САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО

Институт «Компьютерных наук и кибербезопасности»

«Высшая школа технологий искусственного интеллекта»

Направление «Математическое обеспечение и администрирование информационных систем»

**Отчёт по заданию**

«Введение в науки о данных и визуализации»

Преподаватель: Э.Р. Хольгер

Выполнили:

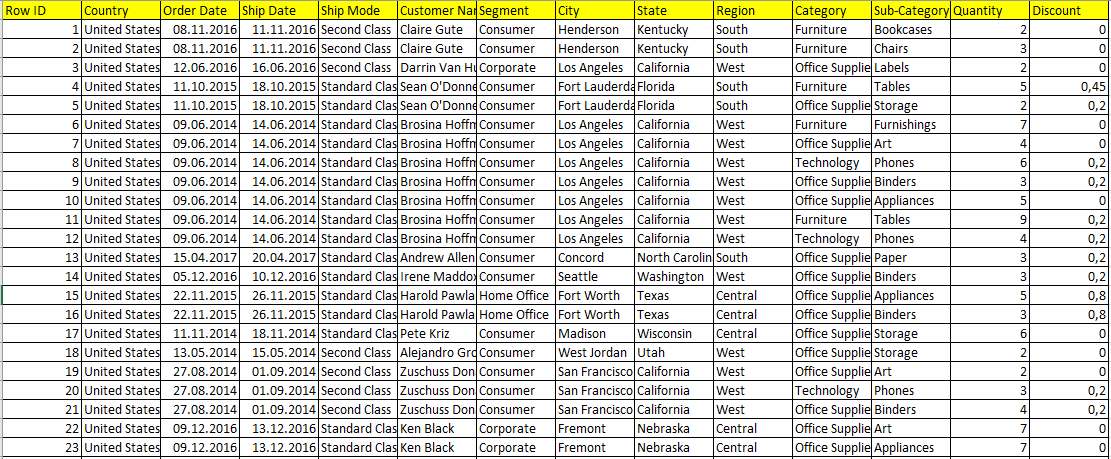
Соколова Елизавета группа: 5130203/40002 семестр: 2

Большаков Леонид группа: 5130203/40002 семестр: 2

Ксиротирис Константин группа: 5130203/40002 семестр: 2

Кардашев Алияр группа: 5130203/40002 семестр: 2

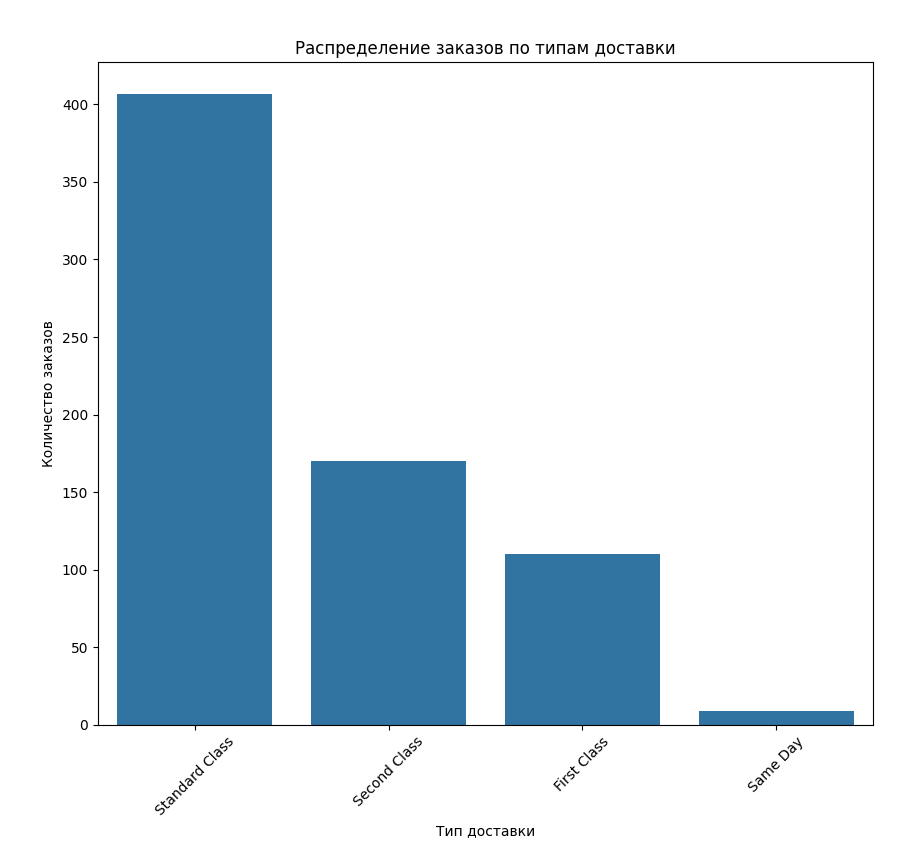
В качестве базы данных для выполнения данной работы мы взяли набор данных с табличной структурой данных, содержащий информацию о продажах за последние годы в одном из торговых центров в Соединенных Штатах с числовыми и категориальными видами данных. Выглядит он таким образом:



Наш исходный код написан на python и работает в VS Code.

Ниже представлена визуализация данных этой таблицы в различных формах.

1) Визуализация категориальных переменных (по типу доставки)

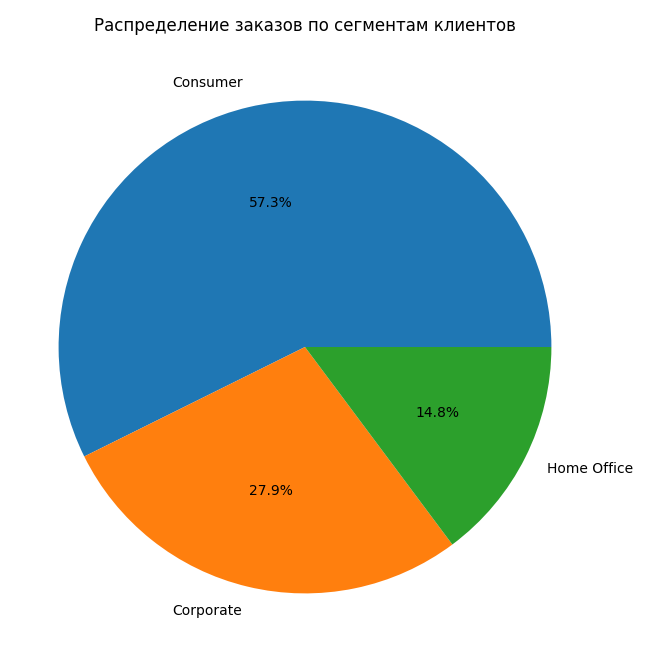


Ключевые выводы:

1. Стандартная доставка — явный фаворит (400+ заказов), так как наиболее экономична.
2. Другие классы (Второй, Премиум) значительно менее популярны из-за высокой стоимости.
3. Рекомендации:
   * Улучшить стандартную доставку (нагрузка велика).
   * Снизить цены/добавить бонусы для других классов, чтобы стимулировать спрос.

Итог: Клиенты выбирают бюджетный вариант — это нужно учитывать в логистике и ценообразовании.

## 2) Визуализация категориальных переменных (по сегменту клиентов)



Анализ данных показывает четкое распределение клиентской базы:

- Основной сегмент - Consumer (53.7%) - физические лица формируют более половины всех заказов, что указывает на ориентацию бизнеса на массового потребителя.

- Корпоративный сектор (27.9%) занимает второе место, демонстрируя значительный, но не основной объем продаж.

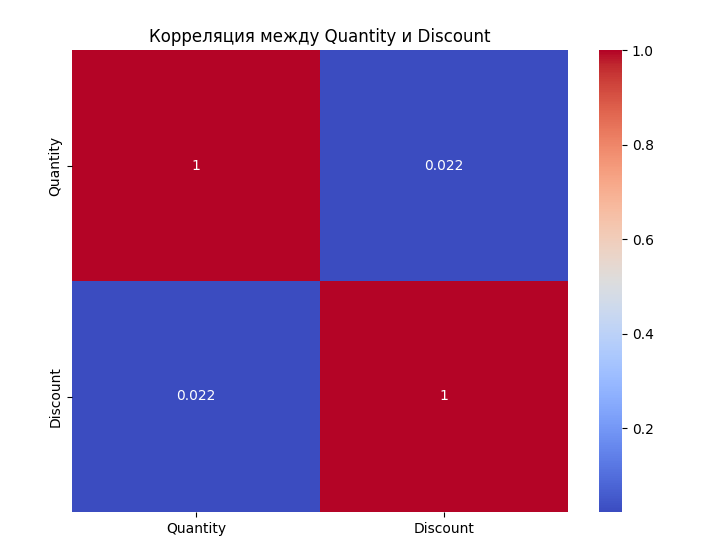
- Самозанятые (14.8%) представляют наименьшую, но заметную долю рынка.

