

Документация по проекту

Дипломная работа по теме:

"Исследование продаж в супермаркетах (поиск инсайтов, составление рекомендаций стейкхолдерам, построение дашборда)"

Профессия “Аналитик BI”, ABU-58

Хуснутдиновой Наталии Алексеевны

Пермь, 2024

Оглавление

1. Постановка задачи	3
1.1. Описание бизнес-задачи и ее актуальности	3
1.2. Круг стейкхолдеров, сбор бизнес-требований по задаче, гипотезы для проверки	3
1.3. Метрики для проверки гипотез, источник информации для сбора данных для расчета таких метрик:	4
2. Анализ исходных данных.....	4
2.1. Описание исходных данных	4
2.2. Алгоритмы и техники, применяемые для решения задачи, их требования к исходным данным	5
2.3. Форма представления итоговых результатов.....	5
3. Методика решения.....	6
3.1. Шаги преобразования и очистки данных.....	6
3.2. Выбор методов работы с данными и метрик для решения бизнес-задачи, построена система метрик	7
3.3. Сценарий проверки гипотез	7
3.4. Модель данных	7
4. Результаты.....	8
5. Выводы	11

1. Постановка задачи

1.1. Описание бизнес-задачи и ее актуальности

Заказчик дашборда – компания – собственник нескольких крупных супермаркетов, решила выйти на новый для неё рынок Мьянмы. Руководство хочет увеличить объём продаж, улучшить управление запасами товаров, повысить качество обслуживания клиентов. Для этого необходимо проанализировать реалии, существующие на рынке Мьянмы:

- динамику торговли,
- покупательское поведение потребителей супермаркетов,
- тенденции в области категорий продуктов.

1.2. Круг стейкхолдеров, сбор бизнес-требований по задаче, гипотезы для проверки

Стейкхолдеры:

- собственник и топ-менеджмент компании,
- маркетологи компании,
- отдел аналитики компании.

Бизнес-требования:

- на дашборде должны быть отражены продажи в супермаркетах по категориям продуктов, способам оплаты в разрезе по городам и месяцам / неделям. А также необходимо определить характеристики потенциальной целевой аудитории.

Гипотезы:

- покупки в категориях: модные аксессуары, дом и образ жизни, здоровье и красота – чаще совершают женщины, а покупки в категориях: электронные аксессуары, спорт и путешествия, еда и напитки – мужчины;
- кредитная карта – самый часто используемый способ оплаты;
- при оплате кредитной картой клиенты совершают покупки на большую сумму, нежели при оплате другими способами;
- в столице Нейпидо совершается большее количество продаж на большую сумму, нежели в других городах Мьянмы;
- постоянные клиенты приносят супермаркетам больше прибыли, чем новые;
- среди женщин больше постоянных покупателей, чем среди мужчин;
- в столице Нейпидо большее число покупателей с высоким рейтингом удовлетворённости, чем в других городах;
- у женщин выше уровень удовлетворённости покупками, нежели у мужчин;
- в утреннее и в вечернее время совершается большее число транзакций / покупок, чем днём.

1.3. Метрики для проверки гипотез, источник информации для сбора данных для расчета таких метрик:

- выручка;
- прибыль;
- количество транзакций (покупок);
- количество купленных товаров.

2. Анализ исходных данных

2.1. Описание исходных данных

Источник данных: [super-market-sales](#).

Табл. Исходные данные

Показатель	Описание	Проблемы
Invoice ID (идентификатор счёта)	Уникальный идентификатор для каждого счёта или транзакции	-
Branch (филиал)	Буквенное обозначение филиала города, где произошла транзакция (A, B, C)	-
City (город)	Город, в котором расположен филиал (Янгон, Мандалай, Нейпидо)	-
Customer type (тип клиента)	Указывает: является ли клиент постоянным или новым клиентом (member – постоянный, normal – новый)	-
Gender (пол клиента)	Male – мужской, female – женский	-
Product line (продуктовая линейка)	Категория/тип приобретённого продукта: модные аксессуары, еда и напитки, электронные аксессуары, спорт и путешествия, дом и образ жизни, здоровье и красота	-
Unit price (цена за единицу товара)	Цена одной единицы товара	- разделитель целой и дробной части - точка - значение указано в валюте Мьянмы – кьят. Для удобства восприятия в визуализациях заменены на у.е.
Quantity (количество)	Количество единиц приобретённого продукта	-

