# ASSIGNMENT 1

In [1]:

```
import pandas as pd
```

In [2]:

```
data = {'Name': ['Jai', 'Princi', 'Gaurav',
                 'Anuj', 'Ravi', 'Natasha', 'Riya'],
        'Age': [17, 17, 18, 17, 18, 17, 17],
        'Gender': ['M', 'F', 'M', 'M', 'M', 'F', 'F'],
        'Marks': [90, 76, 'NaN', 74, 65, 'NaN', 71]}
```

In [3]:

```
df = pd.DataFrame(data)
df
```

Out[3]:

| | Name | Age | Gender | Marks |
|---|---|---|---|---|
| **0** | Jai | 17 | M | 90 |
| **1** | Princi | 17 | F | 76 |
| **2** | Gaurav | 18 | M | NaN |
| **3** | Anuj | 17 | M | 74 |
| **4** | Ravi | 18 | M | 65 |
| **5** | Natasha | 17 | F | NaN |
| **6** | Riya | 17 | F | 71 |

In [4]:

```
c = avg = 0
for ele in df['Marks']:
    if str(ele).isnumeric():
        c += 1
        avg += ele
avg /= c
```

In [5]:

```
df = df.replace(to_replace="NaN",
                value=avg)
```

In [6]:

```
df
```

Out[6]:

| | Name | Age | Gender | Marks |
|---|---|---|---|---|
| **0** | Jai | 17 | M | 90.0 |
| **1** | Princi | 17 | F | 76.0 |
| **2** | Gaurav | 18 | M | 75.2 |
| **3** | Anuj | 17 | M | 74.0 |
| **4** | Ravi | 18 | M | 65.0 |
| **5** | Natasha | 17 | F | 75.2 |
| **6** | Riya | 17 | F | 71.0 |

In [7]:

```
df['Gender'] = df['Gender'].map({'M': 0,
                                 'F': 1, }).astype(float)
```

In [8]:

```
df
```

Out[8]:

| | Name | Age | Gender | Marks |
|---|---|---|---|---|
| **0** | Jai | 17 | 0.0 | 90.0 |
| **1** | Princi | 17 | 1.0 | 76.0 |
| **2** | Gaurav | 18 | 0.0 | 75.2 |
| **3** | Anuj | 17 | 0.0 | 74.0 |
| **4** | Ravi | 18 | 0.0 | 65.0 |
| **5** | Natasha | 17 | 1.0 | 75.2 |
| **6** | Riya | 17 | 1.0 | 71.0 |

In [9]:

```
df = df[df['Marks'] >= 75]
```

In [10]:

```
df = df.drop(['Age'], axis=1)
```

In [11]:

```
df
```

Out[11]:

| | Name | Gender | Marks |
|---|---|---|---|
| **0** | Jai | 0.0 | 90.0 |
| **1** | Princi | 1.0 | 76.0 |
| **2** | Gaurav | 0.0 | 75.2 |
| **5** | Natasha | 1.0 | 75.2 |

In [12]:

```
details = pd.DataFrame({
    'ID': [101, 102, 103, 104, 105, 106,
           107, 108, 109, 110],
    'NAME': ['Jagroop', 'Praveen', 'Harjot',
             'Pooja', 'Rahul', 'Nikita',
             'Saurabh', 'Ayush', 'Dolly', "Mohit"],
    'BRANCH': ['CSE', 'CSE', 'CSE', 'CSE', 'CSE',
               'CSE', 'CSE', 'CSE', 'CSE', 'CSE']})
```

In [13]:

```
print(details)
```

```
    ID     NAME BRANCH
0  101  Jagroop    CSE
1  102  Praveen    CSE
2  103   Harjot    CSE
3  104    Pooja    CSE
4  105    Rahul    CSE
5  106   Nikita    CSE
6  107  Saurabh    CSE
7  108    Ayush    CSE
8  109    Dolly    CSE
9  110    Mohit    CSE
```

