|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №3**

**«Анализ и прогнозирование временных рядов. Метод ARIMA»**

**ДИСЦИПЛИНА: «Технологии анализа данных»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-82Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Карельский М.К. )  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Ерохин И.И. )  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

Калуга, 2024

**Цель:** формирование практических навыков анализа и прогнозирования временных рядов, а также применения метода ARIMA.

**Задачи:**

1. Ознакомиться с функциональными возможностями Python для анализа временных рядов.
2. Изучить метод ARIMA.

**Вариант 5**

Считать данные в структуру Dataframe. Выбрать из набора данные, относящиеся ко всем подкатегориям. Определить какая из подкатегорий будет лучше всего продаваться через 5 лет на основе графических данных модели прогноза по методу ARIMA. Оценить качество модели (дисперсия и СКО от реальных данных за какой-либо период) на примере любых двух подкатегорий.

**Листинг:**

import pandas as pd

from statsmodels.tsa.arima.model import ARIMA

import matplotlib.pyplot as plt

data = pd.read\_excel("Superstore.xls")

sub\_categories\_data = data.groupby('Sub-Category')

for sub\_category, sub\_category\_df in sub\_categories\_data:

    yearly\_sales = sub\_category\_df.groupby(pd.Grouper(key='Order Date', freq='Y'))['Sales'].sum()

    model = ARIMA(yearly\_sales, order=(5,1,0))

    model\_fit = model.fit()

    forecast = model\_fit.forecast(steps=5)

    print("Sub-Category:", sub\_category)

    print("Forecast for next 5 years:")

    print(forecast)

    plt.figure(figsize=(10,6))

    plt.plot(yearly\_sales.index, yearly\_sales, label='Actual Sales')

    plt.plot(pd.date\_range(start=yearly\_sales.index[-1], periods=6, freq='Y')[1:], forecast, label='Forecast')

    plt.title("ARIMA Forecast for {}".format(sub\_category))

    plt.xlabel("Year")

    plt.ylabel("Sales")

    plt.legend()

    plt.show()

    residuals = model\_fit.resid

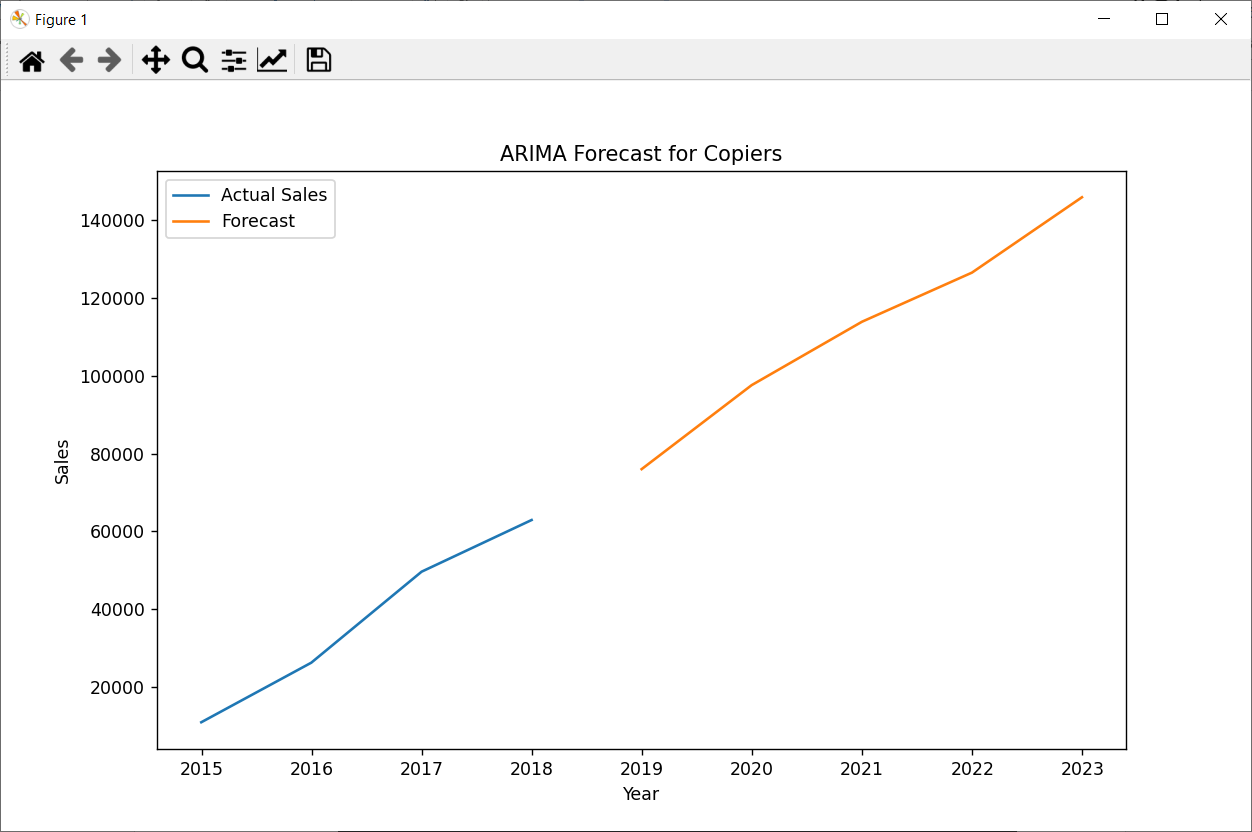
    variance = residuals.var()

    std\_deviation = residuals.std()

    print("Variance:", variance)

    print("Standard Deviation:", std\_deviation)

**Результат:**



**Рис. 1.** Подкатегория, которая будет лучше всего продаваться через 5 лет



**Рис. 2.** Дисперсия и СКО подкатегории Tables



**Рис. 3.** Дисперсия и СКО подкатегории Accessories

**Вывод:** в ходе выполнения лабораторной работы были получены практические навыки анализа и прогнозирования временных рядов, а также применения метода ARIMA.