

*Everybody likes chicken*

# 치렐루야 치킨공화국에서 살아남는법

*I like chicken*

1조 BA40

(김민호, 최현우, 유소희, 박지환, 박주은, 김민우)

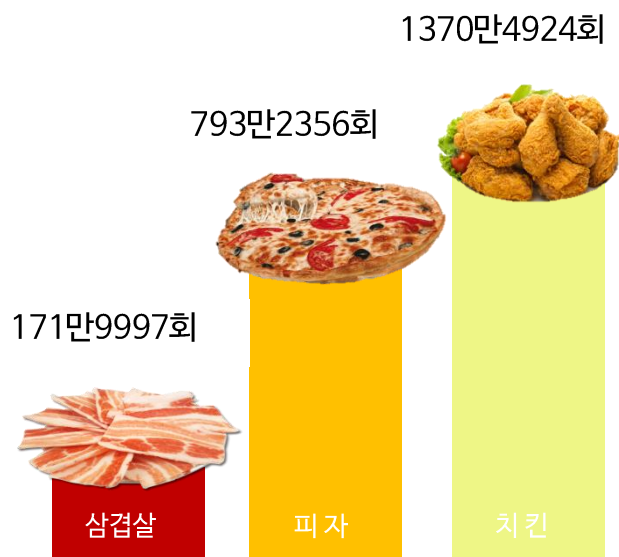
# 목 차

1. 분석 배경
2. 데이터 소개
3. 분석 내용
4. 모델링
5. 결론



# 1. 분석 배경

외식 메뉴별 언급량



자료. 다음소프트 분석 2011

SNS (사회관계서비스망)상 언급량을 분석해보면 치킨이 다른 외식메뉴보다 언급 횟수가 높은 것으로 확인되었다.

'치느님' '치얼루야'  
치킨 사랑이 유별난 대한민국은  
역시  
'치킨 공화국'



## 2. 데이터 소개

### 서울시 치킨 판매업종 이용 통화량

A	B	C	D	E	F	G	H	I	J
	기준일	요일	성별	연령대	시도	시군구	읍면동	업종	통화건수
0	20190901	일	여	30대	서울특별시	강남구	개포동	치킨	5
1	20190901	일	남	30대	서울특별시	강남구	논현동	치킨	20
2	20190901	일	남	30대	서울특별시	강남구	삼성동	치킨	20
3	20190901	일	남	30대	서울특별시	강남구	세곡동	치킨	5
4	20190901	일	남	30대	서울특별시	강남구	수서동	치킨	5
5	20190901	일	여	60대이상	서울특별시	강남구	수서동	치킨	5
6	20190901	일	여	60대이상	서울특별시	강남구	도곡동	치킨	5
7	20190901	일	여	30대	서울특별시	강남구	논현동	치킨	24

자료. SK 빅데이터 허브 (2019.09)

데이터
기준일
요일
성별
연령대
시도
시군구
읍면동
업종
통화건수

종속변수





## 2. 데이터 소개

### 최종 데이터

	A	B	C	D	E
1	day	gender	age	gu	call
2	1	2	30	1	1
3	1	1	30	1	2
4	1	1	30	1	2
5	1	1	30	1	1
6	1	1	30	1	1
7	1	2	60	1	1
8	1	2	60	1	1
9	1	2	30	1	2
10	1	2	30	1	1

