Nairobi AirBnb Capstone Project

(GROUP 3 PRESENTATION)

Group Members

- 1. Maureen Auka
- 2. Brian Kipchumba
- 3. Ochieng' Ouma
- 4. Andrew Chege

Introduction

- This project seeks to analyze Airbnb listings and reviews in Nairobi City to uncover key determinants of revenue and guest satisfaction.
- The project would use combined structured data modeling with NLP techniques to: build a price prediction model, extract review topics, and perform sentiment analysis for actionable insights.

Business Understanding

- Key question - "What are the key determinants of revenue and customer satisfaction in Airbnb?"
- According Riungu et al. (2025), Airbnb is becoming a trendy and lucrative investment in Kenya.
- The project is applicable in the real estate industry.

- The targeted audience:-

a) Investors in the Airbnb industry

b) Guests booking through this platform.

- Project's key goal:-

a) to have a model that would allow for investors in Airbnb business to identify ideal locations for acquiring and/or leasing apartments for their investment.

b) to have a model that will assist customers by showing the safe neighborhood, identify good hosts, and provide price range across the market.

- Pre-existing research in this field:-
  a) Capstone Project - Airbnb Cape Town Business Analysis by Salrahim21
  b) Airbnb Investment Project by amtherwi.

Data Understanding

- Import Libraries
- Loading the Data:-
  a) Data/Future Calendar Rates.csv
  b) Data/Listings (2).csv
  c) Data/Past Calendar Rates.csv
  d) Data/Reviews (2).csv

- check the first 5 rows of each data
- check the columns and the shape of the data
- Summary Statistics
- Check for Missing Values in both Counts and Percentage
- Check for unique values
- Check data types

Data Cleaning

- Remove duplicates
- Handle missing values
- Convert data types
- Merge the data

Exploratory Data Analysis

- Univariate Analysis
- Bivariate Analysis
- Time-based Analysis
- Correlation Analysis
- Geospatial Analysis
- Time Series / Seasonality

Modelling

Price Prediction Model & Occupancy Prediction Model:

- Linear Regression Baseline Model
- Random Forest
- XGBoost Baseline
- XGBoost Hyperparameter Tuning

Conclusion

- **Location, amenities, and property characteristics** strongly influence pricing and occupancy.
- **Sentiment analysis** reveals that positive reviews often correlate with higher occupancy and better ratings.
- **Geospatial mapping** highlights distinct clusters of high-performing Airbnb listings across Nairobi.
- **The Random Forest model** demonstrates strong predictive performance, with predictions closely aligned to actual values. This validates its suitability for deployment in the Airbnb price prediction task.