

AI MENTORSHIP

Final Project

DATASET A



Context

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019.

Content

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

DATASET B



Context

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

This hotel booking dataset can help you explore those questions!

Content

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

All personally identifying information has been removed from the data.

WHAT TO DO WITH DATA A

1. Get the big picture of what the data is about.
2. Inspect the properties of the data in each column (type, missing value, etc)
3. Show the amount of data, mean, median, and std.
4. How much is the min and max price?
5. Get rid of the missing value:
 - a. Drop columns which has missing value that cannot be filled with any logical value (like name, host_id, etc)
 - b. Fill the missing value in **last_review** with the most common date
 - c. Fill the missing value in **reviews_per_month** with the mean
6. Make category based on the **price**: Cheap, Medium, and Expensive
7. Filter all data which has **last_review** in > 2019
8. Visualize the **price** to compare the price of available rooms

WHAT TO DO WITH DATA B

1. Get the big picture of what the data is about.
2. Inspect the properties of the data in each column (type, missing value, etc)
3. Show the amount of data, mean, median, and std.
4. How much is the min and max for column with numerical value?
5. Get rid of the missing value:
 - a. Drop columns which has missing value more than 50% of the total data
 - b. Fill the missing value in **agent** with the mean
 - c. Fill the missing value in **country**, **children** with the most common name
6. Make category based on the **lead time**: Short, Medium, and Long
7. Filter all data which has **reservation_status_date** in >2017
8. Visualize the **arrival_date_month** to see in what month the most visitors come

DATA A

1. Ariadi
2. Nugraha
3. Lili
4. Heri
5. Eriko
6. Hano

DATA B

1. Ferdi
2. Fakhri
3. Daffa
4. Safira
5. Rizki
6. Hasbi