

I. ÖDEVİN TANIMI

1. Python'un "Seaborn" kütüphanesinden "mpg.csv" veri setini kullanın. Veri setini 0.9/0.1 olarak eğitim ve test setlerine ayırın.
2. "horsepower", "acceleration" ve "weight" atributlarını kullanarak "mpg" değerini tahmin etmek için K-en yakın komşu (KNN), Rassal Orman (Random Forest) ve Yapay Sinir Ağı (ANN) modellerini oluşturun.
3. Modelleri test setinde deneyerek algoritmaların başarılarını karşılaştırın. Başarının karşılaştırılmasında kullandığınız kriterin formülünü ve açıklamasını raporda yazın.
4. En başarılı modeli kullanarak horsepower = 130, acceleration=13, weight=3500 olan bir otomobilin "mpg" değerini tahmin edin.
5. Programın kodunu içeren .py dosyasını ve Rapor.doc rapor dosyasını Sakai'ye yükleyin.
6. Rapor dosyasında modellerin karşılaştırma sonuçlarını içeren yorumlara, tablolara ve grafiklere yer verilmelidir.

II. KULLANILAN YÖNTEMLER VE TEKNOLOJİLERİ

Bu projede, metin verilerinin analizi ve kümeleme amacıyla çeşitli veri işleme, modelleme ve görselleştirme yöntemleri kullanılmıştır. Aşağıda, bu süreçte kullanılan temel yöntemler ve teknolojiler açıklanmıştır:

1. **Veri Seti ve Ön işleme:** Proje, Seaborn kütüphanesinden alınan "mpg" veri setini kullanmaktadır. Bu veri seti, otomobillerin performans özelliklerini içermektedir. Ön işleme adımları şunlardır:
 1. Eksik değerlerin temizlenmesi: dropna() fonksiyonu ile eksik değerler kaldırılmıştır.
 2. "horsepower", "acceleration" ve "weight" değişkenleri bağımsız; "mpg" değişkeni bağımlı değişken olarak belirlenmiştir.
 3. Veri seti, %90 eğitim ve %10 test oranında ikiye ayrılmıştır.
2. **Kullanılan Modeller:** Proje kapsamında aşağıdaki modeller kullanılmıştır:
 - *K-en Yakın Komşu (KNN):*
KNN algoritması, tahmin yapılacak örneklerin en yakın k komşusunun ortalamasını kullanır.
KNeighborsRegressor sınıfı ile oluşturulmuştur.
 - *Rassal Orman (Random Forest):*
Birden fazla karar ağacının ortalamasını alarak tahmin yapar.
RandomForestRegressor ile 100 ağaç kullanılarak modellenmiştir.
 - *Yapay Sinir Ağı (ANN):*
Çok katmanlı bir sinir ağı (MLP) kullanılmıştır.
MLPRegressor sınıfı ile iki gizli katmanda 50 nöron içeren bir model tasarlanmıştır.

3. Model Başarısının Ölçülmesi

Başarı ölçütü olarak Mean Squared Error (MSE) ve R² Skoru kullanılmıştır:

- Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- y_i : Gerçek değer
- \hat{y}_i : Tahmin edilen değer
- Küçük MSE değerleri daha iyi bir model performansını gösterir.

III. UYGULAMA

Otomobil Performans Tahmini için Makine Öğrenimi Modellerinin Karşılaştırılması

I. Giriş

Bu çalışma, **Seaborn** kütüphanesindeki **mpg** veri seti üzerinde, otomobillerin performansını belirleyen "**mpg**" (**mil başına galon**) değerini tahmin etmek amacıyla gerçekleştirilmiştir. Çalışmada, **K-en Yakın Komşu (KNN)**, **Rassal Orman (Random Forest)** ve **Yapay Sinir Ağı (ANN)** modelleri kullanılmıştır. Bu modellerin başarımı test edilmiş ve en iyi model belirlenmiştir. Ayrıca, seçilen en iyi model ile belirli bir otomobil için **mpg** tahmini yapılmıştır.

