

# Coal

yyaadet2002

2022-10-25

## Coal analysis

library

```
library(knitr)
library(dplyr)
```

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

```
library(ggplot2)
library(pinyin)
```

initial

```
pydic(only_first_letter = T)
```

```
## <environment: 0x1107fb158>
```

```
df = read.csv("coal.csv")
df$timestamp = as.Date(df$timestamp)
df$name = py(df$name, dic = pydic(only_first_letter = F, method = "toneless"), sep = "_")
df$market_value = df$close * df$volume
head(df)
```

```
##   X               id      entity_id timestamp provider  code
## 1 0 stock_sz_000723_2005-01-04 stock_sz_000723 2005-01-04 joinquant 723
## 2 1 stock_sz_000983_2005-01-04 stock_sz_000983 2005-01-04 joinquant 983
## 3 2 stock_sh_600157_2005-01-04 stock_sh_600157 2005-01-04 joinquant 600157
## 4 3 stock_sh_600188_2005-01-04 stock_sh_600188 2005-01-04 joinquant 600188
## 5 4 stock_sz_000723_2005-01-05 stock_sz_000723 2005-01-05 joinquant 723
## 6 5 stock_sz_000983_2005-01-05 stock_sz_000983 2005-01-05 joinquant 983
##           name level  open close  high   low  volume turnover change_pct
## 1  mei_jin_nai_yuan   1d 13.69 13.28 13.72 13.25   47160   629031         NA
## 2  shan_xi_jiao_mei   1d 15.61 15.19 15.61 15.14 1069190 16304143         NA
## 3  yong_tai_nai_yuan   1d 17.20 16.82 17.43 16.82   53796   916412         NA
## 4  yan_kuang_nai_yuan   1d 13.20 12.92 13.20 12.73 2692892 34732570         NA
## 5  mei_jin_nai_yuan   1d 13.12 13.33 13.51 13.12   30246   401927         NA
```

```
## 6 shan_xi_jiao_mei 1d 15.19 15.38 15.38 15.09 1345112 20524755 NA
## turnover_rate market_value
## 1 NA 626284.8
## 2 NA 16240996.1
## 3 NA 904848.7
## 4 NA 34792164.6
## 5 NA 403179.2
## 6 NA 20687822.6
```

codes.

```
df %>%
  group_by(entity_id, name) %>%
  summarise(count=n(),
            list_date = min(timestamp),
            open = mean(open),
            close = mean(close),
            last_date = max(timestamp),
            market_value = mean(market_value)) %>%
  arrange(desc(market_value))
```

## `summarise()` has grouped output by 'entity\_id'. You can override using the `.groups` argument.

```
## # A tibble: 6 x 8
## # Groups:   entity_id [6]
##   entity_id      name      count list_date    open close last_date market_value
##   <chr>         <chr>    <int> <date>      <dbl> <dbl> <date>      <dbl>
## 1 stock_sh_601088 zhong_gu~ 3675 2007-10-09 30.5 30.5 2022-11-15 701317155.
## 2 stock_sh_601225 shan_xi_~ 2141 2014-01-28 10.1 10.1 2022-11-15 500417726.
## 3 stock_sz_000983 shan_xi_~ 4341 2005-01-04 69.0 69.1 2022-11-15 474707969.
## 4 stock_sz_000723 mei_jin_~ 4341 2005-01-04 36.9 37.0 2022-11-15 340090367.
## 5 stock_sh_600188 yan_kuan~ 4341 2005-01-04 41.0 41.0 2022-11-15 293973514.
## 6 stock_sh_600157 yong_tai~ 4341 2005-01-04 111. 112. 2022-11-15 219133937.
```

summary.

