



网易出品
PRODUCE BY NETEASE

短租房案例分析

《数据分析师》微专业



大数据案例分析--短租房数据分析

短租房数据介绍

1、数据来源和信息介绍

本案例中的数据来自于爱彼迎（Airbnb）网站2018-2019年度的多伦多市的真实数据。

数据集中包含listings数据集，约有2万条数据，记录着所有的房屋信息，包括价格在内的几十项信息字段。

数据集中的另一个数据集是calendar，包含约650万条的租房交易数据，拥有每一天每一所住房的入驻信息。

数据分析思路

2、数据案例分析思路步骤1

常规数据分析，数据字段载入和常见数据ETL四板斧的清洗处理，

招数1 `.isnull().sum()` 检查空值情况

招数2 `.shape` 检查数据尺寸

招数3 `.describe` 查看数据字段数据类型

招数4 `.value_counts()` 查看数据集合数据分布情况

数据分析思路

2、数据案例分析思路步骤1案例截图

```
In [5]: calendar.isnull().sum()
Out[5]: listing_id      0
        date            0
        available      0
        price      4069436
        dtype: int64

In [6]: calendar.shape
Out[6]: (6326545, 4)

In [7]: calendar.head()
Out[7]:
```

	listing_id	date	available	price
0	349827	2019-06-17	t	\$200.00
1	349827	2019-06-16	t	\$200.00
2	349827	2019-06-15	t	\$200.00
3	349827	2019-06-14	t	\$200.00
4	349827	2019-06-13	t	\$200.00

```
In [8]: calendar.available.value_counts()
Out[8]: f      4069436
        t      2257109
        Name: available, dtype: int64
```

