

# PYTHON ile VERİ ANALİZİ

Hafta 06

# PYTHON ile VERİ ANALİZİ

- pandas, PYTHON Programlama Dili için geliştirilen açık kaynak kodlu bir veri analizi kütüphanesidir.

# PYTHON ile VERİ ANALİZİ

- pandas veri yapıları:
- Series
- DataFrame

# PYTHON ile VERİ ANALİZİ

- pandas veri yapıları:
- Panel veri yapısı ise pandas kütüphanesinin üçüncü veri yapısıdır.

# PYTHON ile VERİ ANALİZİ

- Series
- pandas veri yapılarından birisi olan Series, değişik veri türlerinden aynı veri türünü saklayabilen tek boyutlu dizilerdir.

# PYTHON ile VERİ ANALİZİ

- pandas'da her veri sütunu sadece bir tek veri türünü saklar.

# PYTHON ile VERİ ANALİZİ

- Python, finans, sinirbilim, ekonomi, istatistik, reklam ve web analitiği alanlarında akademik ve ticari olarak kullanılmaktadır.
- Veri bilimcilerin veri analizi için verilerin temsilinde en çok tercih edilen araçlardan biri haline gelmiştir.

# PYTHON ile VERİ ANALİZİ

- Bilişim alanında, birçok veri işleme aracı vardır.
- Bu araçlar arasında, ilişkisel veritabanları (MS SQL Server, Oracle), elektronik tablolar (Excel), olay(event) işleme sistemleri (Apache Spark gibi) ve R Programlama Dili ve pandas kütüphanesi gibi araçlar bulunmaktadır.



# PYTHON ile VERİ ANALİZİ

- **Veri Analizi**
- Veri analizi, veriden anlam çıkarma sürecidir.

# PYTHON ile VERİ ANALİZİ

- pandas veri yapıları:
- Series bir nesne (object)
- DataFrame de bir nesne(object)

# PYTHON ile VERİ ANALİZİ

- Görsel olarak, bir pandas DataFrame'in görünümü, indeks, satır ve sütunlardan oluşmaktadır.
- Üç bileşeni(component) vardır:
- DataFrame'in indeks, sütunlar ve veriler (değerler) olmak üzere 3 bileşeni vardır.

# PYTHON ile VERİ ANALİZİ

pandas öncelikle, `read_csv` fonksiyonunu kullanarak diskten belleğe(memory) ve DataFrame'e veri okur.

İndeks etiketi ve sütun adı, sırasıyla indeksin ve sütunların tek tek üyelerini ifade eder.

İndeks terimi, tüm indeks etiketlerini, sütun terimi ise tüm sütun adlarını gösterir.

# PYTHON ile VERİ ANALİZİ

- DataFrame yapısının Series'den, Series'in de liste(list) nesnesi ile oluşturulması:

# PYTHON ile VERİ ANALİZİ

- `import pandas as pd`
- `ürün = ['PC Bilgisayar', 'Yazıcı', 'Tablet', 'Laptop', 'Mouse']`
- `ürün__seri = pd.Series(ürün)`
- `marka = ['X', 'Y', 'Z', 'K', 'L']`
- `marka__seri = pd.Series(marka)`
- `fiyat = [10700, 3000, 5000, 18000, 20]`
- `fiyat__seri = pd.Series(fiyat)`

# PYTHON ile VERİ ANALİZİ

- `df__ürün= pd.DataFrame(ürün__seri)`
- `df__ürün = df__ürün.rename(columns = {o:'Ürün'})`
- `df__marka = pd.DataFrame(marka__seri)`
- `df__marka = df__marka.rename(columns = {o:'Marka'})`
- `df__fiyat = pd.DataFrame(fiyat__seri)`
- `df__fiyat = df__fiyat.rename(columns = {o:'Fiyat'})`

# PYTHON ile VERİ ANALİZİ

- `df__malzeme = pd.concat([df__ürün, df__marka, df__fiyat ], axis=1)`
- `print(df__malzeme)`
- `print("\n")`
- `print(type(df__malzeme))`
- `# concat fonksiyonu bir eksenini dikkate alarak DataFrame'leri birleştirir.`
- `# axis=1 sütunları göstermektedir`
- `# axis= 0 satırları gösterir`



# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:

- |   | Ürün          | Marka | Fiyat |
|---|---------------|-------|-------|
| 0 | PC Bilgisayar | A     | 10700 |
| 1 | Yazıcı        | B     | 3000  |
| 2 | Tablet        | C     | 5000  |
| 3 | Laptop        | D     | 18000 |
| 4 | Mouse         | E     | 20    |

- `<class 'pandas.core.frame.DataFrame'>`

# PYTHON ile VERİ ANALİZİ

- `print(marka__seri)`
- `print("\n")`
- `print(type(marka__seri))`

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
  - 0 X
  - 1 Y
  - 2 Z
  - 3 K
  - 4 L
  - dtype: object
- 
- <class 'pandas.core.series.**Series**'>

# PYTHON ile VERİ ANALİZİ

- #Series'nin DataFrame haline dönüşmesi
- `df__ürün= pd.DataFrame(ürün__seri)`
- `df__ürün = df__ürün.rename(columns = {o:'Ürün'})`
- `print (df__ürün )`
- # {o:'Ürün'} sütun sıfır ve başlık Ürün

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- Ürün
- 0   PC Bilgisayar
- 1       Yazıcı
- 2       Tablet
- 3       Laptop
- 4       Mouse

# PYTHON ile VERİ ANALİZİ

- # DataFrame'nin csv dosyası olarak çalışılan klasörde kaydedilmesi
- `df__malzeme.to_csv('Satış.csv')`

# PYTHON ile VERİ ANALİZİ

	A	B	C
1	,ÄœerÄ¼n,Marka,Fiyat		
2	0,PC Bilgisayar,X,10700		
3	1,YazÄ±cÄ±,Y,3000		
4	2,Tablet,Z,5000		
5	3,Laptop,K,18000		
6	4,Mouse,L,20		
7			

# PYTHON ile VERİ ANALİZİ

The diagram illustrates a Pandas DataFrame with the following structure and annotations:

- columns** (axis=1): Points to the header row.
- column name**: Points to the `director_name` column header.
- more columns to display**: Points to the ellipsis (...) in the header row.
- index label**: Points to the index values (0, 1, 2, 3, 4).
- index** (axis=0): Points to the index values.
- missing values**: Points to the `NaN` values in the `director_name` and `num_critic_for_reviews` columns for index 4.
- data (values)**: Points to the data values in the rows.

	color	director_name	num_critic_for_reviews	duration	...	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes
0	Color	James Cameron	723.0	178.0	...	936.0	7.9	1.78	33000
1	Color	Gore Verbinski	302.0	169.0	...	5000.0	7.1	2.35	0
2	Color	Sam Mendes	602.0	148.0	...	393.0	6.8	2.35	85000
3	Color	Christopher Nolan	813.0	164.0	...	23000.0	8.5	2.35	164000
4	NaN	Doug Walker	NaN	NaN	...	12.0	7.1	NaN	0



# PYTHON ile VERİ ANALİZİ

- NaN : Not a Number
- Missing Values: Eksik Veriler

# PYTHON ile VERİ ANALİZİ

- `# TTKOM Geçmiş Verileri__3__yıllık`
- `import pandas as pd`
- `import numpy as np`
- `# csv dosyasından verilerin yüklenmesi`
- `veri =`  
`pd.read_csv("C:/Users/cedim/Desktop/NLP/TTKOM__Geç`  
`miş__Verileri__3__yıllık.csv")`
- `#veri seti başlangıcından 2 adet veri okuma`
- `print("\n")`
- `print (veri.head(2))`

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- "Tarih","Şimdi","Açılış","Yüksek","Düşük","Hac.,""Fark %"
- 0 02.03.2023,"17,70","18,04","18,31","17,65","52...
- 1 01.03.2023,"18,03","18,16","18,28","18,00","43...

