

**Технически университет – София**

**Факултет по приложна математика и информатика**

**Курсова работа**

**по**

**Анализ на бизнес данни в социални мрежи**

**на тема**

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

**Изготвили: Савина Вълчанова, Денис Цолов, Цветомир Цветков**

**Фак. номера:** **961324002, 961324006, 961324011**

**Група: 252**

Съдържание

1. **Въведение**
2. **Цел**
3. **Експериментална рамка – методи и техника на анализ**
4. **Обработка и анализ на данните**
5. **Представяне и визуализация на резултатите**
6. **Допълнителен анализ на резултатите и извеждане на кл. изводи**
7. **Заключение – изводи и препоръки**
8. **Сорс код**
9. **Информационни източници**
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 съдържа контекст за отделните кампаниим реклами от тях и времевата им рамка,
* **Разходи**: Използваният бюджет за отделните кампании
* **Демографски данни**: Данните за възраст и пол на аудиторията дават възможност за анализ на целевите групи на таргетирането,
* **Импресии, кликвания и конверсии**: Тези метрики измерват обхвата и успеха на кампаниите в задържане на ангажираността на потребителите,
* **Интереси**: допълнителни характеристики на таргерираните профили.

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 както и да постави оценка на ефективността на рекламната кампания и формулиране на препоръки за оптимизация на маркетинговите стратегии. Целта може да се разгледа като комбинация от следните точки:

* Демография:
  + Колони: age, gender
  + Цел: Анализиране на възрастовите групи и пола на потребителите, които взаимодействат с кампанията.
* Анализ на ангажираността:
  + Колони: impressions, clicks, total\_conversion, approved\_conversions, interest1, interest2, interest3
  + Цел: Изучаване на интересите на потребителите, за да се види кои типове съдържание са най-популярни и ефективни.
* Времеви анализ:
  + Колони: reporting\_start, reporting\_end
  + Цел: Изследване на времевите периоди на кампанията и идентифициране на пиковите времена на активност.
* Идентификация на кампаниите:
  + Колоната campaign\_id
  + Цел: Разграничаване на различните кампании и проследяване на тяхната ефективност и въздействие.

1. **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, add\_id, age, gender, interest1, interest2, interest3, impressions, clicks, spent, total\_conversion, approved\_conversion
* Период на събиране: Данните обхващат периода от 17 до 30 август, 2017 година.
  + - 1. 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.
  + - 1. 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 ангажираността:
  + Метод: Изследване на цялостната ангажираност спрямо възрастовите групи и броя кликвания върху реклама.
  + Техника: Създаване на бар графики и проследяване ан разпределението с помощтта на библиотеката Seaborn.
* 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.
* Вътрешен анализ за всяка кампания:
  + Метод: Изследване на ангажираност на таргетираните потребители и ефективността на кампанията спрямо похарчения и бюджет.
  + Техника: Използване на спайдър диаграми и таблица за лесно проследяване на получените резултатни метрики.
    - 1. 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.
  + - 1. Техника за интерпретиране на резултатите
* Описание на резултатите: Подробно описание на наблюдаваните тенденции и модели в данните. Определяне на ключовите открития спрямо резултатите.
* 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

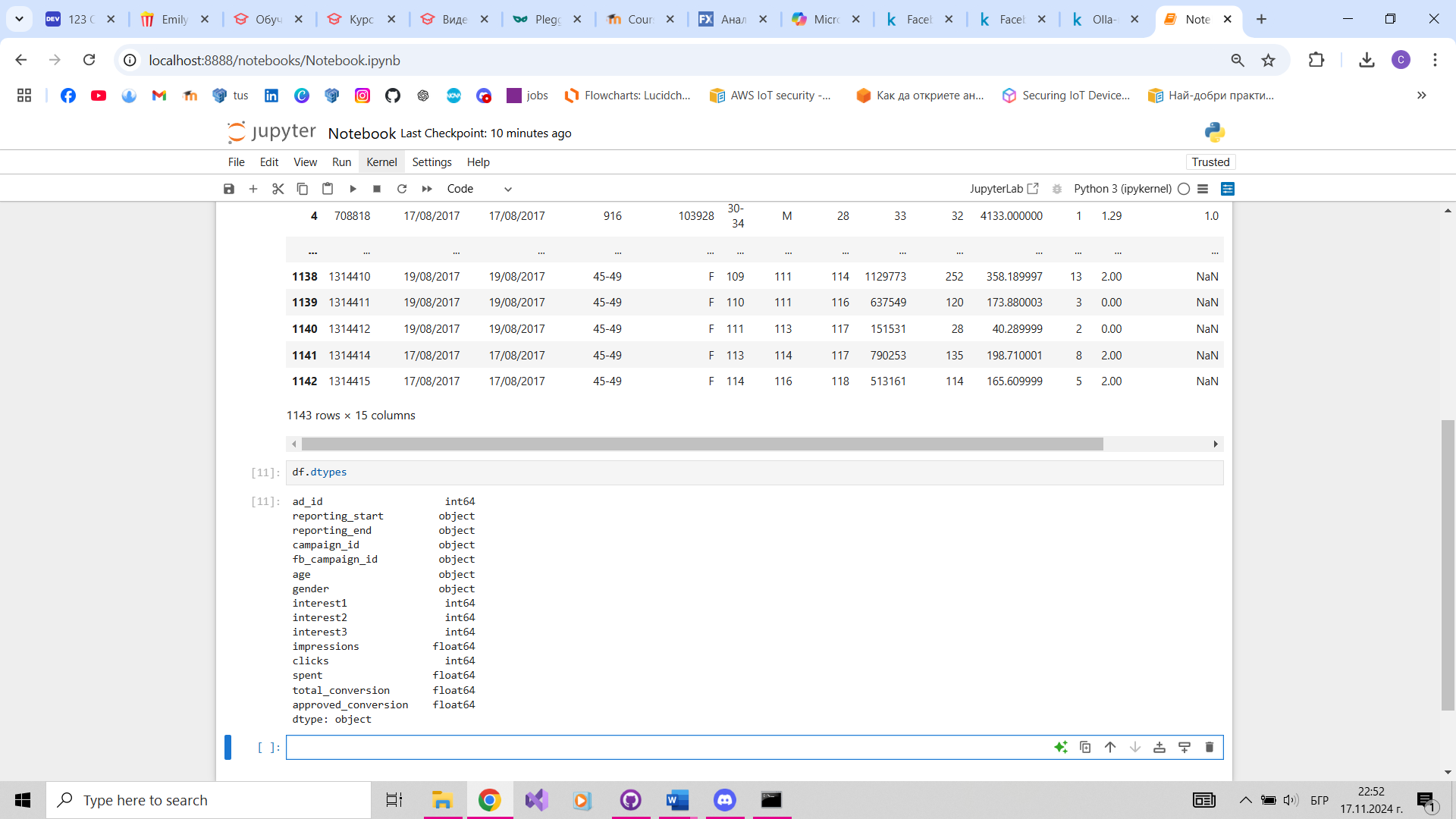
На *Фигура 1* виждаме несъответствие при стойностите на колоните при campaign\_id, fb\_campaign\_id, age, gender. Има липсващи данни и в колоните total\_conversions, approved\_conversion. Тези липси в данните ор ред 762 напред в датасета са непоправими и поради това за проекта ще се използва само първата половина с коректно структурираните данни.

A screenshot of a computer

Description automatically generated

*Фигура 1.Използван набор от данни*

* 1. Анализ на набора от данни



*Фигура 2. Типове данни*

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

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*

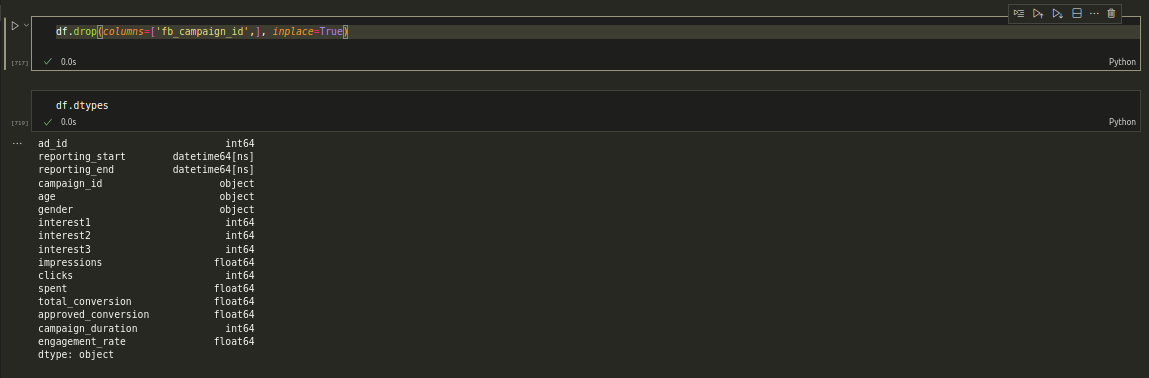
Проверяваме за дубликати (*Фигура 7*) и установяваме, че няма.

A screenshot of a computer

Description automatically generated

*Фигура 8. Безкрайни и NaN стойности*

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

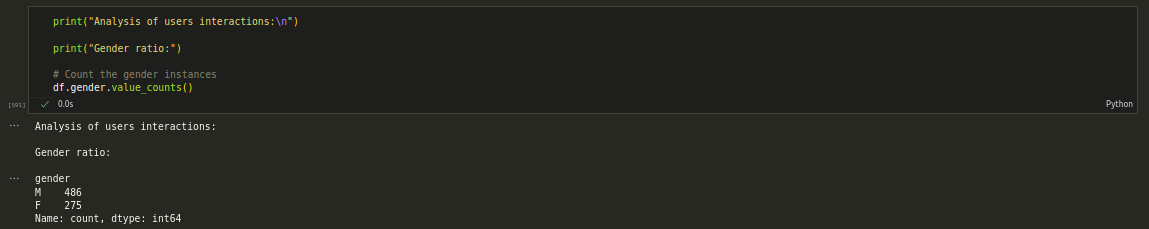


*Фигура 9. Изчистен и готов файл за анализ*

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

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

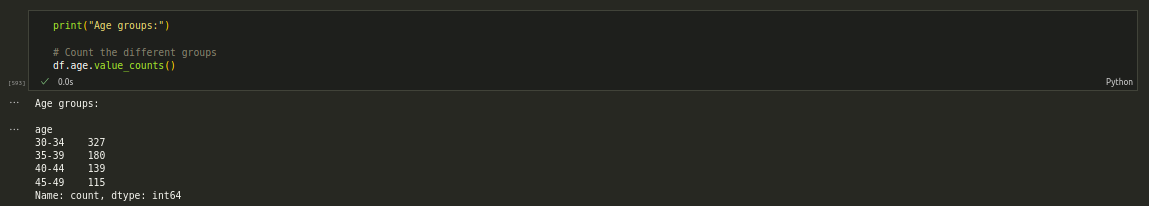
Броят на участващите мъже е близо **1.8** пъти повече от този на жените. (*Фигура 9, 10*)

*Фигура 9. Демографско разделение*

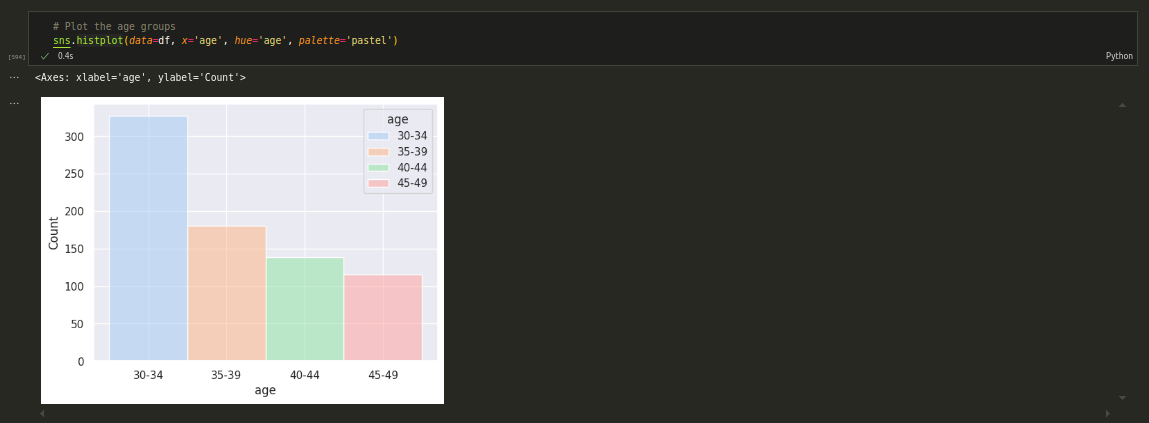


*Фигура 10. Демографско разделение (графично)*

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*)

