

파뿌리 컨퍼런스 발표

2021314031 최민서

목차

1. 데이터셋 설명
2. EDA과정에서 알아낸 인사이트 소개
3. 연간소득, 구매량에 대한 회귀예측
4. 캠페인 참여여부 분류예측

데이터셋 설명

"days"

데이터셋: 마케팅 데이터

Data columns (total 29 columns):

#	Column	Non-Null Count	Dtype
0	ID	2240 non-null	int64
1	Year_Birth	2240 non-null	int64
2	Education	2240 non-null	object
3	Marital_Status	2240 non-null	object
4	Income	2216 non-null	float64
5	Kidhome	2240 non-null	int64
6	Teenhome	2240 non-null	int64
7	Dt_Customer	2240 non-null	object
8	Recency	2240 non-null	int64
9	MntWines	2240 non-null	int64
10	MntFruits	2240 non-null	int64
11	MntMeatProducts	2240 non-null	int64
12	MntFishProducts	2240 non-null	int64
13	MntSweetProducts	2240 non-null	int64
14	MntGoldProds	2240 non-null	int64
15	NumDealsPurchases	2240 non-null	int64
16	NumWebPurchases	2240 non-null	int64
17	NumCatalogPurchases	2240 non-null	int64
18	NumStorePurchases	2240 non-null	int64
19	NumWebVisitsMonth	2240 non-null	int64
20	AcceptedCmp3	2240 non-null	int64
21	AcceptedCmp4	2240 non-null	int64
22	AcceptedCmp5	2240 non-null	int64
23	AcceptedCmp1	2240 non-null	int64
24	AcceptedCmp2	2240 non-null	int64
25	Complain	2240 non-null	int64
26	Z_CostContact	2240 non-null	int64
27	Z_Revenue	2240 non-null	int64
28	Response	2240 non-null	int64

mark_df.shape #행렬개수

(2240, 29)

[https://github.com/minseochoi-AI/BreakPython_Data-analysis-study/blob/main/Untitled%20\(1\).ipynb](https://github.com/minseochoi-AI/BreakPython_Data-analysis-study/blob/main/Untitled%20(1).ipynb)

내가 하려는 Task는?

마케팅 데이터이므로
매출증대를 위한 방안을
생각해보자!

인사이드 설명

카탈로그 구매 횟수와 육류 구매량의 관계

와인 구매형태의 특성

자녀가 있는 가정의 소득분포특성

연간 소득과 물품 구매량의 관계

EDA과정에서 나온 인사이트들

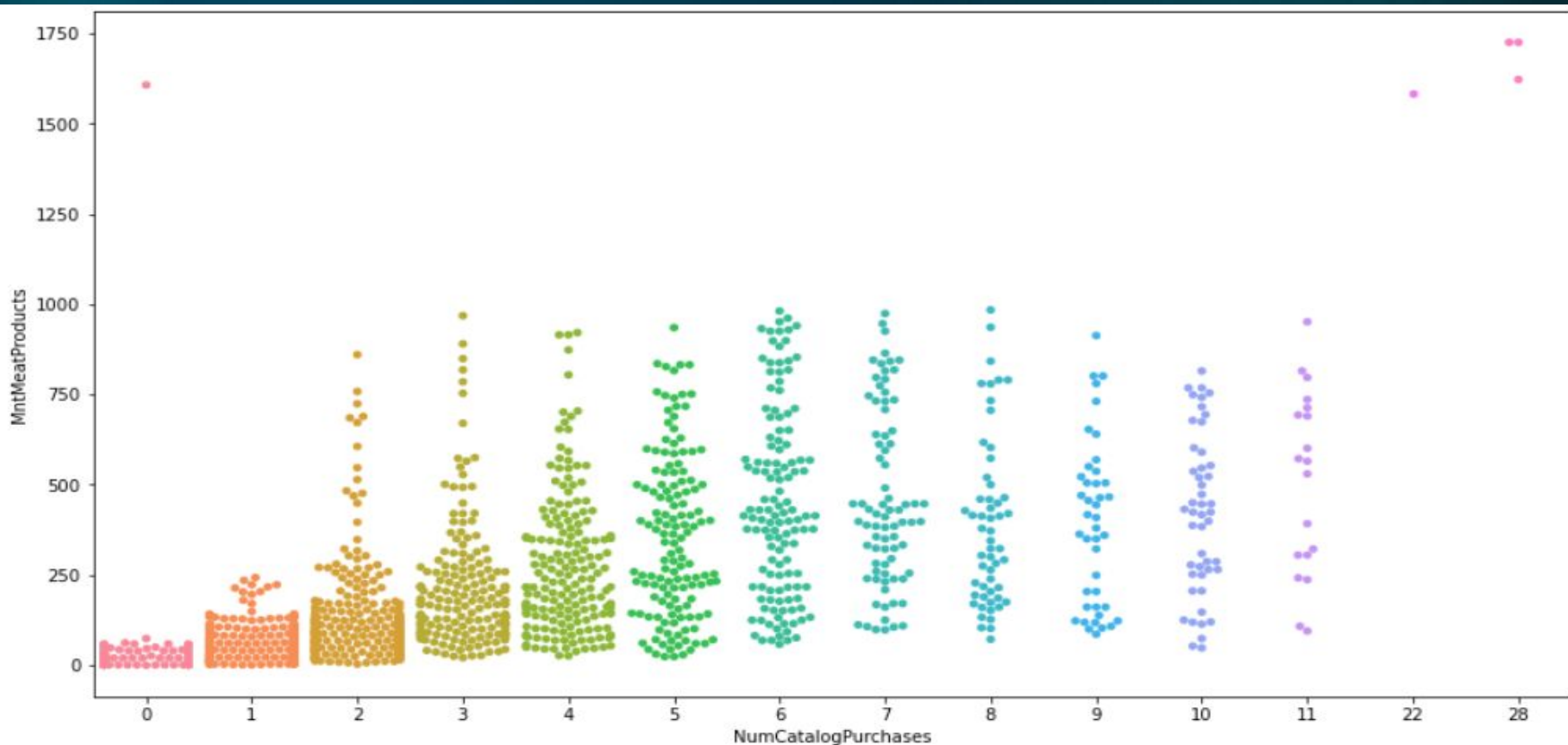
카탈로그 구매 횟수와 육류 구매량의 관계

	NumCatalogPurchases	MntMeatProducts
NumCatalogPurchases	1.000000	0.723827
MntMeatProducts	0.723827	1.000000

Correlation 결과 카탈로그(안내책자) 구매 횟수와 고기
구매량 사이에 강한 선형적 관계!!

Scatter는 값이 겹쳐 제대로 보여주지 못하니
이를 방지해주는 Swarmplot을 찍어 봅시다!

카탈로그 구매횟수 0에서는 고기 구매량도 0~100부근에 뭉쳐있지만 카탈로그 구매 횟수가 늘수록 고기 구매량도 증가하는 경향을 확인해 볼 수 있습니다!



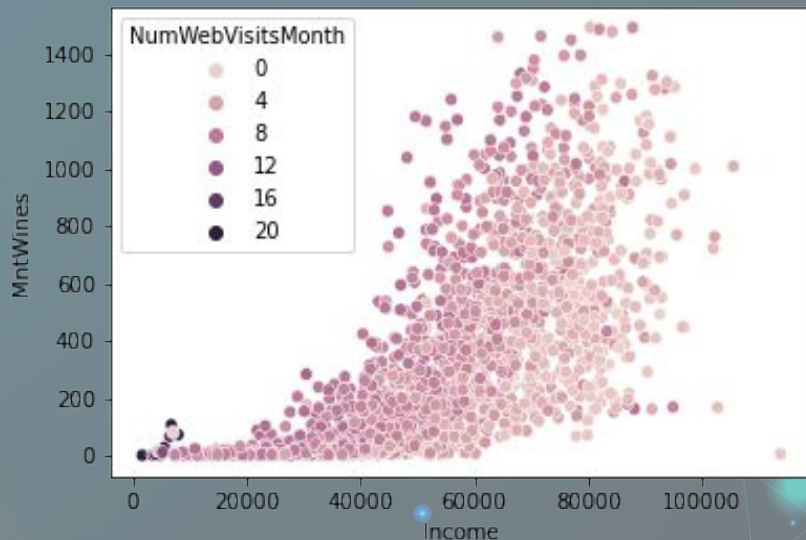
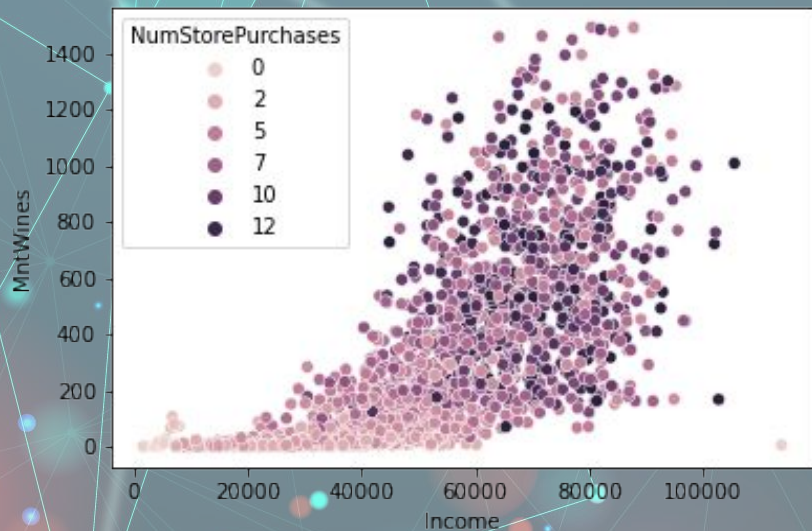
매출증대 방안1

