



Высшая школа бизнеса НИУ ВШЭ

Бизнес-аналитика и системы
больших данных

Москва, 2025

Бизнес-дашборд для ритейла

Прикладная наука о данных

Выполнили:

Лашхия Софья
Гаджибегова Кристина
Рощина Надежда
Васенёва Валерия



Введение

Целью является *разработка* бизнес-дашборда *для анализа продаж* в розничной торговле и *формулировка* аналитических гипотез на основе данных.

Основой исследования послужил датасет *Superstore Sales Dataset*, содержащий информацию о продажах товаров в различных штатах США за период 2015–2018 гг.

Процесс работы:

1. Сбор и первичная обработка данных
2. Разведочный анализ и генерация гипотез (EDA)
3. Визуализация и интерактивная аналитика
4. Проверка гипотез
5. Формулировка выводов



Описание данных и EDA

Финансовые показатели

- Общий объем продаж составил \$2.26 миллиона при среднем чеке \$230
- Высокая вариативность покупок — от \$0.44 до \$22,638
- 75% всех продаж не превышают \$210

Бизнес-метрики

- 4,922 уникальных заказа от 793 клиентов
- Ассортимент включает 1,861 уникальный товар
- Средняя корзина — 2 товара на сумму \$460
- Стандартное время доставки — 4 дня

В основе анализа — 9,800 транзакций за 4-летний период (2015-2018 гг.)

Данные охватывают продажи по всем регионам США с детализацией до уровня отдельного товара в заказе

Структура данных:

- Содержит полную информацию о заказах, клиентах, товарах и финансовых показателях
- Включает 3 ключевые категории товаров: Furniture, Office Supplies и Technology
- Каждая запись представляет собой товарную позицию в составе заказа



Формулировка гипотез

Гипотезы сфокусированы на ключевых бизнес-вопросах: *клиентской базе, товарных категориях, временных паттернах и логистике.*

Статистический анализ

Принцип Парето для клиентов

Сезонность технологий

День недели для максимального чека

Логистика и стоимость заказа

Лояльность и география

Совместные покупки

Визуальная аналитика

Сезонный пик продаж



Проверка гипотез 1-6: Результаты

	Гипотеза	Статус	Ключевой факт
1	Принцип Парето для клиентов	✗	20% клиентов = 48,5% выручки
2	Сезонность технологий	✓	Коэф. сезонности = 0,601(макс)
3	День недели для максимального чека	✗	Лидер: Четверг (\$264), а не понедельник
4	Логистика и стоимость заказа	✗	p-value = 0,325 (разница незначима)
5	Лояльность и география	✗	Заказов/клиент: крупные города - 1,02, малые - 1,00 при $p < 0,001$
6	Совместные покупки	⚠	Комбинация встретилась 901 раз(2 место)

Итог:

2 подтвержденные гипотезы дали четкие инсайты
4 опровергнутых — скорректировали понимание бизнеса.



Дашборд Power BI





Гипотеза 7





Гипотеза 7





Гипотеза 7



Вывод: гипотеза частично подтверждена

«Сентябрь действительно является одним из пиковых месяцев, однако главный пик потребительской активности приходится на ноябрь».



Результаты

1. **Провели полный цикл анализа**
 - а. От первичной обработки данных до проверки гипотез и визуализации
2. **Получили проверенные выводы**
 - а. Из 7 гипотез подтвердились 3 (1 полностью, 2 частично)
3. Главное открытие: Клиентская база сбалансирована, а ноябрь — ключевой месяц для продаж
4. **Создали рабочий инструмент**
 - а. Дашборд в Power BI для мониторинга и углубленного анализа
5. **Предоставили бизнесу:**
 - а. Понимание структуры продаж и клиентского поведения
 - б. Основу для данных решений в маркетинге и ассортименте



Спасибо за внимание!

