

```
In [1]: #library lubridate for manipulate date
#library ggplot2 for plot
#library dplyr for manipulate data
#library plyr for manipulate data
#library forecast for forecaste
#library plotly for plot
#library glmnet for multiple regression
library("lubridate")
library("ggplot2")
library("dplyr")
library("plyr")
library("plotly")
library("mlbench")
library("glmnet")
library("stringr")

Attaching package: 'lubridate'
```

```
The following object is masked from 'package:base':

  date

Registered S3 methods overwritten by 'ggplot2':
  method      from
[.quosures    rlang
c.quosures    rlang
print.quosures rlang

Attaching package: 'dplyr'

The following objects are masked from 'package:lubridate':

  intersect, setdiff, union

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

-----
You have loaded plyr after dplyr - this is likely to cause problems.
If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
library(plyr); library(dplyr)
-----

Attaching package: 'plyr'

The following objects are masked from 'package:dplyr':

  arrange, count, desc, failwith, id, mutate, rename, summarise,
  summarize

The following object is masked from 'package:lubridate':

  here

Attaching package: 'plotly'

The following objects are masked from 'package:plyr':

  arrange, mutate, rename, summarise

The following object is masked from 'package:ggplot2':

  last_plot

The following object is masked from 'package:stats':

  filter

The following object is masked from 'package:graphics':

  layout

Loading required package: Matrix
Loading required package: Foreach
Loaded glmnet 2.0-16
```

```
In [2]: #Load data
xs<- "C:/Users/jessi/OneDrive/Desktop/ALY 6070/Final Project/D24 MAB FY19P1to7.csv"
df <- read.csv(file = xs)
df
```

MBWHSE	MBREG_NUM	MBTRN_NUM	FISCAL_PERIOD	MBDATE	MBTIME	MBITEM	MBDSUN	MBDSEL	MBRESL	...	IMDES2	REGION
423	12	398	2	10/26/2018 0:00	1721	1234449	1	549.99	...	I5-8250U/12GB/1TB	BA	
770	2	108	2	10/26/2018 0:00	1226	1235362	1	469.99	...	I3-8130U/12GB/1TB	NW	
770	2	209	2	10/26/2018 0:00	1442	1238985	1	899.99	...	I7-8550U/16GB/1TB	NW	
770	51	52	2	10/26/2018 0:00	1458	1238985	-1	-899.99	...	I7-8550U/16GB/1TB	NW	
29	9	482	2	10/26/2018 0:00	1908	1275609	1	799.99	...	RYZEN 5/8GB/256SSD	BA	
1070	4	410	2	10/10/2018 0:00	1827	1235362	1	469.99	...	I3-8130U/12GB/1TB	NW	
113	8	170	2	10/10/2018 0:00	1341	1234449	1	549.99	...	I5-8250U/12GB/1TB	NE	
126	51	40	2	10/10/2018 0:00	1321	1181376	1	449.97	...	I5-8250U/12GB/1TB P=87	LA	
126	82	88	2	10/10/2018 0:00	1317	1235362	-1	-469.99	...	I3-8130U/12GB/1TB	LA	
