

BigData Visualization & MachineLearning

February 2018

Group3

Chanyeong Min

Jeheon Kwon

Minsung Kim

Younghyun Lee

| Group 3



인하대학교
아태물류

민찬영



인하대학교
산업경영공학

권제현



서울여자대학교
정보보호

이영현



단국대학교
데이터사이언스

김민성



Visualization (Jan 11 - Feb 01)

“Oracle Analytic Cloud로
빅데이터에서 인사이트 도출 ”

Machine Learning (Jan 28 - Feb 18)

“Oracle Big Data Cloud로
맞춤형 영화 추천 시스템 구축 ”

ORACLE®



Big Data Cloud Service

Agenda

- 데이터셋 및 전처리 과정
- 시나리오 소개
- 데이터 시각화

- 데이터셋 및 전처리 과정

• 시나리오 소개

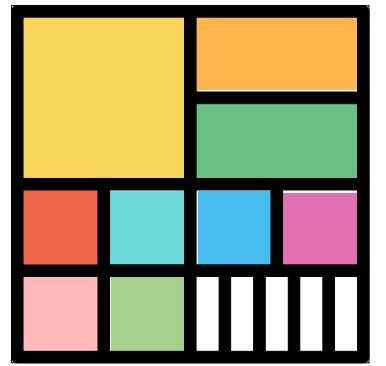
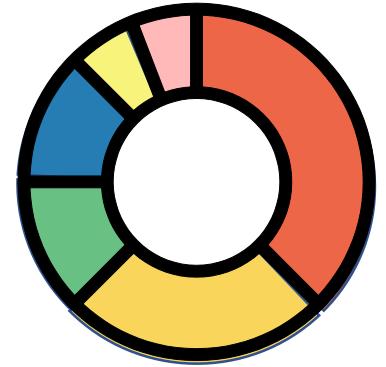
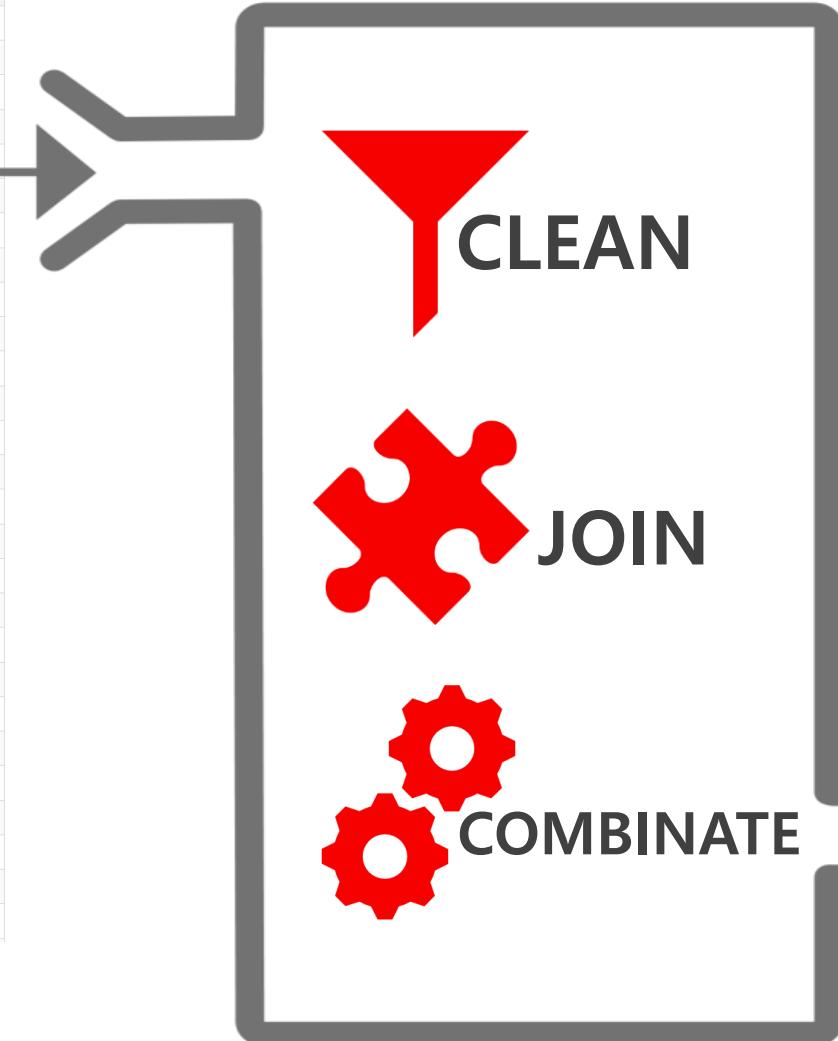
• 데이터 시각화

ORACLE®

데이터셋 소개 및 전처리 과정

품목명◆	HS코드◆	수출증량◆	수출금액◆	수입증량◆	수입금액◆	무역수지◆
전기기기와 그 부분품, 녹음기...	85	17,156,710.8	49,223,098	53,222,309.6	45,588,531	3,634,567
원자로 · 보일러 · 기계류와 이...	24	205,329.2	1,694,314	144,593.4	7,278,635	7,415,679
철도용이나 궤도용 의의 차량과...	77	40,734.7	46,021	185,913.3	5,816,660	329,362
광물성 연료 · 광물유(礦物油)와 ...	27	53,016.8	4,854,171	2,024.2	1,389,334	3,469,138
플라스틱과 그 제품	39	6,225,940.7	3,704,550	34,062,591.9	11,773,651	-8,069,101
선박과 수상 구조물	89	1,334,532.2	2,988,441	248,366.7	972,206	2,016,235
광학기기 · 사진용 기기 · 영화용 ...	90			83,097.9	84,772	2,304,408
유기화학품	29			22,748.3	1,925,507	430,080
철강	72			666,789.2	1,181,863	876,098
철강의 제품	73			1,940,876.4	1,366,267	674,543
고무와 그 제품	40			229,503.2	544,074	236,261
정유(essential oil)와 레진노이드(...	33			92,034.1	222,910	466,647
무기화학품, 귀금속 · 희토류(稀土...)	28			8,212.9	167,470	293,182
구리와 그 제품	74			420,766.0	675,893	-216,959
각종 화학공업 생산품	38			435,701	80,622.2	557,566
알루미늄과 그 제품	76			351,384	738,439	-387,055
메리야스 편물과 뜨질 편물	34	72,541.6	257,669	236,681.0	619,458	-321,836
종이와 판지, 제지용 풀(풀종)	46	30,711.2	21,569	55,461.3	21,86	262,383
비금속(非金屬)으로 만든 공구	91	259,750	16,690	203,792	15,587	78,680
인조필라멘트, 인조광직용 섬유...	54	45,804.3	249,574	55,461.3	44,440	168,612
의료용품	30	1,818.2	22,517	6,014.8	561,984	-335,367
유연용 · 염색용 추출물(extract), ...	32	2,678.	20,745	3,313.3	221,472	-20,429
천연진주 · 양식진주 · 귀석 · 반...	71	624.	19,219	4,331.1	273,128	-74,909
인조스테이플섬유	55	90,177.1	171,067	21,645.9	68,442	102,625
유리와 유리제품	70	30,708.4	154,532	116,353.4	193,022	-38,490
항공기와 우주선, 이들의 부품	88	435.7	154,073	230.7	114,620	39,453

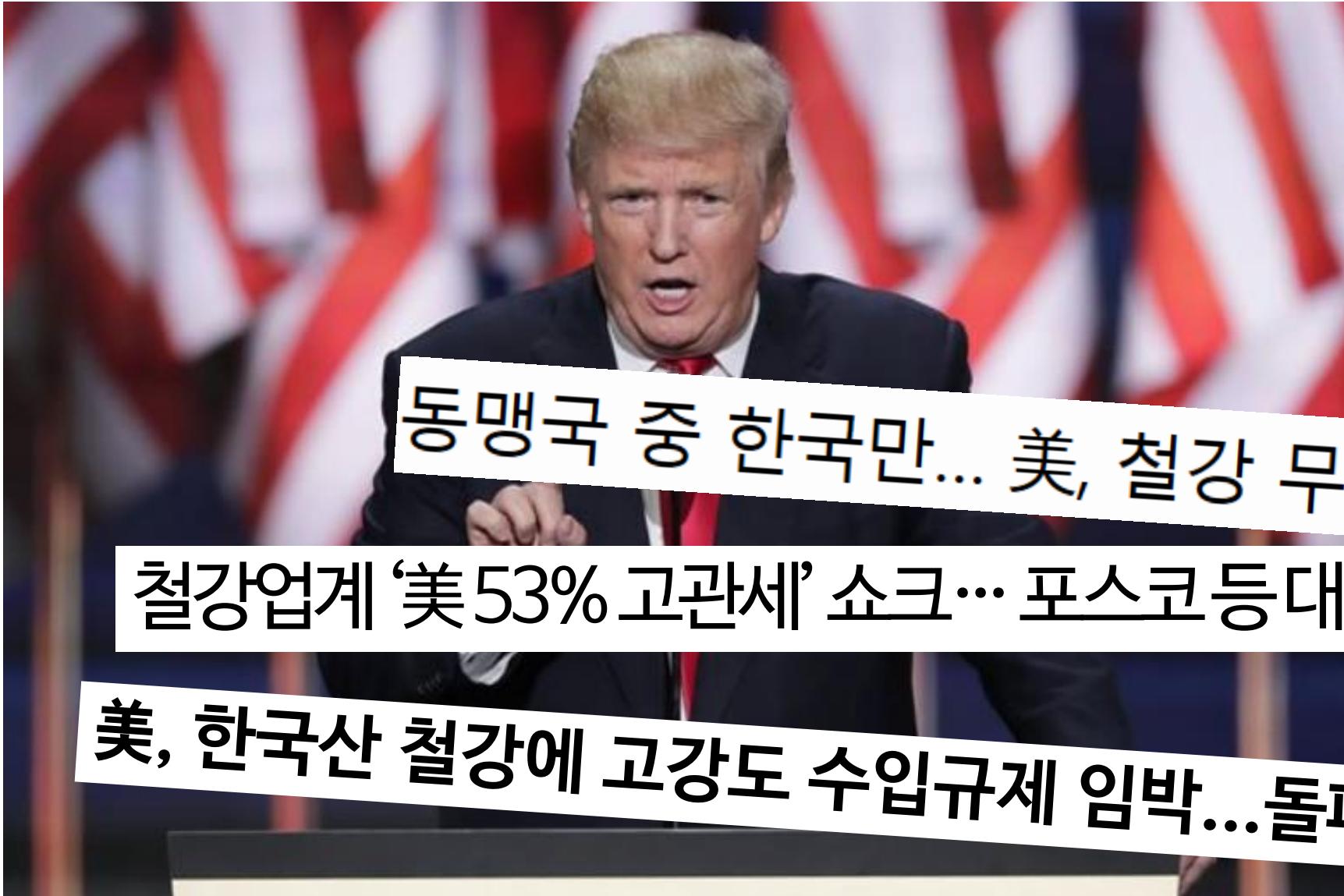
수출입무역통계
1800만 건



- 데이터셋 및 전처리 과정
- 시나리오 소개
- 데이터 시각화



WHAT IS
THREAT



동맹국 중 한국만... 美, 철강 무역제재

철강업계 '美 53% 고관세' 쇼크... 포스코 등 대미 수출 비상

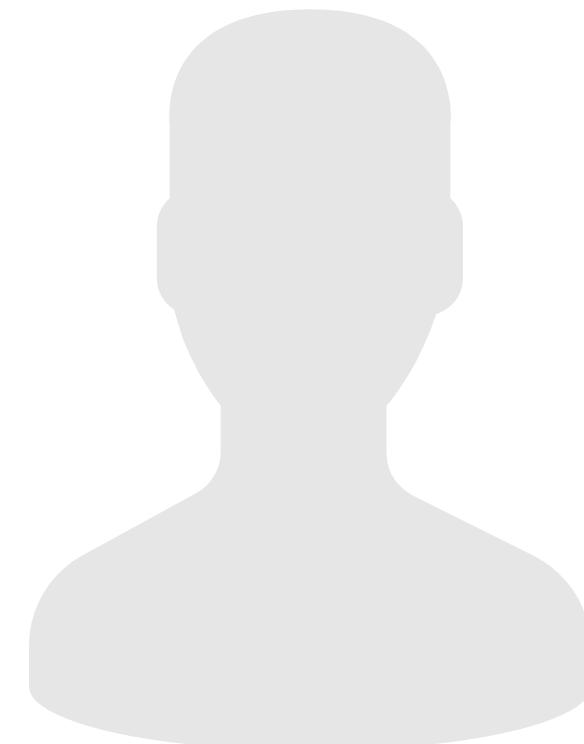
美, 한국산 철강에 고강도 수입규제 임박... 돌파구 있나?

