Using Data Science to Find an Apartment in Amsterdam

Rafael V. Pierre
Github: rafaelpierre
Medium: @rafaelpierre

Welcome & Intro

Rafael V. Pierre

- Brazilian/Italian
- Track record in **Software Engineering** within **Stock Market, High Frequency Trading (HFT), Private Banking, Foreign Exchange, Credit Risk**
- **Data Scientist** @ **Weet Analytics**

Before we start...

- conda install -c pyviz/label/dev pyviz
- pip install ipywidgets
- jupyter nbextension enable --py widgetsnbextension
- pip install folium
- git clone https://github.com/rafaelpierre/JupyterAMS.git

Agenda

- *How it all started*
- *Why Jupyter?*
- *PyViz*
- *IPython Widgets*
- *Let's Practice!*

How it all started

Let’s go!

#10yearchallenge

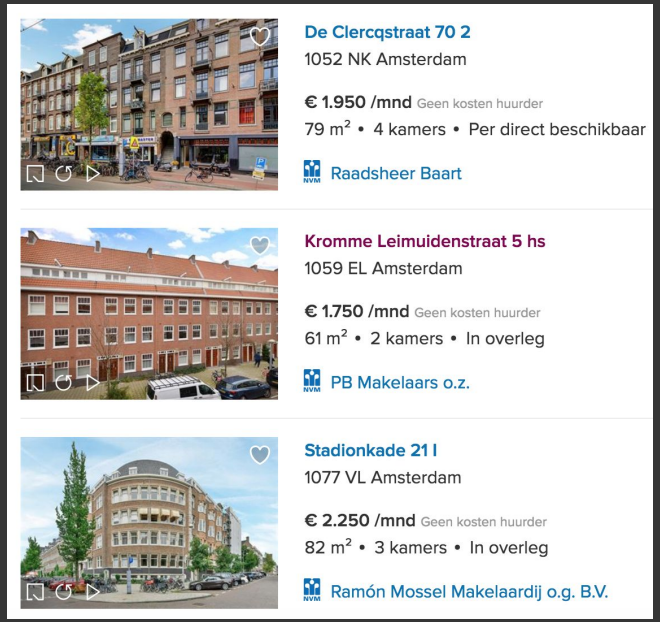
#6yearchallenge

#6yearchallenge

Where do we start?

The classic way...

...is not the easy way



- *Are these expensive?*
- *Are these good areas?*
- *Can I find better deals?*
- *Kromme Leimuidenstraat!?*

Really...

Hi,

Thank you for the interest shown in our offer.
We request that you fill in the below “get to know you questions” and send this per email to [redacted]

-

- What is your current occupation?
- What is your study background?
- What is your age?
- What is your family situation?
- What is your nationality?
- What is your income (net)?
- What is your contract situation?
- What is the reason for looking a rental property?
- What is your current adress?
- Do you have any pets?
- Do you smoke?
- When would you like to rent?
- For how long do you intent to rent?
- Do you give permission to share the above information, if applicable, with the landlord?
- What is your phone number?

Guess what... there's a solution!

Fundamental Analysis

Fundamental analysis is a method of evaluating a security in an attempt to assess its intrinsic value, by examining related economic, financial, and other qualitative and quantitative factors.

- Return on Equity
- Price to Earnings
- Price to Equity
- Compare with other companies within the sector
- Find out if asset is *overvalued* or *undervalued*

Fundamental Analysis Applied to Real Estate

Intrinsic and extrinsic parameters

- Apartment size
- Number of kamers
- Neighborhood
- Price per square meter
- Price comparison with other apartments in the neighborhood

The smart way

So... where do we start?

So... where do we start?

Ask questions!

Ask Questions...

- How prices are distributed?
- Which are the hottest areas?
- Where to start hunting?
- How to find a good deal?

