# Satış Verileri Analizi ve Kümeleme Projesi

## 1. Proje Amacı

Bu projenin temel amacı, satış verilerini analiz ederek anlamlı içgörüler elde etmek ve bu veriler üzerinden **K-Means kümeleme** ve **zaman serisi regresyon** gibi temel makine öğrenmesi yöntemlerini uygulayarak gruplandırma ve tahminleme işlemleri yapmaktır.

Proje kapsamında:

* Verilerin ön işlenmesi (eksik veriler, tarih formatları)
* Satışların zamanla nasıl değiştiği,
* Ürün kategorileri bazında satış performansı,
* Ortalama satış miktarlarına göre ürünlerin gruplandırılması (K-Means)
* Gelecek satışların tahmini (Linear Regression)

adımları gerçekleştirilmiştir.

Kullanılan kütüphaneler: pandas, matplotlib, seaborn, sklearn, numpy

## 2. Veri Yükleme ve Temel İnceleme

Veri bir .csv dosyasından yükleniyor. Pandas kullanılarak içeri aktarılıyor:

import pandas as pd

df = pd.read\_csv("train.csv")

### 2.1. Tarih Kolonlarının Dönüştürülmesi

Tarihsel analiz yapılabilmesi için Order Date ve Ship Date sütunları datetime formatına çevriliyor:

df['Order Date'] = pd.to\_datetime(df['Order Date'], dayfirst=True)

df['Ship Date'] = pd.to\_datetime(df['Ship Date'], dayfirst=True)

### 2.2. Eksik Değerlerin Kontrolü

Veride boş (NaN) alan var mı diye kontrol ediliyor:

df.isnull().sum()

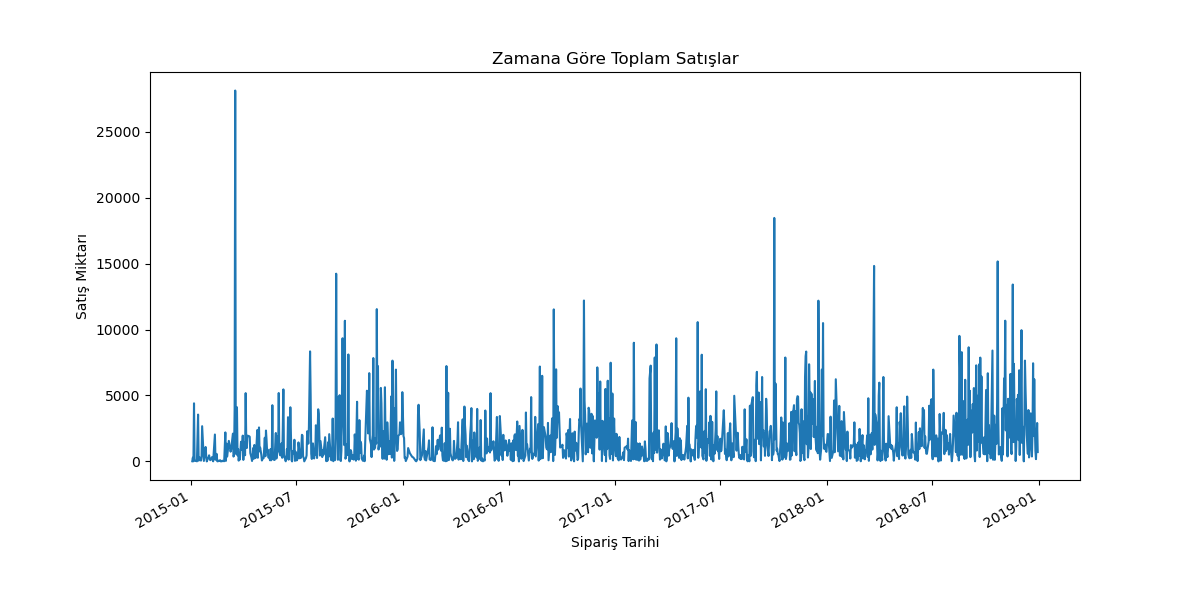
## 3. Zaman Serisi Analizi ve Görselleştirmeler

