

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ
BİLGİSAYAR ORGANİZASYONU
İSTATİSTİK RAPORU

İÇİNDEKİLER

| | |
|--|----|
| ÖZET | 1 |
| İKİNCİ EL ARABA FİYAT TAHMİNİ VE İSTATİSTİK RAPORU | 2 |
| Veri Setimiz ve Grafikler | 2 |
| Veri seti çekme işlemi yapıldı..... | 2 |
| Veri Setimizi Tanıyalım..... | 2 |
| ARABALAR İLE İLGİLİ BAZI GRAFİKLER | 3 |
| Grafik_1 | 3 |
| Çubuk Grafik..... | 3 |
| Grafik_2 | 3 |
| Çubuk Grafik..... | 4 |
| Grafik_3 | 4 |
| Boxplot Grafik..... | 4 |
| Grafik_4 | 4 |
| Violinplot Grafik | 4 |
| Grafik_5 | 5 |
| Pasta (Pie) Grafik..... | 5 |
| Grafik_6 | 5 |
| Histogram Grafik | 5 |
| Grafik_7 | 6 |
| Çubuk Grafik..... | 6 |
| Grafik_8 | 6 |
| Serpme Grafik..... | 6 |
| Serpme Grafik..... | 7 |
| ARABA FİYATI TAHMİNİ MODELLERİ | 7 |
| MODELİMİZ | 9 |
| HİPOTEZLER..... | 9 |
| LINEAR REGRESSION MODELİ | 10 |
| DECISION TREE REGRESSION MODELİ | 11 |
| RANDOM FOREST REGRESSION MODELİ..... | 11 |
| MODELLERİN KARŞILAŞTIRILMASI | 12 |
| KAYNAKÇA..... | 12 |

ÖZET

Veri seti ile ilgili bir hikâyeye değinildikten sonra kullanılan veri setini tanıdık. Grafik ve grafiğe ait yorumlarına yer verildi. Modellere geçmeden önce bazı istatistiksel çalışmalardan bahsedildi. Daha sonra sırasıyla Doğrusal, Karar Ağacı, Rastgele Orman modelleri grafiklerle anlatılmıştır. Ardından üç model karşılaştırılmıştır. Veri seti ile yapılan çalışmaların bazıları incelenmiştir. Bu incelen sayfaları da kaynakça kısmına eklenmiştir.

İKİNCİ EL ARABA FİYAT TAHMİNİ VE İSTATİSTİK RAPORU

Veri Setimiz ve Grafikler

Araba alırken birçok faktör etkili olabiliyor. Buna bir örnekle anlatıyım. Babamla araba almaya gittiğimizde arabanın üst kısmındaki güneş yanığının internette yazılardan fazla olduğundan fiyatın biraz daha düşmesini istedi. Yani küçük bir güneş yanığı var diyerek. Fiyatı biraz daha yüksek tutmuş.

Veri setimizi indirdikten sonra **Python** da yazacağımız kodla ilgili kütüphaneleri **import** etmemiz gerekiyor. Bunların bazıları **Pandas**; Bir tanesi veri setini okutmak için. **Seaborn**; Bazı grafikleri çizdirmek için. **LinearRegression**; Doğrusal regresyon modeli için. **DecisionTreeRegressor**; Karar ağacı modeli için. **RandomForestRegressor**; Rastgele orman modeli için.

| | |
|--|--|
| import pandas as pd | from sklearn import linear_model |
| import patsy as ps | from sklearn.model_selection import train_test_split, cross_val_score |
| import numpy as np | from sklearn.linear_model import LinearRegression |
| import seaborn as sns | from matplotlib import pyplot as plt |
| | from sklearn.tree import DecisionTreeRegressor |
| import statsmodels.api as sm | from sklearn.ensemble import RandomForestRegressor |
| import statsmodels.formula.api as smf | from sklearn.datasets import make_regression |

Veri seti çekme işlemi yapıldı.

```
IEAGrafik=pd.read_csv('M:/MEHMET ARICI/NotebooksPython/BilgisayarOrganizasyonu/Final/IkinciElAraba.csv')
```

Veri Setimizi Tanıyalım

| Orijinal İsmi | Değiştirilmiş İsmi | Açıklama |
|----------------------|-------------------------|---|
| v.id | Değişken silindi | Python kendi ID atması yaptığı için silindi. |
| on road old | EskiToplam | Arabaların belli bir zamandaki toplamı |
| on road now | GuncelToplam | Arabaların en güncel toplamı |
| years | Yasi | Arabaların kaç yaşında olduğu |
| km | Km | Arabaların kilometresi |
| rating | MPuani | Arabaların motor durumu |
| condition | DDurumu | Arabaların değişen parçaları |
| economy | KParca | Arabaların konfor parçaları |
| top speed | EnYuksekHiz | Arabaların en yüksek hızları |
| hp | BeygirGucu | Arabaların belli zamandaki hızları |
| torque | Tork | Arabaların motorunun gücü |
| current price | Fiyat | Arabaların fiyatları |

Modelimizi iki ayrı veri seti gibi düşünüp. Biri grafik için kullanıldı. Diğeri ise model için kullanıldı. Grafik verisini kategorize edildiği için bu yöntemle başvurulmuştur. V.id değişkeni silindi. Head() ile ilk beş veriye bakıldı. Değişken ismi Türkçeleştirildi.

| IEAGrafik.head() | | | | | | | | | | | | |
|------------------|------|-------------|-------------|-------|--------|--------|-----------|---------|-----------|----|--------|---------------|
| | v.id | on road old | on road now | years | km | rating | condition | economy | top speed | hp | torque | current price |
| 0 | 1 | 535651 | 798188 | 3 | 78945 | 1 | 2 | 14 | 177 | 73 | 123 | 351318.0 |
| 1 | 2 | 591911 | 861056 | 6 | 117220 | 5 | 9 | 9 | 148 | 74 | 95 | 285001.5 |
| 2 | 3 | 686990 | 770762 | 2 | 132538 | 2 | 8 | 15 | 181 | 53 | 97 | 215386.0 |
| 3 | 4 | 573999 | 722381 | 4 | 101065 | 4 | 3 | 11 | 197 | 54 | 116 | 244295.5 |
| 4 | 5 | 691388 | 811335 | 6 | 61559 | 3 | 9 | 12 | 160 | 53 | 105 | 531114.5 |

Veri setini; IEAGrafik ve IEAModel olarak iki farklı şekilde kullanılabilir hale getirildi.

```
IEAGrafik.drop(["v.id"], axis=1, inplace=True)
IEAGrafik.columns = ['EskiToplam', 'GuncelToplam', 'Yasi', 'Km', 'MPuani', 'DDurumu', 'KParca', 'EnYuksekHiz', 'BeygirGucu', 'Tork', 'Fiyat']
IEAModel = IEAGrafik[['EskiToplam', 'GuncelToplam', 'Yasi', 'Km', 'MPuani', 'DDurumu', 'KParca', 'EnYuksekHiz', 'BeygirGucu', 'Tork', 'Fiyat']]
```

