

ESTEVAN RAFAEL TOMAZINI

OFICINA MAKER: Um projeto de Data Science com o dataset “Sistema E-saúde - Perfil de atendimento médico nas Unidades municipais de saúde de Curitiba”

CURITIBA

2023

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
DEPARTAMENTO DE TECNOLOGIA DA INFORMAÇÃO  
BIGDATA E INTELIGÊNCIA ANALÍTICA  
ESTEVAN RAFAEL TOMAZINI

OFICINA MAKER: Um projeto de Data Science com o dataset “*Sistema E-saúde - Perfil de atendimento médico nas Unidades municipais de saúde de Curitiba*”

Atividade somativa apresentada ao professor-tutor Galbas Milleo Filho como requisito para aprovação na disciplina oficina maker (11100010573\_20232\_01) do curso de bigdata e inteligência analítica da Pontifícia Universidade Católica do Paraná.

CURITIBA

2023

## 1. INTRODUÇÃO

Na atividade somativa 1 dessa disciplina, foi introduzido alguns conceitos do SUS, dos dados escolhidos para análise, quais seriam as atividades realizadas e o cronograma de desenvolvimento. Nessa entrega final, serão abordadas algumas modificações de cronograma e adição da ferramenta PowerBI, uma vez que Inteligência de mercado faz parte do currículo e o programa permite ETF e análise de dados em grandes volumes de forma satisfatória, muitas vezes considerado a melhor ferramenta para esse fim, permitindo validar a análise exploratória de dados realizados via a Stack pynb/pandas/matplotlib.

Além disso, serão mostradas as análises de *machine learning* e *bigdata*, como descrito no cronograma anteriormente.

Novamente, lembrando, temos um arquivo CSV contendo 1175733 linhas e 42 colunas. Nesses dados, os dados N/A e/ou NaN são praticamente nulos, representando menos de 10% dos valores. Devido à natureza desses dados serem ordinais, a imputação utiliza a moda, para não alterar a estrutura da coluna. Ao final da etapa de ETF restaram 27 colunas, interessante frisar que os dados faltantes desses dados estavam concentrados nessas colunas retiradas.

Ao final, o dicionário de dados ficou como mostrado a seguir, onde os dados grifados em vermelho foram retirados.

1. **Data do Atendimento:** Data e hora do atendimento médico. DD/MM/YYYY HH:MM. Dado quantitativo contínuo.
2. **Data de Nascimento:** Data de nascimento do paciente. DD/MM/YYYY HH:MM. Dado quantitativo contínuo.
3. **Sexo:** Gênero do paciente (M = Masculino, F = Feminino). Dado qualitativo nominal.
4. **Código do Tipo de Unidade:** Código que representa o tipo de unidade de saúde. (1, 2, 3). Dado qualitativo ordinal.
5. **Tipo de Unidade:** Tipo da unidade de saúde (BASICO, UPA). Dado qualitativo nominal.

6. **Código da Unidade:** Código único da unidade de saúde. Dado qualitativo ordinal.
7. **Descrição da Unidade:** Descrição da unidade de saúde. Dado qualitativo nominal.
8. **Código do Procedimento:** Código do procedimento médico realizado. Dado qualitativo ordinal.
9. **Descrição do Procedimento:** Descrição do procedimento médico. Dado qualitativo nominal.
10. **Código do CBO:** Código de Classificação Brasileira de Ocupações do profissional de saúde. Dado qualitativo ordinal.
11. **Descrição do CBO:** Descrição da ocupação do profissional de saúde. Dado qualitativo ordinal.
12. **Código do CID:** Código Internacional de Doenças para diagnósticos médicos. Dado qualitativo ordinal.
13. **Descrição do CID:** Descrição do diagnóstico médico. Dado qualitativo ordinal.
14. **Solicitação de Exames:** Indicação se exames foram solicitados (Sim/Não). Dado qualitativo nominal.
15. **Quantidade Prescrita Farmácia Curitiba:** Quantidade de medicamentos prescritos pela Farmácia Curitiba. Dado quantitativo discreto.
16. **Quantidade Dispensada Farmácia Curitiba:** Quantidade de medicamentos dispensados pela Farmácia Curitiba. Dado quantitativo discreto.
17. **Quantidade de Medicamento Não Padronizado:** Quantidade de medicamentos não padronizados. Dado quantitativo discreto.
18. **Encaminhamento para Atendimento Especialista:** Indicação se houve encaminhamento para especialista (Sim/Não). Dado qualitativo nominal.
19. **Área de Atuação:** Área de atuação do profissional de saúde. Dado qualitativo ordinal.
20. **Desencadeou Internamento:** Indicação se o caso desencadeou internamento (Sim/Não). Dado qualitativo nominal.
21. **Data do Internamento:** Data do internamento, se aplicável. Dado quantitativo contínuo.

22. **Estabelecimento Solicitante:** Estabelecimento de saúde que solicitou o atendimento. Dado qualitativo ordinal.
23. **Estabelecimento Destino:** Estabelecimento de saúde para o qual o paciente foi encaminhado. Dado qualitativo ordinal.
24. **CID do Internamento:** Código Internacional de Doenças para diagnósticos médicos relacionados ao internamento. Dado qualitativo ordinal.
25. **Tratamento no Domicílio:** Indicação se o tratamento foi realizado no domicílio (Sim/Não). Dado qualitativo nominal.
26. **Abastecimento:** Tipo de abastecimento de água na habitação do paciente. Dado qualitativo nominal.
27. **Energia Elétrica:** Disponibilidade de energia elétrica na habitação do paciente. Dado qualitativo nominal.
28. **Tipo de Habitação:** Tipo de habitação do paciente. Dado qualitativo nominal.
29. **Destino Lixo:** Método de descarte de lixo do paciente. Dado qualitativo nominal.
30. **Fezes/Urina:** Método de descarte de fezes e urina do paciente. Dado qualitativo nominal.
31. **Cômodos:** Número de cômodos na habitação do paciente. Dado quantitativo discreto.
32. **Em Caso de Doença:** Ação tomada em caso de doença. Dado qualitativo nominal.
33. **Grupo Comunitário:** Participação em grupo comunitário (Sim/Não). Dado qualitativo ordinal.
34. **Meio de Comunicação:** Meio de comunicação utilizado pelo paciente. Dado qualitativo nominal.
35. **Meio de Transporte:** Meio de transporte utilizado pelo paciente. Dado qualitativo nominal.
36. **Município:** Município de residência do paciente. Dado qualitativo nominal.
37. **Bairro:** Bairro de residência do paciente. Dado qualitativo nominal.
38. **Nacionalidade:** Nacionalidade do paciente. Dado qualitativo nominal.

39. **Código de usuário (cod\_usuario):** Código único do usuário. Dado qualitativo ordinal.

40. **Origem do usuário (ori\_usuario):** Origem do usuário. Dado qualitativo ordinal.

41. **Residente:** Indicação de se o paciente é residente no local (1 para Sim, 0 para Não). Dado qualitativo nominal.

42. **Código do profissional (cod\_profissional):** Código único do profissional de saúde. Dado qualitativo ordinal.

## 1.1. TÉCNOLOGIAS

Nessa sessão serão abordadas as tecnologias aplicadas nesse projeto.

### 1.1.1. Apache Spark

Devido à entrega prévia, essa sessão será resumida para não gerar auto plágio. Apache Spark é um framework multi linguagem e multi propósito na esfera dos dados. Embora Funcione para maquinas individuais, foi pensado para a computação distribuída e paralelas (APACHE, 2022). O ecossistema Spark está mostrado na figura 1, segundo a documentação Apache Spark. Figura 1:

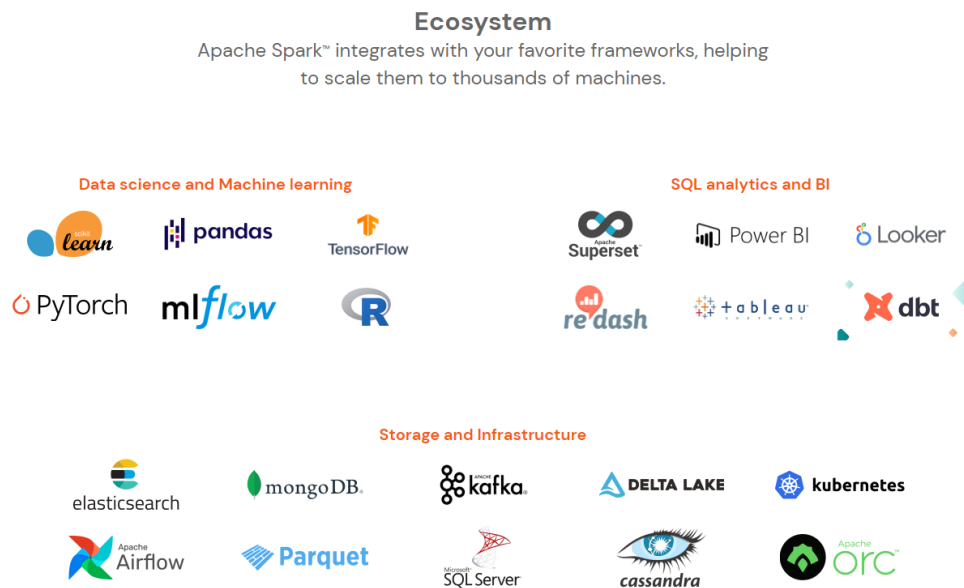


Figura 1: Ecosystema Spark

Spark foi desenvolvido sob o algoritmo *MapReduce*, desenvolvido para o Apache Hadoop, porém chega a ser cerca de 100 vezes mais rápido do que o Hadoop devido ao modelo de processamento da memória RAM e armazenamento, e funcionalidades integradas como GraphX, MLlib e SparkSQL (APACHE, 2022; DEVMEDIA, 2022).

Conta com 4 componentes principais (figura 2): O **SparkCore** - componente central, provê as funções fundamentais map, reduce, filter e collect; **Driver program** (contém o SparkContext, o componente central que gerencia o processamento vindo dos escravos); **Cluster manager** (necessário para a execução distribuída, administra os Workers; as aplicações podem ser administradas via *StandAlone* (local), Apache Mesos, Hadoop YARN ou Kubernetes) e **Workers** (executam as tarefas enviadas pelo *Driver Program*).

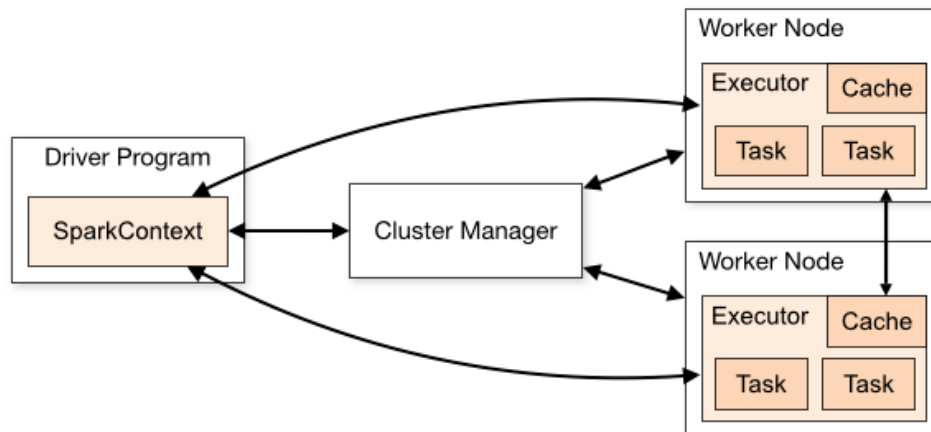


Figura 2: Arquitetura Spark.

Spark funciona a partir dos RDDs - *Resilient Distributed Datasets*: Objeto principal do Spark, faz a transformação dos dados, gerando novos RDDs, imutáveis depois de criados. Sua partição é feita em nós compondo clusters que podem ser paralelizados. Os RDDs podem ser criados a partir de HDFS (*Hadoop supported file system*) ou coleção Scala no *Drive Program* e podem ser persistidos na memória, para que seja utilizado eficientemente nas outras máquinas do cluster. Os RDDs se recuperam automaticamente de falhas nos nós, aumentando o grau de resiliência.

Segunda abstração do *Spark* são as variáveis compartilhadas para operações paralelas, agilizando as análises em várias vezes. As operações são divididas em transformações (por exemplo: agrupamentos, filtros e mapeamentos; são de *lazy evaluation* – não processadas até que comando de ação seja dado) ou ações (por exemplo: contagem e persistência) realizadas nos RDDs. **SparkContext**: é o objeto que conecta o spark ao programa sendo desenvolvido

### 1.1.2. Hadoop

É um framework em Java desenvolvido a partir da abstração *mapreduce*. É bastante semelhante ao Apache Spark descrito acima, sendo um projeto de software de código aberto que pode ser usado para processar de modo eficiente



grandes conjuntos de dados, foi desenvolvido para funcionar de modo distribuído com computação paralela, permitindo a análise e processamento de grande volume de dados. É customizável, apresentando diversos componentes que podem ser adicionados a ele conforme a necessidade da equipe. Alguns exemplos são o Hive, Pig, Hue, Ganglia, Oozie e HBase.

A arquitetura básica do Hadoop inclui o MapReduce (algoritmo), o YARN (gerente de recursos) e o HDFS (sistema de armazenamento distribuído). O Hadoop MapReduce, divide o trabalho em componentes menores que podem ser distribuídos entre os nós no seu cluster. Se um servidor falhar, o Hadoop reexecuta essa tarefa em outra máquina do cluster até sua conclusão. O YARN, monitora todos os recursos no cluster, garantindo alocação dinâmica para o processamento de tarefas e o Hadoop Distributed File System (HDFS) armazena dados em discos locais do cluster em grandes blocos e permite replicação de dados para gerar tolerância a falhas mencionado no trecho do mapReduce.

### 1.1.3. PowerBI

O Power BI é uma ferramenta de inteligência de negócios desenvolvida pela Microsoft, projetada para permitir transformações de grandes volumes de dados brutos em informações visuais e estratificáveis (MICROSOFT, 2023).

Conta com interface relativamente simples, o Power BI facilita a importação de ampla gama de tipos dados. Talvez sua característica mais interessante é uma visualização estratificada dos dados, que não é possível com outras ferramentas, onde os *dashboards* permitem divisões complexas dos dados, possíveis com a simples adição de um botão. Embora não permita gráficos detalhados, como aqueles desenhados em Excel, a análise exploratória de dados fica muito simples e objetiva, permitindo obtenção de *insights* valiosos e diferenciados.

O Power BI também permite integração com basicamente todos os outros produtos Microsoft, embora esse não seja o foco desse projeto. Diria que o processo de tratamento dos dados e a capacidade de trabalhar com grande

volume de dados são os aspectos mais úteis e comumente bem avaliados pela comunidade dos dados, razão pela qual escolhi utilizá-lo.

#### 1.1.4. Zeppelin

O Apache Zeppelin é uma plataforma de código aberto que fornece ambiente de notebook interativo para análise de dados bastante ampla. Ele suporta várias linguagens de programação, como Python, Scala, SQL, entre outras, permitindo que os usuários integrem diferentes etapas de análise em um único documento interativo. Nesse caso, permite utilização do PySpark, sendo a opção de use devido a precariedade das VMs disponibilizadas.

Os notebooks Zeppelin podem conter código, gráficos, texto explicativo e até mesmo widgets interativos. Os notebooks são divididos em células, inclusive podendo rodar linguagens específicas em cada célula.

Zeppelin facilita a visualização de dados através da geração de gráficos interativos, como seria o papel do software “Voila”, comumente associado ao ambiente anaconda.

O aspecto mais importante do Zeppelin diz respeito a Integração com Big Data. Zeppelin é conhecido por sua integração com diversos frameworks de big data, como Apache Spark, Apache Flink e Apache Hadoop. Isso permite realizar análises em conjuntos de dados massivos.

Além desses aspectos, ainda permite colaboração e compartilhamento para grupos, extensões para customizar conforme suas necessidades. Além de representar um ótimo ambiente de desenvolvimento.

## 2. APRESENTAÇÃO DO PROBLEMA

O objetivo se manteve: é realizar uma análise abrangente dos dados de saúde disponíveis para melhorar o atendimento ao paciente nas unidades de saúde. Os dados contam com grande volume, provenientes de diferentes fontes, incluindo registros de pacientes, informações de tratamento, dados

demográficos e muito mais. O Big Data nos oferece uma oportunidade única de extrair *insights* significativos desses dados para aprimorar a eficiência operacional, otimizar os recursos e, o mais importante, melhorar a qualidade do atendimento ao paciente.

### **Problema:**

1. **Eficiência Operacional:** Entender os padrões de demanda dos pacientes ao longo do tempo para otimizar o agendamento de consultas e evitar sobrecargas em determinados períodos.
2. **Atendimento ao Paciente:** Identificar padrões em dados clínicos para melhorar o diagnóstico precoce, garantindo que os pacientes recebam o tratamento adequado no momento certo.
3. **Uso de Recursos:** Otimizar o uso de recursos, como equipamentos médicos e pessoal, para garantir que estejam disponíveis quando e onde são mais necessários.

### **Metodologia Proposta:**

1. **Análise de Padrões:** Utilizar análise exploratória de dados e *machine learning* para entender as variações na demanda dos pacientes ao longo do dia, da semana e do ano.
2. **Predição de Dados Clínicos:** Utilizar técnicas de mineração de dados para identificar padrões em dados clínicos que possam indicar condições de saúde específicas para tratamento precoce.
3. **Otimização de Recursos:** Utilizar algoritmos de otimização para distribuir recursos de forma eficiente, evitando gargalos e maximizando a utilização.

## **Benefícios Esperados:**

1. **Melhoria na Experiência do Paciente:** Redução nos tempos de espera, consultas mais pontuais e melhor atendimento ao paciente devido a diagnósticos mais precisos.
2. **Eficiência Operacional:** Melhoria na alocação de recursos, reduzindo custos operacionais e aumentando a eficiência da equipe.
3. **Resultados de Saúde Aprimorados:** Identificação precoce de condições de saúde, melhorando as taxas de tratamento bem-sucedido e a qualidade de vida dos pacientes.

## **3. RELEVÂNCIA**

A análise desses dados nos permite encontrar tendências e falhas de alocação de recurso em determinados setores, permitindo correção e melhorias contínuas de processo de modo a minimizar desperdício de recursos.

## **4. OBJETIVO**

Objetivo é realizar uma análise abrangente dos dados de saúde disponíveis para melhorar o atendimento ao paciente em nossos estabelecimentos de saúde. Possuímos uma grande quantidade de dados provenientes de diferentes fontes, incluindo registros de pacientes, informações de tratamento, dados demográficos e muito mais. O Big Data nos oferece uma oportunidade única de extrair insights significativos desses dados para aprimorar a eficiência operacional, otimizar os recursos e, o mais importante, melhorar a qualidade do atendimento ao paciente.

- 1 – Análise exploratória dos dados através de notebooks Python
- 2 – Análise de desenvolvimento de Dashboards em PowerBI
- 3 – Análise e comparação com Hadoop
- 4 – Análise e comparação com Spark
- 5 – Elaboração de relatório/ Storytelling

## 5. CRONOGRAMA

O cronograma leva em consideração o período para a elaboração da ATP I, com as atividades a serem realizadas bem como um cronograma de Gantt, para facilitar e melhor resumir os processos e o fluxo de trabalho.

### 1. SEMANA 5 – Clean up e análise em PowerBI:

- 1.1. Clean up;
- 1.2. Análise exploratória;
- 1.3. Dashboards;

### 2. SEMANA 6 – Análise de *machine learning*:

- 2.1. seleção das variáveis relevantes;
- 2.2. Escolha do método;
- 2.3. Ajuste de features;

### 3. SEMANA 7 - Análise em BigData:

- 3.1. Utilização da ferramenta Hadoop;
- 3.2. Utilização da ferramenta PySpark;

### 4. SEMANA 8 – Desenvolvimento do StoryTelling:

- 4.1. Entrega do relatório.

TABELA 1: CRONORAMA DE GUNTT DO PROJETO DA SOMATIVA 1

	Semana 5					Semana 6					Semana 7					Semana 8				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<b>1. SEMANA 5 – CleanUp e BI</b>																				
1.1. CleanUp																				
1.2. Análise exploratória																				
1.3. DashBoards																				
<b>2. SEMANA 6 - Machine Learning</b>																				
2.1. Seleção das variáveis																				
2.2. Escolha do método																				
2.3. Ajuste de Features																				
<b>3. SEMANA 7 - BigData</b>																				
3.1. Hadoop																				
3.2. PySpark																				
<b>4. SEMANA 8 – StoryTelling</b>																				
4.1. Entrega do Relatório																				

Tabela 1: cronograma de Gunnt. Em Azul -> atividade antes do prazo; Em vermelho -> atividade atrasada; em verde -> atividade entregue no prazo.

## 6. RESULTADOS E DISCUSSÃO

Nessa etapa, serão entregues o restante das análises indicadas na etapa I, com adição da análise via PowerBI. Gerando 4 arquivos: esse PDF, o HTML contendo o *python notebook*, dicionário de dados e os *dashboards*.

Vamos iniciar com as comparações e análise de dados que não foram possíveis de se analisar no ambiente Python diretamente. Essa análise leva em consideração 1,1 milhão de linhas em 27 colunas, apresentando perdas mínimas de dados, ao contrário do observado anteriormente.

### 6.1. PERFIL DO PACIENTE DAS UNIDADES DE SAÚDE

Para iniciar, temos o número de atendimentos por sexo. É possível verificar que 62% das consultas são feitas para o sexo feminino. Dado

interessante, corroborando para o desenvolvimento de frentes de combate ao descaso com a saúde masculina.

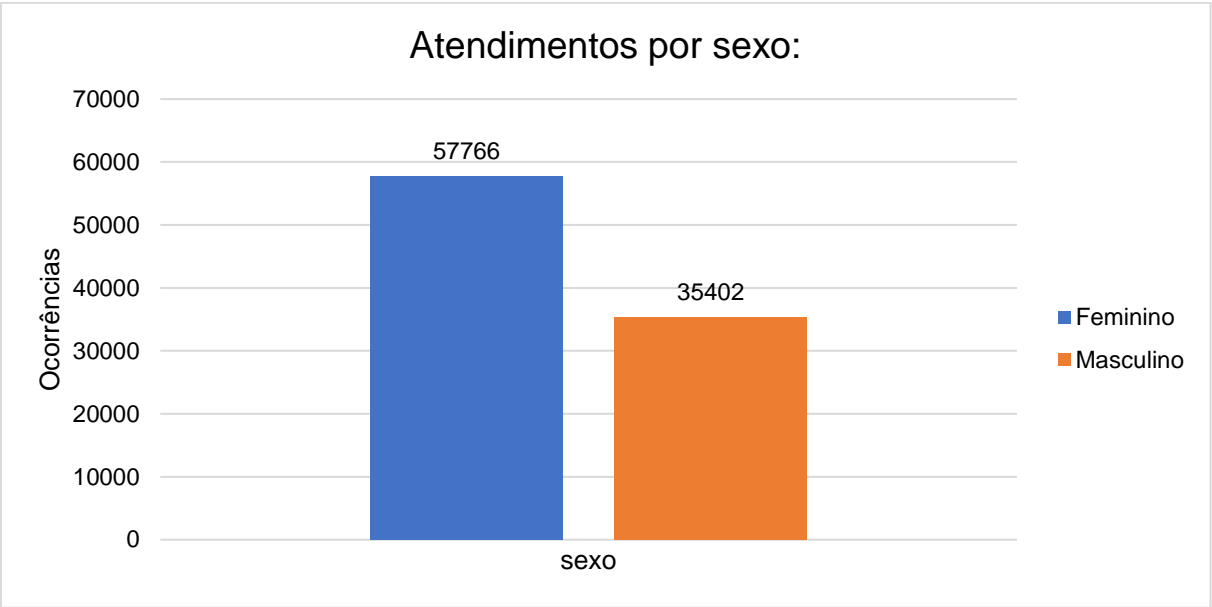


Figura 3: Atendimento segundo o sexo do paciente.

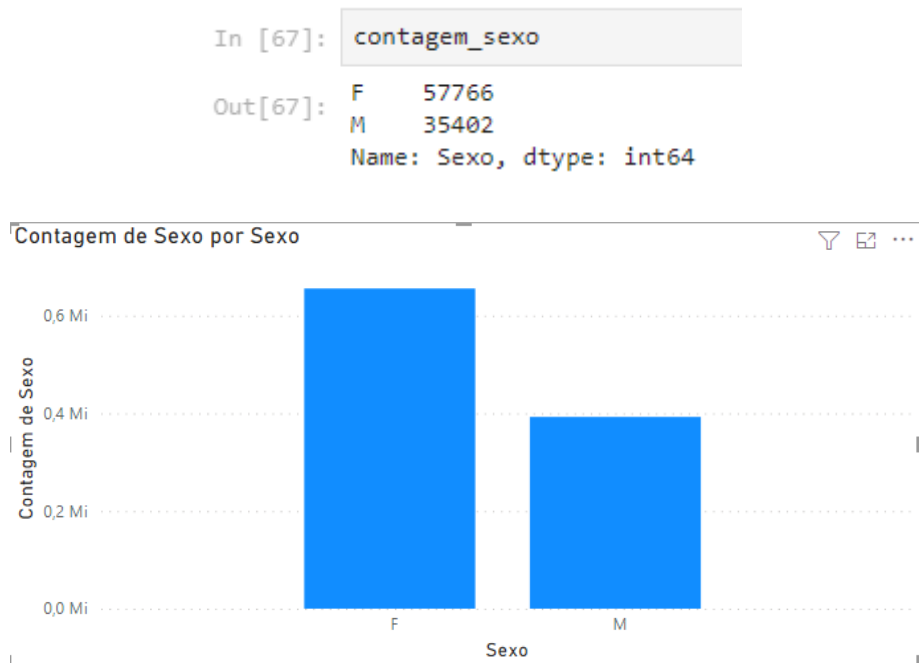


Figura 4: Gráfico de barras do dataset completo feito em PowerBI.

Com o *dataset* completo, foi possível verificar leve discrepância através da análise realizada via PowerBI, onde verificamos 655844 atendimentos do sexo

feminino (60%) e 392731 atendimentos do sexo masculino (40%), apresentando uma leve diferença de 2% em relação a amostra anterior, mostrando que o teorema central do limite foi respeitado para esse espaço amostral.

A próxima informação extraída diz respeito a nacionalidade dos pacientes atendidos pelo sistema único de saúde em Curitiba, uma vez que o SUS preza pela universalidade irrestrita de seu atendimento. Pode ser verificado que 96% dos pacientes são brasileiros natos, dado interessante se considerarmos que curitiba conta com 19 mil estrangeiros oficialmente segundo a prefeitura (CURITIBA, 2023), tornando o grupo superrepresentado nos dados (2,5% representando 1% da população), indicando que realmente o princípio universalista está sendo cumprido, sem deixar moradores à mingua.

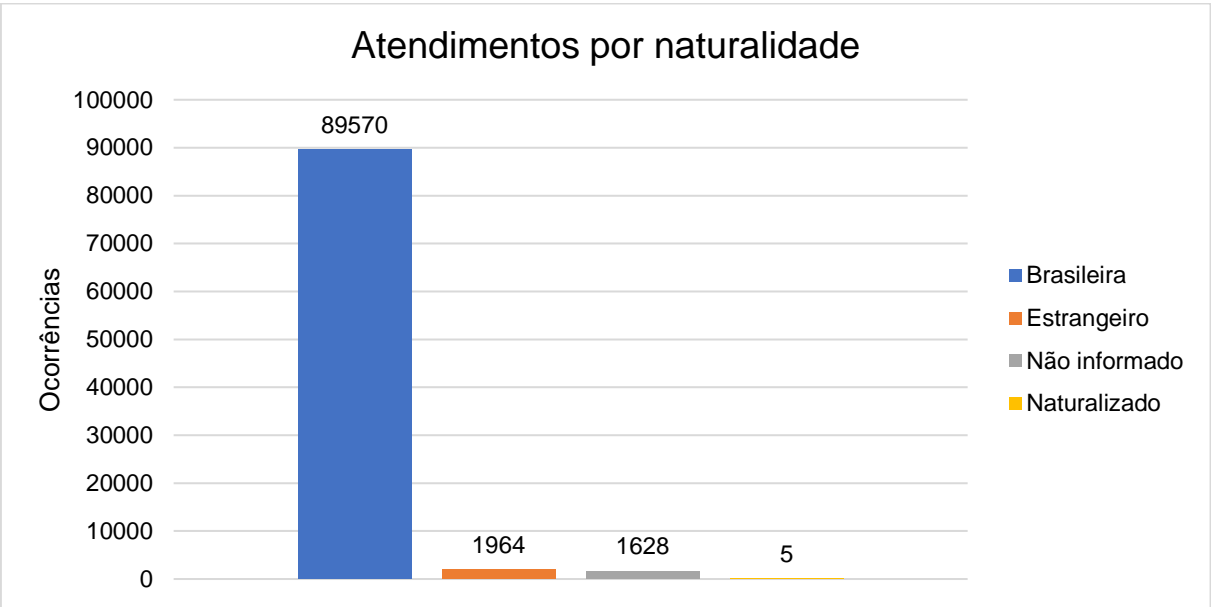


Figura 5: Nacionalidade dos pacientes do SUS em Curitiba.

```
In [76]: contagem_n
Out[76]: Brasileira      89570
          Estrangeiro    1964
          Não informado   1628
          Naturalizado     5
          Name: Nacionalidade, dtype: int64
```



Contagem de Nacionalidade por Nacionalidade



Figura 6: Gráfico de nacionalidade dos pacientes com o dataset completo.

Novamente, a tendência dos dados se manteve próximo da amostra anterior, com 10135630 (96,6%) dos pacientes Brasileiros natos.

Já para o perfil de atendimentos por localidade. Esses pacientes são 96,5% de Curitiba, em seguida vem os pacientes de Pinhais, com 1% dos atendimentos em Curitiba e o restante das cidades apresentam menos do que 0,5% do volume de pacientes. Esses pacientes comumente representam pacientes de alta complexidade alocados em hospitais para tratamentos e exames. O gráfico a seguir demonstra os dados:



Figura 7: Atendimento seguindo a cidade de origem.

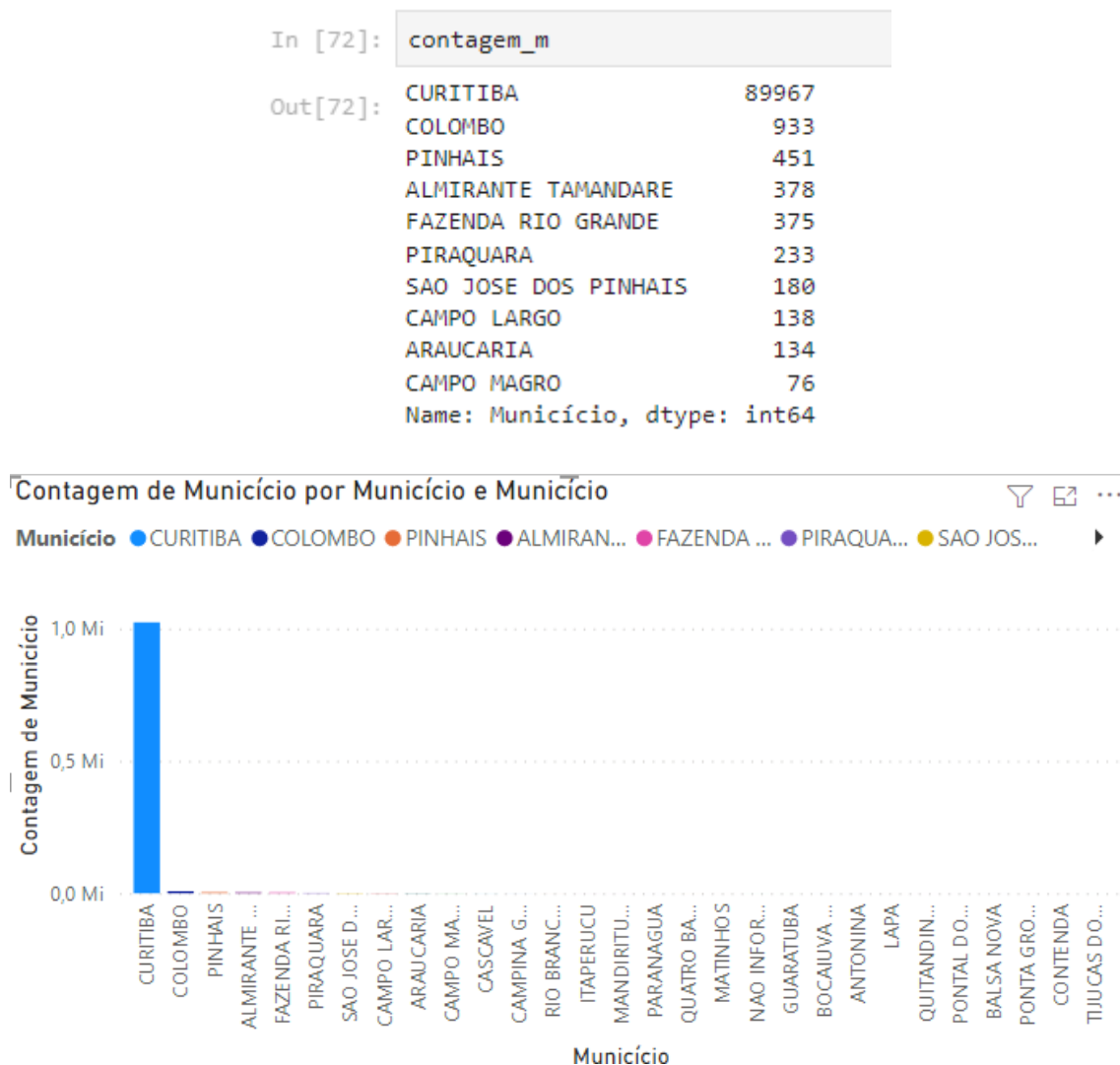


Figura 8: Atendimentos por município pelo SUS Curitiba, Dados completos.

Com o *dataset* completo, os atendimentos de residentes de Curitiba ficaram levemente mais elevado, com 97,7% dos atendimentos (1024004).

10240A próxima informação diz respeito ao tipo de unidade que presta o atendimento aos pacientes. É possível verificar que 65% dos pacientes são atendidos em unidades básicas de saúde, 32% são atendidos em Unidades de pronto atendimento e 3% dos pacientes são atendidos em unidades especializadas, demonstrando o perfil de atendimento do SUS, onde existe a necessidade de triagem médica para que possam ser encaminhados para tratamentos de maior complexidade, como demonstrado no gráfico a seguir:

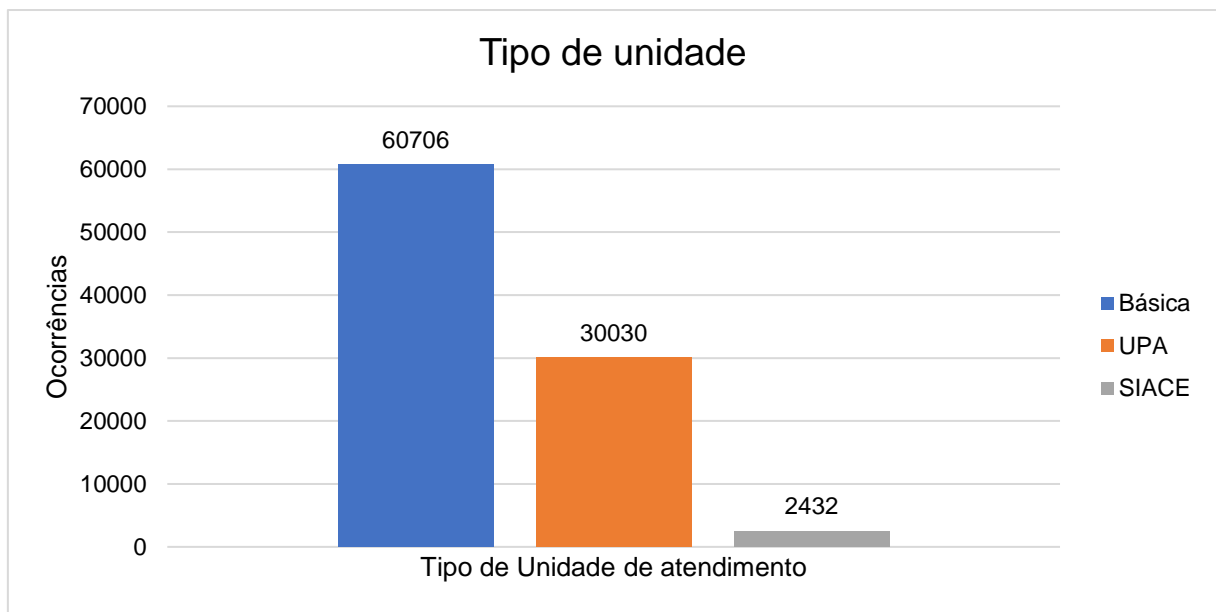


Figura 9: Atendimento segundo o tipo de unidade de atendimento ao cliente.

```
In [68]: contagem_upa
```

```
Out[68]:
```

BASICO	60706
UPA	30030
SIACE	2432

Name: Tipo de Unidade, dtype: int64

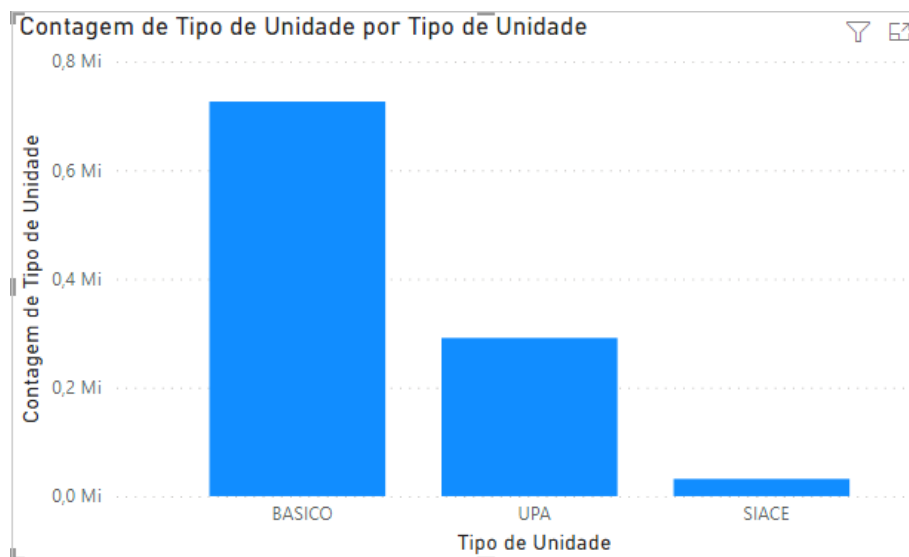


Figura 10: Gráfico de atendimentos por tipo de unidade de atendimento completo.

Para o perfil de atendimento, o dataset completo mostra acumulo nos atendimentos nas unidades básicas em relação as unidades de pronto atendimento com 69% (725940), 28% e 3%, repectivamente.

Alguns dados relacionados com o perfil do paciente não foram mostradas na etapa I devido à omissão dos dados pela biblioteca pandas, esses são “tipo de habitação”, “Numero de comodoss em casa”, “tratamento de agua”, “Tratamento de esgoto”, “Energia elétrica”, “Origem do usuário”, “Destino do lixo”, “Sistema de esgoto” e “Destino em caso de doença”.

Esses dados serão discutidos a partir daqui, iniciando com o tipo de habitação, onde 78,8% dos pacientes atendidos em Curitiba contam com habitação em alvenaria (728000), 5% casas de madeira e menos de 2% habitam em formas insalubres de moradia. 12% não informaram.

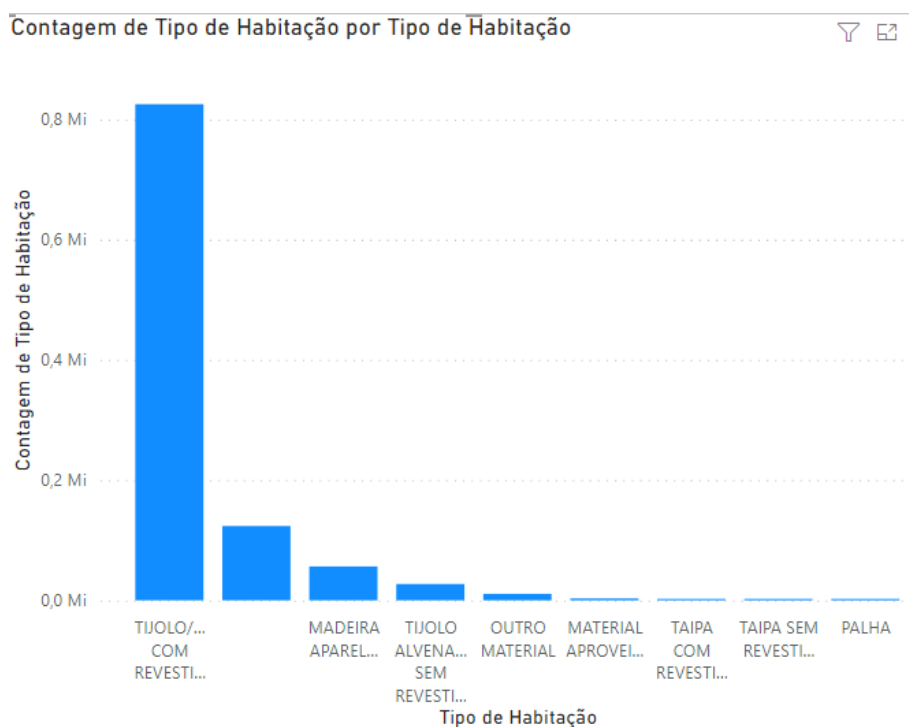


Figura 11: Perfil de moradias em Curitiba. dataset completo.

### Contagem de Cômodos por Cômodos

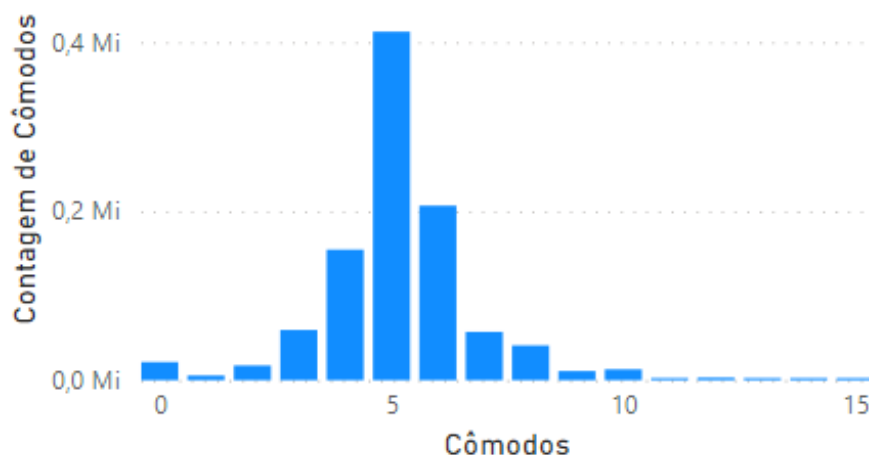


Figura 12: Média do número de cômodos dos usuários do SUS Curitiba. Dataset completo.

O número de cômodos presente nas casas dos pacientes do SUS Curitiba segue uma distribuição normal com a média ligeiramente acima de 5 cômodos, valor o qual representa 40% das casas (405 mil).

Temos o perfil de água nas residências dos usuários do SUS Curitiba, aqui é possível observar que 46% (480 mil) dos pacientes não tratam sua água, enquanto que 20% realizam a cloração e 20% filtram sua água.

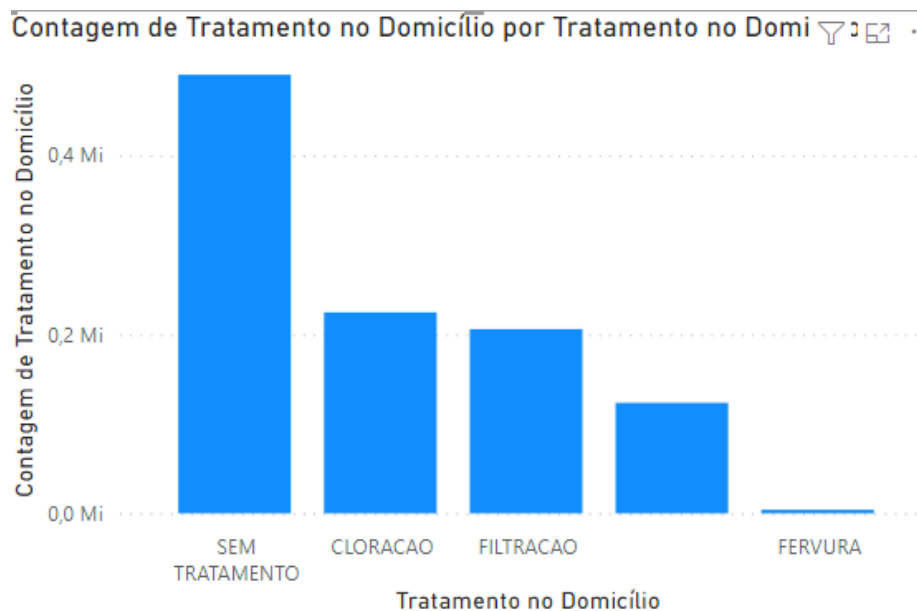


Figura 13: Tratamento da água para consumo nos domicílios dos usuários do SUS em Curitiba. Dataset completo.

Quanto ao perfil de abastecimento de água, 906 mil pacientes recebem água tratada pela sanepar, 86,4%. 12% não responderam e as demais opções somam 1,6%.

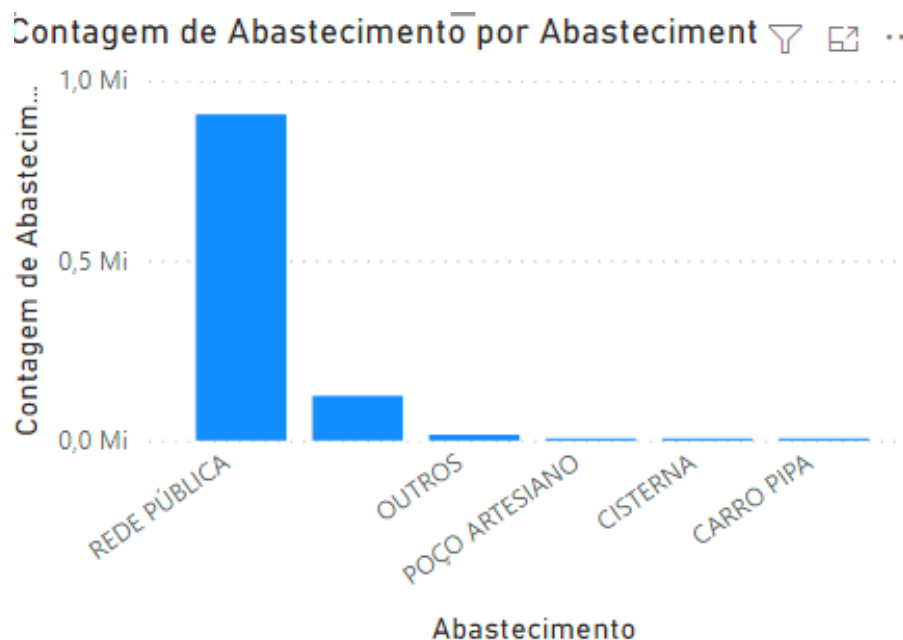


Figura 14: Perfil de Abastecimento de água dos usuários do SUS em Curitiba. Dataset completo.

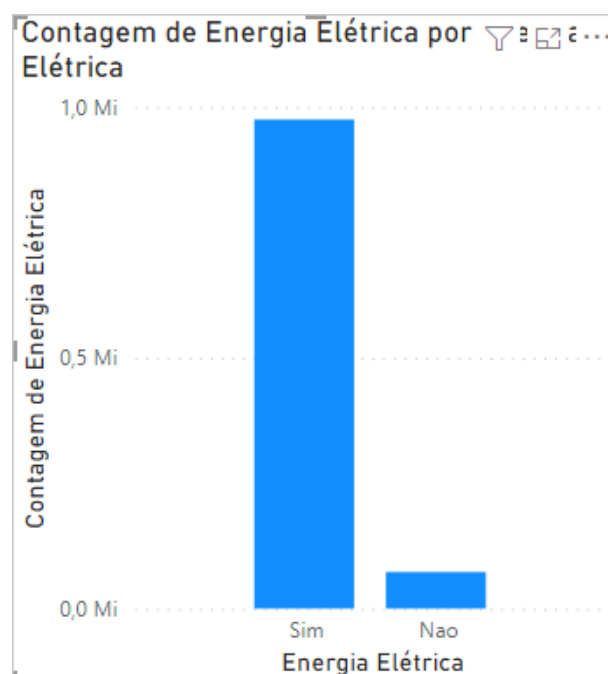


Figura 15: Disponibilidade de luz elétrica dos pacientes do SUS Curitiba. Dataset completo.

93% dos pacientes do SUS contam com luz elétrica em casa (975 mil).

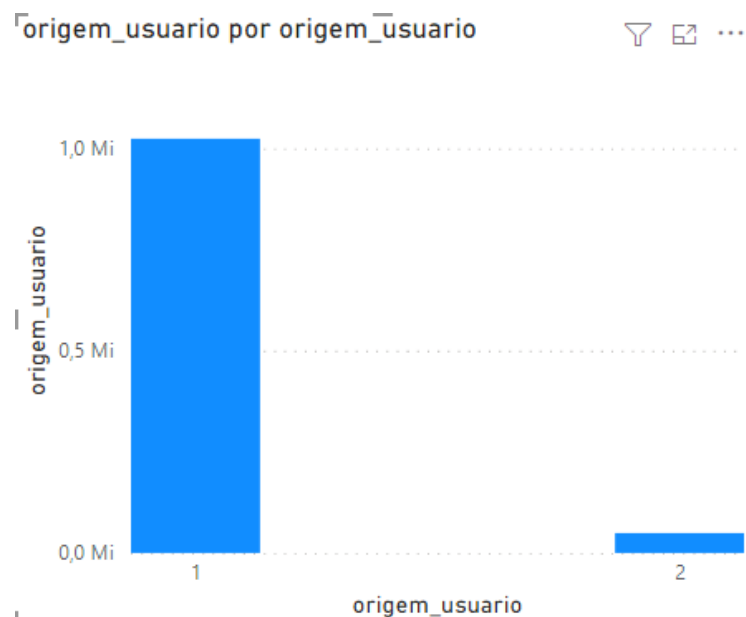


Figura 16: Origem do usuário. 1 - Rede pública, 2 - Planos de saúde. Dataset completo.

