Университет онлайн профессий Нетология

**ДИПЛОМНАЯ РАБОТА**

на тему:

"Исследование продаж британского e-commerce (поиск инсайтов, составление рекомендаций стейкхолдерам, построение модели)".

Профессия: Аналитик данных

Выполнила: студентка потока DA-21

Савченкова Екатерина Юрьевна

Москва 2021

Содержание

1. Введение……………………………………………………………………………………….3
2. Этапы предобработки…………………. ……………………………………………………4
3. Exploratory Data Analysis (EDA) ……………………………………………………………5
4. Заключение …………………………………………………………………………………...9

**I. Введение**

Нашему вниманию представлена база данных крупного Британского ритейлера электронной коммерции, которая содержит транзакции, совершенные за период 01/12/2010 - 09/12/2011. Основной акцент компания делает на продаже уникальных подарков на все случаи жизни. Многие клиенты компании являются оптовиками.

Фрейм данных состоит из восьми столбцов:

InvoiceNo: номер транзакции, состоит из 6 цифр. Номер, содержащий “С” является отмененной транзакцией (canceled)

StockCode: код товара (позиции). Уникальное 5-значное целое число, присваиваемое каждому отдельному продукту.

Description: описание товара

Quantity: количество

InvoiceDate: дата транзакции

UnitPrice: цена товара за единицу в фунтах стерлингов

CustomerID: номер клиента. Уникальное 5-значное целое число, присваиваемое каждому покупателю.

Country: название страны, в которой проживает каждый покупатель.

Основной **бизнес - целью** работы данного ритейлера является увеличение объемов продаж в ряде ключевых стран, а также расширение доли рынка.

Для достижения поставленной бизнес - цели была сформулирована **бизнес - задача:** исследовать и визуализировать ключевые показатели эффективности работы для того, чтобы выявить текущую ситуацию с объемами продаж за исследуемый период. На основе проведенного анализа мы разработаем модель, которая позволит спрогнозировать объемы на последующие 3 месяца.

Помимо описанной выше ключевой бизнес - цели можно также выделить маркетинговую цель, которая заключается в том, чтобы продумать программы поощрения для наиболее лояльных покупателей, а именно: триггерные или персонализированные email рассылки, накопительные карты, начисление бонусов за рекомендацию товара другу и т.п.

Хочется также сказать несколько слов, а какие же все-таки преимущества сможет получить бизнес по итогу достижения целей? Ими могут быть:

* Увеличится узнаваемость бренда;
* Возрастет авторитет компании;
* Улучшится конверсия сайта и др.
* Увеличится посещаемость сайта и т.д

Основными стейкхолдерами данного анализа являются генеральный директор, финансовый директор, а также маркетолог.

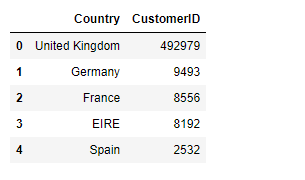
Очистка, предобработка и EDA анализ данных были проведены в Python.

**II. Предобработка** данных включала в себя следующие этапы:

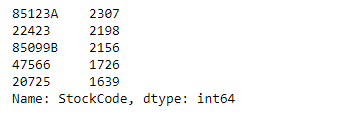
1. Меняем верхний регистр в написании слов в столбце Description на нижний регистр;
2. Выводим основные статистики и смотрим количество пропущенных значений;
3. Около 25% пропусков в столбце CustomerID. Принимаем решение заменить пропущенные значения на 0;
4. В столбце Description отсутствуют описания некоторых товаров. Предполагаем, что информация просто не была занесена в базу, либо случайным образом удалена. Поэтому заменяем пропущенные значения фразой no\_info. Прекрасно! Более пропусков в нашем датасете нет.
5. Транзакции, начинающиеся на “C” являются отмененными. Поскольку мы приняли решение не использовать их в нашем анализе, мы их удаляем.
6. Меняем тип колонок: CustomerID - на 'int64', колонку InvoiceDate приводим к типу datetime. Это принципиально важно для дальнейшего анализа и вывода нужных нам метрик;
7. Добавляем к нашему датасету дополнительную колонку Revenue, которая отражает общую сумму по каждому проданному товару. Таким образом, мы выводим первую ключевую метрику, которую будем также использовать для дальнейших вычислений и построения графиков;

**III**. Следующим этапом следует **Exploratory Data Analysis (EDA)**

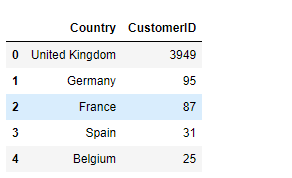
1. Выводим *TOP-5 best selling countries* (отсортировали 'UnitPrice > 0')



2. Хотим посмотреть на *TOP-5 most ordered products,* для того, чтобы понимать какие продукты пользуются наибольшим спросом среди пользователей сайта



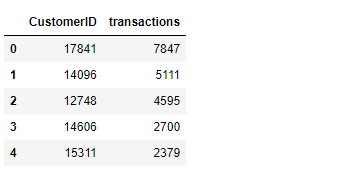
3. Выводим *TOP-5 countries by unique users.* Эта метрика дает понять, какие страны лидируют по числу пользователей, в первый раз посетивших сайт магазина



