**Министерство Образования и Исследований Молдовы**

**Технический Университет Молдовы**

**Факультет Вычислительной техники, Информатики и Микроэлектроники**

**Департамент Информатики и Системной Инженерии**

**Предмет:** Analiza Datelor

**Научная статья**

**На тему: «Анализ рынка недвижимости в Италии»**

**Выполнил:** Grama Andrei

**Группа:** MI-213

**Кишинёв 2023**

**Введение**

**1.1 Актуальность темы**

Рынок недвижимости занимает весомую позицию в экономической структуре любой страны, являясь важным индикатором её экономического благосостояния и стабильности. В Италии, стране с богатейшей историей и культурным наследием, которое привлекает миллионы туристов ежегодно, рынок недвижимости приобретает особую актуальность. Он становится зеркалом социально-экономических трансформаций, отражая изменения в предложении и спросе, что делает его анализ значимым не только для инвесторов и застройщиков, но и для государственного сектора, способствуя разработке эффективной градостроительной политики и регулированию рынка. Частные лица, заинтересованные в покупке или аренде жилья, также зависят от тенденций и колебаний этого рынка, что делает его анализ актуальным для широкого круга заинтересованных сторон.

**1.2 Цель и задачи**

Основная цель данной работы заключается в комплексном исследовании рынка недвижимости Италии. Особое внимание уделяется анализу ценовых категорий и географическому размещению объектов недвижимости. Это позволит выявить основные тенденции и определить факторы, оказывающие влияние на формирование цен, предоставляя таким образом ценную информацию потенциальным инвесторам и покупателям. В рамках исследования будут выполнены следующие задачи:

- Анализ корреляции между такими параметрами, как площадь жилья, его местоположение и стоимость, для определения степени их взаимного влияния.

- Применение методов множественной регрессии для количественной оценки влияния разнообразных факторов на ценообразование.

- Использование кластерного анализа для идентификации групп регионов с сходными особенностями рынка недвижимости.

Поставленные задачи позволят проверить гипотезы о наличии значимой корреляции между размером жилья и его стоимостью, оценить влияние локализации объектов недвижимости на их ценность, особенно в туристических зонах.

**Методология**

**2.1 Описание набора данных**

Исследование основано на использовании данных с одного из ведущих национальных веб-сайтов по недвижимости в Италии (Immobiliare.it), который предоставляет обзор факторов, влияющих на рынок недвижимости. Анализ включает в себя данные, собранные за 2023 год, предоставляющие широкую картину рыночных тенденций и цен на жилье по всей Италии. В исследование включены данные от всех основных региональных центров, что обеспечивает представительную выборку объектов недвижимости на рынке. Собранные данные дают возможность провести многоуровневый анализ динамики и тенденций рынка. Текущий анализ направлен на изучение изменений на рынке недвижимости Италии за 2023 год, что дает уникальное понимание текущих рыночных условий и потенциальных трендов.

**2.2 Переменные**

В рамках исследования рынка недвижимости в Италии за 2023 год были использованы различные переменные и методы измерения для получения полной и точной картины рынка. Основными переменными в данном исследовании являются цена недвижимости в евро (Price(€)), количество комнат (Rooms), площадь в квадратных метрах (mq) (Figura 1), тип объекта недвижимости (Type), описание объекта (Desc), город расположения (City), район города (Zone) и улица (Street). Цена недвижимости измеряется в евро и представляет собой один из ключевых показателей, отражающих стоимость объекта на рынке. Количество комнат и площадь в квадратных метрах используются для определения размера и вместительности объекта, что также оказывает значительное влияние на его стоимость. Тип объекта недвижимости классифицируется по определенным критериям, включая апартаменты, дома, виллы и другие категории. Описание объекта содержит дополнительную информацию, такую как состояние недвижимости, наличие удобств и другие характеристики. Город и район расположения используются для анализа географического распределения и рыночных тенденций в различных частях Италии. Эти переменные в совокупности позволяют провести глубокий и многоаспектный анализ рынка недвижимости, выявляя ключевые факторы, влияющие на цены и предпочтения покупателей.

Также была добавлена переменная `Viewport`, которая в анализе рынка недвижимости используется для географического позиционирования и визуализации объектов на карте. Она не является прямым показателем ценовых тенденций, но обеспечивает важный контекст, позволяя сравнивать различные рынки недвижимости и оценивать влияние местоположения на стоимость. `Viewport` может служить основой для стратегического планирования, инвестиционных решений и понимания местных рыночных условий.

