

```
#load packages
```

```
library(readr)
```

```
library(plyr)
```

```
library(dplyr)
```

```
##
```

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

```
## The following objects are masked from 'package:plyr':
```

```
##
```

```
##      arrange, count, desc, failwith, id, mutate, rename, summarise,
```

```
##      summarize
```

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

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(tsibble)
```

```
##
```

```
## Attaching package: 'tsibble'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, union
```

```
library(fable)
```

```
## Loading required package: fabletools
```

```
library(zoo)
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following object is masked from 'package:tsibble':
```

```
##
```

```
##      index
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```

df <- read_csv(file = "data/data_shipping_weather_interpolated_ca_az_or.csv")

## Warning: Missing column names filled in: 'X1' [1]

##
## -- Column specification -----
## cols(
##   X1 = col_double(),
##   Mode = col_character(),
##   ORegionDAT = col_character(),
##   DRegionDAT = col_character(),
##   yw = col_character(),
##   sanitized_cost = col_double(),
##   approx_cost = col_double(),
##   prcp = col_double(),
##   tmax = col_double()
## )

#interpolate the prcp data
df <- df %>%
  mutate(prcp = na.approx(prcp)) %>%
  mutate(tmax = na.approx(tmax))

#sum up the prcp over the last n weeks
#find the percipitation with a lag of n weeks, here we chose 12 weeks
#cumsum(vector)= for all i, take cumulative sum of first i entries, outputting a vector of same length
df <- df %>%
  group_by(Mode,ORegionDAT, DRegionDAT) %>%
  mutate(prcp_lag12 = (cumsum(prcp)-lag(cumsum(prcp), n=12))/12) %>%
  mutate(tmax_lag12 = (cumsum(tmax)-lag(cumsum(tmax), n=12))/12) %>%
  mutate(prcp_lag8 = (cumsum(prcp)-lag(cumsum(prcp), n=8))/8) %>%
  mutate(tmax_lag8 = (cumsum(tmax)-lag(cumsum(tmax), n=8))/8) %>%
  mutate(prcp_lag4 = (cumsum(prcp)-lag(cumsum(prcp), n=4))/4) %>%
  mutate(tmax_lag4 = (cumsum(tmax)-lag(cumsum(tmax), n=4))/4) %>%
  mutate(prcp_lag2 = (cumsum(prcp)-lag(cumsum(prcp), n=2))/2) %>%
  mutate(tmax_lag2 = (cumsum(tmax)-lag(cumsum(tmax), n=2))/2)

write_csv(df, file="data/ca_az_or_weather_lag.csv")

```