Índice

[Introdução 3](#_Toc134487236)

[Preparação de dados 4](#_Toc134487237)

[Extração 4](#_Toc134487238)

[Limpeza 4](#_Toc134487239)

[Seleção de variáveis 5](#_Toc134487240)

[Transformação 5](#_Toc134487241)

[Organização 5](#_Toc134487242)

[Ferramentas de exploração e visualização de dados 5](#_Toc134487243)

[Técnicas de visualização de dados 5](#_Toc134487244)

[Análise exploratória de dados 5](#_Toc134487245)

[Casos de uso de exploração e visualização de dados 5](#_Toc134487246)

[Desafios e tendências 5](#_Toc134487247)

[Conclusão 5](#_Toc134487248)

[Bibliografia 5](#_Toc134487249)

[Webgrafia 6](#_Toc134487250)

# Introdução

No âmbito da unidade curricular de seminário foi pedido para de fazer um relatório sobre um de 15 temas dados a escolher, no qual este grupo optou pelo tema número 15: “Exploração e visualização de dados”. Ao longo deste relatório iremo-nos debruçar sobre os seguintes temas:

* Preparação de dados;
* Ferramentas de exploração e visualização de dados;
* Técnicas de visualização de dados;
* Análise exploratória de dados;
* Casos de uso de exploração e visualização de dados;
* Desafios e tendências.

Com o crescimento das novas tecnologias, o surgimento da internet das coisas, a rapidez com que a informação circula e a capacidade crescente de criação, assim como, com o armazenamento de dados, estima-se que em 2025 o tamanho da Esfera de Dados Global seja de 175 ZB [1], [2]. Tendo em conta que a exploração e visualização de dados é utilização de técnicas para apresentar e analisar informação, por vezes extremamente complexa, tanto pela sua estrutura como pela sua quantidade, como referido anteriormente, de forma transparente. Isso faz com que a exploração de dados e a sua visualização, nos próximos anos, assuma um papel cada vez mais importante por esta ser uma ferramenta poderosa para explorar, entender e comunicar informação complexa. Uma imagem com gráfico

Descrição gerada automaticamente

Como referido anteriormente, devido a grande quantidade de dados que possuímos atualmente, pode ser difícil extrair conhecimento significativo apenas olhando para números e tabelas. É aí que a visualização de dados entra em jogo, permitindo que os usuários transformem informações em gráficos e diagramas facilmente compreensíveis. Além de ajudar a entender os dados, a visualização também tem o poder de revelar padrões, tendências e anomalias que podem não ser aparentes em formatos de dados brutos [3]. Ao visualizar dados, podemos descobrir relações e informação que não seriam detetáveis de outra forma.

# Preparação de dados

Muitas vezes os dados que possuímos são derivados de texto, tabelas ou base de dados e são nos apresentado de uma forma bruta, isto é, com valores em falta, erros, distorções, entre outros problemas. É por isso que a preparação de dados é um passo importante para a exploração e visualização de dados, pois é nesta fase que se trabalha a qualidade dos dados, que por vezes pode ser um processo demoroso e chato. É nesta fase que vamos determinar o que são os dados, melhorar a sua qualidade, padronizar, consolidar e transformá-los para que estes sejam uteis para posterior analise. A preparação de dados pode ser dividida por vários passos [4], [6], [7] :

1. Extração;
2. Limpeza;
3. Seleção de variáveis;
4. Transformação;
5. Organização.

## Extração

A extração é onde os dados são obtidos, seja através da recolha de dados brutos, importação de dados de fontes externas ou através de arquivos armazenados em sistemas de gestão de base de dados. É importante garantir que a fonte dos dados esteja fiável e que os dados sejam recolhidos de forma consistente [7] .

## Limpeza

Após a extração, vem a fase de limpeza, onde os dados são submetidos a uma série de tratamentos para eliminar dados duplicados, corrigir valores incorretos, preencher valores em falta e eliminar valores discrepantes. Este é um passo crucial na preparação de dados, pois dados incorretos ou incompletos podem levar a análises imprecisas e conclusões erradas [4], [5], [6], [7] .

## Seleção de variáveis

Na seleção de variáveis, o foco é identificar as variáveis que são relevantes para a análise, eliminando as que não têm importância ou são redundantes. Isso ajuda a simplificar a análise, tornando-a mais eficiente[8].

## Transformação

A transformação de dados pode incluir a normalização, discretização e agregação de dados, bem como a conversão de formatos de dados. Este passo é importante para garantir que os dados sejam comparáveis e possam ser analisados em conjunto [4], [5], [6], [7] .

## Organização

Por fim, a organização dos dados é feita para garantir que os dados estejam num formato que seja fácil de trabalhar, incluindo a ordenação dos dados, a definição de tipos de dados e a criação de tabelas de base de dados. Esta fase é importante para garantir que os dados sejam acessíveis e possam ser facilmente integrados em ferramentas de análise e visualização [4], [5], [6], [7] .

# Ferramentas de exploração e visualização de dados

# Técnicas de visualização de dados

# Análise exploratória de dados

# Casos de uso de exploração e visualização de dados

# Desafios e tendências

# Conclusão

# Bibliografia

[1] C. S. Rosa, «UNIVERSIDADE FEDERAL DE UBERLÂNDIA FACULDADE DE ENGENHARIA ELÉTRICA-PATOS DE MINAS ENGENHARIA ELETRÔNICA E DE TELECOMUNICAÇÕES Estudo sobre as técnicas e métodos de análise de dados no contexto de Big Data», 2018.

[2] D. Reinsel, J. Gantz, e J. Rydning, «The Digitization of the World From Edge to Core», 2018.

[3] F. S. Tsai e K. L. Chan, «Dimensionality Reduction Techniques for Data Exploration».

[4] D. Stodder, «Improving Data Preparation for Business Analytics Applying Technologies and Methods for Establishing Trusted Data Assets for More Productive Users BEST PRACTICES REPORT Q3 2016», 2016.

[5] G. Mansingh, K. M. Osei-Bryson, L. Rao, e M. McNaughton, «Data preparation: Art or science?», em *Proceedings of the 2016 International Conference on Data Science and Engineering, ICDSE 2016*, Institute of Electrical and Electronics Engineers Inc., Jan. 2017. doi: 10.1109/ICDSE.2016.7823936.

[6] J. Brownlee, «Data Preparation for Machine Learning», 2020.

[7] B. R. Santos, P. T. Fonseca, M. Barata, R. A. Ribeiro, e P. A. C. Sousa, «New data preparation process - A case study for an exomars drill», em *2006 World Automation Congress, WAC’06*, IEEE Computer Society, 2006. doi: 10.1109/WAC.2006.376041.

[8] L. Soibelman, M. Asce, e H. Kim, «Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases», 2002, doi: 10.1061/ASCE0887-3801200216:139.

# 

# Webgrafia