**Цель проекта:**  
Проанализировать, влияет ли наличие промо-акций (Promo = 1) на ежедневные продажи в ритейл-магазине, используя симулированные данные. Провести статистический A/B-тест и выявить закономерности.

**Ограничения и методология:**  
Данные являются наблюдательными, а не результатом контролируемого эксперимента. Поэтому проведённый анализ представляет собой **квазиэксперимент**. Мы использовали **Welch t-test**, чтобы компенсировать неравенство размеров групп и потенциальные различия в дисперсиях.

**Гипотеза:**

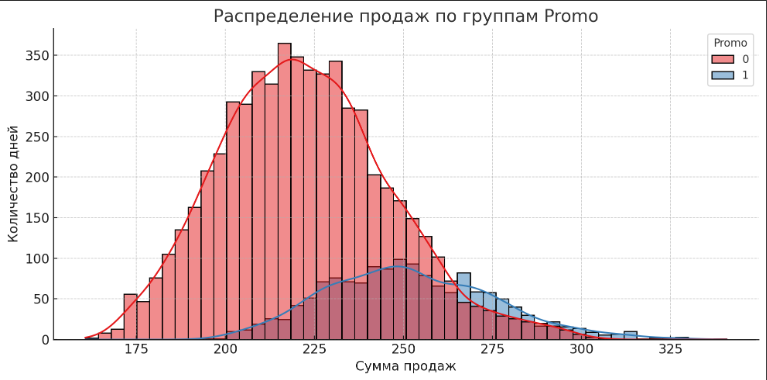
* H₀: Средние продажи в дни с акцией и без неё не отличаются.
* H₁: Продажи в дни с акцией выше.

