Data Analytics-ENGM 182
Final Exam Report
Winter 2023

Prof. Erin Mayfield   **Submitted By- Supritha S Rao**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

There has been a constant discussion on whether or not the U.S. is headed for a recession. There are various opinions and thoughts divided over it, many believe that these are mere fears. The National Bureau of Economic Research(NBER), who are the arbiter of recessions, haven't declared any recession as yet. The technical definition of a recession is two consecutive quarters of declining gross domestic product(GDP), but U.S.' NBER define a recession as widespread contraction in economic growth across various macro-economic indicators, such as income, personal consumption, employment, and industrial production. Table 1 from Federal Reserve Bank shows the percentage change, on average, across recessions since 1947 for (a) the period prior to the recession and (b) the first two quarters subsequent to the peak.

Table 1: Percentage change of economic indicators across historical recessions

Indicator	Prior to Recession	Two Subsequent Quarter
Real personal income	0.62%	-0.83%
Household employment	0.30%	-1.42%
Real personal consumption	0.48%	-2.92%
Wholesale-retail sales	0.28%	-0.95%
Industrial Production	0.05%	-2.79%

I.I Report Purpose

This report is to assess the possibility of a recession by supervising and evaluating historical and future trends in U.S. economic indices, in addition to the state-level variation in recession-related indicators. We will extract information from the patterns of clustering the state date over economic indicators and further develop a time series analysis among these indicators from 2020 to 2022, spread across 12 quarters and project what it may look like for the year 2023 and draw conclusions as to whether the fear of the country entering into recession is true or not.

II. DATA

II.I Dataset Description

The two datasets used for computation are "Quarter.csv" and "Qarter.csv". The former dataset includes quarterly economic indicator data for the entire U.S. from the first quarter(Q1) of 2020 to the fourth and the last quarter(Q4) of 2022, spread across 6 variables as given in Table 2. Whereas, the later dataset includes economic data for each state in the U.S, and the data reported for each variable are the most recent values that were publicly reported, across 5 variables as shown in Table 2.

Table 2: Data descriptions for Quarter.csv

Variable Name	Variablel Description
date	Quarter of observation (e.g., 1/1/2020 is Q1 2020)
income	Percentage change (relative to previous quarter) in real personal income excluding current transfer receipts
employment	Percentage change (relative to previous quarter) in employment (nonfarm)
expenditures	Percentage change (relative to previous quarter) in real personal consumption
manufacturing	Percentage change (relative to previous quarter) in real manufacturing & trade industries sale
production	Percentage change (relative to previous quarter) in industrial production index

Table 3: Data descriptions for State.csv

Variable Name	Variablel Description
state	State name
income	Percentage change in personal income from Q2 2022 to Q3 2022
employment	Percentage change in private employment from Q1 2022 to Q2 2022
expenditures	Percentage change in real personal consumption expenditures from 2020 to 2021
gdp	Percentage change in real gross domestic product from Q2 2022 to Q3 2022

