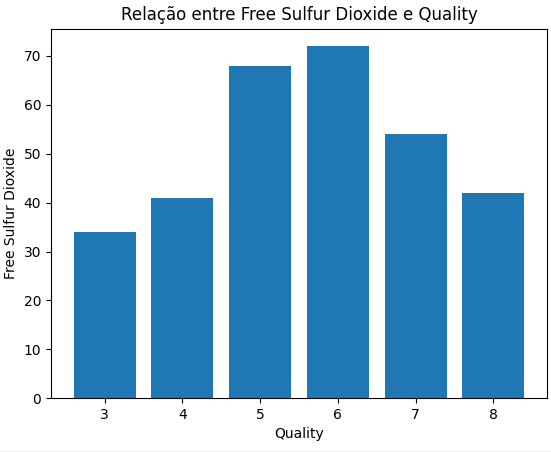
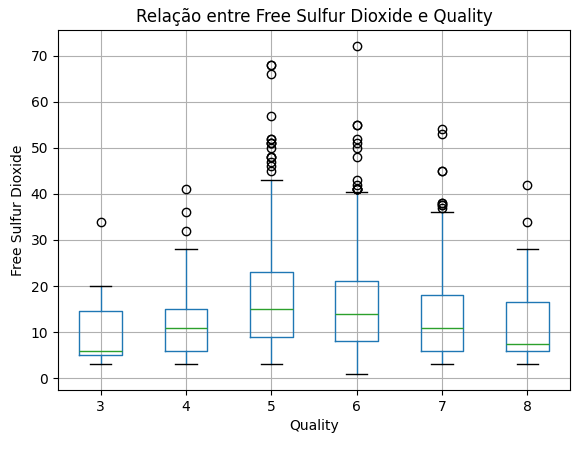


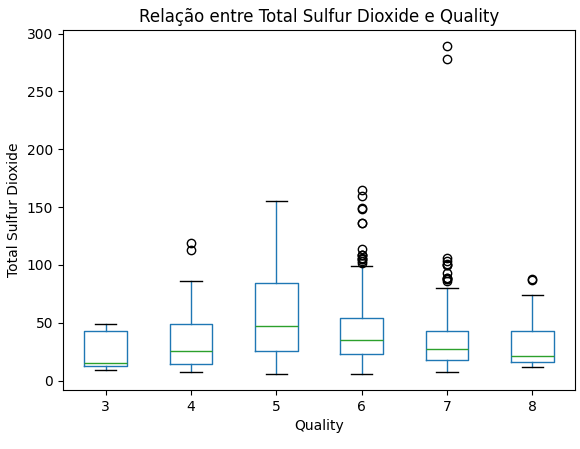
1. **Tendência Geral**:
   * **Gráfico de Barras**: Podemos observar que os níveis médios de dióxido de enxofre livre variam conforme a qualidade do vinho. Isso sugere que há uma relação entre a quantidade de dióxido de enxofre livre e a qualidade percebida do vinho. 
2. **Variabilidade e Consistência**:
   * **Box Plot**: A variabilidade dos níveis de dióxido de enxofre livre é maior em vinhos de qualidade inferior (qualidade 3 e 4), como indicado pelas caixas e whiskers mais longos. Isso pode indicar inconsistências na produção desses vinhos.
   * Vinhos de qualidade superior (qualidade 6, 7 e 8) mostram uma menor variabilidade, sugerindo uma produção mais consistente e controlada.



1. **Outliers**:
   * **Box Plot**: Existem vários outliers em diferentes níveis de qualidade, indicando que alguns vinhos têm níveis de dióxido de enxofre livre significativamente diferentes da maioria dos vinhos na mesma categoria de qualidade. Isso pode ser relevante para identificar vinhos que não seguem o padrão esperado.
2. **Distribuição dos Dados**:
   * **Box Plot**: A distribuição dos níveis de dióxido de enxofre livre é mais ampla em vinhos de qualidade inferior, enquanto vinhos de qualidade superior tendem a ter uma distribuição mais concentrada. Isso pode indicar que a qualidade do vinho está associada a um controle mais rigoroso dos níveis de dióxido de enxofre livre.

