

Data Mining and Practical Machine Learning

Lab Sheet 6

1. Load the dataset 'BigmartData.csv'.
2. Print the following contents of the dataset.
 - (a) Whole data set
 - (b) First 5 rows
 - (c) Last 5 rows
3. Print the following characteristics of the dataset.
 - (a) Size
 - (b) Column names
 - (c) Data types of the column names
4. Find the rows with missing values and handle them.
5. Obtain the value counts of all data columns.
6. Sort any two numerical data columns in ascending and descending order.
7. Obtain the descriptive statistics and the correlation of the dataset.
8. Slicing out the following portions of the dataset.
 - (a) Item_Type, Outlet_Size, Outlet_Type columns
 - (b) [100:110, 5:10] (from row 100 to 110 & col 5 to 10)
9. Draw the histogram and the boxplot for the 'Item_Weight' column.
10. Draw a bar chart and pie chart for the value count of the 'Item_Type'.
11. Apply the aggregation using groupby () to the following situations.
 - (a) Count of the 'Outlet_Location_Type' from Outlet_Identifier' and 'Outlet_Location_Type'.
 - (b) Mean of the 'Item_Outlet_Sales' from 'Item_Weight' and 'Item_Outlet_Sales'.
12. Apply filters for the following situations.
 - (a) Item_Outlet_Sales \geq 2000.0
 - (b) Item_Fat_Content with 'Low Fat' content.