



건강검진 정보를 활용한 보험사의 전략



목 차



01 스토리텔링



02 활용할 데이터 소개

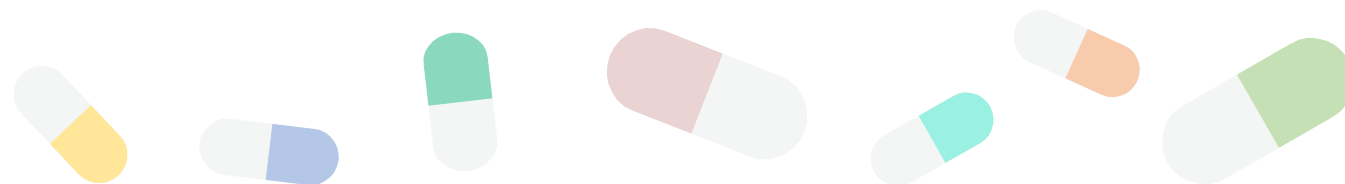


03 시각화 구현 방법



04 시각화 설명 및 해석

01 스토리텔링





01 스토리 텔링

최근 상승세를 보이는 신생 보험사 A는 더욱 성장하기 위해 고민 중이다. 때문에 어떤 전략을 세워야 더욱 회사를 성장시킬 수 있을지 전략을 세우려고 한다.

이 보험사에게 가장 필요한 정보는 2가지이다

1. 연령대별로 어떤 전략을 세워야 하는가?

2. 어느 질병에 혜택을 높여야 소비자들이 좋아할까?

이 보험사에게 필요한 정보를 데이터 시각화를 통해 얻을 수 있을까?



02 활용할 데이터 소개





02 활용할 데이터 소개

국민건강정보 (출처:국민건강보험)

<https://www.data.go.kr/emphasisData/show.do;jsessionid=WyOMIPgJtDIZl2nefZWfSH.node20>

국민건강정보는 국민건강보험자격자(전국민)의 약 2%에 해당하는 100만명에 대한 2002년부터 2013년까지의 진료내역정보, 약품처방내역정보, 건강검진정보이다.

국민건강정보 안에있는 여러 정보들 중 진료내역 정보를 이용할 것이다



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	기준년도	가입자 일련	진료내역일	성별코드	연령대코드	시도코드	요양개시일	서식코드	진료과목코	주상병코드	부상병코드	요양일수	입내원일수	심결가산율	심결요양급
2	2017	1	4661608	1	5	11	20170316	3	13	J060	R51	1	1	0.15	16420
3	2017	1	14468123	1	5	11	20170123	3	1	R51	K297	1	1	0.15	16420
4	2017	1	22980223	1	5	11	20170404	3	13	J0190	J060	1	1	0.15	14650
5	2017	1	23050697	1	5	11	20170407	3	13	J0190	J060	1	1	0.15	10620
6	2017	1	29517148	1	5	11	20170516	3	14	L309	L301	1	1	0.15	21760
7	2017	1	31437896	1	5	11	20170525	3	14	L309	L301	1	1	0.15	10620
8	2017	1	40255948	1	5	11	20170726	3	13	J060	J343	1	1	0.15	16500
9	2017	1	41427169	1	5	11	20170918	3	1	R611	B07	1	1	0.15	18360
10	2017	1	44725321	1	5	11	20170915	3	1	R611	B07	1	1	0.15	122790
11	2017	1	44725322	1	5	11	20170901	3	1	R611	B07	1	1	0.15	122790
12	2017	1	47924163	1	5	11	20170928	3	13	J060	J40	1	1	0.15	14860
13	2017	1	53808162	1	5	11	20171030	3	13	H6800	J304	1	1	0.15	23240
14	2017	2	13554560	2	2	28	20170206	3	1	J0190	J209	1	1	0.15	15570
15	2017	2	15145058	2	2	28	20170320	3	1	J0190	J209	1	1	0.15	15570
16	2017	2	23621437	2	2	28	20170422	3	1	J0190	J209	1	1	0.15	13230
17	2017	2	26835493	2	2	28	20170427	3	1	J0190	J209	1	1	0.15	10910
18	2017	2	31464692	2	2	28	20170516	3	1	J0190	J209	1	1	0.15	12950
19	2017	2	32251105	2	2	28	20170605	3	1	B_	J0190	1	1	0.15	10620
20	2017	2	32455449	2	2	28	20170608	3	1	J0190	J0391	1	1	0.15	10620
21	2017	2	45878641	2	2	28	20170811	3	1	J209	K529	1	1	0.15	14860
22	2017	2	61216285	2	2	28	20171110	3	1	J0190	J209	1	1	0.15	14860
23	2017	2	62002381	2	2	28	20171218	3	24	K5909	R1049	1	1	0.25	95670
24	2017	3	662717	2	8	48	20170207	3	5	R111	M6584	1	1	0.15	15550
25	2017	3	662718	2	8	48	20170206	3	5	R111	M6584	1	1	0.15	15550
26	2017	3	662719	2	8	48	20170209	3	5	R111	M6584	1	1	0.15	15550
27	2017	3	662720	2	8	48	20170213	3	5	R111	M6584	1	1	0.15	15550
28	2017	3	705916	2	8	48	20170208	3	5	R111	M6584	1	1	0.15	17330
29	2017	3	715445	2	8	48	20170203	3	5	M6584	R111	1	1	0.15	27630
30	2017	3	4653936	2	8	48	20170324	3	12	H0411	H101	1	1	0.15	10620
31	2017	3	14338419	2	8	48	20170104	3	12	H0411	H101	1	1	0.15	19430
32	2017	3	25008062	2	8	48	20170512	3	12	H0411	H108	1	1	0.15	12920



02 활용할 데이터 소개

활용할 요소들

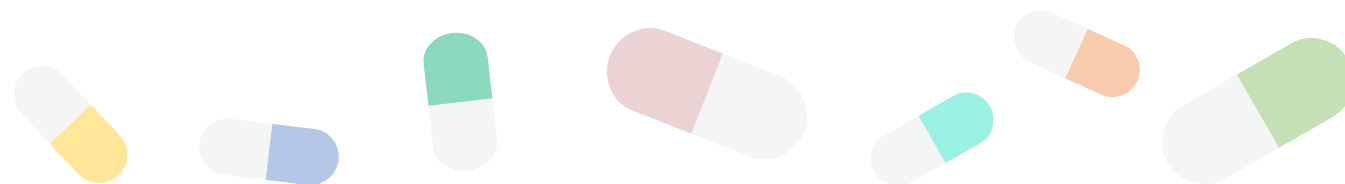
주상병코드 : 정해진 규칙을 가지고 영어와 숫자만으로 질병을 구분한 코드,
A~Z로 질병을 큰 유형으로 나눈 뒤 세부사항은 뒤에 붙는 번호로 결정한다

심결본인부담금: 진료 후 총 진료자에게 청구되는 비용

연령대코드: 5단위로 진료자의 나이를 알 수 있는 코드

A00-B99	I. 특정 감염성 및 기생충성 질환 Certain infectious and parasitic diseases
C00-D48	II. 신생물 Neoplasms
D50-D89	III. 혈액 및 조혈기관의 질환과 면역메커니즘을 침범한 특정 장애 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90	IV. 내분비, 영양 및 대사 질환 Endocrine, nutritional and metabolic diseases
F00-F99	V. 정신 및 행동 장애 Mental and behavioural disorders
G00-G99	VI. 신경계통의 질환 Diseases of the nervous system
H00-H59	VII. 눈 및 눈 부속기의 질환 Diseases of the eye and adnexa
H60-H95	VIII. 귀 및 유도의 질환 Diseases of the ear and mastoid process

03 시각화 구현방법





03 시각화 구현 방법

파이썬을 이용해 주어진 CSV 파일을 가공했다.

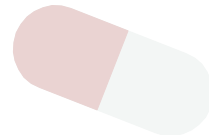
시각화를 위한 2가지 테이블을 재구성했고 이를 **mysql**에 저장하였다

Mysql에 저장된 데이터들을 **Google Chart**로 웹으로 볼 수 있도록 구현했다

전처리는 가공과 동시에 진행했다. 형식에 맞지 않는 데이터거나 빈값이면 계산에서 제외되는 형식으로 코딩했다

(자세한 코드는 첨부해서 넣어놨습니다.)

04 시각화 설명 및 해석





04 시각화 설명 및 해석

연령대별로 어떤 전략을 세워야 하는가?

