

Master of Science in Applied Data Science (MS-ADS) Self-assessment (Answer Key)

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.

General Math

1. $\log_2(8) = x$ What is x ?

$$x=3 \quad (2^3 = 8)$$

2. Write $15^{3/4} = 8$ in logarithmic form.

$$\log_{15} 8 = \frac{3}{4}$$

More information on logarithms [here](#).

3. How many permutations can you make from the letters d,e,f?

6 permutations def, dfe, edf, efd, fde, fed

4. How many permutations can you make from the letters a,b,c,d,e,f?

720 permutations $(6*5*4*3*2*1 = 720)$

5. 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?

$$\binom{10}{3} = 120 \text{ sundae combinations}$$

More information on permutations [here](#).

6. If $A = \{3,4,5,6,7\}$ and $B = \{2,3,4,5,6,7,8\}$, what is $A \cap B$ and $A \cup B$?

$$A \cap B = \{3, 4, 5, 6, 7\}$$

$$A \cup B = \{2, 3, 4, 5, 6, 7, 8\}$$

More information on sets [here](#).

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

$$m = \frac{8 - 2}{-4 - (-2)} = \frac{6}{-2} = -3$$

More information on slopes [here](#).

8. True or False, $a^0 = 1$?

True

More information on the Zero Power Rule and exponents [here](#).

9. Simplify using the quotient rule $\frac{x^4}{x^9} = ?$

$$\frac{1}{x^5}$$

More information on the Quotient Rule for Exponents [here](#).

Calculus

1. For $x=4$, differentiate $f(x) = \frac{x^2}{4-x}$

$$f(x) = \frac{8x - x^2}{(4-x)^2}$$

More information on basic differentiation [here](#).

2. Evaluate $\lim_{x \rightarrow 8} \frac{2x^2 - 17x + 8}{8-x}$

$$\lim_{x \rightarrow 8} \frac{2x^2 - 17x + 8}{8-x} = \lim_{x \rightarrow 8} \frac{(2x-1)(x-8)}{-(x-8)} = \lim_{x \rightarrow 8} \frac{2x-1}{-1} = -15$$

More information on limits [here](#).

3. Find the derivative of $f(y) = \frac{4y^3 - 7y + 8}{y}$

$$f(y) = \frac{4y^3}{y} - \frac{7y}{y} + \frac{8}{y} = 4y^2 - 7 + 8y^{-1} \quad f'(y) = 8y - 8y^{-2}$$

More information on derivative concepts [here](#).

4. Evaluate the following indefinite integral: $\int 6x^5 dx - 12x^2 + 8$

$$x^6 + c - 12x^2 + 8$$

More information on indefinite integrals [here](#).

Probability

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

$$p(C) = p(H,T,H) + p(T,H,H) + p(H,H,T) = 3 \times \frac{1}{8} = 37.5\%$$

For more information please review [here](#).

2. 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.

$$P(G \text{ or } S) = 1 - 0.60 = 0.40$$

$$P(G \text{ or } S) = P(G) + P(S) - P(G \text{ and } S)$$

$$0.40 = 0.20 + 0.30 - P(G \text{ and } S)$$

$$= P(G \text{ and } S) = 0.10$$

More information on probability, intersection, and union of sets [here](#).

3. 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?)

$$\text{Face Card} = 12/52 = 0.231 \text{ (red card)} = 26/52 = 0.5 \text{ (red face card)} =$$

$$\text{Probability of Face Card} \times \text{Probability of Red Card} = 0.231 \times 0.5 = 0.116$$

For more information please review [here](#).

Linear Algebra

1. Consider matrix $A = \begin{bmatrix} -4 & 0 \\ 0 & 1 \end{bmatrix}$ What are the eigenvalues of A?

$$-4, 1$$

For more information please review [here](#).

2. Consider matrix $B = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 6 & 7 \\ 4 & 5 & 9 \end{bmatrix}$ What is the transpose of B denoted by B^T ?

$$\begin{bmatrix} 2 & 3 & 4 \\ 3 & 6 & 5 \\ 4 & 7 & 9 \end{bmatrix}$$

For more information please review [here](#).

3. Consider matrices $C = \begin{bmatrix} 2 & 1 & 3 \\ 0 & 8 & 0 \\ 1 & 5 & 3 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 & 7 \\ 3 & 4 & 9 \\ 0 & 8 & 0 \end{bmatrix}$ What is the sum of $C + D$?

$$\begin{bmatrix} 3 & 1 & 10 \\ 3 & 12 & 9 \\ 1 & 13 & 3 \end{bmatrix}$$

For more information please review [here](#).

