The Impact of Investing at Various Points in your Life

Ashley Mayo

2024-03-01

Introduction

What age did you start investing in the stock market? The average person in the United States does not start investing until they are 33.3 years old (Royal, J., 2023). Did you start putting money aside for your retirement or general investment account before then? Why does the average American wait until their thirties to start investing? This research study intends to uncover the impacts of investing later in your life compared to in your twenties. Data science will be used to evaluate the importance of investing early by analyzing previous market data and building predictive models. Would you want to know the impact on your portfolio if you started investing today instead of in five years? Many Americans have not received the proper education to understand the importance of saving for retirement or other goals. The following research questions were developed to explore the affects of delaying investments and shine a light on how crucial investing is.

1.) How does someone quantify the compounding effect of investing early on an portfolio?

2.) Can demographic factors influence someone’s decision?

3.) How does the timing of entering into the market influence generational wealth?

4.) What are the consequences of delaying investing until further in life for retirement savings?

5.) What are the patterns and timing of investors exits from the market during downturns do to the overall outlook of their portfolio?

To address the problem, the first step of the analysis involved gathering the past performance of a benchmark. Benchmarks are indexes or averages that are tracking a particular stock market or market segment. Some of the well known include S&P 500, Russell 2000, and Dow Jones Industrial Average. By obtaining the past performance of one of these benchmarks, it would allow for the comparison of various investment time horizons. This information would provide the basis of the research study. From there, there was an exploration of the impact societal and economic factors have on investing in the stock market. The observation of economic indicators such as gross domestic product (GDP) and the average income in the United States were used to evaluate potential fluctuations of the benchmark and investing. Additionally, I took the data of the past performance of the benchmark and compared it to the percentage of Americans invested directly or indirectly in stocks, which allowed me to evaluate if people were trying to time the market. Analyzing the data sets partially addressed the importance of investing early. There were other factors such as personal debt, income limitations, and family issues that impact someone’s ability to invest early or late in their life.

After the initial data collection and analysis of benchmark performance, the methodology explores broader societal and economic influences on investment behaviors. This includes examining economic indicators like GDP and average income levels in the United States to gauge potential correlations with market fluctuations. Moreover, the investigation extends to assessing the level of market participation among Americans, comparing it with benchmark performance to discern trends in market timing behaviors. While this analysis provides valuable insights into the importance of early investing, it acknowledges that additional factors such as personal financial constraints and life circumstances may impact individuals’ ability to invest at different stages of life. This comprehensive approach lays the groundwork for the research study, blending quantitative analysis with contextual understanding to inform strategies for early investment planning.

Overview of the Analysis

The analysis begins with a focus on the historical performance of the S&P 500, highlighting recessionary periods such as those in 2001, 2007-2009, and during the COVID-19 era. Despite these downturns, the graph illustrates resilience and upward trends, offering valuable insights for investors and policymakers alike. Further exploration into investment outcomes over varying time horizons emphasizes the importance of early investment and the power of compounding returns. Additionally, a detailed examination of income trends, stock ownership rates, and GDP fluctuations provides a nuanced understanding of investment behaviors across different demographic segments. However, predicting future S&P 500 performance presents challenges, as machine learning models often fail to incorporate human behavior and complex economic factors accurately.

Overview of the Implications

The analysis reveals valuable insights for the target audience, emphasizing the benefits of early investment and the importance of understanding market dynamics. By grasping the power of compounding returns and considering demographic factors, consumers can tailor their investment strategies to achieve long-term financial goals. Additionally, recognizing the consequences of delaying investment underscores the urgency of proactive financial planning, empowering individuals to make informed decisions and secure their financial future.

Overview of the Limitations

The research study involved compiling data frames, creating visualizations, and analyzing results, but encountered limitations primarily due to the quality and accessibility of historical financial data, as well as constraints on time and resources. Predicting future investment outcomes proved challenging due to market sentiment and dynamic benchmarks, with historical performance not guaranteeing future results. Additionally, personal limitations in using RStudio affected the accuracy of investment compounding tables, and the data set on American stock ownership lacked data on individuals selling out during recessions, which highlighted behavioral biases impacting investment decisions.

