

Surabaya's Public Health Center Visit Trend Analysis

Pandu Gumelar Pratama

2023-01-20

Introduction

Public Health Center (or “Puskesmas” in Bahasa), is a health center built by government to ease the health access and fees for Public especially with People that don't have enough finance to get to private health center. As one of the major city in Indonesia, Surabaya itself have 63 Public Health Center accross the city (Reference)

In this study case we will analyze the trend between new and recurring patients on Surabaya's Public Health Center From January to November 2022 using the data from Jatim Prov Open Data (Link). The tools that will be used in this study case is RStudio that using R Language as its programming language. The output of this research are the trend starting January 2022 to November 2022 and where is the Public Health Center that have most visitor in average.

Data Preparation

In this section, we will gather the data provided, checking any inconsistent data if any and merging into single file so we can start analyze the data. As previously stated in Introduction section, the data will be gathered from Jatimprov Open Data website (Link). From the link provided there are 11 links indicated each CSV file are based on month-year the data are gathered.

Based on the random sampling of the header of all files downloaded (January, April, August and November) there are 8 columns on each file which are

- Periode (Indicating what month period the data are gathered)
- Wilayah (Indicating the region the data are gathered)
- Kecamatan (Indicating the distict the data are gathered)
- Nama Puskesmas (Indicating the name of the Public Health Center)
- Poli (Indicating the region the data are gathered)
- Baru (Indicating the number of new patients visit)
- Lama (Indicating the number of recurring patients visit)
- Total Kunjungan (Indicating the total of the visit combined)

After collecting all the CSV file, the next step are merging the CSV File into single dataframe. To read CSV files we need to import the “readr” and “dplyr” libraries and using lapply to combine all the csv file in the folder.

```
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(lubridate)
```

```
## Loading required package: timechange
```

```
##
## Attaching package: 'lubridate'
```

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

```
df <- list.files(full.names = TRUE) %>%
  lapply(read.csv, sep=";") %>%
  bind_rows
glimpse(df)
```

```
## Rows: 15,751
## Columns: 11
## $ Periode <chr> "Agustus", "Agustus", "Agustu~
## $ Wilayah <chr> "Surabaya Barat", "Surabaya B~
## $ Kecamatan <chr> "Asem Rowo", "Asem Rowo", "As~
## $ Nama.Puskesmas <chr> "Puskesmas Asemrowo", "Puskes~
## $ Poli <chr> "BATRA", "GIGI", "GIZI", "INS~
## $ Baru <int> 2, 59, 0, 1, 142, NA, NA, NA, ~
## $ Lama <int> 5, 282, 1, 0, 568, NA, NA, NA~
## $ Total.Kunjungan <int> 7, 341, 1, 1, 710, 0, 0, 0, 0~
## $ X..DOCTYPE.html. <chr> NA, NA, NA, NA, NA, NA, NA, N~
## $ X... <chr> NA, NA, NA, NA, NA, NA, NA, N~
## $ X..Options.for.packages.loaded.elsewhere <chr> NA, NA, NA, NA, NA, NA, NA, N~
```

As we can see here there are 65,168 rows based on combined CSV file into single, now we will remove two unused columns to clean up the dataframe which are "X..DOCTYPE.html." and "x.." Using subset function

```
df<-subset(df, select=-c(X..DOCTYPE.html.,X...))
glimpse(df)
```

```
## Rows: 15,751
## Columns: 9
## $ Periode <chr> "Agustus", "Agustus", "Agustu~
## $ Wilayah <chr> "Surabaya Barat", "Surabaya B~
## $ Kecamatan <chr> "Asem Rowo", "Asem Rowo", "As~
## $ Nama.Puskesmas <chr> "Puskesmas Asemrowo", "Puskes~
## $ Poli <chr> "BATRA", "GIGI", "GIZI", "INS~
## $ Baru <int> 2, 59, 0, 1, 142, NA, NA, NA, ~
## $ Lama <int> 5, 282, 1, 0, 568, NA, NA, NA~
## $ Total.Kunjungan <int> 7, 341, 1, 1, 710, 0, 0, 0, 0~
## $ X..Options.for.packages.loaded.elsewhere <chr> NA, NA, NA, NA, NA, NA, NA, N~
```

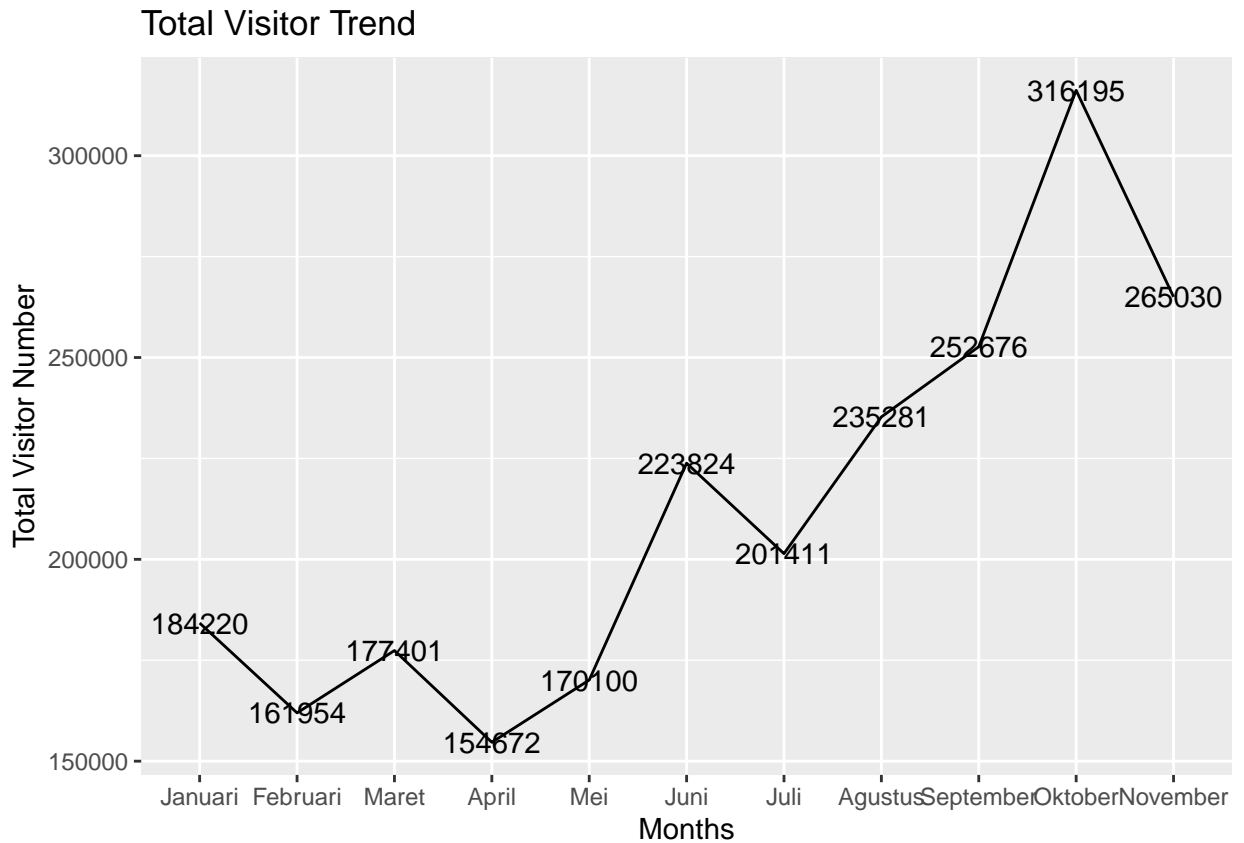
Data Transformation

After two unused columns are removed, we can now progress to the next step, now we want to check how many visits total in all surabaya based on periode/month. To do this we can use group by function combined with summarize and sum function then arrange in custom order since the not all months are in english, starting from Januari(January) to November

```
Summarized_Data_Month_Based<-df%>%
  group_by(Periode)%>%
  summarize(New_Visitor=sum(Baru,na.rm = TRUE),Old_Visitor=sum(Lama,na.rm = TRUE),Total_Visit=sum(Total
  arrange(match(Periode,c("Januari","Februari","Maret","April","Mei","Juni","Juli","Agustus","September"
Summarized_Data_Month_Based
```

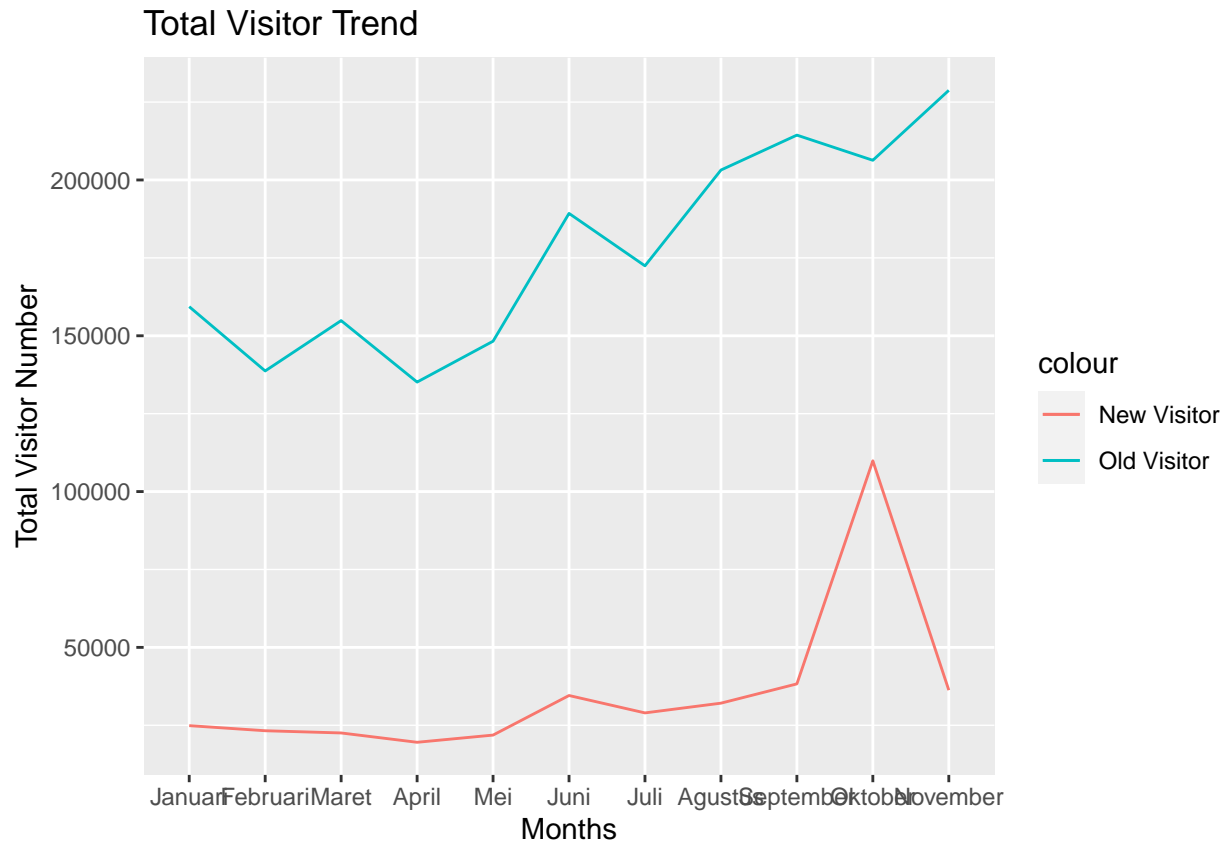
```
## # A tibble: 12 x 4
##   Periode   New_Visitor Old_Visitor Total_Visit
##   <chr>         <int>         <int>         <int>
## 1 Januari         24878         159342        184220
## 2 Februari        23243         138711        161954
## 3 Maret           22539         154862        177401
## 4 April           19536         135136        154672
## 5 Mei             21847         148253        170100
## 6 Juni            34554         189270        223824
## 7 Juli            28970         172441        201411
## 8 Agustus         32096         203185        235281
## 9 September       38280         214396        252676
## 10 Oktober        109878         206317        316195
## 11 November        36265         228765        265030
## 12 <NA>              0              0             NA
```

Now, we will delete the NA row using omit NA Function and using ggplot2 library we will create a line graph using the summarized data to see the trend from January to November 2022. In this process we will change to X axis Legend Name to Months and the Y axis legend name to Total Visitor to Total Visitor Trend for easier

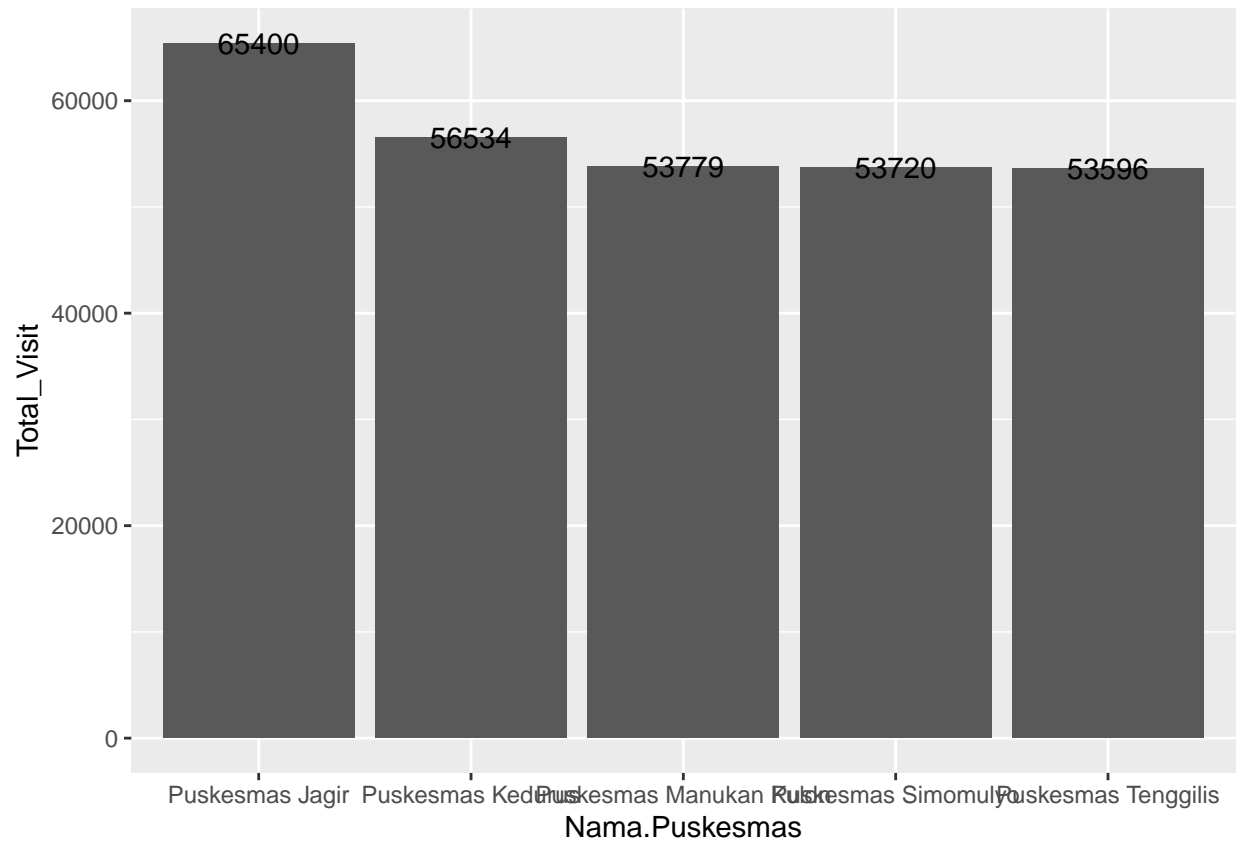


reading

As we can see here the trends of total visitor from January to November is increasing while not stable from 184220 to 265030 with minimum 154672 and peak 316495. Now Using the same method now we will create graph consisting of Old and New Visitor trend from January to November. and in



As We can see here, most of the contributor of total visitor are old/recurring visitor. While the new visitor in the lower number, due to the significant contributor in october, it affect the total number of visitor that reached the peak with significant gap comparing to other months. Now we will get to know which public health care visited the most by using the same method but this time with bar graph and we will limit to top 5 most visited public health care using `top_n` function



As we can see here from the graph above currently the most visited Public Health Care in Surabaya is Puskesmas Jagir with total of 65400 visitors from January to November 2022

Github Links(Link)