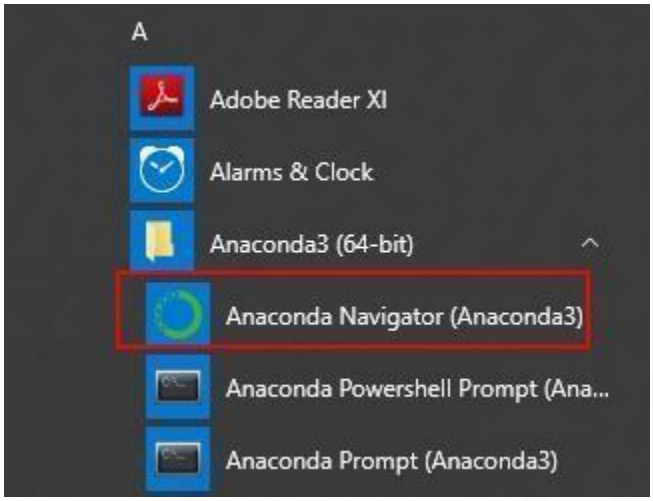
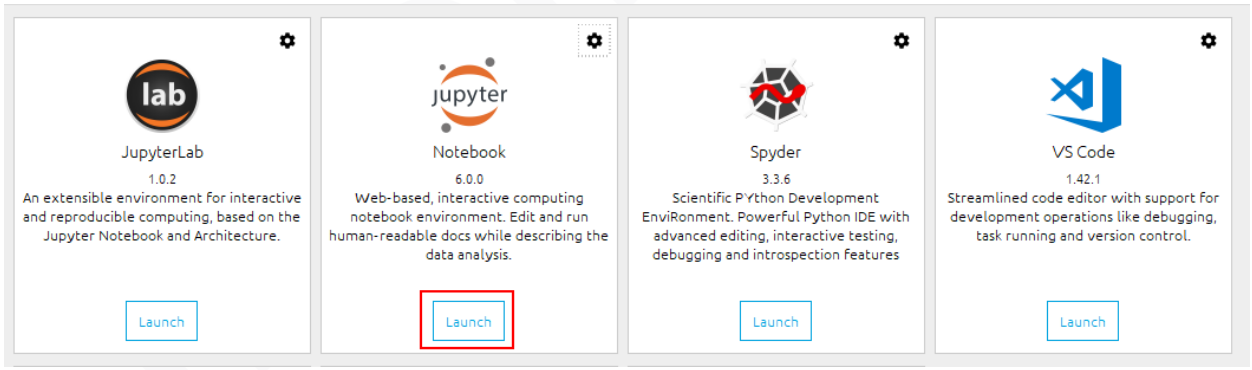
Module 7: Hands-On: 2

Data Manipulation:

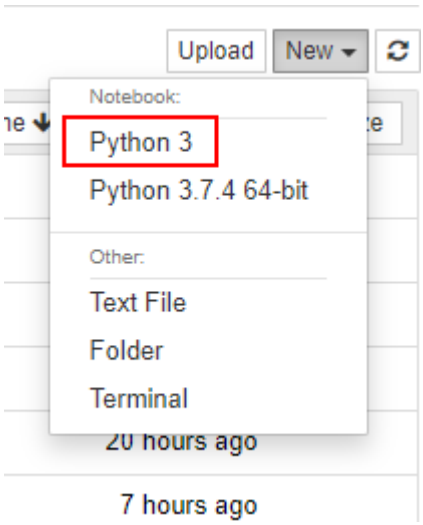
Step 1: Open Anaconda Navigator



Step 2: Click on Launch button under Jupyter Notebook



Step 3: After the notebook opens click on New and Python 3



Step 4: Import the required packages and read the data

```
In [1]: import pandas as pd
        import numpy as np

In [9]: data = pd.read_csv('housing.csv')

In [12]: data.head()
```

Out[12]:

	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: Analyze the shape of data

```
In [13]: data.shape
Out[13]: (489, 4)
```

Step 6: Extract a subset of data using iloc

```
In [16]: data.iloc[:5, :]
```

Out[16]:

	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

```
In [21]: data.iloc[2:5, :2]
```

Out[21]:

	RM	LSTAT
2	7.185	4.03
3	6.998	2.94
4	7.147	5.33

```
In [19]: data.iloc[2:4, 1:2]
```

Out[19]:

	LSTAT
2	4.03
3	2.94

Step 7: Extract a subset of data using loc

```
In [25]: data.loc[:5, 'LSTAT':'MEDV']
```

Out[25]:

	LSTAT	PTRATIO	MEDV
0	4.98	15.3	504000.0
1	9.14	17.8	453600.0
2	4.03	17.8	728700.0
3	2.94	18.7	701400.0
4	5.33	18.7	760200.0
5	5.21	18.7	602700.0

Step 8: Change all values in LSTAT column to 1

```
In [26]: data['LSTAT'] = 1
         data.head()
```

Out[26]:

	RM	LSTAT	PTRATIO	MEDV
0	6.575	1	15.3	504000.0
1	6.421	1	17.8	453600.0
2	7.185	1	17.8	728700.0
3	6.998	1	18.7	701400.0
4	7.147	1	18.7	760200.0

Step 9: Apply function to change LSTAT value to its double (multiply by 2)

```
In [28]: data['LSTAT'] = data['LSTAT'].apply(lambda x : x * 2)
```

```
In [29]: data
```

Out[29]:

	RM	LSTAT	PTRATIO	MEDV
0	6.575	2	15.3	504000.0
1	6.421	2	17.8	453600.0
2	7.185	2	17.8	728700.0
3	6.998	2	18.7	701400.0
4	7.147	2	18.7	760200.0