II.II Dataset Manipulation & Cleaning

The dataset from State.csv, consisting all the U.S. states are divided based on the 4 regions, the west, northeast, south and Midwest states as in Fig 1. This is done just to have a clear cut idea about the various regions of the state and not to pull out interpretations or for computational purposes. Further, no missing values from the data were encountered. However, from the same dataset, an initial exploration showed that there was a need of data cleaning with the state record of "Alaska", as it was a constant outlier while the formation of clusters and this data was disrupting. Hence, this state was totally omitted as the need to clean this data became necessary for the above reason and also because it geographically positioned way towards the north of U.S. and alongside the top of Canada. Hence, cleaning this data was a necessary step in order to further compute and infer from the datasets.
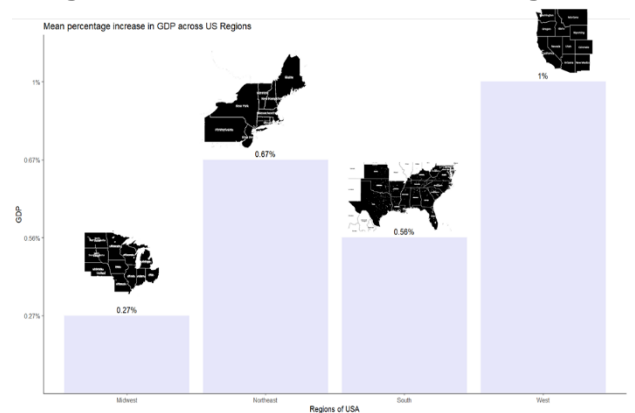
Fig 1: States divided into regions
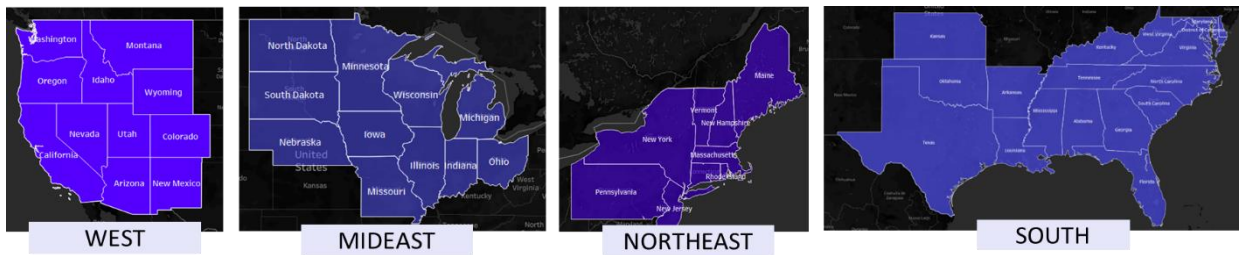
III. MODEL

III.I Regional Analysis

With the conventional belief of defining a recession as a decrease in GDP, Fig 2, shows the mean percentage of GDP across the four regions. We can see the there's a lowest increase of 0.27% in GDP for the Mideast states, followed by an increase of 0.56% in south region and 0.67% in the Northeast states. The highest increase of mean percentage across all the regions was seen in the west region of U.S. This might be because of various tech industries having their major hubs located in this region and hence the contribution to gross domestic product can be seen as the highest.

This specific approach might look good for further interpretations but the best option would be to cluster

Fig 2: Mean % increase in GDP across regions

all the states and the various macro-economic indicator variables, using k-means clustering to know the trends in the data and predict if there will be recession in the year 2023.

III.II K-Means Clustering Model

For the K-Means clustering we consider our "State.csv", the variables and their percentage between quarters. The initial data of all the 5 variables with their values spread across this time frame is shown geographically in Fig 3.

Fig 3: Economic Indicators

2

After the data cleaning & manipulation, when we perform K-Means clustering, various methods like bar graph, nbClust & silhouette, suggests 2 clusters would be the best for the model as seen in Fig 4 below.

Fig 4: Number of clusters
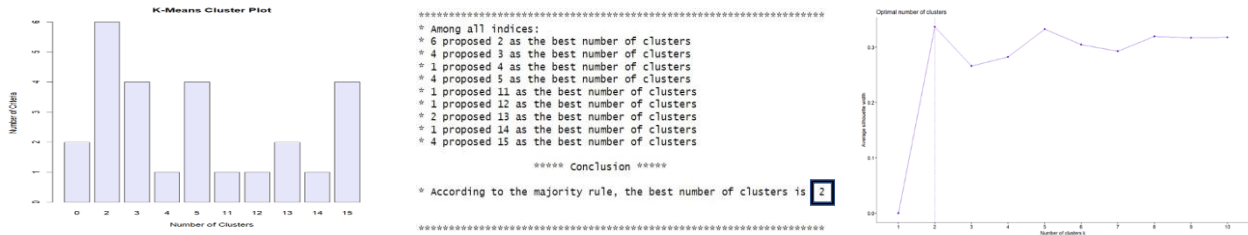
So, on going ahead with the number suggested by the above methods, the data is split across as 2 clusters as shown in Fig 5.