IEAGrafik isimli veri seti için sınıflama yapıldı. Yorumlamayı kolaylaştırıldı. Motor puanı; düşük hasarlı motorları (**Düşük** şeklinde) kodlandı. Motor puanı; normal hasarlı motorları (**Normal** şeklinde) kodlandı. Motor puanı; yüksek hasarlı motorları (**Yüksek** şeklinde) kodlandı. Bir diğer sınıflama ise arabaların parça değişimi için kodlama yapıldı. (**Çok İyi**) Çok az parça değişimi olan

MEHMET ARICI
2018165002

arabalar. (İyi) Az parça değişimi olan arabalar. (Normal) Orta seviyede parça değişimi olan arabalar. (Kötü) Fazla parça değişimi olan arabalar. (Çok Kötü) Çok fazla parça değişimi olan arabalar. Sınıflamadan sonra oraya ait (sayısal) değişken silindi.

```
IEAGrafik.loc[(IEAGrafik['MPuani']<3), 'MotorPuanı'] = 'Düşük'  
IEAGrafik.loc[(IEAGrafik['MPuani']==3), 'MotorPuanı'] = 'Normal'  
IEAGrafik.loc[(IEAGrafik['MPuani']>3), 'MotorPuanı'] = 'Yüksek'  
  
IEAGrafik.loc[(IEAGrafik['DDurumu']<=2), 'APDDurumu'] = 'Çok İyi'  
IEAGrafik.loc[(IEAGrafik['DDurumu']>2) & (IEAGrafik['DDurumu']<=4), 'APDDurumu'] = 'İyi'  
IEAGrafik.loc[(IEAGrafik['DDurumu']>4) & (IEAGrafik['DDurumu']<=6), 'APDDurumu'] = 'Normal'  
IEAGrafik.loc[(IEAGrafik['DDurumu']>6) & (IEAGrafik['DDurumu']<=8), 'APDDurumu'] = 'Kötü'  
IEAGrafik.loc[(IEAGrafik['DDurumu']>8), 'APDDurumu'] = 'Çok Kötü'  
  
IEAGrafik.drop(["MPuani","DDurumu"], axis=1, inplace=True)  
IEAGrafik.head(3)
```

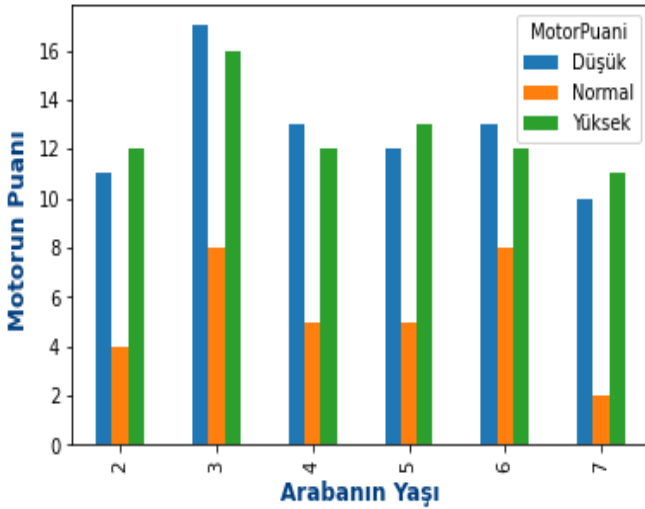
| | EskiToplam | GuncelToplam | Yasi | Km | KParca | EnYukseHiz | BeygirGucu | Tork | Fiyat | MotorPuanı | APDDurumu |
|---|------------|--------------|------|--------|--------|------------|------------|------|----------|------------|-----------|
| 0 | 535651 | 798186 | 3 | 78945 | 14 | 177 | 73 | 123 | 351318.0 | Düşük | Çok İyi |
| 1 | 591911 | 861056 | 6 | 117220 | 9 | 148 | 74 | 95 | 285001.5 | Yüksek | Çok Kötü |
| 2 | 686990 | 770762 | 2 | 132538 | 15 | 181 | 53 | 97 | 215386.0 | Düşük | Kötü |

ARABALAR İLE İLGİLİ BAZI GRAFİKLER

Grafik_1

```
mask=(IEAGrafik.APDDurumu=='Çok İyi')  
freq=pd.crosstab(IEAGrafik[mask].Yasi,IEAGrafik[mask].MotorPuanı)  
freq.plot(kind='bar')  
plt.xlabel("Arabanın Yaşı",color="#003f7f", fontsize = 12, weight='bold')  
plt.ylabel("Motorun Puanı ",color="#003f7f", fontsize = 12, weight='bold')  
plt.title("ARABANIN PARÇA DURUMUNU (ÇOK İYİ), YAŞINA GÖRE MOTOR PUANI", fontsize = 12, weight='bold')  
plt.show()
```

ARABANIN PARÇA DURUMUNU (ÇOK İYİ), YAŞINA GÖRE MOTOR PUANI



Çubuk Grafik

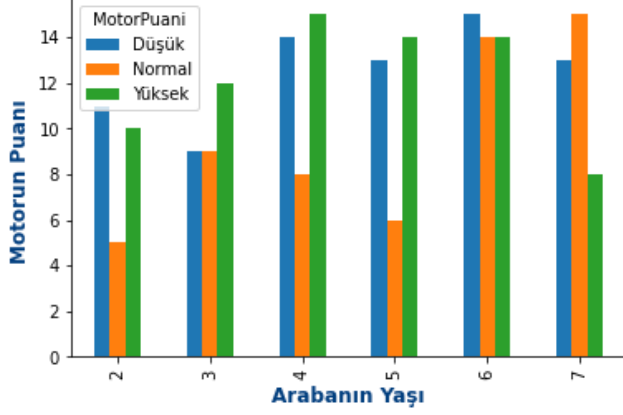
Araba parça değişim durumu çok az olan arabaların. İki yaşındaki arabadan; 11 adet motor hasarı düşük, 4 adet motor hasarı normal, 12 adet motor hasarı yüksek araba varmış.

Grafik_2

```
mask=(IEAGrafik.APDDurumu=='Çok Kötü')  
freq=pd.crosstab(IEAGrafik[mask].Yasi,IEAGrafik[mask].MotorPuanı)  
freq.plot(kind='bar')  
plt.xlabel("Arabanın Yaşı",color="#003f7f", fontsize = 12, weight='bold')  
plt.ylabel("Motorun Puanı ",color="#003f7f", fontsize = 12, weight='bold')  
plt.title("ARABANIN PARÇA DURUMUNU (ÇOK KÖTÜ), YAŞINA GÖRE MOTOR PUANI", fontsize = 12, weight='bold')  
plt.show()
```

MEHMET ARICI
2018165002

ARABANIN PARÇA DURUMUNU (ÇOK KÖTÜ), YAŞINA GÖRE MOTOR PUANI



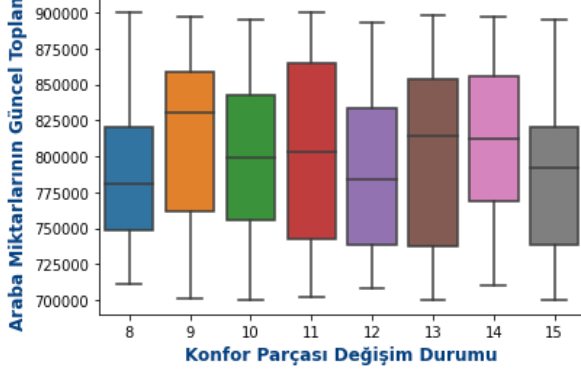
Çubuk Grafik

Araba parça değişim durumu çok fazla olan arabaların. Beş yaşındaki arabadan; 11 adet motor hasarı düşük, 6 adet motor hasarı normal, 15 adet motor hasarı yüksek araba varmış.

Grafik_3

```
mask=( IEAGrafik.MotorPuani=="Yüksek")
sns.boxplot(x="KParça", y="GuncelToplam", data=IEAGrafik[mask])
plt.xlabel("Konfor Parçası Değişim Durumu",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("Araba Miktarlarının Güncel Toplamı",color="#003f7f", fontsize = 12, weight='bold')
plt.title("MOTOR PUANI (YÜKSEK) KONFOR PARÇASI DEĞİŞİMİYLE GÜNCEL TOPLAMI", fontsize = 12, weight='bold')
plt.show()
```

MOTOR PUANI (YÜKSEK) KONFOR PARÇASI DEĞİŞİMİYLE GÜNCEL TOPLAMI



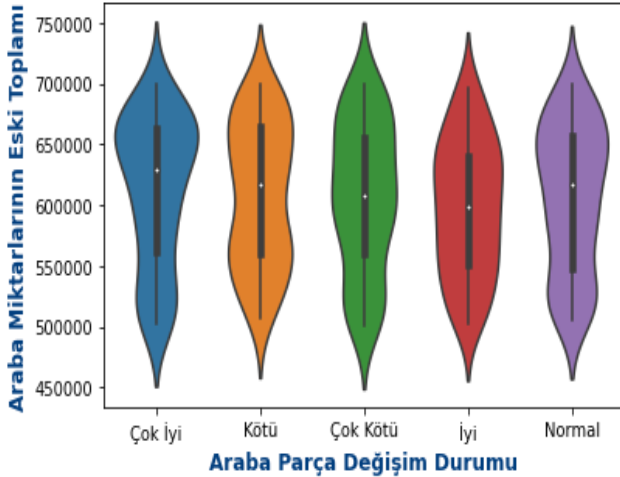
Boxplot Grafik

Arabanın motor hasarının çok fazla (Yüksek) olan arabaların. 9 numaralı konfor parçasının ortalaması 830000 civarındadır. Aykırı değer yoktur. Diğer parçalar için de aykırı değer yoktur.

Grafik_4

```
mask=( IEAGrafik.MotorPuani=="Düşük")
sns.violinplot(x="APDDurumu", y="EskiToplam", data=IEAGrafik[mask])
plt.xlabel("Araba Parça Değişim Durumu",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("Araba Miktarlarının Eski Toplamı",color="#003f7f", fontsize = 12, weight='bold')
plt.title("MOTOR PUANI (DÜŞÜK) ARABA PARÇASI DEĞİŞİMİYLE ESKİ TOPLAMI", fontsize = 12, weight='bold')
plt.show()
```

MOTOR PUANI (DÜŞÜK) ARABA PARÇASI DEĞİŞİMİYLE ESKİ TOPLAMI



Violinplot Grafik

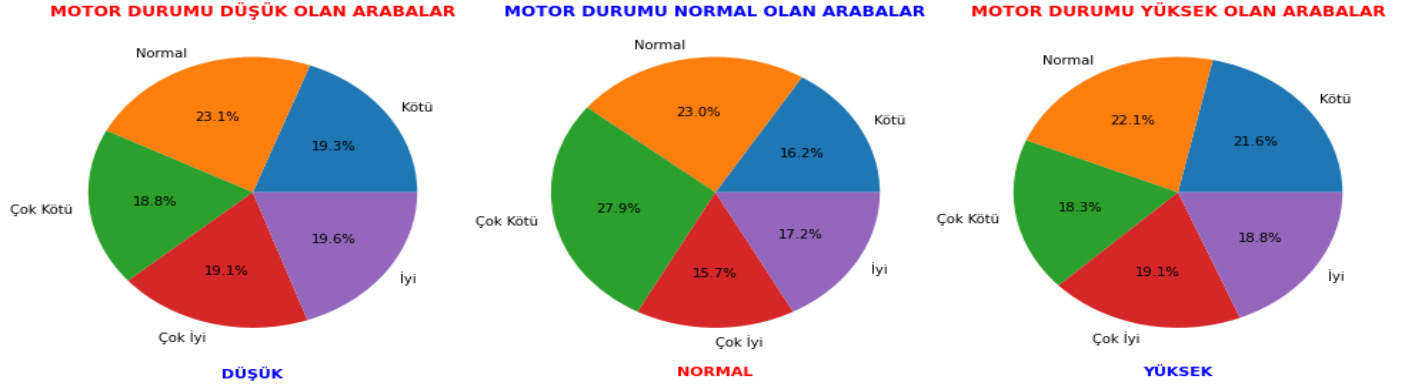
Arabanın motor hasarının çok az (Düşük) olan arabaların. Araba parça değişim durumu çok az parça değişimi (Çok İyi) olan arabaların. Eski toplam ortalaması 640000 civarındadır. 650000 ile 700000 arasında eski araba toplamı fazlamış.

Grafik_5

```
freq=pd.crosstab(IEAGrafik.APDDurumu,IEAGrafik.MotorPuanı)
plt.axes([0.00,0.05,1,1])
freq['Düşük'].plot(kind='pie', autopct='%1.1f%%')
plt.ylabel("")
plt.xlabel('DÜŞÜK',color="b", fontsize = 11, weight='bold')
plt.title('MOTOR DURUMU DÜŞÜK OLAN ARABALAR',color="r", fontsize = 12, weight='bold')

plt.axes([0.75,0.05,1,1])
freq['Normal'].plot(kind='pie', autopct='%1.1f%%')
plt.ylabel("")
plt.xlabel('NORMAL',color="r", fontsize = 11, weight='bold')
plt.title('MOTOR DURUMU NORMAL OLAN ARABALAR',color="b", fontsize = 12, weight='bold')

plt.axes([1.5,0.05,1,1])
freq['Yüksek'].plot(kind='pie', autopct='%1.1f%%')
plt.ylabel("")
plt.xlabel('YÜKSEK',color="b", fontsize = 11, weight='bold')
plt.title('MOTOR DURUMU YÜKSEK OLAN ARABALAR',color="r", fontsize = 12, weight='bold')
plt.show()
```

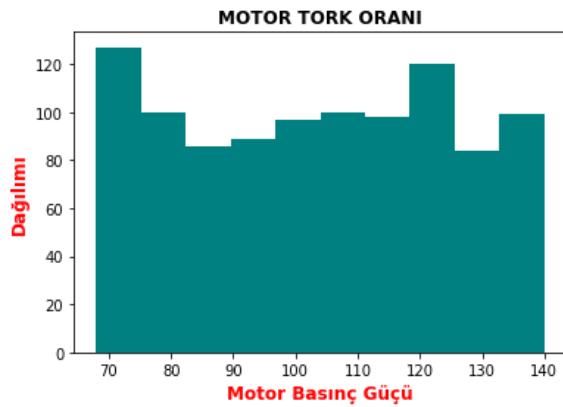


Pasta (Pie) Grafik

Araba parça değişim durumu normal, arabanın motor hasarının düşük olduğu %23 araba vardır. Araba parça değişim durumu çok iyi arabanın motor hasarının normal olduğu %16 araba vardır. Araba parça değişim durumu kötü arabanın motor hasarının yüksek olduğu %22 araba vardır.

