Changers

Pet-проєкт

На тему: аналіз даних про покупки клієнтів в 10 найпопулярніших торгівельних центрах міста Стамбул

Виконав: Мальчев Павло

Зміст

Вступ	3
Мета проєкту	3
Важливість проєкту	3
Методологія	4
Вибір теми	4
Використані інструменти	4
Методи аналізу	5
Збір та очищення даних	5
Джерела даних	5
Основні етапи очищення	5
Аналіз даних	10
Методи аналізу	10
Отримані результати	11
Візуалізація та дашборди	12
Створені візуалізації	12
Приклади використання дашбордів	15
Висновки та рекомендації	18
Головні висновки	18
Рекомендації для бізнесу	19

Вступ

Мета проєкту є аналіз даних про покупки у 10 торгових центрах Стамбула. Ця тема є надзвичайно важливою, оскільки дозволяє отримати глибоке розуміння поведінки споживачів, що є критичним для розробки ефективних маркетингових стратегій та оптимізації асортименту товарів. Вивчення вподобань клієнтів, популярності різних категорій товарів, а також методів оплати надає можливість торговим центрам не лише підвищити свої продажі, але й значно покращити загальний клієнтський досвід, що є важливим фактором у сучасних умовах конкуренції

Важливість проєкту полягає у можливості виявлення ключових трендів та сезонних змін у покупках. Це допомагає торговим центрам ефективно планувати свої маркетингові кампанії, проводити акції та спеціальні пропозиції у найбільш оптимальні періоди. Наприклад, знання про те, що грудень є піковим місяцем для покупок, дозволяє зосередити маркетингові зусилля саме в цей час, що може значно збільшити обсяги продажів і залучити більше клієнтів. Крім того, розуміння щомісячних і тижневих патернів покупок дозволяє краще управляти запасами і персоналом, дозволяє краще управляти запасами і персоналом, забезпечуючи високу якість обслуговування у будь-який час року

Отже, цей проєкт має на мету не лише підвищення продажів, але й покращення загального досвіду клієнтів, що є ключовим для довгострокового успіху торгових центрів. Використання отриманих даних допоможе приймами більш обґрунтовані рішення у сфері торгівлі, що сприятиме сталому розвитку бізнесу. Застосування аналітичних

інструментів для розуміння поведінки покупців дозволяє отримати конкурентні переваги, забезпечуючи більш високу ефективність маркетингових заходів та управління запасами, що в кінцевому рахунку сприяє зростанню прибутковості та зміцнення позицій на ринку

Методологія

Вибір теми

Причина вибору цієї теми пов'язана з прагненням детально вивчити вплив різних факторів на поведінку покупців. Аналіз гендерних та вікових груп, популярність різних торгових центрів, категорій товарів та методів оплати дає змогу створити більш цілеспрямовані та персоналізовані пропозиції для клієнтів. Це, в свою чергу, сприяє підвищенню лояльності клієнтів та зростанню їхньої задоволеності від покупок. Розуміння того, як різні групи соживачів взаємодіють з різними категоріями товарів та послуг, дозволяє точніше сегментувати ринок та адаптувати маркетингові стратегії під конкретні потреби різних аудиторій

Використані інструменти

- Мова програмування **Python** та її бібліотеки **Pandas**, **Matplotlib**, **Seaborn** для первинної обробки та аналізу даних
- **Jupyter Notebook** використувався як середа розробки з можливістю створення графіків та таблиць, що є дуже зручним при аналізі
- **PostgreSQL** використовувся для роботи з базою даних, написання додаткових запитів
- **Tableau Public** для створення графіків та в кінцевому рахунку цілісних дашбордів по кожному направленню дослідження

- Word для написання детального звіту по проєкту
- Google Slides для створення презентації

