A THESIS REPORT

On

Association Rule Mining for Finding the Relationship between Each Category of Work,Pricing & Sales of Gigs in Fiverr Marketplace

Submitted By

MD. Samim Ferdous

16CSE037

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Bangabandhu Sheikh Mujibur Rahman Science and Technology University

January,2022

Association Rule Mining for finding the relationship between each category of work, pricing & sales of gigs in fiverr marketplace

Submitted By
Md. Samim Ferdous
16CSE037

Supervised by
F. M. Rahat Hasan Robi
Assistant Professor
Department of Computer Science & Engineering
Bangabandhu Sheikh Mujibur Rahman Science and Technology University

Thesis Approval

Students Name: Md. Samim Ferdous

Student's ID: 16CSE037

Thesis Title: Association Rule Mining for Finding the Relationship between Each Category of Work,Pricing & Sales of Gigs in Fiverr Marketplace

I the undersigned, recommend that the the thesis completed by the student listed above, in partial fulfillment of Bachelor of the requirements for the Degree of Bachelor of Science in Computer Science and Engineering be accepted by the Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University for deposit.

Supervisor Approval

....
F. M. Rahat Hasan Robi
Assistant Professor

Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj

STATEMENT OF ORIGINITY

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. Have acknowledged all main sources of help.

Signature of the Candidate
Date:

Signature of the Supervisor
Date:

Dedicated
to
My Parents

Acknowledgement

It gives me a great sense of pleasure to present the report of the thesis undertaken during BSc Engineering in CSE fourth year second semester. The research in itself is an acknowledgement to the inspiration, drive and technical assistance contributed to it by many individuals. This research would never have seen the light of the day without the help and guidance that I have received.
Firstly, all praise to the Great Allah for whom my thesis have been completed without any major interruption.
Secondly, I owe special debt of gratitude to F. M. Rahat Hasan Robi, Assistant Professor Department of CSE, for his constant support and guidance throughout the course of my work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for me. He has showered me with all his extensively experienced ideas and insightful comments at virtually all stages of the project & has also taught us about the latest industry-oriented technologies. He helped us whenever I needed.
Thirdly,I also do not like to miss the opportunity to acknowledge the contribution of google, wiki, stackoverflow. These websites helped me a lot during the development of this project.
And finally to my parents without their throughout sup-port it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Abstract

Fiverr has become a large marketplace for selling/buying small services. Jobs listed in fiverr are classified into different categories and subcategories. Those who are willingly to join fiverr may need to know if there is available relationships between this categories and subcategories. Also for a particular service if pricing has any influence on sales/rating or vice-versa. The purpose of this study is to extract the maximum amount of information and providing an explained analysis of the available relations, which may be very useful for people inside and outside of fiverr. Association rule mining, an unsupervised learning method is used in this research to find the relationships between each category and subcategory of work. In addition, Pearson Correlation coefficient is applied to explore available correlation between pricing and sales/rating and vice-versa. Finally, several association rules between services was gained within a given criteria. From this research, it will be very useful for people to decide whether he/she should develop a new skill to join freelancing or to stick with his/her current job. In addition, for a particular service it would be easier to fix the best price to increase sales.

Keywords: fiverr; freelancing; online marketplace; association rule mining; apriori

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Overview

1.1 Introduction

Online freelancing marketplaces have grown quickly in recent years [8]. Fiverr is one of the largest of them which is an Israeli digital platform for freelance services was founded by Micha Kaufman and Shai Wininger. The founders came up with the concept of a online marketplace that would provide a two-sided market for people to buy and sell a variety of digital services which are typically offered by freelance contractors[14]. Jobs listed on the platform are diverse and classified into several categories and sub-categories. Fiverr offers services like Graphics & Design, Digital Marketing, Writing & Translation, Video & Animation, Music & Audio, Programming & Tech, Business and Lifestyle which are divided into several sub-categories. These categories bring global clients and the freelancers together, helps to decrease unemployment rate, and improve their lives [10].

With the widespread access and use of Internet, nowadays many people switch to freelance career and join online marketplace like fiverr day by day. Factors like personal and family life, side interests and hobbies, the desire to travel and many more are reasons to choose freelancing. Also work-life balancing, Desire to choose where they work and when they work, be their own boss, freedom, choosing projects that suit themselves, being more productive and time saving are the other factors. Freelancers who provide selling services in fiverr work in a variety of workplaces, ranging from home to office. Each service offered is called a "gig". Fiverr's services start at 5 USD and can go up to thousands of dollars with gig extras[14].

Many of the freelancers who want to join fiverr has limited knowledge. The need for freelancers with curremt skill set is of constant concern. Before joining fiverr many questions may arise in a freelancer mind like whether he or she is enough skilled for choosing this track or they should develop new skills to do well in the marketplace. Also who are already in fiverr selling their services may want to know how they can increase sales, providing which services along with their current ones may grow their earnings .If providing services in combine with others increase the sales then which are them! Like if certain type of services has association in between that may increase sales. For example providing Digital Marketing services with Writing and Translation or if we say specifically like providing article writing service with Search Engine Optimization may increase their sales cause the buyer may need these services at the same time. There may be hundreds of them like these those have relationship between them. Also for a certain service if decreasing the price always increase the sales/rating or not likely if there has correlation between them or not. So this research is very much needed and has a lot of significance.

Since selling small services is a raising business domain, it is wise to think about investing time and effort in this market if anyone has a value to share.This research is likely a proper market study of the fiverr marketplace. The aim is to provide explained analysis of the available relations between each category and subcategory of work which are listed in fiverr. Also the correlation between pricing and sales for each category, the correlation between and ratings for each category and the correlation between ratings and sales need to be found. It will be an overview about what is going on out there. Will answer many questions to those who are willingly to join fiverr and also who are currently selling their small services in fiverr. In this way this work will be too much useful for the sellers.

Association Rule Mining, which is an unsupervised learning method of machine learning, is used for this research which discover frequent patterns , relationship among a set of items in the database [17]. Apriori algorithm of association rule which is widely used for finding frequent itemsets [6]provide a new dimension in labour market research and Pearson correlation test for finding correlation between several variables have been used in this research, will be discussed in Methodology section. By analyzing the data the relationships between the several variables like category, sub-category we find several services that has association between them which will be discussed in Results & Discussion section.

1.2 Motivation and Aim

Basically a lot of people are out there who wants to join fiverr but may not have vast knowledge. They may want to know which skills have high demand and if demeaned ones matches with their skill. And then it will certainly help themselves to understand whether they should have develop a new skill or to stick with their previous work. Selling services that has association in between can certainly increase their sales. Also seller inside fiverr may want to know if pricing has any influence on sales or ratings and vice-versa. So this research is needed to find out the answer of these questions for people inside and outside of fiverr.

