Lab	Practical
LAB-1	**Introduction to library for Data Science** 1. Introduction to various library : numpy, pandas, matplotlib, scipy, scikit-learn, Keras and TensorFlow, NetworkX, Beautiful Soup. 2. Load data from a CSV file. 3. Introduction to Viewing Data, Data Selection, Data Cleaning, Filtering and Sorting, Modifying Data, Grouping and Aggregation. 4. Viewing Data function: head(), tail(), shape(), info(), describe(). 5. Data Selection function: select single/multiple column, iloc(), loc().
LAB-2	**Perform Exploratory Data Analysis on User Demographics using Pandas library functions** 1. Import the necessary libraries. (A) 2. Import the dataset from this address. (A) https://raw.githubusercontent.com/justmarkham/DAT8/master/data/u.user 3. Assign it to a variable called users and use the 'user_id' as index. (A) 4. Display the first 25 entries. (A) 5. Display the last 10 entries. (A) 6. What is the number of observations in the dataset? (A) 7. What is the number of columns in the dataset? (A) 8. Print the name of all the columns. (A) 9. How is the dataset indexed? (A) 10. What is the data type of each column? (A) 11. Print only the occupation column. (B) 12. How many different occupations are in this dataset? (B) 13. What is the most frequent occupation? (B) 14. Summarize the Data Frame. (B) 15. Summarize all the columns. (B) 16. Summarize only the occupation column. (B) 17. What is the mean age of users? (C) 18. What is the age with least occurrence? (C)

LAB-3	**Demonstrate basic operations and time series creation using the Pandas library** 1. Import the necessary libraries. (A) 2. Create a Series from a list. (A) 3. Accessing elements from series. (A) 4. Performing arithmetic operation on series. (A) 5. Create Series with custom index labels. (A) 6. Create a Series with data, index, and dtype parameters. (B) 7. Create a Pandas Series using two lists: (B) ▪ One list for the **index labels** ▪ One list for the **data values** 8. Creating Time Series using Pandas Series: (C) ▪ Use pd.to_datetime() to create a starting date. ▪ Use pd.to_timedelta() to generate a range of time delta (days). ▪ Add the delta to the starting date to generate a datetime index. ▪ Create a Pandas Series using the datetime index.
LAB-4	**Create and manipulate Series and DataFrames using the Pandas library with a student dataset** 1. Import the necessary libraries. (A) 2. Create a data dictionary (name, roll_no, grade, email, department). (A) 3. Assign it to a variable called student. (A) 4. Reorder the columns as: name, department, roll_no, grade, email. (B) 5. Add another column called city and insert any values you like. (B) 6. Display the data types of each column. (B)
LAB-5	**Perform Exploratory Data Analysis on Global Alcohol Consumption using Pandas library** 1. Import the necessary libraries. (A) 2. Import the dataset from this address. (A) https://raw.githubusercontent.com/justmarkham/DAT8/master/data/drinks.csv 3. Assign it to a variable called drinks. (A) 4. Which continent drinks more beer on average? (A) 5. For each continent print the statistics for wine consumption. (B) 6. Print the mean alcohol consumption per continent for every column. (B) 7. Print the median alcohol consumption per continent for every column. (B) 8. Print the mean, min and max values for spirit consumption. (C)

LAB-6	**Perform data analysis on an employee salary dataset** 1. Import the necessary libraries. (A) 2. Read Salaries.csv as a DataFrame called sal. (A) 3. Check the head of the DataFrame. (A) 4. Use the .info() method to find out how many entries there are. (A) 5. What is the average BasePay? (A) 6. What is the highest amount of OvertimePay in the dataset? (A) 7. What is the job title of JOSEPH DRISCOLL? Note: Use all caps, otherwise you may get an answer that doesn't match up (there is also a lowercase Joseph Driscoll). (A) 8. How much does JOSEPH DRISCOLL make (including benefits)? (B) 9. What is the name of highest paid person (including benefits)? (B) 10. What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid? (B) 11. What was the average (mean) BasePay of all employees per year? (2011-2014)? (B) 12. How many unique job titles are there? (B) 13. What are the top 5 most common jobs? (B) 14. How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?). (C) 15. How many people have the word Chief in their job title? (C)
LAB-7	**Perform data analysis on E-commerce Purchase dataset** 1. Import the necessary libraries. (A) 2. Read in the Ecommerce Purchases csv file and set it to a DataFrame called ecom. (A) 3. Check the head of the DataFrame. (A) 4. How many rows and columns are there? (A) 5. What is the average Purchase Price? (A) 6. What were the highest and lowest purchase prices? (A) 7. How many people have English 'en' as their Language of choice on the website? (A) 8. How many people have the job title of "Lawyer"? (A) 9. How many people made the purchase during the AM and how many people made the purchase during PM? (B) 10. What are the 5 most common Job Titles? (B) 11. Someone made a purchase that came from Lot: "90 WT", what was the Purchase Price for this transaction? (B) 12. What is the email of the person with the following Credit Card Number: 4926535242672853. (B) 13. How many people have American Express as their Credit Card Provider *and made a purchase above $95? (B) 14. How many people have a credit card that expires in 2025? (C) 15. What are the top 5 most popular email providers/hosts (e.g. gmail.com, yahoo.com, etc...). (C)