Методи аналізу

- Первинна обробка у Python, ознайомлення з датасетом, очищення, фільтрація та обробка необхідних значень
- Первинний аналіз даних у Python, виявлення основних статистичних показників датасету, розрахунок основних метрик
- Побудова графіків за допомогою бібліотек Matplotlib та Seaborn
- SQL запити у PostgreSQL для написання запитів до бази даних та отримання додаткових таблиць для аналізу
- Використання Tableau Public для створення основних візуалізацій для демонстрації цілісної картини аналізу
- Обробка всіх отриманих інсайтів виявлених при аналізі та створення висновків на основі отриманої інформації

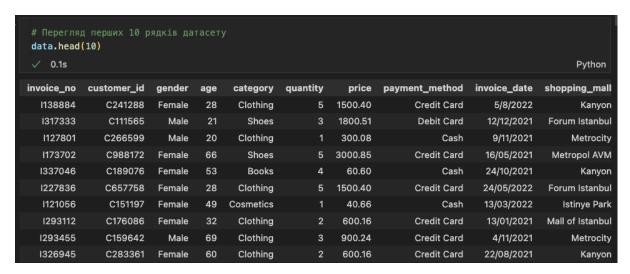
Збір та очищення даних

Джерела даних

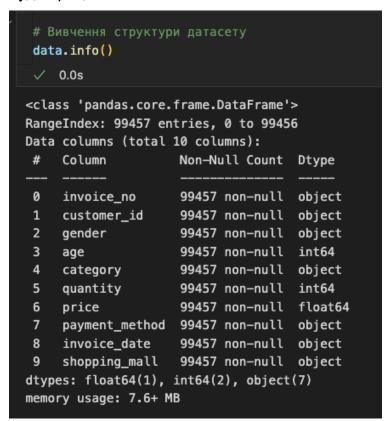
Дані були отримані з джерела <u>Kaggle</u>, яке містить інформацію про покупки клієнтів у різних торгових центрах Стамбула

Основні етапи очищення

• Ознайомлення з датасетом у Python, для первинного ознайомлення. Для цього на самому початку було виведено 10 перших рядків таблиці:



 Вивчення структури датасету, розуміння з якими типами даних потрібно буде працювати:



• Перевірка даних на null-значення та дублікати:

```
# Перевірка на наявність null-значень
 data.isnull().sum()
 √ 0.0s
invoice_no
                  0
customer_id
                  0
gender
                  0
                  0
age
category
                  0
quantity
                  0
                  0
price
payment_method
                  0
                  0
invoice_date
                  0
shopping_mall
dtype: int64
 # Перевірка на наявність дублікатів
 data.duplicated().sum()
 ✓ 0.3s
```

• Проведення необхідних перетворень даних для подальшого аналізу

```
# Перетворення колонки "invoice_date" у формат datetime64

data['invoice_date'] = pd.to_datetime(data['invoice_date'], dayfirst=True)

# Додавання стовпця "price_usd" який показує ціну у доларах замість турецьких лір

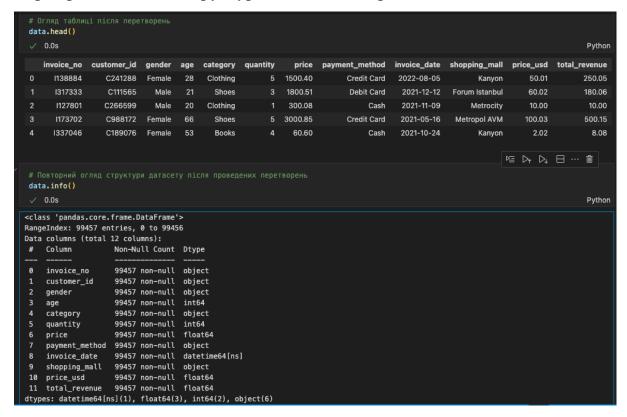
data['price_usd'] = round(data['price'] / 30, 2)

# Додавання стовпця "total_revenue"

data['total_revenue'] = round(data['quantity'] * data['price_usd'], 2)

✓ 0.8s
```

