



# [ 아마존 신서비스 개발 전략 기획서 ]

## <The world's largest bookstore>의 탄생

“  
온라인에서 무언가를 팔아야 한다면,  
책이 최초이자 최고의 상품이 될 것”

-아마존 창업자 제프 베조스



# amazon books

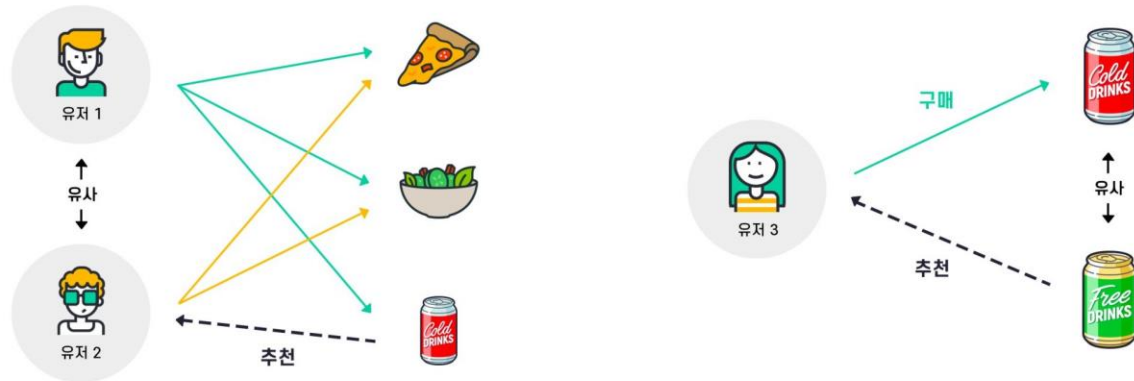
Frequently Wished-For Travel  
Books on Amazon.com



책을 구매하기 위해 방문한 소비자들을 위해  
**아마존은 어떤 서비스 전략을 취하고 있을까?**



검색, 구매 등의 행동 패턴 데이터들을 기반으로  
소비자에게 '함께 구매한 상품' 혹은 '유사한 상품'을 추천



\*상품 간 유사도는 함께 구매되는 경우의 빈도를 분석하여 측정

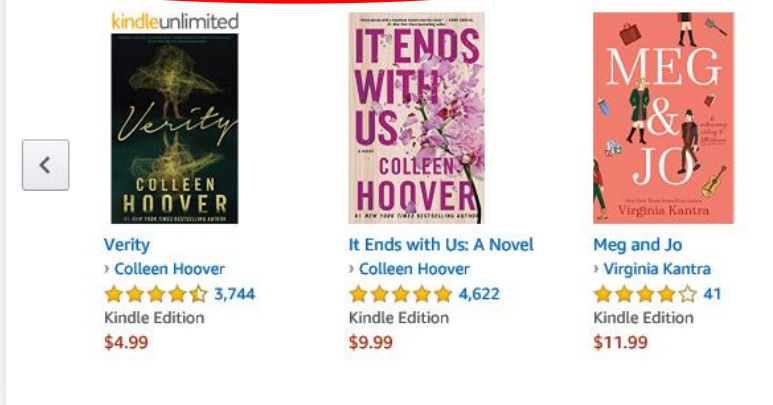
그러나 아마존의 기존 추천시스템에는 **한계**가 존재



### Inspired by your browsing history



### Customers who bought this item also bought



책에서 가격, 무게, 디자인보다 중요한 건, 바로 **내용!**  
근데, 이거 내가 원하는 내용 맞아?

관심사 이외의 책을 클릭했거나  
검색한 것도 집계해 추천한다.

구매패턴에 의해 추천될 뿐,  
원하는 내용이 아닐 수 있다.



Books

Business

Business

Business

# Most-Wished-For Books on Amazon.com

Most

즉, 개인의 취향에 꼭 맞는 책을 추천할 수 있는  
직접적인 접근법의 새로운 서비스 기획이 필요!



Books

Business

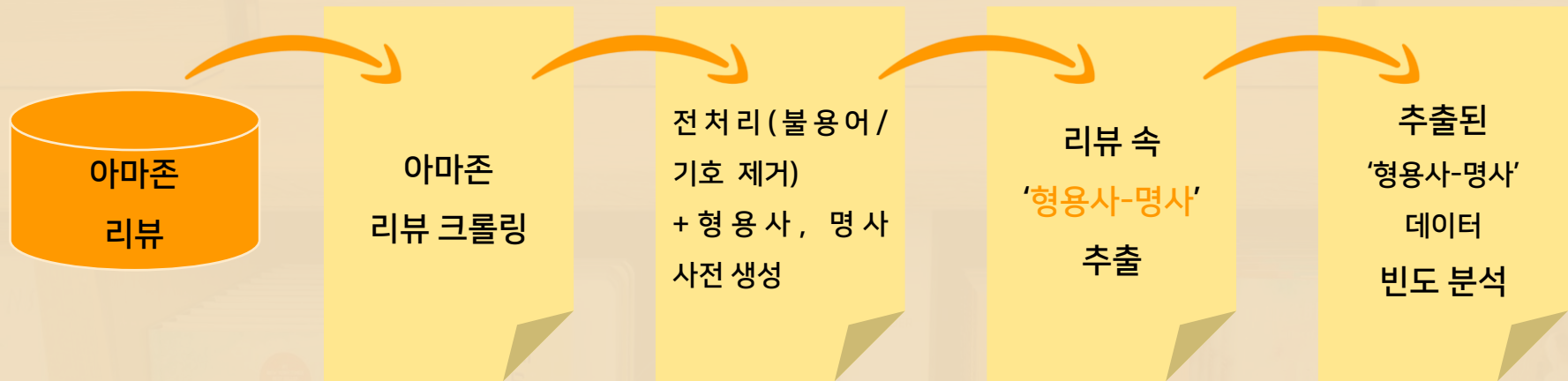
Business

Business

# Most-Wished-For Books

Most

## 독자 리뷰 활용한 'Choose your favorite' 서비스



<리뷰 분석 절차>



# 리뷰 크롤링

## 1) 함수 amazon\_scraper 생성

> 아마존 리뷰 스크랩을 위한 함수 `amazon_scraper` 생성  
제품의 아이디와 주소를 넣으면 그에 해당하는 제품 아이디, 책 제목, 별점, 리뷰를 추출해 데이터프레임으로 만들어 반환하는 기능을 수행

## 2) 리뷰 크롤링 시작

- > 앞서 만든 `amazon_scraper` 함수와 반복문을 이용해 전체 리뷰 페이지수만큼 크롤링하여 csv 파일로 저장
- > **[로맨스] 베스트셀러 100권**의 csv 파일 전체를 통합하여 하나의 리뷰데이터 파일을 만들

	A	B	C	D
	pid	title	stars	comment
1				Last year, I read a book called How Not to Die Alone, and I couldn't help but draw comparisons to it while reading this. Both feature lonely, middle-aged guys working in industries that deal with death, who develop feelings for married women, and make the realization that they need to make some changes if they don't want to, well, die alone. In a lot of ways, the comparisons end there, but the tones of the book remain similar - as serious and dark as they are funny and poignant. This story follows Oliver, a 39-year-old funeral director who has had feelings for a woman for years. He waited too long, and I still feel short.
2	B07SDRG638	Serious with shots of humor	3	
	B07SDRG638	Boring and Dreadful	1	I got this book for free through Prime, and I still feel short-changed. Should have picked literally anything else. If you think middle-aged, boring men having the world's most polite rebellion where he firmly and kindly tells his mum to accept his judgement sounds like a worth-while read, then read this! Otherwise, give it a miss. It was not only painfully spare on any interesting detail of actual mortuary arts or gallows humor, but the prose was painful and the narrator's voice took pains to explain the broader allegory / metaphor for the reader. Just in case you didn't get it the first time, First Reads can be hit or miss; let's cautiously crack this one open and see what it has to offer. They say old habits die hard, which is true, but I also think that new habits are born easily when you live on your own. Like talking to yourself. Ordering pantry items in alphabetical order. Storing balled-up socks in colour-coordinated rows in a drawer. Buying microwave meals for one when you should be cooking from scratch, because you really do love to cook. And fantasising <U+2013> writing lists of things you'd like to do, imagining a life that isn't yours, all from the comfort of your
3	B07SDRG638	Perfect first book of the year and decade	5	
4	B07SDRG638		5	I almost put 4 stars, however it feels like people only rate stuff 5 stars or 1 star and I thoroughly enjoyed this read. The main character was quirky and oddball yet relatable. His daily issues were ones that many struggle with, like family expectations, goals and relationships. The others in his life were likable, even the occasional annoying characters. Was not sorry I picked this.

