SEEK

Amish Verma

2023-09-01

```
# Load necessary libraries
library(readxl)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(stats)
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.2.3
```

```
## corrplot 0.92 loaded
```

```
# Read the Excel sheets into data frames
jobdf <- readxl::read_excel(path = '../Data/Seek_NZ_data_01-07-2023.xlsx',
                            sheet = 'SEEK Job Ad Index')

appdf <- readxl::read_excel(path = '../Data/Seek_NZ_data_01-07-2023.xlsx',
                            sheet = 'SEEK Applications per Ad Index')


# Converting dates into dates
head(jobdf)
```

```
## # A tibble: 6 x 9
##   DATE                COUNTRY     STATE ADS_SA_INDEX ADS_TREND_INDEX
##   <dttm>              <chr>       <chr>        <dbl>           <dbl>
## 1 2002-01-01 00:00:00 New Zealand Total         23.4            23.0
## 2 2002-02-01 00:00:00 New Zealand Total         22.8            23.3
## 3 2002-03-01 00:00:00 New Zealand Total         23.8            23.8
## 4 2002-04-01 00:00:00 New Zealand Total         21.7            24.4
## 5 2002-05-01 00:00:00 New Zealand Total         24.8            25.2
## 6 2002-06-01 00:00:00 New Zealand Total         28.4            26.2
## # i 4 more variables: ADS_SA_GROWTH_MONTH <dbl>, ADS_SA_GROWTH_PCP <dbl>,
## #   ADS_TREND_GROWTH_MONTH <dbl>, ADS_TREND_GROWTH_PCP <dbl>
```

```
jobdf$DATE <- as.Date(jobdf$DATE)
appdf$DATE <- as.Date(appdf$DATE)
```

```r
# Check for missing values
sum(is.na(jobdf))
```

```
## [1] 0
```

```r
sum(is.na(appdf))
```

```
## [1] 0
```

```r
# Check for duplicates
sum(duplicated(jobdf))
```

```
## [1] 0
```

```r
sum(duplicated(appdf))
```

```
## [1] 0
```

```r
# Check data types
str(jobdf)
```

```
## tibble [1,806 x 9] (S3: tbl_df/tbl/data.frame)
##  $ DATE                   : Date[1:1806], format: "2002-01-01" "2002-02-01" ...
##  $ COUNTRY                : chr [1:1806] "New Zealand" "New Zealand" "New Zealand" "New Zealand" ...
##  $ STATE                  : chr [1:1806] "Total" "Total" "Total" "Total" ...
##  $ ADS_SA_INDEX           : num [1:1806] 23.4 22.8 23.8 21.7 24.8 ...
##  $ ADS_TREND_INDEX        : num [1:1806] 23 23.3 23.8 24.4 25.2 ...
##  $ ADS_SA_GROWTH_MONTH    : num [1:1806] 0 -0.0235 0.0423 -0.0895 0.1458 ...
##  $ ADS_SA_GROWTH_PCP      : num [1:1806] 0 0 0 0 0 0 0 0 0 0 ...
##  $ ADS_TREND_GROWTH_MONTH: num [1:1806] 0 0.0137 0.0213 0.0273 0.0338 ...
##  $ ADS_TREND_GROWTH_PCP   : num [1:1806] 0 0 0 0 0 0 0 0 0 0 ...
```

```r
str(appdf)
```

```
## tibble [1,295 x 9] (S3: tbl_df/tbl/data.frame)
##  $ DATE                   : Date[1:1295], format: "2008-02-01" "2008-03-01" ...
##  $ COUNTRY                : chr [1:1295] "New Zealand" "New Zealand" "New Zealand" "New Zealand" ...
##  $ STATE                  : chr [1:1295] "Total" "Total" "Total" "Total" ...
##  $ CA_SA_INDEX            : num [1:1295] 28.6 44.6 49.4 46.7 49.1 ...
##  $ CA_TREND_INDEX         : num [1:1295] 38.6 40.7 43.3 46.3 49.5 ...
##  $ CA_SA_GROWTH_MONTH     : num [1:1295] 0 0.5581 0.107 -0.0546 0.0507 ...
##  $ CA_SA_GROWTH_PCP       : num [1:1295] 0 0 0 0 0 0 0 0 0 0 ...
##  $ CA_TREND_GROWTH_MONTH: num [1:1295] 0 0.0555 0.0644 0.0692 0.0697 ...
##  $ CA_TREND_GROWTH_PCP    : num [1:1295] 0 0 0 0 0 0 0 0 0 0 ...
```

```r
# Summary statistics
summary(jobdf)
```

```
##       DATE              COUNTRY             STATE            ADS_SA_INDEX
##  Min.   :2002-01-01   Length:1806        Length:1806        Min.   :  7.706
##  1st Qu.:2007-05-01   Class :character   Class :character   1st Qu.: 71.472
##  Median :2012-09-16   Mode  :character   Mode  :character   Median :101.018
##  Mean   :2012-09-15                                         Mean   :110.946
##  3rd Qu.:2018-02-01                                         3rd Qu.:136.855
##  Max.   :2023-06-01                                         Max.   :368.162
##  ADS_TREND_INDEX   ADS_SA_GROWTH_MONTH ADS_SA_GROWTH_PCP ADS_TREND_GROWTH_MONTH
##  Min.   :  8.383   Min.   :-0.657057   Min.   :-0.7810   Min.   :-0.114924
##  1st Qu.: 72.372   1st Qu.:-0.020224   1st Qu.: 0.0000   1st Qu.:-0.001837
##  Median :100.698   Median : 0.008866   Median : 0.1062   Median : 0.009577
##  Mean   :110.762   Mean   : 0.015126   Mean   : 0.1670   Mean   : 0.010223
##  3rd Qu.:137.128   3rd Qu.: 0.042346   3rd Qu.: 0.2673   3rd Qu.: 0.024074
##  Max.   :361.387   Max.   : 1.065244   Max.   : 4.4008   Max.   : 0.133068
##  ADS_TREND_GROWTH_PCP
##  Min.   :-0.5954
##  1st Qu.: 0.0000
##  Median : 0.1117
##  Mean   : 0.1513
##  3rd Qu.: 0.2633
##  Max.   : 1.6089
```

```
summary(appdf)
```

```
##       DATE              COUNTRY             STATE            CA_SA_INDEX
##  Min.   :2008-02-01   Length:1295        Length:1295        Min.   :  0.00
##  1st Qu.:2011-12-01   Class :character   Class :character   1st Qu.: 97.32
##  Median :2015-10-01   Mode  :character   Mode  :character   Median :106.95
##  Mean   :2015-10-01                                         Mean   :117.01
##  3rd Qu.:2019-08-01                                         3rd Qu.:123.39
##  Max.   :2023-06-01                                         Max.   :438.39
##  CA_TREND_INDEX   CA_SA_GROWTH_MONTH  CA_SA_GROWTH_PCP   CA_TREND_GROWTH_MONTH
##  Min.   :  0.00   Min.   :-1.000000   Min.   :-1.00000   Min.   :-1.000000
##  1st Qu.: 97.59   1st Qu.:-0.032277   1st Qu.:-0.06326   1st Qu.:-0.011023
##  Median :106.66   Median : 0.004153   Median : 0.01998   Median : 0.004995
##  Mean   :116.19   Mean   : 0.009392   Mean   : 0.12595   Mean   : 0.004443
##  3rd Qu.:122.78   3rd Qu.: 0.043004   3rd Qu.: 0.16546   3rd Qu.: 0.023062
##  Max.   :363.39   Max.   : 1.023829   Max.   : 3.04266   Max.   : 0.136726
##  CA_TREND_GROWTH_PCP
##  Min.   :-1.00000
##  1st Qu.:-0.06602
##  Median : 0.01893
##  Mean   : 0.11816
##  3rd Qu.: 0.15302
##  Max.   : 2.12448
```

