네이버 쇼핑 상품 리뷰 수 예측

훈훈한 2조_훈2 이은서 / 김재정 류가연 / 한상훈



1 프로젝트 개요

2 프로젝트 진행 과정

3 분석 결과 및 평가

extstyle 4 Q&A

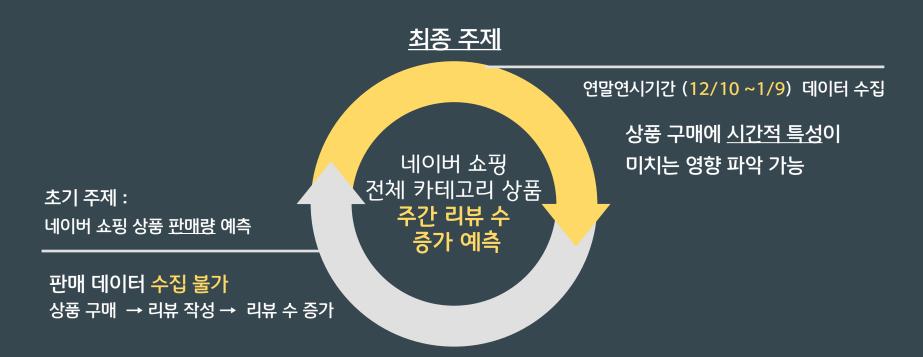
1. 프로젝트 개요

<u> 1 프로</u>젝트 주제 및 선정 배경

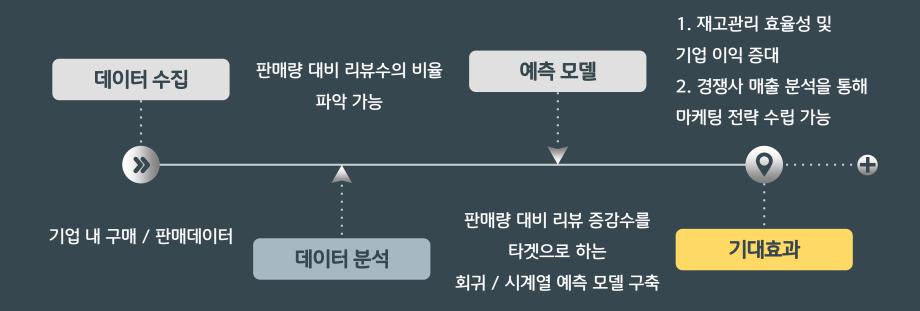
2 프로젝트 목표

3 프로젝트 기대효과

프로젝트 주제 및 선정배경



프로젝트 목표 및 기대효과



2. 프로젝트 진행과정

│ 데이터 수집

2 데이터 전처리 및 EDA

3 분석 Dataset

4 모델링 - 회귀/시계열

데이터 수집-구성

- 기간: 2021년 12월 10일 ~ 2022년 1월 9일 (30일간)
- 대상 : 네이버 쇼핑 11개 카테고리 별 랭킹 1~400위 상품 (매일 4,400개)



▲ 네이버 쇼핑 內 네이버페이 상품목록 페이지

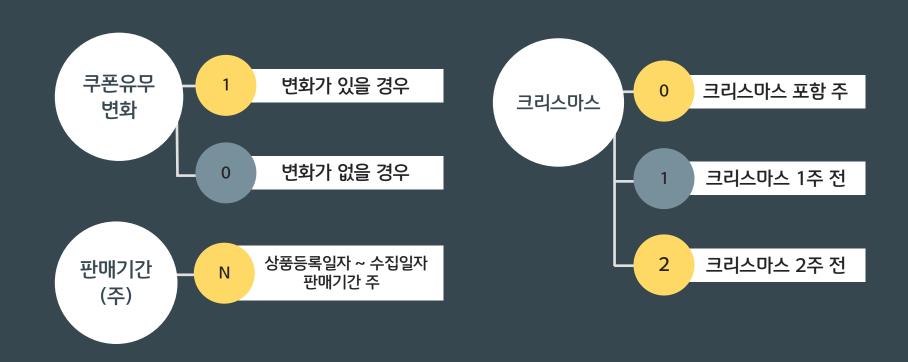
데이터 수집 - 분석 Dataset

	수집일자	카테 카티 고리 리즘 대분 리즘	I고 카타 분 고리 류 소를	카테그 리세는	고 상위	à	강품명	상품아이디	판매 자	판매자등	굿서비 스	가격	등록일자	리뷰 건수	구매 건수	평하기 수	마하다	벤 유 무	배송	배송정보	적림포인:	편이저미	검색 키워 드	태그	추가 이미 지개 수	비디 오유 무	이미지url	이미지사이 즈	이미지 컬러코 드	리뷰건 수합
						국내매장판 내셔널지요. 2021 FW 경량 패딩 정패 여성 숏패딩 겨물														V2I#!^~覧網 ^^^3000^6000	^990^990^^^0	0 4.8	NaN	간절기패딩,남자경량패딩, 남자패딩,남자숏패딩,깔깔 이	6	0	https://shopping- phinf.pstatic.net/main_834778	1000x1000	7	284
1	2021-12-10 17:10:54.575132	패션 남성 의류	의 류 정프	i Nat	N 2	209 에선셜 람포지퍼 덕다 퍼 숏 패딩 자켓 - 헤비다 장패딩 겨울 커플	운푸 운대 83 점퍼	3548281046	자바 나스	M44002	True	109700 20	0211117115618	968	0	517.0	1	0	0	v2i#i^小型網	^1097^1097^^^0	0 4.9	NaN	남자숏패딩,겨울패딩,남성 패딩,패딩점패	2	0	https://shopping- phinf.pstatic.net/main_835482	840x840	8	1011
	2021-12-10 17:10:54.575132					[무배] 엘로 부드러운 골지 로이 용기모 반당 조거														v2i#i^~戰배	^148^148^^0	0 4.7	NaN	용기모바지,여자기모바지	0	0	https://shopping- phinf.pstatic.net/main_834855	2000x2500	33	868
3	2021-12-10 17:10:54.575132	남성 패션 ^C 외류 어	전 시즌 위 성니 장옷	Lya	美 4	BYC 남성 여성 발열 기모 남자 여자 성인 내의 보온 타이즈 상의 하의	내복 겨울 81 서트	1193080648	관제 리아	M44002	True	9900 20	0181012105120	8516	6823	776.0	1	0	2500	V2I#I^^監H ^^^2500^5000	^99^99^^^0	0 4.6	NaN	남성내복,남성내의,발열내 복,발열내의,히트텍,겨울내 복,성인내복	1	0	https://shopping- phinf.pstatic.net/main_811930	1000x1000	0	8574
4	2021-12-10 17:10:54.575132	패션 남성 의류	취 점프	i Nai	N 5	국내애장판 내셔널지오그 21FW 홈페딩 숏페딩 리 타루가 여성 남성 구스다운	1래픽 이만 83 점패	3542825647	국내 정품	M44006	False	254000 20	0211109165513	31	210	346.0	1	0	3000	V2i#!^^택비	^2540^2540^^^0	0 4.9	NaN	구스패딩,겨울점퍼,캐주얼 점퍼,겨울아우터	5	0	https://shopping- phinf.pstatic.net/main_835428	800x800	37	33
5	2021-12-10 17:10:54.575132	여성 패션 G 의류 어	전 작용 의 음악	/ I Nat	N 6	수면잠옷 겨울 극세사 공주 파자마 곰물이 원피스 여성 면 순면 바지 I	잠옷 긴팔 82 디지	2674727081	멀키 대박	M44003	False	10900 20	0200922210810	2511	4883	3718.0	1	0	3000	v2(비오늘즐발^13:00^택배 ^^^3000^6000^^2.6^^^^	^109^109^^^0	0 4.7	NaN	여성실크잡옷,긴팔잠옷,커 풀잠옷,잠옷,홈웨어	7	0	https://shopping- phinf.pstatic.net/main_826747	1080×1080	32	2541
6	2021-12-10 17:10:54.575132	패션 여성 의류	의 네트	/ II Nat	N 7	부드러운 캐시미어 물 여성 겨울 목 폴라 터톨넥 스웨 트티 브라운 블랙 아이	기본 터니 82 보리	2237690733	코이 블리	M44003	False	18900 20	0210929163509	1339	1392	1740.0	1	0	2500	V2I#I^^턴비	^189^189^^^0	0 4.6	NaN	어리어리니트,데일리로,베 이직록,부드러운니트,이너 티셔츠,고급스러운니트,여 자겨물코디,여_	9	0	https://shopping- phinf.pstatic.net/main_822376	640x640	32	1350
