An Exploratory Analysis of Online Adult Content Sharing In Pakistan Through Social Media: A Case Study (Jan-2021 To Oct-2021)

By

Babar Ali

(Spring19 - MS CS&E - 00000281861)

Supervisor

Dr. Zamir Hussain

A thesis submitted in conformity with the requirements for
the degree of *Master of Science* in
Computational Science and Engineering
School of Interdisciplinary Engineering & Sciences (SINES)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
January 2023

An Exploratory Analysis of Online Adult Content Sharing In Pakistan Through Social Media: A Case Study (Jan-2021 To Oct-2021)

By

Babar Ali
00000281861

Supervisor

Dr. Zamir Hussain

A thesis submitted in conformity with the requirements for
the degree of *Master of Science* in
Computational Science and Engineering
School of Interdisciplinary Engineering & Sciences (SINES)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
January 2023

DEDICATION

I dedicate this dissertation to my Family.
For their endless love, support, and encouragement

Declaration

I, *Babar Ali* declare that this thesis titled "An Exploratory Analysis of Online Adult Content Sharing In Pakistan Through Social Media: A Case Study (Jan-2021 To Oct-2021)" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
5. I have acknowledged all main sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

_____

Babar Ali,
00000281861

Copyright Notice

Copyright in the text of this thesis rests with the student author. Copies (by any process) either in full or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of SINES, NUST. Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.

• The ownership of any intellectual property rights which may be described in this thesis is vested in SINES, NUST, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of SINES, which will prescribe the terms and conditions of any such agreement.

•Further information on the conditions under which disclosures and exploitation may take place is available from the Library of SINES, NUST, Islamabad

Acknowledgments

I would first like to thank my thesis supervisor Dr. Zamir Hussain of the School of Interdisciplinary Engineering & Sciences, NUST. The door to his office was always open whenever I was stuck in a problem related to my research. He always guided me in the right direction whenever he thought I needed it. I would also like to pay special thanks to Principal SINES Dr. Hammad Mehmood Cheema. He remained a source of motivation for this project with his timely support. I would also like to thank my other GEC members including Dr. Shahzad Rasool and Dr. Ammar Mushtaq for their support and help throughout the research phase.

I would also like to thank the Management of SINES, NUST was very supportive and was there to assist for arranging presentations and demos of this research work. During the research phase, my fellow researchers of Data Analyst Group (DAG) also helped me during research phase.

Table of Contents

List of Abbreviations

ID	Identity
ESP	Electronic service Provider
NCMEC	National Centre for Missing & Exploited
FIA	Federal investigation agency
COPINE	Combating Pedophile Information Networks in Europe Center
CSAM	Child sexual abuse material
IC	Industrial classification
FR/FRV	File relevance/File relevance value
IP/ ISP	Internet protocol / Internet service provider
Mb	Mega bits
PTCL	Pakistan telecommunication authority
KP	Khyber-Pakhtunkhwa
GB	Gilgit Baltistan
SD	Sindh
BA	Baluchistan
PB	Punjab
IS	Islamabad
AJK	Azad Jammu & Kashmir
Fo	Observed Count
Fe	Expected Count
StDev	Standard Deviation

List of Tables

List of Figures

Abstract

Sharing adult content online has always been difficult to regulate for the global community. An increase in criminal activity has coincided with the adoption and use of information and communication technologies (ICTs). Recent web search pattern shows a sharp rise in demand for explicit / adult content. According to estimates, up to 90% or more of people between the ages of 12 and 18 have access to the Internet. Greater accessibility to internet results in a rise in young people seeking out porn and the exposures to pedophiles increase. This then has a detrimental effect on their mating development. Therefore, there is a need to analyze the adult content sharing through social media in Pakistan. An exploratory analysis in Pakistan can provide useful information for combating the increasing rate in the child abuse and exploitations. It can also be helpful in identifying the pedophile/offender of the crime. The data regarding sharing of adult content material is gathered from NCMEC, an international nonprofit organization working with the law enforcement agencies to end this crime. When a user shares any adult content using social media an automated Tipline report is generated by the respective platform and sent to the concerned country. 55740 Tiplines of different social media platforms which are reported from various regions of Pakistan are used in this study. Tipline reports are very detailed and can go up to several pages. NCMEC therefore provides the data in excel format too. Originally the data consisted of 17 variables, out of which only 8 are used. The variables used are 1-Report ID, 2-Electronic Service Provider (social media platforms e.g., Facebook), 3-Industrial Classification (type of data shared), 4-File Relevance Value (frequency of reported ID), 5-City, 6-Region, 7-IP/ISP (internet providers e.g. Zong, Telenor etc.), 8-File size (Mb). The remaining variables such as Mobile number, IP address etc., were omitted due to privacy concerns. Analysis has shown that 82% of people in each region (province) are sharing lascivious exhibition content, 90% of which is through Facebook. The average file size (mb) of adult content shared in all the regions is about the same. 79% of

user IDs in the Tipline are reported only a couple of times. While the frequency of user IDs reported between 11-335 times are 0.55%. Highest reports are from Karachi (28.07%), Lahore (23.75%) and Islamabad (15.78%), while Zong (26.11) and Jazz (25.01) are the most used networks for it. There is also no correlation found between the amount of data shared and the number of times it is reported.

Chapter 1

Introduction

Social networking sites have attracted millions of users to use their services to link, reply, follow, like, comment and message each other by forming a certain electronic connection. Apart from genuine legitimate users, social media also includes cyber-criminals, who exploit the networks by spreading abusive / explicit / adult content. Adult content sharing and its commercialization is on a rise in cyberspace, which compels all nations to enact strong policies and work together worldwide to combat this problem. The nature of adult content has drastically changed over the past 20 years in terms of production, dissemination, and possession of photos. The reason behind sharing abusive content varies from user to user and there hasn't been much empirical research done on this subject. Therefore, laying a foundation for future research and the creation of programmes and policies that attend to the needs of victims is important. The current study's objective is to examine participants who shared or posted a media that was meant to be private.

1.1 What is adult content?

Any medium currently known or subsequently developed that incorporates scenes of a graphic abusive nature is known as adult content. Adult content may include movies, videos, software, pictures, images, sound recordings, audiovisual works, video games, websites, or any other form of media, that depicts abusive behaviors in any way that goes above and beyond to what is typically aired. As per the definition of Oxford Reference, adult content is stated to be explicit depiction, in writing, pictures, or other material, of sexual activity or subjects in a manner intended to arouse. It doesn’t include material intended for other artistic, aesthetic, or

educational purposes [1]. Similarly, the content produced by victimizing children also comes under adult content. The offenders of such crimes are called pedophile.

1.2 Why share adult content?

There are different reasons for sharing adult content. Some users share adult content purposely on demand of their lover. While some share to take revenge from an ex-lover, for hurting their reputation in the public. Most social media users share content for entertainment purposes only, however there are some spammers and scammers who share data for the sake of earning from it. Some take it as business to create adult content and share it publicly to gain the attention of the people. While some use adult content advertisement to generate traffic towards their websites to generate extra revenue.

When it comes to children, pedophiles not only create child abusive material for business purpose, but they abuse children for their own satisfaction too. They then share it on dark web and other platforms where they can find people of their own type. Children generally are an easy target due to their innocence and lack of exposure.

Cyber Sextortion: Sextortion, a combination of the words “se*ual” and “extortion” is the threat to disseminate explicit abusive content if a victim does not comply with demands. In conventional extortion, criminals threaten to leak private information to intimidate their victims. This technique is used by offenders to control and coerce their victims' behavior. Furthermore, technology has globalized cyber sextortion, facilitating greater access to victims and increased anonymity for offenders. This has created substantial investigative challenges for law enforcement, such as the need for multiagency cooperation, specialized cybercrime units, and advanced technological resources.

Revenge explicit content victimization: There has been an increase in conversation over sharing or publishing adultly suggestive or explicit images of a former partner without their

permission to humiliate, harass, or punish the victim. This is often referred to as revenge porn. An act whereby the perpetrator satisfies his anger and frustration for a broken relationship through publicizing false, adultly provocative portrayal of his / her victim to publicly defame them [2]

1.3 Why social media?

Social media is the easiest way to share or publish any documents, video, picture, and story regardless of the nature of content available in the shared material. At the time of uploading the material, social media platforms are unable to identify its nature. This is the reason why adult content is shared through social media at a very high rate. Social media is the biggest internet industry that has an overflow of shared adult content. It is not surprising that the adult industry including abusive material and explicit spam is generating its own traffic through social media.

There are 4.59 billion social media users in the world, which is 58% of the total population. Whereas in Pakistan, there were 46.0 million social media users as of January 2021 which is 20.6% of total population [3], and by January 2022 these increased to 71.70 million, which is 31.5% of the total population. The increase in one year is 25.7 million social media users. These include, 43.55 million Facebook users (19.2% of total users), 13.75 million Instagram users (6% of total users), 18.80 million Snapchat users (8.3% of total users) etc. [4]. This rapid increase in the number of users has also increased the cybercrimes in Pakistan.

1.4 Pakistan combating adult content

If someone has shared a sexually suggestive or explicit media that is intended to be private, not only does it disrupt the life of the victim in the present but continues to do so forever. Since the images shared continues to reappear on the internet at multiple places. Many sites now allow

the victims to request the images to be taken down, but once the picture has been posted the damage is already done [5].

As mentioned earlier, adult content shared through social media is done for earning purpose by spammers and scammers. However, there are some serious offenders as well who victimize people to create abusive content. This heinous crime is nonbailable and highly punishable in Pakistan PECA-2016. Child abuse and child trafficking is also highly condemned in Pakistan, but despite the highest punishment as per the Cyber Law-2016, it is still growing [6].

Child pornography is a grave threat to the future and morality of children in the country. In 2016 for the first time, Pakistan had criminalized child pornography, making the offence punishable with 7 years in prison and a fine of Rs700,000 [7]. This law was made stricter by increasing the punishment to 14 years to 20 years in prison and a fine of 1 million Rupees [8]. This initiative came after a major pedophilia scandal in August 2015 rocked the country. It was revealed that hundreds of child pornographic videos from Hussain Khanwala village in the Punjab province had been made and were being circulated online [9] .

1.4.1 Laws on adult content

To address the prevention of cybercrimes and unlawful acts with respect to Pakistan's information system, the Prevention of Electronic Crimes Act 2016 (PECA 2016), was passed. It is applicable to every Pakistani citizen as well as the foreigner. It also applies to any conduct carried outside of Pakistan, which violates this Act and has an impact on a Pakistani person, piece of property, information system, or data. Two PECA 2016 acts related to our research are attached in the Appendix I.

Despite such harsh punishment there are many people who still circulate adult content through social media. It not only harms the reputation of the Pakistan internationally, but it is also a

shameful act as a human being. Pakistan is an underdeveloped country; it does not have resources to deal with such issues itself. It also does not have MOUs signed with social media sites like Facebook and Instagram for regular provision of data. Therefore, data is collected from external sources to deal with such issues. In this study, the data is provided by a private non-profit organization NCMEC (National center for Missing and exploited children).

1.5 Data and Source

The National Centre for Missing & Exploited Children (NCMEC) was established in 1984 as a private, non-profit organization by John and Revé Walsh and other child activists to act as the country's information hub for missing and exploited children. The goal of NCMEC is to assist in the finding of missing children, decrease child abuse and exploitation, and stop child victimization. When any type of content is reported on the website of NCMEC a Tipline is generated (attached in Appendix II). In this study, a total of 55470 abusive/explicit content tiplines are used. The data is a closed source, and only the registered FIA officials have access to it. This Tipline includes some basic information like the culprit's name, material shared, IP, city, region, country, Gmail. mobile number etc. (discussed in detail in Section 3.2).

1.6 Motivation of the study

As part of the Federal Investigation Agency, cybercrime wing, it has been observed that the adult content in the form of images or videos is immensely prevalent on social media (Facebook Instagram twitter etc.) despite of their policies classifying such content as spam. Adult content is different from other spam messages as its content is very specific and revolves around explicit material. Social media keeps changing its policies regarding abusive content. On March 11, 2015, twitter has changed its abuse policy (Update in Twitter rules to prevent revenge porn 2015) to specifically forbid the publishing of intimate images or videos that are

taken or transmitted without the user's consent. However, despite the publication of these standards and restrictions, adult content is still a thriving component on social media. The goal of this research is to explore users who share adult and abusive content on social media by analyzing their activities and frequency of their ID being reported.

1.7 Research Question

According to the data revealed by Google, Pakistan is on top of the list for searching porn videos [10]. On daily basis, explicit data of people is being shared in Pakistan. So far, there is no scientific and empirical research to prove its adequacy, therefore, this study is designed to explore and analyze the online adult content sharing in Pakistan through social media. Detailed data which is provided by an authentic source is used for the analysis. It provides useful information regarding the user who shares the data. Analysis will be done on these sub-questions:

1. Adult content sharing is increasing in Pakistan. There is no statistical analysis to examine or take measures for it. Therefore, its general trend and tendencies are yet to be explored.

1.8 Objectives:

Based on the above stated questions, the main objectives are:

1. To analyze general trends and tendencies with respect to different available variables i.e., region, city, ESP, amount of data shared etc.,
2. Analyzing associations between multiple characteristics with respect to files size and available qualitative variables and testing the significant differences in categorical variables.
3. Recommendations to deal with issue of adult content sharing via social media.

1.9 Limitations of the study

The proposed study will proceed with the following limitations:

1. The study results would be limited to the available data and factors only from Jan 2021 to Oct 2021.
2. This study will only focus on statistically analyzing and exploring the variables rather than modeling the information for prediction purpose.
3. The data used in this study is a closed source. Its accessibility is difficult.

1.10 Structure of thesis

The rest of this thesis is organized as follows. In chapter 2, there is literature review concerning the related studies worldwide and a link of empirical verification of the work. Chapter 3 discusses the applied methods, both conceptual and theoretical, used to analyze and explore. Chapter 4 discusses the findings of the data which are concluded in Chapter 05.

Chapter 2

Literature Review

Technology has played a predominant role in the circulating of adult content. Different policies and regulations have been devised to control its misuse. This research builds on the existing literature of the experiences of social media users and the techniques applied by the researchers to analyze it.

