>>>

Project Title: Project Thrive – Dubai Real Estate Analysis

Author:

Name: Charisma Augustine

Email: rhizconnect@gmail.com

LinkedIn: http://linkedin.com/in/rhizaustin/

Github: https://github.com/August-96

Date of Publication:

January 30, 2024.

Introduction

Hello and welcome to Project Thrive! My name is Charisma Augustine, and I am passionate about data analysis and insights generation. In this project, I have explored a dataset containing Dubai real estate transactions, aiming to uncover trends, patterns, and valuable insights. This project is a work in progress!

About the Author

I am Entry-Level Data analyst and this is my debut project in data analytics. I enjoy leveraging data to extract meaningful insights that drive informed decision-making. Feel free to connect with me on LinkedIn or reach out via email if you have any questions or if you'd like to discuss this project further, or if you want to collaborate on any other project.

Acknowledgments

I would like to express my gratitude to [Any Acknowledgments or Collaborators], who contributed to the success of this project.

Conclusion

Thank you for exploring Project Thrive with me. I hope you find the insights valuable and applicable. Feel free to reach out if you have any feedback or if there are additional aspects, you'd like me to explore.

Welcome to Project Thrive by Charisma!

**Project Thrive – Dubai Real Estate Analysis**

Theme: Optimizing Real Estate Investments Through Data-Driven Insights

**Introduction:** The real estate market is dynamic and influenced by various factors. To make informed investment decisions and maximize returns, it's essential to extract actionable insights from available data. This project aims to analyze a comprehensive real estate dataset and highlight some insights that would be useful for investors, developers, and stakeholders in the industry. It is a continuous project which will be updated periodically.

**Problem Statement:**

The real estate market needs more data-driven decision-making and insights that would make one stay one step ahead of market challenges leading to suboptimal investment strategies and potential missed opportunities, to say ‘a few’. Investors and stakeholders could add to their wealth of understanding of transaction trends, property types, location-based dynamics, and pricing factors to enhance their investment portfolios.

**Objectives:**

Temporal Analysis: Understand how real estate transactions fluctuate over time to identify trends and seasonality.

Property Type Insights: Analyze the distribution and characteristics of different property types to guide investment preferences.

Location-based Strategy: Investigate the geographical distribution of transactions to identify lucrative areas and potential growth regions.

Pricing and Property Size Relationship: Explore the correlation between sale prices and property sizes to guide pricing strategies.

**Expected Impact:** This project aims to empower stakeholders in the real estate industry with actionable insights derived from data analysis. By addressing the identified objectives, we anticipate providing a foundation for informed decision-making, risk mitigation, and the optimization of real estate investments.

**Scope of Work:** The project will involve data cleaning, exploratory data analysis (EDA), statistical analysis, and the creation of visualizations to highlight findings effectively. The insights gained will contribute to the development of a data-driven approach for real estate investment.

**Success Criteria:** The success of the project will be measured by the relevance of the insights provided, the practicality of recommendations, and the potential positive impact on real estate investment strategies.

**Stakeholders:**

Real estate investors

Property developers

Financial analysts

Consultants

Government bodies involved in urban planning and development

Public.

**Deliverables:**

Analysis of real estate transactions.

Visualizations highlighting insights.

Foundation for optimized investment strategies.

**Project Questions:**

1. What is the overall structure and content of the dataset?
2. How do real estate transactions vary over time?
3. What are the distribution and characteristics of different property types?
4. How are transactions distributed across different locations?
5. What is the relationship between sale price and property area?

**Collection of data**

Data source: Open-Source Data from official website of Dubai Land Department.

https://www.dubaipulse.gov.ae/

<https://dubailand.gov.ae/>

**DATA CLEANING**

Most cleaning processes were carried out using Microsoft excel.

Refer to *‘CleanProx’* file for data cleaning processes.

**Hypotheses to Test:**

Two types of hypotheses were tested in this project:

Pricing Variation Hypothesis (Hypothesis 1):

Alternative Hypothesis (H1): There is a significant difference in average sale prices across different property types.

Null Hypothesis (H0): There is no significant difference in average sale prices across different property types.

Property Size and Price Relationship Hypothesis (Hypothesis 2):

Null Hypothesis (H0): There is no correlation between property size and sale price.

Alternative Hypothesis (H1): There is a significant correlation between property size and sale price.

**Steps to test Hypothesis 1:**

Data Preparation:

Ensured that "prop\_area(sq/m)" and "sale\_price(sq/m)" columns are numeric.

Checked for missing data or outliers.

Calculate Correlation Coefficient:

In Python, we used the ‘corr’ function with pandas to run Pearson correlation statistical method to calculate the correlation coefficient between property size and sale price.

Hypothesis Testing:

Formulate the null hypothesis (H0) and alternative hypothesis (H1).

Null Hypothesis (H0): There is no significant correlation between property size and sale price.

Alternative Hypothesis (H1): There is a significant correlation between property size and sale price.

Determine Significance Level (α):

A significance level of 0.05 was used to determine the threshold for statistical significance.

Statistical Test:

In Python, we used scipy library to perform a statistical test - correlation test to assess the significance of the correlation coefficient.

Evaluate Results:

Compared the p-value obtained from the test with the chosen significance level.

If the p-value is less than α, reject the null hypothesis and conclude that there is a significant correlation.

**Steps to test Hypothesis 2:**

Data Preparation:

Ensured that the "prop\_type" column contains categorical data representing different property types.

Ensured “sale\_price(sq/m)" column is numeric.

Hypothesis Formulation:

Formulated the null hypothesis (H0) and alternative hypothesis (H1).

Null Hypothesis (H0): There is no significant difference in average sale prices across different property types.

Alternative Hypothesis (H1): There is a significant difference in average sale prices across different property types.

Determine Significance Level (α):

A significance level of 0.05 was used to determine the threshold for statistical significance.

ANOVA Test:

In Python, we used the ‘f\_oneway’ function from the scipy library to perform ANOVA test.

Evaluate Results:

Examine the p-value obtained from the ANOVA test.

If the p-value is less than α, reject the null hypothesis and conclude that there is a significant difference in average sale prices.

If the p-value is greater than or equal to α, fail to reject the null hypothesis.

**Exploratory Data Analysis (EDA) Tasks:**

Temporal Analysis:

Explore transaction trends over time (monthly, yearly) to identify seasonality or trends.

Create a time series plot to visualize transaction volumes.

Property Type Insights:

Analyse the distribution of different property types in the dataset.

Compare average sale prices and transaction counts for each property type.

Location-based Analysis:

Examine the geographical distribution of transactions.

Identify areas with the highest transaction counts and average sale prices.

Pricing and Property Size Relationship:

Create a scatter plot to visualize the relationship between property size and sale price.

Use statistical measures to quantify the strength and direction of this relationship.

Categorical Analysis:

Explore the distribution of transaction types (trans\_type) and their impact on sale prices.

Use box plots or violin plots to visualize the spread of sale prices for different transaction types.

These exploratory data analysis tasks and hypotheses testing will provide valuable insights into the patterns, relationships, and potential influencing factors within the real estate dataset.