f (false) 代表已经被租用，t(true) 代表可以被出租

数据分析思路

3、数据案例分析思路步骤2

由于我们主要研究的目标是价格的依赖因素，所以我们采用价格和另一个因素的对比作图来观察。常见可视化招数包括：

招数1 . 柱状图，来观察数据的分布情况

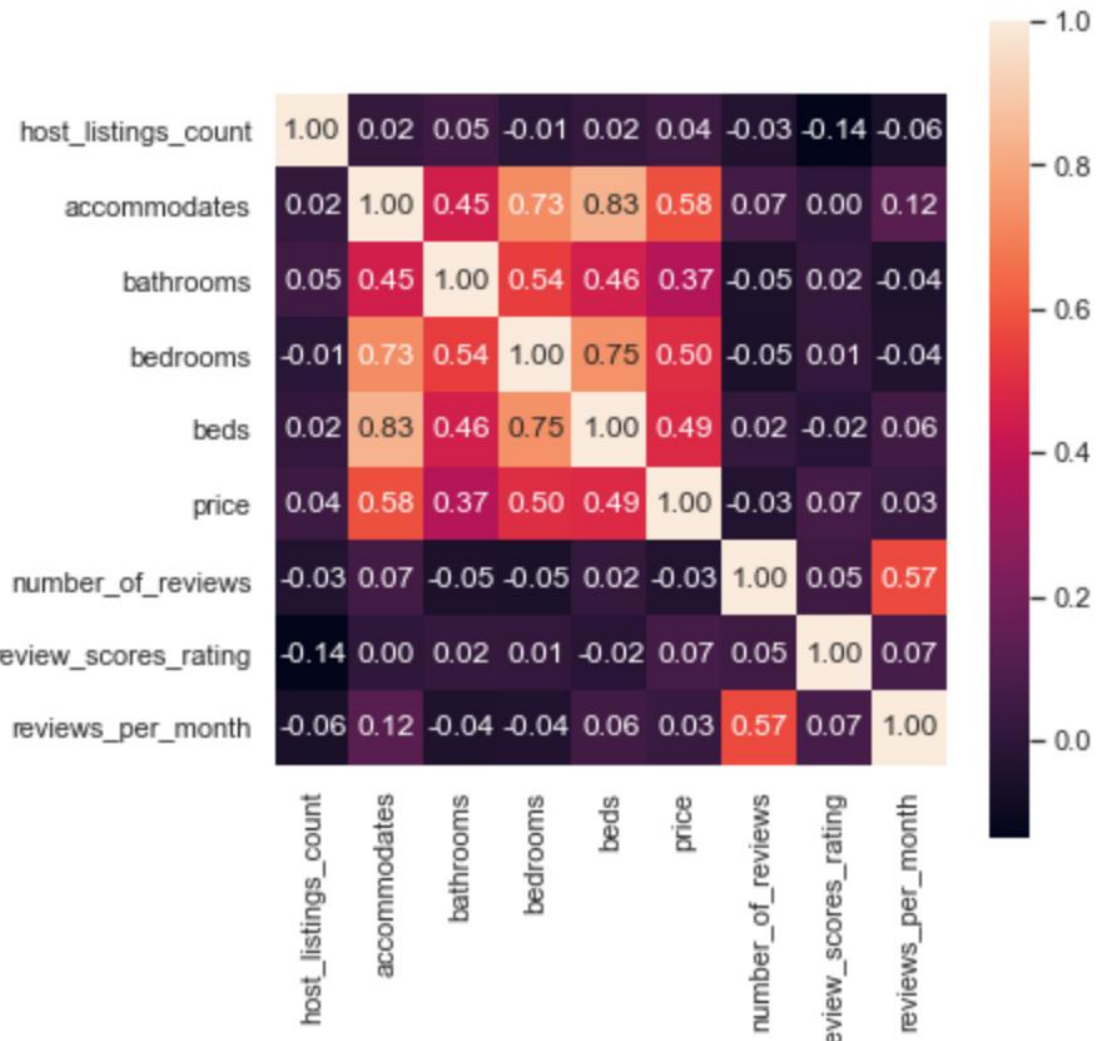
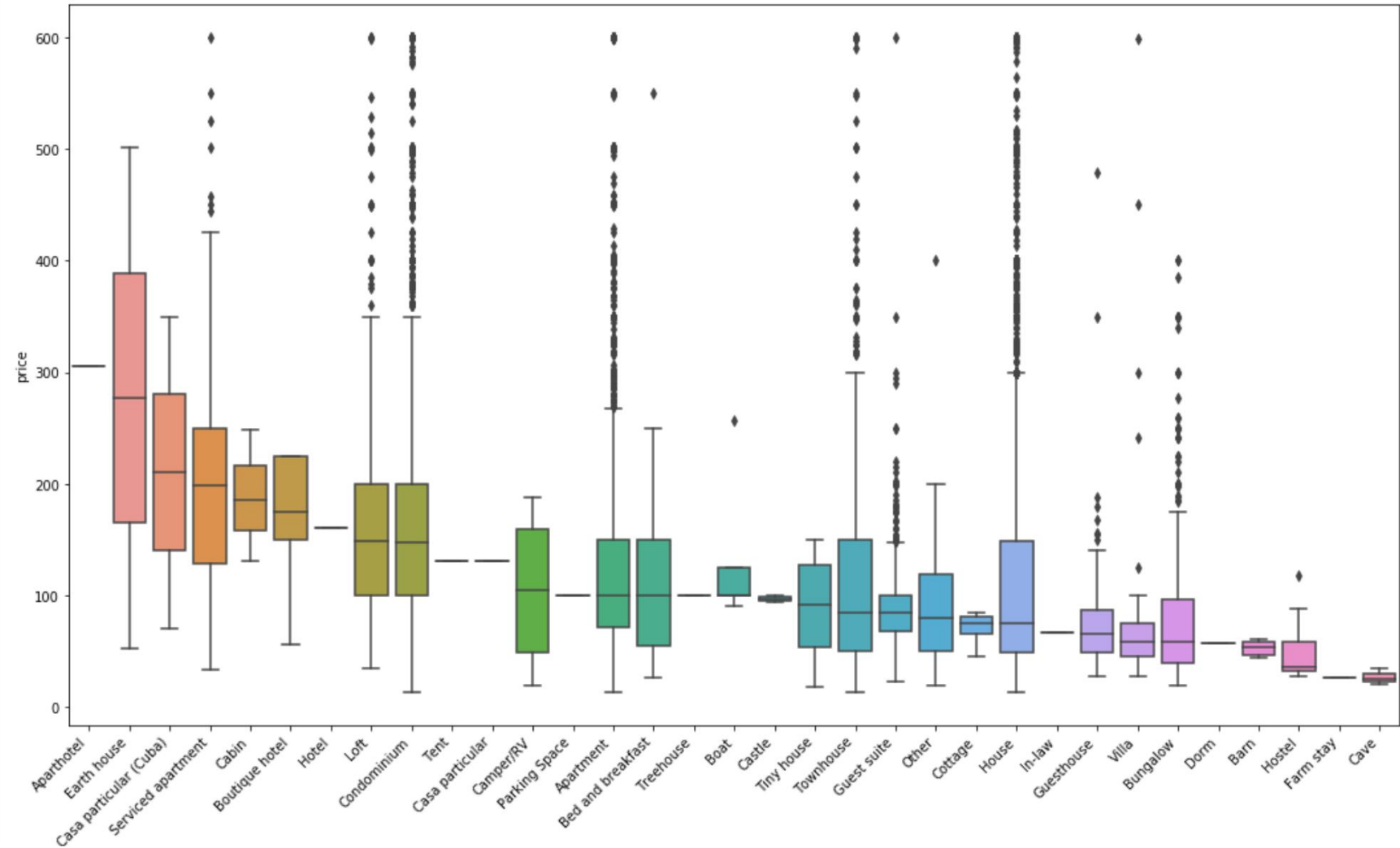
