## **Тема**

**Дослідження ефективності рекламних кампаній у сфері діджитал маркетингу шляхом аналізу поведінки клієнтів та маркетингових метрик**

## Опис проєкту

Моя робота буде зосереджена на дослідженні даних про взаємдію користувачів з рекламними кампаніями, їх демографічних особливостей і поведінки. Аналіз демографічних особливостей користувачів і їх доходу допомагає краще зрозуміти їх потреби й відповідно поведінку, а дослідження поведінки й взаємодії з рекламними кампаніями дає змогу обрати найкращі маркетингові стратегії для найвищої ефективності в роботі з користувачами.

Робота включатиме розподіл користувачів за сегментами за статтю, віковими й доходовими групами, аналіз присутності чи відсутності впливу бюджету, витраченого на рекламну кампанію, на її показники ефективності, створення рейтингу лояльності клієнтів і дослідження поведінки найлояльніших з них, а також сегментацію за лояльністю для використання в подальших рекламних кампаніях.

**Метою** роботи є не тільки дослідити наведені вище аспекти, а й надати рекомендації бізнесу щодо найкращих каналів і типів рекламних кампаній для досягнення якнайвищої маркетингової ефективності.

## Стек технологій

**Power Query:** очищення даних і визначення правильних типів даних для кожної змінної;

**Excel:** описова статистика, box plots, кореляційний аналіз;

**Google BigQuery (SQL):** підготовка даних до побудови візуалізацій, створення вікових та доходових категорій, створення рейтингового списку та агрегація метрик по каналам та типах кампаній;

**Power BI:** візуалізація даних;

**Jupyter Notebook (Python):** сегментація клієнтів за лояльністю для її подальшого використання в маркетинговій стратегії.

## Вибір датасету

Для вибору датасету була обрана платформа Kaggle за її багатогранність та різноманітність матеріалу. Обраний для проєкту датасет:

Rabie El Kharoua. (2024). 📈 Predict Conversion in Digital Marketing Dataset [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/8742946>

Датасет сподобався за об’ємність даних та можливість їх багатогранного аналізу.

Було досліджено також такі бази даних, як: [BigQuery public datasets](https://cloud.google.com/bigquery/public-data) та [*UCI*](https://archive.ics.uci.edu/)

[*Machine Learning Repository*](https://archive.ics.uci.edu/), але вони не були обрані через доволі специфічну тематику.

### Опис датасету

**Демографічна інформація**

* **CustomerID**: унікальний ключ кожного клієнта.
* **Age**: вік клієнта.
* **Gender**: стать клієнта (Male/Female).
* **Income**: річний дохід клієнта в USD.

**Marketing-specific Variables**

* **CampaignChannel**: канал поширення маркетингової кампанії (Email, Social Media, SEO, PPC, Referral).
* **CampaignType**: тип маркетингової кампанії (Awareness, Consideration, Conversion, Retention).
* **AdSpend**: бюджет витрачений на рекламу в USD.
* **ClickThroughRate**: метрика, що характеризує частоту кліків на маркетинговий вміст.
* **ConversionRate**: метрика, що характеризує частоту, з якою кліки перетворюються на бажані дії (наприклад, покупки).

**Customer Engagement Variables**

* **WebsiteVisits**: кількість відвідувань вебсайту.
* **PagesPerVisit**: середня кількість відвіданих сторінок за візит.
* **TimeOnSite**: середній час, проведений на сайті (у хвилинах).
* **EmailOpens**: кількість відкриттів листів.
* **EmailClicks**: кількість кліків по лінках в листах.

**Historical Data**

* **PreviousPurchases**: кількість попердніх покупок, зроблених клієнтом.
* **LoyaltyPoints**: кількість балів лояльності, зібраних клієнтом.

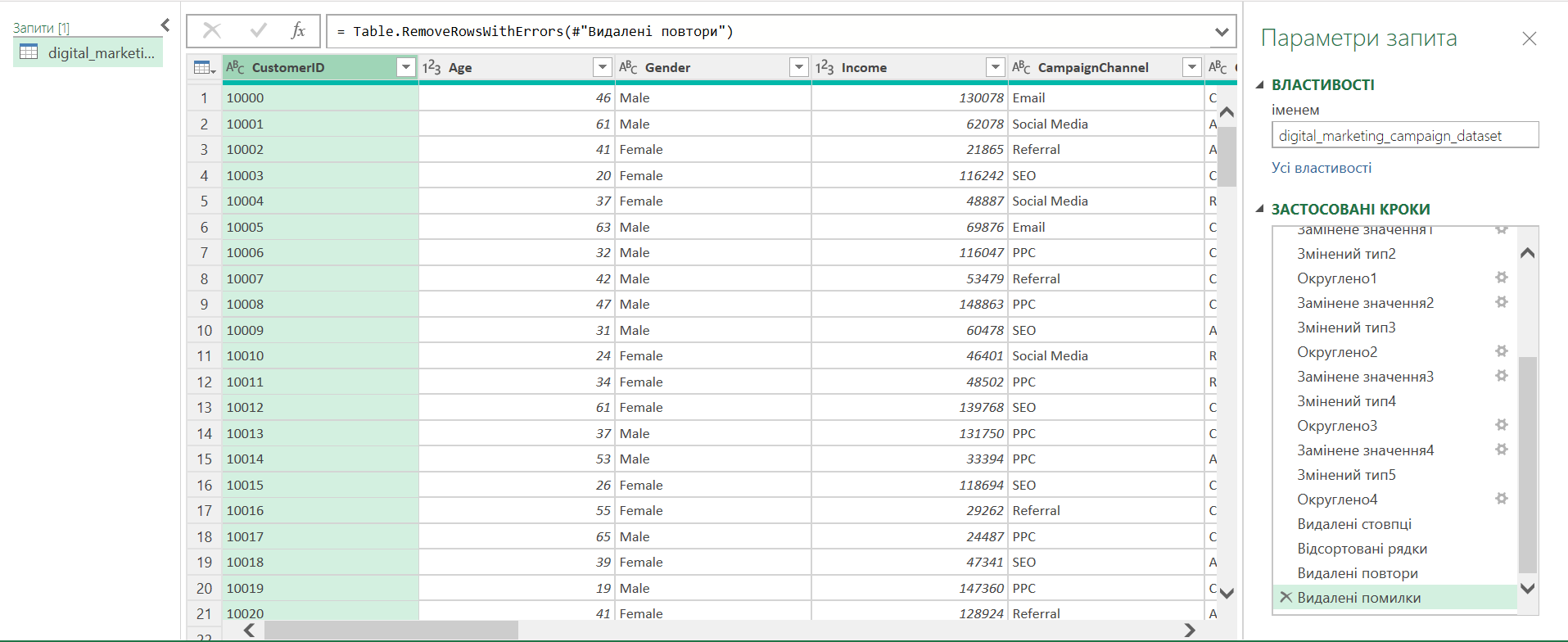
**Target Variable**

* **Conversion**: двійкова змінна, що вказує, чи здійснив клієнт конверсію (1) чи ні (0).

## План

1. Очищення й перетворення даних, підготовка до подальшого аналізу в Power Query.
2. Використання описової статистики для кращого розуміння даних, пакету аналізу даних для проведення кореляційного аналізу в Excel.
3. Загрузка даних в Google BigQuery.
4. Підготовка даних до візуалізації за допомогою додавання нових стовпців та агрегації.
5. Завантаження даних та створення візуалізацій у Power BI.
6. Проведення сегментації в Python.

### Очищення даних



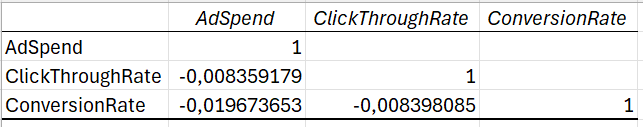
У процесі підготовки й очищення даних було визначено правильні типи даних у стовпцях, усунено пусті записи й помилки. Результати показали чистоту датасету, адже кількість записів після цього не змінилась.

### Description Statistics

<https://docs.google.com/spreadsheets/d/1y8BaWVfx7fVvt-8Hk62EePpUo2WdJTHL/edit?usp=sharing&ouid=104763776962176246792&rtpof=true&sd=true> (оскільки робота була створена в Excel, Google Sheets не відображають деякі елементи, зокрема box plots).

Було створено описову статистику по кожному числовому полю та побудовано box plots для виявлення викидів у даних, якщо вони наявні. Статистика допомогла згодом визначити рамки для груп користувачів, а box plots показали відсутність викидів у даних.

### Correlation Analysis



Кореляційний аналіз показав дуже слабку від’ємну кореляцію між бюджетом, витраченим на рекламні кампанії, та їхньою ефективністю (основними метриками CR та CTR). Отже, у цьому випадку кількість витрачених грошей не впливає на ефективність.

