

Jcp_Inm Dataset

Pablo Edronkin, 2019

<https://orcid.org/0000-0001-8690-7030>

Abstract.

This is a dataset based in Sqlite3 database format containing data on real estate values for the downtown area of the city of Jose C. Paz, Buenos Aires, Argentina, gathered in 2019.

It is presented in Sqlite3 database format and contains values in US dollars per square meter averaged (built and not built), plus data on the number of pedestrians strolling thorough the streets of the commercial district of the city.

Keywords.

Real estate, database, dataset, data, properties, value, commercial, residential, pedestrian, traffic

Acknowledgements.

This project could not have taken place without the work done by:

- The developers of Sqlite3^[6.1.].

License and info for contributors.

Please read the following files included with this project:

- `README.md`: contains info on setting up your system and installing the files related to this module.
- `CONTRIBUTING.md`: if you want to contribute to this project.
- `COPYING`: for license information.

Table of Contents

Abstract.....	1
Keywords.....	1
Acknowledgements.....	1
License and info for contributors.....	1
1. Introduction.....	4
2. Conventions, caveats and base conditions for development.....	5
2.1. Terms and abbreviations used.....	5
2.2. Caveats.....	5
2.3.1. Samples per AB.....	5
2.3.2. Samples per TB.....	6
2.3.3. Properties on sale per AB.....	6
2.3.4. Urban codes.....	6
2.3.5. The effect of public activities.....	7
2.3.6. Area of sampling blocks.....	8
2.3.7. Currency used for appraisal.....	9
2.3.8. RO and CO properties.....	9
2.3.9. The effect of urban design.....	9
3. Development.....	11
4. Structure and data.....	12
4.1 Data tables in the dataset.....	12
4.1.1. <i>trf_pers</i>	12
4.1.2. <i>trf_ppcm</i>	13
4.1.3. <i>trf_streets</i>	14
5. Use.....	16
6. Sources.....	17

1. Introduction.

This is a dataset presented in Sqlite3 database format containing data on real estate values for the downtown area of the city of Jose C. Paz, Buenos Aires, Argentina, gathered in 2019.

It is presented in Sqlite3 database format and contains values in US\$ per square meter averaged (built and not built), plus data on the number of pedestrians strolling thorough the streets of the commercial district of the city.

SQL queries useful for processing the data contained in the data set are also provided.

2. Conventions, caveats and base conditions for development.

Take into consideration these factors.

2.1. Terms and abbreviations used.

AB: Address block.

C: Commercial zone.

CO: Commercial zone, outside the commercial district.

R: Residential zone.

RO: Residential zone, outside the commercial district.

SQL: Structured query language.

TB: Time block.

2.2. Caveats.

You should be aware of:

2.3.1. Samples per AB.

The number of samples per address block might be not be the same in all cases, but in all cases the number of samples performed is high enough to be statistically significant.

2.3.2. Samples per TB.

The number of samples per time block might not be the same in all cases. Since this data set is geared among other things, towards gather information on pedestrian traffic in relation to the commercial value of different address blocks samples were mostly obtained between 7:00 and 19:00 hours.

However, in some cases night hours might make a difference, so a few samples were taken during night hours as well, when they imply business activities.

2.3.3. Properties on sale per AB.

The number of sampled properties on sale per address block might not be the same in all cases. This is due to market reasons and depends on the decision of owners, which its outside the control possibilities of the authors of this work.

Simply put, it might be possible to find one or more properties per address block being sold over a sufficiently long period of time, but for the purposes of this dataset and the time allotted to building it, such a homogeneity of data is not achievable.

2.3.4. Urban codes.

Urban codes establish that middle to intensive commercial activities shall only take place on blocks of streets that have been declared commercial zones. There are streets that can be partially commercial and partially residential, or being commercial altogether, display a very different pedestrian density depending on factors like the presence of schools or bus stops.

Urban codes guarantee that there is no middle to heavy commercial, or industrial development in residential areas. In this part of the city, there is no allowance for industrial activities so they are not being considered in this dataset.

The price of properties per square meter is influenced by urban codes: indeed, residential areas have lower prices since there are less options to build and develop. Commercial areas, on the other side, allow for both commercial as well as residential construction. Of course, few people tend to prefer living in a commercial zone but they can.

This is the reason why in areas that have been declared commercial only recently and there are still private residents, some people might want to use as a reference the residential square meter price instead of a commercial value.

But since this policy does not make well with common sense, once home owners realize the difference between commercial square meter prices and those for homes, they start selling their properties at the commercial market value.

