**Master of Science in Applied Data Science (MS-ADS) Self-assessment Questions**

**A Note to Prospective Students in Data Science:**

Hello! Welcome, and congratulations on taking your first step towards this fast-paced and rewarding field. This self-assessment is designed to help you identify your level of preparedness for the Master of Science in Applied Data Science program. You will be able to get a good sense of your basic technical background and time management by completing this test. You should aim to answer all the questions in the assessment offline with limited effort in 1-4 hours.

The purpose of this self-assessment is only to highlight specific areas that may require extra preparation **before** beginning the program. Please do not be discouraged if you are unable to answer all the questions; this only indicates that you need to **refresh** and **prepare** those concepts and topics further. Additional information is embedded throughout the document, and we *strongly* encourage you to take advantage of the free resources on the Internet.

We wish you good luck!

# **General Math**

1. (8) = x What is x?
2. Write in logarithmic form.

More information on logarithms [here](https://www.khanacademy.org/math/algebra2/x2ec2f6f830c9fb89:logs/x2ec2f6f830c9fb89:log-intro/v/logarithms).

1. How many permutations can you make from the letters d,e,f?
2. How many permutations can you make from the letters a,b,c,d,e,f?
3. An ice cream parlor offers ten different toppings for their sundaes. How many different 3-topping sundae combinations (not allowing for double toppings) are there?

More information on permutations [here](https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:prob-comb/x9e81a4f98389efdf:combinatorics-precalc/v/permutation-formula).

1. If A = {3,4,5,6,7} and B = {2,3,4,5,6,7,8}, what is and ?

More information on sets [here](http://discrete.openmathbooks.org/dmoi2/sec_intro-sets.html).

1. What is the slope of the line that includes the points (-2, 2) and (-4, 8)?

More information on slopes [here](https://www.khanacademy.org/math/algebra/x2f8bb11595b61c86:linear-equations-graphs/x2f8bb11595b61c86:slope/v/introduction-to-slope).

1. True or False, ?

More information on the Zero Power Rule and exponents [here](https://www.khanacademy.org/math/cc-eighth-grade-math/cc-8th-numbers-operations/exponents-with-negative-bases/v/raising-a-number-to-the-0th-and-1st-power).

1. Simplify using the quotient rule ?

More information on the Quotient Rule for Exponents [here](https://www.youtube.com/watch?v=vJ7RbHPsDL0).

# **Calculus**

1. For x≠4, differentiate f(x)=

More information on basic differentiation [here](https://www.khanacademy.org/math/old-differential-calculus/basic-differentiation-dc).

1. Evaluate

More information on limits [here](https://www.khanacademy.org/math/ap-calculus-ab/ab-limits-new/ab-1-2/v/introduction-to-limits-hd).

1. Find the derivative of

More information on derivative concepts [here](https://www.khanacademy.org/math/ap-calculus-ab/ab-differentiation-1-new/ab-2-1/v/derivative-as-a-concept).

1. Evaluate the following indefinite integral:

More information on indefinite integrals [here](https://www.khanacademy.org/math/old-integral-calculus/indefinite-integrals#indefinite-integrals-intro-ic).

# **Probability**

1. If a fair coin was flipped 3 times and it landed on heads twice. What is the probability of this happening?

For more information please review [here](https://www.khanacademy.org/math/statistics-probability/probability-library/probability-sample-spaces/v/coin-flipping-example?modal=1).

1. 60% of swimmers do not wear goggles nor a swim cap. 20% wear goggles, and 30% wear swim caps. If one swimmer is chosen at random, what is the probability that he/she is wearing both goggles and a swim cap? It may be helpful to draw a Venn Diagram to visualize.

More information on probability, intersection, and union of sets [here](https://www.khanacademy.org/math/statistics-probability/probability-library).

1. If one card is drawn from a standard 52 card deck, what is the probability it will be a red (hearts or diamonds) face card (king, queen, or jack?)

