DIGITAL ADVERTISING FRAUD PREDICTION USING OLS REGRESSION

NG KAH MUN

BACHELOR OF INTERNATIONAL BUSINESS (HONS)

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF ACCOUNTANCY AND MANAGEMENT
DEPARTMENT OF INTERNATIONAL BUSINESS

JUNE 2023

DIGITAL ADVERTISING FRUAD PREDICTION USING OLS REGRESSION

BY

NG KAH MUN

A final year project submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF INTERNATIONAL BUSINESS (HONS)

UNIVERSITI TUNKU ABDUL TAHMAN

FACULTY OF ACCOUNTANCY AND MANAGEMENT
DEPARTMENT OF INTERNATIONAL BUSINESS

JUNE 2023

DECLARATION

I hereby declare that:

(1) This undergraduate FYP is the end result of my own work and that due acknowledgement has been given in the references to ALL sources of information be they printed, electronic, or personal.

(2) No portion of this FYP has been submitted in support of any application for any other degree or qualification of this or any other university, or other institutes of learning.

(3) Sole contribution has been made by me in completing the FYP.

(4) The word count of this research report is ____11855__________________.

Name of student:	Student ID:	Signature:
Ng Kah Mun	1903811	

Date: 5 May 2023

ACKNOWLEGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this research project.

First and foremost, I would like to express my sincere gratitude to my research supervisor, Dr. Seah Choon Sen, for his important advice, constant support, and inspiring words. My research strategy and direction have greatly benefited from their advice and input.

I want to express my sincere gratitude to Ms. Tai Lit Cheng, my second examiner, who genuinely gives me guidance when my research is lacking. This study would not have been feasible without their openness to share their opinions and experiences.

On the other hand, I am also grateful to Universiti Tunku Abdul Rahman (UTAR), which provided me with access to necessary resources and facilities that made this research possible. I would like to acknowledge the contribution of my friends who have provided me with valuable feedback and suggestions during the course of this research.

Lastly, we would like to express our appreciation to our families for their unwavering support and understanding throughout this project. Thank you all for your valuable contributions.

DEDICATION

I would like to dedicate this Final Year Project to my respected supervisor, Dr Seah Choon Sen, and my respected second examiner, Ms Tai Lit Cheng, who gave me valuable feedback, guidance, insight and encouragement throughout my Final Year Project Journey. I would like to express my gratitude to my beloved family members, friends, lecturers, seniors, and each respondent who assisted me a lot in my Final Year Project. Without all of your supports, encouragements and assistances, it would be challenging for me to accomplish this Final Year Project.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

PREFACE

Every student must complete the Universiti Tunku Abdul Rahman (UTAR) Final Year Project 'UKMZ2016 Research Project' in order to receive a degree for Bachelor of International Business (Honours). As a Bachelor of Science student concentrating in International Business, the knowledge and good awareness of global concerns is a significant source of competitive advantage. However, international business is a complicated and broad topic with numerous dimensions. Among the numerous global concerns, digital advertising conversion fraud, a contemporary issue that every organisation faces, has been identified as one important trend that has severely disrupted the digital advertising landscape. It is attracting more and more firms to invest in digital advertisement due to the effectiveness and efficiency in terms of reaching target audience and cost savings. However, the problem of conversion fraud is yet to emphasize among the firms who just started to invest in digital advertisement. Therefore, the author is inspired to explore the importance of the problem of conversion fraud.

ABSTRACT

Digital advertising has become the essential tools for every business. The digital advertising budget is increasing over the years. There are more businesses transform to digital advertising during the pandemic, however, the beginner of advertiser might lack of knowledge on digital advertisement at the same time pouring extra capital into digital advertising. Furthermore, the fraudulent activity in digital advertisement is also increasing which is harming the current digital marketing environment as well as every party involved. This study examines the factors affect conversion fraud in digital advertising. A sample of 956 observation of computing generate data is used to examine the variables towards conversion fraud. The result shows that advertiser, ad log, item, goal, and ad slot have positive relationship towards conversion fraud.

Chapter 1: RESEARCH OVERVIEW

1.0 Introduction

This research is designed to predict digital advertising conversion fraud for business to avoid conversion fraud and improve digital advertising environment. In this chapter 1, background, problem statement, research questions, research problems, research objectives, research significant, are covered.

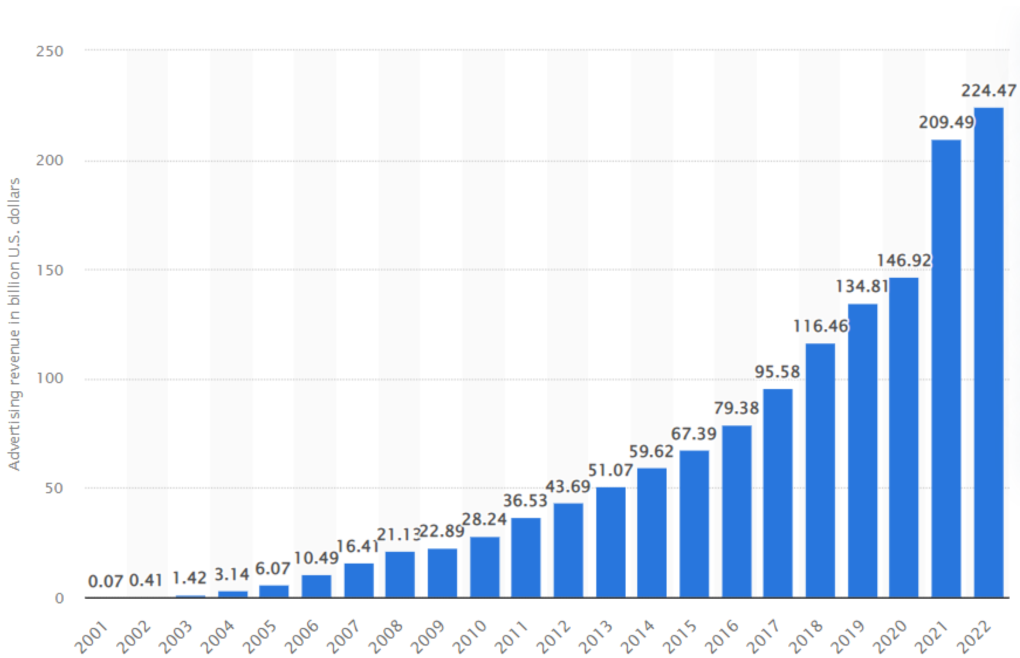
1.1 Background of study

Advertisement is an important tool to every business, as it is the main tool to get exposure to their target customer. From traditional advertisement such as printed magazines, newspapers, and television, until today, digital advertisement has specific places are assigned to be sold for advertisement based on users search results, posts, or the content of the website (Gharibshah, & Zhu, 2021). Nowadays, people are heavily relying on internet to get information and thus, digital advertisement is widely used by businesses to promote their goods and services. Digital advertisement can help business to promote their goods and services in efficient and effective way. Advertising serves as a way of communication to inform and persuade audience to make purchasing decision regarding a goods or services (Wuisan, & Handra, 2023).

There are different stakeholders act in the background in order to present advertisement to internet users via numerous platforms, including search engines, news sites, social networks, all of these platforms could use for display and deliver advertisement (Gharibshah, & Zhu, 2021). However, when business start to pour money into digital advertising, business might face losses due to conversion fraud. Conversion fraud is an action fraud that taken by the users which does not bring value to advertiser and causing misleading information to advertiser (Zhu *et al*, 2017). Conversion fraud can cause an increase in advertising budget for advertiser (Sadeghpour, & Vlajic, 2021). It could also cause advertiser get charge even it does not bring advertiser any valuable data, at the same time, the measurability of the charge to advertiser is unclear.

Google as the largest search engine website, the main revenue of google is through advertisement. Based on Bianchi, 2023, the annual advertising revenue is rising every year from 2001 to 2022, in the latest data, Google's ad revenue amounted to $224.47 billion. Google generates advertising revenue from advertiser through Google Ads platform and search advertising is the mainstream of Google advertising revenue.

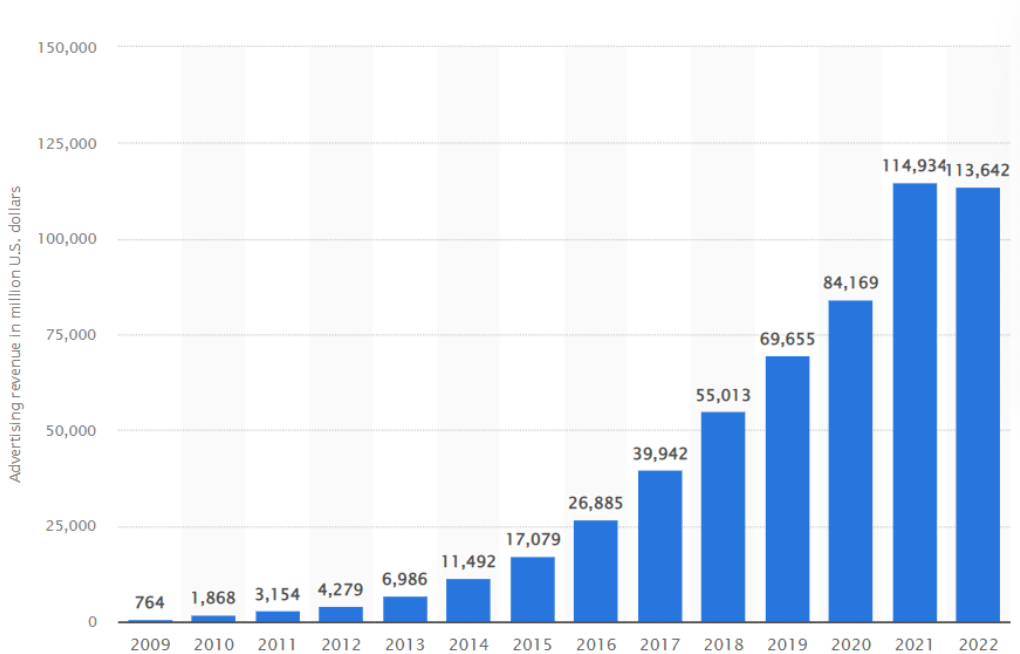
Figure 1.1 (I): Advertising revenue of Google from 2001 to 2022 (in billion USD)



Source: Statista (2023)

On the other hand, another platform which is also a powerful tool for advertiser to publish their ad is Meta platform. Meta, (formerly known as Facebook Inc.) owned Facebook and Instagram which both social media platforms have the largest number of users compare to other social media platforms. Meta business model is immensely relying on advertising revenue, and it is also the main source of Meta income. The chart shown the advertising revenue of Meta from 2009 to 2022, the number is increasing year by year, and the highest advertising revenue is $114,934 million in the year 2021.

Figure 1.1 (II): Annual advertising revenue of Meta Platforms worldwide from 2009 to 2022 (in million USD)



Source: Statista (2023)

As Figure 1 and 2 has shown that the advertising revenue of Google and Meta platforms has risen sharply in the 2021, which is also during the outbreak of pandemic. This data also indicates that more and more businesses is pouring their money into digital advertising platforms.

Prediction and measurable of user response are important for every advertiser since it could relate to recommendation systems to understand user’s action such as purchasing a product or subscribing a service (Gharibshah, & Zhu, 2021). Prediction in digital advertising plays a significant role to advertiser to have more detailed information on the interest of target audience and thus, maximize the effectiveness of advertisement with minimize budget of advertisement.

1.2 Problem Statement

In recent year, digital advertising has turn to one of the most prominent and profitable form of marketing for many businesses. Due to the trend of digital advertising, unethical problem occurs which cause to unhealthy digital advertising environment, that is conversion fraud (Sadeghpour, & Vlajic, 2021).

Conversion fraud is a potential threat which could not be control and measure by advertiser. It brings negative impact to advertisers in terms of cost and data information. This also indirectly affect customer's response to the advertisement. Conversion fraud can be cause by few factors, such as unethical website publisher, and users' actions by accident. unethical website publisher behavior affects the digital advertising environment and cause unfair to advertiser (Sadeghpour, & Vlajic, 2021); publisher charging advertiser for every user's action, such as click, view, and leads, but users' action by accident brings zero value and increasing advertiser budget. On the other hand, the **failure of conversion** is out of advertiser's control. It might describe to user's action increase advertiser's budget but did not convert to sales. For example, users accidently click on an ad, which does not convert sales to advertiser but increase advertiser's budget (Daswani, *et al*, 2008).

Due to the outbreak of Covid-19 pandemic, **more and more firms are switching their businesses from traditional business model to digitalization**. Since physical operation mode has been forced to shut down temperately and also permanently, in order to survive during the downward economic, many businesses started to switch to e-commerce and invest on digital marketing. Businesses advertised through internet such as social media, and other digital platforms to increase conversion that led to revenue. However, firms might face many issues if firms do not aware of the challenges of digital advertising.

Firms might **lack of knowledge on digital advertising** and pouring too much money in advertising that become expenses instead of investment. Intellectual technology (IT) knowledge is important for ability of a company to adopt or not adopt ICT and e-commerce among Small and Medium Enterprises (SEMs). The research found that IT and e-commerce knowledge among employees has significant impact of the use of ICT and e-commerce by SMEs (Religia *et al*, 2021). When more businesses switch to digital business model, the more money flow in digital platforms, and the misuse of digital advertising might occur more often, that is conversion fraud.

1.3 Research Question

In the light of aforementioned problem statement, the present study aims to answer the following research question:

1. What are the factors that led to conversion fraud?
2. How would conversion fraud affect advertiser?

1.4 Research Objectives

With the research questions raised, the present study aims to determine the most significant variables

1. To identify the factors that led to conversion fraud in digital advertising.
2. To propose the prediction framework in determining the conversion fraud.

1.5 Research Significance

This study could provide an insight of conversion fraud in digital advertising to advertiser, regulators, and future academic researchers.

Advertiser is the party who create advertisement to promote their product or service to target audience through invest in digital advertising. However, advertiser is the victim in the cases of conversion fraud as it has direct impact on advertiser. Conversion fraud creates fake information and lead to costly advertising budget to advertiser. Hence, it is necessary to provide the literature review of digital advertising to advertiser to understand the digital advertising and conversion fraud.

Moreover, prediction of conversion fraud could enforce rules and regulation. It helps **regulators** identify potential risks and develop targeted regulations since the rules and regulations for digital advertising environment is lack of completeness and it falls into a legal grey area (Lim, 2020). According to Henk concept, digital world consists of networks and interconnected, however, digital law seems to be lack of frontiers and jurisdiction. While the concept of good governance is important and necessary to be evaluate the improvement of digital law and society. The digital law is lack of implementation in terms of the principles of properness, transparency, accountability, and human right (Marwan, & Bonfigli, 2022).

Apart from that, it contributes a cross-disciplinary study to **future academic**. A cross-disciplinary study which combine knowledge of more than one academic disciplines from varies field of study. While this study involving business and computer science that has the potential to contribute to the development of solutions in complex problems in various fields, including business, technology, and society. It can be very beneficial in many ways such as leading to a more comprehensive understanding, innovative solution to complex problem, and broader perspectives and challenges assumption. It can be a powerful tool for researchers as it allows them to generate new insights and develop innovative solutions that have practical benefits.

