Data Analytics Project 1 Data Handling Summary

Tyler Watson

1. Removed all duplicates
2. Changed all “Noord-Holland” variants under “State” to “Noord-Holland”
3. Standardized spelling of “Amsterdam” under “City”
4. Corrected other misspellings in “City” to standardize city names
5. Using Wikipedia, standardized neighborhood names. For cities other than Amsterdam, used city name in both “City” and “Neighborhood” fields since neighborhood names were not given
6. Added “host\_name\_id\_concatenate” column to combine “host\_id” and “host\_name”
7. Added “host\_since\_concatenate” column to combine “host\_since\_anniversary” and “host\_since\_year”
8. Added “days\_since\_host\_join” column that calculates days since join date for each member
9. Added “daily\_revenue” column to calculate amount listing earns per night using assumptions given in prompt
10. Added “revenue\_per\_booking” column that multiplies “daily\_revenue” by “minimum\_nights”
11. Added “total\_revenue” column that calculates total revenue earned by listing since joining (based on assumption that one review is equal to two stays at the listing)
12. Added “occupancy\_rate” column that calculates what percentage of time listing has been occupied since joining
13. Added “annualized\_revenue” column that uses “daily\_revenue,” “occupancy\_rate,” and 365 days a year to calculate annualized revenue since listing joined
14. Created Pivot Table 1 showing average of annualized\_revenue for every neighborhood
15. Using top 10 values, created Chart 1 of average of annualized\_revenue
16. Using INDEX/MATCHes, determined top 5 hosts by annualized\_revenue in every neighborhood (“Support Data 1”, “Support Data 2”)
17. Used information above to calculate percentage of total neighborhood annualized\_revenue earned by top 5 hosts in each neighborhood (“Support Data 2”)
18. Using that data, created Chart 2 showing which ten neighborhoods have the highest market share controlled by the top 5 hosts
19. Created Chart 3 showing percentage of market share controlled by top 5 hosts in 10 neighborhoods with highest average of annualized\_revenue (from Chart 1)
20. Created Chart 5 showing average of annualized\_revenue for each property\_type
21. Created Chart 6 showing occupancy\_rate for beds