4. Perform the matrix multiplication: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} -2 & 3 \\ 4 & 5 \end{bmatrix} = i$

$$\begin{bmatrix} 6 & 13 \\ 10 & 29 \end{bmatrix}$$

For more information please review [here](#).

5. If $A = \{2, 4, 6, 8, 10, 12\}$. Which of the following are subsets of A ?

a. $U = \{3, 5\}$

b. $V = \{2, 8, 10\}$

c. $W = \{0\}$

d. $X = \{12, 10, 8, 6, 4, 2\}$

e. $Y = \{6, 10, 2\}$

f. $Z = \{4, 5, 6, 12\}$

For more information please review [here](#).

6. What is the L2 vector norm of vector $A = [10, 2, 8, 5, 4]$?

$$\sqrt{10^2 + 2^2 + 8^2 + 5^2 + 4^2} = 14.036$$

For more information please review [here](#).

Fundamentals of Statistics

1. Match the statistical concept to the correct definition:

Concept	Definition
---------	------------

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

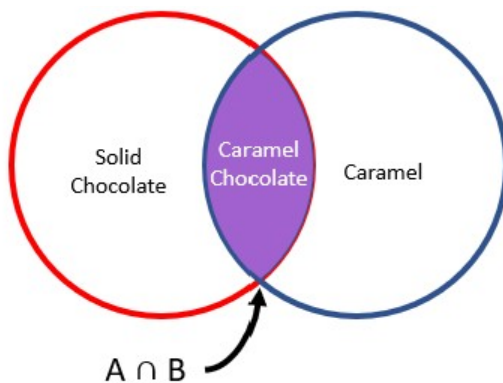
For more information please review [here](#).

2. Find the median in the following list of numbers: 1, 3, 4, 5, 7, 8, 9, 12

$$(5 + 7) / 2 = 6$$

$$\text{Median} = 6$$

3. 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 $A \cap B$ on the Venn Diagram, where A is solid Chocolate, and B is Caramel.



For more information please review [here](#).

4. If A and B are independent, and $P(A) = 0.3$ and $P(B) = 0.5$, please find $P(A \cup B)$.

$$P(A \cup B) = 0.30 + 0.50 - 0.30 \times 0.50 = 0.65$$

For more information please review [here](#).

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

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

For more information please review [here](#).

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

6. Please refer to the table above. What are the individuals in this dataset?

- a) The Cafeteria's customers
- b) Entrée**
- c) Type
- d) Menu

7. 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?

- a) 5 variables, 2 categorical
- b) 5 variables, 1 categorical

c) 4 variables, 1 categorical

d) 4 variables, 0 categorical

For more information please review [here](#).

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

Univariate example: heartbeat measured by beats per minute (1 variable)

Bivariate example: systolic and diastolic blood pressure measurement (2 variables)

Multivariate example: 2 or more observations. Multivariate can be bivariate.


For more information please review [here](#).

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.

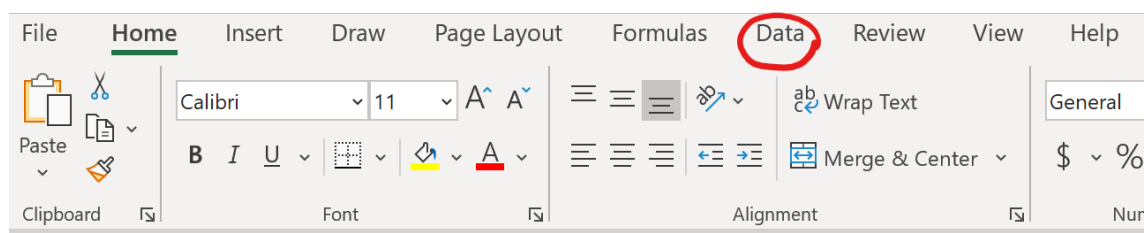
application_id	budget_start	project_title	support_year	total_cost
9223315	2/5/2019	Dietary factors, biomarkers, metabolic pathways related to non-alcoholic fatty liver disease	1	 249000
9334591	12/1/2018	P2 Study of Clofazimine for Pulmonary M. avium Disease IND 125,266 (1/5/2015)	2	495835
9485561	2/1/2019	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

