



Trabalho

Neste trabalho vamos manipular uma base de dados e extrair informações úteis dela usando Python. Ele poderá ser feito individualmente ou em dupla, **sem exceções**.

Você(s) terá(ão) que escolher uma de duas opções: encontrar uma base de dados por conta própria (que tenha ao menos três tabelas e que eu terei que aprovar previamente) ou então utilizar a base que eu estou disponibilizando, que contém dados do Censo Demográfico de 2010. Detalhes de cada tabela disponibilizada encontram-se na Seção 1.

Você(s) deverá(ão) desenvolver seu próprio *notebook* no Jupyter, e este será o único arquivo que deve ser entregue no Tidia com a sua solução, caso você(s) utilize(m) os dados do Censo. Se você(s) utilizou(aram) outra base de dados, deve(m) submetê-la juntamente com o seu notebook. Apenas um dos alunos da dupla precisa submeter os arquivos.

Para criar um novo notebook, basta clicar no botão *New* (ao lado do botão *Upload*) e em seguida escolher *Python 3*. Lembre-se que um notebook possui células de texto e células de código. Uma célula nova pode ser criada por meio do menu *Insert* e o seu tipo pode ser modificado no menu *Cell* (o tipo *code* mantém a célula como código e o tipo *markdown* transforma a célula no tipo texto).

A primeira célula do seu notebook deve obrigatoriamente conter o seu nome (e da sua dupla), o(s) seus(s) RA(s) e, caso você(s) use(m) uma base de dados diferente, uma descrição dos dados que ela está armazenando com um link de onde ela foi retirada.

O trabalho deve conter *pelo menos* um gráfico de cada um dos seguintes tipos: linha, barra, barra empilhada, histograma, pizza e boxplot. Você(s) pode(m) utilizar colunas já existentes nos arquivos **ou então criar colunas novas com outras informações interessantes (médias, porcentagens, etc.).** No caso dos dados do censo, utilize(m) um arquivo diferente para gerar cada gráfico. **Seus gráficos devem fazer sentido e por isso devem ser acompanhados de uma justifica/descrição simples sobre quais informações eles apresentam.**

Além dos gráficos, você(s) deve(m) elaborar e responder pelo menos 10 (dez) questões que envolvam cálculos de estatísticas como máximo, mínimo, média, mediana, variância e correlação. Você(s) pode utilizar informações dos gráficos anteriores ou então plotar novos gráficos para responder essas questões também.

Critérios de correção. O trabalho será avaliado em dois quesitos: corretude do código gerado (se seu código executa corretamente) e qualidade das informações obtidas sobre os dados (se seus gráficos e as questões fornecem informações relevantes sobre a base de dados). Isso significa que você(s) deve(m) ter como objetivo criar gráficos e questões que sejam bem representativos dos dados.

1 Descrição dos arquivos CSV disponíveis

Os censos populacionais produzem informações imprescindíveis para a definição de políticas públicas e a tomada de decisões de investimento, sejam eles provenientes da iniciativa privada ou de qualquer nível de governo, e constituem a única fonte de referência sobre a situação de vida da população nos municípios e em seus recortes internos, como distritos, bairros e localidades, rurais ou urbanas, cujas realidades dependem de seus resultados para serem conhecidas e terem seus dados atualizados.

O Censo 2010 compreendeu um levantamento minucioso de todos os domicílios do país. Nos meses de coleta de dados e supervisão, 191 mil recenseadores visitaram 67,6 milhões de domicílios nos 5.565 municípios brasileiros para colher informações sobre quem somos, quanto somos, onde estamos e como vivemos.

Fonte: IBGE.

De todas as tabelas disponíveis no site do IBGE, você irá lidar com cinco. Elas estão disponíveis para download no Tidia nos seguintes arquivos:

`pop_residente.csv` Descreve a nacionalidade dos indivíduos entrevistados.

`nivel_instrucao.csv` Descreve o nível de instrução dos indivíduos entrevistados.

`ocupacao.csv` Descreve a condição de atividade e ocupação dos indivíduos entrevistados.

`renda.csv` Descreve o rendimento mensal, em salários mínimos (SM), dos indivíduos entrevistados com mais de 10 anos.

`agua.csv` Descreve a existência de água canalizada e a forma de abastecimento de água nos domicílios dos indivíduos entrevistados.

Cada arquivo possui informações sobre **todos** os municípios que foram considerados no Censo. Você(s) poderá(ão) considerar todos os estados quando estiver(em) desenvolvendo o trabalho ou pode(m) focar em um subconjunto deles. Veja bem: você(s) pode(m) criar um gráfico que utilize dados do Brasil todo e ao mesmo tempo criar outro gráfico que utilize apenas dados de um subconjunto de estados. O ponto é que, se você(s) observar(em) que o gráfico não ficou tão informativo/legível com os dados do Brasil todo, talvez valha a pena utilizar um conjunto menor de estados.

As seções a seguir descrevem o conteúdo de cada coluna presente nos arquivos, de acordo com seu rótulo.

1.1 pop_residente.csv

Rótulo da coluna	Conteúdo
Estado	Nome de uma Unidade da Federação.
Município	Nome de um município/área de ponderação visitado.
Total Residente	Quantidade de residentes entrevistados naquele município.
Residente Brasileira	Dentre os residentes, quantidade de brasileiros entrevistados.
Brasileira Nata	Dentre os residentes brasileiros, quantidade dos que são brasileiros natos.
Brasileira por naturalização	Dentre os residentes brasileiros, quantidade dos que são brasileiros por naturalização.
Residente Estrangeira	Dentre os residentes, quantidade de estrangeiros entrevistados.
Código da Unidade Geográfica	Código identificador do município/área de ponderação.

