

## **Лабораторная работа №6**

### **Визуализация данных.**

**Цель работы:** изучить процесс визуализации данных.

**Задачи работы:**

- изучить методику визуализации данных в Deductor Studio;
- изучить приведенные в лабораторной работе примеры;
- выполнить контрольное задание.

#### **1. Краткая теория**

##### **1.1. Визуализация данных**

Визуализация — представление данных в виде, который обеспечивает наиболее эффективную работу пользователя. Способ визуализации должен максимально полно отражать поведение данных, содержащуюся в них информацию, тенденции, закономерности и т.д. При этом выбор способа визуализации зависит от характера исследуемых данных и от задачи анализа, а также от предпочтений пользователя.

**Визуализация источников данных.** В источнике данных перед их загрузкой в аналитическую систему аналитику требуется визуально оценить:

- характер, тип и поведение данных;
- динамический диапазон значений;
- степень гладкости;
- наличие факторов, снижающих качество данных, таких как шумы, аномальные и пропущенные значения.

Визуальный анализ источника данных позволяет:

- увидеть, соответствуют ли данные ожидаемым;
- оценить степень пригодности данных к анализу;
- выдвинуть гипотезы о закономерностях процессов, описываемых данными;
- определить, какие виды очистки и предобработки необходимо применить к данным.

**Визуализация данных, загруженных в аналитическое приложение.** После загрузки данных из источника в аналитическое приложение работа с выборкой также начинается с визуального анализа. Однако теперь цели, задачи и методы визуального анализа будут несколько другими.

Нужно убедиться, что данные загрузились правильно: не появились пропуски, сохранилась структура строк и столбцов и т. д. Искажение данных при загрузке может произойти из-за несоответствия их типов, неправильной настройки параметров загрузки и т. д.

**Визуализация данных в процессе их аналитической обработки.** Сложные аналитические процедуры являются многошаговыми. Это означает, что в процессе анализа к данным последовательно применяется несколько алгоритмов или моделей.

Например, сначала данные подвергаются предобработке с целью сглаживания и нормализации, затем к результирующей выборке применяется та или иная модель. При этом выборка, формируемая на выходе каждого алгоритма или модели, может подаваться на вход следующего этапа обработки. Очевидно, что если данные, поступившие с предыдущего этапа, окажутся некорректными, то дальнейшая обработка теряет смысл. Поэтому очень важно предусмотреть визуализацию промежуточных результатов анализа с целью проверки корректности используемых моделей и алгоритмов.

**Визуализация результатов анализа.** После получения конечных результатов аналитической обработки на первый план выходит задача их интерпретации и оценки достоверности. И здесь не обойтись без визуализации. Следует заметить, что, даже если в процессе анализа были получены достоверные и ценные результаты, неудачный выбор визуализации не позволит их интерпретировать, увидеть в них зависимости и закономерности.

### 1.2. Визуализаторы общего назначения

**Графики** представляют собой линии, отображающие зависимость между несколькими переменными в некоторой системе координат. Линия на графике состоит из множества точек, положение каждой из которых определяется значениями зависимой и независимой переменной (переменных). Чаще всего используется декартова система координат ( $X$ ,  $Y$ ,  $Z$ ). Также может применяться полярная система координат ( $z$ ,  $\theta$ ), где положение точки на координатной плоскости зависит от расстояния до начала координат  $z$  и угла  $\theta$ .

Чтобы построить график, достаточно задать таблично значения зависимой и независимой переменной, отметить соответствующие точки на координатной плоскости и соединить их линиями. Линии, соединяющие узлы графика, могут быть прямыми или сглаженными. На рисунке 1 представлены примеры обычного (слева) и сглаженного графика (справа).

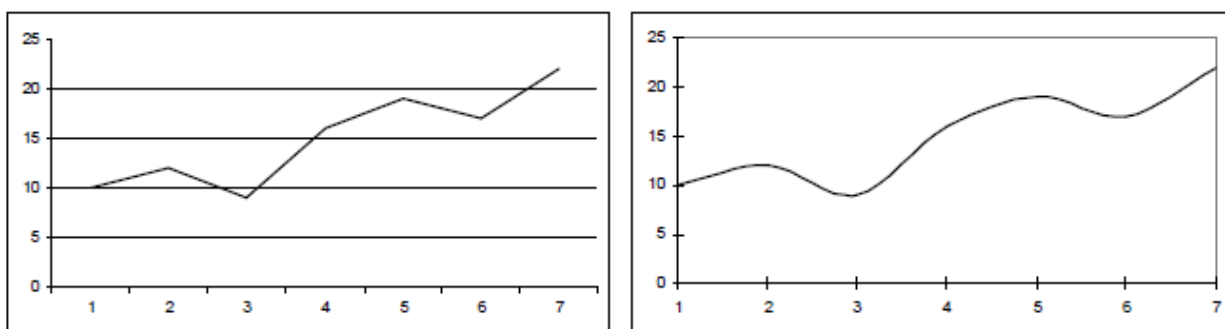
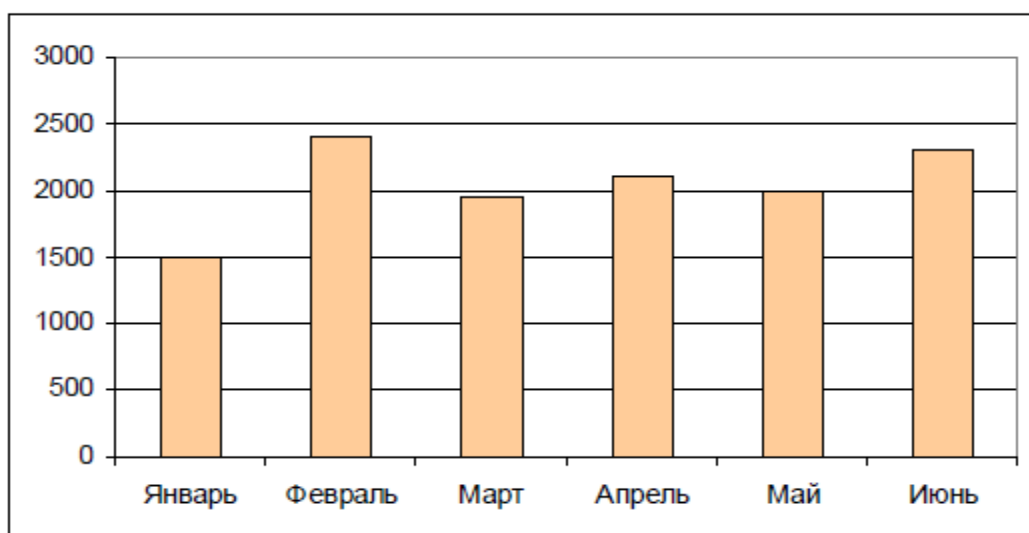


Рисунок 1 – Ломаный и гладкий графики

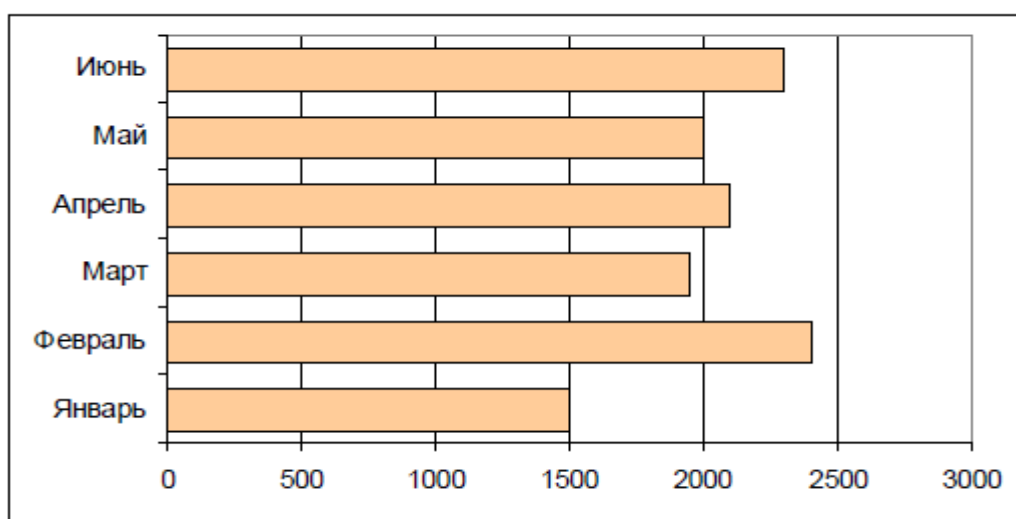
С помощью графика удобнее всего отображать непрерывные (числовые) величины, поскольку можно получить достаточное число точек, чтобы его построить. Если же речь идет о категориальных (дискретных)

