MGT 6203 Group Project Proposal

Group 14

Please edit the following template to record your responses and provide details on your project plan.

TEAM INFORMATION (1 point)

Team #: 14

Team Members:

1. Aaditya Vinnakota, Aaditya316

   My name is Aaditya Vinnakota. I am currently working as a Program Manager leading a data team in one of the Insurance companies in California. I have around 14 years of experience working on various databases, ETL and Reporting tools. I was born and brought up in India, I have a Bachelor's degree in Computer Science Engineering. I am taking the Micromasters program to enhance my skills and also to prepare for the OMSA program.

2. Bao Vo, baovo022293
   My name is Bao Vo. I grew up in Vietnam and immigrated to the U.S about 8 years ago. I studied Architecture in Vietnam and after moving to the U.S, I started another degree and attended Northeastern University, majoring in IT. I have worked as a Data Analyst at various Fortune 500 tech companies in Silicon Valley for the past 5 years. I am taking the MicroMaster to prepare myself for the OMSA program and am also looking forward to learning from the course and others as well.

3. Frank Corrao, fcorrao
   My name is Frank Corrao. I am a quantitative developer and team leader at a quantitative hedge fund with 20 years of professional experience. I manage the team responsible for our simulation platform and portfolio optimization software. I have a computer science background and am self-taught in various aspects of quantitative trading and analytics. I am admitted to the Analytics masters program in the fall and am studying to build a more rigorous knowledge base in analytics.

4. Michael Obin, Obin_michael
   My Name is Michael Obin. My professional experience entails roles in financial analysis, quantitative analysis, and forecasting. I completed my undergraduate studies in Actuarial Science & Commerce, and I am a CFA charterholder. I am enrolled in the MircoMasters program with the aim for expanding my analytical skill set, and applying to join the next cohort of the OMSA program.

5. Nicholas Coyle, nscoyle (Team Leader)

   My name is Nick and I work in Strategy & Analysis for a SVOD platform. I work across Finance, Data Science, and Content to better understand our core business drivers and optimize our content investment strategy. Prior to this job, I was a management consultant (focusing on M&A) and studied Economics in college. I am enrolled in this course (and the MicroMaster program) to hone my technical skills and prepare for OMSA.

OBJECTIVE/PROBLEM (5 points)

Project Title: Recession-Proof Portfolio

Background Information on chosen project topic:

Our goal is to identify a portfolio of stocks which will outperform the market in a recession. To do so, we plan to leverage data for securities in the S&P 500 during the 'Great Recession' (period of December 2007 – June 2009), to better understand factors which are indicative of resilience during economic downturns. We will then use these findings to infer which securities may be strong investments during the present period of economic uncertainty.

Problem Statement (clear and concise statement explaining the purpose of your analysis and investigation):

The purpose of our analysis is to identify stock attributes and performance trends which are indicative of resilience during economic downturns. In doing so, we would be better positioned to design a portfolio to minimize losses during times of uncertainty.

State your Primary Research Question (RQ):

Which stock attributes and performance trends (e.g. company industry, profit margin, volatility, TTM growth, etc) were most predictive of 'market-outperformance' during the 'Great Recession'?

Add some possible Supporting Research Questions (2-4 RQs that support problem statement):

1. Which securities were part of the S&P 500 during the 'Great Recession' (period of December 2007 – June 2009)?
2. Which of those securities had the highest alphas during that period? The lowest?
3. Which stock attributes and performance trends were most predictive of high alphas during the period? low alphas?
4. Which of those indicators can be extrapolated to periods outside of the 'Great Recession' for use in future portfolio design?

Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)

By identifying factors which are related to resilience during economic downturns, we can be better positioned to design a portfolio to minimize losses during times of uncertainty.