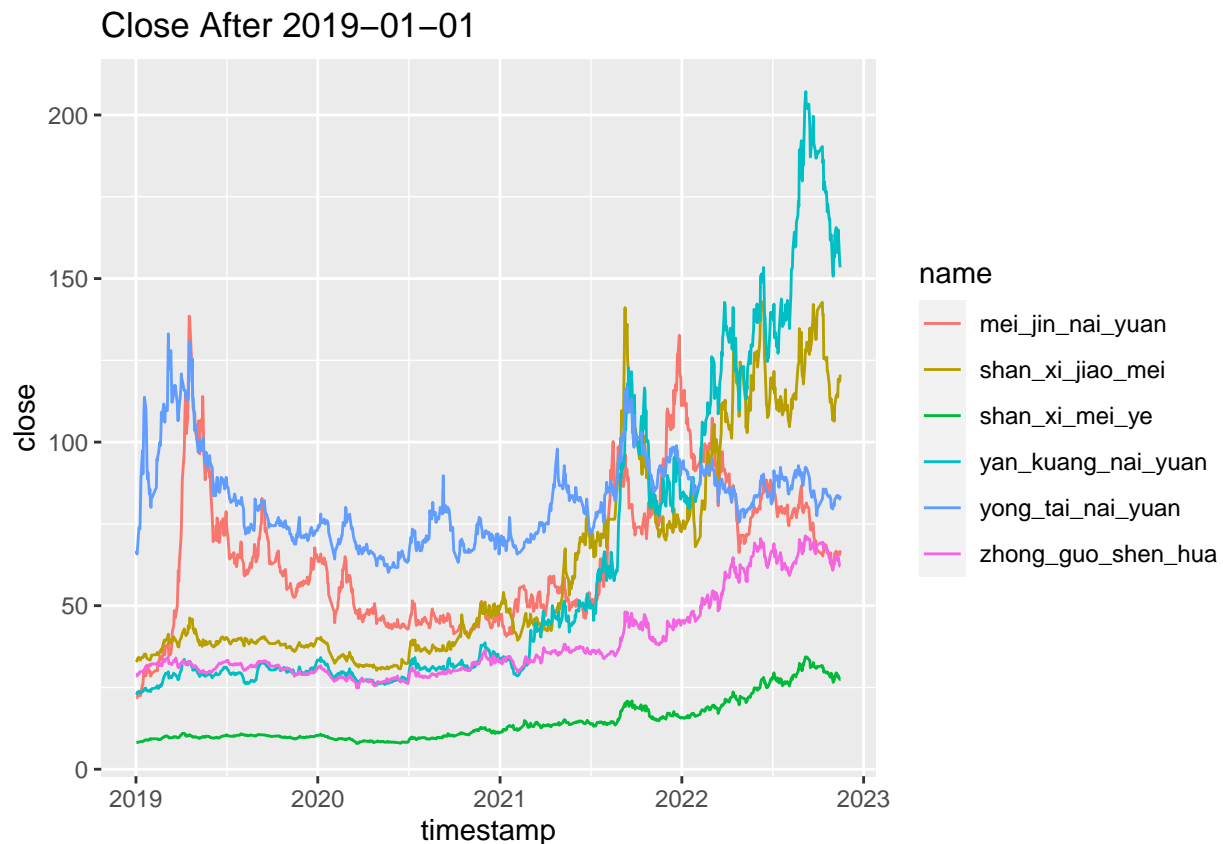
```
summary(df)
```

```
##      X      id      entity_id      timestamp
## Min.   : 0   Length:23180   Length:23180   Min.   :2005-01-04
## 1st Qu.: 5795 Class :character Class :character 1st Qu.:2010-04-29
## Median :11590 Mode  :character Mode  :character Median :2014-12-08
## Mean   :11590                                     Mean   :2014-07-31
## 3rd Qu.:17384                                     3rd Qu.:2018-11-22
## Max.   :23179                                     Max.   :2022-11-15
##   provider      code      name      level
## Length:23180   Min.   : 723   Length:23180   Length:23180
## Class :character 1st Qu.: 983   Class :character Class :character
## Mode  :character Median :600157 Mode  :character Mode  :character
##                                     Mean  :375942
##                                     3rd Qu.:601088
##                                     Max.   :601225
##   open      close      high      low
## Min.   : 3.84   Min.   : 3.85   Min.   : 4.03   Min.   : 3.73
## 1st Qu.: 22.29   1st Qu.: 22.31   1st Qu.: 22.74   1st Qu.: 21.80
## Median : 35.29   Median : 35.29   Median : 35.80   Median : 34.67
## Mean   : 54.14   Mean   : 54.21   Mean   : 55.28   Mean   : 53.12
```

```
## 3rd Qu.: 67.45    3rd Qu.: 67.56    3rd Qu.: 69.03    3rd Qu.: 66.27
## Max.   :490.68    Max.   :480.60    Max.   :517.88    Max.   :457.43
##      volume      turnover      change_pct      turnover_rate
## Min.    :      0    Min.    :0.000e+00    Mode:logical    Mode:logical
## 1st Qu.: 1576526    1st Qu.:6.745e+07    NA's:23180      NA's:23180
## Median : 5048656    Median :2.004e+08
## Mean    : 11852364    Mean    :4.057e+08
## 3rd Qu.: 13279986    3rd Qu.:5.059e+08
## Max.    :465128280    Max.    :3.225e+10
## market_value
## Min.    :0.000e+00
## 1st Qu.:6.731e+07
## Median :2.000e+08
## Mean    :4.061e+08
## 3rd Qu.:5.074e+08
## Max.    :3.223e+10
```

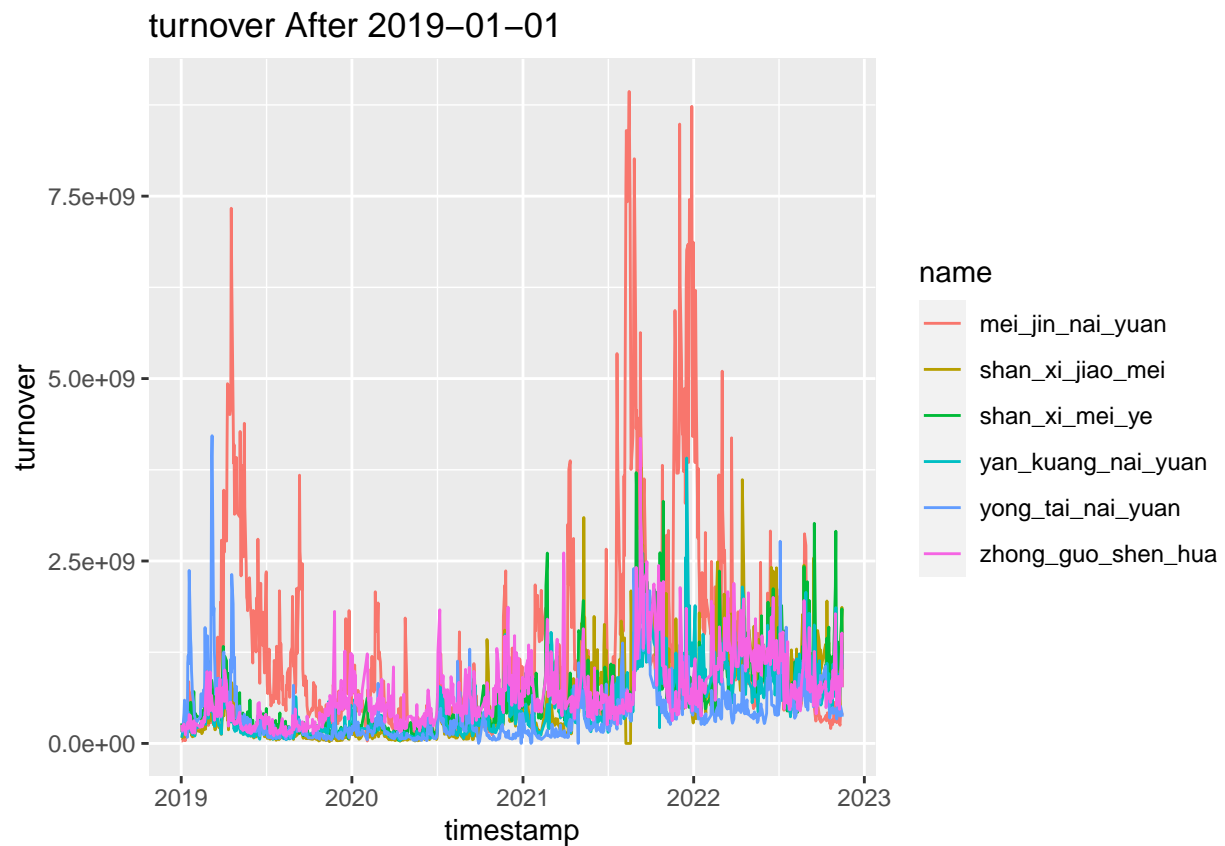
plot

```
# close
df %>%
  filter(timestamp > as.Date("2019-01-01")) %>%
  ggplot(aes(x = timestamp, y=close, color=name)) +
  geom_line() +
  labs(title="Close After 2019-01-01")
```