데이터
<del>기준일</del>
요일
성별
연령대
<del>시도</del>
시군구
<del>읍면동</del>
<del>업종</del>
통화건수

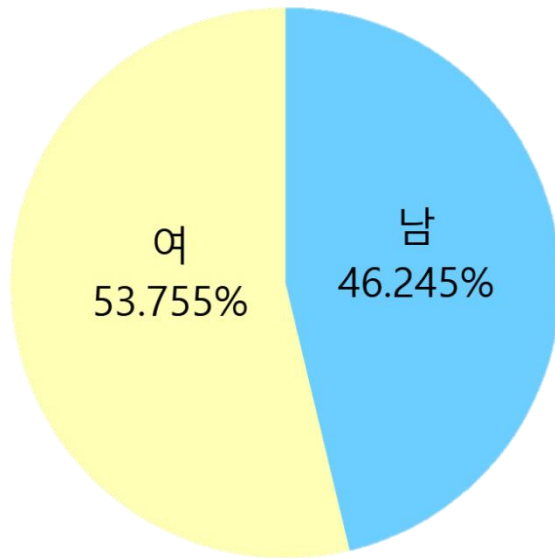
종속변수



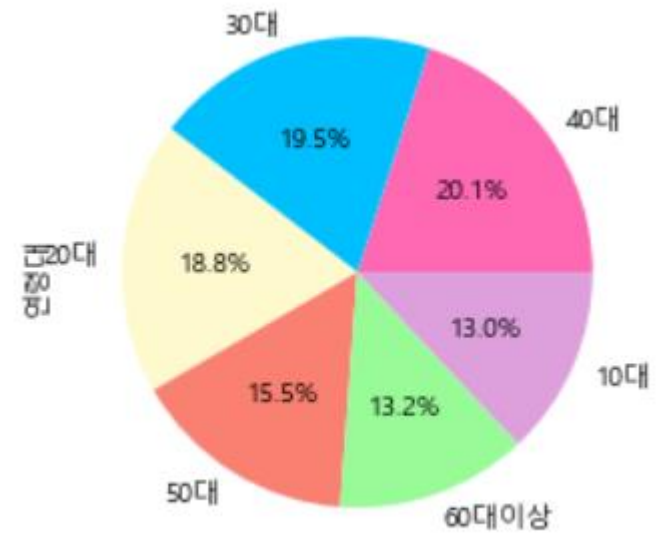


### 3. 분석 내용

성별 통화건수



연령대별 통화건수





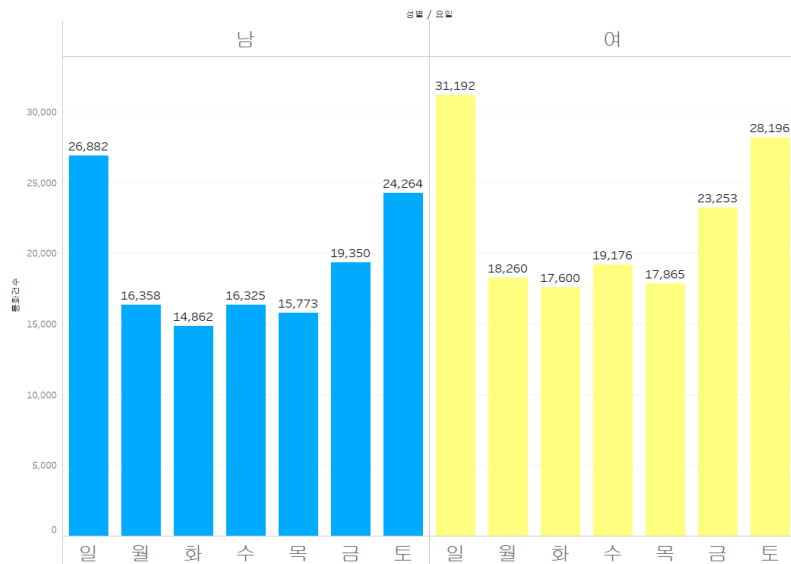
### 3. 분석 내용



#### 성별 요일별 통화건수

남 여

성별 요일별 통화건수

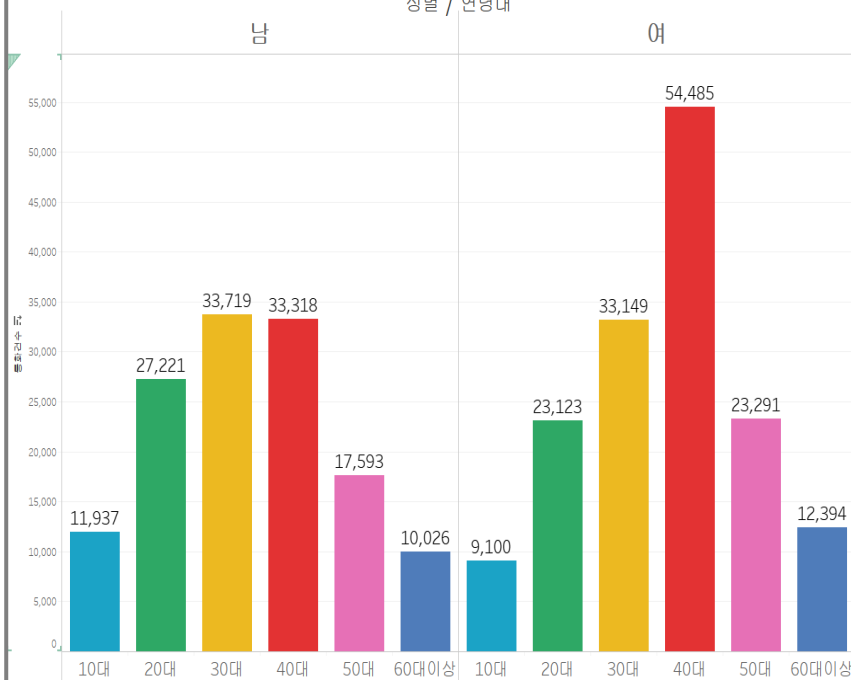


#### 성별 연령별 통화건수

성별 / 연령대

남

여



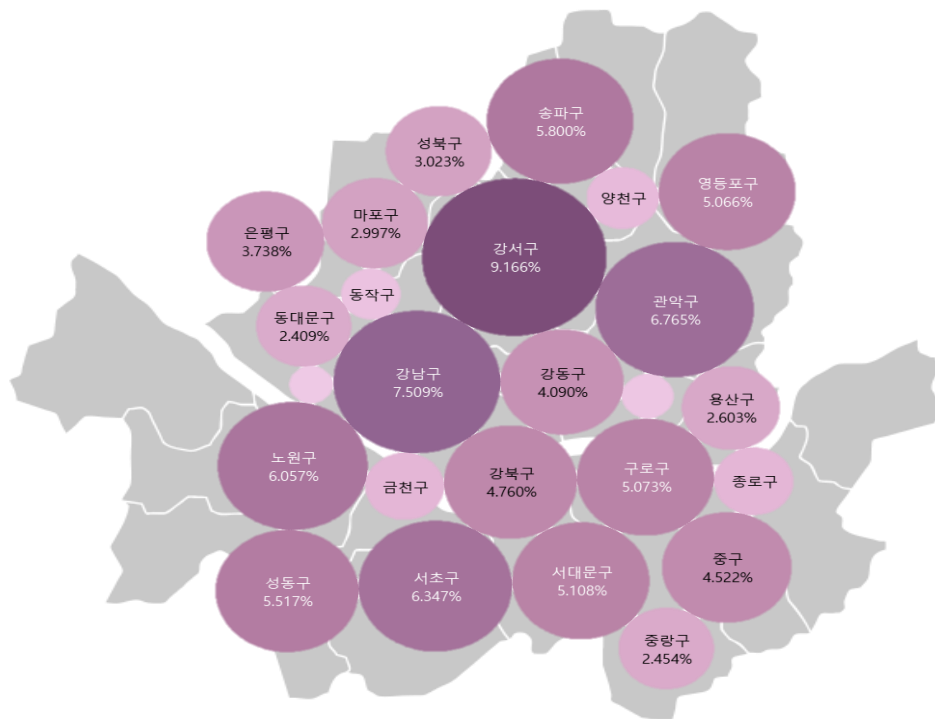




### 3. 분석 내용



서울시 구별 통화건수



#### TOP 3

1. 강서구
2. 강남구
3. 관악구





## 4. 모델링



data\$day	data\$call 1	2	Row Total
1	2714 8.633 0.654 0.163 0.113	1433 19.435 0.346 0.193 0.059	4147  0.172
2	2768 17.277 0.749 0.166 0.115	926 38.893 0.251 0.125 0.038	3694  0.153
3	2241 3.856 0.722 0.134 0.093	864 8.680 0.278 0.116 0.036	3105  0.129
4	2325 1.867 0.712 0.139 0.096	939 4.202 0.288 0.127 0.039	3264  0.135
5	2180 0.591 0.704 0.131 0.090	917 1.329 0.296 0.124 0.038	3097  0.128