1.3 Objectives

The main objectives include finding the association and correlation between variables. The specific objectives I tried to find includes finding the association between each category and subcategory of work that are listed in fiverr marketplace. And another one is to find the correlation between prices ->sales, prices ->ratings and ratings ->sales for each category of work.

Chapter 2

Related Works

In the past few years a lot of works have been done mainly the job recommender system for freelancers. Sabir et al. [11] propose a recommender system to find out appropriate jobs for freelancers using client's feedback classification and Association rule mining technique. After collecting the previous work history of freelancers, they analyze the sentiment of client's feedback. Then, they apply the Association rule mining technique to find out freelancer's frequent skillsets used in both categories of completed jobs. However, they did not show that which category of works are related in online marketplace. Like if, a freelancer has gigs (services) in graphics design category how much possibility increases if he adds a gig in digital marketing category or if the category of work has any relations with pricing or sales, which we are supposed to show in my research.

Frederick at el. [12] proposed a methodology for identifying and analyzing skill-job relationships using frequency word occurrences of skills as a requirement of the job. It employed association rule mining, which aims to discover frequent patterns, relationships among a set of items in the database. Their research, could provide insights on the gap between the school acquired skills and actual IT industry skill needs and as the basis for curriculum enhancement and policy-making interventions by the Philippine government in its educational system. They did the research for finding skill-job relationship and our research is based on finding associations between skills.

Zheng et al. [7] proposed a method to match buyers-freelancers where they used bid pricing dispersion for removing buyer indecision and freelancer regret.

Chapter 3

Data & Methods

3.1 Data Collection

The data has been collected by a scraping tool. The tool has run for a huge time and managed to crawl the necessary data needed. The scraper has also generated some unusual information too, like the visited links. Therefore, after collecting the data cleaning has been done. Later the dataset has been preprocessed. In this manner, it was able to collect 1500 freelancers services which they offer in fiverr, mainly the category and sub-category of work and each gig's price, number of sales and the ratings they got in that particular gig/service.

3.2 Structured Dataset Construction

From the scrapped data two separate xlxs file is made. One include the profile name of the seller, category and sub-category of services they provide. Another file include the gigs category, subcategory and their pricing, sales and ratings. Dataset is split and properly wrangled to perform association rule mining in one file and to find out correlation between price and rating/sales or vice versa on the other. The first ten raw's of the separate files are cataloged in table 3.1 and table 3.2.

3.3 Association Rule Mining

Extracting important and hidden information from a large dataset by mining association rules is one of the most common tasks in data mining [9]. The association rule mining can be described as a two-step process [1].

- Generating frequent item sets—find all frequent item sets whose support value is equal to or greater than the minimum support value;
- Generating association rules—generate association rules from frequent item sets under the condition of minimum confidence.;

The association rules mining algorithms include Apriori[3], SETM [2], MAFIA [5], and Pincer Search [4], which are based on support-confidence framework proposed by Agrawal and Srikant. The Apriori algorithm is succinct and clear, which adopts an iterative method of layer-by-layer search. In this study, the Apriori algorithm was used to discover the significant rules between the factors. The figure 3.1 shows the process the Association Rule Mining.

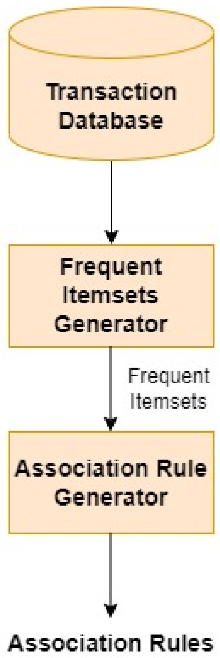

Figure 3.1: Association Rule Mining Process

3.4 Association Measures

There are three association measures generally. By which we may be find out associations. They are support, confidence and lift. Before applying the Apriori algorithm, the threshold values of support and confidence should be taken into consideration. These thresholds are nothing but a minimum criterion that will be used in pruning for picking the popular Itemsets and strong association rules.

Support: Support measures the frequency of association. It is calculated by,

$$Support(X \Rightarrow Y) = \frac{NumberofrecordswithXandY}{Totalnumberofrecordsinitemset}$$

Confidence: Confidence is the strength of association. It is calculated by,

$$Confidence(X \Rightarrow Y) = \frac{NumberofrecordswithXandY}{NumberofrecordswithX}$$

Lift: Lift is nothing but a ratio of confidence and expected confidence. The equation of Lift is,

$$Lift(X) = \frac{Confidence(X \Rightarrow Y)}{ExpectedConfidence}$$

Table 3.1: Sellers with corresponding category, subcategory of work they provide (Top 10 rows)

Fiverr profle	**Category**	**Subcategory**
aaimran	Data,Business	Data Processing,Data Entry,E-Commerce Management
aaliyaan	Programming & Tech,Digital Marketing	E-Commerce Development,WordPress,Video Marketing,Music Promotion,Search Engine Optimization (SEO),Search Engine Marketing (SEM)
abantikabose	Digital Marketing,Writing & Translation	Book & eBook Marketing,Book & eBook Writing,Book Editing,Beta Reading
abdelhamid19	Digital Marketing	Web Analytics,Web Traffic
abidhussain1918	Digital Marketing,Data,Business	Influencer Marketing,Data Processing,Data Entry,Sales
abigaelseptembe	Programming & Tech,Writing & Translation	Online Coding Lessons,Proofreading & Editing
abihere	Digital Marketing,Programming & Tech	Search Engine Optimization (SEO),WordPress,Web Analytics
abuzar_mudassar	Digital Marketing,Writing & Translation,Programming & Tech	Search Engine Optimization (SEO),Website Content,WordPress
activecomputech	Graphics & Design	Photoshop Editing,Flyer Design,Signage Design,Business Cards & Stationery,Logo Design,Social Media Design,Brochure Design
adanab	Graphics & Design	Catalog Design,Other,Brand Style Guides,Logo Design,Packaging & Label Design

Table 3.2: Gigs with their corresponding category, subcategory and other values (Top 10 rows).

