Dutra e na divisa leste do município aparecem com baixa representação, possivelmente indicando:

* Desigualdade na oferta de serviços entre regiões.
* Populações invisibilizadas (ex.: assentamentos informais não mapeados).