



DÜZCE
ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

2023-2024 AKADEMİK YILI

BAHAR DÖNEMİ

BİLGİ MÜHENDİSLİĞİ VE BÜYÜK VERİYE GİRİŞ

Ders Sorumlusu:

Unvan. Ahmet ALBAYRAK

KAYSERİ İLİ KAZA VERİ SETİ ANALİZ RAPORU

Hazırlayan:

Mücahit ÖLEZ

Öğrenci No:

201001005

GİRİŞ

Trafik kazaları, dünya genelinde her sene yüzbinlerce kişinin yaşamını ve ekonomisini etkileyip zarara uğratan önemli bir unsurdur. Günümüzde gün geçtikçe artan araç sayısı trafik yoğunluğunu arttırmakta olup bu durum da trafik kazalarının sayısının artmasına sebebiyet vermektedir. Trafik kazaları sadece kazalara karışan kişilerin değil toplumun da sağlık, güvenlik ve ekonomik düzeyine zarar verme ihtimali barındırmaktadır. Bu sebeple, trafik kazalarına ait verilerin incelenip analizinin gerçekleştirilmesi alınabilecek önlemler için büyük önem taşımaktadır.

Bu raporda ise Kayseri il sınırları içerisinde son 10 yılda gerçekleşen kazalar ile ilgili oluşturulmuş veri setinin analizi yapılmaktadır. Veri seti çeşitli faktörlerin trafik kazalarının gerçekleşmesine olan etkisini incelemek için kullanılmaktadır. Raporun amacı veri seti üzerinde gerekli analiz işlemlerinin uygulanarak trafik kazalarının sıklık, dağılım ve trendleri gibi dağılımlarını anlayıp, muhtemel risk faktörlerini belirleyip gelecekteki gerçekleşebilecek trafik kazalarına dair olumsuzlukları belirlemektir.

Kayseri il sınırları içerisinde son on yılda ölümlü ve yaralanmalı olmak üzere toplam 12804 kaza gerçekleştirilmiştir. Gerçekleşen bu kazalar ise veri setine toplam 24 farklı niteliğe dayalı olarak kaydedilmiştir. Gerçekleşen 12804 kazanın 12746 adedi yaralanmalı kaza olarak kaydedilmişken kalan 58 adedi de ölümlü kaza olarak kaydedilmiştir. Gerçekleşen 12746 yaralanmalı kazanın 4451'inde bir kişiden fazla kişi yaralanırken, ölümlü gerçekleşen 58 kazanın da 8 tanesinde de bir kişiden fazla kişi ölmüştür.

Veri setinde bulunan 24 sütunun 11 tanesi “object” türünde veri barındırırken 13 tanesi de “int64” tipinde veri barındırmaktadır. Veri setinin büyüklüğü ise 2.3+ MB boyutundadır. Veri setindeki sütunların ismi, tipi, içerdiği nitelik değer sayısı ve içerdiği veri sayısı aşağıdaki tabloda verilmiştir.

Tablo 1 Veri Seti Nitelikleri Hakkında Genel Bilgi

Sütün İsmi	Veri Tipi	Nitelik Değer Sayısı	İçerdiği Veri Sayısı
KAZA TURU	object	2	12804
KGM ciddiyeet seviyesi	int64	26	12804
Kaza Ciddiyet Seviyesi	int64	6	12804
AY	int64	12	12804
MEVSİM	object	4	12804
KAZA SAAT	int64	26	12804
MAHALLE	object	116	12804
MAHALLE yoğunluęu	int64	4	12804
YOL	object	1498	12804
OLU SAYISI	int64	6	12804
YARALI SAYISI	int64	26	12804
TASIT SAYISI	int64	10	12804
SURUCU KUSURU	object	78	12804
OLUS NEDENİ	object	15	12804
KAPLAMA CINSİ	object	6	12804
SERİT SAYISI	int64	11	12804
AYDINLATMA	object	3	12804
Unnamed: 17	int64	4	12804
Yasal Hız Limit	int64	11	12804
ARAC CINSİ	object	19	12804
ARAC MODELİ	int64	56	12804
SURUCU CİNSİYET	object	2	12804
SURUCU EĞİTİM DURUMU	object	12	12804
EHLİYET YILI	int64	55	12804

İÇİNDEKİLER DİZİNİ

GİRİŞ	1
TABLO DİZİNİ	5
VERİ SETİ ANALİZİ	6
KAZA VERİLERİNİN ANALİZİ	6
2.1 Keşifsel Veri Analizi	6
2.1.1 Tek değişkenli analiz.....	8
2.1.2 Sayım grafiği	13
2.1.3 Çok değişkenli analiz	14
2.1.3 İki değişkenli analiz.....	15
SONUÇ	18

TABLO DİZİNİ

Tablo 1 Veri Seti Nitelikleri Hakkında Genel Bilgi.....	3
Tablo 2 Veri Setinin İstatistiksel Değerleri.....	7
Tablo 3 "KGM ciddiye seviyesi" Tek Değişkenli Analiz Grafiği.....	8
Tablo 4 "Kaza Ciddiyet Seviyesi" Tek Değişkenli Analiz Grafiği	9
Tablo 5 "AY" Tek Değişkenli Analiz Grafiği.....	9
Tablo 6 "KAZA SAAT" Tek Değişkenli Analiz Grafiği	9
Tablo 7 "MAHALLE yoğunluğu" Tek Değişkenli Analiz Grafiği	9
Tablo 8 "OLU SAYISI" Tek Değişkenli Analiz Grafiği.....	10
Tablo 9 "YARALI SAYISI" Tek Değişkenli Analiz Grafiği.....	10
Tablo 10 "TASIT SAYISI" Tek Değişkenli Analiz Grafiği.....	10
Tablo 11 "SERIT SAYISI" Tek Değişkenli Analiz Grafiği.....	10
Tablo 12 "Yasal Hız Limit" Tek Değişkenli Analiz Grafiği.....	11
Tablo 13 "ARAC MODELİ" Tek Değişkenli Analiz Grafiği	11
Tablo 14 "EHLİYET YILI" Tek Değişkenli Analiz Grafiği.....	11
Tablo 15 "Araba_Yaşı" Tek Değişkenli Analiz Grafiği.....	11
Tablo 16 "KAPLAMA CINSİ" ve "SURUCU KUSURU" Niteliklerinin Sayım Grafiği.....	13
Tablo 17 "MEVSİM" ve "ARAC CINSİ" Niteliklerinin Sayım Grafiği.....	13
Tablo 18 "SURUCU CINSİYET" Niteliğinin Sayım Grafiği	13
Tablo 19 Niteliklerin Çok Değişkenli Analiz Matrisi	14
Tablo 20 "OLU SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği	15
Tablo 21 "YARALI SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği	15
Tablo 22 "KAPLAMA CINSİ" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği	16
Tablo 23 "SERIT SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği	16
Tablo 24 "MAHALLE yoğunluğu" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği.....	16

VERİ SETİ ANALİZİ

KAZA VERİLERİNİN ANALİZİ

2.1 Keşifsel Veri Analizi

Keşifsel veri analizi (Exploratory Data Analysis), bir veri setini anlamak, özelliklerini keşfetmek ve içindeki desenleri ortaya çıkarmak için kullanılan bir veri analizi yöntemidir. Genellikle veri bilimi ve istatistik alanlarında kullanılan bir ilk adımdır ve veri setinin derinlemesine incelenmesine dayanır. Keşifsel veri analizi (KVA) adımlarını uygulamak veriyi anlama, ilişkileri ortaya çıkarma, aykırı değerleri belirleme, modelleme ve raporlama gibi bir çok görev ve iş için bizlere imkan ve kolaylık sağlar.

KVA birkaç adımdan oluşan standart prosedürleri olan bir süreçtir. Bu çalışmada ilk olarak tek değişkenli analiz yapılarak keşifsel veri analizine başlanmıştır. Ancak tek değişkenli analiz bölümüne geçmeden önce veri seti ile ilgili son düzenlemeler ve hesaplamaları yapacağız.

Veri setinde analiz gerçekleştirilmeden önce veri bütünlüğünü sağlamak için “Unnamed: 17” isimli sütun silinmiştir. “ARAC MODELİ” isimli sütun kullanılarak da “Araba_Yaşı” isimli sütun oluşturulmuştur. “Araba_Yaşı” sütunundaki değerler “Arac MODELİ” isimli sütundaki mevcut değerlerin 2024’den çıkartılması ile bulunmuştur. Detaylı incelemeden önce son olarak da veri setinin istatistiksel değerleri hesaplanmıştır. Veri setinin ortalama, standart sapma, medyan, mod vb. değerleri aşağıdaki tablo da bulunmaktadır.

Tablo 2 Veri Setinin İstatistiksel Değerleri

Sütun İsmi	Count (adet)	mean (ortalama)	std(standart sapma)	min	25%	50%	75%	max
KGM ciddiye seviyesi	12804.0	5.079663	4.566233	3.0	3.0	3.0	6.0	180.0
Kaza Ciddiyet Seviyesi	12804.0	0.226257	0.642644	0.0	0.0	0.0	0.0	5.0
AY	12804.0	6.741721	3.254241	1.0	4.0	7.0	9.0	12.0
KAZA SAAT	12804.0	14.176351	5.530597	0.0	11.0	15.0	18.0	45.0
MAHALLE yoğunluğu	12804.0	2.435567	1.044351	1.0	2.0	2.0	3.0	4.0
OLU SAYISI	12804.0	0.005936	0.116427	0.0	0.0	0.0	0.0	8.0
YARALI SAYISI	12804.0	1.675414	1.436218	0.0	1.0	1.0	2.0	36.0
TASIT SAYISI	12804.0	1.841612	0.709604	1.0	1.0	2.0	2.0	12.0
SERIT SAYISI	12804.0	2.696657	1.232542	1.0	2.0	2.0	3.0	13.0
Yasal Hız Limit	12804.0	50.820056	7.410005	20.0	50.0	50.0	50.0	120.0
ARAC MODELİ	12804.0	2003.451656	8.675636	1962.0	1997.0	2005.0	2011.0	2021.0
EHLİYET YILI	12804.0	12.584271	9.951433	1.0	4.0	10.0	19.0	61.0
Araba Yaşı	12804.0	20.548344	8.675636	3.0	13.0	19.0	27.0	62.0

Yukarıda yer alan tablodaki istatistik özetinden şu sonuçlara varabiliriz:

- Kazaların gerçekleştiği yerlere baktığımızda, o lokasyonlar için olabilecek kaza ciddiye seviyesi KGM tarafından 3.0 ile 180.0 arası belirlenmişken aynı lokasyonlarda gerçekleşmiş kazaların ciddiye seviyesi 0.0 ile 5.0 arasında belirlenmiştir.
- Gerçekleşen kazaların ortalama olarak ne zaman ve nerede gerçekleştiğine bakmak için “AY”, “KAZA SAAT” ve “MAHALLE yoğunluğu” incelendiğinde kazaların ortalama olarak ~2 değerinde yoğunluğuna sahip mahallelerde, haziran ayının sonlarına doğru ve genellikle sabah ~6,5-~7 saatleri arasında gerçekleştiğini görüyoruz.
- “KAZA SAAT” değeri ortalama olarak ~14 gösterirken max değerinin 45.0 göstermesi, “OLU SAYISI” değeri ortalama olarak ~0 gösterirken max değerinin 8.0

göstermesi, “YARALI SAYISI” değeri ortalama olarak ~1 gösterirken max değerinin 36.0 göstermesi, “TASIT SAYISI” değeri ortalama olarak ~1 gösterirken max değerinin 12.0 göstermesi, “SERIT SAYISI” değeri ortalama olarak ~2 gösterirken max değerinin 13.0 göstermesi, “EHİYET YILI” değeri ortalama olarak ~12 gösterirken max değerinin 61.0 göstermesi, bu değerlerin aykırı bir değer olarak sağa çarpık olduğunu ve bu yüzden bu değerlerin kaldırılabilir olduğunu bizlere sunmaktadır.

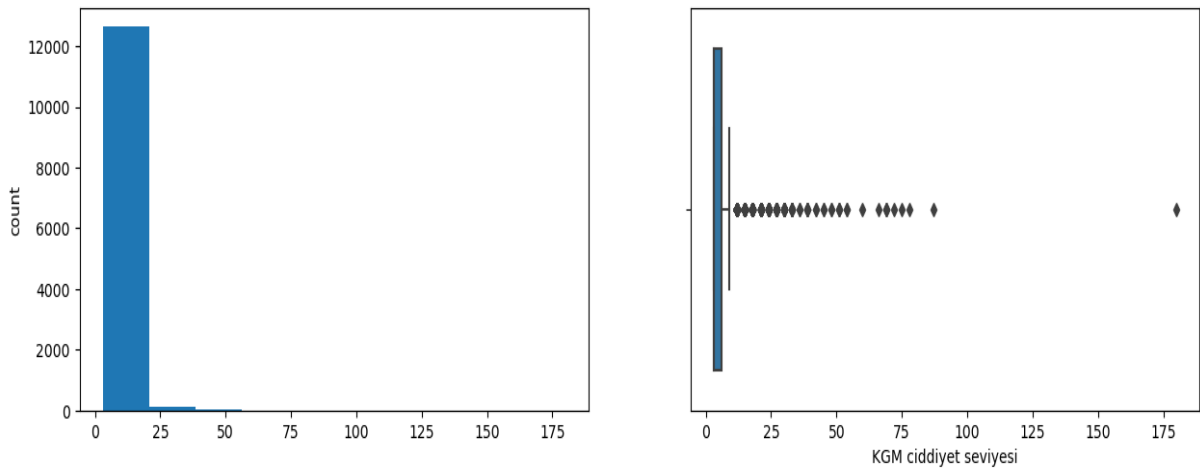
- “OLU SAYISI” ve “YARALI SAYISI” niteliklerinin değerleri incelendiğinde ise ortalama olarak, gerçekleşen kazalarda ~283 kişiden birinin öldüğünün diğerlerinin yaralı kurtulduğunu görüyoruz.

2.1.1 Tek değişkenli analiz

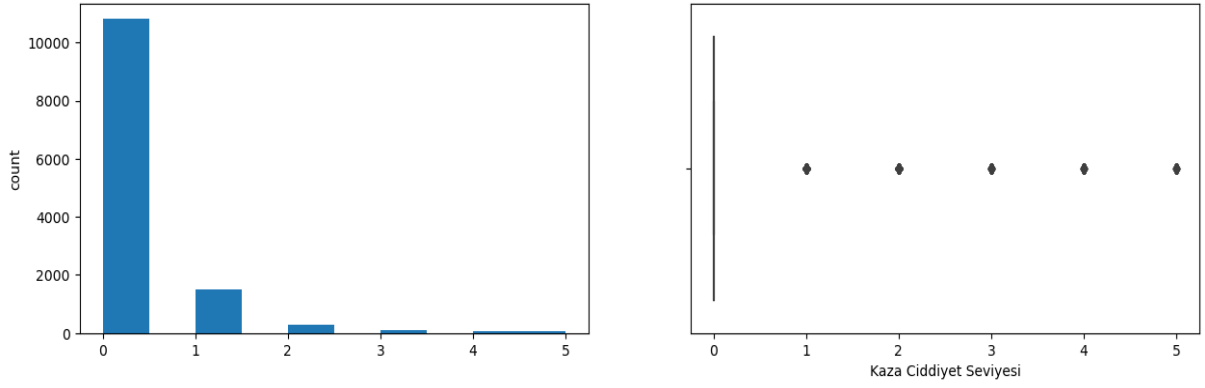
Tek değişkenli analiz, veri setindeki her bir değişkenin tek tek ele alınarak analiz edilmesidir. Bu yüzden önemli bir analiz adımı olup daha nicel sonuç yorumlaması yapabilmek içinde veri görselleştirilmesi önem sarf etmektedir. Burada veri görselleştirme içinse Matplotlib ve Seaborn kütüphaneleri kullanılarak 2D çizim ve grafikler elde ediyoruz.

Sayısal değere sahip niteliklerin tek değişkenli analiz grafikleri aşağıdaki gibidir:

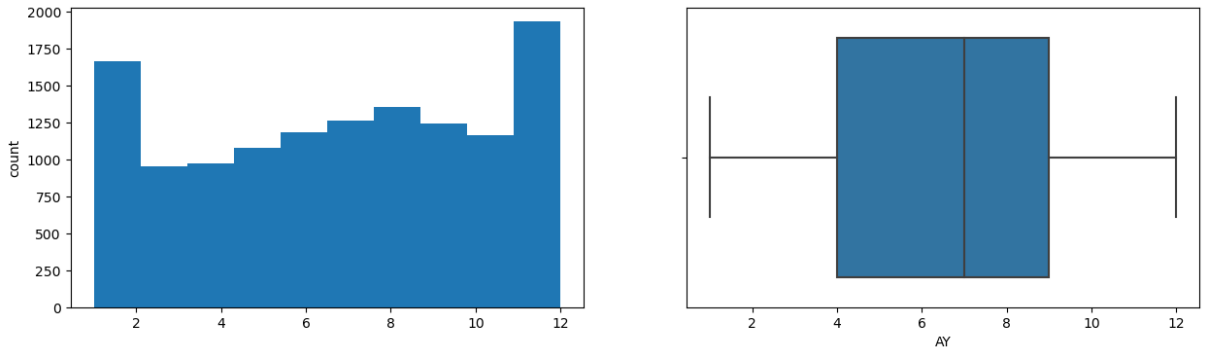
Tablo 3 "KGM ciddiye seviyesi" Tek Değişkenli Analiz Grafiği



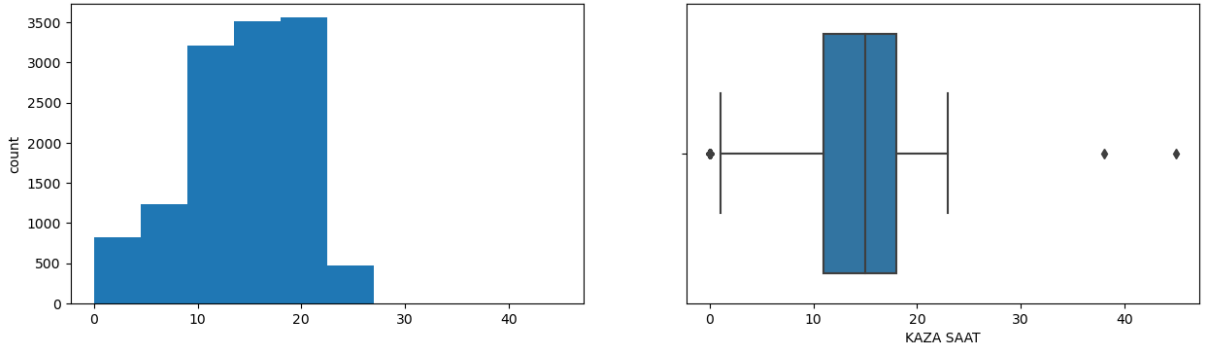
Tablo 4 "Kaza Ciddiyet Seviyesi" Tek Değişkenli Analiz Grafiği



