



2018 **FAST CAMPUS**  
**DATA SCIENCE**  
**SCHOOL PROJECT(1)**

REGRESSION ANALYSIS

DATA SCIENCE SCHOOL

# 목차

- 01. 주제설명
- 02. EDA
- 03. CORRELATION PLOT
- 04. FEATURE SELECTION
- 05. OLS MODELING
- 06. CONCLUSION
- 07. MOVING AVERAGE

## 01.

### Walmart Recruiting II : Sales in Stormy Weather

Predict how sales of weather – sensitive products are affected by snow and rain



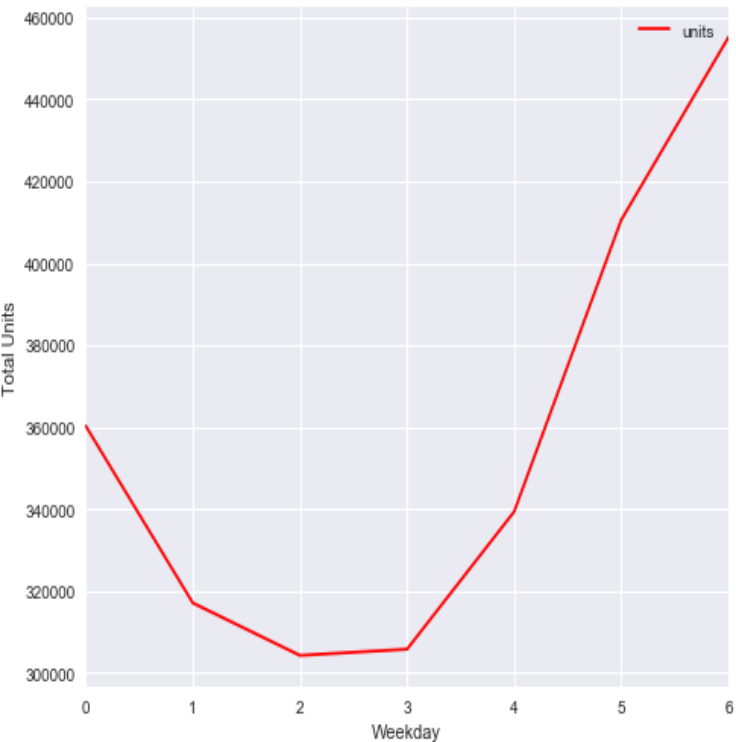
### Objective

45개의 store에서 weather-sensitive products are affected by snow and rain

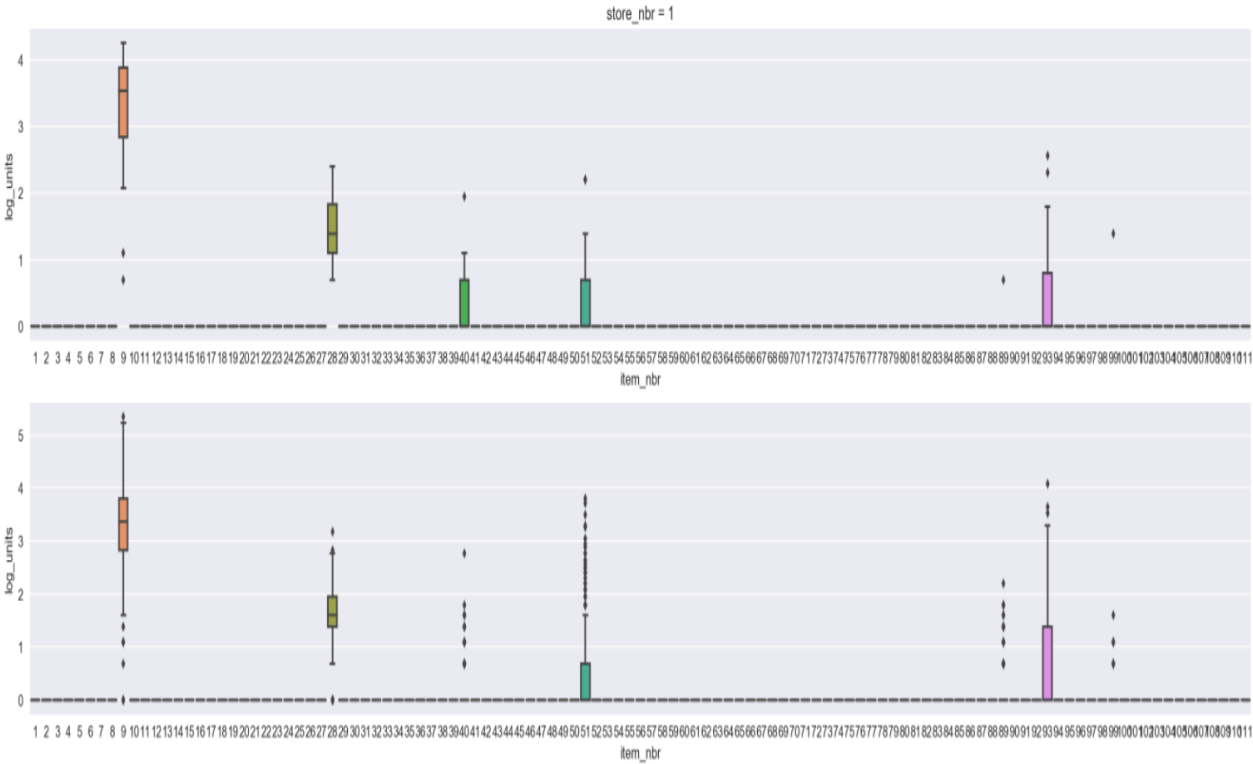
### Data Set

- Weather: 2012.01.01 – 2014.10.31의 각 statio날씨
- Key: Store과 Weather Station간의 관계 Mapping
- Train: 2012.01.01 – 2014.10.31의 각 Store, Item 별 Units Data( test날씨 제외)
- Test: 2013.04.01 이후 Weather Event가 발생한 전후 3일

