NUS Invest QF Department Round 1 Assessment

For **both** Quant Researcher and Data Engineer applicants, please **attempt all questions** from **Section 1** and **Section 2**.

Those interested in applying for the Quant Researcher role need **only attempt Section 3** and may **skip Section 4** while Data Engineers need **only attempt Section 4** and may **skip Section 3.**

If using Python, please submit your findings/results for Section 3/4 in a Jupyter Notebook.

If you are also applying to the NUS Fintech Society Quant department, for Section 3, you may submit the same model/strategy in both applications.

Section 1:
Q1: What is the expected number of rolls of a fair 6-sided die until you see two 5s in a row?
Q2: You continually roll a fair 10-sided die. What is the expected number of rolls until the lowest common multiple of all numbers that have appeared is greater than 2000? (You may give an approximate answer which is close to the true value.)
Q3: Suppose the true value of V a car is uniformly distributed between 0 and 1000. You can bid any amount for the car, and if you bid the true value or more then you pay your bid and get the car. You know a very good salesman, and are confident in your ability to sell the car for 50% more than its true value. What should you bid to maximize your expected profit? Now suppose V ~ Uniform[100, 1000], what should you bid?

Section 2:
Q4: There are 50 ants on a 1m line. The 25 left-most ants are moving right and the 25 right-most ants are moving left. When 2 ants collide, they will both reverse direction. How many collisions will there have been in total once all ants have fallen off the end of the line?
Q5: Joe flips a fair coin 26 times, while Tim flips a fair coin 25 times. What is the probability that Joe gets strictly more heads than Tim?
Q6: You are with your spouse at a party, along with 6 other couples. Some people shake hands at this party, although nobody shakes hands with themselves or their spouse. You observe that everyone (excluding yourself, but including your spouse) shook hands with a different number of people. How many people did your spouse shake hands with?

Section 3 (Quant Researchers only):
Refer to the Round 1 Quant Researcher Coding Questions Excel file for more details. You need to **only attempt 1** of the 4 questions provided.

Section 4 (Data Engineers only):
You need **only** attempt **2** questions: Q1 **and** either Q2 **or** Q3.

Q1: Briefly write about a project/internship that you have done that is relevant to Data Engineering/Machine Learning and explain your understanding of **only one** of the listed topics: ETL process, Unsupervised/Supervised Learning, Gradient Descent, Ensemble Learning.

Q2: This question will consist of three small tasks. You may scrape data from the FRED website (or use the API if you have an account):
Part (a):
You will start by gathering the table in the base url, https://fred.stlouisfed.org/release/tables?rid=53&eid=13416&od=#, which consists of data for Q1 2024, Q4 2023 and Q1 2023 of gross value added by sector.
Your result should (but not necessarily) look like this:

Time	Gross domestic product	Business	Nonfarm	Farm	Households and institutions	Households	Nonprofit institutions serving households	General government	Federal	State and local	Gross housing value added
Q1 2024	28,255.928	21,678.178	21,494.105	184.073	3,593.059	2,068.862	1,524.197	2,984.692	956.593	2,028.099	2,600.752
Q4 2023	27,956.998	21,491.702	21,307.177	184.525	3,535.921	2,031.491	1,504.430	2,929.376	937.753	1,991.623	2,555.891
Q1 2023	26,813.601	20,657.951	20,435.618	222.333	3,357.194	1,916.742	1,440.452	2,798.456	893.423	1,905.033	2,414.062

Part (b):
Similar to part (a), now you will have to gather data starting from '1950-01-01' to '2024-01–01'. Save your data frame in a csv file.
Your result should (but not necessarily) look like this:

Time	Gross domestic product	Business	Nonfarm	Farm	Households and institutions	Households	Nonprofit institutions serving households	General government	Federal	State and local	Gross housing value added
1950-01-01	280.828	232.582	214.879	17.703	17.258	12.575	4.684	30.988	19.098	11.890	15.993
1950-04-01	290.383	241.558	223.656	17.902	17.585	12.815	4.770	31.241	19.173	12.068	16.419
1950-07-01	308.153	257.435	238.324	19.111	18.096	13.247	4.849	32.622	20.237	12.385	16.878
1950-10-01	319.945	266.085	245.899	20.186	18.764	13.555	5.209	35.096	22.405	12.691	17.372
1951-01-01	336.000	278.530	257.330	21.200	19.431	14.162	5.269	38.039	24.947	13.092	17.903

Part (c): Data visualization
Plot your data and give some insights about the data.

Q3:
Provided is a DTCC dataset containing derivatives trade data submitted to the CFTC for swaps, forwards, options (put, call, combination) and other derivatives in different currencies. You may refer to page 19 of the RT PPD Quick Reference Guide to get a better understanding of the different columns.

The main task is to clean this dataset and to categorize the transactions into the different types of derivatives (swaps, forwards, options, etc.). For this dataset, there is no one best approach to clean and categorize it.