**Data Visualization with Python**

28.04.2023

**Data Visualization introsu**

**Matplotlib**

**Seaborn**

**Plotly-3D**

**Visualization project**

**Explorative Data Analysis(EDA) icin kullanilan gorsellestirme kutuphaneleri**

**Matplotlib:** create figure, axes, subplots (import matplotlib.pyplot as plt)

**Seaborn:** built on matplotlib and can be used together with it (import seaborn as sns)

2002’de John Hunter Matlab arayuz icin ilk surumunu yayinliyor.

**Functional:**

Plt.plot(age, salary) : burda age x ekseni, y ekseni salary

Plt.xlabel(“age”)

Plt.ylabel(“salary”)

Plt.title(“Salary by Age”)

\*

\*plotlamak kavrami kullanilir.

**Object-oriented:**

Fig, ax **=** plt.subplots() figure axes giriyorum.

ax.plot(age, salary, “r”)

ax.set\_xlabel(“Age”)

ax.set\_ylabel(“Salary”)

Ax.set\_title(“Salary by Age”)

Fig=plt.figure()

Burasi axes ve fgure ile oynayabilecegim tek yer olacak.

Object oriented’de functional’dan farkli olarak bir atama var.

Her bir x,y ekseni birer axis, bunlardan olusan grafik axes olacak benim icin

Figure siniftaki tahta, tahtaya asilan panolar axes gibi dusun.

Axes axis’in cogulu.

**Seaborn:** Matplotlib uzerine insa edilmistir. Istatisteksel analizlerde kullanilir. Yazilimi daha kolay. Gorselleri daha kuvvetlidir.

seaborn’de subplots olacak.

Figure=fisize(6,4)

**Scatter Plot**

Bir datanin normal dagilim gosterip gostermediginin goruntusunu verir.

X eksenini ve y eksenini array olarak tanimladim diyelim

import matplotlib.pyplot as plt

import numpy as np

x = np.array([5,7,8,7,2,17,2,9,4,11,12,9,6])

y = np.array([99,86,87,88,111,86,103,87,94,78,77,85,86])

plt.scatter(x,y)

plt.show

Dedigim zaman, scatter grafigini veriyor

Plt.show yerine, **;** kullanirsam da bir satirlik aciklama yazisini gormem.

Indekse gore siralama yaptigindan, x ve y eksenindeki eleman sayilarinin esit olmasi gerekiyor

Yani bire bir orten.

**Bar Chart**

Kategorik veri kullaniyorsam bir data icinde bar chart kullaniyorum.

En azindan x ekseninde class olan yerlerde kullaniyorum. Classlama var yani

Unique degerlerine, frekanslarina, moduna, Count, meanine, bakabilirim ben bu kategorik datanin bakabilirim.

**Histogram**

Histogram da bar chart’la cok benziyor. Birbiri ile karistiriliyor. Ama histogram numeric

Aralik olarak datalar varsa.

Histogram normal dagilimi gosteren grafiktir. Mesela can egrisi Normal dagilimdir.

**Box Plot**

x normal dagilimli bir array olsun diyelim

plt.boxplot(x,)

plt.show

boxplot’da outlier’a bakarim. Cok guzel gorunur.

ML=prediction

Target

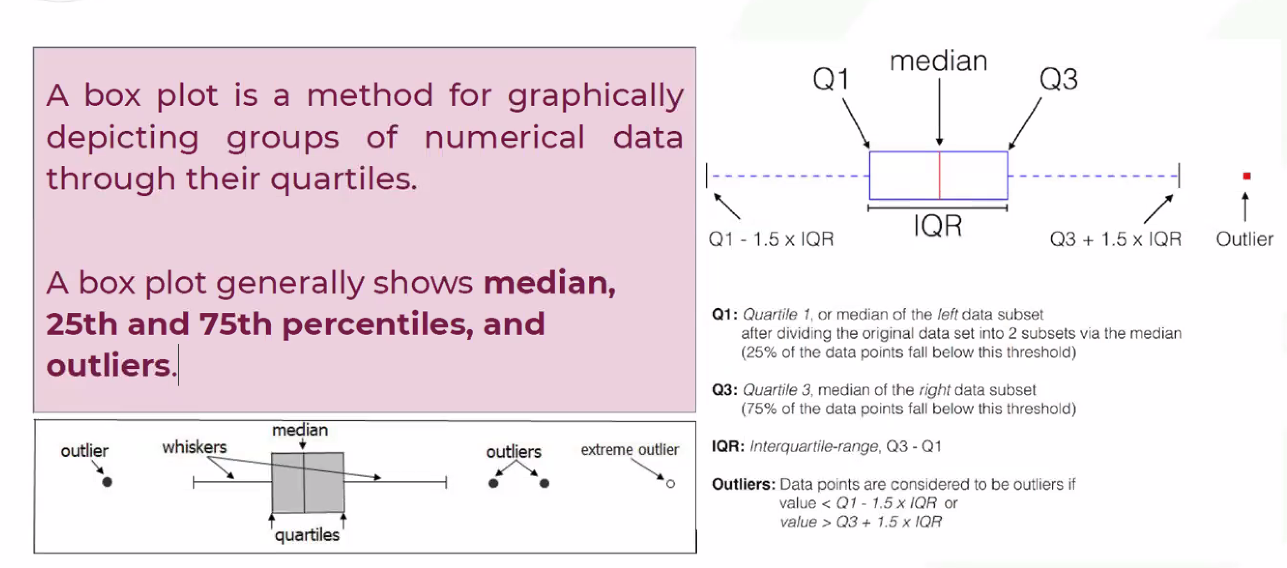
Egitim datasi ile test datasini kiyasliyorum.