2.1 Different Categories of Adult content pictures

10 categories of pictures are identified by Combating Pedophile Information Networks in Europe Center (COPINE), Taylor and Quayle [11] that may be sexualized by an adult. COPINE scale is shown in **Table 1**. To define the extent to which someone is caught red-handed with abusive / explicit imagery, the COPINE scale is pivotal.

Some appropriate legislation on the theme can also be driven using the COPINE studies. Researchers have been working on detection techniques that center on the users. The COPINE method is effective in detecting suspicious content actively. However, it raises serious questions regarding high false positive rates and censorship, oftentimes blocking otherwise legitimate content. [12].

Table 1 COPINE scale for adultery exploitative imagery of children [20]

S. No	Level	Description
1	L1	Indicative (non-erotic pictures);
2	L2	Nudist (naked or semi-naked in legitimate settings);
3	L3	Erotica (secretive photographs showing underwear/nakedness);
4	L4	Posing (intentional posing suggesting sexual content);
5	L5	Erotic Posing (intentional sexual or provocative poses);
6	L6	Explicit Erotic Posing (emphasis on genital areas);
7	L7	Explicit abusive Activity (explicit activity with no adult involved);
8	L8	Assault (mating assault involving adult);
9	L9	Gross Assault (penetrative assault involving adult);
10	L10	Sadistic/Bestiality (imagery involving pain or animal).

2.2 Sharing Sexually Explicit Photos or "Sexting"

Today, majority of people own a handheld device with a camera that makes it incredibly simple to share digital pictures. According to estimates, more than 75% of teenagers have a cell phone, and more than 25% have a smartphone with Internet access [13]. The proliferation of cell phones and smartphones togethers with access to the Internet has contributed to the development of "sexting". This is commonly defined as sending explicit photos, videos, or messages through electronic means, like cell phones, email, or posts on social media sites (e.g., Facebook, SnapChat, Instagram, etc.) [14]. According to research, many young adults in the United States share photos of themselves or others in their undies or semi-undies. Estimates on the number of teens and young adults engaging in sexting vary, ranging anywhere from a low of 5% to a high of 40%. Overall, women and girls have been shown to engage in more sexting behaviors than men and boys [24].

In one of the first studies conducted on teen sexting by the National Campaign to Prevent Teen and Unplanned Pregnancy (2008), it was discovered that nearly 18% of teen boys felt pressure from girls to send these kinds of photos. Over half of teen girls felt that they send adultery suggestive pictures and messages because of pressure from a guy. A large majority of

adolescent boys and girls claimed that they sent sexually explicit content to be "fun" or "flirtatious". Additionally, about 44% of male and female teenagers send adultly explicit messages or photographs in reaction to adultly explicit information that they received. While about 50% of teen girls claimed that they send explicit messages as a "romantic present" for their partner [15].

Similar findings are observed in more recent research. Over half of the students participating in the study admitted to sending sexually explicit photos to appear "flirtatious" and 17% admitted to doing so under pressure from a lover [16]. Sending and receiving abusive/ explicit photographs may be a way for men and women to "do gender" given the pressure and motivation noted by teens and young adults.

2.3 Impact of Sharing Explicit Images

Many people who decide to post an explicit graphic photo, do so in the hope that their information won't be made public. Recent studies, however, suggest that this might not always be the case. According to research conducted in 2012, 25% of those who received text messages transmitted the content to other people. In the study, only 50% of women and 33% of men agreed that it was "always improper" to share a private explicit photograph with another person [17].

Sharing explicit images to harass, control, manipulate or habitually disparage an individual is considered as cyber bullying and cyber harassment. These terms are often used interchangeably because of the difficulty in making a clear distinction between what behaviors should be labeled harassment and what behaviors should be labeled stalking/bullying [13]. It has been observed that student victims of cyberbullying had higher rates of academic issues, substance usage, and school-related issues than their classmates who had not experienced cyberbullying.

Abusive photographs and remarks keep resurfacing online, due to which the impacts of cyberbullying and cyber harassment can be worse than those of physical harassment. If the victim's identifying information is made public, cyberbullying may also raise the possibility of in-person victimization or other potentially violent behavior like stalking, abusive harassment, or abusive assault. The impact of a former partner posting abusive explicit images of a former lover without that person's consent has hardly ever been the subject of empirical investigation.

2.4 Doing Gender

A complex set of socially regulated perceptional, interpersonal, and micropolitical actions known as "doing gender" depict behaviors as manifestations of male and feminine "natures". Doing gender more precisely refers to societally induced disparities between women and men and between girls and boys that are not essential, biologically necessary, or natural. These socially manufactured disparities are then utilized to support the need for gender in society. Sharing images that are abusive explicit has also evolved into a component of "doing gender." The relationship between nudity and femininity is strengthened by the normalizing of nudity in images of well-known female celebrities on magazine covers or on Instagram posts [14]. While men are taught to do gender through the consumption of such material, women are taught gender through the practice of making their bodies visible in sexually explicit ways. Sharing graphic images of women could also be a method for guys to "do gender" and show that they have mating control over their bodies [14].

2.5 Prevalence of intentional Internet exposures

Teenagers' deliberate exposure to explicit content online has been questioned in several studies. A Midwest private urban high school's survey of students found that 21% had spent at least three minutes on a pornographic website [18]. 15% of youth between the ages of 12 and 17

who participated in a nationwide telephone poll admitted to lying about their age to access websites. It's interesting to note that this percentage is comparable to the equivalent figure found for adults in the same poll. Males, older children, people with more months or years of internet experience, and people who use the internet more frequently (i.e., for more than five hours per day) are more likely to say they deliberately seek out these websites [18].

2.6 Adolescent beliefs about Internet exposures to se*ual material

National research on older teenagers sheds some light on the relationship between the exposure to adult content and ensuing behavioral changes. Most respondents (59%) think that viewing internet pornography might tempt young people to engage in abusive activity early. Nearly half (49%) of the respondents said that unfavorable attitudes about women are encouraged by internet pornography, and a similar number (49%) said that the images encourage the idea that abusive activity is "acceptable". Beyond perception and belief, nothing is known about the results or connections between intentional internet exposure to porn and developmental or psychosocial difficulties [19].

2.7 Gaps in current literature

An important question remains unanswered despite the enormous volume of literature about adult purposeful exposure to pornography. That what are the factors that go beyond demographics and what kind of young people are attracted to porn. With the introduction of the Internet, it is also critical to start identifying the distinctive traits of children and adolescents who look for online adult content and determine whether these features set them out significantly from other young people. This study will present a statistical analysis for significant cross-sectional associations required to support and guide future, more intricate longitudinal studies.

Chapter 3

Methodology

In this section, we are going to discuss the methodology used in this research to achieve the objectives stated in chapter 1. The main objective of this research is to find the relationship between different variables by using correlation analysis and investigate the other facts about adult content sharing. For this purpose, A secondary dataset containing 55741 entries with 8 variables is analyzed. The methodology of this research study is discussed in detail in the upcoming subsections.

3.1 Data collection

3.1.1 A brief about data source

In 1984, National Centre for Missing & Exploited Children (NCMEC) was established as a non-profit, private organization by Revé and John Walsh and other child activists to act as the country's information hub for missing and exploited children. The goal of NCMEC is to help find the missing children, stop child victimization, and decrease child exploitation and abuse. NCMEC collaborates with law enforcement, victims, families, and the public to locate missing children, help prevent kidnappings, and offer services to combat child exploitation.

NCMEC's CyberTipline offers the Public and online ESP an easy way to quickly report suspected incidents of exploitation online. Since CyberTipline's inception in 1998, millions of reports have been received and hundreds of millions of videos and images of suspected Child sexually abuse material (CSAM) have been reviewed. They work to disrupt the trading of child se*ual abuse images and videos online and help survivors begin to rebuild their lives.

A variety of things can be indicative of higher numbers of reports, including platform has large number of users or how robustly an abusive content is identified and removed from their platforms. CyberTipline reports provided by ESPs in 2021, included 44.8 million videos, of which 5.1 million are unique and 39.9 million images, of which 16.9 million are unique.

NCMEC often receives CyberTipline reports about imagery of child abuse exploitation that is being shared online. In some cases, the reports are made by the child victims themselves or their guardians/caregivers. The CyberTipline is a lifeline for families who are struggling to have explicit images of their teenage child taken down. After visually reviewing the reported imagery, NCMEC staff members notify the electronic service provider where the image or video has been shared. Based on how a company or their platform operates, content may be removed, or users blocked or banned from their services in response to a notification. Once a notice has been sent to the company, NCMEC staff manually track the response and will generate additional notices until they can confirm that the content is addressed

3.1.2 CyberTipline Files

Reports made to CyberTipline by ESPs may include videos, images and other files related to suspected explicitly / abusive adult material. This content is analyzed along with the report. Unfortunately, explicitly / abusive material and videos are often shared and circulated online repeatedly. Even after years of initial abuse, CSAM of a single child victim can keep circulating. One of CyberTipline's critical functions is to identify unique images by using technology and the work of analysts.

3.1.3 File Review

NCMEC makes CyberTipline reports, available to law enforcement around the world, including NCMEC's additional analysis. Given the incredible volume of reports, the goal is to

help law enforcement prioritize the most urgent cases. NCMEC successfully tagged over 30 million files to date and over 22 million files in 2021 alone. It uses this information to ensure that the most urgent cases, those where a child may be suffering ongoing abuse, get immediate attention.

To report instances of suspected abusive material, NCMEC operates CyberTipline, a national mechanism for the public and electronic service providers. CyberTipline has received more than 116 million reports since its inception in 1998.

These include: 29,309,106 (CSAM), 44,155 Online enticements, including "sextortion", 16,032 Child abuse trafficking, 12,458 Child abuse molestation, 15,930 Other, including child abuse tourism, misleading domain name, misleading words or images and unsolicited obscene material sent to a child.

3.1.4 Hash tagging

Another benefit of the recent enhancements to CyberTipline process is the ability to grow the list of hash values which is part of an important initiative to stop the spread of CSAM. Hash values are unique digital fingerprints applied to data like videos and images. When an image or video is identified as containing known CSAM, NCMEC adds the hash value to a list that is shared with technology companies.

The companies can then use these hash values to identify, report and remove abusive content by voluntary scanning of their system. Likewise, NCMEC can instantly identify whether a picture of child abuse or sexual content has already been reported and whether the child in the image has been recognized when it gets a report with a known hash value.

NCMEC assisted with more than 22,170 requests in 2021 to help locate noncompliant offenders, 4037 out of those noncompliant offenders were subsequently located/arrested as per the feedback [20].

NCMEC compliments the ESPs that makes reporting and identifying of this content a priority and encourages all other companies to do the same. These reports are helpful in stopping further victimization of children and remove them from harmful situations [21].

There has been an increased focus over the past several decades, on various forms of adultly aggressive behaviors and abusive violence (e.g., harassment, stalking). New opportunities have arisen due to technology for adultly aggressive behavior and involvement in revenge porn and sexting. Information technology MOU is signed with the different social media / Chatting Companies / Organization who provide them data regarding adult content sharing. Adult content at any place of the world will get reported through the corporative partners if data is transmitted using social media platform [22].

3.2 Data Description

FIA is in collaboration with the NCMEC to combat the abusive material shared through social media in Pakistan. NCMEC has provided access to the data to registered officers of FIA only. It provides data to FIA on daily basis regarding the abusive material being shared in Pakistan. Data provided by the NCMEC contains the necessary information regarding offenders. You can access or analyze one report at a time. One Tipline report may have pages from one to many, depending upon the activities of the offender. A tipline report is attached in the annexure.

When anyone shares adultery / abusive content through social media, a report called Tipline is generated, which contains numerous information regarding the alleged activity. Tipline can be generated in two ways, one is made by the public, the second is made by ESP (Electronic

Services Provider) / Social Media Platforms. It is an automatically generated report whenever any ID of the ESP is reported. It goes into the system and is analyzed against its shared media and hence Tipline is generated with the necessary information of the offender like name, material shared, IP, City etc. Fields of the Tipline that have been used in this research are discussed below.

It can be seen in the report that very detailed information is shared by NCMEC for a single report. It is very difficult to read every report and find out the offender who is sharing maximum data among them or to look for the habitual offender among them. There is another way to access the NCMEC data called Triage. In a triage, data can be accessed in an excel format, which contains 50 different Tiplines in a single sheet. In excel sheet data is described in rows and columns. Columns contain the attributes, while rows define the record. Originally, the excel sheet contained the following attributes in the column as shown in **Table 2**

Table 2 Attributes in Tipline Report

Sr. No	Variables
1	Report ID
2	ESP Name
3	IP Address
4	Industry Classification
5	IP / ISP
6	Country
7	Assigned Date
8	Report Date
9	Postal Code.
10	Email
11	Phone Number
12	Type of Files
13	Region
14	City
15	Screen Name
16	Suspect
17	File Relevance

Few attributes have been excluded due to privacy purposes such as mobile number, Gmail and IP address etc. While few are not required in this case study, such as assigned date, screen name, and suspect postal code. Country has been excluded too since our case study is based on the data of Pakistan only from Jan-2021 to Oct-2021. The attributes used for this exploratory analysis are Report ID, ESP, Industry Classification, File Relevance, City, Region, IP ISP and Total File Size - (Mb).

3.2.1 Report ID:

Report ID is automatically generated by tipline when adult content is shared. It is unique and numeric in nature.

3.2.2 ESP:

Electronic Service Provider is the medium which is used to share adult content. Our data contains tiplines from Facebook, Instagram, WhatsApp, Google, Snap Chat, Microsoft Online Operation, and Drop Box.

3.2.3 Industry Classification:

Industry classification includes the extent and type of data which is shared on the ESP. These are divided into 4 categories by NCMEC. The 4 classifications are, Prepubescent Minor A1, Prepubescent Minor A2, Pubescent Minor B1, Pubescent Minor B2. Low ranking content (A1& B1) is termed as Abuse Act. While High ranking content (A2 & B2) is termed as Lascivious Exhibition. The type of content in these 4 categories are mentioned in **Table 3**.

Table 3 Description of Industrial Classification

Name	Description
Prepubescent Minor A1 Pubescent Minor B1	Any image of sexually explicit conduct (simulated or actual sexual intercourse including anal-genital, genital-genital, or oral-anal between same or opposite sex person), sadistic or masochistic abuse, bestially, degradation, masturbation, or any such depiction that lacks serious political, literary, scientific or artistic value.
Prepubescent Minor A2 Pubescent Minor B2	Any image that depicts nudity and one or more of: inappropriate touching, sexually suggested poses, adult arousal, focus on genital, spreading of limbs or genitals, and such depiction lacks serious political, literary, scientific or artistic value.

3.2.4 File Relevance Value

This is the number of times an ID is reported for adult content sharing. These range in between 1 to 335. We classified them into 11 categories. From 1-10 there are 10 categories, and the rest are combined into 11^{th} category (File Relevance 10+). This is done for convenience of handling the data since 99% of the file relevance values are lower than 10.

3.2.5 City

Name of the city from which the content is shared from. Tiplines in the data are from 240+ cities of Pakistan. For convenience we selected to represent the data from top 8 cities which are Karachi, Lahore, Islamabad, Rawalpindi, Peshawar, Multan, Faisalabad, Quetta, while the remaining cities are combined into ‘other cities’.

3.2.6 Region

Name of the region / province from which the content is shared from. Tiplines in the data are from 7 regions of Pakistan. These included Punjab, Sindh, Baluchistan, Khyber Pakhtunkhwa, Gilgit Baltistan, Islamabad, and Azad Jammu & Kashmir.

3.2.7 IP / ISP

Internet Protocol / Internet Service Provider are the names of the Internet services Provider using which adult content is shared on internet. There are 220+ Internet Service Providers in the data. We choose to represent the data from top 6 ISP's which are CM Pak Limited (Zong), Mobilink Mobile internet (Jazz), Telenor Pakistan, PTCL, Ufone, and Cybernet, while the remaining ISPs are combined into Other ISPs.

3.2.8 Total File Size – (Mb)

This indicates the total amount of data shared through a Tipline in Mbs. The data ranges from 1Mb to 121Mbs.

3.3 Data Pre-processing

3.3.1 Data cleaning & coding

It is beneficial to preprocess the data for getting authentic results and digging into the analysis [23]. Data is first analyzed for any missing values and outliers. Before analysis, the minimum and maximum of all the variables are checked to make sure that all data values are in the desired range. As discussed earlier, the data is downloaded in the form of rows and columns in excel sheet from the NCMEC website. The downloaded data contained strings, digits, discrete values, etc. in different formats. These are all brought to the same format for convenience in

coding. The missing values are also dealt with. The data is coded at SPSS in order to apply different tests and analysis.

3.3.2 Variable transformation

The names of a few attributes are changed for ease of understanding. As discussed earlier in Table 3, industrial classifications contained 4 types of data being shared. However, some data was a combination of these 4 categories, so 3 new categories are created which contained data from multiple categories as shown in **Table 4**.

Table 4 Industrial classification Names

S. No	New Category	Old Category
1	Pre-Minor Low	A1
2	Pre-Minor High	A2
3	Pub Minor Low	B1
4	Pub Minor High	B2
5	Pre High/Low Comb	A1 + A2
6	Pub High/low Comb	B1 + B2
7	Pre /Pub Comb	A1 + A2 + B1 + B2

File relevance value contained a string and numeric value. It was in the format REPORTED - 2, REPORTED - 3 etc. The string 'REPORTED' is removed and only the number is used in the analysis. In the field of cities, there are more than 220 cities. For convenience, we considered the top 8 cities only, and accumulated the rest of them as 'Other Cities'. This is done because most cities have very less frequency distribution, even lower than 1%. Same is done with the field of IP / ISP. In the field of the IP/ISP there are around 300 IP/ISP. We

considered only the top six IP/ISP used in Pakistan; the rest are clustered as one variable because their frequency is very less.

3.4 Scheme of study

Statistics are generally classified into two types: descriptive and inferential statistics. Inferential statistics are used to see the trends and tendencies of the data, while descriptive statistics are used to identify data's fundamental characteristics. Both descriptive and inferential statistics are used in this study. Their results are discussed in Chapter 4.

3.4.1 Descriptive statistics

It provides a quick summary of the sample and the metrics along with simple graphical analysis. The three measures used in this study are as follows.

- **Location**

To find the location of the data, mean, median and mode is used.

- **Dispersion**

To find the dispersion of the data, variance and standard deviation is used

- **Shape**

To find the shape of the data, skewness and kurtosis is used.

3.4.2 Inferential statistics

It infers properties of an underlying distribution of probability by driving estimates and testing hypothesis. The three measures used in this study are as follows

- **Chi Square**

The discrepancy between the observed and predicted frequencies of the results of a group of events or variables is measured by the chi-square (x^2) statistic. Such categorical differences can

be examined using chi-square, especially for qualitative variables**.** To find the association between variables, a formulation of hypothesis is performed.

1. State your null and alternate hypothesis

Null hypothesis	Two attributes are independent
Alternative hypothesis	Two attributes are associated
Equal variances were assumed for the analysis.	

2. Significance level (α = 0.05)

3. Calculate the test statistic

$$x^2 = \sum \left(\frac{(f_0 - f_e)^2}{f_e} \right)$$

f_0 = *Observed frequency*

f_e = Expected frequency

4. Calculate probability value (p-value)

5. Conclusion: Accept or reject the null hypothesis based on the comparison of p-value with the stated level of significance

- **Point Biserial Correlation**

The strength and direction of the association between one quantitative variable and one qualitative variable are assessed using a point-biserial correlation.

$$r_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{pq}$$

M_1 = mean of the group that received the positive binary variable
M_0 = mean of the group that received the negative binary variable.
S_n = standard deviation for the entire test.
p = Proportion of cases in the "0" group.
q = Proportion of cases in the "1" group.

- **Analysis of Variance (ANOVA)**

Analysis of Variance (ANOVA) is procedure used to test the equality of means for more than two population. The steps of ANOVA are as follows

1. Formulate a hypothesis

Null hypothesis	All population means are equal
Alternative hypothesis	All population means are not equal
Equal variances were assumed for the analysis.	

2. Set a significance level (α = 0.05),

3. Compute F-Statistic

$$f = \frac{Larger\ sample\ variance}{Smaller\ sample\ vaiance} = \frac{\sigma_1^2}{\sigma_2^2}$$

4.Use the F-Statistic to derive p-value

5. Conclusion: Accept or reject the null hypothesis based on the comparison of p-value with the stated level of significance

Chapter 4

Results and Discussion

4.1 Uni variate analysis of data

An important aspect of this exploratory analysis research is the introduction of certain variables in the Tipline report to understand the information with respect to its characteristics. For this study, eight variables are used namely, Report ID, ESP Name, Industry Classification-Category, File Relevance Value, City, Region, IP ISP, Total File Size – Mb. Details of these are already discussed in the above section. Frequency distributions of these variables are illustrated in **Figure 1** to **Figure 6**.

4.1.1 Frequency distribution of electronic service provider (ESP)

Figure 1 illustrates the percentages of ESPs used by the users while sharing adult content. Results show that Facebook is on top, with 95% of Tiplines generated. The reason is that the number of Facebook users in Pakistan are greater as compared to any other social media. Moreover, it is also easy to use and create new/fake profiles using the same physical and virtual address compared to other social media platforms. Instagram is in the second position with 2.68%, and WhatsApp is at third with 1.05%. The percentage associated to remaining platforms are almost negligible with lower than 1%, which includes, Google 0.55%, Snap Chat 0.08%, Microsoft Online Operation 0.02%, and Drop Box 0.01%.

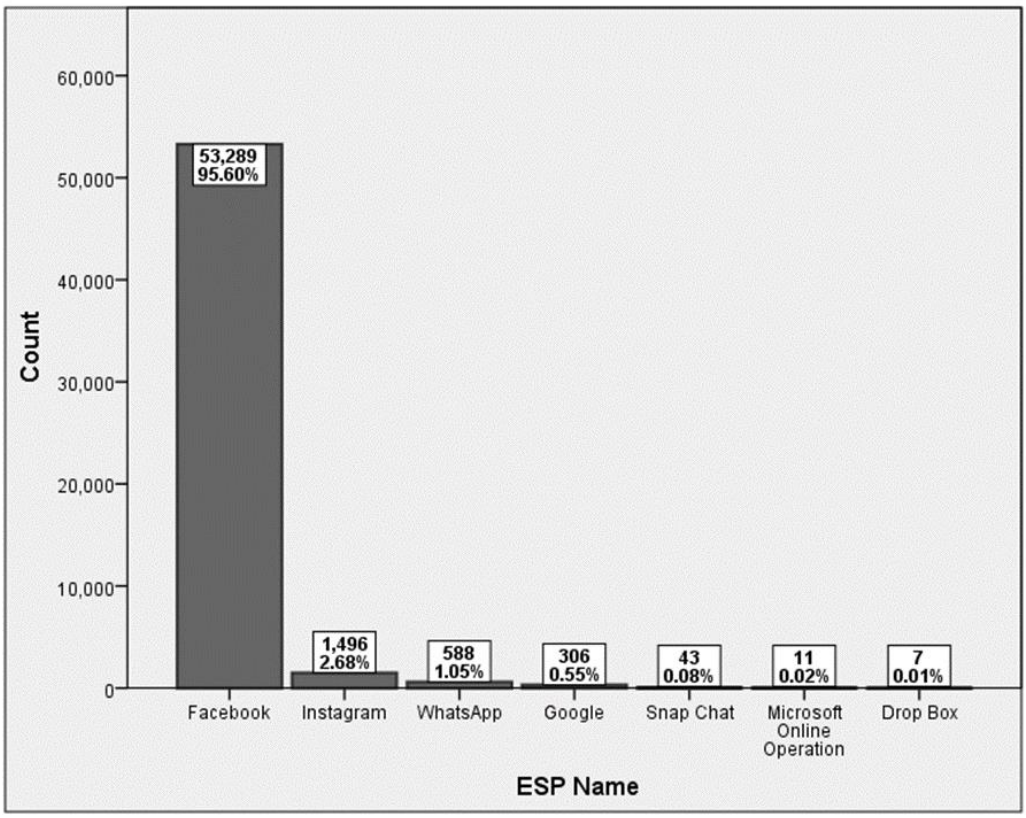

Figure 1 Frequency Distribution of ESP

4.1.2 Frequency Distribution of Industrial Classification

Figure 2 illustrates the different types of adult content shared. The data is divided into 7 different categories. It shows that Pub Minor High 81.19% is the highest type of shared data followed by Pre-Minor High 8.05% and Pub Minor Low 6.95%. High category data consists of content which includes "*abusive explicit conduct, (simulated or actual adult intercourse including anal-genital, genital-genital, or oral-anal between same or opposite gender person), bestially, sadistic or masochistic abuse, degradation, masturbation, or any such depiction that lacks serious political, literary, artistic, or scientific value*". While low category data includes *"Any image that depicts nudity and one or more of: inappropriate touching, sexually suggested poses, adult arousal, focus on genital, spreading of limbs or genitals, and such depiction lacks serious political, literary, scientific or artistic value.".* Data shared in the rest of the categories is low, Pre/Pub Comb 1.47%, Pub High/Low Comb 1.37%, Pre-Minor Low 0.96%, and Pre

High/low Comb 0.01%. These categories are systematically defined in the algorithm for detection. The algorithm divides and arranges itself according to the data sharing as defined in the literature review.

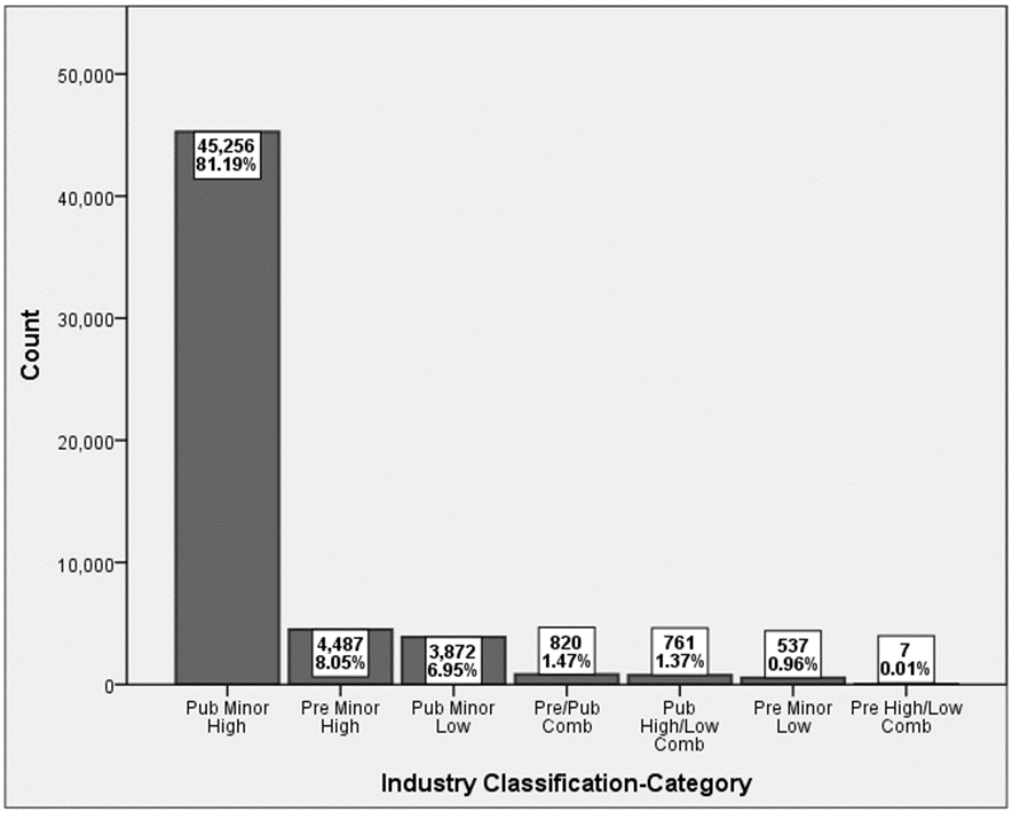

Figure 2 Frequency Distribution of IC

4.1.3 Frequency Distribution of File Relevance

Figure 3 shows the frequency distribution of File Reference Values. FRV is the number of times an ID has been reported. This value varies from Tipline to Tipline. For the dataset used in this study, it ranges from 1 to 335. The most shared type of File relevance value is FRV2, which consumes 79.15% of the total. This shows that 79% of Tiplines are of IDs reported twice. It can also be seen in the data that the FRV of 10+ is only 0.55%, Which shows us that very few IDs have been reported more than ten times. However, these are more likely to be the serious offenders of adult content sharing. Other File relevance values are as follows FRV1 4.40%, FRV3 9.39%, FRV4 2.85%, FRV5 1.76%, FRV6 0.83%, FRV7 0.48%, FRV8 0.33%,

FRV9 0.17%, FRV10 0.11%. From the distribution of the percentages, it's obvious that the data is positively skewed, with majority of the Tiplines created for IDs being reported lower than five times.

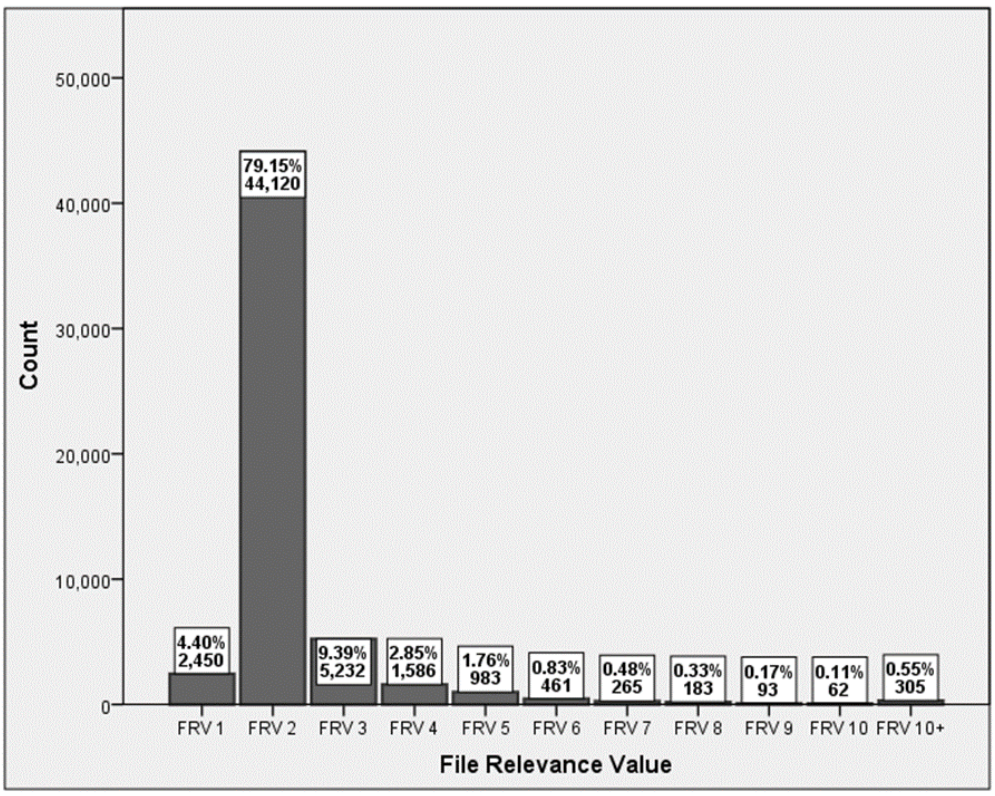

Figure 3 Frequency Distribution of FRV

Further investigation has been performed on FRV10+ (0.55%) Tiplines, which are 305 in total. The values of these Tiplines are analyzed in terms of industrial classification, region, city, and file size. Results suggest that pub minor high and pub high/low comb are the most shared type of media. Most of the data shared is from Punjab (Lahore), Sindh (Karachi), and Islamabad. Descriptive analysis of these 305 tiplines shows that the mean of Khyber Pakhtunkhwa (11.42) is the highest while AJK (8.91) is next. Standard deviation of Khyber Pakhtunkhwa (19.45) suggests that the data is highly spread. In cities Peshawar, Rawalpindi and Karachi has the highest mean value of shared data. While the highest standard deviation in cities is of Peshawar, Lahore and Karachi. Further details are shown in **Table 5-Table 7**.

Table 5 IC of FRV10+

S. No	Industrial Classification	No of files shared
1	Pre High/Low Comb	2
2	Pre Minor High	7
3	Pre Minor Low	3
4	Pre/Pub Comb	72
5	Pub High/Low Comb	60
6	Pub Minor High	117
7	Pub Minor Low	44
8	Total	305

Table 6 Region wise file size of FRV10+

S. No	Region	Mean	n	StDev
1	Sindh	7.17	101	10.66
2	Punjab	6.08	135	9.64
3	Khyber Pakhtunkhwa	11.42	21	19.45
4	Baluchistan	2.29	1	0.0
5	Islamabad	5.80	39	4.33
6	Gilgit Baltistan	5.035	2	1.81
7	AJK	8.91	6	11.39
8	Total	6.81	305	10.47

Table 7 City wise file size of FRV10+

S. No	City	Mean	n	StDev
1	Karachi	7.26	90	11.10
2	Lahore	6.29	77	11.55
3	Islamabad	5.80	39	4.33
4	Rawalpindi	7.72	11	10.97
5	Peshawar	17.02	9	28.20
6	Multan	3.90	8	3.64
7	Faisalabad	6.15	9	6.76
8	Quetta	2.29	1	0.0
9	Other cities	6.33	61	6.31
10	Total	6.81	305	10.47

4.1.4 Frequency Distribution of Cities

Figure 4 shows the 8 cities having the highest amount of reported Tiplines. The remaining cities are clustered as ‘other cities. Karachi has the highest percentage of Tiplines with 28.07%. It is expected as Karachi has the highest population among all the cities in Pakistan [24]. The reported Tiplines in Karachi alone are higher than the collective Tiplines of the remaining 220 cities which are clustered into ‘Other cities’ having 19.46%. Next in line is Lahore with 23.75%, Islamabad 15.78%, Multan 3.76%, Peshawar 2.89%, Rawalpindi 2.86%, Faisalabad 2.52%, Quetta 0.91%. We can see that the major of the city, higher the Tiplines reported.

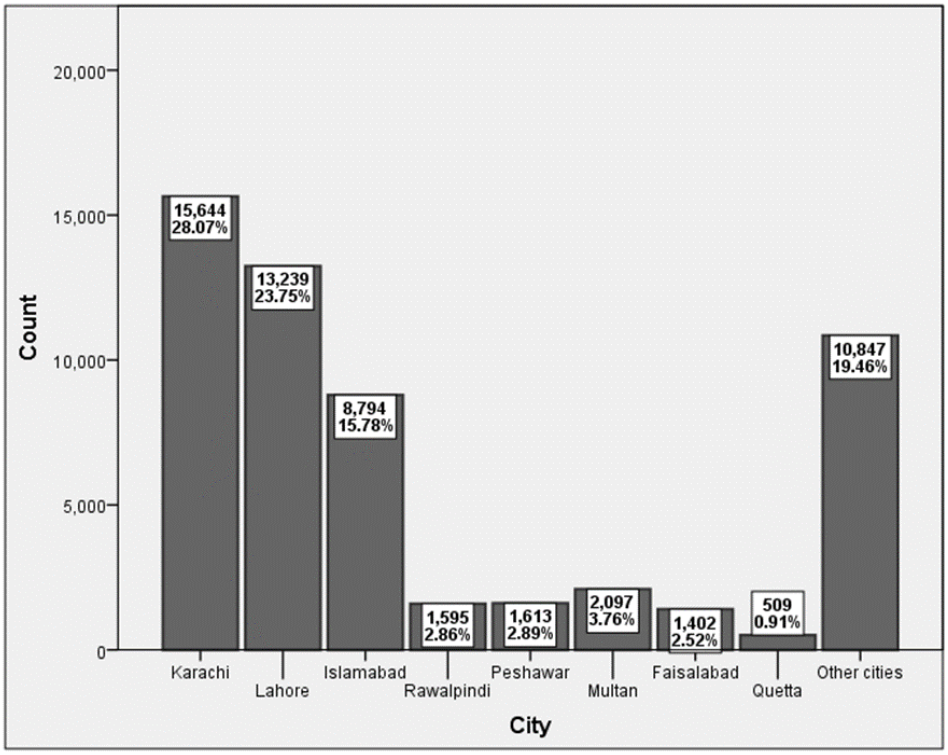

Figure 4 Frequency Distribution of Cities

4.1.5 Frequency Distribution of Region

Figure 5 shows the frequency distribution of the 7 regions of Pakistan. These include Punjab 47.38%, Sindh 28.60%, Khyber Pakhtunkhwa 6.68%, Baluchistan 0.99%, Islamabad 15.08%, Gilgit Baltistan 0.74%, AJK 0.53%. Punjab has the highest frequency followed by Sindh among other regions, since these two are the most populated provinces of Pakistan with greater

amount of IT infrastructure. Islamabad showed the most alarming results. Since being a small city with less population compared to other regions [25], its percentage of data shared is about half of Sindh, which is the second most populated region of Pakistan.

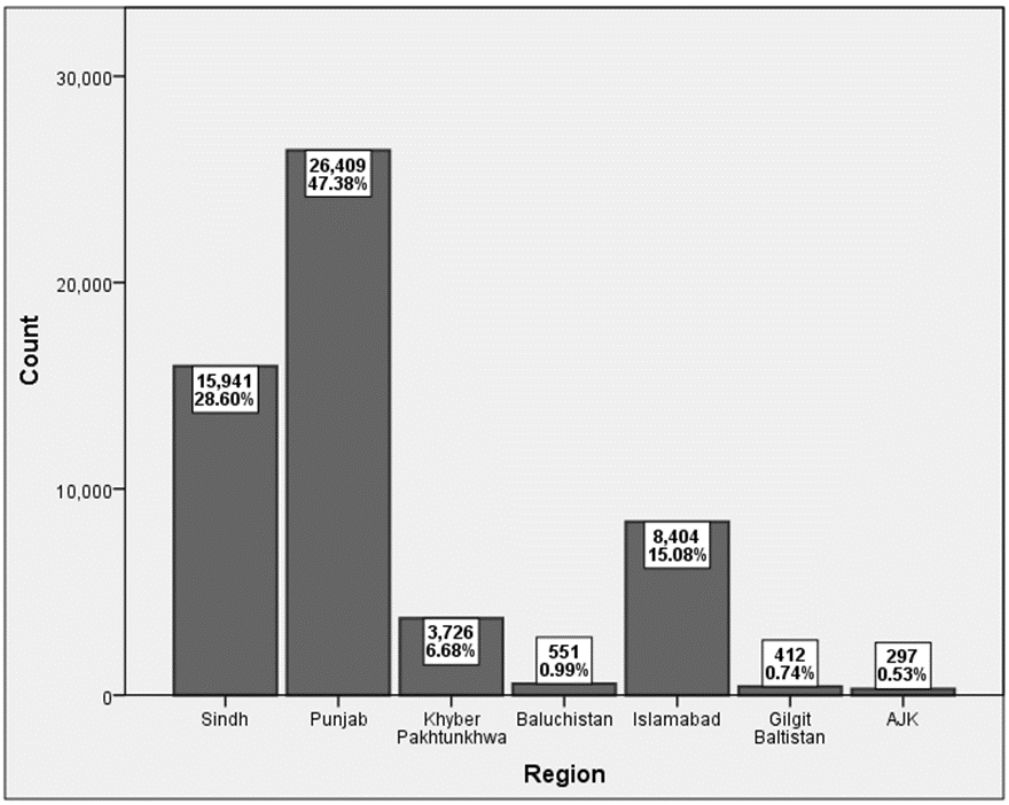

Figure 5 Frequency Distribution of Region

4.1.6 Frequency Distribution of IP/ ISP

Figure 6 shows the frequency distribution of IP/ISP (internet service provider). There are numerous service providers in Pakistan. We have considered the top 6 ISPs with the highest frequency, while the rest are grouped as 'Others'. The percentages associated with different ISPs are, Zong 26.11%, Jazz 25.01%, Telenor 18.07%, PTCL 13.93%, Ufone 3.12%, and Cybernet 2.52%. While the 'Other' ISP is 11.24%. Zong, Jazz and Telenor have a higher frequency then others since these are the most used networks in Pakistan with a better data service [26].

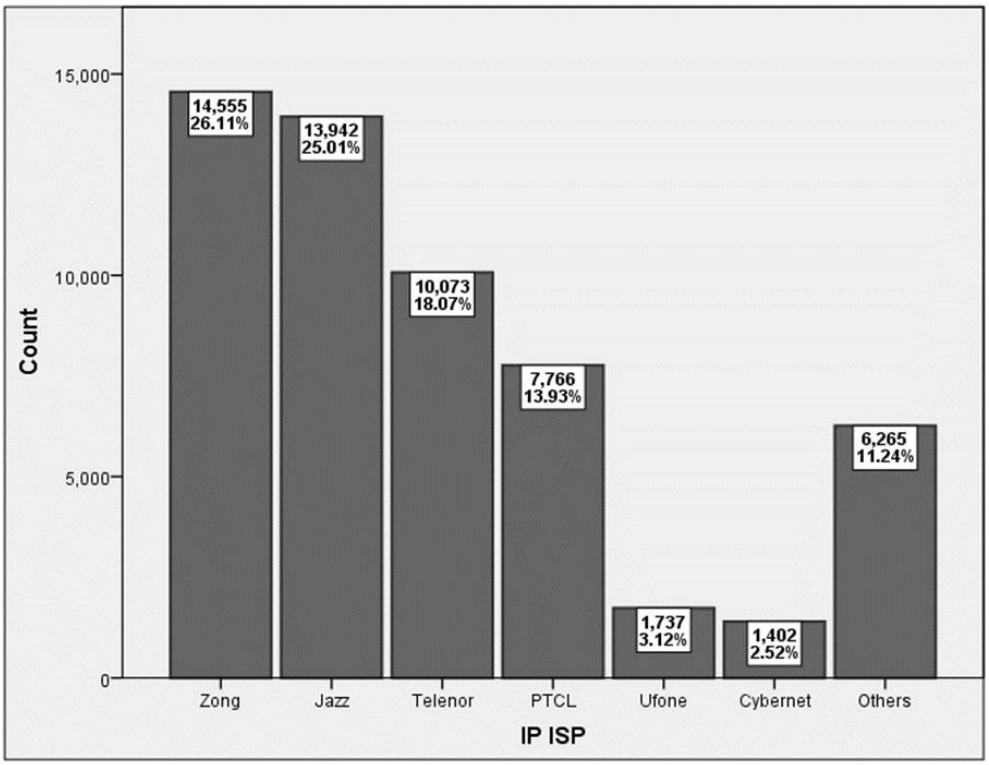

Figure 6 Frequency Distribution of IP/ISP

4.1.7 Descriptive analysis of File size

Table 8 shows the descriptive analysis of the file size. File size is a discrete value which shows how much data has been shared through the Tipline. The maximum shared data is 335Mbs while 0.01Mb is the lowest. On average, 5.88Mbs of data is shared in 55740 reported Tiplines. The slope of the data distribution is skewed with sharp peak as shown by the kurtosis.

