**ВВЕДЕНИЕ**

В данной лабораторной работе необходимо изучить возможности прикладного программного обеспечения Excel для организации данных и их анализа путём создания реляционной базы данных выбранной транзакции.

**ПРЕДМЕТНАЯ ОБЛАСТЬ**

Предметная область: прокат экипировки для катания на лыжах и сноуборде.

На горнолыжных курортах России, Московской и Ленинградской областей, расположены пункты проката, осуществляющие аренду необходимой экипировки отдыхающим. В основной базе данных отображается дата, место и вид предоставленной экипировки, а также информация о стоимости услуги и сроке аренды для того, чтобы можно было определить выручку прокатного пункта за определённый срок. Кроме того, база данных содержит информацию о количестве оборудования, остающегося на складе на момент аренды, для последующего анализа этих значений и, в случае необходимости, восполнения недостающих предметов, пользующихся спросом у клиентов. Более того, фиксируется ID клиента для выявления постоянных арендаторов и предоставления им определённых льгот на услуги пункта.

**АТРИБУТЫ И СУЩНОСТИ**

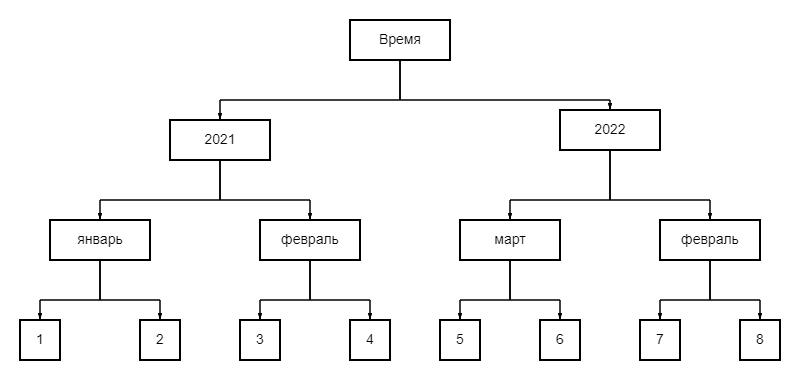
Первой сущностью является время осуществления транзакции, которая включает в себя такие атрибуты как год (числовое значение), месяц (текстовое значение), дата (числовое значение) (Рис.1).

Рис.1 Графический пример объекта «Время»

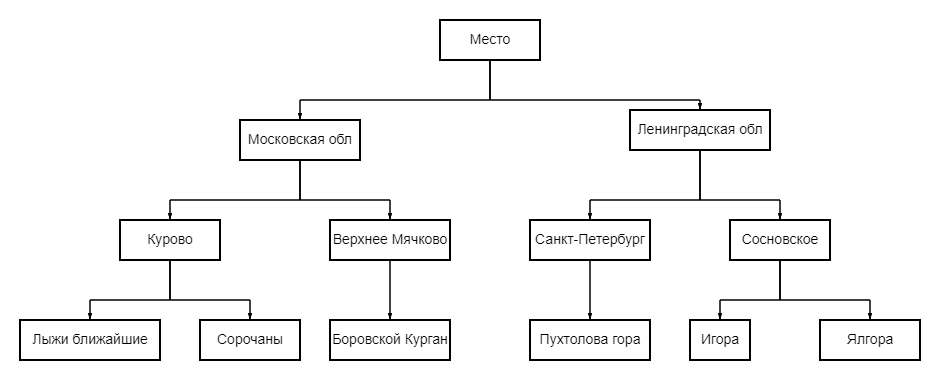
Второй сущностью является место транзакции со следующими атрибутами: область (текстовое значение), населённый пункт (текстовое значение) и название горнолыжного курорта (текстовое значение). (Рис.2)

Рис.2 Графический пример объекта «Место»

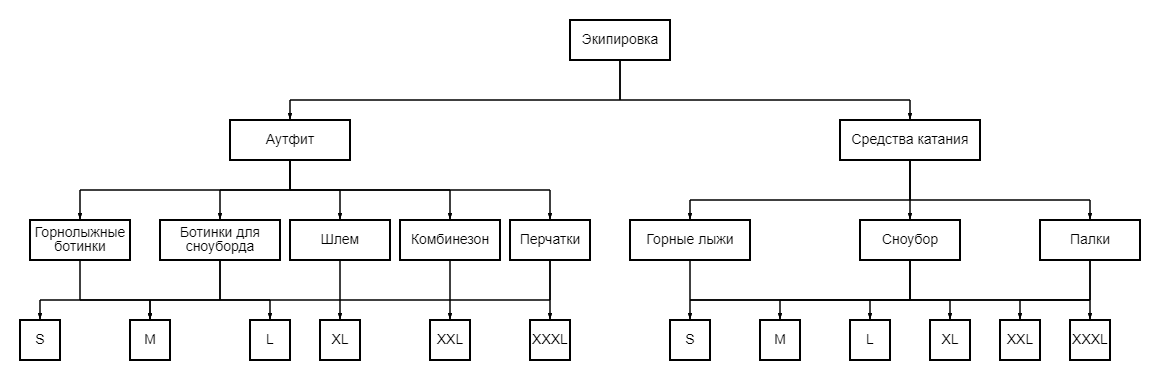
Третья сущность – предмет проката с атрибутами: тип экипировки (текстовое значение), название (текстовое значение) и размер (текстовое значение). (Рис.3)

Рис.3 Графический пример объекта «Экипировка»

**ВИЗУАЛЬНОЕ ОПИСАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ**

**СРУКТУРА СПИСКА**

Изучив предметную область и определив задачи для которых необходима реляционная база данных определяем названия полей для будущих записей: год, месяц, дата, область, населённый пункт, название курорта, тип экипировки, наименование, размер, залог, стоимость, срок аренды, остаток на складе, ID клиента и статус возврата.

**РАСКРЫВАЮЩИЕСЯ СПИСКИ**

Для удобства заполнения таблицы разумнее использовать раскрывающиеся списки со следующими диапазонами значений (Рис.4):