Makineye ogretiyorum, %98 accuracy ile predict ediyor mesela.

Piskolojiyi olcmek anketlerde ozellikle piskometri ile yapiliyor. En gercekci sonuclari veriyor

Box’in icindeki cizgi median

Q1-1.5IQR ve Q3+ 1.5 IQR disindakiler outlier.



**Pie Chart**

En fazla 5-6’yi gecmeyecek sekilde kategorik data icin kullanirim.

**Distributions plots (dagilimlarda bakiyorum)**

kdeplot

rugplot

displot

Histplot

**Categorical Plots**

Barplot

Countplot

Boxplot

Swarmplot

Violinplot

**Comparison Plots**

Joinplot

**Pairplot (en cok bunu kullaniyoruz)**

Catplot

Matrix plot

Grid plot

|  |
| --- |
| Import seaborn as sns  tips=sns.load\_dataset(“tips”)  **sns.rugplot(x=”total\_bill”, data=tips)**  **sns.rugplot(tips[”total\_bill”])**  **#bu ikisi ayni gorseli verir.** |

**sns.rugplot(x,y,data)**

(y yerine tek degiskenle bakmak istersem sadece x veriyorum. Bu durumda bana y ekseninde yogunlugu veriyor (density))

Datamin icinde bulunan

**columns,feature**

**Row, index, satir, observation**

Displot kutuphanede dagilimlari gostermek icin ilk cikan fonksiyon. Zamanla gelistikce yerine hisplot gibi baskaca seyler geldi. Displot yazinca icerisinde hangi func calissin istedigimi yazabildigim baskaca plot’lar var.

**Rugplot’da y ekseninde yogunluk cikiyordu.**

**hisplot’da ise y ekseninde count cikiyor.**

**sns.hisplot(x=”total\_bill”, data=tips, kde=True)**

Default da olsa icine yazmakta fayda var projeyi bir yerde paylasacaksan

sns.hisplot(tips[“total\_bill”])

**Distribution plots-kde**

kdeplot da rugplot gibi calisiyor. Sadece x verdigimde, y’de yogunlugu gosteriyor

**Categorical Plots -countplot**

Sns.countplot (x= ‘day’, data=tips)

Sns.countplot(tips[‘day’]

Bir de **hue= ‘sex’** dedigimde mesela, icine verdigimiz kategorik degiskene gore renklendiriyor datayi.

Eger targetimiz kategorik bir degiskense, hue icin mutlaka o target’taki kategorik degiskeni veririrm. Yani train ettigimiz sutuna gore olmali.

Yani mesela smoker’lari tahmin edeceksem modelimde. hue= “smoker” derim.

**Categorical Plots- barplots**

sns.barplot(x=’day’, y=’total\_bill’, data=tips)

**sns.barplot(**x=’day’, y=’total\_bill’, data=tips, hue=’sex’) bu sekilde yazdigimda cinsiyete gore renklendirerek getiriyor grafigi

**Categorical Plots- barplots**

Gunlere gore toplam bahsisin dagilimini gosteren grafik

**sns.boxplot(**x=’day’, y=’total\_bill’, data=tips)

Median cizgisi bir tarafa daha yakinsa, carpiklik var demektir. Tam ortadaysa normal dagiliyor demektir.

Outlier tespit edince napiyorsun:

Gercek dunya datalari outlierlardan olusur. Hemen atamazsin.

**Comparison Plots-pairplot**

Birseyin aynisiyla kendisiyle korelasyonu 1’dir. Fahreneit ile celcius korelasyonu da 1 cikar mesela. Ayni seyi olcenlerden hangisinde daha az missing value varsa onu tutabiliriz aslinda.

sns.pairplot(tips, hue= “smoker”)

#corner=True

**Grids**

**Pairgrid**

Pairplottan farki ustte hangisi olsun, altta hangisi olsun secebiliyorum.

Yukariya scatter koy diyorum mesela

**g=g.map\_diag(sns.histplot)**

**FacetGrid**

Satirda ne yer alsin, sutunda ne yer alsin onu seciyorum bir de bunda.

**#facetgrid ile pairgrid cok kullanilmaz.**

**Matrix Plots**

**Heatmap (isi haritasi) (onemli grafik)**

Korelasyon tablolarinda bakilir. Hangi feature’un hangi feature ile core

1’e yaklasanin correlation’i yuksek demektir. Hem 1 hem -1 icin.

Birlikte hareket etme durumu.