Assignment 5: Perscriptive Analytics

Edom Solomon

2025-11-10

Load data

```
load("~/Downloads/NYC RE DATA Fa25.RData")
```

```
#Typecasting data
#Add all joined
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
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v forcats   1.0.0     v readr     2.1.5
v ggplot2   3.5.2     v stringr   1.5.1
v lubridate 1.9.4     v tibble    3.3.0
v purrr     1.1.0     v tidyr     1.3.1
```

```
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
all_nyc <- NYC_TRANSACTION_DATA %>%
    left_join(BUILDING_CLASS, by = c("BUILDING_CLASS_FINAL_ROLL" = "BUILDING_CODE_ID")) %>%
    left_join(NEIGHBORHOOD, by = c("NEIGHBORHOOD_ID" = "NEIGHBORHOOD_ID"))
```

```
# filter by type and neighborhood
my_hood <- filter(all_nyc, NEIGHBORHOOD_NAME == "UPPER EAST SIDE (59-79)")

UES_residential <- filter(my_hood, TYPE == "RESIDENTIAL")
# install.packages("lubridate")
library(lubridate)
library(dbplyr)
```

```
Attaching package: 'dbplyr'

The following objects are masked from 'package:dplyr':

    ident, sql
```

```
# filter by date (day, month, year)
UES_residential$SALE_DATE <- as.Date(UES_residential$SALE_DATE, "%Y-%m-%d")
UES_residential <- mutate(UES_residential, SALE_YEAR = format(UES_residential$SALE_DATE, "%Y'

UES_residential <-UES_residential %>%
  mutate(SALE_DATE = ymd (SALE_DATE)) %>%
  mutate(SALE_YEAR = year (SALE_DATE)) %>%
  mutate(SALE_MONTH = month (SALE_DATE)) %>%
  mutate(SALE_DAY = day (SALE_DATE))
```

```
# Observation of data
library(utils)
head(UES_residential)
```

```
  SALE_ID NEIGHBORHOOD_ID BUILDING_CLASS_FINAL_ROLL                     ADDRESS
1 1013689             243                        C6    332 EAST 77TH STREET, 19
```

```
2 1013690           243                        D4      781 FIFTH AVENUE, 1104
3 1013691           243                        D4 333 EAST 66TH   STREET, 14J
4 1013692           243                        D4    315 EAST 68TH STREET, 5T
5 1013693           243                        R4             610 PARK AVENUE
6 1013694           243                        R4      401 EAST 60TH   STREET
  APARTMENT_NUMBER ZIP_CODE RESIDENTIAL_UNITS COMMERCIAL_UNITS
1               NA    10075                 0                0
2               NA    10022                 0                0
3               NA    10065                 0                0
4               NA    10065                 0                0
5               8B    10065                 1                0
6              18B    10022                 1                0
  GROSS_SQUARE_FEET YEAR_BUILT SALE_PRICE  SALE_DATE
1                 0       1910     240000 2013-05-14
2                 0       1928    8500000 2013-05-14
3                 0       1964     645000 2013-05-14
4                 0       1931     540000 2013-05-14
5                 0          0          0 2013-05-14
6                 0       1999          0 2013-05-14
                              DESCRIPTION        TYPE       NEIGHBORHOOD_NAME
1                        WALK-UP COOPERATIVE RESIDENTIAL UPPER EAST SIDE (59-
79)
2                       ELEVATOR COOPERATIVE RESIDENTIAL UPPER EAST SIDE (59-
79)
3                       ELEVATOR COOPERATIVE RESIDENTIAL UPPER EAST SIDE (59-
79)
4                       ELEVATOR COOPERATIVE RESIDENTIAL UPPER EAST SIDE (59-
79)
5 CONDO; RESIDENTIAL UNIT IN ELEVATOR BLDG. RESIDENTIAL UPPER EAST SIDE (59-
79)
6 CONDO; RESIDENTIAL UNIT IN ELEVATOR BLDG. RESIDENTIAL UPPER EAST SIDE (59-
79)
  BOROUGH_ID SALE_YEAR SALE_MONTH SALE_DAY
1          1      2013          5       14
2          1      2013          5       14
3          1      2013          5       14
4          1      2013          5       14
5          1      2013          5       14
6          1      2013          5       14
```

