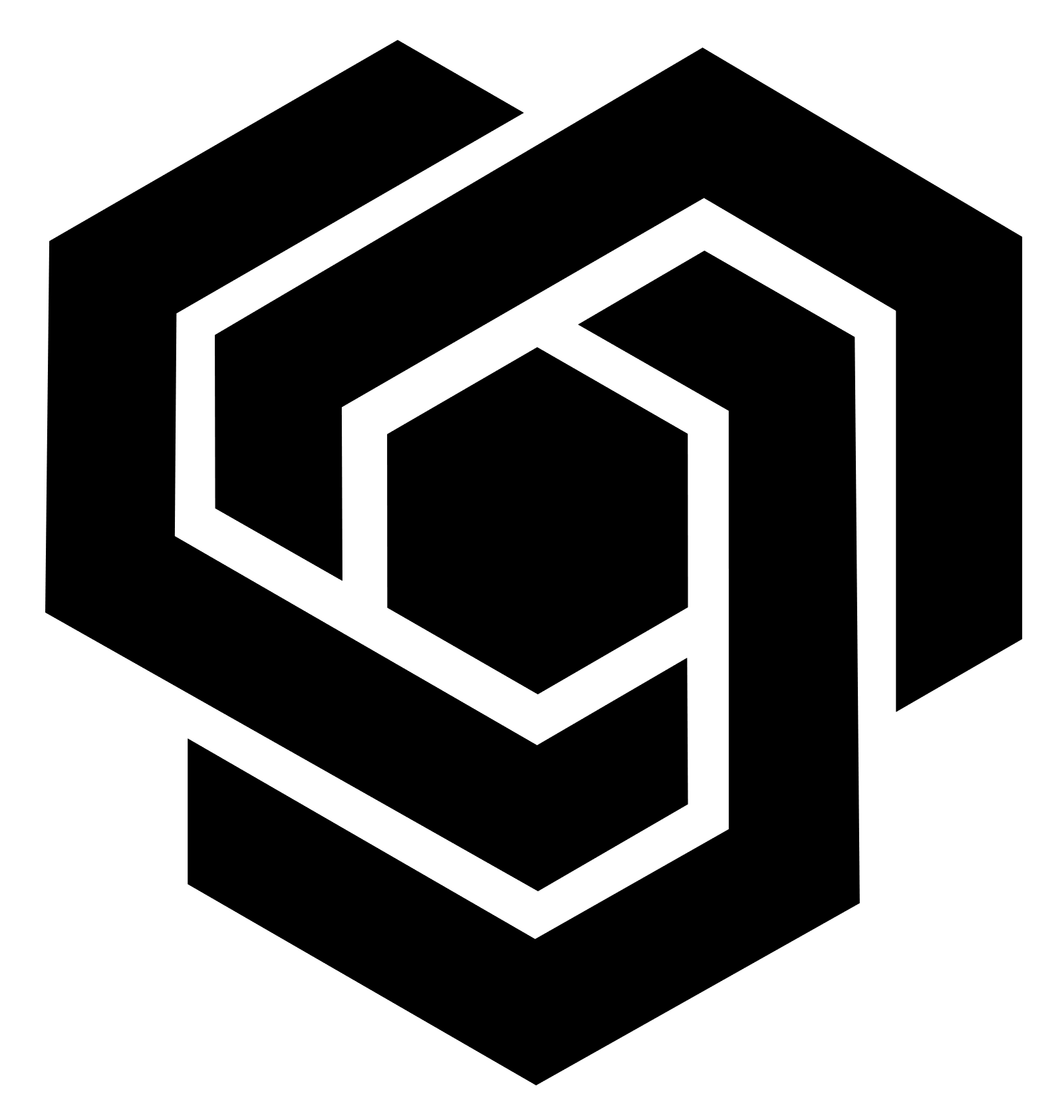
*Технически университет - София*



**Анализ на бизнес данни в социалната мрежа Facebook**

Изготвили: Савина Вълчанова, Денис Цолов, Цветомир Цветков

Специалност: АГМПД, ФПМИ

Съдържание:

1. Въведение……………………………………………….…………………3
2. Цел………………………………………………………………….………3
3. Експериментална рамка – методи и техника на анализ………………..4
4. Обработка и анализ на данните………………………………….……….6
5. Представяне и визуализация на резултатите……………………….…..10
6. Допълнителен анализ на резултатите и извеждане на кл. изводи…….14
7. Заключение и изводи……………………………………...………………22
8. Сорс код……………………………………………………………………23
9. Информационни източници………………………………………………23
10. Въведение

Анализът на маркетинговите кампании и взаимодействията на потребителите с тях е от съществено значение за разбиране на ефективността и въздействието на тези усилия. Представените данни обхващат информация за различни рекламни кампании, като се фокусират върху показатели като импресии, кликвания, разходи и конверсии. Чрез детайлен анализ на тези данни можем да извлечем ценна информация за поведението на потребителите и ефективността на рекламните послания.

Прегледът на данните разкрива различни аспекти на кампаниите, като:

* **Рекламни идентификатори и периоди на отчитане**: Информацията за уникалните идентификатори на рекламите и периодите на отчитане предоставя контекст за времевата рамка на кампаниите.
* **Демографски данни**: Данните за възраст и пол на аудиторията дават възможност за анализ на целевите групи.
* **Интереси**: Информацията за интереси на потребителите позволява по-детайлен анализ на таргетирането.
* **Импресии и кликвания**: Тези метрики измерват обхвата и ангажираността на кампаниите.
* **Разходи и конверсии**: Данните за разходите и конверсиите показват икономическата ефективност на рекламите.

Чрез анализа на тези данни можем да идентифицираме тенденции, да оценим въздействието на различни фактори и да направим информирани решения за оптимизация на бъдещите кампании.

1. Цел

Целта на тази разработка е да анализира маркетингови кампании в мрежата Facebook. Оценка на ефективността на рекламната кампания и формулиране на препоръки за оптимизация на маркетинговите стратегии.

Подцели с използване на предоставените колони:

1. Демография:
   * Колони: age, gender
   * Цел: Анализиране на възрастовите групи и пола на потребителите, които взаимодействат с кампанията.
2. Анализ на ангажираността:
   * Колони: interest1, interest2, interest3
   * Цел: Изучаване на интересите на потребителите, за да се види кои типове съдържание са най-популярни и ефективни.
3. Времеви анализ:
   * Колони: reporting\_start, reporting\_end
   * Цел: Изследване на времевите периоди на кампанията и идентифициране на пиковите времена на активност.
4. Идентификация на кампаниите:
   * Колони: campaign\_id, fb\_campaign\_id
   * Цел: Разграничаване на различните кампании и проследяване на тяхната ефективност и въздействие.
5. Експериментална рамка – методи и техника на анализ

В тази секция ще опишем методите и техниките, използвани за анализа на данни от рекламната кампания във Facebook.

1. Набори данни (Datasets):

* Източник на данни: <https://www.kaggle.com/datasets/madislemsalu/facebook-ad-campaign/data>
* Обхванати колони: reporting\_start, reporting\_end, campaign\_id, fb\_campaign\_id, age, gender, interest1, interest2, interest3, impressions, clicks, spent, total\_conversion, approved\_conversion
* Период на събиране: Данните обхващат различни периоди от време.

2. Обработка на данни:

* Почистване на данни: Премахване на дублирани записи, справяне с липсващи стойности и коригиране на грешки в данните.
* Преобразуване на данни: Преобразуване на времеви данни и форматиране на числови и категорийни стойности за по-лесен анализ.

3. Избор на метод и техника за анализ:

* Демографски анализ:
  + Метод: Използване на описателни статистики за анализ на възрастовите групи и пола на потребителите.
  + Техника: Честотен анализ и визуализация чрез хистограми и бокс плотове.
* Анализ на интересите:
  + Метод: Изследване на предпочитанията на потребителите по различни интереси.
  + Техника: Създаване на бар графики и разпределение на интересите с помощта на библиотеки като Seaborn и Matplotlib.
* Времеви анализ:
  + Метод: Анализ на времевите серии за идентифициране на пикове в активността на потребителите.
  + Техника: Използване на времеви графики и анализ на продължителността на кампаниите.

4. Използвани библиотеки и софтуерни средства:

* Pandas: За манипулиране и анализ на данните.
* NumPy: За математически изчисления и операции.
* Matplotlib: За създаване на графики и визуализации на резултатите.
* Seaborn: За по-сложни и изящни визуализации.
* Sklearn: За конвертиране на категорийни променливи в числови стойности.
* Jupyter Notebook: За организиране, изпълнение и представяне на анализа.

5. Техника за интерпретация на резултатите:

* Описание на резултатите: Подробно описание на наблюдаваните тенденции и модели в данните.
* Формулиране на препоръки: На базата на анализа ще се изготвят препоръки за оптимизация на бъдещите рекламни кампании.

1. Обработка и анализ на данните
   1. Набор от данни (*Фигура 1*)

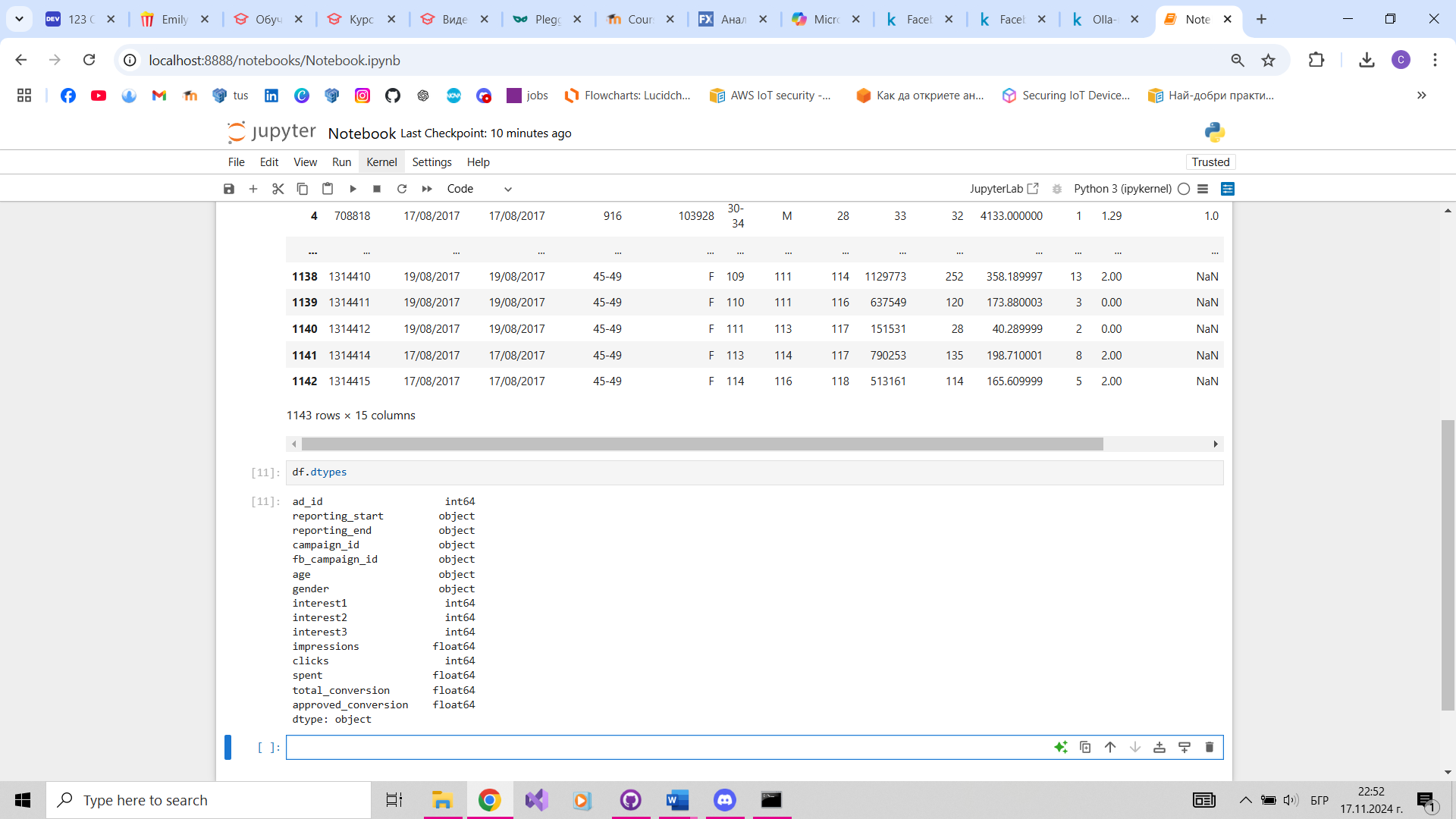
Тук виждаме несъответствие при стойностите на колоните при campaign\_id, fb\_campaign\_id, age, gender, т.е. за целта на проекта ще работим с данните до ред 762 от датасета.

