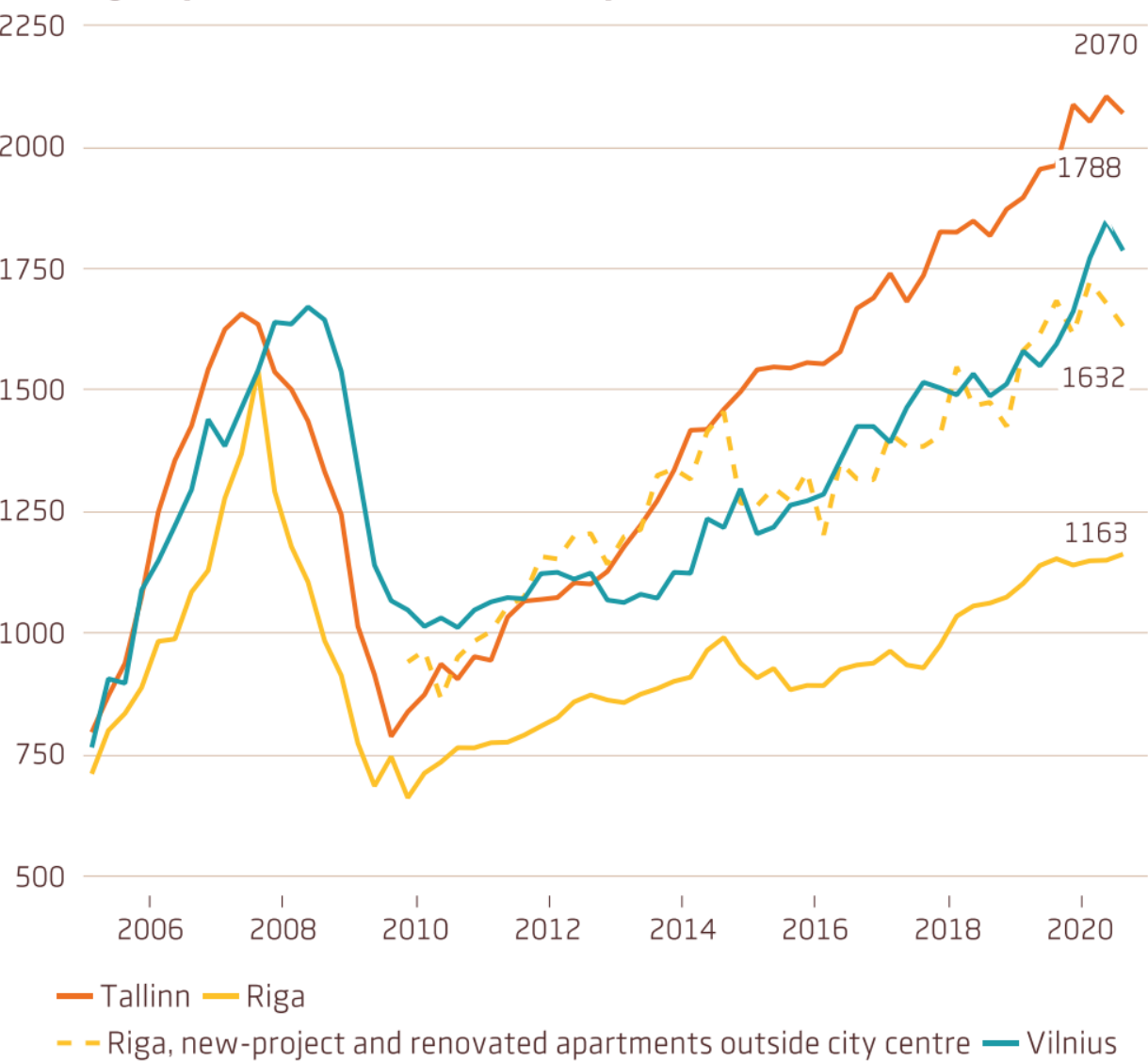




How can Data Analytics help in
predicting house prices?
Cracking Riga real estate market



Average apartment transaction price, EUR/m2



Sources: Swedbank Research & Macrobond



Maskavas iela 46 - 3, Centrs

Rīga

120 000 EUR

1 809.95 EUR/m²

Monthly payment 403 EUR

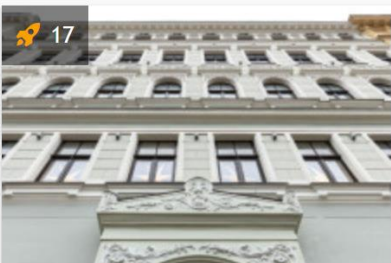
66.3 m²

4 Rooms

3/4 On Floor

Apartments, For Sale

Rēdera Nami, Amela Al- Asbahi



Krišjāņa Barona iela 6 - 13, Centrs

Rīga

Project of the month

112 960 EUR

3 200 EUR/m²

Monthly payment 379 EUR

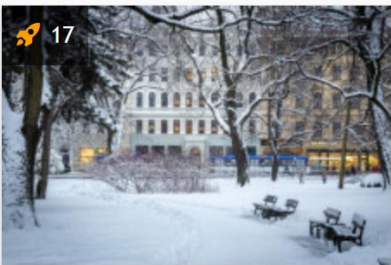
35.3 m²

2 Rooms

5/5 On Floor

Apartments, For Sale

Riverproperties, Laura Semjonova



Krišjāņa Barona iela 6 - 12, Centrs

Rīga

Project of the month

99 900 EUR

3 000 EUR/m²

Monthly payment 336 EUR

33.3 m²

2 Rooms

2/5 On Floor

Apartments, For Sale

Riverproperties, Laura Semjonova



Krišjāņa Barona iela 6 - 11, Centrs

Rīga

Project of the month

102 300 EUR

3 000 EUR/m²

Monthly payment 344 EUR

34.1 m²

2 Rooms

2/5 On Floor

Apartments, For Sale

Riverproperties, Laura Semjonova



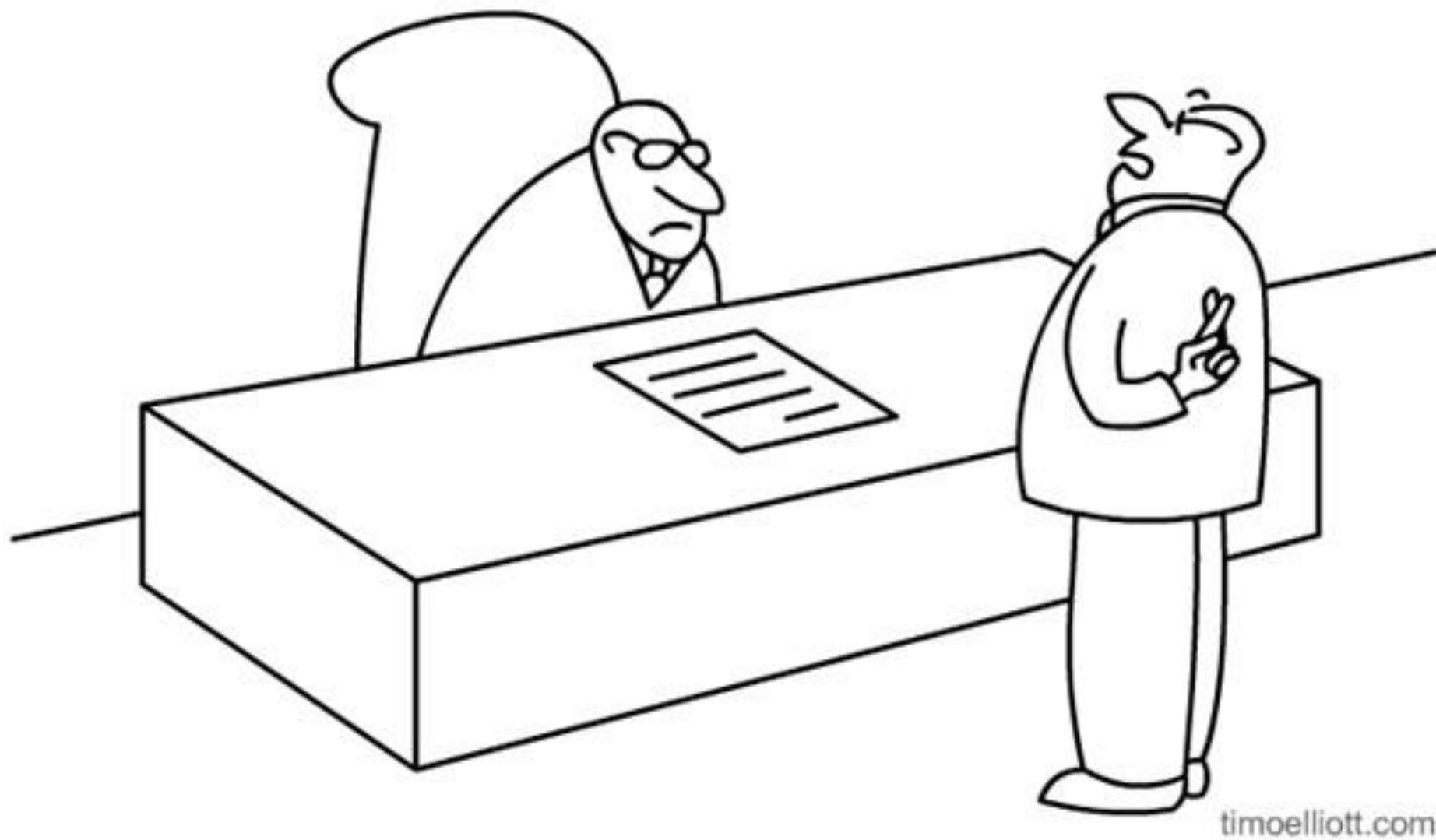




Benchmarking of houses







timoelliott.com

"Yes sir, you can absolutely trust those numbers"



www.timoelliott.com

“When you two have finished arguing your opinions, I actually have data!”



“If you don't reveal some insights soon, I'm going to be forced to slice, dice, and drill!”







**KEEP
CALM**

AND

**LET`S SEE WHAT WE
FOUND OUT**

AGENDA

1 What we analyzed

Setting the context | Defining the problem statement

2 What we found

Conclusions and unexpected insights

3 How we analyzed

What goes in the background in making all of this...



1 What we analyzed

2 What we found

Analysis of data regression analysis

3 How we analyzed


Conclusions and notes





Setting the context


ČETRAS RĪGAS

Biežākais mājokļa celšanas laiks (01.03.2011., %)

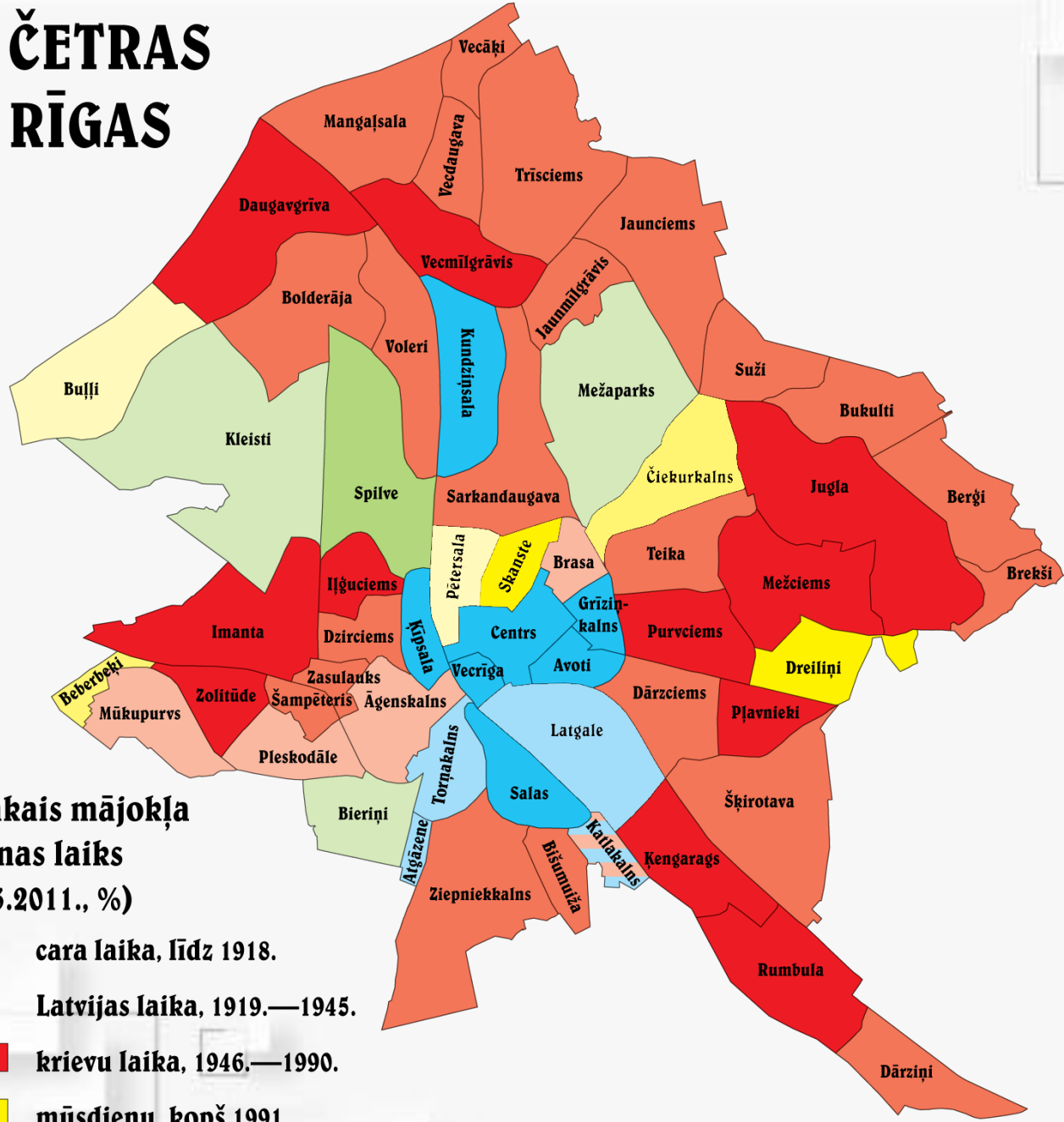
 cara laika, līdz 1918.

 Latvijas laika, 1919.—1945.

 krievu laika, 1946.—1990.

 mūsdienu, kopš 1991.

<50
50...90
>90



Examples of house types

Classical



Stalin era houses



New houses



Neo Classical



Soviet mass housing

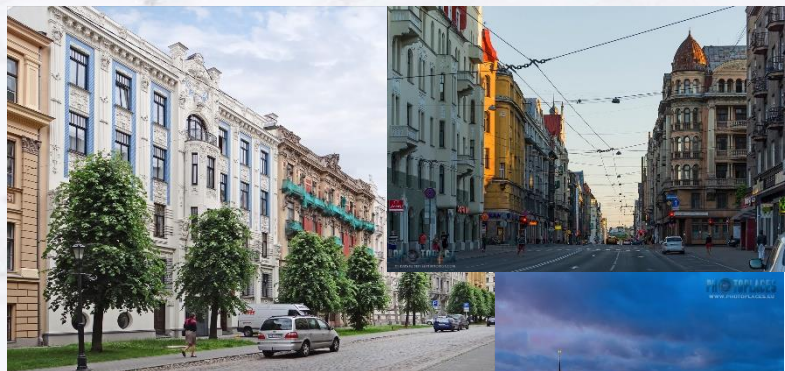


New houses exclusive



Most common areas

City center



Historic heighbourhoods



Soviet built areas



Other smaller regions

Private residences within city limits



Seaside region



Criminal, run down areas



New area



1

What we analyzed

Setting the context and framing
the problem statement

2 What we found

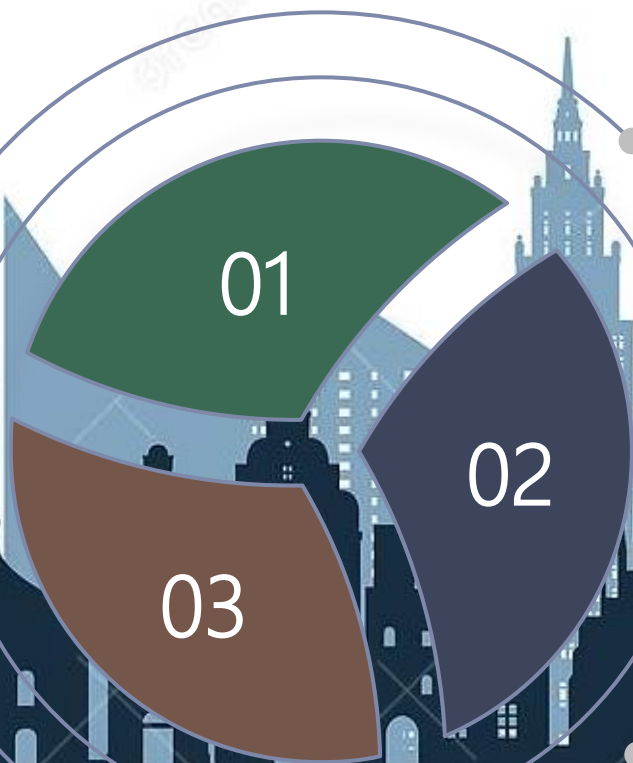
3

How we analyzed

Conclusions and overviews



Key Findings



How large is the apartment?



Is it a new house or not?



Is it in city center or not?

