Personality Analysis of Sharing Economy Consumers: An Application of Text Mining

Murat Acar
Bogazici University
Aysegul Toker
Bogazici University

Cite as:
Acar Murat, Toker Aysegul (2019), Personality Analysis of Sharing Economy Consumers: An Application of Text Mining. *Proceedings of the European Marketing Academy*, 48th, (8813)

Paper presented at the 48th Annual EMAC Conference, Hamburg, May 24-27, 2019.

EUROPEAN MARKETING ACADEMY

Personality Analysis of Sharing Economy Consumers:

An Application of Text Mining

Abstract:

Data-driven marketing has given rise to the search for new data sources to extract valuable insights. Researching consumer behaviors through data of segments like demographics and transactions is now obsolete since the idiom of "you" has become popular in marketing phenomena, especially in Sharing Economy. Airbnb is one of the disruptive examples where guests (i.e., consumers) and hosts (i.e., suppliers) reside in the humanization of consumer-supplier relationship under a new form of contractual agreement. This service context offers a great opportunity to analyze user-generated content like guests' textual reviews on the hosts' listings (i.e., rooms or apartments) upon the service encounters. Our main goal is to extract Airbnb guests' Big5 personalities using their personal texts by utilizing an Artificial Intelligence application. The significant findings include that the Airbnb consumers score high in Extraversion and Openness dimensions of Big5.

Keywords: Sharing Economy, User-generated Content, Text Mining

Track: Digital Marketing & Social Media

1. Introduction

Sharing Economy (SE) has revealed new horizons for service marketing research, which includes the shift from ownership to access, the change in value mindsets, and humanization of consumer-supplier relationship, especially in technology-based service encounters. Specifically, Airbnb offers a new form of contractual relationship, and the analysis of its consumers is to be fine-grained by data-driven marketing that uses Big Data and Artificial Intelligence (AI) in the era of computational social science.

Marketing researchers can extract information from structured and unstructured data sources. The common types of structured data are demographics and transactions, which do not necessarily yield data-driven insights. With that, psychometric data might be used to better model consumer behavior and perceptions (Liu et al., 2016). For example, human language and profile are psychologically rich, which underlie many personality traits (Kosinski, Matz, Gosling, Popov, and Stillwell, 2015). Psychographic findings including psychological states (variability within consumers over time) and traits (variability across consumers) might be used to attract consumers (Matz & Netzer, 2017; Sarli & Tat, 2011). Also, Wedel and Kannan (2016) reported that the personalization of the marketing mix by using Big Data and AI techniques is a promising field for marketing analytics research, which is one of the baselines of this study.

Psychographic or psychologically-personalized marketing has emerged in response to conventional segmentation, targeting, and positioning. In general, personality analysis is mostly carried out by survey-based assessments where the responses depend on self-reporting. A user-generated text might accompany and be even superior to structured data alternatives for personality analysis. The purpose of this study is to extract Airbnb guests' textual review data on Airbnb listings, which is in free text format, and to demonstrate personality findings of guests by text mining.

2. Dataset

We collected the Airbnb guests' textual review data from 45 cities all around the globe including 9,982,450 distinct reviews written in 74 distinct languages. In the data set, there are 7,264,026 distinct reviewers across 417,395 distinct listings (see Appendix-1 for details). Two top review languages are selected: English and Spanish. All the Airbnb room types are included in data: Entire Room (ER), Private Room (PR), and Shared Room (SR). Appendix -1 also shows the percentages of room types per Airbnb location. After data cleansing, 528 guests are selected having 1500+ words within the entire data repository. 16 distinct guests were excluded due to weak significance results and multi-language reviews that do not sum up to desired word counts per English and Spanish language. These exclusion criteria resulted in our final dataset which includes 512 guests having 1500+ words across all listings with clear language according to the data quality analysis made by R-Studio. We have cleared out all automated Airbnb responses like «*The host canceled this reservation X days before arrival. This is an automated posting*. »