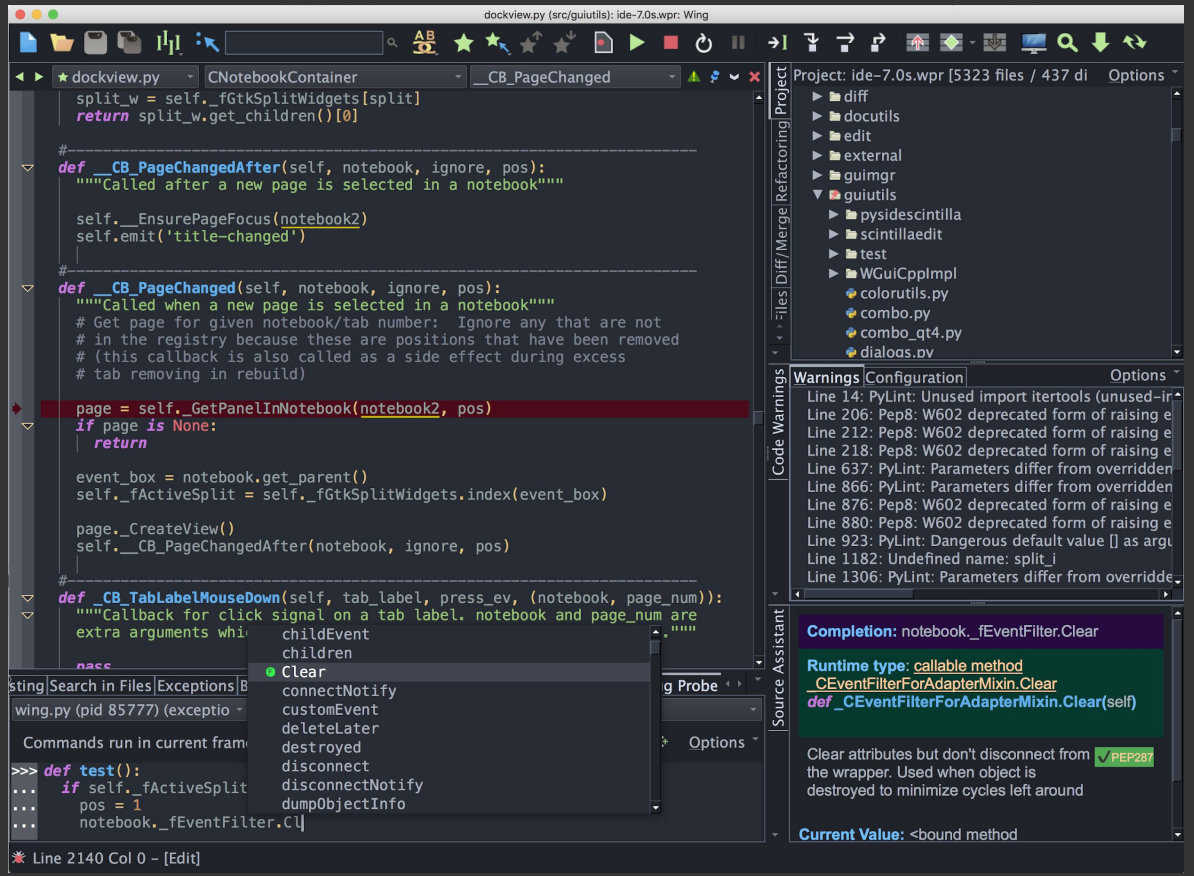
Iterative Pipeline

Which tools?