### SQL

SELECT

CustomerID,

Gender,

Age,

--Creating age categories

CASE

WHEN Age >= 18 AND Age < 31 THEN "young"

WHEN Age >=31 AND Age < 51 THEN "middle"

ELSE "senior"

END age\_category,

Income,

--Creating income categories

CASE

WHEN Income >= 20000 AND Income < 50000 THEN "low"

WHEN Income >= 50000 AND Income < 100000 THEN "moderate"

ELSE "high"

END income\_category,

WebsiteVisits,

PagesPerVisit,

TimeOnSite,

EmailOpens,

EmailClicks,

--Counting personal activity rate (shows the ratio of email clicks to email opens)

ROUND(CASE

WHEN EmailOpens = 0 THEN 0 --considering the presence of 0 values in the divider

ELSE (EmailClicks / EmailOpens) \* 100

END, 2) AS personal\_activity\_rate,

PreviousPurchases,

LoyaltyPoints,

--ranking the customers by loyalty points without gaps in rating values caused by the same points values

DENSE\_RANK() OVER(ORDER BY LoyaltyPoints DESC) as loyalty\_rating,

Conversion

FROM `pet-project-428819.marketing\_data.data`

ORDER BY loyalty\_rating

**Results**: <https://drive.google.com/file/d/1Zv8aTfffYoIdsp02Z4oGy898FBjY2-4o/view?usp=sharing>

SELECT

CampaignChannel,

--counting average spendings on the campaign channel

ROUND(AVG(AdSpend), 2) as avg\_ad\_spend,

--counting average CTR for each channel

ROUND(AVG(ClickThroughRate), 2) as avg\_ctr,

--counting average CR for each channel

ROUND(AVG(ConversionRate), 2) as avg\_cr

FROM `pet-project-428819.marketing\_data.data`

GROUP BY CampaignChannel

**Results:** <https://drive.google.com/file/d/1xvJ5x9_mhpdWYVi4eXx1NA-nfljD_zx5/view?usp=sharing>

SELECT

CampaignType,

--counting average spendings on the campaign type

ROUND(AVG(AdSpend), 2) as avg\_ad\_spend,

--counting average CTR for each type

ROUND(AVG(ClickThroughRate), 2) as avg\_ctr,

--counting average CR for each type

ROUND(AVG(ConversionRate), 2) as avg\_cr

FROM `pet-project-428819.marketing\_data.data`

GROUP BY CampaignType

**Results**: <https://drive.google.com/file/d/1B_ZZFPcBBwPCqFNZEQdIDbBnEq43LJpB/view?usp=sharing>

SELECT

--coverted users

SUM(Conversion) as total\_conversions,

--not converted users

COUNT(CustomerID) as total\_customers,

--full ratio of converted users

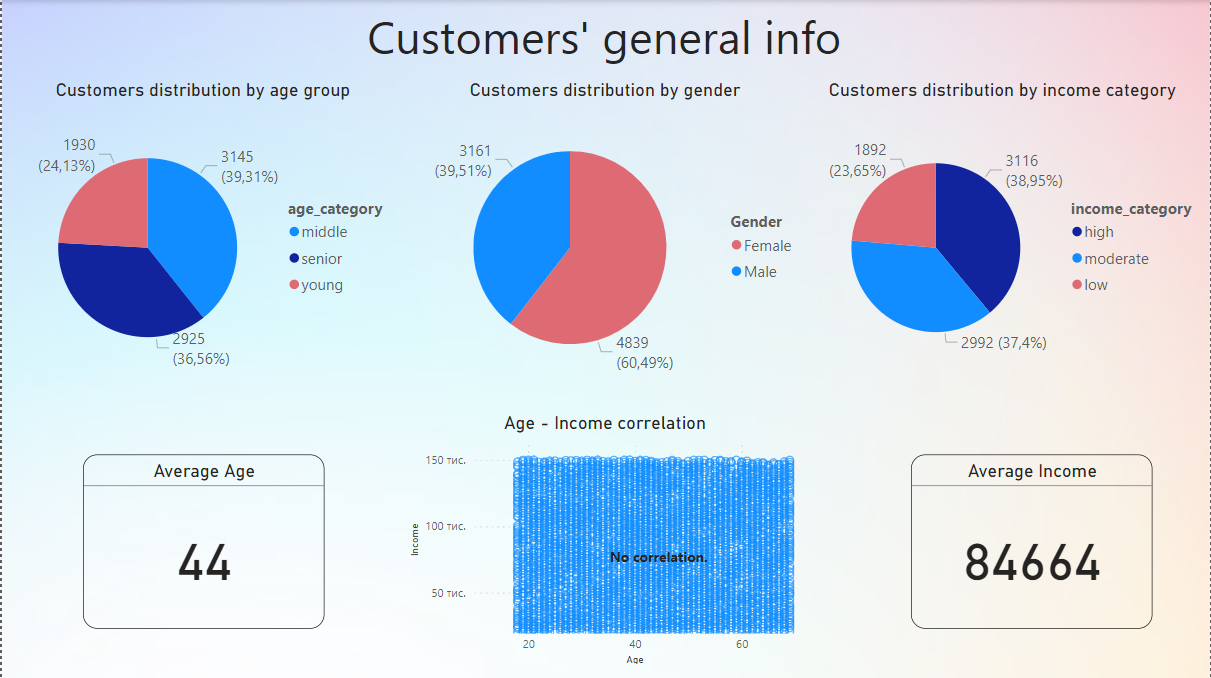
ROUND(SUM(Conversion)/(COUNT(CustomerID)), 2) as users\_conversion\_rate

