

**UNIVERSITY OF ECONOMICS AND LAW**

**VNUHCM**

**MID-TERM TEST:**

**PROGRAM PACKAGE IN FINANCE 2**

**YEAR 2023**

**<TRADE CREDIT (RECEIVABLE) OF FIRMS LISTED ON HNX>**

**Syllabus: K20414C\_ Fintech**

**Course: 222CN0901**

**Student: Lương Thị Mỹ Tâm**

**MSSV: K204141929**

**UNIVERSITY OF ECONOMICS AND LAW**

**VNUHCM**

**MID-TERM TEST:**

**PROGRAM PACKAGE IN FINANCE 2**

**YEAR 2023**

**<TRADE CREDIT (RECEIVABLE) OF FIRMS LISTED ON HNX>**

**DECLARATION**

I hereby declare that the mid-term test "Trade credit (Receivable) of firms listed on HNX" is the result of my work under the guidance of Dr. Nguyen Thanh Liem, within the framework of the subject 'Program package in finance 2'.

**TABLE OF CONTENTS**

**[DECLARATION 1](#_Toc26220)**

**[TABLE OF CONTENTS 2](#_Toc681)**

**[LIST OF TABLES/FIGURES 3](#_Toc8818)**

**[TASK 1: Perform literature review 5](#_Toc23307)**

**[TASK 2: Create Dataset 7](#_Toc18709)**

**[TASK 3: Report 7](#_Toc22636)**

[A) 5 firms with highest trade credit 8](#_Toc31540)

[B)5 firms with lowest trade credit 8](#_Toc18655)

[C) Descriptive statistics 8](#_Toc1162)

**[TASK 4: Data Visualization 10](#_Toc20345)**

[1. The histogram of trade credit 10](#_Toc28935)

[2. The scatter plot of trade credit with the continuous variable 11](#_Toc20665)

[3. The boxplot of trade credit with the discrete variable 12](#_Toc16962)

[4. The plot that allow the combination of continuous, discrete variables and trade credit 12](#_Toc25279)

**[TASK 5: Regression 13](#_Toc27950)**

[A) Regression Analysis 13](#_Toc26786)

[B) Test of multicollinearity 14](#_Toc19121)

[C)Test of heteroskedasticity 15](#_Toc31025)

[D)Improving model 16](#_Toc14810)

**[TASK 6: USING LOOP 18](#_Toc29907)**

**[REFERENCE 19](#_Toc26812)**

**[APPENDIX 20](#_Toc2927)**

# **LIST OF TABLES/FIGURES**

|  |  |  |
| --- | --- | --- |
|  | **Name of tables/figures** | **Page** |
| 1 | Table 1: The Data set | 7 |
| 2 | Table 2: Top 5 highest trade credit | 8 |
| 3 | Table 3: Top 5 lowest trade credit | 8 |
| 4 | Table 4: Descriptive Statistics of discrete variable | 9 |
| 5 | Table 5: Descriptive Statistics of continuous variable | 10 |
| 6 | Figure 6: The histogram chart | 11 |
| 7 | Figure 7: The scatter plot | 11 |
| 8 | Figure 8: The box plot | 12 |
| 9 | Figure 9: The 3-variable combination plot | 13 |
| 10 | Figure 10: The result of regression model | 14 |
| 11 | Figure 11: Checking multicollinearity | 14 |
| 12 | Figure 13: The plots | 16 |
| 13 | Figure 14: The result of regression model | 17 |
| 14 | Figure 15: The result of checking model | 18 |
| 15 | Figure 15: Predicting new customer | 18 |
| 16 | Figure 16: The plots | 18 |
| 17 | Figure 17: The number of firms each industry | 19 |
| 18 | Figure 18: The number of firms | 19 |

**TASK 1: Perform literature review**

Nadiri (1969) studied the trade credit of manufacturing firms in the United States through quarterly data analysis, from the first quarter of 1949 to the fourth quarter of 1964. Regression analysis results of log-linear model have identified factors that have a statistically significant influence on trade credit. Size (revenue) is a factor that has a very positive impact on A/R. Furthermore, liquidity (cash and negotiable securities/short-term debt) has a negative impact on A/R.

Petersen & Rajan (1997) studied the trade credit of 3,404 small businesses, from the National Survey of Small Business Finances (NSSBF) data in the United States from 1988 to 1989. The pooled OLS analysis shows that many factors have a significant influence on trade credit. Factors affecting A/R: Size (assets), bank loans and revenue growth (investment opportunities) have a positive impact, while fertility Net profit (profit after tax/sales) has a negative impact, besides, the number of years of operation and profit margin (gross profit/sales) have a nonlinear effect.

García-Teruel & Martínez-Solano (2010) analyzed the determinants of trade credit granted and received of 47,197 small and medium enterprises in Europe between 1996 to 2002. The results show that there is a strong homogeneity in the determinants of trade credit among European countries. For trade credit (receivable), company size is considered a determinant factor of receivable. The results show that there is a positive relationship between size and granted trade credit. As a result, large companies finance their clients more than small ones.

Vaidya (2011) conducted a study to analyze data for the period 1993 - 2006 of 1,522 companies in India to determine the factors affecting trade credit. Through the Generalized Method of Moments (GMM), the study has demonstrated that many factors have significant (statistical) influence. The factors that positively affect A/R are fixed assets (fixed assets/total assets), liquidity (cash and negotiable securities/revenue) and size (assets). Negative factors are Inventory (inventories/sales), profitability (profit before depreciation and taxes/sales) and short-term bank loans (bank loans/sales) .

Giannetti et al. (2011) studied the data of 3,489 small businesses, from the National Survey of Small Business Finances (NSSBF) data in the United States in 1998, transportation belongs to the service industry group. The results of the pooled OLS analysis show that many factors have a significant influence on trade credit. Factors affecting A/R including size (assets) and bank loans have a positive impact; fixed assets and distance to the bank have a negative effect. Furthermore, the Service sector (including transportation) grants more trade credits than the goods manufacturing industry.

Akinlo (2012) analyzed panel data of 66 non-financial listed companies on the Nigerian stock market for the period 1999 - 2007, to study the factors affecting trade credit. The results of the pooled OLS analysis show that many factors have a significant influence on A/R. In which, bank loans (short-term loans/assets) and liquidity (money and negotiable securities/assets) have a positive impact. Return on assets (ROA), size (sales) and inventory (value of inventory/assets) have a negative effect. On the other hand, through the Hausman test, the study shows that the FEM method gives better results than the REM method. The results of the FEM analysis only confirm two factors that determine A/R are Size has a negative impact and liquidity has a positive impact.

Santos & Silva (2014) conducted a study with 11,040 industrial enterprises in Portugal between 2003 and 2009 using the FEM method to determine the factors affecting trade credit. Regression analysis results confirm many factors affecting A/R. Number of years of operation, profit margin (gross profit/revenue), equity, revenue growth have a positive impact. Besides, accounts payable and official credit shortfall are factors that have a negative impact on A/R of surveyed companies.

Chou et al. (2015) analyzed data from the annual balance sheets of 300 Vietnamese non-financial listed companies in both the Hanoi Stock Exchange (HNX) and the Ho Chi Minh City Stock Exchange (HOSE) for the period 2005 to 2012 to specify how the determinants affect trade credit. The article used regression analysis and the results showed that factors including cash, equity, net sales, short-term debt, and size positively affect accounts receivable. Meanwhile, EBIT and inventory have a negative link with accounts receivable.

Based on literature review above, I decided to choose two independent variables are expected to positively affect Trade credit (Receivable) for the model as follows:

+ 1 Categorical variable: Size. This is the variable that is transformed from the continuous variable totalasset based on the quartile of the totalasset variable. Therefore, the size variable will include 4 values including if the companies have total assets below the value Q1 they will be 1, if the total assets are less than mean they will be 2, if the total assets are less than the value of Q3 they will be 3 and the remaining companies will be 4.

