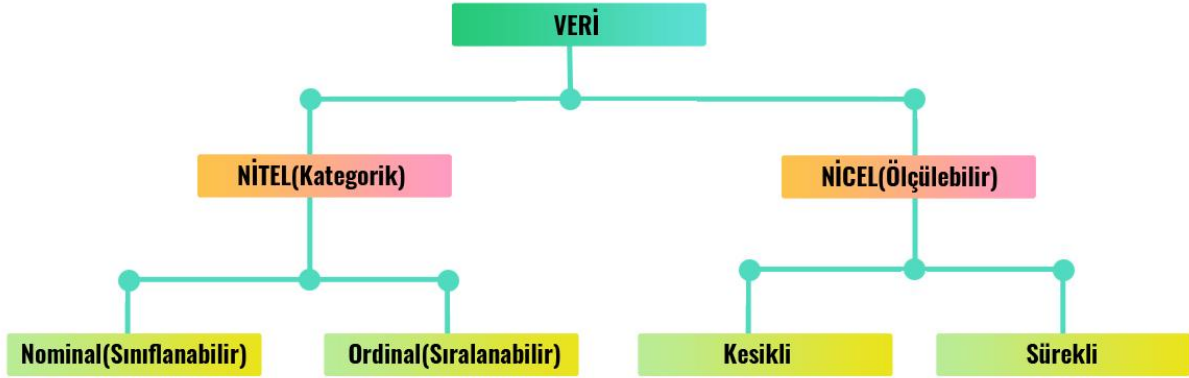


## Veri Görselleştirme

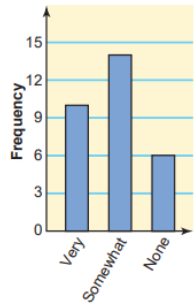
Veri görselleştirme, karmaşık veri kümelerini anlamlandırmak ve verileri daha kolay yorumlanabilir hale getirmek için kullanılan güçlü bir araçtır. Doğru çizilmiş bir grafik ,uzun açıklamalardan daha etkilidir.Bu yazıda kategorik ve nicel verilerin nasıl görselleştirileceği ve bu amaçla kullanılacak çeşitli grafik türleri ele alınacaktır. Ayrıca, Python programlama dilinde veri görselleştirmek için yaygın olarak kullanılan Matplotlib, Seaborn ve Plotly kütüphanelerine değinilecektir. Yazının sonunda, bu kütüphanelerin basitçe kullanım kodları sunulacaktır.



- Verilerimizi grafiğe dökmeden önce , görselleştirmek istediğimiz değişkenin türünü tanımamız gerekiyor.

### 1. Kategorik Verilerin Görselleştirilmesi

- Kategorik verilerin görselleştirilmesi, temel olarak **frekans** ve **oranlar** üzerine kuruludur. Kategorik veriler, belirli bir kategoriye düşen veri noktalarının sayısını (frekansını) veya bu frekansların toplam içindeki oranını gösterir.
- Çubuk grafik (bar graph) ve pasta grafik (pie chart), nitel verileri göstermek için yaygın olarak kullanılan iki grafik türüdür.

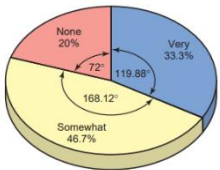


#### I. Bar (Çubuk) Grafiği

Çubuk Grafik (Bar Graph): Yükseklikleri, ilgili kategorilerin frekanslarını temsil eden çubuklardan oluşan, her çubuğun genişliği eşit olan bir grafiğe çubuk grafik denir. Burada önemli olan çubuğun yüksekliği yani frekansdır.

#### II. Pie(Pasta) Grafiği

Pasta grafik , genellikle yüzdeleri göstermek için kullanılır, ancak frekansları veya göreceli frekansları göstermek için de kullanılabilir. Tüm pasta (veya daire) toplam örnekleme veya popülasyonu temsil eder. Ardından, pastayı farklı kategorileri temsil eden dilimlere böleriz.