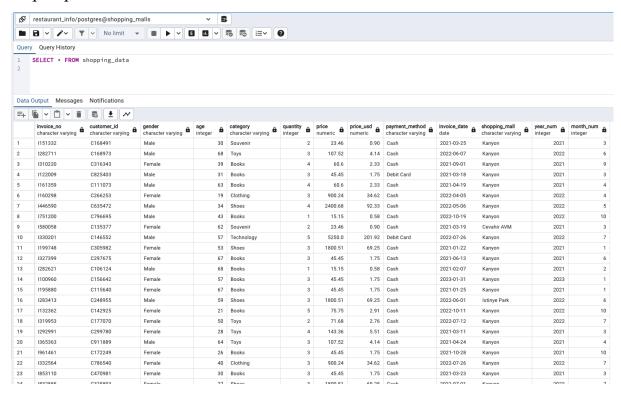
• Перевірка таблиці та структури даних після проведених змін:



• Наступним кроком було завантаження даних у PostgreSQL:

```
Query Query History
1 ~ CREATE TABLE shopping_data_new AS
     SELECT invoice_no,
3
             customer_id,
             gender,
4
5
            age,
6
            category,
7
            quantity,
8
            price,
            ROUND(price / 26.0, 2) AS price_usd,
9
            payment_method,
10
11
             invoice_date,
12
             shopping_mall,
13
            year_num,
14
            month_num
15
     FROM shopping_data;
16
17
     DROP TABLE shopping_data;
18
19
     ALTER TABLE shopping_data_new RENAME TO shopping_data;
20
Data Output Messages Notifications
ALTER TABLE
Query returned successfully in 1 secs 484 msec.
```

• Перевірка завантажених даних:



В результаті була отримана таблиця з наступною структурою:

Назва стовпця	Тип даних	Опис
invoice_no	text	номер транзакції
customer_id	text	айді клієнта
gender	text	стать
age	integer	вік
category	text	категорія товару
quantity	integer	кількість товарів
price	decimal	ціна за одиницю
price_usd	decimal	конвертація у долари
payment_method	text	платіжний метод
invoice_date	date	дата транзакції
shopping_mall	text	торговий центр

Аналіз даних

Методи аналізу

• Розрахунки основних статистичних показників та метрик у Python:

# Розрахунок основних статистичних показників data.describe() ✓ 0.4s						
	age	quantity	price	invoice_date	price_usd	
count	99457.000000	99457.000000	99457.000000	99457	99457.000000	
mean	43.427089	3.003429	689.256321	2022-02-04 02:46:59.783424	22.975059	
min	18.000000	1.000000	5.230000	2021-01-01 00:00:00	0.170000	
25%	30.000000	2.000000	45.450000	2021-07-19 00:00:00	1.520000	
50%	43.000000	3.000000	203.300000	2022-02-05 00:00:00	6.780000	
75%	56.000000	4.000000	1200.320000	2022-08-22 00:00:00	40.010000	
max	69.000000	5.000000	5250.000000	2023-03-08 00:00:00	175.000000	
std	14.990054	1.413025	941.184567	NaN	31.372892	

```
Загальна кількість продажів: 298712 одиниці
Загальний дохід (в доларах): 8383429.0$

Середній чек (в доларах): 84.29$

Найпопулярніша категорія за сумою продажів (в доларах): Clothing — 3799847.17$

Найменш популярна категорія за сумою продажів (в доларах): Souvenir — 21139.95$

Найпопулярніший торговий центр за сумою продажів (в доларах): Mall of Istanbul — 1695730.8$

Найбільша аудиторія (жінки чи чоловіки): Female — 59482

Середній вік клієнтів: 43.43

Мінімальний вік клієнтів: 18

Максимальний вік клієнтів: 69

Медіана віку клієнтів: 43.0

Найпопулярніший спосіб оплати: Cash — 44447 покупок, 144655.01$

Торговий центр з найбільшою кількістю покупців: Emaar Square Mall — 4811
```

- Побудова графіків розподілу середнього чеку, загальної суми покупок та карту кореляцій:
- Використання SQL для витягення додаткових таблиць з бази даних
- Побудова основних графіків та дашбордів для більш поглибленого аналізу даних

