**Desenvolvimento de um modelo de análise de dados**

**Sumário do problema abordado na 1ª parte do projeto**

No cenário em questão, um ou mais compradores possuem uma lista de produtos que pretendem comprar. Simultaneamente, vários vendedores anunciarão a venda de diversos produtos.

Cada vendedor fixa um preço mínimo para cada produto que vende, sendo este o valor inicial do produto em leilão (English auction). O mesmo produto, oferecido por diferentes vendedores, pode ter preços diferentes.

Cada comprador estará disposto a gastar um valor máximo por cada produto associado ao respetivo vendedor. O valor máximo que cada comprador se dispõe a gastar será também influenciado por diferentes métricas. Assim, o preço inicial do leilão, a reputação do vendedor e o tempo estimado de entrega do produto influenciarão a decisão do comprador.

**Análise de dados preditiva a resolver**

Para um produto e dado um vendedor, o seu atraso na entrega do mesmo, o número de potenciais interessados no produto, o número de outros vendedores do mesmo produto, o valor inicial do leilão (valor mínimo da venda), pretende-se saber se o produto é vendido ou não e por quanto (classificação e regressão).

**Variável dependente**

A variável dependente em estudo na análise preditiva será a venda (ou não) do produto. O grupo reconhece este caso como sendo um de classificação.

O grupo pretende ainda fazer uma análise preditiva por regressão, não só vendo se o produto é vendido ou não mas por quanto (valor contínuo).

**Alguns exemplos de variáveis independentes**

Como variáveis independentes, o grupo pensou em utilizar:

1. **valor inicial do leilão**: dado por uma distribuição Gaussiana com um certo valor médio e desvio padrão, sendo estas variáveis independentes por si
2. **nome do vendedor**: para efeitos de identificação do mesmo apenas
3. **atraso de entrega do vendedor**: o atraso que um vendedor tem em média a entregar o produto
4. **número de leilões do mesmo produto (de outros vendedores)**: basicamente impondo mais ou menos concorrência na venda do produto
5. **número de potenciais compradores interessados no nosso produto**: aumentando ou diminuindo a procura

CENAS:

1. desvio percentual entre a média e o valor inicial de venda (media/ (inicial \* 100))
2. desvio entre o valor maximo medio dos buyers e o initial value