

EMOS Statistical Production Systems – Datasets

There are two datasets available for the assignment: the buildingsregister (BAG) and a dataset with sale prices of houses.

Buildingsregister

The buildings register (BAG – Basisadministratie Gebouwen) is an official register used by all government agencies. It is the official source for information about buildings and addresses.

This dataset consists of multiple files. Each of the datasets contains information about a different object type. The files can be linked to each other using shared linkage keys. The names of the files and variables are in Dutch following the official names.

The dataset used for the assignment only contains information for the area around the city of Tilburg.

The dataset can be found at: https://research.cbs.nl/emos/buildings_register.zip

panden.csv: buildings

This file contains information about buildings.

Variable	Type	Description
id	string	id of building (numeric; but too long (16 digits) to store in int/double)
bouwjaar	integer	year built
pandstatus	string	status of building
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty
x	double	x coordinate of centroid of building (Rijksdriehoek; EPSG:28992; unit metre)
y	double	y coordinate

Possible values in the field ‘pandstatus’:

Value	Description
Bouwvergunning verleend	Building permit granted
Pand ten onrechte opgevoerd	Building incorrectly listed
Pand in gebruik	Building in use
Pand gesloopt	Building demolished

Pand in gebruik (niet ingemeten)	Building in use (not measured)
Bouw gestart	Construction started
Sloopvergunning verleend	Demolition permit granted
Verbouwing pand	Building renovation
Niet gerealiseerd pand	Building not realized

openbareruimte.csv: public spaces

Contains information about public spaces such as streets, squares, roads.

Variable	Type	Description
id	string	id of public space (numeric; but too long (16 digits) to store in int/double)
name	string	name public space
type	string	type of public space
woonplaats	integer	town (links to id in woonplaatsen.csv)
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty

Possible values for the field ‘type’:

Value	Description
Weg	Road
Terrein	Terrain
Landschappelijk gebied	Landscape area
Administratief gebied	Administrative area
Kunstwerk	Infrastructure work (e.g. bridge or tunnel)
Water	Water

verblijfsobjecten.csv: dwellings

Contains information about dwellings. Dwellings are contained in buildings and are the smallest unit. This is where persons live or companies/shops are located. Each dwelling has a unique address.

Variable	Type	Description
----------	------	-------------

id	string	id of use object (numeric; but too long (16 digits) to store in int/double)
gebruiksdoel	string	use goal
oppervlakte	integer	area (square metres)
status	string	status
hoofdadres	string	id of main address (links to id in nummeraanduidingen.csv)
pand	string	id of related building; the building in which the dwelling is located (links to id in panden.csv)
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty
x	double	x coordinate of centroid (rijksdriehoek)
y	double	y coordinate of centroid (rijksdriehoek)

Possible values for the field 'gebruiksdoel':

Value	Description
Woonfunctie	Residential function
Industriefunctie	Industrial function
Winkel functie	Retail function
Kantoorfunctie	Office function
Bijeenkomstfunctie	Assembly function
Sportfunctie	Sports function
Gezondheidszorgfunctie	Healthcare function
Logiesfunctie	Lodging function
Overige gebruiksfunctie	Other usage function
Onderwijsfunctie	Educational function
Celfunctie	Prison function

Possible values for the field 'status':

Value	Description
Verblijfsobject gevormd	Dwelling formed
Verblijfsobject ten onrechte opgevoerd	Dwelling incorrectly listed

Verblijfsobject in gebruik	Dwelling in use
Verbouwing verblijfsobject	Renovation of dwelling
Verblijfsobject in gebruik (niet ingemeten)	Dwelling in use (not measured)
Verblijfsobject ingetrokken	Dwelling withdrawn
Verblijfsobject buiten gebruik	Dwelling out of use
Niet gerealiseerd verblijfsobject	Dwelling not realized

nummeraanduidingen.csv: addresses

Contains the actual addresses. An address consists of a postcode and number and addition (e.g. 1234AB 110A). These can be linked to a public space (street).

Variable	Type	Description
id	string	id of number designation (numeric; but too long (16 digits) to store in int/double)
postcode	string	postcode
huisnummer	integer	address number
huisnummertoevoeing	string	address number addition
openbareruimte	string	id of corresponding public space (links to id in openbareruimte.csv)
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty

woonplaatsen.csv: towns

Contains information about towns (e.g. 'Amsterdam'). This is different from a municipality. A municipality can contain multiple towns (for example the municipality Amsterdam contains the towns Amsterdam, Weesp, Sloterdijk etc.)

Variable	Type	Description
id	integer	id of town
name	string	name
status	string	status
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty
x	double	x coordinate of centroid (rijksdriehoek)

y	double	y coordinate of centroid (rijksdriehoek)
---	--------	--

gemeente_woonplaats.csv: municipalities

Contains information about the municipalities the towns belong to.

Variable	Type	Description
woonplaats_id	integer	id of town (links to id in woonplaatsen.csv)
gemeente_id	integer	id of municipality (official id of municipality)
status	string	status
begin_geldigheid	date	start date of validity record (format: YYYY-MM-DD)
eind_geldigheid	date	end date of validity; often empty

Possible values for the field 'status':

Value	Description
Woonplaats aangewezen	Town designated
Woonplaats ingetrokken	Town withdrawn

Salesprices

The file sales.csv is a fictive dataset containing data about housing prices that CBS obtained from an external source.

Variable	Type	Description
address	string	Address of sold dwelling ('Mainstreet 1-b')
postcode	string	Postcode of the sold dwelling ('1234AB')
city	string	The city/town in which the dwelling is located.
sales_price	integer	Price in k€ for which the dwelling was sold.
sales_data	date	Date at which the dwelling was sold (format YYYYMMDD)