Project Summary

Batch details	PGPDSE – FT Chennai Jul21
Team members	Mohanraj Kandhasamy, Rajendran Karthick Sharan, Yogeshwar Mohan, Dineshkumar Anbalagan, Pydipati Praneeth Kumar
Domain of Project	Retail
Proposed project title	Price prediction of Airbnb Accommodation
Group Number	Group 4
Team Leader	Mohanraj Kandhasamy
Mentor Name	Mupidi Srikar

Dataset name

Introduction to the problem/domain/background details: The project will be based on the retail domain. In this project we would like to analyze the factors that affect the pricing of Airbnb accommodations and to help hosts to give the best price to the customers and also make good profit in this post pandemic resurgence. We would like to do this with the help of a Machine learning algorithm to find hidden information and relations.

Problem Statement: Predicting the price of stay in Airbnb, with the help of a suitable Machine learning algorithm.

Business problem/ Impact in business of your problem/Need for this study/Abstract (Executive summary): Since 2008, Airbnb has helped guests and hosts to travel in a more unique, personalized way. The company went from a single air mattress for rent to global cooperation valued at more than 30 billion dollars all thanks to its energetic founder- Brian Chesky. 2020 was supposed to be the golden year for Airbnb as it would go public and issue the worlds' most sought after stocks. Tragically, Coronavirus happened. The travel sector was gutted by the pandemic. Airbnb now faces burning cash, angry hosts and an uncertain future as 2000 employees could potentially be discharged from their positions and the billion dollars debts with a high-interest rate that is being built to refund their customers. With the help of this project, we would like to help predict the prices of the Airbnb Accommodation based on several factors and help the hosts to quote a suitable price that will satisfy the travelers and to ensure that the hosts with healthy profits, which in-turn will help Airbnb get back on its feet.

Variable identification: Independent variables and Target

Variable information/Data description:

id	integer	Airbnb's unique identifier for the listing

scrape_id	bigint	Inside Airbnb "Scrape" this was part of
last_scraped	datetime	UTC. The date and time this listing was "scraped".
name	text	Name of the listing
description	text	Detailed description of the listing
neighborhood_overview	text	Host's description of the neighbourhood
picture_url	text	URL to the Airbnb hosted regular sized image for the listing
host_id	integer	Airbnb's unique identifier for the host/user
host_url	text	The Airbnb page for the host
host_name	text	Name of the host. Usually just the first name(s).
host_since	date	The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
host_location	text	The host's self-reported location
host_about	text	Description about the host
host_response_time		
host_response_rate		
host_acceptance_rate		That rate at which a host accepts booking requests.
host_is_superhost	boolean [t=true; f=false]	
host_thumbnail_url	text	
host_picture_url	text	
host_neighbourhood	text	
host_listings_count	text	The number of listings the host has (per Airbnb calculations)
host_total_listings_count	text	The number of listings the host has (per Airbnb calculations)
host_verifications		
host_has_profile_pic	boolean [t=true; f=false]	
host_identity_verified	boolean [t=true; f=false]	
neighbourhood	text	
neighbourhood_cleansed	text	The neighbourhood as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital

neighbourhood_group_cleansed	text	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.
latitude	numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
longitude	numeric	Uses the World Geodetic System (WGS84) projection for latitude and longitude.
property_type	text	Self selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field
room_type	text	[Entire home/apt|Private room|Shared room|Hotel]
accommodates	integer	The maximum capacity of the
bathrooms	numeric	The number of bathrooms in the listing
bathrooms_text	string	The number of bathrooms in the listing.
bedrooms	integer	The number of bedrooms
beds	integer	The number of bed(s)
amenities	json	
price	currency	daily price in local currency
minimum_nights	integer	minimum number of night stay for the listing (calendar rules may be different)
maximum_nights	integer	maximum number of night stay for the listing (calendar rules may be different)
minimum_minimum_nights	integer	the smallest minimum_night value from the calender (looking 365 nights in the future)
maximum_minimum_nights	integer	the largest minimum_night value from the calender (looking 365 nights in the future)
minimum_maximum_nights	integer	the smallest maximum_night value from the calender (looking 365 nights in the future)

maximum_maximum_nights	integer	the largest maximum_night value from the calender (looking 365 nights in the future)
minimum_nights_avg_ntm	numeric	the average minimum_night value from the calender (looking 365 nights in the future)
maximum_nights_avg_ntm	numeric	the average maximum_night value from the calender (looking 365 nights in the future)
calendar_updated	date	
has_availability	boolean	[t=true; f=false]
availability_30	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_60	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_90	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
availability_365	integer	avaliability_x. The availability of the listing x days in the future as determined by the calendar. Note a listing may not be available because it has been booked by a guest or blocked by the host.
calendar_last_scraped	date	
number_of_reviews	integer	The number of reviews the listing has
number_of_reviews_ltm	integer	The number of reviews the listing has (in the last 12 months)
number_of_reviews_l30d	integer	The number of reviews the listing has (in the last 30 days)
first_review	date	The date of the first/oldest review
last_review	date	The date of the last/newest review
review_scores_rating		

review_scores_accuracy		
review_scores_cleanliness		
review_scores_checkin		
review_scores_communication		
review_scores_location		
review_scores_value		
license	text	The licence/permit/registration number
instant_bookable	boolean	[t=true; f=false]. Whether the guest can automatically book the listing without the host requiring to accept their booking request. An indicator of a commercial
calculated_host_listings_count	integer	The number of listings the host has in the current scrape, in the city/region geography.
calculated_host_listings_count_entire_homes	integer	The number of Entire home/apt listings the host has in the current scrape, in the city/region
calculated_host_listings_count_private_rooms	integer	The number of Private room listings the host has in the current scrape, in the city/region
calculated_host_listings_count_shared_rooms	integer	The number of Shared room listings the host has in the current scrape, in the city/region
reviews_per_month	numeric	The number of reviews the listing has over the lifetime of the listing

Future Work/Methodology (Details of algorithms): Once this process has been implemented we can collect the reviews of the customers and perform sentiment analysis using NLP. We can also deploy the price prediction algorithm.

Timeline Chart (Weekly plan): Tentative weekly plan that you will be following.

WEEK 1	21-11-2021	SYNOPSIS - Understanding the Dataset and Business Aspects, Performing
WEEK 2	28-11-2021	EDA and various Feature engineering techniques on dataset
WEEK 3	02-12-2021	Interim presentation & Report
WEEK 4	12-12-2021	Working Progress Status I

WEEK 5	19-12-2021	Working Progress Status II
WEEK 6	30-12-2021	Final Report
WEEK 7	31-12-2021	Final Presentation

References (Data set source/Journals/articles)

Data set Source: https://public.opendatasoft.com/explore/dataset/airbnb-listings/table/?disjunctive.host_verifications&disjunctive.amenities&disjunctive.features

Declaration: This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated to it and can be used to showcase the work we do on it as a presentation in Great Learning.