

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [37]: ab_music = pd.read_csv('ab_music.csv')
ab_music = ab_music.drop('recipient',1)
ab_music.head()
```

Out[37]:

	picture	clickthrough
0	beach	no
1	concert	no
2	concert	no
3	concert	yes
4	concert	no

```
In [38]: ab_music.head()
clickthrough = ab_music[(ab_music.clickthrough == "yes")]
clickthrough.head()
```

Out[38]:

	picture	clickthrough
3	concert	yes
6	beach	yes
14	beach	yes
29	beach	yes
30	beach	yes

```
In [41]: clickthrough.groupby(['picture']).describe()
```

Out[41]:

	clickthrough			
	count	unique	top	freq
<b>picture</b>				
<b>beach</b>	2118	1	yes	2118
<b>concert</b>	2705	1	yes	2705

```
In [42]: clickthrough_no = ab_music[(ab_music.clickthrough == "no")]
clickthrough_no.head()
```

Out[42]:

	picture	clickthrough
0	beach	no
1	concert	no
2	concert	no
4	concert	no
5	concert	no

```
In [43]: clickthrough_no.groupby(['picture']).describe()
```

Out[43]:

	clickthrough			
	count	unique	top	freq
picture				
beach	7882	1	no	7882
concert	7295	1	no	7295

```
In [46]: #clickthrough rate for picture which is concert
c_rate1 = 2705/10000
print(c_rate1)

#clickthrough rate for picture which is beach
c_rate2 = 2118/10000
print(c_rate2)
```

```
0.2705
0.2118
```

We see that the clickthrough rate from the email with picture of "concert" is higher than the one with the picture of beach. Using the chi-square test, it is then our task to check whether or not the difference is significant.

Our null hypothesis says that there is no significant difference in clickthrough rate obtained by two different emails.

```
In [ ]: #Now, we will build the contingency table for our hypothesis

#E(1,1) = ((2705+7295)*4823)/20000 = 2411.5
E11 = 2411.5

#E(1,2) = 7588.5
E12 = 7588.5

#E(2,1) = 2411.5

E21 = 2411.5

#E(2,2) = 7588.5
E22 = 7588.5

#Now we will calculate the Test statistics of chi-square test

Test statistics value of chi-square test = 94.14.
```

We have calculated the test statistics. now, we need to compare it with the table value of probability.

Reference - <https://medium.com/bukalapak-data/meet-the-engine-of-a-b-testing-chi-square-test-30e8a8ab44c5>  
(<https://medium.com/bukalapak-data/meet-the-engine-of-a-b-testing-chi-square-test-30e8a8ab44c5>)

Our degree of freedom is  $2-1*2-1 = 1$ .

Therefore, we know that our test statistics (94.14) is greater than the table value (3.84). Thus we reject the null hypothesis. There is enough evidence to state that there is a significant difference in clickthrough rates obtained by the two designs of emails.

Since the clickthrough rate of the concert email is higher than beach as per the above, we can conclude that the concert email design is the winner of this A/B testing — the concert email design is better for Winterland email than the beach email design.