Computational Social Science
Project Report

Gender Gap In Electronic Music Scene

Basitur Rahman Chowdhury
Matriculation Number- *218100976*
bchowdhury@uni-koblenz.de

Muhammad Momotazul Islam
Matriculation Number- *218203293*
momotazulislam@uni-koblenz.de

Md Tariqul Islam
Matriculation Number- *218100921*
tariquldipu@uni-koblenz.de

Universität Koblenz-Landau
Koblenz, Germany

1. Abstract
With the introduction of new digital services and platforms to enjoy music, the global music consumption is skyrocketing in almost every part of the world. New artists are going into production and new genres of music are being introduced everyday. In order to ensure that listeners can always keep themselves updated with music related information, websites like `musicbrainz`, `songkick` or `discogs` are regularly storing various information on artist's personal profile, their recorded releases, live shows and collaborations. This report will attempt to give a comprehensive idea about the gender inequality that persists in the Electronic Music (EM) scene; based on the data that were obtained from these websites. The aim is to provide a clear perception about the existing gaps between male and female EM artists in terms of their participation, opportunity, remuneration and output in the industry. We also try to demonstrate the trend of gender inequality over the past few decades. The findings can help us to acknowledge the contribution of female artists in electronic music and inspire gender equity in order to form a competitive and innovative music industry around the world.

2. Introduction
People of the modern world are now realizing gender equality is a very fundamental aspect for sustainable development. However, in every part of the world, almost in every sector, women are still subjected to gender discrimination, often barring them to flourish their capabilities and practice their skills. Artistic world is not exempted from these stereotypes as we can observe. Music industry is an inseparable member of this world and this too needs an equal participation from the two dominant gender groups in order to provide justice to itself. There are a lot of genres in today's world music and Electronic Music (EM) is one of them. Our focus is primarily on the gender bias that exists in this sector i.e. how female EM artists are doing compared to their male counterparts.

From clubs to our very own kitchen, people are enjoying music every moment. Online streaming providers are paving the way to access the world music with just a single tap on the listener's smart devices. Thus We can enjoy any music created from any part of the world. By successful circulation of these productions, music industry worldwide is now worth US$1.4 trillion[1]. Electronic Music is a widely popular genre and values up to US$7.4 billion[2]. This colossal industry comprises of both male and female artists and also artists from other gender groups. However, very few tried to address the discrimination between the two major gender groups. Our quest is to answer if there is actually a gender bias existing in EM scene.

We should also assess the importance of investigating these questions. When and Where gender equality is not inspired, half of the population is being deprived of their basic human rights[3]. When an artist is tackled only because of his/her gender orientation, it affects not only him/her, it also affects the possibility of future artistic endeavors of others. A healthy competition in order to produce quality work is also hampered by this. Gender equality is thus very crucial when it comes to building a music industry where every artist gets absolute equal opportunity to practice and demonstrate his/her skill.

3. Research Questions

In this project, we will investigate- How big is the gender gap in the EM scene? And how it has evolved over the last decade?

For investigating this question some aspects will be considered. For instance,

Representation Gap: ratio between the number of male and female artists

Productivity Gap: ratio between number of releases by male and female artists

Opportunity Gap: ratio between number of gigs and festivals participation by male and female artists

Earning Gap: difference between the average album release prices by male and female artists

4. Methodology

For answering the main questions of this project, we had to work on various datasets extracted from different websites. There are several websites which contain a wide range of information regarding artists working in the electronic music scene. A dataset that is used here were extracted from `musicbrainz`. Other datasets include `songkick`, `junodownloads` and `discogs`. `Junodownloads` and `discogs` have a broad spectrum of artists' release data such as artist names, their album names , release year and album price. Here, we had different types of data distributed in different datasets. So, we had to merge these release information and incorporate them during the compilation process. However, `junodownloads` and `discogs` do not contain information like artist gender or place of birth.

[1] "the broader music industry - IFPI." https://www.ifpi.org/content/library/the-broader-music-industry.pdf. Accessed 19 Jul. 2019.

[2] "Electronic music industry sees significant growth and is worth $7.4 billion." 24 May. 2017, https://mixmag.net/read/electronic-music-industry-sees-significant-growth-and-is-worth-7-4-billion-news. Accessed 19 Jul. 2019.