Отримані результати

- **Розподіл за віком** гістограма розподілу віку клієнтів показала, що більшість клієнтів знаходятся в діапазоні від 20 до 40 років. Це свідчить, що основна цільова аудиторія торгових центрів складається з молоді.
- **Розподіл за статтю** аналіз виявив, що жінки складають більшу частину клієнтів торгових центрів
- Розподіл за категоріями товарів одяг, взуття та косметика виявилися найпопулярнішими категоріями серед покупців. Це вкає на те, що ці категорії є основними драйверами продажів у торгових центрах
- Розподіл за методами оплати банківські картки є
 найпопулярнішим методом оплати серед клієнтів, що свідчить про
 зручність та перевагу безготівкових розрахунків
- **Кореляційний аналіз** показав, що існує слабка кореляція між ціною товару, кількістю та віком клієнтів. Проте середня кореляція була виявлена між ціною та середнім чеком
- Продажі за торговими центрами результати показали, що найбільша кількість покупок здійснюється у таких торгових центрах, як Kanyon, Mall of Istanbul та Metrocity

- Середня сума покупок по торговим центрам варіюється в залежності від торгового центру, причому найвищі значення спостерігаються в Kanyon та Mall of Istanbul
- Розподіл покупок за категоріями товарів показав, що в більшості торгових центрів найбільш популярними категоріями є взуття та одяг

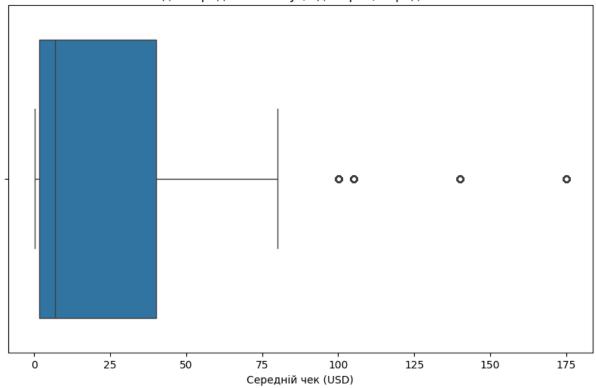
Візуалізація та дашборди

Створені візуалізації

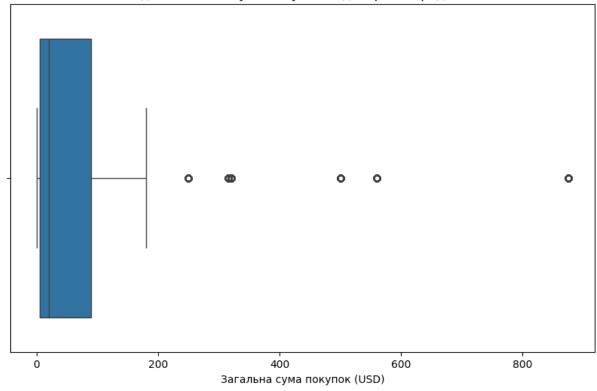
• Розподіл середнього чеку серед клієнтів та розподіл загальної суми покупок серед клієнтів:

```
# Групування даних за customer id
customer_group = data.groupby('customer_id').agg({
  'price_usd': 'mean',
                          # Середній чек у доларах
  'quantity': 'sum'
}).reset_index()
# Перейменування колонок
customer_group.columns = ['customer_id', 'average_purchase_value', 'total_quantity']
# Обчислення загальної суми покупок для кожного клієнта в доларах
customer_group['total_purchase_value'] = customer_group['average_purchase_value'] *
customer_group['total_quantity']
print(customer_group.head())
# Боксплот за середнім чеком у доларах
plt.figure(figsize=(10, 6))
sns.boxplot(x=customer_group['average_purchase_value'])
plt.title('Розподіл середнього чеку (в доларах) серед клієнтів')
plt.xlabel('Середній чек (USD)')
plt.show()
# Боксплот за загальною сумою покупок у доларах
plt.figure(figsize=(10, 6))
sns.boxplot(x=customer_group['total_purchase_value'])
plt.title('Розподіл загальної суми покупок (в доларах) серед клієнтів')
plt.xlabel('Загальна сума покупок (USD)')
plt.show()
```

Розподіл середнього чеку (в доларах) серед клієнтів



Розподіл загальної суми покупок (в доларах) серед клієнтів



На основі даних графіків можна зробити висновок, що медіана середнього чека знаходиться приблизно на рівні 25 доларів. Це означає, що половина клієнтів витрачає в середньому за одну покупку менше цієї суми, а

половина — більше. А медіана загальної суми покупок знаходиться приблизно на рівні 100 доларів.

