

```
1 https://s6.dosya.tc/server16/kwwh3b/insurance_.csv.html
```

```
In [1]: 1 import pandas as pd
        2 import numpy as np
```

```
In [2]: 1 # csv formatlı datayı dataframe çevirmek
        2 df = pd.read_csv('insurance_.csv')
```

```
In [3]: 1 #İlk beş satır data
        2 df
```

Out[3]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
1336	50	male	30.970	3	no	northwest	10600.54830
1337	18	female	31.920	0	no	northeast	2205.98080
1338	18	female	36.850	0	no	southeast	1629.83350
1339	21	female	25.800	0	no	southwest	2007.94500
1340	61	female	29.070	0	yes	northwest	29141.36030

1341 rows × 7 columns

```
In [4]: 1 #Datanın sütunları
        2 df.columns
```

```
Out[4]: Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'], dtype='object')
```

```
In [5]: 1 #Datanın sütunlarının isimlerini değiştirebiliriz
        2 df.columns = ['yaş', 'cinsiyet', 'body_maas_index', 'çocuk_sayısı', 'sigara_içiyor', 'bölge', 'sigorta_pirim_tutarı']
```

```
In [6]: 1 df
```

```
Out[6]:
```

	yaş	cinsiyet	body_maas_index	çocuk_sayısı	sigara_içiyor	bölge	sigorta_pirim_tutarı
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
1336	50	male	30.970	3	no	northwest	10600.54830
1337	18	female	31.920	0	no	northeast	2205.98080
1338	18	female	36.850	0	no	southeast	1629.83350
1339	21	female	25.800	0	no	southwest	2007.94500
1340	61	female	29.070	0	yes	northwest	29141.36030

1341 rows × 7 columns

In [7]:

```
1 #Data tip bilgisi
2 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1341 entries, 0 to 1340
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   yaş                   1341 non-null   int64
1   cinsiyet              1341 non-null   object
2   body_maas_index       1340 non-null   float64
3   çocuk_sayısı         1341 non-null   int64
4   sigara_içiyor         1341 non-null   object
5   bölge                 1340 non-null   object
6   sigorta_pirim_tutarı  1341 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.5+ KB
```

In [8]:

```
1 #Datenin boyutları
2 df.shape
```

Out[8]: (1341, 7)

In [9]:

```
1 sayısal_sütunler = ['yaş', 'body_mass_index', 'çocuk_sayısı', 'sigorta_pirim_tutarı']
2 sayısal_olmayan_sütunlar = ['cinsiyet', 'sigara_içiyor', 'bölge']
```

```
In [10]: 1 #Integer ve float tip datanın ayırt edici istatistiksel özeti
        2 df.describe()
```

Out[10]:

	yaş	body_maas_index	çocuk_sayısı	sigorta_pirim_tutarı
count	1341.000000	1340.000000	1341.000000	1341.000000
mean	39.241611	30.669019	1.093960	13264.401302
std	14.055373	6.097738	1.204519	12097.383457
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.308750	0.000000	4746.344000
50%	39.000000	30.400000	1.000000	9377.904700
75%	51.000000	34.700000	2.000000	16586.497710
max	64.000000	53.130000	5.000000	63770.428010

```
In [11]: 1 #Object ve boolean tip datanın ayırt edici istatistiksel bilgisi
        2 df.describe(include=["object", "bool"])
```

Out[11]:

	cinsiyet	sigara_içiyor	bölge
count	1341	1341	1340
unique	2	2	4
top	male	no	southeast
freq	679	1067	366

```
In [12]: 1 #sayısal olmayan datanın içerdiği değerler
        2 for i in sayısal_olmayan_sütunlar:
        3     print(df[i].unique())
```

```
['female' 'male']
```

```
['yes' 'no']
```

```
['southwest' 'southeast' 'northwest' 'northeast' nan]
```

```
In [13]: 1 #sayısal olmayan datanın içerdiği değerlerin yüzdesi
          2 for sütun in sayısal_olmayan_sütunlar:
          3     print(df[sütun].value_counts(normalize=True))
```

```
male      0.506339
female    0.493661
Name: cinsiyet, dtype: float64
no        0.795675
yes       0.204325
Name: sigara_içiyor, dtype: float64
southeast 0.273134
southwest 0.242537
northwest 0.242537
northeast 0.241791
Name: bölge, dtype: float64
```

```
In [14]: 1 #missing value olup olmadığını kontrol ediyoruz
          2 df.isnull().sum()
```

```
Out[14]: yaş      0
cinsiyet      0
body_maas_index  1
çocuk_sayısı  0
sigara_içiyor  0
bölge         1
sigorta_pirim_tutarı  0
dtype: int64
```

```
In [15]: 1 #missing valueları düşürüyoruz
        2 df.dropna(subset=['body_maas_index','bölge'],inplace=True)
        3 df.isnull().sum()
```

```
Out[15]: yaş                0
        cinsiyet            0
        body_maas_index     0
        çocuk_sayısı       0
        sigara_içiyor       0
        bölge               0
        sigorta_pirim_tutarı 0
        dtype: int64
```

```
In [16]: 1 #duplicated datanın olup olmadığını kontrol ediyoruz
        2 df.duplicated().sum()
```

```
Out[16]: 2
```

```
In [17]: 1 #duplicated dataları siliyoruz
        2 df.drop_duplicates(inplace=True)
```

```
In [18]: 1 # eğer data silme işlemi yaparsak indeks serimiz bozulacal(1337 adet datamız var ama indeks serimiz 1340'da bitiyö
        2 df.index
```

```
Out[18]: Int64Index([ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
        ...
        1331, 1332, 1333, 1334, 1335, 1336, 1337, 1338, 1339, 1340],
        dtype='int64', length=1337)
```

```
In [19]: 1 # Datanın indeksini reset_index() ile düzeltebiliriz
        2 df.reset_index(inplace=True)
```

```
In [20]: 1 # eğer datanın indeksini tekrar kontrol edersek düzeldiğini göreceğiz
        2 df.index
```

```
Out[20]: RangeIndex(start=0, stop=1337, step=1)
```

```
In [21]: 1 #Datayı spesifik sütuna bağlı olarak sıralamak
        2 df.sort_values(by="sigorta_pirim_tutarı", ascending=False)
```

```
Out[21]:
```

	index	yaş	cinsiyet	body_maas_index	çocuk_sayısı	sigara_içiyor	bölge	sigorta_pirim_tutarı
543	543	54	female	47.410	0	yes	southeast	63770.42801
1299	1300	45	male	30.360	0	yes	southeast	62592.87309
1229	1230	52	male	34.485	3	yes	northwest	60021.39897
577	577	31	female	38.095	1	yes	northeast	58571.07448
818	819	33	female	35.530	0	yes	northwest	55135.40209
...
22	22	18	male	34.100	0	no	southeast	1137.01100
662	663	18	male	33.660	0	no	southeast	1136.39940
1243	1244	18	male	33.330	0	no	southeast	1135.94070
807	808	18	male	30.140	0	no	southeast	1131.50660
939	940	18	male	23.210	0	no	southeast	1121.87390

1337 rows × 8 columns

```
In [22]: 1 #ortalama sigorta primin bulmak istersek
        2 df['sigorta_pirim_tutarı'].mean()
```

```
Out[22]: 13279.121486655948
```

```
In [23]: 1 #sigara içenlerin ödediği ortalama sigorta primini görmek istersek
        2 df[df['sigara_içiyor']=='yes']['sigorta_pirim_tutarı'].mean()
```

Out[23]: 32050.23183153285

```
In [24]: 1 #eğer sigara içen ve kadın olanların ortalama sigorta primini görmek istersem
        2 df[(df['cinsiyet']=='female') & (df['sigara_içiyor']=='yes')]['sigorta_pirim_tutarı'].mean()
```

Out[24]: 30678.996276260867

```
In [25]: 1 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes') & (df['bölge']=='southeast')]['sigorta_pirim_tutarı'].i
```