+ 1 Continuous variable: totalequity. This is the variable that represents the total equity of the company.

**TASK 2: Create Dataset**

|  |
| --- |
| df <- data.frame(select(X040522\_Data\_Mid\_term\_test\_Final, c(2,3,4,5,9,23,26,8))  %>% filter(exchangename == 'HANOI STOCK EXCHANGE'))  set.seed(929)  df <- df[sample(1:nrow(df),100), ]  df <- rename(df,'Trade\_credit'= receivable)  df$Trade\_credit[is.na(df$Trade\_credit)]=median(df$Trade\_credit,na.rm=T)  df$Trade\_credit[df$Trade\_credit == 0] <- mean(df$Trade\_credit[df$Trade\_credit != 0])  df$totalasset[is.na(df$totalasset)]=median(df$totalasset,na.rm=T)  df$totalequity[is.na(df$totalequity)]=median(df$totalequity,na.rm=T)  df$totaldebt[is.na(df$totaldebt)]=median(df$totaldebt,na.rm=T)  df$totaldebt[df$totaldebt == 0] <- mean(df$totaldebt[df$totaldebt != 0])  df$size <- ifelse(df$totalasset < quantile(df$totalasset, probs = 0.25),'1' ,  ifelse(df$totalasset < median(df$totalasset),'2', ifelse(df$totalasset < quantile(df$totalasset, probs = 0.75) ,'3' , '4')))  options(scipen=999)  View(df) |

|  |
| --- |
|  |
| Table 1: The Data set |

**TASK 3: Report**

1. **5 firms with highest trade credit**

|  |
| --- |
| df\_max<-df %>% slice\_max(Trade\_credit, n = 5)  options(tibble.print\_max = Inf)  View(df\_max) |

|  |
| --- |
|  |

Table 2: Top 5 highest trade credit

**B)5 firms with lowest trade credit**

|  |
| --- |
| df\_max<-df %>% slice\_min(Trade\_credit, n = 5)  options(tibble.print\_max = Inf)  View(df\_min) |

|  |
| --- |
|  |

Table 3: Top 5 lowest trade credit

**Comment:**

From the two tables above, we can see that the top 5 companies with the highest trade credit belong to the industries of Industrials, Energy, Basic Materials, and Consumer Cyclicals, in which Vegetexco Port JSC of the Industrials industry has the highest trade credit. The top 5 companies with the lowest trade credits belong to the Consumer Non-Cyclicals, Real Estate, Consumer Cyclicals, and Financials industries, in which OHC Hospitality and Service JSC belongs to the Consumer Non-Cylicals industry which has the lowest trade credit. This result is different from in Europe, the businesses with the highest trade credit will be in the Constructions, Whole sales, and the lowest trade credit will be in the Retail Trade industry. However, only at the scope of small and medium enterprises in Europe according to research by García-Teruel & Martínez-Solano (2010).

1. **Descriptive statistics**

**-The discrete variable**

|  |
| --- |
| sta<-df %>%  group\_by(size) %>%  summarize(mean\_receivable = mean(Trade\_credit),  median\_receivable = median(Trade\_credit),  min\_receivable = min(Trade\_credit),  max\_receivable = max(Trade\_credit),  std\_receivable=sd(Trade\_credit),  count = n())  View(sta) |

|  |
| --- |
|  |

Table 4: Descriptive Statistics of discrete variable

**Comment:**

Based on the above descriptive statistics, we can see that the number of companies in each class is the same without much difference. Size (total assets) is 2 has the highest min\_receivable value and max\_receivable value is the lowest so its trade credit standard deviation is smallest, while Size (total assets) is 4 is the opposite of size (total asset) is 2 so there is the biggest standard deviation of the 4 classes. Looking at the mean and median values of the trade credit of the 4 categories, it is easy to see that the larger the size (total assets), the larger the trade credit. This relationship supports the results of Vaidya (2011) and many other research papers reviewed above.