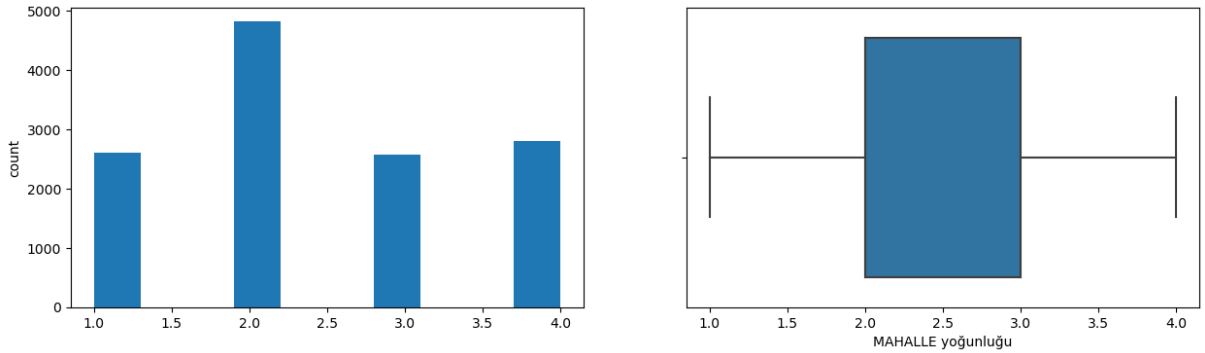
Tablo 5 "AY" Tek Değişkenli Analiz Grafiği



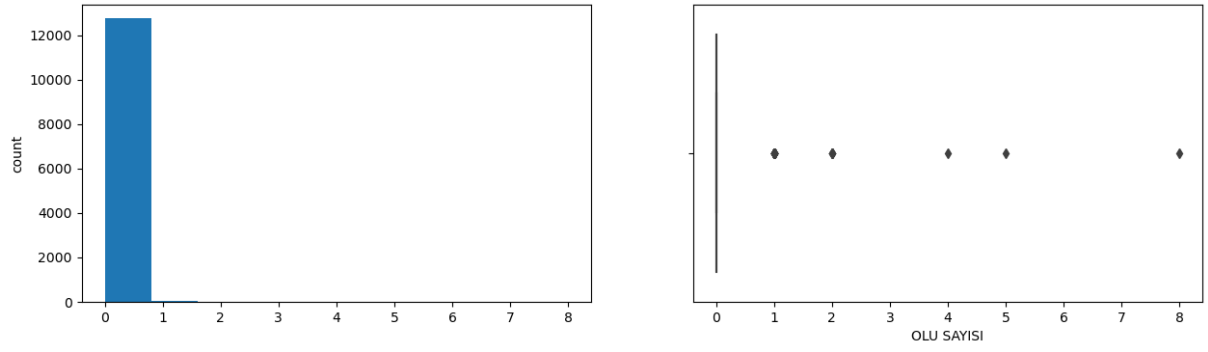
Tablo 6 "KAZA SAAT" Tek Değişkenli Analiz Grafiği



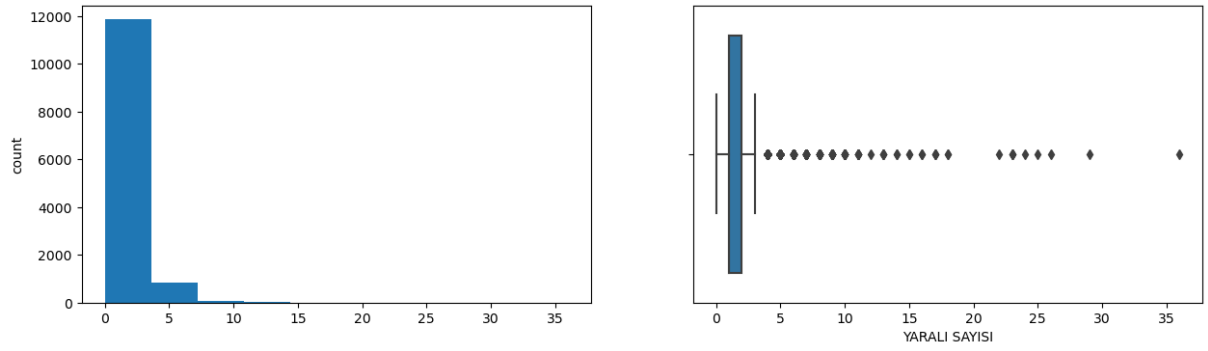
Tablo 7 "MAHALLE yoğunluğu" Tek Değişkenli Analiz Grafiği



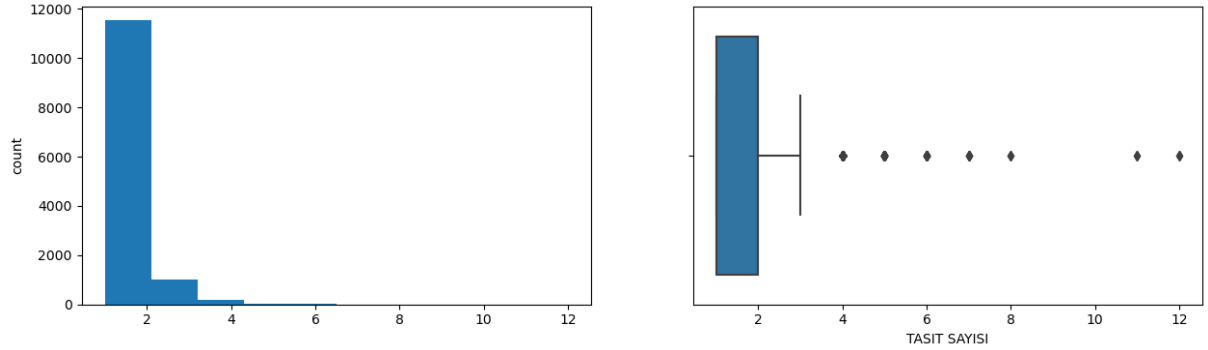
Tablo 8 "OLU SAYISI" Tek Değişkenli Analiz Grafiği



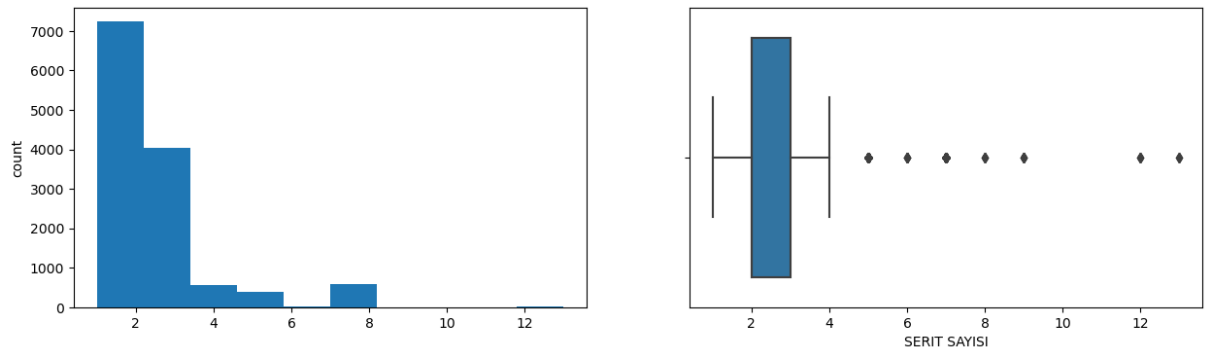
Tablo 9 "YARALI SAYISI" Tek Değişkenli Analiz Grafiği



