STATS/CSE 780

Homework Assignment 1

Seyed Mohammad Mehdi Hassani Najafabadi(Student ID: 400489126)

04 February, 2023

Contents

Shiny_link: [https://mehdihassani.shinyapps.io/Assign1_SID_400489126/][https://mehdihassani.shinyapps.io/Assign1_SID_400489126/]

Introduction

The Consumer Price Index (CPI) represents changes in prices as experienced by Canadian consumers. It measures price change by comparing, through time, the cost of a fixed basket of goods and services . The goods and services in the CPI basket are divided into 8 significant components including Food, energy and etc. In this report, the data of price indexes of Canadian customers before(i.e.Jan 2020) and after the pandemic(i.e.Nov 2021) has been extracted from statics Canada [1] analyzed. Additionally, the analysis includes some interpretation and comparison in terms of the fluctuation of price indexes during the pandemic period will be discussed.

Data Transformation and Preprocessing

At the first step, the functional data such table of data has been extracted form other data.

Based on the tidiness rules, all the months name from column names should be summarize in a column with months names. On the other hand and due to the importance of energy and food prices, in this report they are analyzed in a separate columns and for all other items price the *All-items excluding food and energy* will be used for the underlying analysis.Therefore, our tidy table are consisted of four columns of ***Months***, ***food***,***energy***,***All-items excluding food and energy*** which is shown as follows:

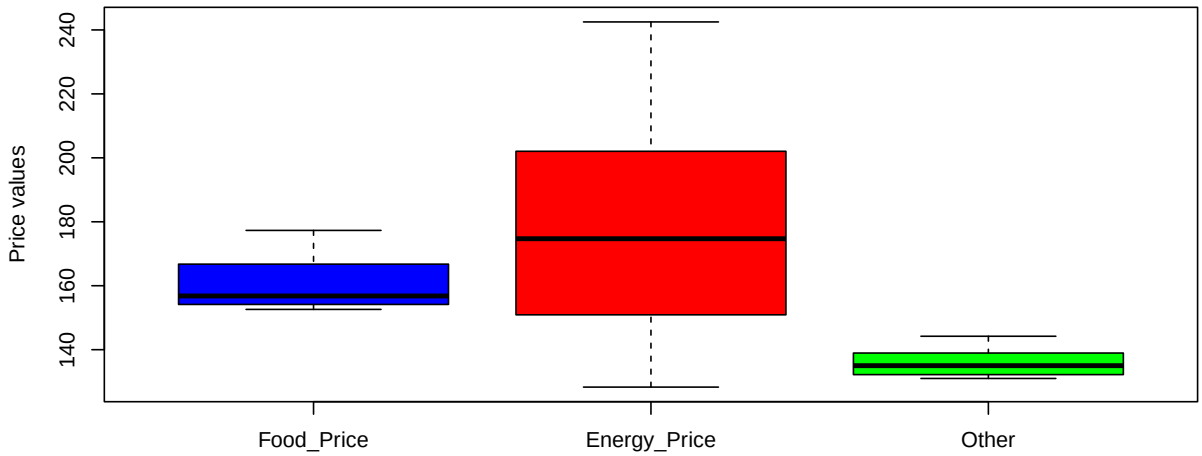
Visualization and Statistical Transfromation

In this section, the preprocessed table is used to create two graphs of how the variables are changed and what are the patterns behind their changes in all across Canada and Ontario province:

Price Index Variation in Ontario

In this section, three retrieved variables has been visualized to reveal general trends as the following figures. Due to limited ranges between the observations indexes (100-200) statistical transformation has not been applied to the data. Regarding outlines detection the below figure show that all the data listed in the three variables are within the boxes, so there are not any outline in the data based.

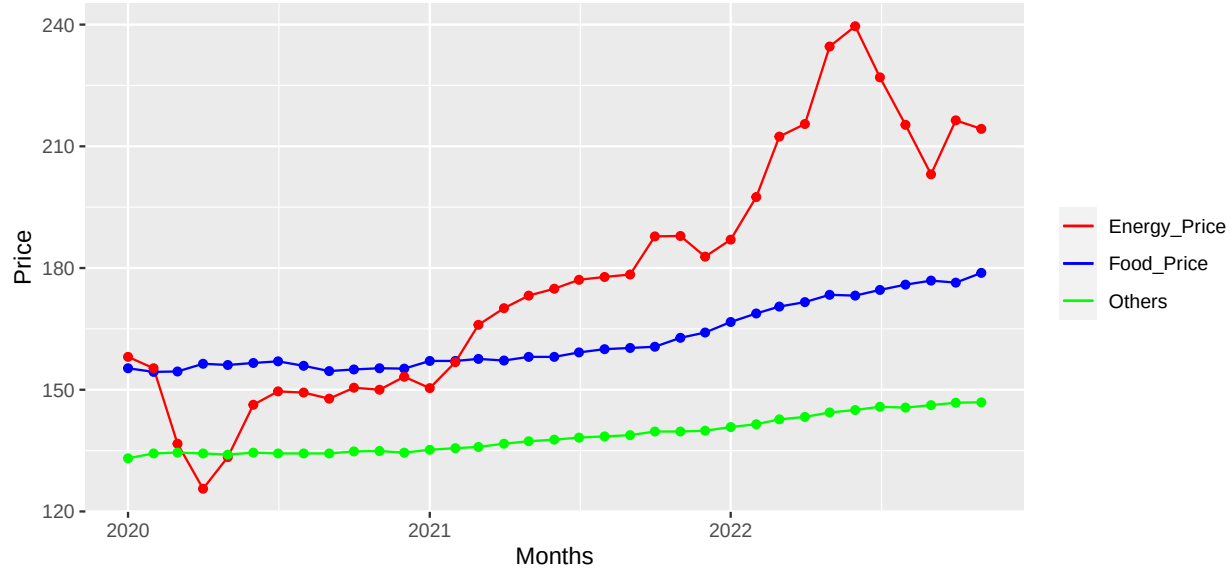
Figure(1) :Box Plot of Prices for Different Variables



Price Index Variation for CANADA

In this section, three retrieved variables has been visualized to reveal general trends for CANADA as the following figures. Due to limited ranges between the observations indexes (100-200) statistical transformation has not been applied to the data.

Figure(2) : Food, Energy and all other Price Trends in Ontario Jan 2020 to Nov 2022



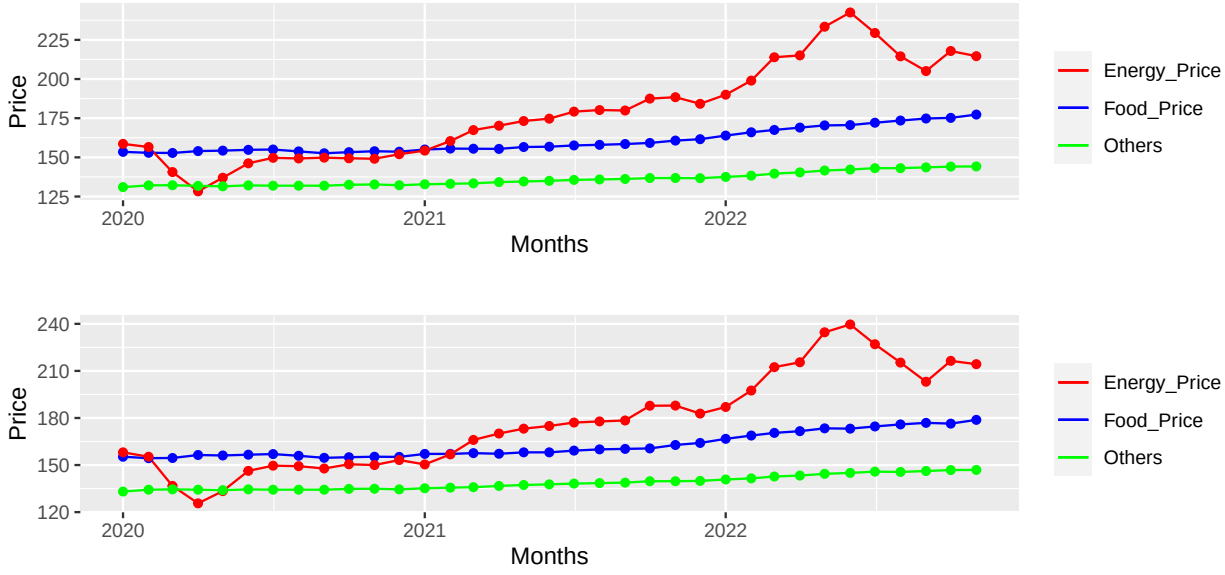
According to figure(),General trends show that COVID-19 pandemic has had the impact on price rising . After 2021, price escalation experienced more growing that prior year. It could be mainly

