**CLO1**

**Question 1: Define the purpose of data analysis for the chosen dataset.**

**Answer:**

**Purpose: Find which factors (film, price, time, cinema) drive ticket sales and seat occupancy, so you can optimize scheduling and pricing.**

**Question 2: Identify and Justify the type of programming used for data analysis.**

**Answer:**

**Programming: Python—using pandas/NumPy for data wrangling, Matplotlib/Seaborn for charts, and scikit-learn for modeling.**

**Question 3: Identify the type and purpose of the machine learning algorithm to be implemented for the chosen dataset.**

**Answer:**

**ML Algorithm: Regression (e.g. linear or random forest) to predict total sales or occupancy rate from your features.**

**Question 4: Identify and Justify the independent and dependent variables for the chosen dataset.**

**Answer:**

**Variables:**

**Features (independent): ticket\_price, show\_time, capacity, film\_code, cinema\_code**

**Target (dependent): total\_sales (or occupancy\_rate)**

**Question 5: Justify why you want to perform the descriptive analysis for the chosen dataset.**

**Answer:**

**Why Descriptive Analysis: Quickly summarize distributions, catch errors/outliers, and inform feature choices before building models.CLO2**

**Loading the selected data.**

To answer this question, we will start reading data and doing some processing of that data. Now I will share some of the screenshots of the data loading and processing. 

Now we are done with the initial analysis of the data after loading it, this data loading will help us for later analysis.

**Question 6: Create a script to develop a Python function for** **descriptive statistics. The input for the function should be the sample and the field to perform the descriptive statistics.**

**Answer:**

Now we will write the script for the descriptive statistics.

A screen shot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

**Question 7: Create a program to random sampling of size 150 and find the descriptive statistics for the dependent variable from the sample [Apply the descriptive function which you created.**

**Answer:**

**A white background with black text

AI-generated content may be incorrect.**

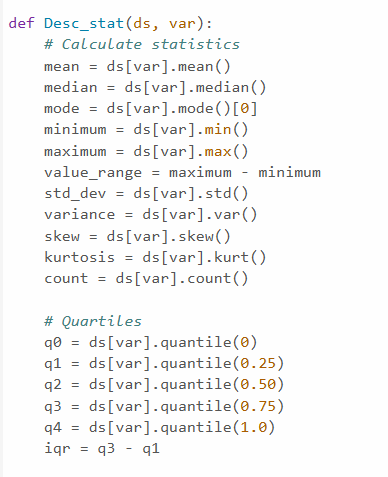
**This is the results of the script that we write.**

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**Question 8: Create a script for systematic sampling by giving certain conditions and finding the desc stat for the dependent variable from the sample [Apply the descriptive function which you created].**

**Answer:**

****

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Question 9: Create a detailed descriptive statistics report about the dependent variable of the chosen dataset.**

**Answer:**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Question 10: Visualize the dependent variable by the Graph/Chart of the following using Python Program:**

* Scatter plot
* Box Plot
* Histogram
* Heat Map

Hint: Use Matplot or Ski-learn library

**Answer:**

1. Script for the scatter plot

**A screen shot of a computer code

AI-generated content may be incorrect.**

**A graph with blue dots

AI-generated content may be incorrect.**

1. Script for the Box Plot

A screenshot of a computer code

AI-generated content may be incorrect.

A graph with lines and dots

AI-generated content may be incorrect.

1. Script for the Histogram.

A computer code with text

AI-generated content may be incorrect.

A graph with a bar

AI-generated content may be incorrect.

1. Script for of Heat Map.

A screenshot of a computer code

AI-generated content may be incorrect.

A diagram of a heatmap

AI-generated content may be incorrect.

**Question 11: Perform the hypothesis test to find the correlation (Pearson and Spearman for numerical variable and chi-square test for categorical variable) between the independent variable and the dependent variable.**

**Note:** If you have more than one independent variable, then chose any one of the independent variables.

**Answer:**

We will:

* Use Pearson and Spearman correlation for numerical variables
* Use Chi-Square test for categorical variables

Let’s say:

* Dependent Variable: total\_sales
* Independent Variable: tickets\_sold (numeric)
* Optional Categorical: You could use something like cinema\_code or month if you want to run a chi-square test.

