PARTE I

1. **Quais as principais limitações de modelos de correlações?**

   R.: É possível que, na matriz de correlação, apareça um valor forte , indicativo de uma correlação. No entanto, tal valor não faz sentido, e é pura coincidência que os dois atributos estejam intimamente relacionados. Assim, é importante entender e aceitar que há limitações nestes modelos.

2. **O que é um coeficiente de correlação e como é interpretado?**

   R.: Um coeficiente de correlação é uma medida do grau de correlação entre duas variáveis de escala métrica. Este coeficiente assume apenas valores entre -1 e 1. Assim, se o valor for 1 (ou próximo), significa que existe uma correlação perfeita (ou quase perfeita) positiva entre as duas variáveis. Se o coeficiente for 0, significa que as duas variáveis não dependem linearmente uma da outra (No entanto, pode existir uma dependência não linear, e este resultado deve ser investigado). Por fim, o coeficiente -1 (ou próximo) indica uma correlação negativa perfeita (ou quase perfeita) entre as duas variáveis – se uma aumenta, a outra diminui sempre.

3. **Qual a diferença entre uma correlação negativa e uma correlação positiva?**

   R.: Uma correlação positiva indica que se uma variável aumenta, a outra também (ou se uma diminui, a outra igualmente). Já uma correlação negativa explica que se uma variável aumenta, a outra diminui (e vice-versa).

   a. **Se dois atributos diminuem essencialmente à mesma taxa é uma correlação positiva ou negativa? Explique.**

      R.: Não, se ambos diminuem, independentemente da taxa, é uma correlação positiva.

4. **Como é medida a força de uma correlação? Quais os limites para essa força?**

   R.: A força de uma correção é medida pela proximidade aos valores extremos possíveis: -1 e 1 (e ao 0). Por exemplo, um valor próximo de -1 ou próximo de 1 indica uma correlação forte, enquanto que um valor próximo de 0 indica uma correlação fraca.

5. **Consegue pensar em atributos que poderiam ser interessantes incluir no dataset estudado no exemplo da aula?**

   R.: Atributos como “Altitude” (da casa) ou “Classificação energética” (conversão de A,B,C… para 1,2,3… Valores numéricos mais altos para as primeiras letras) seriam interessantes de serem incluídos no dataset.

PARTE II

Dataset utilizado: wine_quality_red (apenas para vinho tinto)
Informação relativa dados químicos e compostos do vinho

2. **Executar a operação de Data Understanding.**

R.: Input variables (based on physicochemical tests)
1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 – chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 – density
9 – pH
10 – sulphates
11 - alcohol

Output variable (based on sensory data):
12 - quality (score between 0 and 10)

4. **Documentar quais os atributos que podem influenciar ou explicar o consumo de combustível num determinado veículo (mpg). – Neste caso, influenciar a qualidade.**

R.: Analisando a qualidade, vê-se que o que mais a influencia são as variáveis "alcohol"e "volatile acidity".

Para a restante matriz, existe, por exemplo, uma forte correlação (negativa) entre o "ph" e o "fixed acidity", o que faz sentido, já que quanto maior o ph, menos acidez (e vice-versa).

Attributes	fixed acidity	volatile acidity	citric ...	resid...	chlori...	free ...	total sul...	density	pH	sulphates	alcohol	quality
fixed acidity	1	-0.256	0.672	0.115	0.094	-0.154	-0.113	0.668	-0.683	0.183	-0.062	0.124
volatile acidity	-0.256	1	-0.552	0.002	0.061	-0.010	0.076	0.022	0.235	-0.261	-0.202	-0.391
citric acid	0.672	-0.552	1	0.144	0.204	-0.061	0.036	0.365	-0.542	0.313	0.110	0.226
residual sugar	0.115	0.002	0.144	1	0.056	0.187	0.203	0.355	-0.086	0.006	0.042	0.014
chlorides	0.094	0.061	0.204	0.056	1	0.006	0.047	0.201	-0.265	0.371	-0.221	-0.129
free sulfur dioxide	-0.154	-0.010	-0.061	0.187	0.006	1	0.667	-0.022	0.071	0.052	-0.070	-0.051
total sulfur dioxide	-0.113	0.076	0.036	0.203	0.047	0.667	1	0.071	-0.066	0.043	-0.206	-0.185
density	0.668	0.022	0.365	0.355	0.201	-0.022	0.071	1	-0.342	0.149	-0.496	-0.175
pH	-0.683	0.235	-0.542	-0.086	-0.265	0.071	-0.066	-0.342	1	-0.197	0.206	-0.058
sulphates	0.183	-0.261	0.313	0.006	0.371	0.052	0.043	0.149	-0.197	1	0.094	0.251
alcohol	-0.062	-0.202	0.110	0.042	-0.221	-0.070	-0.206	-0.496	0.206	0.094	1	0.476
quality	0.124	-0.391	0.226	0.014	-0.129	-0.051	-0.185	-0.175	-0.058	0.251	0.476	1