MET AD571 Business Analytics Foundations

Assignment 4

Student name: **Kunfei Chen**

Student number: **U15575304**

Table of Contents

1. Executive Summary....3
2. Time Series Analysis on Sales....3
3. Multiple Regression Model for Forecast....4
4. Multiple Regression Model for Property Price Analysis....5
5. Appendix....6

1. Executive Summary

This report aims to apply R-Studio to import the NYC Real Estate data from the Boston University SQL server, and do time series analysis on the total dollar amount by quarters as well as use two multiple regression models to forecast for the sale and determine the sale of properties. The analyzed neighborhood is targeted at Bedford-Stuyvesant. The historical data is filtered by "Sale Price >100 and Gross Square Feet > 100" so as to remove meaningless data. The time series analysis applied Holt-Winters model for forecasting and estimated that the pattern in the next 8 quarters could be upward and with high sales at the fourth quarter but low sales at the next year's first quarter. Estate company is suggested to properly improve their sale price and conduct more publicity and marketing activities at quarter 4. The first multiple regression model select time series and four quarters as independent variables, its forecasted pattern is generally similar to that resulted by Holt-Winters model. The multiple regression model for property price analysis tested different formulas and variables, and finally identified the multiplicative pattern and five independent variables, which resulted in 0.6481 Multiple R-Squared. Based on this model, the comparison of each property's residuals finally determined the most underestimated estate and the most overpriced one. Both of them are elevator apartments, their difference might attribute to their location, so the estate company need to consider the factor of surrounding facilities when sell property.

2. Time Series Analysis on Sales

The total dollar amount of residential real estate in Bedford-Stuyvesant is aggregated into total sales per quarter. The time period for sales analysis is 10 years original from 2009 to 2019. According to the historical data of real estate, quarter sales of Bedford-Stuyvesant generally fluctuated in each quarter. Exponential smoothing model is firstly applied to fit the total dollar amount (See Figure 1), it can be seen that the level line showed an upward trend among this period, which means the average sales level steadily increased as time goes by. The behind reason might be the economy in this area continually developed so that resulted in the growth of sale price. In addition, the season line became more and more significant, which means the season effect is more important years, especially in recent years from 2016 to 2019. On observed pattern is that the value in current Quarter 1 is always lower than the value in last Quarter 4.

Since the analysis above shows that the pattern of sales has both trend and seasonality, the options of forecasting model could be Holt-Winters additive or multiplicative

seasonality mode with trend. In this case, the Holt-Winters additive seasonality model with trend is settled (See Figure 2, Figure 3) and the parameters are set as “alpha = 0.6, beta = 0.2, gamma = 0.6, seasonal = "additive"”. The value of alpha is relatively low, which means the estimated level in current period is based on data recently. The value of beta is relatively low, which means the estimation of the slop of the trend component is less updated on the time series. The value of gamma is also high, which indicates the estimation of seasonality is based on recent observed data. The forecast point for the next 8 quarters are 213383073, 264943855, 27188670, 365413133 in 2020 and 251876680, 303437461 in 2021 respectively (See Figure3, Figure 4). This pattern is similarly to the situation in 2018 and 2019, which displays an upward trend among a year but with a decrease in next year’s Quarter 1. Based on the analysis above, the recommendation for real estate company is that, since the generally economy environment estate is great, the sold price in Bedford-Stuyvesant can be properly improved to satisfy increasing demand. In addition, based on the pattern of high sales in Quarter 4 and low sales in Quarter 1, estate company should seize the opportunity in Quarter 4, conducting more publicity and marketing activities during this period.

3. Multiple Regression Model for Forecast

Time-series t (1, 2, 3, 4…) and quarter (Q1, Q2, Q3, Q4) are considered when running multiple regression model. The estimated coefficients are 4884292 for timeseries, 17180934 for Q2, 5301605 for Q3, and 15581968 for Q4 (See Figure 5). There is no coefficient for Q1 since Q1 is represented by number of 0. It is clear that the third quarter is relatively insignificant to regression compared with Q2 and Q4. The Multiple R-squared is 0.6159, which means that 61.59% observation can be explained by regression model. The equation of regression model (Quarter is dummy variable) is: Sales per Quarter = 32645468 + 4884292 * t + 17180934 * Q2 + 5301605 * Q3 + 15581968 * Q4.

According to the equation of regression model, forecast in next quarters is showed in Figure 6. The sales in the next 8 quarters basically follows the pattern discussed in the Holt-Winters model, which is the sales will decrease from Quarter 4 to next year’ Quarter 1 but remain a genera upward trend through a whole year. Since the coefficients for Q3 and Q4 are relatively high, it is recommended that estate company need to pay more attention to these two quarters and offer appropriate discount in another two quarters so as to attract more transactions.

4. Multiple Regression Model for Property Price Analysis

This analysis aims to run multiple regression model in terms of each property's price via the following independent variables: Sale Date, Year built, Building type, Gross Square Feet, Residential Units, and Commercial Units. The initial formula for this models is set as additive, assuming the sale price and other independent variables are linear correlated. Whereas its Multiple R-squared is 0.3642 (See Figure 7), showing that 36.42% observe data can be explained by this regression model. The low Multiple R-squared indicates this model is not great enough to fit these independent variables to the sale price. In other words, the relationship between sale price and other six independent variables is not significantly linear correlated. Therefore, the formula was changed as multiplicative, and the Multiple R-squared increased to 0.6297 (See Figure 8). However, the Building type variable seems to be redundant due to the fact that this model only aims to determine the residential property's sale price. Therefore, variable of type was removed in latter analysis and the analyzed data were filtered by "TYPE == 'RESIDENTIAL". Finally, the settled formula for this model was set as

"SALE_PRICE~SALE_DATE*YEAR_BUILT*GROSS_SQUARE_FEET*RESIDENTIAL_UNITS*COMMERCIAL_NUITS",

and the final Multiple R-squared is 0.6481 (See Figure 9). It can be seen that the among these five predictors, gross square feet have lowest estimated coefficient, which indicates gross square feet will not significantly affect the price since the unit price per gross square footage are totally different in different areas. The highest one is "commercial units", which means those residential estates with commercial units will significantly increase their price. In addition, it is clearly that the "residential units" variable is also statistically significant but the "sale date" variable is relatively meaningless. To some extent, the "year built" can affect the price, this might because public prefer to choose new building rather old one.

According to the regression model, range of residuals is from -11732859 to 24650057. It is interested that both the property with highest residuals the one with lowest residuals have the same building code id which is D3, its description is "elevator apartment; fireproof without stores". The reason of their difference could be their location. The most overpriced property is located in "1867-1875 Atlantic Avenue" (See Figure 12), there are only a few restaurants and shopping stores in its surrounding areas. On the

contrary, the address of the most underestimated property is “902 Bedford Avenue”, which is located in a prosperous area with abundant facilities and restaurants.