6	2322 0.063 0.689 0.139 0.096	1049 0.142 0.311 0.141 0.043	3371  0.140
7	2149 22.646 0.625 0.129 0.089	1290 50.980 0.375 0.174 0.053	3439  0.143
Column Total	16699 0.692	7418 0.308	24117

## 결과분석

### 교차분석 - 요일별 통화건수

Statistics for All Table Factors

Pearson's Chi-squared test

---

Chi^2 = 178.5941      d.f. = 6      p = 6.74783e-36



## 4. 모델링



data\$age	data\$call		Row Total
	1	2	
10	2814	323	3137
	189.688	427.016	
	0.897	0.103	0.130
	0.169	0.044	
	0.117	0.013	
20	3154	1376	4530
	0.096	0.216	
	0.696	0.304	0.188
	0.189	0.185	
	0.131	0.057	
30	2809	1894	4703
	61.477	138.394	
	0.597	0.403	0.195
	0.168	0.255	
	0.116	0.079	
40	2732	2105	4837
	113.745	256.057	
	0.565	0.435	0.201
	0.164	0.284	
	0.113	0.087	
50	2472	1263	3735
	5.041	11.347	
	0.662	0.338	0.155
	0.148	0.170	
	0.103	0.052	
60	2718	457	3175
	122.798	276.437	
	0.856	0.144	0.132
	0.163	0.062	
	0.113	0.019	
Column Total	16699	7418	24117
	0.692	0.308	

## 결과분석

### 교차분석 - 연령별 통화건수

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 1602.311      d.f. = 5      p = 0



## 4. 모델링



### 결과분석

#### 교차분석 - 성별 통화건수

data\$gender	data\$call		Row Total
	1	2	
1	8155	3640	11795
	0.018	0.040	
	0.691	0.309	0.489
	0.488	0.491	
	0.338	0.151	
2	8544	3778	12322
	0.017	0.038	
	0.693	0.307	0.511
	0.512	0.509	
	0.354	0.157	
Column Total	16699	7418	24117
	0.692	0.308	

Statistics for All Table Factors

Pearson's Chi-squared test

-----  
chi^2 = 0.1131015      d.f. = 1      p = 0.7366404

Pearson's Chi-squared test with Yates' continuity correction

-----  
chi^2 = 0.103909      d.f. = 1      p = 0.7471882



## 4. 모델링



```
> summary(model)
```

Call:  
glm(formula = call ~ day + gender + age + gu, family = "binomial",  
data = data)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0388	-0.8256	-0.4585	0.9395	2.8803

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-2.41941	0.08686	-27.855	< 2e-16	***
day2	-0.62602	0.05595	-11.188	< 2e-16	***
day3	-0.44452	0.05792	-7.675	1.66e-14	***
day4	-0.36531	0.05684	-6.428	1.30e-10	***
day5	-0.32498	0.05746	-5.656	1.55e-08	***
day6	-0.22222	0.05568	-3.991	6.58e-05	***
day7	0.17715	0.05459	3.245	0.00118	**
gender2	-0.02676	0.03141	-0.852	0.39423	
age20	1.65318	0.07088	23.324	< 2e-16	***
age30	2.19007	0.07017	31.212	< 2e-16	***
age40	2.34037	0.06997	33.450	< 2e-16	***
age50	1.76883	0.07210	24.532	< 2e-16	***
age60	0.44179	0.08066	5.477	4.33e-08	***
gu2	0.07190	0.08650	0.831	0.40582	
gu3	0.94367	0.09033	10.447	< 2e-16	***
gu4	1.33433	0.07524	17.734	< 2e-16	***
gu5	1.84665	0.11187	16.507	< 2e-16	***
gu6	-16.71371	181.87041	-0.092	0.92678	
gu7	0.88924	0.08354	10.645	< 2e-16	***
gu8	-0.54680	0.11562	-4.729	2.25e-06	***
gu9	1.76561	0.09224	19.142	< 2e-16	***
gu10	-16.70404	215.55624	-0.077	0.93823	
gu11	-0.16037	0.09936	-1.614	0.10650	
gu12	-16.78017	161.14573	-0.104	0.91707	
gu13	-1.00701	0.10082	-9.988	< 2e-16	***
gu14	0.57258	0.07928	7.222	5.11e-13	***
gu15	0.51792	0.07781	6.656	2.81e-11	***
gu16	0.88975	0.09137	9.738	< 2e-16	***
gu17	-0.48741	0.09629	-5.062	4.15e-07	***
gu18	0.55645	0.08525	6.528	6.68e-11	***
gu19	-1.60880	0.15791	-10.188	< 2e-16	***
gu20	0.11599	0.08260	1.404	0.16026	
gu21	-1.08671	0.10622	-10.231	< 2e-16	***
gu22	0.37061	0.09634	3.847	0.00012	***
gu23	-0.54174	0.11481	-4.718	2.38e-06	***
gu24	-0.25024	0.08393	-2.982	0.00287	**
gu25	0.30804	0.10070	3.059	0.00222	**

--  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 29768 on 24116 degrees of freedom  
Residual deviance: 24210 on 24080 degrees of freedom  
AIC: 24284

결과분석

로지스틱 회귀분석



## 4. 모델링



```
> model3<-step(model,direction='backward')
```

```
Start:  AIC=24284.24
```

```
call ~ day + gender + age + gu
```

	Df	Deviance	AIC
- gender	1	24211	24283
<none>		24210	24284
- day	6	24479	24541
- age	5	26568	26632
- gu	24	27806	27832

```
Step:  AIC=24282.97
```

```
call ~ day + age + gu
```

	Df	Deviance	AIC
<none>		24211	24283
- day	6	24480	24540
- age	5	26568	26630
- gu	24	27809	27833

### 결과분석

로지스틱 회귀분석  
(backward)



