Consumer Profiles: Regression

Gabe Anoia, Harrison Hubbard,
Levi Sessions

Customer Personality Analysis Dataset

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

- 2,450 samples over 28 features
- Customer demographic, past spending, responses to marketing
- Data collected from undisclosed European retail company
- Collected from sales data and customer survey responses

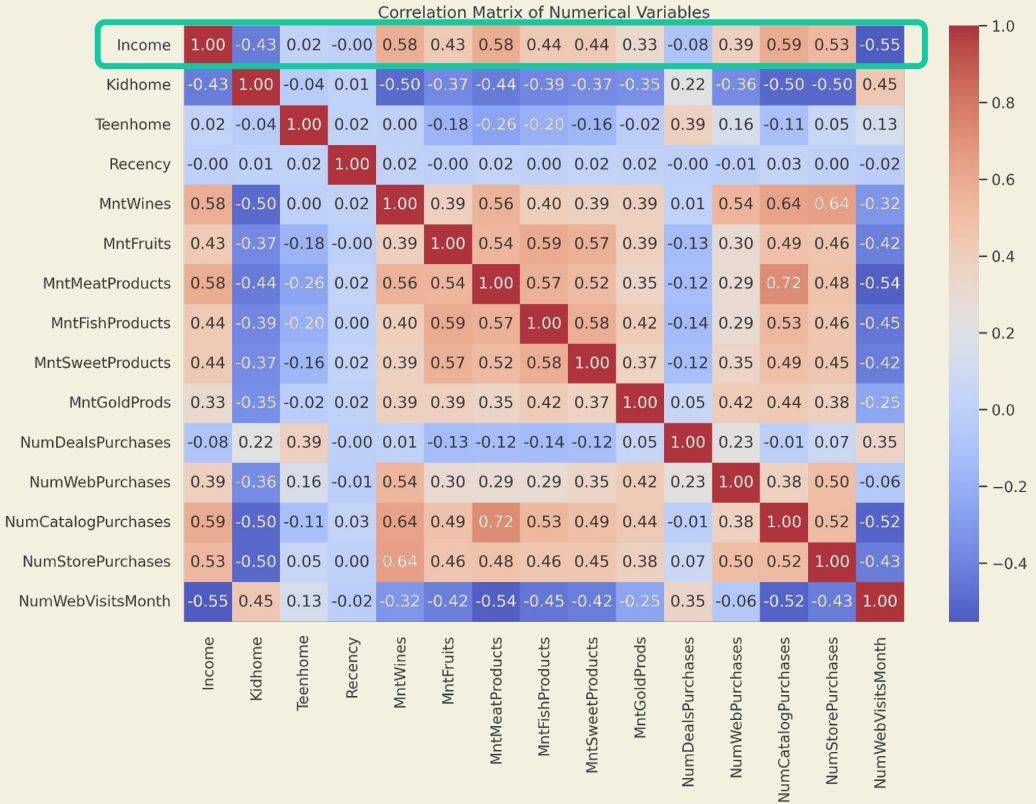
Regression Analysis

- Statistical methodology of estimating the relationship between a dependent variable and one or more independent variables
- Finds linear equation that most closely fits existing data
- Used to describe relationships and predict dependent variables from unseen data

Our Goals

- Create and analyse regression models to:
  - Describe the relationships of customer income with their demographic information and past spending
  - Find underlying interactions between different factors
  - Find a model to predict customer income from different factors

Variable Selection



Numerical Variables:

- Examine correlation between covariates and Income
- Choose those with “high” correlation (>|.30|)

Variable Selection

Variable Name	Description	Variable Type
Birth_Year	Customer's birth year	Numerical
Education	Customer's highest level of education obtained	Categorical
Marital_Status	Customer's marital status	Categorical
Income	Customer's yearly household income	Numerical
Kidhome	Number of children in customer's household	Numerical
Teenhome	Number of teenagers in customer's household	Numerical
Dt_Customer	Date of customer's enrollment with the business	Categorical
Recency	Number of days since customer's last purchase	Numerical
Complain	Whether or not customer has complained in the last two years	Categorical
MntWinesProducts	Amount spent on wine in last two years	Numerical
MntFruitsProducts	Amount spent on fruits in last two years	Numerical
MntMeatProducts Spending	Amount spent on meat in last two years	Numerical
MntFishProducts Spending	Amount spent on fish in last two years	Numerical
MntSweetProducts Spending	Amount spent on sweets in last two years	Numerical
MntGoldProds Spending	Amount spent on gold in last two years	Numerical
NumDealPurchases	Number of purchases made with a discount applied	Numerical
AcceptedCpn$N$ (1-5)	Whether or not customer accepted the offer in the Nth marketing campaign	Categorical
Response	Whether or not customer accepted the offer in the latest marketing campaign	Categorical
NumWebPurchases	Number of purchases made through the business's website	Numerical
NumCatalogPurchases	Number of purchases made using a catalogue	Numerical
NumStorePurchases	Number of purchases made directly in stores	Numerical
NumWebsiteVisitsMonth	Number of visits to business's website in the last month	Numerical

Model Creation

Model Type	Variables Used
Simple Linear	MntWines, MntMeat, MntFish, MntSweet, MntGold, NumCatPurch, NumStorPurch, NumWebVisit
Base ML	All Relevant Variables
Multiple Linear	M1: MntWines, MntMeat, MntFish, MntFruit M2: MntWines, MntMeat, NumCatPurch, NumWebVisit M3: M2 + Teenhome
ML Comparison	All 3 ML Models

Simple Linear

Highly Correlated Variables

The spending categories and the purchasing types were of higher r-values, showing correlation with Income.

