



Cumhuriyetimizin 100. Yılı Kadınlara Özel

Data Science with Python Bootcamp

Final Projesi Raporu:

Airlines Customer Satisfaction Prediction

Nursena Ertuğrul

Eğitmen Adı: Çağla Şahin

05.01.2024

Table of Contents

1. Giriş.....	3
1.1 Projeye Genel Bakış	4
2. EDA	4
3. Veri Görselleştirmesi	15
3.1 Satisfaction	15
3.2 Gender	16
3.3 Customer Type	17
3.4 Age.....	17
3.5 Type of Travel	18
3.6 Class	19
3.8 Flight Distance	20
3.9 Derecelendirme Sütunları	21
4. Correlation Analizi	23
5. Modelleme	25
4.1 Decision Tree	27
4.2 Logistic Regression	28
5. Conclusion	32

1. Giriş

Günümüzde birden fazla ulaşım yolu vardır. Bunlardan en kolay ve hızlı olanı bilindiği üzere havayoludur. Yıllar geçtikçe bu ulaşım yoluna talep giderek artmaktadır. Artan talep sonrasında havayolu şirketlerinin de rekabeti giderek artmaktadır. Çok yaygın olan bu ulaşım sektörünün de müşteri memnuniyeti elbette çok büyük öneme sahiptir. Havayolları, müşterilerine yolculuklarının bütün serüvenine konfor, rahatlık ve kolaylık sağlaması şirketlerin amacı olmaktadır. Sadık müşterilerin yanında yeni potansiyel müşterinde ilk tercihi olmak havayolu şirketleri için çok önemlidir. Şirketler, müşteri memnuniyetlerini inceleyerek bu amacı daha net analiz edip yeni politikalar belirleyebilir. Bunu için bu projede Invistico Airlines şirketini tercih eden müşterilerin memnuniyet verilerini inceleyerek hem şirketin müşteri memnuniyetleri hakkında tahmin yapmasını hem de gelecek dönemlerde nasıl iyileştirmeler yapılabilir gözlemleri sağlanacaktır. Veri seti [kaggle](#) üzerinden elde edildi. Veri setinin linkine bu [bağlantı](#) üzerinden ulaşılabilir.

1.1 Projeye Genel Bakış

Veri setinin içeriği bu havayolu şirketini kullanan müşterilerin uçuşları hakkındaki değerlendirmelerini içermektedir. Bu projede müşteri memnuniyetlerini tahmin etmek için makine öğrenimi kullanılmaktadır. Bu projede kullanılan veri seti, müşterinin uçuş bilgileri (cinsiyet, müşteri tipi, yaşı, yolculuk türü, sınıfı, uçuş mesafesi) ve hizmet alanlarına dair değerlendirmeleri yer almaktadır. Verilmiş olan bu bilgileri müşterinin uçuşundan genel olarak memnun kalıp kalmadığını tahmin etmek üzere kullanılacaktır.

İlk olarak veri setini tanımayla başlayacağız. Veri setinin boyutuna, verilerin dağılımına, eksik değerlere verinin analizi için bakacağız. Daha sonra modeli eğitmek ve daha doğru değerlendirmeler almak için iki farklı modelleme yapacağız. Son olarak iki model karşılaştırılıp en iyi olanı tercih edip tahminleri yapacağız.

Modelin sınıflandırma raporlarına bakılacak. Elde edilen sonuçlar havayolu şirketinin eski müşterilerinin memnun kalıp kalmadıkları, memnun kalmayan müşteriler için nasıl iyileştirmeler yapılabileceği ve gelebilecek potansiyel müşteriler için bir politika izlenimi vermesi için kullanılacaktır.

2. EDA

Veri setini incelemek için öncelikle notebook’a veri setinin çağırılması gerekmektedir. Bunun için pandas kütüphanesine bulunan “read_csv” fonksiyonu kullanılır. Bu fonksiyonun projede kullanımı Figür 1’de gösterilmiştir.

```
df = pd.read_csv("Invistico_Airline.csv")
```

Fig.1. Dosya Okuması

Tüm veri setini çağırmak yerine veri setinin ilk beş satırını yazdırmak daha kolay ve okunabilir bir tercihtir. Bunun için “head” fonksiyonu kullanılır. Fonksiyonun proje içinde kullanımı ve elde edilen sonucu Figür 2’de gösterilmiştir.

```
df.head()
```

	satisfaction	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Seat comfort	Departure/Arrival time convenient	Food and drink	...	Online support	Ease of Online booking	On-board service	Leg room service	Baggage handling	Check-in service
0	satisfied	Female	Loyal Customer	65	Personal Travel	Eco	265	0	0	0	...	2	3	3	0	3	
1	satisfied	Male	Loyal Customer	47	Personal Travel	Business	2464	0	0	0	...	2	3	4	4	4	
2	satisfied	Female	Loyal Customer	15	Personal Travel	Eco	2138	0	0	0	...	2	2	3	3	4	
3	satisfied	Female	Loyal Customer	60	Personal Travel	Eco	623	0	0	0	...	3	1	1	0	1	
4	satisfied	Female	Loyal Customer	70	Personal Travel	Eco	354	0	0	0	...	4	2	2	0	2	

Fig.2. İlk Beş Satır

Veri setinde kaç adet satır ve sütun bulunmakta öğrenmemiz gerekmektedir. Bunun için “shape” fonksiyonu kullanılacaktır. Fonksiyonun kullanımı ve çıktısı Figür 3’de gösterilmiştir.

```
df.shape
```

```
(129880, 23)
```

Fig.3 Satır ve Sütun Sayısı

Görünen o ki veri setinde 129880 adet veri ve 23 adet sütun bulunmaktadır. Bu demek oluyor ki havayolu şirketinin 129880 adet müşterisi ve her müşterinin 23 adet bilgi ve değerlendirmesi mevcuttur.

Bu 23 adet değişkenin veri setinde hangi veri tipinde olduğu bilinmelidir. Veri tiplerini bilmeden analiz etmemiz mümkün değildir.

değişkenleri integer veya kabul etmek yanlış bir öğrenmek için “info” kullanacağız. Bu çıktısı Figür 4’de

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129880 entries, 0 to 129879
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   satisfaction                             129880 non-null object
1   Gender                                   129880 non-null object
2   Customer Type                             129880 non-null object
3   Age                                       129880 non-null int64
4   Type of Travel                           129880 non-null object
5   Class                                    129880 non-null object
6   Flight Distance                           129880 non-null int64
7   Seat comfort                             129880 non-null int64
8   Departure/Arrival time convenient         129880 non-null int64
9   Food and drink                           129880 non-null int64
10  Gate location                             129880 non-null int64
11  Inflight wifi service                     129880 non-null int64
12  Inflight entertainment                   129880 non-null int64
13  Online support                             129880 non-null int64
14  Ease of Online booking                   129880 non-null int64
15  On-board service                         129880 non-null int64
16  Leg room service                         129880 non-null int64
17  Baggage handling                         129880 non-null int64
18  Checkin service                         129880 non-null int64
19  Cleanliness                             129880 non-null int64
20  Online boarding                         129880 non-null int64
21  Departure Delay in Minutes               129880 non-null int64
22  Arrival Delay in Minutes                 129487 non-null float64
dtypes: float64(1), int64(17), object(5)
```

Bütün object veri tipi olarak adım olur. Veri tiplerini fonksiyonunu fonksiyonun kullanımı ve gösterilmiştir.

Fig.4. Veri Tipi İncelemesi

Çıktı incelendiği zaman görünüyor ki 17 adet integer, 5 adet object, ve 1 adet float veri tipinde değişkenimiz mevcut. Bu işlem sonrasında veri setinde herhangi bir null değer var mı kontrol edilmelidir. Bunun için “isna” ve “sum” fonksiyonları kullanılacaktır. Kullanımları ve sonuçları Figür 5’de yer almaktadır.