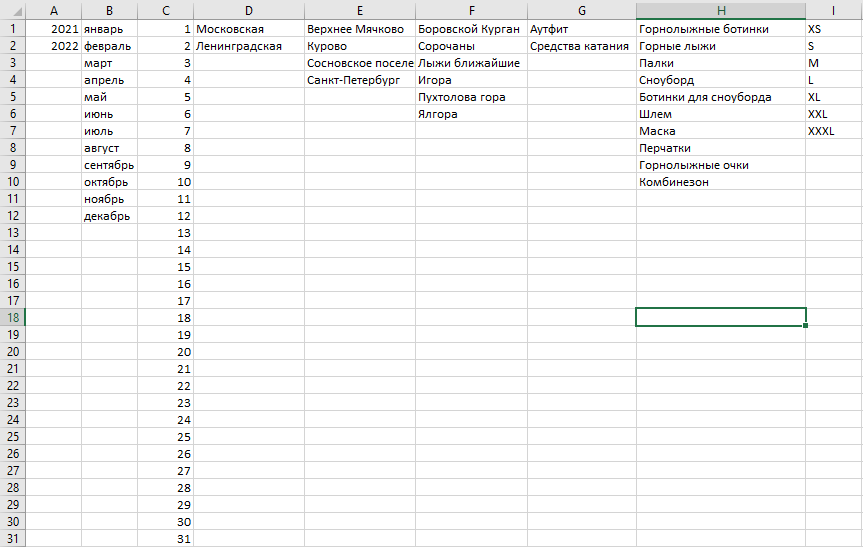


Рис.4 Диапазон значений раскрывающихся списков

**ЗАПОЛНЕННАЯ БАЗА ДАННЫХ**

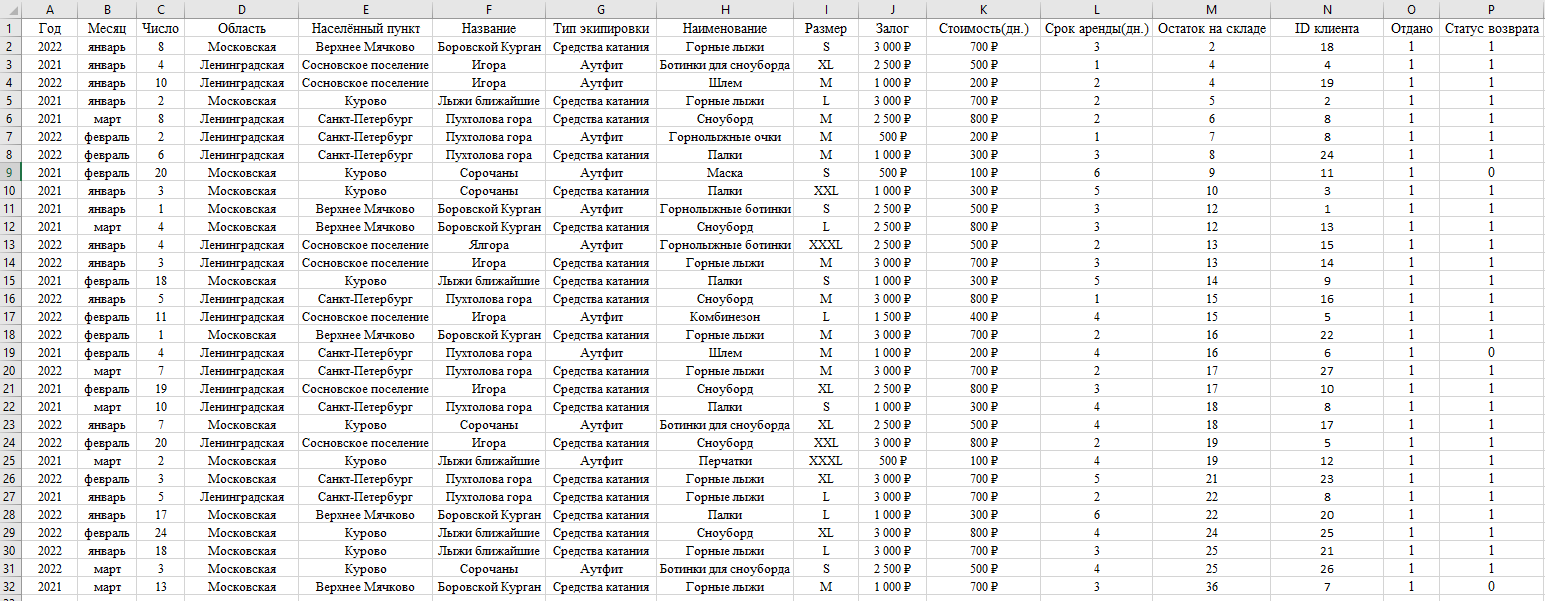
 Добавим 31 запись в базу данных, соответствующую выбранной предметной области (Рис.5).

Рис.5 Заполненная БД

**МАНИПУЛЯЦИИ С ТАБЛИЦЕЙ**

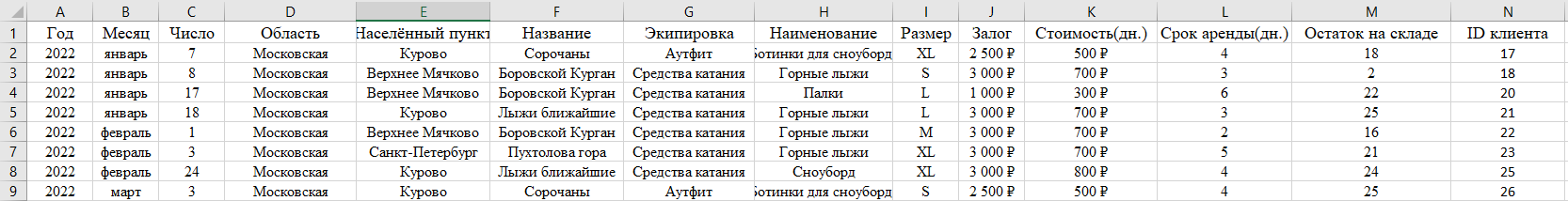
Выберем из таблицы все записи, хранящие информацию об услугах 2022 года в Московской области (Рис.6).

Рис.6 Выборка по «2022» и «Московская»

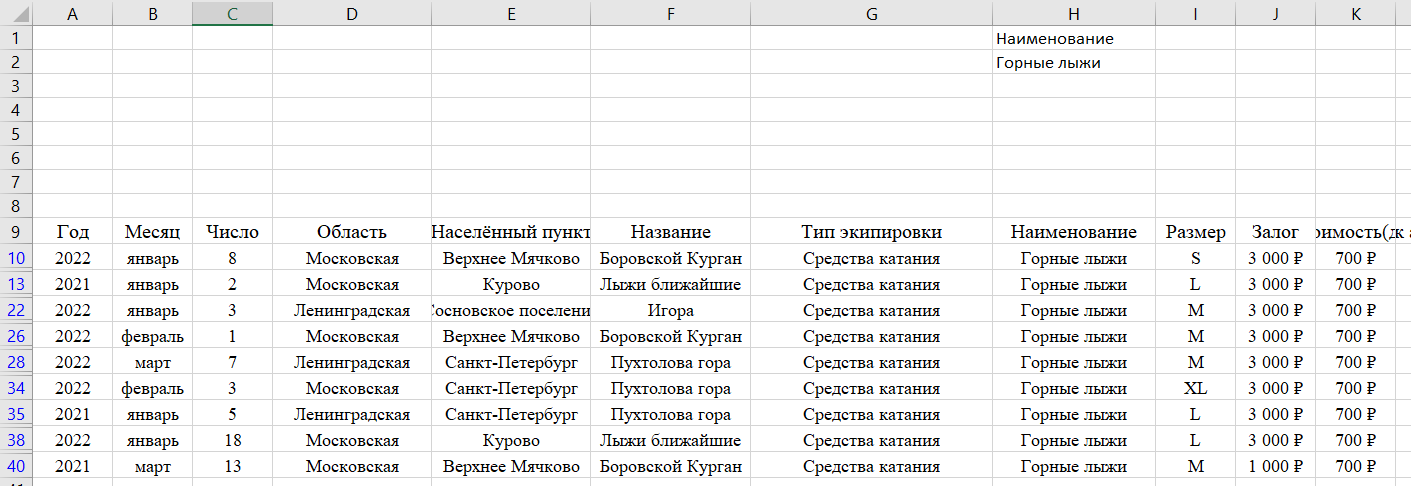
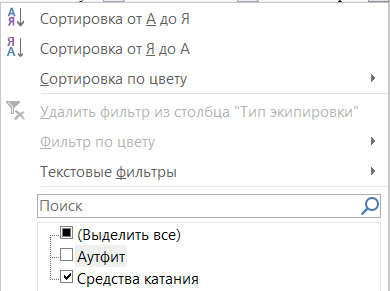
Применим расширенный фильтр для Наименования (Рис.7).