95% dos usuários do SUS Curitiba (1 milhão) não foram encaminhados por planos de saúde. Lembrando que alguns tratamentos são de monopólio do SUS, principalmente prevenção e tratamento do HIV, tuberculose, Hanseníase e boa parte dos tratamentos antineoplásicos.

Contagem de Destino Lixo por Destino Lixo

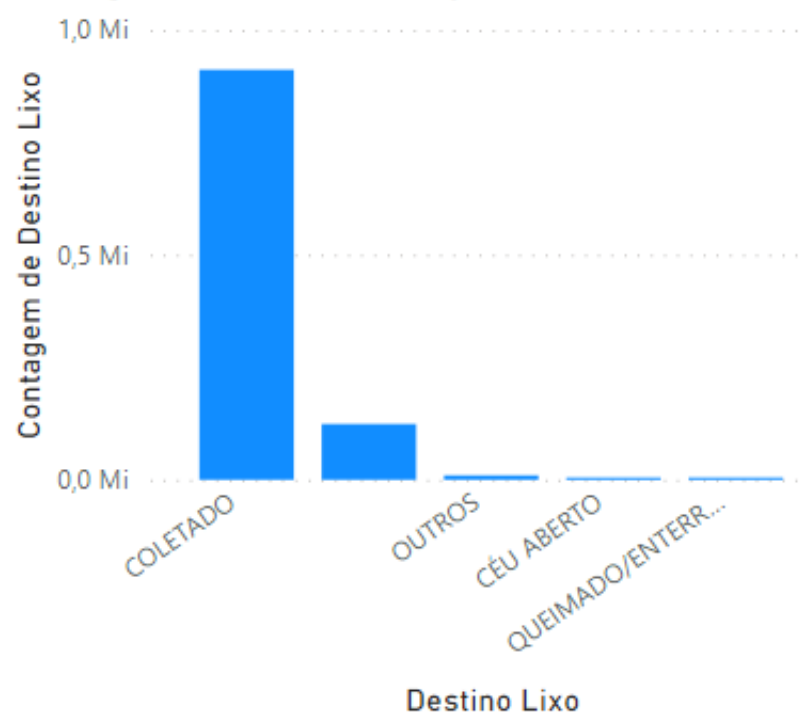


Figura 17: Coleta de lixo. Dataset completo.

87% (911 mil) dos pacientes do SUS Curitiba tem acesso a coleta de lixo. 12% não informaram. Numero bastante parecido com aqueles que contam com acesso ao saneamento básico, serviço de esgoto, com 83% do total (864 mil). Esses dados mostram descaso público em relação ao saneamento básico dessa parcela da população que utiliza o SUS, visto que a média de saneamento da cidade está em torno de 97% (IBGE 2019).



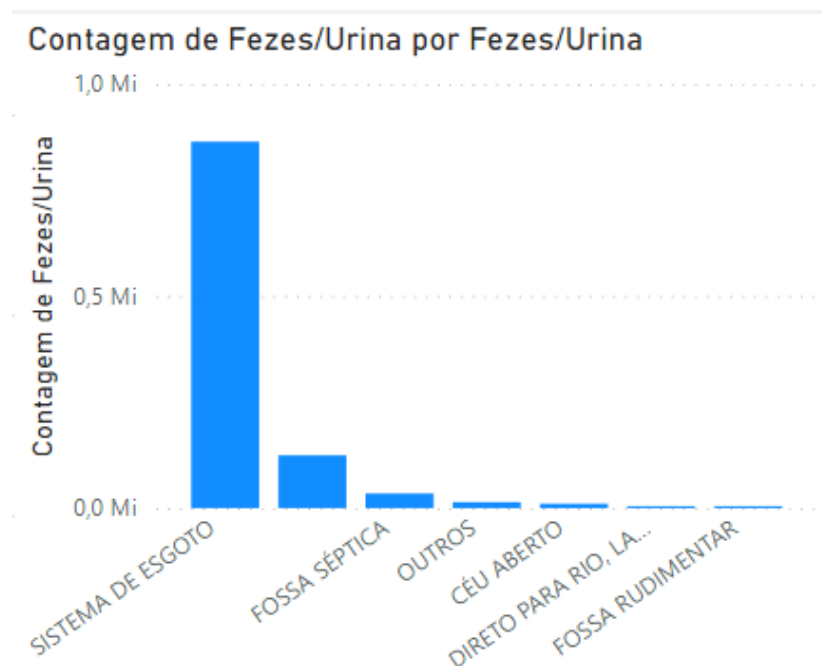


Figura 18: saneamento básico. perfil do Esgoto dos usuários do SUS Curitiba. Dataset completo.

Para onde o paciente será enviado em caso de doença, temos que 85% dos pacientes buscam unidade básica, UPA e emergência, reforçando a dependência que essa porção da população tem em relação ao SUS.

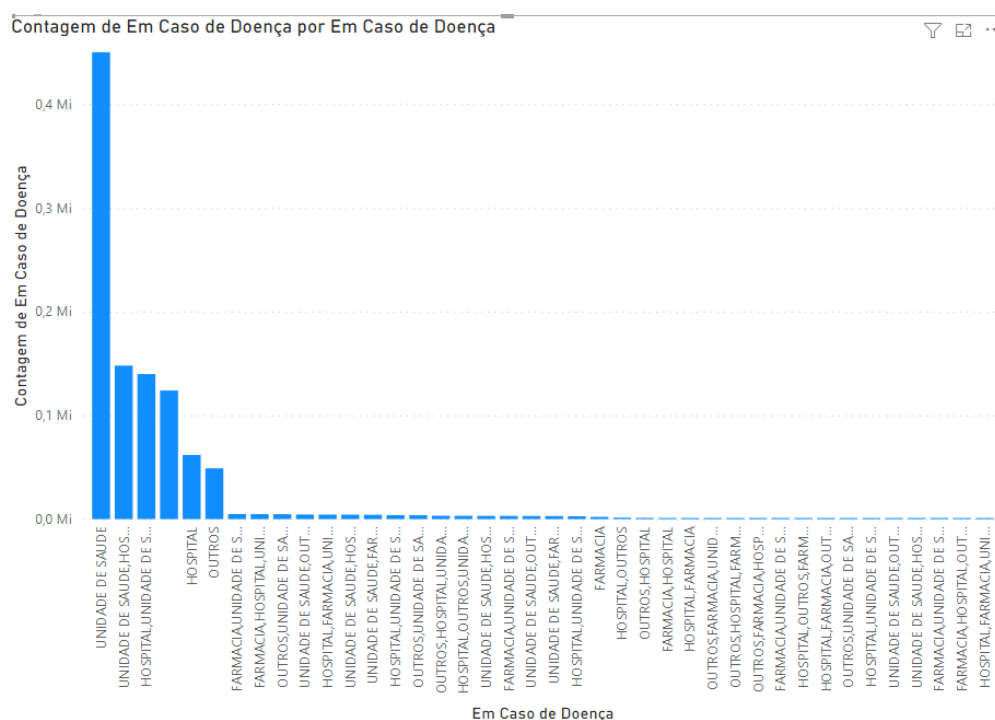


Figura 19: Locais de atendimento em caso de doença. dataset completo.

Por fim, o perfil de transporte dos pacientes. É possível verificar que 54% deles são dependentes exclusivamente do transporte público. 34 % dos pacientes tem acesso a veículo porém sem autonomia de utilização total. 6 % dos pacientes são usuários de outras formas de transporte, como motocicletas, bicicletas e caminhada e apenas 5% dos pacientes do SUS tem acesso integral a carro próprio. Indicando que embora universalista, grupos de maior poder aquisitivo evitam o sistema único de saúde, como mostrado a seguir:

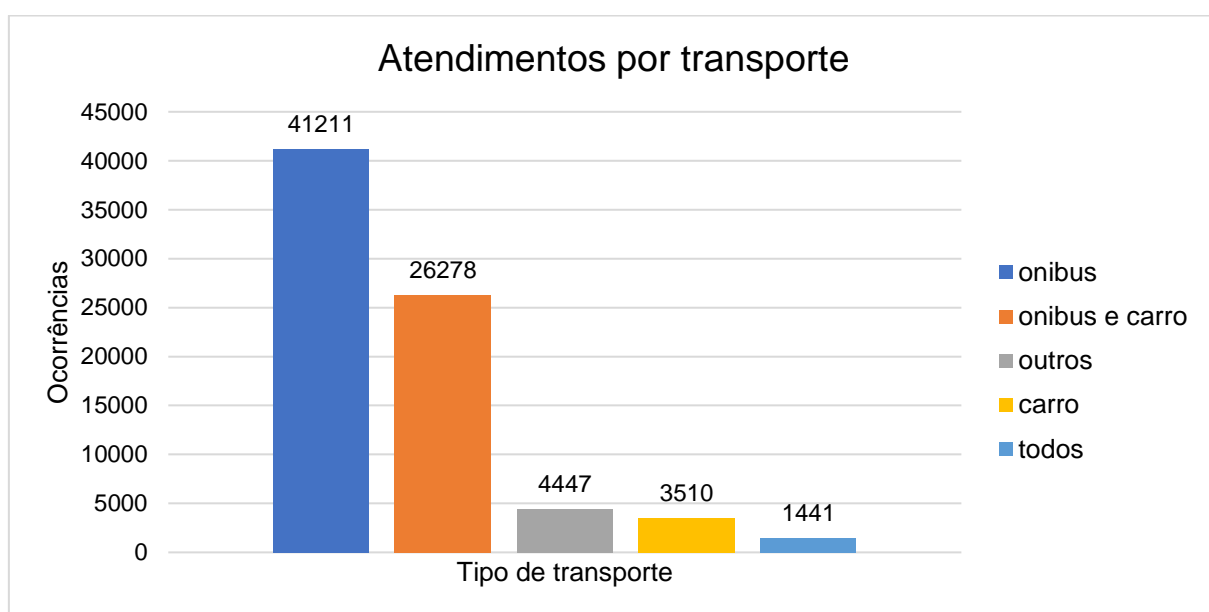


Figura 20: perfil de deslocamento dos usuários do SUS em Curitiba.

```
In [70]: contagem_t

Out[70]: ONIBUS          41211
ONIBUS,CARRO      26278
OUTROS            4447
CARRO             3510
OUTROS,ONIBUS,CARRO 1441
Name: Meio de Transporte, dtype: int64
```

## Contagem de Meio de Transporte por Meio de Transporte

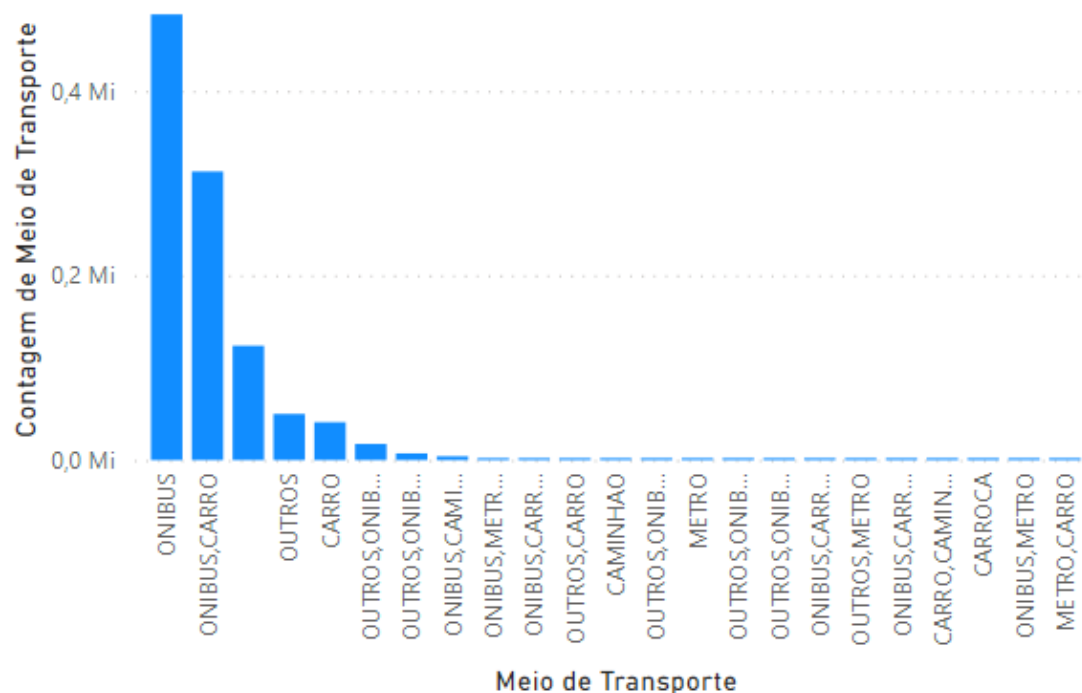


Figura 21: Perfil do meio de transporte dos usuários do SUS Curitiba. Dataset completo.

Nesse caso a distribuição teve sua dispersão aumentada, com 46% dos pacientes utilizando exclusivamente o transporte público (483 mil). Nesse caso, esse aumento na dispersão se deu devido a amostra utilizada na etapa I não conter em torno de 10% de respostas omitidas como verificado aqui.

## 6.2. PERFIL DE ATENDIMENTO NA UNIDADES DE SAÚDE

Agora, veremos o perfil de atendimento das unidades de saúde. A seguir, está presente o gráfico dos 10 bairros com maior número de atendimentos pelas unidades de saúde. A CIC é o bairro com maior número de atendimentos, representando 13,5% de todos os atendimentos realizados pelo SUS de Curitiba, seguido pelo Sitio Cercado com 8% dos atendimentos e Cajuru com 5% dos atendimentos totais de Curitiba, mostrando uma distribuição relativamente uniforme entre os 75 bairros. As posições correlacionam positivamente com a população geral do bairro, mas não diretamente, mostrando alguns focos de

demanda que poderiam receber foco. Exemplos positivos disso são o CIC , que embora seja o bairro mais populoso de Curitiba, está sobre representado em número de atendimentos (13,5%) visto conter 5,7% da população. Também o Tatuquara, que apresenta em torno de 5% dos atendimentos, mas apenas 2% da população. Os demais bairros estão sub representados, refletindo em superlotação e dificuldade de alocação de consultas.

