

DOĞRUSAL REGRESYON MODELİ İLE OTOMOTİV SATIŞ TAHMİNİ

Presented By Mert Dil

25.04.2023

AGENDA

Giriş	3
Veri Keşfi ve Ön İşleme	4-5
Varsayımların Test Edilmesi	6
Model Oluşturma	7
Mevsimsel Değişkenlerle Model Oluşturma	8
Modellerin Karşılaştırılması	9
Modelin Servis Haline Getirilmesi ve Test Edilmesi	10-11



Çalışmanın Arka Planı



- Otomotiv sektörü, Türkiye ekonomisi için önemli bir yerdedir.
- Bu sektördeki Otomotiv Satış tahmini, firmaların planlama ve karar verme süreçlerinde kritik öneme sahiptir.
- Bu nedenle, otomotiv satış tahmini için veri bilimi yöntemleri ile bir model oluşturmak, sektördeki firmalar için büyük bir avantaj sağlayabilir.
- Bu çalışma, çoklu doğrusal regresyon modeli kullanarak Haziran 2022 - Haziran 2023 dönemi için otomotiv satışlarının tahminini amaçlamaktadır.
- Veri keşfi, model oluşturma ve modelin servis haline getirilmesi adımları takip edilerek bir çözüm geliştirilmiştir.

VERİ KESFİ VE ÖN İŞLEME

[Back to Agenda](#)

	Otomotiv Satis	OTV Orani	Faiz	EUR/TL	Kredi Stok
count	149.000000	149.000000	149.000000	149.000000	1.490000e+02
mean	65902.617450	43.697987	16.034380	4.808703	1.787554e+06
std	26175.801077	5.846811	5.319518	3.321767	1.176366e+06
min	14373.000000	37.000000	9.607500	1.927305	3.412441e+05
25%	50008.000000	40.000000	12.715000	2.464018	7.755455e+05
50%	63044.000000	45.000000	14.485000	3.276861	1.474594e+06
75%	79819.000000	50.000000	17.475000	6.380609	2.447479e+06
max	156173.000000	65.000000	32.782500	16.560000	4.862887e+06

Table 1. Verimizin İstatiksel Tablosu

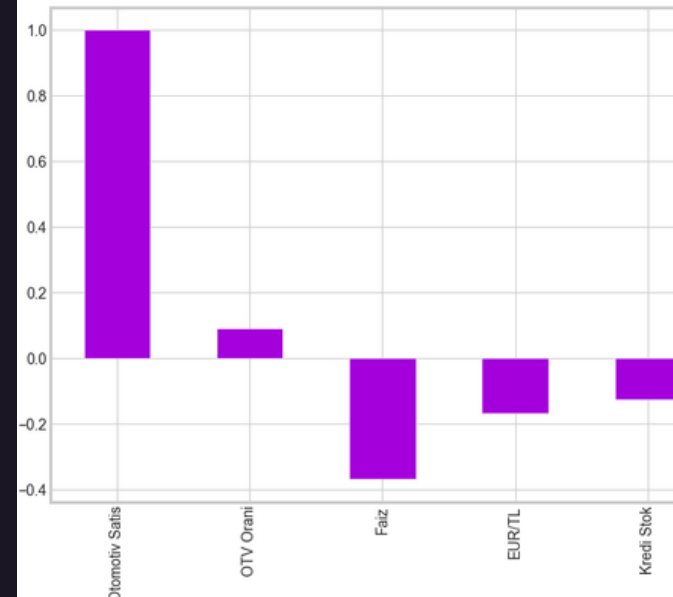


Fig 1. Verimizin Korelasyon Dağılımları



Fig 2. Verimizin Zamana Bağlı Grafikleri

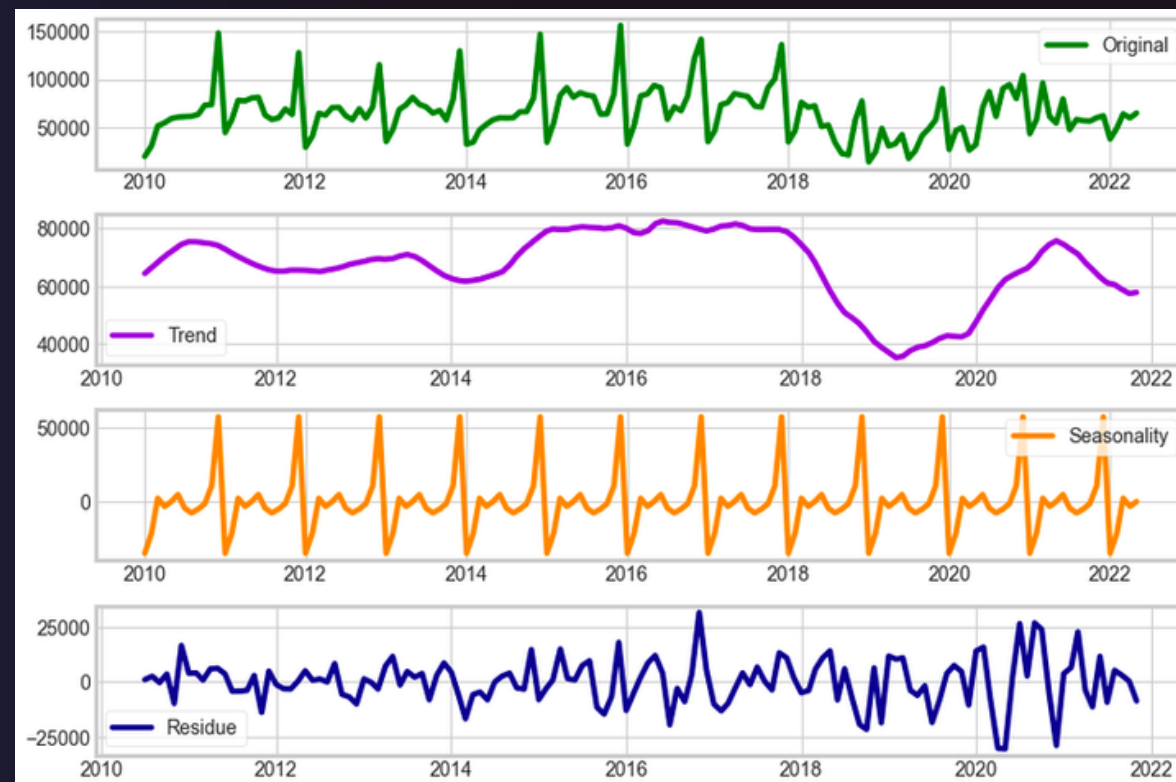


Fig 3. Sezonsal Grafikler

- Veri seti, 2010-2022 yılları arasında Türkiye'deki otomotiv satışlarına ait aylık verileri içermektedir.
- Veri seti, 5 farklı değişken içermektedir: "Otomotiv Satis" (hedef değişken), "OTV Oranı", "Faiz", "EUR/TL" ve "Kredi Stok".
- Veri keşfi aşamasında, veri setinin istatistiksel özellikleri (ortalama, standart sapma, vb.) incelenmiş ve değişkenler arasındaki ilişkiler görsel olarak gösterilmiştir.(Table 1)
- Veri keşfi sonuçlarına göre, "Otomotiv Satis" değişkeninin diğer değişkenlerle OTV verisi dışında negatif korelasyon gösterdiği tespit edilmiştir. (Fig. 1)
- Otomotiv satışlarındaki artış ve azalışlar, tarihlerine göre bir timeseries grafiği ile daha iyi anlaşılabilir. (Fig. 2)
- Verilere sezonsal değişkenleri eklemek için öncelikle hangi zaman aralıklarının sezonsal özellikler sahip olduğu analizi yapıldı. (Fig. 3)