Category	Subcategory	Price	Sales	Stars
Programming & Tech	Data Analysis & Reports	155	10	5
Lifestyle	Greeting Cards & Videos	5	1000	5
Programming & Tech	Website Builders & CMS	75	68	5
Lifestyle	Cooking Lessons	20	5	4.9
Writing & Translation	Legal Writing	10	2	5
Writing & Translation	UX Writing	5	0	0
Writing & Translation	Podcast Writing	10	0	0
Writing & Translation	Email Copy	35	57	5
Writing & Translation	Legal Writing	55	309	5
Digital Marketing	Local SEO	100	4	5

Table 3.3: Test case for Installation

Serial No.	Test Case ID	Test Description	Input test data	Expected Result	Actual Result	Remarks
1	TC-INS-01	Install Find My Thing app in android phone	Transfer Find My Thing app	Open application with it's splash screen	Application executed with splash screen	Pass

3.5 Apriori Algorithm

An unsupervised learning algorithm is Apriori Algorithm. From the given data set it generates association rules. If an item occurs Association rule implies. With a certain probability, then item B also occurs. By IF THEN format most of associate rules are generated. It is an important concept for finding relations between two variables in a large database. Now the working steps of Apriori is described below,

Step 1: A. Create 1-Itemset candidates and calculate support for all the items. B. Perform pruning to create L1 Frequent Itemset. In pruning, we will filter out all items with Support less than the minimum support value.

Step 2: A. Create 2-Itemset candidates from L1 Frequent Itemset and calculate support for all of them. B. Perform Pruning to create L2 Frequent Itemset. As before, we will again filter out all the Itemsets with Support less than minimum support value.

A. Create 3-Itemset candidates from L2 Frequent Itemset and calculate support for all of them. B. Perform pruning to create L3 Frequent Itemset. But here if you see, Support is less than minimum support value for all the 3-itemsets so we cannot go any forward and need to find Association Rules from L2 Frequent Itemset only.

Final Step: Create Association Rules A. We need to calculate the Confidence for all combinations of items in the L2 Frequent Itemset. B. Perform pruning for again, this time to filter out all those association rules that have Confidence(%) less than the min_conf.

Additionally,Lift values were also calculated to better understand the kind of impact this association rule is going to make on the sale of individual items. However, we are not using it as a rule selection criteria in this research.

So, as far as we discussed, the Apriori algorithm was developed by Agrawal and Srikant [1] and has become a classical technique for discovering frequent item sets. It uses a bottom-up iterative method, improving the computing efficiency because of its connection steps and pruning steps. The Apriori algorithm usually involves two steps: in the first step, given minSup, all of the frequent items in the database are filtered out; in the second step, given minConf, all of the strong association rules are mined based on the frequent item results.

In this research while choosing the support of categories and subcategories of work we choose support 0.024 while calculating the association of categories and support 0.020 for the association between the subcategories. We choose lower support [13] [15] because in the dataset the number of different items are too many. They appear less frequently. While for categories we took the frequency of the lowest categories that is Lifestyle in consideration. And for Subcategory we decided to choose a support criteria from which we can get a number of 70 itemsets for generating strong association rules later. The complete working procedure is defined by a flowchart in figure 3.2 .

3.6 Pearson Correlation Coefficient

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables[16]. It is known as the best method of measuring the association between variables of interest because it is based on the method of co variance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship. We conduct this test to find out correlation between the variables pricing, sales and the ratings. By which, we can know for increasing 1 US Dollar how many sale is increasing or decreasing also how many ratings is increasing or decreasing for a certain category of work in fiverr. For increasing one sale the ratings the freelancer's received is increasing or decreasing. And of course the vice-versa of the correlation described can also be found. Below the formula of calculating the Pearson Correlation Coefficient is given. If we think all the pricing as a variable x_i and all the sales as a y_i variable then we calculate the correlation in between pricing and sales by the below formula stated below. For this the mean value of both x and y variable is also needed.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x}) \sum(y_i - \bar{y})}}$$

$$\begin{aligned} where, & \\ r &= correlation coefficient \\ x_i &= values of x variable in a sample \\ \bar{x} &= mean of the values of x variable \\ y_i &= values of y variable in a sample \\ \bar{y} &= mean of the values of the y variable \end{aligned}$$

3.6.1 Properties of Correlation

Limit: Coefficient values can range from +1 to -1, where +1 indicates a perfect positive relationship, -1 indicates a perfect negative relationship, and a 0 indicates no relationship exists.

Pure number: It is independent of the unit of measurement. For example, if one variable's unit of measurement is in inches and the second variable is in quintals, even then, Pearson's correlation coefficient value does not change.

Symmetric: Correlation of the coefficient between two variables is symmetric. This means between X and Y or Y and X, the coefficient value of will remain the same.

3.6.2 Degree of correlation

Perfect: If the value is near ± 1, then it said to be a perfect correlation: as one variable increases, the other variable tends to also increase (if positive) or decrease (if negative).

High degree: If the coefficient value lies between ± 0.50 and ± 1, then it is said to be a strong correlation.

Moderate degree: If the value lies between ± 0.30 and ± 0.49, then it is said to be a medium correlation.

Low degree: When the value lies below + .29, then it is said to be a small correlation.

No correlation: When the value is zero.

