WHICH FACTORS CONTRIBUTE MOST TO USER SATISFACTION OF PHARMACEUTICAL DRUGS

Vrushali Patel

Vallesia Pierre Louis

Ryan Chen

DATA DESCRIPTION AND MOTIVATION

MOTIVATION

Motivation -Prescription and over the counter drug are used everyday, we hope to find out the factors related to medicine that contribute the most to user satisfaction as consumers

DATA DESCRIPTION

- Dataset- Drug Performance Evaluation

- The predictors of this dataset are the condition, the drug name , the ease of use, the effectiveness, form, indication, price, reviews, satisfaction, and type.

- The data was collected through performance evaluation, on how different variables affect patient evaluation.

- The diverse variables are a valuable tool for satisfaction.


BASIC
DATA
ANALYSIS

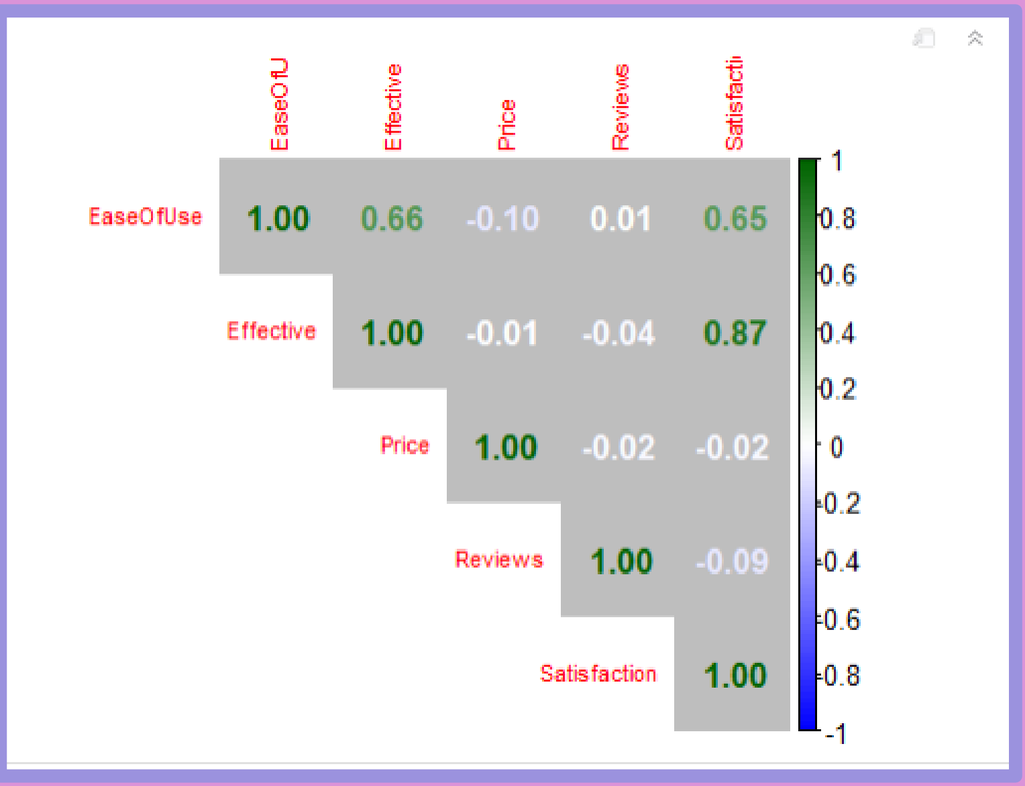
DATA CLEANING

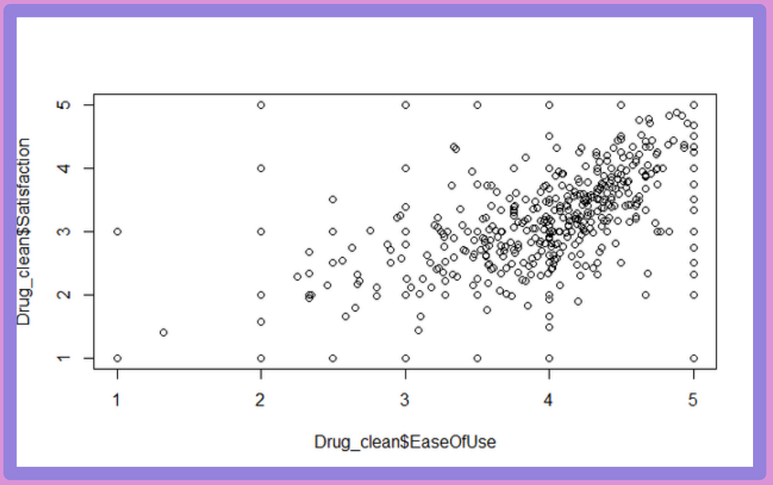
- The variables, Condition and Drug Name will be removed.
- Omit missing values
- Removed “other” from Indication and Type
- Indication and Type were turned into dummy variables

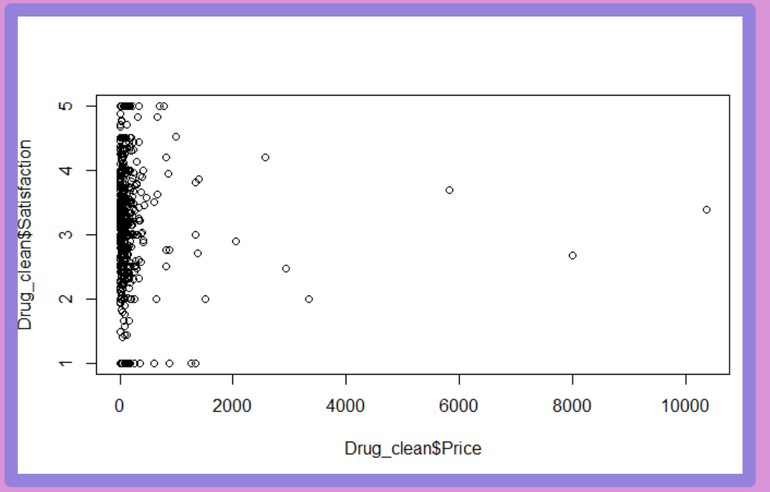
```
'data.frame':   645 obs. of  8 variables:
 $ EaseOfUse   : num  3.85 3.47 3.12 2 3.25 ...
 $ Effective   : num  3.66 3.29 2.96 3 3 ...
 $ Form        : chr  "Capsule" "Liquid (Drink)" "Tablet"
"Capsule" ...
 $ Indication  : Factor w/ 2 levels "Off Label","On Label": 2
2 2 1 2 2 2 2 ...
 $ Price       : num  12.6 287.4 70.6 12.6 125.2 ...
 $ Reviews     : num  86.3 43 267.3 1 15 ...
 $ Satisfaction: num  3.2 2.59 2.25 1 3 ...
 $ Type        : Factor w/ 2 levels "OTC","RX": 2 2 2 2 2 2 2
2 ...
```

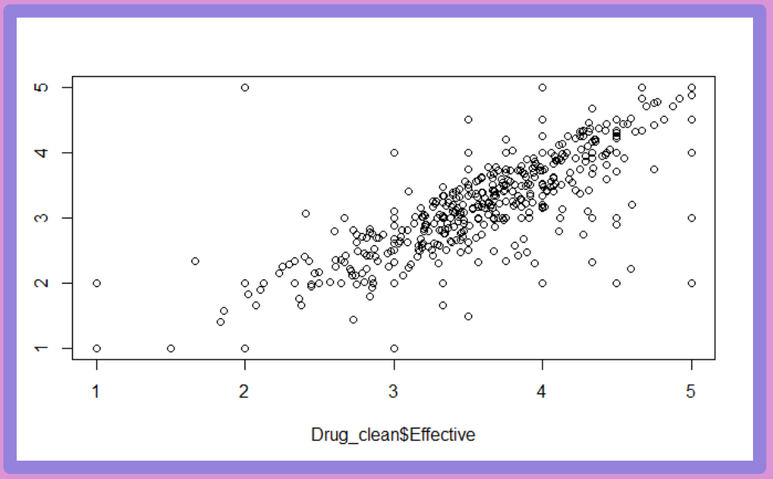
OUTLIERS

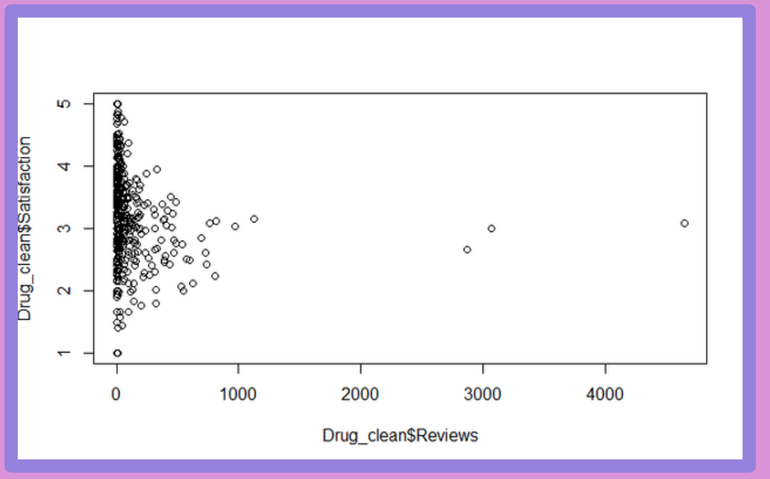
Correlation PLot











OUTLIERS (PART 2)

- All cook's Distance values were below 1
- Jackknife value is 1.963707
- Leverage threshold: 0.02790698 with 74 outliers
- studentized residuals have 34 outliers
- 8 in common

```
[1] "Indices of potential outliers based on leverage:"
 39  41  55  62  74  78  84  93  99 104 108 114 115 119 129 140 155 177 179 180 183 184 207 208 223 225 229 235 236 238 255 259 261 264 272 274 287 290 294 325 326 335 337 340 347 349 357 366 375 380
 38  40  52  57  69  73  77  86  92  97 101 107 108 112 122 133 148 170 172 173 176 177 197 198 213 215 219 225 226 228 245 249 251 254 262 264 276 279 283 309 310 319 321 324 331 333 341 349 358 363
381 397 424 447 462 464 466 468 477 487 514 515 516 552 577 592 623 627 630 659 671 677 683 685
364 378 399 421 435 436 437 438 446 456 481 482 483 517 542 557 588 592 595 622 633 638 643 645
[1] "Indices of potential outliers based on studentized residuals"
  4  71  72  83 102 108 146 149 179 183 189 194 251 255 256 259 290 293 294 297 301 303 306 320 327 359 372 476 478 479 515 556 600 617
  4  66  67  76  95 101 139 142 172 176 182 187 241 245 246 249 279 282 283 286 290 292 295 307 311 343 355 445 447 448 482 521 565 582
[1] "common outliers"
[1] 101 172 176 245 249 279 283 482
```

MODEL DIAGNOSTIC

Collinearity

All VIF values below 10

```
 [1] 5.000000 1.920784 1.843834 1.762109 1.593556 1.385924
 [7] 1.357879 1.262361 1.135076 1.065399 1.058287 1.055687
[13] 1.027466 1.018473 1.009194 1.000000 1.000000 1.000000
[19] 1.000000 1.000000 1.000000
```

NORMALITY

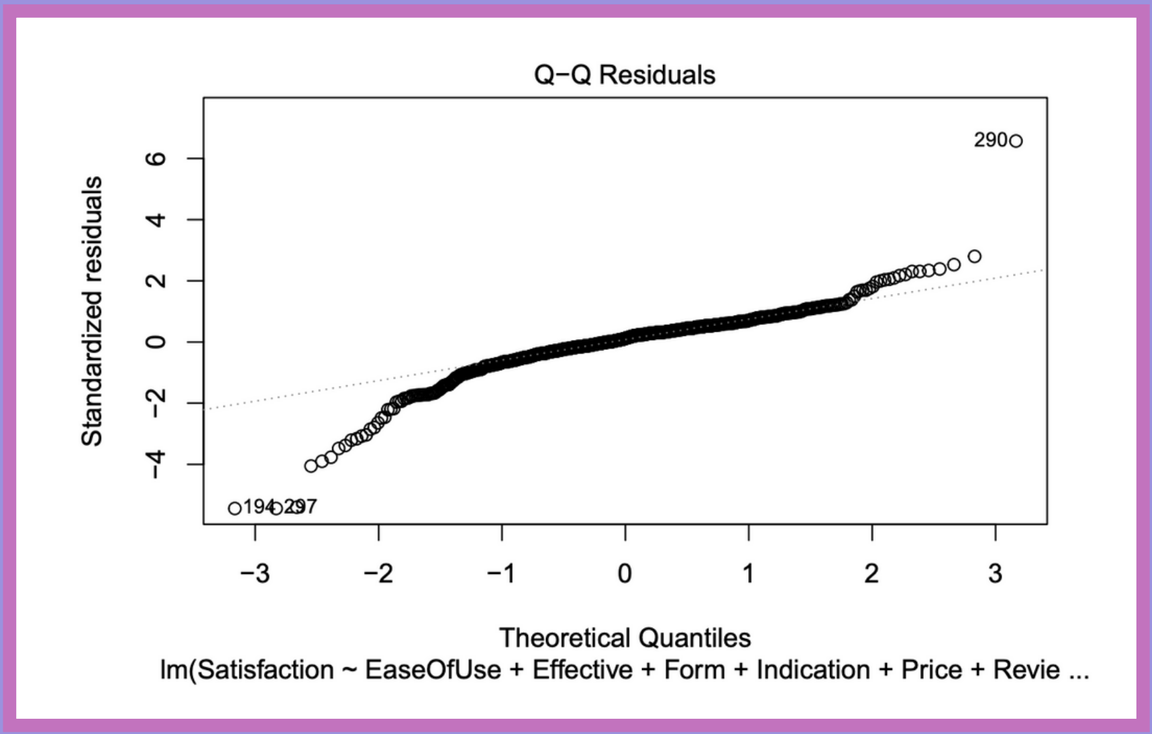
After Shapiro Test, we decided to transform our model, but our data did not improve enough to justify it.

```
Call:
lm(formula = (Satisfaction)^2 ~ EaseOfUse + Effective + Form +
    Indication + Price + Reviews + Type, data = Drug_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-16.6100  -1.6269  -0.1366   2.0796  20.5742

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        -9.2925724  0.9136996 -10.170  < 2e-16 ***
EaseOfUse           0.8545328  0.2080862   4.107 4.54e-05 ***
Effective           4.9592018  0.1907103  26.004  < 2e-16 ***
FormCream           1.8874461  0.5854144   3.224  0.00133 **
FormLiquid (Drink)  1.6921420  0.5761295   2.937  0.00343 **
FormLiquid (Inject) 1.4762462  0.6434006   2.294  0.02209 *
FormOther           1.6213164  0.7093004   2.286  0.02260 *
FormTablet         -0.1474817  0.4657920  -0.317  0.75163
IndicationOn Label  0.4716583  0.3453862   1.366  0.17255
Price              -0.0001051  0.0002247  -0.468  0.64017
Reviews            -0.0012340  0.0004938  -2.499  0.01271 *
TypeRX             -1.6752496  0.3857167  -4.343 1.63e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.425 on 633 degrees of freedom
Multiple R-squared:  0.7316,    Adjusted R-squared:  0.727
```

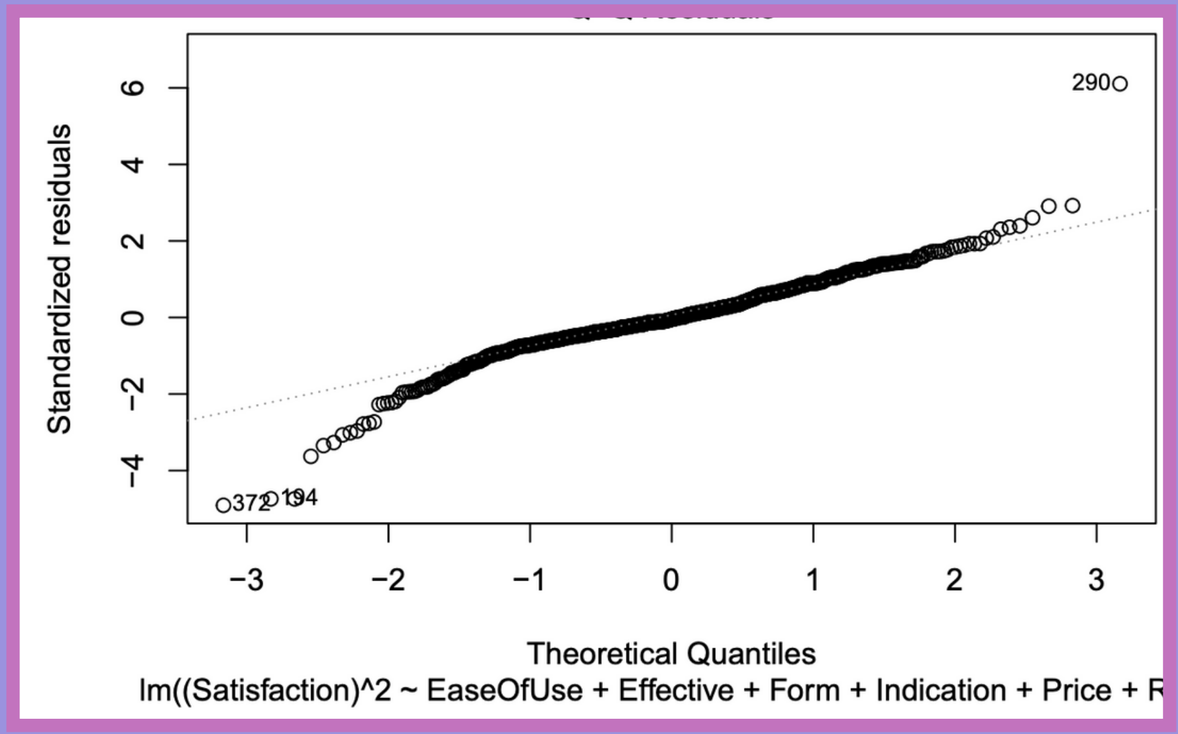
```
Residual standard error: 3.425 on 633 degrees of freedom
Multiple R-squared:  0.7316,    Adjusted R-squared:  0.727
F-statistic: 156.9 on 11 and 633 DF,  p-value: < 2.2e-16
```

```
        Shapiro-Wilk normality test

data:  model$residuals
W = 0.88328, p-value < 0.00000000000000022
```

Normality



Transformed



MODEL SELECTION

Decided to use Backwards elimination because of our F-statistic and P-Value.

```
## Residual standard error: 3.425 on 633 degrees of freedom
## Multiple R-squared:  0.7316, Adjusted R-squared:  0.727
## F-statistic: 156.9 on 11 and 633 DF,  p-value: < 2.2e-16
```

Model Selection

```
Call:
lm(formula = (Satisfaction)^2 ~ EaseOfUse + Effective + Form +
    Indication + Price + Reviews + Type, data = Drug_clean)

Residuals:
     Min       1Q   Median       3Q      Max
-16.6100  -1.6269  -0.1366   2.0796  20.5742

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -9.2925724  0.9136996 -10.170  < 2e-16 ***
EaseOfUse           0.8545328  0.2080862   4.107 4.54e-05 ***
Effective           4.9592018  0.1907103  26.004  < 2e-16 ***
FormCream           1.8874461  0.5854144   3.224  0.00133 **
FormLiquid (Drink)  1.6921420  0.5761295   2.937  0.00343 **
FormLiquid (Inject) 1.4762462  0.6434006   2.294  0.02209 *
FormOther           1.6213164  0.7093004   2.286  0.02260 *
FormTablet         -0.1474817  0.4657920  -0.317  0.75163
IndicationOn Label  0.4716583  0.3453862   1.366  0.17255
Price              -0.0001051  0.0002247  -0.468  0.64017
Reviews            -0.0012340  0.0004938  -2.499  0.01271 *
TypeRX             -1.6752496  0.3857167  -4.343 1.63e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.425 on 633 degrees of freedom
Multiple R-squared:  0.7316, Adjusted R-squared:  0.727
F-statistic: 156.9 on 11 and 633 DF,  p-value: < 2.2e-16
```

```
AIC(model)

## [1] 916.8616
BIC(model)

## [1] 974.881
```

```
##
## Call:
## lm(formula = Satisfaction ~ EaseOfUse + Effective + Form + Indication +
##     Type, data = new_data_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65646 -0.14876  0.04888  0.22962  1.09341
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.49258    0.16733  -2.944  0.00355 **
## EaseOfUse           0.11392    0.04258   2.675  0.00796 **
## Effective           0.90651    0.03747  24.191  < 2e-16 ***
## FormCream           0.20885    0.07738   2.699  0.00743 **
## FormLiquid (Drink)  0.16081    0.07950   2.023  0.04417 *
## IndicationOn Label  0.17707    0.07396   2.394  0.01740 *
## TypeRX             -0.21956    0.06627  -3.313  0.00106 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.448 on 249 degrees of freedom
## Multiple R-squared:  0.8461, Adjusted R-squared:  0.8424
## F-statistic: 228.1 on 6 and 249 DF,  p-value: < 2.2e-16
```

Variables removed:
FormLiquid(Inject), FormOther
FormTablet, Price, Reviews

```
AIC(model_6)

## [1] 324.3032
BIC(model_6)

## [1] 352.6646
```

PRELIMINARY RESULTS

Backwards Elimination

Using backwards elimination a final model we selected consist of 6 predictors: EaseOfUse, Effective,FormCream,Form Liquid(Drink),Indication(On Label),Type(RX)

Potential Outliers

No major outliers from cooks distance, the ones from Jackknife and leverage we deemed to be feasible and decided not to remove them

No collinearity

All VIF values < 10.

Final Model

From our model we can say that the EaseOfUse, Effectiveness, Form(cream and liquid), Indication(on Label), Type(Rx) factors that contribute most to user satisfaction of pharmaceutical drugs

84 % of the customer's satisfaction is accounted for by Ease OfUse, Effectiveness, Form, Indication and Type.

FINAL MODEL

Satisfaction = -049258 + 0.11392(EaseOfUse) +0.90651(Effective) + 0.20885 (FormCream) + 0.16081(FormLiquid (Drink)) + 0.17707 (IndicationOn Label) - 0.21956(TypeRX)

DIFFICULTY ENCOUNTERS

Normality violation want to look back and try to find a solution

Rcode was having issues maybe due to operating system between us

RESOURCES

HIsch, Irl. “Insulin Pricing in the USA: The Saga Continues.” The Lancet, www.thelancet.com/journals/landia/article/PIIS2213-8587(22)00251-0/fulltext

“Most Expensive Drugs in the US in 2023.” Fierce Pharma, www.fiercepharma.com/special-reports/priciest-drugs-2023.

https://www.kaggle.com/datasets/thedevastator/drug-performance-evaluation