

УДК: 004.8

EDN: ELAMMI

DOI: <https://doi.org/10.47813/2782-5280-2024-3-2-0212-0218>



## Разработка модели K-Means для выявления наиболее выгодных предложений на рынке недвижимости Москвы

М. А. Зуев, В. М. Шibaев, К. С. Баланев

*ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия*

**Аннотация.** В статье рассматривается применение модели кластеризации K-Means для анализа рынка недвижимости Москвы. Основное внимание уделяется сегментации рынка с целью выявления наиболее выгодных предложений. Используемые данные включают параметры стоимости, площади, близости к метро, год постройки и другие характеристики объектов недвижимости. Метод "локтя" был применен для определения оптимального числа кластеров, которое впоследствии было увеличено до восьми для более точного сегментирования. Полученные результаты показали, что кластер 0 представляет собой наиболее доступные и выгодные предложения. Модель K-Means, разработанная в ходе исследования, может быть использована покупателями для оптимизации процесса выбора жилья, снижая временные и финансовые затраты.

**Ключевые слова:** кластеризация данных, K-Means, анализ недвижимости, оптимизация выбора, машинное обучение, рынок недвижимости Москвы.

**Для цитирования:** Зуев, М. А., Шibaев, В. М., & Баланев, К. С. (2024). Разработка модели K-Means для выявления наиболее выгодных предложений на рынке недвижимости Москвы. Информатика. Экономика. Управление - Informatics. Economics. Management, 3(2), 0212–0218. <https://doi.org/10.47813/2782-5280-2024-3-2-0212-0218>

---

## Development of the K-Means model to identify the most profitable offers on the Moscow real estate market

M. A. Zuev, V. M. Shibaev, K. S. Balanev

*National Research University «Moscow Power Engineering Institute», Moscow, Russia*

**Abstract.** The article discusses the application of the K-Means clustering model to analyze the Moscow real estate market. The main focus is on market segmentation in order to identify the most profitable offers. The data used includes parameters of cost, area, proximity to the subway, year of construction and other characteristics of real estate. The elbow method was used to determine the optimal number of clusters, which was subsequently increased to eight for more accurate segmentation. The results showed that cluster 0 represents the most affordable and profitable offers. The K-Means model developed during the study can be used by buyers to optimize the housing selection process, reducing time and financial costs.

**Keywords:** data clustering, K-Means, real estate analysis, selection optimization, machine learning, Moscow real estate market.

**For citation:** Zuev, M. A., Shibaev, V. M., & Balanov, K. S. (2024). Development of the K-Means model to identify the most profitable offers on the Moscow real estate market. Informatics. Economics. Management, 3(2), 0212–0218. <https://doi.org/10.47813/2782-5280-2024-3-2-0212-0218>

## ВВЕДЕНИЕ

Рынок недвижимости Москвы является одним из наиболее динамично развивающихся и конкурентных в России. Высокая стоимость жилья, разнообразие предложений и быстрое изменение цен требуют от покупателей тщательного анализа и выбора наиболее выгодных предложений. В этих условиях необходимо использовать современные методы анализа данных, которые могут учитывать множество факторов и обеспечивать точное сегментирование рынка [1].

Традиционные методы анализа рынка недвижимости, такие как простая статистика или экспертные оценки, часто оказываются недостаточно эффективными при большом объеме данных и множестве переменных факторов. Применение методов машинного обучения, таких как кластеризация, становится особенно актуальным [2-4]. Кластеризация позволяет разделить объекты недвижимости на группы с похожими характеристиками, что упрощает анализ и принятие решений для покупателей.

## МАТЕРИАЛЫ И МЕТОДЫ

Кластеризация является одной из ключевых ветвей машинного обучения без учителя. Она позволяет автоматически группировать объекты, основываясь на их сходстве, даже если у них нет явной классифицирующей характеристики. Алгоритм K-Means широко используется благодаря своей простоте и эффективности [5].

