

George Brown College

BUS 4045 – Capstone Project (Fiera Real Estate)

Final Report

Submitted By:

Ling Cheng

Chengcheng Cheng

Mehdi Hassanianesfahani

Yi Hu

MingQian Zhou

April 18, 2021

# Contents

1. Background: Introduction and statement of business problem	1
2. Methodology/Approach	2
3. Summary of creation of analytical file	3
3.1 Merge all Files and Change the Format	3
3.2 Create Derived Variables	3
3.3. Identified Target Variable	3
4. Summary of Data Audit	4
5. Analytical Results	5
5.1 Correlation Report	5
5.2 EDA's of top 5 correlated variables	6
5.3 Final model report	8
5.4 Decile/gains/AUC chart if building model	9
5.5 key insights/findings from all analytical results	10
6. Conclusions/Next steps	11

## 1. Background: Introduction and statement of business problem

Fiera Real Estate was founded as a global real estate investment management company almost three decades ago in the UK to provide direct real estate investment opportunities to institutional investors and high net worth investors. They equip their clients with an ongoing range of services, including property, financial and legal expertise, and assistance with structuring transactions, sourcing finance, and advising with negotiations and management matters and now are considering operating globally.

Fiera Real Estate Canada is a relatively new division of Fiera Real Estate Global with offices in Toronto, Montreal, and Halifax. Their objective is to protect and grow the money they manage on behalf of the clients and partners by investing directly in exceptional commercial real estate assets globally. With the help of David Bender, the associate director of Fiera Real Estate, we are using advanced analytical tools to explore the available data of Canada's demography and commercial real estate data and help them make informed investment decisions.

As an investment management company, Fiera Real Estate's first challenge is determining the best geographical location for their next investment. Canada is vast, and each city and province is different. In the initial data exploration, we analyze the historical data of all major cities in Canada to compare their population, consumer price index, and the population growth rate of the past ten years. We are aiming to find a meaningful trend and select our top 10 cities of this report.

The second challenge for Fiera Real Estate is determining the type of property their stakeholders can invest in. We narrowed down the scope of this study to 'office buildings' and studied seven types of office space, namely: Class A, Class AA, Class AAA, Class B, Class C, Class RC & Class C + D. Using analytical tools, we study the variables of the detailed existing data of office properties in selected cities, to identify high propensity target markets, and design a predictive model to predict the weight of each variable in the profitability of an investment.

## 2.Methodology/Approach

In this report, we use Statistics Canada datasets and the exclusive data provided by Fiera Real Estate (assumed to be valid and accurate) through a statistical approach to create a propensity model for future investments. Our objective is to calculate the significance of each related variable in a successful investment project and design a predictive model to assess available projects and help stakeholders make an intelligent decision.

We calculate the Return on investment for the stakeholders as 'net rental rate' minus 'taxes and operation costs' divided by 'taxes and operation costs,' disregarding the initial investment of acquiring the office building. In other words, we would like our Model to compare two office buildings in different cities and tell us which one brings more rental profit to the owner. The metric ROI shows how much the investors earn from rental properties.

$$ROI = \frac{\text{Net Rental Rate} - \text{Taxes and Operating Costs}}{\text{Taxes and Operating Costs}}$$

We use Microsoft Excel for preliminary statistical analysis and data cleaning and use Python Jupyter to calculate and analyze the correlation coefficient of each variable to ROI. Correlation analysis defines vital variables that are relevant to our target variable and could be used for modeling or reporting.

Finally, we create a correlation–response model of all independent variables. Our correlation–response model should compare different cities or types of office buildings and find the best and worst investments.

Lastly, we audit the Model and check the decile and grains graphs to determine the difference between investing with and without our Model to see whether and to what extent it is beneficial to Fiera Real Estate's stakeholders.

In this report, we use the following Python packages for our primary data exploration and cleaning: NumPy, pandas, matplotlib, pyplot, seaborn, scipy.stats & scipy. To fit a regression model, we use the following Python packages: pandas, sklearn.model\_selection, sklearn.linear\_model, statsmodels.formula.api, matplotlib.pyplot & dmba. To validate the performance of our Model, we use the following Python packages: pathlib, pandas, sklearn.model\_selection, sklearn.linear\_model, sklearn.metrics, matplotlib.pyplot & dmba.