**2.3 Результативные переменные**

Основными результативными переменными в вашем исследовании рынка недвижимости в Италии за 2023 год являются изменения в ценах на недвижимость, динамика спроса и предпочтений покупателей. Эти переменные отражают, как факторы, такие как местоположение объекта, его размер и количество комнат, влияют на рыночную стоимость. Важной частью анализа является оценка того, как различные районы и города Италии демонстрируют уникальные тенденции в ценообразовании. Кроме того, исследование фокусируется на понимании изменений в спросе, учитывая экономические и социальные факторы, что позволяет выявить актуальные тенденции рынка и потенциальные направления его развития. Отсутствие данных о типах недвижимости и дополнительных удобствах компенсируется более детальным изучением географических, структурных и экономических характеристик объектов, что способствует глубокому пониманию рыночной динамики.

**Результаты**

**3.1 Визуализация данных**

Визуализация данных в исследовании рынка недвижимости в Италии представлена серией графиков, которые наглядно демонстрируют связь между различными характеристиками объектов недвижимости и их ценами. Матрица корреляций (Figura 2) выявляет от умеренной до сильной положительной связи между площадью и ценой, а также между количеством комнат и ценой, что указывает на важность этих параметров при оценке стоимости недвижимости. Гистограмма, отражающая количество недвижимости в разных городах (Figura 3), показывает, что распределение объектов недвижимости неравномерно по стране, с концентрацией предложений в крупных городах. Это подчеркивает разнообразие рыночных условий и потенциальную зависимость цены от географического положения. Дополнительно, столбчатая диаграмма, отображающая среднюю стоимость недвижимости по городам (Figura 4), предоставляет представление о том, как цены варьируются в зависимости от местоположения, с наибольшими средними ценами в крупнейших и наиболее экономически развитых центрах. Распределение количества комнат в недвижимости (Figura 5) подчеркивает популярность определенных размеров жилья, что может служить индикатором спроса на рынке. Наконец, диаграмма рассеяния для цены и площади (Figura 6) подтверждает предположение о том, что более крупные объекты недвижимости имеют тенденцию быть дороже, хотя стоит отметить значительный разброс цен, особенно для объектов большей площади, что может указывать на наличие других факторов, влияющих на стоимость кроме размера. Диаграмма рассеяния (Figura 7) иллюстрирует зависимость цен на недвижимость от количества комнат, показывая, что с увеличением числа комнат наблюдается тенденция к увеличению стоимости, хотя этот рост не всегда пропорционален и сопровождается заметным разбросом цен внутри каждой категории комнат. Наличие аномально дешёвых объектов с большим количеством комнат может указывать на уникальные особенности этих объектов или на ошибки в данных.

**Результаты ANOVA**, представленные на изображении (Figura 9), показывают следующее:

* sum\_sq (сумма квадратов): Это мера общей вариации в данных. Для фактора "City" (C(City)) сумма квадратов составляет около 5.35e+14, что указывает на общую вариацию цен, объяснимую различиями между городами. Residual показывает оставшуюся вариацию, которая не объясняется моделью, и составляет около 2.59e+15.
* df (степени свободы): Для C(City) степени свободы равны 15, что предполагает, что в анализе было 16 уникальных городов (15 дополнительных категорий плюс одна базовая категория). Residual имеет 15982 степени свободы, что соответствует количеству наблюдений минус количество городов.
* F (F-статистика): Значение F-статистики равно 219.688612 для городов, что очень велико. Это статистический тест, используемый для сравнения вариаций между группами и внутри групп. Высокое значение F указывает на то, что средние цены на недвижимость значительно различаются между городами.
* PR(>F) (p-значение для F-статистики): Почти равно 0, что говорит о том, что есть очень сильные доказательства того, что средние цены на недвижимость статистически значимо различаются между городами. В стандартных исследованиях p-значение меньше 0.05 обычно считается достаточным для отклонения нулевой гипотезы (о том, что нет различий между группами), так что здесь различия между городами по ценам на недвижимость можно считать статистически подтвержденными.

В общем, эти результаты указывают на то, что местоположение в разных городах является значимым фактором, влияющим на цены на недвижимость, и что эти различия статистически подтверждены.

