# Anakha R Menon

# CH.EN.U4CSE20103

# CSE-B

# carmpg

In [3]:
```
import pandas as pd
Normal = pd.read_csv("mpg_raw.csv")
```

In [4]:
```
Normal.shape
```

Out[4]:
```
(398, 9)
```

In [5]:
```
Normal.head()
```

Out[5]:

| | mpg | cylinders | displacement | horsepower | weight | acceleration | model_year | origin | name |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 18.0 | 8 | 307.0 | 130.0 | 3504 | 12.0 | 70 | usa | chevrole chevell malibu |
| **1** | 15.0 | 8 | 350.0 | 165.0 | 3693 | 11.5 | 70 | usa | buicl skylarl 32( |
| **2** | 18.0 | 8 | 318.0 | 150.0 | 3436 | 11.0 | 70 | usa | plymoutl satellit |
| **3** | 16.0 | 8 | 304.0 | 150.0 | 3433 | 12.0 | 70 | usa | am rebel ss |
| **4** | 17.0 | 8 | 302.0 | 140.0 | 3449 | 10.5 | 70 | usa | forc torinc |

In [6]:
```
Normal.dtypes
```

Out[6]:
```
mpg             float64
cylinders         int64
displacement    float64
horsepower      float64
weight            int64
acceleration    float64
model_year        int64
origin           object
name             object
dtype: object
```

In [11]:
```
Normal.weight.unique()
```

Out[11]:
```
array([3504, 3693, 3436, 3433, 3449, 4341, 4354, 4312, 4425, 3850, 3563,
       3609, 3761, 3086, 2372, 2833, 2774, 2587, 2130, 1835, 2672, 2430,
       2375, 2234, 2648, 4615, 4376, 4382, 4732, 2264, 2228, 2046, 2634,
       3439, 3329, 3302, 3288, 4209, 4464, 4154, 4096, 4955, 4746, 5140,
       2962, 2408, 3282, 3139, 2220, 2123, 2074, 2065, 1773, 1613, 1834,
       1955, 2278, 2126, 2254, 2226, 4274, 4385, 4135, 4129, 3672, 4633,
       4502, 4456, 4422, 2330, 3892, 4098, 4294, 4077, 2933, 2511, 2979,
       2189, 2395, 2288, 2506, 2164, 2100, 4100, 3988, 4042, 3777, 4952,
       4363, 4237, 4735, 4951, 3821, 3121, 3278, 2945, 3021, 2904, 1950,
       4997, 4906, 4654, 4499, 2789, 2279, 2401, 2379, 2124, 2310, 2472,
       2265, 4082, 4278, 1867, 2158, 2582, 2868, 3399, 2660, 2807, 3664,
       3102, 2875, 2901, 3336, 2451, 1836, 2542, 3781, 3632, 3613, 4141,
       4699, 4457, 4638, 4257, 2219, 1963, 2300, 1649, 2003, 2125, 2108,
       2246, 2489, 2391, 2000, 3264, 3459, 3432, 3158, 4668, 4440, 4498,
       4657, 3907, 3897, 3730, 3785, 3039, 3221, 3169, 2171, 2639, 2914,
       2592, 2702, 2223, 2545, 2984, 1937, 3211, 2694, 2957, 2671, 1795,
       2464, 2572, 2255, 2202, 4215, 4190, 3962, 3233, 3353, 3012, 3085,
       2035, 3651, 3574, 3645, 3193, 1825, 1990, 2155, 2565, 3150, 3940,
       3270, 2930, 3820, 4380, 4055, 3870, 3755, 2045, 1945, 3880, 4060,
       4140, 4295, 3520, 3425, 3630, 3525, 4220, 4165, 4325, 4335, 1940,
       2740, 2755, 2051, 2075, 1985, 2190, 2815, 2600, 2720, 1800, 2070,
       3365, 3735, 3570, 3535, 3155, 2965, 3430, 3210, 3380, 3070, 3620,
       3410, 3445, 3205, 4080, 2560, 2230, 2515, 2745, 2855, 2405, 2830,
       3140, 2795, 2135, 3245, 2990, 2890, 3265, 3360, 3840, 3725, 3955,
       3830, 4360, 4054, 3605, 1925, 1975, 1915, 2670, 3530, 3900, 3190,
       3420, 2200, 2150, 2020, 2595, 2700, 2556, 2144, 1968, 2120, 2019,
       2678, 2870, 3003, 3381, 2188, 2711, 2434, 2110, 2800, 2085, 2335,
       2950, 3250, 1850, 2145, 1845, 2910, 2420, 2500, 2905, 2290, 2490,
       2635, 2620, 2725, 2385, 1755, 1875, 1760, 2050, 2215, 2380, 2320,
       2210, 2350, 2615, 3230, 3160, 2900, 3415, 3060, 3465, 2605, 2640,
       2575, 2525, 2735, 2865, 3035, 1980, 2025, 1970, 2160, 2205, 2245,
       1965, 1995, 3015, 2585, 2835, 2665, 2370, 2790, 2295, 2625],
      dtype=int64)
```

In [16]:
```
Normal.replace(to_replace = "Nrmal",value = "Normal", inplace=True)
Normal
```

Out[16]:

| | mpg | cylinders | displacement | horsepower | weight | acceleration | model_year | origin | na |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 18.0 | 8 | 307.0 | 130.0 | 3504 | 12.0 | 70 | usa | chevı che ma |
| **1** | 15.0 | 8 | 350.0 | 165.0 | 3693 | 11.5 | 70 | usa | b sky |
| **2** | 18.0 | 8 | 318.0 | 150.0 | 3436 | 11.0 | 70 | usa | plym sate |
| **3** | 16.0 | 8 | 304.0 | 150.0 | 3433 | 12.0 | 70 | usa | rebe |
| **4** | 17.0 | 8 | 302.0 | 140.0 | 3449 | 10.5 | 70 | usa | to |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **393** | 27.0 | 4 | 140.0 | 86.0 | 2790 | 15.6 | 82 | usa | must |
| **394** | 44.0 | 4 | 97.0 | 52.0 | 2130 | 24.6 | 82 | europe | pic |
| **395** | 32.0 | 4 | 135.0 | 84.0 | 2295 | 11.6 | 82 | usa | dc ramp |
| **396** | 28.0 | 4 | 120.0 | 79.0 | 2625 | 18.6 | 82 | usa | rai |
| **397** | 31.0 | 4 | 119.0 | 82.0 | 2720 | 19.4 | 82 | usa | chev |

398 rows × 9 columns

In [17]:
```
Normal['mpg'].value_counts()
```

Out[17]:
```
13.0    20
14.0    19
18.0    17
15.0    16
26.0    14
        ..
31.9     1
16.9     1
18.2     1
22.3     1
44.0     1
Name: mpg, Length: 129, dtype: int64
```

In [24]:
```
null = (Normal.isnull().sum()/len(Normal))*100
null
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [24], in <cell line: 3>()
      1 null = (Normal.isnull().sum()/len(Normal))*100
      2 null
----> 3 d=df.isnull().sum()

NameError: name 'df' is not defined
```

In [25]:
```
d=Normal.isnull().sum()
```

In [30]:
```
# checking for missing values
mis_val = Normal.isnull().sum().any()
dff = d.fillna(d.mean())
dff
```

Out[30]:
```
mpg             0
cylinders       0
displacement    0
horsepower      6
weight          0
acceleration    0
model_year      0
origin          0
name            0
dtype: int64
```

In [ ]:
```

```