![Brandeis University]

Advanced Topics in Healthcare Data Analytics and Data Mining

Case Study: Coronavirus Stock Market Crash

Erin Cao

Bo Chen

Qingyue Su
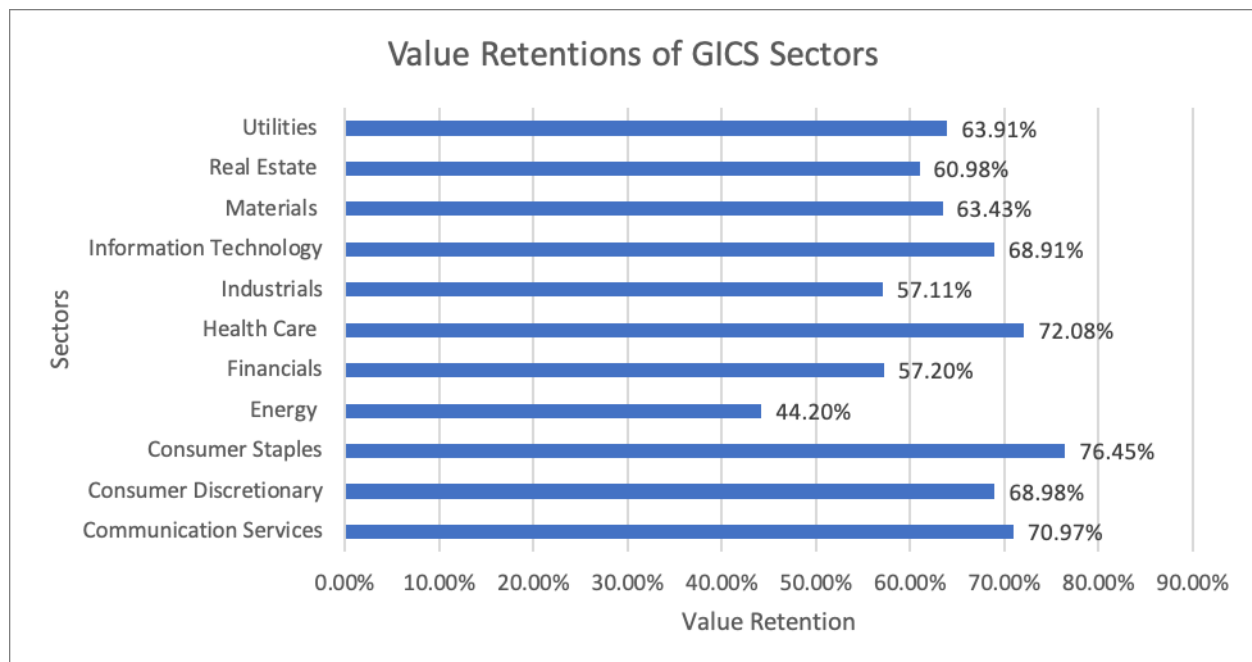
Peihan Tian

Jiujun Zhang

Kaihang Zhao

04/14/2020

Introduction

Undoubtedly, the Coronavirus crisis has shocked global economic and financial markets. From observing the value retention index of S&P 500 from Feb 14 to Mar 23, we broadly explored its negative impact on each industry, from consumer discretionary, energy, financial services, to healthcare. We also discussed some government rescue packages and social implications of the market crash.

I. Making sense of sector trend movements using Microeconomics theory

To better know the sector performance on protecting value when faced with the economic strike of coronavirus, we calculated the weighted average percentage value retentions of GICS sectors based on the market cap and the historical stock prices of 505 companies within the S&P 500 respectively on February 14th and March 23rd, which is shown as the graph below.

Value Retentions of GICS Sectors

Sector	Value Retention
Utilities	63.91%
Real Estate	60.98%
Materials	63.43%
Information Technology	68.91%
Industrials	57.11%
Health Care	72.08%
Financials	57.20%
Energy	44.20%
Consumer Staples	76.45%
Consumer Discretionary	68.98%
Communication Services	70.97%

Graph1 Value Retentions of GICS Sectors

From the graph above, we can draw many conclusions behind it. Here, we take the consumer discretionary and consumer staples as the first example to test how important the income elasticity plays. As we all know, in economics, income elasticity of demand measures the responsiveness of the quantity demanded for a good or service to a change in income. It is calculated as the ratio of the percentage change in quantity demanded to the percentage change in income, whose mathematical function is shown as below.

