|  |  |
| --- | --- |
| Data Quality **DATA CLEANING CHEAT SHEET - 2**  1) Duplicate Values 2) Missing Values 3) Invalid Values 4) Outliers  **3. Invalid Values** Remove rows based on specific conditions | **if** df['Rating']**.**all()**<**0 **&** df['Rating']**.**all()**>**5:  print(df[['Rating']])  **if** df['ROM']**.**all()**<**0 **&** df['ROM']**.**all()**>**256:  print(df[['ROM']])  **if** df['Display (cm)']**.**all()**>**30.00:  print(df[['Display (cm)']])  df['ROM'] **=** df**.**iloc[:,2]**.**replace(1,1024) |
| **4. Outliers using IQR** Identify and remove outliers based on a specific column | Q1 **=** df['ROM']**.**quantile(0.25)  Q3 **=** df['ROM']**.**quantile(0.75)  IQR **=** Q3 **-** Q1  print(IQR)  lower\_bound **=** Q1 **-** 1.5 **\*** IQR  upper\_bound **=** Q3 **+** 1.5 **\*** IQR  df **=** df[(df['ROM'] **>=** lower\_bound) **&** (df['ROM'] **<=** upper\_bound)] |