II. Uygulama Süreci

1. Veri İşleme ve Temizleme

- **Veri Seti:** Seaborn kütüphanesindeki **mpg** veri seti kullanılmıştır. Eksik veriler temizlenmiş, bağımsız değişkenler olarak **horsepower**, **acceleration**, ve **weight**, bağımlı değişken olarak **mpg** seçilmiştir.
- **Eğitim ve Test Setleri:** Veri, %90 eğitim ve %10 test olmak üzere iki gruba ayrılmıştır.

```
X = df[['horsepower', 'acceleration', 'weight']]
```

```
y = df['mpg']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42)
```

2. Modellerin Eğitilmesi

- **KNN Modeli:** Komşu sayısı **5** olarak belirlenmiş ve model eğitilmiştir.
- **Random Forest Modeli:** 100 ağaçtan oluşan bir model kullanılmıştır.
- **ANN Modeli:** İki gizli katmanlı, her biri 50 nörona sahip bir yapay sinir ağı modeli eğitilmiştir.

```
knn_model = KNeighborsRegressor(n_neighbors=5)
```

```
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
```

```
ann_model = MLPRegressor(hidden_layer_sizes=(50, 50), max_iter=1000,  
random_state=42)
```

3. Modellerin Test Edilmesi

Test seti üzerinde, **K-en Yakın Komşu (KNN)**, **Random Forest**, ve **Yapay Sinir Ağı (ANN)** modellerinin başarıları karşılaştırılmıştır. Bu karşılaştırmada kullanılan performans kriterleri:

- **Mean Squared Error (MSE):** Tahmin edilen değerler ile gerçek değerler arasındaki hata miktarını ölçer.
- **R2 Score:** Modelin, bağımlı değişken üzerindeki değişiklikleri açıklayabilme yeteneğini ölçer (0 ile 1 arasında bir değer alır, 1 en iyi durumdur).

Model	MSE	R2 Score
KNN	16.6229	0.7206
Random Forest	17.7765	0.7012
ANN	30.3521	0.4899

Yukarıdaki tabloya göre, en düşük MSE ve en yüksek R2 Score değerine sahip olan **KNN modeli**, bu veri setinde **en başarılı model** olarak belirlenmiştir.

4. Tahminler ve Karşılaştırmalar

Eğitilen modellerden en başarılı olan KNN modeli, belirli bir otomobilin yakıt tüketimini (mpg) tahmin etmek için kullanılmıştır. Tahmin edilen otomobilin özellikleri:

- horsepower = 130
- acceleration = 13
- weight = 3500

Bu özelliklerle yapılan tahmin sonucunda, KNN modeli bu otomobilin mpg değerini 17.84 olarak öngörmüştür.

IV. SONUÇLAR ve GÖRSELLEŞTİRME

Modellerin performansı, Gerçek mpg ile Tahmini mpg değerlerinin karşılaştırıldığı grafikler ve performans metrikleriyle analiz edilmiştir.

1. Modellerin Performans Grafikleri

- Gerçek ve Tahmini mpg: Üç modelin test setindeki tahminleri, gerçek değerlerle kıyaslanmış ve görselleştirilmiştir.
- MSE Karşılaştırma Grafiği: Modellerin hata oranları (MSE) çubuk grafiğiyle karşılaştırılmıştır.
- R2 Score Karşılaştırma Grafiği: Modellerin açıklayıcılık yetenekleri (R2 Score) çubuk grafiğiyle ifade edilmiştir.

2. Veri Görselleştirme

- *Korelasyon Matrisi*: Bağımsız değişkenler ve mpg arasındaki ilişkiler bir ısı haritası ile görselleştirilmiştir. Özellikle horsepower ve weight ile mpg arasında güçlü bir negatif korelasyon görülmüştür.
- *Boxplot ve Histogram*: MPG değerinin ağırlık ve menşei bazında dağılımı incelenmiştir.
- *Yoğunluk Grafiği*: Ağırlık ile MPG arasındaki ilişki yoğunluk grafiği ile görselleştirilmiş ve verilerin bu iki değişken arasında eğilim gösterdiği görülmüştür.