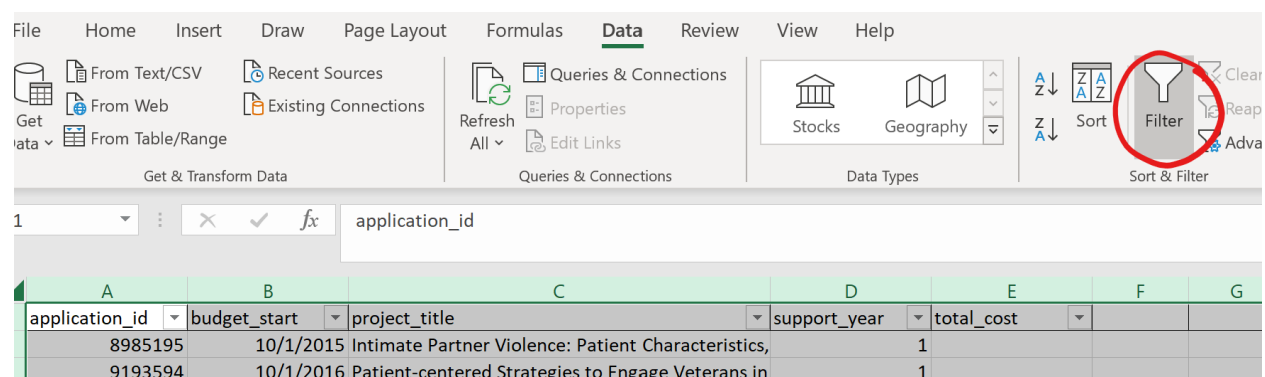
9613856	1/1/2019	Symptom burden in End-Stage Liver Disease patient-caregiver dyads	4	385000
9616316	1/1/2019	Fetal Brain Development in Congenital Heart Disease	3	189954
9617236	12/1/2018	Integrative genomic, epigenetic and functional studies in diabetic kidney disease	4	725688
9617699	1/1/2019	Disease-Induced 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.

• Click on data.



• Click on filter.



• Click on the dropdown menu for support_year, then sort by smallest to largest.

	A	B	C	D	E
1	application_id	budget_start	project_title	support_year	total_cost
2	8985195	10/1/2015	Intimate Partner Violence: Patient Characteristics,		
3	9193594	10/1/2016	Patient-centered Strategies to Engage Veterans in		
4	9204638	10/1/2016	Optimizing Function and Independence QUERI		
5	9223315	2/5/2019	Dietary factors, biomarkers, metabolic pathways related to non-alcoholic fatty liver disease		249000
6	9284011	7/1/2017	Determining and targeting reasons for low statin use		
7	9284247	5/1/2017	Promoting Advance Care Planning as a Healthy Behavior		
8	9397857	1/1/2019	Improving Sleep as a Strategy to Reduce Suicide Risk		
9	9448740	6/1/2019	Intelligent Wearable Analyzer for Vapor Exposure		513191

2. Format the budget_start column as YYYY-MM-DD.

- Select the budget_start column

	A	B	C	D	E
1	application_id	budget_start	project_title	support_year	total
2	8985195	10/1/2015	Intimate Partner Violence: Patient Characteristics,	1	
3	9193594	10/1/2016	Patient-centered Strategies to Engage Veterans in	1	
4	9204638	10/1/2016	Optimizing Function and Independence QUERI	1	
5	9223315	2/5/2019	Dietary factors, biomarkers, metabolic pathways related to non-alcoholic fatty liver disease	1	
6	9284011	7/1/2017	Determining and targeting reasons for low statin use	1	
7	9284247	5/1/2017	Promoting Advance Care Planning as a Healthy Behavior	1	
8	9397857	1/1/2019	Improving Sleep as a Strategy to Reduce Suicide Risk	1	
9	9448740	6/1/2019	Intelligent Wearable Analyzer for Vapor Exposure	1	
10	9505197	6/1/2019	Noninvasive monitoring and evaluation of anti-tuberculosis	1	
11	9524130	9/30/2019	Transformation of Paraplegic Paralysis to Overgrowth	1	
12	9525527	2/15/2019	Multi-method signature of risk for compulsive substance use	1	
13	9552989	10/1/2018	Purinergic control of calcium flux in podocytes	1	
14	9575621	6/1/2019	Advanced recombinase-based gene expression technology	1	
15	9605870	1/1/2019	Improving patient safety by promoting guideline-concordant care	1	
16	9607386	10/1/2018	Optimizing Outcomes for Older Veterans with Chronic Pain	1	
17	9607763	3/1/2019	Evaluating Care Coordination for Community Hospital Patients	1	
18	9608243	1/1/2019	Project 1-Cohort and Ethical Analysis of Patients using a Digital Health Platform	1	
19	9608244	1/1/2019	Project 2-Optimization of Post-Transplant care via a Digital Health Platform	1	
20	9608274	6/6/2019	Measuring out-of-pocket problem drinking: An intervention for	1	

- Right-click and select format cells.

id	budget_start	project_title
8985195	10/1	
9193594	10/1	
9204638	10/1/2016	Optimizing Function and
9223315	2/5	
9284011	7/1	
9284247	5/1	
9397857	1/1	
9448740	6/1	
9505197	6/1	
9524130	9/30	
9525527	2/15	
9552989	10/1	
9575621	6/1	
9605870	1/1	
9607386	10/1	
9607763	3/1	
9608243	1/1	
9608244	1/1	
9608374	6/6	
9612421	2/4	
8978569	10/1	

- Select "date."