\*로맨스 분야 베스트셀러 100권  
리뷰 50,838개 merged.csv



# 2

## 전처리 & 형용사, 명사 사전 생성

### 1) 리뷰 데이터 전처리

> 전체 리뷰 데이터 파일을 데이터프레임으로 불러온 뒤  
제품아이디(pid)와 코멘트만 남기고 나머지 열 제거 &  
NA값 제거

### 2) 형용사, 명사 사전 생성

> 형용사, 명사 dictionary list를 받아와서 형용사 사전과  
명사 사전을 생성함

	word1		word2
1	abandoned	1	abbreviation
2	able	2	abbreviations
3	absolute	3	abettor
4	adorable	4	abettors
5	adventurous	5	abilities
6	academic	6	ability
7	acceptable	7	abrasion
8	acclaimed	8	abrasions
9	accomplished	9	abrasive
10	accurate	10	abrasives
11	aching	11	absence
12	acidic	12	absences
13	acrobatic	13	abuse
14	active	14	abuser
15	actual	15	abusers
16	adept	16	abuses

<형용사 dic>

<명사 dic>



## 바이그램 생성

### 1) 바이그램 생성

> 형용사, 명사 사전 과 전처리한 데이터를 비교해서 '형용사-명사' 바이그램을 생성함.(전체 문장에서 형용사, 명사 조합을 추출한 것)

> 로맨스 소재별 핵심 키워드들이 추출됨

- lonely
- aged
- married
- .
- .
- .

	word1	word2
1	book	called
2	draw	comparisons
3	feature	lonely
4	lonely	middle
5	middle	aged
6	aged	guys
7	develop	feelings
8	married	women
9	book	remain
10	remain	similar
11	poignant.this	story
12	funeral	director
13		

# ‘형용사-명사’ 추출하는 이유?



영어 리뷰를 단순히 명사(noun)로만 추출하는 것이 아니라 **형용사(adjective)**를 함께 포함했기 때문에 보다 세밀한 추천이 가능해짐



대중의 일반화된 선택이 아닌, **개인의 취향**에 따라 책을 선택할 수 있음



실제 독자들의 **생생한 리뷰**에서 추출된 유의미한 키워드를 바탕으로 취향을 선택하기 때문에 추천 정확도가 높아짐.

# 4

## 추출된 '형용사-명사' 데이터 빈도 분석

### 1) 추출한 바이그램 빈도 분석

> 워드클라우드, 히스토그램, 연결망을 통해  
형용사-명사의 빈도를 시각화함.

### 2) 후처리(post-processing)

> 중복되는 의미의 단어 및 불필요한 단어를  
제거하여 유의미한 결과 도출!

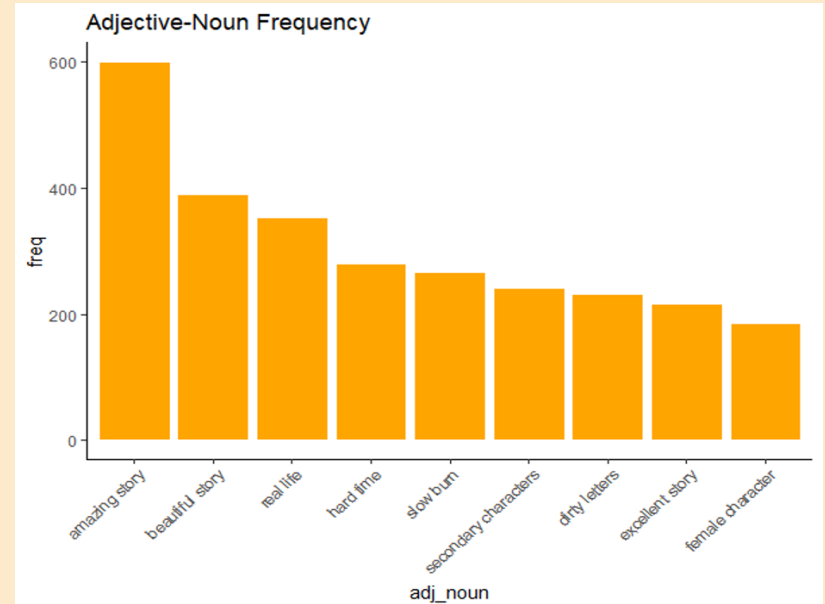
	adj_noun	freq
1	amazing story	601
2	beautiful story	390
3	real life	353
4	hard time	281
5	slow burn	266
6	secondary characters	242
7	dirty letters	232
8	excellent story	216
9	female character	185
10	sweet story	183
11	true love	177
12	previous storys	171