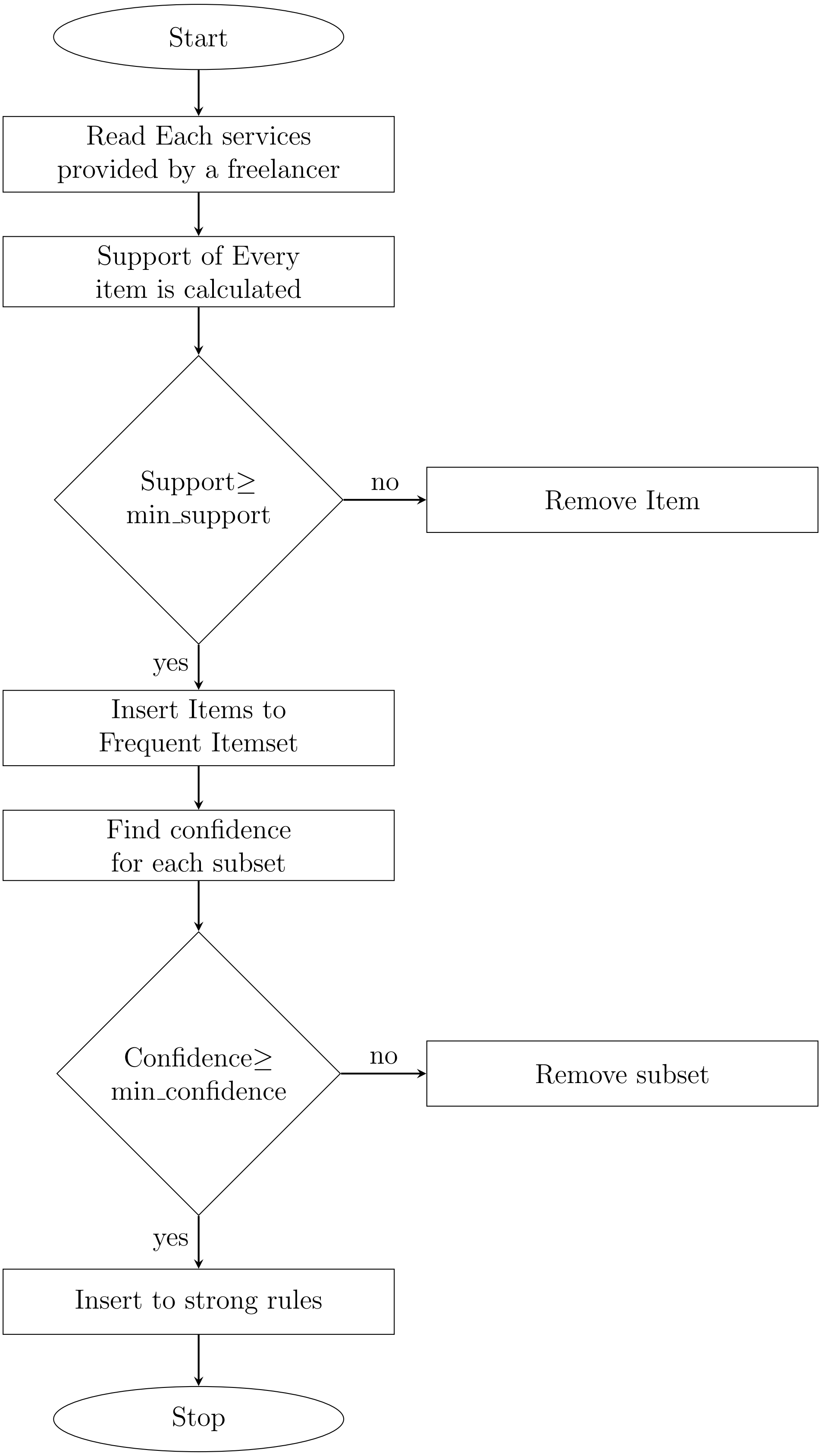

Figure 3.2: Apriori Algorithm Working Procedure

Chapter 4

Results and Discussion

4.1 Rules Mining Using Apriori Algorithm

The Apriori algorithm is used in association rule mining to explore relationships among all category and subcategory of work and identify the most suitable indices that have the strongest associations within the study area.

First, the Apriori algorithm was used to explore frequent item sets. Based on the dataset constructed, this study found that the mining results were satisfactory when minSup = 0.024 was used for the category indices and 0.20 was used for the subcategory indices. Frequent item sets for both the category and subcategory inddices is listed in table 4.1, table 4.2 and table 4.3.

Second,the Apriori algorithm was further applied to investigate strong association rules in the database, and by setting minSup = 0.024 and minConf = 0.50 for the category indices & minSup = 0.020 and minConf = 0.50 for the subcategory indices. From these criteria, 7 strong association rules were generated for the each category of work and 15 rules were generated for each subcategory. Generated strong association rules are listed in table 4.4 and table 4.5. In both the cases confidence 50% was used as a criterion for the minimum support value. Distribution of the rules using scatter plots are presented in Figure 4.1 and 4.4. Heatmaps of these rules for each category and subcategory are visualised in Figure 4.2 and Figure 4.5. Figure 4.3 and 4.6 visualizes the strong association rules for each category and subcategory with a network tree.

4.1.1 Frequent Item Sets for the Category Indices

As listed in table 4.1, 33 frequent item sets were found in total and the support indicates the frequency of each category of work in the dataset. Out of the 32 we got 9 frequent item sets that were the single ones and 23 multiple indices of which 17 were double and rest are triple indices. The results show that the top three frequent item sets of single indices Digital Marketing,Graphics & Design and Programming & Tech were accounting 37.1%,37.2% and 39% respectively. Furthermore,Programming & Tech, Graphics & Design and Digital Marketing,Writing & Translation accounted for the largest percentage among multiple frequent item sets having 10.8% and 11% respectively which indicates that these two category services are usually provided together by the freelancers.

Table 4.1: Frequent item sets for the category indices

Frequent Item Set (Single Indices)	**Support**	**Frequent Item Set (Multiple Indices)**	**Support**	**Frequent Item Set (Multiple Indices)**	**Support**
Business	0.154	Business, Digital Marketing	0.074	Programming & Tech, Graphics & Design	0.108
Data	0.174	Business, Graphics & Design	0.044	Video & Animation, Graphics & Design	0.076
Digital Marketing	0.371	Business, Programming & Tech	0.054	Graphics & Design, Writing & Translation	0.052
Graphics & Design	0.372	Business, Writing & Translation	0.052	Programming & Tech, Video & Animation	0.030
Lifestyle	0.024	Digital Marketing, Data	0.044	Programming & Tech, Writing & Translation	0.058
Music & Audio	0.046	Graphics & Design, Data	0.024	Business, Programming & Tech, Data	0.028
Programming & Tech	0.390	Programming & Tech, Data	0.088	Business, Programming & Tech, Digital Marketing	0.026
Video & Animation	0.144	Data, Writing & Translation	0.040	Business, Writing & Translation, Digital Marketing	0.032
Writing & Translation	0.218	Graphics & Design, Digital Marketing	0.076	Programming & Tech, Graphics & Design, Digital Marketing	0.036
		Programming & Tech, Digital Marketing	0.146	Digital Marketing, Graphics & Design, Writing & Translation	0.030
		Video & Animation, Digital Marketing	0.026	Digital Marketing, Programming & Tech, Writing & Translation	0.034
		Digital Marketing, Writing & Translation	0.110		

4.1.2 Frequent Item Sets for the Subcategory Indices

As listed in table 4.2 and 4.3, 80 frequent item sets were found in total and the support indicates the frequency of each category of work in the dataset. Out of the 80 item sets we got 59 frequent item sets that were the single ones and 21 multiple indices. The results show that the top three frequent item sets of single indices Other, Data Entry, Article & Blog Posts were accounting 9.6%,9.2% and 8.2% respectively. Furthermore, Search Engine Optimization(SEO), Wordpress; Web Programming & Mobile Apps;Virtual Assistant, Data Entry and Brand Stylle Guides, Logo Design accounted for the largest percentage among multiple frequent item sets having 3%, 3%, 3.2% and 4.6% respectively which indicates that these four subcategory of services are usually provided together by the freelancers.

4.1.3 Association Rules between Each Category of Work

There are 7 strong association rules generated between each category of work. During the generation of these rules the support value was chosen as 0.24% for frequent itemset generation and the confidence value was chosen as 50% for strong rules generation criteria.

Out of these 7 rules the top two were Writing & Translation ->Digital Marketing and Data ->Programming & Tech accounting support of 11.0% and 8.8% respectively. On the contrary the rules Business, Programming & Tech ->Data and Graphics & Design, Writing & Translation ->Digital Marketing are the ones having lowest support of 2.8% and 3.0% respectively.