Out[25]: 36029.83936654546

```
In [26]: 1 #eğer sigara içen ve kadın olanlardan en yüksek sigorta primini ödeyen kişi
        2 df[(df['cinsiyet']=='female') & (df['sigara_içiyor']=='yes')].max()
```

Out[26]:

index	1340
yaş	64
cinsiyet	female
body_maas_index	47.6
çocuk_sayısı	5
sigara_içiyor	yes
bölge	southwest
sigorta_pirim_tutarı	63770.42801
dtype:	object


```
In [27]: 1 #eğer sigara içen ve kadın olanlardan en düşük sigorta pirimini ödeyen kişi
          2 df[(df['cinsiyet']=='female') & (df['sigara_içiyor']=='yes')].min()
```

```
Out[27]: index                0
         yaş                  18
         cinsiyet            female
         body_maas_index      17.195
         çocuk_sayısı         0
         sigara_içiyor        yes
         bölge               northeast
         sigorta_pirim_tutarı  13844.506
         dtype: object
```

```
In [28]: 1 #eğer sigara içen ve erkek olanların ortalama sigorta pirimini görmek istersem
          2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes')]['sigorta_pirim_tutarı'].mean()
```

```
Out[28]: 33042.00597528302
```

```
In [29]: 1 #eğer sigara içen ve erkek olanlardan en yüksek sigorta pirimini ödeyen kişi
          2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes')].max()
```

```
Out[29]: index                1321
         yaş                  64
         cinsiyet            male
         body_maas_index      52.58
         çocuk_sayısı         4
         sigara_içiyor        yes
         bölge               southwest
         sigorta_pirim_tutarı  62592.87309
         dtype: object
```

```
In [30]: 1 #eğer sigara içen ve erkek olanlardan en düşük sigorta pirimini ödeyen kişi
        2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes')].min()
```

```
Out[30]: index          14
        yaş            18
        cinsiyet       male
        body_maas_index 17.29
        çocuk_sayısı    0
        sigara_içiyor   yes
        bölge          northeast
        sigorta_pirim_tutarı 12829.4551
        dtype: object
```

```
In [31]: 1 #eğer sigara içen ve erkek olanların ve souteast'de oturanların ortalama sigorta pirimini görmek istersem
        2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes') & (df['bölge']=='southeast')]['sigorta_pirim_tutarı'].i
```

```
Out[31]: 36029.83936654546
```

```
In [32]: 1 #eğer sigara içen ve kadın olanların ve southwest'de oturanların ortalama sigorta pirimini görmek istersem
        2 df[(df['cinsiyet']=='female') & (df['sigara_içiyor']=='yes') & (df['bölge']=='southwest')]['sigorta_pirim_tutarı']
```

```
Out[32]: 31687.988429999998
```

```
In [33]: 1 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes') & ((df['bölge']=='southwest')|(df['bölge']=='southeast
```

```
Out[33]: 34649.99011695652
```

```
In [34]: 1 #eğer sigara içen ve erkek olanların ve southwest'de veya southeast'da oturanların ortalama sigorta pirimini görn
        2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes') & (df['bölge']=='southwest') | (df['bölge']=='southeas
```

```
Out[34]: 16383.660072044897
```

```
In [35]: 1 #eğer sigara içen ve kadın olanların ve nothwest'de veya northeast'da oturanların ortalama sigorta primini görme  
2 df[(df['cinsiyet']=='male') & (df['sigara_içiyor']=='yes') & (df['bölge']=='northwest') | (df['bölge']=='northhea
```

Out[35]: 30713.181419310342

```
In [36]: 1 # ortalama sigorta priminin üzerinde ödeme yapanların ortalama yaşları nedir  
2 df[df['sigorta_pirim_tutarı']>df['sigorta_pirim_tutarı'].mean()]['yaş'].mean()
```

Out[36]: 42.464285714285715

```
In [37]: 1 # ortalama sigorta priminin altında ödeme yapanların ortalama yaşları nedir  
2 df[df['sigorta_pirim_tutarı']<df['sigorta_pirim_tutarı'].mean()]['yaş'].mean()
```