<adj\_noun freq>

# '형용사-명사' 데이터 워드클라우드



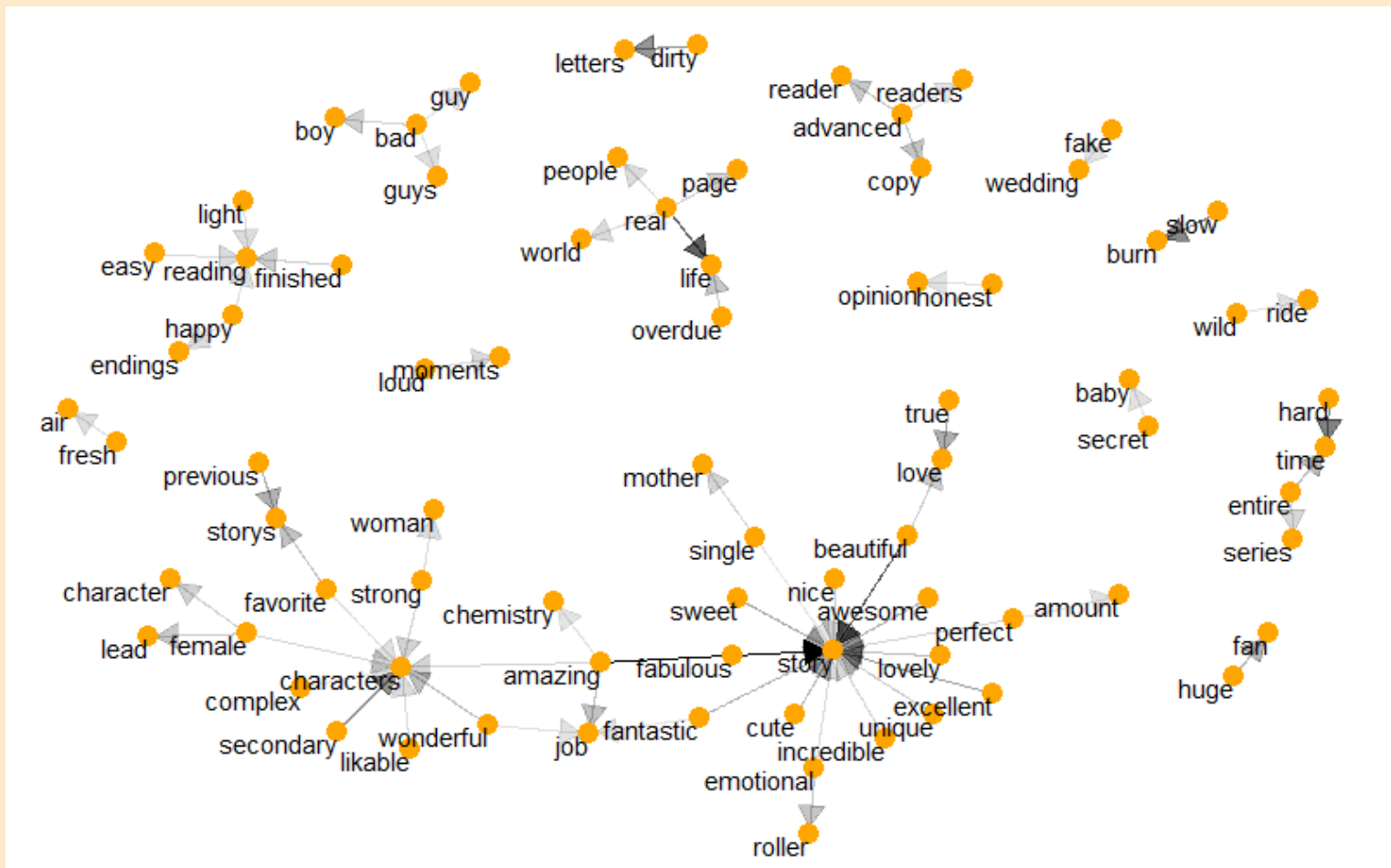
## '형용사-명사' 데이터 빈도그래프





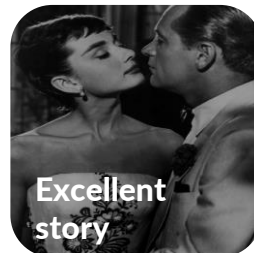
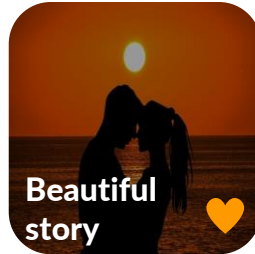
## 리뷰 분석 결과 시각화

## '형용사-명사' 데이터의 연결망



# ‘Choose your favorite’ 서비스의 활용

당신의 [로맨스] 취향을 알려주세요!