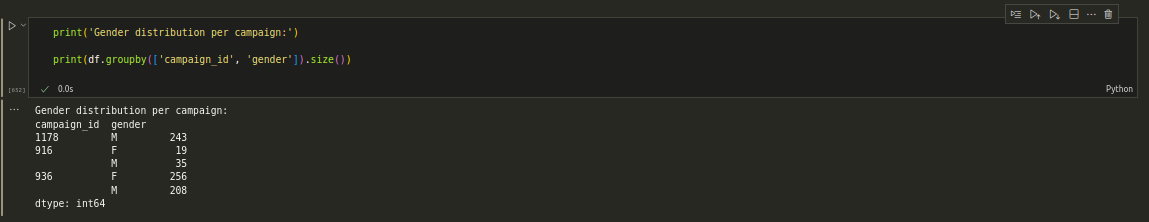


*Фигура 11. Възрастово разделение*

**

*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*)

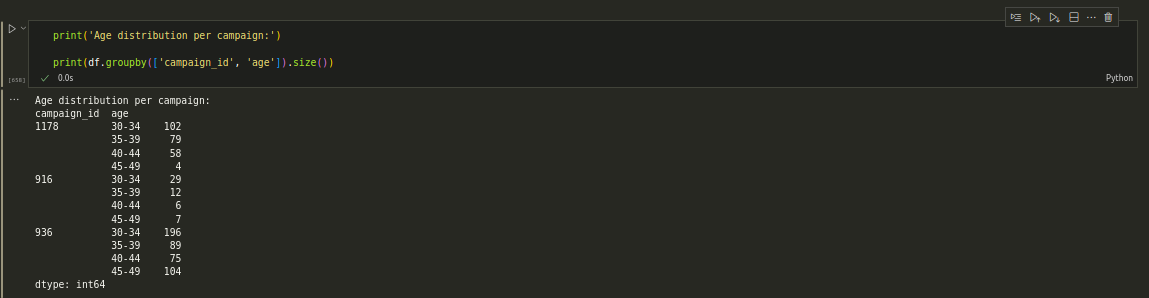


*Фигура 13. Полово разпределение спрямо кампаниите*

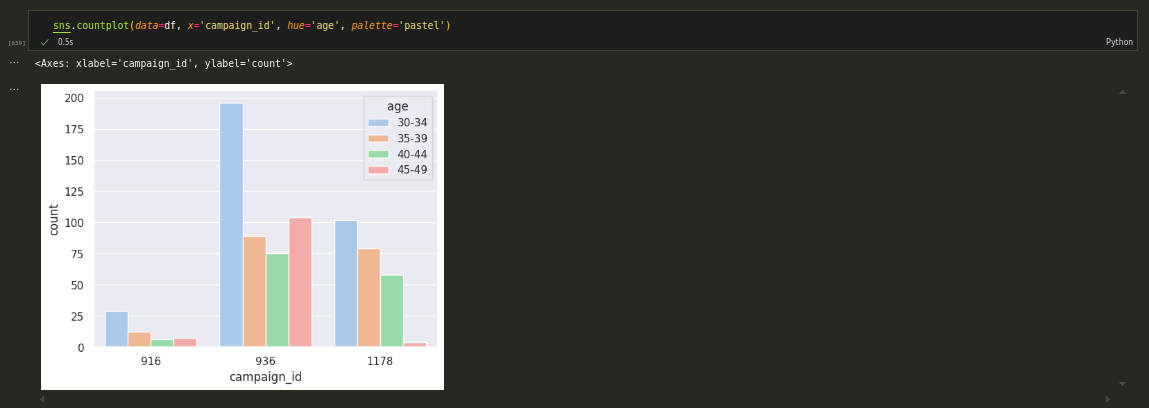


*Фигура 14. Полово разпределение спрямо кампаниите (графично)*

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*)



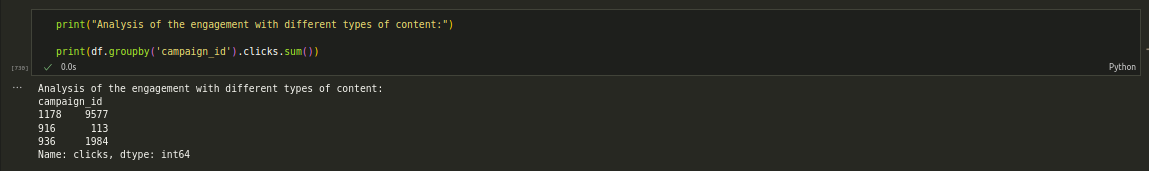
*Фигура 15. Възрастово разпределение спрямо кампаниите*

**

*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*)



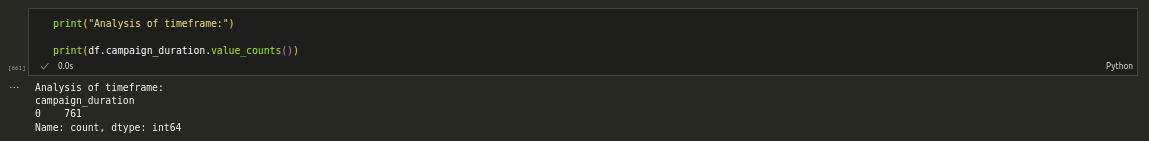
*Фигура 17. Брой кликове за всяка кампания*

**

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

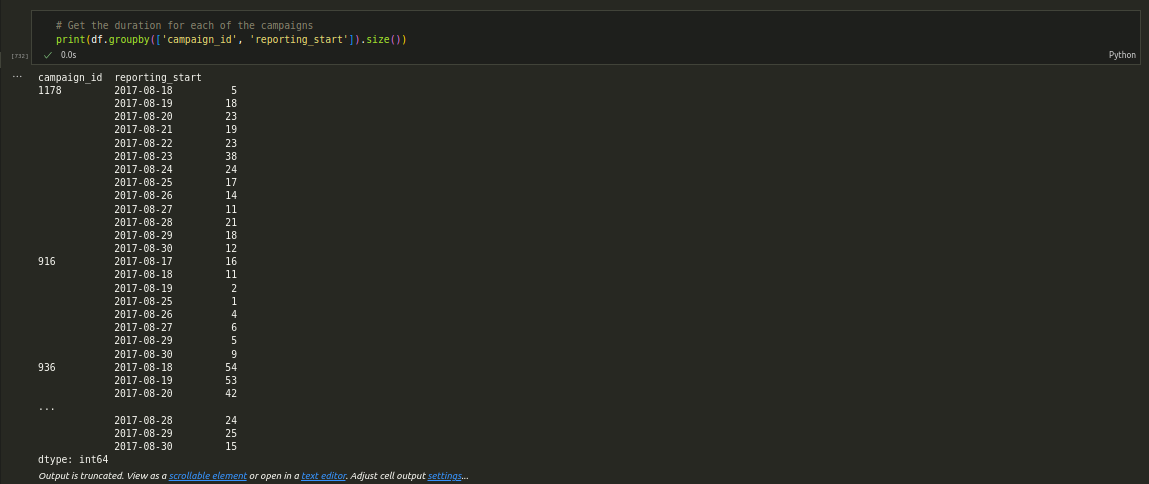
* + - 1. Времеви анализ

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)*

**

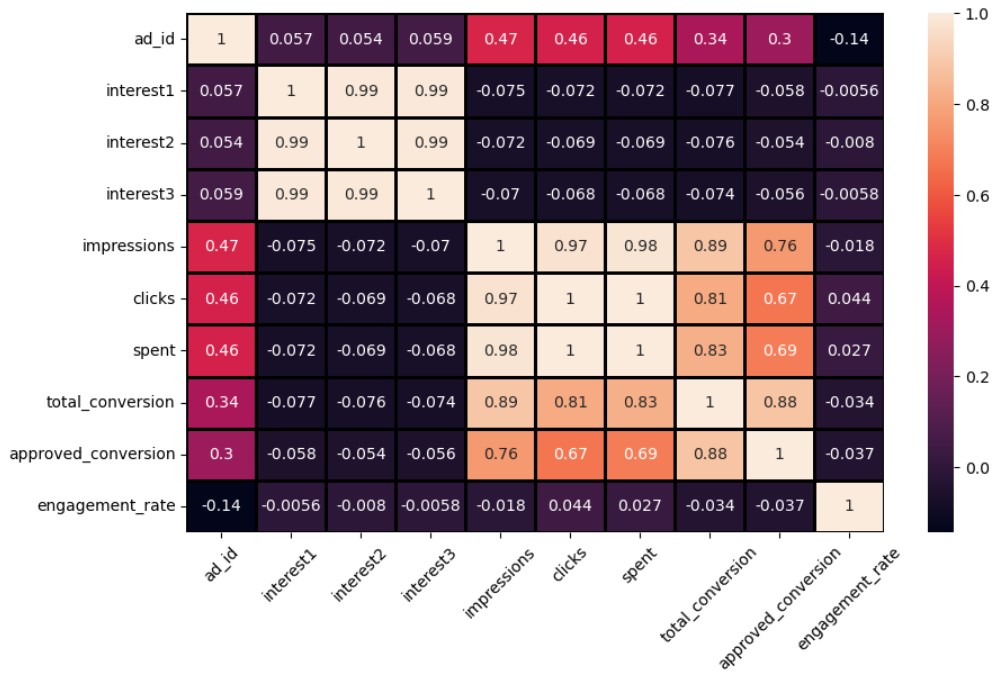
*Фигура 21. Времево разпределение за кампаниите по дни (графично)*

1. **Допълнителен анализ на резултатите и извеждане на ключови изводи**

Допълнителният анализ на резултатите от рекламните кампании цели да идентифицира ключови метрики и фактори, влияещи върху ефективността за всяка една кампания по отделно.

### 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. Затова те биват изключени от по-нататъшния анализ.

### Описание на основните атрибути:

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

## Събиране на контекст за това какво представлява агенция:

### За да разберем структурата на анализа, първо трябва да можем да разграничaваме различните термини и да знаем тяхното значение:

### Кампания (Add Campaign): Това е структурата от най-високо ниво, която представлява общата цел или задача (напр. увеличаване на трафика, генериране на потенциални клиенти или повишаване на осведомеността за марка).

### Набор от реклами (Add set): Средното ниво в рамките на кампанията, което определя аудиторията, бюджета, графика и разположението на рекламите. Всеки набор от реклами има уникален идентификатор на набора от реклами.

### Реклама (Add): Най-долният слой, който съдържа действителните видове креативи (например изображения, видеоклипове и текст). Всяка реклама има уникален идентификатор на рекламата.

### Въображаем примерза на кампания с цел превръщане на термини в реалните им еквиваленти – въображаемата кампанията „Представяне на нашата дизайнерска лятната колекция“ има:

### Ad Set 1: Аудитория: Жени, на възраст 25-35 години, интересуващи се от мода; Бюджет: 50 USD/ден; График: 1 юни - 15 юни; Насоченост: Instagram Stories и Facebook Feed

### Реклама 1: креатив - изображение на лятна рокля с текст „20% отстъпка“. CTA: „Пазарувай сега“ ID на рекламата: 334455

### Реклама 2: креатив - видеоклип, представящ модел, облечен в лятната колекция. CTA: „Вижте колекцията“ ID на рекламата: 334456

### Ad Set 2: Мъже на възраст 25-35 години, интересуващи се от ежедневно облекло; 30 USD на ден; 1 юни - 15 юни; Facebook Feed

### Реклама 1: ...

### Реклама 2: ...

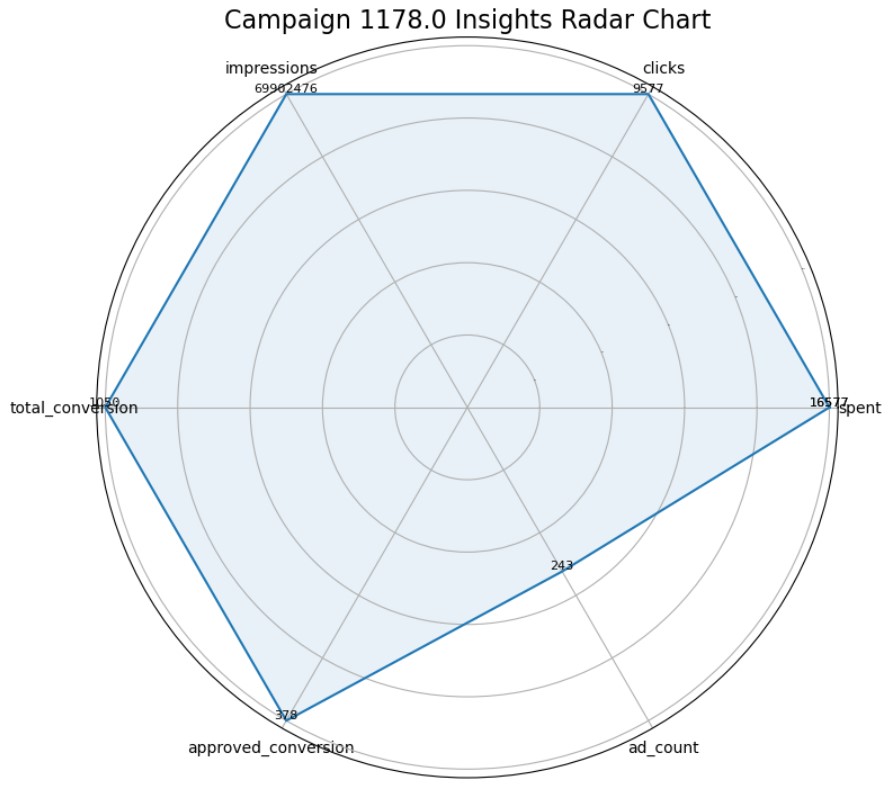
### Анализ на разходи, кликове и импресии