значениях, то более подходящим средством визуализации является диаграмма. Принципиального различия между понятиями «график» и «диаграмма» нет. Просто под графиком традиционно понимают представление зависимостей в виде линий, тогда как в **диаграмме** значения отображаются с помощью самых разнообразных объектов и фигур. Как правило, в диаграммах, по горизонтальной оси X откладываются категории, а по вертикальной оси Y — значения.

Самые простые и часто используемые диаграммы — столбиковые. В них значение каждой категории представляется в виде столбика, высота которого пропорциональна соответствующему значению (рисунок 2, а). Разновидностью столбиковой диаграммы является линейчатая диаграмма, которая отличается от столбиковой тем, что ось категорий откладывается вертикально, а ось значений — горизонтально (рисунок 2, б).



а



б

Рисунок 3 - Столбиковая (а) и линейчатая (б) диаграммы

Еще одним распространенным видом диаграмм является круговая диаграмма. Ее очень удобно использовать, если нужно показать долю, которую вносит то или иное значение в общую сумму.

Эта доля может быть выражена как в абсолютных единицах, так и в процентах (например, процент от выручки, который обеспечил товар А). Примеры круговой диаграммы распределения месячных продаж бытовой техники и электроники по категориям товаров представлен на рисунке 3.

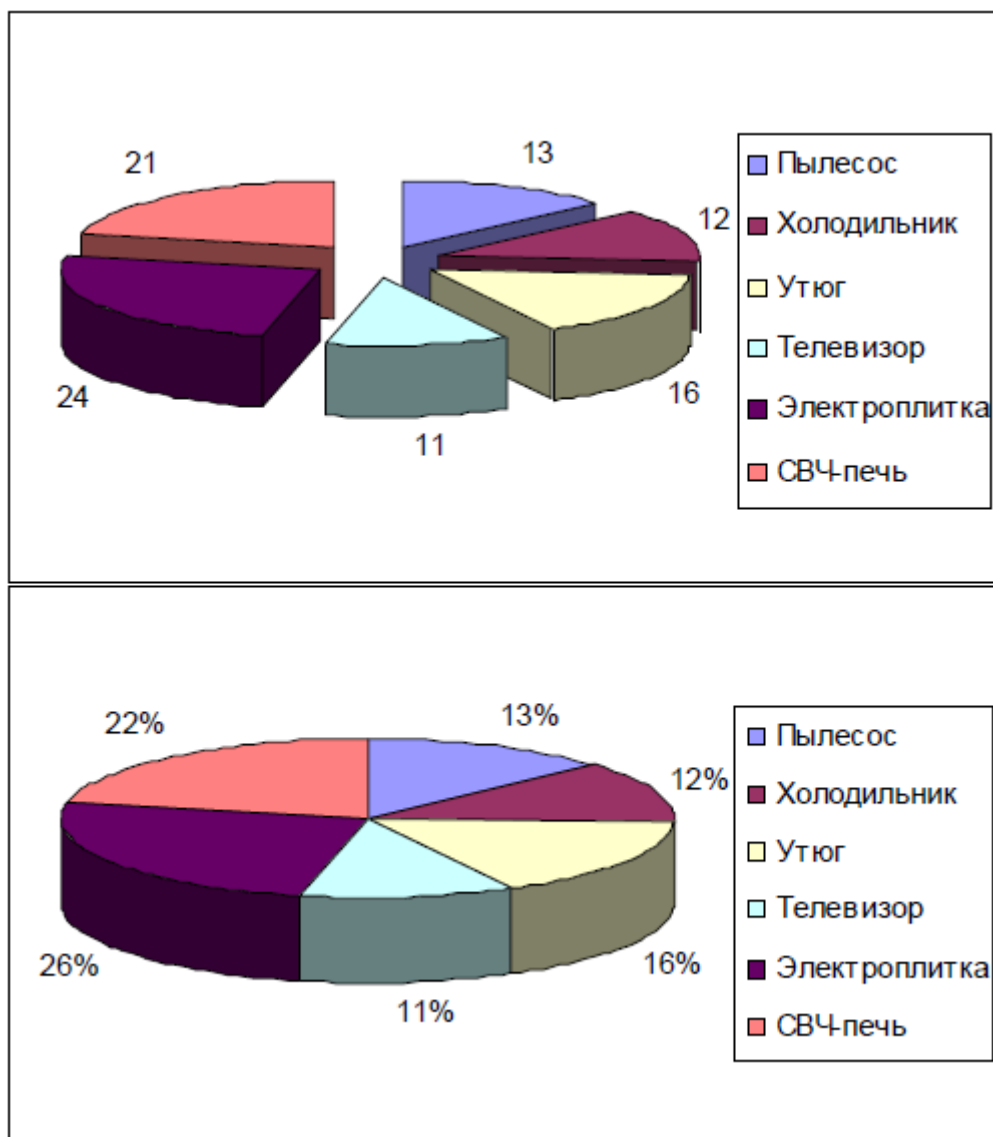


Рисунок 3 – Круговые диаграммы

Хорошую наглядность обеспечивают **лепестковые диаграммы**. На них каждая категория данных представлена в виде отдельной оси (лепестка), на которой отображается соответствующее значение. Затем значения на всех осях соединяются линиями. Если в ряду данных все точки имеют одинаковые значения, то лепестковая диаграмма приобретает вид круга. Наличие на диаграмме нескольких рядов позволяет сравнивать, например, динамику изменения структуры продаж в группах товаров по месяцам (рисунок 4).

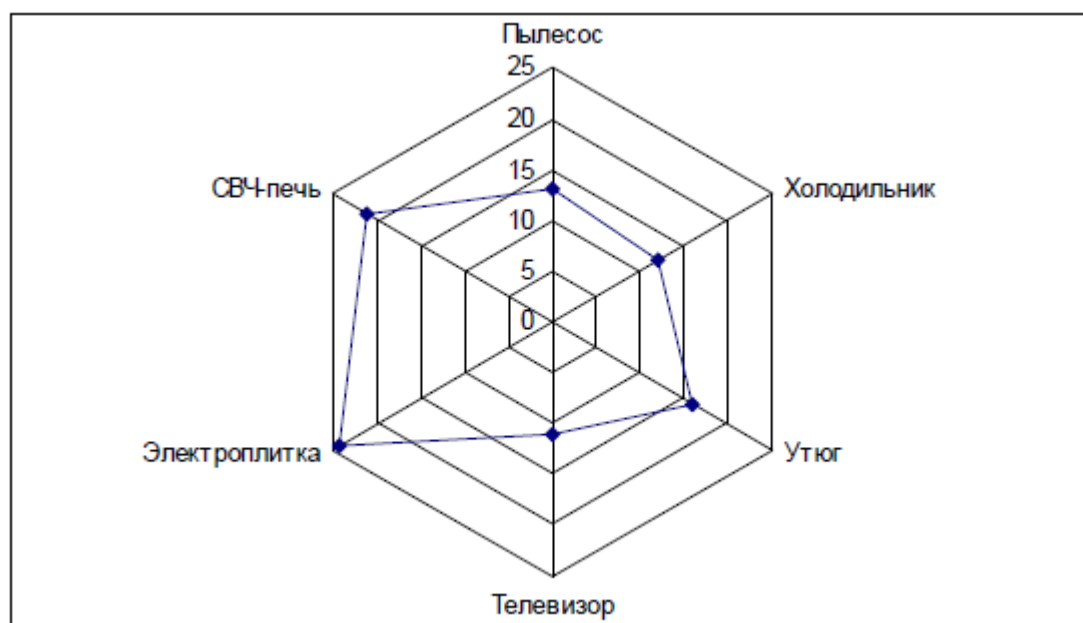


Рисунок 4 – Лепестковая диаграмма

Существует множество других видов диаграмм — диаграммы с накоплением, с областями, пузырьковые, кольцевые и т. д. При этом разновидность диаграммы подбирается исходя из той информации, которую желает извлечь из нее пользователь.