The term "seasoanlly adjusted over here means the data has been modified to eliminate the effect of seasonal patterns, making it easier to observe the fundamental trends over the time

Jobs

, the mean is 0.015 or 1.5% which means that there has been an average increase of 1.5% in job ad volumes from month to month over the period of time

- In the ADS_TREND_GROWTH_MONTH, the mean is 1.0%, indicating that the average month on month growth rate in the trend component of the job ad volumes. The trend component is a smooth version of original data, which only captures long term changes and filter out seasonal fluctuations and random noise

Applicatins per job ad

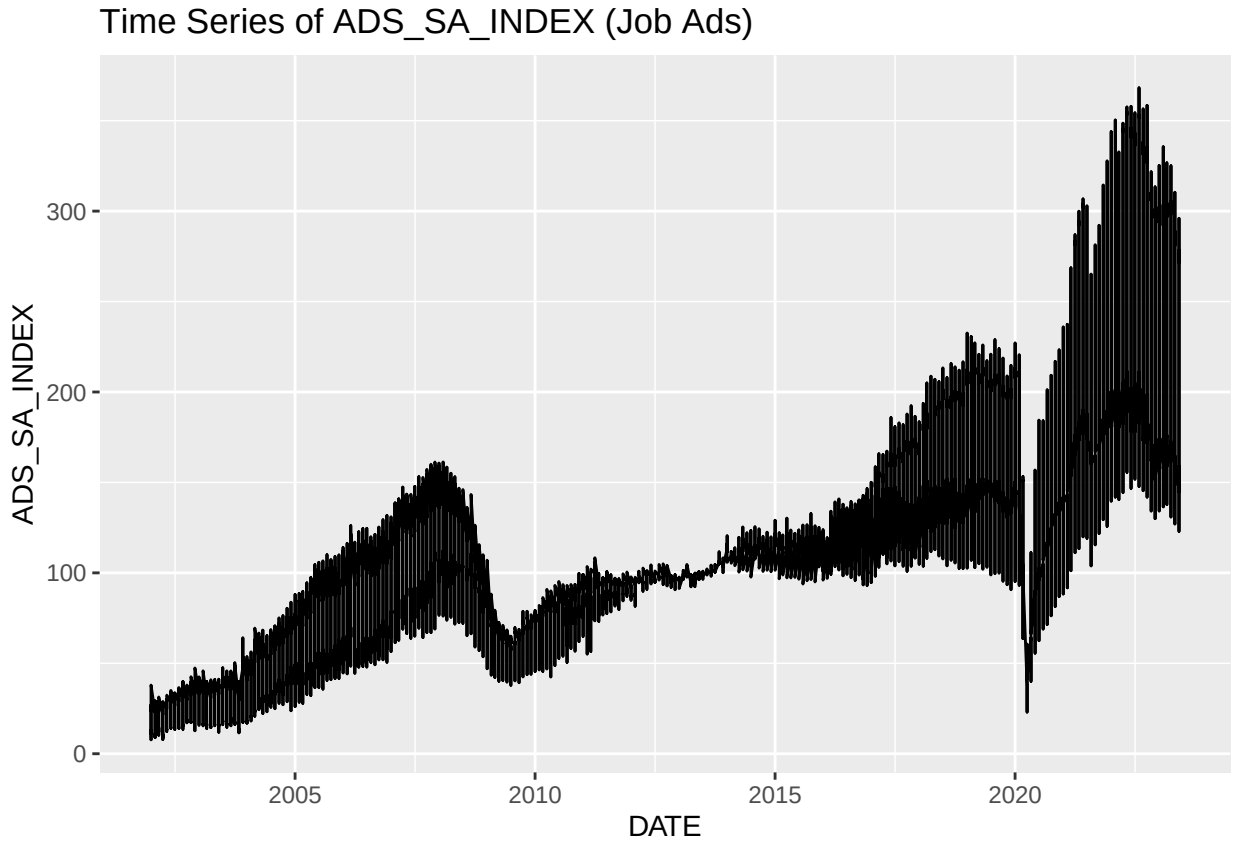
CA_SA_GROWTH_MONTH: 0.009 or 0.9%

- Seasonally adjusted growth : This metric represents the average monthly growth rate in the seasonally adjusted "Applications per Ad" index. Seasonal adjustments remove the effects of seasonal patterns in the data, making it easier to observe fundamental trends and changes over time.