RIGA



Size matters, every 1%↑ in house size increases price by 1.1%



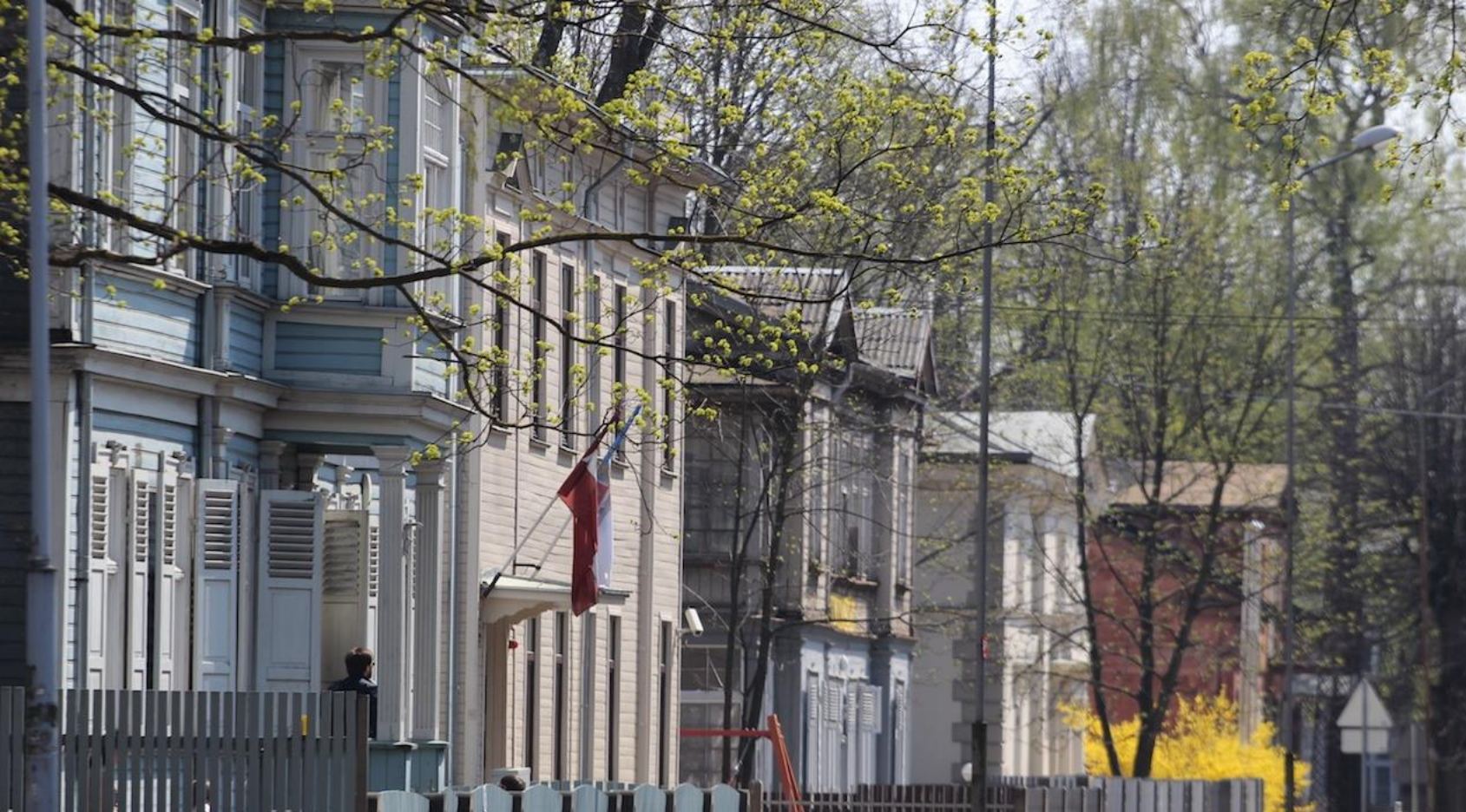
Whilst similar, there is 40% price difference between both house types



Apartments in city center are more expensive than suburbs



Historic center discount 0%



Historic regions discount -20%





Soviet housing regions discount -40%

City limits discount -60%



Criminal & run-down
areas discount -80%





Private residences premium +10%





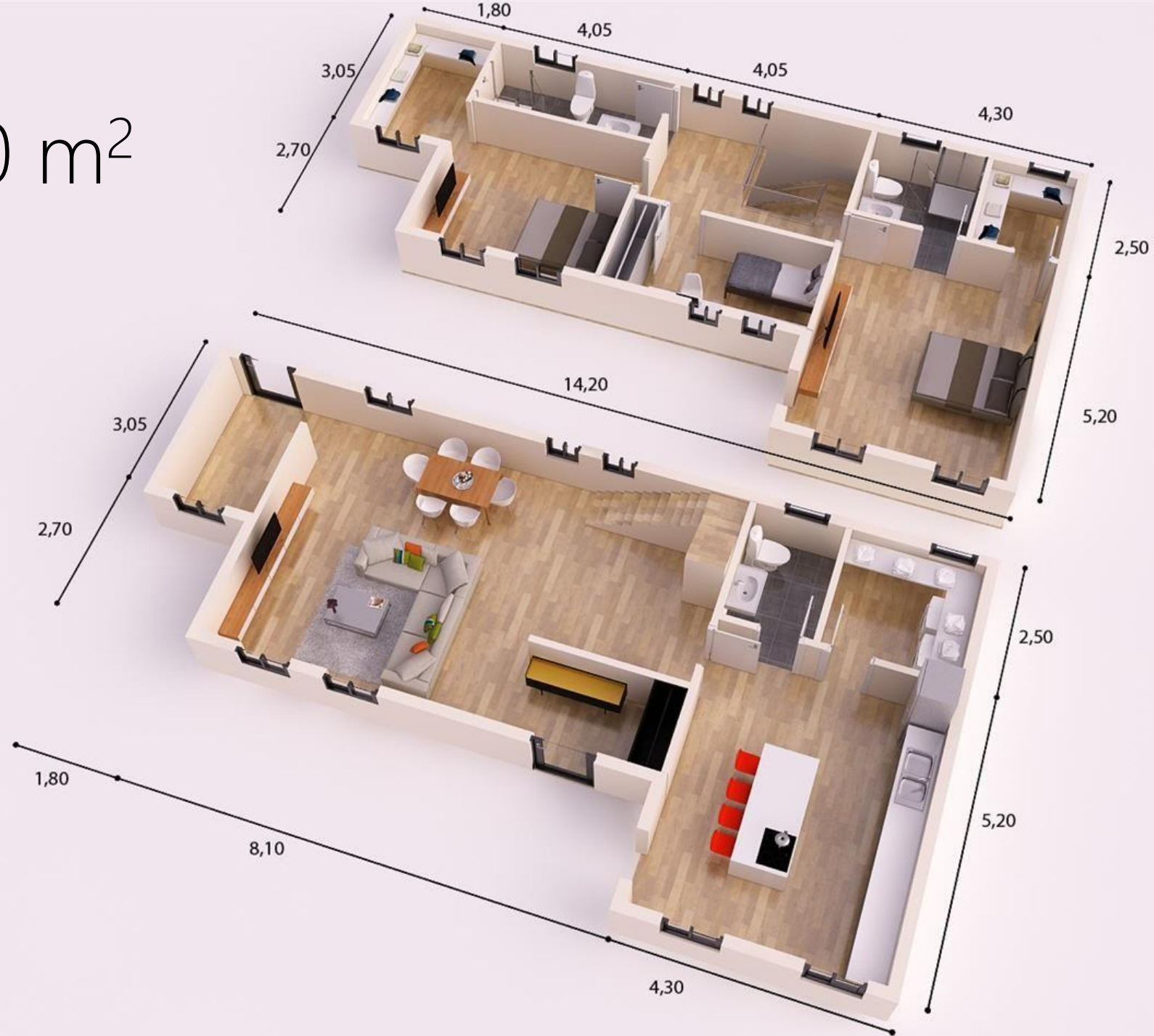
Historic houses in city center were
2x more valuable than other regions

1ST

FLOOR



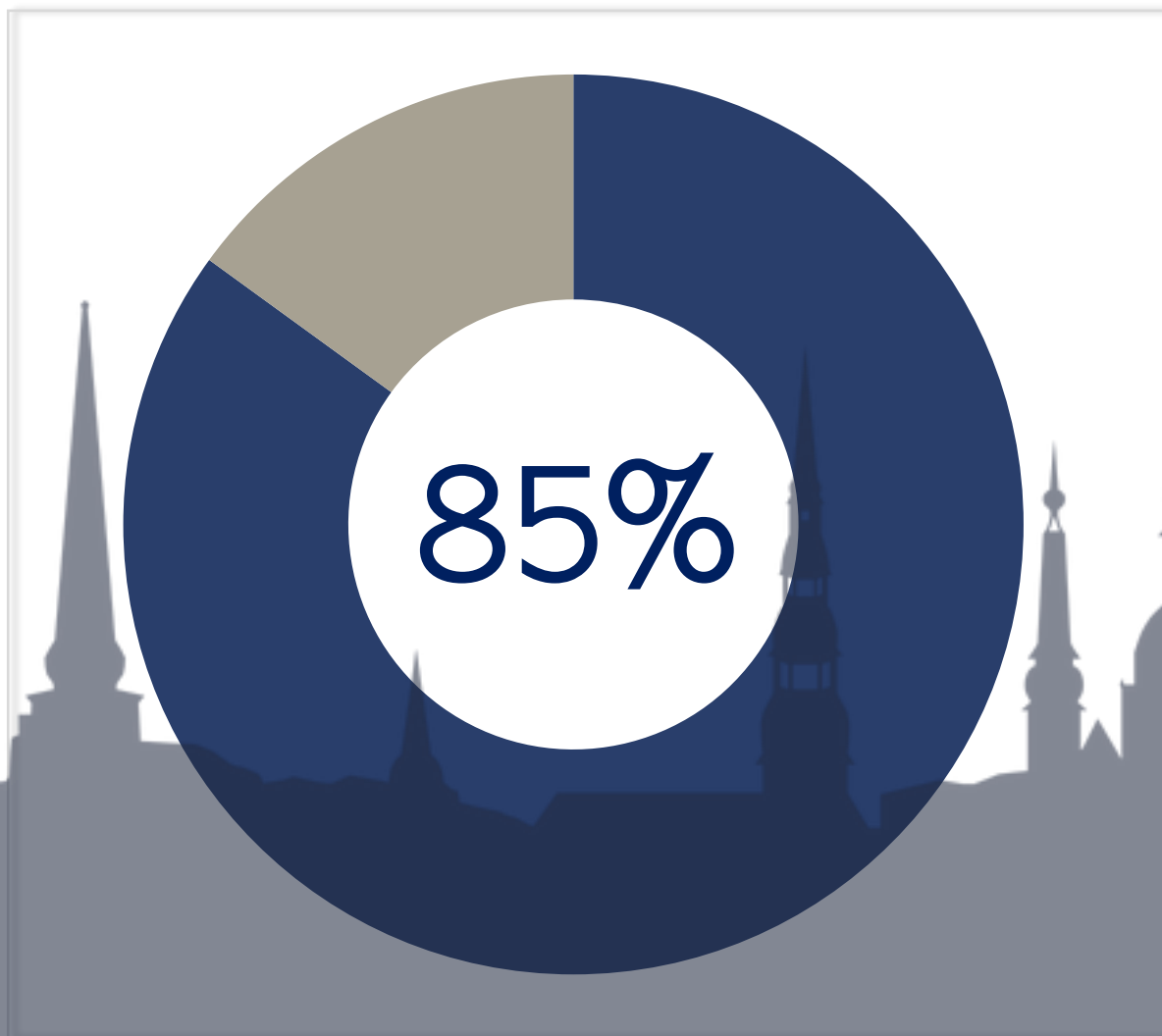
+200 m²







How much we can understand the price?



What to do with rest of it?
Well, you make the call!



1 What we analyzed






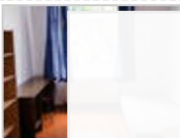
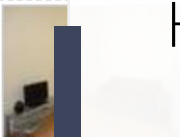
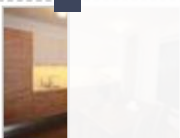

Setting the context | Defining the problem statement


2 What we found

Cleaning of data | Regression analysis

3 How we analyzed



Sludinājumi	<u>datums</u>	Iela	Ist.	m2	Stāvs	Sērija	Cena, m2	
<input type="checkbox"/> 	Izīrē 3 istabu dzīvokli	Bruņinieku 39	3	87	4/4	P. kara	7.47 €	65
<input type="checkbox"/> 	Izīrē, bez	Lāčplēša 100	1	34	1/6	P. kara	8.24 €	28
<input type="checkbox"/> 	Gāzes apkure.	Stabu 29	3	72	4/7	P. kara	2,146 €	15
<input type="checkbox"/> 	Просторная квартира	Matīsa 41	5	101	3/5	P. kara	3.76 €	38
<input type="checkbox"/> 	Деньги в тот же	-	-	-	-	-	-	pē
<input type="checkbox"/> 	Izīrē 3 istabu dzīvokli,	Stabu 102	3	73	2/5	P. kara	4.66 €	34
<input type="checkbox"/> 	Izīrējam kompak...	Dzirnavu 115	3	108	1/6	Jaun.	3,870 €	41
<input type="checkbox"/> 	Izīrē trīsstabu dzīvokli	Grostonas 25	3	90	12/24	Jaun.	8.89 €	80
<input type="checkbox"/> 	Magdelenas Kvartals	Antonijas 17A	3	108	1/6	Jaun.	3,870 €	41



840 000 €

4492 € per m²

Sludinājumi centrs rīga dzīvokli nekust...

Rīga,centrs,Elizabetes 8

Продается квартира в тихом центре Риги, на улице Элизабетес 8. Большая квартира

5 rooms


187 m²

stāvs: 4 no 5

Updated: 29. janvāris

Found: 29. janvāris

DETAILS



290 000 €

2071 € per m²

Sludinājumi centrs rīga dzīvokli nekust...

Rīga,centrs,Antiņģes 8a

Elegants četrp istabu dzīvoklis Rīgas centrā ar mēbelēm un nepieciešamo tehniku, pēc

4 rooms


140 m²

stāvs: 1 no 3

Updated: 1. februāris

Found: 1. februāris

DETAILS



15 500 €

304 € per m²

Sludinājumi ventspils raj dzīvokli neku...

Tārgales 72

Pārdod siltu saulainu dzīvokli Pārventā. 51m2, istabas izolētas, lodžija, liela virtuve, tīra kāpņu


2 rooms

51 m²

stāvs: 5 no 5

Updated: today at 14:24

Found: 24. janvāris



175 000 €

1215 € per m²

Sludinājumi jugla rīga dzīvokli nekusta...

Rīga,Jugla,Ciemaupes 1

Plaša, moderns, labiekārtots dzīvoklis jaunā projekta Juglā, ļoti ērta atrašanās vieta. Blakus

4 rooms

144 m²

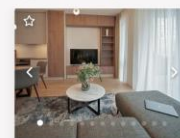
stāvs: 9 no 10

Updated: 1. februāris

Found: 1. februāris

Rīga, Centrs, vecrīga

Kartot pēc: Cenā Cenā/m2 Platības




€700/mēn.

2 ist. 3/5 st.

€15,5/m2

€52 m2

9 Rūpniecības iela 25, 3zeme...




€85 000

2 ist. 3/7 st.

€19,478/m2

€46 m2

9 Aleksandra Čaka iela 33, La...




€418 000

3 ist. 1/6 st.

€3881,2/m2

€107,7 m2

9 Antonijas iela 17a, Central D...




€380/mēn.

5 ist. 3/5 st.

€3,7/m2

€101,4 m2

9 Matīsa iela 41, Latgale Subu...




€520/mēn.

3 ist. 4/5 st.

€6,9/m2

€75 m2

9 Elizabetes iela 4, Centra raj...



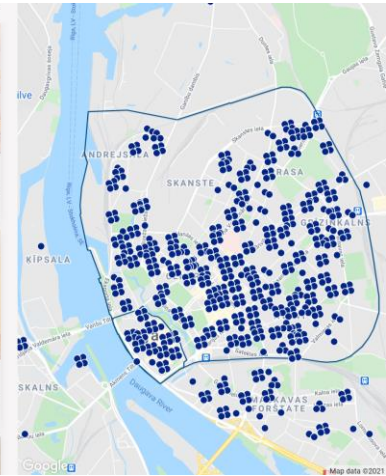
€200/mēn.

1 ist. 3/5 st.

€73/m2

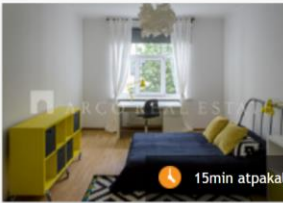
€28 m2

9 laboratorijas iela 11



Dzīvoklis, Pārdod

ARCO REAL ESTATE, Māris Paeglis



Marijas iela 16, Centrs

Rīga

350 EUR (mēn)

2,87 EUR/m²

122 m²

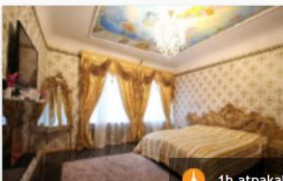
4 Istabas

5/6 Stāvs

15min atpakaļ

Dzīvoklis, Izīrē

ARCO REAL ESTATE, Daiga Bondare



Dzirnavu iela 115, Centrs

Rīga

700 EUR (mēn)

100 EUR 800-EUR-(mēn)

6,09 EUR/m²

115 m²

3 Istabas

2/5 Stāvs

1h atpakaļ

How to make sense of all this?

Data Analytics knowledge can do wonders!

edureka!



Web Scrapping with Python



From Web

Best-rated

Coming Soon

Game Demos

Game Previews

Most Played

New

Top free

Showing 1 - 46 of 46 results



UNO®

★★★★★

\$9.99+



**MONOPOLY
PLUS**

★★★★★

\$14.99+



**ARCADE GAME
SERIES: PAC-
MAN**

★★★★★

\$3.99



Rare Replay

★★★★★

\$29.99+

[Site feedback](#)

