

Assignment Specification

Kim, Yang Sok
Dept. of MIS, Keimyung University

Context



- A grocery store sells various products such as tea, coffee, soft drink, water, alcohol, milk etc.

	A	B	C	D	E	F	G
1	Classification	Product Code	Product Name	Price	Quantity	Sales	Date
2	Tea	7031763.0	광동)헛개차500ML	1500.0	1.0	1500.0	20170101.0
3	Coffee	7019784.0	칠성)칸타타아메리카노175ML	1200.0	1.0	1200.0	20170101.0
4	Soft Drink	8006731.0	풀무원)아임리얼토마토190ML	2900.0	1.0	2900.0	20170101.0
5	Water	8006911.0	코레버)행복할수500ML(진로)	500.0	1.0	500.0	20170101.0
6	Tea	7012629.0	남양)17차340ML	1200.0	1.0	1200.0	20170101.0
7	Soft Drink	7019607.0	코카)코카콜라500ML	2000.0	3.0	6000.0	20170101.0
8	Coffee	7014363.0	동서)스타벅스모카병281ML	3000.0	1.0	3000.0	20170101.0
9	Coffee	7016193.0	동서)스타벅스더블샷200ML	1500.0	1.0	1500.0	20170101.0
10	Water	7013526.0	진로)석수500ML	800.0	1.0	800.0	20170101.0
11	Alcohol	1000007.0	오비)카스캐255ML	1800.0	1.0	1800.0	20170101.0



- The owner of this store thinks that weekend and holiday impact on the sales.
- The owner wants to get insightful finding by analyzing the sales data. For example, he wants to know whether sales amount weekend is really larger than the other days. This may different between different product types(classifications).



- In addition, the owner of this store thinks that weather and air quality impact on the sales.

Weather Condition

	A	B	C	D	E	F
1	YearDayTime	Temp	Rain(mm)	Wind(m/s)	Humidity(%)	Visibility(10m)
2	2017-01-01 1:00	2.1		1.1	80	1949
3	2017-01-01 2:00	1.6		1	82	1847
4	2017-01-01 3:00	1.3		1.9	83	1857
5	2017-01-01 4:00	1.2		1.9	83	1899
6	2017-01-01 5:00	0.7		1.9	85	1680
7	2017-01-01 6:00	0.7		3	84	1596
8	2017-01-01 7:00	0.8		3.6	83	1717
9	2017-01-01 8:00	1.5		1	80	1744
10	2017-01-01 9:00	3.9		2.4	70	2056

Air Quality

Air Quality				D	E	F	G
1	YearDayHour	SO2	CO	O3	NO2	PM10	PM25
2	2017010101	0.003	0.7	0.019	0.022	35	23
3	2017010102	0.004	0.6	0.021	0.019	36	25
4	2017010103	0.003	0.6	0.023	0.016	42	28
5	2017010104	0.002	0.6	0.018	0.019	39	27
6	2017010105	0.003	0.6	0.02	0.018	42	27
7	2017010106	0.003	0.6	0.02	0.018	40	26
8	2017010107	0.003	0.7	0.026	0.017	43	27
9	2017010108	0.003	0.6	0.027	0.018	41	27
10	2017010109	0.003	0.7	0.021	0.023	45	30

Sales



- The owner wants to know sales amount in the next day by using the models derived from data using data mining techniques.
- Although the owner wants to predict sales by each product, this project only wants to predict sales amount by classifications (e.g., Tea, Coffee, Soft Drink, Water, Alcohol, Milk),
- You can evaluate the model by using the given data and also you need to provide the predicted sales by classifications on the 1st of December, 2017.
- We assume that the next day's weather, air quality and disease are known when we try to predict sales.

Tasks



- Try to convert three different dataset into single dataset for the project.
- Transform or derive attributes appropriate for the analysis.
- Summarizes what you did to prepare dataset.



- Summarizes sales characteristics by classifications, by month, by week of days.
- Using visualization tools, analyze trends of sales.
- Report any interesting findings.



- Design modeling for test.
- Create prediction models using various data mining techniques.
- Note that various techniques can be applied to the same problem.
- Explaining modeling results. In particular, if you use explainable models such as linear regression, you need to explain how the model can be explained.



- You need to predict the sales amount for the 1st of December, 2017 with the best model.

Classification	Sales
Alcohol	
Coffee	
Milk	
Soft Drink	
Tea	
Water	
Total	

Submission



- **Project Plan**
 - Describe your project and work plan
- **Data Preparation Report**
 - Describe how you prepared for data mining
- **Data Understanding Report**
 - Summarize data exploration results
- **Modeling Results**
 - Test design, Model Description, Performance Comparison, Model explanation
- **Prediction Results**

Write your report
with Power Point.
(About 30 pages)



- Due date: Week 13, 11/28 (Thursday) 12:00
- Submit the assignment submission page of Teaching and Learning System (교수학습시스템)
- You also need to submit the printed documents at the class.
- Team will be created during the class.

No pains, no gains.



QUESTIONS?