

SA app user data Latin America 1Dec20-1Jan22

Dani Behonick

2/2/2022

```
install.packages("readr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)

install.packages("dplyr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)

install.packages("ggplot2")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)

library(readr)
library(dplyr)

##
## 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

library(ggplot2)
```

NEW USERS

```
LAnewusers_Dec20Jan22 <- read.csv("newusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")
```

Data cleaning

```
#columns renamed
colnames(LAnewusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_06", "2021_11", "2021_05", "2021_02", "2021_08",
"2021_12", "2021_07", "2021_10", "2021_09", "2021_04", "2021_03",
"2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAnewusers_Dec20Jan22 <- LAnewusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |
```

```

Country_ID == "EC")

#transposed rows/columns
LNewusers_Dec20Jan22_2 <- t(LNewusers_Dec20Jan22[-1])
colnames(LNewusers_Dec20Jan22_2) <- LNewusers_Dec20Jan22[, 1]

#making time a recognized column variable
LNewusers_Dec20Jan22_2 <- cbind(rownames(LNewusers_Dec20Jan22_2),
  LNewusers_Dec20Jan22_2)
rownames(LNewusers_Dec20Jan22_2) <- NULL

#Totals row removed
LNewusers_Dec20Jan22_2 <- LNewusers_Dec20Jan22_2[-c(15), ]

#make it a dataframe
LNewusers_Dec20Jan22_2 <- as.data.frame(LNewusers_Dec20Jan22_2)

install.packages("tidyr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
library(tidyr)

#pivot to make CountryID a callable variable
LNewusers_Dec20Jan22_3 <- pivot_longer(LNewusers_Dec20Jan22_2, cols=2:8, names_to = "Country", values_to = "New_Users")

#rename columns
colnames(LNewusers_Dec20Jan22_3) <- c("Year_Month", "CountryID", "New_Users")

#make it a dataframe
LNewusers_Dec20Jan22_3 <- as.data.frame(LNewusers_Dec20Jan22_3)

#coerce New_Users to numeric variable rather than character
typeof(LNewusers_Dec20Jan22_3$New_Users)

## [1] "character"
LNewusers_Dec20Jan22_3$New_Users <- as.numeric(LNewusers_Dec20Jan22_3$New_Users)
typeof(LNewusers_Dec20Jan22_3$New_Users)

## [1] "double"

#make Month_Year a date/time object
install.packages("zoo")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
library(zoo)

##
## Attaching package: 'zoo'

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

```

```

Year_Month <- LAnewusers_Dec20Jan22_3$Year_Month
Year_Month_2 <- as.yearmon(as.character(Year_Month), "%Y_%m")
LAnewusers_Dec20Jan22_3[, 'Year_Month'] <- Year_Month_2
typeof(LAnewusers_Dec20Jan22_3$Year_Month)

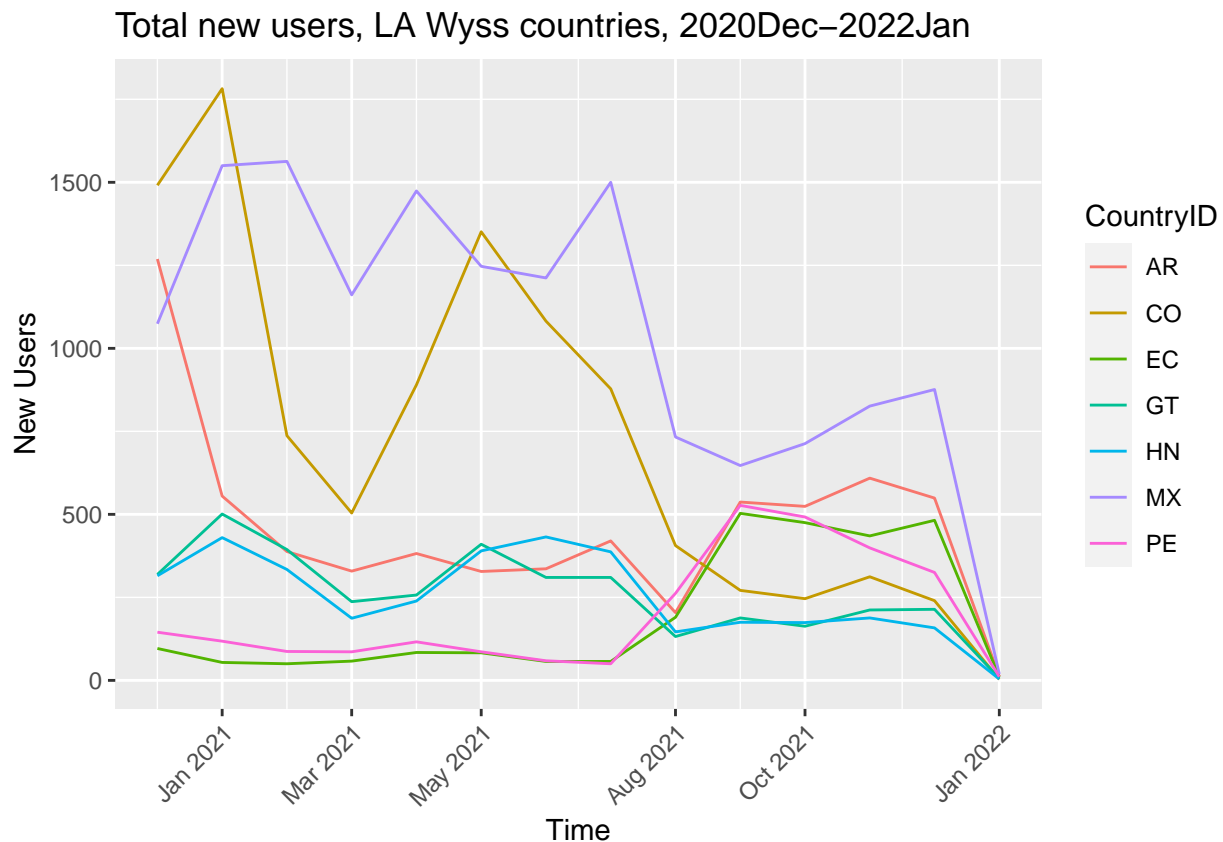
```

```
## [1] "double"
```

```

#jpeg('LAnewusers_Dec20Jan22_3.jpg')
ggplot(data = LAnewusers_Dec20Jan22_3, aes(x=Year_Month, y=New_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Time", y = "New Users",
  title = "Total new users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))

```



RETURN USERS

```

LAreturningusers_Dec20Jan22 <- read.csv("returningusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")

```

Data cleaning

```

#columns renamed
colnames(LAreturningusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_02", "2021_05", "2021_06", "2021_04", "2021_03",
"2021_12", "2021_10", "2021_07", "2021_11", "2021_09", "2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAreturningusers_Dec20Jan22 <- LAreturningusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |

```

```

Country_ID == "EC")

#transposed rows/columns
LAreturningusers_Dec20Jan22_2 <- t(LAreturningusers_Dec20Jan22[-1])
colnames(LAreturningusers_Dec20Jan22_2) <- LAreturningusers_Dec20Jan22[, 1]

#making time a recognized column variable
LAreturningusers_Dec20Jan22_2 <- cbind(rownames(LAreturningusers_Dec20Jan22_2),
  LAreturningusers_Dec20Jan22_2)
rownames(LAreturningusers_Dec20Jan22_2) <- NULL

#Totals row removed
LAreturningusers_Dec20Jan22_2 <- LAreturningusers_Dec20Jan22_2[-c(14), ]

#make it a dataframe
LAreturningusers_Dec20Jan22_2 <- as.data.frame(LAreturningusers_Dec20Jan22_2)

#pivot to make CountryID a callable variable
LAreturningusers_Dec20Jan22_3 <- pivot_longer(LAreturningusers_Dec20Jan22_2, cols=2:8, names_to = "Country_ID", values_to = "Returning_Users")

#rename columns
colnames(LAreturningusers_Dec20Jan22_3) <- c("Year_Month", "CountryID", "Returning_Users")

#make it a dataframe
LAreturningusers_Dec20Jan22_3 <- as.data.frame(LAreturningusers_Dec20Jan22_3)

#coerce Returning_Users to numeric variable rather than character
LAreturningusers_Dec20Jan22_3$Returning_Users <- as.numeric(LAreturningusers_Dec20Jan22_3$Returning_Users)
typeof(LAreturningusers_Dec20Jan22_3$Returning_Users)

## [1] "double"

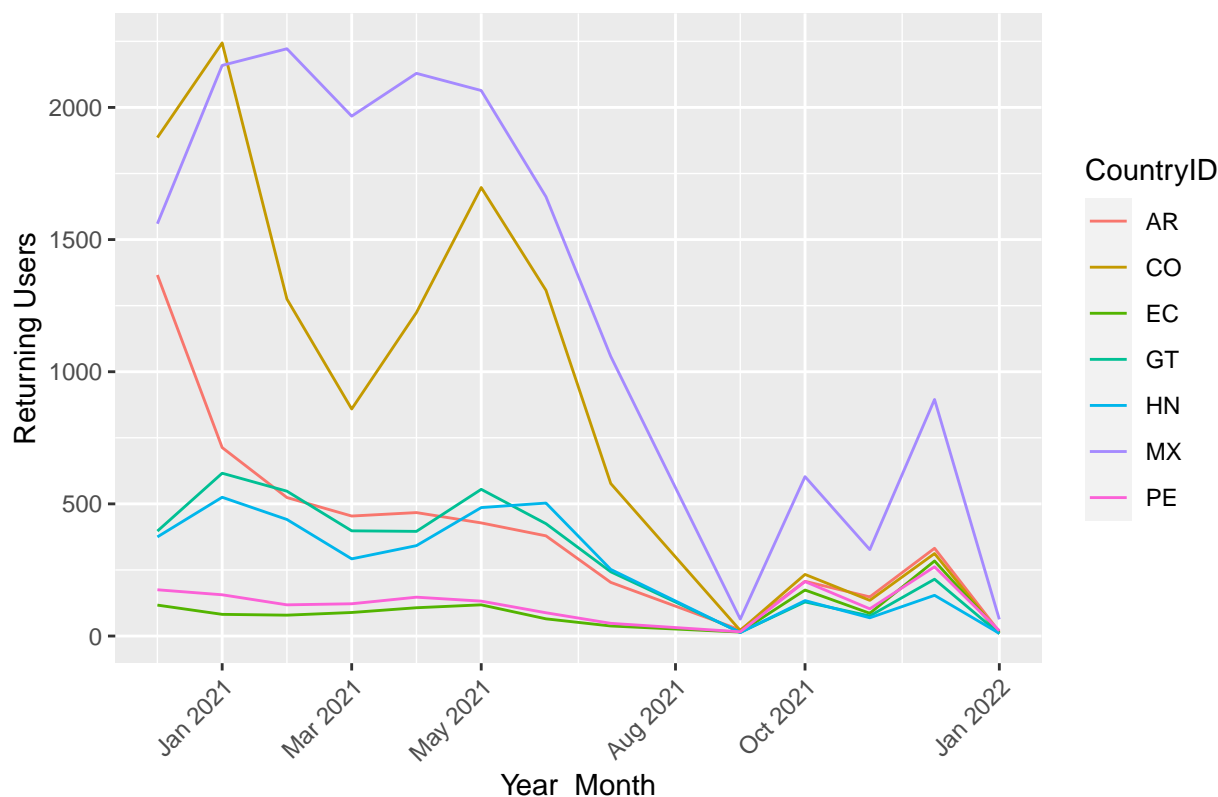
#make Month_Year a date/time object
Year_Month_returning <- LAreturningusers_Dec20Jan22_3$Year_Month
Year_Month_returning_2 <- as.yearmon(as.character(Year_Month_returning), "%Y_%m")
LAreturningusers_Dec20Jan22_3[, 'Year_Month'] <- Year_Month_returning_2
typeof(LAreturningusers_Dec20Jan22_3$Year_Month)

## [1] "double"

#jpeg('LAreturningusers_Dec20Jan22_3.jpg')
ggplot(data = LAreturningusers_Dec20Jan22_3, aes(x=Year_Month, y=Returning_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Year_Month", y = "Returning Users",
  title = "Total returning users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))

```

Total returning users, LA Wyss countries, 2020Dec–2022Jan



TOTAL USERS

```
LAtotalusers_Dec20Jan22 <- read.csv("totalusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")
```

Data cleaning

```
#columns renamed
colnames(LAtotalusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_02", "2021_08", "2021_12", "2021_11", "2021_06",
"2021_07", "2021_05", "2021_09", "2021_10", "2021_04", "2021_03", "2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAtotalusers_Dec20Jan22 <- LAtotalusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |
Country_ID == "EC")

#transposed rows/columns
LAtotalusers_Dec20Jan22_2 <- t(LAtotalusers_Dec20Jan22[-1])
colnames(LAtotalusers_Dec20Jan22_2) <- LAtotalusers_Dec20Jan22[, 1]

#making time a recognized column variable
LAtotalusers_Dec20Jan22_2 <- cbind(rownames(LAtotalusers_Dec20Jan22_2),
LAtotalusers_Dec20Jan22_2)
rownames(LAtotalusers_Dec20Jan22_2) <- NULL

#Totals row removed
```

```

LATotalusers_Dec20Jan22_2 <- LATotalusers_Dec20Jan22_2[-c(15), ]

#make it a dataframe
LATotalusers_Dec20Jan22_2 <- as.data.frame(LATotalusers_Dec20Jan22_2)

#pivot to make CountryID a callable variable
LATotalusers_Dec20Jan22_3 <- pivot_longer(LATotalusers_Dec20Jan22_2, cols=2:8, names_to = "Country", va

#rename columns
colnames(LATotalusers_Dec20Jan22_3) <- c("Year_Month", "CountryID", "Total_Users")

#make it a dataframe
LATotalusers_Dec20Jan22_3 <- as.data.frame(LATotalusers_Dec20Jan22_3)

#coerce Total_Users to numeric variable rather than character
LATotalusers_Dec20Jan22_3$Total_Users <- as.numeric(LATotalusers_Dec20Jan22_3$Total_Users)
typeof(LATotalusers_Dec20Jan22_3$Total_Users)

## [1] "double"

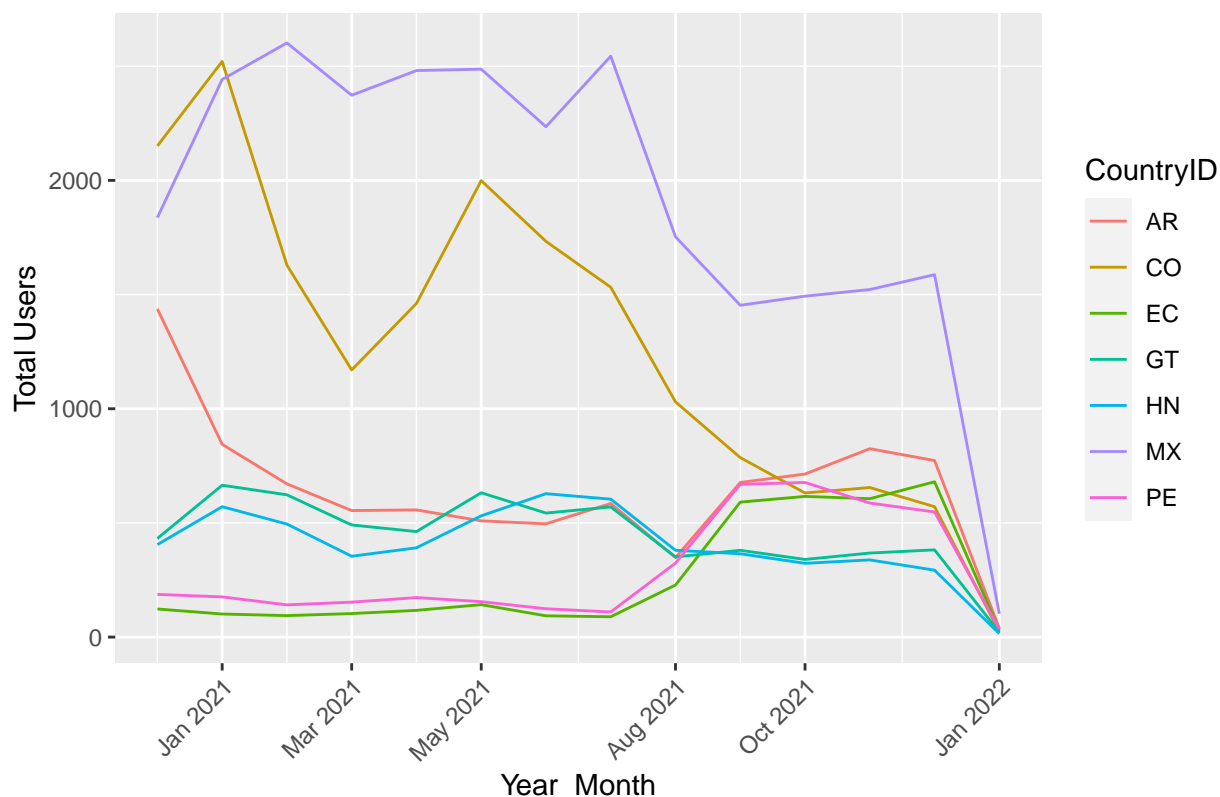
#make Month_Year a date/time object
Year_Month_total <- LATotalusers_Dec20Jan22_3$Year_Month
Year_Month_total_2 <- as.yearmon(as.character(Year_Month_total), "%Y_%m")
LATotalusers_Dec20Jan22_3[, 'Year_Month'] <- Year_Month_total_2
typeof(LATotalusers_Dec20Jan22_3$Year_Month)

## [1] "double"

#jpeg('LATotalusers_Dec20Jan22_3.jpg')
ggplot(data = LATotalusers_Dec20Jan22_3, aes(x=Year_Month, y=Total_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Year_Month", y = "Total Users",
  title = "Total users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))