02. EDA



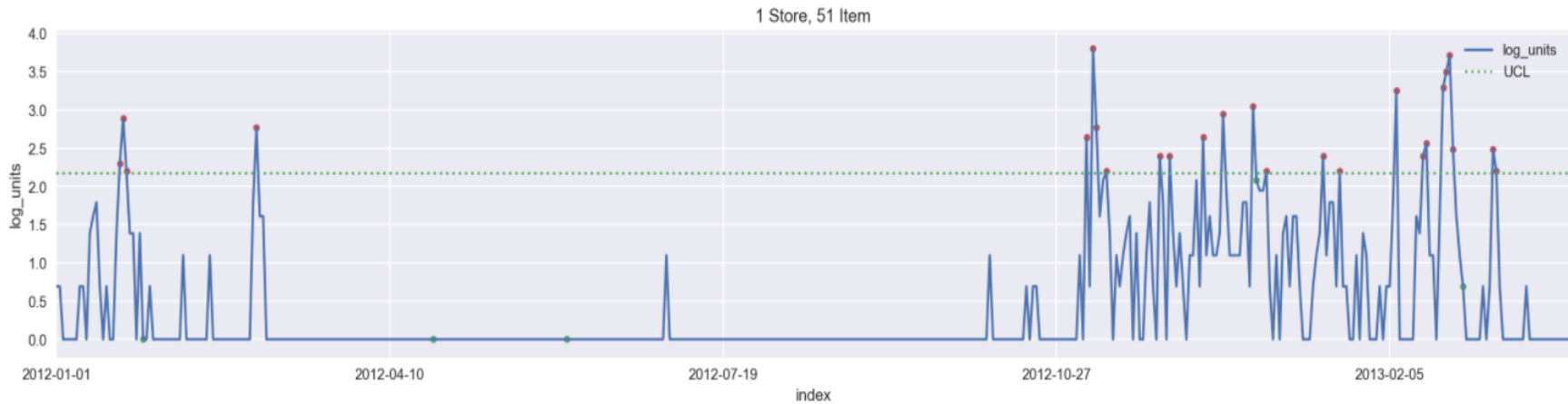
요일별 유닛총량



Store nbr 1의 아이템별 공휴일/비공휴일 판매량

02. EDA

Unit Data



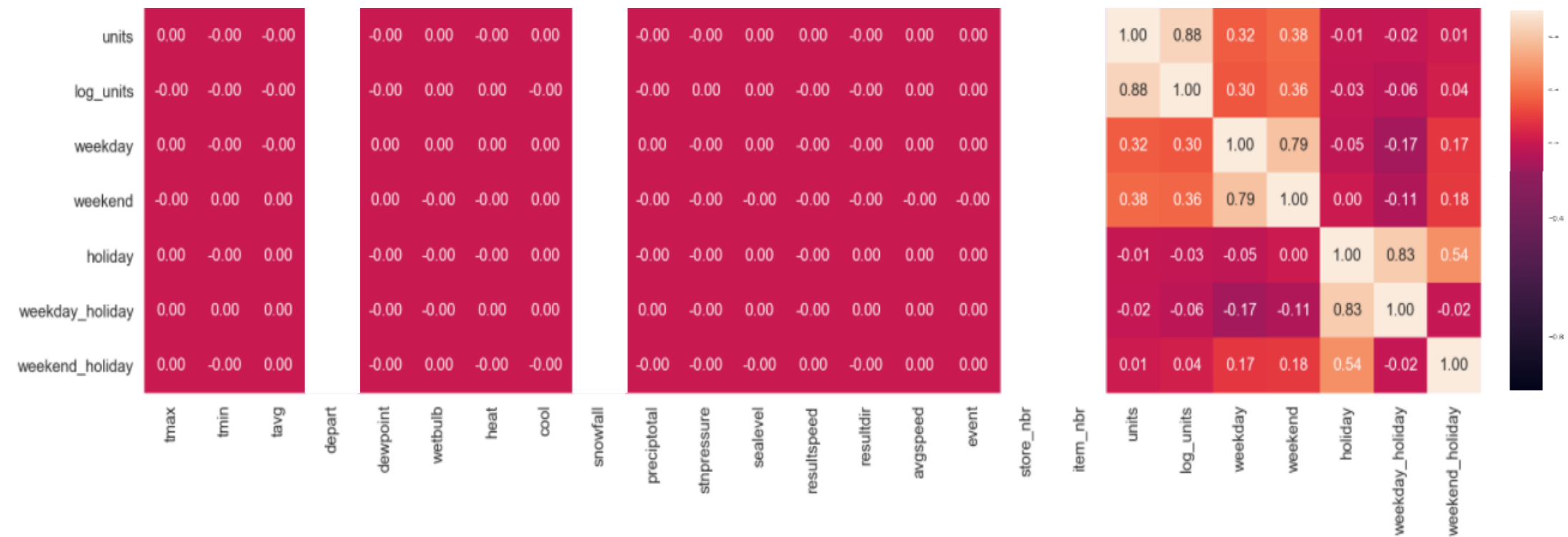
Warning! : 85221 2012-03-19  
177573 2012-10-13  
Name: date, dtype: object, Match: 2

Unit Data event



### 03. CORRELATION PLOT

1번 스토어 9번 아이템 Correlation plot



DATA SCIENCE SCHOOL

## 04. FEATURE SELECTION

feature selection

- 각 store\_nbr, item\_nbr별로 나누어 modeling 해야한다
- weather와 log\_units(또는 units)는 큰 상관관계가 없어 보인다
- weekday와 holiday가 log\_units(또는 units)와 약간의 상관관계가 있어 보인다 (item\_nbr에 따라 다름)

## 04. OLS MODELING

Your most recent submission				
Name	Submitted	Wait time	Execution time	Score
sub_test.csv	a few seconds ago	3 seconds	6 seconds	1.78595
Complete				
<a href="#">Jump to your position on the leaderboard ▾</a>				

Your most recent submission				
Name	Submitted	Wait time	Execution time	Score
sub_test3.csv	a few seconds ago	2 seconds	5 seconds	0.44340
Complete				
<a href="#">Jump to your position on the leaderboard ▾</a>				

Your most recent submission				
Name	Submitted	Wait time	Execution time	Score
