Airbnb Listings Dashboard

This project provides a comprehensive analysis of Airbnb listings data through an interactive dashboard built with Streamlit and includes data generation and Spark analysis components.

Files:

main.py: Contains SQL queries and creates the airbnb\_queries.db database. This script processes the raw Airbnb data and prepares it for visualization.

airbnb\_dashboard.py: The main Python file for the Streamlit dashboard. It creates an interactive web interface to visualize and explore the Airbnb data.

airbnb\_queries.db: SQLite database containing pre-processed query results. This database is created by main.py and read by airbnb\_dashboard.py.

fake\_data.py: Generates fake Airbnb listing data and saves it as JSON files. It creates both a full dataset and a sample dataset.

spark\_analysis.py: Performs analysis on the fake Airbnb data using PySpark. It calculates the top 5 cities with the highest average listing prices for each property type and saves the results to a SQLite database.

airbnb\_listings.json: (Generated by fake\_data.py) Contains the full fake Airbnb listing dataset.

airbnb\_listings\_sample.json: (Generated by fake\_data.py) Contains a sample of 50 fake Airbnb listings.

airbnb\_analysis\_results.db: SQLite database containing the results of the Spark analysis.

requirements.txt: List of required Python packages for running the project.

airbnb-listings-mini.csv – Short version of the big CSV file.

Purpose:

The purpose of this project is to generate, analyze, and visualize Airbnb listing data, providing insights into pricing trends, property types, guest satisfaction, and geographical distribution of listings. The project demonstrates the use of various data processing and analysis tools including Faker for data generation, PySpark for big data analysis, and Streamlit for interactive data visualization.

Setup and Running:

1. Ensure you have Python 3.7+ installed

2. Install required packages: pip install -r requirements.txt

3. Generate fake data: python fake\_data.py

4. Run Spark analysis: python spark\_analysis.py

5. Run the main script to process data: python main.py

6. Launch the Streamlit dashboard: streamlit run airbnb\_dashboard.py

Contributors:

Aviv Erez - 211356027

Lior Yakobovich - 207258625