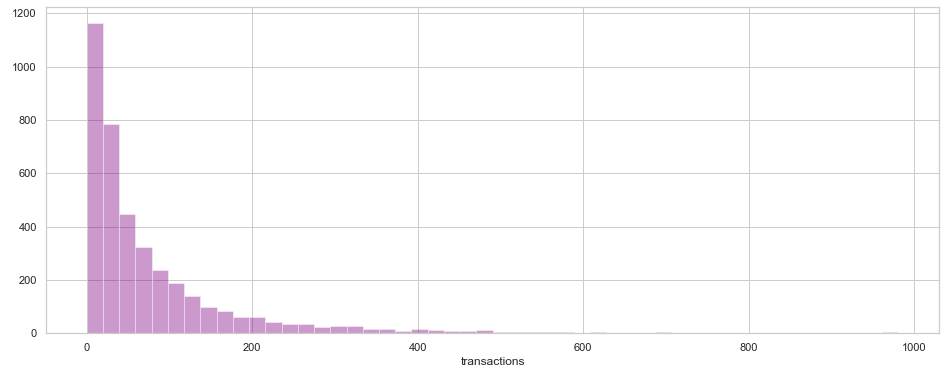
Сгруппировав наши данные по странам и выявив количество уникальных пользователей мы обнаружили, что более 90% покупателей приходится на Великобританию, и поэтому дальнейший анализ будем проводить на продажах магазина этой страны.

****

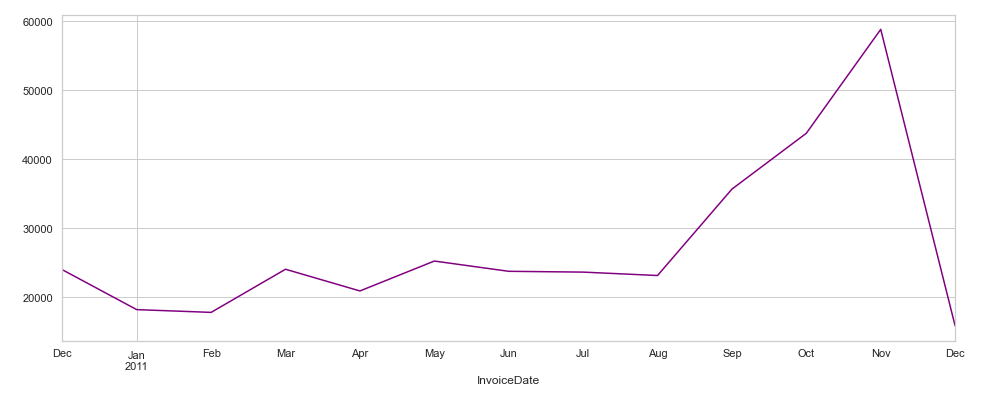
4. Выводим *Number of transactions per each user*. Данная метрика позволит нам выявить кто из пользователей совершил больше всего покупок за указанный период, и соответственно, принять решение о возможном поощрении. Как было озвучено выше, это могут быть разного рода программы лояльности, накопительная/бонусная карты, скидка постоянного клиента и т.д.



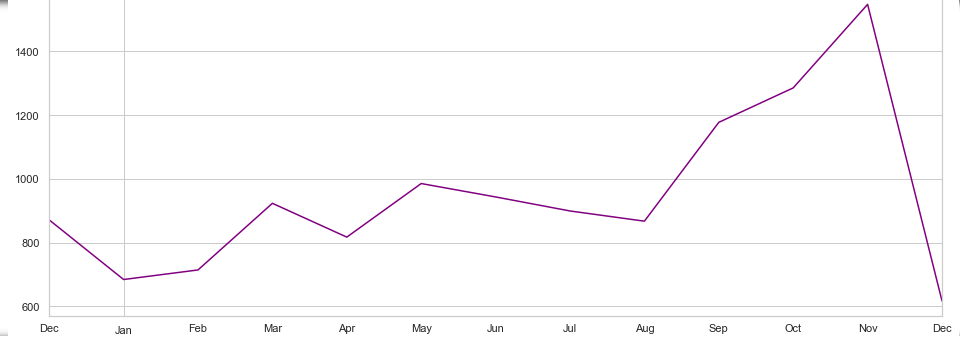
5. Из графика ниже видим, что более 1000 покупателей совершили около 20 транзакций