We know that income elasticity for luxury items is much higher than essential goods and services, which means that income elasticity of consumer discretionary is much higher than that of consumer staples. However, what value retention shows is the opposite, consumer staples have a stronger ability to maintain values than consumer discretionary, which makes sense when it comes to a crisis threatening people's lives.

Since customer discretionary means that something is not a must in our daily life, so it is usually more expensive than the commodities, which we can tell from the graph below. The range of consumer discretionary index is between 700 and nearly 1100, which almost hit the peak on February 14th (1044.57), and got to the bottom on March 23rd (715.35). During this period (February 14th - March 23rd), the index dropped nearly 31.5%, which is a very huge drop actually.

Graph2 S&P 500 Customer Discretionary (Sector)

By contrast to what customer discretionary shows, the customer staples index is relatively more stable. But it still decreased a lot because of the outbreak of Coronavirus. Since customer staple represents some common stuff we buy in our daily lives, the range here (500 - 665) is much smaller than that in customer discretionary (700 - 1100), and the overall value is much higher than that in customer discretionary. Since no matter how severe the flu is, people still have to eat salt, use toilet paper, drink milk, or something like that, the price of customer staple is more stable than that of customer discretionary.

Graph3 S&P 500 Customer Staples (Sector)

II. Energy Sector: Story of a Deadly Crash

From the graph in question one, we can tell that the energy sector possesses the lowest value retention among all the sectors, which only maintains 44.2% of the original stock price after experiencing the sudden strike of coronavirus. To explore the reason why this happens, we can figure out that actually the energy stock index has already dropped before the outbreak of Coronavirus. What Coronavirus does to the energy sector just to make the decrease faster and sharper, which drops from 411.57 on February 14th to 183.24 on March 23rd (55.48%).
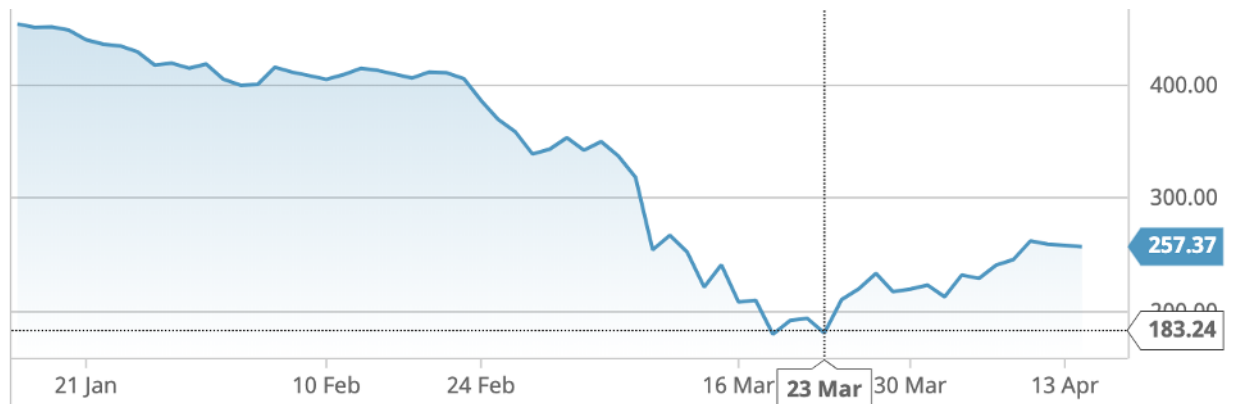
Graph4 S&P 500 Energy (Sector)

Speaking of the reason why the oil price dropped before the Coronavirus outbreak, it is something relevant to the demand part of raw oil. Since Russia said no to Saudi Arabia on the strategy of decreasing the production of crude to increase the price together, Saudi Arabia decided to enlarge its production in a short time to drag the price down as a revenge to Russia, which eventually led to the sharp drop in the price of oil.

Then, we moved back to the period after the coronavirus broke out. Worries about fallout from the coronavirus are pummeling already battered energy stocks, upending a nascent recovery that followed a spate of better-than-expected earnings and dividend increases and signaling the magnitude of investors' concern about the virus's effect on crude demand, which is also shown in the stock price. Energy shares in the S&P 500 have fallen significantly more than the broader index's other groups.

