МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет»
(национальный исследовательский университет)»
Высшая школа экономики и управления
Кафедра «Информационные технологии в экономике»

ПРОЕКТ ПРОВЕРЕН
Рецензент, старший менеджер ООО НПФ «РД Технология»
_____ (М.М. Игнатов)
«___»_____ 2019 г.

ДОПУСТИТЬ К ЗАЩИТЕ
Заведующий кафедрой, д.т.н., с.н.с.
_____ Б.М. Суховилов
«___»_____ 2019 г.

Разработка модели анализа и прогнозирования продаж торгового предприятия с использованием метода ARIMA

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА К ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ

ЮУрГУ–38.04.05.2019.132 ПЗ ВКР

Руководитель работы, к.т.н., доцент
_____ Е.В. Бунова
_____ 2019 г.

Автор работы
студент группы ЗЭУ-356
_____ С.Ю. Выдря
_____ 2019 г.

Нормоконтролер, к.т.н., доцент

_____ Е.В. Бунова
_____ 2019 г.

Челябинск 2019

АННОТАЦИЯ

Выдря С.Ю. Разработка модели анализа и прогнозирования продаж торгового предприятия с использованием метода ARIMA. – ЮУрГУ, ЗЭУ-356, 80 с., 4 табл., 20 рис., библиогр. список – 37 наим., 2 прил.

Данная работа посвящена построению модели анализа и прогнозирования объема продаж торгового предприятия (на примере ООО НПФ «РД Технология»).

Введение содержит описание актуальности темы, формулировку целей и задач работы, раскрывает основные научные методы, используемые в работе, и определяет содержательную часть работы.

Первая глава работы описывает теоретический аспект методов прогнозирования, применяемых на сегодняшний день, их взаимосвязь с методами машинного обучения, рассматриваются классы и задачи машинного обучения и аргументируется выбор алгоритмов для прогнозирования.

Во второй главе описано формирование модели прогнозирования и обработки исходных данных, описаны основные методы, применяемые при обучении прогнозирования объемов продаж и продемонстрирована работа с программными модулями.

В третьей главе разработаны основные направления коммерциализации проекта.

В результате работы с помощью языка программирования Python и программного обеспечения Anaconda разработана интегрированная модель авторегрессии. На основе результатов построенных моделей сформулированы рекомендации по повышению объема продаж предприятия розничной торговли, в частности ООО НПФ «РД Технология».

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ……..…….....7
1 ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ……..…….…….....11
1.1 Анализ терминов и определений по проблематике прогнозирования…….…….....11
1.2 Анализ факторов, влияющих на объем продаж торговой организации…….....18
1.3 Анализ существующих методов прогнозирования…….....23
1.3.1 Машинное обучение: основные понятия и обозначения…….....23
1.3.2 Модель временных рядов…….....26
1.3.3 Регрессия на основе нейросетевого подхода…….....30
2 ПОСТРОЕНИЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ..... .24
2.1 Выбор инструментов реализации задачи прогнозирования…….....36
2.1.1 Язык программирования R…….....36
2.1.2 Сервис Amazon Machine Learning…….....36
2.1.3 Библиотека scikit-learn……...…….....37
2.1.4 Язык программирования Python…….....39
2.2 Построение модели анализа и прогнозирования продаж ООО НПФ «РД Технология»….....42
3 КОММЕРЦИАЛИЗАЦИЯ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ…..…….....58
3.1 Конкурирующие решения…….....62
3.2 Web-сайт…….....64
3.2.1 Структура сайта……...…….....64
3.2.2 Основные функциональные страницы…….....62
ЗАКЛЮЧЕНИЕ…….…….,….….....67
БИБЛИОГРАФИЧЕСКИЙ СПИСОК……..…….....68
ПРИЛОЖЕНИЕ А. МОДУЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ С ПРИМЕНЕНИЕМ МОДЕЛИ ARIMA НА ЯЗЫКЕ PYTHON…….…….....72

ВВЕДЕНИЕ

Актуальность. В современном динамично развивающемся мире предприятиям приходится постоянно сталкиваться с изменяющейся рыночной конъюнктурой. Изменение спроса на продукцию может быть вызвано различными факторами, такими как изменение емкости рынка, ухудшение жизненного уровня населения, законодательные нововведения и др. Прогнозировать продажи исключительно методами экспертных оценок, которые основываются на субъективном мнении специалистов, являющихся экспертами в определенной области, становится затруднительно. Точный прогноз позволяет снизить издержки по хранению продукции, оптимизировать загрузку производственных мощностей, сократить расходы на транспорте. Все большее число компаний применяют специальное программное обеспечение для прогнозирования продаж. Использование предприятиями информационных технологий является необходимостью для их успешной деятельности.

Правильно составленный прогноз увеличивает эффективность ведения бизнеса путем контроля и оптимизации расходов, что, в свою очередь, помогает сформировать оптимальные (а не завышенные или заниженные) запасы продукции на складе. Таким образом в условиях рыночной экономики становится актуальным вопрос организации оперативного контроля и управления запасами материальных ресурсов на предприятии. Также на предприятии существует необходимость обработки большого количества номенклатуры. Решением данных проблем в определенной степени способствует внедрение автоматизированных систем, например таких как 1С–предприятие, которые позволяют наладить учет движения материальных ресурсов, то есть спрогнозировать какое количество товара из конкретной категории предприятие может реализовать потребителям в различные временные промежутки при условии сохранения интереса потребителей.

Задачи снижения товарного запаса и повышения оборачиваемости напрямую связаны с точностью прогнозирования продаж. При расчете страхового запаса среднее отклонение продаж от прогнозов является одной из основных составляющих. Поэтому повышение точности прогноза на 30–40% может дать повышение оборачиваемости на 15–20%, а также позволит увеличить продажи из-за снижения количества отсутствующего запаса или распроданного товара.

Нормативно-правовое регулирование развития малого и среднего предпринимательства в Российской Федерации основывается на Конституции Российской Федерации и осуществляется Федеральным законом от 24.07.2007 № 209-ФЗ (ред. от 29.06.2015) «О развитии малого и среднего предпринимательства в Российской Федерации», другими федеральными законами, принимаемыми в соответствии с ними иными нормативными правовыми актами Российской Федерации, законами и иными нормативными правовыми актами субъектов Российской Федерации, нормативными правовыми актами органов местного самоуправления.

Основные нормативные документы, регулирующие развитие малого и среднего предпринимательства:

Федеральный закон от 24.07.2007 N 209-ФЗ (ред. от 28.11.2018) "О развитии малого и среднего предпринимательства в Российской Федерации" [1];

Конституция РФ, которая закрепляет общие принципы правового регулирования предпринимательской деятельности, устанавливает минимум гарантий прав и интересов участников предпринимательских правоотношений, который не может быть ограничен, закрепляет предметы ведения Российской Федерации и субъектов РФ [2].

Таким образом, тема исследования является актуальной в соответствии с действующим законодательством РФ в области развития малого и среднего предпринимательства.

Цель работы: формирование модели прогнозирования объема продаж на предприятии розничной торговли на основе методологии Бокса-Дженкинса.

Для достижения поставленной цели необходимо решить следующие задачи:

1) проанализировать существующие методы анализа и прогнозирования данных на предмет набора показателей;

2) собрать данные для формирования модели прогнозирования продаж;

3) сформировать методологию анализа и прогнозирования объема продаж торгового предприятия;

4) построить модель прогнозирования;

5) описать основные направления коммерциализации разработанной методологии.

Объектом исследования является предприятие розничной торговли как система товарных потоков.

Предметом исследования является объем продаж ООО НПФ «РД Технология» за 2009-2018 гг.

Информационной базой исследования составляют данные по объему продаж ООО НПФ «РД Технология» за 2009-2018 гг., материалы официальных сайтов сети Интернет и электронных средств массовой информации по исследуемой теме.

Рекомендации, предложенные в ходе исследования, могут быть использованы в практической деятельности в ходе прогнозных расчетов экономической деятельности предприятий.

1 ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

1.1 Анализ терминов и определений по проблематике прогнозирования

Прогнозирование является актуальной задачей в наше время. Это объясняет большое разнообразие исследований на данную тему.

Прогнозирование социально-экономических и иных показателей, которые можно охарактеризовать таким общим понятием как временной ряд, описано в исследованиях ряда зарубежных и отечественных ученых: Р. Хиндмана [3, 4], Дж. Атанасопулоса [5], J. S. Racine [6], Q. Li [6], T. Hayfield [7], С. А. Айвазяна [8, 9], Б. Б. Демешева [10], И. С. Светунькова [11, 12], С. Г. Светунькова [13], N. Kourentzes [13], А. И. Орлова [14], В. Н. Афанасьева [15], М. М. Юзбашева [15], О. А. Мишулиной [16], О. М. Писаревой [17], Н. А. Садовниковой [18], Р. А. Шмойловой [18], Т. А. Дубровой [19], В. К. Семёнычева [20], Е. В. Семёнычева [20]. Исследования посвящены изучению динамики социально-экономических показателей, их свойствам и математическому моделированию процессов и временных рядов. Представлен широкий обзор тренд-сезонных и авторегрессионных инструментов, но практически не рассматривается сложное многоуровневое моделирование динамических процессов.

Для однозначного определения понятия прогнозирования необходимо отразить общее понятие прогнозирования в современной научной среде. В энциклопедических источниках прогнозирование характеризуют как «определение тенденций и перспектив развития тех или иных процессов на основе анализа данных об их прошлом и нынешнем состоянии». В целом, это является достаточно емким и цельным определением процесса прогнозирования, отражающим его основную суть.

Еще одним емким определением, сформулированным в форме вопроса, является определение из Энциклопедии эпистемологии и философии науки: прогнозирование – это «рассуждение, направляемое вопросом: «Какое событие (явление, ситуация, свойство, процесс) возможно, если имеет место ситуация q?». Здесь важным моментом является, то что ситуация q определена набором

характеристик, известных исследователю. Соответственно, для того чтобы понять вероятность наступления события, необходимо формализовать его модель, куда включена ситуация q в связке с указанными характеристиками.

В экономико-математической терминологии прогнозирование – это «система научных исследований качественного и количественного характера, направленных на выяснение тенденций развития народного хозяйства или его частей (отраслей, регионов, предприятий и т.п.) и поиск оптимальных путей достижения целей этого развития». Подобное понятие отражает конкретные аспекты прогнозирования – наличие качественных и количественных выводов, на основании которых осуществляется прогноз, но при этом включает в себя также процедуру управления – в этом случае поиск способов достижения значений полученного прогноза при осуществлении задач. В указанной трактовке понятие прогнозирования становится довольно близко к понятию планирования.

Современное понятие прогнозирования определено Светуньковым И. Г. и Светуньковым С. Г., как «результат именно индуктивного вывода, когда по характеру ограниченного множества значений показателей или взаимосвязи факторов делается вывод о том, что и остальные, еще не наблюдаемые значения этих показателей или взаимосвязи будут обладать аналогичными свойствами». Как и в предыдущих определениях, понятие подразумевает под собой построение модели, которая характеризует исследуемый объект, на основе имеющейся информации.

Между тем, для определения понятия прогнозирования спроса недостаточно представлять понятие прогнозирования и знать об объекте прогнозирования – спросе на товар (товарный запас) в розничной торговле. Это происходит из-за достаточно большого разнообразия подходов к самой методологии прогнозирования спроса.

Например, существует маркетинговый подход к прогнозированию спроса, который может быть отражен в следующем определении Котлера и Келлера: «прогнозирование — искусство предвидения поведения покупателей в определенных обстоятельствах на основе анализа результатов опросов» [40].

Следовательно, оценка будущего спроса здесь видится как величина, которую можно оценить только с помощью инструментов маркетингового анализа. Для этого производится планирование опроса, выборки, формулируются части опросника, проводится непосредственно опрос и на основании его результатов делаются выводы о рыночном спросе на данный товар. Этот способ может быть очень полезен, так как гибко выявляет предпочтения покупателей в зависимости от изменения условий (цены, упаковки и т.п.). Тем не менее, он является достаточно затратным и не всегда оправданным для его использования при автоматизации управления товарными запасами.

Классическое определение прогнозирования спроса дает Лопатников: «Прогнозирование спроса (англ. forecasting of demand) — исследование будущего (возможного) спроса на товары и услуги в целях лучшего обоснования соответствующих производственных планов». При этом уточняется, что для осуществления прогноза используется статистика о реализации товаров, и, как следствие, предполагается анализ предыдущих продаж товара.

Соответственно, из определений видно, что основной сутью прогнозирования спроса является характеристика объекта прогнозирования – в данном случае спрос на товары и услуги – и подход к прогнозированию. В рамках диссертационной работы подход определяется как прогнозирование на основе статистики о предыдущих продажах и условиях, в рамках которых они осуществлялись. Здесь следует выделить пространство классических методов прогнозирования спроса: анализ временных рядов и регрессионно-эконометрический анализ и продвинутые методы машинного обучения. Учитывая выбранные техники, закладывается математическая природа прогнозирования спроса. При этом под математическим прогнозированием понимается формализованный вид прогнозирования, который основан на построении математической модели процесса, улавливающего большую часть его закономерностей.

Просветов Г.И. рассмотрел основные методы экономического прогнозирования и планирования производственной деятельности. Особое

внимание уделил применению вероятностных методов при прогнозировании и планировании производственных программ.

В книге Антохоновой И.В. «Методы прогнозирования социально-экономических процессов» определена классификация и описаны методы прогнозирования для различных социально-экономических объектов. Изложены формализованные и интуитивные методы прогнозирования. Основное внимание уделено статистическим методам прогнозирования и прикладным аспектам разработки прогнозов на различных уровнях иерархии. [28]

У Бутаковой М.М. в работе «Экономическое прогнозирование: методы и приемы практических расчетов» в единой системе излагаются методы разработки экономических прогнозов, включая методику социально-экономического прогнозирования. [29]

Теоретико-прикладной анализ основ социального проектирования провел Леньков Р.В. в книге «Социальное прогнозирование и проектирование». Им рассмотрены основные направления реализации социального прогнозирования и применяемые при этом научные методы. Дана структура модели обоснования и апробации проектных решений. [30]

Автор более 650 научных и учебно-методических работ в области экономики и организации приборостроительного производства, маркетинга и коммерции, методов исследования операций Минько Э.В. в книге «Методы прогнозирования и исследования операций» рассмотрел применение методов прогнозирования и исследования операций в формализации и решении экстремальных задач в экономике, менеджменте, маркетинге и коммерции, ориентированных на оптимизацию управленческих решений с использованием экономических критериев и ограничений. [31]

Черданцева И.В. в рамках научного исследования по проблемам прогнозирования в региональных социально-экономических системах в условиях модернизации экономики России рассмотрела ряд вопросов научной и практической методологии прогнозирования в регионе в современных условиях.

В таблице 1 приведены основные формулировки определения понятия «прогноз».

Таблица 1 – Формулировки определения «прогноз»

№ п/п	Формулировка определения	Источник
1	Научно обоснованное суждение о возможных состояниях объекта в будущем и (или) об альтернативных путях и сроках их осуществления	Сборник терминов Комитета научно-технической терминологии АН СССР. Вып. 92.
2	Предсказание исхода болезни на основании наблюдаемых симптомов	Малый энциклопедический словарь Брокгауза-Ефрона. М.: «ТЕРРА», 1994. – С.1088.
3	Предсказание о развитии и исходе событий или явлений на основании имеющихся данных	Энциклопедический словарь. М.: «БСЭ». – Т. 3. – С. 18.
4	Всякое конкретное предсказание, суждение о состоянии какого-либо явления в будущем	БСЭ, 3-е изд. – М.: Энциклопедия, 1975, Т. 21. – С. 16.
5	Предвидение, предсказание о развитии чего-либо, основанное на определенных данных	Словарь иностранных слов. – М.: «Русский язык», 1988. – С. 410.
6	Заключение о предстоящем развитии и исходе чего-нибудь на основании каких-нибудь данных	Ожегов С.И. Словарь русского языка. – М.: «Русский язык», 1978. – С. 558.
7	Результат прогнозирования, означающий опережающее отображение действительности, основанное на познании законов и механизмов развития исследуемого явления либо объекта природы, общества и мышления	Управление организацией: Энциклопедический словарь. – М.: ИНФРА-М, 2001. – С. 516.

Из определений видно, что основной сутью прогнозирования спроса является характеристика объекта прогнозирования – в данном случае спрос на товары и услуги – и подход к прогнозированию. В рамках диссертационного исследования подход определяется как прогнозирование на основе статистики о предыдущих продажах и условиях, в рамках которых они осуществлялись. Здесь

следует выделить пространство классических методов прогнозирования спроса: анализ временных рядов и регрессионно-эконометрический анализ и продвинутые методы машинного обучения [32]. Учитывая выбранные техники, закладывается математическая природа прогнозирования спроса. При этом под математическим прогнозированием понимается формализованный вид прогнозирования, основанный на построении математической модели процесса, учитывающего большую часть его закономерностей [33].

Торопов А.С. в статье «Прогнозирование почасового электропотребления региональной энергосистемы с использованием искусственных нейронных сетей» использует методический инструментарий нейронных сетей для краткосрочного прогнозирования почасового электропотребления. Для анализа и построения модели были включены следующие факторы: температура, тип дня (праздничный, выходной, рабочий).

П. Ю. Трофимов, В. Ю. Носков в статье «Прогнозирование временных рядов методом ARIMA» описывают метод прогнозирования ARIMA и представляют основные этапы разработки программного обеспечения, предназначенного для прогнозирования временных рядов. [21] В данной статье авторы используют набор данных «Атмосферное содержание CO2 из образцов воздуха в обсерватории Мауна-Лоа, Гавайи, США», не рассматривая другие параметры.

