Course Name	Analytical Techniques for Business (ATB)	Course Code	**20821**
Max. Time	**2 hours**	Max. Marks	**40 MM**

INSTRUCTIONS:

a. Attempt all the questions on a single Jupyter Notebook
b. The data for the case is available on Moodle.
c. Write down your Roll no., course name and course code on top of Jupyter Notebook
d. Save your Jupyter notebook with **.ipynb** extension and as **pdf** file
e. Upload both the files on Moodle.
f. Label the files as ATB_roll no (for example: ATB_23)
g. This is an open book exam. Students may refer to the codes available on Moodle.

Read the case below and answer the questions given by analyzing the data using Python.

Case Study: Ratings of Bollywood Movies

The Indian film industry produces the maximum number of movies per year, higher than any other country's movie industry. However, very few movies taste commercial success. With 3.3 billion tickets sold annually, India also has the highest number of theater admissions. With so much at stake and highly uncertain nature of returns, it is of commercial interest to develop a model which can predict the success of a movie. Indian Hindi Movie industry popularly known as Bollywood has reached staggering proportions in terms of volume of business, employment, movies produced (more than 100 in a year) and its reach (more than 100 countries worldwide).

In a BBC article, film director Karan Johar is quoted as saying "only 45 of the 300 million (India's middle class) is reached by the movie industry. A vested effort in using data analytics to improve India's film industry could be the way to reach the remaining 255 million.

The data of 140 Bollywood movies is given in Excel file labelled "ATB_Bollywood_2023". Karan Johar has hired you as the data scientist to extract insight by analyzing the data using descriptive analytics.

Variable(s)	Description
MovieName	Name of the Movie
Release_festive	Whether the movie is released during festive or holiday season 1 if released during festive season 0, otherwise
StatusStar	Status of lead actor/ actoress; debut, star or superstar
Sequel	Whether the movie is sequel 1 if yes 0, otherwise
Genre	1. Action/adventure 2. Family/children 3. Comedy 4. Drama 5. Horror 6. Mystry/Suspense 7. Sci-fi/ fantasy
BudgetCrores	Budget in crore
CriticsRating	Critics' rating
IMDbRating	IMDB movie rating
1stWeekBoxOfficeCollection	First week box office collection
TotalBoxOffice	Total Box-office collection

Analyze the data and create a managerial report by writing answers to the following questions:

1. Import the data file   **(2 marks)**
2. What is the dimension of the imported data   **(2 marks)**
3. Identify the numeric and categorical variables.   **(2 marks)**
4. Develop the graphical summary of numerical and categorical variables.   **(10 marks)**
5. Does there exist association amongst the numeric variables. Analyze and comment upon the degree of association.   **(10 marks)**
6. Determine the highest and lowest IMDB Rating.   **(2 marks)**
7. Estimate the average Box office collection of a sci-fi movie?   **(2 marks)**
8. Does there exist outlier in Total Box Office collection. Analyze and list all the outlier values.   **(5 marks)**
9. Suggest two more findings which are not covered in the above questions.   **(5 marks)**

Note: Interpretation of all the outputs should be written by putting comments on the Jupyter notebook.

***** *