ISDS 577 MASTER OF SCIENCE CAPSTONE SEMINAR

Spring 2018 session- PHASE 2-DELIVERABLE

Musical Preferences with Cognitive Characteristics & Other Variables

Professor: Dr Daniel Soper

Team Members

Harshal Saptarshi

Malvika Kulkarni

Shashwat Khambhatta

Taranjot Singh

Vishal Vatnani

TABLE OF CONTENTS

Executive Summary

What causes us to like music? Who likes what kind of music? Why do you and your friends have different taste in music? Research, time and again, has proved that musical choices and personality is linked to each other, yet we know so little of what is this relationship and what influences or mediates this relationship. To address this curiosity and urge for learning, a scientific investigation was conducted in order to learn more about the cognitive styles or personas linked with musical preferences. By collecting real data, a research was orchestrated and the findings were reported. Music is a prominent feature on a day to day life and based on this, a survey with questions pertaining to personality characteristics, mood, education level, age, and intelligence were framed. The data collected underwent rigorous data pre-processing and exploration methods and several machine learning concepts were employed. The study revealed that out of five personality types, openness, extraversion, and agreeableness in a person related to which musical genre that person chose and mood also influenced what kind of music a person likes or rates the musical genres. By successfully answering the research questions, the study could be used for further enhancement.

Introduction

The data for the study were gathered using Qualtrics survey designing platform and consisted of 989 records and 28 variables after cleaning. Among these variables contained measures of music genre preference, ratings on six music genres, ratings on six audio clips spanning each of the six genres that the participants were not aware of, personality items, mood scale items, and demographic variables. The personality types and mood inventory were directly adapted from the Ten Item Personality Inventory Scale [3] and the Brief Mood Introspection Scale [4]. The study found that there are indeed relationships between some of the personality dimensions and music genres along with mood and other demographic variables.

Exploratory Data Analysis

Missing rows combinations

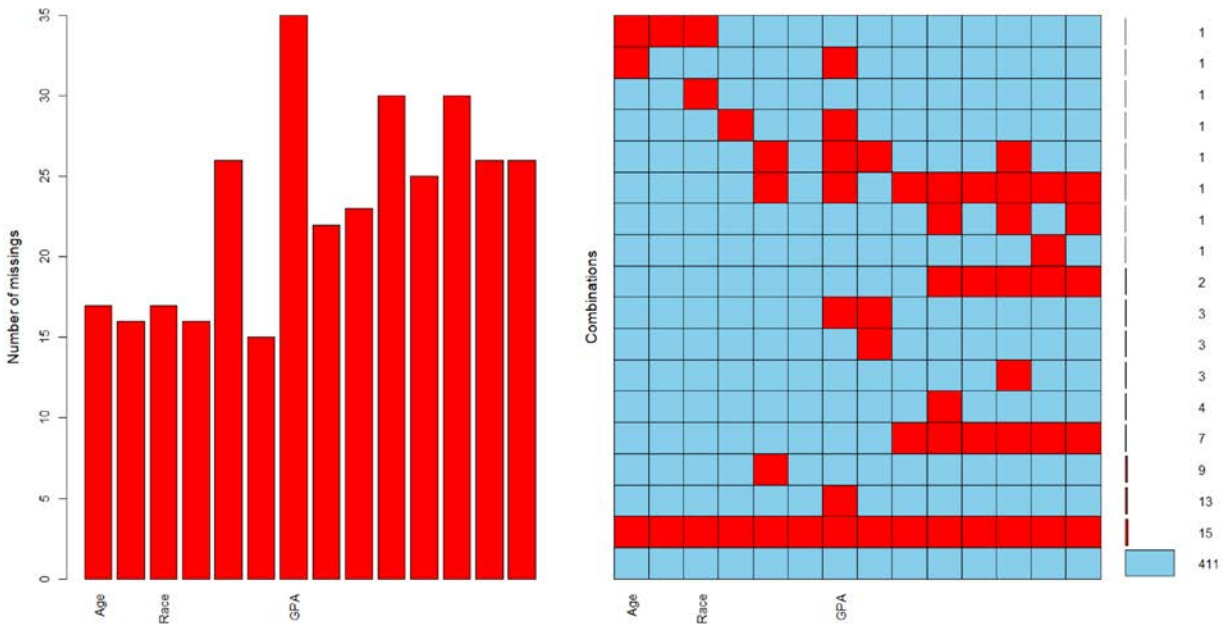

Figure 1: Data Visualization to exhibit missing rows

With the raw data that we had collected there were 67 rows with missing data in at least one column. The total number of rows was 978. If we had gone for listwise deletion where the whole row is deleted if there is a single missing data, new row count would have been 911. The data in the image shows 411 at the end – that is the data excluding the online surveys.

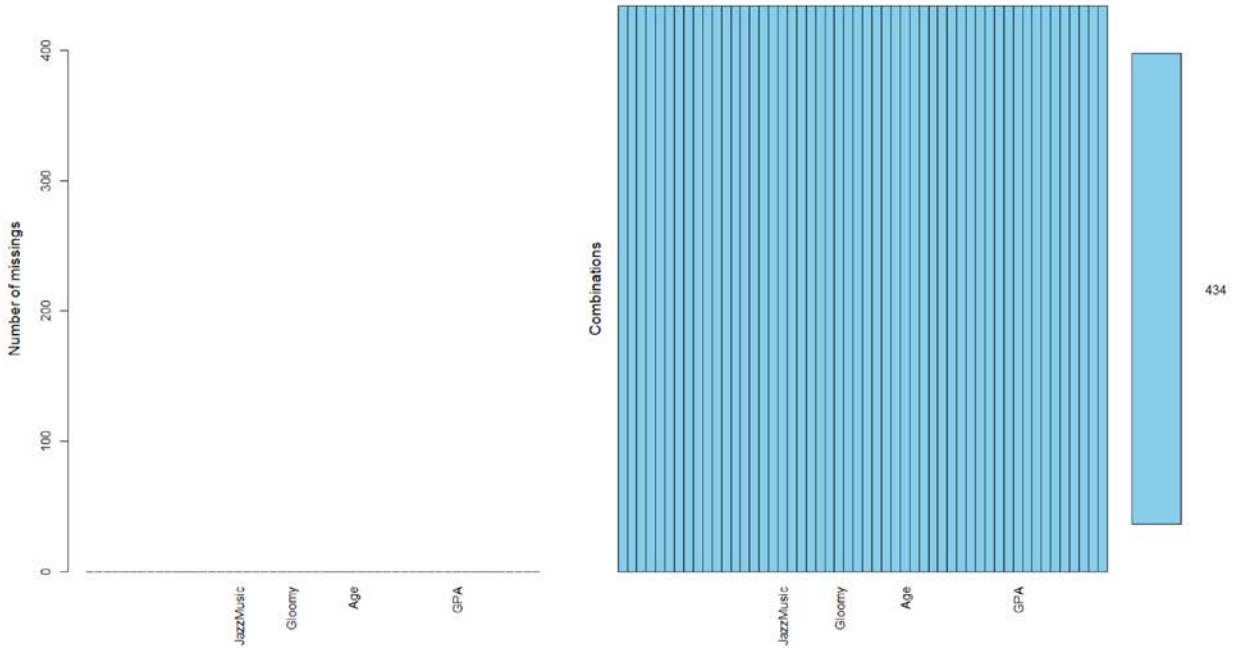

Figure 2: Data visualization to exhibit cleaned data after pairwise deletion

After careful pairwise deletion, and mean imputing rows which had less than 5 columns missing except the Dependent Variable which is Personal Music Choice, and deleting rows with more than 5 columns missing, we get the cleaned data. We are left with 934 rows. So the pairwise deletion minimizes the loss that occurs for listwise deletion.

