#Importando as Bibliotecas  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
import numpy as np  
import numpy as np  
  
# Lendo arquivo  
vinhos = pd.read\_csv('wine.csv')  
print(vinhos.info())  
print(vinhos.isna().sum()) # Somando quantidade de nulos no dataframe  
  
#removendo valores nulos caso nao tenha nenhum dado faltante  
vinhos.dropna(inplace=True, thresh=1 , axis=1)  
vinhos.fillna(value=0, inplace=True) # Alterando dados nulos para 0!  
  
# Alterando o nome das colunas  
vinhos.columns = ['id','pais','descricao','designacao','pontos','preco','provincia','regiao1','regiao2','nome\_provador','twitter\_provador','nome\_vinho','variedade','vinicola']  
print(vinhos.info())  
  
#Nova coluna onde tiramos os anos da coluna do titulo do vinho.  
vinhos['ano'] = vinhos['nome\_vinho'].str.extract(r'(\d{4})', expand=False)  
vinhos['ano'] = vinhos['ano'].fillna(0).astype(int)  
print(vinhos.info())  
  
#Criando outra coluna e filtrando dados de 1950 á 2024  
vinhos['anos'] = vinhos['ano'].where((vinhos['ano'] >= 1950) & (vinhos['ano'] <= 2024))  
vinhos['anos'] = vinhos['anos'].fillna(0).astype(int)  
print(vinhos['anos'])  
  
sns.set\_style("whitegrid") # Inserindo um tema nos graficos  
  
  
#==============================================================================================================  
#1 Qual é o país com o maior número de vinhos no dataset?  
  
# Criar a contagem dos vinhos por país  
contagem = vinhos['pais'].value\_counts().head(5)  
contagem = pd.DataFrame(contagem).reset\_index()  
contagem.columns = ['pais', 'total']  
print(contagem)  
# Criar o gráfico de barras  
plt.title('(TOP 5) - Total de Vinhos por País', color="firebrick", fontsize=16)  
ax = sns.barplot(x='total', y='pais', data=contagem, palette='rocket')  
# Adicionar as etiquetas nas barras  
for container in ax.containers:  
 ax.bar\_label(container)  
# Ajustar labels dos eixos  
plt.xlabel("Total", fontsize=14, color="firebrick")  
plt.ylabel("Países", fontsize=14, color="firebrick")  
# Ajustar o layout para evitar sobreposição  
plt.tight\_layout()  
plt.show()  
  
  
#-------------------------------------------  
#2 Qual é a média de pontos dos vinhos por país?  
  
media= vinhos[['pontos','pais']].groupby('pais').mean().sort\_values(by='pontos', ascending=False).reset\_index()  
media = media[media['pais'] != 0] # Tirando os dados com 0 na tabela pais  
fig, ax = plt.subplots(figsize=(6,4))  
# Removendo os eixos  
ax.axis('tight')  
ax.axis('off')  
# Criando a tabela  
table = ax.table(cellText=media.values, colLabels=media.columns, cellLoc='center', loc='center')  
# Personalizando a tabela  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(14)  
table.scale(1.2, 1.2)  
plt.show()  
  
  
#===================================  
#3 Existe uma correlação entre o preço e os pontos dos vinhos?  
  
sns.residplot(data=vinhos, x='pontos', y='preco', lowess=True, color="midnightblue")  
plt.title('Correlação De Preços e Pontos Dos Vinhos', color = "firebrick", fontsize= 16)  
plt.xlabel("Pontos",fontsize=14,color = "firebrick")  
plt.ylabel("Preços",fontsize=14,color = "firebrick")  
plt.show()  
  
#4 Quais regiões têm os vinhos com as maiores classificações?  
maxpontos= vinhos[['pontos','regiao1']].groupby('regiao1').max().sort\_values(by='pontos', ascending=False).reset\_index().head(11)  
maxpontos = maxpontos[maxpontos['regiao1'] != 0]  
fig, ax = plt.subplots(figsize=(6,4))  
# Removendo os eixos  
ax.axis('tight')  
ax.axis('off')  
# Criando a tabela  
plt.title('(TOP 10) - Pontuação por Região', color = "firebrick", fontsize= 16)  
table = ax.table(cellText=maxpontos.values, colLabels=maxpontos.columns, cellLoc='center', loc='center')  
# Personalizando a tabela  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(14)  
table.scale(1.2, 1.2)  
plt.show()  
  
