Assignment-1

Aim:

- 1. Get the nrows, ncolumns, datatype, summary stats of each column of a dataframe? Also get the array and list equivalent.
- 2. Extract the row and column number of a particular cell with given criterion?
- **3.** Rename a specific columns in a dataframe?
- **4.** Count the number of missing values in each column?
- **5.** Replace missing values of multiple numeric columns with the mean?
- **6.** Replace one missing value with Imputer class.
- 7. Apply function on existing columns with global variables as additional arguments?
- **8.** Change the order of columns of a dataframe?
- **9.** Set the number of rows and columns displayed in the output?
- **10.** Create a primary key index by combining relevant columns?
- 11. Get the row number of the nth largest value in a column?
- **12.** Find the position of the nth largest value greater than a given value?
- 13. Get the last n rows of a dataframe with row sum > 100?
- **14.** Create a column containing the minimum by maximum of each row?
- **15.** Convert categorical values into numerical using one hot and label encoding?
- **16.** Normalize some columns in a dataframe using minmax normalization?
- 17. Normalize some columns in a dataframe using zscore normalization?
- **18.** Import a file from a url into a dataframe link: http://www.fdic.gov/bank/individual/failed/banklist.html
- 19. Print count of unique values in a column
- 20. Sort values by col2 in descending order

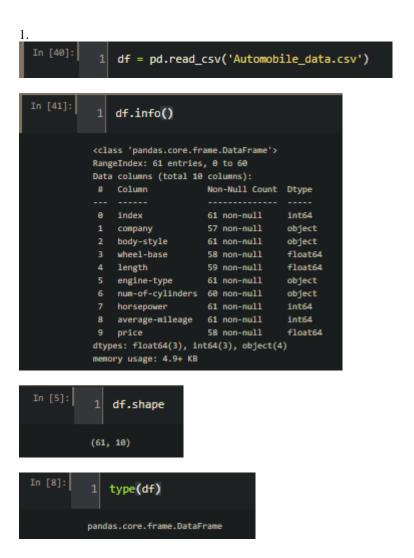
Description:

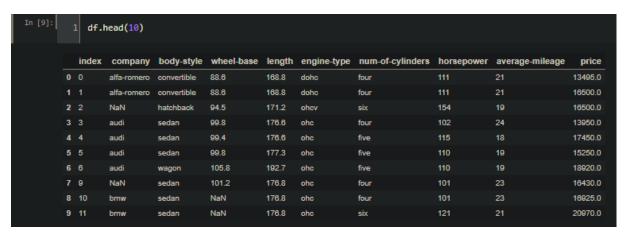
In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating nu-

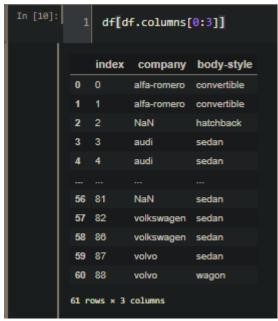
merical tables and time series. It is <u>free</u> software released under the <u>three-clause BSD</u> license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself.

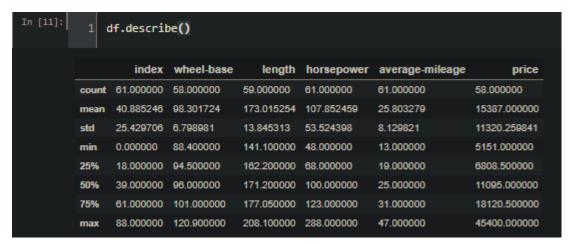
Pandas DataFrame is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns.

Series: Series is a one-dimensional labeled array capable of holding any data type.





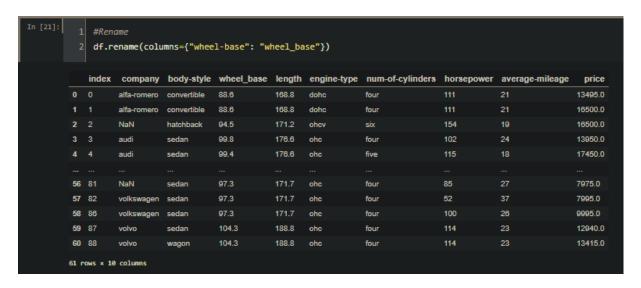




```
# Python
   ## Extract the third row
   df.iloc[2]
   ### or ###
index
company
                     NaN
wheel-base
              hatchback
length
                   171.2
engine-type
                   ohcv
num-of-cylinders
horsepower
average-mileage
                     19
price
                    16500
Name: 2, dtype: object
```

```
In [20]: 1 df[df['company']=='audi']['horsepower']

3 102
4 115
5 110
6 110
Name: horsepower, dtype: 1nt64
```

