# Drug Abuse EDA

Project Submission-  DSI Immersive – Pramod Paul

# Goal/Objective

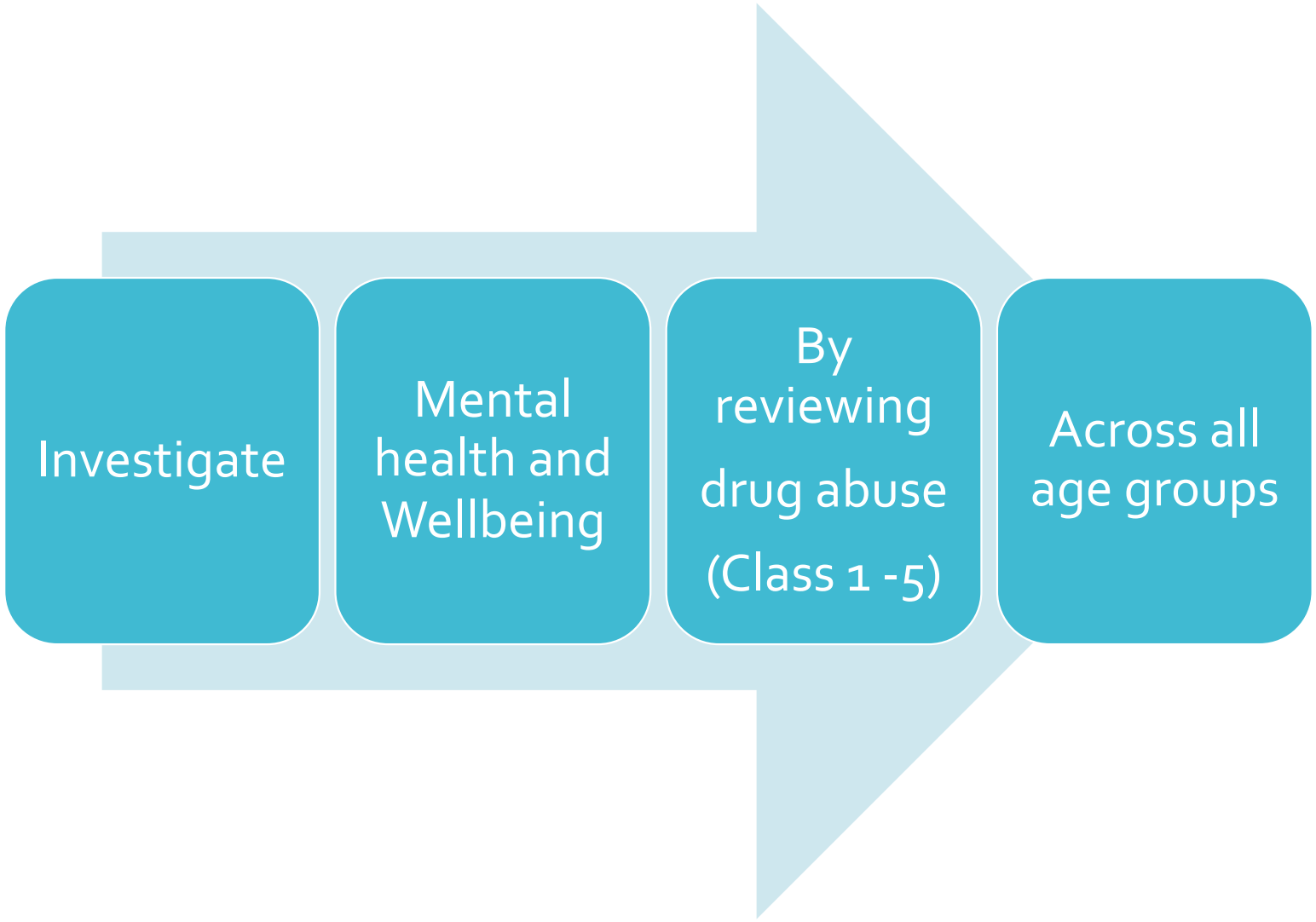
- EDA for Drug abuse CSV
- EDA  for SAT score CSV (not covered in this ppt)