```
df.isna().sum()
satisfaction          0
Gender                0
Customer Type         0
Age                  0
Type of Travel        0
Class                 0
Flight Distance       0
Seat comfort          0
Departure/Arrival time convenient  0
Food and drink        0
Gate location         0
Inflight wifi service  0
Inflight entertainment  0
Online support        0
Ease of Online booking  0
On-board service      0
Leg room service      0
Baggage handling      0
Checkin service       0
Cleanliness           0
Online boarding       0
Departure Delay in Minutes  0
Arrival Delay in Minutes 393
dtype: int64
```

Fig.5. Null Değer Kontrolü

Elde ettiğimiz bu değişkenler adları ve tanımları bize havayolu şirketi tarafından verilmiştir.

Sırayla bu değişkenlerin veri setindeki anlamlarına bakalım.

1. **satisfaction:** Müşterinin genel memnuniyet düzeyi. "satisfied" veya "dissatisfied" seçeneklerinin bulunduğu kategorik bir değişkendir.
2. **Gender:** Müşterinin cinsiyeti. "Male" veya "Female" seçeneklerinin bulunduğu kategorik bir değişkendir.
3. **Customer Type:** Müşterinin "loyal customer" mı yoksa "disloyal customer" mı olduğu değişkendir.

4. **Age:** Müşterini yaşı
5. **Type of Travel:** "Personal Travel" veya "Business travel" gibi seyahat türü.
6. **Class:** "Eco" veya "Business" gibi seyahat sınıfı.
7. **Flight Distance:** Uçuş Mesafesi
8. **Seat Comfort:** Müşterinin koltuk konforuna ilişkin değerlendirmesi.
9. **Departure/Arrival time convenient:** Bu sütun, müşterinin kalkış ve varış saatlerinin rahatlığından duyduğu memnuniyeti ifade eder.
10. **Food and Drink:** Müşterinin yiyecek ve içecek kalitesine ilişkin değerlendirmesi.
11. **Gate Location:** Müşterinin kapı konumuna ilişkin değerlendirmesi.
12. **Inflight Wifi Service:** Uçak içi Wi-Fi hizmetine ilişkin müşteri derecelendirmesi.
13. **Inflight Entertainment:** Uçak içi eğlence seçeneklerine ilişkin müşteri derecelendirmesi.
14. **Online support:** Çevrimiçi müşteri desteğinin müşteri değerlendirmesi.
15. **Ease of Online booking:** Müşterinin çevrimiçi rezervasyon kolaylığına ilişkin değerlendirmesi.
16. **On-board service:** Havayolu tarafından sağlanan uçakta hizmete ilişkin müşteri derecelendirmesi.
17. **Legroom service:** Uçuş sırasında sağlanan bacak odası hizmetine ilişkin müşteri derecelendirmesi.
18. **Baggage handling:** Müşterinin bagaj taşıma derecelendirmesi.
19. **Checkin service:** Müşterinin check-in hizmetine ilişkin değerlendirmesi.
20. **Cleanliness:** Kabin temizliğine ilişkin müşteri derecelendirmesi.
21. **Online boarding:** Çevrimiçi binış sürecine ilişkin müşteri derecelendirmesi.
22. **Departure Delay in Minutes:** Her uçuş için dakika cinsinden kalkış gecikmesi.
23. **Arrival Delay in Minutes:** Her uçuş için dakika cinsinden varış gecikmesi.

Alınan çıktıya bakıldığında “Arrival Delay in Minutes” değişkeninde 393 adet null değer olduğu görülmektedir. Null değerleri veri setinden silmeden önce veri setinde bulunan geçerli değerlere olan oranına bakmalıyız. Oran küçükse silme işlemi yapılabilir. Bu incelemeler Figür 6’yı inceleyerek yapılabilir.

```
df.isna().sum() / df.shape[0]
```

satisfaction	0.000000
Gender	0.000000
Customer Type	0.000000
Age	0.000000
Type of Travel	0.000000
Class	0.000000
Flight Distance	0.000000
Seat comfort	0.000000
Departure/Arrival time convenient	0.000000
Food and drink	0.000000
Gate location	0.000000
Inflight wifi service	0.000000
Inflight entertainment	0.000000
Online support	0.000000
Ease of Online booking	0.000000
On-board service	0.000000
Leg room service	0.000000
Baggage handling	0.000000
Checkin service	0.000000
Cleanliness	0.000000
Online boarding	0.000000
Departure Delay in Minutes	0.000000
Arrival Delay in Minutes	0.003026

Fig.6. Null Değer Oranı

Figüre bakıldığında görülüyor ki null değerlerin oranı yaklaşık olarak %0,3 olarak görülüyor. Bu oran oldukça düşük olduğu için null değerleri drop edebiliriz. Bu işlemten sonra analizler daha doğru şekilde yapılacaktır.

```
df.dropna(inplace=True)
print("Null Değer Kontrolü")
df.isna().sum()
```

Fig.7. Null Değer Drop Etme ve Null Değer Kontrolü

Null değerleri drop ettikten sonra kontrol etmek için tekrar null değer kontrolü yapmak sağlıklı bir adım olacaktır. Bu adım Figür 7’de gösterilmiştir.


```

satisfaction      0
Gender            0
Customer Type     0
Age              0
Type of Travel    0
Class            0
Flight Distance   0
Seat comfort      0
Departure/Arrival time convenient 0
Food and drink    0
Gate location     0
Inflight wifi service 0
Inflight entertainment 0
Online support    0
Ease of Online booking 0
On-board service 0
Leg room service 0
Baggage handling 0
Checkin service  0
Cleanliness      0
Online boarding   0
Departure Delay in Minutes 0
Arrival Delay in Minutes 0
dtype: int64

```

Fig.8. Null Değer Kontrol Çıktısı

Yapılan adımdan sonra elde edilen çıktı incelendiği zaman görülüyor ki veri setinde artık herhangi bir null değer bulunmamaktadır.

Drop işleminden sonra veri setinde kaç adet satır kaldığını tekrar gözlemlemek gerekmektedir. Figür 9’da bu adım gösterilmektedir.

```

df.shape
(129487, 23)

```

Fig.9. Veri Seti Boyutu

Doğru değerlere ulaşıldıktan sonra artık istatistiksel analiz kısmına geçebiliriz.

Öncesinde biliyoruz ki veri setinde 17 integer, 5 object ve 1 float veri tipi bulunmaktaydı. Yani veri setinde numerik ve kategorik olarak iki farklı sınıf bulunmaktadır. Bu sınıflandırmanın veri setine manuel olarak girilmesi gerekmektedir. Bu işlem Figür 10’da gösterilmektedir.

```

numeric_col = df.select_dtypes(exclude = object) #
categoric_col = df.select_dtypes(include = object)

```

Fig.10. Veri Tipi Sınıflandırma

Sınıflandırmadan sonra kontrol ediyoruz ki numerik veriler 18 adet olarak incelediğimiz üzere ayrılmışlardır. Figür 11’de kontrol ve ayrımı görülmektedir.

```
print("Numeric Columns")
numeric_col.info()

Numeric Columns
<class 'pandas.core.frame.DataFrame'>
Int64Index: 129487 entries, 0 to 129879
Data columns (total 18 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Age                                       129487 non-null  int64
1   Flight Distance                         129487 non-null  int64
2   Seat comfort                            129487 non-null  int64
3   Departure/Arrival time convenient       129487 non-null  int64
4   Food and drink                          129487 non-null  int64
5   Gate location                           129487 non-null  int64
6   Inflight wifi service                   129487 non-null  int64
7   Inflight entertainment                  129487 non-null  int64
8   Online support                           129487 non-null  int64
9   Ease of Online booking                  129487 non-null  int64
10  On-board service                        129487 non-null  int64
11  Leg room service                        129487 non-null  int64
12  Baggage handling                        129487 non-null  int64
13  Checkin service                         129487 non-null  int64
14  Cleanliness                             129487 non-null  int64
15  Online boarding                         129487 non-null  int64
16  Departure Delay in Minutes              129487 non-null  int64
17  Arrival Delay in Minutes                129487 non-null  float64
```

Fig.11. Numerik Değişkenler

Aynı şekilde bahsedilen adım kategorik değişkenler için de uygulanmıştır. Figür 12’de de kategorik ayrım görülmektedir.

```
print("Categoric Columns")
categoric_col.info()

Categoric Columns
<class 'pandas.core.frame.DataFrame'>
Int64Index: 129487 entries, 0 to 129879
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   satisfaction    129487 non-null  object
1   Gender          129487 non-null  object
2   Customer Type   129487 non-null  object
3   Type of Travel  129487 non-null  object
4   Class           129487 non-null  object
```

Fig.12. Kategorik Değişkenler

Veri tipi ayrımı yaptıktan sonra artık istatistiksel analizi yapabiliriz. Bu analiz “describe” fonksiyonu ile yapılır. Fonksiyon kullanımı ve çıktısı Figür 13’de görülmektedir.

```
numeric_col.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Age	129487.0	39.428761	15.117597	7.0	27.0	40.0	51.0	85.0
Flight Distance	129487.0	1981.008974	1026.884131	50.0	1359.0	1924.0	2543.0	6951.0
Seat comfort	129487.0	2.838586	1.392873	0.0	2.0	3.0	4.0	5.0
Departure/Arrival time convenient	129487.0	2.990277	1.527183	0.0	2.0	3.0	4.0	5.0
Food and drink	129487.0	2.852024	1.443587	0.0	2.0	3.0	4.0	5.0
Gate location	129487.0	2.990377	1.305917	0.0	2.0	3.0	4.0	5.0
Inflight wifi service	129487.0	3.249160	1.318765	0.0	2.0	3.0	4.0	5.0
Inflight entertainment	129487.0	3.383745	1.345959	0.0	2.0	4.0	4.0	5.0
Online support	129487.0	3.519967	1.306326	0.0	3.0	4.0	5.0	5.0
Ease of Online booking	129487.0	3.472171	1.305573	0.0	2.0	4.0	5.0	5.0
On-board service	129487.0	3.465143	1.270755	0.0	3.0	4.0	4.0	5.0
Leg room service	129487.0	3.486118	1.292079	0.0	2.0	4.0	5.0	5.0
Baggage handling	129487.0	3.695460	1.156487	1.0	3.0	4.0	5.0	5.0
Checkin service	129487.0	3.340729	1.260561	0.0	3.0	3.0	4.0	5.0
Cleanliness	129487.0	3.705886	1.151683	0.0	3.0	4.0	5.0	5.0
Online boarding	129487.0	3.352545	1.298624	0.0	2.0	4.0	4.0	5.0
Departure Delay in Minutes	129487.0	14.643385	37.932867	0.0	0.0	0.0	12.0	1592.0
Arrival Delay in Minutes	129487.0	15.091129	38.465650	0.0	0.0	0.0	13.0	1584.0

Fig.13. İstatistiksel Analiz

Elde edilen bütün değişkenleri sırayla inceleyelim.