because of inflation rate started at 2021 [2]. Besides, growing percentages in food price and all other categories excluding Food and energy are steady, however, Energy price undergo two abrupt changes.By the earlier of 2020 [3], the combination of falling demand, rising supply, and diminishing storage space caused such a pronounced crude petroleum price plunge. On the other hand Ukrainian war was the mere factor to the sharp growth at the mid of 2022.

Price Index changes national and provincial comparison(Canada-Ontario Province)

AS it is shown in the figure(), general patterns at both are almost completely identical. It is mightily because of Ontario percentage share in Canada economy.Ontario is the largest economy in Canada, making up around 38% of Canadian GDP [4].

Figure(3) : Food, Energy, and all other Price Trends in Ontario in comparison with Canada



It is worthy to cite again that due to bounded ranges of index values in the values in both Canada and Ontario table, Statistical transformation has not been utilized.

Shiny App Report

Developed ***Shiy app*** is used to show the change of each category in the selected time interval started from the start of Pandemic at January 2020 to the November 2022 as the pandemic end time. By clicking on the link you can externally access to the prepared Shiny App.

References

1. Table 18-10-0004-01 Consumer Price Index, monthly, not seasonally adjusted DOI: https://doi.org/10.25318/1810000401-eng

2. https://fred.stlouisfed.org/graph/?g=XQGK&utm_campaign=myfred_referrer&utm_medium=exported-chart&utm_source=direct

3. https://www.bls.gov/opub/mlr/2020/article/from-the-barrel-to-the-pump.htm

4. https://www.bbc.com/news/business-63855030

STATS/CSE 780

Technical Supplemental 1

Seyed Mohammad Mehdi Hassani Najafabadi(Student ID: 400489126)

04 February, 2023

```r
library(nycflights13)
library(astsa)
library(lubridate)
library(dplyr)
library(magrittr)
library(readr)
library(stringr)
library(ggplot2)
library(reshape2 )
library(gridExtra)
```

```r
knitr::opts_chunk$set(
  comment = '', fig.width = 8, fig.height = 4)
```

```r
data <- read.csv("1810000401_CA.csv")
my_data <- dplyr::slice(data,11:25)
```

```r
col_names <- dplyr::slice(data,9 )
col_names <- gsub("-", "- 20", col_names)
col_names <-gsub("Jan", "1-1", col_names)
col_names <-gsub("Feb", "1-2", col_names)
col_names <-gsub("Mar", "1-3", col_names)
col_names <-gsub("Apr", "1-4", col_names)
col_names <-gsub("May", "1-5", col_names)
col_names <-gsub("Jun", "1-6", col_names)
col_names <-gsub("Jul", "1-7", col_names)
col_names <-gsub("Aug", "1-8", col_names)
col_names <-gsub("Sep", "1-9", col_names)
col_names <-gsub("Oct", "1-10", col_names)
col_names <-gsub("Nov", "1-11", col_names)
col_names <- gsub("Dec","1-12", col_names)
```

```
colnames(my_data) <- col_names
va <- my_data$`Products and product groups 3 4`
cat_name <- gsub(".*[0-9]\\. ", "", va)
cat_name <- lapply(stringr::str_split(va,"[0-9]"),
       unlist(function(x){x[1]}))
cat_name <-make.names(cat_name)
my_data$`Products and product groups 3 4`<- cat_name
```

```
Food_Price_1 <-
  melt(dplyr::filter(my_data,my_data$`Products and product groups 3 4`=="Food."))
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

```
Energy_Price_1 <-
  melt(dplyr::filter(my_data,my_data$`Products and product groups 3 4`=="Energy."))
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

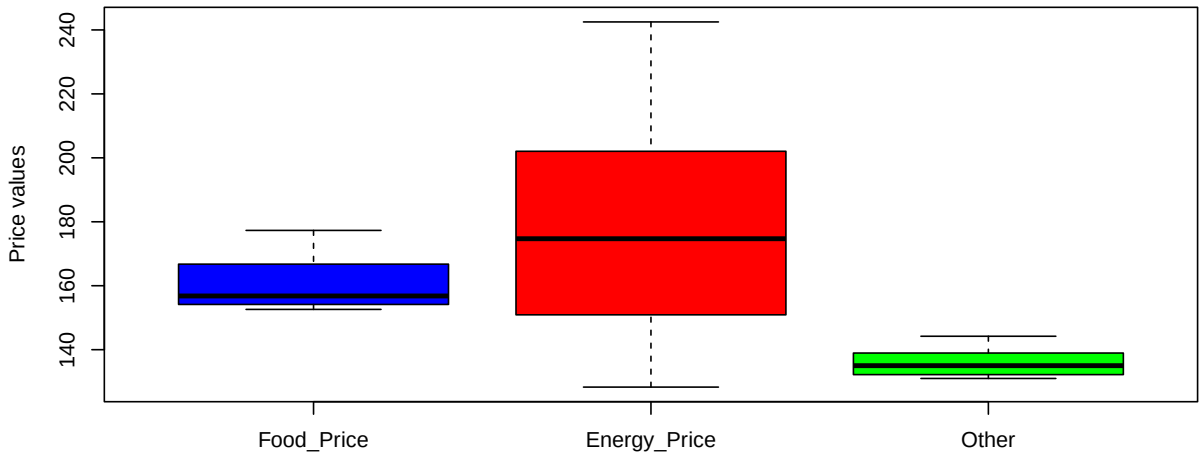
```
Others_1 <- melt(dplyr::filter(my_data,my_data$`Products and product groups 3 4`=="All.items.ex
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

```
mytable <-data.frame("Months" = col_names[2:length(col_names)]%>% as.Date(format = "%d-%m-%Y")
```

```
par(cex = 0.8)
boxplot(mytable[, -1], main = "Figure(1) :Box Plot of Prices for Different Variables",
        xlab = "", ylab = "Price values", col = c("blue", "red", "green"),
        names = c("Food_Price", "Energy_Price", "Other")
        )
```

Figure(1) :Box Plot of Prices for Different Variables




```
p1 <- ggplot(data = mytable) +
  geom_line(
    aes(x = Months, y = Food_Price, color = "Food_Price")
  ) +
  geom_line(
    aes(x = Months, y = Energy_Price, color = "Energy_Price")
  ) +
  geom_line(
    aes(x = Months, y = Others, color = "Others")
  ) +
  geom_point(
    aes(x = Months, y = Food_Price),
    col = "blue1"
  ) +
  geom_point(
    aes(x = Months, y = Energy_Price),
    col = "red1"
  ) +
  geom_point(
```

```
    aes(x = Months, y = Others),
    col = "green1"
  ) +
  labs(x = "Months", y = "Price", color = "Variable") +
labs(x = "Months", y = "Price", color = "Variable",subtitle="Figure(3) : Food, Energy, and all
  theme(legend.title = element_blank())
```

```
data_ON <- read.csv("1810000401_ON.csv")
# Print the name of the file
my_data_ON <- dplyr::slice(data_ON,11:25)
```

```
col_names_ON <- dplyr::slice(data_ON,9 )
col_names_ON <- gsub("-", "- 20", col_names_ON)
col_names_ON <-gsub("Jan", "1-1", col_names_ON)
col_names_ON <-gsub("Feb", "1-2", col_names_ON)
col_names_ON <-gsub("Mar", "1-3", col_names_ON)
col_names_ON <-gsub("Apr", "1-4", col_names_ON)
col_names_ON <-gsub("May", "1-5", col_names_ON)
col_names_ON <-gsub("Jun", "1-6", col_names_ON)
col_names_ON <-gsub("Jul", "1-7", col_names_ON)
col_names_ON <-gsub("Aug", "1-8", col_names_ON)
col_names_ON <-gsub("Sep", "1-9", col_names_ON)
col_names_ON <-gsub("Oct", "1-10", col_names_ON)
col_names_ON <-gsub("Nov", "1-11", col_names_ON)
col_names_ON <- gsub("Dec","1-12", col_names_ON)

colnames(my_data_ON) <- col_names_ON
va_ON <- my_data_ON$`Products and product groups 3 4`
cat_name_ON <- gsub(".*[0-9]\\. ", "", va_ON)
cat_name_ON <- lapply(stringr::str_split(va_ON,"[0-9]"),
       unlist(function(x){x[1]}))
cat_name_ON <-make.names(cat_name_ON)
```