466	11	79	2	10/10/2018 0:00	1152	1235362	1	469.99	...	I3-8130U/12GB/1TB	SD	
29	5	44	2	10/10/2018 0:00	1141	1234449	1	549.99	...	I5-8250U/12GB/1TB	BA	
341	82	60	6	2/3/2019 0:00	1349	1222491	-1	-1499.99	...	I7-8550U/16GB/512SSD P=48	MW	
241	701	64	6	2/3/2019 0:00	1144	1157726	-1	-399.99	...	I3-7100U/12GB/1TB P=87	NE	
130	10	60	6	2/3/2019 0:00	1124	1235362	1	469.99	...	I3-8130U/12GB/1TB	LA	
130	51	33	6	2/3/2019 0:00	1129	1235362	-1	-469.99	...	I3-8130U/12GB/1TB	LA	
423	701	68	6	2/3/2019 0:00	1442	1275609	-1	-799.99	...	RYZEN 5/8GB/256SSD	BA	
130	3	113	6	2/3/2019 0:00	1324	1238985	1	899.99	...	I7-8550U/16GB/1TB	LA	
130	7	11	6	2/3/2019 0:00	1024	1241016	1	1299.99	...	I7-850U/16GB/1TB+256SSD	LA	
130	81	6	6	2/3/2019 0:00	957	1241016	-1	-1299.99	...	I7-850U/16GB/1TB+256SSD	LA	
670	13	180	6	2/3/2019 0:00	1411	1235362	1	469.99	...	I3-8130U/12GB/1TB	NW	
659	12	236	6	2/3/2019 0:00	1555	1222491	1	1499.99	...	I7-8550U/16GB/512SSD P=48	BA	
1059	9	7	6	2/3/2019 0:00	1447	1234449	1	479.97	...	I5-8250U/12GB/1TB	NW	
1	5	8	6	2/3/2019 0:00	1011	1222491	1	1499.99	...	I7-8550U/16GB/512SSD P=48	NW	
333	204	98	6	2/3/2019 0:00	1404	1228070	1	999.99	...	I7-8550U/16GB/512SSD	NE	
1163	5	65	6	2/3/2019 0:00	1044	1234449	1	479.97	...	I5-8250U/12GB/1TB	TE	
1163	82	5	6	2/3/2019 0:00	1039	1234449	-1	-479.97	...	I5-8250U/12GB/1TB	TE	
1030	2	419	6	2/3/2019 0:00	1559	1272966	1	399.99	...	I3-8130U/4GB/128SSD	SD	
1030	6	341	6	2/3/2019 0:00	1532	1228011	1	699.99	...	I5-8250U/12GB/1TB	SD	
766	12	308	6	2/3/2019 0:00	1658	1234449	1	479.97	...	I5-8250U/12GB/1TB	NW	
766	10	35	6	2/3/2019 0:00	1025	1234449	1	479.97	...	I5-8250U/12GB/1TB	NW	
...	...	...	...	...	...	...	...	...	...	...	...	...
738	6	5	3	11/19/2018 0:00	1026	1272966	1	399.99	...	I3-8130U/4GB/128SSD	SD	
738	5	45	6	1/25/2019 0:00	1056	1234449	1	479.97	...	I5-8250U/12GB/1TB	SD	
69	10	3	4	12/2/2018 0:00	1013	1235362	1	469.99	...	I3-8130U/12GB/1TB	NW	
642	10	300	4	12/7/2018 0:00	1620	1272966	1	399.99	...	I3-8130U/4GB/128SSD	NW	
642	10	300	4	12/7/2018 0:00	1620	1272966	1	399.99	...	I3-8130U/4GB/128SSD	NW	
847	999	17487	4	11/29/2018 0:00	102	1278582	1	299.99	N ...	4/64/FHD/WN10S	EC	
847	999	16723	4	11/29/2018 0:00	102	1278524	1	549.99	N ...	12/1/HD+7TOUCH	EC	
122	82	2	2	10/14/2018 0:00	951	1234449	-1	-549.99	...	I5-8250U/12GB/1TB	LA	
145	82	17	2	10/22/2018 0:00	1027	1234449	-1	-549.99	...	I5-8250U/12GB/1TB	NW	
784	51	23	5	11/9/2019 0:00	1324	1228070	1	859.99	...	I7-8550U/16GB/512SSD	MW	
1089	10	202	6	21/2/2019 0:00	1352	1235362	1	469.99	...	I3-8130U/12GB/1TB	NE	
