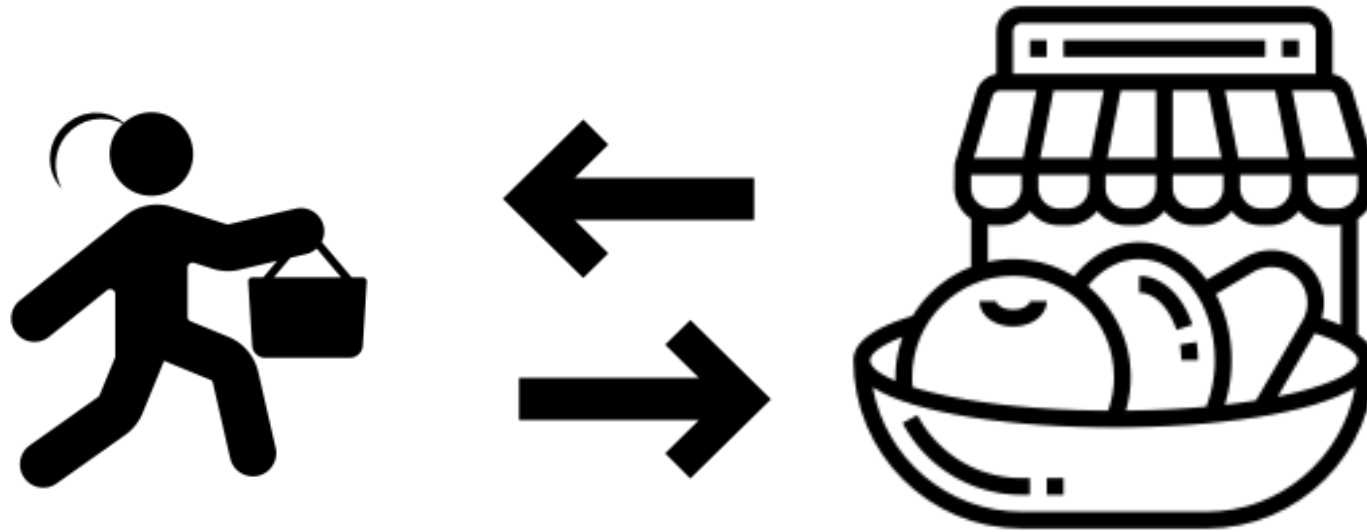


# 추천 시스템

## <Recommender System>

## 이 수업의 주제 : 추천

---



고객

상점

핵심 이슈 : 고객에게 **어떠한 제품을 권할 것인가?**

# 이 수업의 구성

---

## 1. 고객이 첫 방문을 했을 때

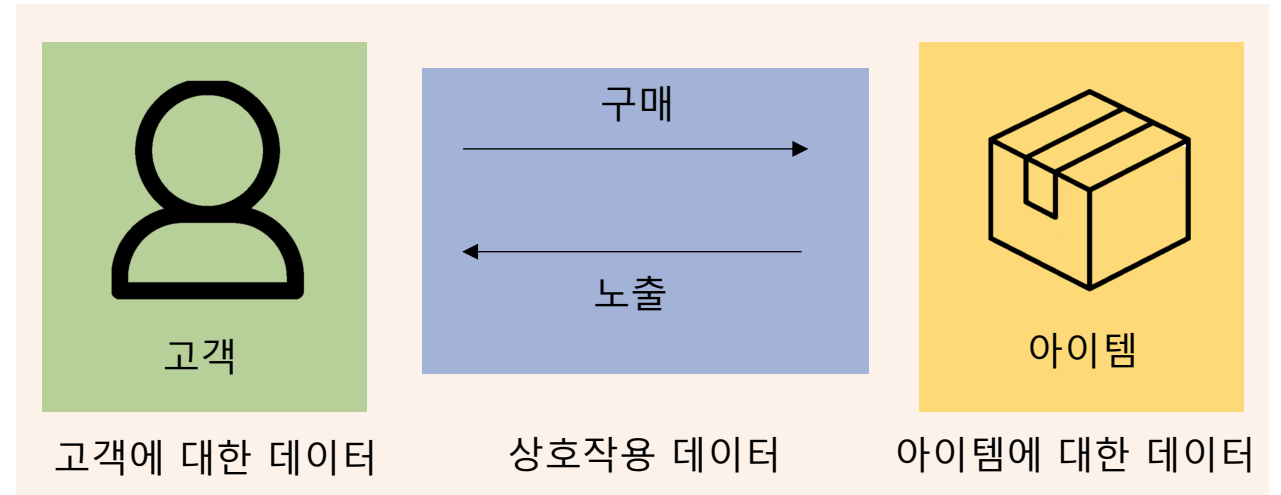
: 고객에 대한 정보가 없을 때

## 2. 고객이 재차 방문을 했을 때

: 고객의 이전 구매 정보가 있을 때

## 3. 고객이 단골이 되었을 때

: 고객 신원 정보가 있을 때



# 이 수업의 주요 토픽

---

## 1. 고객이 첫 방문을 했을 때

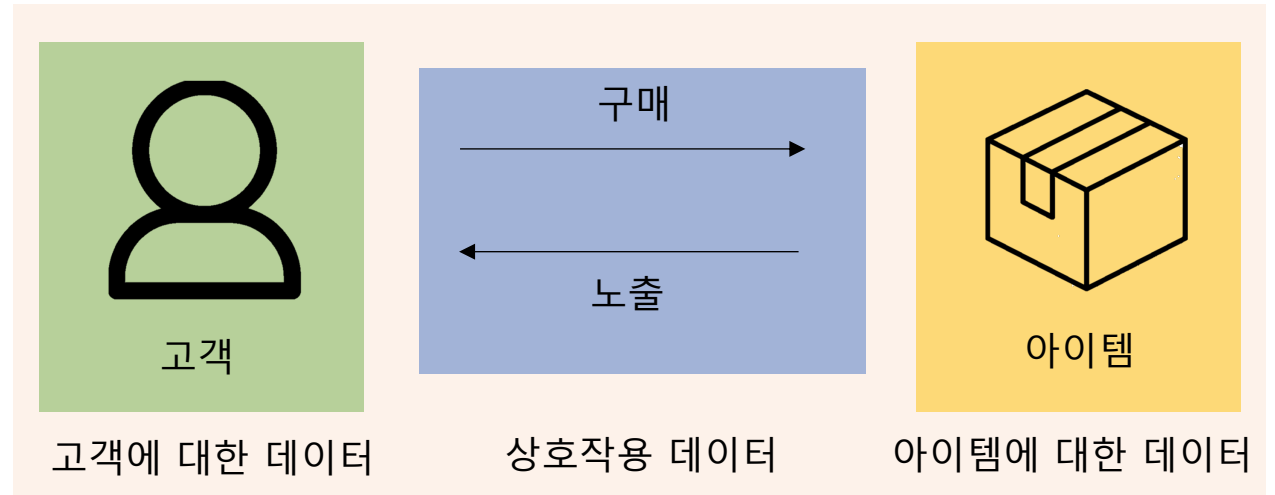