Рис.7 Расширенный фильтр по «Горные лыжи»

Применим автофильтр для типа экипировки «Средства катания»



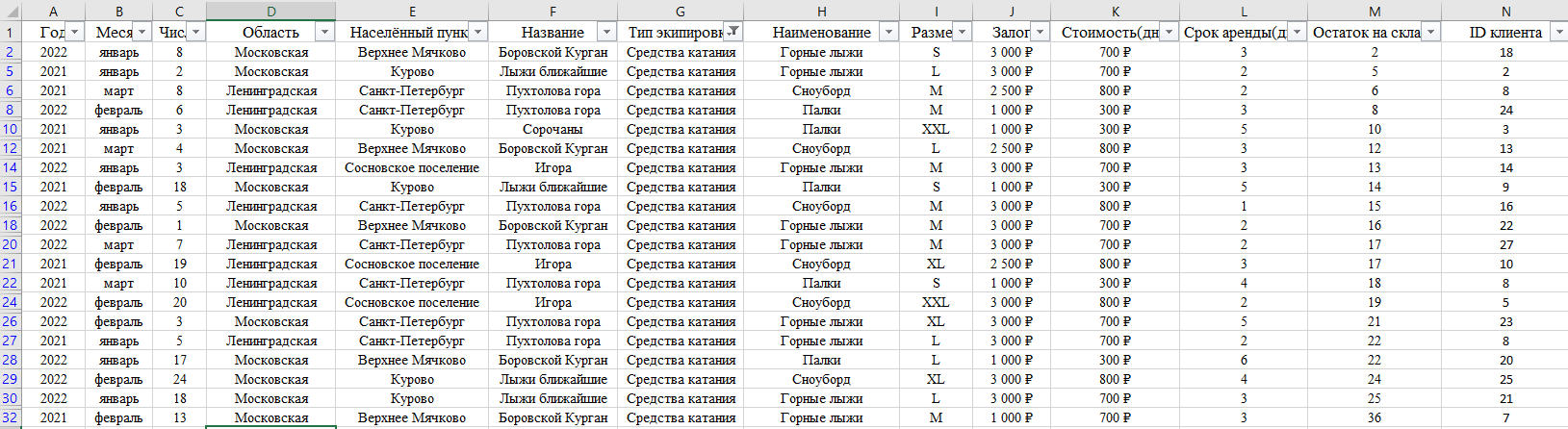
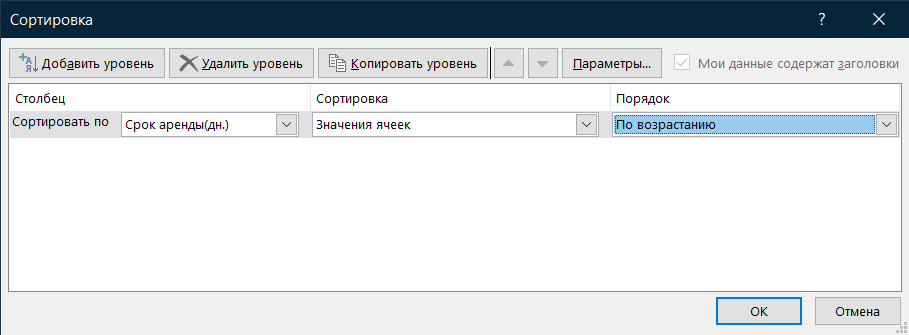
Рис.8 Условие автофильтра

Рис.9 Результат автофильтра

Осуществим сортировку записей по сроку аренды.

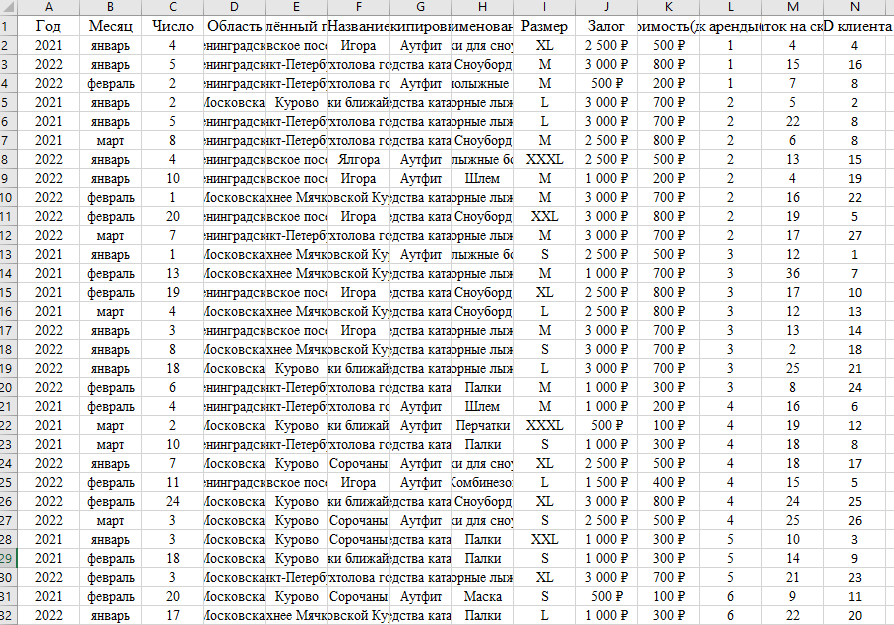
Рис.10 Условие сортировки

Рис. 11 Результат сортировки

**Лабораторная работа №2**

Проанализировав таблицу из первой лабораторной работы, определим необходимые атрибуты и факты для текущей базы данных (рис. 12).

