

Kopi:

d. 21.2.2011

Kathrine Lausted Veie

Dok. nr.

## **Dokumentation: Identifikation af markeder**

**Dette notat beskriver inddelingen af undersøgelsesområdet i 9 forskellige delmarkeder. Denne inddeling er en forudsætning for udførelsen af første trin af husprisanalysen, hvor markedsspecifikke marginalpriser estimeres. Første trin af husprisanalysen er beskrevet i notatet ”Dokumentation for husprisanalysens første trin: Marginalpriser”, mens andet trin af analysen er beskrevet i ”Dokumentation: Husprisanalysens andet trin: Efterspørgsel efter fravær af støj”.**

Identifikationen af del-markeder indenfor undersøgelsesområdet i Storkøbenhavn er vigtig for korrekt estimation af implicitte priser på støj. Den hedoniske metode afhænger af, at undersøgelsesområdet udgør et enkelt marked i estimeringen af første trin, da de estimerede priser skal afspejle markedsprisen på markedet og ikke et vægtet gennemsnit af flere forskellige markeder.

I litteraturen er flere forskellige metoder anvendt til markedsidentifikation. I nogle tilfælde er administrative inddelinger anvendt (f.eks. kommuner), mens andre har benyttet sig af samtaler med ejendomsmæglere i området eller statistiske metoder til segmentering, som clusteranalyser. Vi har taget udgangspunkt i et speciale omkring bosætningsmønstre i Storkøbenhavn, hvor husstandenes socioøkonomiske karakteristika er beskrevet, jf. Poulsen (2008). Dette har givet anledning til en foreløbig inddeling af undersøgelsesområdet i markeder. Herefter er der udført en clusteranalyse for at nuancere inddelingen.

En clusteranalyse identificerer grupper i data, hvor observationerne i en gruppe ligner hinanden mere, end de ligner observationer i de øvrige grupper. Et delmarked udgøres af boliger, der tiltaler et bestemt markedssegment og derfor formentlig har en del fælles træk. Her spiller formentlig både boligens type og beliggenhed en rolle. Ideelt set burde man tage udgangspunkt i efterspørgselsiden og identificere grupper af købere. Imidlertid spiller geografi en væsentlig rolle i forhold til markedssegmenteringen, og vi havde ikke mulighed for at koble detaljerede oplysninger om den enkelte boligs beliggenhed direkte med detaljerede oplysninger om husstandens sammensætning på grund af Dan-

marks Statistiks krav til datas anonymitet. Derfor har vi baseret markedsinddelingen på oplysninger om boligens type og dens beliggenhed alene.

### Antallet af delmarkeder

For at udføre den type clusteranalyse, vi har benyttet, er det nødvendigt at bestemme antallet af delmarkeder i udgangspunktet. Det drejer sig primært om at finde ud af, hvor mange delmarkeder, området skal inddeles i for på en rimelig måde at beskrive den heterogenitet, der er på boligmarkedet i Storkøbenhavn. Her har vi afprøvet forskellige inddelinger mellem 6 og 18 markeder. Tabellen nedenfor viser dog alene inddelingen for 6, 8, 9, 10 og 12 markeder. Der er angivet en startinddeling i områder baseret på Poulsen (2008) og egne vurderinger. Disse er lagt til grund for clusteranalysen som initiale clustre, jf. tabel 1. I den senere clusteranalyse er det testet, om startinddelingen har haft betydning for resultatet ved at lave clusteranalyser med tilfældigt startpunkt og samme antal clustre. Der er ganske lidt variation i de clustre, der kommer ud af analysen med tilfældigt startpunkt og de clustre, som kommer ud ved brug af de på forhånd definerede startpunkter. Vi holder os til sidstnævnte, da resultaterne så kan reproducere.<sup>1</sup>

*Tabel 1: Initial opdeling af undersøgelsesområdet på baggrund af postnumre, Poulsen (2008) og egne vurderinger*

