

Patika.dev & Kız Başına Veri Analizi Bootcamp 1. Bitirme Projesi

Bu rapor, bitirme projesinin genel çerçevesini ve temel bulgularını özetlemektedir. Detaylı veri analizleri, istatistiksel yorumlar ve grafiksel görselleştirmeler ilgili Jupyter Notebook (Colab) dosyasında sunulmuştur.

1. Proje Amacı

Bu projenin temel amacı, havayolu yolcu memnuniyetini etkileyen faktörleri analiz etmek ve müşteri memnuniyetini tahmin etmeye yönelik veri odaklı içgörüler elde etmektir. Özellikle Kaggle üzerinden sağlanan “Airline Passenger Satisfaction” veri seti kullanılarak, yolcuların demografik özellikleri (cinsiyet, müşteri türü), seyahat türü, uçuş sınıfı ve rezervasyon deneyimleri gibi değişkenlerin memnuniyet düzeyleri üzerindeki etkisi araştırılmıştır.

Bu bağlamda, proje kapsamında:

- Eksik ve aykırı değerler tespit edilip uygun yöntemlerle işlenmiş,
- Verinin genel yapısı incelenerek temel istatistiksel analizler gerçekleştirilmiş,
- Cinsiyet, müşteri türü, seyahat türü, uçuş sınıfı ve çevrim içi rezervasyon kolaylığı gibi değişkenlerle memnuniyet düzeyleri arasındaki ilişkiler görselleştirilmiş,
- Elde edilen bulgular yorumlanarak işletmelerin stratejik kararlar almasına katkı sağlayacak çıkarımlar yapılmıştır.

Yürütülen bu analiz, veri analitiği ve istatistiksel yöntemler aracılığıyla işletmelere müşteri memnuniyetini artırmaya yönelik veri temelli öneriler geliştirme olanağı sunmaktadır. Aynı zamanda bu çalışma, başlangıç seviyesinden ileri seviyeye geçiş yapan veri analist adaylarının pratik becerilerini geliştirmeyi amaçlayan “Kız Başına Veri Analizi Bootcamp” programının bir parçası olarak gerçekleştirilmiştir.

2. Veri Seti Tanıtımı

Bu projede analiz edilen veri seti, **Kaggle** platformundan temin edilen “**Airline Passenger Satisfaction**” veri setidir. Bu veri seti, havayolu yolcularına ait çeşitli demografik bilgiler ve yolculuk deneyimlerine ilişkin memnuniyet verilerini içermektedir.

Veri setinde toplamda **129,880 satır** ve **25 sütun** bulunmaktadır. Her bir satır, bir yolcunun uçuş deneyimine dair bir gözlemi temsil etmektedir. Bu veri seti, sınıflandırma problemleri ve memnuniyet tahmini gibi uygulamalar için yaygın olarak kullanılmaktadır.

Veri setinde yer alan bazı temel sütunlar şunlardır:

- **Gender:** Yolcunun cinsiyeti (Male, Female)
- **Customer Type:** Yolcunun sadakat durumu (Loyal Customer, disloyal Customer)
- **Age:** Yolcunun yaşı
- **Type of Travel:** İş ya da kişisel amaçlı seyahat (Business, Personal)
- **Class:** Uçuş sınıfı (Eco, Eco Plus, Business)
- **Flight Distance:** Uçuş mesafesi

- **Inflight wifi service - Cleanliness:** Yolculuk sırasında sunulan çeşitli hizmetlere verilen puanlar (0–5)
- **Ease of Online Booking:** Çevrim içi rezervasyon deneyimi kolaylığı (0–5)
- **Satisfaction:** Hedef değişken. Yolcunun memnuniyet durumu:
 - **satisfied**
 - **neutral or dissatisfied**

Veri Türleri:

Veri seti **kategorik ve sayısal (numerik)** değişkenlerin karışımından oluşmaktadır. Kategorik değişkenler sayısallaştırılarak grafiksel analizlere ve modelleme adımlarına uygun hale getirilmiştir.

Bu veri seti, müşteri memnuniyetini etkileyen değişkenlerin analiz edilmesine ve görselleştirilmesine olanak sağladığı için bu projede tercih edilmiştir.