# Drug Abuse CSV

## CSV file – initial thoughts

- 13 drugs, 17 age groups
- Median percent use for each drug in sample
- Number of samples for each age group

# Story behind the data



# EDA

## Step 1

a) Replace missing values with np.nan (zero not used to avoid skew)

b) Convert numeric values to float

c) Alcohol values removed as this skews the results

EDA

Step 2a

Weighted Median frequency for each drug

**Convert median freq to weighted median value as shown below**

$$\frac{\text{Marijuana-Frequency}}{\text{Marijuana percent use} * n}$$

n = sample size

(Had to convert zero values to np.nan to avoid infinity values- (division by zero) )

EDA step 2b:

Classify drugs by prescription drugs category in USA

Controlled Substances Act

**Class 1 (Type 1):** Illegal because they have high abuse potential

Marijuana, Heroin, Crack, Cocaine, Hallucinogen, Stimulant, Tranquilizer

**Class 2 (Type 2):** High potential for abuse and dependence, an accepted medical use, and the potential for severe addiction.

Oxycontin, Meth, Sedative

EDA step 2b:

Classify drugs by prescription drugs category in USA

Controlled Substances Act

**Class 3:** Lower potential for abuse than type 1 and 2, accepted medical use, and mild to moderate possible addiction.

None in our data

**Class 4:** Even lower abuse potential than type 3 drugs, accepted medical use, and limited addiction potential.

None in our data

EDA step 2b:
Class 5

Step2c:
Add all median weights by class

Step2d: Clean up age values

**Class 5:** Low abuse potential, accepted medical use, and a very limited addiction potential. Inhalant & Pain-reliever

**Step 2c:**Add all median-wts by class

**Step2d:**

a)Convert age to float.

b) Take average of values where two are given.

jupyter project2_New Last Checkpoint: 12 minutes ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3

Run | Code

In [191]:
```
1 dfDrugFiltN
```

Out[191]:

| | age | Class1_med_wt | Class2_med_wt | Class5_med_wt |
|---|---|---|---|---|
| 0 | 12.0 | 0.000000 | 11.079342 | 1.067727 |
| 1 | 13.0 | 0.000000 | 23.576351 | 0.385685 |
| 2 | 14.0 | 0.000000 | 11.953797 | 0.179083 |
| 3 | 15.0 | 3.987784 | 3.848106 | 0.135933 |
| 4 | 16.0 | 0.000000 | 4.533563 | 0.069622 |
| 5 | 17.0 | 28.295754 | 3.202295 | 0.100685 |
| 6 | 18.0 | 6.026432 | 2.151383 | 0.142834 |
| 7 | 19.0 | 16.661103 | 12.932973 | 0.153822 |
| 8 | 20.0 | 2.943761 | 1.250205 | 0.161456 |
| 9 | 21.0 | 3.886416 | 1.857177 | 0.131488 |
| 10 | 22.5 | 1.514697 | 7.371166 | 0.116847 |
| 11 | 24.5 | 3.228928 | 2.894459 | 0.090757 |
| 12 | 27.5 | 4.138423 | 2.711187 | 0.313277 |
| 13 | 32.0 | 7.510633 | 7.371198 | 0.435713 |
| 14 | 42.0 | 40.032110 | 8.027782 | 0.489656 |
| 15 | 57.0 | 21.563416 | 17.397400 | 1.842977 |
| 16 | 72.0 | 0.000000 | 0.000000 | 1.633987 |

In [188]:
```
1 dfDrugFiltN.dtypes
```

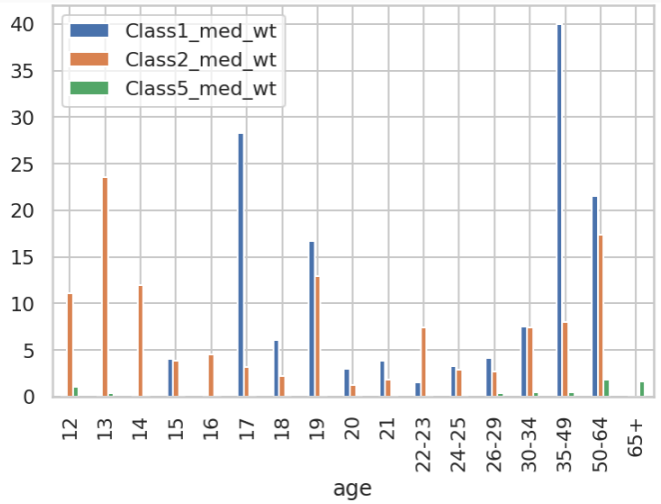
Out[188]:
```
age             float64
```