Tax 5% (налог 5%)	Сумма налога (5% от общей стоимости), применённая к транзакции	- разделитель целой и дробной части - точка - значение указано в валюте Мьянмы – кьят. Для удобства восприятия в визуализациях заменены на у.е.
Total (итого)	Общая стоимость транзакции, включая налоги. Посчитана как количество * цену за единицу * налог 5%.	- разделитель целой и дробной части - точка - значение указано в валюте Мьянмы – кьят. Для удобства восприятия в визуализациях заменены на у.е.
Date (дата)	Дата совершения транзакции	- указана в формате: мм/дд/гггг (локаль не России)
Time (время)	Время, когда произошла транзакция	-
Payment (Оплата)	Используемый способ оплаты (банковская карта, наличные, электронный кошелек)	-
Cogs (себестоимость проданных товаров)	Прямые затраты, связанные с производством или приобретением проданной продукции	- разделитель целой и дробной части - точка - значение указано в валюте Мьянмы – кьят. Для удобства восприятия в визуализациях заменены на у.е.
Gross margin percentage (процент валовой прибыли)	Процент прибыли от транзакции	-
gross income (валовой доход)	Общая прибыль, полученная от транзакции. Посчитана как: итого минус себестоимость проданных товаров.	- разделитель целой и дробной части - точка - значение указано в валюте Мьянмы – кьят. Для удобства восприятия в визуализациях заменены на у.е.
Rating (рейтинг)	Удовлетворённость клиентов, отзывы о сделке. Значение указано в баллах (максимум 10).	- разделитель целой и дробной части - точка

2.2. Алгоритмы и техники, применяемые для решения задачи, их требования к исходным данным

Алгоритмы и техники: дашборд.

Требования к исходным данным:

- корректное указание типов данных;
- отсутствие ошибок в данных.

2.3. Форма представления итоговых результатов

Дашборд, выполненный в MS Power BI Desktop.

3. Методика решения

3.1. Шаги преобразования и очистки данных

Очистки данных не потребовалось.

Табл. Преобразование данных

Показатель	Шаги
Параметры текущего файла	Локаль заменена с России на Английский (США) для интерпретации чисел, дат и времени в импортированном тексте из файла с исходными данными
Total / gross income / Rating	Установлен тип данных – десятичное число
Unit price / Tax 5% / cogs/ Gross margin percentage	Показатели исключены из дашборда с целью оптимизации модели (так как для расчётов негодились)
	Все оставшиеся столбцы с данными разделены на 3 таблицы для удобства восприятия и целей визуализации. Названия столбцов переименованы на русском языке.
Таблица «Справочник городов»	- Добавлены столбцы «Филиал» и «Город» - В каждом столбце оставлены лишь уникальные значения
Таблица «Клиенты»	- Добавлены столбцы «Идентификатор», «Тип клиента», «Пол», «Категория продукта», «Способ оплаты», «Рейтинг удовлетворённости»
Таблица «Клиенты»	- Добавлен новый расчётный столбец, который выводит для каждого покупателя его уровень удовлетворённости, в соответствии с рейтингом: «Низкий» - от 4 до 5,9 баллов; «Средний» - от 6 до 7,9 баллов; «Высокий» - от 8 до 10 баллов
Таблица «Календарь»	- Дополнительно создана таблица-справочник «Календарь» для удобства создания визуализаций - Она содержит все даты, без пропусков, имеющиеся в таблице «Продажи» - Выведены в отдельные столбцы данные по месяцам, номерам недель и дней
Таблица «Клиенты»	- Из столбца «Time» выделены в отдельный столбец данные с часом совершения продажи – «Час продажи» - Добавлен новый расчётный столбец «Время суток», который выводит для каждого покупателя время, в которое он совершил покупку: «Утро» - с 10 до 12 часов; «День» - с 13 до 16 часов; «Вечер» - с 17 до 20 часов - Столбец «Time» удалён из модели
Таблица «Продажи»	- Добавлены столбцы «Идентификатор», «Филиал», «Количество», «Выручка (включая налог)», «Дата», «Прибыль»
	Все неиспользуемые в визуализациях столбцы с данными скрыты для удобства визуализирования

3.2. Выбор методов работы с данными и метрик для решения бизнес-задачи, система метрик

- 1) Выручка считается как сумма по столбцу «Выручка (включая налог)» в таблице «Продажи», при заданных фильтрах;
- 2) Прибыль считается как сумма по столбцу «Прибыль» в таблице «Продажи», при заданных фильтрах;
- 3) Количество транзакций / покупок считается как количество уникальных значений в столбце «Идентификатор» в таблице «Продажи» при заданных фильтрах;
- 4) Количество товаров считается как сумма по столбцу «Количество» в таблице «Товары», при заданных фильтрах.

Все 4 метрики считаются по неделям и месяцам (с возможностью детализации по дням) (с января по март 2019г.);

Все 4 метрики распределены по трём городам: Янгон, Мандалай, Нейпидо;

Все 4 метрики распределены по способам оплаты: банковская карта, наличные, электронный кошелёк.

Для характеристики потенциальной целевой аудитории выведены следующие метрики:

- 1) количество совершённых транзакций / покупок в разное время суток по полу покупателей;
- 2) количество купленных товаров по категориям по полу покупателей;
- 3) количество покупателей с разным уровнем удовлетворённости по городам;
- 4) размер прибыли, полученной от постоянных / и новых покупателей, по городам.

Все 4 метрики считаются с возможностью разбивки по месяцам (с января по март 2019г.);

Все 4 метрики распределены по трём городам: Янгон, Мандалай, Нейпидо;

Все 4 метрики считаются с возможностью выделить: только мужчин / только женщин / обе категории вместе.

3.3. Сценарий проверки гипотез

Гипотезы проверяются путем сравнения значений метрик при установке необходимых фильтров в дашборде.

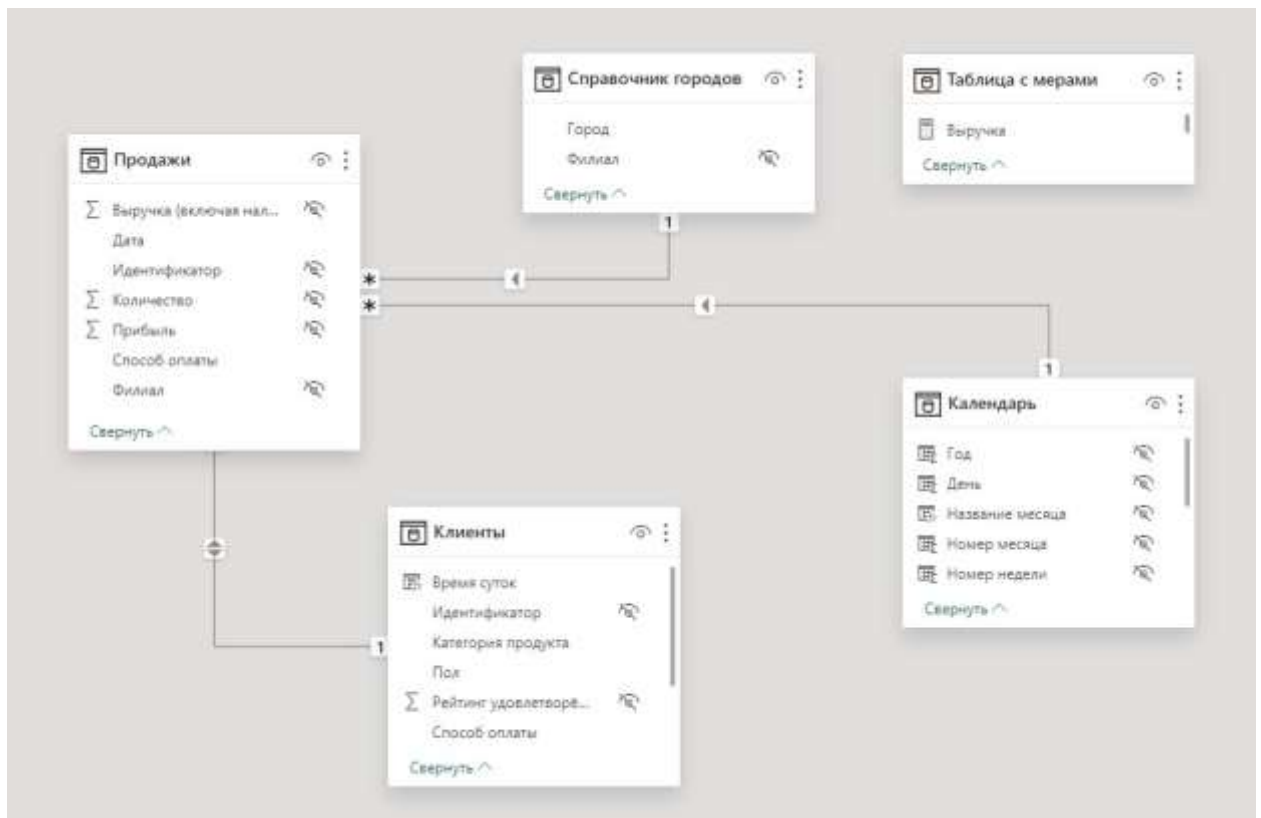
3.4. Модель данных

Модель данных включает 5 таблиц: «Справочник городов», «Клиенты», «Продажи», «Календарь» и таблицу мер.