이 문제를 해결하기 위해서 주어진 데이터를 이용해 다음과 같은 테이블을 만들었다

	idx	age	A_B	C_D	D	E	F	G
	1	0	56621	782	802	998	753	451
	2	10	46761	2153	808	7111	8491	8491
	3	20	40789	4241	1170	4293	24056	7882
	4	30	54942	10298	1416	12766	29464	9172
	5	40	61118	26352	3015	37035	39046	16957
	6	50	66680	47974	3763	95372	44878	28084
	7	60	66402	62908	2101	170...	51860	39608

age는 연령대, 그 뒤에 알파벳들은 질병코드를 알파벳으로 구분해 연령대 별로 얼마나 카운팅 됐는지를 넣어준 것이다

질병 코드는 알파벳 기준으로 질병의 유형을 구분하기 때문에 그림처럼 테이블을 구성하면 연령대별로 어떤 유형의 질병에 많이 걸렸는지 알 수 있다.



04 시각화 설명 및 해석

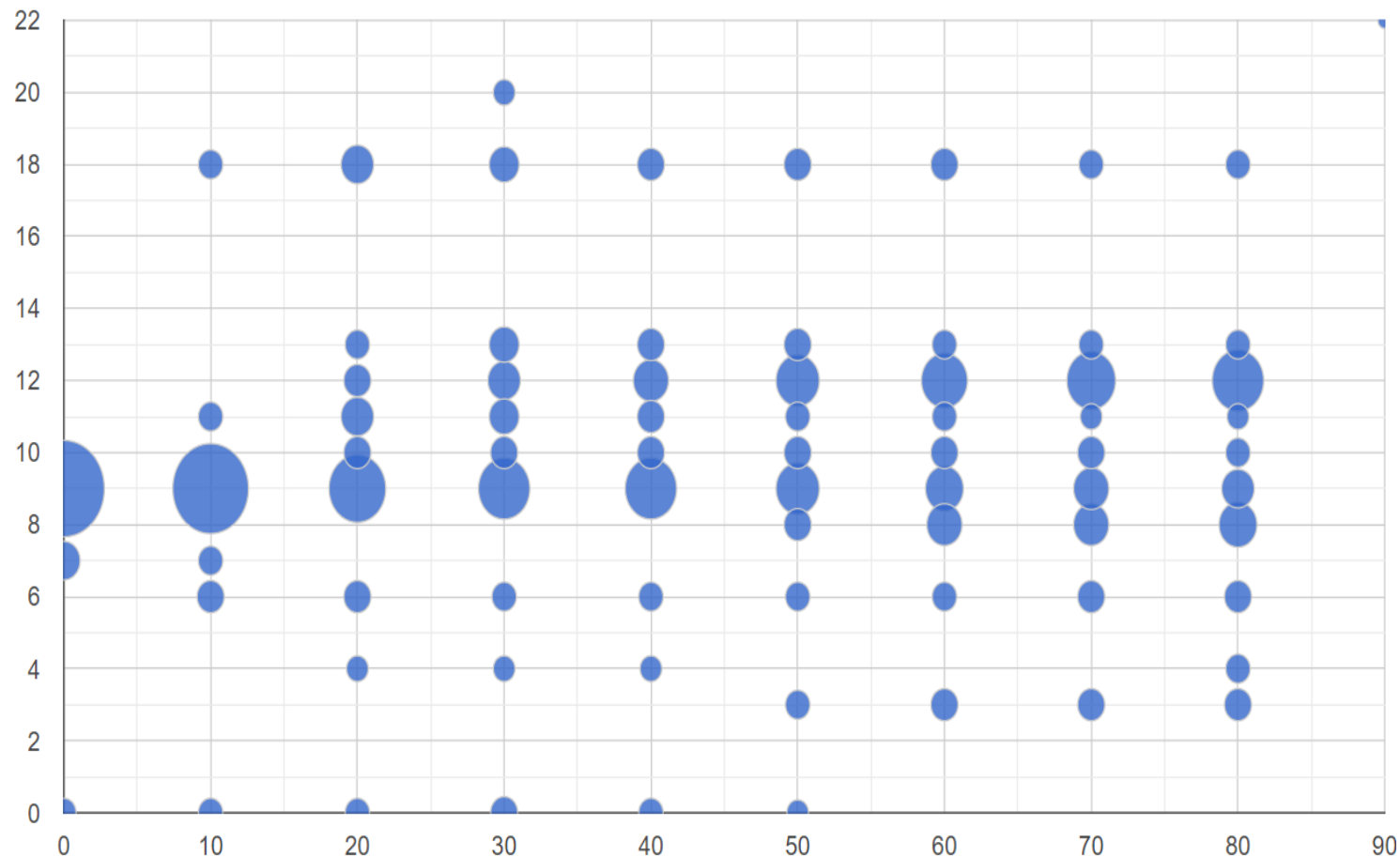
X축: 10단위로 나눈 연령대이며

Y축: 질병코드를 알파벳으로 나눈 22가지 유형이다

점의 크기 : 각 질병유형이 연령대에서 차지하는 비중, 각 유형의 진단횟수를 백분율로 표현해

(가시성을 위해 4%미만인 질병 유형은 그리지 않았다)

연령별 질병코드 진단 비율 (X축:연령, Y축:질병코드 숫자, 크기:각 연령에서 질병이 진단된 비율)





04 시각화 설명 및 해석

해석

1. 0대~10대는 특정 종류의 질병이 큰 비중을 차지한다
2. 20대 이후부터는 점점 다양한 질병으로 비율이 나뉘어지기 시작해 50대 이후로는 큰 변동이 없다.

=> 0~10대 보험은 혜택을 특정 질병에 집중!

20대 이후의 보험은 특정 질병에 혜택을 집중하기 보단 넓고 다양하게 혜택을 제공한다.