# PYTHON ile VERİ ANALİZİ

- `import os`
- `# Klasördeki dosyalar os.listdir() methodu ile listelenir`
- `for sayaç in os.listdir("C:/Users/cedim/Desktop/NLP/"):`
- `print(sayaç)`

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- 1-Intro NLP.pptx
- AI\_\_tools.txt
- amazon\_\_Alexa.docx
- analyzing-and-comparing-deep-learning-models.docx
- BTK YapayZeka\_\_Atölyesi.txt
- CNN\_\_Best.txt
- CNN\_\_G.docx
- CNN\_\_step\_\_by\_\_step.docx
- CNN\_\_step\_\_by\_\_step\_\_good\_\_analytics.docx

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- transformer.txt
- transformers.docx
- transformers\_\_vision\_\_mastery.docx
- transformer\_\_deep\_\_quara.docx
- transformer\_\_g.docx
- TTKOM\_\_Geçmiş\_\_Verileri\_\_3\_\_yıllık.csv
- wordCloud.txt
- yapay\_\_zeka\_\_ankara\_\_üniversitesi.docx
- YOLO\_\_Ödev\_\_Link.txt

# PYTHON ile VERİ ANALİZİ

- # TTKOM Geçmiş Verileri\_\_3\_\_yıllık
- 
- import pandas as pd
- import numpy as np
- # csv dosyasından verilerin yüklenmesi
- veri =  
pd.read\_csv("C:/Users/cedim/Desktop/NLP/TTKOM\_\_Geçmiş  
\_\_Verileri\_\_3\_\_yıllık.csv")
- #veri setinin listelenmesi
- print("\n")
- print(veri.to\_string())

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:

- "Tarih", "Şimdi", "Açılış", "Yüksek", "Düşük", "Hac.", "Fark %"
- 0 02.03.2023, "17,70", "18,04", "18,31", "17,65", "52,37M", "-1,83%"
- 1 01.03.2023, "18,03", "18,16", "18,28", "18,00", "43,21M", "0,17%"
- 2 28.02.2023, "18,00", "17,82", "18,23", "17,81", "62,40M", "1,52%"
- 3 27.02.2023, "17,73", "17,64", "17,87", "17,13", "63,54M", "0,62%"
- 4 24.02.2023, "17,62", "17,90", "18,02", "17,62", "40,41M", "-1,23%"
- 5 23.02.2023, "17,84", "18,10", "18,40", "17,74", "63,75M", "-4,55%"
- 6 22.02.2023, "18,69", "18,31", "19,13", "18,31", "62,26M", "2,30%"
- 7 21.02.2023, "18,27", "18,44", "19,24", "18,27", "59,60M", "-2,30%"
- 8 20.02.2023, "18,70", "19,50", "19,50", "18,37", "51,24M", "-2,91%"
- 9 17.02.2023, "19,26", "20,20", "20,40", "19,24", "41,32M", "-6,69%"



# PYTHON ile VERİ ANALİZİ

- # TTKOM Geçmiş Verileri\_\_3\_\_yıllık
- 
- import pandas as pd
- import numpy as np
- # csv dosyasından verilerin yüklenmesi
- veri =  
pd.read\_csv("C:/Users/cedim/Desktop/NLP/TTKOM\_\_Geçmiş  
\_\_Verileri\_\_3\_\_yıllık.csv")
- #veri setinin listelenmesi
- print("\n")
- print(veri)

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- "Tarih","Şimdi","Açılış","Yüksek","Düşük","Hac.,""Fark %"
- 0 02.03.2023,"17,70","18,04","18,31","17,65","52...
- 1 01.03.2023,"18,03","18,16","18,28","18,00","43...
- 2 28.02.2023,"18,00","17,82","18,23","17,81","62...
- 3 27.02.2023,"17,73","17,64","17,87","17,13","63...
- 4 24.02.2023,"17,62","17,90","18,02","17,62","40...
- .. ...
- 745 09.03.2020,"6,35","6,86","6,86","6,21","16,40M...
- 746 06.03.2020,"7,06","7,30","7,38","7,01","11,46M...

