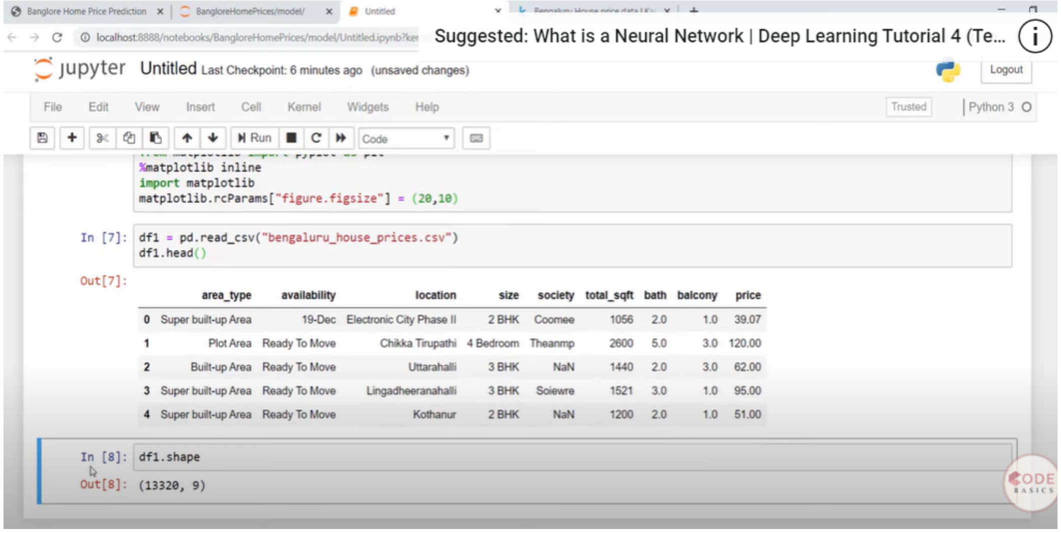
We found this cool video on youtube where he is doing data cleaning using different techniques on his real Estate Price Prediction Project. The following link can take you to that video. It shows the model built by sklearn and linear regression. The pandas data frames are loaded and then handled.
https://www.youtube.com/watch?v=_drqJ9SFCgU
Some pictures from the videos are shown below which shows the different techniques used.

