Kod, alışveriş merkezindeki müşterilerin (mall) yaş (Age) ve harcama puanı (Spending Score) özelliklerine göre kümeleme yapmayı, cinsiyet (Gender) özelliğini de dikkate almayı amaçlar.

Kullanılan Modüller:

* pandas: Python programlama dilinde veri manipülasyonu ve analizi için kullanılır. Veriyi dış kaynaklardan okuma ve analiz sonuçlarını dosyaya kaydetme işlemleri yapar.
* numpy (Numerical Python): Python programlama dilinde sayısal işlemler ve dizi manipülasyonları için kullanılır.
* matplotlib: Python programlama dilinde grafik, çizim ve diğer görselleştirmeleri oluşturmak için kullanılır.
* seaborn: Python programlama dilinde daha çekici ve bilgilendirici istatistiksel görselleştirmeler oluşturmak için kullanılır. Matplotlib üzerine inşa edilmiş olup, daha basit kodlarla ilgi çekici grafikler üretmek için daha yüksek bir arayüz sağlar.
* from sklearn.cluster import OPTICS: Bu ifade, sklearn (scikit-learn) paketinde bulunan cluster modülünden OPTICS sınıfını içe aktarır.
* OPTICS sınıfı, scikit-learn içinde OPTICS (Ordering Points To Identify the Clustering Structure) algoritmasının bir uygulamasıdır.
* from sklearn.preprocessing import StandardScaler: Bu ifade, sklearn (scikit-learn) paketinde bulunan preprocessing modülünden StandardScaler sınıfını içe aktarır.
* StandardScaler sınıfı, veri ön işleme için scikit-learn tarafından sağlanan bir dönüştürücüdür.

**'data'** değişkeni, verilen veri çerçevesini depolamak ve okumak için kullanılan bir değişkendir.

Kod satırı "**data = pd.read\_excel(r'C:/python/Python/Mall\_Customers.xlsx')"** bir Excel dosyasını okuyarak "data" adında bir değişkene atar.

Bu kod, "C:/python/Python/Mall\_Customers.xlsx" yolundaki Excel dosyasını pandas kütüphanesini kullanarak okur. Dosya yolu, dosyanın bilgisayarınızdaki gerçek konumuna göre değiştirilmelidir. Dosya okunduktan sonra veriler "data" değişkenine atanır ve bu değişken üzerinden ilgili veri setine erişebilirsiniz.

Kod satırında **data['Gender'] = data['Gender'].map({'Male': 0, 'Female': 1})** ifadesiyle 'Gender' sütunu string olan verilerin sayısal değerlere dönüştürülmesi sağlanır. OPTICS algoritması sayısal verileri kullanır.

Kod satırlarında **scaler = StandardScaler()** ve **data\_scaled = scaler.fit\_transform(data)** ifadeleri, sklearn.preprocessing modülünden StandardScaler sınıfı kullanılarak verilerin standartlaştırılması veya normalleştirilmesi işlemini gerçekleştirir

Kod satırında **optics = OPTICS(min\_samples=5, xi=0.05, min\_cluster\_size=0.1)** ifadesiyle sklearn.cluster modülünden OPTICS sınıfı kullanılarak OPTICS (Ordering Points to Identify the Clustering Structure) algoritması veri üzerine uygulanır.

**min\_samples**: Bir kümenin geçerli kabul edilmesi için içermesi gereken minimum nokta sayısını belirler. Bu değerin altında nokta sayısına sahip olan bir küme, gürültü veya anormallik olarak kabul edilir.

**xi**: Mesafe ve yoğunluk arasındaki bağımlılığı etkileyen bir kontrol parametresidir. Daha düşük bir değer, daha yoğun kümeler üretir.

**min\_cluster\_size**: Oluşturulabilecek minimum küme boyutunu belirler. Bu değerin altında nokta sayısına sahip olan bir küme, gürültü olarak kabul edilir.

**optics.fit(data\_scaled)** ifadesi, önceden işlenmiş veriyle OPTICS algoritmasını eğitmek için kullanılır ve elde edilen sonuçlarla kümeleme analizini gerçekleştirmemizi sağlar.

**data['Cluster'] = optics.labels**\_ ifadesi, önceden yüklediğimiz DataFrame 'data'ya 'Cluster' adında yeni bir sütun eklemek için kullanılır.

* -1 gürültü olarak kabul edilen noktaları temsil eder.
* 0, 1, 2 ise kümeleri temsil eder.

**plt.figure(figsize=(8, 6))** ifadesi, 8x6 inçlik bir görüntü nesnesi (figure) oluşturmak için kullanılır. Bu kod, çizimi göstermek için kullanılacak görüntünün boyutunu ayarlar.

**sns.scatterplot(x='Age', y='Spending Score (1-100)', hue='Cluster', data=data, palette='rainbow')** ifadesi, dağılım grafiği oluşturmak için kullanılır. Bu durumda, seaborn modülünden sns.scatterplot kullanılır.

* x='Age' ve y='Spending Score (1-100)': Bu, grafiğin x ve y eksenlerini belirler. Veri çerçevesindeki 'Age' ve 'Spending Score (1-100)' özelliklerini temsil eder.
* hue='Cluster': Bu, çizimdeki veri noktalarının renklerini belirlemek için kullanılan değişkeni tanımlar, yani 'Cluster' sütunu. Her küme farklı bir renk alır.
* data=data: Bu, çizimde kullanılacak olan DataFrame'i belirtir.
* palette='rainbow': Bu, kümeler için kullanılacak renk şemasını belirleyen isteğe bağlı bir argümandır. Bu durumda, 'rainbow' renk şemasını kullanıyoruz.

**plt.show()** : Oluşturulan çizimi görüntülemek için kullanılan bir komuttur.