**Aplicações Práticas:**

* **Para Produtores de Vinho**: Esses insights podem ajudar a ajustar os processos de produção para melhorar a consistência e a qualidade do vinho. Reduzir a variabilidade nos níveis de dióxido de enxofre livre pode ser uma meta para alcançar uma qualidade mais alta.
* **Para Consumidores**: Entender a relação entre dióxido de enxofre livre e qualidade pode ajudar na escolha de vinhos com base em preferências pessoais e preocupações de saúde. Consumidores podem preferir vinhos com níveis mais baixos de dióxido de enxofre livre, associados a uma qualidade superior.

Este **boxplot** que mostra a relação entre o dióxido de enxofre total (Total Sulfur Dioxide) e a qualidade (Quality) do vinho.  
 

Aqui estão os principais pontos para entender o gráfico:

1. **Eixo X (Quality)**: Representa a qualidade do vinho, variando de 3 a 8.
2. **Eixo Y (Total Sulfur Dioxide)**: Representa a quantidade total de dióxido de enxofre presente no vinho, variando de 0 a 250.

**Componentes do Boxplot:**

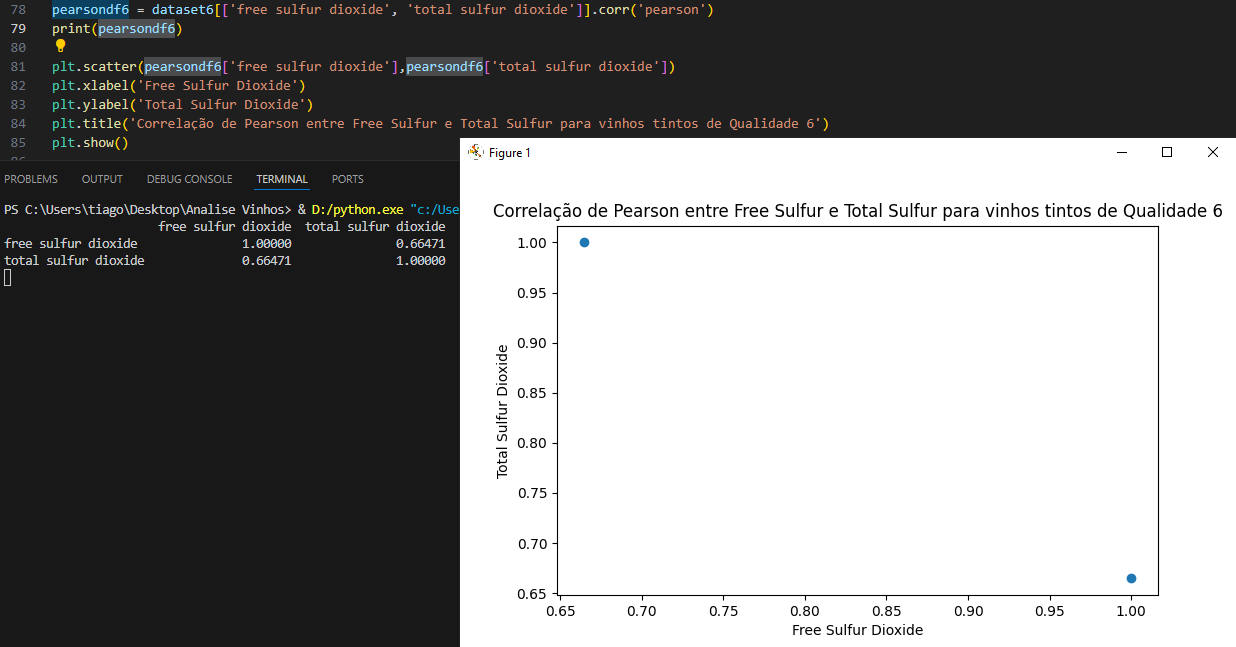
* **Caixa (Box)**: Representa o intervalo interquartil (IQR), que é a diferença entre o 1º quartil (Q1) e o 3º quartil (Q3). A caixa contém 50% dos dados.
* **Linha no meio da caixa**: Indica a mediana dos dados.
* **Extensões (Whiskers)**: Mostram a extensão dos dados, excluindo os outliers.
* **Pontos fora da caixa**: São os outliers, valores que estão fora do intervalo esperado.