В статье «Использование моделей ARIMA для прогнозирования валютного курса» [22] Пилюгина А.В. и Бойко А.А. описывают модель ARIMA и методику ее применения для краткосрочного прогнозирования среднемесячного курса доллара США к российскому рублю. Приведена классификация временных рядов с точки зрения их стационарности. Описана процедура проверки временных рядов на стационарность с помощью критерия Дики-Фуллера. Показано, что рассматриваемый временной ряд среднемесячного валютного курса доллар/рубль с января 2009 г. по август 2015 г. является «интегрированным первого порядка». Операция дифференцирования обращает среднемесячный курс в стационарный в

широком смысле временного ряда. Авторы статьи рассматривают динамику курса валюты без влияния других факторов.

Гребенников А.В. в статье «Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA» подбирает модель ARIMA с минимальным числом параметров, имеющую адекватный прогноз. Также представлена процедура оценки параметров модели ARIMA и выбора модели с минимальным числом параметров. Приведены сравнения оценок качества прогноза для полученных моделей. [23]

Стоит отметить, что лишь малое количество вышеупомянутых авторов учитывает в расчётной части, а также на этапе выбора оцениваемых показателей факторы, влияющие на исследуемый процесс и результат прогнозирования.

В нашей же работе для оценки уровня точности прогнозирования, а также изучения факторов влияния на его значение, предлагается использование набора факторов, влияющих на объем продаж торгового предприятия, что позволит производить грамотное и эффективное управление им.

1.2 Анализ факторов, влияющих на объем продаж торговой организации

Необходимость исследования такого экономического показателя как прибыль определяется основной целью функционирования организации - осуществление деятельности, которая дает положительный финансовый результат. Полученная прибыль создает необходимые условия для устойчивости и дальнейшего развития организации. На величину прибыли в той или иной степени оказывают влияние различные факторы. Анализ воздействия отдельных факторов на финансовый результат дает возможность своевременно обнаружить ошибки и разработать соответствующие меры, направленные на решение возникших проблем.

Финансовый результат деятельности организации характеризуется показателем прибыли или убытка. Положительный финансовый результат свидетельствует об эффективном и рациональном использовании средств организации.

Прибыль, как конечный финансовый результат, является одним из основных показателей оценки хозяйственной деятельности организации. Без получения прибыли организация не может развиваться, деятельность, приносящая прибыль, - необходимое условие устойчивого функционирования организации. Прибыль формируется под воздействием многих факторов, что создает предпосылки для анализа и оценки степени их влияния на конечный результат. Финансовый анализ позволяет выявить уязвимые места, требующие особого внимания, и выработать меры по борьбе с ними.

Исследование явлений и процессов, влияющих на получение прибыли, приобретает все большую значимость и актуальность на современном этапе развития экономики, так как деятельность организации зависит от ее способности получать положительный финансовый результат.

Выявление причинно-следственных связей, возникающих в ходе формирования финансового результата, определение и анализ факторов, влияющих на величину прибыли, позволяют обосновать направления развития

организации.

Анализ финансовых результатов деятельности организации предполагает измерение воздействия отдельных составляющих факторов на формирование итогового показателя. Основной задачей факторного анализа является поиск путей и способов максимизации прибыли организации.

Первым шагом по анализу факторов внешней среды является сбор информации о социальных, политических, экономических и технологических тенденциях в изменениях среды существования организаций. Для этого очень часто назначают отдельных сотрудников в организации или приглашают внешних консультантов. Перед ними ставится специальная задача: следить за различными источниками информации (профессиональные журналы, книги и газеты, информационные системы, Интернет, библиотеки, научные исследования, ведущиеся в университетах и НИИ, поставщики, распространители, покупатели, конкуренты и др.). Эти специалисты представляют периодические обзорные доклады руководству (топ-менеджерам), которое отвечает за организацию исследований факторов внешней среды. В крупных организациях подобная работа ведётся постоянно.

Покупательский спрос является наиболее случайной компонентой любой модели управления запасами в торговом предприятии. Чаще всего, при недостаточном его изучении, спрос принимается как детерминированная составляющая. В общем случае это приводит к отклонениям в оценках необходимых товарных запасов. Ранее, было обозначено, что для качественного управления запасов необходима более точная оценка покупательского спроса. Поэтому для создания четкой методологии прогнозирования необходимо изучить как понятие покупательского спроса, так и круг возможных детерминант, определяющих его динамику.

На характер спроса влияет множество факторов, различающихся по природе. Например, авторы [30] выделяют следующие группы факторов:

1. Социальные: социальная структура общества, уровень развития культуры, мода, профессиональный состав населения, уровень безработицы;

2. Экономические: уровень развития экономики страны, региона, размер денежных доходов, уровень и динамика розничных цен, соотношение товарной и нетоварной форм потребления, степень обеспеченности покупателей товарами, величина налогов, культура торговли, уровень цен на коммунальные услуги;

3. Демографические: численность, половозрастной состав потребителей, размер и состав семьи, миграция населения;

4. Природно-климатические: специфика климата, национальные традиции, сезонность продаж;

5. Политические: меры по поддержке малоимущих потребителей, минимальный уровень заработной платы и прожиточного минимума, индексация доходов, международные кризисы, войны, политические конфликты;

6. Товарная политика: качество товаров, насыщенность рынка товарами, современная техника и технология производства, широта ассортимента, наличие товаров-заменителей, взаимодополняющих товаров;

7. Прочие факторы: потребительские ожидания, предпочтения потребителей, сегмент рынка, реклама.

Как видно из списка, факторов, влияющих на потребительский спрос, довольно много. Но также необходимо понимать, что на практике задача значительно усложняется:

• Затруднен поиск информации обо всех предлагаемых факторах в достаточном количестве и объеме. Это настолько сложная деятельность, что стоимость совершаемых действий при этом, может оказаться значительно выше, чем экономический эффект от создания сверхточной математической модели прогнозирования спроса;

• Внутренние бизнес-процессы розничного предприятия оказывают специфическое влияние на товарный спрос. Например, реализуемая маркетинговая программа в праздничный день может принести магазину дополнительный доход или, наоборот, снизить эффект от праздника в зависимости от качества проводимых мероприятий;

• Ограничения задачи управления запасами. Прогнозирование спроса в

рамках задачи управления запасами в общем случае осуществляется на краткосрочный период. Это значит, что, во-первых, спрос оценивается на каждодневной основе, во-вторых, прогноз совершается и используется при принятии решений на горизонт не более 2-х недель или месяца. Это говорит о том, что в качестве основных детерминант могут быть использованы только те показатели, которые обновляются достаточно часто и могут являться своевременными индикаторами изменения спроса.

Методика выбора основных факторов, влияющих на покупательский спрос определяется с помощью корреляционного анализа. В этом случае предполагается цельный анализ рассматриваемых данных, который включает в себя графический и количественный анализ зависимостей.

На работу торговой сети в значительной мере влияет общее экономическое положение как в стране, так и за рубежом. Можно выделить несколько важных факторов, влияющих на товарный спрос:

1. Реальные располагаемые доходы населения.

2. Монетарные индикаторы: ключевая процентная ставка, ставка по потребительским кредитам.

3. Индекс потребительских и производственных цен.

4. Курс иностранной валюты на отечественную и многие другие.

К сожалению, в целях создания динамичной системы прогнозирования спроса, которая работает на ежедневной основе очень сложно подобрать макропоказатели, которые сигнализируют об изменении потребления достаточно быстро. В открытых источниках информация о большинстве показателей формируется с лагом от 3-х и более месяцев, они также агрегированы по месяцам, кварталам и более широким периодам. Исключением в данном случае является такой индикатор как валютный курс.

При этом влияние валютного курса на потребительский спрос является неоспоримым – повышение международной стоимости национальной денежной единицы (падение валютного курса) сопровождается удешевлением иностранных товаров и удорожанием отечественных. Снижение международной стоимости

отечественной валюты приводит к обратной ситуации.

На графике можно рассмотреть динамику курса доллара США к рублю (устанавливался Центральным Банком России) – основной международной валюте:

Рисунок 1 – Динамика курса доллара США к российскому рублю

Как видно по рисунку 1, в конце 2014 – начале 2015 гг. произошел резкий скачок валютного курса. Это привело к удорожанию импортных товаров, а также отечественных товаров, которые производятся на основе импортного сырья. В значительной мере это влияние (совместно с запретом на ввоз импортных товаров) ощутил на себе рынок бытовой техники.

Исходя из вышеизложенного, можно сделать вывод о том, что предварительный выбор факторов должен осуществляться на индивидуальной основе с учетом описанных выше особенностей. Ниже приведен список факторов, которые могут оказывать влияние на спрос и по которым имеется достаточно полная информация, а также краткое обоснование их включения в модель:

1) Количество покупателей в розничном магазине. Данный фактор позволяет оценить как уровень покупательской активности в магазине влияет на характер и размер спроса на тот или иной товар. Данные по чекам фиксируются в информационной системе предприятия розничной торговли и могут быть использованы для достаточно сложных видов экономико-математического

анализа;

2) Влияние макроэкономических параметров. Из известных макроэкономических параметров, оказывающих достаточно сильное влияние на торговлю и имеющих динамичную природу, стоит выбрать валютный курс. С учетом значительных колебаний курса, наблюдающихся с 2014 года, можно изучить влияние этого показателя на уровень продаж и сделать выводы об его включении в модель прогнозирования;

3) Влияние календарных праздников. Наступление праздников, предположительно, имеет значительное влияние на размер товарного спроса, а также на состав покупательской корзины. Можно предположить, что накануне 23 февраля активнее ведутся продажи крепкого алкоголя или мужской парфюмерии, в то время как перед 8 марта покупатель скорее возьмет вино, конфеты и духи;

5) Влияние погодно-климатических условий. Погодный фактор сказывается больше всего на товарах повседневного спроса. Учитывая характер данных применяемых в исследовании, использовать данный фактор – обоснованная мера. В качестве иллюстрации влияния климатических факторов, можно, допустим, предположить, что в жаркую погоду растут продажи пива и мороженного.

1.3 Анализ существующих методов прогнозирования

1.3.1 Машинное обучение: основные понятия и обозначения

Машинное обучение широко применяется во многих областях, которые, так или иначе, занимаются сбором и анализом данных. Машинное обучение является, по сути, развитием методов аппроксимации функции, но в качестве точек выступают более сложные объекты, элементы сложно-описанных пространств, а ответами могут быть не только числа, но и множества. Задачи машинного обучения сводятся к задаче нахождения неизвестной зависимости между известным множеством объектов и множеством ответов. Т.е. необходимо построить такую функцию, которая бы достаточно точно приближала значения

множества ответов в точках множества объектов и на всем остальном пространстве.

Модель алгоритмов (модель зависимости) – параметрическое семейство отображений, в котором нужно найти функцию, хорошо приближающую целевую зависимость, функциональную или стохастическую зависимость между объектами и ответами. Задача нахождения модели зависимости сводится к построению алгоритма, который бы одинаково точно приближал неизвестную целевую зависимость, как на элементах выборки, так и на всем пространстве объектов. Это задача получила название – обучение с учителем (supervised learning).

Как правило, зависимость параметров строится исходя из законов исследуемой области, например физическая модель. В простейшем случае можно использовать линейную модель.

Метод обучения – отображение, которое ставит в соответствие обучающей выборке алгоритм из заданной модели. Другими словами, алгоритм строится по выборке.

Процесс обучения подразделяется на два этапа:

- Этап обучения – процесс выявления зависимостей в эмпирических данных;
- Этап применения – процесс оценки точности, выявленной зависимости.

На этапе обучения для выявления зависимости используется обучающая выборка (training sample), по ней производится оптимизация параметров. Для оценки качества модели используется тестовая, или контрольная, (test sample) выборка. Во избежание смещения оценки, обучающая и тестовая выборки должны быть независимыми. Для выбора наилучшей модели из множества моделей, построенных на обучающей выборке, используется проверочная выборка (validation sample). Важным условием является то, чтобы обучающая выборка обладала достаточной полнотой, т.е. должна покрывать все возможные случаи.

Для оценки алгоритма, а точнее его применение на выборке, следует, сначала оценить применение алгоритма на отдельном объекте. Для формализации понятия величины ошибки алгоритма на объекте вводится понятие – функция потерь. Функция потерь характеризует величину отклонения ответа модели от правильного ответа на произвольном объекте. Эмпирический риск – функционал алгоритма на обучающей выборке, характеризует качество приближения заданной функции на выборке и равна среднему значению потерь (сумма значений функции потерь по обучающей выборке, деленная на длину выборки). Другими словами, эмпирический риск – средняя величина ошибки алгоритма на обучающей выборке. Таким образом, задачи обучения сводятся к задачам оптимизации, нахождению функций приближения с минимальным значением функции потерь, т.е. сводить задачи к численным методам оптимизации.

Обучающая (обобщающая) способность алгоритма характеризуется соотношением точности приближений на тестовой и обучающей выборках. Если вероятность возникновения ошибок на тестовой выборке мала и близка к значению вероятности ошибок на обучающей выборке, то можно говорить о хорошей обучающей способности алгоритма, иначе мы сталкиваемся с такими явлениями, как недообучение и переобучение.

Недообучение возникает при использовании недостаточно сложных моделей и характеризуется большой величиной ошибки на обучающей выборке.

Переобучение возникает при использовании чрезмерно сложных моделей. Переобучением называют состояние, когда вероятность ошибки на объектах тестовой выборки существенно выше вероятности на обучающей. Переобучение возникает из-за избыточной сложности пространства параметров, лишние степени свободы в модели «тратятся» на чрезмерно точную подгонку под обучающую выборку. Переобучение есть всегда, когда есть оптимизация параметров по конечной (заведомо неполной) выборке. Существуют различные методы обнаружения переобучения.

Эмпирически переобучение можно обнаружить с помощью скользящего контроля.

Совсем избавиться от переобучения нельзя, можно его лишь минимизировать. Один из способов - введение ограничений на пространство параметров.

1.3.2 Модель временных рядов

Наиболее часто используемым подходом в прогнозировании спроса является инструментарий анализа временных рядов. Временным рядом обычно называют последовательность измерений $y_{t1}, y_{t2}, \ldots, y_{tN}$ случайной величины ξ_t, которые произведены в последовательные моменты времени $t_1, t_2, \ldots, t_N$ [16]. В рамках рассматриваемой задачи речь идет о временном ряде спроса на тот или иной товар. Соответственно, с помощью методов анализа временных рядов, возможно формализованное представление математической модели спроса как изменяющейся во времени случайной величины ξ_t с определенным законом распределения.

Наиболее популярными видами моделей временных рядов для прогнозирования спроса являются:

- Тренд-сезонные модели;
- Модели экспоненциального сглаживания;
- Модели авторегрессии и скользящего среднего.

Это объясняется тем, что данные модели хорошо аппроксимируют специфические элементы временного ряда спроса, например, сезонность.

Тренд-сезонные модели являются фундаментальными моделями временных рядов, которые определяют основные составляющие временного ряда: T – тренд временного ряда или так называемую тенденцию (направление), в которой движется временной ряд, S – сезонная компонента временного ряда, которая отражает амплитудные колебания вокруг линии тренда, и E (или ε) – ошибка

модели. В зависимости от характеристик процесса элементы временного ряда могут быть объединены двумя способами:

• аддитивно:

$$y_t = T_t + S_t + \varepsilon_t \quad (1)$$

• мультипликативно:

$$y_t = T_t * S_t * \varepsilon_t \quad (2)$$

В данном случае главной процедурой по созданию модели временного ряда является декомпозиция или разложение исходного временного ряда на составляющие T, S и E. При этом в зависимости от спецификации модели ошибка ε_t распределена следующим образом:

$\varepsilon_{t\sim N}(0,\sigma_\varepsilon^2)$ – ошибка нормально распределена для аддитивных моделей и $\varepsilon_t \sim LogN(0,\sigma_\varepsilon^2)$ – логнормально для мультипликативных. Существует большое количество вариантов декомпозиции: классическая, декомпозиция X-12, декомпозиция с помощью создания локальных регрессий. Однако, следует обозначить, что, несмотря на довольно простую интерпретацию процесса, методы, основанные на декомпозиции, в среднем отличаются более низким качеством прогнозирования. Несмотря на это, принцип разложения ряда на составляющие используется во многих современных моделях временных рядов, что будет показано в рамках данного исследования.

Метод экспоненциального сглаживания является адаптивным методом прогнозирования. Это значит, что с получением новой информации о временном ряде, модель способна подстраиваться под изменения, тем самым повышая свою прогностическую способность. Идею экспоненциального сглаживания отражает самая простая спецификация – модель простого экспоненциального сглаживания:

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}, \quad (3)$$

где $\hat{y}t$ – прогнозируемое значение временного ряда в момент t, y_{t-1} – фактическое значение временного ряда в момент $t-1$, α – параметр сглаживания,

принимающий значения от 0 до 1. Следовательно, речь идет о процедуре «сглаживания» исходного ряда, которая позволяет увидеть изменения в тенденциях.

Существует множество модификаций исходной модели простого экспоненциального сглаживания, в том числе которые позволяют учесть компоненты временного ряда. Здесь следует остановиться наиболее подробно на подходе пространства состояний (в англ. state space approach). Важной характеристикой подхода является то, что он базируется на декомпозиции временного ряда, о которой было сказано ранее. Отсюда модель экспоненциально сглаживания может быть аддитивной (1) и мультипликативной (2). Составляющие ряда (E,T,S) – будь то E, T или S – адаптируют составляющие ряда с учетом их характера. В некотором смысле существуют два глобальных класса моделей, которые имеют собственную форму. Внутри каждого класса находятся подвиды в зависимости от адаптации конкретных компонент – каждая компонента также может быть либо аддитивной, либо мультипликативной. Аддитивный класс моделей имеет следующий общий вид:

$$\begin{cases} y_t = \omega' v_{t-l} + \varepsilon_t \\ v_t = F v_{t-l} + g\varepsilon_t \end{cases}, \quad (4)$$

где y_t – фактическое значение временного ряда в момент t, v_t – это вектор состояний, который содержит в себе компоненты (E, T, S), w′ – заданный измерительный вектор, F – матрица переходов, g – вектор, который содержит в себе постоянные сглаживания и ε_t – ошибка модели, распределенная нормально $N(0,\sigma_\varepsilon^2)$.

