

DATA 624: PREDICTIVE ANALYTICS Project 1

Gabriel Campos

Last edited March 23, 2024

```
library(fpp3)
library(dplyr)
library(ggplot2)
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```
library(tsibble)
library(psych)
library(tidyr)
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.3.3
```

Description

This project consists of 3 parts - two required and one bonus and is worth 15% of your grade. The project is due at 11:59 PM on Sunday Apr 11. I will accept late submissions with a penalty until the meetup after that when we review some projects.

Part A

ATM Forecast [ATM624Data.xlsx](#)

In part A, I want you to forecast how much cash is taken out of 4 different ATM machines for May 2010. The data is given in a single file. The variable 'Cash' is provided in hundreds of dollars, other than that it is straight forward. I am being somewhat ambiguous on purpose to make this have a little more business feeling. Explain and demonstrate your process, techniques used and not used, and your actual forecast. I am giving you data via an excel file, please provide your written report on your findings, visuals, discussion and your R code via an RPub link along with the actual.rmd file Also please submit the forecast which you will put in an Excel readable file.

Part B

Forecasting Power [ResidentialCustomerForecastLoad-624.xlsx](#)

Part B consists of a simple dataset of residential power usage for January 1998 until December 2013. Your assignment is to model these data and a monthly forecast for 2014. The data is given in a single file. The variable 'KWH' is power consumption in Kilowatt hours, the rest is straight forward. Add this to your existing files above.

Part C

BONUS, optional (part or all), [Waterflow_Pipe1.xlsx](#) and [Waterflow_Pipe2.xlsx](#)

Part C consists of two data sets. These are simple 2 columns sets, however they have different time stamps. Your optional assignment is to time-base sequence the data and aggregate based on hour (example of what this looks like, follows). Note for multiple recordings within an hour, take the mean. Then to determine if the data is stationary and can it be forecast. If so, provide a week forward forecast and present results via Rpubs and .rmd and the forecast in an Excel readable file.

Data Load

https://github.com/GitableGabe/Data624_Data/raw/main/ATM624Data.xlsx

```
atm_coltype<-c("date","text","numeric")

atm_import<-read_xlsx('ATM624Data.xlsx', col_types = atm_coltype)
# Ommitting Extra Credit as I won't be working on it
# WP1_df<-read_xlsx('Waterflow_Pipe1.xlsx')
# WP2_df<-read_xlsx('Waterflow_Pipe2.xlsx')

power_raw<-read_xlsx('ResidentialCustomerForecastLoad-624.xlsx')
```

Part A

EDA & Cleanup

```
head(atm_import%>%
  filter(ATM=="ATM4"))
```

```
## # A tibble: 6 x 3
##   DATE                ATM    Cash
##   <dtm>              <chr> <dbl>
## 1 2009-05-01 00:00:00 ATM4  777.
## 2 2009-05-02 00:00:00 ATM4  524.
## 3 2009-05-03 00:00:00 ATM4  793.
## 4 2009-05-04 00:00:00 ATM4  908.
## 5 2009-05-05 00:00:00 ATM4   52.8
## 6 2009-05-06 00:00:00 ATM4   52.2
```

```
atm_range<-range(atm_import$DATE)
atm_range[1]
```

```
## [1] "2009-05-01 UTC"
```

```
atm_range[2]
```

```
## [1] "2010-05-14 UTC"
```