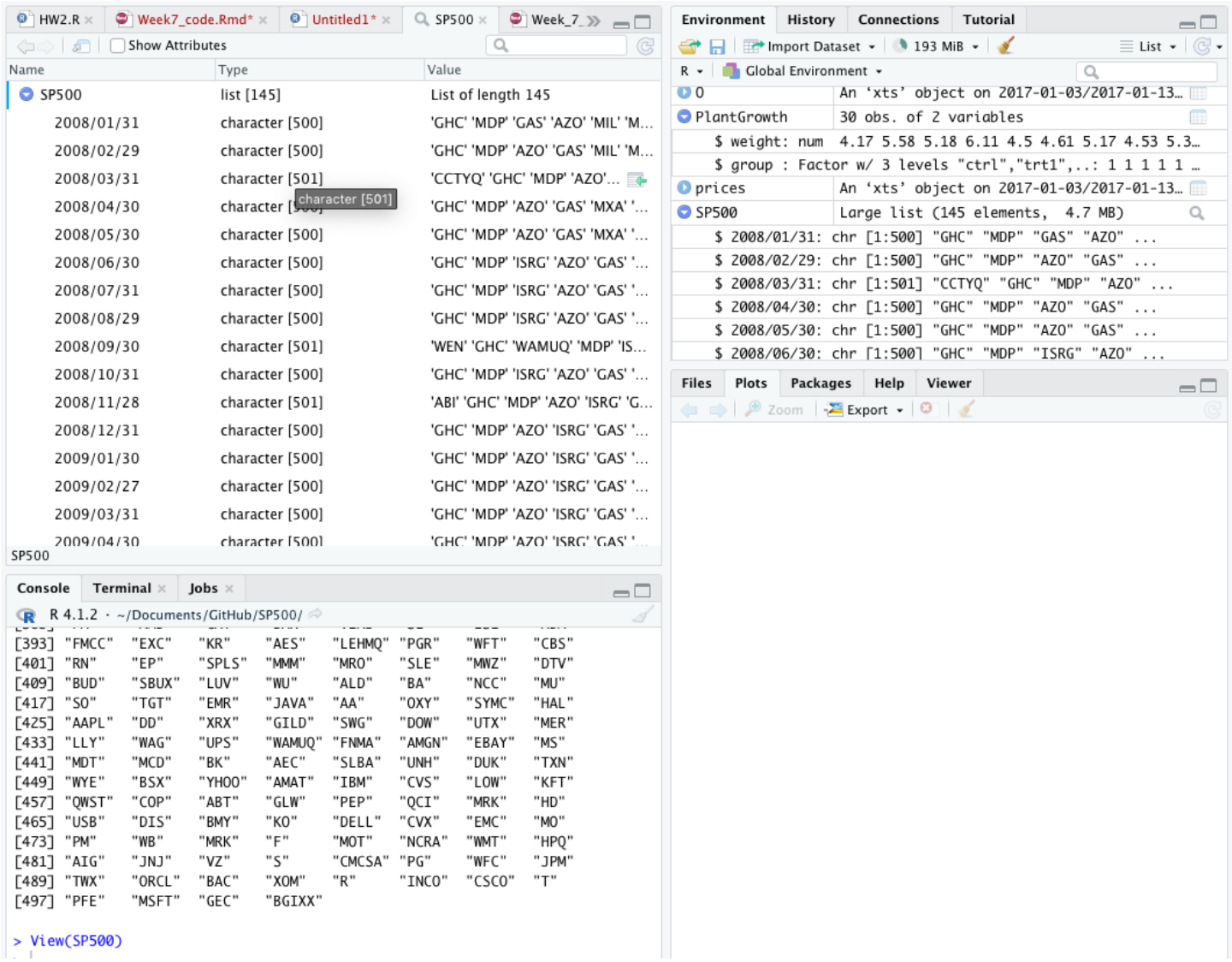
Additionally, these insights may be useful to individual companies looking to optimize their financial structure such that it is resilient to macro-economic factors.