**Interpretação:**

* **Distribuição dos Dados**: A maioria dos valores de dióxido de enxofre total está concentrada dentro das caixas, com alguns outliers acima e abaixo.
* **Mediana**: A linha dentro de cada caixa mostra a mediana do dióxido de enxofre total para cada nível de qualidade.
* **Outliers**: Os pontos fora das caixas indicam valores atípicos de dióxido de enxofre total.

**Observações:**

* Para vinhos de qualidade 5, há uma maior variação nos valores de dióxido de enxofre total, com alguns outliers significativos.
* Vinhos de qualidade 6 e 7 parecem ter uma distribuição mais concentrada, com menos outliers.



 A saída do print(pearsondf6) confirma a correlação de Pearson entre free sulfur dioxide e total sulfur dioxide para vinhos de qualidade 6:

* **Correlação entre**free sulfur dioxide**e**total sulfur dioxide: 0.66471

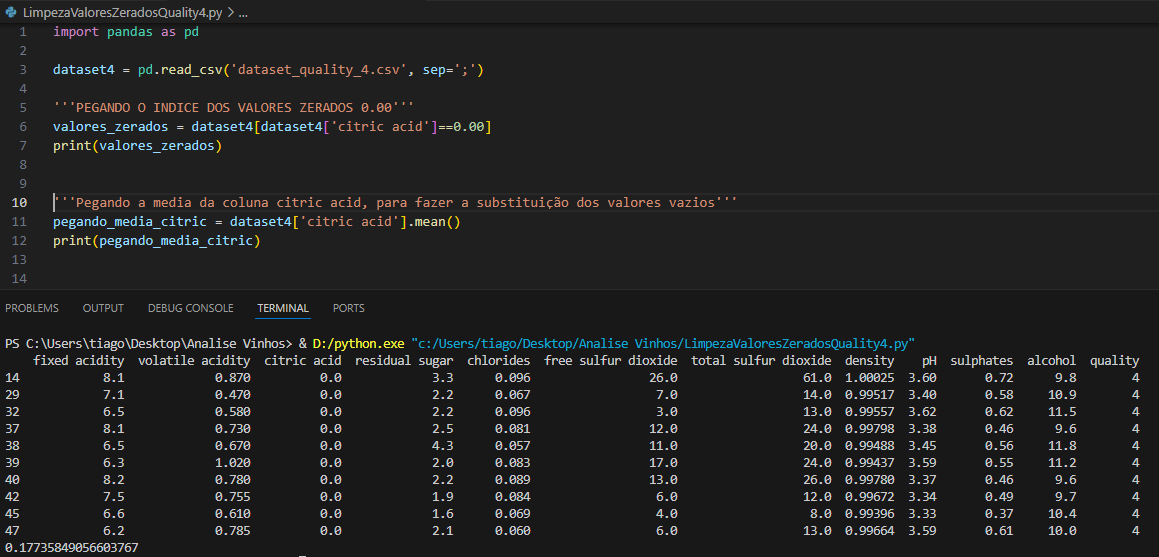
**Interpretação**

* **Correlação Moderada**: O valor de 0.66471 indica uma correlação positiva moderada entre free sulfur dioxide e total sulfur dioxide. Isso significa que, em geral, à medida que o dióxido de enxofre livre aumenta, o dióxido de enxofre total também tende a aumentar, mas não de forma perfeita.

**Conclusão**

Essa correlação moderada pode explicar a presença de outliers no boxplot para vinhos de qualidade 6, pois há variabilidade nos valores de total sulfur dioxide que não é completamente explicada por free sulfur dioxide.

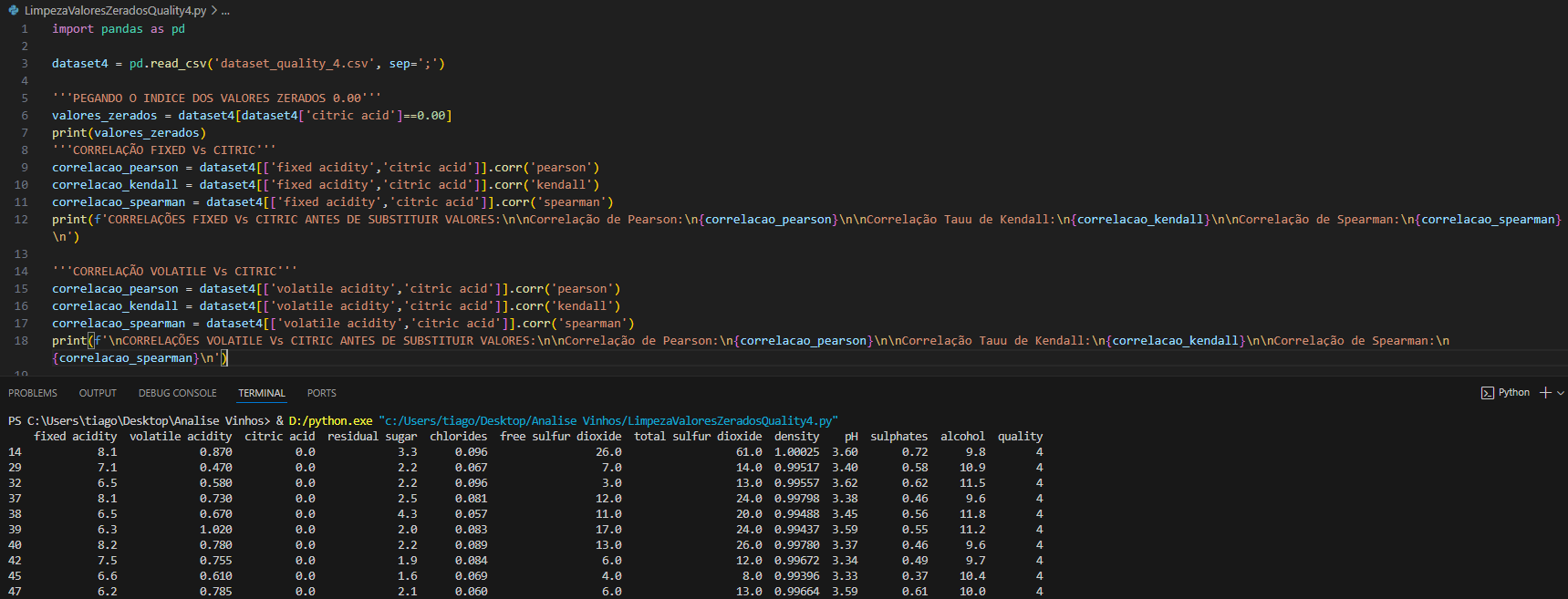
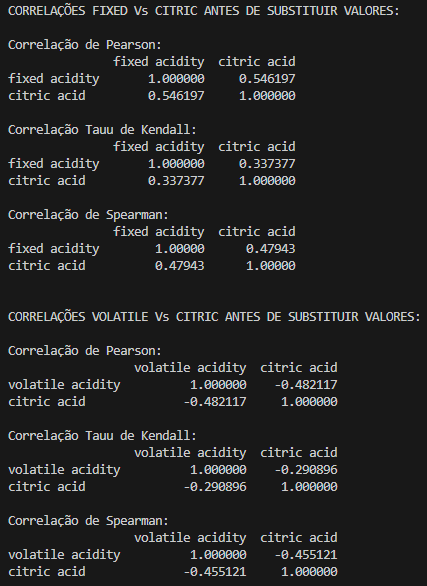
Analise sobre vinhos de qualidade 4:

Efetuando algumas analises sobre o dataset (*winequality-red.csv*), acabamos por encontrar (Eu, Felipe), que a coluna “*Citric Acid*”, apresentava valores zerados (0.00). Então seguimos pesquisando na internet qual a melhor maneira de tratar esses dados, como exclusão das linhas, ou substituir por uma média em comum, ou ate mesmo substituindo os valores individualmente para cada nova média adquirida.  


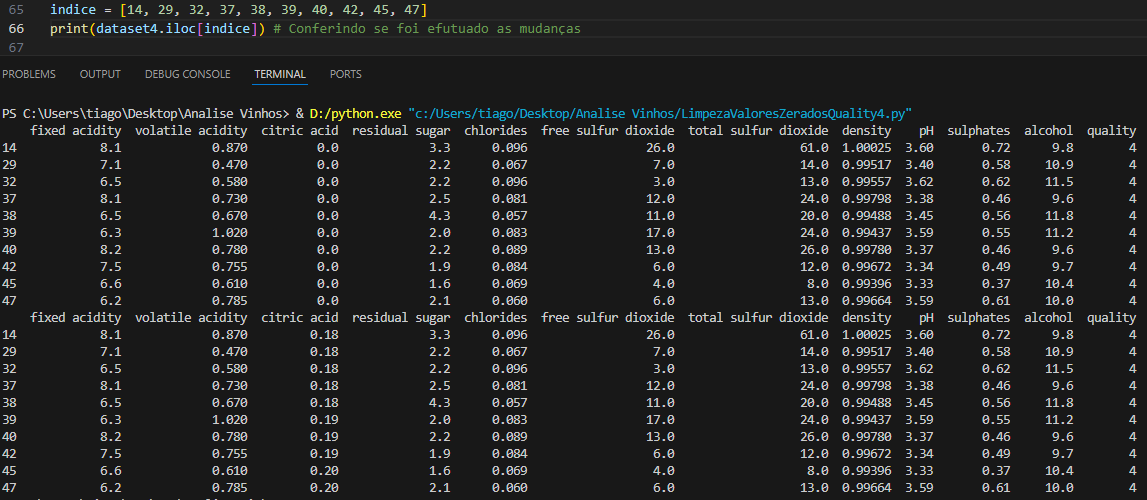
**Legenda**:  
**1º. Seta Verde**: Definindo a coluna citric acid e os valor que estou buscando que neste caso é o valor (=0).

**2º. Seta Laranja**: O retorno da pesquisa com todas as linhas que apresentam o valor 0’zero’, na coluna citric acid.

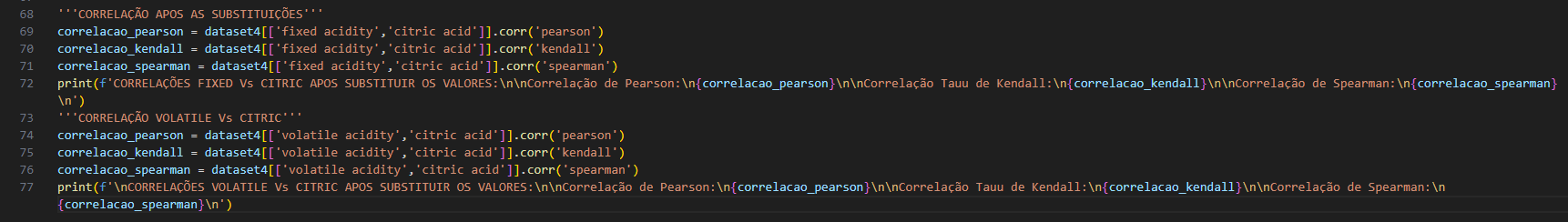
**3º. Seta Amarelo**: Retorno do valor médio contido na coluna citric acid, o valor será arredondado para 0.18 e inserido no índice 14.

