# Exploratory Data Analysis

## Data: 08/21/2022  
## Objective: Exploring data through visualization  
# Description: In this session, we will explore various techniques to explore the data through visualization  
# We will be discussing various techniques such as bar chart, histograms, scatter plots, etc   
# Exploratory Data Analysis helps in identifying hidden patterns in the data.  
## We will use pandas() libraries  
# PANDAS Documentation  
# https://pandas.pydata.org/docs/getting\_started/intro\_tutorials/01\_table\_oriented.html  
# Each column in a DataFrame is a Series.  
# Pandas dataframe is similar to Excel/database/Table  
# You have rows and columns

import pandas as pd  
import sys  
import os

income = pd.Series([2000,4000,6000,1000,2200,3400,2030,2310,4300.2300])

income.head()

0 2000.0  
1 4000.0  
2 6000.0  
3 1000.0  
4 2200.0  
dtype: float64

income.describe()

count 9.000000  
mean 3026.692222  
std 1537.090831  
min 1000.000000  
25% 2030.000000  
50% 2310.000000  
75% 4000.000000  
max 6000.000000  
dtype: float64

income.max()

6000.0

income.min()

1000.0

income.sum()

27240.23

income.mean()

3026.692222222222

stocks=pd.read\_csv("E:/Umesh-MAY2022/Personal-May2022/BA2ndEdition/2ndEdition/Book Chapters/Chapter 5 - EDA/stocks3.csv")

stocks.head()

Day Stock1 Stock2 Stock3 Stcok4 Stock5 Stcok6 Stock7 Stock8 \  
0 1 17.219 50.500 18.750 43.000 60.875 26.375 67.750 19.000   
1 2 17.891 51.375 19.625 44.000 62.000 26.125 68.125 19.125   
2 3 18.438 50.875 19.875 43.875 61.875 27.250 68.500 18.250   
3 4 18.672 51.500 20.000 44.000 62.625 27.875 69.375 18.375   
4 5 17.438 49.000 20.000 41.375 59.750 25.875 63.250 16.500   
  
 Stock9 Stock10 Ratings   
0 48.750 34.875 High   
1 48.750 35.625 Low   
2 49.000 36.375 Medium   
3 49.625 36.250 High   
4 47.500 35.500 Low

stocks.describe()

Day Stock1 Stock2 Stock3 Stcok4 Stock5 \  
count 950.000000 950.000000 950.000000 950.000000 950.000000 950.000000   
mean 475.500000 37.925592 43.955526 18.704474 45.352895 60.863553   
std 274.385677 10.759764 11.377308 2.730364 5.606898 14.299412   
min 1.000000 17.219000 19.250000 12.750000 34.375000 27.750000   
25% 238.250000 27.781000 35.406250 16.125000 41.375000 49.656250   
50% 475.500000 38.922000 49.062500 19.375000 43.937500 61.750000   
75% 712.750000 46.875000 53.250000 20.875000 48.125000 71.843750   
max 950.000000 61.500000 60.250000 25.125000 60.125000 94.125000   
  
 Stcok6 Stock7 Stock8 Stock9 Stock10   
count 950.000000 950.000000 950.000000 950.000000 950.000000   
mean 24.122632 70.673421 23.294211 44.214211 46.994079   
std 5.530962 6.773101 2.970152 4.270098 6.539420   
min 14.125000 58.000000 16.375000 31.500000 34.000000   
25% 18.000000 65.625000 21.250000 41.750000 41.375000   
50% 25.750000 68.625000 22.500000 44.750000 46.687500   
75% 28.875000 76.375000 26.375000 47.625000 52.125000   
max 35.250000 87.250000 29.250000 53.000000 62.000000

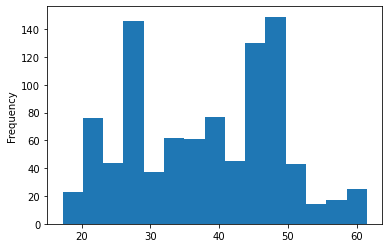
stocks.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 950 entries, 0 to 949  
Data columns (total 12 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Day 950 non-null int64   
 1 Stock1 950 non-null float64  
 2 Stock2 950 non-null float64  
 3 Stock3 950 non-null float64  
 4 Stcok4 950 non-null float64  
 5 Stock5 950 non-null float64  
 6 Stcok6 950 non-null float64  
 7 Stock7 950 non-null float64  
 8 Stock8 950 non-null float64  
 9 Stock9 950 non-null float64  
 10 Stock10 950 non-null float64  
 11 Ratings 950 non-null object   
dtypes: float64(10), int64(1), object(1)  
memory usage: 89.2+ KB

