Starbucks Sugar

AUTHOR
C. Fan, L. Adzobu, S.Pritchard

Starbucks

Overview

Starbucks is an international breakfast branch, serving high quality breakfast food and drink. The company was founded in 1971 in Seattle, Washington. In the past 53 years, Starbucks has expanded into 86 different countries and is valued at roughly **$110 Billion**. Starbucks centers on its core mission:

"To inspire and nurture the human spirit - one person, one cup, and one neighborhood at the time".

Starbucks has a wide range of menu items, however its backbone is the high quality coffee. For the first 10 years of operation, Starbucks operated as a coffee bean retailer. They have since expanded into a fast-casual model of business selling more than just coffee and making it in house. But coffee remains at the forefront of its identity.

As consultants our focus was to observe specific commodity inputs, being **coffee** and **sugar**. These commodities were given to us as they are a staple of the Starbucks business model. Sugar and coffee will always be a backbone of Starbuck’s purchases order. We forecast into the future to help Starbucks best prepare for the cost of coffee and sugar in the near future, and help make important buying decisions for their commodities.

Gathering, Cleaning, and Early Observations of the Data

We began by using data sets that reflect the price of sugar and coffee. We found this data through the *Federal Reserve Economic Data*. The following code reflects how we loaded in our csv, cleaned it, ran some simple descriptive statistics, and manipulated it into working data. We ran the mean and median for coffee and sugar, which encapsuled the time of 1990-today. We also fitered the data to reflect more accurate data for our decision making, which we deemed to be the last 10 years

```
library(tidyverse)
library(fpp3)
library(ggthemes)

# Load data:
## data: prices of coffee, prices of sugar
sugar<-read_csv("PSUGAUSAUSDQ.csv")
coffee <- read_csv("PCOFFOTMUSDM(1).csv")

mean(sugar$PSUGAUSAUSDQ)
```

```
[1] 25.17457
```

```
median(sugar$PSUGAUSAUSDQ)
```

```
[1] 22.5264
```

```
mean(coffee$PCOFFOTMUSDM)
```

```
[1] 140.2551
```

```
median(coffee$PCOFFOTMUSDM)
```

```
[1] 134.661
```

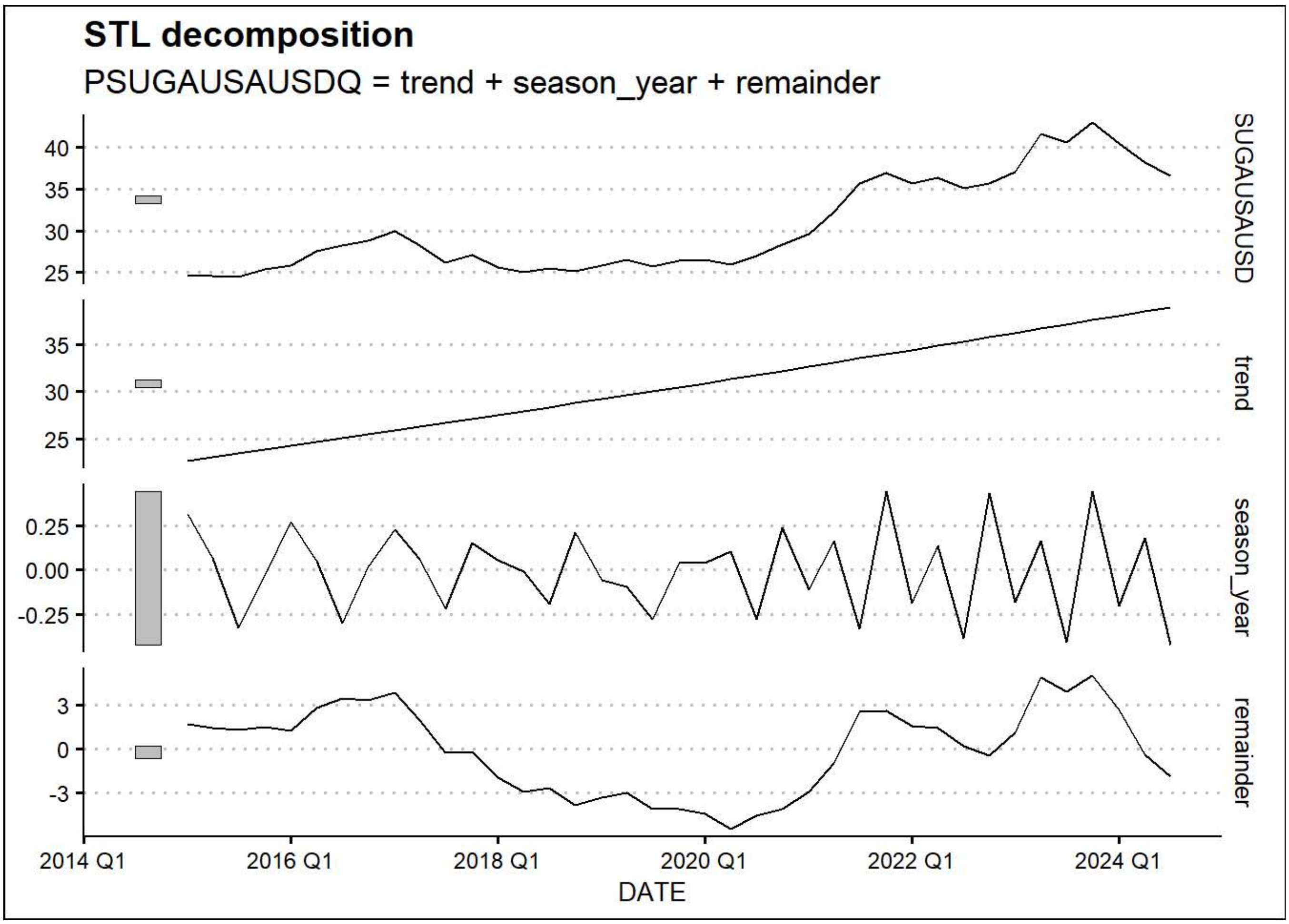
```
# Make Tsibble
sugar %>%
  mutate(DATE=yearquarter(DATE)) %>%
  as_tsibble(index=DATE) %>%
  mutate(Quarter = factor(quarter(DATE))) %>%
  filter(DATE > yearquarter("2014 Q4")) ->sugar_ts

coffee %>%
  mutate(DATE=yearquarter(DATE)) %>%
  as_tsibble(index=DATE) %>%
  mutate(Quarter = factor(quarter(DATE))) %>%
  filter(DATE > yearquarter("2014 Q4"))->coffee_ts
```