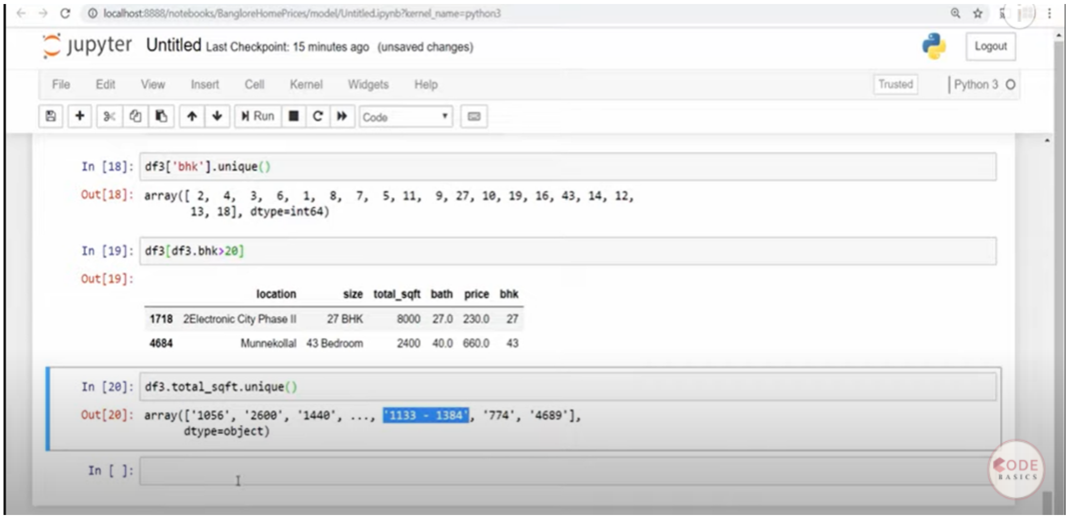

Banglore Home Price Prediction
BangloreHomePrices/model/
Untitled
localhost:8888/notebooks/BangloreHomePrices/model/Untitled.ipynb?ker
Suggested: What is a Neural Network | Deep Learning Tutorial 4 (Te...
jupyter Untitled Last Checkpoint: 6 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

```
%matplotlib inline
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
```

```
In [7]: df1 = pd.read_csv("bengaluru_house_prices.csv")
        df1.head()
Out[7]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

```
In [8]: df1.shape
Out[8]: (13320, 9)
```


localhost:8888/notebooks/BangloreHomePrices/model/Untitled.ipynb?kernel_name=python3
jupyter Untitled Last Checkpoint: 15 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code

```
In [18]: df3['bhk'].unique()
Out[18]: array([ 2,  4,  3,  6,  1,  8,  7,  5, 11,  9, 27, 10, 19, 16, 43, 14, 12,
                13, 18], dtype=int64)

In [19]: df3[df3.bhk>20]
Out[19]:
```

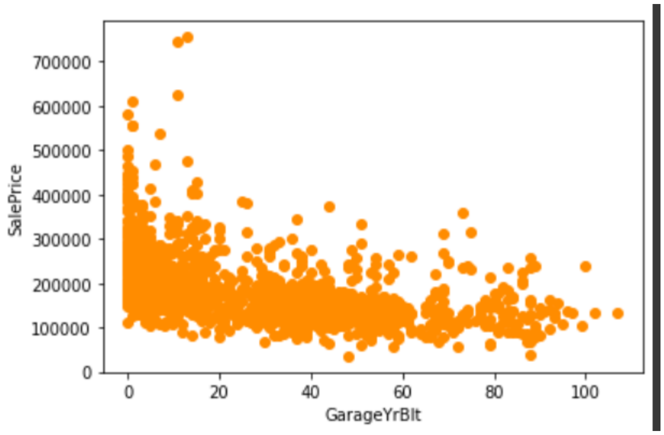
	location	size	total_sqft	bath	price	bhk
1718	2Electronic City Phase II	27 BHK	8000	27.0	230.0	27
4684	Munnekollal	43 Bedroom	2400	40.0	660.0	43

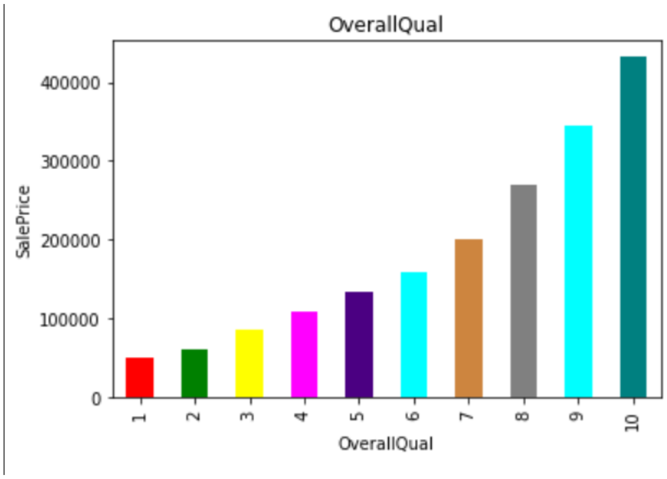
```
In [20]: df3.total_sqft.unique()
Out[20]: array(['1056', '2600', '1440', ..., '1133 - 1384', '774', '4689'],
               dtype=object)

In [ ]:
```

Machine Learning & Data Science Project - 3 : Feature...
codebasics
73K views • 2 years ago

Mix - codebasics
YouTube

Different high level visualization libraries were also used. Some of the interactive graphs were plotted using it.


SalePrice
GarageYrBlt


OverallQual
SalePrice
OverallQual