A screenshot of a computer

Description automatically generated

*Фигура 1. Датасета*

* 1. Анализ на набора от данни



*Фигура 2. Типове данни*

Визуализирайки типовете, виждаме, че имаме int, obj и float. (*Фигура 2*)

Проверяваме дали има липсващи стойности (*Фигура 3*) и виждаме, че total\_conversion и approved\_conversion има 382 липсващи стойности. Затова попълваме тези стойности със средната стойност (mean) на съответните колони. (*Фигура 4*)Това осигурява разумно предположение за стойностите, които липсват, без да създава големи отклонения в данните.

A screenshot of a computer

Description automatically generated

*Фигура 3. Стойности*

A screenshot of a computer

Description automatically generated

Фигура 4. Запълване със средна стойност

Създаваме campaign\_duration за проследяване на кампаниите. (*Фигура 5*)

Виждаме, че всяка кампания е с продължителност 0, защото кампаниите са в рамките на един ден.

Също създаваме engagement\_rate за ангажираността (*Фигура 6*) като заместваме всички нулеви стойности в колоната impressions с много малко число , за да се избегне деленето на нула, което би довело до грешка.

A screenshot of a computer

Description automatically generated

*Фигура 5. campaign\_duration*

A screenshot of a computer

Description automatically generated

*Фигура 6. engagement\_rate*

A screenshot of a computer

Description automatically generated

*Фигура 7. Дубликати*

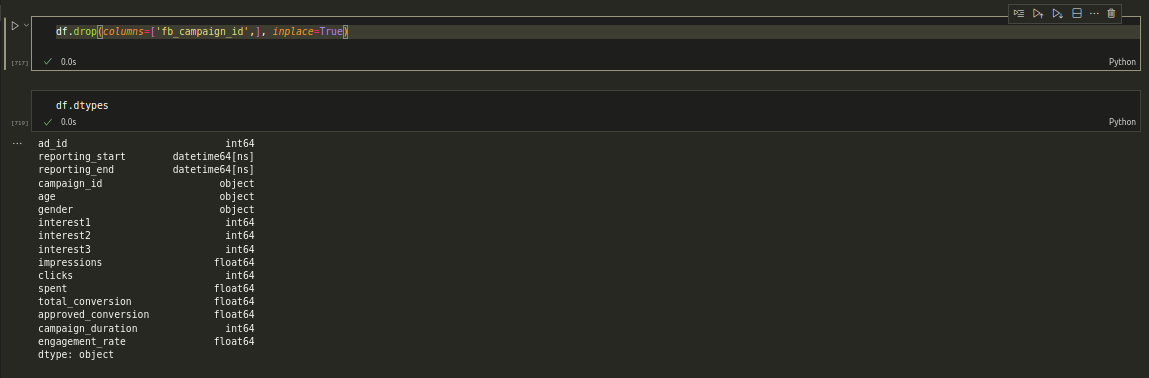
Проверяваме за дубликати (*Фигура 7*) и установяваме, че няма.

A screenshot of a computer

Description automatically generated

*Фигура 8. Безкрайни и NaN стойности*

Последната проверка е за безкрайни и нечислени стойности. (*Фигура 8*)

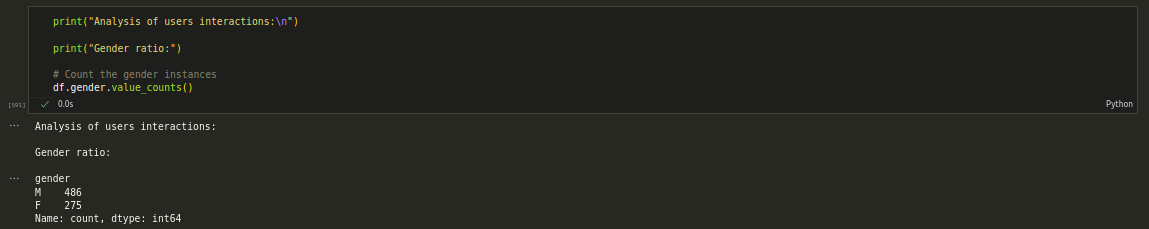


*Фигура 9. Изчистен и готов файл за анализ*

Файлът се подготвя и подрежда за същинския анализ. (*Фигура 9*)

1. Представяне и визуализация на резултатите
   * + 1. Демографски анализ

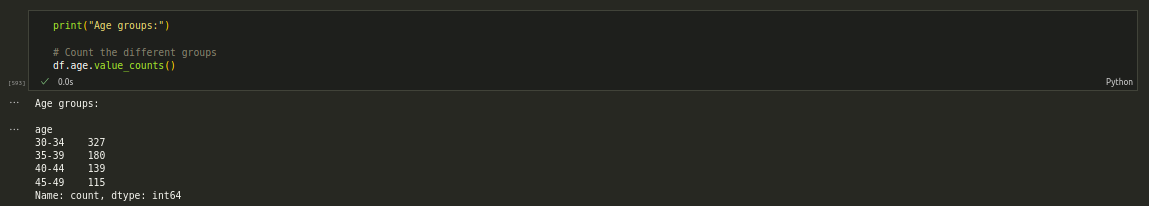
Броят на участващите мъже е близо **1.8** пъти повече от този на жените. (*Фигура 9, 10*)

*Фигура 9. Демографско разделение*

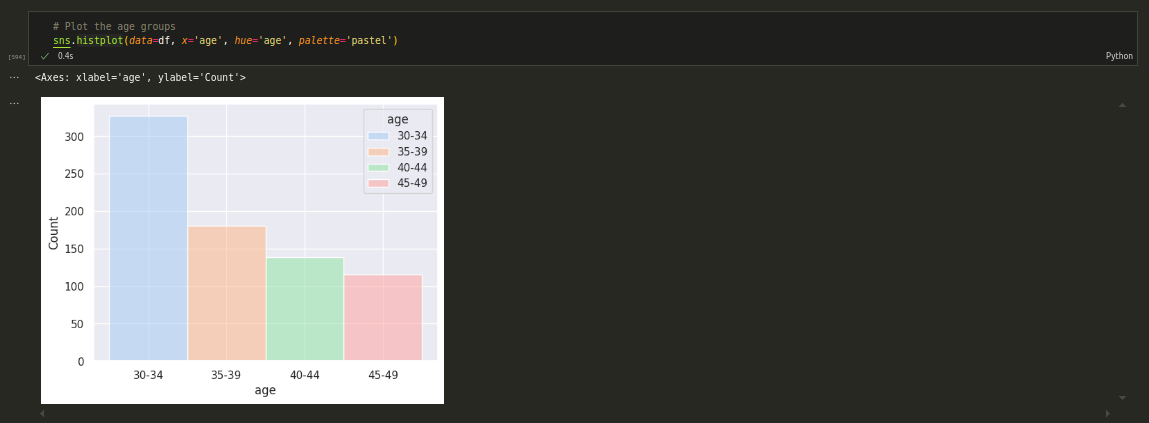


*Фигура 10. Демографско разделение (графично)*

При възрастта на участниците се забелязва, че с най-голям процент е представена възрастовата група на **30-34** годишните, като второто място е с почти 2 пъти по-малко представители. (*Фигура 11, 12*)

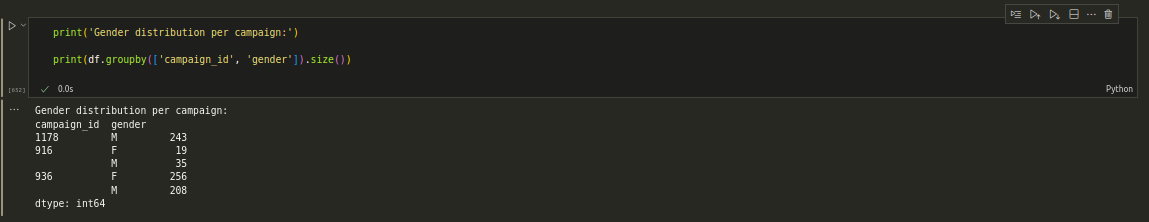


*Фигура 11. Възрастово разделение*

**

*Фигура 12. Възрастово разделение (графично)*

Половото разпределение е очаквано за кампании **916** и **1178** (*Фигура 13*), където мъжете са повече (а по-специално в случая на **1178**, те са единствени представители). Изненадващо е разпределението при кампания **936**, където жените са с близо **⅕** повече от мъжете. (*Фигура 14*)

