1. **Pandas used to read the CSV file**

**DAV MPA 1**

1. **Matplot & seaborn used for graph plot**

**Load the library (1 mark)**

**import** pandas **as** pd

**Open the csv file (1 mark)**

file**=**pd.read\_csv('cancer.csv')

file

**Display the top 5 rows (1 mark)**

file.head()

**Display column information of the dataframe (1 mark)**

file.info()

**Display statistics information (1 mark)**

file.describe()

**Check for null values (1 mark)**

file.isnull().sum()

**Display the column names (1 mark)**

file.columns

**Display the unique value of Level column (1 mark)**

file['Level'].unique()

**Convert the column Level from categorical type to numerical type (1 mark)**

file['Level']**=**pd.factorize(file['Level'])[0]

file['Level'].unique()

**Import the vizualization library (1 mark)**

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**Display cor-relation matrix (1 mark)**

file.corr()

**Display the histogram of the dataframe (2 mark)**

file.plot.hist(bins**=**26, figsize**=**(15,6))

plt.show()

**Display the countplot for Gender at different Level (1 mark)**

sns.countplot(data**=**file, x**=**'Level', hue**=**'Gender')

plt.show()

**Display the countplot for Gender at different Smoking (1 mark)**

sns.countplot(data**=**file, x**=**'Smoking', hue**=**'Gender')

plt.show()

**Display a pie chart the % of different Level (1 mark)**

level\_counts **=** file['Level'].value\_counts()

labels **=** level\_counts.index

sizes **=** level\_counts.values

explode **=** (0.1,0,0)

plt.figure(figsize**=**(8, 6))

plt.pie(sizes, labels**=**labels, explode**=**explode,autopct**=**'%1.1f%%', shadow**=True**, startangle**=**140)

plt.show()

**Display a distribution plot for Level column (1 mark)**

sns.histplot(data**=**file, x**=**'Level', discrete**=True**, kde**=True**)

plt.show()

**Display a line plot between Age and Level (1 mark)**

sns.lineplot(data**=**file, x**=**'Age', y**=**'Level', hue**=**'Level', marker**=**'o', sort**=False**)

plt.show()

**Display a line plot between Chest Pain and Level (1 mark)**

sns.lineplot(data**=**file, x**=**'Chest Pain', y**=**'Level', hue**=**'Level', marker**=**'o', sort**=False**)

plt.show()

**Display a scatter plot between Smoking and Shortness of Breath w.r.t Level (hue) (1 mark)**

sns.scatterplot(data**=**file, x**=**'Smoking', y**=**'Shortness of Breath', hue**=**'Level')

plt.show()

**Display a scatter plot between Age and Gender w.r.t Level (hue) (1 mark)**

sns.scatterplot(data**=**file, x**=**'Age', y**=**'Gender', hue**=**'Level')

plt.show()

**display a histogram plot between OccuPational Hazards and Gender (1 mark)**

sns.histplot(data**=**file, x**=**'OccuPational Hazards',y**=**'Gender',)

plt.show()

**Display a bar plot for obesity and coughing of blood (1 mark)**

plt.bar(file['Obesity'],file['Coughing of Blood'])

plt.xlabel('obesity')

plt.ylabel('coughing of blood')

plt.show()

**Display the boxplot for each feature of the dataframe (2 mark)**

file.plot(kind**=**'box',subplots**=True**,layout**=**(5,5))

plt.show()

​