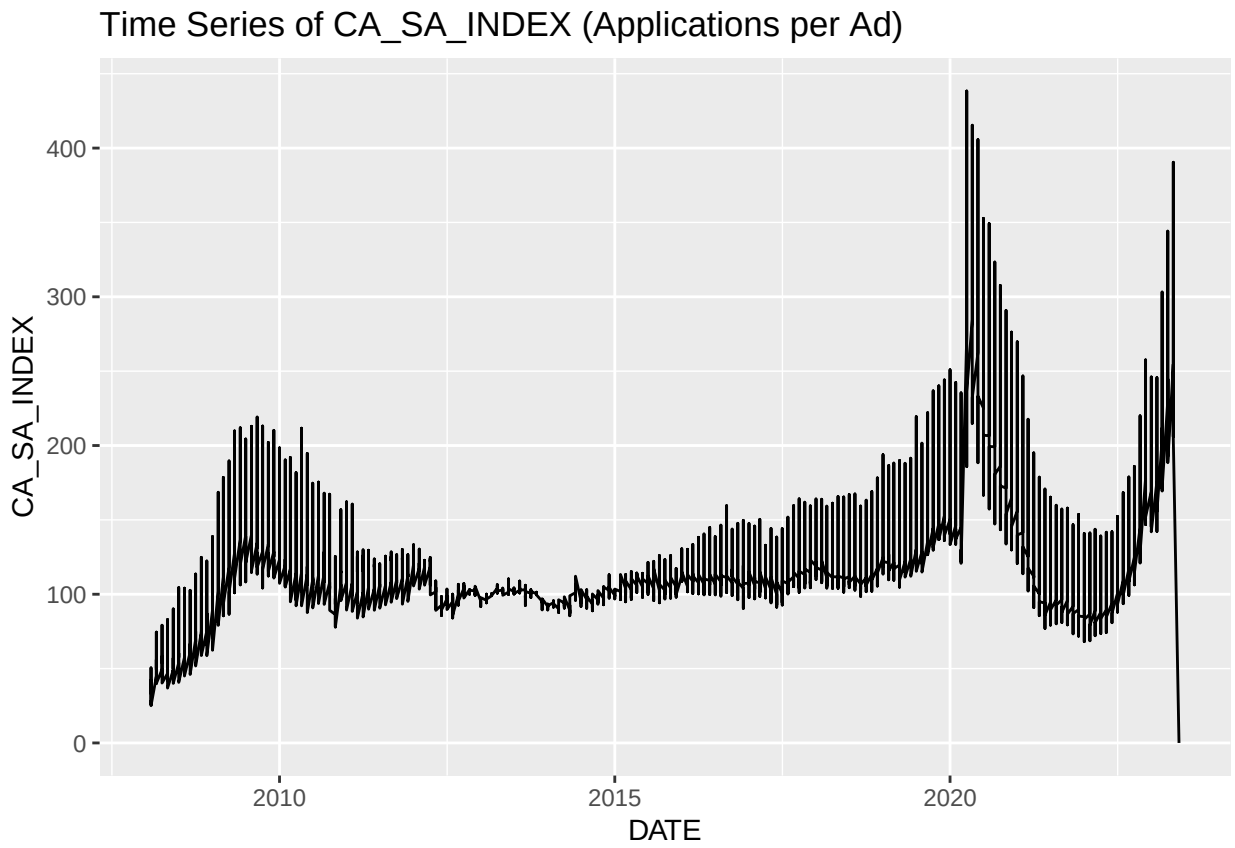
CA_TREND_GROWTH_MONTH: 0.004 or 0.4%

- Trend growth : This metric represents the average monthly growth rate in the trend component of the "Applications per Ad" index. The trend component captures long term changes while filtering out seasonal and random fluctuations.

```
# Plot for jobdf (ADS_SA_INDEX)
ggplot(jobdf, aes(x=DATE, y=ADS_SA_INDEX)) + geom_line() + ggtitle("Time Series of ADS_SA_INDEX (Job Ads
```



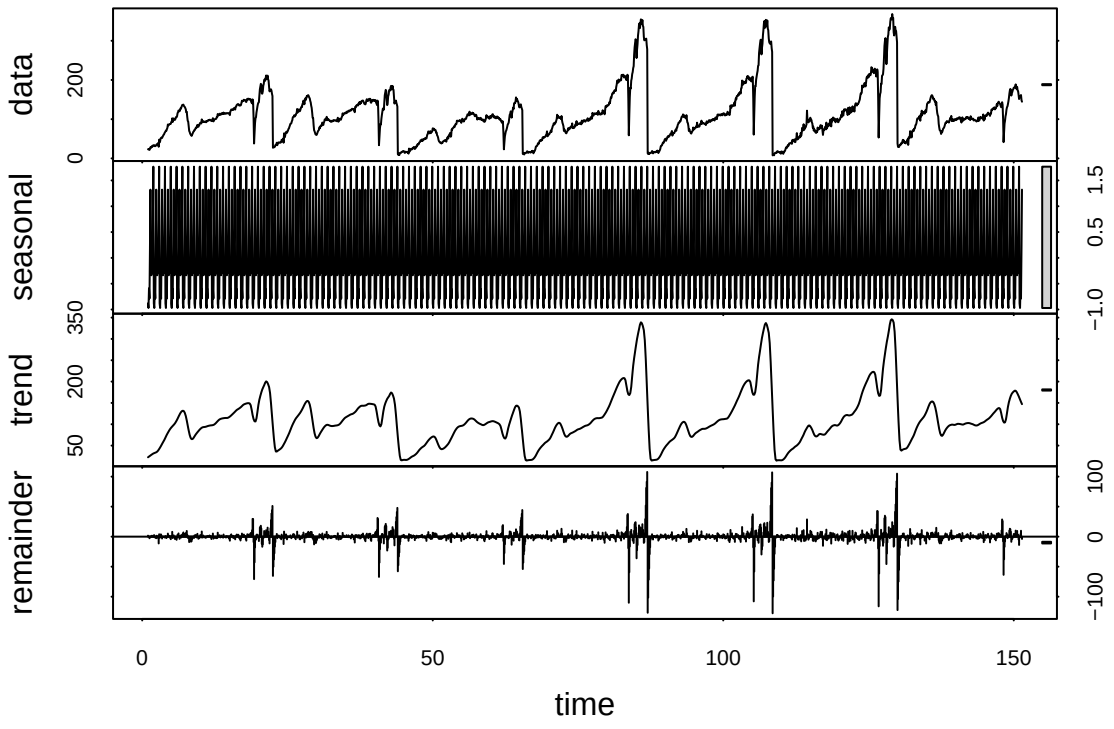
```r
# Plot for appdf (CA_SA_INDEX)
ggplot(appdf, aes(x=DATE, y=CA_SA_INDEX)) + geom_line() + ggtitle("Time Series of CA_SA_INDEX (Applicat
```



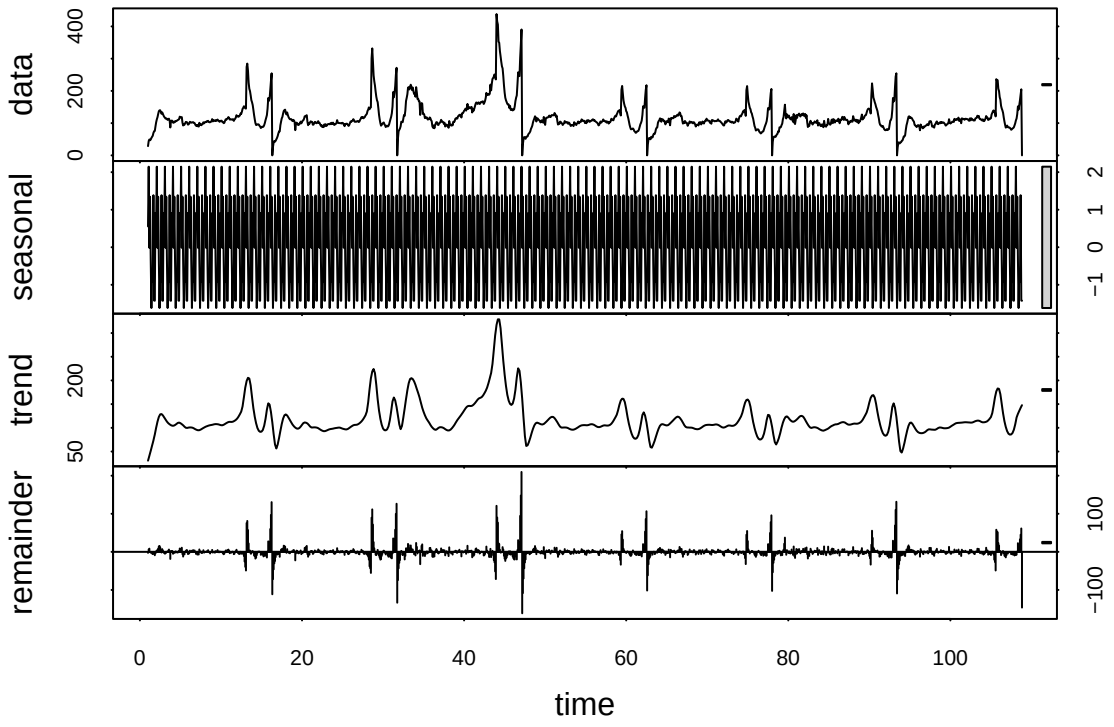
- The two major fluctuations in both the graphs are related due to economic recession in 2018 and the pandemic in 2020.
- Thee jobs ads decreases during the recession and pandemic which leads to the sudden increase in more applications per job ads.

Seasonal decomposition

```r
# Decompose the 'ADS_SA_INDEX' time series from jobdf
job_stl <- stl(ts(jobdf$ADS_SA_INDEX, frequency=12), s.window="periodic")
plot(job_stl)
```




```
# Decompose the 'CA_SA_INDEX' time series from appdf
app_stl <- stl(ts(appdf$CA_SA_INDEX, frequency=12), s.window="periodic")
plot(app_stl)
```



Correlation anaylsis

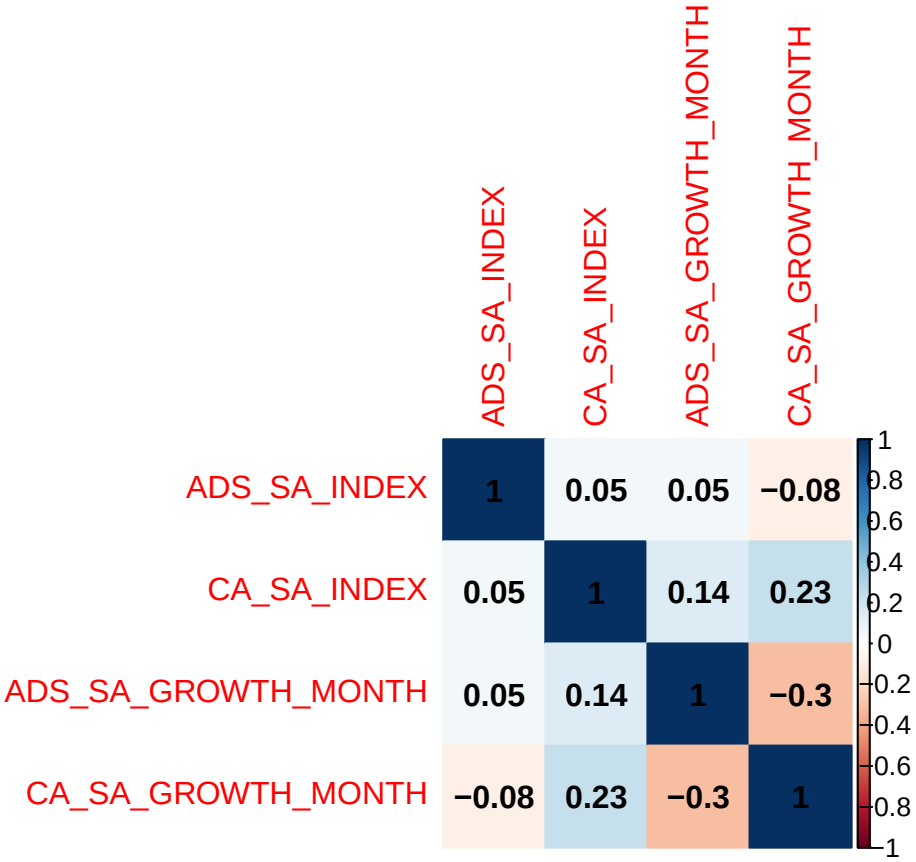
```
# Merge the data frames based on DATE
merged_df <- merge(jobdf, appdf, by="DATE")

# Calculate the correlation matrix
cor_matrix <- cor(merged_df[,c("ADS_SA_INDEX", "CA_SA_INDEX", "ADS_SA_GROWTH_MONTH", "CA_SA_GROWTH_MONTH

# Show the correlation matrix
print(cor_matrix)
```

```
##                     ADS_SA_INDEX CA_SA_INDEX ADS_SA_GROWTH_MONTH
## ADS_SA_INDEX          1.00000000  0.05312508          0.05044988
## CA_SA_INDEX           0.05312508  1.00000000          0.14097266
## ADS_SA_GROWTH_MONTH   0.05044988  0.14097266          1.00000000
## CA_SA_GROWTH_MONTH   -0.08263704  0.23194467         -0.30415249
##                     CA_SA_GROWTH_MONTH
## ADS_SA_INDEX                -0.08263704
## CA_SA_INDEX                  0.23194467
## ADS_SA_GROWTH_MONTH         -0.30415249
## CA_SA_GROWTH_MONTH           1.00000000
```

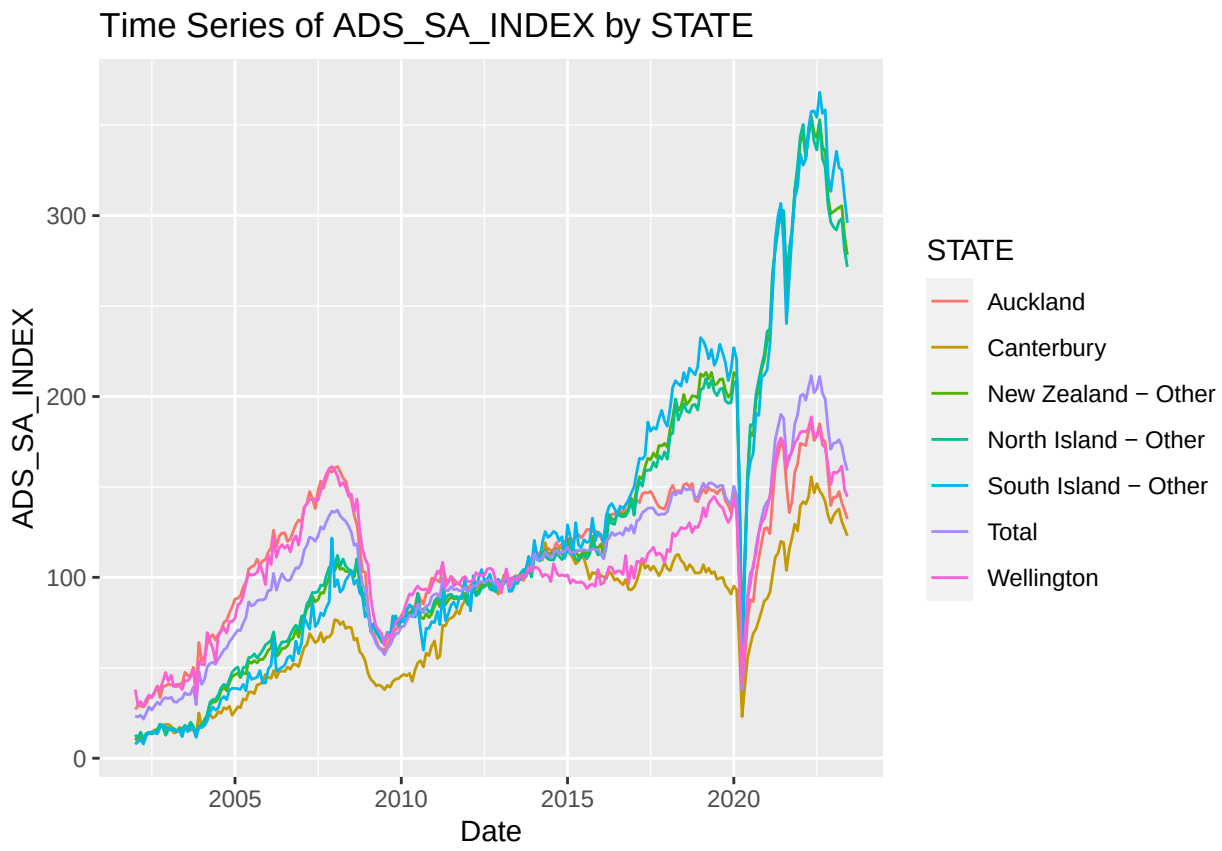
```
# Generate the heatmap
corrplot(cor_matrix, method="color",
         addCoef.col="black") # Add correlation coefficients
```



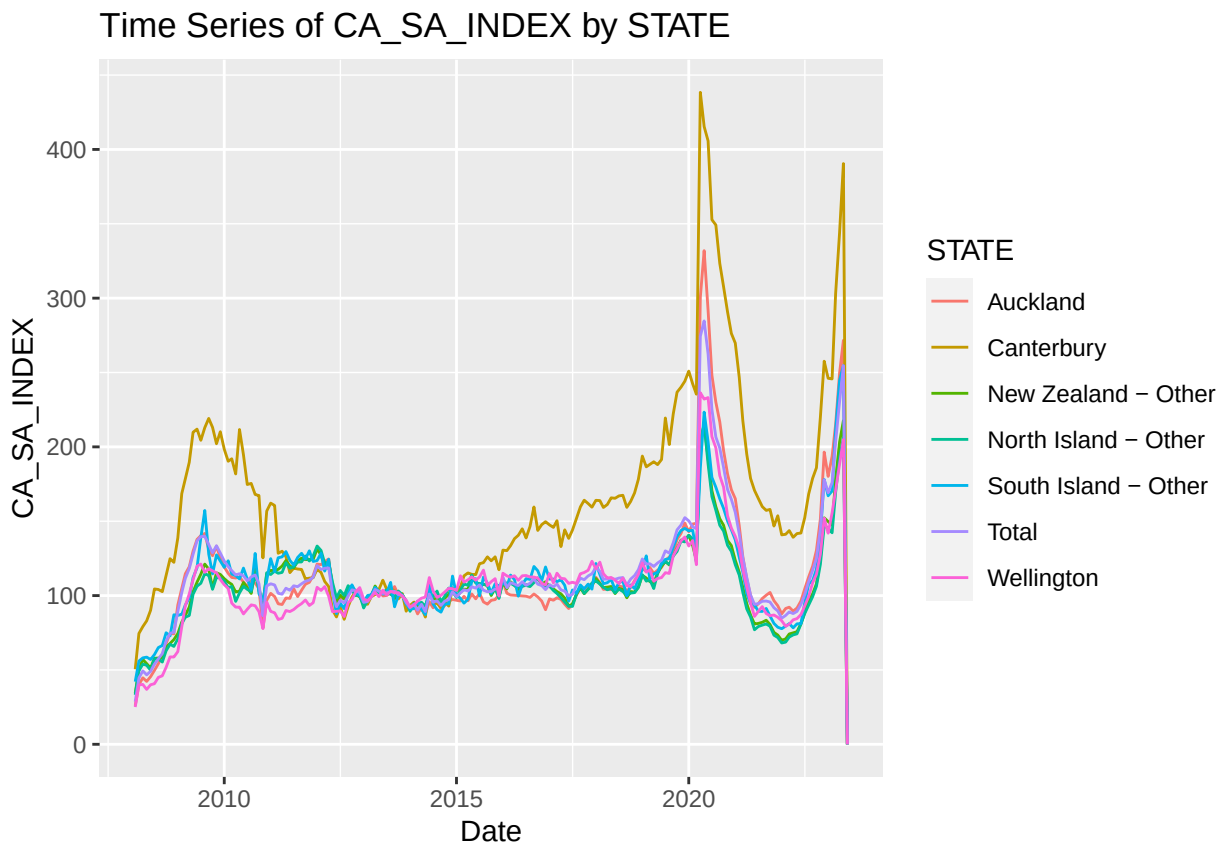
The correlation matrix and heatmap show the following relationships between the selected variables:

- ADS_SA_INDEX and CA_SA_INDEX: A low correlation of 0.05, indicating that the job ad index and the applications per ad index are not strongly related in terms of their levels.
- ADS_SA_GROWTH_MONTH and CA_SA_GROWTH_MONTH: A negative correlation of -0.304, suggesting that when the monthly growth rate of job ads goes up, the monthly growth rate of applications per ad tends to go down, or vice versa.
- ADS_SA_INDEX and ADS_SA_GROWTH_MONTH: A correlation of 0.050, indicating little to no relationship between the level and monthly growth rate of job ads.

```
# Time Series Plot for ADS_SA_INDEX by STATE
ggplot(jobdf, aes(x=DATE, y=ADS_SA_INDEX, color=STATE)) +
  geom_line() +
  ggtitle("Time Series of ADS_SA_INDEX by STATE") +
  xlab("Date") +
  ylab("ADS_SA_INDEX")
```




```
# Time Series Plot for CA_SA_INDEX by STATE
ggplot(appdf, aes(x=DATE, y=CA_SA_INDEX, color=STATE)) +
  geom_line() +
  ggtitle("Time Series of CA_SA_INDEX by STATE") +
  xlab("Date") +
  ylab("CA_SA_INDEX")
```