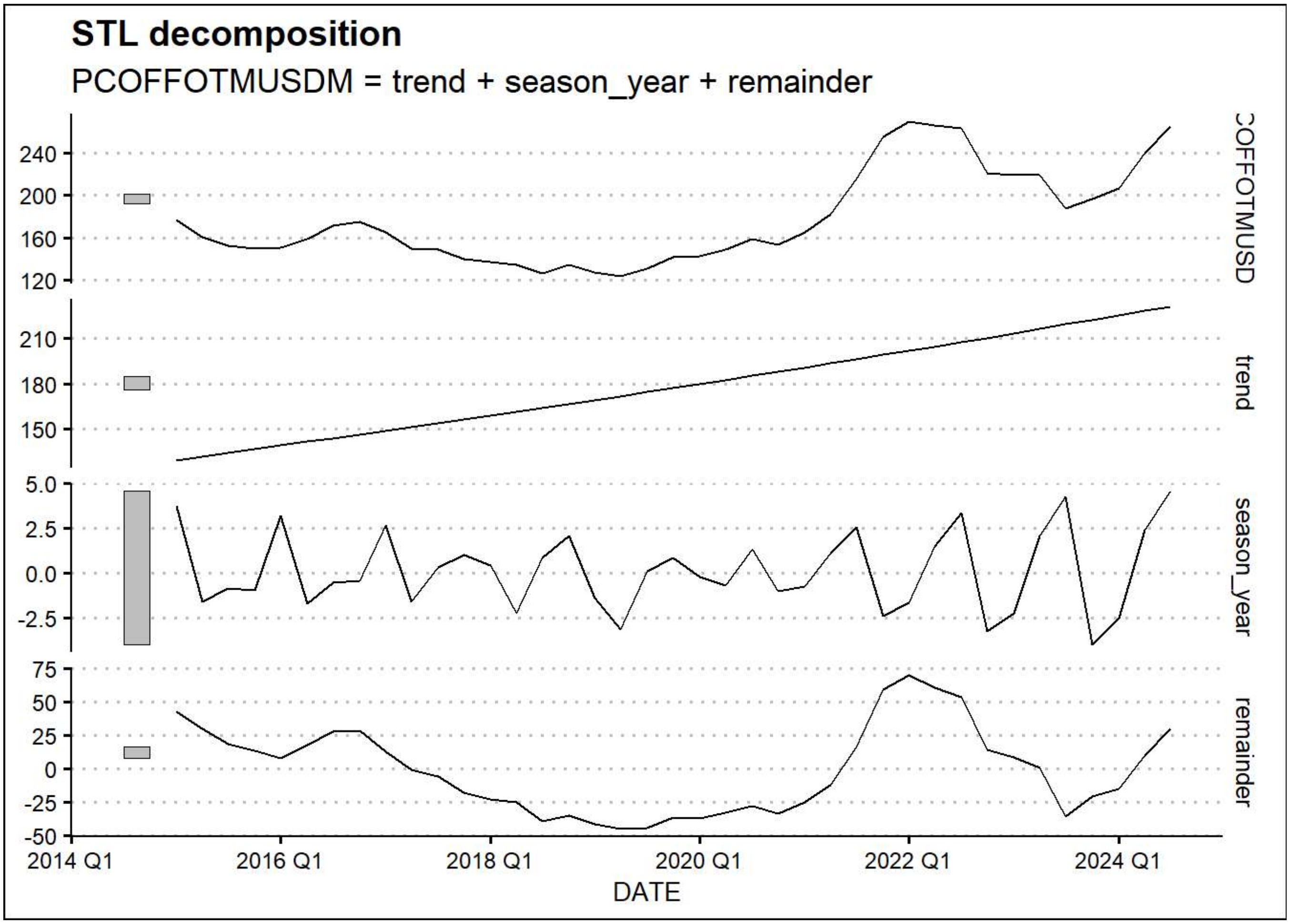
Decompositions

We ran decomposition through the STL model to find seasonality and potential trends. For both their is seasonality present. There seems to be more differences in seasonality in the Sugar data, yet there are peaks and troughs apparent through seasons in both sets.

```
# Decomposition
sugar_ts %>%
  model(STL(PSUGAUSAUSDQ~trend(100)+
              season())) %>%
  components() %>% autoplot() +
  theme_clean()
```




```
coffee_ts %>%
  model(STL(PCOFFOTMUSDM~trend(100)+
              season())) %>%
  components() %>% autoplot() +
  theme_clean()
```



Model Selection

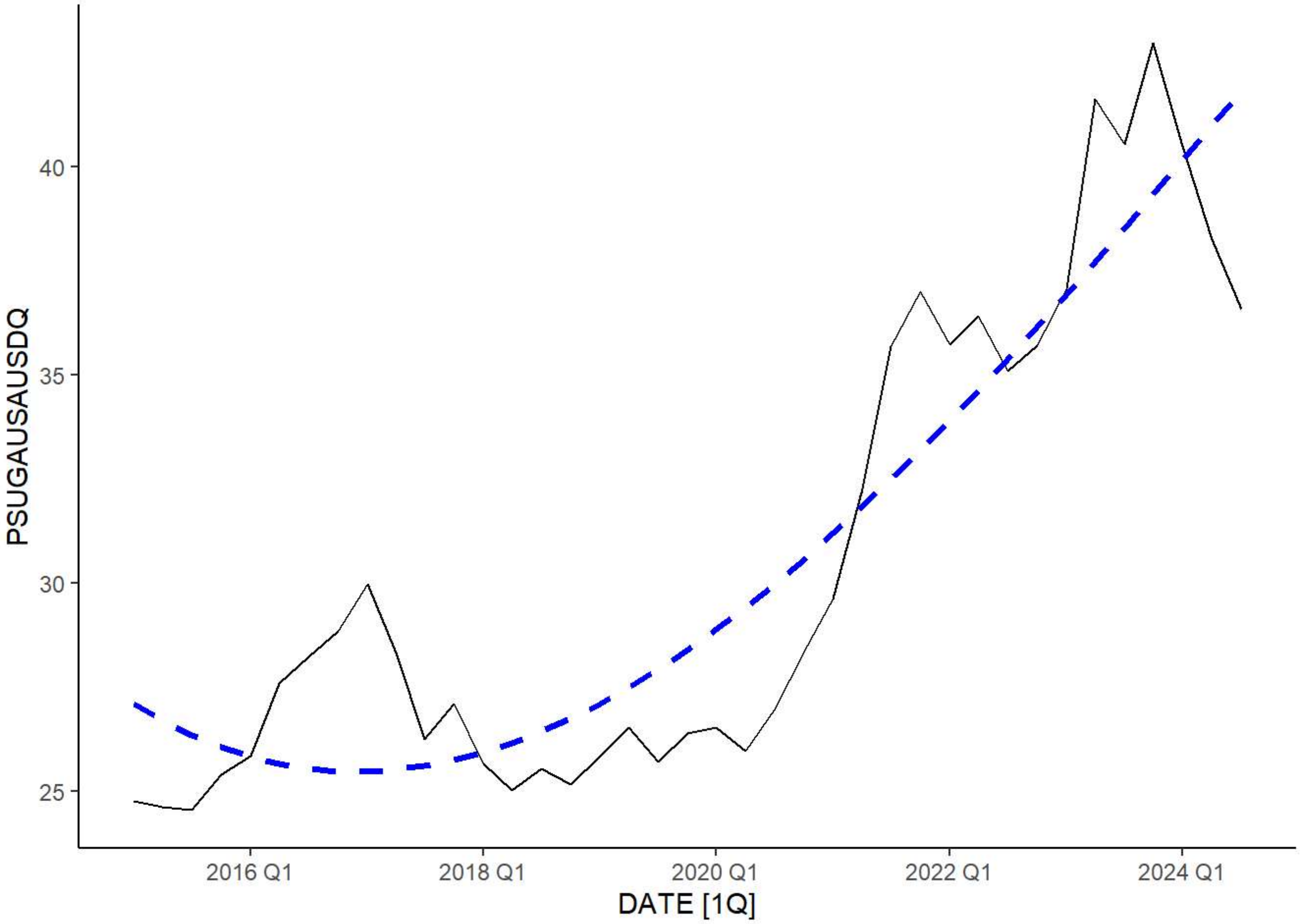
After bringing our data in and observing the decompositons, we looked at 3 different TSLM models. Under these conditions, for both Coffee and Sugar we found our **TRENDN** Model worked best. Understanding that TRENDN was our best model, we showed it as a plot.

```
sugar_ts %>%
  model(TREND=TSLM(PSUGAUSAUSDQ~trend()),
        TREND2=TSLM(PSUGAUSAUSDQ~trend()+I(trend()^2)),
        TRENDN=TSLM(PSUGAUSAUSDQ~I(poly(trend(), 3)))) -> fit2_sugar

fit2_sugar %>% accuracy() #best model, TRENDN
```

```
# A tibble: 3 × 10
  .model .type           ME  RMSE   MAE    MPE  MAPE  MASE RMSSE  ACF1
  <chr>  <chr>        <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 TREND  Training -8.20e-16  3.04  2.63 -0.889  8.95 0.863 0.795 0.864
2 TREND2 Training  2.73e-16  2.37  1.96 -0.579  6.39 0.642 0.620 0.701
3 TRENDN Training  2.73e-16  2.35  1.97 -0.569  6.49 0.647 0.615 0.711
```

```
sugar_ts %>% autoplot(.vars=PSUGAUSAUSDQ) +
  geom_line(data=fit2_sugar%>%augment()%>%filter(.model=="TRENDN"),
            aes(y=.fitted), col="blue", lty=2, lwd=1.1) +
  theme_classic()
```