Fig 5: K-Means Cluster
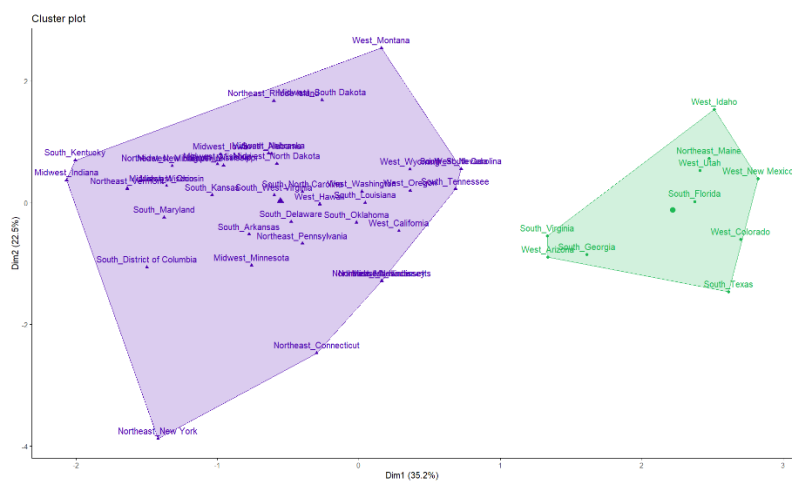
With the cluster plot we see around 58% of variability among the first two variable dimensions of the data. On further computing the mean values of these clusters, we notice that for the first and second cluster, the negative mean percentage value of employment and wages suggest that the country might go into recession. But when you compare it with other variables like GDP, we see that the rates from Q2 2022 TO Q3 2022 is a positive 0.011. When we consider the income economic indicator for cluster 1& 2, we see an increase in the income of 0.023 & 0.0095 while the historical data from Table 1 for income shows that it shot up to 0.62% prior to recession. From the same table when we consider employment, there was a 30% raise prior to recession, but when our mean value table (Table 4) shows negative values for both the clusters with -0.021 and -0.0245% respectively. Overall, we can say that there's no major factor that indicates that there's going to be a recession in the U.S.

Table 4: Mean values of Economic Variables for the clusters

cluster	employment	wages	gdp	income	expenditures
1	-0.021	-0.024	0.011	0.023	0.143
2	-0.0245	-0.0302	0.00475	0.0095	0.122

Fig 6: Geographic representation of cluster along with the mean values


3
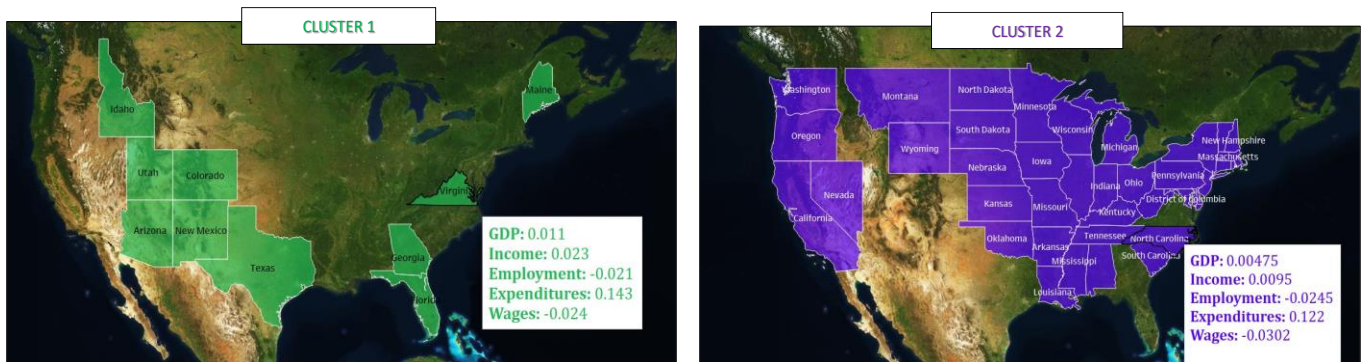
IV. TIME SERIES ANALYSIS

Time series is a statistical methos which analyses the various patterns and trends in the data collected over a period of time and further make predictions about the future values based on the patterns observed. From the graphs (Fig 7), it is evident that there is trend in the time series over the years.

**Fig 7: Changes in economic indicator from 2020 to 2022**

	date	income	employment	expenditures	production	manufacturing
2020						
2020 Q1	1	0.592	0.11871	-1.22374	-1.18424	-1.37232
2020 Q2	7	-5.24168	-11.925	-9.6287	-13.13493	-7.22705
2020 Q3	10	3.8919	5.11537	10.26046	11.25074	11.12107
2020 Q4	4	2.10014	1.44062	1.37144	2.02893	1.06089
2021						
2021 Q1	2	-0.79896	0.7351	3.73941	0.64873	1.97825
2021 Q2	8	1.2735	1.08114	4.50737	1.398	-0.05733
2021 Q3	11	0.62446	1.39105	2.11064	0.93275	-0.08744
2021 Q4	5	0.67372	1.34003	2.29574	1.43447	0.41218
2022						
2022 Q1	3	-0.55908	1.13213	2.15987	0.91378	0.62014
2022 Q2	9	-0.37038	0.79374	2.29099	0.80631	-1.76321
2022 Q3	12	1.07997	0.84668	1.62965	-0.20232	1.28944
2022 Q4	6	0.66141	0.62262	1.25447	-0.76953	0.33494
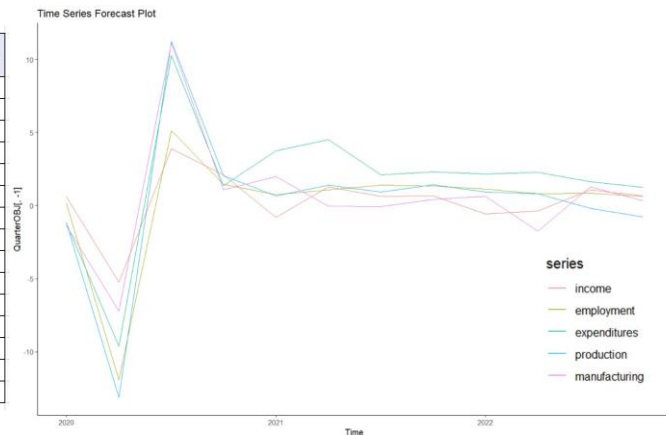
Also, the one very noticeable change that can be observed from the above graph is that the pattern suddenly at the end of 2020 and in the year 2020 we see peaks and troughs varying, the potential reason for this could be because of the pandemic that hit in 2020 which disrupted the complete functionality in and around the world. Everything after the pandemic was to stabilize and get back into the normal functioning and hence over the next 2 years (2021 and 2022) we can observe a constant pattern aiming to not fall further.

