**Report after Project Finalization**

**Title:** Data Analysis with Pandas and Visualization

**Importing**

Pandas as pd

Numpy as np

**Reading Files**

d1=pd.read\_csv("d1.csv")

d2=pd.read\_csv("d2.csv")

d3=pd.read\_csv("d3.csv")

**Checking Data Shape**

d1.shape

d2.shape

d3.shape

**Merging Data Sheet 1 and 2**

merged1d2=pd.merge(d1,d2,on = 'instant',how= 'inner')

**Sorting Data Sheet 3 on basis of specific column**

d3= d3.sort\_values('instant')

**Deleting unwanted column of datasheet 3**

modifiedd3=d3.drop(['yr','holiday','atemp'],axis=1)

**Concatenate Merge sheet with datasheet 3**

combinedata=pd.concat([merged1d2 ,modifiedd3],axis=0)

**Deleting unwanted column of concatenated sheet**

combinedata=combinedata.drop(['holiday','atemp','hr','yr','mnth','Unnamed: 0'],axis=1)

**Finally Save the filtered sheet**

combinedata.to\_csv('Final\_data\_sheet3.csv', index=False)

**Descriptive statistics and Outliers**

combinedata.mode(), combinedata.mean(), combinedata.std. combinedata.describe

**Using Function**

def find\_outliers(combinedata,coln):

q1 = combinedata[coln].quantile(0.25)

q3 = combinedata[coln].quantile(0.75)

iqr = q3 - q1

min\_r = q1 - 1.5 \* iqr

max\_r = q3 + 1.5 \* iqr

outliers\_indices = combinedata.index[(combinedata[coln] < min\_r) | (combinedata[coln] > max\_r)]

return outliers\_indices

find\_outliers(combinedata,coln)

**Data Visualization with matplotlib and seaborn**

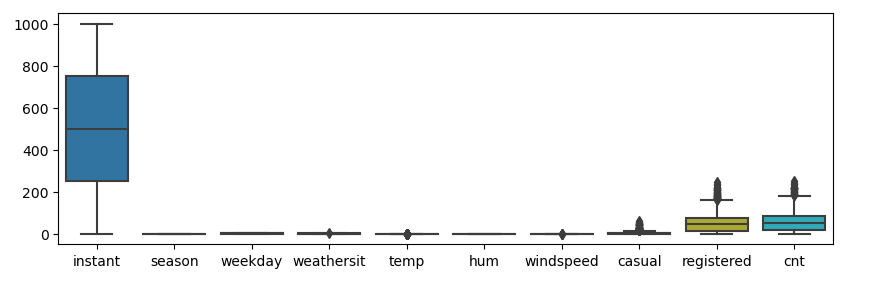
Matplotlib as plt

seaborn as sns

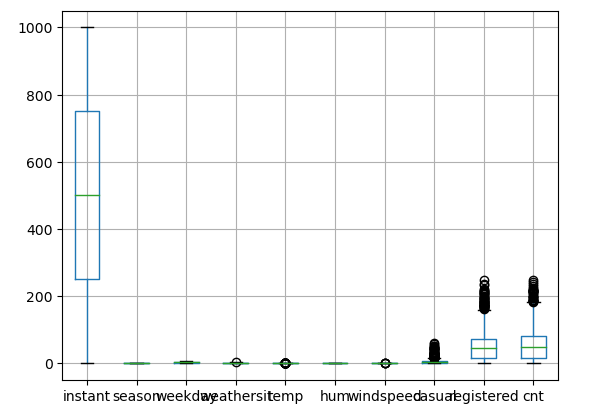
plt.figure(figsize=(10, 3))

sns.boxplot (data = combinedata)

plt.show()



combinedata.boxplot(return\_type= 'axes');



**Skewness and Coherence**

combinedata.skew(axis=0,skipna= True)

corr= combinedata.corr(method = 'pearson')

corr