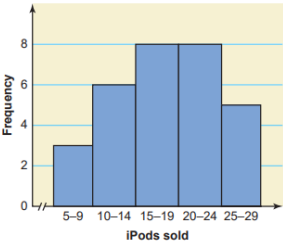


## 2. Nicel Verilerin Görselleştirilmesi

- Nicel değişkenler, sayısal olarak ölçülebilen değişkenlerdir.
- Veriyi ifade etmek için histogram, box plot , violin vb. grafik çeşitleri vardır.
- Nicel veriler için bir frekans dağılımı, tüm sınıfları ve her sınıfa ait olan değerlerin sayısını listeler. Frekans dağılımı şeklinde sunulan verilere gruplandırılmış veri denir.

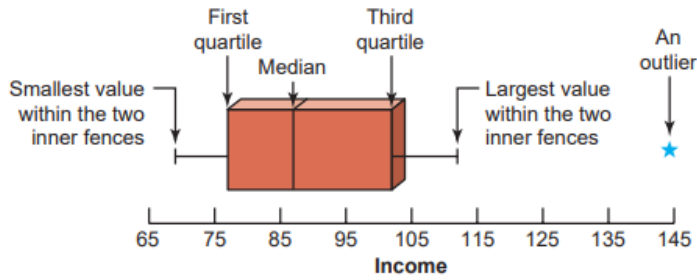
### a. Histogram

Gruplandırılmış (nicel) veriler bir histogram şeklinde gösterilebilir. Histogram, sınıfların yatay ekseninde ve frekansların, göreceli frekansların veya yüzdelerin dikey ekseninde işaretlendiği bir grafik türüdür. Frekanslar, göreceli frekanslar veya yüzdeler çubukların yükseklikleriyle temsil edilir. Histogramda, çubuklar birbirine bitişik olarak çizilir.

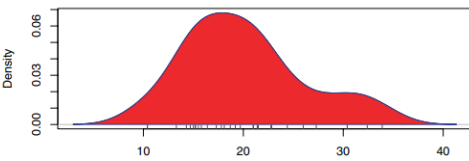


### b. Box Plot

Bir kutu ve saçak grafiği (box-and-whisker plot), verilerin beş ölçümünü kullanarak grafiksel bir sunum sağlar: birinci çeyrek, medyan (ikinci çeyrek), üçüncü çeyrek, ve veri setindeki en küçük ve en büyük değerler kullanılarak bir kutu ve iki saçak çizilerek oluşturulur. Bu kutu ve saçaklar, veri setinin alt ve üst iç çitler (inner fences) arasındaki değerleri temsil eder. Bir kutu ve saçak grafiği, veri setinin merkezini, yayılımını ve çarpıklığını görselleştirmemize yardımcı olabilir. Ayrıca aykırı değerleri tespit etmede yardımcı olur.



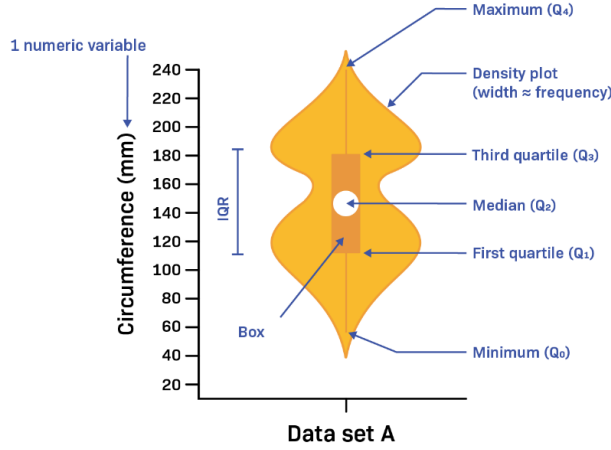
### c. Yoğunluk Grafiği (Density Plot)



**Yoğunluk grafiği** (density plot), sürekli bir değişkenin dağılımını görselleştiren bir grafik türüdür. Bu grafik, verilerin yoğunluk dağılımını pürüzsüz bir eğri ile gösterir. Yoğunluk grafikleri, veri setinin olasılık yoğunluk fonksiyonunu tahmin eder ve histogramlara göre daha düzgün ve kesintisiz bir dağılım görünümü sunar.

