Notes during analysis:   
  
In R,

Outliers exist in the data, however for the purposes of this project, I find that by aggregating the data, the need to clean the data and remove outliers is unnecessary. I could automate outlier removal, but I don’t want to remove data points that may be accurate and influential in the time series analysis.

While I have daily data, I don’t have enough data points to perform a daily time series analysis. Thus, I need to transform the dates into monthly dates for monthly time series analysis.

ANOVA- there is a statistically significant difference in means of holidays vs nonholidays… further exploration to see the strength and directionality of this difference. \*\*\* When total visitors for all restaurants, the ANOVA does not show a significant difference in average visitors between holidays and non holidays

Other types of time series besides ARIMA?

Decompose non-seasonally using a SMA

How to forecast using these results?

Accuracy of these results?

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_PREPARE DATA\_\_\_\_\_\_\_\_\_\_\_\_

#set working directory

setwd("C:/Users/Rachel/Documents/Grad School/MIS581/Portfolio Project")

#Import libraries

install.packages("TTR")

library(TTR)

library(ggplot2)

library(dplyr)

library(forecast)

library(lubridate)

#Load data

visit\_data<-read.csv("~/Grad School/MIS581/Portfolio Project/air\_visit\_data.csv")

date\_data<-read.csv("~/Grad School/MIS581/Portfolio Project/date\_info.csv")

#Clean data

combined\_data <- merge(visit\_data, date\_data, by.x = "visit\_date", by.y = "calendar\_date")

combined\_data$visit\_date <- as.Date(combined\_data$visit\_date)

combined\_data$holiday\_flg <- as.factor(combined\_data$holiday\_flg)

str(combined\_data)

#aggregate

daily\_visitors <- combined\_data %>%

group\_by(visit\_date) %>%

summarise(Total\_Visitors = sum(visitors))

#Time Series

ts\_data<- ts(daily\_visitors$Total\_Visitors, frequency = 365, start = c(2016, 1))

plot.ts(ts\_data)

ggseasonplot(ts\_data) +

labs(title = "Seasonal Plot: Total Visitors", x ="Month", y="Total Visitors")

aov(visitors ~ factor(holiday\_flg))

visitorstsSMA10 <-SMA(ts\_data, n=10)

plot.ts(visitorstsSMA10)