1. 안내책자상에서 사용한 육류관련 광고를 스토어나 웹에서도 연계하여 사용한다.
2. 카탈로그를 통해 육류를 구매 할 시 할인을 해주는 이벤트를 개시한다.

구매 형태에 따른 와인 구매량

와인의 구매량은 매장에 방문해 구매하는 횟수가 많을수록 많아지고 지난달 웹 방문횟수가 많을수록 적어진다.

이는 고객들이 와인의 품질을 직접 확인하고 구매하려는 경향이 있다는 것!!



매출증대 방안 2

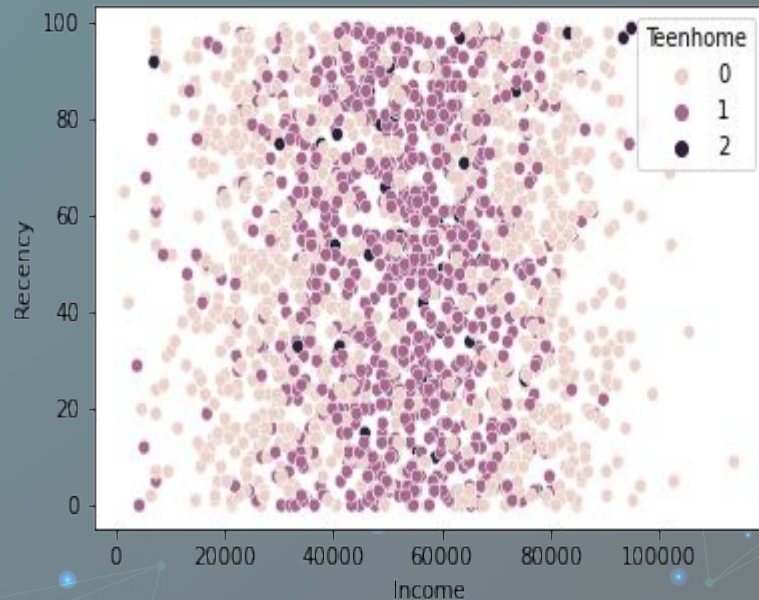
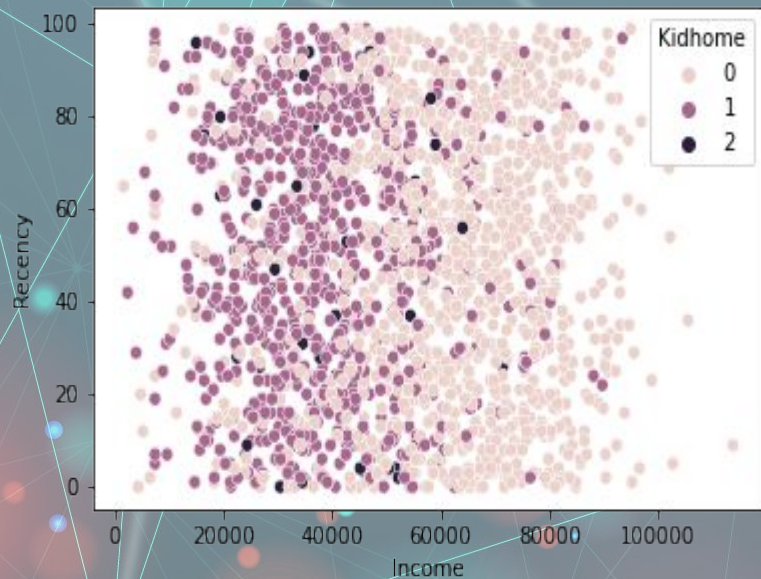
1. 고객들이 웹구매에서도 와인 품질을 신뢰 할 수 있도록 와인 원산지, 연도 인증제 시스템을 도입한다.
2. 고객 방문횟수가 많은 매장을 선별하여 와인 시음회를 개최한다. 이 때 동일한 품질의 와인을 웹으로도 구매가 가능함을 고객들에게 인식시킨다.