**3.2 Описание модели**

Модель визуализации рынка недвижимости Италии, представленная на интерактивной карте, демонстрирует средние цены на жилье в различных городах и регионах страны. Используя цветовые кодировки, модель иллюстрирует градацию стоимости: от более низких цен, отображаемых более холодными оттенками, до более высоких цен, представленных теплыми цветами. Это позволяет наглядно оценить, как экономические, культурные и туристические факторы влияют на стоимость недвижимости в разных областях.

Круги, размещенные на карте, не только обозначают географическое положение каждого рынка недвижимости, но и предоставляют информацию о средней цене недвижимости при наведении курсора. Это делает модель полезным инструментом для инвесторов, государственных органов и частных лиц, интересующихся рынком недвижимости.

**3.3 Визуализация выполнения модели**

**3.3.1 Иллюстрация средних цен на карте**

Из анализа карты (Figura 8), иллюстрирующей средние цены на недвижимость по различным городам и регионам Италии, можно сделать несколько ключевых выводов:

Во-первых, заметны значительные региональные различия в стоимости недвижимости, что может отражать экономическое разнообразие страны. Города с высокой туристической привлекательностью и экономические центры, такие как Рим и Милан, обычно демонстрируют более высокие цены, что, вероятно, связано с большим спросом и высокой стоимостью жизни. Это может быть признаком того, что инвестиции в недвижимость в этих областях потенциально могут предложить более высокую доходность из-за их популярности среди туристов и бизнес-сообщества.

Во-вторых, в сельских и прибрежных районах, где развитие инфраструктуры и туристическая привлекательность могут быть ниже, цены на недвижимость, как правило, более доступны. Это предполагает большую доступность жилья для местных жителей и возможности для тех, кто ищет более уединенные и спокойные районы для жизни или инвестирования.

Наконец, на карте наблюдается кластеризация цен, указывающая на четкое разделение рынка на сегменты в зависимости от географического положения и социально-экономических характеристик регионов. Это подчеркивает важность местоположения в оценке недвижимости и может служить важным инструментом для государственных органов и частных лиц при принятии решений о покупке, продаже или инвестировании в недвижимость.

**3.3.2 Кластеризация регионов**

На Figura 10 представлены результаты кластеризации итальянских городов по средним характеристикам недвижимости. Кластеры сформированы на основе средней цены, количества комнат и размера недвижимости (в квадратных метрах). Выводы по кластеризации:

- Города кластера 0 имеют относительно низкие средние цены и размеры недвижимости, но различное количество комнат.

- Города кластера 1, куда входит Рим, характеризуются более высокими средними ценами и размерами недвижимости.

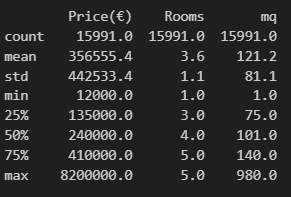
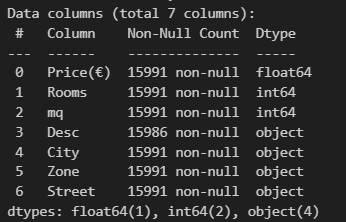
- Города кластера 2 включают Милан, который выделяется высокими средними ценами и меньшим количеством комнат.

- Кластер 3 объединяет города с более низкими средними ценами по сравнению с кластером 2, но с большим количеством комнат, что может указывать на доступность больших жилых площадей по более низким ценам.