- 비개인화 추천
- 연관 분석

## 2. 고객이 재차 방문을 했을 때

- Collaborative Filtering

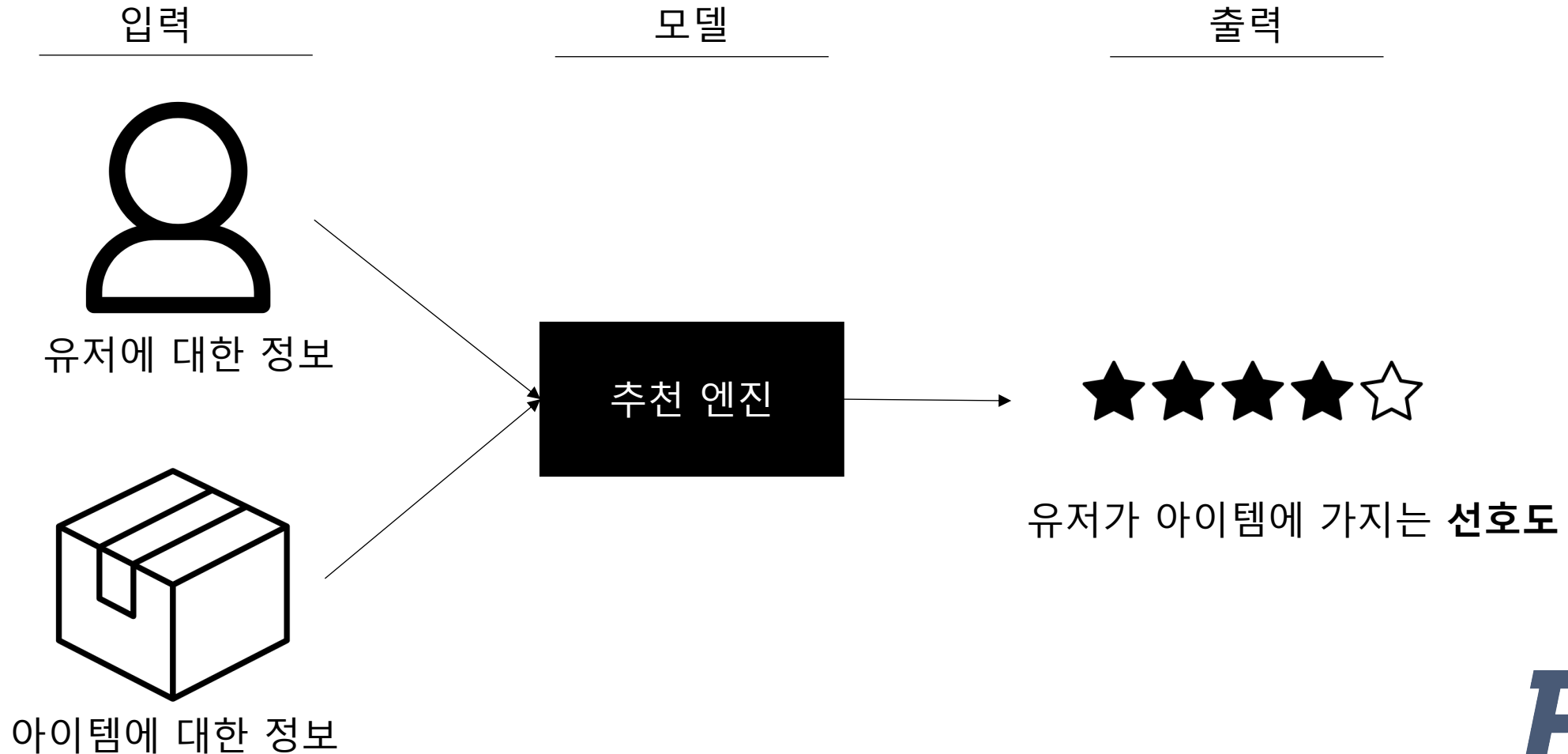
## 3. 고객이 단골이 되었을 때

- Factorization Machine



# 추천 엔진이란

추천 엔진의 과제 : 사용자가 얼마나 해당 아이템을 선호할지를 예측할 수 있는 프로그램을 구성하는 것\*



## 수업 목표

---

현업에서 적용 가능한 수준으로

1. 추천시스템 알고리즘의 이해도 향상
2. 추천시스템 알고리즘의 코드 작성 능력 배양

# 수업 환경

---

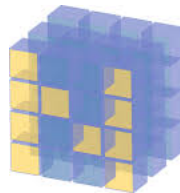
수업 코드 환경 : jupyter notebook



다루는 언어 : Python 3.6



다루는 패키지



NumPy

Pandas



**PAI**  
PUBLIC AI

# 수업 코드 리파지토리

<https://github.com/public-ai/recommender-20201Q>

public-ai / recommender-20201Q

Watch 0

Star 0

Fork 0

<> Code

Issues 0

Pull requests 0

Actions

Projects 0

Wiki

Security

Insights

Settings

러닝스폰즈에서 진행하고 있는 2020년 1분기 추천시스템 수업

[Manage topics](#)

