

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 14.10.2023 07:17:44

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eab0f73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное образовательное

учреждение высшего образования

«Юго-Западный государственный университет»

(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ

Проректор по учебной работе

О.Г. Локтионова

« 5 » 10 2023 г.



## ТЕХНОЛОГИИ ОБРАБОТКИ ИНФОРМАЦИИ

Методические указания по выполнению курсовой работы для студентов направления подготовки 09.03.02 Информационные системы и технологии

Курск 2023

УДК 681.5

Составитель А.В. Ткаченко, Д.О. Бобынцев

Рецензент

Кандидат технических наук, доцент Халин Ю.А.

**Технологии обработки информации:** методические рекомендации по выполнению курсовой работы / Юго-Зап. гос. ун-т; сост А.В. Ткаченко, Д.О. Бобынцев. Курск, 2023. 15 с.

Методические рекомендации разработаны на основе учебного плана подготовки бакалавров направления 09.03.02 «Информационные системы и технологии», нормативных документов по аттестации студентов с учётом накопленного в ЮЗГУ опыта. Приводятся основные сведения нормативно-технического характера. В приложении дана тематика курсовых работ.

Текст печатается в авторской редакции

Подписано в печать 5.10.2023. Формат 60x84 1/16.

Усл.печ. л. 0,87. Уч.-изд. л. 0,79. Тираж 100 экз. Заказ 1179 Бесплатно.

Юго-Западный государственный университет.  
305040, г. Курск, ул. 50 лет Октября, 94.

## Общие положения

Целью выполнения курсовой работы является расширение, углубление и закрепление студентами теоретических знаний, полученных при изучении таких дисциплин как «Теория информационных процессов и систем», «Информатика», «Информационные технологии» и других, путем применения их к комплексному решению поставленной задачи по изучаемой дисциплине «Технологии обработки информации».

Основными задачами при выполнении курсовой работ являются:

- закрепление, углубление и систематизация полученных студентами знаний и выработка умения самостоятельно применять их к решению конкретных задач;
- приобретение и подтверждение наличия навыков исследовательской, расчетной и конструкторской работы;
- закрепление навыков работы с компьютерной и офисной техникой, использования современных информационных технологий;
- воспитание чувства ответственности за принимаемые решения;
- развитие навыков работы с учебной, научной и справочной литературой, нормативно-правовой документацией, периодической печатью, стандартами, типовыми проектами и т.п.
- овладения навыками грамотного, ясного и сжатого изложения результатов работы и аргументированной защиты принятых решений и сделанных выводов;
- формирование научного мировоззрения;
- приобретение навыков регулярной ритмичной работы, развитие самостоятельности и инициативы, воспитание сознательного и творческого отношения к труду.

При разработке курсовой работы необходимо учитывать то обстоятельство, что внедрение автоматизированных технологий решения экономических задач ведет к качественным переменам в труде пользователей: расширяются их профессиональные знания, приобретаются навыки работы в автоматизированной информационной системе.

Внедрение новой информационной технологии на рабочем

месте пользователя может иметь ряд позитивных последствий:

- обработка исходных данных и проведение расчетов поручается не имеющим высокой квалификации и необходимых практических навыков работникам, а высококвалифицированным специалистам отводится время на анализ, выбор вариантов разработки и принятия управленческих решений;

- работа с ЭВМ приводит к повышению квалификации всех исполнителей и общему, довольно высокому уровню их профессиональной культуры;

- сэкономленное в результате автоматизации обработки данных и оформления документов время используется на анализ нескольких вариантов, получение альтернативных оценок ситуаций, что необходимо для принятия обоснованных решений.

Таким образом, создание технологии обработки информации в той или иной предметной области не только выдвигает к персоналу новые требования, но и позволяет качественно изменить их труд.

Наиболее важным требованием к разработчикам информационных технологий является умение осуществить постановку задач, т.е. составить алгоритмы их решения, установить состав информационного наполнения компьютерных процедур обработки информации для получения искомых результатов, сформулировать требования к методам контроля решаемых задач.

### **Рекомендации по выполнению работы**

В работе требуется выбрать предметную область – компанию, ведущую бизнес. Для этой компании необходимо разработать хранилище данных и провести аналитическую обработку информации при помощи аналитической системы Deductor Studio с целью выработки предложений по увеличению прибыли.

### **Создание Проекта в аналитической системе Deductor.**

Для создания Проекта Deductor использует файлы в текстовом формате (обычный текст).

Файл каждым студентом создается самостоятельно.

Например, создадим файл в Excel и преобразуем созданную таблицу в текстовый файл. В нашем случае необходимо в Excel создать таблицу продаж. При создании студентом таблицы она должна включать не менее 50 строк. Пример таблицы показан на рис. 1.

	A	B	C	D	E	F	G	H	I
1									
2		№№	Дата	Точка продажи	Марка телефона	Модель телефона	Объем продаж, шт.	Цена, руб. за шт.	Сумма выручки
3		1	15.05.2020	Магазин №1	Samsung	E2121	4	501	2004
4		2	16.05.2020	Магазин №1	Samsung	I8000	4	2529	10116
5		3	17.05.2020	Магазин №1	Samsung	C5212	2	898	1796
6		4	18.05.2020	Магазин №1	Samsung	E1150	1	335	335
7		5	19.05.2020	Магазин №1	Sony Ericsson	J105	4	897	3588
8		6	20.05.2020	Магазин №1	Samsung	N900	1	3010	602
9		7	21.05.2020	Магазин №2	Samsung	X6	2	2249	4498
10		8	22.05.2020	Магазин №2	Samsung	I5500	4	1556	6224
11		9	23.05.2020	Магазин №2	Sony Ericsson	C903	3	2362	7086
12		10	24.05.2020	Магазин №2	Nokia	6500	1	1783	1783

Рисунок 1 – Пример таблицы исходных данных

Чтобы названия столбцов не вышли за пределы названий в текстовом формате, сократим названия столбцов таблицы, как показано на рисунке 2.

