


```
In [1]: !pip list
```

DEPRECATION: The default format will switch to columns in the future. You can use `--format=(legacy|columns)` (or define a `format=(legacy|columns)` in your `pip.conf` under the `[list]` section) to disable this warning.

- appnope (0.1.0)
- async (0.6.2)
- backports-abc (0.5)
- backports.functools-lru-cache (1.4)
- backports.shutil-get-terminal-size (1.0.0)
- bintrees (2.0.7)
- bleach (2.1.1)
- certifi (2017.11.5)
- chardet (3.0.4)
- cloudpickle (0.5.2)
- configparser (3.5.0)
- cycler (0.10.0)
- dask (0.16.0)
- decorator (4.1.2)
- entrypoints (0.2.3)
- enum34 (1.1.6)
- functools32 (3.2.3.post2)
- html5lib (1.0)
- idna (2.6)
- ipykernel (4.7.0)
- ipython (5.5.0)
- ipython-genutils (0.2.0)
- ipywidgets (7.0.5)
- Jinja2 (2.10)
- jjsonschema (2.6.0)
- jupyter (1.0.0)
- jupyter-client (5.1.0)
- jupyter-console (5.2.0)
- jupyter-core (4.4.0)
- lens (0.4.4)
- locket (0.2.0)
- MarkupSafe (1.0)
- matplotlib (2.1.0)
- mistune (0.8.3)
- nbconvert (5.3.1)
- nbformat (4.4.0)
- notebook (5.2.2)
- numpy (1.13.3)
- pandas (0.21.0)
- pandocfilters (1.4.2)
- partd (0.3.8)
- pathlib2 (2.3.0)
- pexpect (4.3.0)
- pickleshare (0.7.4)
- pip (9.0.1)
- plotly (2.2.3)
- prompt-toolkit (1.0.15)
- ptyprocess (0.5.2)
- Pygments (2.2.0)
- pyparsing (2.2.0)
- python-dateutil (2.6.1)
- pytz (2017.3)
- pyudorandom (1.0.0)
- pyzmq (16.0.3)

```
qtconsole (4.3.1)
requests (2.18.4)
scandir (1.6)
scipy (1.0.0)
seaborn (0.8.1)
setuptools (38.2.4)
simplegeneric (0.8.1)
singledispatch (3.4.0.3)
six (1.11.0)
subprocess32 (3.2.7)
tdigest (0.4.1.0)
terminado (0.8.1)
testpath (0.3.1)
toolz (0.8.2)
tornado (4.5.2)
traitlets (4.3.2)
urllib3 (1.22)
wcwidth (0.1.7)
webencodings (0.5.1)
wheel (0.30.0)
widgetsnbextension (3.0.8)
```

```
In [2]: !python --version
```

```
Python 2.7.12
```

```
In [3]: import lens
```

```
In [4]: import pandas as pd
```

```
#df = pd.read_csv('http://asi-datasets.s3.amazonaws.com/room_occupancy/room_occupancy.csv')
```

```
df = pd.read_csv("https://raw.githubusercontent.com/LuisM78/Occupancy-detection-data/master/datatraining.txt")
```

```
# Split a numerical variable to have additional categorical variables
df['Humidity_cat'] = pd.cut(df['Humidity'], 5,
                             labels=['low', 'medium-low', 'medium',
                                       'medium-high', 'high']).astype('str')
)
```

```
In [5]: print('Number of rows in dataset: {}'.format(len(df.index)))  
df.head()
```

Number of rows in dataset: 8143

Out[5]:

	date	Temperature	Humidity	Light	CO2	HumidityRatio	Occupancy	Humidity_c
1	2015-02-04 17:51:00	23.18	27.2720	426.0	721.25	0.004793	1	medium
2	2015-02-04 17:51:59	23.15	27.2675	429.5	714.00	0.004783	1	medium
3	2015-02-04 17:53:00	23.15	27.2450	426.0	713.50	0.004779	1	medium
4	2015-02-04 17:54:00	23.15	27.2000	426.0	708.25	0.004772	1	medium
5	2015-02-04 17:55:00	23.10	27.2000	426.0	704.50	0.004757	1	medium

```
In [ ]: ls = lens.summarise(df) # This seems to hang...
```