VERİ KEŞFİ VE ÖN İŞLEME

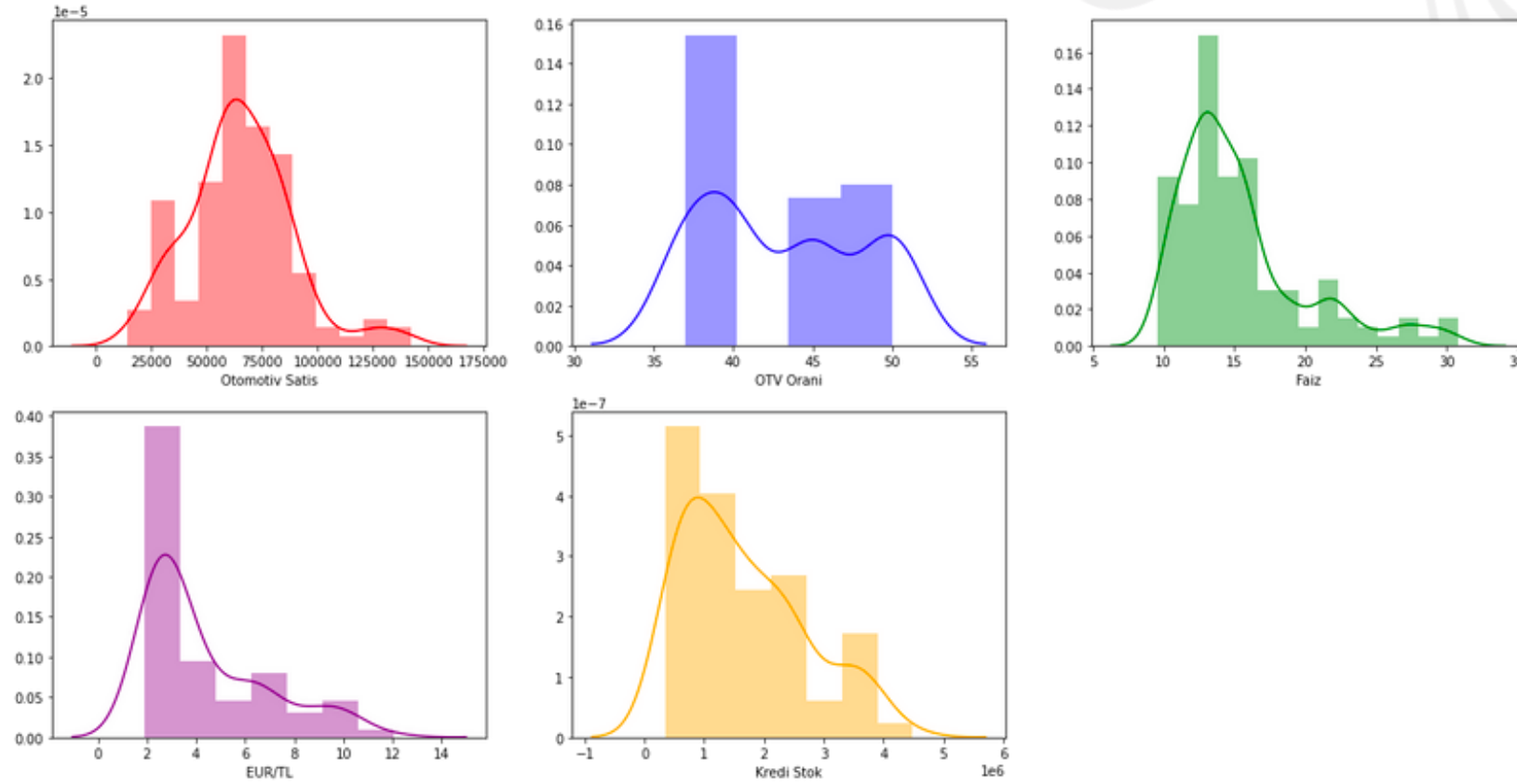
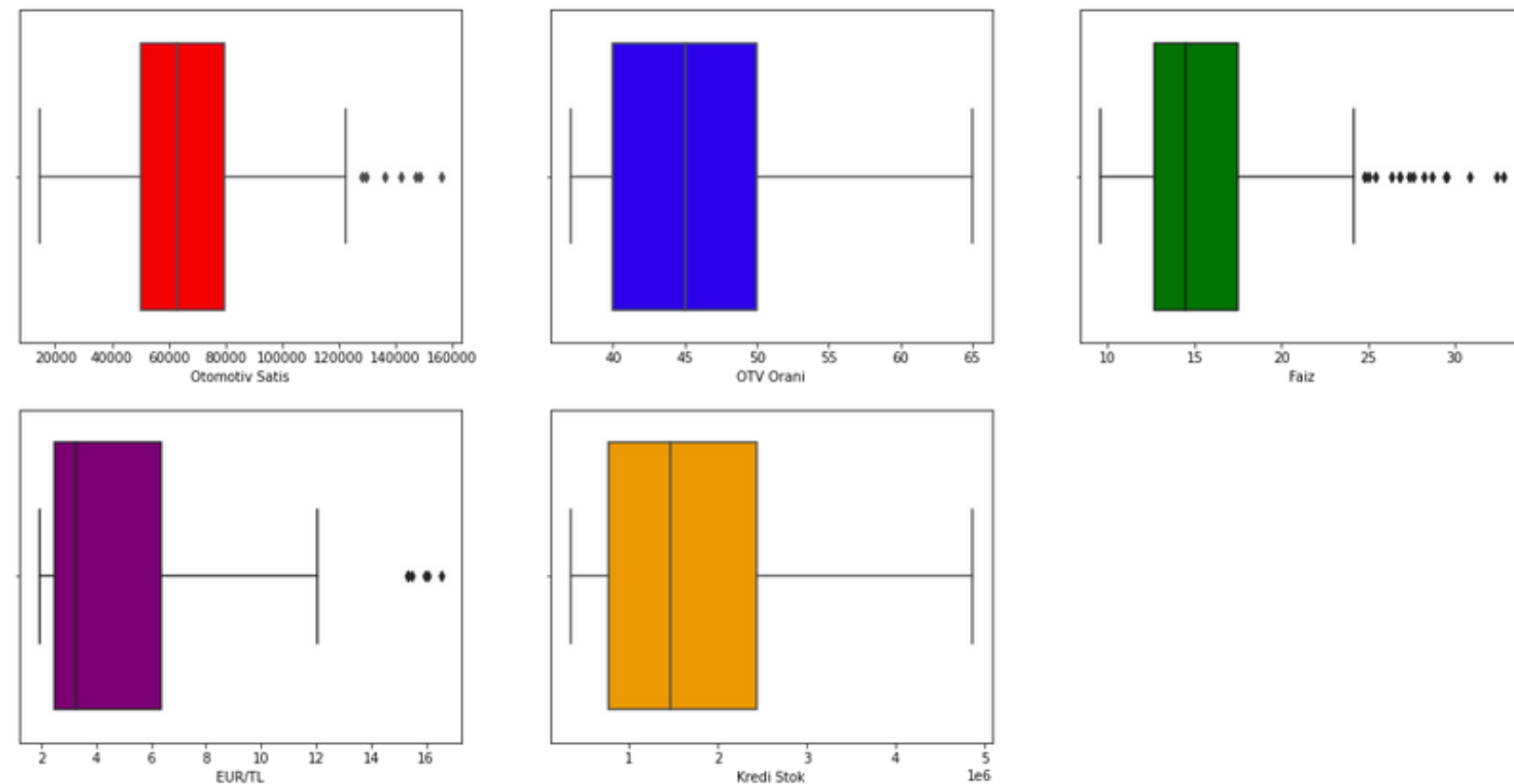
[Back to Agenda](#)