3. Veri Ön İşleme (Eksik ve Aykırı Değerler)

Analiz sürecinin güvenilirliğini artırmak için veri seti üzerinde kapsamlı bir veri ön işleme süreci gerçekleştirilmiştir. Bu aşama, eksik verilerin tespiti, aykırı değerlerin kontrolü ve veri türlerinin uygun hale getirilmesi gibi adımları içermektedir.

3.1 Eksik Değer Analizi

Veri seti incelendiğinde bazı değişkenlerde eksik gözlemler (NaN) tespit edilmiştir. Bu eksik verilerin analizi şu adımlarla yapılmıştır:

- Eksik değer içeren sütunlar belirlenmiş ve toplam eksik değer sayısı her bir sütun için hesaplanmıştır.
- Eksik değerlerin sayısı düşük olan kategorik değişkenlerde **mod (en sık tekrar eden değer)** ile doldurma yöntemi kullanılmıştır.
- Sayısal değişkenlerdeki eksik değerler için **ortalama (mean)** ile doldurma yöntemi tercih edilmiştir.
- Eğer eksik değer oranı çok yüksek olsaydı, ilgili değişkenin analiz dışı bırakılması da değerlendirilecekti; ancak böyle bir duruma rastlanmamıştır.

```
# Eksik değer sayısı ve oranını belirleme
missing_values = df.isnull().sum()
missing_percent = (missing_values / len(df)) * 100

# Sadece eksik değeri olan sütunları filtreler
missing_data = pd.DataFrame({
    'Eksik Değer Sayısı': missing_values,
    'Yüzde (%)': missing_percent
})

missing_data = missing_data[missing_data["Eksik Değer Sayısı"] > 0]
missing_data.sort_values(by="Yüzde (%)", ascending=False)
```

	Eksik Değer Sayısı	Yüzde (%)
Arrival Delay in Minutes	83	0.319526

3.2 Aykırı Değer Analizi

Sayısal değişkenlerde aykırı değer kontrolü aşağıdaki yöntemlerle gerçekleştirilmiştir:

- **Boxplot (Kutu Grafiği)** görselleştirmeleri ile uç değerlerin tespiti yapılmıştır.
- Gözle görülür uç değerler için **IQR (Interquartile Range)** yöntemi uygulanarak sınır değerlerin dışındaki gözlemler aykırı olarak değerlendirilmiştir.
- Aykırı değerlerin sayısı çok fazla olmadığından veri setinden silinmek yerine gerektiğinde dönüştürme işlemleri yapılmıştır (örneğin log dönüşümü).
- Model performansını bozacak düzeyde bir aykırılık gözlemlenmemiştir.

Aykırı Değer Sayıları (IQR Yöntemi)

```
id: 0
Age: 0
Flight Distance: 584
Inflight wifi service: 0
Departure/Arrival time convenient: 0
Ease of Online booking: 0
Gate location: 0
Food and drink: 0
Online boarding: 0
Seat comfort: 0
Inflight entertainment: 0
On-board service: 0
Leg room service: 0
Baggage handling: 0
Checkin service: 3218
Inflight service: 0
Cleanliness: 0
Departure Delay in Minutes: 3569
Arrival Delay in Minutes: 3538
```

3.3 Değişken Türlerinin Dönüştürülmesi

Modelleme ve grafiksel analizler öncesinde:

- Kategorik değişkenler **label encoding** ve **one-hot encoding** gibi yöntemlerle sayısal forma dönüştürülmüştür.
- Hedef değişken olan **satisfaction** sütunu, ikili sınıflandırma için 0 = neutral or dissatisfied, 1 = satisfied şeklinde etiketlenmiştir.
- Sayısal değerlere sahip fakat kategorik anlam taşıyan değişkenlerin (örneğin hizmet puanları) anlamı bozulmadan doğru biçimde işlenmesi sağlanmıştır.

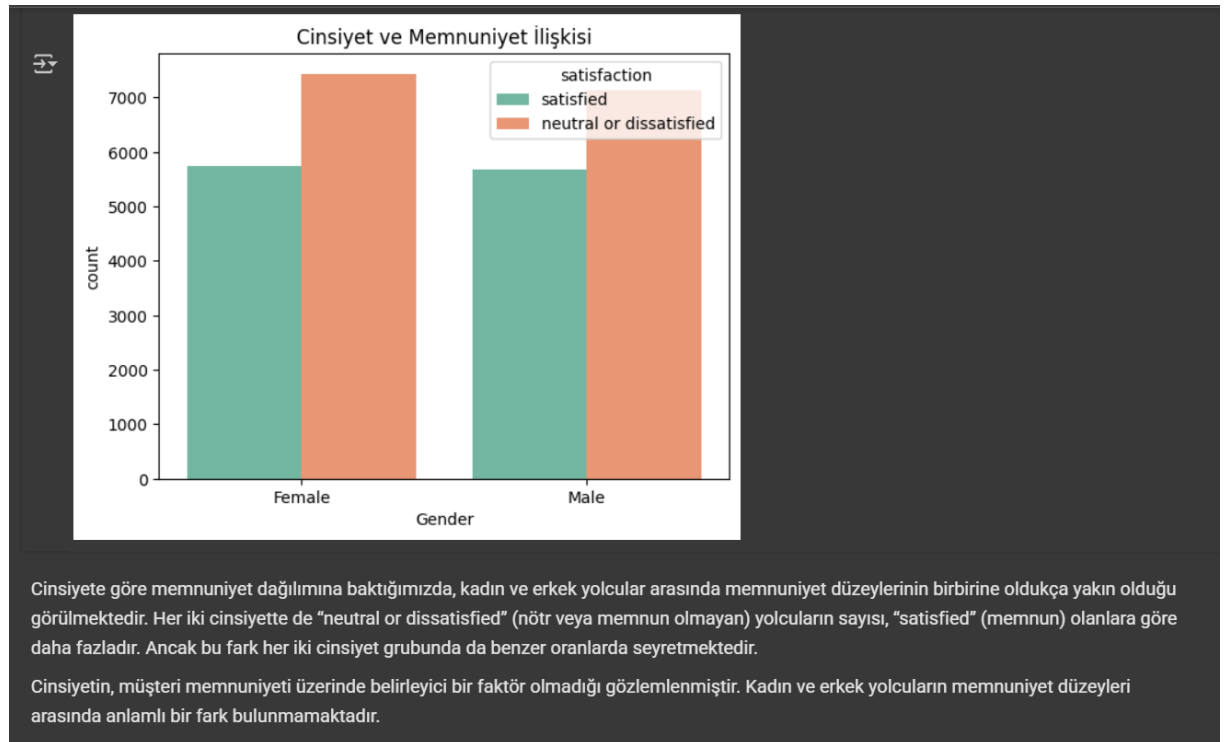
4. Keşifsel Veri Analizi (Exploratory Data Analysis - EDA)

Bu aşamada, veri seti üzerindeki genel eğilimleri, değişkenler arası ilişkileri ve müşteri memnuniyetini etkileyen faktörleri anlamak amacıyla çeşitli istatistiksel analizler ve görselleştirmeler gerçekleştirilmiştir. EDA süreci hem genel dağılımları incelemek hem de memnuniyet değişkeniyle olan ilişkileri anlamak açısından oldukça kritik bir adımdır.

4.1 Demografik Değişkenlerin İncelenmesi

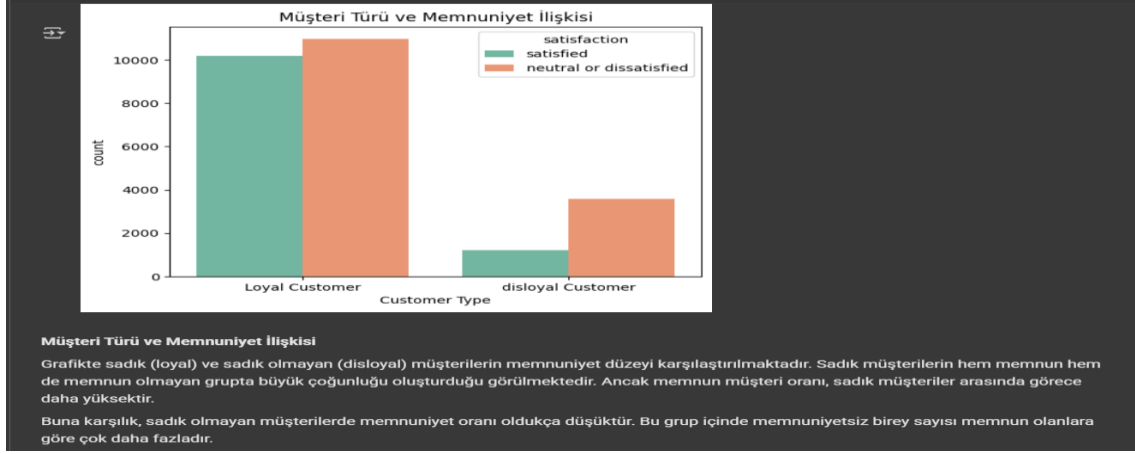
- Cinsiyet ve Memnuniyet İlişkisi:**

Yapılan analizde hem kadın hem erkek yolcular arasında "neutral or dissatisfied" oranının daha yüksek olduğu gözlemlenmiştir. Ancak cinsiyetin memnuniyet üzerinde belirgin bir fark yaratmadığı sonucuna ulaşılmıştır.



- Müşteri Türü ve Memnuniyet İlişkisi:**

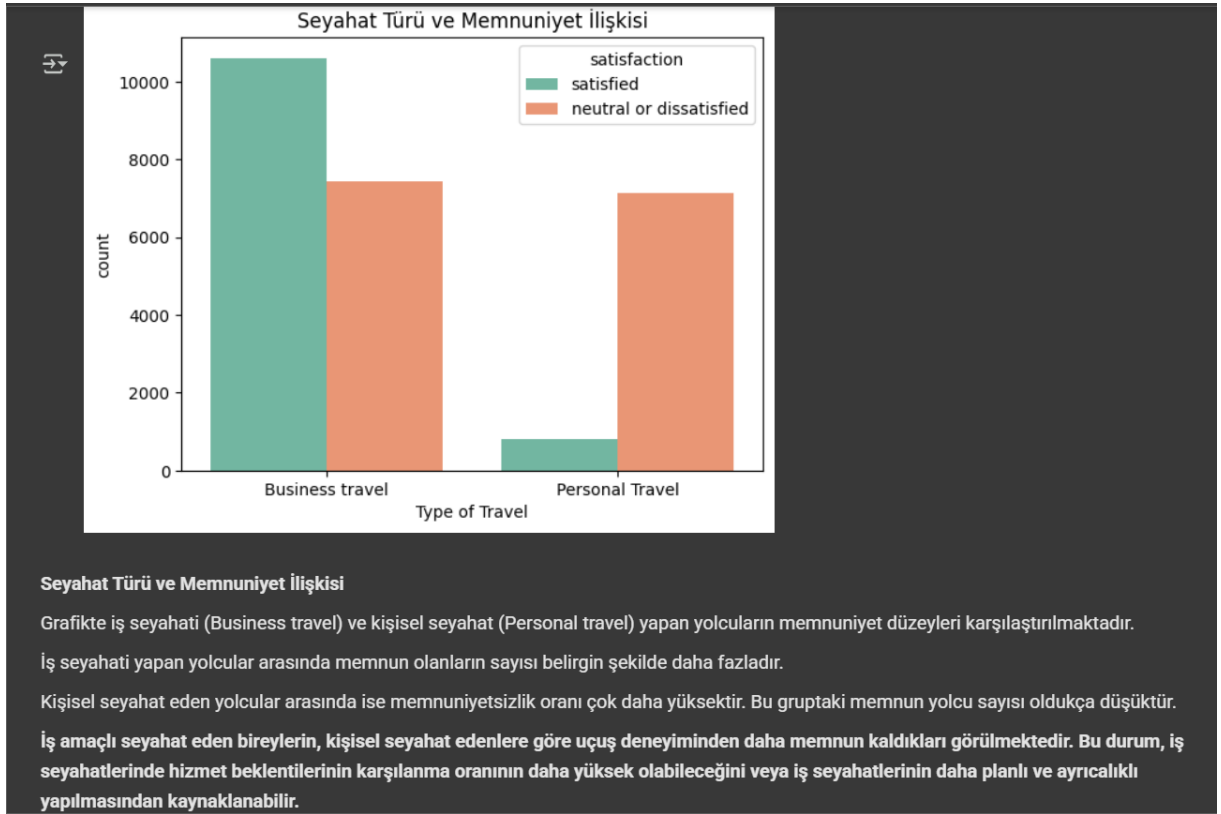
Sadık müşterilerin (loyal customers) memnuniyet oranı disloyal müşterilere göre çok daha yüksektir. Bu durum, havayolu firmasının sadık müşterilerine daha kaliteli hizmet sunduğunu ya da sadık müşterilerin daha az sorun yaşadığını düşündürmektedir.