Box plots and outliers for preprocessed data

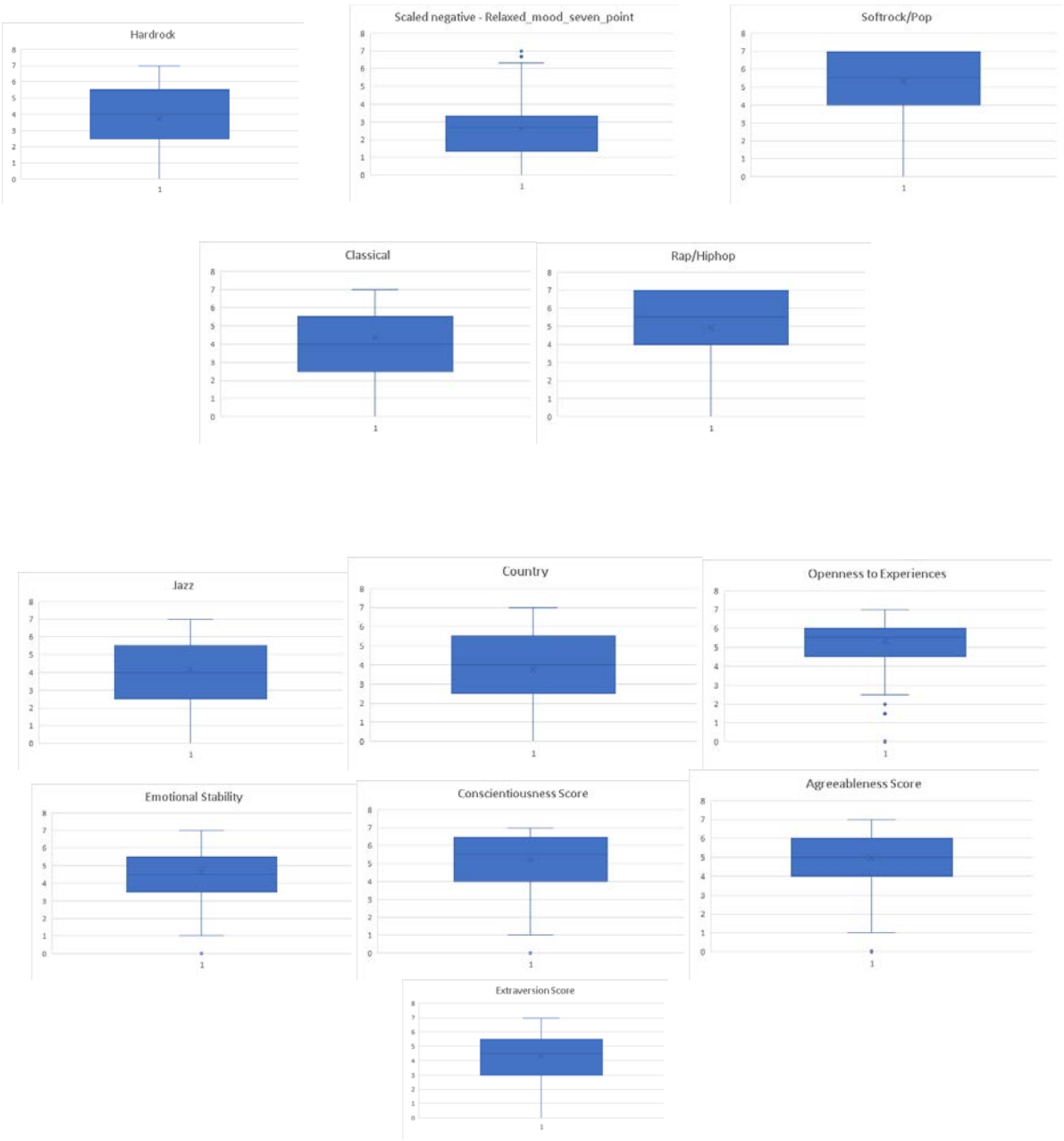

Figure 3: Box plots for missing data

For the unprocessed data, when we found these results after plotting the data using boxplots. We took 12 variables which were continuous. We found multiple counts of outliers for various variables.

Boxplots and outliers for cleaned data

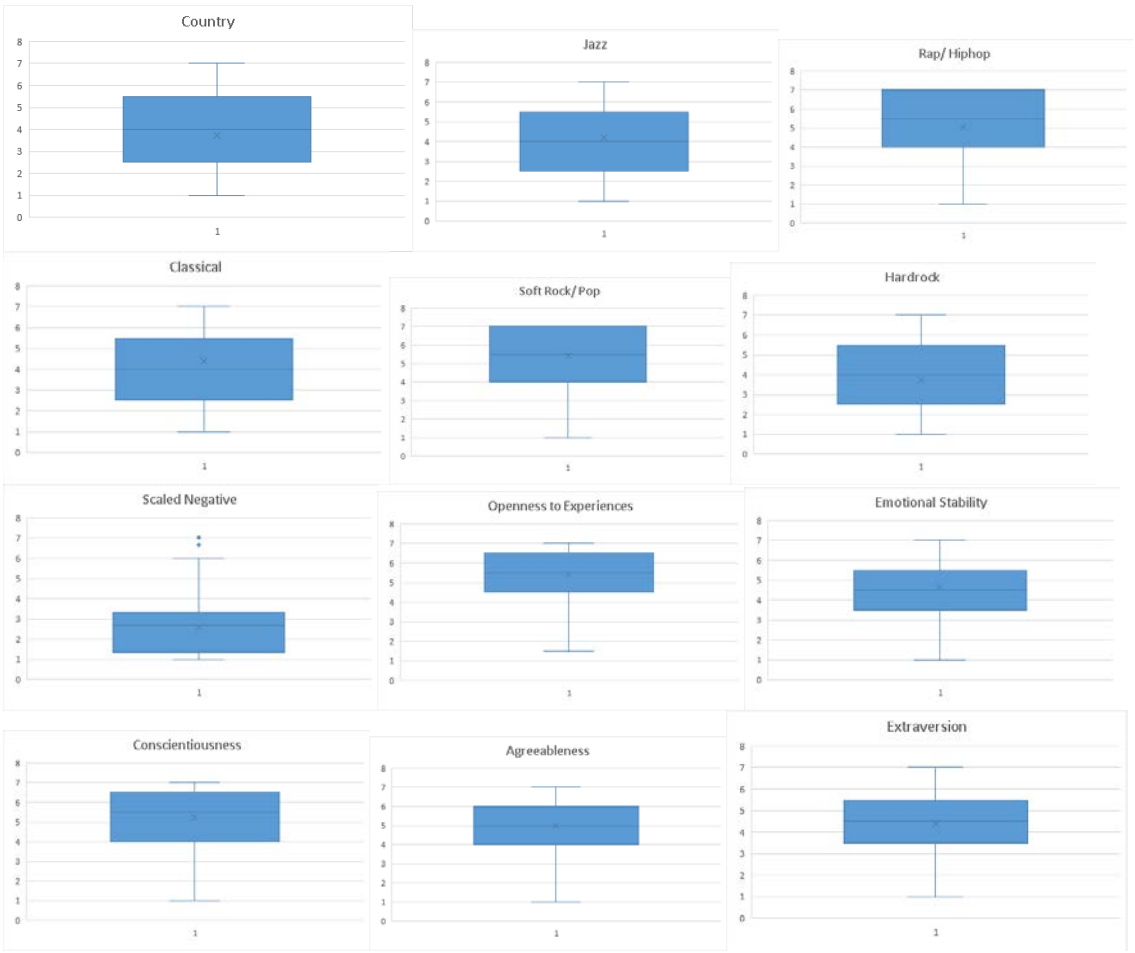

Figure 4: Box plots for cleaned data

For the cleaned data we got the following results for plotting the data using boxplots for these 12 continuous variables.