어떻게 하면 데이터를
쉽고 효과적으로 ”
전달할 수 있을까?

산업장관·철강CEO 긴급 회의…“美 232조 피해 최소화 노력”(종합)

송고시간 | 2018/02/17 19:17

f t , ... | ← + →



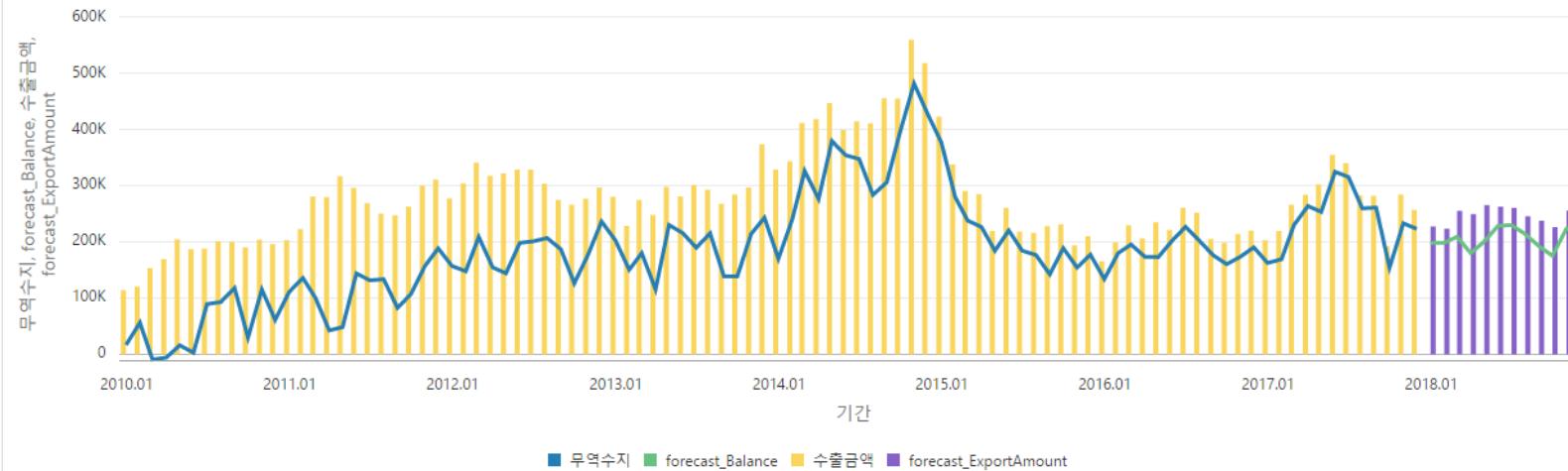
산업통상자원부 공무원

- 데이터셋 및 전처리 과정
- 시나리오 소개
- 데이터 시각화

미국 철강 시장 파악

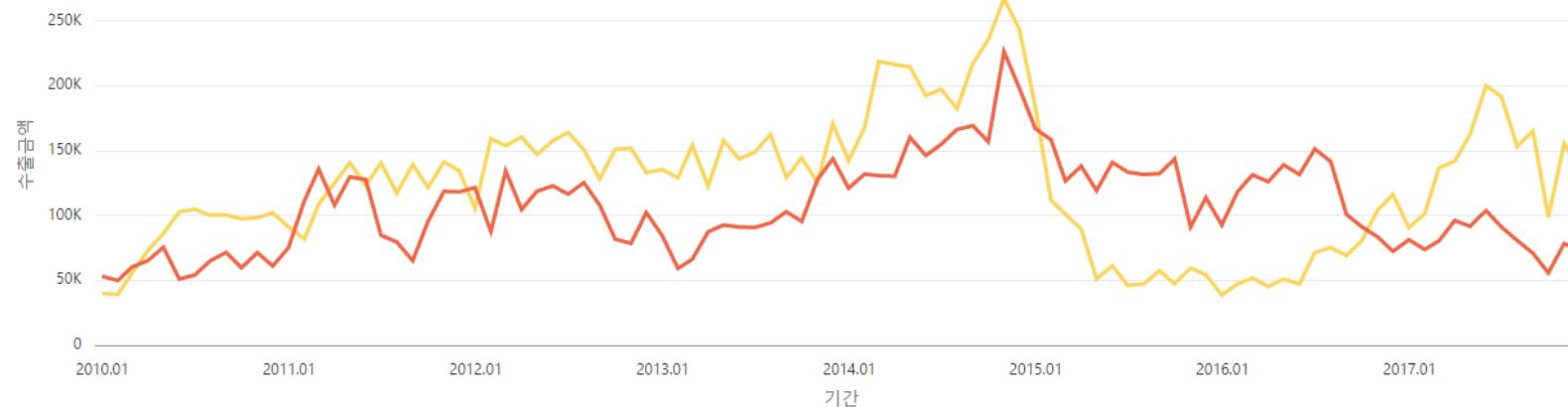
X

기간별 무역수지, forecast_Balance, 수출금액, forecast_ExportAmount



기간, 세분류명별 수출금액

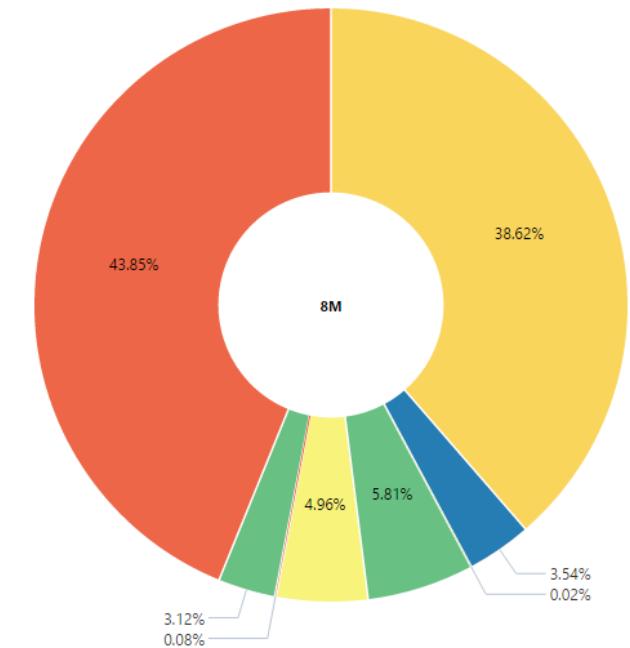
세분류명: -강관, -철강 판매



국가: 미국 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05... +91 중분류명: 라.철강 및 금속제품 소분류명: 1)철강 세분류명: 모두

세분류명별 수출금액

최상위 10 수출금액 기간: 2015.01, 2015.02, 2015.03, 2015.04, 2015.05... +25

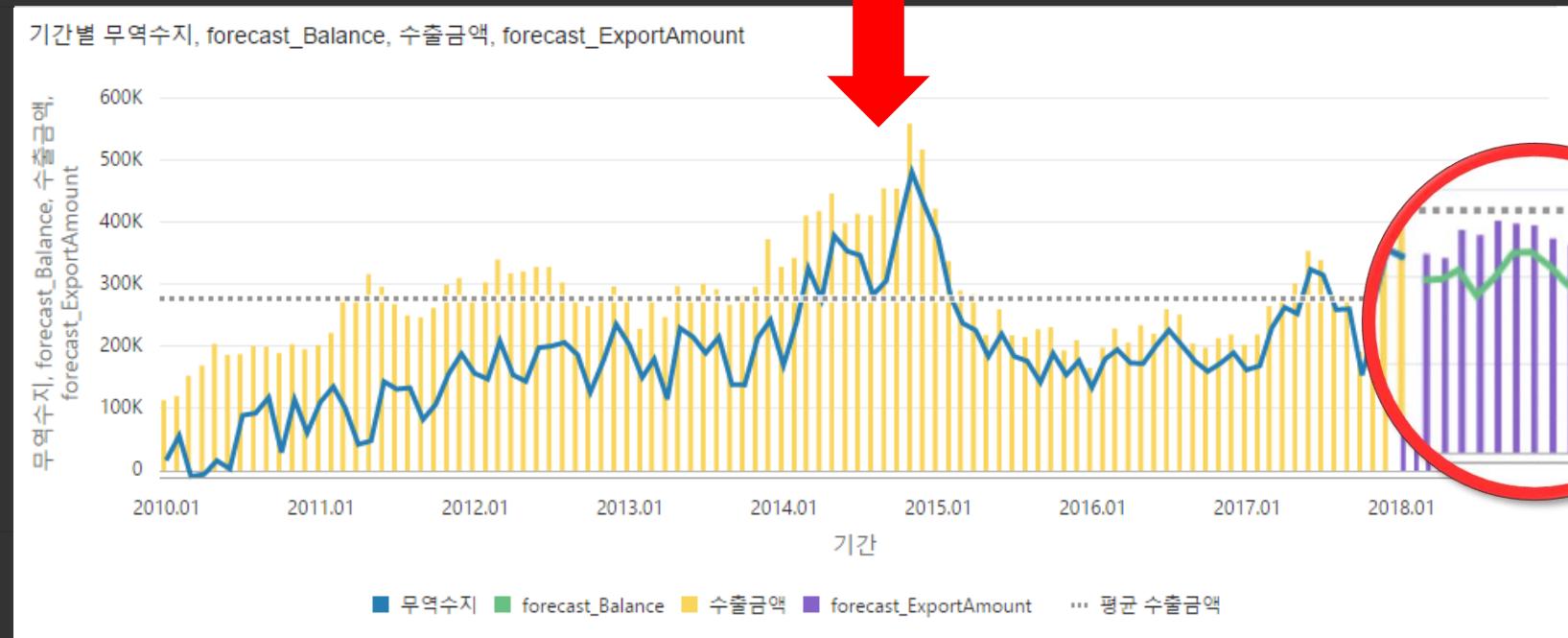


세분류명

-강관	-고철	-주.단강	-철강 원부자재
-강선	-봉.형강	-철강 반제품	-철강 판매

수출 금액 감소 기점

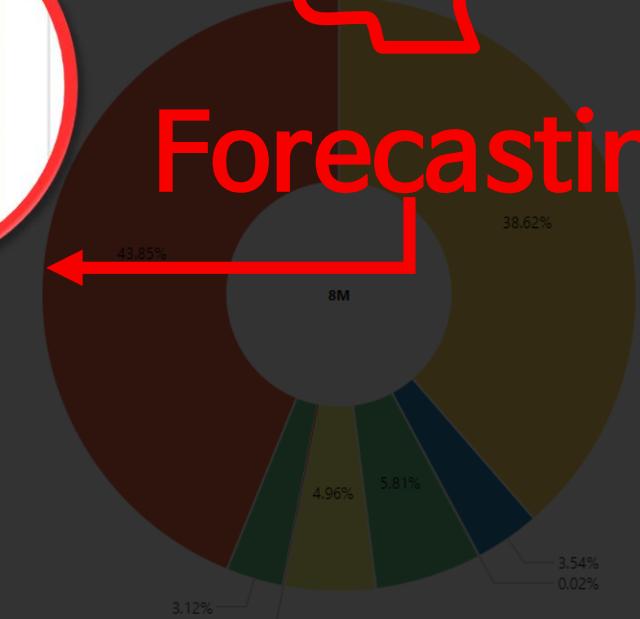
미국 철강 시장 파악



세분류별 수출금액

최상위 10 수출금액 기간: 2015.01, 2015.02, 2015.03...

Forecasting



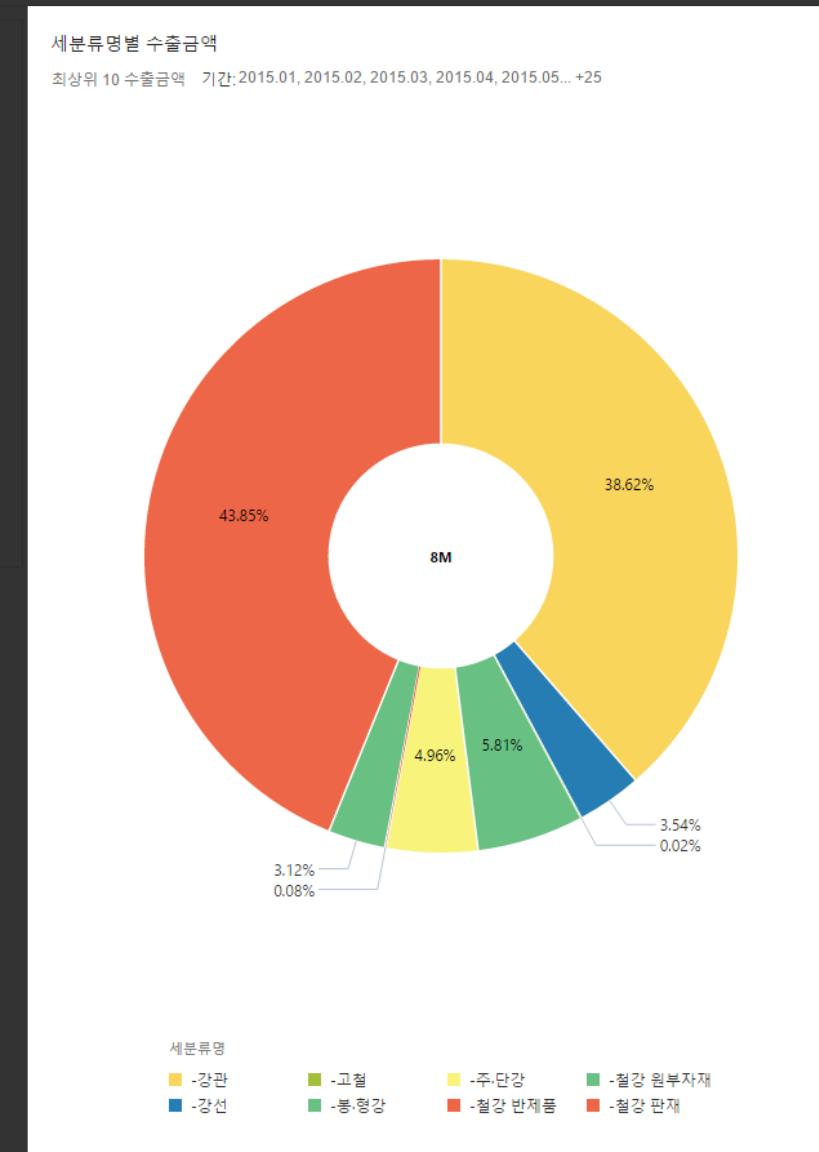
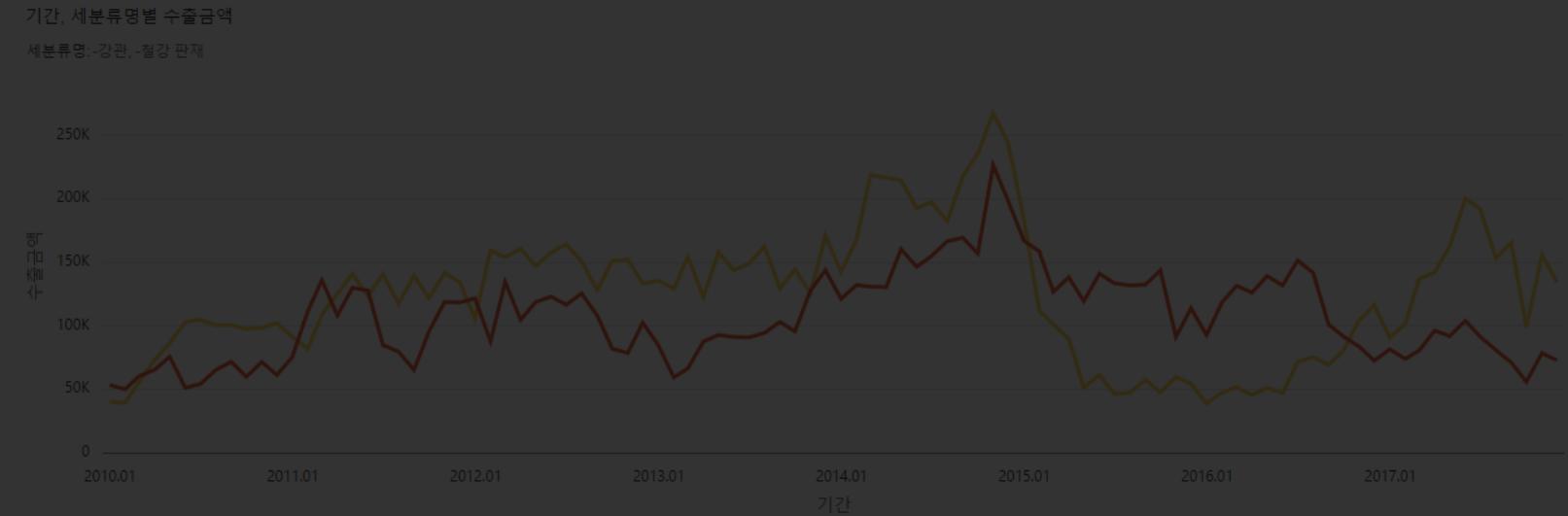
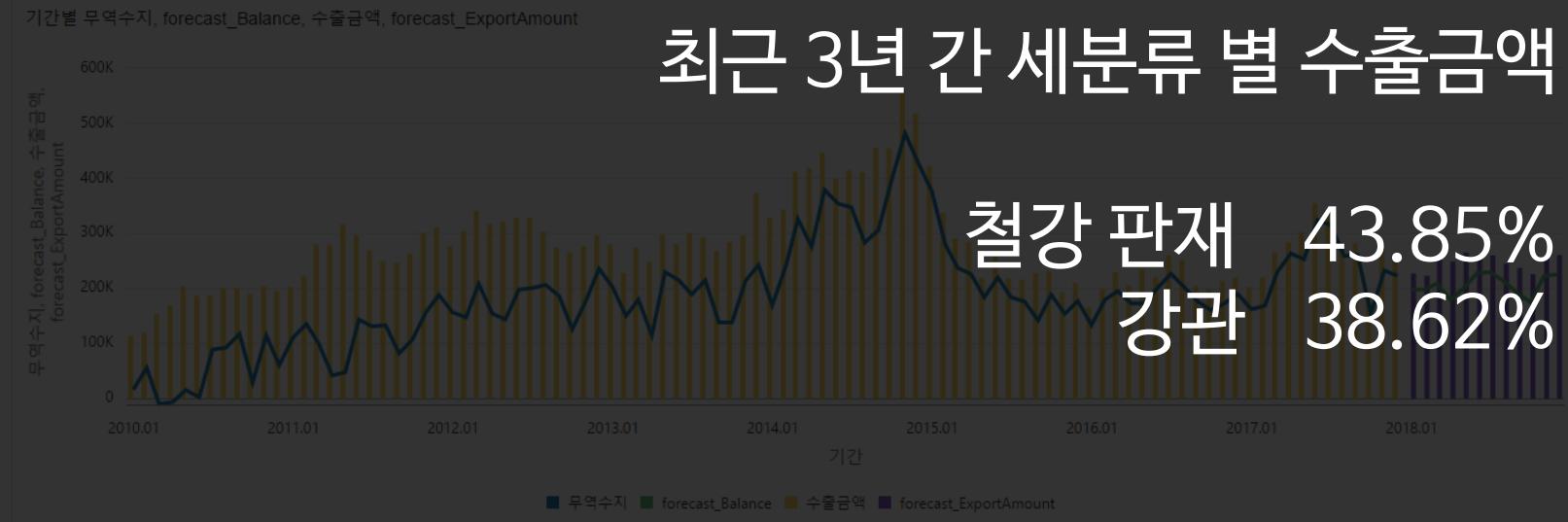
미국 철강 수출 금액 및 무역 수지

2015년을 기점으로 對한국 철강 수입 규제 경향

국내-미국 기간: 2010.01~2010.02, 2010.03~2010.04, 2010.05...+91 중분류명: 라.철강 및 금속제품 소분류명: 1)철강 세분류명: 모든

ORACLE®

최근 3년 간 세분류 별 수출금액

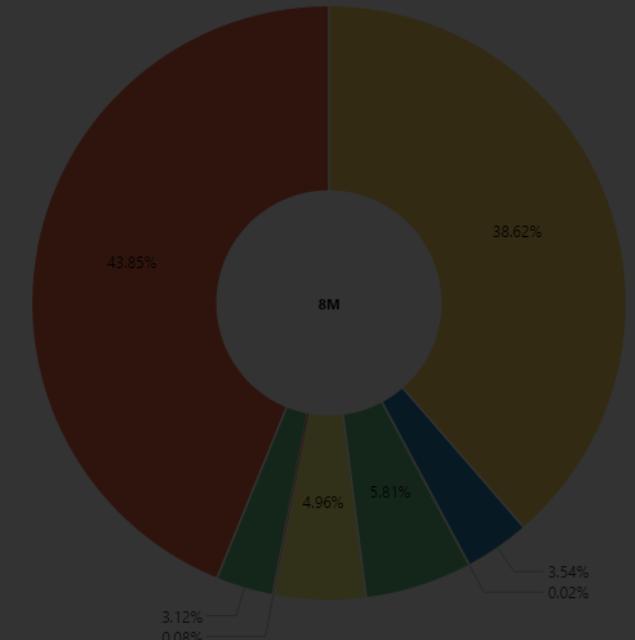


기간별 무역수지, forecast_Balance, 수출금액, forecast_ExportAmount

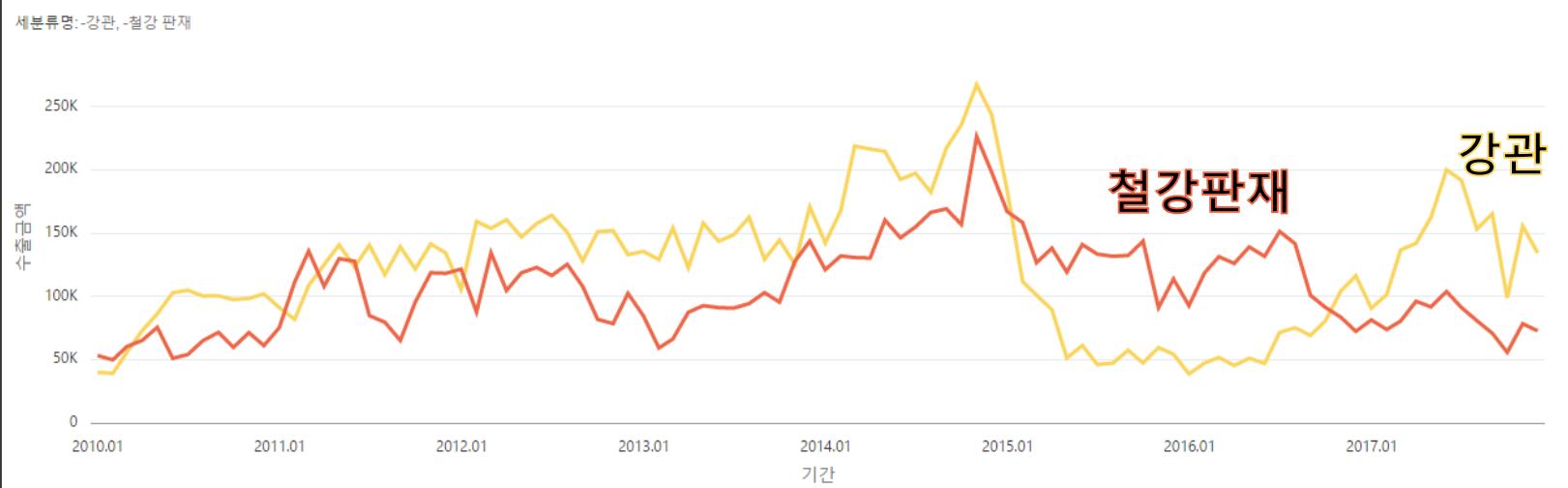


세분류명별 수출금액

최상위 10 수출금액 기간: 2015.01, 2015.02, 2015.03, 2015.04, 2015.05...+25



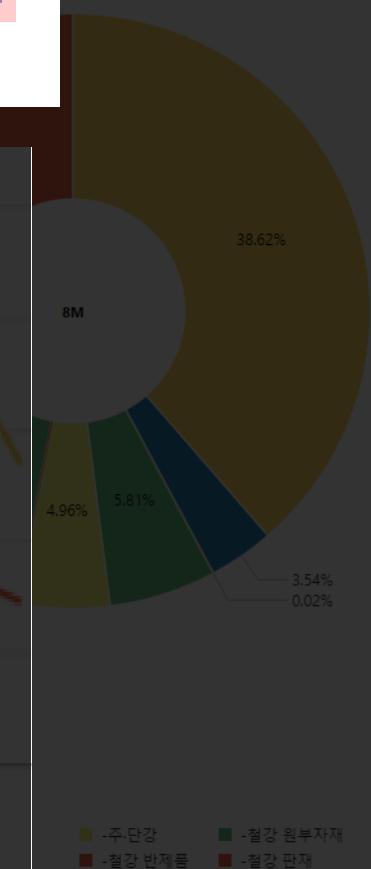
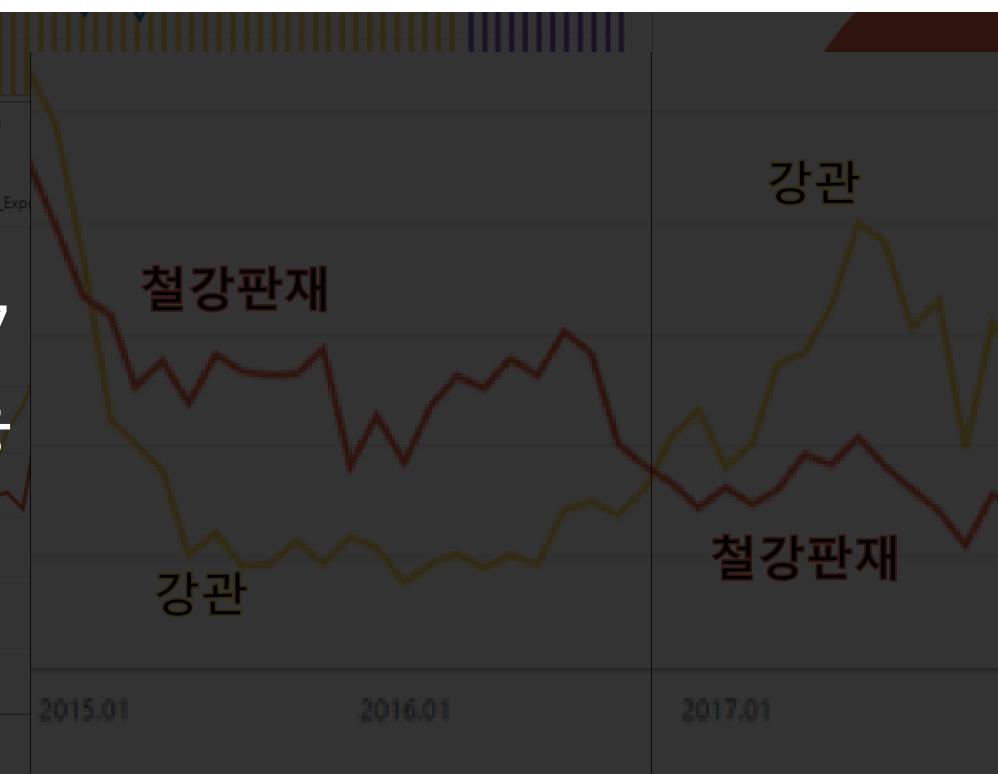
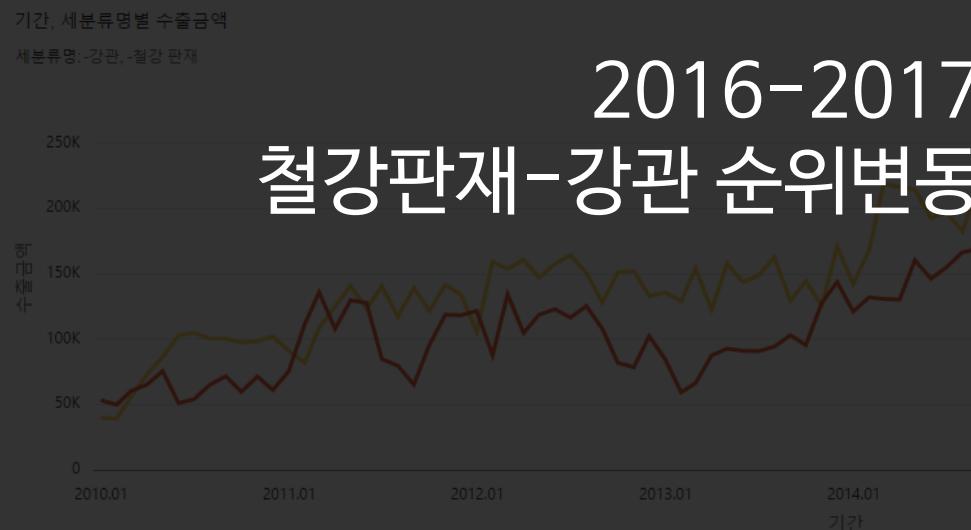
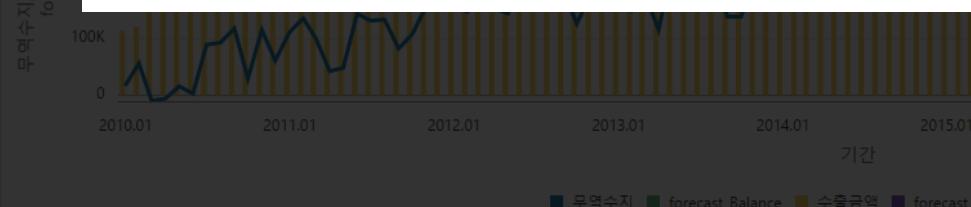
기간별 세분류명별 수출금액



국가: 미국 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05...+91 중분류명: 라.철강 및 금속제품 소분류명: 1)철강 세분류명: 모두

용정보조항(AFA)을 꺼내 들었죠. AFA는 미 정부가 요구하는 자료를 최선을 다해 제출하지 않을 경우 정부 보조금을 받은 기업에나 매길 수 있던 상계관세를 자의적으로 부과할 수 있게 한 조치입니다. 상무부는 평소 같으면 그냥 넘어가던 문제를 집요하게 파고든 뒤, 지난 2016년 9월 포스코의 자료 제출이 미비했다며 열연강판에 57%라는 고율의 상계관세를 부과했습니다.

2015.04, 2015.05...+25





산업통상자원부

“ 트럼프 정부의 승인 전까지 ”
미국을 최대한 설득하겠다





해외 신흥 철강 시장 공략

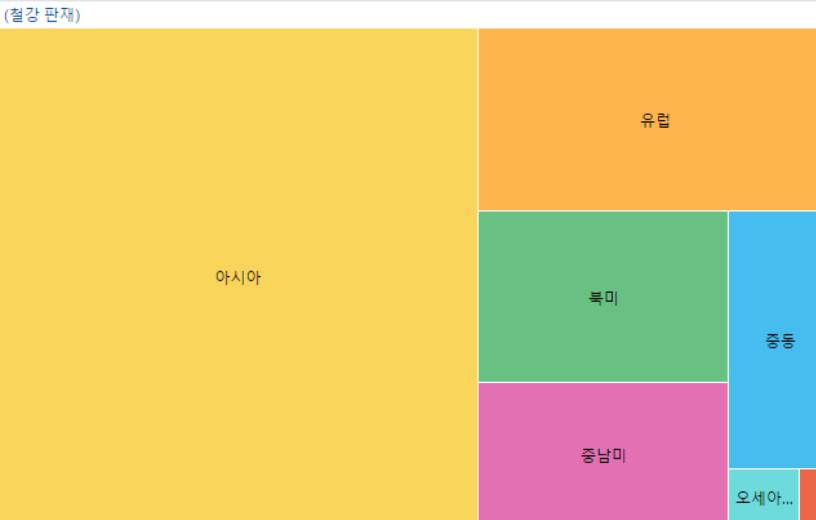
해외 철강 수출 신흥권 분석

▶ ^

세세분류명, 대류별 수출금액

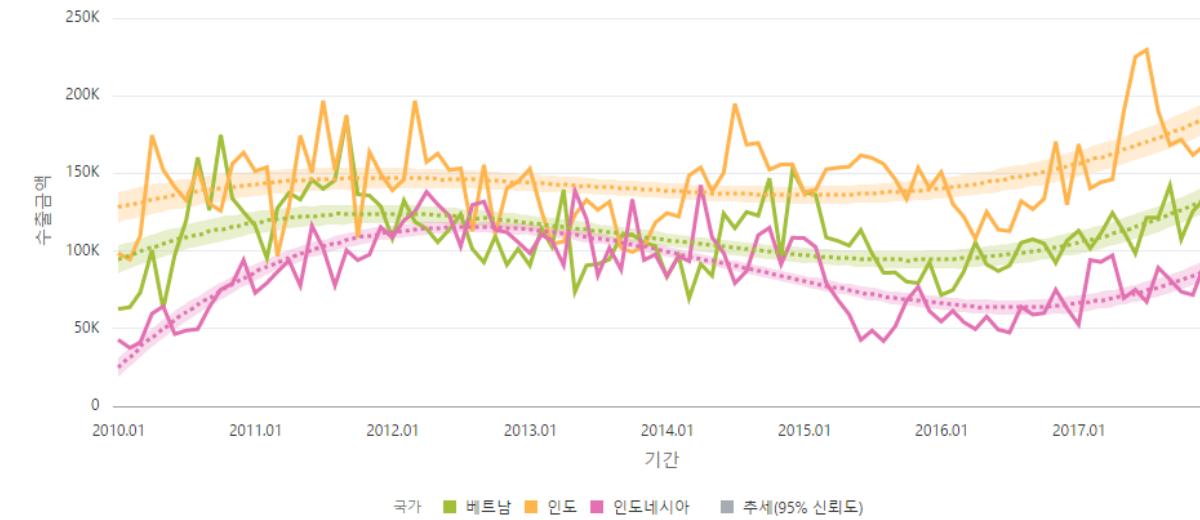
기간: 2015

(철강 판매)



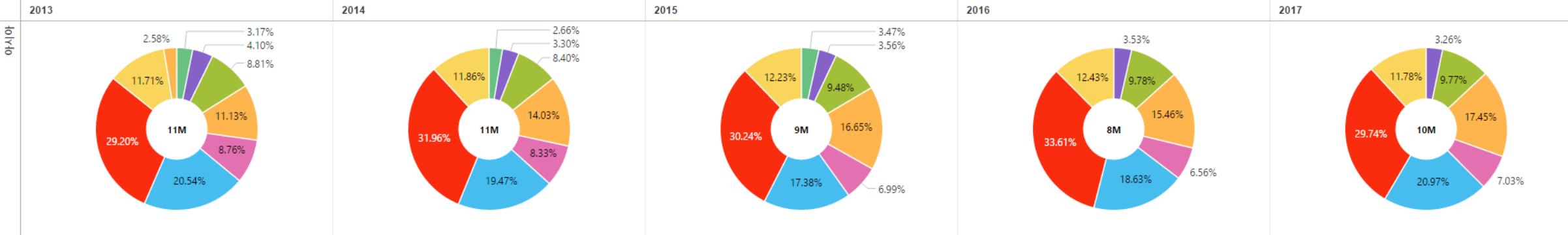
기간, 국가별 수출금액

국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05... +91



국가, 대류, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대류: 아시아 수출금액: ≥ 251,000