Nr.	Område (ud fra postnumre)	6 markeder	8 markeder	10 markeder	12 markeder
1	København K	37.449	22.786	22.786	3.421
2	Østerbro/Hellerup				9.955
3	Vesterbro/Vanløse				6.549
	Frederiksberg C				
4	Nørrebro/NV				8.502
5	Frederiksberg	14.663	14.663	9.022	
6	Gentofte/Lyngby	12.283	12.283	12.283	12.283
7	Amagerbro indre	15.912	10.935	5.471	10.935
	Amagerbro ydre			5.464	
8	Tårnby/Dragør			4.977	
9	Brønshøj/Valby/SV	16.829	16.829	8.931	8.931
10	Rødovre/Hvidovre			7.898	7.898
11	Ballerup/Herlev/Gladsaxe	8.129	8.129	8.129	8.129
12	Brøndby/Albertslund/Glostrup/ Vallensbæk/Høje Taastrup	10.458	10.458	10.458	10.458
	I alt	101.060	101.060	101.060	101.060

<sup>1</sup> Når ikke en startinddeling angives, vælges et tilfældig sæt af observationer som midtpunkter for clusteranalysen. Derfor kan resultaterne variere lidt mellem forskellige analyser på det samme datasæt, hvis ikke startinddelingen angives.

## Metode til clusteranalyse

Til clusteranalysen er anvendt metoden *K-means clustering*.<sup>2</sup> Denne kræver som input, at man specificerer antallet af clusters, jf. ovenfor. Derefter tages der udgangspunkt i en observation som centrum for hvert cluster, og de forskellige observationer allokeres til et cluster ud fra et nærhedskriterium. Der er valgt forskellige variable, som observationerne grupperes efter:

- BBR: Boligareal, grundareal, boligens alder på salgstidspunktet.
- Geografi: Koordinater for beliggenhed, afstand til rådhuset i Kbh, afstand til kysten, afstand til nærmeste tog-/metrostation.

Variablene er standardiseret for at undgå, at en enkelt variabel trækker hele allokeringen på clusters. Det er nødvendigt, da der er meget forskel i størrelse og variation af de forskellige variable, jf. tabel 2. Denne tabel viser beskrivende statistik for de anvendte variable i det samlede datasæt.

Tabel 2: Beskrivende statistik: Variable til dannelse af clusters

Variabel	Obs	Mean	Std. Dev	Min	Max
Nord	101060	6177153.0	4817.6	6163211.0	6189558.0
Øst	101060	720982.4	4968.4	703295.5	731466.5
Boligareal	101060	93.1	40.2	35.0	287.0
Grundareal	101060	212.5	341.8	0.0	1998.0
Alder ved salg	101060	62.5	34.2	-1	426
Afstand til					
- Centrum	101060	6611.0	4301.5	11.8	21256.0
- Kysten	101060	3619.5	2682.6	20.3	15581.5
- Tog/Metro	101060	828.6	684.0	23.8	6460.5

Variablene er standardiserede ved at deres maksimum er sat til 1 for at få variationen mellem observationerne på hver dimension til at variere på samme skala. Da koordinaterne for boligens beliggenhed spænder over et større spektrum end de øvrige variable, betyder det, at den geografiske komponent i allokeringen til clustre spiller en betydelig rolle. Derfor er denne korrigeret ved først at standardisere den til middelværdi 0 og varians 1, og derefter korrigeret i forhold til den største observation på hhv. de standardi-

---

<sup>2</sup> Analysen er udført med brug af STATA v.11. Det er STATAs indbyggede cluster-kommando, der benyttes.

serede længde og breddegrader. Tabel 3 viser beskrivende statistik for de standardiserede variable.

*Tabel 3: Beskrivende statistik, standardiserede variable*

Variable	Obs	Mean	Std. Dev.	Min	Max
Nord	101060	0.0	0.4	-1.1	1
Øst	101060	0.0	0.5	-1.7	1
Boligareal	101060	0.3	0.1	0.1	1
Grundareal	101060	0.1	0.2	0.0	1
Alder ved salg	101060	0.1	0.1	-0.0	1
Afstand til					
- Centrum	101060	0.3	0.2	0.0	1
- Kysten	101060	0.2	0.2	0.0	1
- Tog/Metro	101060	0.1	0.1	0.0	1

Det ses, at standardafvigelserne efter standardiseringen (tabel 3) er mere ensartede end standardafvigelserne for de oprindelige variable (tabel 2). Det betyder, at vægtningen af de forskellige variable i clusteranalysen er mere ensartet, om end koordinaterne for den geografiske beliggenhed stadig spiller den største rolle, da disse indeholder mest variation.

Den anvendte målemetode er den almindelige geometriske afstand mellem to punkter, som anvendes på de forskellige dimensioner på hvilke clusteranalysen udføres:

$$d_{jk} = \sqrt{\sum_i (x_{ij} - x_{ik})^2}$$

Her er  $d$  afstanden for observation  $j$  til cluster  $k$ , som består af afvigelser fra clustrets centrum på hver af de  $i$  variable ( $x$ ), der indgår i clusteranalysen.

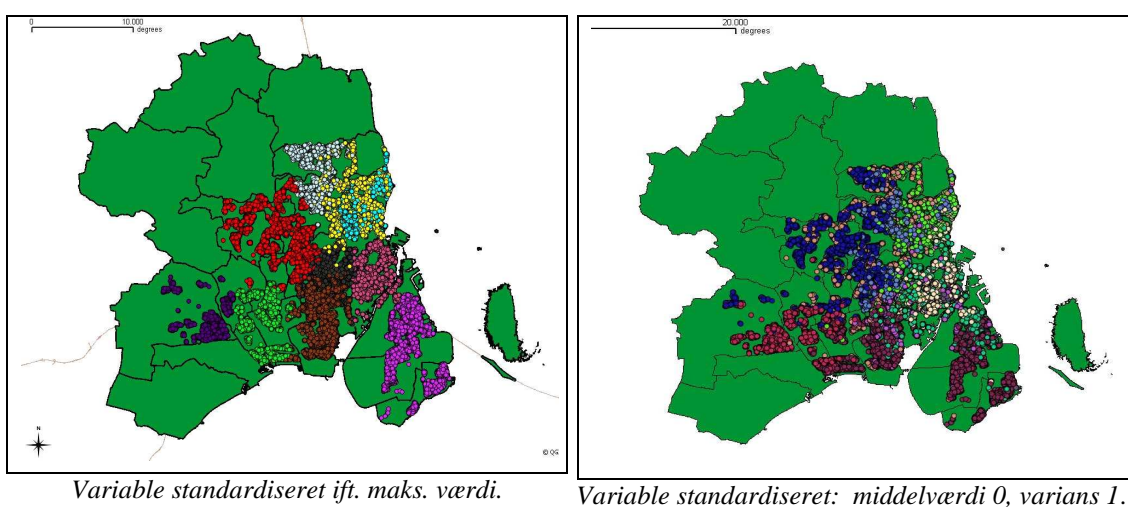
### **Resultater af clusteranalysen**

Indledningsvist er clusteranalyser med forskellige måder at standardisere variablene på afprøvet som en følsomhedsanalyse af de forskellige metoder til standardisering. Der er udført en standard clusteranalyse med et sæt variable standardiseret i forhold til deres eget maksimum (maksimalværdien sat til 1) og et sæt som er standardiseret til at have middelværdien 0 og variansen 1:

$$x_{std1} = x/x_{\max} \quad \text{vs.} \quad x_{std2} = (x - \bar{x})/\sigma_x$$

Disse giver anledning til meget forskellige resultater, idet clusteranalysen med variable standardiseret i forhold til maksimalværdien giver et væsentligt tydeligere mønster, jf. figur 1. Dette skyldes formentlig at vægten på geografien er lidt højere med den pågældende metode end den er på de øvrige variable. Af hensyn til brugen af geografiske fixed effects til håndtering af udeladte variable i første trin af husprisanalysen er det ikke uden betydning, at områderne er sammenhængende. Derfor er standardisering i forhold til max-værdien anvendt i de øvrige clusteranalyser. Dog er koordinaterne for boligens beliggenhed også standardiseret til at have middelværdi nul jf. ovenfor.

