

## Hierarhilise klasterdamise ja partitioning klasterdamise kolme klasteralgoritmi võrdlus (oma 4x3 hoone puhul)

Kasutasin ehv andmete ja 4. ül admete (kuna pikkus ja laius oli kahel hoonetel puudu) põhjal niisuguse andmestiku:

EHV	MaksKorrust	EhitisealunePind	Kõrgus	Pikkus	Laius	LiftideArv	Aasta
120801759	2	650.4	7.1	52.4	11.1	0	2018
120801756	2	650.4	7.1	52.4	11.1	0	2018
120801732	2	650.4	7.1	52.4	11.1	0	2018
120847092	2	433.1	7.2	37.7	11.4	0	2019
120847091	2	433.1	7.2	37.7	11.4	0	2019
120847090	2	433.1	7.2	37.7	11.4	0	2019
101019312	9	598	0	46.8	13.3	2	1974
101014868	9	599	29.8	46.9	13.3	2	1974
101019344	9	674	0	46.2	13.4	2	1974
120857076	12	521.3	38.8	28.3	21.2	2	0
120791952	14	521.3	44.8	29.5	21.2	2	2020
120857095	12	521.3	38.8	28.3	21.2	2	2023

### Meetod K-means:

Eksperiment 1, k=2:

Käivatsin koodi pythonis

Cluster 1 Buildings: [120801759, 120801756, 120801732, 120847092, 120847091, 120847090, 101019312, 101014868, 101019344, 120791952, 120857095] Average Values: MaksKorrust: 5.91 EhitisealunePind: 560.37 Kõrgus: 14.21 Pikkus: 42.55 Laius: 13.63 LiftideArv: 0.91 Aasta: 2006.91	Cluster 2 Buildings: [120857076] Average Values: MaksKorrust: 12.00 EhitisealunePind: 521.30 Kõrgus: 38.80 Pikkus: 28.30 Laius: 21.20 LiftideArv: 2.00 Aasta: 0.00
---	---

On tunne, et 120857076 ehitis on teises klastris, kuna tal puudub aasta. Uurisin teistes allikates, mis aasta see oli. Leidsin <https://kinnisvara24.ee/korter-myyk-tallinn/240408025> põhjal, et ehitusaasta on 2020, seega kasutan seda andmestikkus hiljem.

### Eksperiment 1.2, k=2:

Cluster 1 Buildings: [120847092, 120847091, 120847090, 120857076, 120791952, 120857095] Average Values: MaksKorrust: 7.33 EhitisealunePind: 477.20 Kõrgus: 24.00 Pikkus: 33.20 Laius: 16.30 LiftideArv: 1.00 Aasta: 2020.00	Cluster 2 Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344] Average Values: MaksKorrust: 5.50 EhitisealunePind: 637.03 Kõrgus: 8.52 Pikkus: 49.52 Laius: 12.22 LiftideArv: 1.00 Aasta: 1996.00
--	---

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on keskmiselt umbes 7 korrust, kompaktsema ehitisealuse pinnaga, kuid kõrgemate keskmise kõrguse ja suhteliselt suure laiusega. Need hooned on ehitatud hiljem.
- Cluster 2 hooned on keskmiselt umbes 5.5 korrust, suurema ehitisealuse pinnaga, kuid madalama keskmise kõrguse ja suhteliselt pikema pikkusega. Need hooned on ehitatud varasemalt.

### Eksperiment 2, k=3:

Cluster 1 Buildings: [120857076, 120791952, 120857095]  Average Values: MaksKorrust: 12.67 EhitisealunePind: 521.30 Kõrgus: 40.80 Pikkus: 28.70 Laius: 21.20 LiftideArv: 2.00 Aasta: 2021.00	Cluster 2 Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344] Average Values: MaksKorrust: 5.50 EhitisealunePind: 637.03 Kõrgus: 8.52 Pikkus: 49.52 Laius: 12.22 LiftideArv: 1.00 Aasta: 1996.00	Cluster 3 Buildings: [120847092, 120847091, 120847090]  Average Values: MaksKorrust: 2.00 EhitisealunePind: 433.10 Kõrgus: 7.20 Pikkus: 37.70 Laius: 11.40 LiftideArv: 0.00 Aasta: 2019.00
--	---	--

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on keskmiselt kõrgemad kui teised hooned, ning keskmiselt palju kõrgem kui teised klastrid, kuid keskmiselt pikkus on kõige väiksem. Need hooned on ehitatud hiljem, kui teised, keskmiselt aasta 2021.
- Cluster 2 hooned on keskmiselt 5.50 korrust kõrged, kuuld keskmiselt kõige suurema ehitusaluse pinnaga ning pikemad. Need hooned on ehitatud varasemalt, aasta 1996.
- Cluster 3 hooned on keskmiselt 2.00 korrust kõrged, kõige vähema kõrgusega ning ehitisealuse pinnaga. Need hooned on ehitatud hiljem, aasta 2019, ning neil puudub lift.