Алгоритм K-Means работает по следующему алгоритму [6]:

1. Инициализация центроидов: сначала случайным образом выбираются начальные центроиды кластеров.
2. Назначение объектов кластерам: для каждого объекта рассчитывается евклидово расстояние до каждого из центроидов, и объект присваивается к ближайшему центроиду.
3. Обновление центроидов: центроиды пересчитываются, основываясь на среднем значении характеристик объектов, попавших в данный кластер.

4. Повторение: эти шаги повторяются до тех пор, пока центры не перестанут изменяться или не будет достигнуто заданное количество итераций.

Основное преимущество алгоритма K-Means заключается в его способности быстро и эффективно группировать большие объемы данных, что делает его идеальным инструментом для анализа сложных и многомерных наборов данных. В результате работы алгоритма объекты, близкие по своим характеристикам, оказываются в одном кластере, а объекты с различными характеристиками — в разных [7].

### **Выбор набора данных**

Для исследования был использован набор данных, содержащий информацию о недвижимости Москвы. В нем представлены следующие параметры: цена, время до метро, административный округ, общая площадь, жилая площадь, этаж, количество этажей в здании, год постройки, наличие статуса новостройки, классификация как апартаменты, высота потолков и количество комнат.

### **Предварительная обработка данных**

Для корректной работы модели K-Means были предприняты следующие шаги [8]:

1. Заполнение пропущенных значений: пропущенные значения были заполнены медианными значениями, что позволило минимизировать влияние отсутствующих данных на модель.

2. Преобразование категориальных данных: категориальные данные (например, административный округ) были преобразованы в числовые с использованием метода LabelEncoder. Это позволило учитывать различия между районами Москвы при кластеризации.

### **Определение оптимального числа кластеров**

Оптимальное количество кластеров было определено с помощью метода "локтя", который заключается в анализе инерции модели K-Means при разном числе кластеров и выборе числа, при котором происходит значительное уменьшение инерции [9]. На рис. 1 видно, что оптимальное количество кластеров равно 4. Однако, для более точной сегментации рынка недвижимости и учета большего количества характеристик, было принято решение увеличить количество кластеров до 8. Это позволило выделить более

специфические группы объектов, что улучшило интерпретацию результатов и выявление наиболее выгодных предложений.

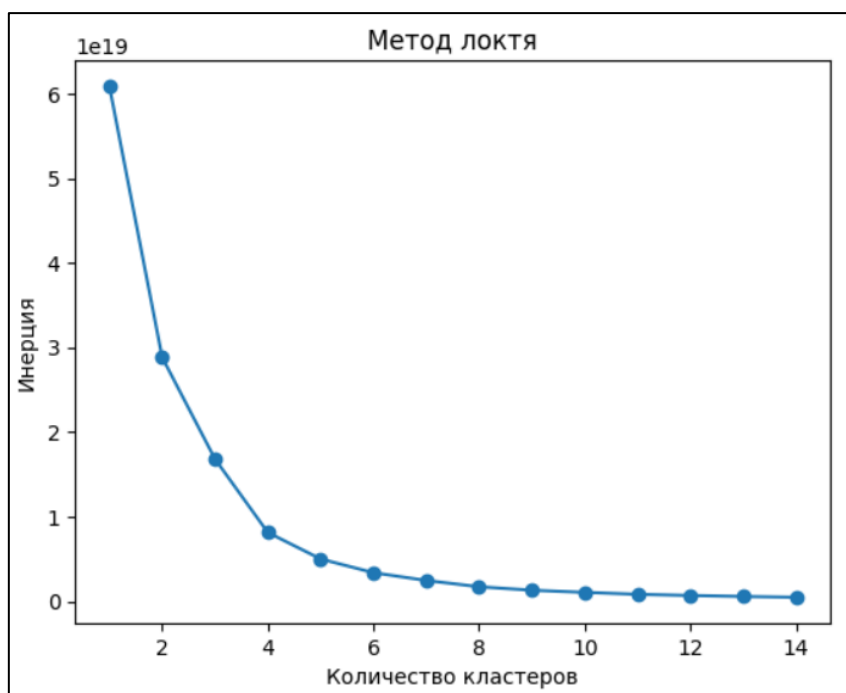


Рисунок 1. Применение метода локтя для набора данных.