7	2021-12-10 17:10:54.575132	여성 패션 다 의류 어	전위 점을	/ Il Nat	N 8	타탄 체크장옷 커플장옷 크 마스 파자마 파티 홀웨어 신혼 실내복 바지 면	크리스 기모 81 선	1729889372	골장 자니	M44003	True	16900 20	0190209231952	821	1055	1554.0	0	0	2500	√2!#!^^₹\#] ^^^2500^5000	^169^169^^^0	0 4.2	NaN	잠옷세트 여성잠옷세트,막 입기중은스타일,깔끔한잠 옷,쾌적한느낌,로맨틱러블 리	9	0	https://shopping- phinf.pstatic.net/main_817298	640x640	1	825
8	2021-12-10 17:10:54.575132	여성 패션 다 의류 어	전 웨 장	ųa.	等 9	유니클로 히트텍 대체 발일 여자 남자 성인 보온 내복 빅사이즈 상하의	내의 기모 82 네트	2678223586	한물 기 빛	M44004	True	9800 20	0200924152920	703	2543	246.0	1	0	0	√2!#!^^€[#]	^98^98^^^0	0 4.5	NaN	패셔너를 베이직코디,따뜻 한내복,막입기중은스타일, 단정,따뜻한소재,편안함을 선사하는,따뜻_	0	0	https://shopping- phinf.pstatic.net/main_826782	600x600	1	721
9	2021-12-10 17:10:54.575132	패션 남성 의류	의 류 점프	Na!	N 10	208 MASUDA 람포지퍼 운 푸퍼 숏 패딩 자켓 -해비 대장패딩 겨울 남지	역다 다운 83 점퍼	3548272147	자바 나스	M44002	True	143900 20	0211202180225	210	0	155.0	1	0	0	√2!#!^~₹(H)	^1439^1439^^^0	0 5.0	NaN	남자숏패딩,겨물패딩,남성 패딩,패딩점패	3	0	https://shopping- phinf.pstatic.net/main_835482	840x840	32	255
		Otto	101			크리스마스 잠옷 신혼부부 면 타탄 체크 커플 파자마 잠옷 바지 콜														v2I#I오늘출발^15:00^학배	^199^199^^^0	0 4.4	NaN	커플잠옷세트,커플홈웨어, 파자마료,귀여운잠옷	9	1	https://shopping- phinf.pstatic.net/main_826540	1000×1000	1	839
11	2021-12-10 17:10:54.575132	패션 여성 의류	류 바지	Na?	N 12	[무배]도톰 기모 골덴 배기 (3C)	팬츠 (F/L) 82	2311352598	오네 스타	M44004	False	26900 20	0211108141814	902	466	528.0	0	0	0	v2i#!^气管	^269^269^^^0	0 4.4	NaN	기모배기팬츠,기모배기바 지,골덴바지,골덴팬츠	7	0	https://shopping- phinf.pstatic.net/main_823113	450x454	520	909
12	2021-12-10 17:10:54.575132	패션 남성 의류 어	연위 편트 참옷	l Naf	N 13	컬빈클라인 언더웨어 CK 팬티 속옷 드로즈 사각 면 도	남자 세트 82 음전	2902367973	<u>78</u>	M44003	True	38800 20	0210126121216	3511	7820	923.0	1	0	0	v2I#I오늘출발^15:00^학배 ^^^2500^5000^1.8^^^^	^388^388^^^0	0 4.7	NaN	남자팬티,남자속옷,남성팬 티,남성속옷,팬티	4	0	https://shopping- phinf.pstatic.net/main_829023	1000×1000	32	3517
13	2021-12-10 17:10:54.575132	패션 여성 의류	의 트리	Nat	N 14	여성 트레이닝 맨투맨 츄리 트 조거팬츠 상하세트 77 즈 키작녀 오트밀 간절기	당세 7사이 82 기모	2786837708	오늘 만스 당	M44003	False	27900 20	0201127133320	1439	3904	3538.0	1	0	0	√2i#!^~₹₫₩! ^^~3000^6000	^279^279^^^0	0 4.6	NaN	조거팬츠록,간절기록	2	0	https://shopping- phinf.pstatic.net/main_827868	1000x1000	0	1473
14	2021-12-10 17:10:54.575132	패션 여성 의류	의 티시	Nat	N 15	겨울 도통하고 텐션 중은 무지 라운드 피지 기모 롱 티 여자목티 폴	이너 긴팔 82 라티	2721911725	호시 탐탐:	M44004	False	13590 20	0211026140043	600	841	645.0	0	0	3000	√2!#!^~₹\H ^^^3000^6000	^135^135^^^0	0 4.5	NaN	포근한티셔츠,질증은기본 티,기모긴팔티,여자목티,레 깅스록	7	0	https://shopping- phinf.pstatic.net/main_827219	2048×2048	32	607

