PYTHON PROGRAMLAMAYA GİRİŞ

Hafta 11

PYTHON PROGRAMLAMA

- pandas kapsamında kullanılan, Pivot tablo, dataframe ile kolayca işlem yapılmasını sağlayan bir tool'dur.

PYTHON PROGRAMLAMA

- pandas, Python programlama dilinde kullanılan açık kaynak kodlu bir veri analizi kütüphanesidir.

PYTHON PROGRAMLAMA

- pandas kütüphanesinde bulunan temel veri yapıları:
- Series
- DataFrame

PYTHON PROGRAMLAMA

- Pivot tablo fonksiyonu pandas kütüphanesinin bir fonksiyonudur.

PYTHON PROGRAMLAMA

- Pandas veri yapısı olan dataframe; satır, sütun ve indeks'ten meydana gelir.
- Dataframe'in her bir sütunu sadece bir tek veri türünü saklamaktadır.

PYTHON PROGRAMLAMA

- Pivot tablo için en az bir index ve bir dataframe kullanılmaktadır.

PYTHON PROGRAMLAMA

- pandas pivot_table fonksiyonu veri analizi için kullanılır.

PYTHON PROGRAMLAMA

- veri seti için link:
- https://raw.githubusercontent.com/SciRuby/daru-io/master/spec/fixtures/csv/sales-funnel.csv

PYTHON PROGRAMLAMA

- veri seti görünümü:
- 714466,Trantow-Barrows,Craig Booker,Debra Henley,CPU,1,30000,presented
- 714466,Trantow-Barrows,Craig Booker,Debra Henley,Software,1,10000,presented
- 714466,Trantow-Barrows,Craig Booker,Debra Henley,Maintenance,2,5000,pending
- 737550,"Fritsch, Russel and Anderson",Craig Booker,Debra Henley,CPU,1,35000,declined
- 146832,Kiehn-Spinka,Daniel Hilton,Debra Henley,CPU,2,65000,won
- 218895,Kulas Inc,Daniel Hilton,Debra Henley,CPU,2,40000,pending
- 218895,Kulas Inc,Daniel Hilton,Debra Henley,Software,1,10000,presented
- 412290,Jerde-Hilpert,John Smith,Debra Henley,Maintenance,2,5000,pending
- 740150,Barton LLC,John Smith,Debra Henley,CPU,1,35000,declined
- 141962,Herman LLC,Cedric Moss,Fred Anderson,CPU,2,65000,won
- 163416,Purdy-Kunde,Cedric Moss,Fred Anderson,CPU,1,30000,presented
- 239344,Stokes LLC,Cedric Moss,Fred Anderson,Maintenance,1,5000,pending
- 239344,Stokes LLC,Cedric Moss,Fred Anderson,Software,1,10000,presented
- 307599,"Kassulke, Ondricka and Metz",Wendy Yule,Fred Anderson,Maintenance,3,7000,won
- 688981,Keeling LLC,Wendy Yule,Fred Anderson,CPU,5,100000,won
- 729833,Koepp Ltd,Wendy Yule,Fred Anderson,CPU,2,65000,declined
- 729833,Koepp Ltd,Wendy Yule,Fred Anderson,Monitor,2,5000,presented

PYTHON PROGRAMLAMA

- import pandas as pd
- import numpy as np

PYTHON PROGRAMLAMA

- # datasetin dataFrame'e alınması:
- df = pd.read_csv("https://raw.githubusercontent.com/SciRuby/daru-io/master/spec/fixtures/csv/sales-funnel.csv")

PYTHON PROGRAMLAMA

- # dataFrame'deki ilk 5 veri:

- df.head()

PYTHON PROGRAMLAMA

- Account Name Rep Manager Product Quantity Price Status
- 0 714466 Trantow-Barrows Craig Booker Debra Henley CPU 1 30000 presented
- 1 714466 Trantow-Barrows Craig Booker Debra Henley Software 1 10000 presented
- 2 714466 Trantow-Barrows Craig Booker Debra Henley Maintenance 2 5000 pending
- 3 737550 Fritsch, Russel and Anderson Craig Booker Debra Henley CPU 1 35000 declined
- 4 146832 Kiehn-Spinka Daniel Hilton Debra Henley CPU 2 65000 won

PYTHON PROGRAMLAMA

- # bütün dataFrame'in listesi
- print(df)

PYTHON PROGRAMLAMA

- Account Name Rep Manager \
- 0 714466 Trantow-Barrows Craig Booker Debra Henley
- 1 714466 Trantow-Barrows Craig Booker Debra Henley
- 2 714466 Trantow-Barrows Craig Booker Debra Henley
- 3 737550 Fritsch, Russel and Anderson Craig Booker Debra Henley
- 4 146832 Kiehn-Spinka Daniel Hilton Debra Henley
- 5 218895 Kulas Inc Daniel Hilton Debra Henley
- 6 218895 Kulas Inc Daniel Hilton Debra Henley
- 7 412290 Jerde-Hilpert John Smith Debra Henley
- 8 740150 Barton LLC John Smith Debra Henley
- 9 141962 Herman LLC Cedric Moss Fred Anderson
- 10 163416 Purdy-Kunde Cedric Moss Fred Anderson
- 11 239344 Stokes LLC Cedric Moss Fred Anderson
- 12 239344 Stokes LLC Cedric Moss Fred Anderson
- 13 307599 Kassulke, Ondricka and Metz Wendy Yule Fred Anderson
- 14 688981 Keeling LLC Wendy Yule Fred Anderson
- 15 729833 Koepp Ltd Wendy Yule Fred Anderson
- 16 729833 Koepp Ltd Wendy Yule Fred Anderson

PYTHON PROGRAMLAMA

- # Pivot tablonun için bir dataFrame ve bir indeks gerekli olmaktadır.
- Bu örnekte, Name indeks olarak olarak belirlenmiştir.

- pd.pivot_table(df,index=["Name"])

PYTHON PROGRAMLAMA

	Account	Price	Quantity
Name			
Barton LLC	740150	35000	1.000000
Fritsch, Russel and Anderson	737550	35000	1.000000
Herman LLC	141962	65000	2.000000
Jerde-Hilpert	412290	5000	2.000000
Kassulke, Ondricka and Metz	307599	7000	3.000000
Keeling LLC	688981	100000	5.000000
Kiehn-Spinka	146832	65000	2.000000
Koepp Ltd	729833	35000	2.000000
Kulas Inc	218895	25000	1.500000
Purdy-Kunde	163416	30000	1.000000
Stokes LLC	239344	7500	1.000000
Trantow-Barrows	714466	15000	1.333333

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Name","Rep","Manager"])

PYTHON PROGRAMLAMA

Account Name	Rep	Manager	Price	Quantity	
Barton LLC	John Smith	Debra Henley	740150	35000	1.000000
Fritsch, Russel and Anderson	Craig Booker	Debra Henley	737550	35000	1.000000
Herman LLC	Cedric Moss	Fred Anderson	141962	65000	2.000000
Jerde-Hilpert	John Smith	Debra Henley	412290	5000	2.000000
Kassulke, Ondricka and Metz	Wendy Yule	Fred Anderson	307599	7000	3.000000
Keeling LLC	Wendy Yule	Fred Anderson	688981	100000	5.000000
Kiehn-Spinka	Daniel Hilton	Debra Henley	146832	65000	2.000000
Koepp Ltd	Wendy Yule	Fred Anderson	729833	35000	2.000000
Kulas Inc	Daniel Hilton	Debra Henley	218895	25000	1.500000
Purdy-Kunde	Cedric Moss	Fred Anderson	163416	30000	1.000000
Stokes LLC	Cedric Moss	Fred Anderson	239344	7500	1.000000
Trantow-Barrows	Craig Booker	Debra Henley	714466	15000	1.333333

PYTHON PROGRAMLAMA

- pd.pivot_table(df, index=["Manager","Rep"])

PYTHON PROGRAMLAMA

Account Manager	Price Rep	Quantity		
Debra Henley	Craig Booker	720237.0	20000.000000	1.250000
	Daniel Hilton	194874.0	38333.333333	1.666667
	John Smith	576220.0	20000.000000	1.500000
Fred Anderson	Cedric Moss	196016.5	27500.000000	1.250000
	Wendy Yule	614061.5	44250.000000	3.000000

PYTHON PROGRAMLAMA

- pd.pivot_table(df, index=["Manager","Rep"],values=["Price"])

PYTHON PROGRAMLAMA

Manager	Rep	Price
Debra Henley	Craig Booker	20000.000000
	Daniel Hilton	38333.333333
	John Smith	20000.000000
Fred Anderson	Cedric Moss	27500.000000
	Wendy Yule	44250.000000