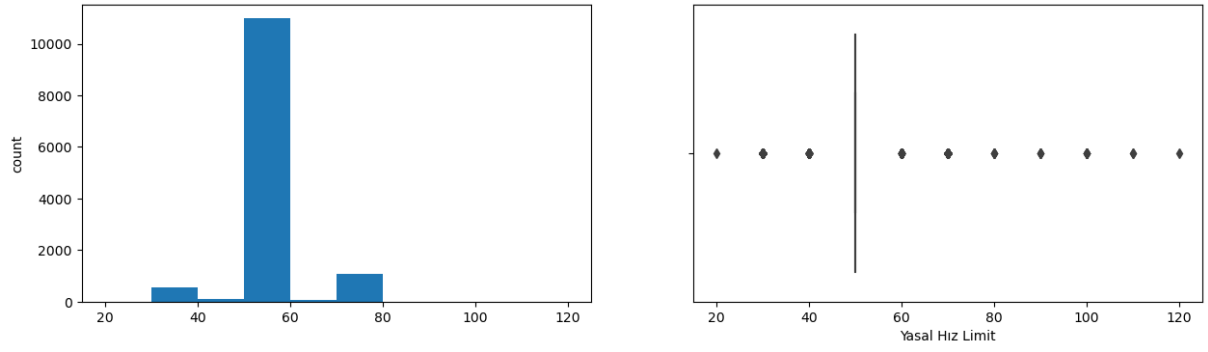
Tablo 10 "TASIT SAYISI" Tek Değişkenli Analiz Grafiği



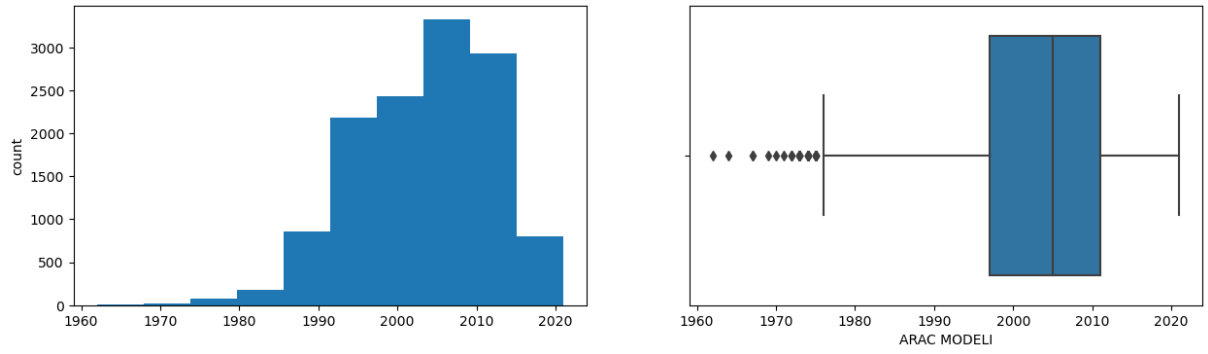
Tablo 11 "SERIT SAYISI" Tek Değişkenli Analiz Grafiği



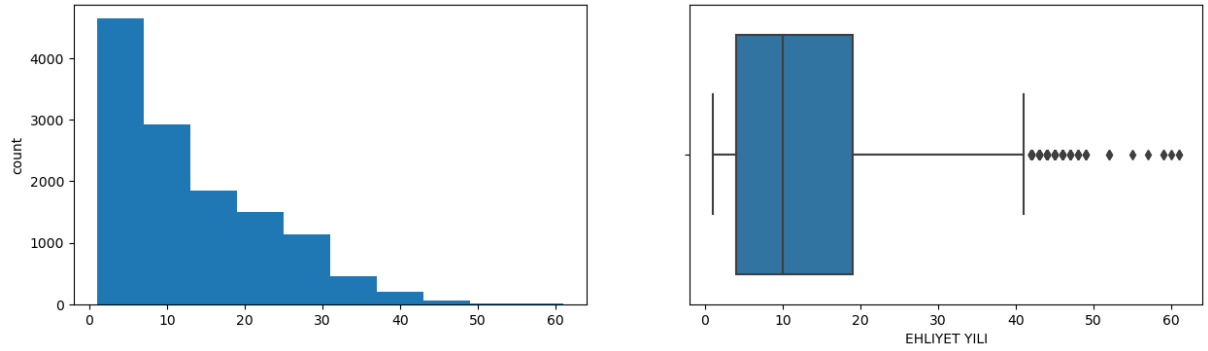
Tablo 12 "Yasal Hız Limit" Tek Değişkenli Analiz Grafiği



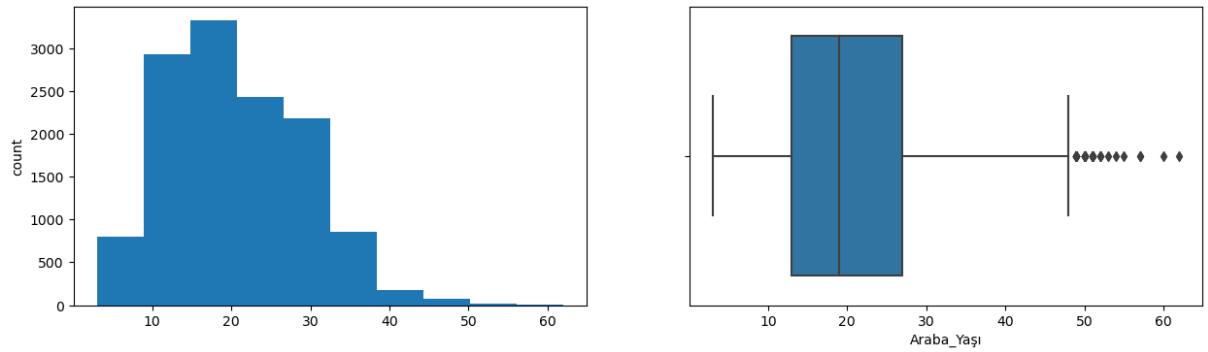
Tablo 13 "ARAC MODELİ" Tek Değişkenli Analiz Grafiği



