Proposta de projeto

# Ferramenta de anotação semi-automática de ficheiros de dados

Aplicação WEB para análise de meta dados.

# Alunos:

Ruben Café

912317910

[A42176@alunos.isel.pt](mailto:A42176@alunos.isel.pt)

Ivo Pereira

926276418

[A42172@alunos.isel.pt](mailto:A42172@alunos.isel.pt)

Duarte Felício

967126787

[A42197@alunos.isel.pt](mailto:A42197@alunos.isel.pt)

# Orientadores:

Nuno Datia ([datia@isel.ipl.pt](mailto:datia@isel.ipl.pt))

Matilde Pato (matilde.pato@isel.pt)

16/03/2020

# Introdução

## Enquadramento

Nos tempos atuais a quantidade de dados recolhidos e armazenados mundialmente aumenta de dia para dia. É estimado que sejam gerados 4.1 terabytes de dados diariamente por quilómetro quadrado de área urbana. Esta recolha de dados aumentou tanto em volume como em detalhe. Os vários fornecedores e/ou agregadores de dados têm diferentes formas de os disponibilizar, originando análises cada vez mais complexas. Em particular, identificar níveis de detalhe e encontrar relações entre os dados tornou-se mais difícil.

É necessário utilizar um processo semiautomático que, sobre um conjunto de dados, gere meta informação para identificar o domínio das variáveis, relações entre os dados e os diferentes níveis de detalhe (Ribeiro, 2019).

O principal foco deste projeto é desenvolver uma aplicação WEB interativa que, recebendo dados em formato CSV (*comma separated values*), utiliza um processo de análise semiautomático de modo a proporcionar ao utilizador uma melhor compreensão dos vários níveis de detalhe e relações entre os dados inseridos.

## Descrição do projeto

Neste projeto será desenvolvida uma aplicação WEB interativa que efetua, autonomamente, análises a conjuntos de dados representados em CSV e identifica possíveis níveis de detalhe e relações presentes nesses dados. Essa informação é possível de ser validada pelo utilizador, num processo iterativo e interativo. Para além da visualização dos níveis de detalhe e das relações entre os dados, deve ser possível exportar essa informação como meta dados.

# Análise

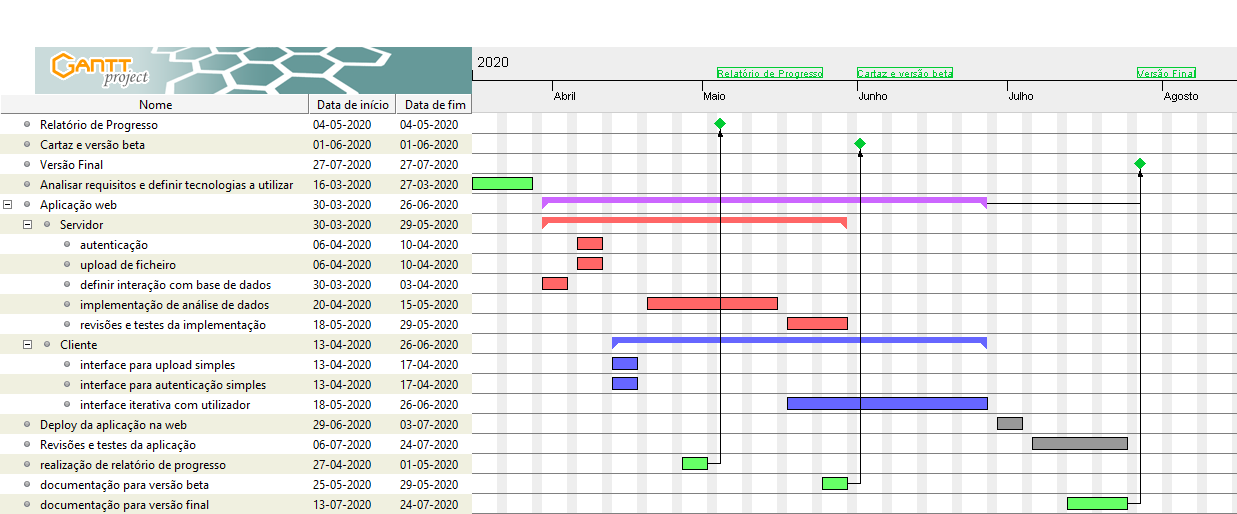
## Problemas a resolver:

* Identificação das melhores praticas e tecnologias a utilizar ao longo do desenvolvimento do projeto;
* Desenhar e implementar uma base de dados;
* Implementação de processo semi-automático para tratamento de meta-dados;
* Extração de meta dados;
* Desenvolvimento de testes com dados dos censos (INE);
* Desenvolvimento de uma interface interativa;
* Documentar o uso da aplicação de modo a fácil utilização;

## Possíveis Técnicas e ferramentas a usar:

* Base de dados em SQL Server ou PostgresSQL;
* Linguagem da aplicação: java, javascript;
* Ferramentas para complementar a interface: Angular, React, etc;
* Ferramentas para melhorar o design da aplicação, como wordpress;

# Plano



# Referências

Ribeiro, N. (2019). *Anotação e Extração Semi-Automática de dados multidimensionais.* Lisboa: Universidade Nova de Lisboa - Faculdade de Ciências e Tecnologia.