



PUC
CAMPINAS
PONTIFÍCIA UNIVERSIDADE CATÓLICA



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS
PROGRAMA DE RESIDÊNCIA TECNOLÓGICA EM CIÊNCIA DE DADOS

LAÍS BARRETO VIEIRA
RIQUELI A. SILVA
VINICIUS DA SILVA ABREU

EXERCÍCIOS DE ANÁLISE EXPLORATÓRIA DE DADOS

Professor: José Guilherme Picolo

Introdução

A mineração de dados desempenha um papel crucial na extração de informações valiosas a partir de conjuntos de dados complexos. O objetivo desse documento é descrever o processo de exploração e análise dos dados, bem como apresentar os resultados encontrados, referente à atividade de Análise Exploratória de Dados da disciplina de Ciência de Dados do Programa de Capacitação com Residência Tecnológica em Ciência de Dados.

Para tal, utilizamos o conjunto de dados proveniente do arquivo "*Dataset: SinaisVitalis003 100Dias DV2 XXX5.txt*." Esta atividade tem como objetivo principal explorar e analisar os sinais vitais, como batimento cardíaco, pressão arterial e temperatura corporal, por meio de técnicas estatísticas avançadas em busca da extração de características e padrões entre esses dados.

Para alcançar os objetivos propostos, desenvolvemos um conjunto de códigos em Python com o uso da metodologia de Orientação a Objetos a fim de garantir a organização do código de forma modular e reutilizável.

Implementação

1. Pré-processamento dos dados

Uma etapa vital na mineração de dados é a preparação da base de dados para garantir a qualidade dos resultados. Desse modo, foram definidas classes para realização de operações de pré-processamento dos dados, em busca de inconsistências e valores fora dos limites pela média dos limites vizinhos (outliers) para serem substituídos antes da exploração dos dados, levando em consideração os critérios de validação para os seguintes intervalos dos parâmetros em repouso (Batimento cardíaco: 0 a 100bpm; Pressão arterial: 0 e 20; Temperatura Corporal: 0 e 40°C). Só depois da preparação dos dados, é criado o *data-frame* com os parâmetros: hora, batimento cardíaco, pressão arterial e temperatura corporal. O arquivo de dados foi segmentado em pacotes de 24 amostras, correspondendo a um dia de amostragem.

2. Correlação e identificação de padrões

Após a segmentação dos dados, foram realizadas as correlações entre os parâmetros: batimento cardíaco, pressão arterial e temperatura corporal. Em seguida, foi instanciada uma classe para classificação das correlações encontradas. O coeficiente de Correlação (r) mede a força de associação entre duas variáveis, quanto menos dispersos os dados, mais forte será a independência.

As correlações foram avaliadas de acordo com os seguintes critérios de classificação para o Coeficiente de Correlação (r) quanto à força de correlação:

- Correlação Muito Forte: 0.8 a 1.0
- Correlação Forte: 0.6 a 0.8
- Correlação Moderada: 0.4 a 0.6
- Correlação Fraca: 0.2 a 0.4
- Correlação Muito Fraca: 0.0 a 0.2

3. Análise Estatística Descritiva

Foram calculadas as medidas de tendência central, como média, mediana e moda, para compreender o comportamento típico dos sinais vitais. A média representou o valor central dos dados, enquanto a mediana e a moda proporcionaram insights sobre a distribuição e a frequência dos valores. Além disso, o desvio padrão foi calculado para avaliar a dispersão dos dados em relação à média, fornecendo informações valiosas sobre a consistência e a variabilidade dos sinais vitais monitorados. Essas análises foram cruciais para uma compreensão abrangente dos padrões e variações nos sinais vitais ao longo do tempo.

4. Plotagem dos gráficos

Os gráficos da amostragem foram elaborados utilizando as bibliotecas *Seaborn* e *Matplotlib.pyplot*, nos formatos de histograma e mapa de calor. Para cada um dos três parâmetros analisados, também foram incluídos os valores mínimo, máximo e médio.

Análise dos resultados

- **Análise estatística descritiva**

A análise estatística dos dados revelou informações essenciais sobre os sinais vitais monitorados. No que diz respeito ao batimento cardíaco, a média registrada foi de 70.13 batimentos por minuto (bpm), com uma mediana de 70.63 bpm, indicando que a maioria das observações se concentra nessa faixa. A moda, que representa o valor mais frequente, foi de 70.69 bpm. O desvio padrão de 2.63 bpm demonstra a dispersão dos dados em relação à média, indicando uma variação moderada. O valor máximo observado foi de 74.00 bpm, enquanto o mínimo foi de 61.73 bpm.

