In [1]:
```
print("Hello Universe")
```

```
Hello Universe
```

In [2]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

In [3]:
```
dataset = pd.read_csv("/Users/dhruvkumar/Desktop/exoplanet.eu_catal
```

In [6]:
```
dataset
```

Out[6]:

	# name	planet_status	mass	mass_error_min	mass_error_max	mass_sini	mass_sini_e
0	11 Com b	Confirmed	NaN	NaN	NaN	16.1284	
1	11 Oph b	Confirmed	21.00	3.00	3.00	NaN	
2	11 UMi b	Confirmed	NaN	NaN	NaN	11.0873	
3	14 And b	Confirmed	NaN	NaN	NaN	4.6840	
4	14 Her b	Confirmed	9.10	NaN	NaN	5.2150	
...	...	...	...	...	...	...	
4974	ups And c	Confirmed	9.10	2.93	5.04	1.8000	
4975	ups And d	Confirmed	23.58	2.29	2.93	10.1900	
4976	ups And e	Confirmed	NaN	NaN	NaN	1.0590	
4977	ups Leo b	Confirmed	NaN	NaN	NaN	0.5100	
4978	zet Del B	Confirmed	40.00	5.00	15.00	NaN	

4979 rows × 98 columns

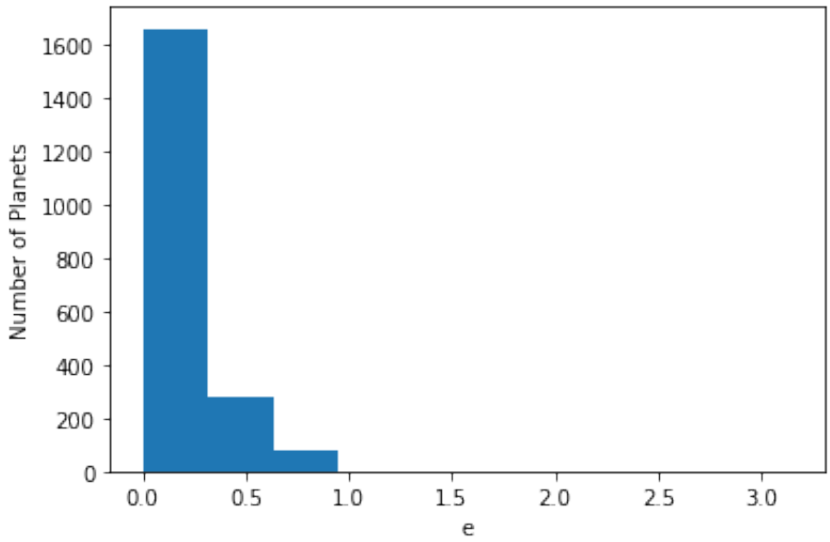
In [8]:
```
e = dataset.eccentricity
```

In [30]:
```
dataset.eccentricity
```

Out[30]:
```
0       0.23100
1           NaN
2       0.08000
3       0.00000
4       0.36900
          ...
4974    0.24450
4975    0.31600
4976    0.00536
4977    0.32000
4978        NaN
Name: eccentricity, Length: 4979, dtype: float64
```

In [87]:
```
plt.hist(e) # ecentricity
plt.xlabel("e")
plt.ylabel("Number of Planets")
```

Out[87]:
```
Text(0, 0.5, 'Number of Planets')
```

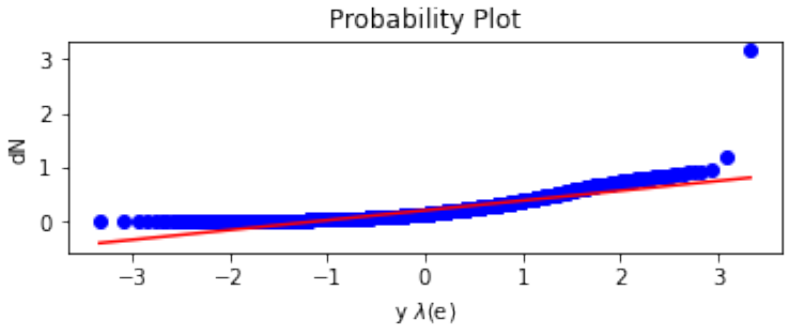


In [88]:
```
import scipy
```

In [89]:
```
from scipy.stats import boxcox
```

In [90]:
```
data = e[e > 0]
fig = plt.figure()
ax1 = fig.add_subplot(211)
e = dataset.eccentricity
prob = stats.probplot(data,dist=stats.norm,plot=ax1)
ax1.set_xlabel('y $\lambda$(e)')
ax1.set_ylabel('dN')
```

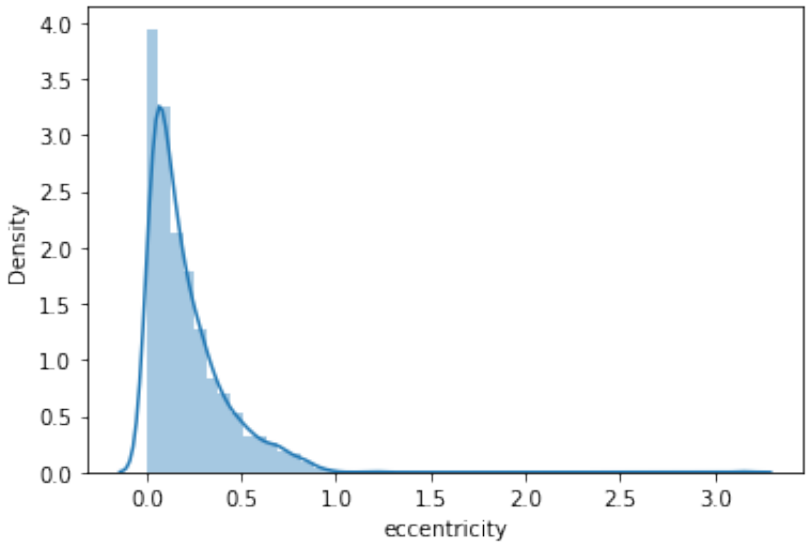
Out[90]:
```
Text(0, 0.5, 'dN')
```



In [105]:
```
#load necessary packages
import numpy as np
from scipy.stats import boxcox
import seaborn as sns

#make this example reproducible
np.random.seed(42)

#generate dataset

#plot the distribution of data values
sns.distplot(data, hist=True, kde=True)
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.p
y:2557: FutureWarning: `distplot` is a deprecated function and wil
l be removed in a future version. Please adapt your code to use ei
ther `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

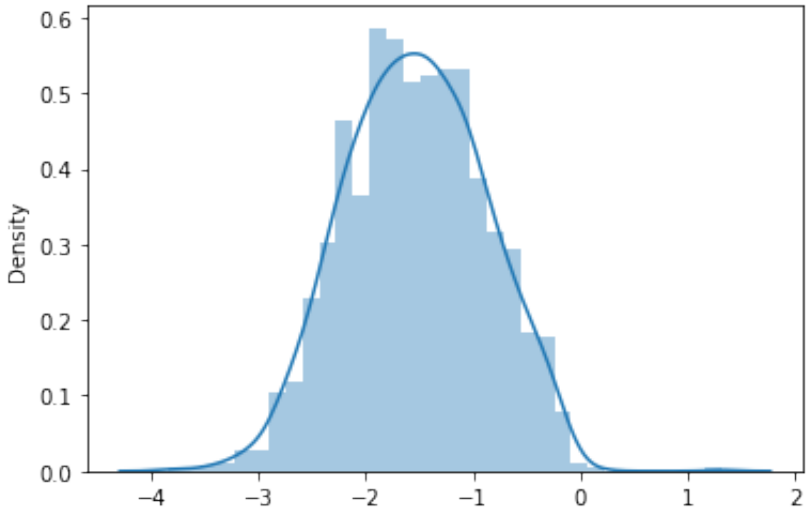
Out[105]:
```
<AxesSubplot:xlabel='eccentricity', ylabel='Density'>
```



In [103]:
```
#perform Box-Cox transformation on original data
transformed_data, best_lambda = boxcox(data)

#plot the distribution of the transformed data values
sns.distplot(transformed_data, hist=True, kde=True)
```

```
/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.p
y:2557: FutureWarning: `distplot` is a deprecated function and wil
l be removed in a future version. Please adapt your code to use ei
ther `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[103]:
```
<AxesSubplot:ylabel='Density'>
```



In [101]:
```
#display optimal lambda value
print(best_lambda)
```

```
0.24271536730189142
```

In [ ]:
```

```