Oesmn Project(Final_Report)

Rajat Singh Saluja

12/5/2016

```
# Submit yelp access keys
consumerKey = "eKmeCNhEchKYe1RsEFojpQ"
consumerSecret = "pA6TMmRET0PYGaIIhe5I507Q5gA"
token = "JkiJNBUHsyXbTidqMOWoRz3n2W8ugzXF"
token_secret = "jd5z_y6BBXECoG8OGp_E2GoVf0o"

#Import Required Packages

require(httr)

## Loading required package: httr

require(httpuv)

## Loading required package: httpuv

require(jsonlite)

## Loading required package: jsonlite

require(ggplot2)

## Loading required package: ggplot2

# authorization
myapp = oauth_app("YELP", key=consumerKey, secret=consumerSecret)
sig=sign_oauth1.0(myapp, token=token,token_secret=token_secret)

##############################################################################
##############################################################################
##############################################################################

#Data extraction for first 40 restraunts in Chicago
offsetx = 0
# 40 restraunts in Chicago

yelpurl <-
paste0("https://api.yelp.com/v2/search/?offset=",offsetx,"&term=reservation&l
ocation=Chicago,%20IL&limit=40&actionlinks=true&category_filter=restaurants")

locationdata=GET(yelpurl, sig)
locationdataContent = content(locationdata)
locationdataList=jsonlite::fromJSON(toJSON(locationdataContent))
```

```
a = data.frame(locationdataList)

#Remove the row names
rownames(a) <- NULL

#data_1 has 3 values Ratings/Buinesses Name/ Reservation
data_1<- a[c(7, 11, 26)]

##############################################################################
##############################################################################
##############################################################################


#Data extraction for 41-80 restraunts in Chicago
offsetx = 41
# 41-80 restraunts in Chicago

yelpurl <-
paste0("https://api.yelp.com/v2/search/?offset=",offsetx,"&term=reservation&l
ocation=Chicago,%20IL&limit=40&actionlinks=true&category_filter=restaurants")

locationdata=GET(yelpurl, sig)
locationdataContent = content(locationdata)
locationdataList=jsonlite::fromJSON(toJSON(locationdataContent))
b = data.frame(locationdataList)

#Remove the row names
rownames(b) <- NULL

#data_2 has 3 values Ratings/Buinesses Name/ Reservation
data_2<- b[c(7, 11, 27)]

##############################################################################
##############################################################################
##############################################################################

#Data extraction for 81-120 restraunts in Chicago
offsetx = 81
# 41-80 restraunts in Chicago

yelpurl <-
paste0("https://api.yelp.com/v2/search/?offset=",offsetx,"&term=reservation&l
ocation=Chicago,%20IL&limit=40&actionlinks=true&category_filter=restaurants")

locationdata=GET(yelpurl, sig)
locationdataContent = content(locationdata)
locationdataList=jsonlite::fromJSON(toJSON(locationdataContent))
c = data.frame(locationdataList)
```

```r
#Remove the row names
rownames(c) <- NULL

#data_3 has 3 values Ratings/Buinesses Name/ Reservation
data_3<- c[c(7, 11, 28)]

##############################################################################
##############################################################################
##############################################################################

#Data extraction for 121-160 restraunts in Chicago
offsetx = 121
# 121-160 restraunts in Chicago

yelpurl <-
paste0("https://api.yelp.com/v2/search/?offset=",offsetx,"&term=reservation&l
ocation=Chicago,%20IL&limit=40&actionlinks=true&category_filter=restaurants")

locationdata=GET(yelpurl, sig)
locationdataContent = content(locationdata)
locationdataList=jsonlite::fromJSON(toJSON(locationdataContent))
d = data.frame(locationdataList)

#Remove the row names
rownames(d) <- NULL

#data_4 has 3 values Ratings/Buinesses Name/ Reservation
data_4<- d[c(7, 11, 27)]

##############################################################################
##############################################################################
##############################################################################

#Data extraction for 161-200 restraunts in Chicago
offsetx = 161
# 161-200 restraunts in Chicago

yelpurl <-
paste0("https://api.yelp.com/v2/search/?offset=",offsetx,"&term=reservation&l
ocation=Chicago,%20IL&limit=40&actionlinks=true&category_filter=restaurants")

locationdata=GET(yelpurl, sig)
locationdataContent = content(locationdata)
locationdataList=jsonlite::fromJSON(toJSON(locationdataContent))
e = data.frame(locationdataList)
#Remove the row names
rownames(e) <- NULL

#data_5 has 3 values Ratings/Buinesses Name/ Reservation
```

```r
data_5<- e[c(7, 11, 27)]

##############################################################################
##############################################################################
##############################################################################

###Bind all the data sets into one set (dataN) of 200 observations.