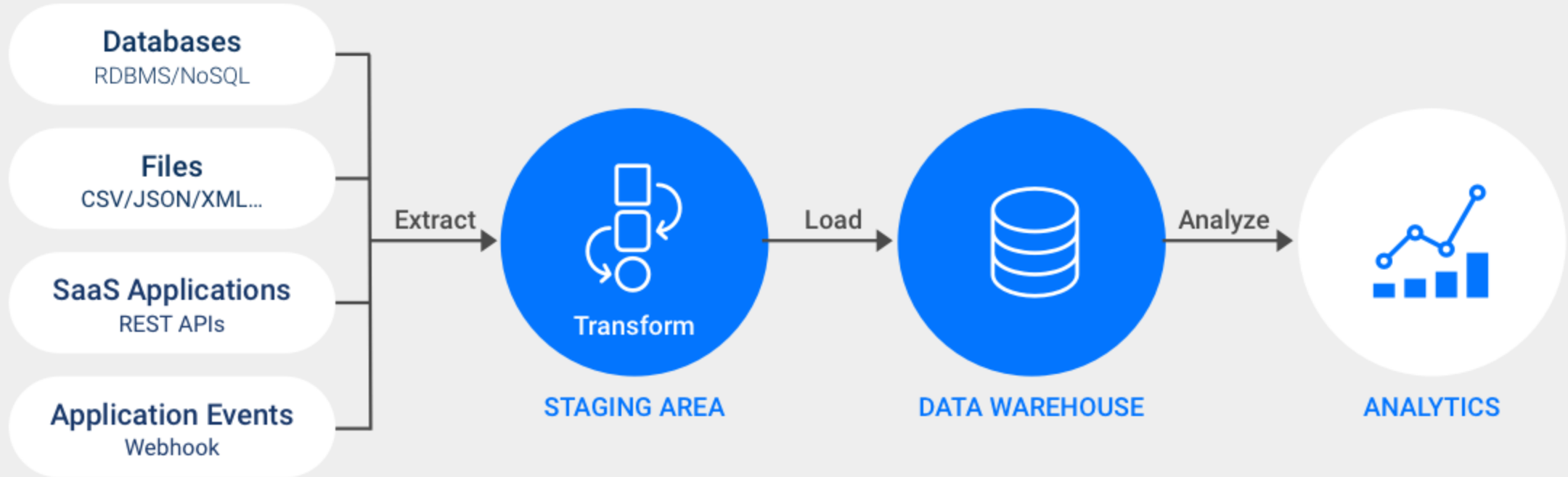
Questions? Talk to an expert

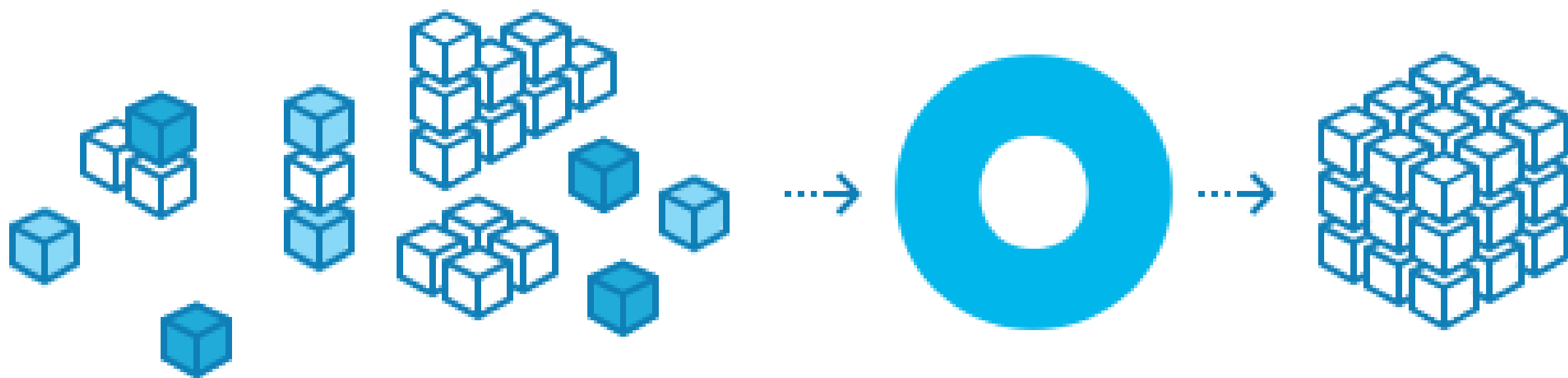
	Column1	Column2	*
1	UNO®	\$9.99	
2	Monopoly Plus	\$14.99	
3	ARCADE GAME...	\$3.99	
4	Rare Replay	\$29.99	
5	MONOPOLY F...	\$19.99	
6	ARCADE GAME...	\$3.99	
7	ARCADE GAME...	\$3.99	
8	Chess Ultra	\$12.49	
9	ARCADE GAME...	\$7.99	
10	Jeopardy!	\$19.99	
11	Babylon 2055 P...	\$4.99	
12	Zombie Party	\$9.99	
13	ARCADE GAME...	\$3.99	
14	The Disney Aft...	\$10.99	

OK

Cancel

ETL PROCESS





Source data | Cleaning and identification of variables

We web extracted data from a largest listings website's real estate section [ss.com/lv/real-estate/flats/riga/].

For the purposes of this analysis, we simplified the data as follows:

- Considered only those listings that are "for sale"
- Re-categorized 'Region' to limit the number of unique/distinct regions to 8.
- Re-categorized 'House_type' to limit the number of unique house types to 6.

Street	Region	Street_name	Rooms	m2	Value_EUR	Price_per_m2	Sale_type	House_type
Valnu 4	City center	Valnu	5	170	790500	4,650	For Sale	New
Vilandes 8	City center	Vilandes	3	164	647010	3,945	For Sale	Neo Classical
Ausekla 6	City center	Ausekla	6	232	638000	2,750	For Sale	Neo Classical
Krisjana Valdemara 23	City center	Krisjana	5	164	620000	3,780	For Sale	Classical

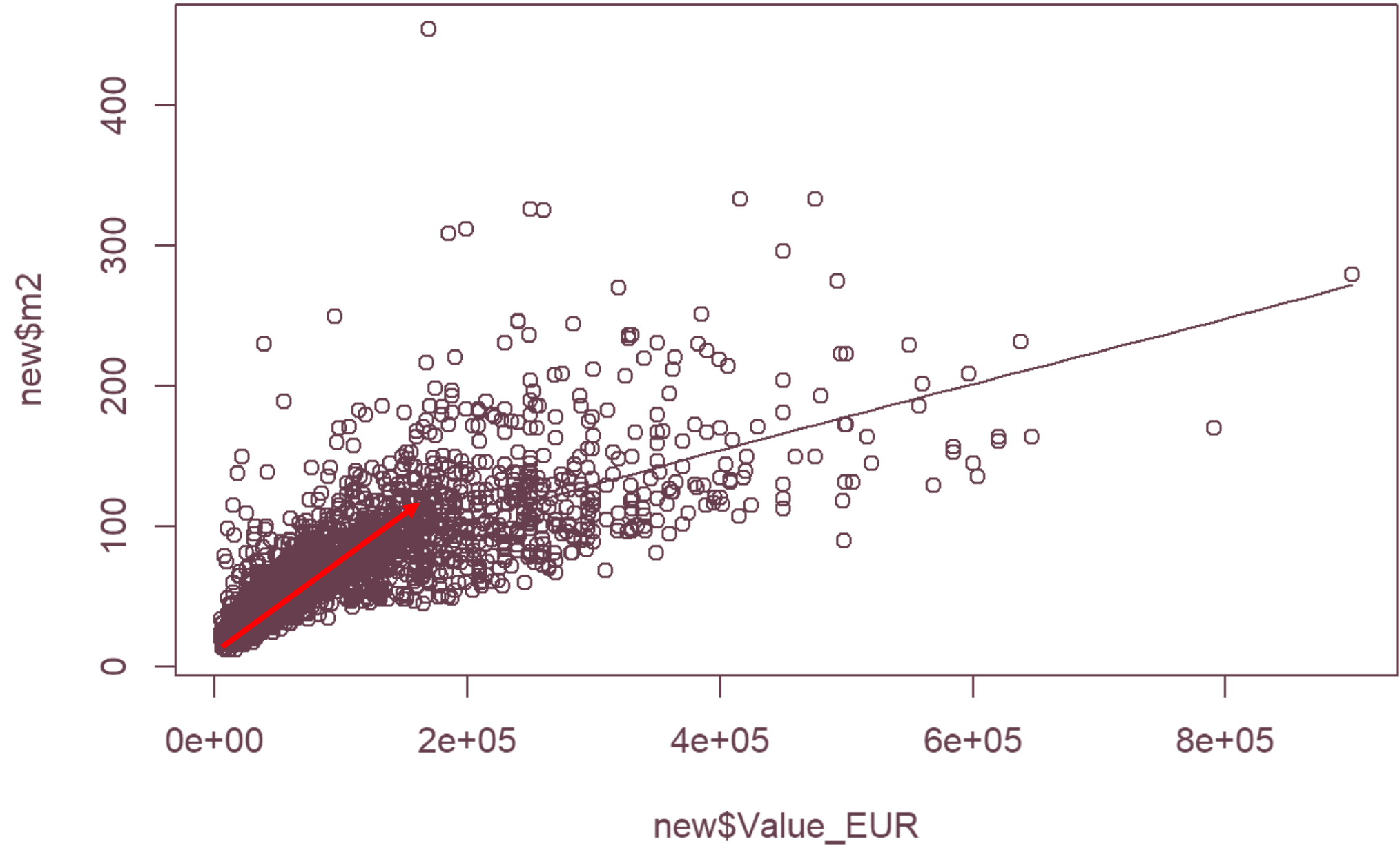
We also recognized that 'Street' and 'Street_name', would be drilling down even further into the region and hence, were not relevant. In addition, we did not expect that few such entries would yield statistically significant result.

Additionally, 'Price_per_m2' would be highly correlated to Property price (i.e. 'Value_EUR') and hence was removed from analysis.

Hence, for the final analysis we retained the dependent variable 'Value_EUR' and the independent variables - 'Region', 'Rooms', 'm2', 'House_type'.