DATASET/PLAN FOR DATA (4 points)

Data Sources (links, attachments, etc.):

Data Description (describe each of your data sources, include screenshots of a few rows of data):

1 - List of S&P500 Historical Constituents from 2008/01/31 - 2019/02/27 - Source



2 - Time Series Stock Price and Return Data From Quantmod - Source

```r
library(tidyquant)
library("rjson")

options("getSymbols.warning4.0"=FALSE)
options("getSymbols.yahoo.warning"=FALSE)

tickers = c("AAPL", "NFLX", "AMZN", "K", "O")
getSymbols(tickers,
           from = "2017-01-01",
           to = "2017-01-15")

prices <- map(tickers,function(x) Ad(get(x)))
prices <- reduce(prices,merge)
colnames(prices) <- tickers

head(prices)

# Give the input file name to the function.
SP500 <- fromJSON(file = "sp500_constituents.json")
head(SP500)

```

13:31 (Top Level) R Script

Console Terminal Jobs

R 4.1.2 · ~/Documents/GitHub/SP500/

```
[1] "AAPL" "NFLX" "AMZN" "K"    "O"
>
> prices <- map(tickers,function(x) Ad(get(x)))
> prices <- reduce(prices,merge)
> colnames(prices) <- tickers
>
> head(prices)
               AAPL   NFLX    AMZN        K        O
2017-01-03 27.25764 127.49 37.6835 60.56914 44.03095
2017-01-04 27.22714 129.41 37.8590 60.41230 44.68961
2017-01-05 27.36559 131.81 39.0225 60.35451 45.89205
2017-01-06 27.67067 131.07 39.7995 60.37101 45.83079
2017-01-09 27.92412 130.95 39.8460 59.54549 45.49379
2017-01-10 27.95229 129.89 39.7950 59.15749 44.95002
>
```

3 - Stock Attribute Data (Sector, P/E Ratio, Market Cap) - Source

Sector	Ticker	Company	Market Capitalization	PE Ratio
Commercial Services	SPGI	S&P Global Inc.	$ 111,125,098,275.50	40.7
Commercial Services	MCO	Moody's Corporation	$ 71,309,376,207.33	35
Commercial Services	CPRT	Copart, Inc.	$ 34,854,843,266.37	36.6
Commercial Services	EFX	Equifax, Inc.	$ 34,648,432,194.28	51.9
Commercial Services	FLT	FleetCor Technologies, Inc.	$ 18,765,806,553.47	23.1
Commercial Services	CRL	Charles River Laboratories International, Inc.	$ 18,273,662,894.10	48.7
Commercial Services	OMC	Omnicom Group Inc.	$ 15,705,948,134.78	11.4
Commercial Services	IPG	Interpublic Group of Companies, Inc. (The)	$ 14,819,801,161.35	21.1
Commercial Services	RHI	Robert Half International Inc.	$ 12,185,090,044.53	23.8
Commercial Services	NLSN	Nielsen N.V.	$ 7,574,271,691.93	29.6
Communications	VZ	Verizon Communications Inc.	$ 217,110,199,611.98	9.8
Communications	T	AT&T Inc.	$ 181,595,632,815.91	202.5
Communications	TMUS	T-Mobile US, Inc.	$ 142,929,192,342.43	43.4
Communications	LUMN	Lumen Technologies, Inc.	$ 13,003,456,094.10	
Consumer Durables	TSLA	Tesla, Inc.	$ 1,204,896,966,220.99	342.2

Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)

Part I - Which securities outperformed the market during the 'Great Recession'?

- Dependent variable: cumulative returns between period of December 2007 – June 2009
  - *from source 1 and source 2*
- Independent variable: S&P composite returns period of December 2007 – June 2009
  - *from source 1 and source 2*
- This will allow us to solve for each stock's alpha

Part II - Which stock attributes and performance trends were most predictive of high alphas during the period?

- Dependent variable: did stock materially outperform the market in the period (yes or no)?
  - *output from Part I*
- Independent variables: stock industry, volatility, TTM cumulative performance (leading into December of 2007), market cap, p/e ratio, profit margin, gross revenue, etc
  - *volatility and TTM cumulative performance from source 1 and source 2*
  - *remaining stock attributes from data source 3*
- This will help us understand which factors are most related to resilience during economic downturns

Sample script that get stock data and S&P500 companies with some other data points

APPROACH/METHODOLOGY (8 points)

Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))

Part I - Which securities outperformed the market during the 'Great Recession'?