Grafik_6

```
plt.hist(IEAGrafik.Tork,color='#008080')
plt.xlabel("Motor Basınç Gücü",color="r", fontsize = 12, weight='bold')
plt.ylabel("Dağılımı",color="r", fontsize = 12, weight='bold')
plt.title("MOTOR TORK ORANI", fontsize = 12, weight='bold')
plt.show()
```



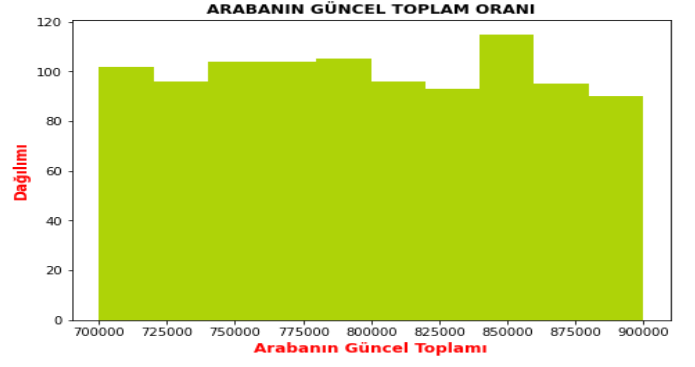
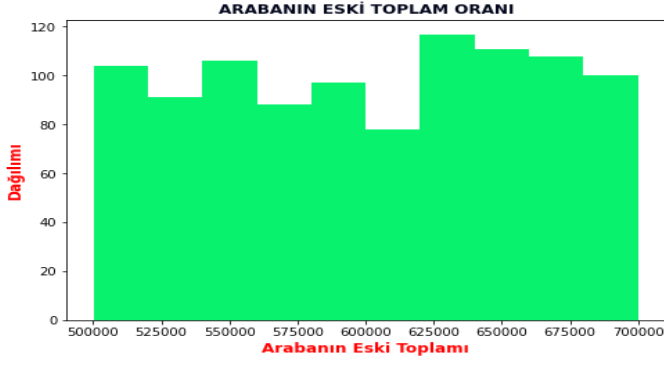
Histogram Grafik

Arabaların motorunun basınç gücü (Tork) göre dağılımının histogram grafiğidir. Birkaç tanesinden bahsedelim. 70 basınç gücüne sahip 130 adet araba vardır. 90 basınç gücüne sahip yaklaşık 85 adet araba vardır. 110 basınç gücüne sahip yaklaşık 95 adet araba vardır.

```
plt.axes([0.01,1.21,1,1])
plt.hist(IEAGrafik.EskiToplam,color='#09f26e')
plt.xlabel("Arabanın Eski Toplamı",color="r", fontsize = 12, weight='bold')
plt.ylabel("Dağılımı",color="r", fontsize = 12, weight='bold')
plt.title("ARABANIN ESKİ TOPLAM ORANI",color="#01071c", fontsize = 12, weight='bold')

plt.axes([1.21,1.21,1,1])
plt.hist(IEAGrafik.GüncelToplam,color='#aed308')
plt.xlabel("Arabanın Güncel Toplamı",color="r", fontsize = 12, weight='bold')
plt.ylabel("Dağılımı",color="r", fontsize = 12, weight='bold')
plt.title("ARABANIN GÜNCEL TOPLAM ORANI", fontsize = 12, weight='bold')
plt.show()
```

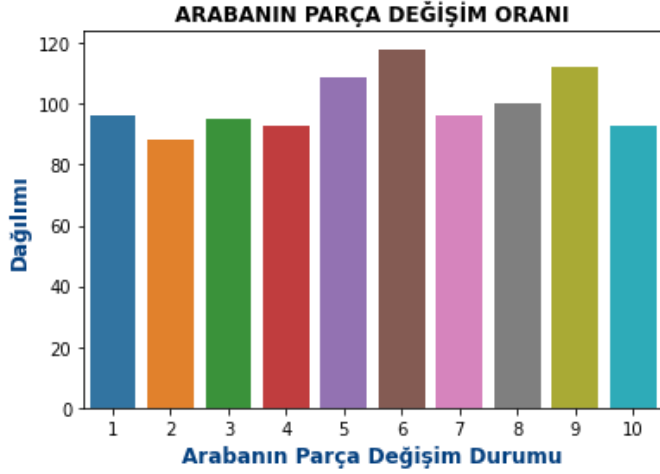
MEHMET ARICI
2018165002



Arabaların eski toplamı ve güncel toplamına göre dağılımının histogram grafiğidir. Eski toplam 550000 olan 110 adet vardır.

Grafik_7

```
sns.countplot(x = 'DDurumu',data = IEAModel)
plt.xlabel("Arabanın Parça Değişim Durumu",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("Dağılımı",color="#003f7f", fontsize = 12, weight='bold')
plt.title("ARABANIN PARÇA DEĞİŞİM ORANI", fontsize = 12, weight='bold')
plt.show()
```

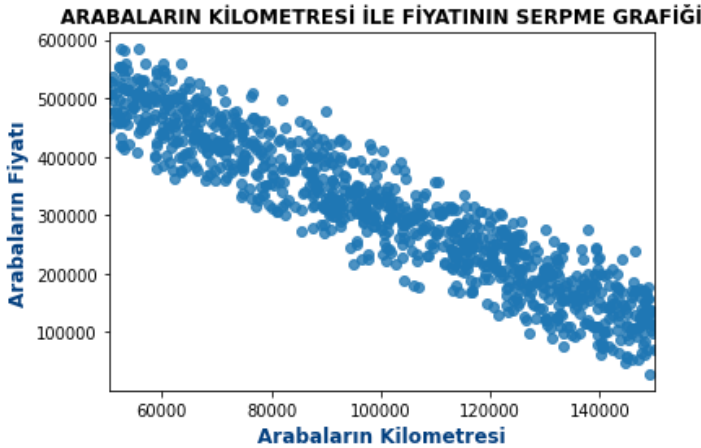


Çubuk Grafik

Araba parçasının değişen; 3 numaralı parçadan yaklaşık 95 adet vardır. 6 numaralı parçadan 115 adet vardır. 10 numaralı parçadan yaklaşık 90 adet vardır.

