Гузенко А.М. Группа 7.2. Вариант 4

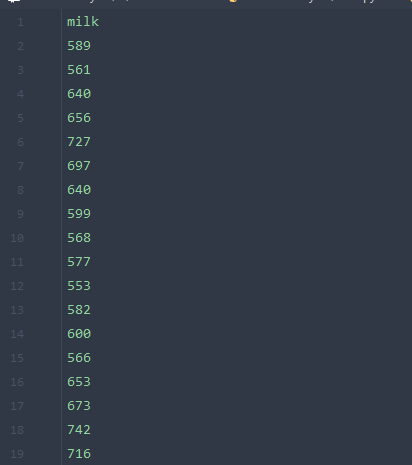
Лабораторная работа № 2

Линейная регрессия

“Количество молока в месяц”

**Цель**

Провести линейный регрессионный анализ, спрогнозировать короткий временной промежуток.

**Описание данных**

Количество произведенного молока в месяц. У ряда есть тренд – с течением времени количество продукции увеличивается. Есть сезонность – чем ближе зима, тем продукции меньше, чем ближе лето, тем продукции становится больше, в мае количество продукции обычно максимальное. Данные не меняют свой характер. Выбросов в данных нет.

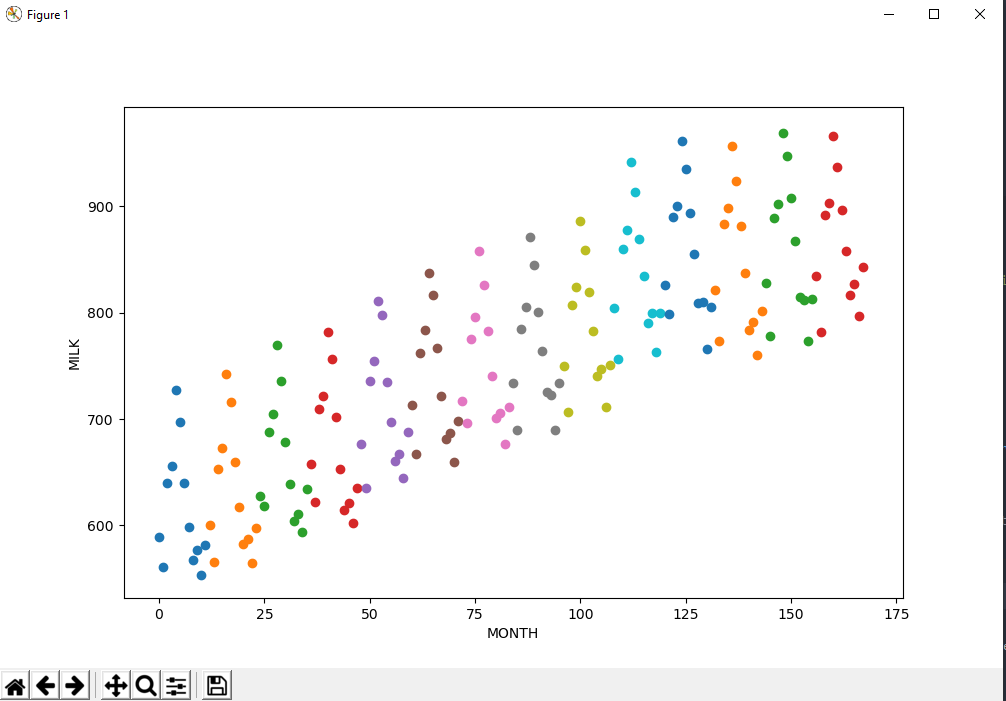
**Выполнение работы**

1. Импортируем нужные нам библиотеки, функции.

import pandas as pd  
import os  
import matplotlib.pyplot as plt  
import numpy as np  
from sklearn.linear\_model import LinearRegression  
from sklearn.preprocessing import normalize

1. Считаем данные и выведем их на графике.

data = pd.read\_csv(PATH + 'data\\milk.csv')  
for i in range(0, 15):  
 plt.scatter([j+(12\*i) for j in range(0, len(data['milk'][i\*12:12+12\*i]))], data['milk'][i\*12:12+12\*i])  
plt.xlabel('MONTH')  
plt.ylabel('MILK')  
plt.show()

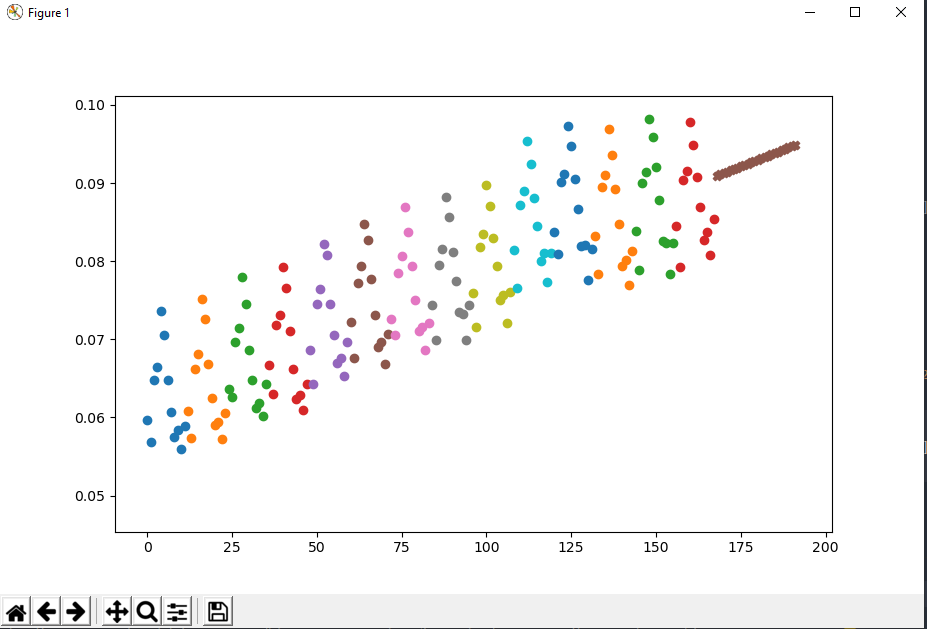


1. Подготовим данные, создадим модель на этих данных и сделаем прогноз.

x = np.array([i for i in range(0, len(data['milk']))]).reshape((-1, 1))  
norm\_data = normalize(np.array(data['milk']).reshape((1, -1)))  
lr = LinearRegression().fit(x, norm\_data[0])  
x\_predict = np.array([i for i in range(len(data['milk']), len(data['milk'])+24)]).reshape((-1, 1))  
y\_predict = lr.predict(x\_predict)

1. Выведем подготовленные данные вместе с прогнозом, на следующие два года, на графике.

for i in range(0, 15):  
 plt.scatter([j+(12\*i) for j in range(0, len(norm\_data[0][i\*12:12+12\*i]))], norm\_data[0][i\*12:12+12\*i])  
plt.scatter(x\_predict, y\_predict, marker='X')  
plt.show()



**Вывод**

При помощи линейной регрессии мы смогли спрогнозировать рост, тренд, молочной продукции на следующие два года, но мы не смогли спрогнозировать сезонность, хотя это можно сделать установив зависимость между месяцами года и наложить ее на спрогнозированные данные.