1. Age:

- Yaş ortalamasına baktığımızda yaklaşık olarak 39 değeri elde edilmiştir. Yaş ortalaması orta yaş olarak belirlenebilir.
- Minimum değer 7 olduğu görülmektedir. Anlaşıyor ki yolcular arasında küçük yaştaki çocuklar da bulunmaktadır.
- Veri setinin ilk çeyreğinde yaş değeri 27'dir.
- Ortanca değerin 40 olduğu belirlenmiştir. Yani yaş değerleri çoğunluk olarak küçük değerlerden oluşmaktadır.
- Veri setinin üçüncü çeyreğinde yaş değeri 51'dir. Veri setinin üst kısımlarında daha büyük yaş değerleri bulunmaktadır.
- Maximum yaş değeri 85 olarak elde edilmiştir.

2. Flight Distance:

- Mesafelerin ortalaması 1981'dir.
- “std” değerinin oldukça yüksek olduğu görülmekte yani uçuş mesafesi oldukça geniş bir yayılıma sahiptir.
- Minimum gidilen mesafe 50'dir.
- İlk çeyrekte gidilen mesafe 1359'dur.

- İkinci çeyrekte 1924 ve üçüncü çeyrekte 2543 mil olarak elde edilmiştir.
 - Maximum gidilen mesafe 6951'dir. Maximum değer ve çeyreklerden de anlaşıldığı üzere uçuş mesafesi genel olarak küçük 2500 ve altı değerlerden oluşmaktadır.
3. Seat Comfort:
- Ortalama değer yaklaşık olarak 3 olduğu görülmektedir. Ortalama bir konfor seviyesi var demektir.
 - Maximum ve minimum değerleri 5 ve 0 olduğunu zaten biliyoruz.
 - İlk çeyrekte değerlendirmeler 2, ikinci çeyrekte 3 ve üçüncü çeyrekte 4 olarak belirlenmiştir.
 - Üçüncü çeyrek değerinin 4 olduğu göz önüne alındığında veri setinin üst yarısının yüksek konfor puanlarına sahip olduğu söylenebilir.
4. Departure/Arrival time convenient:
- Seat Comfort değişkeni ile aynı değerlendirmeler yapılabilir.
5. Food and Drink:
- Seat Comfort değişkeni ile aynı değerlendirmeler yapılabilir.
6. Gate Location:
- Seat Comfort değişkeni ile aynı değerlendirmeler yapılabilir.
7. Inflight wifi service:
- Seat Comfort değişkeni ile aynı değerlendirmeler yapılabilir.
8. Inflight entertainment:
- Ortalama değeri 3.4 olduğu söylenebilir. Gayet yüksek bir ortalamadır.
 - İkinci çeyreğin 4 olduğu gözlemlendi. Yani veri setini üst yarısının daha yüksek eğlence puanlarına sahip olduğu söylenebilir.
9. Online support:
- Ortalama değeri 3.5 olduğu için oldukça yüksek değerlendirmeler vardır.
 - İlk çeyreğinde zaten elde edilen değer 3'tür. Bundan dolayı çoğu kullanıcının orta veya yüksek puanlar verdiği söylenebilir.
10. Ease of Online booking:
- İkinci çeyreği 4 olduğu için kullanıcıların en az yarısı 4 ve 5 değerlendirmeler yapmıştır.
11. On-board service:
- Inflight entertainment değişkeni ile aynı değerlendirmeler yapılabilir.

12. Leg room service

- Ease of Online booking değişkeni ile aynı değerlendirmeler yapılabilir.

13. Baggage handling:

- Burada diğer değişkenlerden farklı olarak minimum değer 1'dir. Yani 0 puan verecek kadar kötü bir hizmet verilmemiş demektir.
- İlk çeyrekteki değer 3 olduğu için müşterilerin yarısından fazlası bagaj taşıma hizmetine 3 üstü değerlendirme yapmışlardır. Genel bir memnuniyet var diyebiliriz.

14. Checkin service:

- Baggage handling değişkeniyle olan farkı minimum değerdir. Burada yine minimum değer 1'dir.
- İlk çeyrek 3 değerinde olduğu için yine müşterilerin yarısından fazlası 3 ve üstü puanlama yapmıştır dolayısıyla müşteriler check-in servisinden genel olarak memnun kalmışlardır.

15. Cleanliness:

- Puanlanabilen değişkenler arasında en yüksek ortalamaya sahip bir değişkendir. Müşterilerin genel olarak en memnun olduğu hizmetin temizlik olduğunu çıkarabiliriz.
- İlk çeyrekteki değer 3 olduğu için yarıdan fazla değerlendirmenin 3 ve üstü puanlama yapmıştır.

16. Online boarding:

- Bu değişken için yine min ve max değerler yine 0 ve 5 olduğu görülmüştür.
- İkinci çeyrekteki değer 4 olduğunu bulduk. Bundan dolayı müşterilerin en az yarısı 4 ve üstü puanlama yapmıştır.

17. Departure Delay in Minutes:

- Max ve min değerlere baktığımızda çok fark olduğu gözlemlendi. Buna dayanarak ortalama değer yaklaşık 15 olduğu bilindiği için küçük gecikmeler yaşandığı çıkarımı yapılabilir.
- İkinci çeyrekliğin 0 olduğu elde edildi. Bunu baz alarak veri setinin ortasında yer alan uçuşların genellikle plana uygun şekilde kalkış yaptığı söylenebilir.

18. Arrival Delay in Minutes:

- Departure Delay in Minutes değişkeni ile yaklaşık olarak aynı değerlendirmeler yapılabilir.

Aynı numerik değişkenlere yapıldığı gibi kategorik değişkenlere de aynı uygulama yapılacaktır. Yapılan uygulama ve çıktısı Figür 14’de gösterilmektedir.

```
categoric_col.astype('object').describe().T
```

	count	unique	top	freq
satisfaction	129487	2	satisfied	70882
Gender	129487	2	Female	65703
Customer Type	129487	2	Loyal Customer	105773
Type of Travel	129487	2	Business travel	89445
Class	129487	3	Business	61990

Fig.14. Kategorik Değişken Analizi

Görülüyor ki elde edilen değerlerde bazı sıklıkla geçen değerler mevcuttur. Sıklık değerleri yerine veri setinde bulunma yüzdeleri bulursak daha net değerlendirme yapabiliriz. Bu inceleme de Figür 15’de yapılmaktadır.

```
def freq_cal(column, category):
    category_data = df[df[column] == category]
    total_category_count = category_data.shape[0]
    percentage = (total_category_count / df.shape[0]) * 100
    return percentage

col_cat = {
    'Gender': 'Female',
    'satisfaction': 'satisfied',
    'Customer Type': 'Loyal Customer',
    'Type of Travel': 'Business travel',
    'Class': 'Business'
}

for column, category in col_cat.items():
    percentage = freq_cal(column, category)
    print(f"The percentage of \"{category.lower()}\" in the \"{column.upper()}\" category is: {percentage:.2f}%")

The percentage of "female" in the "GENDER" category is: 50.74%
The percentage of "satisfied" in the "SATISFACTION" category is: 54.74%
The percentage of "loyal customer" in the "CUSTOMER TYPE" category is: 81.69%
The percentage of "business travel" in the "TYPE OF TRAVEL" category is: 69.08%
The percentage of "business" in the "CLASS" category is: 47.87%
```

Fig.15. Sıklık Oranlarını Bulmak

- "satisfaction" kategorisi en fazla "satisfied" değeri içeriyor. Toplamda 70882 kişi "satisfied" olarak uçuşu sonlandırmış. Veri setindeki müşterilerin %50,74'ü havayolu şirketinden memnun olmuştur.
- "gender" kategorisinde en fazla olan değer "Female" değeridir. Bu değer veri setinin %54,74'üne eş düşmektedir. Müşterilerden 65703 kişi kadındır.
- "Customer Type" kategorisinde en fazla olan değer "Loyal Customer" olarak belirtilmiş. Bu kategori veri setinin %81,69'unu oluşturmaktadır. Yani müşterilerin %81,69'u sadık müşterilerden oluşturmaktadır.