For more information please review [here](https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:prob-comb/x9e81a4f98389efdf:addition-rule-prob-precalc/v/probability-with-playing-cards-and-venn-diagrams).

# **Linear Algebra**

1. Consider matrix A = What are the eigenvalues of A?

For more information please review [here](https://www.khanacademy.org/math/linear-algebra/alternate-bases/eigen-everything/v/linear-algebra-finding-eigenvectors-and-eigenspaces-example).

2. Consider matrix B= What is the transpose of B denoted by BT ?

For more information please review [here](https://www.khanacademy.org/math/linear-algebra/matrix-transformations/matrix-transpose/v/linear-algebra-transpose-of-a-matrix).

3. Consider matrices C= and D= What is the sum of C + D?

For more information please review [here](https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:matrices/x9e81a4f98389efdf:adding-and-subtracting-matrices/v/matrix-addition-and-subtraction-1).

1. Perform the matrix multiplication: x

For more information please review [here](https://www.khanacademy.org/math/precalculus/x9e81a4f98389efdf:matrices/x9e81a4f98389efdf:multiplying-matrices-by-matrices/v/multiplying-a-matrix-by-a-matrix).

1. If A = { 2, 4, 6, 8, 10, 12}. Which of the following are subsets of A?
2. U = {3, 5}
3. V = {2, 8, 10}
4. W = {0}
5. X = { 12, 10, 8, 6, 4, 2}
6. Y = {6, 10, 2}
7. Z = {4, 5, 6, 12}

For more information please review [here](https://www.mathgoodies.com/lessons/sets/subsets).

# **Fundamentals of Statistics**

1. Match the statistical concept to the correct definition:

|  |  |  |
| --- | --- | --- |
| **Concept** |  | **Definition** |
| Mean |  | a) Value that appears most often in a dataset. |
| Mode |  | b) Difference between the lowest and the highest value. |
| Range |  | c) Measurement of dispersion, calculated using the square root of the variance. |
| Median |  | d) Total of all values divided by the number of values. |
| Variance |  | e) Middle value in a list ordered from smallest to largest. |
| Standard Deviation |  | f) Measurement of the spread of values in a dataset, calculated by the average of the squared differences from the mean. |

For more information please review [here](https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/mean-and-median/v/mean-median-and-mode#:~:text=%2D%20Mode%2DThe%20most%20repetitive%20number,THE%20BIGGEST%20minus%20the%20Smallest!).

1. Find the median in the following list of numbers: 1, 3, 4, 5, 7, 8, 9, 12
2. A box of candy contains 6 solid chocolates, 4 chocolates with caramel filling, and 2 caramels. Draw a Venn Diagram to depict the different sets: Chocolate, Caramel + Chocolate, and Caramel. Then shade in A ∩ B on the Venn Diagram, where A is solid Chocolate, and B is Caramel.

For more information please review [here](https://www.khanacademy.org/math/cc-eighth-grade-math/cc-8th-data/two-way-tables/v/two-way-frequency-tables-and-venn-diagrams).

1. If A and B are independent, and P(A) = 0.3 and P(B) = 0.5, please find P(A ∪ B).

For more information please review [here](https://www.khanacademy.org/math/statistics-probability/probability-library/basic-set-ops/v/intersection-and-union-of-sets).

1. Please match the type of variable with the correct definition.

|  |  |  |
| --- | --- | --- |
| **Variable Name** |  | **Definition** |
| Categorical |  | a) Numerical variables that represent a measurement of quantity. |
| Ordinal |  | b) Variables with a numerical value and can be measured along a continuum. |
| Discrete |  | c) Interval variable that has a meaningful zero. |
| Continuous |  | d) Variables that have at least two categories without intrinsic order. |
| Binary |  | e) Numeric variables that take any value in an infinite range. |
| Ratio |  | g) Numeric variables that have a finite number of values. |
| Interval |  | h) A variable that can be put into categories, also known as a nominal variable. |
| Nominal |  | i) Variables that can be ranked. |
| Qualitative |  | j) Variable that is non-numerical with data that fits into categories. |
| Quantitative |  | f) Nominal variables that only have two categories or levels. Also referred to as dichotomous. |

