# Data Science R Capstone Projects Proposal

Daniel Yap

#### **Proposal ONE**

Customer Reviews Analysis
Natural Language Processing
(NLP)





#### Source of Data

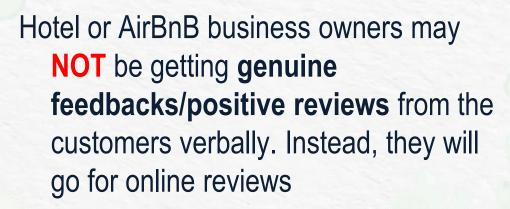


https://www.kaggle.com/datasets/setseries/airbnb-listings-in-california





## Problem Statement





# Sentiment Analysis (NLP)

Analyze reviews to **extract sentiments** and understand
the satisfaction level of guests



#### Audience

- AirBnb Businesses
  - e.g. Hotel Business
  - e.g. Property Owners
- Hotel Booking App Business



# **Proposal TWO**

Booking & Pricing Prediction
Time Series







#### Source of Data

Kaggle "AirBnB Listings in California"







#### Problem Statement

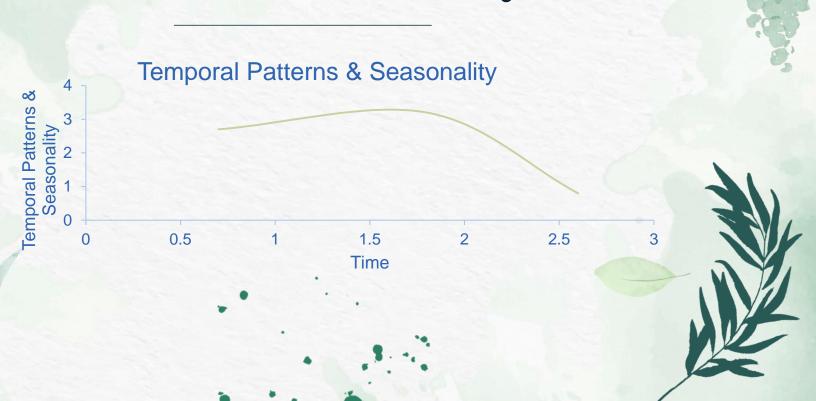




### Time Series Analysis

Explore temporal patterns and seasonality in booking trends and prices. This can help in understanding peak seasons, demand fluctuations, and pricing strategies over time.

# Time Series Analysis



#### Audience

- Hospitality Business

e.g. Hotel

e.g. AirBnB

- Tourism Agencies

e.g. Tourism Board

e.g. Tour Agency Business



## **Proposal THREE**

Property Listings
Categorization
Unsupervised Learning







#### Source of Data

Kaggle "AirBnB Listings in California"







## Problem Statement

No proper categorization of the property listings across different locations/areas





# Clustering Analysis (Unsupervised)



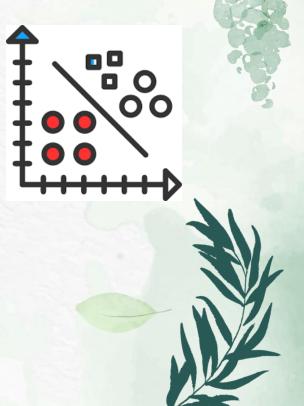
- 1. Location
- 2. Amenities
- 3. Pricing





# Clustering Analysis (Unsupervised)

Using clustering algorithms like **k-means** or **Hierarchical Clustering**. This can help in identifying market segments and targeting specific customer groups.



#### Audience

- Real Estate Agencies
- Property Owners