Investors are still trying to gauge the magnitude of the virus's hit on oil demand, as the epidemic picks up outside China. But, the uncertainty in the energy demand market still makes the potential not that good, which is just like Rebecca Babin, senior energy trader at CIBC Private Wealth Management, said that energy companies are the first thing people look to sell in times like these. That's also the reason why the energy industry hurt the worst among all these industries when faced with the strike from Coronavirus.

III. Financial Sector Crash

In S&P 500, The financial sector includes commercial banks, insurance companies, consumer lenders, investment firms and the like, such as, JP Morgan Chase, Bank of America, Prudential Financial and Discover Financial.

Under the current covid-19 situation, we are trying to explain the reason behind the relationship between value retention of the financial sector and virus. From the dataset, we can see that there are 66 companies in this sector, through calculation, the average value retention of the financial sector is 60.54%, and the average value retention of the weighted financial sector is 57.20%. We can conclude by comparison that the financial sector value retention ranked the 9th during the coronavirus pandemic, which means the virus has great impact on this industry. Next, we are trying to figure out how the financial sector is influenced from 3 aspects.

From a credit perspective, the direct causation is the strong connection between the banking system and business in different industries. As the retail customer and corporate clients of the banking system are struggling with managing their risks, they would use credit lines or skip loan or credit card payments, which would cause damage to the side of banking. This would be a vicious circle, as the situation became worse, credit losses will be spread among most sectors, including small businesses, which will hit the financial sector harder again.

From a liquidity perspective, all industries may face the problems of the simultaneous supply and demand shock, which means the capital market would be less accessible. Lending standards could become higher and liquidity could become an urgent issue. In this case, banks have to increase their internal liquidity-management practices to provide strong backing. There is a ratio of liquidity coverage ratio as a measure of outflows over a one-month period, it may not be enough to capture all the risks to liquidity from a longer period contagion. Banks have to do something to upgrade their risk models and refinance at central banks.

From an interest rate perspective, global interest achieved a historical low point during this period. Regional banks, which interest income explained about two-thirds of their revenue, were already dealing with this before the current crisis when their average ROE slipped in the second half of 2019. On a more fundamental level, banks will be watching to see if the business environment degrades enough to cause real stress for companies with good operating fundamentals. In this way, defaults and bankruptcies will occur for sure. US banks may be faced with the question of whether to pass on negative interest rates as banks have done in many European countries in recent years as well.

Financial sectors always play a vital role in the functioning of the economy. They are not simply commercial enterprises but provide important services and support to individuals and communities. They should carefully draw on the lessons that the current situation offers and build a much higher degree of both operational and financial resiliency, to resist the risk and further deterioration and challenge in the future.

IV.The dilemma of Health Insurance Industry

Health Insurance industry was faced with unprecedented challenges during this tough time. With decreasing revenue and increasing cost, from an income statement view, it is obvious and natural that the economic and financial performance of the health insurance industry went down. From the S&P 500, there are 10 enterprises that provide products and services of health insurance shown below. If we classify them into one new specification, its industry weighted value retention would be 58.8% approximately. That over 2/3 of their assets shrank could demonstrate little confidence and pessimistic attitude of the public on the health insurance industry.

Compa	GICS Sector	GICS Sub Industry	Value Retention	Market Cap(Yahoo Finance)$B
AFL	Financials	Life & Health Insurance	48.65	38.16
ANTM	Health Care	Managed Health Care	58.65	75.15
CNC	Health Care	Managed Health Care	70.14	38.1
CI	Health Care	Managed Health Care	59.03	81.82
GL	Financials	Life & Health Insurance	52.41	11.95
HUM	Health Care	Managed Health Care	57.05	49.65
MET	Financials	Life & Health Insurance	45.01	47.88
PFG	Financials	Life & Health Insurance	42.87	15.48
PRU	Financials	Life & Health Insurance	41.51	37.68
UNH	Health Care	Managed Health Care	65.22	283.41
UNM	Financials	Life & Health Insurance	34.59	6.091

Graph 5 S&P 500 Health Insurance (Sector)

The reason why it suffered a tremendous loss can be explained in two ways; revenue loss and cost rise, as we mentioned before.