For more information please review [here](https://www.statisticshowto.com/probability-and-statistics/types-of-variables/).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cafeteria Menu** | | | | |
| **Entrée** | **Type** | **Total Calories** | **Protein (g)** | **Sugar (g)** |
| Turkey Sandwich | Cold | 400 | 25 | 5 |
| Spaghetti and Meatballs | Hot | 750 | 20 | 15 |
| Cesar Salad | Cold | 350 | 5 | 7 |
| Grilled Cheese | Hot | 625 | 15 | 10 |

1. Please refer to the table above. What are the individuals in this dataset?
2. The Cafeteria’s customers
3. Entrée
4. Type
5. Menu
6. Please refer to the table above. In the *Cafeteria Menu* portion of the table, how many variables are there, and of those variables, how many are categorical?
7. 5 variables, 2 categorical
8. 5 variables, 1 categorical
9. 4 variables, 1 categorical
10. 4 variables, 0 categorical

For more information please review [here](https://www.khanacademy.org/math/statistics-probability/analyzing-categorical-data/one-categorical-variable/v/identifying-individuals-variables-and-categorical-variables-in-a-data-set?modal=1).

1. Give an example of univariate and bivariate observations? Are multivariate and bivariate the same?

For more information please review [here](https://www.geeksforgeeks.org/univariate-bivariate-and-multivariate-data-and-its-analysis/).

# **Excel**

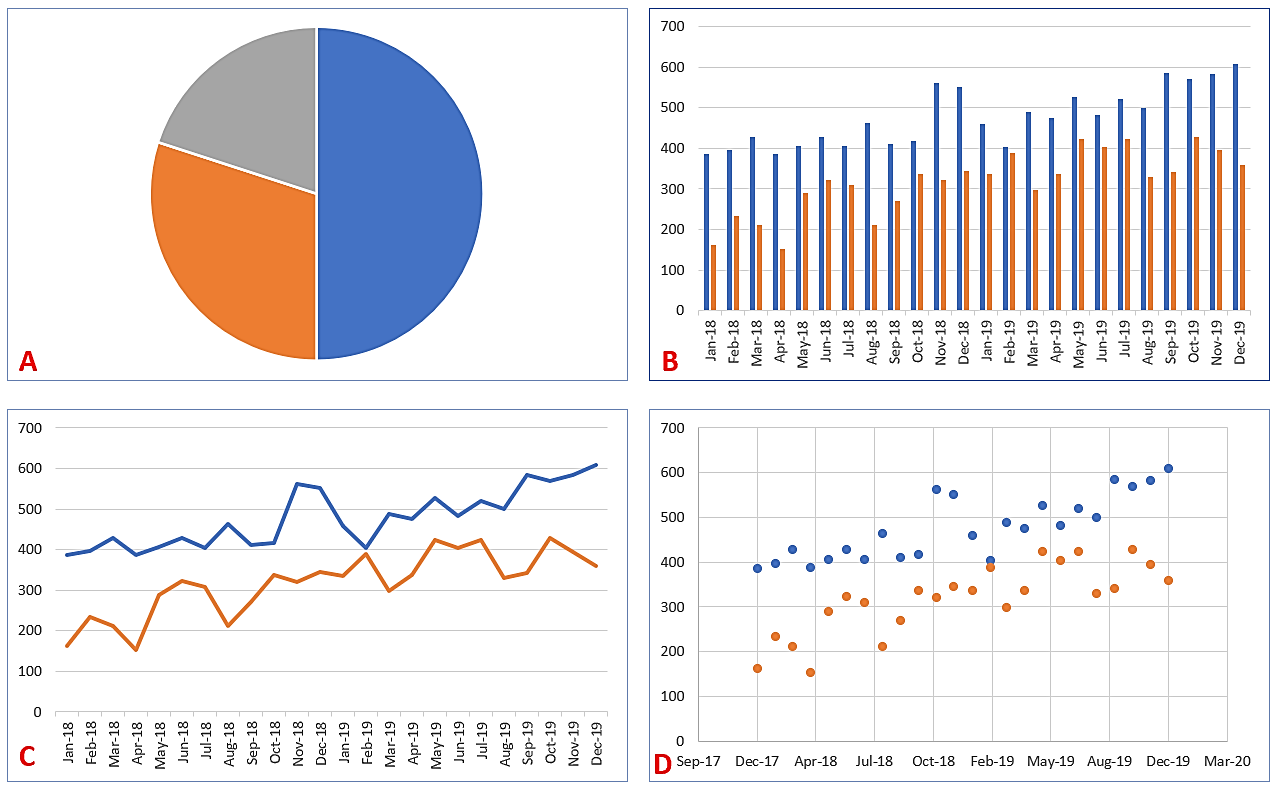
You have been given access to an Excel file containing grant funding applications. The table has 5 columns:

* Application\_id: grant application ID
* Budget\_start: date when the grant payments start
* project\_title: title appearing on the grant application
* Support\_year: number of years the grant will remain active
* Total\_cost: the total dollar amount awarded by the grant

\*Please copy and paste the following dataset into Excel or download data as a Microsoft Excel file from the website.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| application\_id | budget\_start | project\_title | support\_year | total\_cost |
| 9223315 | 2/5/2019 | Dietary factors, biomarkers, metabolic pathways related to non-alcoholic fatty liver disease | 1 | 249000 |
| 9334591 | 12/1/2018 | P2 Study of Clofazimine for Pulmonary M. avium Disease IND 125,266 (1/5/2015) | 2 | 495835 |
| 9485561 | 2/1/2019 | Multidisciplinary Training in Digestive Diseases | 44 | 372638 |
| 9489206 | 12/1/2018 | Early-Stage Chronic Kidney Disease in HIV-infected Individuals | 9 | 625114 |
| 9491549 | 3/1/2019 | Mechanisms of varicella virus-induced multisystem disease using a primate model | 11 |  |
| 9554875 | 8/1/2019 | Probiotics for Prevention of Acute Graft-vs-Host Disease in Children with Cancer | 4 | 744609 |
| 9561532 | 10/1/2018 | The neuropathology of mild traumatic brain injury in Alzheimer disease | 5 |  |
| 9597224 | 12/1/2018 | Hormones in allergic disease | 3 | $ 392,487.00 |
| 9600088 | 12/1/2018 | Molecular Genetic Insight into Neurodegenerative Disease from Drosophila | 3 | 703113 |
| 9601647 | 12/1/2018 | Mechanisms of IL-33 secretion in allergic diseases | 3 | 510759 |
| 9605113 | 12/1/2018 | Non-invasive Intracranial Pressure Sensing with Near Infrared Light for Monitoring the Healthy and Diseased Brain | 2 | 186255 |
| 9605217 | 10/1/2018 | Regulation of extrasynaptic GABA-A receptors in health and disease | 4 |  |
| 9605233 | 10/1/2018 | Dietary Fat and Alcoholic Liver Disease | 3 |  |
| 9606377 | 12/1/2018 | The role of IL10 in early vascularity, endothelial dysfunction, and programming of adult cardiovascular disease | 2 | 178524 |
| 9608053 | 12/1/2018 | Tip60, Ischemic Heart Disease and Regeneration | 3 | 501446 |
| 9608768 | 1/1/2019 | Translational Program in CFTR-Related Airway Diseases | 3 | 917966 |
| 9608791 | 12/1/2018 | VITAMIN D DEFICIENCY, INSULIN RESISTANCE AND CARDIOVASCULAR DISEASE | 8 | 381900 |
| 9608792 | 2/1/2019 | IU training Program in Molecular Physiology and Clinical Mechanisms of Lung Disease | 10 | 245539 |
| 9609465 | 12/1/2018 | Academic Training Program in Pediatric Pulmonary Disease | 30 | 221162 |
| 9610656 | 1/1/2019 | Molecular Mechanisms of Pigmentation in Health and Disease | 2 | 342000 |
| 9610698 | 12/1/2018 | Non-Contrast-Enhanced Velocity-Selective MR Angiography at 3T for Cerebrovascular Diseases | 2 | 404000 |
| 9085817 | 9/1/2016 | STorytelling to Improve DiseasE outcomes in GoUT: The STRIDE-GO Study | 1 |  |
| 9565043 | 1/1/2019 | New approaches to combat CNS inflammation in Veterans: Targeting a metabolic enzyme in demyelinating disease | 1 |  |
| 9588799 | 2/1/2019 | Vascular-associated neuroinflammation in Alzheimer’s disease: differential effects on disease progression modulated by underlying amyloid burden | 2 | 61610 |
| 9591280 | 1/1/2019 | Krüppel-Like Factor 15 is a novel mediator of glucocorticoid-responsive glomerular disease | 2 |  |
| 9598331 | 12/1/2018 | Macrophage differentiation and disease outcome in influenza infection | 3 | 542926 |
| 9613207 | 1/1/2019 | Multivalent Toxoid Vaccine for Prevention of S. aureus Invasive Diseases | 5 | 761776 |
| 9613228 | 1/1/2019 | Identifying a nodal point for G alpha q signaling in eye disease | 4 | 330612 |
| 9613769 | 1/1/2019 | Causal Inference in Infectious Disease Prevention Studies | 9 | 378114 |
| 9613856 | 1/1/2019 | Symptom burden in End-Stage Liver Disease patient-caregiver dyads | 4 | 385000 |
| 9616316 | 1/1/2019 | Fetal Brain Development in Congenital Heart Disease | 3 | 189954 |
| 9617236 | 12/1/2018 | Integrative genomic, epigenetic and functional studies in diabetic kidney disease | 4 | 725688 |
| 9617699 | 1/1/2019 | Disease-Induced Modification in Nanoparticle-Corona Identity and Toxicity | 5 | 249000 |
| 9617787 | 12/1/2018 | microRNAs in vascular diseases | 8 | 425953 |