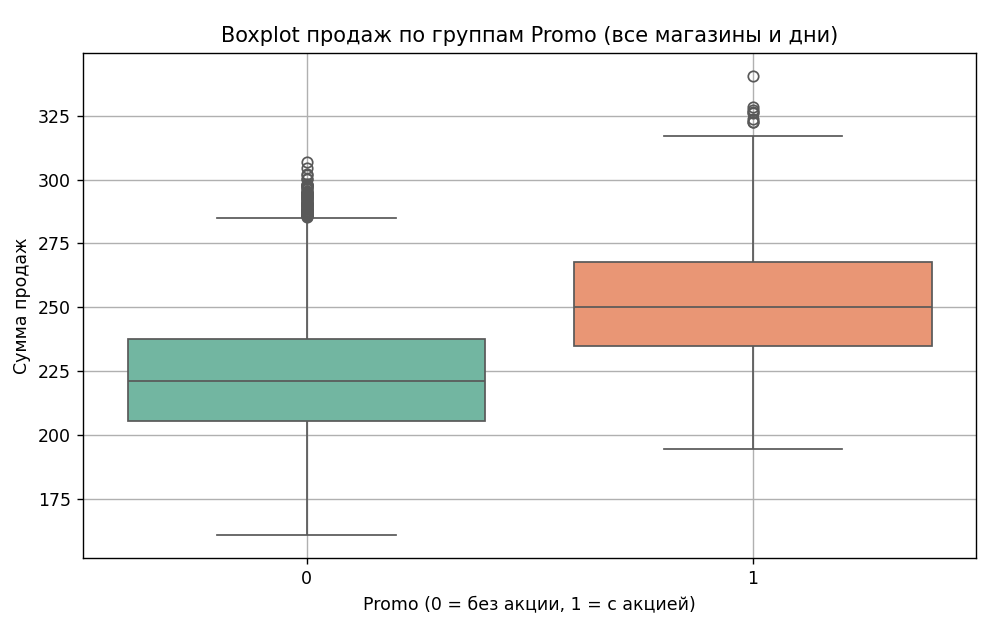
**Данные:**

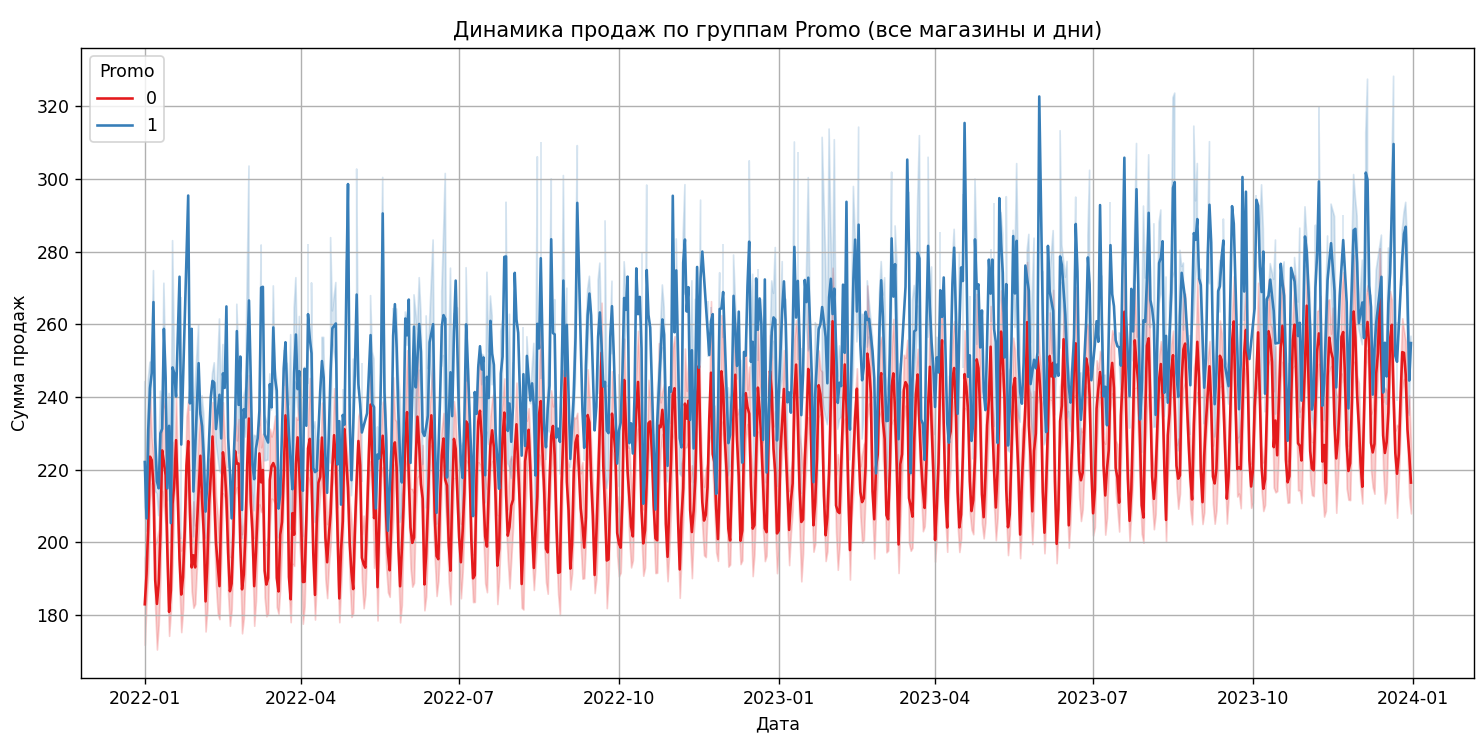
* 7300 строк (по 10 магазинам, ~ за 2 года)
* Поля: Date, Store\_ID, Sales, Promo, Holiday

**Проведём разведочный анализ (EDA), чтобы:**

1. Посмотреть, сколько у нас данных с промо (Promo = 1) и без,
2. Проверить средние продажи в каждой группе,
3. Посмотреть распределения.







| **Группа** | **Средняя продажа** | **Медиана** | **Кол-во записей** |
| --- | --- | --- | --- |
| **Без промо** | 222.47 | 221.12 | 5824 |
| **С промо** | 251.97 | 250.26 | 1476 |

**Вывод по распределению продаж:**

* Распределения похожи на нормальные, особенно в группе без промо.
* Группа с промо имеет **выше среднюю продажу**.
* Размер групп сбалансирован не идеально (меньше записей с Promo = 1), но достаточно для сравнения.

**Вывод по ящику с усами:**

* Группа с Promo = 1 показывает **более высокую медиану**.
* Также у неё выше верхняя граница интерквартильного диапазона.
* Распределения довольно "компактные" и близки по форме, но сдвинуты вверх при Promo = 1.
* Продажи в дни с промо-акцией в среднем и по медиане **значимо выше**, чем в дни без акций, даже без учёта магазинов и праздников. Эффект устойчив и визуально различим.

