In [4]:
```
# import all the necessary libraries. As we will be using linear regression
#here so we will be importing Linear regression from sklearn

import pandas as pd #For data related tasks
import matplotlib.pyplot as plt #for data visualization
import quandl #Stock market API for fetching Data
from sklearn.linear_model import LinearRegression
```

In [5]:
```
# We are using quandl as the data source here by which we are extracting the valu

quandl.ApiConfig.api_key = 'eoX1ZJFGKFXMLo7z3oav'
stock_data = quandl.get('NSE/TCS', start_date='2018-12-01', end_date='2018-12-31'
```

In [6]:
```
# print the data received from quandl API.
print(stock_data)
```

```
                Open     High      Low     Last    Close  Total Trade Quantity
\
Date
2018-12-03  1984.00  1990.00  1968.30  1984.00  1982.40             1610576.0
2018-12-04  1983.00  2019.40  1971.00  2009.85  2010.85             3270615.0
2018-12-05  2006.00  2018.00  1985.00  2003.90  2006.75             2501539.0
2018-12-06  1998.00  2017.00  1979.60  1990.00  1992.70             2321216.0
2018-12-07  1985.05  2003.90  1973.00  1999.85  1995.20             1680420.0
2018-12-10  1975.00  2011.00  1960.00  1961.00  1975.80             2010786.0
2018-12-11  1970.00  2010.00  1961.00  1997.05  2000.00             2942014.0
2018-12-12  2001.10  2022.00  1984.95  2016.50  2016.80             2219993.0
2018-12-13  2024.00  2029.70  1974.50  1981.95  1982.60             3748429.0
2018-12-14  1983.00  1998.95  1975.25  1989.00  1989.75             2473761.0
2018-12-17  1999.00  2004.90  1985.00  1992.90  1994.30             1227921.0
2018-12-18  1991.90  2002.00  1976.40  1987.90  1987.85             1768742.0
2018-12-19  1984.80  1984.80  1960.05  1970.80  1968.45             2498833.0
2018-12-20  1953.80  1974.90  1946.00  1955.00  1954.05             1940277.0
2018-12-21  1948.00  1950.00  1886.55  1905.00  1895.80             3729956.0
2018-12-24  1905.80  1938.90  1905.00  1922.00  1918.50             1864116.0
2018-12-26  1921.80  1921.80  1870.25  1892.00  1889.20             2446614.0
2018-12-27  1909.00  1941.70  1872.10  1909.10  1908.95             4968201.0
2018-12-28  1915.00  1920.00  1893.00  1897.00  1896.05             2239130.0
2018-12-31  1908.00  1909.00  1886.15  1894.75  1893.05             1879740.0

            Turnover (Lacs)
Date
2018-12-03         31868.00
2018-12-04         65568.23
2018-12-05         50131.23
2018-12-06         46373.71
2018-12-07         33463.26
2018-12-10         39990.73
2018-12-11         58636.26
2018-12-12         44663.42
2018-12-13         74802.38
2018-12-14         49094.42
2018-12-17         24482.95
2018-12-18         35137.11
2018-12-19         49198.05
2018-12-20         37945.10
2018-12-21         71360.52
2018-12-24         35878.57
2018-12-26         46112.98
2018-12-27         95411.46
2018-12-28         42708.38
2018-12-31         35647.72
```

In [7]:
```
# converting the quandl received data into pandas dataframe
dataset = pd.DataFrame(stock_data)
```

In [8]:
```
##Now we convert into csv
dataset.to_csv('TCS.csv')
```

In [9]:
```
## We have to read our CSV
data = pd.read_csv('TCS.csv')
```

In [10]:
```
# see the first few data rows
data.head()
```

Out[10]:

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-12-03	1984.00	1990.0	1968.3	1984.00	1982.40	1610576.0	31868.00
1	2018-12-04	1983.00	2019.4	1971.0	2009.85	2010.85	3270615.0	65568.23
2	2018-12-05	2006.00	2018.0	1985.0	2003.90	2006.75	2501539.0	50131.23
3	2018-12-06	1998.00	2017.0	1979.6	1990.00	1992.70	2321216.0	46373.71
4	2018-12-07	1985.05	2003.9	1973.0	1999.85	1995.20	1680420.0	33463.26

