Untitled

Taylor Fourier

2024-11-10

```
# install.packages("readr")
# install.packages("dplyr")
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
ridesharedata <- read_csv("/Users/taylorfourier/Downloads/rideshare_kaggle.csv")
```

```
## Rows: 12583 Columns: 57
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (10): id, timezone, source, destination, cab_type, product_id, name, sh...
## dbl  (46): timestamp, hour, day, month, price, distance, surge_multiplier, l...
## dttm  (1): datetime
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
num_rows <- nrow(ridesharedata)
print(num_rows)
```

```
## [1] 12583
```

```
ridesharedata <- ridesharedata %>% filter(!is.na(price))

num_rows <- nrow(ridesharedata)
print(num_rows)
```

```
## [1] 11604
```

```
colnames(ridesharedata)
```

```
##  [1] "id"                          "timestamp"
##  [3] "hour"                        "day"
##  [5] "month"                       "datetime"
##  [7] "timezone"                    "source"
```

```
##  [9] "destination"                "cab_type"
## [11] "product_id"                 "name"
## [13] "price"                      "distance"
## [15] "surge_multiplier"           "latitude"
## [17] "longitude"                  "temperature"
## [19] "apparentTemperature"        "short_summary"
## [21] "long_summary"               "precipIntensity"
## [23] "precipProbability"          "humidity"
## [25] "windSpeed"                  "windGust"
## [27] "windGustTime"               "visibility"
## [29] "temperatureHigh"            "temperatureHighTime"
## [31] "temperatureLow"             "temperatureLowTime"
## [33] "apparentTemperatureHigh"    "apparentTemperatureHighTime"
## [35] "apparentTemperatureLow"     "apparentTemperatureLowTime"
## [37] "icon"                       "dewPoint"
## [39] "pressure"                   "windBearing"
## [41] "cloudCover"                 "uvIndex"
## [43] "visibility.1"               "ozone"
## [45] "sunriseTime"                "sunsetTime"
## [47] "moonPhase"                  "precipIntensityMax"
## [49] "uvIndexTime"                "temperatureMin"
## [51] "temperatureMinTime"         "temperatureMax"
## [53] "temperatureMaxTime"         "apparentTemperatureMin"
## [55] "apparentTemperatureMinTime" "apparentTemperatureMax"
## [57] "apparentTemperatureMaxTime"
```

```r
ridesharedata$datetime <- as.POSIXct(ridesharedata$datetime, format = "%Y-%m-%d %H:%M:%S", tz = "America

ridesharedata <- ridesharedata %>%
  mutate(isweekend = ifelse(weekdays(datetime) %in% c("Saturday", "Sunday"), 1, 0))

ridesharedata <- ridesharedata %>%
  mutate(
    hour_1 = ifelse(hour == 1, 1, 0),
    hour_2 = ifelse(hour == 2, 1, 0),
    hour_3 = ifelse(hour == 3, 1, 0),
    hour_4 = ifelse(hour == 4, 1, 0),
    hour_5 = ifelse(hour == 5, 1, 0),
    hour_6 = ifelse(hour == 6, 1, 0),
    hour_7 = ifelse(hour == 7, 1, 0),
    hour_8 = ifelse(hour == 8, 1, 0),
    hour_9 = ifelse(hour == 9, 1, 0),
    hour_10 = ifelse(hour == 10, 1, 0),
    hour_11 = ifelse(hour == 11, 1, 0),
    hour_12 = ifelse(hour == 12, 1, 0),
    hour_13 = ifelse(hour == 13, 1, 0),
    hour_14 = ifelse(hour == 14, 1, 0),
    hour_15 = ifelse(hour == 15, 1, 0),
    hour_16 = ifelse(hour == 16, 1, 0),
    hour_17 = ifelse(hour == 17, 1, 0),
    hour_18 = ifelse(hour == 18, 1, 0),
    hour_19 = ifelse(hour == 19, 1, 0),
    hour_20 = ifelse(hour == 20, 1, 0),
    hour_21 = ifelse(hour == 21, 1, 0),
```

```
    hour_22 = ifelse(hour == 22, 1, 0),
    hour_23 = ifelse(hour == 23, 1, 0)
  )

ridesharedata <- ridesharedata %>%
  mutate(across(short_summary, ~ as.factor(short_summary))) %>%
  mutate(
    Mostly_Cloudy = ifelse(short_summary == "Mostly Cloudy", 1, 0),
    Rain = ifelse(short_summary == "Rain", 1, 0),
    Clear = ifelse(short_summary == "Clear", 1, 0),
    Partly_Cloudy = ifelse(short_summary == "Partly Cloudy", 1, 0),
    Overcast = ifelse(short_summary == "Overcast", 1, 0),
    Light_Rain = ifelse(short_summary == "Light Rain", 1, 0),
    Foggy = ifelse(short_summary == "Foggy", 1, 0),
    Possible_Drizzle = ifelse(short_summary == "Possible Drizzle", 1, 0),
    Drizzle = ifelse(short_summary == "Drizzle", 1, 0)
  )

