

# **Содержание:**

- Вводная часть
- Список метрик
- SQL запрос для сбора метрик
  - Код DAG
- Функции для DAG'ов
  - Дашборд
  - Выводы
- Рекомендации

# Вводная часть:

**Задача** - создать инструмент для отслеживания метрик в помощь продакт-менеджеру.

## Описание задачи:

- Сделать DAG в Airflow, который будет собирать сырые данные, агрегировать их и класть в витрину:
  - Построить воронки и посчитать конверсию в добавления и в заказ по разным местам мобильного приложения.
  - Высчитать метрики (GMV за периоды, средний чек, кол-во заказов, кол-во пользователей с заказами, среднее кол-во товаров в заказе).
- Подготовить дашборд для продакт-менеджера:
  - Дашборд выполнен в Yandex DataLens.
  - В дашборде есть фильтры по дате и типу страницы.
  - Метрики названы понятно.
  - Дашборд регулярно обновляется.
- Обновление данных:
  - Настроить расписание DAG в Airflow.
  - Настроить Auto-update в DataLens с интервалом, не привышающим частоту расписания DAG.

# **Список метрик для регулярного подсчета в дашборде.**

- Общая валовая стоимость купленного товара (за весь месяц).
- Общий средний чек (за весь месяц).
- Общее количество заказов (за весь месяц).
- Общее количество пользователей с заказами (за весь месяц).
- Среднее количество товаров в заказе (за весь месяц).
- Воронка пользователей и конверсия за месяц.
- Воронка пользователей и конверсия по дням.
- GMV по дням.
- Средний чек по дням.
- Количество заказов по дням.
- Количество пользователей с заказами по дням.
- Среднее количество товаров в заказе по дням.

# **SQL запрос для сбора и расчета метрик.**

- [SQL запрос для расчета воронки и конверсии.](#)
- [SQL запрос для расчета метрик.](#)

## **Код DAG.**

- [DAG 1](#)
- [DAG 2](#)
- [DAG 3](#)
- [DAG 4](#)
- [DAG 5](#)

## **Функции для DAG'ов.**

[Ссылка](#) на все функции для DAG файлов

## **Дашборд.**

- [Дашборд в Yandex DataLens.](#)

# **ВЫВОДЫ:**

На основе задачи была проделана работа по написанию кода на Python для DAG файла, чтобы автоматизировать процесс сбора метрик.

- Написаны функции для проверки файлов в директории компьютера.
- Написаны функции для создания базы данных в SQLite, а после для агрегации метрик и построения воронки и конверсий в SQLite с последующим сохранением в новые CSV файлы для удобства использования в DataLens.
- Функция для выгрузки новых CSV файлов в Google Sheets через API. Это самый простой способ автоматизировать задачу в MVP-версии проекта.
- В Yandex DataLens установлено соединение с таблицами Google Sheets и подключены датасеты. Соединение устанавливаем именно с Google Sheets, так как это позволяет автоматизировать процесс сбора данных и автоматически обновлять дашборд.
- На основе датасетов был построен дашборд в DataLens.

Задачу по выполнению автоматизации процесса по сбору метрик и отображения их на Дашборде считаю выполненной.

## **Рекомендации:**

Текущая версия дашборда отображает данные за ограниченный период (1.01.2024 - 30.01.2024). Если бизнесу потребуется анализ динамики за несколько месяцев или построение более глубоких когортных отчетов, стоит расширить период хранения и загрузки данных или настроить отдельные витрины для истории.

В DAG файле нужно будет прописать вашу директорию, где будут храниться все CSV файлы и база данных SQLite. Так же рекомендую настроить свой API для выгрузки таблиц в ваш Google Sheets, чтобы у вас всегда был доступ.

Так же рекомендую обратить внимание на собранные данные за январь 2024 года. В январе 31 день, а собранных данных только 30 дней, возможно данные были собраны 31 января, а возможно это баг при сборе информации.