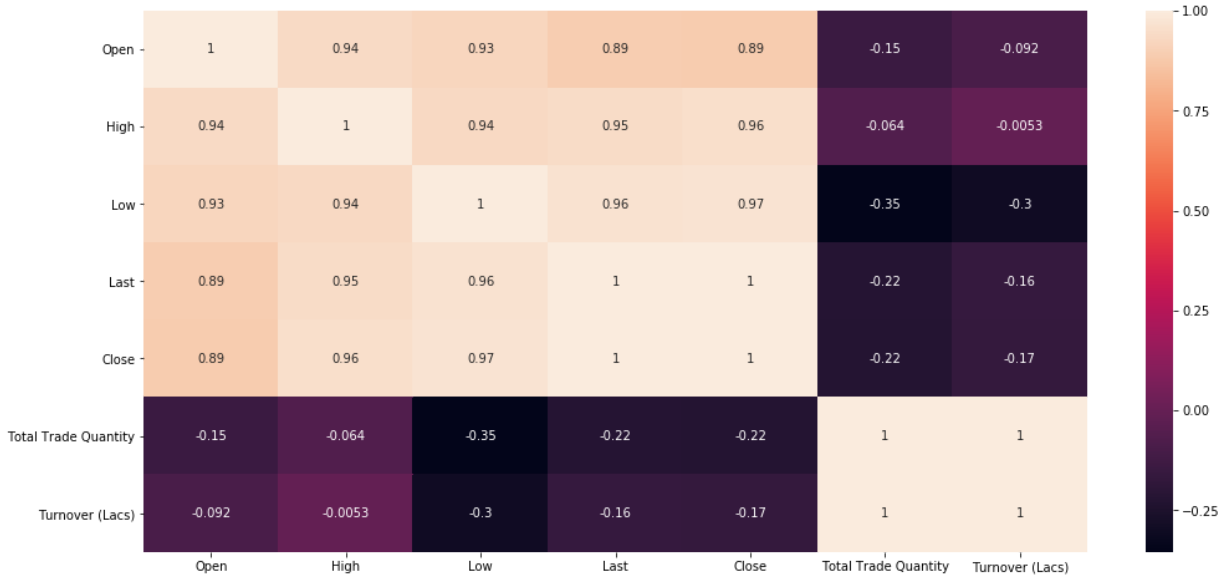
In [11]:
```
#check NULL values

data.isnull().sum()
```

Out[11]:
```
Date                  0
Open                  0
High                  0
Low                   0
Last                  0
Close                 0
Total Trade Quantity  0
Turnover (Lacs)       0
dtype: int64
```

In [12]:
```
#Lets see some correalations between data

import seaborn as sns
plt.figure(1 , figsize = (17 , 8))
cor = sns.heatmap(data.corr(), annot = True)
```



In [13]:
```
# Now we have to divide data in Dependent and Independent variable
# We can see Date column in useul for our prediction but for simplicity we have t
# Now we have to predict open price so this column is out dependent variable beca

x = data.loc[:,'High':'Turnover (Lacs)']
y = data.loc[:,'Open']
```

In [14]:
```
# get top recrods from the dataset
x.head()
```

Out[14]:

	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	1990.0	1968.3	1984.00	1982.40	1610576.0	31868.00
1	2019.4	1971.0	2009.85	2010.85	3270615.0	65568.23
2	2018.0	1985.0	2003.90	2006.75	2501539.0	50131.23
3	2017.0	1979.6	1990.00	1992.70	2321216.0	46373.71
4	2003.9	1973.0	1999.85	1995.20	1680420.0	33463.26

In [15]:
```
# Now we have to split data in training and testing
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.1,random_state
```

In [16]:
```
#  fit our LinearRegression Model
LR = LinearRegression()
LR.fit(x_train,y_train)
```

In [21]:
```
LR.score(x_test,y_test)
```

Out[21]:
```
0.99976474840472
```

In [22]:
```
# ##I given a test data of random day
Test_data = [[2017.0 ,1979.6 ,1990.00 ,1992.70 ,2321216.0 ,46373.71]]
prediction = LR.predict(Test_data)
```

In [24]:
```
print(prediction)
```

```
[2001.75159573]
```

In [25]:
```
# On that day TCS open on 1998.0 price and our model predicted price is 2001.75 s
```