

# Veri Hazırlama ve Önişleme

Dr. Caner Erden [cerden@sakarya.edu.tr](mailto:cerden@sakarya.edu.tr)

Sakarya Üniversitesi, Endüstri Mühendisliği Bölümü



**SAKARYA**  
ÜNİVERSİTESİ

# Veri nedir?

**Özellik** ve veri nesnelerinden oluşan koleksiyon

Özellik veriyi açıklayan veya bir karakteristiğini ortaya koyan bilgi demektir.

Veri, kayıt, nokta, durum, örnek, gözlem, girdi gibi de isimlendirilebilir.

Özellikler, karakteristik, değişken, boyut, alan olarak da isimlendirilebilir.

**Nesneler**

**Özellikler**

No	Indirim	Medeni Hal	Gelir	Sipariş
1	Evet	Bekar	125K	Hayır
2	Hayır	Evli	100K	Hayır
3	Hayır	Bekar	70K	Hayır
4	Evet	Evli	120K	Hayır
5	Hayır	Dul	95K	Evet
6	Hayır	Evli	60K	Hayır
7	Evet	Dul	220K	Hayır
8	Hayır	Bekar	85K	Evet
9	Hayır	Evli	75K	Hayır
10	Hayır	Bekar	90K	Evet

# Veri Tipleri

---

1. **Nominal veriler:** Daha fazla ifadesi ile kullanılmazlar. ID, göz rengi vb.
2. **Sıralama Ölçeği (Ordinal):** Daha fazla ifadesi ile kullanılabilir. Uzunluk (uzun, orta boy, kısa) vb.
3. **Aralıklı (interval):** Fahrenheit, Celcius vb.
4. **Oran (Ratio):** Başlangıç noktası 0'dır ve 0 noktası yokluk ifade eder. Kelvin, uzunluk (cm), zaman gibi.

# Oran ve Aralıklı Veri Tipi

---

- Sıcaklık gibi iki değer arasında katı olarak ifade edilemeyen durumlarda Aralıklı(interval) veri tipi tanımlaması yapılır. Örneğin  $50^{\circ}$   $100^{\circ}$ 'nin yarısıdır diyemeyiz. Yani aralarında belirli bir oran söz konusu değildir.
- Uzunluk ya da ağırlık değişkeninde ise 50 kg 100 kg'nin yarısıdır deriz. Bu nedenle aralarında bir oran söz konusudur.

	Veri Tipi	Açıklama	Örnek
Kategorik (Kalitatif) Veriler	Nominal	Eşitlik ya da eşitsizlik durumu söz konusu ( $=$ , $\neq$ )	PK Kodu, göz rengi, cinsiyet vb.
	Ordinal	Aralarında büyüklük küçüklük ilişkisi var ( $<$ , $>$ , $\leq$ , $\geq$ )	Not (AA, BA, BB), Kalite Derecesi (İyi, Orta, Kötü)
Nümerik (Kantitatif) Veriler	Interval	Toplama çıkarma gibi işlemler yapılabilir (+, -)	Takvim günleri, Santigrat ya da Fahrenheit gibi sıcaklık ölçüsü
	Ratio	İki değer arasındaki oran anlamlıdır. (*, /)	Kelvin sıcaklığı, yaş, ağırlık, uzunluk gibi.

# Veri Dönüşümleri

Veri Tipi		Dönüşüm	Örnek
Kategorik	Nominal	Herbir değer başka bir değerle değiştirilir.	Öğrenci numarasının başka değerle değiştirilmesi
	Ordinal	Değerler sıralama olacak şekilde değiştirilir.	İyi, orta, kötü değerlerinin 3,2,1 ile değiştirilmesi
Nümerik	Interval	$\text{Yeni\_değer} = \text{eski\_değer} + b$	Fahrenayt ile celsius arasındaki dönüşüm
	Ratio	$\text{Yeni\_değer} = a * \text{eski\_değer}$	Metre ve feet arasındaki dönüşüm

# Sürekli ve Kesikli Değişkenler

---

- Kesikli Değişken
  - Belirli sayıda değere sahiptir.
  - Genellikle tam sayılı değerler ile ifade edilir.
  - İkili (Binary) gösterim bu gösterimin bir çeşididir.
- Sürekli Değişken
  - Sınırsız sayıda değere sahiptir.
  - Genellikle küsuratlı değerler ile ifade edilir.

# Temel İstatistik Hesaplamaları

---

- Merkez eğilim ölçüleri
  - Verinin ortası ya da orta noktasını ölçer (mod, medyan, ortalama)
- Dağılım ölçüleri
  - Verinin nasıl dağıldığını ölçer. (aralık, varyans, standart sapma)



# Merkezi Eğilim Ölçüleri

---

- **Ortalama**

- $$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

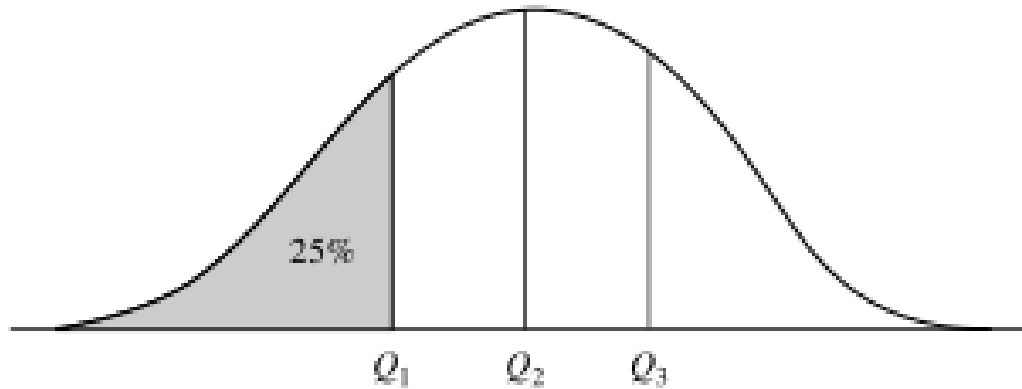
- **Ağırlıklı Ortalama**

- $$\bar{x} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_N x_N}{w_1 + w_2 + \dots + w_N}$$

