



SAKARYA
ÜNİVERSİTESİ

BÜYÜK VERİYE GİRİŞ

Big Market Data - Full Depth Analysis

B181210051 Duhan UZUN 1/A

B181210057 Kadir ÇELİK 1/A

Projenin Amacı

Projemizin amacı, bir mağaza zincirinin verileri üzerinde büyük veri analizi yaparak ürün satışına (gelirine) etki eden etmenleri belirlemek ve satışlarını arttırmak için neler yapılabileceğini incelemektir. Projemizde gıda ürünleri satan bir mağaza zincirini incelemekteyiz. Mağazamızda; dondurulmuş yiyecekler ve fırın ürünleri gibi toplamda 16 kategoriden oluşan 8523 tane ürün bulunmaktadır. İçerisinde en çok ürün bulunduran kategoriler ise atıştırmalık yiyecekler ve meyve-sebzelerdir. Bu iki kategoride bulunan ürünlerin toplamının mağazada bulunan tüm ürünlere oranı yaklaşık %26’dır. İncelememiz, ürünün mağazada satışa çıktığı yıl, mağazanın büyüklüğü ve ürünün ağırlığı gibi 10 tane kategoriden oluşmaktadır.

Projede Kullanılan Araçlar ve Teknolojiler

Projemizde Python programlama dilini kullandık. Projemizde Python dilinde yer alan;

- Numpy,
- Pandas,
- Seaborn,
- Matplotlib,
- Scipy ve
- Missingno kütüphaneleri kullanılmıştır.

Veri seti olarak Big Market’in oluşturmuş olduğu “train.csv” dosyası içinde yer alan verileri kullandık.

Projede Kullanılan Veri Seti

```
train.csv
1 Item_Identifier,Item_Weight,Item_Fat_Content,Item_Visibility,Item_Type,Item_MRP,Outlet_Identifier,Outlet_Establishment_Year,Outlet_Size,Outlet_Location_Type,Outlet_Type,Item_Outlet_Sales
2 FDA15,9.3,Low Fat,0.016047301,Dairy,249.8092,OUT049,1999,Medium,Tier 1,Supermarket Type1,3735.138
3 DRC01,5.92,Regular,0.019278216,Soft Drinks,48.2692,OUT018,2009,Medium,Tier 3,Supermarket Type2,443.4228
4 FDM15,17.5,Low Fat,0.016760075,Meat,141.618,OUT049,1999,Medium,Tier 1,Supermarket Type1,2097.27
5 FDX07,19.2,Regular,0,Fruits and Vegetables,182.095,OUT010,1998,,Tier 3,Grocery Store,732.38
6 NCD19,8.93,Low Fat,0,Household,53.8614,OUT013,1987,High,Tier 3,Supermarket Type1,994.7052
7 FDP36,10.395,Regular,0,Baking Goods,51.4008,OUT018,2009,Medium,Tier 3,Supermarket Type2,556.6088
8 FDO10,13.65,Regular,0.012741089,Snack Foods,57.6588,OUT013,1987,High,Tier 3,Supermarket Type1,343.5528
9 FDP10,,Low Fat,0.127469857,Snack Foods,107.7622,OUT027,1985,Medium,Tier 3,Supermarket Type3,4022.7636
10 FDM17,16.2,Regular,0.016607114,Frozen Foods,96.9726,OUT045,2002,,Tier 2,Supermarket Type1,1076.5986
11 FDU28,19.2,Regular,0.09444959,Frozen Foods,187.8214,OUT017,2007,,Tier 2,Supermarket Type1,4710.535
12 FDY07,11.8,Low Fat,0,Fruits and Vegetables,45.5402,OUT049,1999,Medium,Tier 1,Supermarket Type1,1516.0266
13 FDA03,18.5,Regular,0.045463773,Dairy,144.1102,OUT046,1997,Small,Tier 1,Supermarket Type1,2187.153
14 FDX32,15.1,Regular,0.1000135,Fruits and Vegetables,145.4786,OUT049,1999,Medium,Tier 1,Supermarket Type1,1589.2646
15 FDS46,17.6,Regular,0.047257328,Snack Foods,119.6782,OUT046,1997,Small,Tier 1,Supermarket Type1,2145.2076
16 FDP32,16.35,Low Fat,0.0680243,Fruits and Vegetables,196.4426,OUT013,1987,High,Tier 3,Supermarket Type1,1977.426
17 FDP49,9,Regular,0.069088961,Breakfast,56.3614,OUT046,1997,Small,Tier 1,Supermarket Type1,1547.3192
18 NCB42,11.8,Low Fat,0.008596051,Health and Hygiene,115.3492,OUT018,2009,Medium,Tier 3,Supermarket Type2,1621.8888
19 FDP49,9,Regular,0.069196376,Breakfast,54.3614,OUT049,1999,Medium,Tier 1,Supermarket Type1,718.3982
20 DR111,,Low Fat,0.034237682,Hard Drinks,113.2834,OUT027,1985,Medium,Tier 3,Supermarket Type3,2303.668
21 FDM02,13.35,Low Fat,0.10249212,Dairy,230.5352,OUT035,2004,Small,Tier 2,Supermarket Type1,2748.4224
22 FDM22,18.85,Regular,0.138190277,Snack Foods,250.8724,OUT013,1987,High,Tier 3,Supermarket Type1,3775.086
23 FDM12,,Regular,0.035399923,Baking Goods,144.5444,OUT027,1985,Medium,Tier 3,Supermarket Type3,4064.0432
24 NCB30,14.6,Low Fat,0.025698134,Household,196.5084,OUT035,2004,Small,Tier 2,Supermarket Type1,1587.2672
25 FDC37,,Low Fat,0.057556998,Baking Goods,107.6938,OUT019,1985,Small,Tier 1,Grocery Store,214.3876
```

Şekil 1. Veri setinden bir kesit.

Veri setinde, rastgele bir mağaza zincirinin çeşitli mağazalarındaki ürünlerin satışına etki ettiği düşünülen çeşitli etmenler ve bu etmenlere ait veriler vardır. Veri seti 10 tane kategoriye bölünmüştür. Ürünlere, belirli özelliklerine göre eşsiz bir tanımlayıcı atanmıştır (Örnek: FDC37). Veri setinde ürünler üzerinde ürünlerin;

- Ağırlıklarına,
- Bir ürünün, mağazadaki diğer ürünlere göre bulunma yüzdesine,
- Yağ oranına,
- Bulundukları kategorilere (Örnek: atıştırmalık yiyecekler) ve
- MRP'sine (Malzeme ihtiyaç planlaması) göre sınıflandırma işlemi yapılmıştır.

Veri setinde mağazalar üzerinde ise mağazaların;

- Türü,
- Konumu,
- Kuruluş yılı ve
- Büyüklüklerine göre bir sınıflandırma işlemi yapılmıştır.

Projede Kullanılan Algoritma

Projenin analizinde Exploratory Data Analysis (Keşifsel Veri Analizi) algoritması kullanılmıştır.

Keşifsel Veri Analizi (EDA), veri bilimcileri tarafından veri kümelerini analiz edip araştırmak ve veri görselleştirme yöntemlerini kullanarak bu veri kümelerinin ana özelliklerini özetlemek için kullanılır. İhtiyaç duyulan yanıtları almak için veri kaynaklarını en iyi nasıl işlenebileceğini belirleyerek; veri bilimcilerin kalıpları keşfetmelerini, anormallikleri fark etmelerini, bir hipotezi test etmelerini ya da varsayımları kontrol etmelerini kolaylaştırır.

Keşifsel Veri Analizi, öncelikle verilerin resmi modelleme veya hipotez testi görevinin ötesinde neleri ortaya çıkarabileceğini görmek için kullanılır. Ayrıca bir veri kümesi değişkenleri ve değişkenler arasındaki ilişkilere ilişkin daha iyi bir bakış açısı sağlar. Bunların dışında veri analizi için düşünülen istatistiksel tekniklerin analiz için uygun olup olmadığını belirlemeye de yardımcı olabilir.

EDA'nın karakteristik özellikleri aşağıda sırasıyla verilmiştir;

1. Esneklik: Veriye aktarma konusunda daha fazla esnekliğe sahip teknikler
2. Pratiklik: Veri analiz prosedürleri için tavsiyeler
3. İnovasyon: Sonuçları yorumlama teknikleri

4. Evrensellik: Veri analizi için geçerli olan tüm istatistikleri kullanma
5. Sadelik: Her şeyden önce, sadeliğin altın kural olduğuna inanma

Keşifsel veri analizini oluşturmak için kullanılan en yaygın veri bilimi araçları Python ve R programlama dilleridir.

Uygulama

Adım 1: Veri setinin okunması.

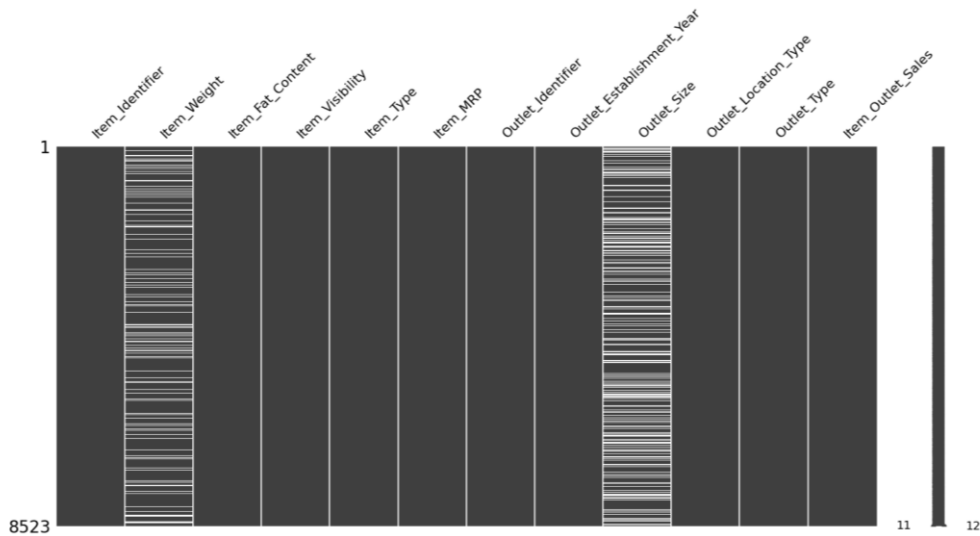
Bu adım sonucunda veri setinde 8523 adet veri ve 12 adet özelliğin bulunduğu öğrenilmiştir.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Item_Identifier Item_Weight Item_Fat_Content ... Outlet_Location_Type Outlet_Type Item_Outlet_Sales
0 FDA15 9.300 Low Fat ... Tier 1 Supermarket Type1 3735.1380
1 DRC01 5.920 Regular ... Tier 3 Supermarket Type2 443.4228
2 FDN15 17.500 Low Fat ... Tier 1 Supermarket Type1 2097.2700
3 FDX07 19.200 Regular ... Tier 3 Grocery Store 732.3800
4 NCD19 8.930 Low Fat ... Tier 3 Supermarket Type1 994.7052
... ... ... ... ...
8518 FDF22 6.865 Low Fat ... Tier 3 Supermarket Type1 2778.3834
8519 FDS36 8.380 Regular ... Tier 2 Supermarket Type1 549.2850
8520 NCJ29 10.600 Low Fat ... Tier 2 Supermarket Type1 1193.1136
8521 FDN46 7.210 Regular ... Tier 3 Supermarket Type2 1845.5976
8522 DRG01 14.800 Low Fat ... Tier 1 Supermarket Type1 765.6700
```

Şekil 2. Veri setinin okunması.

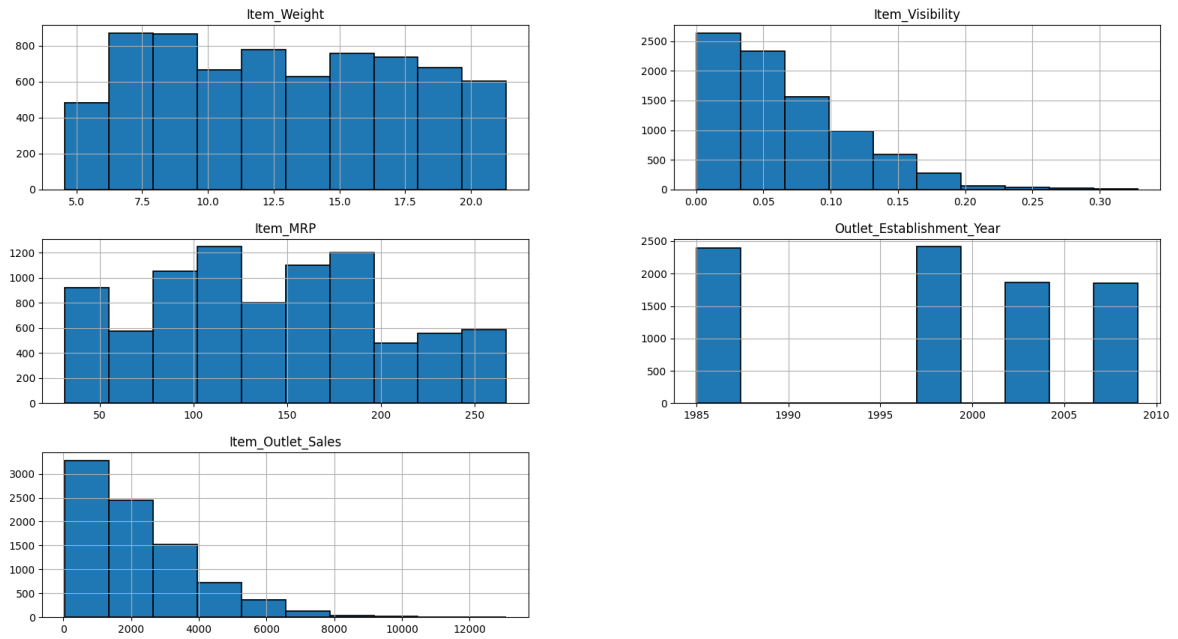
Adım 2: Veri setinin matris grafiğinin çıkarılması.

Bu adımda veri setinin matris grafiği çıkartılarak, veri setinde bulunan boş değerler tespit edilmiştir.



Şekil 3. Veri setinin matris grafiği.

Adım 3: Veri setinde yer alan sütunlara ait histogram grafiklerinin çıkarılması.



Şekil 4. Veri setinde yer alan sütunlara ait histogram grafikleri.

Bu adımda bir ürünün mağazadaki teşhir alanı ne kadar fazla ise ürünün satış sayısının da o kadar fazla olduğunu görmekteyiz. Yani ürün görünürlüğü ile ürün satışı arasında doğru orantı bulunmaktadır.

Adım 4: İlk sütun olan “Item_Identifier” sütununun kontrol edilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
0      FDA15
1      DRC01
2      FDN15
3      FDX07
4      NCD19
...
8518   FDF22
8519   FDS36
8520   NCJ29
8521   FDN46
8522   DRG01
Name: Item_Identifier, Length: 8523, dtype: object
```

Şekil 5. “Item_Identifier” sütunu.

