



gezinomi
her gün tatil olsa...



Otel Satış Fiyatlarının Modellemesi

miuul

İş Problemi

Gezinomi'nin gerçekleştirdiği satış verilerini kullanarak, gelecekteki müşterilerinin ortalama otel ücretini tahmin etmek istiyoruz.

Örneğin:

İzmir'de yoğun bir dönemde yarım pansion bir otele gitmek isteyen bir müşterinin ödeyeceği tahmini ücreti belirlemek istiyoruz.

► İŞ PROBLEMI
VERİ SETİ

DEĞİŞKENLER
GÖREVLER



Veri Seti Hikayesi

gezinomi_miuul.xlsx veri seti Gezinomi şirketinin yaptığı satışların fiyatlarını ve bu satışlara ait bilgiler içermektedir. Veri seti her satış işleminde oluşan kayıtlardan meydana gelmektedir. Bunun anlamı tablo tekilleştirilmemiştir. Diğer bir ifade ile müşteri birden fazla alışveriş yapmış olabilir.



iŞ PROBLEMi

› VERİ SETİ

DEĞİŞKENLER

GÖREVLER

Değişkenler

miuul_gezinomi.xlsx

SaleId: Satış id

SaleDate: Satış Tarihi

CheckInDate: Müşterinin otele giriş tarihi

Price: Satış için ödenen fiyat

ConceptName: Otel konsept bilgisi

SaleCityName: Otelin bulunduğu şehir bilgisi

CInDay: Müşterinin otele giriş günü

SaleCheckInDayDiff: Check in ile giriş tarihi gün farkı

Season: Otele giriş tarihindeki sezon bilgisi

İŞ PROBLEMİ

VERİ SETİ

DEĞİŞKENLER

GÖREVLER



Proje Görevleri

Görev 1: Veri Setine Genel Bir Bakış

1.1 miuul_gezinomi.xlsx dosyasını okutunuz.

1.2 explore_data isminde bir fonksiyon oluşturun. Bu fonksiyon veri seti hakkında aşağıdaki çıktıları vermeli:

- **İlk 5 gözlem birimi**
- **Değişken adları ve türleri**
- **Boş değerlerin sayısı**
- **Bellek kullanımı**
- **Nümerik değişkenler için temel istatistikler**
- **Veri setinin boyutu**

1.3 Boş değerlerin sayısı ile veri setini büyüklüğünü kıyaslayın. Veri setinde önemli bir bilgi kaybı olmayacağı bir durum ise boş değerlerin bulunduğu gözlem birimlerini silin.

Görev 2: Özellik Mühendisliği

2.1 Check-in tarihini kullanarak aşağıdaki değişkenleri türetiniz:

- "CheckInYear": Check-in yapıldığı yıl
- "CheckInMonth": Check-in yapıldığı ay
- "CheckInDay": Check-in yapıldığı gün
- "CheckInWeekday": Check-in yapıldığı gün haftanın kaçinci günü
- "IsWeekend": Check-in yapıldığı günün hafta sonu olup olmadığı
- "HolidayTypes": Veri setindeki check-in yapılan en eski tarih ile en yeni tarih arasındaki Türkiye'deki resmi tatiller. **Örneğin, 2016-08-30 - Zafer Bayramı, 2022-07-09 - Kurban Bayramı**

2.2 Satış ile Check-in tarihleri arasındaki farkı ele alarak "CustomerType" isimli bir değişken oluşturun. "SaleCheckInDayDiff" değişkeni, müşterinin Check-in tarihinden ne kadar önce satın alımını tamamladığını gösterir. Aralıkları aşağıdaki gibi oluşturabilirsiniz:
Örneğin, '**0_7**', '**7_30**', '**30_90**', '**90_max**' gibi aralıkları kullanabilirsiniz. Bu aralıklar için "**Last Minuters**", "**Potential Planners**", "**Planners**", "**Early Bookers**" gibi isimleri kullanabilirsiniz.

2.3 Kategorik değişkenler olan '**ConceptName**', '**SaleCityName**', '**CInDay**', '**Seasons**', '**HolidayType**' ve '**CustomerType**' için one-hot encoding işlemi uygulayın.

Görev 3: PCA ile Boyut Azaltma

3.1 Hedef değişken olan '**Price**' ile bağımsız değişkenleri ayırin.

3.2 Boyut azaltma işlemi için '**sklearn.decomposition**' modülündeki PCA fonksiyonunu kullanın ve Principal Component sayısını **5** olarak belirleyin.

3.3 Boyut azaltılmış veriyi DataFrame tipine dönüştürün.

Görev 4: Lineer Regresyon Modeli Kurma.

4.1 Veriyi '**train_test_split**' fonksiyonunu kullanarak train ve test setlerine ayırin. Test veri setinin boyutunu %20 olarak belirleyin.

4.2 '**sklearn.linear_model**' modülündeki '**LinearRegression**' fonksiyonunu kullanarak modeli kurun ve eğitim seti için modele uydurun.

4.3 Test setini kullanarak tahminler yapın ve modelin başarısını **MSE**, **MAE**, **R2** gibi metriklerle değerlendirin.

4.4 Modelin katsayılarını sırasıyla ekrana yazdırın.

4.5 Derslerde öğrendiğimiz "**en_kucuk_kareler_yontemi**" adlı fonksiyonu kullanarak modelin katsayılarını yeniden hesaplayın ve **LinearRegression** hazır fonksiyonu ile elde ettiğiniz katsayılarla karşılaştırın.

miuul

miuul.com