Tablo 14 "EHLIYET YILI" Tek Değişkenli Analiz Grafiği



Tablo 15 "Araba_Yaşı" Tek Değişkenli Analiz Grafiği



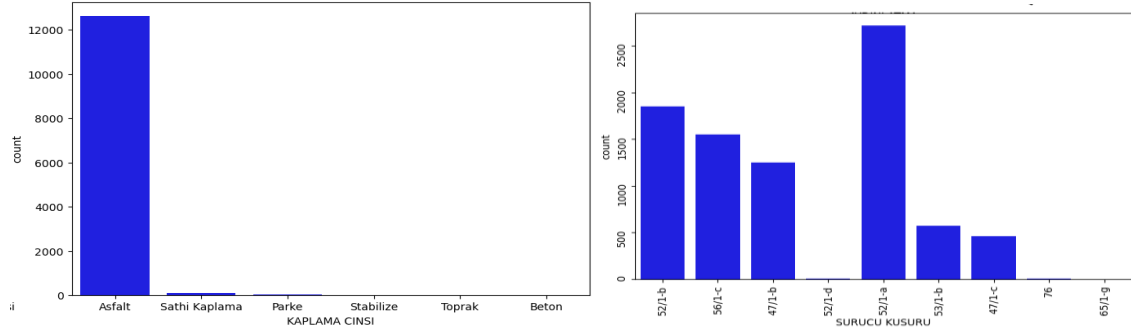
Yukarıdaki tek değişkenli analiz grafiklerini incelediğimiz zaman şu sonuçlara varıyoruz:

- “KGM ciddiye seviyesi” niteliği 0~10 arasında kümelenip sonrasında ~85 değerlerine kadar düzenli ve seyrek görüntü vermektedir.
- “Kaza ciddiye seviyesi” niteliği 0 değerinde yoğunluk göstermekle beraber diğer değerler üzerinde dengeli dağılım gerçekleştirmiştir.
- “AY” niteliği tüm değerleri almış ve devamlılık göstermiş olsa da 4-9 değerlerinde yoğun bir kümelenme olmuştur.
- “KAZA SAAT” niteliği 0~20 değerleri arasında devamlılık gösterip üst değerlerde çok az bulunsa da yoğunluğu ~10~20 değerleri arasında yaşamıştır.
- “MAHALLE yoğunluğu” niteliği de tüm değerler üzerinde devamlılığını korusada asıl yoğunluğu 2.0-3.0 değerleri arasında yaşamıştır.
- “OLU SAYISI” niteliği 0 değerinde ağırlık gösterse de 1,2,4,5,8 değerlerinde de ufak hareketlenmeler oluşturmıştır.
- “YARALI SAYISI” niteliği ise ~0~5 değerleri arasında devamlılığını gösterip ~2.5 civarında da yoğunluk oluşturmıştır. Üst değerlerde ise anlık yoğunluklar yaşamıştır.
- “TASIT SAYISI” niteliği 0-2 değerlerinde kümelenip ~3 değerlerine kadar devamlılık sağlamıştır. Sonrasında belirli değerlerde ufak kümelenmeler oluşturmıştır.
- “SERIT SAYISI” niteliği 0-4 değerlerinde devamlılık sağlayıp ~3 değerlerinde ise yoğunluk sağlamıştır. Sonrasında ise belirli değerlerde ufak kümelenmeler oluşturmıştır.
- “Yasal Hız Limit” niteliği sürekli olarak sabit aralıklarla ufak kümelenmeler oluştursa da ~50 değerinde yoğun artış yaşamıştır.
- “ARAC MODELİ” niteliği 1960~1975 değerleri arasında parçalı devamlılık sağlarken devamında sürekli devamlılık sağlamıştır. ~1995~2015 yılları arasında ise yoğun artış göstermiştir.
- “EHLİYET YILI” niteliği ise ~5~20 değerleri arasında yoğunluk gösterip ~40 değerine kadar sürekli devam sağlayıp sonrasında parçalı devamlılık sağlamıştır.
- “Araba_Yaşı” niteliği ise ~15~30 değerleri arasında yoğunluk gösterip, ~50 değerine kadar sürekli devam sağlayıp sonrasında parçalı devamlılık göstermiştir.

2.1.2 Sayım grafiđi

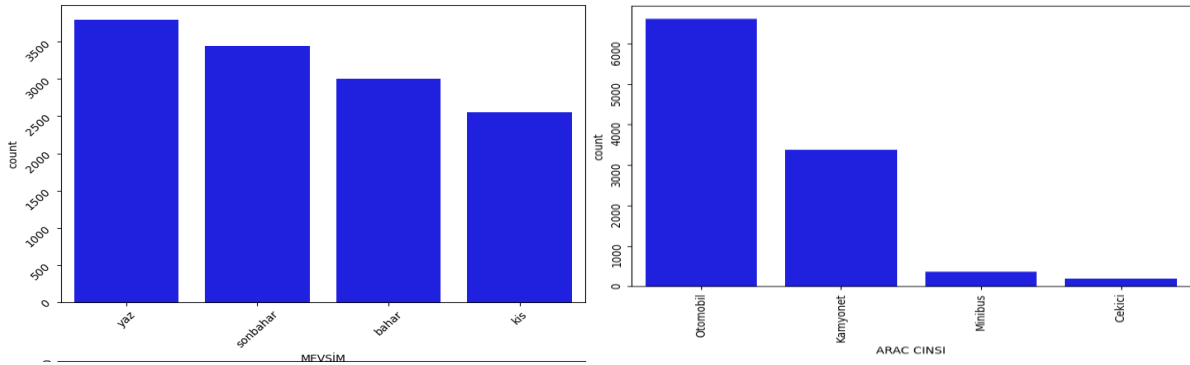
Kategorik değere sahip niteliklerin sayım grafiđi ise ařađıdaki gibidir.

Tablo 16 "KAPLAMA CINSI" ve "SURUCU KUSURU" Niteliklerinin Sayım Grafiđi



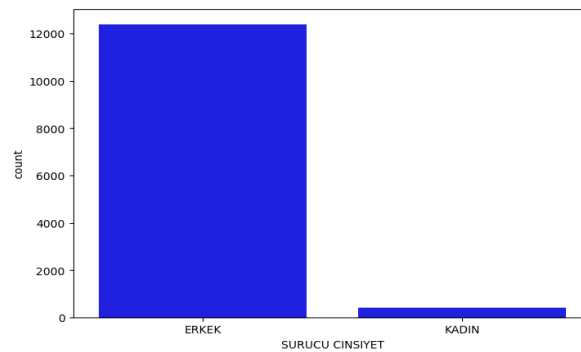
Yukarıdaki grafiklere baktığımızda kazaların çoğunun asfaltta ve 52/1-a kusurundan dolayı gerçekteğini görüyoruz.