[3] "Articles 16-30, United Nations Declaration of Human Rights : Youth ...." https://www.youthforhumanrights.org/what-are-human-rights/universal-declaration-of-human-rights/articles-16-30.html. Accessed 19 Jul. 2019.

The website *musicbrainz* contains general information of most of the artists in the world. At first, we needed a way to filter out data that are only related to electronic music genre. For this, we employed some keywords (e.g. genres) that only refer to electronic music.

At this point, we have crawled artists' meta information (name, gender, origin, birthdate, genre) from *musicbrainz* using the above EM genre keywords. So, finding out the necessary gender ratios from this dataset is now possible. Following, we also merged data from songkick & musicbrainz. This data was necessary because we are also trying to visualize gender inequality in live shows. Songkick data contains live gig information including venue name and country. We found 164 countries and showed a comparative visualization of live gigs in terms of men and women artists.

Feature scaling is a method used here to normalize the range of independent variables or features of data.

Also known as min-max scaling or min-max normalization, is the simplest method and consists of rescaling the range of features to scale the range in [0, 1] or [−1, 1]. Selecting the target range depends on the nature of the data. The general formula for a min-max of [0, 1] is given as[4]

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

5. Data Crawling & Pre-processing

For analysing every aspect, we needed data on artist number, artist gender, their live shows and album releases. We had been provided with some datasets from Mr. Mohsen Jadidi of Gesis which were obtained from his earlier research projects. However, we still needed to build a dataset which provides us with information on artist gender, genre and their type (i.e person or group).

The crawler was developed with python libraries (*BeautifulSoup* and *request*) for making the crawling process simple. For collection of the data we used *musicbrainz* website which contains general information for the music artists around the globe. The crawler was made to extract some specific data (i.e. ID, Name, Gender, Genre, Type, Birthday) for each artist.

Musicbrainz has a downloadable list[5] of all artists. This list contains various information but among these only artist IDs, names and their dates of birth were useful to us. Afterwards, we created a url- musicbrainz.com/artist/ID.

When any specific artist's ID were put in the url, we obtained that artist's html page. After restructuring the page using *soup*, we ran *find* operations in order to obtain their gender, genres and types.

Our main goal is to investigate the gender gap only in the electronic music scene. The dataset was then filtered according to the pre-specified EM genre keywords.

We specifically used the data sets *songkick_all_gigs_extended_version* and *releases_merged* from the already provided datasets of Mr. Jadidi. *songkick* data set contains data like- ID, Name, Date, Concert Venue,

[4] https://en.wikipedia.org/wiki/Feature_scaling

[5] "MusicBrainz Database / Download - MusicBrainz." https://musicbrainz.org/doc/MusicBrainz_Database/Download. Accessed 20 Jul. 2019.

City and Country. And *releases merged* data set contains data like- Name, Artist ID, Album Name, Release ID, Label Name, Genre, Album Price, Year etc. This data set was created by merging two separate data sets from the websites *junodownloads* and *discogs*.

6. Data Cleansing

So finally we have three separate datasets- *musicbrainz*, *songkick* and *releases merged*. We merged these three separate datasets into a single dataset. Here, we list all the data cleaning and preparation steps taken in this stage, that will lead us to the final data set-

1. Drop the ID column from the *musicbrainz* dataset. Now we only have columns for Name, Gender, Genre, Type, Birthday. We transformed all the names to lower case letters.
2. We filtered this dataset with the Type: Person and dropped all the entries that refers to 'band' or other types. Furthermore, we again filtered the dataset with electronic music genres.
3. From the *songkick* dataset, we transformed the Date column into Year (1972-2019). All the gig entries are counted for each artist and then added to a new column GigCount.
4. Also all the concert places in *songkick* dataset for each artist are listed into a new column named Countries. Similarly all the concert years for each artist are listed into a new column ConcertYears. So, now in each row, for each artist, we have all the concert/gig countries and concert years listed in separate cells.
5. We merged the *musicbrainz* and *songkick* datasets into a single dataset named *merge1* on the basis of column *Name*.
6. From the *releases merged*, we dropped the columns artist_id, release_id, label_name, Genre, Release_Name.
7. For each artist, we listed all the album release year (1992-2019) into a single cell under a new column named releaseYear.
8. We also counted all the releases for each artist and put it inside a new column AlbumCount. We also calculated the average price of albums for each artist and listed it inside the column avgPrice.
9. Finally, we merge this modified *releases merged* dataset with the previously made *merge1* dataset and create the final dataset ***finalMerge***.
10. Our final dataset ***finalMerge*** contains the following informations- Name, Concert Countries, Gigs Count, Concert Years, Gender, Birthday, Album Release Years, Albums count, Average of all albums for each artist.
11. Feature scaling is used to normalize the range of gigs and album release years data for male and female.
12. The usable data from the musicbrainz website after filtering with the EM genre list, about 4083 artist's data has been filtered. The final dataset ***finalMerge*** ended up having 549 artist's data. So on the basis of crawled data the percentage of data used is about 13.45%

We will use this ***finalMerge*** dataset for our data analysis and visualization purpose which will be discussed in detail in the following chapters.

7. Data Visualization

After having the prepared and finalized dataset, visualization stage of the data enables to understand the real scenario of the analysis and to make comparisons. For making the visualization of the dataset more efficient and effective Pandas, Numpy, Matplotlib, Scipy, Seaborn, Scikit-learn libraries of Python are used. Besides, Gig year count and Album year count have been normalized to see the proportional insights for male and female participation in Album releases and Concerts. However, function has been introduced for counting the album and gig years to make the dataset more usable to draw the plots, i.e. Pie plot, Bar plot. These data visualization helped to investigate those data in a more efficient and effective way. The detailed discussion and inferences on plots are described in the next chapter Analysis: Revealing the Insights.

8. Analysis: Revealing the Insights

To methodically analyze the gender differences in the EM music scene, we have worked with some aspects of our final dataset that will help us to analyze and interpret the male-female dissimilarity.

Representation Gap Analysis: The simple analysis conducted on the datasets containing 549 artists, we found 485 male and 49 female artists which gives a ratio of 90.8:9.2 respectively. However, 15 artists has been filtered out because of no gender identification.

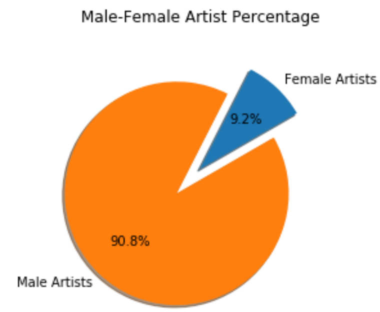

Figure 1 Male-Female Artist Percentage

Productivity Gap Analysis: The following pie chart depicts the average album release ratio of male and female artists. The 49 female artists are responsible for mere ~5% of the total album released from the year 1961 to 2018. The rest of the album has been released by the 485 male artists from the dataset.

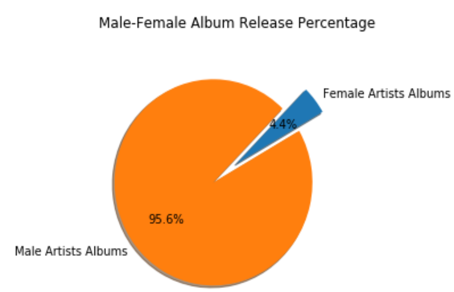

Figure 2 Male-Female Album Release Percentage

Normalized Analysis of Album Releases

1. The above bar plot gives an insight to the number of releases by both gender groups per year where the time frame is from 1992 to 2019.
2. Overall Album release increases almost exponentially for male till the year 2012.
3. After 2012, the number of male albums have declined for the

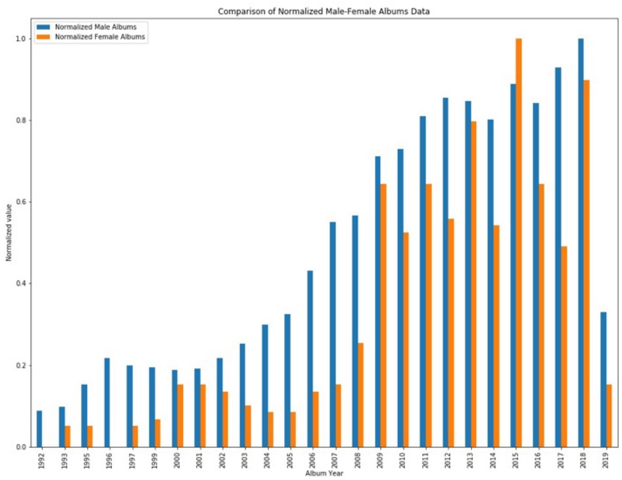

Figure 3 Male-Female Album Release (normalized) Data

following two years. Again, it started to spike up in 2015. After a brief fall in 2016, the number continues to rise till 2018.

4. The number of releases by female follows an irregular pattern throughout the whole time span. The years 1992 and 1996 have no album releases by female artists according to our dataset (it is possible that despite few releases, no data was recorded).
5. However, it is interesting to note that the female artists actually released the highest ever number of albums (60) in 2015 (similar to the trend of album releases of male in 2018).
6. Moreover, it is quite unreliable to predict the trend for 2019 (current year) based on the available data. Firstly, the data is not yet complete and secondly because following the most releases by females in 2015, the number significantly dropped in the following two years, only to see a rise in 2018.

Opportunity Gap Analysis: Opportunity gap reveals the ratio of concerts/gigs conducted by male and female artists. The overall ratio between the male and female in terms of gigs/ concerts is 91.8: 8.2 .

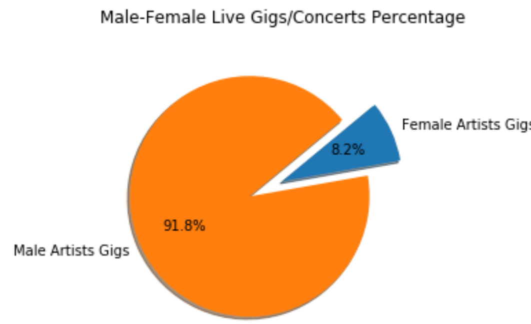

Figure 4 Male-Female Gigs Percentage

Normalized Analysis of Gigs Participation

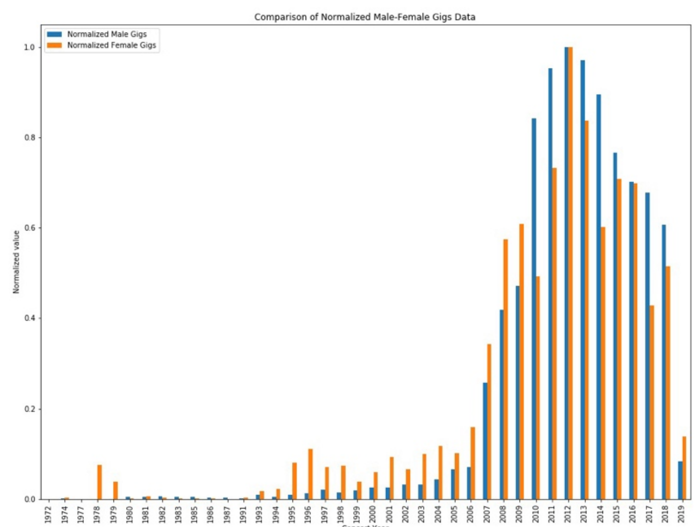

Figure 5 Male-Female Gigs (normalized) Data

1. Above bar plot is a comparative study of the male-female live gigs/concerts by year (1972-2019).
2. It would be quite unconvincing to try to create a picture from the´70's till early ´90's because either there were very few live shows or most of the shows were undocumented.
3. It can be observed that from 1991, the participation of female artists in live shows started to increase. This trend continued almost exponentially till 2012, during

when the ratio assumes an equal value for both male and female.

4. On the other hand, male artists have performed in the live shows in almost a similar pattern. Only after 2012, their participation has been decreasing so far.
5. Moreover, It is very interesting to note that in the year 2019, the proportion of male artists performing in live shows is lower than female artists so far .
6. It is also easily observable that the pattern for female artists is irregular whereas the data on male artists saw a regular pattern (i.e. overall increase before 2012 and decrease afterwards).

Average Album Price Discrimination for Male vs Female: The Average album price for male artist is about €6.76 while the female artists are getting paid about €4.50 on average for their music albums.

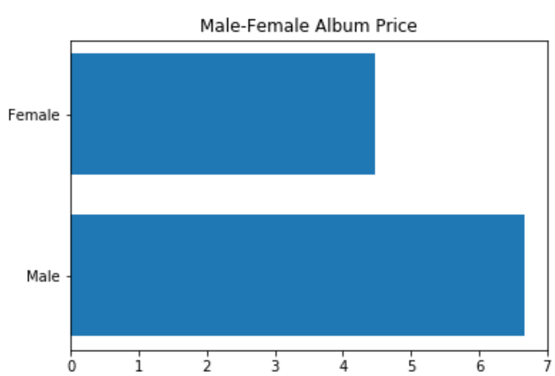

Figure 6 Male-Female Album Price

9. Results

After analyzing the data, it is evident that there exists a significant gender gap in the Electronic Music industry.

1. The representation gap between both gender group is acute as we can see from the pie plot; the number of total male artists is a lot higher than their female counterparts.
2. The productivity gap also persists in almost every year. The ratio of album releases by female artists are lesser than that of males.
3. The remuneration of women is also inferior to men; an average price gap of more than €2 can be observed.
4. However, it is worth mentioning that, the female group were conducting more live shows in the past and their participations are increasing almost every year. As of this year, women are the leading performers in the live gigs/concerts. It can be safely concluded that, female artists are getting more opportunities when it comes to performing live shows.
5. The table depicts the number of album releases and number of concerts for both male and female over the last decade.

Year	Original Male Gigs	Normalized Male Gigs	Original Female Gigs	Normalized Female Gigs
2010	9577	0.841181	504	0.493137
2011	10843	0.952389	748	0.732353
2012	11385	1.000000	1021	1.000000
2013	11051	0.970661	854	0.836275
2014	10187	0.894765	615	0.601961
2015	8723	0.766163	724	0.708824
2016	7998	0.702477	713	0.698039
2017	7713	0.677442	438	0.428431
2018	6918	0.607607	526	0.514706
2019	948	0.083187	142	0.138235

Year	Original Male Albums	Normalized Male Albums	Original Female Albums	Normalized Female Albums
2010	795	0.728440	32	0.525424
2011	884	0.810092	39	0.644068
2012	932	0.854128	34	0.559322
2013	924	0.846789	48	0.796610
2014	874	0.800917	33	0.542373
2015	970	0.888991	60	1.000000
2016	919	0.842202	39	0.644068
2017	1013	0.928440	30	0.491525
2018	1091	1.000000	54	0.898305
2019	361	0.330275	10	0.152542

10. Limitations

Throughout our project development, we faced several difficulties which limited our data processing and analysis. Here are the problems that we have faced--

1. There are numerous sub-genres that fall under Electronic Music. And lots of new genres are being introduced every now and then.

We were not able to find a concise list of all the EM sub-genres.

2. In our dataset, we have many entries where information for artists are not complete. More precisely, for many artists we could not find information like gender, origin, concerts/gigs, album releases etc. These specific informations are crucial for our analysis. So we had to omit those entries with missing data.
3. `Discogs` has half a million data entries. These data entries contain information like concert performances and albums releases for each artist. But compared to `Discogs`, `MusicBrainz` has a very small metadata (like artist name, gender, type, origin, birthdate) collection. So we could not merge this huge `Discogs` data with the small `MusicBrainz` data.
4. Merging `MusicBrainz` data with `Songkick` data resulted in many incomplete entries and discrepancies as `Songkick` dataset is also not complete.
5. Some artists' names that we found on `MusicBrainz`, were not listed in `Songkick`. The vice versa case has also happened. This compelled us to omit many data entries from our consideration.
6. Earning data from album sales are not complete, as album prices are changing regularly. And these changes follow no particular fixed pattern.
7. It would be very useful if we could investigate how the album price evolved over the past years. Unfortunately we did not have the data on album price in different time frames rather we only had their current prices.
8. Different music streaming services are the most prominent source of income for artists in modern times. But we were unable to fetch earning data for artists from the streaming services.

11. Challenges for Afterward Works

1. We only used the data that were extracted from `musicbrainz`. There are a lot of music encyclopedias on the internet that regularly store and organize data on artists. Those data can also be useful for more elaborate works.

2. Newer music genres are coming out everyday. It is possible that the gender bias in the production of those genres can be of a different picture.

3. One analysis can be done in terms of collaborations among the artists. Artists regularly collaborate among themselves. One question can be raised- are these collaborations gender biased? How does this picture evolve over the different parts of the world?

4. It will be useful to draw a clearer picture of remuneration gap if the pricing gap between male and female artists can be analyzed in different time frames. Was the difference of release prices greater in the last century? To analyze this question one approach is to find a dataset where the album price are listed according to their release years.

