Machine Learning Lab

Exercise: 1

Praveshini S B

22PC26

Explore pandas library using the following the dataset

https://drive.google.com/file/d/1Yzw9jJUpbB6ea24RQEoRzceWOgPQHidc/view?usp=drive_link

- Read and load it onto a DataFrame.

```
import pandas as pd
data=pd.read_csv("ipl.csv")
print(data)
```

- Displaying First Few Records of the DataFrame

```
import pandas as pd
data=pd.read_csv("ipl.csv")
print(data.head())

```

- Finding Summary of the DataFrame

```
import pandas as pd
data=pd.read_csv("ipl.csv")
print(data.info())
print(data.describe())
```

- Slicing and Indexing of DataFrame

```
import pandas as pd
data=pd.read_csv("ipl.csv"
print(data.iloc[1:3,1:3])
print(data['COUNTRY'])
```

- Find the occurrences of each unique value in a column.

```
import pandas as pd
df=pd.read_csv("ipl.csv")
for column in df.columns:
    print(f"\nOccurrences of each unique value in column '{column}':")
    print(df[column].value_counts())
```

- Perform cross tabulation between PLAYING ROLE and AGE will give number of players in each age category for each playing role

```
import pandas as pd
df=pd.read_csv("ipl.csv")
print("\nCross tabulation between 'PLAYING ROLE' and 'AGE':")
cross_tab = pd.crosstab(df['PLAYING ROLE'], df['AGE'])
print(cross_tab)
```

- Sorting Data Frame by Column Values

```
import pandas as pd
data=pd.read_csv("ipl.csv")
sortData = data.sort_values(by='AGE')
print("\nSorted DataFrame by 'AGE':")
print(sortData)
```

- Find average SOLD PRICE for each age category, group all records by AGE and then apply mean() on SOLD PRICE column

```
import pandas as pd
data=pd.read_csv("ipl.csv")
avg = data.groupby('AGE')['SOLD PRICE'].mean()
print(avg)
```

- Perform joining of data frame- compare the average SOLD PRICE for different AGE categories with the different age and PLAYING ROLE categories, we need to merge the DataFrames soldprice_by_age and soldprice_by_ age_role. The common column is AGE and this needs outer join. Age is considered as categorical varaibale(as 1,2,3)

```
import pandas as pd
data=pd.read_csv("ipl.csv")
avg = data.groupby(['PLAYING ROLE','AGE'])['SOLD PRICE'].mean()
avg2=data.groupby('AGE')['SOLD PRICE'].mean()
t1=pd.DataFrame(avg)
t2=pd.DataFrame(avg2)
merged = pd.merge(t1,t2, on='AGE', how='outer')
print(merged)
```

- To find whether players carry a premium if they belong to a specific AGE and PLAYING ROLE category. The premium (we will call it change) is calculated in percentage terms and calculated as follows

$$\text{Change} = \frac{\begin{array}{c}(\text{Average } \textit{SOLD PRICE} \text{ for all player in an } \textit{AGE} \text{ and } \textit{PLAYING ROLE} \text{ category} \\ - \text{ Average } \textit{SOLD PRICE} \text{ for all player in an } \textit{AGE} \text{ category})\end{array}}{\text{Average } \textit{SOLD PRICE} \text{ for all player in an } \textit{AGE} \text{ category})}$$