Format Cells ? X

Number Alignment Font Border Fill Protection

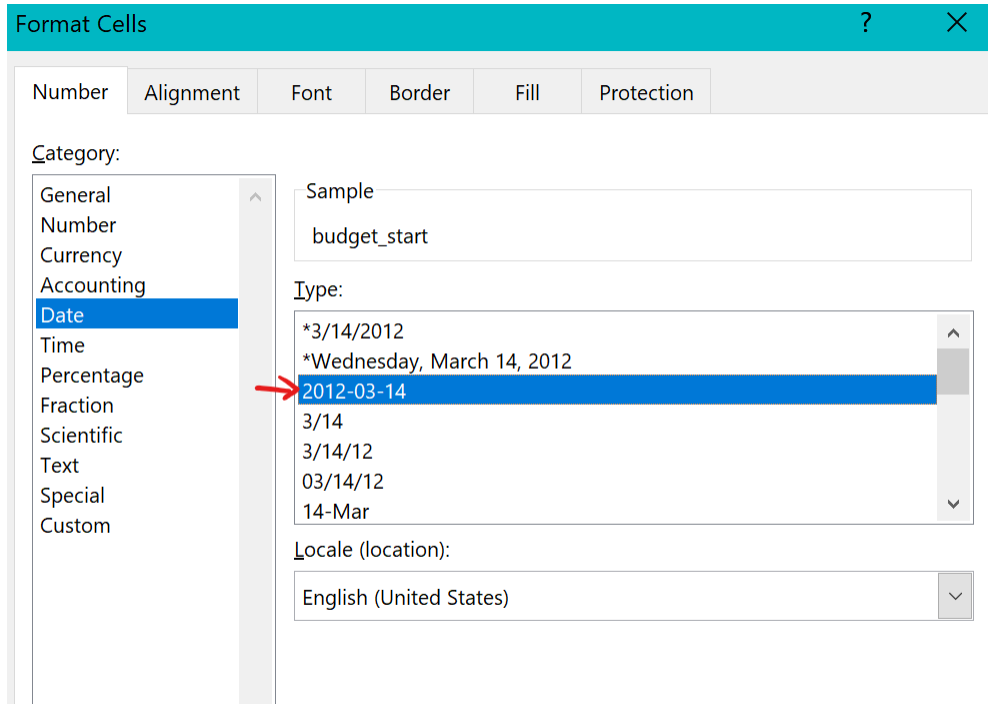
Category:

- General
- Number
- Currency
- Accounting
- Date**
- Time
- Percentage
- Fraction
- Scientific
- Text
- Special
- Custom

Sample

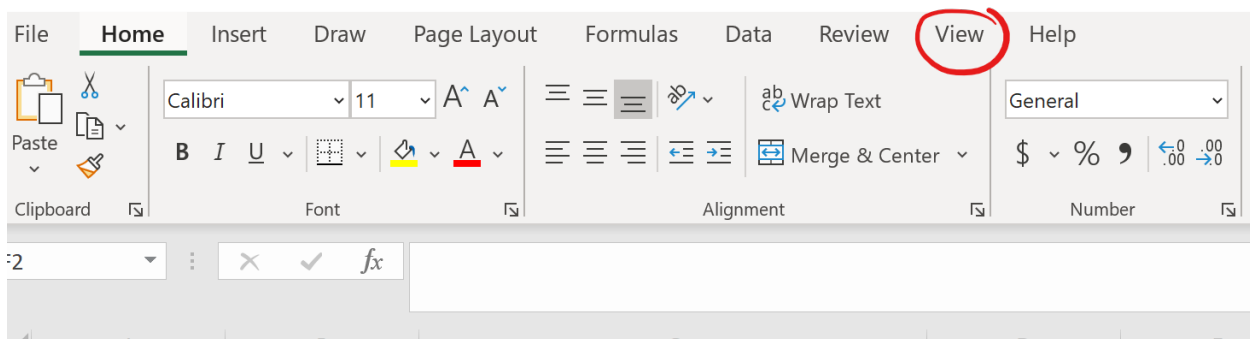
General format cells have no specific number format.