ridesharedata <- ridesharedata %>%
  mutate(
    is_clear = ifelse(short_summary == "Clear", 1, 0),
    is_cloudy = ifelse(short_summary %in% c("Mostly Cloudy", "Partly Cloudy", "Overcast"), 1, 0),
    is_precipitating = ifelse(short_summary %in% c("Rain", "Light Rain", "Drizzle", "Possible Drizzle")
  )

ridesharedata <- ridesharedata %>%
  mutate(
    is_uber = ifelse(grepl("Uber", cab_type), 1, 0),
    is_lyft = ifelse(grepl("Lyft", cab_type), 1, 0)
  )

ridesharedata <- ridesharedata %>%
  mutate(
    is_shared = ifelse(name == "Shared", 1, 0),
    is_lux = ifelse(name == "Lux", 1, 0),
    is_lyft = ifelse(name == "Lyft", 1, 0),
    is_lux_black_xl = ifelse(name == "Lux Black XL", 1, 0),
    is_lyft_xl = ifelse(name == "Lyft XL", 1, 0),
    is_lux_black = ifelse(name == "Lux Black", 1, 0),
    is_uberxl = ifelse(name == "UberXL", 1, 0),
    is_black = ifelse(name == "Black", 1, 0),
    is_uberx = ifelse(name == "UberX", 1, 0),
    is_wav = ifelse(name == "WAV", 1, 0),
    is_black_suv = ifelse(name == "Black SUV", 1, 0),
    is_uberpool = ifelse(name == "UberPool", 1, 0)
  )

ridesharedata <- ridesharedata %>%
  mutate(
    is_luxury = ifelse(name %in% c("Lux", "Lux Black XL", "Lux Black", "Lyft XL", "UberXL"), 1, 0)  # A
  )

ridesharedata$day_part <- cut(ridesharedata$hour,
```

```
                                breaks = c(0, 4, 8, 12, 16, 20, 24),
                                labels = c("Early Morning", "Morning", "Afternoon", "Evening", "Night",
                                right = FALSE)

ridesharedata <- ridesharedata %>%
  mutate(
    Early_Morning = ifelse(day_part == "Early Morning", 1, 0),
    Morning = ifelse(day_part == "Morning", 1, 0),
    Afternoon = ifelse(day_part == "Afternoon", 1, 0),
    Evening = ifelse(day_part == "Evening", 1, 0),
    Night = ifelse(day_part == "Night", 1, 0),
    Late_Night = ifelse(day_part == "Late Night", 1, 0)
  )

ridesharedata1 <- ridesharedata %>%
  select(-id, -timestamp, -hour, -day, -month, -datetime, -timezone, -source, -destination, -latitude,

colnames(ridesharedata1)
```

```
##  [1] "price"              "distance"           "surge_multiplier"
##  [4] "temperature"        "precipIntensity"    "humidity"
##  [7] "windSpeed"          "uvIndex"            "precipIntensityMax"
## [10] "isweekend"          "Mostly_Cloudy"      "Rain"
## [13] "Clear"              "Partly_Cloudy"      "Overcast"
## [16] "Light_Rain"         "Foggy"              "Possible_Drizzle"
## [19] "Drizzle"            "is_clear"           "is_cloudy"
## [22] "is_precipitating"   "is_uber"            "is_lyft"
## [25] "is_shared"          "is_lux"             "is_lux_black_xl"
## [28] "is_lyft_xl"         "is_lux_black"       "is_uberxl"
## [31] "is_black"           "is_uberx"           "is_wav"
## [34] "is_black_suv"       "is_uberpool"        "is_luxury"
## [37] "day_part"           "Early_Morning"      "Morning"
## [40] "Afternoon"          "Evening"            "Night"
## [43] "Late_Night"
```

```
range_price <- range(ridesharedata1$price, na.rm = TRUE)
range_price
```

```
## [1]  2.5 80.0
```

```
# Fit a linear regression model with no ride_service_type i.e.black, xl, wave
model <- lm(price ~ distance +
              surge_multiplier +
              temperature +
              precipIntensity +
              humidity +
              uvIndex +
              isweekend +
              is_luxury +
              is_uber +
              Morning + Afternoon + Evening + Night + Late_Night +
              (isweekend * Morning) + (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Nig
            data = ridesharedata1)
# Summary of the model
summary(model)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + temperature +
##     precipIntensity + humidity + uvIndex + isweekend + is_luxury +
##     is_uber + Morning + Afternoon + Evening + Night + Late_Night +
##     (isweekend * Morning) + (isweekend * Afternoon) + (isweekend *
##     Evening) + (isweekend * Night) + (isweekend * Late_Night),
##     data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.456  -5.398  -2.916   4.455  41.656
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -14.03928    0.96524 -14.545   <2e-16 ***