Основна маса клієнтів витрачає відносно невеликі суми як за одну покупку (середній чек), так і загалом (загальна сума покупок). На обох графіках ϵ значна кількість викидів, що вказу ϵ на наявність різних типів покупців з різними звичками та можливостями витрат.

• Побудова теплової мапи кореляцій:

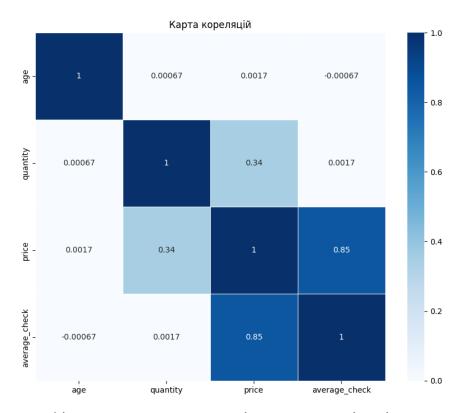
```
# Розрахунок середнього чеку
data['average_check'] = data['price'] / data['quantity']

# Створемо масив з даними для побудови кореляційної мапи
correlation_data = data[['age', 'quantity', 'price', 'average_check']]

# Розрахунок кореляції
correlation_matrix = correlation_data.corr()

# Побудова теплової кореляційної мапи
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='Blues', linewidths=0.5)
plt.title('Карта кореляцій')

plt.show()
```

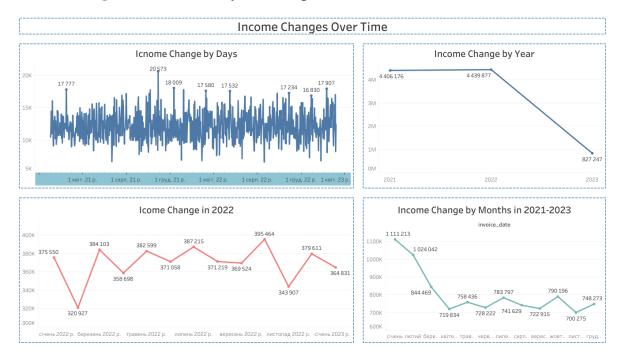


Даний графік показав, що кореляція присутня між ціною покупки та кількістю

Приклади використання дашбордів

Було створено 6 дашбордів:

• **Time Changes** – який показує зміни різних показників з часом:



Даний дашборд допоміг виявити сезонні зміни та оцінити прибутковість торгових центрів у різних роках та місяцях. Наприклад, було виявлено, що січень ϵ найприбутковішим місяцем, а листопад найменш прибутковим.

• **Age Gpoups** – дашборд, який демонструє різні показники по віковим категоріям:



Даний дашборд допоміг виявити основну цільову групу за віком та зрозуміти поведнку клієнтів різних вікових груп

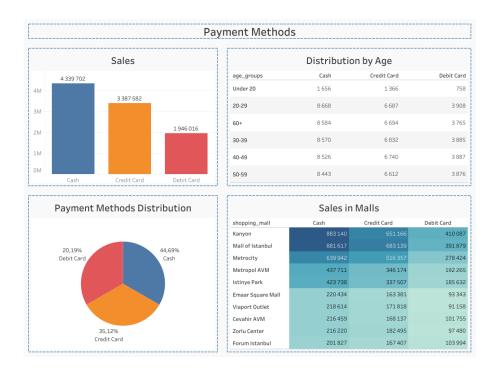
• Categories – дашборд демонструє загальну картину по категоріям, дає розуміння, що продається найчастіше, що є найбільш прибутковим, а що ні, допомагає оцінити вплив сезонності на продажі:



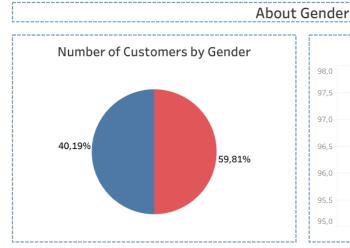
• **Shopping Malls** – дашборд, який демонструє основні показники по кожному торговому центр, такі як кількість клієнтів, продажі за категоріями, гендерний розподіл, прибуток та інші:

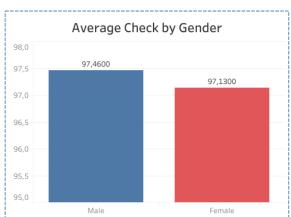


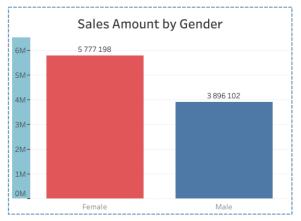
Payment Methods — показує розподіл показників за платіжним методом. Виявилось, що найбільшу долю серед платіжних методів займає банківська кар тка, але й доля готівки виявилась досить високою:



• **About Gender** – демонструє розподіл клієнтів за статтю та графік з основними показниками. Дашборд допоміг виявити, що жінок набагато більше, вони роблять більше покупок та приносять більше доходу, але маються менший середній чек аніж у чоловіків:







Category Sales by Gender				
category	Female	Male		
Clothing	2 625 065	1 759 427		
Shoes	1 516 353	1 043 396		
Technology	1 333 460	892 015		
Cosmetics	156 414	104 850		
Toys	92 925	60 168		
Food & Beverage	19 435	13 239		
Books	18 820	13 278		
Souvenir	14 726	9 729		

Висновки та рекомендації

Головні висновки

1. Демографічний профіль клієнтів

Вік: Основна частина клієнтів торгових центрів Стамбула знаходиться у віці від 20 до 40 років. Це свідчить про те, що молода та середнього віку аудиторія є найбільш активною в покупках. Важливо зазначити, що цей віковий сегмент є ключовим для більшості торгових центрів, оскільки вони є найбільш активними споживачами.

Стать: Жінки складають більшу частину покупців. Це може свідчити про те, що жінки частіше здійснюють покупки в торгових центрах, що ϵ важливою інформацією для маркетингових стратегій, орієнтованих на жіночу аудиторію.

2. Популярні категорії товарів:

Найбільш популярними категоріями товарів ϵ одяг, взуття та косметика. Це вказує на те, що ці категорії ϵ основними драйверами продажів у торгових центрах. Торгові центри можуть збільшити свої доходи, зосереджуючись на розширенні асортименту саме в цих категоріях.

3. Методи Оплати:

Банківські картки ϵ найбільш поширеним методом оплати серед клієнтів. Це свідчить про те, що клієнти віддають перевагу безготівковим розрахункам, що може бути пов'язано з зручністю та безпекою цього методу оплати.

4. Кореляції між показниками:

Аналіз показав середню кореляцію між ціною товару та середнім чеком. Це означає, що клієнти, які купують дорожчі товари, зазвичай витрачають більше за одну покупку. Ця інформація може бути корисною для розробки цінових стратегій та промо-акцій

5. Популярність торгових центрів:

Торгові центри Kanyon, Mall of Istanbul та Меtrосіtу мають найбільшу кількість покупок та високі середні чеки. Це свідчить про їх популярність серед клієнтів, що може бути пов'язано з різноманітністю асортименту товарів, зручністю розташування та якістю обслуговування.

Рекомендації для бізнесу

1. Фокус на цільову аудиторію:

Маркетинг для молоді: Враховуючи, що основна частина клієнтів - це люди віком від 20 до 40 років, рекомендується розробляти маркетингові кампанії, орієнтовані на цей віковий сегмент. Використання соціальних мереж, цифрових рекламних платформ та акцій для молоді може значно підвищити залучення цієї аудиторії.