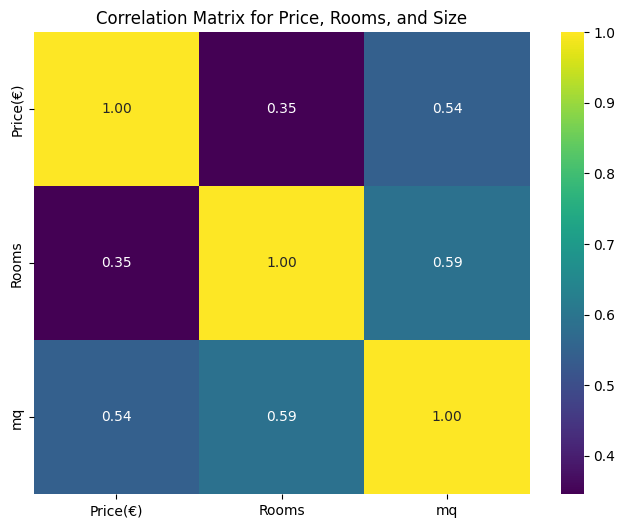
**Вывод**

В моем исследовании рынка недвижимости Италии за 2023 год я провел глубокий анализ, охватывающий ценообразование и географическое распределение объектов. Используя данные с ведущего национального веб-сайта по недвижимости, я сосредоточился на ключевых переменных, таких как цена, размер, тип и местоположение недвижимости. Мой анализ выявил значительные региональные различия в стоимости недвижимости, подчеркивая экономическое и культурное разнообразие страны. Через корреляционный анализ и кластерный анализ, я обнаружил важные связи между ценами и характеристиками недвижимости, выявив различные рыночные сегменты. Визуализация данных, включая интерактивную карту, наглядно демонстрирует эти результаты, делая их доступными для широкого круга заинтересованных сторон. Мое исследование предоставляет ценные сведения о динамике рынка недвижимости Италии, подчеркивая факторы, влияющие на ценообразование.