Revenue loss impressed the dynamics of enterprises drastically. Firstly, as of last week, the US employment rate increased to 13%, which means the unemployed couldn't maintain their high insurance premium and had to escape from it (The Guardian, 2020). Also, as the coronavirus cases expand, many insurance companies announced that they would waive the out-of-network test fee, which belongs to the copayment part previously. At the same time, with the shortage of medical resources, hospitals have to purchase masks, respirators and other equipment as soon as possible and they would not allow a long process, like one month, to receive payment from insurance companies. Under a shorter business cycle, insurance enterprises should guarantee rich cashflow and not invest in other items for investment income. Finally, it is also not plausible to jack up premiums after the coronavirus crisis in the US. Undoubtedly, the global economy, including the US's, is in a depression stage. When the US gets rid of the negative impact of coronavirus, the income of the public would be lower than before. If they raise premiums a lot, other companies and employees could not afford it, thus running away from their insurance plans.

Cost rise affected the industry even more. Firstly, as said before, insurance enterprises waived part of coronavirus-relevant fees. Customers don't have to afford the copayment part, but insurance enterprises themselves. As of April 12, there are 2,935,006 total tests in the US (JHU Coronavirus Map), and their fee is majorly compensated by insurance enterprises. Moreover,

they have to confront a huge amount of medical cost about coronavirus treatment. Both of them might not be forecasted or foreseen by actuarial calculation. They may underestimate the risk premium of this public health crisis and its shock on the economy. Meanwhile, every plan is actuarially calculated in a one-year base. It is not probable for insurance companies to involve short-term emergency events in time.

Combining all elements, health insurance companies were hit by the 'storm' severely and couldn't get recovery in the short run.

V. Cluster Analysis by Value Retention

In order to better understand the effect of Covid-19 on different industries, we plan to cluster companies by their ability to stand the storm using cluster analysis. Specifically, we try to cluster these companies based on the value retention we calculated at the beginning. Additionally, we focus on K-means and K-medians methods and compare these two to get a better model.
K-means:

```
kcca object of family 'kmeans'

call:
kcca(x = df$Value.Retention, k = 3, control = list(initcent =
"kmeanspp"))

cluster sizes:

  1   2   3
154 246 105
```

K-medians:

```
kcca object of family 'kmedians'

call:
kcca(x = df$Value.Retention, k = 3, family =
kccaFamily("kmedians"),
    control = list(initcent = "kmeanspp"))

cluster sizes:

  1   2   3
199 197 109
```
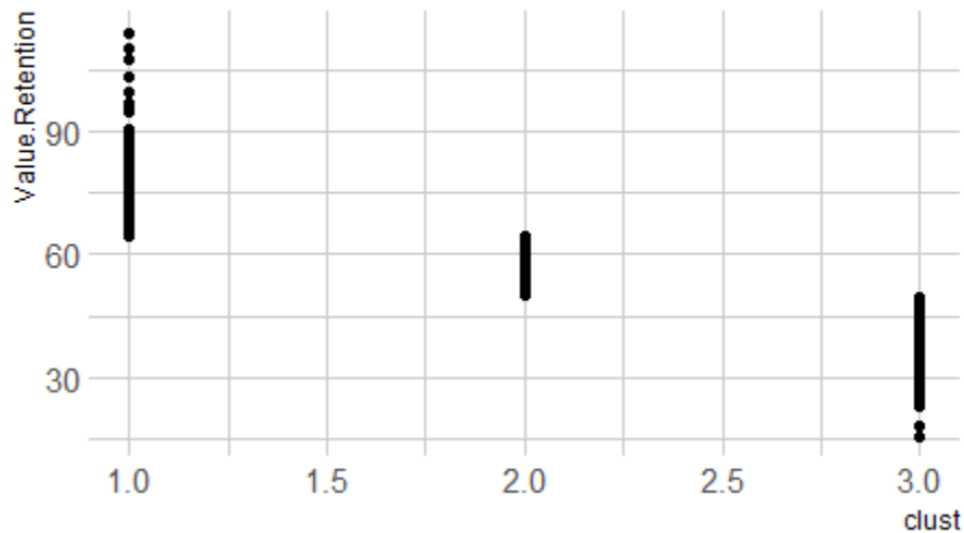
Graph 6 Cluster Analysis

According to the results generated from R, we find that K-medians is a better model in this case because the number of companies in the cluster with the lowest number of companies is larger in the K-medians model (109>105).

Graph 7 Cluster Analysis

Then we just apply the cluster results from K-medians and relabel cluster 1 as HVR, cluster 2 as MVR and cluster 3 as LVR with the HVR being cluster of elite strong companies and LVR the cluster of the most vulnerable companies.

