

# Appendix 10: Hypothesis 2 Test

In [7]:

```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
%matplotlib inline
import matplotlib
import matplotlib.pyplot as plt
import statsmodels.api as sm
```

In [2]:

```
df=pd.read_csv('data_final.csv')
df.head()
```

Out[2]:

	Unnamed: 0	track	artist	uri	danceability	energy
0	0	Wild Things	Alessia Cara	spotify:track:2ZyuwVvV6Z3XJaXIFbspeE	0.741	0.626
1	1	Love Someone	Lukas Graham	spotify:track:2JqnpexIO9dmvjUMCaLCLJ	0.550	0.416
2	2	Here's To Never Growing Up	Avril Lavigne	spotify:track:0qwcGscxUHGZTgq0zcaqk1	0.482	0.876
3	3	Crawling Back To You	Daughtry	spotify:track:6BDtTzjbJ5kKKSWcJT8MIX	0.438	0.916
4	4	Faster	Matt Nathanson	spotify:track:6plKFdrBnKF0y3CRuceTDh	0.742	0.856

5 rows x 32 columns



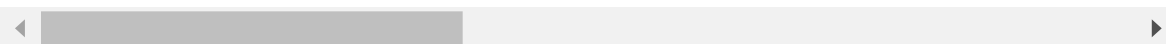
In [3]:

```
df=df.iloc[:,1:]
df.head()
```

Out[3]:

	track	artist	uri	danceability	energy	key	loudness
0	Wild Things	Alessia Cara	spotify:track:2ZyuwVvV6Z3XJaXIFbspeE	0.741	0.626	1	-1.5
1	Love Someone	Lukas Graham	spotify:track:2JqnpexIO9dmvjUMCaLCLJ	0.550	0.415	9	-1.5
2	Here's To Never Growing Up	Avril Lavigne	spotify:track:0qwcGscxUHGZTgq0zcaqk1	0.482	0.873	0	-1.5
3	Crawling Back To You	Daughtry	spotify:track:6BDtTzjbJ5kKKS WCJT8MIX	0.438	0.919	0	-1.5
4	Faster	Matt Nathanson	spotify:track:6pIKFdrBnKF0y3CRuceTDh	0.742	0.853	9	-1.5

5 rows × 31 columns



In [8]:

```
df_1=df
y=df_1['target']
x=df_1.iloc[:,[3,4,10,12]]
x['intercept']=1
x.head()
```

D:\Users\CHRIS\Anaconda3\lib\site-packages\ipykernel\_launcher.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [http://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
after removing the cwd from sys.path.

Out[8]:

	danceability	energy	instrumentalness	valence	intercept
0	0.741	0.626	0.000000	0.706	1
1	0.550	0.415	0.000000	0.274	1
2	0.482	0.873	0.000000	0.737	1
3	0.438	0.919	0.000000	0.195	1
4	0.742	0.853	0.000005	0.950	1

In [9]:

```
logit = sm.Logit(y, x)
result=logit.fit()
result.summary()
```

Optimization terminated successfully.  
Current function value: 0.523064  
Iterations 9

Out[9]:

Logit Regression Results

Dep. Variable:	target	No. Observations:	33354			
Model:	Logit	Df Residuals:	33349			
Method:	MLE	Df Model:	4			
Date:	Sat, 09 May 2020	Pseudo R-squ.:	0.2453			
Time:	22:30:28	Log-Likelihood:	-17446.			
converged:	True	LL-Null:	-23117.			
Covariance Type:	nonrobust	LLR p-value:	0.000			
	coef	std err	z	P> z	[0.025	0.975]
danceability	3.9003	0.096	40.657	0.000	3.712	4.088
energy	0.6304	0.058	10.876	0.000	0.517	0.744
instrumentalness	-7.0114	0.182	-38.485	0.000	-7.368	-6.654
valence	-0.1981	0.062	-3.209	0.001	-0.319	-0.077
intercept	-1.9799	0.057	-34.587	0.000	-2.092	-1.868

In [ ]: