FE03

Universidade do Minho
Escola de Engenharia

Curso: Mestrado em Engenharia Biomédica
U.C.: Aprendizagem e Extração do Conhecimento

Ficha de Exercícios 03	
Docente:	Hugo Peixoto | José Machado
Tema:	AI Studio: Correlações, Regras de Associação e Regressão Linear
Ano Letivo:	2026-2027 – 1° Semestre
Duração da aula:	4 horas

Parte I — Correlações

[1] Quais as principais limitações de modelos de correlações?

[2] O que é um coeficiente de correlação e como é interpretado?

[3] Qual a diferença entre uma correlação negativa e uma correlação positiva?

[4] Se dois atributos diminuem essencialmente à mesma taxa, é uma correlação positiva ou negativa? Explique.

[5] Como é medida a força de uma correlação? Quais os limites para essa força?

[6] Consegue pensar em atributos que poderiam ser interessantes incluir no dataset estudado no exemplo da aula?

[7] Aceda ao ficheiro *mpg_dataset.csv* e tenha em consideração que:

cylinders: número de cilindros no motor
displacement: cilindrada do motor
horsepower: potência do motor
weight: peso do veículo
acceleration: aceleração do veículo, i.e., tempo em segundos para acelerar de 0 a 60
model year: ano do modelo do veículo nos anos 1900s
origin: origem do carro (1: Americana, 2: Europeia, 3: Japonesa)
mpg: consumo/eficiência de combustível (miles per gallon)

[8] Execute a etapa de Data Preparation no AI Studio. Importe o csv e acrescente-o no design.

[9] Documente quais os atributos que podem influenciar ou explicar o consumo/eficiência de combustível num determinado veículo (mpg).

Parte II — Regras de Associação

[1] O que são regras de associação? Para que servem?

[2] Quais são as duas principais métricas calculadas nas regras de associação e como são calculadas?

[3] De que tipo de dados devem ser os atributos de um dataset para usar os operadores Frequent Pattern no AI Studio?

[4] Como é que os resultados das regras de associação são interpretados? No exemplo dos slides desta aula, qual foi a regra mais forte e como se sabe?

[5] Faça download do dataset *orders.csv*, importe-o para o AI Studio e arraste-o para a janela de processo. Proceda à etapa de Data Understanding.

[6] Conforme necessário, execute os passos referentes à etapa de Data Preparation no seu dataset. Certifique-se de que todas as suas variáveis tenham o tipo correto (binomial) para a execução do operador FP-Growth.

[7] Gere regras de associação para o dataset utilizando o método descrito na aula (FP-Growth).

a. Modifique os valores de confiança (min confidence) e suporte (min support) para identificar os níveis ideais, de modo a obter regras interessantes e significativas.

b. Analise as outras medidas de força das regras, como LaPlace ou Lift.

c. Documente as suas descobertas. Que regras encontrou? Que atributos estão mais fortemente associados? Existem produtos frequentemente conectados que possam sugerir oportunidades de negócio?

[8] Crie um modelo de regras de associação usando o mesmo dataset, mas desta vez, use o operador W-FPGrowth no AI Studio.

[9] O algoritmo Apriori é frequentemente usado no processo de Data Mining para associações. Pesquise Apriori (W-Apriori) nos operadores do AI Studio e compare os resultados com os obtidos anteriormente.

Parte III — Regressão Linear

Importe o dataset *insurance.csv*.

[1] Execute o passo Data Understanding da metodologia CRISP-DM.

[2] Repita os passos no AI Studio tal como descritos nos slides da aula e, após executar o seu modelo, examine os coeficientes de regressão e responda às perguntas que se seguem.

[3] Relate os seus resultados:

[a] Que atributos têm maior peso?

[b] Algum atributo foi removido do conjunto de dados por não ter uma boa capacidade de previsão? Em caso afirmativo, quais e porquê?

[c] Que outros atributos considera que poderiam ajudar o modelo a prever melhor o custo anual do seguro de saúde?

[4] Utilizando a visualização do AI Studio avalie os diferentes gráficos e execute uma descrição do que for encontrando. Por exemplo:

[a] Relação entre o Sexo e o Prémio do Seguro

[b] Relação entre o número de dependentes e o Prémio do Seguro

[c] Relação entre se é ou não fumador e o Prémio do Seguro

[d] Relação entre a localização e o Prémio do Seguro

Importe o dataset *happyscore_income.csv* e prepare o processo de Data Mining da forma que considerar mais correta.

[5] Execute os processos de Business e Data Understanding para perceber qual o processo de DM que deverá ser aplicado. Deverá colocar:

[a] 1 frase que caracterize o objetivo do negócio.

[b] 1 frase que caracterize o objetivo do processo de DM.

[c] 1 tabela com o resumo dos atributos encontrados (tipo, min, max, med, nulos, etc).

[6] Crie o modelo de DM no AI Studio e justifique a utilização de todos os operadores de preparação de dados que incluiu no seu modelo.

[7] Descreva as principais conclusões a que chegou.