jupyter project2_New Last Checkpoint: 27 minutes ago (unsaved changes)

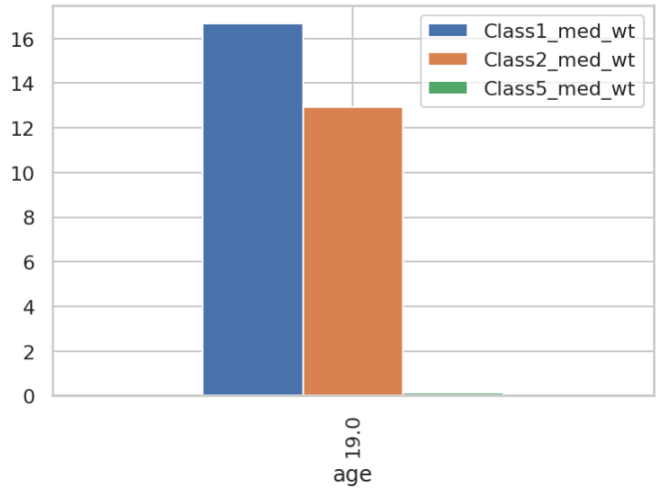
Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3




In [94]:
```
1 dfDrugFiltN.groupby('age').plot(kind='bar', x = 'age', sharey= True)
```

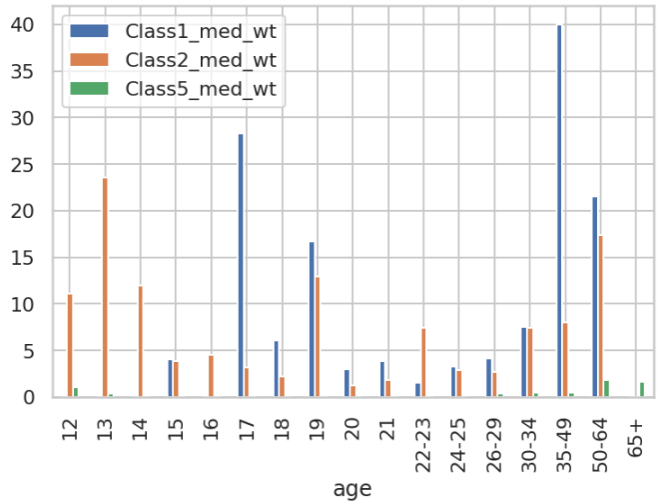


jupyter project2_New Last Checkpoint: 29 minutes ago (autosaved)

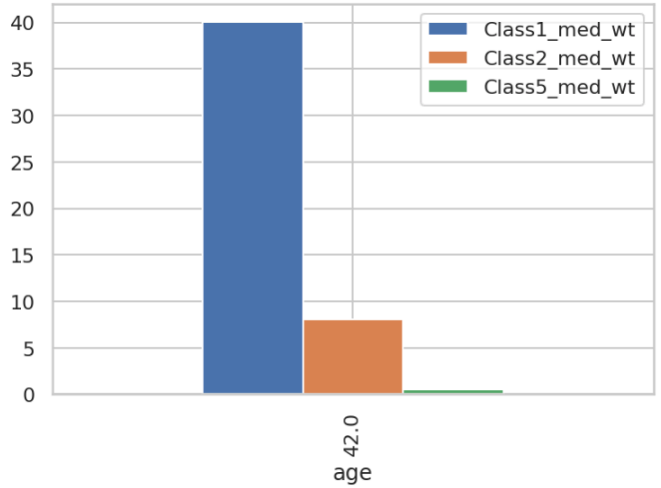
Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3