- "Type of Travel" kategorisinde en fazla olan değer " Business Travel" olara ele alınmış. Bu kategori veri setinin %69.08'ini kaplamaktadır. Sonuç olarak havayolu şirketi ile yapılan uçuşların %69.08'i İş Gezisi olarak yer almıştır.
- "Class" kategorisinde en fazla olan değer "Business" olarak belirlenmiş. Bu kategori veri setinin %47,87'sini oluşturmaktadır. Müşterilerin %47,87'i uçuşlarını Business Class olarak yapmıştır.

Bu analizden sonra herhangi bir duplicate değer var mı kontrol ediliyor. Kontrol aşaması Figür 16'da görülmektedir.

```

duplicates = df[df.duplicated()]
print("Duplicate satır sayısı:", duplicates.shape[0])
Duplicate satır sayısı: 0

```

Fig.16. Duplicate Veri Kontrolü

EDA kısmından sonra artık elimizde temizlenmiş ve analizi yapılmış verinin görselleştirme aşamasına geçebiliriz.

3. Veri Görselleştirme

3.1 Satisfaction

Bu kısımda görünüyor ki müşterilerin %54,7'si havayolu şirketinden memnun ayrılmışlardır. Memnuniyetsizlik oranı da %45,3 büyük bir oranda yer almakta. Oluşturulan görsel ve elde edilen verileri Figür 17'de gösterilmektedir.

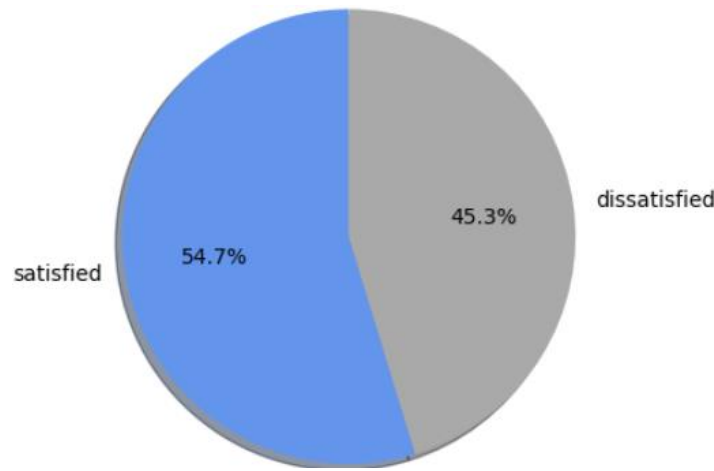


Fig.17. Satisfaction Pie Chart

3.2 Gender

Sırada müşterilerin cinsiyet dağılımı yer almakta. Müşterin %50.7'si kadınlardan, % 49.3'ü de erkeklerden oluşmaktadır. Buradan kadın müşterilerin biraz daha fazla olduğu sonucu ve cinsiyet dağılımının dengede olduğu çıkarımı yapılabilir. Değişkenin görseli ve elde edilen veriler Figür 18'de gösterilmektedir.

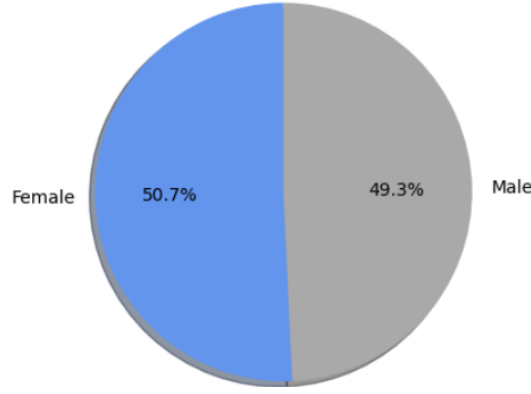


Fig.18. Gender Pie Chart

Şimdi bu cinsiyet dağılımının müşteri memnuniyetine etkisi nasıl olmuş gözlemleyelim. Yani kadınlardan ve erkeklerden kaç kişi memnun olarak veya memnun olmayarak ayrılmıştır cevabını bulacağız. Bu cevap için Figür 19'dan faydalanabiliriz.

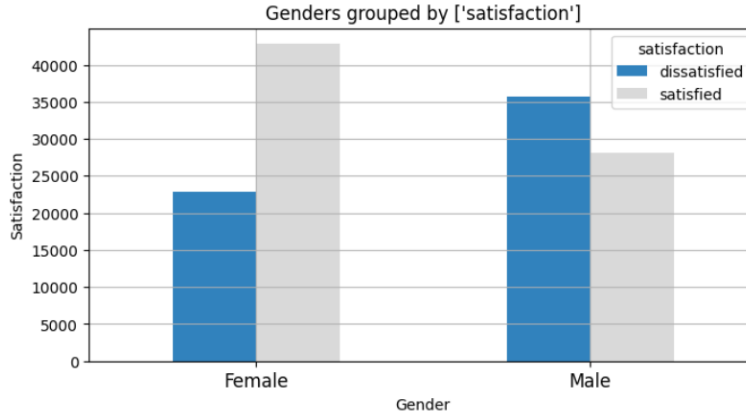


Fig.19. Satisfaction'a göre Gender Dağılımı

Görünen o ki kadınları çoğunlukla havayolu şirketinden memnun kaldığı görülmektedir. Erkeklerin ise tam tersi memnuniyetsizlik sayıları daha fazla olduğu görülmektedir. Kadınların erkeklere göre daha fazla memnun kaldığı yorumunu yapabiliriz.

3.3 Customer Type

Biliyoruz ki “Customer Type” iki değerden oluşmaktadır. Bu değerler “Loyal Customer” ve “Disloyal Customer” olarak yer almaktadır. Sadık müşterilerin kaçı memnun olmuştur kaçı memnun olmamıştır öğrenmek gerekmektedir. Bu incelemeyi yapmak için Figür 20’yi kullanabiliriz.

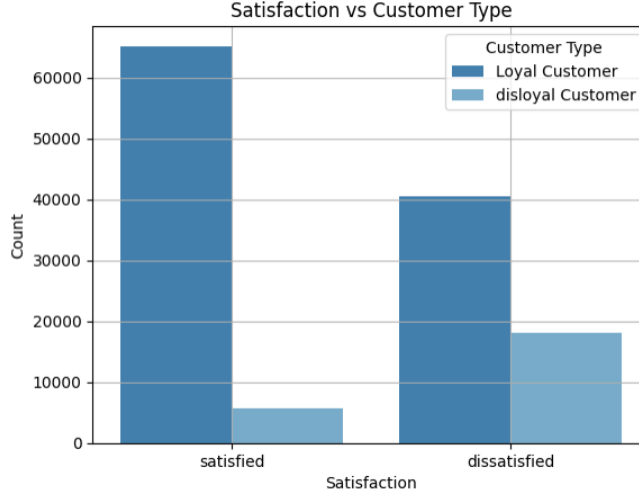


Fig.20. Customer Type vs Satisfaction

Grafiğe bakıldığında göze ilk çarpan şey “Loyal Customer” olan müşterilerin çok yüksek oranda memnun olduğudur. Aynı zamanda memnuniyetsizlik sayısı falan olan müşteri tipi yine “Loyal Customer” müşteri tipidir. “Disloyal Customer” olan müşterilerin ise memnun olmadıkları görülüyor. Ayrıca havayolu şirketinin en çok kazandığı müşterileri yeni müşteriler değil sadık müşterilerdir. Sadık müşteriler için özel teklifler yapılırsa memnuniyet oranlarında artış yakalanabilir. Yeni müşteriler için de farklı kampanyalar veya cazip teklifler yapılması durumunda yeni müşterinin hem bilet satışında hem de memnuniyet artışında yükselişi görmesi mümkün olabilir.

3.4 Age

Müşterilerin yaşları da bu projenin analizi için oldukça gereklidir. Age değişkenini kategorilendirip her kategorinin satisfaction sütununda dağılımını gösterebiliriz fakat elde edilen grafik sonrasında yorumlamak pek mümkün olmayabilir. Bunu yerine “Age” değişkenin “satisfaction” sütununa göre yoğunluk grafiğini elde etmek daha doğru bir yol olacaktır. Elde edilen grafik Figür 21’de gösterilmektedir.

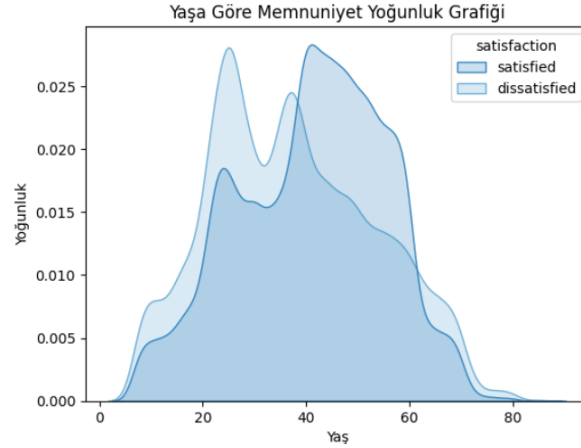


Fig.21. Yaş Göre Memnuniyet Yoğunluk Grafiği

Grafikten anladığımız üzere havayolu şirketinden memnun olan müşterilerin yaşları en çok 40 ve 60 yaşları arasında olan müşteriler olduğu ortaya çıkmıştır. Memnun olmayan müşterilerin ise yaşları genellikle 20 ve 40 yaşları arasında olan müşteriler olduğu gözlemlenmiştir. Genç müşteri sınıfında olan müşteriler için de kampanya, özel teklif veya eğlence seviyesini arttırabilecek hizmetlerin sağlanması genç müşterilerin daha fazla memnuniyet oranına erişebileceği söylenebilir.