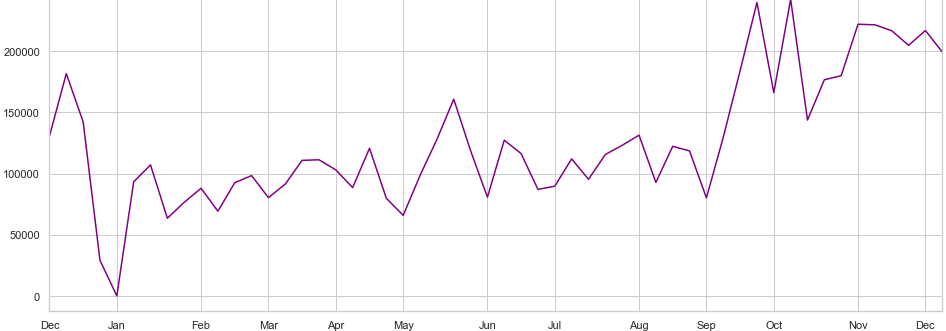
6. На графике ниже мы можем наблюдать как распределяются продажи относительно каждого месяца. Наибольший пик наблюдается в период с октября по ноябрь. Возможно это связано с тем, что люди массово покупают подарки к Рождеству и Новому году. Вместе с тем отмечается резкий спад продаж в конце декабря 2010 г. - январе 2011 г. Вероятнее всего в период Новогодних праздников магазин не работает.



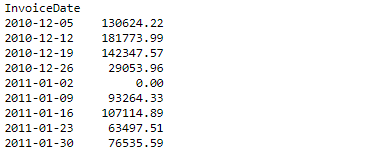
7. Далее посмотрим на то, в каком месяце в магазин пришло наибольшее количество пользователей. Видим, что аналогично графику выше - в ноябре - свыше 1400 чел.



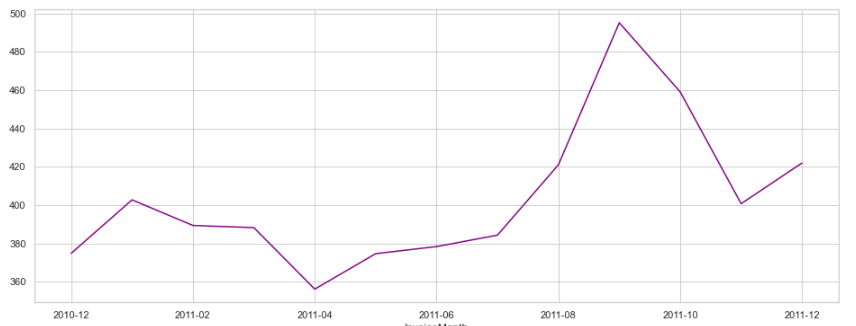
8. Следующий график демонстрирует нам объёмы продаж суммарно за каждый месяц в течение анализируемого периода.



Наибольший спад наблюдается в первую неделю января 2011 г., когда магазин закрыт из-за Новогодних праздников; а также в декабре того же года, когда очевидно, что люди, после многочисленных покупок решили сократили свои расходы.



9. Посчитаем метрику AOV («average order value») — средняя стоимость заказа. Иными словами, это средний чек — сумма, которую компания в среднем зарабатывает с одной покупки за конкретный период времени. Группируем суммарную выручку по месяцам и делим ее на количество заказов в месяце.



Зная средний чек, компания может прогнозировать выручку, которую приводят различные каналы трафика. Месяцем с наиболее высоким показателем среднего чека является сентябрь.

Для прогнозирования объемов продаж в последующие 3 месяца была применена модель линейной регрессии. Метрика оценки качества нашей модели (RMSE) показала, что в среднем модель может ошибаться на 258648 фунтов стерлинга. Сделав прогноз на период 2011.12 - 2012.02 мы видим, что объемы продаж примерно равны аналогичным месяцам предыдущего года. Это значит, что в целом, работа компании стабильна, резкого падения объемов продаж в начале следующего года не предвидится. Увеличению объемов продаж может способствовать выход на новый рынок сбыта.

**IV. Заключение**

Согласно метрике AOV видим, что месяцем с наиболее высоким показателем среднего чека является сентябрь. Показатель AOV можно увеличивать за счет продажи сопутствующих товаров. Как пример, при продаже верхней одежды можно предлагать клиентам купить водонепроницаемый чехол для одежды из инновационной сверх дышащей ткани. Стимулировать рост среднего чека для компании может быть выгоднее, чем привлекать новых клиентов. Готовому к покупке клиенту не нужно еще раз знакомиться с компанией, заново читать отзывы других покупателей или гуглить преимущества товара этого бренда.

Наибольшее количество покупок совершено в октябре - ноябре, то есть перед праздниками. Максимальное число вновь пришедших клиентов приходится как раз на этот же период. Разумеется, фактор сезонности нельзя исключать, но для того, чтобы продажи распределялись более менее равномерно в течение всего года, маркетологам следует уделить внимание поощрению наиболее лояльных клиентов, возможно также сделать акцент на сезонные акции.

Сгруппировав наши данные по странам и выявив количество уникальных пользователей мы обнаружили, что более 90% покупателей приходится на Великобританию. Таким образом, можем заключить, что если руководство компании видит своей целью увеличение объемов продаж, то выход на новые рынки, либо укрепление на уже существующих, однозначно необходим. Возможно маркетологам следует дополнительно проанализировать работу каналов привлечения трафика: какие из них дают наибольшую конверсию.