

Makine Öğrenmesi ile Küresel Yaşam Maliyeti Analizi: PCA ve K-Means Uygulaması

Feyza Gülver

Published in Towards Data Science · August 2026

Dünya genelinde şehirlerin yaşam maliyetleri her geçen gün değişiyor. Peki, yüzlerce şehri kira, ulaşım, gıda ve maaş gibi 50'den fazla harcama kalemine göre gruplandırırsak nasıl bir tablo ortaya çıkar? Bu yazımda, Numbeo yaşam maliyeti veri setini kullanarak dünya şehirlerini makine öğrenmesi algoritmalarıyla nasıl profillendirdiğimi adım adım anlatacağım.

1. Veri Ön İşleme: Gürültüyü Temizlemek

Gerçek dünya verileri genellikle eksiklikler ve aşırı uç değerler (outliers) içerir. Veri setimizde 55 farklı ekonomik değişken bulunuyordu.

- **Eksik Veriler:** Ortalama yerine, uç değerlerden daha az etkilenen **medyan** yöntemi (SimpleImputer) ile dolduruldu.
- **Aykırı Değerler:** Verilerin %1 ve %99'luk persentil sınırlarında **Winsorization** işlemi uygulanarak aşırı uç fiyatlar törpülendi.
- Son olarak, algoritmaların mesafe ölçümelerini doğru yapabilmesi için tüm veriler **StandardScaler** ile standartlaştırıldı.

2. PCA ile Boyut Lanetini Aşmak

50'den fazla değişkeni aynı anda analiz etmek, makine öğrenmesi modellerinde "Boyut Laneti"ne (Curse of Dimensionality) yol açar. Bu durumu çözmek için **Temel Bileşenler Analizi (PCA)** kullandım. Verideki bilgi kaybını minimumda tutarak varyansın %90'ını açıklayacak optimum bileşen sayısını seçtim.

. . .

3. K-Means ile Şehirleri Kümelemek

Şehirleri ekonomik benzerliklerine göre gruplamak için **K-Means** algoritmasını tercih ettim. Doğru küme sayısını (k) bulmak için hem **Elbow (Dirsek)** yöntemini hem de **Silhouette Skorunu** inceledim. Grafikler k=3 noktasında optimum ayrışmayı gösterdi.

. . .

4. Sonuç: Ortaya Çıkan 3 Ekonomik Profil

Kümeleri oluşturduktan sonra sonuçları bir ısı haritası (Heatmap) ve t-SNE ile görselleştirdim. Ortaya şu 3 net profil çıktı:

- **Küme 0 (Yüksek Gelir):** Maaşların ve kiraların zirvede olduğu gelişmiş metropoller.

- **Küme 1 (Orta-Yüksek Gelir):** Gelir ve standart giderlerin dengede olduğu lokasyonlar.
- **Küme 2 (Düşük Maliyet):** Temel barınma ve pazar maliyetlerinin en düşük grupta yer aldığı gelişmekte olan bölgeler.

Eğer kodların detaylarını incelemek veya projeye katkıda bulunmak isterseniz, GitHub repoma göz atabilirsiniz. Okuduğunuz için teşekkürler!