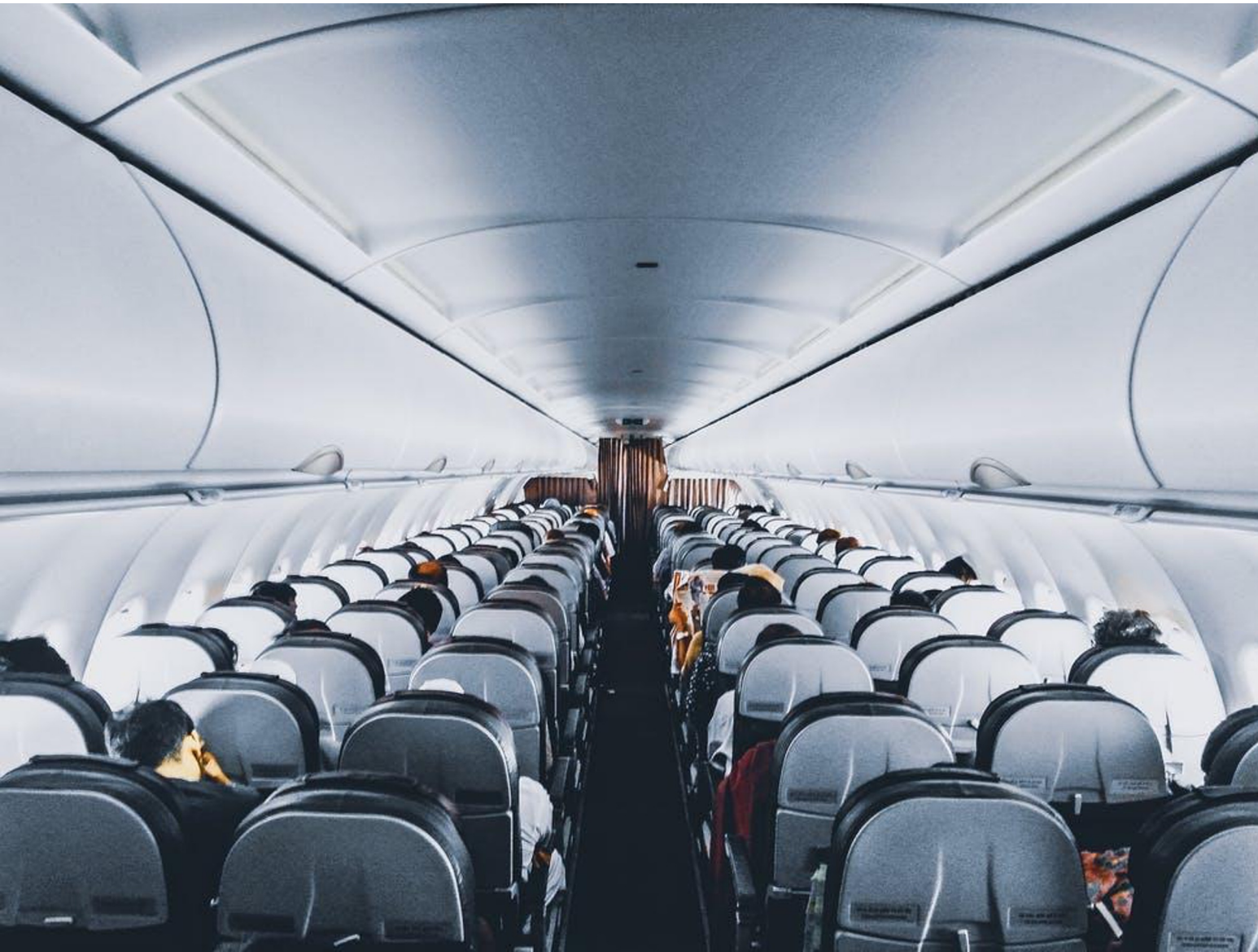




Makine Öğrenmesi ile Yolcu Sayısı Tahmini





İş Problemi

Türk Hava Yolları uçuşlarda bulunan sezonluk rezervasyonlu yolcu sayısını tahmin edebilmek için bir makine öğrenmesi regresyon modeli kurulmasını istemektedir.

Veri Seti Hikayesi

Türk Hava Yollarına ait bu veri seti 1 aktarmalı (2 leg) sezonluk origin & destination bazında uçuş bilgilerini içermektedir.

17 Değişken

18.1173 Gözlem

12 KB

CARRIER	Taşıyıcı firmanın ikili kodudur. 2 leg için de taşıyıcı firma aynıdır.
AIRCRAFT_TYPE	Uçak tipi bilgisidir.
OND_SELL_CLASS	En uzun legin satış sınıfıdır.
LEG1_SELL_CLASS	1.leg in satış sınıfıdır.
OND_CABIN_CLASS	En uzun leg in kabin sınıfıdır.
LEG1_CABIN_CLASS	1.leg in kabin sınıfıdır.
HUB	1.leg uçuşundan 2.leg uçuşuna aktarma yapılan havalimanının IATA kodlarını içerir.
DETUR_FACTOR	Bir yolcunun uçmak istediği noktaya direkt uçmak yerine bağlantılı uçuşu durumunda yolunu ne kadar uzattığını ifade eden bir orandır.
CONNECTION_TIME	1.leg uçuşundan 2.leg uçuşuna aktarma için beklenen süreyi dakika cinsinden ifade eder.
PSGR_COUNT	Toplam rezervasyonlu yolcu sayısıdır.
LEG1_DEP_FULL	1. Leg'in kalkış tarih ve saatidir.
LEG1_ARR_FULL	1. Leg'in varış tarih ve saatidir.
LEG2_DEP_FULL	2. Leg'in kalkış tarih ve saatidir.
LEG2_ARR_FULL	2. Leg'in varış tarih ve saatidir.
LEG1_DURATION	1. Leg'in uçuş süresi.
LEG2_DURATION	2. Leg'in uçuş süresi.
FLIGHT_DURATION	1. Leg'in kalkış saatinden 2. Leg'in iniş saatine kadar geçen zaman.

