# Desafio para o processo seletivo – Consultor Data Science

## Sobre os dados: Diagnóstico de Tecnologia da Informação e Comunicação

O Diagnóstico de Tecnologia da Informação e Comunicação é uma pesquisa que ocorre anualmente para que o Órgão Central possa conhecer a situação enfrentada por cada Órgão Setorial pertencente ao Sistema Municipal de Tecnologia da Informação e Comunicação da Prefeitura de São Paulo. A última versão do questionário, que foi utilizado na edição de 2020 da pesquisa, possui 138 questões, de diversos tipos, e algumas com subquestões específicas.

Para este desafio, 5 questões dessa pesquisa foram selecionadas e separadas para que as respostas sejam trabalhadas e possam revelar algumas informações sobre a situação da TIC na Prefeitura de São Paulo.

## Sobre os arquivos:

Para a realização dessas atividades, estão sendo disponibilizados dois arquivos:

* Um dump da tabela que contém as respostas de 3 anos de realizações da pesquisa. A origem dessa tabela é um banco de dados MariaDB;
* Uma planilha contendo o dicionário de dados dos campos presentes nas tabelas.

## Instruções:

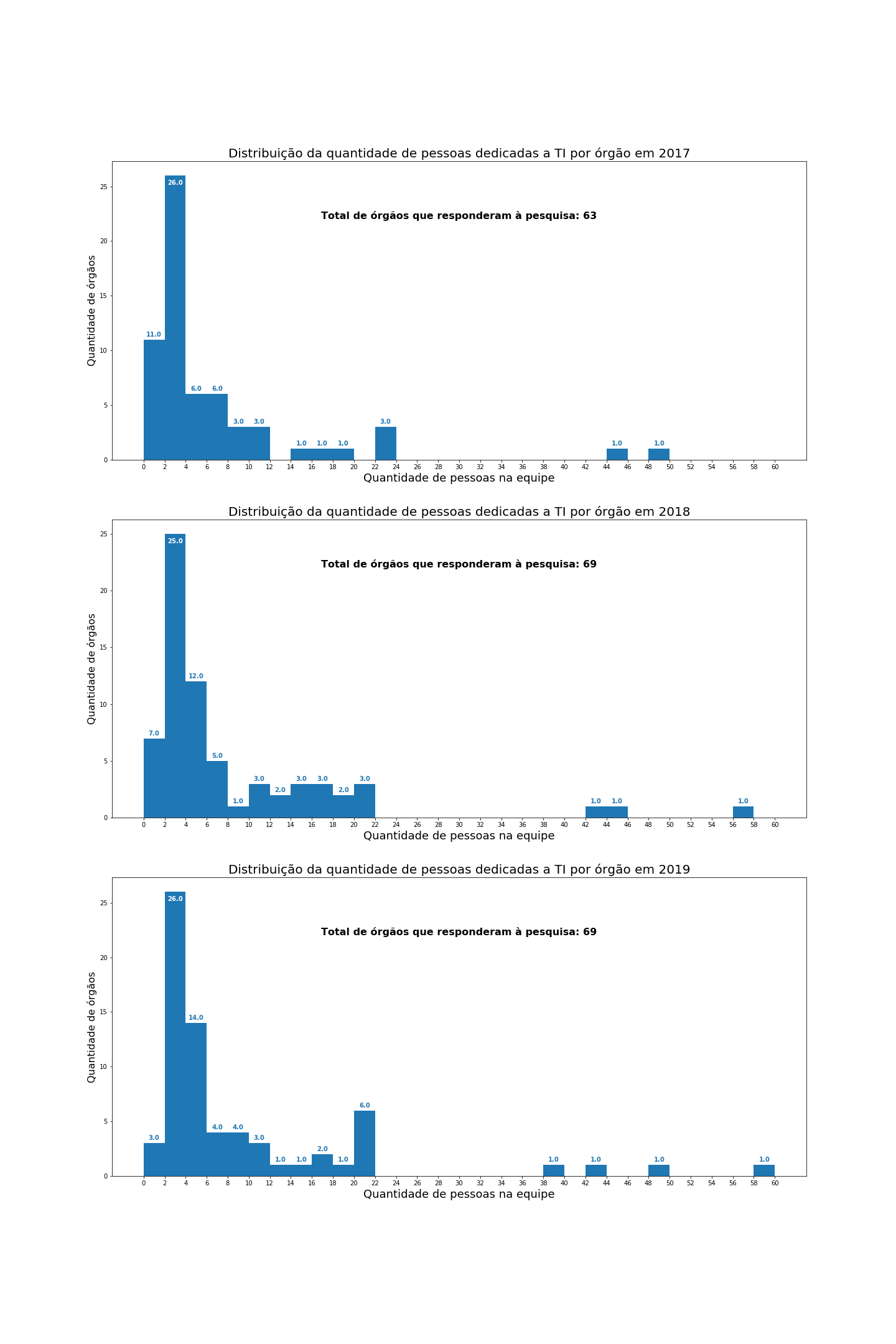
* No dicionário, alguns campos refletem os dados fiéis respondidos no questionário pelos diretores de TI de cada órgão. Quando pertinente, a coluna descrição descreve a pergunta que foi feita.
* Você deve ler os dados do dump e desenvolver as soluções usando **Python**. Para leitura das informações, você pode usar o sqlite3 embedded, que vem em conjunto com as novas versões do Python ou mesmo subir um banco local ou externo em MySQL/MariaDB[[1]](#footnote-1).
* Sugerimos o uso de repositórios públicos (como [GitHub](about:blank) ou [GitLab](about:blank)) e o envio da URL do repositório com o código criado em Python, mas vocês também podem zipar o código e enviar para o email ctd.dados@gmail.com; aos que optarem por outras ferramentas (Google Colab, Gitpod, PythonAnyware), é necessário que o código da solução seja compartilhada através de um link público.
* O teste é razoavelmente extenso e possui algumas tarefas de dificuldade um pouco mais elevada, então, fique tranquilo, não respondê-lo completamente não te desclassifica do processo seletivo. Resolva o máximo de problemas que conseguir.
* É importante ressaltar também que o campo *data\_submissao* é preenchido apenas quando o formulário é efetivamente concluído e aceito. Por isso, **considere somente as respostas que possuam o campo *data\_submissao* preenchido!**

