```
import sys
print(sys.version)
```

```
3.6.9 (default, Apr 18 2020, 01:56:04)
[GCC 8.4.0]
```

```
import platform
print(platform.python_version())
```

```
3.6.9
```

```
!python --version
#Note : In terminal dont use !
```

```
Python 3.6.9
```

```
!python -V
#Note : In terminal dont use !
```

```
Python 3.6.9
```

```
#You can check to see if pip is installed by typing the following command
!python3 -m pip --version
#Note : In terminal dont use !
```

```
pip 19.3.1 from /usr/local/lib/python3.6/dist-packages/pip (python 3.6)
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, ca
```

```
!ls "/content/drive/My Drive"
```

```
'Colab Notebooks'   data_science   housing.csv   housing.tgz
```

```
!ls "/content/drive"
```

```
'My Drive'
```

```
!ls "/content"
```

```
drive  sample_data
```

```
!kaggle datasets list -s cancer
```

```
Warning: Looks like you're using an outdated API Version, please consider u
ref                                                     title
------------------------------------------------------  --------------------
xiaotawkaggle/inhibitors                                Cancer Inhibitors
kmader/skin-cancer-mnist-ham10000                       Skin Cancer MNIST: HA
loveall/cervical-cancer-risk-classification             Cervical Cancer Risk
piotrgrabo/breastcancerproteomes                        Breast Cancer Proteom
uciml/breast-cancer-wisconsin-data                      Breast Cancer Wiscons
merishnasuwal/breast-cancer-prediction-dataset          Breast Cancer Predict
fanconic/skin-cancer-malignant-vs-benign                Skin Cancer: Malignan
sajidsaifi/prostate-cancer                              Prostate Cancer
yuqing01/breast-cancer                                  breast cancer
yusufdede/lung-cancer-dataset                           Lung Cancer DataSet
kmader/mias-mammography                                 MIAS Mammography
kmader/crowds-cure-cancer-2017                          Crowds Cure Cancer 20
roustekbio/breast-cancer-csv                            Wisconsin Breast Canc
crawford/gene-expression                                Gene expression datas
kmader/siim-medical-images                              CT Medical Images
gilsousa/habermans-survival-data-set                    Haberman's Survival D
paultimothymooney/breast-histopathology-images          Breast Histopathology
nodoubttome/skin-cancer9-classesisic                    Skin Cancer ISIC
sarahvch/breast-cancer-wisconsin-prognostic-data-set    Breast Cancer Wiscons
kmader/colorectal-histology-mnist                       Colorectal Histology
```

```
!ls "/"
```

```
bin                                          etc    opt    sys
boot                                         home   proc   tensorflow-1.15.2
content                                      lib    root   tmp
datalab                                      lib32  run    tools
dev                                          lib64  sbin   usr
dlib-19.18.0-cp27-cp27mu-linux_x86_64.whl    media  srv    var
dlib-19.18.0-cp36-cp36m-linux_x86_64.whl     mnt    swift
```

```
!df -h
```

```
Filesystem      Size  Used Avail Use% Mounted on
overlay         108G   31G   72G  31% /
tmpfs            64M     0   64M   0% /dev
tmpfs           6.4G     0  6.4G   0% /sys/fs/cgroup
shm             5.9G     0  5.9G   0% /dev/shm
tmpfs           6.4G   12K  6.4G   1% /var/colab
/dev/sda1       114G   33G   82G  29% /etc/hosts
tmpfs           6.4G     0  6.4G   0% /proc/acpi
tmpfs           6.4G     0  6.4G   0% /proc/scsi
tmpfs           6.4G     0  6.4G   0% /sys/firmware
drive            15G  123M   15G   1% /content/drive
```

```
!du -hs * | sort -rh | head -50
```

```
123M    drive
55M     sample_data
```

```
from google.colab import files
files.upload()  #this will prompt you to upload the kaggle.json
```

```
Choose Files no files selected    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving kaggle.json to kaggle.json
{'kaggle.json': b'{"username":"laxmankusuma","key":"3248ec4caadb0a604cd701f
```

```
!pip install -q kaggle
```

```
!kaggle -v
```

```
Kaggle API 1.5.6
```

```
!pip install --upgrade --force-reinstall --no-deps kaggle
```

```
Collecting kaggle
  Downloading https://files.pythonhosted.org/packages/62/ab/bb20f9b9e24f9a6
     |████████████████████████████████| 61kB 1.7MB/s
Building wheels for collected packages: kaggle
  Building wheel for kaggle (setup.py) ... done
  Created wheel for kaggle: filename=kaggle-1.5.6-cp36-none-any.whl size=72
  Stored in directory: /root/.cache/pip/wheels/57/4e/e8/bb28d035162fb8f17f8
Successfully built kaggle
Installing collected packages: kaggle
  Found existing installation: kaggle 1.5.6
    Uninstalling kaggle-1.5.6:
      Successfully uninstalled kaggle-1.5.6
Successfully installed kaggle-1.5.6
```

```
!ls ~/.kaggle
```

```
!pip install -q kaggle
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!ls ~/.kaggle
!chmod 600 /root/.kaggle/kaggle.json  # set permission
```

```
kaggle.json
```

```
!mkdir -p /etc/laxman
!cp kaggle.json /etc/laxman
!ls /etc/laxman
!chmod 777 /etc/laxman/kaggle.json  # set permission
```

```
kaggle.json
```

```
!df -h
```

```
Filesystem      Size  Used Avail Use% Mounted on
overlay         108G   31G   72G  31% /
tmpfs            64M     0   64M   0% /dev
tmpfs           6.4G     0  6.4G   0% /sys/fs/cgroup
shm             5.9G     0  5.9G   0% /dev/shm
tmpfs           6.4G   12K  6.4G   1% /var/colab
/dev/sda1       114G   33G   82G  29% /etc/hosts
tmpfs           6.4G     0  6.4G   0% /proc/acpi
tmpfs           6.4G     0  6.4G   0% /proc/scsi
tmpfs           6.4G     0  6.4G   0% /sys/firmware
drive            15G  123M   15G   1% /content/drive
```

```
!ls "/etc/laxman"
```

```
ls: cannot access '/etc/laxman': No such file or directory
```

```
!ls "/root"
```

```
import os
import tarfile