First, we will identify the securities which were listed on the S&P 500 during the period of December 2007 – June 2009. We will then run a linear regression between each individual stock's cumulative returns during that period and the S&P 500's composite performance. This will yield 'alpha' values for each security, reflecting 'marked adjusted performance'.

We will use these ‘alphas’ to group the securities into two buckets: ‘did perform the market’ and ‘did not perform the market’.

Part II - Which stock attributes and performance trends were most predictive of high alphas during the period?

Next, we will use logistic regression to identify the relationship between key stock attributes / performance trends and whether they outperformed the market. ‘Market Outperformance’ will be a binary dependent variable and is the output from ‘Part I’. ‘Key Stock Attributes + Performance Trends’ will include things like: stock industry, volatility, TTM cumulative performance (leading into December of 2007), profit margin, gross revenue, etc.

This part of the analysis will require several potential transformations: scaling (as features will have different magnitudes), outlier removal (as some observations may create noise in the data), joining across data sets, etc.

We will be sure to reserve most of our data for training (~70% of observations), with the remainder (~30% of observations) for testing. We can also explore stepwise selection to optimize the features we include.

Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement

We expect companies which serve fundamental needs (e.g. healthcare, consumer, energy, etc) to prove most resilient during downturns. Less resilient players may include providers of luxury goods (e.g. travel, hospitality, etc) or companies with ‘growth stock’ valuations.

The attributes and performance factors which are indicative of these outcomes may be historical volatility, P/E ratios, profit margin, etc. Companies with high values for these factors may be riskier and experience greater losses. Companies with low values for these factors may be more stable and demonstrate more resilience.

What business decisions will be impacted by the results of your analysis? What could be some benefits?

Understanding the factors which are indicative of resilience during an economic downturn could enable us to design a portfolio which minimizes losses during times of uncertainty.

Additionally, these insights may be useful to individual companies looking to optimize their financial structure such that it is resilient to macro-economic factors.

PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention key dates you hope to achieve certain milestones by:

- June 29th - have all data sets pulled, cleaned, and joined
- July 6th - have initial version of analysis completed, with high-level findings summarized
- July 13th - finalize code + complete presentation materials
- July 15th - record final presentation

Appendix (any preliminary figures or charts that you would like to include):

Sample calculations for returns of individual stocks in the S&P 500 (avg returns from ‘08 - ‘22)

```r
head(prices)

# Give the input file name to the function.
# SP500 <- fromJSON(file = "sp500_constituents.json")
# head(SP500)

# install.packages("jsonlite", repos="https://cran.rstudio.com/")
library("jsonlite")

json_file <- 'https://datahub.io/core/s-and-p-500-companies-financials/datapackage.json'
json_data <- fromJSON(paste(readLines(json_file), collapse=""))

# get list of all resources:
print(json_data$resources$name)

# print all tabular data(if exists any)
for(i in 1:length(json_data$resources$datahub$type)){
  if(json_data$resources$datahub$type[i]=='derived/csv'){
    path_to_file = json_data$resources$path[i]
    data <- read.csv(url(path_to_file))
    # print(data)
  }
}

head(data)
head(prices)
merge_data <- merge(x = prices, y = data, by.x="symbol", by.y="Symbol")
head(merge_data)
keep <- c("symbol", "date", "volume", "adjusted", "Name", "Sector")
df1 <- merge_data[keep]
head(df1)

daily_sector = df1 %>% group_by(symbol, Name, Sector) %>%
  tq_transmute(select     = adjusted,
               mutate_fun = periodReturn,
               period     = "daily") %>%
  ungroup()
head(daily_sector)

avg_return =daily_sector %>%
  group_by(Name, Sector) %>%
  summarise(avg_return = round(mean(daily.returns), 4),Volatility =    sd(daily.returns)) %>%
  arrange(desc(avg_return), desc(Volatility))
avg_return %>% head()

avg_return %>% head(20) %>% ggplot(aes(reorder(Name, -avg_return), avg_return, fill = avg_return))+
  geom_col()+
  coord_flip()+
  labs(title = "Some Tickers Average Return in SP500 From 2008 - Present", x = "Ticker", y = "Average Return")+
  theme_classic()+
  theme(legend.position="none")
```