

Atividade 2 – Projeto relacionamento com clientes

Você é um cientista de dados de um Banco no Brasil. Na etapa de atendimento, existe um espaço para os clientes realizarem uma avaliação do processo. No banco de dados “avaliacoes_clientes.csv” temos algumas informações sobre acessos, resultados e avaliações.

Utilizar o arquivo checkpoint2_sem2.ipynb para obter os dados do Banco e carregar os dados do censo de 2010. Faça e responda as seguintes perguntas:

1 – Na tabela do censo de 2010, obter o índice de Gini por UF (state) (criar um dataframe como df_gini_estado).

2 – Realizar o join das tabelas de atendimento e df_gini_estado.

3 – Obter as distribuições de probabilidade (histograma) das variáveis listadas, depois informar qual das distribuições de probabilidade estudadas é mais similar:

- a) TempoAcesso;
- b) Gini.

4 – Assumindo que TempoAcesso (em segundos) e Gini tendem a seguir uma distribuição de NORMAL. Obtenha as seguintes probabilidades:

- a) $P(\text{Tempo Acesso} > 2 \text{ minutos})$
- b) $P(1,5 \text{ minutos} > \text{Tempo Acesso} > 2 \text{ minutos})$
- c) $P(\text{Tempo Acesso} < 0,5 \text{ minuto})$
- d) $P(\text{Gini} > 20\%)$
- e) $P(10\% > \text{Gini} > 15\%)$
- f) $P(\text{Gini} < 12\%)$

5 – Baseando-se na variável TempoAcesso, qual o tamanho de amostra para realizar um estudo com estes dados? Utilizar margem de erro de 15 segundos e nível de confiança de 95%.

6 – Após obter o tamanho de amostra, gerar dois novos *dataframes* conforme o seguinte:

- a) Criar um dataframe chamado df_aas usando amostra aleatória simples.
- b) Criar um dataframe chamado df_aae usando amostra aleatória estratificada. Utilizar como estrato a variável “avaliacao”.

7 – No dataframe df_aas realizar as seguintes análises:

- a) Comparar as medidas descritivas para TempoAcesso com o dataset original. Houve diferenças significativas?
- b) Comparar as medidas descritivas para Gini com o dataset original. Houve diferenças significativas?
- c) Obter as proporções da variável “avaliacao” e compare com o dataset original.
- d) Obter as medidas descritivas para TempoAcesso para cada perfil de “avaliacao” e compare com o dataset original.
- e) Obter as medidas descritivas para GINI para cada perfil de “avaliacao” e compare com o dataset original.

8 – No dataframe df_aae realizar as seguintes análises:

- a) Comparar as medidas descritivas para TempoAcesso com o dataset original. Houve diferenças significativas?
- b) Comparar as medidas descritivas para Gini com o dataset original. Houve diferenças significativas?
- c) Obter as proporções da variável “avaliacao” e compare com o dataset original.
- d) Obter as medidas descritivas para TempoAcesso para cada perfil de “avaliacao” e compare com o dataset original.
- e) Obter as medidas descritivas para GINI para cada perfil de “avaliacao” e compare com o dataset original.