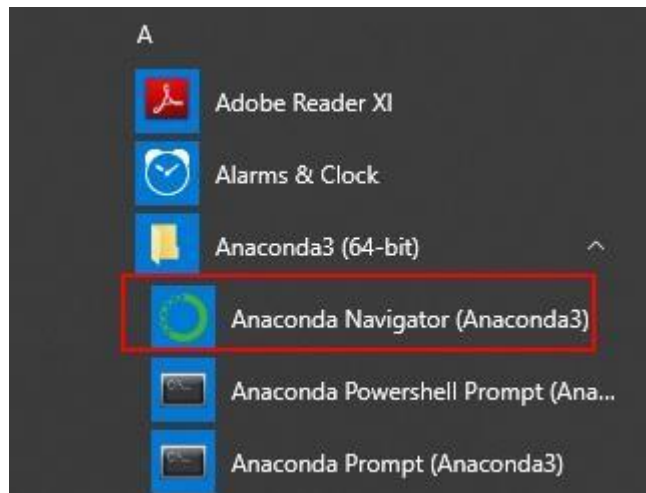


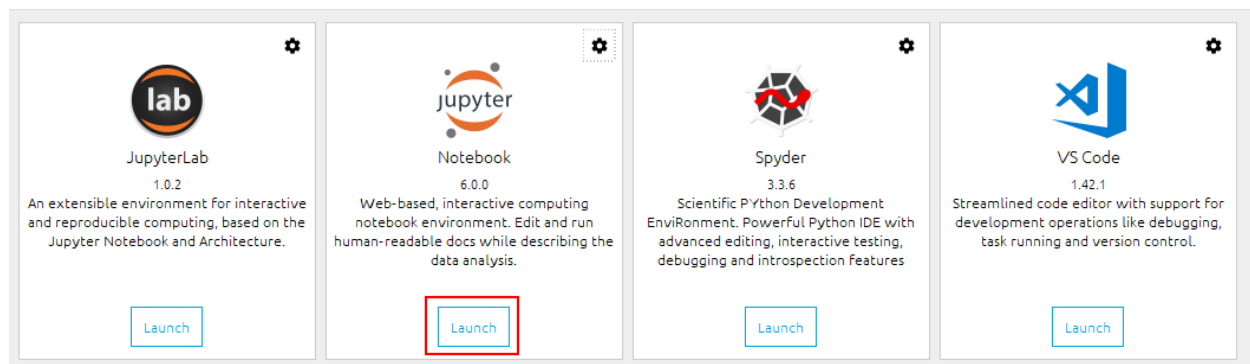
## Module 7: Hands-On: 4

### Data Analysis.

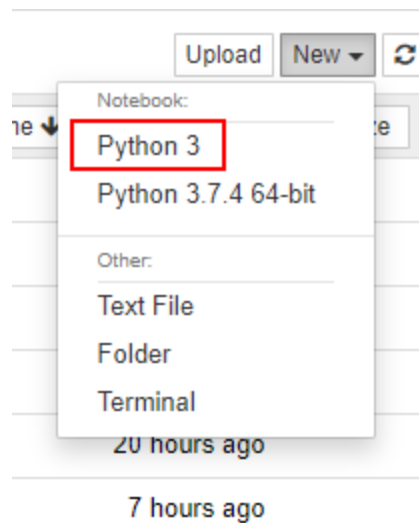
**Step 1:** Open Anaconda Navigator



**Step 2:** Click on Launch button under jupyter notebooks.



**Step 3:** After the notebook opens click on new and Python 3.



**Step 4:** Import the required packages and read data from housing.csv in a dataframe.

```
In [10]: import pandas as pd
```

```
In [11]: data = pd.read_csv('housing.csv')
```

```
In [14]: data.head()
```

```
Out[14]:
```

|   | RM    | LSTAT | PTRATIO | MEDV     |
|---|-------|-------|---------|----------|
| 0 | 6.575 | 4.98  | 15.3    | 504000.0 |
| 1 | 6.421 | 9.14  | 17.8    | 453600.0 |
| 2 | 7.185 | 4.03  | 17.8    | 728700.0 |
| 3 | 6.998 | 2.94  | 18.7    | 701400.0 |
| 4 | 7.147 | 5.33  | 18.7    | 760200.0 |

**Step 5:** Take a look at the shape of data.

```
In [15]: data.shape
```

```
Out[15]: (489, 4)
```

**Step 6:** Take a look at the number of cells that are null in each column.

```
In [16]: data.isnull().sum()
```

```
Out[16]: RM          0  
         LSTAT      0  
         PTRATIO    0  
         MEDV      0  
         dtype: int64
```

**Step 7:** Take a look at the mean, standard deviation, minimum and maximum values is each column.

```
In [17]: data.mean()
```

```
Out[17]: RM          6.240288
         LSTAT       12.939632
         PTRATIO     18.516564
         MEDV      454342.944785
         dtype: float64
```

```
In [18]: data.std()
```

```
Out[18]: RM          0.643650
         LSTAT       7.081990
         PTRATIO     2.111268
         MEDV      165340.277653
         dtype: float64
```

```
In [19]: data.min()
```

```
Out[19]: RM          3.561
         LSTAT       1.980
         PTRATIO     12.600
         MEDV      105000.000
         dtype: float64
```

```
In [20]: data.max()
```

```
Out[20]: RM          8.398
         LSTAT       37.970
         PTRATIO     22.000
         MEDV      1024800.000
         dtype: float64
```

**Step 8:** Use the describe method to check all statistically significant information about data.

```
In [21]: data.describe()
```

```
Out[21]:
```

|       | RM         | LSTAT      | PTRATIO    | MEDV         |
|-------|------------|------------|------------|--------------|
| count | 489.000000 | 489.000000 | 489.000000 | 4.890000e+02 |
| mean  | 6.240288   | 12.939632  | 18.516564  | 4.543429e+05 |
| std   | 0.643650   | 7.081990   | 2.111268   | 1.653403e+05 |
| min   | 3.561000   | 1.980000   | 12.600000  | 1.050000e+05 |
| 25%   | 5.880000   | 7.370000   | 17.400000  | 3.507000e+05 |
| 50%   | 6.185000   | 11.690000  | 19.100000  | 4.389000e+05 |
| 75%   | 6.575000   | 17.120000  | 20.200000  | 5.187000e+05 |
| max   | 8.398000   | 37.970000  | 22.000000  | 1.024800e+06 |