**Вывод Time Series:**

* Промо-акции сопровождаются краткосрочным скачком продаж по всей сети магазинов. Эффект устойчив по времени и работает в различных временных отрезках.

**Результаты A/B-теста**:

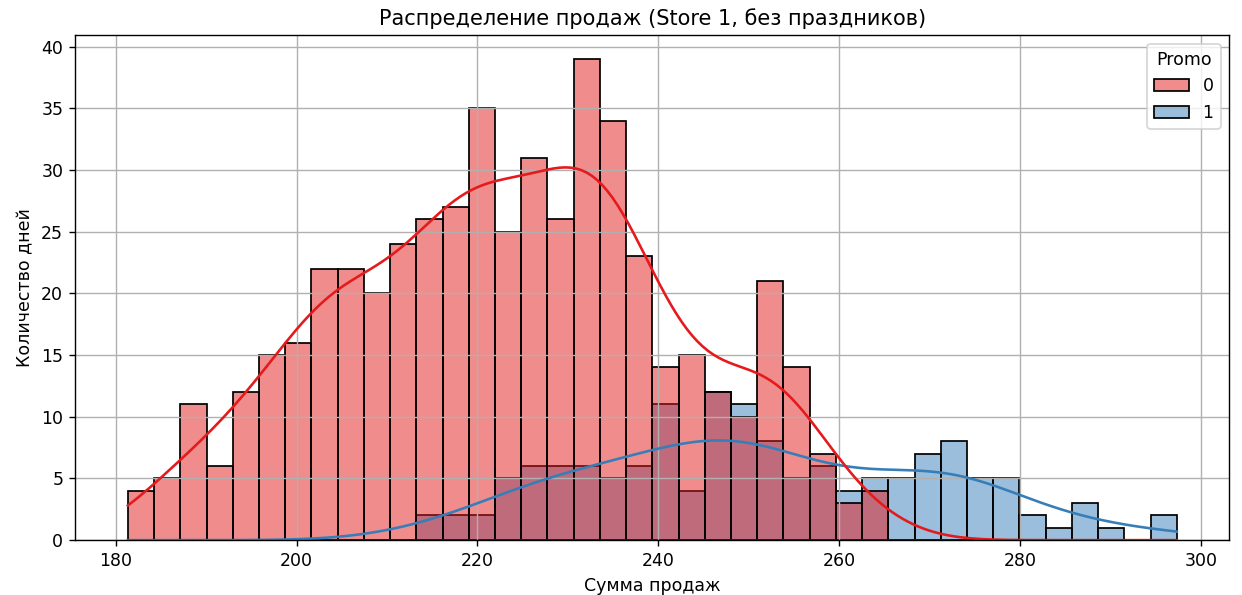
* **t-статистика**: 42.75
* **p-value**: 2.1e-294 (это ~0, означает почти 100% уверенность)
* **Разница средних**: +29.5 (в пользу Promo = 1)

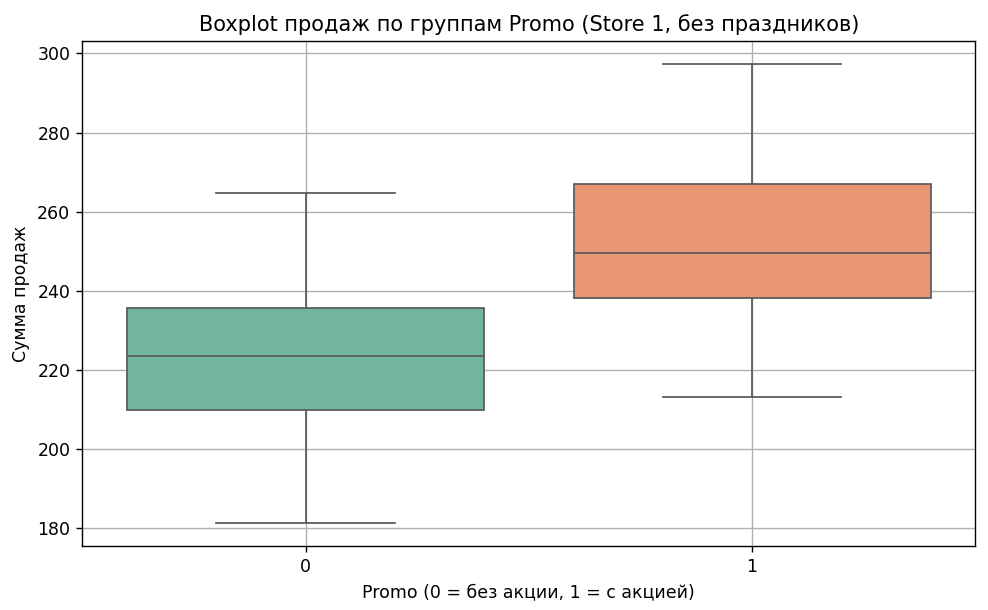
### Итоги:

* Мы **с высокой статистической значимостью отвергаем нулевую гипотезу**.
* **Продажи в дни с промо-акциями действительно выше** в среднем на ~29.5 единиц.
* Такой результат нельзя объяснить случайностью.

**Проверим устойчивость без учёта праздников по отдельному магазину, гипотезы те же.**

Берём Store\_ID = 1 и флаг Holiday = 0 и повторно проводим EDA -> A/B тест.







| **Группа** | **Средняя продажа** | **Медиана** | **Кол-во записей** |
| --- | --- | --- | --- |
| **Без промо** | 223.47 | 223.68 | 523 |
| **С промо** | 251.42 | 249.55 | 145 |

**Вывод по распределению продаж:**

* Группа **без промо** (Promo = 0) содержит больше наблюдений.
* Группа **с промо** — меньше, но достаточно репрезентативна для анализа.
* Средняя продажа в дни **без промо**: около **222**
* Средняя продажа в дни **с промо**: около **250**
* Для обеих групп медиана близка к среднему → **распределение симметрично** (примерно нормальное).
* Это значит, что t-тест применим корректно.
* Распределения двух групп почти нормальные, с похожей формой.
* Кривая для Promo = 1 сдвинута **правее** — это говорит о **более высоких продажах** при наличии акции.
* Перекрытие небольшое → визуально видно, что группы действительно различаются.

**Вывод по ящику с усами:**

* **Медиана и весь ящик у группы Promo = 1 находятся выше**, чем у Promo = 0.
* У группы с промо-акцией также **выше верхний квартиль** — продажи чаще достигают высоких значений.
* Видно некоторое количество **выбросов** в обеих группах, но они не влияют на общую картину.
* В дни с промо-акцией (Promo = 1) продажи статистически и визуально **значимо выше**, чем в обычные дни. Распределение продаж сдвинуто вверх. Это подтверждает гипотезу о положительном влиянии промо

**Вывод Time Series:**

* Во временной динамике видно, что дни с акциями сопровождаются **резким ростом продаж**. Хотя они происходят нечасто, **эффект от промо выражен на фоне общей тенденции**. Это говорит в пользу использования акций как инструмента стимулирования краткосрочного спроса.

В магазине Store\_ID = 1 промо-акции **связаны с ростом продаж**. Это подтверждается как сводной таблицей, так и графическим анализом. Данные распределены достаточно нормально, группы не слишком дисбалансированы — можно уверенно переходить к статистическому тесту.

**Анализ по магазину Store\_ID = 1 (только обычные дни):**

* **t-статистика**: 15.82
* **p-value**: 2.44e-38 — крайне значимо
* **Разница средних**: +27.95
* **Размеры групп**:
  + Promo = 1: 145 дней
  + Promo = 0: 523 дня

**Рекомендации:**

* Промо-акции статистически и визуально приводят к росту продаж.
* Рекомендуется использовать акции как краткосрочный инструмент стимулирования, особенно в непраздничные дни.
* Для более точной оценки влияния акций можно провести A/B-тестирование в реальном времени с рандомизацией.

**Используемые технологии и инструменты:**

* Python (pandas, matplotlib, seaborn, scipy)
* A/B-статистика: Welch t-test
* Визуализация: boxplot, time series
* Источник данных: https://www.kaggle.com/datasets/abhishekjaiswal4896/store-sales-dataset