- **Medyan:** Ortadaki Sayı

- **Mod:** En çok tekrar eden sayı

# Dağılım Ölçüleri



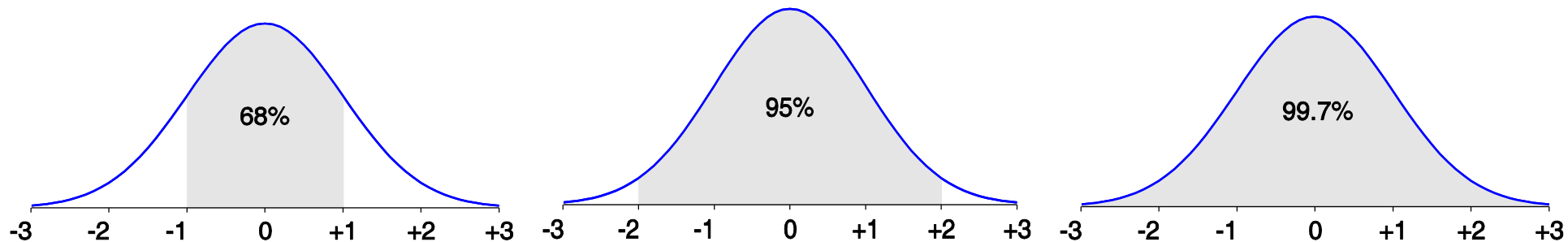
- **Aralık(Range):** Min-Max arasında kalan değer
- **Kantiller(Quantiles):** Eşit aralığa bölünen değer
- **IQR(interquartile range)**
- $IQR = Q_3 - Q_1$

# Normal Dağılım

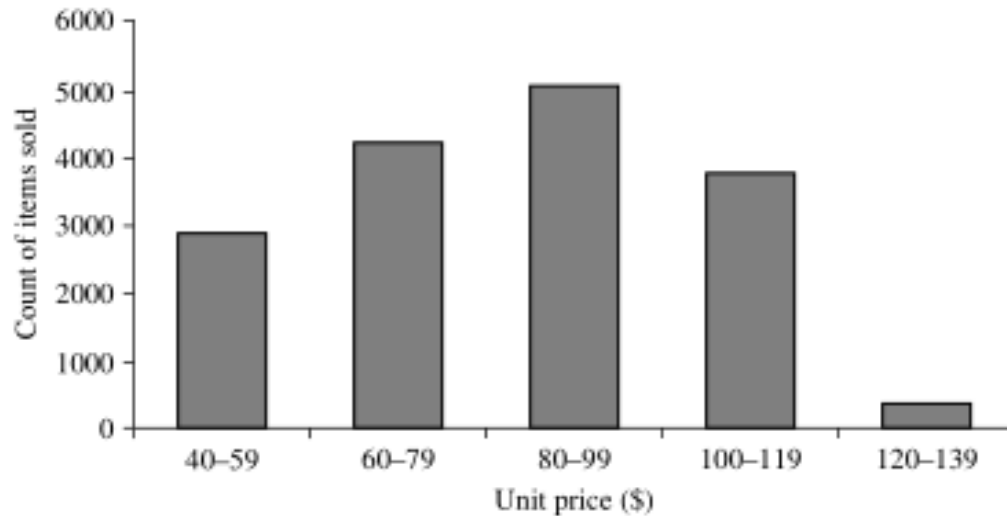
$\mu - \sigma$  ile  $\mu + \sigma$  arasındaki veriler toplam verinin %68'dir.

$\mu - 2\sigma$  to  $\mu + 2\sigma$  arasında %95 veri bulunur.

$\mu - 3\sigma$  to  $\mu + 3\sigma$  arasında %99.7 veri bulunur.



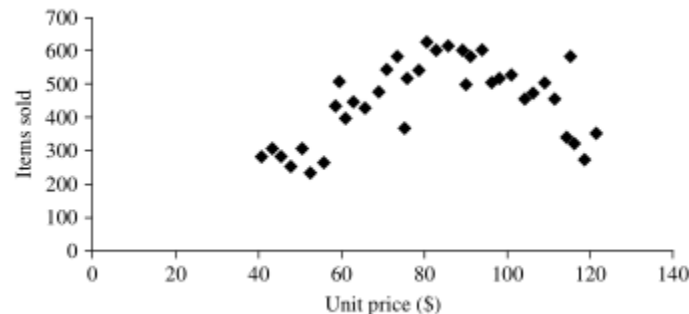
# Histo-gram (Kutup Grafikleri)



