**Week 8**

**Implementation of Second Module**

This phase implements a **content-based filtering** approach, where property recommendations are based on similarity to user preferences (e.g., location or top facilities). The TF-IDF vectorizer transforms textual property features into numerical vectors by calculating how important a term is within a document relative to the entire dataset. This transformation helps the model identify key descriptive terms like "prime location", "gym", or "parking" as distinguishing features. To compare similarity, Cosine Similarity is applied to the TF-IDF vectors. Cosine similarity measures the cosine of the angle between two vectors; a value closer to 1 indicates high similarity. This allows the system to recommend properties that share similar attributes with the user's selected location or desired facilities. Each city’s recommendation logic is stored separately to maintain modularity and accuracy based on local real estate trends. Once the location names are successfully transformed into numerical vectors using the TF-IDF technique, the next crucial step is to evaluate how similar these vectors are to each other. For this purpose, the cosine similarity measure is applied. Cosine similarity is a mathematical technique used to determine the degree of similarity between two non-zero vectors in a multi-dimensional space by measuring the cosine of the angle between them. In the context of text data, this approach is particularly effective because it considers the direction of the vectors rather than their magnitude, which helps in identifying similarity in content regardless of word count or scale. In the Real Estate Explorer system, cosine similarity enables the comparison of user-inputted location names with existing entries in the dataset, even when they are spelled differently or abbreviated. Unlike exact string matching—which would treat “BTM Layout” and “BTM Lyt” as entirely different inputs—cosine similarity focuses on the overall context and structure of the vectorized form of the location name. If the angle between two TF-IDF vectors is small (meaning their cosine similarity is close to 1), it indicates a high degree of similarity between the two locations.