step 1. library importing

In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

step 2. data loading

- data can be of
  - excel file
  - database
  - csv file
  - etc

In [ ]:
```
report = pd.read_excel('report1.xlsx')
```

step 3. understanding the data

- info()
- describe()
- head()
- tail()
- shape
- columns

In [ ]:
```
report.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 6 columns):
 #   Column   Non-Null Count  Dtype 
---  ------   --------------  ----- 
 0   Name     15 non-null     object
 1   Eng      15 non-null     int64 
 2   Hindi    15 non-null     int64 
 3   Science  15 non-null     int64 
 4   Maths    15 non-null     int64 
 5   EVS      15 non-null     int64 
dtypes: int64(5), object(1)
memory usage: 848.0+ bytes
```

In [ ]:
```
report.describe() # default is numeric columns only
```

Out[ ]:

	Eng	Hindi	Science	Maths	EVS
count	15.000000	15.000000	15.000000	15.000000	15.000000
mean	72.200000	68.866667	71.533333	70.066667	64.933333
std	16.725515	16.146060	20.099277	15.396969	17.950196
min	48.000000	40.000000	40.000000	49.000000	41.000000
25%	58.500000	57.500000	52.000000	53.500000	50.500000
50%	72.000000	73.000000	78.000000	75.000000	65.000000
75%	88.000000	80.000000	89.000000	82.000000	75.500000
max	93.000000	93.000000	96.000000	91.000000	96.000000

In [ ]:
```
report.head() # default is first 5 rows
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
0	Aastha Srivastava	48	87	91	75	46
1	Anamika Singh	90	73	76	86	83
2	Anjali Kumari	72	45	40	91	72
3	Anjali Yadav	90	79	54	70	65
4	Areeba Shakeel	93	82	96	54	50

In [ ]:
```
report.tail() # default is last 5 rows
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
10	Muskan Raman	71	40	66	49	51
11	Nandini	86	51	91	85	41
12	Ritakshi Singh	91	64	78	53	96
13	Ritima Sharma	48	48	50	77	67
14	Sarika Rajpoot	72	71	78	49	96

In [ ]:
```
report.shape
```

Out[ ]:
```
(15, 6)
```

In [ ]:
```
report.columns.tolist() # list of column names
```

Out[ ]:
```
['Name', 'Eng', 'Hindi', 'Science', 'Maths', 'EVS']
```

step 4. selecting data

- row selection
  - `iloc` (indexed location)
  - `loc` (location)
- column selection
  - dictonary style
  - object style

In [ ]:
```
report.iloc[6] # single row of data
```

Out[ ]:
```
Name       Gauri Verma
Eng                 80
Hindi               67
Science             90
Maths               79
EVS                 68
Name: 6, dtype: object
```

In [ ]:
```
report.iloc[[7,10]]
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
7	Iram Fatima	49	74	48	76	79
10	Muskan Raman	71	40	66	49	51

In [ ]:
```
report.iloc[[1,2,3,4,5]]
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
1	Anamika Singh	90	73	76	86	83
2	Anjali Kumari	72	45	40	91	72
3	Anjali Yadav	90	79	54	70	65
4	Areeba Shakeel	93	82	96	54	50
5	Armaan Fatima	50	81	87	63	62

In [ ]:
```
report.iloc[5:11] # slice from 5 to 11
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
5	Armaan Fatima	50	81	87	63	62
6	Gauri Verma	80	67	90	79	68
7	Iram Fatima	49	74	48	76	79
8	Jaya Singh	67	93	88	91	57
9	Kirti Mishra	76	78	40	53	41
10	Muskan Raman	71	40	66	49	51

In [ ]:
```
report.iloc[10:15]
```

Out[ ]:

	Name	Eng	Hindi	Science	Maths	EVS
10	Muskan Raman	71	40	66	49	51
11	Nandini	86	51	91	85	41
12	Ritakshi Singh	91	64	78	53	96
13	Ritima Sharma	48	48	50	77	67
14	Sarika Rajpoot	72	71	78	49	96

syntax for iloc

