

Telangana Tourism Project Files

Mukesh Ale

2023-04-29

Import the required library.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.1      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2     3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr       1.0.1
## -- 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
```

Reading the domestic visitors files

```
domestic_visitors_2019 <- read.csv("domestic_visitors/domestic_visitors_2019.csv")
domestic_visitors_2018 <- read.csv("domestic_visitors/domestic_visitors_2018.csv")
domestic_visitors_2017 <- read.csv("domestic_visitors/domestic_visitors_2017.csv")
domestic_visitors_2016 <- read.csv("domestic_visitors/domestic_visitors_2016.csv")
```

Bind those files

```
domestic_visitors <- rbind(domestic_visitors_2016, domestic_visitors_2017, domestic_visitors_2018, domestic_visitors_2019)
```

Check the data thoroughly

```
str(domestic_visitors)
```

```
## 'data.frame':   1512 obs. of  5 variables:
## $ district: chr  "Adilabad" "Adilabad" "Adilabad" "Adilabad" ...
## $ date     : chr  "01-01-2016" "01-02-2016" "01-03-2016" "01-04-2016" ...
## $ month    : chr  "January" "February" "March" "April" ...
## $ year     : int   2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ visitors: int   792136 937820 582946 341948 252887 368237 447562 614285 491279 94184 ...
```

Date coulmn is in chr data type. Need to convert it to date type.

```
domestic_visitors$date <- as.Date(domestic_visitors$date, format = "%d-%m-%Y")
```

Let's check if the changes are live.

```
str(domestic_visitors)
```

```
## 'data.frame': 1512 obs. of 5 variables:
## $ district: chr "Adilabad" "Adilabad" "Adilabad" "Adilabad" ...
## $ date : Date, format: "2016-01-01" "2016-02-01" ...
## $ month : chr "January" "February" "March" "April" ...
## $ year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ visitors: int 792136 937820 582946 341948 252887 368237 447562 614285 491279 94184 ...
```

Date column is successfully changed to date type.

```
summary(domestic_visitors)
```

```
## district          date          month          year
## Length:1512      Min.   :2016-01-01   Length:1512      Min.   :2016
## Class :character  1st Qu.:2017-01-01   Class :character  1st Qu.:2017
## Mode :character  Median :2018-01-01   Mode :character  Median :2018
##                  Mean    :2017-12-24      Mean    :2018
##                  3rd Qu.:2019-01-01      3rd Qu.:2019
##                  Max.    :2019-12-01      Max.    :2019
##
## visitors
## Min.   : 0
## 1st Qu.: 13205
## Median : 119255
## Mean   : 305348
## 3rd Qu.: 315318
## Max.   :20703778
## NA's   :345
```

There are 345 NA's in visitors column. Fill them with the average (mean) value of the visitors table.

```
domestic_visitors$visitors <- ifelse(is.na(domestic_visitors$visitors), mean(domestic_visitors$visitors), domestic_visitors$visitors)
```

Check if the NA's are gone.

```
summary(domestic_visitors)
```

```
##      district      date      month      year
## Length:1512      Min.   :2016-01-01      Length:1512      Min.   :2016
## Class :character 1st Qu.:2017-01-01      Class :character 1st Qu.:2017
## Mode  :character Median :2018-01-01      Mode  :character Median :2018
##                      Mean  :2017-12-24      Mean  :2018
##                      3rd Qu.:2019-01-01      3rd Qu.:2019
##                      Max.   :2019-12-01      Max.   :2019
##      visitors
## Min.   :      0
## 1st Qu.: 22265
## Median : 184508
## Mean   : 305348
## 3rd Qu.: 305348
## Max.   :20703778
```

Lets see the first 40 rows of the data.

```
print(head(domestic_visitors, 40))
```

```
##      district      date      month year      visitors
## 1      Adilabad 2016-01-01      January 2016      792136.0
## 2      Adilabad 2016-02-01      February 2016      937820.0
## 3      Adilabad 2016-03-01      March 2016      582946.0
## 4      Adilabad 2016-04-01      April 2016      341948.0
## 5      Adilabad 2016-05-01      May 2016      252887.0
## 6      Adilabad 2016-06-01      June 2016      368237.0
## 7      Adilabad 2016-07-01      July 2016      447562.0
## 8      Adilabad 2016-08-01      August 2016      614285.0
## 9      Adilabad 2016-09-01      September 2016      491279.0
## 10     Adilabad 2016-10-01      October 2016      94184.0
## 11     Adilabad 2016-11-01      November 2016      99148.0
## 12     Adilabad 2016-12-01      December 2016      53125.0
## 13 Bhadradri Kothagudem 2016-01-01      January 2016      305348.5
## 14 Bhadradri Kothagudem 2016-02-01      February 2016      305348.5
## 15 Bhadradri Kothagudem 2016-03-01      March 2016      305348.5
## 16 Bhadradri Kothagudem 2016-04-01      April 2016      305348.5
## 17 Bhadradri Kothagudem 2016-05-01      May 2016      305348.5
## 18 Bhadradri Kothagudem 2016-06-01      June 2016      305348.5
## 19 Bhadradri Kothagudem 2016-07-01      July 2016      305348.5
## 20 Bhadradri Kothagudem 2016-08-01      August 2016      305348.5
## 21 Bhadradri Kothagudem 2016-09-01      September 2016      305348.5
## 22 Bhadradri Kothagudem 2016-10-01      October 2016      310133.0
## 23 Bhadradri Kothagudem 2016-11-01      November 2016      252127.0
## 24 Bhadradri Kothagudem 2016-12-01      December 2016      326770.0
## 25      Hyderabad 2016-01-01      January 2016      1122510.0
## 26      Hyderabad 2016-02-01      February 2016      778748.0
## 27      Hyderabad 2016-03-01      March 2016      1017794.0
## 28      Hyderabad 2016-04-01      April 2016      1127738.0
```

```
## 29      Hyderabad 2016-05-01      May 2016 1287181.0
## 30      Hyderabad 2016-06-01      June 2016 12032661.0
## 31      Hyderabad 2016-07-01      July 2016 1096754.0
## 32      Hyderabad 2016-08-01      August 2016 1061137.0
## 33      Hyderabad 2016-09-01 September 2016 832987.0
## 34      Hyderabad 2016-10-01     October 2016 901960.0
## 35      Hyderabad 2016-11-01     November 2016 909733.0
## 36      Hyderabad 2016-12-01     December 2016 1225502.0
## 37      Jagtial 2016-01-01      January 2016 305348.5
## 38      Jagtial 2016-02-01     February 2016 305348.5
## 39      Jagtial 2016-03-01      March 2016 305348.5
## 40      Jagtial 2016-04-01      April 2016 305348.5
```

Remove trailing spaces from some district values.

```
domestic_visitors$district <- trimws(domestic_visitors$district, "right")
```

Check if the spaces are gone.

```
print(head(domestic_visitors, 40))
```

```
##      district      date      month year  visitors
## 1      Adilabad 2016-01-01  January 2016  792136.0
## 2      Adilabad 2016-02-01  February 2016  937820.0
## 3      Adilabad 2016-03-01   March 2016  582946.0
## 4      Adilabad 2016-04-01   April 2016  341948.0
## 5      Adilabad 2016-05-01    May 2016  252887.0
## 6      Adilabad 2016-06-01    June 2016  368237.0
## 7      Adilabad 2016-07-01    July 2016  447562.0
## 8      Adilabad 2016-08-01   August 2016  614285.0
## 9      Adilabad 2016-09-01 September 2016  491279.0
## 10     Adilabad 2016-10-01  October 2016   94184.0
## 11     Adilabad 2016-11-01  November 2016   99148.0
## 12     Adilabad 2016-12-01  December 2016   53125.0
## 13 Bhadradi Kothagudem 2016-01-01  January 2016  305348.5
## 14 Bhadradi Kothagudem 2016-02-01  February 2016  305348.5
## 15 Bhadradi Kothagudem 2016-03-01   March 2016  305348.5
## 16 Bhadradi Kothagudem 2016-04-01   April 2016  305348.5
## 17 Bhadradi Kothagudem 2016-05-01    May 2016  305348.5
## 18 Bhadradi Kothagudem 2016-06-01    June 2016  305348.5
## 19 Bhadradi Kothagudem 2016-07-01    July 2016  305348.5
## 20 Bhadradi Kothagudem 2016-08-01   August 2016  305348.5
## 21 Bhadradi Kothagudem 2016-09-01 September 2016  305348.5
## 22 Bhadradi Kothagudem 2016-10-01  October 2016  310133.0
## 23 Bhadradi Kothagudem 2016-11-01  November 2016  252127.0
## 24 Bhadradi Kothagudem 2016-12-01  December 2016  326770.0
## 25      Hyderabad 2016-01-01  January 2016 1122510.0
## 26      Hyderabad 2016-02-01  February 2016  778748.0
## 27      Hyderabad 2016-03-01   March 2016 1017794.0
```

```
## 28      Hyderabad 2016-04-01      April 2016 1127738.0
## 29      Hyderabad 2016-05-01        May 2016 1287181.0
## 30      Hyderabad 2016-06-01        June 2016 12032661.0
## 31      Hyderabad 2016-07-01        July 2016 1096754.0
## 32      Hyderabad 2016-08-01      August 2016 1061137.0
## 33      Hyderabad 2016-09-01 September 2016  832987.0
## 34      Hyderabad 2016-10-01   October 2016  901960.0
## 35      Hyderabad 2016-11-01 November 2016  909733.0
## 36      Hyderabad 2016-12-01 December 2016 1225502.0
## 37      Jagtial 2016-01-01      January 2016  305348.5
## 38      Jagtial 2016-02-01 February 2016  305348.5
## 39      Jagtial 2016-03-01      March 2016  305348.5
## 40      Jagtial 2016-04-01      April 2016  305348.5
```

domestic_visitors data looks fine, now let's get foreign_visitors data and check everything like we did for domestic_visitors.

```
foreign_visitors_2019 <- read.csv("foreign_visitors/foreign_visitors_2019.csv")
foreign_visitors_2018 <- read.csv("foreign_visitors/foreign_visitors_2018.csv")
foreign_visitors_2017 <- read.csv("foreign_visitors/foreign_visitors_2017.csv")
foreign_visitors_2016 <- read.csv("foreign_visitors/foreign_visitors_2016.csv")
```

```
foreign_visitors <- rbind(foreign_visitors_2016, foreign_visitors_2017, foreign_visitors_2018, foreign_visitors_2019)
```

```
str(foreign_visitors)
```

```
## 'data.frame': 1512 obs. of 5 variables:
## $ district: chr "Adilabad" "Adilabad" "Adilabad" "Adilabad" ...
## $ date : chr "01-01-2016" "01-02-2016" "01-03-2016" "01-04-2016" ...
## $ month : chr "January" "February" "March" "April" ...
## $ year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ visitors: int 2 0 2 0 0 0 4 2 0 0 ...
```

```
summary(foreign_visitors)
```

```
##      district      date      month      year
## Length:1512      Length:1512      Length:1512      Min.   :2016
## Class :character Class :character Class :character 1st Qu.:2017
## Mode  :character Mode  :character Mode  :character Median :2018
##                                     Mean  :2018
##                                     3rd Qu.:2019
##                                     Max.   :2019
##
##      visitors
## Min.   : 0.0
## 1st Qu.: 0.0
## Median : 0.0
## Mean   : 889.2
## 3rd Qu.: 0.0
## Max.   :38933.0
## NA's   :320
```

