**2.Aşama – Verinin Görselleştirilmesi ve Sunumu**

Balıkesir ilinin hava kalitesini ölçmek amaçlı oluşturulan verisetini görselleştirmek için **“Anaconda”** programı kuruldu ve **“Spyder”** editörü kullanıldı.

Kütüphanelerin yüklenmesi için;

**import numpy as np, pandas as pd**

**from sklearn.preprocessing import LabelEncoder**

**from sklearn.model\_selection import train\_test\_split as split**

**from sklearn.preprocessing import StandardScaler as SS**

bu kodlar import edildi. Pandas, veri tabanından veri okunabilir ve yazılabilir. Numpy, diziler üzerinde işlem yapılabilir.

Verisetinin yüklenmesi için;

**df=pd.read\_excel('balıkesir.xlsx', header=None, names=["Tarih", "PM10(µg/m³)", "SO2(µg/m³)","NO2(µg/m³)","O3(µg/m³)","Hava\_Sıcaklığı(°C)",**

**"Ruzgar\_Yönü(Derece)"])**

**df=df.drop([0, 1], axis=0)**

Verisetindeki sütun isimleri belirtildi. Ve ilk iki satırın hesaplanmaması için kodlar yazıldı.

Veri bilimi projeleri, verinin keşfedilmesi ve temizlenmesi ile başlar ve bu işlemler projelerin en çok zaman alan kısımlarıdır. Dolayısıyla verinin keşfi ve temizlenmesi sırasında işleri kolaylaştıracak bir takım kütüphanelere ihtiyaç duyulur. Hatırlayacağınız üzere Numpy verilerle çalışmayı oldukça kolaylaştırmıştı. Numpy’ın eksik kaldığı kısımlarda ise Pandas kullanılır. Ancak Pandas Numpy’ ın bir alternatifi olarak değil, uzantısı olarak düşünülmelidir. Pandas, Numpy’ın sütun adları ve homojen olmayan verilerle çalışamama gibi eksik kaldığı kısımlara ve daha fazlasına çözümler üretir. Pandas ile veri analizi yaparken kullanacağımız temel veri yapıları Seriler ve DataFrame’lerdir.

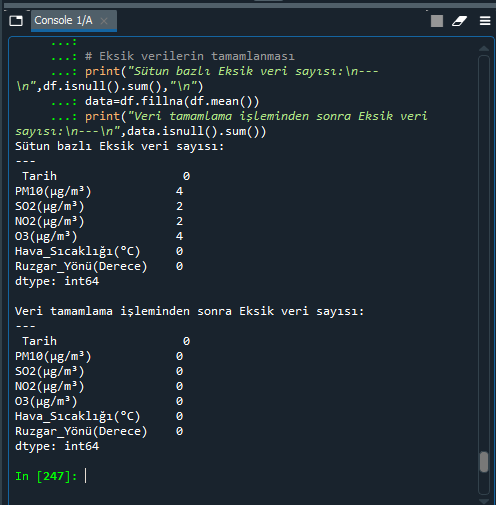
Eksik verilerin tamamlanması için;

**print("Sütun bazlı Eksik veri sayısı:\n---\n",df.isnull().sum(),"\n")**

**data=df.fillna(df.mean())**

**print("Veri tamamlama işleminden sonra Eksik veri sayısı:\n--- \n",data.isnull().sum())**

kodlar çalıştırıldığında Şekil\_2’ deki gibi çıktıyı vermektedir.

****

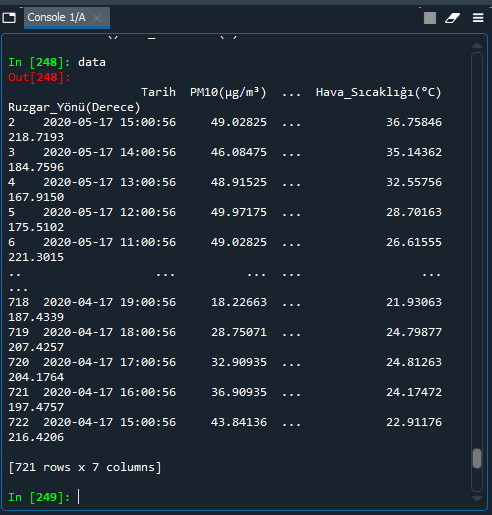
Şekil\_2

Değerlerin standartlaştırılması

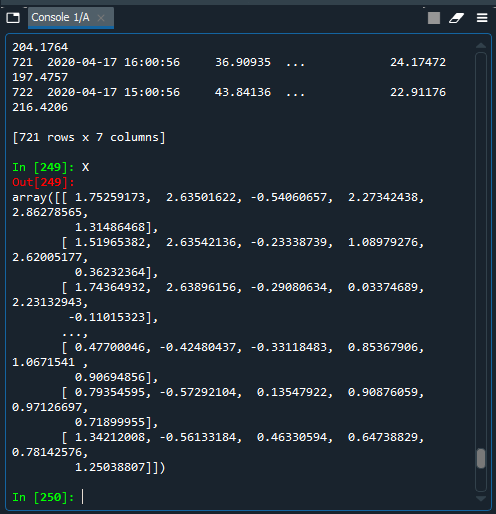
**X=data.drop(["Tarih"],axis=1)**

**X=SS().fit\_transform(X)**

Şekil\_3’ te verilerin ilk hali gösterilmiştir. Verilerin standartlaştırılmış hali Şekil\_4’ te gösterilmiştir.

****

Şekil\_3

****

Şekil\_4