Em relação à pressão arterial, a média registrada foi de 12.18 mmHg, com uma mediana de 12.43. A moda foi de 12.35, indicando o valor mais frequente. O desvio padrão de 1.26 sugere uma dispersão moderada dos dados em torno da média. O valor máximo de pressão arterial observado foi de 14.00, enquanto o mínimo foi de 8.39.

Quanto à temperatura corporal, a média foi de 36.89°C, com uma mediana igualmente de 36.89°C, o que indica uma distribuição bastante uniforme dos dados. A moda foi de 36.46°C, representando o valor mais frequente. O desvio padrão de 0.20°C reflete uma baixa dispersão em relação à média, sugerindo uma estabilidade nas leituras. O valor máximo observado foi de 37.50°C, enquanto o mínimo permaneceu em 36.46°C.

Esses resultados fornecem uma visão abrangente e detalhada dos sinais vitais analisados, permitindo uma compreensão mais aprofundada de seus padrões e variações ao longo do tempo. Eles são fundamentais para avaliações clínicas e monitoramento da saúde, contribuindo para a tomada de decisões embasadas em dados precisos e confiáveis.

- **Análise de correlação entre os dados**

A correlação de 0,91 entre o batimento cardíaco e a pressão arterial evidencia uma relação extremamente forte e positiva entre essas variáveis. Em outras palavras, quando o batimento cardíaco aumenta, a pressão arterial tende a aumentar também, e vice-versa. Essa descoberta é suportada pelos aspectos estatísticos, reforçando a robustez dessa relação. O coeficiente de correlação, próximo de 1, indica uma associação quase linear muito forte entre as variáveis, demonstrando que as mudanças em ambas estão altamente interligadas.

Essa alta correlação pode ser fundamental na compreensão de padrões fisiológicos ou condições médicas em que as variações no batimento cardíaco estão intimamente ligadas a alterações na pressão arterial. Possíveis explicações para essa associação incluem:

1. Fisiologia Humana: O batimento cardíaco (frequência cardíaca) e a pressão arterial estão intrinsecamente relacionados. A contração do coração influencia diretamente a pressão sanguínea.

2. Regulação do Sistema Nervoso Autônomo: O sistema nervoso autônomo desempenha um papel fundamental na regulação do batimento cardíaco e da pressão arterial. Mudanças na atividade simpática (aceleração do batimento cardíaco) e parassimpática (desaceleração do batimento cardíaco) podem influenciar ambos os parâmetros.

3. Resposta ao Estresse: Situações de estresse ou excitação podem resultar em um aumento simultâneo do batimento cardíaco e da pressão arterial. Isso é uma resposta fisiológica natural ao preparar o corpo para lidar com desafios iminentes.

4. Condições Médicas: Certas condições médicas, como hipertensão arterial, podem contribuir para uma correlação forte entre o batimento cardíaco e a pressão arterial.

Essa forte correlação entre o batimento cardíaco e a pressão arterial é um achado significativo e pode ter implicações importantes na compreensão da saúde cardiovascular e em contextos clínicos.

A observação da correlação entre a pressão arterial e a temperatura corporal revela uma relação interessante. O coeficiente de correlação de 0.43 indica uma associação positiva moderada entre esses dois parâmetros. Isso sugere que, em geral, quando a pressão arterial tende a aumentar, a temperatura corporal também apresenta uma elevação correspondente, e vice-versa, em uma medida considerada moderada.

Essa relação pode ser resultado de diversos fatores fisiológicos complexos. Por exemplo, em situações de estresse ou atividade física intensa, o corpo tende a aumentar tanto a pressão arterial quanto a temperatura corporal para garantir um fluxo sanguíneo adequado e a dissipação de calor.

Além disso, outras variáveis, como o estado emocional do indivíduo, podem influenciar essa correlação. Por exemplo, em situações de ansiedade ou excitação, é comum observar um aumento na pressão arterial e, conseqüentemente, na temperatura corporal.

No entanto, é importante notar que a relação entre esses parâmetros é considerada moderada, o que significa que outros fatores também podem influenciar essas medições de forma independente. Portanto, é crucial considerar outros aspectos clínicos e contextuais ao interpretar esses resultados.