Fig 1. Dağılım Grafikleri



- Dağılım grafikleri, verilerin dağılımını görselleştirmek için kullanılan araçlardır. Verilerin türüne göre farklı dağılım grafikleri kullanılabilir.
- **Normal dağılım grafiği**, verilerin merkezi eğilimlerini ve dağılım genişliğini gösterir.(Fig. 1 Otomotiv Satis)
- Bu grafikler, verilerdeki merkezi eğilim ve dağılım genişliğini belirlemenmesinde yardımcı oldu.
- Sağa çarpık bir dağılım grafiği, verilerin sağ tarafta uzun kuyruklar oluşturduğu, sağ tarafta yoğunluğun daha az olduğu bir dağılım şeklidir. (Fig. 1 EUR/TL)
- p-value değeri düşük olan değerler "EUR/TL","Faiz" değişkenlerine Box-Cox dönüşümü uygulanıp normal dağılıma yaklaşması sağlandı.
- Box-Cox dönüşümü modelimizi negatif score olarak etkilediği için Box-Cox dönüşümü yapılmadan model oluşturulmuştur.

- **Aykırı değerler**, veri setinin genel analizini yanıltabilir ve modele zarar verebilir. Bu nedenle, aykırı değerlerinin silinmesi, daha doğru sonuçlar elde etmek için önemlidir.
- Aykırı Değerlerin tesbiti için Kutu Grafikleri kullanıldı.(Fig. 2)
- Aykırı değerlerin elimine edilmesi için kullanılan yöntemlerden
- Standart sapma yöntemi,IQR yöntemi,Z-score yöntemi
- Çalışmada ,"Otomativ Satis","Faiz","EUR/TL" değerleri için Z-score yöntemi kullanılmış, eşik değerimiz 3 olarak belirlenmiş ve bu değer dışında kalan değerler silinmiştir. (Fig. 2)
- Aykırı değerlerin silinmesi modelimizi negatif score olarak etkilediği için aykırı değerler çıkarılmadan model oluşturulmuştur.

