

iFood Customer Insights: EDA for Data-Driven Marketing Optimization

NO.1 DATA

박건, 김정민, 조혜진

NO.1 DATA

서론	3
1. iFood 데이터 셋 개요	3
본론	5
2. EDA for iFood	5
2.1 주요 변수 기본 통계	5
2.2 학력 수준 관련 분석	8
2.3 고객 분석	11
2.3.1. 가족 구조별 제품 소비형태	11
2.3.2. 가족 구성원별 제품 소비 형태	12
2.3.3. 연령별 제품 소비 형태	13
2.3.4. 수입구간별 제품 소비 형태	14
2.3.5. 제품간 상관관계 분석	15
2.3.6. 학력 수준과 관련된 지표와 상품 카테고리 간의 상관관계 분석	15
2.4 채널별 매출 분석	15
2.4.1 카테고리별 지출금액과 채널 구매 빈도 간의 상관관계 분석	15
2.4.2 채널별 매출 기여도 추정	16
2.4.3. 채널별 고객 이용률 분석	17
2.4.4. 웹사이트 방문 빈도와 구매 전환율 분석	18
2.5 캠페인 참여 여부 관련분석	18
2.5.1 나이에 따른 캠페인 참여 의사	18
2.5.2 소득과 학력에 따른 캠페인 참여	19
2.5.3 자녀 보유에 따른 캠페인 참여	20
3. 결론 및 기대효과	20
3.1 마케팅 전략	20
3.1.1. 가족 구성원별 맞춤형 마케팅 전략	20
3.1.2. 가족 구조별 맞춤형 마케팅 전략	22
3.1.3. 수입 구간별 맞춤형 마케팅 전략	23
3.1.4. 연령별 맞춤형 마케팅 전략	24
3.1.5. 마케팅 채널별 맞춤형 마케팅 전략	25
3.1.6.. 마케팅 전략 제안	25
3.2 마케팅 기대효과	25
REFERENCE	26

iFood Customer Insights:

EDA for Data-Driven Marketing Optimization

서론

본 EDA 프로젝트의 목적은 고객 프로필, 제품 선호도, 캠페인 성과, 채널 성과 등의 데이터를 활용하여 마케팅 전략을 최적화하는 것이다. 데이터를 깊이 있게 탐색하고 인사이트를 도출함으로써, 효과적인 고객 타겟팅과 맞춤형 마케팅 전략 수립을 위한 기초를 마련하는 것이다.

1. iFood 데이터 셋 개요

- 도메인 소개:
 - **iFood:** 브라질의 주요 음식 배달 플랫폼으로, 콜롬비아와 멕시코에도 진출했으나, 2020년 이후로는 멕시코에서의 서비스는 종료했다. 2023년 매출액은 약 14억 달러에 달하며, 브라질 내에서 80% 이상의 시장 점유율을 유지하고 있다.
 - 브라질 음식 배달 서비스 흐름: 브라질의 인터넷 보급률은 2014년에 약 40%[1], 2019년에 약 72%, 2020년에 약 80%였다[2]. 배달 서비스 이용자는 2019년에는 인터넷 이용자의 약 32%, 2020년에는 약 54%였다. 배달 서비스 이용자가 시간이 지나면서 늘어난다고 전제하면 2014년의 배달 서비스 이용자는 인터넷 이용자의 32% 미만이었을 것이다. 따라서, 배달 서비스 이용자는 2014년에는 많아도 10% 미만, 2019년에는 약 22%, 2020년은 약 40%였을 것으로 추정된다[3].
- 데이터 소개:
 - 데이터셋[4]: 2018년 9월부터 2020년 8월까지 수집된 2,205명의 고객 데이터다. 공식 문서에는 랜덤 추출한 데이터셋이라고 소개되었다. 그러나 확인 결과, 이 데이터셋에는 가입일로부터 최소 6년, 최대 8년 경과한 고객에 대한 데이터만 있다. 따라서 학력자 비중, 평균 연령 등 각종 지표를 볼 때 2014년 기준임을 고려해야 한다.

NO.1 DATA

○ 변수

속성 (Feature)	설명 (Description)	데이터 타입
Income	고객의 연간 가계 소득	연속형(Continuous)
Kidhome	고객의 가정에 있는 어린 자녀 수	이산형(Discrete)
Teenhome	고객의 가정에 있는 청소년 자녀 수	이산형(Discrete)
Recency	마지막 구매 후 지난 일수	연속형(Continuous)
MntWines	지난 2년 동안 와인에 지출한 금액	연속형(Continuous)
MntFruits	지난 2년 동안 과일에 지출한 금액	연속형(Continuous)
MntMeatProducts	지난 2년 동안 육류 제품에 지출한 금액	연속형(Continuous)
MntFishProducts	지난 2년 동안 생선 제품에 지출한 금액	연속형(Continuous)
MntSweetProducts	지난 2년 동안 과자 제품에 지출한 금액	연속형(Continuous)
MntGoldProds	지난 2년 동안 고가 제품에 지출한 금액	연속형(Continuous)
NumDealsPurchases	할인 구매 횟수	이산형(Discrete)
NumWebPurchases	회사의 웹사이트를 통해 구매한 횟수	이산형(Discrete)
NumCatalogPurchases	카탈로그를 통해 구매한 횟수	이산형(Discrete)
NumStorePurchases	매장에서 직접 구매한 횟수	이산형(Discrete)
NumWebVisitsMonth	지난 한 달 동안 회사의 웹사이트를 방문한 횟수	이산형(Discrete)
AcceptedCmp3	1: 고객이 세 번째 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
AcceptedCmp4	1: 고객이 네 번째 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
AcceptedCmp5	1: 고객이 다섯 번째 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
AcceptedCmp1	1: 고객이 첫 번째 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
AcceptedCmp2	1: 고객이 두 번째 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
Complain	1: 고객이 지난 2년 동안 불만을 제기한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
Z_CostContact	고정된 비용 관련 정보	연속형(Continuous)
Z_Revenue	수익 관련 정보	연속형(Continuous)
Response	1: 고객이 가장 최근의 캠페인에서 제안을 수락한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
Age	고객의 나이	연속형(Continuous)
Customer_Days	고객의 회사 가입 후 경과된 일수	연속형(Continuous)
marital_Divorced	1: 고객이 이혼한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
marital_Married	1: 고객이 기혼인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
marital_Single	1: 고객이 미혼인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
marital_Together	1: 고객이 동거 중인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
marital_Widow	1: 고객이 과부/홀아비인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
education_2n Cycle	1: 교육 수준이 "2nd Cycle"인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
education_Basic	1: 고객의 교육 수준이 "Basic"인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
education_Graduation	1: 고객의 교육 수준이 "Graduation"인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
education_Master	1: 고객의 교육 수준이 "Master"인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
education_PhD	1: 고객의 교육 수준이 "PhD"인 경우, 0: 그렇지 않은 경우	범주형(Categorical)
MntTotal	총 지출 금액 (지난 2년 동안 모든 제품에 대해 지출한 금액의 합계)	연속형(Continuous)
MntRegularProds	MntGoldProds를 제외한 총 지출 금액	연속형(Continuous)
AcceptedCmpOverall	모든 캠페인에 대한 제안을 수락한 횟수 (0~5)	이산형(Discrete)

소득, 나이, 학력, 자녀 유무, 기혼 여부, 구매 채널, 제품 카테고리별 지출액 등 39개 속성으로 구성되어 있다. 학력 사항, 기혼 여부, 캠페인 참여 여부 등의 사항은 바이너리 타입의 속성 여러 개로 구성되어 있다.

NO.1 DATA

- 목적: iFood의 6번째 마케팅 캠페인에 대한 고객 수요를 예측하고, 데이터 기반으로 맞춤형 마케팅 전략을 수립한다.
- 데이터 정제

속성(Feature)	설명(Description)	데이터 타입
income	고객의 연간 가계 소득	연속형(Continuous)
Recency	마지막 구매 후 지난 일수	연속형(Continuous)
Age	나이	연속형(Continuous)
Children	고객의 자녀 정보 ex) {"Teen": 1, "Kid": 2}	이산형(Discrete)
Education	최종 학력	범주형(Categorical)
MaritalStatus	기혼 여부 (미혼, 동거, 기혼, 이혼, 사별 中 1)	범주형(Categorical)
Customer_Days	가입일로부터 지난 일수	연속형(Continuous)
complain	1: 고객이 지난 2년 간 불만을 제기한 경우, 0: 그렇지 않은 경우	범주형(Categorical)
NumPurchases	구매 채널 별 이용 횟수 ex) {"Deals": 3, "Web": 8, "Catalog": 10, "Store": 4}	이산형(Discrete)
NumWebVisitsMonth	지난 한 달 동안 회사의 웹사이트를 방문한 횟수	이산형(Discrete)
MntCategories	상품 카테고리 별 지출액 ex) {"Fish": 172, "Fruits": 88, "Meat": 546, "Sweet": 88, "Wines": 635}	연속형(Continuous)
MntTotal	지난 2년 간의 모든 상품에 대한 지출액 합계	연속형(Continuous)
MntRegularProds	지난 2년 간 일반 제품에 지출한 금액	연속형(Continuous)
MntGoldProds	지난 2년 간 고가 제품에 지출한 금액	연속형(Continuous)
AcceptedCmp	각 캠페인 별 참여 여부 ex) {"1st": 0, "2nd": 0, "3rd": 0, "4th": 0, "5th": 0, "last": 1}	범주형(Categorical)
AcceptedCmpOverall	모든 캠페인에 대한 제안을 수락한 횟수 (0~5)	이산형(Discrete)

Z_Revenue, Z_costcontact의 두 속성은 값이 모두 똑같은 필요없는 속성으로 판단되어서 삭제했다. 결측치는 한 개도 존재하지 않았다. 컬럼의 수가 너무 많아 직관적이지 않기 때문에 바이너리 타입의 여러 개의 컬럼으로 저장되어 있던 속성들을 문자열 타입의 한 개 컬럼으로 병합했다. 가령, 고객의 기혼 여부를 나타내던 marital_divorced, marital_married, marital_single 등의 속성을 marital_status라는 하나의 속성으로 병합하고, 그 값은 single, together, married, divorced, widow의 다섯 개 범주를 갖는다. 또한, 비슷한 의미를 갖는 여러 개의 컬럼을 Json 타입의 한 개 속성으로 병합했다. 대표적으로 상품의 카테고리에 대한 총 지출액을 나타내는 MntFishProducts, MntMeatProducts 등의 컬럼들을 MntCategories라는 하나의 속성으로 병합하며, 각 카테고리를 key, 해당 카테고리에 대한 지출액을 value로 설정했다. 그 외 Children, NumPurchases, AcceptedCmp 컬럼들도 병합의 결과다. 데이터셋의 총 컬럼 수는 39개에서 16개로 감소했다. 반면, 용량은 약 3배 증가했다. 그 원인은 여러 개의 이진 데이터 타입 컬럼을 한 개의 문자열 데이터 타입 컬럼으로 바꿨기 때문으로 추측된다. 데이터셋의 크기가 커질수록 이 문제는 더 심각해질 것이다. 데이터셋을 저장할 때, 데이터셋의 규모를 고려해서 데이터 타입 변환을 고려해야 한다는 것을 알았다. 데이터 변환 sql 코드는 마지막 장 참조[5].

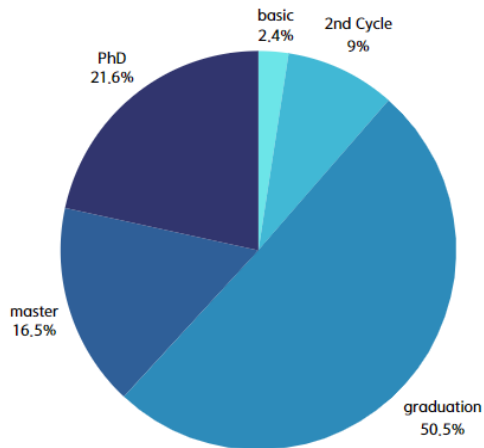
NO.1 DATA

본론

2. EDA for iFood

2.1 주요 변수 기본 통계

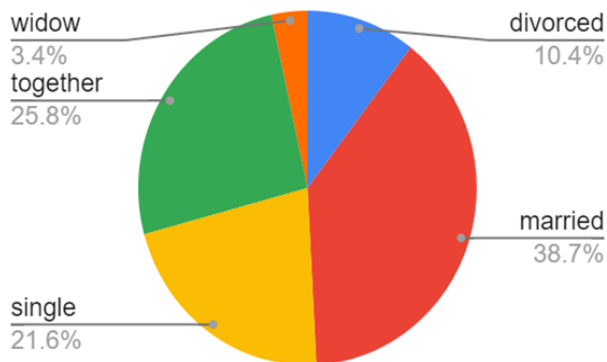
- 주요 변수 선정 기준: 주요 변수는 총 6가지를 꼽았다. 학력, 기혼 여부, 자녀, 연간 가계소득, 나이는 본 서비스 이용으로부터 독립적인 변수이기 때문이다. 총 지출액은 매출 잠재력과 가장 관련 깊은 변수라 판단되어 선택했다.
- 학력 수준



통계: 학사 학위 51%, 박사 학위 22%, 석사 학위 17%

분석: iFood의 고객 중 절반 이상이 대학 졸업 이상의 학력을 보유하고 있다. 이는 인터넷 보급률, 음식 배달 서비스 이용자 수가 매우 적은 2014년 기준 사용자 데이터이기 때문이다.

- 기혼 여부

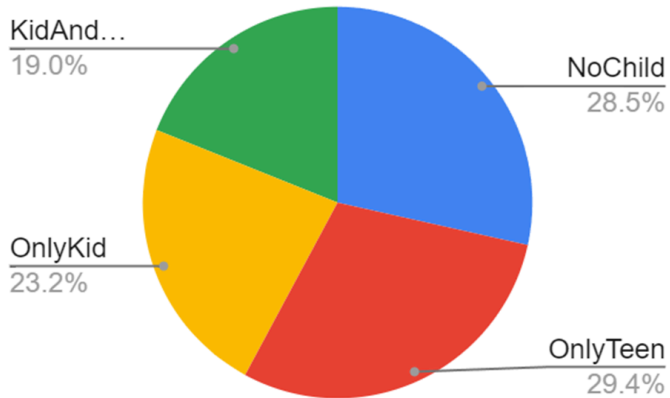


통계: 기혼 38.7%, 동거 25.8%, 독신 21.6%.

분석: 실질적인 single 비율은 single에 divorced, widow를 합한 약 35% 가량을 차지한다. 브라질의 문화적 특성상 동거 비율이 높은 편이다.

NO.1 DATA

- 자녀 유무

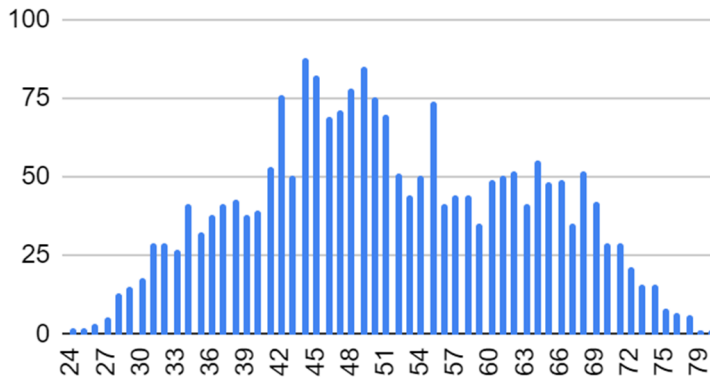


통계: 자녀 없음 28.5%, 10대 자녀만 있음 29.4%, 10세 이하 자녀만 있음 23.2%.

분석: 자녀를 보유한 가정이 약 70%를 차지한다. 자녀 없음, 10대 자녀만 있음, 10세 이하 자녀만 있음, 10대, 10세 이하 자녀가 모두 있음의 네 가지로 구분했을 때, 각 집단이 차지하는 비율이 크게 편향되어있지는 않다.

- 나이

나이 분포도

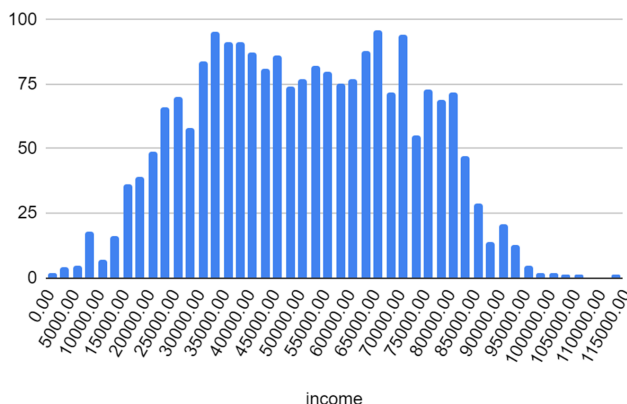


통계: 최솟값 24세, 최댓값 80세, 평균 51세, 중위값 50세.

분석: 주 연령층은 40대와 50대로, 나이가 적어지거나 많아질수록 점점 인구수가 적어진다.

- 가계 소득(단위: BRL)

가계소득



통계: 최솟값 1,730 헤알, 최댓값 113,734 헤알, 평균 51,622 헤알, 중위값 51,277 헤알.

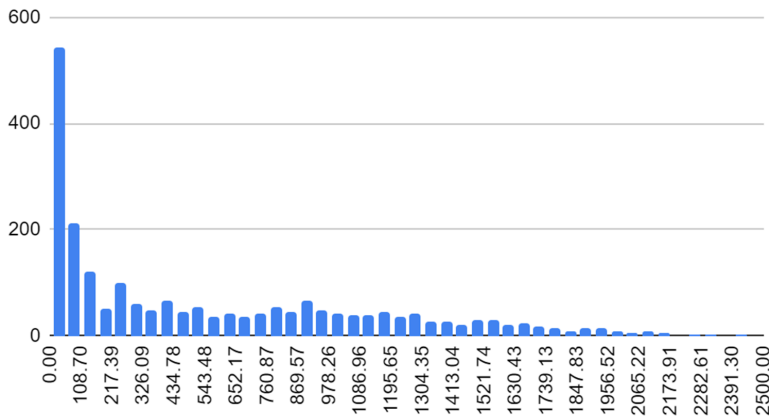
분석: 2020년 기준 브라질의 평균 가계 소득은 약 3만 5천 헤알이며[6], iFood 고객의 평균 소득은 이보다 높은 5만 헤알 수준이다. 데이터셋에 고학력자가

NO.1 DATA

많은 것과 상관성이 있는 것으로 추측된다.

- 지출액

총 지출액



통계: 최솟값 4 헤알, 최댓값 2,491 헤알, 평균 563 헤알, 중위값 343 헤알.

분석: 2년간의 서비스 이용 지출액을 분석한 결과, 지출 규모가 클수록 고객 수는 줄어드는 경향이 나타났다.