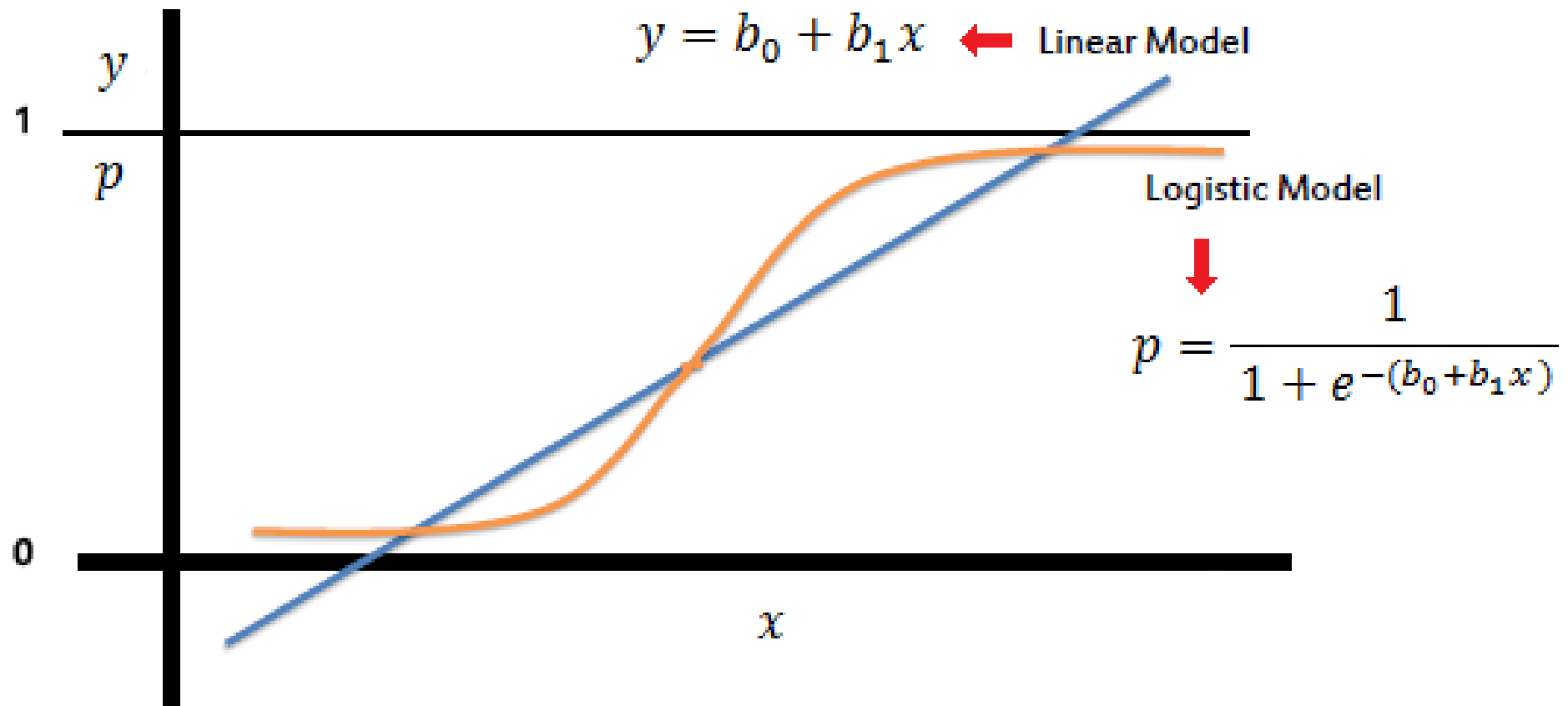
Value to m2



CHECKLIST



$$\text{Ln}\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$



Call:

```
## lm(formula = Value_EUR ~ ., data = train)
```

##

Residuals:

```
##      Min       1Q   Median       3Q      Max
## -433052  -20269   -1388    14468  498864
```

##

Coefficients: (2 not defined because of singularities)

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      -44727.33      8941.15  -5.002 6.03e-07 Reg_Katlakalns      2879.07      27819.48  -0.103 0.917581
```

```
## Rooms              3828.33      1410.26   2.710 0.007581 Reg_Katlakalns      5575.46      9149.23   0.609 0.542315
```

```
## m2              1290.13      39.113   32.984 < 2e-16 *** Reg_Katlakalns     208475.96     16782.35  12.422 < 2e-16 ***
```

```
## Floor              245.43      43.113   5.693 1.13e-06 *** Reg_Katlakalns     81260.83     16656.65   4.879 1.13e-06 ***
```

```
## House_type_Classical      -924.13      48.113  -19.208 0.539809 Reg_Katlakalns     8794.66     14342.76   0.613 0.539809
```

```
## `House_type_Czech type`    -3438.63      48.113  -71.471 0.932944 Reg_Katlakalns     1589.62     18890.62   0.084 0.932944
```

```
## `House_type_Lithuania type` 2121.13      48.113   44.084 -0.780 0.435348 Reg_Katlakalns    -15813.43     20268.81  -0.780 0.435348
```

```
## `House_type_Neo Classical` 53750.99      48.113  1117.41 1.112 0.266325 Reg_Katlakalns     229.51     10999.84   1.112 0.266325
```

```
## House_type_New      34237.96      48.113  710.343 3.616 0.000304 *** Reg_Katlakalns     953.14     11324.63   3.616 0.000304 ***
```

```
## House_type_Other      -154.13      48.113  -3.202 0.819460 Reg_Katlakalns     366.47     10367.35   0.228 0.819460
```

```
## `House_type_Pre Soviet era` -154.13      48.113  -3.202 0.819460 Reg_Katlakalns     366.47     10367.35   0.228 0.819460
```

```
## `House_type_Private house` -896.13      48.113  -18.625 NA NA NA NA
```

```
## `House_type_Series 103`     4137.13      48.113   85.991 0.283 0.776993 Reg_Katlakalns     2595.11     9161.34   0.283 0.776993
```

```
## `House_type_Series 104`    -5155.48      48.113  -107.14 0.486 0.626864 Reg_Katlakalns     4036.73     8302.61   0.486 0.626864
```

```
## `House_type_Series 119`    -4038.72      48.113  -83.944 7.188 8.44e-13 *** Reg_Katlakalns     58680.62     8164.06   7.188 8.44e-13 ***
```

```
## `House_type_Series 467`    -6696.13      48.113  -139.18 1.215 0.224603 Reg_Katlakalns     15845.60     13045.40   1.215 0.224603
```

```
## `House_type_Series 602`    -1606.13      48.113  -33.381 0.254 0.799653 Reg_Katlakalns     2543.90     10022.45   0.254 0.799653
```

```
## `House_type_Soviet era`     18710.13      48.113  388.811 -0.435 0.663580 Reg_Katlakalns     16727.60     38452.29  -0.435 0.663580
```

```
## Reg_agenskalns      18710.13      48.113  388.811 1.739 0.082062 . Reg_Katlakalns     17032.06     9791.50   1.739 0.082062 .
```

```
## Reg_Aplokciems      1116.33      48.113   23.202 0.704 0.481309 Reg_Katlakalns     8550.30     12140.24   0.704 0.481309
```

```
## Reg_Bergi          -8418.22      48.113  -174.97 -0.174 0.861853 Reg_Katlakalns    -2618.14     15044.00  -0.174 0.861853
```

```
## Reg_Biekensala     -46056.49      48.113  -957.30 0.122 0.903119 Reg_Katlakalns     3394.01     27880.56   0.122 0.903119
```

```
## Reg_Bierini         -8492.74      48.113  -176.51 0.100 0.920305 Reg_Katlakalns      999.78     9991.84   0.100 0.920305
```

```
## Reg_Bolderaja      -3991.70      48.113  -82.966 -0.028 0.977776 Reg_Katlakalns    -1074.48     38567.68  -0.028 0.977776
```

```
## Reg_Bukulti        -42244.39      48.113  -877.98 0.427 0.669671 Reg_Katlakalns     23032.94     53986.16   0.427 0.669671
```

```
## Reg_ciekurkalns      9854.93      48.113  204.838 0.181 0.856674 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Darzciems       2530.89      48.113   52.809 -0.001 0.999999 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Daugavgriva     -4808.51      48.113  -99.946 -0.001 0.999999 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Dreilini       -20982.21      48.113  -436.13 -0.389 0.699999 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Dzeguzkalns      5993.45      48.113  124.574 0.445 0.656656 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Ilguociems      5180.10      48.113  107.650 0.533 0.594355 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Imanta         8087.55      48.113  168.100 0.891 0.373184 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

```
## Reg_Jaunciems       337.44      48.113    7.014 0.019 0.985032 Reg_Katlakalns     1623.89     8990.33   0.181 0.856674
```

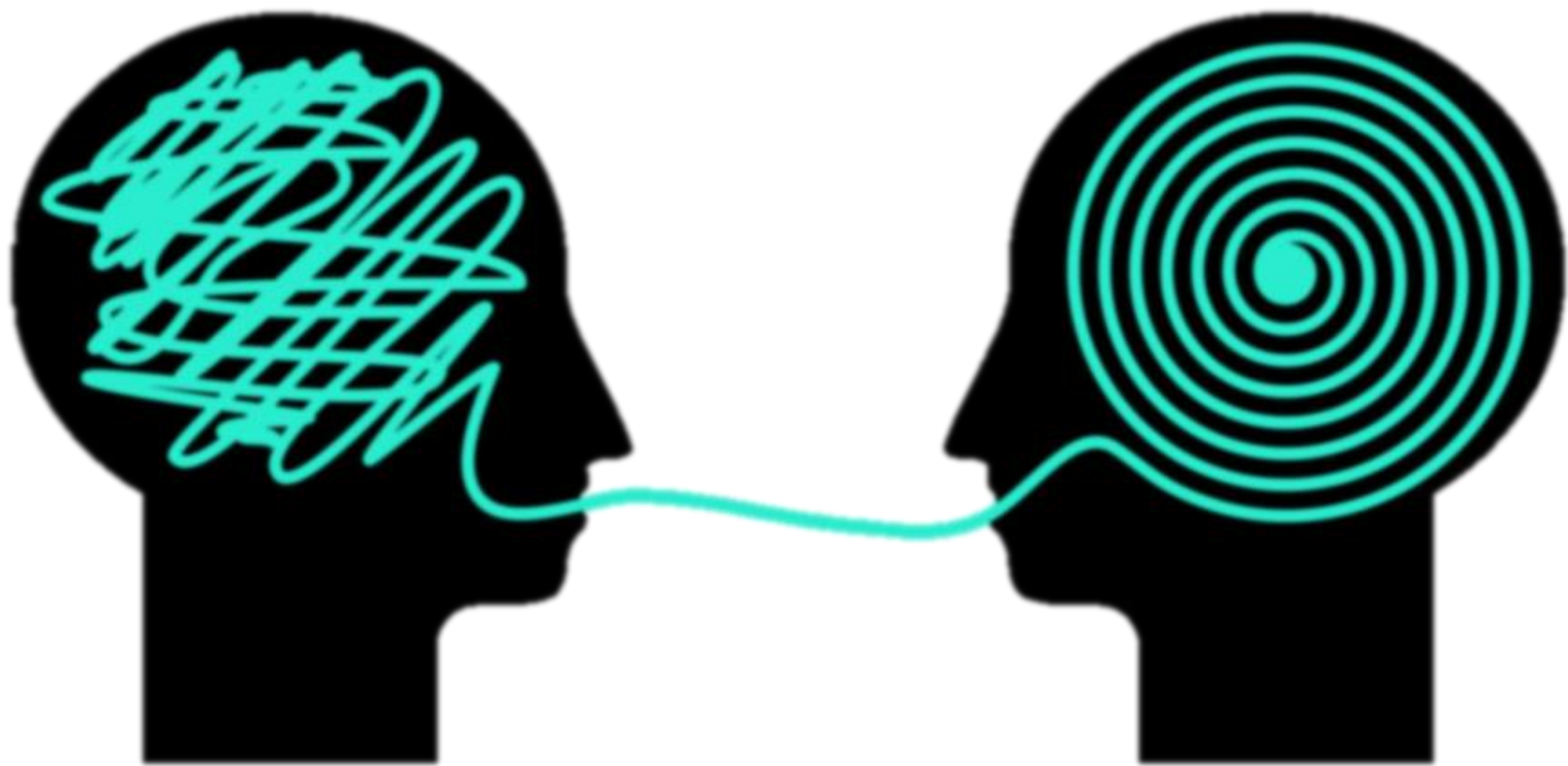
Your head has most probably exploded by this point



