

Data analysis project, checkin analysis for yelp dataset

Dimple Sharma

Wednesday, December 12, 2014

Checkin analysis for yelp dataset

This document contains user checkin analysis. We plot different graphs to analyze check in patterns for a day and for peak hours.

We use time series object to plot hourly timeseries graph and forecast checkins for next 15 hours.

```
#setwd("c:/Dimple/RStats/")
#rm(list=ls())
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.1.1
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.1.1
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 3.1.2
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Loading required package: timeDate
```

```
## Warning: package 'timeDate' was built under R version 3.1.2
```

```
## This is forecast 5.6
```

```
library(xts)
```

```
##
## Attaching package: 'xts'
##
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.1
```

```
checkin_data <- read.csv("datayelp/yelp_academic_dataset_checkin.csv",
                        header = T,
                        stringsAsFactor = F)
# Analyze the data
dim(checkin_data)
```

```
## [1] 31617  170
```

```
summary(checkin_data)
```

```
## checkin_info.9.0 checkin_info.9.1 checkin_info.9.2 checkin_info.9.3
## Min.   : 1      Min.   : 1      Min.   : 1      Min.   : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 1      Median : 1      Median : 1
## Mean   : 3      Mean   : 3      Mean   : 3      Mean   : 3
## 3rd Qu.: 3      3rd Qu.: 2      3rd Qu.: 2      3rd Qu.: 3
## Max.   :522     Max.   :390     Max.   :348     Max.   :425
## NA's   :25130   NA's   :25323   NA's   :25222   NA's   :25269
## checkin_info.9.4 checkin_info.9.5 checkin_info.9.6 checkin_info.20.2
## Min.   : 1      Min.   : 1      Min.   : 1      Min.   : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 2      Median : 2      Median : 2
## Mean   : 3      Mean   : 4      Mean   : 5      Mean   : 4
## 3rd Qu.: 3      3rd Qu.: 3      3rd Qu.: 4      3rd Qu.: 4
## Max.   :485     Max.   :489     Max.   :528     Max.   :284
## NA's   :24440   NA's   :22651   NA's   :24611   NA's   :22778
## checkin_info.20.3 checkin_info.20.0 checkin_info.20.1 checkin_info.20.6
## Min.   : 1      Min.   : 1      Min.   : 1      Min.   : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 2      Median : 2      Median : 2      Median : 2
## Mean   : 4      Mean   : 4      Mean   : 4      Mean   : 5
## 3rd Qu.: 4      3rd Qu.: 4      3rd Qu.: 4      3rd Qu.: 4
## Max.   :466     Max.   :369     Max.   :247     Max.   :517
## NA's   :22485   NA's   :23333   NA's   :23061   NA's   :24268
## checkin_info.20.4 checkin_info.20.5 checkin_info.5.4 checkin_info.4.2
## Min.   : 1      Min.   : 1      Min.   : 1      Min.   : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 2      Median : 2      Median : 1      Median : 1
```

## Mean : 6	Mean : 6	Mean : 3	Mean : 2
## 3rd Qu.: 5	3rd Qu.: 6	3rd Qu.: 2	3rd Qu.: 2
## Max. :565	Max. :382	Max. :241	Max. :93
## NA's :21240	NA's :21777	NA's :29660	NA's :30535
## business_id	checkin_info.12.1	checkin_info.12.0	checkin_info.12.3
## Length:31617	Min. : 1	Min. : 1	Min. : 1
## Class :character	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Mode :character	Median : 2	Median : 2	Median : 2
##	Mean : 4	Mean : 4	Mean : 4
##	3rd Qu.: 4	3rd Qu.: 4	3rd Qu.: 4
##	Max. :388	Max. :499	Max. :419
##	NA's :19798	NA's :19734	NA's :19300
## checkin_info.12.2	checkin_info.12.5	checkin_info.12.4	checkin_info.12.6
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 4	Mean : 5	Mean : 5	Mean : 6
## 3rd Qu.: 4	3rd Qu.: 5	3rd Qu.: 4	3rd Qu.: 5
## Max. :364	Max. :400	Max. :504	Max. :523
## NA's :19635	NA's :17031	NA's :18464	NA's :19092
## checkin_info.15.6	checkin_info.15.4	checkin_info.15.5	checkin_info.15.2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 1
## Mean : 4	Mean : 3	Mean : 4	Mean : 3
## 3rd Qu.: 4	3rd Qu.: 3	3rd Qu.: 4	3rd Qu.: 3
## Max. :560	Max. :549	Max. :403	Max. :370
## NA's :20280	NA's :19940	NA's :18404	NA's :21332
## checkin_info.15.3	checkin_info.15.0	checkin_info.15.1	checkin_info.17.1
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 2
## Mean : 3	Mean : 3	Mean : 3	Mean : 4
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 4
## Max. :453	Max. :433	Max. :315	Max. :353
## NA's :21204	NA's :21491	NA's :21500	NA's :19697
## checkin_info.22.4	checkin_info.22.5	checkin_info.22.6	checkin_info.22.0
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 1	Median : 1
## Mean : 6	Mean : 6	Mean : 4	Mean : 3
## 3rd Qu.: 4	3rd Qu.: 5	3rd Qu.: 3	3rd Qu.: 3
## Max. :591	Max. :688	Max. :381	Max. :250
## NA's :25654	NA's :25732	NA's :27643	NA's :27634
## checkin_info.22.1	checkin_info.22.2	checkin_info.22.3	checkin_info.11.2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 2
## Mean : 3	Mean : 3	Mean : 4	Mean : 4
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 4
## Max. :167	Max. :166	Max. :427	Max. :347
## NA's :27457	NA's :28054	NA's :27005	NA's :20344
## checkin_info.11.3	checkin_info.11.0	checkin_info.11.1	checkin_info.11.6
## Min. : 1	Min. : 1	Min. : 1	Min. : 1

## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 4	Mean : 4	Mean : 4	Mean : 5
## 3rd Qu.: 4	3rd Qu.: 4	3rd Qu.: 3	3rd Qu.: 4
## Max. :432	Max. :493	Max. :385	Max. :624
## NA's :20267	NA's :20463	NA's :20469	NA's :19969
## checkin_info.11.4	checkin_info.11.5	checkin_info.16.5	checkin_info.16.4
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 4	Mean : 5	Mean : 4	Mean : 4
## 3rd Qu.: 4	3rd Qu.: 4	3rd Qu.: 4	3rd Qu.: 4
## Max. :545	Max. :497	Max. :375	Max. :614
## NA's :19405	NA's :17833	NA's :18772	NA's :19483
## checkin_info.16.6	checkin_info.16.1	checkin_info.16.0	checkin_info.16.3
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 4	Mean : 3	Mean : 3	Mean : 3
## 3rd Qu.: 4	3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 3
## Max. :604	Max. :370	Max. :452	Max. :461
## NA's :20513	NA's :20549	NA's :20702	NA's :20357
## checkin_info.16.2	checkin_info.1.0	checkin_info.1.1	checkin_info.1.2
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 1	Median : 1	Median : 1
## Mean : 3	Mean : 3	Mean : 2	Mean : 2
## 3rd Qu.: 3	3rd Qu.: 2	3rd Qu.: 2	3rd Qu.: 2
## Max. :397	Max. :145	Max. :113	Max. :45
## NA's :20656	NA's :30039	NA's :30198	NA's :30127
## checkin_info.1.3	checkin_info.1.4	checkin_info.1.5	checkin_info.1.6
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 3	Mean : 3	Mean : 5	Mean : 5
## 3rd Qu.: 2	3rd Qu.: 2	3rd Qu.: 3	3rd Qu.: 3
## Max. :58	Max. :87	Max. :207	Max. :266
## NA's :30028	NA's :29812	NA's :29088	NA's :29011
## checkin_info.7.3	checkin_info.7.0	checkin_info.4.1	checkin_info.4.0
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 3	Mean : 3	Mean : 2	Mean : 2
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 2	3rd Qu.: 2
## Max. :470	Max. :558	Max. :81	Max. :124
## NA's :27439	NA's :27697	NA's :30582	NA's :30502
## checkin_info.4.3	checkin_info.2.0	checkin_info.4.5	checkin_info.4.4
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 2	Mean : 3	Mean : 3	Mean : 2
## 3rd Qu.: 2	3rd Qu.: 2	3rd Qu.: 2	3rd Qu.: 2
## Max. :171	Max. :65	Max. :177	Max. :108
## NA's :30509	NA's :30528	NA's :30237	NA's :30323

```

## checkin_info.2.5 checkin_info.2.4 checkin_info.3.3 checkin_info.7.4
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 1 Median : 1 Median : 1
## Mean : 4 Mean : 3 Mean : 2 Mean : 4
## 3rd Qu.: 3 3rd Qu.: 2 3rd Qu.: 2 3rd Qu.: 3
## Max. :203 Max. :74 Max. :39 Max. :485
## NA's :29821 NA's :30274 NA's :30842 NA's :27173
## checkin_info.5.3 checkin_info.19.2 checkin_info.19.3 checkin_info.19.0
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 2 Median : 2 Median : 2
## Mean : 3 Mean : 5 Mean : 5 Mean : 4
## 3rd Qu.: 2 3rd Qu.: 5 3rd Qu.: 5 3rd Qu.: 4
## Max. :255 Max. :371 Max. :457 Max. :453
## NA's :29826 NA's :20910 NA's :20569 NA's :21625
## checkin_info.19.1 checkin_info.19.6 checkin_info.19.4 checkin_info.19.5
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 2 Median : 2 Median : 2 Median : 2
## Mean : 5 Mean : 5 Mean : 6 Mean : 6
## 3rd Qu.: 5 3rd Qu.: 5 3rd Qu.: 6 3rd Qu.: 6
## Max. :307 Max. :565 Max. :580 Max. :343
## NA's :21185 NA's :22609 NA's :19720 NA's :20290
## checkin_info.2.3 checkin_info.2.2 checkin_info.2.1 checkin_info.7.2
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 1 Median : 1 Median : 1
## Mean : 2 Mean : 2 Mean : 2 Mean : 3
## 3rd Qu.: 2 3rd Qu.: 2 3rd Qu.: 2 3rd Qu.: 3
## Max. :42 Max. :36 Max. :64 Max. :412
## NA's :30412 NA's :30552 NA's :30629 NA's :27598
## checkin_info.5.5 checkin_info.5.6 checkin_info.7.1 checkin_info.7.6
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 1 Median : 1 Median : 1
## Mean : 2 Mean : 2 Mean : 3 Mean : 3
## 3rd Qu.: 2 3rd Qu.: 2 3rd Qu.: 3 3rd Qu.: 3
## Max. :129 Max. :211 Max. :391 Max. :457
## NA's :29811 NA's :30112 NA's :27608 NA's :27920
## checkin_info.5.1 checkin_info.5.2 checkin_info.7.5 checkin_info.3.2
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 1 Median : 1 Median : 1
## Mean : 3 Mean : 3 Mean : 3 Mean : 2
## 3rd Qu.: 2 3rd Qu.: 2 3rd Qu.: 3 3rd Qu.: 2
## Max. :207 Max. :195 Max. :336 Max. :41
## NA's :29933 NA's :29897 NA's :26892 NA's :30817
## checkin_info.2.6 checkin_info.21.6 checkin_info.21.5 checkin_info.21.4
## Min. : 1 Min. : 1 Min. : 1 Min. : 1
## 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1 1st Qu.: 1
## Median : 1 Median : 2 Median : 2 Median : 2
## Mean : 4 Mean : 4 Mean : 6 Mean : 6
## 3rd Qu.: 3 3rd Qu.: 3 3rd Qu.: 5 3rd Qu.: 5

```

```

## Max. :189      Max. :520      Max. :401      Max. :706
## NA's :29854    NA's :26033    NA's :23528    NA's :23404
## checkin_info.21.3 checkin_info.21.2 checkin_info.21.1 checkin_info.21.0
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 2     Median : 2     Median : 2     Median : 2
## Mean : 4       Mean : 4       Mean : 4       Mean : 4
## 3rd Qu.: 4     3rd Qu.: 4     3rd Qu.: 3     3rd Qu.: 3
## Max. :520      Max. :320      Max. :249      Max. :320
## NA's :24867    NA's :25160    NA's :25317    NA's :25616
## checkin_info.8.5 checkin_info.8.4 checkin_info.8.6 type
## Min. : 1      Min. : 1      Min. : 1      Length:31617
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     Class :character
## Median : 2     Median : 1     Median : 1     Mode :character
## Mean : 4       Mean : 3       Mean : 4
## 3rd Qu.: 3     3rd Qu.: 3     3rd Qu.: 3
## Max. :472      Max. :546      Max. :477
## NA's :24953    NA's :25913    NA's :26397
## checkin_info.8.0 checkin_info.8.3 checkin_info.8.2 checkin_info.5.0
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 1     Median : 1     Median : 1     Median : 1
## Mean : 3       Mean : 3       Mean : 3       Mean : 3
## 3rd Qu.: 3     3rd Qu.: 3     3rd Qu.: 3     3rd Qu.: 2
## Max. :681      Max. :503      Max. :474      Max. :391
## NA's :26520    NA's :26442    NA's :26483    NA's :29958
## checkin_info.17.0 checkin_info.8.1 checkin_info.17.2 checkin_info.17.3
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 2     Median : 1     Median : 2     Median : 2
## Mean : 4       Mean : 3       Mean : 4       Mean : 4
## 3rd Qu.: 4     3rd Qu.: 3     3rd Qu.: 4     3rd Qu.: 4
## Max. :520      Max. :474      Max. :369      Max. :495
## NA's :20066    NA's :26618    NA's :19475    NA's :19192
## checkin_info.17.4 checkin_info.17.5 checkin_info.17.6 checkin_info.14.3
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 2     Median : 2     Median : 2     Median : 1
## Mean : 5       Mean : 5       Mean : 5       Mean : 3
## 3rd Qu.: 5     3rd Qu.: 5     3rd Qu.: 5     3rd Qu.: 3
## Max. :790      Max. :390      Max. :601      Max. :442
## NA's :18548    NA's :18997    NA's :20845    NA's :21182
## checkin_info.14.2 checkin_info.14.1 checkin_info.14.0 checkin_info.14.6
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 1     Median : 1     Median : 1     Median : 2
## Mean : 3       Mean : 3       Mean : 3       Mean : 5
## 3rd Qu.: 3     3rd Qu.: 3     3rd Qu.: 3     3rd Qu.: 4
## Max. :365      Max. :344      Max. :476      Max. :544
## NA's :21373    NA's :21591    NA's :21103    NA's :19683
## checkin_info.14.5 checkin_info.14.4 checkin_info.4.6 checkin_info.13.4
## Min. : 1      Min. : 1      Min. : 1      Min. : 1
## 1st Qu.: 1     1st Qu.: 1     1st Qu.: 1     1st Qu.: 1
## Median : 2     Median : 2     Median : 1     Median : 2

```

## Mean : 5	Mean : 3	Mean : 3	Mean : 4
## 3rd Qu.: 4	3rd Qu.: 3	3rd Qu.: 2	3rd Qu.: 4
## Max. :428	Max. :534	Max. :133	Max. :493
## NA's :17970	NA's :19974	NA's :30358	NA's :19227
## checkin_info.13.5	checkin_info.13.6	checkin_info.13.0	checkin_info.13.1
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 5	Mean : 5	Mean : 3	Mean : 3
## 3rd Qu.: 5	3rd Qu.: 5	3rd Qu.: 3	3rd Qu.: 3
## Max. :407	Max. :554	Max. :466	Max. :352
## NA's :17321	NA's :19255	NA's :20484	NA's :20631
## checkin_info.13.2	checkin_info.13.3	checkin_info.10.6	checkin_info.10.5
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 2	Median : 2	Median : 2	Median : 2
## Mean : 3	Mean : 3	Mean : 5	Mean : 4
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 4	3rd Qu.: 4
## Max. :376	Max. :462	Max. :497	Max. :476
## NA's :20339	NA's :20234	NA's :22289	NA's :20176
## checkin_info.10.4	checkin_info.10.3	checkin_info.10.2	checkin_info.10.1
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 3	Mean : 3	Mean : 3	Mean : 3
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 3
## Max. :534	Max. :446	Max. :375	Max. :419
## NA's :21875	NA's :22699	NA's :22956	NA's :23019
## checkin_info.10.0	checkin_info.3.5	checkin_info.23.1	checkin_info.23.0
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 3	Mean : 4	Mean : 3	Mean : 3
## 3rd Qu.: 3	3rd Qu.: 2	3rd Qu.: 3	3rd Qu.: 3
## Max. :558	Max. :238	Max. :158	Max. :226
## NA's :23021	NA's :30171	NA's :28847	NA's :28935
## checkin_info.23.3	checkin_info.23.2	checkin_info.23.5	checkin_info.23.4
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 2	Median : 2
## Mean : 4	Mean : 3	Mean : 6	Mean : 6
## 3rd Qu.: 3	3rd Qu.: 3	3rd Qu.: 4	3rd Qu.: 4
## Max. :314	Max. :196	Max. :591	Max. :398
## NA's :28435	NA's :28792	NA's :27148	NA's :27300
## checkin_info.23.6	checkin_info.6.5	checkin_info.3.0	checkin_info.0.5
## Min. : 1	Min. : 1	Min. : 1	Min. : 1
## 1st Qu.: 1	1st Qu.: 1	1st Qu.: 1	1st Qu.: 1
## Median : 1	Median : 1	Median : 1	Median : 1
## Mean : 4	Mean : 2	Mean : 3	Mean : 6
## 3rd Qu.: 3	3rd Qu.: 2	3rd Qu.: 2	3rd Qu.: 4
## Max. :322	Max. :237	Max. :67	Max. :305
## NA's :28823	NA's :28712	NA's :30670	NA's :28331
## checkin_info.0.4	checkin_info.0.6	checkin_info.0.1	checkin_info.0.0
## Min. : 1	Min. : 1	Min. : 1	Min. : 1

```
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 2      Median : 1      Median : 1
## Mean : 4        Mean : 6        Mean : 3        Mean : 3
## 3rd Qu.: 3      3rd Qu.: 4      3rd Qu.: 2      3rd Qu.: 3
## Max. :144       Max. :441       Max. :173       Max. :271
## NA's :29119     NA's :28179     NA's :29515     NA's :29283
## checkin_info.0.3 checkin_info.0.2 checkin_info.3.6 checkin_info.6.6
## Min. : 1        Min. : 1        Min. : 1        Min. : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 1      Median : 1      Median : 1
## Mean : 3        Mean : 3        Mean : 4        Mean : 2
## 3rd Qu.: 2      3rd Qu.: 2      3rd Qu.: 3      3rd Qu.: 2
## Max. :131       Max. :92        Max. :192       Max. :343
## NA's :29302     NA's :29493     NA's :30195     NA's :29406
## checkin_info.3.4 checkin_info.6.4 checkin_info.6.3 checkin_info.6.2
## Min. : 1        Min. : 1        Min. : 1        Min. : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 1      Median : 1      Median : 1
## Mean : 2        Mean : 3        Mean : 3        Mean : 3
## 3rd Qu.: 2      3rd Qu.: 3      3rd Qu.: 3      3rd Qu.: 2
## Max. :81        Max. :359       Max. :334       Max. :284
## NA's :30586     NA's :28625     NA's :28846     NA's :28819
## checkin_info.6.1 checkin_info.6.0 checkin_info.3.1 checkin_info.18.6
## Min. : 1        Min. : 1        Min. : 1        Min. : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 1      Median : 1      Median : 1      Median : 2
## Mean : 3        Mean : 3        Mean : 2        Mean : 5
## 3rd Qu.: 3      3rd Qu.: 2      3rd Qu.: 2      3rd Qu.: 5
## Max. :278       Max. :441       Max. :65        Max. :649
## NA's :28888     NA's :29005     NA's :30840     NA's :21589
## checkin_info.18.5 checkin_info.18.4 checkin_info.18.3 checkin_info.18.2
## Min. : 1        Min. : 1        Min. : 1        Min. : 1
## 1st Qu.: 1      1st Qu.: 1      1st Qu.: 1      1st Qu.: 1
## Median : 2      Median : 2      Median : 2      Median : 2
## Mean : 6        Mean : 6        Mean : 5        Mean : 5
## 3rd Qu.: 6      3rd Qu.: 6      3rd Qu.: 5      3rd Qu.: 5
## Max. :364       Max. :589       Max. :525       Max. :416
## NA's :19465     NA's :18769     NA's :19392     NA's :19649
## checkin_info.18.1 checkin_info.18.0
## Min. : 1        Min. : 1
## 1st Qu.: 1      1st Qu.: 1
## Median : 2      Median : 2
## Mean : 5        Mean : 5
## 3rd Qu.: 5      3rd Qu.: 5
## Max. :358       Max. :481
## NA's :19915     NA's :20412
```

```
head(checkin_data, 2)
```

```
## checkin_info.9.0 checkin_info.9.1 checkin_info.9.2 checkin_info.9.3
## 1 NA NA 1 NA
## 2 NA NA NA 1
## checkin_info.9.4 checkin_info.9.5 checkin_info.9.6 checkin_info.20.2
## 1 NA 1 1 NA
```


