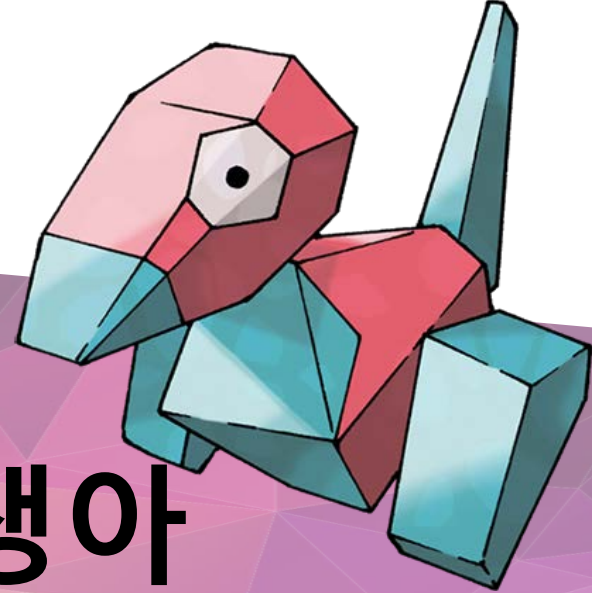


포괄수가제 및 신생아 데이터 분석



2018. 11. 23.

김태형

목 차

1. 건강보험심사평가원 빅데이터 분석 개요
2. 포괄수가제 데이터 분석
3. 신생아 데이터 분석
4. Q&A

건강보험심사평가원에서 제공하는 환자들의 명세서 데이터를 활용

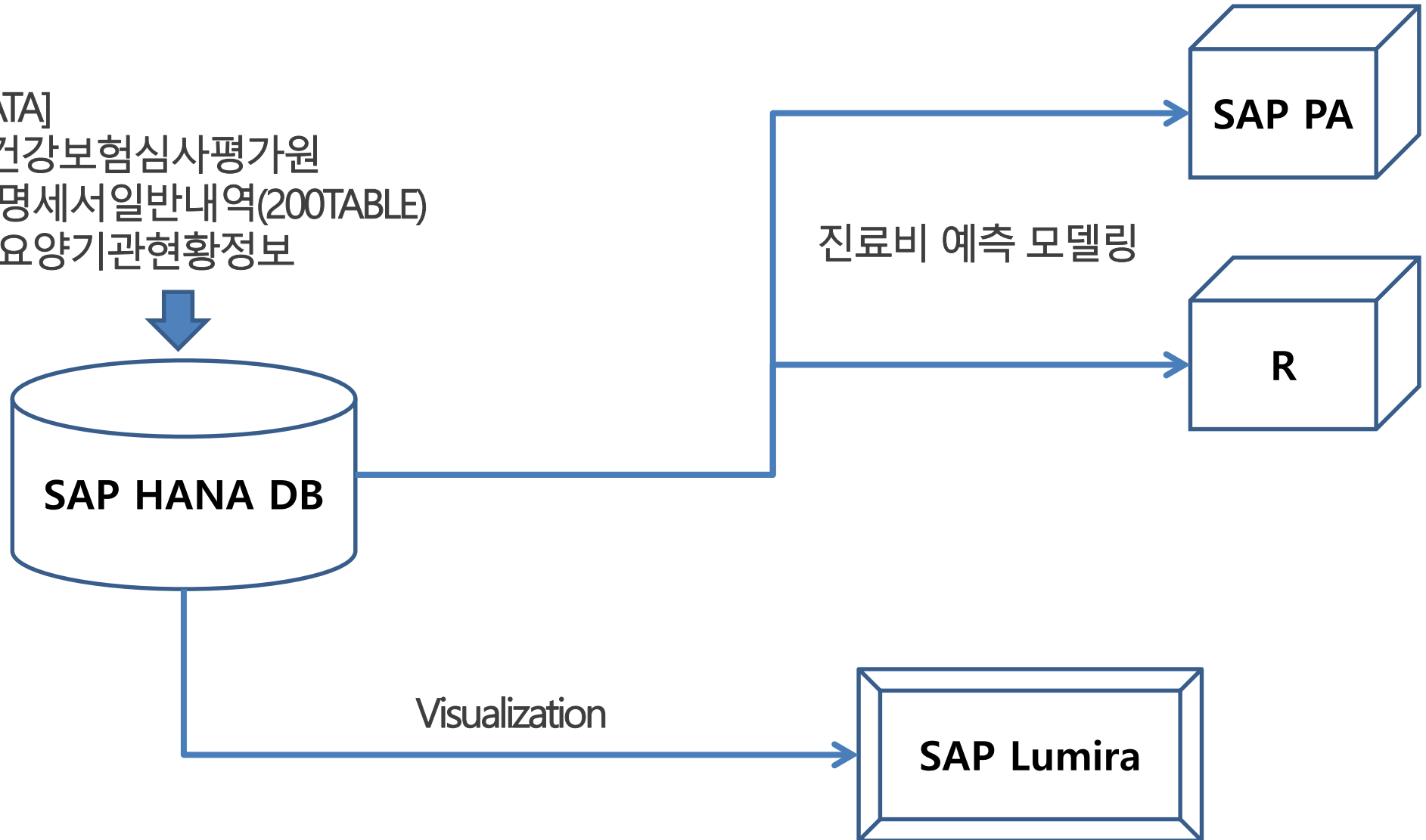
포괄수가제 및 **신생아** 데이터 분석을 주제로

DASHBOARD 및 진료비 예측 모델 구현

시스템 환경

[DATA]