Residual standard error: 53280 on 2771 degrees of freedom

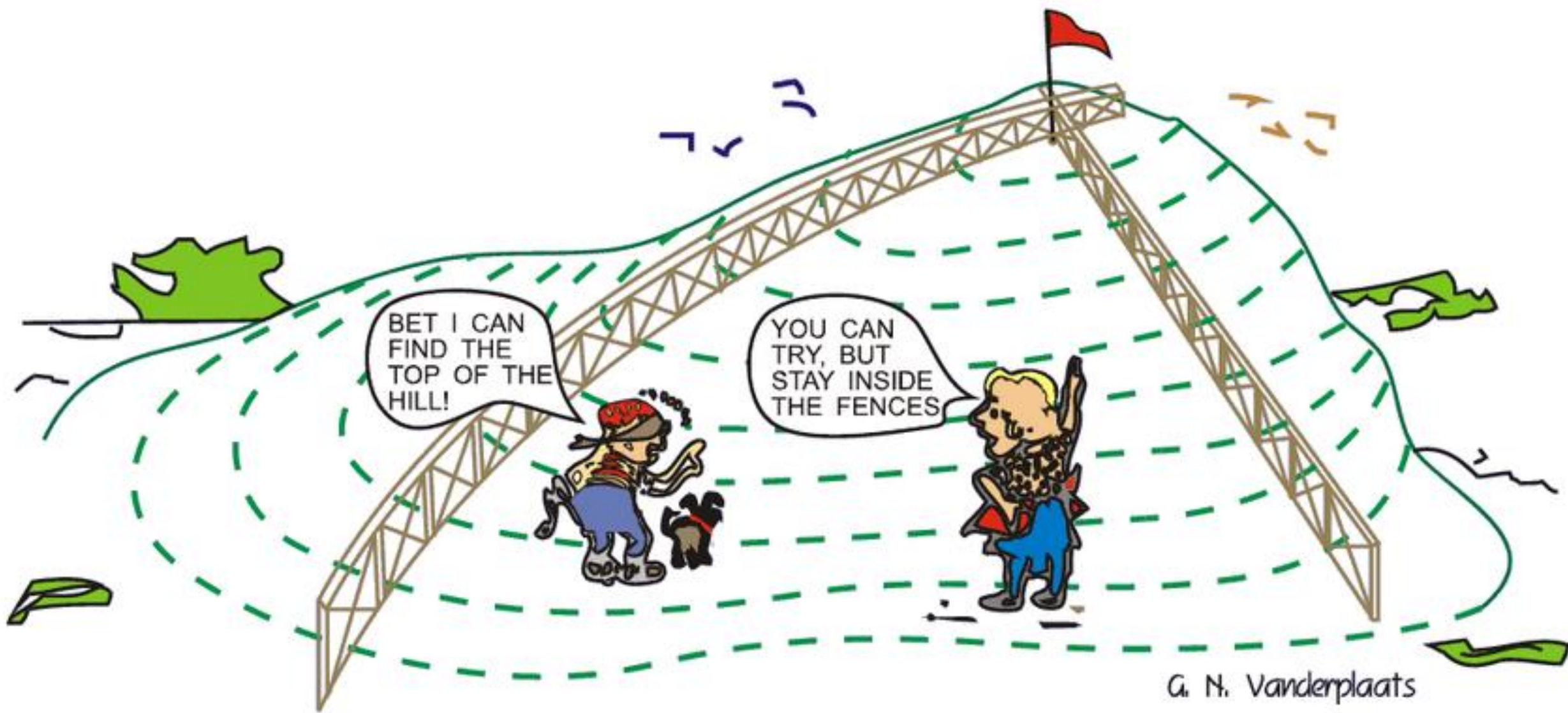
Multiple R-squared: 0.7035, Adjusted R-squared: 0.6974

F-statistic: 115.4 on 57 and 2771 DF, p-value: < 2.2e-16



Machine Learning

The text "Machine Learning" is centered within a cloud-like shape. The cloud is outlined with a thick dark blue line and a thinner light blue line. Inside the cloud, there is a network of light blue circuit lines with small dots at the intersections, resembling a neural network or a computer circuit. The lines are more dense on the right side of the cloud.



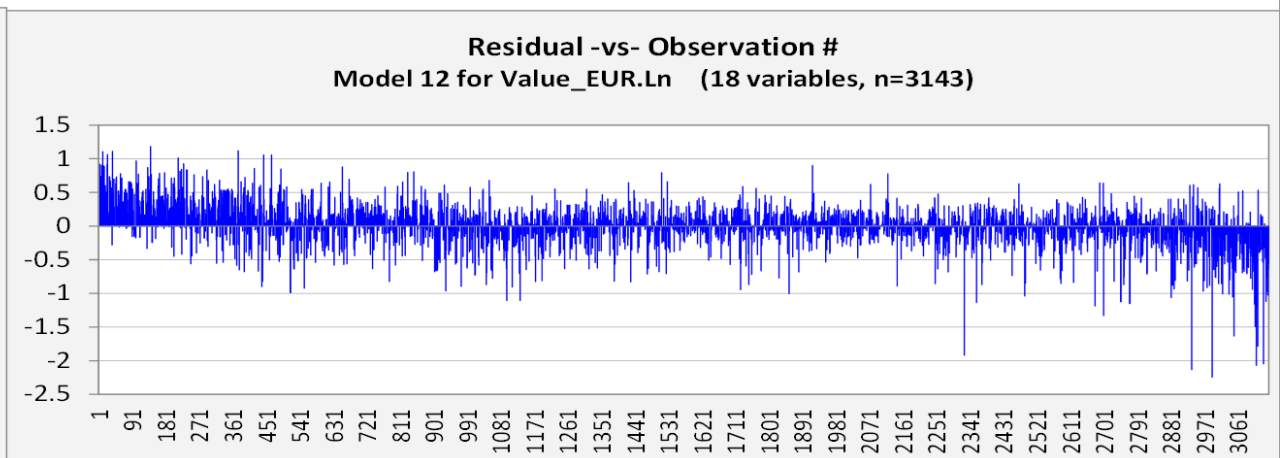
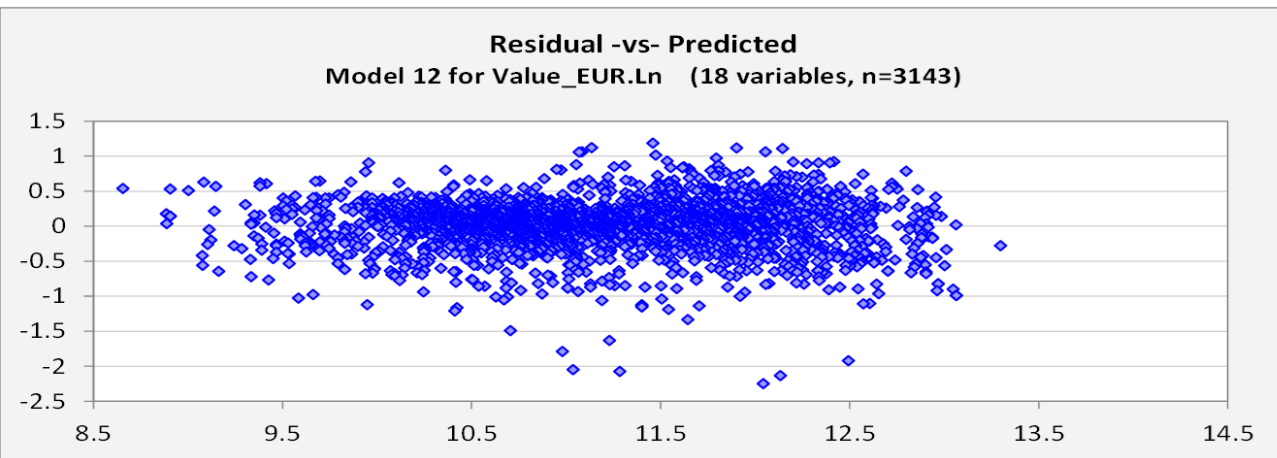
Final regression model