On the otherhand the rules Business,Writing & Translation ->Digital Marketing and Programming & Tech,Writing & Translation ->Digital Marketing having the confidences of 61.54% and 58.62% respectively. We can see Digital Marketing a quiet of times in consequent part of these rules which indicates that it is used most frequently with the others category of work.

It is quite usual of appearing Digital Marketing as consequent most of the times cause nowadays each and every works need online marketing referred as Digital Marketing.

Using the support and confidence Figure 4.1 shows the distribution of rules where 7 blue scatter points represents 7 rules. Figure 4.2 shows the heatmaps of rules which help to understand a large number of rules between a small number of antecedent and consequent. In Figure 4.3 we can see several trees, rules are visualised in this figure with a network tree. Below is the table 4.4 which represents the rules extracted from the frequent item sets.

Table 4.2: Frequent item sets for the Subcategory(Single indices)

Frequent Item Set	**Support**	**Frequent Item Set**	**Support**	**Frequent Item Set**	**Support**
App Design	0.028	Search Engine Optimization (SEO)	0.108	Flyer Design	0.054
Articles & Blog Posts	0.082	Search Engine Marketing (SEM)	0.034	Sales	0.034
Book & eBook Marketing	0.062	Short Video Ads	0.036	Game Art	0.022
Book & eBook Writing	0.030	Social Media Advertising	0.036	Game Development	0.030
Book Design	0.042	Social Media Design	0.078	Graphics for Streamers	0.032
Brand Style Guides	0.050	Social Media Marketing	0.116	Illustration	0.034
Brochure Design	0.032	Support & IT	0.040	Influencer Marketing	0.020
Business Cards & Stationery	0.040	T-Shirts & Merchandise	0.028	Intros & Outros	0.022
Business Consulting	0.022	Translation	0.038	Logo Design	0.162
Content Marketing	0.032	Vector Tracing	0.022	Market Research	0.032
Convert Files	0.032	Video Editing	0.028	Proofreading & Editing	0.030
Data Entry	0.092	Virtual Assistant	0.052	Product Descriptions	0.024
Data Processing	0.034	Voice Over	0.034	Presentation Design	0.020
Data Science	0.054	Web Analytics	0.072	Photoshop Editing	0.030
Data Visualization	0.034	Web Programming	0.066	Packaging & Label Design	0.020
Databases	0.024	Website Builders & CMS	0.034	Other	0.096
Desktop Applications	0.038	Website Content	0.036	Mobile Apps	0.068
E-Commerce Development	0.068	Website Design	0.030	Mobile App Marketing	0.054
E-Commerce Management	0.030	Whiteboard & Animated Explainers	0.060	Marketing Strategy	0.038
E-Commerce Marketing	0.038	WordPress	0.150		

Table 4.3: Frequent item sets for the Subcategory (Multiple Indices)

Frequent Item Set	Support	Frequent Item Set	Support	Frequent Item Set	Support
Articles & Blog Posts, Website Content	0.022	E-Commerce Development, WordPress	0.034	Web Programming, Mobile Apps	0.030
Brand Style Guides, Logo Design	0.046	Flyer Design, Logo Design	0.032	Search Engine Optimization (SEO), WordPress	0.030
Flyer Design, Brochure Design	0.026	Social Media Design, Flyer Design	0.028	Whiteboard & Animated Explainers, Short Video Ads	0.022
Logo Design, Business Cards & Stationery	0.024	Graphics for Streamers, Logo Design	0.022	Social Media Design, Social Media Marketing	0.020
Sales, Data Entry	0.026	Social Media Design, Logo Design	0.034	Social Media Marketing, WordPress	0.024
Virtual Assistant, Data Entry	0.032	T-Shirts & Merchandise, Logo Design	0.020	Web Programming, WordPress	0.026
Data Visualization, Data Science	0.020	Social Media Marketing, Marketing Strategy	0.022	Website Builders & CMS, WordPress	0.022

Table 4.4: Association Rules Between Categories

antecedent	consequent	antecedent support	consequent support	support	confidence	lift
Data	Programming & Tech	0.174	0.390	0.088	0.505747	1.296788
Writing & Translation	Digital Marketing	0.218	0.374	0.110	0.504587	1.349164
Video & Animation	Graphics & Design	0.144	0.372	0.076	0.527778	1.418757
Business, Programming & Tech	Data	0.054	0.174	0.028	0.518519	2.979991
Business, Writing & Translation	Digital Marketing	0.052	0.374	0.032	0.615385	1.645413
Graphics & Design, Writing & Translation	Digital Marketing	0.052	0.374	0.030	0.576923	1.542575
Programming & Tech, Writing & Translation	Digital Marketing	0.058	0.374	0.034	0.586207	1.567398