	A	B	C	D	E	F	G	H	I
1									
2		№№	Дата	Магазин	Телефон	Модель	Продажи, шт.	Цена, руб. за шт.	Сумма выручки
3		1	15.05.2020	Магазин №1	Samsung	E2121	4	501	2004
4		2	16.05.2020	Магазин №1	Samsung	I8000	4	2529	10116
5		3	17.05.2020	Магазин №1	Samsung	C5212	2	898	1796
6		4	18.05.2020	Магазин №1	Samsung	E1150	1	335	335
7		5	19.05.2020	Магазин №1	Sony Ericsson	J105	4	897	3588
8		6	20.05.2020	Магазин №1	Samsung	N900	1	3010	602
9		7	21.05.2020	Магазин №2	Samsung	X6	2	2249	4498
10		8	22.05.2020	Магазин №2	Samsung	I5500	4	1556	6224
11		9	23.05.2020	Магазин №2	Sony Ericsson	C903	3	2362	7086
12		10	24.05.2020	Магазин №2	Nokia	6500	1	1783	1783

Рисунок 2

Для преобразования, выберем «Текстовые файлы (с разделителями табуляции)». Сохраним файл под именем 111.txt в созданную автоматически при установке Deductor папку Документы – BaseGroup – Deductor – Samples. Далее, запускаем Deductor.

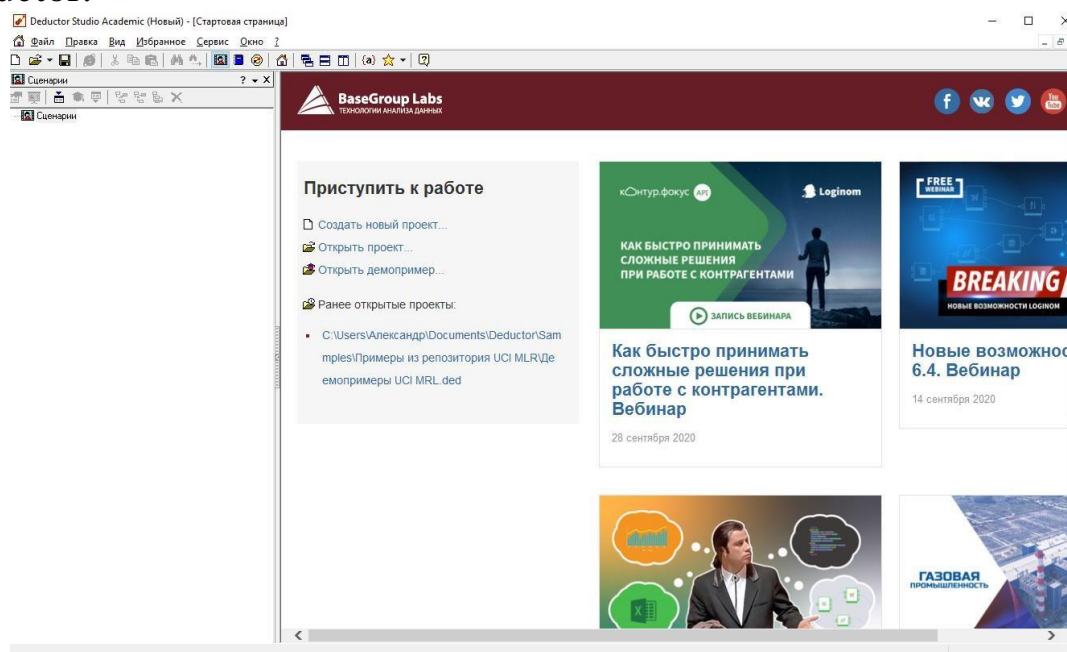


Рисунок 3 – Главное окно

Выбираем из Меню пиктограмму Мастер импорта, расположенную надсловом Сценарии. Запустится Мастер импорта.

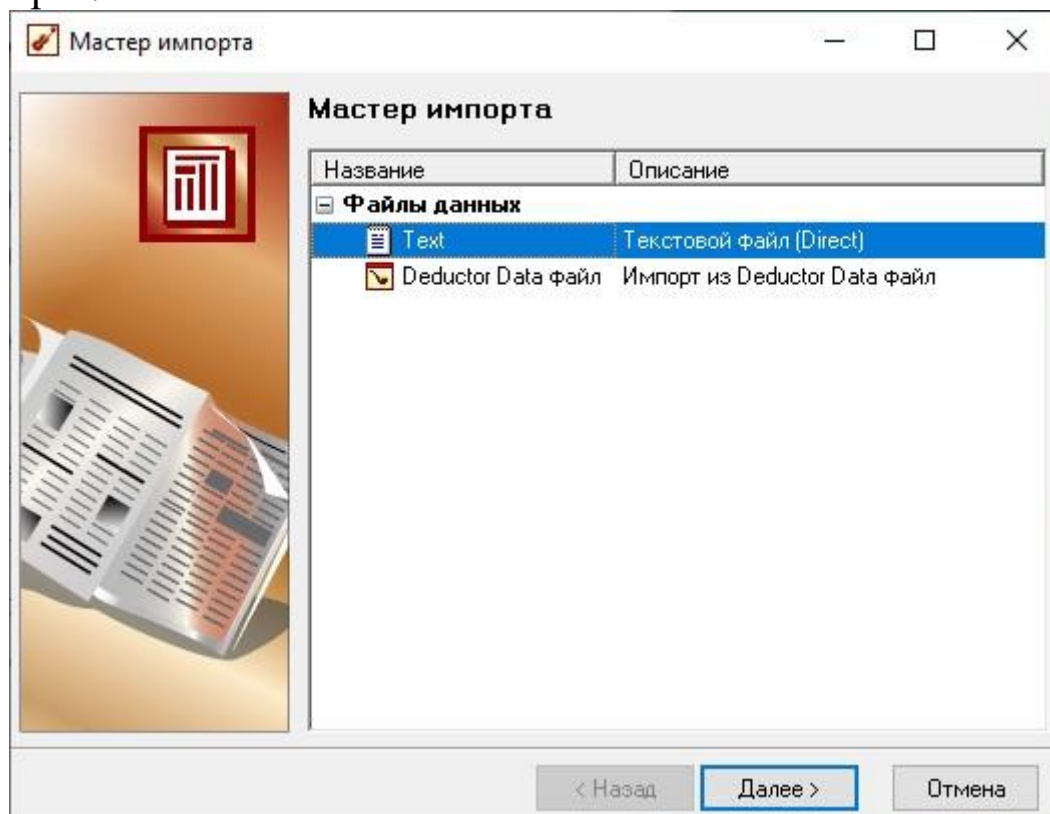


Рисунок 4 – Мастер импорта

В следующем окне необходимо выбрать файл для импорта. Находим наш файл 111.txt при помощи кнопки с тремя точками.

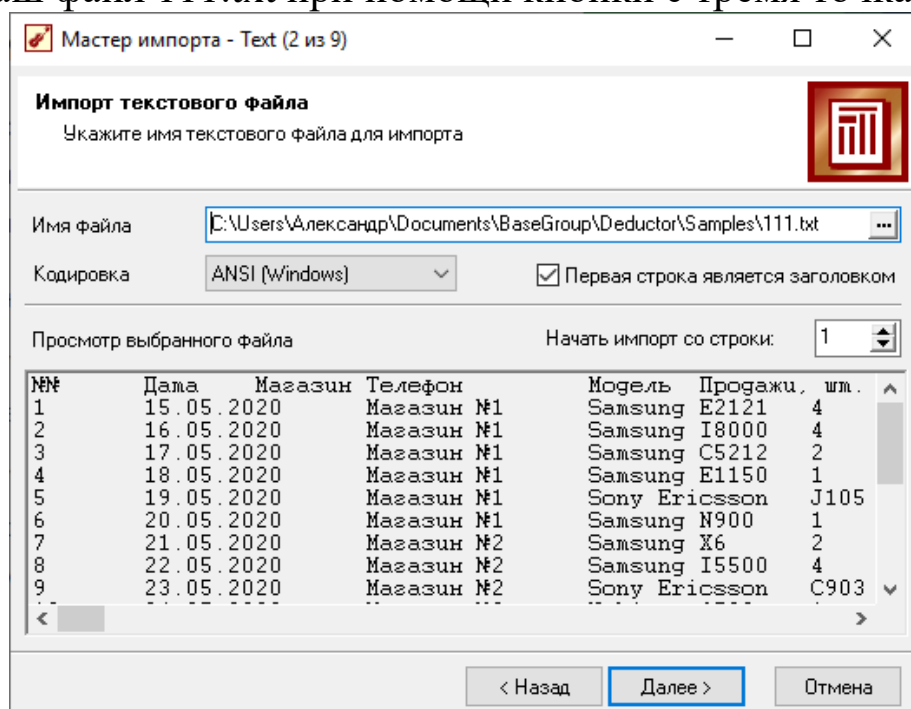


Рисунок 5 – Импорт файла

Далее проводится настройка импорта, где вы можете поменять опции при необходимости. После этого можно запускать процесс импорта. На этапе определения способов отображения выбираем Таблица. После указания имени итогового проекта программа представит данные, которые готовы к обработке.

Для обработки данных предназначен Мастер обработки.

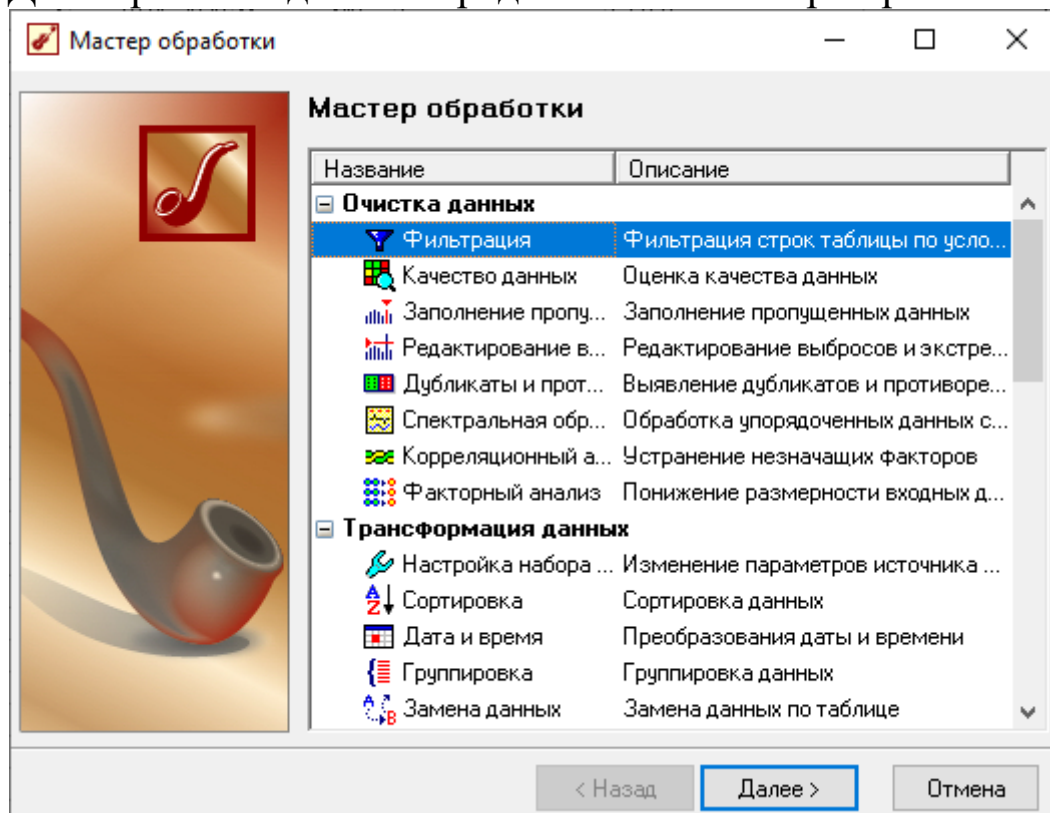


Рисунок 6 – Мастер обработки

Для того, чтобы показать обработку данных, необходимо построить OLAP-отчёт о продажах и выполнить кластерный анализ методом древовидной кластеризации и методом K-средних.