**- The continuous variable**

|  |
| --- |
| df$equity\_class <-ifelse(df$totalequity < median(df$totalequity),'below median' ,'above median')  sta1<-df %>%  group\_by(equity\_class) %>%  summarize(mean\_receivable = mean(Trade\_credit),  median\_receivable = median(Trade\_credit),  min\_receivable = min(Trade\_credit),  max\_receivable = max(Trade\_credit),  std\_receivable=sd(Trade\_credit),  count = n())  View(sta1) |

|  |
| --- |
|  |

Table 5: Descriptive Statistics of continuous variable

**Comment:**

* From the statistics described above, we can see that the number of companies in the two groups is the same, there is not much difference. The above-median group has a lower min\_receivable value and a higher max\_receivable value when compared with the below-median group, so its trade credit standard deviation is higher than that of the below-median group. Looking at the mean and median values of the trade credit of the two groups, we see that the larger the total equity, the larger the trade credit. This relationship is in agreement with the results of Santos & Silva (2014) and Chou et al. (2015).

=> From here, we see that there can be a possible link between the discrete and continuous variables with trade credit that the larger the size (total assets ) and the total equity, the larger the trade credit, and it is in line with our literature review.

**TASK 4: Data Visualization**

1. **The histogram of trade credit**

|  |
| --- |
| ggplot(df, aes(x = Trade\_credit, fill='Trade\_credit')) + geom\_histogram()+ scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |
| Figure 6: The histogram chart |

**Comment:**

Looking at the histogram, we see that the histogram has a right skewed distribution, it shows that the mean is greater than the median and there are outliers. And the value of trade credit is concentrated in the range of 0 to 500 billion VND.

1. **The scatter plot of trade credit with the continuous variable**

|  |
| --- |
| ggplot(df, aes(x = totalequity , y = Trade\_credit)) +  geom\_point()+  scale\_y\_continuous(labels = scales::comma) +  geom\_smooth(method = "lm")+  scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 7: The scatter plot

**Comment:**

* Looking at the scatter chart, we see an upward trend line showing a positive correlation between the two variables total equity and trade credit. It is similar to the comment section in task 2 which suggests that there can be a positive relationship between these two variables. However, the wide dispersion points show that this relationship is weak, so this trade credit variable depends on other factors as well.

1. **The boxplot of trade credit with the discrete variable**

|  |
| --- |
| df %>%  filter(!is.na(Trade\_credit), !is.na(size)) %>%  ggplot(aes(x = size, y = Trade\_credit, fill=size)) +  geom\_boxplot() +  coord\_flip()+  scale\_y\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 8: The box plot

**Comment:**

Looking at the boxplot above, we can see that there are 8 outliers in which Class 4 has the most outliers and Class 2 does not contain outliers. Class 4 is the class with the largest median value and Class 1 is the class with the smallest median value of the 4 classes. When looking at the length of the box body, we see that Class 4 has the longest body length showing the greatest volatility, and Class 2 has the least. Besides, looking at the extreme value at the end of two whiskers of Class 4, we see that the range of scores is highest compared to the rest of the classes which shows that Class 4 has the widest distribution and Class 1 has the narrowest distribution. Regarding skewness, most of the classes have right-skewed, and only class 3 is symmetric.

1. **The plot that allow the combination of continuous, discrete variables and trade credit**

|  |
| --- |
| ggplot(df, aes(x = Trade\_credit, y = totalequity, color = as.factor(size))) +  geom\_point() +  scale\_color\_manual(values = c("#270181", "yellow","green","red"))+  scale\_x\_continuous(labels = scales::comma) |

|  |
| --- |
|  |

Figure 9: The 3-variable combination plot

**Comment:**

* Looking at the combined chart of the 3 variables, we see that the relationship between the 3 variables of size (total assets), total equity, and trade credit is positively correlated. Class 4 is clearly still the Class with the strongest dispersion.

