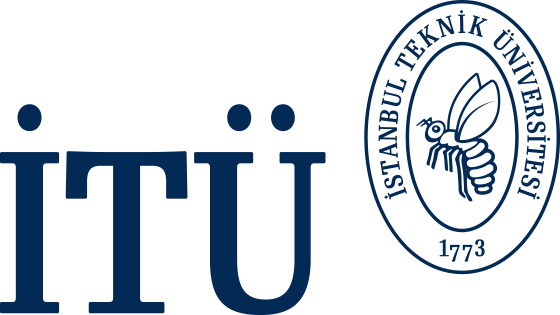


A poster of statistics and graphs

Description automatically generated with medium confidence



*BIG Mart Satış Tahmini ve İstatistiki Analizi*

**İTÜ Büyük Veri ve İş Analitiği Yüksek Lisans Programı**

***Güz 2024/2025*  
İLERİ İSTATİSTİKSEL ANALİZ**

Öğretim Üyesi:  
**Doç. Dr. Umut ASAN**

Öğrenci:  
Göksel Bilici  
528241062

A black text with a white background

Description automatically generated

1. **Giriş**

Bu projede başlangıçta, bireylerin gelir seviyelerini tahmin etmeye yönelik bir analiz gerçekleştirilmesi planlanmıştır. Bankacılık sektöründe gelir bilgisi, kredi değerlendirme, işbirliği fırsatlarını analiz etme ve yatırım kararları gibi birçok alanda hayati öneme sahiptir. Gelir düzeylerinin doğru bir şekilde tahmini, finansal risk yönetimi ve karar destek sistemlerinin etkinliğini artırma potansiyeline sahiptir. Ancak, veri gizliliği ve Kişisel Verilerin Korunması Kanunu (KVKK) gibi yasal düzenlemeler nedeniyle, ideal veri kaynaklarına erişim sınırlıdır. Bu sebeple, analiz için Amerika Birleşik Devletleri'ndeki yetişkin bireylere ait açık bir veri seti tercih edilmişti.

Başlangıç analizlerinde, bu veri seti üzerinden kapsamlı istatistiksel değerlendirmeler yapılmış ve gelir tahmini için etkili bir model geliştirilmesi hedeflenmiştir. Ancak, veri setindeki değişkenlerin çoğunlukla kategorik olması ve beklenen sonuçlarla uyumsuzluk göstermesi nedeniyle, bu yaklaşımdan vazgeçilmiştir.

Güncel durumda ise proje, BigMart tarafından 2013 yılında 10 farklı şehirdeki 1559 ürün ve mağaza satışlarına dair toplanan veriler üzerine odaklanmaktadır. Veri seti, her ürün ve mağazaya ait belirli özellikleri içermektedir. Projenin temel amacı, belirli bir mağazada satılan her bir ürünün satışlarını tahmin edebilecek bir model geliştirmek ve bu model üzerinden ürün ve mağaza özelliklerinin satışları artırmada nasıl bir etki yarattığını anlamaktır. Ayrıca, veri setinde bazı mağazalardan eksik veri içerdiği görülmüştür. Bu eksikliklerin teknik aksaklıklardan kaynaklandığı düşünülmekte olup, bu durumun uygun veri işleme yöntemleriyle ele alınması gerekmektedir.

Projenin genel perspektifi, sadece nihai model geliştirmeyi değil, bu süreci destekleyecek kapsamlı istatistiksel analizler yapmayı ve farklı veri işleme tekniklerini uygulamayı da içermektedir.

1. **Problem Tanımı ve Çalışmanın Amacı**

BigMart, mağazalarındaki satış performansını artırmak ve ürünler ile mağaza özelliklerinin satışlar üzerindeki etkilerini daha iyi anlamayı hedeflemektedir. Farklı şehirlerdeki mağazalar arasında satış dinamikleri, ürün özellikleri ve müşteri tercihleri açısından karmaşık ilişkilerin bulunması, mağaza yönetimi ve strateji geliştirme süreçlerini zorlaştırmaktadır. Ayrıca, veri setinde bazı mağazalara ait eksik verilerin bulunması, analiz süreçlerini ve tahmin doğruluğunu olumsuz etkileyen bir başka önemli faktördür.

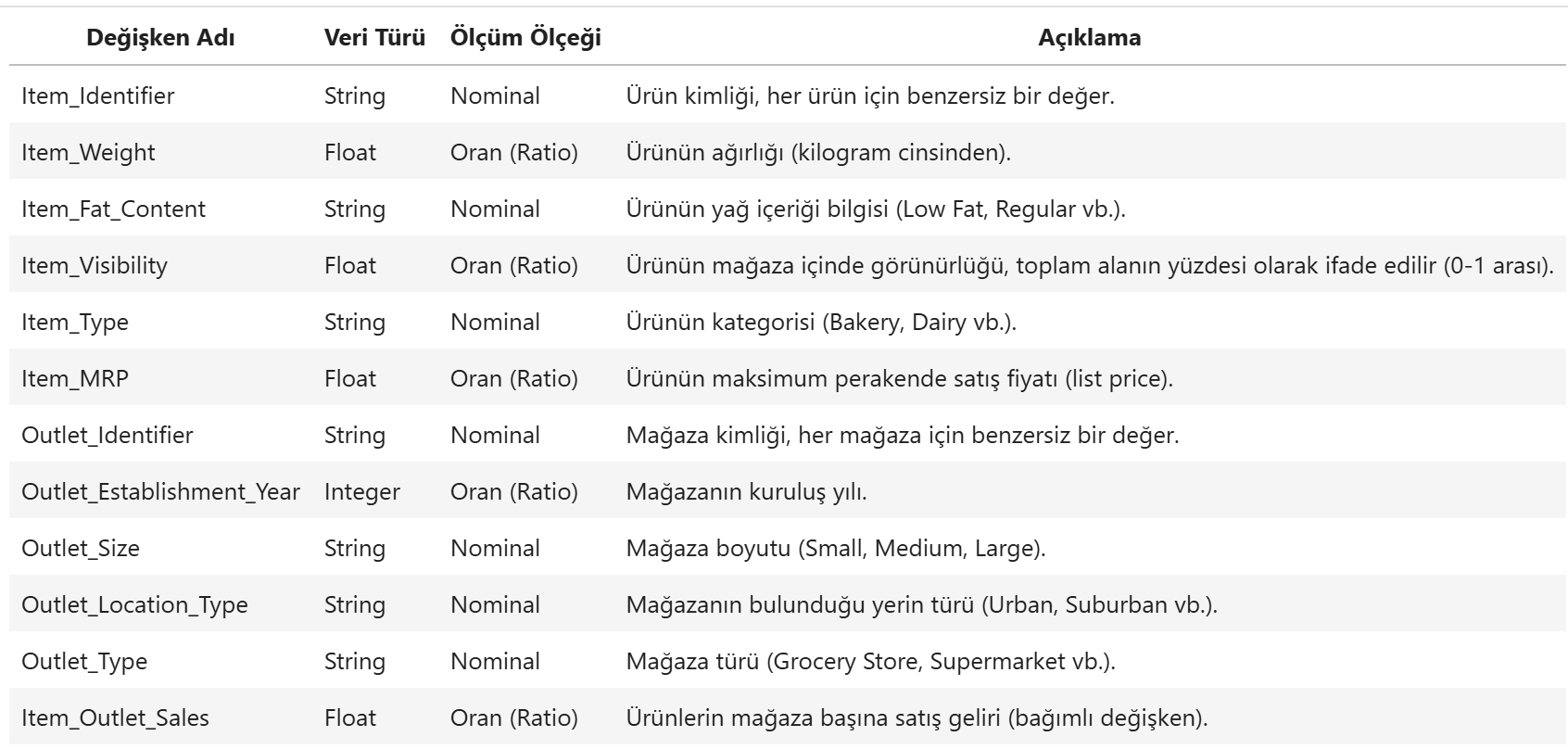
Bu bağlamda çalışmanın amacı, belirli bir mağazada satılan her bir ürünün satışlarını tahmin edebilecek bir makine öğrenimi modeli geliştirmektir. Geliştirilecek model, ürün ve mağaza özelliklerini dikkate alarak satış tahminleri yapacak ve BigMart’ın satış stratejilerini optimize etmesine yardımcı olacaktır. Model geliştirme sürecinde eksik veriler uygun yöntemlerle işlenecek, böylece eksik verilerin analiz ve tahmin süreçlerine olumsuz etkisi en aza indirgenecektir.