Помимо результатов обработки данных, в работе должна быть структура метаданных хранилища. Пример показан на рисунке 7. Сформируем процесс, добавляя атрибуты к измерениям и ссылки на другие измерения. В результате, сформированные метаданные процесса примут вид, показанный на рисунке 8. Это последний этап проектирования структуры метаданных.

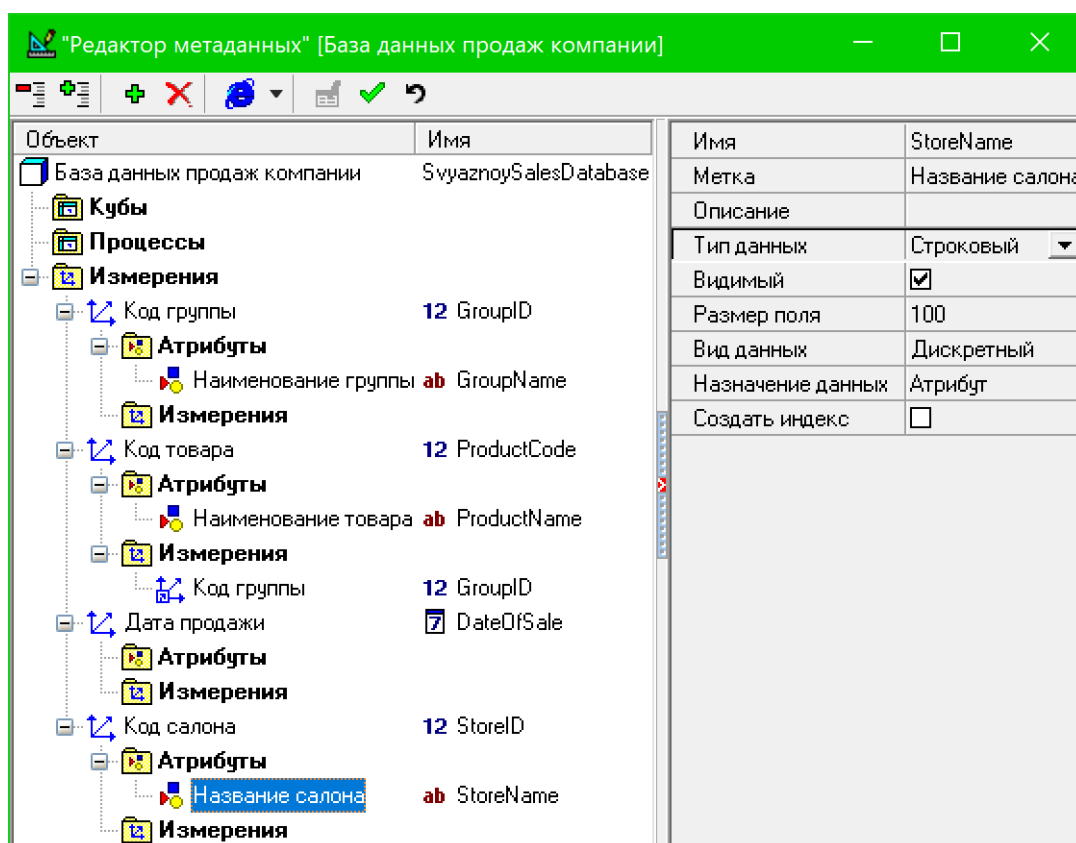


Рисунок 7 – Структура метаданных

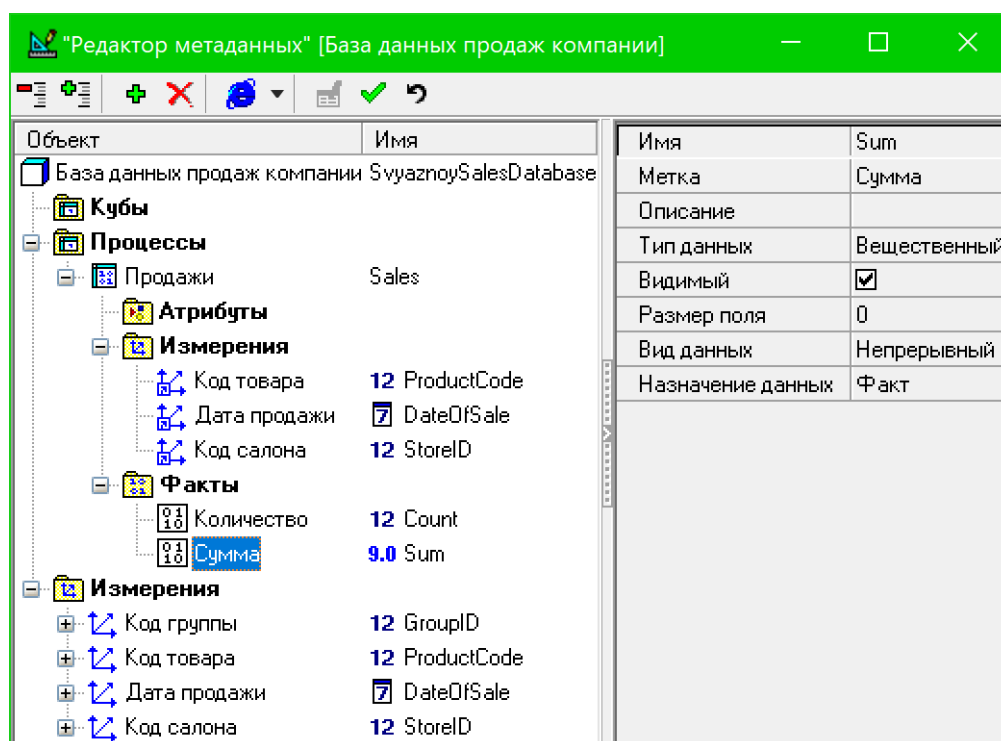


Рисунок 8 – Метаданные процесса

На рисунке 9 представлен пример OLAP-отчёта по работе салонов связи за все месяцы (с января по ноябрь).