**Гистограмма** показывает распределение набора данных внутри выборки (например, количество заемщиков банка в нескольких возрастных группах) в виде столбиков.

Гистограммы широко используются в статистике для определения наиболее вероятных значений, которые может приобретать некоторая величина, а также для выявления законов распределения, которым подчиняется случайная величина.

Пусть исследуемой величиной являются ежедневные продажи торговой точки в течение месяца. При этом минимальное наблюдаемое значение составило 10 тыс., а максимальное — 100 тыс.

Разобьем диапазон изменения величины на 9 поддиапазонов по 10 тыс. и подсчитаем, сколько раз значение продаж попадает в тот или иной поддиапазон. Результаты сведем в таблицу 1.

Таблица 1 – Исходные данные для гистограммы

Диапазон в тыс.	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
Частота	1	1	2	6	10	6	3	1	1

Как видим, продажи на сумму от 10 до 20 тыс. наблюдались только один раз, от 30 до 40 тыс. — два раза и т. д. На основе полученной таблицы строим гистограмму (рисунок 5).

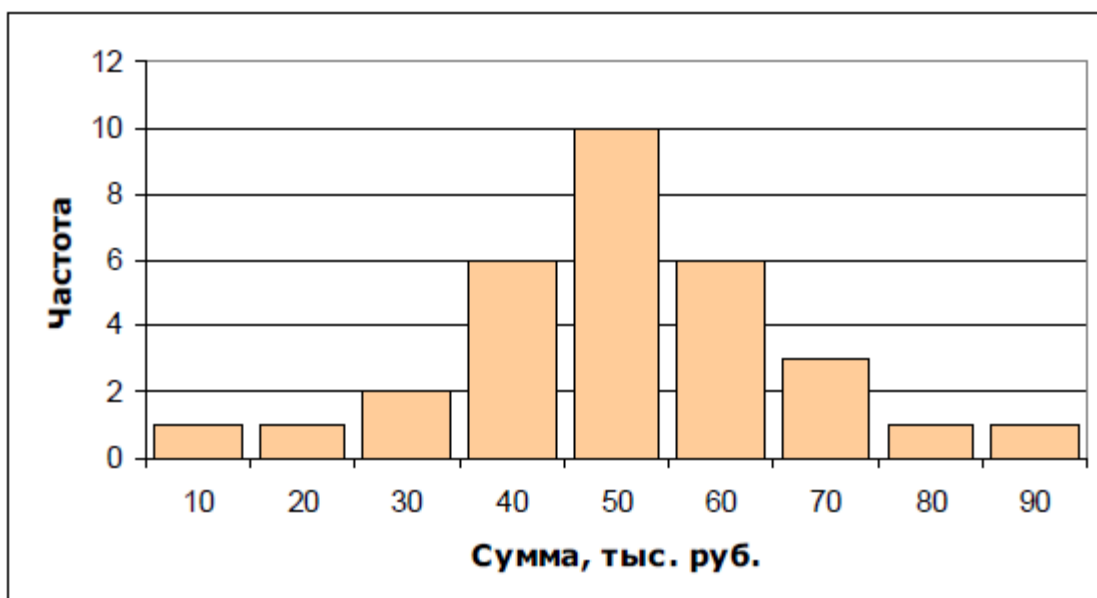


Рисунок 5 – Гистограмма распределения продаж

Иногда используют **нормированную гистограмму**, что позволяет оперировать не значениями наблюдений, а их вероятностями. Для этого каждый элемент гистограммы делится на количество наблюдений, то есть в нашем случае на 31 (число дней в месяце). Рассмотрим рисунок 6.

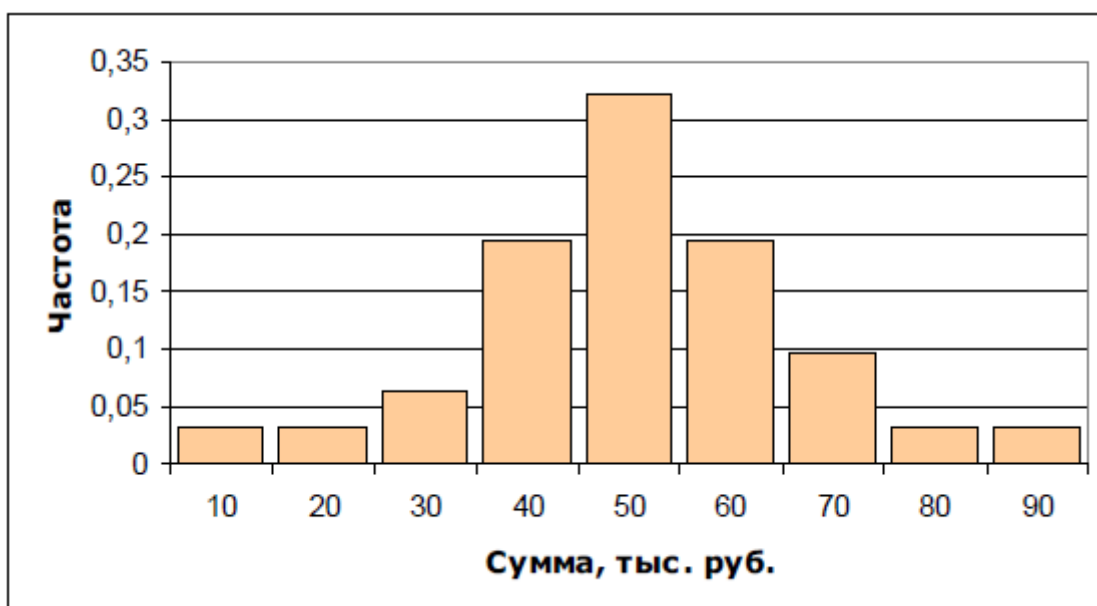


Рисунок 6 – Нормированная гистограмма

Еще одним распространенным средством визуализации, которое входит в состав большинства аналитических приложений, является информация о **статистических характеристиках** исследуемой выборки. Они обычно представляются в табличном виде и указываются для каждого поля выборки. Статистические характеристики позволяют выдвигать

гипотезы о поведении данных и присущих им закономерностях, контролировать результаты обработки данных на различных этапах аналитического процесса. Типичный вид визуализатора статистики представлен в таблице 2.

Таблица 2 – Статистические характеристики

№ п/п	Метка поля	Статистика: Количество значений = 150					
		Минимум	Максимум	Среднее	Стандартное отклонение	Количество уникальных значений	Количество пропусков
1	Длина чашелистика	43,00	79,00	58,43	9,25	–	0
2	Ширина чашелистика	20,00	44,00	30,57	40,34	–	0
3	Длина лепестка	10,00	69,00	37,58	17,59	–	0
4	Ширина лепестка	1,00	25,00	11,99	7,60	–	0
5	Класс цветка	6	9	8	1,14	3	0

### 1.3. OLAP-анализ

Трудности, возникающие при обработке больших и сложных массивов данных, создали предпосылки для появления метода визуализации многомерных табличных данных — OLAP-анализа.