### Проследяваме метриките, които получаваме чрез групиране на набора от данни за всяка една кампания:

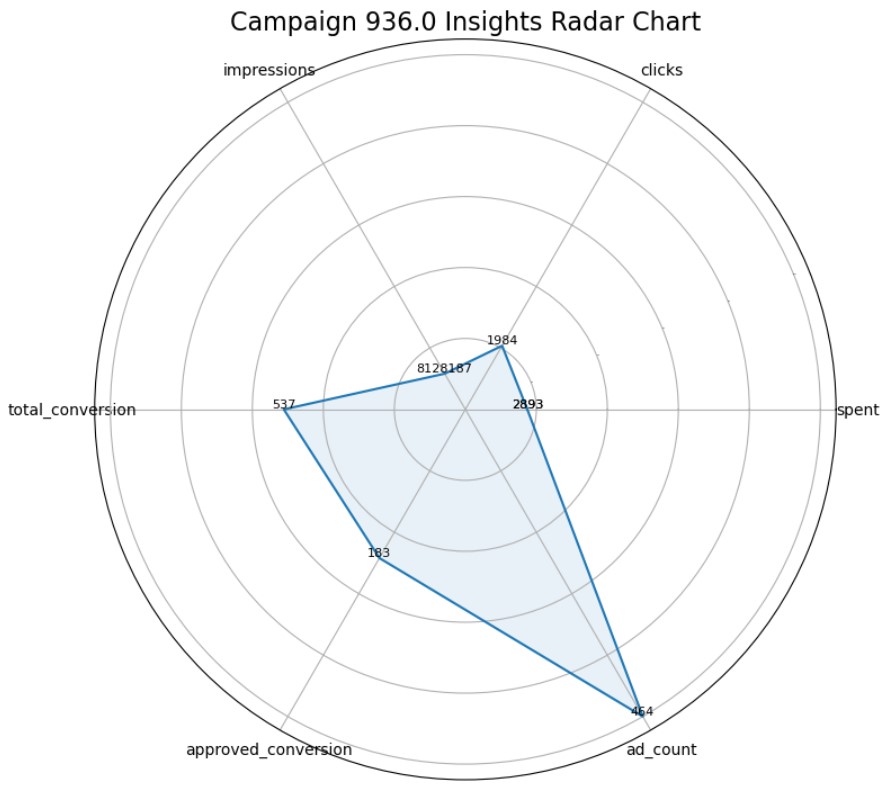
* **Кампания 1178:**
* Разходи: **16,577.16**
* Кликове: **9,577**
* Импресии: **69,902,476**
* Общо конверсии: **1,050**
* Одобрени конверсии: **378**
* Брой реклами: **243**
* **Кампания 936:**
  + Разходи: **2,893.37**
  + Кликове: **1,984**
  + Импресии: **8,128,187**
  + Общо конверсии: **537**
  + Одобрени конверсии: **183**
  + Брой реклами: **464**
* **Кампания 916:**
  + Разходи: **149.71**
  + Кликове: **113**
  + Импресии: **482,925**
  + Общо конверсии: **58**
  + Одобрени конверсии: **24**
  + Брой реклами: **54**

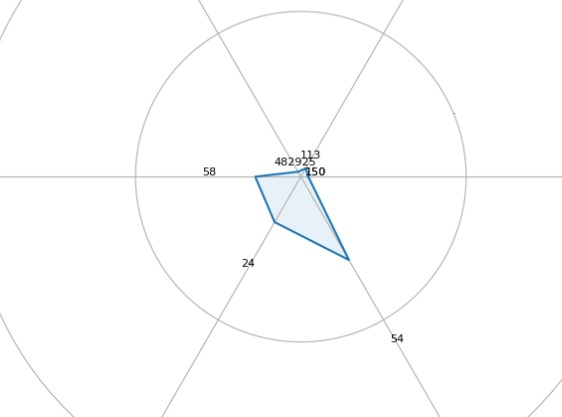
### Визуализация чрез спайдър диаграми

За постигане на максимална разбираемост и проследимост използваме онагледяване на данните чрез спайдър диаграми. За тази цел, данните се нормализират и се използва модула pyplot от библиотеката matplotlib (*Фигура 23, 24, 25*). Накрая отделните спайдър диаграми се комбинират в една обща, където могат да се правят сравнения по лесен визуален начин (*Фигура 26)*.

