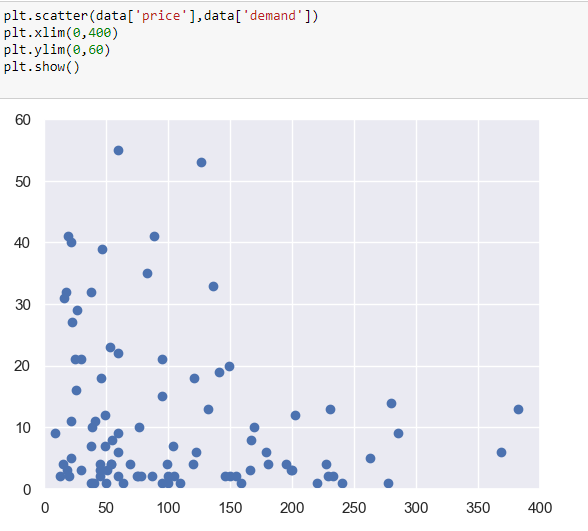
Assignment: Demand Forecasting

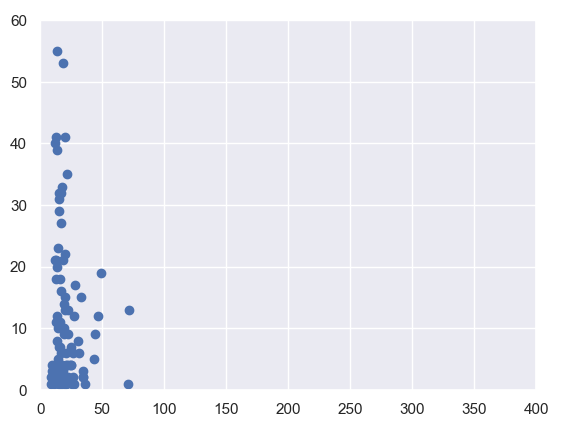
Approach 1: Linear Regression

To begin with, I decided to find the demand (number of purchased) for each product separately.

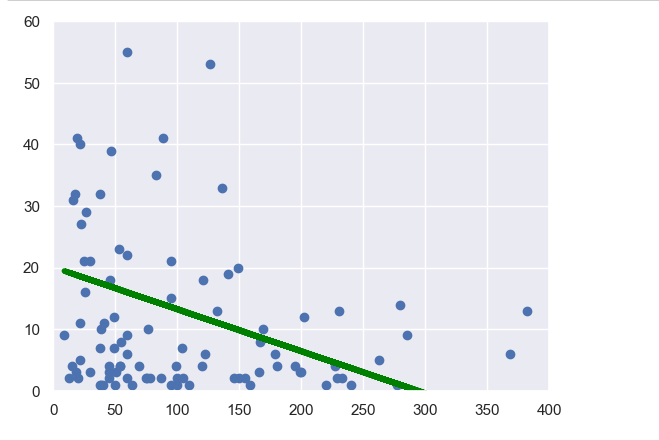
Then counting checkmate. waiting for the price and freight value columns for these goods.

Then, having plotted a graph of demand versus price, I noticed that these data to some extent satisfy the property of linearity



When plotting the dependence of demand on freight\_value, it became clear that this variable is insignificant for forecasting demand

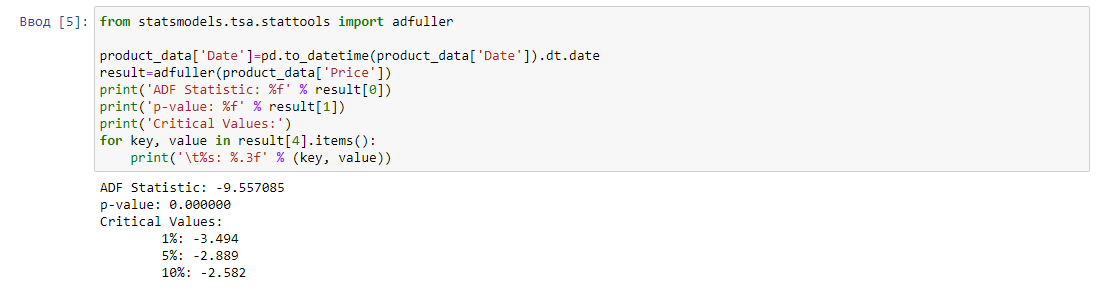
I then built a linear regression model based on a sample of 50 random products,

Having received the coefficients, I built a regression line.

You can clearly see that this model does not capture all the information. The model can be used to forecast demand, however, I believe that this is far from the best way.