데이터 전처리 - 과정



데이터 전처리 - 파생변수 추가



분석 Dataset

	카테고리대분류	카테고리중분류	카테고리소분류	판매자등급	굿서비스	가격	이벤트유무	배송비	배송정보	적립포인트	평점	검색키워드	태그	추가이미지개수	비디오유무	판매기간(주)	쿠폰유무변화	크리스마스
0	8	32	592	3	0	99000	0	1	0	990	4.8	0	5	6	0	7	0	2
1	8	32	592	2	1	109700	0	0	0	1097	4.9	0	4	2	0	3	1	2
2	8	31	381	2	1	9900	0	1	0	99	4.6	0	7	1	0	165	0	2
3	8	32	592	6	0	254000	0	1	0	2540	4.9	0	4	5	0	4	0	2
4	8	111	145	3	0	18900	0	1	0	189	4.6	0	8	9	0	10	1	2
8415	1	86	671	3	1	18000	0	0	1	190	4.8	0	2	8	0	158	1	0
8416	1	163	614	3	1	9900	0	0	1	297	4.8	0	8	1	0	47	0	0
8417	1	86	66	3	0	36900	0	0	0	369	4.7	0	9	0	0	98	1	0
8418	1	157	32	3	0	18000	1	0	1	180	4.9	0	2	0	0	467	1	0
8419	1	12	45	1	0	15190	1	0	0	0	4.8	0	0	0	0	10	0	0

8420 rows x 17 columns

▲ 전처리 후 최종 데이터셋

분석 Dataset

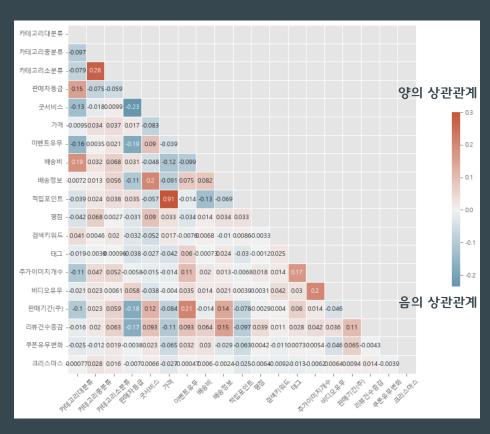
- 19개의 독립변수(범주형 변수 14개 | 연속형 변수 5개) + 1개의 종속변수(리뷰건수증감) = 총 20개
- Y (리뷰건수증감) = 1주일 후 리뷰건수합 ─ 당일 리뷰건수합 ex) 1/10 리뷰건수합 ─ 1/3 리뷰건수합 = 1주일 간 리뷰건수 증감

번호	변수명	번호	변수명	번호	변수명	번호	변수명
1	카테고리대분류	6	가격	11	적립포인트	16	비디오유무
2	카테고리중분류	7	찜하기수	12	평점	17	판매기간
3	카테고리소분류	8	이벤트유무	13	검색키워드	18	쿠폰유무변화
4	판매자등급	9	배송비	14	태그	19	크리스마스
5	굿서비스	10	배송정보	15	추가이미지개수	20	리뷰건수증감 (Y)

데이터 EDA / 모델링 - _{과정}



상관분석



다중공선성

다중공선성(VIF)

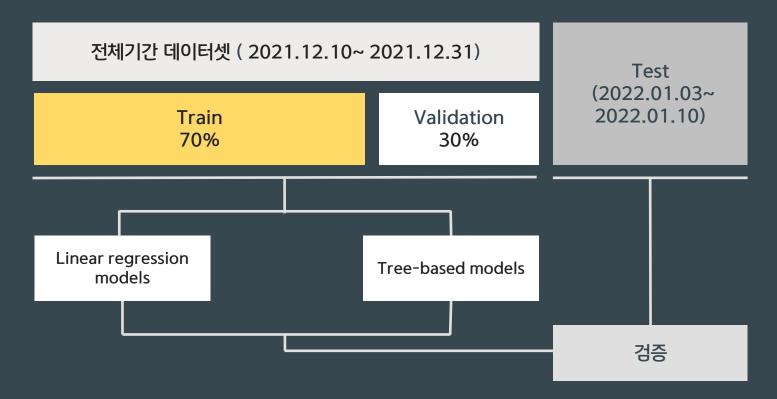
- 독립 변수들 사이에 상관 정도
 가 높을 경우, <u>부정적인 영향</u>
 을 끼칠 수 있음
- VIF 를 확인한 결과,

VIF >= 10 없기 때문에

변수 제거하지 않음

변수	VIF Factor	변수	VIF Factor
const	249.725	적립포인트	5.954
카테고리대분류	1.123	평점	1.020
카테고리중분류	1.105	검색키워드	1.010
카테고리소분류	1.102	태그	1.041
판매자등급	1.145	추가이미지개수	1.101
굿서비스	1.140	비디오유무	1.057
가격	5.974	판매기간(주)	1.123
이벤트유무	1.116	페이지증감	1.008
배송비	1.081	쿠폰유무변화	1.016
배송정보	1.086	크리스마스	1.003