FROM `pet-project-428819.marketing\_data.data`

**Results**: <https://drive.google.com/file/d/1McUrVEyjYPy4PV9KFE7elf-h_xn6I-ue/view?usp=sharing>

### Power BI

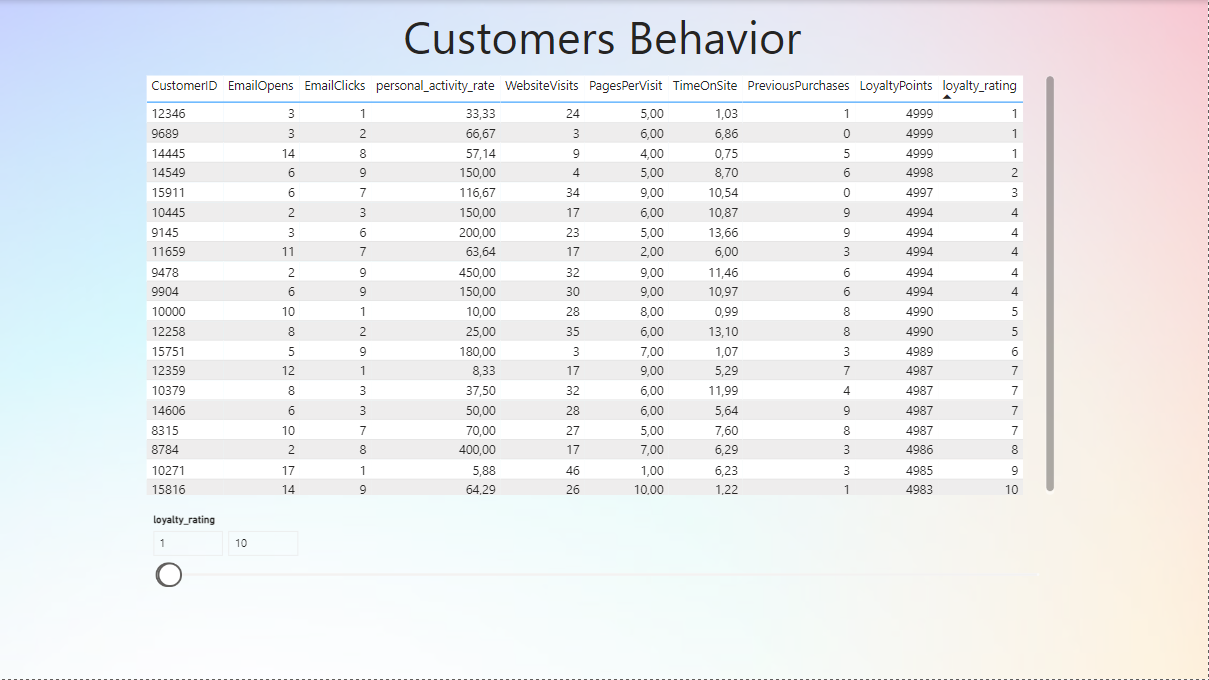
<https://drive.google.com/file/d/17Sh7PXEeHEK9kzAcRgOx6lAY1CFrlcLA/view?usp=sharing>