“Item_Identifier” sütununu incelediğimizde ilk iki harfı birbiriyle aynı olan birçok değerin olduğu görülmektedir. Bu yüzden “Item_Identifier” sütunu içerisinde yer alan değerlerin ilk iki harfini içerisinde barındıran “Item_Code” adında yeni bir sütun oluşturulur.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
0      FD
1      DR
2      FD
3      FD
4      NC
..
8518   FD
8519   FD
8520   NC
8521   FD
8522   DR
Name: Item_Code, Length: 8523, dtype: object
```

Şekil 6. “Item_Code” sütunu.

Adım 5: Veri setinin kontrol edilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Item_Identifier                       8523 non-null   object
1   Item_Weight                          7060 non-null   float64
2   Item_Fat_Content                     8523 non-null   object
3   Item_Visibility                      8523 non-null   float64
4   Item_Type                           8523 non-null   object
5   Item_MRP                            8523 non-null   float64
6   Outlet_Identifier                    8523 non-null   object
7   Outlet_Establishment_Year            8523 non-null   int64
8   Outlet_Size                          6113 non-null   object
9   Outlet_Location_Type                 8523 non-null   object
10  Outlet_Type                          8523 non-null   object
11  Item_Outlet_Sales                    8523 non-null   float64
12  Item_Code                            8523 non-null   object
dtypes: float64(4), int64(1), object(8)
memory usage: 865.7+ KB
```

Şekil 7. Veri seti kontrolü.

Adım 6: Veri setinde önemi kalmayan “Item_Identifier” sütununun silinmesi.

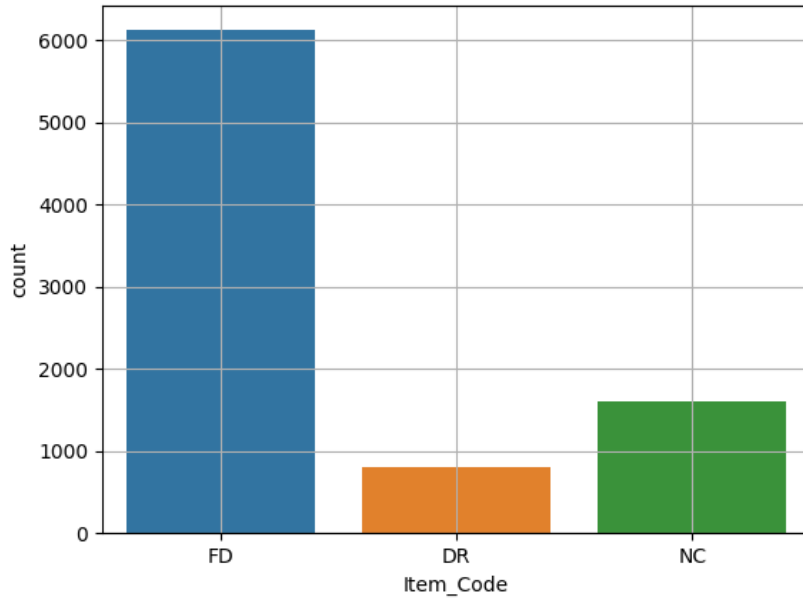
```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Item_Weight Item_Fat_Content Item_Visibility Item_Type ... Outlet_Location_Type Outlet_Type Item_Outlet_Sales Item_Code
0 9.300 Low Fat 0.016047 Dairy ... Tier 1 Supermarket Type1 3735.1380 FD
1 5.920 Regular 0.019278 Soft Drinks ... Tier 3 Supermarket Type2 443.4228 DR
2 17.500 Low Fat 0.016760 Meat ... Tier 1 Supermarket Type1 2097.2700 FD
3 19.200 Regular 0.000000 Fruits and Vegetables ... Tier 3 Grocery Store 732.3800 FD
4 8.930 Low Fat 0.000000 Household ... Tier 3 Supermarket Type1 994.7052 NC
... ..
8518 6.865 Low Fat 0.056783 Snack Foods ... Tier 3 Supermarket Type1 2778.3834 FD
8519 8.380 Regular 0.046982 Baking Goods ... Tier 2 Supermarket Type1 549.2850 FD
8520 10.600 Low Fat 0.035186 Health and Hygiene ... Tier 2 Supermarket Type1 1193.1136 NC
8521 7.210 Regular 0.145221 Snack Foods ... Tier 3 Supermarket Type2 1845.5976 FD
8522 14.800 Low Fat 0.044878 Soft Drinks ... Tier 1 Supermarket Type1 765.6700 DR
[8523 rows x 12 columns]
```

Şekil 8. Veri setinin yeni hali.

Adım 7: Eşsiz “Item_Code” değerleri sayısının tespit edilmesi ve grafiğinin çıkarılması.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
FD 6125
NC 1599
DR 799
Name: Item_Code, dtype: int64
```

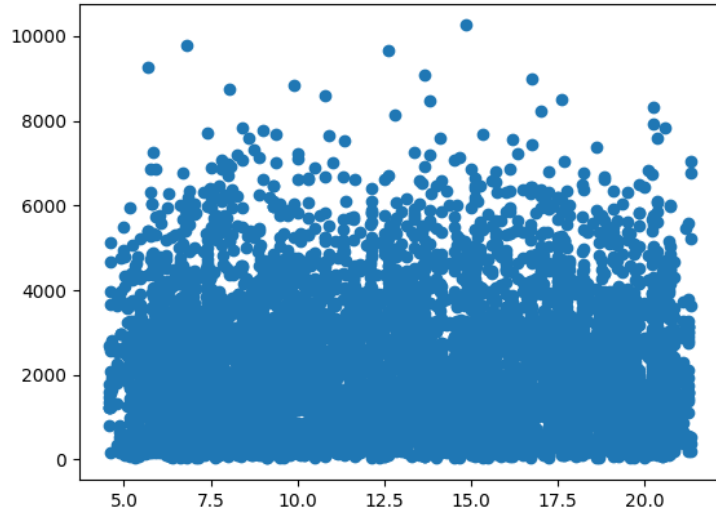
Şekil 9. “Item_Code” içerisindeki değerlerin toplam sayısı.



Şekil 10. “Item_Code” değerlerinin sayısını gösteren grafik.

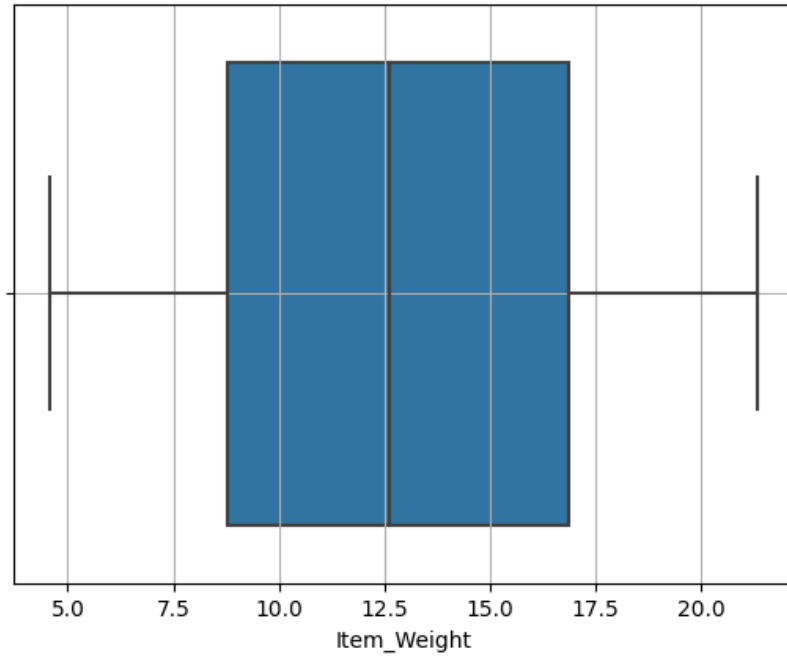
Bu işlemler sonucunda “FD” değerlerinin çoğunluğu oluşturduğu görülmektedir.

Adım 8: “Item_Weight” ve “Item_Outlet_Sales” arasındaki ilişkiyi gösteren serpilme diyagramının çıkarılması.



Şekil 11. Serpilme diyagramı (Item_Weight-Item_Outlet_Sales).

Adım 9: “Item_Weight” alanında aykırı bir değer olup olmadığının kontrol edilmesi ve alanın ortanca değerinin bulunması.



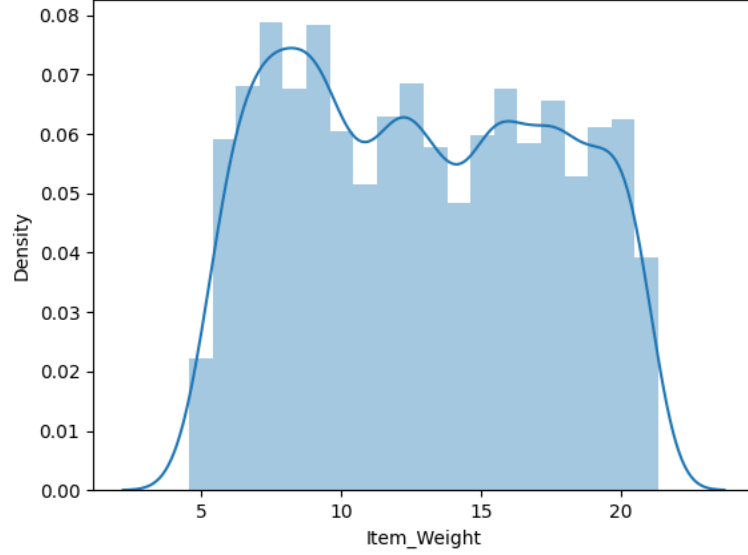
Şekil 12. “Item_Weight” kutu grafiği.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
12.6
```

Şekil 13. “Item_Weight” alanının ortanca değeri.

Bu adımda “Item_Weight” içerisinde aykırı bir değer bulunmadığı öğrenilmiştir.

Adım 10: “Item_Weight” alanının histogram grafiğinin çizilmesi ve çarpıklık – yığılma derecelerinin elde edilmesi.



Şekil 14. “Item_Weight” histogram grafiği.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
Kurtosis: -1.2277664144376634  
Skewness: 0.0824262091221237
```

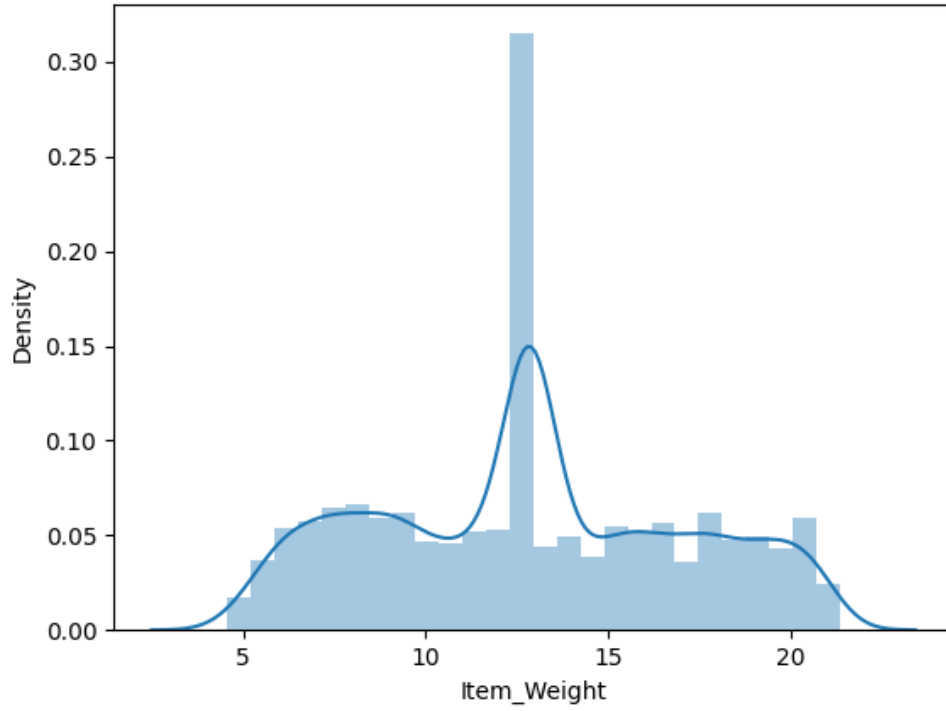
Şekil 15. “Item_Weight” alanının kurtosis (yığılma) ve skewness (çarpıklık) dereceleri.

Bu adımda “Item_Weight” alanında yığılma ve çarpıklığın olmadığı sonucu elde edilmiştir. Bu yüzden “Item_Weight” alanında yer alan boş değerleri alanın ortalama değeri ile doldurabiliriz.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
1463  
0
```

Şekil 16. Alandaki boş alanların doldurulmadan önceki ve doldurulduktan sonraki sayısı.

Adım 10: Boş alanları doldurulan “Item_Weight” alanının histogram grafiğinin çizilmesi ve çarpıklık – yığılma derecelerinin elde edilmesi.