Çalışma, şu temel hedefler etrafında şekillenecektir. Satış tahminlerini etkileyen ürün ve mağaza özelliklerini belirlemek. Eksik verilerin etkilerini gidermek için uygun istatistiksel ve makine öğrenimi yöntemlerini uygulamak. Geliştirilen modelin performansını değerlendirmek ve optimize etmek. BigMart’ın ürün yönetimi ve satış stratejilerini daha etkili bir şekilde yönetmesine katkı sağlamak. Sonuç olarak, bu proje, hem teorik hem de uygulamalı bir yaklaşımla BigMart'ın satış performansını etkileyen temel faktörleri analiz ederek, satış tahminlerine dayalı stratejik karar süreçlerini güçlendirmeyi hedeflemektedir.

**3. Önerilen Model**

Bu çalışmada, BigMart tarafından 2013 yılına ait 1559 ürün ve 10 mağazada toplanan satış verileri analiz edilecektir. Veri seti, ürün ve mağaza özelliklerini içeren çeşitli değişkenlerden oluşmaktadır. Amaç, her bir mağazada satılan ürünlerin satışlarını tahmin etmek ve bu süreçte ürün ve mağaza özelliklerinin satışlar üzerindeki etkilerini anlamaktır. Sonrasında yapılacak işlemlerle de bu perspektiften analiz etmektir.  
**3.1. Kavramlar, Değişkenler ve Ölçekleri**

Aşağıda, çalışmada kullanılan değişkenler, türleri ve ölçüm ölçekleri verilmiştir. Veri setinde, ürünlere ve mağazalara ait çeşitli özellikler yer almakta ve Item\_Outlet\_Sales bağımlı değişken olarak tanımlanmaktadır:

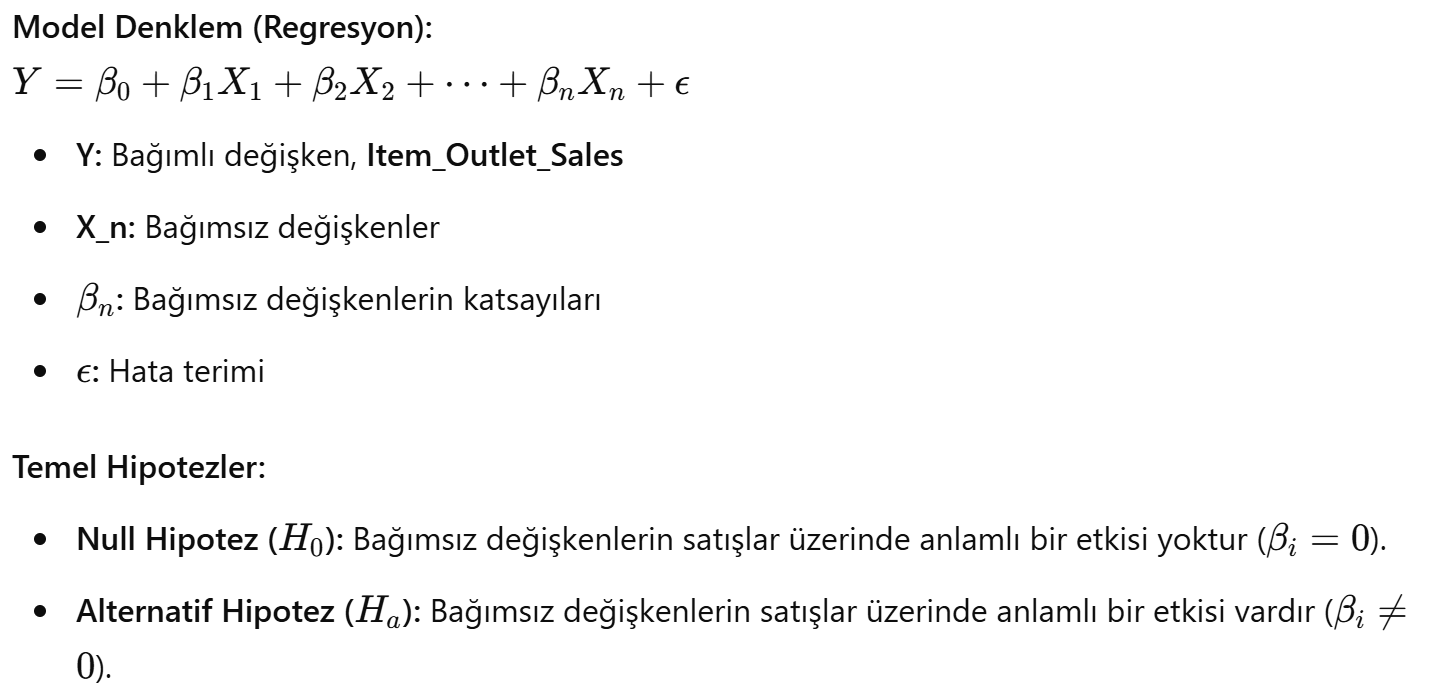
Tablo 1: Modelde kullanacak değişkenler için veri sözlüğü

**3.2. Model (Şekil ve/veya Denklem) ve Temel Hipotezler**

**Çoklu Regresyon Analizi**

Amaç:  
Belirli bir mağazada satılan ürünlerin satışlarını (Item\_Outlet\_Sales) etkileyen faktörleri belirlemek. Bu analizde, bağımlı değişken Item\_Outlet\_Sales olup, bağımsız değişkenler aşağıdaki gibidir:

* Item\_Weight
* Item\_Fat\_Content
* Item\_Visibility
* Item\_Type
* Item\_MRP
* Outlet\_Size
* Outlet\_Location\_Type
* Outlet\_Type



**Temel Hipotezler:**

* **Null Hipotez (H0​):** Bağımsız değişkenlerin satışlar üzerinde anlamlı bir etkisi yoktur (βi=0).
* **Alternatif Hipotez (H1​):** Bağımsız değişkenlerin satışlar üzerinde anlamlı bir etkisi vardır (βi≠0)

**2. İki Yönlü Varyans Analizi (Two-Way ANOVA)**

**Amaç:**  
Mağaza türü (**Outlet\_Type**) ve mağaza boyutunun (**Outlet\_Size**) satışlar üzerindeki etkisini ve bu iki faktörün etkileşimlerini değerlendirmek.

**Hipotezler:**

1. **Faktör 1: Outlet\_Type**
   * **Null Hipotez (H0​):** Mağaza türleri arasında satışlar açısından anlamlı bir fark yoktur.
   * **Alternatif Hipotez (H1​):** Mağaza türleri arasında satışlar açısından anlamlı bir fark vardır.
2. **Faktör 2: Outlet\_Size**
   * **Null Hipotez (H0​):** Mağaza boyutları arasında satışlar açısından anlamlı bir fark yoktur.
   * **Alternatif Hipotez (H1​):** Mağaza boyutları arasında satışlar açısından anlamlı bir fark vardır.
3. **Faktörlerin Etkileşimi (Interaction):**
   * **Null Hipotez (H0​):** Mağaza türü ve boyutu arasında etkileşim yoktur.
   * **Alternatif Hipotez (H1):** Mağaza türü ve boyutu arasında etkileşim vardır.

**4. Yöntem**

**4.1. Çok Değişkenli İstatistiksel Analiz Yöntemleri ve Seçilme Nedenleri**

Bu çalışmada, çok değişkenli analiz yöntemlerinden çoklu regresyon analizi ve iki yönlü varyans analizi (Two-Way ANOVA) kullanılacaktır.

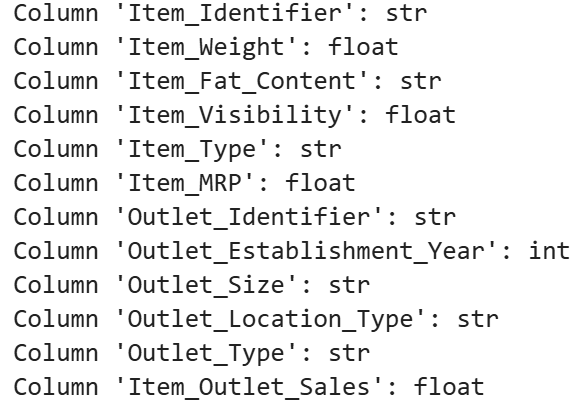
Çoklu Regresyon Analizi: Bu yöntem, bağımlı değişken olan Item\_Outlet\_Sales (outlet başına ürün satışları) üzerinde, kategorik ve nümerik bağımsız değişkenlerin etkisini değerlendirmek ve aralarındaki doğrusal ilişkiyi ölçmek için tercih edilmiştir. Regresyon analizi, veri setinde bulunan sürekli ve kategorik değişkenlerin bağımlı değişken üzerindeki etkilerinin aynı anda incelenmesini sağlar.

İki Yönlü Varyans Analizi: Veri setinde yer alan kategorik değişkenlerin (ör. Outlet\_Type ve Outlet\_Size) bağımlı değişken üzerindeki etkilerini ve bu değişkenler arasındaki etkileşimleri incelemek amacıyla kullanılacaktır. İki yönlü varyans analizi, grupların ortalamaları arasında anlamlı bir fark olup olmadığını test eder. Bu yöntem, farklı mağaza türleri ve boyutlarının satış performansı üzerindeki etkilerini karşılaştırmak için uygundur.