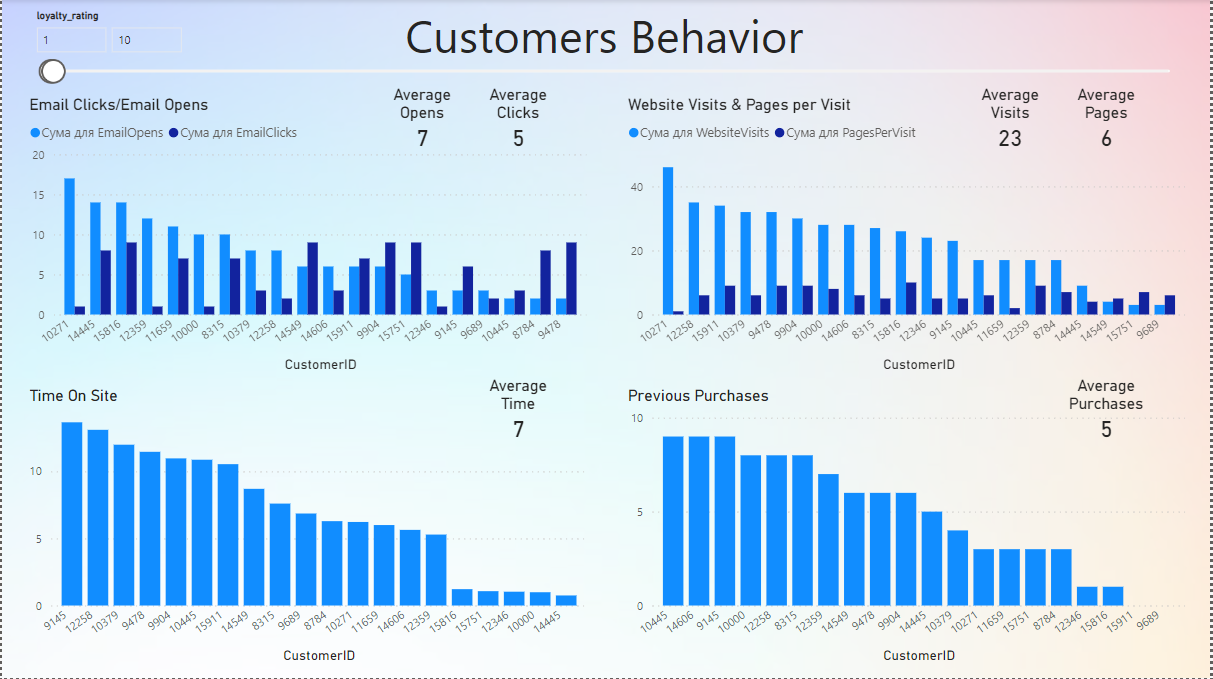
Перший дашборд демонструє загальні відомості про користувачів.

* За віком найбільша частка користувачів - це middle category (31-50 років), на другому місці - senior customers (51-69 років), а найменше корстувачів групи young (18-30 років).
* За статтю серед наших користувачів переважають жінки (60%).
* За рівнем річного доходу переважають користувачі з високим рівнем доходу (вище $100000), майже така ж кількість користувачів із середнім рівнем ($50000-100000) та всього 24% користувачів із низьким рівнем доходу ($20000-50000).
* Середній вік усіх користувачів - 44 роки.
* Середній дохід - $84664.
* Кореляція між віком і доходом відсутня.

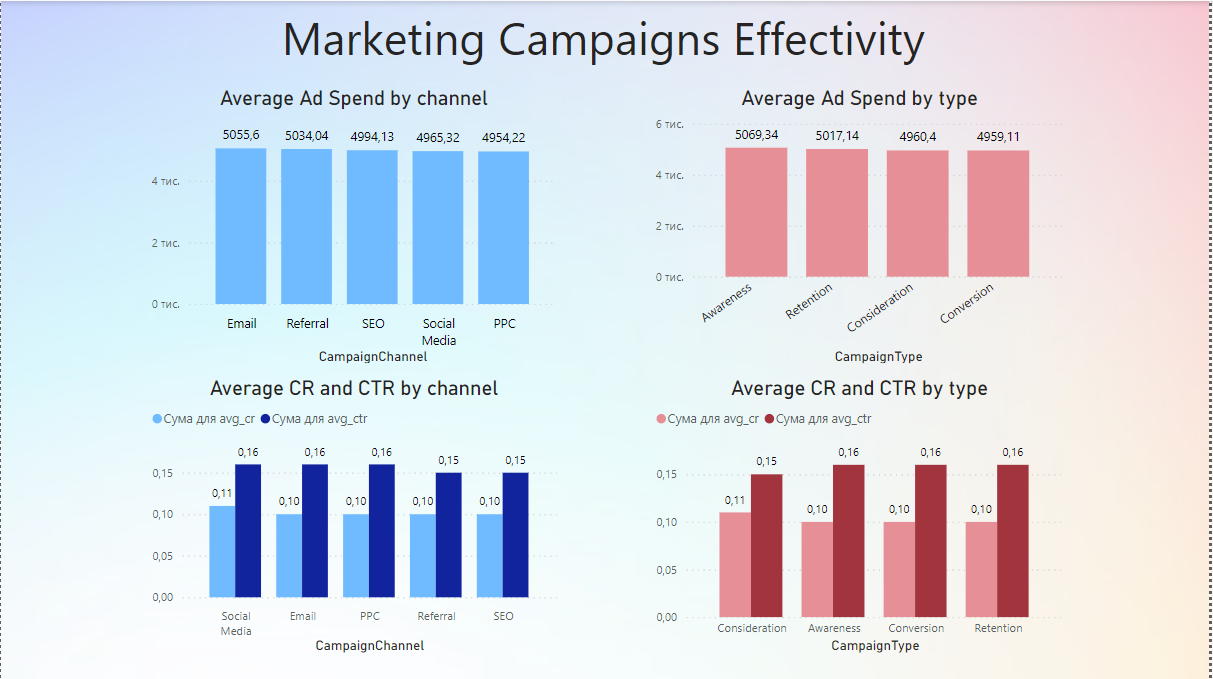
Із наведених даних можна зробити висновок, що при підготовці рекламних кампаній необхідно зосереджувати увагу на інтересах людей старшого віку, у деяких можна робити акцент на жіночу аудиторію, а також можна рекламувати дорогі або середньої ціни продукти, адже аудиторія доволі забезпечена.