We end up the contingency table which well illustrates each sector performance under this pandemic:

GICS.Sector	HVR	LVR	MVR
Communication Services	13	2	11
Consumer Discretionary	11	29	24
Consumer Staples	27	2	4
Energy	1	23	3
Financials	10	21	35
Health Care	39	NA	21
Industrials	21	15	35
Information Technology	44	4	23
Materials	9	3	16
Real Estate	11	9	11
Utilities	13	1	14

Graph 8 Cluster Analysis

Apparently, as we move up the ladder of Value Retention from LVR to MVR and finally to the elite group HVR, the Consumer Discretionary sector, the Financial sector, the Energy sector disappear so the companies in these sectors are mainly vulnerable under this situation.

Additionally, the Communication Services sector, the Health Care Sector, the Consumer Staples sector, and the Information Technology sector come to the surface. We learn that covid-19 negatively impacts different industries in different degrees and sectors like Consumer Staples, Healthcare and Information Technology were slightly influenced and they have relatively high retention value. On the other side, sectors like Energy, Financials, and Consumer Discretionary were significantly influenced and they have relatively low retention value.

VI. Healthcare Sub-Industries

We are motivated to break down Sector Health Care into two sub-industry of Insurance Companies and all other Health Care companies.

We also sort the Value Retention variable from smallest to largest and divide the 505 companies into five equally sized quintiles and labelled them from 1 to 5. Then we derive a contingency table which illustrates the number of companies in each quantile for the Insurance sector and all other Health Care sectors.

Frequency table:

	2	3	4	5
Health Insurance	1	2	2	0
Healthcare Only	9	8	12	26

Proportion table:

	2	3	4	5
Health Insurance	0.2	0.4	0.4	0
Healthcare Only	0.164	0.145	0.218	0.473
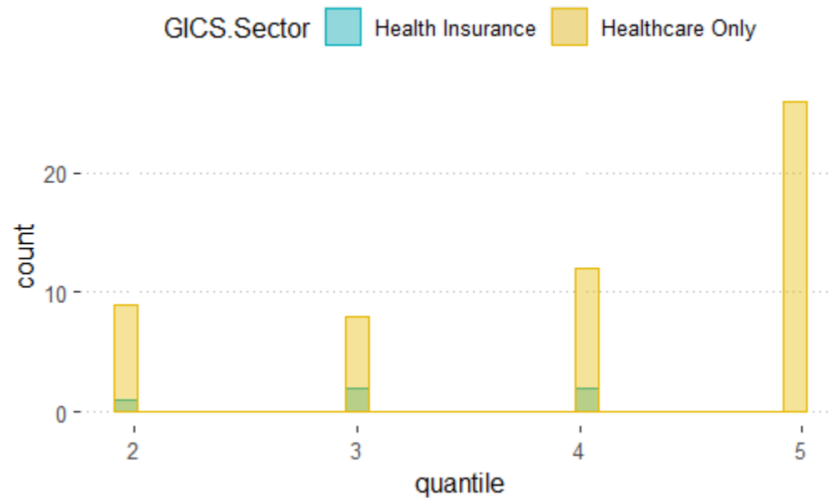
Graph 9

Based on the table and graph, it is easy to infer that Healthcare Only companies have a better appearance in the high retention value group as a large proportion of Healthcare Only companies are in quantile 5 which is the top 101 group. However, most Health Insurance companies are in quantile 3 and 4. In other words, Health Insurance companies literally have worse performances than Healthcare Only companies. As we mentioned in IV, Health Insurance companies are in a dilemma under these uncertain circumstances.

VII. Government Rescue Package: Sector Healthcare Becomes A Darling

Currently, the most difficult challenges for hospitals in responding to COVID-19 are the following:

1, Widespread Shortages of PPE: Hospitals reported that widespread shortages of PPE put staff and patients at risk. Hospitals reported that heavier use of PPE than normal was contributing to the shortage and that the lack of a robust supply chain was delaying or preventing them from restocking PPE needed to protect staff. Hospitals also expressed uncertainty about availability of

PPE from Federal and State sources and noted sharp increases in prices for PPE from some vendors.

2, Increased Costs and Decreased Revenue: Hospitals described increasing costs and decreasing revenues as a threat to their financial viability. Hospitals reported that ceasing elective procedures and other services decreased revenues at the same time that their costs have increased as they prepare for a potential surge of patients. Many hospitals reported that their cash reserves were quickly depleting, which could disrupt ongoing hospital operations.