Şekil 17. “Item_Weight” histogram grafiği.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
Kurtosis: -1.2277664144376634  
Skewness: 0.0824262091221237
```

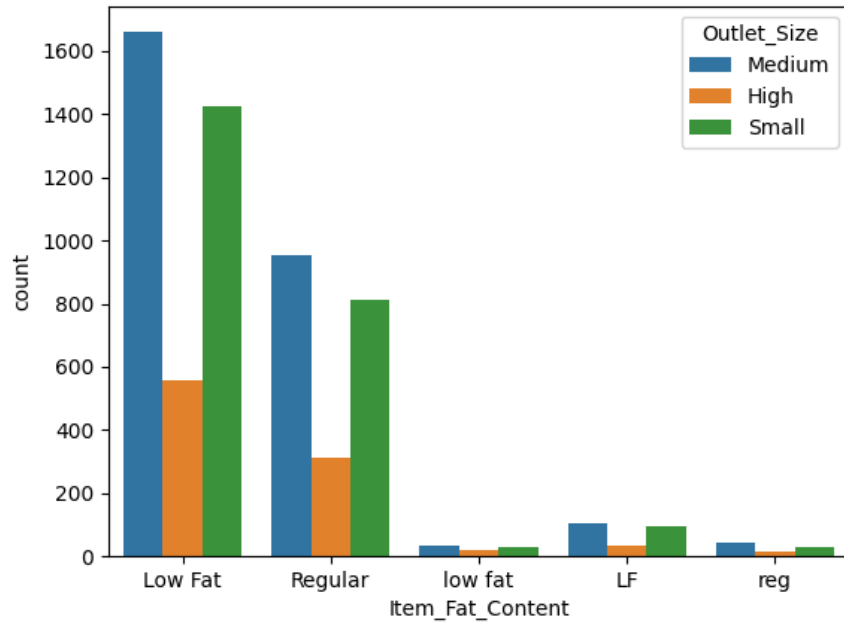
Şekil 18. “Item_Weight” alanının kurtosis (yığılma) ve skewness (çarpıklık) dereceleri.

Adım 11: “Item_Fat_Content” alanının incelenmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
Low Fat    5089  
Regular    2889  
LF         316  
reg        117  
low fat    112  
Name: Item_Fat_Content, dtype: int64
```

Şekil 19. “Item_Fat_Content” alanında bulunan değerlerin toplam sayıları.

Adım 12: “Item_Fat_Content” – “Outlet_Size” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.

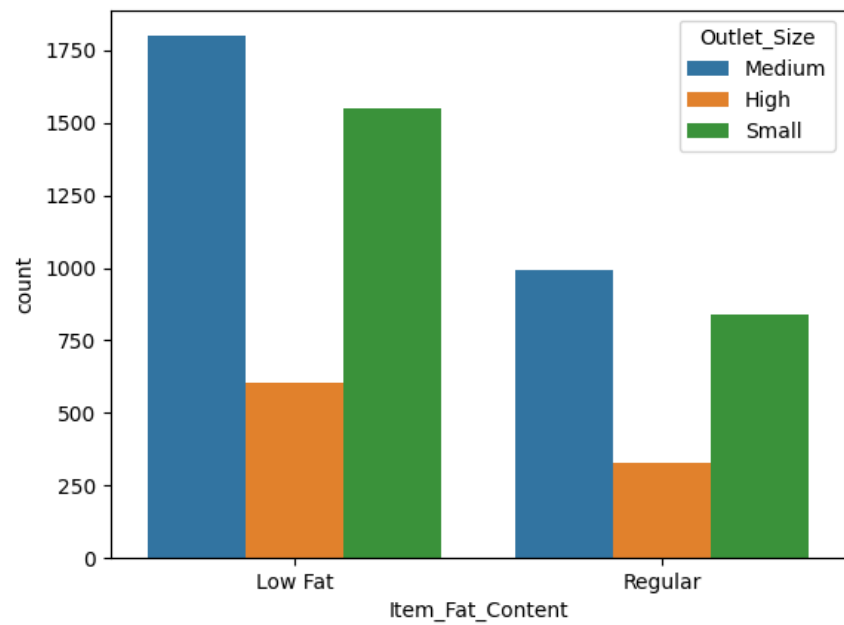


Şekil 20. “Item_Fat_Content” – “Outlet_Size” alanları ile çizilen grafik.

Adım 13: “Low Fat” ve “Regular” tiplerinde aynı türden birden fazla kategori olduğu için isimlerinin tek bir tip altında birleştirilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Low Fat    5517
Regular    3006
Name: Item_Fat_Content, dtype: int64
```

Şekil 21. “Item_Fat_Content” alanının yeni hali.



Şekil 22. “Item_Fat_Content” – “Outlet_Size” alanları ile çizilen grafiğin yeni hali.

Buradan “Low Fat” alanında “Regular” alanına göre daha fazla ürün olduğunu görmekteyiz. Bu durum “Outlet_Size” ile ilişkilidir.

Adım 14: Alanlar arasındaki ilişkileri incelemek için sıcaklık grafiğinin çıkarılması.



Şekil 23. Sıcaklık grafiği.

Burada “Item_MRP” alanının “Item_Outlet_Sales” alanının %50’sini, “Item_Visibility” alanının ise %13’ünü oluşturduğu görülmektedir.

Adım 15: “Item_Fat_Content” alanının nümerik olarak yeniden düzenlenmesi ve korelasyonun tekrar kontrol edilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Item_Weight  Item_Fat_Content  ...  Outlet_Establishment_Year  Item_Outlet_Sales
Item_Weight    1.000000    0.021157  ...                -0.008301    0.011550
Item_Fat_Content  0.021157    1.000000  ...                -0.003151    -0.018719
Item_Visibility  -0.012049   -0.047314  ...                -0.074834    -0.128625
Item_MRP        0.024756   -0.006063  ...                 0.005020    0.567574
Outlet_Establishment_Year -0.008301   -0.003151  ...                 1.000000   -0.049135
Item_Outlet_Sales  0.011550   -0.018719  ...                -0.049135    1.000000

[6 rows x 6 columns]
```

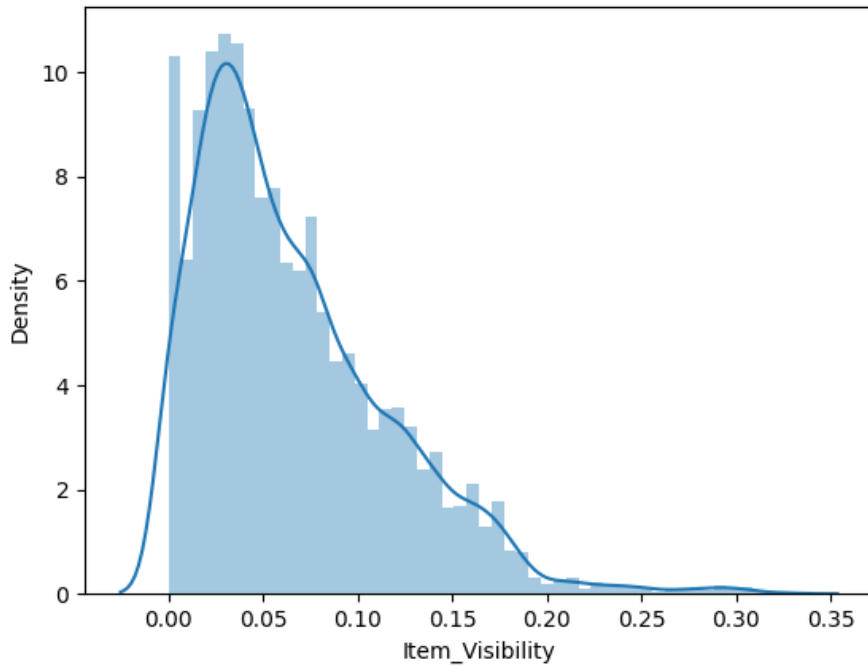
Şekil 24. Korelasyon.

Adım 16: “Item_Visibility” alanının incelenmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
0      0.016047
1      0.019278
2      0.016760
3      0.000000
4      0.000000
...
8518   0.056783
8519   0.046982
8520   0.035186
8521   0.145221
8522   0.044878
Name: Item_Visibility, Length: 8523, dtype: float64
```

Şekil 25. “Item_Visibility” alanı.

Adım 17: “Item_Visibility” alanının histogram grafiğinin çizilmesi ve çarpıklık derecesinin elde edilmesi.



Şekil 26. “Item_Visibility” histogram grafiği.

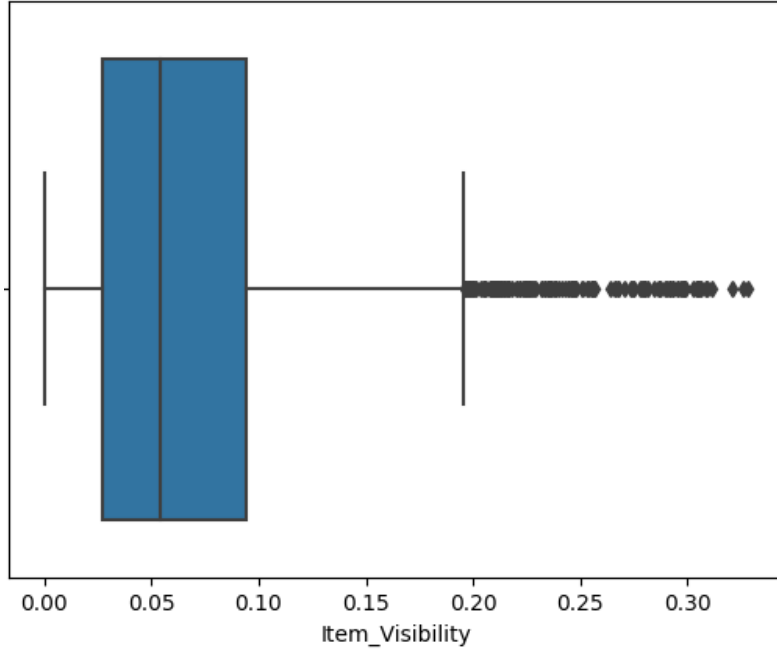
```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Skewness: 1.1670905496918396
```

Şekil 27. “Item_Weight” alanının skewness (çarpıklık) derecesi.

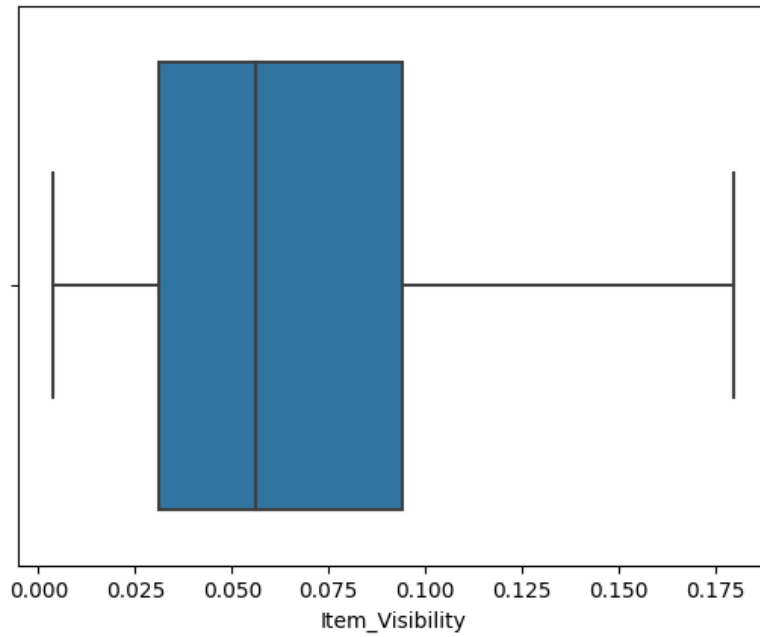
“Item_Visibility” alanının çarpıklık derecesi oldukça yüksek çıkmıştır. Bu durum, alan içerisinde aykırı bir değer olduğunu belirtmektedir.

Adım 18: “Item_Visibility” alanı içerisindeki aykırı değerlerin bulunması ve silinmesi.

İlk olarak alanın kutu grafiği çizilerek aykırı değerler bulunmaktadır. Bulunan aykırı değerler silinerek alan tekrar kontrol edilmektedir.



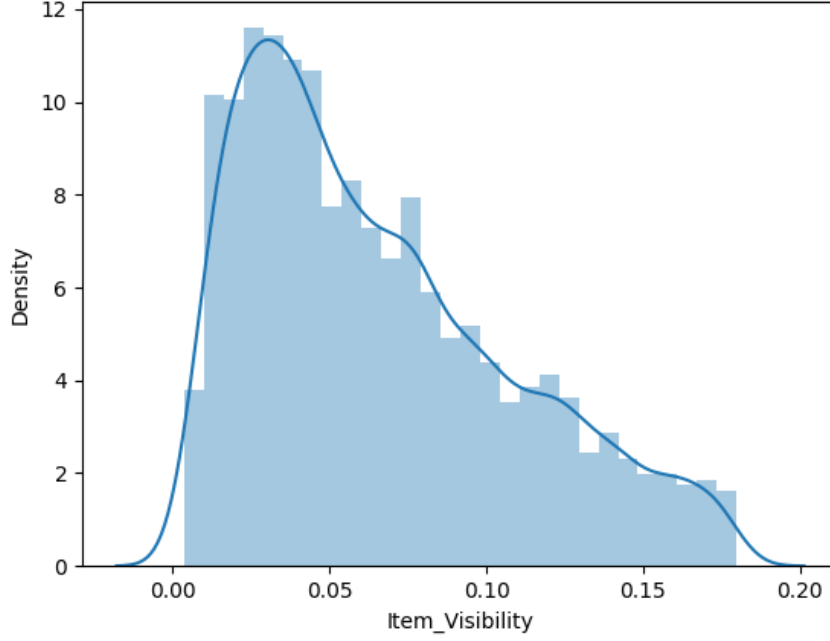
Şekil 28. “Item_Visibility” kutu grafiği.



Şekil 29. Aykırı değerlerin silinmesi sonrası oluşturulan “Item_Visibility” kutu grafiği.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
Skewness: 0.7331544572443054
```

Şekil 30. “Item_Visibility” alanının yeni çarpıklık derecesi.



Şekil 31. “Item_Visibility” alanının yeni histogram grafiği.