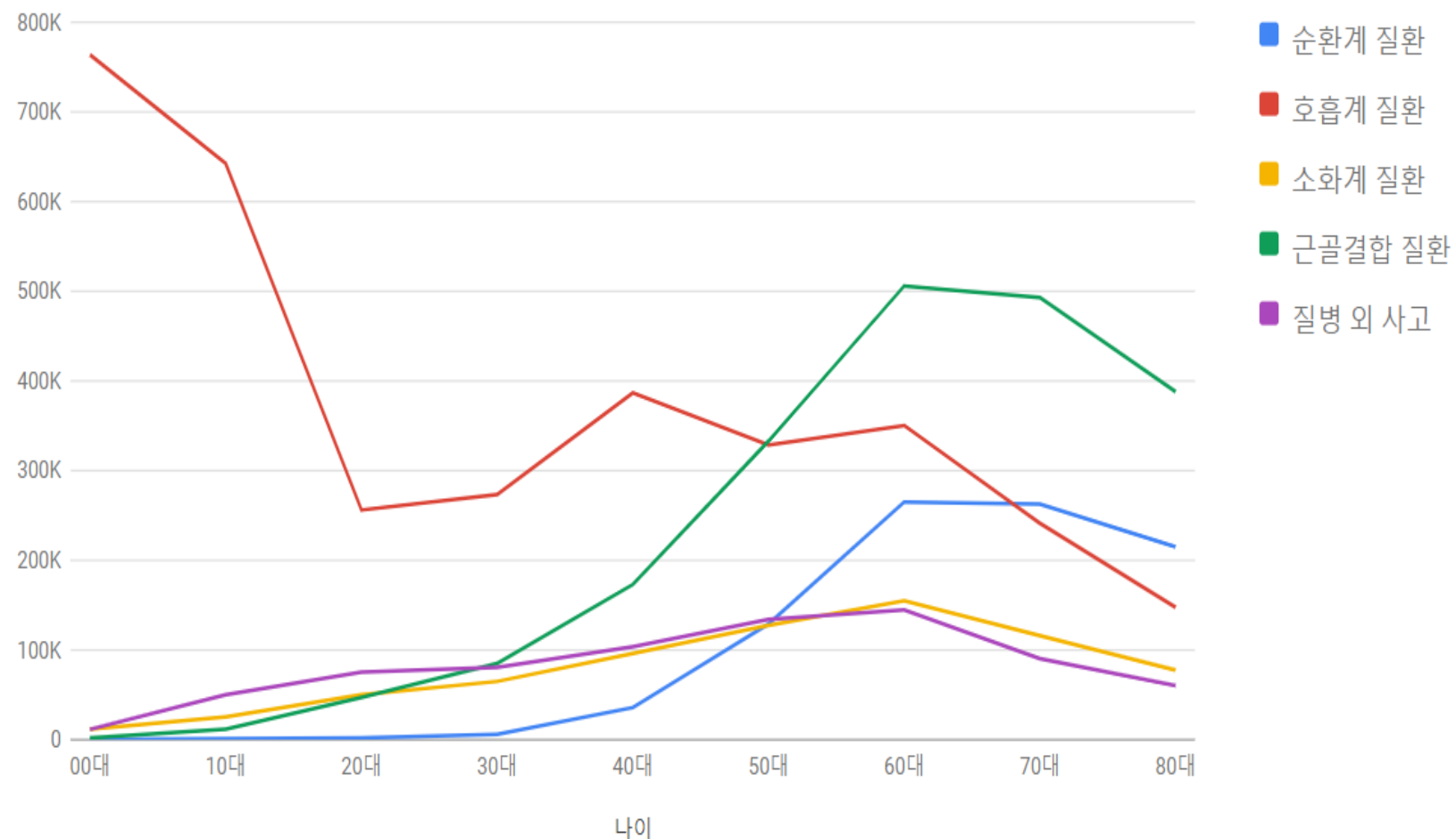
04 시각화 설명 및 해석

연령별로 상위를 가장많이 차지한 질병 5가지를 선 그래프로 그려보았다.

X축: 10단위로 나눈 연령대

Y축: 각 유형의 질병들이 진단된 횟수

각 연령, 상위 5개의 질병의 추이
X축:연령, Y축:질병이 진단된 횟수





04 시각화 설명 및 해석

해석

1. 0대~10대는 호흡기 질환이 가장 큰 비중을 차지한다
2. 20대 이후부터는 질병의 원인이 다양해지기 시작한다.
3. 30대~60대 사이에는 근골결합 질환과, 순환계 질환이 급격하게 증가한다

=> 0~10대는 혜택을 호흡기 질환에 집중!

30대~60대 보험은 근골결합 질환, 순환계 질환에 혜택을 집중하고 이를 강조해 마케팅을 한다!





04 시각화 설명 및 해석

