In [17]:
```
#panda
#https://github.com/GokuMohandas/Made-With-ML/blob/4ad626098aca25db5628fe67895e738d5a5
#https://elitedatascience.com/python-cheat-sheet
#bokeh
#https://docs.bokeh.org/en/latest/index.html
#seborn
#https://seaborn.pydata.org/

import pandas as pd
import chardet

with open('FIFA19.csv', 'rb') as rawdata:
    result = chardet.detect(rawdata.read(100000))
result
oscar_demographics = pd.read_csv('FIFA19.csv',encoding='Windows-1252')
oscar_demographics.head()

data_frame = pd.read_csv('FIFA19.csv',encoding='Windows-1252')
data_frame.shape
```

Out[17]:
```
(18207, 89)
```

In [18]:
```
data_frame.describe()
```

Out[18]:

	Unnamed: 0	ID	Age	Overall	Potential	Special	Internati Reputa
count	18207.000000	18207.000000	18206.000000	18206.000000	18207.000000	18207.000000	18159.00
mean	9103.000000	214298.338606	25.122048	66.237449	71.307299	1597.809908	1.11
std	5256.052511	29965.244204	4.670022	6.907059	6.136496	272.586016	0.39
min	0.000000	16.000000	16.000000	46.000000	48.000000	731.000000	1.00
25%	4551.500000	200315.500000	21.000000	62.000000	67.000000	1457.000000	1.00
50%	9103.000000	221759.000000	25.000000	66.000000	71.000000	1635.000000	1.00
75%	13654.500000	236529.500000	28.000000	71.000000	75.000000	1787.000000	1.00
max	18206.000000	246620.000000	45.000000	94.000000	95.000000	2346.000000	5.00

8 rows × 44 columns

In [25]:
```
df1 = pd.DataFrame(data_frame,columns = ['Name','Wage','Value'])

def value_to_float(x):
    if type(x) == float or type(x) == int:
        return x
    if 'K' in x:
        if len(x) > 1:
            return float(x.replace('K', '')) * 1000
        return 1000.0
    if 'M' in x:
        if len(x) > 1:
            return float(x.replace('M', '')) * 1000000
```

```
        return 1000000.0
    if 'B' in x:
        return float(x.replace('B', '')) * 1000000000
    return 0.0

wage = df1['Wage'].replace('[\€]','',regex=True).apply(value_to_float)
value = df1['Value'].replace('[\€]','',regex=True).apply(value_to_float)

df1['Wage'] = wage
df1['Value'] = value

df1['difference'] = df1['Value'] - df1['Wage']
df1.sort_values('difference',ascending=False)
```

Out[25]:

	Name	Wage	Value	difference
2	Neymar Jr	290000.0	118500000.0	118210000.0
0	L. Messi	565000.0	110500000.0	109935000.0
4	K. De Bruyne	355000.0	102000000.0	101645000.0
5	E. Hazard	340000.0	93000000.0	92660000.0
15	P. Dybala	205000.0	89000000.0	88795000.0
...	...	...	...	...
17752	S. Phillips	1000.0	0.0	-1000.0
12192	H. Sulaimani	3000.0	0.0	-3000.0
3550	S. Nakamura	4000.0	0.0	-4000.0
4228	B. Nivet	5000.0	0.0	-5000.0
864	Hilton	18000.0	0.0	-18000.0

18207 rows × 4 columns

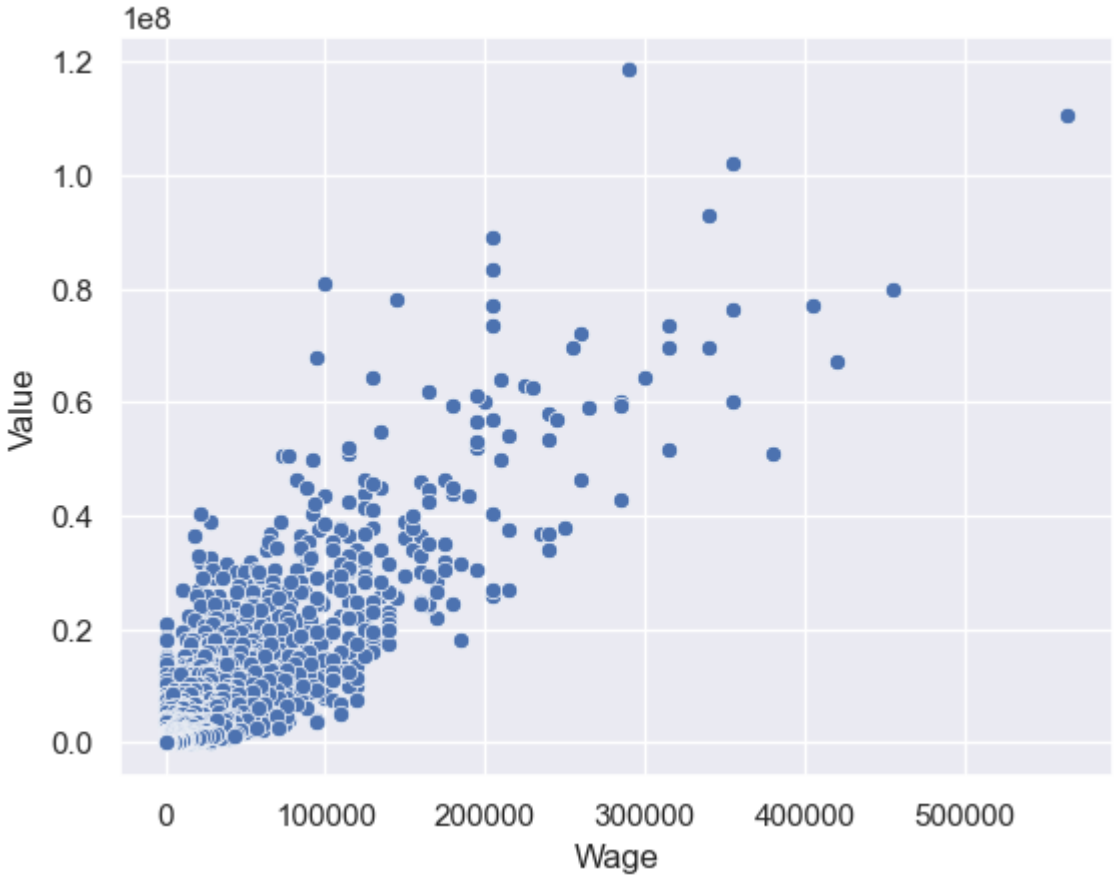
In [27]:

```
import seaborn as sns
sns.set()

graph = sns.scatterplot(x='Wage', y='Value',data=df1)
graph
```

Out[27]:

```
<AxesSubplot:xlabel='Wage', ylabel='Value'>
```



In [33]:
```
from bokeh.plotting import figure,show
from bokeh.models import HoverTool

TOOLTIPS = HoverTool(tooltips=[
    ("index", "$index"),
    ("(Wage,Value)", "(@Wage, @Value)"),
    ("Name", "@Name")]
)

p = figure(title="Soccer 2019", x_axis_label='Wage', y_axis_label='Value', plot_width=
p.circle('Wage', 'Value', size=10, source=df1)
show(p)
```

In [ ]:
```
```