**Solution Exercise Week 3**

**Exercise 1**

1. See python script
2. There is no multicollinearity so we can go ahead with our analyses
3. People on the west side have lower average charges than people on the east side
4. Both smoking and bmi can be used to explain the difference. People on the east side tend to smoke more and have higher bmi.

**Exercise 2**

To start, there is no multicollinearity between any of the relevant variables so we don’t have to remove any variables from the model. The only large correlation was between Entire home/apt and Private room, but if we use one of these as the reference category there is no issue.

If you did everything right you should get the following models presented in APA style.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model | |  |
|  | Coefficient | Standard error |  |
| Latitude | -8.20\*\*\* | 1.57 |  |
| Longitude | -20.8\*\*\* | 1.52 |  |
| Minimum number of nights | -3.67\*\*\* | .96 |  |
| Number of reviews | -9.42\*\*\* | 1.14 |  |
| Total number of listings by host | -2.48\* | .98 |  |
| Year-round availability | 22.46\*\*\* | 1.00 |  |
| Reviews per month | .21 | 1.15 |  |
| Host Listings\*Availability |  |  |  |
| *Manhattan (ref)* |  |  |  |
| Bronx | -33.37\*\*\* | 7.20 |  |
| Brooklyn | -50.23\*\*\* | 3.32 |  |
| Queens | -28.24\*\*\* | 4.77 |  |
| Staten Island | -175.74\*\*\* | 11.91 |  |
| *Private room (ref)* |  |  |  |
| Entire home or apartment | 100.71\*\*\* | 1.95 |  |
| Shared Room | -35.16\*\*\* | 6.53 |  |
| *Model Parameters* |  |  |  |
| R-Squared | 0.1181 |  |  |
| Adjusted R-Squared | 0.1178 |  |  |
| N | 38843 |  |  |

\*\*\* p<0.001 \*\*p<0.01 \*p<0.05

**Interpretation:**

We can see that all of the effects are negative except for availability (which is positive) and reviews per month (which is not significant). There are very large differences between the regions with Manhattan being clearly the most expensive. Private and shared rooms are also much cheaper than whole apartments. From the continuous variables, availability has by far the largest impact on price. So this means that listings which are available more days during the year are also more expensive.