```
print(head(foreign_visitors, 40))
```

```
##           district      date      month year visitors
## 1      Adilabad 01-01-2016   January 2016         2
## 2      Adilabad 01-02-2016  February 2016         0
## 3      Adilabad 01-03-2016    March 2016         2
## 4      Adilabad 01-04-2016    April 2016         0
## 5      Adilabad 01-05-2016     May 2016         0
## 6      Adilabad 01-06-2016     June 2016         0
## 7      Adilabad 01-07-2016     July 2016         4
## 8      Adilabad 01-08-2016    August 2016         2
## 9      Adilabad 01-09-2016   September 2016         0
## 10     Adilabad 01-10-2016   October 2016         0
## 11     Adilabad 01-11-2016   November 2016         0
## 12     Adilabad 01-12-2016   December 2016         0
## 13 Bhadradri Kothagudem 01-01-2016   January 2016        NA
## 14 Bhadradri Kothagudem 01-02-2016   February 2016        NA
## 15 Bhadradri Kothagudem 01-03-2016    March 2016        NA
## 16 Bhadradri Kothagudem 01-04-2016    April 2016        NA
## 17 Bhadradri Kothagudem 01-05-2016     May 2016        NA
## 18 Bhadradri Kothagudem 01-06-2016     June 2016        NA
## 19 Bhadradri Kothagudem 01-07-2016     July 2016        NA
## 20 Bhadradri Kothagudem 01-08-2016    August 2016        NA
## 21 Bhadradri Kothagudem 01-09-2016   September 2016        NA
## 22 Bhadradri Kothagudem 01-10-2016   October 2016         0
## 23 Bhadradri Kothagudem 01-11-2016   November 2016         0
## 24 Bhadradri Kothagudem 01-12-2016   December 2016         0
## 25      Hyderabad 01-01-2016   January 2016      15865
## 26      Hyderabad 01-02-2016   February 2016     29646
## 27      Hyderabad 01-03-2016    March 2016     13019
## 28      Hyderabad 01-04-2016    April 2016      8402
## 29      Hyderabad 01-05-2016     May 2016      8522
## 30      Hyderabad 01-06-2016     June 2016     10284
## 31      Hyderabad 01-07-2016     July 2016     17142
## 32      Hyderabad 01-08-2016    August 2016     12842
## 33      Hyderabad 01-09-2016   September 2016      9397
## 34      Hyderabad 01-10-2016   October 2016     12808
## 35      Hyderabad 01-11-2016   November 2016      8084
## 36      Hyderabad 01-12-2016   December 2016     17620
## 37      Jagtial 01-01-2016   January 2016         NA
## 38      Jagtial 01-02-2016   February 2016         NA
## 39      Jagtial 01-03-2016    March 2016         NA
## 40      Jagtial 01-04-2016    April 2016         NA
```

We have same issues like domestic visitors file. We will solve them by same methods.

```
foreign_visitors$date <- as.Date(foreign_visitors$date, format = "%d-%m-%Y")
```

```
str(foreign_visitors)
```

```
## 'data.frame': 1512 obs. of 5 variables:
## $ district: chr "Adilabad" "Adilabad" "Adilabad" "Adilabad" ...
## $ date : Date, format: "2016-01-01" "2016-02-01" ...
## $ month : chr "January" "February" "March" "April" ...
## $ year : int 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 ...
## $ visitors: int 2 0 2 0 0 0 4 2 0 0 ...
```

```
foreign_visitors$visitors <- ifelse(is.na(foreign_visitors$visitors), mean(foreign_visitors$visitors, na.rm=T), foreign_visitors$visitors)
```