В основе OLAP лежит многомерное представление данных, которые могут быть разделены на количественные и качественные. Качественные данные представляют собой значения, выраженные в категориальной форме. Обычно это наименования товаров, групп товаров, организаций, названия городов, ФИО сотрудников и т. д. С каждым объектом связаны признаки, количественно описывающие его. Для товара это может быть цена, количество или сумма; для города, в котором расположено торговое представительство, — расстояние до него и количество жителей; для сотрудника — заработная плата и стаж работы.

В рамках многомерной модели данные, качественно описывающие исследуемый бизнес-процесс, называются **измерениями**. Измерениями могут быть Товар, Город, Клиент, Организация и др. К ним относят и дату. Данные, количественно описывающие процесс или объект, называются **фактами**. Примеры фактов: Количество, Сумма, Возраст, Доход, Торговая наценка и др. Другими словами, с каждым измерением связаны один или несколько фактов.

Измерения несут смысловую нагрузку, а факты — количественную. Чтобы достоверно отделить измерения от фактов, достаточно сопоставить значение с вопросом. Измерения позволяют ответить на вопросы «Что?» (товар), «Кто?» (клиент), «Когда?» (дата) и «Где?» (город). Факты отвечают на единственный вопрос — «Сколько?».

Чтобы построить OLAP-куб, пользователь должен указать системе следующие параметры:

- какие измерения и факты включать в куб;

– методы агрегации значений фактов.

На рисунке 7 показан пример визуализации данных с помощью OLAP-куба, где представлена информация о количестве и суммах продаж по аптекам в разрезе года и месяца. Эти сведения могут оказаться полезными при определении наиболее загруженных аптек.

Измерения

Факты

Отдел.Наименование

Аптека 1

Аптека 2

Аптека 3

Итого:

- + Год	Месяц	Σ Кол-во	Σ Сумма	Σ Кол-во	Σ Сумма	Σ Кол-во	Σ Сумма	Σ Кол-во	Σ Сумма
2004	01 Январь	573	33 284,1					573	33 284,1
	02 Февраль	623	33 809,2					623	33 809,2
	03 Март	534	32 241,2					534	32 241,2
	04 Апрель	527	33 488,0	353	19 370,6			880	52 858,6
	05 Май	449	22 377,6	224	10 759,3			673	33 136,9
	06 Июнь	425	21 364,1	186	8 160,5			611	29 524,5
	07 Июль	373	13 536,4	164	8 158,1			537	21 694,5
	08 Август	312	14 324,6	227	10 764,9			539	25 089,5
	09 Сентябрь	453	23 436,4	278	15 008,1			731	38 444,6
	10 Октябрь	536	31 328,3	361	21 777,8	566	35 965,4	1 463	89 071,5
	11 Ноябрь	588	33 413,9	281	16 416,8	603	46 399,3	1 472	96 229,9
	12 Декабрь	591	32 596,5	350	21 365,7	655	43 121,3	1 596	97 083,5
	Итого:	5 984	325 200,1	2 424	131 781,8	1 824	125 486,0	10 232	582 468,0
Итого:		5 984	325 200,1	2 424	131 781,8	1 824	125 486,0	10 232	582 468,0

Рисунок 7 – Пример OLAP-куба

#### 1.4. Визуализаторы для оценки качества моделей

Набор визуализаторов для оценки качества моделей следующий:

- матрица классификации;
- диаграмма рассеяния;
- ретропрогноз;
- графики контроля хода обучения.

**Матрица классификации**, или таблица сопряженности (confusion matrix), применяется для оценки качества классификационных моделей.

Для каждого примера классификационная модель формирует на выходе метку класса, к которой относится объект с набором признаков, указанных в примере. Если метка класса, сформированная моделью, совпадает с целевой меткой класса из примера, то такой пример (объект) является правильно распознанным, в противном случае — неправильно распознанным.

Фрагмент классификационной выборки представлен в таблице 3, где имеется четыре признака: Длина чашелистика, Ширина чашелистика, Длина лепестка и Ширина лепестка. На основе значений данных признаков модель



должна отнести каждый предъявленный экземпляр цветка к одному из трех видов (классов) — Versicolor, Virginica или Setosa (данная задача классификации известна как «ирисы Фишера»).

Таблица 3 – Ирисы Фишера

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Класс цветка	Класс цветка (модель)
50	33	14	2	Setosa	Setosa
64	28	56	22	Virginica	Virginica
65	28	46	15	Versicolor	Versicolor
67	31	56	24	Virginica	Virginica
63	28	51	15	Virginica	Versicolor
46	34	14	3	Setosa	Setosa
69	31	51	23	Virginica	Virginica
62	22	45	15	Versicolor	Versicolor
59	32	48	18	Versicolor	Virginica
46	36	10	2	Setosa	Setosa
61	30	46	14	Versicolor	Versicolor
60	27	51	16	Versicolor	Versicolor

Поле Класс цветка является целевым и содержит эталонную метку класса, по которой производится настройка параметров модели. Поле Класс цветка (модель) содержит метку класса, выданную моделью на данном примере. В таблице 3 выделены примеры, в которых модель допустила ошибку, выдав метку класса, не соответствующую эталонной.

Соотношение числа правильно и неправильно распознанных объектов служит критерием качества модели. Чем больше объектов выборки было распознано правильно, тем лучше работает модель.

Визуализатором, который применяется для оценки качества моделей в случае непрерывной выходной переменной, является **диаграмма рассеяния**. Диаграмма рассеяния представляет собой график, по одной оси которого откладываются целевые значения выходной переменной (то есть те, которые заданы в качестве эталона для обучения), а по другой — реальные значения, полученные на выходе модели. На таком графике можно построить линию идеальных значений:  $y = y'$ . На этой линии будет лежать любая точка, для которой реальное выходное значение  $y$ , сформированное моделью, будет равно целевому значению  $y'$  (при этом ошибка  $E = y - y' = 0$ ). Такая линия, очевидно, будет выходить из начала координат и пересекать координатную плоскость по диагонали.

В то же время любая точка, для которой реальное выходное значение будет отлично от эталонного целевого значения, отклонится от линии

идеальных значений. При этом величина отклонения будет равна ошибке, допущенной моделью на данном примере (рисунок 8).

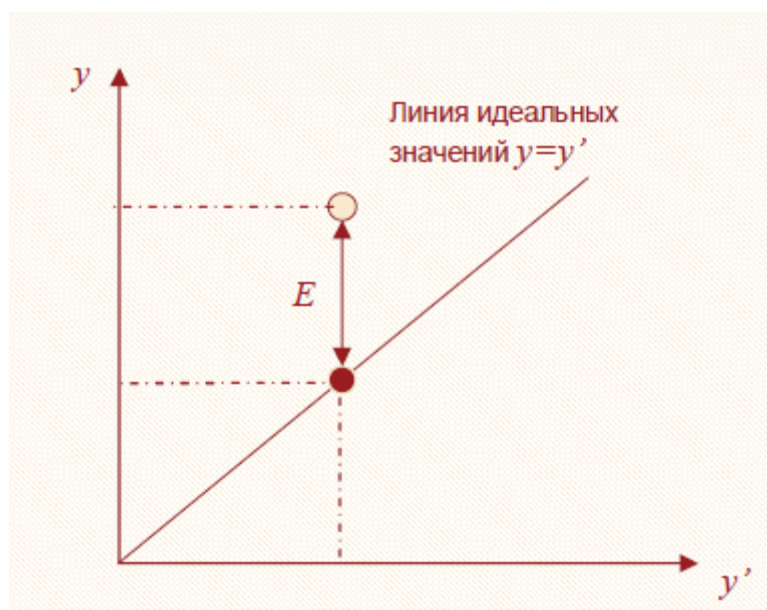


Рисунок 8 – Принцип построения диаграммы рассеяния

Точка, лежащая на линии идеальных значений, соответствует случаю, когда реальный выход модели равен эталонному. На практике модель допускает ошибку  $E$ , которая приводит к отклонению точки, соответствующей реальному выходу модели, от эталонного и, следовательно, от линии идеальных значений. Точка, соответствующая реальному выходу, имеет координаты  $y' = y + E$ .