##	2	NA	NA	NA	NA
##		checkin_info.20.3	checkin_info.20.0	checkin_info.20.1	checkin_info.20.6
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.20.4	checkin_info.20.5	checkin_info.5.4	checkin_info.4.2
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		business_id	checkin_info.12.1	checkin_info.12.0	
##	1	JwUE5GmEO-sH1FuwJgKB1Q	NA	NA	
##	2	uGykseHzyS5xAMWoN6YUqA	NA	NA	
##		checkin_info.12.3	checkin_info.12.2	checkin_info.12.5	checkin_info.12.4
##	1	NA	NA	NA	NA
##	2	NA	NA	1	NA
##		checkin_info.12.6	checkin_info.15.6	checkin_info.15.4	checkin_info.15.5
##	1	1	NA	NA	NA
##	2	NA	1	1	1
##		checkin_info.15.2	checkin_info.15.3	checkin_info.15.0	checkin_info.15.1
##	1	NA	NA	NA	NA
##	2	NA	1	NA	NA
##		checkin_info.17.1	checkin_info.22.4	checkin_info.22.5	checkin_info.22.6
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.22.0	checkin_info.22.1	checkin_info.22.2	checkin_info.22.3
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.11.2	checkin_info.11.3	checkin_info.11.0	checkin_info.11.1
##	1	NA	1	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.11.6	checkin_info.11.4	checkin_info.11.5	checkin_info.16.5
##	1	2	NA	1	NA
##	2	NA	NA	NA	NA
##		checkin_info.16.4	checkin_info.16.6	checkin_info.16.1	checkin_info.16.0
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.16.3	checkin_info.16.2	checkin_info.1.0	checkin_info.1.1
##	1	NA	NA	NA	NA
##	2	1	NA	NA	NA
##		checkin_info.1.2	checkin_info.1.3	checkin_info.1.4	checkin_info.1.5
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.1.6	checkin_info.7.3	checkin_info.7.0	checkin_info.4.1
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.4.0	checkin_info.4.3	checkin_info.2.0	checkin_info.4.5
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.4.4	checkin_info.2.5	checkin_info.2.4	checkin_info.3.3
##	1	NA	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.7.4	checkin_info.5.3	checkin_info.19.2	checkin_info.19.3
##	1	1	NA	NA	NA
##	2	NA	NA	NA	NA
##		checkin_info.19.0	checkin_info.19.1	checkin_info.19.6	checkin_info.19.4
##	1	1	NA	NA	NA

## 2	NA	NA	NA	NA
##	checkin_info.19.5	checkin_info.2.3	checkin_info.2.2	checkin_info.2.1
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.7.2	checkin_info.5.5	checkin_info.5.6	checkin_info.7.1
## 1	NA	1	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.7.6	checkin_info.5.1	checkin_info.5.2	checkin_info.7.5
## 1	1	NA	1	1
## 2	1	NA	NA	1
##	checkin_info.3.2	checkin_info.2.6	checkin_info.21.6	checkin_info.21.5
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.21.4	checkin_info.21.3	checkin_info.21.2	checkin_info.21.1
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.21.0	checkin_info.8.5	checkin_info.8.4	checkin_info.8.6
## 1	NA	1	NA	NA
## 2	NA	2	NA	NA
##	type	checkin_info.8.0	checkin_info.8.3	checkin_info.8.2
## 1	checkin	NA	NA	NA
## 2	checkin	NA	NA	1
##	checkin_info.5.0	checkin_info.17.0	checkin_info.8.1	checkin_info.17.2
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.17.3	checkin_info.17.4	checkin_info.17.5	checkin_info.17.6
## 1	NA	NA	1	NA
## 2	NA	NA	NA	NA
##	checkin_info.14.3	checkin_info.14.2	checkin_info.14.1	checkin_info.14.0
## 1	1	1	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.14.6	checkin_info.14.5	checkin_info.14.4	checkin_info.4.6
## 1	1	1	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.13.4	checkin_info.13.5	checkin_info.13.6	checkin_info.13.0
## 1	NA	1	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.13.1	checkin_info.13.2	checkin_info.13.3	checkin_info.10.6
## 1	NA	1	NA	1
## 2	NA	NA	NA	NA
##	checkin_info.10.5	checkin_info.10.4	checkin_info.10.3	checkin_info.10.2
## 1	1	NA	NA	1
## 2	1	1	NA	NA
##	checkin_info.10.1	checkin_info.10.0	checkin_info.3.5	checkin_info.23.1
## 1	NA	NA	NA	NA
## 2	NA	1	NA	NA
##	checkin_info.23.0	checkin_info.23.3	checkin_info.23.2	checkin_info.23.5
## 1	NA	NA	NA	NA
## 2	NA	NA	NA	NA
##	checkin_info.23.4	checkin_info.23.6	checkin_info.6.5	checkin_info.3.0
## 1	NA	NA	1	NA
## 2	NA	NA	1	NA
##	checkin_info.0.5	checkin_info.0.4	checkin_info.0.6	checkin_info.0.1
## 1	NA	NA	NA	NA

```
## 2          NA          NA          NA          NA
##  checkin_info.0.0 checkin_info.0.3 checkin_info.0.2 checkin_info.3.6
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
##  checkin_info.6.6 checkin_info.3.4 checkin_info.6.4 checkin_info.6.3
## 1          1          1          NA          NA
## 2          2          NA          NA          NA
##  checkin_info.6.2 checkin_info.6.1 checkin_info.6.0 checkin_info.3.1
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
##  checkin_info.18.6 checkin_info.18.5 checkin_info.18.4 checkin_info.18.3
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
##  checkin_info.18.2 checkin_info.18.1 checkin_info.18.0
## 1          NA          NA          NA
## 2          NA          NA          NA
```

```
tail(checkin_data, 2)
```

```
##  checkin_info.9.0 checkin_info.9.1 checkin_info.9.2 checkin_info.9.3
## 31616          NA          NA          NA          NA
## 31617          NA          NA          NA          NA
##  checkin_info.9.4 checkin_info.9.5 checkin_info.9.6 checkin_info.20.2
## 31616          NA          NA          NA          NA
## 31617          NA          NA          NA          NA
##  checkin_info.20.3 checkin_info.20.0 checkin_info.20.1
## 31616          NA          NA          1
## 31617          1          NA          NA
##  checkin_info.20.6 checkin_info.20.4 checkin_info.20.5
## 31616          NA          NA          NA
## 31617          NA          NA          NA
##  checkin_info.5.4 checkin_info.4.2          business_id
## 31616          NA          NA nYer89hXYAoddMEKTxw7kA
## 31617          NA          NA BMjggIgOghBMEXPo8q7q3w
##  checkin_info.12.1 checkin_info.12.0 checkin_info.12.3
## 31616          NA          NA          NA
## 31617          NA          NA          NA
##  checkin_info.12.2 checkin_info.12.5 checkin_info.12.4
## 31616          NA          NA          NA
## 31617          NA          NA          NA
##  checkin_info.12.6 checkin_info.15.6 checkin_info.15.4
## 31616          NA          NA          NA
## 31617          NA          NA          NA
##  checkin_info.15.5 checkin_info.15.2 checkin_info.15.3
## 31616          NA          NA          NA
## 31617          NA          NA          NA
##  checkin_info.15.0 checkin_info.15.1 checkin_info.17.1
## 31616          NA          NA          16
## 31617          NA          1          NA
##  checkin_info.22.4 checkin_info.22.5 checkin_info.22.6
## 31616          NA          NA          NA
## 31617          NA          1          NA
##  checkin_info.22.0 checkin_info.22.1 checkin_info.22.2
## 31616          NA          NA          NA
```

##	31617	NA	NA	NA	
##		checkin_info.22.3	checkin_info.11.2	checkin_info.11.3	
##	31616	NA	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.11.0	checkin_info.11.1	checkin_info.11.6	
##	31616	NA	NA	NA	
##	31617	1	NA	NA	
##		checkin_info.11.4	checkin_info.11.5	checkin_info.16.5	
##	31616	NA	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.16.4	checkin_info.16.6	checkin_info.16.1	
##	31616	NA	NA	1	
##	31617	NA	NA	NA	
##		checkin_info.16.0	checkin_info.16.3	checkin_info.16.2	
##	31616	NA	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.1.0	checkin_info.1.1	checkin_info.1.2	checkin_info.1.3
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.1.4	checkin_info.1.5	checkin_info.1.6	checkin_info.7.3
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.7.0	checkin_info.4.1	checkin_info.4.0	checkin_info.4.3
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.2.0	checkin_info.4.5	checkin_info.4.4	checkin_info.2.5
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.2.4	checkin_info.3.3	checkin_info.7.4	checkin_info.5.3
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.19.2	checkin_info.19.3	checkin_info.19.0	
##	31616	1	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.19.1	checkin_info.19.6	checkin_info.19.4	
##	31616	2	NA	NA	
##	31617	NA	1	NA	
##		checkin_info.19.5	checkin_info.2.3	checkin_info.2.2	checkin_info.2.1
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.7.2	checkin_info.5.5	checkin_info.5.6	checkin_info.7.1
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.7.6	checkin_info.5.1	checkin_info.5.2	checkin_info.7.5
##	31616	NA	NA	NA	NA
##	31617	NA	NA	NA	NA
##		checkin_info.3.2	checkin_info.2.6	checkin_info.21.6	
##	31616	NA	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.21.5	checkin_info.21.4	checkin_info.21.3	
##	31616	NA	NA	NA	
##	31617	NA	NA	NA	
##		checkin_info.21.2	checkin_info.21.1	checkin_info.21.0	
##	31616	NA	NA	NA	

## 31617	NA	NA	NA	
##	checkin_info.8.5	checkin_info.8.4	checkin_info.8.6	type
## 31616	NA	NA	NA	checkin
## 31617	NA	NA	NA	checkin
##	checkin_info.8.0	checkin_info.8.3	checkin_info.8.2	checkin_info.5.0
## 31616	NA	NA	NA	NA
## 31617	NA	NA	NA	NA
##	checkin_info.17.0	checkin_info.8.1	checkin_info.17.2	
## 31616	NA	NA	NA	
## 31617	1	NA	NA	
##	checkin_info.17.3	checkin_info.17.4	checkin_info.17.5	
## 31616	NA	NA	NA	
## 31617	1	NA	NA	
##	checkin_info.17.6	checkin_info.14.3	checkin_info.14.2	
## 31616	NA	NA	NA	
## 31617	1	NA	NA	
##	checkin_info.14.1	checkin_info.14.0	checkin_info.14.6	
## 31616	NA	NA	NA	
## 31617	1	NA	NA	
##	checkin_info.14.5	checkin_info.14.4	checkin_info.4.6	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.13.4	checkin_info.13.5	checkin_info.13.6	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.13.0	checkin_info.13.1	checkin_info.13.2	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.13.3	checkin_info.10.6	checkin_info.10.5	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.10.4	checkin_info.10.3	checkin_info.10.2	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.10.1	checkin_info.10.0	checkin_info.3.5	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.23.1	checkin_info.23.0	checkin_info.23.3	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.23.2	checkin_info.23.5	checkin_info.23.4	
## 31616	NA	NA	NA	
## 31617	NA	NA	NA	
##	checkin_info.23.6	checkin_info.6.5	checkin_info.3.0	checkin_info.0.5
## 31616	NA	NA	NA	NA
## 31617	NA	NA	NA	NA
##	checkin_info.0.4	checkin_info.0.6	checkin_info.0.1	checkin_info.0.0
## 31616	NA	NA	NA	NA
## 31617	NA	NA	NA	NA
##	checkin_info.0.3	checkin_info.0.2	checkin_info.3.6	checkin_info.6.6
## 31616	NA	1	NA	NA
## 31617	NA	NA	NA	NA
##	checkin_info.3.4	checkin_info.6.4	checkin_info.6.3	checkin_info.6.2
## 31616	NA	NA	NA	NA

```
## 31617      NA      NA      NA      NA
##      checkin_info.6.1 checkin_info.6.0 checkin_info.3.1 checkin_info.18.6
## 31616      NA      NA      NA      NA
## 31617      NA      NA      NA      NA
##      checkin_info.18.5 checkin_info.18.4 checkin_info.18.3
## 31616      NA      NA      NA
## 31617      NA      NA      NA
##      checkin_info.18.2 checkin_info.18.1 checkin_info.18.0
## 31616      NA      NA      NA
## 31617      1      NA      NA
```

```
# Convert all na values to 0
n <- ncol(checkin_data)
for(i in 1:n){
  if(colnames(checkin_data) != "business_id" || colnames(checkin_data) != "type" )
    checkin_data[is.na(checkin_data[i]), i] <- 0
}

daily <- seq(from = as.POSIXlt("2014-12-06 00:00:00"),
             to = as.POSIXlt("2014-12-12 23:00:00"), by = "+1 hour")
hours <- as.numeric(format(daily, format("%H")))
total_checkin <- rep( 0, length(daily))

for(i in 1:length(daily)){

  day <- daily[i]
  #day_weekday <- ifelse(wday(day) == 7, 0, wday(day))
  day_weekday <- wday(day) - 1

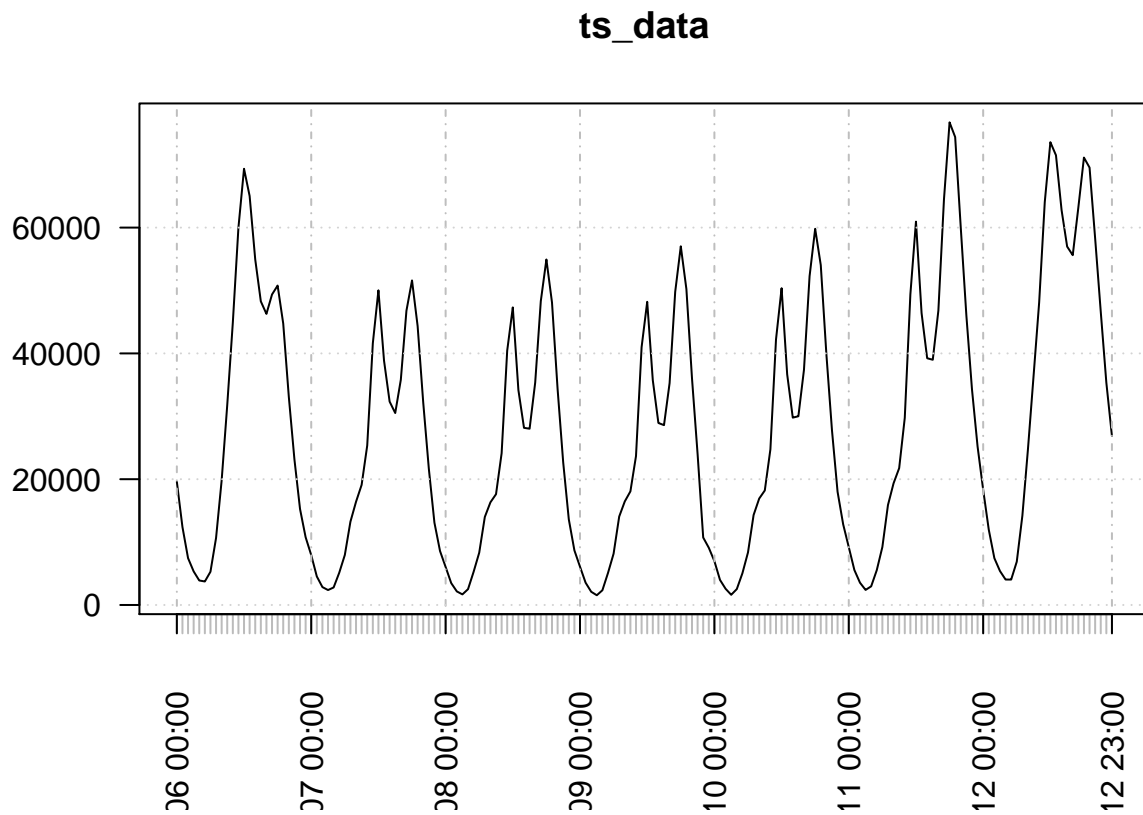
  col_name <- paste("checkin_info.", hours[i], ".", day_weekday, sep = "")
  #print(col_name)
  total_checkin[i] <- sum(checkin_data[,which(colnames(checkin_data) == col_name)])
}

checkin_day_hour <- data.frame(daily, total_checkin)

# Time series and forecast

# Create hourly data time series
ts_data <- xts(checkin_day_hour$total_checkin, order = checkin_day_hour$daily)

# checkin hourly time series plot
plot(ts_data, las = 2)
```

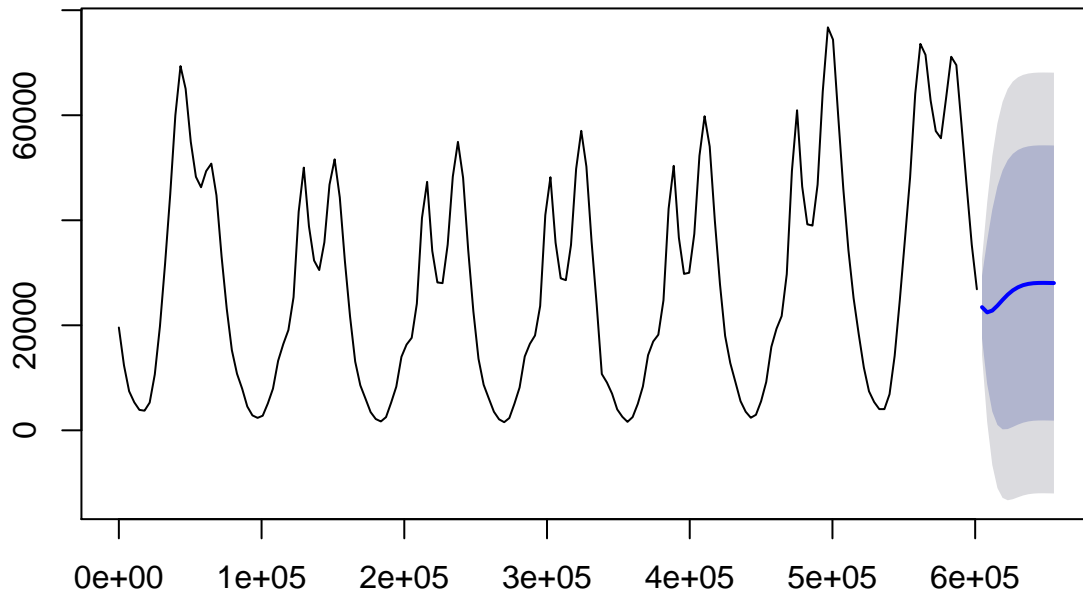


```
summary(ts_data)
```

```
##      Index                ts_data
##  Min.   :2014-12-06 00:00:00  Min.   : 1549
## 1st Qu.:2014-12-07 17:45:00 1st Qu.: 8350
## Median :2014-12-09 11:30:00 Median :24779
## Mean   :2014-12-09 11:30:00 Mean   :28058
## 3rd Qu.:2014-12-11 05:15:00 3rd Qu.:46318
## Max.   :2014-12-12 23:00:00 Max.   :76734
```

```
# Auto arima and forecast checkins for next 15 hours
fit_ts <- auto.arima(ts_data)
hourly_forecast <- forecast(fit_ts, 15)
plot(forecast(fit_ts, 15))
```

Forecasts from ARIMA(2,0,1) with non-zero mean



hourly_forecast

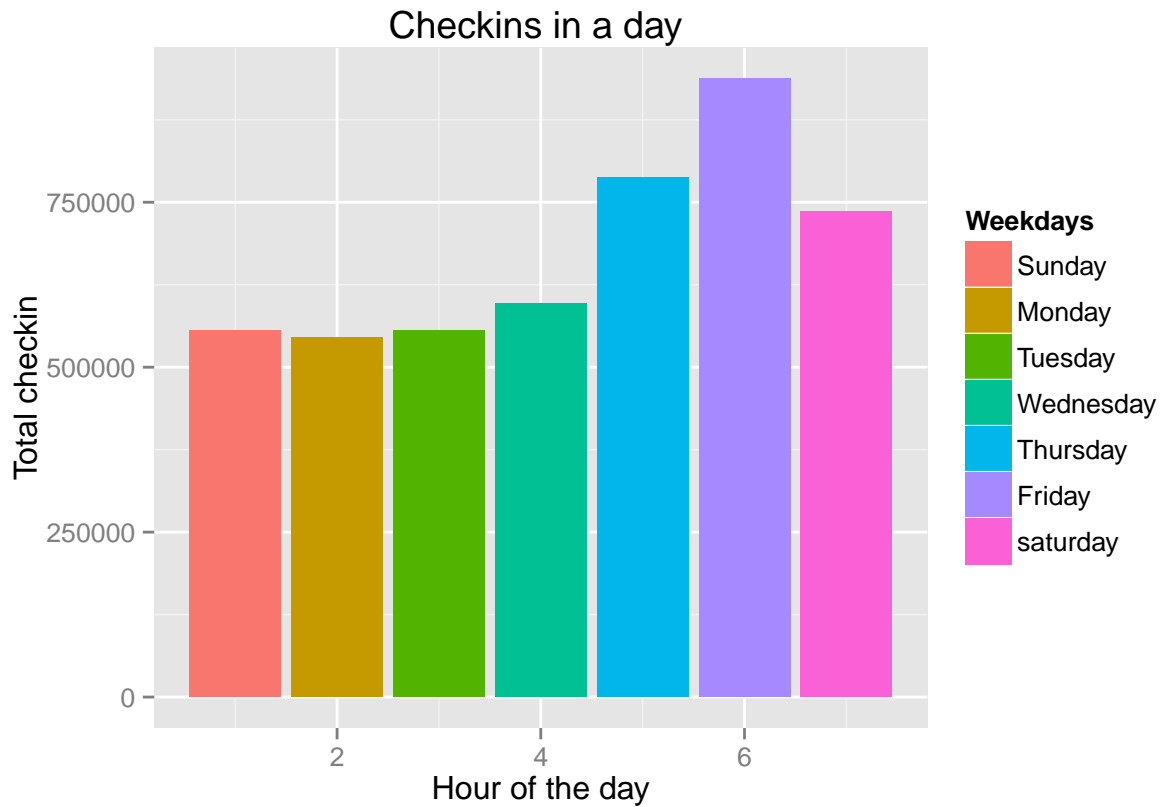
##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## 604801	23457	17488.5	29425	14329	32584
## 608401	22464	8812.0	36117	1585	43344
## 612001	22803	3585.6	42021	-6588	52194
## 615601	23744	1032.8	46455	-10990	58478
## 619201	24838	187.1	49489	-12862	62539
## 622801	25843	241.8	51444	-13311	64997
## 626401	26653	648.9	52657	-13117	66423
## 630001	27246	1099.5	53392	-12741	67232
## 633601	27643	1458.5	53827	-12403	67688
## 637201	27884	1693.9	54074	-12170	67938
## 640801	28013	1822.7	54203	-12041	68067
## 644401	28067	1876.2	54258	-11988	68122
## 648001	28077	1884.1	54269	-11981	68134
## 651601	28063	1868.6	54257	-11998	68123
## 655201	28040	1844.2	54235	-12023	68102