### 3.1. Aylık Satış Trendinin Analizi

Sipariş tarihine göre satışlar gruplanıyor ve zamanla nasıl değiştiği görselleştiriliyor:

df['Order Month'] = df['Order Date'].dt.to\_period('M')

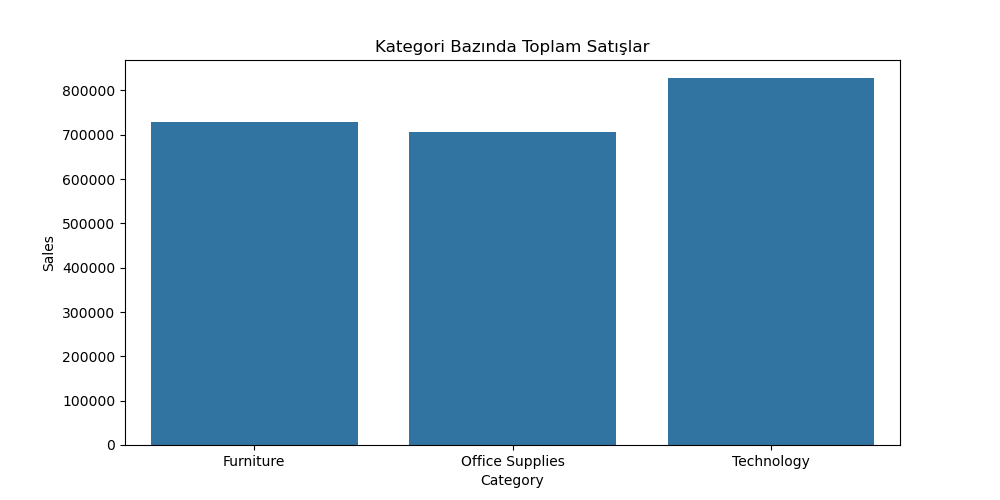
monthly\_sales = df.groupby('Order Month')['Sales'].sum().reset\_index()



### 3.2. Kategori ve Alt Kategorilere Göre Satış

Kategori ve alt kategori bazında toplam satışlar hesaplanıyor ve görselleştiriliyor:

df.groupby(['Category', 'Sub-Category'])['Sales'].sum().sort\_values(ascending=False)



## 4. K-Means Kümeleme

Bu adımda, ortalama satışlara göre kategori ve alt kategori bazında ürünler gruplandırılıyor. Amaç: benzer satış performansına sahip ürün gruplarını belirlemek.

### 4.1. Kümeleme İçin Veri Hazırlığı

Kategorilere göre ortalama satışlar alınır:

df\_cluster = df.groupby(['Category', 'Sub-Category'])['Sales'].mean().reset\_index()

df\_cluster.rename(columns={'Sales': 'Avg\_Sales'}, inplace=True)

### 4.2. Verilerin Ölçeklenmesi (StandardScaler)

Kümeleme algoritmaları, verilerin ölçeğinden etkilenir. Bu nedenle normalizasyon yapılır:

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

df\_cluster['Sales\_scaled'] = scaler.fit\_transform(df\_cluster[['Avg\_Sales']])

### 4.3. Elbow Yöntemi ile Optimal Küme Sayısı

Belirli küme sayılarında (k=1'den 10'a) algoritma çalıştırılır ve her birinin inertia değeri hesaplanır:

from sklearn.cluster import KMeans

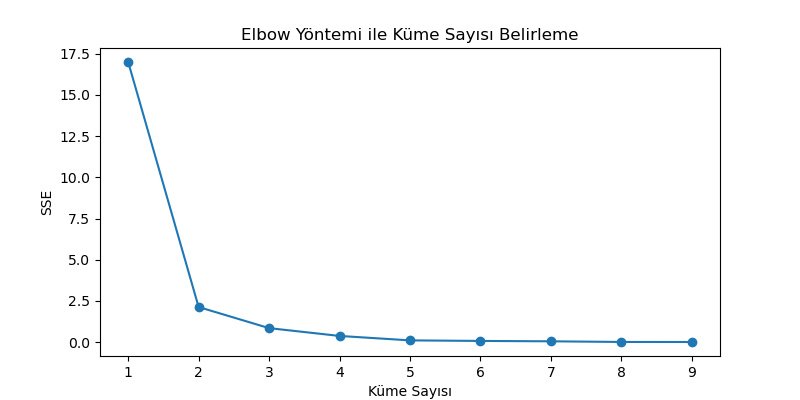
sse = []

for k in range(1, 10):

kmeans = KMeans(n\_clusters=k, random\_state=42)

kmeans.fit(df\_cluster[['Sales\_scaled']])

sse.append(kmeans.inertia\_)

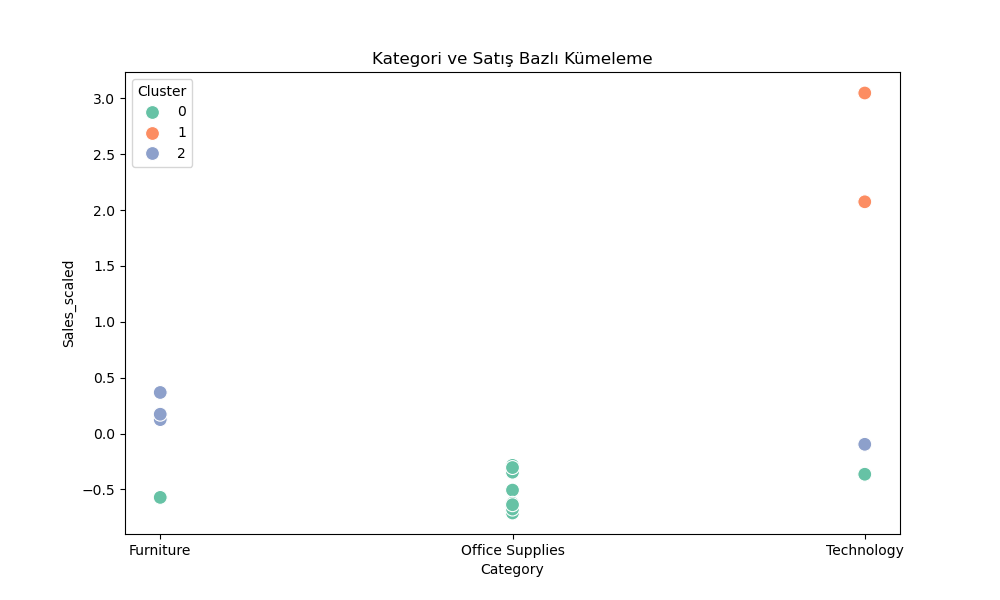


### 4.4. Kümeleme ve Görselleştirme

K=3 seçilerek K-Means modeli uygulanır:

kmeans = KMeans(n\_clusters=3, random\_state=42)

df\_cluster['Cluster'] = kmeans.fit\_predict(df\_cluster[['Sales\_scaled']])



## 5. Zaman Serisi Tabanlı Satış Tahmini

### 5.1. Aylık Satış Tahmini (Linear Regression)

Aylık satışlar, bir zaman serisi olarak ele alınarak tahminlenmiştir.

python

CopyEdit

from sklearn.linear\_model import LinearRegression

monthly\_sales['Month\_Num'] = range(len(monthly\_sales))

