

**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, Цветомир Цветков

**Фак. номера:** 961324002, 961324011

**Група:** 252

# **Съдържание**

[**Съдържание 2**](#_2uvn22gadu7u)

[**1. Въведение 3**](#_krfstm68i8li)

[**2. Цел 3**](#_o08x00ycrio1)

[**3. Експериментална рамка 4**](#_ormkxokr8vou)

[3.1. Набори данни (Datasets): 4](#_qk7yv7gxpdkt)

[3.2. Обработка на данни: 5](#_cyu6rxm2d01i)

[3.3. Избор на метод и техника за анализ: 5](#_p5wpm6rkf8l6)

[3.4. Използвани библиотеки и софтуерни средства: 5](#_41fpdvoha5yj)

[3.5. Техника за интерпретиране на резултатите 6](#_trajl7tsgwla)

[**4. Обработка и анализ на данните 6**](#_8fn045ignzh6)

[4.1. Набор от данни 6](#_e7ar129euro2)

[4.2. Анализ на набора от данни 7](#_b23u4nu8c8qa)

[**5. Представяне и визуализация на резултатите 11**](#_zeun8u4vxys5)

[5.1. Регионален ценови анализ 11](#_77b5yvc4ohld)

[5.2 Анализ по типове 13](#_om8i9nvjipvv)

[5.3 Синтактичен анализ 15](#_1x9ma1w5hbx)

[**6. Заключение 1**](#_g8nvpv1ak0ue)**8**

[**7. Сорс код 18**](#_si9kretzp5xc)

[**8. Информационни източници 18**](#_qn6fkpfv5vgs)

# **1. Въведение**

Анализът на обяви за продажба е от съществено значение за разбиране на тенденциите на пазара и тяхната ефективност. Представените данни обхващат информация за различни обяви за продажба, като се фокусират върху показатели като местонахождение, тип, цена и други. Чрез детайлен анализ на тези данни можем да извлечем ценна информация за ценовите различия, предпочитанията на потребителите и пазарните тенденции в различните градове.

Прегледът на данните разкрива различни аспекти на продажбите, като:

* Най-популярните категории продукти и техните ценови диапазони.
* Разликите в предпочитанията на потребителите в различни градове.
* Влиянието на местоположението върху цените и обема на продажбите.

Чрез анализа на тези данни можем да идентифицираме най-новите тенденции и да оценим въздействието на различни фактори.

# **2.** **Цел**

Целта на тази разработка е да анализира обяви за продажба в пазара на мрежата Facebook - Marketplace, както и да постави оценка на най-търсените активи, техните цени и местата, където те биват предлагани. Целта също може да бъде разгледана и като комбинация от следните точки:

* **Регионален ценови анализ:**
  + **Колони:** City, Price
  + **Цел:** Анализ на цените и изучаване на ценовото разпределение и средните цени за всеки град
* **Анализ на типа:**
  + **Колони:** Title, Item Type
  + **Цел:** Събиране на информация относно най-продаваните активи и техния вид
* **Разпределителен анализ:**
  + **Колони:** City
  + **Цел:** Анализ на броя обяви според местонахождението на продавания артикул

# **3. Експериментална рамка**

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

## **3.1. Набори данни (Datasets):**

* **Източник на данни:**
  + Данните са събрани с помощта на “Instant Data Scraper” Google Chrome Extension.
* **Обхванати колони:**
  + Name, City, Price, Item Type
* **Период на събиране:**
  + Данните обхващат периода от една седмица

## **3.2. Обработка на данни:**

* **Почистване на данни:**
  + Справяне с липсващи стойности и коригиране на грешки в данните.
* **Преобразуване на данни:**
  + Форматиране на числови и категорийни стойности и модифициране на събрани данни с по-лесен анализ

## **3.3. Избор на метод и техника за анализ:**

* **Регионален ценови анализ:**
  + **Метод:** Сравнение на средни цени по региони
  + **Техника:** Анализ и визуализация чрез хистограми и бар плотове
* **Анализ по типове:**
  + **Метод:** Класификация и сравнение на категории
  + **Техника:** Анализ и визуализация чрез пай диаграми и хийт мапове
* **Синтактичен анализ:**
  + **Метод:** Извличане на отделни думи
  + **Техника:** Анализ и визуализация чрез бар плотове

## **3.4. Използвани библиотеки и софтуерни средства:**

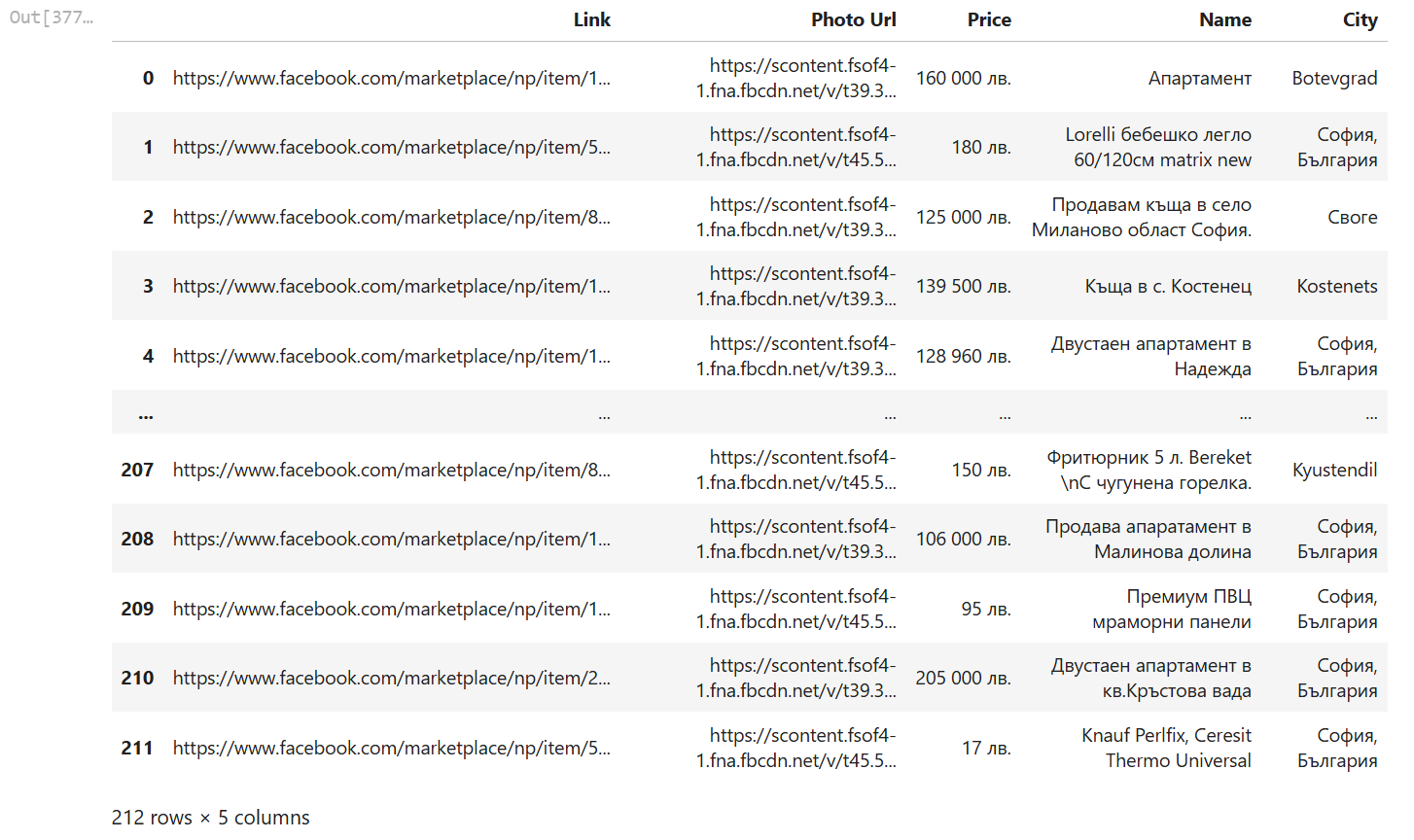
* **Instant Data Scraper**: За създаване и извличане на дейтасета
* **Pandas**: За манипулиране и анализ на данните
* **NumPy**: За математически изчисления и операции
* **Matplotlib**: За създаване на графики и визуализации на резултатите
* **Seaborn**: За по-сложни и изящни визуализации