1. In the Excel file, add filters to the table and sort support\_year from smallest to largest.
2. Format the budget\_start column as YYYY-MM-DD.
3. Freeze the top row, so the column names remain visible while you scroll down the rows.
4. You are asked to create a new column next to “total\_cost” called “priority,” labeling all projects with a “total\_cost” over $20,000 as “high” and anything under $20,000 as “low.” Write an IF statement to quickly accomplish this task.
5. Excel provides quick graphing of data by simply clicking “Insert” and selecting the chart you wish to create. Can you name these different graphs?
6. \_\_\_\_\_\_\_\_\_\_\_\_
7. \_\_\_\_\_\_\_\_\_\_\_\_
8. \_\_\_\_\_\_\_\_\_\_\_\_
9. \_\_\_\_\_\_\_\_\_\_\_\_



# **APA Writing Style**

You will use APA 7 style, a writing format for academic documents, in your reports and presentations in the MS-ADS program.

1. Which of the following is cited correctly in APA in-text citations of an article with three or more authors?

1. Tarshizi, Cooke, Smith, Kim (2020)
2. Tarshizi, et. al. 2020
3. (Tarshizi et al., 2020, p. 155)
4. Tarshizi…& Kim (2020, pp. 155)

2. In APA 7 style, a separate title page is required for a document or report.

1. True
2. False

3. Running heads are optional on all APA 7th edition papers.

a) True

b) False

Purdue has an excellent APA 7th Edition quick [reference guide](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_style_introduction.html).

# **Programming**

1. What is the value of x after the following code snippet is executed?

|  |
| --- |
| n = 5  m = 8  l = 4  x = 0  if (m > l) and (n > m) then      x = 5  else      if (m >= 10) then          x = 6      else          x = 7      end if  end if |

1. x = 0
2. x = 5
3. x = 6
4. x = 7

For more information please review [here](https://www.geeksforgeeks.org/python-if-else/).