**Correlações antes de fazer as substituições:**  


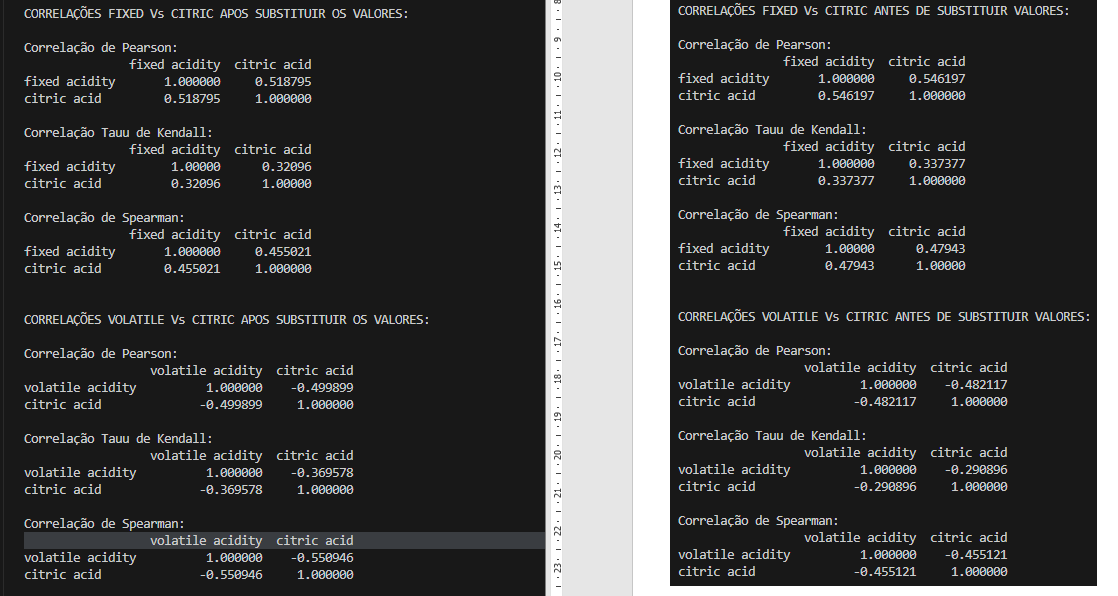
**Conferindo as mudanças feitas:**



**Verificando as correlações com a substituições feitas:**



**Comparando as correlações:**



**Observações entre as Correlações entre Ácidos (Antes e Depois da Substituição de Valores)**

**1.** *Correlações entre Fixed Acidity e Citric Acid*

Antes da Substituição dos Valores:

* **Correlação de Pearson**:
  + Fixed Acidity vs. Citric Acid: 0.546197
* **Correlação Tauu de Kendall**:
  + Fixed Acidity vs. Citric Acid: 0.37377
* **Correlação de Spearman**:
  + Fixed Acidity vs. Citric Acid: 0.47943

Após a Substituição dos Valores:

* **Correlação de Pearson**:
  + Fixed Acidity vs. Citric Acid: 0.518795
* **Correlação Tauu de Kendall**:
  + Fixed Acidity vs. Citric Acid: 0.32096
* **Correlação de Spearman**:
  + Fixed Acidity vs. Citric Acid: 0.45521

**2.** *Correlações entre Volatile Acidity e Citric Acid*

Antes da Substituição dos Valores:

* **Correlação de Pearson**:
  + Volatile Acidity vs. Citric Acid: -0.482117
* **Correlação Tauu de Kendall**:
  + Volatile Acidity vs. Citric Acid: -0.209896
* **Correlação de Spearman**:
  + Volatile Acidity vs. Citric Acid: -0.455121

Após a Substituição dos Valores:

* **Correlação de Pearson**:
  + Volatile Acidity vs. Citric Acid: -0.499899
* **Correlação Tauu de Kendall**:
  + Volatile Acidity vs. Citric Acid: -0.369578
* **Correlação de Spearman**:
  + Volatile Acidity vs. Citric Acid: -0.559046

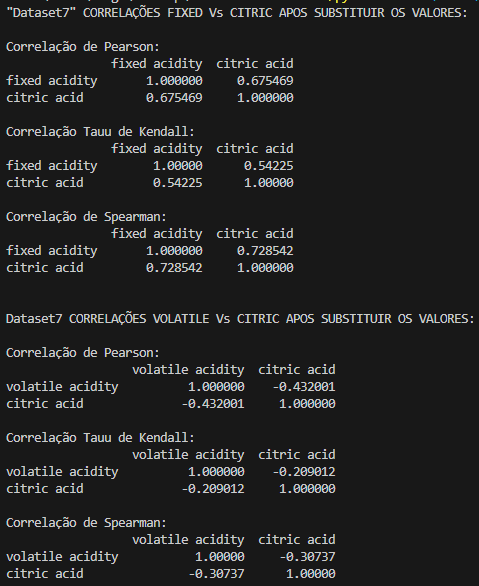
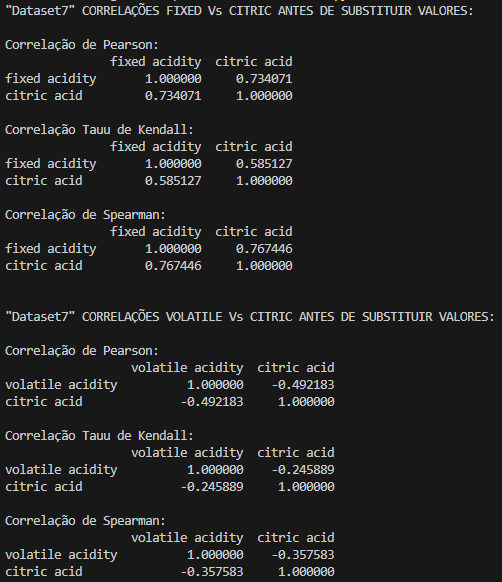
**Análise**

1. *Correlação entre Fixed Acidity e Citric Acid*:
   * A correlação de Pearson diminuiu ligeiramente de 0.546197 para 0.518795 após a substituição dos valores.
   * A correlação Tauu de Kendall também diminuiu de 0.37377 para 0.32096.
   * A correlação de Spearman seguiu o mesmo padrão, diminuindo de 0.47943 para 0.45521.
2. *Correlação entre Volatile Acidity e Citric Acid*:
   * A correlação de Pearson entre Volatile Acidity e Citric Acid se tornou ligeiramente mais negativa após a substituição, de -0.482117 para -0.499899.
   * A correlação Tauu de Kendall apresentou uma diminuição mais acentuada de -0.209896 para -0.369578.
   * A correlação de Spearman também mostrou uma diminuição considerável, de -0.455121 para -0.559046.

**Conclusão**

As substituições de valores influenciaram as correlações entre os ácidos fixos e cítricos, assim como entre os ácidos voláteis e cítricos. No geral, as correlações enfraqueceram após as substituições, tanto para as correlações positivas (fixed acidity e citric acid) quanto para as negativas (volatile acidity e citric acid). Essas mudanças indicam que os valores substituídos têm impacto significativo na relação entre as variáveis analisadas.

**Dataset7:**

 **Antes de substituir valores: Depois de substituir:**