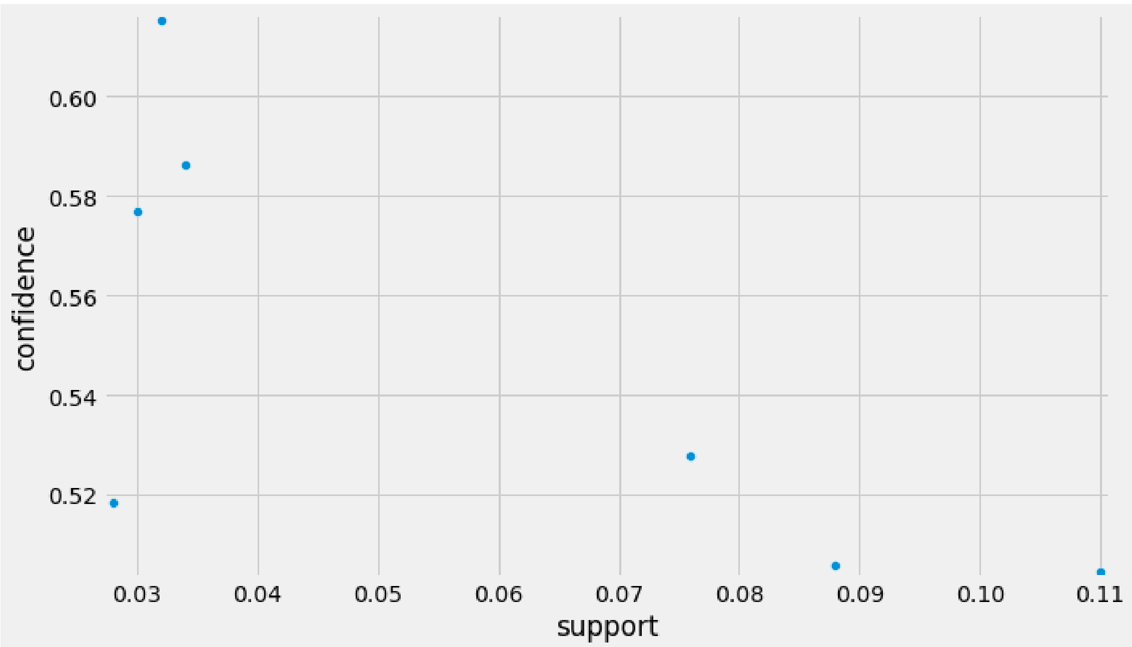

Figure 4.1: Scatterplot of Distribution of Rules for Each Category of Work

Figure 4.2: Heatmaps of Rules for Each Category of Work

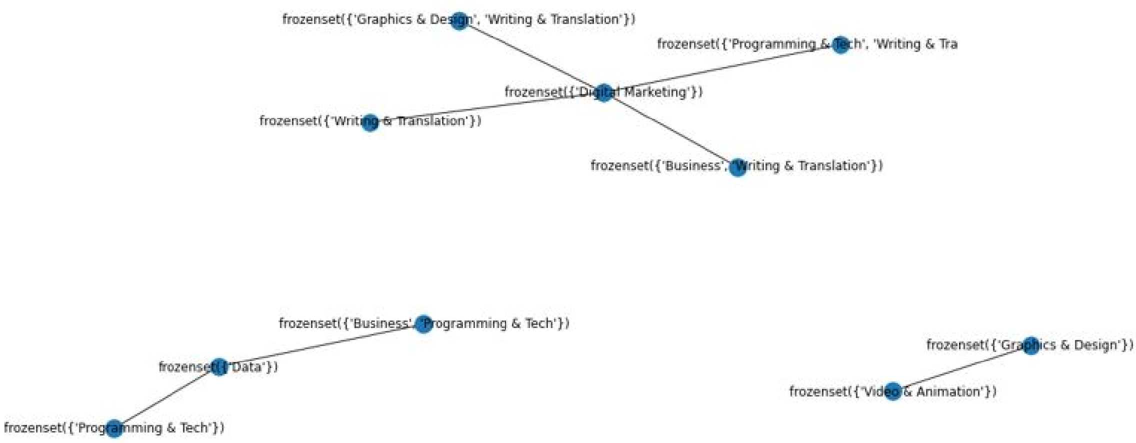

Figure 4.3: Rules Visualization for Each Category of Work with a network Tree

4.1.4 Association Rules between Each Subcategory of Work

There are 15 strong association rules generated between each subcategory of work. During the generation of these rules the support value was chosen as 0.20% for frequent itemset generation and the confidence value was chosen as 50% for strong rules generation criteria.

Out of these 15 rules Brand Style Guides ->Logo Design, E Commerce Development ->Wordpress and Flyer Design ->Logo Design are the top three rules accounting support of 4.6%, 3.4% and 3.2% respectively. On the contrary, Data Visualization ->Data Science, T-shirts & Merchandise ->Logo Design, Website Content ->Article & Blog Posts,Graphics for Streamers ->Logo Design, Marketing Strategy ->Social Media Marketing, Short Video Ads ->Whiteboard & Animated Explainers and Website Builders & CMS ->Wordpress are the ones having lowest support amongest the rules accounting 2% for the first two and 2.2% for the rest of the others respectively.

In contrast the rules Brand Style Guides ->Logo Design, Brochure Design ->Flyer Design and Sales ->Data Entry are top three rules having the confidences of 92.00%,81.25% and 76.47% respectively. We can see Logo Design a quiet of times in consequent part of these rules which indicates that it is used most frequently with the others subcategory of work.

Using the support and confidence Figure 4.4 shows the distribution of rules where 15 blue scatter points represents 15 strong association rules. Figure 4.5 shows the heatmaps of rules which help to understand a large number of rules between a small number of antecedent and consequent. And in Figure 4.3 we can see several trees which represents the rules with a network tree.

4.2 Correlation Analysis

Pearson correlation analysis was utilized to explore the relationship between pricing, sales and reviews of gigs and vice-versa. The table 4.6 represents the results of correlations we got during the research.

4.2.1 Correlation between Price and Sales

The analysis result explained that there is a small negative correlation between pricing and stars where Pearson Coefficient was r= -0.0234 overall. Which indicates, for every 1 USD increase of price, the sales may decrease by 0.025. The values of Pearson coefficient ranges between -0.090 and -0.008 which represents Digital Marketing and Business respectively. The values were always negative (except for the Lifestyle category).There is a small positive correlation between price and sales for the Lifestyle Category indicating increasing of price also increase the sales.

Table 4.5: Association Rules Between Subcategories

antecedent	consequent	antecedent support	consequent support	support	confidence	lift
Website Content	Article & Blog Posts	0.036	0.082	0.022	0.611111	7.452575
Brand Style Guides	Logo Design	0.050	0.162	0.046	0.920000	5.679012
Brochure Design	Flyer Design	0.032	0.054	0.026	0.812500	15.046296
Business Cards & Stationary	Logo Design	0.040	0.162	0.024	0.600000	3.7037041
Sales	Data Entry	0.034	0.092	0.026	0.764706	8.312020
Virtual Assistant	Data Entry	0.052	0.092	0.032	0.615385	6.688963
Data Visualization	Data Science	0.034	0.054	0.020	0.588235	10.893246
E-Commerce Development	Wordpress	0.068	0.150	0.034	0.500000	3.333333
Flyer Design	Logo Design	0.054	0.162	0.032	0.592593	3.657979
Flyer Design	Social Media Design	0.054	0.078	0.028	0.518519	6.647673
Graphics for Streamers	Logo Design	0.032	0.162	0.022	0.687500	4.243827
T-shirts & Merchandise	Logo Design	0.028	0.162	0.020	0.714286	4.409171
Marketing Strategy	Social Media Marketing	0.038	0.116	0.022	0.578947	4.990926
Short Video Ads	Whiteboard & Animated Explainers	0.036	0.060	0.022	0.611111	10.185185
Website Builders & CMS	Wordpress	0.034	0.150	0.022	0.647059	4.313725