▼ 소분류명: 모두 중분류명: 라.철강 및 금속제품 대류: 없음

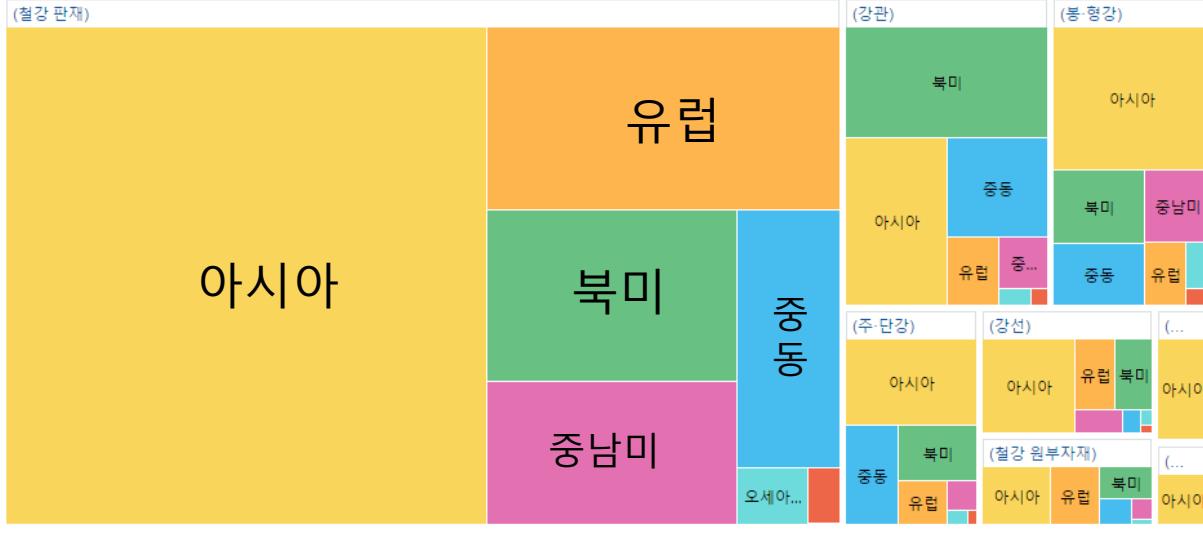
ORACLE®

신흥권 시장 분석

세세분류명, 대륙별 수출금액

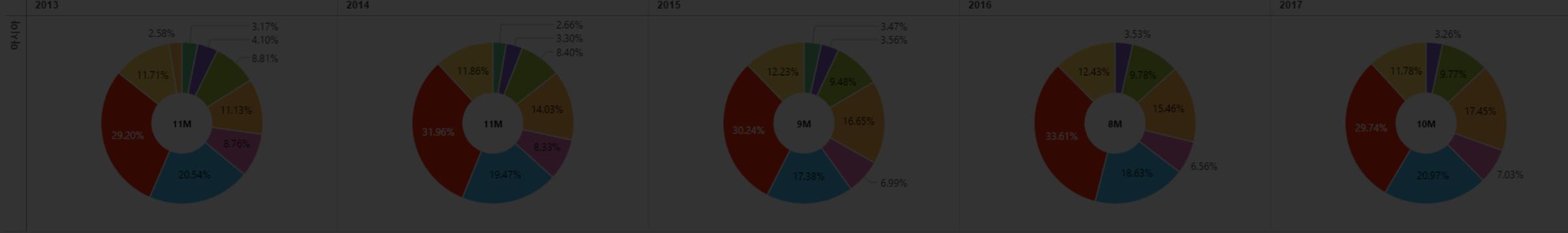
기간: 2015

(철강 판재)



국가, 대륙, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대륙: 아시아 수출금액: ≥ 251,000

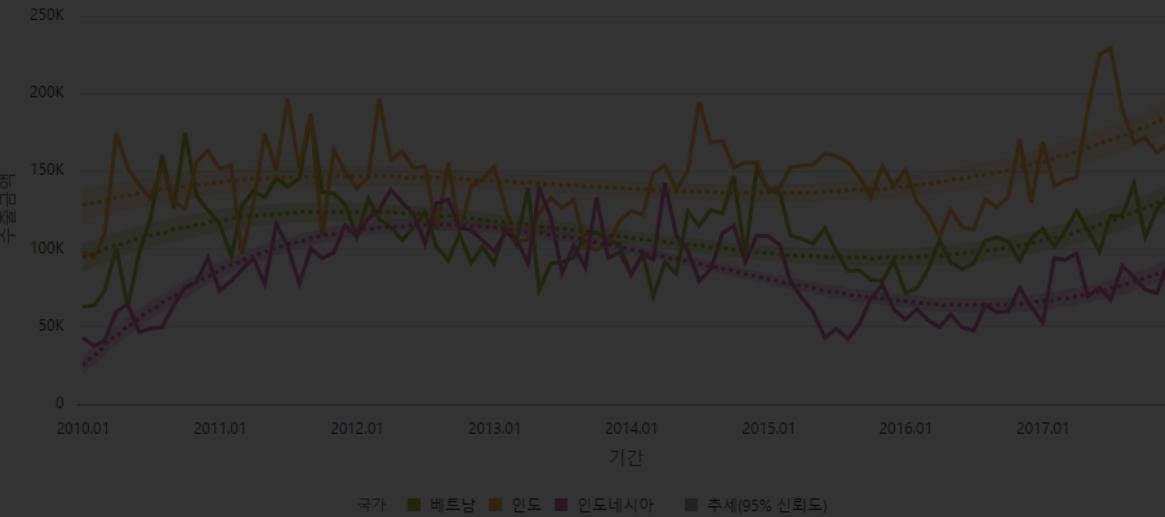


▼ 소분류명: 모두 중분류명: 라.철강 및 금속제품 대류: 없음

대륙별 철강 수출 비중

기간별 수출금액

국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05 ... +91



신흥권 시장 분석

세세분류명, 대륙별 수출금액

기간: 2015

(철강 판재)

유럽

(강판)

(봉형강)

북미

중동

아시아

중동

유럽

기간, 국가별 수출금액

국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05 ... +91

250K

200K

150K

100K

50K

0

기간

국가

베트남

인도

인도네시아

주제(95% 신뢰도)

아시아 철강 수출 상위 9개국

국가, 대륙, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대륙: 아시아 수출금액: ≥ 251,000

한국
수출
증가율

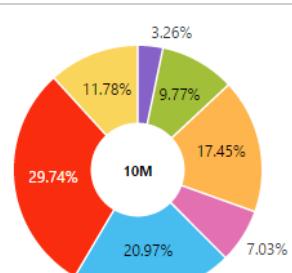
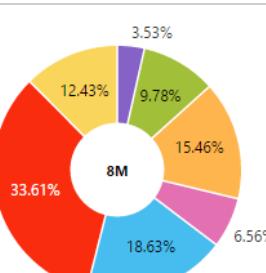
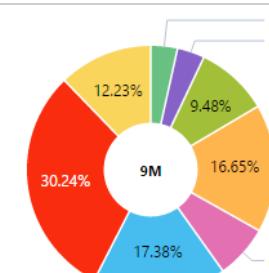
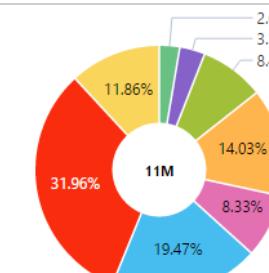
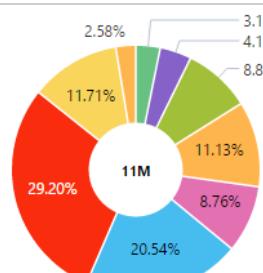
2013

2014

2015

2016

2017



국가

대만

말레이시아

베트남

인도

인도네시아

일본

중국

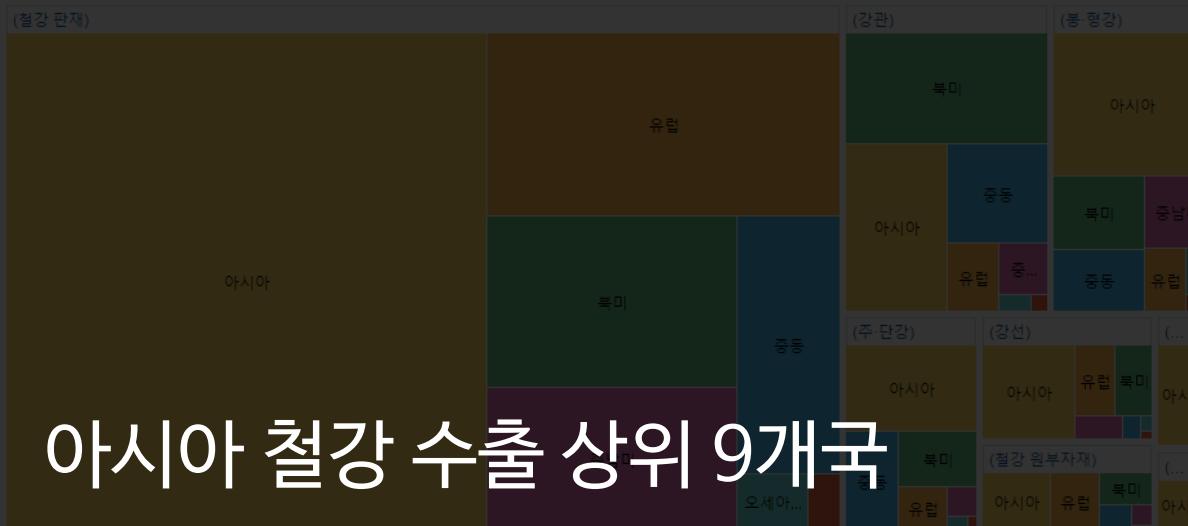
태국

필리핀

소분류명: 모두 중분류명: 라철강 및 금속제품 대륙: 아시아

ORACLE®

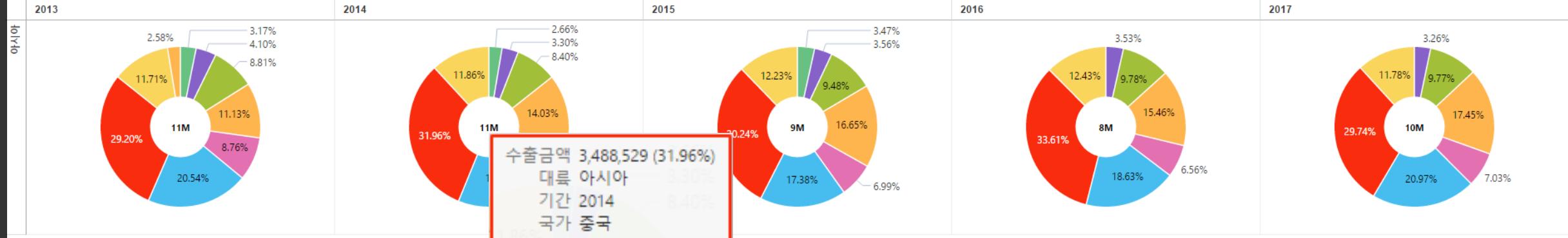
신흥권 시장 분석



아시아 철강 수출 상위 9개국

국가, 대륙, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대륙: 아시아 수출금액: ≥ 251,000



▼ 소부르면·모드 즈부르면·란 척간 미 그속제품

三

1

ANSWER

이지아 ■ ■ ■

1

인도네시아

1

2

1

1

四

1

ORACLE®

신흥권 시장 분석

세세분류명, 대륙별 수출금액

기간: 2015

(철강 판재)

유럽

(강판)

(봉형강)

북미

중동

북미

중동

아시아

유럽

아시아

북미

중동

아시아

유럽

중...

(주 단강)

(강선)

(...)

아시아

유럽

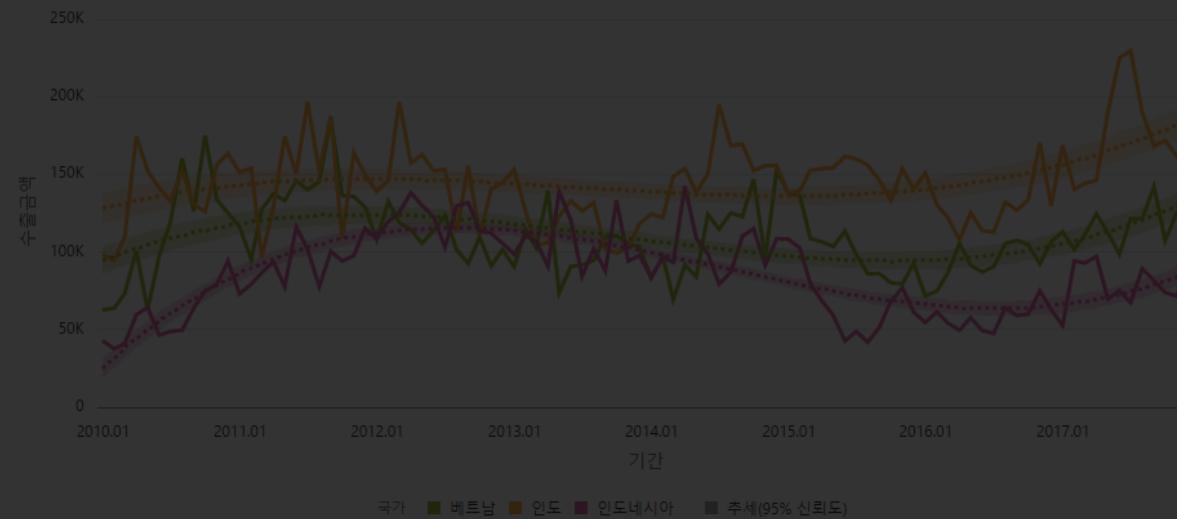
북미

아시아

아시아 철강 수출 상위 9개국

기간, 국가별 수출금액

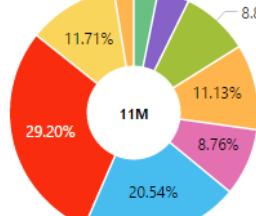
국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05 ... +91



