Lab2_Modeling

2023-12-04

Warning: NAs introduced by coercion

Imports

```
train <- read_csv("./data/processed/train_data.csv")
test<- read_csv("./data/processed/test_data.csv")
```

Basic Model

```
model1 <- lm(total_sales ~ is_vegetarian + ingredeint_count + month_02 + month_03 + month_04 + month_05
summary(model1)
```

```
##
## Call:
## lm(formula = total_sales ~ is_vegetarian + ingredeint_count +
##      month_02 + month_03 + month_04 + month_05 + month_06 + month_07 +
##      month_08 + month_09 + month_10 + month_11 + month_12, data = train)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -1440.80  -573.82   -18.15   541.22  1776.07
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2400.727    253.496   9.470  < 2e-16 ***
## is_vegetarianTRUE  -447.247    103.957  -4.302  2.4e-05 ***
## ingredeint_count    -15.052     33.375  -0.451    0.652
## month_02            -96.828    217.330  -0.446    0.656
## month_03             20.428    233.367   0.088    0.930
## month_04            -99.689    230.853  -0.432    0.666
## month_05            195.278    230.625   0.847    0.398
## month_06            -89.485    233.457  -0.383    0.702
## month_07            -27.790    215.570  -0.129    0.898
## month_08           -162.410    233.280  -0.696    0.487
## month_09           -229.793    221.060  -1.040    0.300
## month_10            -65.517    230.257  -0.285    0.776
## month_11             -3.689    219.600  -0.017    0.987
## month_12           -148.226    223.415  -0.663    0.508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 745.9 on 260 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.0961, Adjusted R-squared:  0.05091
## F-statistic: 2.126 on 13 and 260 DF,  p-value: 0.01317
```

Advanced Model

```
model2 <- lm(total_sales ~ meat_count + ingredeint_count + alt_sauce + alt_cheese + month_02 + month_03
summary(model2)
```

```
##
## Call:
## lm(formula = total_sales ~ meat_count + ingredeint_count + alt_sauce +
##     alt_cheese + month_02 + month_03 + month_04 + month_05 +
##     month_06 + month_07 + month_08 + month_09 + month_10 + month_11 +
##     month_12, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1298.72  -631.46    40.38   532.35  1690.85
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2258.63     269.54   8.380 3.45e-15 ***
## meat_count          148.15      51.72   2.864  0.00452 **
## ingredeint_count    -29.05      36.31  -0.800  0.42444
## alt_sauceTRUE       182.08     130.35   1.397  0.16366
## alt_cheeseTRUE     -259.17     109.17  -2.374  0.01833 *
## month_02            -98.11     215.71  -0.455  0.64963
## month_03             26.35     231.89   0.114  0.90961
## month_04            -89.10     229.13  -0.389  0.69769
## month_05            251.34     229.35   1.096  0.27415
## month_06            -69.81     232.16  -0.301  0.76388
## month_07            -37.13     214.00  -0.174  0.86238
## month_08           -130.67     231.75  -0.564  0.57334
## month_09           -197.69     219.65  -0.900  0.36894
## month_10            -38.25     228.69  -0.167  0.86730
## month_11            -25.40     217.86  -0.117  0.90727
## month_12           -155.95     221.80  -0.703  0.48262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 740 on 258 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.1171, Adjusted R-squared:  0.06576
## F-statistic: 2.281 on 15 and 258 DF,  p-value: 0.004805
```