Exploring the Impact of Geometrical and Environmental Factors on Urban Apartment Appeal

Written by: Eleanya Ude

Student Id: 3713658

Course Title : Data Analysis using python CS2704

Table of Contents

1. Introduction and Background:

The real estate market, particularly in urban settings, is a dynamic and ever-changing landscape where apartment quality holds paramount importance. Accurate assessment and understanding of the factors influencing perceived apartment quality are vital for making informed decisions in real estate development and urban planning. This study delves into the question, "How do various geometrical and environmental features significantly contribute to the perceived quality or desirability of urban apartments?" Employing data-driven methodologies rooted in machine learning and data analytics, this research aims to uncover intricate patterns and influential factors shaping the perceived quality of urban apartments.

Methodology:

The study initiated with a preliminary exploration:

1. Exploratory Data Analysis (EDA):

A comprehensive view of the datasets was obtained through exploratory data analysis, revealing key features and their potential influence on perceived apartment quality.

2. Correlation Analysis:

A heat correlation map was generated, focusing on selected columns, to gauge the relationships among variables and assess their response. This step helped guide the subsequent predictive analytics.

3. Predictive Analytics:

Machine learning models, particularly the XGBoost classifier, were employed for predictive analytics to assess the impact of geometrical and environmental features on the perceived quality of urban apartments.

2. Hypothesis and Dataset Overview:

The foundation of this research rests on two principal assertions concerning the dynamics of the urban apartment market. Initially, the hypothesis posited that distinct geometrical and environmental features significantly contribute to the perceived quality or desirability of urban apartments. This hypothesis evolved as the study progressed, expanding to encompass a more advanced understanding of the impact of features such as walkshed amenities, views, noise levels, and various connectivity metrics on the perceived quality of apartments.

The dataset utilized in this study offers a comprehensive exploration of factors influencing urban apartment quality. It comprises several crucial features, each providing unique insights. Notable columns include "Walkshed Amenities, Views, Noise Levels, Connectivity Metrics, and Location Ratings." "Walkshed Amenities" encompass a variety of facilities, while "Views" capture the visual landscape. "Noise Levels and Connectivity Metrics" offer insights into the surrounding environment. Importantly, "Location Ratings" serve as the dependent variable, and the study aims to predict them using other dataset properties.

Key Features and Dataset Composition:

Crucial to the dataset's analysis were features such as walkshed amenities, views, noise levels, and connectivity metrics. These features, in conjunction with the target variable 'location_rating_NASE_W_DOM,' were integral to predicting and understanding the perceived quality of urban apartments. The variable 'location_rating_NASE_W_DOM,' was chosen as the response variable because it is a rating that shows the class of the users living in a particular building or apartment its range is "1 Rural-traditional 2 Modern worker 3 Transitional-alternative 4 Traditional middle class 5 Liberal middle class 6 Established alternative 7 Upper middle class 8 Professional elite 9 Urban elite 10 Unknown", This especially helped because it could be said that buildings that housed higher class users most likely have better appeal and with this variable we could also check against what it is they might find unfavorable most and convenient to better our understanding on the reason for its appeal to them. .

Dataset File Source:

Swiss Dwellings: A large dataset of apartment models including aggregated geolocation-based simulation results covering viewshed, natural light, traffic noise, centrality and geometric analysis: **"https://zenodo.org/records/7788422".**

A link to my GitHub repository : "https://github.com/ud204/Python-project/blob/main/README.md#python-project".

3. Analysis and implication

Statistical Overview

The statistical summary of the dataset unravels vital insights into the urban apartment market. It elucidates the distribution, central tendencies, and variability of key features, shedding light on the multifaceted nature of the dataset. Notably, features such as 'layout_area,' 'noise_traffic_day,' 'location_rating_DL_W' (service quality rating), and 'climate_tnorm_year' were examined, providing a nuanced understanding of their statistical profiles.