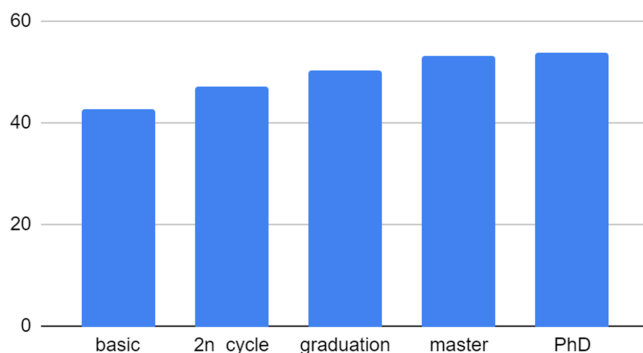
2.2 학력 수준 관련 분석

- 학력 수준

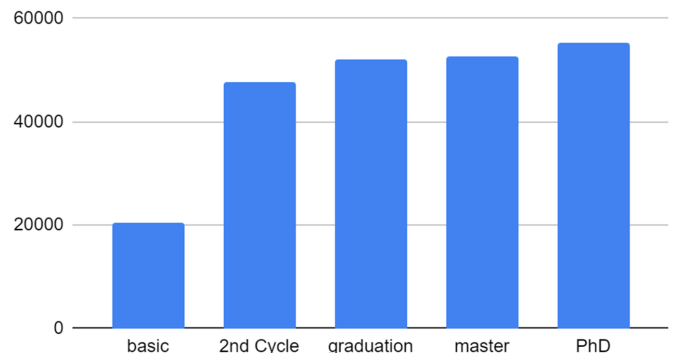
- 개요: 학력 분포를 살펴보면, 약 절반인 50.5%가 학사 학위를 소지하고 있으며, 석사 학위 소지자는 16.5%, 박사 학위 소지자는 21.6%로 나타났다. 특히, 박사 학위 소지자의 비율이 21.6%로 상대적으로 높게 나타난 점은 주목할 만하다. 이는 브라질의 인터넷 보급률이 50%에 미치지 못하던 시절의 가입자들이므로, 비교적 고학력자가 많은 것으로 해석될 수 있다. 학력 수준이 높을수록 소비 성향에 영향을 미칠 가능성이 있으며, 이는 마케팅 전략 수립 시 중요한 요소로 작용할 수 있다.

- 나이, 소득, 지출액

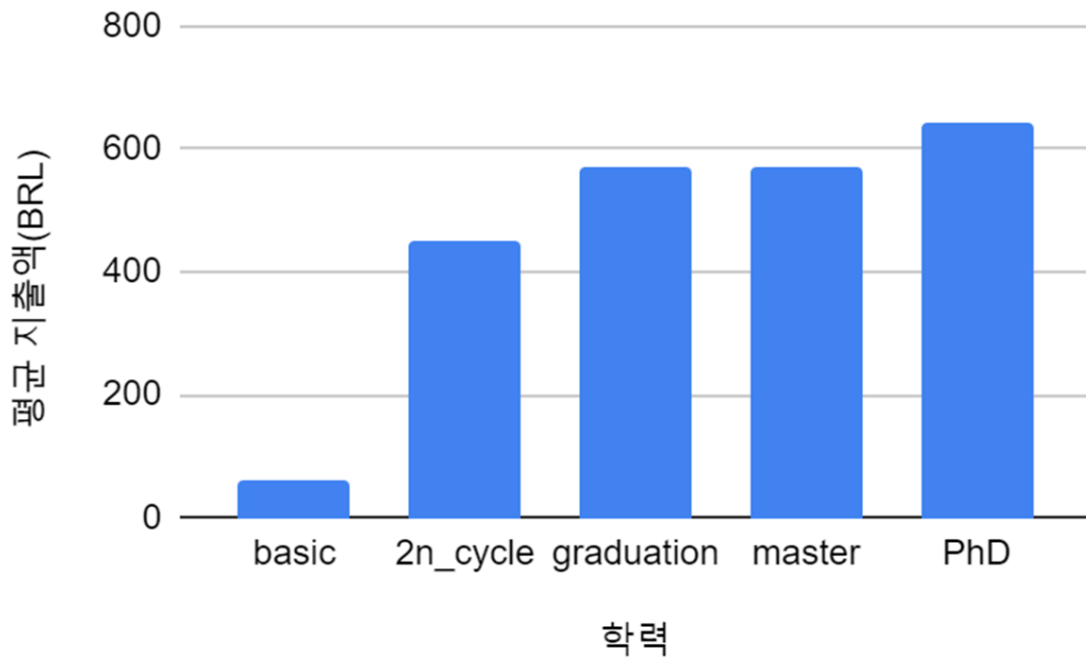
학력 수준 별 평균 연령



학력 수준 별 가계소득 평균



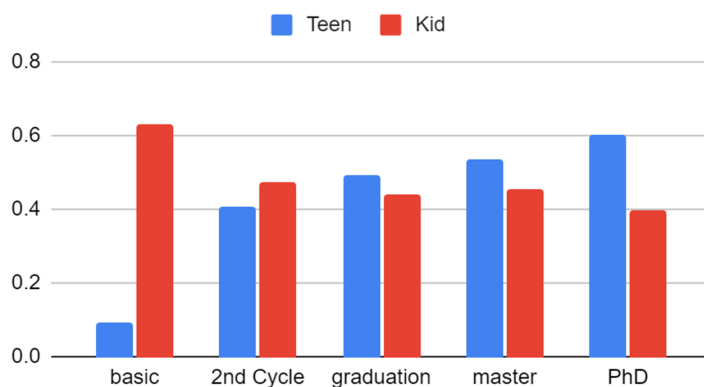
NO.1 DATA



학력 수준이 높을수록 나이, 소득, 지출액이 모두 비례하는 양상을 보인다.

○ 자녀 수

학력 수준 별 평균 자녀 수

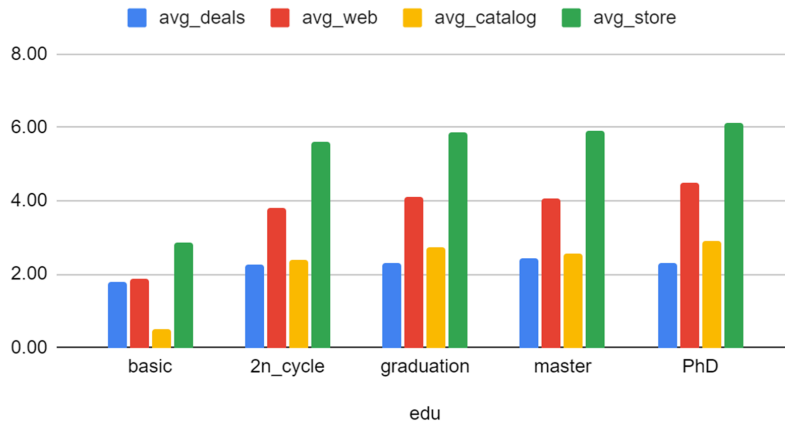


학력 수준이 높아질수록 10대 자녀가 많아지며, 반대로 10세 이하 자녀의 수는 줄어든다. 10세 이상의 자녀가 많을수록 나이가 많을 것이라고 추측되고, 학력 수준과 나이는 비례하는 경향이 있다는 점을 고려할 만하다.

NO.1 DATA

○ 구매 채널 선호

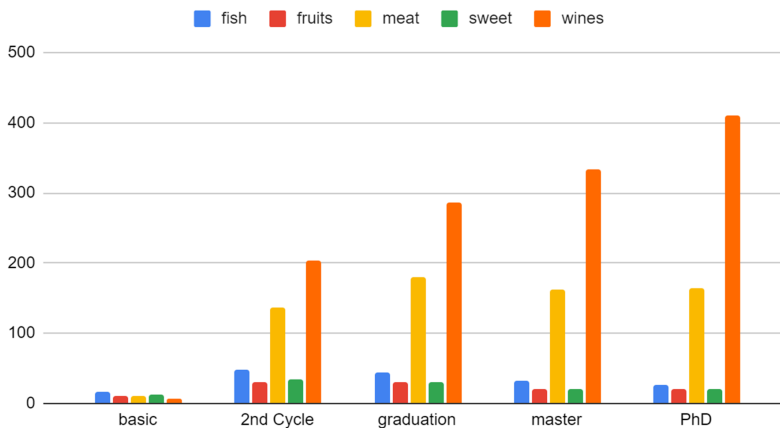
학력 수준 별 구매 경로에 따른 평균 구매 횟수



학력 수준에 따른 구매 채널 선호도에는 큰 차이가 없으며, 전반적으로 매장 구매를 가장 선호한다.

● 카테고리별 지출액

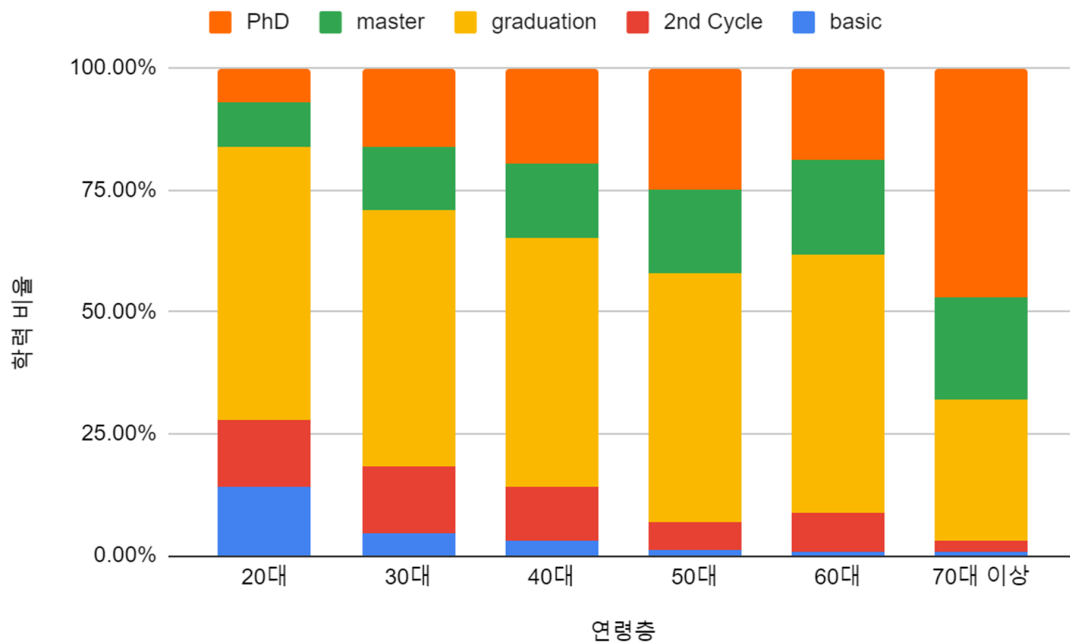
각 카테고리 별 학력 수준에 따른 2년 간 평균 지출액



- 생선, 과일, 디저트: 기초교육 학력자를 제외한 모든 학력자 간에 별 차이는 없으며, 학력이 높아질수록 지출액이 미미하게 감소하는 양상을 띤다.
- 와인: 학력 수준이 높아질수록 와인에 대한 지출이 증가하는 양상이 나타난다.
- 육류: 학사, 석사, 박사 학력에서는 지출액이 비슷하고, 중등 교육 학력자에게서 지출액이 약간 낮게 나타난다.

NO.1 DATA

- 연령층 별 학력 수준 분포



연령층이 높아질수록 고학력자 비율이 높아지는 양상을 보인다. 반면 기초 교육 학력자, 중등 교육 학력자는 연령층이 높아질수록 줄어드는 모습이다. 20대에는 석,박사 학위자가 비교적 적다. 눈여겨볼 것은 70대 이상 연령층에서 약 50%에 달하는 인원이 박사 학위자라는 것이다.

2.3 고객 분석

2.3.1. 가족 구조별 제품 소비형태:

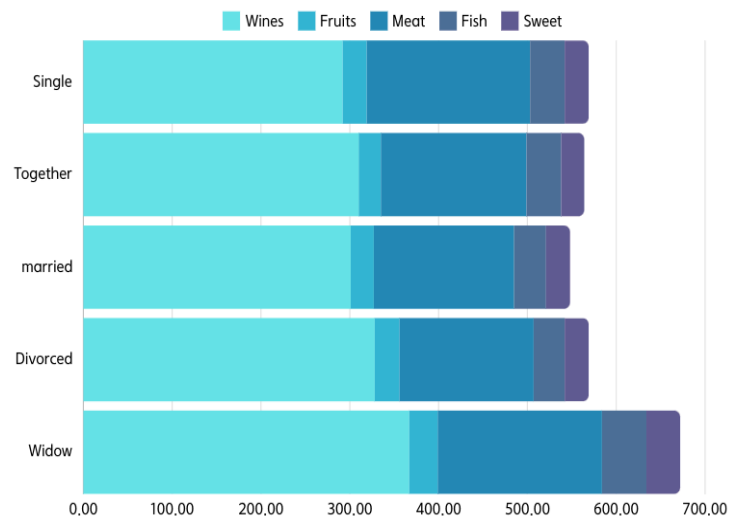
Marital_Status	Wines	Fruits	Meat	Fish	Sweet
Single	292	27	184	39	27
Together	310	25	164	39	26
married	301	26	158	36	27
Divorced	328	28	151	35	27
Widow	367	32	185	50	38

- 와인 소비(**Wines**): Widow 그룹이 가장 높은 평균 소비(367)를 보이며, 그 다음으로 Divorced(328), Together(310), Married(301), Single(292) 순이다.
- 과일 소비(**Fruits**): Widow 그룹이 평균 소비(32)로 가장 높음. 그 다음으로 Divorced(28), Single(27), Married(26), Together(25) 순이다.

NO.1 DATA

- 고기 소비(**Meat**): Single(184)과 Widow(185) 그룹이 가장 높고, Divorced(151) 그룹이 가장 낮다.
- 생선 소비(**Fish**): Widow 그룹이 가장 높은 소비(50)를 보이며, 다른 그룹들은 비슷한 수준(35-39)이다.
- 단맛(**Sweet**) 소비: Widow 그룹이 평균 소비(38)로 가장 높고, 다른 그룹들은 모두 비슷한 수준(26-27)이다.

소비 패턴의 주요 특징



- **Widow** 그룹: 모든 항목에서 가장 높은 소비를 보인다. 특히 와인, 과일, 생선, 단맛의 소비가 두드러진다.
- **Single** 그룹: 고기 소비가 높으며, 와인 소비도 낮지 않다. 그러나 과일과 생선, 단맛 소비는 다른 그룹에 비해 낮은 편이다.
- **Together, Married, Divorced** 그룹: 대체로 비슷한 패턴을 보인다. 이들 그룹은 고기 소비가 낮은 편이고, 과일과 생선, 단맛 소비도 Widow에 비해 낮다. 와인 소비는 상대적으로 높다.

2.3.2. 가족 구성원별 제품 소비 형태:

	Wines	Fruits	Meat	Fish	Sweet	Gold Product
Kid & Teen home	-0.25	-0.24	-0.26	-0.25	-0.23	-0.20
No Kids & Teens	0.35	0.41	0.59	0.45	0.40	0.25
Only Kidhome	-0.36	-0.23	-0.29	-0.23	-0.24	-0.24
Only Teenhome	0.21	0.01	-0.08	-0.01	0.02	0.15

Kid & Teen home (아이와 청소년이 모두 있는 가정): 모든 항목에서 음수의 상관계수를 보인다. 즉, 아이와 청소년이 있는 가정에서는 와인, 과일, 고기, 생선, 단맛, 금 상품 등의

NO.1 DATA

소비가 전반적으로 적은 편이다. 특히, 고기(-0.2622479)와 와인(-0.2544055) 소비가 가장 적다.

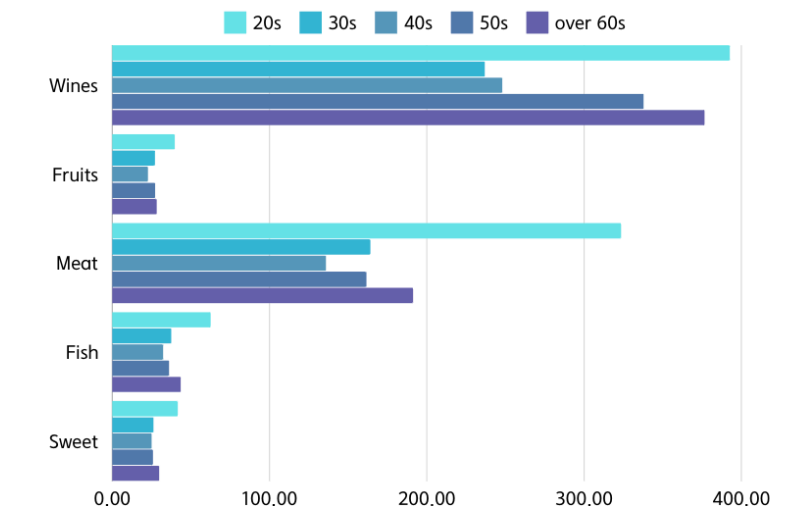
No Kids & Teens (아이와 청소년이 없는 가정): 모든 항목에서 양의 상관계수를 보인다. 아이와 청소년이 없는 가정에서는 모든 항목의 소비가 높다. 특히, 고기(0.58624929)와 생선(0.45215732)의 소비가 두드러진다. 와인(0.34505459)과 과일(0.41297335)의 소비도 비교적 높은 편이다.

Only Kidhome (아이만 있는 가정): 모든 항목이 음수의 상관계수를 가진다. 즉, 아이만 있는 가정에서도 모든 항목의 소비가 전반적으로 적다. 와인(-0.3648032)의 소비 감소가 가장 크다. 다른 항목들은 비교적 덜 감소한다.

Only Teenhome (청소년만 있는 가정): 와인(0.21494156)과 금 상품(0.15061035)은 양의 상관계수를 보이나, 나머지 항목들은 상관계수가 거의 0에 가깝다. 이는 청소년만 있는 가정에서 와인과 금 상품의 소비가 약간 증가할 수 있으나, 다른 항목들은 큰 변화가 없음을 의미한다.

2.3.3. 연령별 제품 소비 형태 :

Age_Group	Wines	Fruits	Meat	Fish	Sweet
20대	392.65	39.84	323.53	62.67	41.67
30대	237.01	27.21	164.14	37.66	26.36
40대	248.11	22.79	136.03	32.34	25.03
50대	337.97	27.35	161.63	36.11	25.90
60대 이상	376.74	28.29	191.29	43.65	29.98

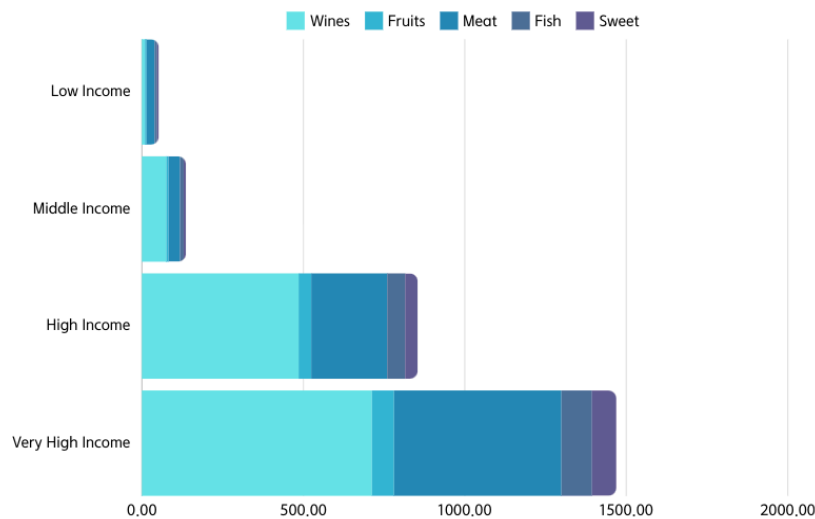


NO.1 DATA

- 와인 소비(**Wines**): 20대가 와인 소비에서 가장 높고(392.65), 그다음은 60대 이상(376.74), 50대(337.97) 순이다. 30대(237.01)와 40대(248.11)는 상대적으로 낮다.
- 과일 소비(**Fruits**): 20대가 가장 높다(39.84). 60대 이상(28.29), 50대(27.35), 30대(27.21), 40대(22.79) 순이다.
- 고기 소비(**Meat**): 20대가 고기 소비에서도 가장 높다(323.53). 60대 이상(191.29), 30대(164.14), 50대(161.63), 40대(136.03) 순이다.
- 생선 소비(**Fish**): 20대가 가장 높다(62.67), 그다음은 60대 이상(43.65), 30대(37.66), 50대(36.11), 40대(32.34) 순이다.

2.3.4. 수입구간별 제품 소비 형태

Income_Group	Wines	Fruits	Meat	Fish	Sweet
Low Income	9.61	5.05	24.76	7.74	5.37
Middle Income	75.61	7.09	35.44	11.22	7.23
High Income	485.29	39.22	235.07	55.53	39.16
Very High Income	712.51	68.15	518.45	94.82	75.32



NO.1 DATA

- 와인 소비(**Wines**): 소득이 높을수록 와인 소비가 크게 증가한다. Low Income 그룹은 9.61, Middle Income은 75.61, High Income은 485.29, Very High Income은 712.51이다.
- 과일 소비(**Fruits**): 소득이 증가할수록 과일 소비도 증가한다. 특히 Very High Income 그룹은 68.15로 과일 소비가 가장 많다.
- 고기 소비(**Meat**): 고기 소비도 소득이 높을수록 증가한다. Very High Income 그룹이 518.45로 가장 많다.
- 생선 소비(**Fish**): 생선 소비는 Very High Income 그룹이 94.82로 가장 높다.
- 단맛(**Sweet**) 소비: 소득이 높을수록 단맛의 소비도 증가하는 경향을 보인다. Very High Income 그룹이 75.32로 가장 많다.

2.3.5. 제품간 상관관계 분석

Correlation	Wines	Fruits	Meat	Fish	Sweet
Wines		0.38	0.60	0.60	0.39
Fruits	0.38		0.55	0.60	0.50
Meat	0.60	0.55		0.59	0.55
Fish	0.40	0.60	0.59		0.55
Sweet	0.39	0.50	0.55	0.55	

- 고기와 와인, 생선과 과일의 조합이 다른 항목들보다 더 많이 소비되는 경향이 있다.
- 단맛은 다른 항목들과 함께 소비되는 경우가 비교적 적다.
- 학력 수준과 관련된 지표와의 상관관계

2.3.6. 학력 수준과 관련된 지표와 상품 카테고리 간의 상관관계 분석

correlations	fish	fruits	meat	sweet	wines
age	0.04	0.01	0.04	0.02	0.16
income	0.55	0.54	0.70	0.56	0.73
spent	0.64	0.61	0.86	0.60	0.90

학력과 상관 있는 것으로 보이던 세 지표, 나이, 가계소득, 총 지출액과의 상관계수를 분석한 결과, 상품 카테고리 5개 모두 소득, 지출액 지표와 상관계수가 0.5 이상으로 나왔고, 나이는 상품 카테고리 5개 모두와 상관계수의 절댓값이 0.2를 넘기는 게 단 하나도 없었다. 나이는 상품 카테고리에 대한 지출과 별 관련이 없다고 결론지을 수 있다. 또한, 육류와 와인의 상관계수 수치가 매우 높게 나타났다. 소득과 육류, 와인의 상관계수는 각각 약 0.75, 0.77이었으며, 지출액과의 그

NO.1 DATA

관계는 약 0.86, 0.9로 나타났다. 이는 육류와 와인의 지출이 전체 지출의 상당수를 차지한다는 것을 강하게 지지한다.

2.4 채널별 매출 분석

2.4.1 카테고리별 지출금액과 채널 구매 빈도 간의 상관관계 분석

다양한 제품 카테고리(MntFish, MntFruits, MntMeat 등)와 각 채널(Deals, Web, Catalog, Store) 간의 상관관계를 분석한 결과, 다음과 같은 패턴이 나타났다.

	Fish	Fruits	Meat	Sweet	Wines	GoldProds
Deals	-0.15	-0.14	-0.17	-0.12	0.01	0.06
Web	0.30	0.30	0.33	0.33	0.55	0.41
Catalog	0.56	0.51	0.71	0.52	0.67	0.47
Store	0.46	0.46	0.52	0.45	0.64	0.39

- 인사이트:
 - Deals(할인)**: 대부분의 카테고리에서 할인 구매와 지출 간의 약한 음의 상관관계가 나타났다.
 - 웹(Web)** 채널: 와인(MntWines, 0.55)과 고급 제품(MntGoldProds, 0.41)에 대해 강한 상관관계를 보였다. 이는 웹 채널에서 프리미엄 제품 마케팅을 강화할 필요가 있음을 시사한다.
 - 카탈로그(Catalog)** 채널: 모든 카테고리에서 가장 강한 상관관계를 보였으며, 특히 고기(MntMeat, 0.71)와 와인(MntWines, 0.67)에 대한 지출에서 큰 기여를 했다.
 - 스토어(Store)** 채널: 전반적으로 높은 상관관계를 보였으며, 특히 와인(MntWines, 0.64)에서 강세를 보였다.

2.4.2 채널별 매출 기여도 추정

분석방법:

각 채널에서의 구매 횟수를 바탕으로 매출 기여도를 추정하였다. 직접적인 매출 데이터는 없었지만, 구매 횟수와 구매 금액 간의 비례 관계를 가정하여 매출액을 추정하였다.

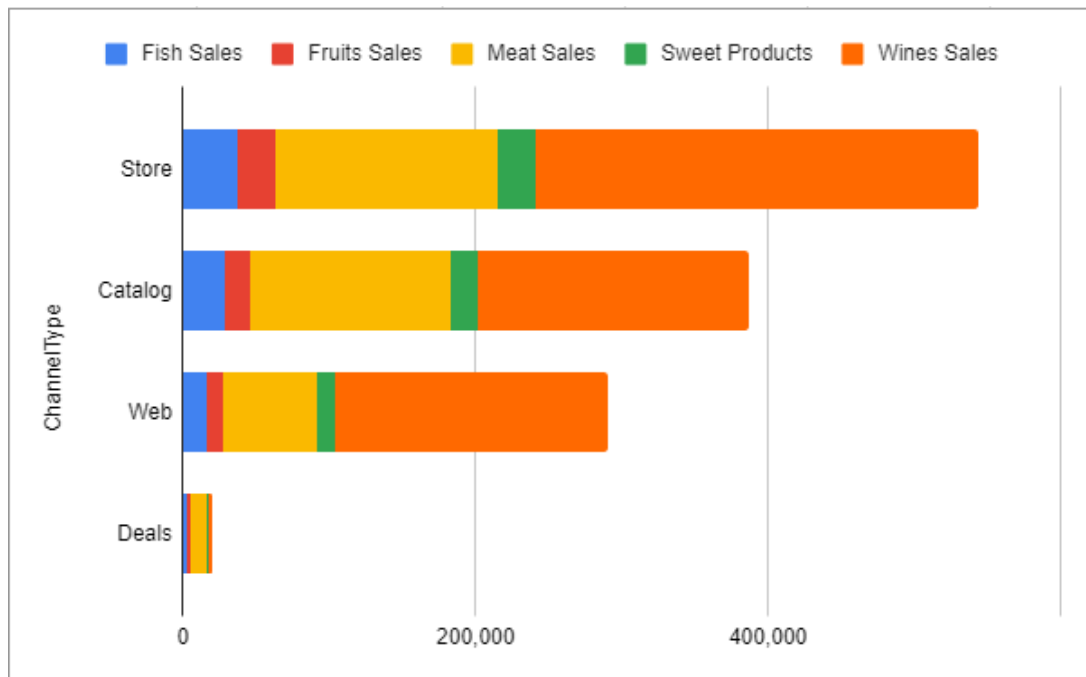
가정:

- 각 채널에서의 평균 구매 금액은 동일하다.

NO.1 DATA

2. 구매 횟수와 구매 금액이 비례한다.

