Data Analysis : Assignment 2

Ghazal

- Filtered the data for **Paris**
- Selected Hotels and Apartments for accommodation type
- Price is less than 600
- Removed null values from *stars*
- Removed duplicates
- Removed Null values from *rating*, *stars*, and *distance*
- Created log of price, *lnprice*
- Created *highly_rated* if rating $>= 4$
- Distance is with in 2 miles

```
## fitting null model for pseudo-r2
## fitting null model for pseudo-r2
```

	Mean	SD	Min	Max	Median	P95	N
highly_rated	0.62	0.48	0.00	1.00	1.00	1.00	11397
distance2	0.69	0.46	0.00	1.00	1.00	1.00	11397
stars	3.22	0.78	1.00	5.00	3.00	4.00	11397

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	1.353**	6.310**		3.459**	
	(0.030)	(0.301)		(0.145)	
lspline(stars, c(3))1	−0.191**	−1.667**	−0.296**	−0.893**	−0.269**
	(0.010)	(0.102)	(0.019)	(0.049)	(0.014)
lspline(stars, c(3))2	−0.347**	−1.603**	−0.285**	−0.968**	−0.292**
	(0.007)	(0.042)	(0.010)	(0.025)	(0.006)
distance2	−0.078**	−0.439**	−0.077**	−0.246**	−0.074**
	(0.009)	(0.049)	(0.009)	(0.029)	(0.008)
Num.Obs.	11 397	11 397	11 397	11 397	11 397

* $p < 0.05$, ** $p < 0.01$

	Model 1	Model 2	Model 3
Constant	1.353**	6.310**	3.459**
	(0.030)	(0.301)	(0.145)
lspline(stars, c(3))1	−0.191**	−1.667**	−0.893**
	(0.010)	(0.102)	(0.049)
lspline(stars, c(3))2	−0.347**	−1.603**	−0.968**
	(0.007)	(0.042)	(0.025)
distance2	−0.078**	−0.439**	−0.246**
	(0.009)	(0.049)	(0.029)
Num.Obs.	11 397	11 397	11 397
R2	0.239		
PseudoR2		0.197	0.196

* $p < 0.05$, ** $p < 0.01$