Checkins each day graph

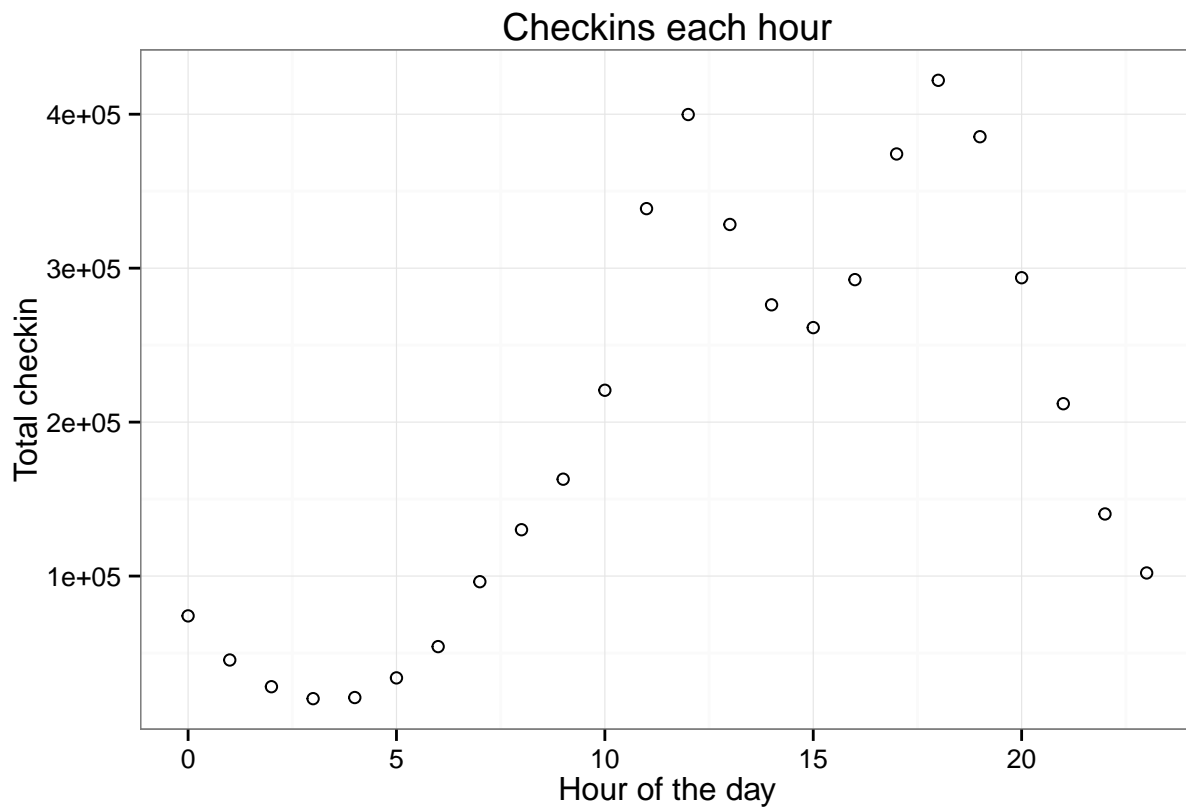
```
checkin_day_hour %>%
  group_by( weekday=yday(daily)) %>%
  summarise( checkins = sum(total_checkin)) %>%
  ggplot(aes(x= weekday, y = checkins, fill=factor(weekday))) +
  geom_bar(stat="identity") +
```



```
labs(title = "Checkins in a day ", x = "Hour of the day", y = "Total checkin") +
scale_fill_discrete(name = "Weekdays",
  labels = c("Sunday", "Monday", "Tuesday", "Wednesday",
    "Thursday", "Friday", "saturday"))
```



```
# Daily hour graph
checkin_day_hour %>%
  group_by( eachhour=hour(daily)) %>%
  summarise( checkins = sum(total_checkin)) %>%
  ggplot(aes(x= eachhour, y = checkins)) +
  geom_point(shape = 1) +
  labs(title = "Checkins each hour", x = "Hour of the day", y = "Total checkin") +
  theme_bw()
```



```
# peek hours of the day
checkin_day_hour %>%
  group_by( weekday=wday(daily), eachhour=hour(daily)) %>%
  summarise( checkins = sum(total_checkin)) %>%
  ggplot(aes(x= eachhour, y = checkins,
    colour = factor(weekday))) +
  geom_point(shape = 1) +
  geom_line() +
  labs(title = "Peak hours of a day", x = "Hour of the day", y = "Total checkin") +
  scale_color_discrete(name = "Weekday number",
    labels = c("Sunday", "Monday", "Tuesday", "Wednesday",
      "Thursday", "Friday", "saturday"))
```