3, Anticipated Shortages of Ventilators: Anticipated shortages of ventilators were identified as a big challenge for hospitals. Hospitals reported an uncertain supply of standard, full-feature ventilators and in some cases used alternatives to support patients, including adapting anesthesia machines and using single-use emergency transport ventilators.

4, Financial Assistance: All types of hospitals, and especially small rural hospitals, requested financial assistance, including faster and increased Medicare payments, and loans and grants.

To solve these challenges, government could assist them in several different ways:

1, government intervention and coordination for distributing limited medical supplies.
Many hospitals noted that they were competing with other providers for limited supplies, and that government intervention is necessary. New York Gov. Andrew Cuomo, a Democrat, suggested with ventilators at a premium and coronavirus numbers peaking along different timelines in different states, loaning resources according to need was the only solution to the problem.

2, government injection of free money into hospital systems.
On April 9, CMS hiked the amount of Medicare accelerated payments. It's getting to providers this week from its original estimate of $34 billion to more than $51 billion. This funding is in the form of loans and is separate from the $30 billion in CARES funding HHS also announced last

week. The Trump administration also plans to roll out the first $30 billion in funding from the latest stimulus package to hospitals. These actions can help hospitals to survive in the COVID-19 temporarily and provide necessary financial support for providers and nurses in the front line. What is more, as the prices of PPE continue to increase, the government can afford these medical supplies. Overall, this stimulus package can greatly relieve stress for the hospital systems.

The Congress also proposed 2 acts to fight against the COVID-19:

1, Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020:

This bill provides $8.3 billion in emergency funding for federal agencies to respond to the coronavirus outbreak.

This division provides FY2020 supplemental appropriations for the Department of Health and Human Services (HHS), the State Department, and the Small Business Administration to respond to the coronavirus outbreak.

The division funds programs that address issues such as developing, manufacturing, and procuring vaccines and other medical supplies; grants for state, local, and tribal public health agencies and organizations; loans for affected small businesses; evacuations and emergency preparedness activities at U.S. embassies and other State Department facilities; and humanitarian assistance and support for health systems in the affected countries.

The supplemental appropriations are designated as emergency spending, which is exempt from discretionary spending limits.

2, Coronavirus Aid, Relief, and Economic Security Act (CARES Act):

The CARES Act contains a number of health-related provisions focused on the outbreak in the United States, including paid sick leave, insurance coverage of coronavirus testing, nutrition assistance, and other programs and efforts. It also includes support for the global response. Highlights are provided below, followed by summaries of provisions in the tables; not included

are provisions that are not closely related to health but which are important aspects of the response to the outbreak, such as those pertaining to support for small businesses and severely stressed sectors of the U.S. economy and childcare.

VIII. Social Implications of An Uneven Market Crash

It has been reported by many mainstream media that jobs in service sectors got slammed early in the coronavirus crisis and very soon the unemployment concern spread out through almost all industries. Countless neighborhood businesses have been shuttered and many local stores are closing.

During the first several months, uncertainty about how things could go pushed many firms to wait and see under this first-time-ever situation. However, businesses are hard to survive with low revenue and previous level labor cost.

To keep the business running, companies are transferring to remote working due to social distancing. White-collar employees are obviously not completely immune from the hardships but their jobs are the simplest to make remote while work-from-home could not work for blue-collar groups so much. On the other side, there was a significant higher training cost for white-collar employees than blue-collar workers which means the company will cost a lot more to hire new white-collar employees when things are back to normal. Therefore, in short term employers tend to lay off blue-collar workers if their business demands an immediate solution.

According to McKinsey's report, we can analyze the inequality implications of the displacement in the market by grouping occupations into low, medium and high vulnerability.

Low-vulnerability jobs are those that are deemed essential (such as doctors), require no physical proximity to others (that is, can be done remotely), or are likely to guarantee pay even if workers are furloughed (such as K–12 public-school teachers). We assume these workers continue to earn their full income during shutdowns. Medium-vulnerability jobs (such as those in manufacturing) require workers to be in proximity to coworkers but not the public; we assume that shutdowns

will affect 30 to 50 percent of these jobs. Finally, high-vulnerability jobs are nonessential roles that involve exposure to the public; we assume that shutdowns affect 70 to 90 percent of these jobs. Examples include restaurant servers or cinema cashiers.