1.2 nivel_instrucao.csv

Rótulo da coluna	Conteúdo
Estado	Nome de uma Unidade da Federação.
Município	Nome de um município/área de ponderação visitado.
Total pessoas > 10 anos	Quantidade de pessoas com mais de 10 anos entrevistadas naquele município.
Sem instrução e fundamental incompleto	Dentre os entrevistados, quantidade dos que não têm instrução ou têm ensino fundamental incompleto.
Fundamental completo e médio incompleto	Dentre os entrevistados, quantidade dos que têm ensino fundamental completo e ensino médio incompleto.
Médio completo e superior incompleto	Dentre os entrevistados, quantidade dos que têm ensino médio completo e ensino superior incompleto.
Superior completo	Dentre os entrevistados, quantidade dos que têm ensino superior completo.
Não determinado	Dentre os entrevistados, quantidade dos que não têm ensino determinado.
Código da Unidade Geográfica	Código identificador do município/área de ponderação.

1.3 ocupacao.csv

Rótulo da coluna	Conteúdo
Estado	Nome de uma Unidade da Federação.
Município	Nome de um município/área de ponderação visitado.
Total pessoas > 10 anos	Quantidade de pessoas com mais de 10 anos entrevistadas naquele município.
Economicamente ativas	Dentre os entrevistados, quantidade de pessoas economicamente ativas.
Ocupadas	Dentre as economicamente ativas, quantidade de pessoas ocupadas.
Desocupadas	Dentre as economicamente ativas, quantidade de pessoas desocupadas.
Não economicamente ativas	Dentre os entrevistados, quantidade de pessoas não economicamente ativas.
Total pessoas 10 a 13 anos	Dentre os entrevistados, quantidade de pessoas com 10 a 13 anos.
Economicamente ativas.1	Dentre as pessoas com 10 a 13 anos, quantidade de pessoas economicamente ativas.
Ocupadas.1	Dentre as pessoas com 10 a 13 anos economicamente ativas, quantidade de pessoas ocupadas.
Desocupadas.1	Dentre as pessoas com 10 a 13 anos economicamente ativas, quantidade de pessoas desocupadas.
Não economicamente ativas.1	Dentre as pessoas com 10 a 13 anos, quantidade de pessoas não economicamente ativas.
Total pessoas > 14 anos	Dentre os entrevistados, quantidade de pessoas com mais de 14 anos.
Economicamente ativas.2	Dentre as pessoas com mais de 14 anos, quantidade de pessoas economicamente ativas.
Ocupadas.2	Dentre as pessoas com mais de 14 anos economicamente ativas, quantidade de pessoas ocupadas.
Desocupadas.2	Dentre as pessoas com mais de 14 anos economicamente ativas, quantidade de pessoas desocupadas.
Não economicamente ativas.2	Dentre as pessoas com mais de 14 anos, quantidade de pessoas não economicamente ativas.
Código da Unidade Geográfica	Código identificador do município/área de ponderação.

1.4 renda.csv

Rótulo da coluna	Conteúdo
Estado	Nome de uma Unidade da Federação.
Município	Nome de um município/área de ponderação visitado.
Total pessoas > 10 anos	Quantidade de pessoas com mais de 10 anos entrevistadas naquele município.
Até 1 SM	Dentre os entrevistados, quantidade de pessoas que recebem até 1 salário mínimo.
De 1 a 2 SM	Dentre os entrevistados, quantidade de pessoas que recebem de 1 a 2 salários mínimos.
De 2 a 3 SM	Dentre os entrevistados, quantidade de pessoas que recebem de 2 a 3 salários mínimos.
De 3 a 5 SM	Dentre os entrevistados, quantidade de pessoas que recebem de 3 a 5 salários mínimos.
De 5 a 10 SM	Dentre os entrevistados, quantidade de pessoas que recebem de 5 a 10 salários mínimos.
De 10 a 20 SM	Dentre os entrevistados, quantidade de pessoas que recebem de 10 a 20 salários mínimos.
Mais de 20 SM	Dentre os entrevistados, quantidade de pessoas que recebem mais de 20 salários mínimos.
Sem rendimento	Dentre os entrevistados, quantidade de pessoas sem rendimento.
Código da Unidade Geográfica	Código identificador do município/área de ponderação.

1.5 agua.csv

Rótulo da coluna	Conteúdo
Estado	Nome de uma Unidade da Federação.
Município	Nome de um município/área de ponderação visitado.
Total domicílios	Quantidade total de domicílios visitados naquele município.
Total com água canalizada	Dos domicílios visitados, quantidade total que possuem água canalizada.
Água em pelo menos um cômodo	Dos domicílios com água canalizada, quantidade dos que têm água em pelo menos um cômodo.
Rede geral de distribuição	Dos domicílios com água canalizada em pelo menos um cômodo, quantidade dos que têm como forma de abastecimento a rede geral de distribuição.
Outra forma	Dos domicílios com água canalizada em pelo menos um cômodo, quantidade dos que têm outra forma de abastecimento.
Água no terreno ou propriedade	Dos domicílios com água canalizada, quantidade total que possuem água no terreno ou propriedade.
Rede geral de distribuição.1	Dos domicílios com água canalizada no terreno, quantidade dos que têm como forma de abastecimento a rede geral de distribuição.
Outra forma.1	Dos domicílios com água canalizada no terreno, quantidade dos que têm outra forma de abastecimento.
Não tinham	Dos domicílios visitados, quantidade total dos que não têm água canalizada.
Código da Unidade Geográfica	Código identificador do município/área de ponderação.