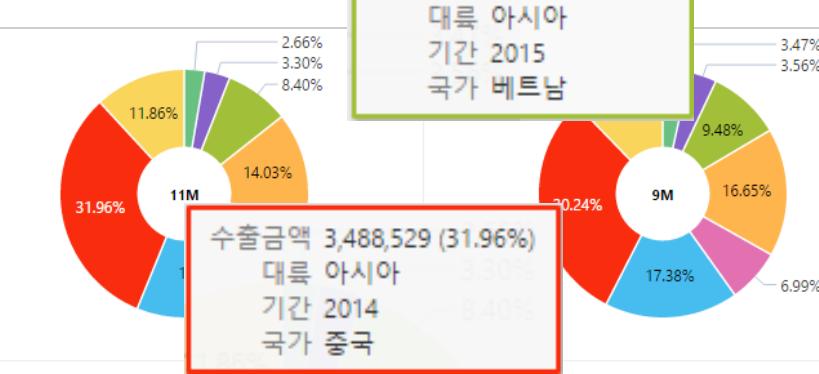
국가, 대륙, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대륙: 아시아 수출금액: ≥ 251,000

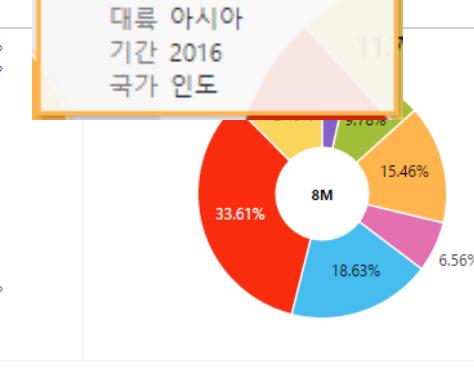
한국
중국
일본
인도
인도네시아
태국
필리핀
타이
마카오
타이완
브라질
미국
한국
중국
일본
인도
인도네시아
태국
필리핀
타이
마카오
타이완
브라질
미국



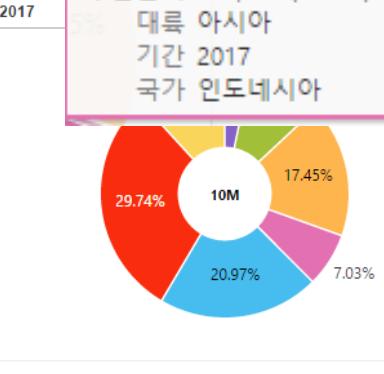
2014



2016



2017



소분류명: 모두 중분류명: 라철강 및 금속제품

국가

대만

말레이지아

베트남

인도

인도네시아

일본

중국

태국

필리핀

신흥권 시장 분석

세세분류명, 대류별 수출금액

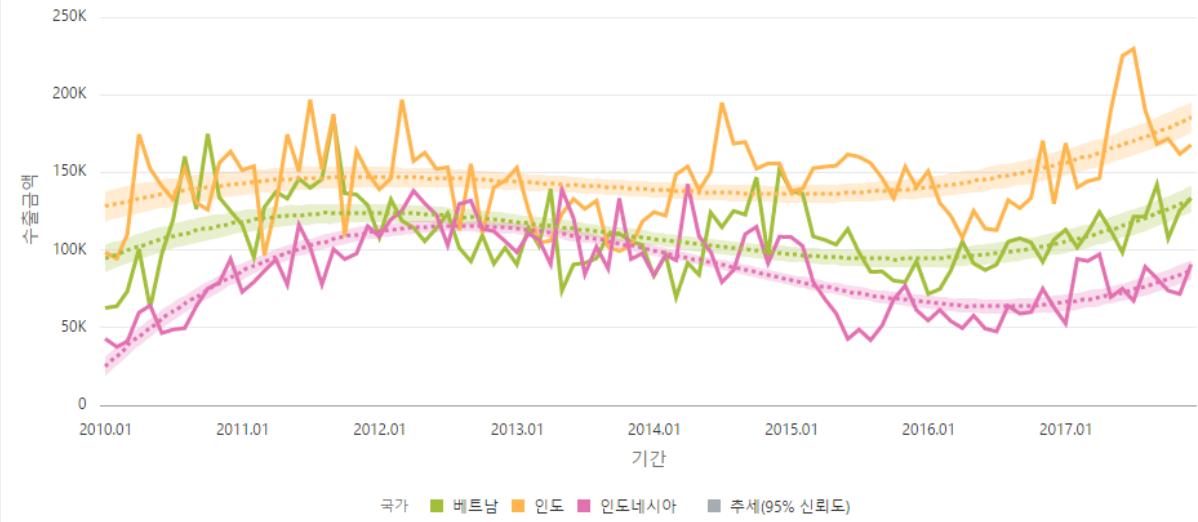
기간: 2015

(철강 판매)



기간, 국가별 수출금액

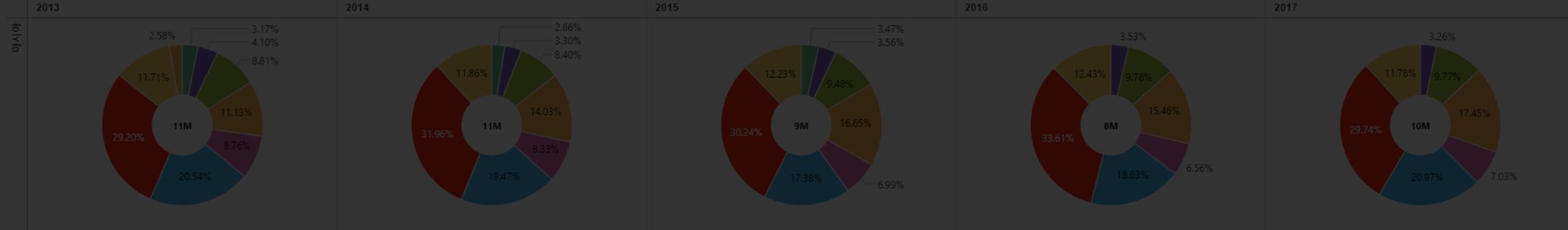
국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05... +91



국가, 대류, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대류: 아시아 수출금액: ≥ 251,000

히
어



소분류명: 모두 중분류명: 라.철강 및 금속제품 대류: 없음

신흥권 시장 분석

세세분류명: 대륙별 수출금액

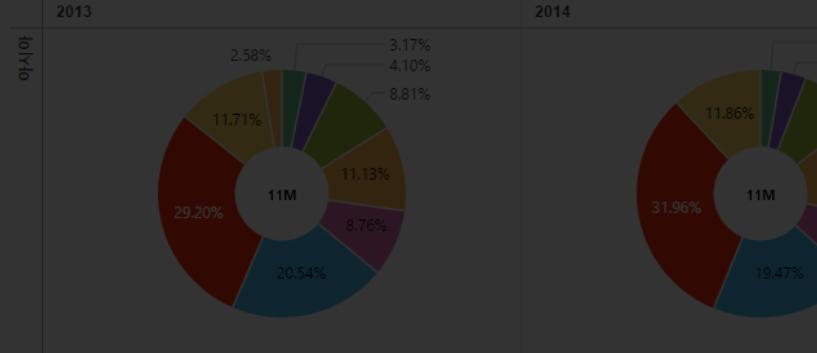
기간: 2015

(질감 판재)