Между таблицами «Продажи» и «Справочник городов» установлена связь «Многие к одному» по полю «Филиал».

Между таблицами «Продажи» и «Клиенты» установлена связь «Один к одному» по полю «Идентификатор».

Между таблицами «Продажи» и «Календарь» установлена связь «Многие к одному» по полю «Дата».



4. Результаты

Дашборд состоит из 2х страниц и имеет 26 визуальных элементов, пронумерованных цифрами красного цвета. На Рисунках 1-2 представлен внешний вид дашборда.



Рис. 1. Внешний вид дашборда

Далее описан каждый из 17-ти визуальных элементов первой страницы дашборда:

1. Заголовок первой страницы дашборда «Анализ и динамика продаж в супермаркетах Мьянмы в 2019г.»
2. Многострочная карточка, которая содержит сумму выручки, прибыли, количество транзакций и количество товаров, с учетом выставленных пользователем фильтров.
3. Срез данных по месяцам, где пользователь может установить интересующий его месяц, либо сразу несколько. Срез синхронизирован со 2ой страницей дашборда, при перемещении между страницами – фильтрация сохранится.
4. Срез данных по категориям товаров, где пользователь может выбрать одну или несколько интересующих его категорий.
5. График демонстрирует общую выручку в у.е. Мьянмы по неделям и месяцам с учетом выставленных пользователем фильтров.
6. График демонстрирует прибыль с продаж в у.е. Мьянмы по неделям и месяцам с учетом выставленных пользователем фильтров.
7. График демонстрирует количество транзакций / совершённых покупок по неделям и месяцам с учетом выставленных пользователем фильтров.
8. График демонстрирует количество проданных товаров (в штуках) по неделям и месяцам с учетом выставленных пользователем фильтров.
9. Гистограмма с накоплением показывает общую выручку в у.е. Мьянмы с распределением по городам с учетом выставленных пользователем фильтров.
10. Гистограмма с накоплением показывает прибыль с продаж в у.е. Мьянмы с распределением по городам с учетом выставленных пользователем фильтров.
11. Гистограмма с накоплением показывает количество транзакций / совершённых покупок с распределением по городам с учетом выставленных пользователем фильтров.
12. Гистограмма с накоплением показывает количество проданных товаров (в штуках) с распределением по городам с учетом выставленных пользователем фильтров.
13. Линейчатая диаграмма с накоплением отражает размер выручки в у.е. Мьянмы с распределением по способу оплаты с учётом выставленных пользователем фильтров.
14. Линейчатая диаграмма с накоплением отражает прибыль с продаж в у.е. Мьянмы с распределением по способу оплаты с учётом выставленных пользователем фильтров.
15. Линейчатая диаграмма с накоплением отражает количество транзакций / совершённых покупок с распределением по способу оплаты с учётом выставленных пользователем фильтров.
16. Линейчатая диаграмма с накоплением отражает количество проданных товаров (в штуках) с распределением по способу оплаты с учётом выставленных пользователем фильтров.
17. Кнопка перехода на следующую страницу – для удобства работы пользователя с отчётом.