358	9	325	1	9/7/2018 0:00	1844	1222491	1	1499.99	...	I7-8550U/16GB/512SSD P=48	SE	
1011	12	303	3	11/11/2018 0:00	1609	1230231	1	699.99	...	I7-8550U/12GB/1TB	BA	
660	3	168	1	9/8/2018 0:00	1200	1234449	1	549.99	...	I5-8250U/12GB/1TB	NW	
631	4	15	5	1/17/2019 0:00	1210	1272966	1	399.99	...	I3-8130U/4GB/128SSD	SE	
784	82	30	5	11/9/2019 0:00	1321	1098376	-1	-849.99	...	1/4G R7/M445/O365P/FHD	MW	
230	11	182	6	21/3/2019 0:00	1854	1234449	1	479.97	...	I5-8250U/12GB/1TB	NE	
13	10	99	7	3/4/2019 0:00	1342	1272966	1	399.99	...	I3-8130U/4GB/128SSD	NW	
341	5	31	6	2/3/2019 0:00	1042	1272966	1	399.99	...	I3-8130U/4GB/128SSD	MW	
660	8	238	6	2/6/2019 0:00	1607	1234449	1	479.97	...	I5-8250U/12GB/1TB	NW	
670	701	51	6	2/7/2019 0:00	1208	1228011	-1	-699.99	...	I5-8250U/12GB/1TB	NW	
670	701	88	6	2/7/2019 0:00	1327	1228011	-1	-699.99	...	I5-8250U/12GB/1TB	NW	
670	16	126	6	2/7/2019 0:00	1339	1230231	1	699.99	...	I7-8550U/12GB/1TB	NW	
1089	10	131	6	2/6/2019 0:00	1411	1235362	1	469.99	...	I3-8130U/12GB/1TB	NE	
1029	13	102	6	21/2/2019 0:00	1403	1234449	1	479.97	...	I5-8250U/12GB/1TB	NW	
674	1	23	3	11/19/2018 0:00	1053	1272966	1	399.99	...	I3-8130U/4GB/128SSD	SD	
144	82	106	2	10/27/2018 0:00	1639	1230231	-1	-699.99	...	I7-8550U/12GB/1TB	BA	
847	999	83151	3	11/5/2018 0:00	30	1275609	1	799.99	N ...	RYZEN 5/8GB/256SSD	EC	
847	999	32163	1	9/9/2018 0:00	6	1231375	1	549.99	N ...	/FHD/BLACK	EC	
110	13	78	3	11/24/2018 0:00	1224	1234449	1	549.99	...	I5-8250U/12GB/1TB	NW	

```
In [3]: str(df)

'data.frame':   832766 obs. of  26 variables:
 $ MBWHSE      : int  423 770 770 770 29 1070 113 126 126 466 ...
 $ MBREG_NUM   : int  12 2 2 51 9 4 8 51 82 11 ...
 $ MBTRN_NUM   : int  398 188 209 52 482 419 170 49 88 79 ...
 $ FISCAL_PERIOD: int  2 2 2 2 2 2 2 2 2 ...
 $ MBDATE      : Factor w/ 196 levels "1/1/2019 0:00",...: 59 50 50 50 33 33 33 33 33 ...
 $ MBTIME      : int  1721 1226 1442 1458 1908 1827 1341 1321 1317 1152 ...
 $ MBITEM      : int  1234449 1235362 1238985 1275609 1235362 1234449 1181376 1235362 1235362 ...
 $ MBDSUN      : int  1 1 1 -1 1 1 1 1 -1 1 ...
 $ MBDSEL      : num  550 470 980 -980 808 ...
 $ MBRESL      : Factor w/ 3 levels "N", "M", "Y": 1 1 1 1 1 1 1 1 1 ...
 $ MBRFND      : Factor w/ 2 levels "N", "Y": 1 1 1 1 1 1 1 1 2 ...
 $ IMCAT1      : Factor w/ 1 level "M": 1 1 1 1 1 1 1 1 1 ...
 $ IMCAT2      : Factor w/ 1 level "A": 1 1 1 1 1 1 1 1 1 ...
 $ IMCAT3      : Factor w/ 1 level "B": 1 1 1 1 1 1 1 1 1 ...
 $ Category    : Factor w/ 1 level "MAB": 1 1 1 1 1 1 1 1 1 ...