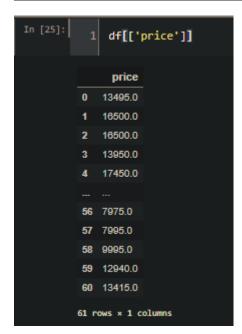


```
In [22]: 1 #Null values in each column

df.isnull().sum()

index 0
company 4
body-style 0
wheel-base 3
length 2
engine-type 0
num-of-cylinders 1
horsepower 0
average-mileage 0
price 3
dtype: int64
```

```
6.
               from sklearn.impute import SimpleImputer
             2 imputer = SimpleImputer(strategy='mean')
               imputer = imputer.fit(df[['price']])
               df[['price']]=imputer.transform(df[['price']])
               df.isnull().sum()
            index
            company
            body-style
            wheel-base
            length
            engine-type
            num-of-cylinders
            horsepower
            average-mileage
            price
            dtype: int64
```



```
In [30]:
              pd.set_option('display.max_rows', 10)
              pd.set_option('display.max_columns', 10)
              pd.set_option('display.width', 6)
             index average-mileage body-style company engine-type horsepower num-of-cylinders price new_col
                                     convertible
                                                alfa-romero doho
                                                                                                       13495.0 26990.0
                                                                                                       16500.0 33000.0
                                     convertible
                                                alfa-romero doho
                                                                                     four
                                                                                                       16500.0 33000.0
                                                                        102
                                                                                                       13950.0 27900.0
                                     sedan
                                                audi
                                                           ohc
                                                                                     four
                                     sedan
                                                            ohc
                                                                                                       17450.0 34900.0
         56 81
                                     sedan
                                                NaN
                                                           ohc
                                                                        85
                                                                                     four
                                                                                                      7975.0 15950.0
         57 82
                                                                        52
                                                                                     four
                                                                                                      7995.0 15990.0
                                                volkswagen ohc
                                     sedan
         58 86
                                                                        100
                                                                                                      9995.0 19990.0
                                                 volkswagen ohc
                                                                                     four
         59 87
                                                                                                       12940.0 25880.0
                                                                                     four
                                                volvo
                                                                                                       13415.0 26830.0
         60 88
                    23
                                                 volvo
                                                           ohc
                                                                        114
                                                                                     four
         61 rows × 9 columns
```

10. In [53]: df2=df.set_index(["index", "company"],append = True, drop = False) 2 df2.head() wheel-base average-mileage index company length ... horsepower price new_col body_style_cat index company 13495.0 -0.273592 0 1 1 alfa-romero alfa-romero 16500.0 -0.271915 0 convertible 88.6 168.8 16500.0 -0.270239 2 NaN NaN 176.6 13950.0 -0.267644 3 3 3 99.8 audi audi 17450.0 -0.267252 3 176.6 18 4 4 audi 99.4 5 rows × 12 columns



```
In [36]:

1     tmp=df
2     tmp['row_sum']=tmp.sum(axis=1)
3     tmp=tmp[tmp['row_sum']>=100].row_sum
4     n=int(input())
5     tmp.tail(n)

2

index company
59 87 volvo 39844.8
60 88 volvo 48470.0
Name: row_sum, dtype: float64
```

14.

```
In [42]:
               tmp=df
               tmp['new_col']=tmp.min(axis=1)/tmp.max(axis=1)
              index company body-style wheel-base length ... num-of-cylinders horsepower average-mileage price new_col
                     alfa-romero convertible 88.6 168.8
                                                                                                     21 13495.0 0.000000
                                                  168.8 ... four 111 21

168.8 ... four 111 21

171.2 ... six 154 19

176.6 ... four 102 24

176.6 ... five 115 18

... ... ... ...

171.7 ... four 85 27

171.7 ... four 52 37

171.7 ... four 100 26

188.8 ... four 114 23
                     alfa-romero convertible 88.6
                                                                                                                       16500.0 0.000061
                                hatchback 94.5
                                                                                                                        16500.0 0.000121
                           sedan 99.8
                                                                                                                     13950.0 0.000215
                     audi
          4 4
                                 sedan 99.4
                                                                                                                        17450.0 0.000229
                     audi
                     NaN sedan 97.3
                    NaN sedan 97.3
volkswagen sedan 97.3
          56 81
                                                                                                                       7975.0 0.003386
          57 82
                                                                                                                       7995.0 0.004628
                     volkswagen sedan
                                                                                                                        9995.0 0.002601
          58 86
                                 sedan 104.3
                                                                                                                     12940.0 0.001777
          59 87
          60 88
                                 wagon
                                                                                                                        13415.0 0.001714
         61 rows × 11 columns
```

```
In [45]:
                1 from sklearn.preprocessing import LabelEncoder
                2 labelencoder = LabelEncoder()
                3 df['body_style_cat'] = labelencoder.fit_transform(df['body-style'])
                4 df.head(5)
                  index company body-style wheel-base length ... horsepower average-mileage price new_col body_style_cat
                           alfa-romero convertible 88.6
                                                                             168.8 ... 111 21
              0 0
                                                                                                                                          13495.0 0.000000 0

        alfa-romero
        convertible
        88.6
        168.8
        ...
        111
        21

        NaN
        hatchback
        94.5
        171.2
        ...
        154
        19

        audi
        sedan
        99.8
        176.6
        ...
        102
        24

        audi
        sedan
        99.4
        176.6
        ...
        115
        18

                                                                                                                                      16500.0 0.000061 0
                                                                                                                                       16500.0 0.000121 2
                                                                                                                                      13950.0 0.000215 3
                                                                                                                                         17450.0 0.000229 3
             5 rows × 12 columns
```

```
In [47]:
          1 from sklearn import preprocessing
          mnorm=preprocessing.MinMaxScaler()
          3 scaled=mnorm.fit_transform(df['price'].values.reshape(-1,1))
          4 pf=pd.DataFrame(scaled)
          5 pf
                 0
         0 0.207309
         1 0.281970
         2 0.281970
         3 0.218614
         4 0.305573
         56 0.070163
         57 0.070660
         58 0.120351
         59 0.193520
         60 0.205322
        61 rows × 1 columns
```

```
import io
     import requests
     url="http://www.fdic.gov/bank/individual/failed/banklist.csv"
  4 s=requests.get(url).content
  5 c=pd.read_csv(io.StringIO(s.decode('utf-8')))
                                       City ST CERT
                                                                   Acquiring Institution Closing Date
                    Bank Name
0
     The First State Bank
                                 Barboursville WV 14361 MVB Bank, Inc.
                                                                                         3-Apr-20
     Ericson State Bank
                                 Ericson
                                            NE 18265 Farmers and Merchants Bank
                                                                                         14-Feb-20
2
     City National Bank of New Jersey Newark
                                            NJ 21111
                                                        Industrial Bank
                                                                                         1-Nov-19
                                                                                         25-Oct-19
     Resolute Bank
                                Maumee
                                            OH 58317 Buckeye State Bank
     Louisa Community Bank
                                 Louisa
                                                58112
                                                        Kentucky Farmers Bank Corporation
                                                                                         25-Oct-19
                                                 32646 Superior Federal, FSB
556 Superior Bank, FSB
                                                                                         27-Jul-01
557 Malta National Bank
                                             OH 6629 North Valley Bank
                                Malta
                                                                                         3-May-01
558 First Alliance Bank & Trust Co.
                                Manchester NH 34264 Southern New Hampshire Bank & Trust 2-Feb-01
559 National State Bank of Metropolis Metropolis IL
                                                        Banterra Bank of Marion
                                                                                         14-Dec-00
                                                 3815
                                            HI 21029 Bank of the Orient
560 Bank of Honolulu
                                Honolulu
                                                                                         13-Oct-00
561 rows × 6 columns
```

In [51]:

1 df['body-style'].value_counts()

sedan 32
hatchback 15
wagon 9
convertible 3
hardtop 2
Name: body-style, dtype: int64

20. srt_df=df.sort_values('average-mileage',ascending=False) srt_df.head(5) index company body-style wheel-base length ... horsepower average-mileage price new_col body_style_cat 13 16 chevrolet hatchback 88.4 5151.0 -0.187685 2 40 53 7099.0 -0.098280 3 nissan sedan 15 18 94.5 158.8 6575.0 -0.197878 3 chevrolet sedan 23 32 155.9 38 5.403161 3 isuzu 94.5 38 NaN 5.225762 3 22 31 isuzu sedan 5 rows × 12 columns