	+ Дата продажи (Год) ▾	Дата продажи (Месяц) ▾
	2020	
Код салона   Название салона ▾	Σ Количество   Сумма	Σ Сумма   Сумма
Cstore, г. Москва, "ТРЦ Калейдоскоп", ул. Сходненская, дом 56	535,00	17 612 550,00
Cstore, г. Москва, "ТРЦ Мозаика", ул. Кожуховская 7-я, дом 9	542,00	21 853 280,00
Cstore, г. Москва, "ТЦ Остров мечты", пр. Андропова, дом 1	520,00	19 951 800,00
Cstore, г. Москва, ул. Тверская, дом 6, строение 1	623,00	24 358 770,00
Связной, г. Москва, Новочеркасский бульвар, дом 10, корпус 1	512,00	8 982 661,00
Связной, г. Москва, пл. Золоторожский Вал, дом 42	502,00	7 908 765,00
Связной, г. Москва, пл. Киевского вокзала, дом 2	501,00	6 833 605,00
Связной, г. Москва, пл. Павелецкая, дом 2, строение 1	650,00	9 382 546,00
Связной, г. Москва, пр. Вернадского, дом 86А	519,00	1 010 753,00
Связной, г. Москва, пр. Пролетарский, дом 30	531,00	3 086 480,00
Связной, г. Москва, ул. Арбат, дом 1	540,00	1 908 008,00
Связной, г. Москва, ул. Декабристов, дом 12	517,00	3 235 496,00
Связной, г. Москва, ул. Зеленодольская, дом 40	504,00	2 008 794,00
Связной, г. Москва, ул. Ленинская слобода, дом 26, строение 2	495,00	3 632 075,00
Связной, г. Москва, ул. Яблочкова, дом 19Г	585,00	3 781 073,00
Связной, г. Москва, ул. Ярцевская, дом 19	500,00	8 456 900,00
Связной, г. Москва, ш. Боровское, дом 51, строение 1	504,00	2 241 006,00
Связной, г. Москва, ш. Рублёвское, дом 62	564,00	7 430 061,00
Связной, г. Москва, ш. Щелковское, дом 75	513,00	10 473 466,00
Связноц, г. Москва, ул. Кировоградская, дом 13А	509,00	1 325 730,00
<b>Итого:</b>	<b>10 666,00</b>	<b>165 533 819,00</b>

Рисунок 9 – OLAP-отчёт работы всех салонов за год

Данные отчёта по продажам различных категорий товаров можно представить в виде кросс-отчёта. Гистограмма иллюстрирует количество проданных товаров каждой группы, пример на рис. 10.

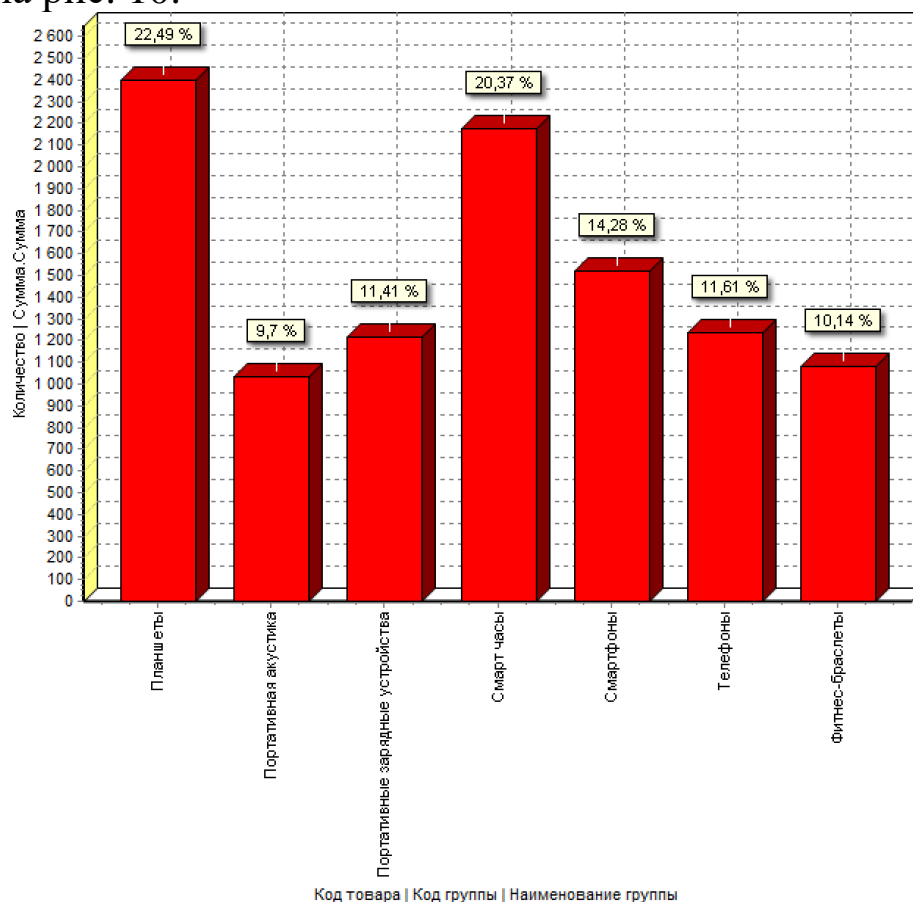


Рисунок 10 – Результаты продаж товаров каждой группы за

год

Из рисунка 22 следует, что группой товаров, проданных в большем количестве, являются «Планшеты», а в меньшем – «Портативная акустика».

### Кластеризация данных. Метод древовидной кластеризации

Исходными данными для каждого из методов является загруженная в программу «Statistica 13.5.0.17» таблица, содержащая информацию о сумме продаж каждого салона за каждый месяц.

	Январь	Февраль	Март	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь
Cstore, г. Москва, "ТРЦ Калейдоскоп", ул. Сходненская, дом 56	1610110	1299480	1774970	1599360	1473660	1224510	2002050	2224500	1324470	3079440
Cstore, г. Москва, "ТРЦ Мозаика", ул. Кожуховская 7-я, дом 9	1908930	1975910	2859480	1473660	2176850	1507050	2529540	3134430	1440070	2847360
Cstore, г. Москва, "ТЦ Остров мечты", пр. Андропова, дом 1	2402460	2369400	2491440	2199600	2313480	1658580	1309920	1446370	2013990	1746560
Cstore, г. Москва, ул. Тверская, дом 6, строение 1	2474550	1399440	1935010	2408890	3409380	3335990	1440070	3311590	1399440	3244410
Связной, г. Москва, Новочеркасский бульвар, дом 10, корпус 1	419580	689540	2669400	249950	975390	1698070	636510	64841	1309920	269460
Связной, г. Москва, пл. Золоторожский Вал, дом 42	727480	264470	209500	839580	713490	2180010	226260	59600	1895520	792855
Связной, г. Москва, пл. Киевского вокзала, дом 2	859570	122145	2529540	138180	895440	90630	919540	77220	1002410	198930
Связной, г. Москва, пл. Павелецкая, дом 2, строение 1	1674500	1124550	742515	991380	609390	478800	815490	870881	1599600	475440
Связной, г. Москва, пр. Вернадского, дом 86А	49500	60445	47952	166336	68954	55944	42957	209580	142749	166336
Связной, г. Москва, пр. Пролетарский, дом 30	391440	82680	489510	103545	117110	122145	866490	143400	671580	98580
Связной, г. Москва, ул. Арбат, дом 1	129060	318010	202420	136950	314370	62750	187530	307560	150220	99138
Связной, г. Москва, ул. Декабристов, дом 12	254490	66515	273890	1195954	148347	54230	149160	815490	242460	34960
Связной, г. Москва, ул. Зеленодольская, дом 40	59600	162720	57770	377460	167944	167580	91290	671520	50490	202420
Связной, г. Москва, ул. Ленинская слобода, дом 26, строение 2	152830	847470	109940	107070	643540	919540	70460	188460	113305	539460
Связной, г. Москва, ул. Яблочкова, дом 19Г	141010	917460	82680	639600	254949	429570	279133	541155	118056	377460
Связной, г. Москва, ул. Ярцевская, дом 19	2239440	1777050	494550	1566930	305030	689540	516530	92220	532590	243020
Связной, г. Москва, ш. Боровское, дом 51, строение 1	103600	43010	243390	69660	202050	41040	160540	1143956	152490	81270
Связной, г. Москва, ш. Рублёвское, дом 62	67158	779400	475440	927420	70453	1610110	1209560	348090	1662990	279440
Связной, г. Москва, ш. Щелковское, дом 75	699500	1364500	289420	65956	1774970	1959510	1299480	76110	719520	2224500
Связной, г. Москва, ул. Кировоградская, дом 13А	42560	242460	98450	234640	74500	108780	46870	230450	47520	199500

Рисунок 11 – Исходные данные

На основе неё с помощью метода Joining (Tree Clustering) строится диаграмма «Vertical icicle plot» (рисунок 12).

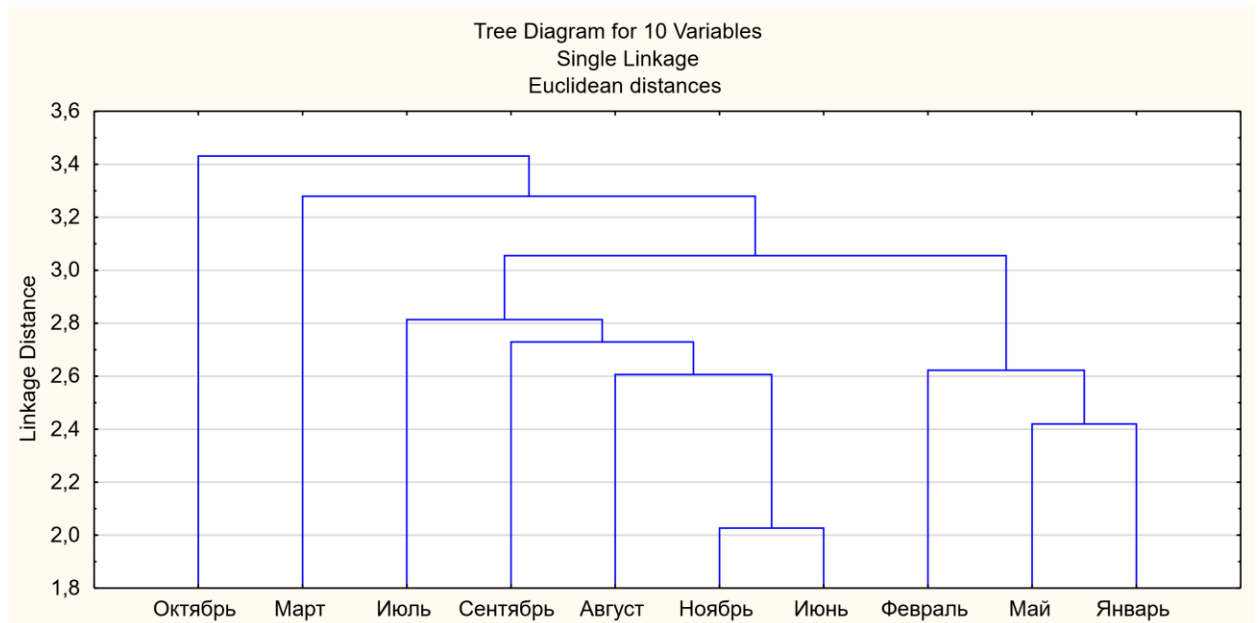


Рисунок 12 – Диаграмма Vertical icicle plot

## Метод К-средних

Суть этого метода состоит в следующем: исследователь заранее определяет количество классов ( $k$ ), на которые необходимо разбить имеющиеся наблюдения, и первые  $k$ -наблюдений становятся центрами этих классов. Для каждого следующего наблюдения рассчитываются расстояния до центров кластеров и данное наблюдение относится к тому кластеру, расстояние до которого было минимальным. После чего для этого кластера (в котором увеличилось количество наблюдений) рассчитывается новый центр тяжести (как среднее по каждому показателю) по всем включённым в кластер наблюдениям.

Результаты разбиения на кластеры К-средних представлены на рисунках 13 – 15.

		Members of Cluster Number 1 (Продажи (для анализа)) and Distances from Respective Cluster Center Cluster contains 4 cases			
		Distance			
Cstore, г. Москва, "ТРЦ Калейдоскоп", ул. Сходненская, дом 56		484610,4			
Cstore, г. Москва, "ТРЦ Мозаика", ул. Кожуховская 7-я, дом 9		416757,1			
Cstore, г. Москва, "ТЦ Остров мечты", пр. Андропова, дом 1		573265,6			
<b>Cstore, г. Москва, ул. Тверская, дом 6, строение 1</b>		690752,9			

Рисунок 13 – Разбиение на кластеры методом К-средних (кластер 1)

		Members of Cluster Number 2 (Продажи (для анализа)) and Distances from Respective Cluster Center Cluster contains 9 cases			
		Distance			
Связной, г. Москва, пр. Вернадского, дом 86А		162768,7			
Связной, г. Москва, пр. Пролетарский, дом 30		326058,3			
<b>Связной, г. Москва, ул. Арбат, дом 1</b>		106313,3			
Связной, г. Москва, ул. Декабристов, дом 12		316108,3			
Связной, г. Москва, ул. Зеленодольская, дом 40		110476,1			
Связной, г. Москва, ул. Ленинская слобода, дом 26, строение 2		352682,1			
Связной, г. Москва, ул. Яблочкова, дом 19Г		238390,7			
Связной, г. Москва, ш. Боровское, дом 51, строение 1		254231,3			
Связной, г. Москва, ул. Кировоградская, дом 13А		130399,6			

Рисунок 14 – Разбиение на кластеры методом К-средних (кластер 2)

		Members of Cluster Number 3 (Продажи (для анализа)) and Distances from Respective Cluster Center Cluster contains 7 cases			
		Distance			
Связной, г. Москва, Новочеркасский бульвар, дом 10, корпус 1		595846,1			
Связной, г. Москва, пл. Золоторожский Вал, дом 42		534290,3			
Связной, г. Москва, пл. Киевского вокзала, дом 2		683591,3			
Связной, г. Москва, пл. Павелецкая, дом 2, строение 1		441683,4			
<b>Связной, г. Москва, ул. Ярцевская, дом 19</b>		696297,6			
Связной, г. Москва, ш. Рублёвское, дом 62		479179,6			
Связной, г. Москва, ш. Щелковское, дом 75		766043,6			

Рисунок 15 – Разбиение на кластеры методом К-средних (кластер 3)

Для получения графика нажмём на кнопку «Graph of Means» в окне «к – Means Clustering Results».

На полученном графике (рисунок 16) показаны средние значения переменных для каждого кластера.

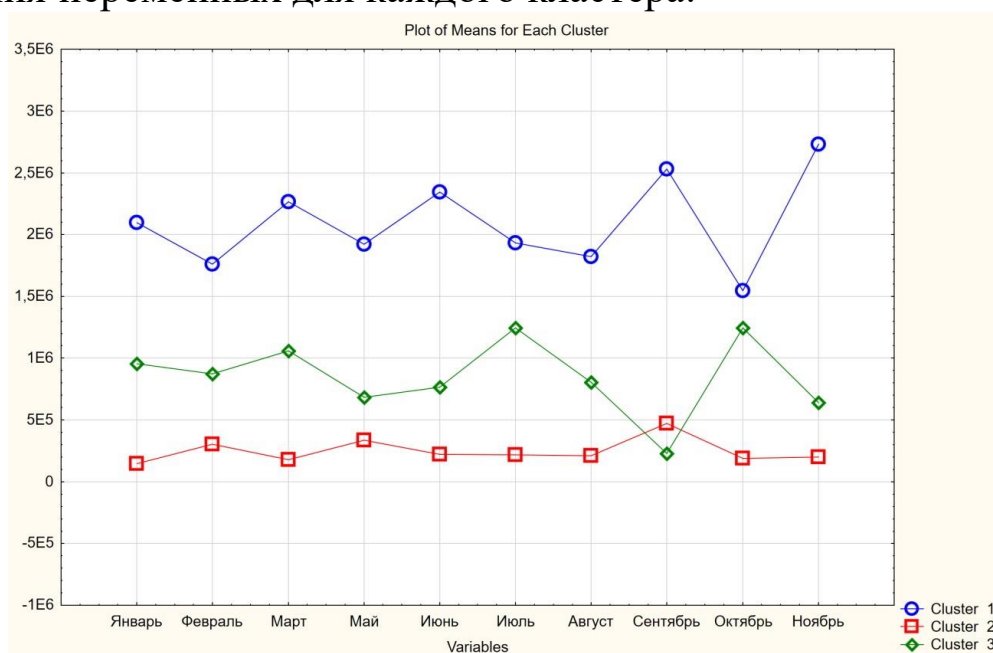


Рисунок 16 – Усреднённые значения переменных

Трендом является функция заданного вида, с помощью которой можно аппроксимировать построенный по имеющимся данным график. Тренд служит для выявления тенденций развития процесса, представленного в виде диаграммы, и обеспечивает прогноз на один заданный период.

На основе данных об объёме продаж сети салонов «Связной», линия тренда будет иметь вид, представленный на рисунке 17.