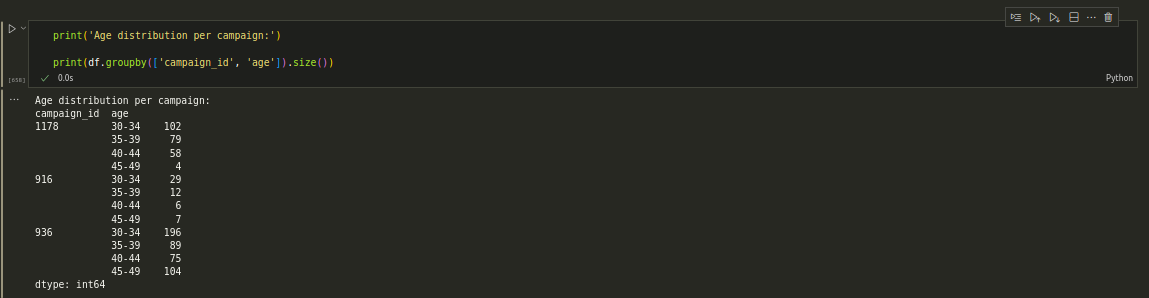


*Фигура 13. Полово разпределение спрямо кампаниите*

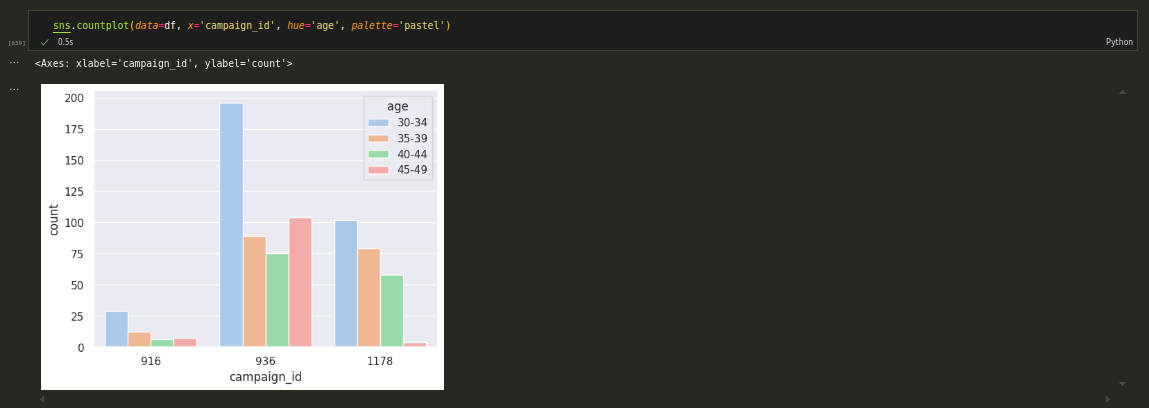


*Фигура 14. Полово разпределение спрямо кампаниите (графично)*

При възрастовото разпределение (*Фигура 15*) няма сериозни изненади - демографската група **30-34** години е водеща за 3-те кампании. Важно е да се отбележи, че **1178** иманай-слаб досег с групата **45-49** години, за които кампания **936** е с най-голям интерес. (*Фигура 16*)



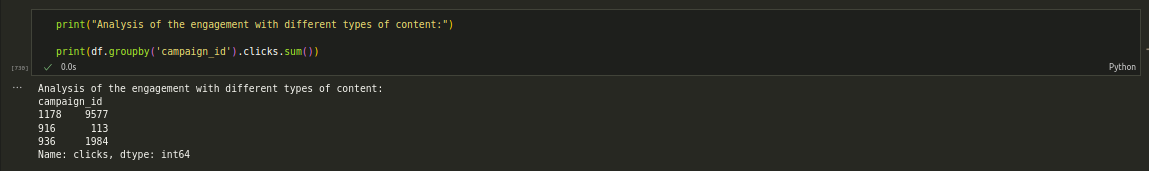
*Фигура 15. Възрастово разпределение спрямо кампаниите*

**

*Фигура 16. Възрастово разпределение спрямо кампаниите (графично)*

* + - 1. Анализ на ангажираността

Въпреки по-малкия брой участници, най-успешна от трите кампании от гледна точка на кликове е **1178**. Кампания **916** е очаквано с най-малък успех.(*Фигура 17, 18*)



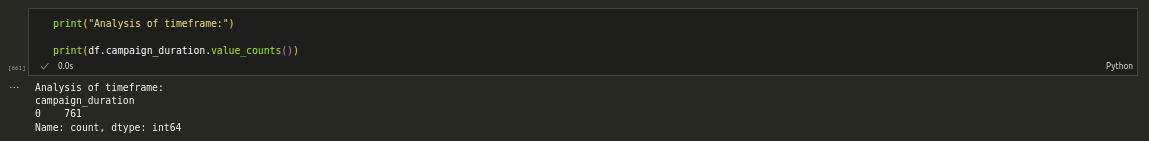
*Фигура 17. Брой кликове за всяка кампания*

**

*Фигура 18. Брой кликове за всяка кампания (графично)*

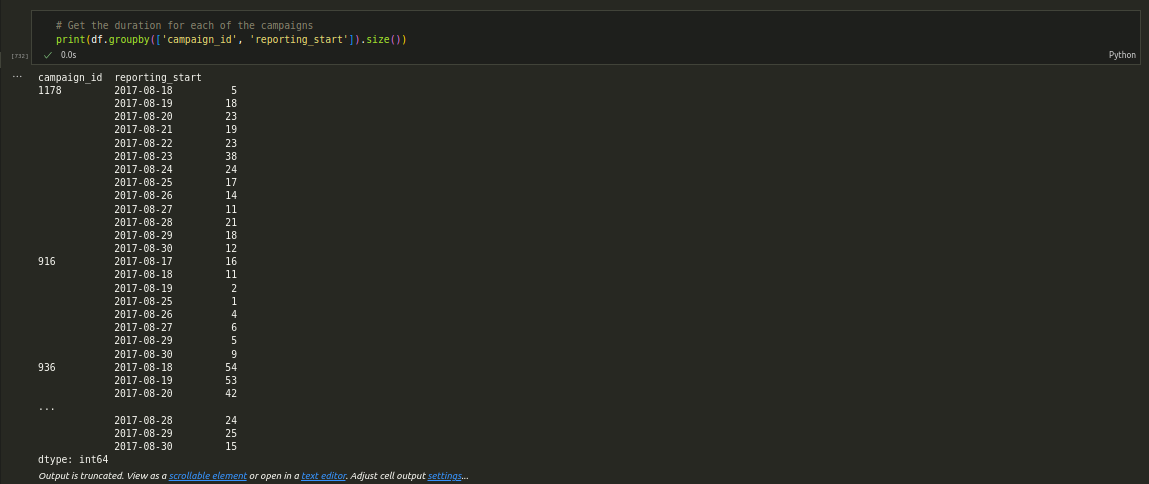
* + - 1. Времеви анализ

Времевият анализ на dataset-а ни показва, че всички записи се случват в рамките на ден. (*Фигура 19*)



*Фигура 19. Времево разпределение*

Можем да проследим в какъв времеви период е продължила всяка една от кампаниите. (*Фигура 20, 21*)