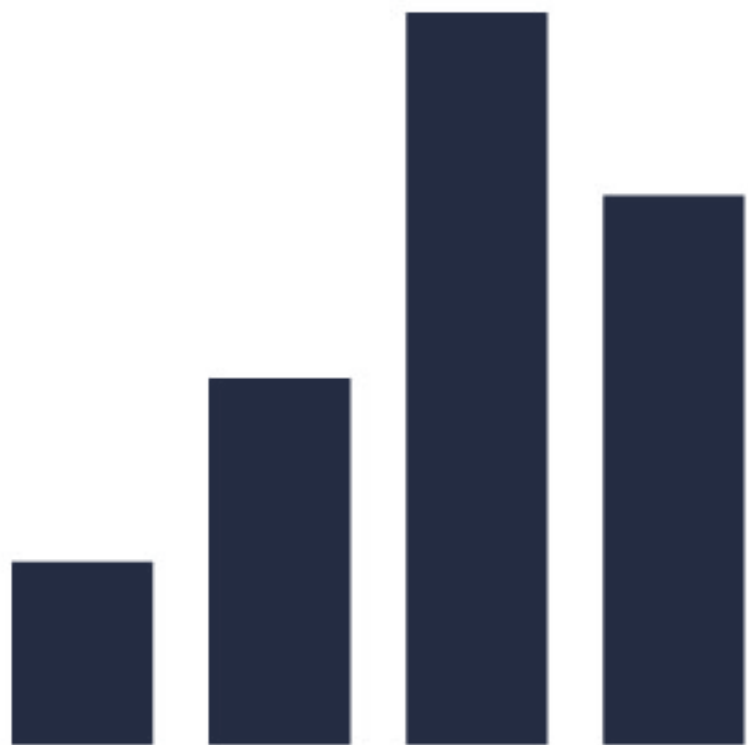
Dependent Variable:	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	Critical t	Confidence
Value_EUR.Ln	0.852	0.852	0.328	0.852	3143	1	1.961	95.0%

Variable	Coefficient	Std.Err.	t-Statistic	P-value	Lower95%	Upper95%	VIF	Std. Coeff.	Explanation
Constant	6.763	0.060	112.381	0.000	6.645	6.881	0.000	0.000	
Floor	0.007925	0.002825	2.805	0.005	0.002385	0.013	1.449	0.023	
Floor_1	-0.161	0.017	-9.313	0.000	-0.194	-0.127	1.389	-0.075	First floor discounted, nobody likes that people from outside can see through your windows
House_type.Eq.Classical	-0.224	0.042	-5.363	0.000	-0.306	-0.142	9.684	-0.115	Classical houses are cheaper overall because they are about 100 years old and not renovated
House_type.Eq.Neo_Classical	0.259	0.047	5.467	0.000	0.166	0.353	5.211	0.086	Renovated classical houses are in high regard as they usually require heavy investments
House_type.Eq.New	0.388	0.018	21.412	0.000	0.352	0.423	1.573	0.185	New houses command much higher price
House_type.Eq.Other	-0.064	0.020	-3.173	0.002	-0.103	-0.024	1.245	-0.024	
m2_Ln	1.097	0.014	79.943	0.000	1.070	1.123	1.530	0.680	m2 has the highest predictive power
m2_200	-0.294	0.050	-5.904	0.000	-0.392	-0.196	1.133	-0.043	large m2 apartments are usually entire floors or rooftops that are in lower value than "actual" apart
Reg.Eq.City_center.Times.House_type.Eq.Class	0.253	0.047	5.371	0.000	0.161	0.345	10.324	0.119	Classical houses in the city center however are much more valued as they are made from bricks v
Reg.Eq.City_center.Times.House_type.Eq.Neo_C	0.221	0.055	4.007	0.000	0.113	0.329	5.779	0.066	This could be greyed out as has similar finding as in the Neo_Classical case (most of renovated h
Reg.Eq.City_outskirts	-0.583	0.087	-6.703	0.000	-0.754	-0.413	1.049	-0.047	Further driving time to city center reduced the value
Reg.Eq.Criminal__run_down_area	-0.839	0.057	-14.691	0.000	-0.951	-0.727	1.194	-0.110	Criminal areas are most potent in negative pricing development
Reg.Eq.historic_neighbourhood	-0.214	0.028	-7.555	0.000	-0.269	-0.158	1.992	-0.073	Historic neighbourhood has the lowest price discount compared to city center among other region
Reg.Eq.New_area	-0.425	0.064	-6.627	0.000	-0.551	-0.300	1.099	-0.048	This is interesting and hard to explain, maybe because those new areas are located far from the c
Reg.Eq.Private_residences_within_city_limits	0.105	0.043	2.438	0.015	0.021	0.189	1.225	0.019	Private houses are in higher value, small number of them make low impact than it could potentially
Reg.Eq.Seaside_region	-0.521	0.047	-11.163	0.000	-0.612	-0.429	1.188	-0.084	This is interesting finding, most likely due to considered "cheap" seaside area (industrial area) cor
Reg.Eq.Soviet_built_region	-0.366	0.021	-17.207	0.000	-0.407	-0.324	3.277	-0.214	Soviet built region typically has social houses that are in low regarded value
Reg.Eq.Soviet_built_region.Times.House_type.E	-0.126	0.054	-2.319	0.020	-0.232	-0.019	2.375	-0.025	

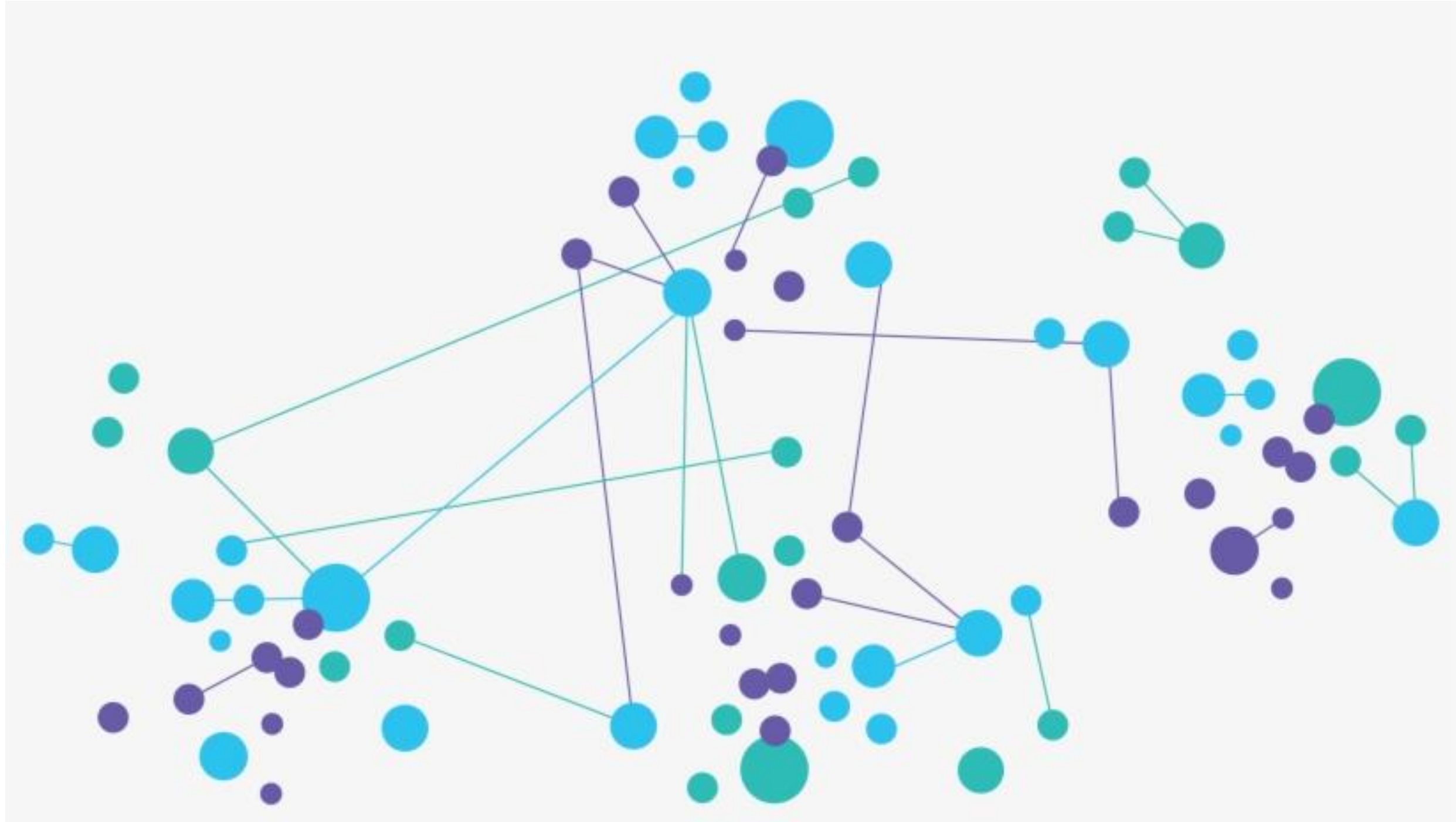
Source	Deg. Freedom	Sum Squares	Mean Square	F-Statistic	P-value
Regression	18	1 944	107.994	1002.028	0.000
Residual	3124	336.692	0.108		
Total	3142	2 281			

	Mean Error	RMSE	MAE	Minimum	Maximum	MAPE	J-B stat
Fitted (n=3143)	0.000	0.327	0.236	-2.249	1.183	2.1%	:371.97 (P=0.000)















SSE RIGA