```
library(tidyverse)
library(haven)
library(dplyr)
library(stargazer)
library(carData)
library(broom)
library(ggplot2)
library(lubridate)
library(quantmod)

comps_final <- read_csv("/Users/ananya/Desktop/490 data.csv")
comp_returns <- read_csv("/Users/ananya/Downloads/comp returns.csv")
factors <- read_csv("/Users/ananya/Downloads/Famafrench factors.csv")


filtered_returns <- comps_final %>%
  left_join(comp_returns, by = "tic")

filtered_returns <- select(filtered_returns, -date.y)

average_returns <- filtered_returns %>%
  group_by(tic, date.x) %>%
  summarise(avg_returns = mean(daily_return, na.rm = TRUE), .groups =
'drop')

filtered_returns <- filtered_returns %>%
  select(-start_date, -end_date, -daily_return)

filtered_returns <- filtered_returns %>%
  left_join(average_returns, by = "tic")

filtered_returns <- filtered_returns %>%
  select(-date.x.y)

filtered_returns <- filtered_returns %>%
  distinct(firm_tic, tic, .keep_all = TRUE)

ticker_rf_data <- comps_final %>%
  rowwise() %>%  # Ensure we iterate over each row individually
  mutate(
    rf_dates = list(factors$date[factors$date >= start_date & factors$date
<= end_date]),
    rf_values = list(factors$rf[factors$date >= start_date & factors$date
<= end_date])
  ) %>%
  ungroup() %>%
  select(tic, start_date, end_date, rf_dates, rf_values) %>%
  tidyr::unnest(c(rf_dates, rf_values))

mean_rf_per_ticker <- ticker_rf_data %>%
  group_by(tic) %>%
  summarise(
    mean_rf = mean(rf_values, na.rm = TRUE),  # Calculate mean, ignoring
NA values
    .groups = 'drop'  # Drop grouping after summarisation
  )

excess_returns <- average_returns %>%
```

```
  left_join(mean_rf_per_ticker, by = "tic") %>%
  mutate(
    excess_return = avg_returns - mean_rf  # Calculate excess return
  )
excess_returns <- excess_returns %>%
  select(tic, excess_return)

factors_mean <- comps_final %>%
  mutate(
    factor_means = map2(start_date, end_date, ~ factors %>%
                          filter(date >= .x & date <= .y) %>%
                          summarise(
                            mean_mktrf = mean(mktrf, na.rm = TRUE),
                            mean_SMB = mean(smb, na.rm = TRUE),
                            mean_HML = mean(hml, na.rm = TRUE)
                          ))
  ) %>%
  unnest(factor_means) %>%
  select(tic, mean_mktrf, mean_SMB, mean_HML)
print(factors_mean)

combined_data <- average_returns %>%
  inner_join(mean_rf_per_ticker, by = "tic") %>%
  inner_join(factors_mean, by = "tic")
combined_data <- combined_data %>%
  distinct(date.x, .keep_all = TRUE)

combined_data <- combined_data %>%
  mutate(
    excess_return = avg_returns - mean_rf,
    market_premium = mean_mktrf - mean_rf,
    adjusted_SMB = mean_SMB,
    adjusted_HML = mean_HML
  )

print(excess_returns)

regression_results <- combined_data %>%
  group_by(tic) %>%
  do({
    model <- lm(excess_return ~ market_premium + mean_SMB + mean_HML, data
= .)
    tidy(model)  # Tidies the output to be more readable
  }) %>%
  ungroup()

write_csv(combined_data, "/Users/ananya/Downloads/comps regression
data.csv")
```