**

*Фигура 20. Времево разпределение за кампаниите (по дни)*

**

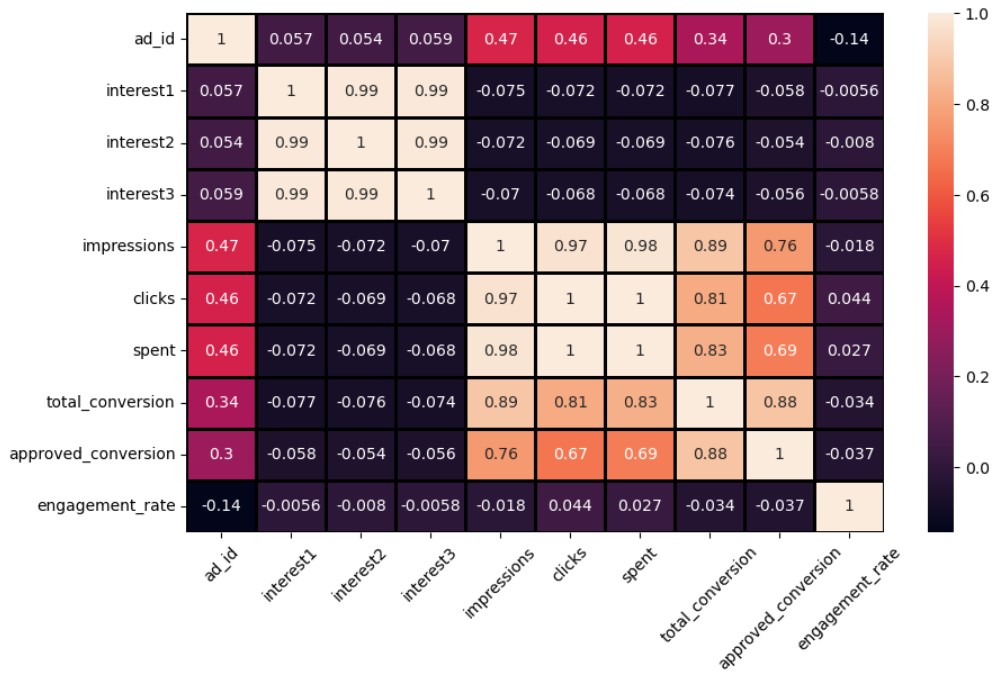
*Фигура 21. Времево разпределение за кампаниите по дни (графично)*

1. Допълнителен анализ на резултатите и извеждане на ключови изводи

Допълнителният анализ на резултатите от рекламни кампании, цели да идентифицира ключови метрики и фактори, влияещи върху ефективността.

### Анализ на числовите данни

След анализ на демографските атрибути (възраст, пол) и времевите данни (начало и край на кампаниите), фокусът е насочен към числовите характеристики. За визуализация на корелациите между числовите атрибути е конструирана карта на корелации (heatmap).

 *Фигура 22. Топлинна карта, отразяваща корелациите в данните*

На картата (*Фигура 22*) се вижда, че атрибутите "interest1", "interest2" и "interest3", представляващи интереси на таргетирани потребители, имат незначителна корелация с останалите характеристики. Те са изключени от по-нататъшния анализ.

### Описание на ключови атрибути

1. **spent** – разходи за всяка кампания.
2. **impressions** – брой показвания на рекламата.
3. **clicks** – брой кликвания върху рекламата.
4. **total\_conversion** – брой потребители, извършили действие (например покупка или инсталация на приложение) след клик.
5. **approved\_conversion** – брой активни потребители след реализация на конверсия.
6. **ad\_id** – уникален идентификатор на рекламата.

## Описание на основните променливи

1. **spent** - Сумата пари, изразходвана за всяка кампания.
2. **impressions** - Броят пъти, когато рекламата е била показана на потребителите.
3. **clicks** - Броят кликвания върху рекламата.
4. **total\_conversion** - Броят потребители, които са кликнали върху рекламата и са извършили действие, като покупка или инсталиране на приложение.
5. **approved\_conversion** - Броят конверсии, довели до активни потребители на приложението.
6. **ad\_id** - Уникален идентификатор на конкретна реклама.

### Обяснение на термините

* **Кампания**: Представлява основната цел (напр. повишаване на трафика).
* **Ad Set**: Определя аудиторията, бюджета, времевия график и местата за показване на рекламите.
* **Ad**: Включва конкретното съдържание (например изображения или видео).

Пример:  
Кампания "Летен Лансиране" има две рекламни групи:

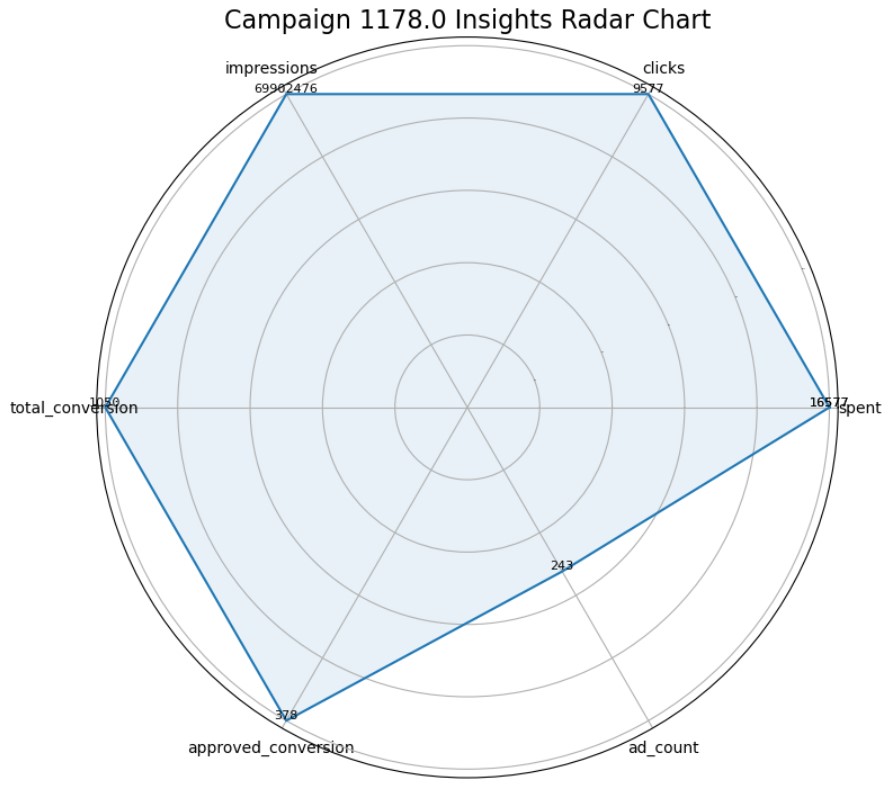
* Ad Set 1: Жени, 25–35 г., бюджет $50/ден, Instagram и Facebook.
* Ad Set 2: Мъже, 25–35 г., $30/ден, Facebook.