# PYTHON ile VERİ ANALİZİ

- # veri sayısı 60 adedi aşarsa **ilk ve son 5 veri** listelenir
- `import pandas as pd`
- `print(pd.options.display.max_rows)`

# PYTHON ile VERİ ANALİZİ

- Ekran Çıktısı:
- 60

# PYTHON ile VERİ ANALİZİ

- `read.csv()` fonksiyonu ile, csv uzantılı dosyalardaki verilerden bir DataFrame oluşturur.

# PYTHON ile VERİ ANALİZİ

- DataFrame'deki bir sütun, bir Seri olarak seçilebilir.
- Bu şekilde indeks ve veri listelenebilir.

# PYTHON ile VERİ ANALİZİ

- `import pandas as pd`
- `import numpy as np`
- `# csv dosyasından verilerin yüklenmesi, movies.csv`
- `veri =  
pd.read_csv('https://raw.githubusercontent.com/Netflix/metaflow/master/R/inst/tutorials/o2-statistics/movies.csv')`
- `print(veri)`

# PYTHON ile VERİ ANALİZİ

•	movie_title	title_year	genre \
• 0	Avatar	2009	Action
• 1	Pirates of the Caribbean: At World's End	2007	Fantasy
• 2	Spectre	2015	Thriller
• 3	The Dark Knight Rises	2012	Thriller
• 4	John Carter	2012	Action
• ...	...	...	...
• 3745	Clean	2004	Romance
• 3746	The Circle	2000	Drama
• 3747	Primer	2004	Thriller
• 3748	El Mariachi	1992	Romance
• 3749	My Date with Drew	2004	Documentary



# PYTHON ile VERİ ANALİZİ

- gross
- 0 760505847
- 1 309404152
- 2 200074175
- 3 448130642
- 4 73058679
- ... ..
- 3745 136007
- 3746 673780
- 3747 424760
- 3748 2040920
- 3749 85222
- [3750 rows x 4 columns]

# PYTHON ile VERİ ANALİZİ

- `movie_title, title_year, genre, gross`
- Avatar,2009,Action,760505847
- Pirates of the Caribbean: At World's End,2007,Fantasy,309404152
- Spectre,2015,Thriller,200074175
- The Dark Knight Rises,2012,Thriller,448130642
- John Carter,2012,Action,73058679
- Spider-Man 3,2007,Romance,336530303
- Tangled,2010,Romance,200807262
- Avengers: Age of Ultron,2015,Action,458991599

# PYTHON İLE VERİ ANALİZİ

- # sütun veri türleri
- `print(veri.dtypes)`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- movie\_title    object
- title\_year    int64
- genre          object
- gross          int64
- dtype: object

# PYTHON İLE VERİ ANALİZİ

- # dataset inceleme
- `print(veri.info())`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- <class 'pandas.core.frame.DataFrame'>
- RangeIndex: 3750 entries, 0 to 3749
- Data columns (total 4 columns):
- # Column Non-Null Count Dtype
- ---
- 0 movie\_title 3750 non-null object
- 1 title\_year 3750 non-null int64
- 2 genre 3750 non-null object
- 3 gross 3750 non-null int64
- dtypes: int64(2), object(2)
- memory usage: 117.3+ KB
- None

# PYTHON İLE VERİ ANALİZİ

- # Çıktının dataFrame olup olmadığını kontrolü
- `print (type(veri))`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- <class 'pandas.core.frame.DataFrame'>



# PYTHON İLE VERİ ANALİZİ

- `# toplam veri sayısı`
- `print('Toplam Veri Sayısı = ', veri.size)`
- `print('Boyut = ', veri.ndim) # boyut sayısı`
- `print('Satır ve Sütun Sayıları = ', veri.shape)`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- Toplam Veri Sayısı = 15000
- Boyut = 2
- Satır ve Sütun Sayıları= (3750, 4)

# PYTHON İLE VERİ ANALİZİ

- `print (veri['movie_title'])`

# PYTHON İLE VERİ ANALİZİ

- 0 Avatar
- 1 Pirates of the Caribbean: At World's End
- 2 Spectre
- 3 The Dark Knight Rises
- 4 John Carter
- ...
- 3745 Clean
- 3746 The Circle
- 3747 Primer
- 3748 El Mariachi
- 3749 My Date with Drew
- Name: movie\_title, Length: 3750, dtype: object