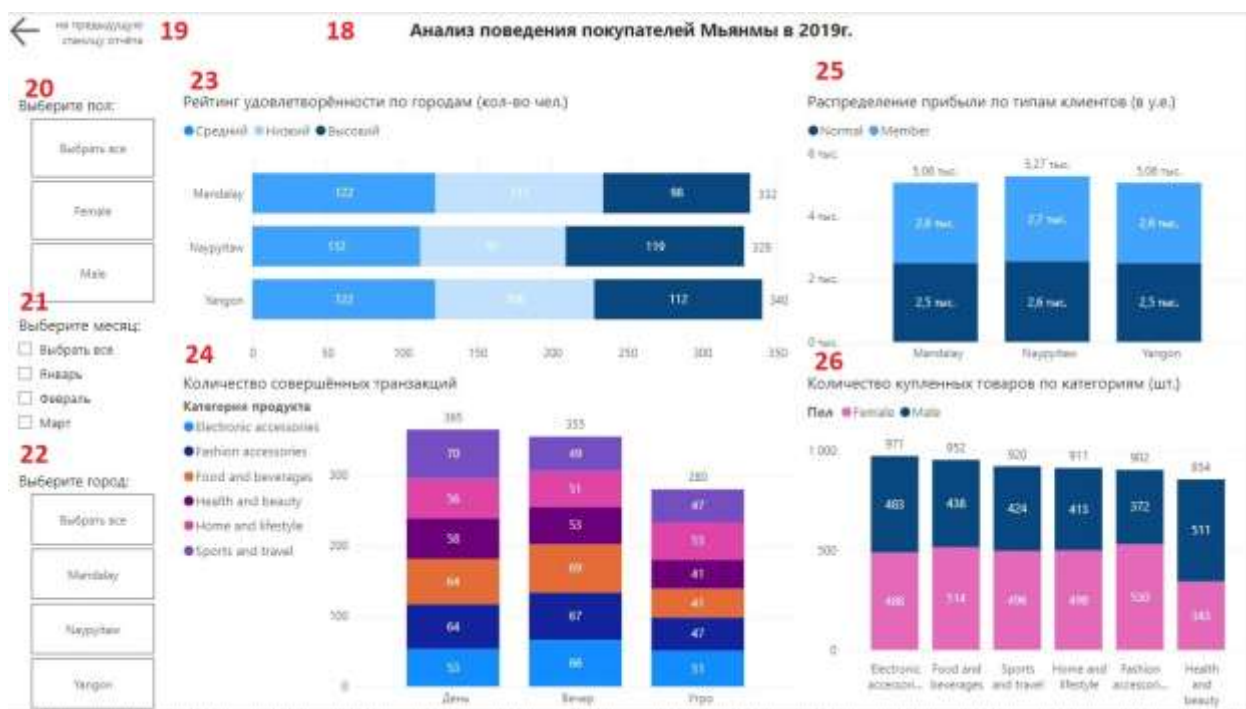


Рис. 2. Внешний вид дашборда

Далее описан каждый из 9-ти визуальных элементов второй страницы дашборда:

- Заголовок второй страницы дашборда «Анализ поведения покупателей Мьянмы в 2019г.»
- Кнопка возврата на предыдущую страницу – для удобства работы пользователя с отчётом.
- Срез данных в формате кнопок по полу покупателя, где пользователь может выбрать пол.
- Срез данных по месяцам, где пользователь может установить интересующий его месяц, либо сразу несколько. Срез синхронизирован с 1ой страницей дашборда, при перемещении между страницами – фильтрация сохранится.
- Срез данных в формате кнопок по городам осуществления продаж, где пользователь может выбрать один или несколько городов.
- Линейчатая диаграмма с накоплением показывает количество человек с разным уровнем удовлетворённости покупками в разбивке по городам с учётом выставленных пользователем фильтров.
- Гистограмма с накоплением показывает количество совершённых транзакций / покупок в разное время суток с разбивкой по категориям товаров с учётом выставленных пользователем фильтров.
- Гистограмма с накоплением показывает распределение прибыли в у.е. Мьянмы по типам покупателей (постоянные / новые) в разбивке по городам с учётом выставленных пользователем фильтров.
- Гистограмма с накоплением показывает количество купленных товаров (в штуках) по категориям товаров и по полу покупателей с учётом выставленных пользователем фильтров.

Элементы имеют следующие взаимодействия:

- Графики (5), (6), (7), (8) являются фильтрами для многострочной карточки.
- Графики (5), (6), (7), (8) являются выделителями для гистограмм с накоплением (9), (10), (11), (12).
- Графики (5), (6), (7), (8) являются выделителями для линейчатых диаграмм с накоплением (13), (14), (15), (16).
- Графики (5), (6), (7), (8) не влияют друг на друга.
- Гистограммы с накоплением (9), (10), (11), (12) являются фильтрами для многострочной карточки, а также для графиков (5), (6), (7), (8).
- Гистограммы с накоплением (9), (10), (11), (12) являются выделителями друг для друга, а также для линейчатых диаграмм с накоплением (13), (14), (15), (16).
- Линейчатые диаграммы с накоплением (13), (14), (15), (16) являются фильтрами для многострочной карточки, а также для графиков (5), (6), (7), (8).
- Линейчатые диаграммы с накоплением (13), (14), (15), (16) являются выделителями друг для друга, а также для гистограмм с накоплением (9), (10), (11), (12).
- Линейчатая диаграмма с накоплением (23) является выделителем для гистограмм с накоплением (24), (25) и (26).
- Гистограмма с накоплением (24) является выделителем для линейчатой диаграммы с накоплением (23), а также для гистограмм с накоплением (25) и (26).
- Гистограмма с накоплением (25) является выделителем для линейчатой диаграммы с накоплением (23), а также для гистограмм с накоплением (24) и (26).
- Гистограмма с накоплением (26) является выделителем для линейчатой диаграммы с накоплением (23), а также для гистограмм с накоплением (24) и (25).