Below shows a chart from the US bureau that describes employed persons by industry and class of worker.

HOUSEHOLD DATA
Table A-8. Employed persons by class of worker and part-time status
[In thousands]

Category	Not seasonally adjusted			Seasonally adjusted					
	Mar. 2019	Feb. 2020	Mar. 2020	Mar. 2019	Nov. 2019	Dec. 2019	Jan. 2020	Feb. 2020	Mar. 2020
CLASS OF WORKER									
Agriculture and related industries	2,243	2,332	2,263	2,336	2,356	2,533	2,412	2,466	2,399
Wage and salary workers(1)	1,558	1,704	1,580	1,671	1,661	1,849	1,750	1,829	1,715
Self-employed workers, unincorporated	667	612	661	647	692	684	647	645	660
Unpaid family workers	18	16	22	-	-	-	-	-	-
Nonagricultural industries	154,197	155,686	152,904	154,346	156,167	156,241	156,337	156,283	153,359
Wage and salary workers(1)	145,493	146,861	144,149	145,578	147,275	147,431	147,467	147,347	144,494
Government	21,077	21,330	21,414	20,706	21,342	21,323	21,390	20,943	21,081
Private industries	124,417	125,531	122,735	124,860	125,996	126,069	126,042	126,282	123,412
Private households	808	781	712	-	-	-	-	-	-
Other industries	123,609	124,750	122,023	124,057	125,157	125,250	125,234	125,487	122,693
Self-employed workers, unincorporated	8,659	8,780	8,717	8,733	8,856	8,821	8,808	8,897	8,818
Unpaid family workers	45	45	37	-	-	-	-	-	-
PERSONS AT WORK PART TIME(2)									
All industries									
Part time for economic reasons(3)	4,621	4,600	5,879	4,517	4,288	4,148	4,182	4,318	5,765
Slack work or business conditions	2,949	3,037	4,159	2,887	2,634	2,657	2,655	2,776	4,043
Could only find part-time work	1,413	1,288	1,404	1,347	1,259	1,215	1,294	1,317	1,321
Part time for noneconomic reasons(4)	21,738	22,902	20,764	21,332	21,532	21,586	22,154	22,175	20,601
Nonagricultural industries									
Part time for economic reasons(3)	4,548	4,501	5,788	4,446	4,221	4,111	4,091	4,225	5,681
Slack work or business conditions	2,895	2,966	4,090	2,834	2,599	2,613	2,580	2,719	3,965
Could only find part-time work	1,410	1,282	1,392	1,343	1,254	1,209	1,308	1,313	1,312
Part time for noneconomic reasons(4)	21,434	22,473	20,392	21,025	21,163	21,180	21,784	21,770	20,236

Graph 10

Employment rate in agriculture and related industries went down by 5.29% and nonagricultural sectors dropped by 1.84% quarterly from December 2019 to March 2020. There tend to be more blue-collar workers in agricultural sectors compared to nonagricultural. Our analysis finds that lockdowns unevenly affect low-income workers. People who were living paycheck to paycheck do not have the financial cushion to absorb a shock of this magnitude. In addition, many of the

lowest-paid Americans who are still working may be risking exposure to the virus as they perform vital services in the economy.

It is also important to try to understand this change eliminating reasons other than economic recession. We noticed that part time workers for noneconomic reasons decreased by 4.56% while part time workers for economic reasons increased significantly, 38.98%. We would say many people have to do part time jobs possibly because they cannot stay full time because of the economic recession.

Below is another chart from the UB bureau that describes unemployed persons by industry and class of worker.

HOUSEHOLD DATA
Table A-14. Unemployed persons by industry and class of worker, not seasonally adjusted

Industry and class of worker	Number of unemployed persons (in thousands)		Unemployment rates	
	Mar. 2019	Mar. 2020	Mar. 2019	Mar. 2020
Total, 16 years and over[1]	6,382	7,370	3.9	4.5
Nonagricultural private wage and salary workers	4,869	5,964	3.8	4.6
Mining, quarrying, and oil and gas extraction	26	52	3.6	6.2
Construction	490	658	5.2	6.9
Manufacturing	475	636	2.9	4.2
Durable goods	254	336	2.5	3.5
Nondurable goods	221	300	3.6	5.3
Wholesale and retail trade	917	955	4.7	4.9
Transportation and utilities	262	380	3.6	5.0
Information	105	47	4.0	1.8
Financial activities	280	230	2.8	2.2
Professional and business services	678	838	3.9	4.8
Education and health services	583	778	2.3	3.2
Leisure and hospitality	827	1,131	5.8	8.1
Other services	224	258	3.4	3.9
Agriculture and related private wage and salary workers	195	139	11.5	8.3
Government workers	405	490	1.9	2.2
Self-employed workers, unincorporated, and unpaid family workers	375	327	3.8	3.4