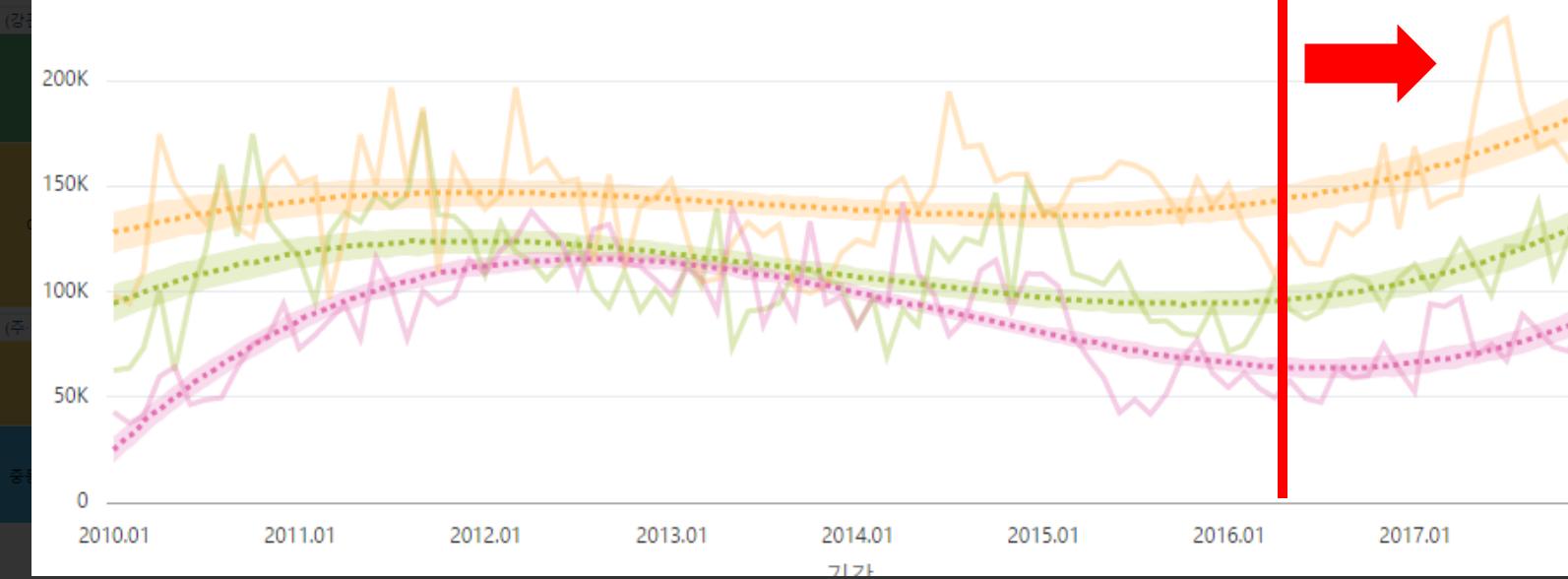
국가, 대륙, 기간별 수출금액

기간: 2013, 2014, 2015, 2016, 2017 대륙: 아시아 수출금액: ≥ 251,000



기간, 국가별 수출금액

국가: 베트남, 인도, 인도네시아 기간: 2010.01, 2010.02, 2010.03, 2010.04, 2010.05...+91



인도, 베트남, 인도네시아 모두 상승 추세

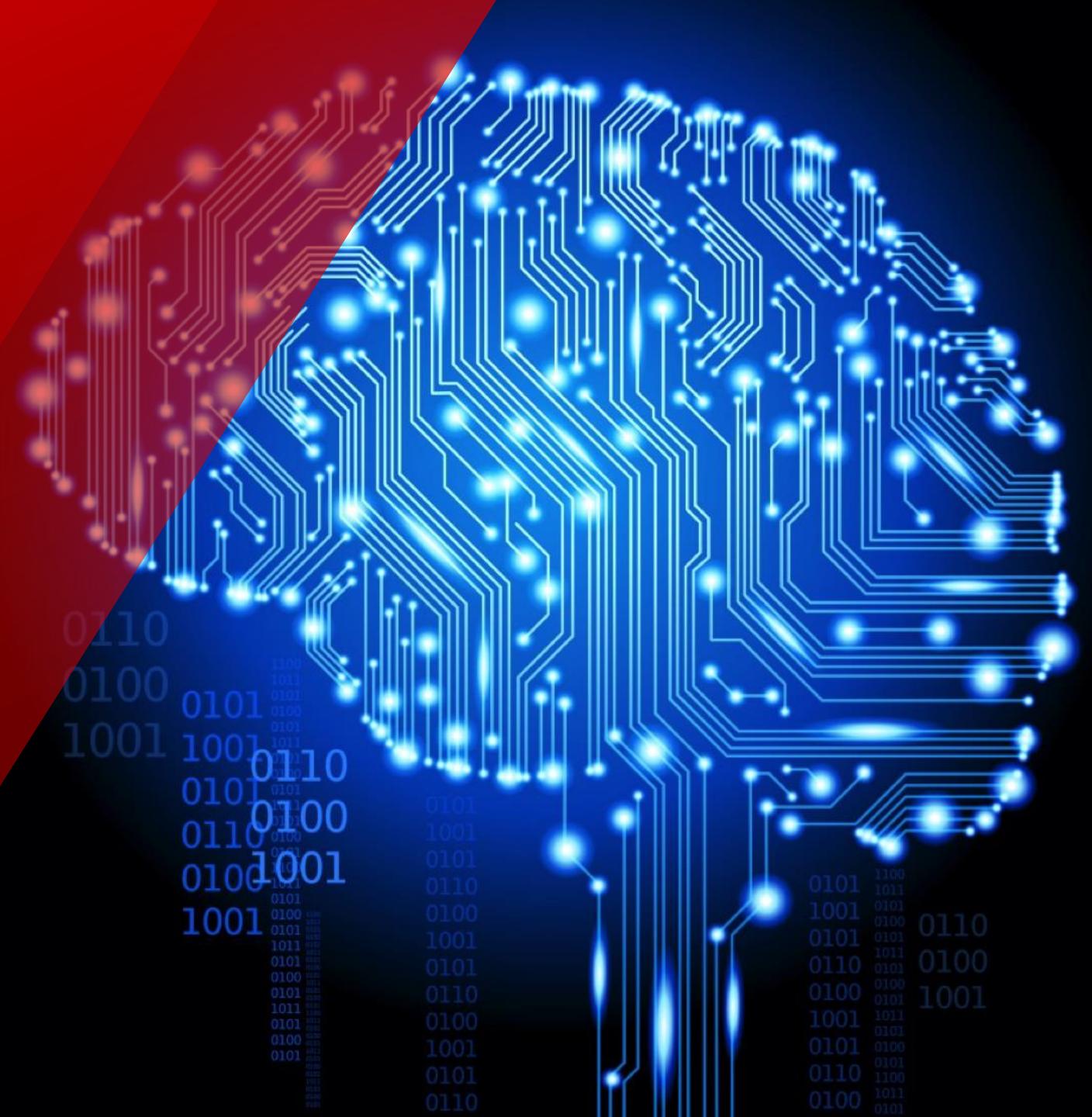
국가 대만 말레이지아 베트남 인도 인도네시아 일본 중국 태국 필리핀

소분류명: 모두 중분류명: 라.철강 및 금속제품 대류: 없음

Conclusion

- OAC와 BDC: 알아보기 어려운 빅데이터 정제 및 가공
- 인사이트를 도출할 수 있는 시각화된 자료 제공
- 다양한 필드의 유저의 데이터 분석을 도와 가치를 찾는 속도를 높여줌

머신러닝을 활용한 영화 추천 모델링



Agenda

- 데이터셋 소개
- 데이터 분석 도구 및 알고리즘
- 데이터 분석 절차 (주요 이슈 및 해결 과정)
- 영화 추천 모델 및 평가 방법 소개
- 영화 추천 모델별 성능 비교

데이터셋 소개



A collage of various movie and television show posters, including X-MEN 3, Battlestar Galactica, The Simpsons, Saturday Night Live, The Office, Prison Break, Heroes, Psych, and Touch. The Oracle MoviePlex logo is overlaid on the bottom left of the collage.

Oracle Big Data Lite
Virtual Machine

Getting Started



Click here for
Oracle MoviePlex

@Oracle Technology Network

데이터셋 소개

Moviedemo 데이터셋을 활용한
영화 추천 머신러닝 모델 개발



데이터 분석 도구



데이터 분석 도구 데이터 전처리

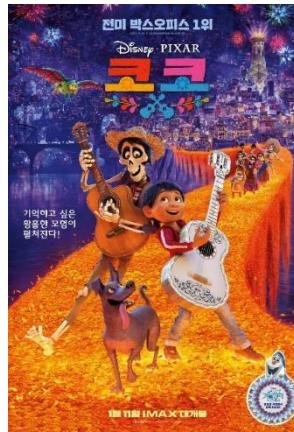


데이터 분석 도구



데이터 분석 알고리즘

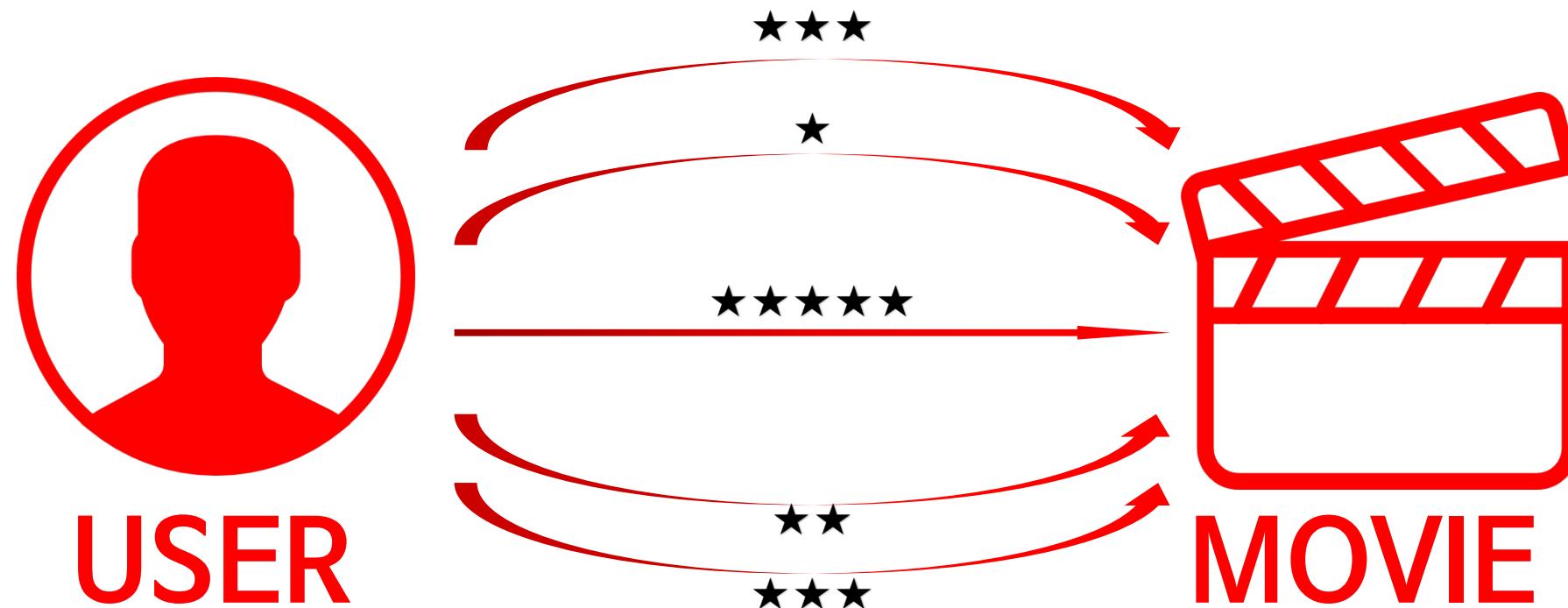
비슷한 성향



USER1	5	1	3	4	3
USER2	2	4	5	2	3
USER3	5	2	3	4	4

협업 필터링 (Collaborative Filtering, CF)

데이터 전처리 - 1. 중복 데이터 제거

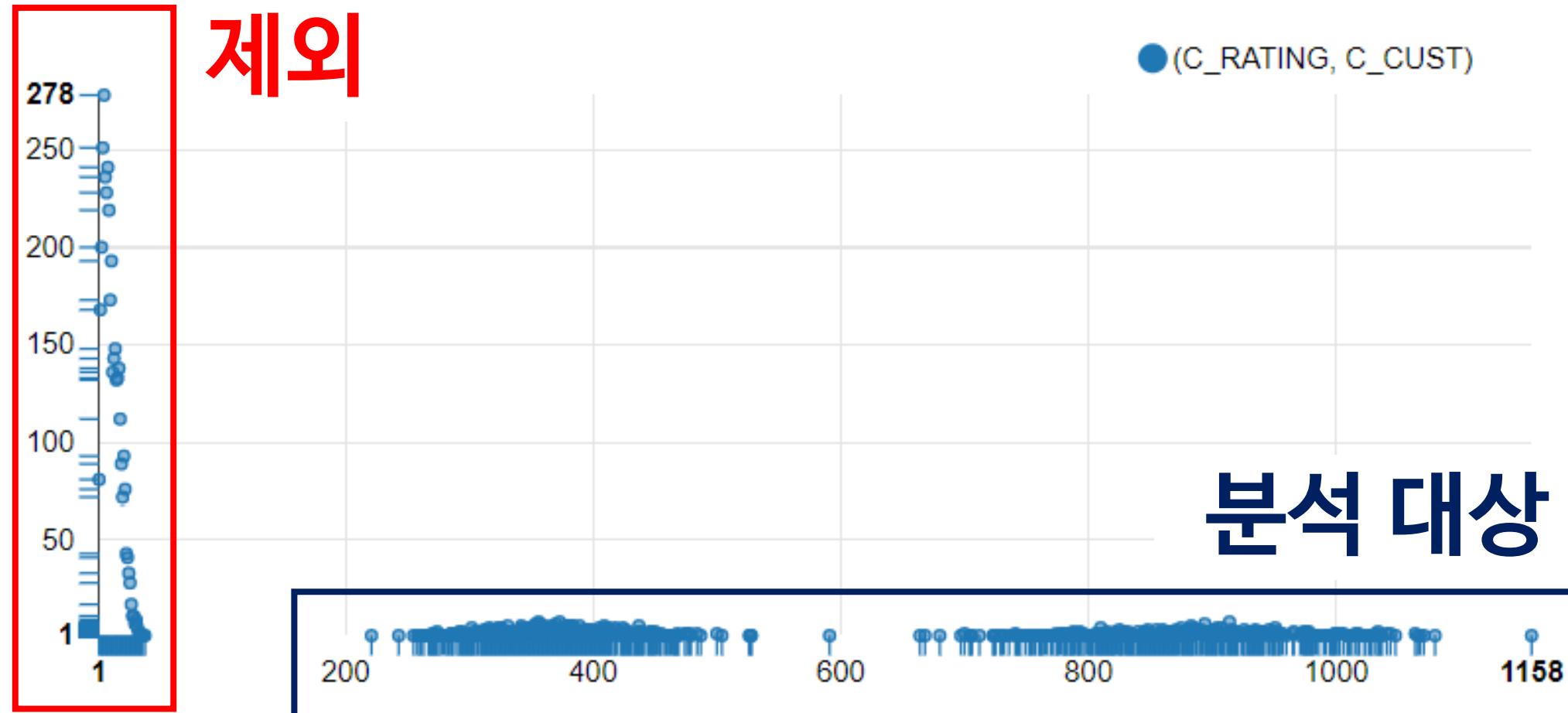


데이터 전처리 - 2. NULL 값 처리



기본값
(3점)

데이터 전처리 – 3. 최소 RATING 건수 필터링



영화 추천 모델의 성능 개선

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF)

1. (유저별 모든 영화) 평점 예상

2. (유저의 취향 반영) 영화 추천

영화 추천 모델의 성능 개선

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF)

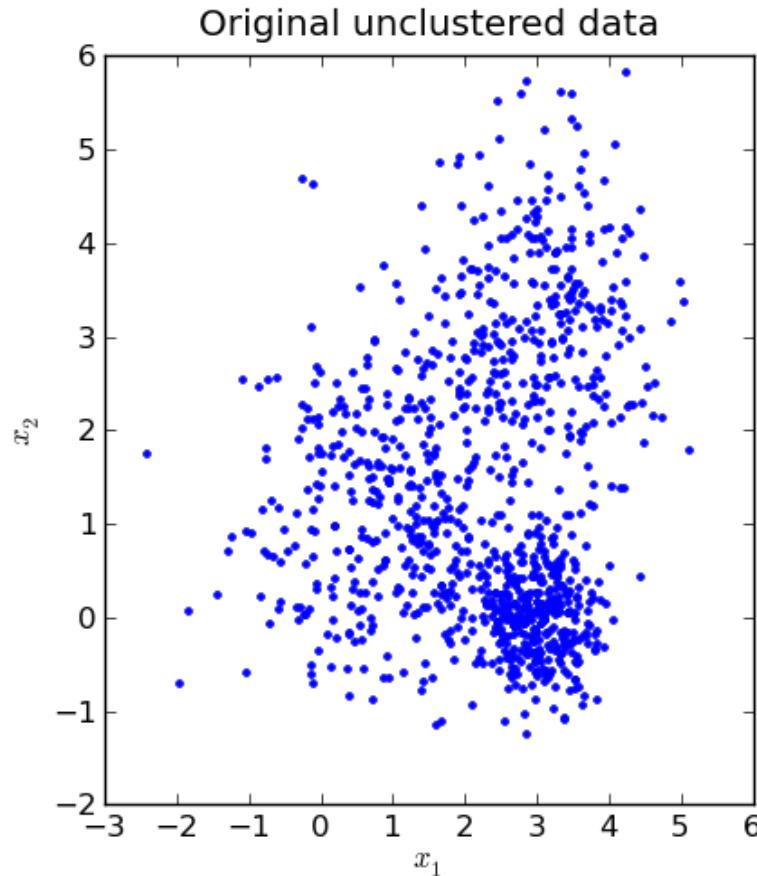
유저 클러스터링(UC)



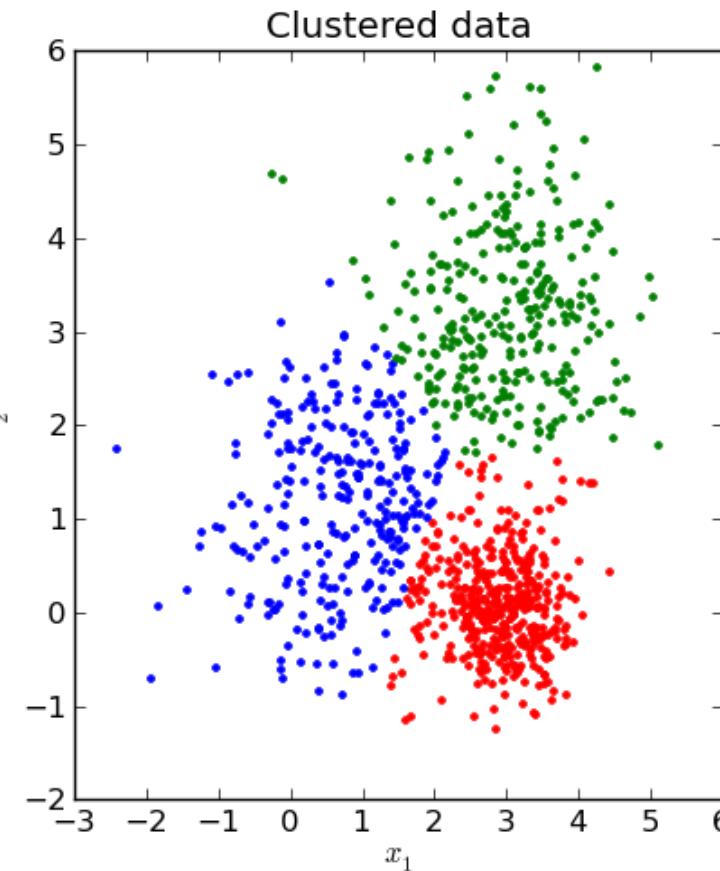
영화 클러스터링(MC)



데이터 분석 알고리즘



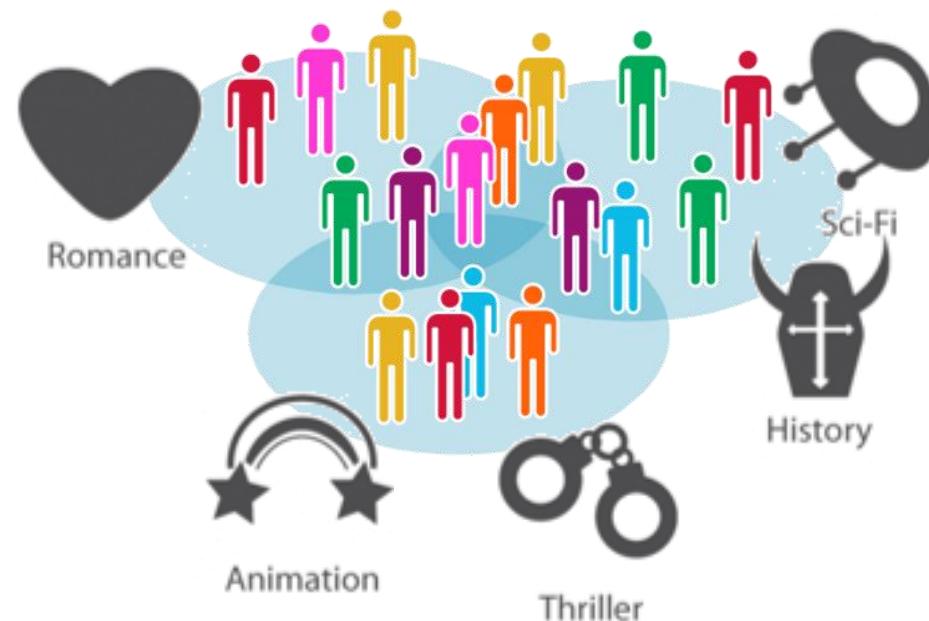
클러스터링



K-means

유저 및 영화 클러스터링

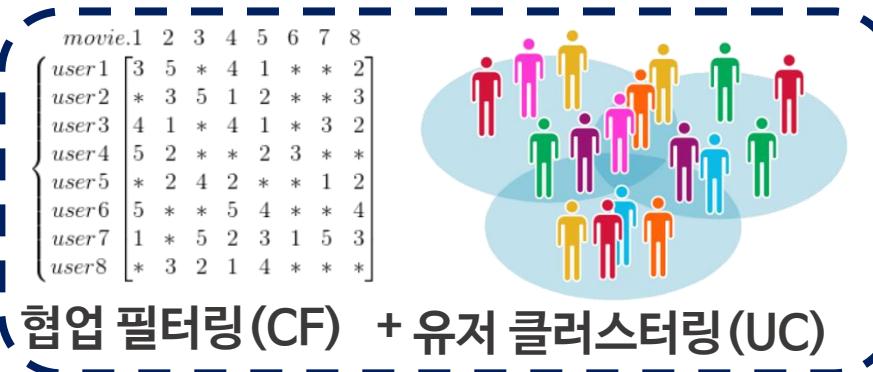
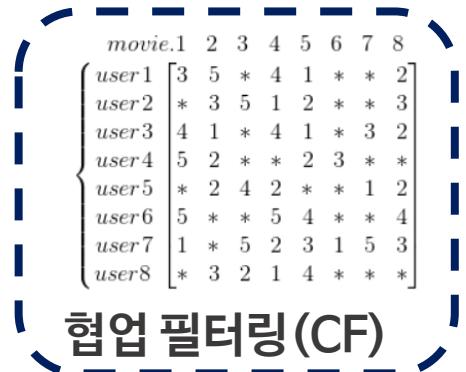
유저 클러스터링(UC)



영화 클러스터링(MC)



영화 추천 모델 및 평가 방법 소개



모델 평가

영화 10편씩 추천

추천 점수

- ✓ 부정(-1) / 긍정(+1)
- ✓ (-10점) ~ (+10점)

추천 적중률

- ✓ 로그 기록 0 → 적중
- ✓ 0% ~ 100%

추천 적합도

- ✓ 추천 점수 × 추천 적중률
- ✓ (정규화) 0점 ~ 100점

영화 추천 모델별 성능 비교

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF)

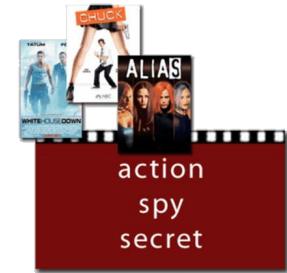
	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF) + 유저 클러스터링(UC)



	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*



추천 점수

5.26점

6.39점

5.87점

추천 적중률

74%

85%

82%

추천 적합도

69점

77점(+8점)

74점(+5점)

영화 추천 모델별 성능 비교

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF)

	movie.1	2	3	4	5	6	7	8
user1	3	5	*	4	1	*	*	2
user2	*	3	5	1	2	*	*	3
user3	4	1	*	4	1	*	3	2
user4	5	2	*	*	2	3	*	*
user5	*	2	4	2	*	*	1	2
user6	5	*	*	5	4	*	*	4
user7	1	*	5	2	3	1	5	3
user8	*	3	2	1	4	*	*	*

협업 필터링(CF) + 유저 클러스터링(UC)



5.26점

6.39점

5.87점

추천 점수

74%

85%

82%

추천 적중률

69점

77점(+8점)

74점(+5점)

추천 적합도

Conclusion

- 다양한 ML 기법 활용한 추천 시스템 개선
- 빅데이터를 활용한 인사이트 제공
- Oracle Cloud를 활용한 데이터 분석의 A to Z



Q&A



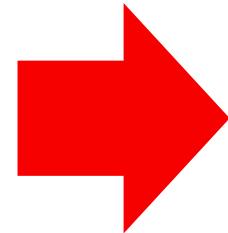
ORACLE®

[부록1]

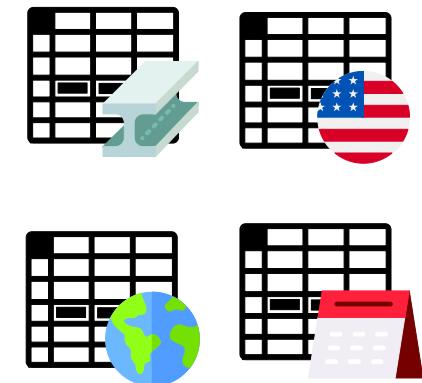
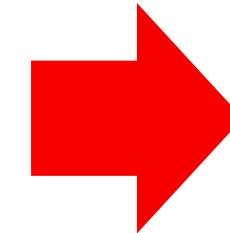
Big Data Cloud System 활용



관세청 수출입무역통계 데이터



전처리



전처리 된 데이터

...

[부록1]



STEEL_EXPORT



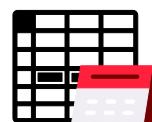
STEEL_BALANCE



COUNTRY_USA



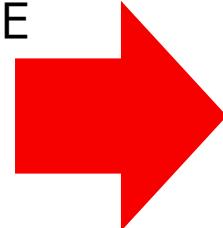
CONTINENTS



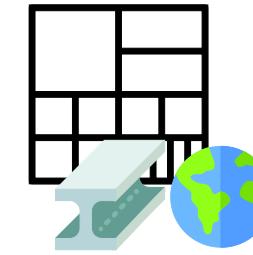
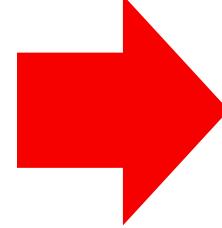
MONTH

...

전처리 된 데이터



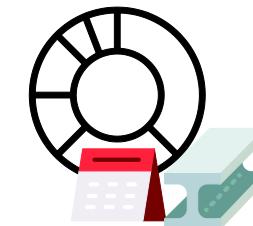
Oracle Analytics Cloud 활용
Data 시각화



STEEL_BALANCE
+
CONTINENTS



STEEL_BALANCE
+
CONTINENTS
+
COUNTRY_USA

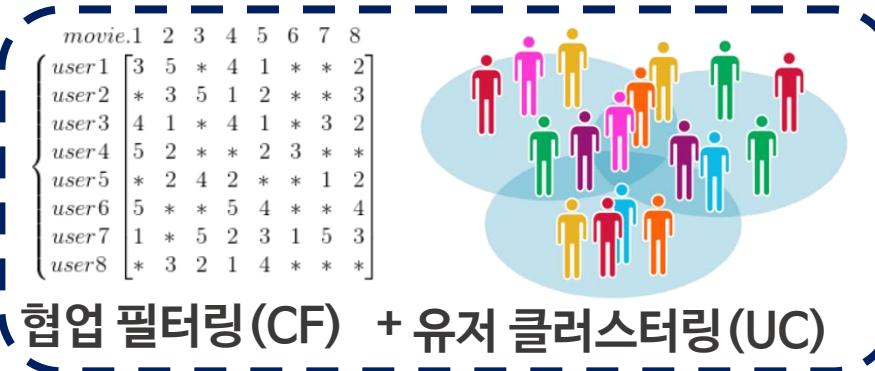
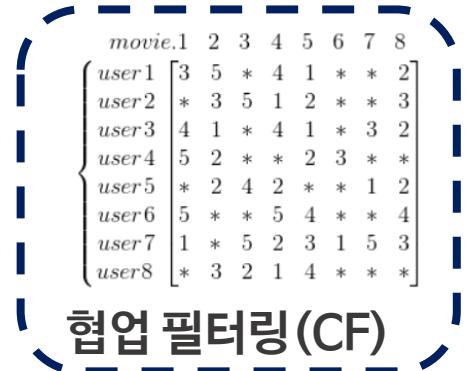


MONTH
+
STEEL_EXPORT

...

시각화 데이터

[부록2] 영화 추천 모델 및 평가 방법 소개



모델 평가

영화 10개씩 추천

추천 점수

- ✓ 부정(-1) / 긍정(+1)
- ✓ (-10점) ~ (+10점)

추천 적중률

- ✓ 로그 기록 0 → 적중
- ✓ 0% ~ 100%

추천 적합도

- ✓ 추천 점수 X 추천 적중률
- ✓ (정규화) 0점 ~ 100점