Ayrıca, veri setinde tek yönlü varyans analizi de anlamlı sonuçlar verebilirdi; ancak, çalışmanın gereklilikleri doğrultusunda iki yönlü varyans analizi tercih edilmiştir. İki yönlü analiz, kategorik değişkenler arasında potansiyel etkileşimlerin değerlendirilmesine olanak tanır. Ana odak, ürünlerin özellikleri ve outlet bilgileri üzerinden satış performansını değerlendirmektir.

**4.2. Örnekleme Yaklaşımı ve Veri Toplama**

A screenshot of a graph

Description automatically generated 

Resim 2: Verilere ait istatistiki veriler

Bu çalışma kapsamında kullanılan veri seti, BigMart tarafından toplanmıştır ve 2013 yılına ait 1559 ürün ve 10 mağazanın satış performansını içermektedir. Veri seti, farklı mağaza türleri, boyutları ve ürün özelliklerini kapsayacak şekilde geniş bir örneklemden oluşturulmuştur.

Veri seti, ürünler ve mağazalarla ilgili çeşitli değişkenler içermekte olup, Item\_Outlet\_Sales bağımlı değişken olarak tanımlanmıştır. Ayrıca, veri setinde bazı mağazalardan eksik veri bildirimleri olduğu belirtilmiş ve bu durumun teknik aksaklıklardan kaynaklandığı açıklanmıştır. Eksik veriler, uygun yöntemlerle ele alınacaktır.

Bu veri seti, araştırmacılar tarafından doğrudan üretilmemiştir; ancak örnekleme sürecinde dikkat edilen noktalardan biri, mağaza ve ürünlerin farklı özelliklerini temsil edecek şekilde dengeli bir dağılım sağlanmasıdır. Bu da veri setinin analiz sürecinde genelleştirilebilirlik açısından önemli bir avantaj sunmaktadır.

**4.3. Verilerin Gözden Geçirilmesi**

* Log dönüşümünden önceki dağılım, uç değerlere oldukça duyarlıydı ve sağ tarafa doğru uzun bir kuyruk içeriyordu.
* Log dönüşümü ile bu kuyruk sıkıştırılarak daha simetrik bir dağılım elde edildi. Ancak, sol tarafa bir miktar kayma mevcut, bu da verinin tamamen normalleşmediğini ancak daha dengeli hale geldiğini gösteriyor. Bu durum genelde log dönüşüm sonrası doğal bir sonuçtur.
* Item\_Outlet\_Sales kolonunda log dönüşüm uygulandıktan sonra, uç değerlerin etkisi azaltılmış ve verinin dağılımı daha dengeli hale getirilmiştir. Dönüşüm sonrası veri, daha kompakt bir aralıkta ve daha az çarpıklıkla temsil edilmektedir. Bu işlem, analiz sürecinde regresyon gibi yöntemlerin varsayımlarını daha iyi karşılamak ve modelin doğruluğunu artırmak için yapılmıştır.

**4.3.1. Dışa Düşen Veriler**

Z Dönüşümü Sonuçları:

- Item\_Weight: Hiçbir dışa düşen veri tespit edilmedi. Bu, bu kolonun genel dağılımının merkezi değerlere yakın olduğunu ve uç değer içermediğini gösteriyor.

- Item\_Visibility: 95 dışa düşen veri tespit edildi. Bu, bazı ürünlerin mağaza görünürlüğünde aşırı uç değerlere sahip olduğunu gösteriyor.

- Item\_MRP: Hiçbir dışa düşen veri bulunmadı. Ürünlerin maksimum perakende satış fiyatlarının beklenen aralıkta dağıldığını gösteriyor.

- Item\_Outlet\_Sales: 90 dışa düşen veri tespit edildi. Bu, bazı ürünlerin satışlarının çok yüksek değerlere ulaştığını (örneğin, 7968.29 gibi) ve uç değer oluşturduğunu gösteriyor.

IQR Sonuçları:

- Item\_Weight: Z dönüşümüne benzer şekilde hiçbir dışa düşen veri tespit edilmedi.

- Item\_Visibility: 144 dışa düşen veri tespit edildi. Bu, Z dönüşümüne kıyasla daha fazla dışa düşen veri bulunduğunu gösteriyor, çünkü IQR yöntemi genellikle daha geniş bir veri aralığını kapsar.

- Item\_MRP: Hiçbir dışa düşen veri tespit edilmedi. Z dönüşümü ile tutarlı.

Item\_Outlet\_Sales: 186 dışa düşen veri tespit edildi. Z dönüşümüne kıyasla çok daha fazla dışa düşen veri bulundu. Bu durum, satış değerlerinin geniş bir varyansa sahip olabileceğini gösteriyor.

Analiz ve Öneriler

- Item\_Visibility: Hem Z dönüşümü hem de IQR yöntemi bu kolonda dışa düşen veriler olduğunu tespit etti. Ancak IQR yöntemi daha fazla dışa düşen veri buldu. İki yöntem arasındaki fark, Z dönüşümünün standart sapma tabanlı olması, IQR'nin ise çeyrekler arası mesafeyi kullanmasından kaynaklanıyor. Burada, dışa düşen verilerin işlemesine gerek olup olmadığını belirlemek için bu verilerin analize etkisini değerlendirebiliriz.

- Item\_Outlet\_Sales: Z dönüşümü ve IQR yöntemi bu kolonda da tutarlı şekilde dışa düşen veriler tespit etti. Ancak IQR yöntemi daha fazla dışa düşen veri buldu. Bu, satış değerlerinin geniş bir aralıkta değiştiğini ve bu uç değerlerin analize etkili olabileceğini gösteriyor.

**4.3.2. Eksik Veriler ve Tamamlanması**

Veri setindeki eksik değerler, analiz sürecinin doğruluğunu ve modelleme aşamasını olumsuz etkilememesi için uygun yöntemlerle doldurulmuştur. Sürekli bir değişken olan Item\_Weight kolonu, veri setinin genel eğilimini korumak amacıyla kolonun ortalama değeri kullanılarak tamamlanmıştır. Öte yandan, kategorik bir değişken olan Outlet\_Size kolonu, veri setindeki mağaza türlerine göre en sık görülen değer (mod) kullanılarak doldurulmuştur. Bu yöntemler, eksik verilerin neden olabileceği önyargıları azaltırken, veri setinin analize uygun hale getirilmesini sağlamıştır.

**4.4. Varsayımların Gözden Geçirilmesi**

**4.4.1. Normal Dağılıma Uygunluk**

Bu adımda Item\_Outlet\_Sales kolonunun (log dönüşümü yapılmış haliyle) normal dağılıma uygun olup olmadığını test edeceğiz. Şu testleri gerçekleştireceğiz:

- Kolmogorov-Smirnov (KS) Testi

- Shapiro-Wilk Testi

- Jarque-Bera (JB) Testi

- Q-Q Plot Görselleştirmesi

Çıktı Sanucu:

KS Testi: Test Statistiği = 0.999, P-değeri = 0.0

- P-değerinin 0 olması, log dönüşümlü verinin normal dağılımdan önemli ölçüde sapma gösterdiğini ifade eder

Shapiro-Wilk Testi: Test Statistiği = 0.947, P-değeri = 1.0149037533873771e-47

- Shapiro-Wilk testinde P-değerinin çok düşük (0.05'ten küçük) olması, verinin normal dağılıma uygun olmadığını gösterir

Jarque-Bera Testi: Test Statistiği = 1206.239, P-değeri = 1.1707428026972933e-262

- Çarpıklık ve basıklık açısından log dönüşümlü verinin normal dağılımdan oldukça uzak olduğunu gösterir.

A graph with a red line

Description automatically generated

Resim 1: Q-Q Plot'ta verinin çeyrekleri teorik normal dağılım çizgisinden (kırmızı çizgi) sapmalar göstermektedir

**4.4.3. Yöntemlere Özel Diğer Varsayımlar**

**Varyans Homojenliği Testleri**

Levene Testi:

- Test Statistiği = 139.254, P-değeri = 4.376e-88

- P-değeri 0.05’ten çok küçük olduğu için, varyansların homojen olduğu varsayımı reddedilir.

Bartlett Testi:

- Test Statistiği = 500.176, P-değeri = 4.369e-108

- Benzer şekilde, P-değerinin 0.05’ten küçük olması, gruplar arasında varyans eşitliği olmadığını gösterir.

Varyans homojenliği sağlanmadığı için, ANOVA gibi varyans homojenliğine dayalı yöntemler doğru sonuçlar vermeyebilir. Regresyon ile devam.