sub_test2.csv	a few seconds ago	1 seconds	5 seconds	0.44604
Complete				
<a href="#">Jump to your position on the leaderboard ▾</a>				

### 1' OLS. 2등(뒤에서)

$\log\_units \sim C(\text{store\_nbr}) + C(\text{item\_nbr}) + C(\text{week day}) + C(\text{holiday}) + C(\text{event}) + 0$

### 2' OLS. 485명중 350등

$C(\text{store\_nbr}):C(\text{item\_nbr}) + C(\text{weekday}) + C(\text{holiday}) + C(\text{event}) + 0$

### 2-1' OLS.

$C(\text{store\_nbr}):C(\text{item\_nbr}) + C(\text{weekday}) + C(\text{holiday}) + \text{snowfall} + \text{preciptotal} + 0$







## 04. OLS MODELING

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
sub_test4.csv	a few seconds ago	3 seconds	5 seconds	0.15898

Complete

[Jump to your position on the leaderboard](#)

272	▼14	63614		0.15659	20	3y
273	▲3	81010		0.15800	69	3y
274	▼2	mnrl		0.15944	2	3y
275	▲3	sna		0.15984	8	3y

3번

log\_units ~

C(station\_nbr):C(store\_nbr):C(item\_nbr) +  
C(weekday) + C(holiday) + C(event) + 0

memory error 발생

station\_nbr별로 나눠서 ols

4번 log\_units ~ C(store\_nbr):C(item\_nbr) +  
C(weekday) + C(holiday) + C(event) + 0