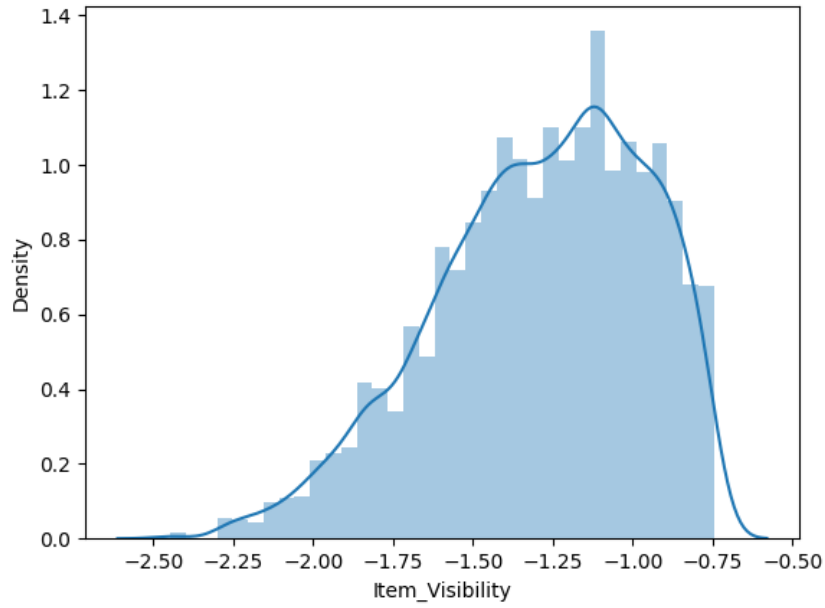
Bu işlemlerden sonra log dönüşümü yapılır ve alan tekrar kontrol edilir.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
0      -1.794598  
1      -1.714933  
2      -1.775724  
6      -1.894793  
7      -0.894593  
...  
8518   -1.245779  
8519   -1.328065  
8520   -1.453627  
8521   -0.837972  
8522   -1.347964  
Name: Item_Visibility, Length: 7769, dtype: float64
```

Şekil 32. Log dönüşümü sonrası “Item_Visibility” alanı.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py  
Skewness: -0.5193282394757215
```

Şekil 33. Log dönüşümü sonrası “Item_Visibility” alanının çarpıklık değeri.



Şekil 34. Log dönüşümü sonrası “Item_Visibility” histogram grafiği.

Adım 19: Korelasyonun kontrol edilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
      Item_Weight  Item_Fat_Content  ...  Outlet_Establishment_Year  Item_Outlet_Sales
Item_Weight      1.000000      0.025202  ...                -0.008069      0.008839
Item_Fat_Content  0.025202      1.000000  ...                -0.006849     -0.021578
Item_Visibility  -0.011904     -0.057360  ...                -0.022496     -0.059361
Item_MRP         0.027181     -0.012756  ...                 0.004344      0.577356
Outlet_Establishment_Year -0.008069     -0.006849  ...                 1.000000     -0.069994
Item_Outlet_Sales  0.008839     -0.021578  ...                -0.069994      1.000000

[6 rows x 6 columns]
```

Şekil 35. Korelasyon.

Adım 20: “Item_Type” alanının incelenmesi.

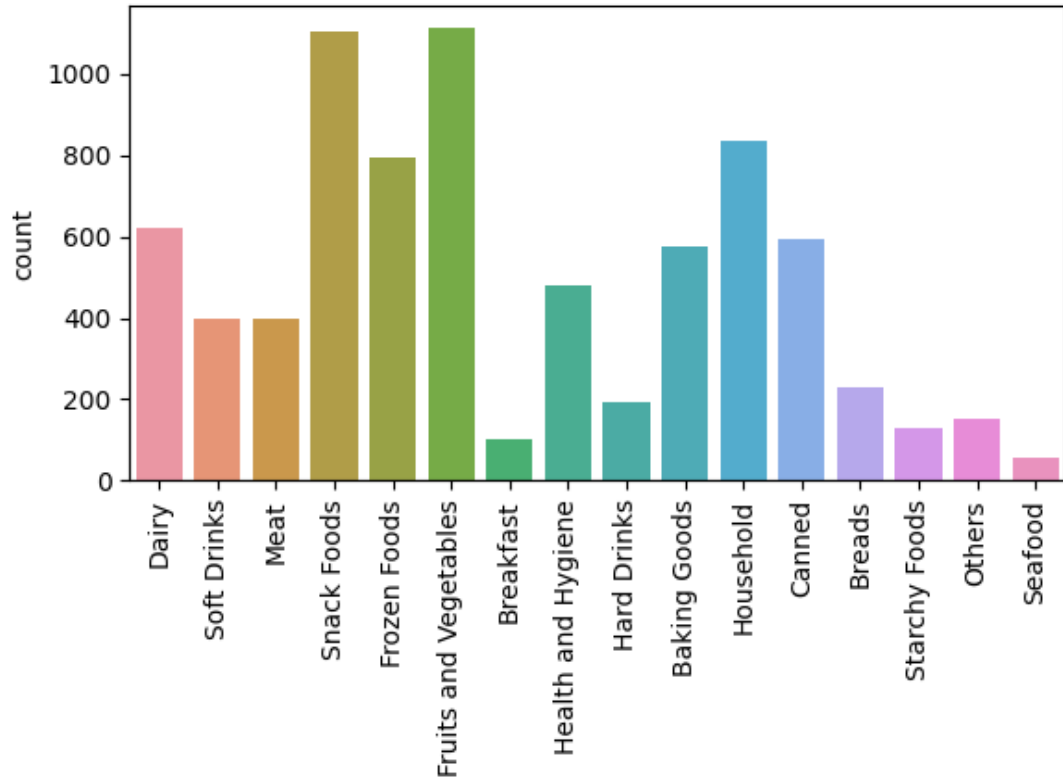
```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
0      Dairy
1      Soft Drinks
2      Meat
6      Snack Foods
7      Snack Foods
...
8518    Snack Foods
8519    Baking Goods
8520    Health and Hygiene
8521    Snack Foods
8522    Soft Drinks
Name: Item_Type, Length: 7769, dtype: object
```

Şekil 36. “Item_Type” alanı.

Adım 21: “Item_Type” alanındaki eşsiz değerlerin toplam sayılarının bulunması ve grafik olarak çizdirilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Fruits and Vegetables      1113
Snack Foods                1104
Household                  833
Frozen Foods               792
Dairy                     622
Canned                     593
Baking Goods               577
Health and Hygiene         479
Meat                       398
Soft Drinks                397
Breads                     229
Hard Drinks                193
Others                     151
Starchy Foods              128
Breakfast                  103
Seafood                     57
Name: Item_Type, dtype: int64
```

Şekil 37. “Item_Type” içerisindeki eşsiz değerler ve toplam sayıları.

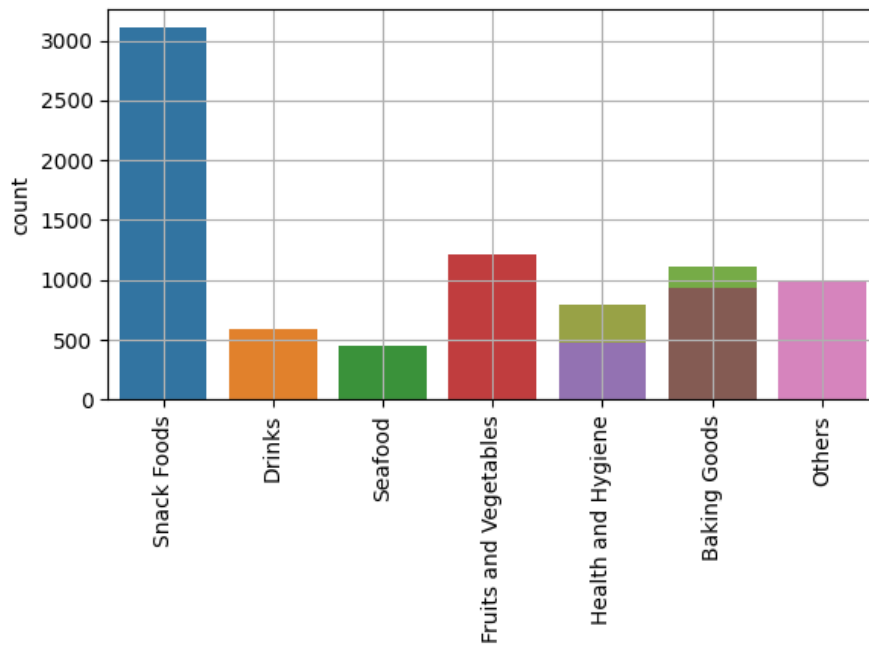


Şekil 38. “Item_Type” içerisindeki değerleri gösteren grafik.

Adım 22: “Item_Type” içerisinde yer alan farklı türdeki tiplerin kombine edilmesi.

```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
Snack Foods      3111
Fruits and Vegetables  1216
Others           984
Baking Goods     934
Drinks           590
Health and Hygiene  479
Seafood          455
Name: Item_Type, dtype: int64
```

Şekil 39. Kombine edilmiş “Item_Type” alanı.

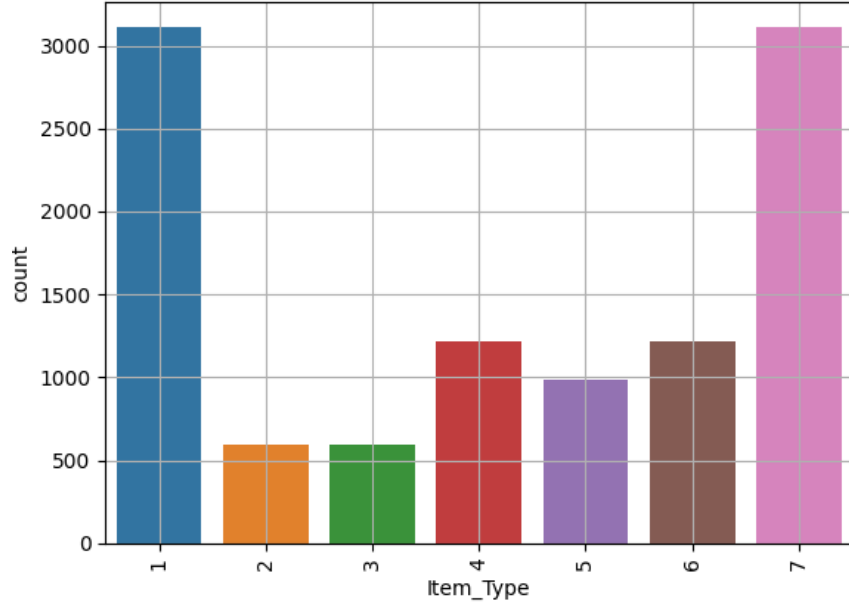


Şekil 40. Kombine edilmiş “Item_Type” değerlerini gösteren.

Adım 23: “Item_Type” alanının sayısal olarak yenilenmesi ve yüksek miktardaki ürüne daha yüksek değerin verilmesi.

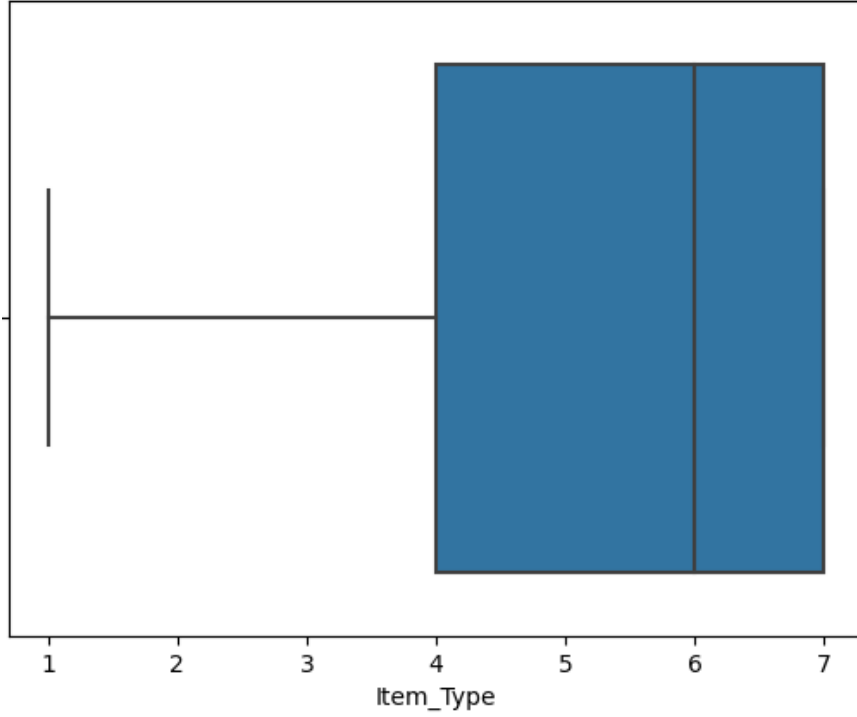
```
C:\Users\Duhan\Desktop\Big Data>python bigmarket.py
7      3111
6      1216
5       984
4       934
3       590
2       479
1       455
Name: Item_Type, dtype: int64
```

Şekil 41. Yenilenmiş “ItemType” alanı.



Şekil 42. Yenilenmiş “Item_Type” grafiği.

Adım 24: “Item_Type” alanının kutu grafiğinin çizilmesi.



Şekil 43. “Item_Type” kutu grafiği.

Bu grafikten yola çıkarak çoğu öge türünün grafiğin sağ tarafında yoğunlaştığını görebiliriz.

Adım 25: Veri seti içerisindeki ikili ilişkilerin çizdirilmesi.



Şekil 44. Veri setindeki ikili ilişkileri gösteren grafik.

Bu grafikten yola çıkarak MRP arttıkça, ürün satışlarının da arttığını görebiliriz.