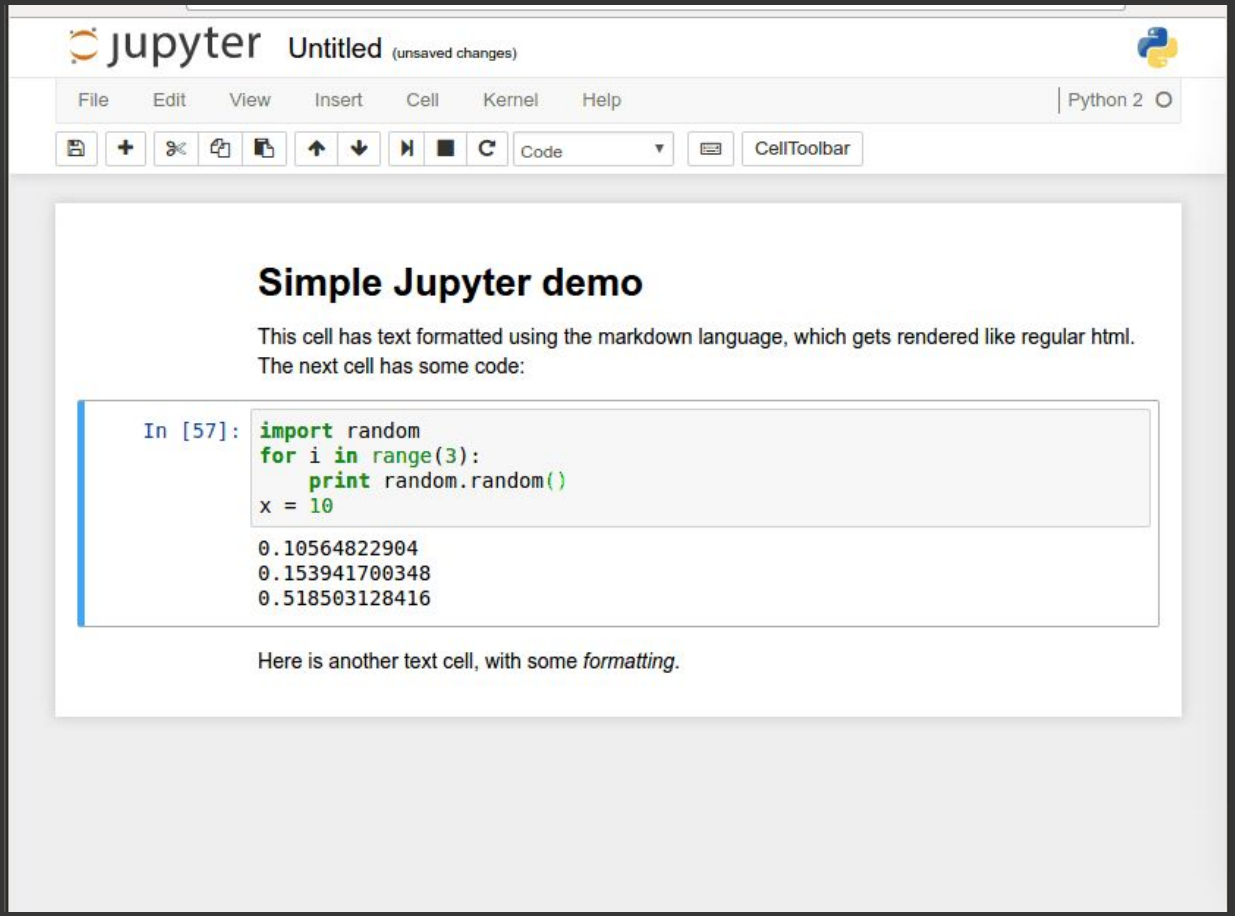
jupyter
Lua
php
ERLANG
C++
Spark
C#
F#

Why Jupyter?

This...


dockview.py (src/guiutils): ide-7.0s.wpr: Wing
dockview.py
CNotebookContainer
__CB_PageChanged
split_w = self._fGtkSplitWidgets[split]
return split_w.get_children()[0]
def __CB_PageChangedAfter(self, notebook, ignore, pos):
"""Called after a new page is selected in a notebook"""
self.__EnsurePageFocus(notebook2)
self.emit('title-changed')
def __CB_PageChanged(self, notebook, ignore, pos):
"""Called when a new page is selected in a notebook"""
# Get page for given notebook/tab number: Ignore any that are not
# in the registry because these are positions that have been removed
# (this callback is also called as a side effect during excess
# tab removing in rebuild)
page = self._GetPanelInNotebook(notebook2, pos)
if page is None:
return
event_box = notebook.get_parent()
self._fActiveSplit = self._fGtkSplitWidgets.index(event_box)
page._CreateView()
self.__CB_PageChangedAfter(notebook, ignore, pos)
def _CB_TabLabelMouseDown(self, tab_label, press_ev, (notebook, page_num)):
"""Callback for click signal on a tab label. notebook and page_num are
extra arguments whi
childEvent
children
Clear
connectNotify
customEvent
deleteLater
destroyed
disconnect
disconnectNotify
dumpObjectInfo
Search in Files
Exceptions
wing.py (pid 85777) (exceptio
Commands run in current fram
Options
>>> def test():
if self._fActiveSplit
pos = 1
notebook._fEventFilter.Cl
Line 2140 Col 0 - [Edit]
Project: ide-7.0s.wpr [5323 files / 437 di
Options
diff
docutils
edit
external
guimgr
guiutils
pysidescintilla
scintillaedit
test
WGuiCppImpl
colorutils.py
combo.py
combo_qt4.py
dialogs.py
Warnings
Configuration
Options
Line 14: PyLint: Unused import itertools (unused-im
Line 206: Pep8: W602 deprecated form of raising e
Line 212: Pep8: W602 deprecated form of raising e
Line 218: Pep8: W602 deprecated form of raising e
Line 637: PyLint: Parameters differ from overridden
Line 866: PyLint: Parameters differ from overridden
Line 876: Pep8: W602 deprecated form of raising e
Line 880: Pep8: W602 deprecated form of raising e
Line 923: PyLint: Dangerous default value [] as argu
Line 1182: Undefined name: split_i
Line 1306: PyLint: Parameters differ from overridde
Completion: notebook._fEventFilter.Clear
Runtime type: callable method
_CEventFilterForAdapterMixin.Clear
def _CEventFilterForAdapterMixin.Clear(self)
Clear attributes but don't disconnect from the wrapper. Used when object is destroyed to minimize cycles left around
PEP287
Current Value: <bound method
Project
Refactoring
Merge
Diff
Files
Code Warnings
Source Assistant

Versus this...


jupyter Untitled (unsaved changes)
File Edit View Insert Cell Kernel Help
Python 2
Code
CellToolbar
Simple Jupyter demo
This cell has text formatted using the markdown language, which gets rendered like regular html.
The next cell has some code:
In [57]:
import random
for i in range(3):
print random.random()
x = 10
0.10564822904
0.153941700348
0.518503128416
Here is another text cell, with some formatting.