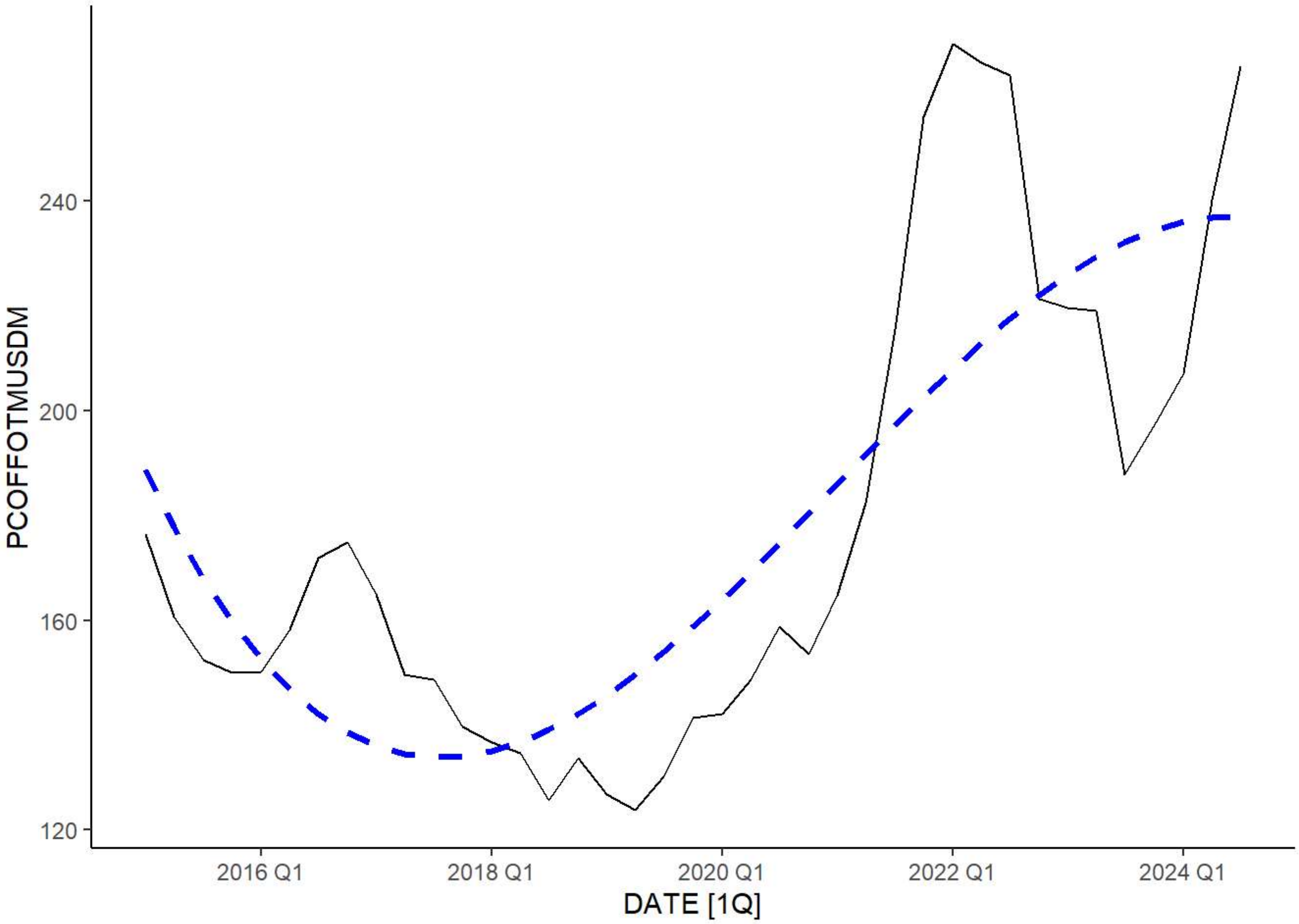

```
coffee_ts %>%
  model(TREND=TSLM(PCOFFOTMUSDM~trend()),
        TREND2=TSLM(PCOFFOTMUSDM~trend()+I(trend()^2)),
        TRENDN=TSLM(PCOFFOTMUSDM~I(poly(trend(), 3)))) -> fit2_coffee

fit2_coffee %>% accuracy() #best model, TRENDN
```

```
# A tibble: 3 × 10
  .model .type           ME  RMSE   MAE   MPE  MAPE  MASE RMSSE  ACF1
  <chr>  <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 TREND  Training -1.09e-15  33.0  28.6 -3.21  16.7 0.880 0.781 0.843
2 TREND2 Training -3.65e-16  28.3  20.7 -2.20  11.2 0.637 0.670 0.855
3 TRENDN Training  3.28e-15  25.9  21.0 -1.76  11.5 0.645 0.614 0.807
```

```
coffee_ts %>% autoplot(.vars=PCOFFOTMUSDM) +
  geom_line(data=fit2_coffee%>%augment()%>%filter(.model=="TRENDN"),
```

```
            aes(y=.fitted), col="blue", lty=2, lwd=1.1) +
    theme_classic()
```



After becoming familiar with the data, we returned to our main focus of forecasting the price of commodities into the future. First, we had to split the data into two sets, training and testing. This allowed us to run different models and understand which was best reflected in our data.

```
train_del_sugar<-filter_index(.data=sugar_ts,.~"2023 Q3")
test_del_sugar<-filter_index(.data=sugar_ts,"2023 Q4"~.)
train_del_coffee<-filter_index(.data=coffee_ts,.~"2023 Q3")
test_del_coffee<-filter_index(.data=coffee_ts,"2023 Q4"~.)
```

Once setting up our new intervals, we ran our training sets through all the same models as prior, TREND, TREND2, TRENDN. Under these circumstances, the **TREND** simulation worked best for both Coffee and Sugar.

```
train_del_sugar %>%
  model(TREND=TSLM(PSUGAUSAUSDQ~trend()),
        TREND2=TSLM(PSUGAUSAUSDQ~trend()+
                      I(trend()^2)),
        TRENDN=TSLM(PSUGAUSAUSDQ~I(poly(trend(), 3)))) -> fit3_sugar

fit3_sugar %>% forecast(new_data = test_del_sugar) %>% accuracy(sugar_ts)
```

```
# A tibble: 3 × 10
  .model .type    ME  RMSE   MAE    MPE  MAPE  MASE RMSSE   ACF1
  <chr>  <chr> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>  <dbl>
1 TREND  Test   2.48  3.77  3.03   5.86  7.35  1.06  1.02 0.254
2 TREND2 Test  -4.30  5.80  4.84 -11.5  12.8   1.69  1.58 0.254
3 TRENDN Test  10.1  18.6  18.2   23.5  45.6   6.36  5.05 0.0710
```

```
train_del_coffee %>%
  model(TREND=TSLM(PCOFFOTMUSDM~trend()),
        TREND2=TSLM(PCOFFOTMUSDM~trend()+
                      I(trend()^2)),
        TRENDN=TSLM(PCOFFOTMUSDM~I(poly(trend(), 3)))) -> fit3_coffee

# Passing deli_ts gets the accuracy on the test set: sugar

fit3_coffee %>% forecast(new_data = test_del_coffee) %>% accuracy(coffee_ts)
```

```
# A tibble: 3 × 10
  .model .type     ME  RMSE   MAE    MPE  MAPE  MASE RMSSE     ACF1
  <chr>  <chr>  <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl>    <dbl>
1 TREND  Test    6.66  25.3  22.9   1.70  9.77 0.707 0.600  0.286
2 TREND2 Test  -57.1   58.9  57.1 -26.2  26.2  1.76  1.40   0.275
3 TRENDN Test   55.6   95.2  84.8  27.6  38.6  2.62  2.26  -0.00956
```