```
my_data_ON$`Products and product groups 3 4`<- cat_name_ON
```

```
Food_Price_1 <- melt(dplyr::filter(my_data_ON,my_data_ON$`Products and product groups 3 4`=="Fo
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

```
Energy_Price_1 <- melt(dplyr::filter(my_data_ON,my_data_ON$`Products and product groups 3 4`=="
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

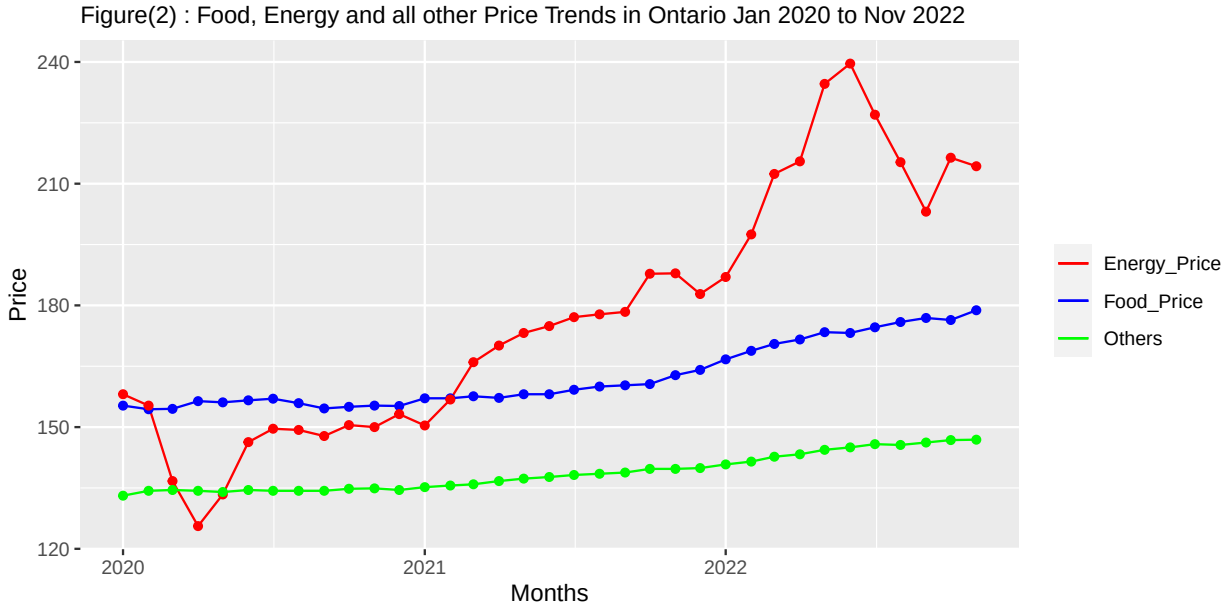
```
Others_1 <- melt(dplyr::filter(my_data_ON,
                               my_data_ON$`Products and product groups 3 4`=="All.items.excludi
```

```
Using Products and product groups 3 4, 1-1- 2020, 1-2- 2020, 1-3- 2020, 1-4- 2020, 1-5- 2020,
```