2.3.5. The effect of public activities.

In some cases, commercial prices per square meter have significant variations between fairly close or even contiguous properties. This is mostly due to certain activities like.

2.3.5.1. Banking. The presence of banks tend to markedly increase the number of pedestrian in a given street block and those found in their way from the train station or bus stops. Normal banking hours go from 10:00 to 16:00.

However, the presence of people using an ATM or waiting in queues at the banks – particularly pensioners and people receiving unemployment subsidies – is significant almost around the clock.

2.3.5.2. Schools and a local university. They have a similar effect as banks, but with reduced hours. Many schools run from 7:00 to 23:00.

Such a schedule includes lessons for primary and secondary school children, up to 17:00, and from then on courses for adults, night lessons for working people, vocational courses, etc.

2.3.5.3. Trains and buses. By far, the train station in the city produces the heaviest pedestrian density in nearby areas, but bus stops also produce significant pedestrian numbers. Moreover: legislation established that within cities there must be a bus stop every 200 m along the route of any bus line.

In commercial areas this guarantees a significant and homogeneous pedestrian density, especially if banks or schools are found nearby.

2.3.5.4. Large commercial establishments: a shopping center, a couple of discos, etc. are also responsible for a relatively large number of regular, visiting pedestrians.

2.3.5.5. Visitors from nearby towns. Since some towns in the region do not offer certain services, people from those usually visit the city to shop, run errands, etc.

2.3.6. Area of sampling blocks.

In Argentina, square blocks tend to measure 100 x 100 meters and streets are laid in a square grid pattern for the most part. In this case, since the urban area of the city includes railway tracks, there are some diagonal streets but since the data set includes data from all the streets in the commercial zone of the town, the numbers represent well the presence and distribution thorough the day and days of the week of pedestrians (customers) in the whole area.

This means that samples were obtained in segments of 100 m on each street. The total number of pedestrians per minute per block indicate the number of pedestrian per each segment of a hundred meters on each street, regardless of the urban characteristics of each segment. That is, if in a commercial street one block instead of being a built up area is a park, pedestrians were accounted for on that segment too.

In some cases, streets have been partially zonified as commercial areas, while in others the same streets remains being residential.

Blocks located within the commercial area of the city were sampled regardless of whether they were defined as commercial or residential blocks based on the fact that such characteristics of a street do influence the value of properties and the number of pedestrians passing thorough.

2.3.7. Currency used for appraisal.

In Argentina, the currency used for real estate appraisals has been traditionally the US dollar. This is mainly a collateral effect of the chronic high inflation of the Argentine peso. Dollars provide a better framework for buyers, investors and real estate agents.

In more recent times, the Bitcoin has been acquiring some relevance in real estate operations in the country. However, the volatility and speculative values surrounding this crypto currency has put a limit to this growth, at least for now.

Hence, for the purposes if this work we will stick with the US dollar.

2.3.8. RO and CO properties.

As expressed in [2.1.], there are some properties classified as RO and CO. These lie outside the commercial district studied but were taken as external control references for the purposes of this study. Their values and characteristics are not part of the statistic itself.

2.3.9. The effect of urban design.

The way in which accesses for pedestrians, bus stops, etc. are distributed also seems to have an effect of pedestrian transit, affecting the number of people that:

2.3.9.1. Visit each street. Undeniably, the town's architecture, good or bad, has an effect on how many people walk on different areas. Some places are better suited for pedestrians than others on the basis of their characteristics such as vehicle traffic, street lights for crossing, viaducts, etc.

This study does not involve itself into the details of urban design except for the fact that the count of pedestrians on each sampling block might vary, for this, among other reasons.

2.3.9.2. Walk on each side of some streets. Aside from what is stated in [2.3.9.2.], the characteristics of urban design, development, etc. influence even how people walk on one side or the other of some streets.

This study takes into account this fact but is not designed to make such differentiation. Statistics on each block include the sum of pedestrians from both sides of each involved street.

3. Development.

Regarding the numbers representing customers in the area, data was gathered in each street, from each block, by counting the pedestrians – people on foot, not vehicles – passing on a period of approximately one minute. People inside shops or buildings were not counted.

Data on commercial and residential values in the area was obtained from the properties being offered publicly for sale at the time.

In order to better pinpoint the date and time in which data was added to the database, each record was provided with a time stamp field.

4. Structure and data.

This section should be considered as the data dictionary of the database.

4.1 Data tables in the dataset.

4.1.1. *trf_pers*

This table contains data regarding the number of persons measured on each street block.