V. DEĞERLENDİRME

Çalışma sonucunda aşağıdaki değerlendirmelere ulaşılmıştır:

1. Model Performansları:

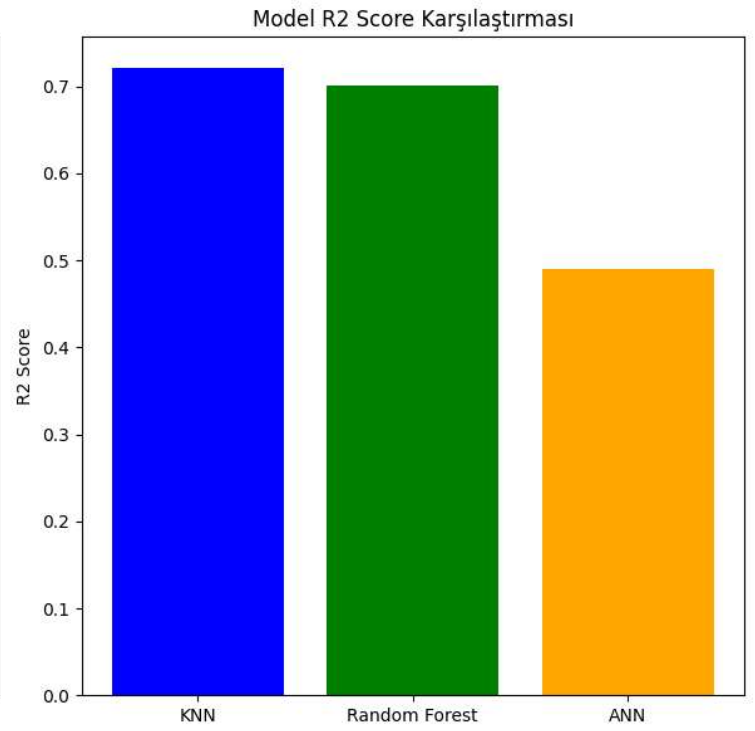
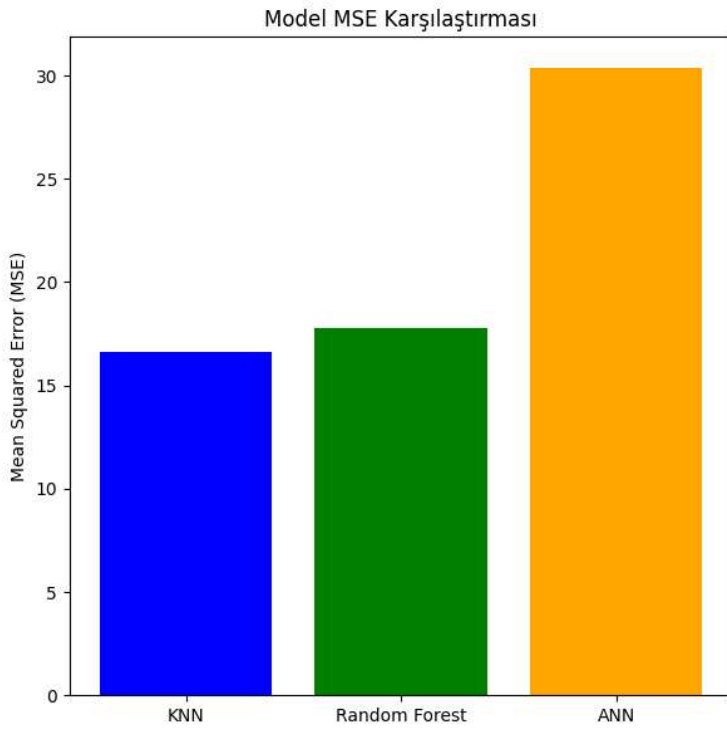
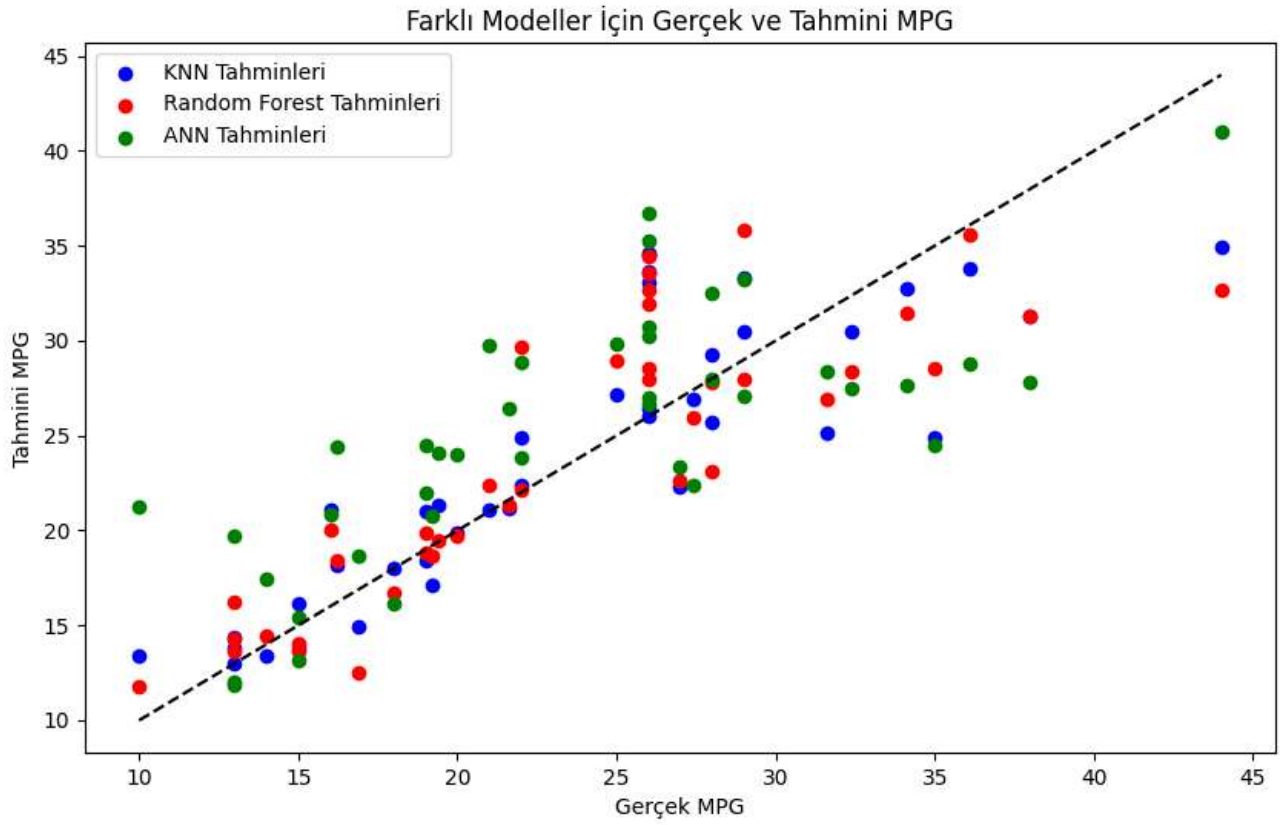
- **KNN modeli, bu veri seti ve problem için en başarılı algoritma olmuştur.** Random Forest modeli, yakın bir başarı göstermiştir. ANN modeli ise, özellikle daha fazla veriye veya parametre optimizasyonuna ihtiyaç duymaktadır.
-

