Мохаммад Каріна

Прогнозування цін на житло в Києві: первинка

1. **Збір даних**

У процесі здійснення дослідження щодо прогнозування цін на житло була здійснена збір та організація даних. Дані для аналізу були взяті з веб-сайту нерухомості [Лун Статистика](https://misto.lun.ua/stat/kyiv) та включають інформацію про середні ціни на житло на первинку в місті Київ протягом періоду з жовтня 2014 року по жовтень 2023 року. Для зручності та подальшого аналізу ці дані були внесені в електронну таблицю формату Excel (розширення .xlsx). Додатково, дані розділені на чотири сегменти: "економ", "комфорт" та "бізнес"", для подальшого розгляду та порівняння цін у різних категоріях житла.

Цей підхід до збору та організації даних забезпечує необхідну основу для подальшого аналізу та створення моделі машинного навчання з метою прогнозування цін на житло в місті Київ. Використання електронної таблиці дозволяє зручно керувати та аналізувати дані, а розподілення на сегменти дозволяє досліджувати ціни в різних категоріях житла, роблячи дослідження більш деталізованим та обґрунтованим.

Отримані дані складаються з трьох аркушів: "економ", "комфорт" та "бізнес", кожен із яких містить 110 стовпчиків і 2 рядки. Інформація на цих аркушах включає в себе ціни на житло за місяць за квадратний метр.

1. **Завантаження та Обробка Даних**

Для початку, дані було завантажено з кожного аркуша Excel за допомогою бібліотеки pandas. Кожен аркуш містив інформацію про ціни на житло за квадратний метр на протязі років. Однак, перед аналізом, була виконана додаткова обробка даних.

Конкретно, стовпець 'Month', що містив дати, було перетворено у формат дати та часу за допомогою функції pd.to\_datetime(). Це дозволило нам працювати з датами як зі спеціальним типом даних і здійснювати аналіз, пов'язаний із часовими рядами.

Після конвертації дат, ми провели аналіз якості даних. Для цього було виявлено пропущені значення, дублікати та викиди.

Після обробки даних ми провели аналіз кожного аркуша окремо. Ми вивели загальну інформацію про дані, включаючи типи даних та кількість рядків, статистичні показники, такі як середнє значення, мінімальне та максимальне значення, що дозволило нам краще зрозуміти розподіл цін на житло.

Крім того, було використано статистичний підхід за допомогою з-оцінок (z-scores) для виявлення викидів. Викиди були визначені як ті дані, для яких з-оцінка перевищує певний поріг (зазвичай, 3).

Изображение выглядит как текст, снимок экрана, документ, Шрифт

Автоматически созданное описание

Аркуш "Econom" містить дані про ціни на житло в місті Київ на первинному ринку протягом декількох років. Загальний розмір аркуша складає 109 рядків та 2 стовпці, де перший стовпець містить інформацію про дати у форматі дати та часу, а другий стовпець відображає ціни на житло за квадратний метр. Аналіз статистичних показників показав, що середнє значення цін становить 20,321.50 у валюті, зі значним стандартним відхиленням, що свідчить про велику варіабельність цін. Дублікатів та викидів не виявлено, що робить цей аркуш готовим для подальшого аналізу та використання у дослідженні цінових динамік на ринку житла.

Изображение выглядит как линия, График, диаграмма, Шрифт

Автоматически созданное описание

Аналіз статистичних показників аркушу "Comfort" показав, що середнє значення цін становить 23,165.42 у валюті, із стандартним відхиленням, що свідчить про високий рівень варіабельності в цінах. На аркуші "Comfort" не було виявлено дублікатів та викидів, що робить ці дані готовими для подальшого аналізу та використання в контексті вивчення цінових тенденцій на ринку житла в Києві.

Изображение выглядит как линия, График, текст, Шрифт

Автоматически созданное описание

Аналіз статистичних показників аркушу "Бізнес" вказує на високий середній рівень цін у цьому сегменті ринку, де середнє значення становить 40,130.84 у валюті, зі значним стандартним відхиленням, що вказує на значну варіабельність в цінах на житло. Аркуш "Бізнес" не містить дублікатів чи викидів, що позитивно впливає на якість даних та їх готовність для подальшого використання в аналізі та прогнозу цін на житло в місті Київ.

