### **MES Abasaheb Garware College, Karve Road, Pune**

### **Department of Computer Science**

### **M.Sc. (computer Science)**

### **Semester IV Internship Evaluation Sheet (2022-2023)**

Name of the student: - Vaishnavi Anil Khandekar

Mobile No: - 9284910295

Name of the Company offering internship: - Expertrons Company

Name of the Guide: - Bharti Karpatti

Contact Email of the Manager: -  [trainer@foxmula.com](mailto:trainer@foxmula.com)

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Date** | **Description of the task completed** | **Status** |
| 1 | 10 February to 16 February | Following session were conducted:  1.Introduction of statistics and Machine learning  2. Installation of Anaconda | **Assignment**  **Completed** |
| 2 | 19 February to 24 February | Following session were conducted:  1. Introduction of jupyter workbench  2. Basics of Python  \*Data Structures  -Operation on Integer  -Operation on String  -Operation on Float  -Operation on Boolean | **Assignment**  **Completed** |
| 3 | 26 February to 2March | Following session were conducted:  1.Python  -Operation on List  -Operation on Dictionaries  \* Loops  -Operation on While Loops  -Operation on For Loops  -Operation on Break and Continue | **Assignment**  **Completed** |
| 4 | 4 March to 9March | Following session were conducted:  1.Pandas  2.SQL Joins  -Operation on Inner Joins  -Operation on Left Joins  - Operation on Right Joins  - Operation on Full Outer Joins  3.Merge  - Operations on Concatenating Data Frames  4.Converting Data From Long to Wide  Format using pivot()  5.Transforming Data  -Operations on Removing Duplicates  - Operations on Binning Numeric variables to Categorical  6. Task  - Create a data frame with 2 variables called 'Age' and 'Income'.  - Fill these with random integers between (1, 100) and (10k to 100k) for 'Income'. Use 10k rows.  - Use cut to bin Age into 5 bins.  - Use qcut to bin Income in to 10 bins.  - Assign meaningful labels to each.  - Convert both these cut variables into Dummies.  - Report the mean and sum of each dummy variable. | **Assignment**  **Completed** |
| 5 | 11March to 16March | Following session were conducted:  1.Plotting and Visualization  -Matplotlib Basics  -Plotting in Pandas  -Operation on Bar Plots  -Operation on Histogram and Density plots  - Operation on Scatter Plots  *2.Group by*  *3.Task*  -Create a 100x4 Data Frame filled with random numbers (from a normal distribution.) Ensure that there's 2 categorical columns with 5 and 3 categories each. - Create the groupby object using both keys and find the mean, max, median for each group.  4. Column-wise aggregation and UDFs  5. Groupby Practice Tasks - Baseball Data a. Import the data from this link http://bit.ly/144sh7t (hint: use read\_csv) Call it `baseball` b. Check column types, dataframe shape c. How many rows have missing data? d. Find the proportion of missing values in each column e. Find - The number of rows in every league - The count of records per year - Average, Median experience of players participating in each year. | **Assignment**  **Completed** |
| 6 | 18March to 23March | Following session were conducted:  1.Data Science Life Cycle  - Understanding the problem statement  - Understand The Data  - Data Preprocessing  - Exploratory Data Analysis(EDA)  - Feature Engineering  - Selection of variables  - Test and Train Data  - Building Model  - Evaluate the Model  -Prediction  2.Machine Learning  -Supervised vs Unsupervised  3.Tree Algorithm  -Decision Tree  -Random Forest  4.SVM(Support Vector Machine)  5.Basics Statistics  -Linear Regression  -Logistic Regression  6.Neural Networks |  |
|  |  |  |  |
|  |  |  |  |