

Northeastern University

Project Draft 1 Report

IE 7374 Data Warehousing and Business Intelligence

Spring 2024

Team 18

Team Member's Name

Meet Vasani

Mihir Kakadiya

Project Proposal Summary:

We propose to transform the Boston Airbnb dataset, consisting of 95 columns, into dimension and fact tables. This structured approach will involve grouping related columns into dimension tables such as Listing, Host, Location, Property, Availability, and Review, while aggregating numerical data into a Fact Table capturing pricing details. By implementing this transformation, we aim to streamline data analysis, enhance decision-making processes, and provide valuable insights for Airbnb hosts and guests.

Analysis Objectives:

In addition to structuring the dataset into dimension and fact tables, we aim to conduct in-depth analysis on property listings based on geographical areas, pricing dynamics, and neighbourhood-specific trends. This analysis will involve:

Geographical Analysis:

- Segmenting property listings by geographical areas such as neighbourhoods, cities, and states.
- Analysing the distribution of listings across different areas to identify popular and emerging markets.
- Investigating spatial patterns and correlations between listing characteristics and geographical features.

Price Analysis:

- Studying the pricing variations across different geographical areas and neighbourhoods.
- Identifying factors influencing pricing decisions, such as property type, amenities, and seasonal demand.
- Conducting comparative analysis to determine price competitiveness and value propositions for listings in specific areas.

Neighbourhood Analysis:

- Examining property listings within individual neighbourhoods to understand localized demand and preferences.
- Assessing neighbourhood-specific amenities, attractions, and accessibility factors influencing listing popularity and pricing.
- Utilizing sentiment analysis or reviews data to gauge neighbourhood reputation and guest satisfaction levels.

Extra Charges Analysis:

- Investigating additional charges such as cleaning fees, security deposits, and fees for extra guests.
- Analysing the impact of extra charges on booking patterns, guest satisfaction, and overall listing performance.
- Identifying optimal pricing strategies for extra charges to maximize revenue while maintaining competitiveness.

Here's a description of each dimension and fact table along with their columns that we will make from raw dataset:

1. Listing Dimension Table:

- Listing_id: Unique identifier for each listing.
- Name: Name of the listing.
- Description: Description of the listing.
- Transit: Information about transportation options nearby.
- House_rules: Rules set by the host for guests.

- Interaction: Information about interaction with guests provided by the host.
- Guest_included: Number of guests included in the listing.
- Min_night: Minimum number of nights required for booking.
- Max_night: Maximum number of nights allowed for booking.
- Require_license: Indicates if a license is required for booking.
- Instant_bookable: Indicates if the listing supports instant booking.
- Cancellation_policy: Policy regarding cancellation of bookings.

2. Host Dimension Table:

- Id: Unique identifier for each host.
- Name: Name of the host.
- Host_since: Date when the host joined Airbnb.
- Host_location: Location of the host.
- Host_response_time: Response time of the host to inquiries.
- Host_response_rate: Rate of response by the host.
- Host_total_listings: Total number of listings managed by the host.
- Host_identity_verified: Indicates if the host's identity is verified.

3. Location Dimension Table:

- Street: Street address of the listing.
- Neighbourhood: Neighbourhood where the listing is located.
- City: City where the listing is located.
- State: State where the listing is located.
- Country: Country where the listing is located.
- Latitude: Latitude coordinates of the listing.
- Longitude: Longitude coordinates of the listing.
- Is_location_exact: Indicates if the location is exact.

4. Property Dimension Table:

- Property_type: Type of property (e.g., apartment, house, etc.).
- Room_type: Type of room (e.g., entire home, private room, shared room).
- Accommodates: Maximum number of guests the property can accommodate.
- Bathrooms: Number of bathrooms in the property.
- Bedrooms: Number of bedrooms in the property.
- Beds: Number of beds in the property.
- Amenities: Amenities provided in the property.

5. Availability Dimension Table:

- Availability_30: Number of days the property is available for booking in the next 30 days.
- Availability_60: Number of days the property is available for booking in the next 60 days.
- Availability_90: Number of days the property is available for booking in the next 90 days.
- Availability_365: Number of days the property is available for booking in the next 365 days.
- Calendar_updated: Date when the calendar was last updated.

6. Review Dimension Table:

- Number_of_reviews: Number of reviews received for the listing.
- Review_score_rating: Overall rating score given by guests.
- Cleanliness_score: Score given for cleanliness.
- Checking_score: Score given for the check-in process.
- Communication_score: Score given for communication with the host.
- Location_score: Score given for the location of the listing.
- Value_score: Score given for the value provided by the listing.

7. Fact Table:

- Price: Price per night for the listing.
- Weekly_price: Price for a week's stay.
- Security_deposit: Amount of security deposit required.
- Cleaning_fees: Fees for cleaning the property.
- Extra_people_price: Price for additional guests beyond the included number.

Citation

<https://www.kaggle.com/datasets/airbnb/boston>