



Predict Future Sales

인사이트 도출



주제



데이터 관찰



정리

---

주 제 .

What I choose is..

Dealing with real data of a firm at

The Kaggle logo, featuring the word "kaggle" in a blue, lowercase, sans-serif font.

**Welcome to Kaggle Datasets**











The best place to discover and seamlessly analyze open data

Launch  
4 months ago

Close  
7 months

711 804 5,237  
Teams Competitors Entries

Points This competition does not award standard [ranking points](#)  
Tiers This competition does not count towards [tiers](#)

#	△1w	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	cumicumi			0.86039	10	1mo
2	▲3	Zoltan Szabo			0.87197	128	4d
3	▼1	Pieter Voloshyn			0.87218	1	2mo
4	▼1	timeray			0.87966	1	4mo
5	▼1	terr01			0.88306	6	3mo
6	new	Rustem Bekmukhametov			0.88869	1	6d
7	▲2	Share			0.89757	12	4h
8	▼2	Janos			0.89776	1	25d
9	▲76	Ho Yin Chan			0.89909	59	9h
10	▼3	miewmiewman			0.89911	1	2mo

# Predict Future Sales

본 Competition의 목적.

In this competition you will work with a challenging time-series dataset consisting of daily sales data, kindly provided by one of the largest Russian software firms - **1C Company**.

We are asking you to predict total sales for every product and store in the next month. By solving this competition you will be able to apply and enhance your data science skills.

:회사 데이터를 제공->미래(다음달) 판매량  
예측



데이터 관찰

.

# 데이터 관찰

## File descriptions

- sales\_train.csv - the training set. Daily historical data from January 2013 to October 2015.
- test.csv - the test set. You need to forecast the sales for these shops and products for November 2015.
- sample\_submission.csv - a sample submission file in the correct format.
- items.csv - supplemental information about the items/products.
- item\_categories.csv - supplemental information about the items categories.
- shops.csv- supplemental information about the shops.

---

## column descriptions

- ID - an Id that represents a (Shop, Item) tuple within the test set
- shop\_id - unique identifier of a shop
- item\_id - unique identifier of a product
- item\_category\_id - unique identifier of item category
- item\_cnt\_day - number of products sold. You are predicting a monthly amount of this measure
- item\_price - current price of an item
- date - date in format dd/mm/yyyy
- date\_block\_num - a consecutive month number, used for convenience. January 2013 is 0, February 2013 is 1,..., October 2015 is 33
- item\_name - name of item
- shop\_name - name of shop
- item\_category\_name - name of item category



# 데이터 ; 관찰

```
sales_train.head()
```

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0

```
shops.head()
```

	shop_name	shop_id
0	!Якутск Орджоникидзе, 56 фран	0
1	!Якутск ТЦ "Центральный" фран	1
2	Адыгея ТЦ "Мега"	2
3	Балашиха ТРК "Октябрь-Киномир"	3
4	Волжский ТЦ "Волга Молл"	4

```
item_categories.head()
```

	item_category_name	item_category_id
0	PC - Гарнитур/Наушники	0
1	Аксессуары - PS2	1
2	Аксессуары - PS3	2
3	Аксессуары - PS4	3
4	Аксессуары - PSP	4

```
items.head()
```

	item_name	item_id	item_category_id
0	! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
1	IABBY FineReader 12 Professional Edition Full...	1	76
2	***В ЛУЧАХ СЛАВЫ (UNV) D	2	40
3	***ГОЛУБАЯ ВОЛНА (Univ) D	3	40
4	***КОРОБКА (СТЕКЛО) D	4	40