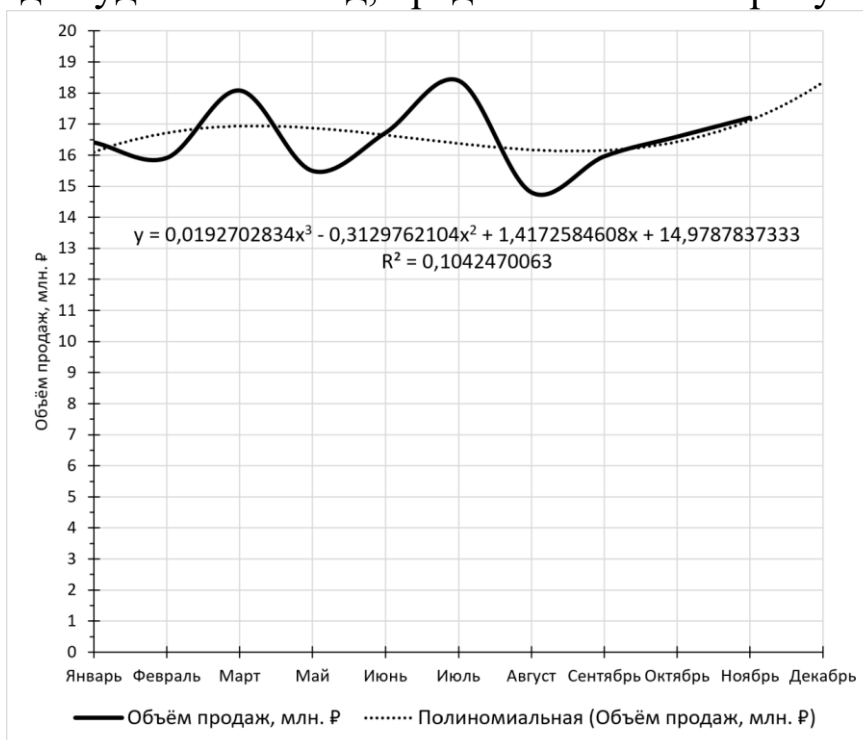


Рисунок 17 – Линия тренда

## Оформление курсовой работы

Требования к оформлению определяются стандартом Юго-Западного государственного университета СТУ 02.030-2023 «Курсовые работы (проекты). Выпускные квалификационные работы. Общие требования к структуре и оформлению», доступном на официальном сайте [www.swsu.ru](http://www.swsu.ru) в разделе «Сотруднику – Документация системы менеджмента качества – Стандарты университета». При оформлении пояснительной записки к курсовой работе необходимо руководствоваться данным стандартом. Титульный лист и бланк задания на курсовую работу должны соответствовать стандарту. Особенное внимание необходимо обратить на поле «Шифр» на бланке задания – в данном случае под шифром подразумевается индивидуальный шифр студента, указанный в зачётной книжке.

Примерный перечень разделов работы:

1. Введение
2. Теоретические основы создания хранилищ данных.
3. Разработка хранилища данных для аналитической обработки информации ... (объект исследования)
4. Результаты аналитической обработки.
5. Заключение.
6. Список использованных источников.
7. Приложение (если есть)

*Список использованных источников* содержит перечень источников, использованных автором, оформленный по ГОСТ 7.1-2003. Пример:

Бычкова, С.М. Планирование в аудите [Текст] / С.М. Бычкова, А.В. Газорян. – М.: Финансы и статистика, 2001. – 263 с.

В списке обязательно должны присутствовать упомянутый выше стандарт университета и библиографическая ссылка на данные методические указания. **Минимальный объём работы – 30 страниц основного текста.**

## Защита курсовой работы

Выполненную работу студент представляет на проверку руководителю. При наличии замечаний студент вносит в работу исправления.

По результатам работы проводится ее устная защита в форме собеседования. Работа должна быть сдана и защищена до экзамена по дисциплине. В процессе защиты выясняется:

- умение студента проанализировать задание, последовательно, полно и корректно излагать результаты выполненной работы;

- умение студента анализировать предметную область и выявлять бизнес-правила;

- умение студента проектировать хранилища данных;

- наличие практических навыков в области методов аналитической обработки данных.

Представленная руководителю, но не защищенная студентом работа не засчитывается как выполненная.

Критерии оценки работы:

1. Работа оценивается отлично, если содержание работы соответствует упомянутым выше требованиям, нет претензий к оформлению, анализ предметной области в достаточной мере отражает автоматизируемые информационные процессы, хранилище данных спроектировано без замечаний, снижающих качество работы, проведена аналитическая обработка, а студент успешно ответил на все вопросы собеседования.

2. Работа оценивается хорошо, если содержание работы соответствует упомянутым выше требованиям, анализ предметной области в достаточной мере отражает автоматизируемые информационные процессы, однако имеются замечания по оформлению, или не полностью проведена аналитическая обработка. Оценка «хорошо» так же выставляется, если работа соответствует критериям оценки «отлично», но сдаётся после истечения установленного срока, либо студент неуверенно отвечает на вопросы собеседования.

3. Работа оценивается удовлетворительно, если в работе присутствует только хранилище данных и один из методов аналитической обработки, а также имеются замечания по оформлению работы. Оценка «удовлетворительно» также выставляется, если работа соответствует критериям оценки «хорошо», но сдаётся после истечения установленного срока, либо студент неуверенно отвечает на вопросы собеседования.

4. Работа оценивается неудовлетворительно, если содержание работы не соответствует теме, работа оформлена не по стандарту, объём текстовой части без приложений менее 30 страниц, а также если хранилище спроектировано некорректно и не позволяет говорить об адекватности результатов аналитической обработки.

### **Примеры тем работы**

1. Разработка хранилища данных кафе
2. Разработка хранилища данных автовокзала
3. Разработка хранилища данных магазина электроники
4. Разработка хранилища данных туристического клуба
5. Разработка хранилища данных аэропорта
6. Разработка хранилища данных зоопарка
7. Разработка хранилища данных спортивного клуба
8. Разработка хранилища данных кинотеатра
9. Разработка хранилища данных нефтяной компании
10. Разработка хранилища данных магазина автозапчастей
11. Разработка хранилища данных автосалона
12. Разработка хранилища данных магазина мебели
13. Разработка хранилища данных аптеки
14. Разработка хранилища данных магазина хлебобулочных изделий
15. Разработка хранилища данных розничной торговой сети.
16. Разработка хранилища данных магазина канцтоваров.
17. Разработка хранилища данных гостиницы.
18. Разработка хранилища данных магазина бытовой техники.
19. Разработка хранилища данных сети магазинов «Спортмастер».
20. Разработка хранилища данных сети пекарен «Добропек».