Table 8 Descriptive analysis of the File Size (Mbs)

S. No	Statistics	Values
1	n	55740
2	Range	121.06
3	Minimum	0.001
4	Maximum	335
6	Mean	5.88
8	Standard Deviation	6.34
10	Skewness	3.61
12	Kurtosis	27.55

4.2 Bi variate analysis of data

4.2.1 File Relevance Value vs File Size

To identify the habitual offender of adult/explicit content sharing the FRV and file size needed to be further analyzed. *A habitual offender is the user who has previously committed crimes of similar gravity* [27]. Therefore, habitual offender's ID should either be reported multiple number of times, or the file size of their shared content should be high or both. However, we can see in **Figure 7** that there is no direct relation between an ID reported higher number of times with the size of shared file. So, we can assume that only the outlier values in **Figure 7** are the habitual offenders.

Table 9 shows that max and min values of File size is 121.06 and 0.01 while its mean is only 5.88. Similarly, the max and min values of FRV are 335 and 1 while its mean is only 2.39. The standard deviation of FRV and file size is 3.226 and 6.34 respectively. It can be observed in the figure too that the data of file size is more dispersed compared to FRV.

Table 23 shows the correlation value to be 0.046, which is too little to be considered. Hence, this is another proof that there is no linear relation between FRV and file size.

Table 9 Descriptive Statistic of FRV & File size

	n	Min	Max	Mean	StDev	Skewness	Kurtosis
Total File Size (mb)	55740	0.010	121.06	5.88	6.34	3.614	27.551
FRV	55740	1	335	2.39	3.226	46.478	3333.318

4.2.2 Region wise File size

In **Table 10** the quantity of data shared in different regions is shown. It is obvious that even though the regions have different number of reported Tiplines, all the regions have comparable means. This indicates that all the regions have the same intensity of sharing the explicit / adult content. However, the means are not statistically equal as shown in **Table 23**. The standard deviation of the regions is approximately similar, which shows that average data dispersion is around the mean.

Table 10 Descriptive Analysis of Region & File size

S. No.	Region	Mean	n	StDev	Min	Max
1	Sindh	5.99	15941	6.50	0.01	96.46
2	Punjab	5.84	26409	6.36	0.01	121.06
3	KP	5.99	3726	6.63	0.01	93.74
4	Baluchistan	5.42	551	5.71	0.01	48.58
5	Islamabad	5.86	8404	6.00	0.01	96.05
6	Gilgit Baltistan	5.21	412	5.25	0.01	41.67
7	AJK	5.03	297	4.68	0.01	31.59
8	Total	5.88	55740	6.35	0.01	121.06

4.2.3 Region wise ESP

Table 11 shows the observed and Fe of electronic service providers in all the regions. It is visible that the medium used mostly in Pakistan to share adult / explicit content is Facebook. As discussed earlier in the literature review, the number of Facebook users is more compared to any other social/chatting platform, similar findings are reflected here as well. In each region, Facebook is the topmost used platform for this activity. The Chi square test results are shown

in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables Region and ESP are associated with each other.

Table 11 Crosstabulation of Region and ESP

			Region							
			AJK	BA	GB	IS	KP	PB	SD	Total
ESP	Dropbox	Fo	0	0	0	1	0	3	3	7
		Fe	0.0	0.1	0.1	1.1	0.5	3.3	2.0	7.0
	Facebook	Fo	282	532	402	8197	3633	25121	15122	53289
		Fe	283.9	526.8	393.9	8033.5	3561.2	25243.9	15240	53289
	Google	Fo	3	2	0	23	9	150	119	306
		Fe	1.6	3.0	2.3	46.1	20.4	145.0	87.5	306
	Instagram	Fo	12	15	8	156	74	810	421	1496
		Fe	8.0	14.8	11.1	225.5	100.0	708.7	427.8	1496
	Microsoft	Fo	0	0	0	2	0	5	4	11
		Fe	0.1	0.1	0.1	1.7	0.7	5.2	3.1	11.0
	Snapchat	Fo	0	2	1	5	6	20	9	43
		Fe	0.2	0.4	0.3	6.5	2.9	20.4	12.3	43.0
	WhatsApp	Fo	0	0	1	19	4	301	263	588
		Fe	3.1	5.8	4.3	88.6	39.3	278.5	168.2	588.0
Total		Fo	297	551	412	8403	3726	26410	15941	55740
		Fe	297	551	412	8403	3725	26405	15941	55740

Table 12 Association of attributes

Sr. No	Attributes	Estimated Chi-Square Value	p-value
1	ESP, Region	253.588	0.000
2	Region, IC	211.745	0.000
3	Region, IP/ISP	13363.033	0.000
4	FRV, IC	15376.817	0.000
5	City, IC	203.554	0.000
6	IP/ISP, IC	216.809	0.000

4.2.4 Region wise Industry Classification

Table 13 shows the observed and Fe of the type of data shared in different regions of Pakistan. Pub Minor High is the most shared data category in every region. While the second most shared category is pre minor high. In both the cases, the shared data is related to mating content rather

than venereal act. The Chi square test results are shown in **Table 12** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables Region and IC are associated with each other.

Table 13 Crosstabulation of Region and IC

			Region							
			AJK	BA	GB	IS	KP	PB	SD	Total
IC	Pre High/Low Comb	Fo	0	0	0	2	0	1	4	7
		Fe	.0	.1	.1	1.1	.5	3.3	2.0	7.0
	Pre Minor High	Fo	13	36	48	610	250	2241	1289	4487
		Fe	23.9	44.4	33.2	676.4	299.9	2125.6	1283.2	4487
	Pre Minor Low	Fo	1	2	2	31	7	248	246	537
		Fe	2.9	5.3	4.0	81.0	35.9	254.4	153.6	537.0
	Pre/Pub Comb	Fo	4	6	10	110	43	406	241	820
		Fe	4.4	8.1	6.1	123.6	54.8	388.4	234.5	820.0
	Pub High/Low Comb	Fo	5	15	2	90	54	379	216	761
		Fe	4.1	7.5	5.6	114.7	50.9	360.5	217.6	761.0
	Pub Minor High	Fo	255	456	321	6966	3175	21295	12788	45256
		Fe	241.1	447.4	334.5	6822.5	3024.4	21438.5	12942.7	45256
	Pub Minor Low	Fo	19	36	29	594	197	1840	1157	3872
		Fe	20.6	38.3	28.6	583.7	258.8	1834.2	1107.3	3872
Total		Fo	297	551	412	8403	3726	26410	15941	55740
		Fe	297	551	412	8403	3725	26405	15941	55740

4.2.5 Region wise File Relevance Value

In **Table 14,** FRV is analyzed with respect to different regions of Pakistan. The average reported Tiplines in all the regions are approximately similar. However, their standard deviations are different. Sindh has the highest standard deviation, followed by AJK.

Baluchistan on the other hand Baluchistan has the lowest standard deviation since the spread around mean in its data is less with a maximum FRV of only 11.

Table 14 Descriptive Analysis of FRV & Region

S. No.	Region	Mean	n	Sum	Min	Max	StDev
1	Sindh	2.44	15941	38950	1	335	4.257
2	Punjab	2.37	26409	62597	1	160	2.537
3	KPK	2.31	3726	8618	1	28	1.371
4	Baluchistan	2.36	551	1303	1	11	1.073
5	Islamabad	2.39	8404	20079	1	186	3.572
6	Gilgit Baltistan	2.40	412	989	1	46	2.393
7	AJK	2.72	297	808	1	55	3.684
8	Total	2.39	55740	133344	1	335	3.226

4.2.6 Region Wise IP/ ISP

Table 15 shows different mobile/internet service providers being used by the users. In Pakistan commonly 4 mobile networks are used, namely Zong, Mobilink, Telenor and Ufone. There are several wired and wireless networks also available all-around Pakistan, in which the most popular one is PTCL. Mobile networks can be used as Mobile data networks. Consumption of the mobile / wireless network depends upon the different reasons. These include the availability of service, geographical location, mobile packages etc. In Baluchistan PTCL is the most used network among all, the reason is its geographical location. Baluchistan has mountainous terrain; wireless towers cannot cover all the regions. Therefore, they use PTCL wired/Wireless network. AJK and GB have fewer mobile networks activated because of the geographical location and less advancement compared to the other regions. Sindh, Punjab, and Islamabad are the only regions where all the networks are being used properly. In totality, Zong, Telenor, and Jazz are the most used mobile networks. The Chi square test results are shown in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables Region and IP/ISP are associated with each other.

Table 15 Crosstabulation of Region & IP/ISP

			IP/ISP							Total
			Cyber net	Jazz	Others	PTCL	Telenor	Ufone	Zong	
Region	AJK	Fo	0	9	170	27	46	5	40	297
		Fe	7.5	74.3	33.4	41.4	53.7	9.3	77.6	297
	BA	Fo	42	3	47	445	8	0	6	551
		Fe	13.9	137.8	61.9	76.8	99.6	17.2	143.9	551
	GB	Fo	0	3	403	0	3	0	3	412
		Fe	10.4	103.1	46.3	57.4	74.5	12.8	107.6	412
	IS	Fo	0	1856	342	640	2261	261	3043	8403
		Fe	211.4	2101.8	944.5	1170.8	1518.5	261.9	2194.2	8403
	KP	Fo	220	1031	329	1824	215	6	101	3726
		Fe	93.7	931.7	418.7	519.0	673.2	116.1	972.7	3725
	PB	Fo	479	7772	2771	3389	4741	816	6442	26410
		Fe	664.2	6604.6	2967.8	3678.9	4771.8	822.8	6895.0	26405
	SD	Fo	661	3268	2203	1441	2799	649	4920	15941
		Fe	401.0	3987.3	1791.7	2221.0	2880.8	496.8	4162.6	15941
Total		Fo	1402	13942	6265	7766	10073	1737	14555	55740
		Fe	1402	13942	6265	7766	10073	1737	14555	55740

4.2.7 Industrial Classifications of FRV

Table 16 shows the number of times an ID of industrial classification is reported. In each row of industrial classification, Pub Minor High is the highest. Moreover, the percentage of industrial classification content pub minor is decreasing with the increasing FRV. The frequency of other industrial classification content sharing increases from FRV 9 to 10+. With the increase in FRV we can also see an increase in multiple industrial classification categories. This means that an ID which is reported a greater number of times shares data from multiple categories. Although, the number of IDs reported with higher FRV is less, the intensity of their shared content is more alarming. If an ID is reported 9 or more times, there is a possibility it's

a habitual offender. With the increase in FRV there is an exponential decrease in the number of reported IDs. The Chi square test results are shown in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables FRV & IC are associated with each other.

Table 16 Crosstabulation of FRV & IC

			IC							
			Pre High/Low Comb	Pre Minor High	Pre Minor Low	Pre/Pub Comb	Pub High/Low Comb	Pub Minor High	Pub Minor Low	Total
FRV	FRV 1	Fo	0	192	384	0	0	1536	338	2450
		Fe	0.3	197.2	23.6	36.0	33.4	1989.2	170.2	2450
	FRV 2	Fo	0	3797	76	31	27	37088	3101	44120
		Fe	5.5	3551.6	425.1	649.1	602.4	35821.6	3064.8	44120
	FRV 3	Fo	1	200	34	424	382	4048	143	5232
		Fe	0.7	421.2	50.4	77.0	71.4	4247.9	363.4	5232
	FRV 4	Fo	4	112	14	126	113	1151	66	1586
		Fe	0.2	127.7	15.3	23.3	21.7	1287.7	110.2	1586
	FRV 5	Fo	0	106	11	64	53	675	74	983
		Fe	0.1	79.1	9.5	14.5	13.4	798.1	68.3	983.0
	FRV 6	Fo	0	45	6	36	43	291	40	461
		Fe	0.1	37.1	4.4	6.8	6.3	374.3	32.0	461.0
	FRV 7	Fo	0	14	2	31	29	157	32	265
		Fe	0.0	21.3	2.6	3.9	3.6	215.2	18.4	265.0
	FRV 8	Fo	0	5	5	13	22	118	20	183
		Fe	0.0	14.7	1.8	2.7	2.5	148.6	12.7	183.0
	FRV 9	Fo	0	1	2	11	18	53	8	93
		Fe	0.0	7.5	0.9	1.4	1.3	75.5	6.5	93.0
	FRV 10	Fo	0	8	0	12	14	22	6	62
		Fe	0	5.0	0.6	0.9	0.8	50.3	4.3	62.0
	FRV 10+	Fo	2	7	3	72	60	117	44	305
		Fe	0.0	24.6	2.9	4.5	4.2	247.6	21.2	305.0
Total		Fo	7	4487	537	820	761	45256	3872	55740
		Fe	7	4487	537	820	761	45256	3872	55740

4.2.8 Industrial classification of City

Figure 7 shows the type of content shared with respective to different cities of Pakistan. The top cities with the highest shared content are Karachi Lahore and Islamabad respectively and Pub Minor High has the highest frequency in each city. The combination of other cities shows a higher count too, but the values are negligible because it's a cluster of 300+ cities. The Chi square test results are shown in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables City and IC are associated with each other.

