1. **This question requires you to do some coding and show results. You can cut and paste your code and the results into the exam using a snipping tool. Review the dplyr exercise (module: R Data Management; Exercise: Dplyr #4) in which a function is used to calculate the min and max values of ni for each firm along with the associated year. Using this as foundation, create and implement a function (using the same dataset) to output the following values for each firm:**
    a. **Min value of net margin that is positive (i.e., exclude negative values)**
    b. **Year in which the above value is achieved**
    c. **Max value of net margin that is positive (i.e., exclude negative values)**
    d. **Year in which the above value is achieved**

```r
library('dplyr')

df <- read.csv("/Users/dana/Downloads/Compustat data small 2000-17.csv", na.strings = '')
df <- df[! (df$tic == "BAC"),] # taking professor's data cleaning from class code

get_nm <- function(df) {
  df <- df %>% mutate(nm = ni / revt)  %>% filter(nm >= 0)
  YrMinNM <- df$fyear[which.min(df$nm)]
  YrMaxNM <- df$fyear[which.max(df$nm)]
  data.frame('Year_Min_NM' = YrMinNM, 'Min_NM' = min(df$nm),
             'Year_Max_NM' = YrMaxNM, 'Max_NM' = max(df$nm))
}
df %>% select(tic, fyear, ni, revt) %>% group_by(tic) %>% do(get_nm(.))

```

14:1 (Top Level)

Console | Terminal | Jobs

~/

```
# A tibble: 11 x 5
# Groups:   tic [11]
   tic   Year_Min_NM  Min_NM Year_Max_NM Max_NM
   <chr>       <int>   <dbl>       <int>  <dbl>
 1 AAPL         2003 0.0111         2012 0.267
 2 AMAT         2012 0.0125         2000 0.216
 3 BA           2002 0.00910        2007 0.0614
 4 F            2003 0.00301        2011 0.148
 5 HON          2004 0.0500         2015 0.124
 6 IBM          2002 0.0441         2013 0.165
 7 MSFT         2015 0.130          2000 0.410
 8 PAYX         2003 0.234          2000 0.293
 9 PG           2001 0.0745         2009 0.170
10 STZ          2009 0.0295         2013 0.399
11 XOM          2016 0.0397         2006 0.118
> 
```

1. **This question requires you to do some coding and show results. You can cut and paste your code and results into the exam using a snipping tool. Find the value (or price) of an American-type call option using the following details. The stock's most recent price is 67.25. The stock's return has a standard deviation of 40%. The stock has a dividend yield of 3%. The option expires in 6 months and has a strike price of 60. The risk-free rate of interest is 1.25%.**

```
1  library('fOptions')
2
3  TypeFlag<-"c"
4  X <- 60
5  t <- .5
6  S <- 67.25
7  R <- .0125
8  B <- -0.0175
9  sg <- .4
10
11 fOpBAW<-BAWAmericanApproxOption(TypeFlag, S, X, t,
12                                 R, B, sg,title = NULL,description = NULL)
13 fOpBAW # option price = $10.9545
14
```

14:1 (Top Level)

Console | Terminal | Jobs

~/

```
Title:
 BAW American Approximated Option

Call:
 BAWAmericanApproxOption(TypeFlag = TypeFlag, S = S, X = X, Time = t,
     r = R, b = B, sigma = sg, title = NULL, description = NULL)

Parameters:
          Value:
 TypeFlag c
 S        67.25
 X        60
 Time     0.5
 r        0.0125
 b        -0.0175
 sigma    0.4

Option Price:
 10.9545

Description:
 Wed Jun 30 10:41:03 2021
```

Dana DiVincenzo
6/30/2021

1. **What is the Fama-French 3-factor model? How do people use this model? Discuss one of the three factors in detail making references to real firms. (You can refer to the class material on Microsoft as guidance but use other firms in your example)**

The Fama-French 3-Factor Model is a performance evaluator for firms, named after finance professors Eugune Fama and Dan French. The model is based on the idea that stock prices and returns are sensitive to a certain set of factors. In this model, the 3 factors used are:

- The Market
- SMB - This factor is about the size of the firm. Small firms and large firms were found to respond differently to markets
- HML- This factor is about growth firms vs value firms. Low Book-to-Market (ie growth) firms seemed to perform differently compared to high book - to-market (ie value) firms

Let's look at SMB in the context of real companies. SMB, or small minus big, measures the extra return investors tend to see from investing in stocks of companies with smaller market capitalization, rather than companies with larger market capitalization. Some examples of large-cap firms include Apple and Amazon, huge powerhouses in the financial market. A positive SMB value tells us that small cap stocks performed better than large cap stocks during that month or chosen time period. The logic of the FF 3 factor model tells us that an investor would likely see extra returns from investing in a less expensive or slowing growing stock, rather than a huge company such as Apple or Amazon.

You can use this model to look at and explain company performance in the past X years. To do this, you can conduct a regression analysis, where the dependent variable is the excess return of a stock, and the independent variables are the 3 factors listed above. Once the regression is complete, you can look at the summary to find out information on the company's performance. Within the summary, one way you can measure alpha, which refers to performance, is by looking at the intercept. A positive intercept represents an over performance in the past X years, rather than an underperformance. We can also look at the coefficients column to see which of the dependent variables were significant in the regression. The R^2 value in the regression tells us how much variation in the dependent variable is explained by the independent variables.

You can also use this model to look at the performance of a company currently, relative to its past performance. To do this, we first calculate a benchmark, which is found by multiplying the current factor values times that variable's past coefficient calculated within the regression, and then taking the sum of all values. Next, you can find alpha by taking the difference between exRET and the benchmark. Again, a positive alpha value here represents a current over performance relative to the past X years, and a negative value would indicate an underperformance.

Although the FF 3 Factor model is important for investment managers, it is also important for corporate managers, because it allows you to understand the ways in which outsiders are analyzing the performance of your firm.

1. **What is tf-idf? Discuss an application using posted class material. Next, calculate tf-idf using the following information. A corpus has 10 documents. In total, these documents have 85,000 words. In one of these documents, the word "bankruptcy" occurs 12 times. The total word count in this particular document is 4,800. This word appears in 7 of 10 documents in the corpus.**

TF-IDF stands for Term Frequency, Inverse Document Frequency. This method of text mining uses word frequency in a document, relative to work frequency within other documents in the corpus, to understand the significance of the word. A word will be considered more important or relevant to a specific document if it appears more frequently within that document, but less frequently within other documents in the corpus. For example, in class we performed some text mining on Jane Austin books. The collection of books was the corpus and each individual book is a document within that corpus. We found that certain words, like the word "the" appeared very often in each document. However, that word appeared in every document, so the IDF = 0, making TF-IDF = 0. Therefore, the word 'the' is not particularly important/relevant within any of the documents. TF-IDF can be used on a wide variety of corpuses, not just books. For example, you could download news articles about a specific topic, such as COVID19, from a website such as Google News or CNN. By using TF-IDF, you can find the most relevant keywords, and see which topics are being discussed in the news related to COVID19.

Mathematically, TF-IDF is calculated by finding TF (the number of times a word appears in a document divided by the total number of words in that document) and multiplying it by IDF (the natural log of the number of documents in the corpus divided by the number of documents containing your chosen word).

Calculation of TF-IDF in question:

TF = number of occurrences / number of word = 12/4800 = .0025

IDF = ln(number of documents / number of docs containing chosen word) = ln(10/7) = 0.35667494393

TF*IDF = .0025*0.35667494393 = **0.00089168735**