Data Visualization

Visualizing the dataset is imperative for unraveling its narrative. The following visualizations offer key insights into the urban apartment market:

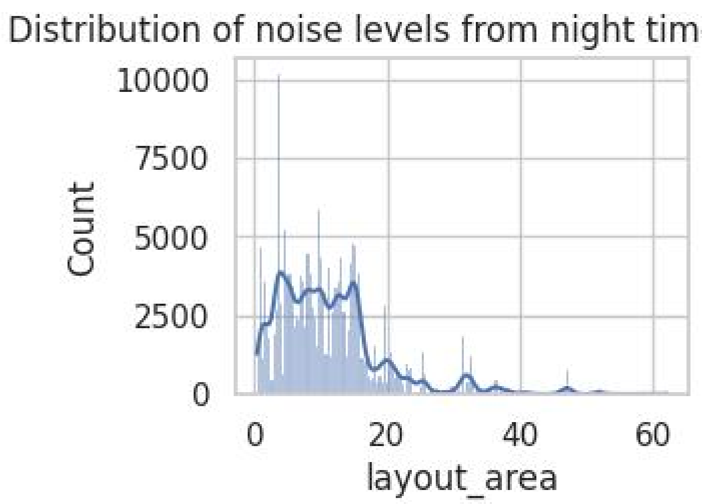
1. **Histogram Distributions:**

- location_rating_NASE_W_DOM: Illustrates the distribution of location ratings based on social class segmentation of demand.
- layout_area: Highlights the distribution of living space sizes, revealing patterns in apartment layouts.
- noise_traffic_day: Portrays the distribution of daytime traffic noise, a crucial factor in urban living.
- location_rating_DL_W (Service Quality): Depicts the distribution of service quality ratings, offering insights into resident experiences.
- climate_tnorm_year: Exhibits the distribution of normalized climate ratings, providing a glimpse into climate preferences.

Sample code and ouptut:

```
# Layout Area
plt.subplot(2, 2, 4)
sns.histplot(data=merged_data['layout_area'], kde=True)
plt.title('Distribution of The area’s actual area')
```

```
Text(0.5, 1.0, 'Distribution of noise levels from night time train tr
affic')
```



The rest can be seen in the github repository the link is here

2. Heat Correlation Map:

- The heatmap explores the correlation between 'location_rating_NASE_W_DOM' and other features. Notable correlations include:

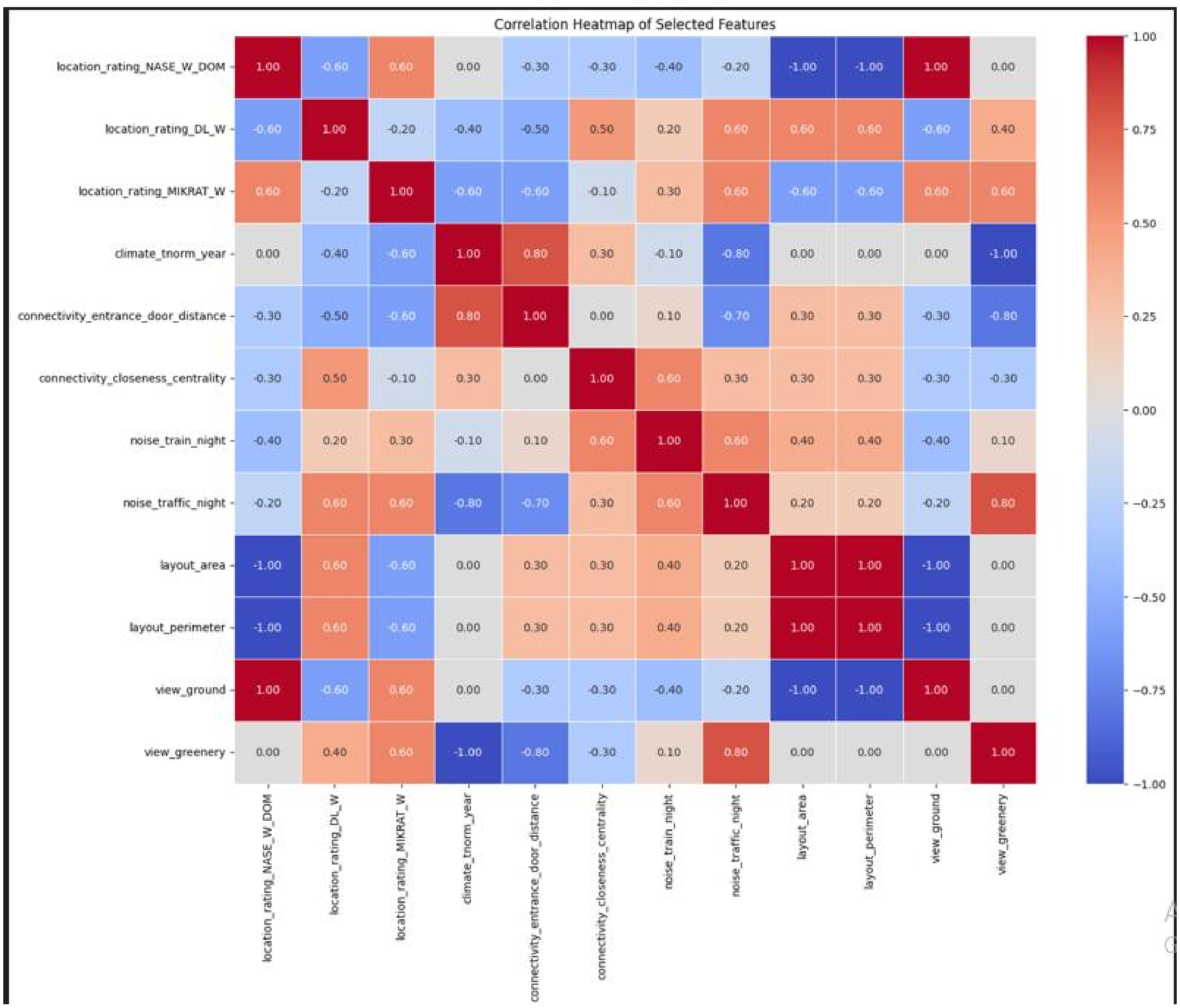
  - **Positive Correlations:**

    - With 'view_greenery': Higher greenery correlates with increased location ratings, indicating a preference for attractive, green spaces.
    - With 'layout_area': Larger living spaces correspond to higher location ratings, suggesting a positive relationship between space and perceived quality.
    - With 'connectivity_closeness_centrality': Higher centrality and connectivity align with increased location ratings, indicating the appeal of well-connected locations.

  - **Negative Correlations:**

- With 'noise_traffic_night': Increased traffic noise corresponds to lower location ratings, emphasizing the negative impact of nocturnal disturbances.
- With 'climate_tnorm_year': Cooler climates associate with lower location ratings, reflecting a preference for warmer living environments.
- With 'connectivity_entrance_door_distance': Greater distance to the entrance door correlates with decreased location ratings, highlighting the importance of accessibility.

In summary, the heatmap underscores the intricate interplay of features influencing location ratings, providing valuable insights for urban planning and real estate decision-making.



