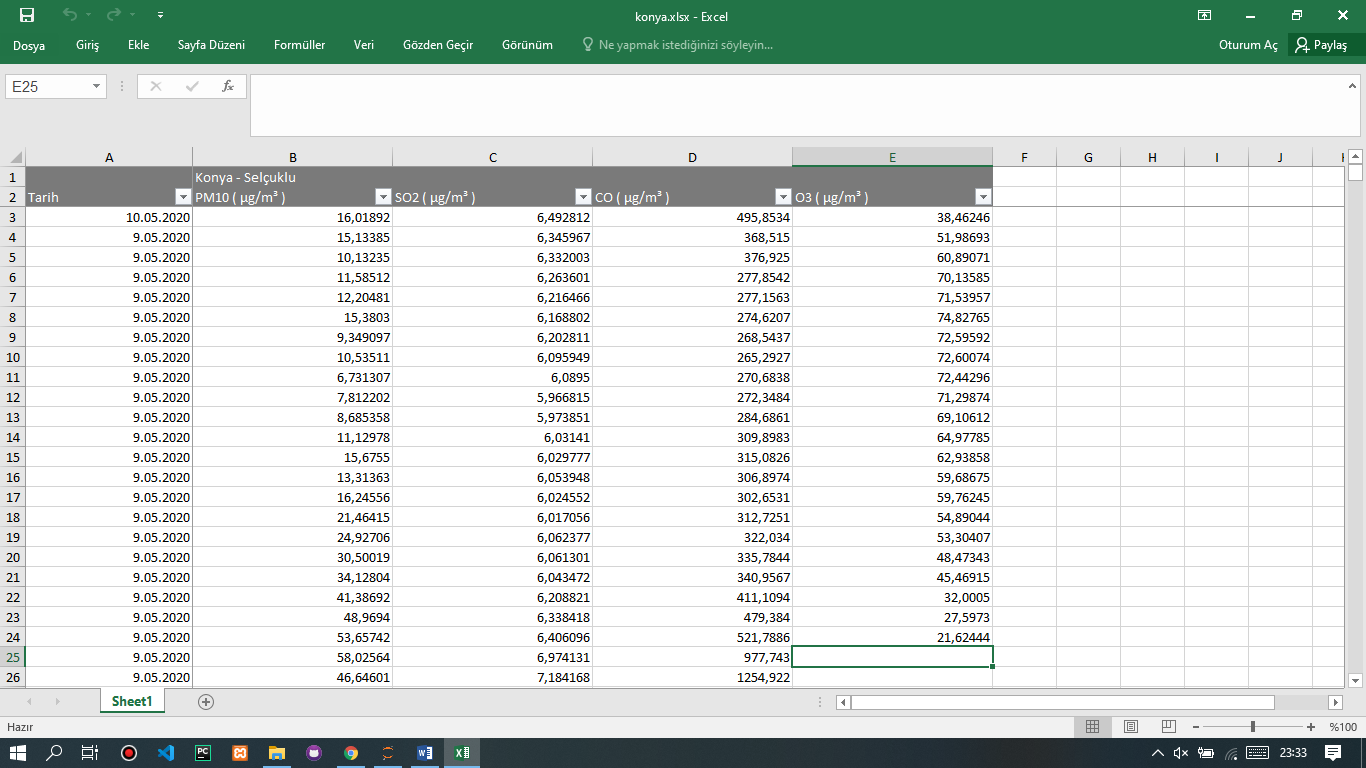
**1.Aşama – Veri Toplama**

Veri seti seçilirken öncelikle Balıkesir ili seçilmişti. Ancak veri setindeki değerler birbirine yakın olduğundan iyi bir sonuç alınmayacağı düşünüldüğü için veri seti değiştirildi. Yani değerler birbirine yakın olduğu için ilerde yapay zekâ sinir ağı ile oluşturulacak modele sadece bu değerdeki kayıtlar öğretilir. Daha büyük değerler girildiğinde model bunu tanımayacağı için yanlış değerler üretilecektir. Bu dersten kazanılan bilgilere göre daha büyük şehir seçilmesi yönünde karar verildi. Konya ili seçildi. Bütün değerleri kapsayan kayıtlar olduğu için bu veri seti ile proje devam edilecektir.

Çevre Bakanlığı portalında halka açık sunulan hava kalitesi izleme istasyonları bulunmaktadır. 81 ilde bulunan istasyonların, hava kalitesi günlük, saatlik olarak detaylı olarak göstermektedir. Burada “veri bankası” > “istasyon veri indirme” kısmında **“tipi=Hava Kalitesi”, “il=Konya”, “istasyon=Konya”** ve son olarak **“10.01.2020 – 10.05.2020”** tarih aralığındaki veriler, “**PM10 ( µg/m³ ), SO2 ( µg/m³ ), CO ( µg/m³ ), O3 ( µg/m³ )** ” olmak üzere 4 tane nitelik seçilmiştir.Bu kirleticiler havada asılı kalması hava kalitesinin değişimine sebep olur.

Bu özelliklerin seçilmesi ile sorgulanan veri seti **Şekil\_1’** deki gibi oluşturulup indirildi. Veri seti **2905 satır, 4 sütundan** oluşmaktadır. Veri setinde bulunan **4 özellik** ile Hava Kalite indeksi hesaplanarak **”Index”** sütunu oluşturulması hedefleniyor. Sınıflandırma işlemi bu sütuna göre yapılacaktır. Veri setinde eksik verilerin geldiği de gözlemlenmiştir.

**Veri setinin Görünümü**



**Şekil\_1 Veri setinin görünümü**