=> From the analysis of the 4 plots in the visualization part, we summarize that there is a positive correlation relationship between the 3 variables, which is the larger the size of the business (total assets) and the larger the total equity, the larger the trade credit. However, the correlation between these 3 variables is not strong. And with large-scale companies, there will be very strong fluctuations in trade credit. This can be explained that Larger firms finance their customers more than small ones, since large firms can obtain finance more easily and hence can act as financial intermediaries according to research by García-Teruel, P. J., & Martinez-Solano, P. (2010). This result supports and clarifies the comment in task 2 about the relationship between the 3 variables and this result is in agreement with the results of the above review articles, but disagree with the study by Akinlo (2012) conducted with companies on the Nigerian stock market.

**TASK 5: Regression**

1. **Regression Analysis**

|  |
| --- |
| summary(Trade\_credit.lm<-lm(Trade\_credit ~ size +totalequity +totaldebt , data = df)) |

|  |
| --- |
|  |

Figure 10: The result of regression model

**Comment:**

In running the regression model, we will add the total debt variable. Regression results show that the adjusted R-squared coefficient is 0.4315 which means 43.15% variation in trade credit explained by the independent variables. In task 1, we see that there are many papers that give results that the size (total asset) will have an effect on Trade credit, but the regression results show that the categorical variable of size is not statistically significant (p-value > 0.05). The two variables total equity and total debt are statistically significant (p-value < 0.05) and positively correlated with trade credit. The results of these two variables support the results of many studies reviewed in task 1, but disagree with the research results of Vaidya (2011) on the totaldebt variable. In terms of coefficient significance, we can explain that when the total equity of the company increases by 1 VND, the trade credit of the company will increase by 0.22117 VND, other things being equal (Similar to total debt). However, the model has a very large residual standard error.

1. **Test of multicollinearity**

|  |
| --- |
| car::vif(Trade\_credit.lm) |

|  |
| --- |
|  |

Figure 11: Checking multicollinearity

**Comment:**

* The VIF coefficients of all three variables are less than 2, so the model does not have multicollinearity.

**C)Test of heteroskedasticity**

|  |
| --- |
| bptest(Trade\_credit.lm)  shapiro.test(resid(Trade\_credit.lm)) |

|  |
| --- |
|  |

Firgure 12: Checking heteroskedasticity

**Comment:**

* The p-value of Breusch-Pagan test is greater than 0.05 so the model is not heteroskedasticity. Besides, the p-value of Shapiro-wilk normality test is less than 0.05 so we conclude that the data set is not normally distributed.

**D)Test**

|  |
| --- |
| par(mfrow=c(2,2))  plot(Trade\_credit.lm) |

|  |
| --- |
|  |

Figure 13: The plots

**Comment:**

The first graph (Residuals vs Fitted) shows that the assumption of linearity of the data is slightly violated. And the assumption of the mean of residuals is also violated because the red line is far from the horizontal line (corresponding to residual =0). The Normal Q-Q plot shows that the assumption of normally distributed residuals is satisfied. The (Scale - Location) plot shows that the assumption of uniformity of variance is violated because the red line curves and the residuals are unevenly dispersed around this line. The fourth graph shows that the 198th, 199th, and 223rd observations may be highly influential points in the dataset.

=> From here, we will improve the model by using a log-log regression model to minimize the effect of outliers and remove the variable size from the model since it is not statistically significant.

**E)Improving model**

|  |
| --- |
| summary(Trade\_credit.lm1<-lm(log(abs(Trade\_credit)) ~ log(abs(totalequity))+ log(abs(totaldebt)), data = df)) |

|  |
| --- |
|  |

Figure 14: The result of regression model

**Comment:**

The regression results show that the adjusted R-squared coefficient of 0.1664 is lower than that of the previous model. The two variables total equity and total debt are statistically significant (p < 0.05) and positively correlated with trade credit. In terms of coefficient significance, it can be explained that when a company's total equity increases by 1%, the company's trade credit will increase by 0.4306%, other things being equal. (Similar to total debt). And the model has improved the problem of large residual standard error.