Adım 26: Veri setinin kontrol edilmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
Item_Weight  Item_Fat_Content  Item_Visibility  ...  Outlet_Type  Item_Outlet_Sales  Item_Code
0      9.300000         1      -1.794598  ...  Supermarket Type1      3735.1380      FD
1      5.920000         0      -1.714933  ...  Supermarket Type2       443.4228      DR
2     17.500000         1      -1.775724  ...  Supermarket Type1     2097.2700      FD
6     13.650000         0      -1.894793  ...  Supermarket Type1      343.5528      FD
7     12.857645         1      -0.894593  ...  Supermarket Type3     4022.7636      FD
...
8518     6.865000         1      -1.245779  ...  Supermarket Type1     2778.3834      FD
8519     8.380000         0      -1.328065  ...  Supermarket Type1       549.2850      FD
8520    10.600000         1      -1.453627  ...  Supermarket Type1     1193.1136      NC
8521     7.210000         0      -0.837972  ...  Supermarket Type2     1845.5976      FD
8522    14.800000         1      -1.347964  ...  Supermarket Type1       765.6700      DR

[7769 rows x 12 columns]
```

Şekil 45. Veri seti.

Adım 27: “Item_MRP” alanının incelenmesi.

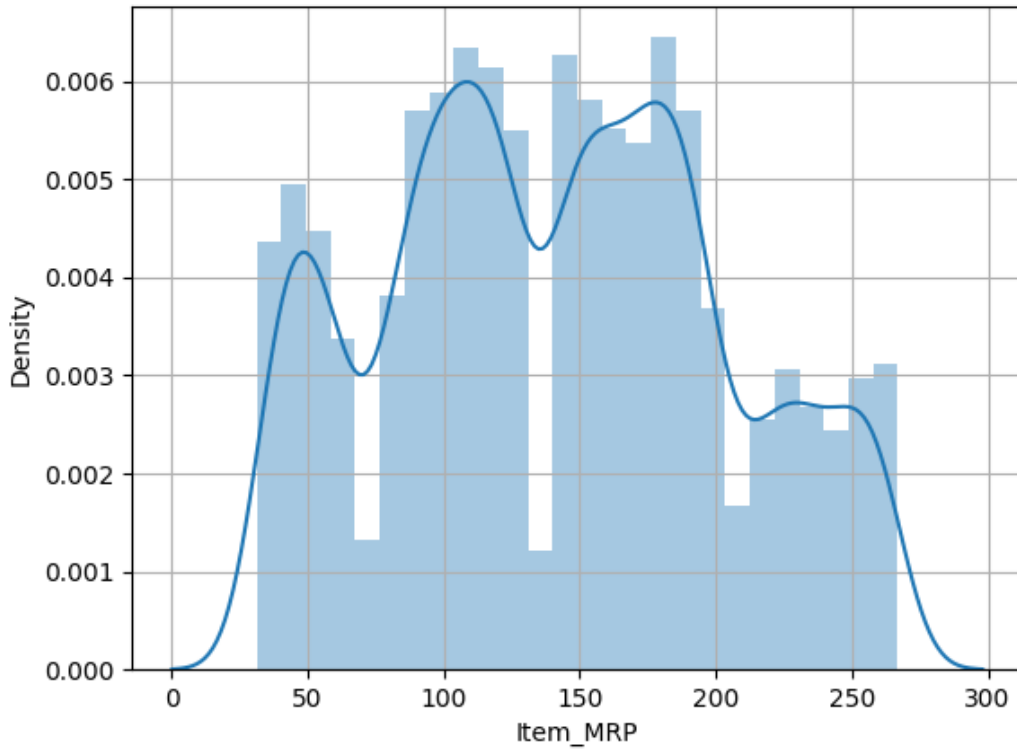
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py  
count      7769.000000  
mean       141.284440  
std         62.122423  
min         31.290000  
25%         94.175200  
50%        143.215400  
75%        185.826600  
max         266.888400  
Name: Item_MRP, dtype: float64
```

Şekil 46. “Item_MRP” alanı hakkında özet bilgiler.

Adım 28: “Item_MRP” alanının çarpıklık ve yığılma derecelerinin bulunması, histogram grafiğinin çizilmesi.

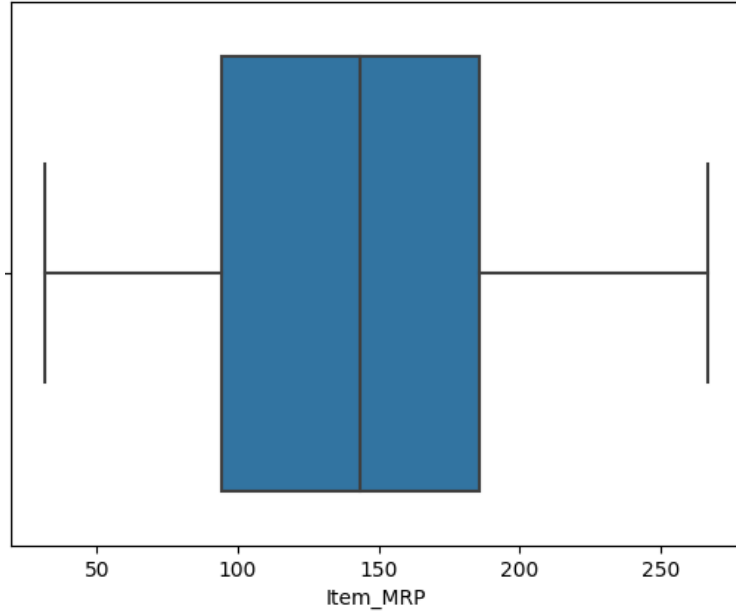
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py  
Kurtosis:  -0.8856077097078825  
Skewness:  0.12276214388421557
```

Şekil 47. “Item_MRP” alanının çarpıklık ve yığılma dereceli.



Şekil 48. “Item_MRP” histogram grafiği.

Histogram grafiđi, arpıklık ve yığılma dereceleri incelendiđinde “Item_MRP” alanında hi aykırı veri olmadıđı sonucu elde edilmektedir.

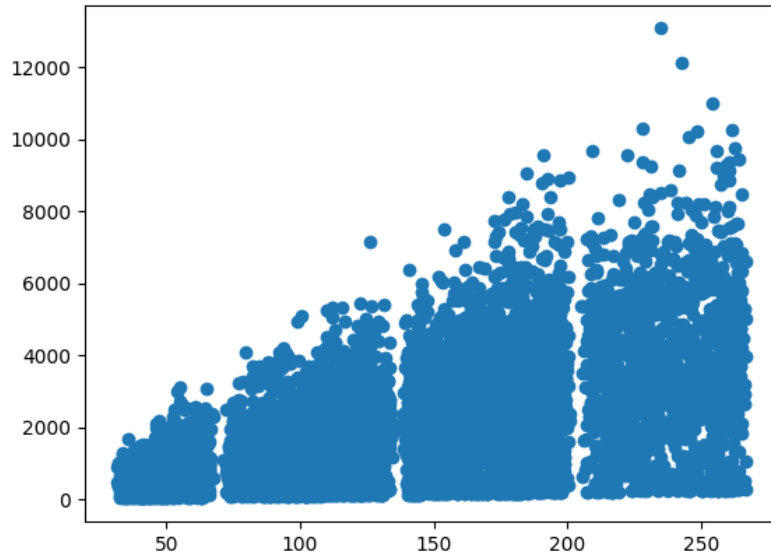


Şekil 49. “Item_MRP” kutu grafiđi.

Adım 29: “Item_Outlet_Sales” ve “Item_MRP” alanları arasındaki korelasyonun incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py  
Co-relation of sales and mrp is : 0.5773559921342624
```

Şekil 50. “Item_Outlet_Sales” ve “Item_MRP” alanları arasındaki korelasyon katsayısı.



Şekil 51. “Item_Outlet_Sales” ve “Item_MRP” alanları arasındaki korelasyonu gösteren grafik.

Adım 30: “Outlet_Identifier” alanının incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
0      OUT049
1      OUT018
2      OUT049
6      OUT013
7      OUT027
...
8518   OUT013
8519   OUT045
8520   OUT035
8521   OUT018
8522   OUT046
Name: Outlet_Identifier, Length: 7769, dtype: object
```

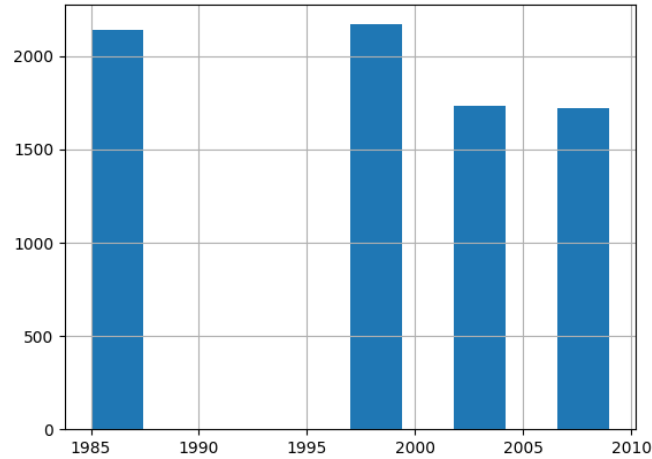
Şekil 52. “Outlet_Identifier” alanı.

Bu alan herhangi bir anlam içermediği için veri setinden silinir.

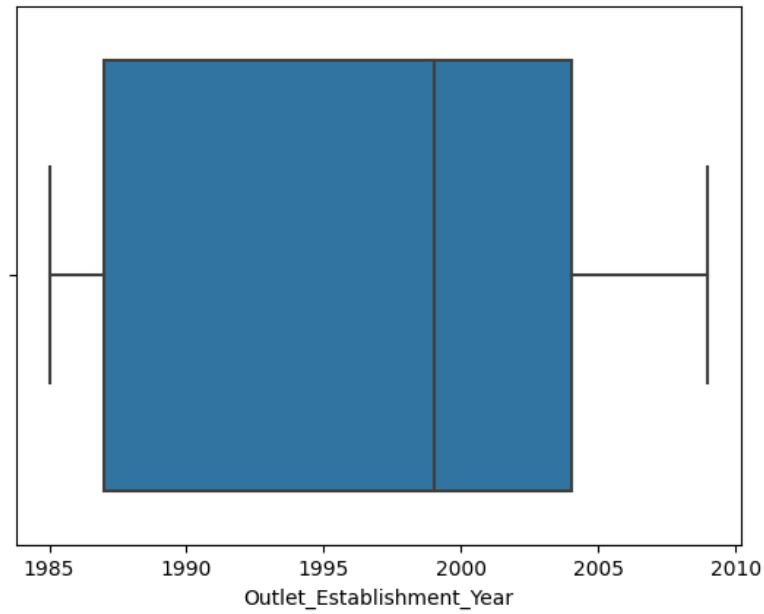
Adım 31: "Outlet_Establishment_Year" alanının incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
0      1999
1      2009
2      1999
6      1987
7      1985
...
8518   1987
8519   2002
8520   2004
8521   2009
8522   1997
Name: Outlet_Establishment_Year, Length: 7769, dtype: int64
```

Şekil 53. “Outlet_Establishment_Year” alanı.



Şekil 54. “Outlet_Establishment_Year” grafiği.

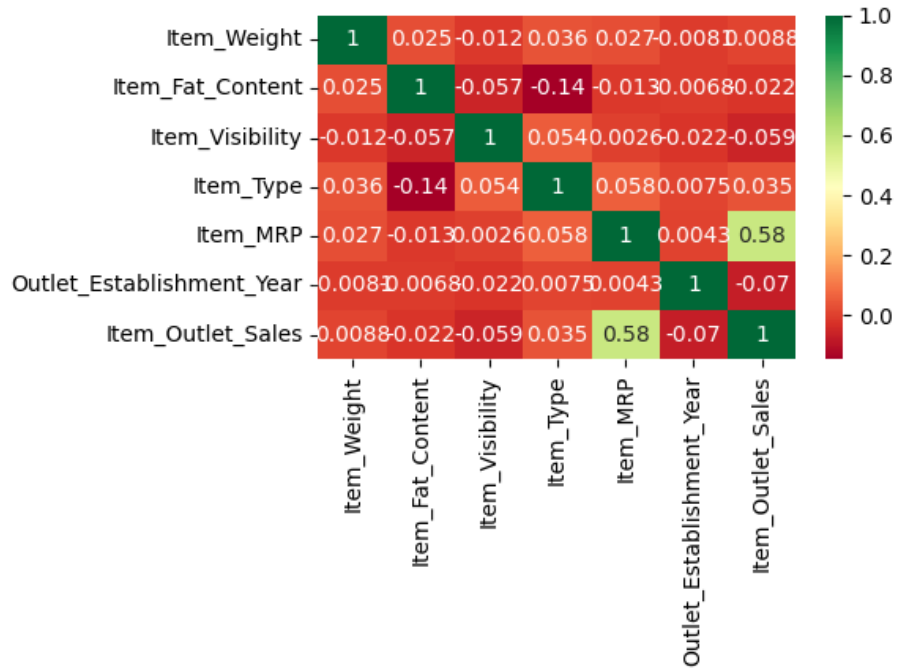


Şekil 55. “Outlet_Establishment_Year” kutu grafiği.

Adım 32: “Outlet_Establishment_Year” ve “Item_Outlet_Sales” alanları arasındaki korelasyonun incelenmesi.

