# Plotting Graphs

September 4, 2019

[3]:
```
import pandas as pd
data=pd.read_csv("/home/sakil/Desktop/DataScience/Udemy/Module2/airlinedata.
 ↪csv")
data.shape
```

[3]:
```
(1936758, 30)
```

[4]:
```
data.head()
```

[4]:

| | Unnamed: 0 | Year | Month | DayofMonth | DayOfWeek | DepTime | CRSDepTime | ArrTime | CRSArrTime | UniqueCarrier | ... | TaxiIn | TaxiOut | Cancelled | CancellationCode | Diverted | CarrierDelay | WeatherDelay | NASDelay | SecurityDelay | LateAircraftDelay |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 2008 | 1 | 3 | 4 | 2003.0 | 1955 | 2211.0 | 2225 | WN | ... | 4.0 | 8.0 | 0 | N | 0 | NaN | NaN | NaN | NaN | NaN |
| 1 | 1 | 2008 | 1 | 3 | 4 | 754.0 | 735 | 1002.0 | 1000 | WN | ... | 5.0 | 10.0 | 0 | N | 0 | NaN | NaN | NaN | NaN | NaN |
| 2 | 2 | 2008 | 1 | 3 | 4 | 628.0 | 620 | 804.0 | 750 | WN | ... | 3.0 | 17.0 | 0 | N | 0 | NaN | NaN | NaN | NaN | NaN |
| 3 | 4 | 2008 | 1 | 3 | 4 | 1829.0 | 1755 | 1959.0 | 1925 | WN | ... | 3.0 | 10.0 | 0 | N | 0 | 2.0 | 0.0 | 0.0 | 0.0 | 32.0 |
| 4 | 5 | 2008 | 1 | 3 | 4 | 1940.0 | 1915 | 2121.0 | 2110 | WN | ... | 4.0 | 10.0 | 0 | N | 0 | NaN | NaN | NaN | NaN | NaN |

[5 rows x 30 columns]

[9]:
```
data1=data.loc[0:2,["Year","Month"]]
data1
```

[9]:
```
   Year  Month
0  2008      1
1  2008      1
2  2008      1
```

[10]:
```
data1.shape
```

[10]:
```
(3, 2)
```

[11]:
```
data1.plot()
```

[11]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5768020b00>
```

[15]:
```
#concept of joins
dataframe1=pd.DataFrame({
    "employee":["ABC","XYZ","MNO"],
    "age":["20","30","40"]
}

)
dataframe1
```

[15]:
```
  employee age
0      ABC  20
1      XYZ  30
2      MNO  40
```

[14]:
```
dataframe2=pd.DataFrame({
    "employee":["PQR","XYZ","MNO"],
    "salary":["10000","20000","30000"]
}

)
dataframe2
```

[14]:
```
  employee salary
0      PQR  10000
1      XYZ  20000
2      MNO  30000
```

[17]:
```
#inner join
dataframe3=pd.merge(dataframe1,dataframe2,on="employee")
dataframe3
```

[17]:
```
  employee age salary
0      XYZ  30  20000
1      MNO  40  30000
```

[18]:
```
#outer join
dataframe4=pd.merge(dataframe1,dataframe2,on="employee",how="outer")
```

```
dataframe4
```

[18]:
```
  employee  age salary
0      ABC   20    NaN
1      XYZ   30  20000
2      MNO   40  30000
3      PQR  NaN  10000
```

[19]:
```
#left join
dataframe5=pd.merge(dataframe1,dataframe2,on="employee",how="left")
dataframe5
```

[19]:
```
  employee age salary
0      ABC  20    NaN
1      XYZ  30  20000
2      MNO  40  30000
```

[20]:
```
#right join
dataframe6=pd.merge(dataframe1,dataframe2,on="employee",how="right")
dataframe6
```

[20]:
```
  employee  age salary
0      XYZ   30  20000
1      MNO   40  30000
2      PQR  NaN  10000
```

[29]:
```
#pivoting
dataframe1.pivot(index="employee",columns="age")
```

[29]:
```
Empty DataFrame
Columns: []
Index: [ABC, MNO, XYZ]
```

[30]:
```
#pivoting table
dataframe1.pivot_table(index="employee",aggfunc="sum")
```

[30]:
```
          age
employee
ABC        20
MNO        40
XYZ        30
```

[33]:
```
dataframe1.pivot_table(index="employee",aggfunc="count")
```

[33]:
```
          age
employee
ABC         1
MNO         1
XYZ         1
```

Working with shift function and writing to CSV file

[34]:
```
#working with shift function
```