ChannelType	Fish Sales	Fruits Sales	Meat Sales	Sweet Products	Wines Sales	Gold Sales	Total
Store	36,663	26,790	151,702	26,459	302,297	38,599	545,841
Catalog	27,944	17,894	137,680	18,462	185,411	26,834	386,281
Web	15,843	11,634	63,563	13,160	185,744	29,820	303,927
Deals	2,803	1,901	11,567	1,738	1,641	1,893	18,733
Total	43,787	29,529	201,244	31,621	371,155	56,654	690,201

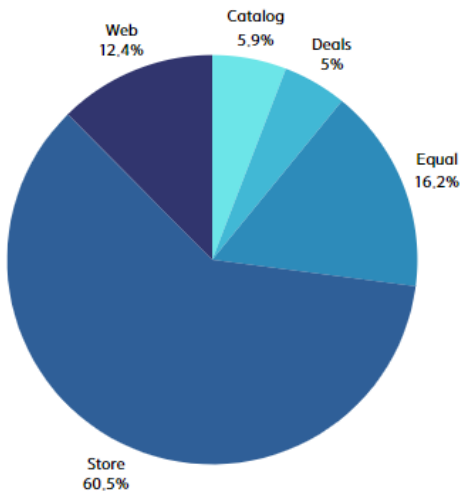


- 인사이트:
 - 스토어(**Store**) 채널: 와인(302,297.14)과 고기(151,702.41) 등 고가 제품에서 가장 높은 매출을 기록했다. 이는 고객들이 스토어에서 직접 제품을 보고 구매하는 선호가 크다는 점을 시사한다. 따라서 스토어에서의 고객 경험을 강화하는 마케팅 전략이 필요하다.
 - 웹(**Web**) 채널: 와인(185,744.30)과 고급 제품(29,819.65)에서 강세를 보였지만, 다른 제품군에서는 상대적으로 낮은 매출을 기록했다. 웹 채널의 매출 증대는 UX/UI 개선과 맞춤형 프로모션을 통해 달성할 수 있다.
 - 카탈로그(**Catalog**) 채널: 고기(137,680.30)와 와인(185,410.94)에서 높은 매출을 기록했으며, 고가 제품군에서 특히 강세를 보였다. 이는 카탈로그 채널을 통해 VIP 고객을 타겟으로 한 맞춤형 마케팅이 효과적일 수 있음을 나타낸다.

NO.1 DATA

- 할인(Deals) 채널: 할인 채널은 전반적으로 매출 기여도가 낮았으며, 특히 고가 제품에서는 매우 낮은 매출을 기록했다. 이는 저가 제품 판매 촉진에 더 효과적이며, 고가 제품에 대한 할인 전략이 재검토되어야 한다는 점을 시사한다.

2.4.3. 채널별 고객 이용률 분석



- 분석방법:

고객의 채널별 구매 횟수를 기준으로 선호하는 채널을 분석했다. 각 고객이 가장 많이 이용한 채널을 기준으로 그룹화하여 채널별 고객 분포를 도출했다.

- 결과:

스토어 채널이 60% 이상의 고객이 이용하는 주요 채널로 분석되었으며, 웹 채널은 15%로 상대적으로 적었다. 스토어 기반 마케팅을 우선적으로 고려하면서, 웹 고객 유입 전략을 강화할 필요가 있다.

2.4.4. 웹사이트 방문 빈도와 구매 전환율 분석

- 분석방법:

웹사이트 방문 빈도와 웹 구매 횟수를 바탕으로 구매 전환율을 계산하였다. 구매 전환율은 "2년간 웹 구매 수 / (월평균 방문자 수 * 24)"로 정의되었다.

- 결과

- 평균 구매 전환율: 4.5%
- 중위수 구매 전환율: 3.31%

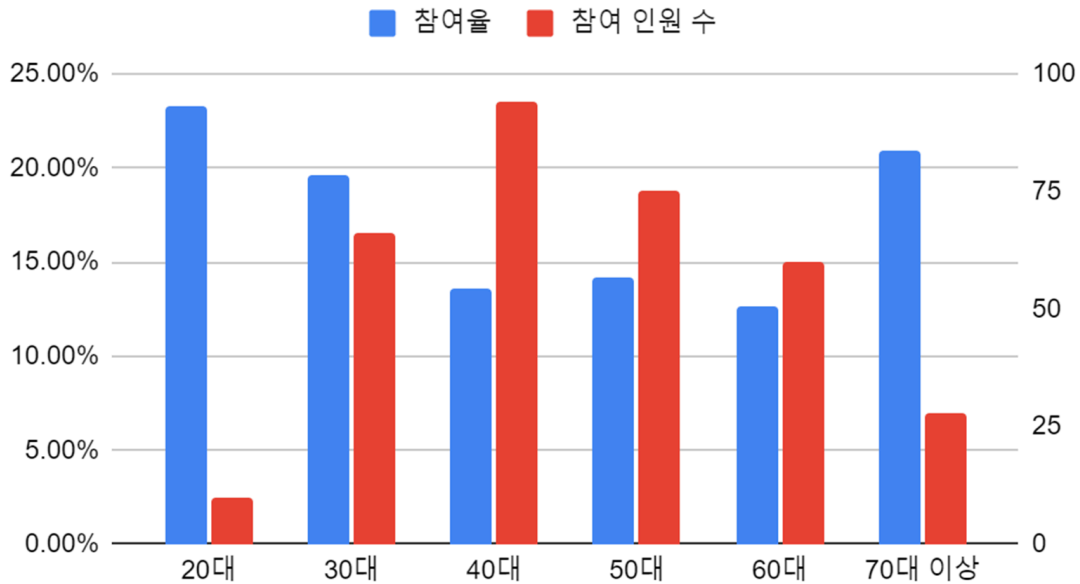
웹사이트 방문 대비 구매 빈도가 매우 낮은 것으로 나타났다. 이는 웹사이트의 사용자 경험(UX/UI)을 개선하고, 개인화된 추천 시스템 도입, 웹 전용 프로모션 강화를 통해 구매 전환율을 개선할 수 있는 여지가 크다.

NO.1 DATA

2.5 캠페인 참여 여부 관련분석

2.5.1 나이에 따른 캠페인 참여 의사

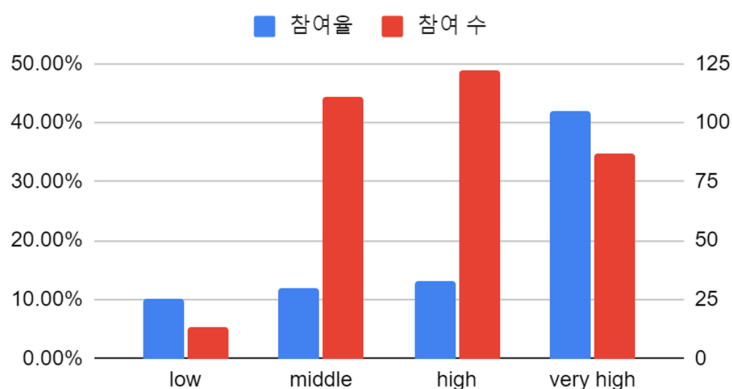
연령층 별 캠페인 참여 의사



- 결과: 2030대와 7080대 고객의 캠페인 참여율이 높았으나, 참여 인원은 적었다. 반면, 데이터셋의 주 연령층인 40대~60대의 캠페인 참여율은 낮았지만, 참여 인원은 많았다.
- 전략 제안: 중간 연령층(40대~60대)을 대상으로 캠페인 참여를 독려하기 위한 맞춤형 전략이 필요하다.

2.5.2 소득과 학력에 따른 캠페인 참여

소득 구간 별 참여 의사

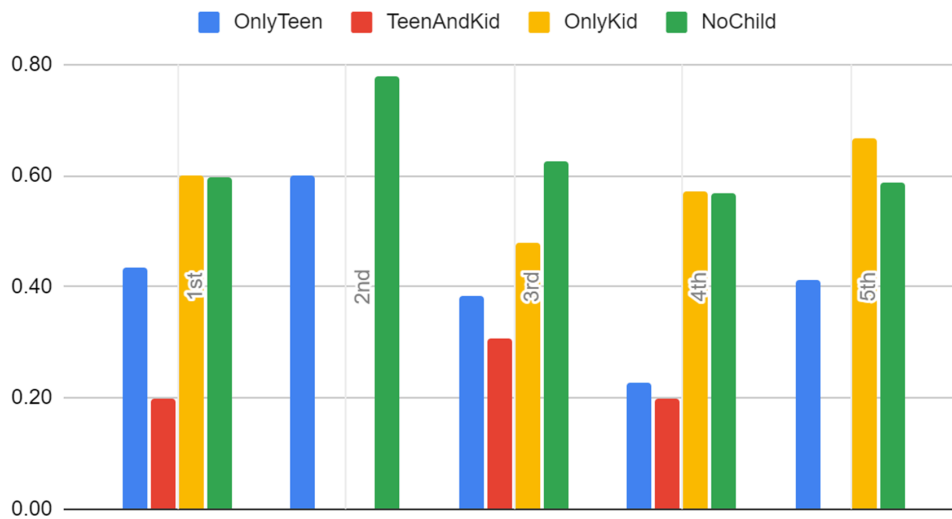


결과: 초고소득층과 고학력 고객일수록 캠페인 참여율이 높다. 학력 수준으로 나누었을 때의 참여 인원 수는 학력 분포와 같은 양상을 띈다.

전략 제안: 이들 그룹을 타겟으로 한 맞춤형 캠페인을 통해 참여를 유도할 수 있다.