Concluding Remarks

This research study illuminates the critical importance of early investment and informed decision-making in achieving long-term financial security. By analyzing historical market data and exploring demographic influences, we have gained valuable insights into the dynamics of investment behavior and its implications for consumers. The findings underscore the transformative potential of compounding returns and the need for proactive financial planning. While limitations exist in data availability and predictive modeling, this study lays a solid foundation for further research and underscores the ongoing need for education and awareness in financial literacy. Moving forward, empowering individuals with the knowledge and tools to navigate the investment landscape will be essential in fostering resilience and promoting financial well-being for generations to come.

```r
# Required Packages
library(readxl)
library(readr)
library(ggplot2)
library(caret)
library(dplyr)
library(DataEditR)
library(collapse)
library(purrr)
library(tidyr)
library(lubridate)
```

```r
# Original and adjusted data sets

averages <- read_excel("C:/Users/ashle/Downloads/MEHOINUSA646N.xls")
```

```
## New names:
## • `` -> `...2`
```

```r
# The average income in the United states for the past twenty four years was obtained from: http
s://fred.stlouisfed.org/series/MEHOINUSA646N. It was imported into R using the readxl package. T
he original data from possessed the average income from every country. I categorized it to only
include the United States, as my research studying is focusing on people living in the U.S.. The
data was collected on September 12, 2023 and the purpose was to gather the average annual wages
globally. The data ranges from 1985 to 2022. The data frame had linear interpolation to combat t
he standard errors between 1976 and 2011 regarding the median household income.
Dates <- c('1984','1985','1986','1987','1988','1989','1990','1991','1992','1993','1994','199
5','1996','1997','1998','1999','2000','2001','2002','2003','2004','2005','2006','2007','2008','2
009','2010','2011','2012','2013','2014','2015','2016','2017','2018','2019','2020','2021','2022')
Household_Income <- averages %>%
  slice(-1, -2, -3, -4, -5, -6, -7, -8, -9, -10) %>%
  rename(Average_Income = ...2, Year = `FRED Graph Observations`) %>%
  mutate(Year = Dates) %>%
  mutate(Average_Income = as.numeric(Average_Income))
# The average income data set "averages" was modified using the package dplyr. This allowed for
the first 10 rows to be excluded from the data frame because they had various links and irreleva
nt verbiage included. This allowed for the relevant information to remain. When the data was imp
orted into R, there was an issue with the presentation of the observation dates. It showed them
as 011984, instead of 01/1984.The final edit that was having the average_income be considered a
numeric value because I was having issues creating a ggplot as the column was not considered to
be numeric.
head(Household_Income)
```

```
## # A tibble: 6 × 2
##   Year  Average_Income
##   <chr>          <dbl>
## 1 1984           22420
## 2 1985           23620
## 3 1986           24900
## 4 1987           26060
## 5 1988           27230
## 6 1989           28910
```

```
stock_percent <- read.csv("C:/Users/ashle/Downloads/data-r1GiV.csv", header=FALSE)
# This data was obtained through (https://news.gallup.com/poll/266807/percentage-americans-owns-
stock.aspx). It was imported into R using the readr package and function read.csv(). The origina
l purpose of the data was to explore the percentage of people in the United States that have mon
ey invested in the stock market. The data included directly holding stock, mutual funds, exchang
e traded funds, and other securities. The article then proceeded to break down that data by demo
graphic indicators. The data was collected and updated on May 24, 2023.
US_own_stock <- stock_percent %>%
  slice(-1) %>%
  rename(Year = V1, Percentage_Own_Stock = V2) %>%
  mutate(Percentage_Own_Stock = as.numeric(Percentage_Own_Stock)) %>%
  mutate(Percentage_Dont_Own  = 100 - Percentage_Own_Stock) %>%
  arrange(order(Year))
# The percentage of Americans that were invested into the stock market can be seen in the data s
et "stock_percent" and I modified that data to "US_own_stock" using the dplyr package. I removed
the first row in the data frame by slicing it because it repeated the column name, which is redu
ndant. I then renamed the columns to allow the reader to understand the purpose of the column. F
rom there, I rearranged the presentation of the data frame to be ordered from the oldest to the
most recent year. Additionally, I added a column to show the percentage of citizens that don't o
wn stock.
head(US_own_stock)
```

```
##   Year Percentage_Own_Stock Percentage_Dont_Own
## 1 1998                   60                  40
## 2 1999                   59                  41
## 3 2000                   60                  40
## 4 2001                   62                  38
## 5 2002                   62                  38
## 6 2003                   61                  39
```

```
GDP <- read_excel("C:/Users/ashle/Downloads/P_Data_Extract_From_World_Development_Indicators
(2).xlsx")
# The world development indicator data was obtained from: https://databank.worldbank.org/report
s.aspx?source=2&series=NY.GDP.MKTP.CD&country=#. It was imported into R using the readxl package
and the function read_excel().The data was originally gathered in 2023 to provide a comparison o
f 266 countries GDP in US dollars. The original data frame provided the GDP for 266 other countr
ies, I narrowed it down to the United States because that is what I am focusing on.
US_GDP <- data.frame(
  Country_Name = "United States",
  X1990 = 5.96,
  X2000 = 10.25,
  X2001 = 10.58,
  X2002 = 10.92,
  X2003 = 11.45,
  X2004 = 12.21,
  X2005 = 13.03,
  X2006 = 13.81,
  X2007 = 14.47,
  X2008 = 14.71,
  X2009 = 14.41,
  X2010 = 14.96,
  X2011 = 15.51,
  X2012 = 16.15,
  X2013 = 16.69,
  X2014 = 17.39,
  X2015 = 18.12,
  X2016 = 18.71,
  X2017 = 19.48,
  X2018 = 20.56,
  X2019 = 21.43,
  X2020 = 21.22,
  X2021 = 24.90,
  X2022 = 25.12
)
GDP_long <- US_GDP %>%
  pivot_longer(cols = -Country_Name, names_to = "Year", values_to = "GDP") %>%
  mutate(Year = as.numeric(substring(Year, 2)))
# The GDP data frame that was obtained from the world bank was cleaned using dplyr and renamed U
S_GDP. This data frame needed the most cleaning compared to the other three. The years were list
ed as the headers instead of in a row to make it easier to create a ggplot with the data. Thus,
I manually typed the data from the past twenty plus years to create a new data frame. Additional
ly, there were six rows in the data frame that had no information in them, thus, it made the dat
a look clean and was easier to read. The last adjustment I made was to the e+13 related to the G
DP. It can be very confusing when numbers are listed with a e+ at the end so I wanted to clarify
for the user that it is trillions instead of including that figure.
head(GDP_long)
```

```
## # A tibble: 6 × 3
##   Country_Name   Year   GDP
##   <chr>         <dbl> <dbl>
## 1 United States  1990  5.96
## 2 United States  2000 10.2
## 3 United States  2001 10.6
## 4 United States  2002 10.9
## 5 United States  2003 11.4
## 6 United States  2004 12.2
```

```
SP500 <- read_excel("C:/Users/ashle/Downloads/Data (1).xlsx")
# The WSJ Markets is a commonly known organization owned by Dow Jones. They report news relating
to politics, economics, finance, real estate, and more. The data can be found on: https://www.ws
j.com/market-data/quotes/index/SPX/historical-prices. This last data frame was imported using th
e package: readxl and function read_excel(). The intention of the data set was to report histori
cal performance of a well known benchmark, the S&P 500. The data is collected every day that the
market is open, thus the last time it was updated was February 9th, 2024.
Values <- c(4804.49, 3960.66, 4573.82, 3793.75, 3278.10, 2607.39, 2789.80, 2275.12, 1918.60, 202
8.18, 1822.36, 1480.40, 1300.58, 1282.62, 1123.58, 865.58, 1378.76, 1424.16, 1278.73, 1181.41, 1
132.52, 895.84, 1140.21, 1335.63, 1425.59, 1248.77, 963.36, 766.22, 614.42, 465.25, 472.99, 435.
23, 416.08, 325.49, 339.97)
Performance_SP500 <- SP500 %>%
  slice(-1) %>%
  rename(Year = Date) %>%
  mutate(SP500_Performance = Values) %>%
  select(Year, SP500_Performance) %>%
  arrange(order(Year))
# There were multiple adjustments made to the SP500 data frame using the same package as the oth
ers: dplyr. The first adjustment was removing the first row of data. The reason I did this was i
t had the performance of February 2024. Although this is helpful information, all of the perform
ance was one year apart from each other and this one data point was out of sync. I then renamed 
the "Date" column to year to be consistent across all of the data frames. Additionally, I had to
manually type of the performance for the S&P500 to not include the comma because when I was goin
g through the rate of return estimations it would not recognize the row as a numeric because of 
the comma. The last two adjustments that were made was to only include the Year and SP500 perfor
mance because that was the data I am focusing on, and then I arranged the year from oldest to ne
west. Similar to the other data frames.
head(Performance_SP500)
```

```
## # A tibble: 6 × 2
##   Year                SP500_Performance
##   <dttm>                          <dbl>
## 1 1990-01-01 00:00:00              340.
## 2 1991-01-01 00:00:00              325.
## 3 1992-01-01 00:00:00              416.
## 4 1993-01-01 00:00:00              435.
## 5 1994-01-01 00:00:00              473.
## 6 1995-01-01 00:00:00              465.
```