Выводы для бизнеса:

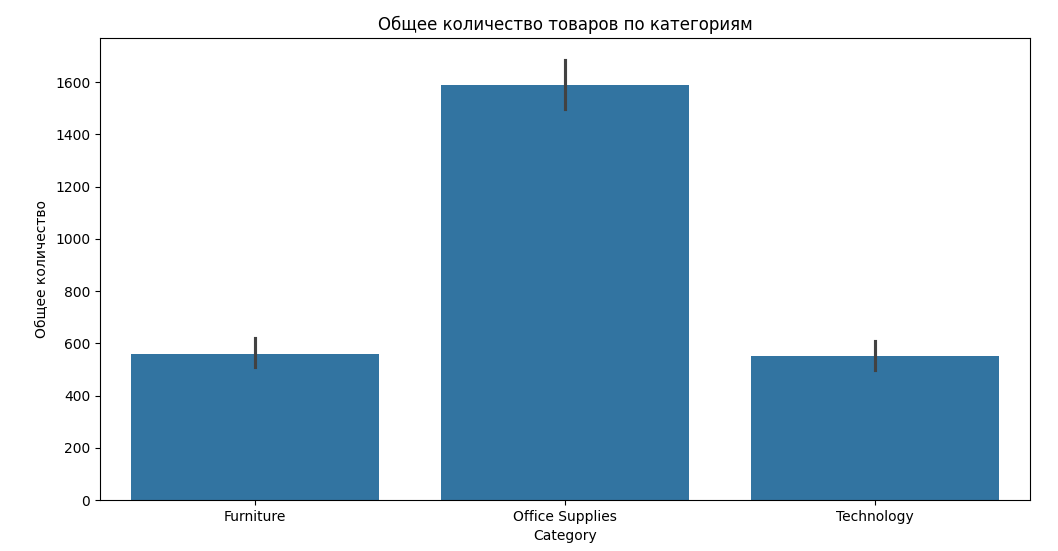
* Маркетинговые стратегии должны быть ориентированы прежде всего на частных потребителей
* Корпоративный сегмент требует отдельного подхода и специальных условий
* Самозанятые могут представлять перспективную нишу для развития

Такое распределение помогает правильно распределить рекламный бюджет и разрабатывать таргетированные предложения для каждой группы клиентов.

3) Таблица корреляции числовых характеристик



4) Столбчатая диаграмма по категориям товаров, анализ их количества



Из данной диаграммы можно сделать несколько выводов: во-первых, офисные принадлежности, или канцтовары, больше чем втрое превышают остальные категории товаров ; во-вторых, количество канцтоваров, проданных за данный промежуток времени, превышает 1500 и почти доходит до 1600, что позволяет оценить их спрос у клиентов.

5) Анализ и распределение скидок по категориям.

