

**Decoding Real Estate Trends with PySpark**

**Massive Mining of Datasets - BCS-8A**

**Instructor:**

Ms Zareen Alamgir

**Group Members:**

Sohaib Iqbal 19L-0958

Mohammad Muneeb 19L-2216

National University Of Computer and Emerging Sciences

Department of Computer Science

Lahore, Pakistan

**1. Introduction:** The real estate market plays a pivotal role in economic dynamics, reflecting both societal trends and financial opportunities. In this report, we present an exploratory analysis of the real estate market in Toulouse and its suburbs using PySpark, a powerful tool for large-scale data processing. Our analysis aims to uncover key insights into property prices, sales dynamics, and market segmentation to provide stakeholders with valuable information for decision-making.

**2. Dataset Description:** The dataset used for this analysis contains information on real estate transactions in Toulouse and its suburbs, including sales dates, property types, prices, and locations. It comprises 80,118 records, providing a comprehensive view of the local real estate market.

**3. Methodology:** We employed PySpark, a distributed computing framework for processing large datasets, to conduct our analysis. The methodology involved several steps, including data loading, preprocessing, exploratory data analysis (EDA), and visualization.

**4. Key Findings:**

**Total Sales Orders:** The dataset comprised 80,118 sales orders, indicating significant activity in the real estate market.

**Price Trends Over Time:** Analysis of price trends revealed fluctuations over time, with a notable spike in early 2015. This suggests dynamic market conditions influenced by various factors such as economic trends and policy changes.

**Price Distribution:** The majority of properties had low prices, highlighting affordability challenges in the market. This distribution can impact market dynamics and buyer behavior.

**Sales Outliers:** Outliers were identified based on Z-scores, indicating transactions with unusually high or low prices. Understanding these outliers is crucial for detecting anomalies and identifying potential investment opportunities or market inefficiencies.

**Price Difference Between Property Types:** Apartments had a higher average price compared to houses, reflecting differing preferences and demand for residential properties.

**Difference Between Types of Sales:** The analysis revealed differences in average prices between different types of sales transactions, indicating varying market dynamics and buyer preferences.

**Difference Between Toulouse and Suburbs:** Toulouse had the third-highest average price compared to its suburbs, suggesting higher demand and potentially different market dynamics.

**Market Segmentation:** Cities were categorized based on price thresholds into high-cost, low-cost, or emerging markets. This segmentation provides insights into market dynamics and investment opportunities in different areas.

**5. Conclusion:** Our exploratory analysis of the real estate market in Toulouse and its suburbs provides valuable insights for stakeholders, including real estate agencies, investors, and policymakers. By understanding price trends, sales dynamics, and market segmentation, stakeholders can make informed decisions regarding investment, development, and policy interventions. Further research could focus on predictive modeling and forecasting to anticipate future market trends and opportunities.

**6. Limitations and Future Directions:** While our analysis provides valuable insights, it has certain limitations. The dataset may not capture all relevant factors influencing the real estate market, such as socio-economic trends, infrastructure development, or regulatory changes. Future research could incorporate additional data sources and advanced modeling techniques to enhance the accuracy and robustness of the analysis.

In conclusion, our exploratory analysis sheds light on the dynamic nature of the real estate market in Toulouse and its suburbs, offering actionable insights for stakeholders navigating this complex and ever-changing landscape.