Information that is not self-evident, slice and dice, and plots and tables

The data was appropriately cleaned to be more comprehensible for the reader to understand. Although, the plots for the data frames have not been created yet, so there could be additional adjustments made to the data. For instance, additional rows may need to be excluded to ensure the data frames are evaluating the same time frame. I am also planning on creating a rate of return calculation for the performance of the S&P500. This will allow me to compare the impact of the investment time horizon of ten, twenty, and thirty years. To uncover the information that is currently not self-evident, the next step is to analyze and display the data into various plots. This will present additional questions, which may lead to including additional data frames into the research study. Currently the data frames have been cleaned using the dplyr package. A couple other ways to evaluate the data is to use the purrr package if I need to decrease the number of years in the data frame to accurately compare two data frames against each other or use the summary function to uncover the mean, median, maximum, and minimum. I believe the summary function would be most helpful for the Performance of S&P500 data set.

To effectively answer the research questions, the four data frames will be used in various plots and tables to explore the impact of investing. The data frame of the historical performance of S&P 500 will be used to demonstrate the growth over the past thirty years of one of the most well known benchmarks in a ggplot line graph. From there, regression analysis will be used to show the rate of return based off a time horizon of ten, twenty, and thirty years. This plot and table will assist in answering the first, third, and fourth research question. This will show the impacts of investing earlier in someones life. Although, this data will not show potential withdrawals and the impact of taxation of the account. If the investments are held within a qualified plan, there would not be concerns of capital appreciation due to the tax benefits, but if the position is held in a non-qualified plan, what implications would capital gains have on the appreciation of the investment? This will be important to evaluate once the table is created.

Another ggplot that will be created is a bar graph to display the average income in the United States. This graph will be compared to a ggplot histogram of the percentage of American citizens that owned stocks directly or indirectly. The purpose of these charts will be to uncover any correlations between salary increases and an increase of investing among United States citizens. These plots will partially answer research question two by observing an economical indicator's impact on investing. After evaluating this data, additional economic factors could be evaluated. For instance, demographic factors such as race and age. Older individuals that are in retirement typically shift to a conservative portfolio to ensure limited fluctuation of their accounts. This is due to the potential need of using the funds to supplement their income in retirement. Thus, if there is a larger population of retired individuals than newer generations, this could be weighing the statistic of how many people are invested in the stock market. An observation to make in the future is the percentage of retired citizens in the United States. How would this impact the results of the percentage of those invested in the stock market?

The final plot will be another ggplot line graph that displays the current GDP in the United States. The previous statement goes over how the median household income may impact the percentage of Americans that own stock. Gross domestic product (GDP) is considered to be another economic indicator. This plot will help answer question two and five after the plots have been appropriately evaluated. This will help demonstrate the outlook Americans have on the United States economy and how it impacts their investment decisions. It will be important to observe any decreases in GDP and stock investment percentages. These factors may correlate with each other because if GDP decreases, consumer outlook on the stock market may decline leading to a decrease in the percentage invested.

Questions for Future Steps

1. Are there any disparities when trying to access to investment opportunities based on income levels in the United States?
2. Are there any correlations between income levels, stock ownership percentage, and GDP? If so, how do changes in one variable affect the others?
3. How do income levels and stock ownership vary across different demographic groups?
4. Are there additional data sources that could provide meaningful insights into societal impacts of delaying investments.

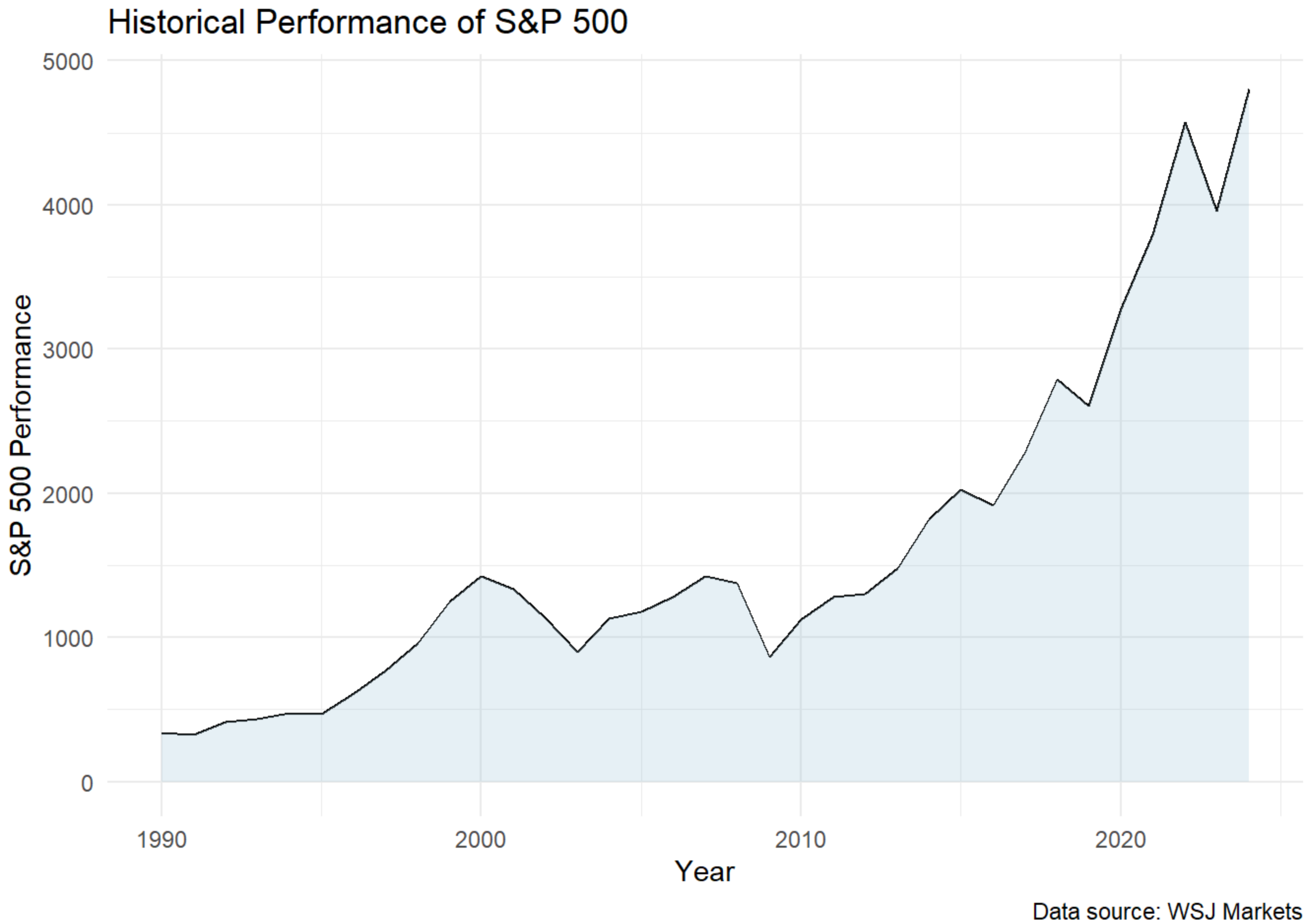
```
# Descriptive Statistics
summary(Performance_SP500)
```

```
##       Year                           SP500_Performance
##  Min.   :1990-01-01 00:00:00.00   Min.   : 325.5  
##  1st Qu.:1998-07-02 12:00:00.00   1st Qu.: 880.7  
##  Median :2007-01-01 00:00:00.00   Median :1282.6  
##  Mean   :2007-01-01 02:44:34.28   Mean   :1632.7  
##  3rd Qu.:2015-07-02 12:00:00.00   3rd Qu.:1973.4  
##  Max.   :2024-01-01 00:00:00.00   Max.   :4804.5
```

```
summary(GDP_long)
```

```
##  Country_Name            Year           GDP       
##  Length:24          Min.   :1990   Min.   : 5.96  
##  Class :character   1st Qu.:2005   1st Qu.:12.82  
##  Mode  :character   Median :2010   Median :15.23  
##                     Mean   :2010   Mean   :15.92  
##                     3rd Qu.:2016   3rd Qu.:18.90  
##                     Max.   :2022   Max.   :25.12
```

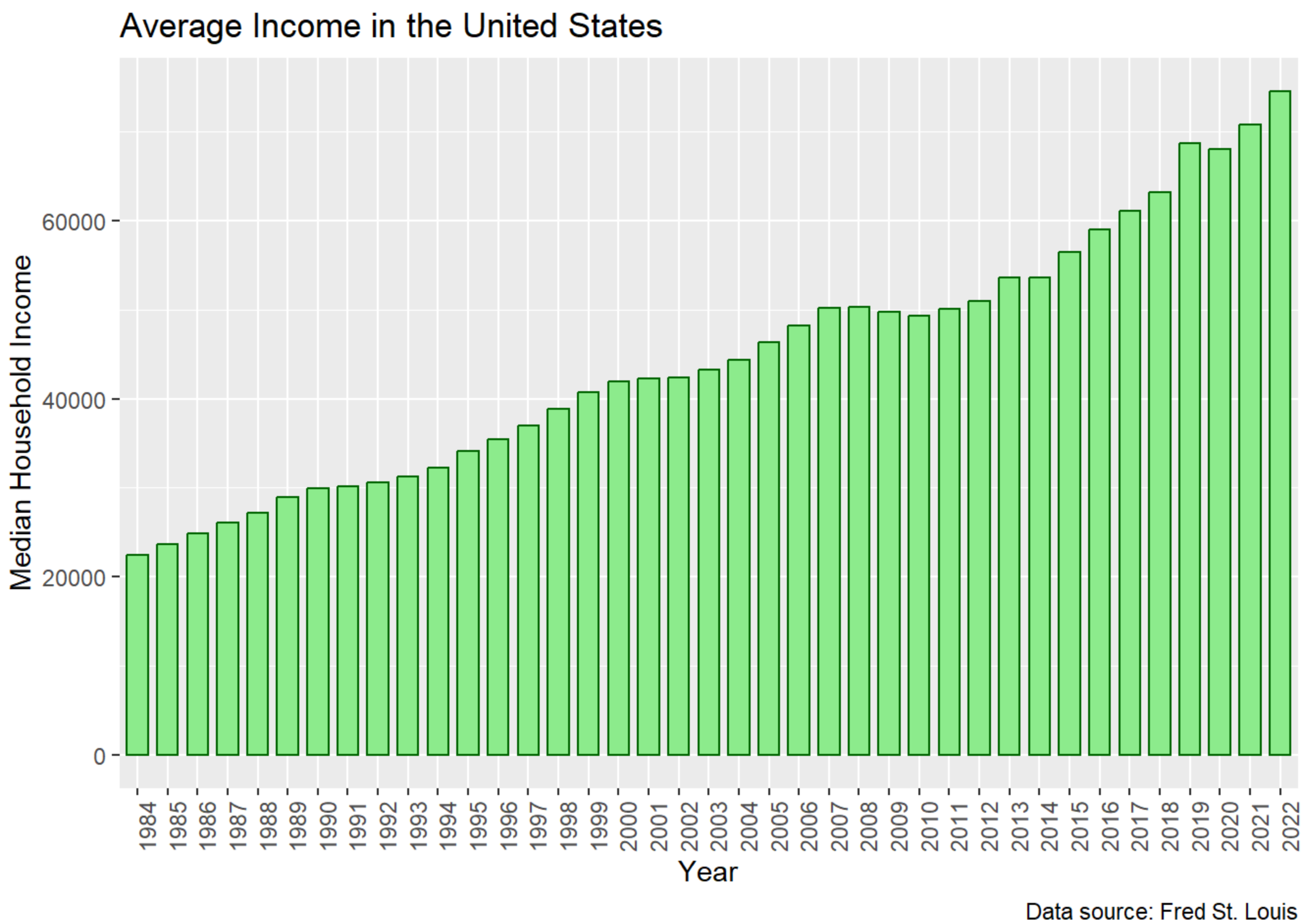
```
# Plots and Tables

## Past performance of S&P500 chart
ggplot(Performance_SP500, aes(x = Year, y = SP500_Performance)) +
  geom_line(color = "black") +
  geom_ribbon(aes(ymin = 0, ymax = SP500_Performance), fill = "lightblue", alpha = 0.3) +
  labs(title = 'Historical Performance of S&P 500', y = 'S&P 500 Performance', caption = "Data s
ource: WSJ Markets") +
  theme_minimal()
```




```
## Hypothetical growth of $1200 after 10, 20 and 30 years
initial_investment <- 1200

growth_10_years <- initial_investment * (Performance_SP500$SP500_Performance[nrow(Performance_SP
500)] / Performance_SP500$SP500_Performance[nrow(Performance_SP500) - 9])
growth_20_years <- initial_investment * (Performance_SP500$SP500_Performance[nrow(Performance_SP
500)] / Performance_SP500$SP500_Performance[nrow(Performance_SP500) - 19])
growth_30_years <- initial_investment * (Performance_SP500$SP500_Performance[nrow(Performance_SP
500)] / Performance_SP500$SP500_Performance[nrow(Performance_SP500) - 29])


growth_table <- data.frame(
  Time_Horizon = c("10 years", "20 years", "30 years"),
  Growth = c(growth_10_years, growth_20_years, growth_30_years))

print(growth_table)
```