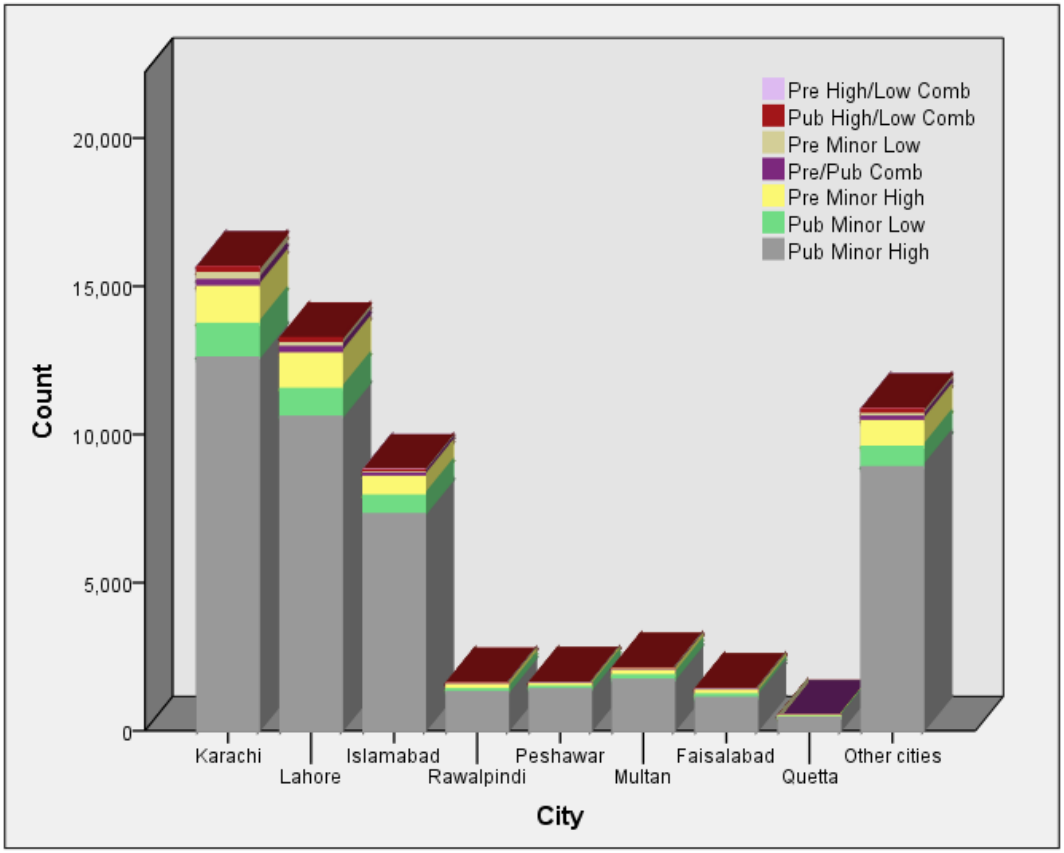

Figure 7 Industrial Classification of Cities

4.2.9 Industrial classification of IP/ISP

Figure 8 shows the type of content shared using different telecommunication networks. It's visible that Zong, Jazz and Telenor have the highest frequencies of shared data. This is not a surprise though, since these networks have the highest number of telecom users in Pakistan.

Pub Minor High is the most shared type of data followed by Pre-Minor High and Pub-Minor Low. The Chi square test results are shown in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables IP/ISP and IC are associated with each other.

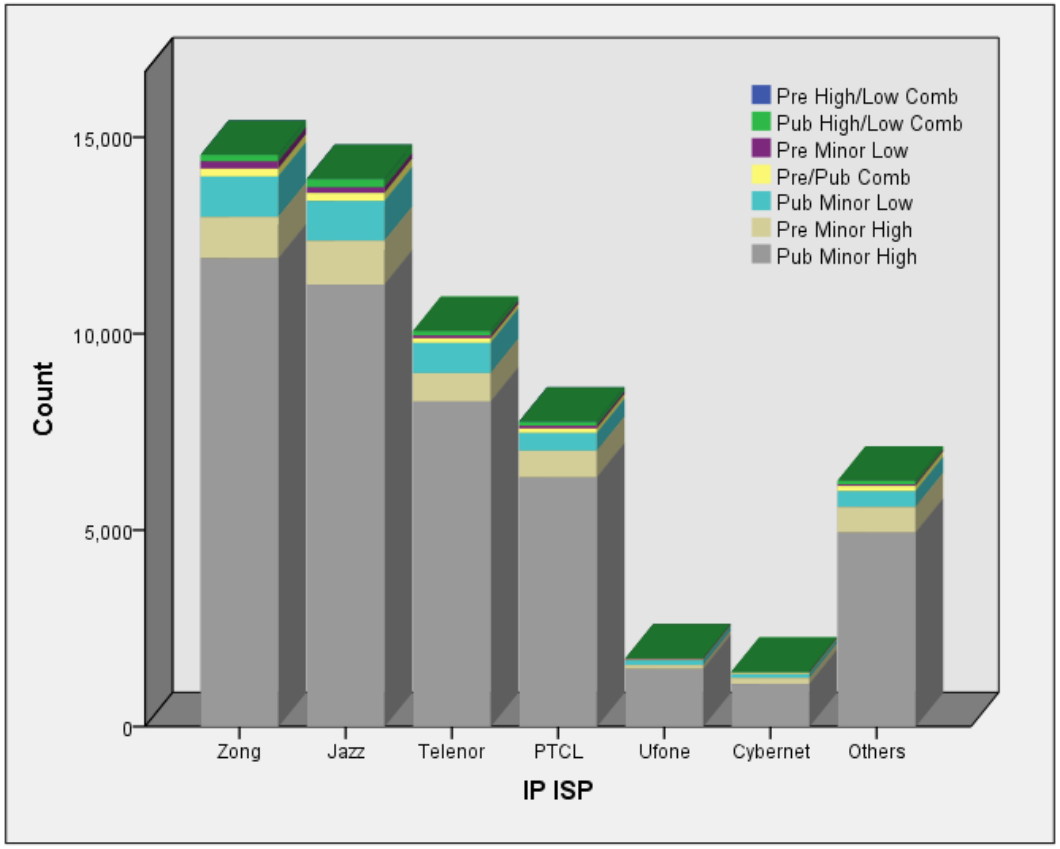

Figure 8 Industrial classification of IP / ISP

4.2.10 City wise IP/ISP

Figure 9 shows the networks used in different cities to share adult content. It's visible that Zong, Jazz and Telenor have the highest frequencies of shared data in Karachi, Lahore and Islamabad. The Chi square test results are shown in **Table 12.** Since the p-value is less than 5%, we reject the null hypothesis and conclude that the variables IP/ISP and City are associated with each other.

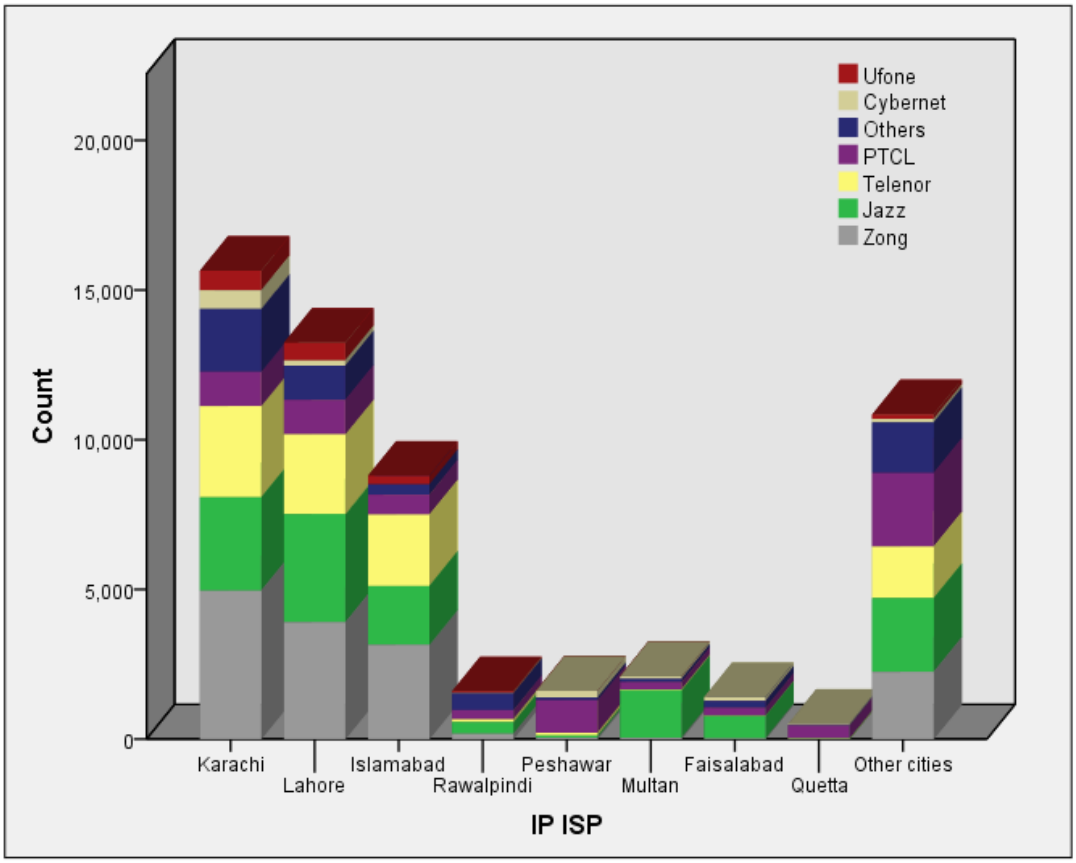

Figure 9 City wise IP/ISP

4.2.11 Point biserial correlation

Point biserial correlation is used since File Size is a quantitative variable. The relation between file size and other variables is checked whether the increment in one has any effect on the other. It can be observed in **Table 17** that there exists no correlation between file size and other variables. The values are closer to 0 suggesting a very weak linear relation.

Table 17 Point-biserial correlation of file size (mb) and qualitative variables

Variables	Estimates of Correlation	Interpretation
FRV	0.046	Weak positive correlation
IP ISP	0.007	Weak positive correlation
IC	-0.014	Weak negative correlation
Region	-0.01	Weak negative correlation
ESP Name	-0.008	Weak negative correlation
City	-0.006	Weak negative correlation

4.3 Analysis of Variance (ANOVA)

ANOVA is used in a variety of situations to discover whether there are any differences between the means of various groups. In this study we have used one-way ANOVA of File size with respect to different variables.

Table 18 to 21 shows the means of file size with respect to different variables, their f-value and p-value. Only the variable City and IP/ISP has p-values greater than the level of significance as shown in **Table 20 and Table 21**. Their null hypothesis cannot therefore be rejected, suggesting that the means of these two variables are statistically equal. Apart from these 2 variables, the null hypothesis for all the other variable is rejected. Suggesting that the mean values of ESP, Region, IC and FRV are not equal as shown in **Table 18-Table 21**. The tables also mention the class interval of the independent variable which is assumed to be 95% for this ANOVA test. The class interval of 95% states that if 100 times similar data is used for this study, 95% of the time their mean values will remain in between class interval.

Table 18 ANOVA of ESP

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
ESP	1324	220.65	5.48	0.000
Error	2243143	40.25		
Total	2244467			
Descriptives				
ESP	n	Mean	StDev	95% CI of mean
Dropbox	7	8.37	17.27	(3.67, 13.07)
Facebook	53289	5.91	6.36	(5.86, 5.96)
Google	306	6.14	7.74	(5.43, 6.85)
Instagram	1496	5.09	5.22	(4.77, 5.41)
Microsoft	11	4.83	3.67	(1.08, 8.57)
Snapchat	43	4.48	5.96	(2.59, 6.38)
WhatsApp	588	5.33	6.70	(4.81, 5.84)
Total	55740			

Table 19 ANOVA of IC

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
IC	4672	778.68	19.38	0.000
Error	2239795	40.19		
Total	2244467			
Descriptives				
IC	n	Mean	StDev	95% CI of mean
Pre High/Low Comb	7	2.40	1.74	(-2.29, 7.10)
Pre Minor High	4487	5.81	6.39	(5.62, 5.99)
Pre Minor Low	537	5.20	5.37	(4.66, 5.73)
Pre/Pub Comb	820	7.62	10.05	(7.19, 8.06)
Pub High/Low Comb	761	7.00	8.58	(6.55, 7.45)
Pub Minor High	45256	5.89	6.21	(5.83, 5.95)
Pub Minor Low	3872	5.41	6.34	(5.21, 5.61)
Total	55740			

Table 20 ANOVA of FR

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
FRV	7199	719.93	17.93	0.000
Error	2237268	40.15		
Total	2244467			
Descriptives				
FRV	n	Mean	StDev	95% CI of mean
FRV - 1	2450	5.67	5.98	(5.42, 5.92)
FRV – 2	44120	5.74	6.10	(5.68, 5.79)
FRV – 3	5232	6.38	6.83	(6.21, 6.55)
FRV – 4	1586	7.01	7.68	(6.70, 7.33)
FRV – 5	983	6.49	7.63	(6.09, 6.89)
FRV – 6	461	7.15	7.93	(6.58, 7.73)
FRV – 7	265	7.83	9.52	(7.06, 8.59)
FRV – 8	183	7.31	9.45	(6.39, 8.23)
FRV – 9	93	6.97	8.73	(5.68, 8.26)
FRV – 10	62	5.78	8.39	(4.21, 7.36)
FRV - 10+	305	6.82	10.48	(6.10, 7.53)
Total	55740			

Table 21 ANOVA of region

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
Region	803	133.8	3.32	0.003
Error	2243664	40.26		
Total	2244467			
Descriptives				
Region	n	Mean	StDev	95% CI of mean
AJK	297	5.03	4.68	(4.31, 5.76)
BA	551	5.42	5.71	(4.89, 5.95)
GB	412	5.21	5.25	(4.59, 5.83)
IS	8403	5.86	6.00	(5.73, 5.99)
KP	3726	5.99	6.63	(5.79, 6.19)
PB	26410	5.84	6.36	(5.76, 5.91)
SD	15941	5.99	6.50	(5.89, 6.09)
Total	55740			

Table 22 ANOVA of city

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
City	443	55.38	1.38	0.202
Error	2244024	40.27		
Total	2244467			
Descriptives				
City	n	Mean	StDev	95% CI of mean
Faisalabad	1402	5.77	6.27	(5.45, 6.11)
Islamabad	8794	5.86	6.01	(5.73, 5.99)
Karachi	15644	5.98	6.44	(5.88, 6.08)
Lahore	13238	5.82	6.41	(5.72, 5.94)
Multan	2097	5.67	6.00	(5.40, 5.94)
Other Cities	10851	5.85	6.36	(5.74, 5.98)
Peshawar	1610	6.07	7.16	(5.77, 6.39)
Quetta	509	5.51	5.71	(4.96, 6.06)
Rawalpindi	1595	5.87	6.47	(5.56, 6.19)
Total	55740			

Table 23 ANOVA of IP/ISP

Analysis of Variance				
Source of Variation	Adj SS	Adj MS	F-Value	P-Value
IP/ISP	374	62.27	1.55	0.158
Error	2244094	40.27		
Total	2244467			
Descriptives				
IP/ISP	n	Mean	StDev	95% CI of mean
Cybernet	1402	6.36	7.25	(6.03, 6.69)
Jazz	13942	5.87	6.35	(5.76, 5.97)
Others	6265	5.92	6.35	(5.77, 6.08)
PTCL	7766	5.85	6.45	(5.71, 5.99)
Telenor	10073	5.89	6.25	(5.77, 6.02)
Ufone	1737	5.90	6.11	(5.59, 6.19)
Zong	14555	5.83	6.29	(5.73, 5.94)
Total	55740			

Chapter 4

Conclusion

This study is an exploratory analysis of first of its kind in Pakistan to explore the adult content sharing of various regions from Jan 2021 to Oct 2021. The data used for this study is a closed source and only the officials of law enforcement agencies have access to it. The analysis is based on frequency distribution, mean, point biserial correlation and association of attributes using chi-square analysis and analysis of variance to test the equality of sample population means have been used. Observatories of different variables are based on the data provided by the NCMEC, which is a nonprofit American organization that works with law enforcement agency around the world. 55741 Tiplines provided by NCMEC are used to perform the exploratory analysis.