На другому дашборді можна переглянути всю інформацію по поведінці клієнтів у зручному форматі фільтруючи по рейтингу користувачів, а також виокремити користувачів із необхідним рейтингом.



На третьому дашборді можна переглянути візуальне відображення поведінки користувачів фільтруючи за рейтингом. Із першого графіку видно, що лояльні користувачі часто повертаються до листів і повторно переходять за посиланнями. Середня кількість відкритих листів серед ТОП-10 лояльних користувачів становить 7 штук, а середня кількість кліків - 5. Середня кількість відвідувань сайту на користувача становить 23 рази, а середня кількість переглянутих сторінок за візит - 6 штук. Середня кількість часу проведеного на сайті - 7 хвилин, а середня кількість попередніх покупок - 5 штук.

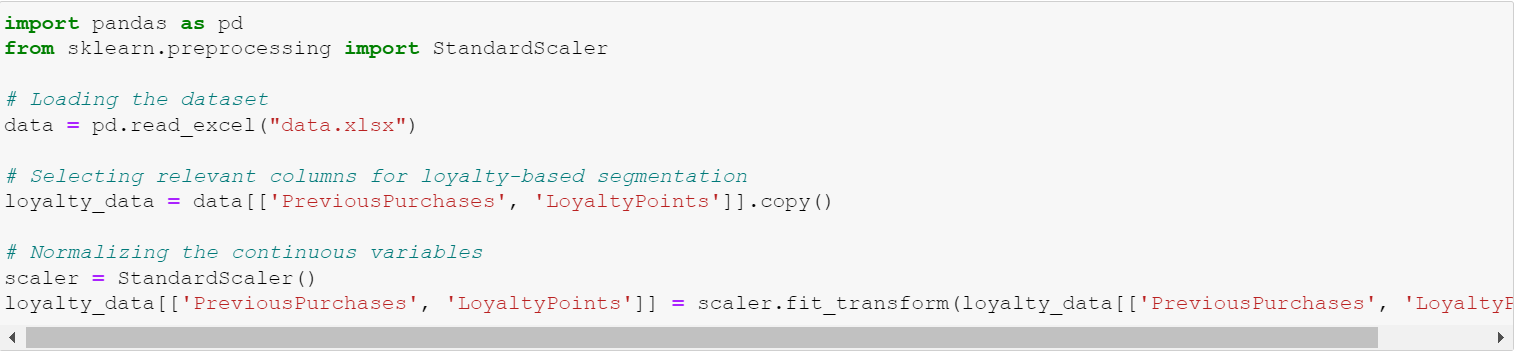


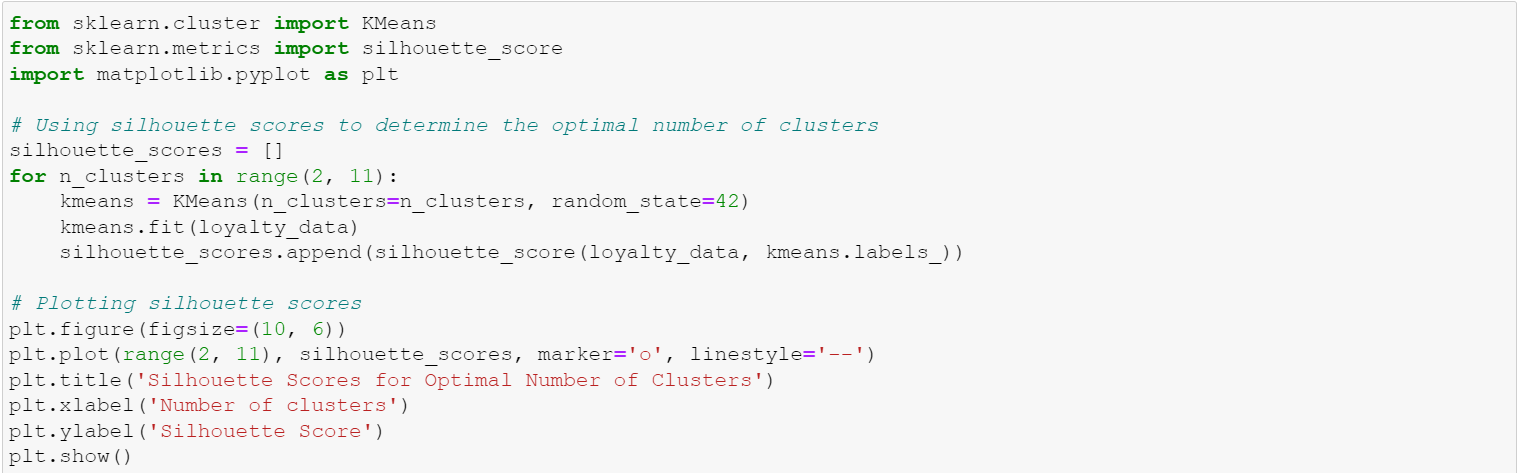
На останньому дашборді можна побачити ефективність рекламних каналів і типів кампаній за основними маркетинговими метриками Click-Through Rate (CTR) та Conversion Rate (CR).

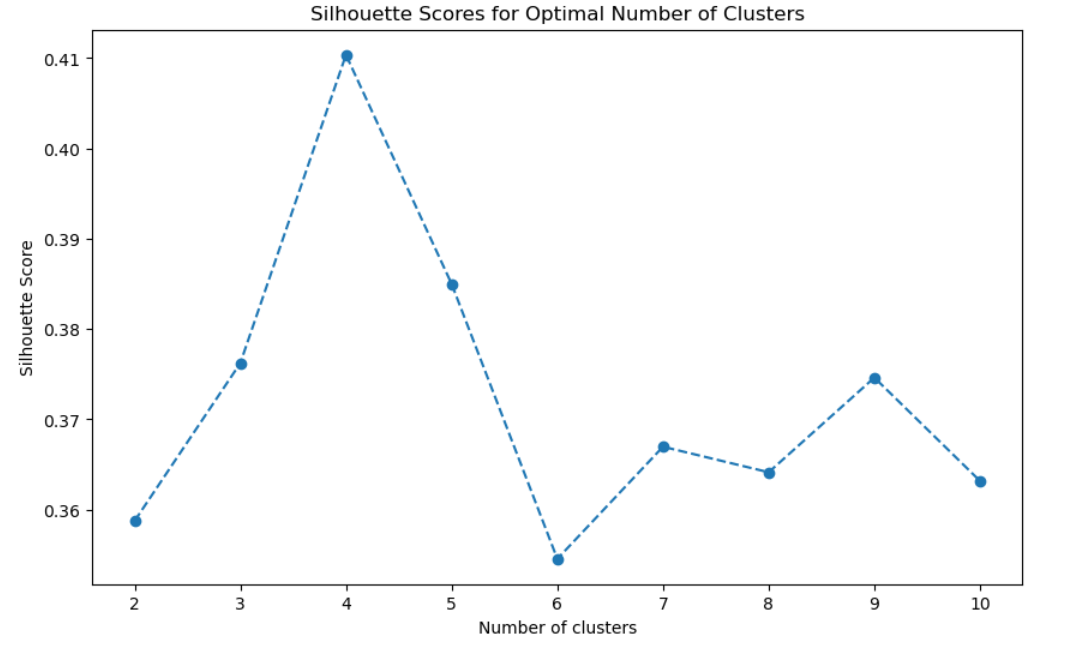
Можна зробити висновок, що найкращими каналами для рекламних кампаній за вартістю і ефективністю є Social Media (найвища ефективність і невисока вартість) та PPC (найнижча вартість і високий CR), а типами - Conversion (найнижча вартість і високий CR) та Consideration (найвищий CTR і невисока вартість).