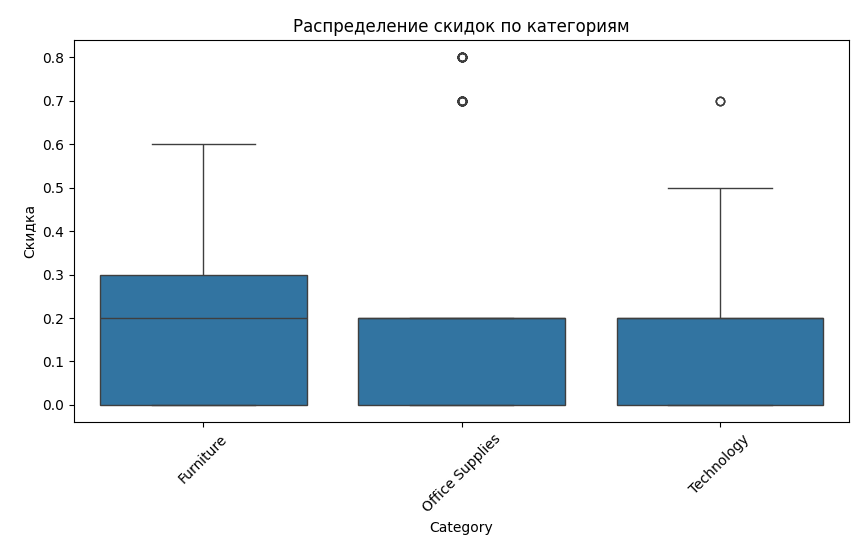


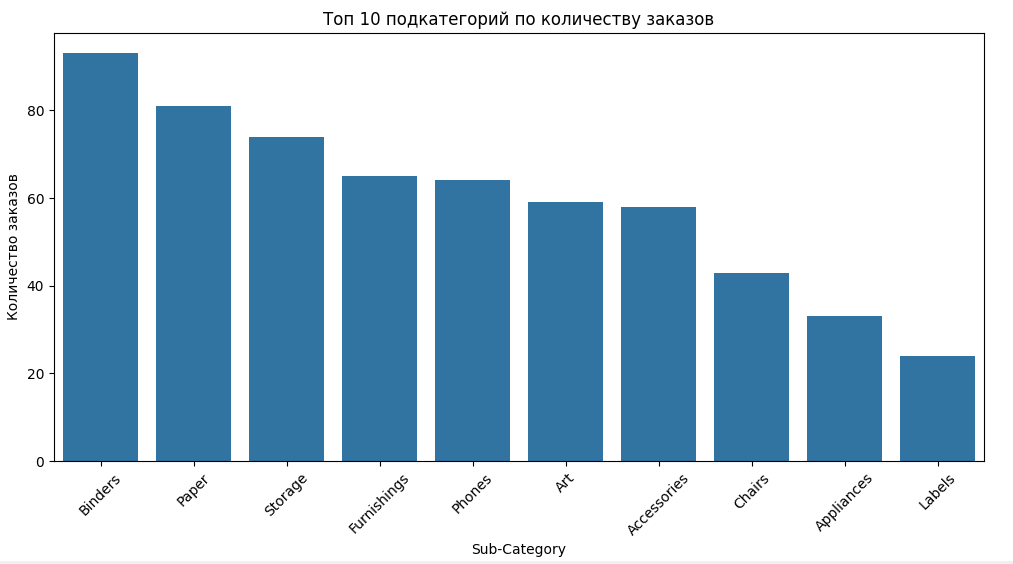
Диаграмма показывает явный перекос в распределении скидок: фурнитура лидирует с большим отрывом, тогда как технологии и офисные товары имеют скромные и примерно равные показатели.

Ключевые выводы:

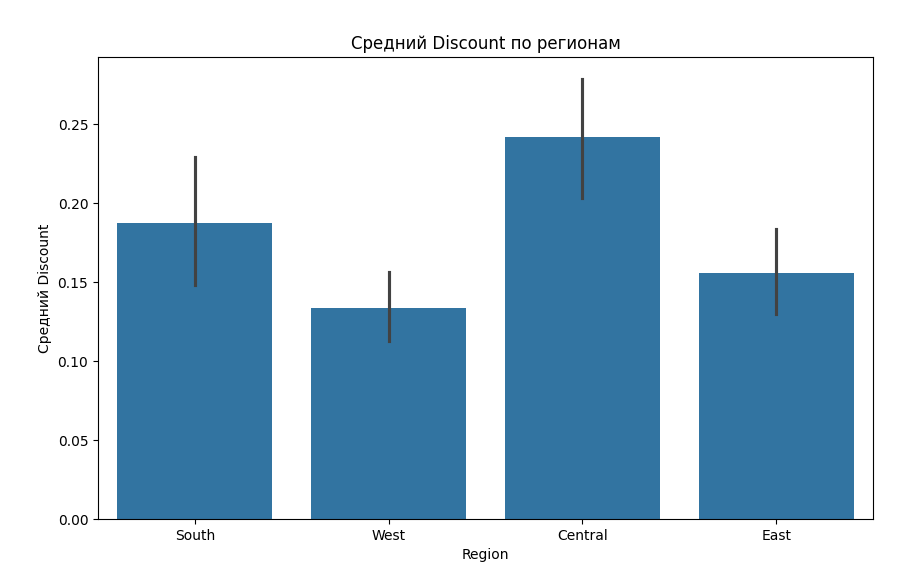
1. Фурнитура — главный драйвер продаж или наиболее конкурентная категория
2. Технологии и офисные товары не требуют активного продвижения через скидки
3. Возможность перераспределить маркетинговый бюджет в пользу приоритетных категорий

Такие данные помогают оптимизировать ассортиментную и ценовую политику.