**Checking model**

|  |
| --- |
| car::vif(Trade\_credit.lm1)  bptest(Trade\_credit.lm1)  shapiro.test(resid(Trade\_credit.lm1)) |

|  |
| --- |
|  |

Figure 15: The result of checking model

|  |
| --- |
| par(mfrow=c(2,2))  plot(Trade\_credit.lm1) |

|  |
| --- |
|  |

Firgure 16: The plots

**Comment:**

* Based on the above firgures, we see that the model has no multicollinearity and heteroskedasticity. The first graph (Residuals vs Fitted) shows that the assumption of the linearity of the data and the assumption of the mean of the residuals have both been improved. The normal Q-Q plot shows that the assumption of normally distributed residuals is satisfied. The graph (Scale - Location) shows that the assumption of uniformity of variance has also been improved to no longer be violated.

**TASK 6: USING LOOP**

|  |
| --- |
| for(i in df$industry)  {  print(count(df, industry))  break  }  #Visualize  ggplot(df, aes(x = industry, fill=as.factor(industry))) +  geom\_bar() |

|  |
| --- |
|  |

Figure 17: The number of firms each industry

|  |
| --- |
| for(i in df$industry)  { print(df %>%  filter(Trade\_credit > 6.074e+10) %>%  group\_by(industry) %>%  count(industry))  break  } |

|  |
| --- |
|  |

Firgure 18: The number of firms

**REFERENCE**

Akinlo, A. (2012). The Determinants of Trade Credit: Evidence from Nigerian Manufacturing Firms. Economic Journal of A 2 Z, 1(1), 31-42.

Chou, J. H., Horng, M. S., Wu, W. M., & Phoung, N. T. (2015). The Determinants of trade credit: Vietnam experience. In Proceedings of the Second Asia-Pacific Conference on Global Business, Economics, Finance and Social Sciences (AP15Vietnam Conference).

García-Teruel, P. J., & Martinez-Solano, P. (2010). Determinants of trade credit: A comparative study of European SMEs. International Small Business Journal, 28(3), 215-233.

Giannetti, M., Burkart, M., & Ellingsen, T. (2011). What you sell is what you lend? Explaining trade credit contracts. The Review of Financial Studies, 24(4), 1261-1298.

Nadiri, M. I. (1969). The determinants of trade credit in the US total manufacturing sector. Econometrica: Journal of the Econometric Society, 408-423.

Petersen, M. A., & Rajan, R. G. (1997). Trade credit: theories and evidence. The review of financial studies, 10(3), 661-691.

Santos, J., & Silva, A. (2014). The determinants of trade credit: A study of Portuguese industrial companies. International Journal of Financial Research, 5, 128-138.

Vaidya, R. R. (2011). The determinants of trade credit: Evidence from Indian manufacturing firms. Modern Economy, 2(05), 707.

**APPENDIX**

#task2

library(dplyr)

df <- data.frame(select(X040522\_Data\_Mid\_term\_test\_Final, c(2,3,4,5,9,23,26,8))

%>% filter(exchangename == 'HANOI STOCK EXCHANGE'))

set.seed(929)

df <- df[sample(1:nrow(df),100), ]

as\_tibble(df)

df <- rename(df,'Trade\_credit'= receivable)

View(df)

#Check and replace NA

summary(df)

sum(is.na(df))

which(is.na(df))

df$Trade\_credit[is.na(df$Trade\_credit)]=median(df$Trade\_credit,na.rm=T)

df$Trade\_credit[df$Trade\_credit == 0] <- mean(df$Trade\_credit[df$Trade\_credit != 0])

df$totalasset[is.na(df$totalasset)]=median(df$totalasset,na.rm=T)

df$totalequity[is.na(df$totalequity)]=median(df$totalequity,na.rm=T)

df$totaldebt[is.na(df$totaldebt)]=median(df$totaldebt,na.rm=T)

df$totaldebt[df$totaldebt == 0] <- mean(df$totaldebt[df$totaldebt != 0])

sum(is.na(df))

#create size variance

unique(df$industry)

df$size <- ifelse(df$totalasset < quantile(df$totalasset, probs = 0.25),'1' ,

ifelse(df$totalasset < median(df$totalasset),'2',

ifelse(df$totalasset < quantile(df$totalasset,

probs = 0.75) ,'3' , '4')))

options(scipen=999)