Tablo 17 "MEVSİM" ve "ARAC CINSI" Niteliklerinin Sayım Grafiđi



Yukarıdaki grafiklere baktığımızda kazaların gerçekteğini mevsimlerin değeri arasında pek fark olmadığını ama kazaya karışan araç cinslerinde otomobillerin ey yakın rakibinin iki katı değeri yakaladığını görüyoruz.

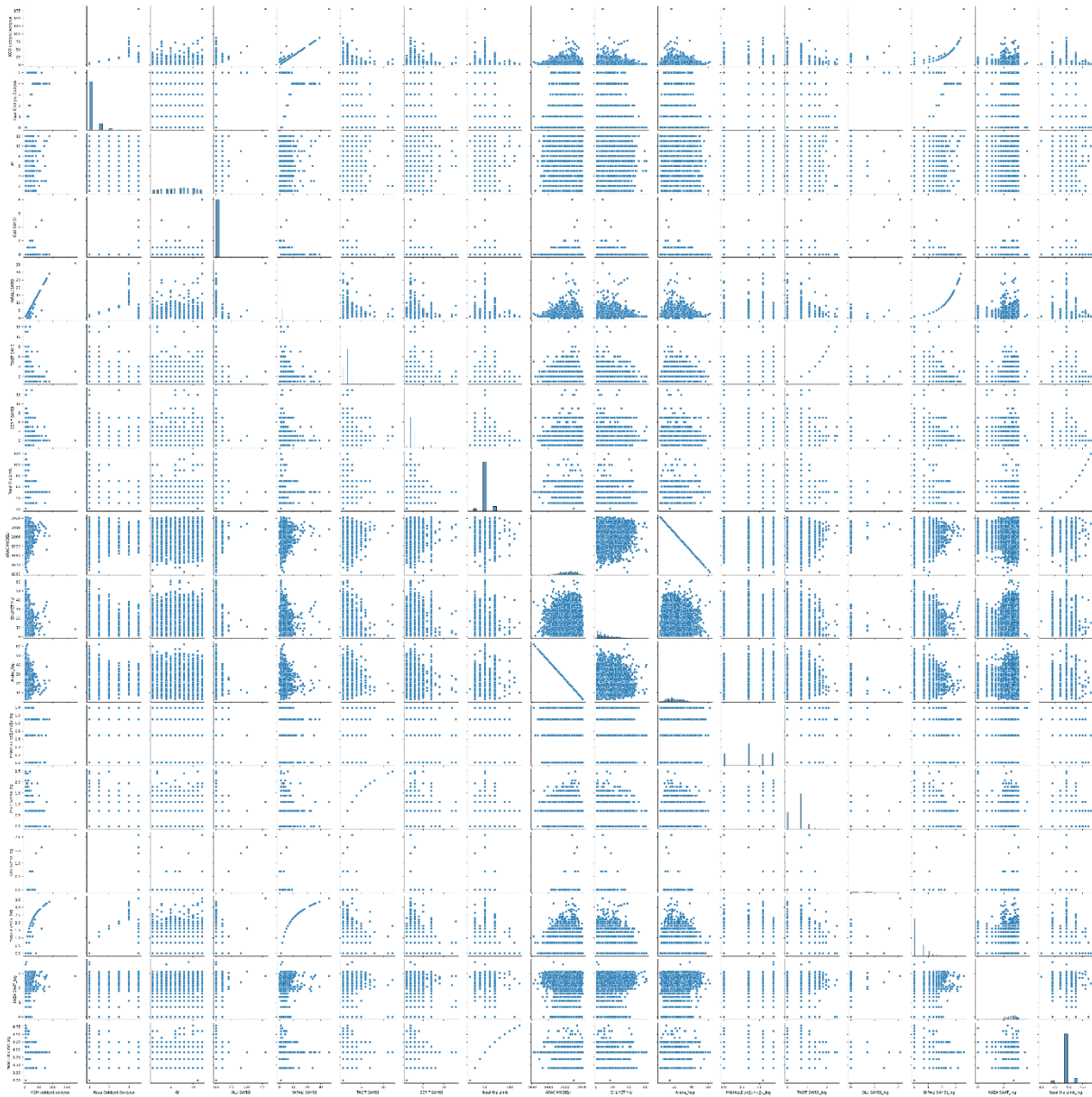
Tablo 18 "SURUCU CINSİYET" Niteliğini Sayım Grafiđi



2.1.3 Çok değişkenli analiz

Veri setimizdeki niteliklerin çok değişkenli analiz tablosunun grafiği ise aşağıdaki şekilde olduğu gibidir. Şekle baktığımızda matrisin soldan sağa yükselen köşegeninde yoğunlaşmalar gerçekleşirken soldan sağa inen köşegeninde yoğunlaşma pek görünmemektedir. Bunun haricinde matrisin orta noktalarına doğru gelindiğinde genellikle yoğunluk artmış bulunurken kenarlarda daha seyrek görünmektedir.

Tablo 19 Niteliklerin Çok Değişkenli Analiz Matrisi

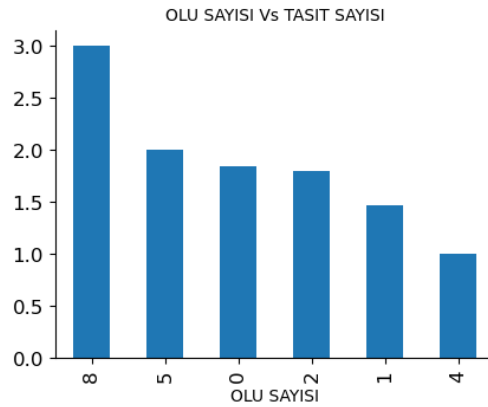


2.1.3 İki değişkenli analiz

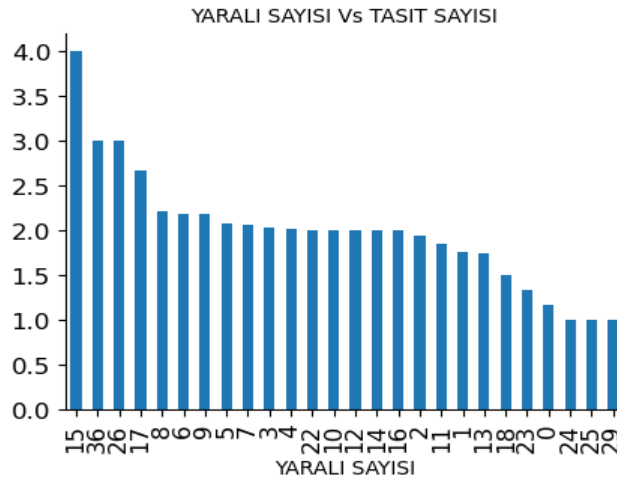
İki değişkenli analiz, iki farklı değişken arasındaki ilişkiyi veya etkileşimi inceleyen istatistiksel bir analiz türüdür. Bu analiz, iki değişken arasındaki ilişkinin doğasını anlamak, değişkenler arasındaki ilişkileri belirlemek veya bir değişkenin diğerine etkisini değerlendirmek için kullanılır.

