TSA Final Project

Shubhangi Gupta, Aditi Jackson, David Robinson

2024-03-29

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Setup

```
# loading packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(lubridate)
library(readxl)
library(ggplot2)
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.3.2
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(Kendall)
library(tseries)
library(outliers)
library(smooth)
```

```
## Warning: package 'smooth' was built under R version 4.3.3

## Loading required package: greybox

## Warning: package 'greybox' was built under R version 4.3.3

## Package "greybox", v2.0.0 loaded.
##
##
## Attaching package: 'greybox'
##
## The following object is masked from 'package:lubridate':
##
##     hm
##
## The following object is masked from 'package:tidyr':
##
##     spread
##
## This is package "smooth", v4.0.0
```

```
library(readr)
library(zoo)
```

```
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Data Wrangling

```
# importing data
transmission_RAW <- read_excel("Data/Raw/Transmission_Lines_1711480217358.xlsx",
                               col_names=TRUE)

#renaming columns, subsetting date and line length columns and dropping NAs
transmission_CLEAN <- transmission_RAW %>%
  select("Line Length (cKM)","Month Of Completion","Year Of Completion (FY)") %>%
  rename("LineLength"= "Line Length (cKM)",
         "Month"="Month Of Completion",
         "Year"="Year Of Completion (FY)")

# checking for N/As
any(is.na(transmission_CLEAN))
```

```
## [1] TRUE
```

```r
# replacing N/As with 0's to account for 0 transmission built in that period
transmission_CLEAN <- transmission_CLEAN %>%
  mutate(LineLength = ifelse(is.na(LineLength),0, LineLength))

#parsing the year column
transmission_CLEAN$Year <- substr(transmission_CLEAN$Year,start=1,stop=4)

#combining month and year columns into a new column
transmission_CLEAN$Date <-  paste0(transmission_CLEAN$Month, "-", transmission_CLEAN$Year)

#converting date column into a date object
transmission_CLEAN$Date <- mdy(transmission_CLEAN$Date)
```

```
## Warning: 5 failed to parse.
```

```r
# grouping by date and summing up line length
transmission_CLEAN <- transmission_CLEAN %>%
  mutate(Month = month(Date), Year = year(Date)) %>%
  group_by(Date)%>%
  summarise(TotalLineLength = sum(LineLength))

# CHECK FOR MISSING DATES
# Days <- as.data.frame(seq.Date(from=as.Date("2015/01/16"),to=as.Date("2023/12/23"), by ="day"))
# colnames(Days) <- "date"
# merge with the data with missing rows
# date_check <-left_join(Days, transmission_CLEAN_Date, by="Date")

# creating time series object
transmission_TS <- ts(transmission_CLEAN,start=c(2015,1),end = c(2023,12),frequency = 12)
tail(transmission_TS,12)
```

```
##           Date TotalLineLength
## Jan 2023 19381              938
## Feb 2023 19470              476
## Mar 2023 19500              528
## Apr 2023 19531             1792
## May 2023 19561              531
## Jun 2023 19592              330
## Jul 2023 19623             1549
## Aug 2023 19653             1820
## Sep 2023 19684              818
## Oct 2023 19714             1203
## Nov 2023    NA                0
## Dec 2023 16451             1522
```

```r
# initial plot
#ggplot(transmission_CLEAN_Date,aes(x=Date,y=TotalLineLength))+
 # geom_line()
```

[**ADD**]

[**ADD**]

[**ADD**]

[**ADD**]

[**ADD**]

[**ADD**]

[**ADD**]