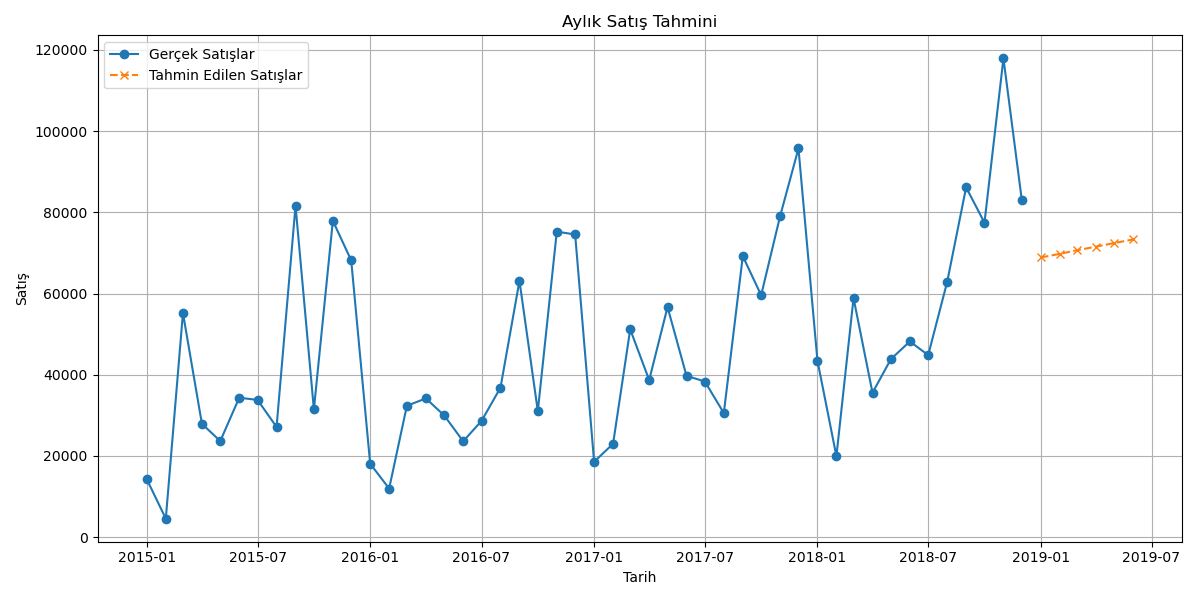
X = monthly\_sales[['Month\_Num']]

y = monthly\_sales['Sales']

model = LinearRegression()

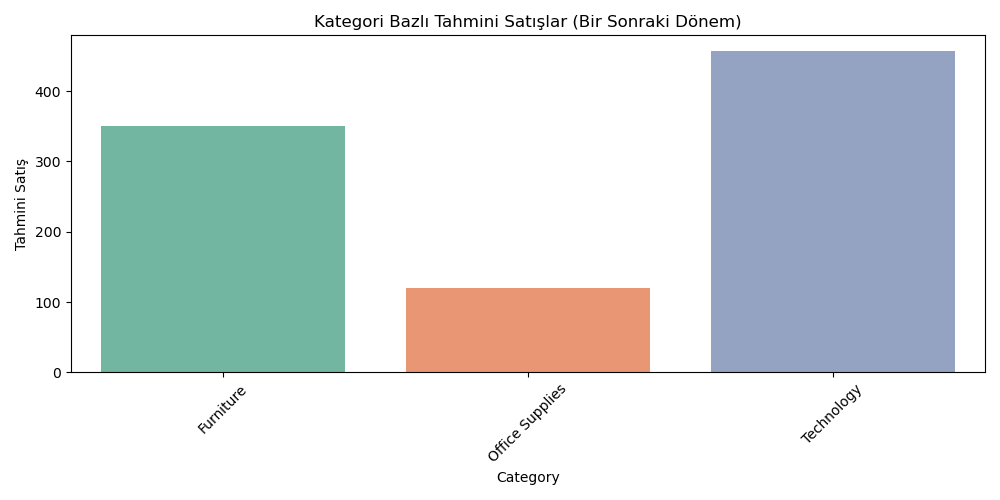
model.fit(X, y)

Model, geçmiş satış trendine göre gelecekteki satışları tahmin etmektedir.



### 5.2. Kategori Bazlı Satış Tahmini

Ayrıca, her kategori için ayrı ayrı gelecek satış tahminleri de yapılmıştır. Bu da stratejik planlama için oldukça faydalı olabilir.



## 6. Sonuç ve Değerlendirme

Bu projede satış verileri üzerinden hem istatistiksel analizler hem de temel makine öğrenmesi teknikleri uygulanmıştır:

* Satışların zamanla değişimi analiz edilmiş, hangi dönemlerde satışların arttığı görülmüştür.
* Kategori ve alt kategori bazlı satış hacimleri ortaya konmuştur.
* Ortalama satışlara göre ürünler gruplandırılarak pazar segmentasyonu yapılmıştır.
* Basit bir regresyon modeliyle gelecekteki satışlar tahmin edilmiştir.

### Genel Değerlendirme:

* Şirketler bu analizlerle **düşük performanslı ürünleri tespit edebilir**.
* Gelecekteki satış hedeflerini daha gerçekçi belirleyebilir.
* Pazarlama stratejilerini küme yapısına göre özelleştirebilir.

Yasin Deniz Cüre 132230035