### Сегментація за лояльністю

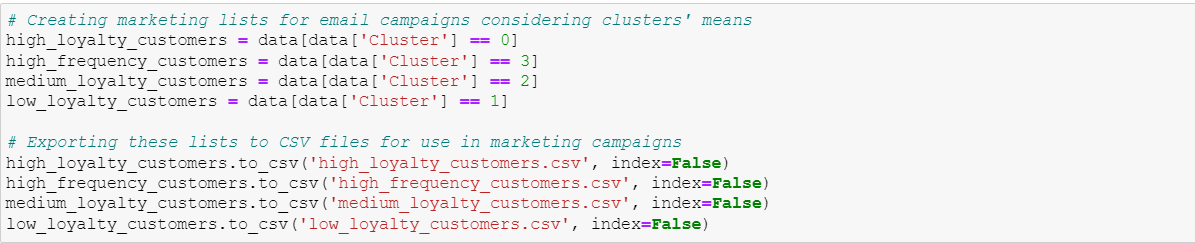
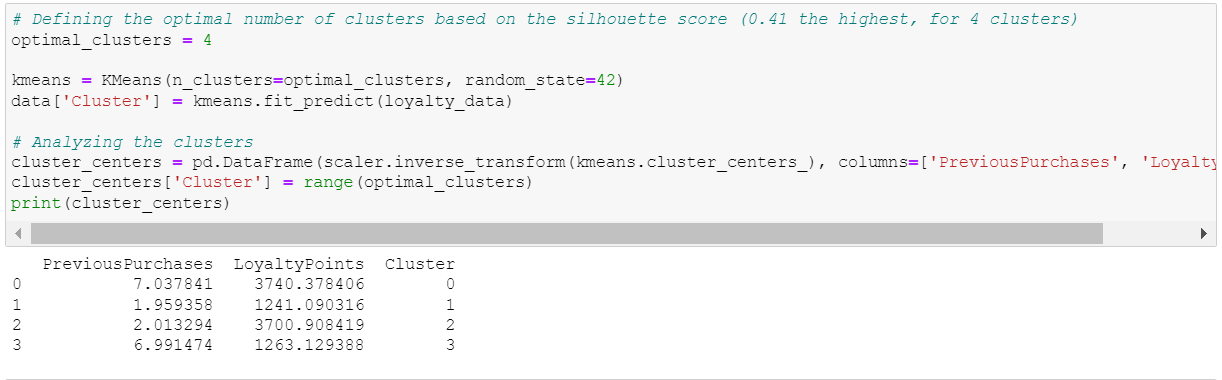
Для проведення сегментації було обрано метод KMeans, оскільки він є доволі універсальним і одним із найпопулярніших та найефективніших. Метод прекрасно справляється з великими об’ємамами даних, розподіляючи їх на кластери зі значень які найбільше підходять до середнього цього кластеру, аніж іншого.







Для цього набору даних найефективнішою виявилась сегментація на 4 кластери, адже цей варіант має найвищий silhouette score. Тому далі розподілення даних на 4 кластери (сегменти).



У результаті сегментації було виділено 4 сегменти:

1. High loyalty customers: високий рівень балів лояльності та багато попередніх покупок;   
   Для цього сегменту можна впроваджувати спеціальну систему лояльності зі знижками та спеціальними пропозиціями.
2. High frequency customers: багато попередніх покупок, але не високі значення балів лояльності;   
   Для цього сегменту доцільно часто надсилати пропозиції зі знижками.
3. Medium loyalty customers: середні значення балів лояльності і попередніх покупок;   
   Періодично надсилати спеціальні пропозиції та знижки.
4. Low loyalty customers: низькі значення балів лояльності і попередніх покупок;   
   Знижки на першу покупку після довгого затишшя, акції по типу 1+1 = 2.

## Висновки

* Основна аудиторія бізнесу - люди середнього та старшого віку (від 30 років) із середнім та високим рівнем річного доходу (від $50000), більше половини з них - жінки.
* Бюджет витрачений на рекламу не впливає на її ефективність.
* Залежність рівня доходу користувачів від їхнього віку відсутня.
* Не всі користувачі з ТОП-10 за рейтингом лояльності демонструють високу активність в плані кліків, відвіданих сторінок, часу проведеного на сайті тощо. Можливо, вони просто робили дорогі покупки, за що отримали багато поінтів.
* На всі канали й типи рекламних кампаній витрачено приблизно однакову кількість грошей.
* Усі кампанії показали майже однакову ефективність за CR та CTR. Проте, найкращими каналами для рекламних кампаній за вартістю і ефективністю є Social Media (найвища ефективність і невисока вартість) та PPC (найнижча вартість і високий CR), а типами - Conversion (найнижча вартість і високий CR) та Consideration (найвищий CTR і невисока вартість).
* Користувачів можна поділити на 4 сегменти за лояльністю задля диверсифікації рекламних кампаній.
* Бізнес має високий показник конвертації користувачів (88%).

## Рекомендації бізнесу

* Під час створень рекламних кампаній орієнтуватися на інтереси й потреби основної аудиторії (люди середнього та старшого віку (від 30 років) із середнім та високим рівнем річного доходу (від $50000), більше половини з них - жінки).
* Спробувати покращити якість рекламних кампаній через Social Media за рахунок підвищння бюджету й проаналізувати, чи підвищить це ефективність, адже канал є перспективним за показниками.
* Перевірити якість усіх рекламних кампаній та покращувати її задля підвищення показників ефективності, адже вони насправді доволі низькі.
* Використовувати сегментацію за лояльністю задля сворення більш диверсифікованих і націлених на користувача рекламних кампаній, аби підтримати лояльність клієнтів або підвищити її.