4.1.1.1. *Id*: Primary key. Unique integer that identifies each record in the table.

4.1.1.2. *Context*: describes the area or district within the town.

4.1.1.3. *Status*: ‘*enabled*’ for records to be considered, ‘*disabled*’ for records on hold.

4.1.1.4. *Abr*: A two to four letter code that identifies each street. Correspondence between *Abr* and the actual name of the street is contained in table ‘*trf_streets*’ between fields *Item* and *Abr*.

4.1.1.5. *Alt*: Street coordinate - i.e. street number. In Argentina normally each street block carries numbers by a century or hundred meters counted from what is considered to be the starting point of the street, for a given town. For example, being at the corner of Lavalle 1700 means that this is the 17th block of that street.

Houses are identified by meters. So, Lavalle 1723 is the address that stands at 1723 meters from the beginning of the street. For the purposes of this study, addresses represent the starting point of each street block from which a count of pedestrians is made.

4.1.1.6. *Value*: contains the number of pedestrians counted in the $[Alt, Alt + 100)$ interval for that particular street. That is, the values from the whole block, for a given street as expressed in *Abr*.

4.1.1.7. *Hr*: hour interval at which the sample is taken.

4.1.1.8. *Dayn*: number corresponding to the day of the month.

4.1.1.9. *Dayw*: day of the week, as a three letter abbreviation.

4.1.1.10. *Monthn*: number corresponding to the month.

4.1.1.11. *Year*: self explanatory.

4.1.1.12. *Tstamp*: timestamp, provided by the RDBMS itself.

4.1.2. *trf_ppcm*

Thus table contains data regarding known prices asked for actual properties in the area of interest.

4.1.2.1. *Id*: see [4.1.1.1].

4.1.2.2. *Context*: see [4.1.1.2.].

4.1.2.3. *Status*: see [4.1.1.3.].

4.1.2.4. *Abr*: see [4.1.1.4.].

4.1.2.5. *Alt*: see [4.1.1.5.].

4.1.2.6. *Value*: see [4.1.1.6.].

4.1.2.7. *Zone*: a one or more letter long string that identifies commercial and residential areas relevant to the study. “C” identifies an active commercial zone, “R” identifies an active residential zone.

4.1.2.8. *Monthn*: see [4.1.1.10.].

4.1.2.9. *Tstamp*: see [4.1.1.12.].

4.1.3. *trf_streets*

4.1.3.1. *Id*: see [4.1.1.10.].

4.1.3.2. *Abr*: see [4.1.1.2.].

4.1.3.3. *Item*: official, current street name.

4.1.3.4. *Alias*: unofficial or not current names or aliases by which the street might also be identified.

4.1.3.5. *Avg_pers*: average number of people per block, per sample instance.

4.1.3.6. *Samples_total*: total number of samples per street.

4.1.3.7. *Avg_ppmc*: average price per square meter, commercial.

4.1.3.8. *Avg_ppmr*: average price per square meter, residential.

4.1.3.9. *Value_ppers*: value per person.

4.1.3.10. *Est_final_ppmc*: estimated final price per square meter.

This table contains data considered to be part of the conclusion of the study.

5. Use.

SQLite3 databases are fully self-contained. Users need to have Sqlite3 installed on their system, and preferably a visual database editor. Using SQL queries it is possible to obtain any section of the gathered data, and it can be converted to any data format used by libraries written in R, Python, C++ etc.

It is recommendable for users to review the structure of the data tables as well as the data itself in order to understand the whole package and what can be done with it. In general, tables contain data either as strings, integers or real numbers.

6. Sources.

6.1. Sqlite.org. (2000). SQLite Home Page. [online] Available at: <https://www.sqlite.org/index.html> [Accessed 26 Aug. 2019].

Alphabetical Index

A

ab.....	1, 3pp., 8, 10pp., 15pp.
Ab.....	1, 3, 12pp.
AB.....	3, 5p.
abbreviation.....	3, 5, 13
Abr.....	12pp.
address.....	6, 12
Address.....	5
alias.....	15
Alias.....	15
alt.....	6
Alt.....	12pp.
architecture.....	10
Argentina.....	1, 4, 8p., 12
ATM.....	7
average.....	1, 4, 15
Avg_pers.....	15
Avg_ppmc.....	15
Avg_ppmr.....	15

B

bank.....	7p.
Bank.....	7
Bitcoin.....	9
block.....	3, 5pp., 15
Block.....	9
Buenos Aires.....	1, 4
build.....	6p., 11
building.....	6, 11
bus.....	6pp.
business.....	6

