**VERİ MADENCİLİĞİ VE BÜYÜK VERİ ANALİZİ DÖNEM ÖDEVİ RAPORU**

**Giriş**

Bu raporda bazı özelliklerin öğrencilerin başarı durumlarına etkilerini gösteren bir veri seti kullanılmıştır. Bu özellikler cinsiyet, ebeveyn eğitim durumu, okulda öğlen yemeği yeme durumu, öğrencinin aldığı özel ders ve öğrencinin matematik, okuma ve yazma becerilerinden oluşmaktadır.

**Kütüphanelerin Eklenmesi**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

Ödevimizde numpy, pandas, matplotlib ve seaborn kütüphanelerini kullandık. Matplotlib ve seaborn kütüphanelerini görselleştirmeler için kullandık. Numpy ve pandas’ı da veri çekme ve veri analizi için kullandık

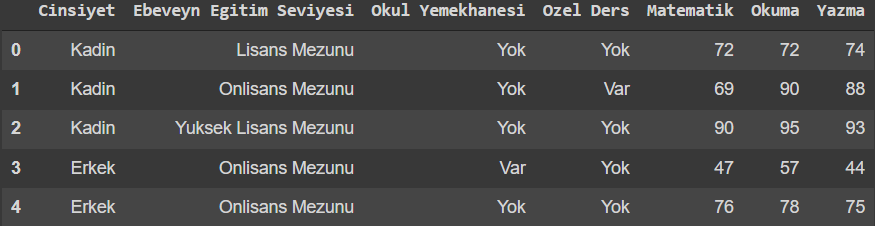
import sklearn

Sklearn kütüphanesinde veriye uygun model çıkarma ve modeli kullanma işleminde kullandık.

**Verinin İncelenmesi**

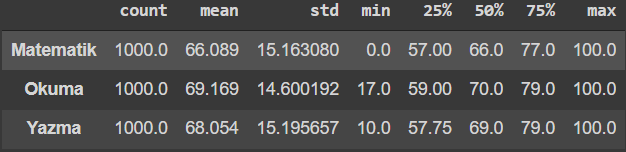
df\_raw = pd.read\_csv("Ogrenci\_Performans.csv", header="infer")

df\_raw.head()

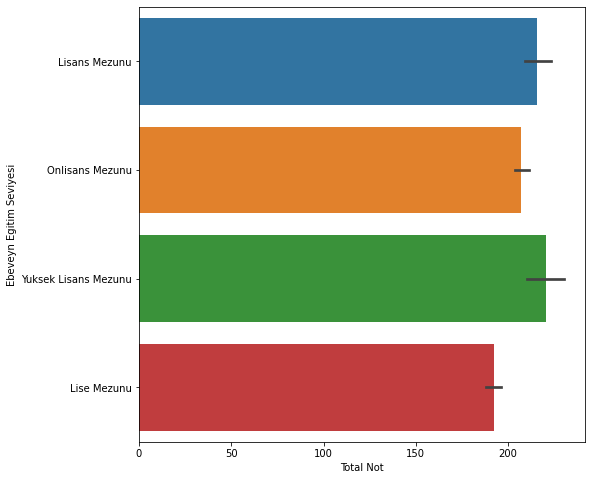


df = df\_raw.copy()

df.describe().T



sns.barplot(x=df['Total Not'], y=df['Ebeveyn Egitim Seviyesi'])



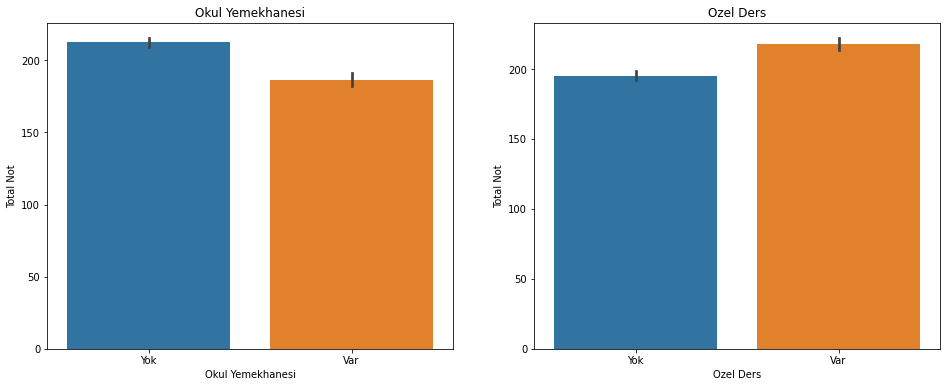
fig, (ax1,ax2) = plt.subplots(1,2, figsize=(16,6))

ax1.set\_title('Okul Yemekhanesi')

sns.barplot(x = 'Okul Yemekhanesi', y = 'Total Not', data= df, ax=ax1)

ax2.set\_title('Ozel Ders')

sns.barplot(x='Ozel Ders', y='Total Not', data=df, ax=ax2)



df.groupby('Cinsiyet')['Total Not'].mean()

Cinsiyet

Erkek 197.512448

Kadin 208.708494

df.groupby('Ebeveyn Egitim Seviyesi')['Total Not'].mean().sort\_values(ascending = True)

Ebeveyn Egitim Seviyesi

Lise Mezunu 192.170667

Onlisans Mezunu 207.053571

Lisans Mezunu 215.771186

Yuksek Lisans Mezunu 220.796610

df.groupby('Okul Yemekhanesi')['Total Not'].mean().sort\_values(ascending = True)

Okul Yemekhanesi

Var 186.597183

Yok 212.511628

df.groupby('Ozel Ders')['Total Not'].mean().sort\_values(ascending = True)

Ozel Ders

Yok 195.116822

Var 218.008380

df\_corr = df.corr()

df\_corr



df\_corr['Total Not'].sort\_values(ascending=False)

Total Not 1.000000

Okuma 0.970331

Yazma 0.965667

Matematik 0.918746

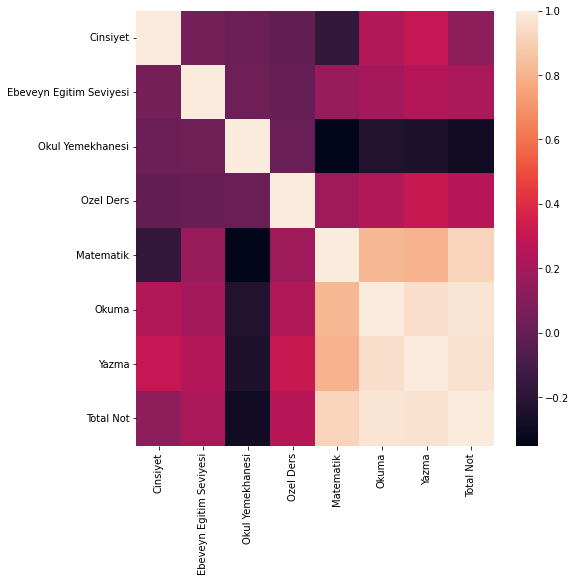
Ozel Ders 0.256710

Ebeveyn Egitim Seviyesi 0.214121

Cinsiyet 0.130861

Okul Yemekhanesi -0.290064

sns.heatmap(df\_corr)



**Modelleme**

feature\_vectors = df[['Cinsiyet', 'Ebeveyn Egitim Seviyesi','Okul Yemekhanesi','Ozel Ders', 'Okuma','Matematik','Yazma']].values

target = df['Notlar'].values

feature\_vectors.shape, target.shape

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test = train\_test\_split(feature\_vectors, target,test\_size = 0.25, random\_state = 45)

from sklearn.preprocessing import MinMaxScaler

scale = MinMaxScaler()

x\_train = scale.fit\_transform(x\_train)

x\_test = scale.transform(x\_test)

from sklearn.linear\_model import  LogisticRegression

model = LogisticRegression( max\_iter=7000)

model.fit(x\_train, y\_train)

y\_pred = model.predict(x\_test)

print("Training Accuracy :", model.score(x\_train, y\_train))

print("Testing Accuracy :", model.score(x\_test, y\_test))

Training Accuracy : 0.7546666666666667

Testing Accuracy : 0.708

import sklearn

model\_svm = sklearn.svm.LinearSVC(penalty='l1', dual=False,loss='squared\_hinge',  max\_iter=7000)

model\_svm.fit(x\_train, y\_train)

print("Training Accuracy :", model\_svm.score(x\_train, y\_train))

print("Testing Accuracy :", model\_svm.score(x\_test, y\_test))

Training Accuracy : 0.7546666666666667

Testing Accuracy : 0.68

from sklearn.ensemble import RandomForestClassifier

model\_rm = RandomForestClassifier()

model\_rm.fit(x\_train, y\_train)

y\_pred = model\_rm.predict(x\_test)

print("Training Accuracy :", model\_rm.score(x\_train, y\_train))

print("Testing Accuracy :", model\_rm.score(x\_test, y\_test))

Training Accuracy : 1.0

Testing Accuracy : 0.908

**Sonuç**

1. Özel ders almanın genel başarıya olan etkisi nedir?

df\_corr['Total Not'].sort\_values(ascending=False)

Ozel Ders 0.256710

Lojistik Regresyon modelinden çıkan sonuç;

Training Accuracy : 0.2813333333333333

Testing Accuracy : 0.3

SVM modelinden çıkan sonuç;

Training Accuracy : 0.2813333333333333

Testing Accuracy : 0.3

Raporumuzda başarının tanımını matematik, yazma ve okuma notlarından yola çıkarak hesapladığımız genel bir not (total not) olarak yaptık. Özel ders almanın başarıya etkisi bulunmaktadır. Fakat verilerden elde ettiğimiz korelasyon tablosuna bakarsak başka değişkenlere göre özel ders almanın etkisi çok da büyük değildir.

1. Hangi faktörlerin başarıya etkisi yoktur?

df\_corr['Total Not'].sort\_values(ascending=False)

Cinsiyet 0.130861

Okul Yemekhanesi -0.290064

Cinsiyetin ve okulda öğlen yemeği yemenin genel başarıya etkisi bulunmamaktadır. Okul yemekhanesinde ters orantı ortaya çıkmıştır. Okulda öğlen yemeği yemeyen öğrencilerin notları daha yüksektir.

1. Hangi faktörler başarıyı en çok etkilemektedir?

df\_corr['Total Not'].sort\_values(ascending=False)

Okuma 0.970331

Yazma 0.965667

Matematik 0.918746

Ozel Ders 0.256710

Ebeveyn Egitim Seviyesi 0.214121

Genel başarıyı tahmin ederken önemli olan faktörler matematik, okuma ve yazma notlarıdır. Ek olarak özel ders almanın ve ebeveynlerin eğitim seviyeleri de genel başarıya etki etmektedir.

1. Okuma becerisinin yazma ve matematik üzerindeki etkisi nedir?

df\_corr['Okuma'].sort\_values(ascending=False)

Yazma 0.954598

Matematik 0.817580

Lojistik Regresyon modelinden çıkan sonuç;

Training Accuracy : 0.104

Testing Accuracy : 0.1

SVM modelinden çıkan sonuç;

Training Accuracy : 0.09466666666666666

Testing Accuracy : 0.068

Eğer okuma becerisi yüksekse matematik ve yazma becerisi de ona göre yüksektir.