3. Research Method

IBM Watson Personality Insights AI application is a cutting-edge tool for personality analysis by using social media, enterprise data, or other digital communications. From targeted marketing to customer acquisition, Personality Insights might have a wide range of use cases. This tool resides on the commonly accepted hypothesis that human language implicitly reveals personality. The language text can include opinions, experiences, attitudes, and sentiments. With a vector representation of the words in the input text, the application employs a machine-learning algorithm (*which do not input user demographics*) that output a personality profile with the following Big5 dimensions:

- Agreeableness: Tendency to be compassionate and cooperative toward others
    - Facets: Altruism, Cooperation, Modesty, Morality, Sympathy, Trust
- Conscientiousness: Tendency to act in an organized or thoughtful way
    - Facets: Achievement-striving, Cautiousness, Dutifulness, Orderliness, Self-discipline, Self-efficacy
- Extraversion: Tendency to seek stimulation in the company of others.
    - Facets: Activity-level, Assertiveness, Cheerfulness, Excitement-seeking, Friendliness, Gregariousness
- Emotional range, namely Neuroticism: Tendency to be sensitive to the environment
    - Facets: Anger, Anxiety, Depression, Immoderation, Self-consciousness, Vulnerability
- Openness: Tendency to experience different activities
    - Facets: Adventurousness, Artistic interests, Emotionality, Imagination, Intellect, Liberalism

The application is pre-trained with thousands of Twitter users to position the ground-truth. We have used Curl command line tool and library for transferring text data to the AI application that we customized for our needs. Using Personality Insights' API key and URL within Curl commands, we have received JSON outputs. The outputs include the personality findings and significance results for each Big5 dimension and the facets (6 per dimension, total of 30) under those five dimensions separately. Figure 1 depicts our research method.

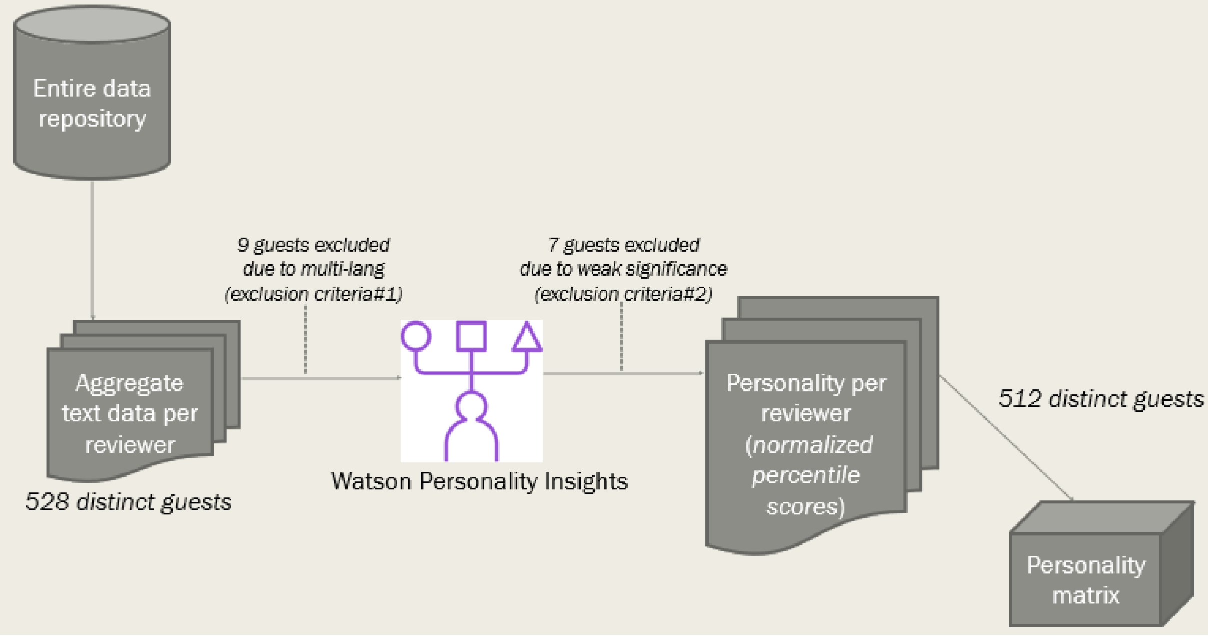


Figure 1 Our research method

English and Spanish are the top two performing languages based on the MAE (Mean Absolute Error) and average correlation results (see Table 1) reported by the application that compared actuals with inferred scores. MAE scores that are closer to 0 are better. For correlation, scores greater than 0.2 are acceptable.

Language	**Big5 dimensions**	**Big5 facets**
English average MAE	0.12	0.12
English average correlation	0.33	0.28
Spanish average MAE	0.10	0.12
Spanish average correlation	0.35	0.21

Table 1 Precision of Personality Insights for English and Spanish text

We have also used many open-source tools, especially in data cleansing and preparation:

- Google language detector 2 and 3 (*CLD2* and *CLD3*) for detecting text languages

- R-Studio packages like *textcat* for text categorization, *hunspell* for spell checking, *sqldf* for SQL-like querying on the dataset, *xlsx* and *openxlsx* for individual personality scores, *readr* for csv operations, *qdap* for aggregating data by grouping and visualization of text

4. Major Results

4.1. Interpretation of results

For each request, Personality Insights application reports a normalized score as a percentile (as a double in the range of 0 to 1) for each Big5 personality trait, which is based on the qualities that the application infers from the input text. For all the upcoming results, we have used the following criteria as shown by IBM researchers in the Personality Insights documentation:

- A percentile score at or above 0.75 is considered as high. Any score above the mean of 0.5 indicates an above average tendency of our sample for a characteristic.
- The reverse statements are true of scores below 0.50 and 0.25, which are considered as below average and low respectively.

Since we want to know how Airbnb guests' characteristics compare with a large population, we use the normalized percentile scores instead of raw scores. There is no mathematical relationship between the percentiles that are reported for Big5 dimensions and facets, which are calculated independently. With that, even the facets are sub-descriptors of Big5 dimensions, adding the scores of facets does not necessarily give the results for dimensions. We graphed the percentiles for 512 guests within a histogram for all the personality traits. Skewness itself yields how our sample set resides for a single characteristic.

4.2 Results

IBM Watson Personality Insights AI application outputs the significance results per each Big5 dimension and facet. For our sample set, we calculate the mean, standard deviation and variance measures for the normalized scores of application's outputs. The statistically significant results (see Appendix-2) based on the interpretation criteria include that Airbnb guests score high in Altruism, Cooperation (i.e., Accommodating), Sympathy (i.e., Empathetic), Trust (i.e., Trusting of others), Cautiousness (i.e., Deliberate), Dutifulness (i.e., Respectful in rules and obligations), Activity-level (i.e., Energetic), Extraversion, Artistic interests (i.e., Appreciative of art), Intellect (i.e., Philosophical), Liberalism (i.e., Authoritychallenging), and Openness (i.e., Open to experiencing). On the other hand, they score low in Excitement-seeking (i.e., they are calm-seeking), Gregariousness (i.e., they feel independent), Anger (i.e., they are mild-tempered), and Self-consciousness (i.e., they are hard to embarrass).

5. Conclusion

"Understanding and knowing consumers themselves" has become more important than "Predicting and knowing more about consumers" in marketing analytics. The psychology of consumer behavior implicitly precedes the acquisition of consumers' relevance in multichannel digital marketing. Psychographic or psychologically-personalized marketing that utilizes AI and Big Data might be a game changer in today's human-centered engagement models and disruptive innovations where experience is at the center. In this study, there are pioneering implications for travel marketing and service researchers to reveal traveler personalities inferred from user-generated content in SE Airbnb service context. Linear combinations of personality scores might be used to come up with distinct organic behaviors of consumers (i.e., associations between separate psychological traits). As Yoo and Gretzel (2011) reported, the influence of personality on travel-related user-generated content creation can be further researched from this study's point of view.

References

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543.

Liu, Z., Wang, Y., Mahmud, J., Akkiraju, R., Schoudt, J., Xu, A., & Donovan, B. (2016, November). To buy or not to buy? understanding the role of personality traits in predicting consumer behaviors. In *International Conference on Social Informatics* (pp. 337-346).

Matz, S. C., & Netzer, O. (2017). Using big data as a window into consumers' psychology. *Current opinion in behavioral sciences*, *18*, 7-12.

Sarli, A., & Tat, H. H. (2011). Attracting Consumers by Finding out Their Psychographic Traits. *International Journal of Fundamental Psychology & Social Sciences*, *1*(1), 6-10.

Yoo, K. H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior*, *27*(2), 609-621.

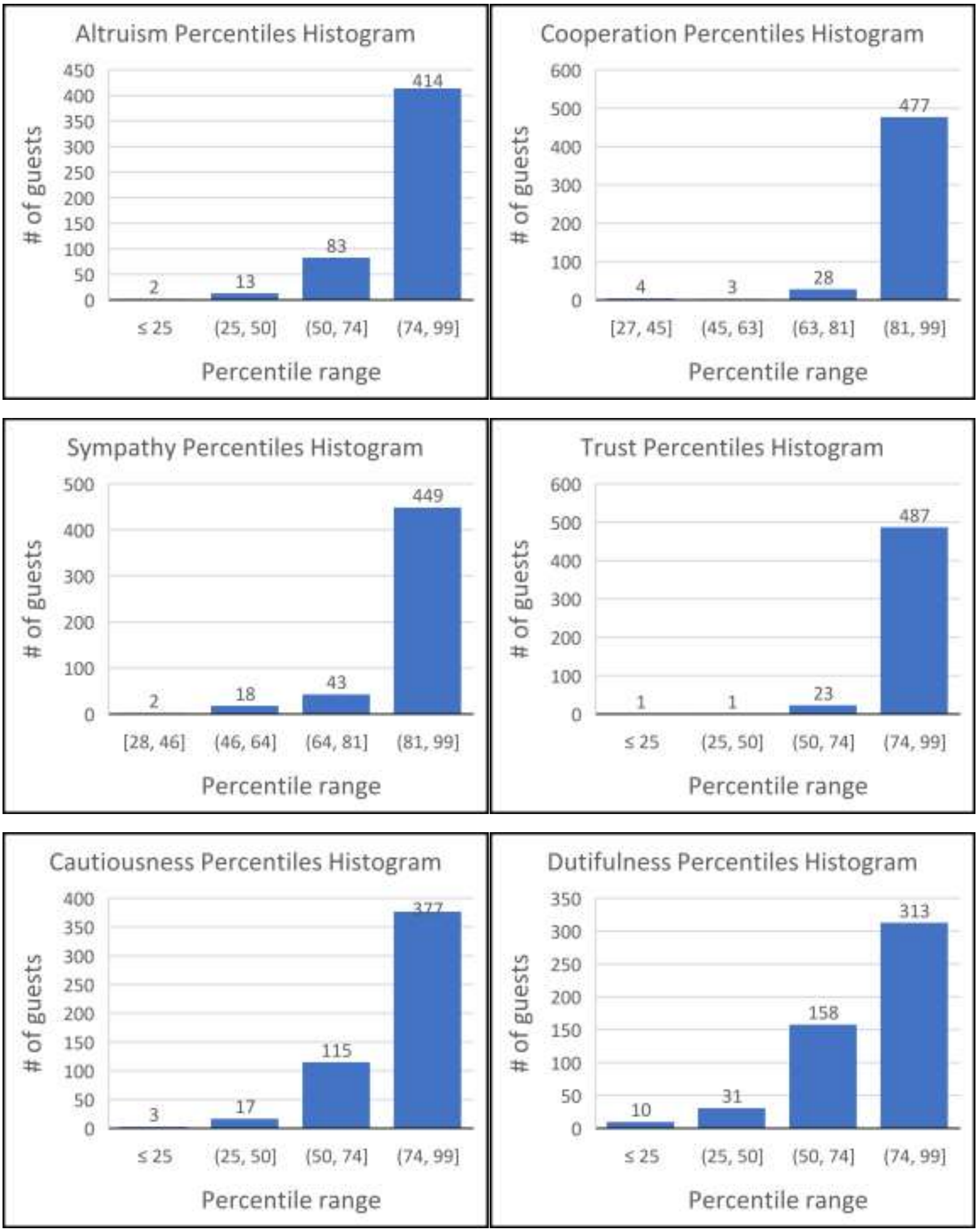
Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, *80*(6), 97-121.

Appendix – 1 Airbnb locations within our data set and statistical results

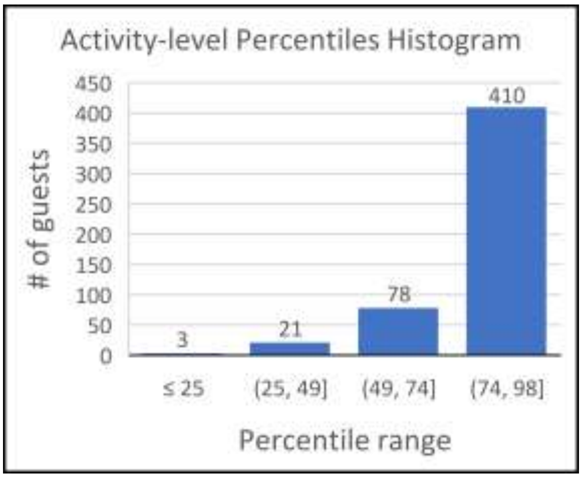
(*ER%: % of entire room listings, PR%: % of private room listings, SR: % of shared room listings*)

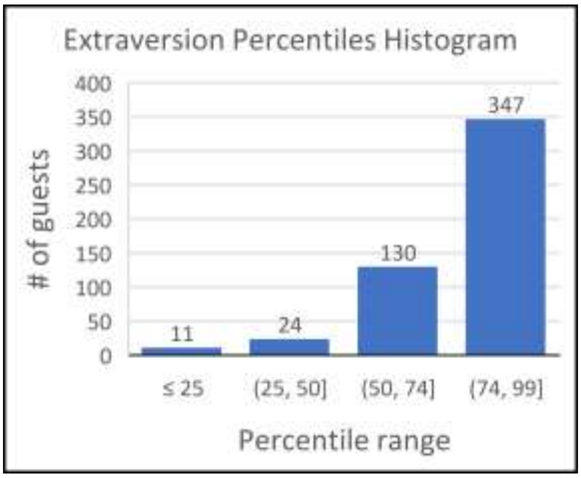
Locations in Airbnb Data	Number of reviews	Number of distinct reviewers	Number of distinct listings	EH %	PR %	SR %	Average word count per review	Standard Deviation of word count
Amsterdam, The Netherlands	337816	323133	16157	80.1	19.7	0.2	56.45886	47.39348
Antwerp, Belgium	26644	25112	1024	71	27.9	1.1	48.50977	39.43244
Asheville, United States	27721	25669	742	61.5	37.9	0.6	67.5114	50.00364
Athens, Greece	124377	112047	3927	83.2	15.8	1	63.55369	52.37973
Austin, Texas, United States	134550	117591	6007	69.4	28.3	2.3	57.46183	47.94543
Barcelona, Spain	500413	473204	14838	40.1	58.9	1	55.13578	50.45597
Berlin, Berlin, Germany	266555	245576	16180	52.2	46.5	1.3	50.93166	42.95077
Boston, Massachusetts, US	120787	112540	3986	62.2	36.6	1.2	53.42348	50.01038
Brussels, Belgium	112060	103056	4904	64.5	34.2	1.3	48.21616	41.57828
Chicago, Illinois, United States	132353	121572	4497	59.7	37.2	3.1	57.75287	51.369
Copenhagen, Denmark	221047	206585	16445	80.9	18.7	0.5	56.17079	44.86127
Denver, Colorado, US	128834	116788	3406	68.1	30.1	1.8	48.27096	43.61672
Dublin, Leinster, Ireland	141152	130051	5361	47.3	50.1	2.6	59.24268	47.81057
Edinburgh, Scotland	259295	239798	8557	57.2	42.5	0.3	52.86119	43.70517
Geneva, Switzerland	46583	40780	2359	66.1	33	0.9	43.57983	41.27622
Hong Kong, China	82394	73759	4617	50.1	44.7	5.2	54.88736	61.3998
London, England	672760	571730	37438	51.5	47.1	1.4	58.641	51.35805
Los Angeles, California, US	651938	529730	24030	63	32.9	4.1	54.56894	51.44898
Madrid, Spain	444774	399996	13261	64.1	34.8	1.1	48.5384	45.02056