dataN<- rbind(data_1,data_2,data_3,data_4,data_5)

### Conversion of NULL values in " businesses.reservation_url " to NA.

dataN$businesses.reservation_url[sapply(dataN$businesses.reservation_url,
is.null)] <- NA

### Make new dataset(newdataN) & remove rows with NA Values in the column
"businesses.reservation_url" [45 Data Observations]
newdataN <- dataN[!is.na(dataN[,3]),]
Yelpdata_WITH_Reservation<- newdataN

#Intialization of all reservation based restraunts as "With Online
Reservation" by adding another column
Yelpdata_WITH_Reservation$index<- seq.int(nrow(newdataN))

Yelpdata_WITH_Reservation$index <- "With Online Reservation"

#Convert rating into numeric for analysis

newrating<- as.numeric(unlist(Yelpdata_WITH_Reservation$businesses.rating))

#Adding numeric Rating values in the DATA SET

Yelpdata_WITH_Reservation$NumericRating<-
seq.int(nrow(Yelpdata_WITH_Reservation))
Yelpdata_WITH_Reservation$NumericRating<- newrating




### Make new dataset(newdataN) WITHOUT Online Reservation(NA Values)
restaurantso in the column "businesses.reservation_url"
Yelpdata_NO_Reservation<- dataN[-c(2, 4:11, 13, 15, 17, 19, 22, 28, 31:32,
37, 44:46, 49, 66, 68, 75, 77, 91, 95, 98, 103, 108, 127, 128, 131, 141, 148,
153, 165, 169, 176, 179, 181, 183, 189, 199),]

#Dropping data observations to create an equal dataset of 45 Observations
Yelpdata_NO_Reservation$Sno<- seq.int(nrow(Yelpdata_NO_Reservation))
Yelpdata_NO_Reservation<- Yelpdata_NO_Reservation[-c(46:155),]
```

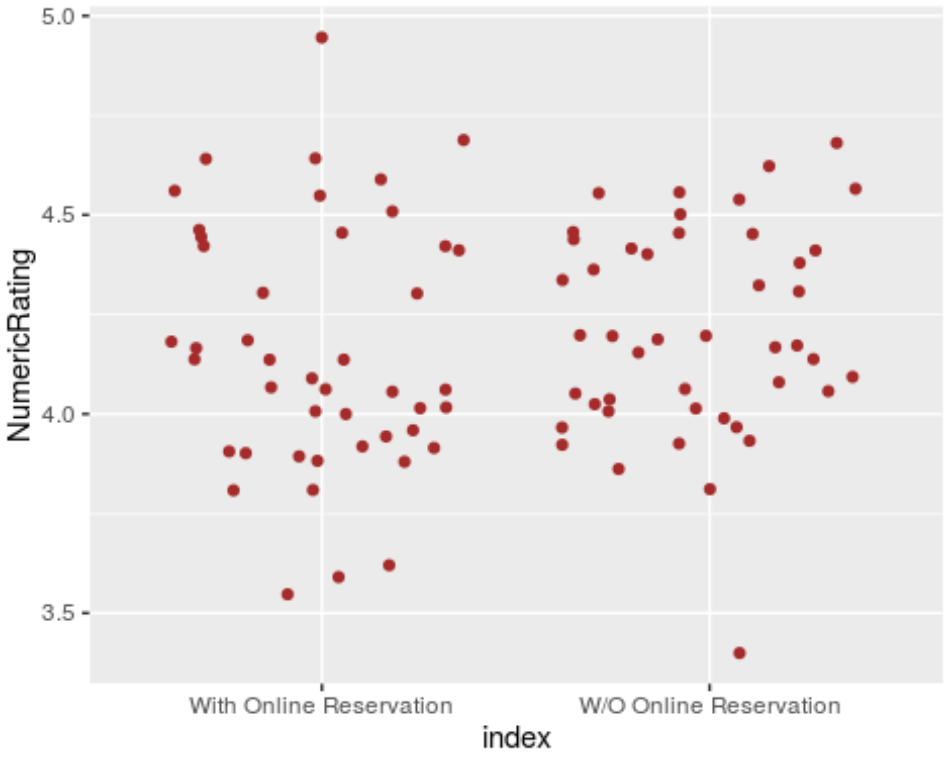
```r
#Intialization of NON reservation based restraunts as "W/O Online
Reservation" by adding another column
Yelpdata_NO_Reservation$index<- seq.int(nrow(Yelpdata_NO_Reservation))
Yelpdata_NO_Reservation$index<- "W/O Online Reservation"

#Again Delete "Yelpdata_NO_Reservation$Sno"
Yelpdata_NO_Reservation<- Yelpdata_NO_Reservation[c(1:3, 5)]

#Convert rating into numeric for analysis
newratingNO<- as.numeric(unlist(Yelpdata_NO_Reservation$businesses.rating))


#Adding numeric Rating values in the DATA SET
Yelpdata_NO_Reservation$NumericRating<-
seq.int(nrow(Yelpdata_NO_Reservation))
Yelpdata_NO_Reservation$NumericRating<- newratingNO


#Combine Yelpdata WITH & WITHOUT Reservations
Yelpdata<- rbind(Yelpdata_WITH_Reservation, Yelpdata_NO_Reservation)

##############################################################################