Base Multiple Linear

All Variables

Base model for multiple linear regression was made using all of the main variables in the study.

Multiple Linear

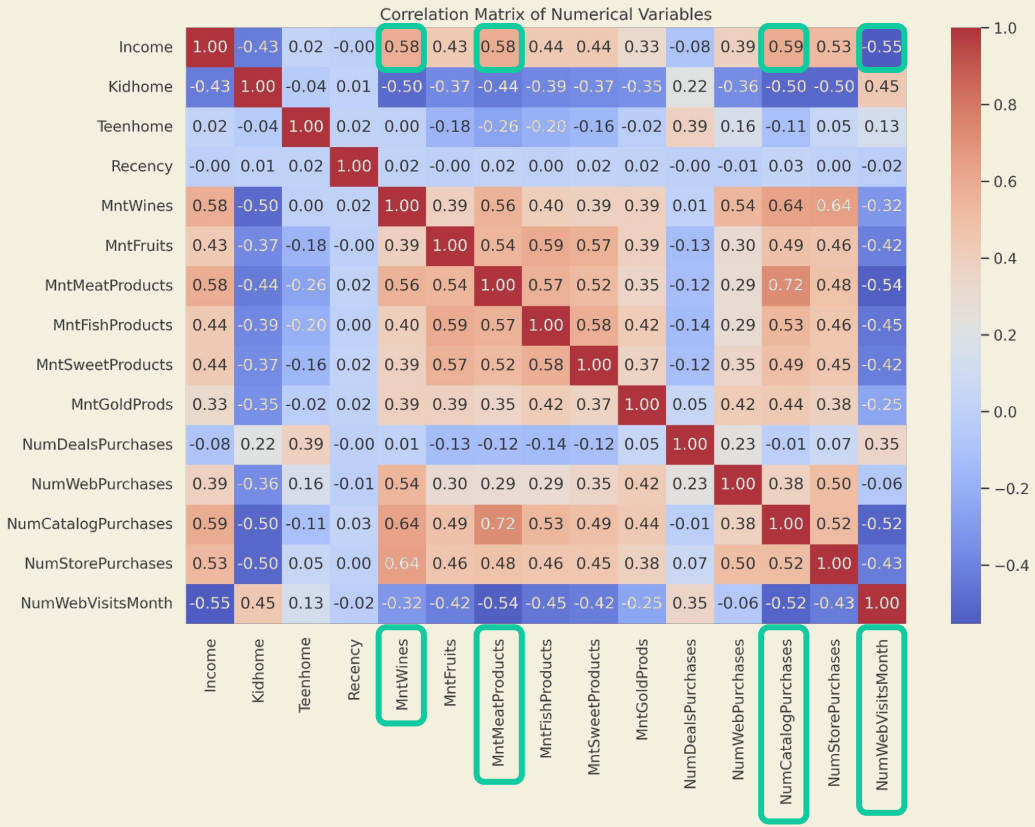
Variable Groups

Uses 3 different models based on spending options, highly correlated variables, and highly correlated variables accounting for home life.

ML Comparison

Comparison Models

Uses ANOVA tables, interaction terms, and log-transformed models to compare all models to see which best explains Income.



Choosing Covariates

- Pick four variables with highest correlation
- MntWines
- MntMeatProducts
- NumCatalogPurchases
- NumWebVisitsMonth

Income ~ MntWines

```
Residuals:
   Min     1Q Median     3Q    Max
-40749  -9937  -1291   8259 120070

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.860e+04  4.476e+02   86.22   <2e-16 ***
MntWines    4.388e+01  9.837e-01   44.60   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15610 on 2210 degrees of freedom
Multiple R-squared:  0.4738,    Adjusted R-squared:  0.4735
F-statistic:  1990 on 1 and 2210 DF,  p-value: < 2.2e-16
```

Intercept: **38,600**

- Base salary

Coefficient: **43.88**

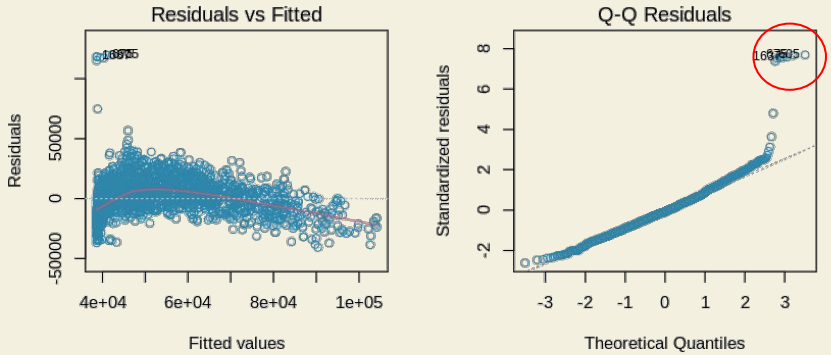
- +43.88 euros/euro spent on wine

R^2: **0.4738**

Both terms show significance

F-statistic p-value shows significance

Income ~ MntWines



- Slight curve in Residuals vs Fitted plot
- Notable deviation on right side of Q-Q plot
  - Skewed-ness in distribution
- Scale-Location plot shows mostly uniform shape
- No points close to Cook's distance in Residuals vs Leverage plot

Income ~ MntMeatProducts

```
Residuals:
   Min     1Q Median     3Q    Max
-39043 -10074   -282  10280 120731

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     40567.333    403.758  100.47   <2e-16 ***
MntMeatProducts    68.652      1.455   47.17   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15190 on 2210 degrees of freedom
Multiple R-squared:  0.5017,    Adjusted R-squared:  0.5015
F-statistic:  2225 on 1 and 2210 DF,  p-value: < 2.2e-16
```

Intercept: **40,567.33**

- Base salary

Coefficient: **68.65**

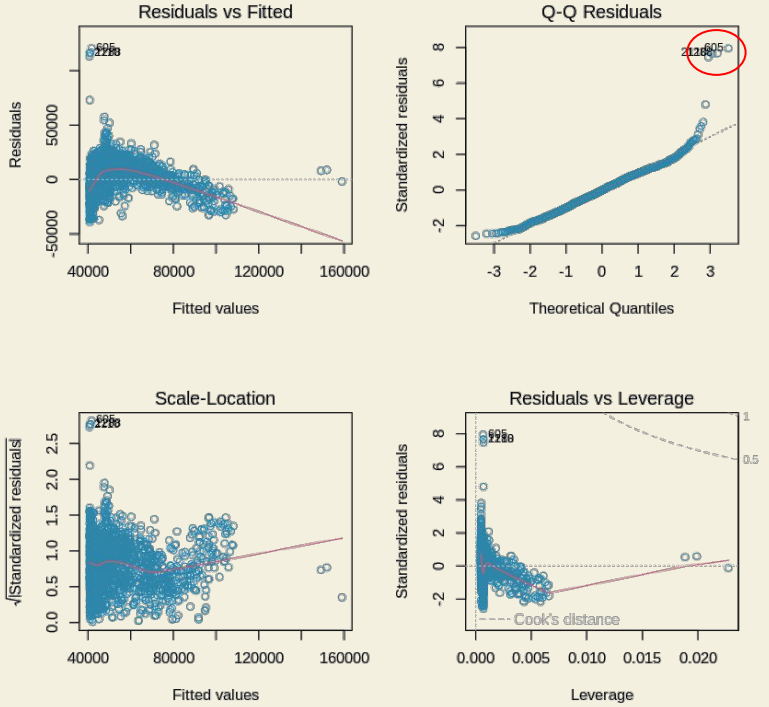
- +68.65 euros/euro spent on meat

R^2: **0.5017**

Both terms show significance

F-statistic p-value shows significance

Income ~ MntMeatProducts



- Curve in Residuals vs Fitted plot
- Notable deviation on right side of Q-Q plot
  - Skewed-ness in distribution
- Scale-Location plot shows slight cone shape
- No clear high-leverage points in Residuals vs Leverage plot

Income ~ NumCatalogPurchases

```
Residuals:
   Min     1Q Median     3Q    Max
-47305  -9599    117   9776 124699

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         37698.3      433.4   86.99   <2e-16 ***
NumCatalogPurchases  5371.1      110.5   48.59   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14960 on 2210 degrees of freedom
Multiple R-squared:  0.5165,    Adjusted R-squared:  0.5163
F-statistic:  2361 on 1 and 2210 DF,  p-value: < 2.2e-16
```

Intercept: **37,698.30**

- Base salary

Coefficient: **5,371.10**

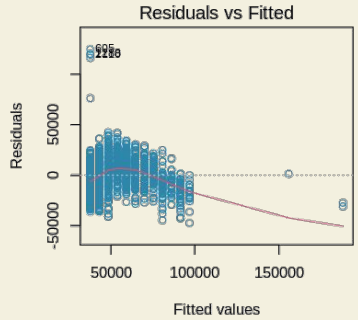
- +5,371.10 euros/each purchase via catalog

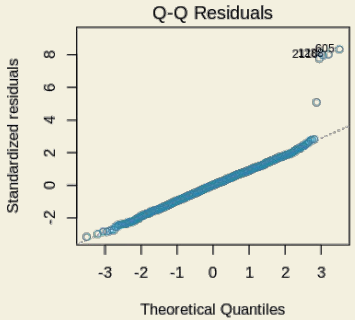
R^2: **0.5165**

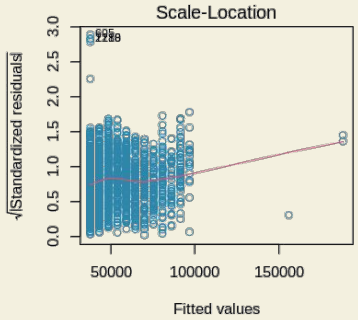
Both terms show significance

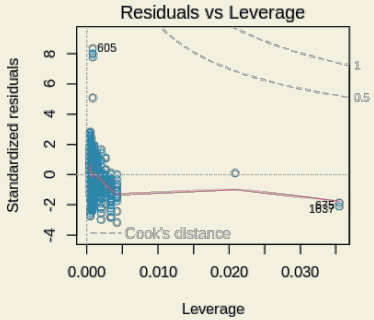
F-statistic p-value shows significance

Income ~ NumCatalogPurchases









- Some downward curve in Residuals vs Fitted plot
- Deviation on upper right side of Q-Q plot
  - Skewed-ness in distribution
- Scale-Location plot shows mostly uniform distribution
- No clear high-leverage points in Residuals vs Leverage plot

Income ~ NumWebVisitsMonth

```
Residuals:
   Min     1Q Median     3Q    Max
-75711 -10614    718  10080  85345

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       82855.0      834.8   99.26   <2e-16 ***
NumWebVisitsMonth  -5802.6      142.8  -40.63   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16280 on 2210 degrees of freedom
Multiple R-squared:  0.4275,    Adjusted R-squared:  0.4273
F-statistic:  1651 on 1 and 2210 DF,  p-value: < 2.2e-16
```

Intercept: **82,855.00**

- Base salary

Coefficient: **-5,802.60**

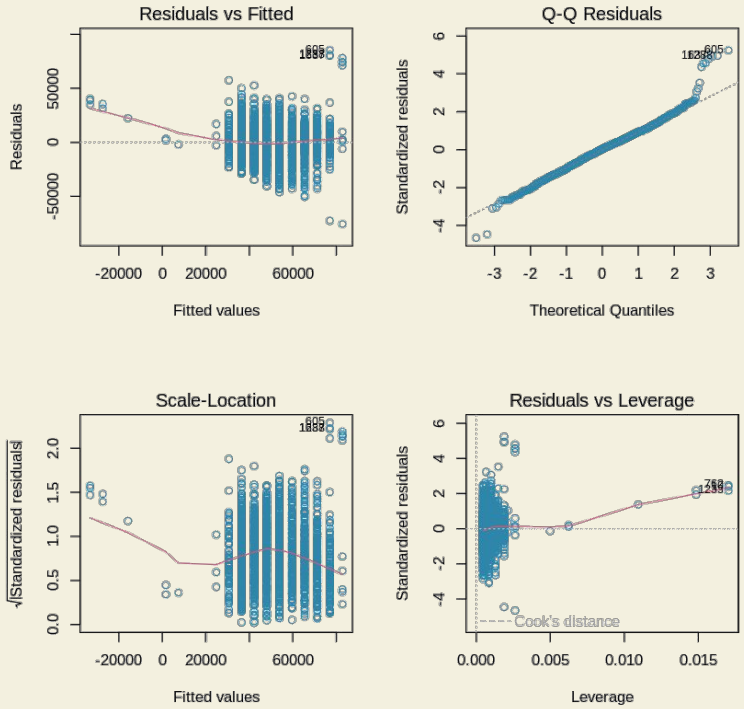
- -5,802.60 euros/website visit in a month

R^2: **0.4275**

Both terms show significance

F-statistic p-value shows significance

Income ~ NumWebVisitsMonth



- Slight upward curve in Residuals vs Fitted plot
- Deviation on upper right side of Q-Q plot
- Scale-Location plot shows uniform distribution
- No clear high-leverage points in Residuals vs Leverage plot

“Base” Model

```
Residuals:
   Min     1Q Median     3Q    Max
-76638  -5805   -135   5477 108928

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             50256.450   7376.772   6.813 1.23e-11 ***
EducationBasic         -10448.429   1581.847  -6.605 4.97e-11 ***
EducationGraduation       758.418    786.927   0.964 0.335267
EducationMaster          1255.406    913.429   1.374 0.169462
EducationPhD             2146.400    893.553   2.402 0.016384 *
Marital_StatusDivorced  -2888.356   7300.836  -0.396 0.692424
Marital_StatusMarried   -3433.096   7274.632  -0.472 0.637026
Marital_StatusSingle    -3840.489   7280.916  -0.527 0.597918
Kidhome                  1975.078    545.277   3.622 0.000299 ***
Teenhome                 5895.962    475.663  12.395  < 2e-16 ***
Recency                   -10.397      7.483  -1.389 0.164873
MntWines                   14.818      1.047  14.154  < 2e-16 ***
MntFruits                  13.510      7.567   1.786 0.074312 .
MntMeatProducts            22.138      1.661  13.328  < 2e-16 ***
MntFishProducts             4.165      5.750   0.724 0.468982
MntSweetProducts           27.121      7.291   3.720 0.000204 ***
MntGoldProds               -7.078      5.138  -1.378 0.168474
NumDealsPurchases        -515.367    145.162  -3.550 0.000393 ***
NumWebPurchases           999.450    109.659   9.114  < 2e-16 ***
NumCatalogPurchases      1018.332    130.341   7.813 8.61e-15 ***
NumStorePurchases         378.049    101.500   3.725 0.000201 ***
NumWebVisitsMonth       -2983.893    129.025 -23.127  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10150 on 2190 degrees of freedom
Multiple R-squared:  0.7794,    Adjusted R-squared:  0.7773
F-statistic: 368.5 on 21 and 2190 DF,  p-value: < 2.2e-16
```

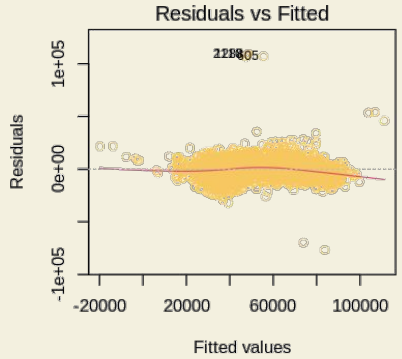
- All numerical variables + Education and Marital_Status
- Four highest correlated variables are significant

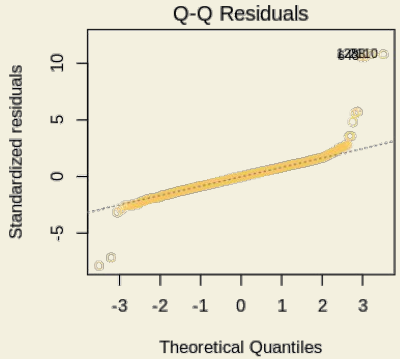
Adj. R^2: **0.7773**

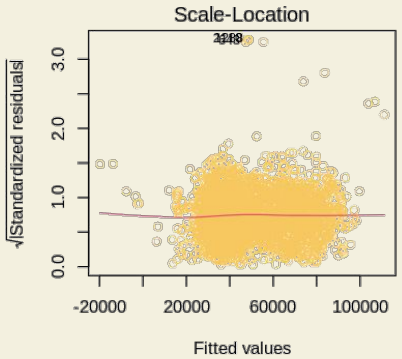
RSE: **10150**

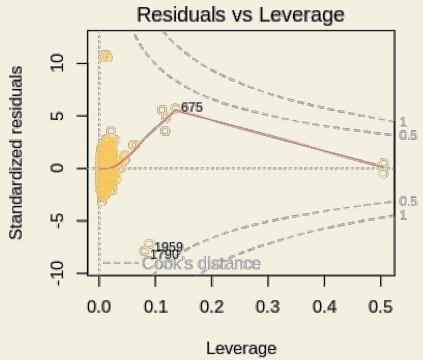
p-value: **<0.05**

“Base” Model









- Residuals vs Fitted plot shows linear, straight line
- Q-Q Residuals shows tight fit to line, except endpoints
- No clear cone shape in Scale-Location plot
- A few observations close to Cook’s distance line in Residuals vs Leverage plot