```
summary(foreign_visitors)
```

```
## district          date          month          year
## Length:1512      Min.   :2016-01-01 Length:1512      Min.   :2016
## Class :character 1st Qu.:2017-01-01 Class :character 1st Qu.:2017
## Mode :character  Median :2018-01-01 Mode :character  Median :2018
##                  Mean    :2017-12-24              Mean    :2018
##                  3rd Qu.:2019-01-01              3rd Qu.:2019
##                  Max.    :2019-12-01              Max.    :2019
## visitors
## Min.   : 0.0
## 1st Qu.: 0.0
## Median : 0.0
## Mean   : 889.2
## 3rd Qu.: 230.0
## Max.   :38933.0
```

```
foreign_visitors$district <- trimws(foreign_visitors$district, "right")
```

```
print(head(foreign_visitors, 40))
```

```
## district          date          month year visitors
## 1 Adilabad 2016-01-01 January 2016 2.0000
## 2 Adilabad 2016-02-01 February 2016 0.0000
## 3 Adilabad 2016-03-01 March 2016 2.0000
## 4 Adilabad 2016-04-01 April 2016 0.0000
## 5 Adilabad 2016-05-01 May 2016 0.0000
## 6 Adilabad 2016-06-01 June 2016 0.0000
## 7 Adilabad 2016-07-01 July 2016 4.0000
## 8 Adilabad 2016-08-01 August 2016 2.0000
## 9 Adilabad 2016-09-01 September 2016 0.0000
## 10 Adilabad 2016-10-01 October 2016 0.0000
## 11 Adilabad 2016-11-01 November 2016 0.0000
## 12 Adilabad 2016-12-01 December 2016 0.0000
## 13 Bhadradi Kothagudem 2016-01-01 January 2016 889.1745
## 14 Bhadradi Kothagudem 2016-02-01 February 2016 889.1745
## 15 Bhadradi Kothagudem 2016-03-01 March 2016 889.1745
## 16 Bhadradi Kothagudem 2016-04-01 April 2016 889.1745
## 17 Bhadradi Kothagudem 2016-05-01 May 2016 889.1745
## 18 Bhadradi Kothagudem 2016-06-01 June 2016 889.1745
```

## 19	Bhadradri	Kothagudem	2016-07-01	July	2016	889.1745
## 20	Bhadradri	Kothagudem	2016-08-01	August	2016	889.1745
## 21	Bhadradri	Kothagudem	2016-09-01	September	2016	889.1745
## 22	Bhadradri	Kothagudem	2016-10-01	October	2016	0.0000
## 23	Bhadradri	Kothagudem	2016-11-01	November	2016	0.0000
## 24	Bhadradri	Kothagudem	2016-12-01	December	2016	0.0000
## 25		Hyderabad	2016-01-01	January	2016	15865.0000
## 26		Hyderabad	2016-02-01	February	2016	29646.0000
## 27		Hyderabad	2016-03-01	March	2016	13019.0000
## 28		Hyderabad	2016-04-01	April	2016	8402.0000
## 29		Hyderabad	2016-05-01	May	2016	8522.0000
## 30		Hyderabad	2016-06-01	June	2016	10284.0000
## 31		Hyderabad	2016-07-01	July	2016	17142.0000
## 32		Hyderabad	2016-08-01	August	2016	12842.0000
## 33		Hyderabad	2016-09-01	September	2016	9397.0000
## 34		Hyderabad	2016-10-01	October	2016	12808.0000
## 35		Hyderabad	2016-11-01	November	2016	8084.0000
## 36		Hyderabad	2016-12-01	December	2016	17620.0000
## 37		Jagtial	2016-01-01	January	2016	889.1745
## 38		Jagtial	2016-02-01	February	2016	889.1745
## 39		Jagtial	2016-03-01	March	2016	889.1745
## 40		Jagtial	2016-04-01	April	2016	889.1745

Both the dataset looks good and now we are good to go. Let's download the datasets.

```
write.csv(domestic_visitors, "domestic_visitors.csv", row.names = FALSE)
write.csv(foreign_visitors, "foreign_visitors.csv", row.names = FALSE)
```