Capstone Project: Fiera Real Estate

How to use data to make better decisions in Canada real estate industry

BUS 4045
Group # 7
Winter 2021

By:
Yi Hu
Ling Cheng
Mingqian Zhou
Chengcheng Cheng
Mehdi Hassanianesfahani

Statement of Problem

FIERA CAPITAL

- UK based
  global real estate investment
  management company
- Expanding to Canada's real estate industry

Business problem:

- Where to invest next:
- Geographical location
- The future value of the building
- What type of building is more profitable

Objective:

- Profitability / Return on Investment

Research Methodology

Data: Statistics Canada & internal data (assumed to be valid and accurate)

Tools: Microsoft Excel
Python Jupyter

Correlation Analysis Report
Target variable: Return on investment (ROI)

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

Approach:

Basic Statistics, label variables and define target variable

Create a correlation model to select related data

EDA, Create the final model, check decile/gains

Summary of Creation of Analytical File

Merge all Files and Change the Format

We change the format of the data files from time-based to region-based.

Merged the selected eight files into one analytical file.

Create Derived Variables

We create some derived variables based on the source variables.

Identified Target Variable

Our goal is to determine the Return on Investment (ROI) among all types of office properties across Canada.

Delete Variables Being Part of the Derived Variable

Deleted ‘Net Rental Rate’, ‘Taxes and Operating Costs’, ‘Net Effective Rent’, ‘Gross Effective Rent’, and ‘Tenant Inducements’, which are part of the ‘ROI’.

Summary of Data Audit: (1) Sample Data

	Region	Year	Class Type	ROI	Number of Buildings	Occupied	Absorption YTD	New Supply YTD	Direct Vacancy	Sublease Vacancy	Under Construction
0	Montreal	2011	A	0.10	NaN	32303128.0	579560.0	187181.0	2307229.0	409036.0	634642.0
1	Montreal	2012	A	0.08	NaN	32309302.0	52555.0	353965.0	2609575.0	483135.0	1581613.0
2	Montreal	2013	A	0.08	NaN	32314695.0	5393.0	678781.0	2869355.0	896743.0	1730547.0
3	Montreal	2014	A	0.12	NaN	33492685.0	-241755.0	915071.0	4018696.0	1084792.0	1563967.0
4	Montreal	2015	A	0.12	NaN	33700015.0	-116627.0	630067.0	4477727.0	1019955.0	2130512.0
...	...	...	...	...	...	...	...	...	...	...	...
275	Vancouver	2016	C	0.40	156.0	5635742.0	-14516.0	0.0	432901.0	9746.0	0.0
276	Vancouver	2017	C	0.31	170.0	6387341.0	44374.0	0.0	356706.0	41567.0	0.0
277	Vancouver	2018	C	0.51	168.0	6333740.0	127227.0	0.0	253644.0	17402.0	0.0
278	Vancouver	2019	C	0.62	182.0	6690245.0	-272877.0	0.0	407085.0	117216.0	0.0
279	Vancouver	2020	C	0.69	183.0	7022157.0	19622.0	0.0	233220.0	18204.0	0.0

280 rows × 11 columns

* A sample from the database: there are 11 variables in total, including 10 dependent variables and 1 target variable.

(2) Diagnostic Report

* The summary of the diagnostic database with 11 variables.

The total of records is 280 and there are only 2 variables with missing value. i.e., “Number of buildings“ contains 78 missing values, but it does not affect our target variable in the analytical file.

	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

(3) Frequency Distributions

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

* Notice that each variable is different as the database contains different types of variables.

Analytical Results: Correlation Analysis Report

Total 21 independent variables.

Relevant: 16 variables

Irrelevant: 5 variables

Strong positive relationship with ROI:

Region_Vancouver

Class Type_Class RC

Strong negative relationship with ROI:

Direct Vacancy

No. of Buildings

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

Exploratory Data Analysis Reports (EDA)

Region - Vancouver	# of Observations	ROI
YES	40	63.0%
NO	239	11.0%
Total/Average	279	68.5%

- The value of ROI are 63% and 11%, respectively.
- ROI and the variable Region-Vancouver are in a positive relationship.
- It is consistent with the result of the positive relationship.

Class Type_Class RC	# of Observations	ROI
YES	19	73.0%
NO	260	15.0%
Total/Average	279	44.0%

- The ROI of Class RC is much larger than the ROI of other Class types.
- ROI and the variable Class Type_Class RC are in a positive relationship.
- Compared with the value of 0.46 of the Correlation Coefficient.

Analytical Results: Final Model Variable Report

Total 12 variables in the final model.

The top five variables contribute to 60% of the variance of ROI.

Most contributing variable:

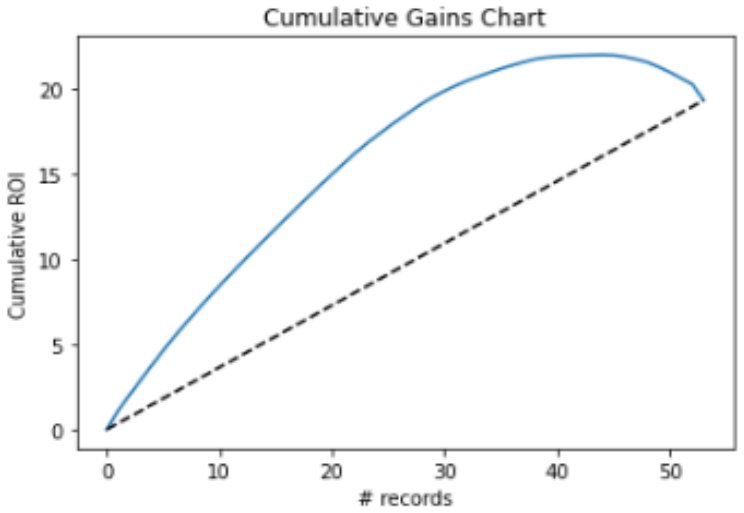
Region_Vancouver

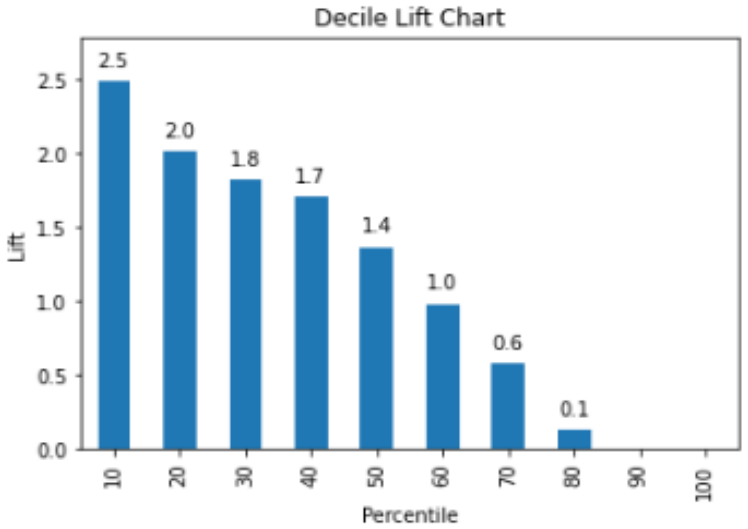
Least contributing variable:

Occupied

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%

Decile / Gains Chart





The gains chart shows the model performs well because the blue line is higher than the black dotted line.

The Lift charts show randomly selecting a dataset, 10% of the dataset performs 2.5 times better with our model than without our model.

Conclusion

- Vancouver is the hottest spot to invest in Office Space,

  since Vancouver yields the best return among all the regions.
- Ottawa will be the riskiest place to invest in the Office Space.
- Class AAA and Class AA are the two classes that collect the best ROI among all the class type.
- Class C + D office properties should be avoided as they are the riskiest types & less profitable.

Next Step

- Expanding the scope of data collection
- More variables should be taken into consideration, and to be analysed
    - Population and demographic of a specific area
    - Develop a specific model to that specific area
- Investigating the relationship between more variables in the market
- The initial capital required for the investment is also a factor for the future analysis report