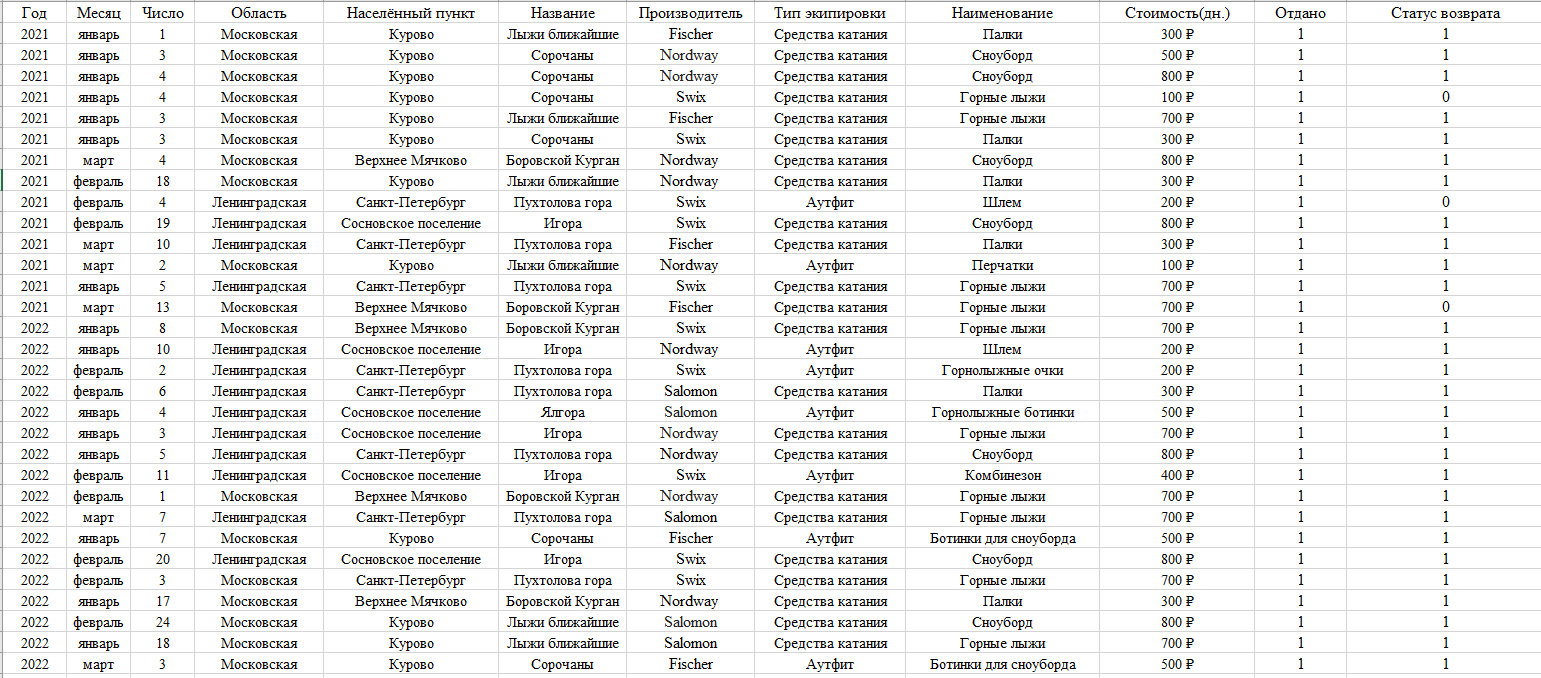


Рис. 12. База данных

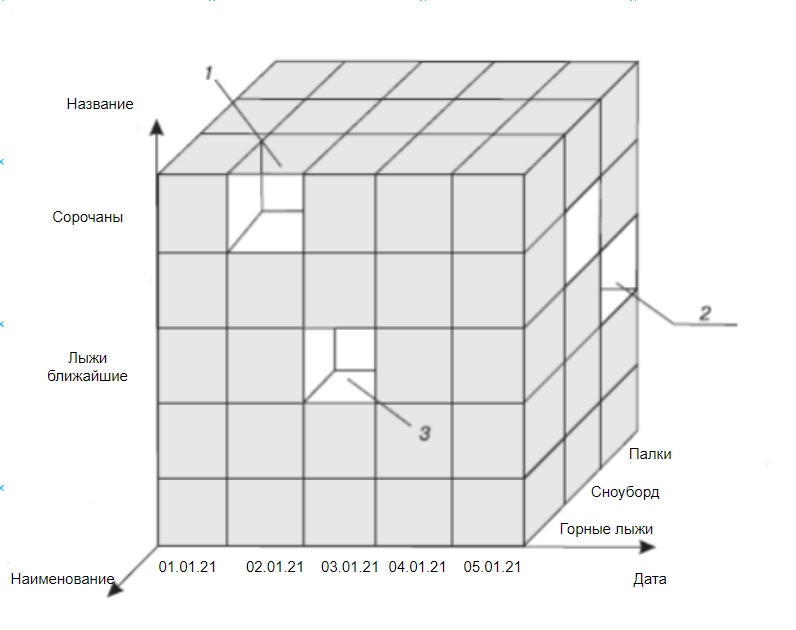
Спроектируем OLAP-куб на основе данных БД согласно нижнему уровню иерархий (рис. 13).

Рис. 13. OLAP-куб

* Измерение «Объект»: Производитель/ Тип / Наименование.

• Измерение «Время»: Год/ Месяц / День.

• Измерение «Место»: Область / Населённый пункт / Название.

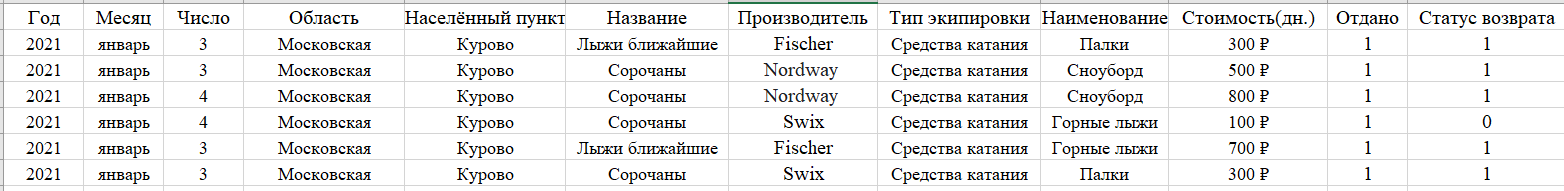
Продемонстрируем связь между кубом и таблицей.

Рис. 14. Связь куба с таблицей

Приведём пример одного простого среза по дате (рис. 15).

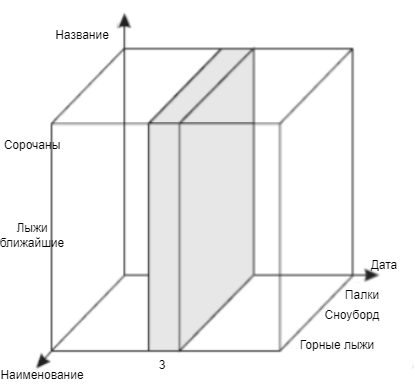


Рис. 15. Срез по дате

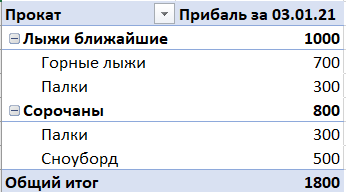
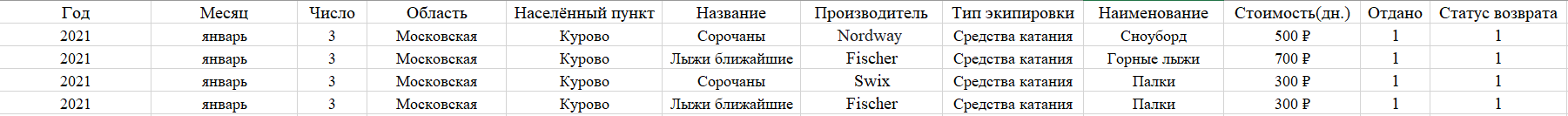


Рис. 16. Сводная таблица

Рис. 17. Связь между кубом и таблицей

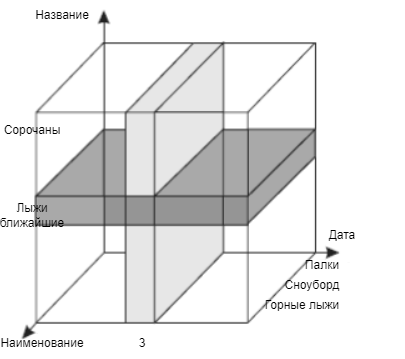
Приведём пример двойного среза (по дате и названию курорта).

Рис. 18. Двойной срез

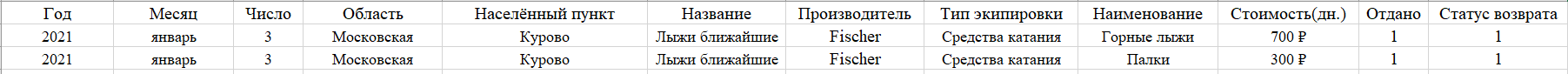


Рис. 19. Связь куба с таблицей.

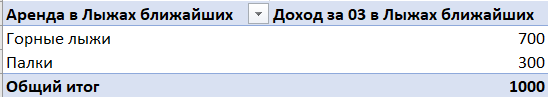


Рис. 20. Сводная таблица

Произведём транспонирование куба.

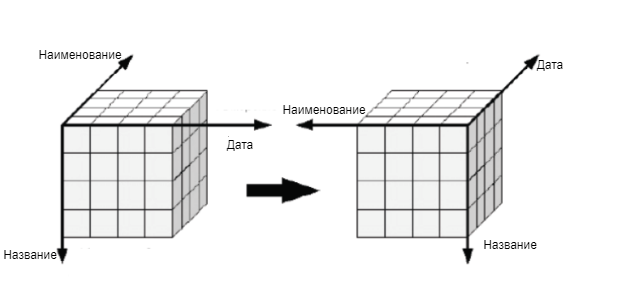


Рис. 21. Транспонирование

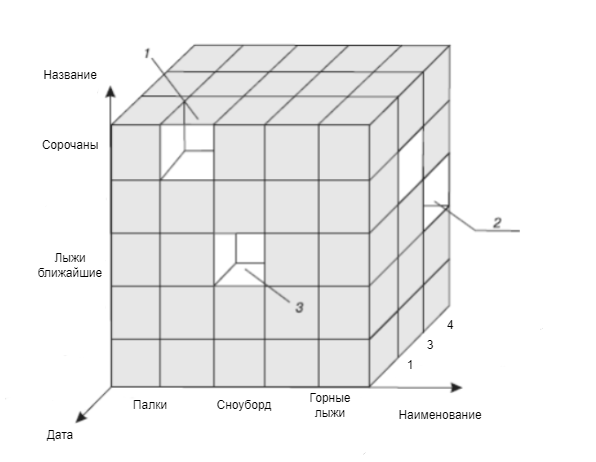


Рис. 22. Транспонированный OLAP-куб

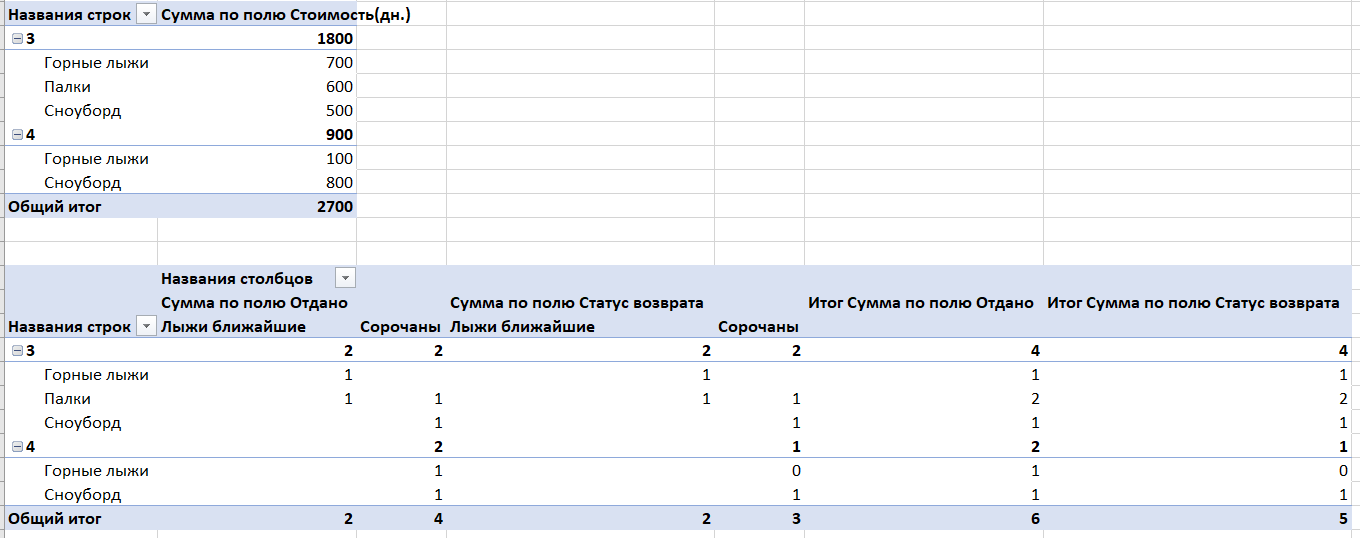


Рис. 23. Сводные таблицы.

Произведём свёртку куба.



Рис. 24. Свёртка куба



Рис. 25. Связь таблицы с кубом

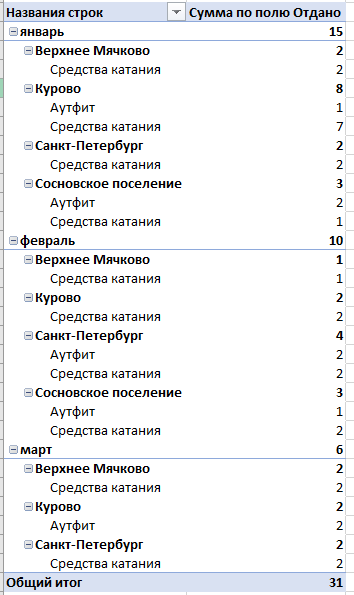


Рис. 26. Сводная таблица

Построим OLAP-куб, опираясь на верхние уровни иерархий.

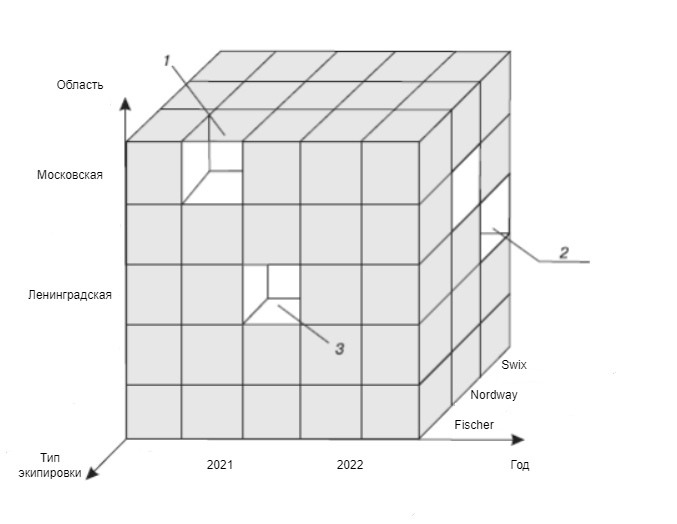


Рис. 27. OLAP-куб

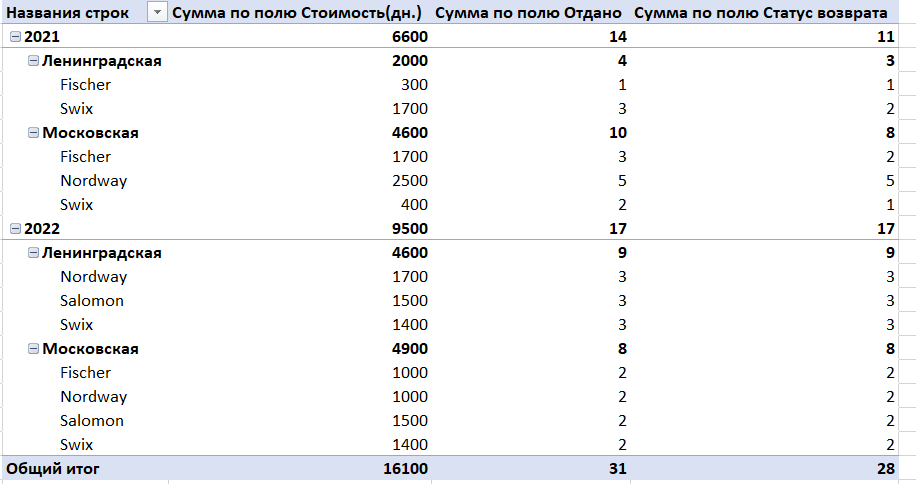


Рис. 28. Сводная таблица

**Лабораторная работа №3**

Модернизируем таблицу из предыдущей лабораторной работы для текущих задач. Будем проверять гипотезу зависимости утраченной экипировки от времени.

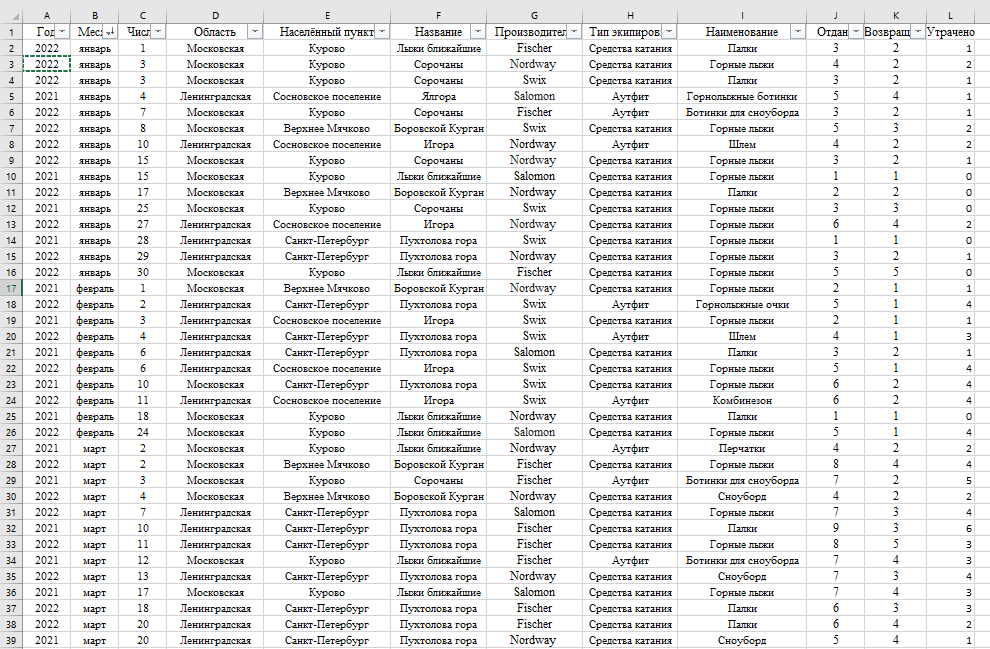
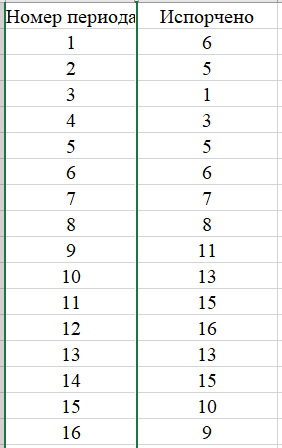


Рис. 29. База данных

Гипотеза состоит в том, что во время зимнего сезона, с каждым днем лыжная экипировка изнашивается, что ведёт к её поломке или появлению дефектов во время использования.

Независимой переменной (X) будет являться временной период, а зависимой (Y) – число испорченных единиц экипировки.

Создадим таблицу с данными, включёнными в регрессионную модель.



Построим диаграмму рассеивания для выбранной пары переменных.

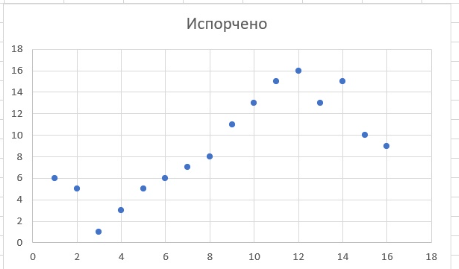


Рис. 30. Диаграмма рассеивания

На основе диаграммы можно сделать вывод, что формой связи будет положительная регрессия.

Добавим линию тренда на диаграмму.