5. With the increasing popularity and usage of music streaming platforms like Spotify or Youtube Music, artists are getting a considerable portion of their royalties through these services. Are Spotify and Youtube Music prone to prioritize male electronic artists? This can

also be interesting to investigate in the future.

12. Related Work

Here are some important related research works that we were able to pull from the World Wide Web. These research articles helped us to understand the current scenario in the music industry and assisted us in conducting our analysis.

1. Women in Audio: Contributions and Challenges in Music Technology and Production-- Marlene Mathew, Jennifer Grossman, and Areti Andreopoulou- https://www.researchgate.net/publication/312116630_Women_in_Audio_Contributions_and_Challenges_in_Music_Technology_and_Production
2. Who gets to play the electronic music- Carolina Lindén- http://www.diva-portal.org/smash/get/diva2:1092021/FULLTEXT01.pdf
3. Gender differences in computer- and instrumental-based musical composition- Shibazaki, K. and Marshall, N.A.- https://www.researchgate.net/publication/271946349_Gender_differences_in_computer-_and_instrumental-based_musical_composition
4. Gender Differences in the Global Music Industry: Evidence from MusicBrainz and The Echo Nest- Yixue Wang, Emőke-Ágnes Horvát-- https://www.aaai.org/ojs/index.php/ICWSM/article/view/3249
5. Where Is She? Finding the Women in Electronic Music Culture- Freida Abtan- https://www.tandfonline.com/doi/full/10.1080/07494467.2016.1176764
6. Payne, Barbara. "The Gender Gap: Women on Music Faculties in American Colleges and Universities 1993-1994." *College Music Symposium*, vol. 36, 1996, pp. 91–102. *JSTOR*,
7. Pink Noises: Women on Electronic Music and Sound- Tara Rodgers-DOI: https://doi.org/10.1215/9780822394150-ISBN (print): 978-0-8223-4661-6-ISBN (electronic): 978-0-8223-9415-0-Publisher: Duke University Press-Published: 2010

13. Acknowledgement

We would like to express our sincere gratitude to Mr. Mohsen Jadidi from Gesis for his consistent support and mentorship throughout this project development. Mr. Jadidi has provided us some valuable datasets from his research works. In this project, we have explicitly and extensively used his two data sets- `songkick_all_gigs_extended_version` and `releases_merged`. These data sets helped us to process, manipulate and prepare our final datasets and its subsequent analysis.

14. References

1. Lecture slides of Computational Social Science by J.Prof Claudia Wagner, Universität Koblenz-Landau

2. Bit By Bit: Social Research in the Digital Age- Matthew J. Salganik, Princeton University Press

3. Computational Social Science: Discovery and Prediction- R Michael Alvarez, Cambridge University Press
4. Generative Social Science: Studies in Agent-Based Computational Modeling (Princeton Studies in Complexity)- Joshua M. Epstein, Princeton University Press
5. Practical Statistics for Data Scientists- Peter Bruce, O'Reilly
6. Think Stats- Allen B. Downey, Green Tea Press
7. Gender Differences in the Global Music Industry: Evidence from MusicBrainz and The Echo Nest- Yixue Wang, Emőke-Ágnes Horvát-- https://www.aaai.org/ojs/index.php/ICWSM/article/view/3249

15. Team Work Distribution

Code and Programming Work Distribution-

Data Crawling	Basitur Rahman Chowdhury
Data Processing & Preparation	Md Tariqul Islam
Data Analysis	Muhammad Momotazul Islam

Git Repository- https://github.com/tariquldipu/CSS_GenderGap
Binder- https://mybinder.org/v2/gh/tariquldipu/CSS_GenderGap/master

Report Preparation
All three of us jointly worked and contributed on this report preparation. However, here we have listed the names of the persons who was mainly responsible for each specific chapter writing and text preparation.

Author Name	Chapter Number & Name
Basitur Rahman Chowdhury	2. Introduction 5. Data Crawling & Preprocessing
Md Tariqul Islam	4. Methodology 6. Data Cleansing 10. Limitations
Muhammad Momotazul Islam	7. Data Visualization 8. Analysis: Revealing the Insights 9. Results
All three of us	1. Abstract 3. Research Questions 11. Challenges for Afterward Works 12. Related Work 13. Acknowledgement 14. References