3.5 Type of Travel

Seyahat türleri iki değerden oluşmaktadır. Müşteriler ya iş seyahati ya da kişisel seyahat yapmaktadır. Memnuniyetin en fazla sağlandığı seyahat türü bulunmalıdır. İstenilen grafik Figür 22’de gösterilmektedir.

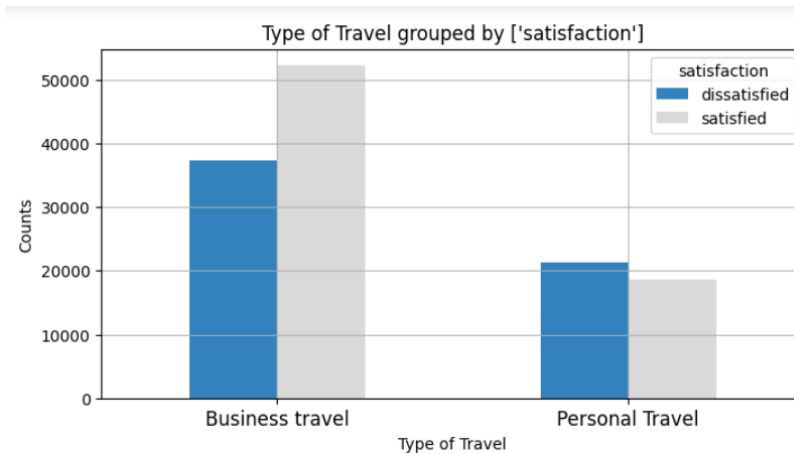


Fig.22. Satisfaction’a göre Type of Travel Dağılımı

Seyahatlerin çoğunluğunun iş seyahatleri olduğu görülüyor. Kişisel seyahat yapan müşterilerin neredeyse yarı yarıya memnun ve memnun olmadıklarını görebiliyoruz. İş seyahati yapan kişilerin de çoğunlukla seyahatlerinden memnun kaldığı görülmektedir. İş seyahatlerini yapan müşterilerin kaçının "loyal customer" sınıfında olduğuna bakarak bu müşterilerin memnuniyet oranlarının nasıl daha fazla arttırılabileceği konusunda daha iyi fikirler üretebiliriz. İstenilen analizin yapılması için gereken grafik Figür 23’de görülmektedir.

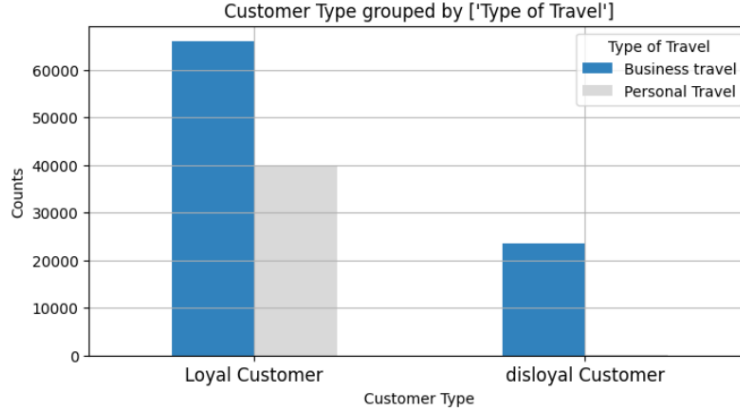


Fig.23. Type of Travel’a göre Customer Type Dağılımı

Görünen o ki sadık olmayan müşterilerin hepsi iş seyahati yapmış. Buradan sadık olmayan müşteriler için kişisel geziden daha çok iş seyahatleri için iyileştirmeler yapılabilir.

3.6 Class

Uçuş sınıflarının memnuniyet oranlarını bulmak analizi sağlamlaştıracaktır. Memnuniyet oranının uçuş sınıfına bir etkisi olmakta mı alınan hizmet farklılığı memnuniyet oranını etkilemekte mi gibi sorulara cevap bulmak gerekmektedir. Soruların cevaplarını Figür 24’e bakarak cevaplandırılabilir.

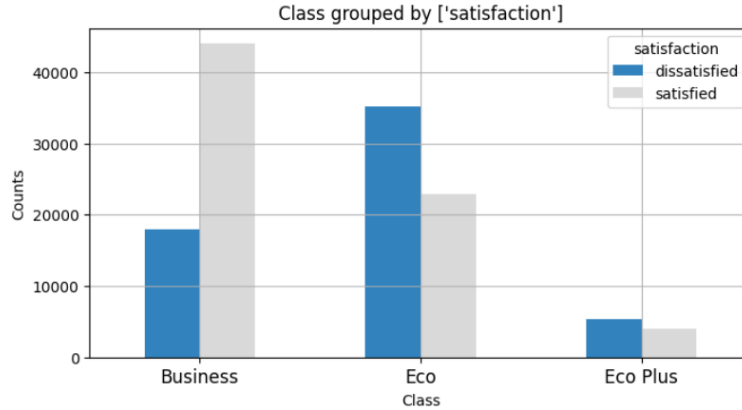


Fig.24. Satisfaction’a göre Class Dağılımı

Sorulan sorular üzere bakıldığında evet alınan hizmetin farklılığı da müşteri memnuniyetini etkilemektedir. Görülmektedir ki “Business Class” tercih eden müşterilerin memnuniyet oranları hayli yüksek görülmektedir. Aynı mantık çerçevesinden bakıldığında da “Eco Class” için de tam tersi bir yorum yapılabilir. Bu sınıfta da memnuniyetsizlik oldukça fazladır.

3.8 Flight Distance

Uçuş mesafesinin de incelenmesi gereken bir değişken olduğuna bazı sorularla varabiliriz. Uçuş mesafesi uzadıkça müşteri memnuniyeti ne şekilde değişmektedir, mesafe kısaldıkça memnuniyet oranı ne yönde ilerlemiştir gibi soruların cevaplarını Figür 25’de görebiliriz.

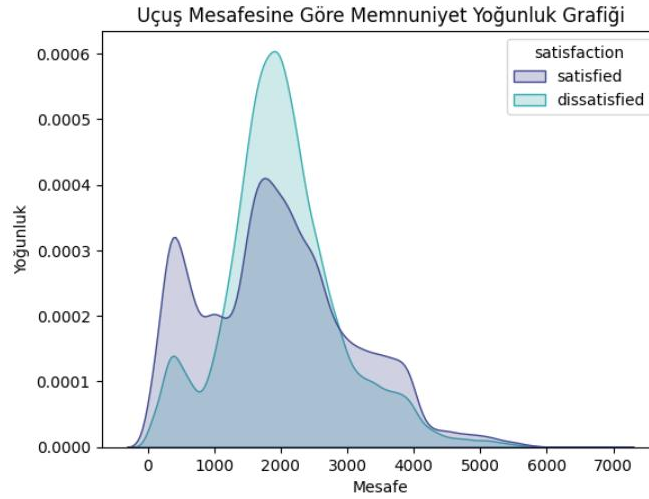


Fig.25. Uçuş Mesafesine Göre Memnuniyet Yoğunluğu

Grafikten anladığımız üzere memnuniyetin en fazla yoğunlukta olduğu aralık 2000 olduğu gözlemleniyor. Fakat aynı zamanda memnuniyetsizliğin de en yüksek olduğu aralık bu aralıktır. Bu aralıktan sonra mesafe arttıkça memnuniyet seviyesinin giderek azaldığını görebiliriz. Tahminler üzerine uçuş mesafesi kısaldıkça memnuniyet oranı artmakta ve yine aynı şekilde mesafe uzadıkça memnuniyetsizlik oranı yükselişe geçmektedir.

3.9 Derecelendirme Sütunları

Bahsettiğimiz üzere veri setimiz müşterilerin uçuş bilgileri ve buna dair bazı değerlendirmeleri mevcuttur. Fakat bu değerlendirmelerden 14 tanesi 0 ile 5 arasında derecelendirilme yapılan değerlendirmelerdir. Bu değişkenlerin sadece müşteri memnuniyeti üzerinde olan etkisini bilmek bu proje için yeterli olacaktır. Fakat bundan önce bir encoding işlemi yapılması gerekmektedir. Veri setinde artık satisfaction değişkeninde yer alan “dissatisfaction” artık “0” ve “satisfaction” ise “1” olarak yer alacaktır. Burada sadece birkaç örnek grafik göstermek yeterli olacaktır. Detaylı inceleme için proje sonunda verilen proje linklerinden inceleme yapılabilir.