Пример диаграммы рассеяния представлен на рисунке 9. Диагональная линия на рисунке — это линия идеальных значений. Точками, рассеянными вдоль линии идеальных значений, обозначены реальные выходные значения модели.

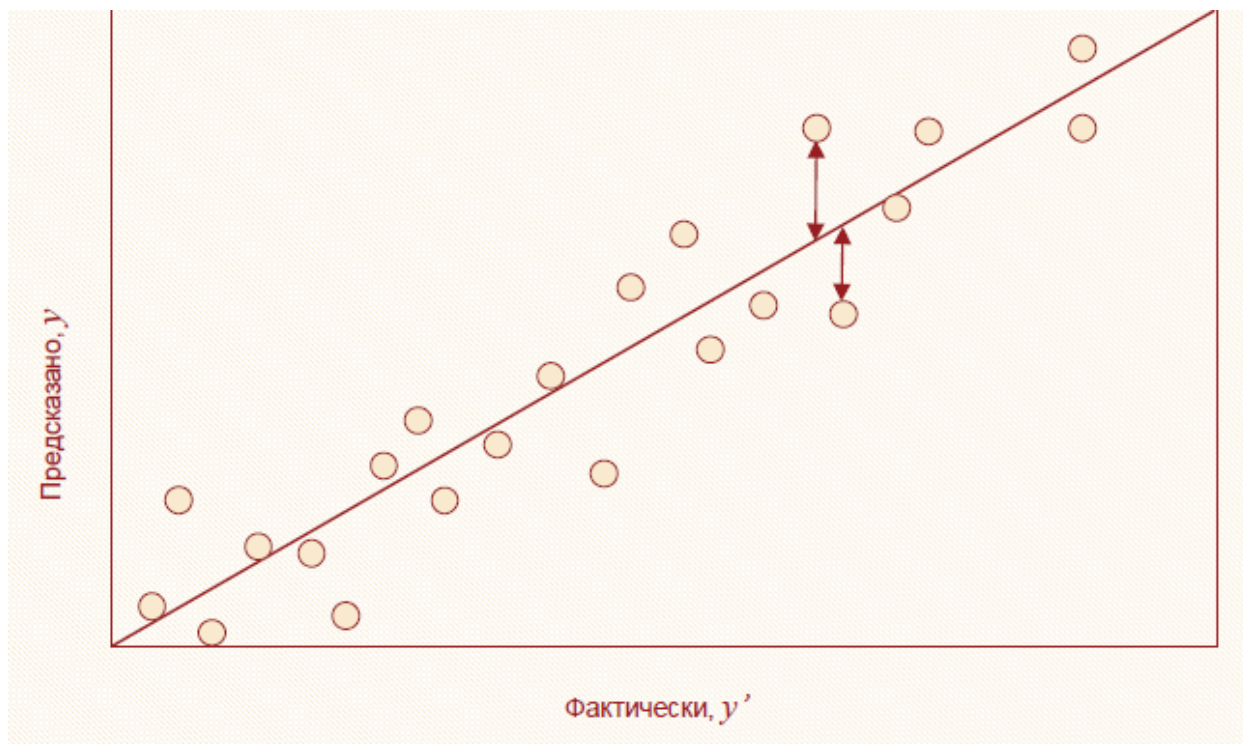


Рисунок 9 – Диаграмма рассеяния

Если все точки (или хотя бы основная масса), представляющие реальные выходные значения модели, сосредоточены вблизи линии идеальных значений, то модель работает хорошо.

Частный случай задачи регрессии — прогнозирование временных рядов. Чтобы составить прогноз, строится регрессионная модель, которая на основе прошлых значений ряда рассчитывает прогнозируемые значения. Она имеет некоторый набор параметров, позволяющих получить прогноз с учетом поведения ряда в прошлом.

Для построения **ретропрогноза** нужно выбрать некоторое подмножество данных из прошлого, чтобы использовать их в качестве исходных. Затем к этому подмножеству применяется прогностическая модель с заданными параметрами. Модель формирует набор прогнозных значений, которые затем сравниваются с данными, реально имевшими место в прошлом. Если в результате такого сравнения обнаружится, что между значениями ретропрогноза и реальными данными существует большое расхождение, это даст повод усомниться в корректности прогностической модели. Сам характер расхождения (его величина, знак и т. д.) позволяет выработать методику коррекции модели.

Принцип работы ретропрогноза поясняется с помощью диаграммы на рисунке 10.

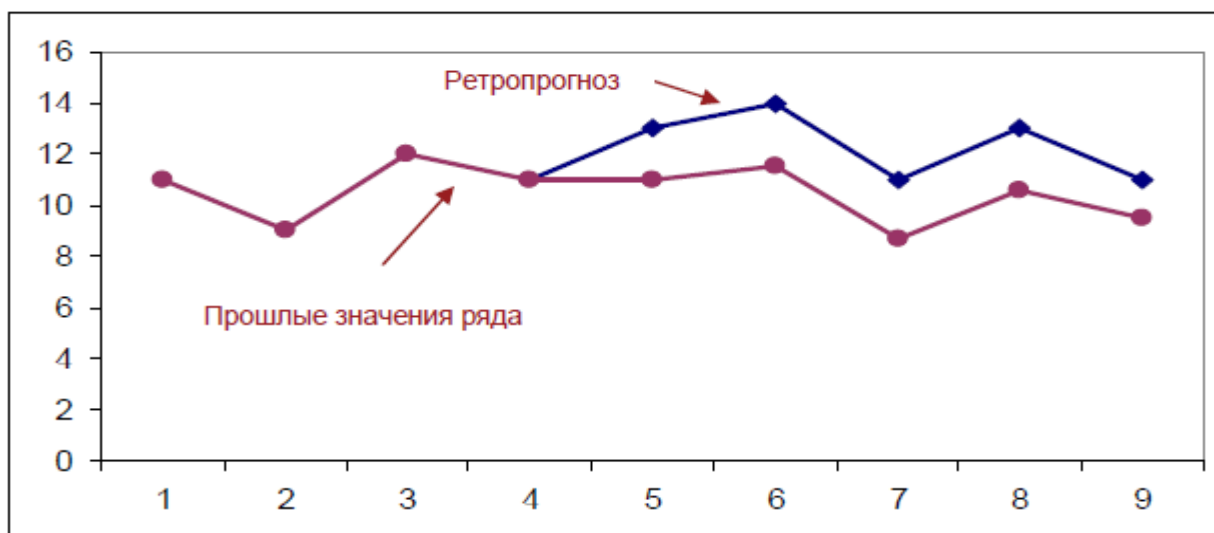


Рисунок 10 - Ретропрогноз

Ретропрогноз направлен не в будущее ряда, а формируется параллельно его прошлым значениям и сравнивается с ними.

В корпоративных аналитических системах широко используются модели, основанные на **машинном обучении**, например нейронные сети, деревья решений и т.д. В процессе настройки своих параметров (или обучении) модель приобретает свойства, которые обеспечивают выполнение ею требуемого преобразования данных.

При обучении модели большой интерес представляет наблюдение за изменением выходной ошибки на каждой итерации. По изменению ошибки можно сделать вывод о том, как идет процесс обучения, есть ли смысл его продолжать и каковы его предполагаемые результаты.

Реализация такого визуализатора представлена на рисунке 11.

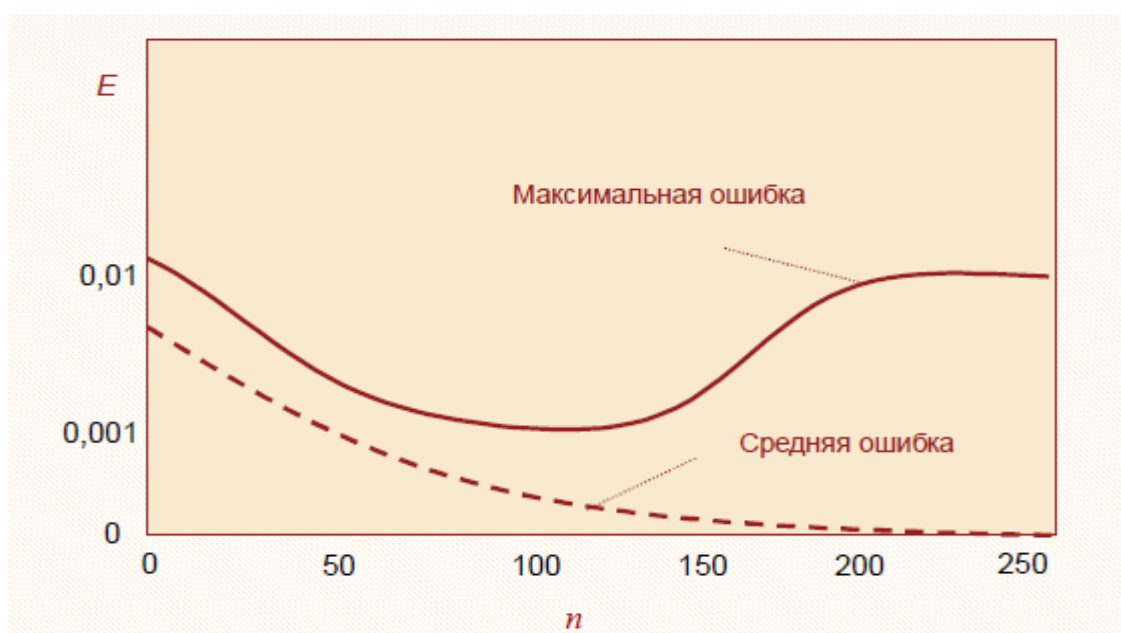


Рисунок 11 – Визуализатор наблюдения за ошибками в ходе обучения (n – номер итерации)

### 1.5. Визуализаторы для интерпретации результатов анализа

Существуют два основных способа описания данных. Первый заключается в построении иерархической структуры данных, то есть в разбиении данных на классы, подклассы и т.д., второй — в сокращении размерности исходного пространства данных. Наиболее распространенные типы визуализаторов, применяемых для интерпретации результатов анализа:

- древовидные визуализаторы;
- визуализаторы связей;
- карты.

**Деревья.** Одной из наиболее популярных классификационных моделей являются деревья решений. Они представляют собой иерархическую последовательность правил вида «если...то...».

Для интерпретации результатов классификации с помощью дерева решений используются визуализаторы, показывающие структуру дерева, а также сформулированные в нем правила.

Пример дерева решений для классификационной задачи об ирисах Фишера представлен на рисунке 12. Для компактности классификационные признаки цветков обозначены следующим образом: SL — длина лепестка, SW — ширина лепестка, CL — длина чашелистика, CW — ширина чашелистика. При таком способе визуализации дерева решений узлы и листья представлены в виде прямоугольников и расположены в порядке их иерархической подчиненности. В узле указывается действующее в нем правило.

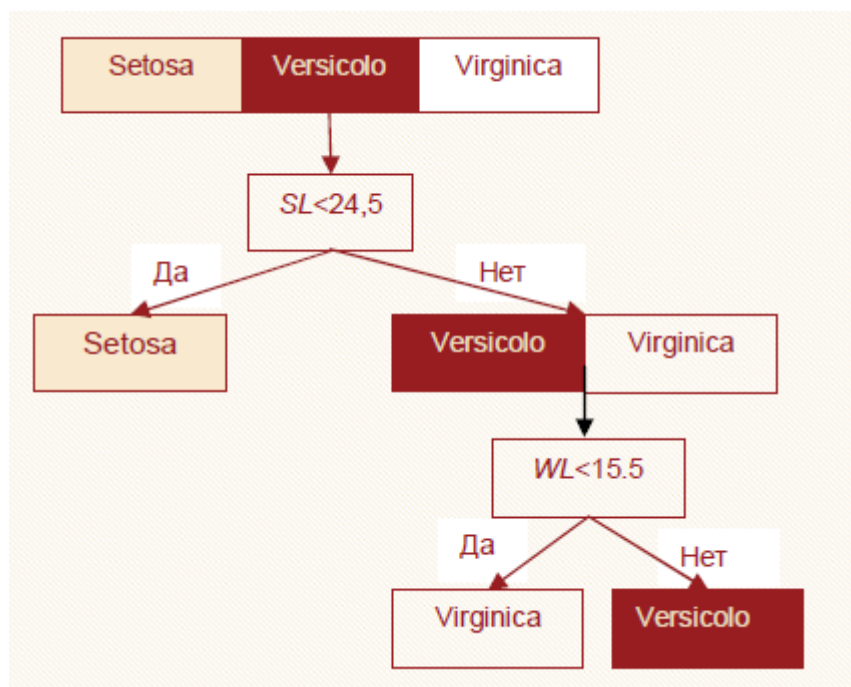


Рисунок 12 – Дерево решений

Для визуализации результатов иерархической кластеризации используют специальную разновидность диаграмм, называемую



**дендрограммой.** Дендрограмма показывает степень близости отдельных объектов и кластеров, а также наглядно представляет в графическом виде последовательность их объединения или разделения. Количество уровней дендрограммы соответствует числу шагов слияния или разделения кластеров. Пример дендрограммы представлен на рисунке 13. В нижней части рисунка расположена шкала, на которой откладывается расстояние между объектами в пространстве признаков.

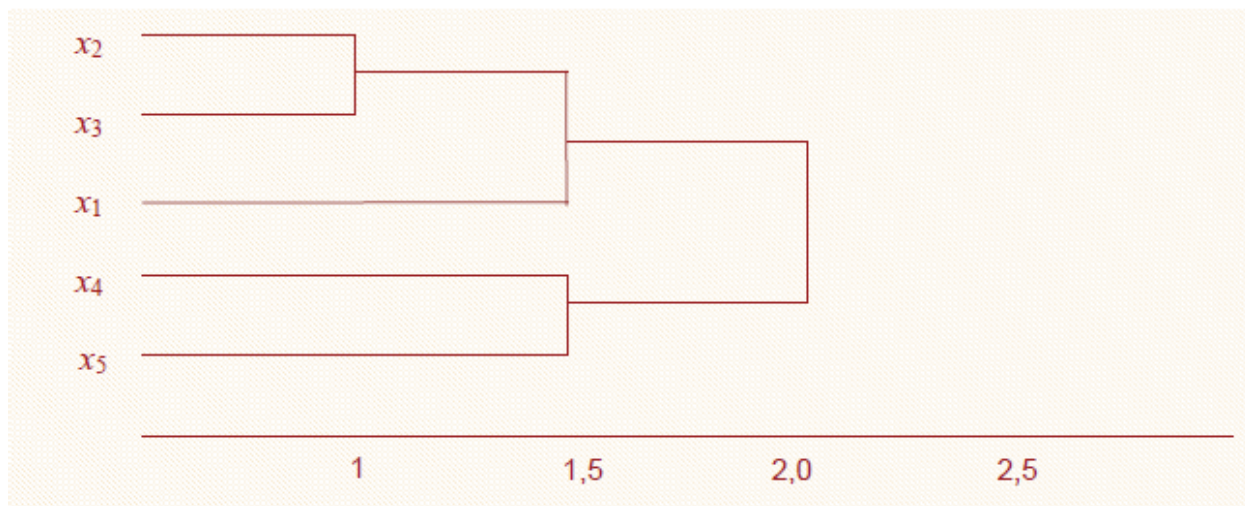


Рисунок 13 - Дендограмма

Нередко требуется исследовать характер и степень взаимной зависимости между различными объектами. Для анализа можно использовать **визуализацию связей**, когда объекты представляются в виде некоторых значков, а связи между ними — в виде линий, соединяющих соответствующие значки. При этом сила связи, то есть степень взаимной зависимости объектов, может показываться различными способами.

Объекты сравниваются между собой на основе какого-либо признака. Если признаков несколько, выбирается способ расчета критерия связи. Чаще всего это метрика — правило вычисления расстояния в многомерном пространстве признаков — евклидово расстояние, расстояние Манхэттена и др.

Пример диаграммы связей между несколькими регионами РФ представлен на рисунке 14. В качестве признаков используются:

- численность постоянного населения;
- удельный вес городского населения;
- зафиксированное за последний период число умерших на 1000 человек.

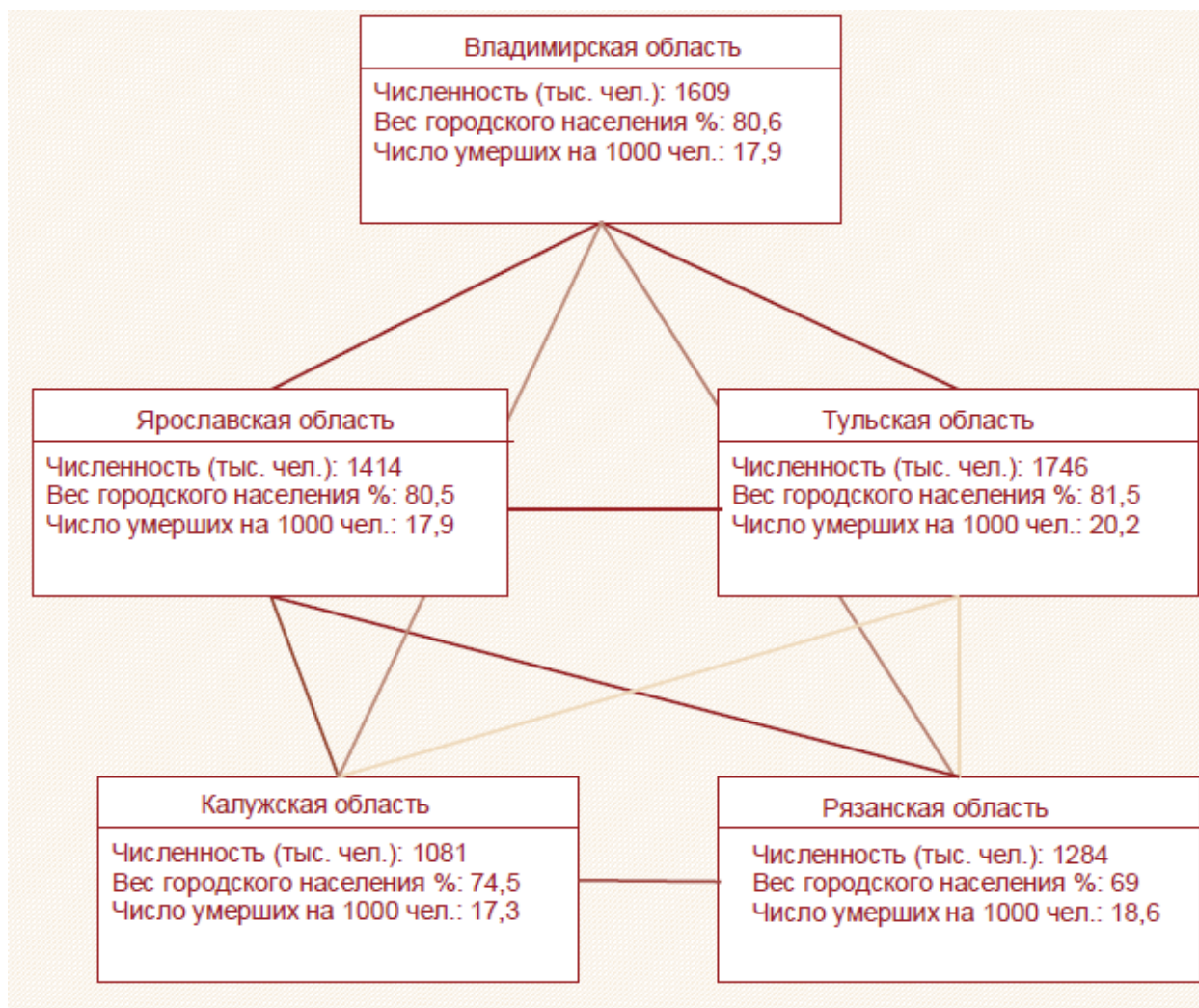


Рисунок 14 – Пример диаграммы связей между регионами РФ

Критерий связи рассчитывался с использованием манхэттенновского расстояния. Видно, что Владимирская и Ярославская области очень близки по сравниваемым признакам (расстояние равно 0,85), а между Калужской и Тульской областями, наоборот, различие самое большое (расстояние — 0,14).

**Карты** позволяют наглядно представить данные, связанные с географическим расположением исследуемых объектов и процессов. Это могут быть демографические данные (например, распределение показателей смертности или рождаемости по различным регионам), а также данные, отражающие миграцию населения и рабочей силы, обеспечение регионов энергоносителями, динамику распространения эпидемий и т.д. В бизнес-аналитике это информация, связанная с уровнем потребления в регионах, характером спроса и предложения по различным видам товаров, сведения о продажах, осуществляемых региональными дилерами, о логистических и транспортных потоках и т.д.

В таблице 4 содержатся данные об отгрузках продукции, произведенных компаниям мелким оптовикам из различных субъектов Центрального федерального округа (ЦФО).

Таблица 4 – Объем продаж по субъектам ЦФО РФ

№ п/п	Область	Продажи, млн. руб.	№ п/п	Область	Продажи, млн. руб.
1	Белгородская	162	10	г. Москва	664
2	Брянская	61	11	Московская	328
3	Владимирская	54	12	Орловская	54
4	Воронежская	154	13	Рязанская	151
5	Ивановская	209	14	Смоленская	137
6	Калужская	117	15	Тамбовская	29
7	Костромская	90	16	Тверская	178
8	Курская	57	17	Тульская	186
9	Липецкая	50	18	Ярославская	120

Имея карту ЦФО, можно в соответствии с палитрой раскрасить области и получить сравнительный отчет о продажах по регионам (рисунок 15).