## Desafio:

Seu grande desafio aqui será montar uma API Rest, em Python, usando as bibliotecas que desejar e consumindo o dump anexado.

Para as questões 1, 2, 3, 4 e 6, cada necessidade deve ser resolvida criando um endpoint na API **apenas** com os dados solicitados ou informados. A questão 7 é uma pergunta teórica que pode ser respondida no próprio formulário.

1. Gostaria de listar os órgãos que responderam a pesquisa, passando o ano referência do diagnóstico como parâmetro. Eles devem ser apresentados como uma lista de objetos e conter minimante o nome do órgão e o tipo de órgão.
2. Passando o órgão como parâmetro opcional e ano como parâmetros obrigatório na chamada, gostaria que saber quantas pessoas trabalharam de forma dedicada à TI na Prefeitura de São Paulo.
3. Considerando que todas as pessoas que trabalharam de forma dedicada a TI receberam R$ 12.500,00/mês, gostaria de saber qual a proporção de custo com pessoal de TI por tipo de órgão.
4. Gostaria de listar a quantidade de desktop próprios e desktop locados, por secretaria.
5. Criar uma tabela cópia da tabela respostas\_diagnostico apenas com os formulários que foram devidamente concluídos e aceitos. Incluir uma coluna do tipo datetime para gravar a data de última atualização de cada registro.
6. Criar um único endpoint para atualização das seguintes informações na tabela criada no item 5: Quantidade de pessoas que trabalham de forma dedicada, utilização de metodologia para gerenciamento de projetos, desktops próprios, locados e antigos.
7. **Questão teórica**: Considerando que os dados sobre quantidade de pessoas que trabalham dedicadas à Tecnologia da Informação no ano de 2019 apresenta a seguinte distribuição no ano de 2019, complete a frase que aparece abaixo dos histogramas:



OBS: Os intervalos do histograma são fechados à esquerda (ex: 0 até 1,999...; 2 até 3,999...; etc.)