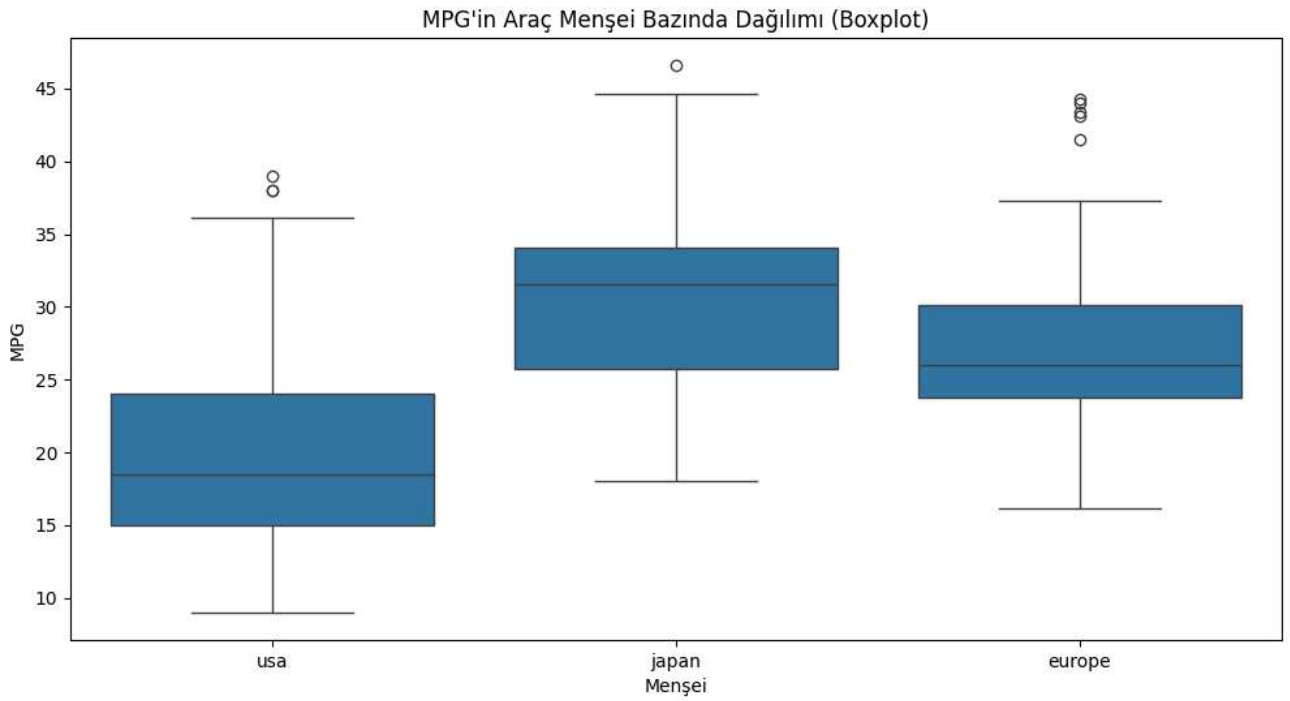
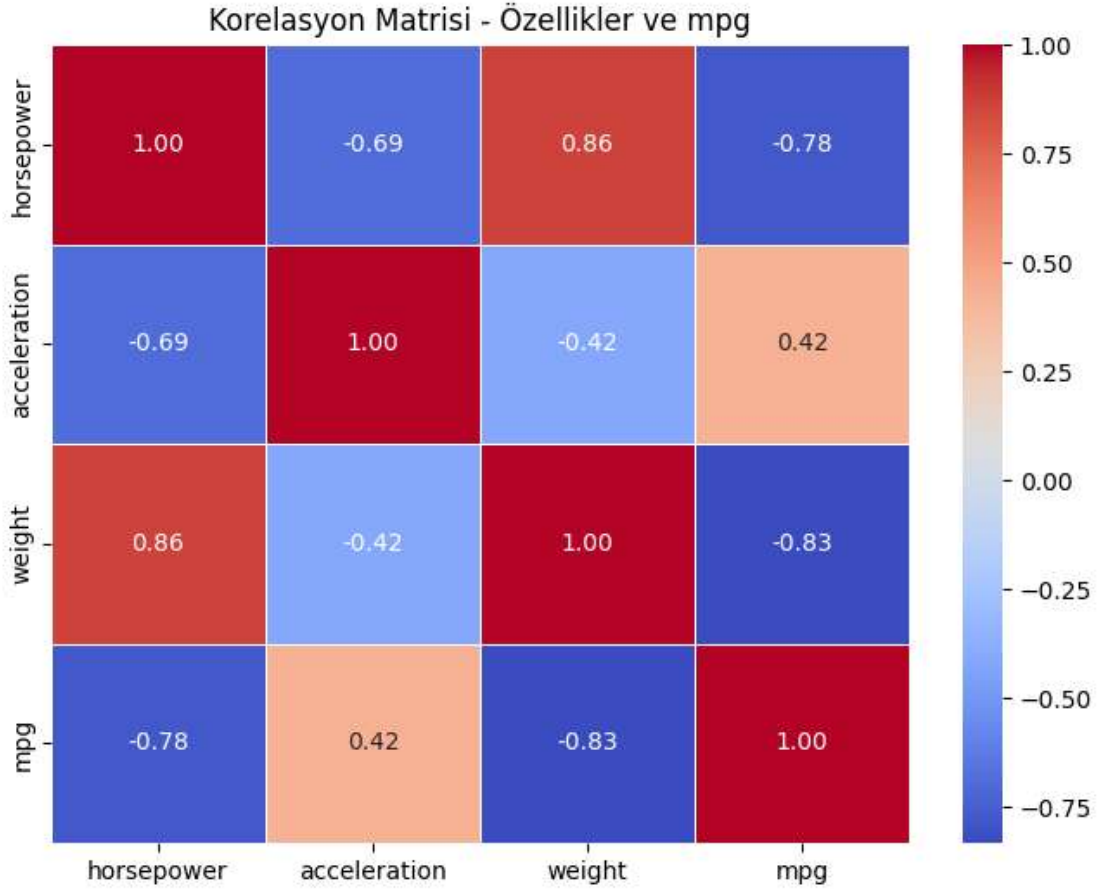
2. Özelliklerin Etkisi:

- **Weight** ve **horsepower** değişkenleri, mpg tahmininde önemli bir role sahiptir. Bu değişkenlerin negatif korelasyonu, aracın ağırlığı ve beygir gücü arttıkça yakıt tüketiminin azaldığını göstermektedir.
-

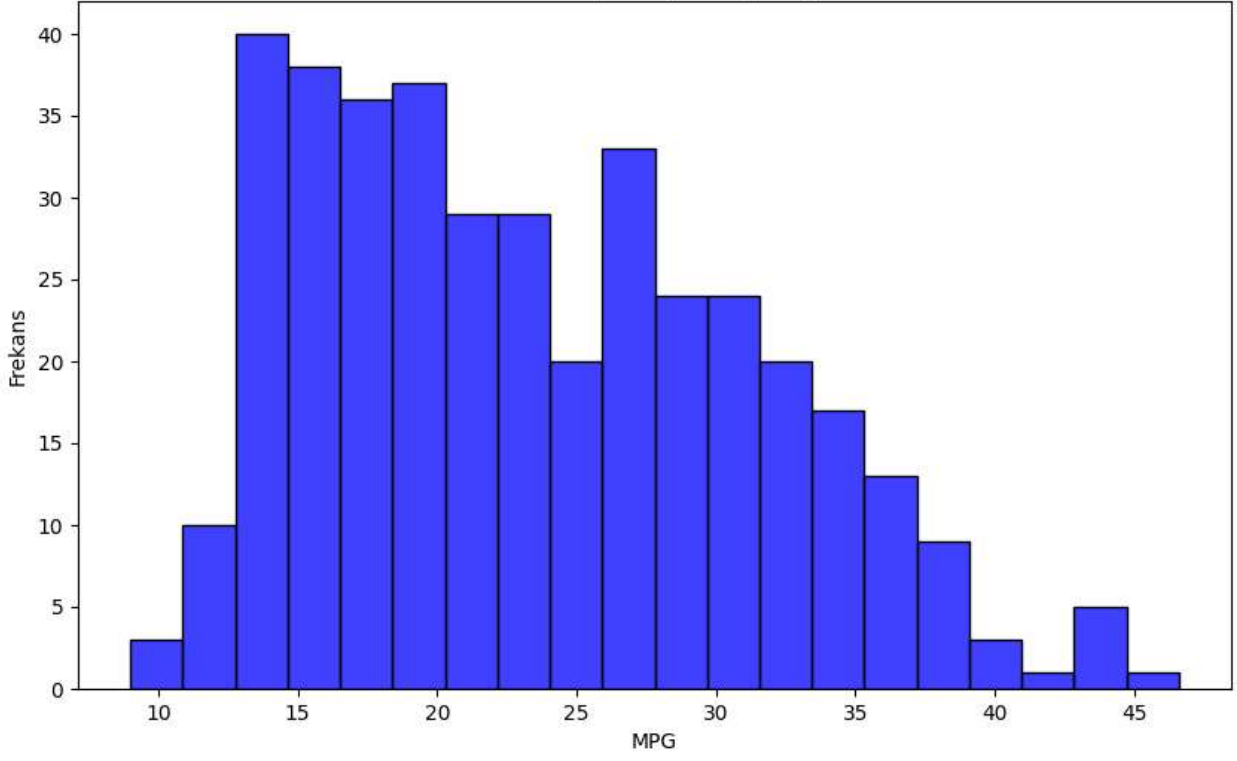
3. Görselleştirme ve Analizler:

- Histogram ve yoğunluk grafikleri, veri dağılımını ve kümelenme eğilimlerini anlamada yardımcı olmuştur. Korelasyon matrisi, bağımsız değişkenlerin mpg üzerindeki etkilerini görselleştirerek daha anlamlı yorumlar yapılmasını sağlamıştır.





MPG Dağılımı (Histogram)



Ağırlık ve MPG ilişkisi (Yoğunluk Grafiği)