DOĞRUSAL REGRESYON İÇİN VARSAYIMLARIN TEST EDİLMESİ

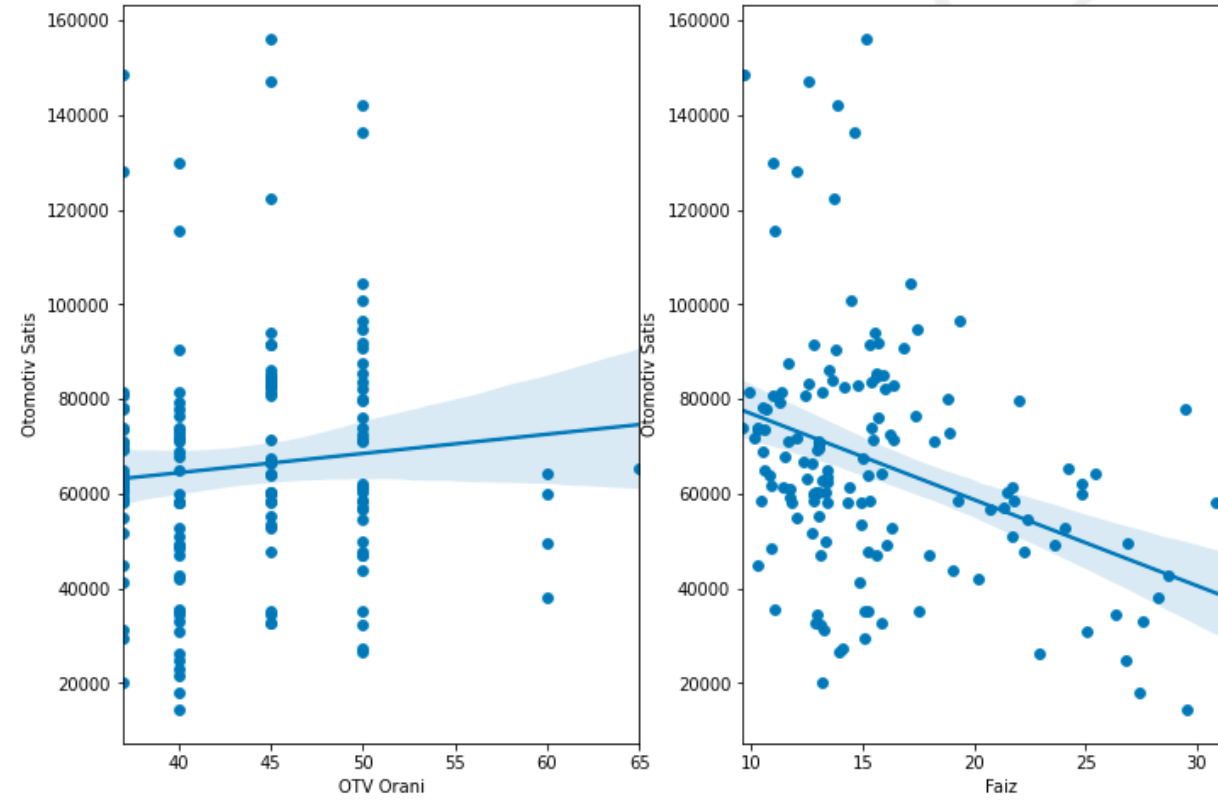


Fig 3. Doğrusallık Varsayımı

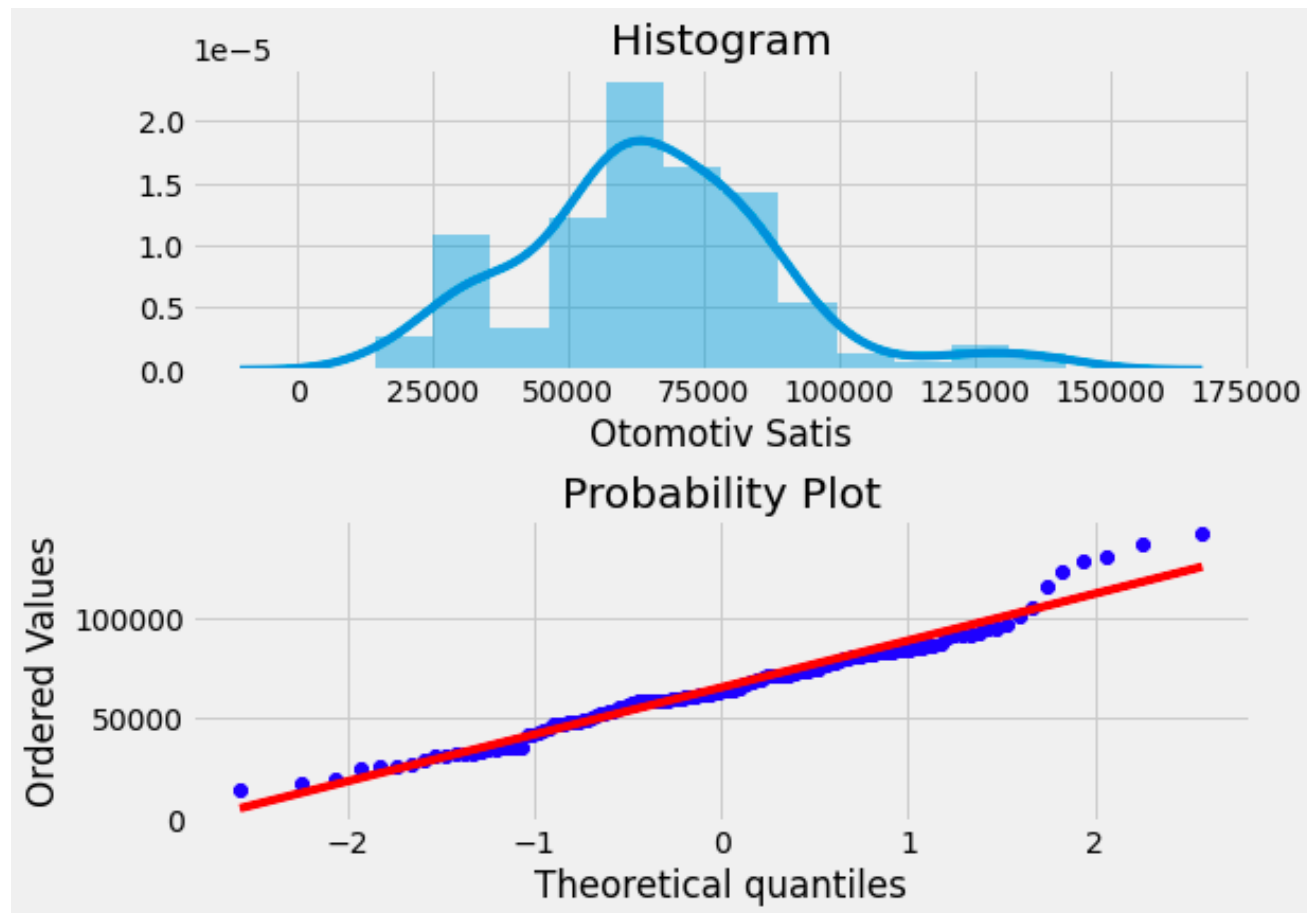


Fig 4. Normallik Kontrolü

- Doğrusal bir model kurduğumuz için, hedef "Otomotiv Satis" ile diğer özellikler arasındaki ilişkinin doğrusal olduğunu varsayıyoruz.
- **1.Doğrusallık Varsayımı**, bağımsız değişken ile bağımlı değişken arasındaki ilişkinin doğrusal olmasını gerektirir. Bu varsayımı bazı dağılım grafikleri ve regresyon çizgileri ile test edebiliriz (Fig. 3):
- "Otomotiv Satış",değişkenine bağlı iki farklı korelasyon vektörleri nedeniyle 'OTV Oranı' ve 'Faiz' özellikleri ile doğrusallık varsayımı yapıldı.
- **2.Normallik Kontrolü**, bağımlı değişken olan 'Otomotiv Satis 'ın normal dağılımlı olmasını gerektirir.
- Uyum iyiliği testi, örneğin Kolmogorov-Smirnov testi, bağımlı değişkendeki normalliği kontrol edebilir.
- Hedef değişkenimiz olan 'Otomotiv Satis'in nasıl davrandığını göstermek için iki grafiği de görüntüledik (Fig. 4):
- Hedef değişkenimiz olan 'Otomotiv Satis' normal dağılıma sahiptir
- Uyum testimiz,doğrusal,uyumlu bir davranış sergilemiştir.

MODEL OLUŞTURMA

Mevsimsel Değişkenler Dahil Edilmeden Yapılan Model

	Actual value	Predicted value
Date		
2021-02-01	58504.0	68774.450564
2019-02-01	24875.0	42184.474305
2014-12-01	146989.0	73785.920703
2016-09-01	67593.0	71159.462082
2010-08-01	61764.0	66637.816550
2018-09-01	23028.0	26989.231761
2021-09-01	57141.0	63692.886734
2017-12-01	136240.0	75989.569828
2019-11-01	58176.0	68427.557375

Table 1. Gerçek ve tahmin verileri Tablosu



Fig 1. Model Grafiğimiz

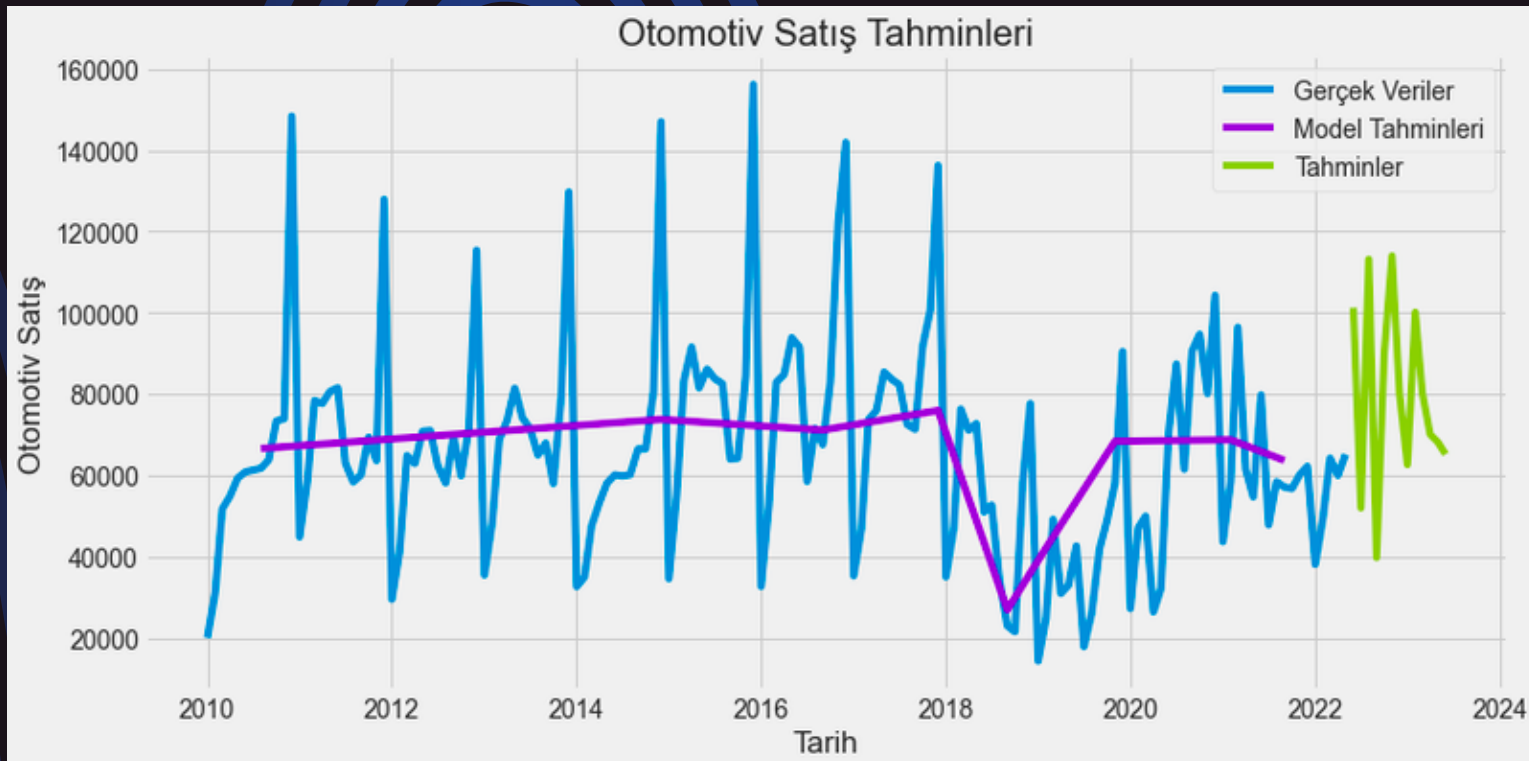


Fig 2. Time Series Forecast Model

MSE: 1066010217.2154276
RMSE: 32649.811901685247
R2 Score: 0.3632194068204203
determination of prediction Score: 0.3632194068204203