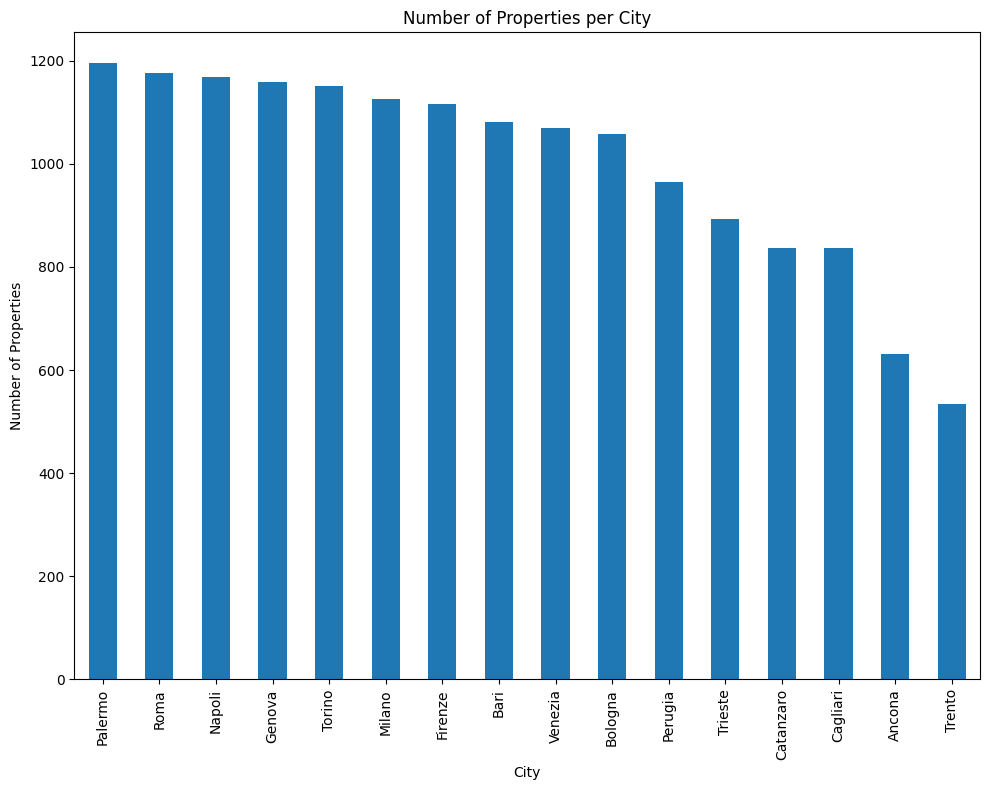
**Изображения**

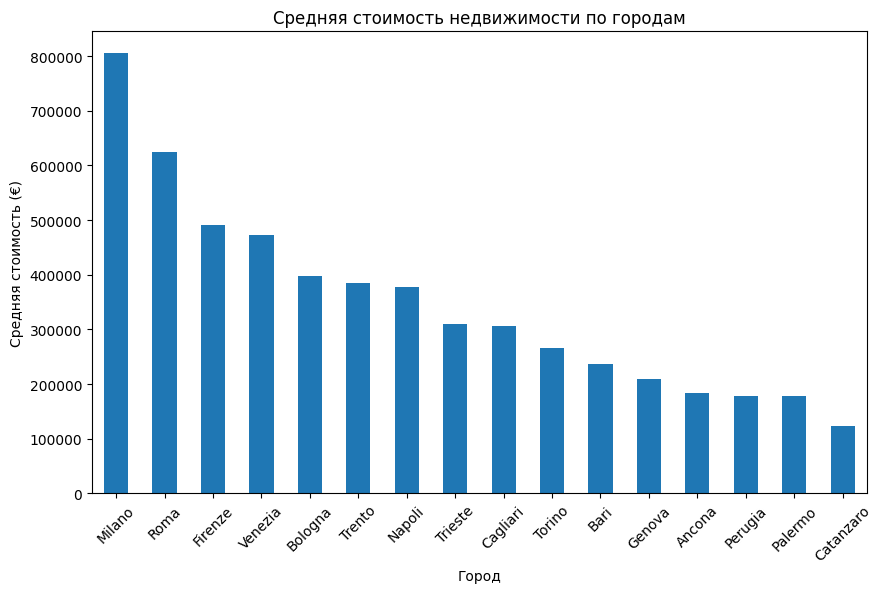
**Figura 1. Статистические данные**

****

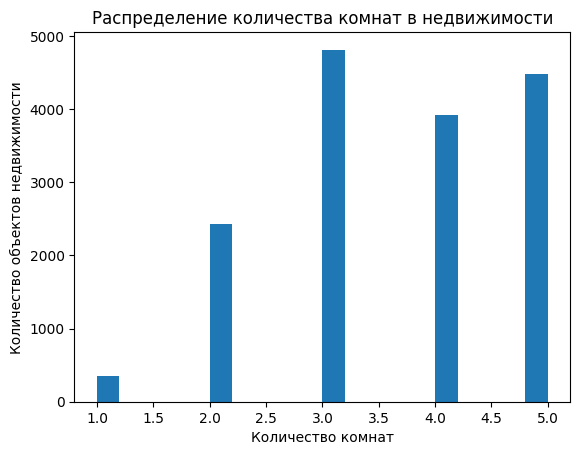
**Figura 2. Корреляционная матрица**

****

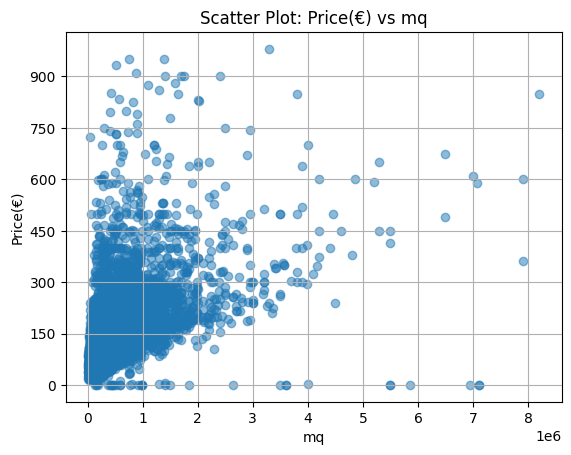
**Figura 3. Гистограмма “количество недвижимости в разных городах”**



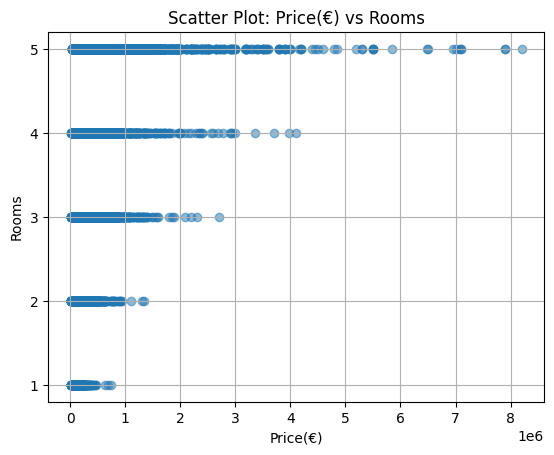
**Figura 4. Гистограмма “Средняя стоимость недвижимости по городам”**