View(df)

#task3

#top 5 highest and lowest trade credit

df\_max<-df %>% slice\_max(Trade\_credit, n = 5)

df\_min<- df %>% slice\_min(Trade\_credit, n = 5)

options(scipen=999)

View(df\_max)

View(df\_min)

#descriptive statistics

sta<-df %>%

group\_by(size) %>%

summarize(mean\_receivable = mean(Trade\_credit),

median\_receivable = median(Trade\_credit),

min\_receivable = min(Trade\_credit),

max\_receivable = max(Trade\_credit),

std\_receivable=sd(Trade\_credit),

count = n())

View(sta)

df$equity\_class <-ifelse(df$totalequity < median(df$totalequity),'below median'

,'above median')

sta1<-df %>%

group\_by(equity\_class) %>%

summarize(mean\_receivable = mean(Trade\_credit),

median\_receivable = median(Trade\_credit),

min\_receivable = min(Trade\_credit),

max\_receivable = max(Trade\_credit),

std\_receivable=sd(Trade\_credit),

count = n())

View(sta1)

#task4

library(ggplot2)

#histogram of trade credit

ggplot(df, aes(x = Trade\_credit, fill='Trade\_credit')) +

geom\_histogram()+

scale\_x\_continuous(labels = scales::comma)

#scatter plot of trade credit with the continuous variable

ggplot(df, aes(x = totalequity , y = Trade\_credit)) +

geom\_point()+

scale\_y\_continuous(labels = scales::comma) +

geom\_smooth(method = "lm")+

scale\_x\_continuous(labels = scales::comma)

#boxplot of trade credit with the discrete variable

df %>%

filter(!is.na(Trade\_credit), !is.na(size)) %>%

ggplot(aes(x = size, y = Trade\_credit, fill=size)) +

geom\_boxplot() +

coord\_flip()+

scale\_y\_continuous(labels = scales::comma)

#plot for combination of continuous, discrete variables and trade credit

ggplot(df, aes(x = Trade\_credit, y = totalequity, color = as.factor(size))) +

geom\_point() +

scale\_color\_manual(values = c("#270181", "yellow","green","red"))+

scale\_x\_continuous(labels = scales::comma)

#task5

library(tidyverse)

#check linearity

plot(Trade\_credit ~ totalequity, data=df)

plot(Trade\_credit ~ size, data=df)

#check correlation

df$size <- as.numeric(df$size)

cor.test(df$size, df$totalequity)

install.packages('car')

library(car)

install.packages("stargazer")

library(stargazer)

#regression model

summary(Trade\_credit.lm<-lm(Trade\_credit ~ size +totalequity +totaldebt , data = df))

stargazer(Trade\_credit.lm, type ="text")

#check multicollinearity

car::vif(Trade\_credit.lm)

par(mfrow=c(2,2))

plot(Trade\_credit.lm)

#check heteroskedasticity

install.packages('lmtest')

library(lmtest)

bptest(Trade\_credit.lm)

shapiro.test(resid(Trade\_credit.lm))

#improving model

summary(Trade\_credit.lm1<-lm(log(abs(Trade\_credit)) ~ log(abs(totalequity))+ log(abs(totaldebt)), data = df))

car::vif(Trade\_credit.lm1)

bptest(Trade\_credit.lm1)

shapiro.test(resid(Trade\_credit.lm1))

par(mfrow=c(2,2))

plot(Trade\_credit.lm1)

#task6

#Count the number of firms in an industry

for(i in df$industry)

{

print(count(df, industry))

break

}

#Visualize

ggplot(df, aes(x = industry, fill=as.factor(industry))) +

geom\_bar()

#Count the number of firms each industry and with trade credit above a certain value

for(i in df$industry)

{ print(df %>%

filter(Trade\_credit > 6.074e+10) %>%

group\_by(industry) %>%

count(industry))

break

}