2.5.3 자녀 보유에 따른 캠페인 참여

n차 캠페인에 참여했을 때, 새 캠페인에 참여(r)할 확률, $P(r|n)$



- 결과: 자녀가 없는 가정과 10세 이하 자녀만 있는 가정에서 기존 캠페인에 참여한 경우 새 캠페인 참여율이 상대적으로 높게 나타났다.
- 분석: 나이가 비교적 젊은 층에서 캠페인 참여율이 높았던 것과 일관된 결과를 보인다. 자녀가 없는 가정이나 어린 자녀를 둔 가정이 새로운 서비스나 제품에 대한 관심이 높을 수 있으며, 이들이 캠페인에 더 적극적으로 참여하는 경향이 있다.
- 전략 제안: 이러한 고객층을 대상으로 한 맞춤형 캠페인을 통해 캠페인 참여율을 더욱 높일 수 있다. 자녀가 없는 고객이나 청소년 자녀를 둔 고객을 타겟으로 한 캠페인 혜택을 강화하는 것이 효과적일 수 있다.

3. 결론 및 기대효과

3.1 마케팅 전략

3.1.1 가족 구성원별 맞춤형 마케팅 전략

NO.1 DATA

Kid & Teen Home (아이와 청소년이 모두 있는 가정)

- 특징: 모든 항목에서 소비가 낮다. 특히, 고기와 와인 소비가 적다.
- 마케팅 전략:
 - 건강과 영양 중심의 마케팅: 아이와 청소년이 있는 가정은 건강한 식단을 중요시할 가능성이 높다. 저칼로리, 고영양 제품과 함께 건강식을 강조하는 프로모션을 통해 접근한다.
 - 가족 단위 패키지 상품 제공: 가족 전체가 즐길 수 있는 식사 키트나 묶음 상품(예: 과일과 생선, 간편 조리 세트 등)을 기획하여 판매한다. 아이들에게도 어필할 수 있는 건강한 스낵을 포함한다.
 - 비알코올 대체 제품 마케팅: 와인 대신 가족과 함께 즐길 수 있는 무알코올 음료나 어린이 및 청소년이 선호하는 음료로 대체할 수 있는 마케팅 전개한다.

No Kids & Teens (아이와 청소년이 없는 가정)

- 특징: 모든 항목의 소비가 높다. 특히, 고기와 생선의 소비가 두드러진다.
- 마케팅 전략:
 - 고급화 전략: 아이와 청소년이 없는 가정은 상대적으로 경제적 여유가 있을 가능성이 높으므로, 프리미엄 제품 마케팅을 강화한다. 고급 와인, 고품질 고기, 해산물 등의 고가 상품 추천한다.
 - 식재료와 와인의 페어링 제안: 고기, 생선, 와인 소비가 높다는 점을 활용해 식재료와 와인의 페어링을 추천하는 캠페인을 진행한다. 레시피 제안 및 페어링 키트 판매한다.
 - 경험 중심 마케팅: 미식 경험이나 요리 클래스를 함께 제공하여 특별한 경험을 중시하는 소비자층을 타겟팅하면 좋다.

Only Kidhome (아이만 있는 가정)

- 특징: 모든 항목에서 소비가 감소했고, 특히 와인 소비가 크게 감소했다.
- 마케팅 전략:
 - 어린이를 위한 건강한 간식과 간편식 제안: 아이가 있는 가정은 간편하고 건강한 식사에 관심이 많다. 따라서 어린이를 위한 건강 간식, 영양가 높은 간편식을 강조한다.
 - 가족 참여형 요리 키트: 아이와 함께 요리할 수 있는 간단한 레시피와 재료가 포함된 키트를 제공하여 아이와의 즐거운 시간을 강조한다.
 - 비알코올 음료 홍보: 와인 소비가 낮은 점을 고려해 아이들이 즐길 수 있는 다양한 맛의 비알코올 음료나 과일 주스 중심의 프로모션을 강화한다.

Only Teenhome (청소년만 있는 가정)

- 특징: 와인과 금 상품 소비는 약간 증가했고, 나머지 항목들은 큰 변화가 없다.
- 마케팅 전략:
 - 청소년과 부모를 모두 타겟팅: 청소년은 부모의 영향을 많이 받기 때문에, 부모와 청소년 모두를 겨냥한 가족 패키지 상품을 제안한다.
 - 트렌디한 비알코올 음료 마케팅: 청소년이 관심을 가질 만한 비알코올 음료나 건강 음료를 트렌디한 방식으로 홍보한다.
 - 와인 대체제와 프리미엄 비식품 상품 마케팅: 청소년만 있는 가정에서는 부모가 자신을 위한 고급 제품을 구매할 가능성도 있으므로, 프리미엄 비식품(금 상품 등)과 와인 대체 상품을 함께 마케팅한다.

3.1.2. 가족 구조별 맞춤형 마케팅 전략

Widow 그룹 대상 마케팅 전략

- 프리미엄 제품 및 고급 패키지 상품 제공: Widow 그룹은 모든 항목에서 소비가 높기 때문에 고급화 전략을 사용할 수 있다. 프리미엄 와인, 과일 세트, 고급 해산물, 고급 디저트 등의 상품을 추천하며, 패키지 상품을 제안해 고급스러운 느낌을 강조한다.
- 건강 및 웰빙 중심의 제품 홍보: 이 그룹은 상대적으로 여유 있는 소비 패턴을 보일 가능성이 크다. 건강식과 웰빙을 강조한 고급 제품을 제안하여 이들의 관심을 유도한다.
- 편리한 쇼핑 옵션 제공: Widow 그룹은 단독 생활의 비율이 높을 가능성이 있으므로, 편리한 배달 서비스와 구독형 모델을 제안하여 구매 편의를 제공하는 것이 좋다.

Single 그룹 대상 마케팅 전략

- 간편한 요리 키트와 식사 대용 제품 마케팅: Single 그룹은 고기 소비가 높은 것으로 나타났다. 간편하게 조리할 수 있는 고기 요리 키트, 샐러드 패키지, 고단백 식사 대용품 등을 강조한다.
- 와인과 함께 즐길 수 있는 소규모 파티용 제품 추천: 와인 소비도 높으므로 소규모 파티나 친구와 함께 즐길 수 있는 와인과 간단한 안주 세트 상품을 기획한다.
- 할인 프로모션과 쿠폰 전략: 싱글 라이프스타일을 즐기는 이들에게 매력적인 할인 프로모션과 적립 혜택 등을 통해 재구매를 유도한다.

Together, Married, Divorced 그룹 대상 마케팅 전략

NO.1 DATA

- 가족 단위 상품 및 경제적인 패키지 제공: Together, Married, Divorced 그룹은 전반적으로 안정된 소비 패턴을 보인다. 가족 단위로 즐길 수 있는 경제적인 패키지 상품을 추천하여 가성비를 강조한다.
- 일상 식료품 및 건강 제품 홍보: 과일, 생선, 고기 등 건강에 좋은 식재료를 강조한 홍보를 통해 가족의 건강을 중시하는 마케팅을 전개한다.
- 구독형 서비스 제안: 이들 그룹은 상대적으로 꾸준한 소비 패턴을 보이므로, 정기적으로 배송받을 수 있는 구독형 서비스와 같이 편리함을 강조한 마케팅 전략을 사용할 수 있다.

3.1.3. 수입 구간별 맞춤형 마케팅 전략

Low Income 그룹 대상 마케팅 전략

- 저렴한 가격과 가치 중심 마케팅: 이 그룹은 모든 항목에서 소비가 낮다. 따라서 가격에 민감한 소비자들을 대상으로 한 가성비 좋은 상품과 프로모션이 효과적이다. 대용량 패키지나 저렴한 가격대의 상품을 묶어서 할인 제공한다.
- 기본 필수품 위주의 상품 기획: 과일, 고기, 생선 등 기본적인 식재료를 중심으로, 필수적인 생활 용품 패키지를 기획하여 제공한다.
- 편의성 강조: 저렴한 가격대의 간편식이나 즉석 조리식품을 강조하여 경제적이고 실용적인 상품을 홍보한다.

Middle Income 그룹 대상 마케팅 전략

- 중저가 프리미엄 제품 홍보: 이 그룹은 중저가에서 고급으로 넘어가는 제품에 관심을 가질 가능성이 크다. 중저가의 프리미엄 와인, 과일, 고기 등을 강조하며, 건강과 품질을 내세운 마케팅을 진행한다.
- 가족 및 친구 모임 패키지 제공: 이 그룹은 중간 소득층으로, 가족 및 친구와의 모임에 관심이 많을 수 있다. 소규모 모임에 적합한 와인, 고기, 생선 패키지 상품을 제안한다.
- 다양한 프로모션과 할인 혜택: 충성도 높은 고객을 확보하기 위해 할인 쿠폰과 멤버십 혜택 제공한다. 정기 구매 고객에게 추가 혜택을 부여하는 등 프로모션 강화한다.

High Income 그룹 대상 마케팅 전략

- 프리미엄 제품 및 럭셔리 마케팅 전략: 이 그룹은 고급 와인, 고기, 생선 소비가 높기 때문에, 프리미엄 제품을 타겟팅한 마케팅이 효과적이다. 고급 와인, 프리미엄 해산물 세트, 고품질 고기 등 고가의 제품을 적극적으로 홍보한다.
- 고객 경험 중심의 마케팅: 고급 레스토랑 체험, 미식 여행, 와인 테이스팅 이벤트 등을 통해 프리미엄 소비 경험을 제공한다.

NO.1 DATA

- 고객 맞춤형 서비스 제공: 개인 맞춤형 추천 시스템을 도입하여 고객이 선호하는 제품을 추천하고, VIP 전용 상담 서비스 및 혜택 제공한다.