4.2 Seyahat Özelliklerinin İncelenmesi

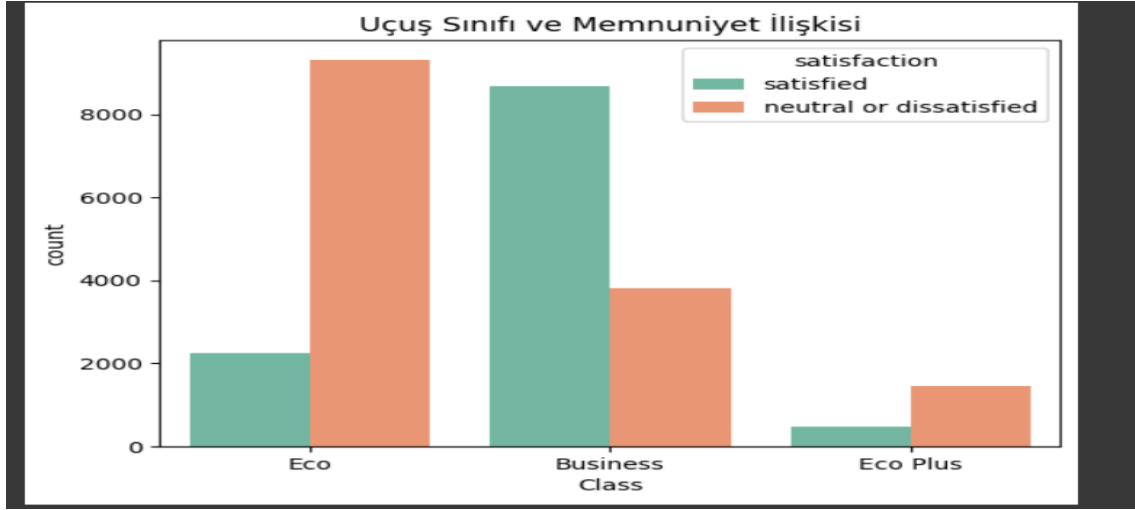
- Seyahat Türü ve Memnuniyet İlişkisi:**

İş seyahati yapan bireylerin memnuniyet düzeyi, kişisel seyahat gerçekleştiren bireylere göre çok daha yüksektir. Bu durum, iş seyahati kullanıcılarına sunulan hizmetlerin daha standart ve profesyonel olabileceğini göstermektedir.



