

WHO is fatty?

Team#12

김성연 심성호 황세현

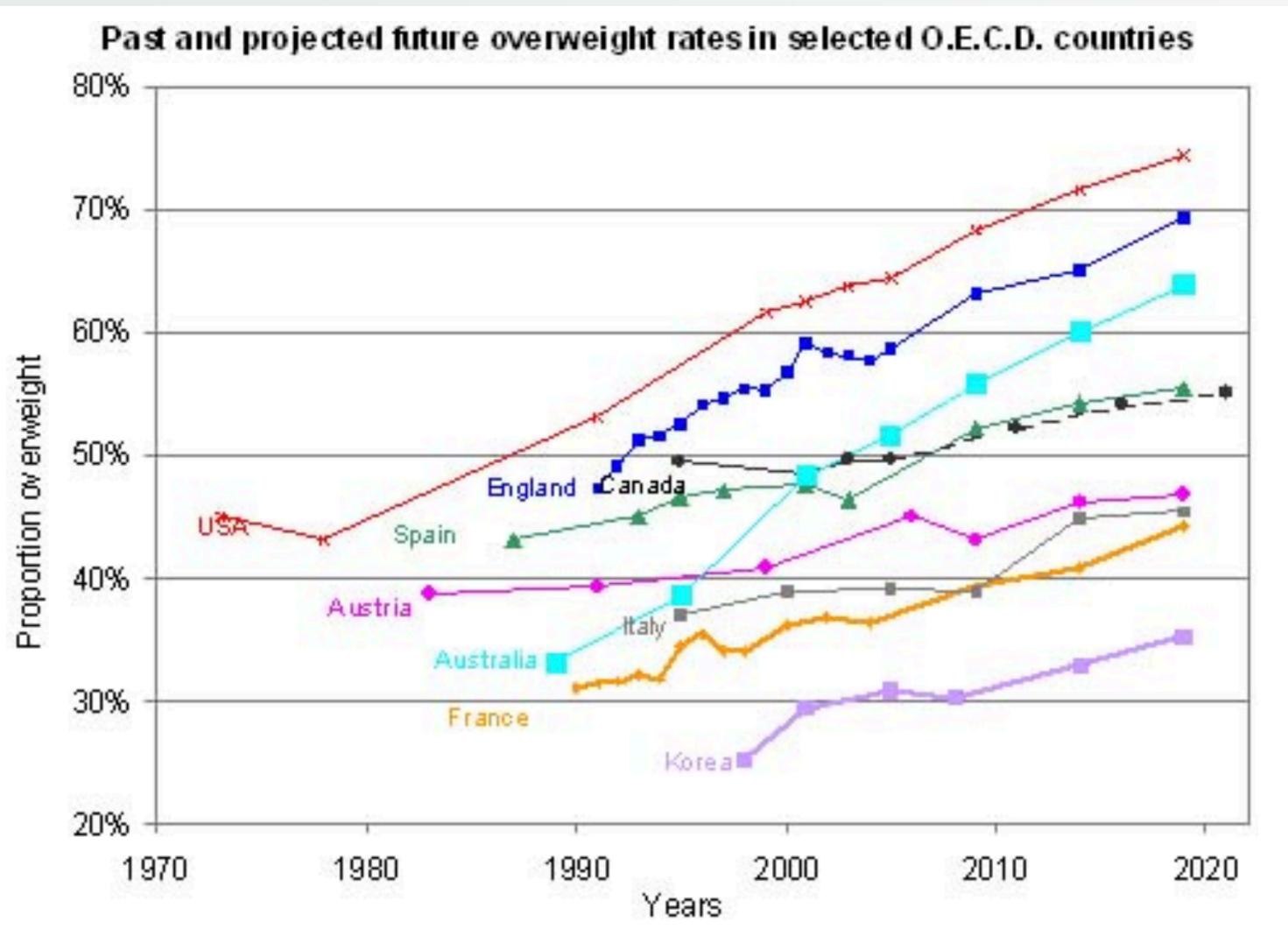




Contents

- Our goal
- Data sets
- Feature selection
 - ✓ Correlation, Random forest, ANOVA
- Prediction
 - ✓ MLP, Random forest, SVM with Balanced bagging
 - ✓ AUROC
- Poverty & BMI
- Conclusion

Problem



Who we are?

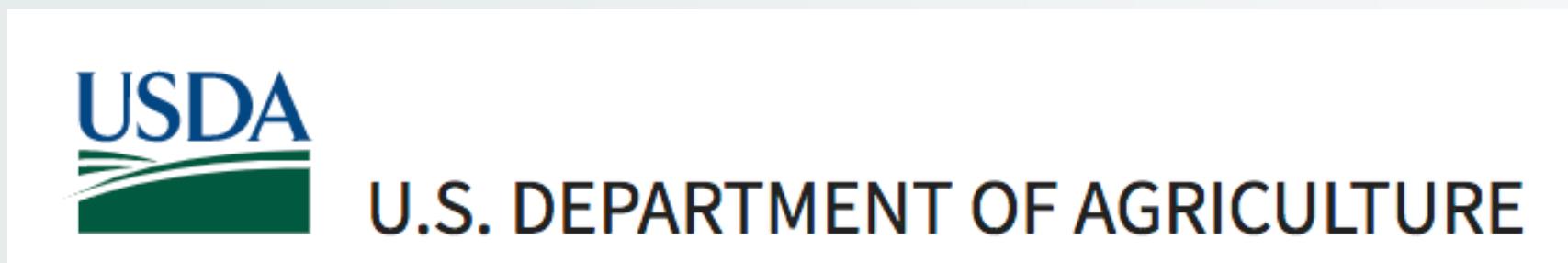
- 세계 보건 기구의 입장에서 빅데이터 분석을 통해 점점 증가하는 비만율을 관리하기 위한 정책을 제안하자 !



Data sets

- ‘American Time Use Survey (ATUS) eating and health module’ data in 2014 & 2015 from ‘**United States Department of Agriculture (USDA)**’

21,838 Rows & 37 Columns



Data refinement

- Remove meaningless columns
- Remove outlier of weight & height
- Imbalanced data handling (down sampling)
- Null value handling



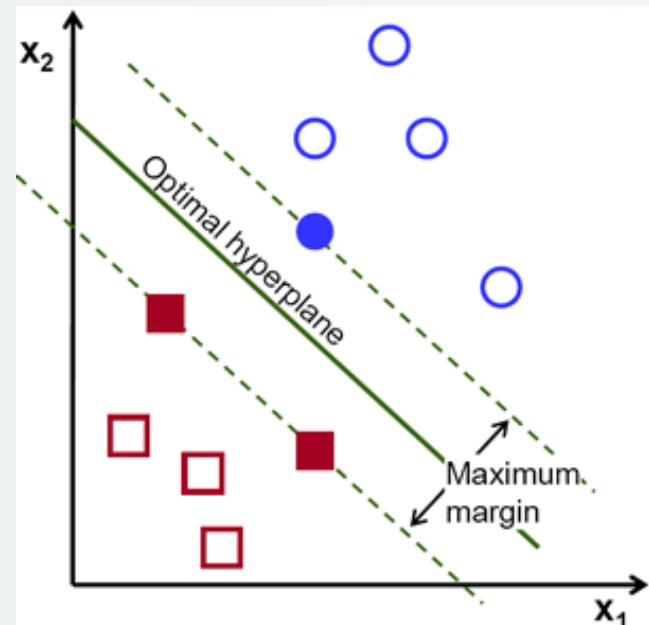
Feature selection

- Correlation, Random forest, ANOVA 를 통해 찾은 ‘**BMI**’(체질량 지수)와 연관성이 가장 높은 feature 5개
- ‘**eugenht**’ : 건강 자신감
- ‘**euexercise**’ : 운동 유무
- ‘**erincome**’ : 소득 분위
- ‘**eusoda**’ : 탄산음료 음용 유무
- ‘**eusnap**’ : 영양 보충 보조 프로그램 수혜 유무

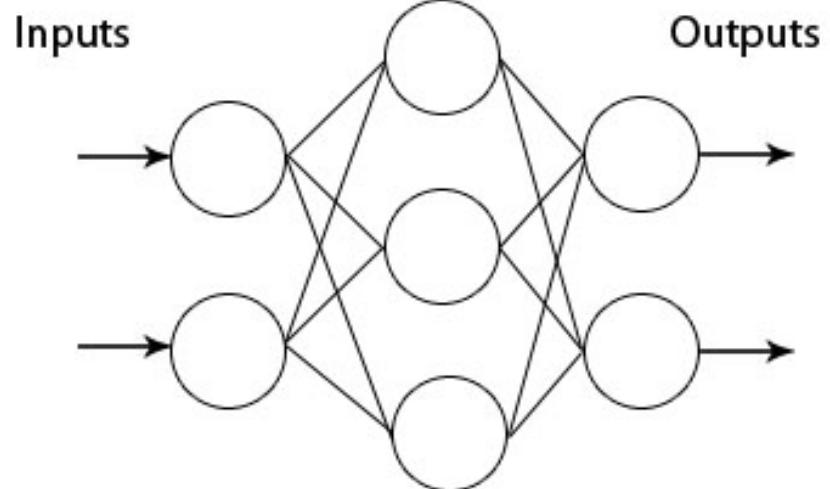
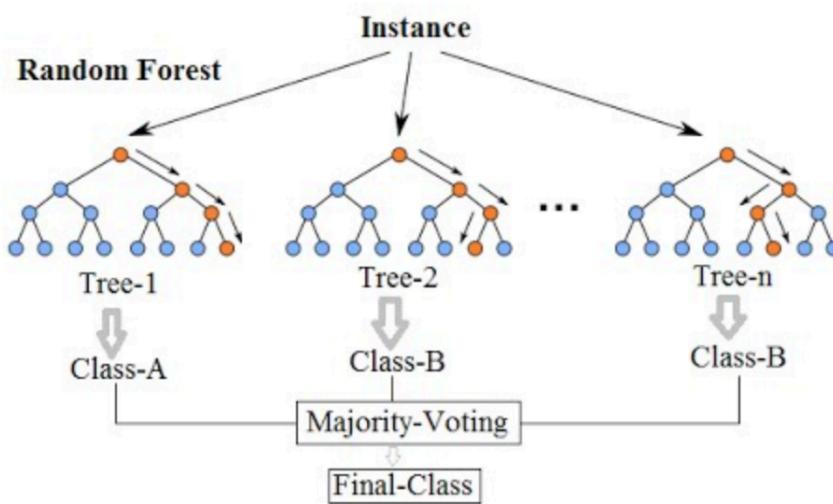