### Eksperiment 3, k=4:

Cluster 1 Buildings: [120801759, 120801756, 120801732, 101019344] Average Values: MaksKorrust: 3.75 EhitisealunePind: 656.30 Kõrgus: 5.32 Pikkus: 50.85 Laius: 11.67 LiftideArv: 0.50 Aasta: 2007.00	Cluster 2 Buildings: [120857076, 120791952, 120857095] Average Values: MaksKorrust: 12.67 EhitisealunePind: 521.30 Kõrgus: 40.80 Pikkus: 28.70 Laius: 21.20 LiftideArv: 2.00 Aasta: 2021.00	Cluster 3 Buildings: [120847092, 120847091, 120847090] Average Values: MaksKorrust: 2.00 EhitisealunePind: 433.10 Kõrgus: 7.20 Pikkus: 37.70 Laius: 11.40 LiftideArv: 0.00 Aasta: 2019.00	Cluster 4 Buildings: [101019312, 101014868] Average Values: MaksKorrust: 9.00 EhitisealunePind: 598.50 Kõrgus: 14.90 Pikkus: 46.85 Laius: 13.30 LiftideArv: 2.00 Aasta: 1974.00
---	---	---	--

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on omavahel sarnased, kuna neil on madal keskmine kõrgus (umbes 5.32 meetrit), kõige suurem ehitisealune pind (umbes 656.30 m<sup>2</sup>) ja nad on ehitatud hiljem (keskmiselt aastal 2007). Neile on iseloomulik madal liftide arv (keskmiselt 0.50) võrreldes teiste klastritega.
- Cluster 2 hooned erinevad teistest klastritest kõrge keskmise kõrguse (umbes 40.80 meetrit) ja suure laiuse (umbes 21.20 meetrit) poolest. Need hooned on ehitatud hiljuti (aastal 2021) ning neil on kõrge liftide arv (keskmiselt 2).
- Cluster 3 hooned on samuti omavahel sarnased madala keskmise kõrguse (umbes 7.20 meetrit) ja väikese ehitisealuse pinna (umbes 433.10 m<sup>2</sup>) poolest. Need hooned on ehitatud suhteliselt hiljuti (aastal 2019) ning neil puudub täielikult lift.
- Cluster 4 hooned eristuvad teistest kõrge keskmise kõrguse (umbes 14.90 meetrit) ja suure ehitisealuse pinna (umbes 598.50 m<sup>2</sup>) poolest. Need hooned on ehitatud varakult (aastal 1974) ning neil on keskmiselt 2 lifti.

## Meetod 2 spectral clustering:

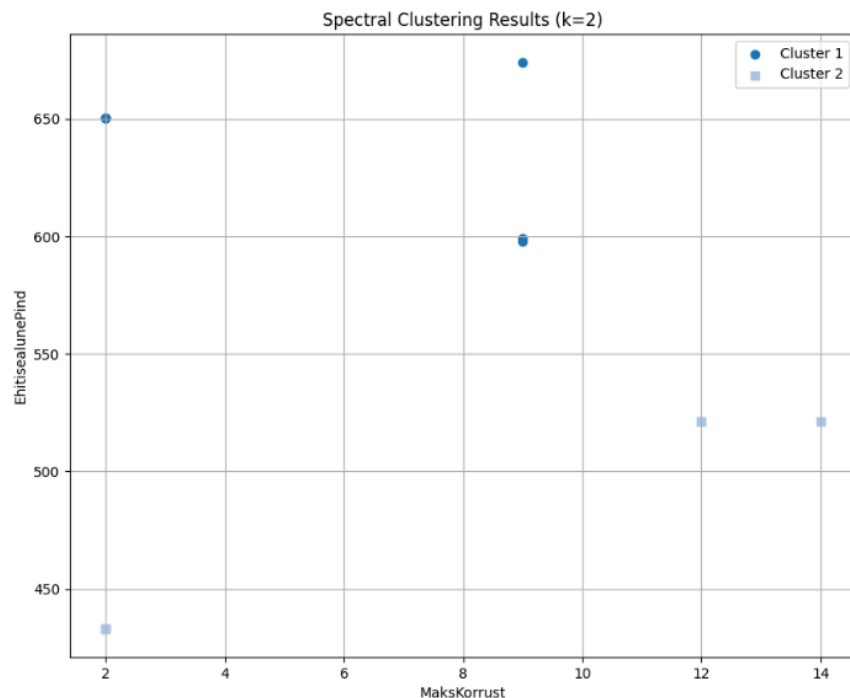
Miks võtsin: vaatasin internetis mis see on ja mulle meeldis see fakt, et sellel meetodil on graafipõhine lähenemine. Muutsin natukene antud pythoni skripti, et visualiseerida andmed.