In [14]:

```
fees_status = pd.DataFrame(
    {'ID': [101, 102, 103, 104, 105,
            106, 107, 108, 109, 110],
     'PENDING': ['5000', '250', 'NIL',
                 '9000', '15000', 'NIL',
                 '4500', '1800', '250', 'NIL']})
```

In [15]:

```
print(fees_status)
```

```
    ID PENDING
0  101    5000
1  102     250
2  103     NIL
3  104    9000
4  105   15000
5  106     NIL
6  107    4500
7  108    1800
8  109     250
9  110     NIL
```

In [16]:

```
print(pd.merge(details, fees_status, on='ID'))
```

```
    ID     NAME BRANCH PENDING
0  101  Jagroop    CSE    5000
1  102  Praveen    CSE     250
2  103   Harjot    CSE     NIL
3  104    Pooja    CSE    9000
4  105    Rahul    CSE   15000
5  106   Nikita    CSE     NIL
6  107  Saurabh    CSE    4500
7  108    Ayush    CSE    1800
8  109    Dolly    CSE     250
9  110    Mohit    CSE     NIL
```

In [17]:

```
car_selling_data = {'Brand': ['Maruti', 'Maruti', 'Maruti',
                              'Maruti', 'Hyundai', 'Hyundai',
                              'Toyota', 'Mahindra', 'Mahindra',
                              'Ford', 'Toyota', 'Ford'],
                    'Year':  [2010, 2011, 2009, 2013,
                              2010, 2011, 2011, 2010,
                              2013, 2010, 2010, 2011],
                    'Sold': [6, 7, 9, 8, 3, 5,
                             2, 8, 7, 2, 4, 2]}
```

In [18]:

```
df = pd.DataFrame(car_selling_data)
```

In [19]:

```
print(df)
```

```
       Brand  Year  Sold
0     Maruti  2010     6
1     Maruti  2011     7
2     Maruti  2009     9
3     Maruti  2013     8
4    Hyundai  2010     3
5    Hyundai  2011     5
6     Toyota  2011     2
7   Mahindra  2010     8
8   Mahindra  2013     7
9       Ford  2010     2
10    Toyota  2010     4
11      Ford  2011     2
```

In [20]:

```
grouped = df.groupby('Year')
print(grouped.get_group(2010))
```

```
       Brand  Year  Sold
0     Maruti  2010     6
4    Hyundai  2010     3
7   Mahindra  2010     8
9       Ford  2010     2
10    Toyota  2010     4
```

In [21]:

```
non_duplicate = df[~df.duplicated('Year')]
```

In [22]:

```
print(non_duplicate)
```

```
    Brand  Year  Sold
0  Maruti  2010     6
1  Maruti  2011     7
2  Maruti  2009     9
3  Maruti  2013     8
```

In [23]:

```
non_duplicate = df[~df.duplicated('Brand')]
```

In [24]:

```
print(non_duplicate)
```

```
      Brand  Year  Sold
0    Maruti  2010     6
4   Hyundai  2010     3
6    Toyota  2011     2
7  Mahindra  2010     8
9      Ford  2010     2
```

In [25]:

```
non_duplicate = df[~df.duplicated('Sold')]
```

In [26]:

```
print(non_duplicate)
```

```
      Brand  Year  Sold
0    Maruti  2010     6
1    Maruti  2011     7
2    Maruti  2009     9
3    Maruti  2013     8
4   Hyundai  2010     3
5   Hyundai  2011     5
6    Toyota  2011     2
10   Toyota  2010     4
```

In [27]:

```
df.isnull()
```

Out[27]:

| | Brand | Year | Sold |
|---|---|---|---|
| **0** | False | False | False |
| **1** | False | False | False |
| **2** | False | False | False |
| **3** | False | False | False |
| **4** | False | False | False |
| **5** | False | False | False |
| **6** | False | False | False |
| **7** | False | False | False |
| **8** | False | False | False |
| **9** | False | False | False |
| **10** | False | False | False |
| **11** | False | False | False |