Grafik_8

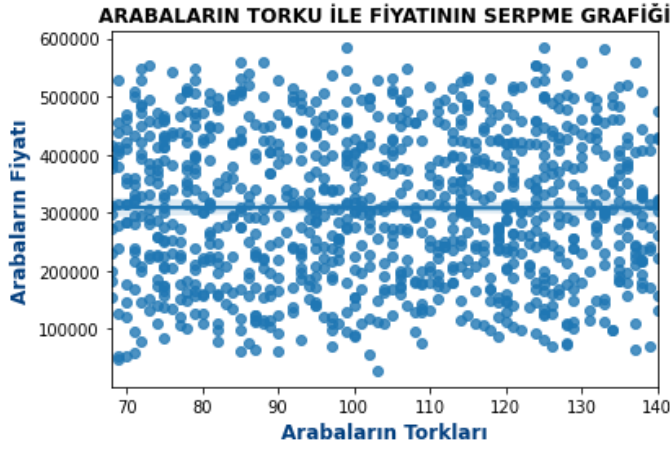
```
sns.regplot(x='Km',y='Fiyat',data=IEAGrafik)
plt.xlabel("Arabaların Kilometresi",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("Arabaların Fiyatı",color="#003f7f", fontsize = 12, weight='bold')
plt.title("ARABALARIN KİLOMETRESİ İLE FİYATININ SERPME GRAFİĞİ", fontsize = 12, weight='bold')
plt.show()
```



Serpme Grafik

Arabaların kilometresi ile Fiyatının serpme grafiğinde gördüğümüz. Negatif yönde doğrusal ilişki vardır.


```
sns.regplot(x='Tork',y='Fiyat',data=IEAGrafik)
plt.xlabel("Arabaların Torkları",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("Arabaların Fiyatı",color="#003f7f", fontsize = 12, weight='bold')
plt.title("ARABALARIN TORKU İLE FİYATININ SERPME GRAFİĞİ", fontsize = 12, weight='bold')
plt.show()
```



Serpme Grafik
Arabaların basınç gücü (Tork) ile Fiyatının
serpme grafiğinde gördüğümüz. Nötr ilişki vardır.

ARABA FİYATI TAHMİNİ MODELLERİ

Modellere geçmeden önce bazı işlemlere bakalım. Describe ile veri setimizin. Bazı istatistiksel değerlere ulaşıyoruz. Bunlar; Ortalaması, Standart sapması, Min. – Max. Değerleri ve çeyrekler öğrenebiliyoruz.

```
IEAModel.describe()
```

| | EskiToplam | GuncelToplam | Yasi | Km | MPuani | DDurumu | KParca | EnYuksekHiz | BeygirGucu | Tork | Fiyat |
|-------|---------------|--------------|-------------|---------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| count | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 | 1000.000000 |
| mean | 601648.286000 | 799131.3970 | 4.561000 | 100274.430000 | 2.988000 | 5.592000 | 11.625000 | 166.89300 | 84.54600 | 103.423000 | 308520.2425 |
| std | 58407.246204 | 57028.9502 | 1.719079 | 29150.463233 | 1.402791 | 2.824449 | 2.230549 | 19.28838 | 20.51694 | 21.058716 | 126073.2591 |
| min | 500265.000000 | 700018.0000 | 2.000000 | 50324.000000 | 1.000000 | 1.000000 | 8.000000 | 135.00000 | 50.00000 | 68.000000 | 28226.5000 |
| 25% | 548860.500000 | 750997.7500 | 3.000000 | 74367.500000 | 2.000000 | 3.000000 | 10.000000 | 150.00000 | 67.00000 | 85.000000 | 206871.7500 |
| 50% | 601568.000000 | 798168.0000 | 5.000000 | 100139.500000 | 3.000000 | 6.000000 | 12.000000 | 166.00000 | 84.00000 | 104.000000 | 306717.7500 |
| 75% | 652267.250000 | 847563.2500 | 6.000000 | 125048.000000 | 4.000000 | 8.000000 | 13.000000 | 184.00000 | 102.00000 | 121.000000 | 414260.8750 |
| max | 699859.000000 | 899797.0000 | 7.000000 | 149902.000000 | 5.000000 | 10.000000 | 15.000000 | 200.00000 | 120.00000 | 140.000000 | 584267.5000 |

Info ile boş (null) değerimiz var mı? Diye kontrol ettik.

```
IEAModel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column              Non-Null Count  Dtype  
--  --   --
 0   EskiToplam          1000 non-null   int64  
 1   GuncelToplam        1000 non-null   int64  
 2   Yasi                1000 non-null   int64  
 3   Km                  1000 non-null   int64  
 4   MPuani              1000 non-null   int64  
 5   DDurumu             1000 non-null   int64  
 6   KParca              1000 non-null   int64  
 7   EnYuksekHiz         1000 non-null   int64  
 8   BeygirGucu          1000 non-null   int64  
 9   Tork                1000 non-null   int64  
10   Fiyat               1000 non-null   float64
dtypes: float64(1), int64(10)
memory usage: 86.1 KB
```

Korelasyonlarını tablo halinde çıktısını aldık. Sıcaklık grafiğiyle daha iyi görebiliriz.

```
IEAModel.corr()
```

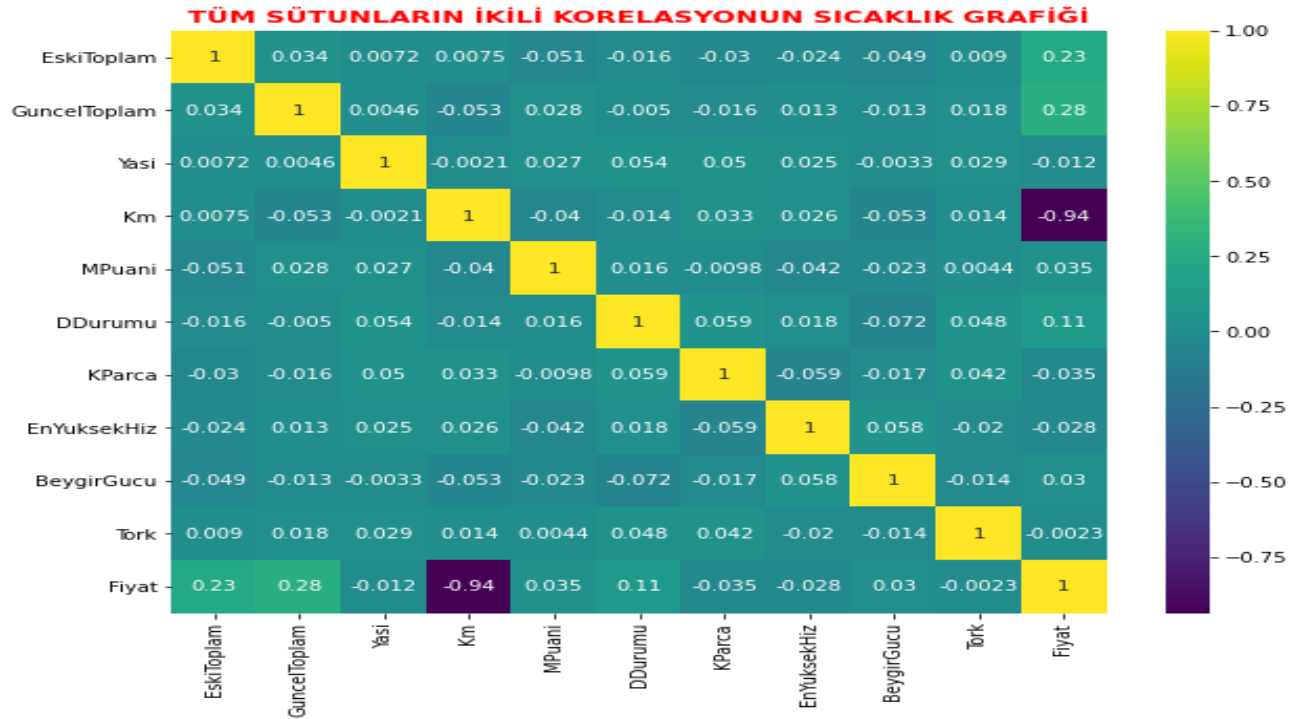
| | EskiToplam | GuncelToplam | Yasi | Km | MPuani | DDurumu | KParca | EnYuksekHiz | BeygirGucu | Tork | Fiyat |
|--------------|------------|--------------|-----------|-----------|-----------|-----------|-----------|-------------|------------|-----------|-----------|
| EskiToplam | 1.000000 | 0.034113 | 0.007207 | 0.007488 | -0.050717 | -0.015682 | -0.030097 | -0.023816 | -0.049266 | 0.008950 | 0.233035 |
| GuncelToplam | 0.034113 | 1.000000 | 0.004609 | -0.053202 | 0.028280 | -0.005043 | -0.015880 | 0.012699 | -0.012719 | 0.017955 | 0.282793 |
| Yasi | 0.007207 | 0.004609 | 1.000000 | -0.002089 | 0.027285 | 0.053579 | 0.050220 | 0.025148 | -0.003272 | 0.028859 | -0.011854 |
| Km | 0.007488 | -0.053202 | -0.002089 | 1.000000 | -0.039930 | -0.013640 | 0.032680 | 0.026450 | -0.052918 | 0.013566 | -0.935924 |
| MPuani | -0.050717 | 0.028280 | 0.027285 | -0.039930 | 1.000000 | 0.015943 | -0.009757 | -0.042222 | -0.022623 | 0.004408 | 0.035038 |
| DDurumu | -0.015682 | -0.005043 | 0.053579 | -0.013640 | 0.015943 | 1.000000 | 0.058788 | 0.018472 | -0.071552 | 0.047805 | 0.110108 |
| KParca | -0.030097 | -0.015880 | 0.050220 | 0.032680 | -0.009757 | 0.058788 | 1.000000 | -0.059402 | -0.016782 | 0.041632 | -0.034711 |
| EnYuksekHiz | -0.023816 | 0.012699 | 0.025148 | 0.026450 | -0.042222 | 0.018472 | -0.059402 | 1.000000 | 0.057827 | -0.019697 | -0.027993 |
| BeygirGucu | -0.049266 | -0.012719 | -0.003272 | -0.052918 | -0.022623 | -0.071552 | -0.016782 | 0.057827 | 1.000000 | -0.013817 | 0.030238 |
| Tork | 0.008950 | 0.017955 | 0.028859 | 0.013566 | 0.004408 | 0.047805 | 0.041632 | -0.019697 | -0.013817 | 1.000000 | -0.002290 |
| Fiyat | 0.233035 | 0.282793 | -0.011854 | -0.935924 | 0.035038 | 0.110108 | -0.034711 | -0.027993 | 0.030238 | -0.002290 | 1.000000 |