* 1. "De acordo com o Diagnóstico de 2019, metade dos órgãos setoriais da prefeitura possui, no máximo, \_\_\_ pessoas trabalhando de forma dedicada à área de TI".

Metade dos Orgaos possui **no maximo 5 pessoas trabalhando na area de TI**. Para isso usamos a frequencia acumulada. Como 50% (69/2 = 34.5) esta entre 42% (29) e 62% (43) deve-se fazer a interpolacao entre 4 e 6 (5) pessoas. Notar ainda que na verdade a metade estaria em 52% (ou seja, um pouco acima de 5 pessoas). Como se tratam de dados discretos deve-se entao usar o inteiro mais proximo.

|  |  |  |  |
| --- | --- | --- | --- |
| Pessoas | Orgaos | Acum | % |
| 0 - 2 | 3 | 3 | 4% |
| 2 - 4 | 26 | 29 | 42% |
| 4 - 6 | 14 | 43 | 62% |
| 6 - 8 | 4 | 47 | 68% |
| 8 - 10 | 4 | 51 | 74% |
| 10 - 12 | 3 | 54 | 78% |
| 12 - 14 | 1 | 55 | 80% |
| 14 - 16 | 1 | 56 | 81% |
| 16 - 18 | 2 | 58 | 84% |
| 18 - 20 | 1 | 59 | 86% |
| 20 - 22 | 6 | 65 | 94% |
| 22 - 24 | 0 | 65 | 94% |
| 24 - 26 | 0 | 65 | 94% |
| 26 - 28 | 0 | 65 | 94% |
| 28 -30 | 0 | 65 | 94% |
| 30 - 32 | 0 | 65 | 94% |
| 32 - 34 | 0 | 65 | 94% |
| 34 - 36 | 0 | 65 | 94% |
| 36 - 38 | 0 | 65 | 94% |
| 38 - 40 | 1 | 66 | 96% |
| 40 - 42 | 0 | 66 | 96% |
| 42 - 44 | 1 | 67 | 97% |
| 44 - 46 | 0 | 67 | 97% |
| 46 - 48 | 0 | 67 | 97% |
| 48 - 50 | 1 | 68 | 99% |
| 50 - 52 | 0 | 68 | 99% |
| 52 -54 | 0 | 68 | 99% |
| 54 - 56 | 0 | 68 | 99% |
| 56 - 58 | 0 | 68 | 99% |
| 58 -60 | 1 | 69 | 100% |

# Anexo - Alterando o nome da base de dados no dump

Caso não seja possível utilizar o nome padrão da base de dados por questões de permissão do usuário conectado ao banco ou outro imprevisto, basta abrir o dump em um editor de texto e alterar as linhas 17 e 18, a seguir:

CREATE DATABASE IF NOT EXISTS `novo\_nome\_da\_base` /\*!40100 DEFAULT CHARACTER SET utf8 COLLATE utf8\_unicode\_ci \*/;  
USE `novo\_nome\_da\_base`;

A linha 17 também é opcional, caso a base de dados já exista ela pode ser excluída.

1. MariaDB é um sistema de banco de dados open source, baseado e compatível com o MySQL. Ele pode ser facilmente instalado através do XAMPP, que agrega servidores web e de banco de dados num mesmo instalador. Caso você não tenha a possibilidade de instalar uma instância em seu computador, existem alternativas (como [br.000webhost.com](about:blank), [www.freewebhostingarea.com](about:blank) ou [www.awardspace.com](about:blank)) onde se pode obter uma base de dados gratuita. Nesse caso, duas linhas de código do dump precisam ser alteradas. Confira o anexo para mais detalhes. [↑](#footnote-ref-1)