## **3.5. Техника за интерпретиране на резултатите**

**Описание на резултатите:** Подробно описание на наблюдаваните тенденции и модели в данните. Определяне на ключовите открития спрямо резултатите.

# **4. Обработка и анализ на данните**

## **4.1. Набор от данни**

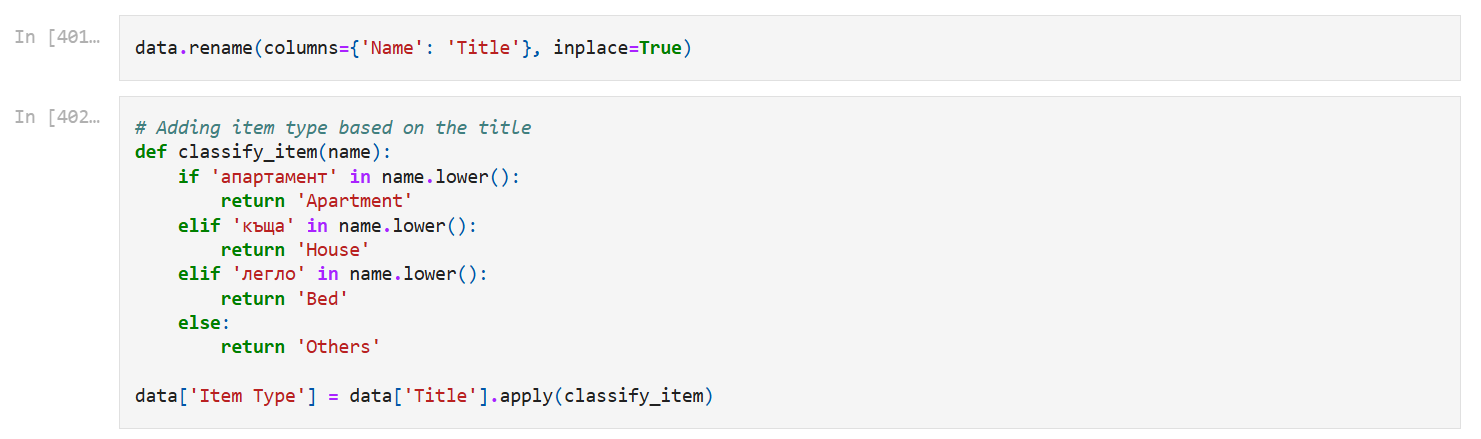
*Фигура 1.Използван набор от данни*

## **4.2. Анализ на набора от данни**

*Фигура 2. Типове данни*

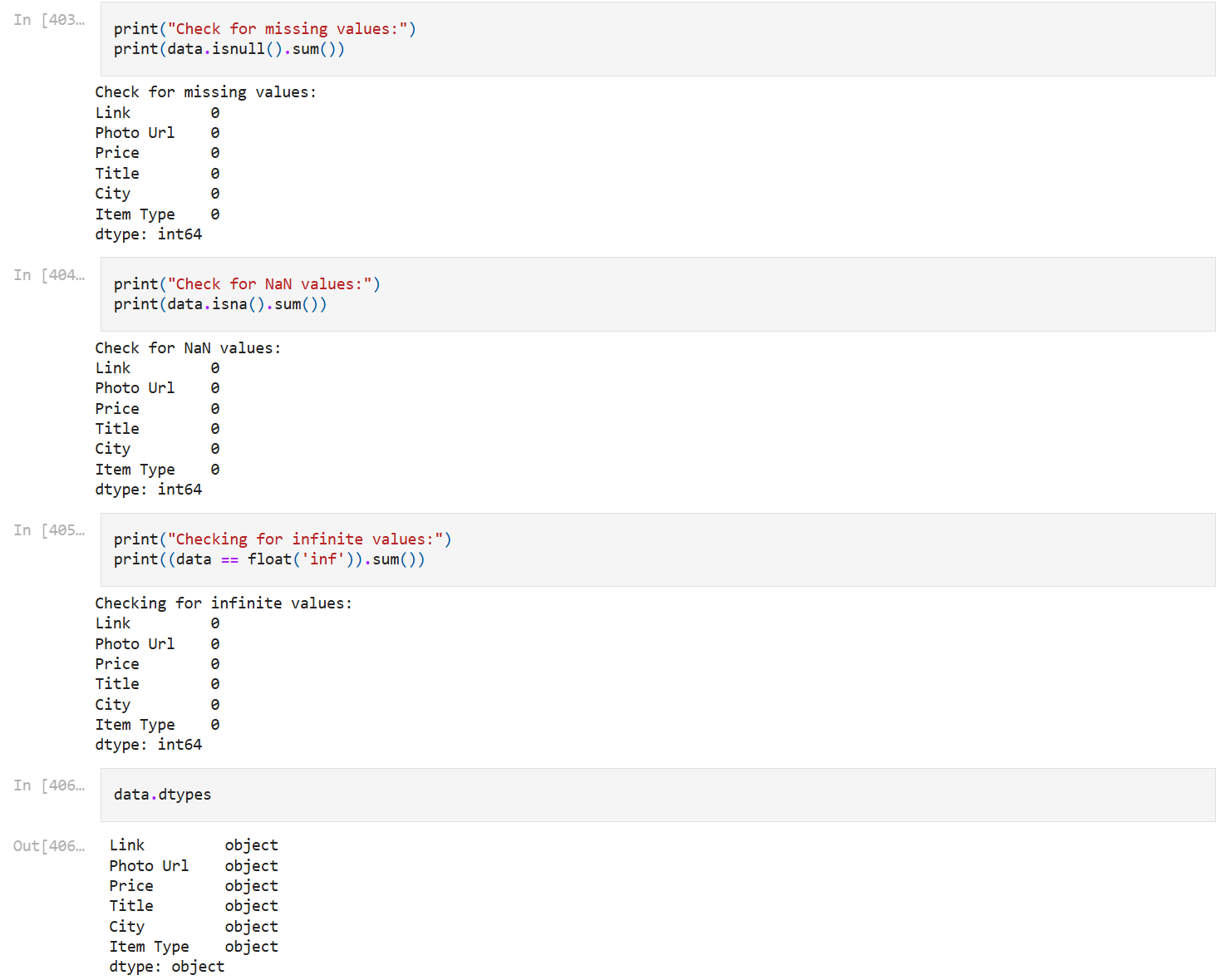
Визуализирайки типовете (*Фигура 2*), забелязваме, че всички колони са от тип object и можем да пристъпим към предварителната обработка.