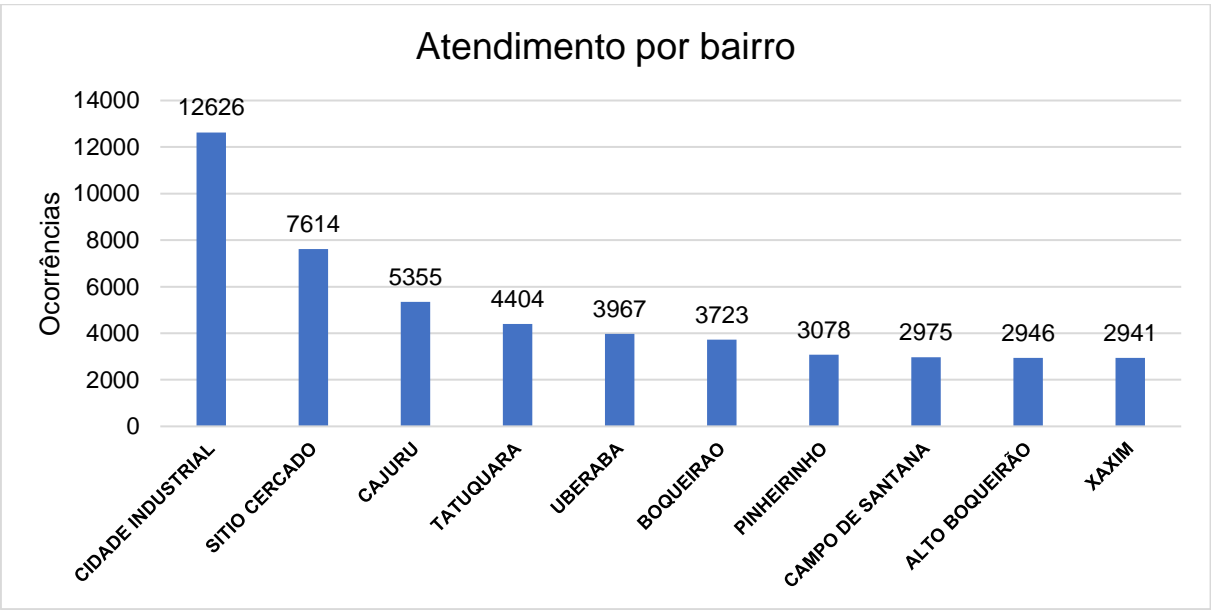


Figura 22: Atendimento por bairro em Curitiba

```
In [74]: contagem_b

Out[74]: CIDADE INDUSTRIAL    12626
          SITIO CERCADO       7614
          CAJURU              5355
          TATUQUARA           4404
          UBERABA              3967
          BOQUEIRAO           3723
          PINHEIRINHO         3078
          CAMPO DE SANTANA     2975
          ALTO BOQUEIRAO       2946
          XAXIM                2941
          Name: Bairro, dtype: int64
```

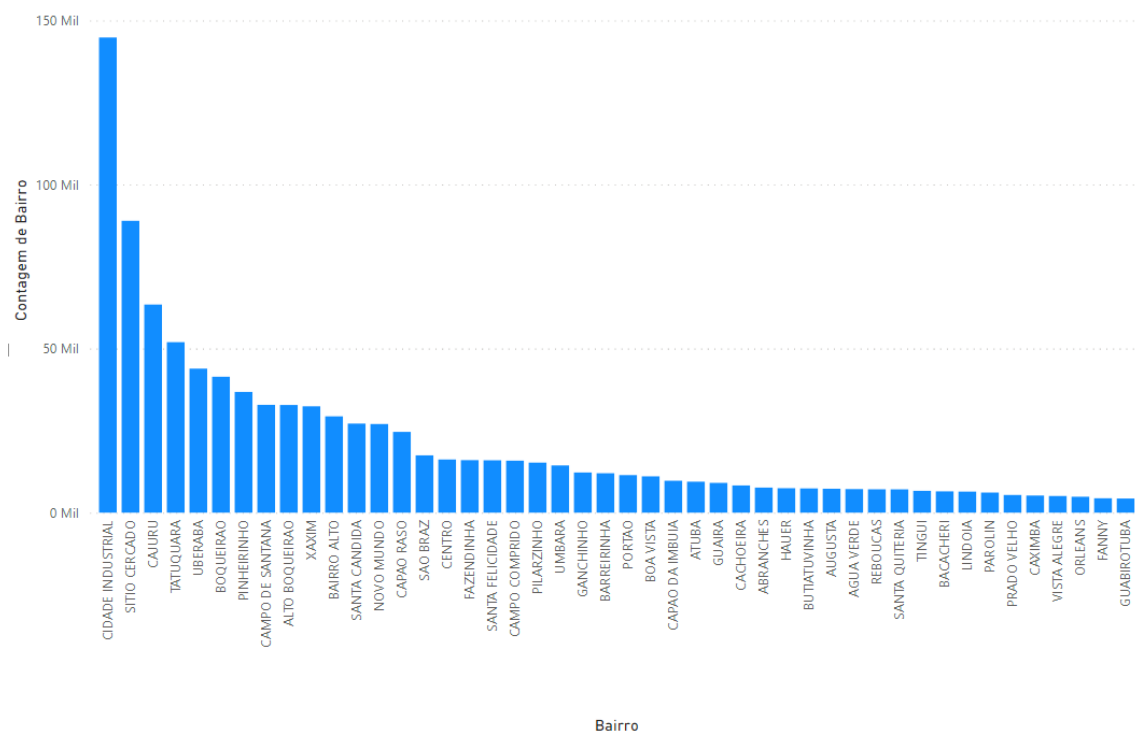


Figura 23: Atendimentos por Bairro com dataset completo.

Perfil dos dados mantido para o dataset completo, onde CIC manteve 13,8% dos atendimentos totais da cidade de Curitiba, porém dessa vez não foram omitidos os bairros além dos 10 principais, como mostrado anteriormente.

Outro ponto interessante diz respeito as Unidades de saúde que mais atendem, indiferente do bairro que faz parte. Verifique que as Unidades de saúde com maior número de atendimentos não seguem diretamente o perfil de atendimento por bairro. A UPA Cajurú é a unidade com maior numero de atendimentos de Curitiba, com 4% dos atendimentos, representando mais atendimentos isolados do que as 5 unidades presentes no Boqueirão. Também é possível verificar que o CIC, embora conte com o maior numero de atendimentos, não apresenta unidades grandes ou super lotadas, indicando demanda excessiva na UPA Cajurú e nas demais ranquadas até a UPA Boqueirão.

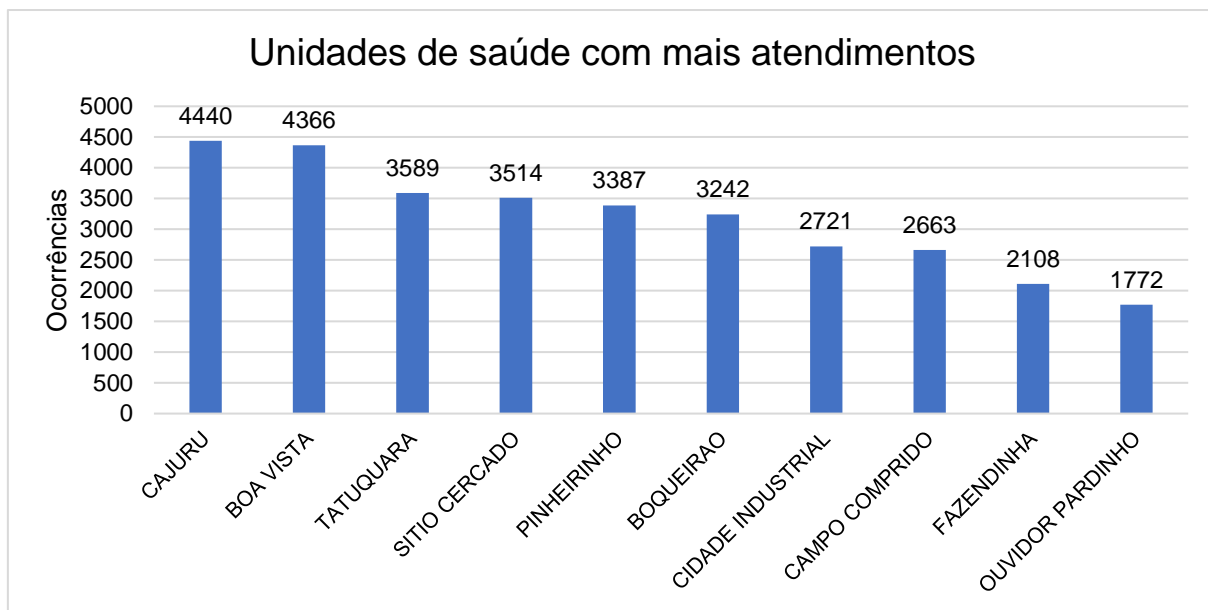


Figura 24: Atendimento segundo as maiores unidades de saúde de Curitiba

```
In [80]: contagem_ud
```

```
Out[80]: UPA CAJURU          4440
UPA BOA VISTA        4366
UPA TATUQUARA        3589
UPA SITIO CERCADO    3514
UPA PINHEIRINHO      3387
UPA BOQUEIRAO        3242
UPA CIDADE INDUSTRIAL 2721
UPA CAMPO COMPRIDO   2663
UPA FAZENDINHA       2108
UMS OUIDOR PARDINHO  1772
Name: Descrição da Unidade, dtype: int64
```

## Contagem de Descrição da Unidade por Descrição da Unidade

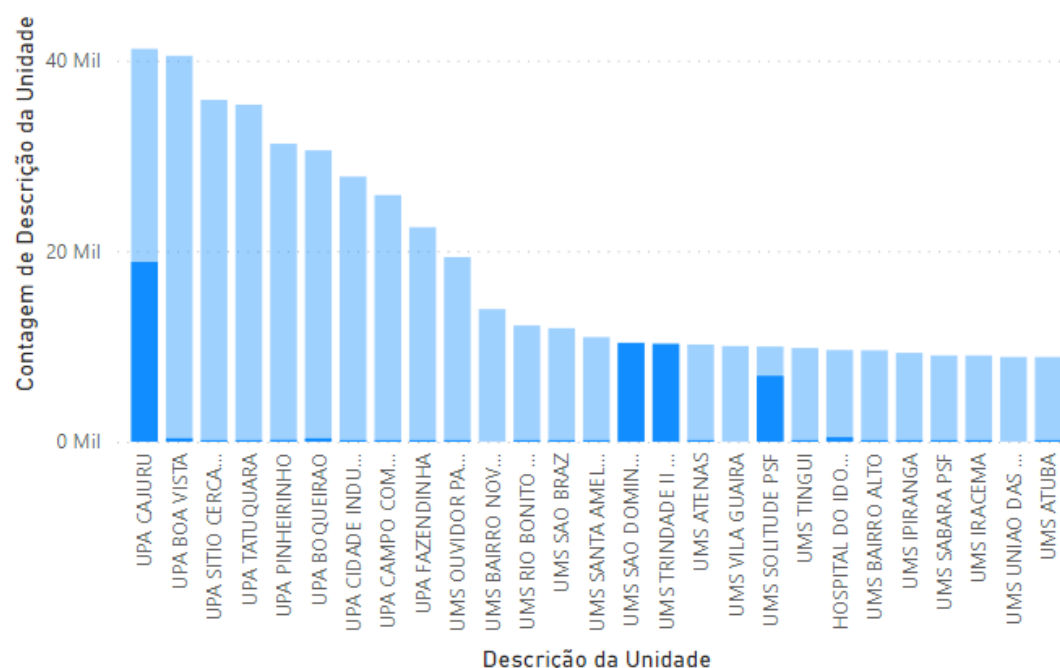


Figura 25: Unidades de Saúde com o maior número de atendimentos em Curitiba. Dataset completo.

Nesse caso, o perfil se manteve inalterado em relação ao espaço amostral da etapa I, porém a análise através do PowerBI permite verificar de onde veio os dados que compõe a barra. Nesse caso, é possível verificar que a UPA do CIC é bem distribuído, sem acumular demandas de vários outros bairros, porém UPAs como a Sitio cercado e Cajuru atendem vários outros bairros, muitas vezes assumindo totalmente essa demanda, como o UPA Sitio Cercado que atende a totalidade do Bairro novo, indicando grandes falhas estratégicas na demanda.

Essa característica também é verificada em relação aos maiores prescritores, poucos profissionais atendem a maior parte da demanda, eram 1444 prescritores durante aquele período, levando-se em consideração a média de 0,069% das prescrições por médico, temos que os maiores prescritores tem média mais 4,5 vezes maior do que a média, indicando má alocação de recursos e pouca capacidade de avaliar os serviços prestados dos médicos contratados, levando a desperdício de recursos, como será mostrado na figura a seguir.

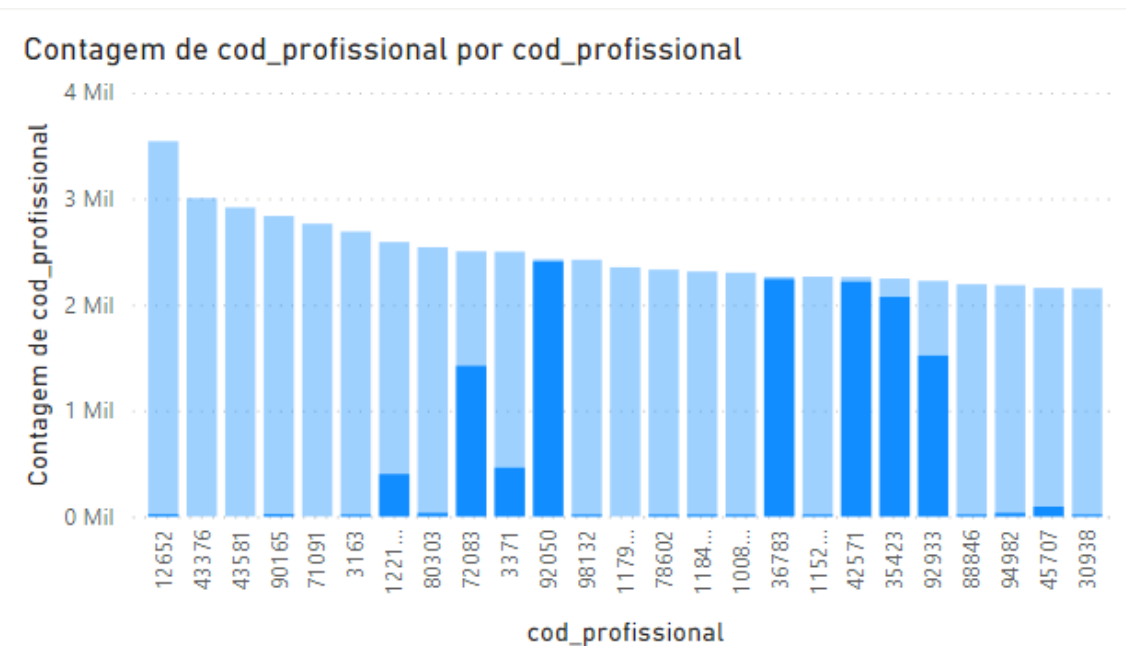


Figura 26: Maiores prescritores do SUS Curitiba. Dataset completo.

Talvez o melhor exemplo da má alocação de recursos e concentração das responsabilidades sobre alguns profissionais, permitido pelo PowerBI, esteja mostrado no gráfico 26 acima. Perceba que o maior bairro de Curitiba (CIC) é atendido praticamente por 8 profissionais prescritores.

Continuando o perfil do atendimento médico, a figura a seguir mostra que 980 mil atendimentos (88%) foram realizados por “residentes”.

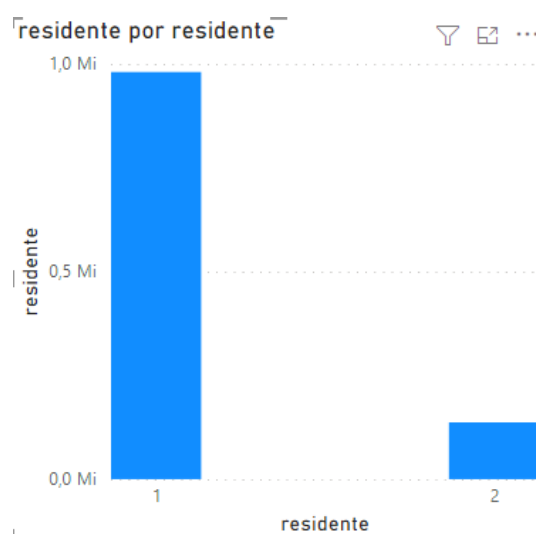


Figura 27: Perfil do profissional que atende no SUS Curitiba. Dataset completo.



O termo residente possivelmente está sendo utilizado incorretamente. Residentes são profissionais formados realizando uma pós graduação do tipo especialização profissional, e nesse caso devem estar incluído estagiários de graduação, o que poderia explicar parcialmente a carga extra sobre os prescritores mais assíduos, indicando que grande parte desse 88% compões profissionais não capacitados para prescrever ou tomar ação. Isso pode representar uma forma de embelezar os dados de pessoas/médico da cidade, visto que esse índice compõe o cálculo de índice de desenvolvimento humano (IDH).

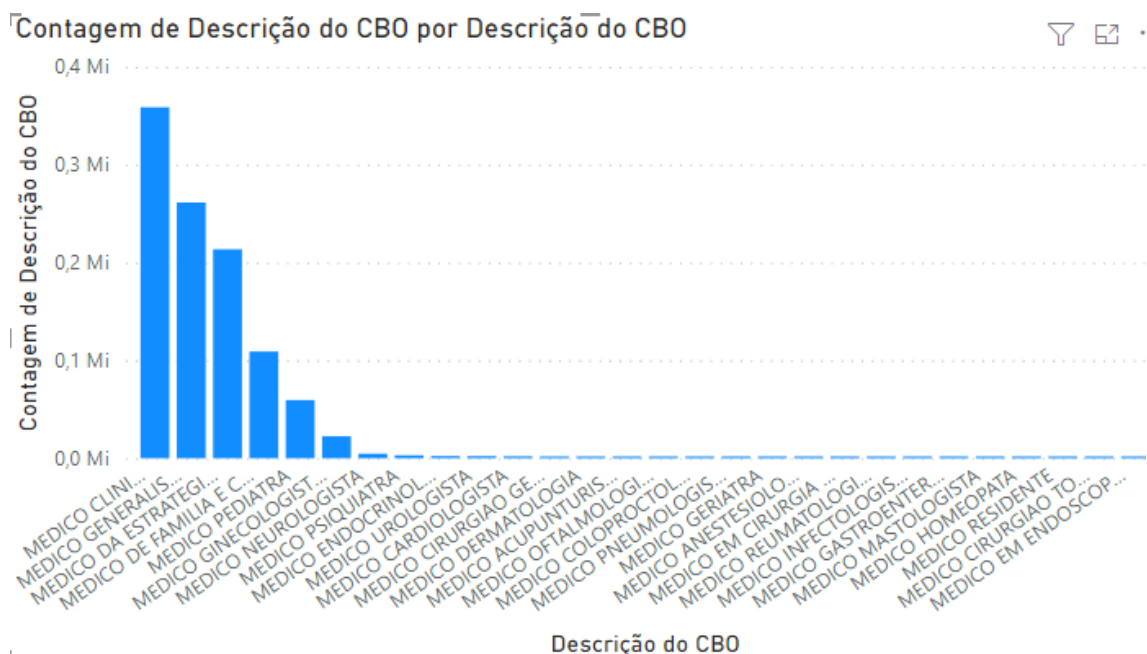


Figura 28: Perfil dos atendimentos médicos do SUS Curitiba. Dataset completo.