Prediction

- Random forest, SVM, MLP
- 비만유무($BMI \geq 30$) 예측

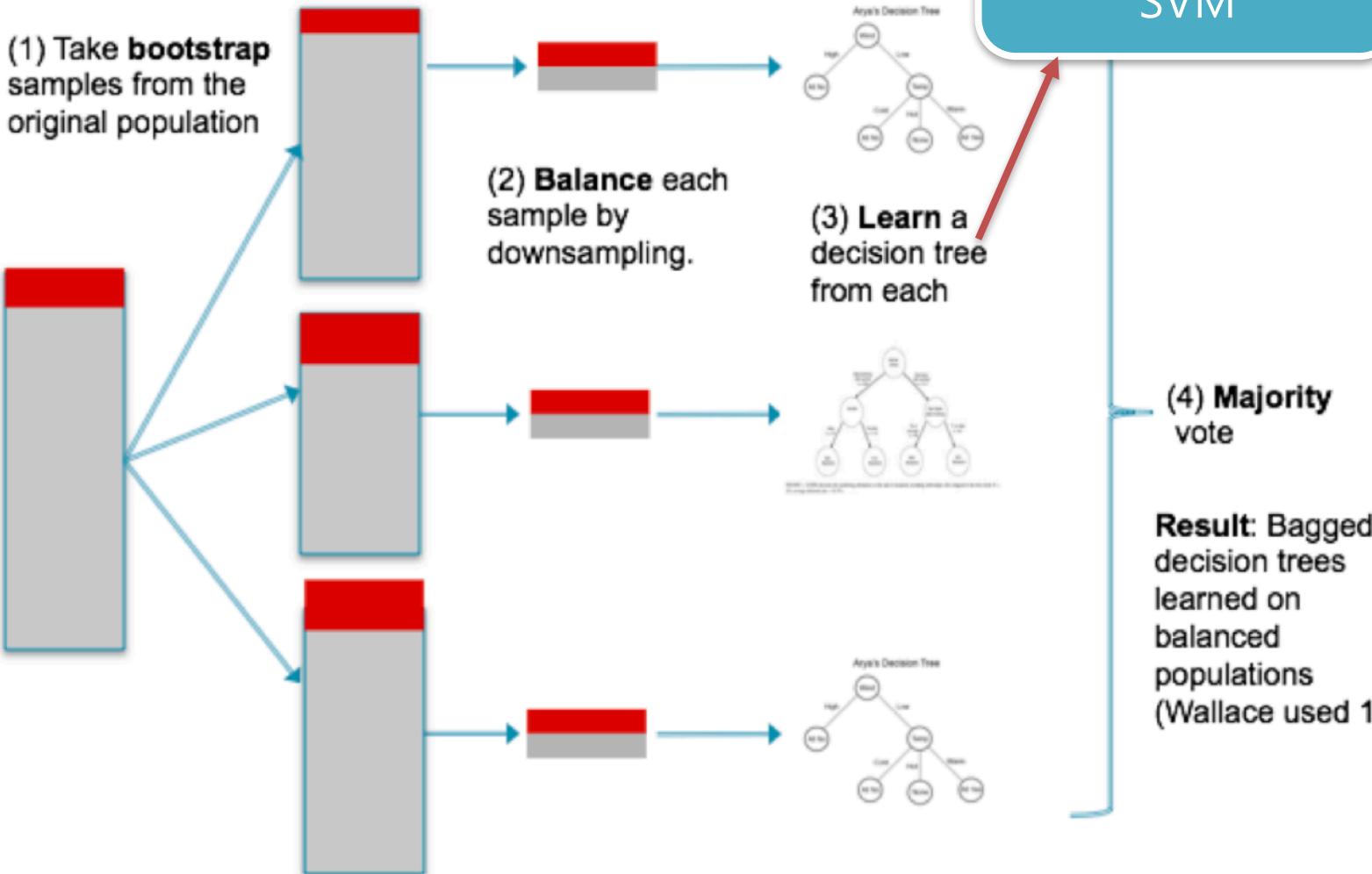


Random Forest Simplified



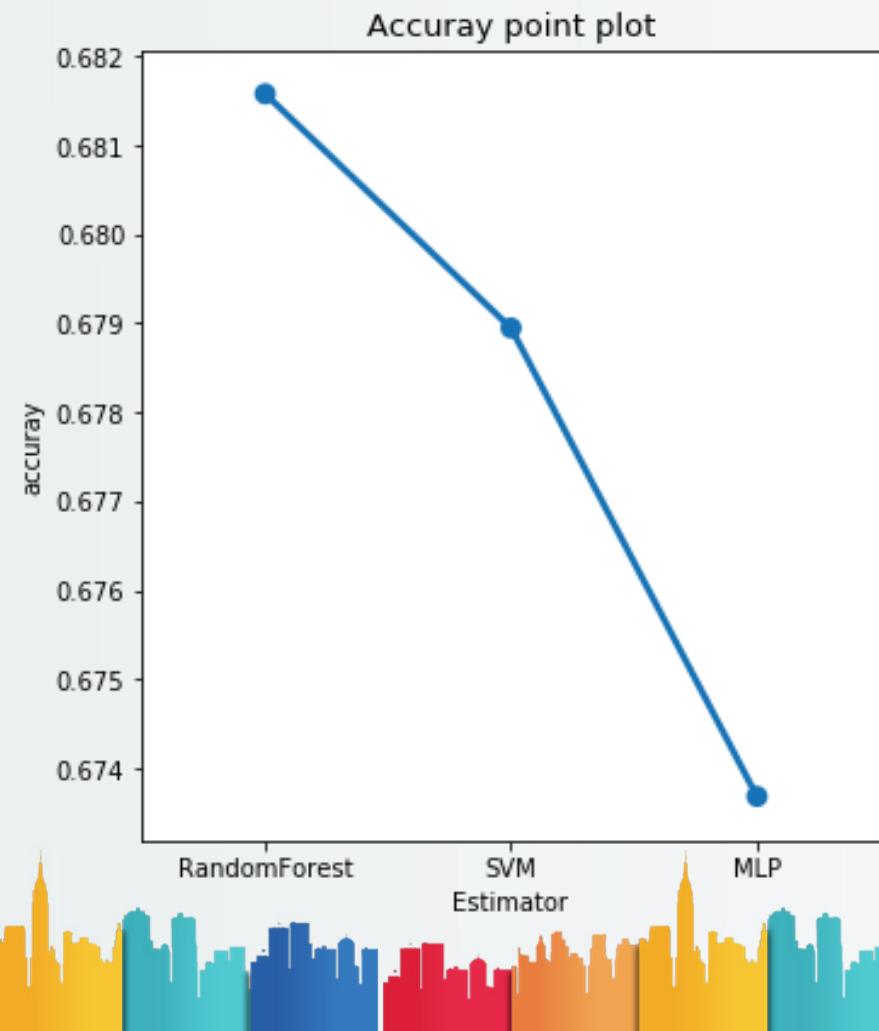
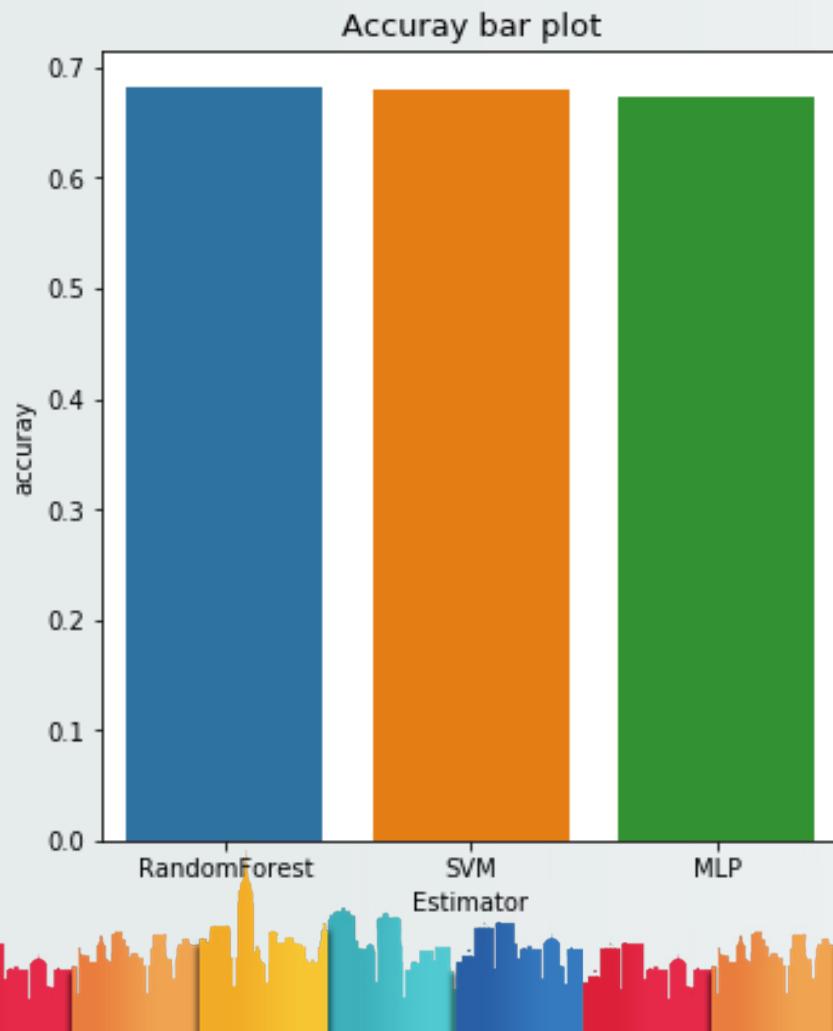
Balanced Bagging

MLP,
Random forest,
SVM



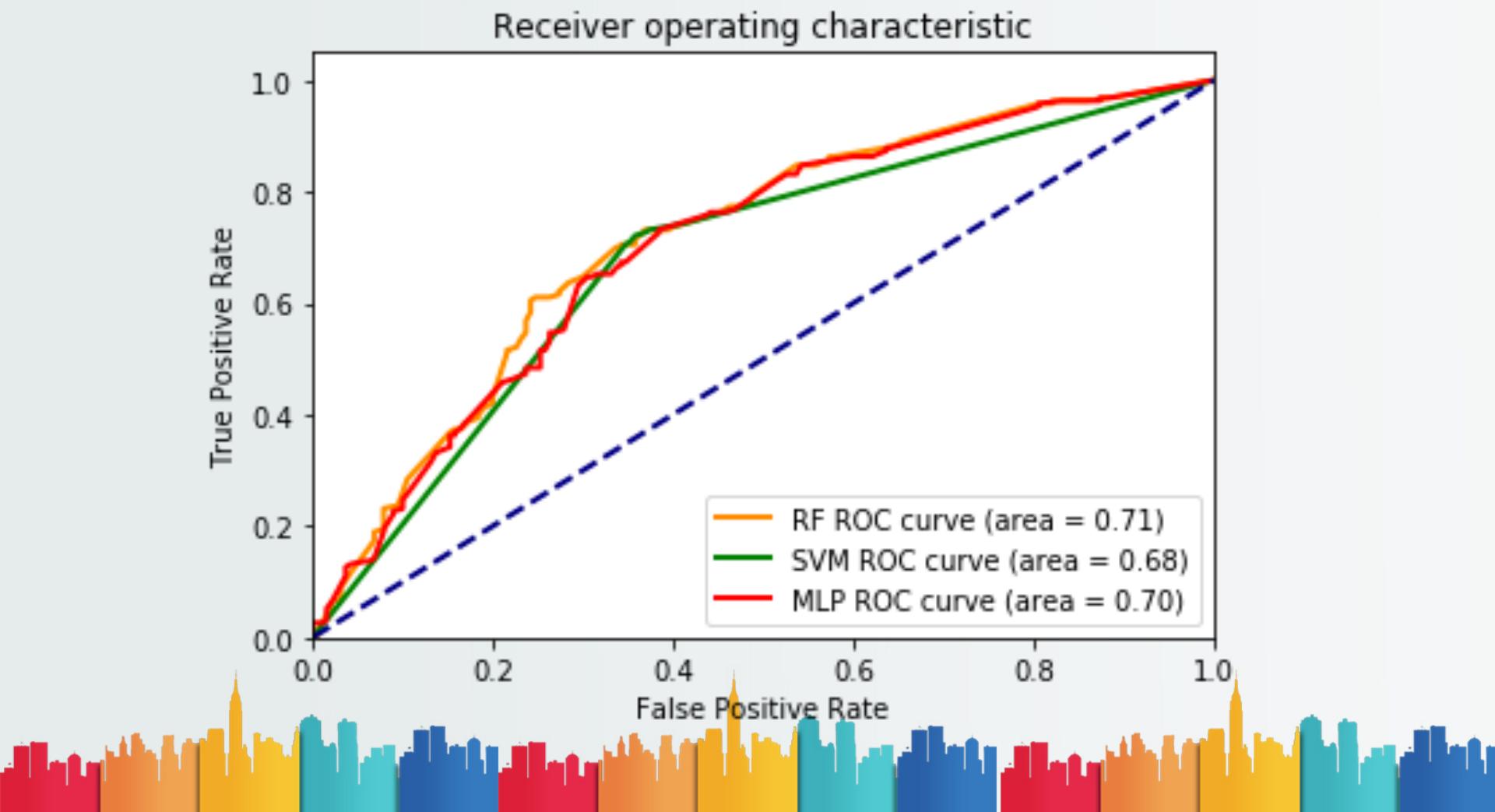
Prediction

- Random forest, SVM, MLP

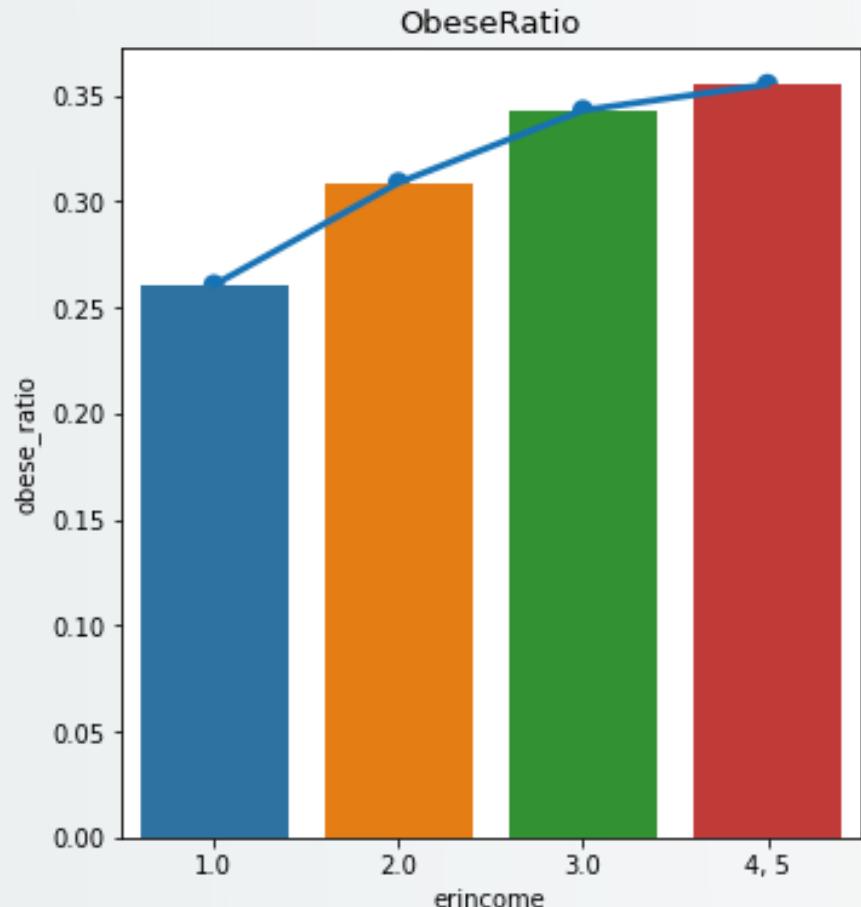
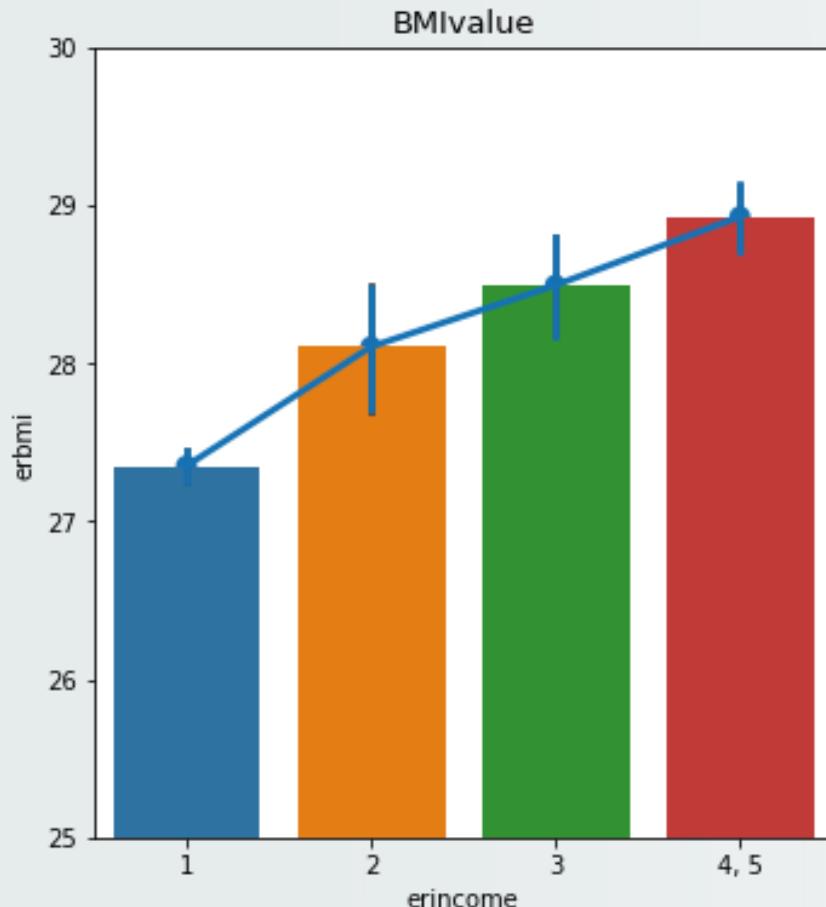


Prediction

- AUROC (Random forest, SVM, MLP)



Income-BMI & Income-Obese Ratio

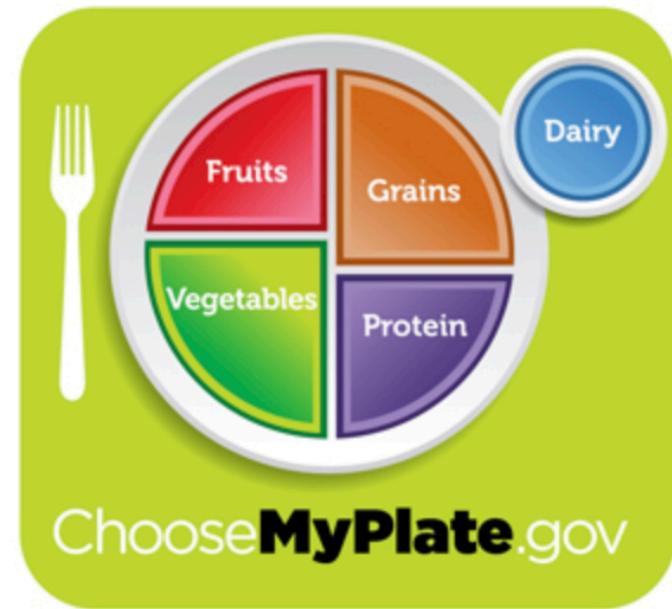


- Income이 적을 수록 BMI와 비만율이 높아진다.