- Uçuş Sınıfı ve Memnuniyet İlişkisi:**

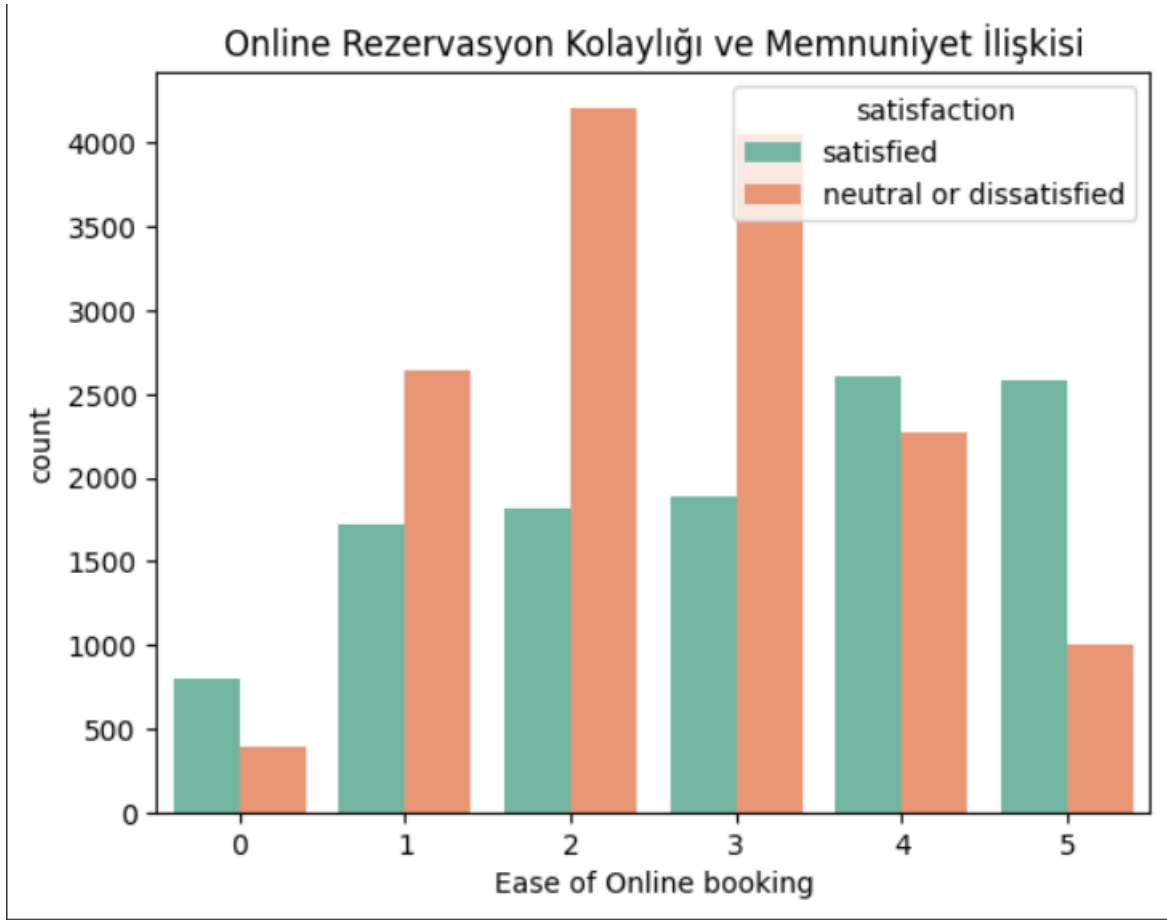
Business class yolcularının memnuniyet oranı oldukça yüksektir. Buna karşılık ekonomi sınıfı (Eco) yolcular arasında memnun olmayanların oranı daha fazladır. Bu sonuç, hizmet kalitesi ve konforun memnuniyet üzerinde belirleyici bir faktör olduğunu ortaya koymaktadır.



4.3 Hizmet Kalitesi ve Online Deneyimin İncelenmesi

- **Online Rezervasyon Kolaylığı ve Memnuniyet İlişkisi:**

Online rezervasyon kolaylığı arttıkça memnuniyet oranının da arttığı görülmüştür. Özellikle 4 ve 5 puan alan yolcularda memnuniyet düzeyi gözle görülür şekilde yükselmiştir. Bu da kullanıcı deneyiminin dijital platformlarda önemli bir memnuniyet belirleyicisi olduğunu göstermektedir.



4.4 Genel Gözlemler

- "Neutral or Dissatisfied" sınıfına ait yolcu sayısı tüm kategorilerde daha fazladır.

- İş seyahati yapan ve Business Class'ta seyahat eden yolcular genellikle daha memnundur.
- Online sistemlerin kullanıcı dostu olması memnuniyetin artırılması açısından önemlidir.

5. Veri Görselleştirme ve Sonuçların Yorumlanması

Proje kapsamında yapılan keşifsel analizler, çeşitli grafikler aracılığıyla görselleştirilmiş ve önemli eğilimler yorumlanmıştır. Hem sayısal hem de kategorik değişkenler için grafik temelli analizler gerçekleştirilmiştir.

5.1 Sayısal Değişkenlerin Görselleştirilmesi

Sayısal değişkenler için dağılım grafikleri, histogramlar ve boxplot görselleri kullanılarak veri setinin genel eğilimleri analiz edilmiştir. Bu görselleştirmeler sayesinde özellikle aşağıdaki bulgular öne çıkmıştır:

- **Flight Distance:** 4000 km üzeri uçuşlar aykırı değer olarak tespit edilmiştir. Bunlar nadir ama anlamlı uçuşlardır.
- **Arrival Delay in Minutes:** 200 dakikayı geçen gecikmeler önemli aykırılıklar içermektedir. Ancak bunlar gerçek hayatta karşılaşılabılır durumlardır.
- **Age:** Yaş dağılımı dengeli görünmektedir, dikkat çekici bir aykırılık bulunmamaktadır.