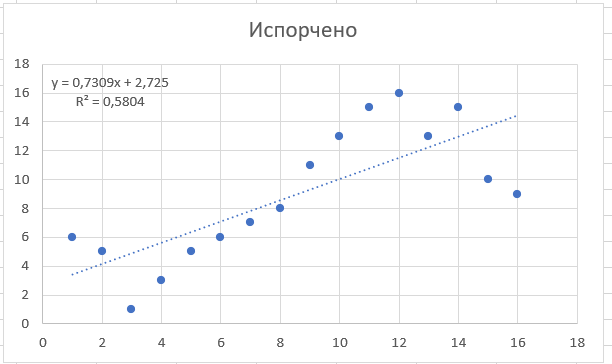


Рис. 31. Линия тренда

Подключим «Пакет анализа» для последующих шагов лабораторной работы.

Воспользуемся анализом данным, указав значения независимой и зависимой переменных в качестве входных данных.

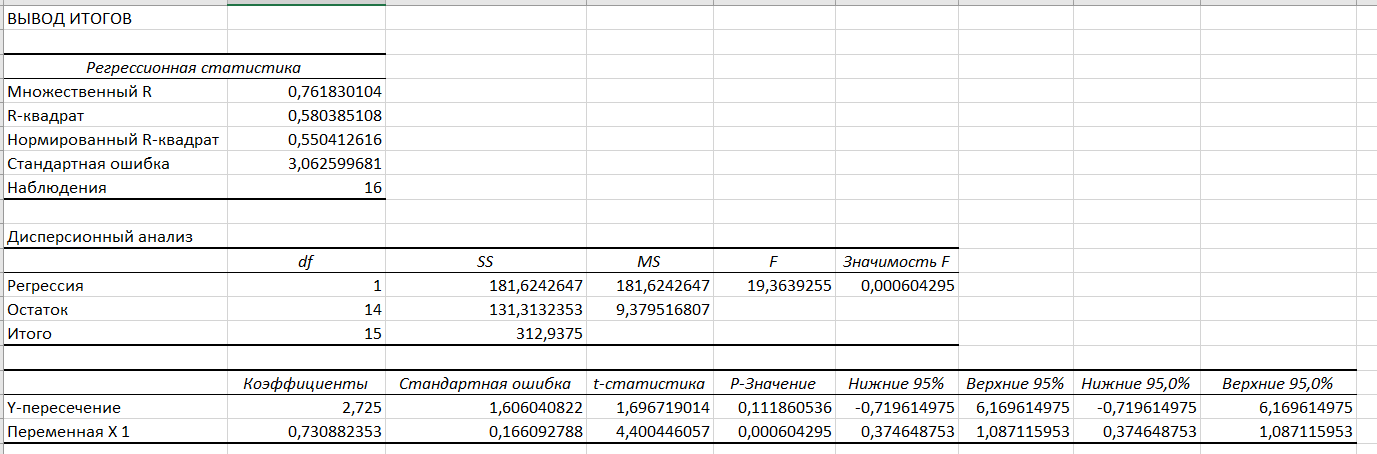


Рис. 32. Результат анализа данных

Значения коэффициентов уравнения и R-квадрата совпадают со значениями, полученными при построении линии тренда.

Значения, полученные с помощью функций НАКЛОН и ОТРЕЗОК также совпадают с ранее полученными.



Рис. 32. Результаты функций

Можно сделать вывод, что количество испорченных единиц экипировки увеличивается к концу сезона, так как оборудование успевает поизноситься за предыдущее время. То есть владельцам прокатного центра следует закупать новое и заменять старое оборудование в середине сезона, чтобы не возникало дефицита.

Полученное значение R-квадрат говорит о том, что 58% вариаций зависимой переменной объясняется независимой переменной, а 42% - другими факторами. На мой взгляд, это недостаточные показатели, чтобы говорить о высокой точности прогнозирования.

С помощью уравнения регрессии постараемся спрогнозировать количество испорченных единиц на ближайшие временные периоды. На мой взгляд, нельзя полностью доверять полученным данным, так как существуют другие факторы, влияющие на зависимость.



**Лабораторная работа №4**

Создадим сводную таблицу а основе базы данных пункта проката.

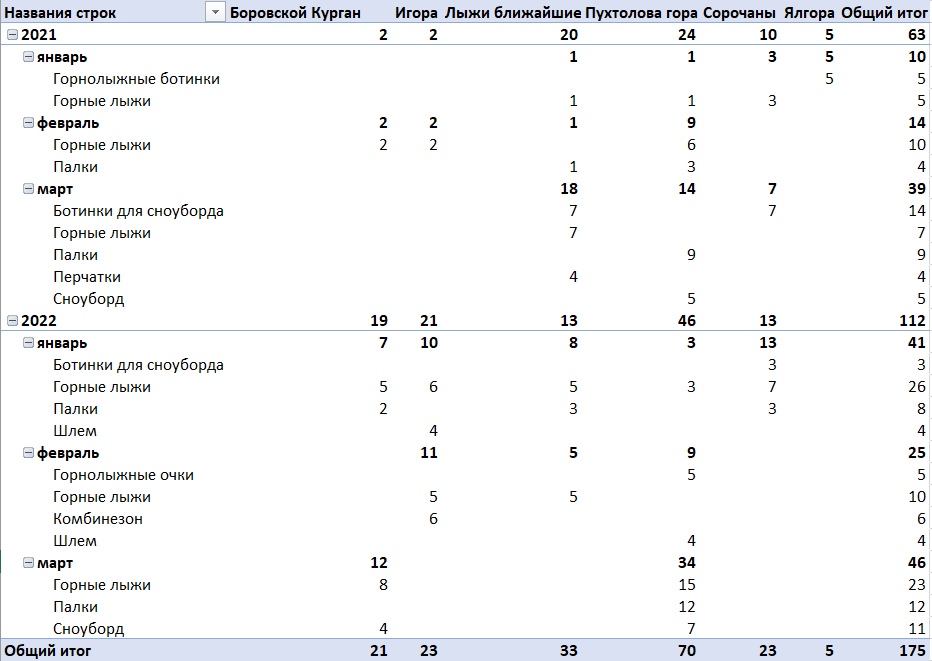


Рис. 33. Сводная таблица

Сделаем срез по наименованию арендуемой экипировки.

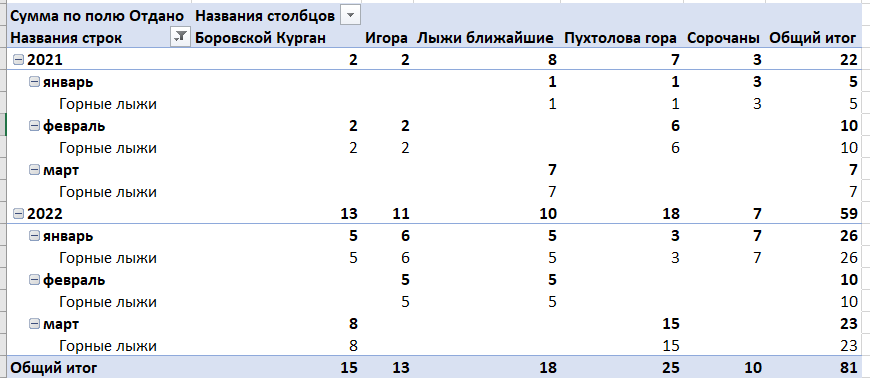


Рис. 34. Одинарный срез

Сделаем двойной срез по наименованию и названию курорта.

