

MSDA Challenge Exam

Here is the challenge exam for your CUNY MSDA application. You have one week to complete the exam and submit to Rokshana Ali at Rokshana.Ali@cuny.edu.

You may consult the Internet, calculators, and textbooks. You may not ask other people for help.

The questions on the exam are designed to help us evaluate your background and experience in a number of areas, so that we are able to determine your level of preparedness to enter the program. If your answers demonstrate less proficiency in an area, you may be accepted contingent upon successful completion of one or more “bridge courses,” which are free to admitted students.

Please do your best. And know that most successful MSDA candidates will not be able to answer every question.

We care about correct answers (when questions have correct answers!), but we’re also interested in understanding your approach to solving problems, and your ability to demonstrate that you can deliver work in agreed upon timeframes.

1. The chart below shows the on-time and delayed arrivals for two competing railway companies to five cities. What would you conclude about the relative performance of these two railway companies?

		Atlantis	El Dorado	Hyperborea	Narnia	Valhalla
SHINKANSEN	on time	497	221	212	503	1841
	delayed	62	12	20	102	305
TGV	on time	694	4840	383	320	201
	delayed	117	415	65	129	61

2. If you roll four dice at once, what is the probability that the same value appears on all four dice? Generalize the formula that you used for n dice.
3. The table below shows four small datasets, each with an x variable and y variable. What can you say about these four datasets?

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

4. Write a program in a language of your choice to determine how many numbers from 1 to 1000 are not divisible by any of 3, 7, and 11.
5. Create a normalized SQL database that shows a many to many relationship between information in two tables. Combine information from both tables into a single result set, and export its result to a CSV file. So that your work is *reproducible*, please provide scripts to show how you created the tables, populated the tables with a few sample records, combined information from the two tables into a result set, and exported the result set into a .CSV file.

6. Write code that uses latitude and longitude to calculate the distance between New York and Mumbai. Briefly explain why the concept of “distance” is important in data mining.
7. Box A contains one white ball and two red balls. Box B contains one white ball and three red balls. A ball is picked at random from box A and put in box B. A ball is then picked at random from box B. What is the probability that the final ball picked is white?
8. Find $\int t^2 e^t dt$
9. Find \mathbf{A}^{-1} if $\mathbf{A} = \begin{pmatrix} 5 & -1 \\ 2 & 3 \end{pmatrix}$
10. Find the eigenvalues and corresponding eigenvectors of $\mathbf{A} = \begin{pmatrix} 4 & -6 \\ 3 & -7 \end{pmatrix}$
11. At a movie theater, customers wait in line to buy tickets for an average of 5 minutes, and they wait to buy popcorn for an average of 8 minutes. Assuming that the wait times are independent, find the probability that a customer waits a total of less than 20 minutes before taking his or her seat.
12. Briefly (in 200 words or less) compare business intelligence and predictive analytics.
13. If you were to play a match of 100 “rock-paper-scissors” games against a well-written computer program, who do you expect would win? Briefly explain your reasoning.
14. <https://xkcd.com/882/>. Please read and explain.
15. Have you ever used data to solve an interesting problem? If so, please describe.
16. What do you see as the most important skills in the day to day work of a successful data scientist?
17. Describe what you see as the biggest obstacle for you to become a successful data scientist, and your plan to overcome this obstacle.
18. If you were to ask one additional question for this challenge exam, what would that question be?