## distance               2.73628    0.06316  43.325   <2e-16 ***
## surge_multiplier      18.61479    0.74520  24.979   <2e-16 ***
## temperature            0.01100    0.01341   0.820   0.4121
## precipIntensity        5.27278    3.01742   1.747   0.0806 .
## humidity               0.04585    0.70123   0.065   0.9479
## uvIndex                0.17651    0.23031   0.766   0.4435
## isweekend              0.48242    0.39112   1.233   0.2174
## is_luxury              8.28232    0.16762  49.411   <2e-16 ***
## is_uber                3.23503    0.16598  19.491   <2e-16 ***
## Morning                0.05713    0.29620   0.193   0.8471
## Afternoon              0.01888    0.30218   0.062   0.9502
## Evening               -0.27851    0.31217  -0.892   0.3723
## Night                  0.10731    0.34830   0.308   0.7580
## Late_Night            -0.03736    0.29685  -0.126   0.8999
## isweekend:Morning     -0.18204    0.55766  -0.326   0.7441
## isweekend:Afternoon   -0.78615    0.56463  -1.392   0.1638
## isweekend:Evening     -0.51665    0.56879  -0.908   0.3637
## isweekend:Night       -0.49785    0.56016  -0.889   0.3742
## isweekend:Late_Night   0.39268    0.55056   0.713   0.4757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.64 on 11584 degrees of freedom
## Multiple R-squared:  0.3103, Adjusted R-squared:  0.3092
## F-statistic: 274.3 on 19 and 11584 DF,  p-value: < 2.2e-16
```

```
#STEPWISE
stepwise_model <- step(model, direction = "both")
```

```
## Start:  AIC=47212.02
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     humidity + uvIndex + isweekend + is_luxury + is_uber + Morning +
##     Afternoon + Evening + Night + Late_Night + (isweekend * Morning) +
##     (isweekend * Afternoon) + (isweekend * Evening) + (isweekend *
##     Night) + (isweekend * Late_Night)
##
##                        Df Sum of Sq    RSS   AIC
## - humidity              1         0 676209 47210
## - isweekend:Morning     1         6 676215 47210
```

```
## - isweekend:Late_Night  1        30 676238 47211
## - uvIndex               1        34 676243 47211
## - temperature           1        39 676248 47211
## - isweekend:Night       1        46 676255 47211
## - isweekend:Evening     1        48 676257 47211
## - isweekend:Afternoon   1       113 676322 47212
## <none>                              676209 47212
## - precipIntensity       1       178 676387 47213
## - is_uber               1     22176 698384 47584
## - surge_multiplier      1     36424 712633 47819
## - distance              1    109572 785780 48953
## - is_luxury             1    142520 818729 49429
##
## Step:  AIC=47210.03
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     uvIndex + isweekend + is_luxury + is_uber + Morning + Afternoon +
##     Evening + Night + Late_Night + isweekend:Morning + isweekend:Afternoon +
##     isweekend:Evening + isweekend:Night + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - isweekend:Morning      1         6 676215 47208
## - isweekend:Late_Night   1        30 676239 47209
## - uvIndex                1        34 676243 47209
## - isweekend:Night        1        46 676255 47209
## - isweekend:Evening      1        49 676258 47209
## - temperature            1        53 676262 47209
## - isweekend:Afternoon    1       115 676324 47210
## <none>                              676209 47210
## - precipIntensity        1       205 676414 47212
## + humidity               1         0 676209 47212
## - is_uber                1     22176 698385 47582
## - surge_multiplier       1     36424 712633 47817
## - distance               1    109579 785788 48951
## - is_luxury              1    142522 818731 49427
##
## Step:  AIC=47208.13
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     uvIndex + isweekend + is_luxury + is_uber + Morning + Afternoon +
##     Evening + Night + Late_Night + isweekend:Afternoon + isweekend:Evening +
##     isweekend:Night + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - Morning                1         0 676215 47206
## - uvIndex                1        34 676250 47207
## - isweekend:Night        1        41 676256 47207
## - isweekend:Evening      1        44 676259 47207
## - temperature            1        50 676266 47207
## - isweekend:Late_Night   1        59 676274 47207
## <none>                              676215 47208
## - isweekend:Afternoon    1       119 676334 47208
## - precipIntensity        1       215 676430 47210
## + isweekend:Morning      1         6 676209 47210
## + humidity               1         0 676215 47210
## - is_uber                1     22170 698385 47580
```

```
## - surge_multiplier      1     36437 712652 47815
## - distance              1    109579 785794 48949
## - is_luxury             1    142555 818770 49426
##
## Step:  AIC=47206.14
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     uvIndex + isweekend + is_luxury + is_uber + Afternoon + Evening +
