**Dr. Ambedkar Institute of Technology, Bengaluru-56**

**(An Autonomous Institute, Approved by AICTE-I, Affiliated to VTU)**

**Master of Computer Applications**

**ASSIGNMENT-I**

Subject Name : Data Science Subject Code : 22MCA23

Semester : II Max. Marks : 10

|  |  |  |  |
| --- | --- | --- | --- |
| Sl.No |  | CO | BTL Levels |
| 1 | Generate a **Patient data set** with following attributes and perform different operations using Python   1. **Data Creation using Data Frame**   PatientID, Age  hemoglobin (Range 8-14)  bloodpressure (100-180)   1. **Data description** 2. Display first 3 records & Display last 2 records 3. Get structure of the dataset 4. Find the summary of the attributes 5. Remove the headers and display the dataset 6. **Data Preprocessing**   Data cleaning-  Find and treat missing values   1. Find if missing values exists 2. Deletion 3. Assuming 0 4. Imputation with mean & median 5. Write the treated(imputed) dataset to .csv file 6. **Data manipulation** 7. Extract first 1st, 3rd rows and 2nd, 4th columns (attributes) 8. Remove 2,4,6 records from the dataset 9. Display a subset whose **age** is >40 10. Write the subset data to CSV file 11. Compute mean of hemoglobin values 12. **Data Visualization** 13. Plot a graph with age and hemoglobin levels 14. Plot a pie chart on blood pressure levels 15. Plot box plot with age and blood pressure levels. | CO1  CO2  CO3 | L3  L4 |
|  |  |  |  |
| 2  333 | Generate **a Student Result data set** with following attributes and perform operations using Python   1. **Data Creation**  * StudentID * Gender * University Affiliation * UG Percentage * PG Percentage  1. **Data Preprocessing**   Data cleaning-   1. To identify duplicates & remove them   Find and treat missing values   1. Find if missing values exists 2. Deletion 3. Assuming 0 4. Imputation with mean & median 5. Write the treated(imputed) dataset to .csv file 6. **Data Description** 7. Read a CSV or Excel dataset 8. Display first 3 records & Display last 2 records 9. Get structure of the dataset 10. Find the summary of the attributes 11. Remove the headers and display the dataset 12. **Data manipulation** 13. Extract only 2 columns (attributes) 14. Remove 1,3 5 records from the dataset 15. Display a subset whose **UG percentage** is > 65 16. Display a subset whose **University Affiliation =Bangalore** 17. Create a subset with **StudentID, University Affiliation & PG Percentage** 18. Write the subset data to CSV file 19. **Data Visualization** (Include axis label, title, color, legend) 20. Plot a graph with **Gender Vs UG Percentage** 21. Plot a pie chart on **University Affiliations** 22. Draw a line graph (overlaying graphs) on **UG percentage & PG percentage**. | CO1  CO2  CO3  CO4 | L3  L4 |
| 3 | Create 2 data frames and perform different operations -merge(), groupby(), concat()  Dataframe1={EmpID, EmpName, Designation}  Dataframe2={EmpID, Designation, Qualification}   1. Display both Data frames 2. Merge Dataframe1 with Dataframe2 3. Merge with data frames along the column EmpID (on=empid) 4. Merge with data frames along the Designation with matching records from both sides (outer) 5. Merge two dataframes using keys from left dataframe only(left) 6. Merge two dataframes using keys from right dataframe only(right) 7. Append either columns or rows from one DataFrame to another(concat) 8. Grouping the data based on distinct values in rows and columns(groupby) 9. Count the number of occurrences for each unique value after grouping the data(groupby with count) |  |  |

**Signature of the faculty Signature of the HOD**

**Dr. Ambedkar Institute of Technology, Bengaluru-56**

**(An Autonomous Institute, Approved by AICTE, Affiliated to VTU)**

**Master of Computer Applications**

**ASSIGNMENT-II**

Subject Name : Data Science Subject Code : 22MCA23

Semester : II Max. Marks : 10

|  |  |  |  |
| --- | --- | --- | --- |
| Sl.No |  | CO | BTL Levels |
| 1 |  | CO3  CO4 | L4  L5 |
| 2 |  | CO3 | L3  L4 |
| 3 |  | CO3  CO4 | L4  L5 |

**Signature of the faculty Signature of the HOD**

**Dr. Ambedkar Institute of Technology, Bengaluru-56**

**(An Autonomous Institute, Approved by AICTE, Affiliated to VTU)**

**Master of Computer Applications**

**GROUP ACTIVITY**

Subject Name : Data Science Subject Code : 22MCA23

Semester : II Max. Marks : 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sl.No** | **USN** | **Name of the Student** | **Topic Assigned** | CO | BTL Levels |
|  |  |  | Build a Linear Regression Model on **Blood Pressure Vs Age** | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Build Linear Regression Model on **Student Performance** | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Analyse **House price using suitable regression model** | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Build a model to predict **blood glucose levels with respect to age** | CO1 | L1 |
|  |  | CO1 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Analyse **the purchase pattern by an emplyee based on salary** using suitable analytical model | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |
|  |  |  | Build a model to **predict price of car** based on suitable features of it | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Perform prediction of **disease(any) using suitable features of patient** | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  | Model build to measure of **countries GDP per year** | CO1 | L1 |
|  |  | CO2 | L2 |
|  |  | CO3 | L3 |
|  |  | CO4 | L4, L5 |
|  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Signature of the faculty Signature of the HOD**