```
fb_finance=pd.read_csv("/home/sakil/Desktop/DataScience/Udemy/Module2/finance.
 ↪csv")
fb_finance.shape
```

[34]:
```
(4, 7)
```

[35]:
```
fb_finance.head()
```

[35]:

| | Date | open | high | low | close | adjclose | volume |
|---|---|---|---|---|---|---|---|
| 0 | 2 | 4 | 30 | 1 | 8 | 1 | 11 |
| 1 | 4 | 8 | 40 | 2 | 4 | 2 | 22 |
| 2 | 6 | 10 | 50 | 3 | 10 | 3 | 33 |
| 3 | 8 | 12 | 60 | 4 | 12 | 4 | 44 |

[36]:
```
fb_finance.shift(1)
```

[36]:

| | Date | open | high | low | close | adjclose | volume |
|---|---|---|---|---|---|---|---|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 2.0 | 4.0 | 30.0 | 1.0 | 8.0 | 1.0 | 11.0 |
| 2 | 4.0 | 8.0 | 40.0 | 2.0 | 4.0 | 2.0 | 22.0 |
| 3 | 6.0 | 10.0 | 50.0 | 3.0 | 10.0 | 3.0 | 33.0 |

[37]:
```
fb_finance=pd.read_csv("/home/sakil/Desktop/DataScience/Udemy/Module2/finance.
 ↪csv",index_col="Date")
fb_finance
```

[37]:

| Date | open | high | low | close | adjclose | volume |
|---|---|---|---|---|---|---|
| 2 | 4 | 30 | 1 | 8 | 1 | 11 |
| 4 | 8 | 40 | 2 | 4 | 2 | 22 |
| 6 | 10 | 50 | 3 | 10 | 3 | 33 |
| 8 | 12 | 60 | 4 | 12 | 4 | 44 |

[38]:
```
fb_finance.shift(1)
```

[38]:

| Date | open | high | low | close | adjclose | volume |
|---|---|---|---|---|---|---|
| 2 | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | 4.0 | 30.0 | 1.0 | 8.0 | 1.0 | 11.0 |
| 6 | 8.0 | 40.0 | 2.0 | 4.0 | 2.0 | 22.0 |
| 8 | 10.0 | 50.0 | 3.0 | 10.0 | 3.0 | 33.0 |

[39]:
```
fb_finance.shift(-1)
```

[39]:

| Date | open | high | low | close | adjclose | volume |
|---|---|---|---|---|---|---|
| 2 | 8.0 | 40.0 | 2.0 | 4.0 | 2.0 | 22.0 |
| 4 | 10.0 | 50.0 | 3.0 | 10.0 | 3.0 | 33.0 |
| 6 | 12.0 | 60.0 | 4.0 | 12.0 | 4.0 | 44.0 |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN |

[40]:
```
fb_finance.to_csv("/home/sakil/Desktop/DataScience/Udemy/Module2/writeto.csv")
```

[42]:
```
fb_finance.to_csv("/home/sakil/Desktop/DataScience/Udemy/Module2/
↪writetoSomeColumns.csv",columns=["Date","open"])
```

```
/home/sakil/anaconda/lib/python3.7/site-packages/pandas/core/indexing.py:1494:
FutureWarning:
Passing list-likes to .loc or [] with any missing label will raise
KeyError in the future, you can use .reindex() as an alternative.

See the documentation here:
https://pandas.pydata.org/pandas-docs/stable/indexing.html#deprecate-loc-
reindex-listlike
  return self._getitem_tuple(key)
```