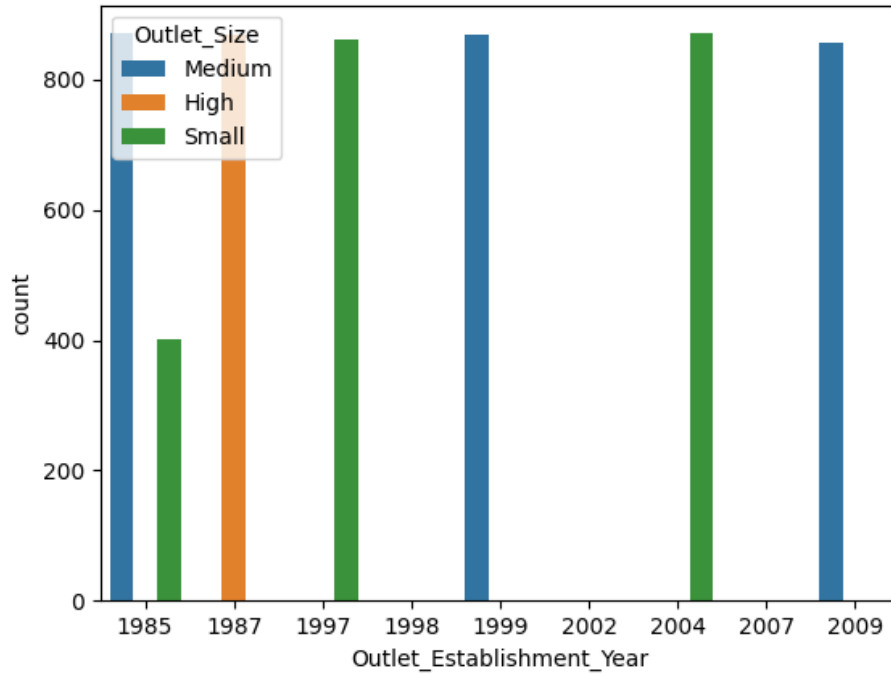
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
Co-relation of sales and mrp is : -0.06999353412667209
```

Şekil 56. “Outlet_Establishment_Year” ve “Item_Outlet_Sales” alanları arasındaki korelasyon katsayısı.



Şekil 57. Sıcaklık grafiği.

Adım 33: “Outlet_Establishment_Year” ve “Outlet_Size” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 58. “Outlet_Establishment_Year” ve “Outlet_Size” alanları ile çizilen grafik.

Grafikten anlaşılabacağı üzere zaman ilerledikçe mağazaların büyüklüğü artmıştır ve belli bir zamandan sonra sabit bir büyüklükte kalmıştır.

Adım 34: Veri setinin kontrol edilmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales	Item_Code
0	9.300000	1	-1.794598	7	249.8092	1999	Medium	Tier 1 Supermarket	Type1	3735.1380	FD
1	5.920000	0	-1.714933	3	48.2692	2009	Medium	Tier 3 Supermarket	Type2	443.4228	DR
2	17.500000	1	-1.775724	1	141.6180	1999	Medium	Tier 1 Supermarket	Type1	2097.2700	FD
6	12.050000	0	-1.894703	7	57.6588	1987	High	Tier 3 Supermarket	Type1	343.5528	FD
7	12.857645	1	-0.894593	7	107.7622	1985	Medium	Tier 3 Supermarket	Type3	4022.7636	FD
...
8518	6.865000	1	-1.245779	7	214.5218	1987	High	Tier 3 Supermarket	Type1	2778.3834	FD
8519	8.380000	0	-1.328065	4	108.1570	2002	NaN	Tier 2 Supermarket	Type1	549.2850	FD
8520	10.600000	1	-1.453627	2	85.1224	2004	Small	Tier 2 Supermarket	Type1	1193.1136	NC
8521	7.210000	0	-0.837972	7	103.1332	2009	Medium	Tier 3 Supermarket	Type2	1845.5976	FD
8522	14.800000	1	-1.347964	3	75.4670	1997	Small	Tier 1 Supermarket	Type1	765.6700	DR

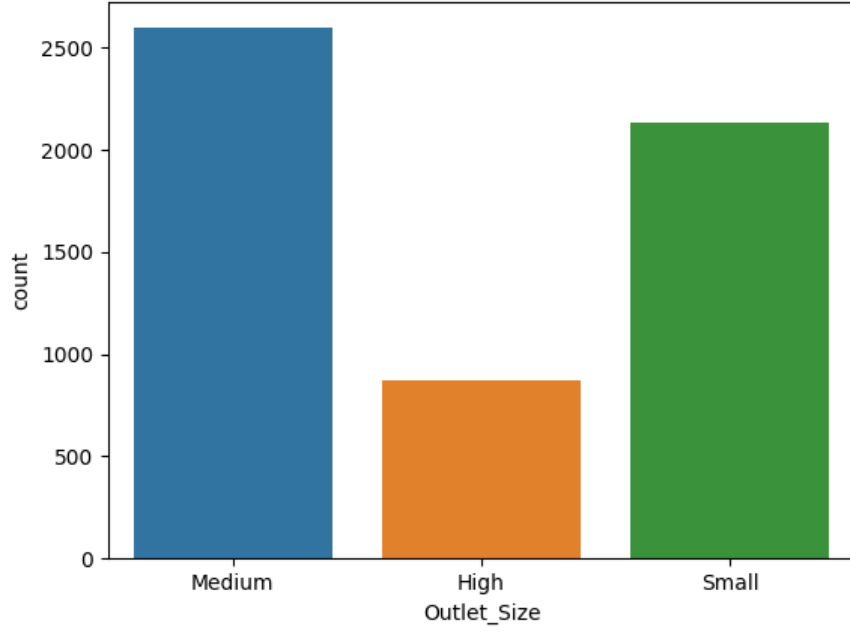
```
[7769 rows x 11 columns]
```

Şekil 59. Veri setinin son hali.

Adım 35: “Outlet_Size” alanının incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
2168
```

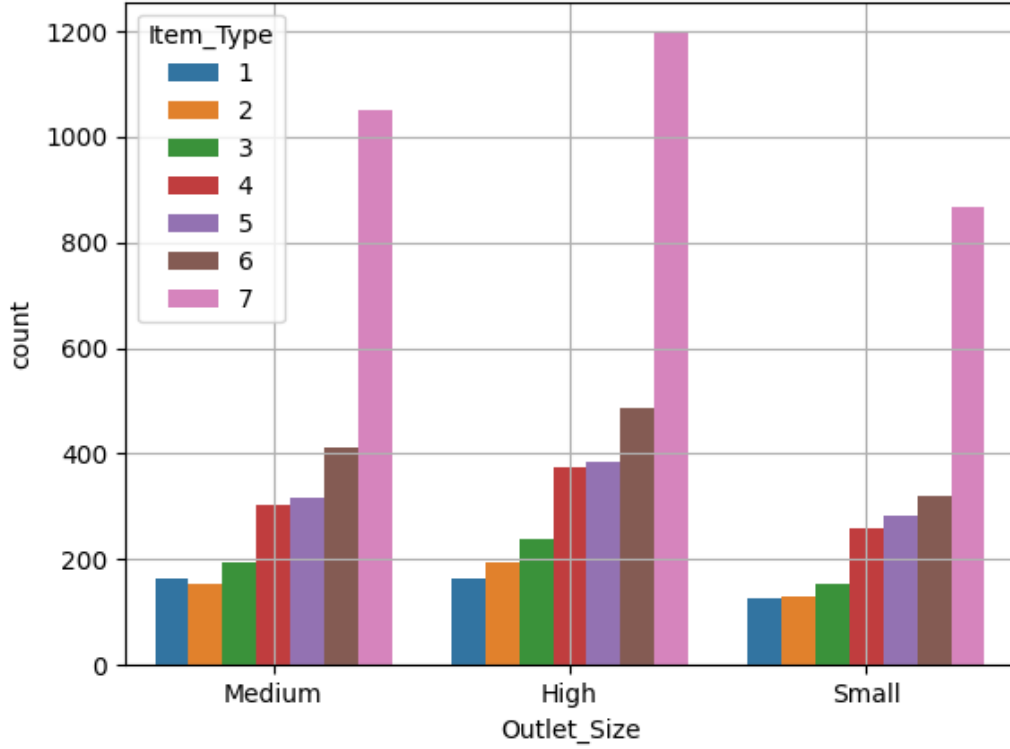
Şekil 60. “Outlet_Size” alanındaki boş değerlerin sayısı.



Şekil 61. “Outlet_Size” grafiği.

Büyük olan mağaza sayısı az olduğu için boş olan değerler “High” değeri ile doldurulur.

Adım 36: “Outlet_Size” ve “Item_Type” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 62. “Outlet_Size” ve “Item_Type” alanları ile çizilen grafik.

Grafikten büyük mağazaların, yüksek miktarda ürüne sahip olduğu anlaşılmaktadır.

Adım 37: “Outlet_Size” alanının sayısal olarak yenilenmesi ve alanlar arasındaki korelasyon ilişkisine bakılması.

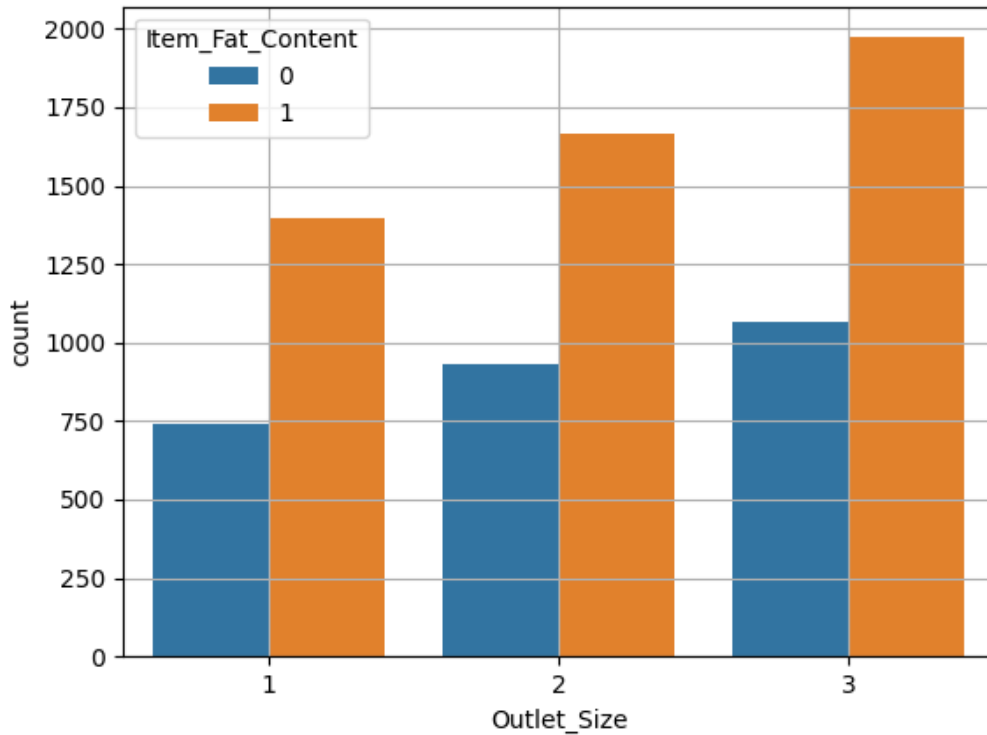
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Establishment_Year	Outlet_Size	Item_Outlet_Sales
Item_Weight	1.000000	0.025202	-0.011904	0.036343	0.027181	-0.008069	0.012255	0.008839
Item_Fat_Content	0.025202	1.000000	-0.057360	-0.144768	-0.012756	-0.006849	-0.006122	-0.021578
Item_Visibility	-0.011904	-0.057360	1.000000	0.054497	0.002599	-0.022496	-0.043187	-0.059361
Item_Type	0.036343	-0.144768	0.054497	1.000000	0.057840	0.007480	-0.001228	0.034615
Item_MRP	0.027181	-0.012756	0.002599	0.057840	1.000000	0.004344	-0.004199	0.577356
Outlet_Establishment_Year	-0.008069	-0.006849	-0.022496	0.007480	0.004344	1.000000	-0.328695	-0.069994
Outlet_Size	0.012255	-0.006122	-0.043187	-0.001228	-0.004199	-0.328695	1.000000	0.109193
Item_Outlet_Sales	0.008839	-0.021578	-0.059361	0.034615	0.577356	-0.069994	0.109193	1.000000

Şekil 63. Alanlar arasındaki yeni korelasyon ilişkisi.

Alanlar arasındaki korelasyon ilişkisini incelediğimizde satış noktasının boyutunun, ürün satışlarına %2 oranında katkı sağladığını görebiliriz.

Adım 38: “Outlet_Size” ve “Item_Fat_Content” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



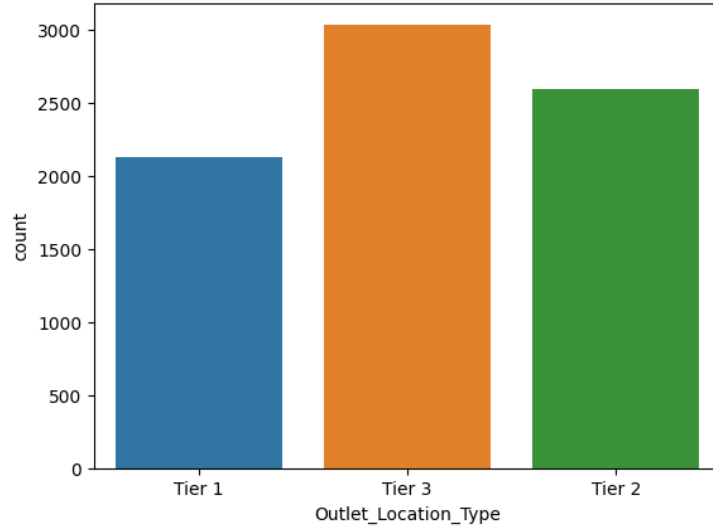
Şekil 64. “Outlet_Size” ve “Item_Fat_Content” alanları ile çizilen grafik.

Yukarıdaki grafikten yola çıkılarak mağaza boyutunun eşit oranda yağ (1) ve normal öge (0) içerdiği sonucu elde edilebilir.

Adım 39: “Outlet_Location_Type” alanının incelenmesi.

