Mid-semester Test

September 20, 2024: 16:00 – 18:00

Marks: 30, Weightage: 30%

INSTRUCTIONS:

1. You are allowed to access your notes, slides, Moodle, Internet and other Internet based tools during the test. However, the following restrictions will strictly apply:
    a) Tools such as MS Teams, Google Docs, WhatsApp, etc. SHOULD NOT be accessed during the test.
    b) Use of any other online collaboration tool / other means of collaboration is also NOT ALLOWED at any time through the duration of the test.
    c) Use of mobile phone is allowed, but **ONLY** for connecting to the internet, only if there are Wi-fi problems. No other application (including calculators) should be used on the mobile phone during the test.
2. The test should be answered using Jupyter Notebook, or similar Notebook environment. If you use a 'coding only' environment, you will have to submit a separate document (PDF) containing your analysis and report.
3. File names of the Notebook / code / report files SHOULD have your roll number.
4. The first cell of the Notebook SHOULD be a markup cell and it SHOULD list all the major decisions and steps you have taken to arrive at the solution(s). Subsequently, include markup cells containing your analysis as and when required.
5. The code should flawlessly execute with your data file(s) kept in the same folder.
6. Upload **all created data files, Notebooks, source code files** (if separately created), **PDF of your report** (if separately created) to the **MidsemTest** submission point on Moodle.
7. **MidsemTest** will stop accepting submissions at 18:15 Hrs.
8. **MALPRACTICES / CHEATING: ANY MALPRACTICE OR VIOLATION OF ABOVE STATED RESTRICTIONS OR ANY CHEATING WILL RESULT IN THE SEVEREST POSSIBLE DISCIPLINARY ACTION – WHETHER DETECTED DURING THE TEST OR LATER / DURING EVALUATION.**

Read the following statements carefully and work out your solutions following Data Science best practices. The assessment scheme will be as follows:

Problem definition and solution approach: completeness and correctness	10 marks
Results, analysis and conclusions: completeness and correctness	15 marks
Adequate and proper documentation of analysis and results	05 marks

---

You have applied for the post of a Data Scientist (DS) at FitCo Pvt. Ltd., a company with a pan-India presence, and engaged in the manufacture, marketing, sales, and support of fitness equipment. With Indians focussing on fitness goals, the company has been growing steadily since the Pandemic. As per the advertisement for the DS post, it now wants to introduce and emphasize **data orientation** in all its activities. The company's Marketing and Sales Head is tasked with the responsibility of building the team of Data Scientists and Analysts.

On the day of the interview the Head addresses the candidates as follows: "We are giving you one of the files - **fitco.csv** - from our data archives. See what best you can do with it! It contains the total sales value (Sales), and unit Sales price (SellingPrice) related to the sales of some of our products. It also tells us if any product return happened related to a particular sale. We always wanted tools to help us predict the total sales value of our products given their SellingPrice. We would also like to know if product returns are likely at selected price points. Can you help create this capability?"

He further sheepishly said, "But, you know, I really do not know how many products' data is captured in this file!" He continued, "BTW, there is some relevant information for you. For some of the products we followed pricing policies that were different for Class-A cities as compared with the other towns and cities. People in Class-A cities showed

some weird purchasing behaviour!" Finally, he said, "And one more point, while my team will be interested in reviewing and understanding your code, I will need complete, detailed, and good explanation of your approach, steps and analysis. It should be thorough, and amply supported with plots, tables and appropriate explanations. I've heard about many undesirable behaviours of ML models – you will have to convince me that your models are the best, well behaved, and reliable!"

He continued: "Once you develop the required models, see if you can provide us the following information: for SellingPrice ranging between Rs 60 and Rs 150, incremented at Rs 5, predict the product(s) sales value at each selling price, and also let us know if any product 'returns' can be expected at those price points. It will be nice if you can create Table(s) and plot(s) to illustrate the results, and explain them."

"Wish you the very best, all of you"

Do your best to satisfy the executive, and get the job. It will help if you create logical, complete yet precise and well-organized material for explanation and discussion. Good luck!

***

Remember:

- Your submissions should be uploaded to Moodle by 18:15 Hrs, September 20, 2024.
- Cheating, and use of any collaboration tools on the computer / phones is completely prohibited. Violations WILL attract disciplinary action.

Only in case of difficulty in submitting to Moodle by 18:10 Hrs:

- Using your **IITB login**, upload your submission to:

  https://docs.google.com/forms/d/e/1FAIpQLSe50vwoGMH5B6gnuAz2_xd8U-tNesfPSULPQSfJGiWAg2Wxyw/viewform

- **You will receive a confirmation email upon successful submission.**
- **Note:**
  - **If you have submitted to Moodle, DO NOT use this method to create a duplicate submission. Submissions to Moodle will be treated as final.**
- **If you are unable to submit even using this alternate method, immediately contact one of the invigilating TAs before 18:15 Hrs.**

Under no circumstances will submissions be accepted after 18:15 Hrs.

***