

A/B Test 실험 주제 2

[List] 기본 정렬 필터 변경 실험

1-Pager 작성 - [List] 기본 정렬 필터 변경 실험

[List] 기본 정렬 필터 변경 실험 - 1 Pager	
Target User	첫 화면의 기본 정렬(=판매순)의 제휴점들이 마음에 들지않아 이탈하는 고객
Observation	<div> <div> <p>[List Page] Churn Rate</p> </div> <div> <p>[List Page] Map@K 정렬 평가</p> </div> <div> <p>[List Page] 정렬별 CTR 평가</p> </div> </div>
Problem Statement	첫화면의 기본정렬(=판매순)의 제휴점 노출이 고객을 이탈시켜, 리스트 페이지의 구매전환이 떨어진다.
Hypothesis	기존의 기본정렬을 다른 정렬 기준으로 바꾼다면 리스트 페이지의 구매전환이 상승할 것이다.

Target User

타겟 고객은 누구인가?

Observation

데이터적으로 관찰되는 현상은 무엇인가?

Problem Statement

현상으로써 발생하는 문제는 무엇인가?

Hypothesis

기능 개선을 위한 가설 설립

1-Pager 작성 - [List] 기본 정렬 필터 변경 실험

Experiments Group	<div> <div> <div>A – Feb 15, 2022</div> <div> <div>거친배달</div> <div>관매순</div> <div> <div> <div>거친샐롱(모란점)</div> <div>★ 4.2(6+)</div> <div>최소주문 17,000원, 배달비 3,500원</div> <div>평균 40~45분 소요</div> </div> <div> <div>거친샐롱(수내점)</div> <div>★ 4.9(50+)</div> <div>최소주문 20,000원, 배달비 2,500원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친치킨(관교점)</div> <div>★ 4.8(50+)</div> <div>최소주문 18,000원, 배달비 1,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친샐롱(분당점)</div> <div>★ 4.4(60+)</div> <div>최소주문 19,000원, 배달비 2,000원</div> <div>평균 40~45분 소요</div> </div> <div> <div>거친강점(태평점)</div> <div>★ 4.7(20+)</div> <div>최소주문 15,000원, 배달비 3,000원</div> <div>평균 30~35분 소요</div> </div> </div> </div> <div> <div>B – Feb 15, 2022</div> <div> <div>거친배달</div> <div>배달비 낮은순</div> <div> <div> <div>거친치킨(관교점)</div> <div>★ 4.8(50+)</div> <div>최소주문 18,000원, 배달비 1,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친샐롱(분당점)</div> <div>★ 4.4(60+)</div> <div>최소주문 19,000원, 배달비 2,000원</div> <div>평균 40~45분 소요</div> </div> <div> <div>거친파닭(수내점)</div> <div>★ 4.9(50+)</div> <div>최소주문 20,000원, 배달비 2,500원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친강점(태평점)</div> <div>★ 4.7(20+)</div> <div>최소주문 15,000원, 배달비 3,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친샐롱(모란점)</div> <div>★ 4.2(6+)</div> <div>최소주문 17,000원, 배달비 3,500원</div> <div>평균 40~45분 소요</div> </div> </div> </div> <div> <div>C – Feb 15, 2022</div> <div> <div>거친배달</div> <div>별점 높은순</div> <div> <div> <div>거친파닭(수내점)</div> <div>★ 4.9(50+)</div> <div>최소주문 20,000원, 배달비 2,500원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친치킨(관교점)</div> <div>★ 4.8(50+)</div> <div>최소주문 18,000원, 배달비 1,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친강점(태평점)</div> <div>★ 4.7(20+)</div> <div>최소주문 15,000원, 배달비 3,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친샐롱(분당점)</div> <div>★ 4.4(60+)</div> <div>최소주문 19,000원, 배달비 2,000원</div> <div>평균 40~45분 소요</div> </div> <div> <div>거친샐롱(모란점)</div> <div>★ 4.2(6+)</div> <div>최소주문 17,000원, 배달비 3,500원</div> <div>평균 40~45분 소요</div> </div> </div> </div> <div> <div>D – Feb 15, 2022</div> <div> <div>거친배달</div> <div>배달 빠른순</div> <div> <div> <div>거친강점(태평점)</div> <div>★ 4.7(20+)</div> <div>최소주문 15,000원, 배달비 3,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친치킨(관교점)</div> <div>★ 4.8(50+)</div> <div>최소주문 18,000원, 배달비 1,000원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친파닭(수내점)</div> <div>★ 4.9(50+)</div> <div>최소주문 20,000원, 배달비 2,500원</div> <div>평균 30~35분 소요</div> </div> <div> <div>거친샐롱(모란점)</div> <div>★ 4.2(6+)</div> <div>최소주문 17,000원, 배달비 3,500원</div> <div>평균 40~45분 소요</div> </div> <div> <div>거친샐롱(분당점)</div> <div>★ 4.4(60+)</div> <div>최소주문 19,000원, 배달비 2,000원</div> <div>평균 40~45분 소요</div> </div> </div> </div> </div> </div></div></div></div>				
	Input Metric	List 진입 유저 수 (= List UV)			
	Output Metric	MAP@10 (**MAP : Mean Average Precision)			
		구매 전환율 (= Order Complete UV / Main Home UV)			
		ARPU (= Average Revenue Per User)			
Experiment Period	Base Metric(기존 지표)	Expected Metric(예상 지표)	Alpha(유의수준)	1-Beta(검정력)	Minimum Sample Size
	20%	20.8% (4% ↑)	5%	80%	39780
Trade-Off	Period Calculation	Test Page UV	Required Days	# of Test Group	Period
		20,000	최소 2일(>=39,780)	4개(A,B,C,D)	최소 8일
Andon	추천순과 같이 플랫폼에서 제공하는 광고 비즈니스를 하기 어려운 환경이 될 수도 있다.				
	구매 전환율 및 매출의 유의미한 하락(3일 이상 지속)				

Experiments Group

실험을 위한 집단 구분

Metric

실험 성과를 평가할 수 있는 정량적 지표 설계

Experiments Group

실험을 위한 집단 구분

Trade-Off

기능 도입으로 인해 발생하는 문제 상황 및 예의주시가 필요한 부분

Andon

실험 진행 중 심각한 문제 발생 여부를 파악할 수 있는 기준 정의

* Mean Average Precision (MAP)은 정보 검색 시스템의 평가 모델 중 하나로, 검색 결과의 정확성과 순위를 평가하는 지표

결과 해석 - [List] 기본 정렬 필터 변경 실험

A/B Test Report Sheet

실험 주제	[List] 기본 정렬 필터 변경 실험
실험 기간	2023-01-16 14:34:10 ~ 2023-02-01 14:34:28
실험 위치	List
실험 목표	구매전환율 4% 상승
실험 분배	A 25% : B 25% : C 25% : D 25%
실험 분석가	거친코딩

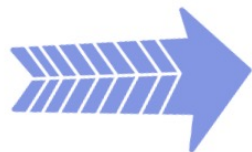
실험 검증에 필요한 지표

지표	Dimension	설명
MAP@10	-	뒷 장에서 상세히 설명
구매 전환율	Menu Category (빵, 한식, 치킨 등)	Buy CVR (= Unique Buyer User / Test User)
매출	Menu Category (빵, 한식, 치킨 등)	ARPU (= Revenue / Test User)

결과 해석 - MAP@K (= Mean Average Precision @ K) 설명



고객



클릭 여부

MAP@5 계산

$$MAP@5 = \frac{\left(\left(\frac{1}{1} \right) * 0 + \left(\frac{1}{2} \right) * 1 + \left(\frac{2}{3} \right) * 1 + \left(\frac{2}{4} \right) * 0 + \left(\frac{3}{5} \right) * 1 \right)}{3}$$

결과 해석 - [List] 기본 정렬 필터 변경 실험

1. MAP@10 (=Mean Average Precision@10)

C그룹에서 A그룹 대비 약 0.6% 더 높은 effect를 보임 (P-Value:0.000001)

D그룹에서 A그룹 대비 약 0.9% 더 낮은 effect를 보임 (P-Value:0.000001)

1. MAP@10 (Mean Average Precision@10)

Test Group	Test User Count	MAP Average	MAP Std	Lift % (vs.A)	P-Value
A	41,057	0.667	0.118	-	-
B	41,081	0.666	0.112	-0.15%	0.21290
C	41,049	0.671	0.127	0.60%	0.000001
D	41,052	0.661	0.121	-0.90%	0.000001

< Python Code Script >

```
clicks = [0,1,0,1,1,0,0,0,0,0]
```

```
def Calculate_MAP_10(clicks):
```

```
    mAP = 0
```

```
    flag = 1
```

```
    for idx, i in enumerate(clicks):
```

```
        if i == True:
```

```
            mAP += (flag/(idx))
```

```
            flag += 1
```

```
    mAP = mAP/(flag - 1)
```

```
    return mAP
```

$$\text{Average MAP@10} = \frac{\text{Each MAP@10}}{\text{Test User}}$$

$$\text{Lift \% (vs. A)} = \frac{(\text{Average MAP@10}_B - \text{Average MAP@10}_A)}{\text{Average MAP@10}_A}$$

$$t \text{ Statistic} = \frac{\overline{X_d}}{S_{\overline{X_B} - \overline{X_A}}}$$