Bu analizler sayesinde hangi değişkenlerin daha fazla varyansa sahip olduğu ve olası veri bozulmalarının nerelerde yoğunlaştığı tespit edilmiştir.

5.2 Kategorik Değişkenlerin Görselleştirilmesi

Kategorik değişkenler için çubuk grafikler (barplot) kullanılarak memnuniyet değişkeni ile olan ilişkiler analiz edilmiştir. Bu analizlerde aşağıdaki gözlemler yapılmıştır:

- **Gender:** Cinsiyete göre memnuniyet farkı gözlenmemiştir.
- **Customer Type:** Sadık müşteriler (Loyal Customers) açık ara daha memnundur.
- **Type of Travel:** İş seyahati yapan yolcular, kişisel seyahat yapanlara göre daha memnundur.
- **Class:** Business Class yolcuları yüksek oranda memnundur, Economy Class yolcuları daha çok memnuniyetsizdir.
- **Ease of Online Booking:** Rezervasyon kolaylığı arttıkça memnuniyet de artmaktadır.

Bu grafikler, müşteri davranışlarının ve tercih edilen hizmet türlerinin memnuniyet üzerindeki etkisini açıkça göstermektedir.

Grafikler detaylı şekilde notebook dosyasında bulunmaktadır. Rapora eklenmemiştir.

6. Genel Değerlendirme ve Sonuç Önerileri

Bu çalışma kapsamında, havayolu yolcularının memnuniyet düzeylerini etkileyen faktörler, veri analitiği ve görselleştirme teknikleri kullanılarak detaylı şekilde incelenmiştir. Kullanılan veri seti, yolculara ait demografik bilgiler, seyahat özellikleri ve hizmet deneyimlerine ilişkin çok boyutlu veriler içermektedir.

Genel Bulgular:

- Eksik Veriler: Sadece Arrival Delay in Minutes deęiřkeninde eksik veri tespit edilmiř olup, eksiklikler medyan deęeri ile doldurularak analize dahil edilmiřtir.
- Aykırı Deęerler: Özellikle uçuř mesafesi ve varıř gecikmeleri gibi deęiřkenlerde istatistiksel olarak aykırı deęerler tespit edilmiřtir. Bu deęerler, veri geręeklięini yansıttıęı iin analiz dıřı bırakılmamıř, yorumlarda dikkate alınmıřtır.
- Sayısal Veriler: Boxplot, histogram ve daęılım grafikleri yardımıyla sayısal deęiřkenlerin genel daęılımları analiz edilmiřtir. Verilerin oęunda simetrik veya normal daęılıma yakın yapılar gzlemlenmiřtir.
- Kategorik Veriler: Barplot yardımıyla memnuniyet deęiřkeni ile olan iliřkileri analiz edilen kategorik deęiřkenler arasında en belirgin etkiler müşteri tipi, seyahat sınıfı ve seyahat türünde gzlemlenmiřtir.

Öne ıkan Sonular:

- Sadık müşteriler (Loyal Customers) ve Business Class yolcuları genel olarak daha yüksek memnuniyet bildirmiřtir.
- Kiřisel seyahat gerekleřtiren yolcuların memnuniyet düzeyi iř seyahatindekilere kıyasla daha düşük kalmıřtır.
- Online rezervasyon kolaylıęı, hizmet kalitesi ve uak ii deneyim gibi hizmete ynelik deęiřkenler memnuniyet düzeyini doęrudan etkilemektedir.

Öneriler:

- Havayolu firmaları, sadık müşteri programlarını artırarak müşteri memnuniyetini yükseltebilir.
- Özellikle ekonomi sınıfı yolculara ynelik hizmet kalitesinin iyileřtirilmesi, genel memnuniyet seviyesini artırabilir.
- Gecikmeleri azaltacak operasyonel iyileřtirmeler, müşteri deneyimi üzerinde doęrudan olumlu etki yaratabilir.
- Online hizmetlerin kullanılabilirlięi ve eriřilebilirlięi artırılarak, memnuniyet düzeyleri daha da yükseltilebilir.