Data Processing and Visualization (P02)

DECISION SUPPORT SYSTEMS, 2022-23

**Pedro Simões (21140), Gonçalo Cunha (21145), João Apresentação (21152)**

<<The goal of this project is to develop a data processing and visualization solution based on the data mart implemented in the project 01. The team may use a different data source, but this must be validated previously by the teacher.

This document has an appropriate structure for the project, however chapters can be deleted, changed, or added. Keep formatting consistent throughout the document.

**Comments**, such as this one, placed between "<<" and ">>", **should be removed**.>>

# Introduction

Processamento e Visualização de Dados é o segundo trabalho prático da cadeira de Sistemas de Apoio à Decisão, do curso de Licenciatura de Sistemas Informáticos (pós-laboral), que tem como objetivo desenvolver uma solução de um software de processamento e visualização de dados, na qual terá de ser implementadas operações de:

* Preparação de dados com Power Query
* Cálculo de colunas e medidas para as tabelas da data mart
* Apresentação e visualização de dados

Outro desafio proposto pelo docente da cadeira, foi a integração de fontes de dados externas, com o objetivo de fazer uma análise a todos os dados e encontrar padrões entre os resultados obtidos entre as diferentes fontes.

<<Falar de todos os pontos abordados do enunciado>>

# Data acquisition and preparation

<<Describe the connections made to the data sources and the major data cleaning and preparation tasks performed.>>

Neste capítulo será mencioado que fontes de dados foram utilizadas e as alterações feitas nas tabelas em causa, segue-se agora a lista de fontes utilizadas neste projeto:

* WWI Data Mart (fonte de dados principal) – data mart fornecida pelo docente, o grupo teria tido bastante interesse em reutilizar o projeto produzido na última entrega desta cadeira, todavia esta não demonstrou ser a melhor solução, portanto foi utilizada a que o docente reproduziu

Fonte: dados de conexão fornecidos pelo docente

* PIB (GDP) per capita nos preços atuais (2017-2020) nos EUA (fonte de dados externa) – dataset composta pelo valor do PIB nos Estados Unidos entre os anos de 2017 e 2020, esta fonte foi escolhida pela relação direta com os gastos por cada pessoa, algo que poderia ter uma possível relação com o número de vendas feitas no WWI  
  Fonte: <https://data.un.org/Data.aspx?d=SNAAMA&f=grID:101;currID:USD;pcFlag:true;crID:840;yr:2017,2018,2019,2020&c=2,3,5,6&s=_crEngNameOrderBy:asc,yr:desc&v=1>
* Taxa de inflação nos EUA (fonte de dados externa) – a mudança da taxa de inflação tem impacto direto no poder de compra dos clientes, bem como nos preços dos produtos que estejam para venda, isto poderá ter algum impacto no número de vendas da WWI  
  Fonte: <https://data.bls.gov/timeseries/CUUR0000SA0L1E?output_view=pct_12mths>
* Número de trabalhadores por estado nos EUA (fonte de dados externa) – o número de trabalhadores num estado poderá ser informação util, por exemplo para prever qual estado possuí o maior valor em compras na WWI, tendo em conta que quantos mais trabalhadores, mais salário é distribuído por essa região/estado  
  Fonte: <https://www.kaggle.com/datasets/justin2028/unemployment-in-america-per-us-state>

Falando agora inicialmente no tratamento e preparação de dados feito nas fontes externas, de uma forma geral as alterações aplicadas foram com o objetivo de adaptar as tabelas obtidas na data mart em causa

# Data modelling and processing

<<Present the **data model**, including the definition/refinement of the relationships between tables, and the creation of **hierarchies**.

Next, describe and present the definition of all the calculated **columns** and **measures**. Include also in this section, other relevant data preparation/calculation operations not mentioned in the previous point.>>

# Data visualization

<<Identify the **goals of the dashboards** and the **profile of end users**. Add screenshot of the dashboards, a short description of the dashboards and other relevant elements, e.g., slicers, KPIs, influencers, etc.>>

# Conclusion

<< Insert here a conclusion about the results accomplished, the development process and the tools used. The structure of the report should be adapted according to each project characteristics. Don’t forget to **remove these comments** (help text)!>>