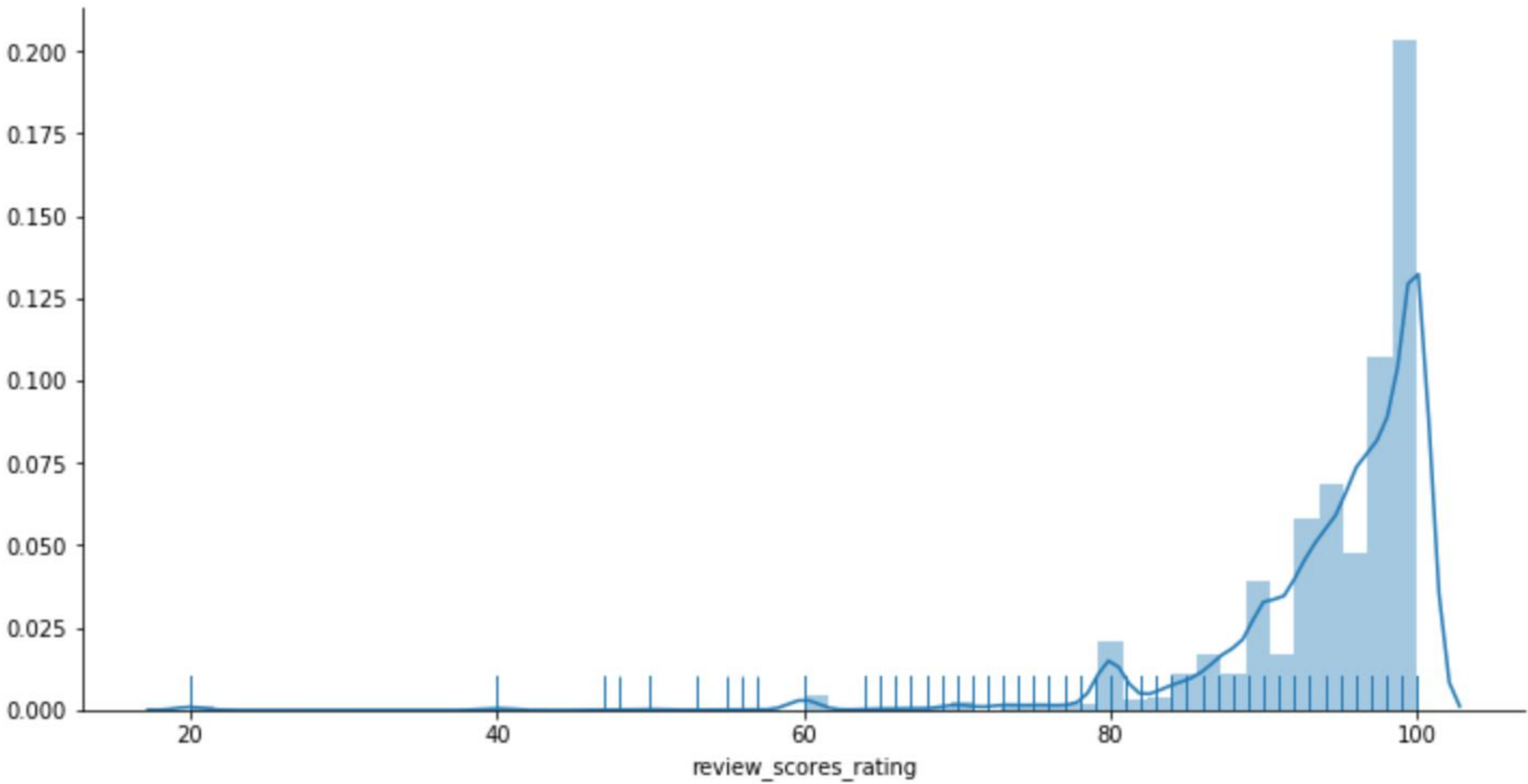
招数2 . 用箱型图，来观察数据的范围

招数3 . 用Pairplot，来观察不同因素间的相互联系

招数4 . 热力图，来快速筛选出高关联度的信息因素。

数据分析思路

3、数据案例分析思路步骤2部分截图



动手实战环节

实战1：数据的加载和基本ETL

实战2：利用数据可视化对数据集进行第一轮分析

谢谢观看