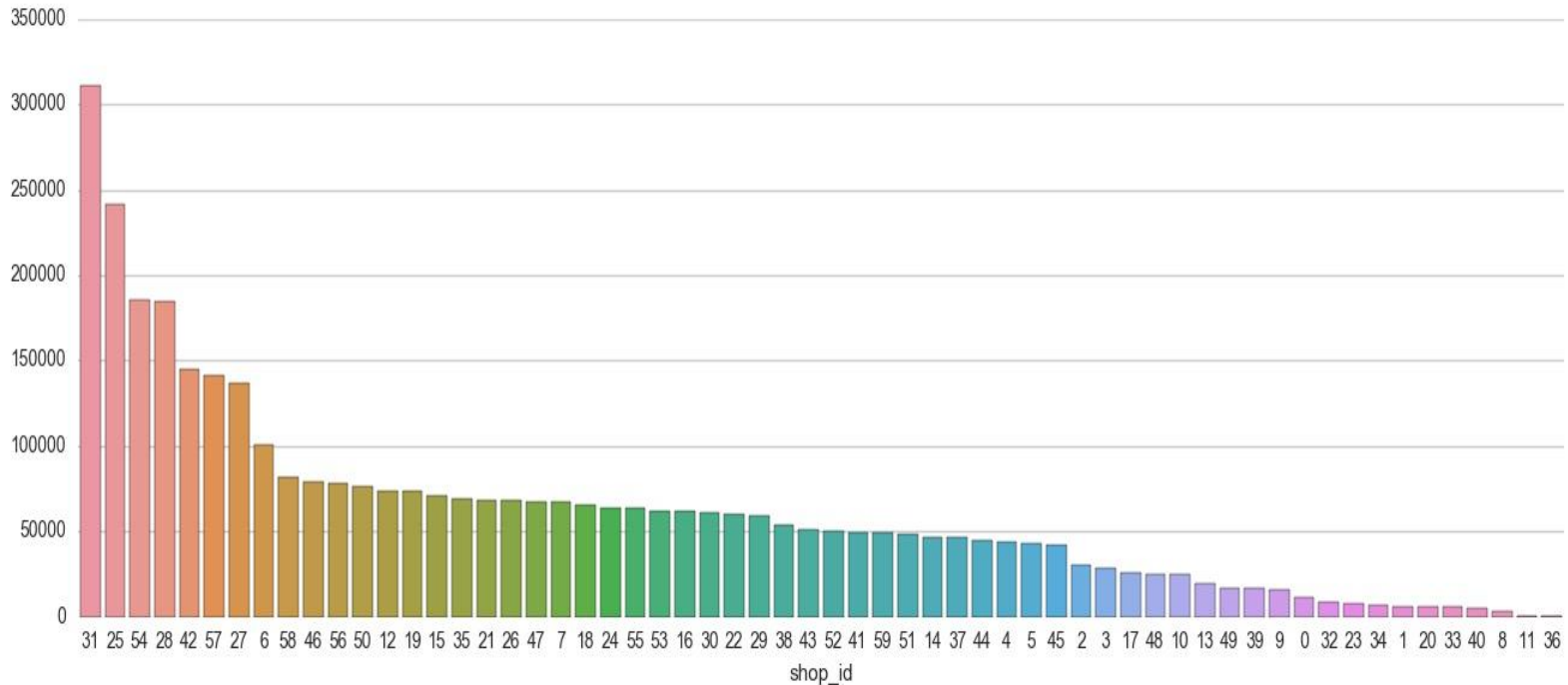
# 데이터 ; 관찰

## 필요한 정보 뭉치기

date	date_block_num	shop_id	item_id	item_price	item_cnt_day	item_name	item_category_id	item_category_name_eng
2013-01-01	0	18	5823	2500.0	1.0	Playstation Store пополнение бумажника: Карта ...	35	Payment cards - PSN
2013-01-01	0	27	5573	849.0	1.0	PS Move Motion Controller Charger (Зарядная ст...	2	Accessories - PS3
2013-01-01	0	7	1006	399.0	1.0	3D Crystal Puzzle Дельфин XL	67	Gifts - Development
2013-01-01	0	19	17707	899.0	1.0	Праздник спорта (Essentials) (только для PS Мо...	19	Games - PS3
2013-01-01	0	14	19548	149.0	1.0	ТАКИЕ РАЗНЫЕ БЛИЗНЕЦЫ (регион)	40	Cinema - DVD

## 데이터 ; 관찰

상점마다 판매량 크게 다를까?  
=> 상점마다 판매 격차가 있는 편



데이터 ;  
추론

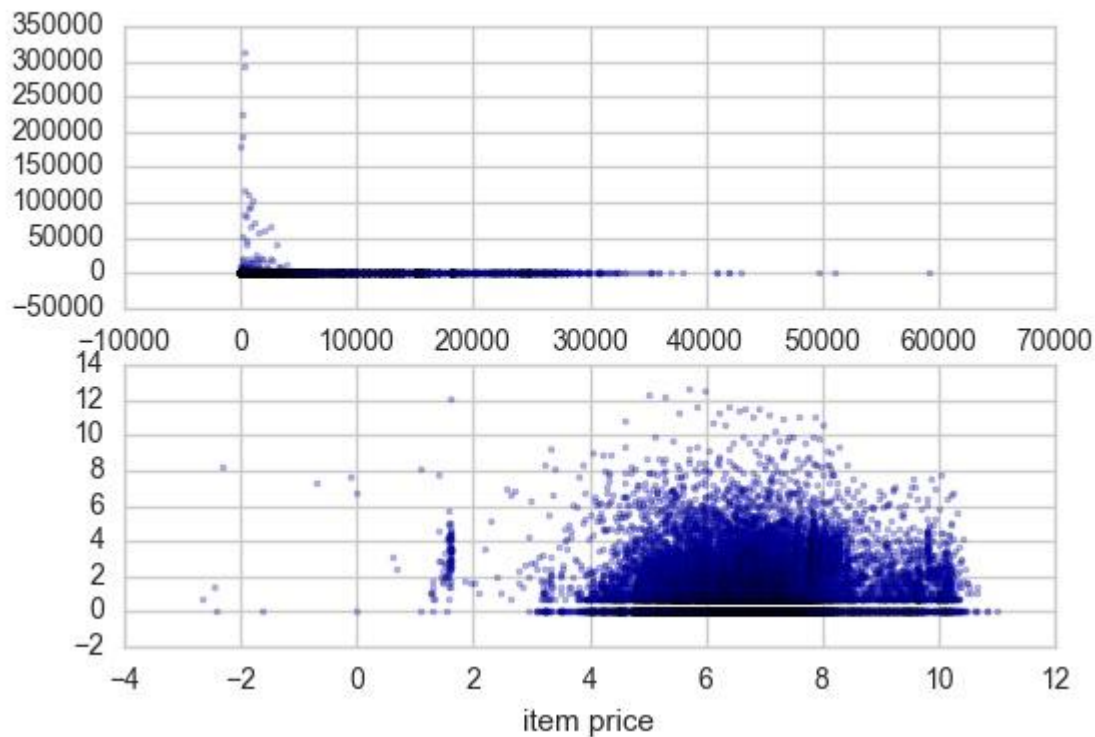
가격에 따른 판매량도 다르지 않을까?

추측1: 비싸면 덜 팔린다..

추측2: 싼 게 아무래도 많이 팔리겠지..

데이터 ;  
관찰

가격은 판매량과 뚜렷한 상관관계를 보이지 않음을 알아냄!!



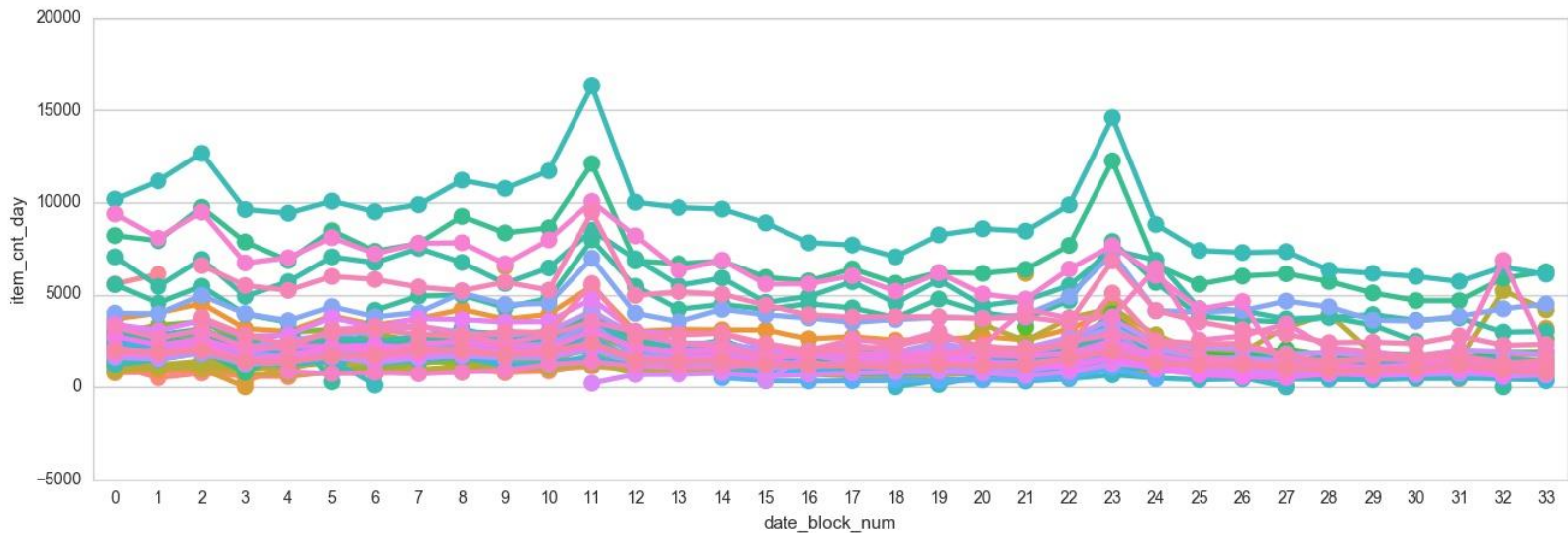
데이터 ;  
관찰

상점의 월별 판매량도 다른지 보고 싶어..

	date_block_num	shop_id	item_cnt_day
0	0	0	5578.0
1	0	1	2947.0
2	0	2	1146.0
3	0	3	767.0
4	0	4	2114.0