1. 건강보험심사평가원
 - 명세서일반내역(200TABLE)
 - 요양기관현황정보



DATA 설명

명세서일반내역(200TABLE)

AB	SPEC_ID_SNO	AB	DMD_TP_CD	AB	DMD_FRM_CD	AB	FOM_TP_CD	AB	JID	AB	AGG	AB	SEX_TP_CD	AB	PAT_AGE	AB	INSUP_TP_CD
	1915644		0		3		031		380782		11		2		53.0		4
	2774633		0		3		031		987614		12		1		58.0		4
	3600438		0		3		031		1134515		13		1		61.0		4
	4142252		0		3		031		1229133		12		2		59.0		4
	4380515		0		3		031		840435		12		2		57.0		4

요양기관현황정보

AB	YID	AB	SIDO_CD	AB	RECU_CL_CD	AB	ORG_TP_CD	AB	BED_GRADE	12	PER_50BED_DOCTOR	12	PER_50BED_D_DENTAL	12	PER_50BED_D_DENTAL
	3		11		31		12		0		0		0		
	7		11		31		12		0		0		0		
	9		11		31		12		0		0		0		
	12		11		31		12		0		0		0		
	16		11		31		12		0		0		0		

건강보험 심사평가원의 명세서일반내역 및 요양기관현황정보 테이블을 활용하여 분석

2014년부터 2016년까지 36개의 월별 테이블을 년도를 기준으로 Union하고

기준연도와 기준월 Column 생성 후 Information View 생성

포괄수가제 설명

포괄수가제란?

환자가 입원해서 퇴원할 때까지 발생하는 진료에 대하여 질병마다 미리 정해진 금액을 내는 제도

4개 진료과 7개 질병군

- 안과 : 백내장수술(수정체 수술)
- 이비인후과 : 편도수술 및 아데노이드 수술
- 외과 : 항문수술(치질 등), 탈장수술(서혜 및 대퇴부), 맹장수술(충수절제술)
- 산부인과 : 제왕절개분만, 자궁 및 자궁부속기(난소, 난관 등)수술(악성종양 제외)

포괄수가제 VIEW 설명

CVCP_DRG_MED_KU17

AB	DMD_FRM_DETAIL	AB	JID	AB	SEX_DETAIL	AB	AGG_DETAIL	AB	PAT_AGE	AB	SICK	AB	KOREAN	AB	DGSBJT_CD_DETAIL	AB	DMD
	DRG		238521		여		35~39세		37.0		Z348		기타 정상임신...		산부인과		O01600
	DRG		288318		여		40~44세		40		Z349		상세불명의 정...		산부인과		O01600
	DRG		1435048		여		35~39세		36		Z355		고령 초임녀 ...		산부인과		O01600
	DRG		139853		여		35~39세		36		Z355		고령 초임녀 ...		산부인과		O01600
	DRG		893231		여		30~34세		33.0		Z359		상세불명의 고...		산부인과		O01600

Value	Count	Distribution
DRG	99,192	100.0 %

총 약 8378만 건의 명세서 데이터에서 99192건의 포괄수가제(DRG) 데이터 추출

주상병코드 4자리(ex. H250)로 데이터 전처리

EDA 후 영향을 끼치지 않는 필요 없는 변수들 제거

포괄수가제 주상병 분석

포괄수가제 주상병 분석

기준년도

0 selected

기준월

0 selected

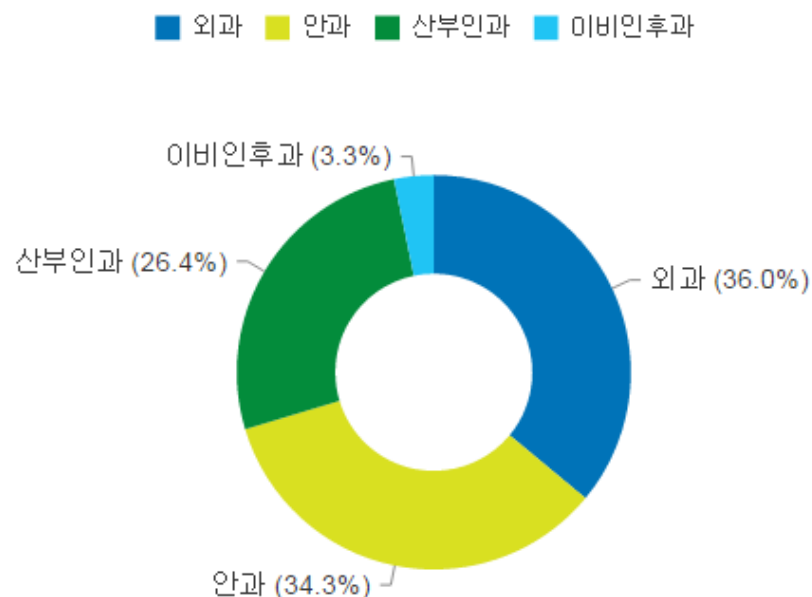
성별

0 selected

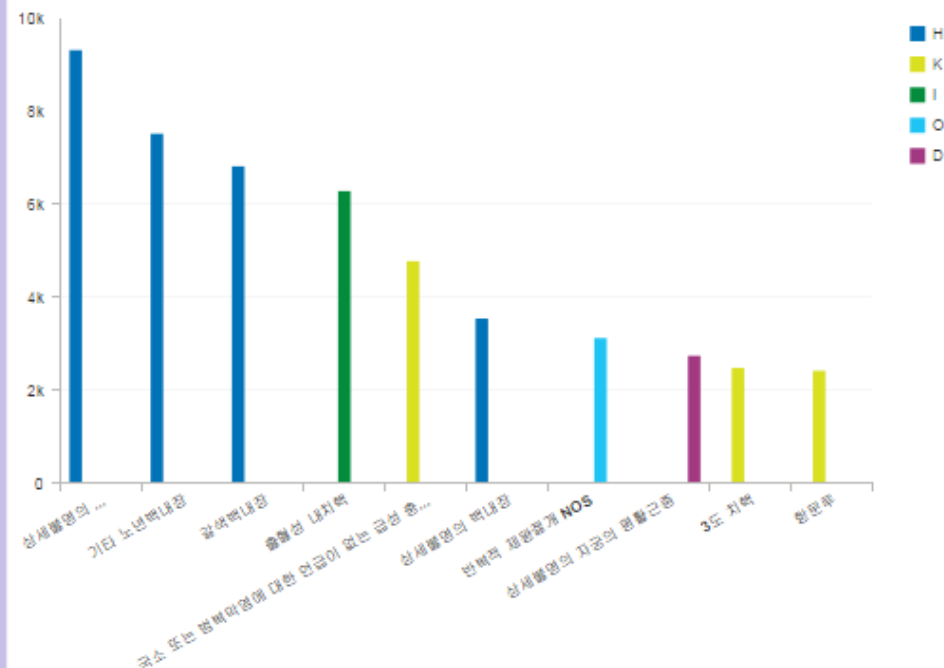
연령군

0 selected

포괄수가제 진료과목



포괄수가제 질병군



포괄수가제 진료비 분석

포괄수가제 진료비 분석

기준년도

0 selected

기준월

0 selected

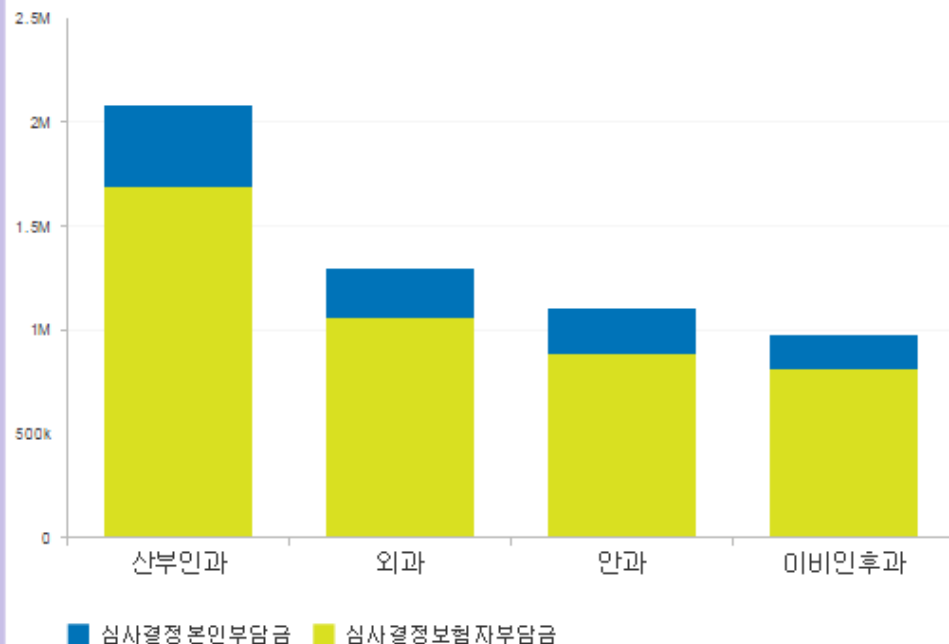
성별

0 selected

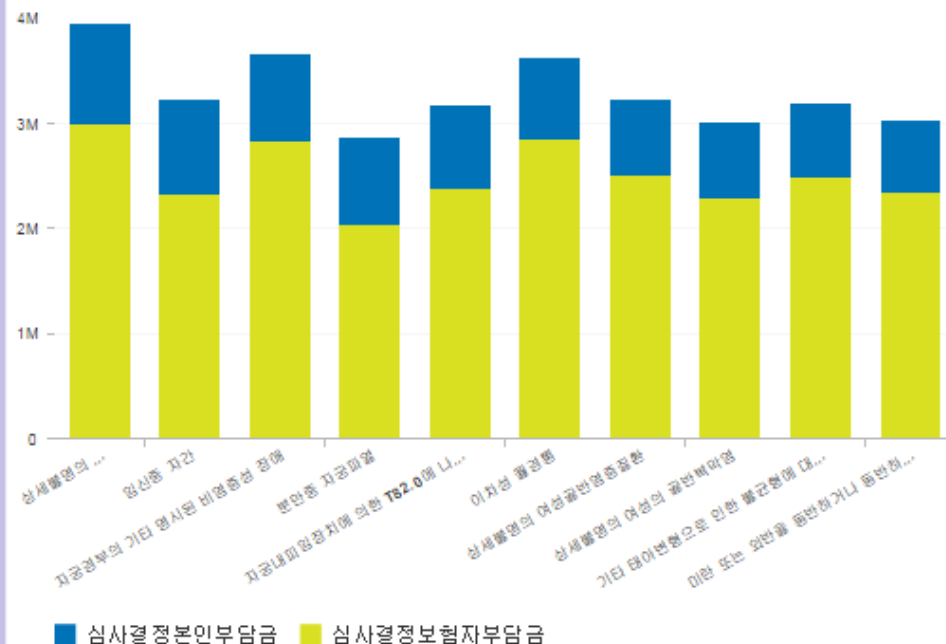
연령군

0 selected

포괄수가제 진료과목별 진료비



포괄수가제 주상병별 진료비



포괄수가제 진료비 예측 모형

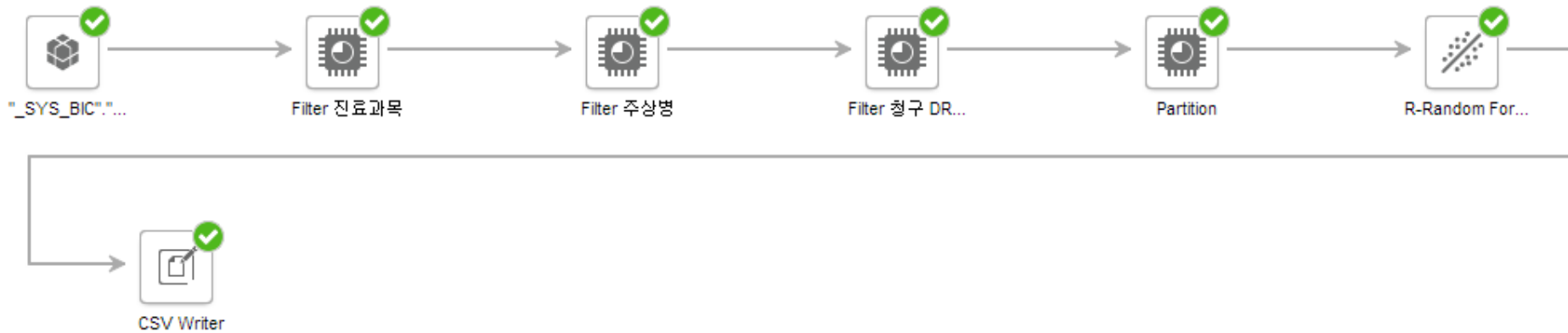
(단위:원)

수술명	수정체(백내장) 수술 ▼
입원일자 (요양개시일)	<input type="text"/> <input type="button" value="입력 예) 20170701"/> <p>※ 수술 시행을 위해 입원한 날을 선택하세요</p>
입원일수	<div>1 ▼</div> <p> ▪ 1일부터 30일 입원까지가 포괄수가에 적용되며 31일 이후 입원은 포괄수가에서 제외됩니다. ▪ 입원일수에 따라 진료비가 달라집니다. </p>
본인부담금 (자격기준)	<div>본인부담금 20% ▼</div> <p> 본인부담금 20%: 일반환자, 2016.7.1.이전 제왕절개분만 본인부담금 14%: 차상위만성질환자 18세 미만 본인부담금 10%: 희귀난치성질환자(6세미만 포함) 본인부담금 5%: 중증질환자, 2016.7.1. 이후 제왕절개분만 본인부담금 3%: 차상위 18세미만 본인부담경감대상자 중 6세 이상 15세 이하 </p>