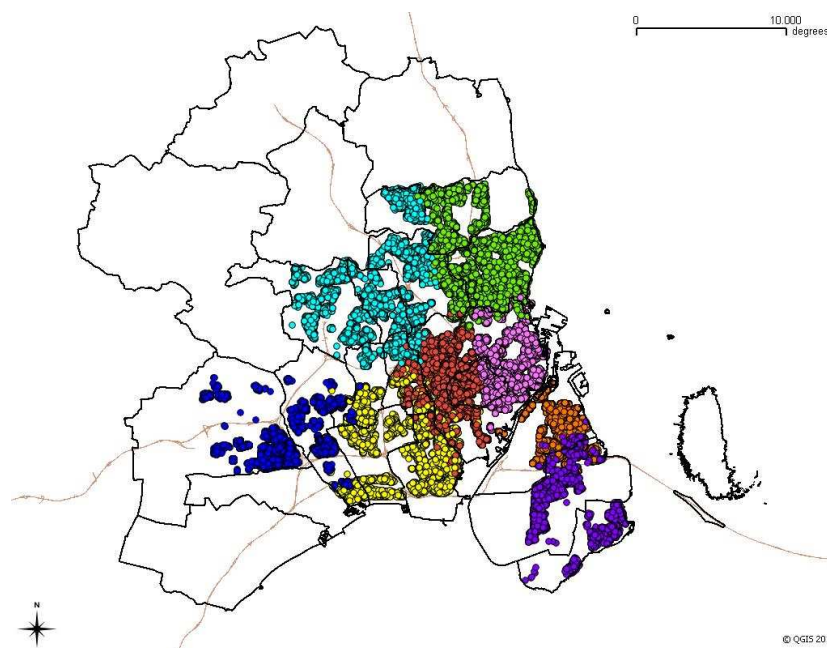
*Figur 1: Sammenligning af standardiseringer af variable, 10 markeder – tilfældig initial opdeling.*



Der er lavet clusteranalyser på 6 til 18 markeder i trin af to (6,8,10 etc.). Grundlæggende sker der det, når man går fra 6 til 8 til 10 markeder mfl., at de eksisterende markeder, som var mest heterogene, splittes op. I Day mfl. (2007) anvendes 8 markeder på et område af samme størrelse som Storkøbenhavn. Bl.a. på den baggrund vurderes 18 markeder at være i overkanten. Derfor tager vi her en inddeling, som stemmer nogenlunde overens med vores forståelse af Storkøbenhavns bolig mønstre, og som vurderes at fange en rimelig grad af kompleksiteten på det storkøbenhavnske boligmarked.

Den foretrukne markedsinddeling er en opdeling af området på 9 markeder, som er foretaget dels ud fra clusteranalysen og dels ud fra justeringer af dennes resultat i overensstemmelse med egne indtryk af København. Det vurderes, at clusteranalysens inddeling i 8 markeder, inddeler Storkøbenhavn på en måde, der stemmer nogenlunde overens med, hvordan byområdet opfattes – den er gengivet nedenfor i figur 2:

Figur 2: Inddeling i 8 markeder.



De 8 markeder identificeret ovenfor er beskrevet i tabellen herunder:

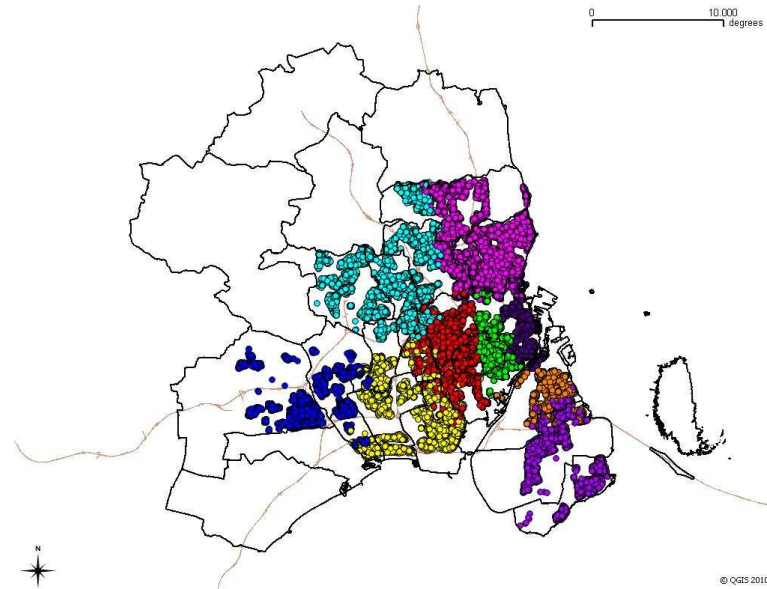
Markedsnr.	Beskrivelse	Antal obs.
1 (grøn)	Området omkring Gentofte	14484
2 (rød)	Udkanten af KBH city	18968
3 (pink)	Centrum af KBH	26908
4 (orange)	Indre Amager inkl. Christianshavn	13830
5 (gul)	Forstæderne Hvidovre, Glostrup, Brøndby	7679
6 (turkis)	Ballerup/Herlev/Gladsaxe	9005
7 (blå)	Høje Taastrup/Albertslund	6771
8 (lilla)	Ydre Amager/Dragør	3415