Malaga, Spain	98114	91698	3834	79.1	20.2	0.7	50.81432	45.93124
Mallorca, Spain	109662	97733	8407	87.9	12	0.1	66.89309	56.74921
Manchester, England	14880	13717	676	42	56.6	1.4	54.7132	43.69469
Melbourne, Australia	231550	193179	11223	61.3	37.2	1.5	50.63239	43.66092
Montreal, Canada	97208	86871	7028	60.7	38	1.3	61.55244	50.13677
Nashville, Tennessee, US	170343	154694	4570	76.5	22.5	1	52.25317	44.8812
New Orleans, Louisiana, US	188329	174798	4723	83.1	16.3	0.6	62.0787	55.3435
New York City, New York, US	896208	781742	37916	51	46,9	2.1	57.15083	53.04253
Northern Rivers, Australia	24951	22630	1726	77.1	22.7	0.2	62.58451	47.47128
Oakland, California, US	26736	23346	1311	56.1	40.5	3.4	66.50814	53.94048
Paris, France	969581	854834	45797	87.3	12	0.7	53.2519	51.12878
Portland, Oregon, US	224755	196178	4204	68.1	30.6	1.3	55.31222	46.70338
Quebec City, Canada	50396	47884	1963	63.7	35.2	1.1	47.32303	42.77186
Rome, Italy	572969	540706	18758	60.3	39	0.7	66.21876	55.64346
San Diego, California, US	92862	85263	4590	66.1	31.1	2.8	67.04618	55.21456
San Francisco, California, US	256635	230331	7123	57.6	40.4	2	57.94606	50.78153
Santa Cruz, California, US	22121	20455	702	65.5	33.1	1.4	69.98422	54.60817
Seattle, Washington, US	84849	75730	3191	66.8	30.2	3	68.49241	54.06384
Sydney, Australia	335036	277825	22685	60.7	37.6	1.7	46.92218	44.10497
Tasmania, Australia	138532	97564	3909	76.5	22.9	0.6	45.95982	39.0521
Toronto, Ontario, Canada	203887	170733	9895	65.1	33.2	1.7	50.43339	46.34209
Vancouver, Canada	149595	135048	5660	69	30.1	0.9	52.92683	46.33905
Venice, Italy	216305	210898	5243	77.2	21.9	0.9	65.64307	55.45585
Victoria, Canada	31243	28226	1441	76.3	23	0.7	64.94363	49.11254
Vienna, Austria	191102	180407	7424	72.1	26.9	1	51.62907	44.55493
Washington, D.C., US	152548	139705	5660	68.7	28.8	2.5	58.26675	51.48371