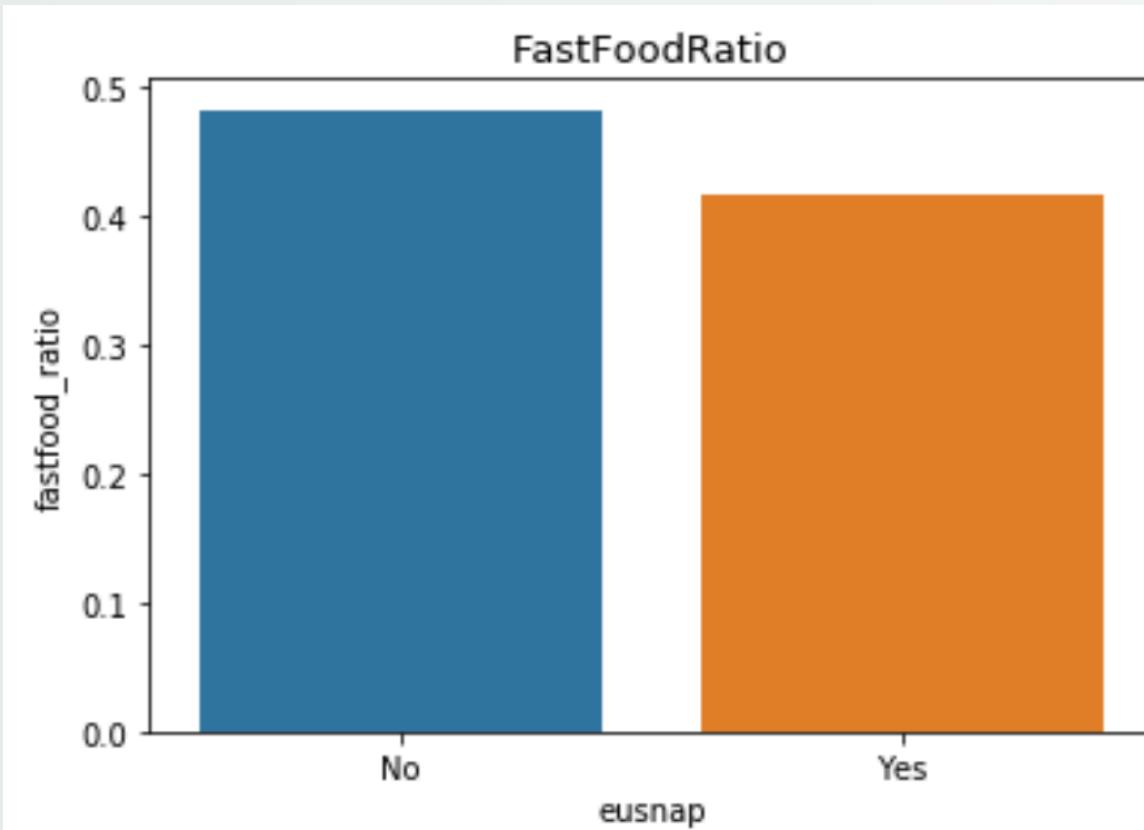
*erincome이 높을 수록 낮은 소득 분위이다.

SNAP

- Supplemental Nutrition Assistance Program is a federal nutrition program that helps low income people buy healthy food



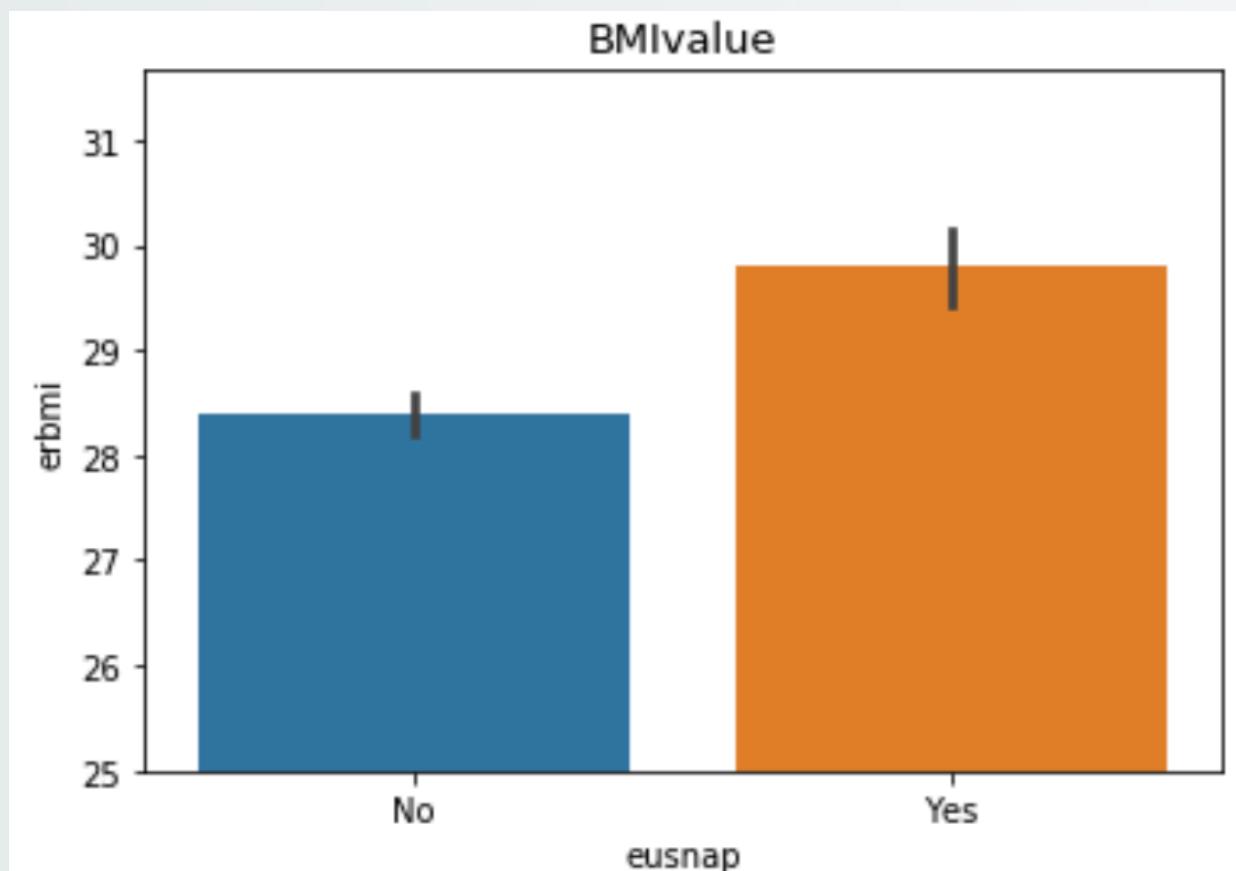
SNAP-Fast food ratio



- SNAP을 지원받는 사람들은 Fast food를 더 적게

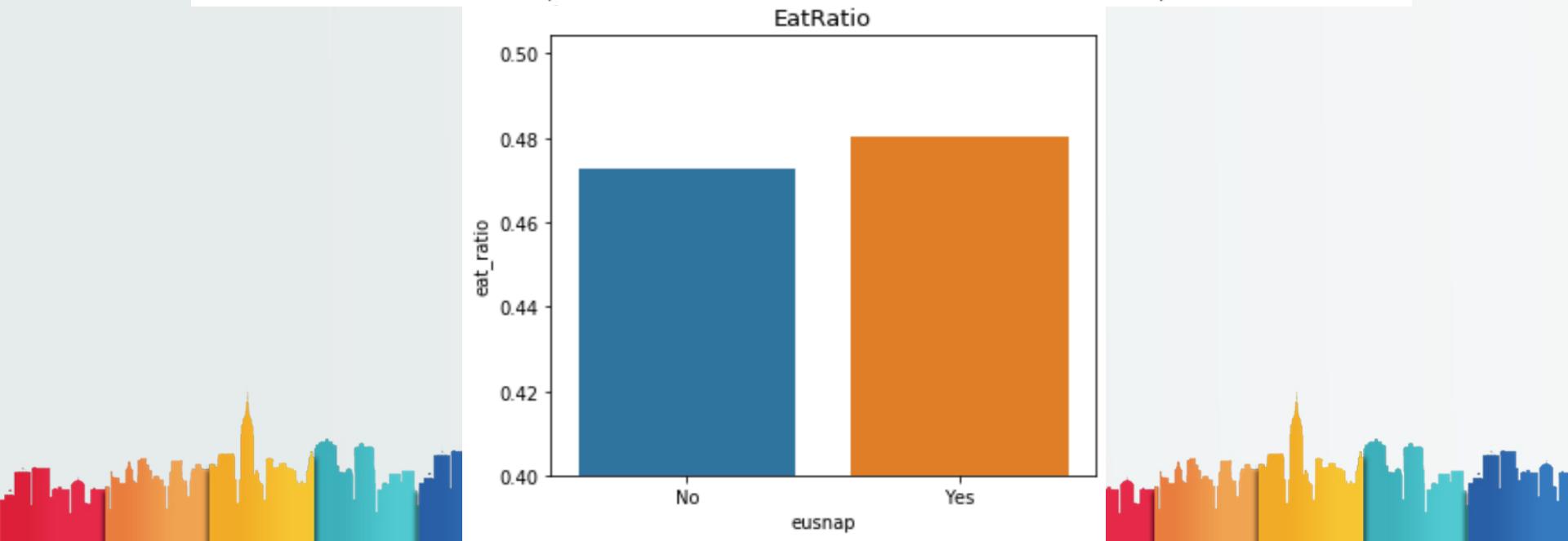
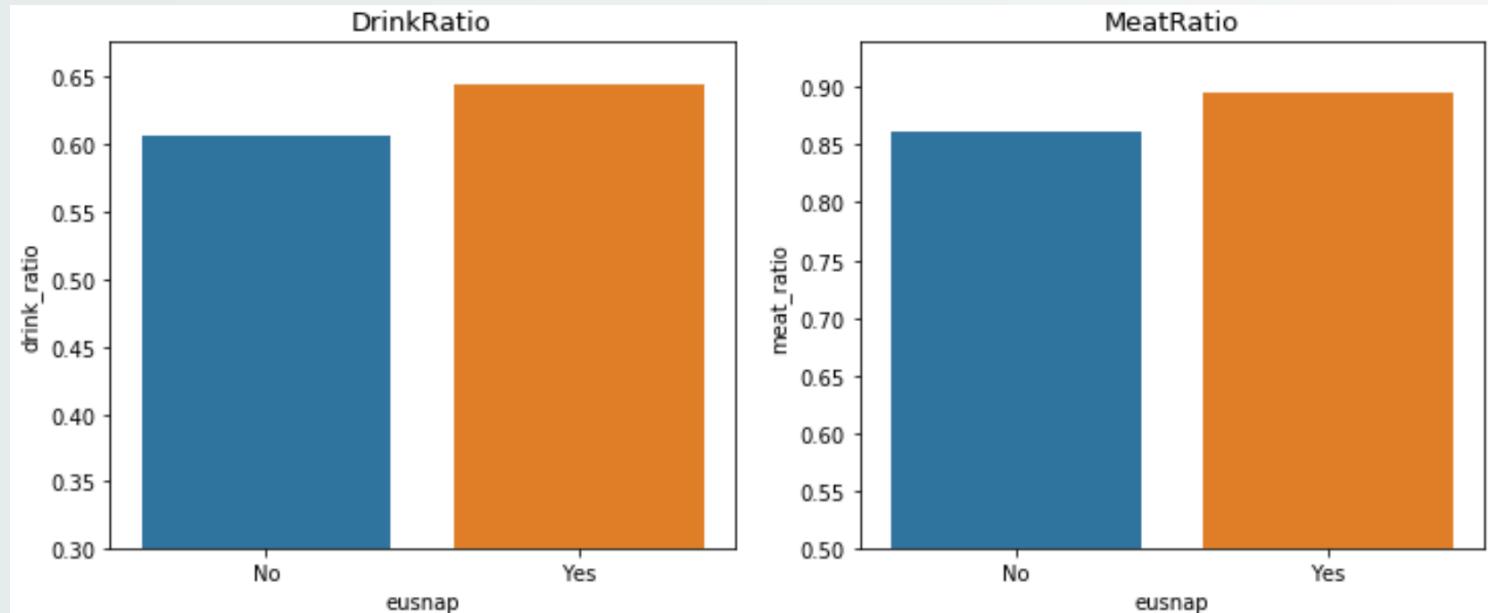
먹는다.

SNAP-BMI

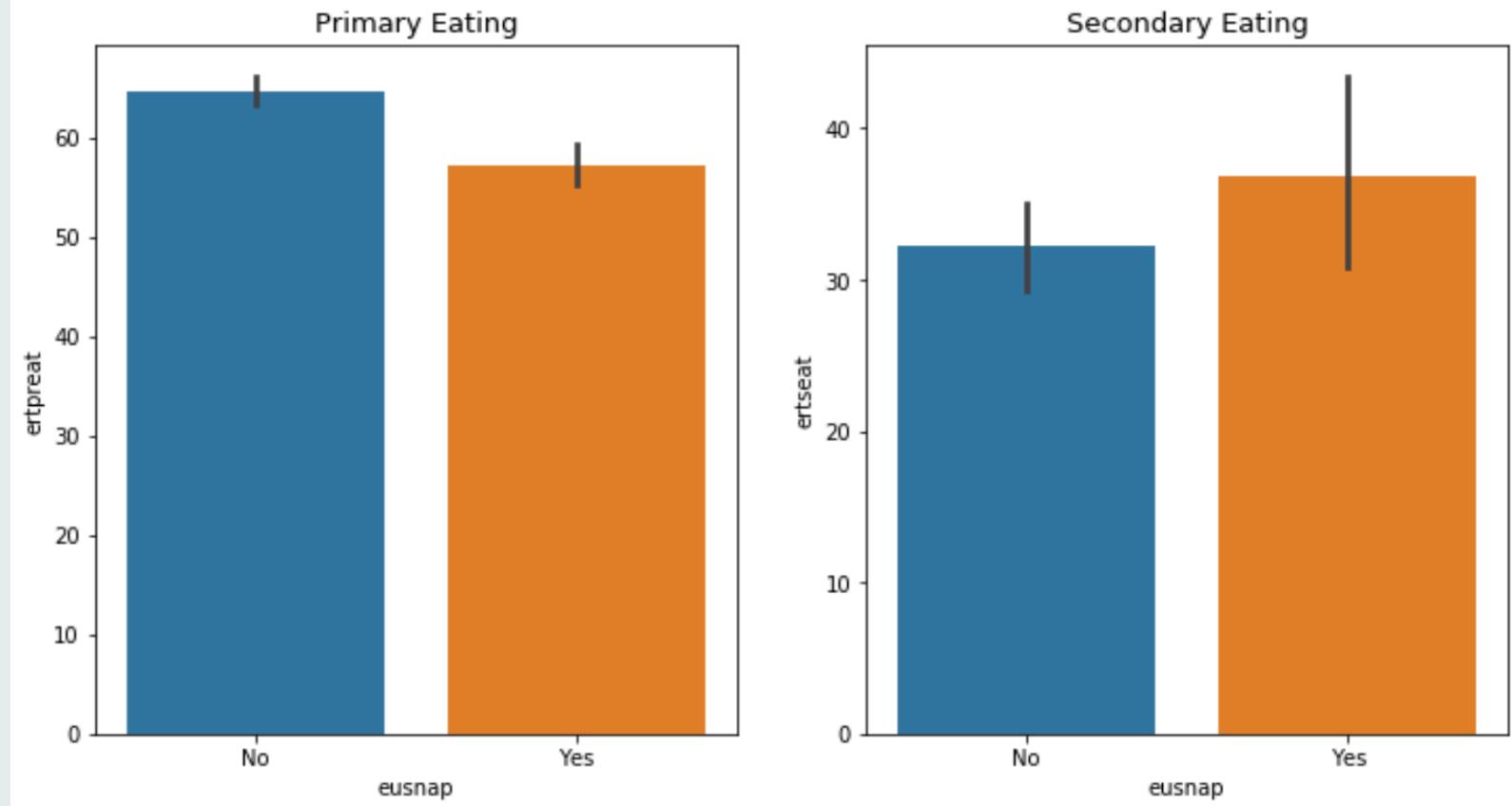


- SNAP을 지원받는 사람들이 오히려 BMI가 더 높다

SNAP-Drink & Meat ratio & Eat ratio



SNAP-Primary & Secondary eating

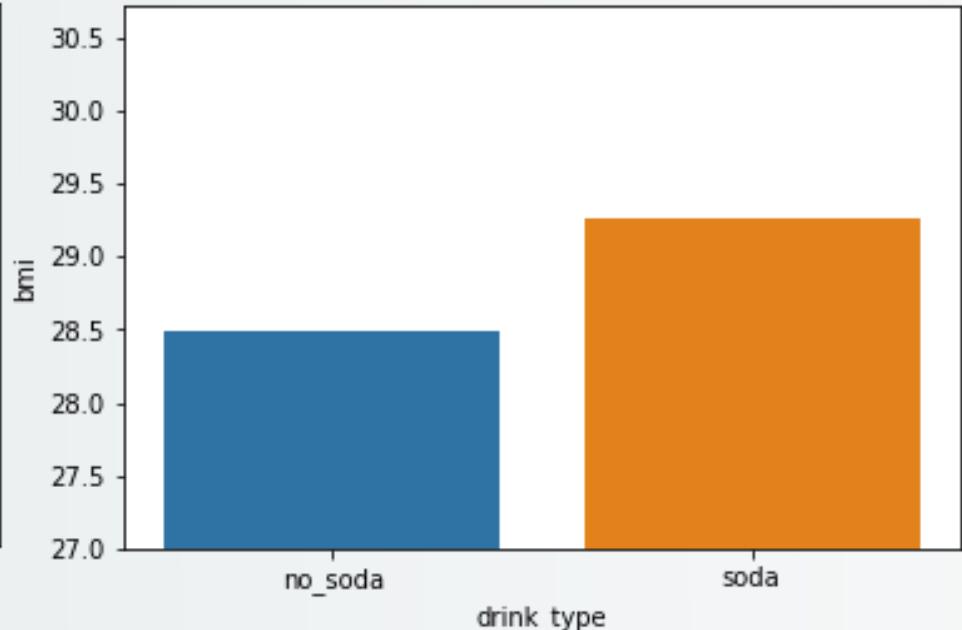
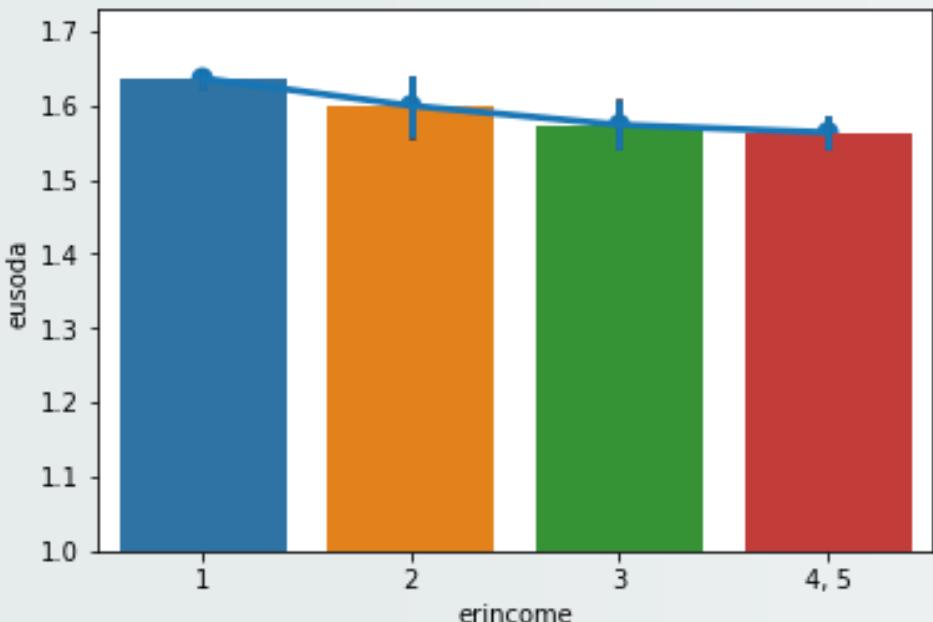


- SNAP을 지원받는 사람들은 식사보다 간식을 더 많

이 먹는다

Drink type-BMI

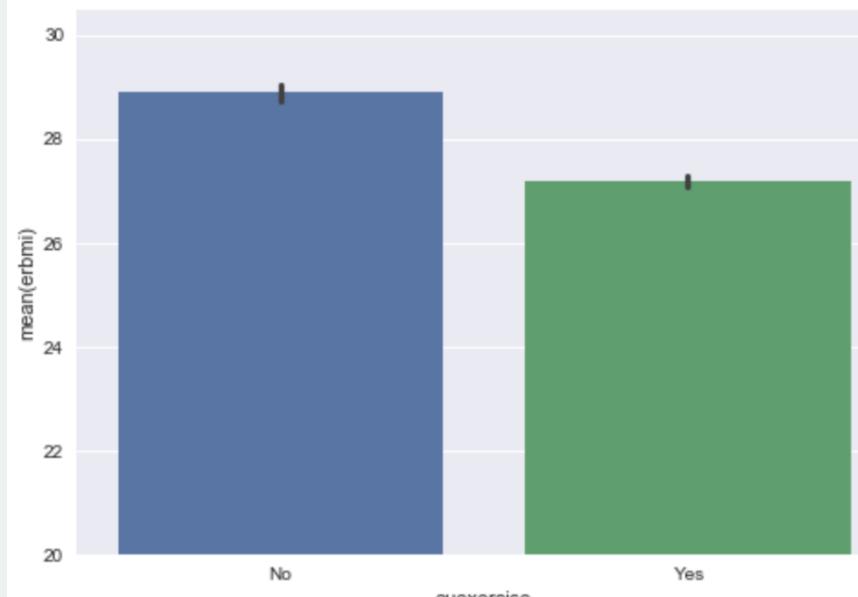
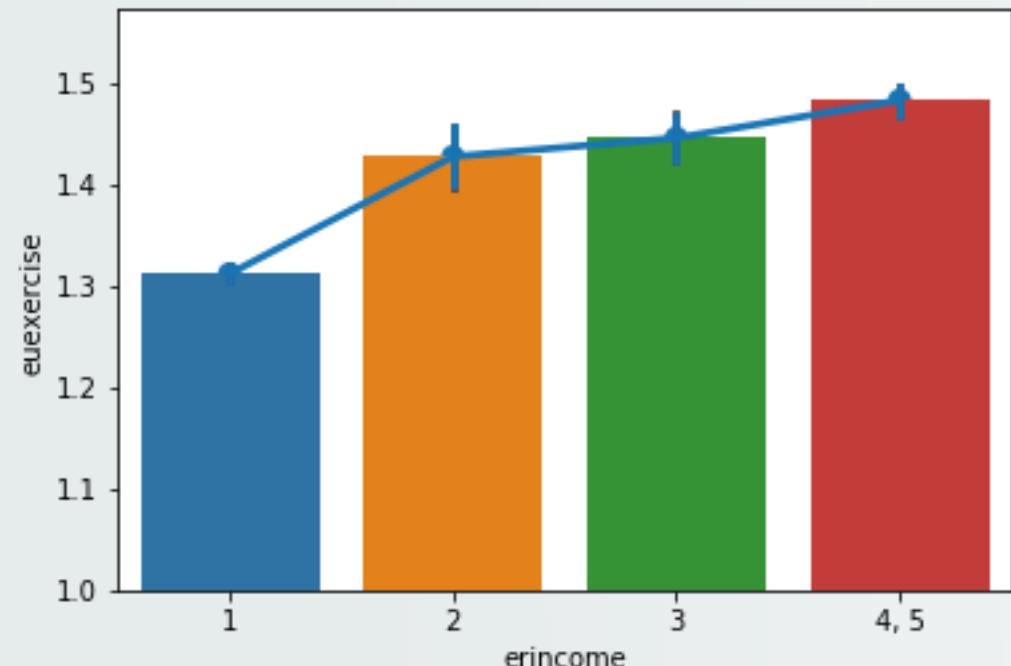
*eusoda는 높을 수록 탄산 음료를 덜 먹는 것이다.



- Income이 낮은 그룹의 탄산음료 소비량이 높다

Exercise-BMI

*euexercise는 높을 수록 운동을 덜 하는 것이다.



- Income이 낮은 그룹의 운동량이 적다

Conclusion – Policy suggestion

- 비만과 관련된 feature 5개: ‘eugenht’,
‘euexercise’, ‘erincome’, ‘eusoda’, ‘eusnap’
- ✓ Policy #1 - SNAP과 비슷한 정책을 실시하되, 지
원금으로 구매할 수 있는 Secondary eating량에
제한을 두자



Conclusion – Policy suggestion

- ✓ Policy #2 - SNAP과 비슷한 정책을 실시하되, 지원금으로 구매할 수 있는 Soda구매량에 제한을 두자



Conclusion – Policy suggestion

- ✓ Policy #3 - SNAP과 비슷한 정책을 실시하되, 지원금으로 헬스장을 다닐 수 있도록 하는 등의 운동 활동을 지원해주자



Limitation & Future work

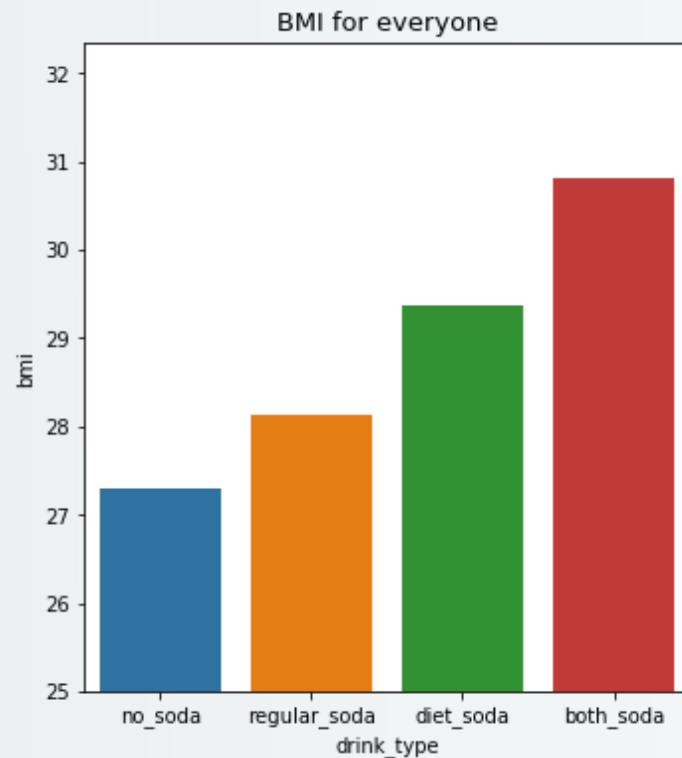
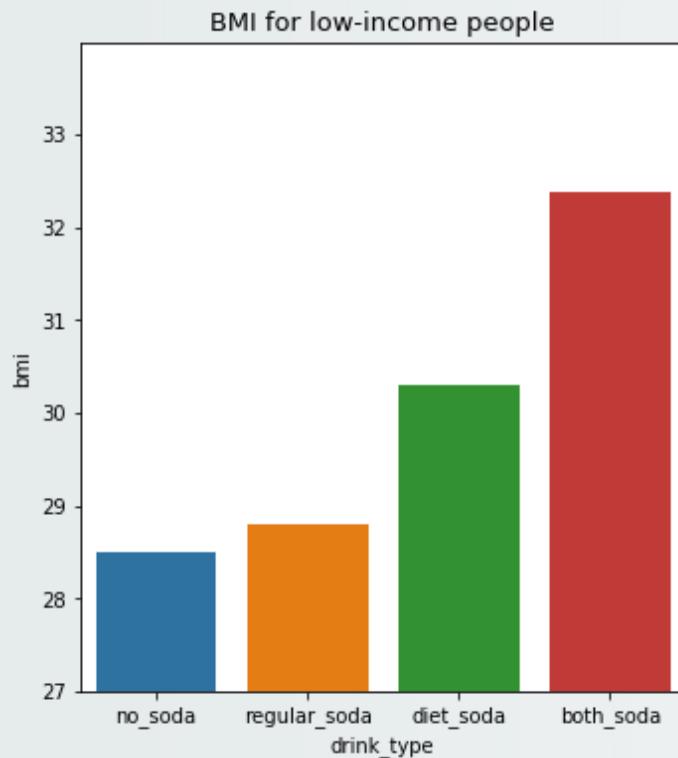
2014

2015

1	tucaseid	tulineno	eeincome	erbmi	erl	1	TUCASEID	TULINENO	EEINCOME1	ERBMI
2	20140101140007.0	1	-2	33.2	2	2	20150101150018.0	1	1	31.4
3	20140101140011.0	1	1	22.7	3	3	20150101150053.0	1	2	25.7
4	20140101140028.0	1	2	49.4	4	4	20150101150071.0	1	1	29.6
5	20140101140063.0	1	-2	-1	5	5	20150101150146.0	1	3	23.4
6	20140101140168.0	1	2	31	6	6	20150101150147.0	1	1	35.9
7	20140101140559.0	1	1	30.7	7	7	20150101150184.0	1	1	32.1
8	20140101140610.0	1	1	33.3	8	8	20150101150500.0	1	1	30.1
9	20140101140614.0	1	1	27.5	9	9	20150101150523.0	1	1	33.9
10	20140101140639.0	1	1	25.8	10	10	20150101150535.0	1	2	26.3
11	20140101140665.0	1	1	28.3	11	11	20150101150539.0	1	1	22.6
12	20140101140685.0	1	1	40.5	12	12	20150101150572.0	1	3	28.3
13	20140101140702.0	1	2	28	13	13	20150101150578.0	1	2	22.9
14	20140101140725.0	1	1	27.9	14	14	20150101150601.0	1	1	-1
15	20140101140759.0	1	2	30.4	15	15	20150101150607.0	1	2	31.4
16	20140101140792.0	1	2	26.8	16	16	20150101150615.0	1	2	26.6
17	20140101140804.0	1	1	32.9	17	17	20150101150628.0	1	1	21.8
18	20140101140836.0	1	3	25.8	18	18	20150101150641.0	1	2	29.5
19	20140101140852.0	1	1	26.5	19	19	20150101150677.0	1	1	25.1



Limitation & Future work



- 비만이어서 다이어트 소다를 많이 먹는다
- 다이어트 소다를 많이 먹어서 비만이다

Thank you!
Any questions?