- **Monitoramento e correlações entre os sinais vitais e possíveis implicações para a saúde.**

As informações relacionadas à pressão arterial, batimento cardíaco e temperatura corporal podem oferecer insights valiosos sobre a saúde de uma pessoa. Aqui estão algumas correlações e informações relevantes:

1. Pressão Arterial:

- Pressão arterial elevada (hipertensão) pode indicar um risco aumentado de doenças cardíacas, derrame e problemas renais.
- Pressão arterial baixa pode causar tonturas, desmaios e indicar problemas como desidratação ou condições cardíacas subjacentes.

2. Batimento Cardíaco (Frequência Cardíaca):

- Um batimento cardíaco regular em repouso geralmente indica um coração saudável.
- Frequência cardíaca elevada em repouso pode indicar estresse, ansiedade ou condições cardíacas.
- Frequência cardíaca muito baixa pode ser um sinal de problemas no nó sinusal ou outras condições cardíacas.

3. Temperatura Corporal:

- A temperatura corporal normal é em torno de 37°C (98.6°F).
- Febre (temperatura corporal elevada) geralmente indica uma resposta do corpo a uma infecção ou outra condição médica.
- Hipotermia (temperatura corporal muito baixa) pode ocorrer em exposição prolongada ao frio e pode ser perigosa.

Correlações:

- **Febre e Frequência Cardíaca:** A febre muitas vezes leva a um aumento na frequência cardíaca, pois o corpo trabalha mais para combater a infecção.
- **Pressão Arterial e Estresse:** O estresse crônico pode levar a hipertensão, aumentando o risco de problemas cardíacos.
- **Temperatura Corporal e Hipotermia:** Em casos de hipotermia, a temperatura corporal baixa pode levar à diminuição da frequência cardíaca e da pressão arterial, colocando a pessoa em risco de complicações graves.

É importante monitorar esses sinais vitais regularmente para identificar alterações que possam indicar problemas de saúde e procurar orientação médica quando necessário.

- **Critérios para tomada de decisão**

Na análise dos pacotes de dados, estabelecemos critérios fundamentais para a tomada de decisão em situações que demandam atenção especial.

Alerta de Frequência Cardíaca Elevada: Caso a média dos batimentos cardíacos em um pacote persista consistentemente acima de 90 bpm, este pode ser um indicador de taquicardia, uma condição que necessita de avaliação médica imediata.

Alerta de Pressão Arterial Elevada: Se a média da pressão arterial em um pacote se mantiver consistentemente acima de 15 mmHg (ou outro valor limiar predefinido), isso pode indicar hipertensão, exigindo acompanhamento médico para controle e avaliação adequada.

Alerta de Temperatura Corporal Anormal: Será acionado um alerta se a média da temperatura corporal em um pacote estiver fora da faixa considerada normal (por exemplo, abaixo de 36.5°C ou acima de 37.5°C). Esta condição requer atenção e avaliação médica para determinar possíveis causas e intervenções necessárias.

Esses critérios estabelecidos são cruciais para a detecção precoce de potenciais problemas de saúde, permitindo uma intervenção rápida e adequada para garantir a segurança e bem-estar do indivíduo monitorado.

Conclusão

Diante das análises realizadas, foi possível obter insights valiosos que promovem uma compreensão mais aprofundada e um monitoramento mais eficaz dos sinais vitais ao longo do tempo. A implementação e utilização desses códigos representam um avanço considerável na exploração e interpretação de conjuntos de dados clínicos, trazendo consigo um grande potencial de aplicação em cenários médicos e de saúde. Esses resultados destacam a relevância e a promissora contribuição dessa abordagem para a área da saúde e indicam caminhos promissores para pesquisas futuras.

Referências Bibliográficas

1. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.
2. Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc.
3. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
4. VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc.
5. Seaborn Development Team. (2021). Seaborn: statistical data visualization. <https://seaborn.pydata.org/index.html>

6. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95. <https://matplotlib.org/>
7. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. <https://scikit-learn.org/stable/>
8. Wes McKinney. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56. <https://www.scipy.org/>
9. Millman, J., & Aivazis, M. (2011). Python for Scientists and Engineers. Computing in Science & Engineering, 13(2), 9-12. <https://www.nature.com/cise/>