**Пояснительная записка**

Датасет по покупкам интернет-магазина состоит из 8 столбцов и 54214 строк. Есть номер платежа, код покупки, описание, количество единиц заказа по каждой категории, дата покупки, цена за единицу товара, ID покупателя и страна (скорее всего доставки).

Полный анализ датасета производился через Python (оболочка Jupiter Notebook). Производился анализ на момент наличия отрицательных значений, уникальных значений, пустых строк или неверно введенных данных. Оценивались доли данных в общем датасете, формировались гипотезы на основе данных и производились расчеты используемых метрик.

Пустые или незаполненные строки присутствуют в CustomerID и Description. Потеря идентификационного номера (CustomerID) пользователя более 25%, что много. Возможно ошибки произошли в результате сбоя ПО. Также замечено, что в столбце Quantity(Количество) имеются отрицательные значения,что можно расценивать или как ошибку сбора датасета или как возврат товара, скидку, брак. Данных значений более 34000, что является большой долей из общего количества. Разъяснений по данному моменту не было. Стоило бы уточнить у заказчика, чем именно являются отрицательные данные, чтобы их неправильная интерпретация не повлияла на анализ. Но так как у нас нет данной возможности, а доля данных отрицательных значений большая, то будем отталкиваться от второго определения (брак, возврат, скидки - т.е. все, что дает отрицательный доход).

В общем можно расценивать достаточность количества данных для проведения анализа, с допущением вышеуказанной поправки.

Основная бизнес задача - максимизация прибыли. Следовательно, будем искать точки роста бизнеса, куда можно направить силы компании. В данных отсутствует информация о затратах на производство, логистику, маркетинг, конкурентное сравнение по рынку и т.п. следовательно построить цельную иерархию метрик у нас не получится. В реальной ситуации рекомендовано обратиться к бизнес-заказчику для предоставления отсутствующих данных. Но будем исходить из того, что есть.

Описание заказчика исходя из датасета - это онлайн-магазин, специализирующийся на продажах, преимущественно собственного рынка, а именно Великобритании. Компания также отправляет заказы посредством доставки по всему миру, преимущественно в страны Европы.

Гипотетическая проблема - ориентация на внутренний рынок.

Цель - расширение компании посредством логистики, маркетинга на новые рынки (страны). Приоритетнее делать ставку на рынки с большим средним чеком и объемом заказа.

Используемые метрики для проведения анализа:

* Доход, отдельно по каждой стране (Revenue) - покажет топ стран с выручкой/средней выручкой
* Доход с пользователя за месяц (ARPU) - динамику/сезонность покупок
* Средний чек заказов (AOV) - в каких странах он больше (где выгоднее развиваться компании)
* Количество заказов по клиентам - какие клиенты лояльнее всего и чем их можно наградить, сделать из них “адвокатами” компании/бренда
* Средняя цена товара или средний чек (APP) - чем больше средний чек по заказам, тем лучше или стараемся его поднять, привлекаем для этого маркетологов.

Все эти метрики необходимы, чтобы мы могли дать заказчику ответ, в какой стране больше средний чек и объем заказов, и где меньшая доля покупок. Мы сможем указать, где приоритетнее присутствовать компании дабы увеличить выручку, где имеется спрос на продукцию компании, но с малой долей людей (например, в отличии от Великобритании). С помощью развития сети заказчик может поставить более амбициозные бизнес-цели, укрупнить свой бизнес, повысить доходность компании.

Часть метрик сделаем при расчете в Python, а часть через интерактивный дашборд Google Studio (ссылка будет в приложении). Этапы анализа, расчеты показателей и выводы можно посмотреть в файле Final\_work\_metrics.ipynb, но финальный вывод также продублируем в пояснительной записке.

Менеджменту компании стоит обратить внимание на тот факт, что страны с наивысшим средним чеком составляют малую долю, менее 2-х процентов. По средней сумме заказа в разрезе стран также входят с наименьшей долей. Это большая упущенная выгода для компании. Следует поднимать объемы заказов с данных стран, использовав маркетинг или другие способы. Следует обратить внимание на такие страны как: Гонконг (большой средний чек!), Португалия, Сингапур, Норвегия, Австралия, Нидерланды, Швеция. Графическое представление рекомендуется посмотреть в дашборде. Там же видно, если выставить фильтр по первой половине периода и второй половине, что наблюдается небольшой рост среднего чека и связано это в основном за счет спроса на новых рынках (странах), следовательно гипотеза подтверждается историческими данными по компании, но стоит уделить этому большее внимание.

Если имеются проблемы с логистикой в мировых масштабах и нет возможности точечно развиваться в странах, то можно предложить развитие рынка через ближайшие европейские страны. Также стоит более детально узнать у заказчика, что означают данные по объемам заказа с отрицательным значением. Если это возвраты, брак, то необходимо уделить внимание анализу и формулированию гипотез по уменьшению их объемов, поработать над качеством предоставляемой продукции (например, наладить взаимодействие с поставщиками).

Данный анализ можно рекомендовать менеджменту компании, ген. директору, отделу маркетинга.

Ссылка на Data Studio Google: <https://datastudio.google.com/reporting/8dbd7dd5-aec0-465b-a770-78c6a378df37>

Файл Final\_work\_metrics.ipynb в архиве отчета.