#################################T - Test#####################################
##############################################################################

t.test(newrating,newratingNO,var.equal = T,paired = F)

##
##  Two Sample t-test
##
## data:  newrating and newratingNO
## t = -0.71685, df = 88, p-value = 0.4754
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.16765634  0.07876746
## sample estimates:
## mean of x mean of y
##  4.155556  4.200000

#P value is more than 0.5 therefore their is no significant difference
between mean values of Ratings in both
#the types of restraunts i.e. with and without Reservations & NULL hypothesis
is correct

##############################################################################
#################################Wilcox - Test################################
##############################################################################
```

```r
wilcox.test(newrating,newratingNO, correct = F)
```

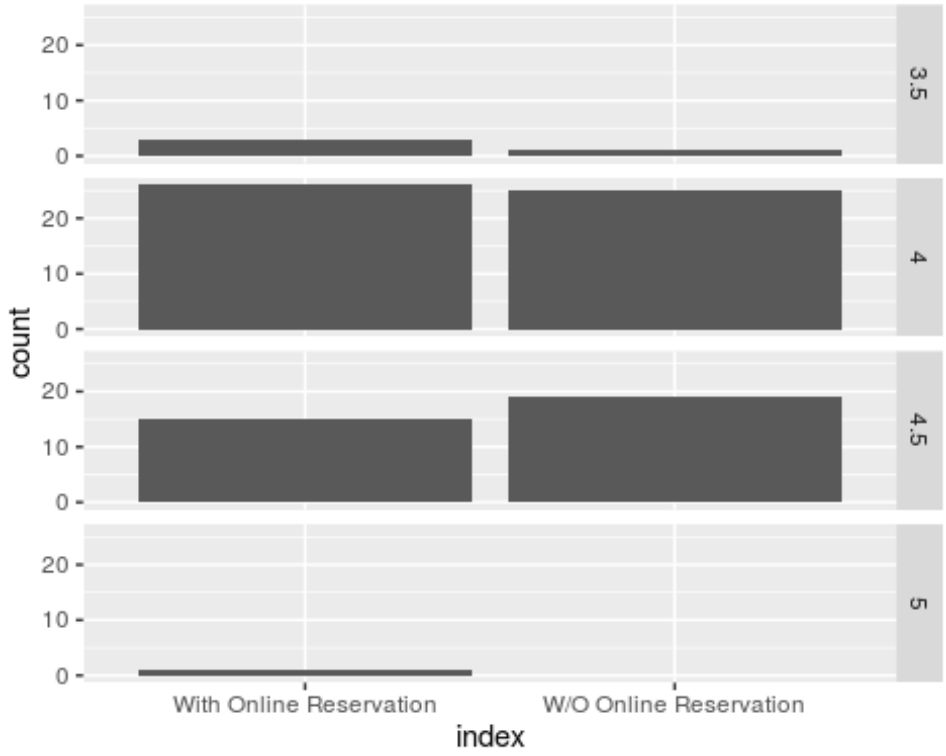
```
## Warning in wilcox.test.default(newrating, newratingNO, correct = F):
cannot
## compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test
##
## data:  newrating and newratingNO
## W = 930, p-value = 0.4463
## alternative hypothesis: true location shift is not equal to 0
```

```r
#Again, P value is here more than 0.5 therefore their is no significant
difference between mean values of Ratings in both
#the types of restraunts i.e. with and without Reservations & NULL hypothesis
is correct.


##############################################################################
##############################################################################
##############################################################################




##################################Scatter Plot#################################

ggplot(Yelpdata, aes(index, NumericRating) ) +  geom_jitter(colour="Brown")
```




```
##################################BAR GRAPH#######################################

ggplot(Yelpdata, aes(index) ) +  geom_bar() + facet_grid(NumericRating ~ .)
```



```
###################################LINE GRAPH#####################################

ggplot(Yelpdata, aes(x= index, y= NumericRating))+geom_line(colour="purple",
linetype= "dotted", size=3)+geom_point()
```