Graph 11

If we look at the statistics in that last column, we can conclude that agriculture and related private wage and salary workers suffered the most from unemployment. In terms of nonagricultural industries, more people in Mining, quarrying, and oil and gas extraction, Construction and Leisure and hospitality are losing jobs compared to other sectors. This is also

consistent with our insights since these industries require more blue-collar workers than those highly educated and elite workforce centered industries like Information, Financial activities and Government workers.

In conclusion, in months and quarters we see a white-collar vs blue-collar contrast in the groups with better chance of survival vs. certain sectors already at the risk of collapse and blue-collar tends to suffer more from the recession.

Brandeis
UNIVERSITY

Reference:

The Guardian.2020.US unemployment rises 6.6m in a week as coronavirus takes its toll.

https://www.theguardian.com/business/2020/apr/09/us-unemployment-filings-coronavirus

JHU.2020. Coronavirus Map https://coronavirus.jhu.edu/map.html

Coronavirus Aid, Relief, and Economic Security Act (CARES Act)
https://www.kff.org/global-health-policy/issue-brief/the-coronavirus-aid-relief-and-economic-security-act-summary-of-key-health-provisions/

Coronavirus Preparedness and Response Supplemental Appropriations Act of 2020
https://www.congress.gov/bill/116th-congress/house-bill/6074

Lives and Livelihoods: Assessing the Near-term Impact Of Covid-19 on Us Workers About the author(s) Susan Lund is a McKinsey Global Institute (MGI) partner based in the Washington - https://www.mckinsey.com/industries/public-sector/our-insights/lives-and-livelihoods-assessing-the-near-term-impact-of-covid-19-on-us-workers

Leadership in the time of coronavirus: COVID-19 response and implications for banks
https://www.mckinsey.com/industries/financial-services/our-insights/leadership-in-the-time-of-coronavirus-covid-19-response-and-implications-for-banks

Coronavirus (COVID-19) impacts banking and capital markets: PwC

https://www.pwc.com/us/en/library/covid-19/coronavirus-banking-and-capital-markets.html

Appendix R script of Cluster Analysis

```
set.seed(3)
kmedian=kcca(df$Value.Retention,k=3,family=kccaFamily("kmedians") ,control=list(initcent="
kmeanspp"))
kmedian
set.seed(3)
kmean=kcca(df$Value.Retention,k=3, control=list(initcent="kmeanspp"))
kmean
clust=clusters(kmedian)
df=df%>%mutate(clust=clust)
#ggplot(df, aes(x=clust, y=Value.Retention)) + geom_point()+theme_ipsum()
df=df%>%mutate(clust=case_when((clust==1)~"HVR",(clust==2)~"MVR",(clust==3)~"LVR"))
df%>%group_by(GICS.Sector,clust)%>%summarise(n=n())%>%
  spread(clust, n)%>%
  kable()
df2=df%>%mutate(GICS.Sector=ifelse(Company=="ANTM"|Company=="CNC"|Company=="
CI"|Company=="HUM"|Company=="UNH","Health Insurance",as.character(df$GICS.Sector)))
df2=df2%>%mutate(GICS.Sector=ifelse(GICS.Sector=="Health Care","Healthcare
Only",as.character(GICS.Sector)))
df2=df2%>%mutate(quantile=ntile(Value.Retention,5))
df3=df2%>%filter(GICS.Sector=="Health Insurance"|GICS.Sector=="Healthcare Only")
t<-table(df3$GICS.Sector,df3$quantile)
ftable(t)
prop.table(t,1)
df3%>%ggplot(aes(x=quantile))+ geom_histogram(aes(color = GICS.Sector, fill =
GICS.Sector),
               alpha = 0.4, position = "identity") +
  scale_fill_manual(values = c("#00AFBB", "#E7B800")) +
  scale_color_manual(values = c("#00AFBB", "#E7B800"))  +theme_pubclean()
```