##     Night + Late_Night + isweekend:Afternoon + isweekend:Evening +
##     isweekend:Night + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - uvIndex                1        34 676250 47205
## - isweekend:Night        1        41 676256 47205
## - isweekend:Evening      1        44 676259 47205
## - temperature            1        50 676266 47205
## - isweekend:Late_Night   1        59 676274 47205
## <none>                               676215 47206
## - isweekend:Afternoon    1       119 676334 47206
## - precipIntensity        1       215 676431 47208
## + humidity               1         0 676215 47208
## + Morning                1         0 676215 47208
## - is_uber                1     22170 698385 47578
## - surge_multiplier       1     36445 712660 47813
## - distance               1    109614 785829 48947
## - is_luxury              1    142557 818772 49424
##
## Step:  AIC=47204.73
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     isweekend + is_luxury + is_uber + Afternoon + Evening + Night +
##     Late_Night + isweekend:Afternoon + isweekend:Evening + isweekend:Night +
##     isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - isweekend:Evening      1        37 676287 47203
## - isweekend:Night        1        38 676288 47203
## - temperature            1        54 676304 47204
## - isweekend:Late_Night   1        59 676309 47204
## <none>                               676250 47205
## - isweekend:Afternoon    1       119 676369 47205
## + uvIndex                1        34 676215 47206
## - precipIntensity        1       211 676460 47206
## + Morning                1         0 676250 47207
## + humidity               1         0 676250 47207
## - is_uber                1     22151 698400 47577
## - surge_multiplier       1     36433 712683 47812
## - distance               1    109625 785875 48946
## - is_luxury              1    142589 818839 49423
##
## Step:  AIC=47203.36
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     isweekend + is_luxury + is_uber + Afternoon + Evening + Night +
##     Late_Night + isweekend:Afternoon + isweekend:Night + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
```

```
## - isweekend:Night       1        19 676306 47202
## - temperature           1        54 676341 47202
## - isweekend:Afternoon   1        89 676376 47203
## - isweekend:Late_Night  1       105 676392 47203
## <none>                              676287 47203
## - Evening               1       125 676411 47204
## - precipIntensity       1       190 676477 47205
## + isweekend:Evening     1        37 676250 47205
## + uvIndex               1        28 676259 47205
## + humidity              1         0 676287 47205
## + Morning               1         0 676287 47205
## - is_uber               1     22215 698501 47576
## - surge_multiplier      1     36463 712749 47811
## - distance              1    109700 785986 48946
## - is_luxury             1    142706 818993 49423
##
## Step:  AIC=47201.7
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     isweekend + is_luxury + is_uber + Afternoon + Evening + Night +
##     Late_Night + isweekend:Afternoon + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - Night                  1        16 676322 47200
## - temperature            1        51 676357 47201
## - isweekend:Afternoon    1        75 676381 47201
## <none>                              676306 47202
## - Evening                1       128 676434 47202
## - isweekend:Late_Night   1       138 676444 47202
## - precipIntensity        1       187 676493 47203
## + uvIndex                1        27 676279 47203
## + isweekend:Night        1        19 676287 47203
## + isweekend:Evening      1        19 676288 47203
## + Morning                1         0 676306 47204
## + humidity               1         0 676306 47204
## - is_uber                1     22206 698512 47575
## - surge_multiplier       1     36477 712783 47809
## - distance               1    109704 786011 48944
## - is_luxury              1    142690 818996 49421
##
## Step:  AIC=47199.97
## price ~ distance + surge_multiplier + temperature + precipIntensity +
##     isweekend + is_luxury + is_uber + Afternoon + Evening + Late_Night +
##     isweekend:Afternoon + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - temperature            1        73 676395 47199
## - isweekend:Afternoon    1        74 676397 47199
## <none>                              676322 47200
## - isweekend:Late_Night   1       140 676462 47200
## - precipIntensity        1       171 676494 47201
## - Evening                1       187 676509 47201
## + uvIndex                1        43 676279 47201
## + isweekend:Evening      1        17 676305 47202
## + Night                  1        16 676306 47202
```

```
## + Morning               1         3 676320 47202
## + humidity              1         1 676321 47202
## - is_uber               1     22195 698517 47573
## - surge_multiplier      1     36471 712793 47807
## - distance              1    109711 786033 48942
## - is_luxury             1    142675 818997 49419
##
## Step:  AIC=47199.23
## price ~ distance + surge_multiplier + precipIntensity + isweekend +
##     is_luxury + is_uber + Afternoon + Evening + Late_Night +
##     isweekend:Afternoon + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - isweekend:Afternoon   1        70 676465 47198
## <none>                              676395 47199
## - isweekend:Late_Night  1       135 676531 47200
## + temperature           1        73 676322 47200
## + uvIndex               1        68 676327 47200
## + Night                 1        38 676357 47201
## - Evening               1       195 676591 47201
## + isweekend:Evening     1        17 676378 47201
## + Morning               1        10 676385 47201
## + humidity              1         3 676392 47201
## - precipIntensity       1       232 676627 47201
## - is_uber               1     22178 698574 47572
## - surge_multiplier      1     36458 712854 47806
## - distance              1    109672 786067 48941
## - is_luxury             1    142676 819072 49418
##
## Step:  AIC=47198.42
## price ~ distance + surge_multiplier + precipIntensity + isweekend +
##     is_luxury + is_uber + Afternoon + Evening + Late_Night +
##     isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## - Afternoon             1       107 676572 47198
## <none>                              676465 47198
## + uvIndex               1        71 676394 47199
## + isweekend:Afternoon   1        70 676395 47199
## + temperature           1        69 676397 47199
## - isweekend:Late_Night  1       184 676649 47200
## + Night                 1        37 676428 47200
## - Evening               1       199 676664 47200
## + Morning               1         9 676456 47200
## + humidity              1         5 676460 47200
## + isweekend:Evening     1         4 676461 47200
## - precipIntensity       1       237 676702 47200
## - is_uber               1     22139 698604 47570
## - surge_multiplier      1     36466 712931 47806
## - distance              1    109711 786177 48941
## - is_luxury             1    142615 819081 49416
##
## Step:  AIC=47198.26
## price ~ distance + surge_multiplier + precipIntensity + isweekend +
```

```
##     is_luxury + is_uber + Evening + Late_Night + isweekend:Late_Night
##
##                         Df Sum of Sq    RSS   AIC
## <none>                                  676572 47198
## + uvIndex                1       113 676459 47198
## + Afternoon              1       107 676465 47198
## + temperature            1       106 676466 47198
## - Evening                1       141 676713 47199
## + Night                  1        86 676486 47199
## - isweekend:Late_Night   1       183 676755 47199
## + isweekend:Evening      1         4 676568 47200
## + Morning                1         0 676572 47200
## + humidity               1         0 676572 47200
## - precipIntensity        1       244 676816 47200
## - is_uber                1     22103 698675 47569
## - surge_multiplier       1     36554 713126 47807
## - distance               1    109713 786285 48940
## - is_luxury              1    142546 819118 49415
```

```
stepwise_model
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + precipIntensity +
##     isweekend + is_luxury + is_uber + Evening + Late_Night +
##     isweekend:Late_Night, data = ridesharedata1)
##
## Coefficients:
##          (Intercept)              distance      surge_multiplier
##            -13.51788               2.73710               18.63975
##      precipIntensity             isweekend             is_luxury
##              5.42267               0.14554                8.27744
##              is_uber               Evening            Late_Night
##              3.22663              -0.29949               -0.09467
## isweekend:Late_Night
##              0.75969
```

```
# Filter the dataset for Uber
ridesharedata_uber <- ridesharedata %>%
  filter(cab_type == "Uber")

# Filter the dataset for Lyft
ridesharedata_lyft <- ridesharedata %>%
  filter(cab_type == "Lyft")
```

```
# Fit a linear regression model with no cab_type i.e. uber or lyft
model1 <- lm(price ~ distance +
               surge_multiplier +
               temperature +
               precipIntensity +
               humidity +
               uvIndex +
               isweekend +
               Mostly_Cloudy + Rain + Partly_Cloudy + Overcast + Light_Rain + Foggy + Possible_Drizzle +
               + is_lux + is_lux_black_xl + is_lyft_xl + is_lux_black + is_uberxl + is_black + is_uberx +
                Morning + Afternoon + Evening + Night + Late_Night +
```

```
              (isweekend * Morning) + (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Nig
            data = ridesharedata1)

# Summary of the model
summary(model1)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + temperature +
##     precipIntensity + humidity + uvIndex + isweekend + Mostly_Cloudy +
##     Rain + Partly_Cloudy + Overcast + Light_Rain + Foggy + Possible_Drizzle +
##     Drizzle + +is_lux + is_lux_black_xl + is_lyft_xl + is_lux_black +
##     is_uberxl + is_black + is_uberx + is_wav + is_black_suv +
##     is_uberpool + Morning + Afternoon + Evening + Night + Late_Night +
##     (isweekend * Morning) + (isweekend * Afternoon) + (isweekend *
##     Evening) + (isweekend * Night) + (isweekend * Late_Night),
##     data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.974  -1.447  -0.195   1.327  37.521
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -15.782953   0.332523 -47.464   <2e-16 ***
## distance               2.750178   0.020613 133.420   <2e-16 ***
## surge_multiplier      17.487665   0.243376  71.854   <2e-16 ***
## temperature            0.004198   0.004517   0.929   0.3527
## precipIntensity        1.478688   3.507833   0.422   0.6734
## humidity              -0.349112   0.271305  -1.287   0.1982
## uvIndex               -0.016596   0.075707  -0.219   0.8265
## isweekend              0.124924   0.130169   0.960   0.3372
## Mostly_Cloudy         -0.020645   0.086000  -0.240   0.8103
## Rain                  -0.496597   0.455735  -1.090   0.2759
## Partly_Cloudy         -0.134173   0.089146  -1.505   0.1323
## Overcast              -0.089374   0.088032  -1.015   0.3100
## Light_Rain            -0.120592   0.232025  -0.520   0.6033
## Foggy                 -0.167557   0.228221  -0.734   0.4628
## Possible_Drizzle      -0.027902   0.169793  -0.164   0.8695
## Drizzle                0.262757   0.258526   1.016   0.3095
## is_lux                 9.459167   0.099101  95.450   <2e-16 ***
## is_lux_black_xl       23.988344   0.099008 242.287   <2e-16 ***
## is_lyft_xl             7.033843   0.099224  70.888   <2e-16 ***
## is_lux_black          14.630768   0.099437 147.136   <2e-16 ***
## is_uberxl              8.073452   0.098137  82.267   <2e-16 ***
## is_black              12.984434   0.097988 132.511   <2e-16 ***
## is_uberx               2.215446   0.098083  22.588   <2e-16 ***
## is_wav                 2.219460   0.098328  22.572   <2e-16 ***
## is_black_suv          22.694781   0.098157 231.208   <2e-16 ***
## is_uberpool            1.246270   0.097991  12.718   <2e-16 ***
## Morning               -0.022071   0.099556  -0.222   0.8246
## Afternoon             -0.092746   0.104364  -0.889   0.3742
## Evening               -0.075120   0.104019  -0.722   0.4702
## Night                  0.029988   0.117187   0.256   0.7980
```

```
## Late_Night            -0.172875   0.100057  -1.728   0.0841 .
## isweekend:Morning     -0.057266   0.185058  -0.309   0.7570
## isweekend:Afternoon   -0.055388   0.188976  -0.293   0.7695
## isweekend:Evening     -0.092182   0.188463  -0.489   0.6248
## isweekend:Night        0.044011   0.186033   0.237   0.8130
## isweekend:Late_Night   0.060184   0.182107   0.330   0.7410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.492 on 11568 degrees of freedom
## Multiple R-squared:  0.9267, Adjusted R-squared:  0.9265
## F-statistic:  4180 on 35 and 11568 DF,  p-value: < 2.2e-16
```

```
# Fit a linear regression model with no cab_type i.e. uber or lyft
model2 <- lm(price ~ distance +
               surge_multiplier +
               temperature +
              humidity +
              uvIndex +
              isweekend +
              is_cloudy + is_precipitating +
               is_luxury + Morning + Afternoon + Evening + Night + Late_Night +
              (isweekend * Morning) + (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Nig
            data = ridesharedata1)

# Summary of the model
summary(model2)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + temperature +
##     humidity + uvIndex + isweekend + is_cloudy + is_precipitating +
##     is_luxury + Morning + Afternoon + Evening + Night + Late_Night +
##     (isweekend * Morning) + (isweekend * Afternoon) + (isweekend *
##     Evening) + (isweekend * Night) + (isweekend * Late_Night),
##     data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.378  -5.320  -3.220   4.892  42.457
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -10.234518   0.975542 -10.491   <2e-16 ***
## distance               2.766990   0.064172  43.118   <2e-16 ***
## surge_multiplier      16.948591   0.752466  22.524   <2e-16 ***
## temperature            0.010604   0.013805   0.768    0.442
## humidity               0.350366   0.750799   0.467    0.641
## uvIndex                0.129892   0.234435   0.554    0.580
## isweekend              0.425854   0.398680   1.068    0.285
## is_cloudy              0.039804   0.224346   0.177    0.859
## is_precipitating       0.146960   0.315057   0.466    0.641
## is_luxury              6.657007   0.147747  45.057   <2e-16 ***
## Morning               -0.009309   0.301132  -0.031    0.975
## Afternoon             -0.068932   0.304543  -0.226    0.821
```

```
## Evening               -0.347507   0.312230  -1.113    0.266
## Night                  0.020349   0.353543   0.058    0.954
## Late_Night            -0.103298   0.301172  -0.343    0.732
## isweekend:Morning     -0.087885   0.567330  -0.155    0.877
## isweekend:Afternoon   -0.606266   0.578079  -1.049    0.294
## isweekend:Evening     -0.597372   0.578195  -1.033    0.302
## isweekend:Night       -0.429944   0.570601  -0.753    0.451
## isweekend:Late_Night   0.385420   0.560408   0.688    0.492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.766 on 11584 degrees of freedom
## Multiple R-squared:  0.2875, Adjusted R-squared:  0.2863
## F-statistic:   246 on 19 and 11584 DF,  p-value: < 2.2e-16
```