상세수술명	<p> <input type="checkbox"/>수정체(백내장) 수술 소절개, 단안(C0510) <input type="checkbox"/>수정체(백내장) 수술 소절개, 양안(C0520) <input type="checkbox"/>수정체(백내장) 수술 대절개, 단안(C0530) <input type="checkbox"/>수정체(백내장) 수술 대절개, 양안(C0540) </p> <p>소절개, 대절개는 수정체 렌즈에 따라 결정되는 것입니다.</p> <p>양안이란 양쪽 눈을 동시에 수술하는 경우를 말합니다.</p>
야간·공휴 여부	<p><input type="checkbox"/> 야간, 공휴 수술</p> <p>※ 야간, 공휴수술은 18시~09시 또는 공휴일에 응급진료가 불가피하여 수술을 행한 경우를 말합니다. 해당하는 경우에만 체크하시기 바랍니다.</p>

Q 검색

질병 군 번호	명칭	상급종합병원		종합병원		병원		의원	
		총 진료 비	본인 부담 금	총 진료 비	본인 부담 금	총 진료 비	본인 부담 금	총 진료 비	본인 부담 금
C05100	수정체 소절개 수술 (유리체 절제술 유무와 무관), 단안,심각하거나 중증 혹은 중등도의 합병증이나 동반상	1,292,460	256,330	1,147,770	227,630	930,280	184,500	870,960	172,730
C05101	수정체 소절개 수술 (유리체 절제술 유무와 무관), 단안,중증 혹은 중등도의 합병증이나 동반상병 동반	1,375,580	234,020	1,208,190	205,540	980,640	166,830	920,240	156,550
C05102	수정체 소절개 수술 (유리체 절제술 유무와 무관), 단안,심각한 합병증이나 동반상병 동반	1,857,150	283,330	1,432,820	218,590	1,175,410	179,320	1,075,930	164,140

포괄수가제 진료비 예측 모형 (SAP PA)