As per the analysis, Facebook is the most used medium in Pakistan for sharing the adult content since it has the highest number of users compared to other social media platforms. While the type of content shared the most belonged to the Pub Minor High category which includes abusive explicit conduct. Most of the users are reported only a couple of times, while the users reported higher number of times are very less. However, we can assume these to be serious criminals. The users reported higher number of times are mainly from Peshawar, Karachi, Rawalpindi and Lahore

Karachi has the highest reported users due to its large population, while Lahore and Islamabad followed up. These are the largest metropolitan cities with greater amount of IT infrastructure, maybe due to this, there file relevance values are also the highest. In regions, most of the tiplines are generated from Punjab, Sindh, and Islamabad respectively. Punjab and Sindh are to two most populated provinces of Pakistan with greater advancements compared to other

regions. However, Islamabad ranking in top 3 regions is a bit alarming. Since the population of Islamabad is lower than other regions. An important thing to note here is that the average data shared in each region is about the same regardless of the number of users. This again is an alarming situation for regions with a smaller number of users e.g., Islamabad.

Zong and Jazz are the two most used networks by the users since their service and coverage in Pakistan is better. However, this can also be a reason why most data shared has low file size, because mobile networks have limited packages. The highest reported file size in the data is shared through PTCL which usually has unlimited data packages. PTCL is used mostly in Baluchistan, maybe due to its geographical location. AJK and GB have fewer mobile network activated users due to its location and less advancement in infrastructure as compared to another region. Sindh, Punjab, and Islamabad are the only regions where all the networks are operated properly.

There is no linear relation between any of the variables. Even File relevance and file size has no correlation. Which means that a user being reported greater number of times does not necessarily share more data. It is also discovered that the user who is reported more is sharing content from multiple categories. All the variables are associated with each other. While only the means of city and IP/ISP are statistically equal.

5.1 Recommendation

By using the above information, law enforcement agencies, government officials and associated researchers can easily create plans and make policies to combat adult content sharing, especially in the regions where more data is shared. PTA and Telecommunication companies can also play their role in reducing adult content sharing and restricting fake users. They can also revise their data packages to limit adult content sharing. Digital media can raise awareness on the topic to educate people that sharing such content even for entertainment

purposes is also a crime. Educating people about the ramification of sharing pictures that are intended to be private may go a long way in preventing this behavior. If people realize that posting nude or even semi-nude photos of others, or sharing photographs meant to be private have serious legal consequences and psychological consequences perhaps they will deter from engaging in these behaviors.

5.2 Limitations

This is an exploratory analysis study and only highlights useful statistics, however further exploration and data is required to explore reasons behind it. This is a limited data of 9 months, to develop a model we need data of more years.

5.3 Future recommendation

- This is an exploratory analysis study and only highlights useful statistics, however further exploration and data is required to explore reasons behind it.
- This is a limited data of 9 months, to develop by model for testing and training to identify the serious offenders, more data is required.

References

[1] "Adult Content," defination.net, [Online]. Available: https://www.definitions.net/definition/adult+content.

[2] Jaishankar & Halder, "Revenge porn by teens in the United States and India: A socio-legal analysis.," *International Annals of Criminology,* pp. 51, 85-111, 2013.

[3] datareportal, "DIGITAL 2022: PAKISTAN," 2021. [Online]. Available: https://datareportal.com/reports/digital-2021-pakistan.

[4] datareportal, "DIGITAL 2022: PAKISTAN," 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-pakistan.

[5] Monika Singh & Divya Bansal, "Behavioral analysis and classification of spammers distributing," *cross mark,* 2016.

[6] Ahmed, Kathy Gannon & Munir, "Aid group reports surging numbers in child abuse in Pakistan," AP News, 2022.

[7] Pakistan penal code 1860, "Pakistan parliament passes law against child sexual abuse," News 18, 2016.

[8] Express tribune, "NA approves harsher punishments for child abuse," Express tribune, 2018.

[9] "Hussain Khanwala: Village scarred by child abuse scandal," Al Jazeera, 2015.

[10] Dnpindia, [Online]. Available: https://www.dnpindia.in/.

[11] Quayle, M. Taylor , "Child pornography: An internet crime,," *Brunner-Routledge,* 2003.

[12] L. Edwards, "Content ltering and the new censorship," in *IEEE Intl. Conference on Digital Society, 2010*, 2010.

[13] Bauman & Sinclair, "Cyber and biasbased harassment: Associations with academic, substance use, and mental health problems," *Gender and Society,,* 2012.

[14] Halder & Jaishankar, "Teen sexting: A critical analysis on the criminalization vis-à-vis victimizationn.," 2014.

[15] Megan, "Schools crack down on teen 'sexting'," 2008. https://phys.org/news/2009-09-schools-teen-sexting.html (A*ccessed on 8/9/2022).*

[16] Vandiver & Martinez-Prather, "Sexting among teenagers in the United States: A retrospective analysis of identifying motivating factors, potential targets, and the role of a capable guardian.," *International Journal of Cyber Criminology,* 2014.

[17] Schrock, & Schwalbe, "Men, masculinity, and manhood acts.," *Annual Review of Sociology, 35, 277-295.,* 2009.

[18] Fritz & Stahl, "Internet safety: adolescents' self-report.," *Journal of Adolescent Health 285: 3011–3014.,* 2002.

[19] Rideout, "Generation Rx.com: how young people use the Internet for health information.," *Kaiser Family Foundation,* 2001.

[20] "https://www.missingkids.org/," NCMEC, 2021. [Online]. Available: https://www.missingkids.org/content/dam/missingkids/pdfs/2021-notifications-by-ncmec-per-esp.pdf.

[21] "https://www.missingkids.org/," NCMEC, 2021. [Online]. Available: https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf .

[22] "Missing Kids," NCMEC, [Online]. Available: https://www.missingkids.org/supportus/our-corporate-partners.

[23] Iqbal & Marrington, "A study of detecting child pornography on smart phone.," in *InInternational Conference on Network-Based Information Systems*, 2017.

[24] "Karachi, Pakistan Metro Area Population 1950-2022," 2019. [Online]. Available: https://www.macrotrends.net/cities/22044/karachi/population.

[25] "Information about Population of Pakistan and its Provinces," [Online]. Available: https://www.pakinformation.com/population.html. [Accessed 23 2 2022].

[26] "MOBILE NETWORKS," PAKISTAN TELECOMMUNICATION AUTHORITY, [Online]. Available: https://www.pta.gov.pk/assets/media/qos_survey_29062021.pdf.

[27] Philip, Jenkins & John, "habitual offender," Britannica, https://www.britannica.com/topic/habitual-offender, (A*ccessed on 16/11/2022).*

[28] Government of Pakistan, "Pakistan electronic crime act 2016," 2016.

[29] Bahadur, "Victims of 'revenge porn' open up on reddit about how it impacted their lives.," 2014, http://www.huffingtonpost.com/2014/01/09/revenge-porn-stories-realimpact_n_4568623.html, (A*ccessed on 12/12/2022).*

[30] D. Zillmann, " Influence of unrestrained access to erotica on adolescents' and young adults' dispositions towards sexuality," *Journal of Adolescent Health 27:41–44.*, 2000.

[31] "https://www.opensignal.com," opensignal, 2022. [Online]. Available: https://www.opensignal.com/reports/2022/02/pakistan/mobile-network-experience .

[32] "DIGITAL 2022: PAKISTAN," datareportal.com, 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-pakistan/.

[33] A. Alasa, "A Legal Analysis of Cybercrimes and Cybertorts: Lessons for Nigeria.," *accessed [12/1/2021].*, 2019.

[34] V. Dupont, "Celebrity photo hackers 'committed sex crime," 2014. [Online]. Available: http://phys.org/news/2014-09-celebrity-photo-hackers-committedsex.html.

[35] J. Garcia, "Adult Content Sharing and the Technological," *ResearchGate DOI:10.13140/RG.2.1.2065.0644*, 2016.

[36] S. Becker, "Is sexual erotica associated with sexual deviance in adolescent males?," *International Journal of Law and Psychiatry 14:85–95.*.

[37] Huberman & Berne, "European approaches to adolescent sexual behavior and responsibility," *Advoctaes for youth*, 1999.

[38] Wolak & Finkelhor, "Online victimization: a report on the nation's young people.," *National Center for Missing & Exploited Children.*, 2000.

[39] Frank & Jacobs, "Cyber civil rights initiative., 2017" https://www.cybercivilrights.org/definitions/Jordan (A*ccessed on 12/09/2022).*

[40] Henry & Powell, "Technology-facilitated sexual violence victimization.," *Journal of Interpersonal Violence, 34, 3637–3665.*, 2019.

Appendix I

Under section 21 of PECA: Offences against dignity of a natural person. –

(l) Whoever intentionally and publicly exhibits or displays or transmits any information through any information system, which he knows 1o be false, and intimidates or harms the reputation or primacy of a natural person, shall be punished with imprisonment for a term which may extended to three years or with fine which may extent to one million rupees or with both:

(2) Any aggrieved person or his guardian, where such person is a minor, may apply to the Authority for removal, destruction of or blocking access to such information referred to in sub-section (l) and the Authority on receipt of such application, shall forthwith pass such orders as deemed reasonable in the circumstances including an order for removal, destruction, preventing transmission of or blocking access to such information and the Authority may also direct any of its licenses to secure such information including traffic data,

Under section 22 of PECA*: *Offences against modesty of a natural person and minor.

–

(l) Whoever intentionally and publicly exhibits or displays or transmits any information which-

(a) superimposes a photograph of the face of a natural person over any sexually explicit image or video; or

(b) includes a photograph or a video of a natural person in sexually explicit conduct; or

(c) intimidates a Natural person with any sexual act, or any sexually explicit image or video of a natural person; or

(d) cultivates, entices, or induces a natural person to engage in a sexually explicit act, through an information system to harm a natural person or his reputation, or to take revenge, or to create hatred or to blackmail, shall be punished with imprisonment for a term which may extend to five years or with fine which may extend to five million rupees or with both.

Under section 22A of PECA: Child pornography. –

(l) Whoever intentionally produces, offers or makes available, distributes or transmits through an information system or procures for himself or for another person or without lawful justification possesses material in an information system, that visuality depicts-

(a) a minor engaged in sexually explicit conduct:

(b) a person, appearing to be a minor engaged in sexually explicit conducti or

(c) realistic images representing a minor engaged in sexually explicit conduct; or

(d) discloses the identity of the minor. shall be punished with imprisonment for a term which may extend to seven years, or with fine which may extend to five million rupees or with both.

(2) Any aggrieved person or his guardian, where such person is a minor, may apply to the Authority for removal, destruction of or blocking access to such information referred to in sub-section (l) and the Authority, on receipt of such application, shall forthwith pass such orders as deemed reasonable in the circumstances, including an order for removal, destruction, preventing transmission of or blocking access to such information and the Authority may also direct any of its licenses to secure such information including traffic data." [28]

Appendix II

CyberTipline Report 83113748

Priority Level: E

(Report submitted by a registered Electronic Service Provider)

Received by NCMEC on 11-25-2020 12:33:10 UTC

All dates are displayed as MM-DD-YYYY

Except for times provided in Additional Information sections, all time zones are displayed in UTC

Executive Summary

The following is a brief overview of information contained in this CyberTipline report:

Incident Type: Auto-referred International
Files Not Reviewed by NCMEC
Total Uploaded Files: 9

The National Center for Missing & Exploited Children (NCMEC) was incorporated in 1984 by child advocates as a private, non-profit
501(c)(3) organization to serve as a national clearinghouse and resource center for families, victims, private organizations, law
enforcement, and the public on missing and sexually exploited child issues. To further our mission to help find missing children, reduce child sexual exploitation, and prevent future victimization, NCMEC operates the CyberTipline and Child Victim Identification Program.
NCMEC makes information submitted to the CyberTipline and Child Victim Identification Program available to law enforcement and
also uses this information to help identify trends and create child safety and prevention messages. As a clearinghouse, NCMEC also works with Electronic Service Providers, law enforcement and the public in a combined effort to reduce online child sexual abuse
images. NCMEC performs its programs of work pursuant to its own private mission and independent business operations. NCMEC does not act in the capacity of or under the direction or control of the government or law enforcement agencies. NCMEC does not investigate and cannot verify the accuracy of the information submitted by reporting parties.

