Добро пожаловать в нашу команду – “Delivery Dron”. Мы молодая и амбициозная команда, которая только начинает свое большое путешествие в бизнесе, но с серьезными инвестициями и грамотным бизнес планом.

Если у нас не получилось тебя напугать, тогда погнали!

Мы жаждем увидеть в своем составе классных аналитиков, которые возьмут на себя роль наших оракулов и сотворят чудо! Немного о чуде: наши пользователи совершают заходы на наш сайт по рекламе из разных рекламных каналов, а также есть органические пользователи, которые перешли на сайт самостоятельно (без затрат на рекламу) – по совету знакомых, увидев рекламу в любом рекламном источнике, но не воспользовались ссылкой и др.Стоимость привлечения пользователей разная от бесплатной (“органические пользователи”) и почти бесплатной - email-рассылки, до самой дорогой - контекстная реклама и реклама в социальных сетях; привлечение пользователей через блогеров по стоимости привлечения находится посередине.

Таким образом рекламные источники можно расположить в следующем порядке по убыванию стоимости привлечения:

1. контекстная реклама
2. социальные сети
3. реклама у блогеров
4. email-рассылки

Т.к. маркетингом занимается другая компания, нам важно провести собственное расследование и мы предлагаем вам стать нашими детективами. Вашей задачей будет разобраться с имеющимися данными о продажах и помочь нам принять обоснованное решение относительно маркетинговых кампаний в следующем году.

**Описание задачи:**

[**Ссылка на файл**](https://drive.google.com/file/d/1iQ273No2o7pe-zkh2Vkw3p0U65y7anjq/view?usp=sharing)**.**

Вы работаете продуктовым аналитиком в иностранной компании “Delivery Dron”, которая занимается продажей дронов доставщиков через интернет. Компания запустила продажи в четырех регионах в мае 2019 года. Инвесторы требуют отчета о вложенных средствах и необходимо строить планы на следующий год.

**Вам необходимо проанализировать имеющиеся данные и ответить на главный вопрос заказчика: в какой регион и какой рекламный канал инвестировать больше денег на привлечение клиентов в 2020 году и обосновать свои предложения?**

* **Описание данных:**Файл ecom\_yl.csv содержит данные о визитах и покупках пользователей сайта интернет магазина “Delivery Dron”. Каждая строка представляет визит пользователя. Данные отсортированы по Session Date.  
    
   **Столбцы набора данных:** 0: User Id - id пользователя  
    
   1: Region - регион визита  
    
   2: Device - устройство, с которого производится вход на сайт  
    
   3: Channel - рекламный канал перехода на сайт  
    
   4: Session Start - время начала пользовательской сессии (входа на сайт)  
    
   5: Session End - время окончания пользовательской сессии (уход с сайта)  
    
   6: SessionDurationSec - длительность сессии в секундах  
    
   7: Session Date - дата визита  
    
   8: Month - месяц визита  
    
   9: Day - день дня недели визита  
    
   10: Hour of day - час (от 0 до 24) визита  
    
   11: Order Dt - дата покупки  
    
   12: Revenue - сумма покупки  
    
   13: Payment Type - тип оплаты  
    
   14: Promo code - использование промокода на скидку 10% (есть-1, нет-0).
* **Исследовательский анализ данных и предобработка:** Для проведения исследования необходимо:  
  + названия столбцов привести в соответствии с PEP8
  + проверить данные на пропуски и заполнить их при необходимости. Обосновать решение.
  + проверить данные на дубликаты, в том числе на неявные дубликаты в категориальных данных
  + перевести тип данных столбцов с датой и временем в соответствующий формат при необходимости ([pd.to\_datetime()](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.to_datetime.html))
  + добавить столбец итоговой суммой покупки с учетом применения промокода на скидку 10%
  + определить исследуемый период, весь ли период брать для исследования или нет?
  + проверить данные на выбросы и адекватность данных
  + добавить столбец с указанием времени суток визита (утро 06:00-09:59, день 10:00-16:59, вечер 17:00-21:59, ночь 22:00-05:59)
  + добавить столбец “payer” с информацией о том, является ли пользователь платящим или нет.
  + Провести аналитический и графический анализ данных:
    - Доля продаж по регионам
    - Доля продаж по источникам
    - Доля продаж по устройствам
    - Количество пользователей с разбивкой на платящих/не платящих по регионам
    - Количество пользователей с разбивкой на платящих/не платящих по устройствам
    - Количество пользователей с разбивкой на платящих/не платящих по источникам
    - Графики, показывающие есть ли сезонность в продажах по месяцам, дням недели, времени суток
    - Диаграмма количества покупок по типу оплаты
* **Проведение расчётов**
  + Рассчитать средний чек.
  + Сколько покупок в среднем совершает 1 пользователь?
  + Рассчитать среднюю продолжительность сессии по рекламным каналам.
  + Рассчитать среднюю продолжительность сессии по типу устройства.
  + Определить топ-3 рекламных канала по среднему чеку.
  + Определить топ-3 региона по среднему чеку.
  + Определить топ-3 месяца по среднему чеку с разбивкой по регионам.
  + Посчитать MAU\* по каждому месяцу с разбивкой по рекламным каналам и выделить топ-3 рекламных каналов по количеству уникальных пользователей в месяц.

\*MAU (Monthly Active Users) — это количество уникальных пользователей за месяц.

* + Составить таблицу в которой будет указано по рекламным каналам: количество пользователей, количество уникальных пользователей, количество платящих пользователей, сумма продаж и определить какой источник “принес” больше всего платящих пользователей и большую сумму продаж.
* **Проверка гипотез**
  + Влияет ли тип устройства на количество покупок в день по каждому региону?
  + Влияет ли тип рекламного канала на количество покупок в день по каждому региону?
  + Проверить гипотезу о том, что средний чек отличается в зависимости от региона?
  + Проверить гипотезу о том, что средний чек отличается в зависимости от рекламного канала?
  + Проверить гипотезу о том, что средний чек отличается в зависимости от времени суток?
  + Есть ли взаимосвязь между продолжительностью сессии с суммой покупок?
  + Придумать и проверить еще не менее 2 гипотез, на основании гипотез команды.
* **Регрессионное моделирование**
  + Построить регрессионную модель продаж товаров, исходя из факторов продаж. При этом выбор и аргументация факторов - остается за командой.
* **Дашборд**
  + Создать дашборд, включающий основные метрики работы интернет магазина: индикаторы, графики и диаграммы (не менее 5), которые будут отвечать на следующие вопросы инвесторов:
  + сумма продаж, средний чек, количество пользователей, количество платящих пользователей, средняя продолжительность сессии
  + распределение пользователей по регионам и устройствам
  + график(и) количества пользователей и суммой продаж по месяцам
  + ключевые диаграммы - сумма продаж по рекламному каналу и количество привлеченных платящих пользователей с разбивкой по региону.
  + Добавить на дашборд фильтр по рекламному каналу, дате, типу региона и типу оплаты.
* **Выводы**
  + Необходимо оформить ключевые выводы и найденные инсайты в презентацию с наглядной визуализацией (до 10 слайдов не считая титульного и заключительного слайда).