Преименуваме колоната *Name* на *Title* за по-акуратно репрезентиране на информацията и добавяме нова колона *Item Type*, която да съдържа информацията за типа на обявата (*Фигура 3*).???????



*Фигура 3. Преименуване и създаване на колона за тип*

Правим проверка за липсващи, безкрайни или нечислови стойности (*Фигура 4*) и виждаме, че в дейтасета такива няма.



*Фигура 4. Проверка за липсващи, безкрайни или нечислови стойности*

Правим проверка, дали успешно може да конвертираме цената към числен тип, като преди това почистваме редовете от ненужни символи (*Фигура 5*).



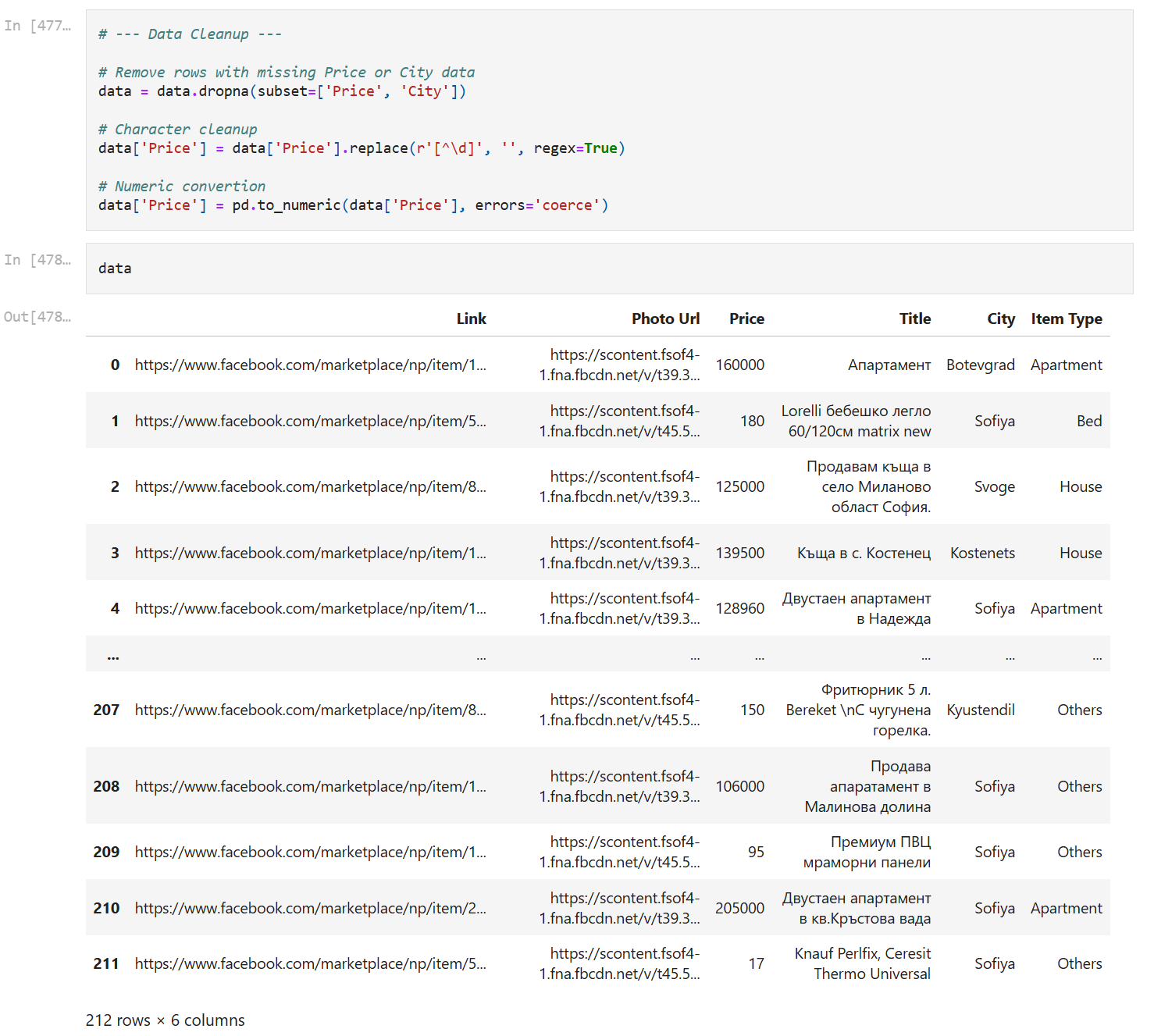
*Фигура 5. Почистване на Price колоната*

Уеднаквяваме стойностите на колоната *City* с цел събиране на обявите, където местоположението съвпада, но градът в обявата е написан на кирилица вместо на латиница. (*Фигура 6*).