5. Appendix

Figure 1

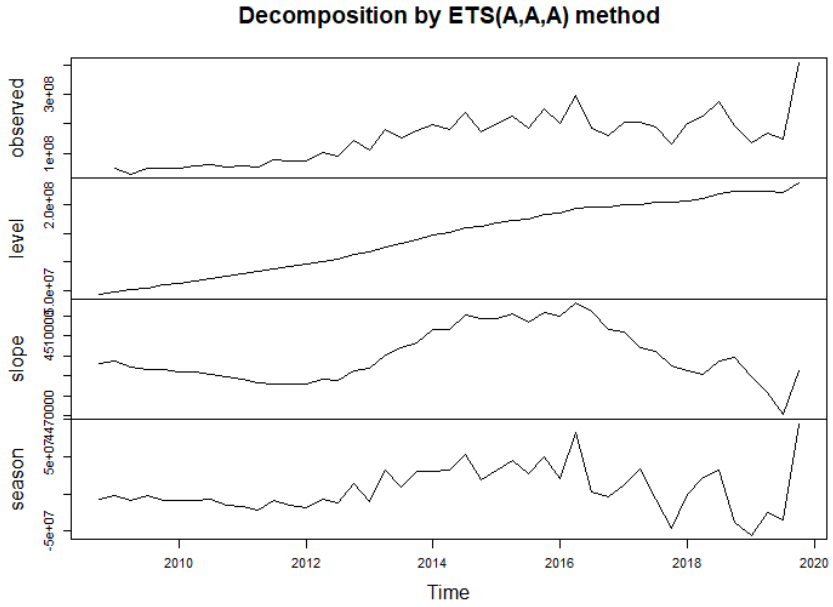


Figure 2

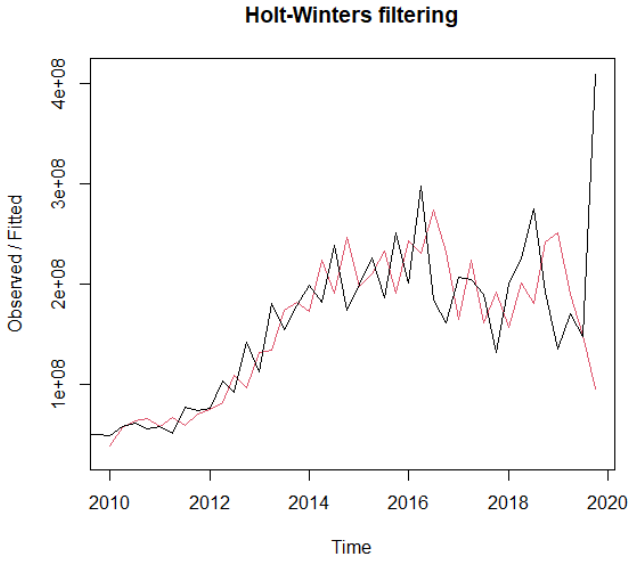


Figure 3

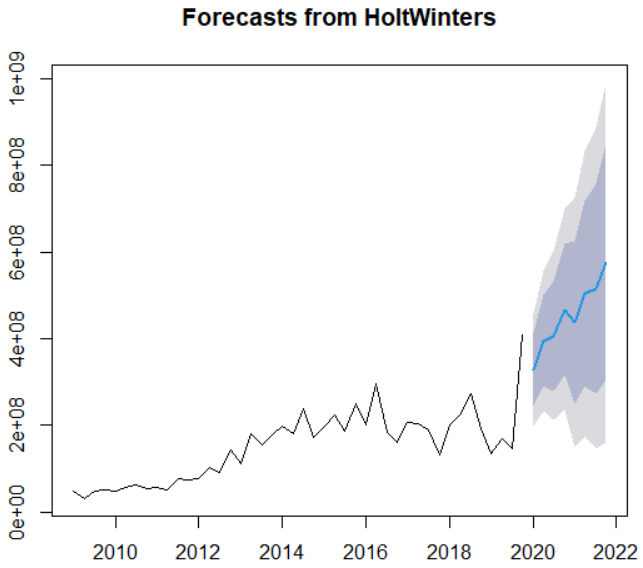


Figure 4

```
Forecast method: HoltWinters

Model Information:
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = timeseries_Bedford_Stuyvesant, alpha = 0.3, beta = 0.2,     gamma = 0.6, seasonal = "additive")

Smoothing parameters:
 alpha: 0.3
 beta : 0.2
 gamma: 0.6

Coefficients:
        [,1]
a  244540643
b    9623402
s1 -40780971
s2   1156409
s3  -1522177
s4  82378885

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 2454294 62576326 38527602 -3.01001 21.28785 0.8816072 -0.1166638

Forecasts:
        Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
2020 Q1      213383073 132229145 294537002  89268851 337497296
2020 Q2      264943855 178691316 351196394 133031980 396855730
2020 Q3      271888670 179145671 364631669 130050493 413726847
2020 Q4      365413133 264821538 466004729 211571564 519254703
2021 Q1      251876680 124643469 379109890  57290276 446463083
2021 Q2      303437461 167205176 439669746  95088161 511786762
2021 Q3      310382276 163998945 456765608  86508290 534256263
2021 Q4      403906740 246292522 561520957 162856596 644956883
```

Figure 5

```
Call:
lm(formula = TotalSales ~ t + Quarter, data = fore_Bedford_Stuyvesant)

Residuals:
       Min         1Q     Median         3Q        Max
-100809859  -20255571   -8929534   24940255  145921408

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 32645468   20633967   1.582    0.122
t            4884292     627581   7.783 1.85e-09 ***
QuarterQ2   17180934   22461795   0.765    0.449
QuarterQ3    5301605   22488081   0.236    0.815
QuarterQ4   15581968   22531824   0.692    0.493
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 52660000 on 39 degrees of freedom
Multiple R-squared:  0.6159,    Adjusted R-squared:  0.5765
F-statistic: 15.63 on 4 and 39 DF,  p-value: 1.027e-07
```

Figure 6

```
> predict.lm(regression_time, x, interval='confidence')
        fit       lwr       upr
1 252438599 208173014 296704183
2 274503824 230238240 318769409
3 267508788 223243203 311774372
4 282673442 238407857 326939027
5 271975766 224073682 319877850
6 294040991 246138907 341943076
7 287045955 239143871 334948039
8 302210609 254308525 350112694
```

Figure 7

```
Call:
lm(formula = SALE_PRICE ~ SALE_DATE + YEAR_BUILT + TYPE + GROSS_SQUARE_FEET +
    RESIDENTIAL_UNITS + COMMERCIAL_UNITS, data = Bedford_Stuyvesant_reg)

Residuals:
      Min        1Q    Median        3Q       Max
-23143915   -380145    -67864    226056  48844265

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -5.221e+06  3.352e+05 -15.575  < 2e-16 ***
SALE_DATE          4.186e+02  1.475e+01  28.374  < 2e-16 ***
YEAR_BUILT         2.146e+02  1.008e+02   2.128  0.03333 *
TYPEMIXED         -1.671e+06  1.123e+05 -14.888  < 2e-16 ***
TYPEOTHER         -6.288e+05  2.133e+05  -2.948  0.00321 **
TYPERESIDENTIAL   -1.536e+06  1.097e+05 -14.005  < 2e-16 ***
GROSS_SQUARE_FEET  4.914e+01  2.243e+00  21.909  < 2e-16 ***
RESIDENTIAL_UNITS  9.950e+04  3.541e+03  28.098  < 2e-16 ***
COMMERCIAL_UNITS   1.848e+05  3.512e+04   5.262 1.46e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1368000 on 7702 degrees of freedom
Multiple R-squared:  0.3642,    Adjusted R-squared:  0.3635
F-statistic: 551.5 on 8 and 7702 DF,  p-value: < 2.2e-16
```

Figure 8

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1051000 on 7597 degrees of freedom
Multiple R-squared:  0.6297,    Adjusted R-squared:  0.6241
F-statistic: 114.3 on 113 and 7597 DF,  p-value: < 2.2e-16
```

Figure 9

```
Call:
lm(formula = SALE_PRICE ~ SALE_DATE * YEAR_BUILT * GROSS_SQUARE_FEET *
    RESIDENTIAL_UNITS * COMMERCIAL_UNITS, data = Bedford_Stuyvesant_reg)

Residuals:
      Min        1Q    Median        3Q       Max
-11732859   -284958    -15196    231119  24650057

Coefficients:
                                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                     4.349e+07  9.708e+06   4.479 7.61e-06 ***
SALE_DATE                                                      -2.642e+03  5.584e+02  -4.731 2.28e-06 ***
YEAR_BUILT                                                     -2.264e+04  5.022e+03  -4.507 6.67e-06 ***
GROSS_SQUARE_FEET                                              -2.153e+03  3.524e+03  -0.611 0.541231
RESIDENTIAL_UNITS                                              -3.603e+07  5.545e+06  -6.498 8.69e-11 ***
COMMERCIAL_UNITS                                                1.028e+09  2.951e+08   3.482 0.000500 ***
SALE_DATE:YEAR_BUILT                                            1.380e+00  2.889e-01   4.776 1.82e-06 ***
SALE_DATE:GROSS_SQUARE_FEET                                     1.878e-01  2.142e-01   0.877 0.380742
YEAR_BUILT:GROSS_SQUARE_FEET                                    4.930e-01  1.831e+00   0.269 0.787697
SALE_DATE:RESIDENTIAL_UNITS                                     2.196e+03  3.375e+02   6.506 8.26e-11 ***
YEAR_BUILT:RESIDENTIAL_UNITS                                    1.839e+04  2.888e+03   6.368 2.04e-10 ***
GROSS_SQUARE_FEET:RESIDENTIAL_UNITS                             1.330e+03  9.792e+01  13.580  < 2e-16 ***
SALE_DATE:COMMERCIAL_UNITS                                     -6.594e+04  1.817e+04  -3.629 0.000287 ***
YEAR_BUILT:COMMERCIAL_UNITS                                    -5.296e+05  1.517e+05  -3.491 0.000484 ***
GROSS_SQUARE_FEET:COMMERCIAL_UNITS                              2.560e+05  6.310e+04   4.058 5.01e-05 ***
RESIDENTIAL_UNITS:COMMERCIAL_UNITS                             -3.633e+08  4.936e+07  -7.360 2.05e-13 ***
SALE_DATE:YEAR_BUILT:GROSS_SQUARE_FEET                         -5.267e-05  1.113e-04  -0.473 0.636123
SALE_DATE:YEAR_BUILT:RESIDENTIAL_UNITS                         -1.119e+00  1.758e-01  -6.364 2.10e-10 ***
SALE_DATE:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS                  -8.662e-02  5.907e-03 -14.663  < 2e-16 ***
YEAR_BUILT:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS                 -6.819e-01  4.947e-02 -13.784  < 2e-16 ***
SALE_DATE:YEAR_BUILT:COMMERCIAL_UNITS                           3.398e+01  9.342e+00   3.637 0.000277 ***