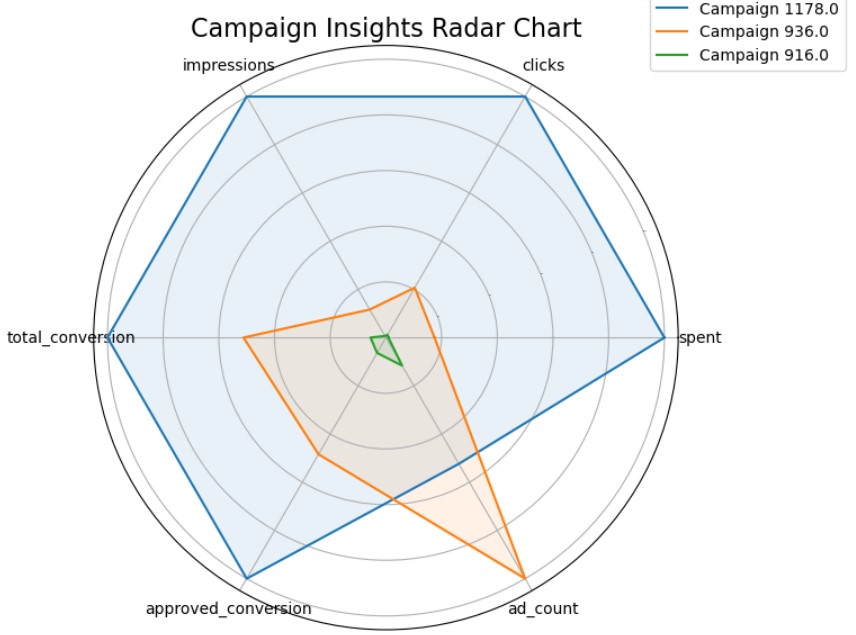


*Фигура 23. Спайдър диаграма за кампания 1178*

 *Фигура 24. Спайдър диаграма за кампания 936*



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

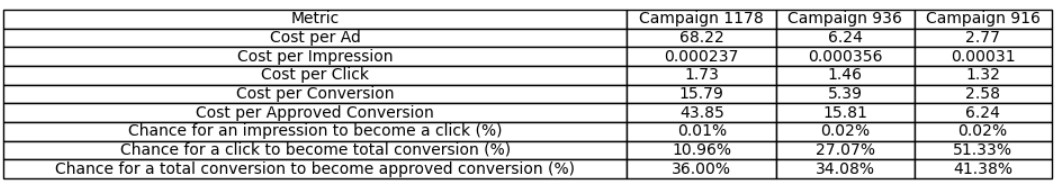


*Фигура 26. Обединена спайдър диаграма за всичките кампании*

### Анализ на ефикасността спрямо разходите и проследяване на ангажираността по веригата на преобразуване на посетител на рекламата в клиент

### Чрез допълнителни прости изчисления върху данните от анализа на разходи, кликове, импресии, комбинирани с колоните засягащи ангажираността (impressions, clicks, total\_conversions, approved\_conversions), достигаме до следните заключения за всяка отделна кампания:

* **Кампания 1178:**
  + **Цена на клик: 1.73**
  + **Цена на конверсия: 15.79**
  + **Цена на реклама: 68.22**
  + **Цена на одобрена конверсия: 43.85**
  + **Цена на импресия: 0.000237**
* **Кампания 936:**
  + **Цена на клик: 1.46**
  + **Цена на конверсия: 5.39**
  + **Цена на реклама: 6.24**
  + **Цена на одобрена конверсия: 15.81**
  + **Цена на импресия: 0.000356**
* **Кампания 916:**
  + - **Цена на клик: 1.32**
    - **Цена на конверсия: 2.58**
    - **Цена на реклама: 2.77**
    - **Цена на одобрена конверсия: 6.24**
    - **Цена на импресия: 0.00031**

С помощта на matplotlib, тези данни се синтезират до лесно-проследима таблица (*Фигура 27)*.

*Фигура 27. Таблица с допълнителни КПИ*

Метриките в таблицата 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:

#### 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.

#### 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.

#### 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 – изводи и препоръки**

### 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

* **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.
* **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.
* **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.

1. **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>