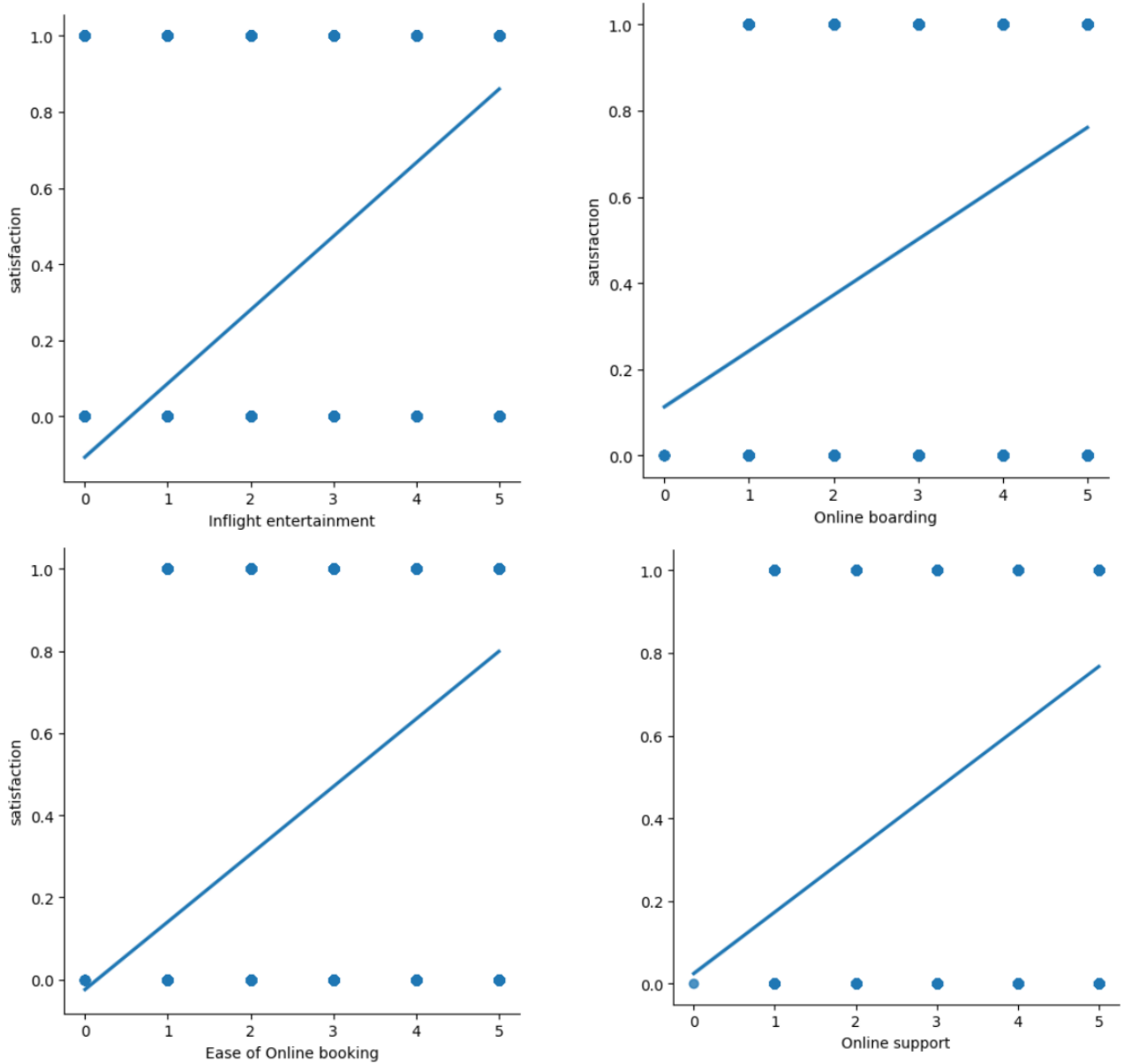


Fig.26. Etkili Derecelendirme Sütunları

Figür 26’da gösterilen 4 farklı grafik satisfaction üzerinde etkileri olan örnek değişkenlerdir. Etkili olan değişkenlerin listesi şu şekilde yapılabilir:

- Seat Comfort
- Inflight wifi service
- Inflight entertainment
- Online support
- Ease of online booking
- On-board service
- Leg room service
- Bagging handling
- Checkin Service
- Cleanliness
- Online Boarding

Etkili olmayan değişkenler ile olan farklılığını gözlemlemek için de etkili olmayan değişkenlerin örnek grafiklerini de Figür 27’de verilmektedir.

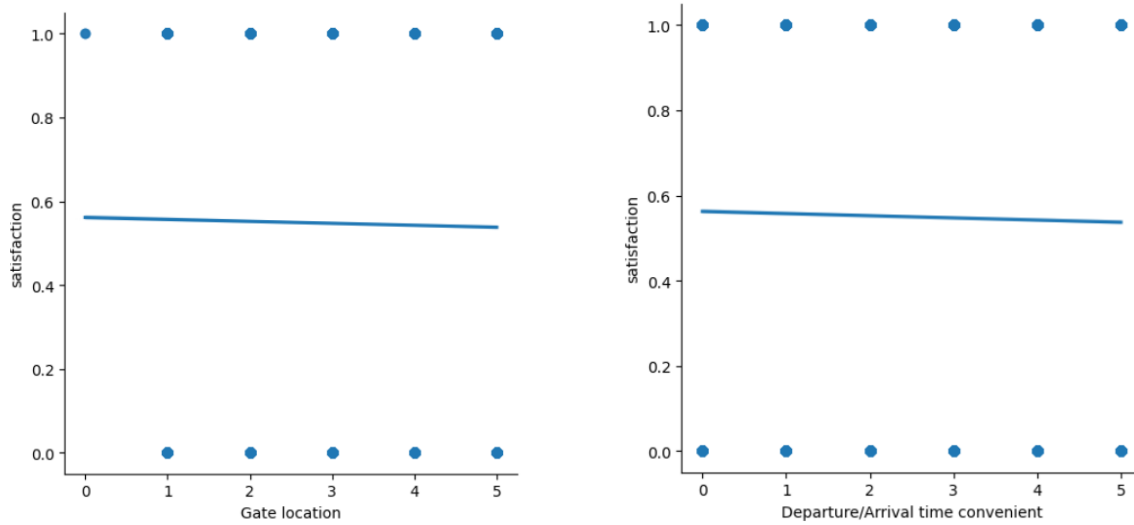


Fig.27. Etkili Olmayan Değişken Örnek Grafikleri

Etkili olmayan sadece 3 adet derecelendirme değişkeni bulunmaktadır. Bunlar:

- Gate Location
- Departure/Arrival Time Convenient
- Food and Drink

4. Correlation Analizi

Korelasyon analizi bize değişkenler arasında olan pozitif korelasyonu gösterecektir. Bu analizi heatmap ile göstereceğiz. Bu adımdan önce biliyoruz ki elimizde object veri tipinde olan değişkenler mevcuttu. Tekrar bahsetmek gerekirse bunlar 5 adetti fakat biz “satisfaction” için önceden encoding yapmıştık. Bu analiz için geri kalan 4 değişken için de yapılması gerekmektedir. Bu değişkenlerin hem kodları hem kendileri Figür 28’de gösterilmektedir.

```
df['Gender'] = pd.factorize(df['Gender'])[0]
df['Customer Type'] = pd.factorize(df['Customer Type'])[0]
df['Type of Travel'] = pd.factorize(df['Type of Travel'])[0]
df['Class'] = pd.factorize(df['Class'])[0]
```

Fig.28. Encoding

Yapılan korelasyon analizi Figür 29’da yer almaktadır. Yukarda bahsettiğimiz etki eden değişkenleri ayrıca heatmap üzerinden de bulabiliriz. "satisfaction" sütunu için pozitif korelasyon gösteren değerler net şekilde görülmektedir.

- | | |
|--------------------------|--------------------|
| • Seat Comfort | • Leg room service |
| • Inflight wifi service | • Bagging handling |
| • Inflight entertainment | • Checkin Service |
| • Online support | • Cleanliness |
| • Ease of online booking | • Online Boarding |
| • On-board service | |