자녀가 있는 가정의 소득분포특성

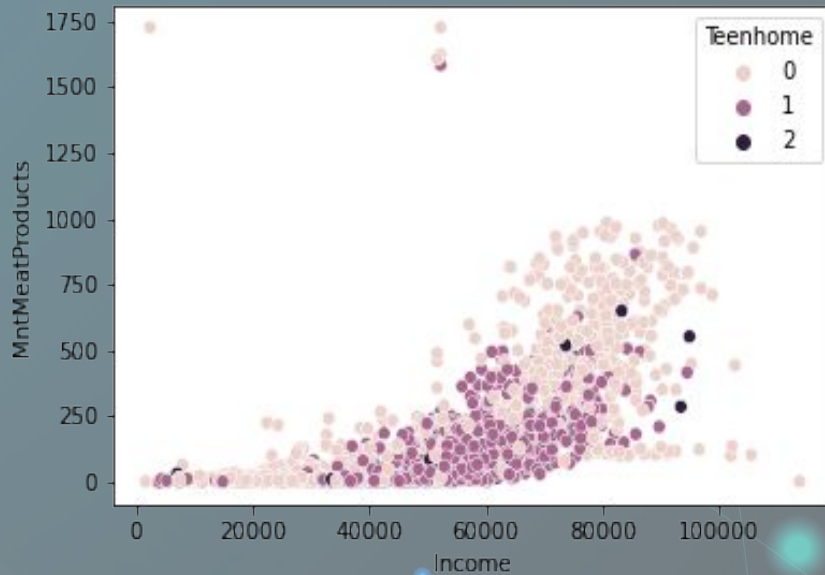
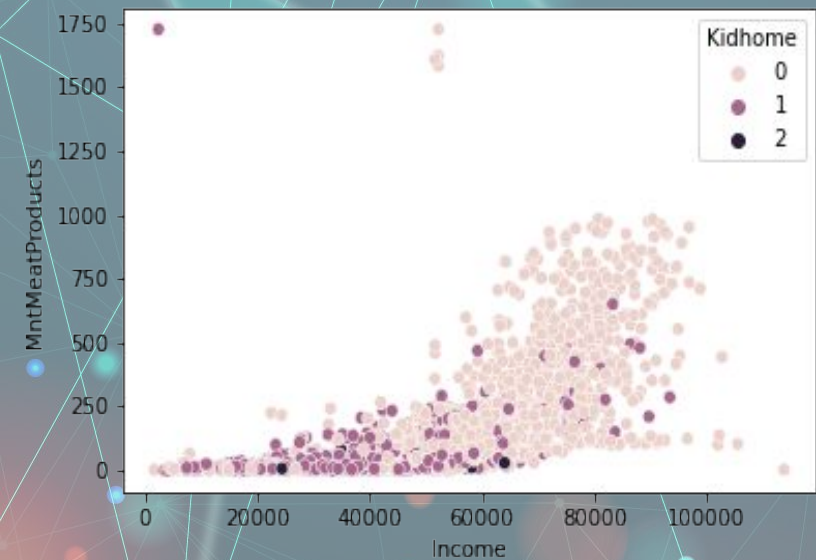
어린이 자녀가 있는 가정은 그렇지 않은 가정보다 대체적으로 연간소득이 적다.

10 대 자녀가 있는 가정의 연간소득은 자녀가 없는 가정보다 분산이 적다.

자녀가 있는 가정이 그렇지 않은 가정보다 더 적은 소비를 할 것이라 예상해봄!



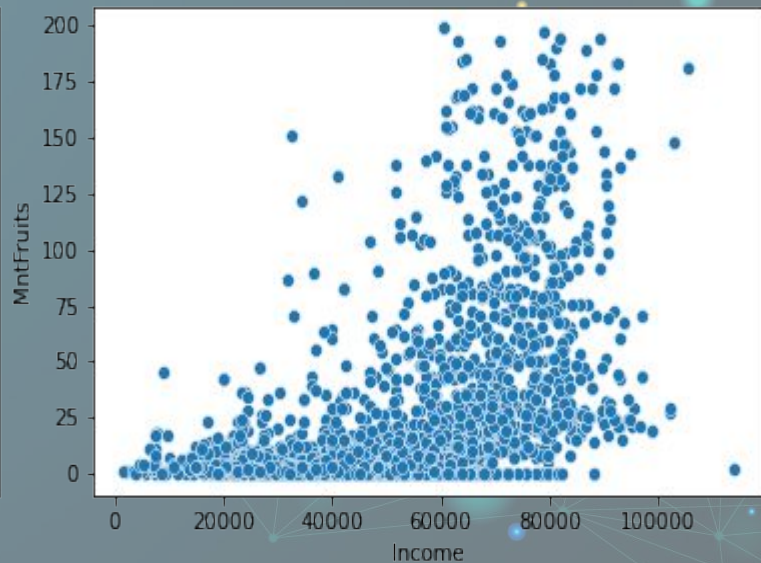
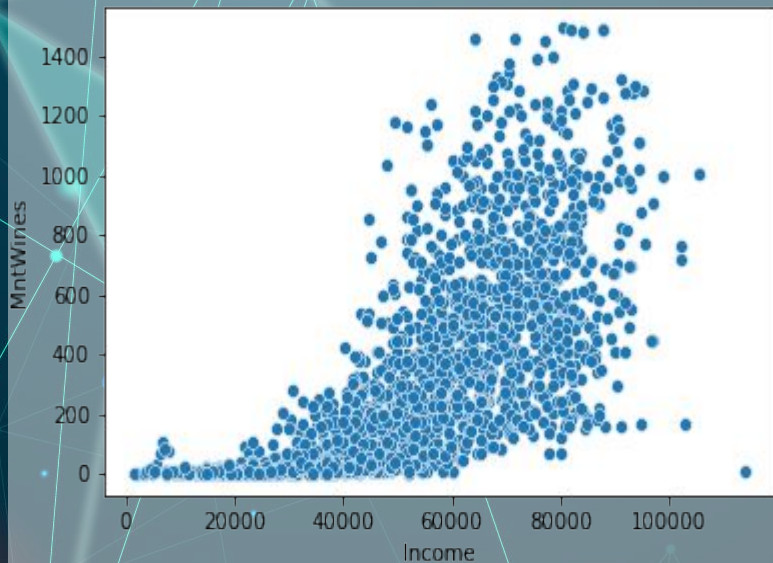
고기 구매량으로 확인해보자