Data Merging:

The dataset after getting cleaned was

- **Anova / t-test** - It was run for continuous variables like Big-5 Personality Dimensions. The variables in personality types like openness, agreeableness, emotional stability and more were tested using t-test to confirm that they were significant for the model and could be used in the research dataset. The t-test were run independently for each continuous variable so as to establish the significance of that variable. An example of such test is shown below
- **Kruskal wallis test** - Kruskal wallis is used to compare 2 or more independent variables to see if they originate from same distribution for range data(ordinal variable) like GPA, Age. The test is valid for ordinal data or range data which basically is the case with variables like GPA and Age since they have been segregated as range data as it is not continuous. The test helps to find the variables which are significant in finding the research questions. An example is shared as a screenshot with Kruskal wallis test in the end.
- **Fisher exact test and Chi square test** - It is a statistical significance test to classify objects in different ways. This test is supposedly more accurate than even the chi square test. The Variables that satisfy this test to take place are nominal variables. These variables do not have any orders or numerical values. We use it for variables like Race, Musical choice.

- **Logistic / chi squared test** -This test helps determine the significance of the variables like Gender. These are binary variables. The variables which prove significant are then merged in the dataset for the analysis.

Variable	question	Analysis Name	p-value	Result	Can we merge?	Notes	Benferoni P-valu
Race	7	Fisher's exact test	0.1146	Not Significant	Yes		false
Personality Types	2.1	Anova	0.139488659		Yes		false
Personality Types	2.1	Anova	0.395530376				false
Personality Types	2.1	Anova	0.525432				false
Personality Types	2.1	Anova	0.644651				false
Personality Types	2.1	Anova	0.296576366				false
Personality Types	2.1	Anova	0.566239				false
Personality Types	2.1	Anova	0.098612				false
Personality Types	2.1	Anova	0.022378				false
Personality Types	2.1	Anova	0.647797				false
Personality Types	2.1	Anova	0.8839024				false
commuter		Chi Square test	0.23		yes		false
Gender		Chi Square test	0.159		yes		false
Musical_Choice		Chi Square test	0.536895		yes		false
commuter		Chi Square test	0		No		true
Gender		Chi Square test	0.032522		no		false
Musical_Choice		Chi Square test	0.020081		no		false
Race	7	Fisher's exact test	2.20E-16	Significant	No	Vishal will run Fis	true
GPA	11	Kruskal-Wallis's test	0.2926	Not Significant	Yes		false
GPA	11	Kruskal-Wallis's test	0.258	Not Significant	Yes		false
Gender		Chi Square test			Yes		true
Commuter		Chi Square test			Yes		true
Musical_Choice		Chi Square test			Yes		true
Personality Types	2.1	ANOVA	0		No		true
Personality Types	2.2	ANOVA	0.04473608		No		false
Personality Types	2.3	ANOVA	0.29138915		Yes		false
Personality Types	2.4	ANOVA	0.02442842		No		false
Personality Types	2.5	ANOVA	0.02839867		No		false
Personality Types	2.6	ANOVA	3.73E-05		No		true
Personality Types	2.7	ANOVA	0.23063996		Yes		false
Personality Types	2.8	ANOVA	0.00790591		No		false
Personality Types	2.9	ANOVA	0.45960336		Yes		false

reddit,online_friends	Time Spent on music	10	Kruskal-Wallis's test	0.0004421		No
reddit,online_friends	GPA	11	ANOVA	0.31423598		Yes
reddit,online_friends	Commuter	12	ANOVA	9.28E-12		No
reddit,online_friends	Music_rating_clip	13.1	ANOVA	0.11511858		Yes
reddit,online_friends	Music_rating_clip	13.2	ANOVA	0.50448861		Yes
reddit,online_friends	Music_rating_clip	13.3	ANOVA	0.64442905		Yes
reddit,online_friends	Music_rating_clip	13.4	ANOVA	0.00119475		No
reddit,online_friends	Music_rating_clip	13.5	ANOVA	0.00109653		No
reddit,online_friends	Music_rating_clip	13.6	ANOVA	0.00062652		No
online_friends, ISDS 577 Second	Race	7	Fisher's exact test	2.20E-16	Significant	No
online_friends, ISDS 577 Second	GPA	11	Kruskal-Wallis's test	0.3955	Not Significant	Yes
mturk, reddit and online friends	Personality types	2.1	ANOVA	0.00000000132	Significant	No

Figure 5: Screenshots from metadata about analyses run during comparing means of all variables across samples

Example of a t-test for continuous variable

The clean data was further needed to be merged and only certain variables which were significant in determining the relation with other continuous variables were selected using the t-test. We did the t-test for each individual continuous variable and then we generated the p-values which we used to determine the significance of the variable. The continuous variables whose p value came out to be more than .05 were declared significant for the research and were further merged in the data with the other variables. The example below shows a similar situation in which one of the Big-5 personality dimensions is being tested for significance.

577	577_survey
5	6
7	5
5	7
7	7
5	7
4	7
5	3
5	4
7	6
5	5
3	5
6	6
7	7
7	5
6	7
6	6
6	7
5	7

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	6.032967033	5.830508475
Variance	0.987789988	1.55698422
Observatio	91	59
Hypothesiz	0	
df	104	
t Stat	1.04907288	
P(T<=t) one	0.148288183	
t Critical on	1.659637437	
P(T<=t) two	0.296576366	
t Critical tw	1.983037526	

Figure 6: T-test performance to merge the data

- Visualizationp; before Data cleaning

The dataset before cleansing had many a number of rows and columns empty which created a lot of inconsistencies during the Visualization process. The Histogram below is an example of such a situation and in spite of numerous attempts to rectify it, it wasn't successful and the histogram continued to have inconsistencies, until the dataset was cleaned

Audio Ratings of surveyors

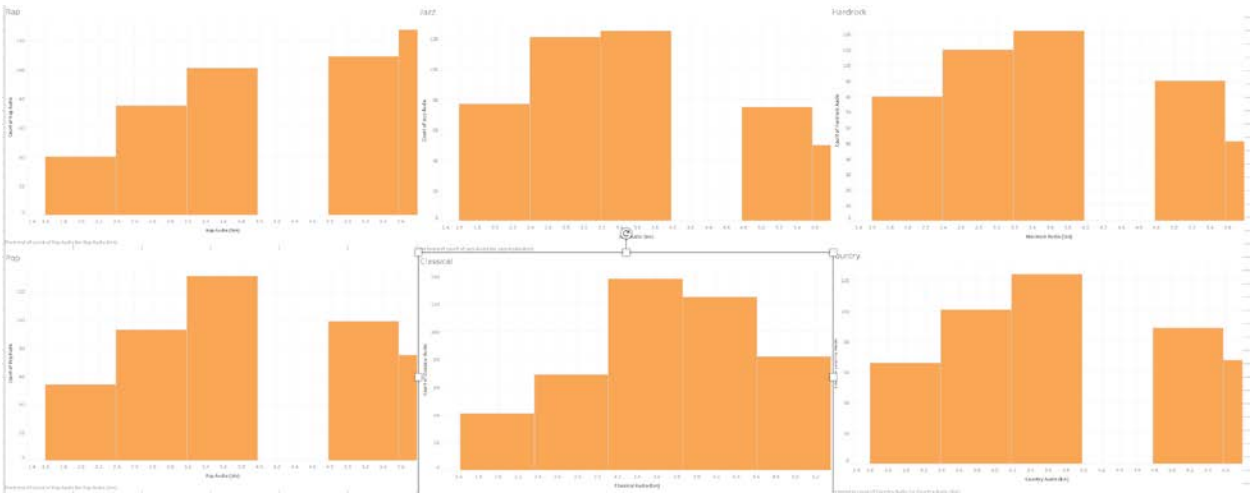

Figure 7: Histogram for Missing data

- **Visualization after Data cleaning**

Once the data was cleaned for inconsistencies and all the missing rows and columns were rectified using the cleaning processes, the final dataset was free from all inconsistencies. When used Tableau for visualization we got a clean and consistent histogram which helped in the visualizing process. The histogram below if of the audio rating provided by the people when made to listen the audio clips in the survey. The histograms represent music genres of classical, soft/pop, hardrock, country, hip hop/rap and jazz. When looked at the histogram in the third column and second row, we see that it looks like the highest rated music choice which is the Rap genre. Even the first histogram seems to be highly liked which is the soft rock. Hence it clearly shows that data cleaning helps in better visualization.

Audio Ratings of surveyors

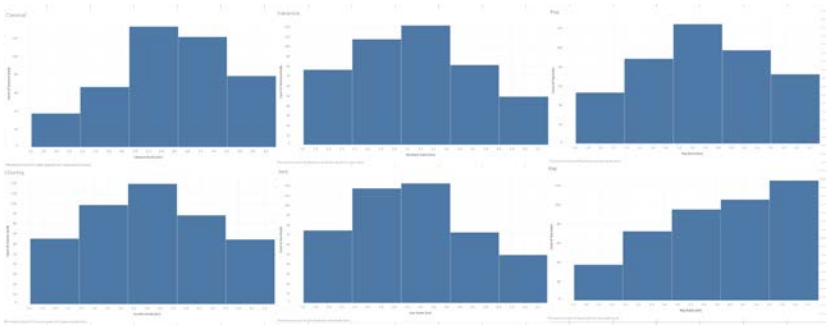

Figure 8: Histogram for audio ratings

Preferred Music genres

The below histograms have been generated using Tableau for visualisation. The continuous variables for selecting the favourite music genre are used to draw a relation between each other. The histograms show that the Rap/Hip hop and Soft/Rock are very highly rated by the surveyors and given high ratings. We see there is a steady decline in the music preference for all other music genres, which means that lesser number of people have higher preference for music genres like jazz, country, hardrock and classic. And the really high ratings for the last columns in the histogram for Hip Hop and soft rock proves that more people gave it higher rating.

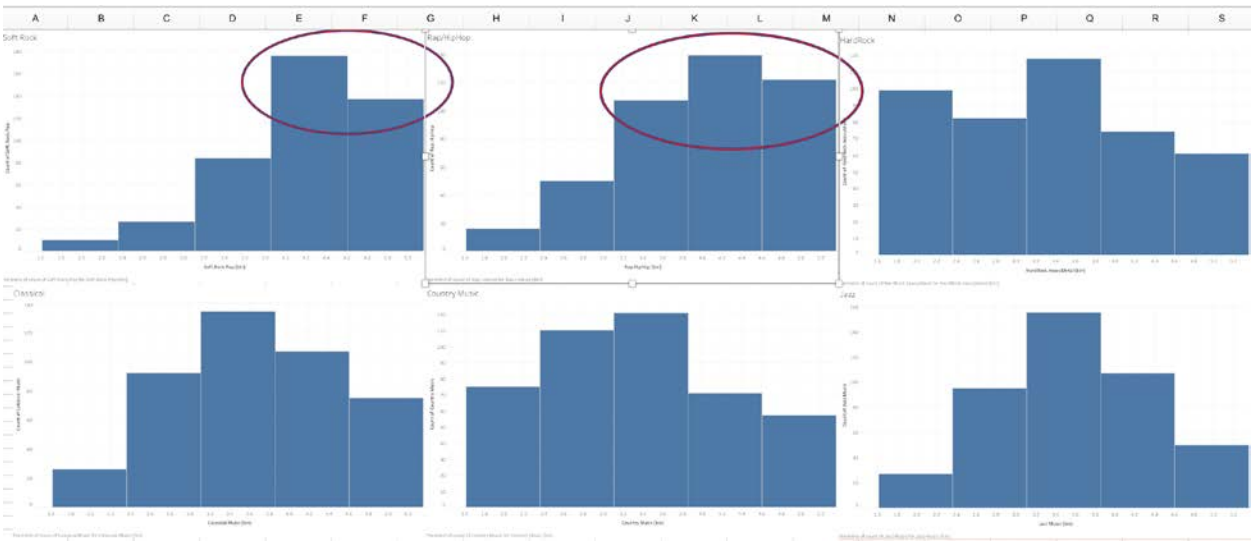

Figure 9: Histogram for Musical genres

Relation between most preferred Music Genre and highest rated Music genre

The histograms below are from preferred music genre and rating of different audio clips. The blue histograms are music preferences of people when asked to rate what intensity they love the genre with. The orange on the other hand is the ratings the surveyors gave to the audio clips they heard. It was after they were made to rate the music genres they liked most. Hence, we tried to figure out if they tend to change their preferences after listing to different songs from different genres. For the analysis we are posting the screenshot for only Rap and Soft rock since they were the most popular and showed good relation from their audio rating histograms as well. What can be visualized using these histograms is that rap is somewhat the more popular or prefered genre and even the ratings of audio clip of rap/hip hop aligns well with it.