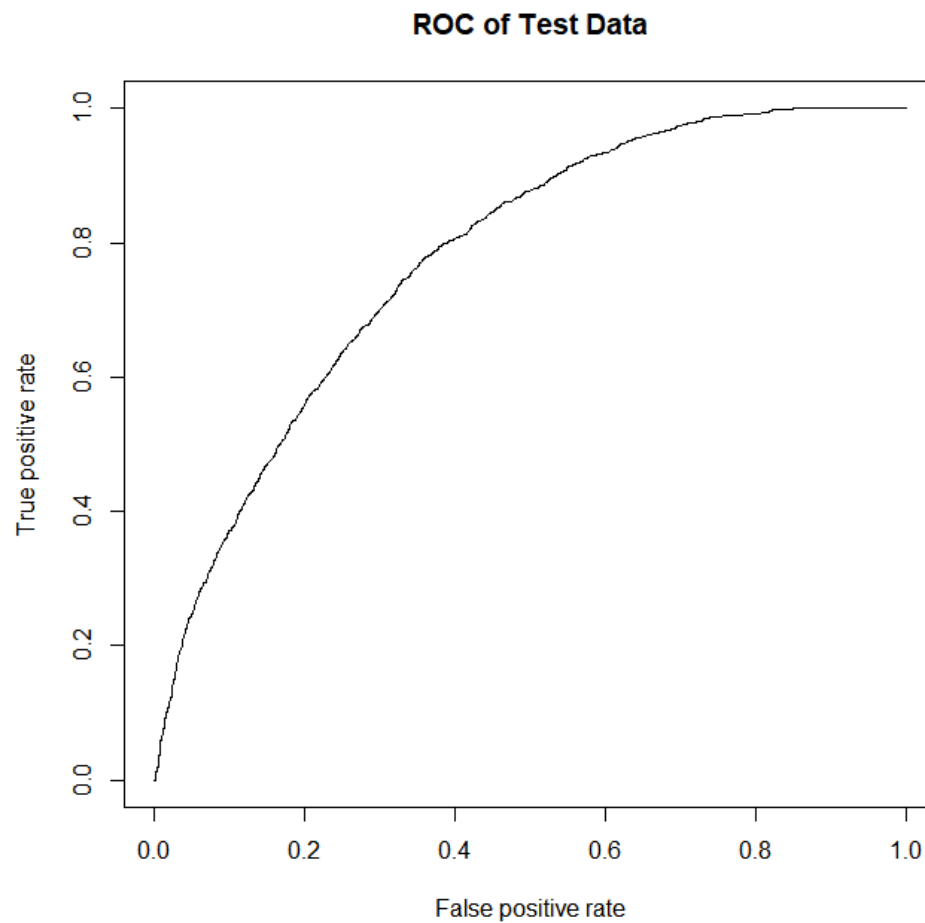
## 4. 모델링



### ROC 커브

로지스틱

```
> auc  
[1] 0.7784048
```



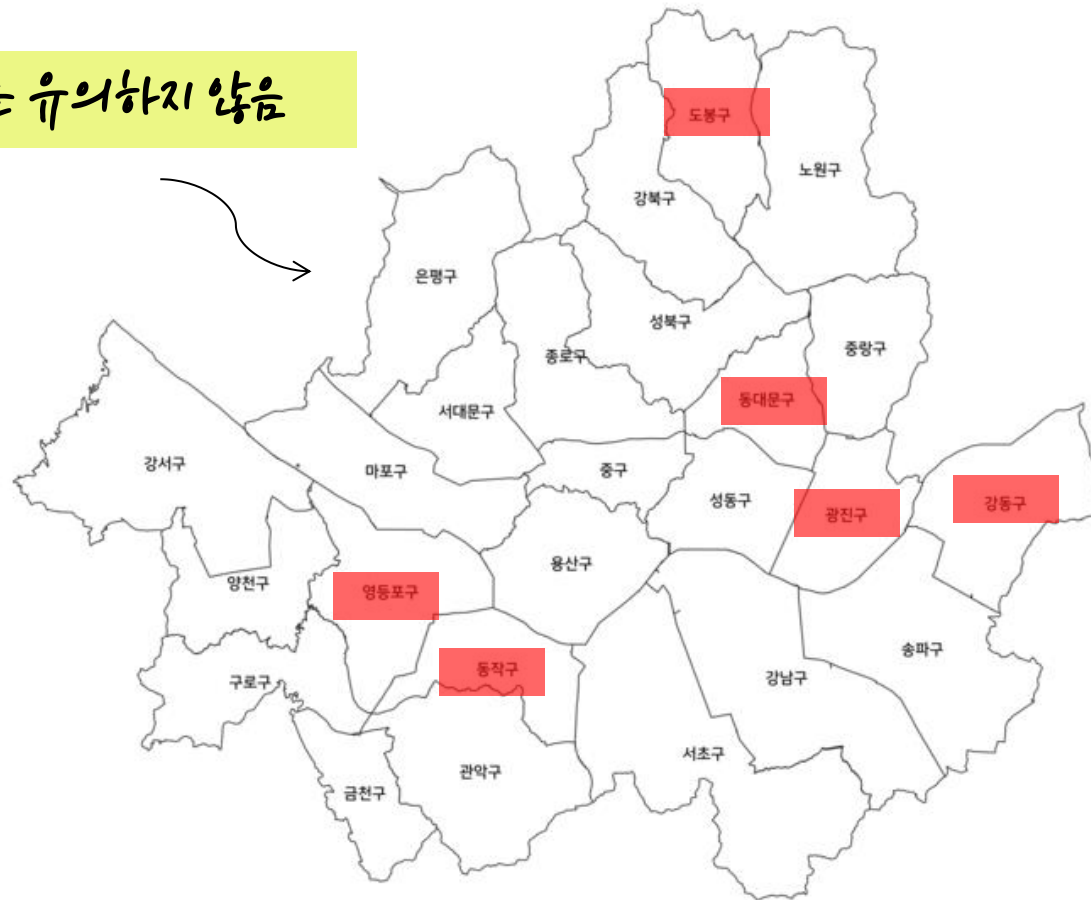


## 4. 모델링



색칠한 구는 유의하지 않음

- 도봉구
- 동대문구
- 강동구
- 광진구
- 동작구
- 영등포구







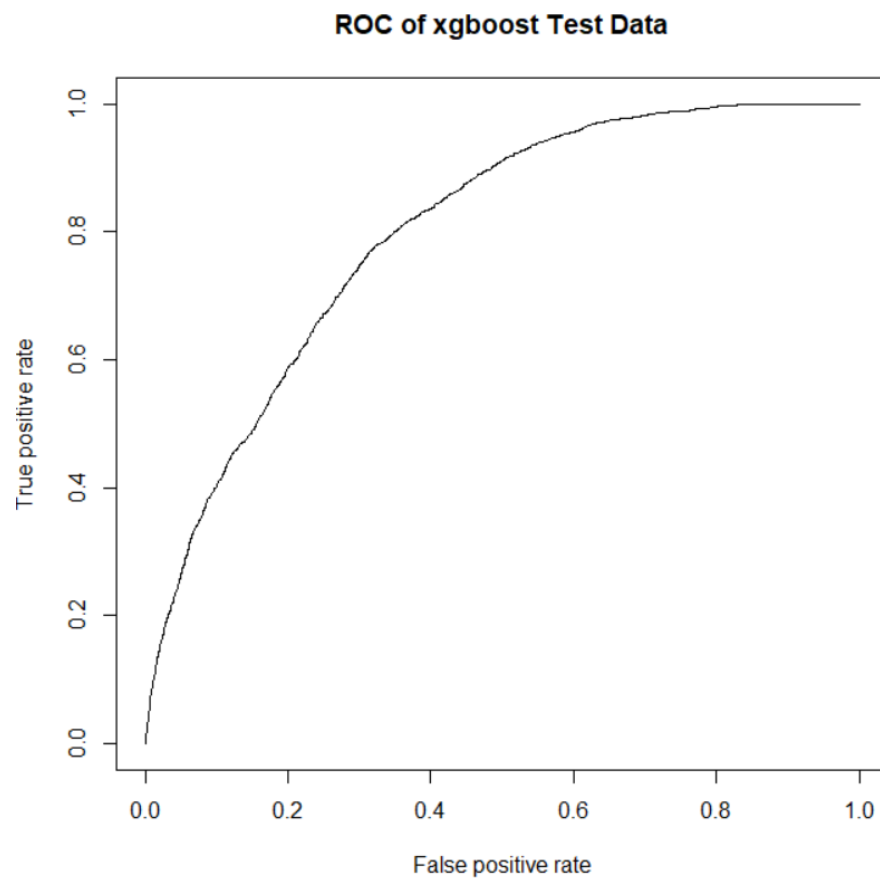
## 4. 모델링



ROC 커브

XGBoost

```
> AUC_xg  
[1] 0.7983607
```





## 5. 결론



### 한계점

- 데이터의 시간적 범위가 9월로 한정되어 있어 일반화하기 어려웠다.
- 파생변수 생성에 어려움이 있었고 외부 변수를 적용하는데 한계가 있었다.
- 시간이 없어서 의견공유를 많이 못했던 점이 아쉬웠다.

### 기대효과

- 기후데이터 (불쾌지수), 경제지표 등과 결합하여 치킨 자영업자의 매출에 영향을 주는 요인을 분석할 수 있을 것 같다.



오늘 저녁 치킨 ㄱ?