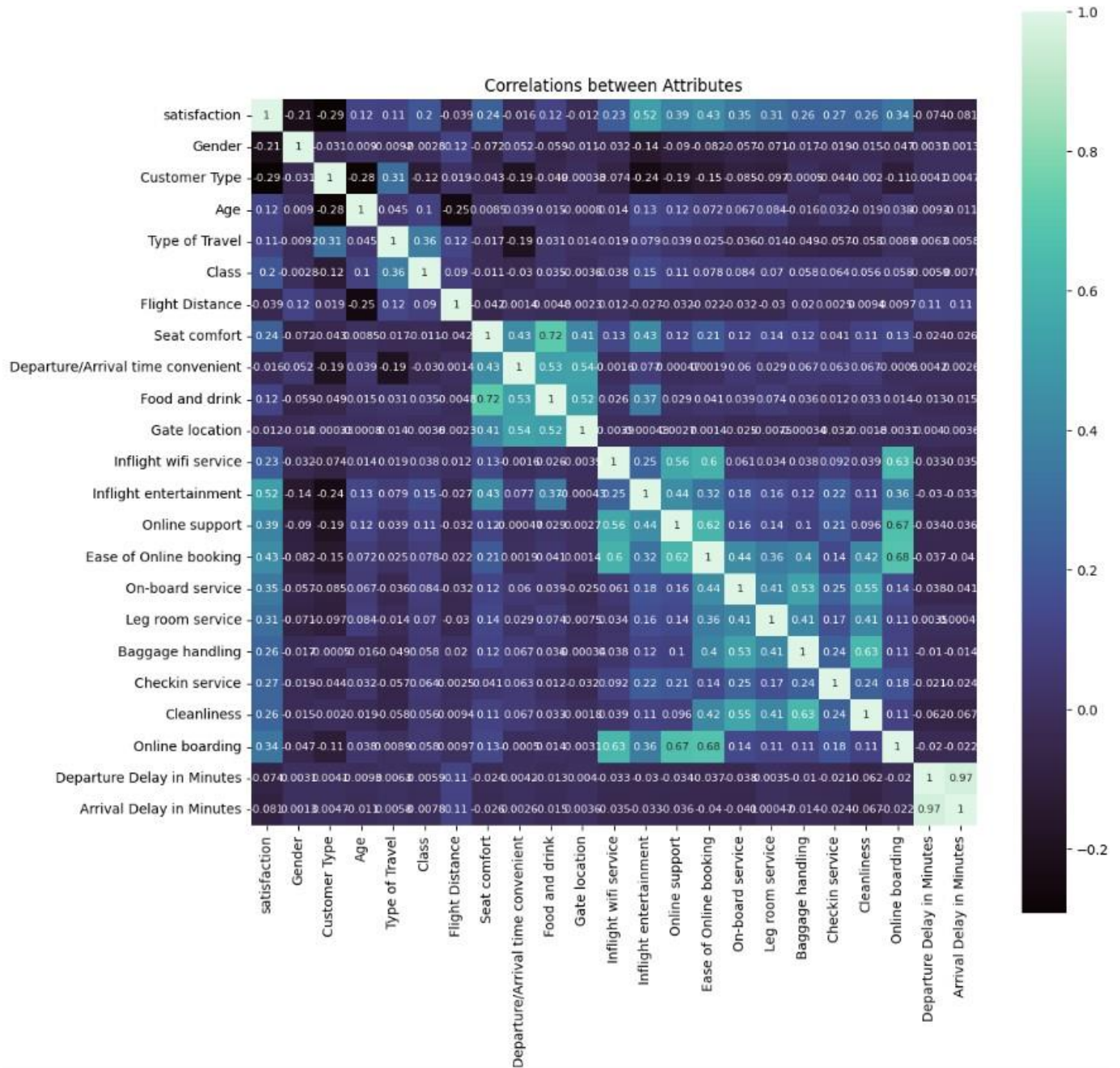


Fig.29. Değişkenler Arasındaki Korelasyon

Bahsettiğimiz değişkenlerin “Decision Tree Classification” modelinde önemli düğümler olmasını beklenebilir.

5. Modelleme

Modellemeye başlamadan önce kullanılacak train ve test setlerin oluşturulması ve “satisfaction” değişkeninin drop edilmiş olması gerekmektedir.

```
X = df.drop('satisfaction', axis=1)
y = df.satisfaction
```

Fig.30. “satisfaction” Drop Etme

Figür 30’da görüldüğü üzere sadece drop etme işlemi gerçekleştirildi. Ve drop edilen sütun da y değişkenine atandı. Bu işlemin doğru şekilde gerçekleşip gerçekleşmediğini öğrenmek için y değişkenini çağırıyoruz. Bu sorunun da cevabı Figür 31’de yer almaktadır.

```
y
0      1
1      1
2      1
3      1
4      1
..
129875  1
129876  0
129877  0
129878  0
129879  0
Name: satisfaction, Length: 129487, dtype: int64
```

Fig.31. y Değişken Kontrolü

Artık veri setinin test set ve train set olarak ayrılması yapılabilir. Bu sayede eğitim seti ve test seti oluşacak ve artık modelin öğrenmesine geçebileceğiz. Bu ayrımın kodunu Figür 32’de görebiliyoruz.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=12)
```

Fig.32. Test ve Train Set Oluşturulması

Veri setinin %30'u test seti ve %70' i train seti yani eğitim seti için ayrılmıştır. Elimizde artın X_test, X_train, y_test ve y_train olarak 4 farklı set bulunmakta. Sırayla bu setleri inceleyelim.

X_train																					
	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Seat comfort	Departure/Arrival time convenient	Food and drink	Gate location	...	Online support	Ease of Online booking	On-board service	Leg room service	Baggage handling	Checkin service				
52358	1	1	25	1	0	2502	3	3	3	1	...	4	2	4	5	5	5				
2840	1	0	32	0	0	1423	1	3	1	3	...	5	5	1	3	3	4				
82848	1	0	25	1	1	3773	2	2	2	2	...	3	3	2	5	5	3				
115679	0	0	53	1	1	1686	3	1	3	3	...	5	5	5	5	5	4				
48470	0	1	36	1	1	2932	2	2	2	4	...	4	4	3	2	3	3				
...				
40320	1	1	22	1	1	1432	0	0	0	3	...	1	1	5	4	5	2				
19785	0	0	10	0	0	2007	2	5	3	3	...	1	5	4	5	4	1				
124671	0	0	67	1	2	814	5	1	1	1	...	4	5	5	5	5	1				
104411	0	0	60	1	1	2642	4	4	2	4	...	5	4	4	4	4	5				
79935	1	0	25	1	1	3119	5	5	5	5	...	3	3	5	3	4	3				

90640 rows × 22 columns

Fig.33. X_train Seti

Figür 33'de de görüldüğü gibi X_train seti tam olarak istediğimiz şekilde oluşmuştur. Dikkat edilirse X_train setinde 90640 adet veri bulunmaktadır. “satisfaction” sütunu da artı X_train ve X_test veri setinin içerisinde yer almamaktadır.

X_test																					
	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Seat comfort	Departure/Arrival time convenient	Food and drink	Gate location	...	Online support	Ease of Online booking	On-board service	Leg room service	Baggage handling	Checkin service				
10408	1	0	9	0	0	1385	2	2	2	4	...	3	3	2	4	3	2				
54961	0	1	25	1	0	2453	3	3	3	4	...	1	1	4	4	4	3				
16796	0	0	35	0	2	1721	1	1	1	1	...	4	4	4	4	4	4				
74884	0	0	39	1	1	292	2	3	3	3	...	3	2	2	2	2	4				
95981	1	0	51	1	1	3026	4	4	4	4	...	5	4	4	4	4	4				
...				
36834	0	0	12	0	0	1168	4	5	4	1	...	1	1	3	4	5	4				
50753	1	1	38	1	0	1943	3	0	3	2	...	1	1	3	4	5	5				
22807	1	0	41	0	0	1413	3	3	3	1	...	4	4	5	4	1	3				
40482	1	1	27	1	1	1986	0	0	0	5	...	4	4	5	4	4	4				
75452	0	0	56	1	0	355	2	3	3	3	...	3	2	2	2	2	2				

38847 rows × 22 columns

Fig.34. X_test Seti

Figür 34'de gördüğümüz gibi yine aynı 22 sütun bulunmakta ve veri setinin geri kalan yani %30'u X_test setinde yer almaktadır. X_test seti içerisinde 38847 adet veri bulunmaktadır.

Aynı şekilde y_train seti içerisinde de daha önce belirttiğimiz üzere “satisfaction” değişkeni yer almaktadır. Aynı şekilde veri setinin %70’i yer almaktadır. Tıpkı X_train setinde olduğu gibi. Bahsedilen bilgiler Figür 35’de yer almaktadır.

y_train	
52358	0
2840	0
82848	1
115679	1
48470	0
..	
40320	1
19785	0
124671	1
104411	1
79935	1
Name: satisfaction, Length: 90640,	

Fig.35. y_train Set

4.1 Decision Tree

Decision Tree modellemesi için accuracy skoruna bakmak gerekmektedir. Decision Tree modellemesi ve elde edilen accuracy skorunun projedeki ekran görüntüsü Figür 36’da yer almaktadır.

```
dtc = DecisionTreeClassifier(random_state=42)

dtc.fit(X_train, y_train)

DecisionTreeClassifier
DecisionTreeClassifier(random_state=42)

y_dtc_test_dt = dtc.predict(X_test)
y_dtc_train_dt = dtc.predict(X_train)

dt_acc= accuracy_score(y_test, y_dtc_test_dt)
print('Accuracy by Decision Tree : ',dt_acc)

Accuracy by Decision Tree : 0.9348984477565835
```

Fig.36. Decision Tree Modelleme ve Accuracy Skoru

Modelin performansı görüldüğü üzere 0,93 olarak elde edilmiştir. Bu demek oluyor ki verilen setin %93’ü doğru tahmin edilecektir. Oldukça yüksek bir değer elde ettik. Bu modelin Classification Report’una Figür 37’de yer almaktadır.

Classification report for Decision Tree (Test set)=				
	precision	recall	f1-score	support
0	0.92	0.93	0.93	17443
1	0.94	0.94	0.94	21404
accuracy			0.93	38847
macro avg	0.93	0.93	0.93	38847
weighted avg	0.93	0.93	0.93	38847

Fig.37. Decision Tree Classification Report

4.2 Logistic Regression

Decision Tree modelinde olduğu gibi aynı şekilde Logistic Regression modeline de aynı adımlar yapılmıştır. Logistic Regression modelinin accuracy skoru Figür 38’de yer almaktadır.