3 commits

1 branch

0 packages

0 releases

1 contributor

Branch: master






New pull request

Create new file

Upload files

Find file

Clone or download

 sangjaekang	temp 파일 삭제	Latest commit d5a3cb9 22 hours ago
 ppt	temp 파일 삭제	22 hours ago
 script/week1	initial commit	22 hours ago
 .gitignore	initial commit	22 hours ago
 README.md	initial commit	22 hours ago

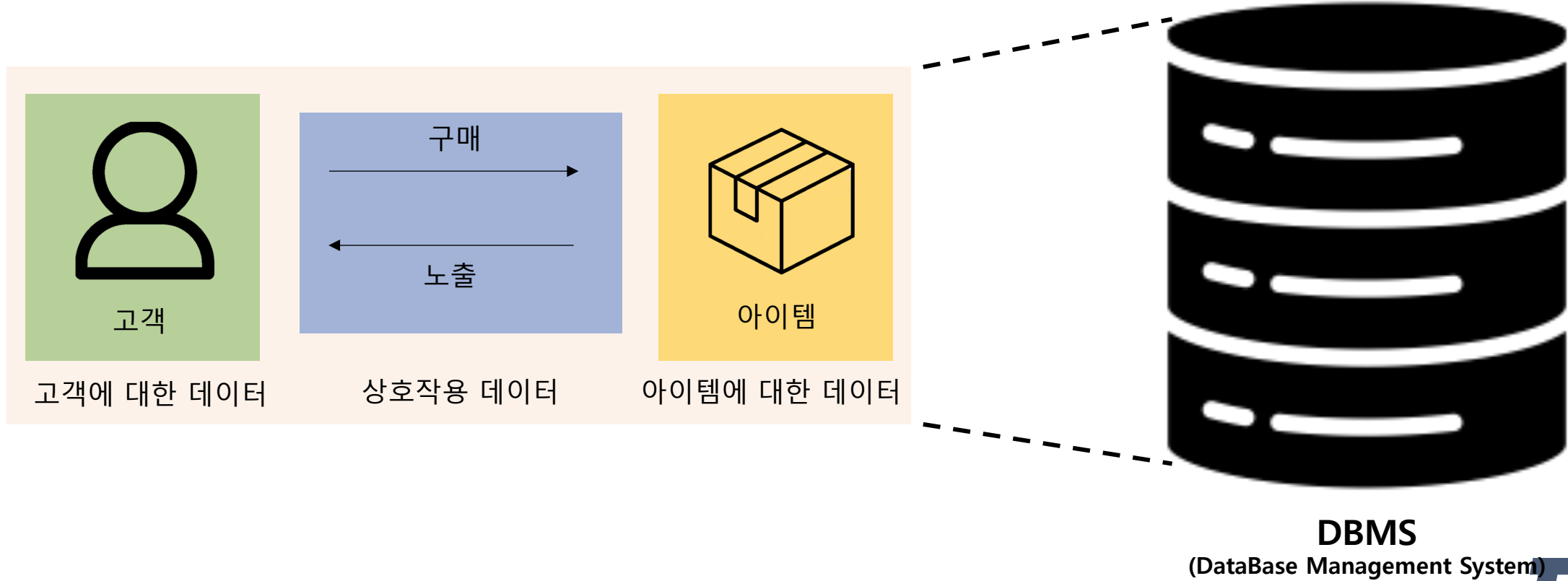
Help people interested in this repository understand your project by adding a README.