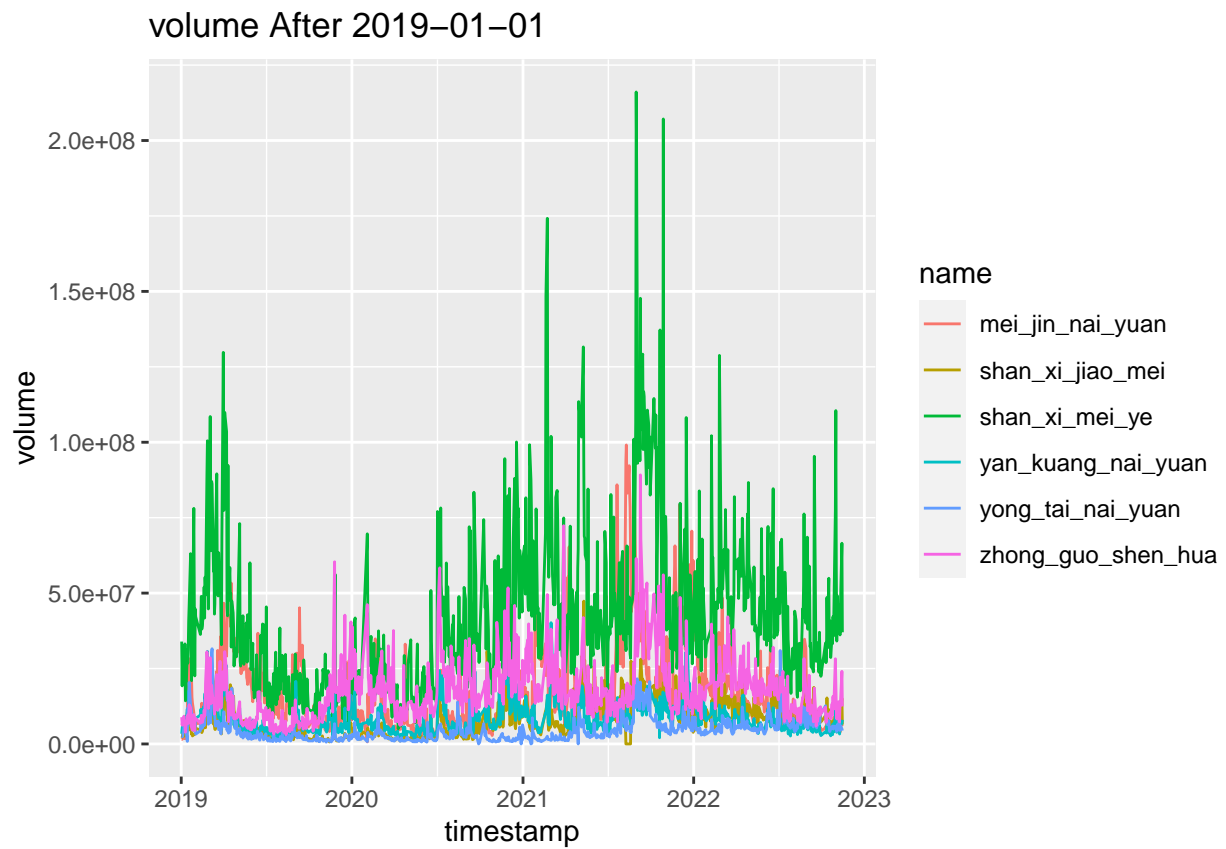


```
# turnover
df %>%
```

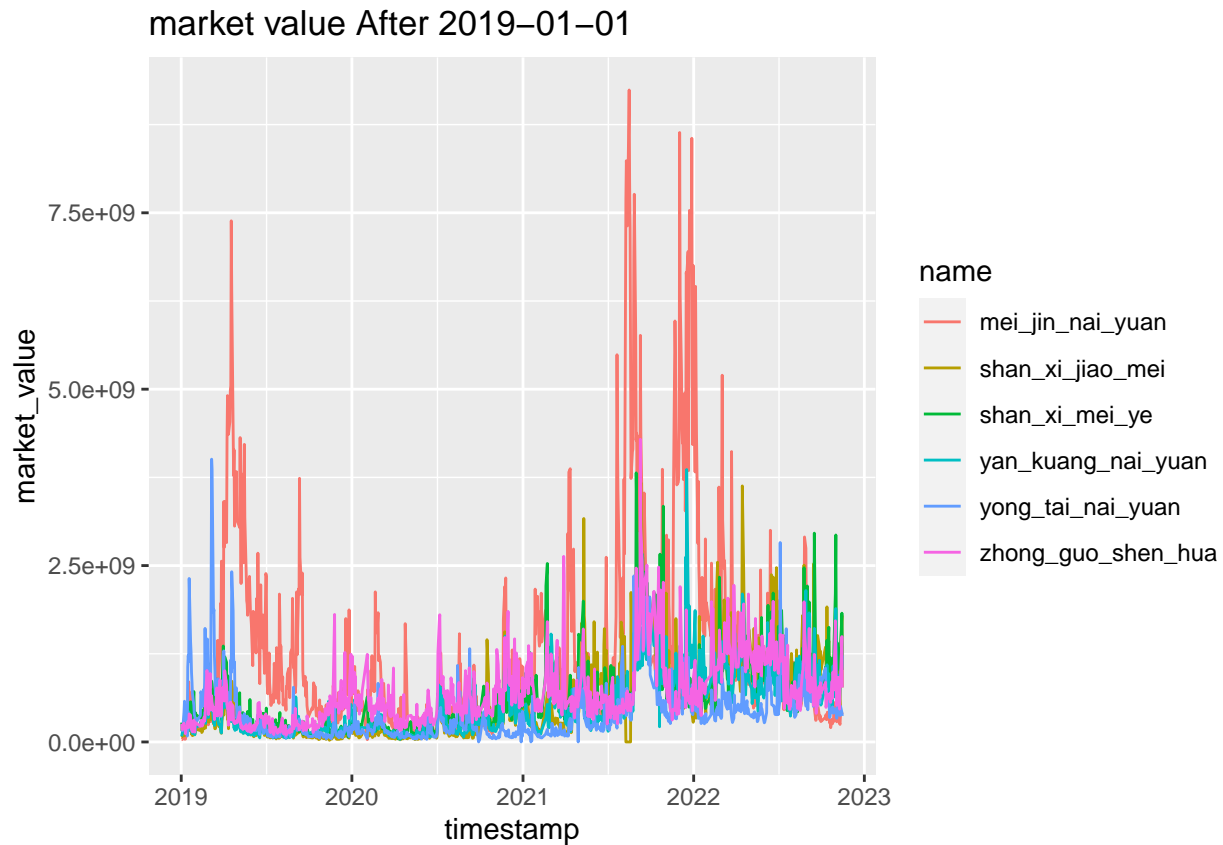
```
filter(timestamp > as.Date("2019-01-01")) %>%
ggplot(aes(x = timestamp, y=turnover, color=name)) +
geom_line() +
labs(title = "turnover After 2019-01-01")
```



```
# volume
df %>%
filter(timestamp > as.Date("2019-01-01")) %>%
ggplot(aes(x = timestamp, y=volume, color=name)) +
geom_line() +
labs(title = "volume After 2019-01-01")
```



```
# market value
df %>%
  filter(timestamp > as.Date("2019-01-01")) %>%
  ggplot(aes(x = timestamp, y=market_value, color=name)) +
  geom_line() +
  labs(title = "market value After 2019-01-01")
```