```
df.iloc[row_indexes, column_indexes]
```

In [ ]:
```
report.iloc[:3,[1,2,3]]
```

Out[ ]:

	Eng	Hindi	Science
0	48	87	91
1	90	73	76
2	72	45	40

In [ ]:
```
report['Eng'] # single column is called as series
```

Out[ ]:
```
0     48
1     90
2     72
3     90
4     93
5     50
6     80
7     49
8     67
9     76
10    71
11    86
12    91
13    48
14    72
Name: Eng, dtype: int64
```

In [ ]:
```
cols_to_see = ['Eng','Maths','Science']
report[cols_to_see]
```

Out[ ]:

	Eng	Maths	Science
0	48	75	91
1	90	86	76
2	72	91	40
3	90	70	54
4	93	54	96
5	50	63	87
6	80	79	90
7	49	76	48
8	67	91	88
9	76	53	40
10	71	49	66
11	86	85	91
12	91	53	78
13	48	77	50
14	72	49	78

In [ ]:
```
report[['Name','Eng','Hindi']]
```

Out[ ]:

	Name	Eng	Hindi
0	Aastha Srivastava	48	87
1	Anamika Singh	90	73
2	Anjali Kumari	72	45
3	Anjali Yadav	90	79
4	Areeba Shakeel	93	82
5	Armaan Fatima	50	81
6	Gauri Verma	80	67
7	Iram Fatima	49	74
8	Jaya Singh	67	93
9	Kirti Mishra	76	78
10	Muskan Raman	71	40
11	Nandini	86	51
12	Ritakshi Singh	91	64
13	Ritima Sharma	48	48
14	Sarika Rajpoot	72	71

step 5. data preprocessing

In [ ]:
```
report[['Eng','Hindi','Science','Maths','EVS']].sum(axis=1) # sum of all
```

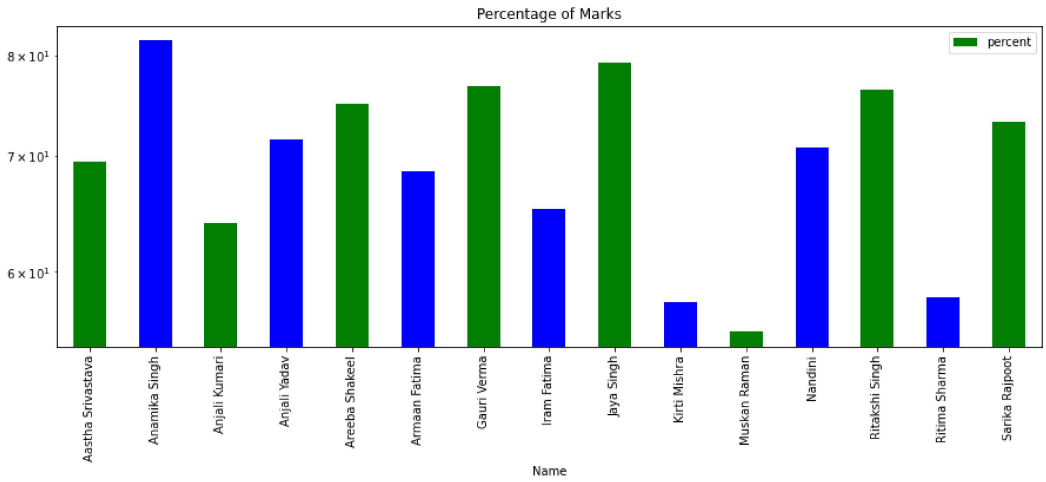
Out[ ]:
```
0     347
1     408
2     320
3     358
4     375
5     343
6     384
7     326
8     396
9     288
10    277
11    354
12    382
13    290
14    366
dtype: int64
```

In [ ]:
```
report['Total'] = report[['Eng','Hindi','Science','Maths','EVS']].sum(ax
```

In [ ]:
```
report['percent'] = report['Total'] / 5
```

In [ ]:
```
report.plot(kind='bar',
            x='Name',
            y='percent',
            figsize=(15,5),
            title='Percentage of Marks',
            color=['green','blue'],)
```

Out[ ]:
```
<AxesSubplot:title={'center':'Percentage of Marks'}, xlabel='Name'>
```



In [ ]: