



# 식습관 스몰 데이터 분석을 통한 장트러블 극복기

## 팀 푸푸

고은솔 <https://github.com/koeunsol>

김민주 <https://github.com/kim4pb>

배예슬 <https://github.com/yeseul0806>

송혜정 <https://github.com/hyejeong11>

이선영 <https://github.com/godkrt>

정보경 <https://github.com/alicebk221>

# | 목차

I. 프로젝트 소개

II. 데이터 소개

III. 분석 방법 / 결과

IV. 가설 검증 / 예측

## | 프로젝트 시작 과정

장트러블로 고생하는 팀원의 제안으로 시작.

“ 배변 만족도, 배변 습관에 영향을 주는 음식이 개인마다 다르지 않을까? ”

배변에 영향을 주는 음식을 알아보고 장트러블을 극복하기 위한 쾌변 프로젝트를 시작.

# | 가설 배변에 영향을 줄 것이라고 예상하는 요인들

Spicy



Cold



Coffee



Alcohol



Milk



## | 본격적인 설명에 앞서

- 이런 한계점을 느꼈습니다

: 음식 외에도 배변에 영향을 미치는 요인이 많음.

- 이런 부분에 집중했습니다

: 살면서 경험적으로 느껴왔던 것처럼,

특정 음식이 배변 패턴과 관련이 있는지를 데이터를 통해 검증하는 것.

- 프로젝트 진행 방향은 이렇습니다

: 팀원 6명 각자의 가설을 검증하는 방향으로 분석.

- 이런 점에 집중해주세요.

: 데이터 활용 방법, 사용한 분석/시각화 도구.

# | 데이터 수집

- 가설을 기준으로 음식을 분류하여 열로 구분
- 직접 구글 스프레드 시트에 입력하여 데이터를 수집
- 수집 기간: 8/2 ~ 10/2 (62일)
- 데이터 개수: 총 1316개

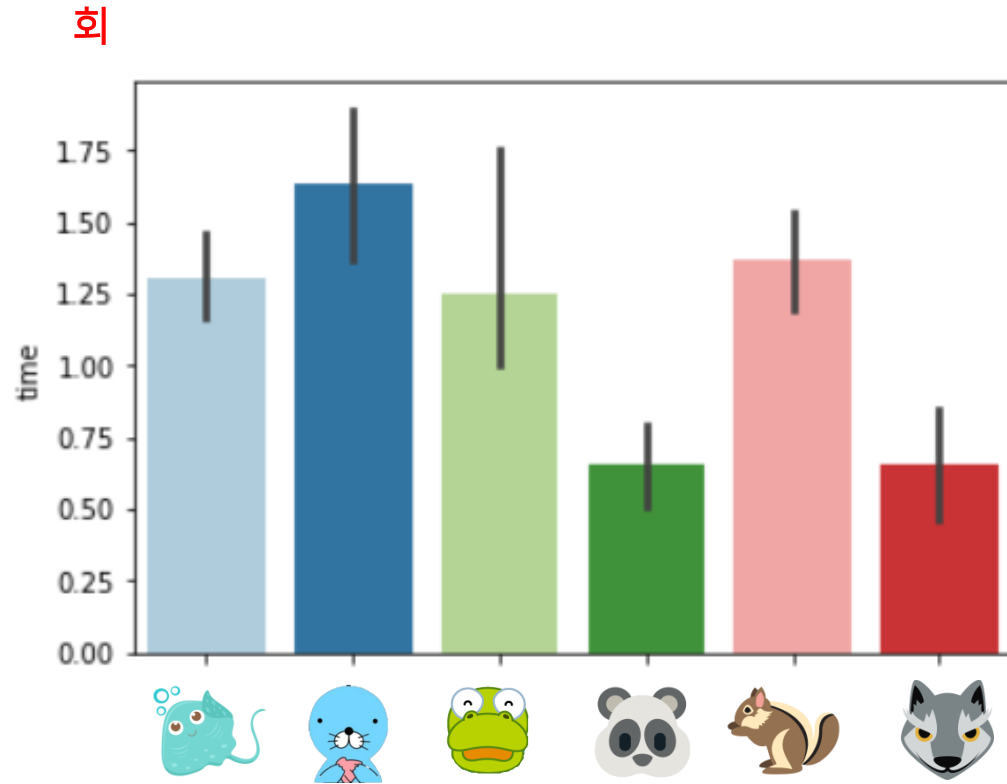
닉네임	날짜	사시	식사 메뉴	맵기	매운거	차가운거	커피	탄산	술	유제품, 두유	물	채소, 과일	유산균	특이사항
가오리	2019-08-02	22:00	오징어, 수박, 새우깡, 맥주 두모금	0	-	-	-	-	맥주 20ml	-	100	수박	0	
가오리	2019-08-02	19:30	만두국, 김치참치치즈밥	2	-	-	-	-	-	슬라이스치즈 1/2	300	-	0	
가오리	2019-08-03	12:00	닭강정	0	-	-	-	-	-	-	500	-	0	
가오리	2019-08-03	15:00	쌀햇도그	0	-	-	-	-	-	-	200	-	0	
가오리	2019-08-03	20:30	생선구이	0	-	-	-	-	-	-	300	오이 콩나물	0	

식사 관련 데이터 | 940개

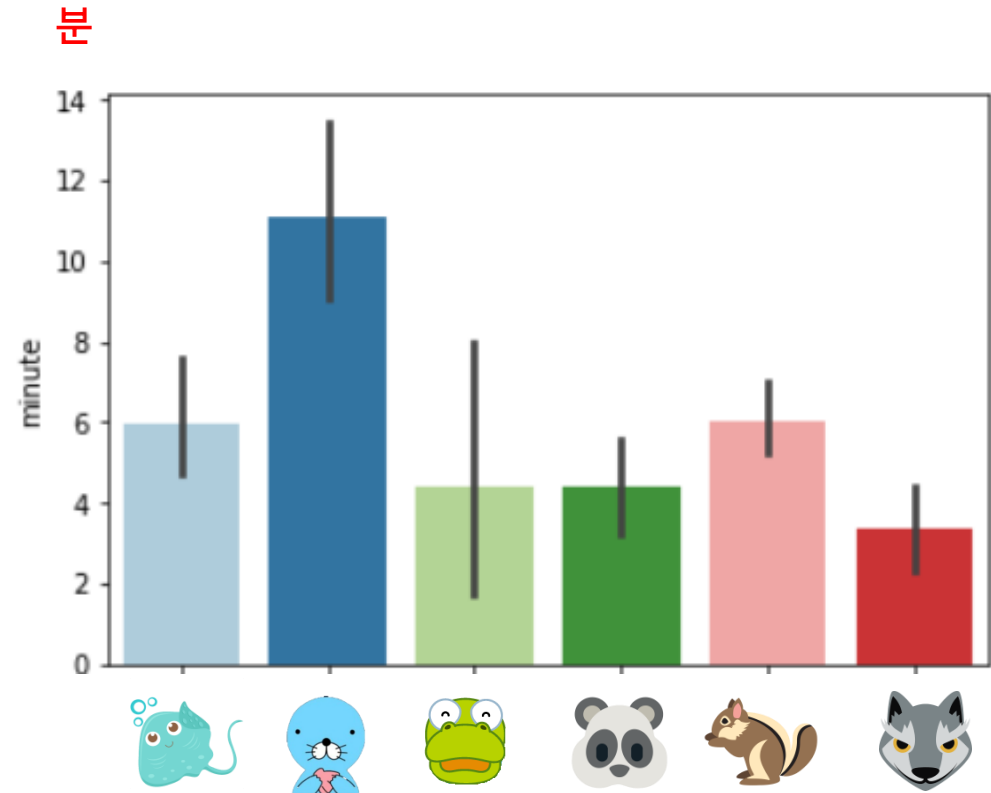
닉네임	날짜	화장실 간 시간	걸린 시간	쾌변 / 애매 / 시도	쾌변 여부	쾌변 패턴	생리
가오리	2019-08-03	9:00	5	애매	아주 약간 배 아팠는데 아주 약간 찜찜		
가오리	2019-08-03	9:50	5	애매	아주 약간 배 아팠고 약간 찜찜		
가오리	2019-08-03	19:00	10	애매	배아픔...약한		
가오리	2019-08-04	10:30	5	애매	낮뻐.. ㅇ 애매		

쾌변 관련 데이터 | 376개

# | Barplot Seaborn 라이브러리의 barplot 이용



일 평균 화장실 간 횟수

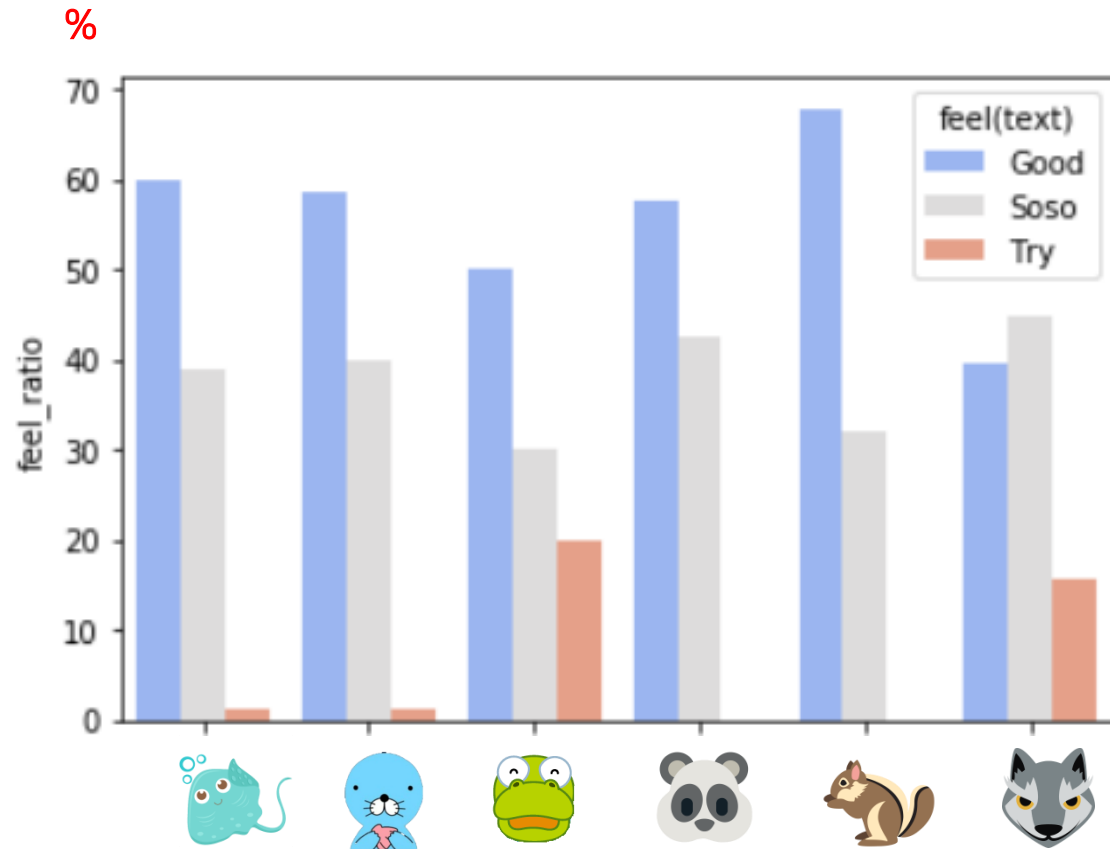


일 평균 화장실에서 보낸 시간

# | Barplot Seaborn 라이브러리의 barplot 이용

배변 만족도

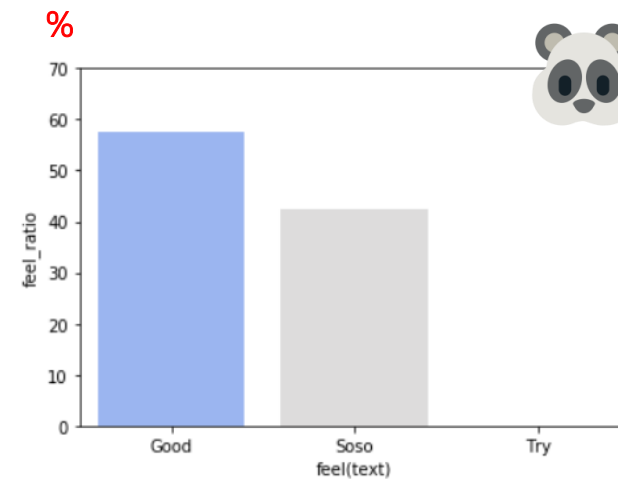
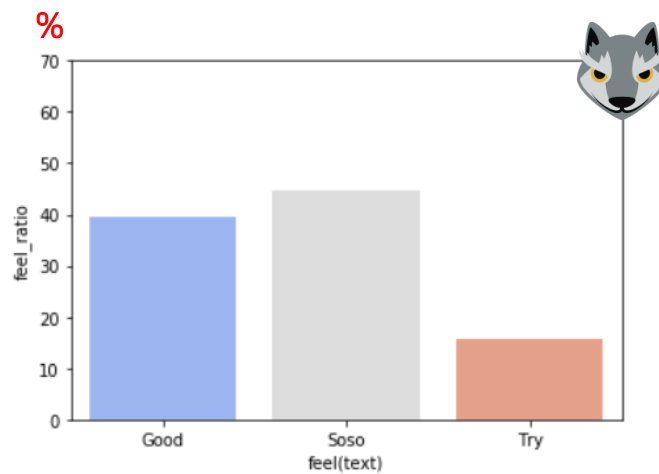
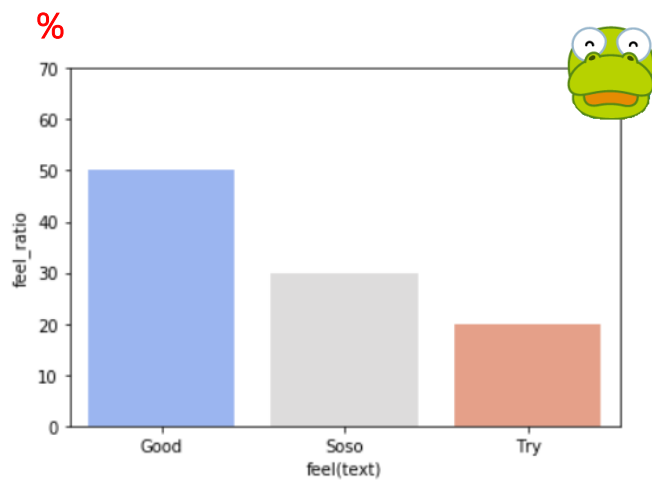
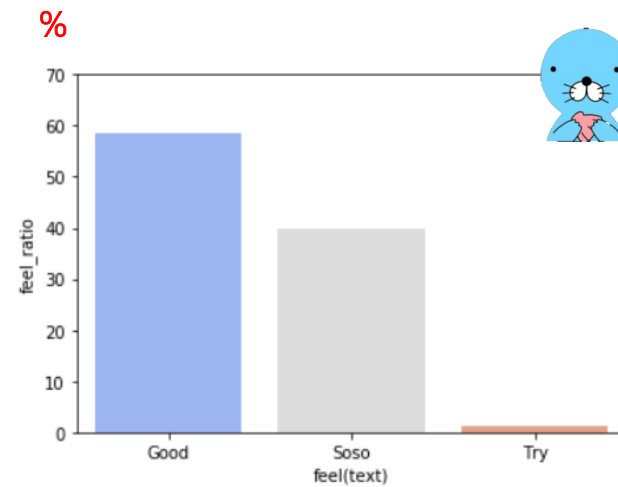
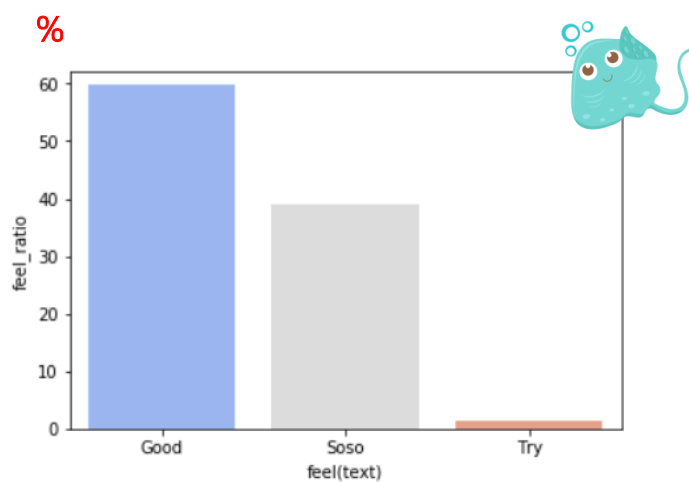
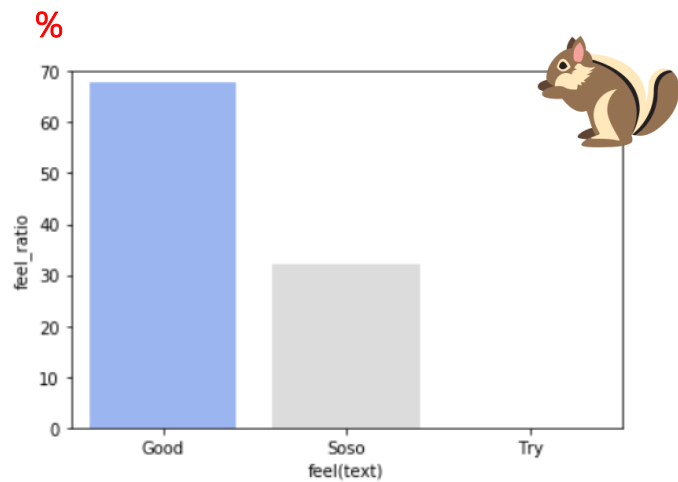
- 1: 시도
- 2: 애매
- 3: 쾌변



개인별 배변 만족 비율



# | Barplot Seaborn 라이브러리의 barplot 이용

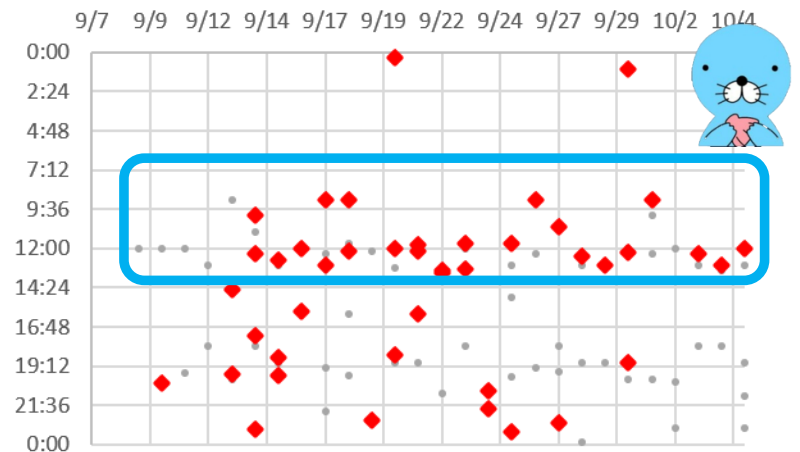
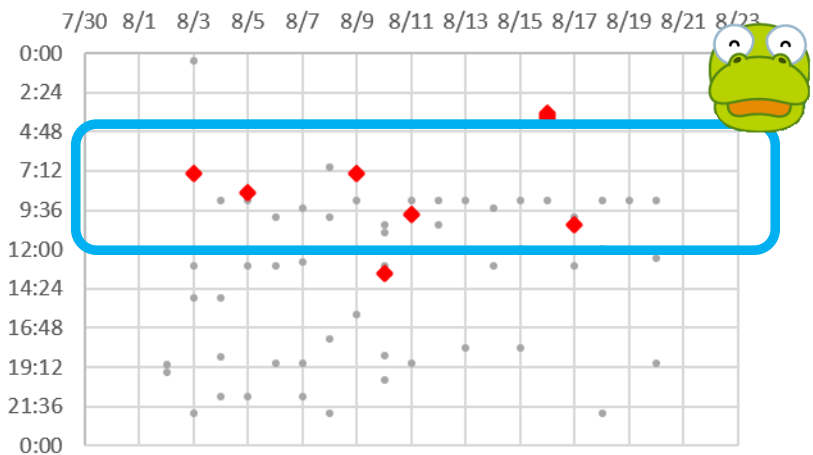
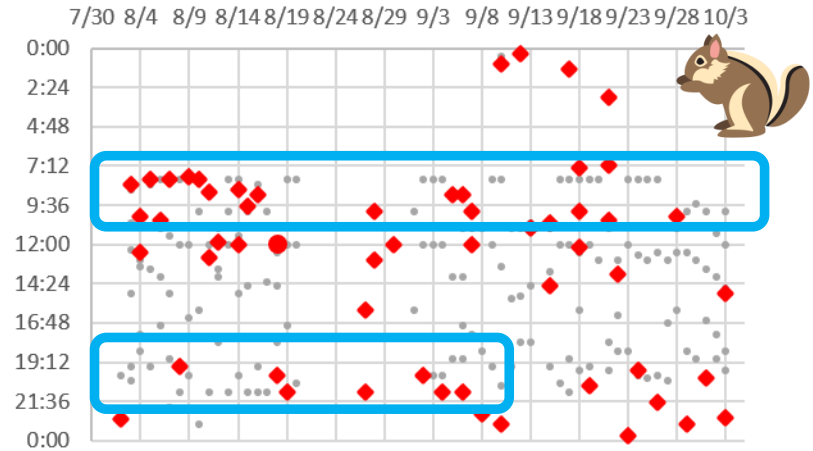
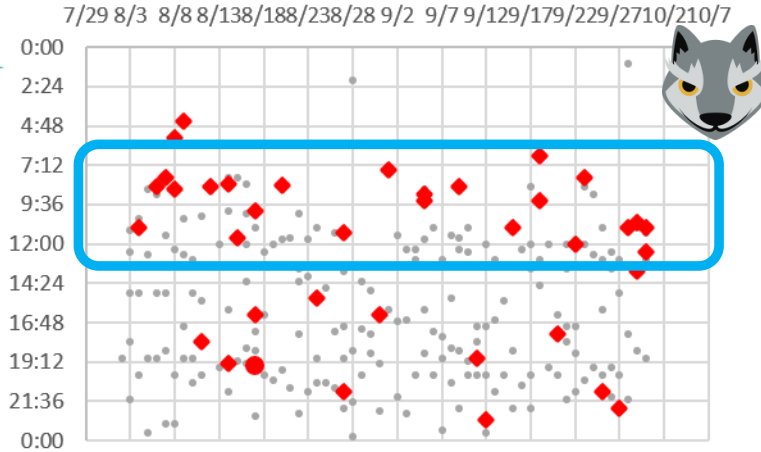
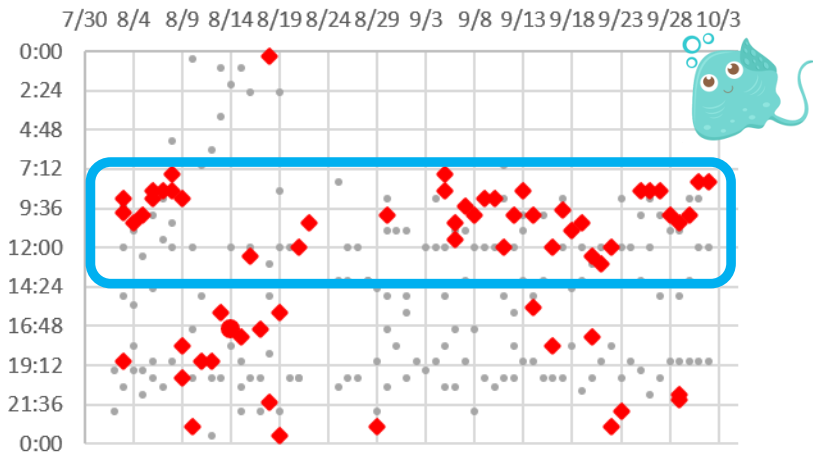


개인별 캐변 만족도 비율

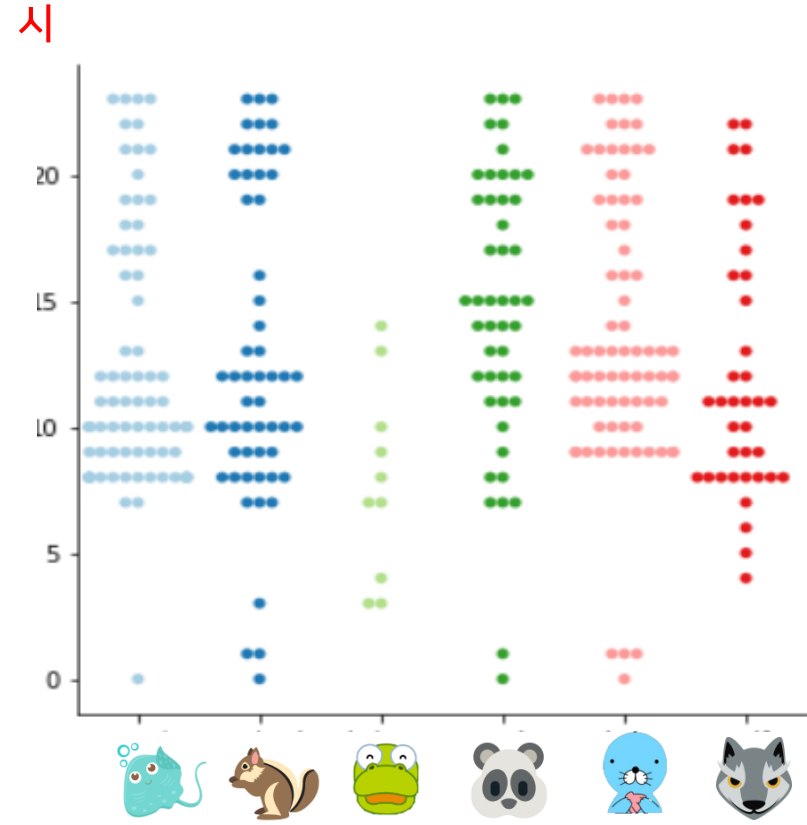
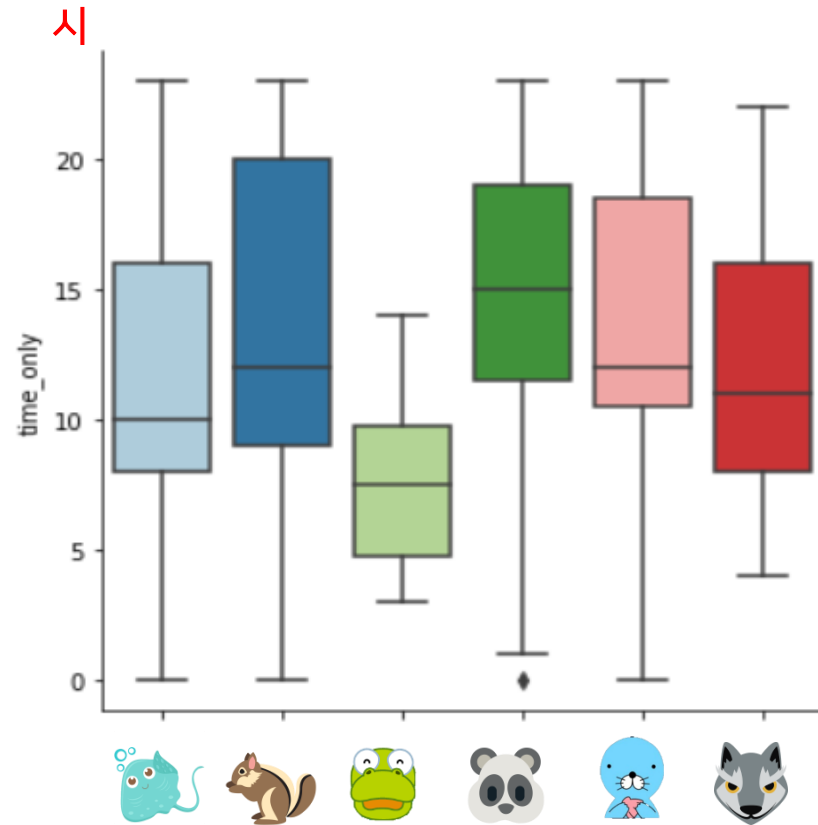
# | Scatterplot MS Excel의 Scatterplot 이용

날짜

시간

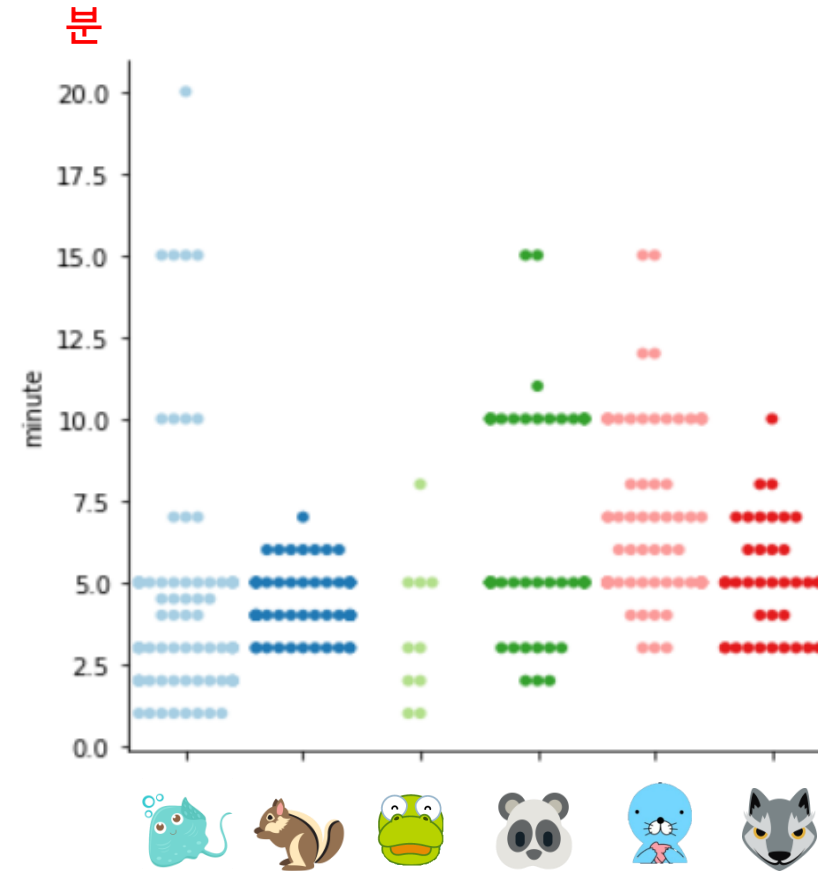
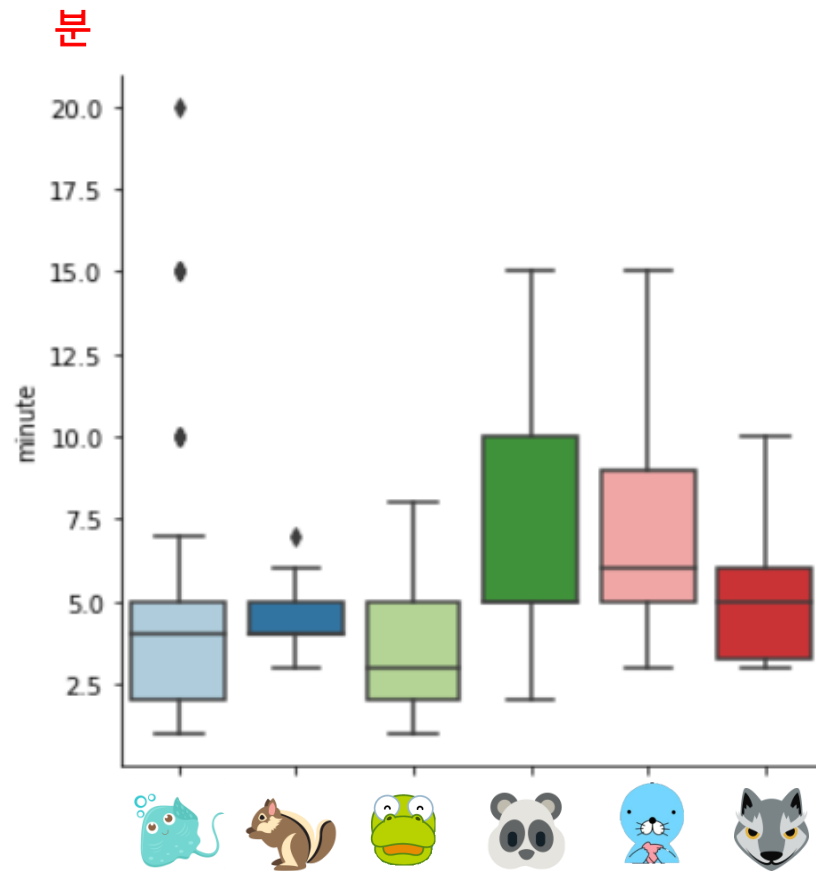


# | Boxplot, Catplot    Seaborn 라이브러리의 barplot, catplot 이용



화장실 간 시간대

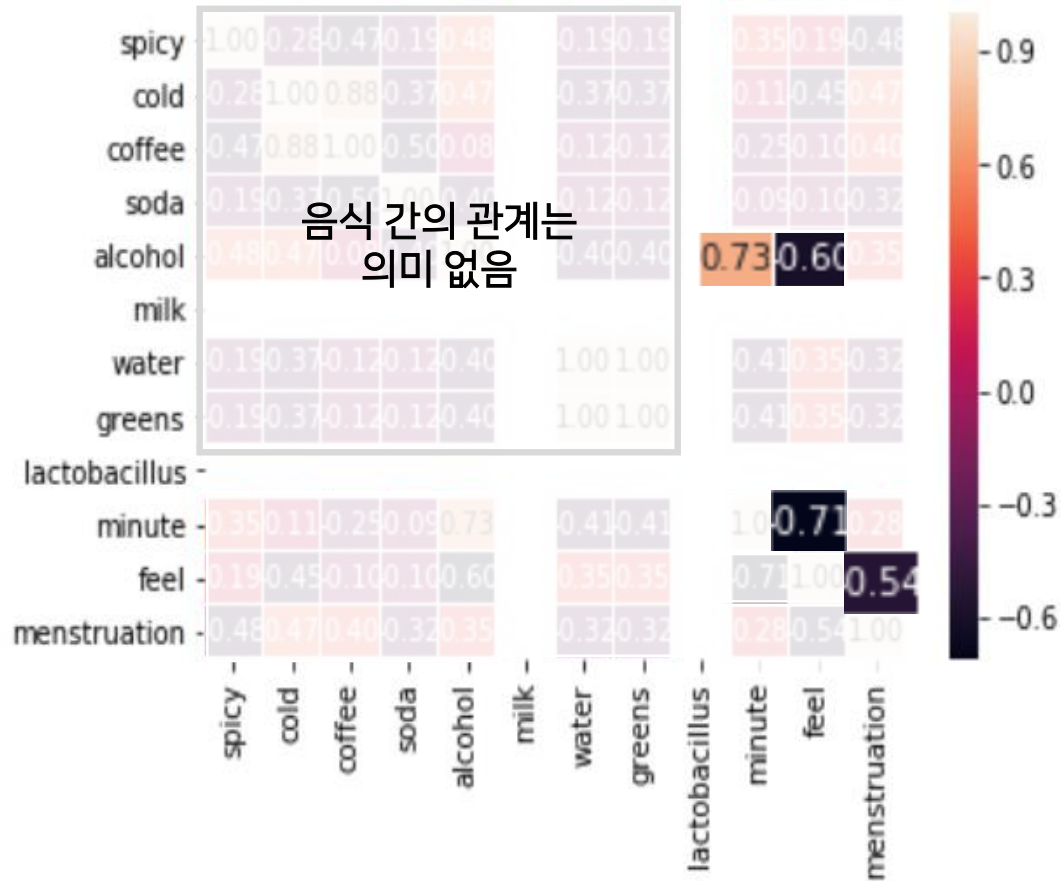
# | Boxplot, Catplot    Seaborn 라이브러리의 barplot, catplot 이용



화장실에서 보낸 시간 (분)

# | Heatmap 🐸

Seaborn 라이브러리의 heatmap 이용

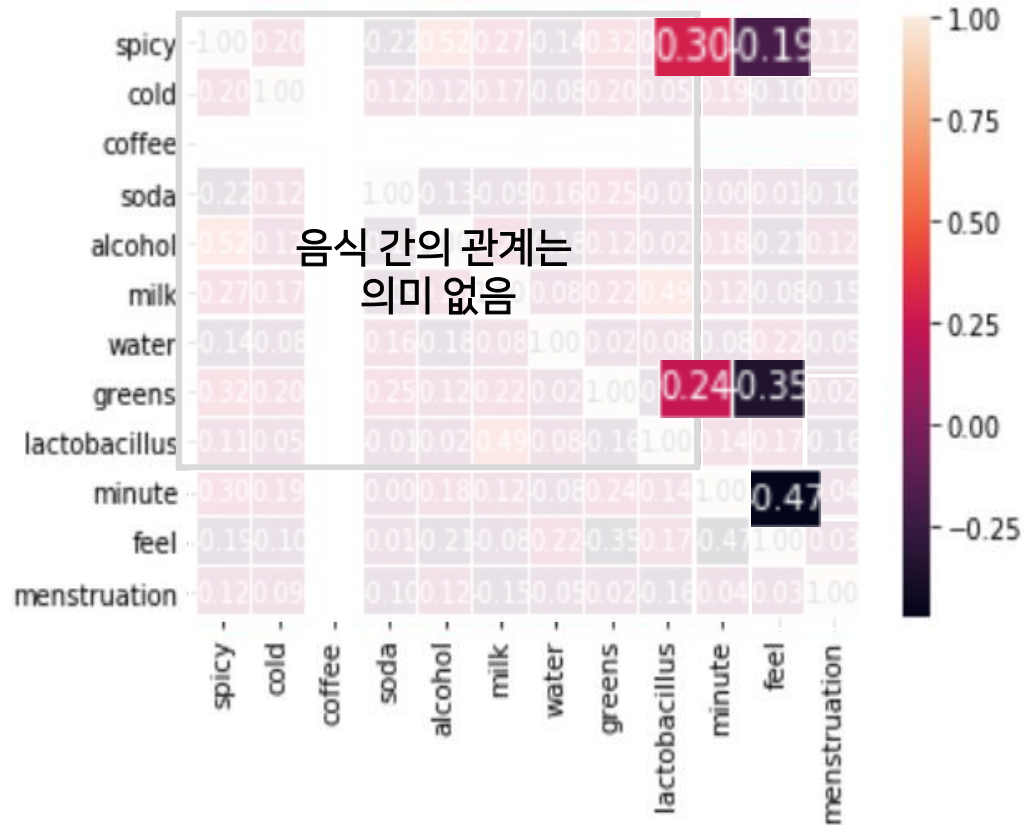


- 술을 마시면 배변 만족감 감소(0.60),  
걸리는 시간 증가(0.73)
- 걸리는 시간이 증가할수록 만족감 감소(-0.71)
- 생리를 하면 만족감 감소(-0.54)

# | Heatmap



Seaborn 라이브러리의 heatmap



- 매운 음식을 먹으면 걸리는 시간 증가(0.30), 만족감 감소(-0.19)
- 야채를 많이 먹으면 만족감 증가(-0.35), 걸리는 시간 증가(0.24)
- 걸리는 시간이 길어질수록 만족감 감소(-0.47)

# | Linear Regression



Statsmodel 라이브러리의 ols 이용

독립 변수: 걸린 시간



종속 변수: 만족감

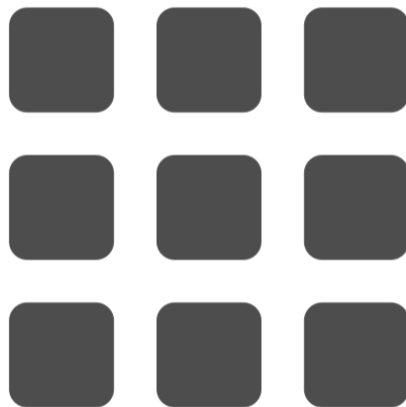
## OLS Regression Results

Dep. Variable:	feel	R-squared:	0.224			
Model:	OLS	<b>Adj. R-squared:</b> 0.205				
Method:	Least Squares					
Date:	Fri, 18 Oct 2019	Prob (F-statistic):	3.05e-05			
Time:	11:22:39	Log-Likelihood:	-52.241			
No. Observations:	85	AIC:	110.5			
Df Residuals:	82	BIC:	117.8			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8841	0.083	34.730	0.000	2.719	3.049
minute	-0.0683	0.014	-4.857	0.000	<b>-0.096</b>	<b>-0.040</b>

- Adj.R-squared: 0.205  
→ 걸린 시간이 만족감의 약 20%를 설명.
- 음의 상관관계  
→ 두 변수가 반비례 관계.

# | Backward Elimination

9개의 변수



유의미한 결과를 찾기에  
적합하지 않음



변수 제거



후방 제거 방법 선택



# | Backward Elimination



- 파이썬을 이용하여 후방 제거를 위한 함수 작성
- 종속변수: 시간

```
pm1 = ['spicy', 'cold', 'soda', 'alcohol', 'milk', 'water', 'greens', 'lactobacillus']  
backward_minute(pm1, fs_1d[fs_1d['id'] == 'gaori'])
```

('spicy+soda+alcohol+milk+water+greens+lactobacillus', 0.0630560144737009)

```
pm2 = ['spicy', 'soda', 'alcohol', 'milk', 'water', 'greens', 'lactobacillus']  
backward_minute(pm2, fs_1d[fs_1d['id'] == 'gaori'])
```

('spicy+soda+milk+water+greens+lactobacillus', 0.07286882643059989)

```
pm3 = ['spicy', 'soda', 'milk', 'water', 'greens', 'lactobacillus']  
backward_minute(pm3, fs_1d[fs_1d['id'] == 'gaori'])
```

('spicy+soda+water+greens+lactobacillus', 0.0813544221919793)

```
pm4 = ['spicy', 'soda', 'water', 'greens', 'lactobacillus']  
backward_minute(pm4, fs_1d[fs_1d['id'] == 'gaori'])
```

('spicy+water+greens+lactobacillus', 0.09271542753997108)

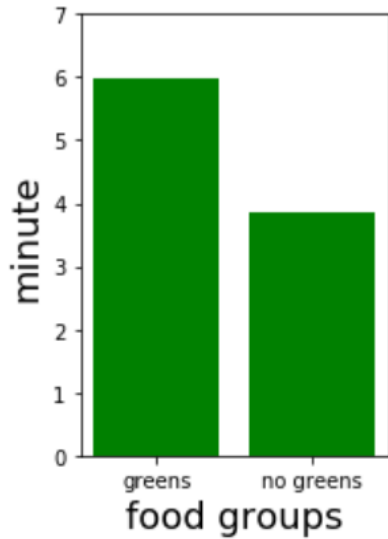
독립변수가 매운 것, 물, 야채/과일, 유산균 일 때 가장 큰 영향을 줌

spicy   water   greens   lacobacillus

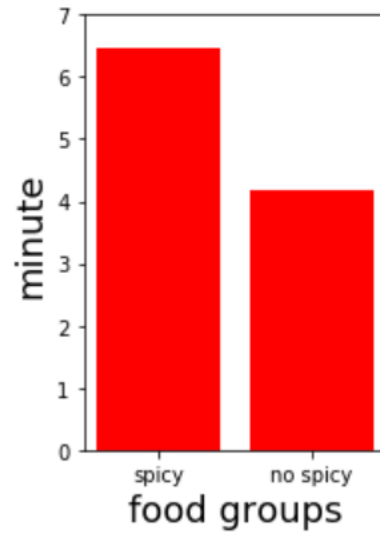
# | 음식 섭취 여부에 따른 결과 비교

걸린 시간

채소

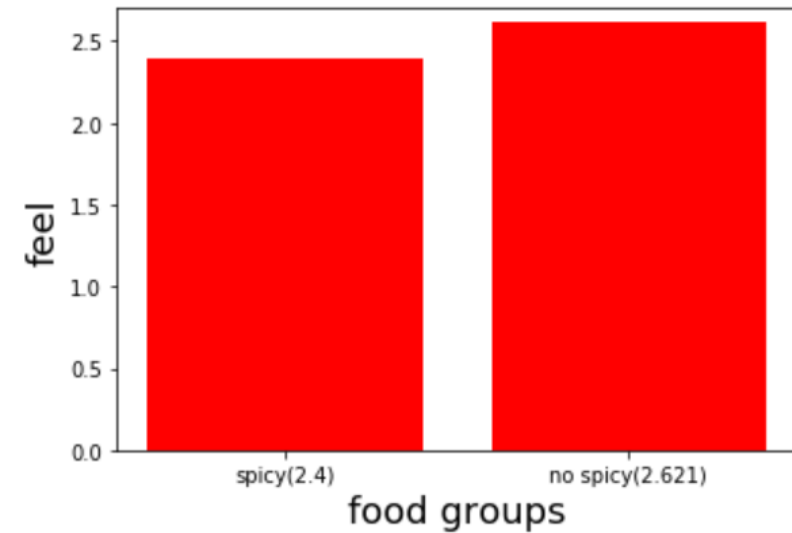


매운 음식



만족도

매운 음식



야채/과일을 먹었을 때의 걸린 시간  
VS  
안 먹었을 때의 걸린 시간

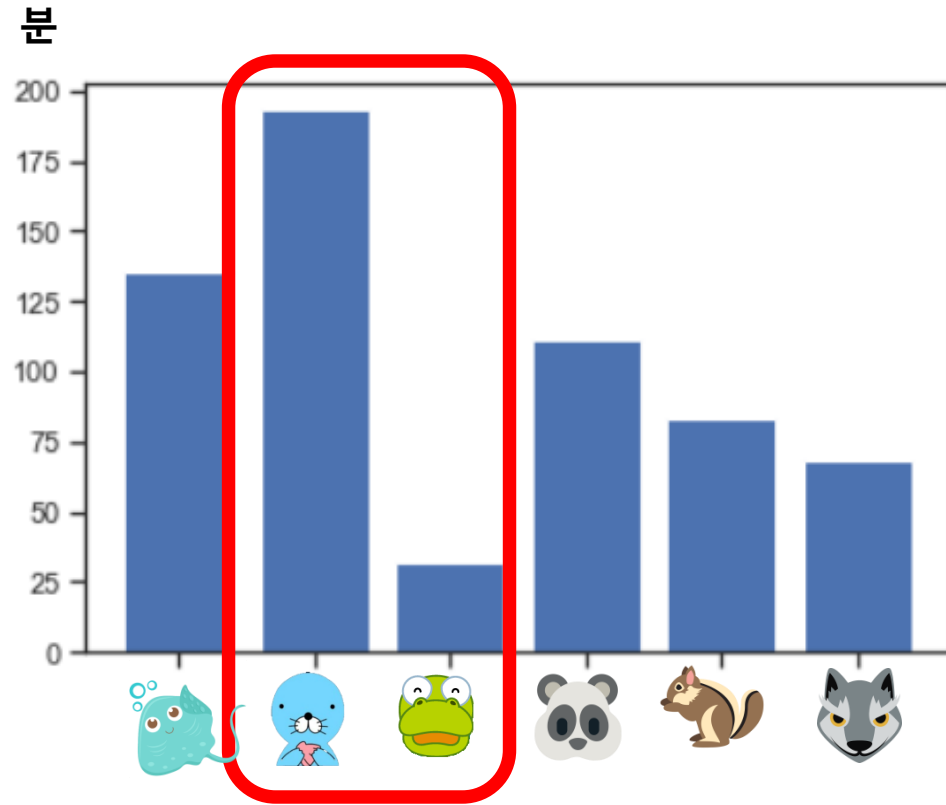
매운 음식을 먹었을 때의 걸린 시간  
VS  
안 먹었을 때의 걸린 시간

매운 음식을 먹었을 때의 배변 만족도  
VS  
안 먹었을 때의 배변 만족도

+ 미래 예측

앞으로 50년 동안 얼마의 시간을  
화장실에서 보내게 될까?

+ 앞으로 50년 동안 화장실에서 보내는 시간



화장실에서 보낸 시간의 총합



VS



화장실에서 보낸 시간이 가장 긴 두 사람을 비교

## | 앞으로 50년 동안 화장실에서 보내는 시간



$$\begin{aligned} &10.7\text{분 /일} \\ &10.7 * 365 * 50 \\ &= \text{약 } 135\text{일} \end{aligned}$$



$$\begin{aligned} &1.25\text{분 /일} \\ &1.25 * 365 * 50 \\ &= \text{약 } 15\text{일} \end{aligned}$$

**120일 차이 발생**



가 식습관을 바꾼다면, 이 차이를 조금은 줄일 수 있지 않을까?

## | 남은 과제

두 달 간의 데이터 분석 결과를 토대로  
실제로 우리의 장 트러블을 극복해보자!

# Thank you