Leg: 1 uçuşu temsil eder. 2 leg uçuş demek, 1 aktarmalı 2 adet uçuşu olan yolculuktur.

Örnek: İstanbul - San Francisco uçuşunda Londra aktarması var ise; İstanbul - Londra 1. leg, Londra - San Francisco 2. legdir.



Proje Görevleri

Adım 1: thyTrain.csv dosyasını okutunuz.

Adım 2: Target değişkeninin (PSGR_COUNT) standart sapma ve ortalama bilgilerini getiriniz.

Adım 3: Target değişkenini büyükten küçüğe sıralayıp veri setinin ilk 10 satırını getiriniz.

Adım 4: LEG1_DEP_FULL, LEG1_ARR_FULL, LEG2_DEP_FULL, LEG2_ARR_FULL değişkenlerinin min ve max değerlerini getiriniz.

Adım 5: Her değişkendeki boş değer sayısını yazdırınız ve boş değer bulunan gözlemleri o değişkenin modu ile doldurunuz.

Görevler

Adım 6: Aşağıda belirtilen değişkenleri oluşturunuz.

Değişken Adı	Değişken Açıklaması
LEG1_DEP_MONTH	LEG1_DEP_FULL değişkeninin ay bilgisi
LEG1_DEP_HOUR	LEG1_DEP_FULL değişkeninin saat bilgisi
LEG1_ARR_MONTH	LEG1_ARR_FULL değişkeninin ay bilgisi
LEG1_ARR_HOUR	LEG1_ARR_FULL değişkeninin saat bilgisi
LEG2_DEP_MONTH	LEG2_DEP_FULL değişkeninin ay bilgisi
LEG2_DEP_HOUR	LEG2_DEP_FULL değişkeninin saat bilgisi
LEG2_ARR_MONTH	LEG2_ARR_FULL değişkeninin ay bilgisi
LEG2_ARR_HOUR	LEG2_ARR_FULL değişkeninin saat bilgisi
LEG1_DEP_DAY	LEG1_DEP_FULL değişkeninin haftanın günü bilgisi (Pazartesi, Salı,...)
LEG1_ARR_DAY	LEG1_ARR_FULL değişkeninin haftanın günü bilgisi (Pazartesi, Salı,...)
LEG2_DEP_DAY	LEG2_DEP_FULL değişkeninin haftanın günü bilgisi (Pazartesi, Salı,...)
LEG2_ARR_DAY	LEG2_ARR_FULL değişkeninin haftanın günü bilgisi (Pazartesi, Salı,...)
LEG1_DURATION_MINUTES	LEG1_DURATION değişkenini dakika cinsine çeviriniz. (01:50 -> 110)
LEG2_DURATION_MINUTES	LEG2_DURATION değişkenini dakika cinsine çeviriniz. (01:50 -> 110)
FLIGHT_DURATION_MINUTES	FLIGHT_DURATION değişkenini dakika cinsine çeviriniz. (01:50 -> 110)
FLIGHT_DURATION_MINUTES_FLIGHTS	Aktarma için havalimanında bekleme süresi hariç uçuş süresi
FLIGHT_DURATION	Aktarma dahil tüm uçuşun dakika cinsinden süresinin, aktarma hariç uçuş süresine oranı
LEG1_RATIO	LEG1_DURATION değişkeninin dakika cinsinden değerinin, tüm uçuşun dakika cinsinden değerine oranı
LEG2_RATIO	LEG2_DURATION değişkeninin dakika cinsinden değerinin, tüm uçuşun dakika cinsinden değerine oranı

Not: Haftanın günü ile ilgili değişkenleri **calendar** kütüphanesinin **weekday()** metodu ile oluşturabilirsiniz.

Değişiklik yapılan değişkenleri datetime’a çevirmeyi unutmayınız.

Adım 6.1: "LEG1_DEP_FULL", "LEG1_ARR_FULL", "LEG2_DEP_FULL", "LEG2_ARR_FULL", "LEG1_DURATION", "LEG2_DURATION", "FLIGHT_DURATION" değişkenlerini dataframeden kaldırınız.

Adım 7: grab_col_names fonksiyonunu kullanarak değişkenleri sınıflandırınız.

Adım 7.1: num_cols değişkeninden PSGR_COUNT target değişkenini list comprehension kullanarak kaldırınız.

Adım 8: Outlier değerleri baskılayınız.

Adım 8.1: Rare sınıfları tek bir sınıf haline getiriniz. (rare_perc = 0.05)

Adım 8.1: Kategorik değişkenlere One Hot Encoding uygulayınız.

Adım 8.3: Standard Scaler işlemini uygulayınız.

Adım 9: Cross Validation kullanarak makine öğrenmesi modellerini eğitiniz ve RMSE değerlerini getiriniz.

Adım 9.1: Belirli algoritmalar üzerinde GridSearchCV uygulayın ve RMSE değerlerini getiriniz.

Adım 10: Değişkenlerin önem düzeyini belirten feature_importance fonksiyonunu kullanarak özelliklerin sıralamasını çizdiriniz.

miuul

miuul.com