### 3.Summary of creation of analytical file

We have created the analytical file with a random extract of five records where variables are labeled as either source, derived, or target.

The data source consists of eight data files, including the information of various class types of offices in eight Canadian regions from 2011 to 2020, respectively. We have done the following step to create the analytical file:

#### 3.1 Merge all Files and Change the Format

We change the format of the data files from time-based to region-based. And then merge all eight files into one analytical file.

#### 3.2 Create Derived Variables

There are 14 source variables in the analytical file. We also create some derived variables based on the source variables. For example, we create dummy variables for the region and the class type. And we use the calculation to create "Change of Vacancy Rate YTY" to track how the vacant rate is changing year over year.

#### 3.3. Identified Target Variable

Our goal is to determine the "Return on investment (ROI) "among all types of properties across Canada. Therefore, our target variable is ROI.

Therefore, we calculated the ROI, which is the ultimate guide to our investors, and the purpose is to maximize it or invest in a market with the highest success ratio. Finally, we deleted the variables that are part of the derived variable, such as 'Net Rental Rate' and 'Taxes and Operating Costs', which are part of the calculation of the target variable 'ROI'. And we also deleted the independent variable related to the rent and cost: ' Net Effective Rent' and 'Gross Effective Rent.'

To sum up, the creation of an analytical file contains cleaning, organizing, creating new variables. We highlighted the best and the worst property class of each city and the best and the worst property class in Canada based on the 'ROI' for a better presentation.

Model Variables	Type	Impact on ROI	Contribution to Overall Equation
Region_Vancouver	Source	Positive	20.68%
Direct Vacancy	Source	Negative	14.10%
Class Type_AAA	Source	Positive	7.97%
Region_Ottawa	Source	Negative	7.54%
Class Type_AA	Source	Positive	6.83%

## 4. Summary of Data Audit

We have conducted the diagnostics report, and the result shows that is a sample from the database. There are total 11 variables, including 10 dependent variables and a target variable.

	Variable	# of Records	Format	# of Unique Vales	# of Missing Value
0	Region	280	object	8	0
1	Year	280	int64	10	0
2	Class Type	280	object	7	0
3	ROI	279	float64	113	1
4	Number of Buildings	202	float64	120	78
5	Occupied	279	float64	279	1
6	Absorption YTD	279	float64	278	1
7	New Supply YTD	279	float64	119	1
8	Direct Vacancy	279	float64	279	1
9	Sublease Vacancy	279	float64	256	1
10	Under Construction	249	float64	121	31

*Figure 1 Summary of diagnostic database*

Figure 1 is the summary of the diagnostic database. The total of records is 280. For most variables in the analytical file, there is only one missing value. Number of buildings contains 78 missing values, but it does not affect our target variable in the analytical file.

Class Type	# of Records	Year	# of Records
A	80	2011	28
B	80	2012	28
C	60	2013	28
AAA	10	2014	28
RC	20	2015	28
C_D	10	2016	28
AA	20	2017	28
		2018	28
		2019	28
		2020	28

*Table 1 Summary of distribution report*

Table 1 shows the frequency distributions, and it shows that each variable is different for each field, due to the database contains different types of variables and some variables have missing values from 2011 to 2020. The year is the only variable in this report because each region collects the same time length of the data.

## 5. Analytical Results

### 5.1 Correlation Report

After recognizing the main variables of our Model and creating a target variable, we conducted a correlation analysis to determine which variables have the highest correlation coefficient and statistical significance with ROI. ROI can be interpreted as Return of Investment and determines whether investing in a particular office building is a success or not.