#Histograms

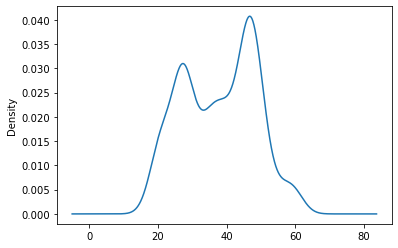
stocks.Stock1.plot.hist(by=None, bins=15)

<AxesSubplot:ylabel='Frequency'>

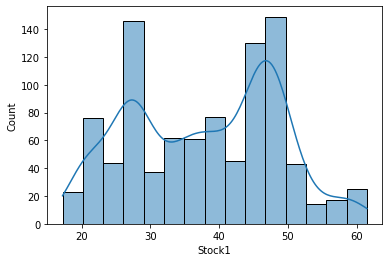


stocks.Stock1.plot.kde(bw\_method='silverman', ind=1200)

<AxesSubplot:ylabel='Density'>



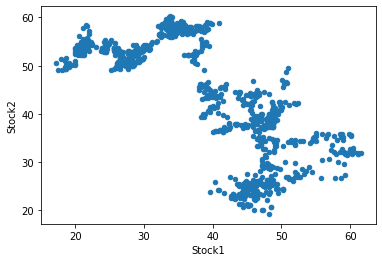
import matplotlib.pyplot as plt  
import seaborn as sns  
sns.histplot(stocks.Stock1, kde=True, bins=15)  
plt.show()



#Scatter plots

stocks.plot.scatter('Stock1', 'Stock2')

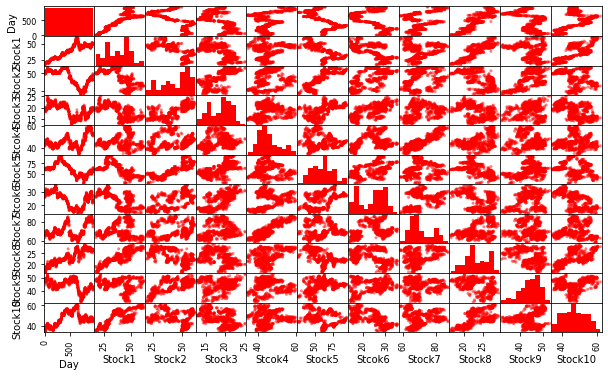
<AxesSubplot:xlabel='Stock1', ylabel='Stock2'>



import matplotlib.pyplot as plt  
from pandas.plotting import scatter\_matrix

scatter\_matrix(stocks,alpha=0.5, figsize=(10,6), ax=None, grid=False,  
 color='red', hist\_kwds={'bins':10, 'color':'red'})

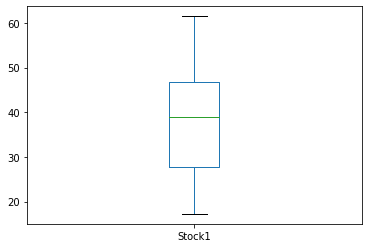
array([[<AxesSubplot:xlabel='Day', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Day'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Day'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock1'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock1'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock2'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock2'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock3'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock3'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stcok4'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stcok4'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock5'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock5'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stcok6'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stcok6'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock7'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock7'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock8'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock8'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock9'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock9'>],  
 [<AxesSubplot:xlabel='Day', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock1', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock2', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock3', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stcok4', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock5', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stcok6', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock7', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock8', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock9', ylabel='Stock10'>,  
 <AxesSubplot:xlabel='Stock10', ylabel='Stock10'>]], dtype=object)



##BOX PLOT

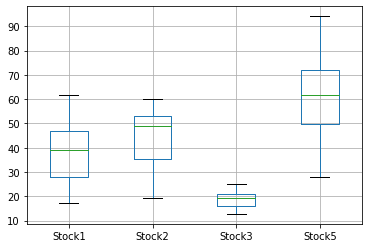
stocks.Stock1.plot.box()