mONTLY GROWTH RATE

```
head(jobdf)
```

```
## # A tibble: 6 x 9
##   DATE       COUNTRY     STATE ADS_SA_INDEX ADS_TREND_INDEX ADS_SA_GROWTH_MONTH
##   <date>     <chr>       <chr>        <dbl>            <dbl>                <dbl>
## 1 2002-01-01 New Zealand Total         23.4             23.0               0
## 2 2002-02-01 New Zealand Total         22.8             23.3              -0.0235
## 3 2002-03-01 New Zealand Total         23.8             23.8               0.0423
## 4 2002-04-01 New Zealand Total         21.7             24.4              -0.0895
## 5 2002-05-01 New Zealand Total         24.8             25.2               0.146
## 6 2002-06-01 New Zealand Total         28.4             26.2               0.145
## # i 3 more variables: ADS_SA_GROWTH_PCP <dbl>, ADS_TREND_GROWTH_MONTH <dbl>,
## #   ADS_TREND_GROWTH_PCP <dbl>
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
jobdf$YEAR <- format(as.Date(jobdf$DATE), "%Y")
appdf$YEAR <- format(as.Date(appdf$DATE), "%Y")

head(jobdf)
```

```
## # A tibble: 6 x 10
##   DATE       COUNTRY     STATE ADS_SA_INDEX ADS_TREND_INDEX ADS_SA_GROWTH_MONTH
##   <date>     <chr>       <chr>        <dbl>           <dbl>                <dbl>
## 1 2002-01-01 New Zealand Total         23.4            23.0              0
## 2 2002-02-01 New Zealand Total         22.8            23.3             -0.0235
## 3 2002-03-01 New Zealand Total         23.8            23.8              0.0423
## 4 2002-04-01 New Zealand Total         21.7            24.4             -0.0895
## 5 2002-05-01 New Zealand Total         24.8            25.2              0.146
## 6 2002-06-01 New Zealand Total         28.4            26.2              0.145
## # i 4 more variables: ADS_SA_GROWTH_PCP <dbl>, ADS_TREND_GROWTH_MONTH <dbl>,
## #   ADS_TREND_GROWTH_PCP <dbl>, YEAR <chr>
```

```
# Calculate mean monthly growth rates for jobdf
mean_growth_jobdf_yearly <- jobdf %>%
  group_by(STATE, YEAR) %>%
  summarise(Mean_ADS_SA_GROWTH_MONTH = mean(ADS_SA_GROWTH_MONTH, na.rm=TRUE))
```

```
## `summarise()` has grouped output by 'STATE'. You can override using the
## `.groups` argument.
```

```
# Calculate mean monthly growth rates for appdf
mean_growth_appdf_yearly <- appdf %>%
  group_by(STATE, YEAR) %>%
  summarise(Mean_CA_SA_GROWTH_MONTH = mean(CA_SA_GROWTH_MONTH, na.rm=TRUE))
```

```
## `summarise()` has grouped output by 'STATE'. You can override using the
## `.groups` argument.
```

```
# Display mean growth rates
print(mean_growth_jobdf_yearly)
```

```
## # A tibble: 154 x 3
## # Groups:   STATE [7]
##    STATE    YEAR  Mean_ADS_SA_GROWTH_MONTH
##    <chr>    <chr>                    <dbl>
##  1 Auckland 2002                   0.0369
##  2 Auckland 2003                   0.0537
##  3 Auckland 2004                   0.0250
##  4 Auckland 2005                   0.0232
##  5 Auckland 2006                   0.0160
##  6 Auckland 2007                   0.0159
##  7 Auckland 2008                  -0.0344
```

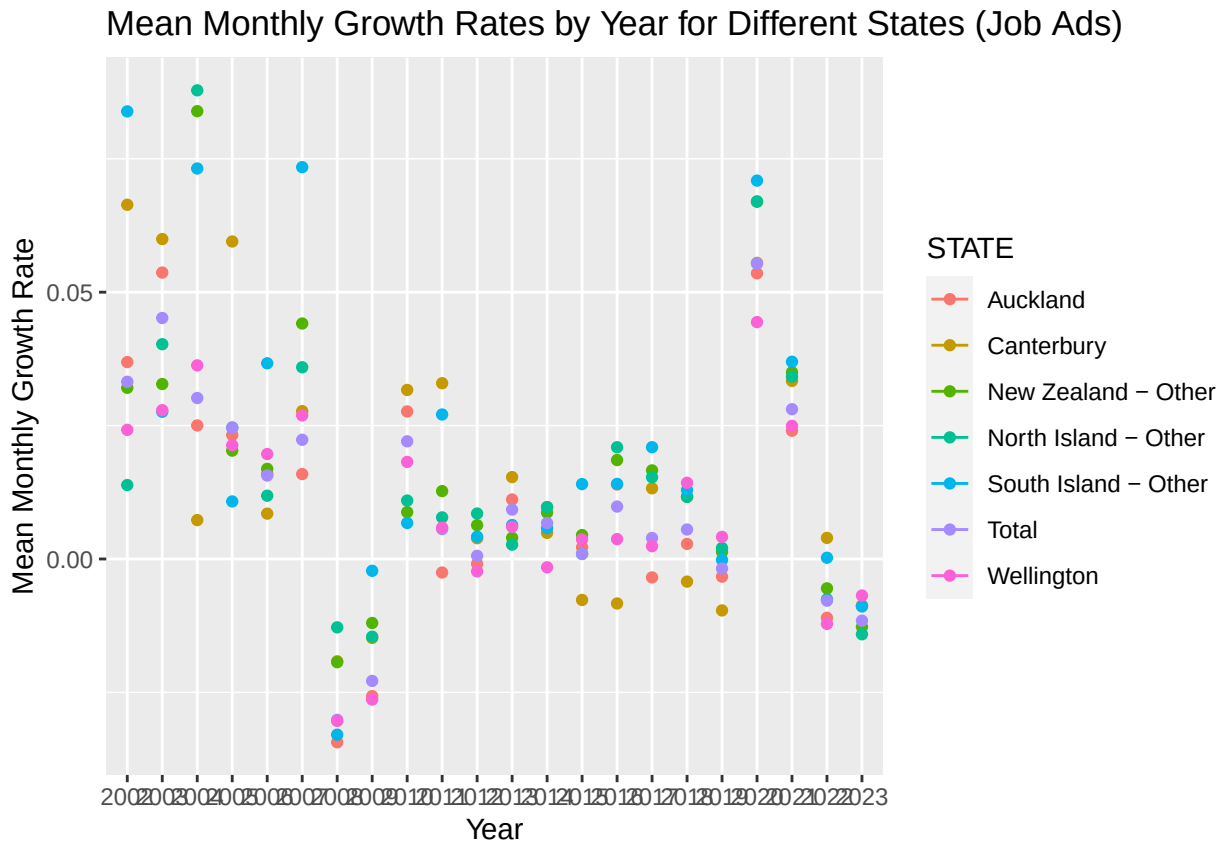
```
##  8 Auckland 2009                  -0.0257
##  9 Auckland 2010                   0.0277
## 10 Auckland 2011                  -0.00254
## # i 144 more rows
```

```
print(mean_growth_appdf_yearly)
```

```
## # A tibble: 112 x 3
## # Groups:   STATE [7]
##    STATE    YEAR  Mean_CA_SA_GROWTH_MONTH
##    <chr>    <chr>                   <dbl>
##  1 Auckland 2008                   0.109
##  2 Auckland 2009                   0.0480
##  3 Auckland 2010                  -0.0158
##  4 Auckland 2011                   0.0139
##  5 Auckland 2012                  -0.00516
##  6 Auckland 2013                  -0.00602
##  7 Auckland 2014                   0.00280
##  8 Auckland 2015                   0.00322
##  9 Auckland 2016                  -0.00807
## 10 Auckland 2017                   0.0151
## # i 102 more rows
```

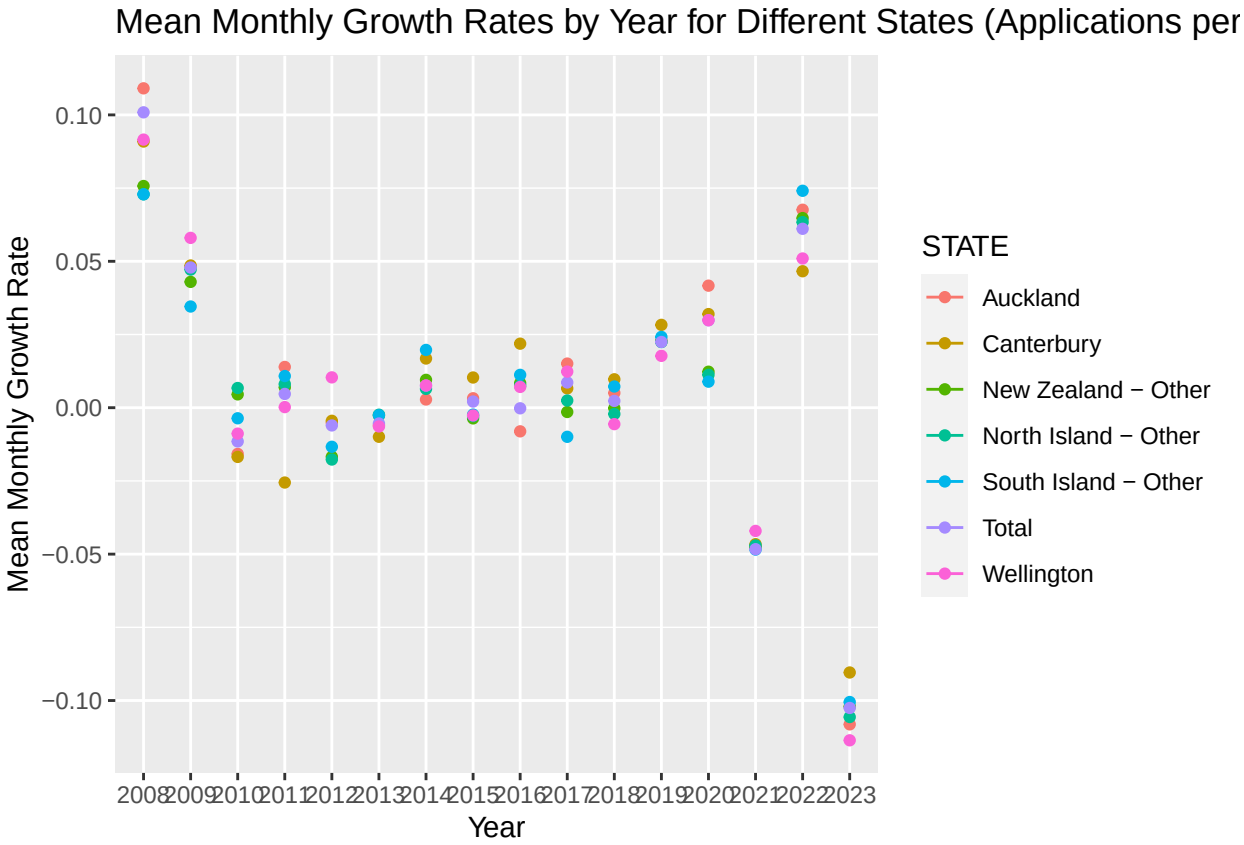
```
# Plot the mean monthly growth rates for each year for different states (Job Ads)
ggplot(mean_growth_jobdf_yearly, aes(x=YEAR, y=Mean_ADS_SA_GROWTH_MONTH, color=STATE)) +
  geom_line() +
  geom_point() +
  ggtitle("Mean Monthly Growth Rates by Year for Different States (Job Ads)") +
  xlab("Year") +
  ylab("Mean Monthly Growth Rate")
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```




```
# Plot the mean monthly growth rates for each year for different states (Applications per Ad)
ggplot(mean_growth_appdf_yearly, aes(x=YEAR, y=Mean_CA_SA_GROWTH_MONTH, color=STATE)) +
  geom_line() +
  geom_point() +
  ggtitle("Mean Monthly Growth Rates by Year for Different States (Applications per Ad)") +
  xlab("Year") +
  ylab("Mean Monthly Growth Rate")
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```



- As we can see in the (Ads_SA_Growth_Month)