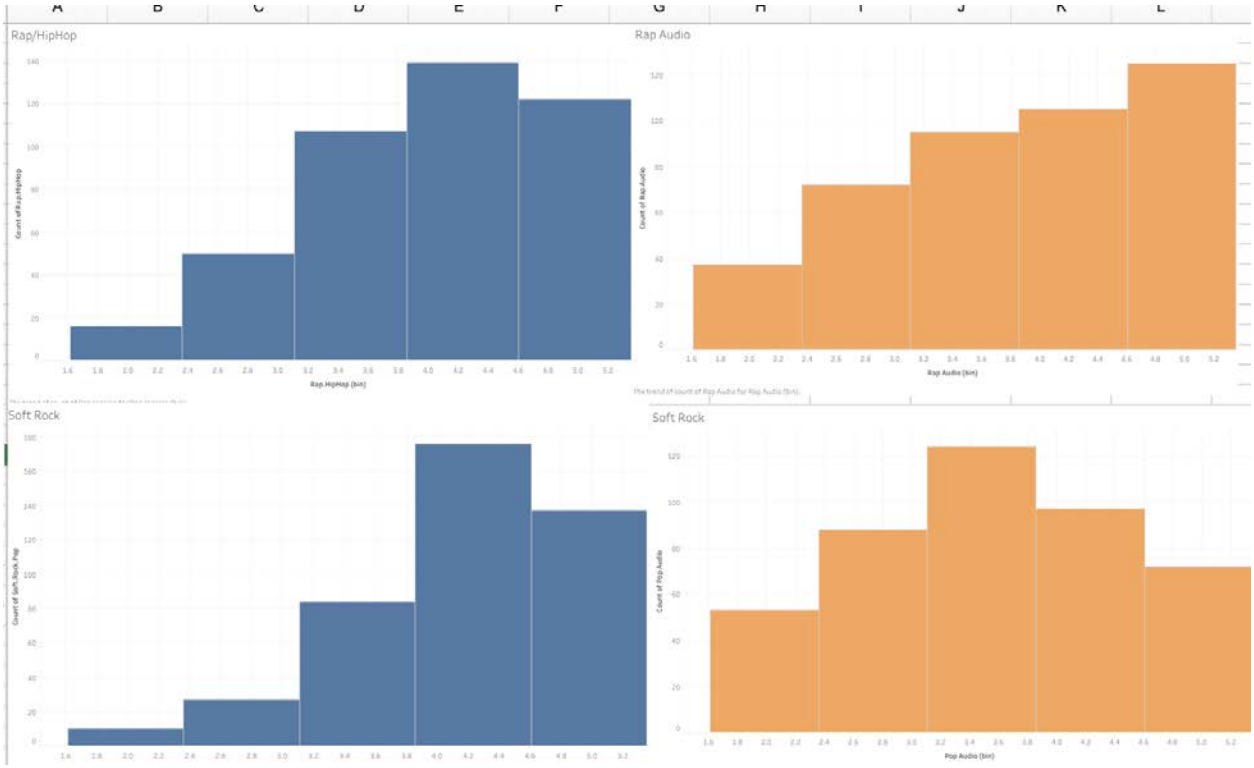

Figure 10: Favorite Musical genre - Softrock/Pop

Exploratory Factor Analysis

We wanted to test for underlying relationships among the variables of interest and also wanted to test the validity of the personality and mood inventories.

Factor Analysis on items adapted from the Brief Mood Introspection Scale revealed the loadings in the figure below, with a cut off of 0.3

```
Loadings:
        Factor1 Factor2 Factor3
Fedup    0.714
Gloomy                   0.989
Jittery
Nervous          1.016
Sad      0.620
Calm    -0.655
```

Figure 11: Factor Loadings for Mood Scale Items

Factor Analysis on the items in the TIPI scale revealed the following loadings, indicating a poor factor structure. This is because of the limited number of items on the Inventory.

```
Loadings:
                                Factor1 Factor2 Factor3 Factor4 Factor5
Extraverted_Enthusiastic                         0.797
Critical_Quarrelsome                             0.322
Dependable_SelfDisciplined                                       -0.316
Anxious_Upset                    1.054
Open.to.new.experiences_complex                          0.403
Reserved_Quiet                                  -0.475
Sympathetic_Warm                                 0.657
Disorganized_Careless                                            0.872
Calm_EmotionallyStable                           0.390
Conventional_Uncreative                  1.019
```

Figure 12: Factor Loadings for Big 5 items

Factor Analysis on Genre and Audio ratings provided the loadings below

```
Loadings:
                  Factor1 Factor2 Factor3
CountryMusic
JazzMusic          0.929
Rap.HipHop                         1.001
CalssicalMusic     0.377
Soft.Rock.Pop              1.014
HardRock.heavyMetal
```

Figure 13: Factor Loadings for Music Genre ratings

```
Loadings:
                Factor1 Factor2 Factor3
Classical_Audio  1.010
Jazz_Audio       0.308
Hardrock_Audio
Country_Audio            1.007
Pop_Audio
Rap_Audio                       0.642
```

Figure 14: Factor Loadings for Music Audio clip ratings

We found that Classical and Jazz audio and genres classical loaded on a common factor.

Assumption Testing for Regression Models

- Linear relationship - independent variables are linearly related to independent variables
- Multivariate normality - all variables belong to the same distribution and QQ-plot is used.
- No or little multicollinearity - VIF(Variation Inflation Factor) used to check this assumption. The independent variables should not be correlated to each othere.
- No autocorrelation - the residuals of independent variables are not correlated.
- Homoscedasticity - residuals are equally distributed across the regression line

Below are the VIF values, correlation matrix and the homoscedasticity plot from the SPSS

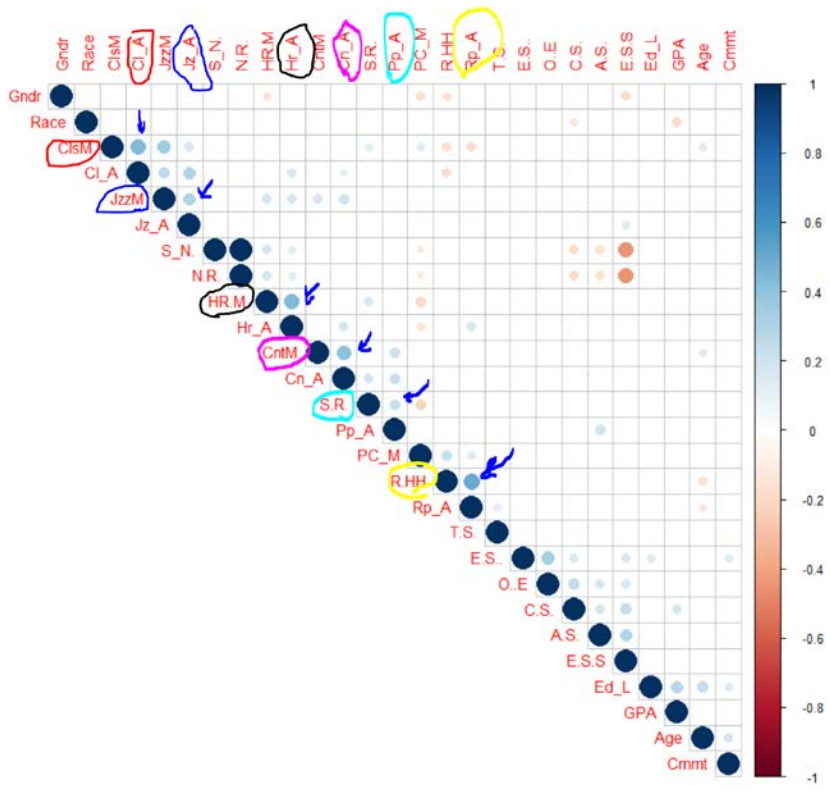

Figure 15: Correlation matrix for all variables

The correlation matrix shows correlation between the musical genre and audio clip variables. And also, interestingly, classical and jazz are correlated, which can be explained by factor analysis. The negative scale mood is correlated to the emotional stability personality. This proves that the factor analysis which was performed has appropriate results.

