

#this jupyter will show some analysis of the pearsons of the project. pic below shows the first 5 lines of pearson.

```
import pandas as pd
import matplotlib.pyplot as plt
pearsons = pd.read_csv("pearson.txt", sep=":", names=[
    "feature", "value"])
pearsons.head()
fig = plt.figure()
pearsons.head()
```

```
<tr style="text-align: right;">
  <th></th>
  <th>feature</th>
  <th>value</th>
</tr>
```

```

<tr>
  <th>0</th>
  <td>location_id_4428</td>
  <td>0.1299</td>
</tr>
<tr>
  <th>1</th>
  <td>creative_id_387049</td>
  <td>0.0464</td>
</tr>
<tr>
  <th>2</th>
  <td>creative_id_version_387049_1</td>
  <td>0.0464</td>
</tr>
<tr>
  <th>3</th>
  <td>local_hour_ru_ru</td>
  <td>0.0352</td>
</tr>
<tr>
  <th>4</th>
  <td>app_bundle_325782</td>
  <td>0.0294</td>
</tr>

```

```

<matplotlib.figure.Figure at 0x10e1f1090>

```

```

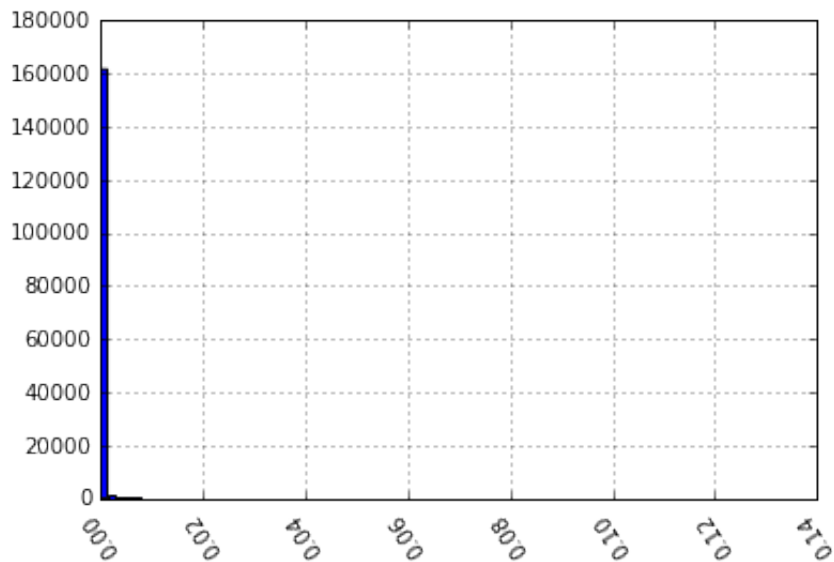
# this give a basic describe of the pearsons.
pearsons["value"] = pearsons.value.apply(lambda x:float(x) if
x!= '0.0d' else 0)
pearsons.value.describe()

```

```
count      164063.000000
mean         0.000133
std          0.000698
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000100
max          0.129900
Name: value, dtype: float64
```

```
# x represents pearson, and y represents count.
pearsons.value.hist(bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x10838ba90>
```



```
#cate represents the column name of feature, and the second
column means the counts of feature appeared.
pearsons['cate'] = pearsons.feature.apply(lambda
x: "_".join(x.split("_")[:2]))
pearsons.groupby('cate').size()
```

```
cate
app_bundle          29459
creative_id         6390
creative_version     45
exchange_id         22
io_id               4418
local_hour          70
location_id        109477
norm_device         14156
order_iab           26
dtype: int64
```

```
# the second column means the sum of pearsons of the cate. this
gives the relation of the column with click.
pearsons.groupby('cate').value.sum()
```

```
cate
app_bundle          4.2744
creative_id         4.4762
creative_version     0.1256
exchange_id         0.0622
io_id               2.9379
local_hour          0.2410
location_id         6.9209
norm_device         2.7626
order_iab           0.0822
Name: value, dtype: float64
```