Very High Income 그룹 대상 마케팅 전략

- 최고급화 및 럭셔리 경험 제공: Very High Income 그룹은 모든 항목에서 가장 높은 소비를 보이기 때문에, 최고급 와인, 희귀 과일, 최고급 고기와 생선 등 최고가 제품을 중심으로 한 마케팅이 적합하다.
- 프리미엄 구독 서비스: 고급 식재료와 와인을 정기적으로 배송받을 수 있는 프리미엄 구독 서비스를 제공하여 편의성과 고급스러움을 동시에 제공한다.
- 개인화된 고급 체험 및 서비스: 프리미엄 요리사 초빙, 맞춤형 이벤트 기획, 고급 호텔이나 리조트와 제휴한 이벤트 등을 통해 독특하고 개인화된 소비 경험을 제공한다.
- 한정판 제품 마케팅: 특정 시즌이나 이벤트에 맞춰 한정판 와인, 고급 과일 바구니, 희귀한 해산물 패키지 등을 판매하여 고유한 경험을 제공한다.

3.1.4. 연령별맞춤형 마케팅 전략

20대 대상 마케팅 전략

- 모던하고 트렌디한 상품과 경험 제공: 20대는 와인, 고기, 생선 등 다양한 식품 소비가 높다. 트렌디한 와인 브랜드, 최신식 요리법을 활용한 고기 및 생선 요리 키트 등을 제안한다.
- 소셜 미디어 기반 마케팅: 인스타그램, 틱톡과 같은 소셜 미디어에서의 영향력이 큰 소비자층이므로, 인플루언서와 협력한 제품 리뷰, 요리 챌린지 등을 통해 브랜드 인지도를 높인다.
- 건강 및 웰빙 트렌드 강화: 과일과 건강한 스낵 제품도 선호하는 경향이 있으므로, 웰빙 트렌드에 맞춘 건강한 식사 대체품과 간식 등을 강조한다.

30대와 40대대상 마케팅 전략

- 실용적이고 편리한 제품 마케팅: 30대와 40대는 전반적으로 소비가 낮다. 가족과 직장 생활로 바쁜 연령층이므로 간편식, 가정간편식(HMR) 등의 실용적인 제품을 추천한다.
- 가성비 좋은 제품과 할인 혜택 제공: 비용 대비 성능을 중시하는 연령대이므로, 가격 대비 품질이 좋은 제품을 강조하고, 할인 프로모션 및 쿠폰 혜택 제공한다.
- 가족 중심 마케팅 전략: 가족과 함께 즐길 수 있는 패밀리 세트나 어린이도 즐길 수 있는 건강한 간식 패키지 등을 기획한다.

50대 대상 마케팅 전략

NO.1 DATA

- 프리미엄 와인 및 고급 식재료 마케팅: 50대는 와인 소비가 높은 편이므로, 프리미엄 와인과 고급 식재료 세트를 제안한다. 와인과 함께 즐길 수 있는 미식 요리 키트를 함께 판매한다.
- 건강과 웰빙 강조: 과일과 생선 등 건강에 좋은 식재료를 활용한 상품을 추천하며, 건강을 강조한 제품 설명과 레시피를 제공한다.
- 문화와 경험 중심의 마케팅: 와인 테이스팅 이벤트, 요리 교실 등 새로운 경험을 제공하는 마케팅 활동을 강화한다.

60대 이상 대상 마케팅 전략

- 건강 중심의 고급화 전략: 이 연령대는 건강을 중시하면서도 프리미엄 제품을 선호하는 경향이 있다. 고급 와인, 프리미엄 생선 및 고기, 고영양 과일 등의 고급화된 제품을 추천한다.
- 정기 구독 서비스 제안: 편리함을 중시하는 경향도 있으므로, 건강 식재료나 와인 정기 구독 서비스를 통해 지속적인 소비를 유도한다.
- 사회적 가치와 관련된 마케팅: 친환경, 사회적 기업 제품 등을 강조하여 이들의 사회적 관심을 반영한 마케팅을 전개한다.

3.1.5. 마케팅 채널별 맞춤형 마케팅 전략

- 스토어 채널: iFood의 주요 매출 채널로서, 프리미엄 제품과 오프라인 고객 경험을 강화하는 것이 중요하다. 와인과 고기 제품에 대한 프로모션과 고급화를 중심으로 마케팅 전략을 수립해야 한다.
- 웹 채널: 웹사이트 구매 전환율이 낮으므로, 이를 개선하기 위한 전략이 필요하다. UX/UI 개선, 개인화된 추천 시스템 도입, 웹 전용 프로모션 강화 등을 통해 고객의 웹 구매 전환율을 높일 수 있다.
- 카탈로그 채널: 고가 제품을 대상으로 한 카탈로그 마케팅이 효과적이므로, 이를 활용한 개인화된 오퍼 제공이 필요하다. 특히, 고소득층 고객을 대상으로 한 프리미엄 마케팅 전략을 강화해야 한다.
- 할인 채널: 할인 캠페인의 효과는 제한적이므로, 저가 제품군에 집중된 할인 전략이 필요하다. 이를 통해 재고 소진 및 신규 고객 유입을 유도할 수 있다.

3.1.6.. 마케팅 전략 제안

- 데이터 기반 맞춤형 타겟팅: 학력, 소득, 나이, 가족 구조 등에 따라 세분화된 고객 그룹에 맞춤형 마케팅 전략을 수립한다.

NO.1 DATA

- 고소득층 및 고학력자: 프리미엄 제품과 럭셔리 서비스를 제공하고, 이들 고객을 위한 구독 서비스나 개인화된 혜택을 강조한다.
- 젊은 층 및 자녀가 없는 가정: 트렌디한 상품과 경험을 제공하며, 소셜 미디어 기반 마케팅을 통해 젊은 층의 참여를 유도한다.
- 가족 단위 고객: 가족 중심의 경제적인 패키지 상품과 건강을 강조한 제품을 제안한다.
- 채널별 마케팅 전략:
 - 스토어: 오프라인 매장에서의 프리미엄 제품 경험을 강화하고, 고급화된 매장 프로모션을 통해 매출 증대를 목표로 한다.
 - 웹: 웹사이트 전용 독점 상품 및 프로모션을 제공하여 웹 구매 전환율을 높입니다. 또한, 웹사이트에서의 사용자 경험을 최적화하여 고객 만족도를 높인다.
 - 카탈로그: 디지털 카탈로그로 전환하여 개인화된 오퍼를 제공하고, VIP 고객을 대상으로 한 프리미엄 마케팅을 추진한다.
 - 할인: 신규 고객 유입을 위한 첫 구매 할인과 같은 전략을 강화하며, 저가 제품군에 집중된 할인 캠페인을 진행한다.

3.2 마케팅 기대효과

- 매출 증대: 고객의 특성과 채널별 특성을 고려한 맞춤형 마케팅 전략을 통해 각 채널에서의 매출을 극대화할 수 있다.
- 고객 만족도 향상: 개인화된 마케팅 전략과 사용자 경험 개선을 통해 고객 만족도와 재구매율을 높일 수 있다.
- 시장 점유율 확대: 맞춤형 마케팅 전략을 통해 iFood의 시장 점유율을 더욱 확대할 수 있으며, 특히 웹 채널에서의 성장이 기대된다.

Github for iFood_EDA : https://github.com/penguin347/iFood_EDA

REFERENCE

1. 브라질 2014 인터넷 보급률
<https://www.metroseoul.co.kr/article/2014031600140>
2. 브라질 2019-2020 인터넷 보급률
<https://www.statista.com/forecasts/292757/brazil-internet-user-penetration>

NO.1 DATA

3. 브라질 배달앱 이용률 및 ifood 시장 점유율

https://www.kati.net/board/globalVillageReportView.do?board_seq=98284&menu_dept2=35

4. ifood_df.csv file directory

<https://github.com/nailson/ifood-data-business-analyst-test>

5. 데이터 정제 sql 코드

```
select income, Recency, Age,  
       json_object('Teen', Teenhome, 'Kid', Kidhome) as Children, -- teenhome과 kidhome  
       컬럼 병합  
       case  
         when education_basic = 1 then 'basic'  
         when education_2n_Cycle = 1 then '2n_cycle'  
         when education_graduation = 1 then 'graduation'  
         when education_master = 1 then 'master'  
         when education_PhD = 1 then 'PhD'  
       end as Education,  
       case  
         when marital_Single = 1 then 'single'  
         when marital_Together = 1 then 'together'  
         when marital_Married = 1 then 'married'  
         when marital_Divorced = 1 then 'divorced'  
         when marital_Widow = 1 then 'widow'  
       end as MaritalStatus,  
       Customer_Days, complain,  
       json_object('Deals', NumDealsPurchases, 'Web', NumWebPurchases, 'Catalog',  
NumCatalogPurchases, 'Store', NumStorePurchases) as NumPurchases,  
       NumWebVisitsMonth, -- purchase 컬럼 병합  
       json_object('Fish', MntFishProducts, 'Fruits', MntFruits, 'Meat', MntMeatProducts,  
'Sweet', MntSweetProducts, 'Wines', MntWines) as MntCategories, MntTotal,  
       MntRegularProds, MntGoldProds, -- Mnt 컬럼 병합  
       json_object('1st', AcceptedCmp1, '2nd', AcceptedCmp2, '3rd', AcceptedCmp3, '4th',  
AcceptedCmp4, '5th', AcceptedCmp5, 'last', Response) as AcceptedCmp,  
       AcceptedCmpOverall -- AcceptedCmp 병합  
from ifood_df
```

NO.1 DATA

6. 브라질 가계소득 평균

<https://tradingeconomics.com/brazil/wages>