어느 질병에 혜택을 높여야 소비자들이 좋아할까?

이 문제를 해결하기 위해서 주어진 데이터를 이용해 다음과 같은 테이블을 만들었다.

dis_code : 질병 코드

num: 각 질병코드가 진단된 횟수

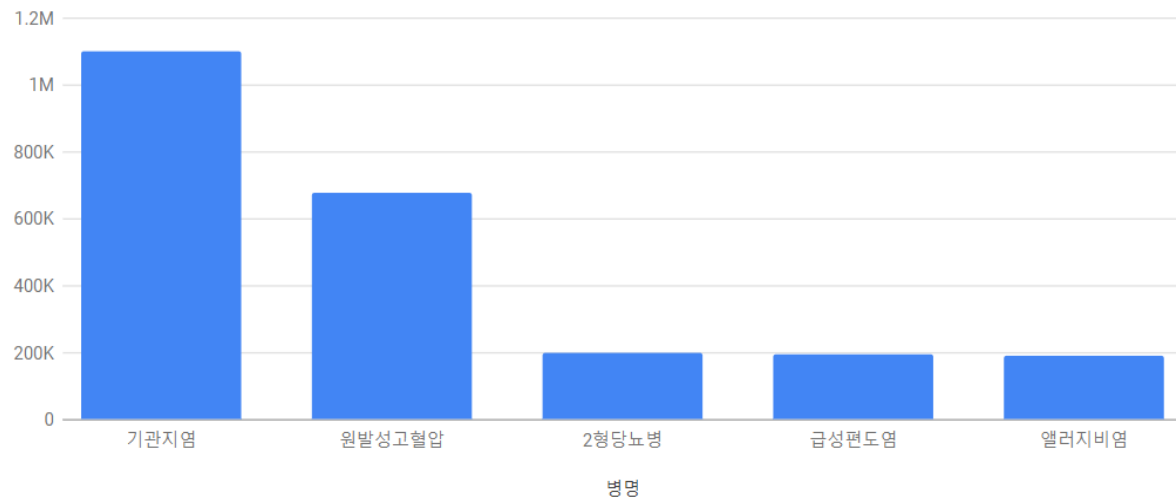
cost: 각 질병코드의 병원비의 평균값

idx	dis_code	num	cost
1	J209	1103359	1700
2	I109	680425	11992
3	F_	334665	192227
4	N_	212523	51858
5	E119	202139	33540
6	J0390	197887	23820
7	J304	193619	21120
8	J029	171936	19604
9	M170	155910	87252
10	J00	145067	15706
11	M4806	135928	62830
12	J060	119770	15303

