

Appendix 6-Corrolation Heat Map

In [1]:

```
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
```

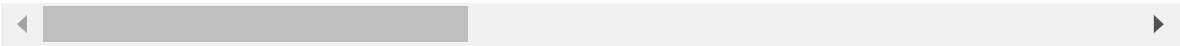
In [2]:

```
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:2JqnpexlO9dmvjUMCaLCLJ	0.550	0.415
2	2	Here's To Never Growing Up	Avril Lavigne	spotify:track:0qwcGscxUHGZTgq0zcaqk1	0.482	0.875
3	3	Crawling Back To You	Daughtry	spotify:track:6BDtTzjbJ5kKKSWcJT8MIX	0.438	0.915
4	4	Faster	Matt Nathanson	spotify:track:6plKFdrBnKF0y3CRuceTDh	0.742	0.855

5 rows x 32 columns



In [3]:

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

Out[3]:

	track	artist	uri	danceability	energy	key	loud
0	Wild Things	Alessia Cara	spotify:track:2ZyuwVvV6Z3XJaXIFbspeE	0.741	0.626	1	-
1	Love Someone	Lukas Graham	spotify:track:2JqnpexlO9dmvjUMCaLCLJ	0.550	0.415	9	-
2	Here's To Never Growing Up	Avril Lavigne	spotify:track:0qwcGscxUHGZTgq0zcaqk1	0.482	0.873	0	-
3	Crawling Back To You	Daughtry	spotify:track:6BDtTzjbJ5kKKSWcJT8MIX	0.438	0.919	0	-
4	Faster	Matt Nathanson	spotify:track:6pIKFdrBnKF0y3CRuceTDh	0.742	0.853	9	-

5 rows × 31 columns

In [4]:

```
df_1=df
y=df_1['target']
x=df_1.iloc[:,[3,4,5,6,7,8,9,10,11,12,13,14,15,16,17]]
x.head()
```

Out[4]:

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	live
0	0.741	0.626	1	-4.826	0	0.0886	0.02000	0.000000	0
1	0.550	0.415	9	-6.557	0	0.0520	0.16100	0.000000	0
2	0.482	0.873	0	-3.145	1	0.0853	0.01110	0.000000	0
3	0.438	0.919	0	-2.910	0	0.0495	0.00674	0.000000	0
4	0.742	0.853	9	-4.147	1	0.0393	0.00743	0.000005	0

In [5]:

```
norm = MinMaxScaler().fit(x)
x_norm = norm.transform(x)
x=pd.DataFrame(data=x_norm,columns=x.columns)

x.head()
```

Out[5]:

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentaln
0	0.734180	0.625906	0.090909	0.838293	0.0	0.071230	0.020080	0.0000
1	0.528627	0.414853	0.818182	0.805631	0.0	0.032086	0.161647	0.0000
2	0.455446	0.872968	0.000000	0.870012	1.0	0.067701	0.011145	0.0000
3	0.408093	0.918980	0.000000	0.874446	0.0	0.029412	0.006767	0.0000
4	0.735256	0.852963	0.818182	0.851105	1.0	0.018503	0.007460	0.0000

In [6]:

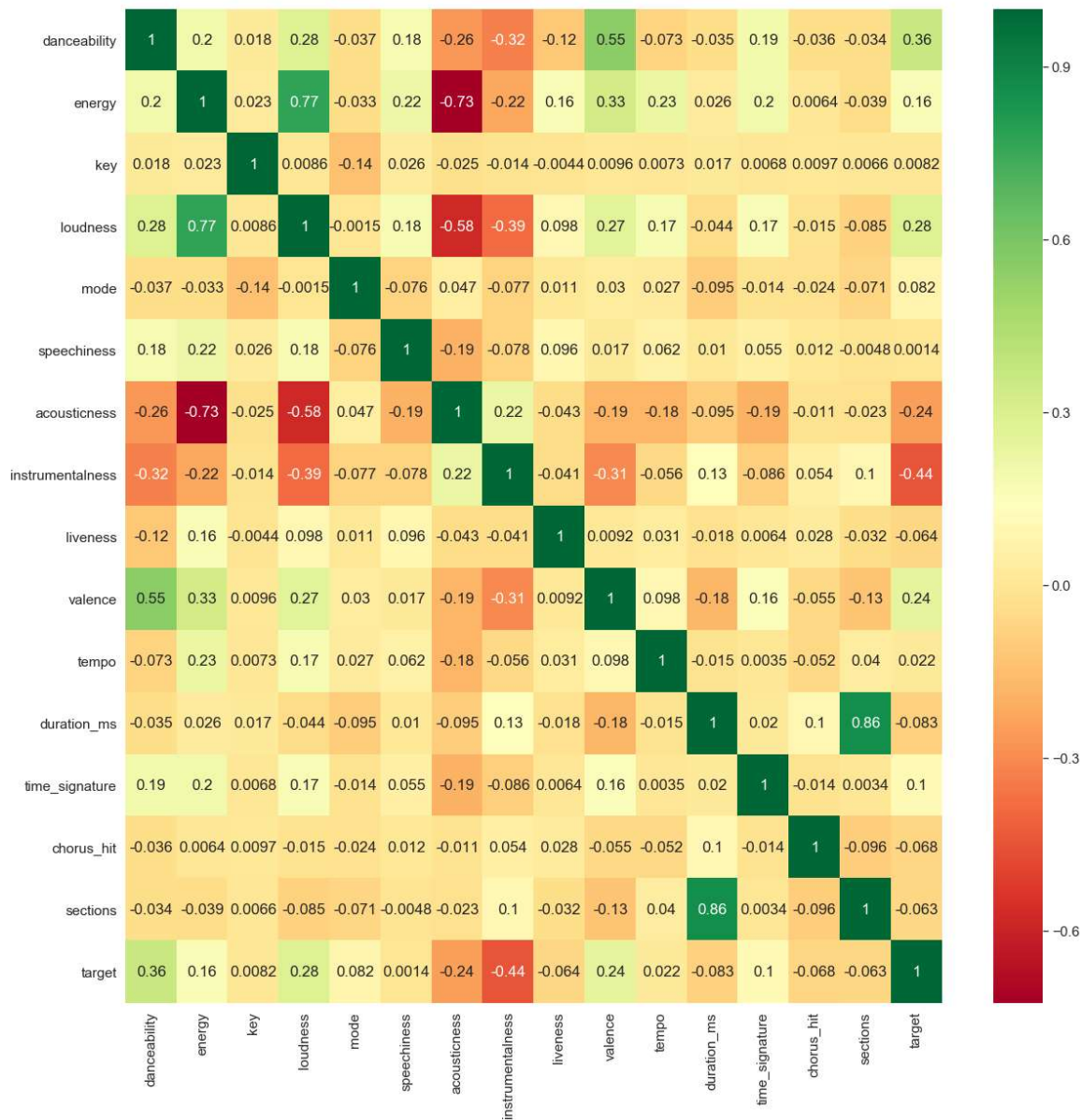
```
data=pd.concat([x,y],axis=1)
data.head()
```

Out[6]:

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentaln
0	0.734180	0.625906	0.090909	0.838293	0.0	0.071230	0.020080	0.0000
1	0.528627	0.414853	0.818182	0.805631	0.0	0.032086	0.161647	0.0000
2	0.455446	0.872968	0.000000	0.870012	1.0	0.067701	0.011145	0.0000
3	0.408093	0.918980	0.000000	0.874446	0.0	0.029412	0.006767	0.0000
4	0.735256	0.852963	0.818182	0.851105	1.0	0.018503	0.007460	0.0000

In [14]:

```
#get correlations of each features in dataset
corrmat = data.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))
sns.set(font_scale=1.4)
#plot heat map
g=sns.heatmap(data[top_corr_features].corr(),annot=True,linewidths=0,cmap="RdYlGn")
```



In [18]:

```
df_1=df
y=df_1['target']
x=df_1.iloc[:,[3,4,5,7,8,10,11,12,13,14,15,16]]
x.head()
```

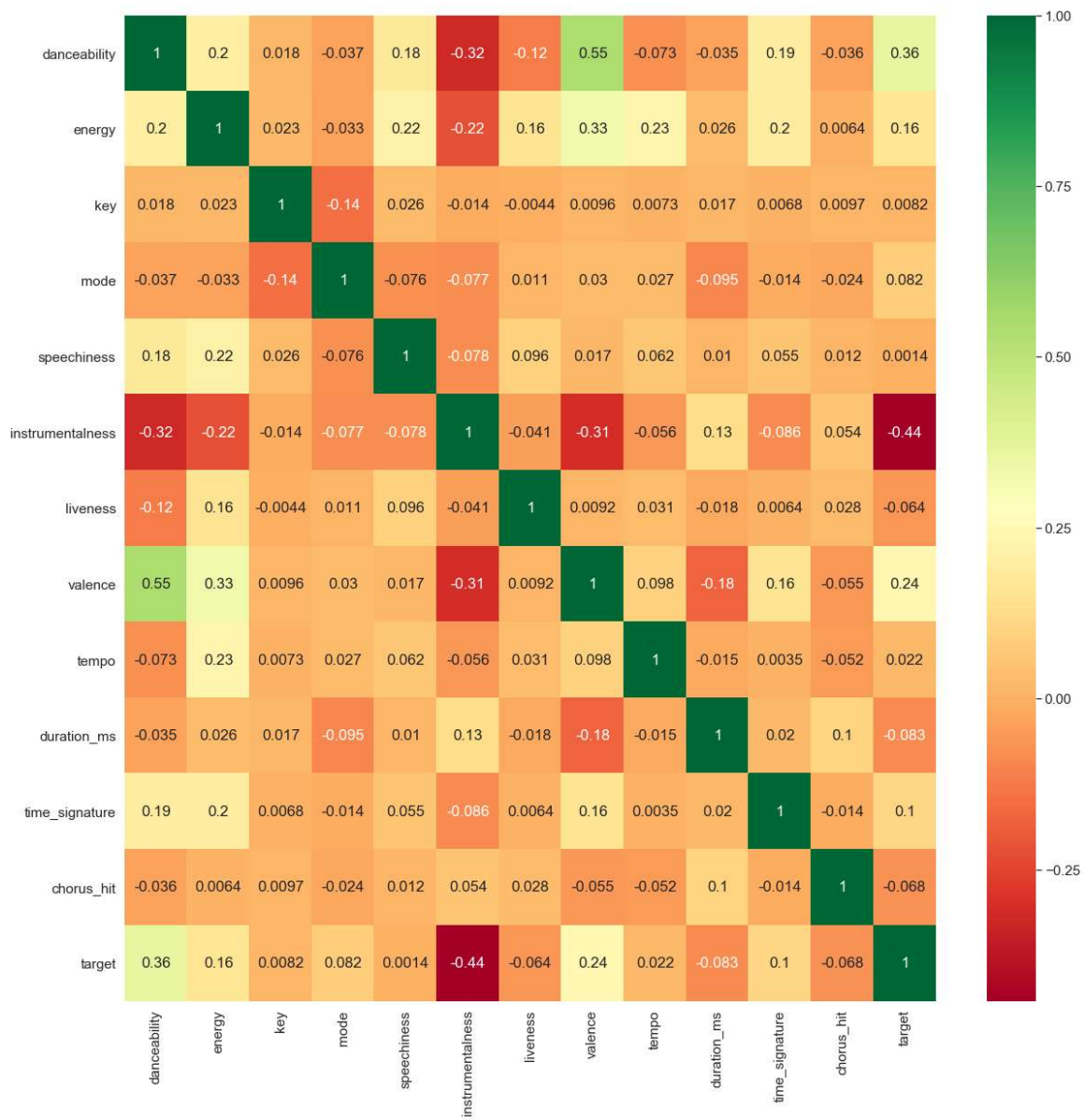
Out[18]:

	danceability	energy	key	mode	speechiness	instrumentalness	liveness	valence	tempo
0	0.741	0.626	1	0	0.0886	0.000000	0.0828	0.706	108.029
1	0.550	0.415	9	0	0.0520	0.000000	0.1080	0.274	172.065
2	0.482	0.873	0	1	0.0853	0.000000	0.4090	0.737	165.084
3	0.438	0.919	0	0	0.0495	0.000000	0.1580	0.195	151.026
4	0.742	0.853	9	1	0.0393	0.000005	0.3320	0.950	107.030



In [19]:

```
norm = MinMaxScaler().fit(x)
x_norm = norm.transform(x)
x=pd.DataFrame(data=x_norm,columns=x.columns)
data=pd.concat([x,y],axis=1)
corrmat = data.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))
sns.set(font_scale=1.4)
#plot heat map
g=sns.heatmap(data[top_corr_features].corr(),annot=True,linewidths=0,cmap="RdYlGn")
```



In []: