

매장 관리 방안

Walmart의 “선택적” & “보편적” 매장 관리 방안



Contents

- 01 [Introduction] Data 소개
- 02 “선택적” 매장 관리 방안
 - Data Preprocessing + EDA
 - Data Modeling
- 03 “보편적” 매장 관리 방안
 - Association Analysis
- 04 [Conclusion] Suggestion to Walmart

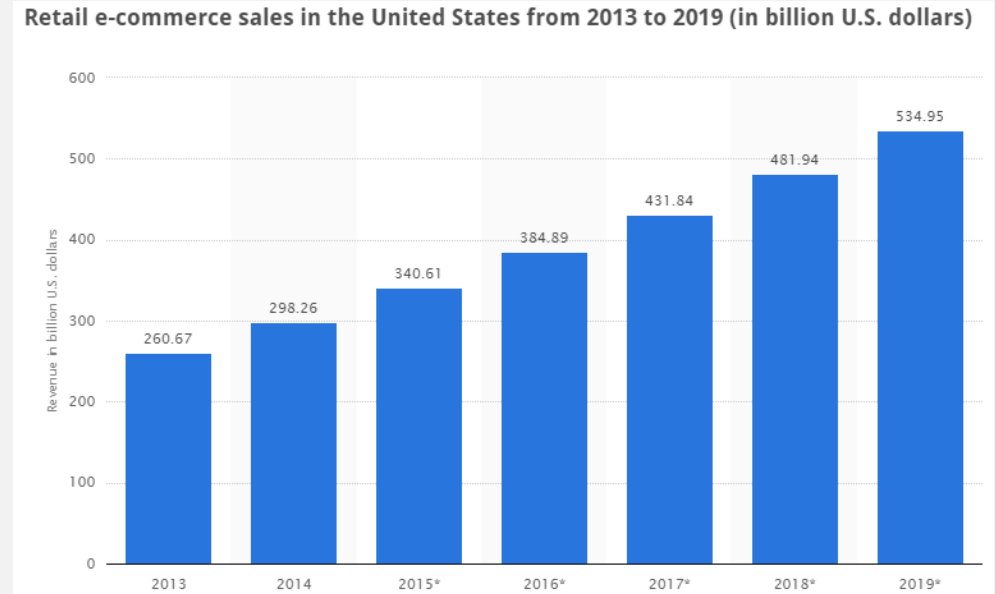


01.Data Explanation



월마트 주식회사는 미국에 본사를 둔 유통 업체이다. 포춘 500 선정 2007년 매출 기준 세계 최대의 기업이다. 1962년 샘 월턴이 아칸소주에 작은 잡화점을 시작한 것에서 월마트의 역사는 시작된다.

Gradually Increase in **ONLINE**... how about **OFFLINE**?



Walmart의 ONLINE상의 매출은 꾸준히 증가하고 있다.

과연 OFFLINE 매장에 대한 관리는 잘 이루어 지고 있는가?



01.Data Explanation

Dataset 1 : Store

- Features : Type & Size

* Type : A,B,C 매장 종류

* Size : 매장의 면적

- data size : (45,3)

Dataset 2 : Sales

-Features : Dept & Date & Sales & IsHoliday

* Dept : 부서 number

* Date : weekly date (2010-02-05 ~ 2012-10-26)

* Sales : weekly sales

* IsHoliday : 공휴일 유무

- data size : (421570,5)

Dataset 3 : Features

-Features : Store & Date & Temp & Fuel_Price & CPI & Unemployment rate & IsHoliday

* Temp : temperature (기온)

* CPI : Consumer Price Index (물가 지수)

- data size : (45,3)

세 개의 Dataset을 Merge하여 활용



01.Data Explanation

```
> head(data)
  Store      Date Dept Weekly_Sales IsHoliday.x Temperature Fuel_Price      CPI Unemployment IsHoliday.y Type      Size
1     1 2010-02-05     1    24924.50        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
2     1 2010-02-05    26    11737.12        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
3     1 2010-02-05    17    13223.76        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
4     1 2010-02-05    45         37.44        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
5     1 2010-02-05    28     1085.29        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
6     1 2010-02-05    79    46729.77        FALSE         42.31        2.572 211.0964         8.106        FALSE     A 151315
```

data size : (421570, 12)

매장 **내&외적인 상황**을 모두 고려하여 분석을 진행!