Jupyter

- Literate Programming (Knuth)
- Interactive Paradigm
- Fast feedback makes it great for **bootstrapping**, **experimenting**, **prototyping ideas** and **visualizing data**

Some bonuses...

PyViz

- Makes data visualization easier, powerful and interactive
- Core libraries: **Panel, hvPlot, HoloViews** and **GeoViews**
- Support for **Bokeh, matplotlib** and **Datashader**

iPython Widgets

- Makes Jupyter even more interactive
- More control over data
- Support for **sliders, toggle buttons, progress bars...**

Our Pipeline

Obtaining Apartment Data

Data on the web comes in different flavors...

- Static HTML
- JSON
- APIs
- Dynamic HTML

Obtaining Apartment Data

For different flavors, different tools

- Requests + Lib2XML2 + JSON
- BeautifulSoup
- Scrapy
- Selenium

UX is your friend!

UX is your friend!

How to generate insights?

Which one is best?

This...

De Clercqstraat 70 2
1052 NK Amsterdam

€ 1.950 /mnd Geen kosten huurder
79 m² • 4 kamers • Per direct beschikbaar

Raadsheer Baart

Kromme Leimuidenstraat 5 hs
1059 EL Amsterdam

€ 1.750 /mnd Geen kosten huurder
61 m² • 2 kamers • In overleg

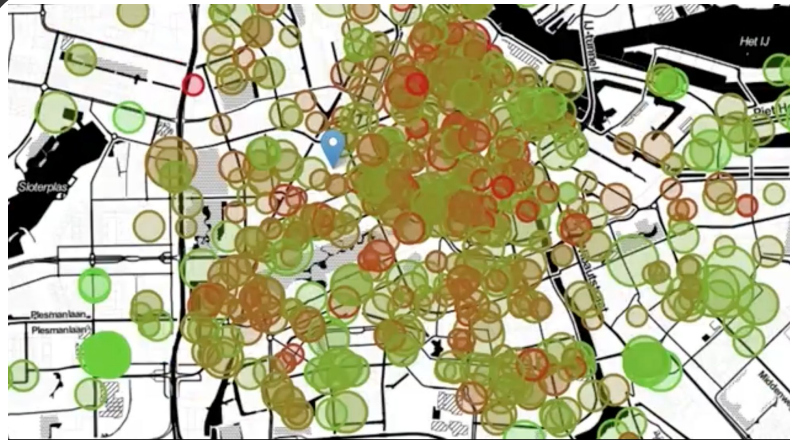
PB Makelaars o.z.

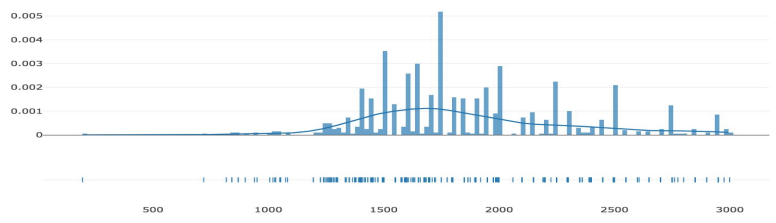
Stadionkade 21 I
1077 VL Amsterdam

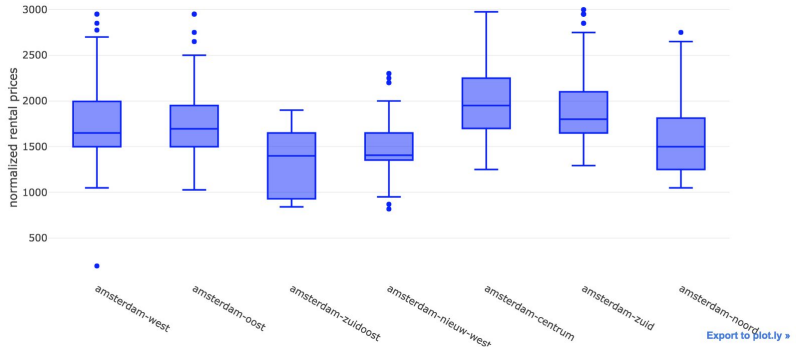
€ 2.250 /mnd Geen kosten huurder
82 m² • 3 kamers • In overleg

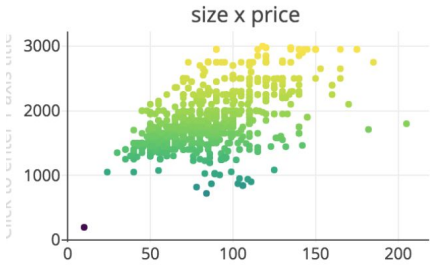
Ramón Mossel Makelaardij o.g. B.V.

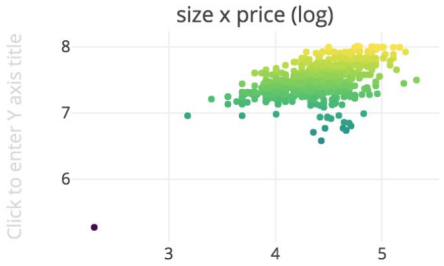
Versus this...


Het IJ
Sloterplas
Plesmanlaan
Plesmanlaan


0.005
0.004
0.003
0.002
0.001
0
500
1000
1500
2000
2500
3000

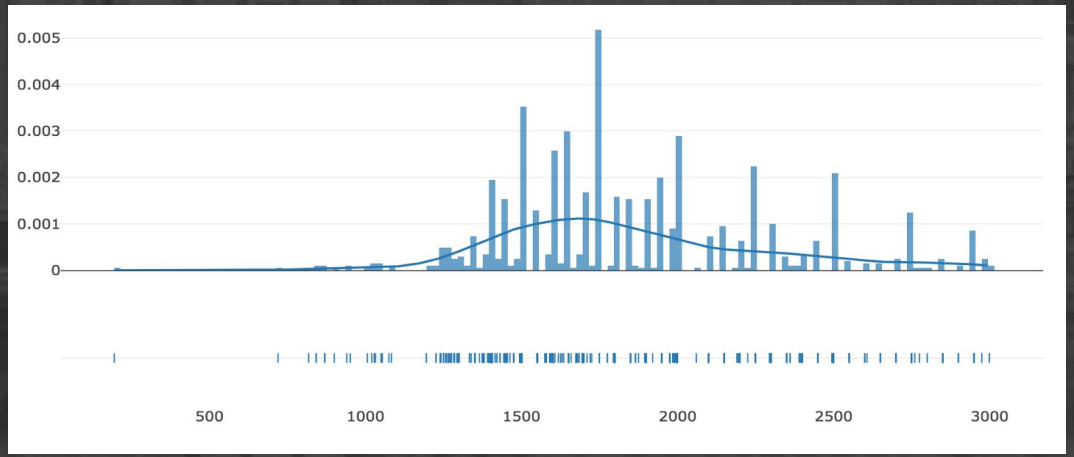

normalized rental prices
3000
2500
2000
1500
1000
500
amsterdam-west
amsterdam-oost
amsterdam-zuidoost
amsterdam-nieuw-west
amsterdam-centrum
amsterdam-zuid
amsterdam-noord
Export to plot.ly »