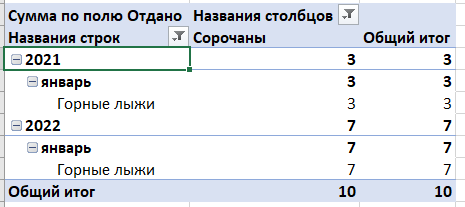


Рис. 35. Двойной срез

Спарклайны — это небольшие диаграммы в ячейках листа, визуально представляющие данные. С помощью спарклайнов можно показывать тенденции в рядах значений (например, сезонные повышения и спады или экономические циклы) и выделять максимальные и минимальные значения. Чтобы добиться максимального эффекта, спарклайны следует располагать рядом с соответствующими данными.

Построим спарклайн-гистограммы по данным из сводной таблицы, где отображены значения предметов, отданных в аренду.

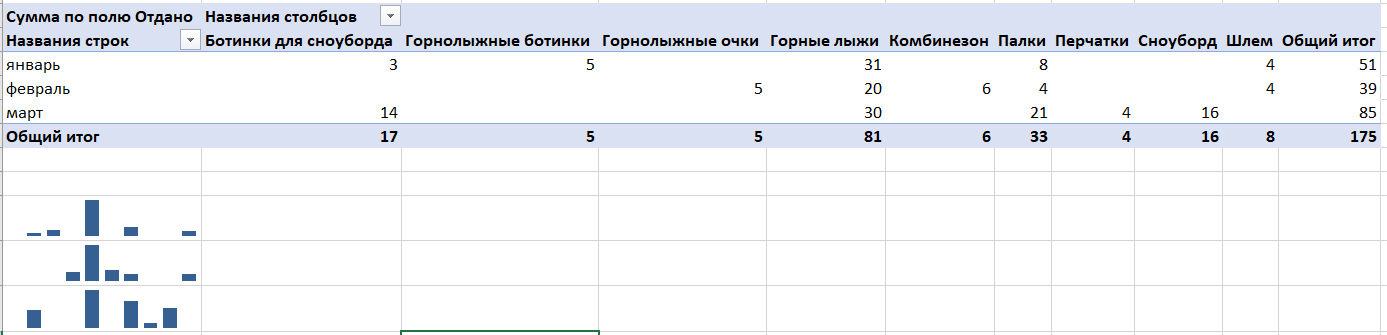
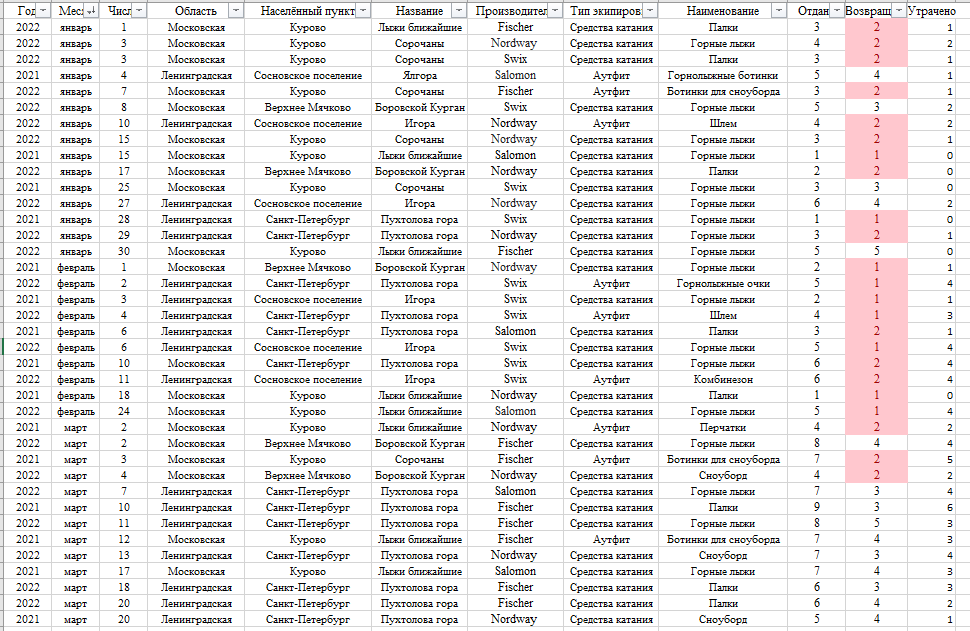
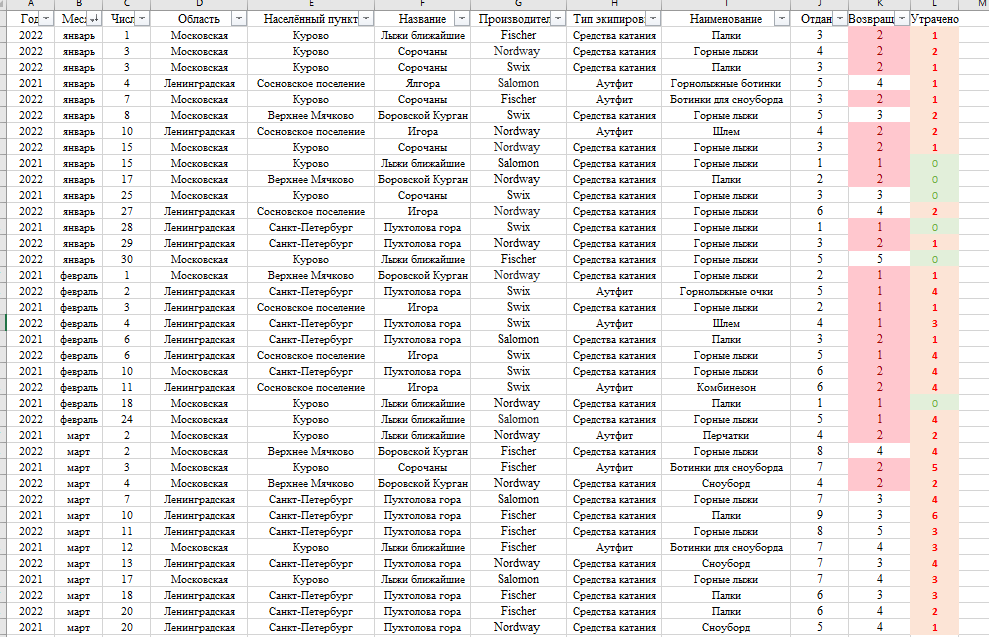


Рис. 36. Спарклайн-гистограммы

Отформатируем таблицу так, чтобы в столбце отдано, выделялись ячейки со значением ниже среднего.



В столбце утрачено, выделим ярким цветом ячейки, значение которых неравно нулю.



Предположим, что компания Salomon прекратило сотрудничество с прокатным центром, необходимо визуально оценить, насколько популярны были их услуги среди клиентов и сильно ли пострадает рейтинг организации после разрыва контракта.