A figura 28 nos mostra que 90% dos atendimentos são de atenção básica, ou de baixa complexidade. 99,69% dos atendimentos não levaram a internamento e apenas 16% dos pacientes foram encaminhados para atendimento com especialista, ou média complexidade.

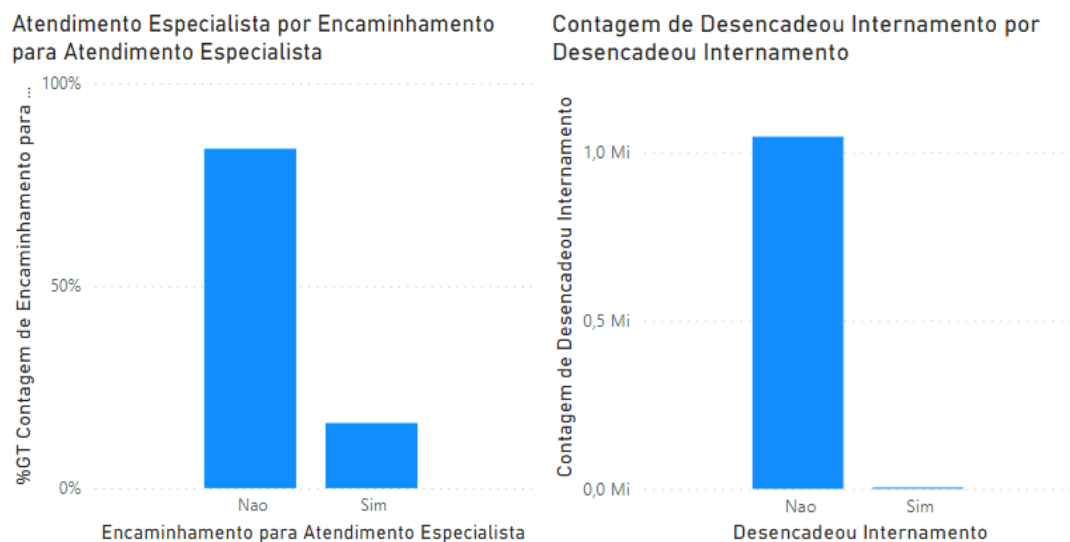


Figura 29: Perfil do encaminhamento e internamento no SUS Curitiba. Dataset completo.

Já as principais atividades desenvolvidas nas unidades de saúde de Curitiba são: Atenção primária com 58% dos atendimentos, pronto atendimento com 25% dos atendimentos, urgência e 24 horas com 7% dos atendimentos, telemedicina 3% dos atendimentos, atenção especializada 3% dos atendimentos, pré-natal 2% dos atendimentos, avaliação do desenvolvimento infantil 1% dos atendimentos, procedimentos administrativos, atendimento domiciliar e pós natal somam menos de 1%.

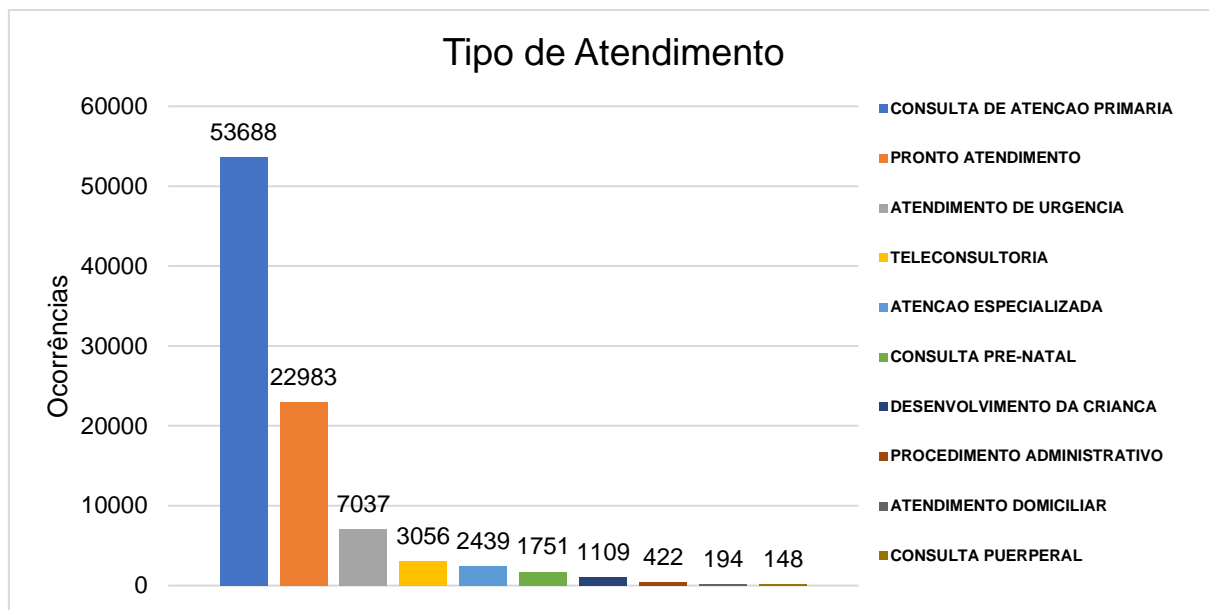


Figura 30: principais atividades fornecidas pela unidades de saúde em Curitiba

```
In [82]: contagem_p
```

```
Out[82]: CONSULTA MEDICA EM ATENCAO PRIMARIA 53688
          ATENDIMENTO MEDICO EM UNIDADE DE PRONTO ATENDIMENTO 22983
          ATENDIMENTO DE URGENCIA C/ OBSERVACAO ATE 24 HORAS EM ATENCAO ESPECIALIZADA 7037
          RESPOSTA TELERREGULAÇÃO/TELECONSULTORIA 3056
          CONSULTA MEDICA EM ATENCAO ESPECIALIZADA 2439
          CONSULTA PRE-NATAL 1751
          AVALIACAO DO DESENVOLVIMENTO DA CRIANCA NA PUERICULTURA 1109
          PROCEDIMENTO ADMINISTRATIVO 422
          CONSULTA/ATENDIMENTO DOMICILIAR 194
          CONSULTA PUERPERAL 148
          Name: Descrição do Procedimento, dtype: int64
```

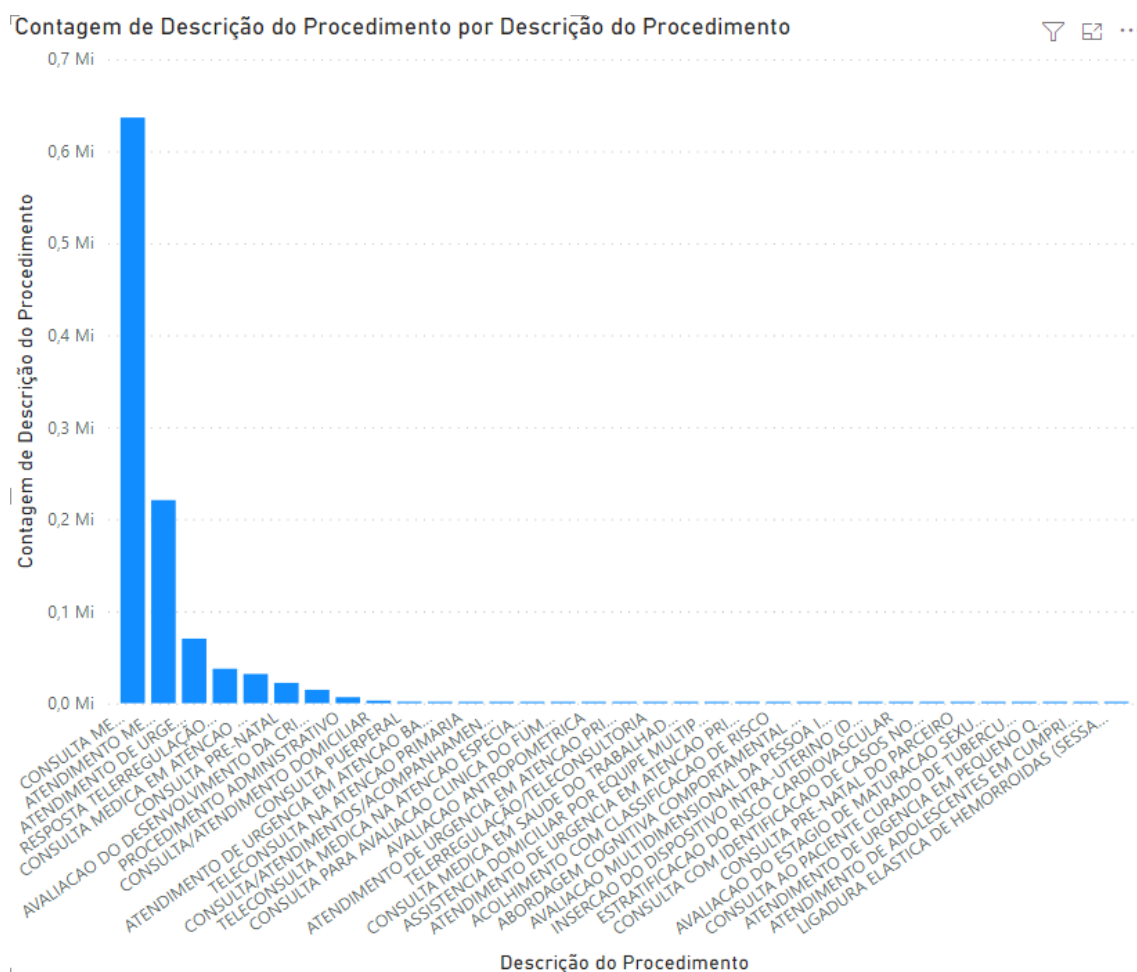


Figura 31: Perfil dos atendimentos por especialidade no SUS Curitiba. Dataset completo.

60% dos atendimentos são de atenção primária, 21% pronto atendimento, 6,7% de urgência e emergência, mantendo o perfil mostrado na etapa I. Na próxima figura, figura 32, mostra os dados sobre razão pela qual os pacientes buscam os serviços do SUS. São 5416 condições relatadas, a partir da condição 16 todas contam com menos de 1% do total observado. 25% dos atendimentos são realizados por agendamento prévio, dos não agendados, a maior proporção (5%) está relacionado com eventos cardiovasculares, hipertensão principalmente.

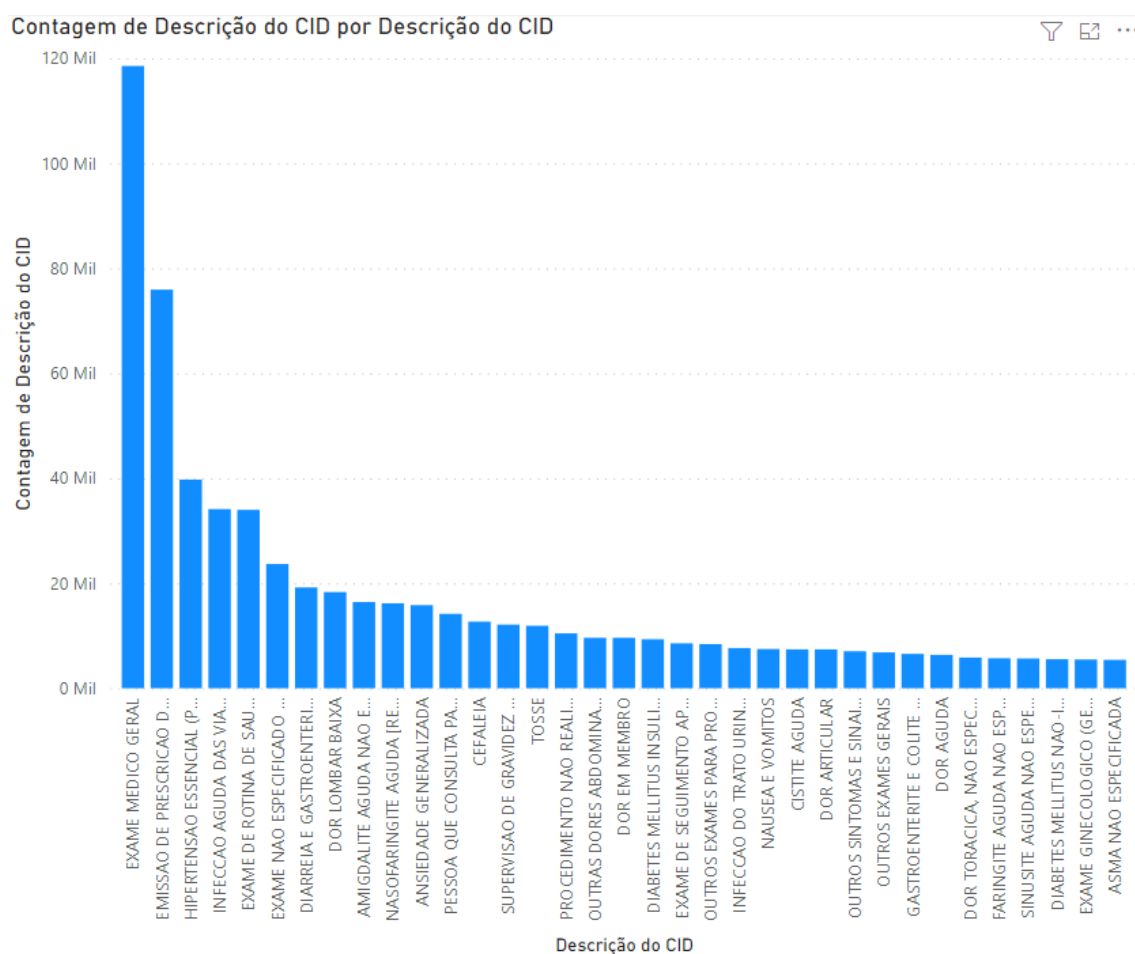


Figura 32: Razão por buscar o SUS Curitiba. Dataset completo.

### 6.3. MODELO DE MACHINE LEARNING – MÉTODO DE CLASSIFICAÇÃO ATRAVÉS DE RANDOM FOREST

Para essa etapa, foram utilizados as bibliotecas pandas, matplotlib.pyplot e scikit-learn. O modelo escolhido foi de classificação através do RandomForest. Os pormenores do modelo estão no documento “Tomazini\_ER\_S8\_OficinaMaker.html” entregue junto à esse relatório.

#### 6.4. ANÁLISE PYSPARK VIA ZEPPELIN/DOCKER

Para a análise em BigData, foi utilizado um container Docker para o Zeppelin notebooks, onde a análise foi feita via pyspark, diretamente utilizando a linguagem Scala. O acesso é feito via prompt de comando, através de:

```
docker run -p 8080:8080 -p 4040:4040 --rm -v c:/zeppelin_example/logs:/logs -v  
c:/zeppelin_example/notebook:/opt/zeppelin/notebook -v  
c:/zeppelin_example/files:/tmp/files -v c:/spark-3.1.3-bin-hadoop3.2:/opt/spark -e c:/spark-  
3.1.3-bin-hadoop3.2=/opt/spark --name zeppelin apache/zeppelin:0.10.0
```

Onde o arquivo “dadosClean.csv” foi adicionado na pasta “temp” do diretório que criei na unidade C. Serão anexados os notebooks em formato zeppelin e py-notebook, porém infelizmente não vem completos como o formato HTML que costumo anexar nesses relatórios, por isso na sessão “ANEXOS”, serão anexados “*printscreens*” das células como prova de trabalho.

A parte mais notável quando utilizamos pyspark, especialmente através da linguagem scala, é a velocidade de processamento e a facilidade que o framework consegue executar os códigos. O que tomava minutos através do ambiente python, agora é executado em poucos segundos e gera resultados confiáveis.

Nessa etapa, foi possível confirmar os números observados no PowerBI e também desambiguar algumas colunas que o PowerBI não consegue mostrar eficientemente, como os dados referentes a farmácia curitibana.

