## 1. 数据加载与初步查看

### 步骤

* 加载数据并检查数据结构，包括缺失值、数据类型、基本统计描述等
* 观察数据的地理分布，初步了解不同区域的出租信息分布情况

### 可视化

* 显示数据的基本统计信息和字段概览，确认所需字段是否完整以及缺失值的情况

## 2. 数据清洗与处理

### 步骤

* 检查数据中的缺失值，并决定是否填补或删除
* 处理与研究问题相关的字段

### 可视化

* 使用缺失值热图，直观展示缺失值的情况
* 使用直方图或箱线图检查价格和评论数量的异常值，以处理数据偏差

## 3. 各区域房源数与评论数分析

### 步骤

* 分析不同**neighbourhood\_group**（区域组）和**neighbourhood**（街区）的房源数量
* 统计不同区域的**number\_of\_reviews**（评论数量）均值，评论数量可以作为流量的一个指标

### 可视化

* 绘制不同区域的房源数量柱状图，以了解各区域的整体分布情况
* 绘制评论数量的条形图，以展示每个区域的流量强弱

## 4. 流量密集区域的地理分布

### 步骤

* 利用地理数据（**latitude**和**longitude**），在地图上显示房源的分布密集度
* 使用不同颜色或大小的点标示高流量（高评论数量）的房源，以识别流量密集区域

### 可视化

* 使用地理散点图（例如散点密度图或热力图）展示房源的分布情况
* 不同区域的流量密集度使用颜色或热力图的方式来标识

## 5.流量差异分析：房间类型、价格与最低入住天数

### 步骤

* 分析不同房间类型（**room\_type**）在不同区域的分布和流量差异，观察哪些房间类型更受欢迎
* 检查价格（**price**）对流量的影响，是否有价格较低的房源流量更大
* 分析最低入住天数（**minimum\_nights**）对流量的影响，尤其是一些热门区域可能有更短的最低入住要求

### 可视化

* 使用箱线图或分组柱状图展示不同房间类型在各区域的评论数量分布
* 绘制散点图，展示价格与评论数量之间的关系
* 绘制柱状图，展示最低入住天数对评论数量的影响

## 6. 结论

### 步骤

* 总结各区域的流量情况，分析哪些区域的流量高并尝试解释原因
* 根据不同区域的房源类型、价格和流量特征提出进一步的洞察，例如某些区域的低价短期房源更具吸引力等