```

Total users, LA Wyss countries, 2020Dec–2022Jan



PREDICTING FUTURE USE New users, MX

```
LAnewusers_Dec20Jan22_3_MX <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "MX")

newusers_MX_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_MX)

New_dates <- data.frame(Year_Month= c("2022_04", "2022_07"))
New_Year_Month <- New_dates$Year_Month
New_Year_Month_2 <- as.yearmon(as.character(New_Year_Month), "%Y_%m")
New_dates[, 'Year_Month'] <- New_Year_Month_2
typeof(New_dates$Year_Month)
```

```
## [1] "double"
```

```
New_users_prediction_col_MX <- predict(newusers_MX_lm, New_dates)
New_users_prediction_MX <- cbind(New_dates, New_users_prediction_col_MX)
```

New users, AR

```
LAnewusers_Dec20Jan22_3_AR <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "AR")

newusers_AR_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_AR)

New_users_prediction_col_AR <- predict(newusers_AR_lm, New_dates)
New_users_prediction_AR <- cbind(New_dates, New_users_prediction_col_AR)
```

New users, CO

```

LAnewusers_Dec20Jan22_3_CO <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "CO")

newusers_CO_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_CO)

New_users_prediction_col_CO <- predict(newusers_CO_lm, New_dates)
New_users_prediction_CO <- cbind(New_dates, New_users_prediction_col_CO)

```

New users, GT

```

LAnewusers_Dec20Jan22_3_GT <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "GT")

newusers_GT_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_GT)

New_users_prediction_col_GT <- predict(newusers_GT_lm, New_dates)
New_users_prediction_GT <- cbind(New_dates, New_users_prediction_col_GT)

```

New users, HN

```

LAnewusers_Dec20Jan22_3_HN <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "HN")

newusers_HN_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_HN)

New_users_prediction_col_HN <- predict(newusers_HN_lm, New_dates)
New_users_prediction_HN <- cbind(New_dates, New_users_prediction_col_HN)

```

New users, PE

```

LAnewusers_Dec20Jan22_3_PE <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "PE")

newusers_PE_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_PE)

New_users_prediction_col_PE <- predict(newusers_PE_lm, New_dates)
New_users_prediction_PE <- cbind(New_dates, New_users_prediction_col_PE)

```

New users, EC

```

LAnewusers_Dec20Jan22_3_EC <- LAnewusers_Dec20Jan22_3 %>% filter(CountryID == "EC")

newusers_EC_lm <- lm(New_Users~Year_Month, LAnewusers_Dec20Jan22_3_EC)

New_users_prediction_col_EC <- predict(newusers_EC_lm, New_dates)
New_users_prediction_EC <- cbind(New_dates, New_users_prediction_col_EC)

```

```

New_users_prediction <- cbind(New_dates, New_users_prediction_AR$New_users_prediction_col_AR, New_users_prediction_CO, New_users_prediction_GT, New_users_prediction_HN, New_users_prediction_PE, New_users_prediction_EC)

colnames(New_users_prediction) <- c("Time", "New users prediction AR", "New users prediction CO", "New users prediction GT", "New users prediction HN", "New users prediction PE", "New users prediction EC")

write.csv(New_users_prediction, "New_users_predictionLA.csv")

```

Returning users, MX

```

LAreturningusers_Dec20Jan22_3_MX <- LAreturningusers_Dec20Jan22_3 %>% filter(CountryID == "MX")

returningusers_MX_lm <- lm(Returning_Users~Year_Month, LAreturningusers_Dec20Jan22_3_MX)

Returning_users_prediction_col_MX <- predict(returningusers_MX_lm, New_dates)
Returning_users_prediction_MX <- cbind(New_dates, Returning_users_prediction_col_MX)

```


Returning users, AR

```
LReturningusers_Dec20Jan22_3_AR <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "AR")  
  
returningusers_AR_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_AR)  
  
Returning_users_prediction_col_AR <- predict(returningusers_AR_lm, New_dates)  
Returning_users_prediction_AR <- cbind(New_dates, Returning_users_prediction_col_AR)
```

Returning users, CO

```
LReturningusers_Dec20Jan22_3_CO <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "CO")  
  
returningusers_CO_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_CO)  
  
Returning_users_prediction_col_CO <- predict(returningusers_CO_lm, New_dates)  
Returning_users_prediction_CO <- cbind(New_dates, Returning_users_prediction_col_CO)
```

Returning users, GT

```
LReturningusers_Dec20Jan22_3_GT <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "GT")  
  
returningusers_GT_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_GT)  
  
Returning_users_prediction_col_GT <- predict(returningusers_GT_lm, New_dates)  
Returning_users_prediction_GT <- cbind(New_dates, Returning_users_prediction_col_GT)
```

Returning users, HN

```
LReturningusers_Dec20Jan22_3_HN <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "HN")  
  
returningusers_HN_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_HN)  
  
Returning_users_prediction_col_HN <- predict(returningusers_HN_lm, New_dates)  
Returning_users_prediction_HN <- cbind(New_dates, Returning_users_prediction_col_HN)
```

Returning users, PE

```
LReturningusers_Dec20Jan22_3_PE <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "PE")  
  
returningusers_PE_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_PE)  
  
Returning_users_prediction_col_PE <- predict(returningusers_PE_lm, New_dates)  
Returning_users_prediction_PE <- cbind(New_dates, Returning_users_prediction_col_PE)
```

Returning users, EC

```
LReturningusers_Dec20Jan22_3_EC <- LReturningusers_Dec20Jan22_3 %>% filter(CountryID == "EC")  
  
returningusers_EC_lm <- lm(Returning_Users~Year_Month, LReturningusers_Dec20Jan22_3_EC)  
  
Returning_users_prediction_col_EC <- predict(returningusers_EC_lm, New_dates)  
Returning_users_prediction_EC <- cbind(New_dates, Returning_users_prediction_col_EC)
```

```
Returning_users_prediction <- cbind(New_dates, Returning_users_prediction_AR$Returning_users_prediction,  
colnames(Returning_users_prediction) <- c("Time", "Returning users prediction AR", "Returning users prediction CO", "Returning users prediction GT", "Returning users prediction HN", "Returning users prediction PE", "Returning users prediction EC")
```

```
write.csv(Returning_users_prediction, "Returning_users_predictionLA.csv")
```

Total users, MX

```
LAtotalusers_Dec20Jan22_3_MX <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "MX")
```

```
totalusers_MX_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_MX)
```

```
Total_users_prediction_col_MX <- predict(totalusers_MX_lm, New_dates)
```

```
Total_users_prediction_MX <- cbind(New_dates, Total_users_prediction_col_MX)
```

Total users, AR

```
LAtotalusers_Dec20Jan22_3_AR <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "AR")
```

```
totalusers_AR_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_AR)
```

```
Total_users_prediction_col_AR <- predict(totalusers_AR_lm, New_dates)
```

```
Total_users_prediction_AR <- cbind(New_dates, Total_users_prediction_col_AR)
```

Total users, CO

```
LAtotalusers_Dec20Jan22_3_CO <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "CO")
```

```
totalusers_CO_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_CO)
```

```
Total_users_prediction_col_CO <- predict(totalusers_CO_lm, New_dates)
```

```
Total_users_prediction_CO <- cbind(New_dates, Total_users_prediction_col_CO)
```

Total users, GT

```
LAtotalusers_Dec20Jan22_3_GT <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "GT")
```

```
totalusers_GT_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_GT)
```

```
Total_users_prediction_col_GT <- predict(totalusers_GT_lm, New_dates)
```

```
Total_users_prediction_GT <- cbind(New_dates, Total_users_prediction_col_GT)
```

Total users, HN

```
LAtotalusers_Dec20Jan22_3_HN <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "HN")
```

```
totalusers_HN_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_HN)
```

```
Total_users_prediction_col_HN <- predict(totalusers_HN_lm, New_dates)
```

```
Total_users_prediction_HN <- cbind(New_dates, Total_users_prediction_col_HN)
```

Total users, PE

```
LAtotalusers_Dec20Jan22_3_PE <- LAtotalusers_Dec20Jan22_3 %>% filter(CountryID == "PE")
```

```
totalusers_PE_lm <- lm(Total_Users~Year_Month, LAtotalusers_Dec20Jan22_3_PE)
```

```
Total_users_prediction_col_PE <- predict(totalusers_PE_lm, New_dates)
```

```
Total_users_prediction_PE <- cbind(New_dates, Total_users_prediction_col_PE)
```

Total users, EC

```

LATotalusers_Dec20Jan22_3_EC <- LATotalusers_Dec20Jan22_3 %>% filter(CountryID == "EC")

totalusers_EC_lm <- lm(Total_Users~Year_Month, LATotalusers_Dec20Jan22_3_EC)

Total_users_prediction_col_EC <- predict(totalusers_EC_lm, New_dates)
Total_users_prediction_EC <- cbind(New_dates, Total_users_prediction_col_EC)

Total_users_prediction <- cbind(New_dates, Total_users_prediction_AR$Total_users_prediction_col_AR, Tot
colnames(Total_users_prediction) <- c("Time", "Total users prediction AR", "Total users prediction CO",
write.csv(Total_users_prediction, "Total_users_predictionLA.csv")

```