Изображение выглядит как линия, График, текст, Шрифт

Автоматически созданное описание

У зв'язку з повномасштабним вторгненням росії в Україну та складними політичними та соціальними обставинами в країні, інформація про ціни на житло за період березень-квітень 2022 року відсутня. Ця відсутність даних є результатом неможливості збирати та систематизувати інформацію в умовах військового стану.

Було використано метод інтерполяції для заповнення пропусків в даних. Цей метод був використаний для забезпечення неперервності даних та підготовки їх до подальшого аналізу. В конкретному випадку, відсутність даних була заповнена за допомогою лінійної інтерполяції за методом, який було виконано в мові програмування Python. Цей підхід був обраний через його ефективність та здатність зберегти неперервність даних.

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Аналіз сезонності

**Функція для візуалізації сезонності (sesonal):** Призначена для візуалізації сезонності у часовому ряді. Ця функція приймає два параметри: **data** (DataFrame або Series) - часовий ряд з датами та значеннями, і **s** (int) - період сезонності у часовому ряді. В результаті виконання функції будується графік, на якому представлена інформація про сезонні коливання у різні роки. Ця функція була застосована для аналізу сезонності в кожній з трьох категорій житла ("Econom," "Comfort," "Business") з вказаним періодом сезонності.

Для кожної категорії житла ("Econom," "Comfort," "Business"), було проведено аналіз сезонності за допомогою графіків, які відображають залежність цін на житло від часу.

Також було використано модель сезонної декомпозиції для аналізу тренду та сезонних складових у даних. Обрано як мультиплікативну, так і адитивну моделі для аналізу сезонності, і результати представлено на графіках в коді.

В результаті було виявлено, що кожен часовий ряд має висхідний тренд та сезонність.

Изображение выглядит как текст, рукописный текст, линия, График

Автоматически созданное описание

1. **Побудова моделі машинного навчання**

* **Вибір методу SARIMA:** В даному аналізі для прогнозування часового ряду цін на житло категорії "Econom" був обраний метод SARIMA, що є абревіатурою для "Seasonal Autoregressive Integrated Moving Average." Цей метод дозволяє враховувати тренди, сезонні зміни та аномалії у часовому ряді. Обрання методу SARIMA базувалося на передбаченні, що в цінах на житло категорії "Econom" існують як сезонні коливання, так і стохастичні зміни, які можуть бути виявлені та передбачені цією моделлю.
* **Навчання моделі SARIMA:**

**Підготовка даних для навчання моделі:** Вибірка для навчання моделі ("train\_‘name of sheet’") була визначена як частина часового ряду з 2014 року до 2022 року. Дані в цій вибірці використовуються для навчання моделі та підбору параметрів SARIMA. Тестова вибірка містить дані за 2023 рік.

**Параметри моделі:** Встановлення параметрів моделі SARIMA, таких як порядок (order) та сезонний порядок (seasonal\_order), що були обрані вручну, зважаючи на статистичну значущість статистичної інформації про підгонку моделі та стаціонарність тренувальних та тестових даних.

**Підгонка моделі:** Побудова та навчання моделі SARIMA на підготовлених даних, використовуючи визначені параметри. В результаті статистична інформація про підгонку моделі виявилась статистично значущою.

* **Аналіз моделі та діагностика:**

**Діагностика моделі:** Після навчання моделі проводиться діагностика, яка включає аналіз залишкових помилок та перевірку поперед покладань моделі.

**Візуалізація діагностики:** Відображення результатів діагностики за допомогою графіків, включаючи графіки розподілу залишкових помилок, автокореляційні графіки, графіки квантиль-квантиль та інші.

* **Прогнозування цін:**

**Отримання прогнозу:** Модель SARIMA, після успішного навчання, використовується для отримання прогнозу цін на житло в майбутньому. Прогноз цін до січня 2026 року. Оскільки ситуація в країні є нестабільною і немає сенсу робити прогнози на великі проміжки.

**Аналіз прогнозу та оцінка якості:** Прогнозовані значення порівнюються з фактичними даними для оцінки точності прогнозу. Виводяться метрики, такі як середня абсолютна похибка (MAD), середньоквадратична похибка (MSE), середня абсолютна відсоткова похибка (MAPE), середня відсоткова похибка (MPE) та стандартна похибка.

* **Візуалізація результатів:**

**Графіки прогнозу та фактичних значень:** Результати прогнозування виводяться на графіку, що відображає фактичні значення і прогнозовані значення, що надає змогу візуально оцінити точність моделі.