5. Выводы

- В столице Нейпидо объём выручки и прибыли несущественно больше, чем в других городах. Можно сделать вывод о достаточно равномерном распределении этих метрик между всеми тремя городами Мьянмы.
- 4 графика с метриками, расположенные рядом друг с другом подтверждают существующую прямую зависимость между всеми 4мя показателями: чем выше объём выручки, тем выше объём прибыли, количество проданных товаров и число транзакций; и наоборот.
- Данные на графиках динамики расположены достаточно равномерно, разброс небольшой. Нулевой выручки и прибыли нет. Так же как нет каких-либо существенных подъёмов и спадов в продажах. Самые минимальные данные мы видим на 5 и 9 неделях продаж (общая выручка за 3 месяца по 3м городам: 6 585, 48 у.е. и 9 194, 67 у.е.; общая прибыль: 313, 59 у.е. и 437, 84 у.е., соответственно). Самые максимальные данные мы видим на 6 и 10 неделях продаж (общая выручка за 3 месяца по 3м городам: 29 428,74 у.е. и 30 108,78 у.е.; общая прибыль 1 401, 37 у.е. и 1 433,75 у.е., соответственно).
- Распределение выручки, прибыли, количества транзакций и объёма продаж по городам довольно равномерное. Нет явных лидеров и аутсайдеров. Можно сделать вывод о том, что продажи во всех трёх городах Мьянмы идут достаточно хорошо.
- Наиболее распространённым способом оплаты в Мьянме являются наличные денежные средства, наименее распространённым – банковская карта.

- При оплате наличными покупатели приносят супермаркетам больше выручки, чем при оплате банковскими картами.
- В столице Нейпидо большее число покупателей с высоким рейтингом удовлетворённости, чем в других городах.
- У женщин несущественно больше количество человек с высоким уровнем удовлетворённости, нежели у мужчин (168 против 161).
- В Нейпидо среди женщин больше постоянных покупателей, чем среди мужчин. В остальных городах распределение по полу среди постоянных и новых покупателей равномерно.
- Прибыль распределяется равномерно среди постоянных и новых клиентов.
- Женщины больше всего товаров приобретают в категориях: модные аксессуары, еда и напитки, дом и образ жизни, спорт и путешествия.
- Мужчины больше всего товаров приобретают в категориях: здоровье и красота, электронные аксессуары, еда и напитки.
- В целом наибольшее количество товаров приобретается в категории: электронные аксессуары, наименьшее количество товаров в категории: здоровье и красота.
- В утреннее время чаще всего покупают товары в категориях: дом и образ жизни, электронные аксессуары; в дневное время: спорт и путешествия, еда и напитки, модные аксессуары; в вечернее время: еда и напитки, модные аксессуары.
- В дневное время совершается большее количество транзакций / покупок, нежели в вечернее / или утреннее.

Рекомендации:

- На основании проведенного исследования рекомендуется открывать супермаркет в любом из трёх городов Мьянмы: Янгон, Мандалай, Нейпидо.
- Возможно отдать предпочтение столице Нейпидо.
- Приоритет следует отдать наличному расчёту с покупателями.
- В дневное время суток прогнозируется самый большой поток покупателей, в связи с чем необходимо обеспечить достаточное количество продавцов и товаров.
- Самый широкий товарный ряд рекомендуется делать у категорий: электронные аксессуары, еда и напитки, спорт и путешествия.
- При проведении акций, распродаж, внедрении программ лояльности – акцент рекомендуется делать на женскую аудиторию.
- Таким образом, рынок Мьянмы является перспективной нишей, и открытие здесь нового супермаркета позволит решить поставленные перед компанией задачи.