Example Air Visit Data Format

Singh, Gurjeet

February 04, 2018

1. Example: Source Data

I am importing and reformatting the source data to a useful format for time series.

```
visit.data.df <- read_csv("Singh_Gurjeet_Example_air_visit_data.csv")
visit.df <-   visit.data.df %>%
              spread(air_store_id, visitors) %>%
              separate(visit_date, c("Year","Month","Day"))
```

2. Example: Define Time Series

I am converting the data into a time series and defining the time series object.

```
newdata=as.ts(visit.df[, -c(1,2,3)], start=c(2016,1,1), frequency=365)
```

3. Example: Simple Imputation

I am performing a quick and simple imputation for demonstration purposes.

```
for(i in 1:ncol(newdata)){
    newdata[is.na(newdata[,i]), i] <- mean(newdata[,i], na.rm = TRUE)
}
```

4. Example: Simple Forecasting (ETS)

I am using the for loop to forecast for each store id for 39 days.

```
myf2=round(forecast(ets(newdata[,1]),39)$mean,0)
#simple ETS
for (i in 2:829) {
  myf= round(forecast(ets(newdata[,i]),39)$mean,0)
  myf2=cbind(myf,myf2)
}
colnames(myf2)=colnames(newdata)
```

5. Example: Final Submission Data

I am converting the forecast result to a data frame with the row id. Next, I format the data into a format of a submission file. Lastly, I save the file in a CSV format.

```
data.result <- rownames_to_column(as.tibble(myf2),var = "RowID")
final_submission <- as.tibble(data.result %>%
                              mutate(Date = seq(from = as.Date("2017-04-23"),
                                                to = as.Date("2017-05-31"),
                                                by= 'day')) %>%
                              gather("StoreID","visitors",
                                     2:ncol(data.result)) %>%
                              unite("id",StoreID,Date, sep = "_")
                             )[,2:3]
write_csv(final_submission, "Singh_Gurjeet_Example_final_submission.csv")
```