Add a README



데이터베이스  
<Database>

# 추천 시스템에서의 데이터



# 오늘 배우게 되는 SQL 구문들

---

원하는 데이터 셋을 가져오기 위한 방법들

**SELECT c1, c2 FROM t;**

Query data in columns c1, c2 from a table

**SELECT \* FROM t;**

Query all rows and columns from a table

**SELECT c1, c2 FROM t**

**WHERE condition;**

Query data and filter rows with a condition

**SELECT DISTINCT c1 FROM t**

**WHERE condition;**

Query distinct rows from a table

**SELECT c1, c2 FROM t**

**ORDER BY c1 ASC [DESC];**

Sort the result set in ascending or descending order

**SELECT c1, c2 FROM t**

**ORDER BY c1**

**LIMIT n OFFSET offset;**

Skip *offset* of rows and return the next n rows

**SELECT c1, aggregate(c2)**

**FROM t**

**GROUP BY c1;**

Group rows using an aggregate function

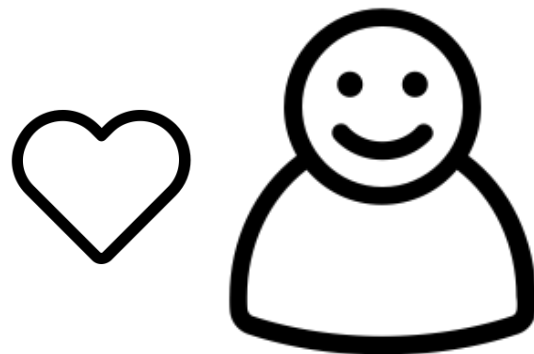
# 연관 분석

## <Association Analysis>

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(1) 지지도 :  $Support(X) = \frac{freq(X)}{N}$

: 전체 고객 중에서 영화 X를 선호하는 사람의 비율은?



지지도를 기준으로 추천할 경우,  
어떤 영화를 선호하던 항상 같은 영화(타이타닉)을  
추천하는 문제가 발생

지지도

4.1%

3.8%

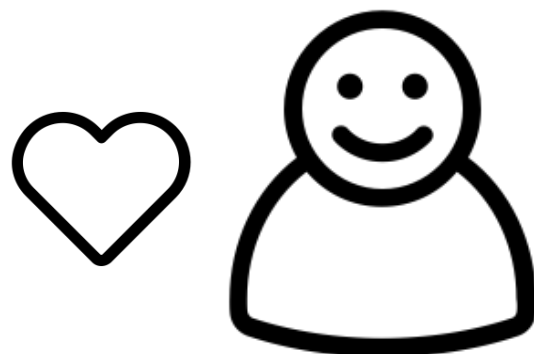
3.8%

10.0%

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(2) 신뢰도 :  $confidence(X \rightarrow Y) = \frac{freq(X,Y)}{freq(X)}$

: 영화 X를 선호한 사람 중에서  
영화 Y를 선호하는 사람의 비율은?(조건부 확률)

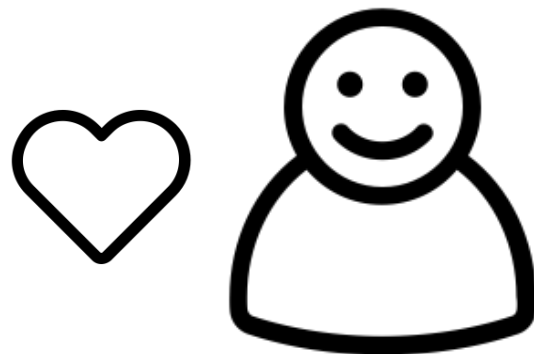


	지지도	신뢰도
스타워즈 3	4.1%	50.2%
스타 트렉	3.8%	17.0%
다들 좋아해요 그 사람들 좋아	3.8%	13.1%
리브 애츨리	10.0%	29.0%

신뢰도를 기준으로 추천할 경우,  
대다수 사람들이 선호하는 영화가  
소수의 사람들이 선호하는 영화보다 우선 추천됨

스타워즈 2를 너무나 재밌게 본 유저에게  
어떤 영화를 추천하는 것이 좋을까?

(3) 리프트 =  $\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)}$   
: 지지도 대비 신뢰도가 얼마나 올라갔는가?



리프트를 기준으로 추천할 경우,  
소수의 사람들이 선호하는 영화이더라도  
강한 연관관계가 있으면 영화가 우선 추천됨



지지도	신뢰도	리프트
4.1%	50.2%	12.2
3.8%	17.0%	4.4
3.8%	13.1%	3.5
10.0%	29.0%	2.9