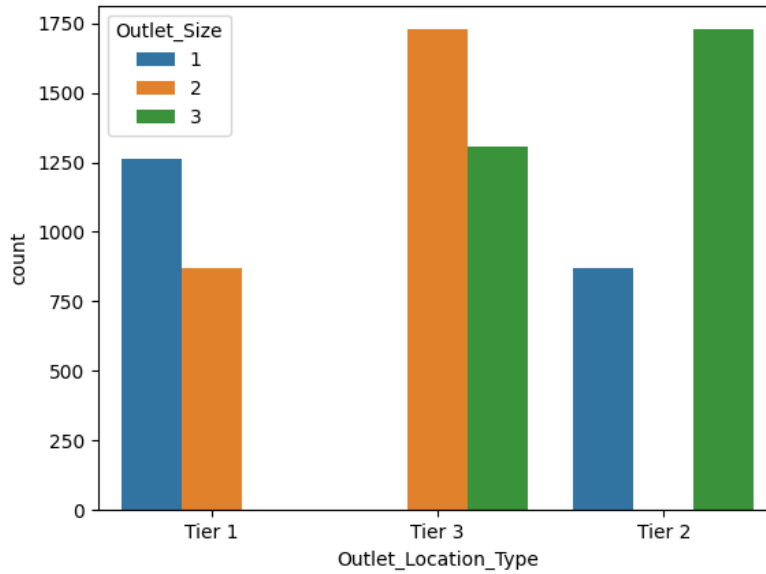
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py  
Tier 3    3037  
Tier 2    2599  
Tier 1    2133  
Name: Outlet_Location_Type, dtype: int64
```

Şekil 65. “Outlet_Location_Type” alanı.



Şekil 66. “Outlet_Location_Type” grafiği.

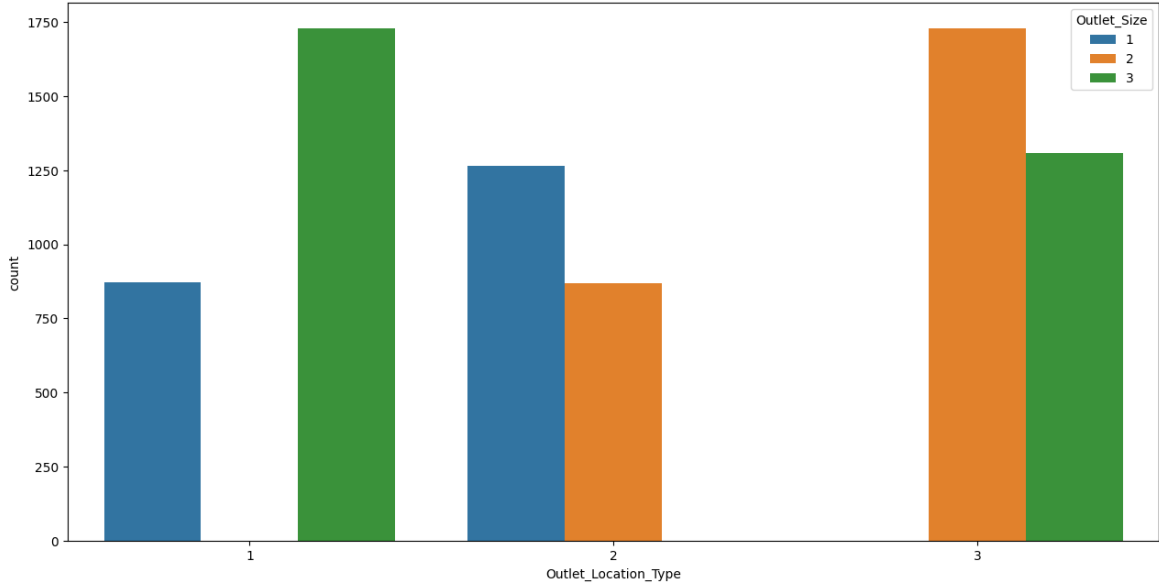
Adım 40: “Outlet_Location_Type” ve “Outlet_Size” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 67. “Outlet_Location_Type” ve “Outlet_Size” alanları ile çizilen grafik.

Grafikten büyük mağazaların, kötü derecedeki konumlarda daha fazla olduğunu görmekteyiz.

Adım 41: “Outlet_Location_Type” alanının yeniden düzenlenmesi ve Şekil 67.’de yer alan grafiğin tekrardan çizdirilmesi.



Şekil 68. “Outlet_Location_Type” ve “Outlet_Size” alanları ile çizilen grafiğin yeni hali.

Adım 42: Alanlar arasındaki korelasyon ilişkisine bakılması.

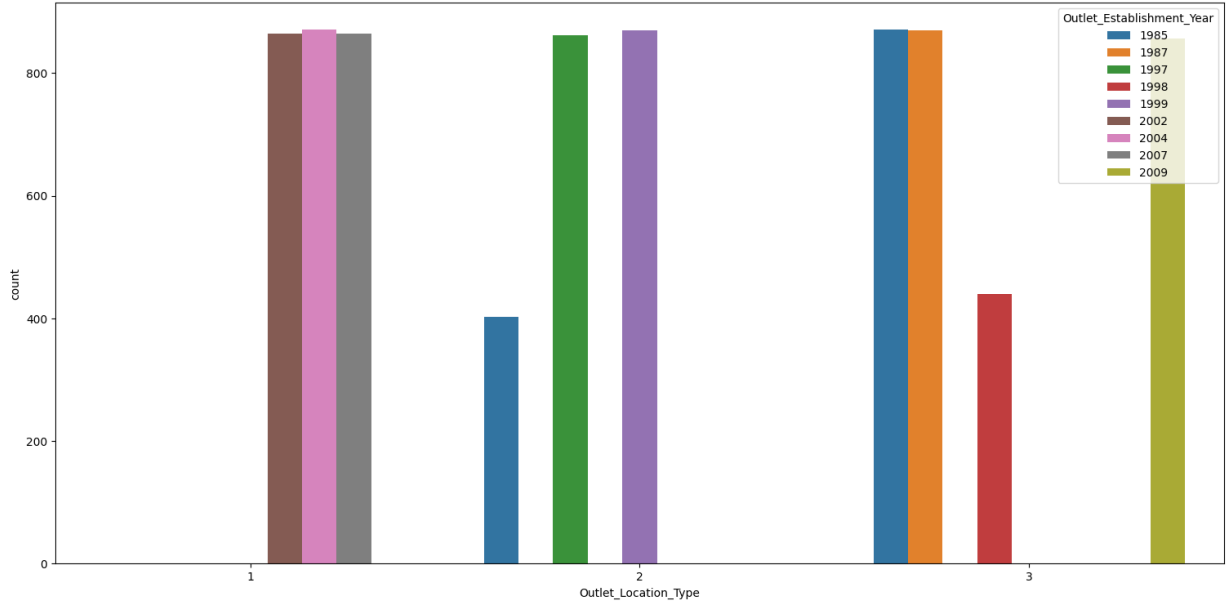
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Item_Outlet_Sales
Item_Weight	1.000000	0.025202	-0.011904	0.036343	0.027181	-0.008069	0.001209	0.012891	0.008839
Item_Fat_Content	0.025202	1.000000	-0.057360	-0.144768	-0.012756	-0.006849	-0.002195	-0.005665	-0.021578
Item_Visibility	-0.011904	-0.057360	1.000000	0.054497	0.002599	-0.022496	-0.007711	0.019084	-0.059361
Item_Type	0.036343	-0.144768	0.054497	1.000000	0.057840	0.007480	-0.003525	-0.008739	0.034615
Item_MRP	0.027181	-0.012756	0.002599	0.057840	1.000000	0.004344	-0.005944	0.003209	0.577356
Outlet_Establishment_Year	-0.008069	-0.006849	-0.022496	0.007480	0.004344	1.000000	0.048901	-0.506451	-0.069994
Outlet_Size	0.001209	-0.002195	-0.007711	-0.003525	-0.005944	0.048901	1.000000	0.075229	-0.018297
Outlet_Location_Type	0.012891	-0.005665	0.019084	-0.008739	0.003209	-0.506451	0.075229	1.000000	0.006432
Item_Outlet_Sales	0.008839	-0.021578	-0.059361	0.034615	0.577356	-0.069994	-0.018297	0.006432	1.000000

Şekil 69. Alanlar arasındaki yeni korelasyon ilişkisi.

Alanlar arasındaki korelasyon ilişkisi incelendiğinde mağaza tipinin kuruluş yılına %50, kuruluş yılının ise çıkış alanı büyüklüğüne %5 oranında bağlı olduğu görülmektedir.

Adım 43: “Outlet_Location_Type” ve “Outlet_Establishment_Year” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 70. “Outlet_Location_Type” ve “Outlet_Establishment_Year” alanları ile çizilen grafik.

Adım 44: Veri setinin incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	...	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales	Item_Code
0	9.300000	1	-1.794598	7	249.8092	...	2	2	Supermarket Type1	3735.1380	FD
1	5.920000	0	-1.714933	3	48.2692	...	2	3	Supermarket Type2	443.4228	DR
2	17.500000	1	-1.775724	1	141.6180	...	2	2	Supermarket Type1	2097.2700	FD
6	13.650000	0	-1.894793	7	57.6588	...	3	3	Supermarket Type1	343.5528	FD
7	12.857645	1	-0.894593	7	107.7622	...	2	3	Supermarket Type3	4022.7636	FD
...
8518	6.865000	1	-1.245779	7	214.5218	...	3	3	Supermarket Type1	2778.3834	FD
8519	8.380000	0	-1.328065	4	108.1570	...	3	1	Supermarket Type1	549.2850	FD
8520	10.600000	1	-1.453627	2	85.1224	...	1	1	Supermarket Type1	1193.1136	NC
8521	7.210000	0	-0.837972	7	103.1332	...	2	3	Supermarket Type2	1845.5976	FD
8522	14.800000	1	-1.347964	3	75.4670	...	1	2	Supermarket Type1	765.6700	DR

[7769 rows x 11 columns]

Şekil 71. Veri setinin son hali.

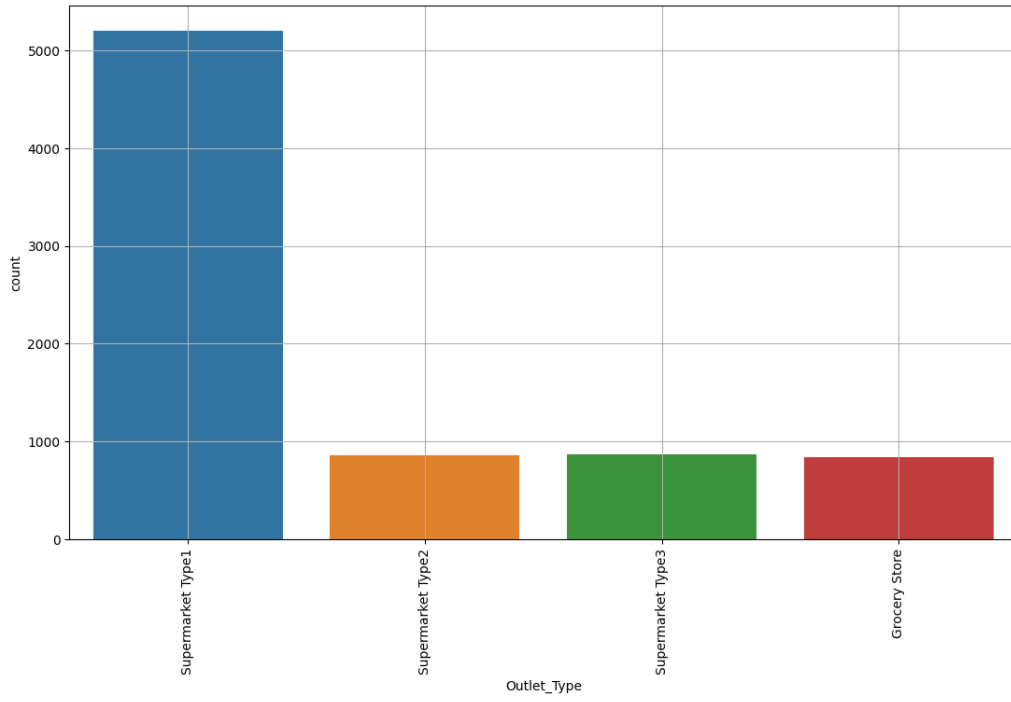
Adım 45: “Outlet_Type” alanının incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

Supermarket Type1	5199
Supermarket Type3	871
Supermarket Type2	857
Grocery Store	842

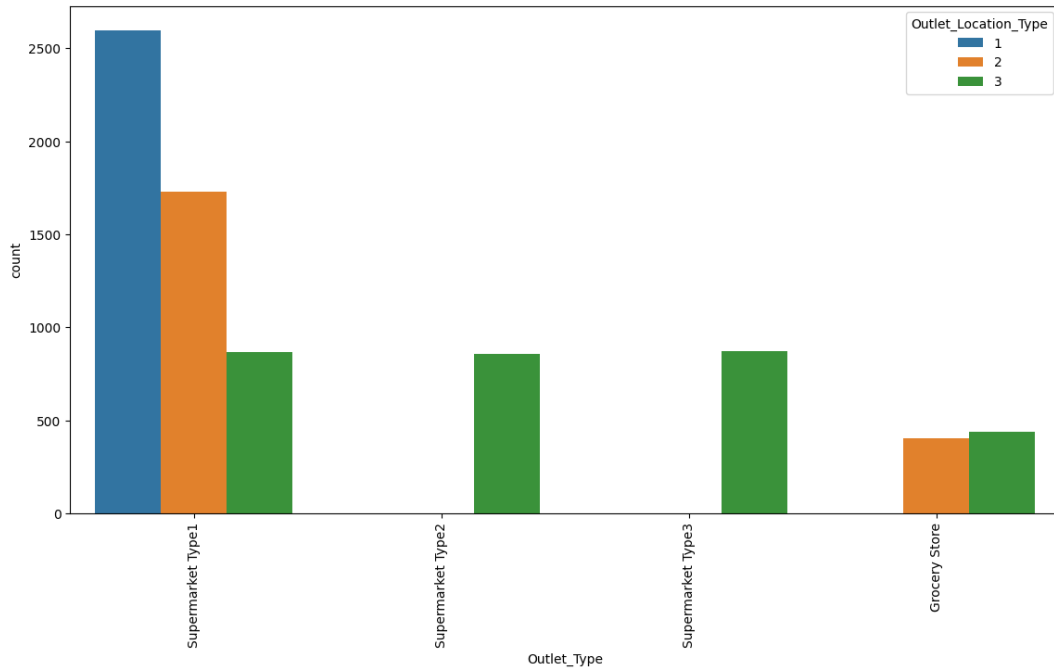
Name: Outlet_Type, dtype: int64

Şekil 72. “Outlet_Type” alanı.



Şekil 73. “Outlet_Type” grafiği.

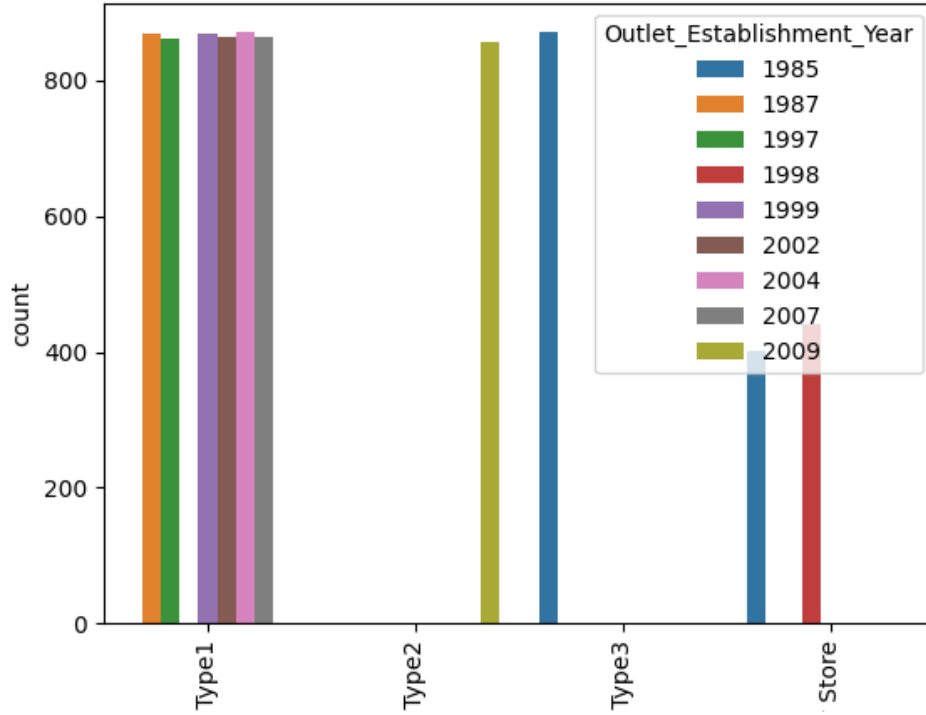
Adım 46: “Outlet_Type” ve “Outlet_Location_Type” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 74. “Outlet_Type” ve “Outlet_Location_Type” alanları ile çizilen grafik.

Grafiklerden Supermarket 1 tipinin çoğunluğu kapsadığını görebiliriz.

Adım 47: “Outlet_Type” ve “Outlet_Establishment_Year” alanları arasındaki ilişkinin incelenmesi ve grafik çıkartılması.



Şekil 75. “Outlet_Type” ve “Outlet_Establishment_Year” alanları ile çizilen grafik.

Adım 48: “Outlet_Type” alanı yeniden düzenlenmesi, veri setinin kontrol edilmesi ve alanlar arasındaki korelasyon ilişkisinin incelenmesi.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales	Item_Code
0	9.300000	1	-1.794598	7	249.8092	2	2	4	3735.1380	FD
1	5.920000	0	-1.714933	3	48.2692	2	3	3	443.4228	DR
2	17.500000	1	-1.775724	1	141.6180	2	2	4	2097.2700	FD
6	13.650000	0	-1.894793	7	57.6588	3	3	4	343.5528	FD
7	12.857645	1	-0.894593	7	107.7622	2	3	2	4022.7636	FD
...
8518	6.865000	1	-1.245779	7	214.5218	3	3	4	2778.3834	FD
8519	8.380000	0	-1.328065	4	108.1570	3	1	4	549.2850	FD
8520	10.600000	1	-1.453627	2	85.1224	1	1	4	1193.1136	NC
8521	7.210000	0	-0.837972	7	103.1332	2	3	3	1845.5976	FD
8522	14.800000	1	-1.347964	3	75.4670	1	2	4	765.6700	DR

[7769 rows x 11 columns]

Şekil 76. Veri setinin son hali.