1. Given the array below, what is the value of “sum” at the end of the execution of the following piece of code: (assume that array index starts at 0 and “loop” increments variable “i”)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 5 | 2 | 7 | 1 | 2 | 6 |

array =

|  |
| --- |
| sum = 0  loop i = 0 to 6      if (i == 0 or i == 3) then          sum = sum + array[i]      end if  end loop |

1. sum = 0
2. sum = 5
3. sum = 6
4. sum = 11

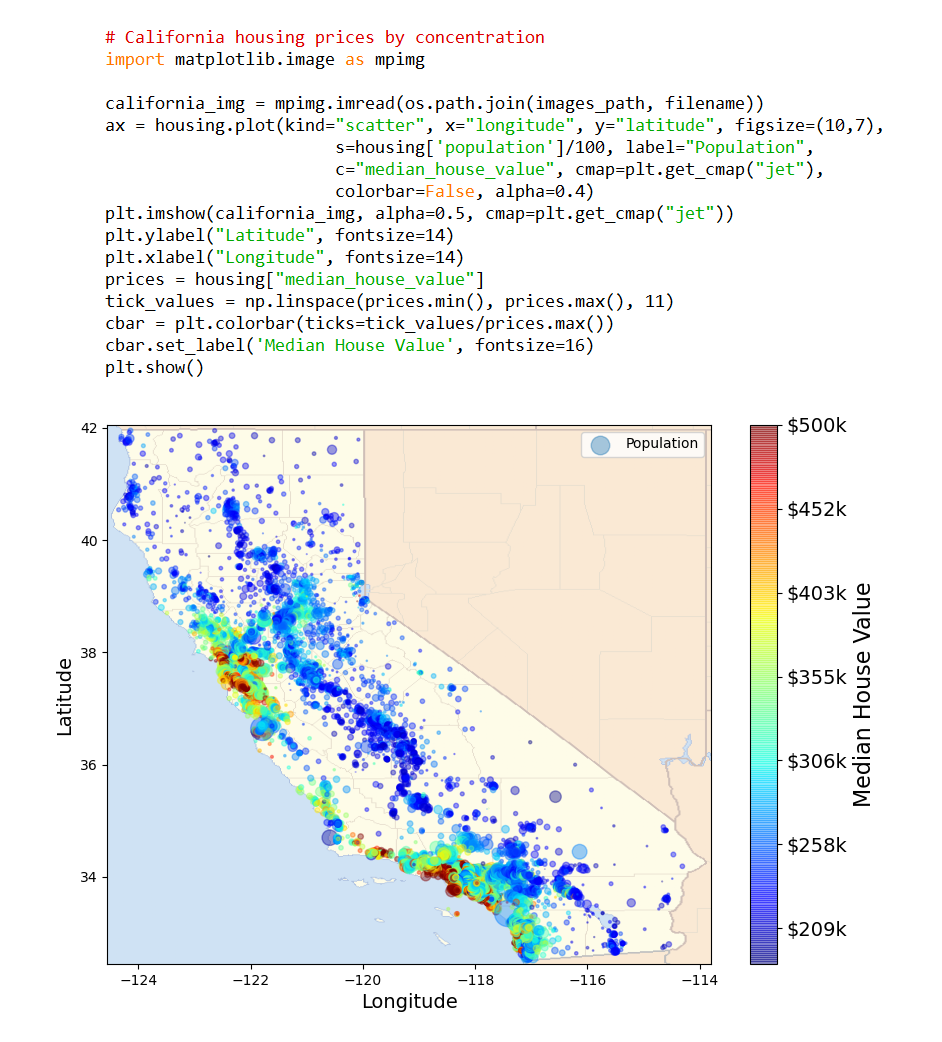
For more information please review [here](https://www.geeksforgeeks.org/python-arrays/).

1. The purpose of this part is to ***review*** sample Python code (as examples) and consider if you would *enjoy* learning and writing similar computer code (programming) in Data Science. In addition, do you know what the name of the plot created in example #3 is?
2. Distribution Plots
3. Scatter Plot Matrix
4. Heatmap
5. Sankey Plots

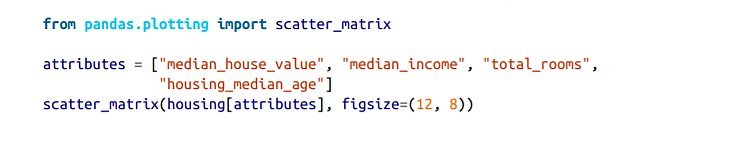
**Example #1:**

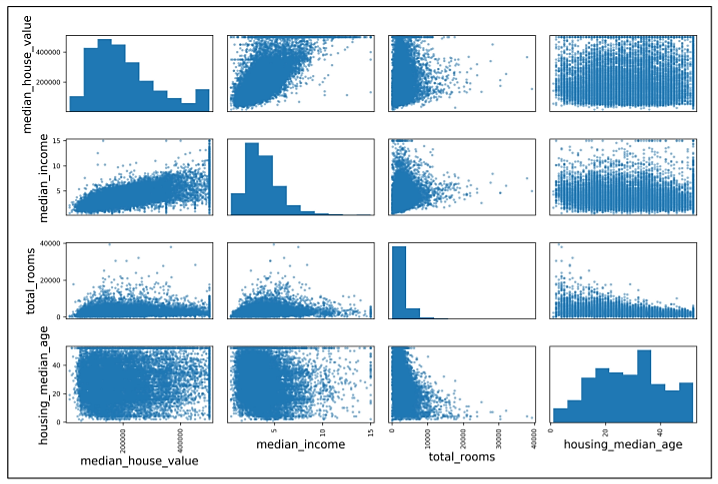


**Example #2:**



**Example #3:**





# **Online Program Readiness Short Evaluation**

There is no “right” or “wrong” answer to these questions. Respond honestly.

1. I am self-motivated and self-disciplined in the online learning environment, and I can allocate appropriate weekly time to obtain a master’s degree in data science.
2. Yes
3. No
4. I’m not sure
5. I am comfortable working and learning independently, and I can maintain a high motivation during the master’s program.
6. Yes
7. No
8. I’m not sure
9. I am good at setting goals and deadlines for myself to learn online technical graduate courses. I usually put a schedule and keep to it. I can turn in assignments and tasks on time without reminders.
10. Yes
11. No
12. I’m not sure
13. I like working in teams and virtual teamwork project settings, and I am responsive to teammates and very comfortable with online communications.
14. Yes
15. No
16. I’m not sure
17. I prefer learning about topics by having them explained rather than *reading* about them. I need to listen to face-to-face lectures on the concepts. I am also better at following oral instructions than written instructions.
18. Yes
19. No
20. I’m not sure

Which option do you prefer (a or b)?

1. I usually need a direct explanation from my professors and face-to-face interaction with my classmates to thoroughly grasp the content. I prefer face-to-face lectures (synchronous or hybrid learning) and meeting my classmates in-person to perform a team project or ask questions regarding assignments, programming, etc.
2. I am an independent learner. I am comfortable learning on my own through reading the assigned textbooks, watching videos, participating in discussions, performing quizzes and assignments, and collaborating with my peers in hands-on projects using a Learning Management System (such as Blackboard). I take responsibility for my learning process and have no issue contacting my professors and classmates via e-mail or any other online tool if I have questions. When faced with difficulties or challenges in different graduate courses, I do not give up or quit. I use my problem-solving and research skills to find a solution.

# **Time Estimation & Management**

Simply use the table below to estimate your time availability to allocate for the MS-ADS program weekly to perform readings, assignments, discussions, quizzes, and exams/projects.

|  |  |
| --- | --- |
| **Days** | **Hours** |
| Tuesday |  |
| Wednesday |  |
| Thursday |  |
| Friday |  |
| Saturday |  |
| Sunday |  |
| Monday (module deadline) |  |
| **Total** |  |