How can you calculate the total residential real estate sales in your neighborhood for each of the past 40 quarters? (2017-2018)

```r
# Time Series Forecasting
# install.packages("forecast")
library(forecast)
```

```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

```r
#Aggregate quarterly sales for past 40 quarters
library(lubridate)
# install.packages("dbplyr")
library(dbplyr)
# install.packages("magrittr")

UES_res_quart_sale <- UES_residential %>%
  mutate(GROSS_SQUARE_FEET = as.numeric(GROSS_SQUARE_FEET),
       RESIDENTIAL_UNITS = as.numeric(RESIDENTIAL_UNITS),
       SALE_PRICE = as.numeric(SALE_PRICE),
       SALE_YEAR = as.numeric(SALE_YEAR),
       SALE_MONTH = as.numeric(SALE_MONTH)
  )
```

```
Warning: There were 2 warnings in `mutate()`.
The first warning was:
i In argument: `GROSS_SQUARE_FEET = as.numeric(GROSS_SQUARE_FEET)`.
Caused by warning:
! NAs introduced by coercion
i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
```

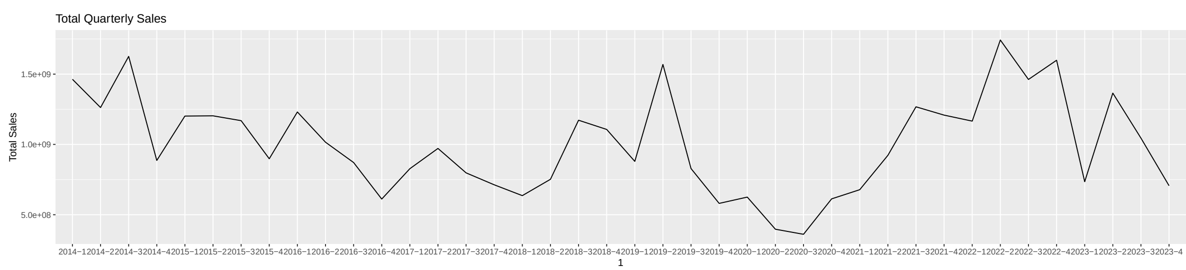
```r
UES_res_quart_sale <- UES_res_quart_sale %>%
  filter(SALE_YEAR >= 2014) %>%
  mutate(SALE_QUARTER = paste0(SALE_YEAR, "-", ceiling(as.numeric(SALE_MONTH)/3))) %>%
         group_by(SALE_QUARTER) %>%
         summarize(TOTAL_SALES = sum(SALE_PRICE, na.rm = TRUE))%>%
         ungroup()
```

When you create a line plot of quarterly sales, do you notice evidence of trend or seasonality? I moreso notice evidence of trend, there isn't a specific season where sales consistently peak.An example of trend is the drop in sales at the first and second quarter marks in 2020, which was the height of the COVD-19 pandemic, where there was a global shutdown; many people had to stay inside in order to protect themselves from potentially catching coronavirus and this impacted many economic sectors, including real estate. There is somewhat of a consistent peak when it comes to seasonality for the second quarter of each year on and after 2017 (sans 2020 due to the COVID-19 pandemic). this makes sense because the second quarter consists of May, June, July, and August. These are the warmest months of the year, which means it is more comfortable for both customers looking to purchase a property and real estate agents showing properties. Also, a lot of leases end between June and August, and a portion of those ending their leases during this time may be buying homes.

I utilized this resource to help utilize scale_x_discrete, labels, and group to label the quarters and classify SALE_QUARTER as since SALE_QUARTER is discrete due to it being a character.https://stackoverflow.com/questions/45342212/making-line-plot-with-discrete-x-axis-in-ggplot2 I also used this site to edit the width and height of the line plot in order to widen the spacing between labels in the x-axis: https://www.andrewheiss.com/blog/2022/06/23/long-labels-ggplot/

```
library(ggplot2)

ggplot(UES_res_quart_sale, aes(x=factor(SALE_QUARTER), y=TOTAL_SALES, group="Sale Quarter"))
  scale_x_discrete(labels("2014-1", "2014-2", "2014-3", "2014-4", "2015-1", "2015-2", "2015-
  geom_line() + labs(title= "Total Quarterly Sales",
  x="Sale Quarter",
  y="Total Sales")
```



Create a time series object

```
library(forecast)
# Create a time series object
UES_ts <- ts(UES_res_quart_sale$TOTAL_SALES, start = c(2014, 1), frequency = 4)



# view(UES_ts)
```

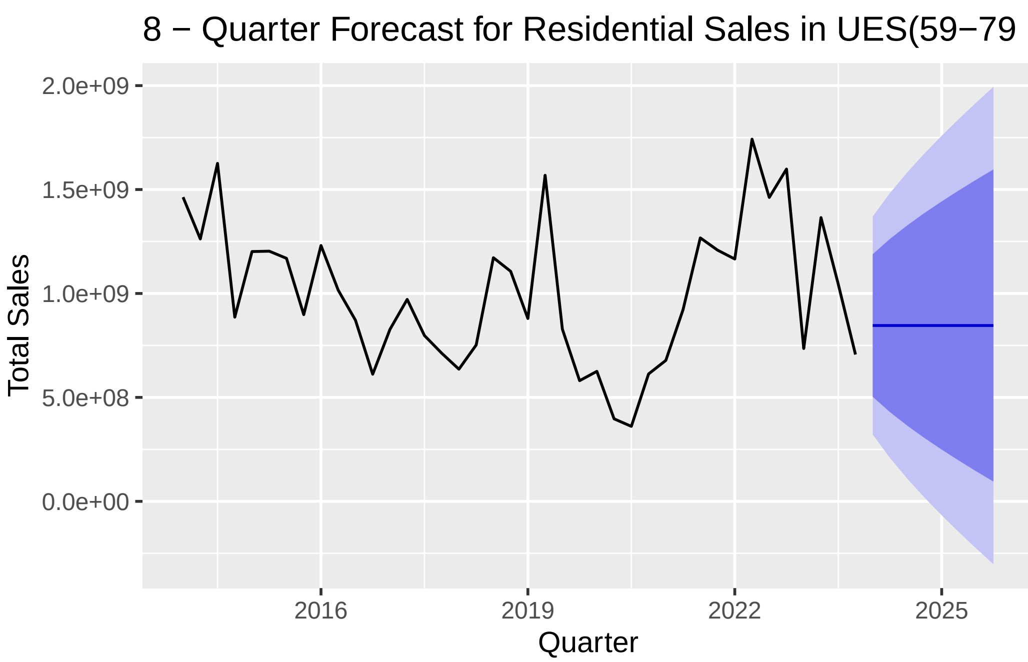
Using this time series, how would you create a forecast for the next 8 quarters? Which forecasting method do you choose, and why?

We can choose between several forecasting methods such as ARIMA, Explonential Smoothing (ETS), or Prophet. For this example, we will utilize ETS because it is effective for capturing trends and seasonality for time series data.

```
# Fit ETS model
ets_model <- (UES_ts)

# Forecast next 8 quarters
forecast_8_quarters <- forecast(ets_model, h = 8)

# Plot the forecast
autoplot(forecast_8_quarters) +
  ggtitle("8 - Quarter Forecast for Residential Sales in UES(59-79") +
  xlab("Quarter") +
  ylab("Total Sales")
```

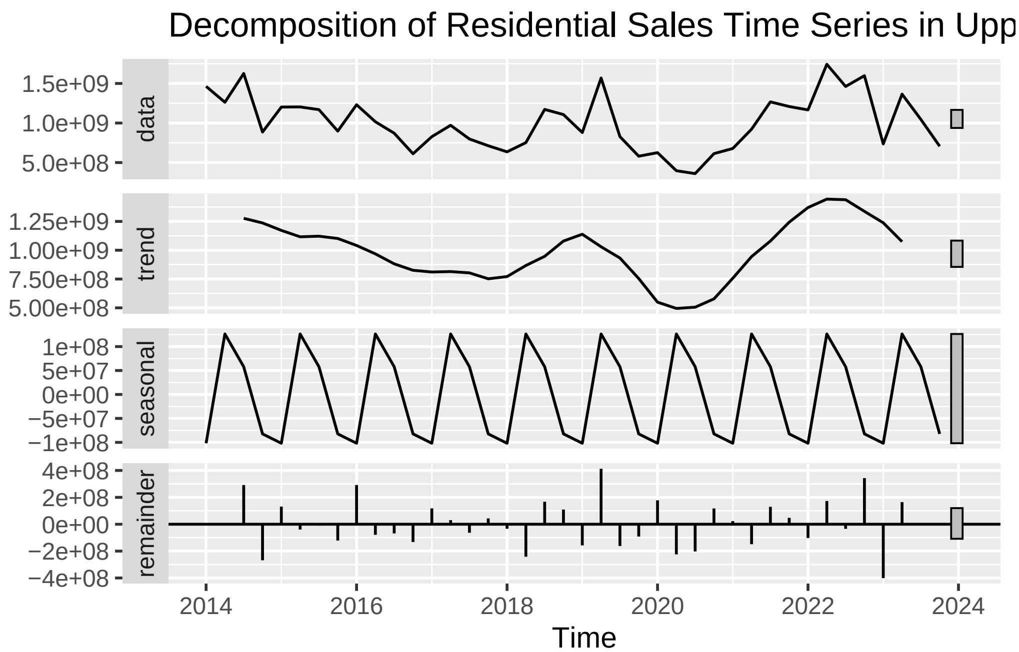


Decompose time series, what insights can you draw about seasonal patterns?

The insights that I can draw about the seasonal patterns are that there seems to be a continuous increase and decrease in total sales across different seasons throughout the years of 2017-2023 in throughout all 4 seasons, total sales peak in the second quarter (May, June, July, and August) consistently then fall in the third and fourth quarters. All months in the second quarter are the warmest months of the year, which means it is more comfortable for both customers looking to purchase a property and real estate agents showing properties. Also, a lot of leases end between June and August, and a portion of those ending their leases during this time may be buying homes.

```r
# Decompose time series
decomposed_ts <- decompose(UES_ts)

#Plot the decomposed componets
autoplot(decomposed_ts) +
  ggtitle("Decomposition of Residential Sales Time Series in Upper East Side (59-79)")
```



Based on your 8-quarter forecast, how would you calculate your company's quarterly profit under these assumptions? I would calculate my company's quarterly profit under these assumptions by multiplying forecasted values by market penetration and multiplying those with the result of commission rate minus variable cost rate. I would then subtract fixed costs from the results from the former (forecasted_values * market_penetration * (commission_rate - variable_cost_rate)). Commission = Variable Based on the neighborhood neighborhood calculate per quarter total sale value - subtract the 100k - subtract variable cost (1% of sales) - subtract commission (5% of sales)

Using these assumptions, what is the company's NPV over the next 8 quarters? The company's NPV over the next 8 quarters is $188,297.9.

```
forecasted_values <- as.numeric(forecast_8_quarters$mean)

# Calculate quarterly profit

commission_rate <- 0.05
variable_cost_rate <- 0.01
market_penetration <- 0.22
fixed_costs <- 100000
cost_of_funds <- 10

quarterly_profit <- (forecasted_values * market_penetration * (commission_rate - variable_cos

print(quarterly_profit)
```

```
[1] 7343619 7343619 7343619 7343619 7343619 7343619 7343619 7343619
```

```
# Calculate NPV
discount_rate <- 0.10 /4

# pv for each quarter
pv_each_quarter <- quarterly_profit / (1 / discount_rate)^(1:8)
pv_each_quarter
```

```
[1] 1.835905e+05 4.589762e+03 1.147441e+02 2.868601e+00 7.171503e-02
[6] 1.792876e-03 4.482190e-05 1.120547e-06
```

```
npv <- sum(pv_each_quarter)
npv
```

```
[1] 188297.9
```

```
# next step, make all of the result numbers positive by increasing your market penetration
```

Calculating Final Profit and Commission

```
# possibility to do these equations
commission <- quarterly_profit * commission_rate

commission
```

```
[1] 367181 367181 367181 367181 367181 367181 367181 367181
```

```r
# go back and reduce quarterly profit from commission for final profit
final_profit <- quarterly_profit - commission
final_profit
```

```
[1] 6976438 6976438 6976438 6976438 6976438 6976438 6976438 6976438
```

Optimization How can you model market penetration as a function of commission when the elasticity of demand is -1.2?

If the company is allowed to set a commission rate between 3% and 8%, which commission rate maximizes NPV? By how much does the optimized commission improve NPV compared to the baseline (5% commission)? The commission rate that maximizes NPV is 7%, as it cased a massive increase from $188,297.9 with the baseline commission at 5% to $1,589,013,415 (when optimizing the commission rate to 7%.

```r
elasticity <- -1.2
baseline_commission <- 0.05
baseline_penetration <- 0.03

#Function to calculate market penetration based on commission
market_penetration_function <- function(commission)
{
  baseline_penetration * (baseline_commission)^elasticity
  }



# Function to calculate NPV based on commission
npv_func <- function(commission){
  penetration <- market_penetration_function(commission)
  quarterly_profit <- (forecasted_values * penetration * (commission - variable_cost_rate)) 
  npv <- sum(quarterly_profit / (1+discount_rate)^(1:8))
  return(npv)
}
 print(npv)
```

```
[1] 188297.9
```

```r
# Optimize commission rate
  opt_result <- optimize(npv_func, interval = c(0.07, 0.25), maximum = TRUE)
opt_result
```

```
$maximum
[1] 0.2499591

$objective
[1] 1589013415
```

```r
print(opt_result)
```

```
$maximum
[1] 0.2499591

$objective
[1] 1589013415
```

Interpretation

What insights does your forecasting model provide about neighborhood real estate sales trends? The insight that my forecasting model provides about neighborhood real estate trends, specifically for my Upper East Side (59-79) neighborhood is that profits continuously increase and decrease depending on seasons. The time series forecasting model displays that there have been continuous increases and decreases. It also shows that there are outside factors that can impact real estate sales trends too. As we see in the year 2020 in the time series forecasting model, there is a huge decrease in sales, from around $1 billion dollars in sale in 2019 to around $500 million in 2020, a $500 million dollar decrease. This can seemingly be correlated to the COVID-19 pandemic, which caused a massive societal shutdown, impacting many economic sectors including real estate.

What managerial implications can be drawn from the optimized commission strategy? The managerial implications that can be drawn from the optimized commission strategy is that it's important to provide monetary support to your agents, these will all play a hand in increasing profits. When optimizing the commission rate from 5% to 7%, between 7 and 25 percent, we saw an extreme rise in NPV (with the former being $188,297.9 and the latter (with the increase in commission) at $1,589,013,415) as a result This shows care for the real estate agents and care for our customers, which boosts morale for those in our company and will have customers wanting to purchase real estate from us, which will sustain and increase our market penetration.

How do changes in commission policy affect both market penetration and profitability in this scenario? Changes in commission policy affect both market penetration and profitability in this scenario because with an increase in commission, it boosts the morale and productivity of our real estate agents, which will then either sustain or increase the NPV. We see that the

NPV has increased dramatically after increasing commission, with the former being \$188,297.9 and the latter (with the increase in commission) at \$1,589,013,415.