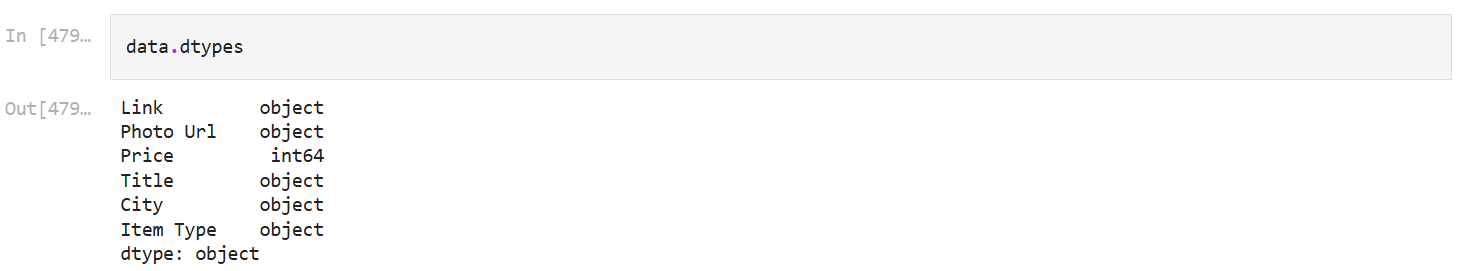


*Фигура 6. Уеднаквяване на градовете в таблицата*

Правим последно почистване на дейтасета като премахваме редове, където липсват цена или град и официално конвертираме цената към числов тип (*Фигура 7*) като по този начин сме приключили с предварителната обработка на данните (*Фигура 8)*.



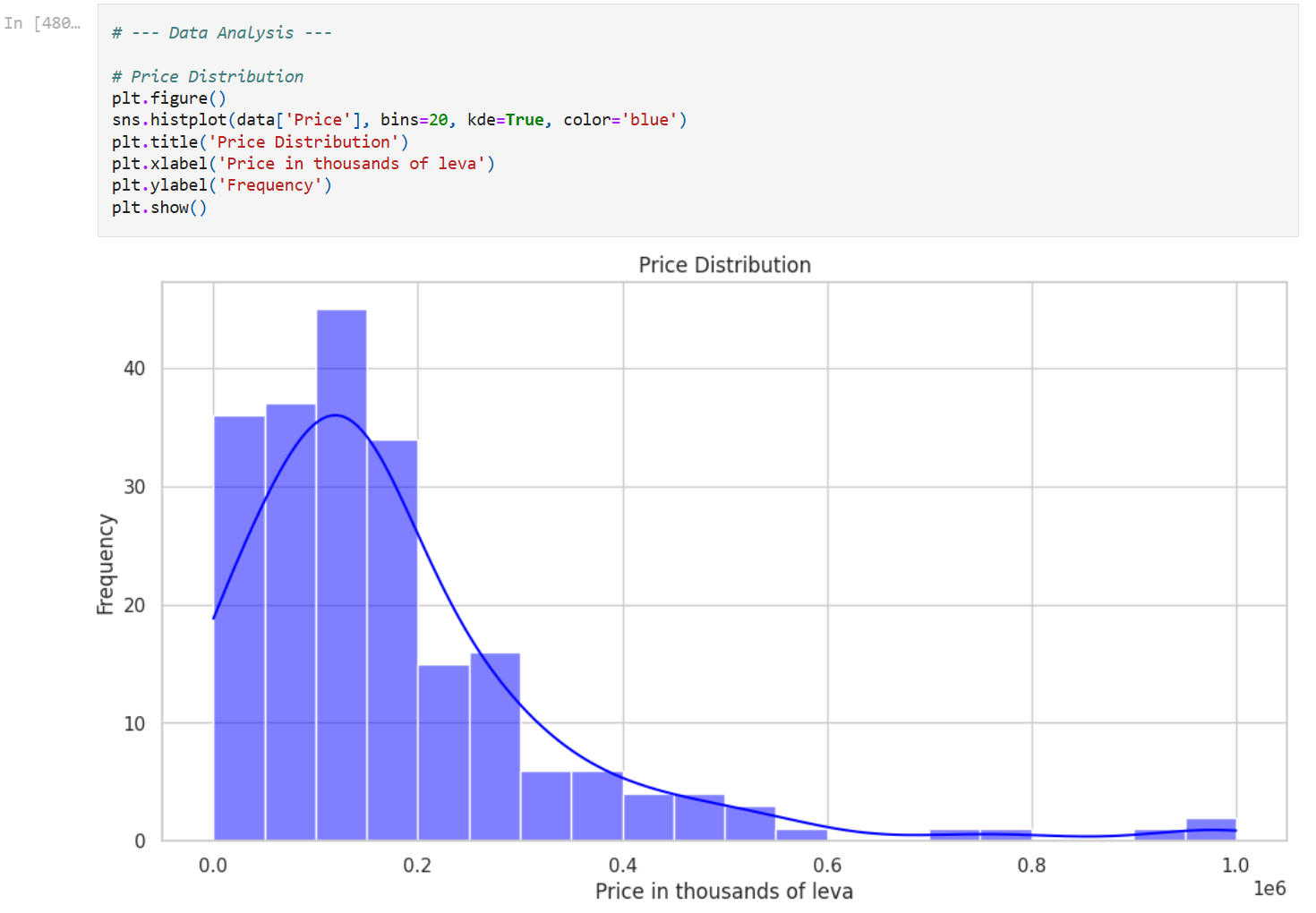
*Фигура 7. Обновен дейтасет*



*Фигура 8. Изчистен и готов файл за анализ*

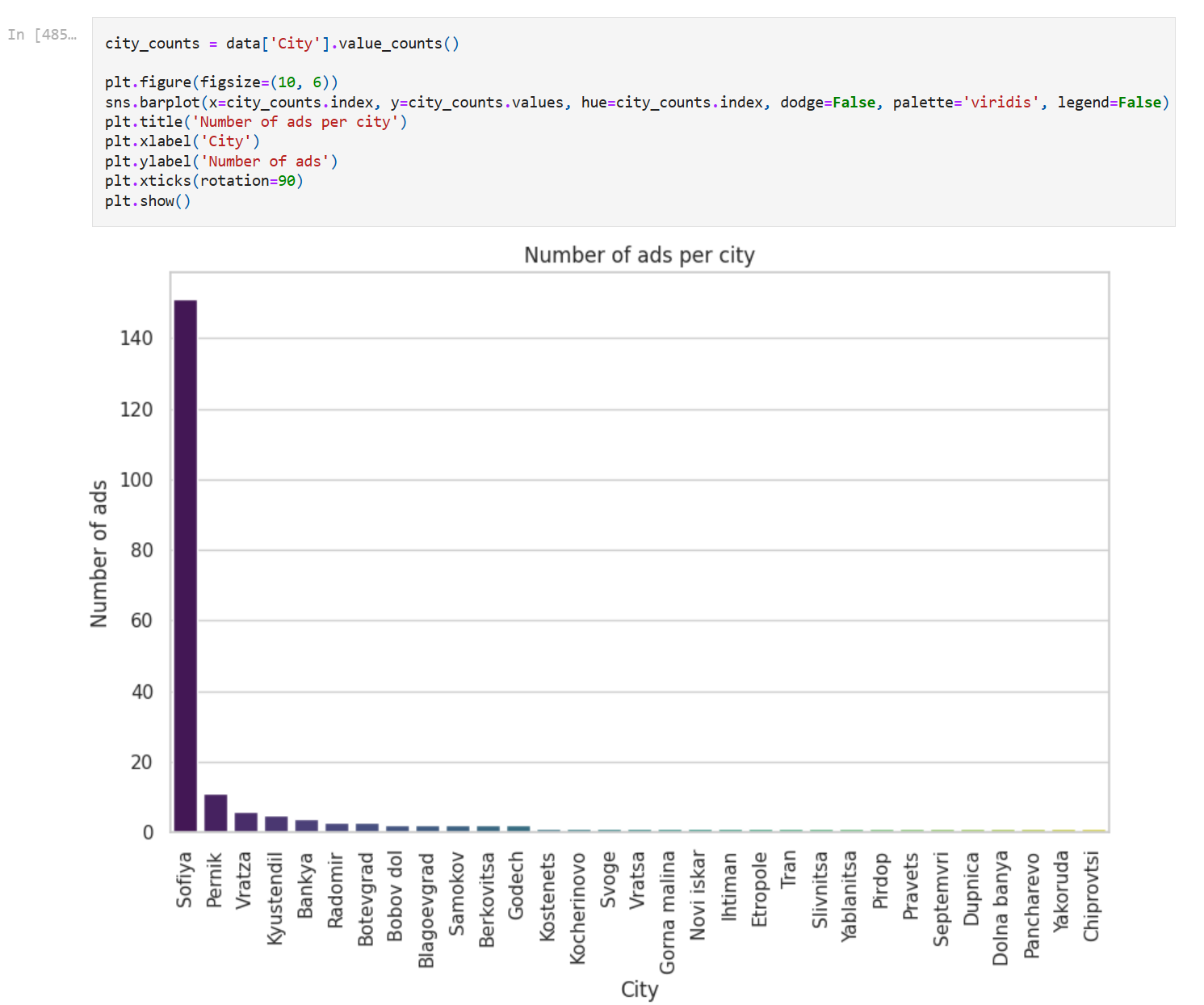
# **5. Представяне и визуализация на резултатите**

## **5.1. Регионален ценови анализ**

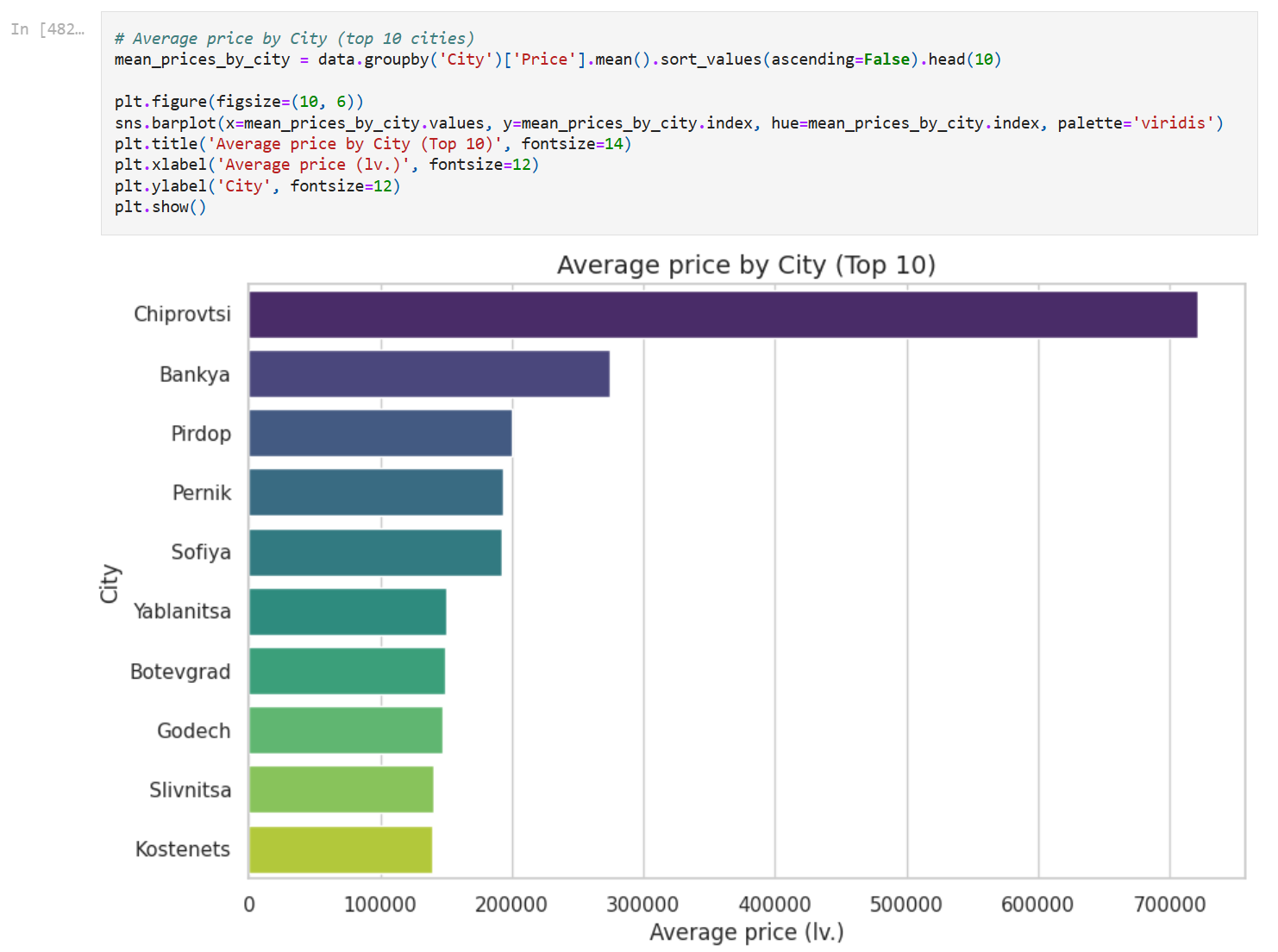


*Фигура 9. Ценово разпределение*

Най-висока е средната цена на обявите в град Чипровци, като това се дължи на единствената обява в размер на *721 721лв*. Изненадваща е позицията на град София, който се намира на едва пето място въпреки огромното превъзходство в броя на обявите с които разполага (*фигура 10)*. Град като Перник, например, постига по-добър резултат от София с над 10 пъти по-малък брой обяви (*Фигура 11*).

**

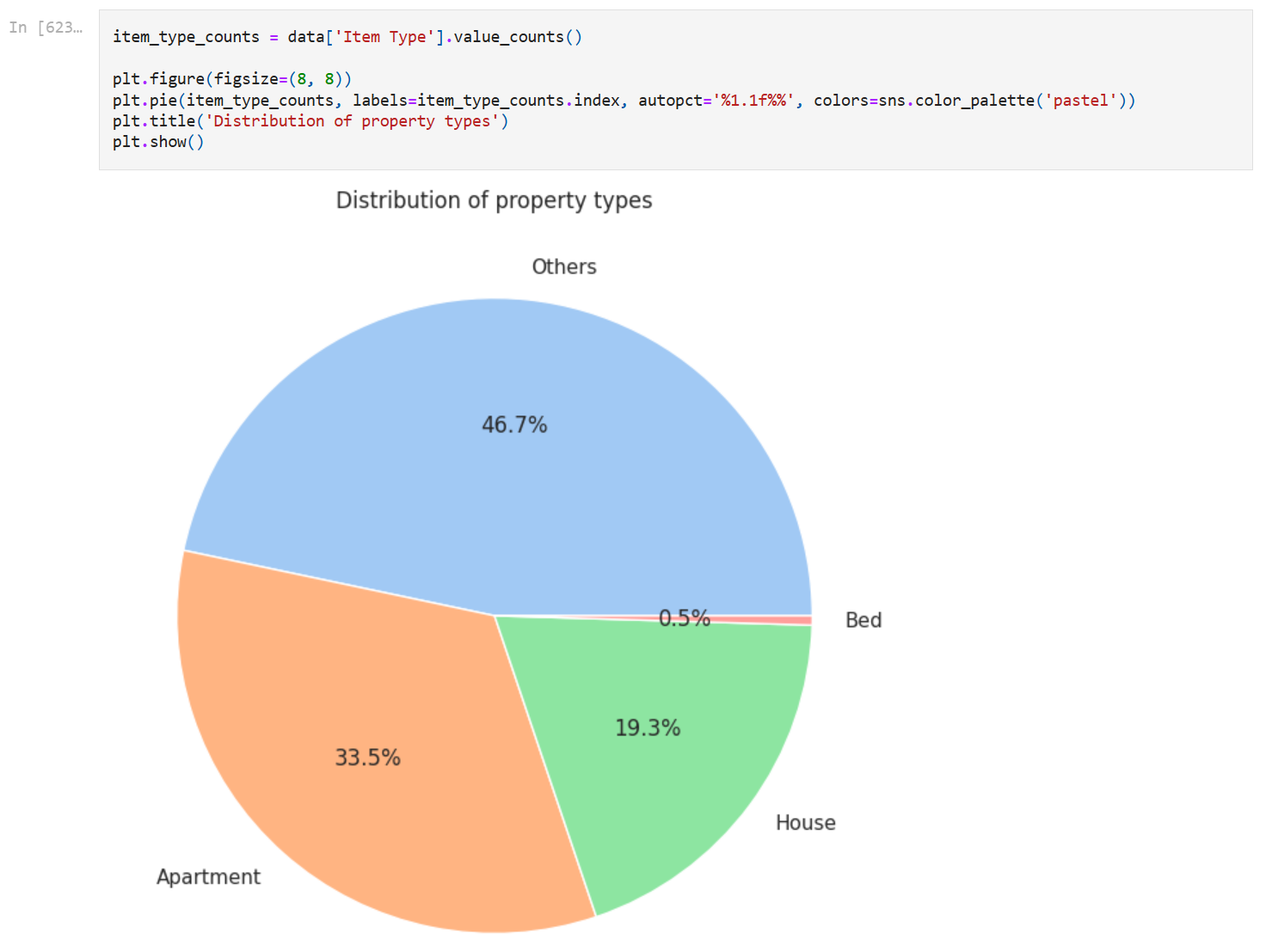
*Фигура 10. Брой обяви за град*

**

*Фигура 11. Средна цена на актив за град*

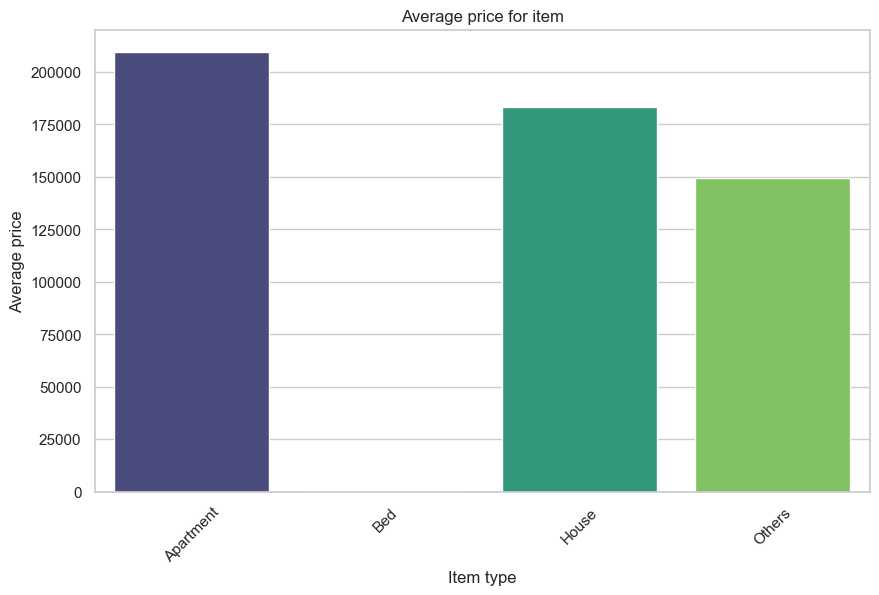
## **5.2 Анализ по типове**

С цел правилната обработка на събраната информация и определянето на типа на обявите, можем да използваме колоната *Item Type*, която е създадена за тази цел (*Фигура 3*). След като начертаем графиката (*Фигура 12*) виждаме, че след некласифицираните “други” в една от 3-те големи кофи обяви, на второ място се подреждат обявите за апартаменти, следвани от тези за къщи и накрая - тези за легла. Вземайки предвид сумата на броя обяви от тип апартамент и тип къща можем да заключим, че най-често срещани са обявите за продажба на жилища, които образуват *52.8%* от всички обяви.