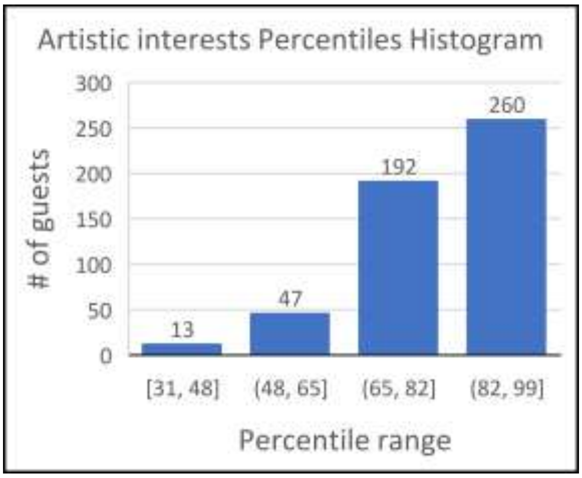
Appendix -2 Personality Findings as the Frequency of Normalized Percentile Scores

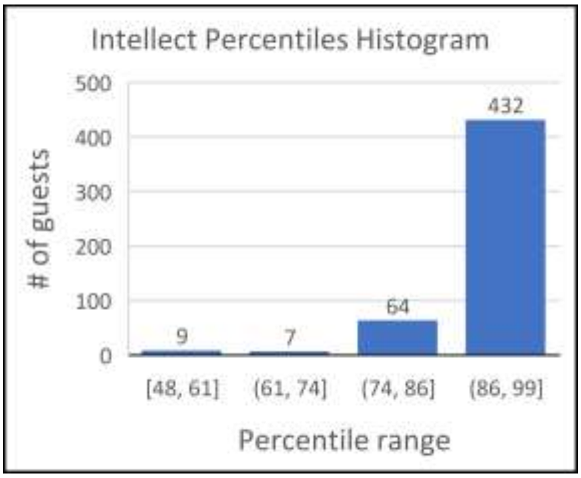


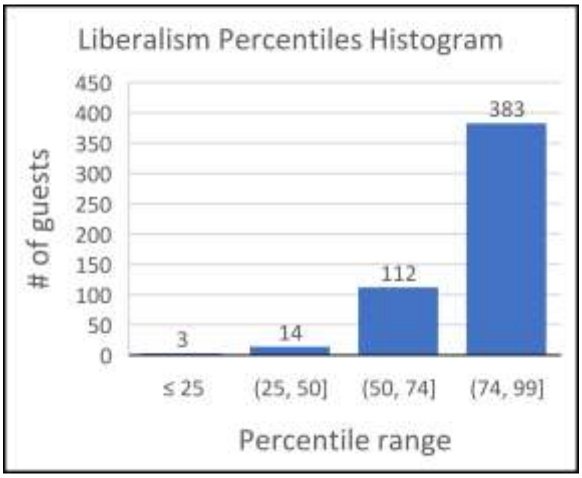
Appendix -2 (continued)

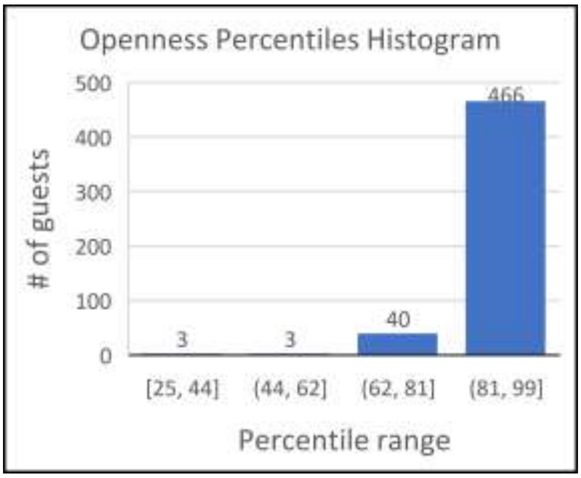












Appendix -2 (continued)