**Multicollinearity Analizi**

Korelasyon matrisine göre, bağımsız değişkenler arasındaki ilişkiyi inceleyebiliriz. Aşağıdaki noktalar dikkatimizi çekiyor:

- Outlet\_Size\_Medium ve Outlet\_Size\_Small:

- Korelasyon: -0.79 (çok güçlü negatif korelasyon).

- Bu durum, multicollinearity işareti. İkisi aynı bilgiyi fazlasıyla temsil ediyor olabilir, bu yüzden yalnızca birini modele dahil etmeyi düşünebiliriz.

- Outlet\_Type\_Supermarket Type1 ve Outlet\_Type\_Supermarket Type2:

- Korelasyon: -0.48 (orta düzey negatif korelasyon).

- Çok güçlü bir multicollinearity olmasa da, dikkat edilmesi gereken bir durum.

- Diğer Değişkenler:

- Numerik değişkenler arasında anlamlı bir korelasyon görünmüyor (Item\_Weight, Item\_Visibility, Item\_MRP arasında düşük korelasyonlar var).

- Outlet türleri ve Outlet konumları arasında ise bazı kategorik bağımlılıklar bulunuyor (ör. Outlet\_Size\_Medium ile Outlet\_Type\_Supermarket Type3: 0.50 korelasyon).

**Performans Metrikleri**

- MSE (Mean Squared Error): 43.84

- Modelin hata oranını gösterir. Daha düşük bir MSE daha iyi bir performans anlamına gelir. Şu an için bu değer makul görünüyor, ancak daha fazla iyileştirme yapılabilir.

- MAE (Mean Absolute Error): 5.18

- Gerçek ve tahmin edilen değerler arasındaki ortalama farktır. Hata metriği olarak daha doğrudan bir anlam taşır.

- R-squared ve Adj. R-squared:

- R2 ve Adjusted R2 ikisi de 0.68: Model bağımlı değişkenin %68’ini açıklıyor. Model performansı için iyi bir değer

**5. Analizler**

Regresyon modeli %68 oranında açıklama gücü sunmuş, MSE 43.84 ve MAE 5.18 olarak bulunmuştur.

Outlet türleri ve Item\_MRP (maksimum perakende fiyat) değişkenlerinin anlamlı etkileri tespit edilmiştir.

**6. Bulgular ve Değerlendirme**

Model Performansı

* R-squared (R²):
  + Modelimiz, bağımlı değişken olan satışların %68’ini açıklayabilmektedir. Bu, modelin veri üzerinde makul bir açıklama gücüne sahip olduğunu göstermektedir.
  + Ancak, geri kalan %32’lik varyans, modele dahil edilmemiş faktörlerden kaynaklanıyor olabilir.
* MSE (Mean Squared Error):
  + Modelin MSE değeri 43.84 olarak hesaplanmıştır. Bu, tahmin edilen ve gerçek değerler arasındaki ortalama karesel farkın büyüklüğünü gösterir.
  + Daha iyi bir model için bu hata oranı azaltılabilir.
* MAE (Mean Absolute Error):
  + Ortalama mutlak hata 5.18 olarak bulunmuştur. Tahmin edilen değerlerin, gerçek değerlere ortalama 5.18 birimlik bir farkla yaklaştığını göstermektedir.

Anlamlı Değişkenler Modelin bağımsız değişkenlerinin etkileri incelendiğinde:

* Item\_MRP (Maksimum Perakende Fiyat):
  + p-değeri <0.05, katsayı 0.1052: Ürünün perakende fiyatı arttıkça, satışların anlamlı bir şekilde arttığı gözlemlenmiştir.
  + Bu, fiyatlandırmanın satışlar üzerinde güçlü bir etkisi olduğunu göstermektedir.
* Outlet Türleri (Supermarket Type1, Type2, Type3):
  + Outlet türlerinin satışlar üzerinde anlamlı etkileri bulunmuştur. Özellikle Supermarket Type3, en yüksek pozitif etkilerden birini göstermektedir (katsayı 26.85).

Anlamlı Olmayan Değişkenler Aşağıdaki değişkenler, ppp-değeri >0.05>0.05>0.05 olduğu için satışlar üzerinde anlamlı bir etkiye sahip değildir:

* Item\_Weight (Ürün Ağırlığı): Satışları anlamlı şekilde etkilemediği görülmüştür.
* Item\_Visibility (Ürün Görünürlüğü): Mağaza içinde ürünün görünürlüğünün, satışlar üzerinde anlamlı bir etkisi yoktur.
* Outlet\_Size: Outlet boyutları (Medium ve Small) anlamlı bir fark yaratmamıştır.
* Outlet Konum Türleri (Tier 2 ve Tier 3): Outlet’in bulunduğu şehir kategorilerinin de satışlar üzerinde anlamlı bir etkisi olmadığı görülmektedir.

