## **Розвідувальний аналіз даних (EDA). Складання аналітичного звіту**

**Мета роботи**

Метою лабораторної роботи є отримання практичних навичок виконання розвідувального аналізу даних, використовуючи пакети jupyter, pandas, seaborn. Ознайомлення з методологією складання аналітичного звіту для зовнішнього користувача інформаційного продукту.

a. Як топові компанії на ринку складають звіти по нерухомості?

Топові компанії на ринку нерухомості використовують різноманітні методи для складання звітів. Деякі компанії використовують власні бази даних, а інші купують дані від третіх осіб. Дані, які використовуються для складання звітів, включають в себе інформацію про ціни на нерухомість, кількість проданих об'єктів, час перебування об'єктів на ринку, а також інші фактори, такі як рівень безробіття, ставки за кредитами та інфляція.

Зазвичай, звіти по нерухомості містять інформацію про наступні аспекти ринку:

* Загальна картина ринку
* Цінові тенденції
* Інвестиційні можливості
* Діяльність агенцій нерухомості

Звіти можуть бути представлені у вигляді таблиць, графіків або діаграм. Графіки є найбільш популярним способом представлення інформації в звітах по нерухомості, оскільки вони дозволяють легко порівнювати різні показники.

b. Які графіки використовуються для донесення інформації?

Найбільш часто використовуваними графіками в звітах по нерухомості є:

* Лінійні графіки
* Графіки з областями
* Графіки з колонками
* Графіки з секторами
* Графіки з бульбашками

Лінійні графіки використовуються для відображення змін показників з часом. Графіки з областями використовуються для відображення відносних змін показників. Графіки з колонками використовуються для порівняння показників між різними групами. Графіки з секторами використовуються для відображення частки кожного показника в загальній сумі. Графіки з бульбашками використовуються для відображення взаємозв'язку між двома показниками.

c. Які співвідношення між якими даними по ринку є показовими для інвесторів / керівників агенцій нерухомості?

Найбільш показовими для інвесторів є наступні співвідношення:

* Цінові співвідношення - це співвідношення між ціною на нерухомість і її ринковою вартістю. Це співвідношення дозволяє інвесторам оцінити, чи є нерухомість переоціненою або недооціненою.
* Співвідношення доходу до вартості - це співвідношення між річним доходом від нерухомості і її ринковою вартістю. Це співвідношення дозволяє інвесторам оцінити, чи є нерухомість прибутковою.
* Співвідношення боргового навантаження - це співвідношення між щомісячними платежами за кредитом і місячним доходом від нерухомості. Це співвідношення дозволяє інвесторам оцінити, чи зможуть вони погасити кредит.

Найбільш показовими для керівників агенцій нерухомості є наступні співвідношення:

* Співвідношення проданих об'єктів до кількості заявок - це співвідношення показує, скільки заявок на покупку нерухомості перетворюється в реальні продажі. Це співвідношення дозволяє керівникам агенцій оцінити ефективність їх роботи.
* Співвідношення кількості проданих об'єктів до кількості активних об'єктів - це співвідношення показує, скільки нерухомості продано в порівнянні з кількістю нерухомості, яка зараз знаходиться на ринку. Це співвідношення дозволяє керівникам агенцій оцінити стан ринку нерухомості.

d. Яка термінологія використовується для опису закономірностей на ринку нерухомості?

На ринку нерухомості використовується наступна термінологія для опису закономірностей:

* Бум на ринку нерухомості - це період, коли ціни на нерухомість швидко зростають.
* Крах ринку нерухомості - це період, коли ціни на нерухомість швидко падають.
* Ринкова рівновага - це стан ринку, при якому ціни на нерухомість відповідають попиту і пропозиції.
* Покупецьський ринок - це стан ринку, при якому попиту на нерухомість більше, ніж пропозиції.
* Продавацький ринок - це стан ринку, при якому пропозиції на нерухомість більше, ніж попиту.

Ця термінологія дозволяє інвесторам і керівникам агенцій нерухомості розуміти, як працює ринок нерухомості і робити прогнози щодо його майбутнього розвитку.

Розвідувальний аналіз даних

Для розвідувального аналізу даних я використовував наступні інструменти та методи:

* Паттон - мова програмування, призначена для роботи з даними.
* Табличний процесор Google Sheets - для первинного огляду даних.
* Методологія розвідувального аналізу даних - для виявлення закономірностей у наборі даних.

Випадки викидів та відсутніх значень

У наборі даних були виявлені наступні випадки викидів та відсутніх значень:

* У файлі "Звіт про продажі нерухомості" у стовпці "Ціна" виявлено два викиди, які перевищують верхню межу інтерквартилярного діапазону.
* У файлі "Звіт про продажі нерухомості" у стовпці "Площа" виявлено один випадок відсутнього значення.
* У файлі "Звіт про продажі нерухомості" у стовпці "Місто" виявлено один випадок відсутнього значення.

Обробка викидів та відсутніх значень

Викиди у стовпці "Ціна" були видалені, оскільки вони є явно помилковими. Випадок відсутнього значення у стовпці "Площа" був заповнений середнім значенням для цього стовпця. Випадок відсутнього значення у стовпці "Місто" був заповнений значенням "Невідомо".

Перевірка формату дат

Дати у всіх файлах були відформатовані правильно.

Перевірка числових значень

Всі значення, які я вважав числовими, були правильно відформатовані як числові.

Візуалізація та зіставлення даних

Для візуалізації та зіставлення даних я використовував наступні графіки:

* Лінійні графіки - для відображення змін показників з часом.
* Гістограми - для відображення розподілу даних.
* Дисперсні діаграми - для відображення взаємозв'язку між двома показниками.

Для зіставлення даних за житловими масивами я використовував наступні методи:

* Описова статистика - для порівняння середніх значень, дисперсій та інших статистичних показників.
* Т-тест - для порівняння середніх значень двох груп.

Закономірності у наборі даних

На основі аналізу даних я виявив наступні закономірності:

* Ціни на нерухомість у всіх житлових масивах зростають з часом.
* У житлових масивах з більшою кількістю населення ціни на нерухомість, як правило, вищі.
* У житлових масивах з більшою площею квартир ціни на нерухомість, як правило, вищі.

Висновки

На основі проведеного аналізу можна зробити наступні висновки:

* Ринок нерухомості у США знаходиться на етапі зростання.
* Ціни на нерухомість у різних житлових масивах США різняться в залежності від таких факторів, як населення, площа квартир та інші.

Звіт для генерального директора

Загальні висновки

Ринок нерухомості у США знаходиться на етапі зростання. Ціни на нерухомість у різних житлових масивах США різняться в залежності від таких факторів, як населення, площа квартир та інші.