- Select YYYY-MM-DD.

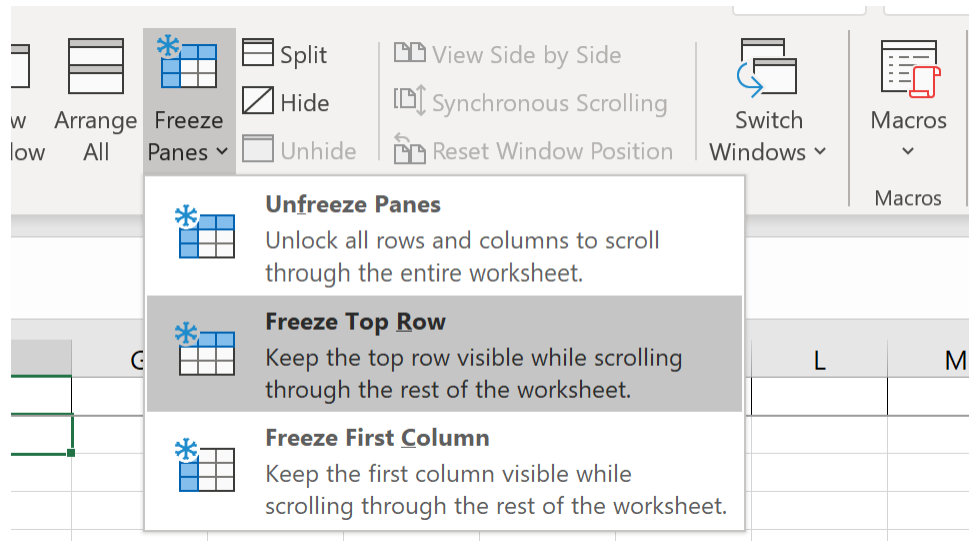


3. Freeze the top row, so the column names remain visible while you scroll down the rows.

- Click on View.



- Click on "Freeze Panes" and select "Freeze Top Row."



- 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.

- Click on the cell under column "F" and type in the column title.

E	F	G
total_cost	priority	
249000		

- Write the IF statement in cell "F2."

E	F	G	H
total_cost	priority		
	=IF(E2>20000, "high", "low")		

- Then double click on the square in the bottom righthand corner of the cell.

E	F
total_cost	priority
	low

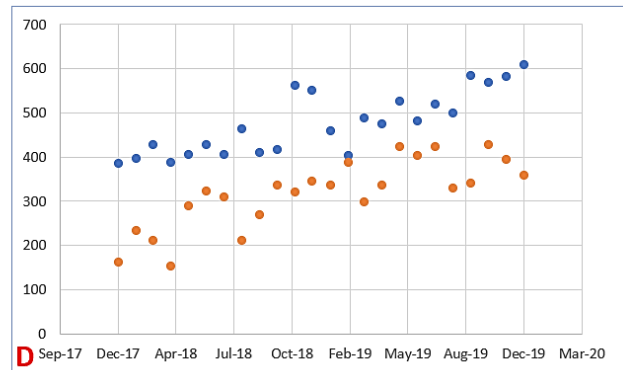
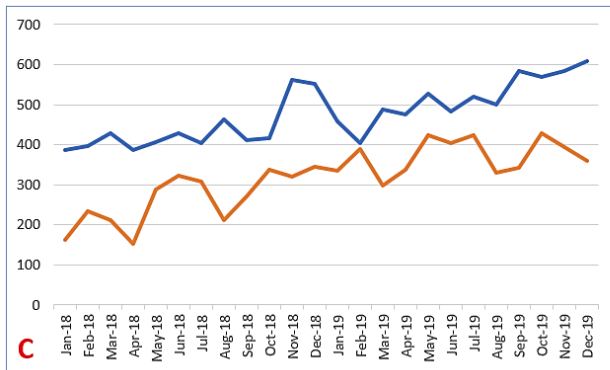
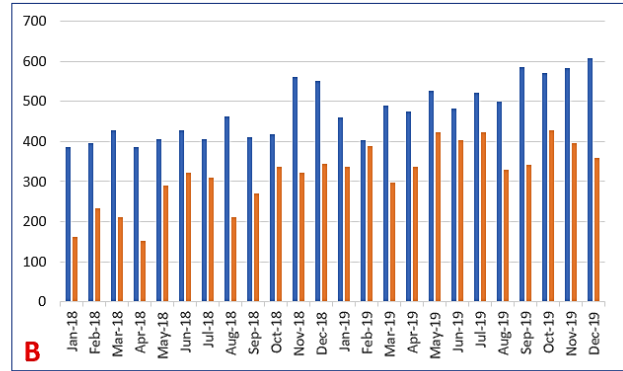
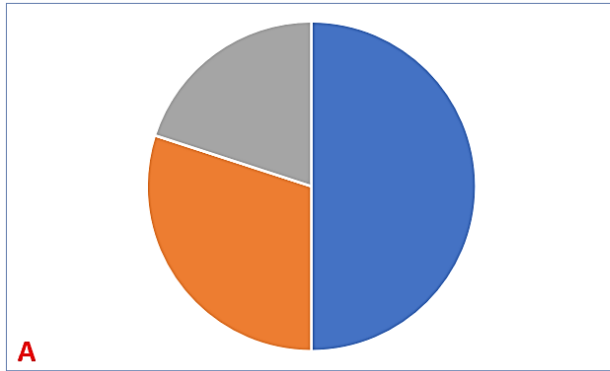
- The values will be filled in for the entirety of the dataset.