```
model4 <- lm(price ~ distance +
               surge_multiplier +
               temperature +
              precipIntensity +
               humidity +
              isweekend +
               is_uber + is_luxury + (is_uber * is_luxury) +
               Morning + Afternoon + Evening + Night + Late_Night +
              (isweekend * Morning) + (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Nig
            data = ridesharedata1)
summary(model4)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + temperature +
##     precipIntensity + humidity + isweekend + is_uber + is_luxury +
##     (is_uber * is_luxury) + Morning + Afternoon + Evening + Night +
##     Late_Night + (isweekend * Morning) + (isweekend * Afternoon) +
##     (isweekend * Evening) + (isweekend * Night) + (isweekend *
##     Late_Night), data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.599  -5.605  -1.182   3.614  40.286
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -16.42526    0.89465 -18.360   <2e-16 ***
## distance               2.73541    0.05845  46.803   <2e-16 ***
## surge_multiplier      17.37273    0.69017  25.172   <2e-16 ***
## temperature            0.01214    0.01234   0.984    0.325
## precipIntensity        3.15412    2.79209   1.130    0.259
## humidity               0.07947    0.64160   0.124    0.901
## isweekend              0.34994    0.36190   0.967    0.334
## is_uber                8.27755    0.19152  43.221   <2e-16 ***
## is_luxury             13.78806    0.19913  69.242   <2e-16 ***
## Morning                0.16747    0.27402   0.611    0.541
## Afternoon             -0.08378    0.27935  -0.300    0.764
## Evening               -0.22646    0.26500  -0.855    0.393
## Night                  0.13795    0.27467   0.502    0.615
```

```
## Late_Night             -0.09302    0.27456  -0.339    0.735
## is_uber:is_luxury     -13.99147    0.31737 -44.086   <2e-16 ***
## isweekend:Morning      -0.34020    0.51604  -0.659    0.510
## isweekend:Afternoon    -0.53181    0.52245  -1.018    0.309
## isweekend:Evening      -0.39765    0.52526  -0.757    0.449
## isweekend:Night        -0.48341    0.51818  -0.933    0.351
## isweekend:Late_Night    0.66471    0.50953   1.305    0.192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.07 on 11584 degrees of freedom
## Multiple R-squared:  0.4094, Adjusted R-squared:  0.4084
## F-statistic: 422.6 on 19 and 11584 DF,  p-value: < 2.2e-16
```

```
model5 <- lm(price ~ distance +
               surge_multiplier +
               isweekend +
               (surge_multiplier * isweekend) +
               is_uber +
               is_luxury +
               (is_uber * is_luxury),
             data = ridesharedata1)
summary(model5)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + isweekend +
##     (surge_multiplier * isweekend) + is_uber + is_luxury + (is_uber *
##     is_luxury), data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.911  -5.623  -1.184   3.623  40.078
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -14.90148    0.83424 -17.862  < 2e-16 ***
## distance                     2.73880    0.05841  46.890  < 2e-16 ***
## surge_multiplier            16.39264    0.79357  20.657  < 2e-16 ***
## isweekend                   -3.87284    1.59826  -2.423  0.01540 *
## is_uber                      8.27015    0.19137  43.215  < 2e-16 ***
## is_luxury                   13.78499    0.19895  69.288  < 2e-16 ***
## surge_multiplier:isweekend   4.05489    1.56676   2.588  0.00966 **
## is_uber:is_luxury          -13.98273    0.31718 -44.084  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.07 on 11596 degrees of freedom
## Multiple R-squared:  0.4088, Adjusted R-squared:  0.4084
## F-statistic:  1145 on 7 and 11596 DF,  p-value: < 2.2e-16
```

```
model <- lm(price ~ distance +
              surge_multiplier +
              (surge_multiplier * isweekend) +
              is_uber +
```

```
              is_luxury +
              (is_uber * is_luxury),
            data = ridesharedata1)
summary(model)
```

```
##
## Call:
## lm(formula = price ~ distance + surge_multiplier + (surge_multiplier *
##     isweekend) + is_uber + is_luxury + (is_uber * is_luxury),
##     data = ridesharedata1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.911  -5.623  -1.184   3.623  40.078
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -14.90148    0.83424 -17.862  < 2e-16 ***
## distance                     2.73880    0.05841  46.890  < 2e-16 ***
## surge_multiplier            16.39264    0.79357  20.657  < 2e-16 ***
## isweekend                   -3.87284    1.59826  -2.423  0.01540 *
## is_uber                      8.27015    0.19137  43.215  < 2e-16 ***
## is_luxury                   13.78499    0.19895  69.288  < 2e-16 ***
## surge_multiplier:isweekend   4.05489    1.56676   2.588  0.00966 **
## is_uber:is_luxury          -13.98273    0.31718 -44.084  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.07 on 11596 degrees of freedom
## Multiple R-squared:  0.4088, Adjusted R-squared:  0.4084
## F-statistic:  1145 on 7 and 11596 DF,  p-value: < 2.2e-16
```

```
# Filter the dataset for Uber
ridesharedata_uber <- ridesharedata %>%
  filter(cab_type == "Uber")

# Filter the dataset for Lyft
ridesharedata_lyft <- ridesharedata %>%
  filter(cab_type == "Lyft")

#MODEL FOR UBER ONLY
model7 <- lm(price ~ distance + temperature +
              precipIntensity + (precipIntensity * precipIntensity) +
              isweekend +
               is_luxury + Morning + Afternoon + Evening + Night + Late_Night +
              (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Night) + (isweekend * Late
            data = ridesharedata_uber)

summary(model7)
```

```
##
## Call:
## lm(formula = price ~ distance + temperature + precipIntensity +
##     (precipIntensity * precipIntensity) + isweekend + is_luxury +
##     Morning + Afternoon + Evening + Night + Late_Night + (isweekend *
```

```
##     Afternoon) + (isweekend * Evening) + (isweekend * Night) +
##     (isweekend * Late_Night), data = ridesharedata_uber)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.411  -6.341  -2.345   4.512  40.623
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            9.34904    0.74457  12.556   <2e-16 ***
## distance               2.44197    0.08901  27.435   <2e-16 ***
## temperature            0.03466    0.01746   1.985   0.0471 *
## precipIntensity        5.92269    4.16685   1.421   0.1553
## isweekend              0.07855    0.41033   0.191   0.8482
## is_luxury             -0.21873    0.28043  -0.780   0.4354
## Morning               -0.23315    0.36806  -0.633   0.5265
## Afternoon              0.15721    0.41632   0.378   0.7057
## Evening               -0.82709    0.40316  -2.052   0.0403 *
## Night                 -0.32224    0.42039  -0.767   0.4434
## Late_Night            -0.59893    0.41623  -1.439   0.1502
## isweekend:Afternoon   -1.17870    0.71009  -1.660   0.0970 .
## isweekend:Evening     -0.22468    0.74914  -0.300   0.7643
## isweekend:Night       -0.83008    0.72664  -1.142   0.2534
## isweekend:Late_Night   1.63825    0.70520   2.323   0.0202 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.015 on 5882 degrees of freedom
## Multiple R-squared:  0.1185, Adjusted R-squared:  0.1164
## F-statistic: 56.47 on 14 and 5882 DF,  p-value: < 2.2e-16
```

```r
#MODEL FOR LYFT ONLY (PRICE)

model8 <- lm(price ~ distance + temperature + surge_multiplier +
               precipIntensity + (precipIntensity * precipIntensity) +
               isweekend +
                is_luxury + Morning + Afternoon + Evening + Night + Late_Night +
               (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Night) + (isweekend * Late
             data = ridesharedata_lyft)

#MODEL FOR LYFT ONLY (surge)
model9 <- lm(surge_multiplier ~ distance + temperature +
               precipIntensity + (precipIntensity * precipIntensity) +
               isweekend +
                is_luxury + Morning + Afternoon + Evening + Night + Late_Night +
               (isweekend * Afternoon) + (isweekend * Evening) + (isweekend * Night) + (isweekend * Late
             data = ridesharedata_lyft)
summary(model9)
```

```
##
## Call:
## lm(formula = surge_multiplier ~ distance + temperature + precipIntensity +
##     (precipIntensity * precipIntensity) + isweekend + is_luxury +
##     Morning + Afternoon + Evening + Night + Late_Night + (isweekend *
##     Afternoon) + (isweekend * Evening) + (isweekend * Night) +
```

```
##     (isweekend * Late_Night), data = ridesharedata_lyft)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05502 -0.03659 -0.03251 -0.01792  1.48666
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.018e+00  1.315e-02  77.406  < 2e-16 ***
## distance              7.644e-04  1.683e-03   0.454    0.650
## temperature          -4.633e-05  3.047e-04  -0.152    0.879
## precipIntensity      -6.738e-03  7.190e-02  -0.094    0.925
## isweekend             7.609e-03  7.091e-03   1.073    0.283
## is_luxury             1.860e-02  3.816e-03   4.874 1.12e-06 ***
## Morning               9.100e-03  6.281e-03   1.449    0.147
## Afternoon            -4.869e-03  7.086e-03  -0.687    0.492
## Evening               1.129e-05  7.040e-03   0.002    0.999
## Night                 4.860e-03  7.110e-03   0.683    0.494
## Late_Night           -2.410e-03  7.320e-03  -0.329    0.742
## isweekend:Afternoon  -9.390e-03  1.268e-02  -0.741    0.459
## isweekend:Evening    -1.390e-02  1.235e-02  -1.125    0.260
## isweekend:Night      -1.387e-02  1.229e-02  -1.128    0.259
## isweekend:Late_Night -6.316e-03  1.215e-02  -0.520    0.603
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1357 on 5692 degrees of freedom
## Multiple R-squared:  0.006043,   Adjusted R-squared:  0.003599
## F-statistic: 2.472 on 14 and 5692 DF,  p-value: 0.001719
```