
Денис Черновол

+7-905-672-76-66

chernovoldv@gmail.com

t.me/chernovoldv

github.com/chernovoldv

ПОРТФОЛИО

Проведение A/B – теста для оценки работы нового алгоритма рекомендаций

Проведение A/B – теста для оценки работы нового алгоритма рекомендаций постов.

Использованы t-тест, u-тест Манна-Уитни, тесты поверх бакетного преобразования, Cuped t-тест, Пуассоновский бутстреп.

Используемые инструменты:

SQL (ClickHouse), Python (PandaHouse, Pandas, NumPy, Matplotlib, Seaborn, SciPy, Statsmodels).

Примечание: в работе оставлены исходные данные, предоставленные karpov.courses.

Ссылки на работу:

[GitHub](#), [Google-Colab](#)

Система алертов для социальной сети

Создание системы алертов для приложения, чтобы своевременно реагировать на технические сбои в работе сервиса, сбои в ETL-процессах, ошибки при обработке данных, существенные изменения метрик и т. д.

Используемые инструменты:

SQL (ClickHouse), Python (PandaHouse, Pandas, NumPy, Matplotlib, Seaborn, Apache Airflow).

Ссылки на работу:

[GitHub](#)

Онлайн-сервис «Поиск ментора»

Анализ данных, поиск закономерностей и прогнозирование на основе запросов продакт-менеджера.

Используемые инструменты:

SQL (PostgreSQL), Python (SQLAlchemy, Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-Learn).

Ссылки на работу:

[GitHub](#), [Google-Colab](#)

Онлайн-сервис «Поиск ментора» (дашборд)

Создание дашборда на основе запросов продакт-менеджера.

Используемые инструменты:

SQL (PostgreSQL), Power BI, DAX.

Ссылки на работу:

Дашборд: [Google-Drive](#)

Описание дашборда: [Google-Drive](#)

[Отчет о доходах от операций на фондовом рынке](#)

Написание функций для автоматизации расчетов прибыли/убытков от операций с валютой

Используемые инструменты:

Python (Pandas, regular expression operations)

Ссылки на работу:

[GitHub](#), [Google-Colab](#)

Примечание

Работы из портфолио содержат измененные или синтетические данные, созданные с помощью библиотек Faker и Synthetic Data Vault.

Проведение A/B – теста для оценки работы нового алгоритма рекомендаций

Ситуация:

В социальной сети (приложении) внедрены алгоритмы рекомендации постов в ленте новостей. Команда дата-сайентистов предложила новый алгоритм рекомендаций, который показывает в рекомендациях похожие на посты с like пользователя.

Основная гипотеза заключается в том, что новый алгоритм во 2-й группе приведет к увеличению CTR.

Необходимо проанализировать результаты эксперимента. В группе 2 был использован один из новых алгоритмов рекомендации постов, группа 1 использовалась в качестве контроля.

Задача:

1. Выбрать метод анализа и сравнить CTR в двух группах (t-тест, Пуассоновский бутстреп, тест Манна-Уитни, t-тест на сглаженном ctr ($\alpha=5$) а также t-тест и тест Манна-Уитни поверх бакетного преобразования).
2. Сравнить данные этими тестами, посмотреть на распределения данных.
3. Напишите рекомендацию, будем ли мы раскатывать новый алгоритм на всех новых пользователей или все-таки не стоит.

Действия:

1. Выгружены данные из базы данных за экспериментальный и пред экспериментальный периоды, проведена проверка на наличие ошибок;
2. Проведен визуальный анализ данных, применены статистические критерии.
3. Дана рекомендация на основе проведенного анализа отказаться от внедрения нового алгоритма.

Результат:

Рекомендации:

От раскатывания нового алгоритма (рекомендации "похожих на лайкнутые постов") на всех пользователей следует отказаться.

Но учитывая, что для части пользователей наоборот стала более вовлеченной (увеличился 90% квантиль CTR), стоит более подробно изучить влияние нового алгоритма на часть пользователей.

Возможно у нас есть категория пользователей, для которой новый алгоритм дает лучшие результаты.

[Главная страница](#)

Система алертов для социальной сети

Ситуация:

Социальная сеть (приложение) имеет 2 раздела лента новостей для просмотра рекомендованных постов и сообщения.

Необходимо написать систему алертов для приложения, чтобы своевременно реагировать на технический сбой в работе сервиса, сбой в ETL-процессах, ошибки при обработке данных, внешние факторы и т.д.

Система должна с периодичность каждые 15 минут проверять ключевые метрики — такие как активные пользователи в ленте / мессенджере, просмотры, лайки, CTR, количество отправленных сообщений.