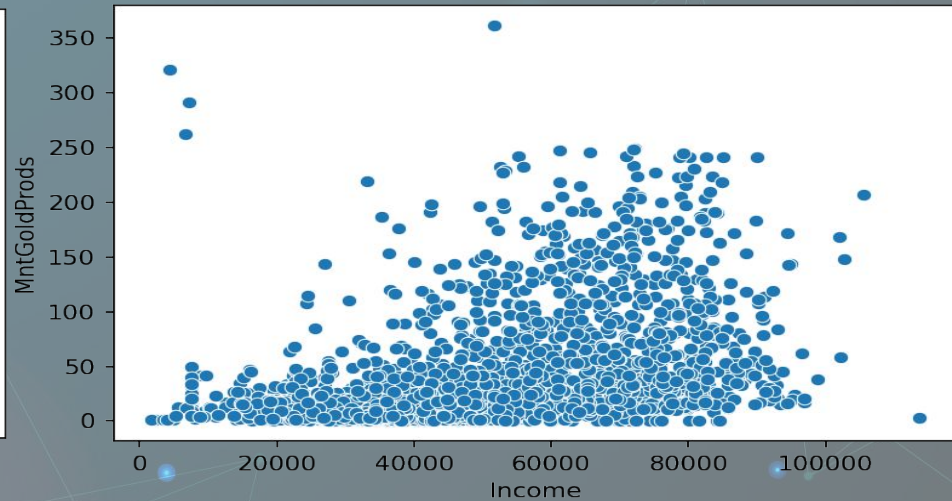
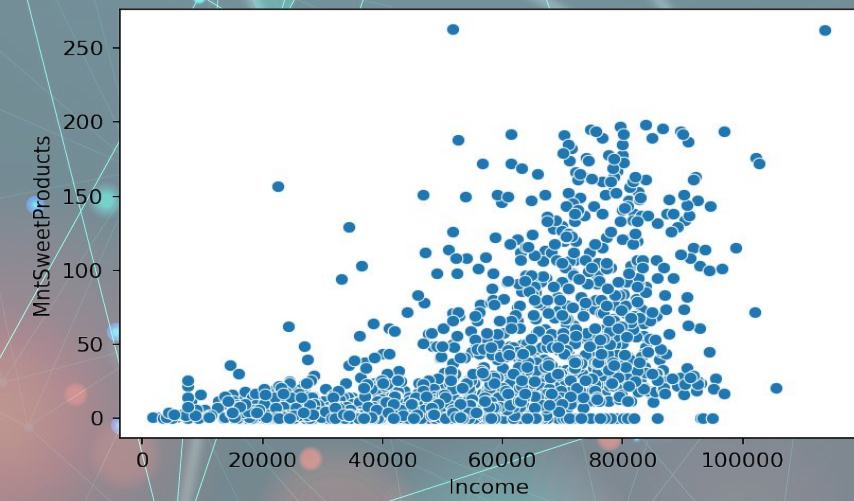
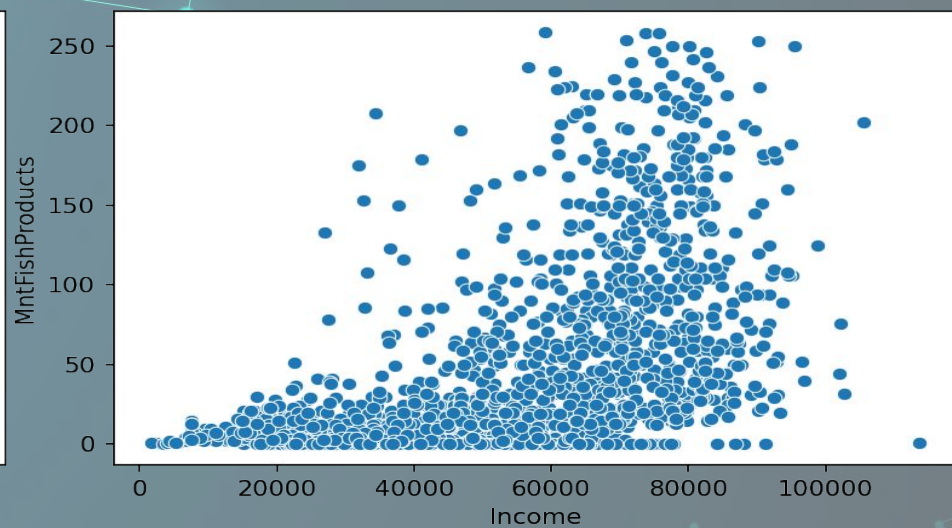
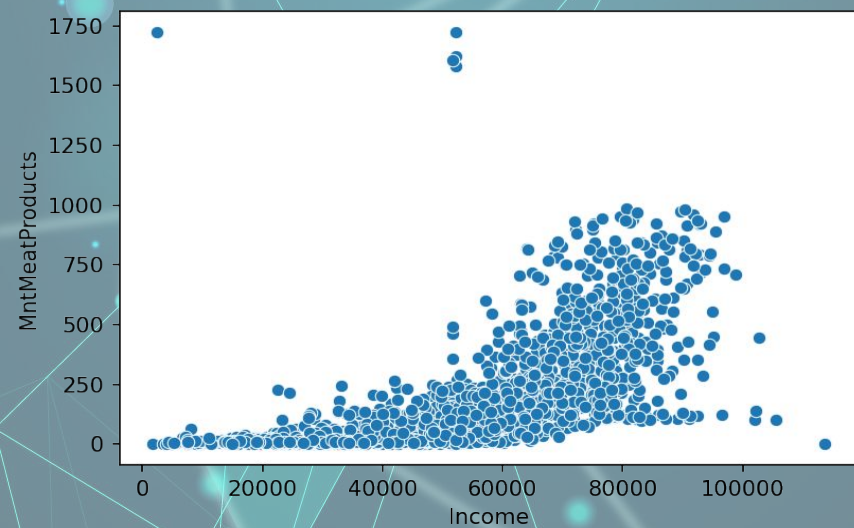


매출증대 방안 3

1. 자녀가 있는 가정은 안정적인 수익에 맞게 적은 금액으로 지속적인 비용을 지불하도록 구독형 상품 (가정용 전자조리도구등)의 소비를 유도한다.
2. 자녀가 없는 가정은 상대적으로 높은비용의 적은수량의 상품을 구매하리라 예상되어 프리미엄 제품위주의 소비를 유도한다.

연간 소득과 물품 구매량의 관계





#고객들은 income(고객 연간 가계소득)이 1.00씩 늘수록 고객이 육류(0.58), 와인(0.57), 단 음식(0.44), 생선(0.438), 과일(0.430), 금(0.32)순으로 더 늘었다
근소한 차이이긴 하나 단 음식과 고기의 관계(0.52)보다 단 음식과 생선(0.57)은 더 큰 선형적 관계를 띄고있다