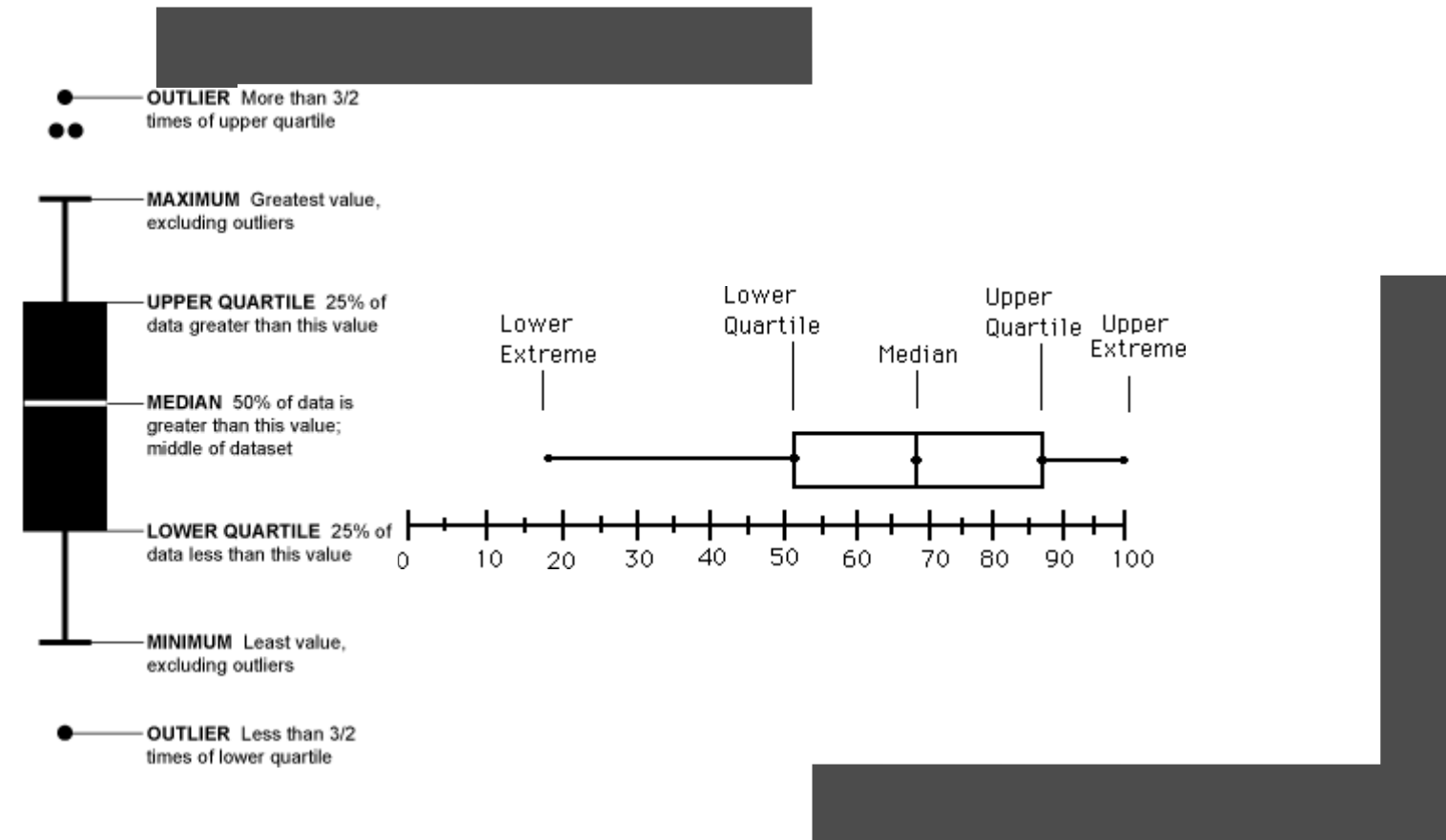
- Verinin dağılımı ile ilgili bilgi verir. Kutucuğun yüksekliği frekansı gösterir.
- Herbir kategoride ne kadar olay gerçekleştiğini gösterir.
- x ekseninde değerler ve y ekseninde değerlerin frekansı

# Dağılım(Scatter) Grafiği

- Veriler arasında ilişki, trend ya da örüntü olup olmadığını kontrol için çizilir.
- Herbir ikili değer x ve y eksenindeki noktada işaretlenir.
- Noktalar ve aykırılıklar hakkında hızlı bir görüş açısı sağlar.
- Her bir nokta veri ve özellik hakkında bilgi verir.



# Kutu Grafikleri (Box Plots)



- Kutu grafikleri aşğıdaki gibi özet bilgiler sağlar;
- 1. Kutu uzunluğu IQR değerini gösterir.
- 2. Medyan ortada işaretlenmiştir.
- 3. İki çizgi(whiskers) en küçük ve en büyük değerleri gösterir.
- 4.  $1.5 \times \text{IQR}$  değerleri aykırı verilerdir.

# Varyans ve Standart Sapma

---

- $\sigma^2 = \frac{1}{N} \sum_i^N (x_i - \bar{x})^2 = \left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right) - \bar{x}^2$
- Varyans Hesabı
- Veriler 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110
- $\bar{x} = \frac{30+36+47+50+52+52+56+60+63+70+70+110}{12} = \frac{696}{12} = 58$
- $\sigma^2 = \frac{1}{12} (30^2 + 36^2 + \dots + 110^2) - 58^2$   
 $= 379,17$

# Kayıt Veriseti

- Bir gözleme ait verilerin tutulduğu verilerden oluşan veri seti.

No	Indirim	Medeni Hal	Gelir	Sipariş
1	Evet	Bekar	125K	Hayır
2	Hayır	Evli	100K	Hayır
3	Hayır	Bekar	70K	Hayır
4	Evet	Evli	120K	Hayır
5	Hayır	Dul	95K	Evet
6	Hayır	Evli	60K	Hayır
7	Evet	Dul	220K	Hayır
8	Hayır	Bekar	85K	Evet
9	Hayır	Evli	75K	Hayır
10	Hayır	Bekar	90K	Evet



# Veri Matrisi

---

- Eğer veri objesi sabit sayıda nümerik verilerden oluşuyorsa bu veri setine veri matrisi denir ve  $n \times m$  boyutuna sahiptir.