E	F
total_cost	priority
	low
	low
	low
249000	high
	low
	low
	low
513191	high
191276	high
589515	high
167079	high
	low
330000	high
	low
	low
	low
	low
	low
	low
	low
15000	low
	low

- Alternatively, you can copy the formula in "E2" using "Control" and "C" then use "Control," "Shift," "End," and hit "Enter," and the formula will be copied down the dataset.

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?

- Pie Chart
- Bar Graph
- Line Graph
- Scatter Plot



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?

- a) Tarshizi, Cooke, Smith, Kim (2020)
- b) Tarshizi, et. al. 2020
- c) (Tarshizi et al., 2020, p. 155)**
- d) Tarshizi...& Kim (2020, pp. 155)

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

- a) True**
- b) False

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

- a) True
- b) False**

Purdue University has an excellent APA 7th Edition quick [reference guide](#).

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
```

- a) x = 0
- b) x = 5
- c) x = 6
- d) x = 7

For more information please review [here](#).

2. 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")

array

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

 =

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

- a) sum = 0
- b) sum = 5
- c) sum = 6
- d) sum = 11

For more information please review [here](#).

3. Match the following Python error messages with a possible solution/explanation:

```
"2" + 2
```

Error Message: TypeError: can only concatenate str (not "int") to str

a.

```
import non_existent_module
```

Error Message: ImportError: No module named 'non_existent_module'

b.

```
my_dict = {"a": 1}  
print(my_dict["b"])
```

Error Message: KeyError: 'b'

c.

```
my_list = [1, 2, 3]  
print(my_list[3])
```

Error Message: IndexError: list index out of range

d.

```
if x == 10  
    print("x is 10")
```

Error Message: SyntaxError: invalid syntax

e.

- i. Add a colon (":") before the line break **(correct match: e)**
- ii. Use pip install or another python installer to add the requested package **(correct match: b)**
- iii. Python uses 0-indexing, so the final value in the list can be accessed with my_list[2] **(correct match: d)**
- iv. Remove the quotation marks to perform numeric functions **(correct match: a)**
- v. Create an entry with my_dict["b"]=x before trying to print it **(correct match: c)**

4. The purpose of this part is to **review** sample code in SQL, R, and Python (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?

- a) Distribution Plots
- b) Scatter Plot Matrix**
- c) Heatmap
- d) Sankey Plots

Example #1 (SQL):

With the sample dataset:

employee_id	first_name	last_name	dept_id	manager_id	salary	expertise
100	John	White	IT	103	120000	Senior
101	Mary	Danner	Account	109	80000	junior
102	Ann	Lynn	Sales	107	140000	Semisenior
103	Peter	O'connor	IT	110	130000	Senior
106	Sue	Sanchez	Sales	107	110000	Junior
107	Marta	Doe	Sales	110	180000	Senior
109	Ann	Danner	Account	110	90000	Senior
110	Simon	Yang	CEO	null	250000	Senior
111	Juan	Graue	Sales	102	37000	Junior

Show all employees with a higher than average salary:

```
SELECT
  first_name,
  last_name,
  salary
FROM employee
WHERE salary > ( SELECT AVG(salary) FROM employee )
```

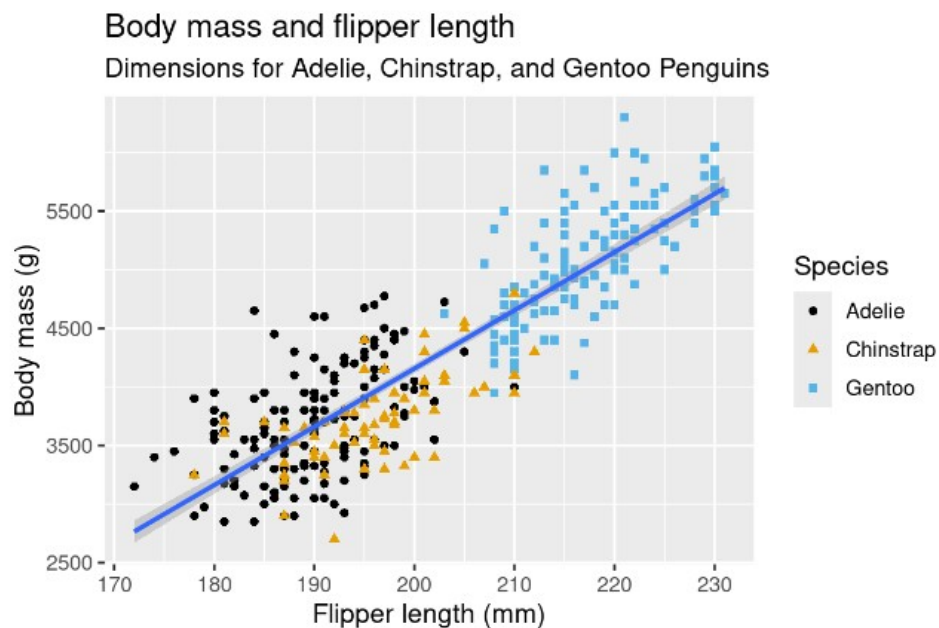
first_name	last_name	salary
Ann	Lynn	140000
Peter	O'connor	130000
Marta	Doe	180000
Simon	Yang	250000

Example #2 (R):

```
library(palmerpenguins)
```

```
library(ggthemes)
```

```
ggplot(
  data = penguins,
  mapping = aes(x = flipper_length_mm, y = body_mass_g)
) +
  geom_point(aes(color = species, shape = species)) +
  geom_smooth(method = "lm") +
  labs(
    title = "Body mass and flipper length",
    subtitle = "Dimensions for Adelie, Chinstrap, and Gentoo Penguins",
    x = "Flipper length (mm)", y = "Body mass (g)",
    color = "Species", shape = "Species"
  ) +
  scale_color_colorblind()
```



Example #3 (Python):

```
# Import necessary libraries
import numpy as np
import pandas as pd
import os
import sklearn

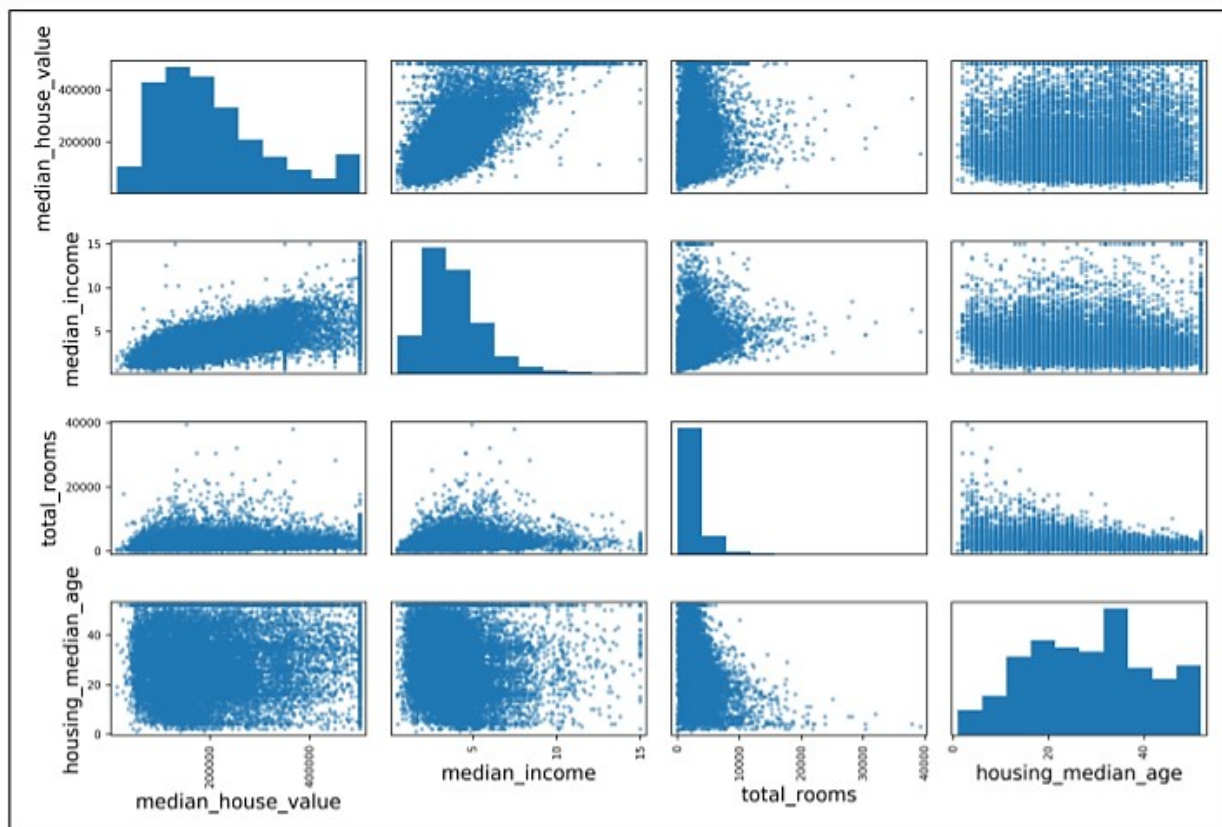
# Libraries to plot pretty figures
import matplotlib as mpl
import matplotlib.pyplot as plt