İki değişkenli analiz, istatistiksel teknikler ve görsel araçlar kullanılarak gerçekleştirilir. Bu veri seti üzerinde yapılan iki değişkenli analiz işlemi içinde 5 karşılaştırma yapılmıştır. Bu karşılaştırmada “TASIT SAYISI” sabit kalırken aynı zamanda sırasıyla “OLU SAYISI”, “YARALI SAYISI”, “KAPLAMA CINSİ”, “SERİT SAYISI” ve “MAHALLE yoğunluğu” ile karşılaştırılmıştır. Bu karşılaştırmaların grafiği aşağıdaki tablolarda yer aldığı gibidir;

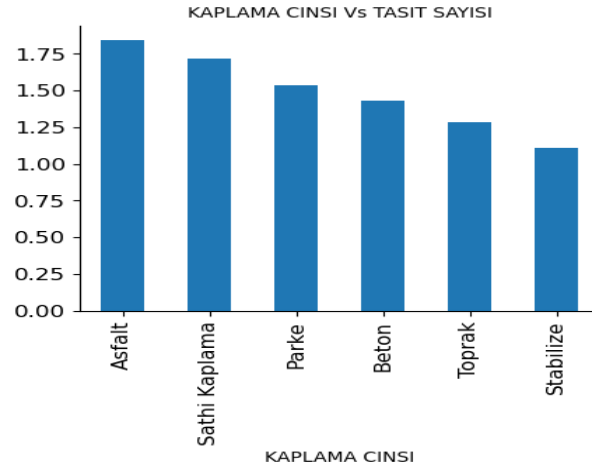
Tablo 20 "OLU SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği



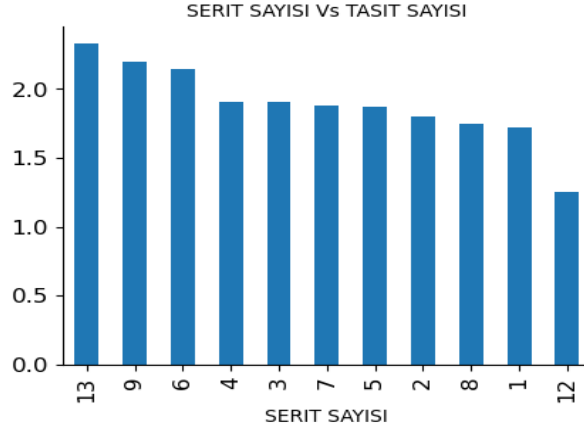
Tablo 21 "YARALI SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği



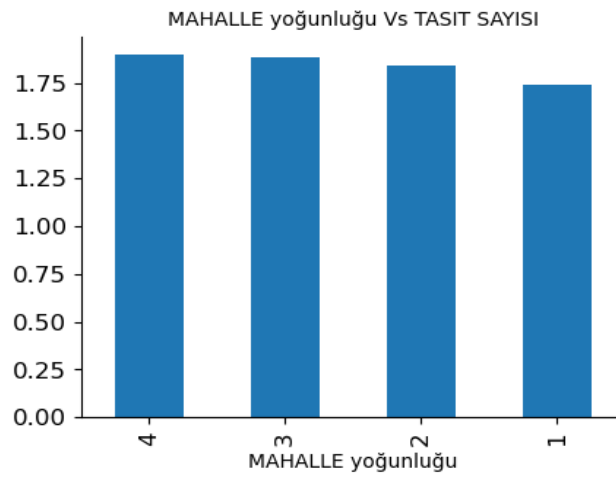
Tablo 22"KAPLAMA CINSI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği



Tablo 23"SERIT SAYISI" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği



Tablo 24"MAHALLE yoğunluğu" ve "TASIT SAYISI"nın İki Değişkenli Analiz Grafiği



Yukarıda yer alan iki değişkenli analiz grafiklerine baktığımız zaman şu sonuçlara varmaktayız:

- "OLU SAYISI" ve "TASIT SAYISI"nın iki değişkenli analiz grafiğine bakınca iki niteliğinde aldığı yüksek değerlerin bir keşişime uğradığını görüyoruz.
- "YARALI SAYISI" ve "TASIT SAYISI"nın iki değişkenli analiz grafiğine bakınca yaralı sayısının 15 olduğunda taşıt sayısının da en yüksek değeri aldığını görüyoruz. 26 ve 36 yaralı varken iki durum içinde taşıt sayısı birbirine çok yakınken diğer değerlerde taşıt sayıları belli bir çizge etrafında irelliyor.
- "KAPLAMA CINSI" ve "TASIT SAYISI"nın iki değişkenli analiz grafiğine bakınca en çok kazanın asfalt yolda olduğunu ve sonrasında stabilize tipindeki yola kadar ortalama aynı oranda azalma gerçekleşiyor.
- "SERIT SAYISI" ve "TASIT SAYISI"nın iki değişkenli analiz grafiğine bakınca 6,9 ve 13 şeritli yollarda aynı oranda ve en çok miktarda kaza gerçekleştiğini görüyoruz. 12 şeritli yollarda hepsinden daha az kaza gerçekleşirken kalan diğer değerlerde yaklaşık aynı miktarda kaza gerçekleşmektedir.
- "MAHALLE yoğunluğu" ve "TASIT SAYISI"nın iki değişkenli analiz grafiğine bakınca mahalleler farklı yoğunluklara sahip olsa da kaza miktarları birbirine oldukça yakın düzeyde bu da gerçekleşen kazaların mahalle yoğunluğu ile çok yakın bir alaka düzeyinin olmadığını bizlere anlatıyor.

SONUÇ

Bu raporda Kayseri il sınırları içerisinde 10 yılda gerçekleşmiş kazalar hakkında detay bilgilerden oluşan veri seti üzerinde yapılan veri analizini sunmaktadır. Bu analiz, veri setinin yapısını anlamak, içerdiği desenleri belirlemek ve trafik kazalarının Kayseri ilindeki durumu hakkında genel bir anlayış edinip gelecekte alınacak önlemler için ışık niteliği taşıması için yapılmıştır. Analiz sonucunda kazaların ay, saat ve mevsimlere göre dağılımı, kazaların türleri ve sonuçları, kazaların konumsal dağılımı, kazaya karışanların nitelikleri gibi bir çok sonucu bizlere vermiştir.