However, after comparing the actual values to predicted, it was **TREND2** for Sugar that proved to be the better model. For Coffee, **TREND** remained the best model.

```
train_del_sugar %>%
  model(TREND=TSLM(PSUGAUSAUSDQ~trend()),
        TREND2=TSLM(PSUGAUSAUSDQ~trend()+
                      I(trend()^2)),
        TRENDN=TSLM(PSUGAUSAUSDQ~I(poly(trend(), 3)))) -> fit3_sugar

fit3_sugar %>% forecast(test_del_sugar) %>% filter(.model=="TREND") %>%
  pull(.mean) -> pred_sugar
mean(test_del_sugar$PSUGAUSAUSDQ[1:2]-pred_sugar[1:2])
```

```
[1] 5.028307
```

```
# Check accuracy on test set Trend 2: sugar: trend 2 is best
fit3_sugar %>% forecast(test_del_sugar) %>% filter(.model=="TREND2") %>%
  pull(.mean) -> pred_sugar_trend2
mean(test_del_sugar$PSUGAUSAUSDQ[1:2]-pred_sugar_trend2[1:2])
```

```
[1] -0.8082319
```

```
fit3_coffee %>% forecast(test_del_coffee) %>% filter(.model=="TREND") %>%
  pull(.mean) -> pred_coffee
mean(test_del_coffee$PCOFFOTMUSDM[1:2]-pred_coffee[1:2])
```

```
[1] -16.21816
```

```
# Check accuracy on test set Trend 2: coffee: trend 2 is best
fit3_coffee %>% forecast(test_del_coffee) %>% filter(.model=="TREND2") %>%
  pull(.mean) -> pred_coffee_trend2
mean(test_del_coffee$PCOFFOTMUSDM[1:2]-pred_coffee_trend2[1:2])
```

```
[1] -71.04644
```

We then added seasonality to our forecast. Here, our best model for Sugar and Coffee was **STREND**.

```
train_del_sugar %>%
  model(STREND=TSLM(PSUGAUSAUSDQ~trend()+Quarter),
        TREND2=TSLM(PSUGAUSAUSDQ~trend()+
                      I(trend()^2)),
        STREND2=TSLM(PSUGAUSAUSDQ~trend()+
                       I(trend()^2)+Quarter)) -> fit4_sugar

fit4_sugar %>% forecast(test_del_sugar) %>% accuracy(sugar_ts)
```

```
# A tibble: 3 × 10
  .model  .type    ME  RMSE   MAE    MPE  MAPE  MASE RMSSE  ACF1
  <chr>   <chr> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 STREND  Test   2.49  3.87  3.02   5.86  7.31  1.06  1.05 0.267
```

```
2 STREND2 Test  -4.34  5.79  4.77 -11.6  12.6   1.67  1.57 0.275
3 TREND2  Test  -4.30  5.80  4.84 -11.5  12.8   1.69  1.58 0.254
```

```
train_del_coffee %>%
  model(STREND=TSLM(PCOFFOTMUSDM~trend()+Quarter),
        TREND2=TSLM(PCOFFOTMUSDM~trend()+
                      I(trend()^2)),
        STREND2=TSLM(PCOFFOTMUSDM~trend()+
                      I(trend()^2)+Quarter)) -> fit4_coffee

fit4_coffee %>% forecast(test_del_coffee) %>% accuracy(coffee_ts)
```