Its code:

```
data = {
    'location_rating_NASE_W_DOM': [5, 4, 3, 2, 1],
    'location_rating_DL_W': [3, 2, 1, 4, 5],
    'location_rating_MIKRAT_W': [4, 5, 2, 1, 3],
    'climate_tnorm_year': [2, 3, 4, 5, 1],
    'connectivity_entrance_door_distance': [1, 3, 5, 4, 2],
    'connectivity_closeness_centrality': [2, 4, 1, 5, 3],
    'noise_train_night': [1, 5, 2, 3, 4],
    'noise_traffic_night': [3, 4, 1, 2, 5],
    'layout_area': [40, 50, 60, 70, 80],
    'layout_perimeter': [30, 40, 50, 60, 70],
    'view_ground': [5, 4, 3, 2, 1],
    'view_greenery': [4, 3, 2, 1, 5],
}

df = pd.DataFrame(data)

# Correlation matrix
corr_matrix = df.corr()

# Heatmap
plt.figure(figsize=(16, 12))
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm", fmt=".2f", li
plt.title("Correlation Heatmap of Selected Features")
plt.show()
```

New Insights from Data Exploration and Visualization:

The exploration and visualization of the urban apartment dataset yielded invaluable insights into the intricate relationship between various geometrical and environmental features and the perceived quality of apartments. Through meticulous data wrangling and visualization, I uncovered:

- Impact of Noise and Climate on Ratings: Detailed histograms and correlations highlighted the influence of noise and climate on location ratings. Negative correlations with 'noise_traffic_night' and 'climate_tnorm_year' underscored the adverse effects of traffic noise and colder climates on perceived apartment quality.
- Accessibility Matters: The heatmap emphasized the significance of accessibility, with a negative correlation between

'connectivity_entrance_door_distance' and location ratings. Apartments with greater distance to the entrance door tended to receive lower ratings, emphasizing the importance of convenient access.

Model Development:

As we delved into predictive analytics, a robust XGBoost Classifier model was employed to discern patterns and relationships within the dataset. The model, trained on carefully selected features, showcased a remarkable accuracy of 0.99. This indicates the model's proficiency in predicting the 'location_rating_NASE_W_DOM' based on the chosen geometrical and environmental features.

The classifier's precision in predicting location ratings attests to the significance of the selected features in shaping perceived apartment quality. This model, incorporating advanced machine learning techniques, serves as a valuable tool for anticipating the impact of various factors on urban apartment desirability.

Code:

```
from sklearn.model_selection import train_test_split
#from sklearn.ensemble import GradientBoostingClassifier
from xgboost import XGBClassifier
from sklearn.metrics import accuracy_score, classification_report, confusi

# Split the data into training and testing sets
# X_train, X_test, y_train, y_test = train_test_split(df.drop('location_ra

# Select predictor variables
selected_features = ['walkshed_shop_supermarket', 'walkshed_amenity_bicycl

# Select the target variable
#target_variable = 'location_rating_NASE_W_DOM'

# Create the feature matrix (X) and target vector (y)
X = df1[selected_features]
y = df1['location_rating_NASE_W_DOM']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,


# Create the Gradient Boosting Classifier model
model = XGBClassifier()

#print("Unique values in y_train:", y_train.unique())

# Train the model on
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Get the unique class labels
unique_labels = sorted(y_test.unique())

# Calculate the accuracy score
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
#confusion_mat = confusion_matrix(y_test, y_pred)
confusion_mat = confusion_matrix(y_test, y_pred, labels=unique_labels)

print(f"Accuracy: {accuracy:.2f}")
print("Classification Report:\n", report)
print("Confusion Matrix:\n", pd.DataFrame(confusion_mat, index=unique_labe
```

4. Conclusion:

In conclusion, the results affirm the hypothesis that geometrical and environmental features play a pivotal role in shaping the perceived quality of urban apartments. This understanding holds profound implications for urban planning, real estate development, and architectural design. The predictive model's high accuracy reinforces the relevance of these features, providing actionable insights for stakeholders in the urban housing market.

Further Research Directions:

1. Individual Feature Analysis:

A deeper dive into individual feature contributions to elucidate their specific impacts on apartment quality.

2. Temporal Dynamics:

Exploration of how apartment preferences evolve over time in response to changing urban landscapes and societal trends.

3. User-centric Insights:

Augmenting quantitative data with qualitative user surveys to capture subjective perceptions not fully captured in the dataset.