 $ IMDES1      : Factor w/ 526 levels "ACER A5 LAPTOP",...: 286 183 287 287 281 183 286 284 183 183 ...
 $ IMDES2      : Factor w/ 517 levels "","S11553 C.2.48GHz W/15","...: 365 334 426 429 462 334 365 366 334 334 ...
 $ REGION      : Factor w/ 10 levels "BA", "BD", "EC",...: 1 7 7 7 1 6 7 4 4 8 ...
 $ LCDIST      : Factor w/ 5 levels "D1", "D2", "D3",...: 3 1 1 2 1 1 3 3 2 ...
 $ LCNAME      : Factor w/ 532 levels "Albany", "Albuquerque",...: 449 431 431 431 384 374 396 163 163 363 ...
 $ LCCITY      : Factor w/ 467 levels "ALBANY", "ALBUQUERQUE",...: 484 299 299 299 345 336 355 143 143 328 ...
 $ LCSTA       : Factor w/ 16 levels "AK", "AL", "AZ",...: 4 42 42 42 4 32 42 4 4 3 ...
 $ MSPIPK      : int  539928 6142209 6139758 6139758 826806 1188080 1880808 4777114 4777114 2640983 ...
 $ ACCTYP      : int  1 1 1 1 1 1 1 1 1 ...
 $ EXECFLG     : Factor w/ 2 levels "N", "E": 2 2 1 1 1 2 2 2 1 ...
 $ Label       : Factor w/ 27 levels "ACER", "ALIEWARE",...: 11 8 11 11 11 8 11 11 8 8 ...
```

```
In [4]: #Edit Date Time
df$date_only<-sapply(as.character(df$MBDATE), FUN = function(x) {strsplit(x, split = '[' ])[[1]][1]})

# create month and year variables
df$month <- sapply(df$date_only, FUN = function(x) {strsplit(x, split = '[' ])[[1]][1]})
df$year <- sapply(df$date_only, FUN = function(x) {strsplit(x, split = '[' ])[[1]][3]})
```

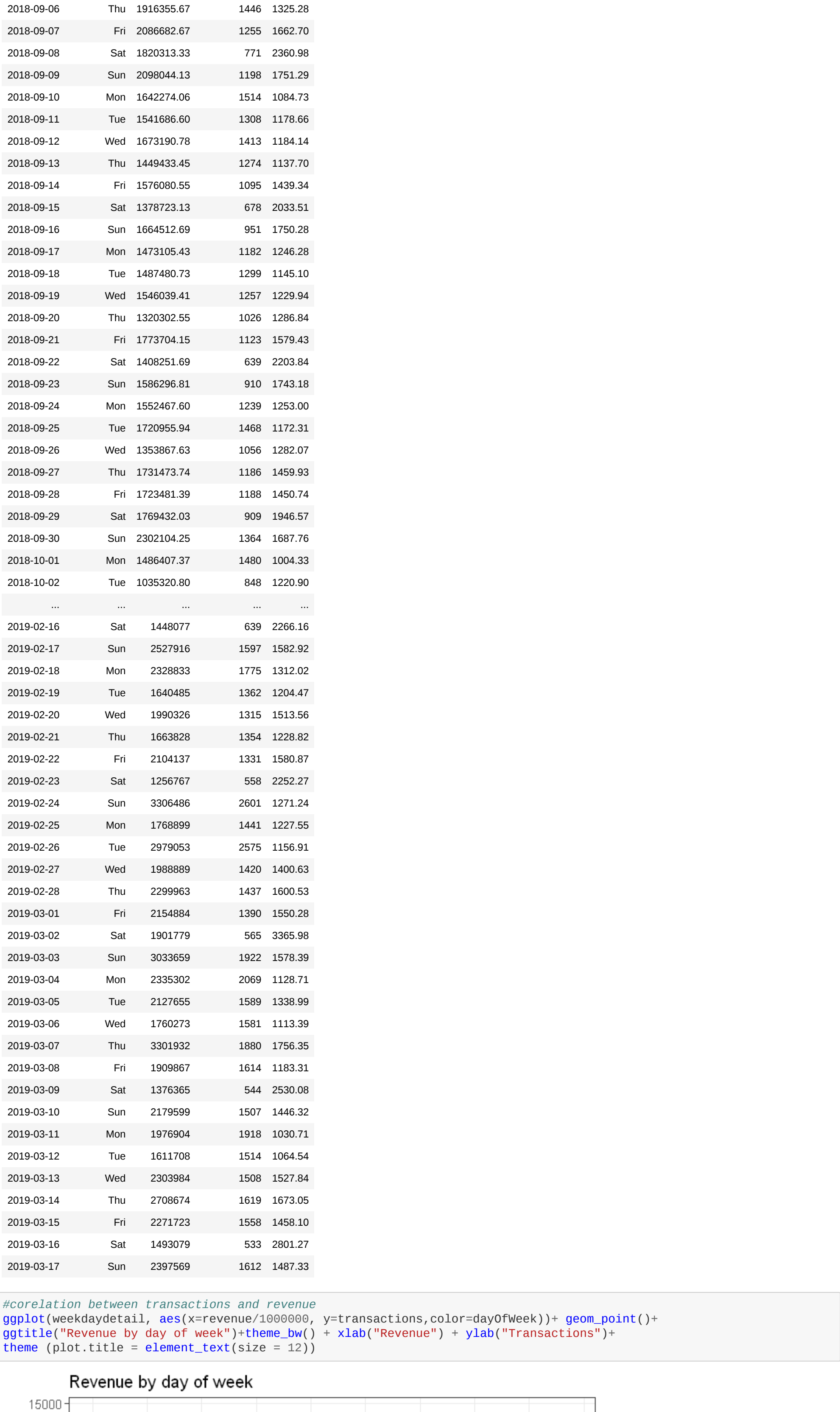
```
In [5]: #adjust date format
df$date_only <- as.Date(df$date_only, "%m/%d/%Y")
df$dayOfWeek <- wday(df$date, label=TRUE)
```

```
In [6]: df$week<-week(df$date)
df$month<-month(ymd(df$date_only),label=FALSE, abbr=FALSE)
```

```
In [8]: # aggregate by Transaction Number
df_new<-df%>%
  group_by(MSPIPK, MBITEM) %>%
  select(MBWHSE, MBREG_NUM, MBTRN_NUM, FISCAL_PERIOD, date_only, year, month, week, dayOfWeek, MBDSUN, MBDSEL,
  MBITEM, MBITEM, MSPIPK, ACCTYP)
head(df_new)
```

MBWHSE	MBREG_NUM	MBTRN_NUM	FISCAL_PERIOD	date_only	year	month	week	dayOfWeek	MBDSUN	MBDSEL	MBITEM	MSPRPK	ACCTYP
423	12	398	2	2018-10-26	2018	10	43	Fri	1	549.99	1234449	539928	1
770	2	108	2	2018-10-26	2018	10	43	Fri	1	469.99	1235362	6142209	1
770	2	209	2	2018-10-26	2018	10	43	Fri	1	899.99	1238985	6139758	1
770	51	52	2	2018-10-26	2018	10	43	Fri	-1	-899.99	1238985	6139758	1
29	9	482	2	2018-10-26	2018	10	43	Fri	1	799.99	1275609	426806	1
1070	4	410	2	2018-10-10	2018	10	41	Wed	1	469.99	1235362	1108808	1

```
In [10]: ##Revenue details by Day of week
weekdaydetail <- df %>%
  group_by(date_only, dayOfWeek) %>%
  summarise(revenue = sum(MBDSEL), transactions = n_distinct(MBTRN_NUM)) %>%
  mutate(aveOrd = (round((revenue / transactions),2))) %>%
  arrange(date_only) %>%
  ungroup()
weekdaydetail
```



```
In [ ]: #export manipulated file
write.csv(df, "C:/Users/jessi/OneDrive/Desktop/ALY 6070/Final Project/SalesFY19P1to7.csv")
```