Задача:

1. Изучите поведение метрик и подберите наиболее подходящий метод для детектирования аномалий.
2. В случае обнаружения аномального значения, в чат должен отправиться алерт — сообщение со следующей информацией: метрика, ее значение, величина отклонения.
3. Автоматизируйте систему алертов с помощью Airflow.

Действия:

4. На основе исторических данных выбран метод выявления аномальных значений (межквартильный размах) и подобраны параметры для каждой метрики.
5. Написан Dag для Airflow, выполняемый каждые 15 минут и включающий функции:
 - 5.1. для извлечения данных из БД Clickhouse;
 - 5.2. для определения аномальных значений;
 - 5.3. для создания отчета (сообщение и графики);
 - 5.4. для отправки отчета.

Результат:

Результат представлен в виде файла Python.

Файл готов для загрузки в Apache Airflow, для работы Dag необходимо указать в Airflow переменные для подключения к БД (host и пароль) и использования telegram-бота (bot token).

[Главная страница](#)

Онлайн-сервис «Поиск ментора»

Ситуация:

Вы работаете аналитиком в онлайн-сервисе “Поиск ментора”. Это сервис, который помогает менторам и менти найти друг друга.

Встречи проходят на площадке сервиса. Назначить встречу можно в личном кабинете после авторизации на сайте.

Продакт-менеджер начал подготовку к новому этапу развития приложения и попросил вас собрать данные о работе сервиса.

Задача:

4. Изучите данные и подготовьте ответы на вопросы продакт-менеджера.
5. Определите точки роста.
6. Спрогнозируйте, сколько новых менторов нужно найти, если в следующие 2 недели количество активных менти увеличится на 250 человек. Объясните результат и ваше решение.

Действия:

6. для каждого вопроса написаны запросы на SQL;
7. для полученных данных построены визуализации;
8. для части вопросов проведен корреляционный анализ;
9. определена потребность в найме менторов.

Результат:

Результат представлен в виде файла Jupyter Notebook.

На основе проведенного анализа даны рекомендации для дальнейшего развития сервиса.

Также определена минимальная потребность в найме менторов на следующие 2 недели при условии увеличения активных менти.

Онлайн-сервис «Поиск ментора» (дашборд)

Ситуация:

Вы работаете аналитиком в онлайн-сервисе «Поиск ментора». Это сервис, который помогает менторам и менти найти друг друга.

Встречи проходят на площадке сервиса. Назначить встречу можно в личном кабинете после авторизации на сайте.

Продакт-менеджер начал подготовку к новому этапу развития приложения и попросил вас собрать данные о работе сервиса.

Задача:

1. Изучите данные и подготовьте ответы на вопросы продакт-менеджера:
 - 1.1. Какое количество менторов и менти взаимодействуют каждый месяц на нашей платформе?
 - 1.2. Сколько у нас менторов и менти, которые еще не приняли участие ни в одной встрече?
 - 1.3. Сколько у каждого ментора в среднем успешных сессий в неделю по месяцам?
 - 1.4. Определите ТОП-5 менторов с самым большим количеством сессий за последний полный месяц.
 - 1.5. Сколько сессий по каждому направлению менторства в месяц обычно отменяется? Как меняется доля отмененных сессий ежемесячно?
 - 1.6. Определите, в какой день недели последнего полного месяца прошло больше всего встреч?
2. Построить дашборд на основе на основе заранее сформулированных вопросов

Действия:

1. В рамках построения дашборда вопросы продакт-менеджера были изменены для отражения показателей месяц к месяцу и год к году;
2. Для каждого вопроса созданы меры и/или таблицы на языке DAX;
3. Для полученных данных построены визуализации.

Результат:

Даны ответы на все вопросы продакт-менеджера. Подготовлен дашборд с подключением к базе данных.

[Главная страница](#)

Отчет о доходах от операций на фондовом рынке

Ситуация:

У меня открыто 2 брокерских счета у Finam и Тинькофф.

Мне необходимо самостоятельно отчитываться прибыли/убытках от продажи валюты.

Расходы на покупку валюты необходимо рассчитать методом FIFO (First In First Out), то есть при продаже валюты, происходит продажа валюты, которая куплена раньше всего.

Задача:

Для корректного расчета расходов на покупку валюты требуется поочередно обработать все сделки с валютой у брокера за все время.

Расчеты нужно произвести для каждого брокера отдельно.

Действия:

1. написано три пользовательские функции:
 - 1.1. tinkoff_def - для обработки отчетов Тинькофф брокера;
 - 1.2. finam_def - для обработки отчетов брокера Finam
 - 1.3. sell_def - функция для расчета прибыли/убытков от проведенных сделок (принимает на обработку датафрейм, полученный от работы функций tinkoff_def или finam_def).

Результат:

Расчет прибыли/убытков от операций на фондовом рынке автоматизирован и не будет занимать значительного времени.

Для каждой функции написана документация, код содержит пояснения.