Model		Collinearity Statistics	
		Tolerance	VIF
1	Agreeableness Score (out of 7)	.773	1.294
	Conscientiousness Score (out of 7)	.775	1.290
	Emotional Stability Score (out of 7)	.619	1.616
	Openness to Experiences Score (out of 7)	.752	1.329
	CountryMusic	.708	1.413
	JazzMusic	.672	1.489
	Rap/HipHop	.619	1.616
	CalssicalMusic	.676	1.479
	Soft Rock/Pop	.848	1.179
	HardRock/heavyMetal	.701	1.427
	Fedup	.576	1.735
	Gloomy	.535	1.871
	Jittery	.602	1.660
	Nervous	.612	1.634
	Sad	.511	1.955
	Calm	.698	1.434
	Age	.808	1.237
	Gender	.824	1.214
	Race	.865	1.156
	Education_Level	.748	1.336
	Time Spent on Music	.913	1.096

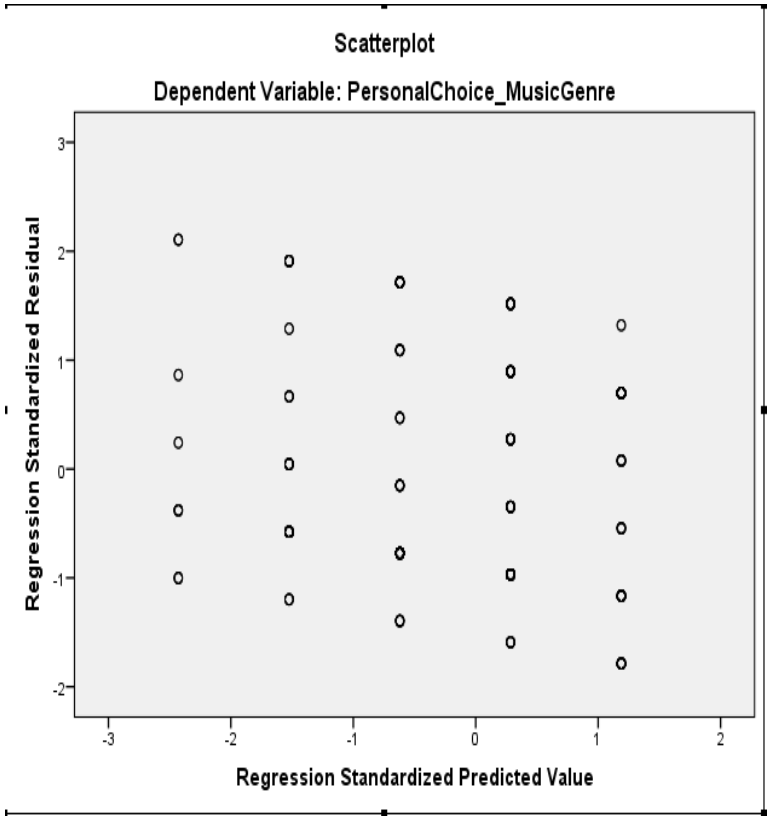

Figure 16: Homoscedasticity and VIF values

Objective

The primary objective of the report is to focus on the causal relationship between an individual's musical preferences, personality, mood and/or demographics. To be more precise, the report answers the following research questions:

1) **Are an individual's Big-5 personality characteristics predictors of their music genre preference?**

Analysis methods: Multinomial Logistic Regression, Simple Linear Regression, Multiple Linear Regression

Procedure: The dependent variable is the most preferred genre preferences of all the six genres of music as well as the ratings on each of the six genres and audio ratings, and the independent variables are the Big-5 personality scores explaining if there is any relation between the dependent and the independent variables.

Results:

```
Coefficients:
          (Intercept) Extraversion.Score..out.of.7. Openness.to.Experiences.Score..out.of.7.
Classical  -1.7312772                  0.017931001                                -0.17280303
Country    -0.3726293                  0.022095832                                -0.37290380
Hard rock  -2.2288703                 -0.138158352                                 0.26436890
Jazz       -1.1963100                  0.009115382                                -0.04863922
Rap        -0.1375152                  0.107981789                                -0.07002090

          Agreeableness.Score..out.of.7. Emotional.Stability.Score..out.of.7.
Classical                     -0.2566871                            0.1003680
Country                       -0.2261086                            0.2494814
Hard rock                     -0.1607237                            0.1080100
Jazz                          -0.3207600                            0.2699459
Rap                           -0.3318626                            0.2990021
          Conscientiousness.Score..out.of.7.
Classical                       0.057063375
Country                         0.129676583
Hard rock                      -0.004419483
Jazz                            0.005059337
Rap                             0.005012801
```

Figure 17. Multinomial Logistic Regression (Music Choice ~ Big 5 scores) with significant coefficients

```
> print(p, cutoff = 0.05)
          (Intercept) Extraversion.Score..out.of.7. Openness.to.Experiences.Score..out.of.7.
Classical   0.1270737                    0.8775806                               0.30190815
Country     0.7548012                    0.8662386                               0.03097859
Hard rock   0.1052122                    0.3167130                               0.19325297
Jazz        0.3684537                    0.9492056                               0.80405698
Rap         0.8637322                    0.2096376                               0.55203944
          Agreeableness.Score..out.of.7. Emotional.Stability.Score..out.of.7.
Classical                    0.091087185                           0.431971621
Country                      0.189751022                           0.100657171
Hard rock                    0.379548418                           0.480196807
Jazz                         0.083750317                           0.097299762
Rap                          0.003291264                           0.002073778
```

Figure 18. Significance values for Multinomial Logistic Regression (Music Choice ~ Big 5 scores)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.253	.508		8.373	.000
	Extraversion.Score..out.of.7	-.082	.054	-.078	-1.527	.127
	Agreeableness.Score..out.of.7	-.144	.070	-.104	-2.061	.040
	Conscientiousness.Score..out.of.7	-.133	.061	-.111	-2.200	.028
	Emotional.Stability.Score..out.of.7	.168	.060	.145	2.822	.005
	Openness.to.Experiences.Score..out.of.7	.172	.074	.121	2.333	.020

a. Dependent Variable: jazz_scale_out_of_7

Emotional.Stability.Score..out.of.7	.148	.062	.123	2.392	.017
Openness.to.Experiences.Score..out.of.7	.016	.077	.011	.213	.832

a. Dependent Variable: Rap_scale_out_of_7

Agreeableness.Score..out.of.7	.126	.064	.100	1.974	.049
Conscientiousness.Score..out.of.7	-.055	.055	-.051	-1.001	.317
Emotional.Stability.Score..out.of.7	-.100	.055	-.095	-1.841	.066
Openness.to.Experiences.Score..out.of.7	.173	.067	.134	2.568	.011

a. Dependent Variable: Soft_Rock_out_of_7

Figure 19: Multinomial logistic Regression(Personality types ~ Musical Genre ratings)

The figure above shows the SPSS output for the multinomial regression between all the personality types and the ratings for each musical genre, this shows the significance of each personality type with the musical genres, and there are both positive and negative

relationships. Personality types and openness like Softrock/Pop, but people who agree more do not like jazz as like people who are conscientious. Similarly, there are regression output for Musical audio ratings too

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.541	.582		6.086	.000
	Extraversion.Score..out.of.7	-.010	.062	-.009	-.167	.867
	Agreeableness.Score..out.of.7	-.016	.080	-.010	-.202	.840
	Conscientiousness.Score..out.of.7	-.227	.069	-.165	-3.273	.001
	Emotional.Stability.Score..out.of.7	.225	.068	.169	3.291	.001
	Openness.to.Experiences.Score..out.of.7	.070	.085	.043	.825	.410

a. Dependent Variable: Jazz_Audio_Out_of_7

1	(Constant)	4.001	.562		7.120	.000
	Extraversion.Score..out.of.7	-.155	.060	-.134	-2.600	.010

a. Dependent Variable: Classical_out_of_7

Figure 20: Multinomial Regression(Personality types ~ Musical Audio ratings)

2) If the relationship in (1) holds true, then is the relationship mediated by their mood (negative-calm) or demographics?