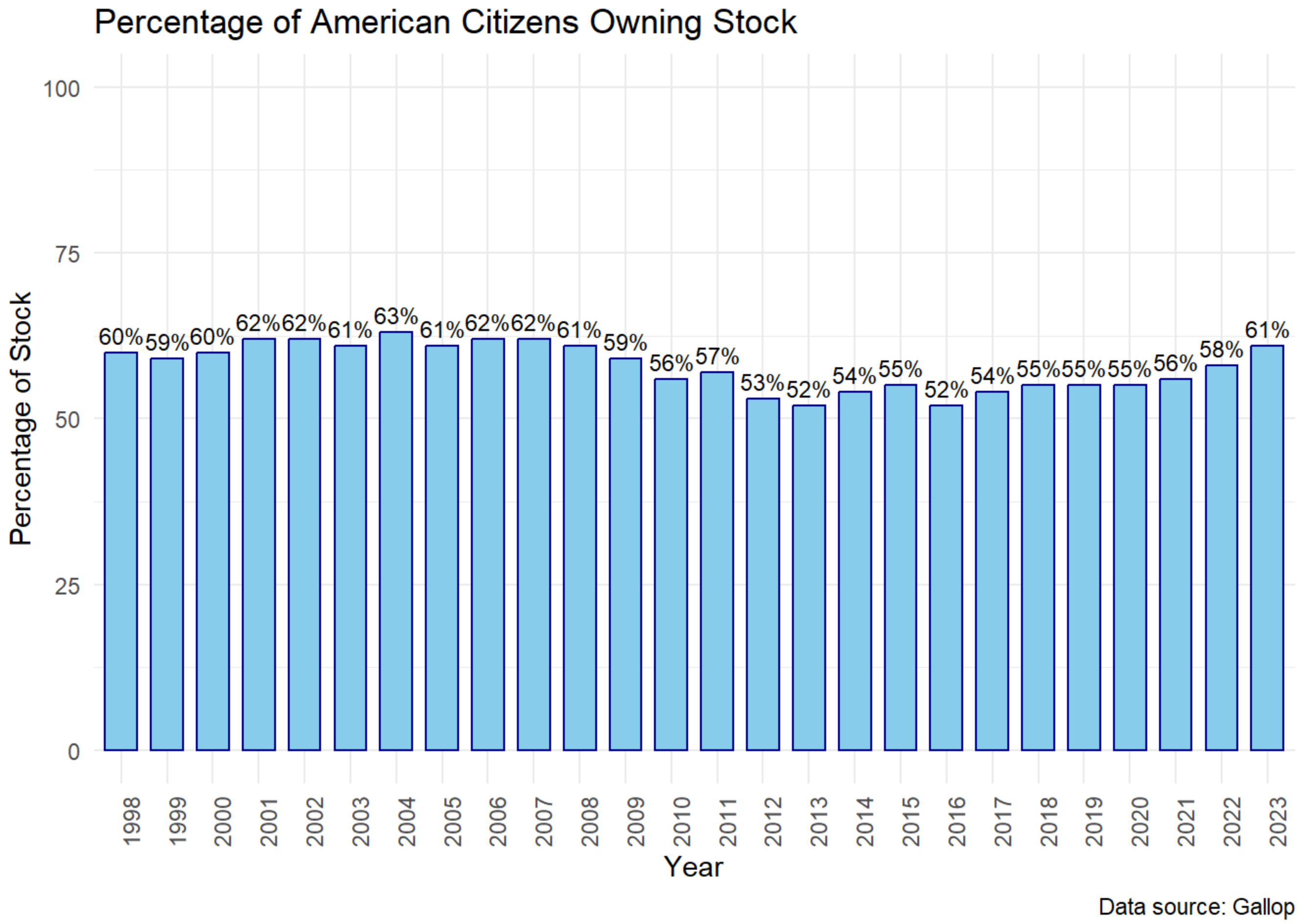
```
##   Time_Horizon    Growth
## 1     10 years  2842.641
## 2     20 years  4880.091
## 3     30 years 12392.021
```

```
## Average household income

ggplot(Household_Income, aes(y = Average_Income, x = Year)) +
  geom_col(fill = 'lightgreen', color = "darkgreen",width = 0.7) +
  labs(y = 'Median Household Income', x = "Year", title = "Average Income in the United States",
caption = "Data source: Fred St. Louis") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

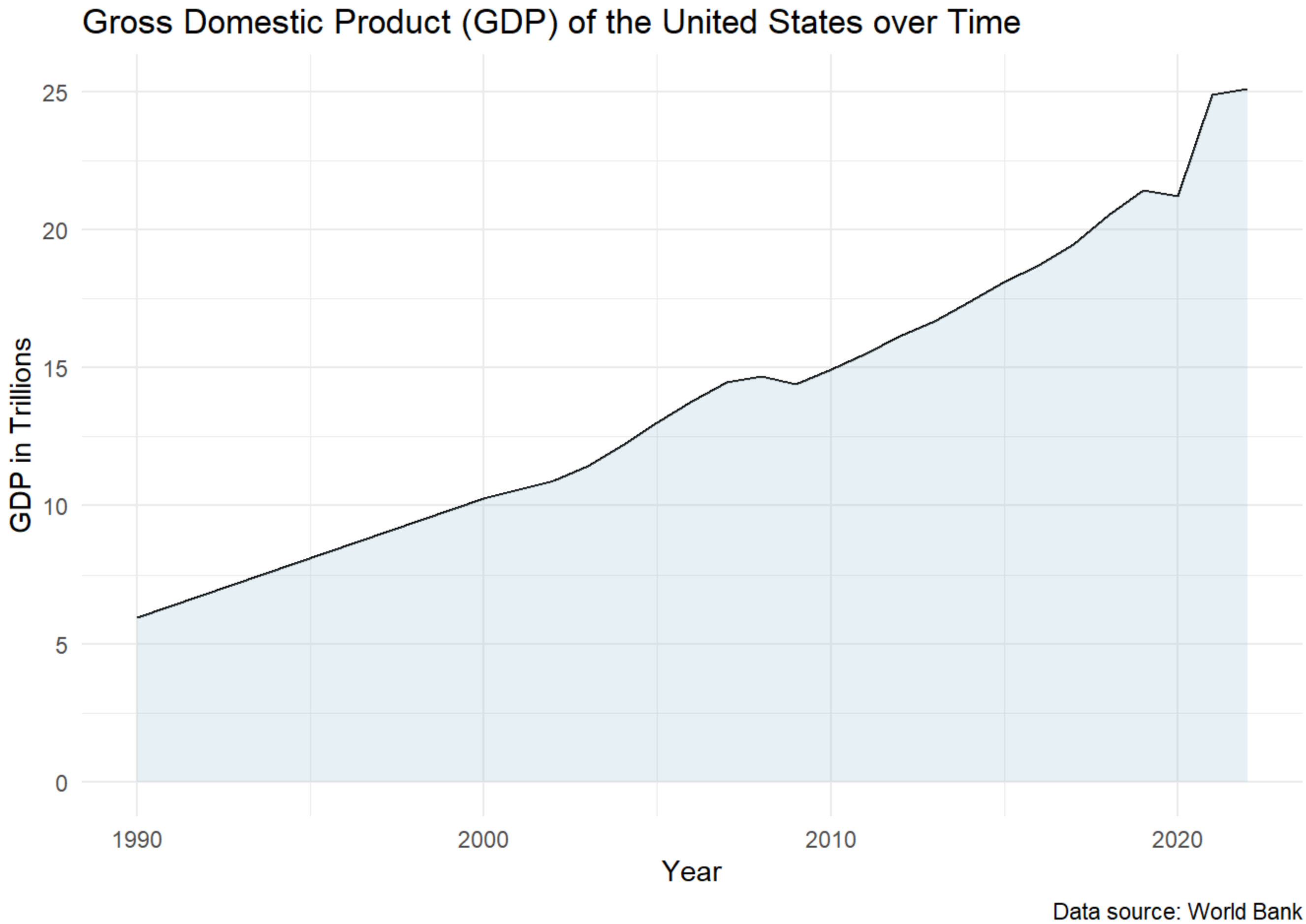