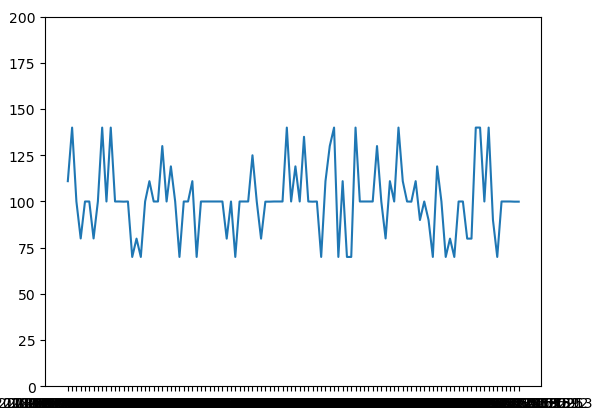
Approach 2: Time Series

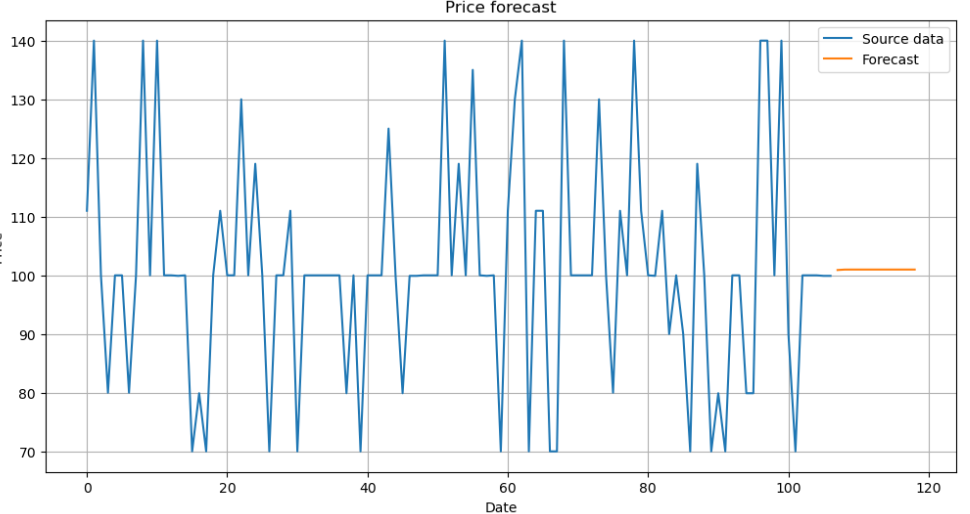
To predict the price, I chose a random product, (product\_id=’ 4fe644d766c7566dbc46fb851363cb3b’)

When predicting any value using a time series, it is necessary to determine whether the original series is stationary. There are several ways to do this. I used the Dickey-Fuller Test and received information that the series is stationary. (p<=0.05)

The stationarity of the series indicates that the series does not have a specific trend.

Then, to predict the price based on previous data, I chose the ARIMA model.

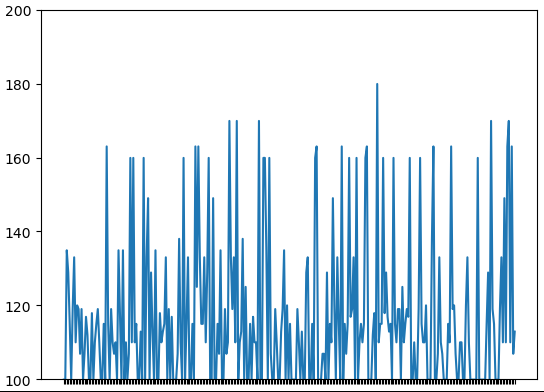
Original row:

Price forecast for 2020:

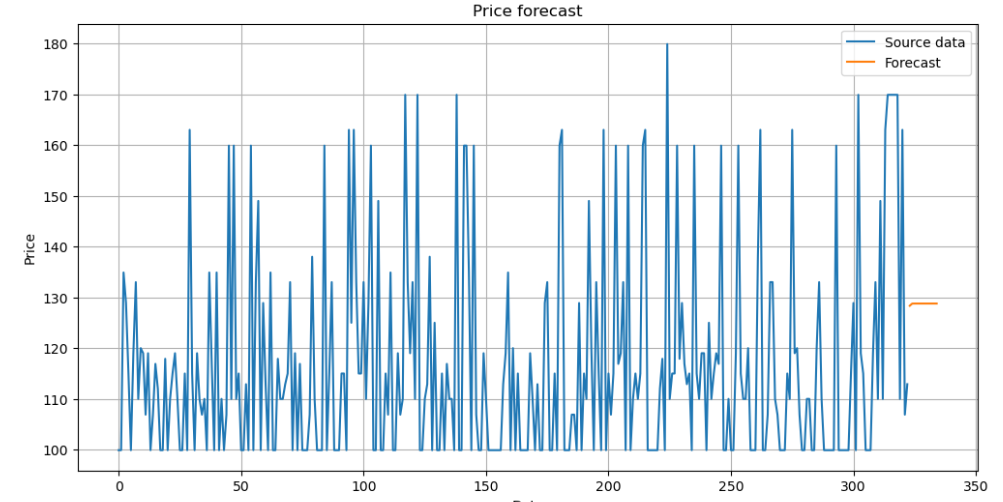
This forecast is quite plausible, since we see that the price of this product fluctuates around $100.

Using another product as an example:

Original row:



Forecast



In this case, the time series of the price of the product has a positive trend, as a result of which the ARIMA model assumes an increase in the price of the product.

The Time Series.ipynb File contains all the calculations I made