Analysis methods: Linear Regression and Multinomial Regression

Procedure: We followed the steps for testing mediation effect using regression analysis as described in [5]. To see whether mood affected the genre choice, we ran a model for the genre choice with the mood scale which gave us the result

below. We find that rap has a negative effect on the negative mood which means it is related with people reporting a better mood

```
Coefficients:
           (Intercept) Scaled_Negative.Relaxed_Mood_seven_point
Classical  -0.6747847                              -0.218509153
Country    -2.0509878                               0.186465484
Hard rock  -1.6896477                              -0.003334952
Jazz       -1.7623700                               0.010980568
Rap         0.3211505                              -0.228201897
```

Figure 21. Multinomial model for Music genre choice ~ Mood

We then tested for a relationship between the genre choice with Openness, Emotional Stability, and Agreeableness. We obtained the results below:

```
Coefficients:
          (Intercept) Openness.to.Experiences.Score..out.of.7. Agreeableness.Score..out.of.7.
Classical -1.54659564                            0.19551673                      -0.2543398
Country    0.01922471                           -0.33288308                      -0.2151655
Hard rock -2.47235357                            0.20136410                      -0.1450952
Jazz      -1.16291936                           -0.04293192                      -0.3231542
Rap        0.08723938                           -0.02170001                      -0.3431090
          Emotional.Stability.Score..out.of.7.
Classical                            0.1127151
Country                              0.2750013
Hard rock                            0.0868882
Jazz                                 0.2728999
Rap                                  0.3158340
```

Figure 22: Regression coefficients of Big 5 dimensions with music genre preference

```
[[1]]
          (Intercept) Openness.to.Experiences.Score..out.of.7. Agreeableness.Score..out.of.7.
Classical  0.14554417                              0.21165053                      0.091735369
Country    0.98598802                              0.03759428                      0.207931404
Hard rock  0.05701621                              0.29036285                      0.427870753
Jazz       0.34874520                              0.81432834                      0.079486142
Rap        0.90685214                              0.84275837                      0.002084356
          Emotional.Stability.Score..out.of.7.
Classical                         0.3664952847
Country                           0.0647217305
Hard rock                         0.5618008084
Jazz                              0.0876197571
Rap                               0.0008870731
```

Figure 23: Regression p values of Big 5 dimensions with music genre preference

We proceeded to test the relationship between Mood and Emotional Stability and found a significant relationship

```
Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           4.4889     0.1919   23.39   <2e-16 ***
Emotional.Stability.Score..out.of.7.  -0.4042     0.0393  -10.28   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.136 on 432 degrees of freedom
Multiple R-squared:  0.1967,    Adjusted R-squared:  0.1948
F-statistic: 105.8 on 1 and 432 DF,  p-value: < 2.2e-16
```

Figure 24: Regression summary for Mood Scale and Emotional Stability

Then, we tested for a relationship between music choice and emotional stability with the results below

```
Coefficients:
           (Intercept) Emotional.Stability.Score..out.of.7.
Classical    -1.553497                           0.07129077
Country      -2.254626                           0.15910065
Hard rock    -2.043605                           0.07564257
Jazz         -2.563269                           0.17831429
Rap          -1.279978                           0.21811426
```

Figure 24: Regression coefficients for Music Genre Choice and Emotional Stability

```
[[1]]
           (Intercept) Emotional.Stability.Score..out.of.7.
Classical 0.0063179456                          0.54761893
Country   0.0006643587                          0.23601637
Hard rock 0.0031002179                          0.59816202
Jazz      0.0004658448                          0.22623254
Rap       0.0028024147                          0.01221671
```

Figure 25: Regression p value for Music Genre Choice and Emotional Stability

At last, we tested for a relationship between music choice with Mood and Emotional Stability together combined. In the results below, we found out that country music was significantly related with Emotional Stability and Mood. In all other cases, Country music was not significant with either of the two variables individually. Thus we conclude the presence of a full mediation effect

```
Coefficients:
           (Intercept) Emotional.Stability.Score..out.of.7. Scaled_Negative.Relaxed_Mood_seven_point
Classical  -0.5742174                          -0.01659637                                -0.22833071
Country    -3.8843906                           0.30535469                                 0.33159520
Hard rock  -2.2362944                           0.09298969                                 0.04198518
Jazz       -3.1350810                           0.22962857                                 0.12256256
Rap        -0.6284182                           0.15971937                                -0.14931489
```

Figure 26: Coefficients for mediation effect of Mood on Country music preference and Emotional stability

```
> p_med_t
[[1]]
          (Intercept) Emotional.Stability.Score..out.of.7. Scaled_Negative.Relaxed_Mood_seven_point
Classical 0.501458571                          0.89985636                                0.1368307
Country   0.000186644                          0.04538003                                0.0318782
Hard rock 0.034732731                          0.56290055                                0.8086395
Jazz      0.004828820                          0.16515970                                0.4841263
Rap       0.324578775                          0.09910067                                0.1721930
```

Figure 27: p values for mediation effect of Mood on Country music preference and Emotional stability

3) Is there a relationship between an individual's mood and their music preference ?

AND

8) Is there a relationship between an individual's mood and their music genre ratings (when they listen to a clip and rate the song)?

Conducting analysis between individual's mood and their music preferences which consisted of the following options

Analysis methods: Multinomial Logistic Regression

i) Scaled mood ~ Musical Genre ratings

ii) Scaled mood ~ Musical Audio Ratings

Procedure: The negative scaled mood with the musical genre ratings by every individual was taken as an independent and a dependent variables in this regression model. The p-value below significance level of 0.05 was considered to have positive or negative relation. The below output from SPSS shows the significant variables

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.282	.207		15.836	.000
	Scaled_Negative. Relaxed_Mood_seven_point	.177	.072	.118	2.465	.014

a. Dependent Variable: country_scale_out_of_7

1	(Constant)	5.057	.160		31.515	.000
	Scaled_Negative. Relaxed_Mood_seven_point	.129	.056	.111	2.327	.020

a. Dependent Variable: Soft_Rock_out_of_7

1	(Constant)	2.959	.217		13.623	.000
	Scaled_Negative. Relaxed_Mood_seven_point	.289	.075	.182	3.847	.000

a. Dependent Variable: Hardrock_out_of_7_A

Figure 28: Multinomial Logistic Regression(Mood scale ~ Musical Genre ratings)

By referring to the figure above, you can see that people who are very sad, gloomy, jittery or highly in a negative mood then they prefer to listen to hardrock music and softrock/pop, and country. The musical audio ratings are also tested for any influence by the mood variable.