C

C++.....	16
cases.....	6p., 9
caveat.....	3, 5
Caveat.....	3, 5
city.....	1, 4, 7pp.
co.....	1pp.
Co.....	3, 5, 7, 12, 14
CO.....	2p., 5, 9
code.....	3, 6p., 12
commercial.....	1, 4pp., 11, 14p.
Commercial.....	5, 7

conclusion.....	15
construction.....	7
Context.....	12, 14
contribute.....	2
CONTRIBUTING.....	2
COPYING.....	2
count.....	8pp.
currency.....	9
Currency.....	3, 9
customer.....	8, 11

D

data.....	1, 3p., 6pp., 11pp., 15p.
Data.....	1, 3, 11p.
dataset.....	1, 3p., 6p., 12
Dataset.....	1
day.....	8, 13
Day.....	13
Dayn.....	13
Dayw.....	13
density.....	6, 8
develop.....	1
developer.....	1
development.....	3, 5, 7, 10
Development.....	3, 11
dictionary.....	12
difference.....	6p.
district.....	1, 4p., 9, 12
dollar.....	1, 9
Dollar.....	9

E

Edronkin.....	1
es.....	1p., 17
Est_final_ppmc.....	15
estate.....	1, 4, 9

F

file.....	1p.
format.....	1p., 4, 6, 16

H

home.....	7
Home.....	17

homogeneous.....	8
hour.....	6p., 13
hr.....	9, 13
Hr.....	13

I

id.....	1, 4pp.
Id.....	12pp.
industrial.....	7
information.....	2, 6
install.....	2
invest.....	9
investor.....	9
Item.....	12, 15

J

Jose C. Paz.....	1, 4
------------------	------

L

language.....	5
libraries.....	16
license.....	2
License.....	1, 3

M

market.....	6p.
meter.....	1, 4, 7p., 12, 15
module.....	2
Monthn.....	13p.

N

night.....	6, 8
norm.....	12
Norm.....	7
number.....	1, 4, 6pp., 11pp., 15p.

O

orcid.....	1
------------	---

P

pedestrian.....	1, 4, 6pp.
people.....	7pp., 15
People.....	11
peso.....	9
policy.....	7
price.....	7, 13, 15
project.....	1p.
properties.....	1, 3, 6p., 9, 11, 13
Properties.....	3, 6
public.....	3, 7, 11
Python.....	16

Q

query.....	5
------------	---

R

README.....	2
real.....	1, 4, 7, 9, 16
Real.....	1
recent.....	7, 9
residential.....	1, 6p., 9, 11, 14p.
Residential.....	5
ro.....	1pp., 6pp.
RO.....	3, 5, 9

S

sale.....	3, 6, 11
sample.....	6, 8p., 13, 15
Sample.....	3, 5p., 15
Samples_total.....	15
sampling.....	3, 8, 10
school.....	6pp.
School.....	7
segment.....	8
service.....	8
setting.....	2
shop.....	8, 11
shopping.....	8
significant.....	6pp.
speculative.....	9
sql.....	17
Sql.....	1, 4, 16p.
SQL.....	4p., 16p.
sqlite.....	17

Sqlite.....	1, 4, 16p.
SQLite.....	16p.
Sqlite3.....	1, 4, 16
SQLite3.....	16
square.....	1, 4, 7p., 15
stamp.....	11, 13p.
station.....	7p.
statistic.....	6, 9
Statistic.....	10
Status.....	12, 14
stop.....	6pp.
street.....	1, 3p., 6pp.
Street.....	12
string.....	14, 16
stroll.....	1, 4
structure.....	16
Structure.....	3, 5, 12
study.....	9p., 12, 14p.
system.....	2

T

TB.....	3, 5p.
Term.....	3, 5
time.....	6, 9, 11, 13
Time.....	5
timestamp.....	13
town.....	1, 4, 8, 10, 12
traffic.....	1, 6, 10
train.....	7p.
Train.....	8
trf_pers.....	3, 12
trf_streets.....	3, 12, 14
Tstamp.....	13p.

U

university.....	7
urban.....	3, 7pp.
Urban.....	3, 6p.

V

value.....	1, 4, 6p., 9, 11, 13, 15
Value.....	13pp.
Value_ppers.....	15
vehicle.....	10p.
viaduct.....	10
volatility.....	9

Y

Year.....13

Z

zone.....5pp., 14
Zone.....14
zonified.....9