


AIRBNB PRICE PREDICTION

Problem Statement

- The purpose of this project is to explore the data and predict the price of an Airbnb listing from features extracted from the listings.
- 
- Several thin, white, parallel diagonal lines are positioned in the bottom right corner of the slide, extending from the right edge towards the center.

Data Source

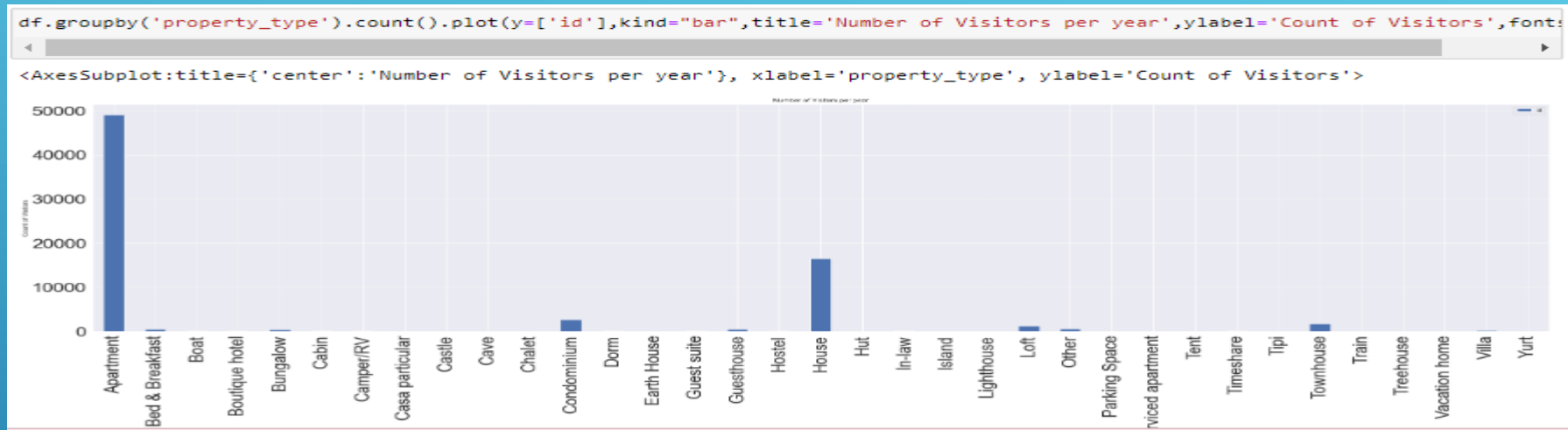
- <https://www.kaggle.com/stevezhenghp/airbnb-price-prediction>
 - Number of columns : 29
 - Number of Rows : 74112
 - Dataset Size: 99MB
- 
- A series of white diagonal lines of varying lengths and thicknesses are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

Data Types of the Dataset variables

```
df.dtypes
```

id	int64
log_price	float64
property_type	object
room_type	object
amenities	object
accommodates	int64
bathrooms	float64
bed_type	object
cancellation_policy	object
cleaning_fee	bool
city	object
description	object
first_review	object
host_has_profile_pic	object
host_identity_verified	object
host_response_rate	object
host_since	object
instant_bookable	object
last_review	object
latitude	float64
longitude	float64
name	object
neighbourhood	object
number_of_reviews	int64
review_scores_rating	float64
thumbnail_url	object
zipcode	object
bedrooms	float64
beds	float64
dtype:	object

Bar graph and pivot table to visualize the count of visitors for each property type




```
table = pd.pivot_table(df, values='id', columns=['property_type'], aggfunc='count', fill_value=0)
```

table

Apartment	Bed & Breakfast	Boat	Boutique hotel	Bungalow	Cabin	Camper/RV	Casa particular	Castle	Cave	...	Serviced apartment	Tent	Timeshare	Tipi	Townhouse	Train
49003	462	65	69	366	72	94	1	13	2	...	21	18	77	3	1692	2

Proposed Solution

- EDA will be performed through pyspark.
 - Conclude whether the price of the Airbnb listing is based on the ratings or not.
 - Check whether the price of the Airbnb listing is based on the property type and amenities.
 - Average rating of the Airbnb listing based on the number of hosts reviewed.
- 
- Three white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, extending from the right edge towards the center.