* where, $\overline{X_d}$: Difference of Each Sample Average : $\overline{X_B} - \overline{X_A}$

* where, $S_{\overline{X_B} - \overline{X_A}}$: Pooled Standard Deviation of Each Sample Average Diff : $\frac{\sqrt{S_A^2}}{n_A} + \frac{\sqrt{S_B^2}}{n_B}$

* where, DF (= Degree of Freedom) : $n_A + n_B - 2$

P - Value_t = **TDIST**(t Statistic, DF, 2)

* where, TDIST: Excel Function

결과 해석 - [List] 기본 정렬 필터 변경 실험

2. 카테고리별 구매 전환율

Category	Test Group	Test User Count	Buyer User	Buy CVR	Lift % (vs.A)	P-Value
전채	A	41057	8317	20.26%	-	-
	B	41081	8119	19.76%	-2.44%	0.07695
	C	41049	8543	20.81%	2.74%	0.15352
	D	41052	8112	19.76%	-2.45%	0.07602
밥	A	41057	192	0.47%	-	-
	B	41081	183	0.45%	-4.74%	0.63729
	C	41049	197	0.48%	2.62%	0.64111
	D	41052	179	0.44%	-6.76%	0.5603
치킨	A	41057	3786	9.22%	-	-
	B	41081	3670	8.93%	-3.12%	0.1512
	C	41049	3891	9.48%	2.79%	0.19227
	D	41052	3671	8.94%	-3.03%	0.1575
햄버거	A	41057	698	1.70%	-	-
	B	41081	689	1.68%	-1.35%	0.79896
	C	41049	681	1.66%	-2.42%	0.80314
	D	41052	679	1.65%	-2.71%	0.68496
한식	A	41057	1219	2.97%	-	-
	B	41081	1188	2.89%	-2.60%	0.51191
	C	41049	1320	3.22%	8.31%	0.54293
	D	41052	1201	2.93%	-1.46%	0.59474
도시락	A	41057	178	0.43%	-	-
	B	41081	181	0.44%	1.63%	0.8783
	C	41049	173	0.42%	-2.79%	0.87608
	D	41052	183	0.45%	2.82%	0.8287
고기	A	41057	365	0.89%	-	-
	B	41081	358	0.87%	-1.98%	0.78764
	C	41049	371	0.90%	1.66%	0.79324
	D	41052	349	0.85%	-4.37%	0.64198
누들	A	41057	495	1.21%	-	-
	B	41081	487	1.19%	-1.67%	0.79021
	C	41049	499	1.22%	0.83%	0.79659
	D	41052	488	1.19%	-1.40%	0.80619
피자	A	41057	599	1.46%	-	-
	B	41081	586	1.43%	-2.23%	0.69615
	C	41049	613	1.49%	2.36%	0.70518
	D	41052	592	1.44%	-1.16%	0.7558
샐러드	A	41057	174	0.42%	-	-
	B	41081	168	0.41%	-3.50%	0.741
	C	41049	178	0.43%	2.32%	0.74453
	D	41052	171	0.42%	-1.71%	0.79493
짜개	A	41057	611	1.49%	-	-
	B	41081	609	1.48%	-0.39%	0.94581
	C	41049	620	1.51%	1.49%	0.95257
	D	41052	599	1.46%	-1.95%	0.80342

관측빈도	Buy	Not Buy	관측빈도	Buy	Not Buy	관측빈도	Buy	Not Buy
A	8,317	32,740	A	8,317	32,740	A	8,317	32,740
B	8,119	32,962	C	8,119	32,506	D	8,112	32,940

기대빈도	Buy	Not Buy	기대빈도	Buy	Not Buy	기대빈도	Buy	Not Buy
A	8,216	32,841	A	8,216	32,841	A	8,216	32,841
B	8,220	32,861	C	8,175	32,450	D	8,214	32,838

$$\text{Buy CVR} = \frac{\text{Buyer User}}{\text{Test User}}$$

$$\text{Lift \% (vs. A)} = \frac{(\text{Buy CVR}_B - \text{Buy CVR}_A)}{\text{Buy CVR}_A}$$

$$\text{Chi - Square Statistic} = \frac{(N * (BU_A * \text{Not } BU_B - \text{Not } BU_A * BU_B))}{(BU_A + \text{Not } BU_A) * (BU_B + \text{Not } BU_B) * (BU_A + BU_B) * (\text{Not } BU_A + \text{Not } BU_B)}$$

