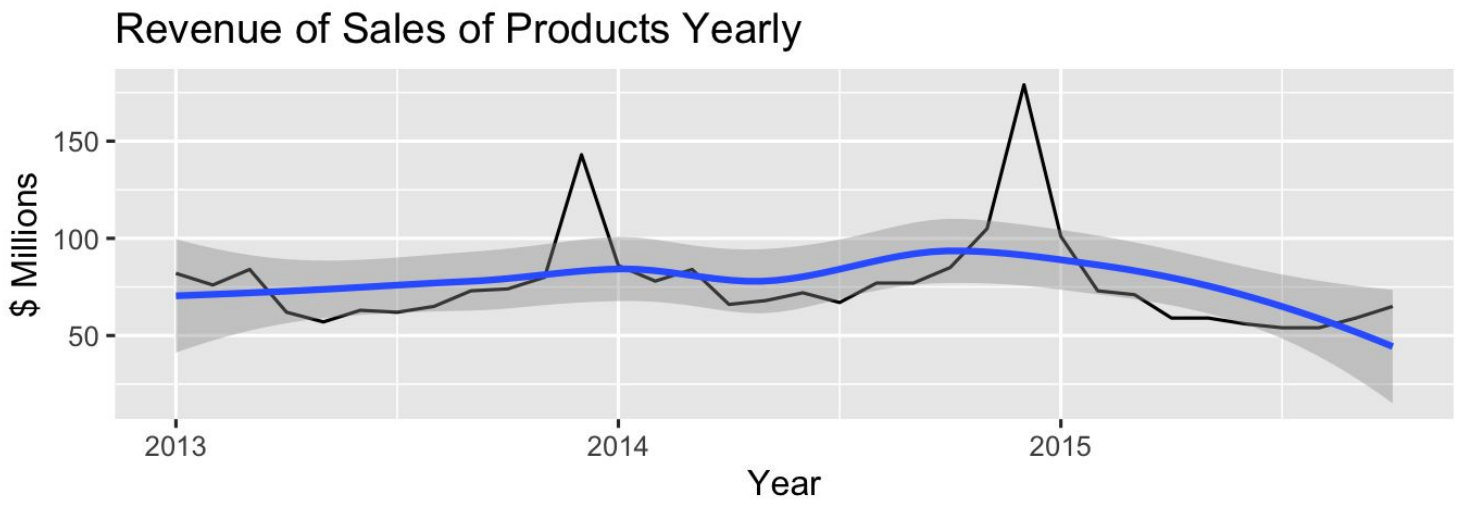
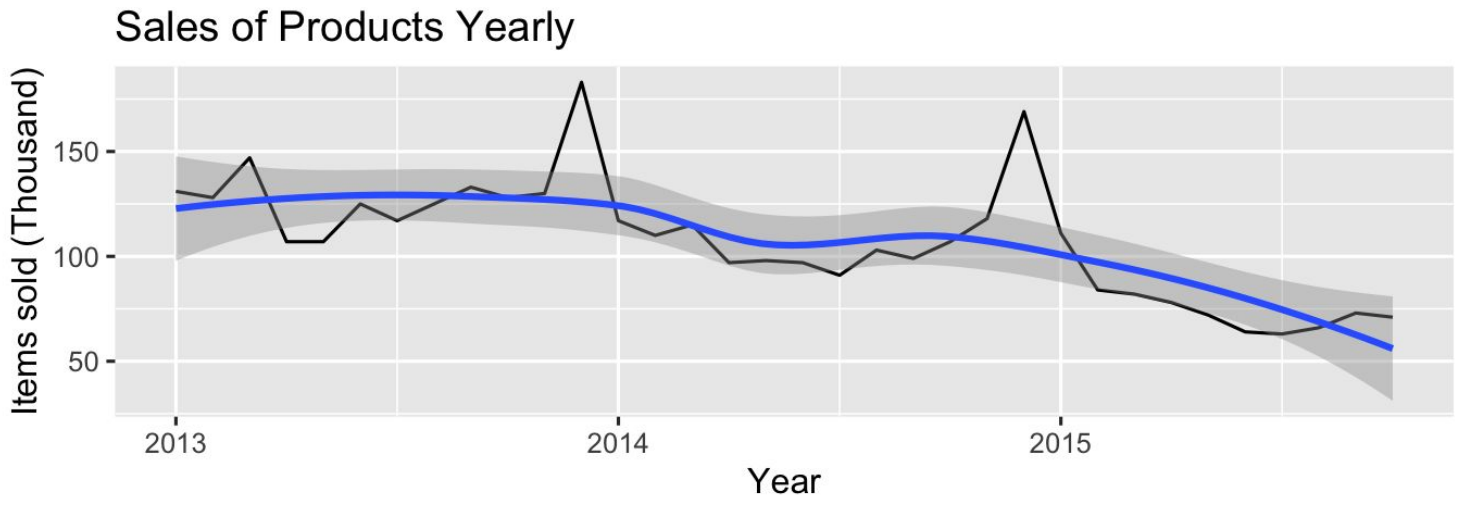
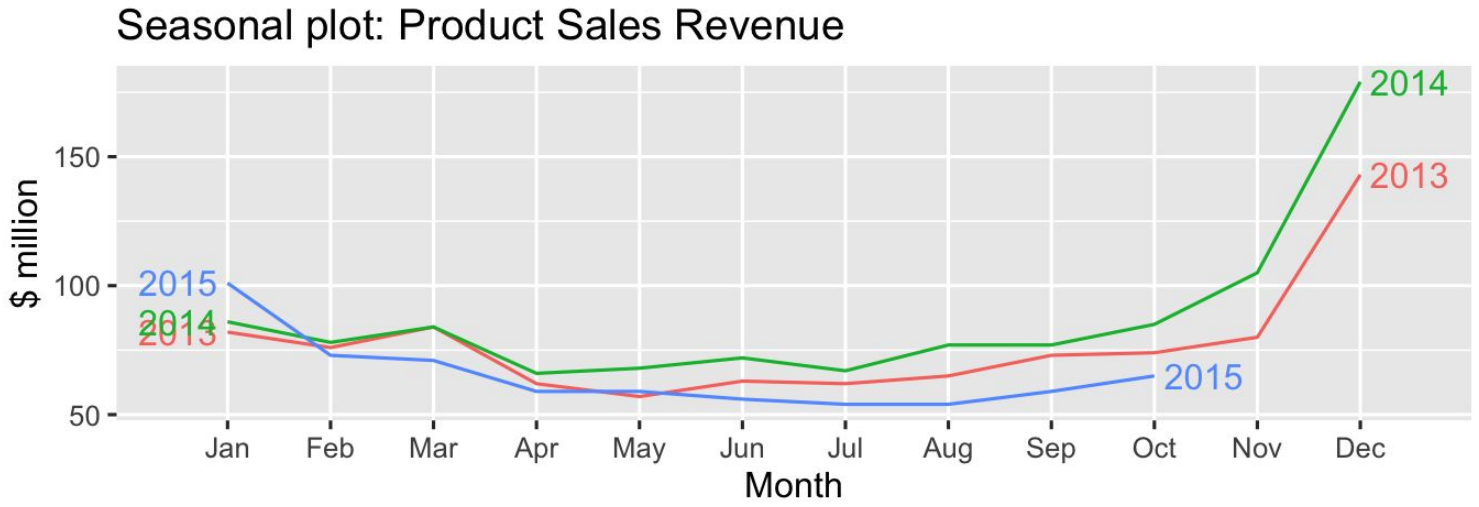
Sales Prediction Project

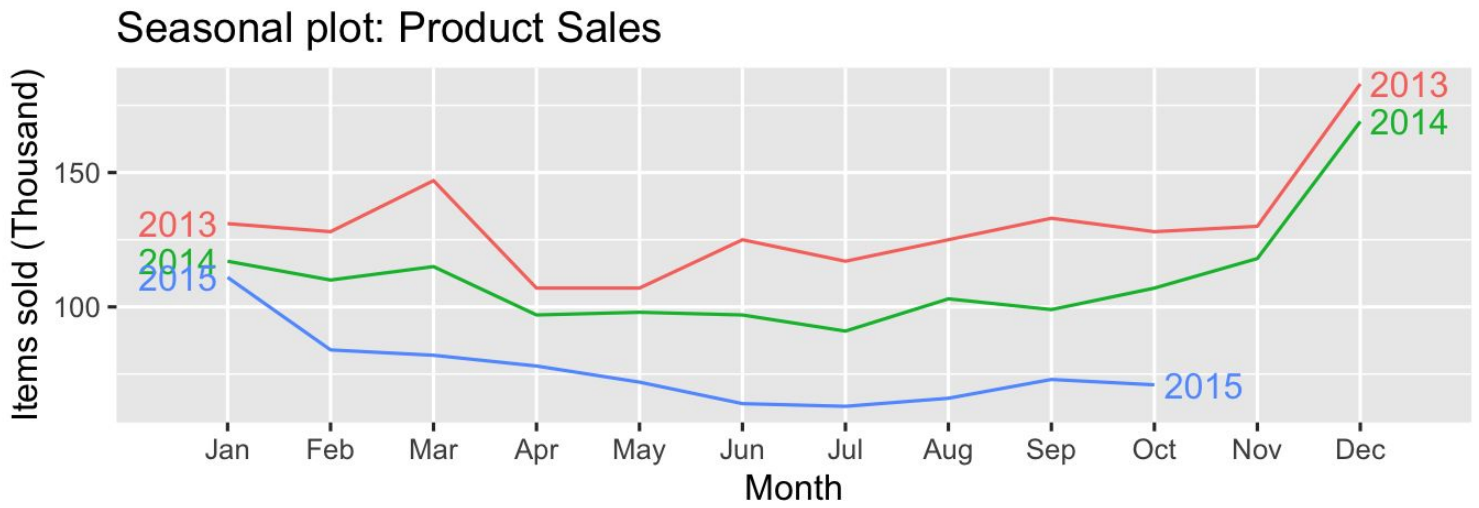
DS636
Pablo Salar Carrera

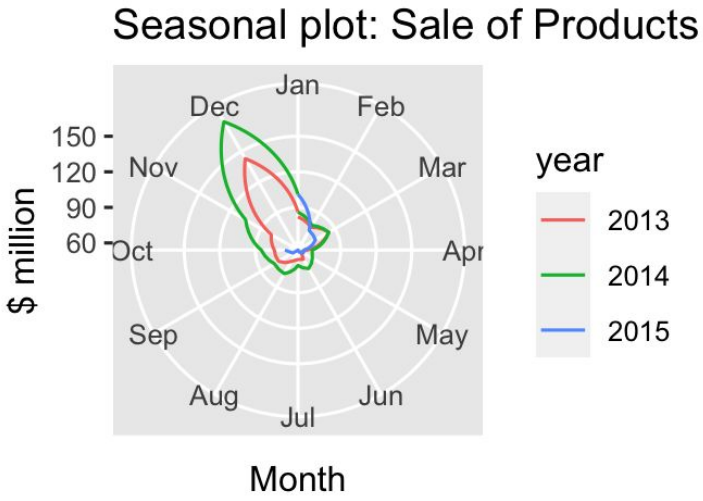
```
> glimpse(sales)
Rows: 2,935,828
Columns: 11
$ date              <date> 2013-01-02, 2013-01-03, 2013-01-05, 2013-01-…
$ date_block_num    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
$ shop_id           <fct> 59, 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 2…
$ item_id           <fct> 22154, 2552, 2552, 2554, 2555, 2564, 2565, 25…
$ item_price        <dbl> 999.00, 899.00, 899.00, 1709.05, 1099.00, 349…
$ item_cnt_day      <dbl> 1, 1, -1, 1, 1, 1, 1, 1, 1, 3, 2, 1, 1, 2, 1,…
$ item_category_id  <fct> 37, 58, 58, 58, 56, 59, 56, 55, 55, 55, 55, 5…
$ day               <int> 2, 3, 5, 6, 15, 10, 2, 4, 11, 3, 3, 5, 7, 8, …
$ month             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
$ weekday           <dbl> 4, 5, 7, 1, 3, 5, 4, 6, 6, 5, 5, 7, 2, 3, 5, …
$ year              <dbl> 2013, 2013, 2013, 2013, 2013, 2013, 2013, 201…
```

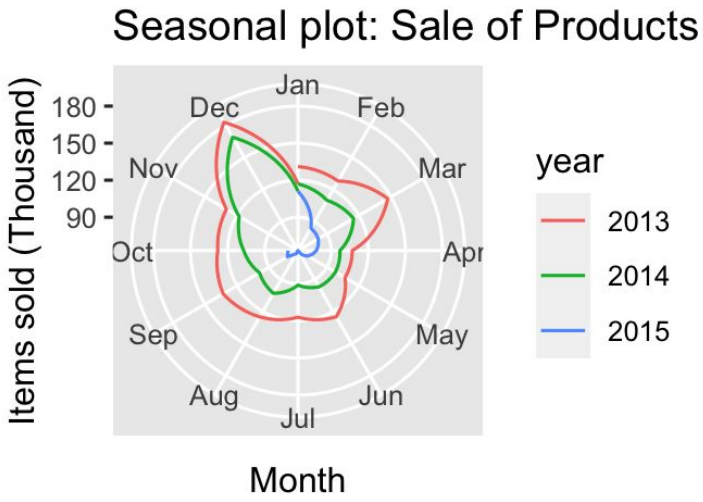

Revenue of Sales of Products Yearly
$ Millions
150
100
50
2013
2014
2015
Year

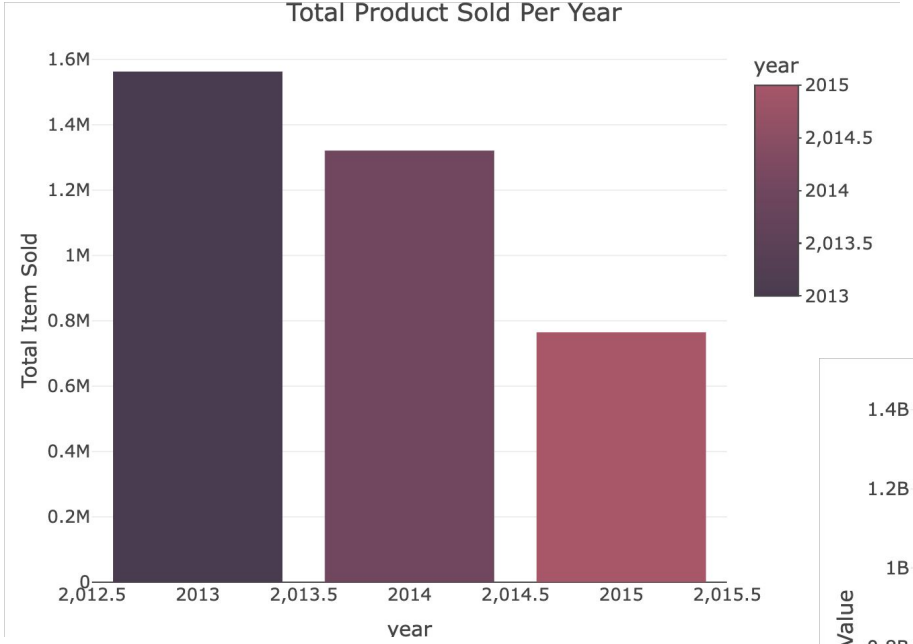

Sales of Products Yearly
Items sold (Thousand)
150
100
50
2013
2014
2015
Year

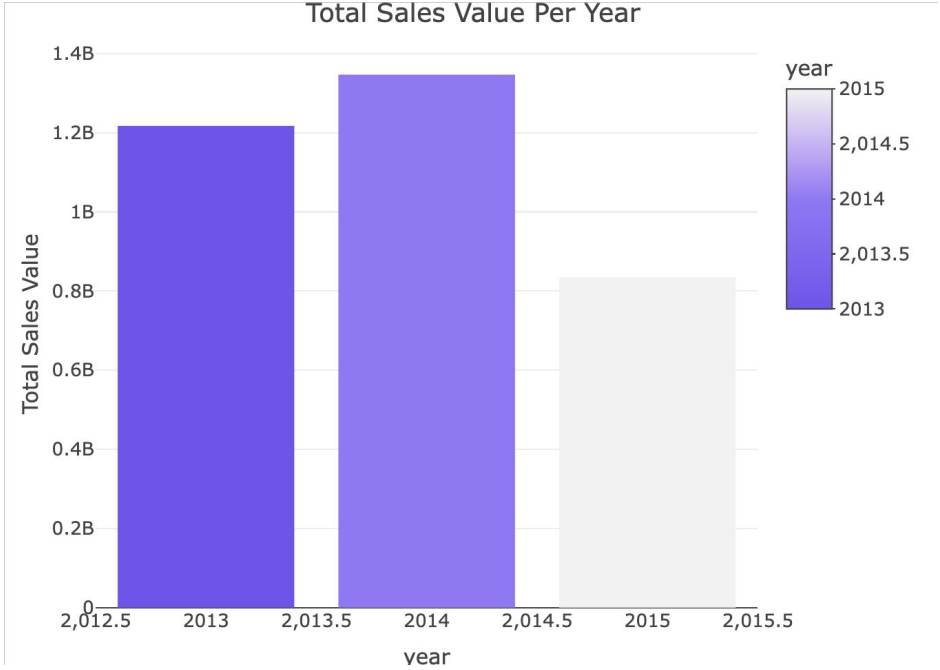

Seasonal plot: Product Sales Revenue
$ million
150
100
50
2015
2014
2013
2014
2013
2015
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

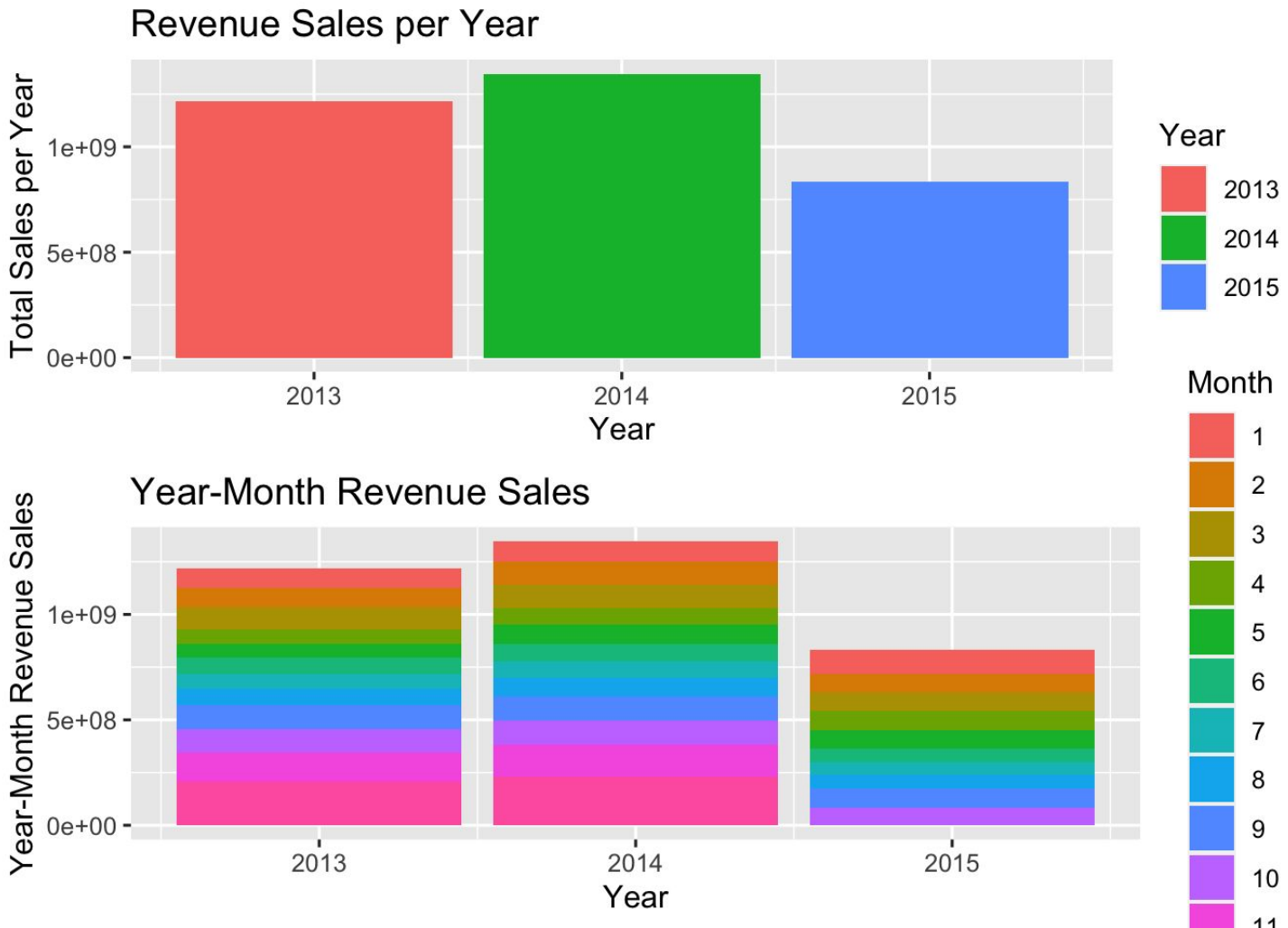

Seasonal plot: Product Sales
Items sold (Thousand)
150
100
2013
2014
2015
2013
2014
2015
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

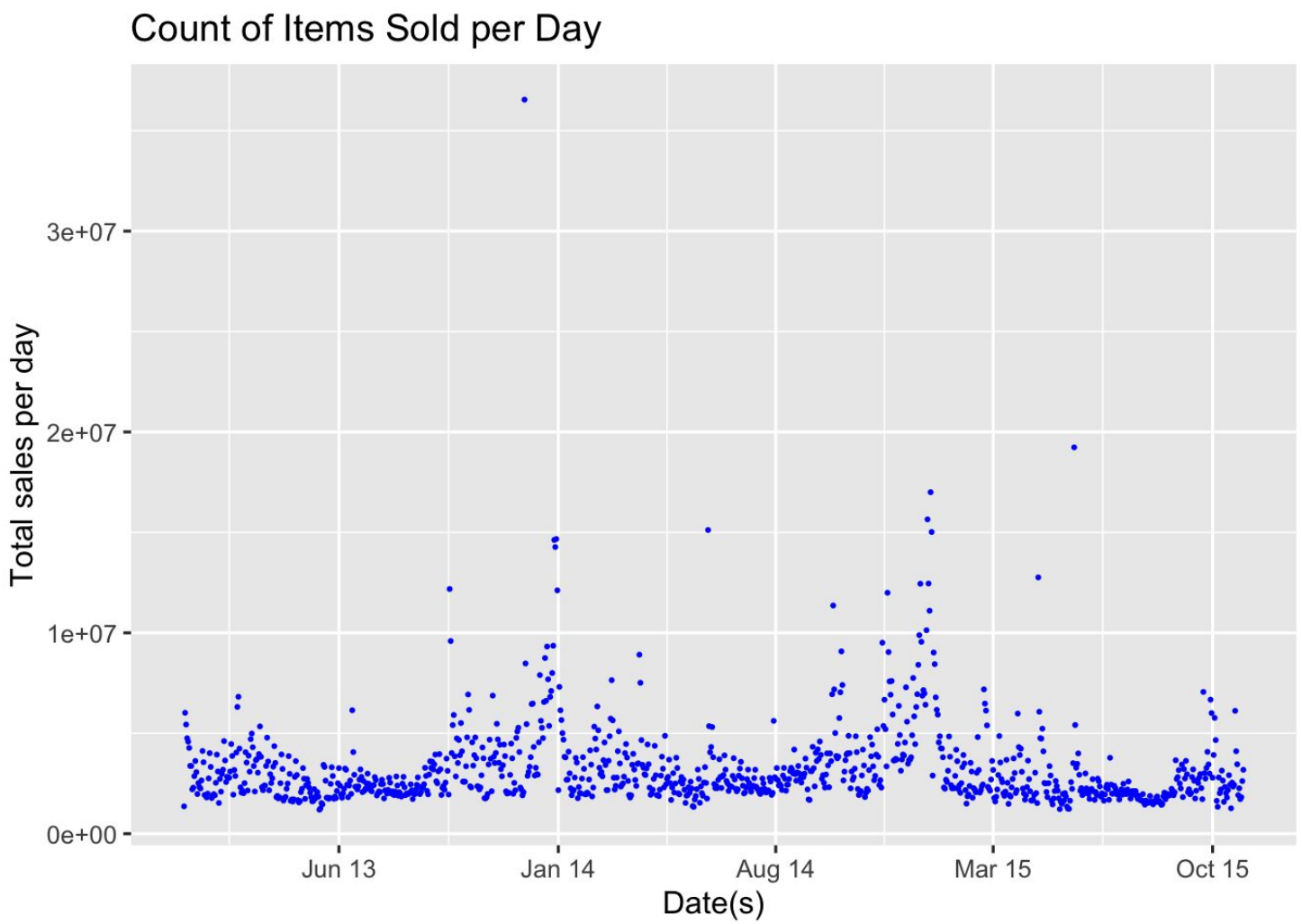

Seasonal plot: Sale of Products Revenue
$ million
150
120
90
60
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
year
2013
2014
2015


Seasonal plot: Sale of Products
Items sold (Thousand)
180
150
120
90
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
year
2013
2014
2015

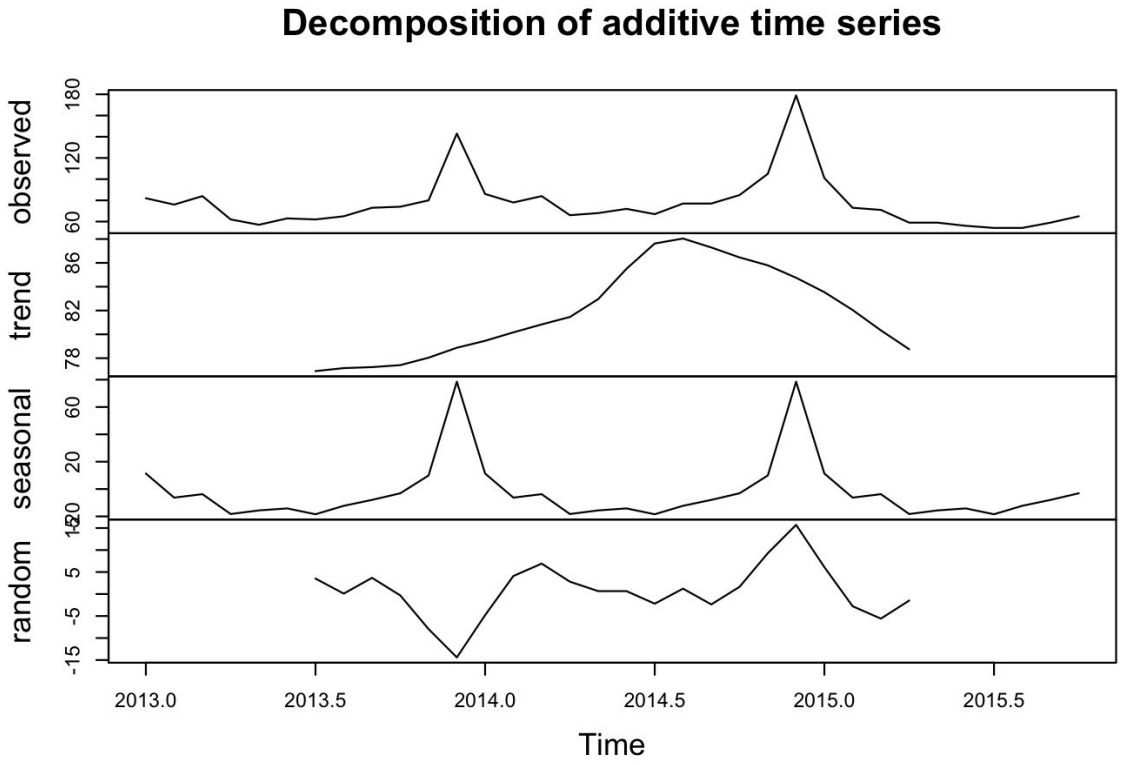

Total Product Sold Per Year
1.6M
1.4M
1.2M
1M
0.8M
0.6M
0.4M
0.2M
0
Total Item Sold
2,012.5
2013
2,013.5
2014
2,014.5
2015
2,015.5
year
year
2015
2,014.5
2014
2,013.5
2013


Total Sales Value Per Year
1.4B
1.2B
1B
0.8B
0.6B
0.4B
0.2B
0
Total Sales Value
2,012.5
2013
2,013.5
2014
2,014.5
2015
2,015.5
year
year
2015
2,014.5
2014
2,013.5
2013

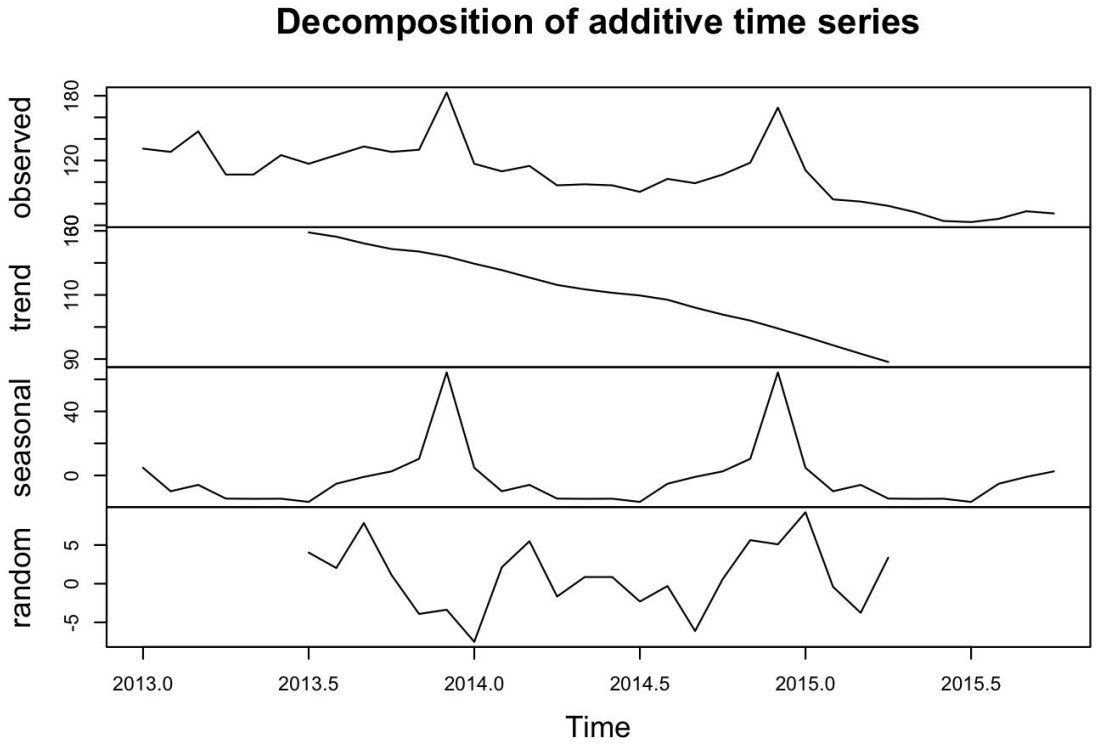

Revenue Sales per Year
Total Sales per Year
1e+09
5e+08
0e+00
2013
2014
2015
Year
Year
2013
2014
2015
Year-Month Revenue Sales
Year-Month Revenue Sales
1e+09
5e+08
0e+00
2013
2014
2015
Year
Month
1
2
3
4
5
6
7
8
9
10
11


Count of Items Sold per Day
Total sales per day
3e+07
2e+07
1e+07
0e+00
Jun 13
Jan 14
Aug 14
Mar 15
Oct 15
Date(s)

Revenue of Sales

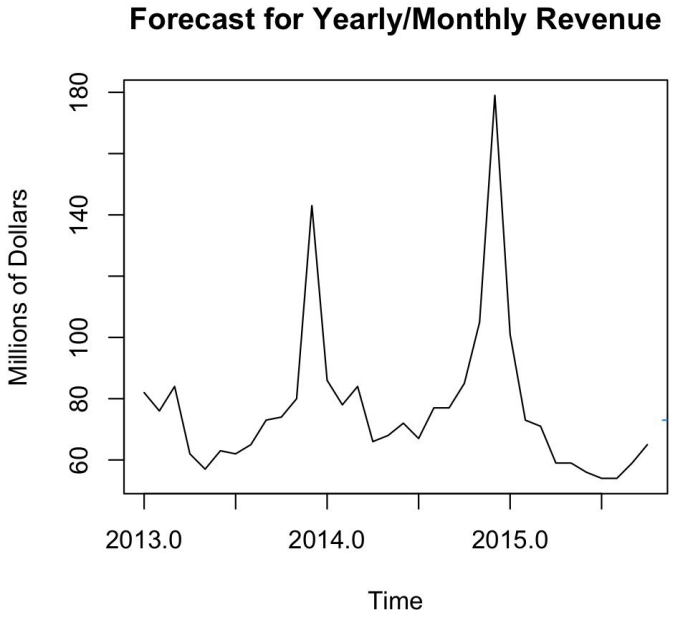


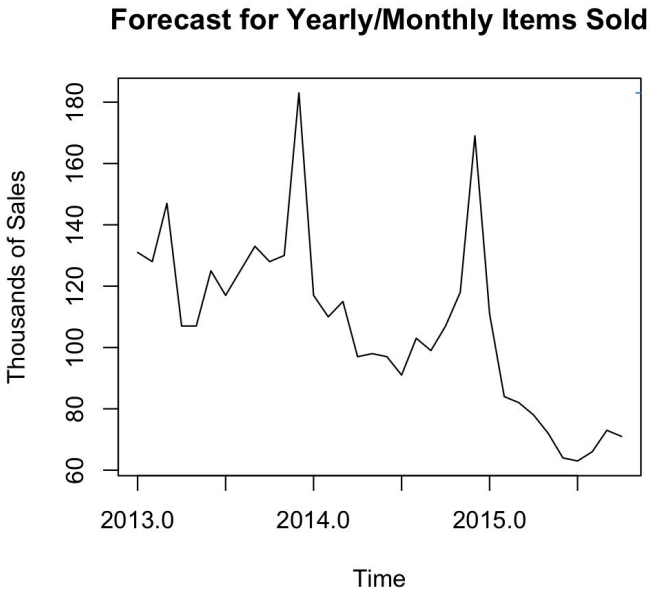
Number of Items Sold

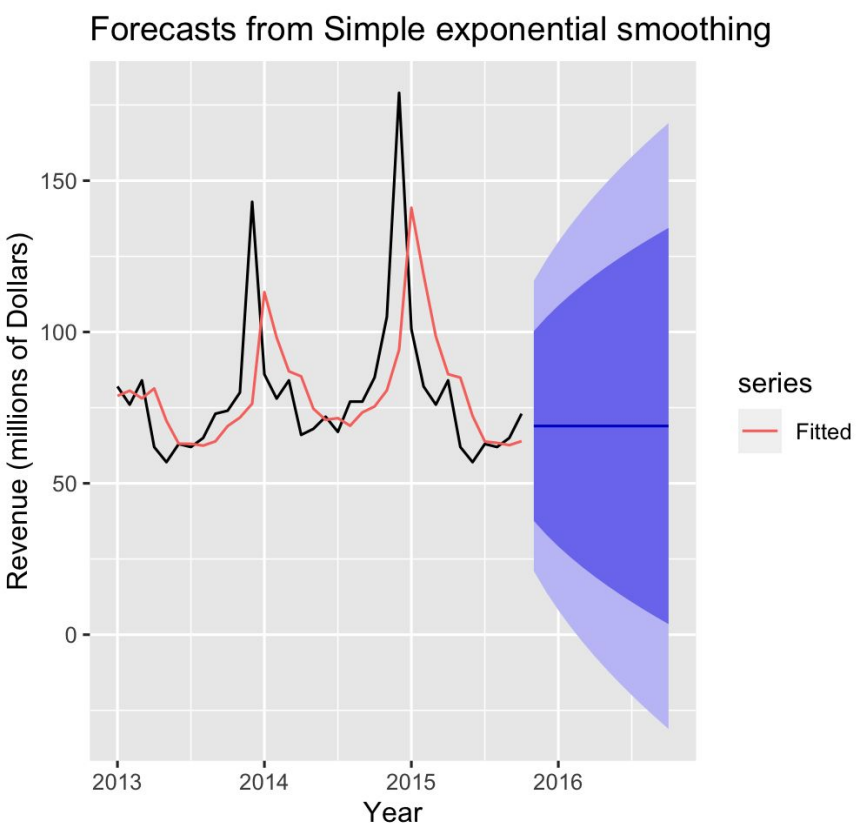


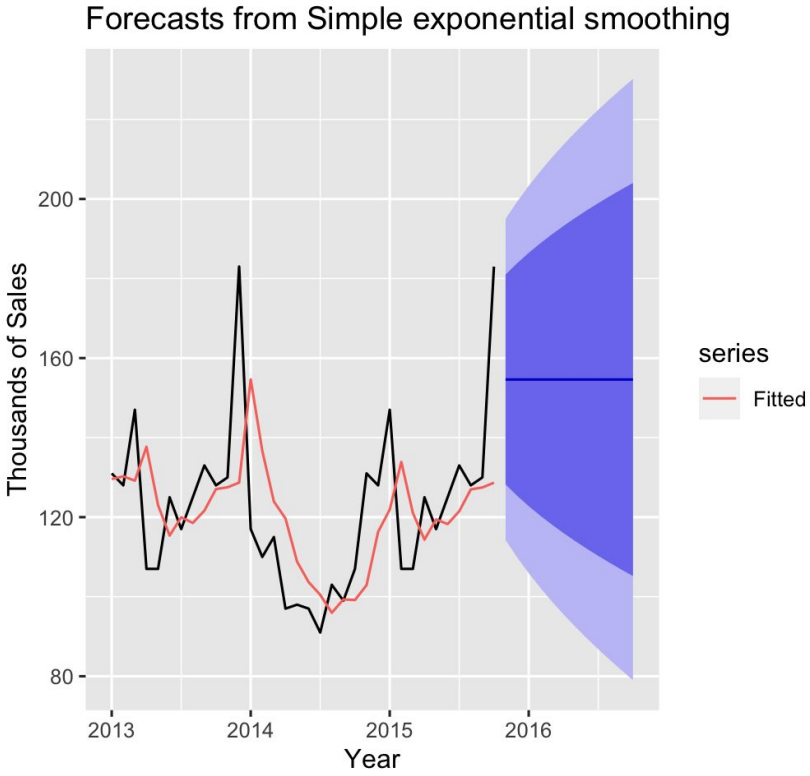
```
> adf.test(df1)

        Augmented Dickey-Fuller Test

data:  df1
Dickey-Fuller = -2.5878, Lag order = 3, p-value = 0.3452
alternative hypothesis: stationary

> adf.test(df2)

        Augmented Dickey-Fuller Test

data:  df2
Dickey-Fuller = -2.7272, Lag order = 3, p-value = 0.2912
alternative hypothesis: stationary
```

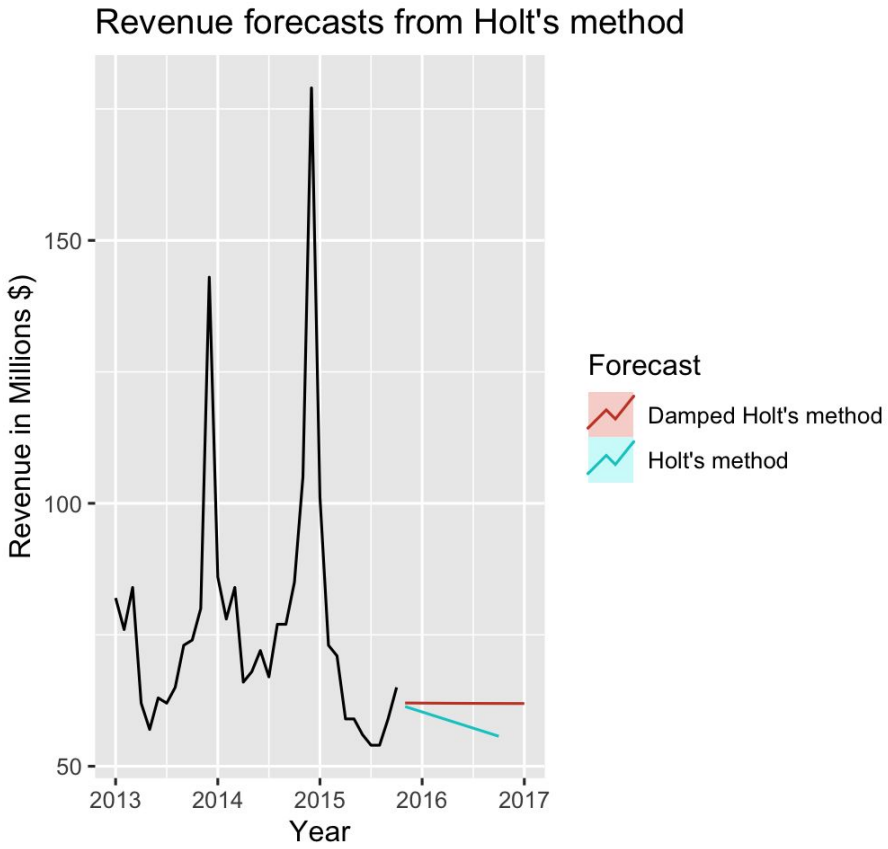
Time Series Prediction

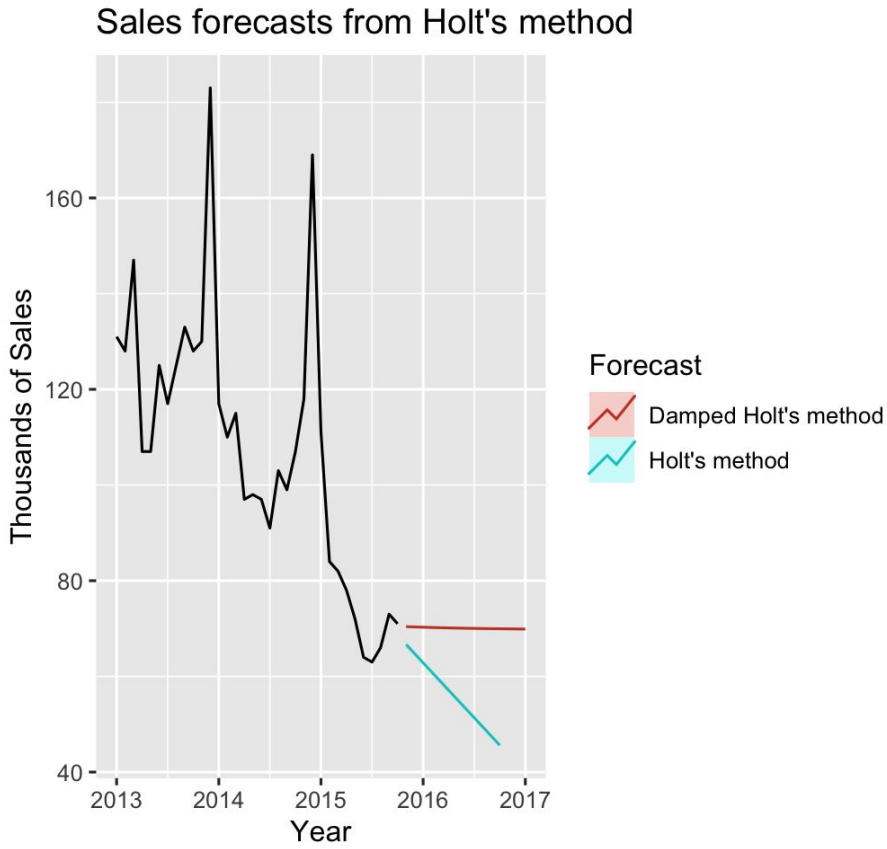





Forecasts from Simple exponential smoothing
Revenue (millions of Dollars)
150
100
50
0
2013
2014
2015
2016
Year
series
Fitted


Forecasts from Simple exponential smoothing
Thousands of Sales
200
160
120
80
2013
2014
2015
2016
Year
series
Fitted


Revenue forecasts from Holt's method
Revenue in Millions $)
150
100
50
2013
2014
2015
2016
2017
Year
Forecast
Damped Holt's method
Holt's method


Sales forecasts from Holt's method
Thousands of Sales
160
120
80
40
2013
2014
2015
2016
2017
Year
Forecast
Damped Holt's method
Holt's method

1 year Time Series Models for Prediction of Revenue

	Model_df1	MAPE_df1
1	Naive	15.9321024419297
2	Simple Exponential Smoothing	12.6278772378517
3	Holt Trend Method	9.66440991958067

1 year Time Series Models for Prediction of Items sold

	Model_df2	MAPE_df2
1	Naive	38.3458646616541
2	Simple Exponential Smoothing	33.5507246376812
3	Holt Trend Method	53.6304581703458

ML Algorithms for Items Sold and Revenue Predictions

Using the test dataset:

Linear Regression:

- Revenue:197.2242M
- Items Sold: 146.4998K

Lightgbm model:

- Revenue: 191.4288M
- Items Sold: 245.2691