

Curso: Engenharia de computação	período: 2016.2
aluno: Filipe Cazuza Cavalcanti	professor: Dr.Paulo Ribeiro Lins
disciplina: probabilidade e estatística aplicada a computação	email: felicavalcanti157@gmail.com

Projeto de probabilidade e estatística
aplicada a computação
manipulação de Dados usando Python
exibição de gráficos (histograma,
boxplot), e medidas Resumo.

Introdução

O projeto descrito neste relatório faz uso de dados requeridos no Google Trends [link](#) toda análise feita neste no mesmo deu-se por meio da contemplação dos gráficos gerados junto com as medidas resumo, os mesmos foram gerados com as bibliotecas Matplotlib.pyplot, pandas. Sobre o que é analisado neste projeto são dados sobre a gripe nos estados Ceará, Minas Gerais, Rio de Janeiro e Paraná e valores nacionais gerando gráficos e medidas resumo de cada estado incluindo o nacional. Analisando tal conjunto de dados de maneira correta consegue-se afirmar onde pode ter havido um provável surto de gripe o estado onde tem uma maior procura sobre este tópico, em que ano houve um provável surto de gripe em todo o país.

Desenvolvimento

Ao falarmos de desenvolvimento primeiramente vamos analisar o código que realiza nosso “trabalho pesado” por assim dizer. A implementação deste projeto for por de funções genéricas onde fale explicar algumas como a função boxplot onde esta recebe dois parâmetros lista e um title, internamente esta função cria uma lista temporaria daí percorre a lista que recebe de parâmetro e faz uma cópia na lista temporaria assim usando a função boxplot nativa do matplotlib.pyplot gerando assim o gráfico de caixas o boxplot, outra função que merece ser descrita é a função histograma em que ela também recebe dois parâmetros que lista e title internamente a função usa a função hist() nativa da biblioteca matplotlib.pyplot, onde esta cria o gráfico histograma, a função describe sendo uma função muito simples fazendo apenas um print do retorno de três outras funções que são DMA() que retorna o desvio médio absoluto, CV() que retorna o coeficiente de variação do conjunto de dados e por último a função describe() nativa da biblioteca pandas em que ela retorna uma tabela com algumas medidas resumo com os quartis o valor mínimo o máximo a média, amplitude e count.

Conclusão

conclui-se ao final deste projeto após ter mencionado minuciosamente o desenvolvimento, que o estado onde pode ter havido um número maior de vezes de surtos de gripe é paraná em contrapartida o estado onde pode ter havido um número menor de surtos de gripe foi Ceará (valendo salientar que, toda a análise está sendo feita com os dados obtidos no Google Trends, então a análise está desconsiderando fatores externos como por exemplo.: a disponibilidade de acesso a internet ser muito diferente entre os estados o que leva a ter estados que tem um número maior de buscas em relação a gripe do que em outros podendo assim mudar a interpretação). A assimetria dos dados nacionais é melhor em relação aos estados separadamente, isso ocorre pelo de fato de ser uma amostra maior de dados se comparados com os estados. Nas próximas páginas estão tabelas e gráficos gerados a partir dos dados passados neste projeto

describe (brazil)

DMA	52.826034678898424
CV	0.0041876381258780706
count	497.000000
mean	199.557344
std	65.864729
min	89.000000
25%	150.000000
50%	192.000000
75%	239.000000
max	432.000000

describe (Ceará)

DMA	39.710723091061425
CV	0.021351310254964683
Count	497.000000
mean	161.742455
std	48.950704
min	83.000000
25%	122
50%	152
75%	192
max	341

describe (Minas Gerais)

DMA	63.89582565817441
CV	0.0079353326864580621
Count	497.000000
mean	218.985915
std	79.229059
min	95.000000
25%	157.000000
50%	208.000000
75%	265.000000
max	472.000000

describe (Paraná)

DMA	61.46449724503971
CV	6.1163926504425277)
count	497.000000
mean	196.865191
std	78.919722
min	68.000000
25%	138.000000
50%	183.000000
75%	239.000000
max	562.000000

describe (Rio de Janeiro)

DMA	49.22989850572248
CV	0.35257426286726146
count	497.000000
mean	209.237425
std	59.427137
min	106.000000
25%	161.000000
50%	204.000000
75%	252.000000
max	387.000000