size x price
3000
2000
1000
0
0
50
100
150
200


size x price (log)
8
7
6
3
4
5

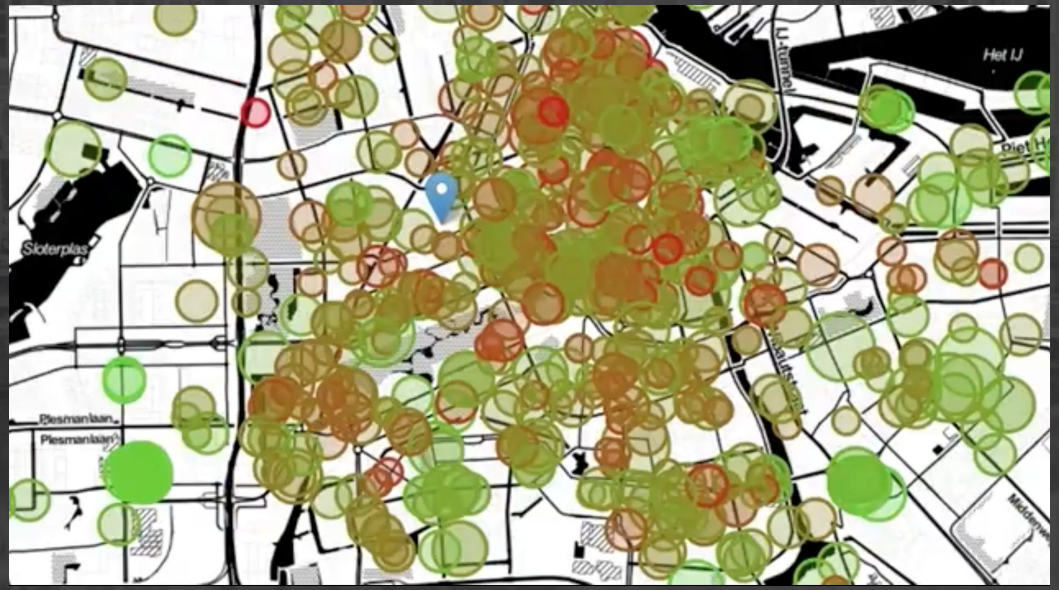
Micro x Macro

How are prices distributed?

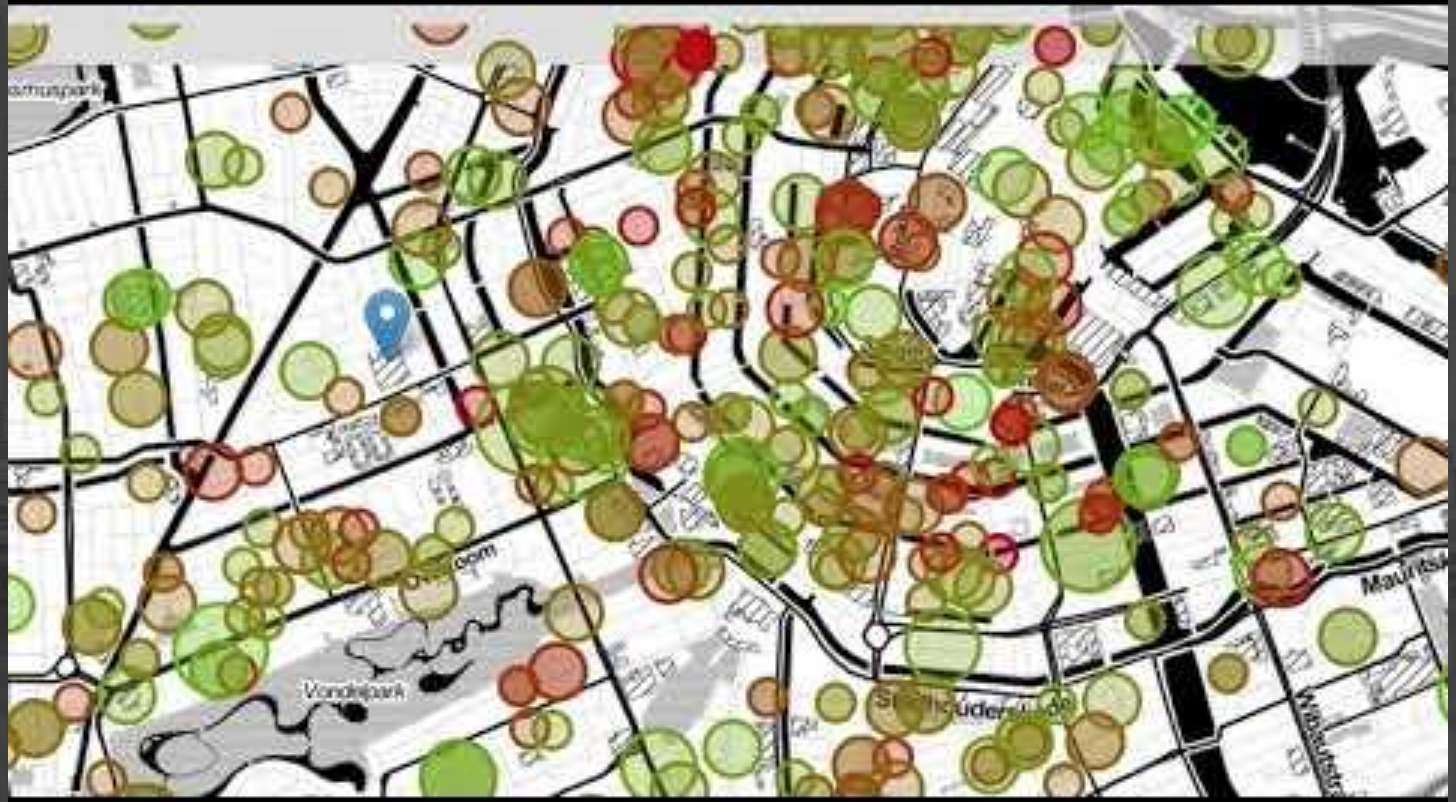


- Prices are concentrated between EUR 1500 and 2000
- Really difficult to find a place for less than EUR 1500!