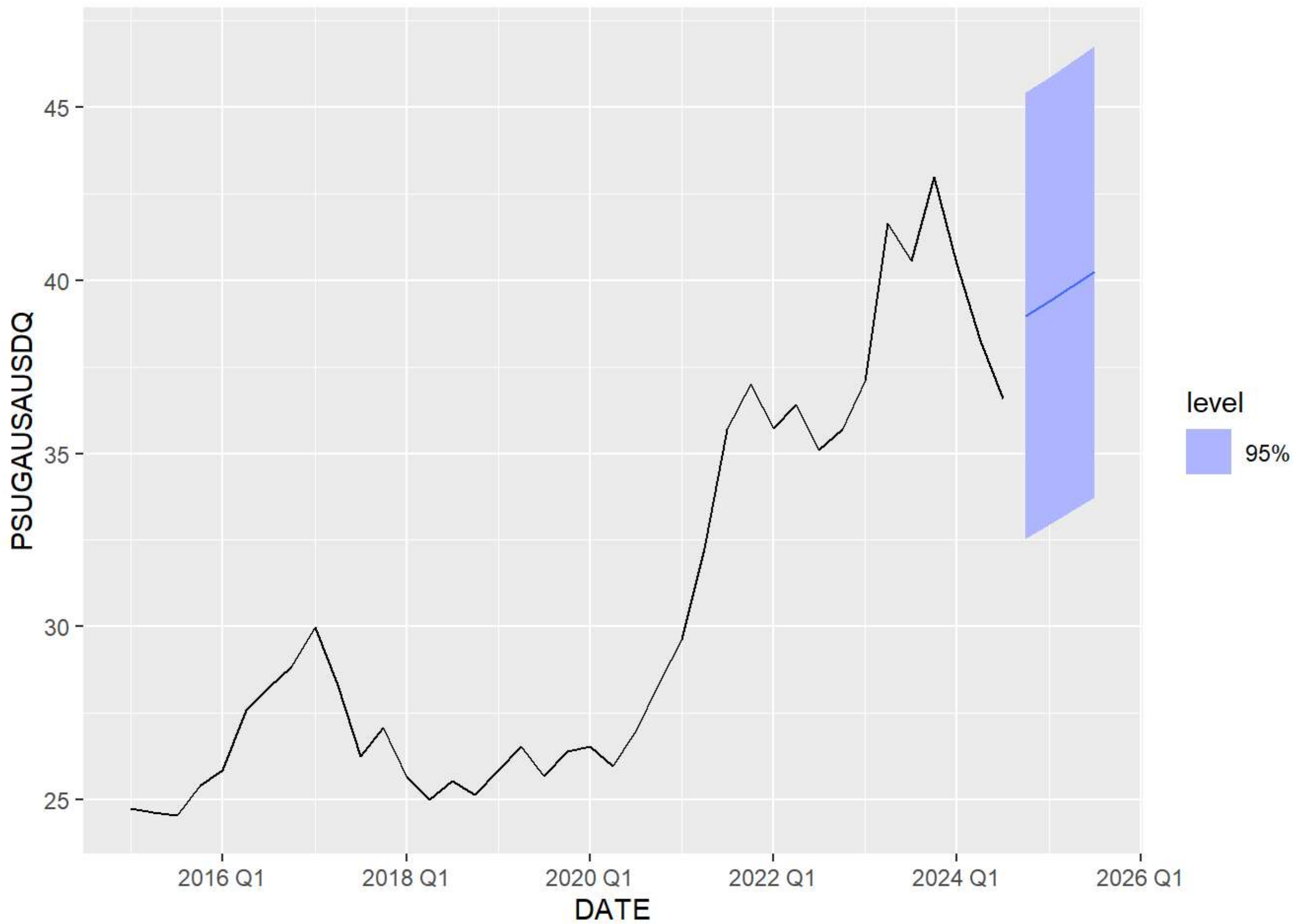
```
# A tibble: 3 × 10
  .model  .type     ME  RMSE   MAE    MPE  MAPE  MASE RMSSE  ACF1
  <chr>   <chr>  <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
1 STREND  Test    6.45  26.3  24.0   1.56  10.3 0.742 0.624 0.283
2 STREND2 Test  -58.0   60.5  58.0 -26.7   26.7 1.79  1.43  0.285
3 TREND2  Test  -57.1   58.9  57.1 -26.2   26.2 1.76  1.40  0.275
```

Results

With two separate models, we then had to choose which would work best for our prediction. Starting with Sugar, we ran our **TREND** model against the STREND model. Through an accuracy test, we concluded the **TREND** model was best fit for our analysis. From the forecast, we are 95% confident that by **Q4 of 2024**, the price of sugar will average around **38.9 USD** and reach about **40.2 USD** by **Q3 of 2025**.