correlations.

```
names = unique(df$name)
start = as.Date("2022-01-01")
df.corr = NULL

for(i in 1:length(names)) {
  df.entity = df[which(df$name == names[i] & df$timestamp > start), ]
  df.entity.sub = subset(df.entity, select = c(timestamp, close))
  colnames(df.entity.sub) = c("ts", names[i])
  if(is.null(df.corr)) {
    df.corr = df.entity.sub
  } else {
    df.corr = merge(df.corr, df.entity.sub, by = "ts")
  }
}

df.corr %>%
  subset(select = -ts) %>%
  cor()
```

```
##           mei_jin_nai_yuan shan_xi_jiao_mei yong_tai_nai_yuan
## mei_jin_nai_yuan           1.0000000      -0.6656913         0.66098558
## shan_xi_jiao_mei          -0.6656913         1.0000000        -0.30110717
## yong_tai_nai_yuan           0.6609856        -0.3011072         1.00000000
## yan_kuang_nai_yuan        -0.6606881         0.8239889        -0.10142972
## zhong_guo_shen_hua        -0.7136521         0.9235554        -0.26469252
```

```
## shan_xi_mei_ye          -0.6924222      0.8147692      -0.09831876
##                        yan_kuang_nai_yuan zhong_guo_shen_hua shan_xi_mei_ye
## mei_jin_nai_yuan        -0.6606881      -0.7136521      -0.69242216
## shan_xi_jiao_mei         0.8239889        0.9235554        0.81476916
## yong_tai_nai_yuan       -0.1014297      -0.2646925      -0.09831876
## yan_kuang_nai_yuan       1.0000000        0.8556054        0.94106462
## zhong_guo_shen_hua       0.8556054        1.0000000        0.90812250
## shan_xi_mei_ye          0.9410646        0.9081225        1.00000000
```

```
head(df.corr)
```

```
##           ts mei_jin_nai_yuan shan_xi_jiao_mei yong_tai_nai_yuan
## 1 2022-01-04          111.55           75.99           94.34
## 2 2022-01-05          115.39           73.71           91.79
## 3 2022-01-06          115.87           73.62           91.79
## 4 2022-01-07          111.07           73.54           91.79
## 5 2022-01-10          111.83           75.55           91.28
## 6 2022-01-11          105.60           75.20           91.28
##   yan_kuang_nai_yuan zhong_guo_shen_hua shan_xi_mei_ye
## 1              84.81              45.57              15.80
## 2              83.29              44.91              15.57
## 3              84.21              44.55              15.55
## 4              81.73              45.13              15.63
## 5              83.53              45.67              16.07
## 6              81.02              45.51              15.92
```

Computer returns.

```
df.returns = df.corr
df.returns$weekday = weekdays(df.returns$ts)
for (i in 2:nrow(df.returns)) {
  df.returns[i, "mei_jin_nai_yuan_return"] =
    df.returns$mei_jin_nai_yuan[i] / df.returns$mei_jin_nai_yuan[i-1] - 1

  df.returns[i, "shan_xi_jiao_mei_return"] =
    df.returns$shan_xi_jiao_mei[i] / df.returns$shan_xi_jiao_mei[i-1] - 1

  df.returns[i, "yong_tai_nai_yuan_return"] =
    df.returns$yong_tai_nai_yuan[i] / df.returns$yong_tai_nai_yuan[i-1] - 1

  df.returns[i, "yan_kuang_nai_yuan_return"] =
    df.returns$yan_kuang_nai_yuan[i] / df.returns$yan_kuang_nai_yuan[i-1] - 1

  df.returns[i, "zhong_guo_shen_hua_return"] =
    df.returns$zhong_guo_shen_hua[i] / df.returns$zhong_guo_shen_hua[i-1] - 1

  df.returns[i, "shan_xi_mei_ye_return"] =
    df.returns$shan_xi_mei_ye[i] / df.returns$shan_xi_mei_ye[i-1] - 1
}

head(df.returns)
```

```
##           ts mei_jin_nai_yuan shan_xi_jiao_mei yong_tai_nai_yuan
## 1 2022-01-04          111.55           75.99           94.34
## 2 2022-01-05          115.39           73.71           91.79
## 3 2022-01-06          115.87           73.62           91.79
```

```
## 4 2022-01-07      111.07      73.54      91.79
## 5 2022-01-10      111.83      75.55      91.28
## 6 2022-01-11      105.60      75.20      91.28
##   yan_kuang_nai_yuan zhong_guo_shen_hua shan_xi_mei_ye   weekday
## 1           84.81           45.57           15.80   Tuesday
## 2           83.29           44.91           15.57 Wednesday
## 3           84.21           44.55           15.55 Thursday
## 4           81.73           45.13           15.63   Friday
## 5           83.53           45.67           16.07   Monday
## 6           81.02           45.51           15.92   Tuesday
##   mei_jin_nai_yuan_return shan_xi_jiao_mei_return yong_tai_nai_yuan_return
## 1                NA                NA                NA
## 2           0.034424025           -0.030003948           -0.027029892
## 3           0.004159806           -0.001221001           0.000000000
## 4           -0.041425736           -0.001086661           0.000000000
## 5           0.006842532           0.027332064           -0.005556161
## 6           -0.055709559           -0.004632694           0.000000000
##   yan_kuang_nai_yuan_return zhong_guo_shen_hua_return shan_xi_mei_ye_return
## 1                NA                NA                NA
## 2           -0.01792241           -0.014483213           -0.014556962
## 3           0.01104574           -0.008016032           -0.001284522
## 4           -0.02945018           0.013019080           0.005144695
## 5           0.02202374           0.011965433           0.028150992
## 6           -0.03004908           -0.003503394           -0.009334163
```

