Problem Statements

Data Import and Data Exploration

1. Read bookings data in a datagrame
2. Explore bookings data
3. Make a bar chart
4. Read rest of the files
5. Explore aggregate bookings
6. Find out unique property ids in aggregate bookings dataset
7. Find out total bookings per property_id
8. Find out days on which bookings are greater than capacity
9. Find out properties that have highest capacity

Data Cleaning

1. Clean invalid guests
2. Outlier removal in revenue generated
3. In aggregate bookings find columns that have null values. Fill these null values with whatever you think is the appropriate subtitute (possible ways is to use mean or median)
4. In aggregate bookings find out records that have successful_bookings value greater than capacity. Filter those records

Data Transformation

1. Create occupancy percentage column
2. Convert it to a percentage value

Insights Generation

1. What is an average occupancy rate in each of the room categories?
2. Print average occupancy rate per city
3. When was the occupancy better? Weekday or Weekend?

4. In the month of June, what is the occupancy for different cities
5. We got new data for the month of august. Append that to existing data
6. Print revenue realized per city
7. Print month by month revenue
8. Print revenue realized per hotel type
9. Print average rating per city
10. Print a pie chart of revenue realized per booking platform