#===========================================================================  
#5 Qual é a variedade de vinho mais comum no dataset?  
  
plt.title('(TOP 3) - Variedades de Vinhos Mais Vendido', color = "firebrick", fontsize= 16)  
contagem = vinhos['variedade'].value\_counts()  
contagem = pd.DataFrame(contagem)  
ax = sns.barplot(contagem.head(3), x= 'variedade', y= 'count', palette='rocket')  
for container in ax.containers:  
 ax.bar\_label(container)  
plt.xlabel("Pontos",fontsize=14,color = "firebrick")  
plt.ylabel("Total",fontsize=14,color = "firebrick")  
plt.subplots\_adjust(left=0.30) # Ajusta 30% para esquerda  
plt.show()  
  
#=======================================================================  
#6 Como o preço dos vinhos varia entre os diferentes países?  
#ajustadar espaço do eixo onde ficam os paises  
  
medianapreco = vinhos[['pais','preco']].groupby('pais').median().sort\_values(by='preco', ascending=False)  
plt.figure(figsize=(8,6))  
plt.title('Visualizando a Mediana Dos Preços por País', color = "firebrick", fontsize= 16)  
sns.barplot(x='preco',y='pais', data=medianapreco,palette='rocket')  
plt.xticks(rotation = 90)  
plt.xlabel("Preço",fontsize=14,color = "firebrick")  
plt.ylabel("Paises",fontsize=14,color = "firebrick")  
plt.tight\_layout() # Ajusta o layout para que tudo caiba bem  
 # Ajusta a parte inferior  
plt.show()  
  
#7 Existe alguma tendência nos pontos dos vinhos ao longo dos anos?  
def anos (ele):  
 for x in ele.split():  
 try:  
 if x.isnumeric() and int(x) < 2024 and int(x) > 1700:  
 return x  
 except:  
 print(f' Não tem numero {ele}')  
  
vinhos['anos'] = vinhos['nome\_vinho'].apply(anos)  
#print(vinhos[['pontos','anos']].groupby('anos').median())  
plt.figure(figsize=(8,8))  
plt.title('Correlação De Pontos Ao Longo dos Anos', fontsize= 16, color = "firebrick")  
df\_vinhos = vinhos[['pontos', 'anos']].groupby('anos').agg(  
 {'pontos':['median','mean']}).sort\_values(by = 'anos', ascending=True)  
print(df\_vinhos)  
sns.lineplot(data = df\_vinhos, x = df\_vinhos.index, y = df\_vinhos['pontos']['median'], color='firebrick', label='Mediana')  
sns.lineplot(data = df\_vinhos, x = df\_vinhos.index, y = df\_vinhos['pontos']['mean'], color='midnightblue', label='Mean')  
plt.xticks(rotation= 90)  
plt.xlabel("Pontos",fontsize=14,color = "firebrick")  
plt.ylabel("Anos",fontsize=14,color = "firebrick")  
plt.tight\_layout()  
plt.show()  
  
#9 Quais são os 5 vinhos com a maior pontuação?  
  
pontosmax = vinhos[['nome\_vinho','pontos']][vinhos['pontos'] == vinhos['pontos'].max()].sort\_values(by='pontos', ascending=False).head(5)  
print(pontosmax)  
#pontosmax = pontosmax[pontosmax['regiao1'] != 0]  
fig, ax = plt.subplots(figsize=(12,2))  
# Removendo os eixos  
ax.axis('tight')  
ax.axis('off')  
# Criando a tabela  
plt.title('(TOP 5) - Os Vinhos Melhores Avaliados', color = "firebrick", fontsize= 16)  
table = ax.table(cellText=pontosmax.values, colLabels=pontosmax.columns, cellLoc='center', loc='center')  
  
plt.title('(TOP 5) Melhores Vinhos')  
# Personalizando a tabela  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(14)  
table.scale(1.2, 1.2)  
plt.tight\_layout()  
plt.show()  
  