# Define a function to load the dataset
def load_housing_data():
    csv_path = os.path.join(housing_path, "housing.csv")
    return pd.read_csv(csv_path)

# Loading the dataset into our environment
housing = load_housing_data()

from pandas.plotting import scatter_matrix

attributes = ["median_house_value", "median_income", "total_rooms",
             "housing_median_age"]
scatter_matrix(housing[attributes], figsize=(12, 8))
```



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.

a) Yes

b) No

c) I’m not sure

2. I am comfortable working and learning independently, and I can maintain a high motivation during the master’s program.

a) Yes

b) No

c) I’m not sure

3. 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.

a) Yes

b) No

c) I’m not sure

4. I like working in teams and virtual teamwork project settings, and I am responsive to teammates and very comfortable with online communications.

a) Yes

b) No

c) I’m not sure

5. 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.

a) Yes

b) No

c) I’m not sure

Which option do you prefer (a or b)?

a) 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.

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

We highly recommend that our students allocate and spend about **18-22** hours per week on readings, assignments, projects, quizzes, etc., in this program.

A Few Notes on Time Management and Reducing Stress:

- Visually prioritize your time with a monthly calendar that shows all major due dates. Create a daily to-do list with study-time goals and assignments you plan to complete.
- Study efficiently through highlighting and taking notes to easily see the main points. Consider even using your time to study while you are waiting for an appointment, etc.
- Learn the time of day you are most the most productive, whether that is early morning or late at night. Use this time to your advantage to complete your work/reading.
- Many students have anxiety about learning the programming languages required for this field. As with learning any new language, at first, it might be difficult, but with **practice** and repetition, you will get the hang of it. MS-ADS understands students with different technical backgrounds are entering into the program and have designed the introductory programming courses to accommodate true beginners.
- Give yourself some grace. Even if you did not accomplish all your goals for the day, know your limits, and give yourself a pass. Make sure you schedule a time for hobbies and socializing to provide yourself with a very important and much needed mental outlet.

A Few Hints on Being a Good Team Player in the MS-ADS Group Projects:

- Actively listen and participate in e-meetings.
- Establish effective communication channels using Slack, Zoom, e-mail, and phone.
- Respect teammates' time and roles. Meet your deadlines.
- Be positive and a problem-solver in challenging cases and data-driven projects.
- Flex to others' work styles. Bring high-quality work to the team.
- Focus on collaboration, not just cooperation.
- Celebrate your teammates' successes.

Free Preparation Resources for Starting a Data Science Graduate Program:

- Khan Academy: [Calculus](#)
- Khan Academy: [Statistics](#)
- Math is Fun: [Statistics](#)
- Khan Academy: [Linear Algebra](#)
- [Calculus eTextbook](#) by Gilbert Strang
- Welcome to [Excel Easy](#)
- [Excel Tutorial](#)
- LearnPython.org: free interactive [Python tutorial](#)
- DataCamp.com: free interactive [Introduction to Python](#)
- DataCamp.com: free interactive [R Introductory Course](#)
- [Introduction to R](#) by Udemy
- learnsqlonline.org: free interactive [SQL tutorial](#)
- [15 MOOCs \(Massive Open Online Courses\) for Data Science](#)
- MS-ADS GitHub [Materials](#)
- MS-ADS [Blog](#)
- Data Science Blog: [Toward Data Science](#)
- Data Science News, Datasets, Tutorials, and Blog: [KDnuggets](#)