Weekday return.

```
df.returns %>%
  na.omit() %>%
  group_by(weekday) %>%
  summarise(
    mei_jin_nai_yuan_return = mean(mei_jin_nai_yuan_return),
    shan_xi_jiao_mei_return = mean(shan_xi_jiao_mei_return),
    yong_tai_nai_yuan_return = mean(yong_tai_nai_yuan_return),
    yan_kuang_nai_yuan_return = mean(yan_kuang_nai_yuan_return),
    zhong_guo_shen_hua_return = mean(zhong_guo_shen_hua_return),
    shan_xi_mei_ye_return = mean(shan_xi_mei_ye_return)
  ) %>%
  arrange(weekday)

## # A tibble: 5 x 7
##   weekday   mei_jin_nai_yuan~ shan_xi_jiao_me~ yong_tai_nai_yu~ yan_kuang_nai_y~
##   <chr>         <dbl>         <dbl>         <dbl>         <dbl>
## 1 Friday       -0.00460         0.00523         0.00112         0.00136
## 2 Monday       -0.00287         0.00642         0.000836        0.00564
## 3 Thursday     -0.00287         0.00561        -0.00281         0.00810
## 4 Tuesday      -0.00294        -0.000960         0.000493         0.00140
## 5 Wednesday     0.00271        -0.00155        -0.00147         0.000732
## # ... with 2 more variables: zhong_guo_shen_hua_return <dbl>,
## #   shan_xi_mei_ye_return <dbl>
```

Weekday increase proportion.

```
df.returns %>%
  na.omit() %>%
  mutate(mei_jin_nai_yuan_return = ifelse(mei_jin_nai_yuan_return > 0, 1, 0)) %>%
```



```

group_by(weekday) %>%
summarise(
  mei_jin_nai_yuan_return = sum(mei_jin_nai_yuan_return) / n(),
  shan_xi_jiao_mei_return = sum(shan_xi_jiao_mei_return) / n(),
  yong_tai_nai_yuan_return = sum(yong_tai_nai_yuan_return) / n(),
  yan_kuang_nai_yuan_return = sum(yan_kuang_nai_yuan_return) / n(),
  zhong_guo_shen_hua_return = sum(zhong_guo_shen_hua_return) / n(),
  shan_xi_mei_ye_return = sum(shan_xi_mei_ye_return) / n(),

) %>%
arrange(weekday) %>%
kable()

```

weekday	mei_jin_nai_yuan_return	shan_xi_jiao_mei_return	yong_tai_nai_yuan_return	yan_kuang_nai_yuan_return	zhong_guo_shen_hua_return	shan_xi_mei_ye_return
Friday	0.4047619	0.0052272	0.0011243	0.0013560	0.0029761	0.0015667
Monday	0.5750000	0.0064176	0.0008358	0.0056391	0.0025930	0.0063842
Thursday	0.3953488	0.0056128	-0.0028086	0.0080985	0.0050024	0.0070206
Tuesday	0.4878049	-0.0009603	0.0004927	0.0013980	0.0014390	0.0011439
Wednesday	0.5238095	-0.0015536	-0.0014710	0.0007318	-0.0030594	-0.0006924