МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

“ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ”

ФГБОУ ВО «ВГУ»

Отчет по теме

Прогнозирование коротких временных рядов

Задание №15

09.03.02 Информационные системы и технологии

Программная инженерия в информационных системах

Зав. кафедрой\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ С. Д. Махортов, д.ф-м.н. \_\_.\_\_.20\_\_

Обучающийся\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М. Д. Шеменев, 4 курс, 5.1 группа

Руководитель\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ И.В. Илларионов, доцент

Воронеж 2023**Содержание**

[Введение 3](#_Toc147095944)

[1 Анализ временного ряда 4](#_Toc147095945)

[1.1 Отбор необходимых данных 4](#_Toc147095946)

[1.2 Проведение анализа временного ряда 4](#_Toc147095947)

[1.2.1 Вычисление тренда, сезонности и выбросов временного ряда 4](#_Toc147095948)

[1.2.2 Анализ графиков временного ряда 5](#_Toc147095949)

[2 Построение прогностической модели 7](#_Toc147095950)

[Заключение 9](#_Toc147095951)

Введение

Целью данной работы является анализ временного ряда на такие компоненты как тренд, сезонность, характер и выбросы. На основании входных данных построить прогностическую модель и сделать предсказание на 8 месяцев. По полученным результатам сделать выводы, отобразив ключевые моменты в отчете.

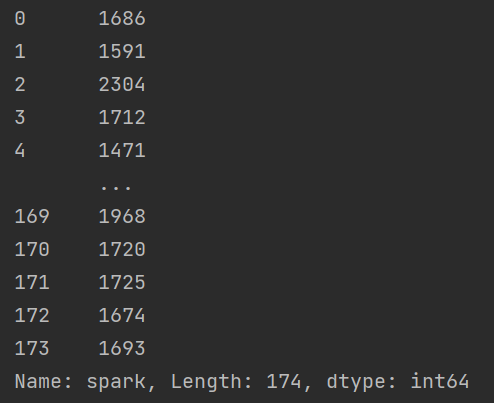
1. Анализ временного ряда
   1. Отбор необходимых данных

Для проведения анализа временного ряда были выбраны данные продажи игристых вин в Австралии в период с января 1980 по июнь 1994.

За выборку данных из файла отвечает строка кода:

spark\_data = pd.read\_csv(file\_name, sep="\t")[key]

После выполнения данного кода, мы получим данные в виде, представленном на Рисунке 1.



1. Входные данные
   1. Проведение анализа временного ряда
      1. Вычисление тренда, сезонности и выбросов временного ряда

Чтобы рассчитать все необходимые параметры временного ряда и затем отобразить полученные результаты на графике для удобного анализа, нам необходимо воспользоваться библиотекой statsmodels, а именно методом seasonal\_decompose.

def show\_time\_series(data):  
 result = seasonal\_decompose(data, model='additive', period=12)  
 result.plot()  
 plt.show()

В данном коде параметр «data» представляет собой входные данные, которые мы получили из файла.

В качестве метода исследования была выбрана аддитивная модель, которая представляет собой математический метод, используемый для анализа временных рядов. Для анализа временных рядов с помощью аддитивной модели производится декомпозиция ряда и оценивается каждый компонент отдельно.

В качестве параметра периода, отвечающего за интервал серии, было выбрано количество месяцев в году.

Результат работы метода представлен на Рисунке 2.