Contents

Section A: Reported Information

The following information was submitted to the CyberTipline by the Reporting Person or Reporting ESP. The information appearing in Section A is information received in the original submission. The reporting of information in Section A, other than the "Incident Type" and "Incident Time," is voluntary and undertaken at the initiative of the Reporting Person or Reporting ESP.

Reporting Electronic Service Provider (ESP)

Submitter:

GoogleEmail: USlawenforcement@google.com
Google Reviewer http://www.google.com/transparencyreport/userdatarequest
Business Address: s/legalprocess/
1600 Ampitheater Parkway
Mountainview, CA 94043 United States

Point of Contact for Law Enforcement:

Company Information

Google identifies and reports child sexual abuse imagery in accordance with federal statutory definition of child pornography as referenced in 18 USC 2256.

Generally, if a report does not contain an upload IP, it may be because Google does not have the information accessible at the time of the report. To the extent that Google does have a record of the upload IP, it can be disclosed in response to valid legal process.

Account information such as IP addresses and user provided information (such as SMS number or secondary e-mail address or Date of Birth) may change over time and/or reflect user provided data that may be inaccurate. This information should be investigated and confirmed independently.

Any account reported in this Cybertip may or may not be disabled in the course of making the report.

For information on which legal process is required for various types of data associated with Google's products and services, please refer to our Transparency Report available here http://www.google.com/transparencyreport/userdatarequests/legalprocess/.

If you intend to seek additional data from Google, please reference this and any related CyberTip number(s) prominently in your legal process.

With respect to the portion of this CyberTip containing the heading: "Was File Reviewed by Company?", when Google responds "Yes" it means the contents of the file reported were viewed by a person concurrently to or immediately preceding the sending of the CyberTip. When Google responds "No", it means that while the contents of the file were not reviewed concurrently to making the report, historically a person had reviewed a file whose hash (or digital fingerprint) matched the hash of the reported image and determined it contained apparent child pornography.

Incident Information

Incident Type: Child Pornography (possession, manufacture, and distribution)

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

Incident Time:	11-24-2020 12:37:25 UTC
Description of Incident Time:	The incident date refers to the approximate date and time Google became aware of the reported material.

Suspect

Name:	anum khan
Mobile Phone:	+923433490554 (Verified 10-17-2020 15:13:23 UTC)
Mobile Phone:	+923170271901 (Verified 10-20-2020 10:01:13 UTC)
Email Address:	anumkhan107@gmail.com (Verified)
Email Address:	nasirabdulqadir53@gmail.com

Additional Information Submitted by the Reporting ESP

Google became aware of the reported content which was stored in Google Photos infrastructure

Uploaded File Information

Number of uploaded files: 9

Uploaded File Information

Filename:	Google-CT-RPT-b47240fc8a9d8b870e80ba0145bef41c-VID-20201107-WA0063.mp4
MD5:	4e2fdc1f653375be46bcfb19b5895d34
Original Filename Associated with File:	VID-20201107-WA0063.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	No
Image Categorization by ESP: (See Section B for further explanation)	B1

Source Information:

Type	Value	Event	Date/Time
IP Address	116.0.56.74	Upload	11-15-2020 12:17:48 UTC

Uploaded File Information

Filename:	Google-CT-RPT-8cdcd287102a25c1a17607ea7d00bb29-VID-20201023WA0067.mp4
MD5:	702ee0108cd5e41ff399417cada94f8a
Original Filename Associated with File:	VID-20201023-WA0067.mp4
Did Reporting ESP view entire contents of uploaded file?	Yes
Did Reporting ESP view the EXIF of uploaded file?	(Information Not Provided by Company)
Were entire contents of uploaded file publicly available?	No
Image Categorization by ESP: (See Section B for further explanation)	B1

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

Source Information:

Type	Value	Event	Date/Time
IP Address	111.119.183.0	Upload	11-10-2020 15:56:00 UTC

Uploaded File Information

Filename:	Google-CT-RPT-07db9cf3f71aac4f391646023f6bdf8c-VID-20201106-WA0019.mp4
MD5:	5504ae6c210dee2812828b0f1177781d
Original Filename Associated with File:	VID-20201106-WA0019.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	No
Reported File Tags:	Potential Meme
Additional Information:	This file is animated, illustrated or appears to be intended for comedic effect.

Source Information:

Type	Value	Event	Date/Time
IP Address	116.0.56.74	Upload	11-15-2020 12:16:35 UTC

Uploaded File Information

Filename:	Google-CT-RPT-14b05f0c199203cd46f3d474e9dcec45-VID-20201025-WA0008.mp4
MD5:	0fcc0c9194170574172ced451ae92d9a
Original Filename Associated with File:	VID-20201025-WA0008.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	(Information Not Provided by Company)
Image Categorization by ESP: (See Section B for further explanation)	B1

Source Information:

Type	Value	Event	Date/Time
IP Address	203.101.162.58	Upload	11-15-2020 11:47:20 UTC

Uploaded File Information

Filename:	Google-CT-RPT-e617d588ed1dfc55a26562bb0a0cd52e-VID-20201029WA0003.mp4

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

MD5:	dbc02cf4b3ba77c3232c4938eae3fe3f
Original Filename Associated with File:	VID-20201029-WA0003.mp4
Did Reporting ESP view entire conte	No
Were entire contents of uploaded fil l publicly available?	No
Additional Information:	Potential Meme This file is animated, illustrated or appears to be intended for comedic effect.

Source Information:

Type	Value	Event	Date/Time
IP Address	111.119.183.0	Upload	11-10-2020 16:07:48 UTC

Uploaded File Information

Filename:	Google-CT-RPT-cd473ee9d87f8262ca172d90e94a6373-VID-20201030-WA0023.mp4
MD5:	7e6de7983a57a5fbdbd636747997af6d
Original Filename Associated with File:	VID-20201030-WA0023.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	No
Image Categorization by ESP: (See Section B for further explanation)	B1

Source Information:

Type	Value	Event	Date/Time
IP Address	116.0.56.74	Upload	11-15-2020 12:27:37 UTC

Uploaded File Information

Filename:	Google-CT-RPT-dff3f7c9c6fd36cf4b29a6456f192772-VID-20201112-WA0072.mp4
MD5:	23ad39a15caef5a0fab77d63268a68cd
Original Filename Associated with File:	VID-20201112-WA0072.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	No
Image Categorization by ESP: (See Section B for further explanation)	A1

Source Information:

Type	Value	Event	Date/Time
IP Address	116.0.56.74	Upload	11-15-2020 12:26:07 UTC

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

Uploaded File Information

Filename:	Google-CT-RPT-69c0a0351a85ce6148b53a36e00709d5-VID-20200925WA0123.mp4
MD5:	3fc8b31d5c58b6f1a23c394a887508c3
Original Filename Associated with File:	VID-20200925-WA0123.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	(Information Not Provided by Company)
Image Categorization by ESP:	B1

(See Section B for further explanation) Source Information:

Type	Value	Event	Date/Time
IP Address	111.119.183.36	Upload	11-09-2020 15:49:58 UTC

Uploaded File Information

Filename:	Google-CT-RPT-2b279747c0b39d0eb39bcdd4e0347aa5-VID-20201023-WA0024.mp4
MD5:	f1ea8fb573f451ca14c9eaefcb110a1d
Original Filename Associated with File:	VID-20201023-WA0024.mp4
Did Reporting ESP view entire contents of uploaded file?	No
Were entire contents of uploaded file publicly available?	No
Image Categorization by ESP:	B1

(See Section B for further explanation) Source Information:

Type	Value	Event	Date/Time
IP Address	111.119.183.0	Upload	11-10-2020 15:54:55 UTC

This concludes Section A. All of the information in this section was submitted electronically to the CyberTipline by the Reporting Person, NCMEC Call Center or Reporting ESP. The information appearing in Section A is information received in the original submission. The reporting of information in Section A, other than the "Incident Type" and "Incident Time," is voluntary and undertaken at the initiative of the Reporting Person or Reporting ESP.

Section B: Automated Information Added by NCMEC Systems

Upon receipt of a CyberTipline report, NCMEC Systems may conduct automated processes on the information submitted in Section A. The information found in Section B of this CyberTipline Report has been automatically generated by NCMEC Systems. If the CyberTipline Report was submitted by a member of the public, Section B will be blank.

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

Explanation of Automated Information (in alphabetical order)

Geo-Lookup: When a Reporting ESP voluntarily reports an IP address for the "Suspect," NCMEC Systems will geographically resolve the IP address via a publicly-available online query. The results of this lookup are displayed.

Geolocation data is approximate and may not display a user's exact location. Please be aware that the geolocation information provided is not exact but is providing a reliable estimate of location based on IP address(es) voluntarily provided by the reporting ESP.

Further Information on Uploaded Files

Number of uploaded files in each categorization category:

IP Address	Country	Region	City	Metro Area	Postal Code	Area Code	Lat/Long	ISP/Org
203.101.162.58	PK	PB	Lahore		54000		31.4888/74.3686	Cyber Internet Services (Pvt)/ Cyber Internet Services (Pvt)

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

116.0.56.74	PK	SD	Karachi	12311	24.9207/ 67.0657	Telecard Limited/ Telecard Limited

A1: 1

B1: 6

The following categorization system was created by various ESPs in January 2014:

	Content Ranking	1	2
A	Prepubescent Minor	A1	A2
B	Pubescent Minor	B1	B2

Rank	Term	Definition
1	Sex Act	Any image of sexually explicit conduct (actual or simulated sexual intercourse including genital-genital, oral-genital, anal-genital, or oral-anal whether between person of the same or opposite sex), bestiality, masturbation, sadistic or masochistic abuse, degradation, or any such depiction that lacks serious literary, artistic, political, or scientific value.
2	Lascivious Exhibition	Any image depicting nudity and one or more of: restraint, sexually suggestive poses, focus on genitals, inappropriate touching, adult arousal, spreading of limbs or genitals, and such depiction lacks serious literary, artistic, political, or scientific value.

Geo-Lookup (Uploaded Files)

111.119.183.0	PK	SD	Karachi	12311	24.9207/ 67.0657	CMPak Limited/ CMPak Limited
111.119.183.36	PK	SD	Karachi	12311	24.9207/ 67.0657	CMPak Limited/ CMPak Limited

Auto Refer

Date: 11-25-2020 12:33:34 UTC The reported IP address appears to originate in a foreign country in which there is a designated law enforcement agency with a VPN connection to NCMECs CyberTipline. In accordance with current ECD policy, this report will be referred automatically to the designated international law enforcement agency for potential evaluation, investigation, and/or prosecution. Law enforcement should review all information and images associated with this report before determining how to proceed. Law enforcement seeking further information relating to this report should submit a request to ecuassistance@ncmec.org.

This concludes Section B

This Report is provided solely for informational purposes pursuant to NCMEC's nonprofit mission. Please treat all information in this Report as confidential.

Section C: Additional Information Provided by NCMEC

Section C contains information collected by NCMEC staff based on the information electronically submitted by the Reporting Person, NCMEC Call Center or Reporting ESP. Section C may contain a variety of additional information, including data gathered from queries on publicly-available, open-source websites. Any queries conducted by NCMEC staff will be documented and any query results will be saved to the electronic filing system when possible. The CyberTipline cannot confirm the accuracy of information found in public records or whether the results are affiliated with any parties relating to this report.

NCMEC Priority Level:	E (Report submitted by a registered Electronic Service Provider)
NCMEC Classification:	Auto-referred International
	Files Not Reviewed by NCMEC
International Country:	Pakistan
NCMEC Date Processed:	11-25-2020 12:33:34 UTC
Made Available to Law Enforcement by NCMEC:	Yes

Uploaded File Information

Files Not Viewed by NCMEC:

NCMEC staff have not viewed the following uploaded files submitted with this report and have no information concerning the content of the uploaded files other than information voluntarily provided in the report by the reporting ESP.

Files Not Viewed by NCMEC

Filename	MD5
Google-CT-RPT-b47240fc8a9d8b870e80ba0145bef41c-VID-20201107-WA0063.mp4	4e2fdc1f653375be46bcfb19b5895d34
Google-CT-RPT-8cdcd287102a25c1a17607ea7d00bb29-VID-20201023-WA0067.mp4	702ee0108cd5e41ff399417cada94f8a
Google-CT-RPT-07db9cf3f71aac4f391646023f6bdf8c-VID-20201106-WA0019.mp4	5504ae6c210dee2812828b0f1177781d
Google-CT-RPT-14b05f0c199203cd46f3d474e9dcec45-VID-20201025-WA0008.mp4	0fcc0c9194170574172ced451ae92d9a
Google-CT-RPT-e617d588ed1dfc55a26562bb0a0cd52e-VID-20201029-WA0003.mp4	dbc02cf4b3ba77c3232c4938eae3fe3f
Google-CT-RPT-cd473ee9d87f8262ca172d90e94a6373-VID-20201030-WA0023.mp4	7e6de7983a57a5fbdbd636747997af6d
Google-CT-RPT-dff3f7c9c6fd36cf4b29a6456f192772-VID-20201112-WA0072.mp4	23ad39a15caef5a0fab77d63268a68cd
Google-CT-RPT-69c0a0351a85ce6148b53a36e00709d5-VID-20200925-WA0123.mp4	3fc8b31d5c58b6f1a23c394a887508c3
Google-CT-RPT-2b279747c0b39d0eb39bcdd4e0347aa5-VID-20201023-WA0024.mp4	f1ea8fb573f451ca14c9eaefcb110a1d

This concludes Section C

If you need further information regarding the contents of this Report, please contact the CyberTipline at ecuassistance@ncmec.org or 1-877-446-2632, ext. 6702.

For more information regarding images containing identified child victims, please contact the Child Victim Identification Program (CVIP) at cvip@ncmec.org.

Section D: Law Enforcement Contact Information

The report was made available to the Law Enforcement Agency listed below.

Federal Investigation Agency of Pakistan

Investigator:

Assigned Officer:	ACCESS VPN
Title:	Assistant Director Imran Haider
City/State:	Islamabad
Country:	Pakistan
Phone Number:	92-342-4177770
Email Address:	imran.haider@nr3c.gov.pk,mohsin.ali@nr3c.gov.pk

Time/Date was made available: 11-25-2020 12:33:34 UTC

This concludes Section D

This concludes CyberTipline Report 83113748