Variable	Correlation Coefficient	t Stat	P-value	Relevant?
Year	-0.20	-3.3905	0.0007988	Relevant
No. of Buildings	-0.27	-4.6649	0.0000048	Relevant
Occupied	-0.14	-2.3134	0.0214294	Relevant
Absorption YTD	0.18	3.0017	0.0029284	Relevant
New Supply YTD	0.10	1.6907	0.0920121	Irrelevant
Direct Vacancy	-0.35	-6.3003	0.0000000	Relevant
Sublease Vacancy	-0.23	-3.8628	0.0001395	Relevant
Under Construction	0.12	1.9317	0.0544146	Irrelevant
Region_Edmonton	0.03	0.5227	0.6016234	Irrelevant
Region_Halifax	-0.14	-2.4055	0.0168021	Relevant
Region_Montreal	0.12	2.0679	0.0395766	Relevant
Region_Ottawa	-0.26	-4.4619	0.0000118	Relevant
Region_Toronto GTA	-0.07	-1.1064	0.2695160	Irrelevant
Region_Vancouver	0.43	7.8966	0.0000000	Relevant
Region_Winnipeg	-0.04	-0.7375	0.4614617	Irrelevant
Class Type_Class AA	0.25	4.2323	0.0000315	Relevant
Class Type_Class AAA	0.27	4.5957	0.0000065	Relevant
Class Type_Class B	-0.21	-3.5305	0.0004853	Relevant
Class Type_Class C	-0.20	-3.4287	0.0006985	Relevant
Class Type_Class RC	0.46	8.5283	0.0000000	Relevant
Class Type_Classes C + D	-0.23	-3.8908	0.0001251	Relevant

Table 2 Correlation Coefficient

Our Correlation Analysis report shows 'Region\_Vancouver', 'Direct Vacancy', and 'Class Type\_Class RC' have high correlation coefficients with ROI, indicating a robust linear relationship between them and ROI with a p-value of next to zero, which makes the prediction quite reliable. It is also noteworthy that the 'Year' and 'No. of Buildings' have a value less than zero, which signifies a negative relationship with ROI.

Consequently, we removed the independent variables with no or very little relevance to our target variable and divided our dataset to be used for training and validation.

## 5.2 EDA's of top 5 correlated variables

The correlation coefficient value of each variable is determined in the Correlation Report. Then the EDA of each variable will confirm the value of the Correlation Coefficient, and the Y value is the variable ROI.

### 5.2.1 Region - Vancouver:

Region - Vancouver	# of Observations	ROI
YES	40	63%
NO	239	11%
Total/Average	279	69%

*Table 3 Region - Vancouver*

There are 279 observed values, which are dummy variables. There are 40 observations data in Montreal and 239 observation data outside Vancouver. We can see that the value of ROI is 63% and 11%, respectively. There is a significant difference, the value of ROI in Vancouver is much larger than the value not in Vancouver. So, ROI and the variable Region-Vancouver are in a positive relationship. Compared with the value of 0.43 of the Correlation Coefficient variable Region-Vancouver, it is consistent with the positive association.

### 5.2.2 Class Type\_Class RC:

Class Type_Class RC	# of Observations	ROI
YES	19	73%
NO	260	15%
Total/Average	279	44%

*Table 4 Class RC*

There are 279 observed values, which are dummy variables. The Class Type of the 19 observations is Class RC, and 260 observations data is the other types. Table 4 shows that the ROI of Class RC is much larger than the ROI of other Class types. They are in a significant difference. So, ROI and the variable Class Type\_Class RC are in a positive relationship. Compared with the value of 0.46 of the Correlation Coefficient variable Class Type\_Class RC, it is consistent with the positive relationship.

### 5.2.3 Class Type\_Class AAA:



Class Type_Class AAA	# of Observations	ROI
YES	10	69.00%
NO	269	11.00%
Total/Average	279	40.00%

*Table 5 Class AAA*

There are ten observations with Class Type AAA and 269 observations are for the other Class Types. We can see a significant difference between the two ROI, where Class Type AAA has a higher ROI than other types. Therefore, the value of the Correlation coefficient variable (0.27) has supported the EDA here.

#### 5.2.4 Class Type\_Class AA:

Class Type_Class AA	# of Observations	ROI
YES	20	48.00%
NO	259	14.00%
Total/Average	279	31.00%

*Table 6 Class AA*

There are 279 observed values. Table 6 shows that the value of ROI for Class type AA is 48%, which is much larger than the average ROI of 31%, while the ROI value of other Class types is 14%. So ROI and the variable Class Type\_Class AA are in a positive relationship. Comparing the value of the Correlation Coefficient, variable Class Type\_Class AA with 0.25, which is consistent with the positive relationship.