회귀 모델링 - 데이터 분할



평가지표

R^2

결정계수, Coefficient of determination

- 선형회귀 모형이 실제 데이터를 얼마나잘 설명하는지 나타내는 지표
- 값이 1에 가까울수록 모형의 설명력이 높음

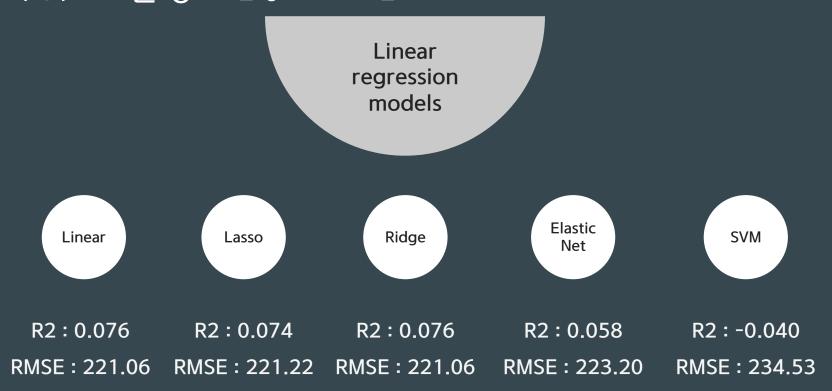
RMSE

'평균제곱근오차, $\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2}$

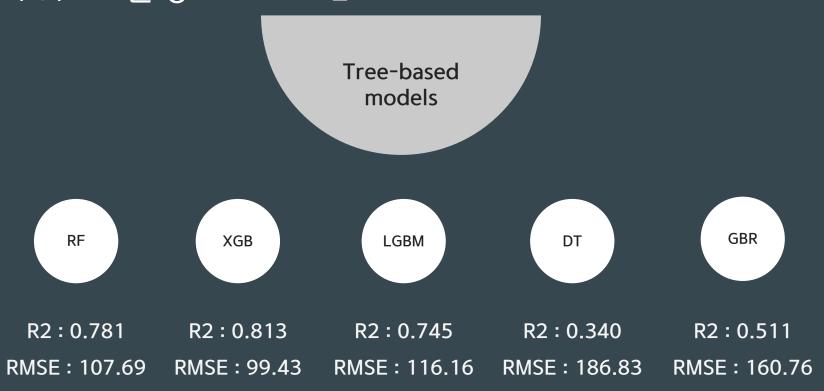
▶ 실제값과 예측값의 차들의 합

▶ 0에 가까울수록 좋은 값

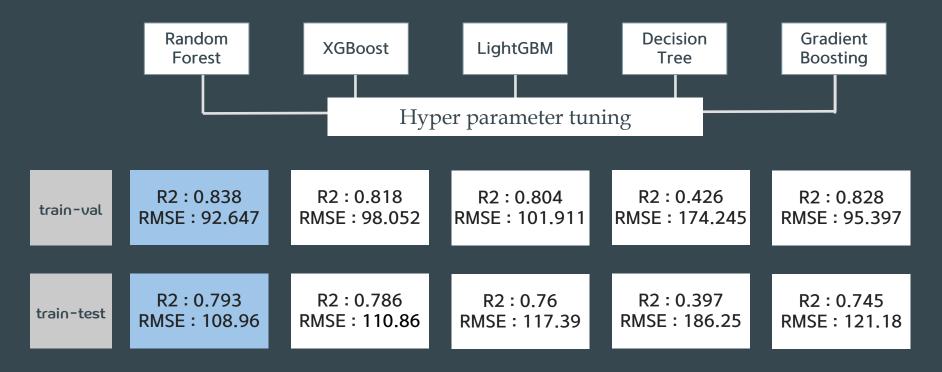
회귀 모델링 - 선형 회귀 기반



회귀 모델링 - 트리기반



회귀 모델링 - 파라미터 튜닝



회귀 모델링 - 앙상블

Averaging



train-val R2:0.848 | RMSE:89.494

train-test R2:0.785 | RMSE:111.052

XGB + GBR + RF / 3

train-val R2:0.847 | RMSE:89.707

train-test R2:0.793 | RMSE:108.945

회귀 모델링 - 앙상블

Stacking



train-val R2

R2:0.849 | RMSE:89.438

train-test

R2:0.801 | RMSE:106.902



train-val

R2:0.815 | RMSE:98.861

train-test

R2:0.822 | RMSE:100.995

시계열 분석 - 구성

[분석 배경]

회귀 모델과의 예측력 비교

[종속 변수]

Y = 리뷰건수합

[분석 과정]

연말연시 기간이 상품 판매량에 얼마나 영향을 미치는지 분석해보기 위하여 크리스마스 상품과 스테디셀러 상품을 선정하여 비교 분석

시계열 분석 - 분석상품

번호	크리스마스 상품	스테디셀러 상품
ı	하루클래스 크리스마스 생일 선물 취미 보석십자수	사무용 의자
2	닌텐도 스위치	애플워치 스트랩
3	보드게임 (커플 젠가)	세제
4	크리스마스 선물 수제쿠키	펩시제로콜라
5	홈파티음식배달	제주 렌트카
6	크리스마스 아기 실내복	유기농 아기과자
7	연말룩/파티룩	무지 반팔티
8	송예교 귀걸이	남자 양말
9	이솝 핸드크림 선물세트	메디힐 마스크팩
10	바라클라바	근육 테이프

시계열 분석 - EDA

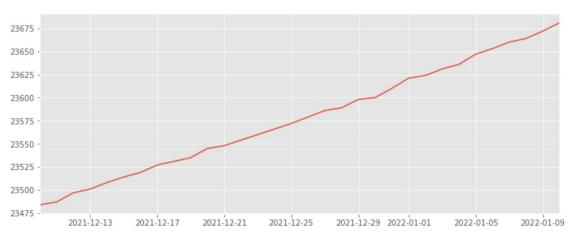
● 스테디셀러 상품: 제주 렌트카

리뷰건수합

수집약	일자
-----	----

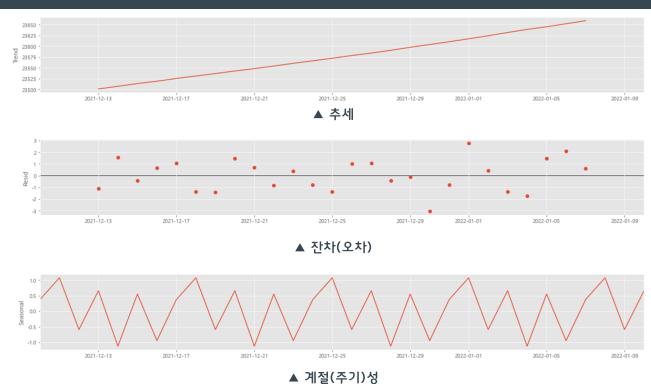
100'	
2021-12-10	23484
2021-12-11	23487
2021-12-12	23497
2021-12-13	23501
:	
2022-01-07	23660
2022-01-08	23664
2022-01-09	23672
2022-01-10	23681

리뷰건수합

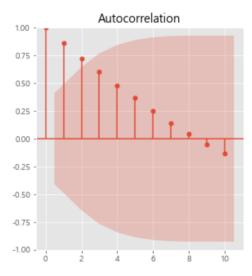


시계열 분석 - EDA

● 스테디셀러 상품: 제주 렌트카



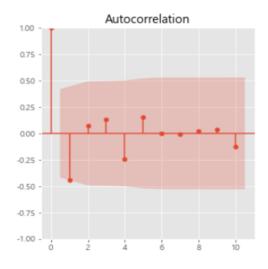
시계열 모델링 - 정상성(Stationary)



▲ 비정상 시계열 데이터

ADF p-value: 0.99

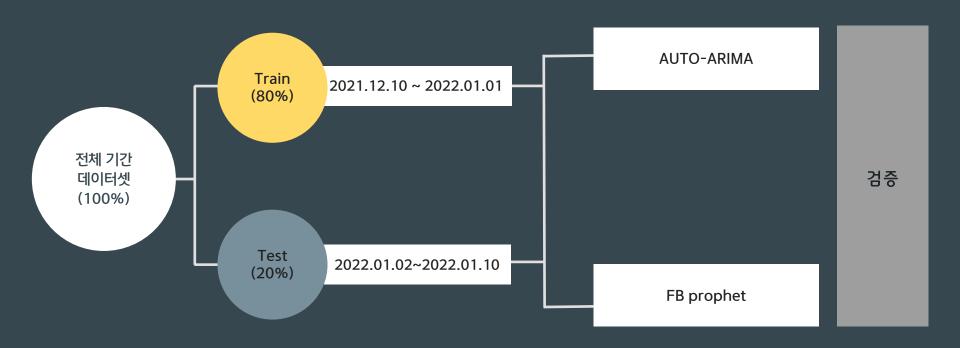




▲ 정상 시계열 데이터

ADF p-value: **0.001**

시계열 모델링 - 데이터 분할



시계열 모델링 - 결과

스테디셀러: 제주 렌트카

Seasonal: True

ARIMA

R2: 0.933

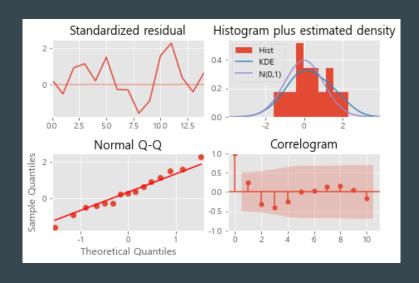
RMSE: 4.678

Holidays = 'KR'



R2: 0.941

RMSE: 4.394



▲ 정규성 확인

시계열 모델링 - 결과

크리스마스 : 십자수 선물세트

Seasonal: True

ARIMA

Holidays = 'KR'

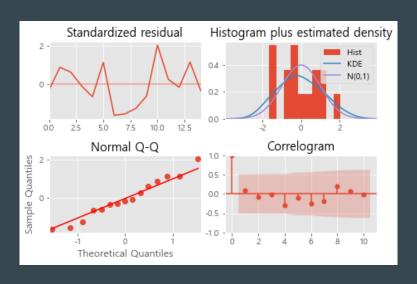
FB Prophet

R2: 0.923

RMSE: 40.635

R2:0.915

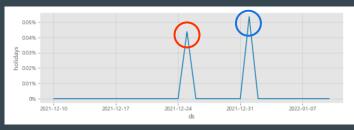
RMSE: 42.813



▲ 정규성 확인

시계열 모델링 - 결과

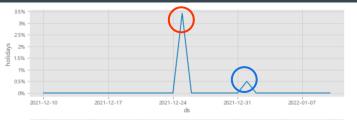
● 스테디셀러: 12/25 < 12/31 휴일 가중치(%) 높은 양상을 보임





● 크리스마스 상품: 12/25 > 12/31 휴일 가중치(%) 높은 양상을 보임





3. 분석 결과 및 평가

1 모델 검증 / 평가

2 모델 예측값 비교

3 향후 활용 방안

모델 검증 / 평가 - 회귀

Regression Best Model

Stacking Regressor

LGBM GBR RF Meta

train-val R2: 0.815 | RMSE: 98

train-test R2: <u>0.825</u> | RMSE: <u>100.99</u>



▲ 최종 모델 학습 곡선

모델 검증 / 평가 - 시계열

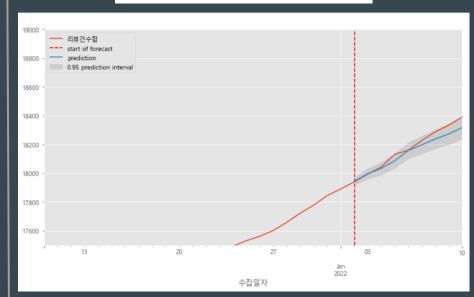
Time series Best Model

스테디셀러 : 여가생활편의-제주렌트카



FB Prophet R2:0.992 | RMSE: 166.594

크리스마스 : 가구인테리어-십자수



Auto-Arima

R2:0.923 | RMSE:40.635

모델 예측값 비교 - 여가/생활편의

● 연말연시 기간에 영향을 많이 받은 상품 vs 스테디셀러 상품 실제값과 예측값 비교

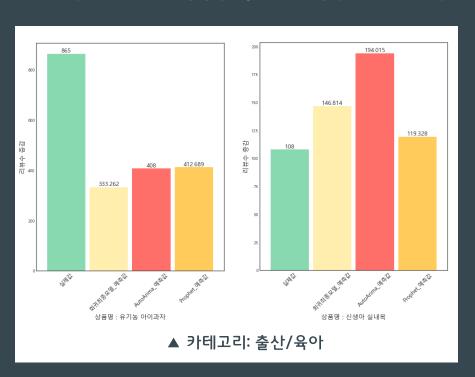




회귀와 시계열 모델 모두 실제값에 <u>근접하게 예측</u>

모델 예측값 비교 - 출산/육아

• 연말연시 기간에 영향을 많이 받은 상품 vs 스테디셀러 상품 실제값과 예측값 비교

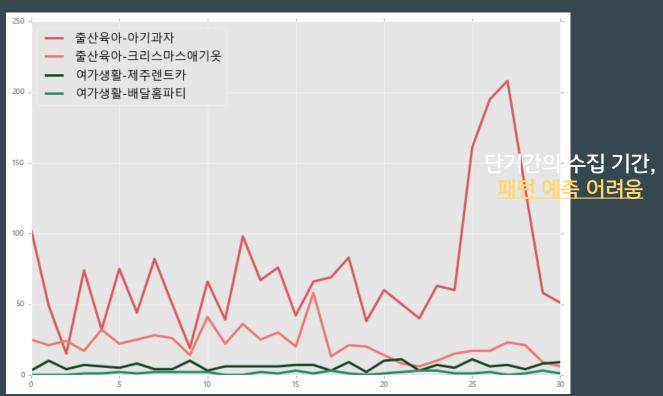




회귀와 시계열 모델 모두 예측력이 <u>떨어짐</u>

모델 예측값 비교 - 추후 분석

• 상품 별 증감치 비교하였을 때, 출산육아 변화량 > 여가생활 변화량



향후 활용 방안



재고 최적화

재고비용 감소로 인한 기업의 이윤 상승 효과 기대 불필요한 제품 생산 또는 제품 폐기로 인한 환경 오염 감소 효과 기대



최적 재고 생산 및 발주

예측을 통해 나온 최적 재고량을 통해 필요한 만큼 발주 및 생산



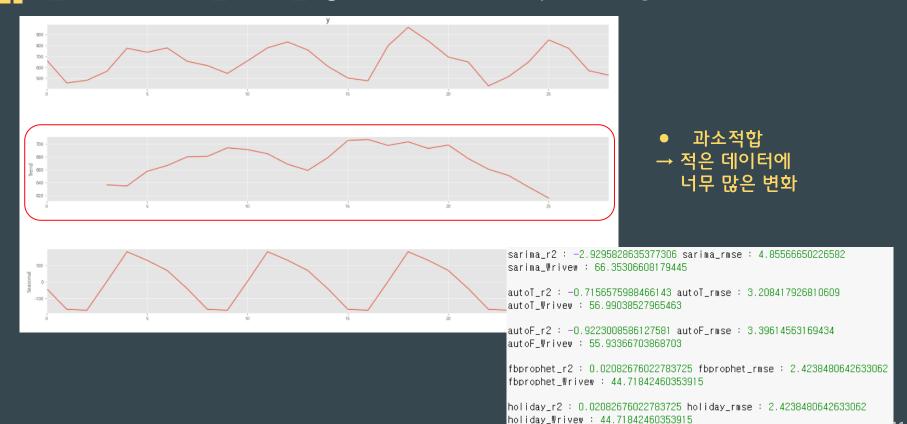
예측 모델 적용

예측 모델을 이용하여 상품의 주간 필요(적정) 재고를 예측

감사합니다.

Q & A

별첨) 시계열 모델링 리뷰 수 누적치 사용 이유



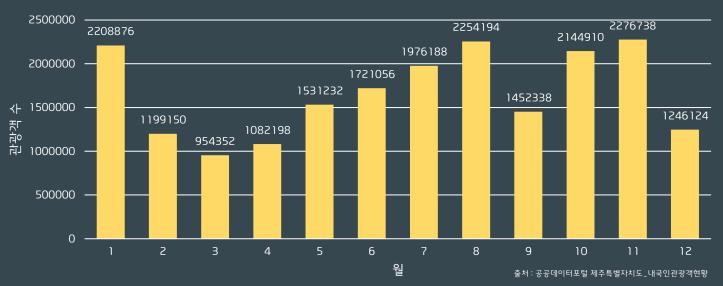
별첨) 네이버 구매건수 사용 못하는 이유

>구매건수 ≠ 판매량 결제건 당 상품 개수(=실제 판매량) 파악이 어려움 구매건수 집계기준 : 상품 1개 구매 - 상품주문번호 수 기준 ☆ ☆ ☆ ☆ : 상품 5개 구매 - 최근 6개월 (어제-181일 ~ 어제) - 추가주문상품 건수 제외 ◆ 하루 판매량 파악 불가능 - 취소/반품 주문건수 차감 12/27일의 구매건수 (취소완료, 반품환불완료,직권취소완료 모두 포 06/27 06/28 12/25 12/26 12/27 함) - 부정거래 주문건수 차감 12/26일의 구매건수 - 매일 1회 집계 12/27일의 구매건수 = 12/26일의 구매건수 - X + Y 출처: 네이버 스마트스토어센터 12/27일의 구매건수 - 12/26일의 구매건수 = - X + Y

즉. 우리는 X와 Y의 각각의 값을 알 수 없음

<u>별첨) 제주 렌터카 스테디셀러 상품 지정 이유</u>

2020년 제주도 내국인 관광객 현황



- 연말연시기간인 12월에 관광객 수가 11월에 비해 감소하였음을 확인
- ▶ 피서철인 6, 7, 8월엔 꾸준히 관광객 수 증가
- → 연말연시기간에 큰 영향을 받지 않는 스테디셀러 상품로 지정 후 분석