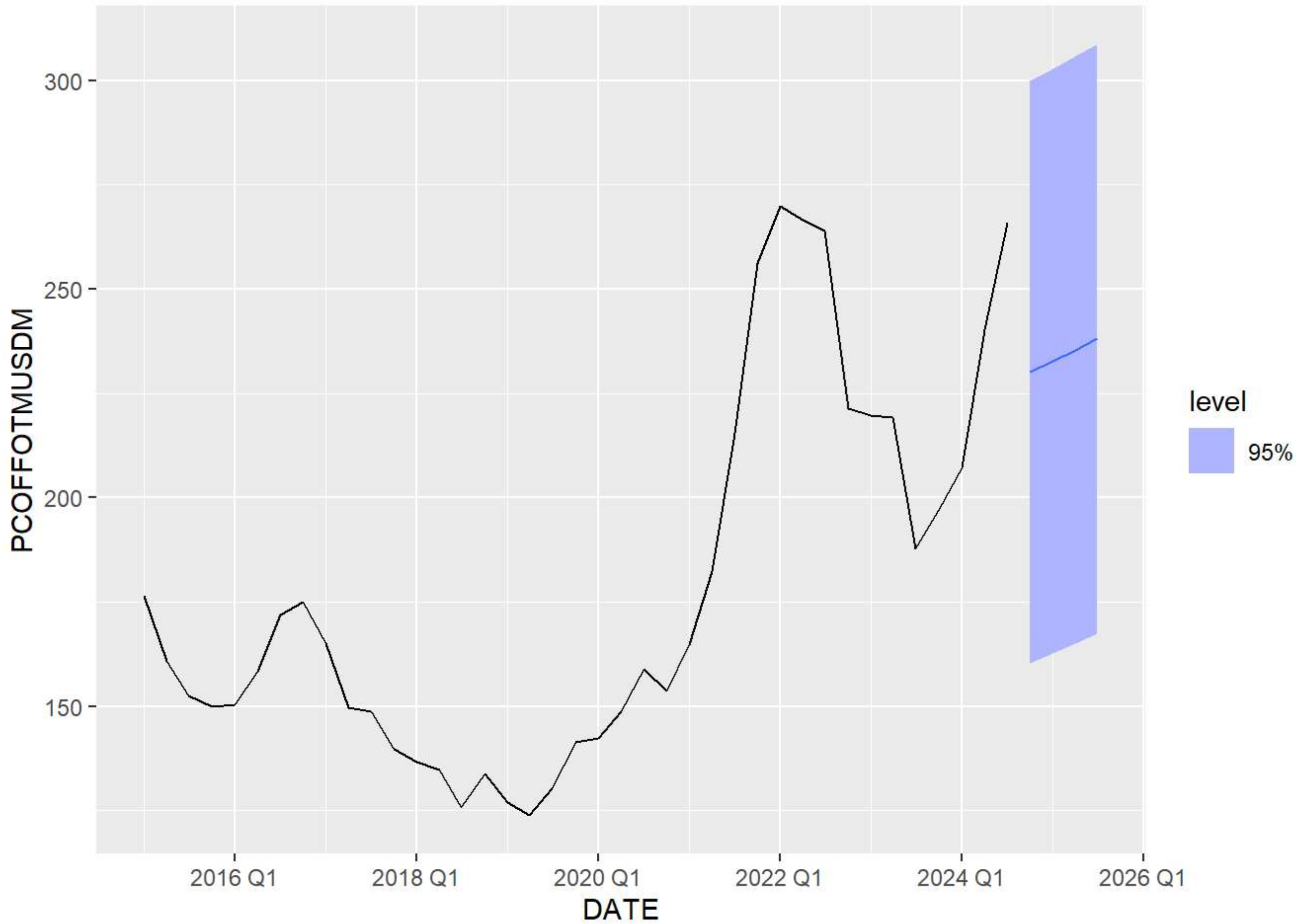
```
sugar_ts %>%
  model(TREND=TSLM(PSUGAUSAUSDQ~trend())) -> fit2_sugar
```

```
fit2_sugar %>% forecast(h = 4) %>% filter(.model=="TREND") %>%
  autoplot(level=95) +
  autolayer(sugar_ts, PSUGAUSAUSDQ)
```



For Coffee, we also chose our **TREND** model. Given our simulations, we predict with 95% confidence that by **Q4 2024** the price will reach at 95% **230.2 USD**. Looking further into the simulation, we predict that by **Q3 2025**, it will reach **238.07 USD**.

```
fit2_coffee %>% forecast(h = 4) %>% filter(.model=="TREND") %>%
  autoplot(level=95) +
  autolayer(coffee_ts, PCOFFOTMUSDM)
```



Recomendation

In this scenario, we recommend Starbucks looks for a future contract where the sugar can be purchased at today’s rate. For coffee, we recommend buying high quantity during Q4 of 2024, and looking for potential future contracts.