* 총 99192개 관측 값 중 진료비 예측 모형을 만들기 위해 필터 적용

1) 진료과목 = {안과, 외과, 산부인과, 이비인후과} : 96349 건

2) 주상병 코드 = (상위 52 계수로 내림차순) : 95815 건

3) 청구 DRG 번호 = (상위 52 계수로 내림차순) : 95728 건

그 후 Training Set : 80 / Validation Set : 10 / Test Set : 10 으로 Partition 적용

포괄수가제 진료비 예측 모형 (SAP PA)

** Random Forest Regression 모델 적용

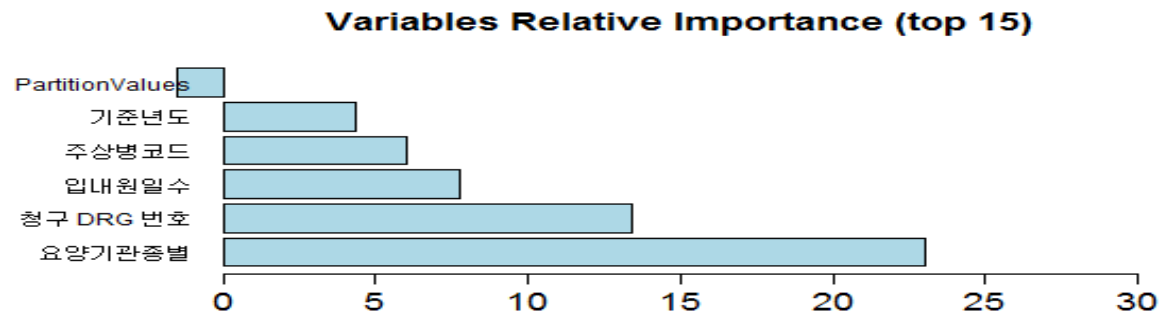
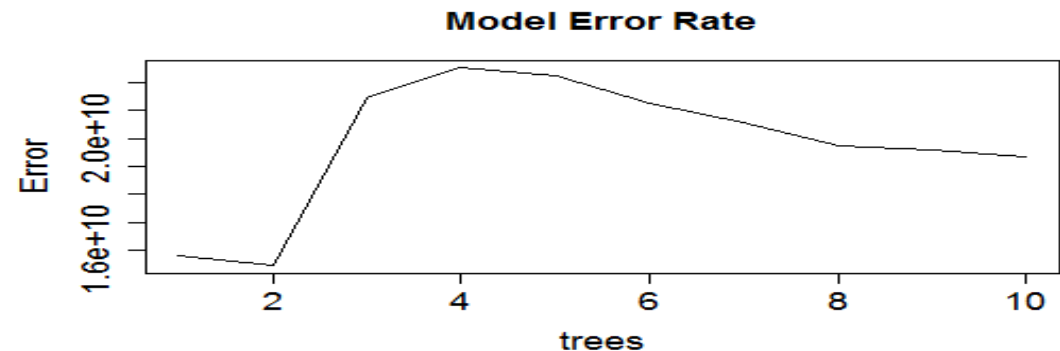
Target Columns : 심사결정요양급여비용총액

Features : 기준년도 + 주상병코드 + 청구 DRG 번호
+ 입내원일수 + 요양기관종별

	Length	Class	Mode
call	9	-none-	call
type	1	-none-	character
predicted	95728	-none-	numeric
mse	10	-none-	numeric
rsq	10	-none-	numeric
oob.times	95728	-none-	numeric
importance	12	-none-	numeric
importanceSD	6	-none-	numeric
localImportance	0	-none-	NULL
proximity	0	-none-	NULL
ntree	1	-none-	numeric
mtry	1	-none-	numeric
forest	11	-none-	list
coefs	0	-none-	NULL
y	95728	-none-	numeric
test	0	-none-	NULL
inbag	0	-none-	NULL
terms	3	terms	call

-- Variable Importance --

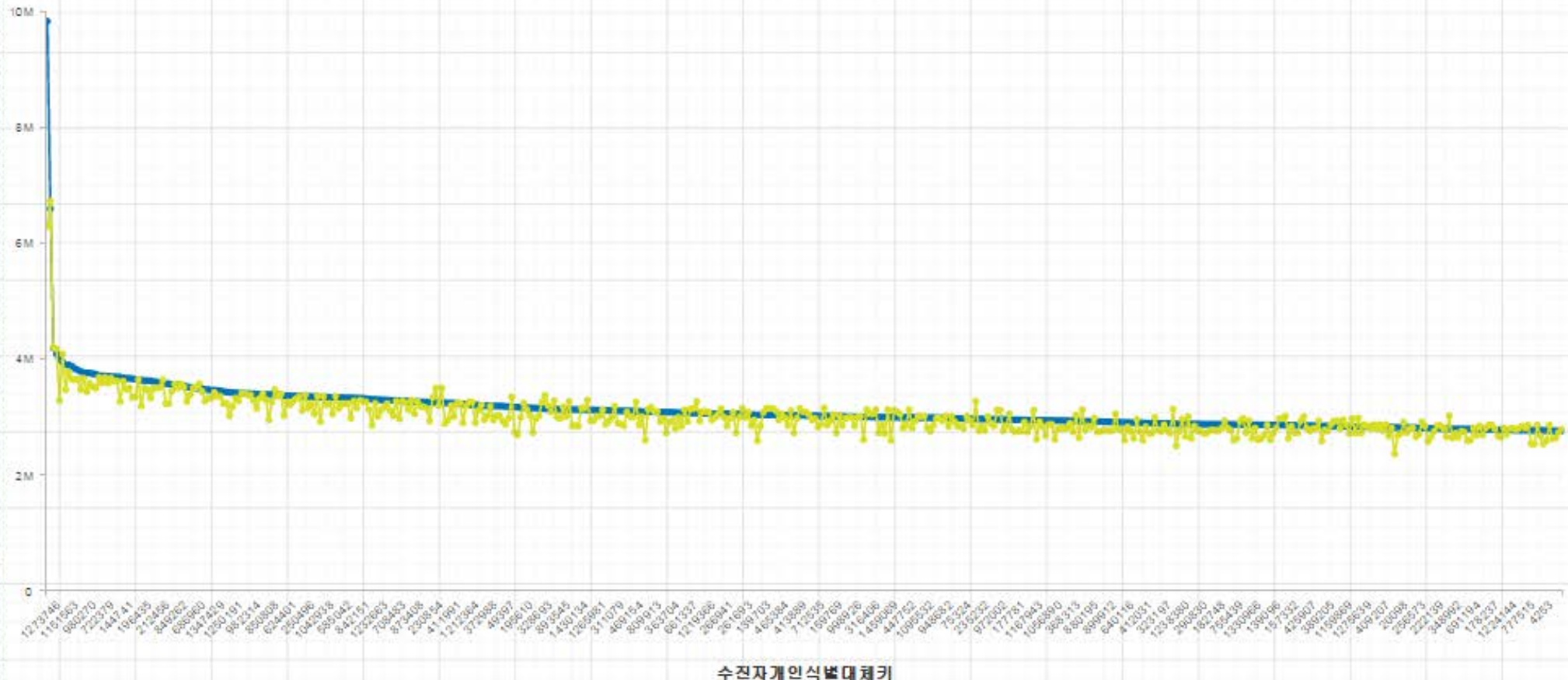
	%IncMSE	IncNodePurity
PartitionValues	-1.489365	2.422557e+13
기준년도	4.383855	2.153514e+14
입내원일수	7.787177	1.028251e+16
주상병코드	6.054447	1.061662e+16
청구 DRG 번호	13.384290	1.722350e+16
요양기관종별	22.995363	8.014392e+15



포괄수가제 진료비 예측 모형 (SAP PA)

포괄수가제 진료비 예측 모형 (sample=500)

● 심사결정요양급여비용총액 ● PredictedValues



포괄수가제 진료비 예측 모형 (R)

진료비 예측 모형을 자세히 알아보기 위해 R에서 더 깊이 있는 분석을 실시

```
1 PG_DN<-read.csv("PG_DN.csv", header=T, stringsAsFactors = F)
2 str(PG_DN)
3
4 library(dplyr)
5 DN <- dplyr::select(PG_DN, c(심사결정요양급여비용총액,기준년도,주상병코드,입내원일수,요양기관종별,청구.DRG.번호))
6
7 nrows <- NROW(DN)
8 set.seed(218)
9 index <- sample(1:nrows, 0.7 * nrows)
10
11 train <- DN[index,]
12 test <- DN[-index,]
13
14 library(randomForest)
15
16 learn_rf <- randomForest(심사결정요양급여비용총액~,
17                           data=train, ntree=500, proximity=T, importance=T)
18 pre_rf <- predict(learn_rf, test[,1])
19 cm_rf <- confusionMatrix(pre_rf, test$심사결정요양급여비용총액)
20 cm_rf
21
22 plot(learn_rf, main="Random Forest (Error Rate vs. Number of Trees)")
23
24 plot(margin(learn_rf,test$diagnosis))
25
26 varImpPlot(learn_rf)|
```



```
> learn_rf <- randomForest(심사결정요양급여비용총액~,
+                           data=train, ntree=500, proximity=T, importance=T)
Error: cannot allocate vector of size 33.5 Gb
In addition: warning messages:
1: In data.matrix(x) : NAs introduced by coercion
2: In data.matrix(x) : NAs introduced by coercion
3: In data.matrix(x) : NAs introduced by coercion
```

노트북 메모리 에러

신생아 VIEW 설명

CVCP_NEWBORN_KU17

AB	JID	AB	DMD_DRG_NO	AB	YEAR	AB	MON	AB	MSICK	AB	SEX_DETAIL	AB	DGSBJT_CD_DETAIL	AB	KOREAN	AB	CT	AB	MRI
1132429	D64010	2015	10월	H66	남	소아청소년과	고막염 동반 ...	0	0										
580832	J69000	2015	10월	L30	남	소아청소년과	기타 피부염	0	0										
137071	D64011	2015	11월	J06	남	소아청소년과	다발성 및 상...	0	0										
327612	E69110	2015	12월	J20	남	소아청소년과	15세 미만에서...	0	0										
1016139	E61500	2015	12월	J21	여	소아청소년과	급성 세기관지...	0	0										

Value	Count	Distribution
0.9	929,781	91.2 %
0.2	89,549	8.8 %

PAT_AGE :