데이터 ;  
관찰

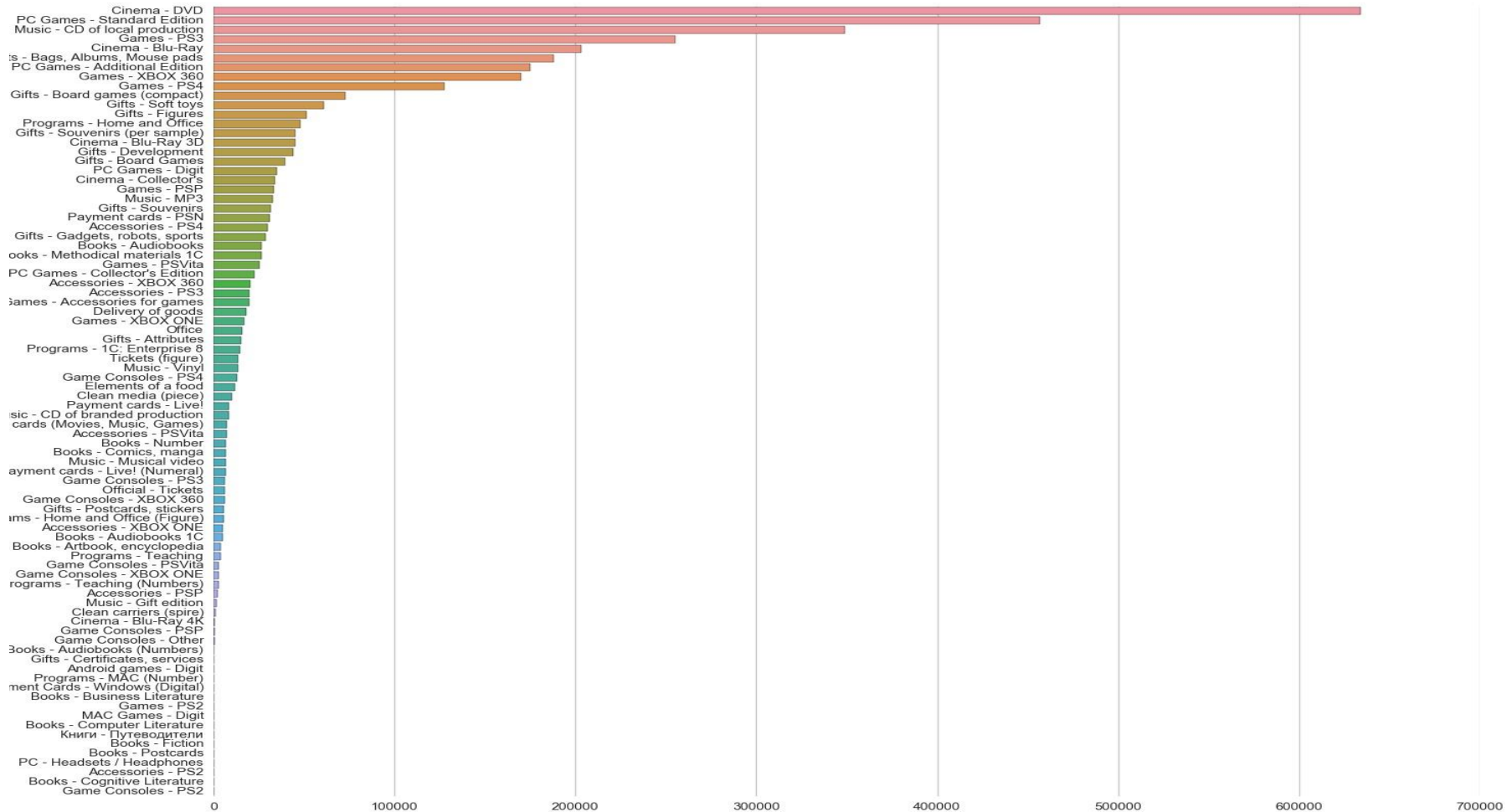
상점별 판매 추이에서 12월에 급등하는 계절성이 보인다. -> 크리스마스? 연말 행사? 선물?



데이터 ;  
관찰

카테고리별 판매도 보자

=>유독 잘 팔리는 품목이 몇 있다!





데이터 ;  
추론

DVD, game, music 카테고리 판매량 다수

게임, 음악, 영화 == 단발 소비성 콘텐츠

출시 이후 판매량 감소하겠지..?

=>제품 출시 이후 시간 경과에 따른 판매량 추이를 보자!