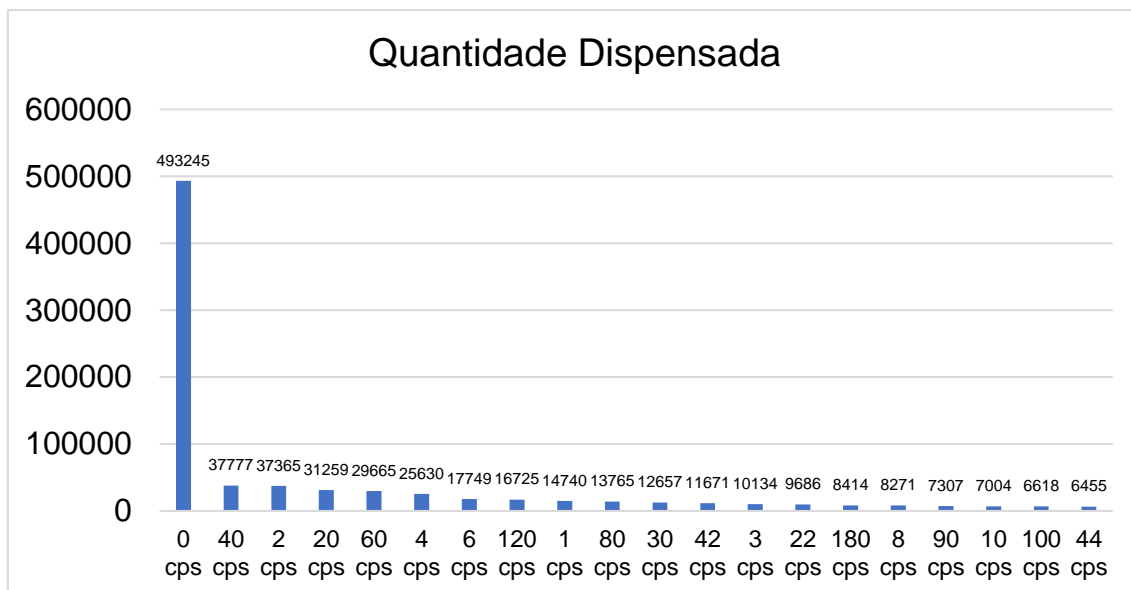


Figura 33: Quantidade dispensada de medicamentos pela farmácia Curitibana. dataset completo.

Podemos observar que 61% dos pacientes do SUS não recebem medicamentos da farmácia curitibana.

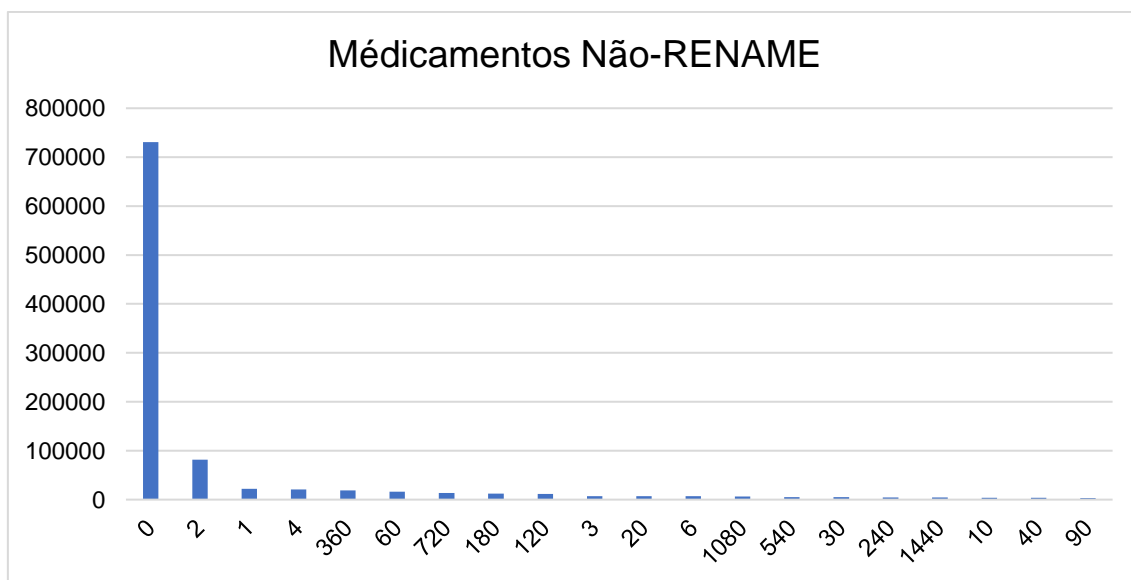


Figura 34: Medicamentos fora do RENAME entregues pelo SUS Curitiba. Dados Completos.

Já para os medicamentos fora do RENAME, lista padronizada do município de Curitiba (que pode ser nacional, estadual ou municipal, conforme a unidade

em questão), 74% dos pacientes não recebem medicamentos e apenas 10% dos pacientes chegam a receber medicamentos para 1 mês ou mais de tratamento.

Os últimos dados que não foi possível verificar satisfatoriamente através das outras ferramentas foi o número de cômodos dos pacientes do SUS Curitiba.



Figura 35: Quantidade de cômodos nas casas dos pacientes do SUS Curitiba. Dados completos.

Pode ser observado que 41% dos pacientes do SUS Curitiba contam com 5 cômodos (sala, cozinha, banheiro e 2 quartos), representando a moda e mediana, com média 5,5.

## 7. CONCLUSÃO

Neste extenso projeto, feito sobre o Sistema Único de Saúde (SUS) em Curitiba, diversas facetas do atendimento médico foram analisadas, proporcionando uma compreensão abrangente do panorama de saúde pública na região. Com um olhar sobre mais de um milhão de registros, tornou-se evidente que o SUS desempenha um papel vital na oferta de atendimento médico à população curitibana, enquanto conta com algumas falhas de demanda.

Uma observação inicial revelou que, embora haja uma participação equilibrada entre os sexos nos atendimentos, é notável a predominância de



consultas para o sexo feminino. Esse achado pode ser crucial para o desenvolvimento de estratégias específicas voltadas à saúde masculina. Quanto a nacionalidade dos pacientes, indicou uma forte representação de brasileiros natos, porém com boa participação de estrangeiros, inclusive acima da proporção que representam da população local, indicando que o SUS em Curitiba cumpre efetivamente seu princípio de universalidade. Entretanto, é intrigante observar uma super representação dessa categoria, indicando uma necessidade de investigar a composição dessa amostra.

Quanto à distribuição geográfica, a concentração de atendimentos em unidades básicas de saúde, seguida por unidades de pronto atendimento, destaca a importância do SUS como provedor de cuidados primários. Essa distribuição é relevante para orientar a alocação de recursos e aprimorar a eficiência do sistema, visto que vários bairros, e até mesmo UPA específicas, acumulam atendimento de outras regionais principalmente os bairros Cajuru e Sitio Cercado, que atendem outros bairros completamente, indicando estratégia de promoção da saúde na região ineficaz.

Aprofundando-se nos detalhes das condições de vida dos pacientes, foram analisados fatores como tipo de habitação, tratamento de água e acesso a serviços básicos. A presença de dados relacionados ao saneamento básico revelou desafios persistentes, destacando a importância de políticas públicas focadas nessa área, visto que aparentemente os dados tratam dos primeiros 2 quartis em relação a renda, visto condições de moradia e saneamento básico abaixo da média do município.

A investigação sobre o transporte dos pacientes ressaltou a dependência significativa do transporte público, sinalizando desafios de acessibilidade. A análise dos bairros mais atendidos e das unidades de saúde com maior demanda aponta para a necessidade de uma distribuição equitativa dos recursos para garantir atendimento eficiente em toda a cidade.

Ao examinar a atuação dos profissionais de saúde, identificou-se presença muito marcante de profissionais “residentes”, muitas vezes não qualificados para o atendimento, levando a uma concentração considerável de prescrições em um número limitado de médicos, indicando possíveis áreas de otimização e alocação mais equitativa de carga de trabalho. No âmbito clínico, a predominância de atendimentos de baixa complexidade e a ênfase na atenção

primária corroboram a natureza preventiva do SUS, mas poderia ser um indicativo da alta carga de profissionais pouco capacitados. Além disso, a prevalência de atendimentos agendados destaca a importância da organização prévia para otimizar os recursos disponíveis, embora as teleconsultas, que poderia representar uma grande fator de melhora na velocidade das consultas, ainda é sub representado.

Por fim, a análise dos serviços farmacêuticos revelou padrões de distribuição e uso de medicamentos, proporcionando insights valiosos para aprimorar a gestão e a entrega desses recursos essenciais. É notório que a grande maioria dos usuários não tem acesso aos medicamentos mesmo quando presentes no RENAME, indicando incapacidade de gestão.

Em síntese, esta análise aprofundada do dataset do SUS em Curitiba forneceu uma visão abrangente do sistema de saúde, destacando áreas de sucesso e identificando oportunidades de melhoria. Essas informações são cruciais para orientar políticas de saúde pública, otimizar recursos e, em última instância, proporcionar um atendimento mais eficiente e abrangente à população.

## REFERÊNCIAS

AMAZON. Apache Hadoop no Amazon EMR. **AWS**. Retirado de: <https://aws.amazon.com/pt/elasticmapreduce/details/hadoop/>

APACHE. Documentação do Apache Hadoop 3.3.6. **Apache**. 2023. Retirado de: <https://spark.apache.org/docs/latest/rdd-programming-guide.html>

APACHE. Documentação do Apache Spark 3.3.1. **Apache**. 2022. Retirado de: <https://spark.apache.org/docs/latest/rdd-programming-guide.html>

APACHE. Documentação do Apache Zeppelin 0.8.0. **Apache**. 2023. Retirado de: <https://zeppelin.gg/docs/introduction>

BRASIL. Lei 8080 de 19 de setembro de 1990. **Justiça brasileira**. 1990. Retirado de: [https://www.planalto.gov.br/ccivil\\_03/leis/l8080.htm](https://www.planalto.gov.br/ccivil_03/leis/l8080.htm)

BRASIL. Lei 8142 de 28 de dezembro de 1990. **Justiça brasileira**. 1990. Retirado de:

[https://www.planalto.gov.br/ccivil\\_03/leis/l8142.htm#:~:text=L8142&text=LEI%20N%C2%BA%208.142%2C%20DE%2028%20DE%20DEZEMBRO%20DE%201990.&text=Disp%C3%B5e%20sobre%20a%20participa%C3%A7%C3%A3o%20da,sa%C3%BAde%20e%20d%C3%A1%20outras%20provid%C3%AAs](https://www.planalto.gov.br/ccivil_03/leis/l8142.htm#:~:text=L8142&text=LEI%20N%C2%BA%208.142%2C%20DE%2028%20DE%20DEZEMBRO%20DE%201990.&text=Disp%C3%B5e%20sobre%20a%20participa%C3%A7%C3%A3o%20da,sa%C3%BAde%20e%20d%C3%A1%20outras%20provid%C3%AAs).

CURITIBA. Estrangeiros em Curitiba no ano de 2023. Retirado de: <https://www.curitiba.pr.gov.br/noticias/curitiba-abriga-estrangeiros-das-mais-diversas-nacionalidades/38093>

DEVMEDIA. Introdução ao apache spark. **Devmedia**. 2022. Retirado de: <https://www.devmedia.com.br/introducao-ao-apache-spark/34178>

IBGE. Censo Demografico 2022 - Microrregiões, distritos, subdistritos e bairros. Retirado de: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html>

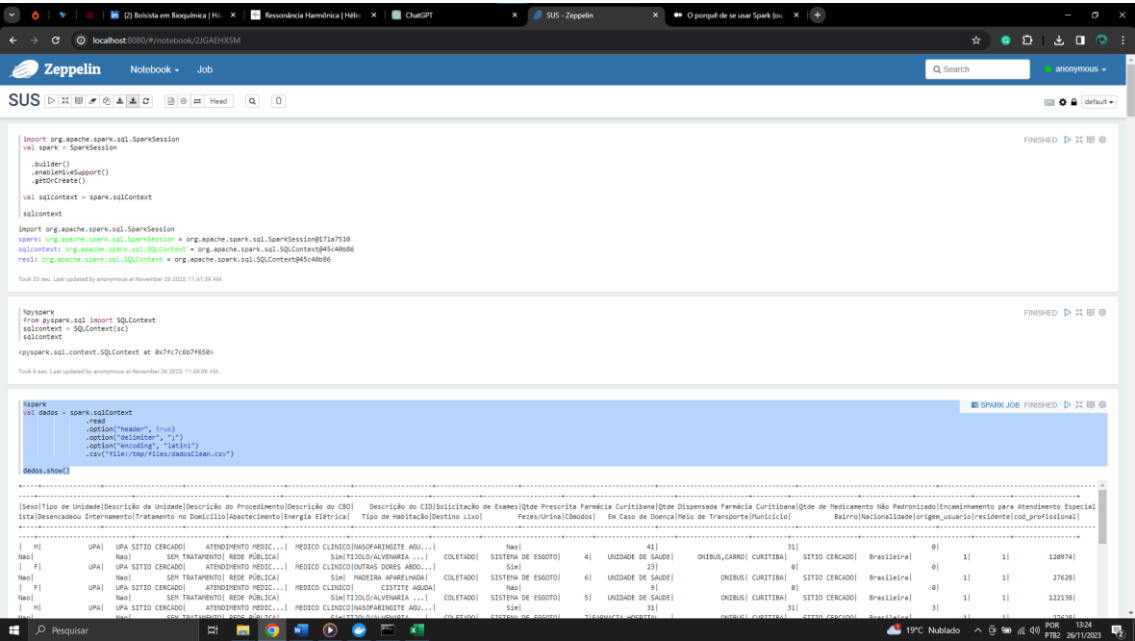
MICROSOFT. Introduction to PowerBI. Retirado de: <https://powerbi.microsoft.com/en-gb/getting-started-with-power-bi/?accordion=getting-started-with-powerbi-accordion&panel=p1&tab=t1>

MONTEIRO, W. O porquê de se utilizar o Spark. **Medium**. Retirado de: <https://medium.com/@wmonteiro/o-porqu%C3%AA-de-se-usar-spark-ou-n%C3%A3o-feat-zeppelin-b05f82ce867b>

SILVEIRA, F. L. Teorema Central do Limite: exemplificações utilizando o Método de Monte Carlo. **IF-UFRGS**. Retirado de: [https://www.if.ufrgs.br/~lang/Textos/Teorema\\_Central\\_Limite.pdf](https://www.if.ufrgs.br/~lang/Textos/Teorema_Central_Limite.pdf)

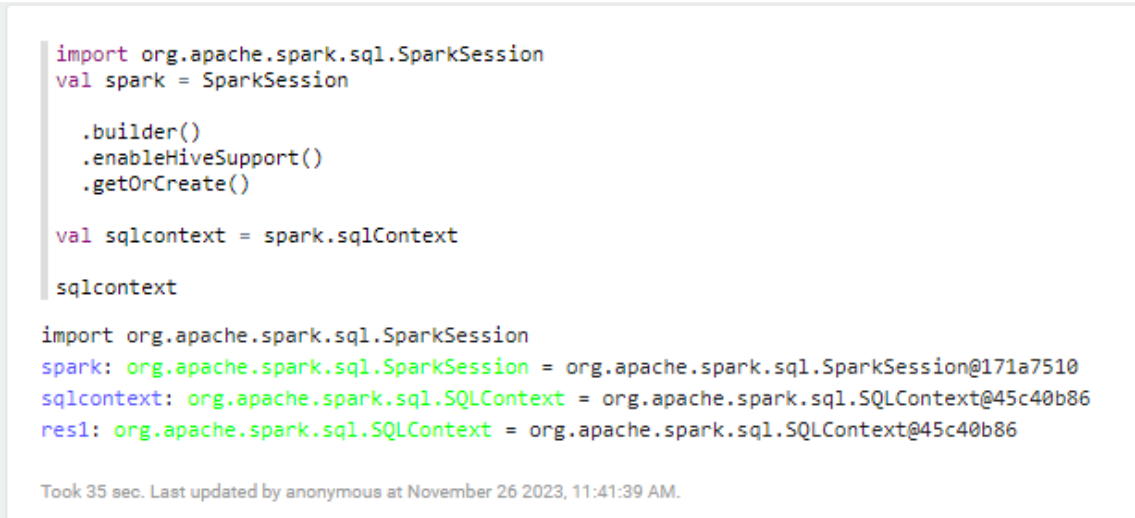
# ANEXOS

## Prova de trabalho:



## CONFIGURAÇÕES

### Importação e SparkSession



### SparkContext

```
%pyspark
from pyspark.sql import SQLContext
sqlcontext = SQLContext(sc)
sqlcontext
```

```
<pyspark.sql.context.SQLContext at 0x7fc7c6b7f650>
```

Took 6 sec. Last updated by anonymous at November 26 2023, 11:48:06 AM.