5번 log\_units ~ C(store\_nbr):C(item\_nbr) +  
C(weekday) + C(holiday) + snowfall +  
preciptotal + 0

store\_nbr별로 나눠서 ols

event 제외

결과안 좋음

04. OLS MODELING


Walmart

6번  
 $\log\_units \sim C(item\_nbr):C(weekday) + C(item\_nbr):C(holiday) + 0$

최종  
 $\log\_units \sim C(item\_nbr):C(weekday)+0$

DATA SCIENCE SCHOOL

Walmart



Walmart Recruiting II: Sales in Stormy Weather

Predict how sales of weather-sensitive products are affected by snow and rain

485 teams · 3 years ago

Overview

Data

Discussion

Leaderboard

Rules

Team

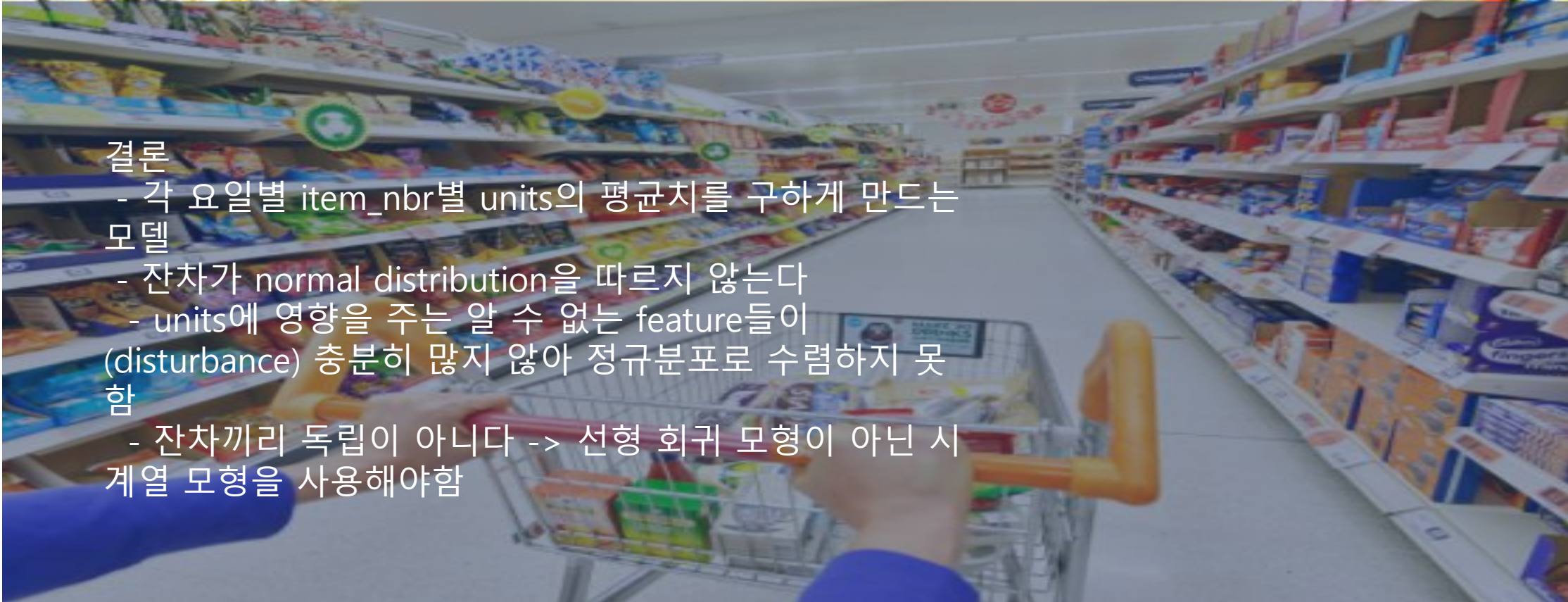
My Submissions

Late Submission

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
------	-----------	-----------	----------------	-------

## 05. CONCLUSION

- 
- 결론
- 각 요일별 item\_nbr별 units의 평균치를 구하게 만드는
- 모델
- 잔차가 normal distribution을 따르지 않는다
  - units에 영향을 주는 알 수 없는 feature들이 (disturbance) 충분히 많지 않아 정규분포로 수렴하지 못함
  - 잔차끼리 독립이 아니다 -> 선형 회귀 모형이 아닌 시계열 모형을 사용해야함



## 06. MOVING AVERAGE



## Walmart Recruiting II: Sales in Stormy Weather

Predict how sales of weather-sensitive products are affected by snow and rain

485 teams · 3 years ago

[Data](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#)

## My Submissions

### Late Submission

recent submission

Submitted

a few seconds ago

### Wait time

5 seconds

### Execution time





4 seconds

Score

0.10568

e

[View your position on the leaderboard ▼](#)

272	▼14	63614		0.15659	20	3
273	▲3	81010		0.15800	69	3
274	▼2	mnrl		0.15944	2	3
275	▲3	sna		0.15984	8	3