데이터 ;  
관찰

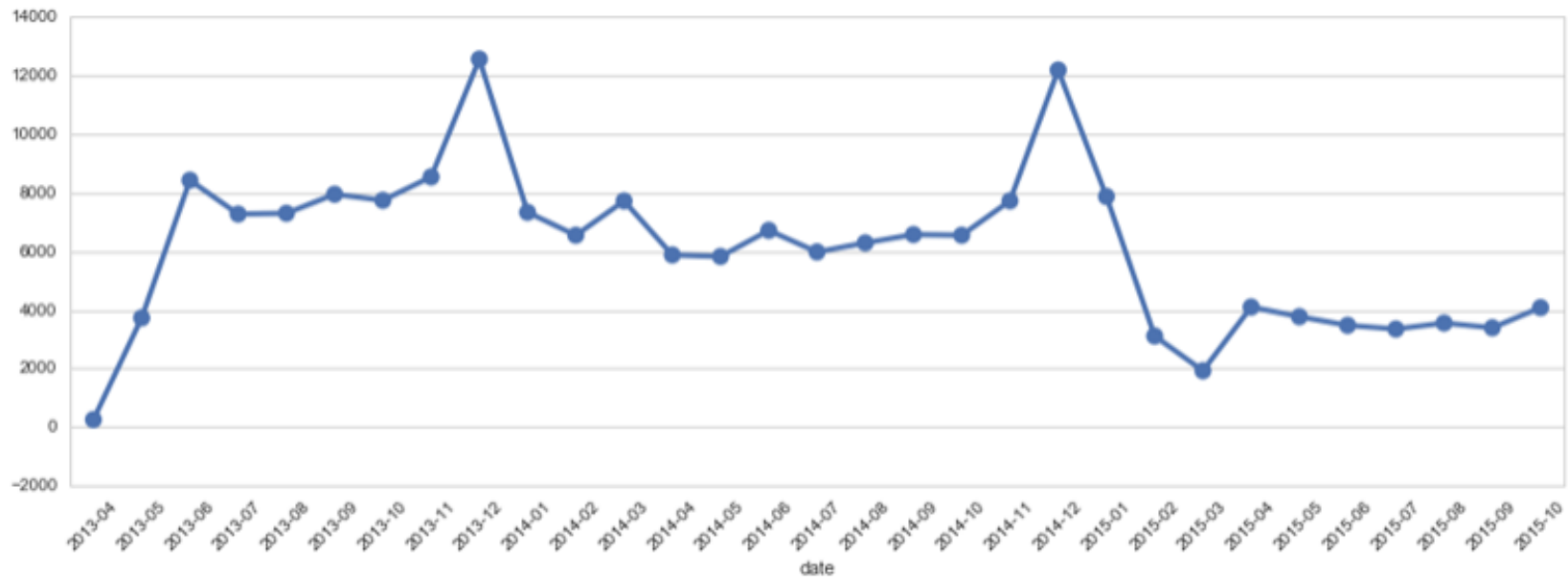
T-shirts  
:기념품

- 대체로 꾸준 판매
- 12월 급등

제품명 : 20949 Corporate package T-shirt 1C Interest white (34 \* 42) ...

카테고리 : 30 PC Games - Standard Edition

Name: item\_category\_name\_eng, dtype: object



데이터 ;  
관찰

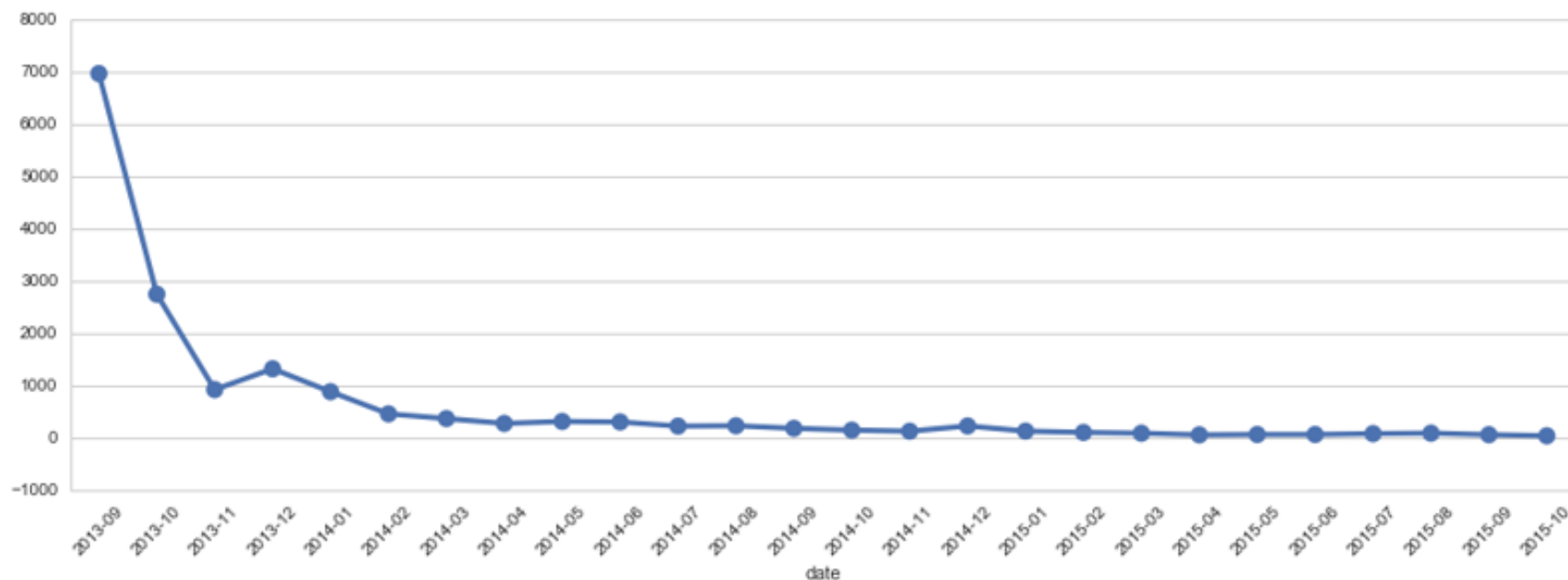
GTA5  
:PS게임

· 전형적 게임 판매 양상

제품명 : 3732 Grand Theft Auto V [PS3, Russian subtitles]

카테고리 : 30 PC Games - Standard Edition

Name: item\_category\_name\_eng, dtype: object

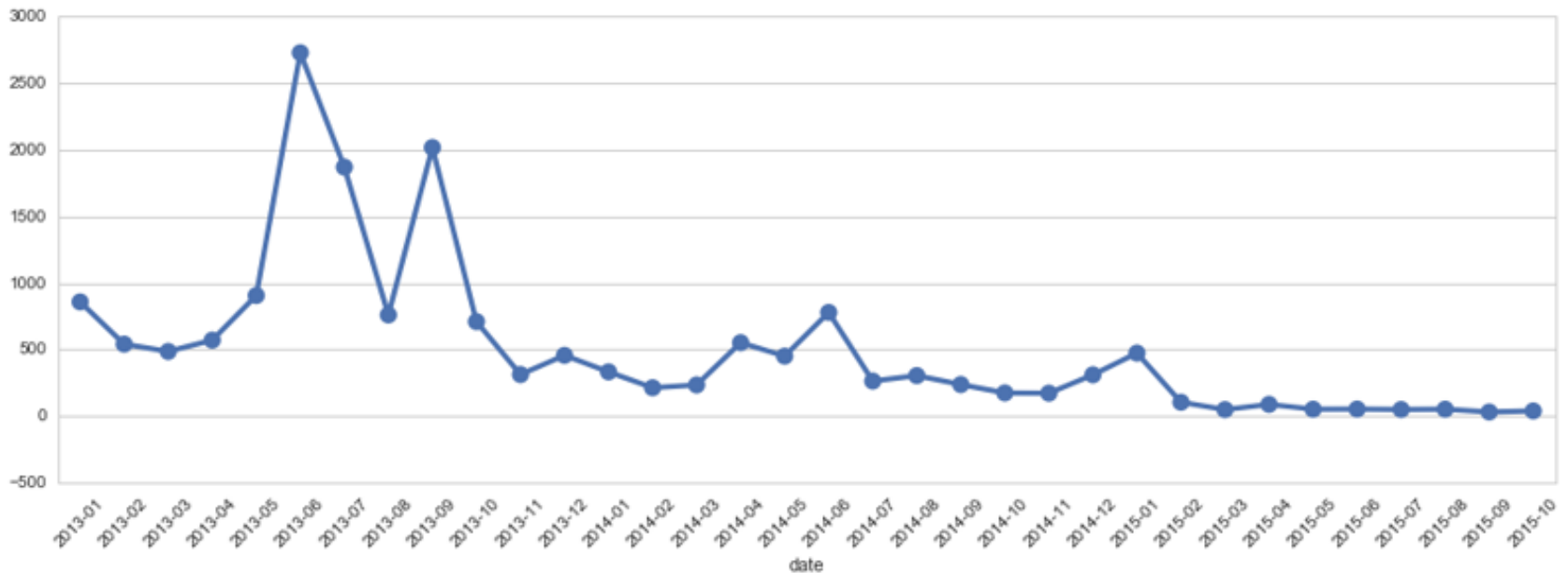


데이터 ;  
관찰

Diablo3  
:PC게임

- 확장 팩 발매 -> 잠시 급등
- 이후 일반적 양상

제품명 : 2808 Diablo III [PC, Jewel, Russian version]  
카테고리 : 30 PC Games - Standard Edition  
Name: item\_category\_name\_eng, dtype: object