```
## Percentage of stocks owned by U.S. citizens

ggplot(US_own_stock, aes(y = Percentage_Own_Stock, x = Year)) +
  geom_bar(stat = "identity", fill = "skyblue", color = "darkblue", width = 0.7) +
  geom_text(aes(label = paste0(Percentage_Own_Stock, "%")), vjust = -0.5, color = "black", size
= 3) +
  labs(title = "Percentage of American Citizens Owning Stock", y = "Percentage of Stock", x = "Y
ear", caption = "Data source: Gallop") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  ylim(0, 100)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



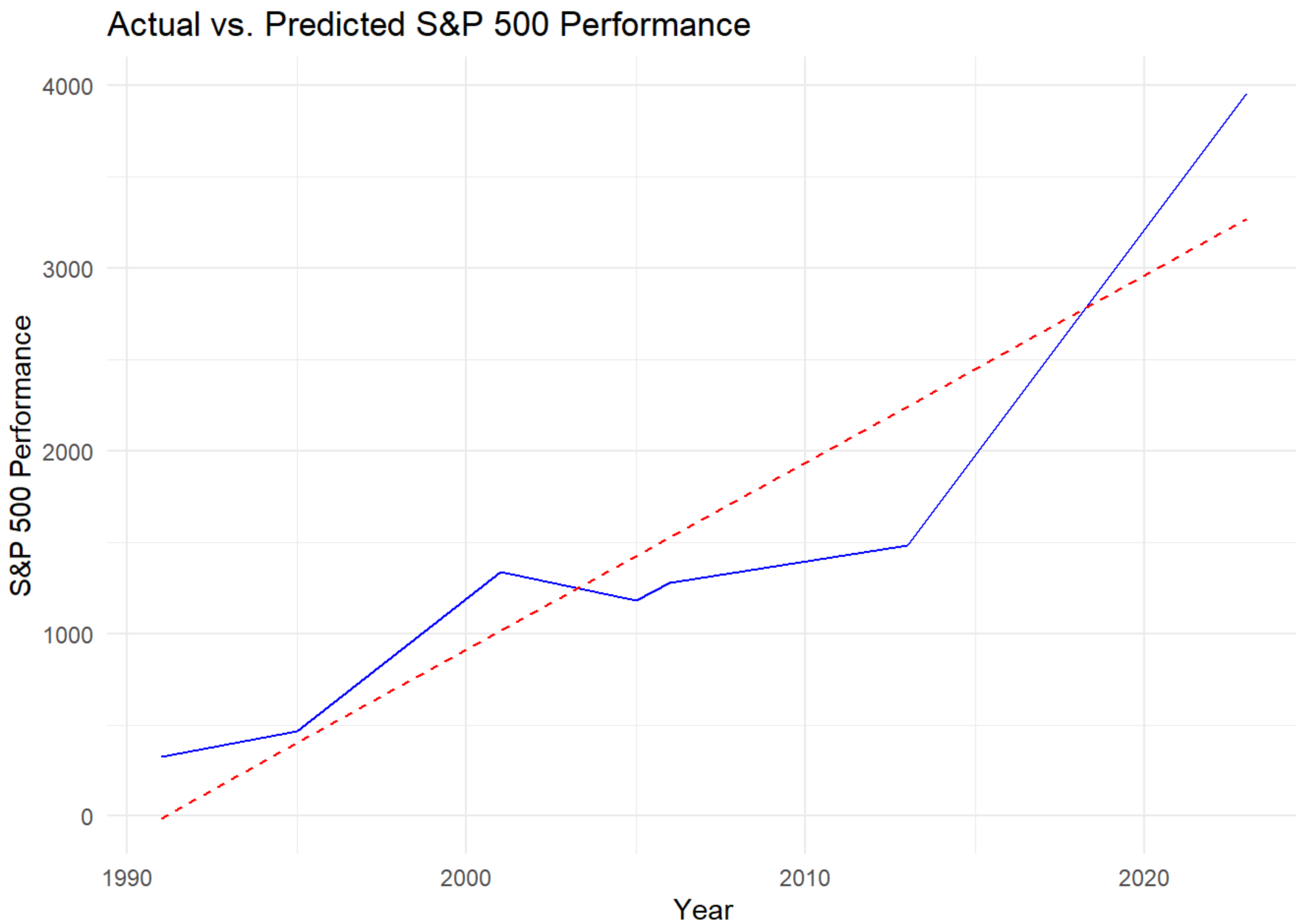

```
## Gross Domestic Product in the United States

ggplot(GDP_long, aes(x = Year, y = GDP)) +
  geom_line(color = "black") +
  geom_ribbon(aes(ymin = 0, ymax = GDP), fill = "lightblue", alpha = 0.3) +
  labs(title = "Gross Domestic Product (GDP) of the United States over Time", x = "Year", y = "G
DP in Trillions", caption = "Data source: World Bank") +
  theme_minimal()
```




```
# Using predictive modeling techniques to predict stock market returns

Performance_SP500$Year <- as.Date(Performance_SP500$Year)

# Split data into training and testing sets
set.seed(123)
train_index <- sample(1:nrow(Performance_SP500), 0.8 * nrow(Performance_SP500))  # 80% for train
ing
train_data <- Performance_SP500[train_index, ]
test_data <- Performance_SP500[-train_index, ]

model <- lm(SP500_Performance ~ Year, data = train_data)

predictions <- predict(model, newdata = test_data)

# Evaluate model performance
mae <- mean(abs(predictions - test_data$SP500_Performance)) # Calculate Mae
print(paste("Mean Absolute Error:", mae))
```

```
## [1] "Mean Absolute Error: 381.498789205677"
```

```
# Visualize the actual vs. predicted values
ggplot() +
  geom_line(data = test_data, aes(x = Year, y = SP500_Performance), color = "blue", linetype =
"solid") +
  geom_line(data = test_data, aes(x = Year, y = predictions), color = "red", linetype = "dashe
d") +
  labs(title = "Actual vs. Predicted S&P 500 Performance",
       x = "Year", y = "S&P 500 Performance") +
  theme_minimal()
```



Analysis

To initiate the data analysis, a line graph depicting the historical performance of the S&P 500 benchmark was generated. Upon examining the graph, it was imperative to identify periods of recession. The first instance occurred from March to November of 2001, constituting a brief recession lasting eight months. However, the benchmark experienced continued decline until 2004. The housing market crash unfolded between December 2007 and June 2009, characterized by a pronounced downturn evident in the chart. Subsequently, the COVID-19 recession commenced in February 2020 and to the present. Notably, the graph illustrates a significant upsurge in market performance, a departure from typical recessionary trends. While a dip in performance is observed from 2021 to 2022, there appears to be a subsequent upward trajectory.

While the graph shows moments of turbulence and recovery, it also underscores the resilience of financial markets and their capacity to adapt and recover in the face of adversity. By synthesizing historical data with broader economic narratives, this analysis offers valuable insights for investors and policymakers navigating an everevolving financial landscape.

Moreover, beyond understanding the dynamics of business cycles, it’s equally crucial to consider the implications for the individual investor. One such consideration is the impact of time horizons on investment outcomes. For instance, exploring how investing $1000 over time frames, such as 10, 20, and 30 years, would influence overall returns. By utilizing historical market trends of the S&P 500, we can discern how early investing can profoundly impact wealth accumulation.

Delving into these scenarios not only shines a light on the power of compounding but also emphasizes the importance of strategic long-term planning in wealth management. As seen in the time horizon table, it becomes apparent the 1,200 dollars initial investment yielded the most favorable return after thirty years, highlighting the significance of initiating investment endeavors earlier in life. For instance, an individual who held the same investment for ten less years found themselves with $7,512 less in returns, starkly illustrating the ramifications of delaying investment decisions.

These observations extend beyond individual financial strategies, they represent the ability to obtain and develop generational wealth. By addressing questions such as the consequences of delaying investment for retirement savings and how the timing of market entry shapes generational prosperity, we can evaluate the dynamics of personal and generational financial security.

To delve deeper into the economic ramifications of investment patterns, an analysis was conducted utilizing several graphical representations. Initially, a bar graph depicting the average income trends in the United States was constructed. Notably, there has been a consistent upward trajectory in median household income, albeit with stagnant growth following the early 2000s recession and a notable downturn during the 2008 financial crisis. This trend was juxtaposed against the percentage of Americans with direct or indirect stock ownership, revealing a decline during the 2008 recession, followed by a gradual recovery to 61% ownership since 2007. Intriguingly, this ownership trend contrasts with the sustained increase in median salary over the years.

Furthermore, an additional analysis involved cross-referencing these income dynamics with the trajectory of the United States’ Gross Domestic Product (GDP) over the same period. The corresponding line graph displayed a consistent positive linear progression, punctuated only by minor fluctuations around the 2008–2009 financial crisis and the onset of the COVID-19 pandemic.

These visualizations aim to address the second research question, which delves into the various demographic factors influencing investment decisions. By juxtaposing income trends, stock ownership rates, and GDP fluctuations, we gain a nuanced understanding of how economic dynamics intersect with individual investment behaviors. This holistic perspective enables us to explore the multifaceted influences shaping investment decisions across diverse demographic segments, thereby contributing to a more comprehensive understanding of economic participation and wealth distribution within society.

However, predicting the future performance of the S&P 500 benchmark faces challenges. Machine learning models often fail to incorporate behavioral reactions to market fluctuations or consider factors like inflation/deflation, limiting their predictive accuracy. Despite their sophistication, these models may overlook the complexities of human behavior and economic dynamics, thus necessitating caution when interpreting their forecasts for future market performance.

Implications of the Study

Understanding the implications of the analysis to the target audience, is pivotal for translating research findings into actionable insights that can empower individuals to make informed financial decisions. The analysis offers several key implications that consumers can leverage to enhance their financial well-being and investment strategies. Firstly, by exploring the historical performance of benchmark indexes like the S&P 500, consumers gain valuable insights into the long-term growth potential of the stock market. Highlighting periods of economic downturns and subsequent recoveries underscores the importance of adopting a resilient investment approach that accounts for market fluctuations. Consumers can use this knowledge to temper short-term market volatility and focus on long-term investment objectives, such as retirement planning or wealth accumulation.

The analysis shows the power of compounding returns and the significance of investing early. By quantifying the compounding effect of early investment on portfolio growth, consumers can grasp the tangible benefits of starting to invest at a younger age. Armed with this understanding, individuals may feel motivated to prioritize savings and investment efforts, recognizing the potential to harness time and maximize returns over the course of their investment journey. Exploring the influence of demographic factors on investment decisions sheds light on the diverse financial needs and preferences of different consumer segments. Recognizing how factors such as age, income level, and life stage impact investment behaviors empowers consumers to tailor their investment strategies accordingly. Whether it involves adjusting risk tolerance, diversifying investment portfolios, or exploring alternative investment vehicles, consumers can make informed decisions that align with their unique financial circumstances and goals.

Additionally, the analysis underscores the consequences of delaying investment for retirement savings, emphasizing the importance of proactive financial planning. By evaluating the long-term implications of deferring investment decisions, consumers can appreciate the potential opportunity costs associated with procrastination. This awareness may prompt individuals to take proactive steps towards building a robust retirement nest egg, such as increasing savings contributions or seeking professional financial guidance.

Limitations of the Study

After compiling the various data frames, creating appropriate visualizations of the data, and analyzing the results based off the problem at hand, there were limitations of this research study. The primary challenge was the quality and accessibility of the historical financial data at my disposal. Given that this project is managed by a single individual, there are constraints on both time and resources for thorough data exploration. Furthermore, certain desired data frames were only available on pay-to-download websites, which didn’t align with my preferred approach.

Data science can be effectively used to provide insights from historical trends of benchmarks and display what the value of that investment would be today. Although, predicting future investment outcomes is extremely challenging due to market sentiment and the dynamic nature of those benchmarks. The table that displayed the difference of investing over ten, twenty, and thirty years used past performance. Although, it was an accurate representation, it does not predict future performance. Additionally, it does not take into consideration human emotion and pulling out of the market during the late business cycle or recessionary period when the market is performing poorly.

One limitation arose concerning the table demonstrating the compounding of a $100 investment would over ten, twenty, and thirty years. This limitation stemmed from my limited proficiency in using RStudio. I aimed to initialize the investment at 1,000 dollars with an annual increment of 100 dollars thereafter. However, despite numerous attempts, I encountered difficulty in incorporating the annual additional investment. Normally, investors continue to contribute funds to their accounts to foster growth. Regrettably, this table lacks such contributions, rendering it less accurate than originally desired.

The data set related to the percentage of Americans that owned stock - directly or indirectly - from 1998 to 2023 possessed various limitations. During this time frame, three recessions occurred. They occurred March to November of 2001, 2007 to 2009, and February 2020 to present day. The importance of this, is there were people pulling out of the market during this time because of the uncertainty, but there was only a clear decrease during the great recession from 2007-2009. A data frame that showed the percentage of people that sold out of the market would have been beneficial to compare to those invested. This would portray if citizens were buying in low and at the same time other Americans may have been selling out of stock investments.This is know as behavioral biases. Human emotions and cognitive biases significantly impact investment decisions, which would create deviations from predicted outcomes.

Conclusion

This research study sheds light on the critical significance of early investment and well-informed decision-making as fundamental pillars for attaining long-term financial security. Through an in-depth analysis of historical market data and a comprehensive exploration of demographic influences, valuable insights have been gleaned regarding the intricacies of investment behavior and its far-reaching implications for consumers. The findings not only highlight the transformative potential of compounding returns but also emphasize the imperative for proactive financial planning to capitalize on such opportunities. Despite inherent limitations stemming from data availability and predictive modeling, this study lays a robust groundwork for further investigation, emphasizing the persistent necessity for enhancing education and awareness in financial literacy.

Looking ahead, it becomes increasingly evident that empowering individuals with the requisite knowledge and tools to navigate the complex investment landscape is paramount for fostering resilience and promoting enduring financial well-being across generations. Recognizing the evolving nature of financial markets and demographic trends, ongoing efforts to equip individuals with the means to make sound financial decisions will be instrumental in safeguarding against economic volatility and ensuring a more secure future. Through continued research, education, and advocacy, we can strive towards a society where individuals are empowered to take control of their financial futures, thus paving the way for greater economic stability and prosperity for all.

Work Cited

Royal, J. (2023, February 7). First Investors Statistics 2023: Bankrate. Bankrate Press. https://www.bankrate.com/investing/statistics-for-beginning-investors/ (https://www.bankrate.com/investing/statistics-for-beginning-investors/)