[Back to Agenda](#)

Table 3 Model Performansı

- Model eğitimi için veri seti, %94 eğitim ve %6 test verisi olarak ayrıldı.
- Modelin performansı, test verisi üzerinde RMSE (Karekök Ortalama Kare Hata) ve R-kare ölçütleri kullanılarak değerlendirildi.(Table 3)
- Model, Haziran 2022 - Haziran 2023 dönemi için otomotiv satış adetlerinin tahmininde kullanılmak üzere eğitildi.(Table 1)
- Tahmin işlemi yapmak için önceden belirlenmesi istediğimiz tarihler için veriye 2022 Haz ve 2023 Haz tarihleri ve 4 bağımsız değişken eklendi.(Table 2)
- Bağımsız değişkenler random olarak eklenirken hepsinin istatistiksel dağılımları özel olarak dikkate alınarak verimiz üretildi.
- Daha sonra ölçeklendirme işlemleri (standart scaler) sağlanarak verimiz tahmin edildi.(Fig. 2)
- Ridge regression ve feature selection ile model geliştirilmeye çalışıldı ancak score'lar düştüğü için modele eklenmediler.
- Elde edilen sonuçlar, modelin performansı ile birlikte sunuldu.(Fig. 1)



```
1 # Verilerin istatistiksel davranışlarını kullanarak rastgele değerler oluşturuyoruz
2 otv_values = np.random.randint(low=37, high=65, size=len(forecast_df))
3 kredi_values = np.random.normal(loc=1787554.29, scale=1176365.97, size=len(forecast_df))
4 faiz_values = np.random.normal(loc=16.03, scale=5.32, size=len(forecast_df))
5 eur_values = np.random.normal(loc=4.81, scale=3.32, size=len(forecast_df))
6
```

Table 2. Hedef Değişkeni Tahmin Etmek için Üretilen Bağımsız Random Değişkenler

MODEL OLUŞTURMA

MSE: 129105609.91228892
RMSE: 11362.46495758244
R2 Score: 0.8342903545377529
determination of prediction Score: 0.8342903545377529

[Back to Agenda](#)

Mevsimsel Değişikler ile Üretilen Model-2

	Actual value	Predicted value
Date		
2021-02-01	58504.0	44362.978884
2019-02-01	24875.0	23356.453710
2014-12-01	146989.0	129406.336438
2016-09-01	67593.0	72414.336936
2010-08-01	61764.0	60018.215106