MEHMET ARICI

2018165002

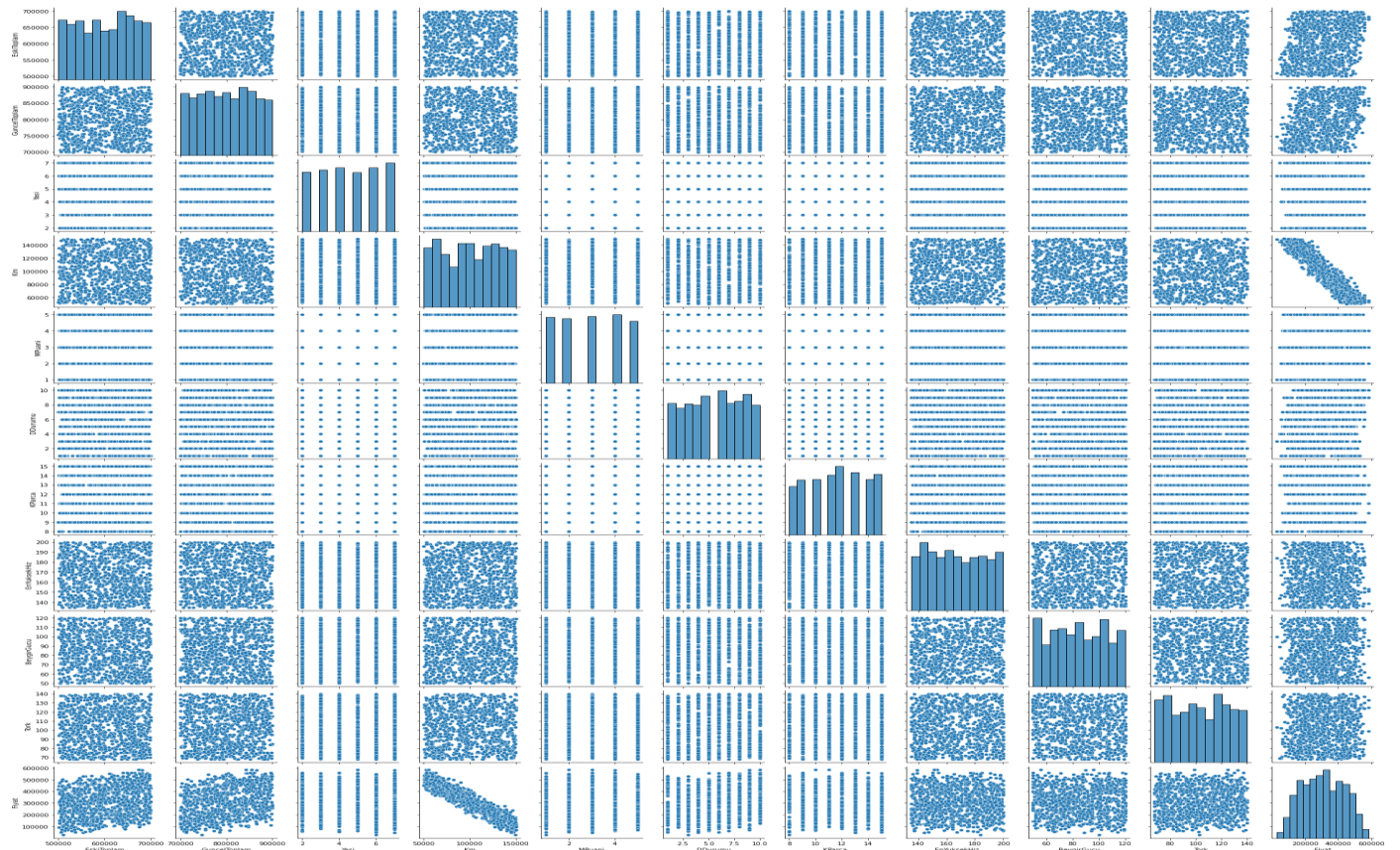
Km ile Fiyat arasında negatif korelasyon %94 'dür. Güncel Toplam ile Fiyat arasında pozitif korelasyon %28 'dir.

```
fig = plt.figure(figsize=(10,8))
sns.heatmap(IEAModel.corr(), annot=True, cmap = 'viridis')
plt.title("TÜM SÜTUNLARIN İKİLİ KORELASYONUN SICAKLIK GRAFİĞİ",color='r', fontsize = 13, weight='bold')
plt.show()
```



Pairplot ile ikili korelasyonlarına baktık. Çoğunluğun nötr ve bir tanesinin negatif yönde korelasyona sahip olduğu gördük.

```
sns.pairplot(IEAModel)
plt.show()
```