Out[37]: 37.73718647764449

```
In [38]: 1 # ortalama sigorta priminin üzerinde ödeme yapanların ortalama body_maas_index nedir  
2 df[df['sigorta_pirim_tutarı']>df['sigorta_pirim_tutarı'].mean()]['body_maas_index'].mean()
```

Out[38]: 31.039249999999992

```
In [39]: 1 # ortalama sigorta priminin üzerinde ödeme yapanların ortalama body_maas_index nedir  
2 df[df['sigorta_pirim_tutarı']<df['sigorta_pirim_tutarı'].mean()]['body_maas_index'].mean()
```

Out[39]: 30.49133042529988

```
In [40]: 1 # ortalama sigorta priminin üzerinde ödeme yapanların ortalama çocuk sayısı  
2 df[df['sigorta_pirim_tutarı']>df['sigorta_pirim_tutarı'].mean()]['çocuk_sayısı'].mean()
```

Out[40]: 1.138095238095238

```
In [41]: 1 # ortalama sigorta priminin üzerinde ödeme yapanların ortalama çocuk sayısı
          2 df[df['sigorta_pirim_tutarı']<df['sigorta_pirim_tutarı'].mean()]['çocuk_sayısı'].mean()
```

Out[41]: 1.0763358778625953

```
In [42]: 1 # ortalama body_mass_index'in üzerinde olanların ödediği ortalama sigorta_pirimi
          2
          3 df[df['body_maas_index']>df['body_maas_index'].mean()]['sigorta_pirim_tutarı'].mean()
```

Out[42]: 15801.788405743035

```
In [43]: 1 # ortalama body_mass_index'in altında olanların ödediği ortalama sigorta_pirimi
          2
          3 df[df['body_maas_index']<df['body_maas_index'].mean()]['sigorta_pirim_tutarı'].mean()
```

Out[43]: 10920.738230895791

```
In [44]: 1 # Kadın ve erkekelerin ortalamaları
          2 df.groupby('cinsiyet').mean()
```

Out[44]:

	index	yaş	body_maas_index	çocuk_sayısı	sigorta_pirim_tutarı
cinsiyet					
female	667.093656	39.503021	30.377749	1.074018	12569.578844
male	670.060741	38.946667	30.943652	1.117037	13974.998864

```
In [45]: 1 # Kadın ve erkelelerin maksimum değerleri
          2 df.groupby('cinsiyet').max()
```

Out[45]:

	index	yaş	body_maas_index	çocuk_sayısı	sigara_içiyor	bölge	sigorta_pirim_tutarı
cinsiyet							
female	1340	64	48.07	5	yes	southwest	63770.42801
male	1336	64	53.13	5	yes	southwest	62592.87309

```
In [46]: 1 # Kadın ve erkelelerin minimum değerleri
          2 df.groupby('cinsiyet').min()
```

Out[46]:

	index	yaş	body_maas_index	çocuk_sayısı	sigara_içiyor	bölge	sigorta_pirim_tutarı
cinsiyet							
female	0	18	16.815	0	no	northeast	1607.5101
male	1	18	15.960	0	no	northeast	1121.8739

```
In [47]: 1 #çapraz tablolalma, kadın ve erkelelerin sigara içenlerin yüzdesi
          2 pd.crosstab(df["cinsiyet"], df["sigara_içiyor"], normalize=True)
```

Out[47]:

	sigara_içiyor	no	yes
cinsiyet			
female	0.409125	0.086013	
male	0.385939	0.118923	

```
In [48]: 1 #çapraz tablolalma, sigara içenler,in bölgeler üzerinde yüzde dağılımı  
2 pd.crosstab(df["bölge"], df["sigara_içiyor"], normalize=True)
```

Out[48]:

sigara_içiyor	no	yes
bölge		
northeast	0.192221	0.050112
northwest	0.198953	0.043381
southeast	0.204188	0.068063
southwest	0.199701	0.043381

In []:

1