Further from the 5 set graphs in Fig 8., shows the economic indicators been decomposed into seasonal, trend and irregular components. We definitely see across all the indicators that the seasonal component captured a cyclical effects around the year . We also encounter a steep decrease trend when it comes to production, which may suggest that the recession could be a possibility if the rate of fall continues. When we look at the last quarter of 2022 the trends of incomes, employment, expenditure and manufacturing don't show a trend that suggests recessions. Again, on comparing the historical data from Table 1, with the last quarter of 2022, we see no such similar trend that indicates that U.S. may be headed into a recession.
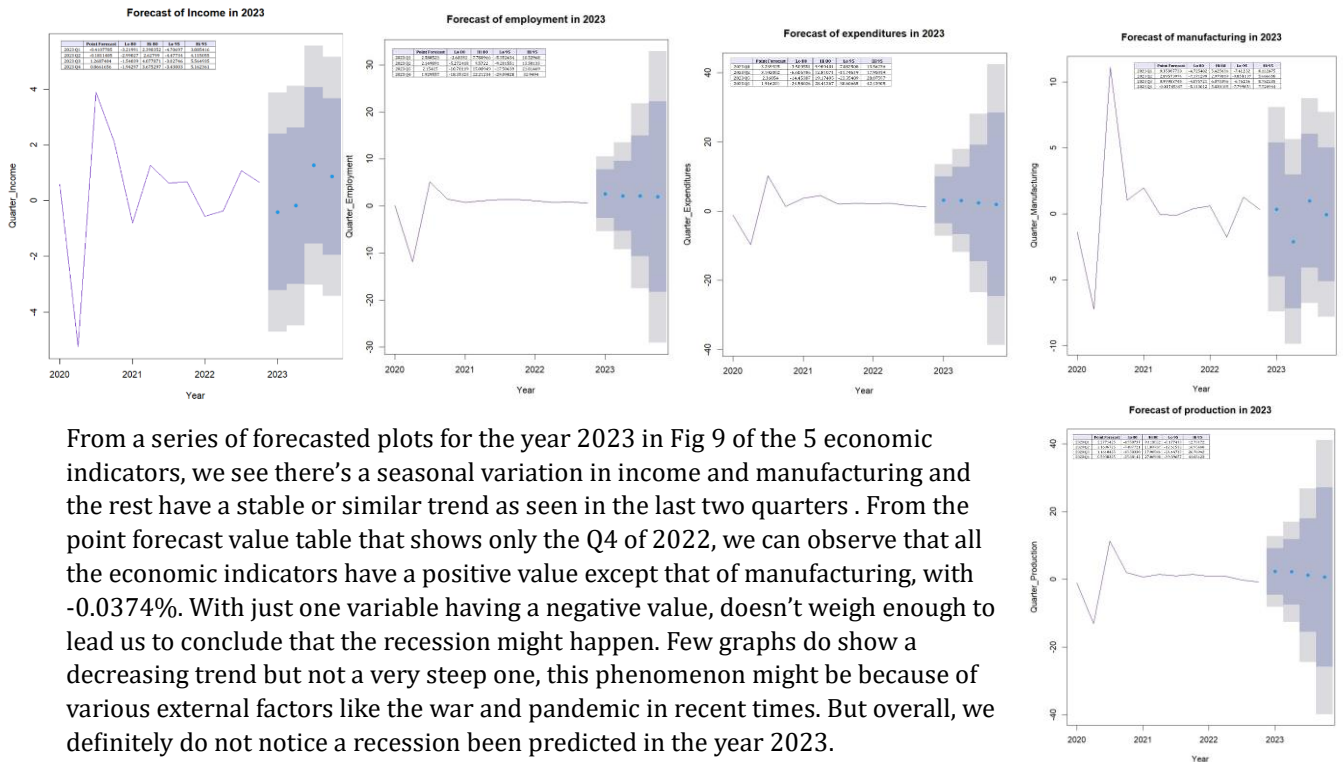
Fig 8: Economic indicators decomposed into seasonal, trend and irregular components
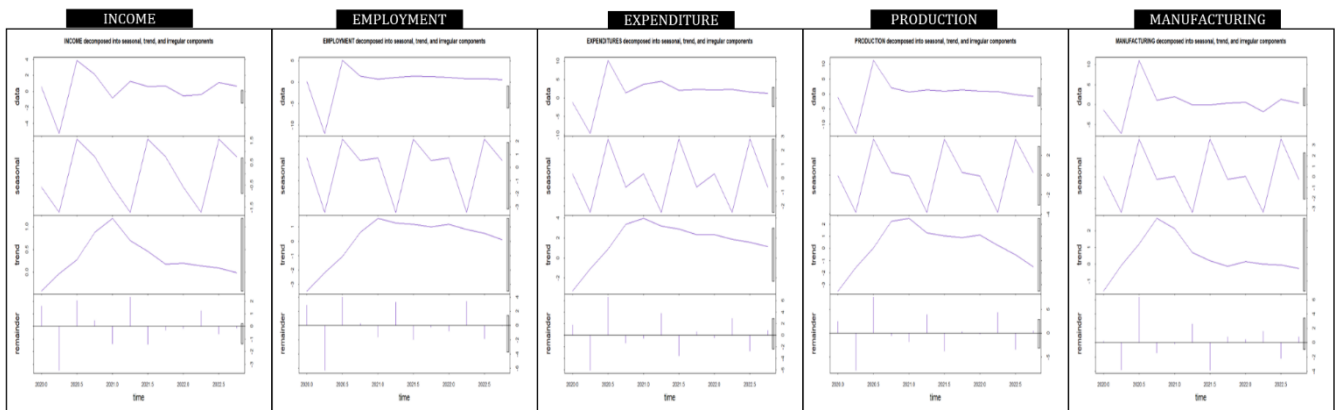
V. PREDICITION OF RECESSION IN FUTURE

The projected future value for all the economic indicators for the United States of America in 2023 is plotted by performing the time series forecasting as seen in the figures below.

Forecast of Income in 2023 | Forecast of employment in 2023 | Forecast of expenditures in 2023 | Forecast of manufacturing in 2023


Forecast of production in 2023

From a series of forecasted plots for the year 2023 in Fig 9 of the 5 economic indicators, we see there's a seasonal variation in income and manufacturing and the rest have a stable or similar trend as seen in the last two quarters . From the point forecast value table that shows only the Q4 of 2022, we can observe that all the economic indicators have a positive value except that of manufacturing, with -0.0374%. With just one variable having a negative value, doesn't weigh enough to lead us to conclude that the recession might happen. Few graphs do show a decreasing trend but not a very steep one, this phenomenon might be because of various external factors like the war and pandemic in recent times. But overall, we definitely do not notice a recession been predicted in the year 2023.

Table 5: Point Forecast values for the Economic Indicators in Q4 of 2022

Economic Indicator	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Income	0.8661656	-1.942965	3.675297	-3.430029	5.162361
employment	1.929557	-18.353227	22.21234	-29.090283	32.9494
expenditure	1.929557	-18.353227	22.21234	-29.090283	32.9494
production	0.5938325	-25.881416	27.069081	-39.896566	41.08423
manufacturing	-0.03745347	-5.113012	5.038105	-7.799851	41.08423

VI. CONCLUSION

The 5 different economic indicators, income, employment, expenditure, production and manufacturing are a great starter to predict recession, but we could also use in additional data for further in-depth analysis, like breakdown the data into state-level and examine each city under them, this gives an opportunity identify if there are any regional variations and it would also help in identifying a particular vulnerable area that could be the major cause of a recession. We could also go ahead and look into various new policies that could have intervened negatively. One more way to go about this would be to take into consideration of the consumer sentiment data which particularly aims at about how a consumer feels about the economy and their financial situation at home. A take at housing market data would also be a good addition, as housing is one the major key sector of the economy and it would provide data on total sales, constructions and so on. We could check the sales of houses and their constructions and if there's a slowdown in this process it might directly imply that the economy is weakening. These additional factors would help in providing additional insights into the U.S. economy.

State-level➔regional factors	New government policies	Consumer sentiment Data	Housing Market Data
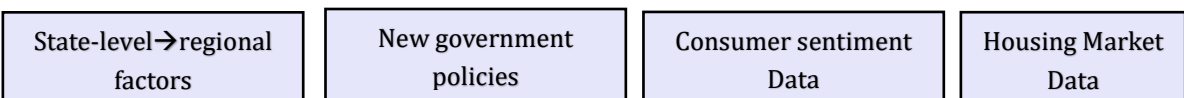
As of now, with our data provided, it's safe to conclude that United States of America isn't getting into a recession in the year 2023!