Второй класс – мультипликативные модели – описываются с помощью логарифмов, аналогично предыдущей форме:

$$\begin{cases} y_t = \exp(\omega' \log(v_{t-l}) + \log(1 + \varepsilon_t)) \\ \log(v_t) = F \log(v_{t-l}) + \log(1 + g\varepsilon_t) \end{cases},$$

(5)

В данном выражении ошибка $1 + \varepsilon_t$ распределена логнормально $\mathrm{Log}N(0,\sigma_\varepsilon^2)$.

Выбор лучшей модели из пространства осуществляется с помощью информационных критериев. Классически используется информационный критерий Акаике, который для класса данных моделей выражен в следующей формуле:

$$AIC = -2Log(L) + 2k, \quad (6)$$

где L – максимизированное значение функции правдоподобия модели, k – общее количество параметров модели. Выбор модели осуществляется по наибольшему значению информационного критерия.

Модель авторегрессии ARIMA является наиболее универсальным инструментом для прогнозирования временных рядов. Их суть заключается в объединении двух смежных процессов. Первый процесс – авторегрессии, который характеризуют зависимость текущих значений ряда от его предыдущих значений:

$$\hat{y}_t = c + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \quad (7)$$

где y_{t-1}, y_{t-p} – предыдущие фактические значения временного ряда, сдвинутые от фактического на лаг от 1 до p, c – константа модели, ε_t – ошибка модели, распределенная нормально. Общую спецификацию модели обозначают как AR (P).

Моделью скользящего среднего порядка q называется модель $MA(q)$

$$\hat{y}_t = c + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}, \quad (8)$$

где $\varepsilon_{t-1},\ldots,\varepsilon_{t-q}$ – предыдущие ошибки модели, сдвинутые от фактической на лаг от 1 до q.

Согласно методологии Бокса-Дженкинса для определения основных параметров интегрированной модели авторегрессии – скользящего среднего используется автокорреляционная (ACF) и частная автокорреляционная функции ($PACF$). Автокорреляционная функция состоит из коэффициентов корреляции текущего значения ряда и p лагов – $corr(y_{t+p}, y_t)$. Отсюда частная автокорреляционная функция принимает следующий вид:

$$pacf(p) = \begin{cases} corr(y_{t+p}, y_t), p = 1 \\ corr(y_{t+p} - y_{t+p}^{p-1}, y_t - y_t^{p-1}), p > 1 \end{cases},$$

(9)

где y_t^{p-1} – линейная регрессия на y_{t+1}, y_{t+2}, …, y_{t+p+1}.

Исходя из характера автокорреляционных функций можно сделать выводы о параметрах моделей $ARIMA$. Сама модель имеет следующую стандартную спецификацию:

$$\hat{y}_t = c + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} + \varepsilon_t \qquad (10)$$

По сути -моделирование заключается в корректном подборе спецификации модели $ARIMA(p,d,q)$, где d – порядок разности временного ряда, который нужен для обращения нестационарного временного ряда в стационарный. Существует два понятия стационарности временного ряда: строгая стационарность и слабая стационарность. В первом случае случайный процесс называется строго стационарным, если сдвиг во времени не меняет ни одну из функций плотности распределения.

1.3.3 Регрессия на основе нейросетевого подхода

Ведущим алгоритмом машинного обучения, который в современном мире используется в качестве теоретической основы многих программ по прогнозированию, является искусственная нейронная сеть. Искусственная нейронная сеть есть адаптация биологического подхода к организации систем высокого порядка. Это значит, что в основе метода лежит взаимодействие нейронов – нервных клеток в живом организме. Нейроны реагируют на раздражители и передают информацию к следующим нейронам. Организация процесса в искусственной нейронной сети аналогична. Как показала практика, применение этого подхода оправдано и успешно, в особенности в задачах распознавания изображений, видеоряда и звука.

Рассмотрим возможности многослойной нейронной сети (от англ. MLP – multi layer perceptron). По сути это однонаправленная сеть, в которой нейроны расположены на разных уровнях, соответствующих понятию слоя. При решении

простых задач машинного обучения используются многослойные нейронные сети с одним скрытым слоем, как это представлено на рисунке 2, реже с двумя.

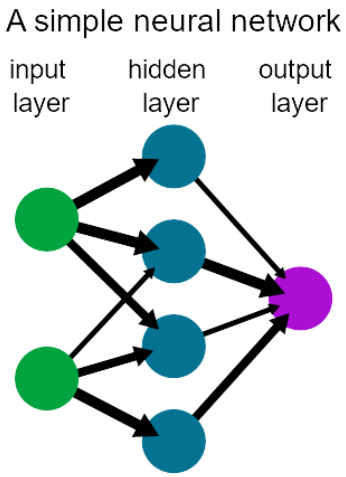


Рисунок 2 - многослойный персептрон

Существует несколько типов функционирования нейронов, т.е. базовых единиц нейронной сети. Это может быть простой персептрон, функция активации которого описывается следующим образом:

$$y_i(u_i) = \begin{cases} 1 \text{ для } u \geq 0 \\ 0 \text{ для } u < 0 \end{cases}, \tag{11}$$

где u_i – выходной сигнал сумматорной функции

$$u_i = \sum_{j=0}^{N} w_{ij} x_j, \tag{12}$$

где x_j – это значение входного сигнала, w_{ij} – значение веса выходного сигнала.

Гораздо чаще при реализации нейронов в нейронной сети используют сигмоидальный вид активационной функции. Он имеет неоспоримые преимущества, так как позволяет использовать градиентные методы при обучении нейронной сети. Сигмоидальный вид функции как правило имеет 2 реализации:

$$f(x) = \frac{1}{1 + e^{-\beta x}}, \quad (13)$$

которую можно обозначить как униполярную функцию. Биполярная функция задается с помощью тангенса:

$$f(x) = \tanh(\beta x) \quad (14)$$

Здесь β – параметр, который выбирается пользователем. По сути он влияет на форму кривой активации. Сигмоидальный нейрон обучается методом обучения с учителем, то есть имеется выборка примеров $\langle x, d \rangle$, где x – набор входящего вектора параметров, d – размеченный примеры напротив каждого входного набора. При этом проводится минимизация целевой функции:

$$E = \frac{1}{2}(y_i - d_i)^2, \quad (15)$$

где $y_i = f(u_i) = f(\sum N w_{ij} x_j)$, x – входной вектор со значениями $x = [x_0, x_1, \dots, x_N]^T$. Здесь x_0 – это единичный вектор и $x_0 = 1$ при поляризации или $x_0 = 0$ в случае ее отсутствия.

Помимо многослойного персептрона существует большое количество иных видов нейросетевых технологий: радиальные базисные сети, рекуррентные нейронные сети, сверточные нейронные сети, нейронные сети с применение нечеткой логики. Выбор того или иного типа нейронной сети, настройка ее топологии, обычно определяется ресурсными возможностями и характеристикой самой задачи.

В целом применение нейронных сетей в задачах экономики приобрело обширный характер. В том числе нейронные сети используются и для прогнозирования уровня спроса и продаж [18]. Очень часто нейронные сети применяются для прогнозирования экономических временных рядов, что также сходно по типу с исходной задачей диссертационной работы по прогнозированию спроса. Все это делает нейронные сети методом с очень

широкой областью применения. При этом они не лишены недостатков, связанных со специальной подготовкой данных и достаточно высокими требованиями на вычислительные ресурсы. Это приводит к тому, что на практике, при решении задач прогнозирования на большом объеме данных, использование нейросетевых технологий становится слишком затратным.

Таблица 2 – Сравнение методов прогнозирования

Модель	Достоинства	Недостатки
Регрессионная модель прогнозирования	• Быстрота получения результата; • Доступность промежуточных вычислений; • Простота моделей;	• Трудоемкость определения параметров; • Возможность моделирования лишь линейных процессов; • Сложность определения вида функциональной зависимости.
Авторегрессионная модель	• Быстрота получения результата; • Доступность промежуточных вычислений; • Простота моделей; • Разнородность решаемых задач.	• Трудоемкость определения параметров.
Модели экспоненциального сглаживания	• Простота моделей; • Быстрота получения результата; • Решение задач долгосрочного прогнозирования.	• Отсутствие гибкости

Окончание таблицы 2

Нейросетевые модели	• Возможность моделирования не линейных процессов; • Адаптивность; • Масштабируемость; • Разнородность решаемых задач.	• Сложность программной реализации; • Отсутствие промежуточных вычислений; • Высокие требования к непротиворечивости обучающей выборки.
Модели на базе цепей Маркова	Простота моделей	• Узкая применимость моделей; • Невозможность решения задач прогнозирования с длинной памятью.

В рамках исследования для построения модели анализа и прогнозирования исходного временного ряда была выбрана интегрированная модель авторегрессии ARIMA.

Выбор модели ARIMA был сделан на основании следующих ее преимуществ:

1. Подход Бокса-Дженкинса к анализу временных рядов является весьма мощным инструментом для построения точных прогнозов с малой дальностью прогнозирования;
2. Модели ARIMA достаточно гибкие и могут описывать широкий спектр характеристик временных рядов, которые встречаются на практике.
3. В подавляющем большинстве случаев прогнозы, сделанные по методике ARIMA, точнее чем прогнозы, сделанные классическими статистическими методами. Это неудивительно – ARIMA более сложный алгоритм, опирающийся на автокорреляцию, очевидно, что он, в целом, не может работать хуже, чем алгоритм, основанный на сглаживании. Кроме того, если анализировать научные публикации последнего десятилетия можно видеть, что в подавляющем количестве работ по прогнозированию классическими методами используется именно ARIMA, как наиболее

обоснованный и надежный алгоритм (из статистических). Результаты нашего теста в очередной раз подтверждают это.

Также выбор модели ARIMA обусловлен тем, что наш временной ряд, представленный в разделе 2.2 является нестационарным, а модель ARIMA прекрасно подходит для анализа, преобразования и прогнозирования таких рядов.

Выводы по главе 1

Подведём итоги проведённого в главе 1 исследования. В данном разделе были рассмотрены различные варианты трактовок понятия «прогнозирование», имеющиеся в данный момент в научной литературе. Анализ показал, что на сегодняшний день нет единого мнения по поводу того, что должно вкладываться в данное понятие.

В результате проделанной работы мы провели обзор актуальности исследуемой темы, дали обоснование важности рассматриваемого вопроса, убедились в важности прогнозирования для торгового предприятия, пришли к выводу о необходимости применения модели авторегрессии в данной сфере. Были сформулированы цели и задачи диссертационного исследования, определена актуальность темы.

2 ПОСТРОЕНИЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ

2.1 Выбор инструментов реализации задачи прогнозирования

Благодаря широкому развитию концепции «больших данных», распространение получил такой раздел информатики, как наука о данных (data science). Наука о данных изучает проблемы анализа, обработки и представления данных больших объемов. Наиболее распространенными инструментами для работы с анализом больших данных на сегодняшний день являются R и Python.

2.1.1. Язык программирования R

R – язык программирования, предназначенный для статистической обработки данных и работы с графикой, но в то же время — это свободная программная среда с открытым исходным кодом, развиваемая в рамках проекта GNU. R получил широкое распространение в сферах, где проводится работа с данными. Основная вычислительная мощь R заключается в статистическом анализе, однако он также обладает обширным функционалом для первичного анализа данных (построение графиков и таблиц сопряженности) и математического моделирования.

2.1.2 Сервис Amazon Machine Learning

Еще один инструмент работы с большими данными – сервис Amazon Machine Learning, в основе которого лежит применение методов машинного обучения. Данный сервис упрощает использование технологий машинного обучения для разработчиков любого уровня квалификации. Amazon предоставляет учебники, руководства, полезные объяснения основных методов машинного обучения и достаточно советов, для совершения первых самостоятельных шагов в области применения этих методов.

Данный сервис предоставляет возможность составления прогноза целевой переменной с помощью заданных начальных условий: набор значений наблюдений и модели.

При инициализации модели в настройках можно задать количество проходов по набору данных и используемый способ регуляризации. В качестве параметра оценки точности предсказания модели используется среднеквадратичное отклонение. Данный сервис также предоставляет возможность отсортировать данные по степени их влияния на целевую переменную. Кроме того, Amazon Machine Learning включает инструменты визуализации для упрощения процесса создания моделей машинного обучения.

2.1.3 Библиотека scikit-learn

Благодаря широкому распространению Python собрал вокруг себя активное сообщество разработчиков, которые в рамках различных проектов разрабатывают модули для узкоспециализированных задач. Одним из таких проектов стал Google Summer of Code 2007, в рамках которого David Cournapeau разработал библиотеку scikit-learn. Разработка данной библиотеки является одной из причин популяризации применения языка Python в области анализа данных с помощью методов машинного обучения.

Библиотека scikit-learn предоставляет реализацию ряда алгоритмов как для обучения с учителем (Supervised learning), так и для обучения без (Unsuoervised learning). Scikit-learn построена на основе стека SciPy (Scientific Python), который включает в себя:

1. NumPy добавляет поддержку больших многомерных массивов и матриц, а также библиотеку высокоуровневых математических функций для операций с ними;
2. SciPy – открытая библиотека высококачественных научных инструментов для языка программирования Python;

3. Matplotlib – библиотека для визуализации двумерной и трехмерной графики;
4. IPython – интерактивная оболочка для языка программирования Python, которая предоставляет расширенную интроспекцию, дополнительный командный синтаксис, подсветку кода и автоматическое дополнение;
5. Sympy – библиотека для работы с символьными вычислениями;
6. Pandas реализует различные структуры данных и анализ. Библиотека scikit-learn состоит из 35 модулей, которые можно подразделить на модули кластеризации, модули оценки модели и количественного определения качества прогнозов, модули работы с наборами данных (предобработка, нормализация), модули работы с признаками (извлечение и выявление наиболее значимых), модули, реализующие различные алгоритмы решения задач классификации и регрессии;

Каждый модуль состоит из классов и функций и решает такие задачи, как:

- Кластеризация (Clustering) – группировка неразмеченных данных;
- Перекрестная проверка (Cross Validation) – оценка эффективности работы модели на независимых данных;
- Наборы данных (Datasets) – для хранения тестовых наборов данных и для генерации наборов данных с определенными свойствами для исследования поведенческих свойств модели;
- Сокращение размерности (Dimensionality Reduction) – набор алгоритмов для уменьшения количества атрибутов для визуализации и отбора признаков (Feature Selection), например, метод главных компонент (Principal Component Analysis);
- Алгоритмические композиции (Ensemble Methods) – набор методов для комбинирования предсказаний нескольких моделей;
- Извлечение признаков (Feature Extraction) – процесс определения атрибутов в данных;

- Отбор признаков (Feature Selection) – набор алгоритмов для выявления значимых атрибутов на основе которых будет построена модель;
- Оптимизация параметров алгоритма (Parameter Tuning) – методы для получения максимально эффективной отдачи от модели;
- Множественное обучение (Manifold Learning) – подход нелинейного сокращения размерности данных.

Отдельно следует выделить методы реализующие обучение с учителем (Supervised Models). Данный набор методов включает в себя:

1. Обобщенные линейные модели (Generalized Linear Models);
2. Методы дискриминантного анализа (Discriminate Analysis);
3. Наивный байесовский классификатор (Naive Bayes);
4. Нейронные сети (Neural Networks);
5. Метод опорных векторов (Support Vector Machines);
6. Деревья принятия решений (Decision Trees).

2.1.4 Язык программирования Python

Python – высокоуровневый универсальный интерпретируемый язык сценариев. При разработке на языке Python большое внимание уделяется простоте и понятности синтаксиса, что не только сокращает время изучения его основ, но и повышает скорость разработки в целом. Это далеко не все преимущества данного языка, основные из них:

- объектно-ориентированность;
- свободное распространение и широкая поддержка;
- кроссплатформенность;
- развитые функциональные возможности.

Язык Python объединяет две парадигмы программирования – объектноориентированную, которая является мощным средством структурированного программного кода для многократного использования, и

процедурную, что расширяет круг решения задач, позволяя использовать средства Python при решении тактических задач с отсутствием фазы проектирования. Объектная модель Python поддерживает понятия полиморфизм, перегрузка операторов и множественное наследование. Кроссплатформенность языка достигается благодаря его реализации на переносимом ANSI C, что позволяет программам, написанным на языке Python, одинаково хорошо компилироваться и выполняться на любых платформах, где установлена совместимая версия Python. Гибридная природа Python объединяет в себе простоту и удобство языков сценариев и мощности компилирующих языков, что делает Python удобным средством разработки приложений различного типа. Однако наибольшая эффективность языка достигается при решении задач анализа данных и автоматизации процессов. Python широко используется в исследовательских работах. Язык программирования Python обладает мощным встроенным инструментарием (встроенные типы объектов и динамическая типизация, автоматическое управление памятью) и возможностью использования внешних библиотек и утилит сторонних разработчиков для решения более узкоспециализированных задач. Python используется не только отдельными пользователями, но и компаниями, включая коммерческое использование.

Например:

- Компания Google широко использует Python в своей поисковой системе и для создания фреймворка App Engine;
- Служба коллективного использования видеоматериалов YouTube в значительной степени реализована на языке Python;
- Популярная программа BitTorrent написана на языке Python;
- Такие компании, как EVE Online и Massively Multiplayer Online Game, широко используют Python в своих разработках;
- Мощная система трехмерного моделирования и создания мультипликации Maya поддерживает интерфейс для управления из сценариев на языке Python;

- Производители электронных устройств и компьютерных компонентов (такие, как Intel, Cisco, Hewlett-Packard, Seagate, Qualcomm и IBM) используют Python для тестирования аппаратного обеспечения;
- Киностудии, занимающиеся созданием визуальных эффектов к кинофильмам (Industrial Light & Magic, Pixar и другие) используют Python в производстве анимационных фильмов;
- Крупнейшие финансовые конгломераты (JPMorgan Chase, Union Bank of Switzerland, Federal Credit Union) применяют Python для прогнозирования финансового рынка;
- Научно-исследовательские центры (NASA, Los Alamos, Fermilab, Jet Propulsion Laboratory и другие) используют Python для научных вычислений;
- Компания по разработке и продаже робото-техники iRobot использует Python в разработке коммерческих роботизированных устройств;
- Компания по производству геоинформационных систем (Environmental Systems Research Institute) использует Python в качестве инструмента настройки своих программных продуктов под нужды конечного пользователя;
- Агентство национальной безопасности Соединенных штатов (National Security Agency) использует Python для шифрования и анализа разведданных.