[매장 **내** 지표] 매장, 부서, 매출량, 매장 면적, 매장 Type

[매장 **외** 지표] 날짜, 휴일 유무, 기온, 기름값, 물가, 실업률

최종 DATASET



02.선택적 매장관리 Explanation

“선택적” 매장 관리

- 모든 Walmart의 매장 중, **특히 어떤 매장에 신경** 써야 할 것인가?
- 해당 매장의 문제점을 해결하기 위해 **어떠한 점을 참고**하여야 하는가?

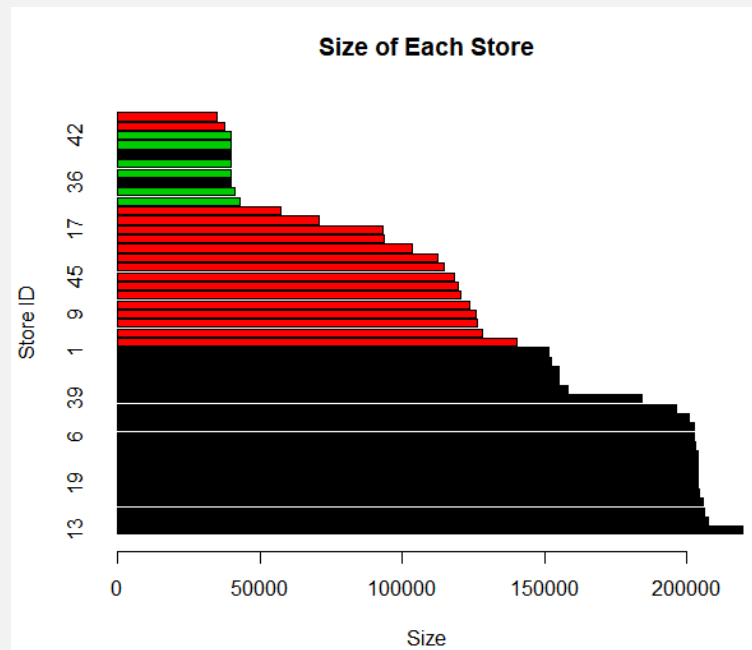
“보편적” 매장 관리

- 모든 Walmart의 매장에게 해줄 수 있는 공통적인 조언은?



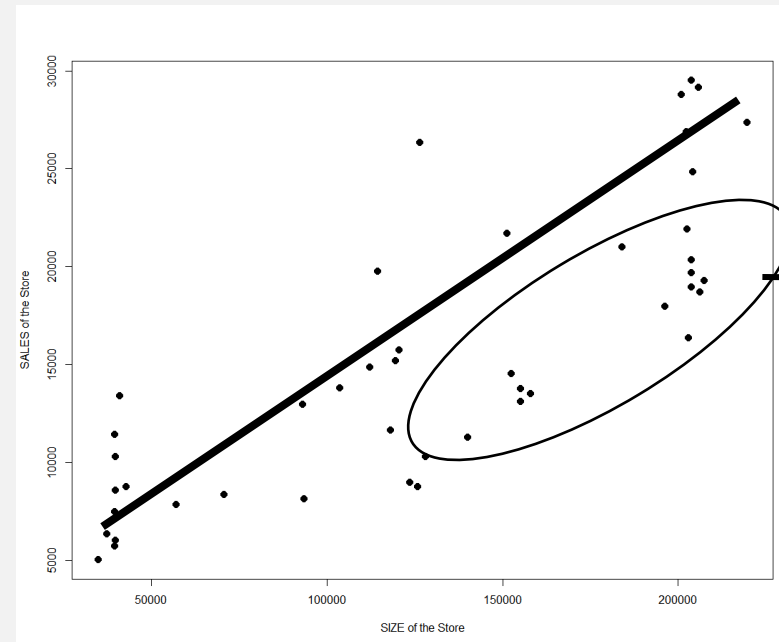
02.Data Preprocessing & EDA

어떤 매장을
특별히 관리할까?



Size of Each Stores

(Black : Type 'A', Red : Type 'B', Green : Type 'C')



Relationship between Sales & Size

한정된 자원을 모든 store를 관리하는데 쓰기 보다, 일부만 선택적으로 관리하는 방향으로!

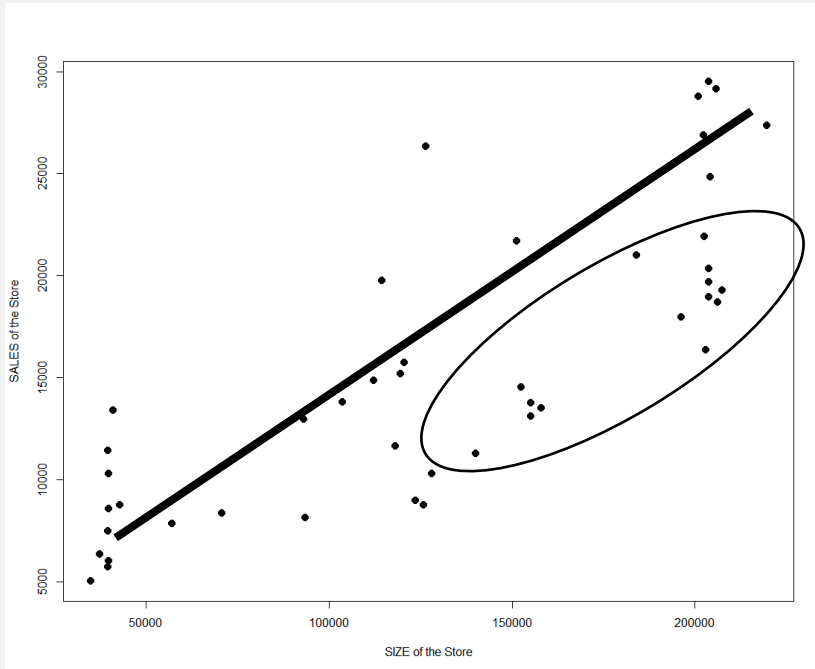
“돈을 많이 끌어올 수 있을 잠재력이 있는 매장!”

“ **LARGE size (큰 매장) + SMALL sales (낮은 매출) = 비효율적인 관리가 이루어지는 매장!** ”

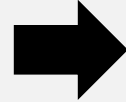




02.Data Preprocessing & EDA



Relationship between Sales & Size



	Store	mean_sales	Size	ratio
14	21	11283.44	140167	0.08049994
26	32	16351.62	203007	0.08054708
44	8	13133.01	155078	0.08468651
28	34	13522.08	158114	0.08552109
35	40	13763.63	155083	0.08875011
21	28	18714.89	206302	0.09071599
36	41	17976.00	196321	0.09156435
3	11	19276.76	207499	0.09290051
17	24	18969.11	203819	0.09306839
19	26	14554.13	152513	0.09542878

매장 규모 당 매출!

(Ratio = 평균 매출 / 면적)

매장이 얼마나 비효율적으로
운영되는지 알 수 있는 지표!

TOP 10 Worst Efficiency Stores

1. 가장 비효율적으로 (ratio가 낮은) 하위 10개 매장 관리하기!
2. 단지 비율만 고려하여 하위 10위를 정할 경우, 매출이 매우 작은 매장이 선별!

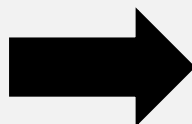
전체 매장 매출의 평균을 넘는 매장만을 대상으로!

“해당 10개의 Store Data로 모델 만들기”



02.Data Preprocessing & EDA

Dept	mean_sales
92	84533.97
95	75048.47
38	57640.22
90	53235.49
72	48438.45
65	45441.71
40	44473.36
2	42686.35
91	40809.97
94	37615.11



Dept	mean_sales
92	84533.97
95	75048.47
38	57640.22
90	53235.49



FOCUS on TOP 4 Department by "SALES"

한 매장 내의 모든 Department를 관리하는 것은
어려우므로, 상위 4개의 매출을 기록하고 있는
Top 4 Department에 집중!

< SUMMARY >

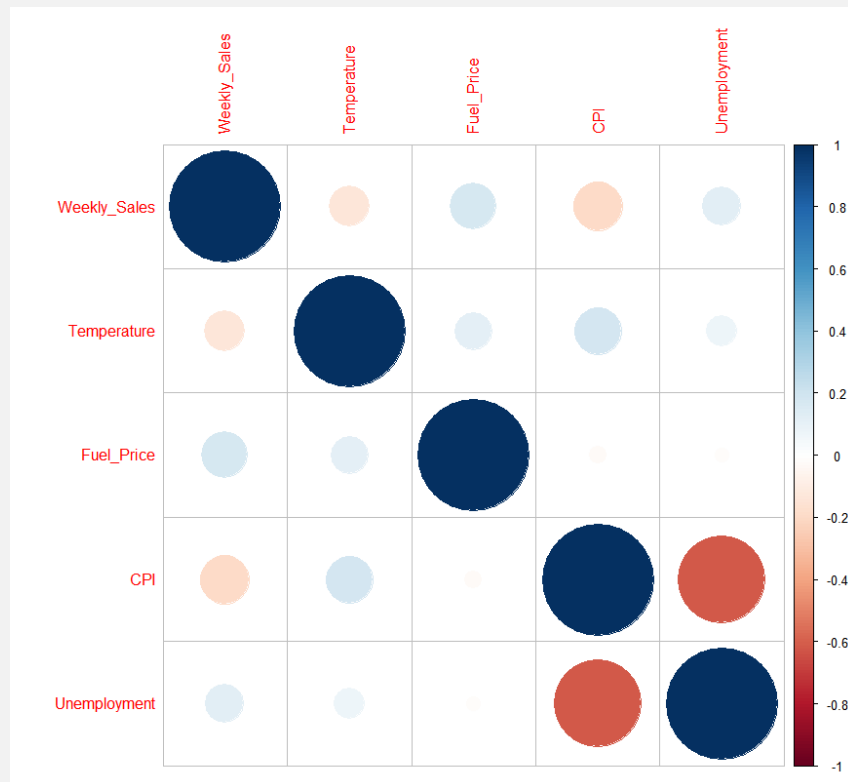
1. 10 개의 매장, 그 안의 4개의 부서 (총 40개의 부서 data)
2. Features : (1) Store (2) Department (3) Weekly_Sales (4) IsHoliday
(5) Temperature (6) Fuel_Price (7) CPI (8) Unemployment rate (9) Size



02.Data Preprocessing & EDA

Before modeling...

MULTI-COLINEARITY! 확인하기



CPI, Unemployment, Fuel_Price

-> 경제 상황과 밀접한 지수이기 때문에, 상관관계 확인 필요!

```
> cor(finaldata$Unemployment, finaldata$CPI)
[1] -0.42041
> cor(finaldata$Fuel_Price, finaldata$CPI)
[1] -0.1826564
> cor(finaldata$Fuel_Price, finaldata$Unemployment)
[1] 0.005432384
```

No Multi-collinearity Problem!

다행히 높은 상관관계를 보이는 변수들은 없었다!



02.Data Preprocessing & EDA

Standard Scaling

“ 변수간의 단위 규모의 차이가 크므로, scaling하여 분석을 진행!

표준정규분포로 모든 column을 scaling하기

Store	Dept	Weekly_Sales	IsHoliday.x	Temperature	Fuel_Price	CPI	Unemployment	Size
11	95	79557.01	FALSE	46.04	2.572	214.4249	7.368	207499
11	90	52189.57	FALSE	46.04	2.572	214.4249	7.368	207499
11	92	90342.72	FALSE	46.04	2.572	214.4249	7.368	207499
11	38	80470.89	FALSE	46.04	2.572	214.4249	7.368	207499
11	92	86857.10	TRUE	48.01	2.548	214.5748	7.368	207499
11	95	78778.26	TRUE	48.01	2.548	214.5748	7.368	207499

Store	Dept	Weekly_Sales	IsHoliday.x	Temperature	Fuel_Price	CPI	Unemployment	Size
11	95	0.4650440	FALSE	-0.6315060	2.572	1.125851	7.368	1.137058
11	90	-0.6006538	FALSE	-0.6315060	2.572	1.125851	7.368	1.137058
11	92	0.8850434	FALSE	-0.6315060	2.572	1.125851	7.368	1.137058
11	38	0.5006308	FALSE	-0.6315060	2.572	1.125851	7.368	1.137058
11	92	0.7493121	TRUE	-0.5267176	2.548	1.129689	7.368	1.137058
11	95	0.4347192	TRUE	-0.5267176	2.548	1.129689	7.368	1.137058





02.Data Preprocessing & EDA

Data Preprocessing SUMMARY

1. 3개의 dataset 합치기
2. '효율성 지표'를 통해 Worst 10 Stores 선별
3. '매출이 높은' 상위 4개의 부서 선별
4. Multi-Collinearity 문제 확인
5. 변수들의 규모 차이문제 해결 위해 Standard Scaling 진행



02.Data Modeling

Goal : Fixed Effect Model을 사용하여, 각 부서의 **“WEEKLY SALES”** 예측!

Model . **Fixed Effect Model**

- “부서” 를 factor로!
- 부서별로 어떠한 차이가 있는지 확인하기 위해서!



02.Data Modeling

result. FIXED-EFFECT MODEL

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.209079   0.085412  -2.448   0.0144 *
as.factor(Dept) 90  -0.171522   0.024346  -7.045 2.07e-12 ***
as.factor(Dept) 92   1.047252   0.024346  43.015 < 2e-16 ***
as.factor(Dept) 95   0.677883   0.024346  27.843 < 2e-16 ***
IsHoliday.xTRUE -0.018056   0.034310  -0.526   0.5987
Temperature    -0.102231   0.009583 -10.668 < 2e-16 ***
Fuel_Price      0.171771   0.019742   8.701 < 2e-16 ***
CPI             -0.321529   0.010687 -30.086 < 2e-16 ***
Unemployment    -0.093310   0.005010 -18.626 < 2e-16 ***
Size            0.534715   0.009643  55.449 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

각 estimate는, 1 standard deviation이 1 만큼
변화할 때 얼마나 Sales가 변화하는지를 나타냄!

(Holiday를 제외하고, 모두 유의미한 결과가 있음을 확
인할 수 있다)

분석 결과

1. 매출 & 매장 크기 : (+) 관계 & 매출 & 물가 : (-) 관계
-> 이는 당연한 결과일 뿐!
2. 공휴일이라는 점이 매출과 크게 관련 있다고 보기는 어렵다!
3. 92번 부서 - high sales / 90번 부서 - low sales



02.Data Modeling

매장에게 할 수 있는 Suggestion

1. 공휴일이라고 사람들이 매장을 자주 방문하는 것으로 보이지 않는다.

오히려, 그 계수는 음수로, 덜 방문하는 경향이 있다고 할 수도 있다.

-> **"With Family"라는 로고로, 공휴일에 가족들끼리 나와서 장을 보도록 유도해도 좋을 것이다.**

2. 부서 관리 차원

비효율적인 운영이 이루어지고 있는 매장 중, 높은 매출을 기록하고 있는 부서 중에서 **보다 신경써야 할 부서는 90번 부서**이다. 이들은 **92번 부서가 매장을 관리하는 방식을 벤치 마킹**하여 개선을 할 필요가 있다.



03.보편적 매장관리

“선택적” 매장 관리

- 모든 Walmart의 매장 중, **특히 어떤 매장에 신경** 써야 할 것인가?
- 해당 매장의 문제점을 해결하기 위해 **어떠한 점을 참고**하여야 하는가?

“보편적” 매장 관리

- 모든 Walmart의 매장에게 해줄 수 있는 공통적인 조언은?



03.Association Analysis

also called **Basket Analysis**

미국인들의 '장바구니' 분석을 통해 **구매 습관 파악!**

(사용 tool : python & Gephi)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon	antioxydant juice	frozen smoothie
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

1. "7501 의 고객 data"
2. "120 여 종의 음식"
3. "1row = 1customer"