6) Анализ самых частых подкатегорий по количеству заказов на основе столбчатой диаграммы.



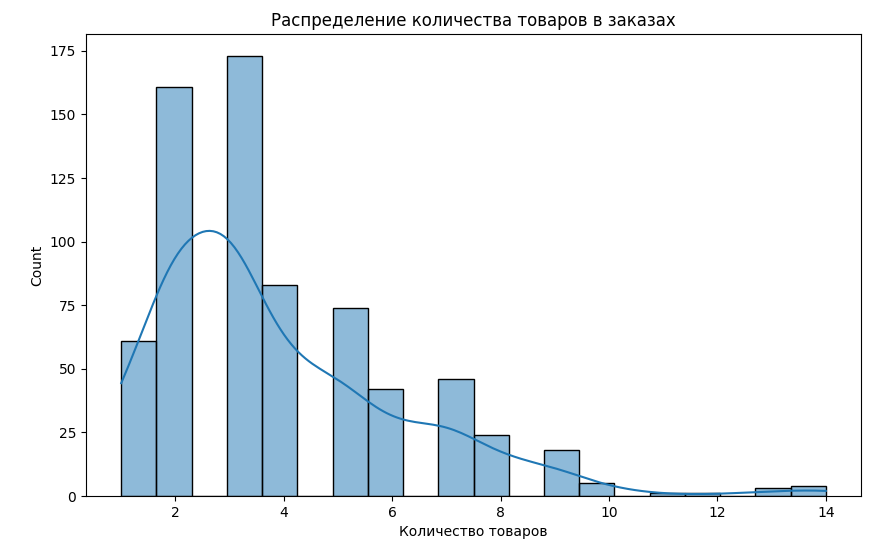
7) Диаграмма скидок по регионам



Эта диаграмма показывает средние скидки по регионам. Центральный регион лидирует по размерам скидок, за ним следуют Юг и Запад. Минимальные скидки предлагаются на Западе — это может быть связано с низкой конкуренцией, высокой логистической нагрузкой или особенностями местного спроса.

Анализ таких данных помогает оптимизировать ценовую политику и маркетинговые стратегии для каждого региона.

8) Гистограмма распределение количества товаров в заказах



1. **Типичный размер заказа**
   * Показывает, какое количество товаров чаще всего заказывают клиенты (модальное значение)
2. **Распределение по объёмам**
   * Выявляет соотношение мелких и крупных заказов

Анализ показывает, что, как правило, количество товаров в заказе – 2-4.

**Общий вывод по анализу данных о продажах**

Проведенное исследование позволило выявить ключевые тенденции в покупательском поведении, эффективности маркетинговых стратегий и логистических процессов. Основные выводы сводятся к следующему:

1. Клиенты предпочитают экономичные варианты

- Стандартная доставка значительно популярнее премиальных вариантов, что указывает на высокую чувствительность покупателей к стоимости услуг.

- Это требует оптимизации логистики для снижения нагрузки на стандартные каналы и разработки стимулов для других способов доставки.

1. Сегмент потребителей определяет структуру спроса

- Физические лица (Consumer) формируют более половины продаж, что подтверждает ориентацию бизнеса на массовый рынок.

- Корпоративный сегмент (Corporate) и самозанятые (Home Office) также представляют значительные доли, но требуют дифференцированного подхода.

1. Ассортимент и ценовая политика влияют на продажи

- Канцтовары являются самой популярной категорией, что говорит о стабильном спросе на офисные принадлежности.

- Фурнитура активно продвигается через скидки, тогда как технологии менее зависят от дисконтов.

- Это позволяет перераспределить маркетинговые усилия между категориями.

1. Региональные различия в скидках и спросе

- Центральный регион лидирует по уровню скидок, возможно, из-за высокой конкуренции.

- Западный регион демонстрирует минимальные скидки, что может быть связано с особенностями местного рынка.

- Такой анализ помогает адаптировать стратегию продвижения под локальные условия.

1. Средний размер заказа помогает в управлении запасами

- Большинство заказов включают 2–4 товара, что важно для планирования складских мощностей и упаковки.