```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
```

	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
Item_Weight	1.000000	0.025202	-0.011904	0.036343	...	0.001209	0.012891	-0.001216	0.008839
Item_Fat_Content	0.025202	1.000000	-0.057360	-0.144768	...	-0.002195	-0.005665	-0.006568	-0.021578
Item_Visibility	-0.011904	-0.057360	1.000000	0.054497	...	-0.007711	0.019084	-0.101738	-0.059361
Item_Type	0.036343	-0.144768	0.054497	1.000000	...	-0.003525	-0.008739	0.018857	0.034615
Item_MRP	0.027181	-0.012756	0.002599	0.057840	...	-0.005944	0.003209	0.002413	0.577356
Outlet_Establishment_Year	-0.008069	-0.006849	-0.022496	0.007480	...	0.048901	-0.506451	0.423132	-0.069994
Outlet_Size	0.001209	-0.002195	-0.007711	-0.003525	...	1.000000	0.075229	0.073130	-0.018297
Outlet_Location_Type	0.012891	-0.005665	0.019084	-0.008739	...	0.075229	1.000000	-0.524445	0.006432
Outlet_Type	-0.001216	-0.006568	-0.101738	0.018857	...	0.073130	-0.524445	1.000000	0.171134
Item_Outlet_Sales	0.008839	-0.021578	-0.059361	0.034615	...	-0.018297	0.006432	0.171134	1.000000

[10 rows x 10 columns]

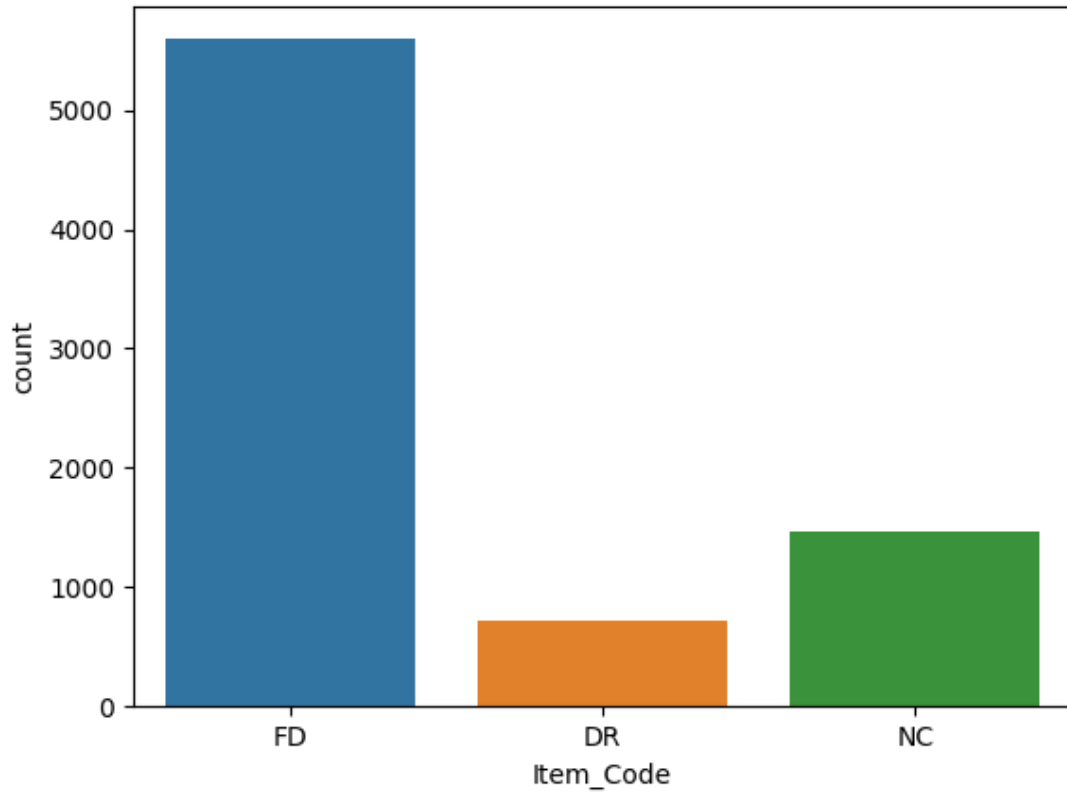
Şekil 77. Alanlar arasındaki yeni korelasyon ilişkisi.

Alanlar arasındaki korelasyon incelendiğinde mağaza tipinin mağaza lokasyonuna %50, satışlara ise %42 oranında katkı sağladığını görebiliriz.

Adım 49: “Item_Code” alanının incelenmesi.

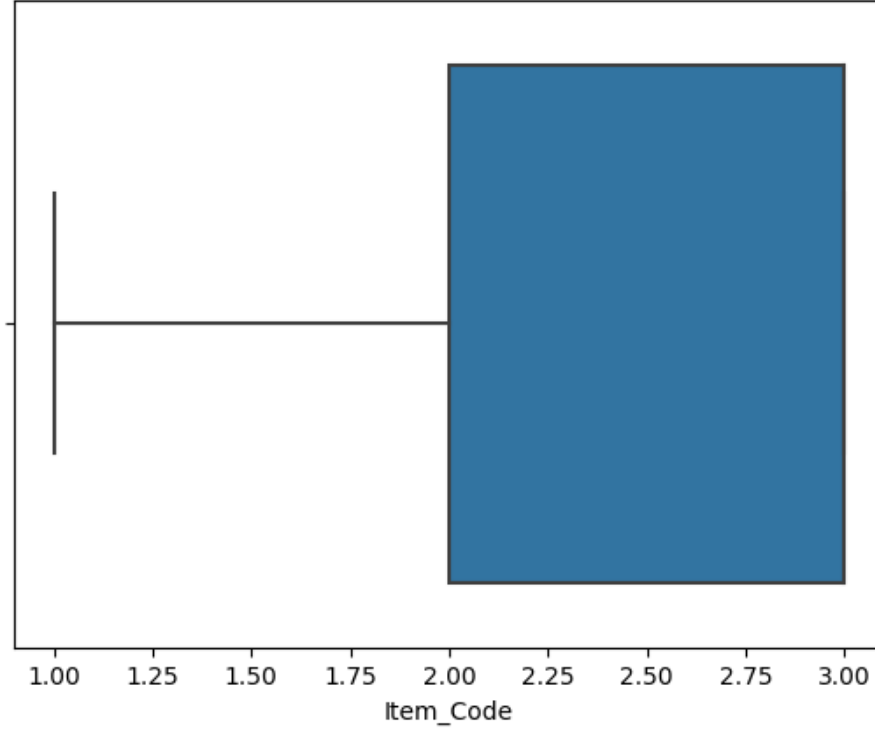
```
C:\Users\kadir\Desktop\Big Data>py bigmarket.py
0      FD
1      DR
2      FD
6      FD
7      FD
..
8518   FD
8519   FD
8520   NC
8521   FD
8522   DR
Name: Item_Code, Length: 7769, dtype: object
```

Şekil 78. “Item_Code” alanı.



Şekil 79. “Item_Code” grafiği.

Adım 50: “Item_Code” alanı yeniden düzenlenir ve kutu grafiği çıkartılır.



Şekil 80. “Item_Code” kutu grafiği.

Büyük Veri Platformu

Projemizde, büyük veri platformu olarak Python programlama dilini kullandık. Analiz işlemleri için Python dili içerisinde yer alan NumPy, Pandas, Matplotlib, Seaborn, SciPy ve Missingno kütüphanelerini kullandık.

Python

Yapısal ve fonksiyonel programlama modellerini destekleyen üst düzey ve nesneye yönelik bir programlama dilidir. Özellikleri nedeniyle Python, büyük ve karmaşık veri kümelerini hızlı ve verimli bir şekilde işleyebilir ve ilişkilendirebilir. Bu yüzden Uber, PayPal ve Facebook gibi önde giden teknoloji girişimleri, markaları, uygulamalarını oluşturmak için Python programlama dilini kullanmaktadır. Ayrıca, daha fazla geliştiricinin veri bilimi için Python’u öğrenmek için zaman ayırmasının nedeni de budur.

Veri bilimcileri büyük ve karmaşık ham veri setlerini organize etmekten, temizlemekten ve basitleştirmekten sorumludur. Ayrıca hem etkili hem de basit çözümler bulmak için verileri analiz etmekten sorumludurlar. Python gibi bir programlama dili kullanmak işlerini kolaylaştırmaya yardımcı olur, çünkü sözdizimi ve programlama gereksinimleri diğer dillerle karşılaştırıldığında kullanımı ve anlaşılması daha kolaydır.

Veri bilimcilerinin veri bilimi için Python’u öğrenmeyi tercih etmesinin bir başka nedeni, Python’un en önemli üç veri bilimi kütüphanesini (Numpy, Pandas, Matplotlib) desteklemesidir.

Biz de projemizde yukarıdaki sebeplerden dolayı Python programlama dilini tercih ettik.

a) NumPy

Büyük, çok boyutlu diziler ve matrisler için destek sağlarken, bu dizilerde çalışmak için yüksek düzeyli matematiksel işlev kolaylığı sağlar. NumPy, temelinde NumPy dizilerinden oluşmakta ve bu diziler Python’daki liste yapısına benzemektedir. Python listelerinden farklı olarak içerisinde yer alan verilerin aynı tipte olması gerekmektedir. NumPy ayrıca Python listelerine göre daha hızlı ve işlevseldir.

b) Pandas

Veri işlemesi ve analizi için Python programlama dilinde yazılmış olan bir yazılım kütüphanesidir. Bu kütüphane temel olarak zaman etiketli serileri ve sayısal tabloları işlemek için bir veri yapısı oluşturur ki bu şekilde çeşitli işlemler bu veri yapısı üzerinde gerçekleştirilebilir olur.

c) Matplotlib

Veri görselleştirmesinde kullandığımız temel Python kütüphanesidir. 2 ve 3 boyutlu çizimler yapmamızı sağlar. Matplotlib, genellikle 2 boyutlu çizimlerde kullanılır.

d) Seaborn

Python’da ilgi çekici ve bilgilendirici istatistiksel grafikler yapmak için kullanılan bir kütüphanedir. Bu kütüphane sayesinde birçok farklı grafik kolaylıkla çizilebilmektedir.

e) SciPy

Bilimsel ve matematiksel problemleri çözmek için kullanılan açık kaynaklı bir Python kütüphanesidir. Numpy uzantısı üzerine inşa edilmiştir ve kullanıcının çok çeşitli üst düzey komutlarla verileri işlemesine ve görselleştirmesine olanak tanır.

f) Missingno

Eksik veri analizi yapmaya yarayan bir Python kütüphanesidir.

Analiz Sonuçları

- MRP, tüm satışlara %57 etki etmektedir.
- Ürünlerin türü baz alındığında %14 yağlı yiyecek, %86 normal yağlı yiyecek bulunmaktadır.
- Mağazanın kuruluş yılı, %50 oranında mağazanın konumunu etkilerken %42 oranında da mağazanın türünü (dükkân, süpermarket) etkilemektedir.
- Mağaza türü, mağazanın konumuyla %53 oranında bağlantılıdır.
- Mağazanın türü, toplam satışları %17 oranında etkilemektedir.
- Büyük mağazalar daha çok ürüne sahiptir.

Daha fazla kar elde edebilmek için;

- Daha fazla mağaza açılmalı,
- Daha fazla normal yağlı yiyeceklerden satılmalı,
- Mağazanın konumu Tier2 alanında, boyutu ise orta ölçekli olmalıdır.
- Ürünlerin görünürlüğü arttırılmalıdır.
- MRP'ye daha çok önem verilmelidir.