Projection of x Load	Projection of y load	Distance	Load	Thickness
10.23	5.27	15.22	2.7	1.2
12.65	6.25	16.22	2.2	1.1

# Belge Veriseti

- Bu veri setinde her belgeye ait bir özellik vardır.

	team	coach	play	ball	score	game	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

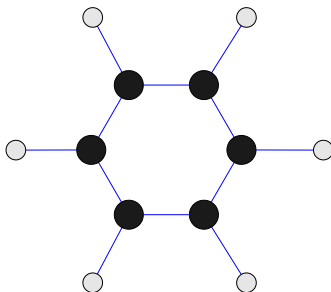
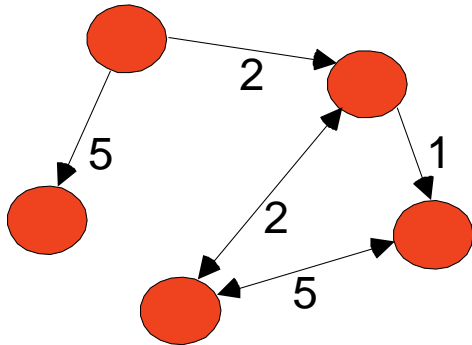
# İşlemler Veriseti

- Özel bir kayıt verisidir. Her kayıt birkaç elemandan oluşan setleri tutar. Örneğin bakkal dükkanının satılan ürünleri sıraladığı liste.

<i>TID</i>	<i>Items</i>
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

# Grafik Veriseti

- Molekül yapısı ya da internet sayfalarının gösterilmesi



## Useful Links:

- [Bibliography](#)
- Other Useful Web sites
  - [ACM SIGKDD](#)
  - [KDnuggets](#)
  - [The Data Mine](#)

## Knowledge Discovery and Data Mining Bibliography

(Gets updated frequently, so visit often!)

- [Books](#)
- [General Data Mining](#)

## Book References in Data Mining and Knowledge Discovery

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/the MIT Press, 1996.

J. Ross Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.  
Michael Berry and Gordon Linoff, "Data Mining Techniques (For Marketing, Sales, and Customer Support)", John Wiley & Sons, 1997.

## General Data Mining

Usama Fayyad, "Mining Databases: Towards Algorithms for Knowledge Discovery", Bulletin of the IEEE Computer Society Technical Committee on data Engineering, vol. 21, no. 1, March 1998.

Christopher Matheus, Philip Chan, and Gregory Piatetsky-Shapiro, "Systems for knowledge Discovery in databases", IEEE Transactions on Knowledge and Data Engineering, 5(6):903-913, December 1993.

# Benzerlik Ölçütleri

---

- Benzerlik ölçüsü
  - İki verinin benzeyip benzemediğinin nümerik ölçüsü
  - [0-1] Arasında bir değerle ölçülür ve 1'e yakın olması daha fazla benzediği anlamına gelir.
- Uzaklık (Benzemezlik) ölçüsü
  - İki verinin birbirinden ne kadar farklı olduğunun ölçülmesi durumu.
  - Eğer benzemez ise değer 0'a yakın çıkar.

# Kümelemede Benzerlik Ölçütleri

---

- Küme bir koleksiyondaki verilerin birbirine benzemesi ve diğer koleksiyondaki verilere benzememesi demektir. Örneğin bir bakkal dükkanı müşterilerini kümelere ayırmak istediğinde müşterilerinin gelir durumları, yaşları gibi özelliklerine bakarak ayırım yapabilir. Bu özellikler arasındaki ilişki ve benzerlik durumu da kişilerin kümeye alınıp alınmayacağını belirler.

# Veri ve Uzaklık Matrisi

---

- Veri Matrisi: 
$$\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$
- Uzaklık Matrisi: 
$$\begin{bmatrix} 0 & d(1, 2) & \cdots & d(1, p) \\ d(2, 1) & 0 & \cdots & d(2, p) \\ \vdots & \vdots & \ddots & \vdots \\ d(n, 1) & d(n, 2) & \cdots & 0 \end{bmatrix}$$
- Benzerlik ölçüsü:  $s(i, j) = 1 - d(i, j)$

# Nominal Verilerin Uzaklık Hesabı

---

- $d(i, j) = \frac{p-m}{p}$
- $m$  : toplam eşleşme sayısı
- $p$  : nesneyi tanımlayan toplam özellik sayısı



# Örnek

No	Özellik-1	Özellik-2	Özellik-3
1	A	Çok İyi	45
2	B	Kötü	22
3	C	İyi	64
4	A	Çok iyi	28

$$\text{Uzaklık Matrisi} : \begin{bmatrix} 0 & & & \\ d(2,1) & 0 & & \\ d(3,1) & d(3,2) & 0 & \\ d(4,1) & d(4,2) & d(4,3) & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 1 & 1 & 0 & \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

$$\text{Benzerlik} : s(i,j) = 1 - d(i,j) = \frac{m}{p}$$

# İkili Verilerin Uzaklık Ölçüsü

- 0 : Olmama Durumu      Simetrik : 2 durum aynı
- 1 : Var Olma Durumu      Asimetrik : 2 durum farklı
- $p = q + r + s + t$

		j nesnesi		
		1	0	Toplam
i nesnesi	1	q	r	q+r
	0	s	t	s+t
	toplam	q+s	r+t	p

# İkili Verilerin Uzaklığını Hesaplama

---

- Simetrik Durum
- $d(i, j) = \frac{r+s}{q+r+s+t}$
- Asimetrik Durum
- $d(i, j) = \frac{r+s}{q+r+s}$
- Benzerlik Oranı (Asimetrik)
- $s(d, j) = 1 - d(i, j) = \frac{q}{q+r+s} \rightarrow \text{Jaccard Katsayısı}$

# İkili Veriler için Örnek

- Aşağıdaki hastalık tablosunu ele alalım.

İsim	Cinsiyet	Öksürük	Ateş	Test1	Test2	Test3	Test4
Ali	E	E	H	P	N	N	N
Faruk	E	E	E	N	N	N	N
Fatma	K	E	H	P	N	P	N
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

# Hesaplamalar

---

- $d(\text{Ali}, \text{Faruk}) = \frac{1+1}{1+1+1} = 0,67$
- $d(\text{Ali}, \text{Fatma}) = \frac{0+1}{2+0+1} = 0,33$
- $d(\text{Faruk}, \text{Fatma}) = \frac{1+2}{1+1+2} = 0,75$
- Bu ölçümler Faruk ve Fatma'nın birbirinden daha uzak olduğunu en yakın ikilinin Ali ve Fatma olduğunu göstermektedir.

# Nümerik Verilerin Uzaklığı

---

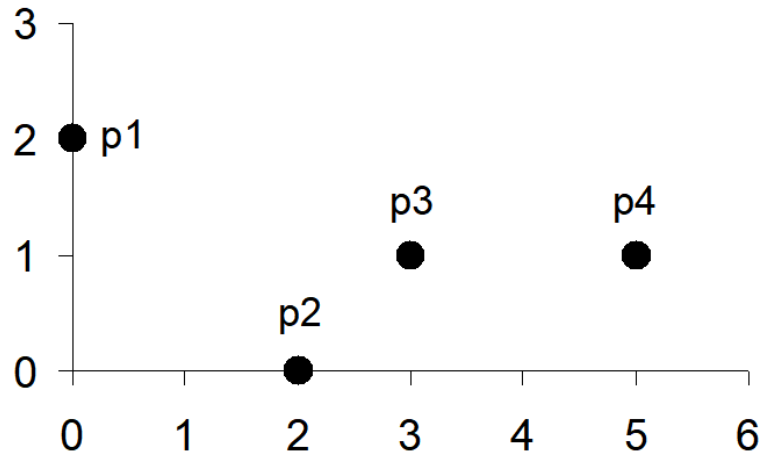
- Öklid: 2 nokta arasındaki uzaklık.

- $$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

- Eğer farklı ölçeklerde ise standardizasyon gereklidir.

- p boyut sayısı,  $i = x_{i1}, x_{i2} \dots x_{ip}$  ve  $j = x_{j1}, x_{j2} \dots x_{jp}$

# Öklid Uzaklığı Hesaplama



point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

	p1	p2	p3	p4
p1	0	2.828	3.162	5.099
p2	2.828	0	1.414	3.162
p3	3.162	1.414	0	2
p4	5.099	3.162	2	0

# Manhattan Uzaklık Hesabı

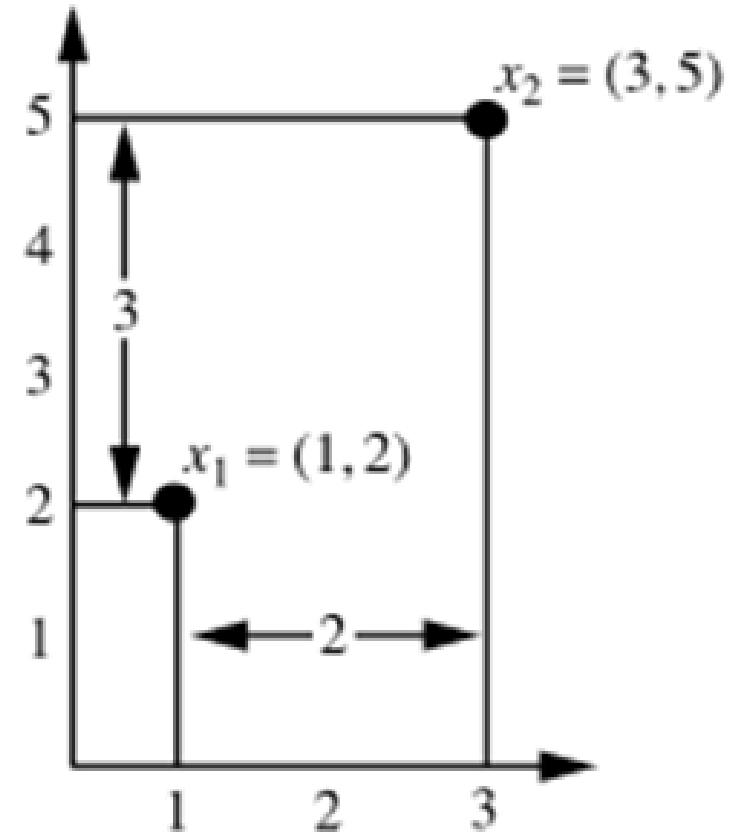
---

- $d(i, j) = \sum_{k=1}^n |i_k - y_k|$
- Özellikleri
  - 1. Negatif olmama koşulu :  $d(i, j) \geq 0$
  - 2. Aynı şehirde olma :  $d(i, i) = 0$
  - 3. Simetrik olma koşulu :  $d(i, j) = d(j, i)$
  - 4. Üçgen eşitsizliği :  $d(i, j) \leq d(i, k) + d(k, j)$
- Eğer bu özellikler sağlanırsa ölçüt **metriktir**.