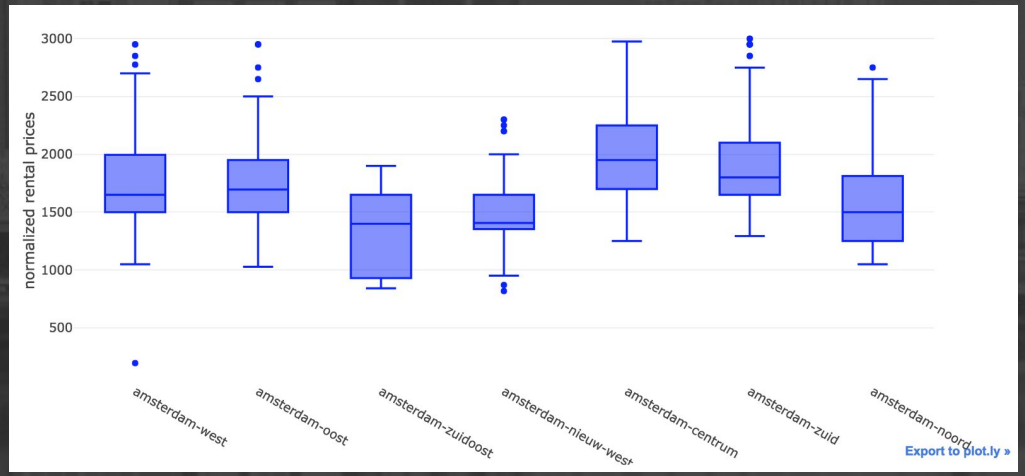
Which are the hottest areas?



- Each circle represents an apartment
- Apartment area is represented by circle radius
- Green or red depending on price per square meter


Vondelpark

Where to start hunting?



- Centrum has the highest median prices
- Niuwe-West and Zuidoost have the lowest median prices

Which is better to make a decision?

Micro x Macro?

What about the surroundings?