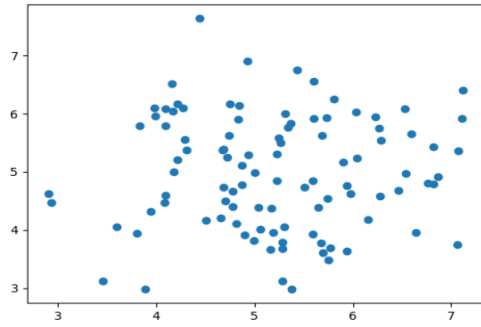
#### d. Violin Grafiği

Violin plot, bir kutu grafiği ve bir çekirdek yoğunluk grafiğinin (kernel density plot) kombinasyonudur. Violin plot'un ortasında, bir kutu grafiği görünümü bulunur. Bu kutu, medyanı, birinci ve üçüncü çeyrekleri (quartiles), ve olası aykırı değerleri gösterir. Violin plot'un iki tarafında, verilerin yoğunluk dağılımını gösteren eğriler bulunur. Bu eğriler, verilerin hangi aralıklarda yoğunlaştığını gösterir ve veri setinin biçimini (örneğin, simetrik veya çarpık olup olmadığını) anlamaya yardımcı olur.



#### e. Scatter Plot

Bir çift gözlemin grafik üzerinde gösterilmesi saçılma diyagramı olarak adlandırılır. Saçılma diyagramı, iki değişken arasındaki ilişkiyi görselleştirmek için kullanılan bir grafik türüdür.



## Python'da Veri Görselleştirme

Kullanılan Kütüphaneler :

#### 1) Matplotlib

Python'da 2D grafikler oluşturmak için kullanılan bir kütüphanedir.

Python'da Kullanım örneği:

```
import matplotlib.pyplot as plt
```

```
# Çubuk Grafiği
```

```
plt.bar(x, y); plt.title('Çubuk Grafiği'); plt.show()
```

```
# Histogram
```

```
plt.hist(data, bins=10); plt.title('Histogram'); plt.show()
```

```
# Saçılma Diyagramı
```

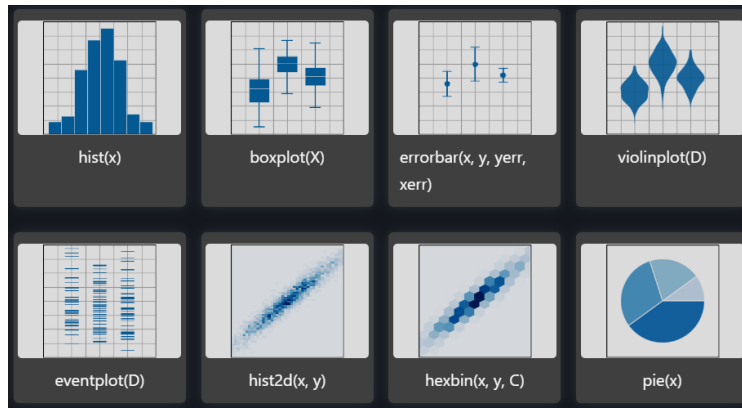
```
plt.scatter(x, y); plt.title('Saçılma Diyagramı'); plt.show()
```

```
# Pasta Grafiği
```

```
plt.pie(sizes, labels=labels); plt.title('Pasta Grafiği'); plt.show()
```