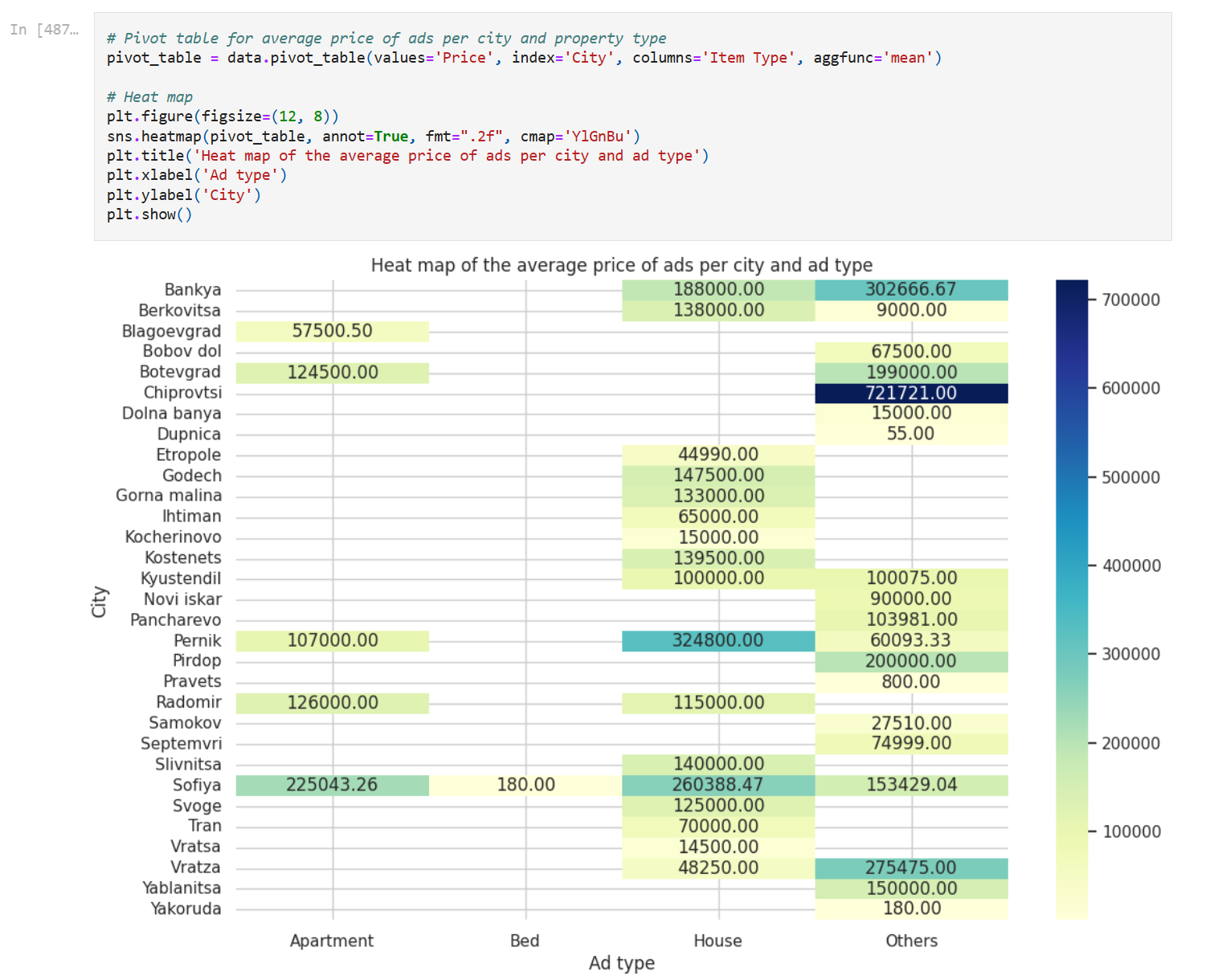
**

*Фигура 12. Разпределение спрямо типа актив*

Чрез помощта на т.н “топлинна карта” и можем да проследим най-скъпите активи в градовете и съответно към кой от 4-те главни типа спадат. (*Фигура 13*). Можем да изведем и графика за средната цена на тип обява. (*Фигура 14*) Като при апартаментите тя е най-висока, следвана от къщите, други и леглата.



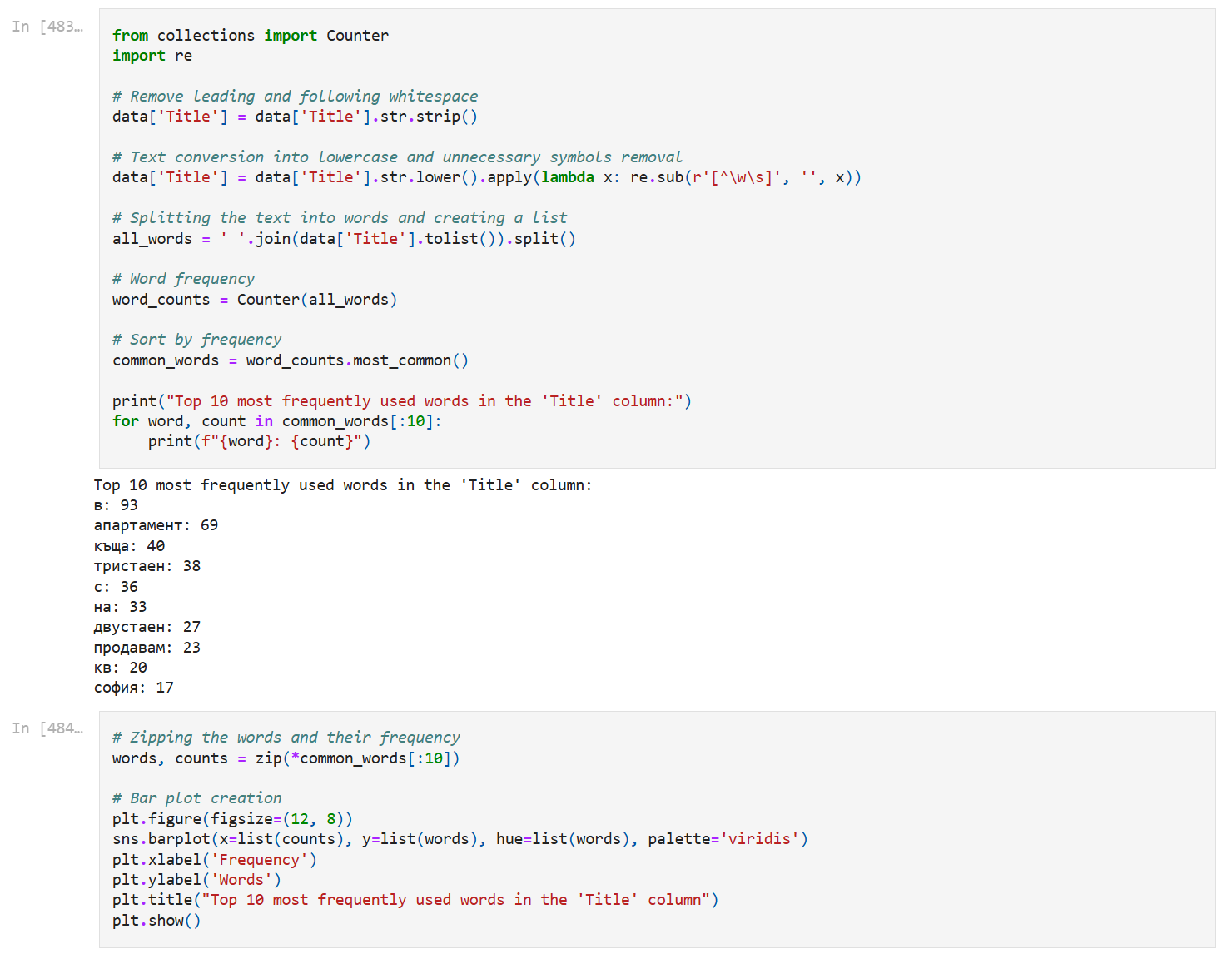
*Фигура 14. Разпределение спрямо цена и тип обява*