0.2 – 28일 미만

0.9 : 28일 이상 1세 미만

총 약 8378만 건의 명세서 데이터에서 1019330건의 신생아 데이터 추출

주상병코드 4자리(ex. H250)로 데이터 전처리

EDA 후 영향을 끼치지 않는 필요 없는 변수들 제거

신생아 진료과목 분석

신생아 진료과목 분석

기준년도

0 selected

기준월

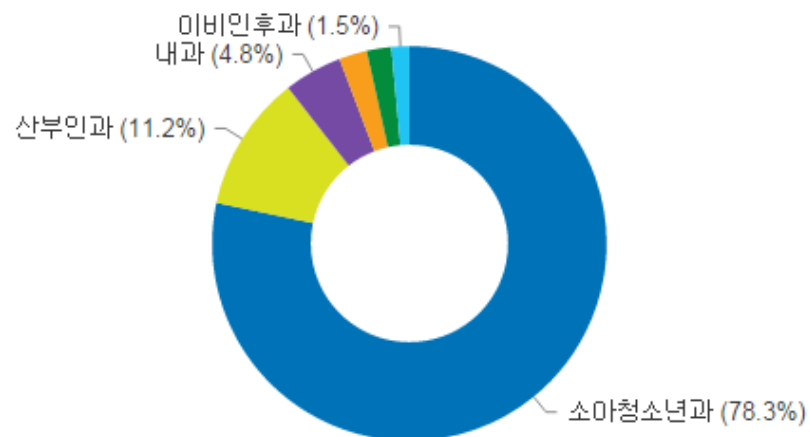
0 selected

성별

0 selected

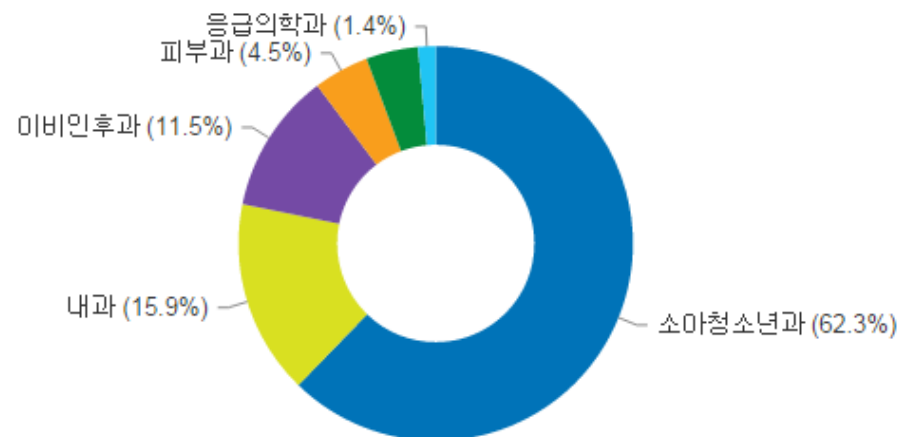
진료과목 (28일 미만)

■ 소아청소년과 ■ 산부인과 ■ 내과 ■ 피부과 ■ 안과 ■ 이비인후과



진료과목 (28일 이상 1세 미만)