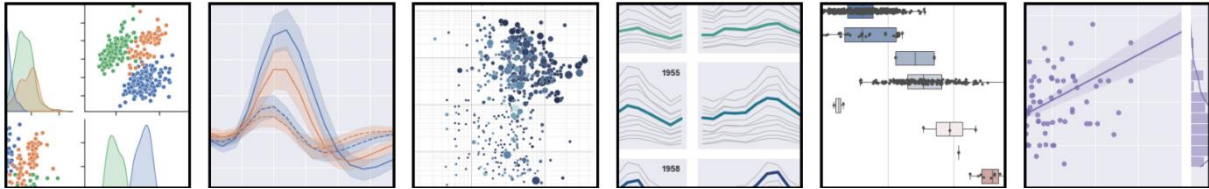
```
# Kutu ve Kuyruk Grafiği
```

```
plt.boxplot(data); plt.title('Kutu ve Kuyruk Grafiği'); plt.show()
```



## 2) Seaborn

Seaborn, Python'da istatistiksel grafikler oluşturmak için bir kütüphanedir. Matplotlib'in üzerine kuruludur.



```
import seaborn as sns
```

# Çubuk Grafiği

```
sns.barplot(x='class', y='fare', data=sns.load_dataset('titanic')); plt.title('Çubuk Grafiği');  
plt.show()
```

# Dağılım Grafiği

```
sns.scatterplot(x='sepal_length', y='sepal_width', hue='species',  
data=sns.load_dataset('iris')); plt.title('Dağılım Grafiği'); plt.show()
```

# Histogram ve Yoğunluk Grafiği

```
sns.histplot(data=sns.load_dataset('iris'), x='sepal_length', kde=True); plt.title('Histogram ve  
Yoğunluk Grafiği'); plt.show()
```

# Kutu ve Kuyruk Grafiği

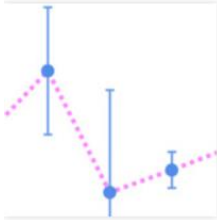
```
sns.boxplot(x='day', y='total_bill', data=sns.load_dataset('tips')); plt.title('Kutu ve Kuyruk  
Grafiği'); plt.show()
```

# Violin Grafiği

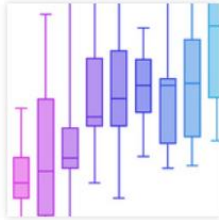
```
sns.violinplot(x='day', y='total_bill', data=sns.load_dataset('tips')); plt.title('Violin Grafiği');  
plt.show()
```

### 3) Plotly

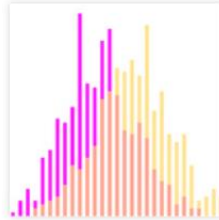
Plotly'nin Python grafik kütüphanesi etkileşimli, yayın kalitesinde grafikler oluşturur. Çizgi grafikleri, dağılım grafikleri, alan grafikleri, çubuk grafikleri, hata çubukları, kutu grafikleri, histogramlar, ısı haritaları, alt grafikler, çoklu eksenler, kutup grafikleri ve balon grafikleri vb. tüm grafikleri çizdirebilir.



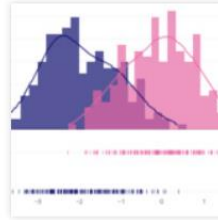
Hata Çubukları



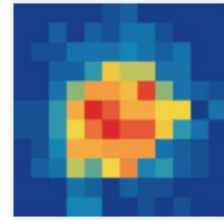
Kutu Grafikleri



Histogramlar



Distplots



2D Histogram

```
import plotly.express as px

# Veri kümesi
df = px.data.tips()

# Çubuk Grafiği
fig1 = px.bar(df, x='day', y='total_bill', color='sex', title='Çubuk Grafiği')
fig1.show()

# Dağılım Grafiği
fig2 = px.scatter(df, x='sepal_length', y='sepal_width', color='species', title='Dağılım Grafiği')
fig2.show()

# Histogram
fig3 = px.histogram(df, x='total_bill', nbins=20, title='Histogram')
fig3.show()

# Kutu Grafiği
fig4 = px.box(df, x='day', y='total_bill', title='Kutu Grafiği')
fig4.show()

# Violin Grafiği
fig5 = px.violin(df, x='day', y='total_bill', box=True, points='all', title='Violin Grafiği')
fig5.show()
```