## **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.

**Хід виконання:**

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.

Візуалізація та зіставлення даних

Ціна з часом

A graph of a line graph

Description automatically generated

Цей графік показує, що ціни на нерухомість у США зростають з часом. Зростання цін було найбільшим у 2022 році, коли вони зросли на 15% порівняно з попереднім роком.

Розподіл ціни

A graph of different colored lines

Description automatically generated with medium confidence

Ця гістограма показує, що розподіл цін на нерухомість у США має вигляд нормальної кривої. Більшість квартир продається за ціною від 500 000 до 1 000 000 доларів США.

Взаємозв'язок між ціною та площею

A screenshot of a video game

Description automatically generated

Ця дисперсна діаграма показує, що між ціною та площею нерухомості існує позитивний кореляційний зв'язок. Чим більша площа квартири, тим вища її ціна.

Середні значення ціни за житловими масивами

A graph showing the growth of a homebuilder

Description automatically generated

Ця діаграма показує, що середнє значення ціни на нерухомість у Нью-Йорку становить 1 200 000 доларів США, а в Лос-Анджелесі - 1 000 000 доларів США. Це означає, що ціни на нерухомість у Нью-Йорку вищі, ніж у Лос-Анджелесі.

Т-тест для порівняння середніх значень ціни між двома житловими масивами

A graph of a number of blue rectangular objects

Description automatically generated with medium confidence

Т-тест показав, що середні значення ціни на нерухомість у Нью-Йорку та Лос-Анджелесі істотно відрізняються. Це означає, що існує значний статистичний доказ того, що ціни на нерухомість у Нью-Йорку вищі, ніж у Лос-Анджелесі.

Рекомендації

* Інвестори можуть розглядати можливість інвестування в нерухомість США, оскільки ринок знаходиться на етапі зростання. Однак важливо враховувати фактори, які можуть впливати на ціни на нерухомість у майбутньому, такі як рівень безробіття, ставки за кредитами та інфляція.
* Керівники агенцій нерухомості можуть використовувати результати аналізу для прогнозування попиту на нерухомість і розробки стратегій продажів. Зокрема, вони можуть використовувати карту США з позначенням житлових масивів, щоб визначити, в яких регіонах США є найбільший потенціал для зростання цін.

import pandas as pd

import matplotlib.pyplot as plt

# Завантаження даних

df\_sales = pd.read\_csv("https://docs.google.com/spreadsheets/d/1MW-zi5SU4VHChqn1heGOi3zQx77dG17\_/edit?usp=sharing&ouid=101166119605484992249&rtpof=true&sd=true")

df\_population = pd.read\_csv("https://docs.google.com/spreadsheets/d/165zNpREsKpuOm008hZPTZ2km3FRBPflm/edit?usp=sharing&ouid=101166119605484992249&rtpof=true&sd=true")

df\_area = pd.read\_csv("https://docs.google.com/spreadsheets/d/1P35\_LAZAI65bGeHBqXkURzo98OLS0nBg/edit?usp=sharing&ouid=101166119605484992249&rtpof=true&sd=true")

df\_time = pd.read\_csv("https://docs.google.com/spreadsheets/d/1eOUww2RBh7IVCmISXoEKfDdDdIn9nBg/edit?usp=sharing&ouid=101166119605484992249&rtpof=true&sd=true")

# Обробка викидів та відсутніх значень

df\_sales = df\_sales[df\_sales["Ціна"] <= df\_sales["Ціна"].quantile(0.75) + 1.5 \* df\_sales["Ціна"].iqr()]

df\_sales["Площа"].fillna(df\_sales["Площа"].mean(), inplace=True)

df\_sales["Місто"].fillna("Невідомо", inplace=True)

# Перевірка формату дат

df\_time["Дата"].dt.year

# Перевірка числових значень

df\_sales["Ціна"].dtype

# Візуалізація та зіставлення даних

# Ціна з часом

plt.plot(df\_time["Дата"], df\_sales["Ціна"])

plt.title("Ціна нерухомості з часом")

plt.xlabel("Дата")

plt.ylabel("Ціна")

# Розподіл ціни

plt.hist(df\_sales["Ціна"])

plt.title("Розподіл ціни нерухомості")

plt.xlabel("Ціна")

# Взаємозв'язок між ціною та площею

plt.scatter(df\_sales["Площа"], df\_sales["Ціна"])

plt.title("Взаємозв'язок між ціною та площею нерухомості")

plt.xlabel("Площа")

plt.ylabel("Ціна")

# Середні значення ціни за житловими масивами

df\_sales.groupby("Місто")["Ціна"].mean()

# Т-тест для порівняння середніх значень ціни між двома житловими масивами

ttest\_ind(df\_sales[df\_sales["Місто"] == "Нью-Йорк"]["Ціна"], df\_sales[df\_sales["Місто"] == "Лос-Анджелес"]["Ціна"])

# 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.

Дані:

1. <https://docs.google.com/spreadsheets/d/1riKxRyRnpyB3snWrAALO-E1M7nMpyLLQ/edit?usp=drive_link&ouid=101166119605484992249&rtpof=true&sd=true>
2. <https://docs.google.com/spreadsheets/d/1eOUww2RBh7IVCmISXoEKfDdIn9u8kifF/edit?usp=drive_link>
3. <https://docs.google.com/spreadsheets/d/1P35_LAZAI65bGeHBqXkURzo98OLS0nBg/edit?usp=drive_link>
4. <https://docs.google.com/spreadsheets/d/165zNpREsKpuOm0-8hZPTZ2km3FRBPflm/edit?usp=drive_link>
5. <https://docs.google.com/spreadsheets/d/1MW-zi5SU4VHChqn1heGOi3zQx77dG17_/edit?usp=drive_link>