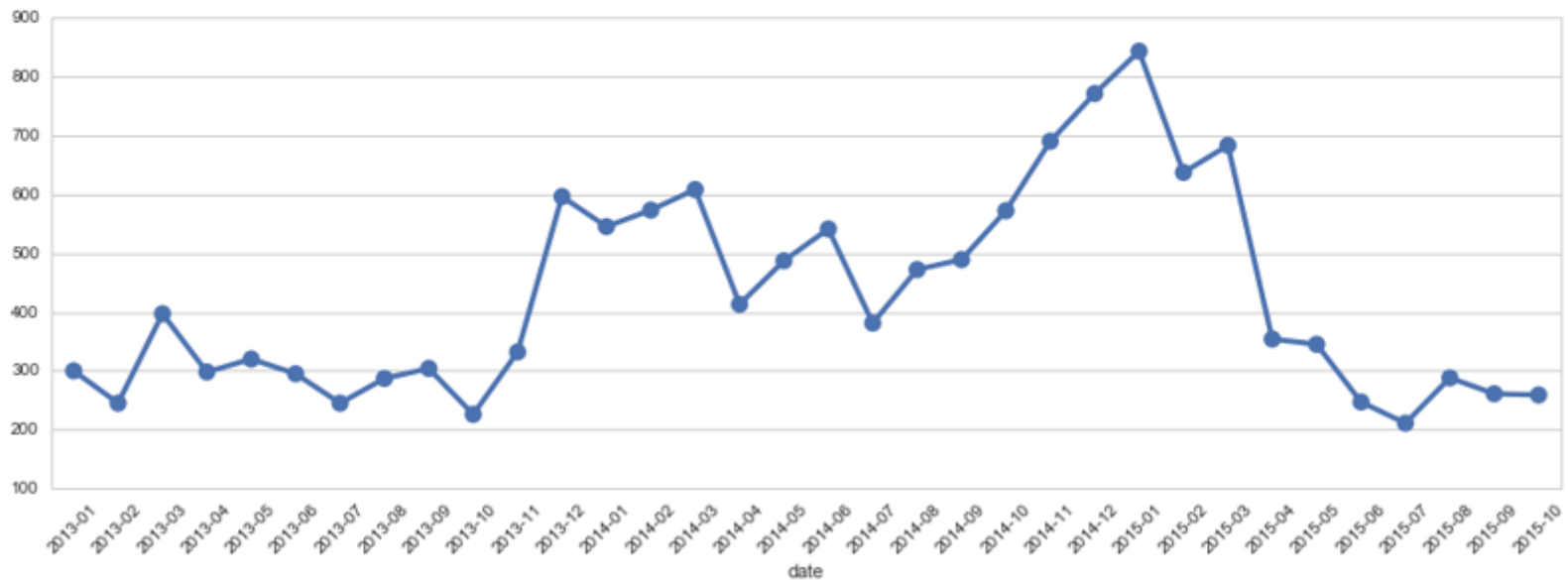


데이터 ;  
관찰

Card  
:캐시템

• 불규칙적 판매

제품명 : 5822 Playstation Store wallet replenishment: Card ...  
카테고리 : 30 PC Games - Standard Edition  
Name: item\_category\_name\_eng, dtype: object



정 리.

## 정리 ; Insight

하나 .

상점 별 판매 격차O  
가격은 별 영향X

둘.

품목 세분화 필요!

셋.

계절성 보임.  
제품별 시기에 따른 경향성.

## 정리 ; 감상

하나 .

데이터 다루는 게 쉬운 일이 아니다.

둘.

데이터를 직접 보지 않으면 잘못된 생각을  
할 수 있다.

셋.

는

큰 규모의 데이터를 관찰하기 위해서 시각화  
선택이 아닌 필수!



---

גב  
גב .

---