- Cafes
- Bars
- Parks
- Transportation
- Schools

De 10 beste plekken in de buurt van Noord, Amsterdam, Noord-Holland

1-10 van 4263 worden weergegeven

Amsterdam > Noord

Alle filters | € | €€ | €€€ | €€€€ | Nu open

1. Winkel 43

580 reviews

€€ · Desserts, Cafés

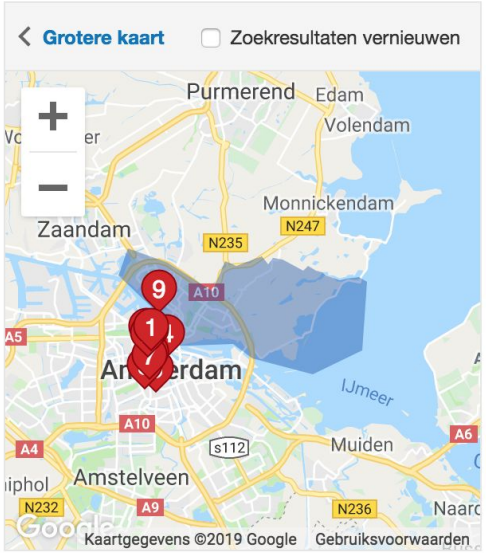
020-6230223
Noordermarkt 43
Jordaan

2. Anne Frank Huis

525 reviews

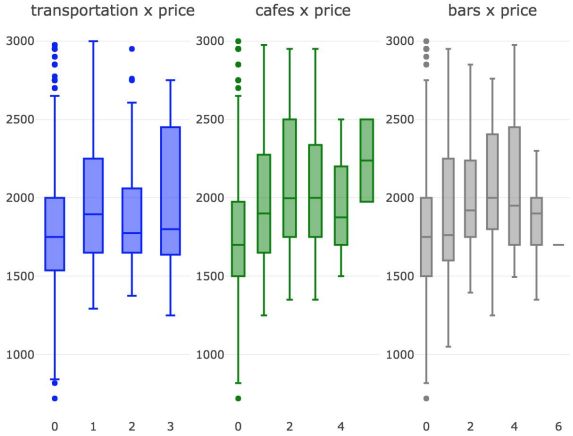
Musea

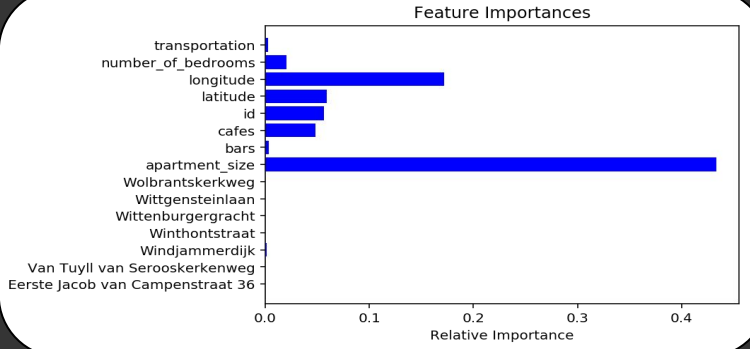
020-5567105
Prinsengracht 267
Centrum

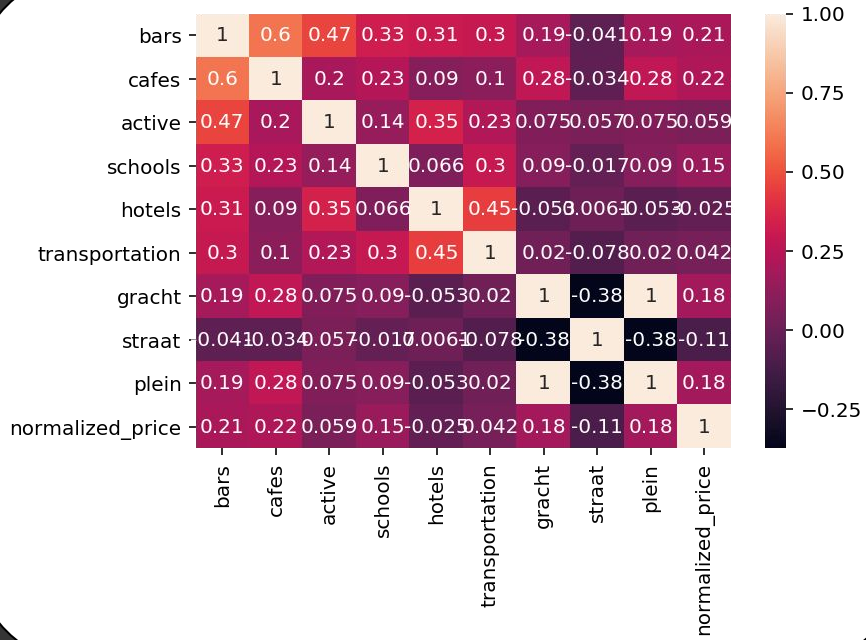


- Over 5 million businesses across 32 countries
- JSON / GraphQL APIs
- Multiple categories: active life, restaurants, education...

How does that affect prices?


transportation x price
cafes x price
bars x price
3000
2500
2000
1500
1000
0 1 2 3
0 2 4
0 2 4 6


Feature Importances
transportation
number_of_bedrooms
longitude
latitude
id
cafes
bars
apartment_size
Wolbrantskerkweg
Wittgensteinlaan
Wittenburgergracht
Winthontstraat
Windjammerdijk
Van Tuyll van Serooskerkenweg
Eerste Jacob van Campenstraat 36
0.0 0.1 0.2 0.3 0.4
Relative Importance


bars 1 0.6 0.47 0.33 0.31 0.3 0.19 -0.041 0.19 0.21
cafes 0.6 1 0.2 0.23 0.09 0.1 0.28 -0.034 0.28 0.22
active 0.47 0.2 1 0.14 0.35 0.23 0.075 0.057 0.075 0.059
schools 0.33 0.23 0.14 1 0.066 0.3 0.09 -0.017 0.09 0.15
hotels 0.31 0.09 0.35 0.066 1 0.45 -0.05 0.0061 0.053 0.025
transportation 0.3 0.1 0.23 0.3 0.45 1 0.02 -0.078 0.02 0.042
gracht 0.19 0.28 0.075 0.09 -0.053 0.02 1 -0.38 1 0.18
straat 0.041 0.034 0.057 0.017 0.0061 0.078 -0.38 1 -0.38 -0.11
plein 0.19 0.28 0.075 0.09 -0.053 0.02 1 -0.38 1 0.18
normalized_price 0.21 0.22 0.059 0.15 -0.025 0.042 0.18 -0.11 0.18 1
bars cafes active schools hotels transportation gracht straat plein normalized_price
1.00
0.75
0.50
0.25
0.00
−0.25

Let's put it to practice!

https://www.github.com/rafaelpierre/JupyterAMS

Thank you!

https://www.github.com/rafaelpierre

https://www.linkedin.com/in/rafaelpierre

https://www.medium.com/@rafaelpierre