MEHMET ARICI
2018165002

```
y, X = ps.dmatrices('Fiyat ~ EskiToplam + GuncelToplam + Yasi + Km + MPuani + DDurumu + KParca + EnYuksekHiz + BeygirGucu + Tork')
LRModeli = sm.OLS(y, X)
fit = LRModeli.fit()
fit.summary()
```

Dmatrices ile veri setimizin R^2 , düzeltilmiş R^2 , condition number ve t (tablo) değerlerimiz gibi Regresyon yapabilmemiz için bazı istatistiksel değerler görmemizi sağladı.

| R^2 | DÜZELTİLMİŞ R^2 | | CONDITION NUMBER | | | | |
|-------|-------------------|------------|------------------|----------|---------|-----------|-----------|
| 0.995 | 0.995 | | 2412.593 | | | | |
| | COEF | STD. ERR. | T | P> T | [0.025 | 0.975] | |
| | INTERCEPT | -1.383e+04 | 6046.330 | -2.287 | 0.022 | -2.57e+04 | -1961.431 |
| | ESKİTOPLAM | 0.5057 | 0.005 | 106.452 | 0.000 | 0.496 | 0.515 |
| | GUNCELTOPLAM | 0.5003 | 0.005 | 102.949 | 0.000 | 0.491 | 0.510 |
| | YASİ | -1622.3890 | 161.327 | -10.057 | 0.000 | -1938.972 | -1305.806 |
| | KM | -3.9961 | 0.010 | -419.521 | 0.000 | -4.015 | -3.977 |
| | MPUANİ | 230.9884 | 197.816 | 1.168 | 0.243 | -157.199 | 619.175 |
| | DDURUMU | 4628.4689 | 98.541 | 46.970 | 0.000 | 4435.096 | 4821.842 |
| | KPARCA | 62.4643 | 124.716 | 0.501 | 0.617 | -182.275 | 307.203 |
| | ENYUKSEKHİZ | -14.5732 | 14.413 | -1.011 | 0.312 | -42.856 | 13.710 |
| | BEYGİRGUCU | 20.3599 | 13.569 | 1.500 | 0.134 | -6.268 | 46.988 |
| | TORK | -1.7328 | 13.157 | -0.132 | 0.895 | -27.551 | 24.086 |

MODELİMİZ

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_9 + \beta_{10} X_{10}$$

HİPOTEZLER

| SÜTUN İSMİ | HİPOTEZLER | KARAR |
|------------|---|--|
| INTERCEPT | $H_0 : \beta_0 = 0$ $H_1 : \beta_0 \neq 0$ | 0.02 < 0.05 olduğu için; İstatistiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| ESKİTOPLAM | $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$ | 0.00 < 0.05 olduğu için; İstatistiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |

| | | |
|--------------|---|---|
| GUNCELTOPLAM | $H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$ | $0.00 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| YASİ | $H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$ | $0.00 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| KM | $H_0 : \beta_4 = 0$ $H_1 : \beta_4 \neq 0$ | $0.00 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| MPUANI | $H_0 : \beta_5 = 0$ $H_1 : \beta_5 \neq 0$ | $0.24 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| DDURUMU | $H_0 : \beta_6 = 0$ $H_1 : \beta_6 \neq 0$ | $0.00 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| KPARCA | $H_0 : \beta_7 = 0$ $H_1 : \beta_7 \neq 0$ | $0.62 > 0.05$ olduğu için; İstatiksel açıdan anlamlı fark yoktur. (H_0 Kabul, H_1 Red) |
| ENYUKSEKHİZ | $H_0 : \beta_8 = 0$ $H_1 : \beta_8 \neq 0$ | $0.31 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| BEYGİRGUCU | $H_0 : \beta_9 = 0$ $H_1 : \beta_9 \neq 0$ | $0.13 < 0.05$ olduğu için; İstatiksel açıdan anlamlı fark vardır. (H_0 Red, H_1 Kabul) |
| TORK | $H_0 : \beta_{10} = 0$ $H_1 : \beta_{10} \neq 0$ | $0.89 > 0.05$ olduğu için; İstatiksel açıdan anlamlı fark yoktur. (H_0 Kabul, H_1 Red) |

LİNEAR REGRESSION MODELİ

```
X = IEAModel.loc[:,['EskiToplam','GuncelToplam','Yasi','Km','MPuani','DDurumu','KParca','EnYuksekhiz','BeygirGucu','Tork']]
y=IEAModel['Fiyat']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state=42)
```

```
LR = linear_model.LinearRegression()
LRModeli = LR.fit(X_train, y_train)
LRpredictions = LR.predict(X_test)
LRpredictions[0:5]
```

```
array([373162.06652492, 513424.22156003, 155733.50826739, 164108.7152749 ,
       510432.21078703])
```

```
scores_LR = cross_val_score(LR, X, y, cv=10)
scores_LR
```

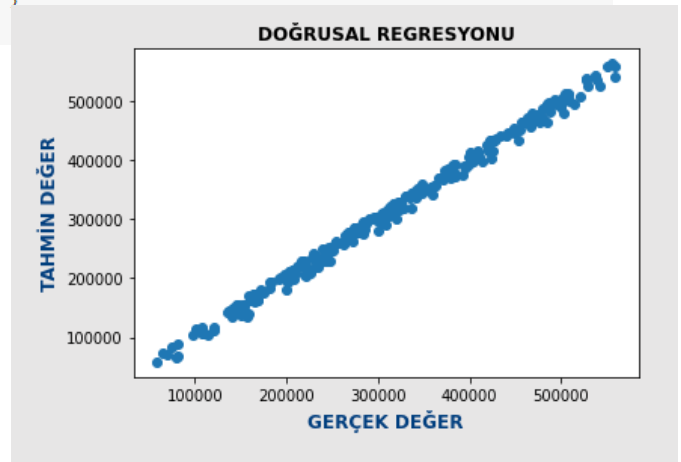
```
array([0.99519019, 0.99451568, 0.99576183, 0.99605278, 0.99510114,
       0.99358522, 0.99545023, 0.9951911 , 0.99536837, 0.99399761])
```

```
scores_LR.mean()
0.9950214147483963
```

```
plt.scatter(y_test, LRpredictions)
plt.xlabel("GERÇEK DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("TAHMİN DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.title("DOĞRUSAL REGRESYONU", fontsize = 12, weight='bold')
plt.show()
```

Doğrusal Regresyon Modeli

Condition number değerimiz 1000 üzerinde çıktığı için modele güvenilemez. Grafikte de gördüğümüz gibi %99 doğrulukla olduğunu söylüyor. Korelasyon çığımıza oldukça uzak değerler tahmin etmiştir.



DECISION TREE REGRESSION MODELİ

```
DTR = DecisionTreeRegressor(random_state=0)
DTR.fit(X,y)
DTRpredictions = DTR.predict(X_test)
```

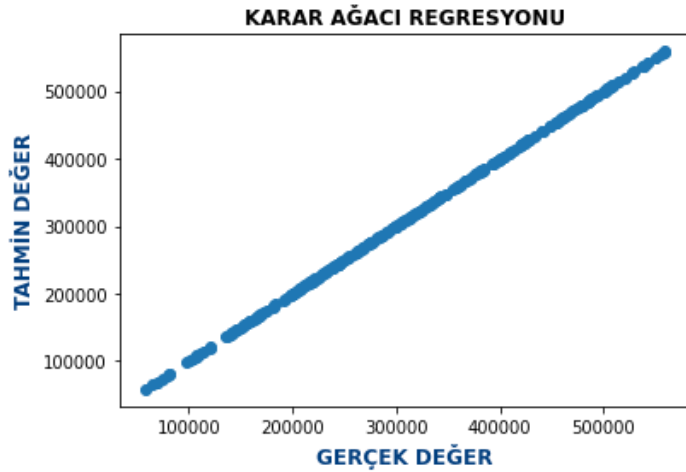
```
scores_DTR= cross_val_score(DTR, X, y, cv=10)
scores_DTR
```

```
array([0.94066896, 0.94803097, 0.95397888, 0.9651149 , 0.9484824 ,
       0.92364486, 0.9620908 , 0.96325461, 0.9465995 , 0.93508497])
```

```
scores_DTR.mean()
```

```
0.9486950854778465
```

```
plt.scatter(y_test, DTRpredictions)
plt.xlabel("GERÇEK DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("TAHMİN DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.title("KARAR AĞACI REGRESYONU", fontsize = 12, weight='bold')
plt.show()
```



Karar Ağacı Regresyon Modeli

Modelimizin puanı %94 doğrulukla ve korelasyon çizgimize daha yakın değerler üretmiştir. En başarılı modelimizdir. Üç modelin karşılaştırmasında bahsedilmiştir.

RANDOM FOREST REGRESSION MODELİ

```
X, y = make_regression(n_features=3, n_informative=2, random_state=0, shuffle=False)
RFR = RandomForestRegressor(max_depth=2, random_state=0)
RFR.fit(X, y)
RFRpredictions = RFR.predict(X)
```

```
scores_RFR = cross_val_score(RFR, X, y, cv=10)
scores_RFR
```

```
array([0.59147537, 0.72737673, 0.8182976 , 0.62461008, 0.59492016,
       0.77460947, 0.69831054, 0.71337219, 0.81882481, -0.03063339])
```

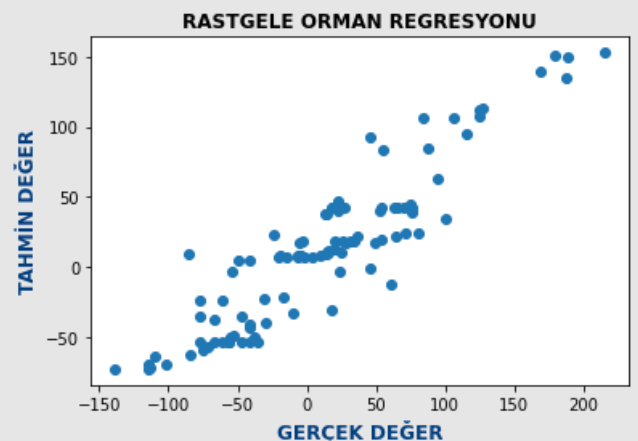
```
scores_RFR.mean()
```

```
0.633116357493986
```

```
plt.scatter(y, RFRpredictions)
plt.xlabel("GERÇEK DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("TAHMİN DEĞER", color="#003f7f", fontsize = 12, weight='bold')
plt.title("RASTGELE ORMAN REGRESYONU", fontsize = 12, weight='bold')
plt.show()
```

Rastgele Orman Regresyon Modeli

Modelimizin puanı %63 doğrulukla ve korelasyon çizgimize daha uzak değerler üretmiş.



MODELLERİN KARŞILAŞTIRILMASI

```
: print('Linear Regression :', scores_LR.mean())
print('Decision Tree Regression :', scores_DTR.mean())
print('Random Forest Regression :', scores_RFR.mean())

Linear Regression : 0.9950214147483963
Decision Tree Regression : 0.9486950854778465
Random Forest Regression : 0.633116357493986

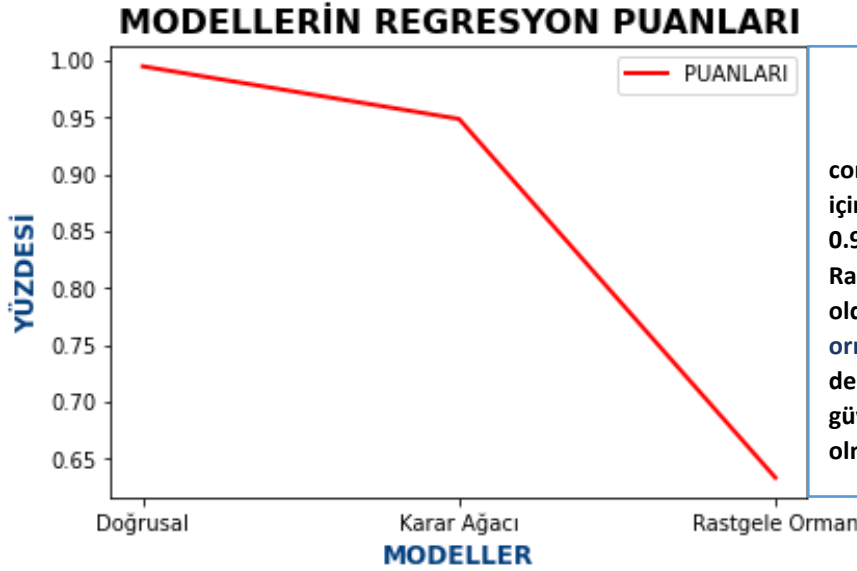
: data = [['Doğrusal', scores_LR.mean()], ['Karar Ağacı', scores_DTR.mean()], ['Rastgele Orman', scores_RFR.mean()]]
modelDataFrame = pd.DataFrame(data, columns = ['Model', 'Scores'])
modelDataFrame

:

```

| | Model | Scores |
|---|----------------|----------|
| 0 | Doğrusal | 0.995021 |
| 1 | Karar Ağacı | 0.948695 |
| 2 | Rastgele Orman | 0.633116 |

```
: plt.plot( 'Model', 'Scores', data=modelDataFrame, marker='r', color='r', linewidth=2, label="PUANLARI")
plt.legend()
plt.xlabel("MODELLER",color="#003f7f", fontsize = 12, weight='bold')
plt.ylabel("YÜZDESİ",color="#003f7f", fontsize = 12, weight='bold')
plt.title('MODELLERİN REGRESYON PUANLARI',fontsize = 16, weight='bold')
plt.show()
```



Karşılaştırma

Doğrusal modelimizin puanı 0.995; condition number değerimiz 1000 üzerinde olduğu için güvenemeyiz. Karar ağacı modelimizin puanı 0.948; Doğrusal modelimiz güvensiz olduğu için ve Rastgele orman modelimizin puanı da düşük olduğundan en başarılı modelimizdir. Rastgele orman modelimizin puanı 0.633; Diğer iki modelden de oldukça düşüktür. Doğrusal modelimize güvenemeyeceğimiz için ikinci başarılı modelimiz olmuş oluyor.

KAYNAKÇA

- <https://www.kaggle.com/mayankpatel14/second-hand-used-cars-data-set-linear-regression/notebooks>
- <https://www.kaggle.com/phaneshwardundigalla/multi-linear-regression-2>
- <https://www.kaggle.com/navdeepsinghhada/used-car-price-prediction>