# Manhattan Uzaklığı Örnek

- Örneğin  $x_1 = (1,2)$  ve  $x_2 = (3,5)$  olsun.
- Öklid uzaklığı :  $\sqrt{2^2 + 3^2} = 3,61$
- Manhattan uzaklığı :  $2 + 3 = 5$



# Ordinal Veriler için Uzaklık Hesabı

---

- $M \rightarrow$  ordinal listenin alabileceği değerler
- Adımlar
- Özellikler arasında  $\{1,2,\dots,M\}$  şeklinde sıralama yap.
- Normalizasyon yap.
- Nümerik veriler için yapılan hesaplamaları burada da yap.

# Ordinal veriler için örnek

- $M=3$  {Çok iyi: 3, İyi: 2, Kötü: 1}
- Atamayı yaparsak özellik-2  $\rightarrow \{3,1,2,3\}$  olur.
- Normalizasyon yaparsak [0-1] Çok İyi: 1, İyi: 0.5, Kötü: 0 olur.

No	Özellik-1	Özellik-2	Özellik-3
1	A	Çok İyi	45
2	B	Kötü	22
3	C	İyi	64
4	A	Çok iyi	28

# Ordinal Veriler Örnek

---

- Sonuç Matrisi  $\rightarrow \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 0,5 & 0,5 & 0 & \\ 0 & 1 & 0,5 & 0 \end{bmatrix}$
- $d(1,2)$  ve  $d(4,2) = 1$  olduğu için en uzak veriler 1-2 ve 4-2 sonucu çıkarılır.
- Benzerlik ölçüsü için  $s(i, j) = 1 - d(i, j)$  formülü kullanılır.

# Kosinüs Benzerliği

	Takım	Koç	Futbol	Oyun	Penaltı	Gol	Sezon
Doküman1	5	3	0	0	3	2	0
Doküman2	3	0	2	1	0	1	1
Doküman3	0	7	0	0	1	3	0

$$s(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

$\|x\| \|y\|$  : x ve y vektörlerinin Öklid formu

$$\begin{aligned} \mathbf{x}^t \cdot \mathbf{y} &= 5 \times 3 + 0 \times 0 + 3 \times 2 + 0 \times 0 + 2 \times 1 + 0 \times 1 + 0 \times 0 + 2 \times 1 \\ &\quad + 0 \times 0 + 0 \times 1 = 25 \end{aligned}$$

$$\|\mathbf{x}\| = \sqrt{5^2 + 0^2 + 3^2 + 0^2 + 2^2 + 0^2 + 0^2 + 2^2 + 0^2 + 0^2} = 6.48$$

$$\|\mathbf{y}\| = \sqrt{3^2 + 0^2 + 2^2 + 0^2 + 1^2 + 1^2 + 0^2 + 1^2 + 0^2 + 1^2} = 4.12$$

$$\text{sim}(\mathbf{x}, \mathbf{y}) = 0.94$$

# Veri Önışleme

---

- Birleřtirme
- Örneklem Alma
- Özellik İndirgeme
- Kesikli ya da İkili Hale Getirme
- Özellik Dönüşümleri
- Veri Görselleřtirme

# Veri Kalitesi

---

- Veri Kalitesini Etkileyen Faktörler
  - Doğruluk
  - Tam olma durumu
  - Tutarlılık
  - Güvenirlik
  - Tahmin edilebilirlik

# Veri Kalitesi

---

- Zayıf veri kümesi verilerin analizi aşamasında sorunlar doğurabilir.
  - Gürültü
  - Eksik Veriler
  - Tekrarlı Veriler
  - Yanlış Veriler



# Eksik Veriler

---

- Tamamen rastgele eksik olan veriler
  - Tüm değişkenlere ait verilerden herhangi birisinde eksik veri yer alabilir.
  - Bu veriler rastgele bir değer ile doldurulabilir.
- Diğer Doldurma Metotları
  - Ortalama ile doldurma
  - Mod, Medyan ile doldurma
  - Akıllı bir algoritma kullanarak doldurma