Imidlertid vurderes det, at indre by er mere heterogent på befolkningssammensætningen end det er afspejlet ved inddelingen i 8 markeder.<sup>3</sup> Specifikt er der dele af indre by, der kan betegnes som værende specielt attraktive. Marked 3 opdeles derfor, så København K, Østerbro og Hellerup er et marked for sig. Der kan argumenteres for, at Frederiksberg C også hører med i denne gruppe. Imidlertid er det vigtigt med sammenhængende områder for brugen af fixed effects i første trin af husprisanalysen, så der er tilstrække-

<sup>3</sup> I bilag 1 ses en oversigt over markedsinddelingen i forhold til postdistrikter.

ligt med observationer indenfor et "kvarter". På den baggrund er det valgt at lade Frederiksberg være i et sammenhængende marked. Opdelingen er vist nedenfor i figur 3.

Figur 3: 9 markeder, egne justeringer.



Oversigten over de nydannede markeder er vist i tabel 4 nedenfor:

Tabel 4: Oversigt over markederne.

Markedsnummer	Beskrivelse	Antal observationer
1 (grøn)	Nørrebro, Frederiksberg, Vesterbro	16.331
2 (rød)	Udkanten af KBH city	18.933
3 (pink)	Området omkring Dyrehaven	14.646
4 (orange)	Indre Amager uden Christianshavn	12.678
5 (gul)	Forstæderne Hvidovre, Glostrup, Brøndby	7.681
6 (turkis)	Ballerup/Herlev/Gladsaxe	8.881
7 (blå)	Høje Taastrup/Albertslund	6.771
8 (lilla)	Ydre Amager/Dragør	3.415
9 (mørk lilla)	Østerbro, København K, Hellerup	11.724

Testregressioner giver indtryk af, at der er forskelle mellem markederne på husprismarkederne. Se også dokumentationsnotatet om første trin.

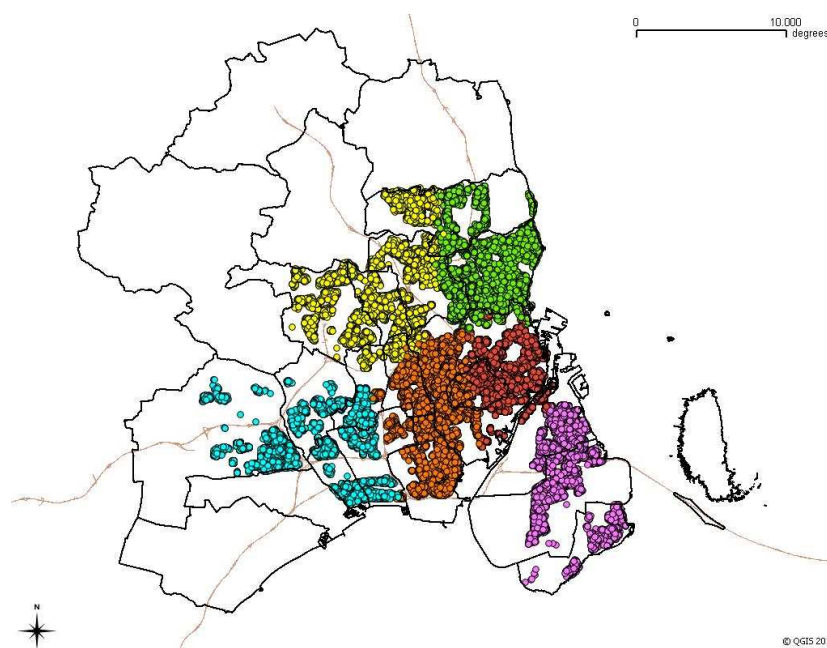
### Øvrige resultater fra clusteranalysen

De forskellige resultater fra clusteranalysen er kommenteret kort nedenfor. Der fokuseres på, hvordan markedsinddelingen tager sig ud for hhv. 6, 8, 10 og 12 markeder, jf. figur 4-7. Forandringerne mellem hver inddeling er kommenteret nedenunder billederne.

Med 6 markeder er Storkøbenhavn tydeligvis inddelt i:

- De nordlige forstæder: Gentofte, Lyngby.
- De nordvestlige forstæder: Ballerup, Herlev.
- De sydvestlige forstæder: Albertslund, Høje Taastrup.
- De nære forstæder: Rødovre, Hvidovre, Vanløse
- Centrum
- Amager

Figur 4: Inddeling i 6 markeder.



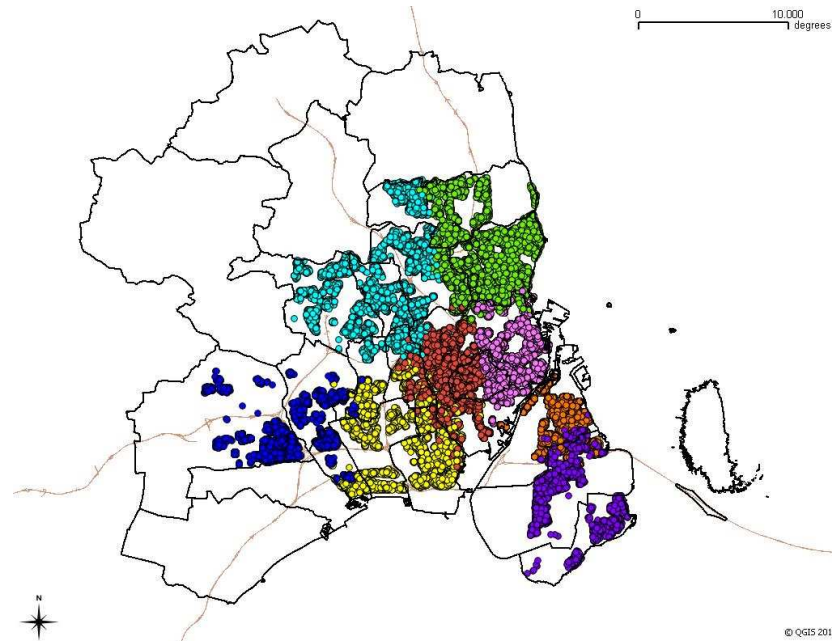
Ved opdeling til 7 markeder opsplittes de sydvestlige og nære forstæder til Albertslund/Høje Taastrup og Brøndby/Glostrup/Rødovre/Hvidovre, mens området fra Utterslev mose i nord til Sydhavnen bliver en del af centrum (ikke vist).

Ved yderligere opdeling i 8 markeder, bibeholdes splittelsen mellem Albertslund/Høje Taastrup (mørkeblå), men der er nu et marked med Brøndby/Hvidovre (gul), og et med Vanløse, ydre Frederiksberg og Brønshøj (rød). Derudover er Amager nu delt i indre (orange) og ydre Amager (lilla). Denne opdeling splitter ligesom opdelingen i seks mar-



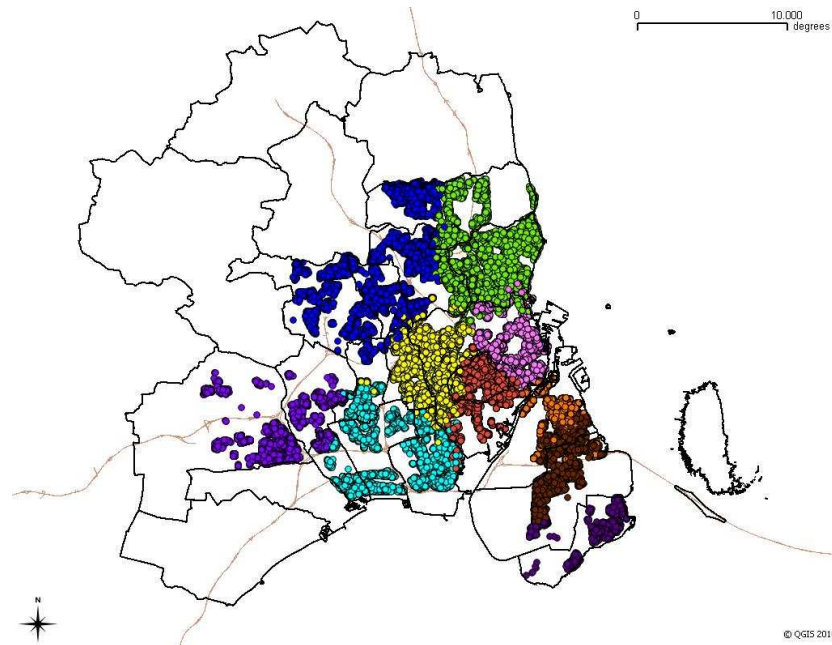
keder, Frederiksberg på midten, hvilket stemmer meget godt overens med idéen om at "centrum" er indenfor Jagtvej/Falconer Allé.

*Figur 5: Inndeling i 8 markeder.*



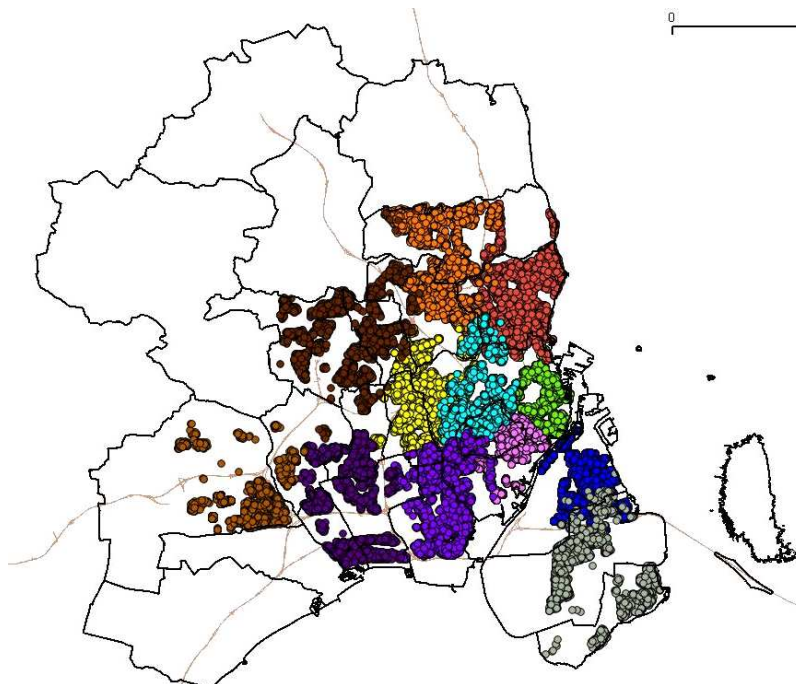
Ved en opdeling i 10 markeder opsplittes Amager yderligere (mørk brun/mørk lilla), så området ved lufthavnen/Dragør er for sig. Derudover opdeles centrum i to, så der er et marked med Vesterbro og Frederiksberg og et med Østerbro/Hellerup/Nørrebro. Her er Frederiksberg dog stadig delt ift. forstæderne, men væsentlig længere ude.

Figur 6: Inndeling i 10 markeder.



I inddelingen til 12 markeder samles det sydlige Amager, mens indre by splittes yderligere. Der er således et område omkring det nordlige Frederiksberg og Brønshøj, et område omkring Frederiksberg/Vesterbro og et omkring Valby/Hvidovre.

Figur 7: Inndeling i 12 markeder.



### **Anvendelse i første trin i husprisanalysen**

Inddelingen i 9 markeder på baggrund af clusteranalysen og egne vurderinger føres videre til husprisanalysens første trin, hvor der estimeres husprisfunktioner for hvert enkelt marked. Resultatet af disse estimationer giver mulighed for at beregne markedsspecifikke marginalpriser på de forskellige karakteristika ved en bolig. Inden estimationerne udføres, er nogle få observationer flyttet fra et marked til et andet. Brugen af fixed effects på "kvarterer" (dvs. sogne og skoledistrikter) kræver, at der er nok observationer i et kvarter til at parametrene i modellen kan identificeres på afvigelser indenfor kvartere. Her har vi fulgt Abbott og Klaiber (2010) og sikret, at der er mindst 30 observationer i hvert kvarter. Det har betydet, at enkelte observationer, som var allokeret til et marked måtte flyttes til nabomarkedet, for at være i en klynge, hvor der var nok observationer.