1	(Constant)	4.075	.202		20.147	.000
	Scaled_Negative. Relaxed_Mood_seven_point	-.159	.070	-.109	-2.270	.024

a. Dependent Variable: Jazz_Audio_Out_of_7

1	(Constant)	3.203	.203		15.772	.000
	Scaled_Negative. Relaxed_Mood_seven_point	.204	.070	.139	2.902	.004

a. Dependent Variable: Hardrock_out_of_7

Figure 29: Multinomial Logistic Regression(Mood scale ~ Musical Audio ratings)

From the figure above, it can be noticed that the negative scale mood is negatively related to jazz, so that means people need to be in good mood or those in good mood will prefer or rate jazz genre better and the hardrock audio rating still shows that people who have negative mood would like to listen to it.

4) Is there a relationship between an individual's demographics, intelligence(GPA) and their music genre preference?

Individual's demographics, intelligence(GPA) and their musical choice has some significance relationship. This is exhibited in the figure below

	Softrock	country	jazz	Rap	classical	Hardrock
Age		0.257 & 0.313		-0.304 & -0.334	0.190 & 0.272	
Education Level		0.302				
GPA						
Race				0.097		
Time Spent listening to music		-0.103		0.136		

Figure 30: Linear Regression (Musical audio ratings ~Demographics and Musical ratings ~Demographics)

From this regression model, education has no influence on musical choices but GPA does come close as 0.08 significance but yet not significant at 0.05 level. The race has relation with just one musical genre and as time listening to music by each person increases, they tend not to go for rap.

5) Is there a relationship between the number of hours an individual spends listening to music per week and their personality characteristics?

From our analysis, we found out that openness to new experience was positively and significantly related with the number of hours people spend listening to music per week

```
Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                           3.53608    0.59753   5.918 6.68e-09 ***
Openness.to.Experiences.Score..out.of.7.  0.18548    0.08405   2.207   0.0279 *
Agreeableness.Score..out.of.7.       -0.09543    0.08295  -1.150   0.2506
Emotional.Stability.Score..out.of.7.  0.13340    0.07068   1.887   0.0598 .
Conscientiousness.Score..out.of.7.   -0.13097    0.07228  -1.812   0.0707 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 31: Linear Regression with Time spent listening to music per week with personality characteristics

6) If the relationship in (5) holds true, is it mediated by the fact that they are a commuter student?

To perform mediation, we first analyzed if the time spent on music was related with commuter student factor. We found a positive relationship as shown below

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.7778     0.1299  29.084   <2e-16 ***
factor(Commuter_text)yes  0.4562     0.1833   2.489   0.0132 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 32: Linear regression with Time spent listening to music per week and commuter student factor

Then we ran an analysis to see if commuter factor could be explained by any of the personality types. The relationship was found to be insignificant, as shown below. And thus we concluded that no mediation effect exists

```
Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                                1.42833    0.63860   2.237   0.0253 *
Openness.to.Experiences.Score..out.of.7. -0.04727    0.08910  -0.531   0.5957
Agreeableness.Score..out.of.7.           -0.12860    0.08825  -1.457   0.1451
Emotional.Stability.Score..out.of.7.      0.05224    0.07496   0.697   0.4859
Conscientiousness.Score..out.of.7.       -0.13647    0.07703  -1.772   0.0764 .
---
```

Figure 33: Commuter factor ~ Personality traits

7) What is the most popular music genre across all individuals? Is this explained by the overall intelligence, age, or personality dimension(s)? [run linear regression on Soft rock/Pop music genre and audio WITH GPA, AGE, AND/OR, Big 5]

As per our findings most popular music genre was **Soft Rock/Pop** across all individuals.

If we consider only one variable across music genre and audio with GPA, AGE, and/or Big 5, the output was not-significant enough, but combining all the variables we were able to predict.

After executing the linear regression analysis, it was found there was significant relationship between one of the Big 5 personalities and Soft rock/Pop.

Openness to new experiences and Soft rock was related with significance level of 0.022 which meant that it is positively related to the Soft Rock/Pop variable.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.110[a]	.012	.010	.976	.012	5.295	1	432	.022	1.884

a. Predictors: (Constant), Openness.to.Experiences.Score..out.of.7

b. Dependent Variable: Soft.Rock.Pop

Figure 34: Linear Regression on Soft rock/Pop music genre with Big-5 Personality

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	3.416	.228		15.007	.000	2.969	3.863					
	Openness.to.Experiences.Score..out.of.7	.095	.041	.110	2.301	.022	.014	.176	.110	.110	.110	1.000	1.000

a. Dependent Variable: Soft.Rock.Pop

Figure 35: Coefficients of Openness to experiences with Soft rock/Pop music genre

Coefficient for each independent gives the size of the effect of that variable is having on the dependent variable, and the sign on the coefficient gives the direction of the effect. Implementing so in our case, we can say that Soft Rock/Pop is likely to be linked with the openness to new experience variable.

Hence, we can deduce that the people who like Soft Rock/Pop music are more open to new experiences as a general.

9) Are the ratings across music genre and prefered music genre significantly different?

There were some of the variables which were significantly different, and they are listed as below.

Genre	Significant	Significant Value
Country	Yes	0.02986
Rap/Hip Hop	Yes	0.0002714
Pop/ Soft Rock	Yes	2.2e-16
Jazz	Yes	1.359e-07
Classical	No	0.3537
Hard Rock/Heavy Metal	No	0.889

Figure 36: Significant Difference between music genre and preferred music genre

10) If the answer to (9) is positive, what factors explain the difference?

Factors explaining the difference

For the variables which were significantly different are Jazz, Pop/Soft Rock and Rap.

The following table describes the factors which were causing the difference in the significant values.

Genre	Factor	p-value	Estimate
Jazz	Openness to New Experience	0.00297	0.123986
Pop/Soft Rock	Openness to Experience and Agreeableness	0.00850 and 0.04148	0.12142 and -0.9268
Rap	Age: 27 to 31	0.0401	0.31818

Figure 37: Factors causing the significant difference

11)Is there a correlation between an individual's geographical location and their musical preference?

To find the correlation between an individual's geographical location and their musical preference we pulled the qualtrics data and tracked down the latitude and longitude required to relate the geographical location and their musical preferences. This required arcGis software and analysis like Spatial Autocorrelation and spatial significance analysis. Further information will be covered in the postmortem report.

Conclusion

After conducting our analysis, we found the following relationships to be significant. Country music preference is negatively related to Openness to New Experiences Score. Agreeableness is negatively related to Rap music preference

Emotional stability is positively related to Rap music preference. Negative mood fully mediates the relationship between Emotional Stability and Country music preference. The more open people are, the more number of hours of music they listen to which means Most favorite among all the individuals is Soft Rock/Pop

Extroversion is negatively related with Softrock/pop and classical genres. People with high agreeableness are more likely to rate pop higher and emotional stability is positively related with Jazz.

References:

1. “Assumptions of Linear Regression.” *Statistics Solutions*, www.statisticssolutions.com/assumptions-of-linear-regression/
2. “Assumptions of Logistic Regression.” *Statistics Solutions*, www.statisticssolutions.com/assumptions-of-logistic-regression/
3. Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. Journal of Research in Personality, 37, 504-528
4. Mayer, J. D., & Gaschke, Y. N. (2013) . Brief Mood Introspection Scale (BMIS) . Measurement Instrument Database for the Social Science. Retrieved from www.midss.ie
5. Baron, Reuben M., and David A. Kenny. "The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations." Journal of personality and social psychology 51.6 (1986): 1173.