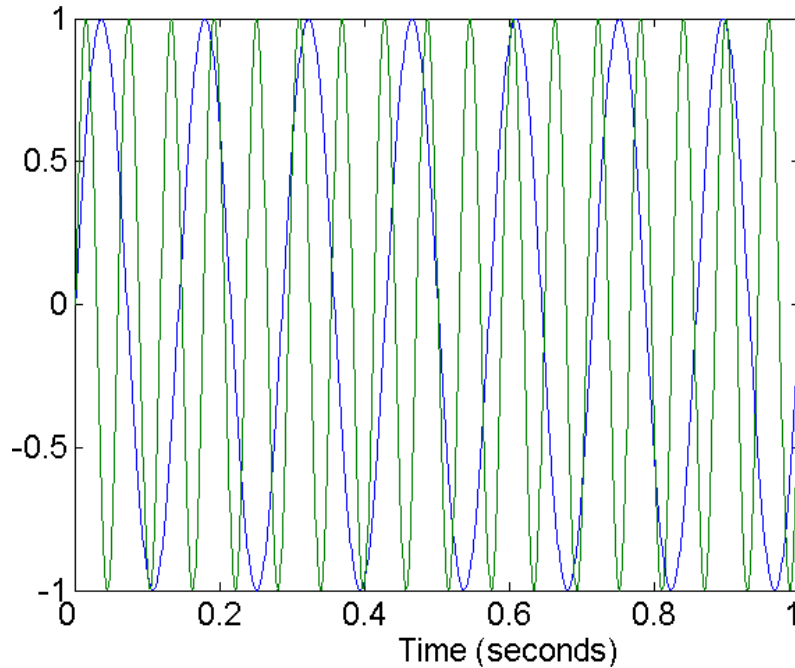
# Eksik Veriler

---

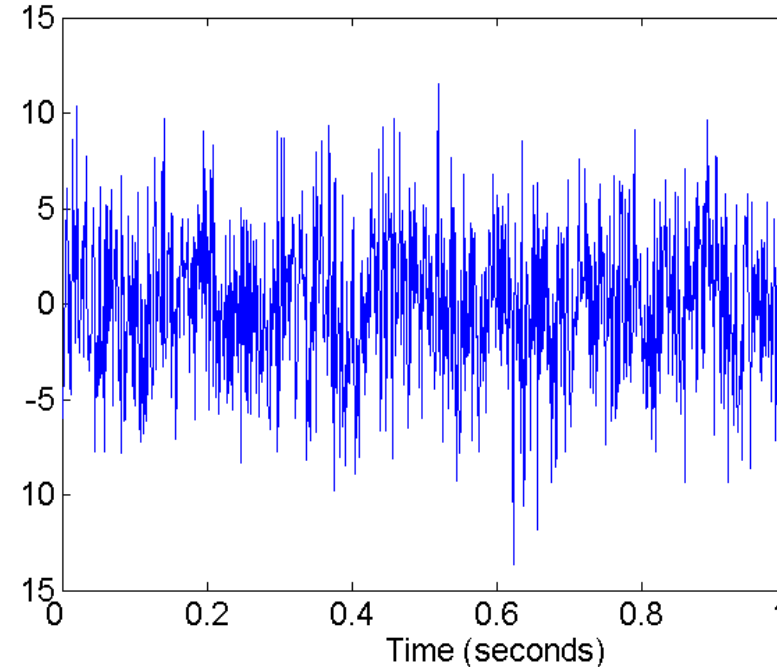
- 1. Eksik verileri gözardı et.
- 2. Elle doldur.
- 3. Global bir değişken kullan.
- 4. Merkezi bir ölçü kullanarak değiştir.
- 5. Aynı sınıfa ait merkezi bir ölçü ile değiştir.
- 6. Olası bir değer ile değiştir

# Gürültülü Veri

- Uç verilerin veri setine girmesi. Örneğin ses kalitesi düşük olan bir telefonda alınan insan sesi frekansları.



**Two Sine Waves**

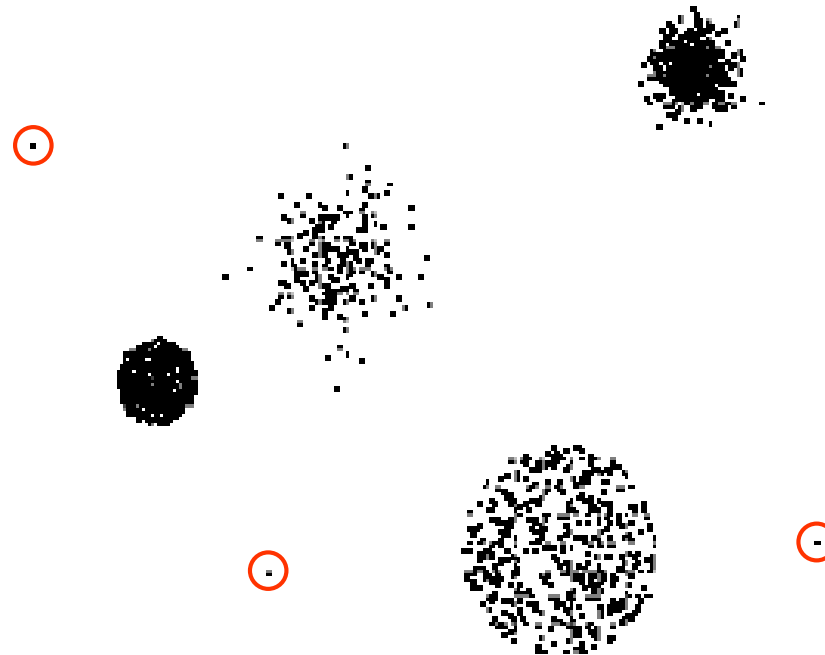


**Two Sine Waves + Noise**

# Aykırı (Uç) Veriler

---

- Veri setinin karakteristiğini yansıtmayan veriler.
- $1.5 \times \text{IQR}$



# Tekrarlı Veriler

---

- Veri setinde aynı olarak geçen birden fazla verinin olması durumu. Örneğin bir kişinin birden fazla mail adresi girmesi.
- Bu tip verilerin temizlenmesi gereklidir.

# Birleştirme

---

- 2 veya daha fazla özelliği bir araya getirme.
- Örneğin gün ve saat olarak verilen özelliklerin tek özellik olan tarih özelliğinde birleştirilmesi
- Ölçek değiştirilmesi
  - Şehirlerin bölgede birleştirilmesi
  - Günlerin ay olarak birleştirilmesi
- Daha stabil veri
  - Birleştirme sayesinde daha az değişken veriler elde edilir.

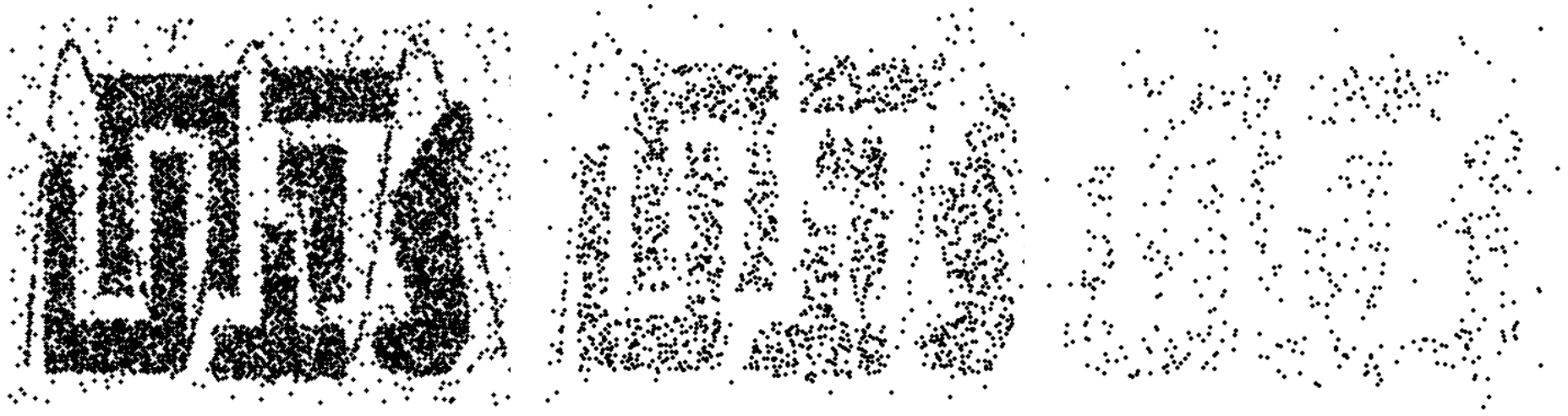
# Örnekleme Alma

---

- Veri azaltımı için gerekli olabilir.
  - Analiz öncesinde veya sonrasında yapılabilir.
- İstatistikçiler evrensel kümeye ulaşmak imkansız olduğundan örneklem üzerinden işlem yaparlar.
- Veri madenciler ise evrensel küme üzerinde işlem yapmak çok yorucu ve maliyetli olduğundan dolayı örneklem alırlar.

# Örnekleme Almaya Örnek

- Örneklemin evrensel kümeyi temsil etmesi gerekir. Yani evrensel küme içerisinde yer alan tüm özellikleri barındırması gerekir.





# Örneklem çeşitleri

---

1. Herhangi bir elemanın seçilmesinin eşit olasılıklı olması
2. Seçerek ayırma
  1. Rastgele seçilen verilerin örneklemden çıkarılması
3. Verileri küçük parçalara bölmek

# Özellik İndirgemesi

---

- Amaç
  - Fazla özelliklerden kurtulmak
  - Gereksiz zaman kaybının önüne geçmek
  - Verilerin daha iyi görselleştirilmesini sağlamak
  - Gereksiz özellik ve gürültülerin elenmesini sağlamak
- Teknikler
  - Birincil Etken Analizi (PCA=Principle Component Analysis)

# Özellik İndirgeme

---

- Gereksiz özellikler
  - Birbirine çok benzeyen iki verinin elenmesi
  - Örneğin araba satış fiyatı ile araba vergi tutarı
- Alakasız veriler
  - Veri madenciliği için kullanılmayacak olan özellikler.
  - Örneğin öğrenci numarası

# Kesikli Hale Getirme

---

- Sürekli verilerin kesikli veriler haline dönüştürülmesi işlemi
  - Sürekli veriler için bir takım kategoriler oluşturarak kesikli veriler kullanılabilir.
  - Özellikle sınıflandırma görevinde kullanılmalıdır.
  - Birçok sınıflandırma algoritması kesikli veriler ile daha iyi sonuçlar vermektedir.

# İkili Sisteme Çevirme

---

- Sürekli ya da kategorik değişkenin ikili(0-1) sistemine dönüştürülmesi işlemidir.
- Genellikle birliktelik kuralları için gerçekleştirilir.
- Sürekli veriler önce kategorik verilere çevrilir ardından ikili değerlere çevrilir.
- Örneğin uzunluğun kısa, orta, uzun diye kategorize edilmesi.

# Özellik Dönüşümleri

---

- Normalizasyon
  - Özellik verileri arasındaki farklılıkları azaltarak bir aralık içerisine alma işlemi
  - Sezonsal etki gibi istenmeyen özellikler ortadan kalkar.
- Standardizasyon
  - Verinin ortalamadan çıkarılarak standart sapmasına bölünmesi ile elde edilen sonuçlardır.

# Uygulamalı Veri Ön işleme

---

- Jupyter Notebook kullanarak öğrendiklerimizi uygulayalım.

# Kaynaklar

---

- Akküçük, Ulaş. “Veri madenciliği: kümeleme ve sınıflama algoritmaları”. *İstanbul: Yalın Yayıncılık* 18 (2011).
- Han, Jiawei, Jian Pei, ve Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- Kantardzic, Mehmed. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.
- Sumathi, Sai, ve S. N. Sivanandam. *Introduction to data mining and its applications*. C. 29. Springer, 2006.
- Tan, Pang-Ning, Michael Steinbach, ve Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.
- Towards Data Science. “Towards Data Science”. Erişim 29 Mart 2020. <https://towardsdatascience.com/>.
- VanderPlas, Jake. *Python Data Science Handbook*. O'Reilly Media. Inc, 2017.