

```

import numpy as np
import pandas as pd

%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
color = sns.color_palette()
sns.set_style('darkgrid')
import warnings
def ignore_warn(*args, **kwargs):
    pass
warnings.warn = ignore_warn

from scipy import stats
from scipy.stats import norm, skew

pd.set_option('display.float_format', lambda x: '{:.3f}'.format(x))

from subprocess import check_output
print(check_output(["ls", "../input"]).decode("utf8"))

train = pd.read_csv('/kaggle/input/house-prices-advanced-regression-techniques/train.csv')
test = pd.read_csv('/kaggle/input/house-prices-advanced-regression-techniques/test.csv')
train.head(5)

```

Id	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Land Contour	Utilities	...	Pool Area	Pool QC	Fence	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type	Sale Condition	Sale Price	
01	60	RL		65.0000	8450	Pave	NaN	Reg	Lvl	All Pub	...	0	NaN	NaN	NaN	0	2	2008	WD	Normal	208500

Id	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Land Contour	Utilities	...	Pool Area	Pool QC	Fence	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type	Sale Condition	Sale Price	
1	2	20	RL	80.000	9600	Pave	NAN	Reg	Lvl	All Pub	...	0	NAN	NAN	NAN	0	5	2007	WD	Normal	181500
2	3	60	RL	68.000	11250	Pave	NAN	IR1	Lvl	All Pub	...	0	NAN	NAN	NAN	0	9	2008	WD	Normal	223500
3	4	70	RL	60.000	9550	Pave	NAN	IR1	Lvl	All Pub	...	0	NAN	NAN	NAN	0	2	2006	WD	Abnormal	140000
4	5	60	RL	84.000	14260	Pave	NAN	IR1	Lvl	All Pub	...	0	NAN	NAN	NAN	0	12	2008	WD	Normal	250000

```
test.head(5)
```

Id	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Land Contour	Utilities	...	Screen Porch	Pool Area	Pool QC	Fence	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type	Sale Condition	
0	1461	20	RH	80.000	11622	Pave	NAN	Reg	Lvl	All Pub	...	120	0	NAN	Mn Priv	NAN	0	6	2010	WD	Normal
1	1462	20	RL	81.000	14267	Pave	NAN	IR1	Lvl	All Pub	...	0	0	NAN	NAN	Gar2	12500	6	2010	WD	Normal
2	1463	60	RL	74.000	13830	Pave	NAN	IR1	Lvl	All Pub	...	0	0	NAN	Mn Priv	NAN	0	3	2010	WD	Normal
3	1464	60	RL	78.000	9978	Pave	NAN	IR1	Lvl	All Pub	...	0	0	NAN	NAN	NAN	0	6	2010	WD	Normal
4	1465	120	RL	43.000	5005	Pave	NAN	IR1	HLS	All Pub	...	144	0	NAN	NAN	NAN	0	1	2010	WD	Normal

```
fig, ax = plt.subplots()
ax.scatter(x = train['GrLivArea'], y = train['SalePrice'])
```

```
plt.ylabel('SalePrice', fontsize=13)
plt.xlabel('GrLivArea', fontsize=13)
plt.show()
```

```
train["SalePrice"] = np.log1p(train["SalePrice"])

sns.distplot(train['SalePrice'] , fit=norm);

(mu, sigma) = norm.fit(train['SalePrice'])
print( '\n mu = {:.2f} and sigma = {:.2f}\n'.format(mu, sigma))

plt.legend(['Normal dist. ($\mu=${:.2f} and $\sigma=${:.2f} )'.format(mu
, sigma)],
           loc='best')
plt.ylabel('Frequency')
plt.title('SalePrice distribution')

fig = plt.figure()
res = stats.probplot(train['SalePrice'], plot=plt)
plt.show()
```