Hotels in Vienna Analysis

Maeva_Braeckevelt

20/12/2020

Introduction

This analysis aimed at analyzing the pattern of association between the price of the hotels in Vienna and the ratings. The data used was gathered in a csv files : The hotelbookingdata.csv. It was download from a comparison website and it was anonymized and slightly altered to ensure confidentiality.The main variables that I used were: the price in dollars (y) and the number of stars (x). My sample is the price of the hotels in Vienna in the weekday of November 2017.

Non-parametric regression

Bin scatter with four bins & Lowess regression
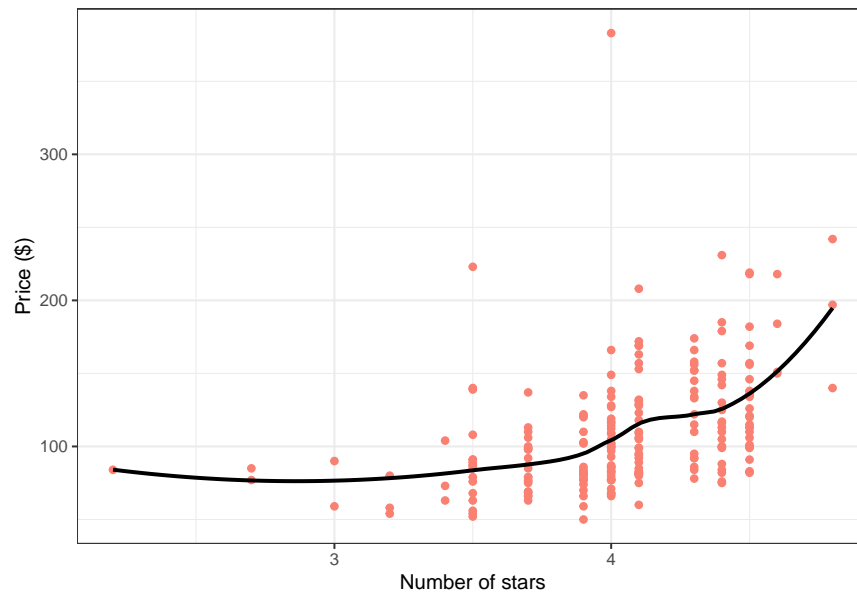
I observed that there is a positive slope in general. More stars the hotels have, more expensive are the hotels. If I look at the hotel with the worse rating and one with one of the best ratings, the price is around the double. So, I did uncover a pattern of association but I don't have any interpretable quantitave answer. I will do a simple linear regression.

Simple Linear regression (A5)
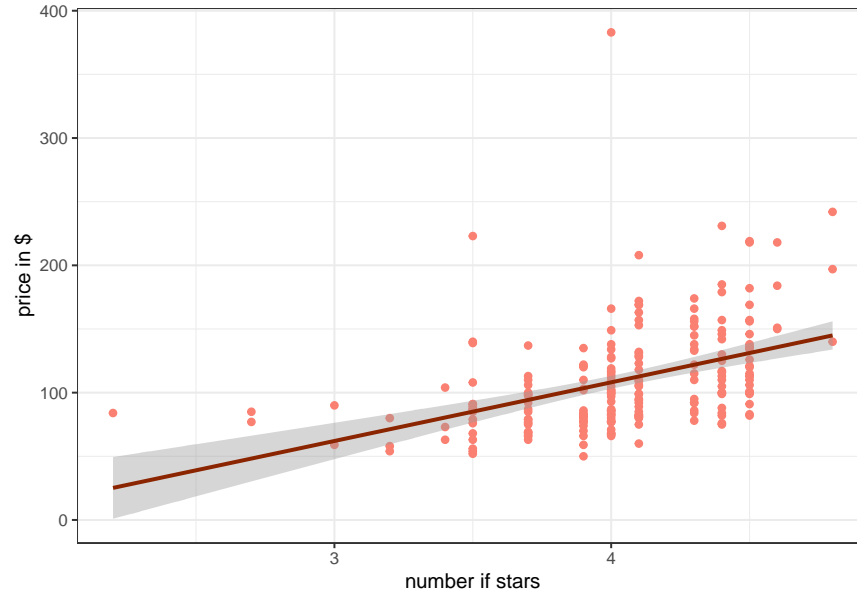
	Linear
(Intercept)	-76.16 **
	(27.60)
ratings	46.07 ***
	(6.88)
nobs	217
r.squared	0.19
adj.r.squared	0.18
statistic	44.86
p.value	0.00
df.residual	215.00
nobs.1	217.00
se_type	HC2.00

*** p < 0.001; ** p < 0.01; * p < 0.05.

This graph represents the simple linear regression between the price of hotels in Vienna and the number of price.

Formula : Price = -76,16 + 46,07 * stars

Alpha : -76,16 not a meaningful interpretation

Beta : the hotels that have 1 more stars are, on average, 46,07$ more expensive

However my R square is 19%, so only 19% of variation of the price is captured in this regression.

	Linear
(Intercept)	-76.16 **
	(27.60)
ratings	46.07 ***
	(6.88)
nobs	217
r.squared	0.19
adj.r.squared	0.18
statistic	44.86
p.value	0.00
df.residual	215.00
nobs.1	217.00
se_type	HC2.00

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.