```
X_test_prediction = model.predict(X_test)
lr_acc = accuracy_score(X_test_prediction, y_test)
print("Accuracy score for test data: ", lr_acc)
```

Accuracy score for test data: 0.834942209179602

Fig.38. Logistic Regression Accuracy Skoru

İki modelin accuracy skorlarını net şekilde karşılaştırmak için bir grafik oluşturmak çok mantıklı bir adım olacaktır. Bu grafiği Figür 39’da görülmektedir.



Fig.39. Modellerin Accuracy Skorları

Bu grafik sonrasında daha net anlaşıyor ki Decision Tree modeli kullanılan veri seti için en doğru bir seçim olacaktır.

Modelimizin testi içi 10 adet sentetik veri oluşturuyoruz. Bu sentetik verilerde 22 sütun için veriler yer almaktadır. Sadece “satisfaction” sütunu için olan veriler girilmemiştir. Bu sentetik veriler Figür 40’da yer almaktadır.

```
sentetic_data = [[0, 1, 24, 0, 1, 500, 4, 2, 3, 4, 2, 4, 3, 4, 2, 2, 3, 4, 4, 5, 0, 0],
[1, 0, 44, 1, 0, 5000, 2, 1, 2, 2, 1, 1, 0, 3, 1, 5, 5, 2, 2, 4, 10, 100],
[0, 1, 20, 1, 2, 1000, 2, 4, 1, 5, 5, 2, 4, 1, 5, 1, 4, 2, 5, 1, 100, 150],
[1, 1, 10, 0, 1, 2500, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 40, 45],
[0, 0, 60, 1, 0, 982, 3, 2, 3, 4, 2, 4, 3, 4, 2, 2, 3, 2, 3, 2, 144, 150],
[0, 1, 34, 0, 1, 754, 2, 4, 1, 4, 0, 2, 4, 0, 4, 3, 5, 2, 3, 4, 0, 0],
[0, 1, 13, 1, 2, 658, 4, 3, 4, 3, 5, 4, 2, 4, 4, 2, 1, 2, 3, 1, 190, 180],
[1, 1, 54, 1, 0, 590, 1, 1, 2, 5, 4, 1, 2, 1, 3, 2, 2, 5, 5, 1, 155, 162],
[1, 0, 37, 0, 2, 940, 5, 5, 4, 4, 2, 5, 2, 4, 3, 2, 4, 4, 4, 2, 1000, 900],
[0, 0, 77, 0, 1, 2000, 2, 5, 4, 4, 4, 4, 3, 5, 5, 2, 2, 4, 4, 4, 2, 0, 0]]
```

Fig.40. Sentetik Veriler

Oluşturduğumuz bu verileri Decision Tree modelinde artık predict edebiliriz. Predict adımı Figür 41’de gösterilmektedir.

```
pred = dtc.predict(sentetic_data)
print(pred)

for i in range(len(pred)):
    print("For ", i+1, ". customer:")
    if pred[i] == 0:
        print("Customer is dissatisfied\n")
    else:
        print("Customer is satisfied\n")
```

Fig.41. Decision Tree Predict

Bu adım sonrasında bir çıktı elde ediyoruz elbette. Buradaki kod bloğunda yapılan işlem verilen sentetik verilerin satisfaction durumlarının tahminidir. Müşterilerin memnun olduğu durumda yazdırılacak olan şey “Customer is satisfied.”dır ve eğer müşteri memnun değilse “Customer is dissatisfied” yazdırılacaktır. Çıktı Figür 42’de gösterilmektedir.

```

[1 1 0 1 0 0 0 0 1 1]
For 1 . customer:
Customer is satisfied

For 2 . customer:
Customer is satisfied

For 3 . customer:
Customer is dissatisfied

For 4 . customer:
Customer is satisfied

For 5 . customer:
Customer is dissatisfied

For 6 . customer:
Customer is dissatisfied

For 7 . customer:
Customer is dissatisfied

For 8 . customer:
Customer is dissatisfied

For 9 . customer:
Customer is satisfied

For 10 . customer:
Customer is satisfied

```

Fig.42. Sentetik Veri Tahmini

Elde ettiğimiz çıktıya baktığımızda 5 müşterinin “satisfied” olduğu geri kalan 5 kişinin de “dissatisfied” olduğu görülmektedir.

Sentetik verilerin dışında gerçek veri setinden alınan iki veri satırını kullanarak modelin doğru çalışıp çalışmadığını kontrol edebiliriz. Bu anlatılan adım Figür 43’de gösterilmektedir.

```

sen2= [[0,1,65,0,0,265,0,0,0,2,2,4,2,3,3,0,3,5,3,2,0,0], ## real datas
        [1,1,47,0,1,2464,0,0,0,3,0,2,2,3,4,4,4,2,3,2,310,305]]

pred2 = dtc.predict(sen2)
print(pred2)

[1 1]

```

Fig.43. Gerçek Veri Tahmini

Figür 43’de görülüyor ki veri setinde bulunan ilk iki müşterinin satifaction durumunu tahmin etmiştir ve öyle görünüyör ki sonuç her iki müşterinin de memnun kaldığı görülmektedir.

```
results = pd.DataFrame(
{
    "Predicted": y_dtc_test_dt,
    "Truth": y_test,
    "Accuracy": y_dtc_test_dt == y_test
})
```

Fig.44. Test Set Prediction Kodu

Figür 44'deki kod bloğunda görülüyor ki test seti modele tahmin etmesi için gönderdik. Ve Figür 45'de bulunan sonucu incelediğimizde çoğunlukla doğru tahmin yaptığı görülmektedir.

	Predicted	Truth	Accuracy
10408	0	0	True
54961	0	0	True
16796	1	1	True
74884	0	0	True
95981	1	1	True
...
36834	1	1	True
50753	0	0	True
22807	0	0	True
40482	1	1	True
75452	0	0	True

38847 rows × 3 columns

Fig.45.Test Set Prediction Sonucu

Bu demek oluyor ki test set içerisinde bulunan 38847 verinin %93'ü doğru tahmin edilmiştir.

Aynı işlemi train set için de uygulandı. Ve elde edilen sonuç ve işlemin kodu Figür 46’da gösterilmiştir. Bu demek oluyor ki train set içerisinde bulunan 90640 verinin %93’ü doğru tahmin edilmiştir.

```
results = pd.DataFrame(  
{  
    "Predicted": y_dtc_train_dt,  
    "Truth": y_train,  
    "Accuracy": y_dtc_train_dt == y_train  
})
```

	Predicted	Truth	Accuracy
52358	0	0	True
2840	0	0	True
82848	1	1	True
115679	1	1	True
48470	0	0	True
...
40320	1	1	True
19785	0	0	True
124671	1	1	True
104411	1	1	True
79935	1	1	True

90640 rows x 3 columns

Fig.46.Train Set Prediction Kodu ve Sonucu

5. Conclusion

Veri setinin analizi için gerekli bütün adımları tamamlamış olduk. Artık herhangi bir müşterinin uçuş bilgilerini girerek memnun olup olmadığını kontrol edebiliriz. Proje boyunca yapılan analizlerle gelecek şirket politikasını ya da müşteri profillerine göre kampanya ve özel teklifler hazırlayarak müşteri memnuniyetini yükseltme adımları atılabilir. Yapılan analizlerde müşteri memnuniyetini etkileyen en büyük özellik “Inflight Entertainment” sisteminin olması ve geliştirilmesidir. Müşteri memnuniyeti bu şekilde daha fazla elde edilebilir. Ayrıca daha uygun maliyetli geliştirme olarak “Seat Comfort” yer alabilir. Koltuk konforunu arttırmak veya koltuğa entegre kafa desteği gibi iyileştirmeler yapılarak bu alanda da müşteri memnuniyeti elde edilebilir. Web sitesi iyileştirmesi yapılarak “Ease of online booking” ve “Online Support” hizmetleri geliştirilebilir. Bu sayede müşteri memnuniyetinde yeniden artış yakalanabilir. Hosteslere verilecek ekstra farklı eğitimler sonrasında “On-boarding service” hizmetinde de yükseliş sağlanabilir.

Proje Bağlantıları:

Veri Seti Linki: <https://www.kaggle.com/datasets/sjleshrac/airlines-customer-satisfaction>

Kaggle Proje Linki: <https://www.kaggle.com/nursenaertugrul/airlines-customer-satisfaction-prediction>

Github Proje Linki: <https://github.com/NursenaErtugrul/Data-Science-With-Python-Bootcamp-Final-Project>