Model 1: Income ~ Product Spending

```
Residuals:
   Min     1Q Median     3Q    Max
-34442  -8050      5   7701 123996

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     35526.731    390.434  90.993  < 2e-16 ***
MntWines           25.595      1.013  25.272  < 2e-16 ***
MntFruits          50.052      9.150   5.470 5.01e-08 ***
MntMeatProducts    36.163      1.796  20.132  < 2e-16 ***
MntFishProducts    34.813      6.796   5.123 3.27e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13030 on 2207 degrees of freedom
Multiple R-squared:  0.6338,    Adjusted R-squared:  0.6332
F-statistic:   955 on 4 and 2207 DF,  p-value: < 2.2e-16
```

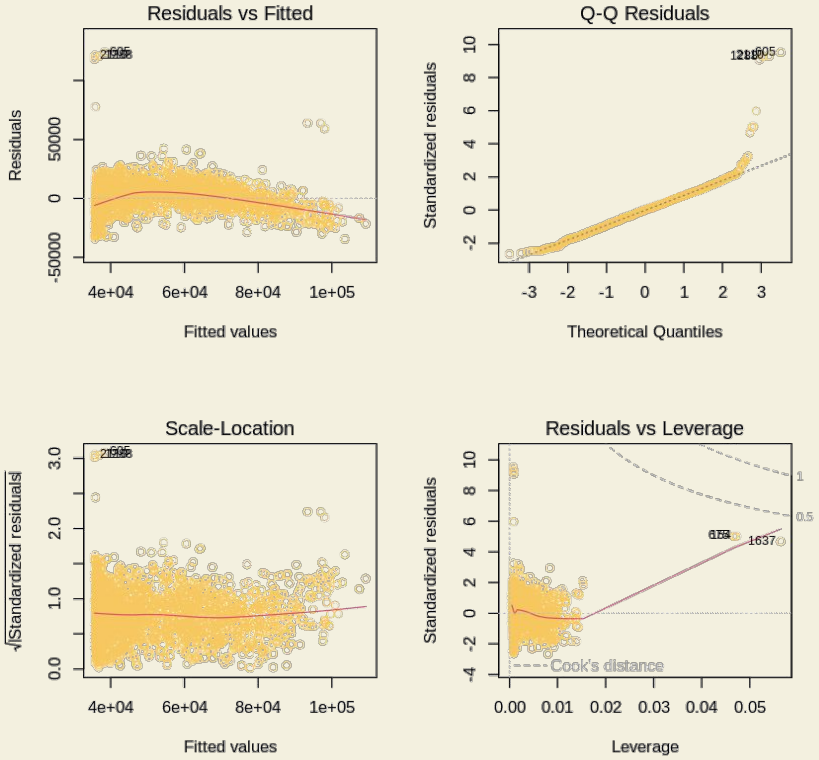
- Left out gold due to low correlation and insignificance in base model
- Meat and wine show smaller p-values than fruit or fish

Adj. R^2: **0.6332**

RSE: **13030**

p-value: **<0.05**

Model 1: Income ~ Product Spending



- Some curve in Residuals vs Fitted plot
- Upper tail of Q-Q plot deviates
- Mostly uniform distribution on Scale-Location plot
- Some potential high leverage points on Residuals vs Leverage plot

Model 2: Income ~ Top 4 Correlated

```
Residuals:
   Min     1Q Median     3Q    Max
-51597  -6780     -2   6783 108179

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         55060.6253   852.5051   64.59   <2e-16 ***
MntWines               22.7904     0.9531   23.91   <2e-16 ***
MntMeatProducts        19.0570     1.6672   11.43   <2e-16 ***
NumCatalogPurchases  1205.5327   136.3679    8.84   <2e-16 ***
NumWebVisitsMonth   -3084.7342   120.5364  -25.59   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11220 on 2207 degrees of freedom
Multiple R-squared:  0.7286,  Adjusted R-squared:  0.7281
F-statistic:  1481 on 4 and 2207 DF,  p-value: < 2.2e-16
```

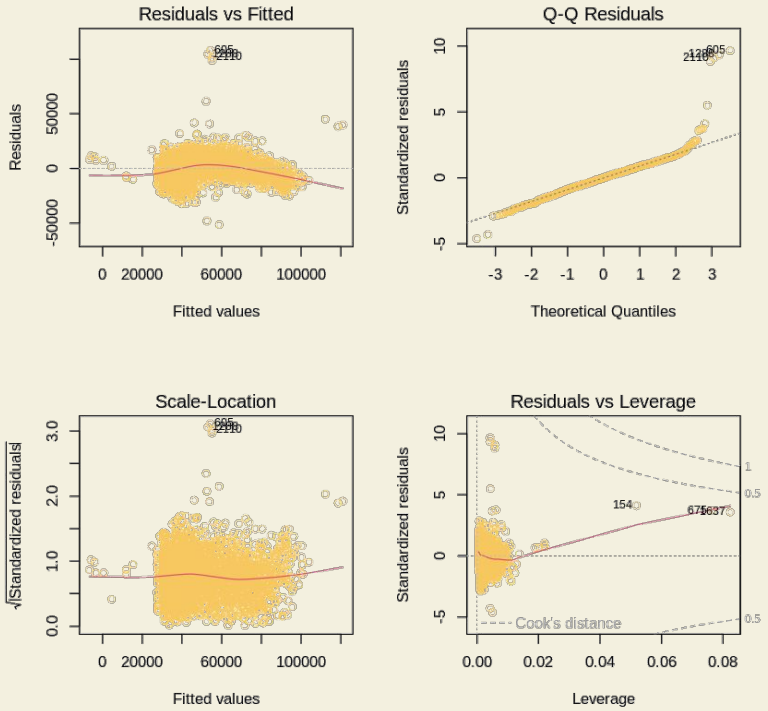
- All of the top 4 highest correlation values
- All of the values are significant

Adj. R^2: **0.7286**

RSE: **11220**

p-value: **<0.05**

Model 2: Income ~ Top 4 Correlated



- Slight curve in Residuals vs Fitted plot
- Upper tail of Q-Q plot deviates, some on lower tail as well
- Mostly uniform distribution on Scale-Location plot
- Some high leverage points on Residuals vs Leverage plot

Model 3: Income ~ Top 4 Correlated + Teens

```
Residuals:
   Min     1Q Median     3Q    Max
-61014  -6074     -5   6041 108259

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         51521.458    848.883  60.693   <2e-16 ***
MntWines               20.608      0.922  22.351   <2e-16 ***
MntMeatProducts        25.903      1.659  15.617   <2e-16 ***
NumCatalogPurchases  1123.999    130.311   8.626   <2e-16 ***
NumWebVisitsMonth   -3084.627    115.078 -26.805   <2e-16 ***
Teenhome             6494.103    442.552  14.674   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10710 on 2206 degrees of freedom
Multiple R-squared:  0.7527,    Adjusted R-squared:  0.7521
F-statistic:  1343 on 5 and 2206 DF,  p-value: < 2.2e-16
```

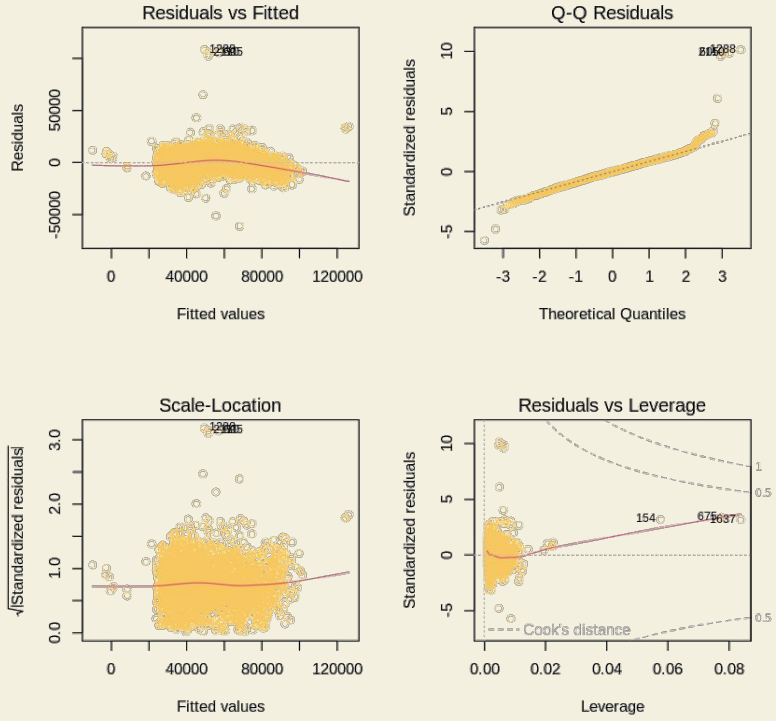
- Four factors covered in simple models + teens in home
- No change in RSE or Adj R^2 removing kids in home

Adj. R^2: **0.7521**

RSE: **10710**

p-value: **<0.05**

Model 3: Income ~ Top 4 Correlated + Teens



- Very little curve in Residuals vs Fitted plot
- Upper tail of Q-Q plot deviates
- Mostly uniform distribution on Scale-Location plot
- Some high leverage points on Residuals vs Leverage plot

Model Comparison

Base Model

All Covariates

Adj. R^2:	**0.7773**
RSE:	**10150**
p-value:	**<0.05**

- First model used as baseline for comparison
- High adjusted r-squared value
- P-value indicates statistical significance

Model 1

Spending Types

Adj. R^2:	**0.6332**
RSE:	**13030**
p-value:	**<0.05**

- Model of all spending options
- Lower adjusted r-squared than baseline
- Higher RSE value than baseline
- P-value indicates statistical significance

Model 2

Top 4 Highest Correlation Values

Adj. R^2:	**0.7286**
RSE:	**11220**
p-value:	**<0.05**

- Model of highest correlation
- Higher r-squared than M1, but less than baseline
- Lower RSE than M1, but higher than baseline
- P-value is significant

Model 3

Correlation Values with TeenHome

Adj. R^2:	**0.7521**
RSE:	**10710**
p-value:	**<0.05**

- Model of highest correlation with teenhome
- Highest adjusted r-squared of the 3 models
- Lowest RSE of the 3, with a significant p-value
- Very close to baseline model

Income ~ MntWines * MntMeatProducts

```
Residuals:
   Min     1Q Median     3Q    Max
-32927  -7393   -147   6691 124638

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.304e+04  3.943e+02   83.81   <2e-16 ***
MntMeatProducts           7.907e+01  2.224e+00   35.55   <2e-16 ***
MntWines                  4.181e+01  1.211e+00   34.52   <2e-16 ***
MntMeatProducts:MntWines -7.410e-02  3.712e-03  -19.96   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12240 on 2208 degrees of freedom
Multiple R-squared:  0.6768,    Adjusted R-squared:  0.6764
F-statistic:  1541 on 3 and 2208 DF,  p-value: < 2.2e-16
```

- Interaction between meat and wine purchases is significant
- In context of our chosen model?

Adj. R^2: **0.6764**

RSE: **12240**

p-value: **<0.05**

Model 3 + MntWines * MntMeatProducts

```
Residuals:
   Min     1Q Median     3Q    Max
-58444  -5918    -62   5491 108612

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                5.036e+04  8.065e+02  62.437   <2e-16 ***
MntWines                   3.505e+01  1.252e+00  27.995   <2e-16 ***
MntMeatProducts            5.622e+01  2.453e+00  22.920   <2e-16 ***
NumCatalogPurchases        2.224e+02  1.355e+02   1.642    0.101
NumWebVisitsMonth         -3.090e+03  1.089e+02 -28.374   <2e-16 ***
Teenhome                   5.231e+03  4.261e+02  12.277   <2e-16 ***
MntWines:MntMeatProducts -5.554e-02  3.453e-03 -16.083   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10140 on 2205 degrees of freedom
Multiple R-squared:  0.7787,    Adjusted R-squared:  0.7781
F-statistic:  1293 on 6 and 2205 DF,  p-value: < 2.2e-16
```

- NumCatalogPurchases no longer significant
  - Removing it

Adj. R^2: **0.7781**

RSE: **10140**

p-value: **<0.05**

Model 4: Without NumCatalogPurchases

```
Residuals:
   Min     1Q Median     3Q    Max
-58631  -5789     62   5522 108287

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 5.066e+04  7.858e+02   64.46   <2e-16 ***
MntWines                    3.619e+01  1.044e+00   34.67   <2e-16 ***
MntMeatProducts             5.857e+01  1.997e+00   29.33   <2e-16 ***
NumWebVisitsMonth          -3.129e+03  1.062e+02  -29.46   <2e-16 ***
Teenhome                    5.204e+03  4.259e+02   12.22   <2e-16 ***
MntWines:MntMeatProducts -5.788e-02  3.145e-03  -18.41   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10140 on 2206 degrees of freedom
Multiple R-squared:  0.7784,    Adjusted R-squared:  0.7779
F-statistic:  1550 on 5 and 2206 DF,  p-value: < 2.2e-16
```

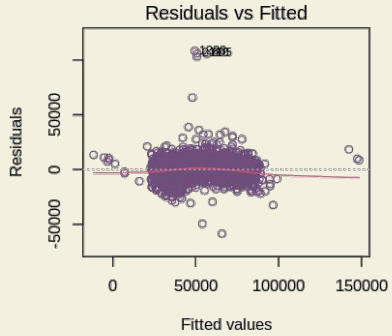
- All factors are significant
- Adj. R^2 matches that of our base model
- RSE is as small as our base model

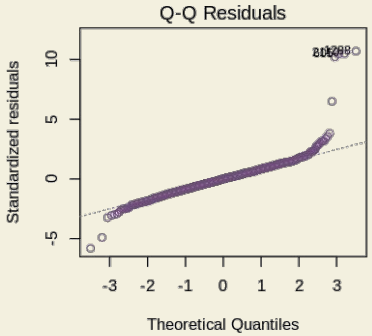
Adj. R^2: **0.7781**

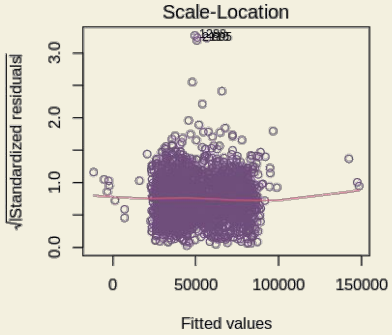
RSE: **10140**

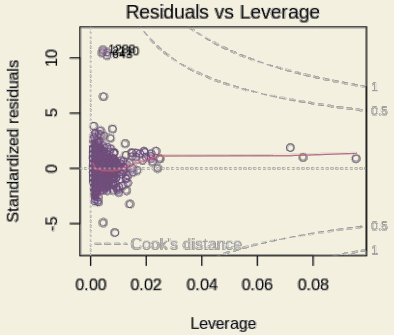
p-value: **<0.05**

Model 4: Without NumCatalogPurchases









- Very little curve in Residuals vs Fitted plot
- Upper and lower tails of Q-Q plot deviates
- Mostly uniform distribution on Scale-Location plot
- Some deviated points on Residuals vs Leverage plot, but none crossing Cook's distance

ANOVA Comparison: Model 3 (no interaction) vs. Model 3 (with interaction)

- Adding the interaction term **significantly improves** the model.
- The p-value is extremely small which is a statistically significant difference between the two models
- The very low p-value suggests that the interaction term MntMeats * MntWines significantly improves the model's ability to explain the variation in Income

Model	Residual DF	RSS	F	p-value
No Interaction	2207	2.616e+12	—	—
With Interaction	2206	2.268e+12	338.79	<2e-16 ***

Income ~ Model 4 (Log-transformed)

```
Call:
lm(formula = Income_log ~ MntWines_log * MntMeatProducts_log +
    NumWebVisitsMonth_scaled + Teenhome_scaled, data = CustomerData)

Residuals:
     Min       1Q   Median       3Q      Max
-2.38346 -0.10450  0.00755  0.12904  1.61738

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     9.819984   0.039302 249.859  < 2e-16 ***
MntWines_log                    0.164666   0.010441  15.771  < 2e-16 ***
MntMeatProducts_log             0.066127   0.013880   4.764 2.02e-06 ***
NumWebVisitsMonth_scaled       -0.178638   0.006217 -28.734  < 2e-16 ***
Teenhome_scaled                 0.045825   0.006180   7.415 1.73e-13 ***
MntWines_log:MntMeatProducts_log -0.005062   0.002488  -2.035    0.042 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2491 on 2206 degrees of freedom
Multiple R-squared:  0.7517,  Adjusted R-squared:  0.7511
F-statistic:  1335 on 5 and 2206 DF,  p-value: < 2.2e-16
```

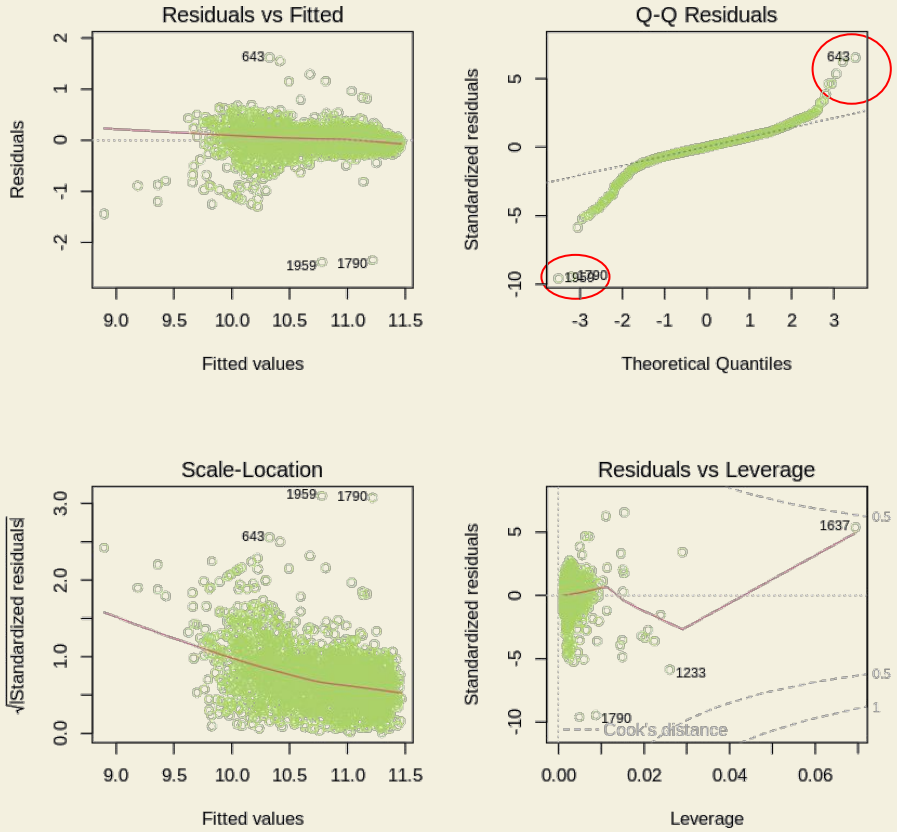
- All of the top 4 highest correlation values log-transformed
- When scaled, NCP loses its significance

Adj. R^2: **0.7511**

RSE: **0.2491**

p-value: **<0.05**

Income ~ Model 4 (Log Transformed)



- Straight line in Residuals vs Fitted plot
- Both tails of Q-Q plot deviate heavily
- Slight cone shape on Scale-Location plot
- Some high leverage points on Residuals vs Leverage plot and points past Cook's Distance

Our Goals

Find a linear regression model that:

- Meaningfully describes customer income
    - Has high degree of variance explained by model
- Uses various demographic and behavioral factors
- Meets LINE assumptions

Our Findings

- Base model:
  - High R^2 and low RSE
  - 16 factors: looking to reduce
- Using product spending categories:
  - Some categories non-significant
  - Lower R^2 and higher RSE than base model
- Using four most correlated variables:
  - Better than using solely product spending categories
  - Adding number of teens in home improved model

Our Findings

- Checked interaction between spending on meat and wine
  - Provided statistically significant improvement to model vs without interaction term
  - Improved model to be on par with base model
    - **4 factors (+ interaction) vs 16 factors**
- Verified that log transformation does **not** improve fit of model via residual plots

Our Chosen Model

Income ~ MntWines * MntMeatProducts + NumWebVisitsMonth + Teenhome

```
Residuals:
   Min     1Q Median     3Q    Max
-58631  -5789     62   5522 108287

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               5.066e+04  7.858e+02   64.46   <2e-16 ***
MntWines                  3.619e+01  1.044e+00   34.67   <2e-16 ***
MntMeatProducts           5.857e+01  1.997e+00   29.33   <2e-16 ***
NumWebVisitsMonth        -3.129e+03  1.062e+02  -29.46   <2e-16 ***
Teenhome                  5.204e+03  4.259e+02   12.22   <2e-16 ***
MntWines:MntMeatProducts -5.788e-02  3.145e-03  -18.41   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10140 on 2206 degrees of freedom
Multiple R-squared:  0.7784,    Adjusted R-squared:  0.7779
F-statistic:  1550 on 5 and 2206 DF,  p-value: < 2.2e-16
```