# PYTHON İLE VERİ ANALİZİ

- `print (veri['movie_title'], "\n", veri['gross'] )`

# PYTHON İLE VERİ ANALİZİ

- 0 Avatar
- 1 Pirates of the Caribbean: At World's End
- 2 Spectre
- 3 The Dark Knight Rises
- 4 John Carter
- ...
- 3745 Clean
- 3746 The Circle
- 3747 Primer
- 3748 El Mariachi
- 3749 My Date with Drew
- Name: movie\_title, Length: 3750, dtype: object

# PYTHON İLE VERİ ANALİZİ

- 0      760505847
- 1      309404152
- 2      200074175
- 3      448130642
- 4      73058679
- ...
- 3745    136007
- 3746    673780
- 3747    424760
- 3748    2040920
- 3749    85222
- Name: gross, Length: 3750, dtype: int64

# PYTHON İLE VERİ ANALİZİ

- #Sütunun Series olup olmadığının kontrolü
- `print (type(veri['gross']))`



# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- `<class 'pandas.core.series.Series'>`

# PYTHON PROGRAMLAMA

- # Çıktının seri olup olmadığını kontrolü
- `print (type(veri['movie_title']))`
- Ekran Çıktısı:
- `<class 'pandas.core.series.Series'>`

# PYTHON İLE VERİ ANALİZİ

- #data set 'den iki farklı veri türünde iki farklı seri oluşturma
- `film = veri['movie_title']`
- `yıl = veri['title_year']`

# PYTHON İLE VERİ ANALİZİ

- `print (film)`
- `print("\n")`
- `print (yıl)`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:

- 0 Avatar
- 1 Pirates of the Caribbean: At World's End
- 2 Spectre
- 3 The Dark Knight Rises
- 4 John Carter
- ...
- 3745 Clean
- 3746 The Circle
- 3747 Primer
- 3748 El Mariachi
- 3749 My Date with Drew
- Name: movie\_title, Length: 3750, dtype: object

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:

- 0 2009
- 1 2007
- 2 2015
- 3 2012
- 4 2012
- ...
- 3745 2004
- 3746 2000
- 3747 2004
- 3748 1992
- 3749 2004
- Name: title\_year, Length: 3750, dtype: int64

# PYTHON İLE VERİ ANALİZİ

- # bütün name space import edilebilir
- from math import \*
- factorial(5)
- Ekran Çıktısı:
- 120

# PYTHON İLE VERİ ANALİZİ

- # kütüphane **m** alias olarak import edildi
- `import math as m`
- `m.factorial(5)`
- Ekran Çıktısı:
- 120



# PYTHON İLE VERİ ANALİZİ

- # isin() metodu kullanımı
- import pandas as pd
- data = {
  - "isim": ["Meltem", "Sema", "Kaya"],
  - "yaş": [30, 40, 50],
  - "Şehir": ["Erzurum", "Uşak", "Kayseri"],
  - }
- df = pd.DataFrame(data)
- print(df.isin([50, 40]))

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- isim yaş Şehir
- 0 False False False
- 1 False True False
- 2 False True False

# PYTHON İLE VERİ ANALİZİ

- # Sütun adları
- df.columns

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- `Index(['isim', 'yaş', 'Şehir'], dtype='object')`

# PYTHON İLE VERİ ANALİZİ

- # Dataframe indeksi
- `df.index`

# PYTHON İLE VERİ ANALİZİ

- Ekran Çıktısı:
- `RangeIndex(start=0, stop=3, step=1)`

# PYTHON PROGRAMLAMA

## Kaynaklar:

Python Data Analysis, Armando Fandango, Packt Publishing, 2017.

Learning pandas, Michael Heydt, Packt Publishing, 2017

Pandas Cookbook, Theodore Petrou, Packt Publishing, 2017.

[https://www.w3schools.com/python/pandas/pandas\\_json.asp](https://www.w3schools.com/python/pandas/pandas_json.asp)

<https://tr.investing.com/equities/turk-telekom-historical-data>