По данным сообщества kaggle.com на рисунке 3 показано, как распределяются решения задач с использованием языков Python и R на протяжении 2017 года .

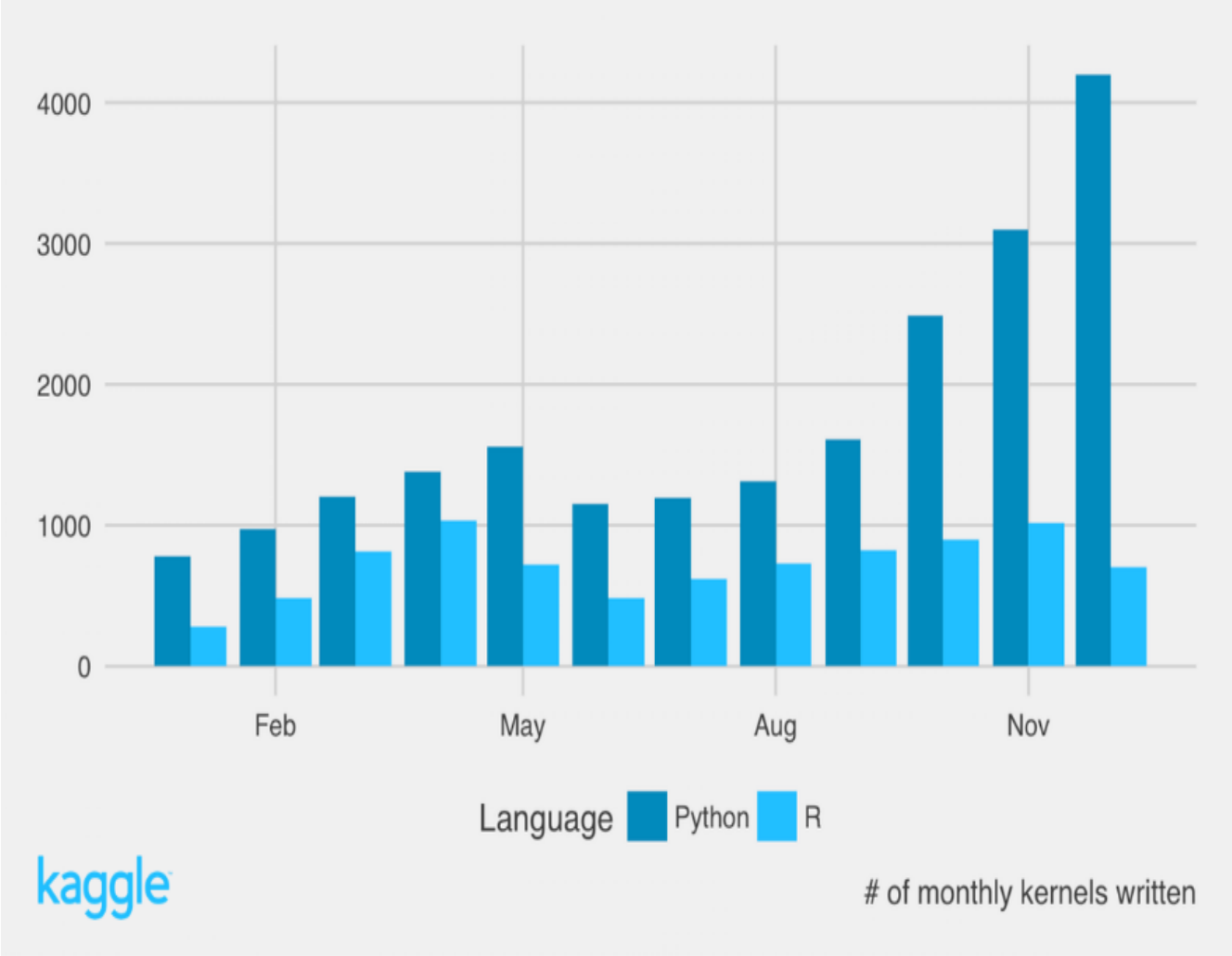


Рисунок 3 – Количество решений задач на языках R и Python на ресурсе Kaggle

На рисунке 3 мы можем наблюдать явное повышение количества решений задач с применением языка Python относительно языка R с течением времени.

На основании вышеизложенного, с учетом всех достоинств и недостатков, а также популярности применения, инструментом для построения нашей модели станет язык программирования Python.

2.2 Построение модели анализа и прогнозирования продаж ООО НПФ «РД Технология»

В данной работе анализ и прогнозирование объема продаж были реализованы с помощью языка программирования Python 3.6 и программного обеспечения Anaconda.

Исходный код разработанного модуля представлен в приложении А.

Таблица с исходными данными представлена в приложении Б и имеет 114 значений о продажах за период с января 2009 г. по июнь 2018 г.

Рассмотрим методологию Бокса-Дженкинса, более известную под названием ARIMA (AutoRegressive Integrated Moving Average), представляющую собой модель авторегрессии и проинтегрированного скользящего среднего. Модель ARIMA может включать в себя сразу обе модели либо каждую по отдельности и обозначается: ARIMA(p,d,q), где p – порядок скользящего среднего. Данная модель является расширением модели ARMA(p,q) (ARMA служит для прогнозирования стационарных временных рядов) и используется, если процесс оказывается нестационарным и для приведения его к стационарному виду потребовалось взять несколько разностей.

Модель ARIMA(p,d,q) для нестационарного временного ряда Yt имеет вид:

$$\Delta^d X_t = c + \sum_{i=1}^{p} a_i \, \Delta^d X_{t-i} + \sum_{j=1}^{q} b_j \, \varepsilon_{t-j} + \varepsilon_t \qquad (16)$$

, где c, ai , Ьj – параметры модели;

Et – стационарный временной ряд;

Δd – оператор разности порядка d.

Для начала импортируем и подключим необходимые для анализа библиотеки.

Далее подключим наш файл с данными в формате .csv с помощью библиотеки Pandas. Здесь основным параметром является index_col – задает номер столбца, который будет использоваться в качестве индекса.

Визуализация играет очень важную роль в анализе временных рядов и прогнозировании. С помощью графика исходных данных можно выявить временные структуры, такие как: тренды, циклы и сезонности, которые могут влиять на выбор модели. Проблема заключается в том, что не всегда используют полный спектр инструментов визуализации, останавливаясь только на линейном графике.

На самом деле инструментов визуализации данных достаточно много, но среди них можно выделить следующие основные виды визуализации:

- Линейный график;
- Гистограммы и график плотности распределения;
- Ящик с усами (Диаграмма размаха);
- Тепловая карта;
- Диаграмма рассеяния;
- График автокорреляции;

Основное внимание уделено одноразмерным временным рядам, но описываемые инструменты можно использовать и на многомерных рядах.

В качестве инструмента визуализации входных данных будет использована библиотека Matplotlib. Matplotlib – библиотека на языке программирования Python для визуализации данных двумерной (2D) графикой (3D графика также поддерживается).

Пакет поддерживает многие виды графиков и диаграмм:

- Графики (line plot);
- Диаграммы разброса (scatter plot);
- Столбчатые диаграммы (bar chart) и гистограммы (histogram);
- Круговые диаграммы (pie chart);
- Контурные графики (contour plot);
- Поля градиентов (quiver);
- Спектральные диаграммы (spectrogram).

Визуализируем наш временной ряд с помощью библиотеки matplotlib:

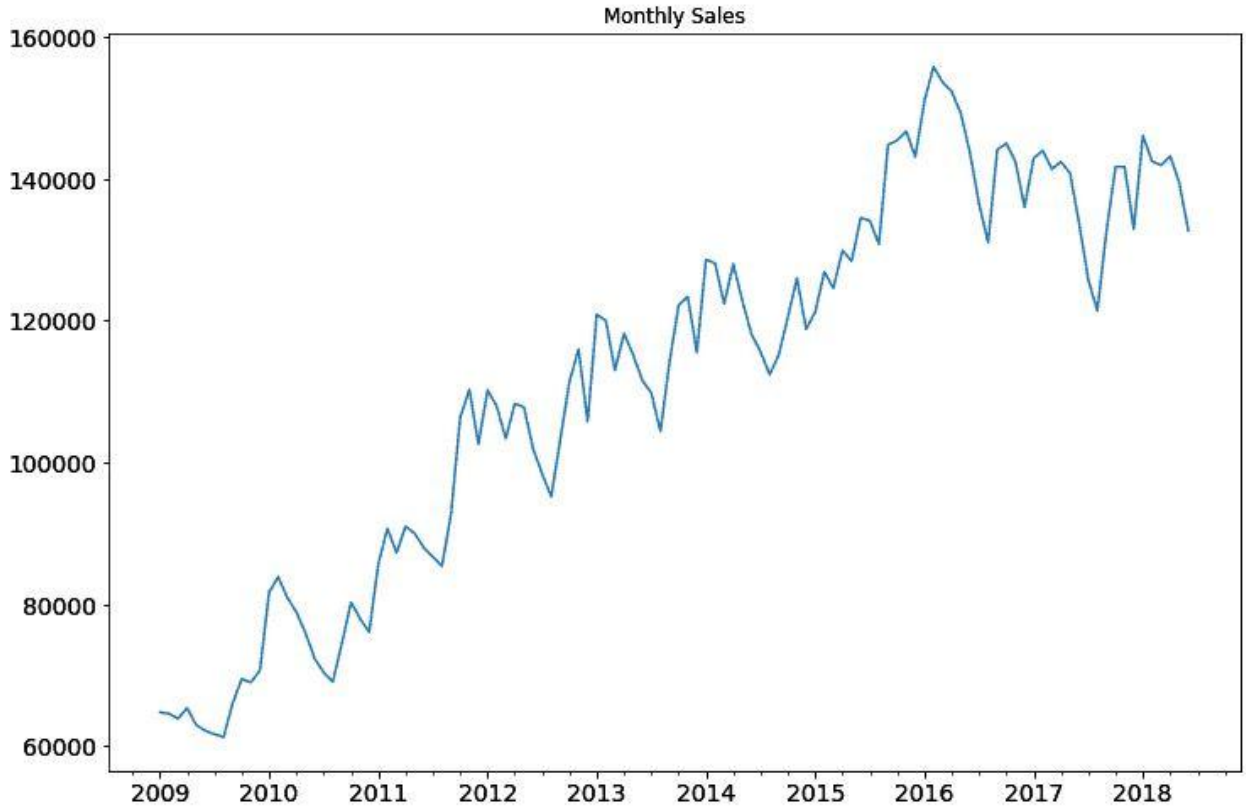


Рисунок 4 – График временного ряда

Проанализировав график, можно сделать вывод о том, что имеется достаточно выраженная сезонность и возрастающий тренд. В 2016 году наблюдается максимальный рост объема продаж.

Другой способ визуализации временного ряда – декомпозиция.

Декомпозиция временного ряда – это выделение тенденции, циклических и случайных колебаний. Проведение декомпозиции позволяет понять структуру ряда, построить его модель и экстраполировать уровни ряда на краткосрочную перспективу. Рассматриваемые компоненты динамического ряда не обязательно имеются в каждом временном ряде. Могут быть ряды динамики, в которых отсутствует тенденция, или периодические колебания, или и то и другое. В этом случае уровни ряда являются функцией случайной компоненты, колебания их значений происходит вокруг среднего уровня, что характерно для стационарного ряда.

Построим модель декомпозиции временного ряда с помощью функции seasonal_decompose, встроенную в библиотеку statsmodels.

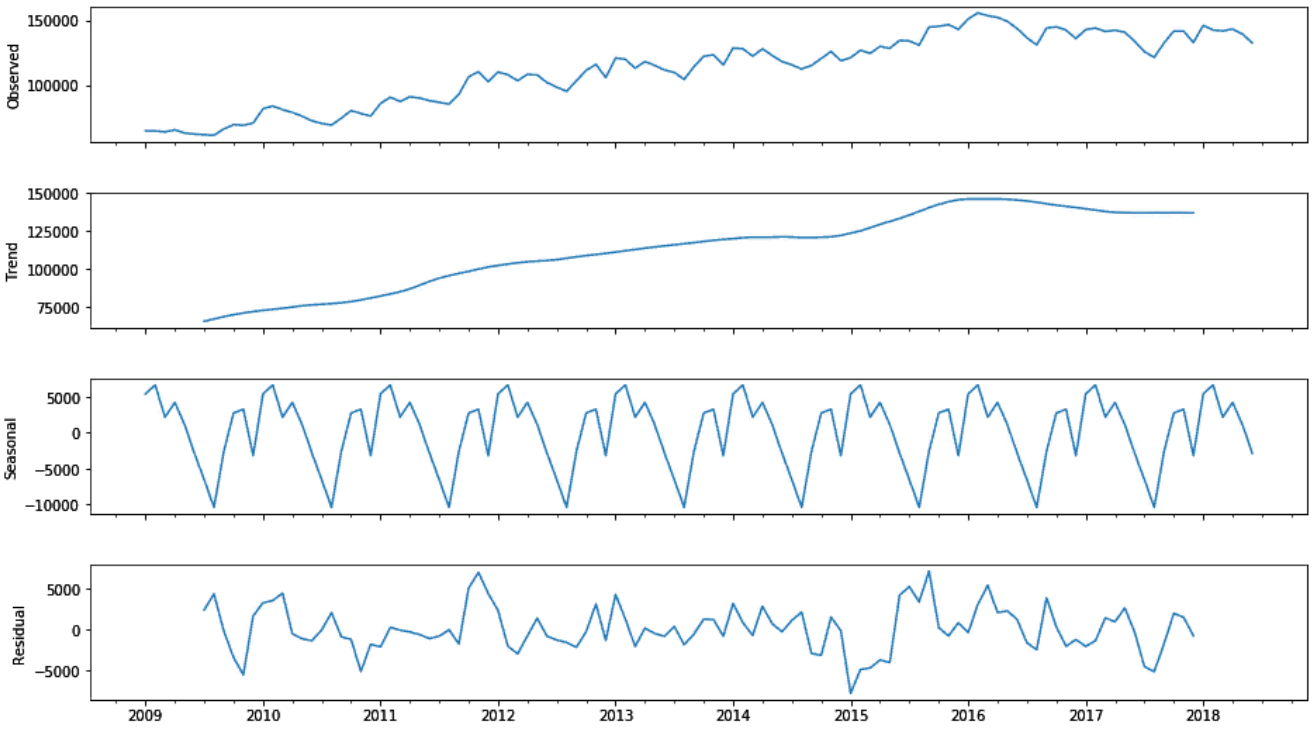


Рисунок 5 – Результат декомпозиции исходного ряда

По результатам декомпозиции ряда можно сделать следующие выводы:

1) наблюдается устойчивый возрастающий тренд (график Trend);

2) сезонный профиль достаточно хорошо выражен и имеет постоянный характер (график Seasonal);

3) остатки распределены равномерно, сильные выбросы отсутствуют. Это говорит о том, что в исходных данных отсутствуют ошибки и выбросы (график Residual).

Для получения прогнозирующей модели ARIMA мы должны быть уверены в стационарности исходных данных, и, если она не наблюдается, применить необходимые преобразования.

Графически можно показать наличие тренда в данном ряде, применив технику сглаживания скользящим средним. Данный тренд является признаком его нестационарности.

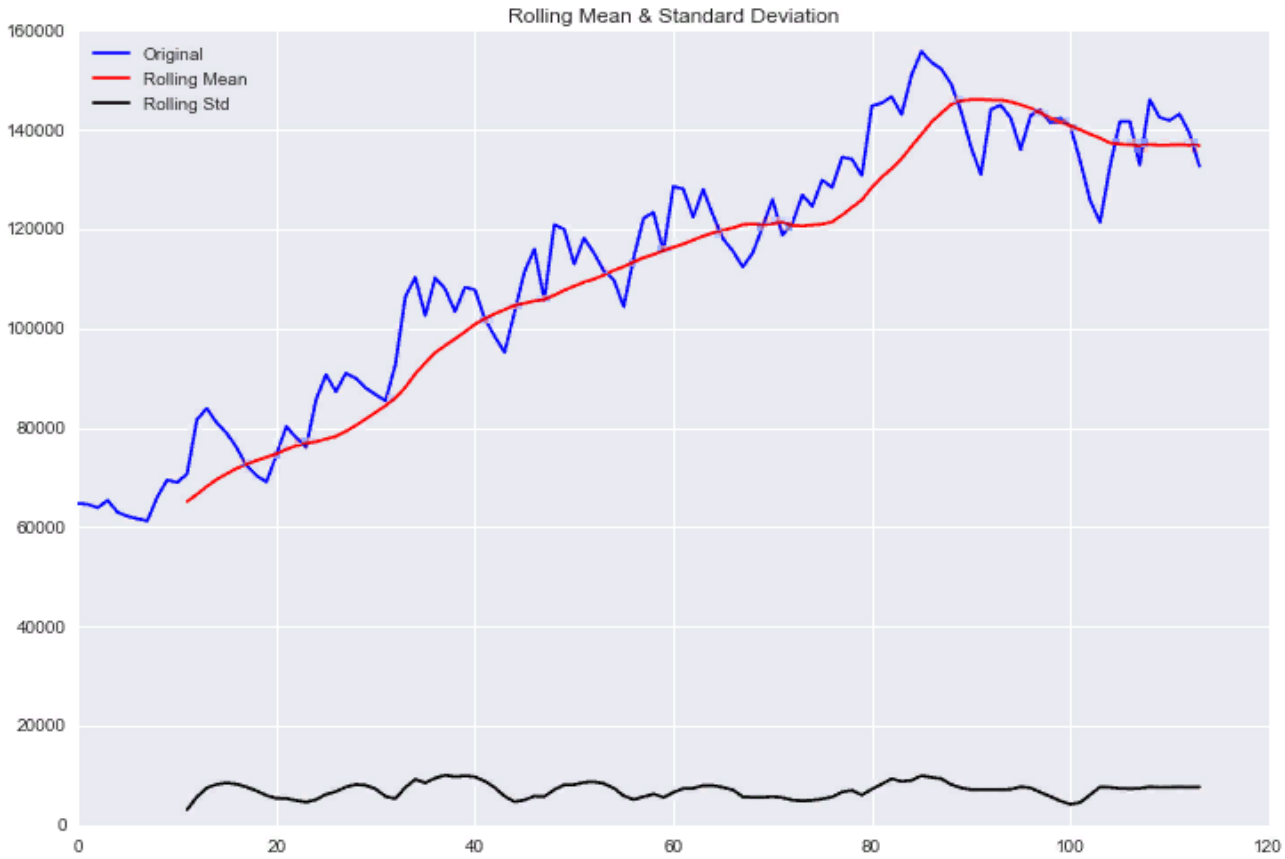


Рисунок 6 – График скользящего среднего и стандартного отклонения

Для проверки ряда на стационарность также можно воспользоваться тестом Дики-Фуллера. Тест Дики-Фуллера (DF-тест, Dickey – Fuller test) – методика, которая используется в прикладной статистике и эконометрике для анализа временных рядов для проверки на стационарность. Является одним из тестов на единичные корни (*Unit root test*). Был предложен в 1979 году Дэвидом Дики и Уэйном Фуллером.

```
Results of Dickey-Fuller Test:
Test Statistic                  -1.536597
p-value                          0.515336
#Lags Used                      12.000000
Number of Observations Used    101.000000
Critical Value (5%)             -2.890611
Critical Value (1%)             -3.496818
Critical Value (10%)            -2.582277
dtype: float64
```

Рисунок 7 – Результаты теста Дики-Фуллера

Как можно видеть из данной таблицы, p-значение значительно превышает пороговый показатель 0.05, отображающий 95%-ый интервал. Это значит, что данный ряд не стационарен.

Для того, чтобы приблизить наш ряд к стационарному, необходимо преобразовать данные. Существует множество способов преобразования данных, с помощью которых можно добиться стационарности исходного ряда. Среди них следует выделить:

- логарифмирование;
- взятие первой разности;
- сезонная разность;
- сезонная корректировка.

Первое, что нам необходимо сделать – применить процедуру взятия разности для наших данных. Это позволит исключить весь тренд из нашего ряда.

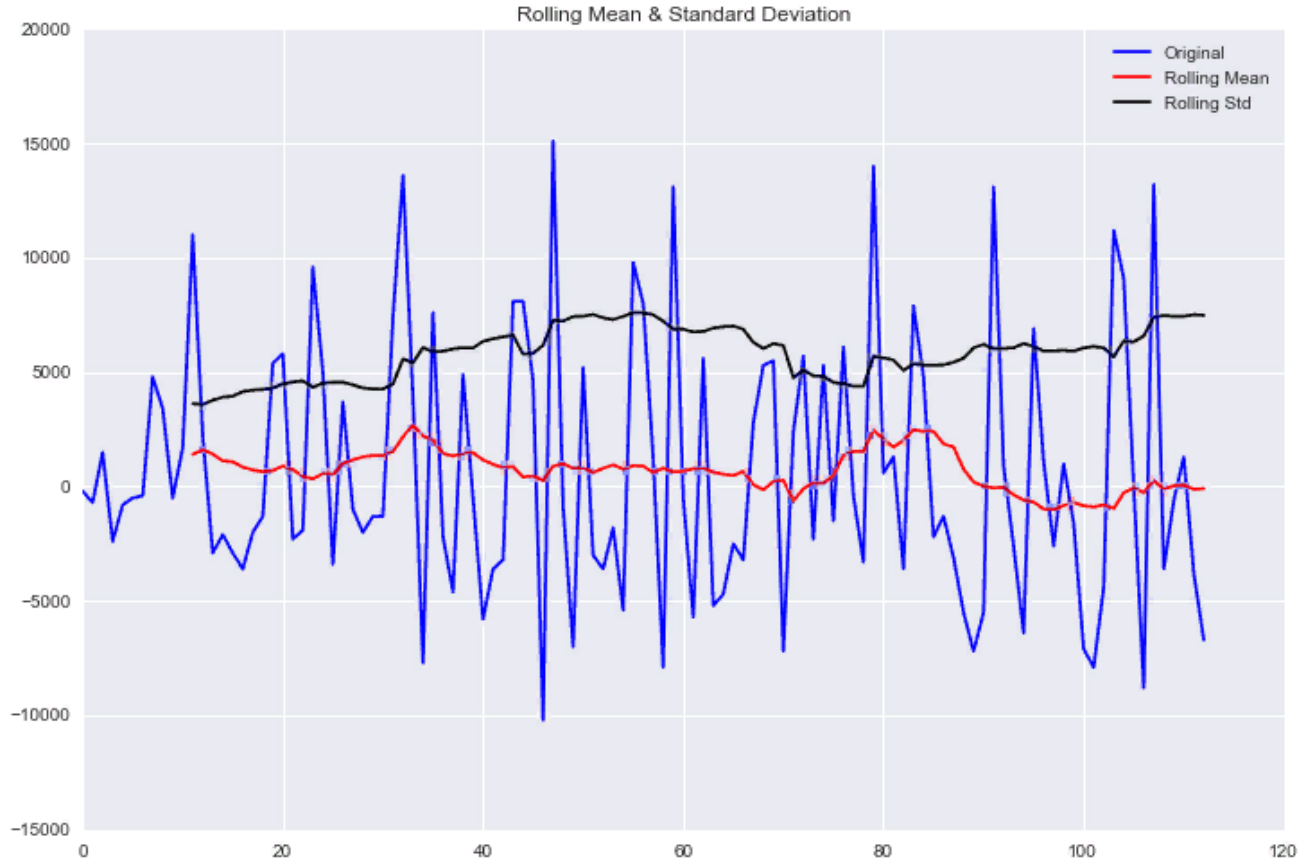


Рисунок 8 – График дифференцированного временного ряда

```
Results of Dickey-Fuller Test:
Test Statistic                  -1.938696
p-value                          0.314082
#Lags Used                      11.000000
Number of Observations Used    101.000000
Critical Value (5%)             -2.890611
Critical Value (1%)             -3.496818
Critical Value (10%)            -2.582277
dtype: float64
```

Рисунок 9 – Тест Дики-Фуллера для дифференцированного ряда

В результате преобразования исследуемый временной ряд приобретает признаки стационарности: тренд не выражен; ряд имеет постоянное среднее и колеблется с постоянной дисперсией. С целью получения лучших результатов применим сезонную разность. Взятие сезонной разности позволяет удалить сезонную компоненту из наших данных и посмотреть, как это повлияет на стационарность.

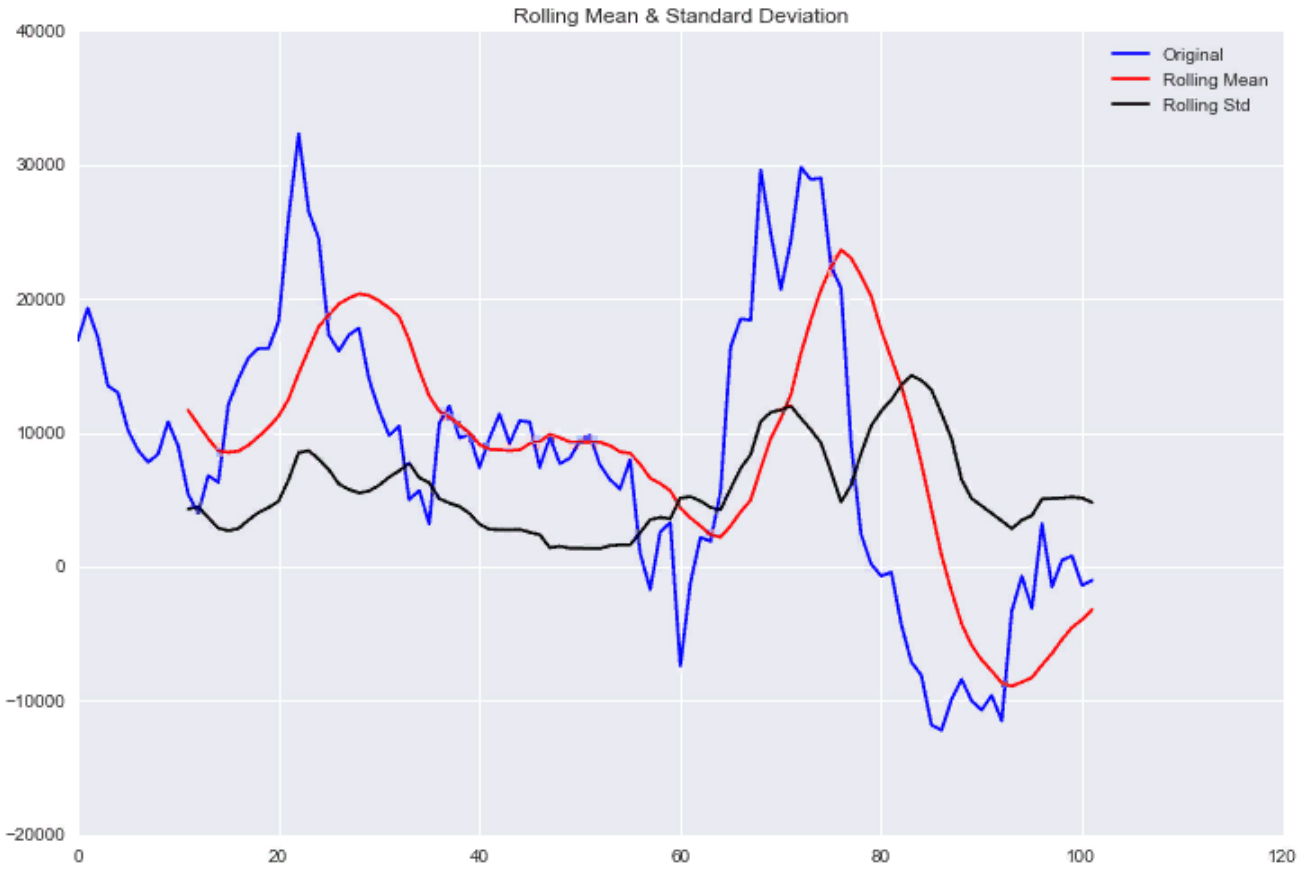


Рисунок 10 – График временного ряда после взятия сезонной разности

```
Results of Dickey-Fuller Test:
Test Statistic                  -2.469741
p-value                          0.123011
#Lags Used                       3.000000
Number of Observations Used     98.000000
Critical Value (5%)             -2.891516
Critical Value (1%)             -3.498910
Critical Value (10%)            -2.582760
dtype: float64
```

Рисунок 11 – Результаты теста Дики-Фуллера после взятия сезонной разности

Следующим шагом преобразования является взятие первой разности от сезонной разности.

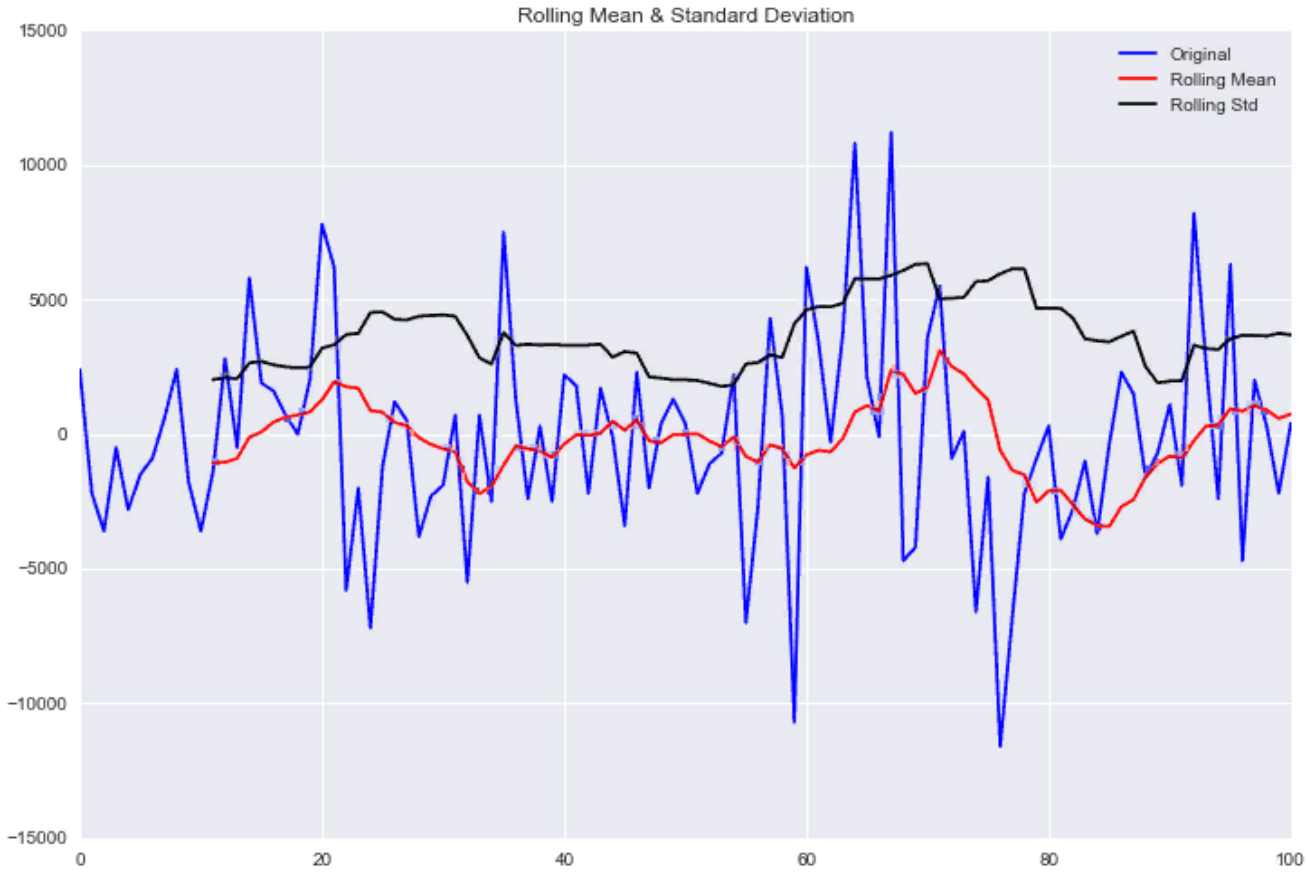


Рисунок 12 – График временного ряда после процедуры взятия второй разности

```
Results of Dickey-Fuller Test:
Test Statistic                  -9.258520e+00
p-value                          1.427874e-15
#Lags Used                       0.000000e+00
Number of Observations Used      1.000000e+02
Critical Value (5%)             -2.890906e+00
Critical Value (1%)             -3.497501e+00
Critical Value (10%)            -2.582435e+00
dtype: float64
```

Рисунок 13 – Результаты теста Дики-Фуллера после взятия второй разности

Высокое значение p-value дает нам возможность утверждать, что нулевая гипотеза о равенстве средних верна, следовательно, теперь наш ряд является стационарным.

Следующим шагом анализа является определение параметров настройки модели путем построения графиков автокорреляции (ACF) и частичной автокорреляции (PACF). Существует множество правил и рекомендаций по выбору подходящих правил для моделей AR, MA, SAR и MAR. Вид модели ARIMA определяется по коррелограмме.

Таблица 2 – отличия функций ACF и PACF для стационарных и нестационарных временных рядов

	Стационарный ряд	Нестационарный ряд
ACF	Независимо от значений в первом лаге, коррелограмма быстро убывает после несколько первых значений	Значения в первом лаге, т.е. ACF(1), близки к единице, а затем коррелограмма медленно убывает по угасающей экспоненте (синусоиде)
PACF	Также быстро убывает после нескольких первых значений	Значение в первом лаге ACF(1)=PACF(1) близко к единице, однако остальные значения коэффициентов корреляции статистически незначимы, т.е. значения функций на коррелограмме не выходят за пределы доверительного интервала

График функций ACF и PACF представлен на рисунке 14.