PYTHON PROGRAMLAMA

- # dataFrame'deki toplam veri sayısı
- df.size

- Ekran Çıktısı:
- 136

PYTHON PROGRAMLAMA

- # dataFrame'deki satır ve sütun sayısı
- df.shape

- Ekran Çıktısı
- (17, 8)

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep"],values=["Price"],aggfunc=np.sum)

PYTHON PROGRAMLAMA

Manager	Rep	Price
Debra Henley	Craig Booker	80000
	Daniel Hilton	115000
	John Smith	40000
Fred Anderson	Cedric Moss	110000
	Wendy Yule	177000

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep"],values=["Price"],aggfunc=[np.mean,len])

PYTHON PROGRAMLAMA

		mean Price	len Price
Manager	Rep		
Debra Henley	Craig Booker	20000.000000	4
	Daniel Hilton	38333.333333	3
	John Smith	20000.000000	2
Fred Anderson	Cedric Moss	27500.000000	4
	Wendy Yule	44250.000000	4

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep"],values=["Price"], columns=["Product"],aggfunc=[np.sum])

PYTHON PROGRAMLAMA

-

		sum Price			
	Product	CPU	Maintenance	Monitor	Software
Manager	Rep				
Debra Henley	Craig Booker	65000.0	5000.0	NaN	10000.0
	Daniel Hilton	105000.0	NaN	NaN	10000.0
	John Smith	35000.0	5000.0	NaN	NaN
Fred Anderson	Cedric Moss	95000.0	5000.0	NaN	10000.0
	Wendy Yule	165000.0	7000.0	5000.0	NaN

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep"],values=["Price"],
columns=["Product"],aggfunc=[np.sum],fill_value=0)

PYTHON PROGRAMLAMA

-

		sum price			
	Product	CPU	Maintenance	Monitor	Software
Manager	Rep				
Debra Henley	Craig Booker	65000	5000	0	10000
	Daniel Hilton	105000	0	0	10000
	John Smith	35000	5000	0	0
Fred Anderson	Cedric Moss	95000	5000	0	10000
	Wendy Yule	165000	7000	5000	0

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep"],values=["Price","Quantity"], columns=["Product"],aggfunc=[np.sum],fill_value=0)

PYTHON PROGRAMLAMA

-

		Sum							
		Price				Quantity			
	Product	CPU	Maintenance	Monitor	Software	CPU	Maintenance	Monitor	Software
Manager	Rep								
Debra Henley	Craig Booker	65000	5000	0	10000	2	2	0	1
	Daniel Hilton	105000	0	0	10000	4	0	0	1
	John Smith	35000	5000	0	0	1	2	0	0
Fred Anderson	Cedric Moss	95000	5000	0	10000	3	1	0	1
	Wendy Yule	165000	7000	5000	0	7	3	2	0

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep","Product"],
- values=["Price","Quantity"],aggfunc=[np.sum],fill_value=0)

Manager	Rep	Product	sum Price	Quantity
Debra Henley	Craig Booker	CPU	65000	2
		Maintenance	5000	2
		Software	10000	1
	Daniel Hilton	CPU	105000	4
		Software	10000	1
	John Smith	CPU	35000	1
		Maintenance	5000	2
Fred Anderson	Cedric Moss	CPU	95000	3
		Maintenance	5000	1
		Software	10000	1
	Wendy Yule	CPU	165000	7
		Maintenance	7000	3
		Monitor	5000	2

PYTHON PROGRAMLAMA

```
pd.pivot_table(df,index=["Manager","Rep","Product"],
               values=["Price","Quantity"],
               aggfunc=[np.sum,np.mean],fill_value=0)
```

PYTHON PROGRAMLAMA

			sum		mean	
			Price	Quantity	Price	Quantity
Manager	Rep	Product				
Debra Henley	Craig Booker	CPU	65000	2	32500	1.0
		Maintenance	5000	2	5000	2.0
		Software	10000	1	10000	1.0
	Daniel Hilton	CPU	105000	4	52500	2.0
		Software	10000	1	10000	1.0
	John Smith	CPU	35000	1	35000	1.0
		Maintenance	5000	2	5000	2.0
Fred Anderson	Cedric Moss	CPU	95000	3	47500	1.5
		Maintenance	5000	1	5000	1.0
		Software	10000	1	10000	1.0
	Wendy Yule	CPU	165000	7	82500	3.5
		Maintenance	7000	3	7000	3.0
		Monitor	5000	2	5000	2.0

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Rep","Product"],
- values=["Price","Quantity"],
- aggfunc=[np.sum,np.mean],fill_value=0,margins=True
)

PYTHON PROGRAMLAMA

			sum	mean			
			Price	Quantity	Price	Quantity	
Manager	Rep	Product					
Debra Henley	Craig Booker	CPU	65000	2	32500	1.000000	
		Maintenance	5000	2	5000	2.000000	
		Software	10000	1	10000	1.000000	
	Daniel Hilton	CPU	105000	4	52500	2.000000	
		Software	10000	1	10000	1.000000	
	John Smith	CPU	35000	1	35000	1.000000	
		Maintenance	5000	2	5000	2.000000	
Fred Anderson	Cedric Moss	CPU	95000	3	47500	1.500000	
		Maintenance	5000	1	5000	1.000000	
		Software	10000	1	10000	1.000000	
	Wendy Yule	CPU	165000	7	82500	3.500000	
		Maintenance	7000	3	7000	3.000000	
		Monitor	5000	2	5000	2.000000	
All			522000	30	30705	1.764706	

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Status"],values=[
"Price"],
- aggfunc=[np.sum],fill_value=0,margins=True)

PYTHON PROGRAMLAMA

-

		sum
	price	
Manager	Status	
Debra Henley	won	65000
	pending	50000
	presented	50000
	declined	70000
Fred Anderson	won	172000
	pending	5000
	presented	45000
	declined	65000
All		522000

PYTHON PROGRAMLAMA

- pd.pivot_table(df,index=["Manager","Status"],columns=["Product"],values=["Quantity","Price"],
- aggfunc={"Quantity":len,"Price":np.sum},fill_value=0)

PYTHON PROGRAMLAMA

		Price			Quantity				
	Product	CPU	Maintenance	Monitor	Software	CPU	Maintenance	Monitor	Software
Manager	Status								
Debra Henley	won	65000	0	0	0	1	0	0	0
	pending	40000	10000	0	0	1	2	0	0
	presented	30000	0	0	20000	1	0	0	2
	declined	70000	0	0	0	2	0	0	0
Fred Anderson	won	165000	7000	0	0	2	1	0	0
	pending	0	5000	0	0	0	1	0	0
	presented	30000	0	5000	10000	1	0	1	1
	declined	65000	0	0	0	1	0	0	0

PYTHON PROGRAMLAMA

- table = pd.pivot_table(df,index=["Manager","Status"],columns=["Product"],values=["Quantity","Price"], aggfunc={"Quantity":len,"Price":[np.sum,np.mean]},fill_value=0)
- table

PYTHON PROGRAMLAMA

- table.query('Manager == ["Debra Henley"]')

PYTHON PROGRAMLAMA

Price		mean				sum				Quantity len			
	Product	CPU	Maintenance	Monitor	Software	CPU	Maintenance	Monitor	Software	CPU	Maintenance	Monitor	Softwa
Manager	Status												
Debra Henley	won	65000	0	0	0	65000	0	0	0	1	0	0	
	pending	40000	5000	0	0	40000	10000	0	0	1	2	0	
	presented	30000	0	0	10000	30000	0	0	20000	1	0	0	
	declined	35000	0	0	0	70000	0	0	0	2	0	0	

PYTHON PROGRAMLAMA

- table.query('Status == ["pending","won"]')

PYTHON PROGRAMLAMA

ice		Quantity									
		mean				sum				len	
	Product	CPU	Maintenance	Monitor	Software	CPU	Maintenance	Monitor	Software	CPU	Maintenance
anager	Status										
Debra Henley	pending	40000	5000	0	0	40000	10000	0	0	1	2
	won	65000	0	0	0	65000	0	0	0	1	0
Fred nderson	pending	0	5000	0	0	0	5000	0	0	0	1
	won	82500	7000	0	0	165000	7000	0	0	2	1

PYTHON PROGRAMLAMA

- Dataset ve Python kodları, Chris Moffitt'in 'Pandas Pivot Table Explained' adlı yazısından 'https://pbpython.com/pandas-pivot-table-explained.html' linkinden alınmıştır.

PYTHON PROGRAMLAMA

Kaynaklar:

https://pbpython.com/pandas-pivot-table-explained.html

https://cmdlinetips.com/2018/12/pivot-table-in-python-pandas/

https://jakevdp.github.io/PythonDataScienceHandbook/03.09-pivot-tables.html

https://datatofish.com/pivot-table-python/

http://www.gregreda.com/2013/10/26/intro-to-pandas-data-structures/

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.describe.html