Homework 1 – BBUS 301 100 points

Material covered: 1. Introduction to R programming
2. Data structures, variables, and data types
3. Descriptive Statistics in R

Follow the directions below to complete the assignment. Once you have completed the assignment, submit your answer on Canvas before the due date.

Submission Guidelines:

Due date: July 3, 2025 at 11:15 am

Submission means: You will submit an .R script containing all code, comments and explanations (remember to use # at the beginning of your comment) for each question in order to receive credit for your work.

Submission details: Students must make sure that their answers are readable, complete, and submitted before the deadline. Late submission for ANY reason, whether in part or in whole, or unreadable files will trigger the late penalties. The time when I receive the last part of your answers will count when computing late penalties, not the time when you sent the answers.

Late Penalties: If an assignment is turned immediately after the due date, your score will be reduced by 10% of the maximum assignment grade. Additional 10% will be deducted from your score for every 24 hours after that.

Academic Integrity: This is an individual assignment. Students need to submit their own answers and understand those answers. I reserve the right to ask you to walk me through the code for your assignment. If you fail to explain the code in detail and the choices you made, you will receive 0 points for the part of the assignment. Students who refuse to provide the explanation for their work will receive 0 points for the entire assignment.

Using code not discussed in class: Students who are submitting code not discussed in the class will temporary receive 0 points for that part of the assignment until they meet with me and they walk me through their code. Students need to meet with me within one week of returning the graded assignment, no later than August 19. If you fail to meet with me or if you cannot explain some of the code & the choices you made, you will keep the 0 points for that part of the assignment.

Homework Questions to Answer:

Dataset: The data for this homework, listings.csv, includes information about Airbnb listings in Seattle. It was scraped from the Airbnb website for public data-sharing and is available via Inside Airbnb, a publicly accessible data project (https://insideairbnb.com).

Variables:

name – Title of the listing
host_id – Unique ID of the host
neighborhood_group – Ballard, Magnolia, etc.
neighborhood – Smaller neighborhood name (Whittier Heights, Lawton Park, etc.)
room_type – Entire home/apt, Private room, etc.
price – Price per night (in USD)
minimum_nights – Minimum number of nights required to book
number_of_reviews – Number of reviews the listing has
availability_365 – Number of available nights in a year
license – do the host have a license (yes or no)

Part 1: Introduction to R programing (25 pts)

1. (2 pts) Create an R script called "hmw_1_name.R" (use your first name instead of name; don't use the quotation marks). This R script will contain all code and explanations (remember to use # at the beginning of your comment) for each question in this homework.
2. (3 pts) Use the RStudio keyboard shortcut Ctrl + Shift + R (or the Mac equivalent) to create a section header in your script. An "Insert Section" dialog box should appear, asking you to input the name of the section header—call it "Homework 1" and press OK.
3. (5 pts) Set your working directory to the folder where you downloaded the data for homework 1 (write the command even if your working directory is already set correctly).
4. (5 pts) Create three new variables: x, y and z. Assign the value 4200 to x, 24 to y, and 6 to z.
5. (10 pts) Working with the three variables created above, use R to determine the value of: $\frac{x^2 + y*z}{x + y + z}$. What is the result?

Part 2: Data Structures, variables and data types (35 pts)

6. You are planning to start an Airbnb business and want to project income.
    a. (2 pts) Create a variable called *expected_bookings* and assign it a value between 80 and 300 (choose your own value). (Hint: don't use a comma as a thousands-separator).
    b. (2 pts) Assume your nightly rate is $120. Create a variable called *rate* and assign it 120 (Hint: don't enter the dollar sign).
    c. (3 pts) Finally, create a variable *projected_income* as *expected_bookings* times *rate*. How much money do you project to make?
    d. (3 pts) Use R to check if *projected_income* is greater than 18000. What answer did you get? What logical operator did you use?
7. Suppose you track actual bookings per month: 5 in Jan, 8 in Feb, 12 in Mar, 10 in Apr, 7 in May, and no other bookings for the rest of the year(i.e.: 0 for Jun - Dec).
    a. (3 pts) Use the combine function in R to create a new variable called *monthly_bookings* that stores all this bookings data. The first number stored should be 5 since you booked 5 books in January, the second should be 8, and so on. Make sure that you record the 0s as well.
    b. (3 pts) Use an R function to find what type of vector is *monthly_bookings*. What did you find?
    c. (3 pts) Coerce the *monthly_bookings* vector into an integer vector.
    d. (3 pts) Since you are earning an exciting $120 per booking, you want to multiply each element in the *monthly_bookings* vector by 120, and store (assign) the result in a vector called *monthly_income*
    e. (3 pts) Use a logical operator on your *monthly_income* vector to check on which months you earned more than $1,000. In which months is the answer TRUE?
8. (5 pts) Create a data frame called *airbnb_df* that has the following two columns:
    a. The first column will be the *monthly_income* vector.
    b. The second column should be called *month*, and needs to include the names of all 12 months.
9. (5 pts) Use the str function to examine your data frame (*airbnb_df*). How many observations and how many variables you have in the data frame?

Part 3: Descriptive statistics in R (40 pts)

10. (5 pts) Import the *listings.csv* file from your downloaded folder, using the read.csv command. Remember to use the assign operator in order to create a new data frame called *airbnb_listings* that will contain the imported data.
11. Find the following descriptive statistics for the *price* variable.
    a. (2 pts) Mean. What is the value?
    b. (3 pts) Variance. What is the value?
12. (5 pts) What is the *price* value such that 75% of the prices lie below this value?
13. (5 pts) Use one descriptive statistics discussed in Module 3 to appropriately describe the variable *room_type*. What do you find? (hint: Start by determining which type of variable is the *room_type*: numerical or categorical? ; once you do that you can look at the handout for Module 3 and find an appropriate statistic). If you use more than one descriptive statistic, I will only grade the first one and disregard the other ones.
14. (10 pts) Create a multidimensional table of frequencies / counts based on the following three categorical variables: neighborhood, room_type and license. How many private rooms, without a license, are offered in Alki?
15. (10 pts) Analyze the relationship/ correlation between the *price* and the *number of reviews*. Does there appear to be an association? Explain.