### Разходи, кликове и импресии

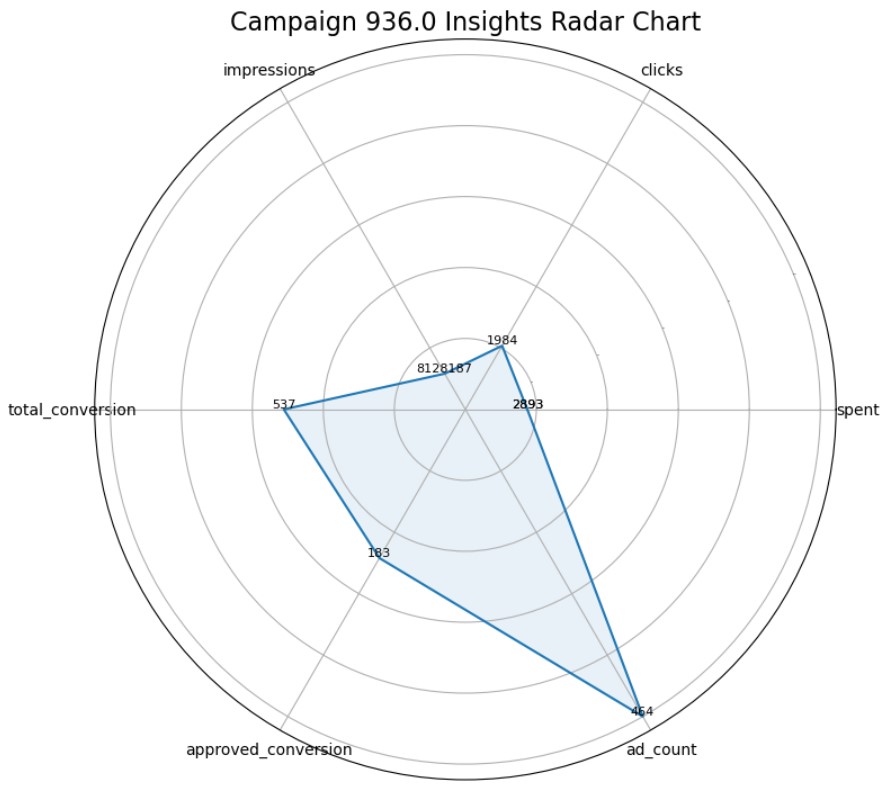
* **Кампания 1178**:
  + Разходи: $16,577
  + Кликове: 9,577
  + Импресии: 69,902,476
* **Кампания 936**:
  + Разходи: $2,893
  + Кликове: 1,984
  + Импресии: 8,128,187
* **Кампания 916**:
  + Разходи: $150
  + Кликове: 113
  + Импресии: 482,925

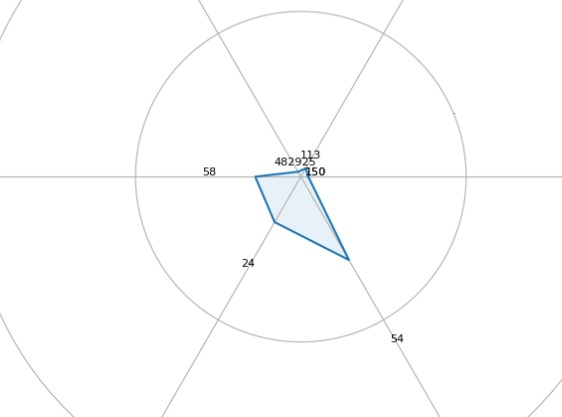
### Спайдър диаграми

Всички метрики са нормализирани за визуализация под формата на диаграма тип паяк.

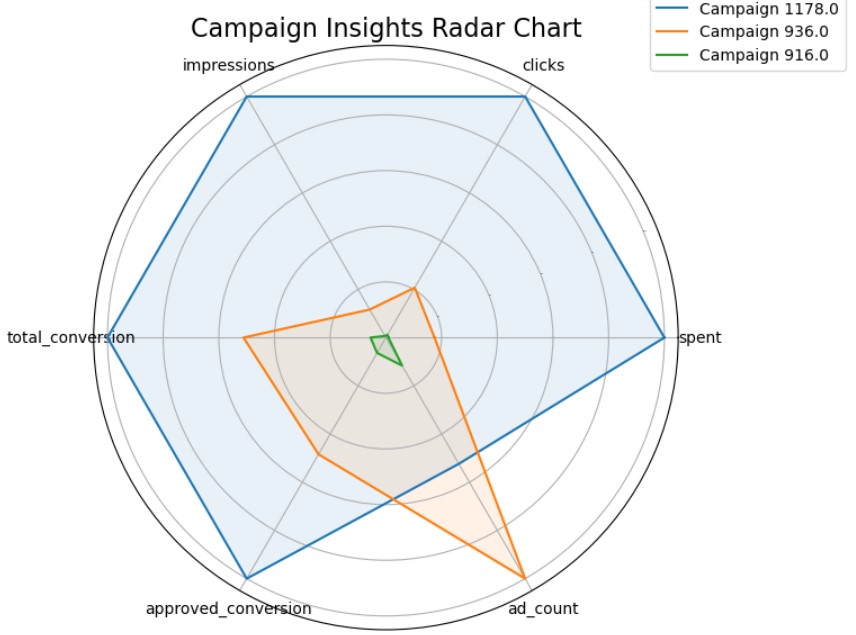


*Фигура 23. Спайдър диаграма за кампания 1178*

 *Фигура 24. Спайдър диаграма за кампания 936*



*Фигура 25. Приближен откъс от спайдър диаграма за кампания 916*



*Фигура 26. Обединена спайдър диаграма за всичките кампании*

### КПИ (ключови показатели за ефективност)

#### Кампания 1178

* Цена за клик: $1.73
* Цена за конверсия: $15.79
* Цена на реклама: $68.22
* Цена за одобрена конверсия: $43.85
* Цена за импресия: $0.00024