Figure 1. Application of the elbow method to a data set.

## РЕЗУЛЬТАТЫ

На основании анализа средних значений параметров для каждого кластера, выделим кластер, который представляет собой наиболее выгодные предложения на рынке недвижимости Москвы.

Кластер 0 характеризуется следующими параметрами: средняя цена объектов составляет около 20,319,055 рублей. Среднее время до ближайшей станции метро составляет 13.02 минут, что является комфортным для большинства покупателей. Объекты недвижимости в этом кластере расположены преимущественно в южном административном округе, что указывает на доступные районы Москвы. Средняя общая площадь объектов в этом кластере составляет 56 кв.м, с жилой площадью 33 кв.м. Средний этаж, на котором расположены эти объекты, составляет 9, а среднее количество этажей в здании – 18. Большинство объектов в данном кластере были построены около 2003 года. Примерно 27.19% объектов являются новостройками. Также 13.17% объектов классифицируются как апартаменты. Средняя высота потолков в этих объектах составляет 3 м. Среднее количество комнат – 2. Таким образом, кластер 0 представляет

собой наиболее выгодные и доступные предложения на рынке недвижимости Москвы, сочетая в себе комфортные условия проживания и оптимальную стоимость.

cluster	price	min_to_metro	region_of_moscow	total_area
0	20,319,055	13.020061	4.524192	55.738995
1	198,544,946	18.424623	1.025126	167.736281
2	3,737,636,000	10.000000	0.000000	530.000000
3	481,139,633	7.646154	0.246154	241.860308
4	118,899,765	8.933333	1.910569	134.626439
5	315,637,948	8.083916	0.741259	202.741259
6	741,009,153	8.222222	0.000000	310.685556
7	57,097,160	14.397742	4.005312	100.058088

cluster	living_area	floor	number_of_floors	construction_year
0	33.470309	9.387302	18.437102	2003.104319
1	60.778894	11.773869	17.597990	2012.140704
2	183.000000	6.000000	7.000000	2003.000000
3	71.541538	11.738462	14.784615	2014.076923
4	58.736423	11.913821	19.988618	2007.338211
5	76.776923	8.783217	12.776224	2011.671329
6	73.505556	6.055556	7.055556	2019.000000
7	50.564807	12.918991	23.930279	2010.918991

cluster	is_new	is_apartments	ceiling_height	number_of_rooms
0	0.271890	0.131697	2.933219	2.031390
1	0.150754	0.195980	3.215201	3.494975
2	0.000000	0.000000	3.000000	4.000000
3	0.153846	0.230769	3.289846	3.661538
4	0.186992	0.232520	3.186699	3.196748
5	0.146853	0.209790	3.353846	3.580420
6	0.277778	0.055556	3.666667	3.777778
7	0.328685	0.174635	3.119900	3.011952

Рисунок 2. Средние значения полученных кластеров.

Figure 2. Mean values of the obtained clusters.

## ЗАКЛЮЧЕНИЕ

В результате исследования была разработана модель на базе алгоритма K-Means, позволяющая выявлять наиболее выгодные предложения на рынке недвижимости Москвы. Модель предоставляет покупателям ценный инструмент для оптимизации процесса выбора недвижимости, что позволяет существенно сократить время и затраты на поиск оптимального варианта. Основная сложность в данной работе заключается в интерпретации полученных кластеров.

## СПИСОК ЛИТЕРАТУРЫ

[1] Shi D., Guan J., Zurada J., Levitan A.S. An Innovative Clustering Approach to Market Segmentation for Improved Price Prediction. Journal of International Technology and

Information Management. 2015; 24(1): 2. <https://doi.org/10.58729/1941-6679.1033>

[2] Гужова О.А. Регулирование рынка жилья с помощью кластер-анализа. Научный журнал НИУ ИТМО. Серия «Экономика и экологический менеджмент». 2015; 3: 201-209.