Multicollinearity Değerlendirmesi

* Outlet\_Size\_Medium ve Outlet\_Size\_Small arasında çok güçlü bir negatif korelasyon (−0.79) bulunmuştur. Bu durum, çoklu doğrusal bağlantı problemlerine yol açabilir.
* Bu değişkenlerden yalnızca birinin modele dahil edilmesi, modelin daha stabil hale gelmesini sağlayabilir.

Modelin Güçlü ve Zayıf Yönleri Güçlü Yönler:

* Model, satışları makul bir şekilde açıklamakta ve fiyatlandırma gibi önemli değişkenlerin etkilerini net bir şekilde ortaya koymaktadır.
* R-squared değeri (%68), modeli pratik uygulamalar için uygun bir seviyede göstermektedir. Zayıf Yönler:
* Multicollinearity (çoklu doğrusal bağlantı) problemi, bazı değişkenlerin çıkarılmasını gerektirebilir.
* Model, R^2 = 0.68 ile iyi bir açıklama oranına sahip olsa da, %32’lik açıklanamayan varyans modelin geliştirilmesi gerektiğini göstermektedir.

**7. Sonuç ve Öneriler**

Sonuç Bu çalışmada, BigMart veri seti kullanılarak mağaza ve ürün özelliklerinin satışlar üzerindeki etkileri analiz edilmiştir. Çoklu doğrusal regresyon modeli, satışları %68 oranında açıklamış ve satışların en güçlü belirleyicilerinin maksimum perakende fiyat (Item\_MRP) ve outlet türleri (Supermarket Type3 gibi) olduğunu göstermiştir. Aşağıda bu çalışmanın temel bulguları ve çıkarımları özetlenmiştir:

* Fiyatlandırma Stratejileri:
  + Item\_MRP (maksimum perakende fiyat), satışlar üzerinde pozitif ve anlamlı bir etkiye sahiptir. Ürün fiyatları arttıkça satışlarda da artış gözlemlenmiştir. Bu durum, fiyatlandırmanın müşteri tercihleri üzerindeki etkisini göstermektedir.
* Outlet Türleri:
  + Outlet türleri arasında Supermarket Type3, satışlar üzerinde en yüksek pozitif etkiye sahiptir. Bu outlet türünün diğer türlere göre daha yüksek performans göstermesi, operasyonel stratejilerde bu tür mağazalara odaklanılmasını gerektirebilir.
* Diğer Değişkenler:
  + Ürün ağırlığı (Item\_Weight) ve ürün görünürlüğü (Item\_Visibility) gibi değişkenlerin satışlar üzerinde anlamlı bir etkisi bulunmamıştır. Bu, operasyonel kararların bu değişkenlere bağlı olmaması gerektiğini göstermektedir.
* Multicollinearity Sorunu:
  + Bazı bağımsız değişkenler arasında güçlü korelasyonlar tespit edilmiştir (ör. Outlet\_Size\_Medium ve Outlet\_Size\_Small). Bu durum modelin istikrarını etkileyebilir ve gelecekte daha sade bir değişken seçimi gerektirebilir.

**8. Kaynaklar**

*BigMart Sales Dataset, Kaggle.*

*Montgomery, D.C., et al. Introduction to Linear Regression Analysis.*

*Seabold, S., & Perktold, J. Statsmodels Documentation.*

**9. Ekler**

**A screenshot of a computer

Description automatically generated**

**Ek 1: Ana Regresyon Sonuçları**

**A graph of a graph

Description automatically generated with medium confidence**

**Ek 2: Box-Cox Öncesi Log manipülasyonluı Satış Dağılımı**

**A graph of a number of blue bars

Description automatically generated with medium confidence**

**Ek 3: Box-Cox Manipülasyonu Sonrası Satış Dağılımı**

Önceden sola yatık bir veri idi

**A graph with a line going up

Description automatically generated**

**Ek 4: Box-Cox Manipülasyonu Sonrası Q-Q grafiği**

A screenshot of a computer

Description automatically generated

**Ek 5: Correlation Matrix that shows some important multicollinearities**