* where, BU_{group} : # of Bearch User(group)

* where, DF (= Degree of Freedom) : $(n \text{ rows} - 1) \times (m \text{ columns} - 1)$

$$P - \text{Value}_{chi} = \text{CHISQ.DIST.RT}(\text{Chi - Square Statistic}, DF)$$

* where, CHISQ.DIST.RT : Excel Function

$$P - \text{Value}_{chi} = \text{CHITEST}(\text{Search}_A: \text{Not Search}_B, \text{Ex Search}_A: \text{Not Ex Search}_B)$$

* where, CHITEST : Excel Function

* where, Element_k : Element_l is Drag Element_k to Element_l

* where, Ex means "Expected Frequency" : $\frac{\text{sum}(\text{row}) * \text{sum}(\text{col})}{\text{sum}(\text{total})}$

결과 해석 - [List] 기본 정렬 필터 변경 실험

2. 카테고리별 ARPU

Category	Test Group	Test User Count	Revenue	Average Revenue Per User	Revenue Std	Lift % (vs.A)	P-Value
전체	A	41057	7,867,398.726	191,621	12,845	-	-
	B	41081	7,834,410.353	190,706	13,434	-0.48%	0.00001
	C	41049	7,963,088.957	193,990	14,596	1.24%	0.00001
	D	41052	7,832,708.252	190,800	14,233	-0.43%	0.00001
밥	A	41057	558,084.531	13,593	12,508	-	-
	B	41081	554,825.930	13,506	11,349	-0.64%	0.29509
	C	41049	560,975.634	13,666	10,930	0.54%	0.37271
	D	41052	554,078.844	13,497	14,576	-0.71%	0.31160
치킨	A	41057	1,033,284.973	25,167	12,145	-	-
	B	41081	1,016,860.676	24,753	12,508	-1.65%	0.00001
	C	41049	1,055,661.634	25,717	13,942	2.19%	0.00001
	D	41052	1,031,341.052	25,123	13,576	-0.18%	0.62229
햄버거	A	41057	658,407.073	16,036	12,487	-	-
	B	41081	657,213.838	15,998	13,101	-0.24%	0.66709
	C	41049	656,239.523	15,987	13,754	-0.31%	0.58792
	D	41052	656,914.104	16,002	13,109	-0.21%	0.70012
한식	A	41057	857,242.428	20,879	11,533	-	-
	B	41081	852,936.691	20,762	12,191	-0.56%	0.15767
	C	41049	906,152.039	22,075	14,249	5.73%	0.00001
	D	41052	852,444.780	20,765	11,980	-0.55%	0.16363
도시락	A	41057	689,516.088	16,794	13,382	-	-
	B	41081	694,186.738	16,898	12,132	0.62%	0.24382
	C	41049	685,161.392	16,691	10,988	-0.61%	0.22893
	D	41052	690,026.380	16,809	11,354	0.09%	0.86727
고기	A	41057	932,828.958	22,720	13,537	-	-
	B	41081	925,513.849	22,529	11,091	-0.84%	0.02672
	C	41049	950,954.262	23,166	12,968	1.96%	0.00001
	D	41052	928,442.312	22,616	12,110	-0.46%	0.24559
누들	A	41057	622,697.833	15,167	12,192	-	-
	B	41081	620,395.479	15,102	13,964	-0.43%	0.47797
	C	41049	624,387.822	15,211	12,678	0.29%	0.61124
	D	41052	619,790.948	15,098	13,231	-0.45%	0.43737
피자	A	41057	1,128,070.452	27,476	12,627	-	-
	B	41081	1,125,459.349	27,396	12,134	-0.29%	0.35692
	C	41049	1,130,936.363	27,551	14,802	0.27%	0.43374
	D	41052	1,124,120.576	27,383	12,342	-0.34%	0.28657
샐러드	A	41057	643,806.234	15,681	14,478	-	-
	B	41081	642,096.030	15,630	13,499	-0.32%	0.60308
	C	41049	647,996.064	15,786	13,331	0.67%	0.27914
	D	41052	641,128.976	15,617	13,477	-0.40%	0.51667
찌개	A	41057	743,460.156	18,108	11,217	-	-
	B	41081	744,921.773	18,133	12,143	0.14%	0.75925
	C	41049	744,624.224	18,140	14,120	0.18%	0.72015
	D	41052	734,420.280	17,890	14,345	-1.20%	0.01529

3. 매출

B그룹(전체 카테고리)에서 **A그룹 대비 약 0.48% 더 낮은** effect를 보임 (P-Value:0.00001)
C그룹(전체 카테고리)에서 **A그룹 대비 약 1.24% 더 높은** effect를 보임 (P-Value:0.00001)
D그룹(전체 카테고리)에서 **A그룹 대비 약 0.56% 더 낮은** effect를 보임 (P-Value:0.00001)
B그룹(치킨 카테고리)에서 **A그룹 대비 약 1.65% 더 낮은** effect를 보임 (P-Value:0.00001)
C그룹(치킨 카테고리)에서 **A그룹 대비 약 2.19% 더 높은** effect를 보임 (P-Value:0.00001)
C그룹(한식 카테고리)에서 **A그룹 대비 약 5.73% 더 높은** effect를 보임 (P-Value:0.00001)
C그룹(고기 카테고리)에서 **A그룹 대비 약 0.84% 더 낮은** effect를 보임 (P-Value:0.02)
D그룹(고기 카테고리)에서 **A그룹 대비 약 1.96% 더 높은** effect를 보임 (P-Value:0.00001)
C그룹(찌개 카테고리)에서 **A그룹 대비 약 1.20% 더 낮은** effect를 보임 (P-Value:0.01)

$$APRU = \frac{Revenue}{Test\ User}$$

$$Lift\ \%(vs.\ A) = \frac{(APRU_B - APRU_A)}{APRU_A}$$

$$t\ Statistic = \frac{\overline{X_d}}{S_{\overline{X_B} - \overline{X_A}}}$$

- * where, $\overline{X_d}$: Difference of Each Sample Average : $\overline{X_B} - \overline{X_A}$
- * where, $S_{\overline{X_B} - \overline{X_A}}$: Pooled Standard Deviation of Each Sample Average Diff : $\frac{\sqrt{S_A^2}}{n_A} + \frac{\sqrt{S_B^2}}{n_B}$
- * where, DF(= Degree of Freedom) : $n_A + n_B - 2$

$$P - Value_t = TDIST(t\ Statistic, DF, 2)$$

* where, TDIST: Excel Function

결과 해석 - [List] 기본 정렬 필터 변경 실험

A/B Test Report Sheet	
실험 주제	[List] 기본 정렬 필터 변경 실험
실험 기간	2023-01-16 14:34:10 ~ 2023-02-01 14:34:28
실험 위치	List
실험 목표	구매전환율 4%상승
실험 분배	A 25% : B 25% : C 25% : D 25%
실험 분석가	거친코딩
결론	MAP@10(=Output Metric)이 Significant Postive -> 특정 정렬필터에 대한 검색 정렬 기능의 긍정 임팩트 확인 ARPU(=Monitoring Metric)이 Significant Postive이기 때문에 -> 특정 정렬필터에 대한 서비스 정량적 성장의 긍정 임팩트 확인

Summary

1. MAP@10 (=Mean Average Precision@10)

C그룹에서 **A그룹 대비 약 0.6% 더 높은** effect를 보임 (P-Value:0.000001)

D그룹에서 **A그룹 대비 약 0.9% 더 낮은** effect를 보임 (P-Value:0.000001)

2. 구매 전환율

유의미한 결과값이 없음

3. 매출

B그룹(전체 카테고리)에서 **A그룹 대비 약 0.48% 더 낮은** effect를 보임 (P-Value:0.00001)

C그룹(전체 카테고리)에서 **A그룹 대비 약 1.24% 더 높은** effect를 보임 (P-Value:0.00001)

D그룹(전체 카테고리)에서 **A그룹 대비 약 0.56% 더 낮은** effect를 보임 (P-Value:0.00001)

B그룹(치킨 카테고리)에서 **A그룹 대비 약 1.65% 더 낮은** effect를 보임 (P-Value:0.00001)

C그룹(치킨 카테고리)에서 **A그룹 대비 약 2.19% 더 높은** effect를 보임 (P-Value:0.00001)

C그룹(한식 카테고리)에서 **A그룹 대비 약 5.73% 더 높은** effect를 보임 (P-Value:0.00001)

C그룹(고기 카테고리)에서 **A그룹 대비 약 0.84% 더 낮은** effect를 보임 (P-Value:0.02)

D그룹(고기 카테고리)에서 **A그룹 대비 약 1.96% 더 높은** effect를 보임 (P-Value:0.00001)

C그룹(짜개 카테고리)에서 **A그룹 대비 약 1.20% 더 낮은** effect를 보임 (P-Value:0.01)