[3] Skovajsa Š. Review of clustering methods used in data-driven housing market segmentation. Real Estate Management and Valuation. 2023; 31(3): 67-74. <https://doi.org/10.2478/remav-2023-0022>

[4] Крапивина А.А. Факторный и кластерный анализ рынка вторичного жилья. Контентус. 2016; 8(49): 29-35.

[5] Дейтел П., Дейтел Х. Python: искусственный интеллект, большие данные и облачные вычисления. М.: Альпина Пабlishер; 2022. 864.

[6] Маккинни Уэс. Python и анализ данных. М.: ДМК Пресс; 2017. 512.

[7] Гитис Л. Х. Статистическая классификация и кластерный анализ. М.: Наука; 2005. 480.

[8] Педрегоса Ф., Варуко Г. Scikit-learn: машинное обучение на Python. Журнал исследований машинного обучения. 2011; 12: 2825-2830.

[9] Баюк И.А., Березин В.В., Иванюк В.И. Практическое применение методов кластеризации, классификации и аппроксимации. М.: Прометей; 2020. 350.

## REFERENCES

[1] Shi D., Guan J., Zurada J., Levitan A.S. An Innovative Clustering Approach to Market Segmentation for Improved Price Prediction. Journal of International Technology and Information Management. 2015; 24(1): 2. <https://doi.org/10.58729/1941-6679.1033>

[2] Guzhova O.A. Regulirovanie rynka zhil'ya s pomoshch'yu klaster-analiza. Nauchnyj zhurnal NIU ITMO. Seriya «Ekonomika i ekologicheskij menedzhment». 2015; 3: 201-209. (in Russian)

[3] Skovajsa Š. Review of clustering methods used in data-driven housing market segmentation. Real Estate Management and Valuation. 2023; 31(3): 67-74. <https://doi.org/10.2478/remav-2023-0022>

[4] Krapivina A.A. Faktornyj i klasternyj analiz rynka vtorichnogo zhil'ya. Kontentus. 2016; 8(49): 29-35. (in Russian)

[5] Dejtel P., Dejtel H. Python: iskusstvennyj intellekt, bol'shie dannye i oblachnye vychisleniya. M.: Al'pina Pablsher; 2022. 864. (in Russian)

- [6] Makinni Ues. Python i analiz dannyh. M.: DMK Press; 2017. 512. (in Russian)
- [7] Gitis L. H. Statisticheskaya klassifikaciya i klasternyj analiz. M.: Nauka; 2005. 480. (in Russian)
- [8] Pedregosa F., Varuko G. Scikit-learn: mashinnoe obuchenie na Python. Zhurnal issledovanij mashinnogo obucheniya. 2011; 12: 2825-2830. (in Russian)
- [9] Bayuk I.A., Berezin V.V., Ivanyuk V.I. Prakticheskoe primenenie metodov klasterizacii, klassifikacii i approksimacii. M.: Prometej; 2020. 350. (in Russian)

#### ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

**Зуев Максим Алексеевич**, студент, кафедра «БИТ», инженерно-экономический институт, направление «Прикладная информатика», ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

**Maxim Zuev**, student, Department of "BIT", Institute of Engineering and Economics, direction "Applied Informatics", National Research University «Moscow Power Engineering Institute», Moscow, Russia

**Шibaев Владимир Михайлович**, студент, кафедра «БИТ», инженерно-экономический институт, направление «Прикладная информатика», ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

**Vladimir Shibaev**, student, Department of "BIT", Institute of Engineering and Economics, direction "Applied Informatics", National Research University «Moscow Power Engineering Institute», Moscow, Russia

**Баланев Кирилл Сергеевич**, ассистент, ФГБОУ ВО «НИУ «МЭИ», г. Москва, Россия

**Kirill Balanov**, assistant Professor, National Research University «Moscow Power Engineering Institute», Moscow, Russia

*Статья поступила в редакцию 14.06.2024; одобрена после рецензирования 27.06.2024; принята к публикации 28.06.2024.*

*The article was submitted 14.06.2024; approved after reviewing 27.06.2024; accepted for publication 28.06.2024.*