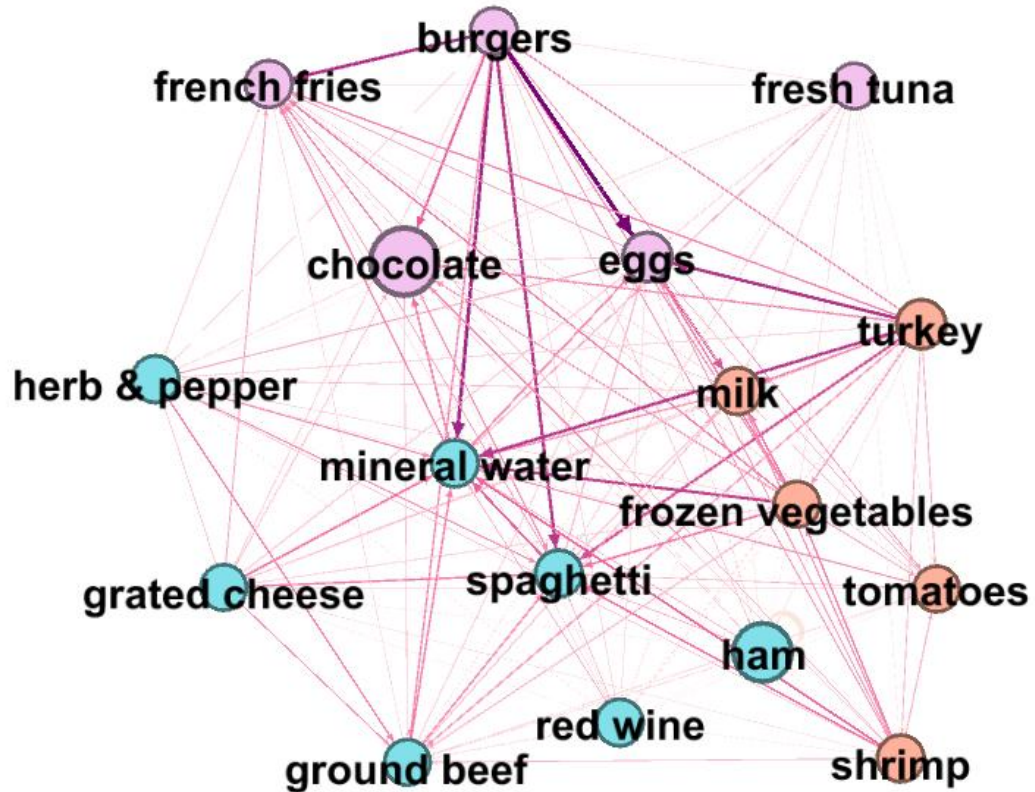
ex) customer id 2 : burger, meatball, egg

```
1 print(len(set(list(data.stack().values))))
2 print(set(list(data.stack().values)))

120
{'tomato juice', 'pet food', 'green grapes', 'milk', 'soda', 'escalope', 'energy bar', 'butter', 'pasta', 'ham', 'protein bar', 'ketchup',
'turkey', 'barbecue sauce', 'mayonnaise', 'soup', 'mineral water', 'whole wheat rice', 'green tea', 'chocolate bread', 'salmon', 'cookie s',
's', 'frozen smoothie', 'chicken', 'magazines', 'french wine', 'blueberries', 'nonfat milk', 'light cream', 'cream', 'yogurt cake', 'frozen vegetables',
'shrimp', 'sandwich', 'yams', 'almonds', 'brownies', 'extra dark chocolate', 'shampoo', 'sparkling water', 'french fries', 'mushroom cream
sauce', 'honey', 'burgers', 'muffins', 'flax seed', 'mint green tea', 'hand protein bar', 'green beans', 'black tea', 'salad', 'asparagu
s', 'bacon', 'fresh bread', 'dessert wine', 'strong cheese', 'bramble', 'antioxydant juice', 'oatmeal', 'light mayo', 'pickles', 'red win
e', 'spaghetti', 'mashed potato', 'pancakes', 'candy bars', 'blueberries', 'parmesan cheese', 'champagne', 'water spray', 'mint', 'cider',
'low fat yogurt', 'body spray', 'whole wheat pasta', 'tea', 'grated cheese', 'napkins', 'vegetables mix', 'zucchini', 'whole weat flour',
'chocolate', 'hot dogs', 'bug spray', 'herb & pepper', 'cooking oil', 'pepper', 'gluten free bar', 'cake', 'spinach', 'olive oil', 'salt',
'toothpaste', 'meatballs', 'energy drink', 'eggplant', 'chili', 'rice', 'shallot', 'tomato sauce', 'white wine', 'carrots', 'babies food',
'strawberries', 'burger sauce', 'oil', 'fresh tuna', 'tomatoes', 'fromage blanc', 'gums', 'clothes accessories', 'melons', 'corn', 'caulif
lower', 'avocado', 'asparagus', 'cottage cheese', 'eggs', 'chutney', 'cereals', 'ground beef'}
```



03.Association Analysis



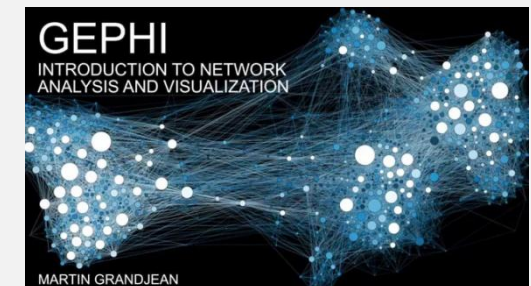
Connection = 함께 사는 물품

THICK edge = 두 물품을 함께 사는 경우가 많음을 의미

SAME color = 같은 부류의 음식

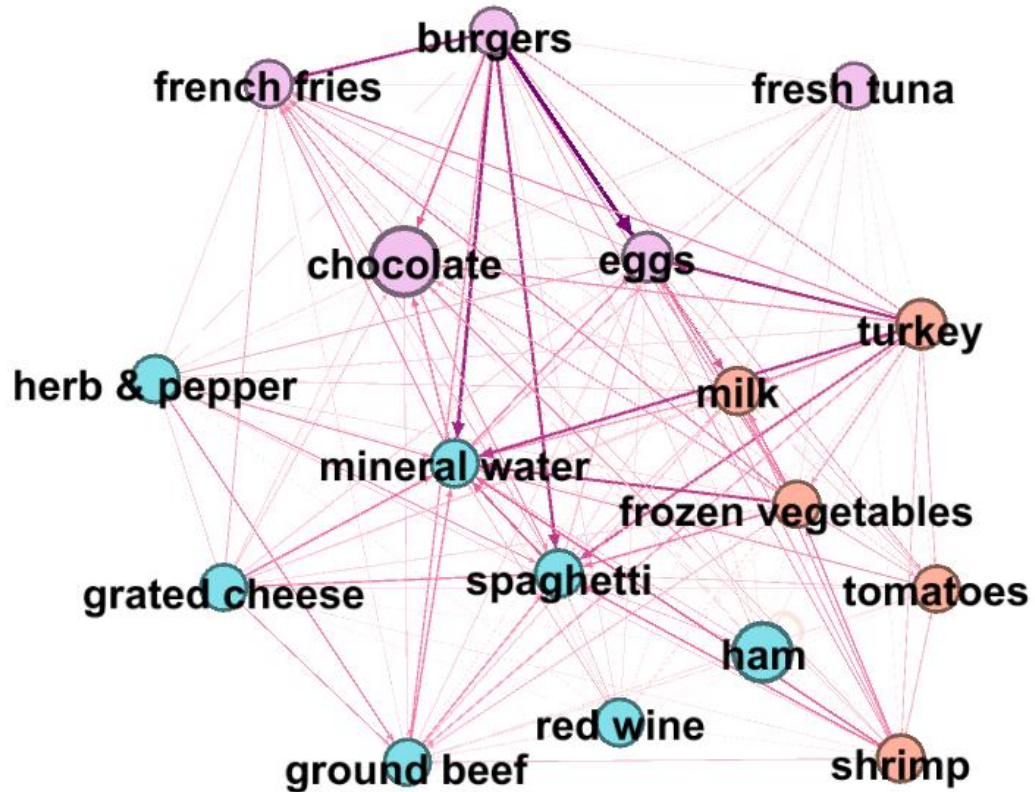
(같은 부류 : 구매 패턴이 비슷한 음식)

BIG node size : 함께 사는 물품의 종류가 많음을 의미!





03.Association Analysis



함께 같이 구매하는 물품/유사물품 끼리

“ 비슷한 곳에 진열하도록! ”

[1] Burger & Egg & Chocolate

[2] Spaghetti & ham

[3] Milk & Turkey & Frozen Vegetables 를

비슷한 곳에 위치시키면 구매를 유도할 수 있을 것이다.



04.Conclusion

Summary

1

Walmart의 “선택적 매장 관리”

- 비효율적으로 관리가 이루어 지고 있는 매장

1. “With Family” 홍보를 통한 공휴일에 가족과 함께 장보기를 유도 (+ 가족 할인)
2. “90번 부서의 문제상황 점검 + 92번 부서의 벤치마킹”



2

Walmart의 “보편적 매장 관리”

-미국인들의 장바구니 분석을 통해

3. 함께 구매하는 물품의 진열을 가까운 곳에! (자세한 품목은 앞 pg)

Solution



04.Conclusion

THANK YOU