1.6 Chapter Summary

In chapter 1, it provides an overview of digital advertising and the existing problems in the digital advertising environment. There is no doubt of that, digital advertising has become a significant topic among the business environments. Most of the firms nowadays are highly relying on digital advertisement however, some might lack of the knowledge on related fields which lead firm less and increase their budget.

Chapter 2: LITERATURE REVIEW

2.0 Introduction

In this chapter, the variables that cause to conversion fraud will be covered. Prediction model is applied to investigate the variables that effect towards digital advertising fraud.

2.1 Digital Advertising

2.1.1 Overview of Digital Advertising

Digital advertising is a fast-paced environment, it occurs in the year 1995, however, the market structures are still changing rapidly until today (Gordon *et al,* 2021). Digital advertising, so-called as internet advertising comprising online advertising and mobile advertising, impact to monetization throughout the internet environment (Chen *et al,* 2016). The internet, social media, mobile apps, and others digital communication technologies has become an essential activity in part of billions of people's daily routine (Stephen, 2016). In today, more and more retailer using social networking sites (SNS) to run business and it create a platform that allow online shopper a convenient way to shop a variety of product and services (Hyun *et al,* 2022). Digital and social media marketing is a new industry-led, research-informed and result driven drive to e-commerce (Heinze *et al* 2016). Digital marketing has become increasingly important to firms because it promotes to the development of a new avenue of reaching target users at the same time satisfy target users with products and services (Patil *et al* 2022). Digital Advertising is one of the major topics under the big marketing umbrella, it studies how consumers react to various aspects of advertising. For instance, consumers purchase intention or behavior will be based on the other consumers' opinions, who purchased a certain brand before, that is review, or friends on social media (Stephen, 2016). The basic concept of digital advertising is to sell space on web pages and apps for advertising (Shaari, & Ahmed, 2020).

Digital advertising is important and useful to business as it offers exceptional innovations to marketers and allow business to advertise to specific target market which best match their target audience. Users can see the advertising that meet their interest and area. Digital advertising offer business a greater adaptability of customer response. Since the digital environment allow the advertisement to spread efficiently, the users could response to advertisement immediately (Ma, & Du, 2018).

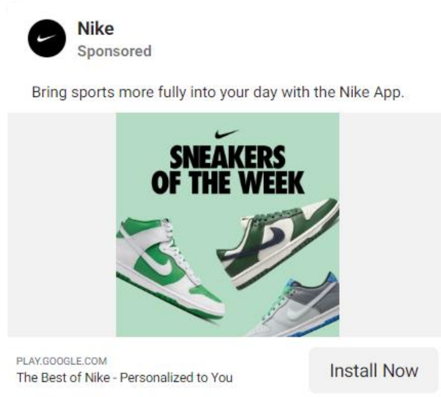
2.1.2 Types of Advertising

2.1.2.1 Meta Ad Format

I. Image and Video

Image and video ad are the most common ad use by advertiser, due to it provide visual aid at the same time its production cost is cheaper. Video movement of ads can be attractive and more eye-catching in social media platform to target users. Advertiser run image or video ad to show off their products, service, or brand in visualization, it able to capture user's attention in order to convey simple message such as encourage user visit advertiser website (Meta, n.d.).

Figure 2.1.2.1 (I): Example of Image and Video Ad



Source: Nike

II. Carousel

Carousel format enables advertisers to display multiple images and/or videos within a single ad, with each item having its own headline, description, link, and call to action. Users can scroll through the carousel by swiping on their mobile device or clicking the arrows on their computer screen. Carousel ad usually used to feature multiple products that link to different landing pages, highlight multiple features of one single product. Due to its scrolling characteristic, some used it to tell a story or process by showing multiple images (Meta, n.d.).

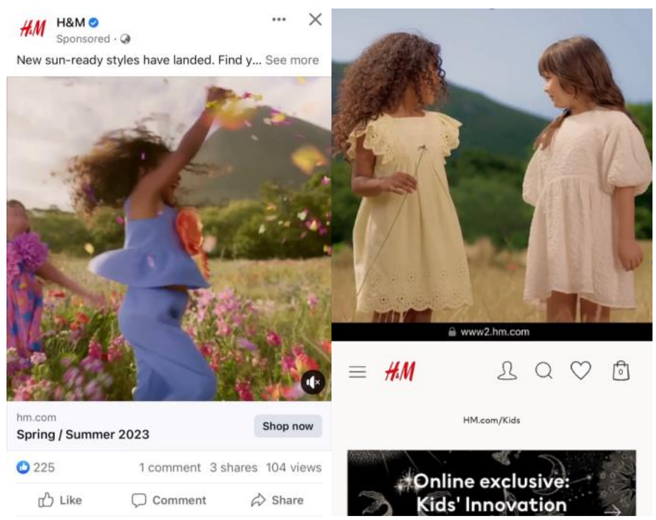
Figure 2.1.2.1 (II): Example of Carousel Ad

Source: Puma

III. Instant Experience

Instant Experience is a full-screen experience which opens once user click on ad on a mobile device. Its able advertiser to highlight their product, service or brand visually. Through instant experience, it able to capture audience attention as it loads instantly and expand into full-screen visualization. It also allows advertiser to include more content to further develop advertiser message by connecting multiple instant experience ads. Moreover, it could combine all ad formats in one single instant experience ad, which user can watch video, scroll carousel, tap buttons that bring user to advertiser websites (Meta, n.d.).

Figure 2.1.2.1 (III): Example of Instant Experience Ad

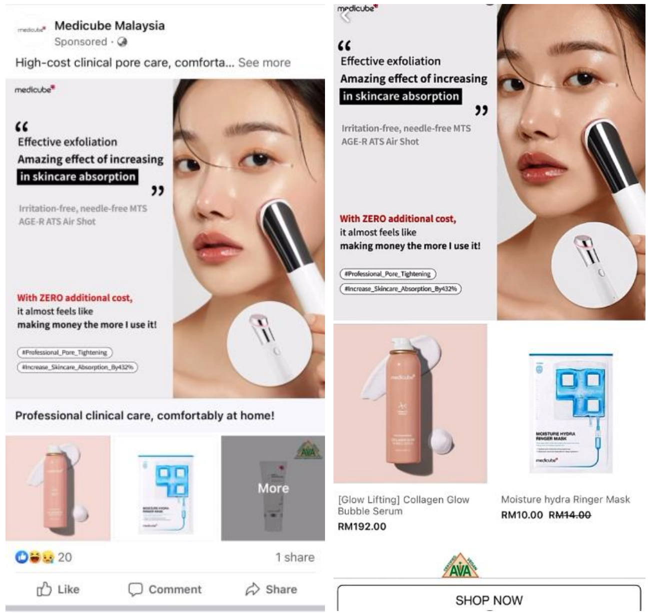


Source: H&M

IV. Collection

Collection ad is quite similar with instant experience ad, it shows in full-screen experience once users tap on the ad. It is an easier way for users to browse through products or services in mobile device with more visual and immersive way. By using collection ad, it allows users to browse multiple products in a single ad and each with its own link (Meta, n.d.).

Figure2.1.2.1 (IV): Example of Collection Ad



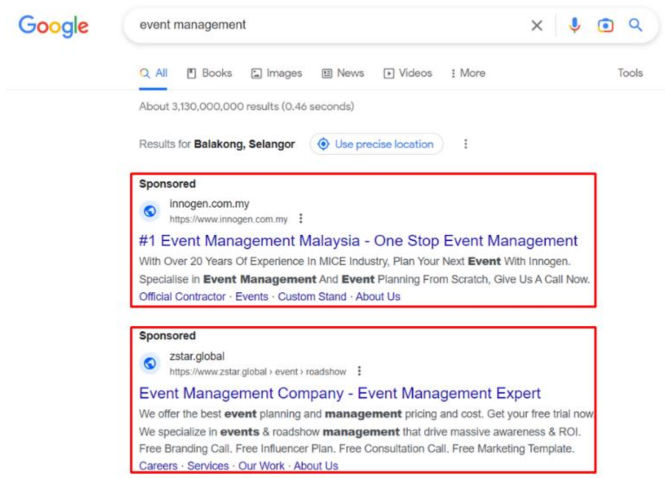
Source: Medicube

2.1.2.2 Google Ad Format

I. Text

Ad only show in words only, which usually used for Google search engine. The main objectives by using text ad format, it able to pops up quickly and easily to reach users. When users search by using search bar, it shows the most related result through text ad with the word "Sponsored" (Google Ads Help, n.d.).

Figure 2.1.2.2 (I): Example of Text Ad

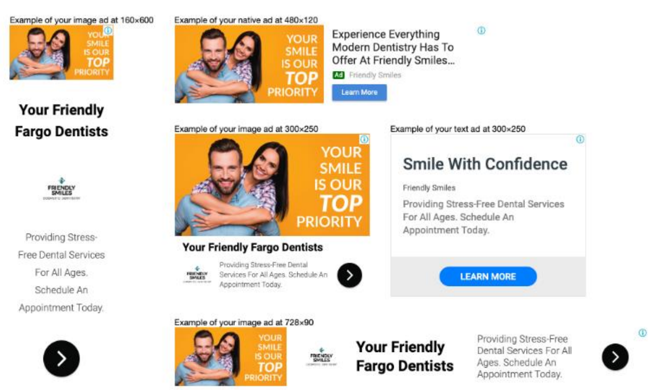


Source: Google

II. Responsive

Responsive ad can be either only text or images, and its flexibility in automatically adjust the ad size, appearance, and format to fit in any ad space. It usually uses in search and display campaign types (Google Ads Help, n.d.).

Figure 2.1.2.2 (II): Example of Responsive Ad



Source: Google

III. Image

Image ad not merely limit a static image, it includes animated image. The file format supported are JPG, PNG, and GIF. Image ad provide featuring details of advertiser brand, product and service. It widely uses for display ad, which appear on webpages in the display network, that is a network of million webpages which allow advertiser to display their ad on (Google Ads Help, n.d.).

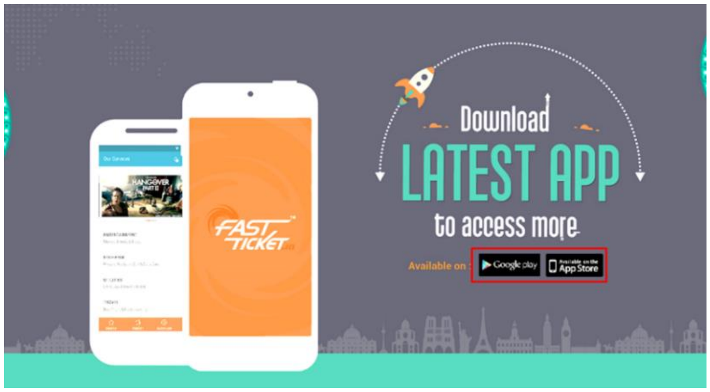
Figure 2.1.2.2 (III): Example of Image Ad

Source: Tabung Pendidikan

IV. App promotion ads

Ad used to drive app downloads and engagements. It normally creates for app campaign, with button shows “Available on Google App and App Store”. It includes a deep link that bring users straight into advertiser app, once user click on their ads (Google Ads Help, n.d.).

Figure 2.1.2.2 (IV): Example of App Promotion Ad

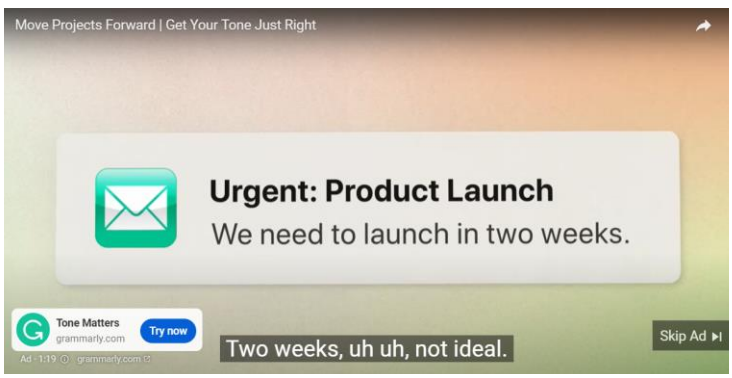


Source: Google

V. Video

Video ad aims to deliver rich and engaging experience to users. It run individual video ads or insert in streaming video content. For example, the video ad on YouTube before the video content user wish to watch (Google Ads Help, n.d.).

Figure 2.1.2.2 (V): Example of Video Ad

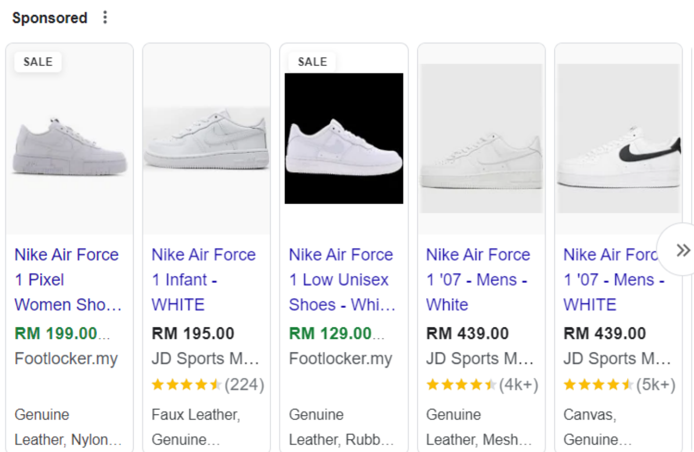


Source: Grammaly

VI. Shopping ads

Shopping ads show users the image of the products with additional details such as title, price, store location, and website. It shows up the most relevant products, with the word "Sponsored" that users search by using the search bar (Google Ads Help, n.d.).

Figure 2.1.2.2 (VI): Example of Video Ad

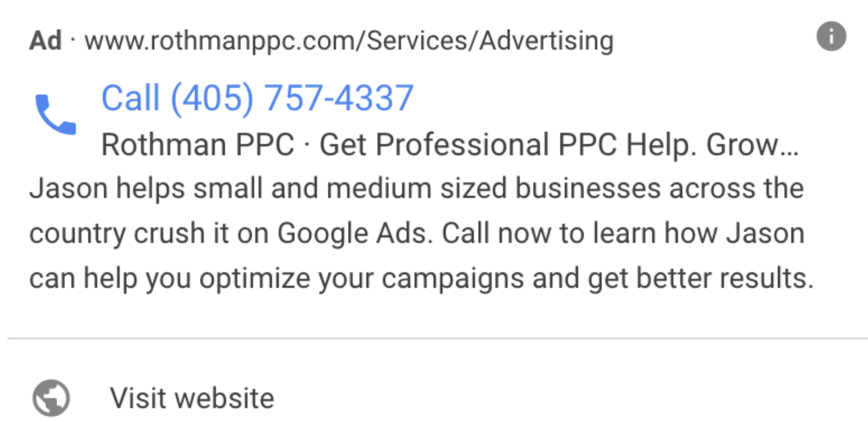


Source: Google

VII. Call-only ads

Ad that drives calls to advertiser which includes business phone number. This ad usually show up when user search for specific word, it appears on mobile device that allow users to make phone call directly (Google Ads Help, n.d.).