SALE_DATE:GROSS_SQUARE_FEET:COMMERCIAL_UNITS                   -1.636e+01  3.864e+00  -4.235 2.31e-05 ***
YEAR_BUILT:GROSS_SQUARE_FEET:COMMERCIAL_UNITS                  -1.337e+02  3.270e+01  -4.088 4.40e-05 ***
SALE_DATE:RESIDENTIAL_UNITS:COMMERCIAL_UNITS                    2.327e+04  3.021e+03   7.700 1.55e-14 ***
YEAR_BUILT:RESIDENTIAL_UNITS:COMMERCIAL_UNITS                   1.889e+05  2.575e+04   7.336 2.45e-13 ***
GROSS_SQUARE_FEET:RESIDENTIAL_UNITS:COMMERCIAL_UNITS            1.685e+03  9.159e+02   1.839 0.065913 .
SALE_DATE:YEAR_BUILT:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS        4.437e-05  2.982e-06  14.878  < 2e-16 ***
SALE_DATE:YEAR_BUILT:GROSS_SQUARE_FEET:COMMERCIAL_UNITS         8.545e-03  2.002e-03   4.267 2.01e-05 ***
SALE_DATE:YEAR_BUILT:RESIDENTIAL_UNITS:COMMERCIAL_UNITS        -1.210e+01  1.576e+00  -7.677 1.86e-14 ***
SALE_DATE:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS:COMMERCIAL_UNITS -1.067e-01  5.624e-02  -1.896 0.057947 .
YEAR_BUILT:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS:COMMERCIAL_UNITS -8.614e-01  4.640e-01  -1.856 0.063432 .
SALE_DATE:YEAR_BUILT:GROSS_SQUARE_FEET:RESIDENTIAL_UNITS:COMMERCIAL_UNITS  5.455e-05  2.850e-05   1.914 0.055670 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 906900 on 6917 degrees of freedom
Multiple R-squared:  0.6481,    Adjusted R-squared:  0.6466
F-statistic:   411 on 31 and 6917 DF,  p-value: < 2.2e-16
```

Figure 10

	BUILDING_CLASS_FINAL_ROLL	ADDRESS	SALE_PRICE	SALE_DATE	YEAR_BUILT	GROSS_SQUARE_FEET	RESIDENTIAL_UNITS	COMMERCIAL_UNITS	Residuals
1	D3	683-691 MARCY AVENUE	1250000	2014-02-04	2015	50929	41	0	-11732859
2	D1	420 TOMPKINS AVENUE	8750000	2017-06-23	2015	40723	44	0	-10320015
3	D3	683-691 MARCY AVENUE	4500000	2014-02-04	2015	50929	41	0	-8482859
4	D1	NA DEKALB AVENUE	3000000	2015-06-22	2015	29131	35	0	-7025671
5	D3	853 LEXINGTON AVENUE	3850000	2019-08-01	1931	26731	31	0	-6532237

Figure 11

	BUILDING_CLASS_FINAL_ROLL	ADDRESS	SALE_PRICE	SALE_DATE	YEAR_BUILT	GROSS_SQUARE_FEET	RESIDENTIAL_UNITS	COMMERCIAL_UNITS	Residuals
6949	D3	902 BEDFORD AVENUE	28600000	2014-09-11	2005	13000	18	0	24650057
6948	D3	902 BEDFORD AVENUE	18200000	2013-09-30	2005	13000	18	0	14556625
6947	B3	340 HART STREET	14187500	2019-12-20	1899	2400	2	0	12809757
6946	B3	903 LAFAYETTE AVENUE	14187500	2019-12-20	1899	3420	2	0	12434499
6945	C0	162 VERNON AVENUE	14187500	2019-12-20	1901	3152	3	0	12369363

Figure 12

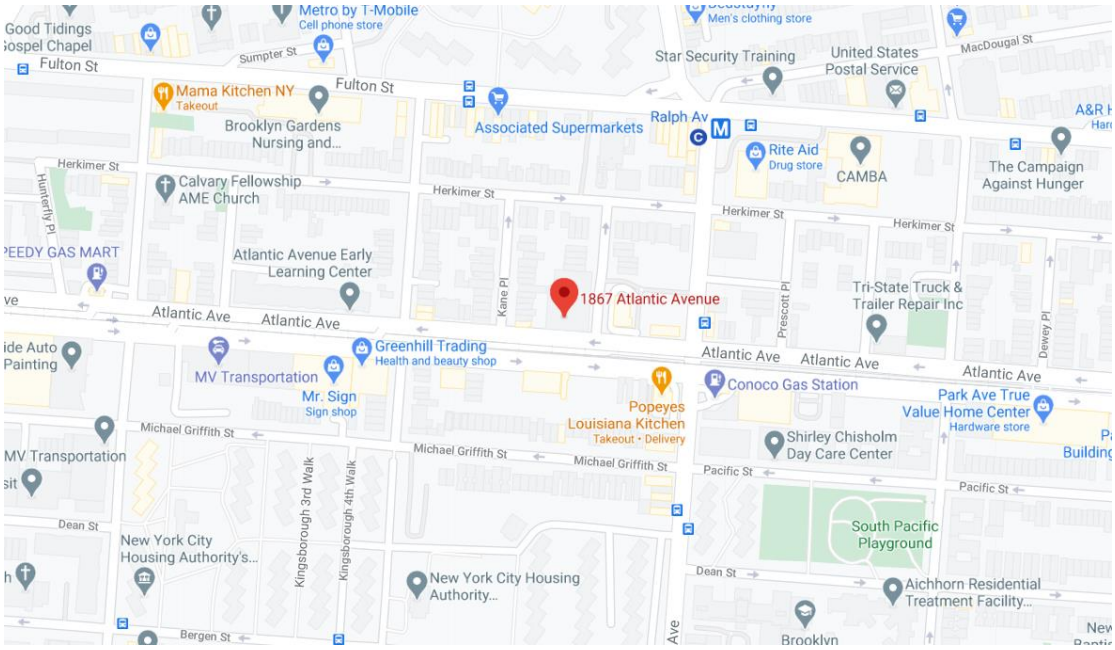
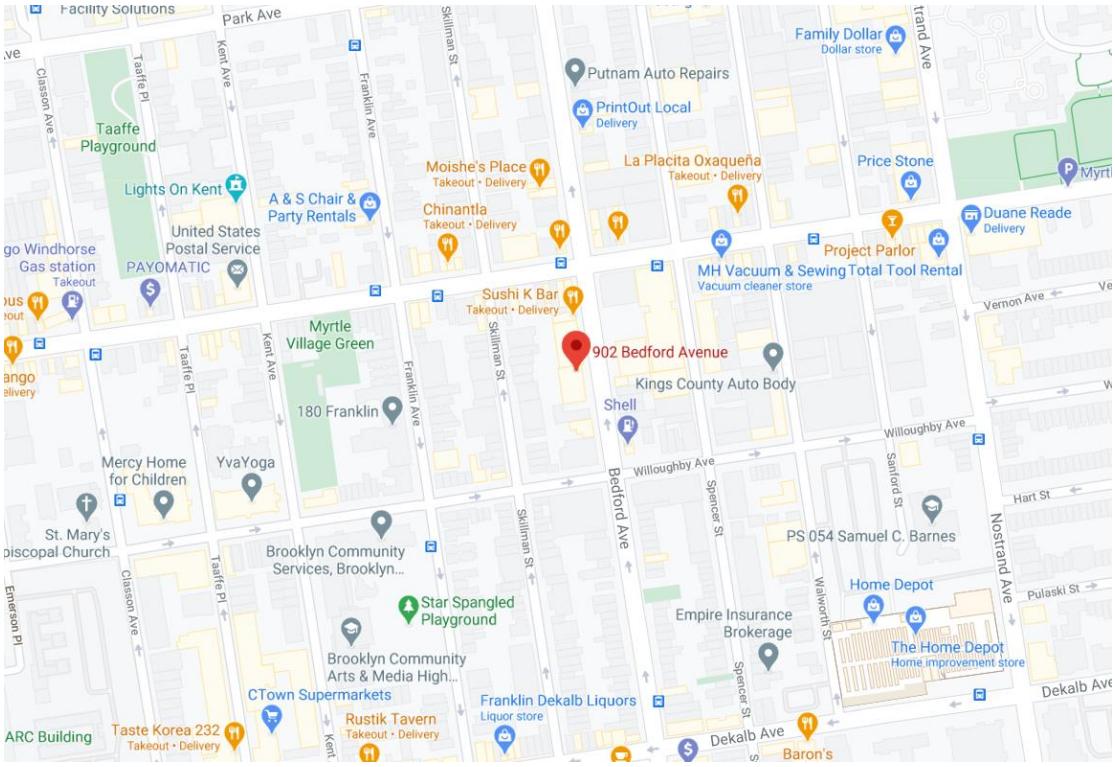

Metro by T-Mobile
Cell phone store
Good Tidings
Fulton St
Sumpter St
Mama Kitchen NY
Takeout
Brooklyn Gardens
Nursing and...
Fulton St
Associated Supermarkets
Ralph Av
Star Security Training
United States
Postal Service
MacDougal St
Rite Aid
Drug store
CAMBA
The Campaign
Against Hunger
Herkimer St
Calvary Fellowship
AME Church
Herkimer St
Herkimer St
Herkimer St
Atlantic Avenue Early
Learning Center
1867 Atlantic Avenue
Tri-State Truck &
Trailer Repair Inc
Atlantic Ave
Atlantic Ave
Greenhill Trading
Health and beauty shop
Atlantic Ave
Atlantic Ave
Atlantic Ave
MV Transportation
Mr. Sign
Sign shop
Conoco Gas Station
Popeyes
Louisiana Kitchen
Takeout · Delivery
Park Ave True
Value Home Center
Hardware store
Michael Griffith St
Michael Griffith St
Michael Griffith St
Shirley Chisholm
Day Care Center
Pacific St
Pacific St
Dean St
New York City
Housing Authority's...
Kingsborough 3rd Walk
Kingsborough 4th Walk
New York City Housing
Authority...
South Pacific
Playground
Dean St
Aichhorn Residential
Treatment Facility...
Bergen St
Brooklyn

Figure 13


Facility Solutions
Park Ave
Family Dollar
Dollar store
Putnam Auto Repairs
PrintOut Local
Delivery
Taaffe
Playground
Lights On Kent
A & S Chair &
Party Rentals
Moishe's Place
Takeout · Delivery
La Placita Oxaqueña
Takeout · Delivery
Price Stone
Chinantla
Takeout · Delivery
Duane Reade
Delivery
United States
Postal Service
PAYOMATIC
Project Parlor
MH Vacuum & Sewing
Vacuum cleaner store
Total Tool Rental
Sushi K Bar
Takeout · Delivery
Vernon Ave
Myrtle
Village Green
902 Bedford Avenue
Kings County Auto Body
180 Franklin
Shell
Willoughby Ave
Willoughby Ave
Mercy Home
for Children
YvaYoga
Hart St
St. Mary's
Episcopal Church
Brooklyn Community
Services, Brooklyn...
PS 054 Samuel C. Barnes
Home Depot
Star Spangled
Playground
Empire Insurance
Brokerage
Pulaski St
The Home Depot
Home improvement store
Brooklyn Community
Arts & Media High...
CTown Supermarkets
Franklin Dekalb Liquors
Liquor store
Dekalb Ave
ARC Building
Taste Korea 232
Takeout · Delivery
Rustik Tavern
Takeout · Delivery
Dekalb Ave
Baron's