```
Mytable_ON <-data.frame("Months" = col_names_ON[2:length(col_names_ON)]%>% as.Date(format = "%d
```

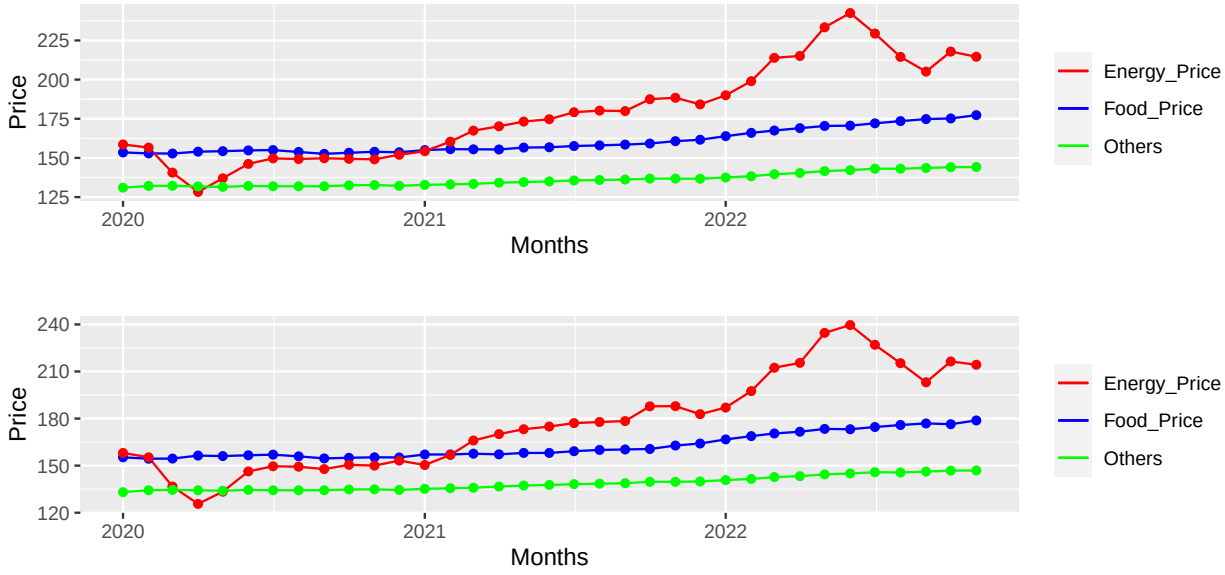
```
p2 <- ggplot(data = Mytable_ON) +
  geom_line(
    aes(x = Months, y = Food_Price, color = "Food_Price")
  ) +
  geom_line(
    aes(x = Months, y = Energy_Price, color = "Energy_Price")
  ) +
  geom_line(
    aes(x = Months, y = Others, color = "Others")
  ) +
  geom_point(
    aes(x = Months, y = Food_Price),
    col = "blue1"
  ) +
  geom_point(
```

```
    aes(x = Months, y = Energy_Price),
    col = "red1"
  ) +
  geom_point(
    aes(x = Months, y = Others),
    col = "green1"
  ) +
  labs(x = "Months", y = "Price", color = "Variable",subtitle="Figure(2) : Food, Energy and al
  scale_color_manual(values = c("Food_Price" = "blue1", "Energy_Price" = "red1", "Others" = "g
  theme(legend.title = element_blank())
p2
```




```
p2 <- p2 + labs(subtitle=" ")
p_combined <- grid.arrange(p1,p2, nrow = 2)
```

Figure(3) : Food, Energy, and all other Price Trends in Ontario in comparison with Canada




```
library(shiny)
load("Mytable_ON.RData")
save(Mytable_ON, file="Mytable_ON")
ui <- fluidPage(
  selectInput(inputId = "y_var",
              label = "Choose desired category",
              choices = c("Food_Price", "Energy_Price", "Others"),
              selected  = "Food_Price"),
  dateInput(inputId = "start_date", label = "Start date", value = "2020-01-01"),
  dateInput(inputId = "end_date", label = "End date", value = "2022-11-01"),
  plotOutput(outputId = "pair_plot")
)
server <- function(input, output) {
  data_filtered <- reactive({
    start_date <- as.Date(paste0(input$start_date, "-01"), format = "%Y-%m-%d")
    end_date <- as.Date(paste0(input$end_date, "-01"), format = "%Y-%m-%d")
    Mytable_ON %>% filter(as.Date(Months, format = "%Y-%m-%d") >= start_date &
                          as.Date(Months, format = "%Y-%m-%d") <= end_date)
  })
```

```
  output$pair_plot <- renderPlot({
    category <- input$y_var
    color <- ifelse(category == "Food_Price", "blue",
                    ifelse(category == "Energy_Price", "red", "green"))
    ggplot(data_filtered(), aes_string(x = "Months", y = category)) +
      geom_line(color=color, size=1) +
      geom_point(color=color) +
      labs(x = "Months", y = category) +
      scale_x_date(limits = as.Date(c(input$start_date, input$end_date), format = "%Y-%m-%d"),
      theme(panel.grid.minor = element_line(size = 0.1),
            panel.grid.major = element_line(size = 0.1))
  })



}
shinyApp(ui = ui, server = server)
```