Figure 2.1.2.2 (VII): Example of Call-only Ad



Source: Google

2.1.3 Digital Advertising Revenue

Digital advertising is a financial pillar that supports both free Web content and services, and free mobile apps. Both web and mobile advertising use a similar infrastructure: the ad library embedded in the web page or mobile app fetches content from ad providers and displays it on the web page or the mobile app's user interface. The ad provider pays the developer for the ads displayed (impressions) and the ads clicked (clicks) by the user. Because web and mobile advertising use a similar infrastructure, they are subject to the same security concerns, such as tracking and privacy infringements (Crussell, Stevens & Chen, 2014).

2.1.3.1 Cost-Per-Click (CPC)

CPC, also known as pay-per-click (PPC), is one of the most common costs which able to apply to all types of ads. CPC refers to a payment method where advertiser is charged for every click action taken by users on ads. In this type of bidding campaign, advertiser usually establish a maximum cost-per-click bid, also known as "max. CPC," which is the highest amount advertise are willing to pay for each click on your ad. Once it has reach max. CPC, publisher is not allowed to charge more than the maximum CPC bid amount that advertiser set (Google Ads Help, n.d.).

2.1.3.2 Cost-Per-Thousand Impression (CPM)

Cost-per-thousand impressions (CPM) is a bidding method that determines the cost an advertiser pays for every one thousand impressions or views of their ad on a publisher's website. An impression refers to each time an ad appears on a webpage, whether the user interacts with it or not. CPM is often used in display advertising, where advertisers pay publishers for ad space on their websites. By tracking CPM, advertisers can measure the effectiveness of their ad campaigns and calculate the return on investment (ROI) for each campaign. This information can help them in decisions making about future advertising efforts and optimize their strategies for better performance. For instance, if an advertiser runs a CPM campaign and pays $10 per thousand impressions, and their ad receives 50,000 impressions, they would pay $500 for the campaign. If they generate $600 in sales from those impressions, they have achieved a positive ROI. However, if they only generate $300 in sales, their ROI would be negative, and they would need to reevaluate their advertising strategy (Google Ads Help, n.d.).

.

2.1.3.3 Cost-Per-View (CPV)

CPV, or cost-per-view, is a type of bidding method that is commonly used for video advertising campaigns. With CPV, advertisers are charged each time a viewer watches their ad. A view is typically counted when a user watches the ad for at least 30 seconds, or for the entire duration of the video if it is less than 30 seconds. Additionally, interactions such as click action on call-to-action overlays, cards, and companion banners are also counted as views. Advertisers can set a maximum CPV bid, which is the highest amount they are willing to pay for each view. This bid is used to inform the publisher website how much they will be paid for showing the ad to viewers. Tracking CPV is an essential measure tool to evaluate the performance of a video ad campaign and assess its return on investment (ROI) potential (Google Ads Help, n.d.).

2.1.3.4 Cost-Per-Action

CPA, or cost per action, is a digital advertising payment model that allows advertisers to pay only when a specific action is taken by a potential customer. The advertiser decides which action is considered valuable, such as a sale, a form submission, or a download. This model helps advertisers to focus on their desired outcomes and optimize their advertising efforts for maximum results. By tracking CPA, advertisers can calculate the cost of each conversion and measure the effectiveness of their advertising campaigns. This data can also help them adjust their strategies and allocate their advertising budget to the most profitable channels. With the rise of digital marketing, CPA has become a popular payment model for online advertising campaigns (Oberlo, n.d.).

2.2 Classification in conversion fraud

2.2.1 Pixel Stuffing

Pixel stuffing, a type of digital ad fraud, involves the insertion of multiple ads into a single pixel space, which results in advertisers being charged for impressions that are not viewable by users due to their tiny size (Zaiceva, 2022). This type of fraud is aimed at generating false impressions and making more money for fraudsters. Like cramming many people into a small space, pixel stuffing creates an environment that is difficult to navigate or observe. This fraudulent practice, also known as in-frame stuffing, is frequently employed by scammers to deceive online marketers. Advertisers are particularly vulnerable to this type of fraud because they are paying for impressions that do not result in any value or return on investment, which ultimately harms their bottom line (Suresh *et al*, 2019).

2.2.2 Ad Stacking

Ad stacking is a form of impression ad fraud where multiple ads are stacked on top of each other in the same ad slot, but only the top ad is visible to users. Despite this, all ad impressions are recorded and counted by the ad server. Fraudster use ad stacking to inflate click and impression counts, leading to higher billing for advertisers. This technique involves hiding the stacked ads from plain sight, making them invisible to naked eye, but still detectable by ad network agents. Ad stacking is commonly used in search engine cloaking to increase visibility and artificially boost metrics. Advertisers should be aware of ad stacking and take measures to prevent it from negatively impacting their campaigns (Zhu et al., 2017; Kumari et al., 2017).

2.2.3 Domain Spoofing

Domain spoofing is a type of fraud where a fraudster deceives users and advertisers by disguising their actual website link of a premium website. This type of fraud can harm not only the users but also the premium domain owners who may lose revenue due to the fraudster's activities (Zaiceva, 2022). Domain spoofing is also one of the major challenges in the digital advertising industry. Scammers take advantage of their ability to manipulate and deceive the parties involved, leading to increased cost per impression and additional expenses for advertisers. This type of fraud can result in significant financial losses for businesses, which is why it is important to take measures to prevent and detect domain spoofing (Bashir *et al*, 2019).

2.2.4 Ad Injection

Ad injection is a form of digital ad fraud that occurs when unauthorized ads are displayed on a website, without the publisher's knowledge or permission. This type of fraud can negatively impact the real ads that are paying for the publisher's ad space, as it can lead to a reduction in their performance metrics (Zaiceva, 2022). Ad injection can be facilitated by malware that serves as an ad injector and compromises users' experience, security, and privacy. By injecting ads without authorization, fraudsters can generate revenue through cost-per-click, cost-per-impression, and cost-per-acquisition models. Ad injection is a serious issue in the digital advertising industry, as it can undermine the trust between advertisers and publishers, and lead to significant financial losses for both parties (Thomas et al., 2015).

2.2.5 Click Injection

Click injection is a form of mobile ad fraud that involves fraudsters injecting fake clicks on ads to charge advertisers for clicks that never actually happened. In this scheme, malware involve triggering the fake click when a genuine ad is displayed. These fake clicks are registered by the ad network and reported to the advertiser as genuine clicks, leading to the advertiser paying for non-existent traffic. Click injection is often accomplished through junk applications that surreptitiously run in the background and generate ad requests, leading to increased ad impressions and clicks (Adjust, n.d.). This type of fraud is especially insidious because it is difficult to detect. The fake clicks generated by click injection can appear to be legitimate, and it is often challenging for advertisers to differentiate between genuine and fake clicks. Moreover, since click injection occurs on the user's device, it is impossible for the advertiser to prevent this type of fraud at the ad network level. Advertisers need to rely on fraud prevention tools that can identify and block fraudulent traffic to avoid wasting their advertising budgets on fake clicks (Zaiceva, 2022).

2.2.6 Geo Masking

Geo masking, also known as location masking, is a fraudulent technique that exploits the notion that traffic from certain regions or countries where is more valuable than others. Fraudsters use this technique to deceive advertisers by disguising low-quality traffic as high-quality traffic, which they can sell for higher price. To carry out geo masking, fraudsters alter or manipulate the information about the IP address to mislead ad platforms and advertisers about the location of the traffic. For example, they may manipulate the IP address to make it appear as though the traffic is coming from the United States, which is considered to be more valuable than traffic from certain developing countries. Advertisers can protect themselves from geo masking by using tools to verify the accuracy of IP addresses and by taking measures to prevent ad fraud in their campaigns. This is important because geo masking can result in advertisers wasting their ad spend on low-quality traffic that does not result in valuable actions or conversions (Publift, 2023).

2.2.7 Bot Fraud

Bot traffic refers to non-human traffic that is generated by software robots, commonly known as bots. Bots can be programmed to perform a variety of tasks such as crawling web pages for search engines, automating repetitive tasks, and even generating fake traffic to websites and apps. This makes bot traffic a significant threat to publishers, as it can disrupt the functioning of websites and apps through activities such as scraping, phishing, or malicious redirects. Bots can impersonate legitimate visitors and generate fake clicks on ads. This can result in the display of ads being banned on monetization programs, which can cause significant losses for publishers and advertisers alike (Zaiceva, 2022). Moreover, bots can also skew website analytics and distort data-driven insights that inform digital marketing strategies. Bots are designed to perform routine and straightforward tasks quickly and efficiently. This capability of bot is valuable due to its automation can help streamline processes and reduce costs. However, when it comes to digital advertising, the use of bots can be highly damaging and deceptive. Advertisers and publishers need to implement measures to detect and prevent bot traffic, such as using advanced analytics and ad fraud detection tools. This can help ensure that digital advertising is transparent, accurate, and delivers real value to all stakeholders (Data Dome, 2022).

2.3 Reviews of Variables

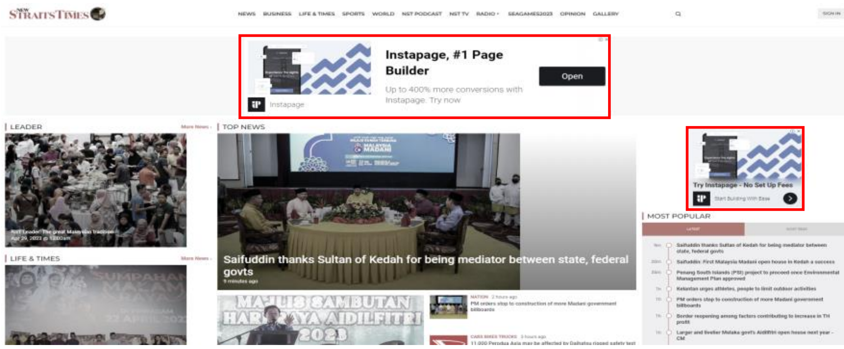
2.3.1 Advertiser

An advertiser is a critical component of the digital advertising industry. Advertisers are typically individuals, companies, or organizations who pay to have their ad for displayed ad on publishers' websites (Stone-Gross, 2011). These ads are priced based on factors such as space or time, and the fees charged are determined by the publishers. Advertisers play a significant role in digital marketing as they fund campaigns and drive traffic to their products or services. Their primary objective is to persuade users to take a specific action, such as installing an app, visiting a website, or making a purchase. Advertisers purchase ad space from publishers, which allows them to reach their target audience with the right message at the right time. For example, a gaming app may purchase ad space from a website publisher to promote a special offer to users who are interested in gaming. The success of the advertising campaign is often measured using metrics such as CPC (cost per click), CPV (cost per view), and CPM (cost per thousand impressions), which can provide insight into how users interact with the ads (Adjust, n.d.).

2.3.2 Ad Slot ID

Ad slot are critical for website publishers and advertisers as they determine where ads are displayed on a webpage. The location of an ad slot can impact the success of an ad campaign. For example, an ad placed in a prominent ad slot, such as above the fold or in the header, is more likely to be noticed by users and generate clicks than an ad placed in a less visible location, such as in the footer or sidebar. Ad slots can also be optimized for different types of ads, such as display ads or native ads, depending on the publisher's goals and audience (Google Ads Help, n.d.). Ad slots may also be sold to advertisers based on different pricing models, such as CPM, CPC, and CPV. Advertisers may compete for a particular ad slot, and the highest bidder may be awarded the slot. In the case of video ads, an ad slot can refer to a specific time segment within a video that is reserved for an advertisement. For example, a 10-second ad slot may be inserted before a video begins playing, or in the middle of a longer video. The success of a video ad campaign can be measured by metrics such as view-through rate and engagement rate. Overall, ad slots are crucial for both publishers and advertisers in ensuring that ads are displayed in the most effective and visible way to generate engagement and conversions (Driskill, n.d.).

Figure 2.3.2: Example of Ad Slot (marked as red rectangles)



Source: New Strait Times website

2.3.3 Item

Item is the ad creatives which are the actual content that users see when an ad is served on a webpage or application. They can include a wide range of media types, such as images, videos, and interactive HTML5 elements (Stone-Gross, 2011). Ad creatives are essential to the success of an advertising campaign because they are what delivers the message to the audience. Advertisers use different creative elements to appeal to their specific target audience and ensure that their message is delivered in an effective and engaging way. Ad creatives can be designed for specific platforms and devices to ensure that they are optimized for the best possible user experience. It can be created using different tools and software, including third-party ad servers, campaign managers, and native ad formats. Overall, ad creatives are an essential part of any advertising campaign and play a crucial role in driving engagement, conversions, and overall success (Google Ads Help, n.d.).

2.3.4 Goal

Advertisers may also have more specific goals that align with their business objectives. For example, an advertiser may want to increase their social media following, drive more sign-ups for their newsletter, or promote a particular product or service. By setting specific goals for their advertising campaigns, advertisers can tailor their strategies and ad creatives to achieve the desired outcomes. Furthermore, advertisers may also have different target audiences that they want to reach with their advertising campaigns. For instance, an advertiser may want to target a specific age range, gender, location, or interest group. By using targeting options available on advertising platforms, advertisers can ensure that their ads are shown to the right audience, which increases the chances of achieving their advertising goals. Ultimately, the success of an advertising campaign depends on a variety of factors, including the chosen advertising platform, the targeting options used, the ad creatives, and the chosen advertising goals. With carefully planning and executing an advertising campaign, advertisers can effectively reach their target audience and achieve their desired outcomes (Business Help Centre, n.d.).

2.3.5 Ad Log

Ad log records chronological documentation of any activities that affected a particular operation or event. In computing, a log refers to an automatic record of relevant events with timestamps for a particular system, which is produced by most software applications and systems (Wigmore, n.d.). it is a record of an ad that occurred at a specific time and additional information to provide context. It captures all system activity, including transactions, errors, and security breaches. The data contained in log files may be structured, or unstructured, it can be transmitted in various ways. Hence, ad log can be useful for keeping track of ad data, emergency recovery, and application improvement. A proper log management is important due to it able to manage the sheer volume of data, absorb and derive valuable insight, and also in terms of digital transformation (Sharif, 2022). By analysing ad logs, advertiser can gain valuable insights into their advertising activities, identify areas for improvement, and take action to prevent fraudulent activities (Rapid7, 2016).

2.3.6 Pricing Type

Setting the right price is crucial for any business, as it directly impacts their profitability and market share. There are 5 common pricing strategies, cost-plus pricing, competitive pricing, price skimming is a strategy, penetration pricing and value-based pricing. Pricing strategies are important not only to the firms themselves but also to their target consumers. It should be well calculated in order to make profit as well as to be align with the products and services (paddle, n.d.). When setting prices, businesses must also consider the elasticity of demand, or how sensitive customers are to changes in price. For example, if demand for a product is highly elastic, a small increase in price may cause a significant drop in demand. Businesses must also consider their costs, including fixed costs such as rent and variable costs like materials and labor, to ensure that their prices are profitable. In summary, pricing is an important aspect of any business strategy and must be carefully considered to ensure profitability and market share. Different pricing strategies, such as cost-plus, competitive, price skimming, penetration, and value-based pricing, can be used depending on the product or service and the target market (BDC, n.d.).