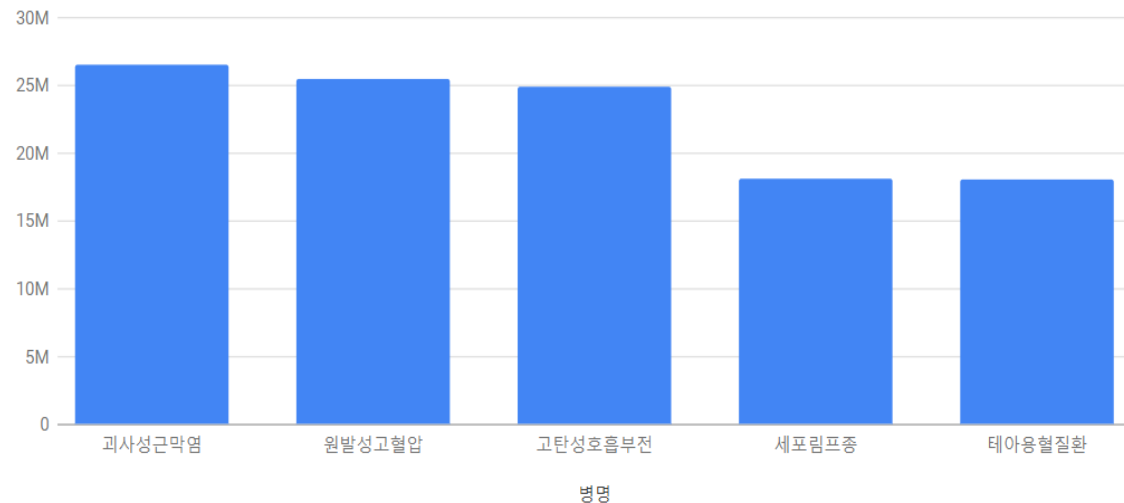


04 시각화 설명 및 해석

가장 많이 진단된 질병
X축:병명, Y축:진단된 횟수



치료비가 가장 많이 나온 질병
X축:병명, Y축:금액(원)



2번 테이블을 이용해서 만든 시각화

오른쪽 그림은 가장 많이 진단된 상위 5가지 질병을 시각화 한것이다. (X축:병명, Y축:진단된 횟수)

왼쪽 그림은 가장 금액이 많이 들어간 질병 상위 5가지를 시각화 한것이다. (X축:병명, Y축:금액)

효율을 찾기 위해서 2가지를 그려봤지만 두 그림을 통해 가장 효율이 좋은 질병이 무엇인지는 찾지 못하였다.



04 시각화 설명 및 해석

가장 많이 진단된 질병이라도 환자가 부담이 되지 않을 정도의 가벼운 질병이면 혜택을 늘려도 가입자가 느끼는 큰 효과는 느끼지 못할것이고

가장 금액이 큰 질병에 혜택을 늘려도 진단된 환자의 수가 적다면 소비자가 큰 체감을 하기가 어렵기 때문이다.

가장 효율이 좋은 질병을 찾기 위해서는 때문에 환자가 부담하는 비용, 진단된 횟수 이 두가지를 고려할 필요가 있었다.

이 두가지를 모두 고려하기 위해서

효율성 = 진단된 횟수 * 병원비 라는 식을 세워서 그러면 이를 해결할 수 있다고 생각했다.

효율성이 좋으려면 진단된 횟수가 많으면서 병원비도 커야 많은 가입자들이 혜택의 체감을 크게 느낄 수 있기 때문이다.



04 시각화 설명 및 해석

효율성을 기준으로 정렬했을때 나오는 상위 5가지를 그려보았다.

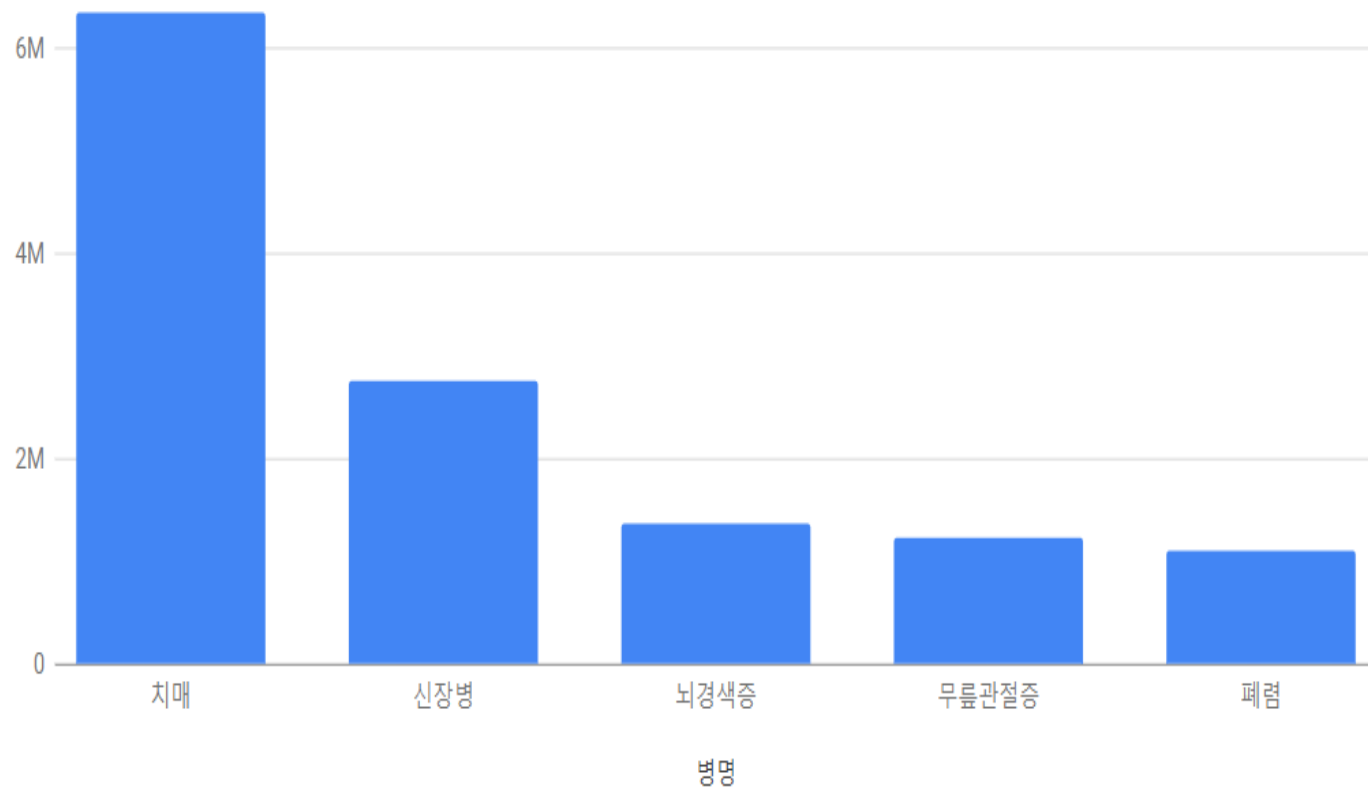
X축: 병명

Y축: 효율성 = 진단수 * 금액 / 10000

(효율성의 값이 너무 커서 10000을 나눠주어 그래프 데이터 저장하는데 무리가 없도록 조정해주었다.)

가장 효율이 좋은 질병

X축:병명, Y축:효율





04 시각화 설명 및 해석

해석

1. 가장 효율성이 높은 질병은 치매이고 이는 2위와 2배차이가 날만큼 압도적이다.

⇒ 특정 질병의 혜택을 강화했을때 사용자들이 체감을 크게 체감할 수 있는 질병은 치매이다.





마무리

이 보험사가 알고싶어 했던 2가지

1. 연령대별로 어떤 전략을 세워야 하는가?
2. 어느 질병에 혜택을 높여야 소비자들이 좋아할까?

이 두가지에 대한 질문에 대한 답을 시각화로 얻은 답을 정리해보면 다음과 같다

1. 0~10대 보험에 집중해야할 혜택은 호흡기 질환에 대한 혜택이다.
2. 20대 이후의 보험은 특정 질병에 혜택을 집중하기 보단 넓고 다양하게 혜택을 제공한다.
3. 30대~60대 보험은 근골결합 질환, 순환계 질환에 혜택을 집중하고 이를 강조해 마케팅을 한다.
4. 특정 질병의 혜택을 강화했을때 사용자들이 체감을 크게 체감할 수 있는 질병은 치매이다.

Thank You 