## Importação da Tabela dadosClean.csv

```
%spark
val dados = spark.sqlContext
    .read
    .option("header", true)
    .option("delimiter", ";")
    .option("encoding", "latin1")
    .csv("file:/tmp/files/dadosClean.csv")

dados.show()
```

Sexo	Tipo de Unidade	Descrição da Unidade	Descrição do Procedimento	Descrição do CBO	Descrição do CID	Solicitação de Exames	Qtde Prescrita Farma	Desencadeou Internamento	Tratamento no Domicílio	Abastecimento	Energia Elétrica	Tipo de Habitação	Destino Lixo	Fezes/Urina	Cômodos
M	UPA	UPA SITIO CERCADO	ATENDIMENTO MEDIC...	MEDICO CLINICO	NASOFARINGITE AGU...	Nao									
Nao			SEM TRATAMENTO	REDE PÚBLICA	Sim	TIJOLO/ALVENARIA ...	COLETADO	SISTEMA DE ESGOTO	4						
F	UPA	UPA SITIO CERCADO	ATENDIMENTO MEDIC...	MEDICO CLINICO	OUTRAS DORES ABDO...	Sim									
Nao			SEM TRATAMENTO	REDE PÚBLICA	Sim	MADEIRA APARELHADA	COLETADO	SISTEMA DE ESGOTO	6						
F	UPA	UPA SITIO CERCADO	ATENDIMENTO MEDIC...	MEDICO CLINICO	CISTITE AGUDA	Nao									
Nao			SEM TRATAMENTO	REDE PÚBLICA	Sim	TIJOLO/ALVENARIA ...	COLETADO	SISTEMA DE ESGOTO	5						
M	UPA	UPA SITIO CERCADO	ATENDIMENTO MEDIC...	MEDICO CLINICO	NASOFARINGITE AGU...	Sim									
Nao			SEM TRATAMENTO	REDE PÚBLICA	Sim	TIJOLO/ALVENARIA ...	COLETADO	SISTEMA DE ESGOTO	7						
F	UPA	UPA TATUQUARA	ATENDIMENTO DE UR...	MEDICO CLINICO	ABSCESSE CUTANEO...	Sim									
Nao			SEM TRATAMENTO	REDE PÚBLICA	Sim	TIJOLO/ALVENARIA ...	COLETADO	DIRETO PARA RIO, ...	5						
F	UPA	UPA SITIO CERCADO	ATENDIMENTO MEDIC...	MEDICO CLINICO	DOR EM MEMBRO	Sim									

Took 1 sec. Last updated by anonymous at November 26 2023, 12:01:06 PM.

## Queries:

### Sexo do paciente

```
dados.groupBy("Sexo").count().orderBy($"count".desc).show()
```

```
+-----+
|Sexo| count|
+-----+
| F | 655844|
| M | 392731|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:09:55 PM.

### Tipo da Unidade básica

```
dados.groupBy("Tipo de Unidade").count().orderBy($"count".desc).show()
```

```
+-----+
|Tipo de Unidade| count|
+-----+
|          BASICO| 725943|
|           UPA| 291014|
|          SIACE| 31618|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:10:10 PM.

Qual foi o procedimento realizado?

```
dados.groupBy("Descrição do Procedimento").count().orderBy($"count".desc).show()
```

```
+-----+
|Descrição do Procedimento| count|
+-----+
| CONSULTA MEDICA E...| 636510|
| ATENDIMENTO MEDIC...| 220752|
| ATENDIMENTO DE UR...| 70170|
| RESPOSTA TELERREG...| 37454|
| CONSULTA MEDICA E...| 31774|
| CONSULTA PRE-NATAL| 22087|
| AVALIACAO DO DESE...| 14541|
| PROCEDIMENTO ADMI...| 6519|
| CONSULTA/ATENDIME...| 2768|
| CONSULTA PUERPERAL| 1643|
| ATENDIMENTO DE UR...| 774|
| TELECONSULTA NA A...| 762|
| CONSULTA/ATENDIME...| 493|
| TELECONSULTA MEDI...| 441|
| CONSULTA PRE-NAT...| 330|
+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:10:19 PM.

Qual foi o código internacional de doença

```
dados.groupBy("Descrição do CID").count().orderBy($"count".desc).show()
```

```
+-----+
| Descrição do CID | count |
+-----+
| EXAME MEDICO GERAL | 118554 |
| EMISSAO DE PRESCR... | 75943 |
| HIPERTENSAO ESSEN... | 39759 |
| INFECCAO AGUDA DA... | 34134 |
| EXAME DE ROTINA D... | 33996 |
| EXAME NAO ESPECIF... | 23674 |
| DIARREIA E GASTRO... | 19214 |
| DOR LOMBAR BAIXA | 18333 |
| AMIGDALITE AGUDA ... | 16429 |
| NASOFARINGITE AGU... | 16193 |
| ANSIEDADE GENERAL... | 15836 |
| PESSOA QUE CONSUL... | 14185 |
| CEFALEIA | 12697 |
| SUPERVISAO DE GRA... | 12130 |
| ... | ... |
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:10:29 PM.

## Qual bairro atendeu?

```
dados.groupBy("Bairro").count().orderBy($"count".desc).show()
```

```
+-----+
| Bairro | count |
+-----+
| CIDADE INDUSTRIAL | 144801 |
| SITIO CERCADO | 88936 |
| CAJURU | 63414 |
| TATUQUARA | 51952 |
| UBERABA | 43887 |
| BOQUEIRAO | 41415 |
| PINHEIRINHO | 36789 |
| CAMPO DE SANTANA | 32850 |
| ALTO BOQUEIRAO | 32809 |
| XAXIM | 32403 |
| BAIRRO ALTO | 29344 |
| SANTA CANDIDA | 27140 |
| NOVO MUNDO | 27008 |
| CAPAO RASO | 24610 |
| ... | ... |
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:12:02 PM.

## Nacionalidade dos pacientes do SUS

```
dados.groupBy("Nacionalidade").count().orderBy($"count".desc).show()
```

```
+-----+
|Nacionalidade| count|
+-----+
|   Brasileira|1013566|
|   Estrangeiro|  22321|
|Não informado|  12570|
| Naturalizado|   118|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:12:07 PM.

## Atendimento por especialista

```
dados.groupBy("Encaminhamento para Atendimento Especialista").count().orderBy($"count".desc).show()
```

```
+-----+
|Encaminhamento para Atendimento Especialista| count|
+-----+
|                                             Nao|879705|
|                                             Sim|168870|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:12:13 PM.

## Descrição do tipo de consulta

```
dados.groupBy("Descrição do CBO").count().orderBy($"count".desc).show()
```

```
+-----+
|  Descrição do CBO| count|
+-----+
|      MEDICO CLINICO|358656|
|  MEDICO GENERALISTA|261324|
|MEDICO DA ESTRATE...|213486|
|MEDICO DE FAMILIA...|109085|
|      MEDICO PEDIATRA| 59431|
|MEDICO GINECOLOGI...| 22431|
| MEDICO NEUROLOGISTA|  4598|
|  MEDICO PSIQUIATRA|  2947|
|MEDICO ENDOCRINOL...|  2155|
|  MEDICO UROLOGISTA|  2081|
|MEDICO CARDIOLOGISTA|  1862|
|MEDICO CIRURGIAO ...|  1666|
| MEDICO DERMATOLOGIA|  1362|
|MEDICO ACUPUNTURISTA|  1154|
|MEDICO OTO-RINO-LA...|   1001|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:14:12 PM.

## Consulta primária levou a especialista?



```
dados.groupBy("Encaminhamento para Atendimento Especialista").count().orderBy($"count".desc).show()
```

```
+-----+
|Encaminhamento para Atendimento Especialista| count|
+-----+
|                                           Nao|879705|
|                                           Sim|168870|
+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:13:23 PM. (outdated)

## Solicitação de exames?

```
dados.groupBy("Solicitação de Exames").count().orderBy($"count".desc).show()
```

```
+-----+
|Solicitação de Exames| count|
+-----+
|                   Sim|782904|
|                   Nao|265671|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:15:14 PM.

## Medicamentos dispensados no SUS

```
dados.groupBy("Qtde Dispensada Farmácia Curitibaana").count().orderBy($"count".desc).show()
```

```
+-----+
|Qtde Dispensada Farmácia Curitibaana| count|
+-----+
|                                0|493245|
|                                40| 37777|
|                                2| 37365|
|                                20| 31259|
|                                60| 29665|
|                                4| 25630|
|                                6| 17749|
|                                120| 16725|
|                                1| 14740|
|                                80| 13765|
|                                30| 12657|
|                                42| 11671|
|                                3| 10134|
|                                22|  9686|
|                                100|  8111|
+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:16:12 PM.

## Medicamentos fora do RENAME

Qtde de Medicamento Não Padronizado	count
0	731028
2	81652
1	22357
4	20983
360	18668
60	16049
720	13284
180	12000
120	11935
3	7192
20	6965
6	6802
1080	6506
540	5149
30	5030

Took 3 sec. Last updated by anonymous at November 26 2023, 12:16:56 PM.

Tratamento adicional da água para consumo?

```
dados.groupBy("Tratamento no Domicílio").count().orderBy($"count".desc).show()
```

```

+-----+-----+
| Tratamento no Domicílio | count |
+-----+-----+
| SEM TRATAMENTO | 490323 |
| CLORACAO | 224696 |
| FILTRACAO | 205996 |
| null | 123537 |
| FERVURA | 4023 |
+-----+-----+

```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:18:05 PM.

De onde vem a água que consome?

```
dados.groupBy("Abastecimento").count().orderBy($"count".desc).show()
```

```
+-----+-----+
| Abastecimento| count|
+-----+-----+
| REDE PÚBLICA|906071|
|           null|124228|
|           OUTROS| 16126|
|POÇO ARTESIANO|  1696|
|           CISTERNA|   246|
|           CARRO PIPA|   208|
+-----+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:18:34 PM.

## Energia elétrica em casa?

```
dados.groupBy("Energia Elétrica").count().orderBy($"count".desc).show()
```

```
+-----+
|Energia Elétrica| count|
+-----+
|              Sim|975668|
|              Nao| 72907|
+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:19:08 PM.

## Tipo de habitação

```
dados.groupBy("Tipo de Habitação").count().orderBy($"count".desc).show()
```

```
+-----+
|Tipo de Habitação| count|
+-----+
|TIJOLO/ALVENARIA ...|825482|
|              null|123623|
|MADEIRA APARELHADA| 56355|
|TIJOLO ALVENARIA ...| 27163|
|OUTRO MATERIAL| 10809|
|MATERIAL APROVEITADO| 3116|
|TAIPA COM REVESTI...| 1439|
|TAIPA SEM REVESTI...|  404|
|              PALHA |   184|
+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:19:34 PM.

## Coleta de lixo?

```
dados.groupBy("Destino Lixo").count().orderBy($"count".desc).show()
```

```
+-----+
|Destino Lixo| count|
+-----+
|COLETADO|911315|
|      null|123875|
|OUTROS| 9662|
|CÉU ABERTO| 2456|
|QUEIMADO/ENTERRADO| 1267|
+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:20:05 PM.

## Foi encaminhados para especialista?

```
dados.groupBy("Encaminhamento para Atendimento Especialista").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|Encaminhamento para Atendimento Especialista| count|
+-----+-----+
|                                           Nao|879705|
|                                           Sim|168870|
+-----+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:20:47 PM.

## Tratamento de esgoto?

```
dados.groupBy("Fezes/Urina").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|          Fezes/Urina| count|
+-----+-----+
|  SISTEMA DE ESGOTO|864152|
|              null|124492|
|    FOSSA SÉPTICA| 34493|
|          OUTROS| 14075|
|      CÉU ABERTO| 10306|
|DIRETO PARA RIO, ...|   785|
|    FOSSA RUDIMENTAR|   272|
+-----+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:21:46 PM. (outdated)

## Onde é atendido em caso de doença?

```
dados.groupBy("Em Caso de Doença").count().orderBy($"count".desc).show()
```

```
+-----+-----+
| Em Caso de Doença | count |
+-----+-----+
| UNIDADE DE SAUDE  | 449999 |
| UNIDADE DE SAUDE,... | 147960 |
| HOSPITAL,UNIDADE ... | 139804 |
| null              | 123940 |
| HOSPITAL          | 61838  |
| OUTROS            | 48957  |
| FARMACIA,UNIDADE ... | 4914   |
| FARMACIA,HOSPITAL... | 4784   |
| OUTROS,UNIDADE DE... | 4746   |
| UNIDADE DE SAUDE,... | 4445   |
| HOSPITAL,FARMACIA... | 4317   |
| UNIDADE DE SAUDE,... | 4290   |
| UNIDADE DE SAUDE,... | 4140   |
| HOSPITAL,UNIDADE ... | 3864   |
| OUTROS,UNIDADE DE... | 3754   |
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:22:21 PM.

## Meio de transporte utilizado

```
dados.groupBy("Meio de Transporte").count().orderBy($"count".desc).show()
```

```
+-----+-----+
| Meio de Transporte | count |
+-----+-----+
| ONIBUS             | 483038 |
| ONIBUS,CARRO       | 312766 |
| null               | 123912 |
| OUTROS             | 50006  |
| CARRO              | 41021  |
| OUTROS,ONIBUS,CARRO | 17507  |
| OUTROS,ONIBUS      | 7271   |
| ONIBUS,CAMINHÃO    | 4105   |
| ONIBUS,METRO,CARRO | 1680   |
| ONIBUS,CARRO,CAMI... | 1018   |
| OUTROS,CARRO       | 1015   |
| CAMINHÃO           | 833    |
| OUTROS,ONIBUS,CAR... | 740    |
| METRO              | 652    |
| OUTROS,ONIBUS,MET... | 531    |
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:22:35 PM.

## Quantidades de cômodos em casa

```
dados.groupBy("Cômodos").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|Cômodos| count|
+-----+-----+
|      5|412336|
|      6|206523|
|      4|154496|
|      3| 59451|
|      7| 57196|
|      8| 41165|
|    null| 39861|
|      0| 21496|
|      2| 17425|
|     10| 12893|
|      9| 10875|
|      1|  5535|
|     12|  3046|
|     11|  1958|
|     13|  1000|
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:23:18 PM.

## Atendimentos por município de Curitiba

```
dados.groupBy("Município").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|      Município| count|
+-----+-----+
|      CURITIBA|1024001|
|      COLOMBO|  7265|
|      PINHAIS|  3650|
|ALMIRANTE TAMANDARÉ|  3023|
|FAZENDA RIO GRANDE|  2947|
|      PIRAQUARA|  1596|
|SAO JOSE DOS PINHAIS|  1351|
|      CAMPO LARGO|  1030|
|      ARAUCARIA|   942|
|      CAMPO MAGRO|   444|
|      CASCAVEL|   217|
|CAMPINA GRANDE DO...|   193|
|      RIO BRANCO DO SUL|   137|
|      ITAPERUCU|   115|
|      MARIPÓLIS|   100|
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:23:38 PM.

## Atendimentos por Bairro

```
dados.groupBy("Bairro").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|      Bairro| count|
+-----+-----+
|CIDADE INDUSTRIAL|144801|
|  SITIO CERCADO| 88936|
|      CAJURU| 63414|
|    TATUQUARA| 51952|
|    UBERABA| 43887|
|    BOQUEIRAO| 41415|
|  PINHEIRINHO| 36789|
|CAMPO DE SANTANA| 32850|
|  ALTO BOQUEIRAO| 32809|
|      XAXIM| 32403|
|    BAIRRO ALTO| 29344|
|  SANTA CANDIDA| 27140|
|    NOVO MUNDO| 27008|
|    CAPAO RASO| 24610|
|  CILLO BRASILEIRO| 23600|
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:24:00 PM.

## Atendimentos por nacionalidade

```
dados.groupBy("Nacionalidade").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|Nacionalidade| count|
+-----+-----+
|  Brasileira|1013566|
|  Estrangeiro|  22321|
|Não informado| 12570|
| Naturalizado|   118|
+-----+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:24:50 PM. (outdated)

Paciente interno n == 1 | externo n == 2;

```
dados.groupBy("origem_usuario").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|origem_usuario| count|
+-----+-----+
|              1|1024001|
|              2| 24574|
+-----+-----+
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:25:09 PM.

Atendimentos feitos por residentes (n == 1)

```
dados.groupBy("residente").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|residente| count|
+-----+-----+
|          1|980270|
|          2| 68305|
+-----+-----+
```

Took 2 sec. Last updated by anonymous at November 26 2023, 12:25:40 PM. (outdated)

Profissionais com maior número de atendimentos:



```
dados.groupBy("cod_profissional").count().orderBy($"count".desc).show()
```

```
+-----+-----+
|cod_profissional|count|
+-----+-----+
|                |12652| 3540|
|                |43376| 3004|
|                |43581| 2915|
|                |90165| 2835|
|                |71091| 2762|
|                |3163 | 2688|
|                |122123| 2589|
|                |80303| 2540|
|                |72083| 2501|
|                |3371 | 2498|
|                |92050| 2428|
|                |98132| 2422|
|                |117997| 2351|
|                |78602| 2329|
|                |110440| 2311|
```

Took 3 sec. Last updated by anonymous at November 26 2023, 12:26:16 PM.