2.3.7 Look Up Form

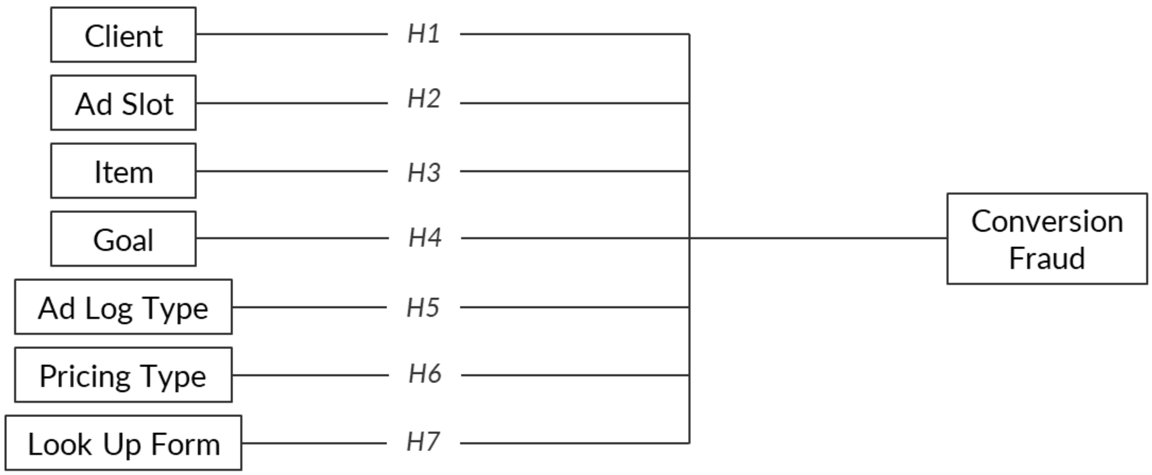
A look up form is a type of form used in a database or software application that allows users to search and retrieve specific data from a database. It typically includes fields or criteria for entering search parameters, including keywords or dates, and displays results. A look up form can be useful for quickly finding and collecting specific information within a large dataset. It usually helps users select records form a related table, while it is automatically added once look up column is added to form (Microsoft, 2022). A lookup form can be a useful tool for advertisers since it enables them to gather and access particular data about their target market or future clients. Advertisers can quickly search through a database using a lookup form to identify pertinent data, such as demographics, interests, or purchase patterns, and utilise that data to construct customised ad campaigns. The table can either be filled with values manually by the programmer, or it can be populated by the program itself while it is calculating those values. In the context of advertising field, it is widely used by advertiser to collect users' information who are target audience or potential consumer. (Computer Hope, 2021).

2.3.8 Conversion Fraud

Conversion fraud including various activities that deceive advertiser with misrepresent advertising inventory or hide machines as humans that lead to misuse advertising expenditures. Conversion fraud takes approximately 30% of total digital advertising revenue estimated by most of the industry (Gordon et al., 2021). A conversion in ad network denotes one or a set of meaningful business actions taken by users which they are potential consumer in converting to sales transaction. Alternatively, a conversion can also be defined as "agreed-upon action taken by a user". For example, a simple conversion event can be a downloading of a file or filling out a form, or a completion of an online purchase order. Such fraud is also called conversion spam. After users clicking on the displayed advertisement, they are normally directed to a landing page which shows summarized information about the advertised product or services (Zhu *et al,* 2017).

2.3 Conceptual Framework

Figure 2.3: Conceptual Framework



Source: Kaggle, 2021

2.4 Hypothesis Development

2.4.1 Advertiser

Advertiser, the main character in most of the cases of conversion fraud. While the advertiser is one of the independent variables towards conversion fraud. In this case, brand image is important which affect consumer's purchase decision. Brand image is a long-term process development and it serve as a weapon to compete in the industry. Based on Malik (2013), assuming brand is managed well, the firms can gain number of consumers and it leads to long-term profitable relation. Nowadays, most people care about status, they use branded products to indicate as their status symbols. This shows people prefer to purchase products that is well known and brand name. Furthermore, brand image is about enhance consumer's memories regarding on brand, which will affect consumer final purchase (Saleem, & Abideen, 2011). Thus, despite the attractiveness of ad, it might not be able to convert to sales since the brand image of advertiser is not build.

H_1: There is a relationship between advertiser and conversion fraud.

2.4.2 Ad Slot

Fraudster take advantage of ad slot to execute fraudulent activities that is conversion fraud. In the cases of ad stacking and pixel stuffing, fraudster overlap multiple ads on top of each other in the same ad slot on the webpages. While ad slot will become vulnerable when it is manipulated by fraudster to serve fraudulent ads to generate fake data. For instance, fake impressions and clicks, while it is paid by advertiser, but it is invalid to advertiser at the same time it might affect in their decision making. Hence, it is important to monitor ad slot on webpages which have potential to happen conversion fraud and harming the digital advertising environment (Zhu *et al*, 2017).

H_2: There is a relationship between ad slot and conversion fraud

2.4.3 Item

Item is the ad creatives that consists of text and video, a typical advertisement contains title, description, and website address (URL) of the advertiser. The creatives context plays important role due to it conveys information about the brand, products and services which aims to reach potential consumer. According to the study, there is a relationship between ad creatives and the effectiveness of listing in the sponsored search market. This shows the direct relationship to the ad performance, the better the ad creative, the higher the effectiveness of listing in the sponsored search market, and thus, increase higher sales and leads (Animesh, 2007).

H_3: There is a relationship between item and conversion fraud

2.4.4 Goal

Goal is the objectives which advertiser wish to achieve, with appropriate setting of ad campaign, it acts effectively to reach target audience. In digital advertising, advertiser set campaign with many goals for their ads (Fowler, 2023). Setting advertising goals can benefit companies in many ways, such as improving brand recognition, increasing sales, building customer loyalty, and driving website traffic. By defining specific goals, businesses can create targeted and effective advertising campaigns that align with marketing strategy. Some common advertising goals include increasing brand awareness, generating leads, boosting sales, promoting a new product or service, building customer trust and loyalty, and increasing website traffic or social media engagement. By understanding these goals and how they can benefit their business, companies can create more successful advertising campaigns and better connect with their target audience (Indeed Editorial Team, 2023).

H_4: There is a relationship between goal and conversion fraud

2.4.5 Ad Log Type

Ad log is the dataset of advertisement, which records the advertisement activities in terms of click, impression, and view. According to Ghosh, 2020, the ad log dataset has been taken as the essential elements that affecting consumer purchasing decision making in the context of previous purchasing habit, and a sequence of advertisement viewed such as number of advertisements click and view sequence. It proves that the higher the click the higher the accuracy of final result of purchase action taken by consumers. It records all the data of an ad, which could also be a powerful tool for detecting conversion fraud, such as number of click does not align the actual sales. The ad log data set provide valuable insight into the performance of an ad campaign.

H_5: There is a relationship between ad log type and conversion fraud

2.4.6 Pricing Type

The amount of money set by company in exchange for a good or service. The price of goods and services may be sensitive to the needs of the consumer. Consumer behaviour is significantly influenced by consumer perception. Price is a key aspect in how consumers perceive products and services. Pricing strategy is significant in this situation and may be a good indicator of a product's quality. While both loyal and non-loyal consumers' price sensitivity is increased by promotional activities and advertising. Although advertising is a key tool for spreading knowledge about a product and shaping public perception, it is not always as effective as it could be (Malik *et al,* 2014).

H_6: there is a relationship between pricing type and conversion fraud

2.4.7 Look Up Form

Look up hold numbers of value that is needed to be calculated. Advertisers gather and access particular data about their target market or future clients though look up form. It allows advertiser to search through a database using a lookup form to identify pertinent data, such as demographics, interests, or purchase patterns, and utilise that data to construct customised ad campaigns in an efficient way. While it might also be used by fraudster for fraudulent activities. Fraudster can use look up form to collect users' personal information or payment details which user interacting with a legitimate business. This not only harming the user but also the reputation of firms, and it can result in financial losses for legal owner (Computer Hope, 2021).

H_7: There is a relationship between look up from and conversion fraud

Chapter 3 METHODOLOGY

3.1 Introduction

This chapter, illustrate the methodology for this study. Secondary data will use to predict the most impactful independent variables towards the conversion fraud in digital advertisement. The secondary data is publicly shared on Kaggle by (author). The entire package is related to digital advertisement conversion fraud. The data consist of 54 variables in total. It is proposed to be analysed using OLS regression.

3.1.1 Accessing secondary data

This study is using secondary data as the main research method. Secondary data is the available sources collected by someone other than the researchers using it (Tantawi, 2023). In this study, the secondary data is accessible through Kaggle, one of the data science platforms. The secondary data is available publicly for Kaggle users. Data that has been gathered and processed by someone else, usually for a different reason, is referred to as secondary data. It can be helpful in research since it gives access to a lot of data that would be challenging or time-consuming to gather independently. The secondary data for this study came from Kaggle, a website that offers a range of datasets to the general public for use in data science projects. It is a raw data which data pre-processing is required.

3.1.2 Sources of secondary data

The source of secondary data is retrieved from Kaggle. Kaggle is an online community of data scientists and machine learning engineers. Kaggle founded in 2010 by providing plenty of dataset in different areas, while those datasets are to help users in building Artificial Intelligent (AI) models, publish datasets, work with other data scientists and machine learning engineers, and enter competition to solve data science challenges. Kaggle also offering a public data and cloud-based business platform for data science and AI education. The contributor of the data is Anoop Prasad H, who is a Python Developer at Mavenir Systems from India. Anoop is also the head of operation in Epsilon Scientific (Mahmoud, 2022).

3.2 Sampling design

In most of the cases, sampling design discuss on target population, sampling frame, and sampling technique. Due to the secondary data of this study has already processed, the target population, sampling frame and location and sampling technique could not be determined, thus remain unknown. While the sampling size of this secondary data is 965 observations.

3.3 Data Processing

Data has been processed and normalize by the author. The computational procedures have been done through OLS regression, and out of 54 variables, the author has concluded 7 variables there are most related to dependent variable. The process of identifying the most relevant variables through regression analysis involves statistical testing and evaluation, which can help to identify the variables with the strongest relationships to the dependent variable.

3.4 Proposed Data Analysis Tool

3.4.1 Descriptive Analysis

Descriptive analysis is a method used to examine data that has been collected, with the goal of presenting it in a clear and straightforward way. This can be done by creating tables, graphs, charts, or other visual representations that help to illustrate patterns and trends within the data. Overall, descriptive analysis is used to provide a basic understanding of the data that has been collected. Overall, descriptive analysis can be a useful tool for researchers, analysts, and decision-makers in a variety of sectors, including business, healthcare, social sciences, and more. It helps to provide context and insight into the data (Borges et al., 2020).

3.4.2 Inferential analysis

In this study, Pearson Correlation, Spearman Correlation, and OLS regression will be used to examine the secondary data.

3.4.2.1 Pearson Correlation

This is a method used to measure the strength of the connection between the independent and dependent variables. If the value is negative, there is an inverse relationship between the variables, while if the value is positive, there is a direct relationship between them. In other words, this technique helps to determine how much influence the independent variable has on the dependent variable (Borges et al., 2020).

Table 3.4.2.1: Rules of thumb about Correlation

Correlation Coefficient	Correlation's strength
$\pm 0.91\ to\ \pm 1.00$	Very high
$\pm 0.71\ to\ \pm 0.90$	high
$\pm 0.41\ to\ \pm 0.70$	Moderate
$\pm 0.21\ to\ \pm 0.40$	Low
$\pm 0.00\ to\ \pm 0.20$	Negligible

Source: Cheong et al., 2021

3.4.2.2 Spearman Correlation

The Spearman correlation coefficient is a statistical measure that helps us understand the strength and direction of the relationship between two variables. It is denoted by "rho" or "rs" and can range from -1 to +1. A value of -1 indicates a perfect negative correlation between the two variables, which means that as one variable increases, the other variable decreases in a perfectly predictable way. A value of +1 indicates a perfect positive correlation between the two variables, which means that as one variable increases, the other variable also increases in a perfectly predictable way. A value of 0 indicates no correlation between the two variables, which means that there is no predictable relationship between them. The formula for calculating the Spearman correlation involves ranking the values of each variable, finding the differences between their ranks, and then using these differences to calculate the correlation coefficient (Frost, 2021).

$$-1 \leq r_s \leq 1$$

Table 3.4.2.2: Rules of thumb about Correlation

Correlation Coefficient	Correlation's strength
0.00-0.19	Very Weak
0.20-0.39	Weak
0.40-0.59	Moderate
0.60-0.79	Strong
0.80-1.00	Very Strong

Source: Frost, 2021

3.4.2.3 OLS Regression

The regression analysis is the most common regression method that used by most of the social science study. Ordinary Least Squares (OLS) regression method of analysis is used for this study. It is a statistical technique which used to investigate the correlation between a dependent variable and one or more independent. The objective of OLS regression aims to identify the line or equation that most suitable to the data, thereby enabling the prediction of the dependent variable based on the values of the independent variables (Burton, 2021).

The regression line used here is:

$$\hat{Y}_i = \text{CONST} + (X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7)$$

$$X_1 = 0.004$$

$$X_2 = 0.001$$

$$X_3 = -0.001$$

$$X_4 = -0.001$$

$$X_5 = 0.396$$

$$X_6 = 0.003$$

$$X_7 = -0.302$$

Figure 4.3.2.3: Computing Coding of processing variables by using OLS Regression

```
X = df_2[['clientid_cr','adslotdimid_cr','itemid_cr','goalid_cr','adLogType_cr','pricingtype_cr','lookUpFrom_cr']] # here we have 2 variables for multiple regression. If you just want to use one variable for simple linear regression, then use X = df['Interest_Rate'] for example.Alternatively, you may add additional variables within the brackets
Y = df_2['conversion_fraud']
regr = linear_model.LinearRegression()
regr.fit(X, Y)

print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)

# with statsmodels
X = sm.add_constant(X) # adding a constant

model = sm.OLS(Y, X).fit()
predictions = model.predict(X)

print_model = model.summary()
print(print_model)
```

Source: Kaggle, 2021

3.5 Chapter Summary

This chapter covered the methodology that was applied in this study by using secondary data. It also covered accessing of secondary data and source of secondary data. In this study, descriptive analysis and inferential analysis are used to examine the secondary data. This including Pearson correlation, Spearman correlation and OLS regression. Furthermore, there are only 7 variables will be used in this study out of 54 variables.

Chapter 4 RESULT AND FINDINGS

4.0 Introduction

This chapter will focus on the results of the data analysis based on the regression model chartered in previous chapter.

4.1 Data Screening

In this study, there is 965 observations in total that have been recorded by the author. All 965 observations that are collected can be continued to be used in this research. Out of 55 variables, it has excluded object and float, this study only discusses on integer variables, there are only 7 variables will be used and discuss in this study.

Float variables, also called floating point is a type of variable that hold decimal numbers with a variable position for the decimal point. Examples are 0.23, 32.568, and 25874.3. The size limits and definitions of floating-point numbers may vary across different programming languages or systems (Computer Hope, 2020).

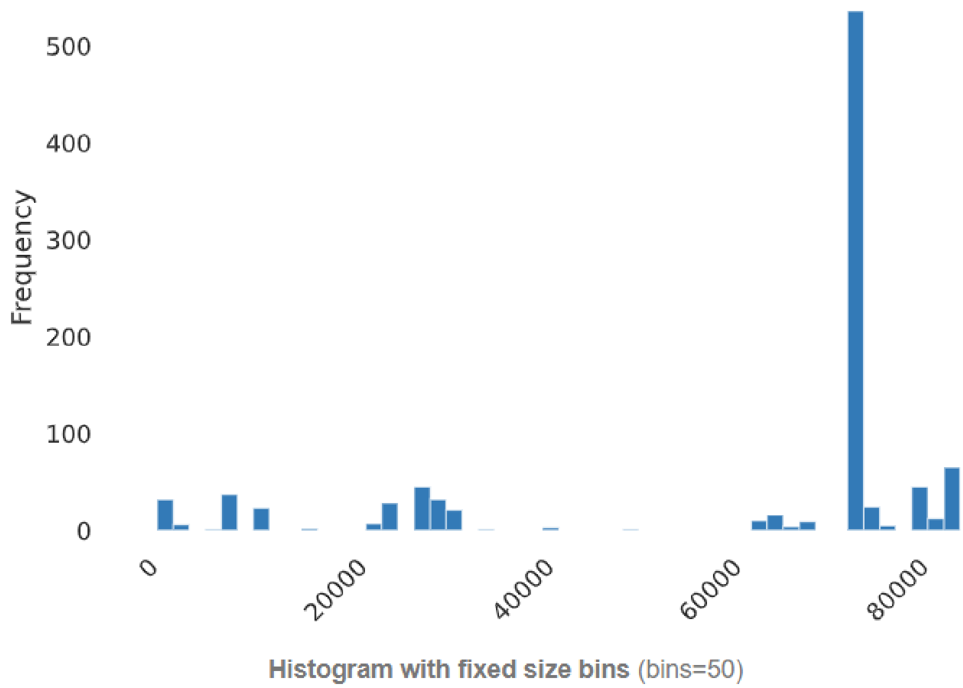
Object variables are simply a collection of data and methods. It can use object variables in expressions, which can be view as the name for a location in computer's memory that holds data. By using object variables, researcher requires to declare and instantiate an object and create a reference to the object (Actian, n.d.).

Integer variables are the binary variables which only take value from 0 to 1. With its code "ints", integer is the values written and stored as tnumbers and run loops (Mattingly, 2023).

4.2 Descriptive Analysis

4.2.1 Advertiser

Figure 4.2.1: Advertiser



Source: Kaggle, 2021

Table: 4.2.1: Advertiser

Quantile Statistic		Descriptive Statistic	
Minimum	802	Standard Deviation	**15 456.7256**
5-th percentile	7 750	Coefficient of variation (CV)	0.4058
Q1	64 588	Kurtosis	0.0792
Median	75 694	Mean	**62 726.9586**
Q3	75 694	Median Absolute Deviation (MAD)	0
95-th percentile	85 572	Skewness	-1.2987

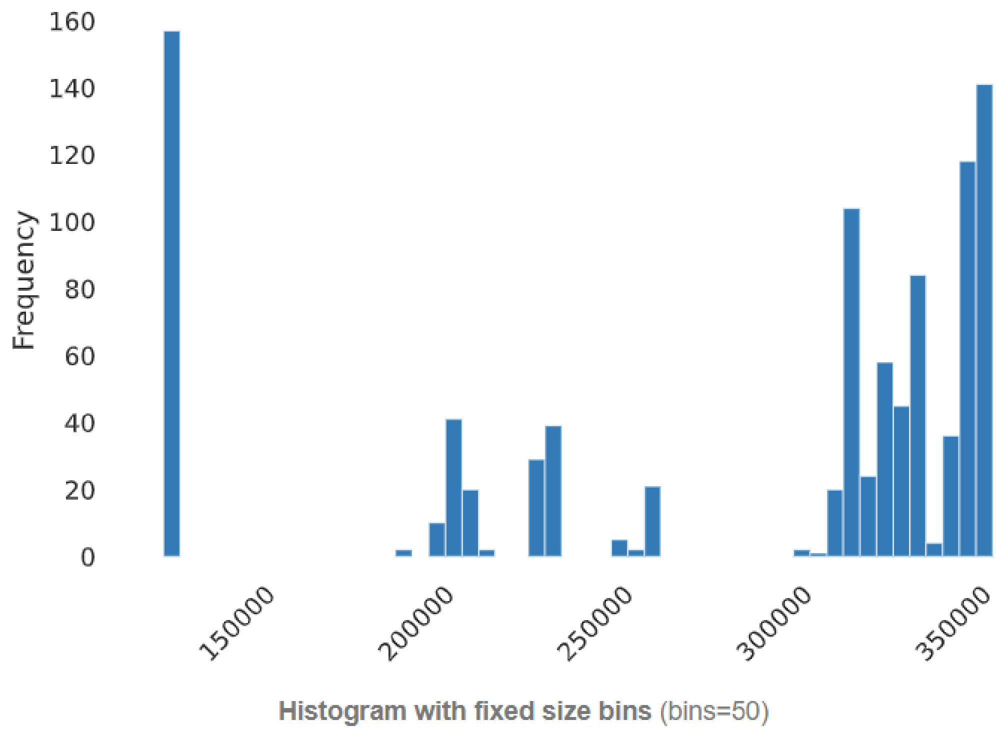
Maximum	86 555	Sum	60 531 515
Range	85 753	Variance	648 044 877.9
Interquartile range (IQR)	11 106	Monotocity	Not monotonic

Source: Kaggle, 2021

Table 4.2.1 shows the quantile statistic and descriptive statistic of advertiser. The standard deviation is 15 456.7256 that is the sampling variability of the parameter; the mean is 62 726.9586.

4.2.2 Ad Slot

Figure 4.2.2 Ad Slot



Source: Kaggle, 2021

Table 4.2.2: Ad Slot

Quantile Statistic		Descriptive Statistic	
Minimum	129 117	Standard Deviation	**82 452.3982**
5-th percentile	129 239	Coefficient of variation (CV)	0.2871
Q1	233 531	Kurtosis	-0.5661
Median	332 775	Mean	**287 159.8145**
Q3	355 670	Median Absolute Deviation (MAD)	22 993
95-th percentile	356 952	Skewness	-0.9955

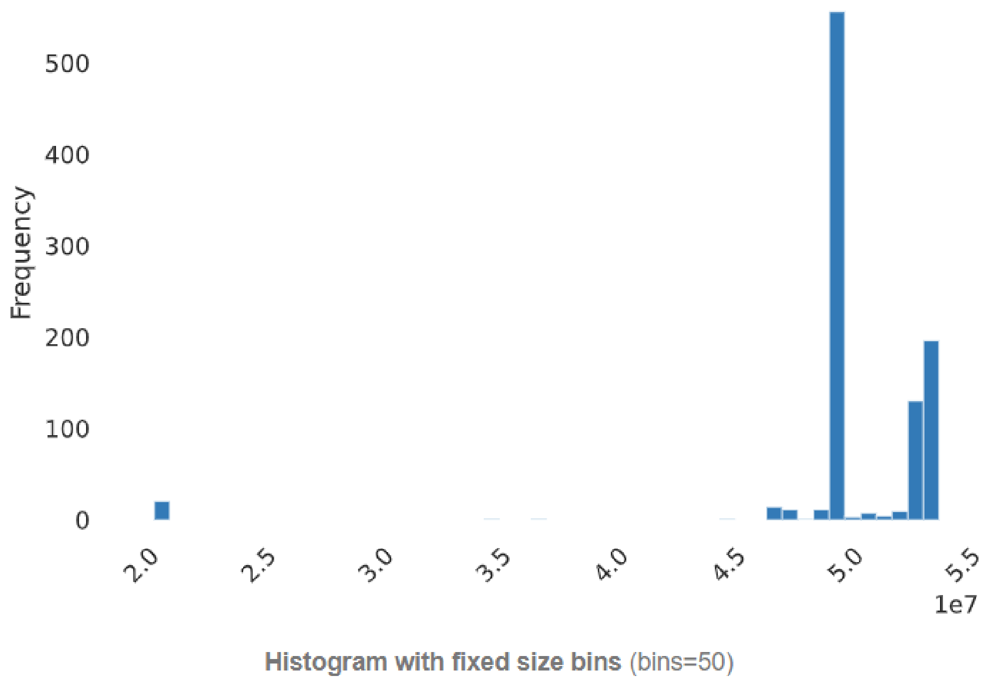
Maximum	360 638	Sum	277 109 221
Range	231 521	Variance	6 798 397 974
Interquartile range (IQR)	122 139	Monotocity	Not monotonic

Source: Kaggle, 2021

Table 4.2.2 shows the quantile statistic and descriptive statistic of ad slot. The standard deviation is 82 452.3982 that is the sampling variability of the parameter; the mean is 287 159.8145.

4.2.3 Item

Figure 4.2.3 Item



Source: Kaggle, 2021

Table 4.2.3: Item

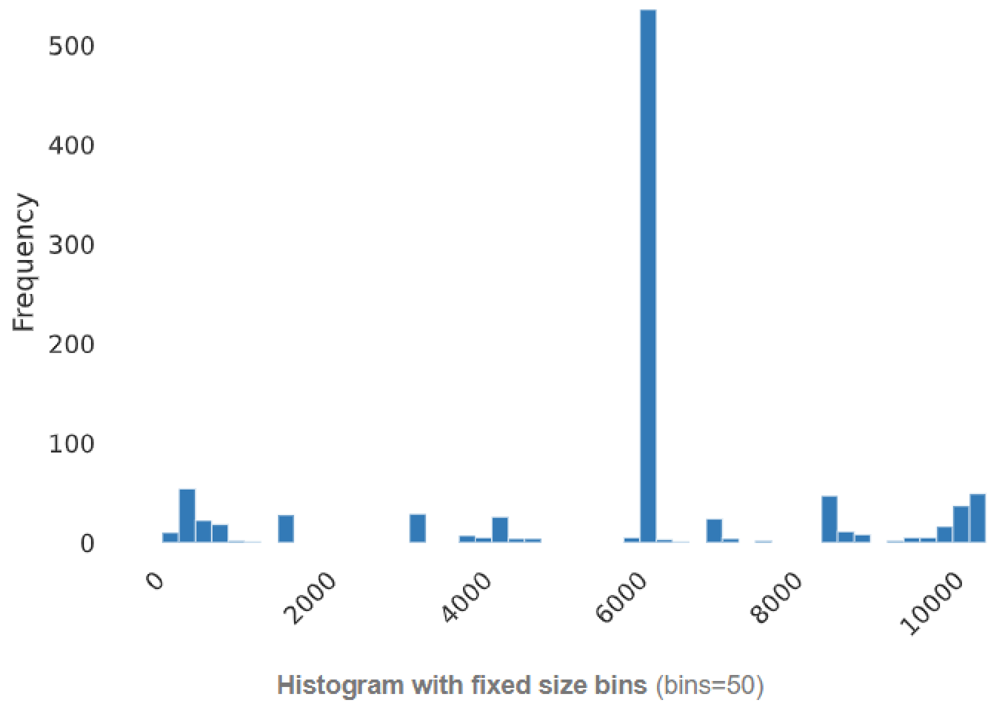
Quantile Statistic		Descriptive Statistic	
Minimum	20 506 010	Standard Deviation	**4 804 771.21**
5-th percentile	48 501 049.8	Coefficient of variation (CV)	0.0957
Q1	49 504 155	Kurtosis	27.7182
Median	49 504 155	Mean	**50 203 296.78**
Q3	53 182 893	Median Absolute Deviation (MAD)	5
95-th percentile	53 930 177.8	Skewness	-4.8595
Maximum	53 930 635	Sum	$4.84461814 \times 10^{10}$
Range	33 424 625	Variance	$2.308582638 \times 10^{13}$
Interquartile range (IQR)	3 678 738	Monotocity	Not monotonic

Source: Kaggle, 2021

Table 4.2.3 shows the quantile statistic and descriptive statistic of item. The standard deviation is 4 804 771.21 that is the sampling variability of the parameter; the mean is 50 203 296.78.

4.3.4 Goal

Figure 4.3.4: Goal



Source: Kaggle, 2021

Table 4.3.4: Goal

Quantile Statistic		Descriptive Statistic	
Minimum	82	Standard Deviation	**2 666.1029**
5-th percentile	469	Coefficient of variation (CV)	0.4394
Q1	6 384	Kurtosis	0.2309
Median	6 384	Mean	**6 067.0124**
Q3	6 384	Median Absolute Deviation (MAD)	0
95-th percentile	10 488.2	Skewness	-0.6115
Maximum	10 679	Sum	5 854 667
Range	10 597	Variance	7 108 104.9

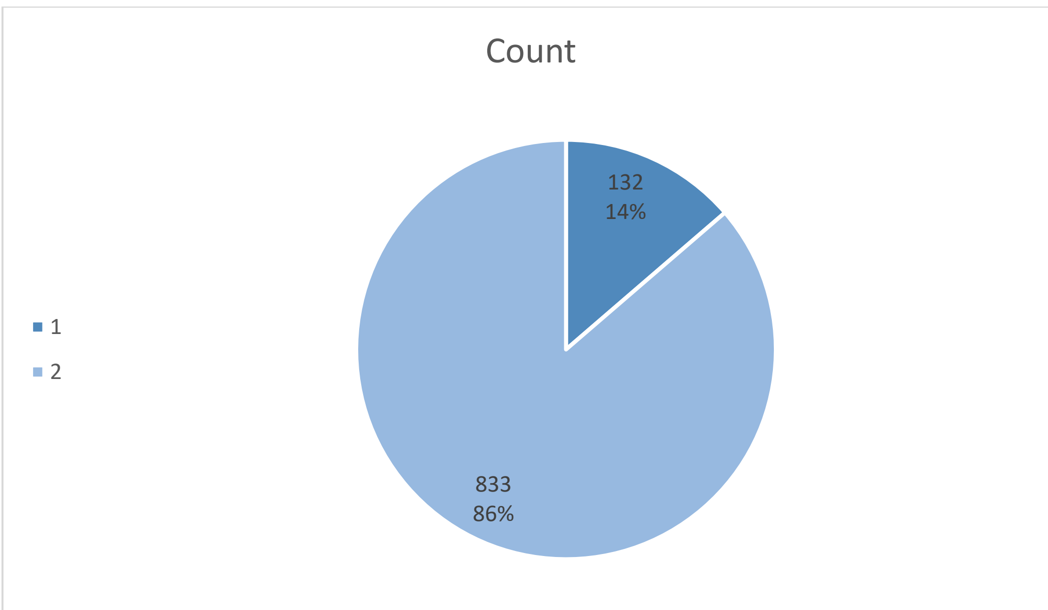
Interquartile range (IQR)	0	Monotocity	Not monotonic

Source: Kaggle, 2021

Table 4.3.4, shows the quantile statistic and descriptive statistic of goal. The standard deviation, 2 666.1029, that is the sampling variability of the parameter. The mean is 6 067.1024.

4.2.5 Ad Log

Figure 4.2.5: Ad Log



Source: Kaggle, 2021

Figure 4.2.5 reflects 833, 86% of the observation lead to conversion fraud, however, 132, 14% of observation did not convert to conversion fraud.

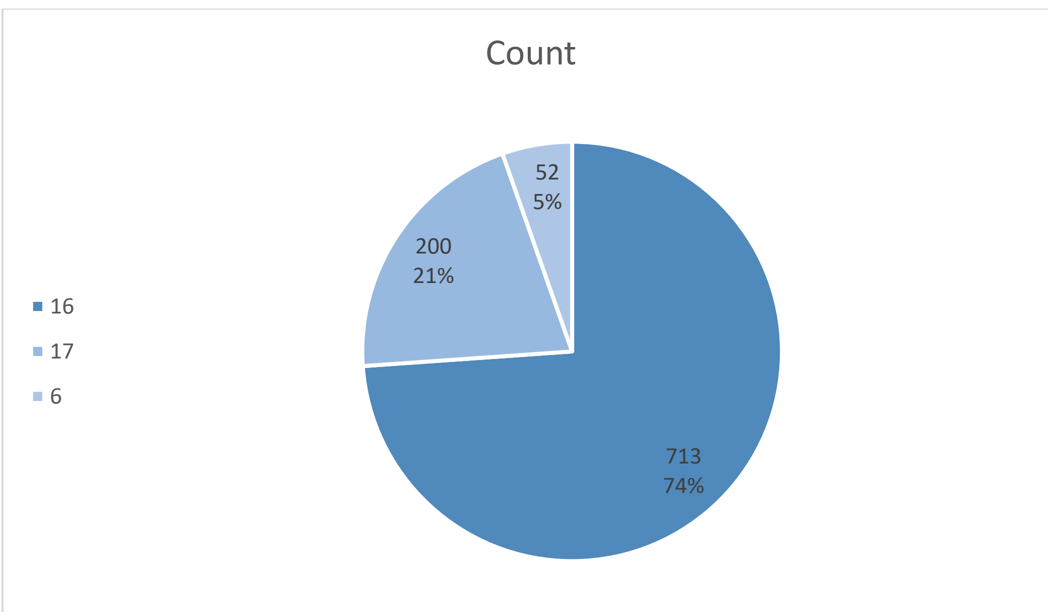
Table 4.2.5: Ad Log

Length		Character and Unicode		Unique		Sample	
Max length	1	Total characters	965	Unique	0	1st row	2
Median length	1	Distinct characters	2	Unique (%)	0.0%	2nd row	2
Mean length	1	Distinct categories	1			3rd row	2
Min length	1	Distinct scripts	1			4th row	2
		Distinct blocks	1			5th row	2

Source: Kaggle, 2021

4.2.6 Pricing Type

Figure 4.2.6: Pricing Type



Source: Kaggle, 2021

Figure 4.2.6 reflects 713, 74% of the observation lead to conversion fraud.

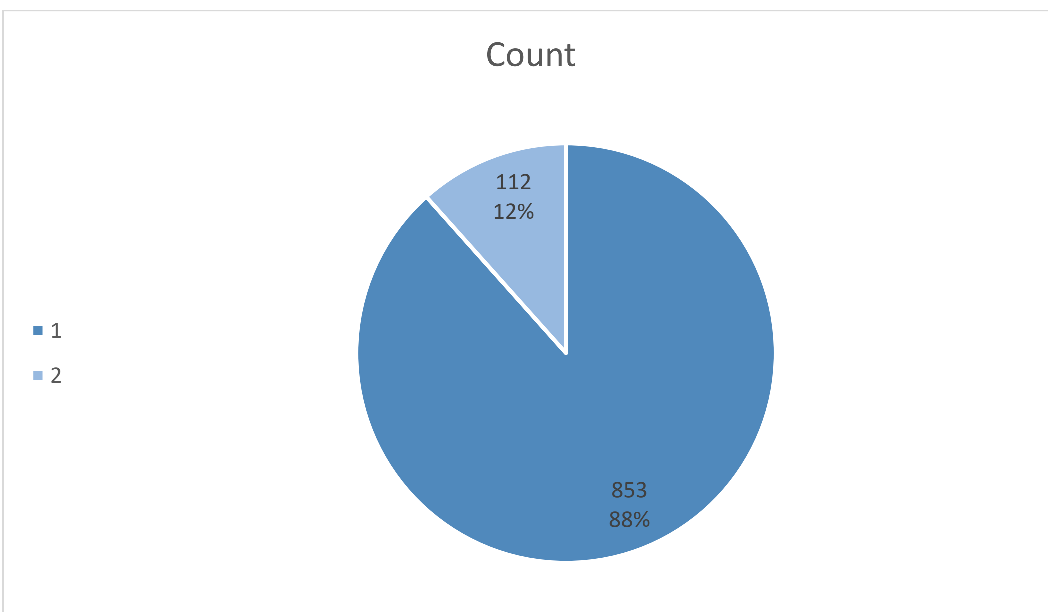
Table 4.2.6: Pricing Type

Length		Character and Unicode		Unique		Sample	
Max length	2	Total characters	1878	Unique	0	1st row	16
Median length	2	Distinct characters	3	Unique (%)	0.0%	2nd row	16
Mean length	1.94611399	Distinct categories	1			3rd row	16
Min length	1	Distinct scripts	1			4th row	16
		Distinct blocks	1			5th row	17

Source: Kaggle, 2021

4.2.7 Look Up Form

Figure 4.2.7: Look Up Form



Source: Kaggle, 2021

Figure 4.2.7 shows there is 853, 88% of the observations has related to conversion fraud, however there is 112, 12% of the observation did not create conversion fraud.

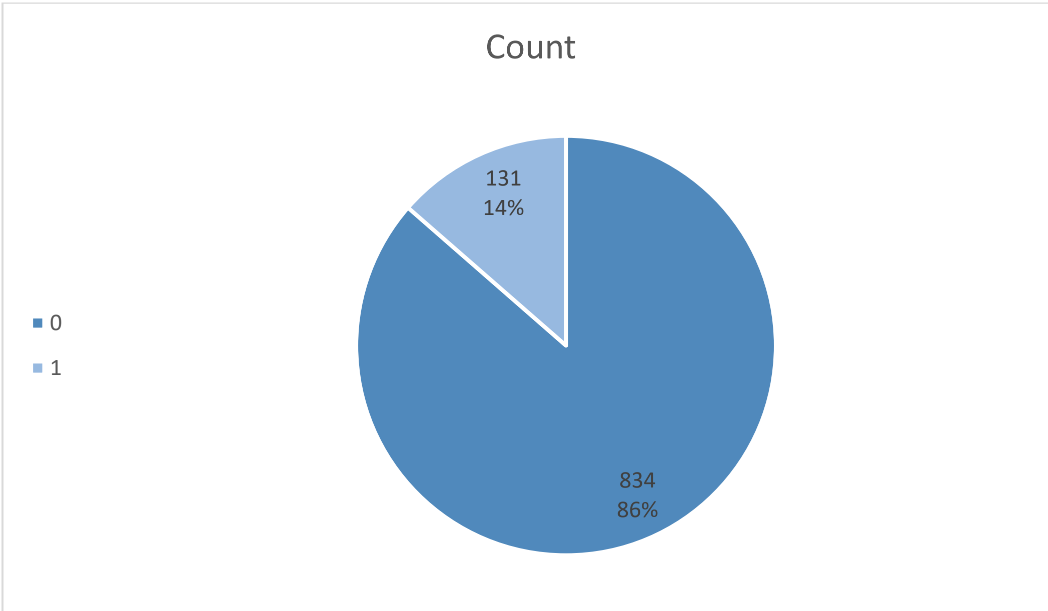
Table 4.2.7: Look Up Form

Length		Character and Unicode		Unique		Sample	
Max length	1	Total characters	965	Unique	0	1st row	2
Median length	1	Distinct characters	2	Unique (%)	0.0%	2nd row	2
Mean length	1	Distinct categories	1			3rd row	2
Min length	1	Distinct scripts	1			4th row	2
		Distinct blocks	1			5th row	2

Source: Kaggle, 2021

4.2.8 Conversion fraud

Figure 4.2.8: Conversion Fraud



Source: Kaggle, 2021

Figure 4.2.8 reflects that 834, 86% of the observations cause to conversion fraud. However, there is 131, 14% of the observations does not convert to conversion fraud.

Table 4.2.8: Conversion Fraud

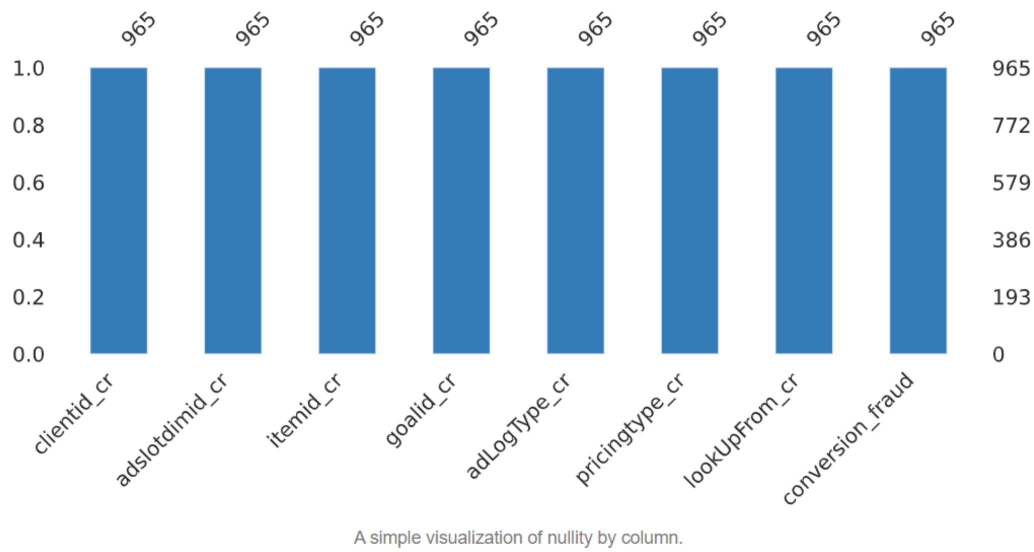
Length		Character and Unicode		Unique		Sample	
Max length	1	Total characters	965	Unique	0	1st row	1
Median length	1	Distinct characters	2	Unique (%)	0.0%	2nd row	1
Mean length	1	Distinct categories	1			3rd row	1
Min length	1	Distinct scripts	1			4th row	1

		Distinct blocks	1			5th row	1

Source: Kaggle, 2021

4.2.9 Missing Value

Figure 4.2.9: Missing Value



Source: Kaggle, 2021

Figure 4.2.9 shows that no missing value from this secondary data. The total of 965 observations data is fully used for this research.

4.4 Inferential analysis

Pearson correlation analysis is the method that is used to test the relationship between independents variables and dependent variables.

4.4.1 Pearson Correlation

The Pearson's correlation coefficient (r) is a measure of linear correlation between two variables. Its value lies between -1 and +1, -1 indicating total negative linear correlation, 0 indicating no linear correlation and 1 indicating total positive linear correlation. Furthermore, r is invariant under separate changes in location and scale of the two variables, implying that for a linear function the angle to the x-axis does not affect r.

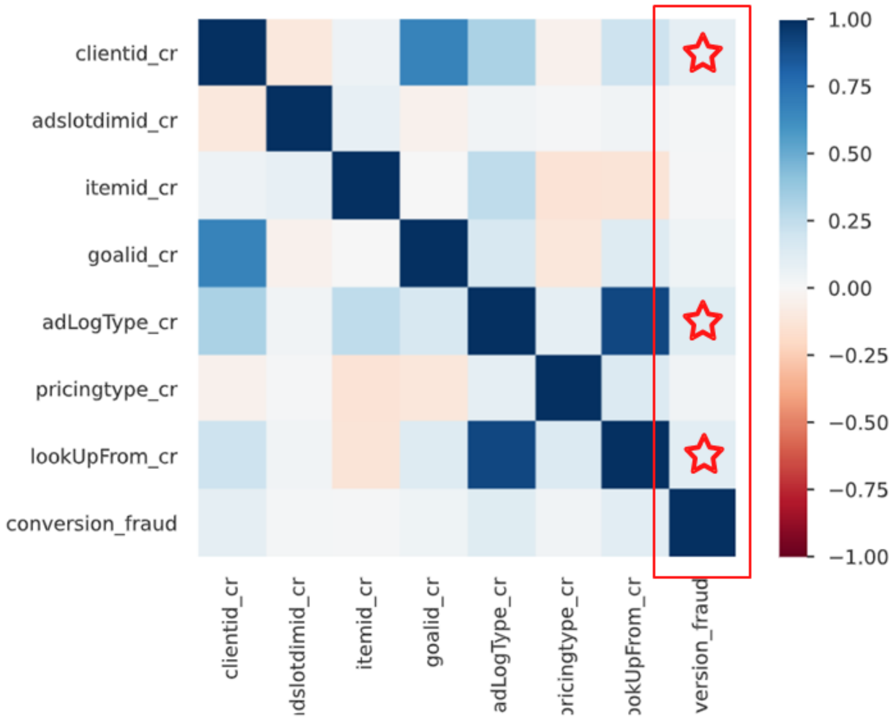
To calculate r for two variables X and Y, one divides the covariance of X and Y by the product of their standard deviations.

Table 4.4.1: Pearson Correlation

		Advertiser	Ad Slot	Item	Goal	Ad Log	Pricing	Look Up Form
Conversion Fraud	Pearson Correlation	**0.25**	0.00	0.00	0.125	**0.25**	0.125	**0.25**
	Sig. (2 tailed)	4.124	0.001	-0.108	-0.001	1.456	0.165	-1.112
	N	965	965	965	965	965	965	965

Source: Kaggle, 2021

Figure 4.4.1: Pearson Correlation heat map



Source: Kaggle, 2021

4.4.2 Spearman Correlation

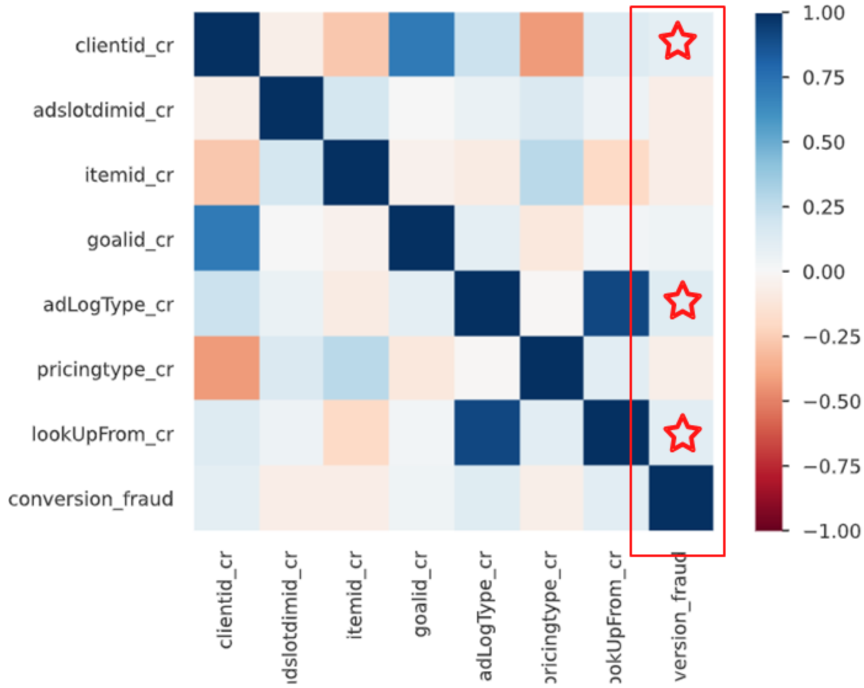
The Spearman's rank correlation coefficient (ρ) is a measure of monotonic correlation between two variables, so it is better in catching nonlinear monotonic correlations than Pearson's r. Its value lies between -1 and +1, -1 indicating total negative monotonic correlation, 0 indicating no monotonic correlation and 1 indicating total positive monotonic correlation. To calculate ρ for two variables X and Y, one divides the covariance of the rank variables of X and Y by the product of their standard deviations (Frost, 2021).

Table 4.4.2: Spearman Correlation

		Advertiser	Ad Slot	Item	Goal	Ad Log	Pricing	Look Up Form
Conversion Fraud	Spearman Correlation	**0.25**	0.00	0.00	0.125	**0.25**	0.00	**0.25**
	Sig. (2 tailed)	0.004	0.001	-0.001	-0.001	1.456	0.165	-1.112
	N	965	965	965	965	965	965	965

Source: Kaggle, 2021

Figure 4.4.2: Spearman Correlation heat map



Source: Kaggle, 2021

4.4.3 OLS Regression

The method of Ordinary Least Squares (OLS) is most widely used model due to its efficiency. This model gives best approximate of true population regression line. The principle of OLS is to minimize the square of errors. In regression analysis, this approach is frequently employed. Its main objective is to estimate the linear regression model's parameters by minimising the sum of the squared discrepancies between the fitted values and the observed data. The OLS approach makes the assumption that the regression model's errors are independent, normally distributed, and have a constant variance. The OLS method can be used to estimate the relationship between two or more variables and make predictions about future values of the dependent variable. It is widely used in various fields, including economics, finance, social sciences, and engineering.

Figure 4.4.3: OLS Regression Result

```
Intercept:
 0.29313503916640654
Coefficients:
 [ 4.52275917e-07  1.23652771e-07 -8.77107102e-09 -1.26592423e-07
  3.95722690e-01  3.31303566e-03 -3.02229936e-01]
                            OLS Regression Results
==============================================================================
Dep. Variable:        conversion_fraud   R-squared:                       0.021
Model:                             OLS   Adj. R-squared:                  0.014
Method:                  Least Squares   F-statistic:                     2.983
Date:                 Tue, 01 Jun 2021   Prob (F-statistic):            0.00422
Time:                         19:07:09   Log-Likelihood:                -324.95
No. Observations:                  965   AIC:                             665.9
Df Residuals:                      957   BIC:                             704.9
Df Model:                            7
Covariance Type:             nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              0.2931      0.368      0.796      0.426      -0.430       1.016
clientid_cr     4.523e-07   7.28e-07      0.621      0.534   -9.76e-07    1.88e-06
adslotdimid_cr  1.237e-07   1.35e-07      0.914      0.361   -1.42e-07    3.89e-07
itemid_cr      -8.771e-09   6.43e-09     -1.363      0.173   -2.14e-08    3.86e-09
goalid_cr      -1.266e-07    5.8e-06     -0.022      0.983   -1.15e-05    1.12e-05
adLogType_cr       0.3957      0.226      1.748      0.081      -0.048       0.840
pricingtype_cr     0.0033      0.005      0.686      0.493      -0.006       0.013
lookUpFrom_cr     -0.3022      0.230     -1.314      0.189      -0.754       0.149
==============================================================================
Omnibus:                      356.324   Durbin-Watson:                   0.046
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              894.132
Skew:                           2.047   Prob(JB):                    6.95e-195
Kurtosis:                       5.341   Cond. No.                     2.20e+09
==============================================================================
```

Source: Kaggle, 2021

Table 4.4.3: OLS Regression Result

	coef	std err	t	P>|t|	[0.025	0.975]
Const	0.2931	0.368	0.796	0.426	-0.430	1.016
Advertiser	0.004	0.007	0.621	0.534	**-0.009***	0.005
Adslot	0.001	0.0010	0.914	0.361	**-0.001***	0.004
Item	-0.001	0.0008	-1.363	0.173	**-0.0007***	0.0005
Goal	-0.001	0.0144	-0.022	0.983	**-0.008***	0.008
Adlog	0.3957	0.226	1.748	0.081	**-0.048***	0.840
Pricing	0.0033	0.005	0.686	0.493	**-0.006***	0.013
Lookupform	-0.3022	0.230	-1.314	0.189	**-0.754***	0.149

Source: Kaggle

R-squared value:

R-squared value is the coefficient of determination that tells us that how much percentage variation independent variable can be explained by independent variable. Here, 2.1% variation in Y can be explained by X. The maximum possible value of R^2 can be 1, means the larger the R-square value better the regression.

All null hypothesises are supported when R-square is positive; F-statistic > 1. While R-square is 0.021, positive value; F-statistic is 2.983, that is more than one.

4.6 Hypothesis Testing

Accept all null hypothesis when R-square is positive, and F-statistic > 1, thus, H_1, H_2, H_3, H_4, H_5, H_6, H_7 are all accepted. However, p value also plays important role in determining significant variables. By using p value, H_7 is not supported.

Table 4.6: Hypothesis Testing Result

Hypothesis Testing	Result	Conclusion
H_1: There is a relationship between advertiser and conversion fraud	0.009 < 0.05	Supported
H_2: There is a relationship between ad slot and conversion fraud	0.001 < 0.05	Supported
H_3: There is a relationship between item and conversion fraud	0.0007 < 0.05	Supported
H_4: There is a relationship between goal and conversion fraud	0.008 < 0.05	Supported
H_5: There is a relationship between ad log type and conversion fraud	0.048< 0.05	Supported
H_6: there is a relationship between pricing type and conversion fraud	0.006 < 0.05	Supported
H_7: There is a relationship between look up form and conversion fraud	0.754 > 0.05	Not supported

Source: Developed for the research

H_1: There is a positive relationship between advertiser and conversion fraud

Supported H_1, if p value < 0.05. Table 4.6 presents there is influential value of advertiser, which the result is 0.009 < 0.05. Therefore, H_1 is supported.

H_2: There is a positive relationship between ad slot and conversion fraud

Supported H_2, if p value < 0.05. Table 4.6 presents there is influential value of ad slot, which the result is 0.001 < 0.05. Therefore, H_2 is supported.

H_3: There is a positive relationship between item and conversion fraud

Supported H_3, if p value < 0.05. Table 4.6 presents there is influential value of item, which the result is $0.0007 < 0.05$. Therefore, H_3 is supported.

H_4: There is a positive relationship between goal and conversion fraud

Supported H_4, if p value < 0.05. Table 4.6 presents there is influential value of goal, which the result is $0.008 < 0.05$. Therefore, H_4 is supported.

H_5: There is a positive relationship between ad log type and conversion fraud

Supported H_5, if p value < 0.05. Table 4.6 presents there is influential value of ad log type, which the result is $0.048 < 0.05$. Therefore, H_5 is supported.

H_6: there is a positive relationship between pricing type and conversion fraud

Supported H_6, if p value < 0.05. Table 4.6 presents there is influential value of pricing type, which the result is $0.006 < 0.05$. Therefore, H_6 is supported.

H_7: There is a negative relationship between look up form and conversion fraud

Not supported H_7, if p value > 0.05. Table 4.6 presents there is influential value of look up form, which the result is $0.754 > 0.05$. Therefore, H_7 is not supported.

Chapter 5 DISCUSSION AND CONCLUSION

5.1 Discussion on 1st Research Objective – To identify the factors that led to conversion fraud in digital advertising

Based on the result in Chapter 4, all 7 variables used in this research are significant. However, p value does plays important role in determining significant variables. By using p value in determining process, there is 6 variables are significant and remaining 1 is not supported.

Advertiser, who is the victim in most of the cases of conversion fraud, advertiser itself is also the significant variable towards conversion fraud. Advertiser also refers to the firm, or the company. Although it is unlikely that advertiser itself would create conversion fraud intentionally that will damage their reputation. While weak management on brand image could lead to conversion fraud. For instance, if advertiser does not monitor it ad campaign closely, fraudster could take advantage by create fake ads and website to trick users that is domain spoofing (Zeiceva, 2022).

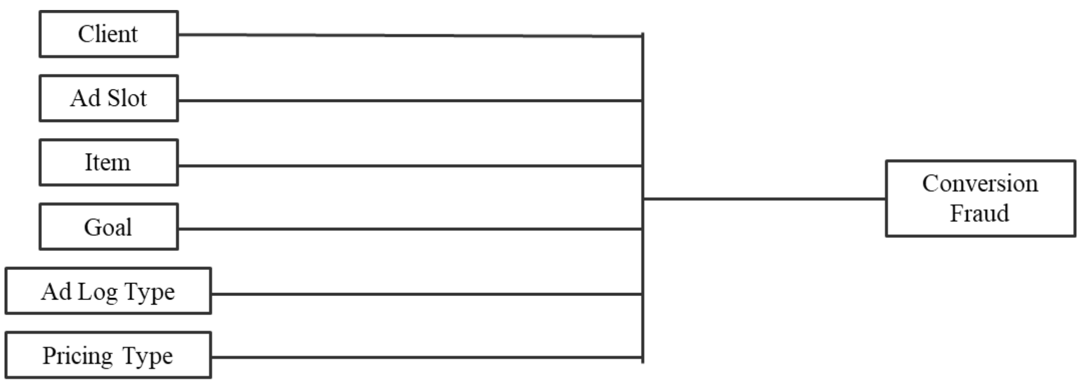
Ad slot is the space on website, advertiser charge based on different ad slot on websites, which indicates specific ad slot generates more data and more exposure to users. Pixel stuffing, ad stacking, and ad injection are the conversion fraud which fraudster use ad slot for fraudulent activities (Zeiceva, 2022). Due to specific ad slot can generate more data such as click and view.

Item is the ad creatives that display as an ad. Ad creative is important as it comprise message of a brand, product, or service, at the same time it has to be attractive to users. While it might be imitated by fraudster to generate click, or impression to trick not only user, but also advertiser (Stone-Gross, 2011).

Goal refers to the ad objectives setting by the advertiser which advertiser wish to achieve (Meta, n.d.). Failure of goal setting such as misleading goals or unrealistic goal setting for ad campaign provide fraudster loopholes to conduct fraudulent activities. For example, advertiser sets a goal of achieving large number of clicks within a short period of time, in this case, fraudster might manipulate the number of data that is not accurately reflect genuine users' action.

5.2 Discussion on 2nd Research Objective - To propose the prediction framework in determining the conversion fraud

Figure 5.2: Prediction Framework



Source: Developed for the research

Figure 5.2 shows the prediction framework of this study. There are 6 significant variables towards one dependent variables. A prediction framework is a system that enables researcher to use historical data to forecast future events, trends, or behaviours. It entails the application of statistical approaches and machine learning algorithms to the development of models that can be used to forecast future outcomes (Google Developers, n.d.).

Overall, predictive frameworks are effective tools for assisting organisations in making more informed decisions, optimising processes, and mitigating risks. However, it is critical to understand that predictive models are only as good as the data and assumptions that they are based on, and that they are not a replacement for human expertise and judgement, therefore, decision making process require critical thinking skill of advertiser.

5.3 Implications

5.3.1 Advertiser

This research could provide an insight to advertiser, in order to **enhance advertiser knowledge on digital advertising and advertisement setting**. Advertiser is always the victim in the entire ecosystem when there is conversion fraud occurs. According to the research by World Federation of Advertisers (WFS), there is estimated marketing spend lost due to conversion fraud, which is likely to reach $50 billion annually by 2025 (Lim, 2020). Prediction can lead advertiser in making strategic budget allocations to reduce conversion fraud.

"Prevention is always better than cure", although conversion fraud is a variable that could not controlled by firms, however, it is always better to have prediction of fraud. It is necessary to every business to study on prediction of fraud, it could help business to spend wisely on digital advertising to avoid conversion fraud. Prediction of conversion fraud could help business in saving advertising budget. Through prediction, advertiser can have more accuracy ad setting for ads, to reach higher ad performance with receiving valuable data. Moreover, it also protects brand image and reputation especially when there is domain spoofing occur. Furthermore, it helps advertiser in data driven decision making, advertiser could make better decision based on valuable data. In terms of ad setting, ad spend, targeting, and so on.

According to Intercept, 2022, it is critical for all parties involved, including advertisers, publishers, and platforms, to work together to tackle conversion fraud as well as protect their investment. Advertiser should stay updated on the latest news and use fraud detection technology and other fraud prevention platform to protect their investment from conversion fraud.

5.3.2 Regulator

Apart from that, this study could help regulator understand the loopholes in the current digital advertising environment. Conversion fraud has been harming the advertising industry and it has become a worldwide problem and expected to remain as an issue in the future. When there is conversion fraud happen, despite advertiser spending on advertisement, but in fact, their ad does not reach to their target audience (Lim, 2020).

The concept of digitalization started since 1950s, it has changed in almost all field in the way of how people work, shop, bank, etc (Press, 2015). However, the current rules and regulation for digital environment is remain unclear, and incomplete (Lim, 2020). Back in 2017, there is a lawsuit between Uber Technologies Inc. and Phunware, Inc. which is a mobile advertising company. Uber accusing there is breach of contract claiming $3 million in unpaid invoices. Uber response to Phunware, Uber accusing Phunware of wire fraud, racketeering, transporting, and other fraudulent activities, and seeking up to $17 million as compensation and punitive damages. After 4 years, Uber won the lawsuit, due to the proof of widespread and continuous fraud unveiled. It has found out that the ad activity involving fraudulent process that is click flooding, which they manipulated the number of clicks to charge higher cost to Uber. Moreover, there is auto-redirects, which malware automatically bring users to app store or play store, no matter user has acted of clicking on ads or not (Fou, 2021).

Through this study, it provides a more details information on conversion fraud which could help regulators to implement a tight anti-fraud standards, such as permitting only trusted sources to post ads, limiting campaign budgets, and routinely monitoring campaigns for suspicious activity (Intercept, 2022). In this study, there are 6 variables show significant towards conversion fraud, and among these variables. Regulators could propose a more completed rules and regulations towards these 6 variables to monitor and reduce conversion fraud in the future.

5.3.3 Academic

In the context of academic, it contributes cross disciplinary study regarding on the field of computing and business.

In research and education, the term "cross-disciplinary" refers to an integrative strategy in which various disciplines work together in any way. The value and necessity of cross-disciplinary learning, and many variations, including multidisciplinary, interdisciplinary, and transdisciplinary, have been proposed and supported. However, the effectiveness of cross-disciplinary learning has been hampered by traditional curricula, which are typically compartmentalised along conventional disciplinary lines (McPhillips, 2018).

Researchers mostly focusing on one field of study when doing complicated research. While cross-disciplinary study applied could provide a more deeply understanding and explanation of complex social challenges and community responses to issues in many fields (Youngs, 2020). This research using computing generate data, and process using OLS regression to find out the relationship between multiple independent variables and conversion fraud.

Conducting a cross-disciplinary study could also be very challenging. First, it is time costly on understanding different practices (Youngs, 2020). Effective communication and common language are also the common challenges in cross-disciplinary study (McPhillips, 2018). In this case, it takes time to understand between business study and computing coding study. While this topic is about digital advertising, marketing field, at the same time it is also a computing generated secondary data, which the result of the data process is using computing coding.

5.4 Limitations and Recommendation

During the progress of this research, it has found numerous limitations which could not have further progress during this research.

5.4.1 Coverage of Variables

The use of only 7 of the 55 variables that are available in this study is a limitation. This indicates that potentially crucial variables were left out of the analysis, which could have provided additional insight into the problem of conversion fraud. To solve this limitation, it is suggested that additional research be conducted by integrating other 48 variables in the analysis. This can be accomplished by collecting extra data or by utilizing other data sources that provide a more comprehensive set of variables. It has been speculated that by integrating additional variables, the study would cover a broader range of parameters that lead to conversion fraud, and the results will be more solid and credible. This can lead to a greater awareness of the problem of conversion fraud in digital advertising and assist marketers in developing more effective prevention tactics.

5.4.2 Sample Size

In this study, there is total 956 sample size used from secondary data. However, due to it is a computing generate data, it is suggested to have larger sample size using big data to conduct further research. Big data is defined as data with greater diversity, coming in higher volumes and with greater velocity (Oracle, n.d.). Structured, semi-structured, and unstructured data generated by humans and computers are all part of big data collection. By using big data collection, it is accurate in quantitative data (Pratt, 2022). In this study, the sample size is 956, which is not a small amount, however, it might be insufficient for numbers of variables in future study. Using big data in future studies may assist to solve the limitation of current study. Researchers can achieve more accurate and representative results with a bigger sample size, which can lead to a better knowledge and prediction of conversion fraud in digital advertising. Furthermore, big data can provide access to more diverse types of data, such as user behaviour data, social media data, and geographic data, which can aid in the identification of patterns and correlations that may not be visible in smaller data sets.

5.4.3 Analysis Method

Apart from that, the analysis methods used in this study is insufficient. In this study, Pearson correlation, Spearman correlation, and OLS regression, are used to examine seven hypothesises with the sample size 956 observations. Thus, it is suggested that to adapt more analysis methods to examine the data. Multiple analysis methods are preferable because they provide a more comprehensive understanding of the data and aid in the validation of the results. Each analysis approach has its own set of strengths and weaknesses, and by employing numerous methodologies, researchers can cross-validate their findings and acquire greater confidence in their conclusions. For example, if one analysis method yields a conclusion that contradicts the others, this could indicate a fault in the analysis or highlight an interesting feature of the data that needs additional research. Furthermore, adapting multiple analysis methods can aid in controlling for any potential biases or limitations inherent in a single method. Overall, employing numerous analysis methodologies can aid in increasing the rigour and validity of study findings.

Reference

Actian. (n.d.). Object Variables. Retrieved from https://docs.actian.com/dataconnect/11.5/index.html#page/User/Object_Variables.htm

Adjust. (n.d.). What is a mobile app advertiser? Retrieved from https://www.adjust.com/glossary/advertiser/

Animesh, A., Viswanathan, S., & Agarwal, R. (2007). Competing'Creatively'in Online Markets: Evidence from Sponsored Search. *Robert H. Smith School Research Paper No. RHS*, 06-064.

Bainchi, T. (2023). Advertising revenue of Google from 2001 to 2022. *Statista.* Retrieved from https://www.statista.com/statistics/266249/advertising-revenue-of-google/

Bashir, M. A., Arshad, S., Kirda, E., Robertson, W., & Wilson, C. (2019, October). A Longitudinal Analysis of the ads. txt Standard. In *Proceedings of the Internet Measurement Conference* (pp. 294-307).

BDC. (n.d.). The 5 most common pricing strategies. Retrieved from https://www.bdc.ca/en/articles-tools/marketing-sales-export/marketing/pricing-5-common-strategies

Borges, W. G., Ng, S. I., Chew, B. C., Lau, T. C., Derek, O. L. T., Devika, N., Lee, C., Goi, M. T., Law, K. A., Mohd Noor, M., Kuppusamy, S., & Hassan, H. (2020), Business Research Methods, (2nd ed), SJ Learning.

Burton, A. L. (2021). OLS (Linear) regression. *The Encyclopedia of Research Methods in Criminology and Criminal Justice*, *2*, 509-514.

Business Help Centre. (n.d.). Hpw to choose the right Meta Ads Manager objective. Retrieved from https://www.facebook.com/business/help/1438417719786914

Chen, G., Cox, J. H., Uluagac, A. S., & Copeland, J. A. (2016). In-depth survey of digital advertising technologies. *IEEE Communications Surveys & Tutorials*, *18*(3), 2124-2148.

Computer Hope. (2020). Floating-point. Retrieved from https://www.computerhope.com/jargon/f/floapoin.htm#:~:text=In%20program ming%2C%20a%20floating%2Dpoint,1.23%2C%2087.425%2C%20and%20 9039454.2

Computer Hope. (2021). Lookup table. Retrieved from https://www.computerhope.com/jargon/l/lookup-table.htm

Crussell, J., Stevens, R., & Chen, H. (2014, June). Madfraud: Investigating ad fraud in android applications. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services* (pp. 123-134).

Daswani, N., Mysen, C., Rao, V., Weis, S., Gharachorloo, K., & Ghosemajumder, S. (2008). Online advertising fraud. *Crimeware: understanding new attacks and defenses*, *40*(2), 1-28.

Data Dome (2022). What is bot fraud? What are scam bots? Your Guide to Bot-Driven Threats. Retrieved from https://datadome.co/learning-center/what-is-bot-fraud-what-are-scam-bots-guide-to-bot-driven-threats/

Driskill, J. (n.d.) Definition: Ad Slot. Retrieved from https://theonlineadvertisingguide.com/glossary/ad-slot/

Fou, A. (2021). One of Uber's lawsuits against ad fraud comes full circle – They won. *Forbes.* Retrieved from https://www.forbes.com/sites/augustinefou/2021/01/17/ubers-lawsuit-against-ad-fraud-comes-full-circlethey-won/?sh=2bf24aa54458

Fowler, E. F., Franz, M. M., Neumann, M., Ridout, T. N., & Yao, J. (2023, February). Digital Advertising in the 2022 Midterms. In *The Forum* (No. 0). De Gruyter.

Frost, J. (2021). Spearman's Correlation Explained. Retrieved from https://statisticsbyjim.com/basics/spearmans-correlation/

Gharibshah, Z., & Zhu, X. (2021). User response prediction in online advertising. *aCM Computing Surveys (CSUR)*, *54*(3), 1-43.

Ghosh, S., & Banerjee, C. (2020, September). A predictive analysis model of customer purchase behavior using modified random forest algorithm in cloud environment. In *2020 IEEE 1st International conference for convergence in engineering (ICCE)* (pp. 239-244). IEEE.

Google Ads Help. (n.d.) Retrieved from https://support.google.com/google-ads?sjid=35934604923049854-AP#topic=10286612

Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, *85*(1), 7-25.

Heinze, A., Fletcher, G., Rashid, T., & Cruz, A. (2016). Digital and social media marketing. *Abingdon: Routledge*.

Hyun, H., Thavisay, T., & Lee, S. H. (2022). Enhancing the role of flow experience in social media usage and its impact on shopping. *Journal of Retailing and Consumer Services*, *65*, 102492.

Indeed Editorial Team. (2023). 11 Goals of Advertising (and Why They're Important). Retrieved from https://www.indeed.com/career-advice/career-development/goals-of-advertising#:~:text=Advertising%20goals%20are%20what%20a,metrics%20to%20measure%20your%20success

Intercept. (2022). Mobile ad fraud | 2023 Predictions. Retrieved from https://interceptd.com/mobile-ad-fraud-2023-predictions/

Kumari, S., Yuan, X., Patterson, J., & Yu, H. (2017, October). Demystifying ad fraud. In *2017 IEEE Frontiers in Education Conference (FIE)* (pp. 1-5). IEEE.

Lim, J. (2020). Cover Story: Fighting digital ad fraud. *The Edge Market.* Retrieved from https://www.theedgemarkets.com/article/cover-story-fighting-digital-ad-fraud

Lumen. (n.d.) Reading: Advertising. Retrieved from https://courses.lumenlearning.com/clinton-marketing/chapter/reading-advertising/#:~:text=Advertising%20has%20three%20primary%20objectives,of%20new%20or%20established%20products.

Ma, J., & Du, B. (2018). Digital advertising and company value: Implications of reallocating advertising expenditures. *Journal of Advertising Research*, *58*(3), 326-337.

Mahmoud, H. (2022). What is Kaggle? Retrieved from https://www.kaggle.com/general/328265

Malik, M. E., Ghafoor, M. M., & Iqbal, H. K. (2014). The impact of advertisement and consumer perception on consumer buying behavior. *International Review of Social Sciences and Humanities*, *6*(2), 55-64.

Malik, M. E., Ghafoor, M. M., Iqbal, H. K., Ali, Q., Hunbal, H., Noman, M., & Ahmad, B. (2013). Impact of brand image and advertisement on consumer buying behavior. *World Applied Sciences Journal*, *23*(1), 117-122.

Marwan, A., & Bonfigli, F. (2022). Detection of Digital Law Issues and Implication for Good Governance Policy in Indonesia. *Bestuur*, *10*(1), 22-32.

Mattingly, J. (2023). What are interger variables? Retrieved from http://support.kodable.com/en/articles/417314-what-are-integer-variables

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., ... & Shafiei Shiva, J. (2018). Defining extreme events: A cross-disciplinary review. *Earth's Future*, *6*(3), 441-455.

Meta. (n.d.). Ad objectives. Retrieved from https://www.facebook.com/business/ads/ad-objectives

Microsoft. (2022). Configure a lookup component on a form. Retrieved from https://learn.microsoft.com/en-us/power-apps/maker/model-driven-apps/form-designer-add-configure-lookup

Minastireanu, E. A., & Mesnita, G. (2019). Light gbm machine learning algorithm to online click fraud detection. *J. Inform. Assur. Cybersecur, 2019*, 263928.

Oberlo. (n.d.) Cost Per Action (CPA). Retrieved from https://www.oberlo.com/ecommerce-wiki/cost-per-action

Oracle. (n.d.) What is big data? Retrieved from https://www.oracle.com/my/big-data/what-is-big-data/#:~:text=Big%20data%20defined,-What%20exactly%20is&text=Put%20simply%2C%20big%20data%20is,just%20can't%20manage%20them.

Paddle. (n.d.). The pricing strategy guide: Choosing pricing strategies that grow (not sink) your business. Retrieved from https://www.paddle.com/resources/pricing-strategy

Patil, A. S., Navalgund, N. R., & Mahantshetti, S. (2022). Digital Marketing Adoption by Start-Ups and SMEs.

Pratt, M, K. (2022). How big data collection works: Process, challenges, techniques. Retrieved from https://www.techtarget.com/searchdatamanagement/feature/Big-data-collection-processes-challenges-and-best-practices#:~:text=Big%20data%20collection%20is%20the,and%20make%20critical%20business%20decisions.

Press, G. (2015). A very short history of digitalization. *Forbes*. Retrieved from https://www.forbes.com/sites/gilpress/2015/12/27/a-very-short-history-of-digitization/?sh=167d84349ac2

Publift. (2023) Ad Fraud: Everything you need to know. Retrieved from https://www.publift.com/blog/ad-fraud#:~:text=Geo%20Masking&text=Fraudsters%20use%20this%20type%20of,sought%2Dafter%20country%20or%20region.

Rapid7. (2016). The 4 steps for creating a log enabled marketing campaign. Retrieved from https://www.rapid7.com/blog/post/2016/03/17/the-4-steps-for-creating-a-log-enabled-marketing-campaign/#:~:text=Each%20time%20a%20marketing%20event,sent%20is%20called%20a%20collector.

Rasool, M. S., Rafique, Y., Naseem, M. A., Javaid, S., Najeeb, M., & Hannan, M. (2012). Impact of advertisement on consumer behavior of FMCG in Lahore city. *Academic Research International*, *2*(3), 571.

Religia, Y., Surachman, S., Rohman, F., & Indrawati, N. (2021, January). E-commerce adoption in SMEs: A literature review. In *Proceedings of the 1st International Conference on Economics Engineering and Social Science, InCEESS 2020, 17-18 July, Bekasi, Indonesia*.

Sadeghpour, S., & Vlajic, N. (2021). Ads and Fraud: A Comprehensive Survey of Fraud in Online Advertising. *Journal of Cybersecurity and Privacy*, *1*(4), 804-832.

Sadeghpour, S., & Vlajic, N. (2021). Click fraud in digital advertising: A comprehensive survey. *Computers*, *10*(12), 164.

Saleem, S., & Abideen, Z. (2011). Effective advertising and its influence on consumer buying behavior. *European Journal of Business and Management*, *3*(3), 55-67.

Shaari, H., & Ahmed, N. (2020). An extensive study on online and mobile ad fraud.

Sharif, A. (2022). Log files Explained. Retrieved from https://www.crowdstrike.com/cybersecurity-101/observability/log-file/

Stephen, A. T. (2016). The role of digital and social media marketing in consumer behavior. *Current opinión in Psychology*, *10*, 17-21.

Stone-Gross, B., Stevens, R., Zarras, A., Kemmerer, R., Kruegel, C., & Vigna, G. (2011, November). Understanding fraudulent activities in online ad exchanges. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference* (pp. 279-294).

Suresh, S., Di Troia, F., Potika, K., & Stamp, M. (2019). An analysis of Android adware. *Journal of Computer Virology and Hacking Techniques*, *15*, 147-160.

Tantawi, R. (2023). Secondary data. *Salem Press Encyclopedia.*

Thomas, K., Bursztein, E., Grier, C., Ho, G., Jagpal, N., Kapravelos, A., ... & Rajab, M. A. (2015, May). Ad injection at scale: Assessing deceptive advertisement modifications. In *2015 IEEE Symposium on Security and Privacy* (pp. 151-167). IEEE.

Wigmore, I. (n.d.). Log (log file) Retrieved from https://www.techtarget.com/whatis/definition/log-log-file#:~:text=A%20log%2C%20in%20a%20computing,and%20systems%20produce%20log%20files.

Wuisan, D. S., & Handra, T. (2023). Maximizing Online Marketing Strategy with Digital Advertising. *Startupreneur Bisnis Digital (SABDA Journal)*, *2*(1), 22-30.

Youngs, A. (2020). Why cross-disciplinary research matters: An interview with Amy Youngs. Retrieved from https://globalartsandhumanities.osu.edu/news/why-cross-disciplinary-research-matters-interview-amy-youngs

Zaiceva, A. (2022). 7 Types of Ad Fraud and How to Prevent Ad Fraud? Retrieved from https://setupad.com/blog/types-of-ad-fraud/#:~:text=There%20are%207%20main%20types%20of%20ad%20fraud%3A&text=Domain%20Spoofing,Geo%20Masking

Zhu, X., Tao, H., Wu, Z., Cao, J., Kalish, K., & Kayne, J. (2017). Ad fraud categorization and detection methods. *Fraud Prevention in Online Digital Advertising*, 25-38.

Zhu, X., Tao, H., Wu, Z., Cao, J., Kalish, K., Kayne, J., ... & Kayne, J. (2017). Ad fraud categorization and detection methods. *Fraud Prevention in Online Digital Advertising*, 25-38.