Table 1. Gerçek ve tahmin verileri Tablosu

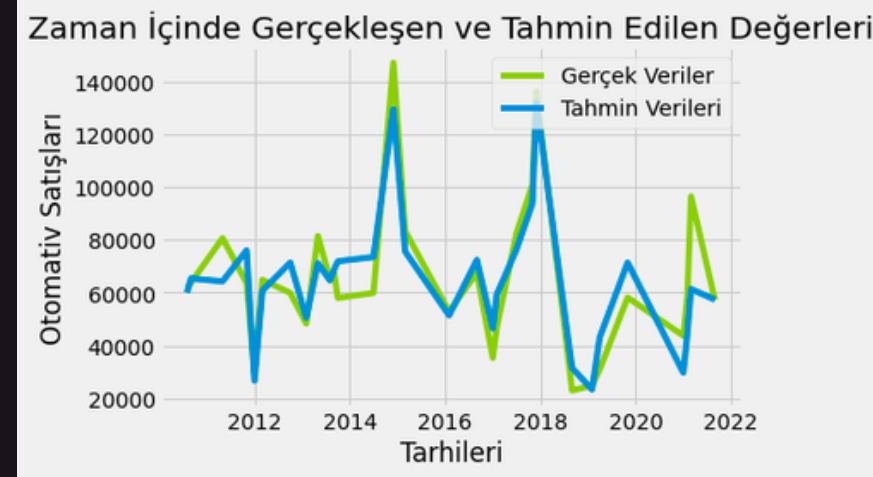


Fig 1. Model Grafiğimiz

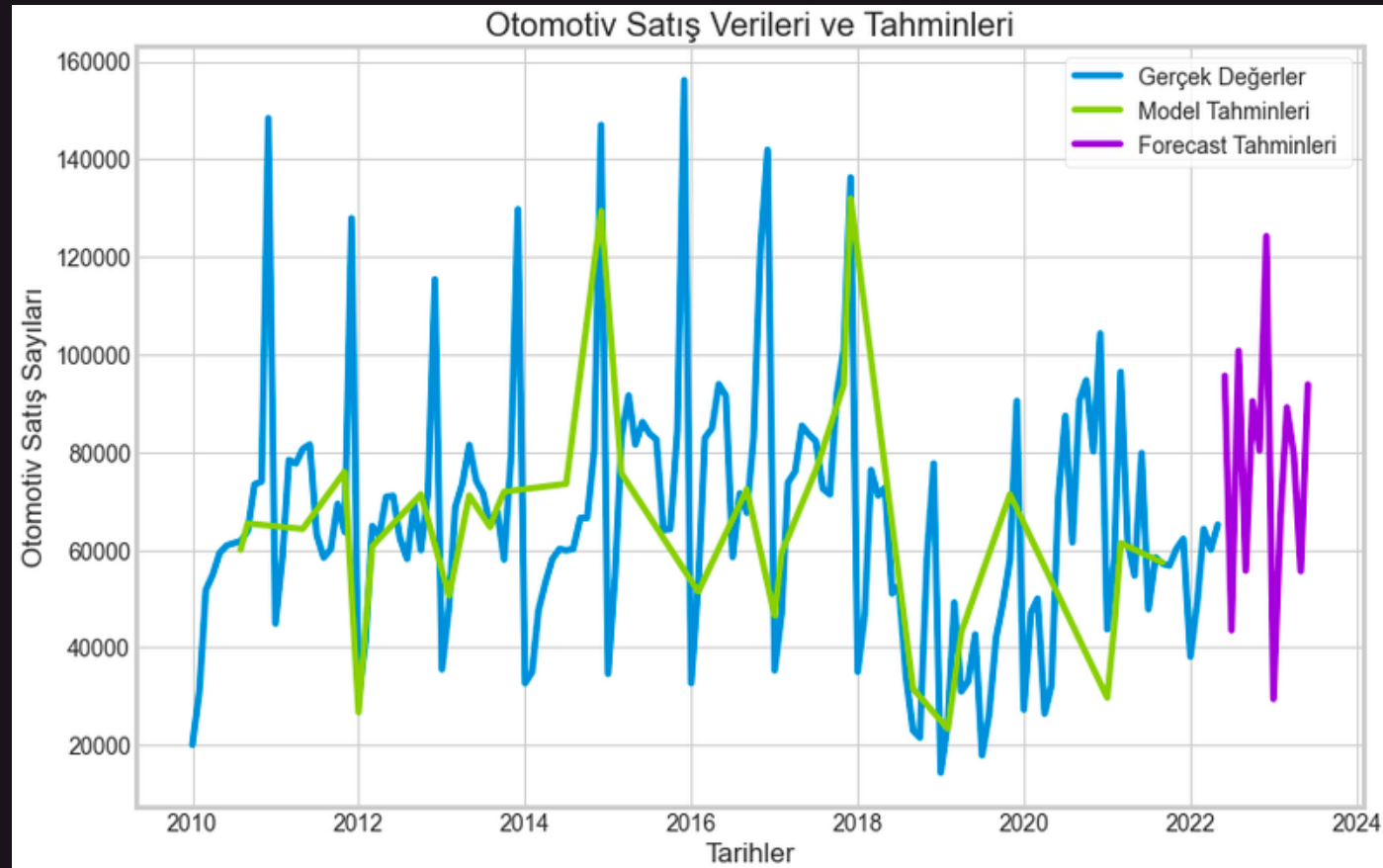


Fig 2. Timeseries Forecast Grafiği

Table 3 Model Performansı

- Çoklu doğrusal regresyon modeli, mevsimsel değişkenler eklenerek geliştirildi.
- Otomotiv satışlarının genellikle yaz aylarında arttığı biliniyor. Bu nedenle, Haziran, Temmuz ve Ağustos aylarını bir sezon olarak düşünebilir. (Table 2)
- Yaz ayları haziran-ağustos dönemi olarak belirlendi.
- Bu karar aşamasından sonra verilerin aylara bölünmesi ve dummy değişkenler oluşturulması sağlandı.
- Dummy aşamasından sonra haziran ve ağustos dönemleri için yeni bir yaz değişkenini tanımlayan bir değer atandı.
- Model eğitimi için veri seti, %80 eğitim ve %20 test verisi olarak ayrıldı. (Table 1)
- Bir önceki model de olduğu gibi bağımsız değişkenler üretildi ve ölçeklendirildi
- Elde edilen sonuçlar, modelin performansı ile birlikte sunuldu. (Fig. 1)
- Modelin performans score'u, 0.36'dan 0.83'e artması sağlandı. (Fig. 2)

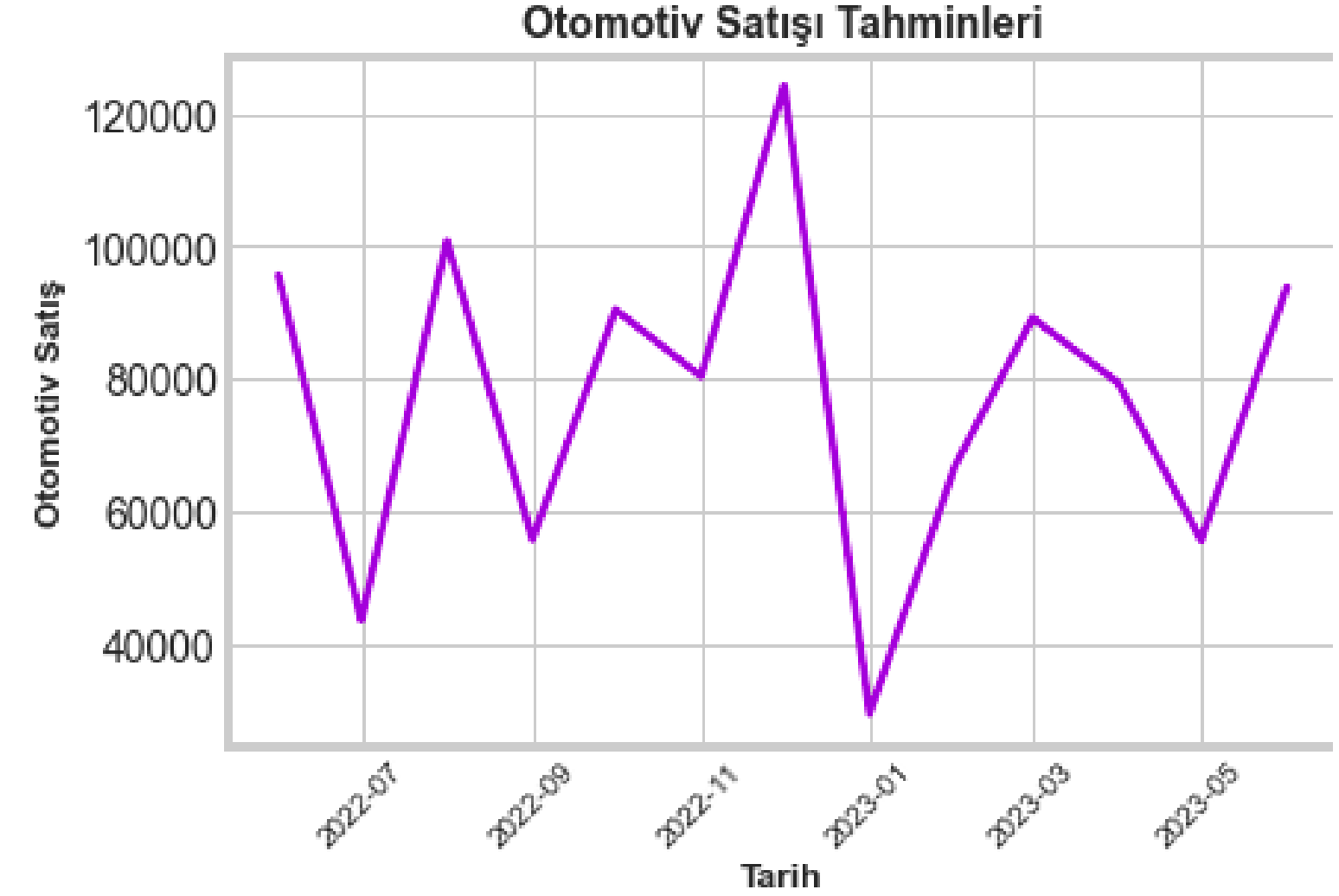
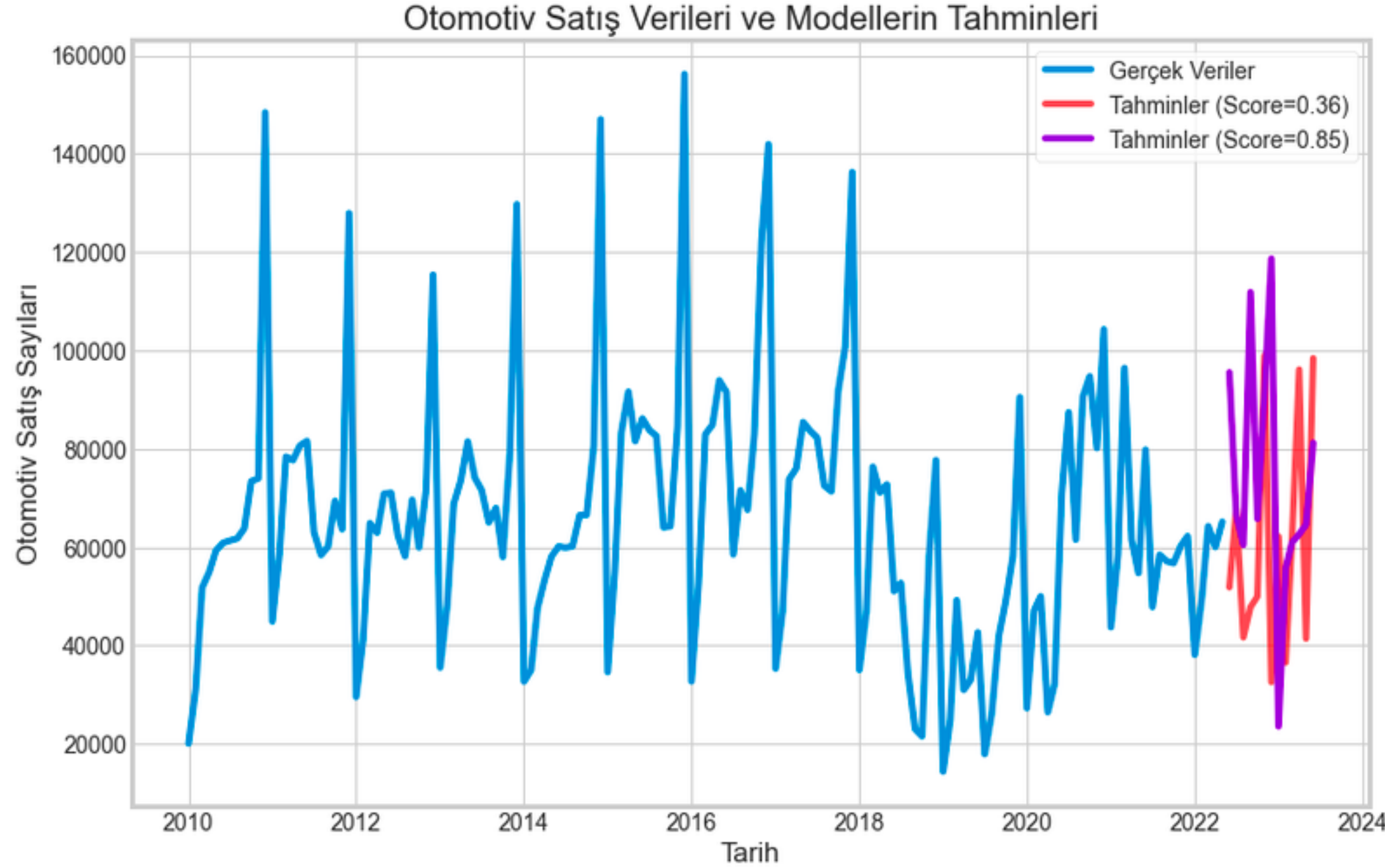


```
1 df['Season_JJA'] = ((df['Month'] == 6) | (df['Month'] == 7) | (df['Month'] == 8)).astype(int)
2 df.columns
3
✓ 0.0s

Index(['Otomotiv Satis', 'OTV Orani', 'Faiz', 'EUR/TL', 'Kredi Stok', 'Month',
      'Month_2', 'Month_3', 'Month_4', 'Month_5', 'Month_6', 'Month_7',
      'Month_8', 'Month_9', 'Month_10', 'Month_11', 'Month_12', 'Season_JJA'],
      dtype=object)
```

Table 2. Hedef Değişkeni Tahmin Etmek için Eklenen Sezonsal Değişkenimiz ve verimizin aylarının artırılmış sütunları

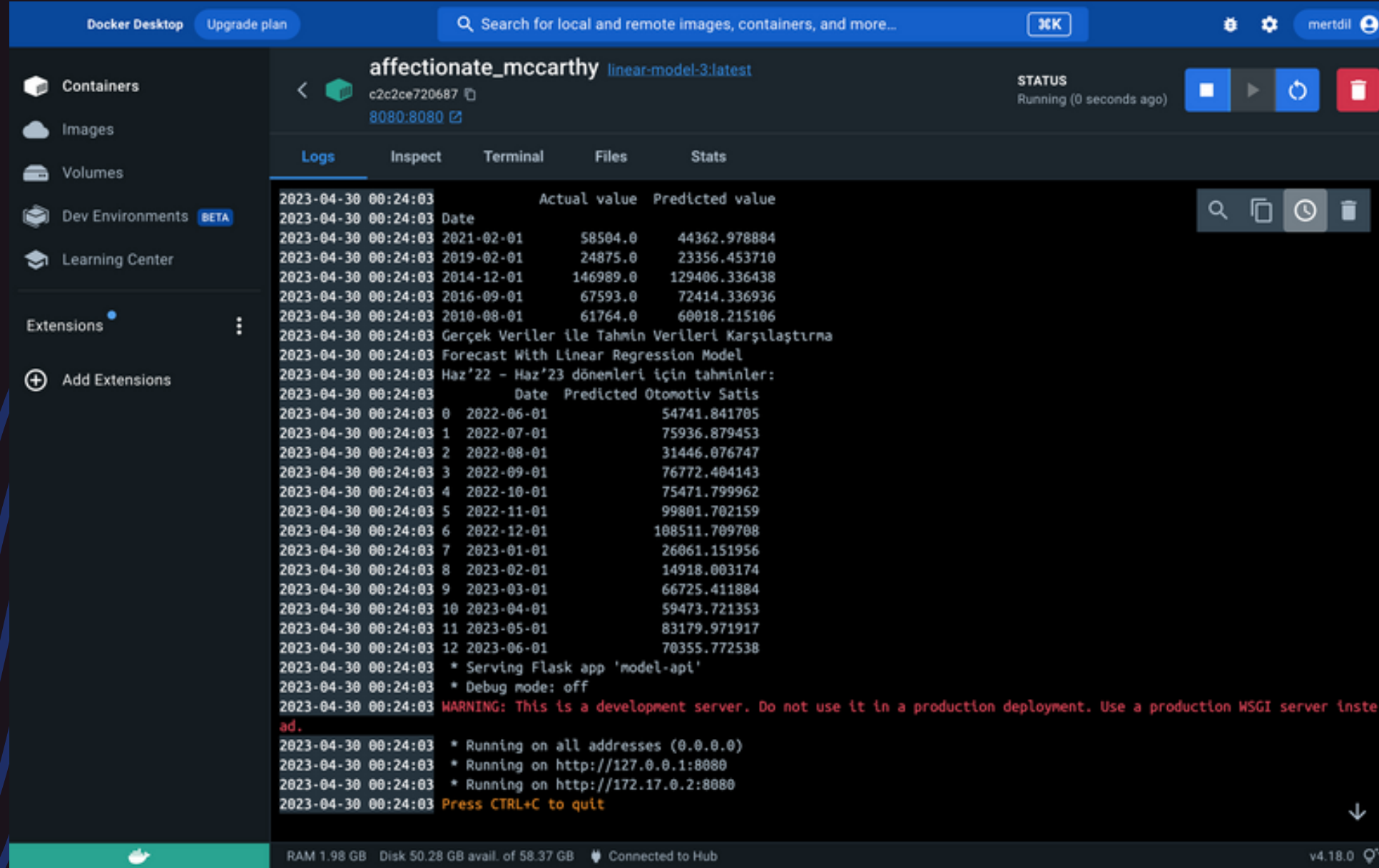
MODEL TAHMINLERİ VE KARŞILAŞTIRMALARI



- Mevsimsel değişikliklerin modellere eklenmesi, özellikle zaman serisi tahminimizde modelin öğrenmesi artırmıştır

- Bu yöntem, Haziran 2022- Haziran 2023 dönemleri arasında Otomotiv Satış değerlerini tahmin etmede önemli bir rol oynamıştır

MODELİN SERVİS HALİNE GETİRİLMESİ (FLASK-DOCKER)