<AxesSubplot:>



stocks.boxplot(column=['Stock1','Stock2','Stock3','Stock5'])

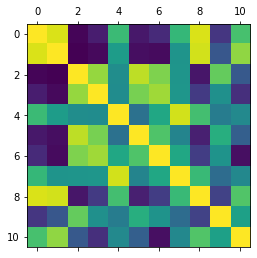
<AxesSubplot:>



import matplotlib.pyplot as plt  
corel=stocks.corr()  
corel.style.background\_gradient(cmap='coolwarm').set\_precision(3)

<pandas.io.formats.style.Styler at 0x1b8c9d62610>

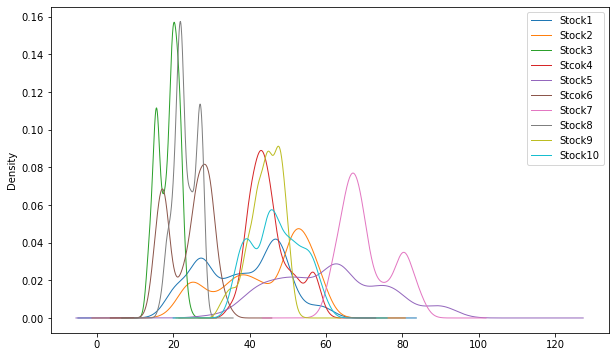
plt.matshow(stocks.corr(method='pearson'))  
plt.show()



stocks1=stocks.drop(columns=['Day'])

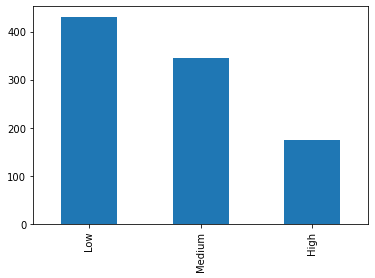
stocks1.plot.density(figsize = (10, 6),  
 linewidth = 1)

<AxesSubplot:ylabel='Density'>



stocks.Ratings.value\_counts().plot.bar()

<AxesSubplot:>



display(stocks)

Day Stock1 Stock2 Stock3 Stcok4 Stock5 Stcok6 Stock7 Stock8 \  
0 1 17.219 50.500 18.750 43.000 60.875 26.375 67.750 19.000   
1 2 17.891 51.375 19.625 44.000 62.000 26.125 68.125 19.125   
2 3 18.438 50.875 19.875 43.875 61.875 27.250 68.500 18.250   
3 4 18.672 51.500 20.000 44.000 62.625 27.875 69.375 18.375   
4 5 17.438 49.000 20.000 41.375 59.750 25.875 63.250 16.500   
.. ... ... ... ... ... ... ... ... ...   
945 946 50.375 46.250 19.375 52.250 61.875 23.500 78.625 26.625   
946 947 50.750 46.375 19.625 50.875 64.625 23.250 77.625 26.500   
947 948 50.625 46.625 19.625 50.875 64.625 23.250 75.000 26.250   
948 949 50.125 47.000 19.875 50.750 62.750 22.875 74.500 25.250   
949 950 49.000 47.000 19.500 49.500 60.875 22.750 75.625 25.500   
  
 Stock9 Stock10 Ratings   
0 48.750 34.875 High   
1 48.750 35.625 Low   
2 49.000 36.375 Medium   
3 49.625 36.250 High   
4 47.500 35.500 Low   
.. ... ... ...   
945 41.875 44.375 Medium   
946 40.750 45.000 High   
947 41.250 44.125 Low   
948 40.625 43.875 Medium   
949 40.500 43.375 High   
  
[950 rows x 12 columns]

stocks.style

<pandas.io.formats.style.Styler at 0x1b8d8e15c40>

os.chdir('C:/Users/u.hodeghatta')

!pandoc BA2-Pandas-EDA-08122022.ipynb -s -o EDA\_notebook.html  
!pandoc BA2-Pandas-EDA-08122022.ipynb -s -o EDA\_n.docx

[WARNING] This document format requires a nonempty <title> element.  
 Defaulting to 'BA2-Pandas-EDA-08122022' as the title.  
 To specify a title, use 'title' in metadata or --metadata title="...".  
pandoc: pandas\_EDA.html: openBinaryFile: does not exist (No such file or directory)