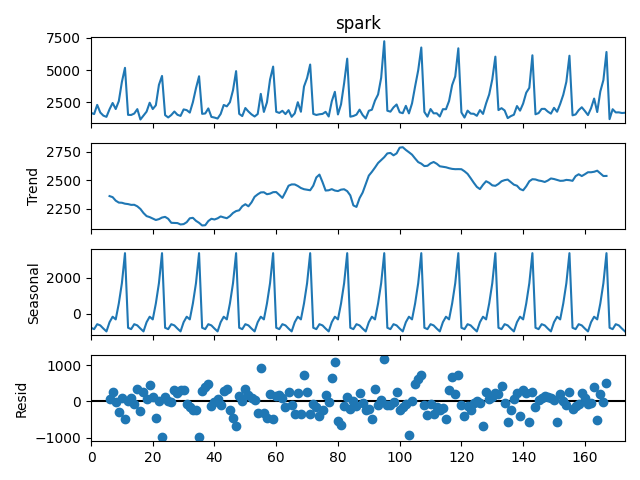


Рисунок 2 - График основных компонентов временного ряда

* + 1. Анализ графиков временного ряда

Первый график представляет из себя исходные данные временного ряда. По оси абсцисс расположены временные промежутки в месяцах. По оси ординат количество продажи игристого вина в тысячах литрах.

Второй график представляет тренд исходного временного ряда. Тренд - это общая тенденция в какой-либо индустрии, которая меняется с течением времени. В данном случае можно точно сказать, что тренд не линейный.

Третий график представляет сезонность исходного временного ряда. Сезонность - это характеристика временного ряда, в которой данные претерпевают регулярные и предсказуемые изменения, повторяющиеся каждый календарный год. Из графика видно, что ряд имеет ежегодную сезонность.

Выбросы во временных рядах – это значения, которые значительно отличаются от закономерностей и трендов других значений ряда. На четвертом графике показаны выбросы во временном ряду.

1. Построение прогностической модели

Чтобы построить прогностическую модель и соответственно сделать прогноз продажи игристого вина в Австралии, нам необходимо сформировать обучающие данные, на основании которых и будет формироваться прогноз.

Проведем анализ выбранных данных и простроим прогноз.

def predict\_the\_result(train):  
 model = SARIMAX(train, order=(3, 0, 0), seasonal\_order=(0, 1, 0, 12))  
 result = model.fit()  
  
 start\_point = len(train)  
 end\_point = start\_point + month  
  
 return result.predict(start\_point, end\_point)

SARIMAX – это сезонная авторегрессионная интегрированная скользящая средняя с моделью экзогенных регрессоров. Обычно применяется для анализа временных рядов. Она позволяет моделировать зависимость между наблюдаемыми значениями и их прошлыми значениями, а также между текущими значениями и разностями между прошлыми значениями. После отработки функции sarimax, мы делаем подгонку данных встроенным методом fit. Применив метод predict, передав в качестве параметров промежуток времени в месяцах, мы получаем прогноз на заявленный срок.

Теперь, когда у нас есть все необходимые данные мы можем провести анализ прогнозирования. Результат представлен на Рисунке 3.

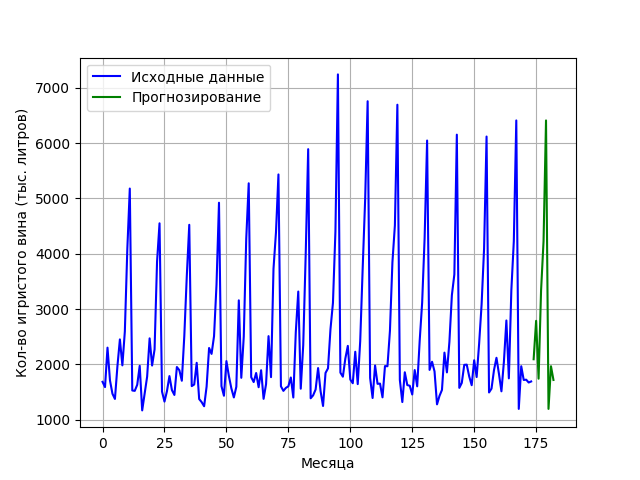


Рисунок 3 - График предсказания на 8 месяцев

Заключение

В результате данной работы мы проанализировали временной ряд на такие компоненты как тренд, сезонность и выбросы. На основании входных данных построили прогностическую модель и сделали предсказание на 8 месяцев.