#10 Como a pontuação média dos vinhos varia por região?  
  
plt.title('Variação de Media de Pontuação Por Região', color = "firebrick", fontsize= 16)  
media = vinhos[['pontos','regiao1']].groupby(by='regiao1').mean().sort\_values(by='pontos', ascending=False)  
media = pd.DataFrame(media)  
sns.lineplot(data=media.head(10), x='regiao1',y='pontos', color = "midnightblue", marker = '^', linestyle = '--', markerfacecolor = 'black', markersize= 8)  
plt.xticks(rotation = 45)  
plt.xlabel("Região",fontsize=14,color = "firebrick")  
plt.ylabel("Pontos",fontsize=14,color = "firebrick")  
plt.grid(visible=False) #Removendo grade grafico  
plt.tight\_layout()  
plt.show()  
  
#11 Existe uma diferença significativa nos preços dos vinhos entre as diferentes variedades?  
  
plt.title('Variedade x Preços', color = "firebrick", fontsize= 16)  
precovar = vinhos[['variedade','preco']].groupby(by='variedade').median().sort\_values(ascending=False, by='preco').head(10)  
sns.barplot(data=precovar, x='preco', y='variedade', palette= "rocket")  
plt.xlabel("Preço",fontsize=14,color = "firebrick")  
plt.ylabel("Varidade",fontsize=14,color = "firebrick")  
plt.tight\_layout()  
plt.show()  
  
  
#12 Qual é a distribuição dos preços dos vinhos?  
  
  
mediapreco = vinhos[['preco', 'regiao1']].groupby('regiao1', as\_index=False).median().sort\_values(by='preco', ascending=False).head(10)  
mediapreco = mediapreco[mediapreco['regiao1'] != 0]  
print(mediapreco)  
  
plt.title('Distribuição de Preço por Regiões', color = "firebrick", fontsize= 16)  
sns.kdeplot(data=mediapreco, x='preco', color="midnightblue")  
plt.title("Distribuição dos Preços dos Vinhos")  
plt.xlabel("preco",fontsize=14,color = "firebrick")  
plt.ylabel("Densidade",fontsize=14,color = "firebrick")  
plt.grid(visible=False)  
plt.tight\_layout() # Ajusta o layout para que tudo caiba bem  
plt.show()  
  
# 13 Como a média de pontos varia entre os diferentes tipos de vinho?  
mediapontos = vinhos[['pontos', 'variedade']].groupby('variedade', as\_index=False).mean().sort\_values(by='pontos', ascending=False).head(5)  
mediapontos = pd.DataFrame(mediapontos)  
print(mediapontos)  
sns.stripplot(  
 data=mediapontos, x="variedade", y="pontos",  
 jitter=False, s=20, marker="D", linewidth=2, alpha=.1, color='black')  
plt.title('Variação Da Media De Pontos Por Tipo de Vinho', color = "firebrick", fontsize= 16) #jitter=True)  
plt.xlabel("Variedade",fontsize=14,color = "firebrick")  
plt.ylabel("Ponto",fontsize=14,color = "firebrick")  
plt.tight\_layout() # Ajusta o layout para que tudo caiba bem  
plt.show()  
  
#14 Quais vinhos têm a maior diferença entre a pontuação e o preço  
  
diferenca = vinhos[['preco','pontos','nome\_vinho']].groupby('nome\_vinho').median().sort\_values(by='preco', ascending=False).head(3)  
print(diferenca)  
plt.title('3 Maiores Diferenças Entre Preços e Pontos', color = "firebrick", fontsize= 16)  
sns.barplot(  
 data=diferenca,  
 x="pontos", y="preco", hue="nome\_vinho", palette='rocket')  
plt.tight\_layout() # Ajusta o layout para que tudo caiba bem  
plt.xlabel("Pontos",fontsize=14,color = "firebrick")  
plt.ylabel("Preço",fontsize=14,color = "firebrick")  
plt.grid(visible=False) #Removendo grade grafico  
plt.show()