#### 5.2.5 Absorption YTD:

Absorption YTD	# of Observations	ROI
-1,761,020 to -150,235	56	5%
-145,922 to -9,989	56	13%
-8,191 to 65,813	56	20%
70,033 to 286,613	56	25%
293,512 to 2,873,734	55	33%
Total/Average	279	19%

*Table 7 Absorption YTD*

There are 279 observed values, and the data for each interval is set to 56. Table 7 shows that the value of ROI is significantly increasing, so ROI and the variable Absorption YTD are in a positive relationship. Compared with the previous Correlation Coefficient variable Net Rental Rate value of 0.18 is consistent with the result of the positive relationship.

### 5.3 Final model report

We conducted a linear regression analysis on the training dataset and produced the MAE to find the best predictive Model. Then computed the common accuracy measures of MAE and MAPE of our predictive models and concluded that Forward Elimination is the best fit. A linear regression analysis on the selected variables confirmed the Model fits well and produced the following results:

Variable	Impact on ROI	Contribution to Overall Equation
Region_Vancouver	Positive	21.48%
Direct Vacancy	Negative	14.65%
Class Type_Class AAA	Positive	8.28%
Region_Ottawa	Negative	7.83%
Class Type_Class AA	Positive	7.10%
Class Type_Classes C + D	Negative	6.05%
Sublease Vacancy	Negative	5.97%
Class Type_Class B	Negative	5.03%
Class Type_Class C	Negative	4.76%
Year	Negative	4.66%
Absorption YTD	Positive	3.68%
Occupied	Positive	2.21%

*Table 8*

*Table 8* has shown the impact and contribution to the Model of 12 variables.

In terms of the contribution to the Overall Equation, 'Region\_Vancouver' appears to be the most driven factor, and the variable 'Occupied' has the most negligible impact on the target variable. It also shows that the contributions of the first five variables are much more significant than other variables, which indicates that they can interpret most of the variance of the target variable, ROI.

Accordingly, we realized Vancouver seems to be the hottest place for investment which yields the best Return, while Ottawa is the riskiest of all with a possible negative profit from investment in the office market.

In short, Vancouver turns out to be the single best place for investment in office space. Furthermore, the type of office yielding more profit is class AAA, and the variable to consider which brings the highest profit is 'Direct Vacancy' (with a negative impact on ROI).

#### 5.4 Decile/gains/AUC chart if building model

Gain and Lift charts are used to evaluate the performance of our final Model. First, we import the modules required from Python libraries, then reorder the predicted values in descending order and calculate the cumulative sum and produce the Gains chart:

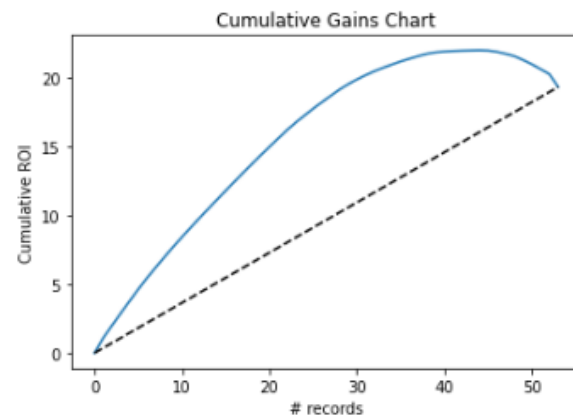


Figure 2 Gains Chart

Gain at a given decile level is the ratio of the cumulative number of targets up to that decile to the total number of targets in the entire data set. The blue line is a predicted ROI using the Model, while the black dotted line is the ROI of random data selected without the Model. Figure 2 shows our predictive Model performs well and increases the investors' ROI. We can also present it in a Decile Lift Chart. The lift Chart measures how much better one can expect to do with the predictive Model compared without a model. It is the ratio of gain % to the random expectation % at a given decile level.

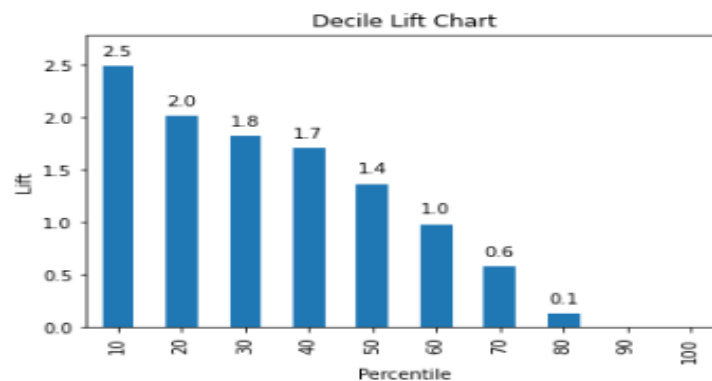


Figure 3 Decile Lift Chart

Figure 3 above shows when randomly selecting a dataset, 10% of the dataset performs 2.5 times better with our Model than without our Model. This number is cumulative, meaning when selecting 20% of the records using our Model, one can expect two times more ROI than selecting 20% of the records from the without-a-model dataset.

## 5.5 key insights/findings from all analytical results

We calculated the correlation coefficient and statistical significance (t-stat and p-value) between the target variable "ROI" and each independent variable in our correlation analysis. We discovered with statistical significance criterion of 95%, 'Net Effective Rent,' 'Net Rental Rate' and 'Gross Effective Rent' have the highest correlation coefficients and the 'Year' and 'No. of Buildings' have a value less than zero which signifies a negative relationship with ROI. Consequently, we removed the independent variables with no or very little relevance to our target variable and divided our dataset to be used for training and validation.

Next, we conducted a linear regression analysis on the training dataset and produced the MAE to find the best predictive Model. Then computed the common accuracy measures of MAE and MAPE of our predictive models and concluded that Forward Elimination is the best fit. A linear regression analysis on the selected variables confirmed the Model fits well and produced the following results. Table 7 displays the top five variables that contribute the most to the equation:

Variable	Impact on ROI	Contribution to Overall Equation
Region_Vancouver	Positive	21.48%
Direct Vacancy	Negative	14.65%
Class Type_Class AAA	Positive	8.28%
Region_Ottawa	Negative	7.83%
Class Type_Class AA	Positive	7.10%

*Table 2 Contribution to the top 5 Variables*

Accordingly, we realized the highest positive impact is by 'Region\_Vancouver,' meaning that Vancouver seems to be the hottest place for investment which yields the best Return, while Ottawa is the riskiest of all with a possible negative profit from investment in the office market. After measuring the accuracy of our Model, the Gain and Lift charts show randomly selecting a dataset. 10% of the dataset performs 2.5 times better with our Model than without our Model.

In conclusion, Vancouver turns out to be the single best place for investment in office space. Furthermore, the type of office yielding more profit is class AAA. The variables to consider which bring the highest profit are net and gross effective rent Direct Vacancy (with a negative impact on ROI).

## 6.Conclusions/Next steps

We have picked Canada Office Data File as our main source of analysis. The data provide us with the information of Office Class, Vacancy Rate, Net Effective Rent, etc. Therefore, our conclusion was based on which place is yielding the best ROI. Although we have concluded that Vancouver is the hottest spot to invest, the data we've been using was still too one-sided. The future challenge would be conducting a more thorough analysis since the real estate investment should not only look at the ROI but also need to consider other variables such as the initial capital required for such an investment and other miscellaneous expenses. To do so, we may analyze a vast database and determine the valuable variables that will be more challenging than this time.

To determine which factors or variables significantly impact ROI, we need to analyze the correlation between these variables and ROI. From the analytical file, ROI is strongly related to the regions. However, using ROI as the only index cannot demonstrate the whole situation of Fiera Real Estate. More variables should be considered, and more variables need to be analyzed to solve this challenge.

This analytical file focuses on the financial aspect of Fiera Real Estate. Although there are plenty of variables to show how the company operates in different scenarios, the market for the real estate itself is not very clear. To solve this challenge, we need more files from vast database containing more information—for instance, the population and demographic of a specific area. In order to getting a more exhaustive report, we can develop a specific model for that notable area. Besides, investigating the relationship between more variables in the market will help the analytical file be more comprehensive. Once conducted a more detailed analysis with the support of the vast database of all other influencing factors, we could combine the result with ROI to do a more in-depth analysis.