+더보기

취향 발견하셨나요?

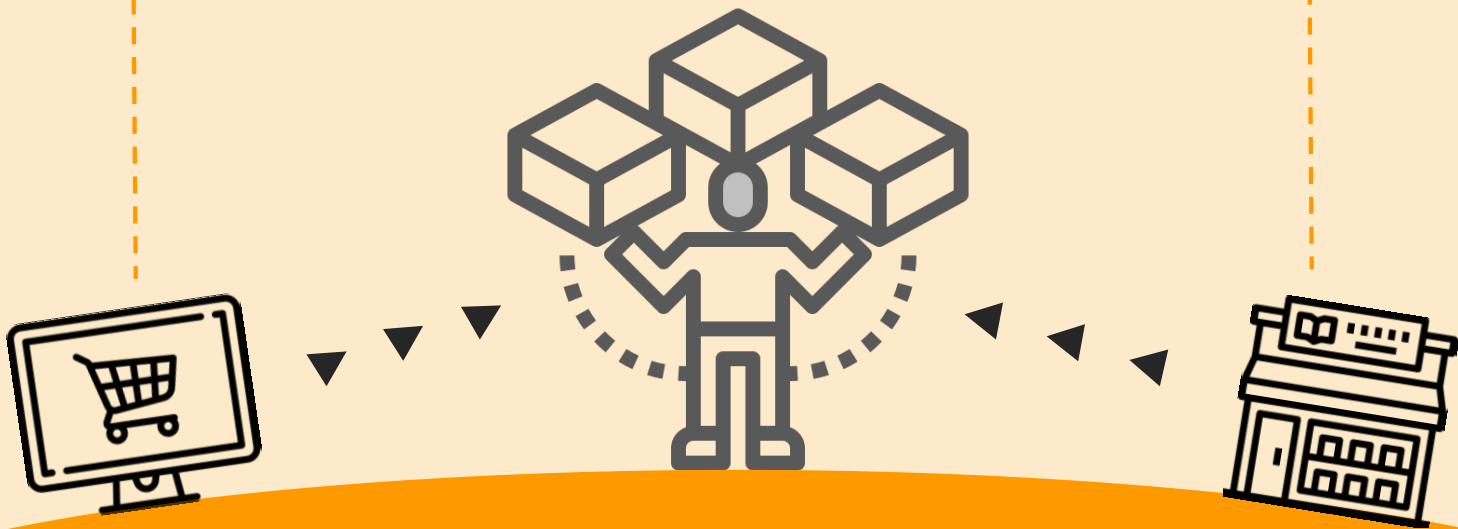
완료

# 효과 INSIGHT

세심한 북 큐레이팅을 더하여+  
기존보다 더 정확히 취향의 책을 추천하는  
**'Choose your favorite'** 서비스

구매 데이터에 의한  
추천 알고리즘

오프라인 서점의  
세심한 '큐레이팅'



amazon books

❄️ Thank you! ❄️

Q & A