Жінки як ключовий сегмент: Оскільки жінки складають більшу частину покупців, варто розробити спеціальні пропозиції та акції, орієнтовані на жіночу аудиторію. Це можуть бути знижки на популярні категорії товарів, такі як одяг, взуття та косметика.

2. Розширення асортименту:

Популярні категорії: Рекомендується розширити асортимент товарів у категоріях одягу, взуття та косметики. Це може включати введення нових брендів, співпрацю з відомими дизайнерами та запуск нових лінійок продукції.

Аналіз попиту: Постійний аналіз попиту на різні категорії товарів допоможе вчасно реагувати на зміни вподобань клієнтів та оптимізувати асортимент.

3. Покращення методів оплати:

Програми лояльності: Впровадження програм лояльності для клієнтів, які використовують кредитні картки, може збільшити частоту покупок та підвищити лояльність клієнтів. Це можуть бути бонусні програми, кешбек або спеціальні знижки.

Безготівкові платежі: Забезпечення максимальної зручності та безпеки безготівкових платежів, включаючи мобільні платежі та онлайн-оплату, сприятиме залученню клієнтів.

4. Оптимізація цінових стратегій:

Аналіз цінової чутливості: Проведення регулярних досліджень цінової чутливості клієнтів допоможе визначити оптимальні цінові рівні для різних категорій товарів. Це дозволить збільшити середній чек та загальні доходи.

Промо-акції: Розробка цільових промо-акцій для категорій з високою маржинальністю та проведення сезонних розпродажів сприятиме збільшенню обсягів продажів.

5. Поліпшення обслуговування у пікові дні та місяці:

- Збільшення персоналу: В пікові дні та місяці варто збільшити кількість персоналу для забезпечення високої якості обслуговування. Це допоможе уникнути черг та покращити загальний досвід покупок.
- Додаткові точки обслуговування: Введення додаткових кас та пунктів обслуговування сприятиме швидкому та зручному обслуговуванню клієнтів.

6. Персоналізація маркетингових кампаній:

Використання даних: Використання аналітичних даних для створення персоналізованих маркетингових кампаній дозволить більш точно визначати потреби та вподобання клієнтів. Це може включати розсилку індивідуальних пропозицій та акцій на основі поведінкових даних.

Сегментація клієнтів: Розподіл клієнтів на сегменти за віком, статтю, частотою покупок та іншими характеристиками дозволить створювати більш цільові та ефективні маркетингові стратегії.

7. Впровадження сучасних технологій:

- **Мобільні додатки:** Розробка мобільних додатків для зручного перегляду асортименту, здійснення покупок та отримання інформації про акції та знижки сприятиме підвищенню лояльності клієнтів.
- Аналітика та прогнозування: Використання аналітичних інструментів для моніторингу та прогнозування попиту дозволить оптимізувати запаси та планувати маркетингові заходи.

8. Розширення маркетингової діяльності у популярних торгових центрах:

Спеціальні заходи: Проведення спеціальних заходів та промо-акцій у популярних торгових центрах, таких як Kanyon та Mall of Istanbul, допоможе ще більше збільшити їх популярність та доходи.

Партнерства: Співпраця з відомими брендами та організація спільних заходів сприятиме залученню нових клієнтів та підвищенню лояльності існуючих.

9. Розширення асортименту та впровадження нових продуктів:

Аналіз продажів: Постійний аналіз продажів по категоріях товарів дозволить виявити нові можливості для розширення асортименту та впровадження нових продуктів.

Тестування нових продуктів: Проведення тестових запусків нових продуктів та категорій товарів допоможе оцінити їх потенціал та прийнятність серед клієнтів.

10. Залучення клієнтів через соціальні мережі та цифрові канали:

Соціальні мережі: Активне використання соціальних мереж для залучення клієнтів та просування акцій і знижок. Розробка контенту, який буде цікавий і корисний для цільової аудиторії.

Цифровий маркетинг: Використання інструментів цифрового маркетингу, таких як таргетинг, ремаркетинг та email-маркетинг, для збільшення залучення клієнтів та підвищення конверсії.