**Use Pearson and Spearman correlation for numerical variables**

A computer code with many colored text

AI-generated content may be incorrect.

The output of this script show below:

A black text with white text

AI-generated content may be incorrect.

**Use Chi-Square test for categorical variables**

A screenshot of a computer program

AI-generated content may be incorrect.

**Question 12: Assess the performance of the dependent variable to know whether the sample is representative of the normal population by a one-sample t-test.**

**Answer:**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**A close up of a text

AI-generated content may be incorrect.**.

**CLO3**

**Question 13: Build, Train, Develop and Evaluate using Simple Regression for chosen dataset.**

**Answer:** We’ll choose one independent variable (e.g., ticket\_price) to predict the dependent variable (e.g., tickets\_sold).

**A screenshot of a computer program

AI-generated content may be incorrect.**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**A graph with a red line and blue dots

AI-generated content may be incorrect.**

**Question 14: Develop a script to forecast the value of the dependent variable from all the relevant independent variables using Multiple Linear Regression.**

**Answer:**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**Question 15: Predict the value of the dependent variable from the different classifier such as Logistic Regression, KNN, Naïve-Bayes and Decision Tree.**

**Answer:**

A computer screen shot of text

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer screen

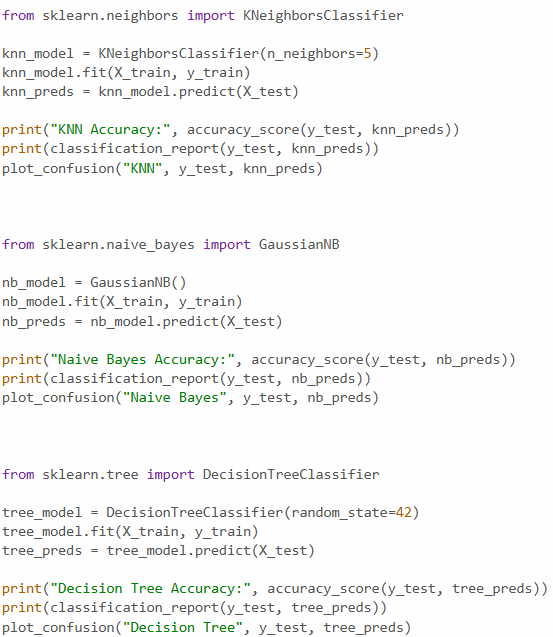
AI-generated content may be incorrect.

**Question 16: Evaluate the performance of each model using confusion matrix and accuracy and identify the best fit classifier for the chosen dataset.**

**Answer:**

**A screenshot of a computer code

AI-generated content may be incorrect.**

****

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**A screenshot of a graph

AI-generated content may be incorrect.**

**A screenshot of a graph

AI-generated content may be incorrect.**

**A screenshot of a graph

AI-generated content may be incorrect.**

**A diagram of a decision tree

AI-generated content may be incorrect.**

**Question 17: Predict the dependent variable by using best-fit classifier.**

**Answer:**

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Question 18: Perform the cluster analysis such as K-means and Horizontal for any field from the chosen dataset  
Answer:**

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

A graph with colored dots

AI-generated content may be incorrect.

A diagram of a clustering graph

AI-generated content may be incorrect.

A graph with red green and blue dots

AI-generated content may be incorrect.

**Question 19: Explain the strategy for improving the system after viewing the cluster diagram.  
Answer:**

**Improvement Strategies Based on Clustering:**

* **Segment Target Groups:** Tailor promotions for low-performing clusters and loyalty programs for high performers.
* **Optimize Resources:** Allocate staff and advertising based on cluster performance and potential.
* **Personalize Marketing:** Customize pricing and services to fit each cluster’s preferences.
* **Analyze Outliers:** Investigate anomalies for errors, fraud, or special cases.
* **Refine Features:** Adjust or create features to better distinguish between groups.
* **Track Continuously:** Regularly update clustering as new data comes in to keep strategies relevant.