LAB-8	**Perform data analysis on Employee Salary dataset** 1. Import the necessary libraries. (A) 2. Read Salaries.csv as a DataFrame called sal. (A) 3. Check the head of the DataFrame. (A) 4. Use the .info() method to find out how many entries there are. (A) 5. What is the average BasePay? (A) 6. What is the highest amount of OvertimePay in the dataset? (A) 7. What is the job title of JOSEPH DRISCOLL? Note: Use all caps, otherwise you may get an answer that doesn't match up (there is also a lowercase Joseph Driscoll). (A) 8. How much does JOSEPH DRISCOLL make (including benefits)? (B) 9. What is the name of highest paid person (including benefits)? (B) 10. What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid? (B) 11. What was the average (mean) BasePay of all employees per year? (2011-2014)? (B) 12. How many unique job titles are there? (B) 13. What are the top 5 most common jobs? (B) 14. How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?). (C) 15. How many people have the word Chief in their job title? (C)
LAB-9	**Demonstration of Pandas and Seaborn library functions for Data Analysis and Visualization** 1. Import the necessary libraries. (A) 2. Load the diabetes.csv dataset into a DataFrame called df. (A) 3. Display the first 5 rows of the dataset using .head(). (A) 4. Use the .info() method to get the number of entries and data types. (A) 5. Find the mean value of the Glucose column. (A) 6. Check for any missing values in the dataset. (A) 7. Display basic statistical details using .describe(). (A) 8. Create a histogram for the Age column. (A) 9. Find out how many patients have a BMI greater than 30 (considered obese). (A) 10. Group the dataset by Outcome and compute the average Glucose level for each group. (A) 11. Find mean, median, and standard deviation of BMI, Glucose, and Insulin. (B) 12. Count patients with Glucose > 120 and BMI > 30. (B) 13. Pair plot of selected features (Glucose, BMI, Age, Outcome). (B)

LAB-10	**Exploring Heart Disease Classification with Multiple Machine Learning Algorithms** 1. Import the necessary libraries. (A) 2. Import the dataset from this address. (A) https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/heart-disease.data 3. Check for missing values. (A) 4. Split the dataset into features (X) and target (y). (A) 5. Split data into training and testing sets (80% training, 20% testing). (A) 6. Standardize the feature data, which is important for models like KNN and Logistic Regression. (A) 7. Initialize the classifiers. (A) 8. Create a function that trains each model, makes predictions, and evaluates them with accuracy, classification report, and confusion matrix. (A) 9. Evaluate the models. (A) 10. Display the results. (A)
LAB-11	**Perform Linear Regression on Sample and Salary Datasets using Scikit-Learn library** **Dataset 1:** 1. Import the necessary libraries. (A) 2. Import the dataset. (A) 3. Import Linear Regression model from sklearn.linear_model. (A) 4. Create a Linear Regression model. (A) 5. Split the dataset into input (X) and output (y). (A) 6. Fit the model (Training). (A) 7. Predict the output values. (A) 8. Visualize the results. (A) **Dataset 2: Salary Dataset** 9. Import the necessary libraries. (B) 10. Read the salary data. (B) 11. View the data using a scatter plot. (B) 12. Split the dataset into input and output variables (X and y). (B) 13. Create Linear Regression Model. (B) 14. Fit the Model (Training). (B) 15. Predicting the Test set results.(B) 16. Visualizing the Training set results. (B) 17. Visualizing the Test set results. (B)

LAB-12	**Implementation of Polynomial Regression using Scikit-Learn library** 1. Import the necessary libraries. (A) 2. Import the dataset. (A) 3. Plot the data using scatter plot. (A) 4. Splitting the dataset into the Training set and Test set. (A) 5. Fitting Linear Regression to the dataset. (A) 6. Predict the x_test using Linear Model. (A) 7. Visualising the Linear Regression results. (A) 8. import PolynomialFeatures. (B) 9. create PolynomialFeatures of degree 4. (B) 10. Apply fit_transform to features. (B) 11. Fitting Polynomial Regression to the dataset. (B) 12. Create Grid for higher resolution and smoother curve. (B) 13. Predict the x_test using Polynomial Regression. (B) 14. Visualising the Polynomial Regression results. (B)
LAB-13	**Implement K-Means and K-Medoids Clustering on Student Performance Data** 1. Load necessary libraries like pandas, matplotlib, seaborn, and clustering modules from sklearn and sklearn_extra. (A) 2. Read the StudentsPerformance.csv file using pandas. (A) 3. Display the first few rows, data types, summary statistics, and check for missing values. (A) 4. Use seaborn to generate a pairplot for visual exploration of relationships between features. (A) 5. Convert categorical data to numerical using encoding (e.g., LabelEncoder or OneHotEncoder). (A) 6. Select relevant features (X) for clustering. (A) 7. Use the KMeans class to cluster data and assign cluster labels. (A) 8. Plot the clusters and centroids for visual understanding. (B) 9. Use KMedoids from sklearn_extra.cluster to perform clustering. (B) 10. Display the clusters formed by K-Medoids for comparison with K-Means. (B)
LAB-14	**Summarize Cloud service models with real time examples.**
LAB-15	**Perform Various steps of utilizing the windows Azure platform.**