#### Кампания 936

* Цена за клик: $1.46
* Цена за конверсия: $5.39
* Цена на реклама: $6.24
* Цена за одобрена конверсия: $15.81
* Цена за импресия: $0.00036

#### Кампания 916

* Цена за клик: $1.32
* Цена за конверсия: $2.58
* Цена на реклама: $2.77
* Цена за одобрена конверсия: $6.24
* Цена за импресия: $0.00031

### Таблица (*Фигура 27)* с резултати обхваща допълнителни КПИ:

*Фигура 27. Таблица с допълнителни КПИ*

Метриките включват разходи (за реклама, показване, клик, конверсия и одобрена конверсия), както и проценти за преобразуване на показване в клик, клик в обща конверсия и обща конверсия в одобрена конверсия. Таблицата дава ясна картина за сравнителната ефективност на всяка кампания:

#### 1. Ефективност на разходите

* **Цена на реклама (CPA):**  
  Кампания 1178 има най-висока цена на реклама - $68.22, което отразява значителен бюджет за индивидуални реклами в сравнение с Кампания 936 ($6.24) и Кампания 916 ($2.77). Кампания 916 постига най-ниската CPA, което показва икономичен подход или по-голям брой реклами спрямо общия бюджет.
* **Цена на показване (CPI):**  
  Кампания 1178 има най-ниска CPI - $0.000237, следвана от Кампания 916 ($0.00031) и Кампания 936 ($0.000356). Това показва, че Кампания 1178 ефективно е използвала бюджета си за генериране на голям обем показвания, демонстрирайки потенциал за мащабиране при по-големи бюджети.
* **Цена на клик (CPC):**  
  Кампания 916 има най-ниска CPC - $1.32, което я прави най-рентабилна в привличането на потребителска ангажираност. Въпреки по-големия бюджет и експозиция на Кампания 1178, нейната CPC е $1.73, докато Кампания 936 се представя малко по-добре с $1.46. Това предполага, че Кампания 1178 може да има слабо качество на рекламите или неточни таргетинг стратегии.
* **Цена на конверсия (CPCO):**  
  Кампания 916 постига най-ниската CPCO - $2.58, следвана от Кампания 936 ($5.39) и Кампания 1178 ($15.79). Това показва, че рекламната стратегия на Кампания 916 ефективно привлича интереса на потребителите, водейки до успешни конверсии.
* **Цена на одобрена конверсия (CPAC):**  
  Кампания 916 запазва предимството си с CPAC от $6.24, надминавайки Кампания 936 ($15.81) и Кампания 1178 ($43.85). Ниската CPAC на Кампания 916 предполага ангажиращ продукт с висока степен на задържане на клиентите.

#### 2. Метрики за ангажираност

* **Вероятност показване да стане клик (CTR):**  
  Кампания 916 води с най-висок CTR - 0.0234%, което отразява добре таргетирана аудитория или привлекателни рекламни послания. Кампания 936 е близка с 0.0244%, докато Кампания 1178 изостава значително с 0.0137%, което предполага възможни проблеми с релевантността на рекламите или таргетирането.

#### 3. Метрики за конверсии

* **Вероятност клик да стане обща конверсия:**  
  Кампания 916 превъзхожда с 51.33% от кликванията, водещи до конверсии, което показва висока ефективност на пост-клик взаимодействията или офертите. Кампания 936 има по-ниска ефективност на конверсия от 27.07%, докато Кампания 1178 изостава с 10.97%, което подчертава необходимостта от оптимизация на потребителския опит след клик.
* **Вероятност обща конверсия да стане одобрена конверсия:**  
  Кампания 916 постига най-висок процент с 41.38% от общите конверсии, които стават одобрени, показвайки по-добро задържане или активиране на потребителите. Кампания 936 има 34.08%, докато Кампания 1178, въпреки големия си бюджет, достига едва 36%, подчертавайки потенциални неефективности в процесите за одобрение на потребители.

1. Заключение и изводи

### Основни изводи

* **Кампания 916** демонстрира изключителна ефективност на разходите във всички метрики, като се отличава в конверсии и ангажираност. Представянето ѝ подсказва добре оптимизирана стратегия за скромния си бюджет.
* **Кампания 936** се представя добре по отношение на разходни метрики като CPC и CPAC, но изостава в ангажираността и успешните конверсии спрямо Кампания 916.
* **Кампания 1178** се възползва от значително по-голям бюджет, генерирайки впечатляващ брой показвания и кликвания. Въпреки това, високата CPAC и ниските проценти на конверсии разкриват неефективности, които изискват допълнителна оптимизация.

### Препоръки

1. **Оптимизиране на Кампания 1178:** Анализирайте факторите, допринасящи за високата CPAC и ниските проценти на конверсии, като подобрите таргетирането, качеството на рекламите и потребителския опит след клик.
2. **Подобряване на ангажираността за Кампания 936:** Провеждайте A/B тестове, за да увеличите CTR и процента на конверсии чрез промени в креативите и прецизиране на аудиторията.
3. **Мащабиране на Кампания 916:** Разпределете допълнителни ресурси за Кампания 916, тъй като нейната изключителна ефективност предполага голям потенциал за мащабиране с минимални загуби в представянето.
4. Сорс код

Предоставяме линк към GitHub:

<https://github.com/savina01/Analysis-of-Facebook-ad-campaign>

1. Информационни източници

* <https://www.webfx.com/blog/social-media/social-media-analysis/>
* <https://www.kaggle.com/datasets/madislemsalu/facebook-ad-campaign>