### Litteratur

Abbott, J.K. og H.A. Klaiber (2010): An Embarrassment of Riches: Confronting Omitted Variable Bias and Multi-Scale Capitalization in Hedonic Price Models. *Review of Economics and Statistics*, forthcoming .

Day, B., I. Bateman og I. Lake (2007): Beyond Implicit Prices: Recovering Theoretically Consistent and Transferable Values for Noise Avoidance from a Hedonic Property Price Model. *Environmental and Resource Economics*, 37 , s. 211-232.

Poulsen, D. (2008): Husholdningernes boligforbrugs- og beliggenhedsvalg i den urbane struktur, i storkøbenhavnsk perspektiv. Speciale ved Økonomisk Institut. Københavns Universitet.

**Bilag 1: Tabel over opdeling af analysens 8 markeder i forhold til postdistrikter**

En tabel sammenholder de 8 markeder ovenfor med postdistrikter i undersøgelsesområdet. De områder, som reallokeres manuelt til marked 9 er markeret med gul.

*Tabel A: Postdistrikter og markedsinddeling.*

Postdistrikt	Markeder								Total
	1	2	3	4	5	6	7	8	
Albertslund	0	0	0	0	0	0	1,308	0	1,308
Bagsværd	0	0	0	0	0	1,570	0	0	1,570
Ballerup	0	0	0	0	0	1,314	0	0	1,314
Brøndby	0	0	0	0	979	0	4	0	983
Brøndby Strand	0	0	0	0	657	0	0	0	657
Brønshøj	0	2,398	0	0	1	493	0	0	2,892
Charlottenlund	0	0	2,879	0	0	0	0	0	2,879
Dragør	0	0	0	0	0	0	0	1,672	1,672
Dyssegård	0	0	631	0	0	0	0		631
Frederiksberg	2,638	3,510	0	0	13	0	0	0	6,161
Frederiksberg C	2,861	0	0	0	0	0	0	0	2,861
Gentofte	0	0	2,389	0	0	0	0	0	2,389
Glostrup	0	0	0	0	1,179	202	657	0	2,038
Hellerup	884	0	1,616	0	0	0	0	0	2,500
Herlev	0	30	0	0	0	1,995	0	0	2,025
Hvidovre	0	1,245	0	0	2,882	0	0	0	4,127
Kastrup	0	0	0	1,604	0	0	0	1,700	3,304
Kgs.Lyngby	0	0	3,402	0	0	362	0	0	3,764
Klampenborg	0	0	388	0	0	0	0	0	388
København K	2,269	0	0	1,152	0	0	0	0	3,421
København N	5,021	0	0	0	0	0	0	0	5,021
København NV	2,726	531	224	0	0	0	0	0	3,481
København S	0	0	0	10,893	0	0	0	43	10,936
København SV	267	1,073	0	181	0	0	0	0	1,521
København V	2,818	0	0	0	0	0	0	0	2,818
København Ø	7,419	0	36	0	0	0	0	0	7,455
Rødovre	0	2,260	0	0	904	607	0	0	3,771
Skovlunde	0	0	0	0	0	758	0	0	758
Søborg	0	44	2,087	0	0	316	0	0	2,447
Taastrup	0	0	0	0	0	0	3,273	0	3,273
Valby	0	4,147	0	0	371	0	0	0	4,518
Vallensbæk	0	0	0	0	9	0	600	0	609
Vallensbæk Str.	0	0	0	0	661	0	929	0	1,590
Vanløse	0	3,695	0	0	25	11	0	0	3,731
Virum	0	0	994	0	0	1,238	0	0	2,232
Værløse	0	0	0	0	0	15	0	0	15
Total	26,903	18,933	14,646	13,830	7,681	8,881	6,771	3,415	101,060