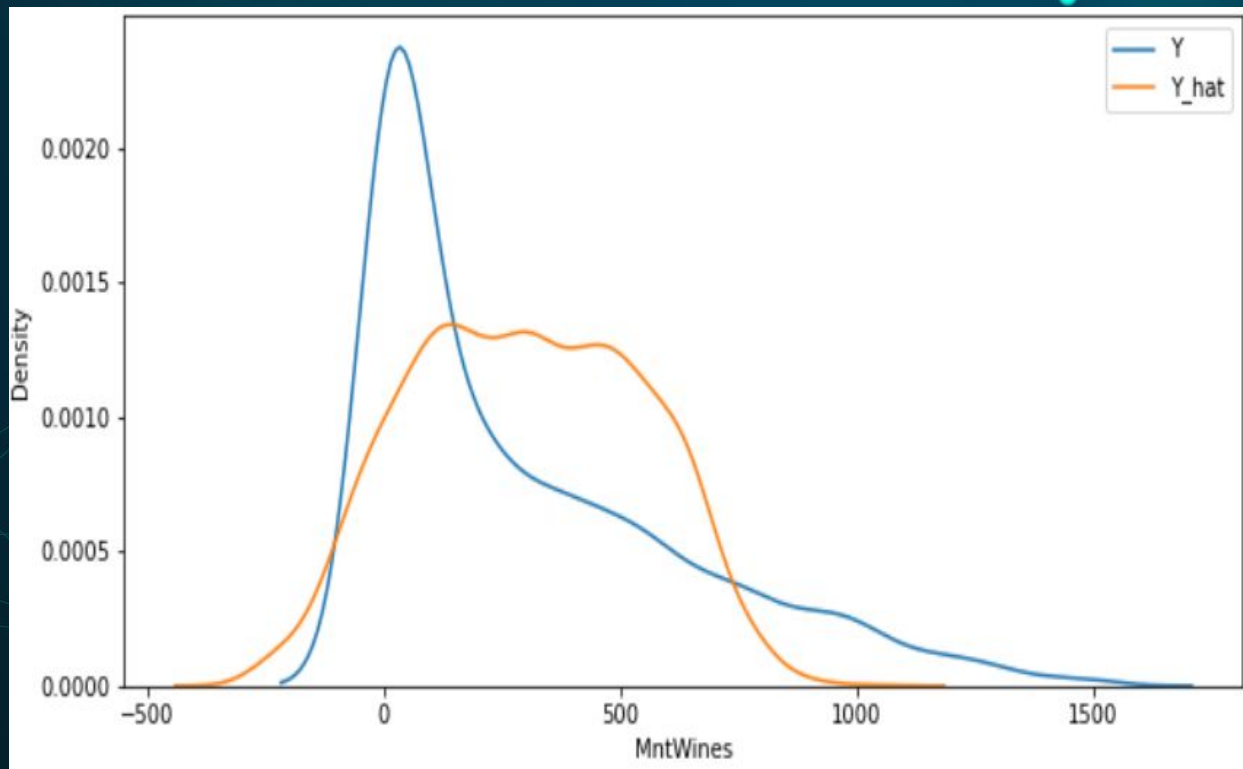
	Year_Birth	Income	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
Year_Birth	1.000000	-0.161791	-0.019871	-0.157773	-0.017917	-0.030872	-0.041625	-0.018133	-0.061818
Income	-0.161791	1.000000	-0.003970	0.578650	0.430842	0.584633	0.438871	0.440744	0.325916
Recency	-0.019871	-0.003970	1.000000	0.016064	-0.004306	0.023056	0.001079	0.022670	0.016693
MntWines	-0.157773	0.578650	0.016064	1.000000	0.389637	0.562667	0.399753	0.386581	0.387516
MntFruits	-0.017917	0.430842	-0.004306	0.389637	1.000000	0.543105	0.594804	0.567164	0.392995
MntMeatProducts	-0.030872	0.584633	0.023056	0.562667	0.543105	1.000000	0.568402	0.523846	0.350609
MntFishProducts	-0.041625	0.438871	0.001079	0.399753	0.594804	0.568402	1.000000	0.579870	0.422875
MntSweetProducts	-0.018133	0.440744	0.022670	0.386581	0.567164	0.523846	0.579870	1.000000	0.369724
MntGoldProds	-0.061818	0.325916	0.016693	0.387516	0.392995	0.350609	0.422875	0.369724	1.000000

구매량,
연간소득
회귀예측

연봉으로 와인 구매량 예측해보기

단순선형 회귀분석

단순회귀모델 회귀식 : $0.01174903793208133 X + -305.302452918962$
단순회귀모델 결정계수 (R^2) : 0.5434817898758919

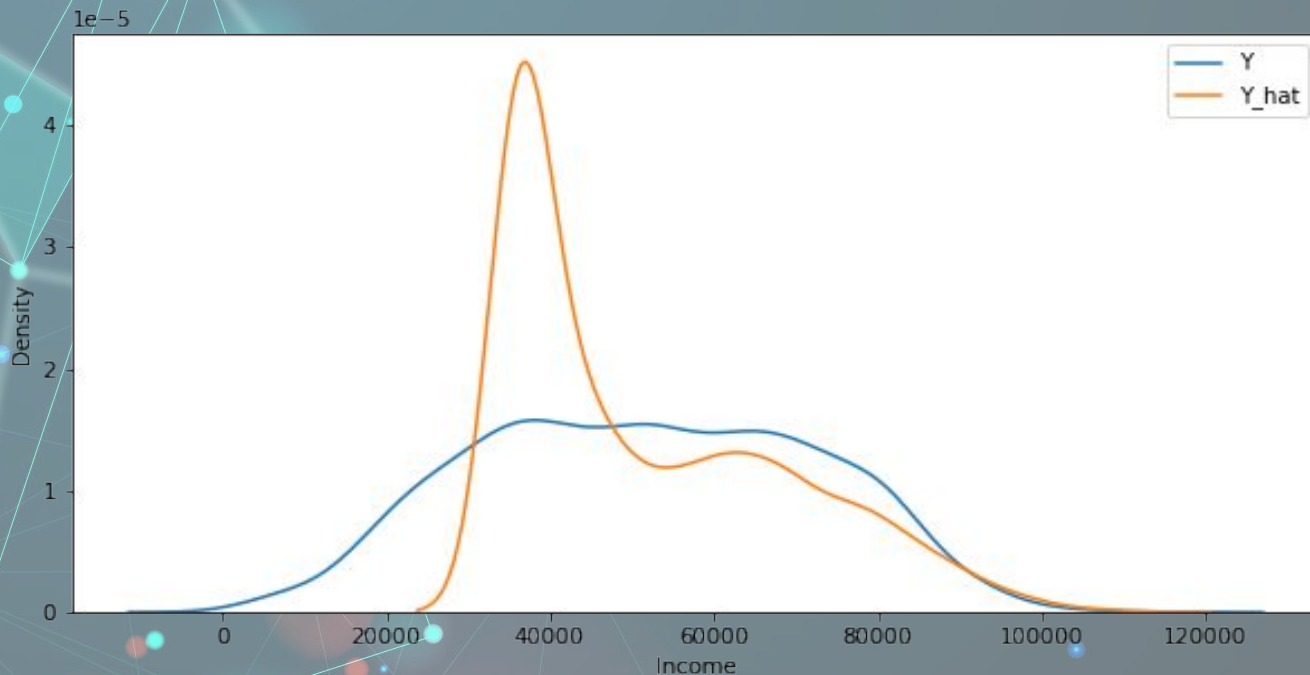


6가지 구매량에
대한 데이터로
연봉을
예측해보자!

다중선형회귀분석(multiple linear regression)

회귀식 : $[28.4455187 \ 41.99462071 \ 22.83120889 \ 33.6484627 \ 67.23256408 \ 8.04863078] X_n + 34603.69606372753$

결정계수 (R^2) : 0.6916989426753315



결과해석

(28.4455187 * 와인 구매량 +

41.99462071* 과일 구매량+

22.83120889* 고기 구매량+

33.6484627* 생선구매량+

67.23256408* 단음식 구매량+

8.04863078* 금 구매량+)

34603.69606372753 (Y절편) >> 새로운 고객의 연간소득 예측 가능!

결정계수(R^2):

이 6가지 품목에 대한 데이터로 연봉데이터를 0.6916989426753315 만큼 설명하는 것이 가능하다.

캠페인 참여여부 분류예측

어떤 결과를 얻고 싶은데?

새로운 고객이 들어왔을때 이 고객이 캠페인에 참여할지에 대해 분류하려고 한다.

1~5번째 까지 캠페인 참여여부데이터와 연간소득, 자녀여부, 마지막 구매 후 일수,고객의 첫 등록으로부터 지금까지 날짜 데이터를 통해

가장 최근 캠페인에 참여여부를(1~5번 이외 다른 캠페인) 분류해보자!

어떤 모델? DecisionTree

"AcceptedCmp1", "AcceptedCmp2", "AcceptedCmp3", "AcceptedCmp4", "AcceptedCmp5", "Income", "Kidhome", "Teenhome", "days", "Recency"



"Response"

confusion_matrix

```
[[565  16]  
 [ 77 14]]
```

	precision	recall	f1-score	support
0	0.88	0.97	0.92	581
1	0.47	0.15	0.23	91
accuracy			0.86	672
macro avg	0.67	0.56	0.58	672
weighted avg	0.82	0.86	0.83	672

Treeplot 링크

[https://github.com/minseochoi-AI/BreakPython_Data-analysis-study/blob/main/Untitled%20\(1\).ipynb](https://github.com/minseochoi-AI/BreakPython_Data-analysis-study/blob/main/Untitled%20(1).ipynb)

감사합니다~