import urllib
```

```
DOWNLOAD_ROOT = "https://raw.githubusercontent.com/ageron/handson-ml2/master/"
# HOUSING_PATH = os.path.join("datasets", "housing") # use this to download data
HOUSING_PATH = "../content/drive/My Drive/" # this is google drive path
HOUSING_URL = DOWNLOAD_ROOT + "datasets/housing/housing.tgz"
```

```
HOUSING_URL
```

```
'https://raw.githubusercontent.com/ageron/handson-ml2/master/datasets/housing/housing.tgz'
```

```
HOUSING_PATH
```

```
'../content/drive/My Drive/'
```

```
def fetch_housing_data(housing_url=HOUSING_URL, housing_path=HOUSING_PATH):
  os.makedirs(housing_path, exist_ok=True)
  tgz_path = os.path.join(housing_path, "housing.tgz")
  urllib.request.urlretrieve(housing_url, tgz_path)
  housing_tgz = tarfile.open(tgz_path)
  housing_tgz.extractall(path=housing_path)
  housing_tgz.close()
```

```
fetch_housing_data()
```

```
import pandas as pd
```

```
def load_housing_data(housing_path=HOUSING_PATH):
  csv_path = housing_path+"housing.csv"
  return pd.read_csv(csv_path)
```

```
load_housing_data
```

```
<function __main__.load_housing_data>
```

```
housing = load_housing_data()
housing.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   longitude           20640 non-null  float64
 1   latitude            20640 non-null  float64
 2   housing_median_age  20640 non-null  float64
 3   total_rooms         20640 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  float64
 6   households          20640 non-null  float64
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  float64
 9   ocean_proximity     20640 non-null  object 
dtypes: float64(9), object(1)
memory usage: 1.6+ MB
```

```
housing.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	popu
0	-122.23	37.88	41.0	880.0	129.0	
1	-122.22	37.86	21.0	7099.0	1106.0	
2	-122.24	37.85	52.0	1467.0	190.0	
3	-122.25	37.85	52.0	1274.0	235.0	
4	-122.25	37.85	52.0	1627.0	280.0	

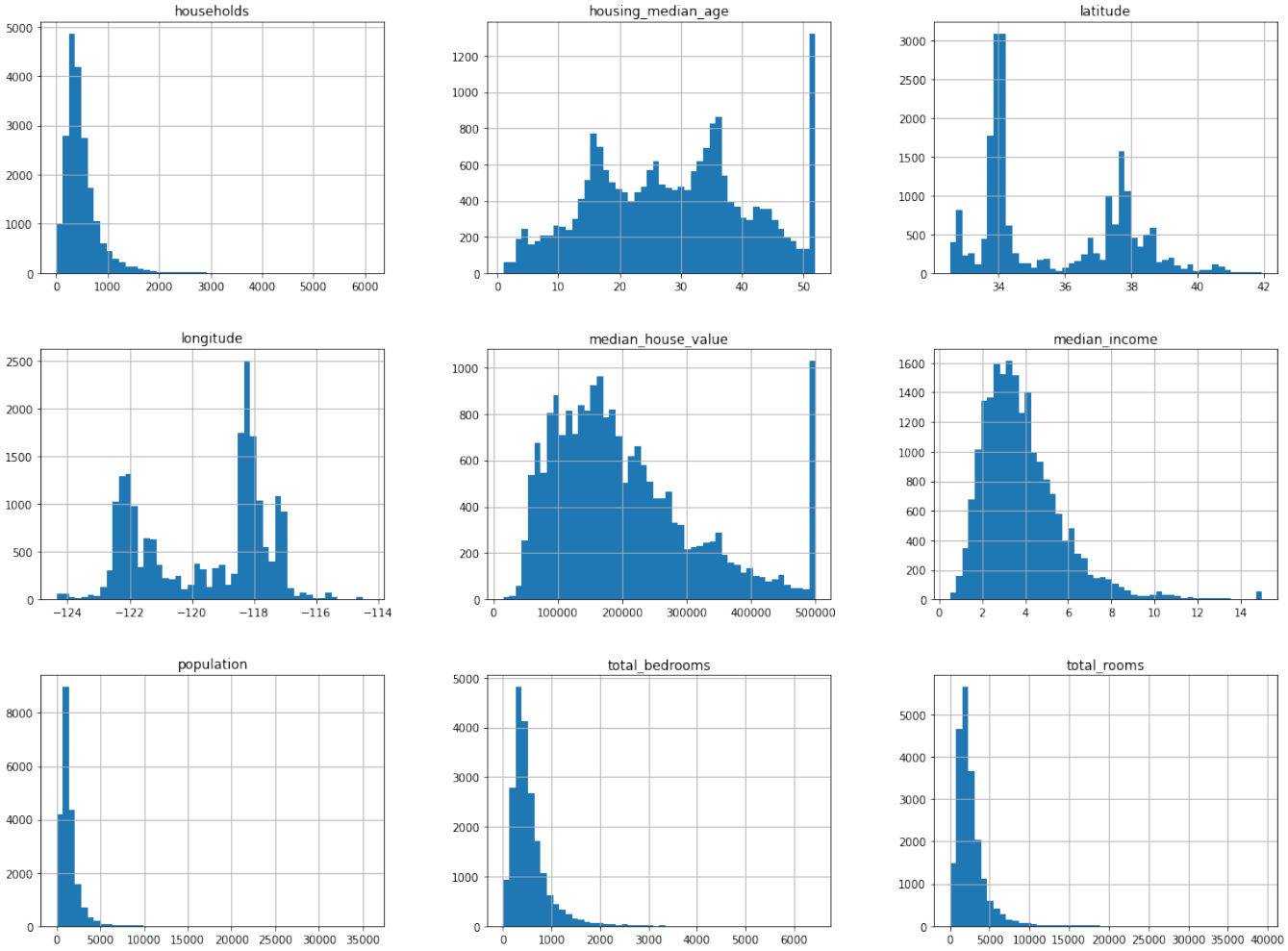
```
housing['ocean_proximity'].value_counts()
```

```
<1H OCEAN     9136
INLAND        6551
NEAR OCEAN    2658
NEAR BAY      2290
ISLAND           5
Name: ocean_proximity, dtype: int64
```

```
housing.describe()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedro
count	20640.000000	20640.000000	20640.000000	20640.000000	20433.000
mean	-119.569704	35.631861	28.639486	2635.763081	537.870
std	2.003532	2.135952	12.585558	2181.615252	421.385
min	-124.350000	32.540000	1.000000	2.000000	1.000
25%	-121.800000	33.930000	18.000000	1447.750000	296.000
50%	-118.490000	34.260000	29.000000	2127.000000	435.000
75%	-118.010000	37.710000	37.000000	3148.000000	647.000
max	-114.310000	41.950000	52.000000	39320.000000	6445.000

```
%matplotlib inline
 # only in a Jupyter notebook
import matplotlib.pyplot as plt
housing.hist(bins=50, figsize=(20,15))
plt.show()
```


households
housing_median_age
latitude
longitude
median_house_value
median_income
population
total_bedrooms
total_rooms