In [94]:
```
1 dfDrugFiltN.groupby('age').plot(kind='bar', x = 'age', sharey= True)
```

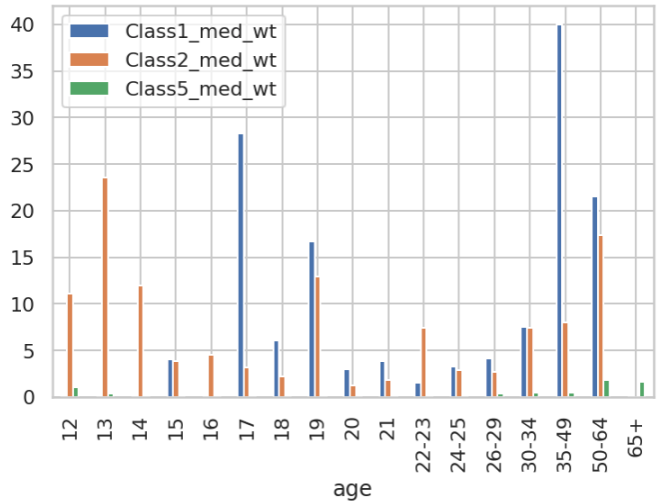


jupyter project2_New Last Checkpoint: 28 minutes ago (autosaved)

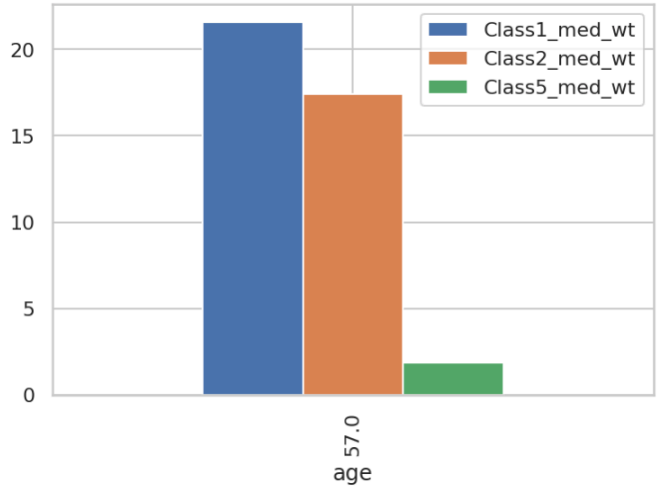
Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3




In [94]:
```
1 dfDrugFiltN.groupby('age').plot(kind='bar', x = 'age', sharey= True)
```



# EDA:

# Null & Alternate HYpothesis

**NULL HYPOTHESIS:**

- **All other things being the same, abuse using Class 2 drugs is same for all age groups than Class 5 drugs**

**ALTERNATE HYPOTHESIS:**

- All other things being the same, abuse using Class 2 drugs is different/greater for all age groups than Class 5 drugs

# EDA:

# Two sample test:

- Two sample test: Comparing two means
- Unpaired data: Data from each of the samples independent of the other
- Goal: Compare responses from each group

# EDA:

# Two sample test:

- Class 2 (experimental) and Class 5 (control) were chosen as they were prescription drugs
- Using Class 1 would have skewed the result.

- stats.ttest_ind(experimental, control)

T- statistic=4.32,

P-value=0.000140

# EDA:

# Cross check by stats.t.cdf

```
lower_tail = stats.t.cdf(-t_stat, (size + size-2), 0, 1)

upper_tail = 1. - stats.t.cdf(t_stat, (size + size -2), 0, 1)

p_value = lower_tail+upper_tail
```

pvalue=0.000140

# Null Hypothesis : Rejected

- All other things being the same, abuse using Class 2 drugs is different/greater for all age groups than Class 5 drugs

High values for age groups 34-49 &

50 -64 similar to age groups 17 and 19

# Recommendations

- For age groups, 35 -49 & 50 -64,
  Doctors to take care to prescribe the class 2 medication in limited quantities and to prescribe only if it is essential.

- Monitor patient behaviour and wellbeing for drug abuse cases.