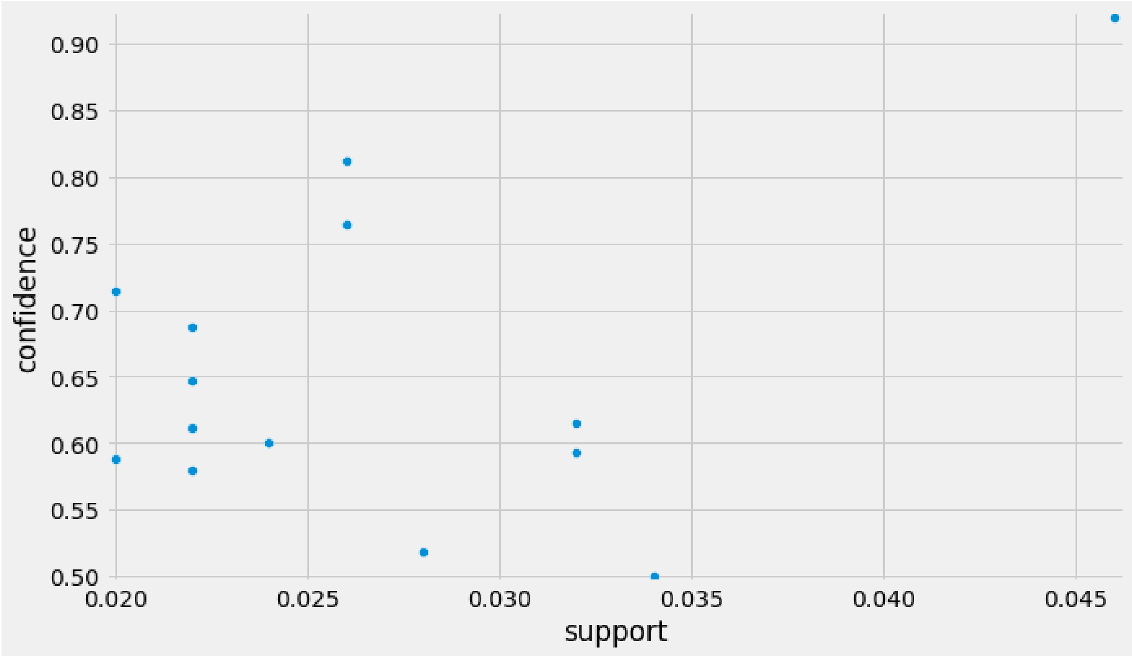

Figure 4.4: Scatterplot of Distribution of Rules for Each Subcategory of Work

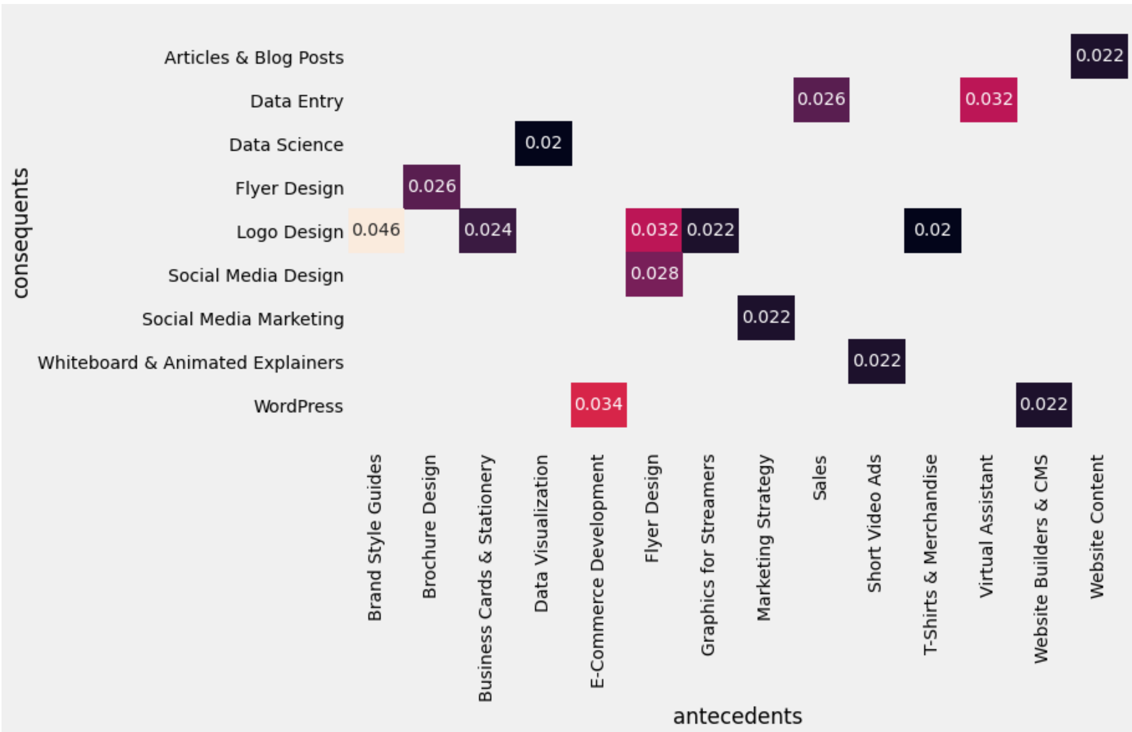

Figure 4.5: Heatmaps of Rules for Each Subcategory of Work

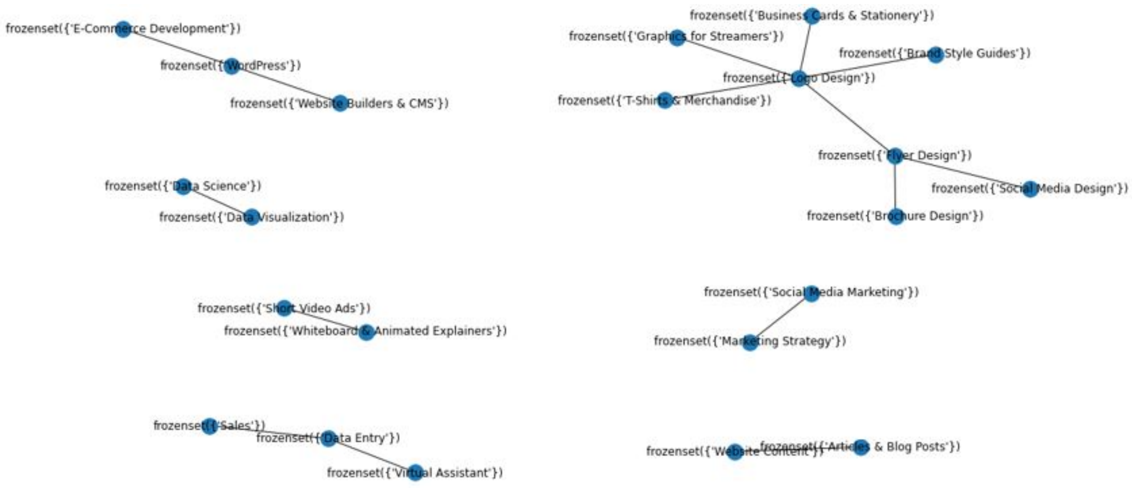

Figure 4.6: Rules Visualization for Each Subcategory of Work using Network Tree

4.2.2 Correlation between Price and Reviews/Stars

Small negative correlation between pricing and sales where Pearson Coefficient was r= -0.10 overall. Which indicates, for every 1 USD increase of price, the reviews/stars may decrease by 0.10. The values of Pearson coefficient ranges between -0.194 and -0.042 which represents Programming & Tech and Digital Marketing respectively. The values were always negative (except for the Lifestyle category). There is a small positive correlation between price and stars/reviews for the Lifestyle Category indicating increasing of price also increase the stars.

4.2.3 Correlation between Sales and Reviews/Stars

Small positive correlation between sales and stars where Pearson Coefficient was r= 0.12 overall. Which indicates, for every sale , the reviews/stars may increase by 0.12. The values of Pearson coefficient ranges between 0.148 and 0.090 which represents Programming & Tech and Lifestyle respectively.

Table 4.6: Correlation between Price, Sales and Stars/Reviews.

Overall Correlation

	Price	Sales	Stars
Price	1.0000	-0.0234	-0.101200
Sales	-0.0234	1.0000	0.122465
Stars	-0.1012	0.122465	1.000000

Programming & Tech

Price	Sales	Stars
1.0000	-0.047587	-0.19415
-0.047587	1.0000	0.148159
-0.194150	0.148159	1.000000

Lifestyle

	Price	Sales	Stars
Price	1.000000	0.073957	0.051180
Sales	0.073957	1.000000	0.090887
Stars	0.051180	0.090887	1.000000

Writing & Translation

Price	Sales	Stars
1.000000	-0.018720	-0.058871
-0.018720	1.000000	0.116646
-0.058871	0.116646	1.000000

Digital Marketing

	Price	Sales	Stars
Price	1.000000	-0.090159	-0.042002
Sales	-0.090159	1.000000	0.141554
Stars	-0.042002	0.141554	1.000000

Video & Animation

Price	Sales	Stars
1.000000	-0.041246	-0.089540
-0.041246	1.000000	0.141531
-0.089540	0.141531	1.000000

Music & Audio

	Price	Sales	Stars
Price	1.000000	-0.015457	-0.061822
Sales	-0.015457	1.000000	0.112850
Stars	-0.061822	0.112850	1.000000

Business

Price	Sales	Stars
1.000000	-0.008695	-0.117306
-0.008695	1.000000	0.111292
-0.117306	0.111292	1.000000

Conclusions

In this era of modern world, selling small services in fiverr has gain vast popularity. In this paper, the strong association rules between each category and subcategory of work listed in fiverr were minded by the Apriori algorithm. Also, correlation between pricing, sales, reviews/stars and vice-versa have been measured. The following specific conclusions were made:

- Using the Apriori algorithm, 7 strong association rules between each category and 15 rules between each subcategory of work were extracted. Of them the rules Business,Writing & Translation ->Digital Marketing and Programming & Tech,Writing & Translation ->Digital Marketing are the strongest twos among the catories of work. And Brand Style Guides ->Logo Design, E Commerce Development ->Wordpress and Flyer Design ->Logo Design are the top three rules among the subcategories. These results are expected to be useful for the freelancers already working in fiverr and to those who are willingly to join.

- Applying the correlation test we discovered small negative correlation between pricing->sales, pricing->stars and small positive correlation between sales->stars/reviews for each category of work except the Lifestyle. For the Lifestyle category we discovered small positive correlation between price->sales, price->stars and sales->stars.

Bibliography

[1] R. Agrawal, R. Srikant, *et al.*, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, Citeseer, vol. 1215, 1994, pp. 487–499.

[2] M. Houtsma and A. Swami, "Set-oriented mining for association rules in relational databases," in *Proceedings of the eleventh international conference on data engineering*, IEEE, 1995, pp. 25–33.

[3] M. J. Zaki, "Scalable algorithms for association mining," *IEEE transactions on knowledge and data engineering*, vol. 12, no. 3, pp. 372–390, 2000.

[4] D.-I. Lin and Z. M. Kedem, "Pincer-search: An efficient algorithm for discovering the maximum frequent set," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 553–566, 2002.

[5] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A maximal frequent itemset algorithm," *IEEE transactions on knowledge and data engineering*, vol. 17, no. 11, pp. 1490–1504, 2005.

[6] J. Dongre, G. L. Prajapati, and S. Tokekar, "The role of apriori algorithm for finding the association rules in data mining," in *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, IEEE, 2014, pp. 657–660.

[7] A. Zheng, Y. Hong, and P. A. Pavlou, "Matching in two-sided platforms for it services: Evidence from online labor markets," *Fox School of Business Research Paper*, no. 16-026, 2016.

[8] A. Hannák, C. Wagner, D. Garcia, A. Mislove, M. Strohmaier, and C. Wilson, "Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr," in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2017, pp. 1914–1933.

[9] M. M. J. Kabir, S. Xu, B. H. Kang, and Z. Zhao, "A new multiple seeds based genetic algorithm for discovering a set of interesting boolean association rules," *Expert Systems with Applications*, vol. 74, pp. 55–69, 2017.

[10] D. D. Green *et al.*, "Gig economy and the future of work: A fiverr. com case study," *Manag Econ Res J*, vol. 4, no. 2018, p. 6844, 2018.

[11] M. S. Hossain and M. S. Arefin, "Development of an intelligent job recommender system for freelancers using client's feedback classification and association rule mining techniques.," *J. Softw.*, vol. 14, no. 7, pp. 312–339, 2019.

[12] M. Acosta and F. Patacsil, “Analyzing the relationship between information technology jobs advertised on-line and skills requirements using association rules,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, pp. 2771–2779, Oct. 2021. DOI: 10.11591/eei.v10i5.2590.

[13] *Apriori - minimum support and minimum confidence in data mining - stack overflow*, https://stackoverflow.com/questions/25322204/minimum-support-and-minimum-confidence-in-data-mining, (Accessed on 02/17/2022).

[14] *Fiverr outsources your small jobs for $5*, https://lifehacker.com/fiverr-outsources-your-small-jobs-for-5-5488130, (Accessed on 02/12/2022).

[15] *In association rules, how do you set the minimum support and minimum confidence? - quora*, https://www.quora.com/In-association-rules-how-do-you-set-the-minimum-support-and-minimum-confidence, (Accessed on 02/17/2022).

[16] *Pearson's correlation coefficient - statistics solutions*, https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/, (Accessed on 02/13/2022).

[17] *What are association rules in data mining (association rule mining)?* https://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining, (Accessed on 02/12/2022).