■ 소아청소년과 ■ 내과 ■ 이비인후과 ■ 피부과 ■ 안과 ■ 응급의학과



신생아 주상병 분석

신생아 주상병 분석

기준년도

0 selected

기준월

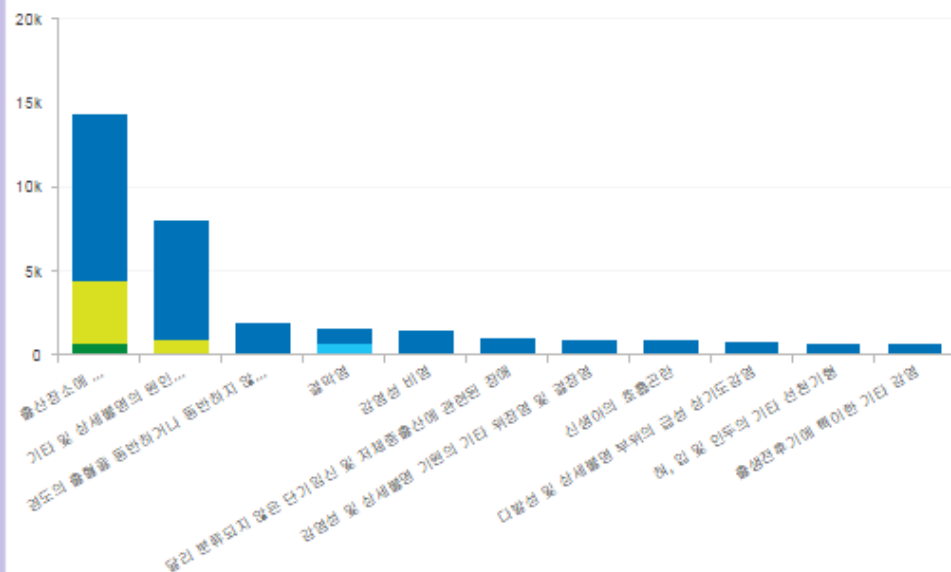
0 selected

성별

0 selected

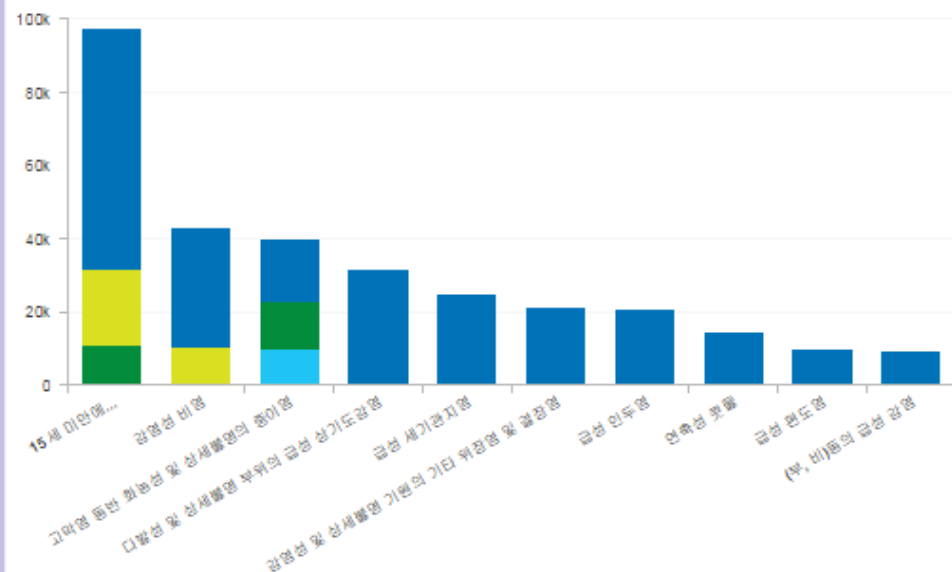
주상병 (28일 미만)

■ 소아청소년과 ■ 산부인과 ■ 내과 ■ 안과



주상병 (28일 이상 1세 미만)

■ 소아청소년과 ■ 내과 ■ 미비인후과 ■ 안과



신생아 진료비 분석

신생아 진료비 분석

기준년도

0 selected

기준월

0 selected

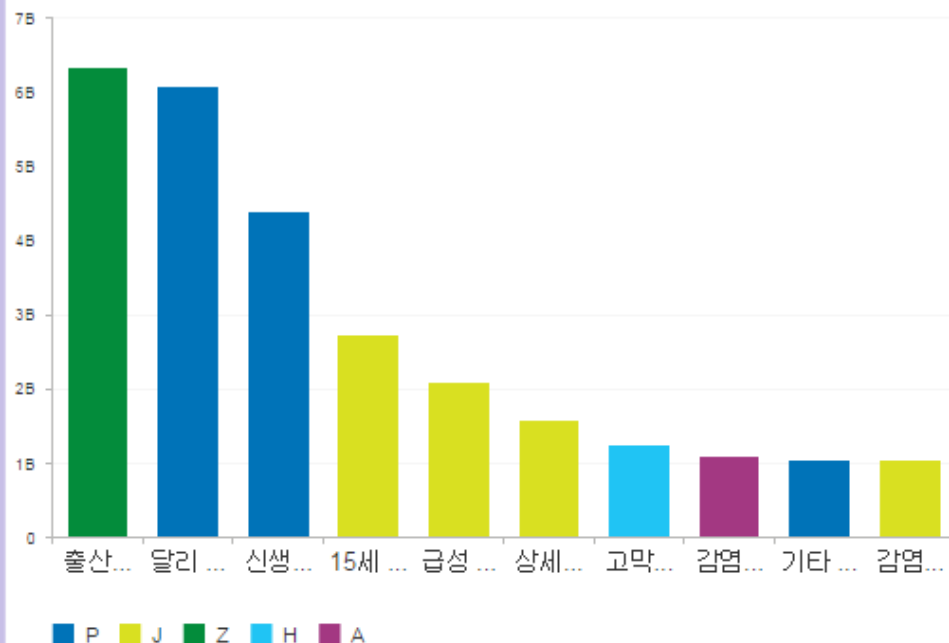
연령

0 selected

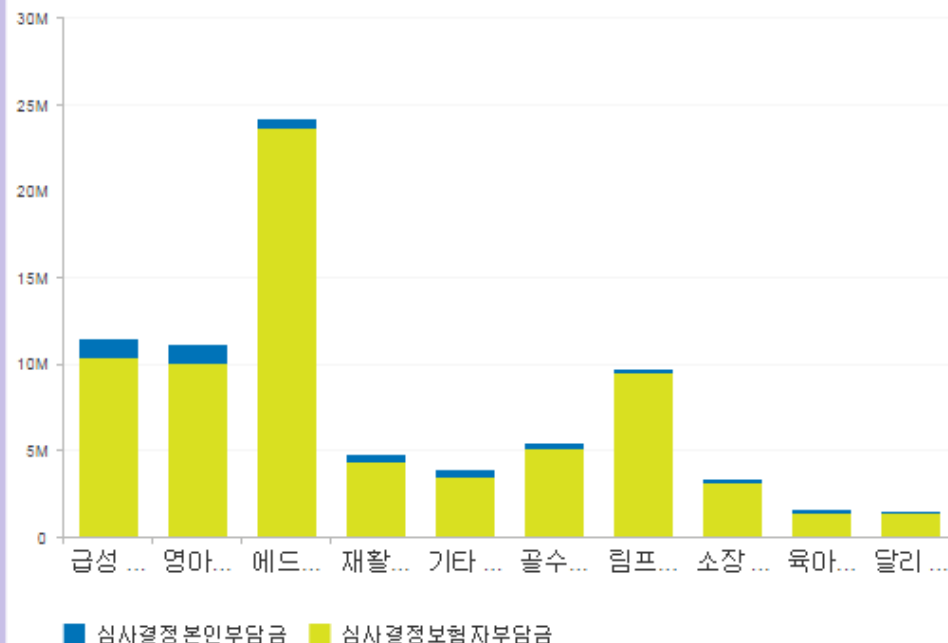
진료과목

0 selected

신생아 진료비 (요양급여비용 총액)



신생아 진료비 (본인 부담금 기준)



Q&A