

Atividade prática 2 - Produtividade Policial

Os dados

Considere os dados no arquivo `produtividade_policial.xlsx`. Estes dados foram obtidos da página de dados estatísticos (<http://www.ssp.sp.gov.br/Estatistica/Pesquisa.aspx>) do Secretaria de Segurança Pública de São Paulo para detalhes sobre a produtividade policial, registrados por mês ao longo de 2021.

Nesta tabela, é possível encontrar os seguintes indicadores para cada uma das regiões do estado de São Paulo:

Variáveis

OCORRÊNCIAS DE PORTE DE ENTORPECENTES

OCORRÊNCIAS DE TRÁFICO DE ENTORPECENTES

OCORRÊNCIAS DE APREENSÃO DE ENTORPECENTES(1)

OCORRÊNCIAS DE PORTE ILEGAL DE ARMA

Nº DE ARMAS DE FOGO APREENDIDAS

Nº DE FLAGRANTES LAVRADOS

Nº DE INFRATORES APREENDIDOS EM FLAGRANTE

Nº DE INFRATORES APREENDIDOS POR MANDADO

Nº DE PESSOAS PRESAS EM FLAGRANTE

Nº DE PESSOAS PRESAS POR MANDADO

Nº DE PRISÕES EFETUADAS

Nº DE VEÍCULOS RECUPERADOS

TOT. DE INQUÉRITOS POLICIAIS INSTAURADOS

(1) São computadas somente as ocorrências em que houve apenas apreensão de entorpecentes (sem pessoas envolvidas).

Manipulação dos dados

O código abaixo carrega as bibliotecas necessárias, carrega o arquivo de dados e exibe uma amostra de linhas da tabela.

```
library(tidyverse)
library(readxl)
library(stringi)
library(factoextra)
library(ggrepel)

dados <- read_xlsx("produtividade_policial.xlsx")

dados %>%
  sample_n(10)
```

Região	Ocorrencia	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
Santos	Nº DE PESSOAS PRESAS POR MANDADO	438	280	236	223	411	310	263	336	360	226	334	195	3612
Bauru	Nº DE FLAGRANTES LAVRADOS	379	342	400	384	387	385	342	417	372	328	302	328	4366

Região	Ocorrencia	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
São José do Rio Preto	OCORRÊNCIAS DE APREENSÃO DE ENTORPECENTES(1)	9	4	4	9	11	3	3	7	7	5	12	8	82
Presidente Prudente	Nº DE INFRATORES APREENSÃO POR MANDADO	4	8	8	5	8	5	1	3	3	6	9	7	67
Piracicaba	Nº DE PRISÕES EFETUADAS	928	921	1033	963	1052	955	972	980	997	882	874	966	11523
Sorocaba	TOT. DE INQUÉRITOS POLICIAIS INSTAURADOS	1819	2193	2065	1960	1941	2118	1966	2124	2008	2175	2066	2062	24497
Sorocaba	OCORRÊNCIAS DE PORTE ILEGAL DE ARMA	20	26	25	18	25	26	29	26	18	22	25	23	283
Presidente Prudente	OCORRÊNCIAS DE PORTE ILEGAL DE ARMA	5	6	7	4	14	7	16	8	17	11	7	8	110
Capital	Nº DE ARMAS DE FOGO APREENSÃO	197	193	241	209	339	193	184	169	173	734	168	271	3071
Grande São Paulo (exclui a Capital)	Nº DE INFRATORES APREENSÃO EM FLAGRANTE	140	148	169	160	146	153	143	140	156	108	142	121	1726

Para facilitar a manipulação dos dados, use o código abaixo para remover caracteres especiais do nome das colunas e deixá-los em letras minúsculas (a função `stri_trans_general` é do pacote `stringi`).

```
dados <- dados %>%
  rename_with(~ stri_trans_general(.x, "Latin-ASCII") %>%
    tolower())
```

Tarefas

1. Para este exercício, manipule os dados para considerar apenas os dados totais de cada indicador para cada região. Forneça o código que deixa os dados no seguinte formato:

regiao	OCORRÊNCIAS DE PORTE DE ENTORPECENTES	OCORRÊNCIAS DE TRÁFICO DE ENTORPECENTES	OCORRÊNCIAS DE APREENSÃO DE ENTORPECENTES(1)	OCORRÊNCIAS DE PORTE ILEGAL DE ARMA	Nº DE ARMAS DE FOGO APREENSÃO	Nº FLAGRANTE LAVRAD
Capital	1587	6237	483	824	3071	231
Grande São Paulo (exclui a Capital)	790	5917	736	610	1747	151

regiao	OCORRÊNCIAS DE PORTE DE ENTORPECENTES	OCORRÊNCIAS DE TRÁFICO DE ENTORPECENTES	OCORRÊNCIAS DE APREENSÃO DE ENTORPECENTES(1)	OCORRÊNCIAS DE PORTE ILEGAL DE ARMA	Nº DE ARMAS DE FOGO APREENDIDAS	Nº FLAGRANT LAVRAD
São José dos Campos	627	2356	92	417	886	4:
Campinas	1260	3502	741	332	972	7
Ribeirão Preto	3208	4284	581	444	1125	8:
Bauru	814	1890	465	183	586	4:

2. Realize o procedimento para obter as componentes principais deste conjunto de dados. Quantas componentes principais são necessárias para se explicar pelo menos 80% da variância dos dados?
3. Obtenha as contribuições das preditoras para a primeira componente principal. Qual nome você daria para esta componente?
4. Obtenha as contribuições das preditoras para a segunda componente principal. Qual nome você daria para esta componente?
5. Faça um gráfico de dispersão com as duas primeiras componentes principais. Com base nas respostas anteriores e neste gráfico, o que pode-se dizer sobre a Capital? E sobre a região de Ribeirão Preto? E Sorocaba?
6. Análise de conglomerados.
 - 6.1. Execute o método k-means para identificar o número ótimo de clusters entre as regiões analisadas;
 - 6.2. Visualize os grupos obtidos em um gráfico de dispersão (utilize o resultado do PCA para a construção dos gráficos);
 - 6.3. Analise os resultados dos métodos de clusterização e interprete os grupos obtidos;
 - 6.4. Discuta as implicações práticas dos grupos identificados, considerando possíveis ações que a Secretaria de Segurança Pública de São Paulo pode realizar.

Entrega: enviar um arquivo .zip contendo o relatório .rmd e a saída .html ou .pdf.