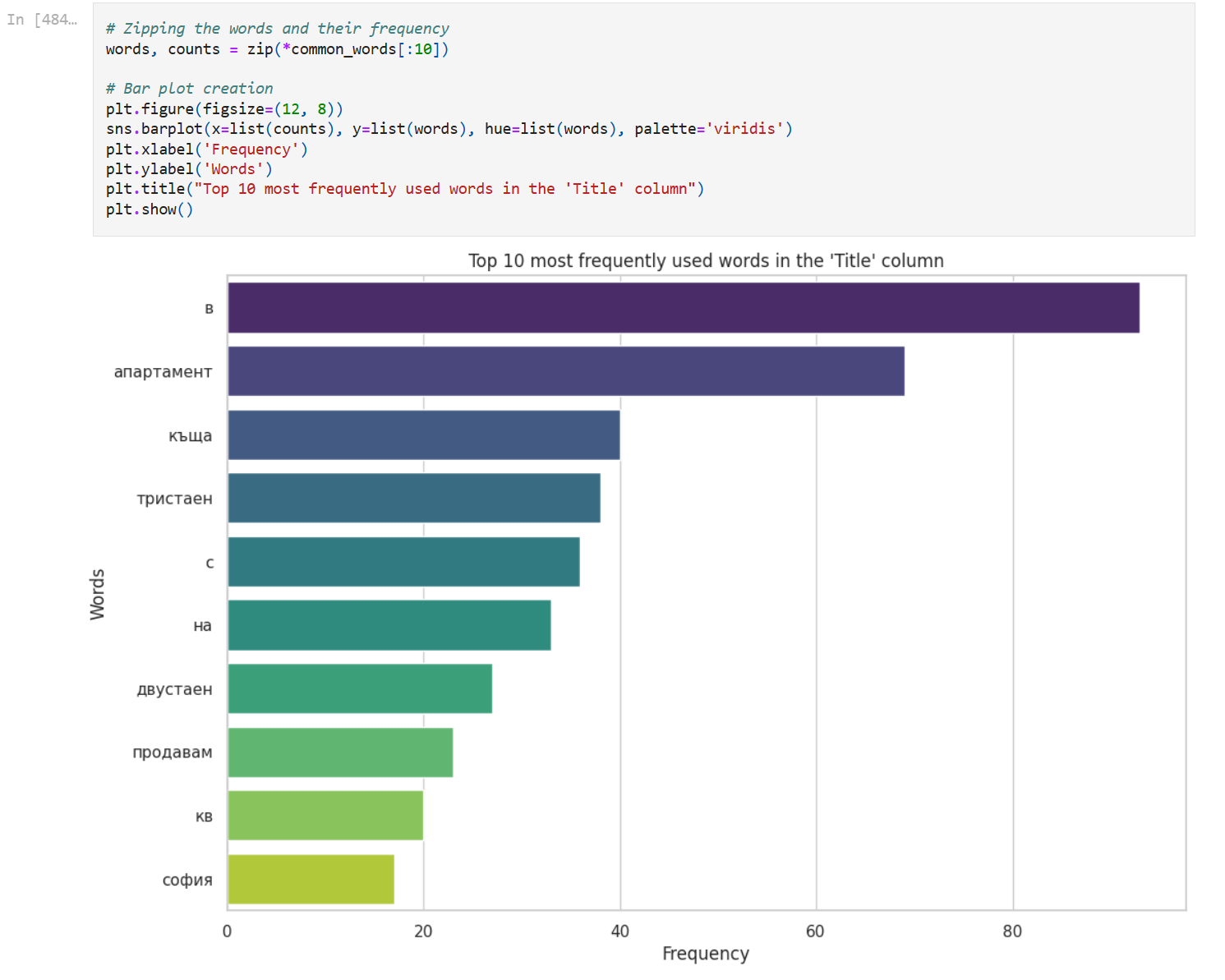
*Фигура 13. Средна цена спрямо населено място и тип на актива*

## **5.3 Синтактичен анализ**

След допълнително почистване на колоната *Title*, можем да създадем списък с всички думи, който да ни даде допълнителна информация за това кои са най-често срещаните (*Фигури 15, 16*). Очаквано предлогът “в” е на първо място, като се среща на цели *93* места. Изненадващо обаче е присъствието на думите “апартамент”, “къща” и “тристаен” съответно на второ, трето и четвърто място. Те разделят “в” от останалите два предлога “с” и “на”, които са на пета и шеста позиция.

**

*Фигура 15. Десетте най-често срещани думи*



*Фигура 16. Графично представяне на десетте най-често срещани думи*

# **6. Заключение**

Анализът на събраните данни от Фейсбук продажби разкрива, че поради ограничения обем и качество на данни, извличането на задълбочени изводи е трудоемка задача. Липсата на ключова информация като детайлни категории за продукти, демографски данни за купувачи или точни времеви периоди ограничава възможността за анализ на тенденции и потребителско поведение.

Въпреки това, наличните данни позволяват идентифицирането на основни модели, като най-продавани категории и средни ценови нива. За по-задълбочен анализ се препоръчва обогатяване на данните с допълнителна информация и повече на брой записи.

# 

# **7. Сорс код**

<https://github.com/savina01/FB_Marketplace_Analysis/>

# **8. Информационни източници**

<https://www.webfx.com/blog/social-media/social-media-analysis/>