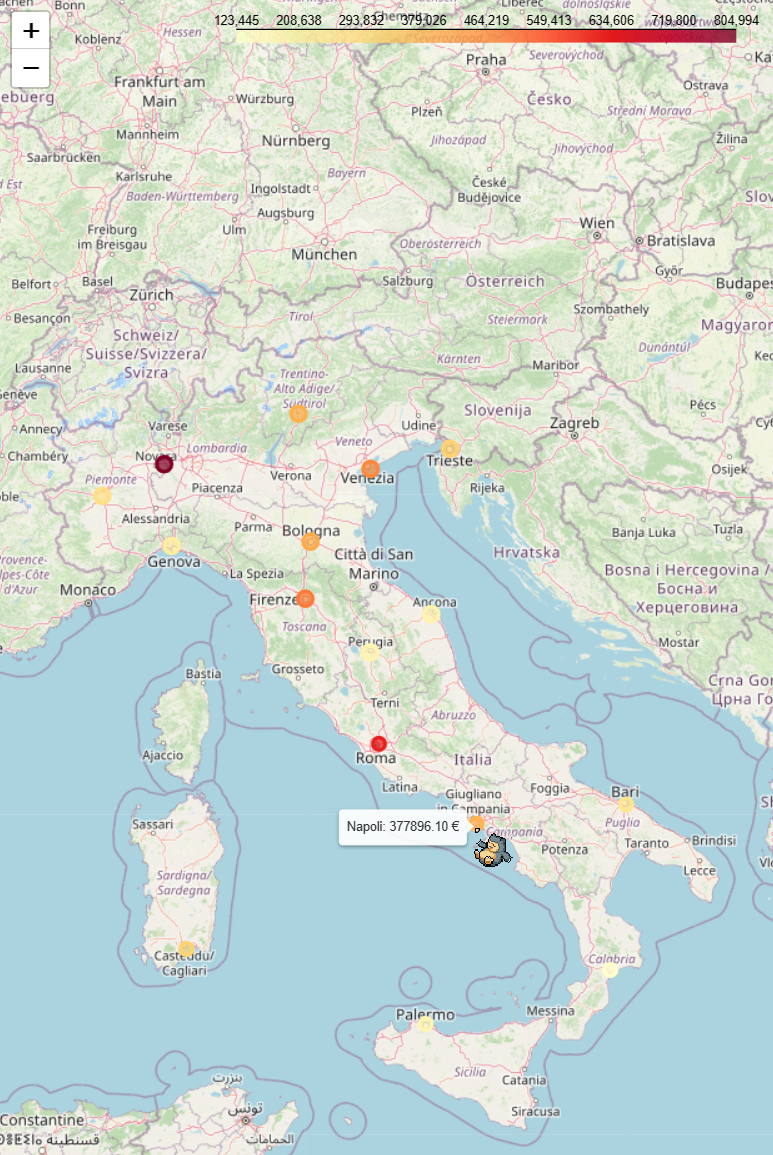
**Figura 5. Гистограмма “Распределение кол-ва комнат”**



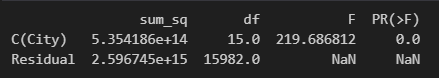
**Figura 6. Диаграмма рассеяния для цены и площади**



**Figura 7. Диаграмма рассеяния цен и количества комнат**



**Figura 8. Иллюстрация средних цен по городам**



**Figura 9. Результаты ANOVA**



**Figura 10. Результаты кластеризации**

**Библиография**

1. Tommaso Ramella, eda\_exploratory\_analysis, Python, [Italy house prices](https://www.kaggle.com/datasets/tommasoramella/italy-house-prices), [eda\_exploratory\_analysis | Kaggle](https://www.kaggle.com/code/tommasoramella/eda-exploratory-analysis#GEOGRAPHICAL-ANALYSIS)
2. [Malamahadevan Mahadevan](https://www.analyticsvidhya.com/blog/author/malamahadevan/), Step-by-Step Exploratory Data Analysis (EDA) using Python, [Step-by-Step Exploratory Data Analysis (EDA) using Python - (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/)
3. Andrea D'Agostino, Exploratory Data Analysis in Python — A Step-by-Step Process, [Exploratory Data Analysis in Python — A Step-by-Step Process | by Andrea D'Agostino | Towards Data Science](https://towardsdatascience.com/exploratory-data-analysis-in-python-a-step-by-step-process-d0dfa6bf94ee)
4. Chao De-Yu, ANOVA Test, with Python, [ANOVA Test, with Python. The Complete Beginner’s Guide to… | by Chao De-Yu | Towards Data Science](https://towardsdatascience.com/anova-test-with-python-cfbf4013328b)
5. ANOVA Test in Python, [ANOVA Test in Python - Javatpoint](https://www.javatpoint.com/anova-test-in-python)
6. [Jason Brownlee](https://machinelearningmastery.com/author/jasonb/), 10 Clustering Algorithms With Python, [10 Clustering Algorithms With Python - MachineLearningMastery.com](https://machinelearningmastery.com/clustering-algorithms-with-python/)
7. Sadrach Pierre, How to Form Clusters in Python: Data Clustering Methods, [How to Form Clusters in Python: Data Clustering Methods | Built In](https://builtin.com/data-science/data-clustering-python)