1. **Результати прогнозування.**

Изображение выглядит как текст, График, линия, диаграмма

Автоматически созданное описание

Рис. 1. Прогноз цін на нерухомість економ-класу

У 2023 році спостерігається зростання середньомісячної ціни, починаючи з 32,714.92 і закінчуючи 38,567.53 в грудні 2023 року. В 2024 році прогнозується подальше зростання цін, протягом цього року середньомісячна ціна зросте з 39,710.04 в січні до 50,819.44 в жовтні. У 2025 році також передбачається зростання цін, і середньомісячна ціна збільшиться з 51,635.57 в січні до 59,669.41 в грудні. В 2026 році прогнозується подальше зростання цін, протягом року середньомісячна ціна зросте з 60,688.82 в січні до нового рівня.

1. **Загальна динаміка цін**: Прогнози показують загальну тенденцію до зростання цін на нерухомість економ-класу протягом періоду від січня 2023 року до січня 2026 року. Це може вказувати на позитивний розвиток ринку нерухомості в цьому сегменті.
2. **Сезонні коливання**: Прогнози враховують сезонні коливання цін на різні місяці. Це може свідчити про вплив сезонних факторів на ціни в цьому сегменті ринку.

Изображение выглядит как текст, График, линия, диаграмма

Автоматически созданное описание

Рис. 2. Прогноз цін на нерухомість комфорт-класу

З січня 2023 року по грудень 2023 року спостерігалося значне зменшення середньомісячної ціни, зі значенням 36,627.97 в січні до 36,563.66 в грудні 2023 року. У 2024 році ціни почали зростати, зі значенням 35,734.50 в січні, до 38,071.54 в грудні 2024 року. У 2025 році також спостерігалося збільшення середньомісячної ціни, з 37,036.05 в січні до 37,730.25 в березні, та подальше збільшення до 36,730.25 в січні 2026 року.

1. **Загальна динаміка цін**: Прогнози показують загальну тенденцію до зростання цін на нерухомість комфорт-класу протягом періоду від січня 2023 року до січня 2026 року. Це може вказувати на позитивний розвиток ринку нерухомості в цьому сегменті.
2. **Сезонні коливання**: Прогнози враховують сезонні коливання цін на різні місяці. Це може свідчити про вплив сезонних факторів на ціни в цьому сегменті ринку.

Изображение выглядит как текст, График, линия, диаграмма

Автоматически созданное описание

Рис. 3. Прогноз цін на нерухомість бізнес-класу

З січня 2023 року по грудень 2023 року спостерігалося збільшення середньомісячної ціни, зі значенням 69,469.47 в січні до 72,826.41 в грудні 2023 року.

У 2024 році спостерігається подальше збільшення цін, починаючи з 71,535.51 в січні і досягаючи 76,938.47 в грудні 2024 року.

У 2025 році знову спостерігається збільшення цін, з 75,360.33 в січні до 77,297.65 в грудні 2025 року.

1. **Зростання цін на квартири**: Прогнози показують тенденцію до зростання цін на квартири бізнес-класу протягом розглянутого періоду (від січня 2023 року до січня 2026 року). Це може вказувати на попит на житло в цьому сегменті ринку та можливі позитивні економічні чи соціокультурні фактори.
2. **Сезонність може бути важливою**: Враховуючи сезонні коливання у прогнозах, це може означати, що ціни на квартири бізнес-класу піддаватимуться впливу сезонних факторів. Такі фактори можуть включати сезонну активність у галузі нерухомості або зміни в попиті в різні часи року.
3. **Планування та інвестиції**: Ці прогнози можуть бути корисними для інвесторів та розробників нерухомості, які планують вкладати кошти у квартири бізнес-класу. Вони можуть використовувати цю інформацію для прийняття рішень щодо розвитку нових проектів або купівлі-продажу нерухомості.

Загальний висновок, який можна зробити на основі наданих прогнозів цін на нерухомість в різних класах (бізнес-клас, комфорт-клас та економ-клас), полягає в тому, що усі три класи нерухомості відзначаються загальною тенденцією до зростання цін протягом розглянутого періоду (з січня 2023 року до січня 2026 року). Але ми маємо розуміти, що на прогноз цін можуть впливати різні зовнішні фактори, які потрібно брати до уваги, провести більш детальний аналіз. Але для цього такі дані мають міститись разом.