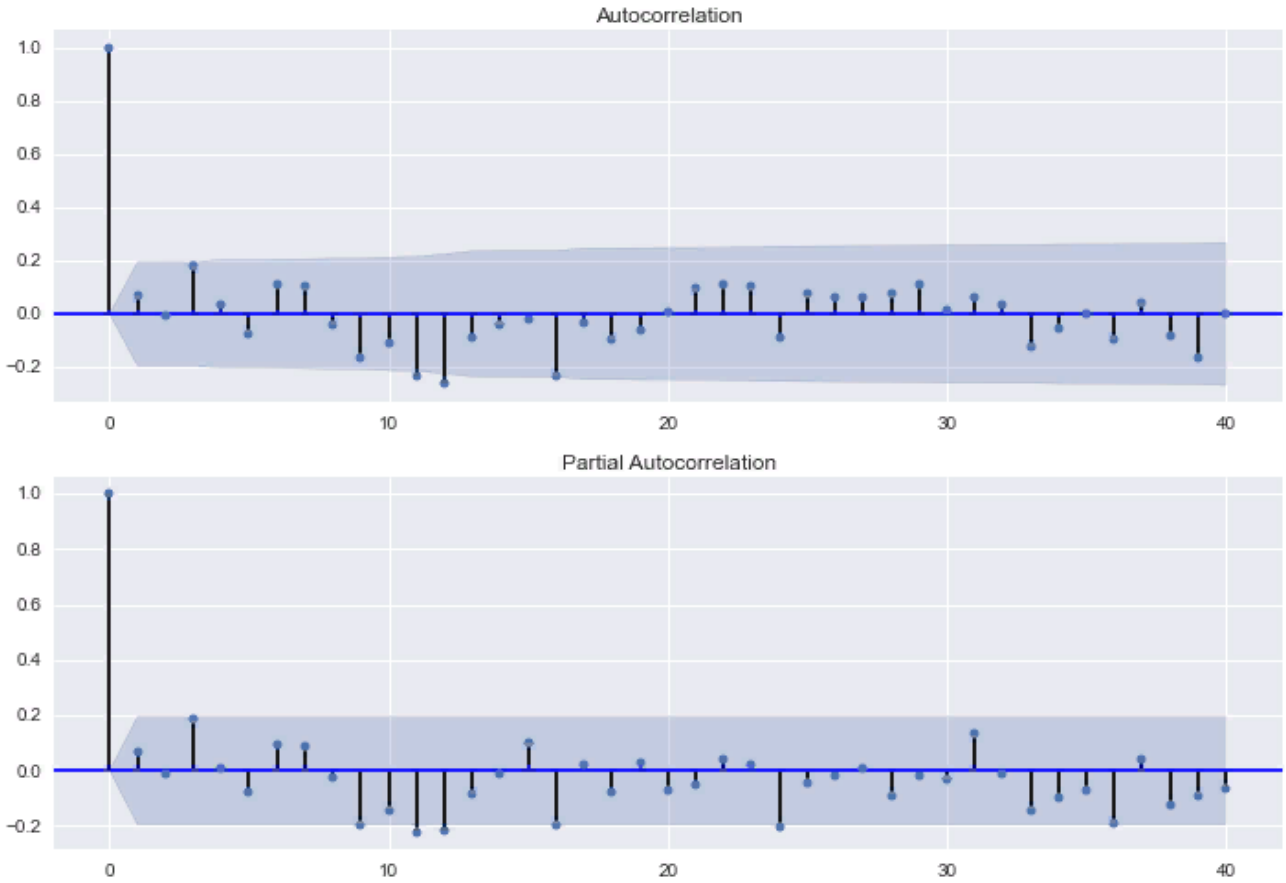


Рисунок 14 – Корреллограммы ACF и PACF

В данном случае затухающий график ACF и выбросы на первых нескольких лагах PACF свидетельствуют о том, что это модель авторегрессии. Также выброс на двенадцатом лаге в PACF показывает наличие 12-звенной сезонности, а значит, необходимо оценить модель SARIMA(p,d,q,)(Ps,Ds,Qs), где p – порядок авторегрессии, d – порядок разности, q – порядок скользящего среднего, Ps – порядок сезонной авторегрессии, Ds – порядок сезонной разности, Qs – сезонный параметр скользящего среднего.

Поскольку автокорреляция дифференцированного ряда отрицательна при лаге 12 (один год спустя), будет использован метод простого скользящего среднего. Пробуя разные методы, было обнаружено, что использование метода простого скользящего среднего повышает точность прогноза на 2018 год. Функция, используемая в дальнейшем, будет подбирать модели, используя все возможные комбинации параметров, эти модели будут применены для прогнозирования результата на несколько периодов вперед.

Теперь построим модель с использованием параметров SARIMAX (0,1,0)x(1,1,1,12):

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                riders   No. Observations:                  114
Model:             SARIMAX(0, 1, 0)x(0, 1, 1, 12)   Log Likelihood                -976.135
Date:                              Wed, 23 Mar 2016   AIC                           1956.271
Time:                                      13:16:18   BIC                           1961.743
Sample:                                  01-01-1973   HQIC                          1958.492
                                       - 06-01-1982
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.S.L12      -0.1377      0.050     -2.757      0.006      -0.236      -0.040
sigma2      1.424e+07   2.62e-10   5.44e+16      0.000    1.42e+07    1.42e+07
===================================================================================
Ljung-Box (Q):                       44.28   Jarque-Bera (JB):                 4.18
Prob(Q):                              0.30   Prob(JB):                         0.12
Heteroskedasticity (H):               1.44   Skew:                             0.20
Prob(H) (two-sided):                  0.29   Kurtosis:                         3.91
===================================================================================
```

Рисунок 15 – Параметры модели

Теперь, после анализа построенной модели выведем наш прогноз на график и сравним исходные значения временного ряда с прогнозными. Для начала используем модель для существующих данных, это поможет нам понять, насколько точно разработанная модель может предсказывать значения временного ряда.

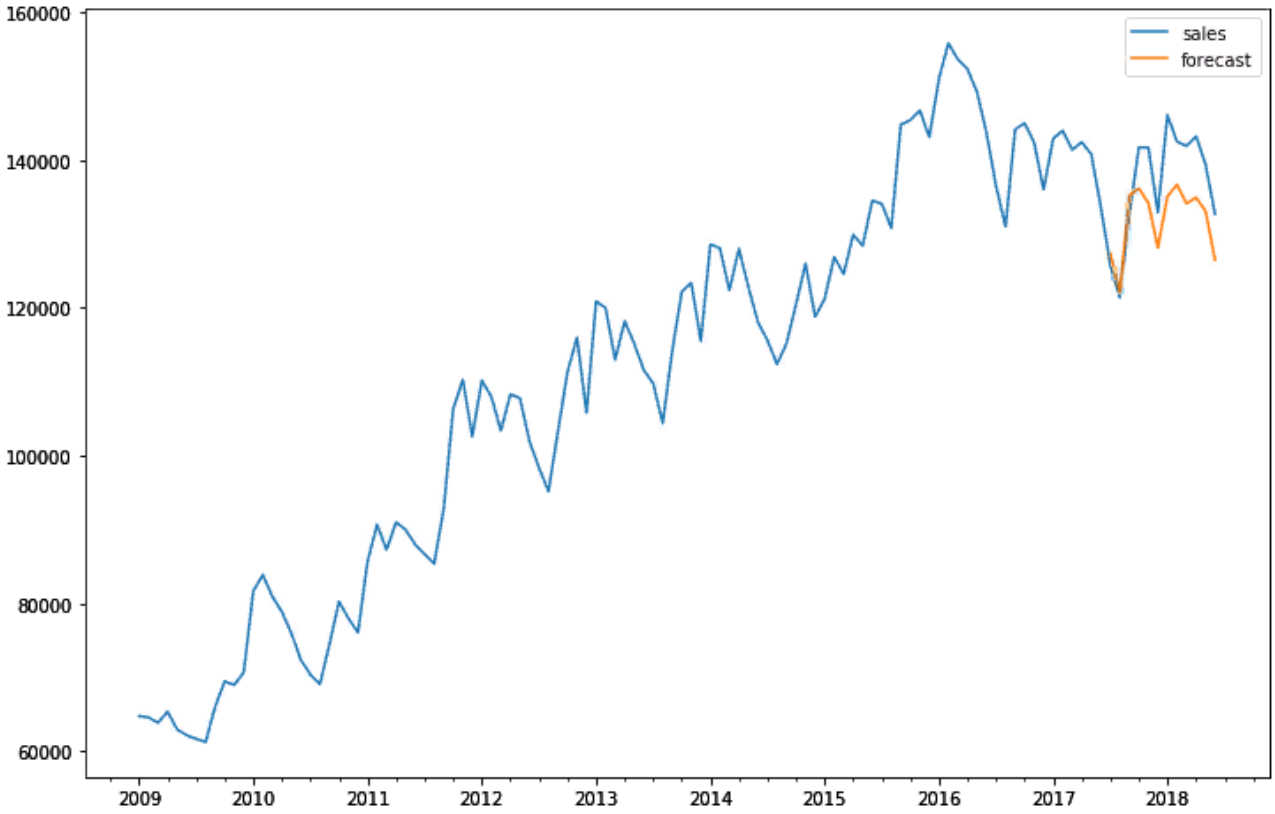


Рисунок 16 — Построение прогнозной модели на исходных данных

Построим более детальный график представления исходных и прогнозных значений:

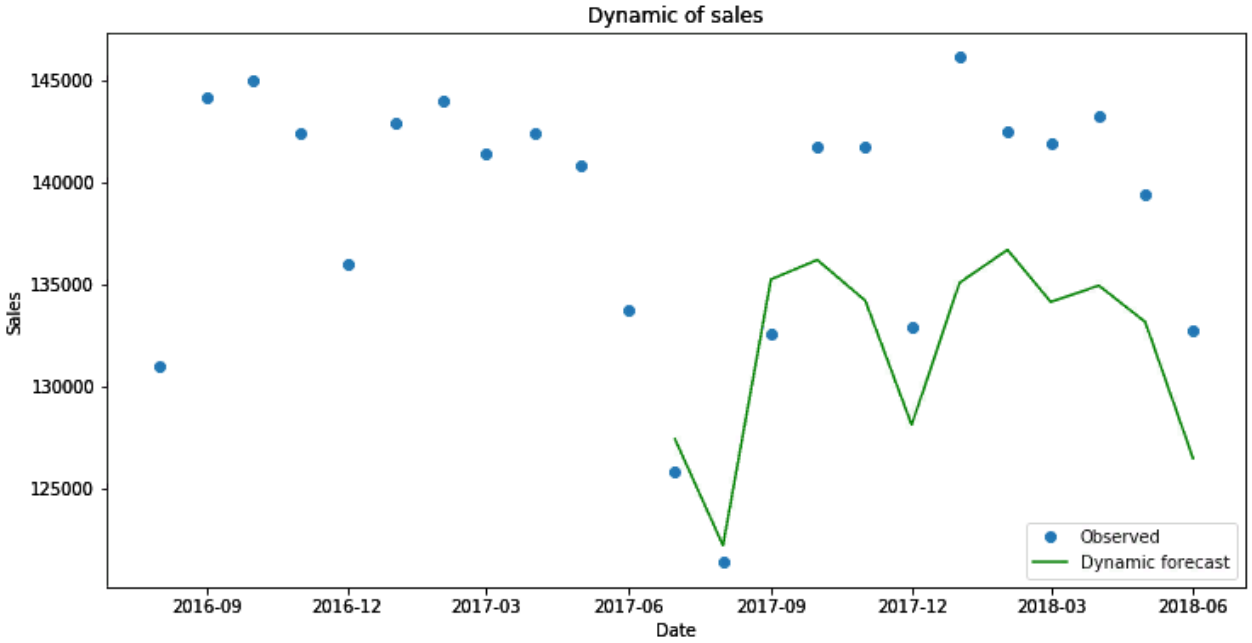


Рисунок 17 - Сравнение прогнозных значений с фактическими

Проанализировав графики, можно сделать вывод, что построенная модель довольно точно предсказывает значения, тренд сохранен, учитывается сезонная компонента.

Теперь с помощью функции predict построим прогнозные значения, добавив в модель анализа новые значения продаж с 01.07.2018 г.

Прогноз представлен на рисунке 18.

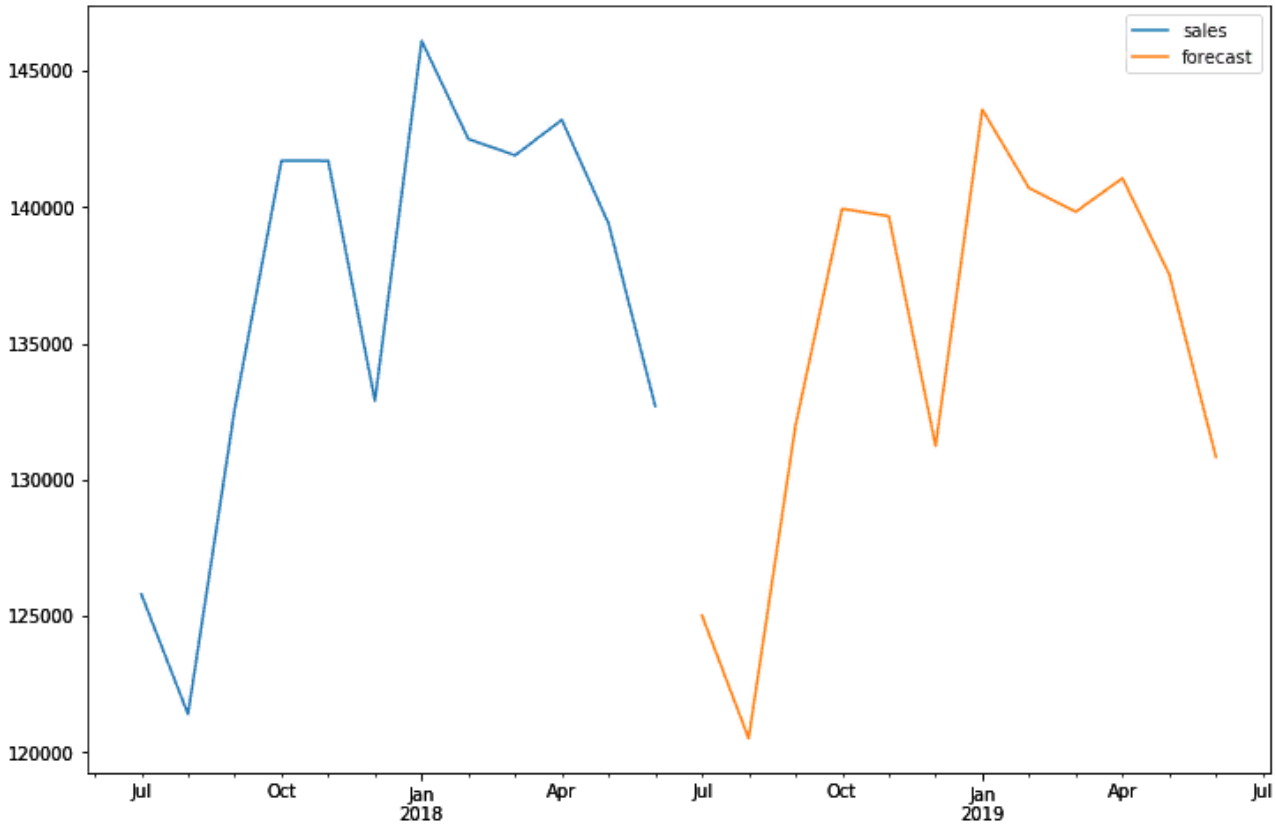


Рисунок 18 — Прогнозные значения объёма продаж

В ходе проведения оценки системы прогнозирования на данных об объеме продаж было выявлено:

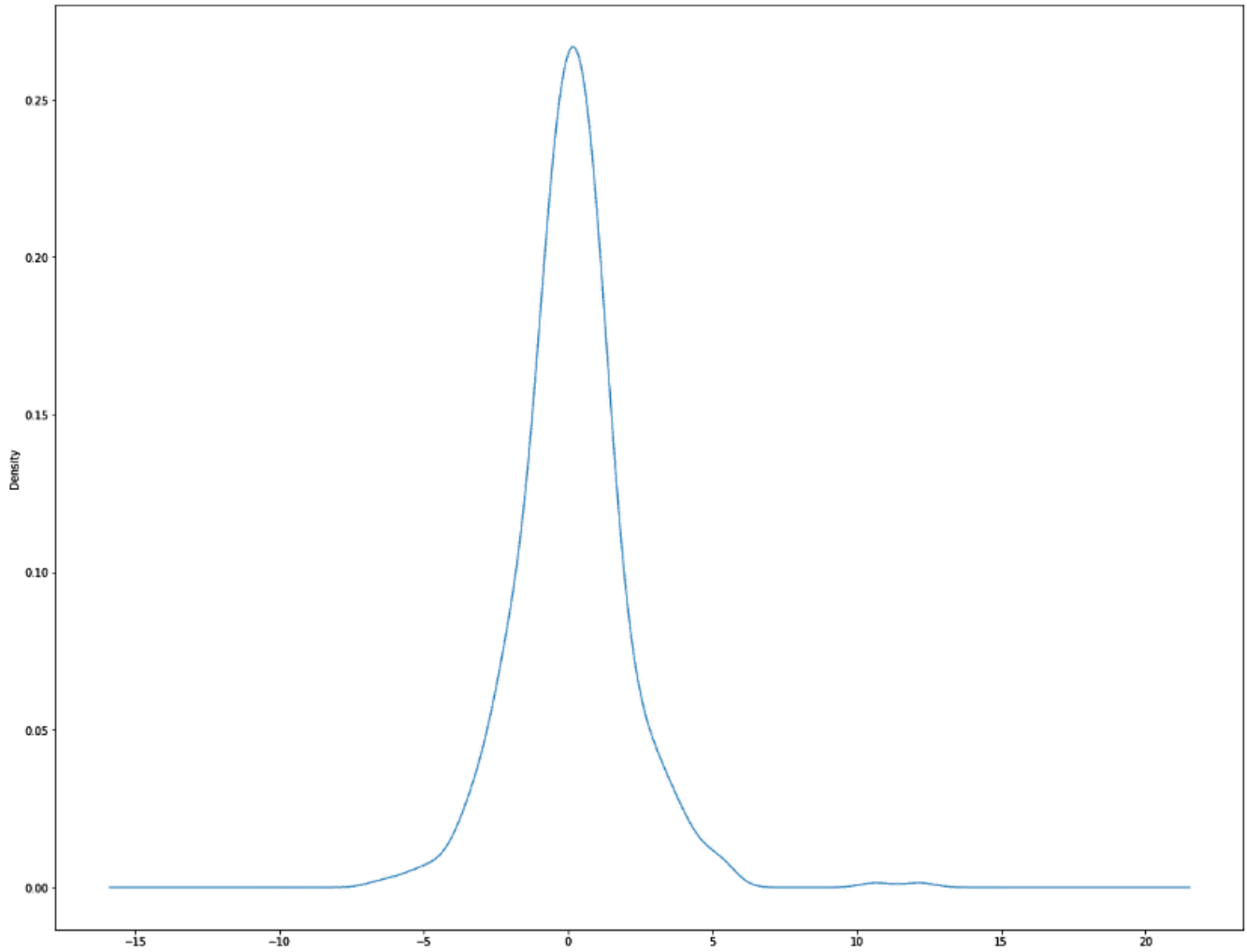
- Прогнозные значения в целом попадают в тренд, циклическую и сезонные компоненты фактического ряда. Направление прогноза угадывается довольно точно;
- Довольно точно прослеживается сезонность: в зимний период объем продаж существенно ниже, чем весной и летом, это обуславливается типом реализуемой продукции;
- Существуют определенные выбросы, связанные в основном с оптовыми и

мелкооптовыми закупками в розничном магазине. Такие случаи должны быть исключены в дальнейшей оценке прогнозной системы.

Для того, чтобы оценить адекватность построенной модели, проведем анализ остатков нашей модели, принятый проводить для моделей ARIMA.

На рисунке 19 представлен график распределения остатков.

Рисунок 19 – график распределения остатков



На графике распределения остатков можно заметить, что они достаточно хорошо ложатся на прямую, имеют характер нормального распределения. На основании этих выводов можно сказать, что построенная нами модель достаточно адекватно строит прогнозные значения и ее можно использовать на других данных.

Выводы по главе 2

В результате исследования были решены следующие задачи:

- Проведен анализ инструментов разработки;
- Были подготовлены данные для построения;
- Визуализирован исходный временной ряд;
- Произведено дифференцирование временного ряда с целью приведения его к стационарному виду;
- Выбраны параметры модели ARIMA для построения прогноза;
- Построен прогноз продаж;
- Произведена оценка результатов прогнозирования.

Также можно сделать вывод, что при имеющейся выборке, а также при получении сторонних данных мы можем успешно прогнозировать объем продаж предприятия розничной торговли с помощью модели авторегрессии ARIMA, а именно с помощью построенной нами модели.

3 КОММЕРЦИАЛИЗАЦИЯ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

В рамках диссертационного исследования будет разработан проект интернет-сервиса «Прогнозирование объема продаж», направленный на создание малого аналитического предприятия для анализа данных, планирования прибыли, оценки финансового состояния предприятия, составления прогноза продаж. Разрабатываемый проект позволит предприятиям малого, среднего и крупного бизнеса осуществлять анализ своей деятельности, получать рекомендации по развитию, принимать оптимальные решения на основе полученного прогноза. В качестве клиентов в основном могут выступать коммерческие предприятия. Сервис позволит проводить анализ и давать адекватную оценку развитию различных предприятий.

В рамках коммерциализации проекта главной задачей является привлечение инвестора с целью финансирования. Перед тем как определить метод, по которому будет наиболее выгодно продвигаться с целью получения наибольшей прибыли, необходимо определиться по какому из трех известных направлений мы можем начать действовать. На рисунке 18 представлены основные способы коммерциализации.

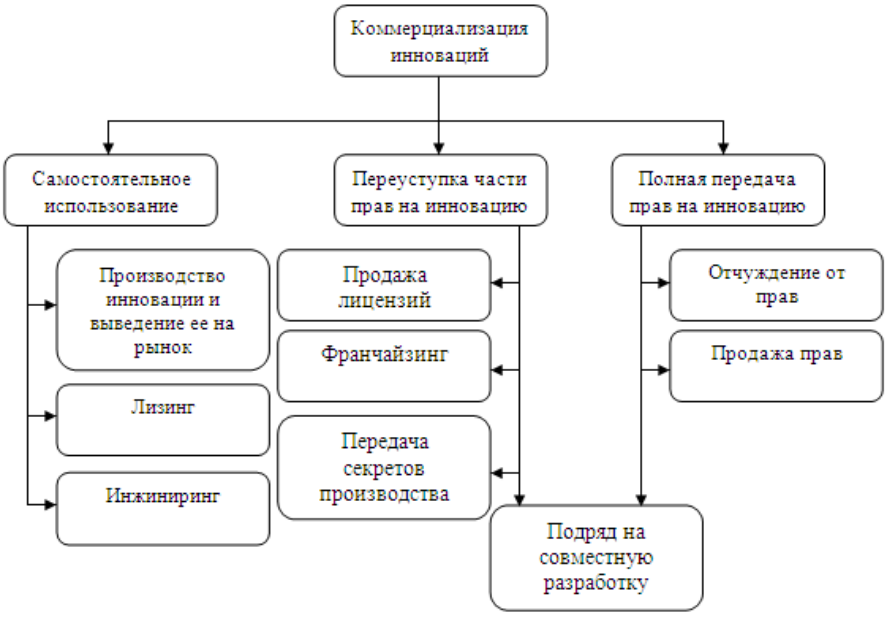


Рисунок 18 - Способы коммерциализации

Так как аналогичных проектов нет, а данная тема весьма актуальна и значима, можно говорить о том, что сервис будет иметь успех. На данный момент возможно найти только некоторые статьи, в которых предоставляется анализ предприятий или рейтинг оценки предприятий, или научно исследовательские работы в данной сфере. Так же, проект еще не имеет рыночной значимости, в связи с этим коммерциализация проекта начинается с самостоятельного использования. Недостатками данного метода являются высокие затраты на первоначальном этапе и долгий срок окупаемости, но при этом, в случае четкого обоснования на рынке, возможны очень высокие доходы.

Предположительно, стоимость одного проектного анализа будет от 25000 рублей в зависимости уровня сложности, охвата и объема данных. Также, будут открыты демо-доступы к готовым исследованиям, часть контента будет платной.

Оценка трудоемкости выполненных работ является важным этапом определения стоимости разработки. Трудоемкость – показатель, характеризующий затраты рабочего времени на выполнение определенного этапа, задачи или операции. Трудоемкость является обратным показателем производительности труда и определяет эффективность использования одного из главных производственных ресурсов – рабочей силы. Величины трудоемкости определяется исходя из следующих факторов: технический уровень производства, квалификация работников, условия труда, сложность выполняемых задач и др.

Расчет трудоемкости ведется на основе плана работ, который отражает этапы проектирования. В таблице 3 приведен перечень работ участников: руководителя и исполнителя с указанием их трудоемкости.

Таблица 3 – Трудоемкость исполнителей

№	Наименование	Трудоемкость (чел./дни)	
		Руководитель	Исполнитель
1	Разработка технического задания	3	3
2	Отбор и изучение литературы	3	21
3	Обработка и анализ информации	-	14
4	Разработка плана решения задачи	2	4
5	Выбор инструментов разработки и реализация алгоритма решения задачи	4	35
6	Анализ результатов исследования	2	7
7	Оформление документации	3	7
8	Проведение предзащиты	1	1
9	Сдача проекта	1	1
	Итого:	19	93

На основе оценки трудоемкости рассчитываются издержки на оплату труда ее исполнителей, являющиеся одной из основных статей калькуляции себестоимости разработки.

Создание малого предприятия по оказанию аналитических услуг предприятиям малого и среднего бизнеса – цель коммерциализации проекта в рамках диссертационного исследования. Безусловно, без первоначальных инвестиций не обойтись. Начальным этапом в предстоящей работе становится поиск инвестора, а также участие в конференциях, грантах, программах по поддержанию деятельности молодых ученых в сфере инноваций. Для первоначальной реализации проекта потребуются затраты на разработку и последующее поддержание Интернет-ресурса, а также маркетинговые исследования. Расходы, связанные с организацией Интернет-ресурса включают:

стоимость доступа к сети Интернет, аренда дискового пространства на сервере и стоимость регистрации доменного имени. Кроме этого, необходимо учитывать затраты на аренду помещения, техническое оснащение рабочих мест, а также заработную плату сотрудников.

В таблице 4 представлены затраты в первый год внедрения проекта с указанием стоимости.

Таблица 4 – Затраты на внедрение проекта в первый год

Затраты	**Период**	**Стоимость, руб.**
Регистрация доменного имени	12 месяцев	3 500,00
Аренда хостинга	12 месяцев	7 000,00
Доступ к сети Интернет	12 месяцев	6 000,00
Маркетинговая кампания (баннерная реклама, коммерческие предложения, деловые встречи)	12 месяцев	150 000,00
Аренда помещения	12 месяцев	30 000,00
Техническое оснащение (рабочее место, ПК, ПО)	12 месяцев	73 000,00
Заработная плата сотрудникам	6 месяцев	200 000,00
Итого:		469 500,00

Суммарные затраты на разработку и внедрение проекта в первый год работы составят 469 500, 00 рублей. Планируется вывести проект на рынок в 2019 году, привлечь не менее 10 клиентов.

3.1 Конкурирующие решения

На рынке имеется достаточное количество онлайн-сервисов и программного обеспечения, предоставляющих услуги прогнозирования объема продаж.

Опишем наиболее близкие аналоги нашего решения.

Сервис SoftServe Business Systems (https://www.softservebs.com/solutions/prognozirovanie) фокусируется на использовании прогнозных методик любой сложности. Онлайн-сервис позволяет:

- Составлять качественный прогноз продаж для производства, что позволит сократить объемы просроченной продукции, минимизировать затраты, связанные с производством и закупкой сырья и материалов;
- Планировать процесс поставок, обеспечивая тем самым нужный объем продукции в точках продаж и сокращая out-of stock;
- Прогнозировать запуск новинок в точках продаж;
- Разрабатывать сценарии реализации промо активностей в сетях;
- Формировать краткосрочные прогнозы продаж для принятия операционных решений;
- Принимать правильные и своевременные решения, используя действующие сценарии развития тех или иных бизнес процессов.

Облачный сервис прогнозирования спроса KORUS Forecast (https://forecast.korusconsulting.ru) позволяет производить расчет и предоставляет прогноз спроса по всем товарам и всем магазинам сети на индивидуально заданной периодической основе. Сервис позволяет отслеживать KPI, а также получать детальную информацию по рассчитанному прогнозу, выявлять причины ошибок. Все отчеты располагаются в облаке с доступом через web-браузер и мобильное приложение.

Среди программных продуктов, позволяющих проводить анализ и прогнозирование продаж, стоит выделить пакет Statistica от компании Statsoft,

Deductor компании BaseGroup. Рассмотрим основные возможности данных программ.

STATISTICA - это система для статистического анализа данных, включающая широкий набор аналитических процедур и методов: более 100 различных типов графиков, описательные и внутригрупповые статистики, разведочный анализ данных, корреляции, быстрые основные статистики и блоковые статистики, интерактивный вероятностный калькулятор, Т-критерии (и другие критерии групповых различий), таблицы частот, сопряженности, флагов и заголовков, анализ многомерных откликов, множественная регрессия, непараметрические статистики, общая модель дисперсионного и ковариационного анализа, подгонка распределений, добыча данных, нейронные сети и многое другое.

Продукты серии STATISTICA основаны на самых современных технологиях, полностью соответствуют последним достижениям в области IT, позволяют решать любые задачи в области анализа и обработки данных, идеально подходят для решения практических задач в маркетинге, финансах, страховании, экономике, бизнесе, промышленности, медицине и т.д.

Deductor является аналитической платформой – основой для создания законченных прикладных решений в области анализа данных. Реализованные в Deductor технологии позволяют на базе единой архитектуры пройти все этапы построения аналитической системы: от создания хранилища данных до автоматического подбора моделей и визуализации полученных результатов.

Возможности платформы:

- Интеграция данных;
- Визуализация;
- Очистка данных;
- Data Mining;
- Сценарный подход.

3.2 Web-сайт

На начальной стадии проекта нами создан сайт, знакомящий клиентов с компанией, а также позволяющий отправлять заявку и данные через форму на сайте. В дальнейшем для увеличения функциональности потребуется его реконструкция.

3.2.1 Структура сайта

Сайт состоит из следующих разделов:

1) Главная страница;
2) О нас:
    a) О компании;
    b) Услуги;
    c) Наши преимущества;
    d) Наши клиенты;
3) Новости;
4) Наши методы прогнозирования;
5) Оценка финансового состояния:
    a) Значение финансового состояния;
    b) Форма заказа.

3.2.2 Основные функциональные страницы

Основной функциональной страницей сайта является страница «Заказать прогноз продаж», содержащая форму заказа.

Главная страница сайта представлена на рисунке 20.

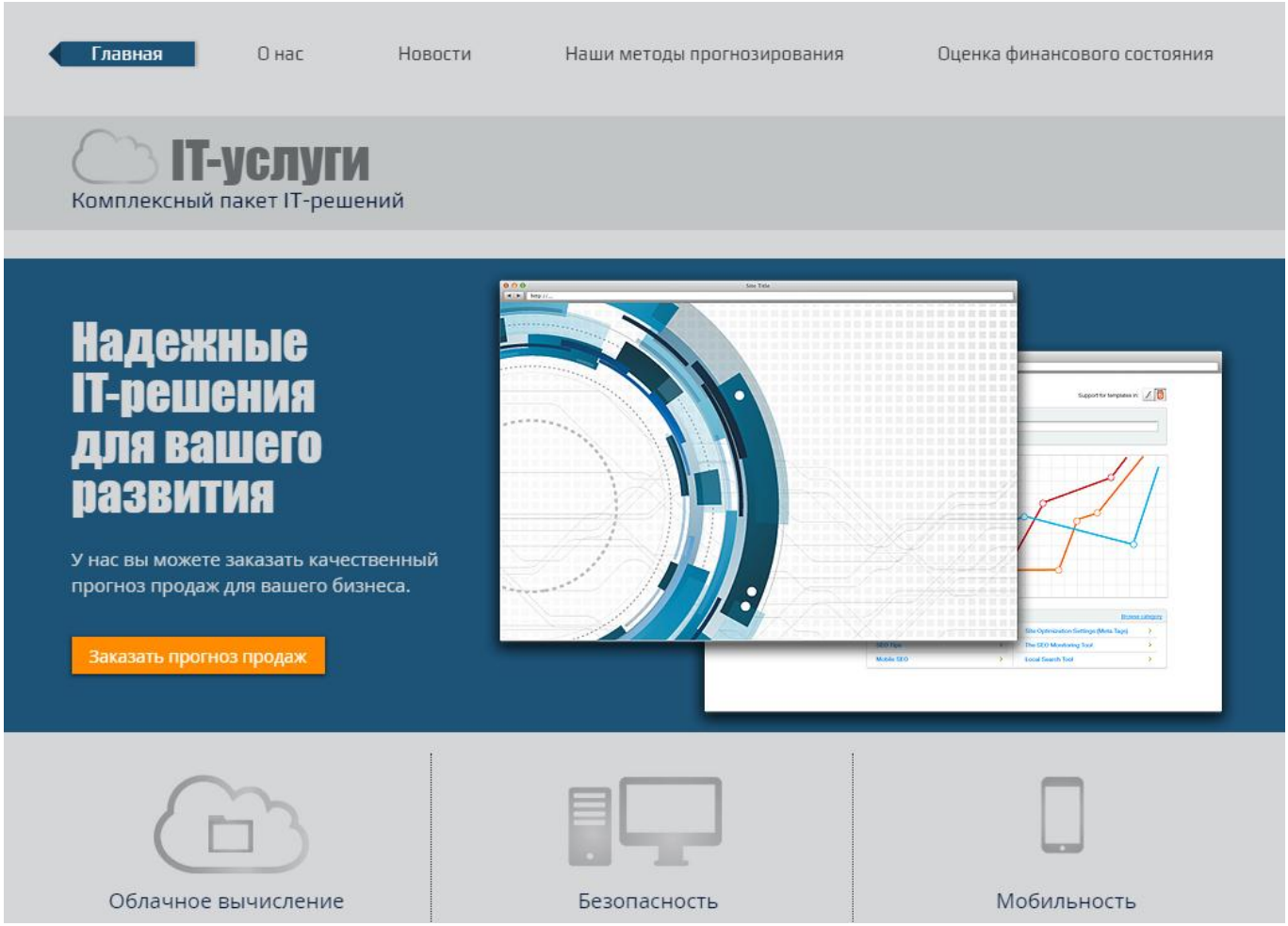


Рисунок 20 – Главная страница сайта

На странице «Форма регистрации» пользователю будут доступны следующие

поля для заполнения:

- Имя;
- Фамилия;
- Отчество;
- E-mail;
- Телефон;
- Вопрос «Вы физическое или юридическое лицо?»;
- Наименование организации (объекта исследования);
- Выбор вида отчёта (из двух предложенных вариантов);
- Поле для прикрепления документа с данными для анализа;

- Комментарий.

Поля «Имя», Фамилия», «E-mail», «Вопрос «Вы физическое или юридическое лицо?», «Наименование организации (объекта исследования)», «Поле для прикрепления документа с данными для анализа» являются обязательными для заполнения.

При нажатии на кнопку «Отправить» на нашу почту будет отправлена вся введённая клиентом информация, а клиент увидит во всплывающем окне сообщение со следующим текстом: «Спасибо за заказ. Ваш анализ будет готов в течение пяти рабочих дней и отправлен на почту, указанную при регистрации. При возникновении дополнительных вопросов относительно присланных данных мы свяжемся с Вами по указанным Вами контактным данным – пожалуйста, будьте на связи».

Выводы по главе 3

В данном разделе нами был осуществлён выбор основных направлений коммерциализации и рассмотрены аналоги нашего решения, в том числе сервисы SoftServe Business Systems, KORUS Forecast, программные продукты Statistica и Deductor.

Также был разработан сайт, позволяющий клиенту получить информацию о наших услугах и отправить онлайн-заявку на заказ прогноза.

ЗАКЛЮЧЕНИЕ

Данное исследование было направлено на построение модели прогнозирования объема продаж в ООО НПФ «РД Технология».

В ходе работы были рассмотрены теоретические основы анализа и прогнозирования временных рядов и проанализированы их преимущества и недостатки.

Представлен проект интернет-сервиса для коммерциализации проекта, который позволит проводить анализ и давать адекватную оценку развитию различных социально-экономических систем. На основе анализа данных, полученных в ходе исследования, представлены приблизительная прибыль и затраты. Таким образом, цель работы достигнута, задачи – решены.

Результатом данной работы послужило решение ряда задач:

1. Сформулирована актуальная в условиях рыночной экономики трактовка понятия «Прогнозирование»;
2. Выявлены факторы, оказывающие наибольшее влияние на уровень продаж предприятия розничной торговли;
3. Проанализированы основные методы и модели прогнозирования данных;
4. Разработана модель прогнозирования объема продаж на основе модели ARIMA, и реализованная на языке программирования Python;
5. Описаны основные направления коммерциализации.

Результаты проделанной работы предлагается использовать предприятиям малого, среднего и крупного бизнеса и позволят осуществлять анализ своей деятельности, получать рекомендации по развитию, принимать оптимальные решения на основе полученного прогноза.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Федеральный закон "О развитии малого и среднего предпринимательства в Российской Федерации" от 24.07.2007 N 209-ФЗ (последняя редакция) [Электронный ресурс]. Режим доступа: http://www.consultant.ru (дата обращения: 01.12.2018)

2. "Конституция Российской Федерации" (принята всенародным голосованием 12.12.1993) (с учетом поправок, внесенных Законами РФ о поправках к Конституции РФ от 30.12.2008 N 6-ФКЗ, от 30.12.2008 N 7-ФКЗ, от 05.02.2014 N 2-ФКЗ, от 21.07.2014 N 11-ФКЗ) [Электронный ресурс]. Режим доступа: http://www.consultant.ru (дата обращения: 01.12.2018)

3. George Athanasopoulos, Rob J Hyndman, Haiyan Song, Doris Wu. The tourism forecasting competition. – International Journal of Forecasting 27(3), 2011

4. Rob J. Hyndman, Yeasmin Khandakar. Automatic Time Series Forecasting: The forecast Package for R. – Journal of Statistical Software. July 2008, Volume 27, Issue 3.

5. George Athanasopoulos, Rob J Hyndman, Haiyan Song, Doris Wu. The tourism forecasting competition. – International Journal of Forecasting 27(3), 2011

6. Li Q., Racine J. S. Nonparametric Econometrics: Theory and Practice. Princeton University Press, 2007 – 746 p.

7. Hayfield T., Racine J. S. Nonparametric econometrics: The np package. Journal of statistical software 27 (5), 1-32 P.

8. Айвазян, С. А. Методы эконометрики: учебник / С.А. Айвазян. – М.: Магистр: ИНФРА-М, 2010. – 512 с.

9. Айвазян С.А., Фантаццини Д. Эконометрика-2. Продвинутый курс с приложениями в финансах. – М.: Магистр: Инфра-М, 2014. – 944 с.

10. Борзых Д. А., Демешев Б. Б. Эконометрика в задачах и упражнениях. – М. : УРСС, 2017. – 304 с.

11. Светуньков, И. С. Методы социально-экономического прогнозирования. В 2 т. Т. 2. Модели и методы : учебник и практикум для академического бакалавриата / И. С. Светуньков, С. Г. Светуньков. – М. :

Издательство Юрайт, 2016. – 447 с.

12. Светуньков, И. С. Методы и модели социально-экономического прогнозирования : учебник и практикум для академического бакалавриата. В 2-х т. Т. 1. Теория и методология прогнозирования / И. С. Светуньков, С. Г. Светуньков. – М. : Издательство Юрайт, 2014. – 351 с.

13. Svetunkov I., Kourentzes N. (February 2015). Complex exponential smoothing. Working Paper of Department of Management Science, Lancaster University 2015:1, 1-31.

14. Орлов, А.И. Эконометрика : учебник для вузов / А.И. Орлов. – Ростов н/Д : Феникс, 2009. – 277 с.

15. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: Учебник. – М.: Финансы и статистика, 2001. – 228 с.

16. Мишулина, О. А. Статистический анализ и обработка временных рядов : учеб. пособие для студентов вузов / О. А. Мишулина ; М-во образования Рос. Федерации, М-во Рос. Федерации по атом. энергии, Моск. инж.-физ. ин-т (гос. ун-т), Экон.-аналит. ин-т, Каф. экон. динамики. – М. : Моск. инж.-физ. ин-т (гос. ун-т), 2004. — 178 с.

17. Писарева, О. М. Методы социально-экономического прогнозирования: Учебник ГУУ – НФПК. – М.: Высшая школа, 2003. – 395 с.

18. Садовникова Н.А., Шмойлова Р.А. Анализ временных рядов и прогнозирование. Вып. 3: Учебно-методический комлекс. – М.: Изд. центр ЕАОИ, 2009. – 264 с.

19. Дуброва Т.А. Статистические методы прогнозирования. – М.: ЮНИТИДАНА, 2003. – 205 с.

20. Семёнычев В. К., Семёнычев Е. В. Параметрическая идентификация рядов динамики: структуры, модели, эволюция: монография. – Самара: Изд-во «СамНЦ РАН», 2011. – 364 с.

21. Трофимов П. Ю. Прогнозирование временных рядов методом ARIMA / П. Ю. Трофимов, В. Ю. Носков // Теплотехника и информатика в образовании, науке и производстве: сборник докладов VI Всероссийской научно-

практической конференции студентов, аспирантов и молодых учёных (ТИМ’2017) с международным участием (Екатеринбург, 11–12 мая 2017 г.). – Екатеринбург: УрФУ, 2017. – с. 260-262.

22. Пилюгина А.В. Использование моделей ARIMA для прогнозирования валютного курса / А.В. Пилюгина, А.А. Бойко // Прикаспийский журнал: управление и высокие технологии №4 (32) - Издательство: Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования "Астраханский государственный университет" (Астрахань) – Астрахань: АГУ, 2015. – с. 249-267.

23. Гребенников А.В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA / А.В. Гребенников, Ю.А. Крюков, Д.В. Чернягин // Электронный журнал «Системный анализ в науке и образовании» №1 – Издательство: Университет "Дубна" (Дубна), 2011. – с. 7-17

24. Кондратьев, Н. Д. План и предвидение (к вопросу о методах составления перспективных планов развития народного хозяйства и сельского хозяйства в частности) / Н. Д. Кондратьев // Вопросы экономики. – 1992. – N.4-6. –С.91-110.

25. Янч Э. Прогнозирование научно-технического прогресса / Издание второе, дополненное. – М.: Прогресс, 1974. – 592 с.

26. Гранберг А.Г. Динамические модели народного хозяйства — Учебное пособие. М.: «Экономика», 1985. – 240 с.

27. Цыгичко В. Н. Прогнозирование социально-экономических процессов. М.: КомКнига. 2007. – 240 с.

28. Антохонова И.В. Методы прогнозирования социально-экономических процессов. Учебное пособие. – Улан-Удэ: Изд-во ВСГТУ, 2005. – 212 с.

29. Бутакова М.М. Экономическое прогнозирование: методы и приемы практических расчетов - Учебное пособие. – 2-е изд., испр. – М.: КНОРУС, 2010. – 168 с.

30. Леньков Р.В. Социальное прогнозирование и проектирование - учебное пособие / Р.В. Леньков. – М.: ЦСП и М, 2013. – 192 с.

31. Минько А.Э., Минько Э.В. Методы прогнозирования и исследования операций - М.: Финансы и статистика, 2010. – 480 с.

32. Уэс Маккинли: Python и анализ данных/ Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 482 с.: ил.

33. Фёрстер Э., Ренц Б. Методы корреляционного и регрессионного анализа: руководство для экономистов. М.: «Финансы и статистика», 1983. – 304 с.

34. Чучуева, И.А. Модель прогнозирования временных рядов по выборке максимального подобия : автореферат дис. кандидата технических наук : 05.13.18 / Чучуева Ирина Александровна; [Место защиты: Моск. гос. техн. ун-т им. Н.Э. Баумана]. – Москва, 2012. – 16 с.

35. Kiser, M. Deploying R Models in Production. URL: https://blog.algorithmia.com/deploying-r-models-production-web-services/

36. Суслов В.И., Ибрагимов Н.М. Эконометрия. – Новосибирск: Изд-во СО РАН, 2005. – 744 с.

37. Сухарев, М. Г. Методы прогнозирования. Учебное пособие – М.: РГУ нефти и газа, 2009 г. – 208 с.

ПРИЛОЖЕНИЕ А

МОДУЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ С ПРИМЕНЕНИЕМ МОДЕЛИ ARIMA НА ЯЗЫКЕ PYTHON

```
#загрузка необходимых для моделирования библиотек
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import datetime
from dateutil.relativedelta import relativedelta
import seaborn as sns
import statsmodels.api as sm
from statsmodels.tsa.stattools import acf
from statsmodels.tsa.stattools import pacf
from statsmodels.tsa.seasonal import seasonal_decompose

#подключение файла с исходными данными для анализа
df = pd.read_csv('sales-average-monthly-.csv', index_col=0)
df.index.name=None
df.reset_index(inplace=True)
df.drop(df.index[114], inplace=True)

#вывод графика временного ряда
start = datetime.datetime.strptime("2009-01-01", "%Y-%m-%d")
date_list = [start + relativedelta(months=x) for x in range(0,114)]
df['index'] =date_list
df.set_index(['index'], inplace=True)
df.index.name=None
```

```
df.columns= ['sales']
df['sales'] = df.sales.apply(lambda x: int(x)*100)
df.sales.plot(figsize=(12,8), title= 'Monthly Sales', fontsize=14)
plt.savefig('month_sales.png', bbox_inches='tight')

#вывод сезонной декомпозиции ряда
decomposition = seasonal_decompose(df.sales, freq=12)
fig = plt.figure()
fig = decomposition.plot()
fig.set_size_inches(15, 8)

#проведение теста Дики-Фуллера
from statsmodels.tsa.stattools import adfuller
def test_stationarity(timeseries):
    #Determing rolling statistics
   rolmean = df.rolling(12).mean()
   rolstd = df.rolling(12).mean()
   #Plot rolling statistics:
   fig = plt.figure(figsize=(12, 8))
   orig = plt.plot(timeseries, color='blue',label='Original')
   mean = plt.plot(rolmean, color='red', label='Rolling Mean')
   std = plt.plot(rolstd, color='black', label = 'Rolling Std')
   plt.legend(loc='best')
   plt.title('Rolling Mean & Standard Deviation')
   plt.show()
   #Perform Dickey-Fuller test:
   print ('Results of Dickey-Fuller Test:')
   dftest = adfuller(timeseries, autolag='AIC')
```

```
    dfoutput = pd.Series(dftest[0:4], index=['Test Statistic','p-value','#Lags
Used','Number of Observations Used'])
    for key,value in dftest[4].items():
        dfoutput['Critical Value (%s)'%key] = value
    print (dfoutput)
test_stationarity(df.sales)

#приведение ряда к стационарному
df.sales_log= df.sales.apply(lambda x: np.log(x))
test_stationarity(df.sales_log)

df['first_difference'] = df.sales - df.sales.shift(1)
test_stationarity(df.first_difference.dropna(inplace=False))

df['seasonal_difference'] = df.sales - df.sales.shift(12)
test_stationarity(df.seasonal_difference.dropna(inplace=False))

df['seasonal_first_difference'] = df.first_difference - df.first_difference.shift(12)
test_stationarity(df.seasonal_first_difference.dropna(inplace=False))

#вывод графиков АКФ и ЧАКФ
fig = plt.figure(figsize=(12,8))
ax1 = fig.add_subplot(211)
fig = sm.graphics.tsa.plot_acf(df.seasonal_first_difference.iloc[13:], lags=40, ax=ax1)
ax2 = fig.add_subplot(212)
fig = sm.graphics.tsa.plot_pacf(df.seasonal_first_difference.iloc[13:], lags=40, ax=ax2)

#определение параметров модели
mod = sm.tsa.statespace.SARIMAX(df.sales, trend='n', order=(0,1,0),
seasonal_order=(0,1,1,12))
```

```
results = mod.fit()
print (results.summary)

#построение прогноза
df['forecast'] = results.predict(start = 102, end= 114, dynamic= True)
df[['sales', 'forecast']].plot(figsize=(12, 8))
plt.savefig('ts_df_predict.png', bbox_inches='tight')

#график сравнения фактических показателей с прогнозными
npredict =df.sales['2017'].shape[0]
fig, ax = plt.subplots(figsize=(12,6))
npre = 12
ax.set(title='Dynamic of sales', xlabel='Date', ylabel='Sales')
ax.plot(df.index[-npredict-npre+1:], df.ix[-npredict-npre+1:, 'sales'], 'o',
label='Observed')
ax.plot(df.index[-npredict-npre+1:], df.ix[-npredict-npre+1:, 'forecast'], 'g',
label='Dynamic forecast')
legend = ax.legend(loc='lower right')
legend.get_frame().set_facecolor('w')
plt.savefig('ts_predict_compare.png', bbox_inches='tight')

#добавление новых значений в модель
start = datetime.datetime.strptime("2018-07-01", "%Y-%m-%d")
date_list = [start + relativedelta(months=x) for x in range(0,12)]
future = pd.DataFrame(index=date_list, columns= df.columns)
df = pd.concat([df, future])
```

```
#построение модели прогноза
df['forecast'] = results.predict(start = 114, end = 125, dynamic= True)
df[['sales', 'forecast']].ix[-24:].plot(figsize=(12, 8))
plt.savefig('ts_predict_future.png', bbox_inches='tight')
```

ПРИЛОЖЕНИЕ Б

ИСХОДНЫЕ ДАННЫЕ ДЛЯ АНАЛИЗА

Month	Sales	Exchange rates (USD/RUB)	Number of buyers	Weather (0-sunny, 1-rainy, 2-snow)
2009-01	648	32,61	95	2
2009-02	646	36,02	101	2
2009-03	639	34,83	100	2
2009-04	654	33,51	115	2
2009-05	630	32,08	102	0
2009-06	622	31,12	114	1
2009-07	617	31,41	116	0
2009-08	613	31,61	120	0
2009-09	661	30,79	125	1
2009-10	695	29,35	128	1
2009-11	960	28,83	110	2
2009-12	707	30,06	100	2
2010-01	817	29,72	95	2
2010-02	839	30,12	99	2
2010-03	810	29,51	135	2
2010-04	789	29,21	150	2
2010-05	760	30,51	165	1
2010-06	724	31,08	168	0
2010-07	704	30,56	170	0
2010-08	691	30,45	152	0
2010-09	745	30,83	143	0
2010-10	803	30,42	132	1
2010-11	780	30,91	114	2
2010-12	761	30,75	108	2
2011-01	857	29,93	101	2
2011-02	907	29,29	92	2
2011-03	873	28,42	110	2
2011-04	910	28,14	166	0
2011-05	900	27,96	178	0
2011-06	880	27,98	180	1
2011-07	867	27,89	174	1
2011-08	584	28,86	154	0

2011-09	928	30,43	133	0
2011-10	1064	31,33	127	1
2011-11	1103	30,84	115	2
2011-12	1026	31,56	107	2
2012-01	1102	31,47	100	2
2012-02	1080	29,91	139	2
2012-03	1034	29,29	148	2
2012-04	1083	29,47	156	2
2012-05	1078	30,95	190	1
2012-06	1020	32,82	164	0
2012-07	984	32,53	158	0
2012-08	952	31,88	141	1
2012-09	1033	31,39	132	1
2012-10	1114	31,12	146	0
2012-11	1160	31,42	102	2
2012-12	1058	30,76	110	2
2013-01	1209	30,24	115	2
2013-02	1200	30,12	118	2
2013-03	1130	30,78	144	2
2013-04	1182	31,32	156	1
2013-05	1152	31,32	172	1
2013-06	1116	32,28	184	1
2013-07	1098	32,64	179	1
2013-08	1044	33	170	0
2013-09	1142	32,51	165	0
2013-10	1222	32,12	143	1
2013-11	1234	32,69	122	1
2013-12	1155	32,86	119	2
2014-01	1286	33,64	110	2
2014-02	1281	35,24	107	2
2014-03	1224	36,23	145	2
2014-04	1280	35,66	199	2
2014-05	1228	34,72	204	1
2014-06	1181	34,39	200	1
2014-07	1156	34,42	188	0
2014-08	1124	36,11	175	0
2014-09	1152	37,99	164	0
2014-10	1205	40,74	152	1

2014-11	1260	46,33	131	1
2014-12	1188	54,43	122	2
2015-01	1212	65,29	114	2
2015-02	1269	64,29	111	2
2015-03	1246	60,66	169	2
2015-04	1299	52,36	185	1
2015-05	1284	50,34	214	1
2015-06	1345	54,37	203	1
2015-07	1341	56,98	198	1
2015-08	1308	65,02	186	0
2015-09	1448	66,59	176	0
2015-10	1454	62,7	154	0
2015-11	1467	64,91	152	2
2015-12	1431	70,22	130	2
2016-01	1510	76,58	135	2
2016-02	1558	77,13	128	2
2016-03	1536	70,23	164	2
2016-04	1523	66,47	187	1
2016-05	1492	65,97	234	1
2016-06	1437	65,13	242	1
2016-07	1365	64,11	210	1
2016-08	1310	64,81	205	0
2016-09	1441	64,76	195	0
2016-10	1450	62,46	188	0
2016-11	1424	64,18	153	1
2016-12	1360	61,64	146	2
2017-01	1429	59,65	134	2
2017-02	1440	58,09	123	2
2017-03	1414	58,24	177	2
2017-04	1424	56,31	195	1
2017-05	1408	56,75	198	1
2017-06	1337	57,44	165	0
2017-07	1258	59,57	174	1
2017-08	1214	59,79	175	0
2017-09	1326	57,71	164	1
2017-10	1417	57,68	151	0
2017-11	1417	59	142	1
2017-12	1329	58,69	130	2

2018-01	1461	56,59	124	2
2018-02	1425	56,62	120	2
2018-03	1419	57,01	186	2
2018-04	1432	61,55	223	2
2018-05	1394	62,31	250	1
2018-06	1327	62,75	241	0