

# L.POINT COMPETITION 제출폴더에 대한 설명입니다.

## 꼭 먼저 읽어주세요!

※ R code, data file에서 깨짐 현상이 발생할 경우, 인코딩 방식을 CP-949나 UTF-8로 맞춰주시길 바랍니다. 해결되지 않을 경우 [ykwon.0407@gmail.com](mailto:ykwon.0407@gmail.com) 으로 연락 주세요.

### 1. data

- 대회에 주어진 자료와 실제 분석에 사용한 가공된 자료로 이루어져있습니다.
- 파일 이름에 (UTF)가 있는 것은 같은 내용이지만 인코딩 방식만 다른 것입니다. (Windows와 Mac에서 같이 작업한 경우에 사용했습니다.)
- Demo.txt, list.txt 는 대회에서 주어진 자료입니다.
- seoul.csv file은 기상청에서 제공하는 서울 날씨를 (2014년 1월 1일 ~ 2015년 12월 1일)분석에 사용할 수 있도록 가공한 파일 입니다. (자료는 일체 변경하지 않고 서식만 가공)
- potato\_ttl.txt, potato\_week.txt, potato\_day.txt 파일은 R 코드 data\_generate.R 파일을 참조해주세요.

### 2. result

- 분석보고서 ppt에 사용된 그림들과 분석을 하면서 필요한 결과들을 모두 담았습니다.
- 파일이름에 'eda\_' 가 앞에 붙은 파일은 potato\_eda.R 에서 생성되는 파일입니다. data\_generate.R 파일에서도 조금 생성됩니다.
- 파일이름에 'eda2\_' 가 앞에 붙은 파일은 potato\_eda2.R 에서 생성되는 파일입니다.
- 파일이름에 'eda3\_' 가 앞에 붙은 파일은 potato\_eda3.R 에서 생성되는 파일입니다.
- 파일이름에 'day\_' 가 앞에 붙은 파일은 일별분석의 결과로 나온 파일입니다.
- 파일이름에 'week\_' 가 앞에 붙은 파일은 주별분석의 결과로 나온 파일입니다.

### 3. doc

- 최종 분석보고서 pdf 파일이 있습니다.

## 4 그 외 R 코드

반드시 `setwd()`를 잘 설정해주세요. Svg 파일에서 한글이 깨지는 현상이 있습니다. 한글로 보고 싶으시면 png 파일로 저장하시면 됩니다.

`gg_color_hue()` 함수는 거의 모든 R file에 있습니다. 이 함수는 ggplot을 이용 시, 상품별로 색깔을 맞춰주기 위한 함수입니다.

- `data_generate.R` 은 주어진 자료로부터 분석에 사용할 자료로 만들고, ppt 4~5쪽에 해당하는 전처리 과정을 합니다. `potato_ttl.txt` 는 이 모든 정보를 담고, `potato_week.txt`, `potat_day.txt`는 상품별로 26주 기간의 주별, 일별합계를 data 폴더에 저장했습니다. 전처리 과정에 사용된 plot은 result 폴더에 저장했으며 'eda\_'라는 prefix가 붙습니다.
- `potato_eda.R` 은 제출보고서의 6~7쪽에 해당하는 코드입니다. 이 R 코드로 생성되는 모든 생성 파일은 'eda\_' 라는 prefix가 붙습니다.
- `potato_eda2.R` 은 제출보고서의 8~11쪽에 해당하는 코드입니다. 이 R 코드로 생성되는 모든 생성 파일은 'eda2\_'라는 prefix가 붙습니다.
- `potato_eda3.R` 은 제출보고서의 12쪽에 해당하는 코드가 입니다. 이 R 코드로 생성되는 모든 생성 파일은 'eda3\_' 라는 prefix가 붙습니다.
- `potato_analysis_day.R` 은 forecast시 일별합계로 분석한 파일입니다. PART 2의 forecast 분석에 사용된 결과와 그림을 만듭니다. 이 R 코드로 생성되는 모든 생성 파일은 'day\_' 라는 prefix가 붙습니다.
- `potato_analysis_week.R` 은 forecast시 주별합계로 분석한 파일입니다. 결과에는 일별합계로 분석한 것이 더 좋게 판단되어 ppt 파일에 들어가는 결과는 없습니다. 이 R 코드로 생성되는 모든 생성 파일은 'week\_' 라는 prefix가 붙습니다.