

DATA PREPARATION

BELLEVUE UNIVERSITY

WEEK 3 & 4: PROJECT MILESTONE 1

Project Milestone 1 Objective: The first milestone of this project will be to select the data you want to work with. We will need to select 3 different data sources that have different file types of information – and the data we will need to have a relationship between them:

- 3 data sources, along with a description of each one (links to each are fine, no need to submit the actual data)
- The relationships between them, or the relationship you will make between them
- What you believe you will have to do to the data to accomplish all 5 milestones and what your interpretation is of what the data means (you could provide a data dictionary or a summary of what the data is) – should be at least 250 words
- Project Subject Area: Describe your project in 1-2 sentences
- Data Sources:
 - Flat File:
 - Description
 - Link or Flat File uploaded

- API:
 - Description
 - Link
- Website:
 - Description
 - Link
- Relationships
 - Describe how the data from each source is connected (see example below).
 - If there isn't an obvious relationship, explain how you will make one
- 250 Words describing how you plan to tackle the project, what the data means, the ethical implications of your project scenario/topic, and what challenges you might face.

Project Description

Many travelers now have access to fantastic, straightforward, and practical lodging from Airbnb. Similar to that, it has provided several people with a means of generating additional income by marketing their homes as places for travelers to stay. For week 3 & 4 milestone I have selected 3 data sources named as (**Listings, Calendar, and Reviews**) to query and design the Airbnb API and websites to easily provide the data to end users and facilitate easy booking and querying from UI.

Data Sources

To complete the project milestone1 I have selected the AIRBNB project and downloaded datasets from Tableau Community. The size of data set meets the project requirements minimum of 1000 rows across all datasets and total of 30 columns across the 3 datasets selected. All the data sources are uploaded to GITHUB link mentioned below.

https://github.com/santoshgithub100/DSC540_TermEndProject

Listings.csv : Summary information and metrics for listings in San Francisco (good for visualizations).

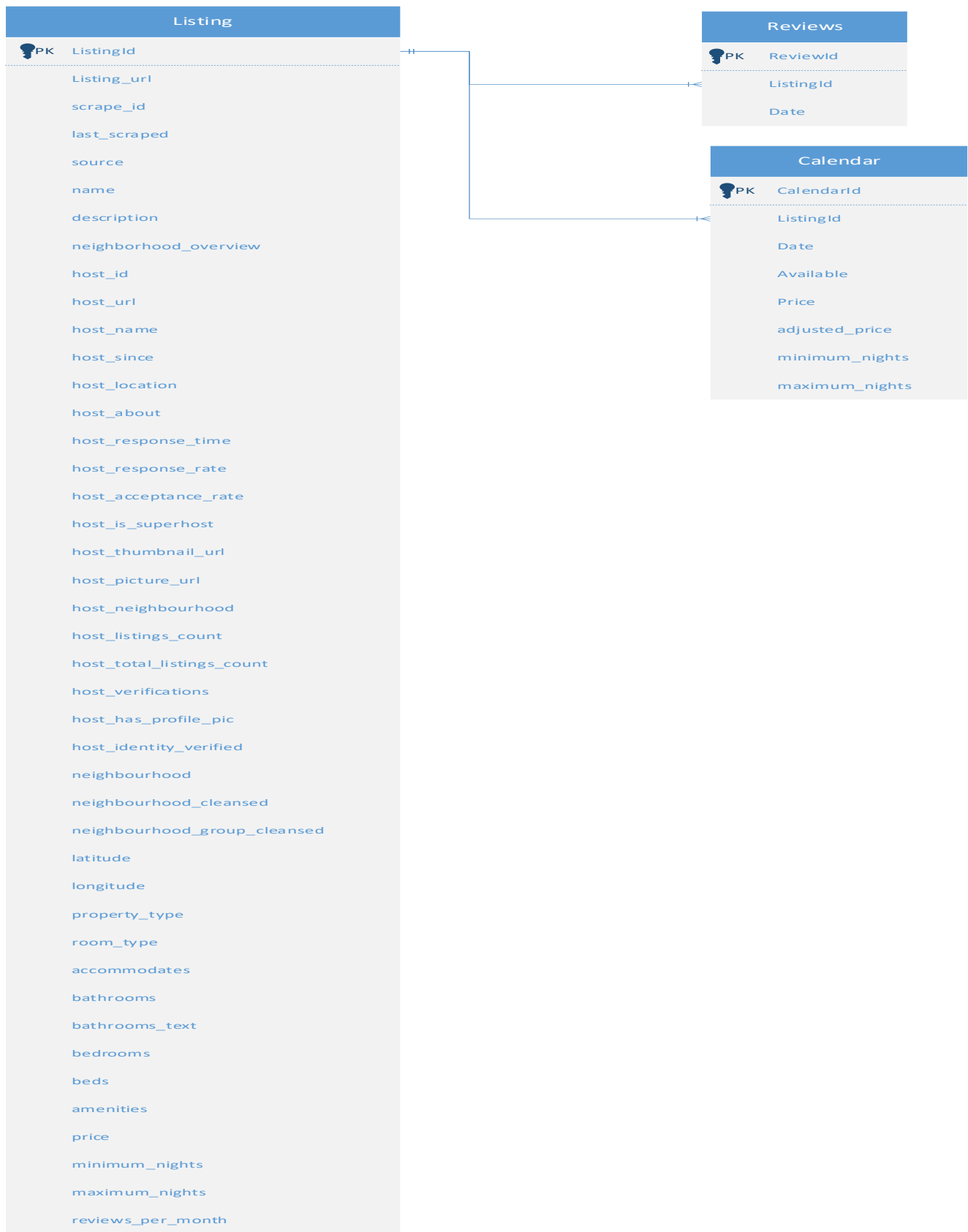
Reviews.csv: Summary Review data and Listing ID (to facilitate time-based analytics and visualizations linked to a listing).

Calendar.csv: details of Calendar data to facilitate rooms availability and customer bookings.

Entities and their association with other entities.

1. The Listing entity has 1: 1 relationship with Calendar entity
2. The Review entity has 1:1 relationship with Calendar and Listing entity.
3. The ListId is a primary key in the Listing entity and Foreign Key in the Reviews and Calendar entity.

The AIRBNB entities relationship



API

- All functions return promises
- Contains the transactions at each Airbnb bookings – contains a transaction ID and store ID. The API response data format is pre-parsed JSON
- All the API will have error reporting and data validation
- **getPublicListingCalendar**: Public availability and price data on a listing, count is the duration in months
- **getCalendar**: Private calendar data regarding your listings. Reservations, cancellations, prices, blocked days
- **getListingInfo**: To provide public facing data on any listing
- **getListingDetails**: To get the Airbnb listing search
- **getListingInfoHost**: To get the private data on one of customer listings
- **getReviews**: To provide the reviews

Website

Provides a list of affordable homes for travelers with great reviews, things available in neighborhood for easy room booking and Additionally, if a traveler wants to find the cheapest listing available but with certain features he can prefer like 'free parking' etc., Thus providing an excellent, easy, and convenient place to stay during travel.

Ethical Implications

While designing websites and API we should be mindful of the content we create. The websites should not promote hate speech, discrimination, or any form of harmful or

offensive content. So, it is very important to create a website that contribute positively to the online community and promote inclusivity, diversity, and respect.

Challenges

Creating APIs and datasets can be both rewarding and challenging tasks, especially when we aim to provide reliable and valuable data or services to users or applications.

Here are some common challenges associated with creating APIs and datasets:

- **Data Quality:** Ensuring that the dataset is accurate, complete, and reliable is one of the primary challenges. Cleaning, validating, and curating data can be time-consuming
- **Data Privacy:** Managing sensitive or private data and adhering to privacy regulations (e.g., GDPR) is crucial. Anonymization and data protection techniques must be applied correctly
- **Security:** Security is a top concern for APIs. Protecting against unauthorized access, data breaches, and attacks such as SQL injections or cross-site scripting requires robust authentication and authorization mechanisms
- **Scalability:** APIs should handle increasing loads and user traffic efficiently. Designing for scalability often involves load balancing, caching, and considering potential bottlenecks in the system