HW 1

Load data

ดาวน์โหลดข้อมูล excel เข้าสู่โปรแกรม R

```
# initial parameter
sheet_name <- "aus_production"
column_name <- c("Quarter", "Cement")

# load data
file_name <- file.choose()
df0 <- read_excel(file_name, sheet=sheet_name)
df1 <- df0 %>% select(column_name)

# Convert to tsibble
df2 <- df1  %>%
  mutate(Quarter = yearquarter(Quarter))  %>%
  as_tsibble(index = Quarter)
```

Train/Test split

แบ่ง dataset เป็น train และ test

```
# split data
train <- df2  %>% filter(year(Quarter) >= 1988 & year(Quarter) <= 2007)
test <- df2  %>% filter(year(Quarter) >= 2008)
```

Train model

ทำ model training ด้วย model แบบต่างๆ

```
# train model
fit <- train %>%
  model(
      Mean = MEAN(Cement),
      Naive = NAIVE(Cement),
      SeasonNaive = SNAIVE(Cement),
      Drift = NAIVE(Cement ~ drift()),
      ETS = ETS(Cement),
      ARIMA = ARIMA(Cement)
  )

# train data evaluation
accuracy(fit)
```

Forecasting

ทำนายผลของข้อมูล test ด้วย model แต่ละตัวที่ train มา

```
# forcasting
fc <- fit %>% forecast(h=nrow(test))
```

Evaluation

วัดประสิทธิภาพของ model ที่ train มาด้วย metric แบบต่างๆ

```
# test data evaluation
acc <- accuracy(fc, test)
```

Result

ผลการทำ forecasting ด้วย **excel** (ไม่มี ETS และ ARIMA)

1	Quarter	Cement	Mean	Naive	SeasonNaive	Drift
208	2007 Q3	2536				
209	2007 Q4	2562				
210	2008 Q1	2183	1843.5	2562	2140	2576.481013
211	2008 Q2	2558	1843.5	2562	2362	2590.962025
212	2008 Q3	2612	1843.5	2562	2536	2605.443038
213	2008 Q4	2373	1843.5	2562	2562	2619.924051
214	2009 Q1	1963	1843.5	2562	2140	2634.405063
215	2009 Q2	2160	1843.5	2562	2362	2648.886076
216	2009 Q3	2325	1843.5	2562	2536	2663.367089
217	2009 Q4	2273	1843.5	2562	2562	2677.848101
218	2010 Q1	1904	1843.5	2562	2140	2692.329114
219	2010 Q2	2401	1843.5	2562	2362	2706.810127
220	**RMSE**	-	**483.9384775**	**360.6740911**	**184.1613423**	**435.9013116**

ผลการทำ forecasting ด้วยโปรแกรม **R** (มี ETS และ ARIMA)

```
> fc_all
# A tibble: 10 × 7
   Quarter  Mean Naive SeasonNaive Drift   ETS ARIMA
     <qtr> <dbl> <dbl>       <dbl> <dbl> <dbl> <dbl>
 1 2008 Q1 1844.  2562        2140 2576. 2248. 2319.
 2 2008 Q2 1844.  2562        2362 2591. 2493. 2502.
 3 2008 Q3 1844.  2562        2536 2605. 2576. 2543.
 4 2008 Q4 1844.  2562        2562 2620. 2542. 2488.
 5 2009 Q1 1844.  2562        2140 2634. 2248. 2296.
 6 2009 Q2 1844.  2562        2362 2649. 2493. 2490.
 7 2009 Q3 1844.  2562        2536 2663. 2576. 2520.
 8 2009 Q4 1844.  2562        2562 2678. 2542. 2463.
 9 2010 Q1 1844.  2562        2140 2692. 2248. 2264.
10 2010 Q2 1844.  2562        2362 2707. 2493. 2472.
>
```

evaluation metric จากโปรแกรม R

```
> acc
# A tibble: 6 × 10
  .model      .type    ME  RMSE   MAE    MPE  MAPE  MASE RMSSE  ACF1
  <chr>       <chr> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARIMA       Test  -161.  216.  186.  -7.71  8.68   NaN   NaN 0.387
2 Drift       Test  -366.  436.  368. -17.3  17.3    NaN   NaN 0.264
3 ETS         Test  -171.  222.  191.  -8.07  8.85   NaN   NaN 0.579
4 Mean        Test   432.  484.  432.  18.2  18.2    NaN   NaN 0.117
5 Naive       Test  -287.  361.  297. -13.7  14.1    NaN   NaN 0.117
6 SeasonNaive Test   -95   184.  166.  -4.64  7.47   NaN   NaN 0.537
>
```

Conclusion

จากผลการทดสอบพบว่า โมเดลที่ดีที่สุดคือ **Seasonal Naïve** โดยให้ค่า RMSE อยู่ที่ 184.16 ดังนั้นการทำ forecasting สำหรับ dataset aus_production เพื่อทำนายข้อมูล Cement โมเดลที่ดีที่สุดจากการทดสอบคือ **Seasonal Naïve**