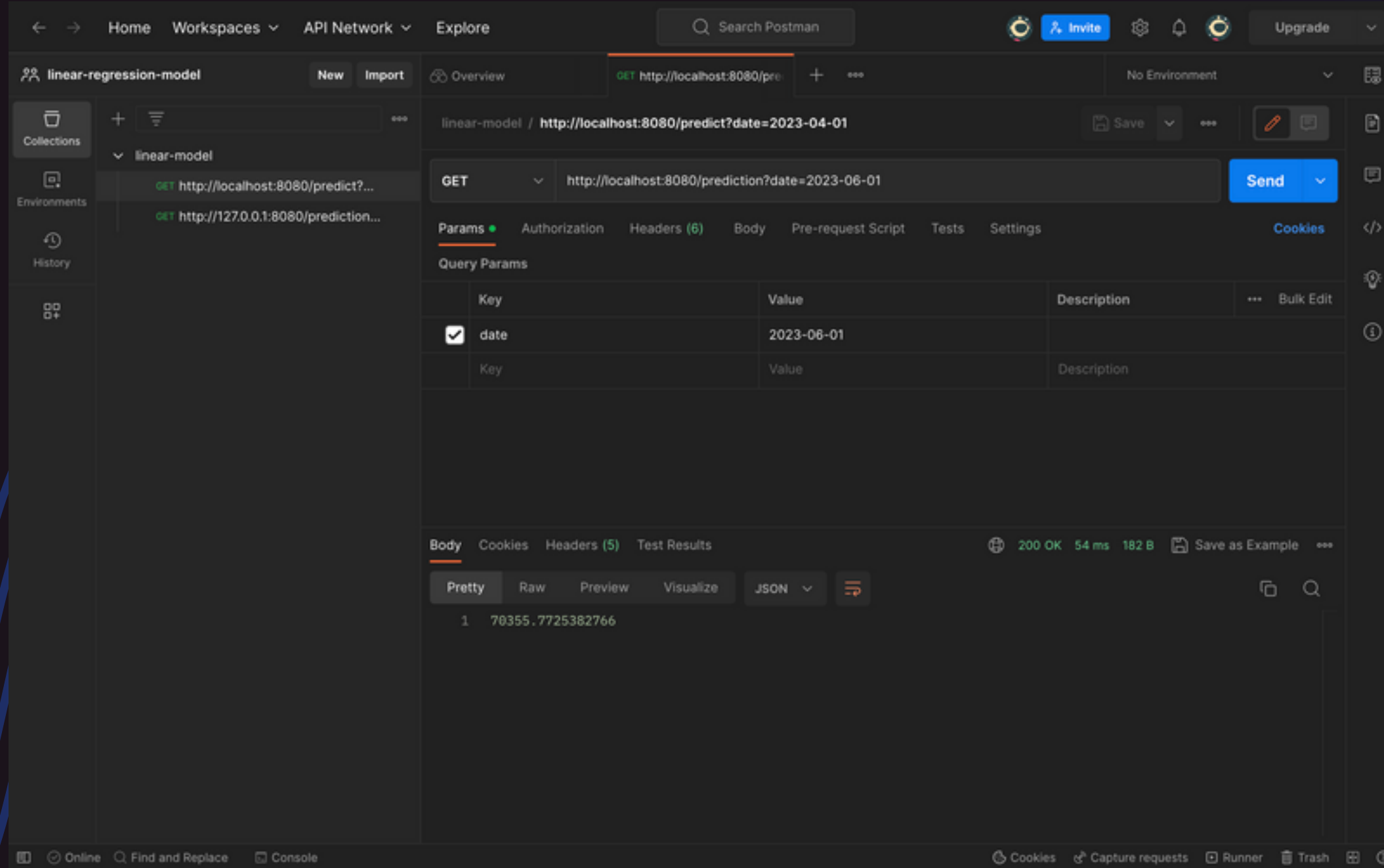


```
2023-04-30 00:24:03 Actual value Predicted value
2023-04-30 00:24:03 Date
2023-04-30 00:24:03 2021-02-01 58504.0 44362.978884
2023-04-30 00:24:03 2019-02-01 24875.0 23356.453710
2023-04-30 00:24:03 2014-12-01 146989.0 129406.336438
2023-04-30 00:24:03 2016-09-01 67593.0 72414.336936
2023-04-30 00:24:03 2010-08-01 61764.0 60018.215106
2023-04-30 00:24:03 Gerçek Veriler ile Tahmin Verileri Karşılaştırma
2023-04-30 00:24:03 Forecast With Linear Regression Model
2023-04-30 00:24:03 Haz'22 - Haz'23 dönemleri için tahminler:
2023-04-30 00:24:03 Date Predicted Otomotiv Satış
2023-04-30 00:24:03 0 2022-06-01 54741.841705
2023-04-30 00:24:03 1 2022-07-01 75936.879453
2023-04-30 00:24:03 2 2022-08-01 31446.076747
2023-04-30 00:24:03 3 2022-09-01 76772.404143
2023-04-30 00:24:03 4 2022-10-01 75471.799962
2023-04-30 00:24:03 5 2022-11-01 99801.702159
2023-04-30 00:24:03 6 2022-12-01 108511.709708
2023-04-30 00:24:03 7 2023-01-01 26061.151956
2023-04-30 00:24:03 8 2023-02-01 14918.003174
2023-04-30 00:24:03 9 2023-03-01 66725.411884
2023-04-30 00:24:03 10 2023-04-01 59473.721353
2023-04-30 00:24:03 11 2023-05-01 83179.971917
2023-04-30 00:24:03 12 2023-06-01 70355.772538
2023-04-30 00:24:03 * Serving Flask app 'model-api'
2023-04-30 00:24:03 * Debug mode: off
2023-04-30 00:24:03 WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
2023-04-30 00:24:03 * Running on all addresses (0.0.0.0)
2023-04-30 00:24:03 * Running on http://127.0.0.1:8080
2023-04-30 00:24:03 * Running on http://172.17.0.2:8080
2023-04-30 00:24:03 Press CTRL+C to quit
```

- **Rest API**, modelin kullanıcılar tarafından kolayca erişilebilir olmasını ve uygulama tarafından modelin sonuçlarına erişebilmesini sağlar.
- **Docker**, modelin izole bir ortamda çalıştırılmasını ve taşınmasını kolaylaştırır. Bu sayede, modelin bağımsız bir servis haline getirilmesi mümkün olur.
- Modelin Dockerize edilmesi, herhangi bir platformda ve herhangi bir cihazda çalışabilmesini sağlar.
- Flask yardımı ile modelin Get metodu ile sorgu yapılması sağlanır.
- dockerfile oluşturularak modelin dockerize edilmesi sağlanmıştır. 8080 portu üzerinde işlemler sürdürülür.
- dockerfile'de bulunan cmd komutu ile .py olarak bulunan model-2 çalıştırılıp image üzerinden kaydedilmiştir.

```
###Postman sorgumuz date=istediğimiz ayın birinci gününü söylediğimizde tahminimizi göstericek şekilde tasarlanmıştır.
"http://localhost:8080/prediction?date=2023-06-01"
"http://0.0.0.0:8080/prediction?date=2023-06-01"
```

MODELİN TEST EDİLMESİ (POSTMAN)



- **Postman** gibi araçlar ise, modelin servis haline getirildikten sonra test edilmesini sağlar ve hataların tespit edilmesine yardımcı olur.
- Bu adım, modelin gerçek hayatta kullanımını mümkün kılar ve uygulama geliştiricilerinin, modeli kolayca entegre etmelerine olanak tanır.
- Docker ile servise alındı ve GET metodu, HTTP isteği ile datamızdan tahmin edilen aylar arasından istekte bulunuyoruz.
- Bu isteğimiz Haziran 2022-Haziran 2023 arasındaki ayları kapsayacak şekilde tahmin edilmiş verimizi kapsar.

Zaman
ayırdığınız
için teşekkürler.

EMAIL

mertdill@outlook.com

TELEFON

05419537658

SOSYAL MEDYA

<https://www.linkedin.com/in/mertdil>

PORTFOLIO

<https://github.com/Mertdil>

Thank
you!