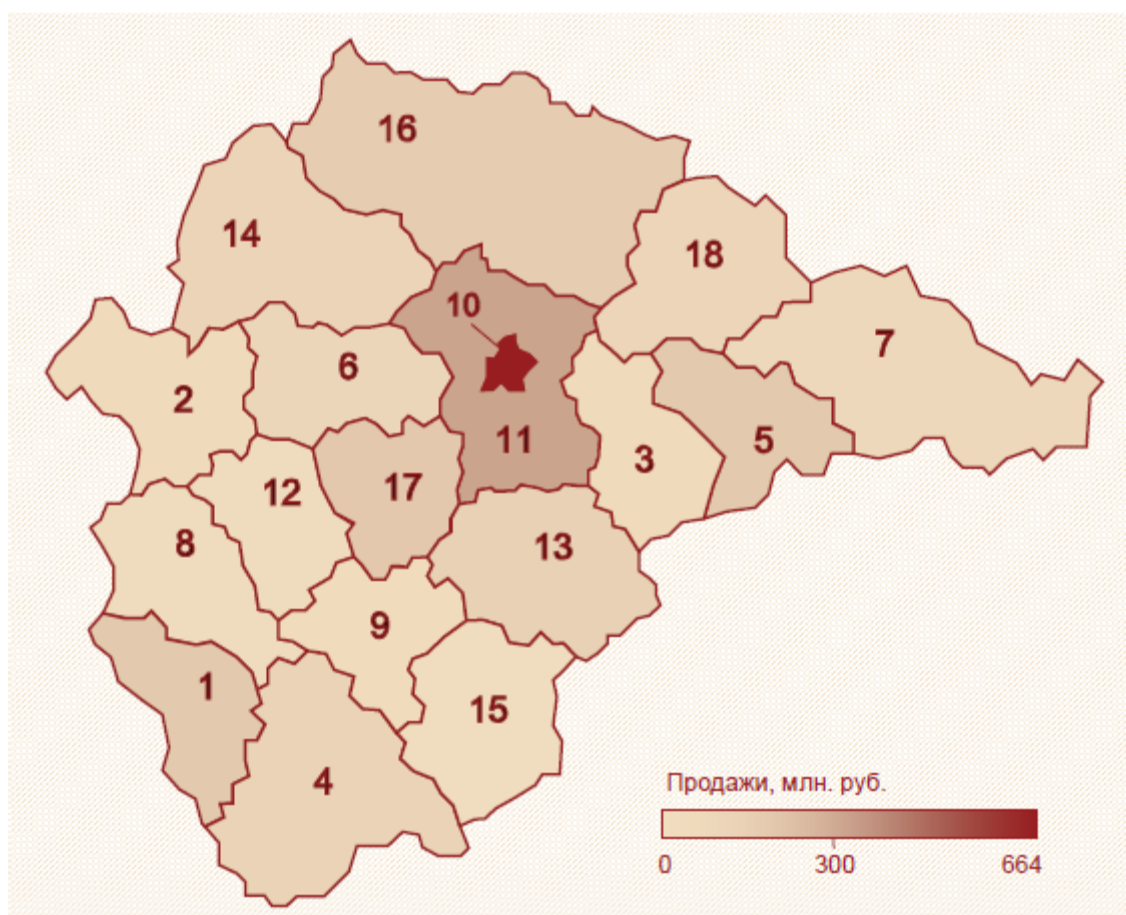


Рисунок 15 – Отчет о продажах по регионам

Видно, что максимальные значения продаж приходятся на Москву и Московскую область, а минимальные — на Тамбовскую. При этом может



использоваться многомерное представление данных. Например, если требуется показать на карте не только суммы продаж, но и прибыль, то последнюю можно отобразить рельефным выделением соответствующей области на карте. Также на карты можно наносить значки, форма, цвет, размер и взаимное положение которых соответствуют свойствам исследуемых объектов.

На рисунке 16 можно наблюдать пример тепловой карты, построенной по двум измерениям — Топливо и Город. Цвет каждой прямоугольной ячейки карты формируется на основе цены топлива: чем темнее оттенок, тем она выше.

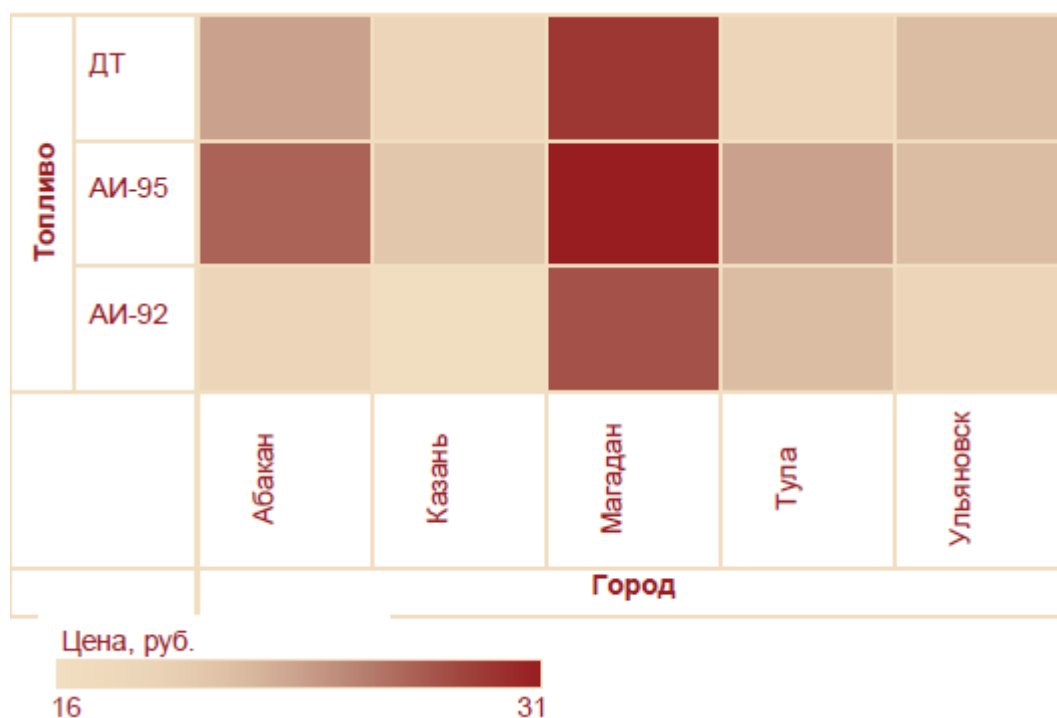


Рисунок 16 – Пример тепловой карты

### Задания для самостоятельной работы:

1. Требуется разработать подсистему аналитической отчетности в Deductor (розничный банковский бизнес). Система отчетности строится на основе прилагаемого к задаче хранилища данных БАНК (BANK.GDB).
2. Все требуемые отчеты должны быть вынесены на Панель Отчетов и разделены на три папки:
  1. Задание 1: Графики;
  2. Задание 2: OLAP отчеты;
  3. Задание 3: Вычисления в кубе.

### Задание 1

1. Постройте **гистограмму** распределения заявленных сумм по потребительским кредитам.

2 Постройте **многомерную диаграмму** зависимости количества договоров отсрочки и суммы кредита. Предварительно квантуйте суммы кредитов на интервалы длиной в 50 тыс. рублей (до 50 тыс., 50-100 тыс. и т.д.).

3 Постройте **временной ряд** количества обращений по кредитам в сентябре (по оси Y откладывается количество заявок).

4 Постройте **диаграмму размещения**, на которой иллюстрируется зависимость количества выданных кредитов и средней суммы по ним от типа кредита.

## **Задание 2**

Постройте следующие OLAP-отчеты:

1 Количество и процентное соотношение Одобренных/Отклоненных на каждом из этапов заявок (измерение Статус заявки), а также кросс-диаграмму в виде круговой диаграммы, иллюстрирующую данное соотношение.

2 Количество одобренных/отклоненных заявок по каждому филиалу и отделению на основе всех имеющихся данных.

3 10 самых злостных неплательщиков по состоянию на 10.11.2009.

4 Договора, на которые приходится 50% суммы от всех выданных в августе кредитов.

5 Среднюю интенсивность поступления заявок по дням недели каждого из отделений банка на основании всех имеющихся данных. Отчет дополните соответствующей кросс-диаграммой.

## **Задание 3**

Импортируйте данные из процесса **Качество обслуживания долга**, который содержит факт Суммарное количество дней просрочки с момента заключения договора по состоянию на расчетную дату.

Используя обработчик Скользящее окно, перейдите от рассмотрения суммарного количества дней просрочки к величине Просрочка за месяц (по сути, это прирост суммарной просрочки по сравнению с предыдущим месяцем). Добавьте калькулятором в сценарии новый столбец Наличие просрочки со значениями:

- True – при наличии месячной просрочки;
- False – просрочка отсутствует.

После того как будет подготовлен такой набор данных, средствами вычисляемого факта в OLAP-кубе каждому договору, который имеет просрочки, проставьте метки на основании следующего правила:

1. Длительные просрочки – суммарное количество просроченных дней по договору за все месяцы больше 10. Иначе:

2. Регулярные просрочки – имеется 2 и более случаев просрочки;

3. 1 просрочка – имеется ровно один случай возникновения просроченной задолженности.