### Eksperiment 4, k=2:

<p>Cluster 1</p> <p>Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344]</p> <p>Average Values:</p> <p>MaksKorrust: 5.50</p> <p>EhitisealunePind: 637.03</p> <p>Kõrgus: 8.52</p> <p>Pikkus: 49.52</p> <p>Laius: 12.22</p> <p>LiftideArv: 1.00</p> <p>Aasta: 1996.00</p>	<p>Cluster 2</p> <p>Buildings: [120847092, 120847091, 120847090, 120857076, 120791952, 120857095]</p> <p>Average Values:</p> <p>MaksKorrust: 7.33</p> <p>EhitisealunePind: 477.20</p> <p>Kõrgus: 24.00</p> <p>Pikkus: 33.20</p> <p>Laius: 16.30</p> <p>LiftideArv: 1.00</p> <p>Aasta: 2020.00</p>
--	---

Klastrite põhjal võib  
järeldata, et:

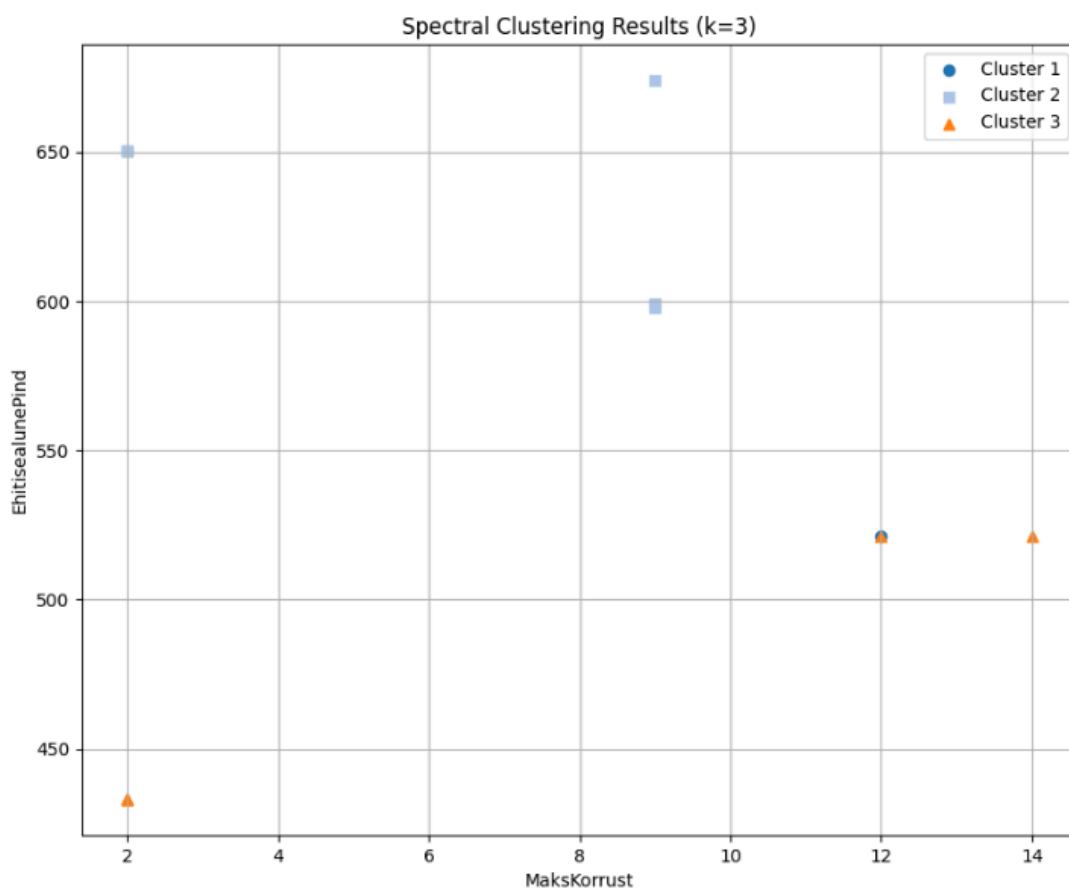
- Cluster 1 hooned on keskmiselt umbes 5.5 korrust, suurema ehitisealuse pinnaga, kuid madalama keskmise kõrguse ja suhteliselt pikema pikkusega. Need hooned on ehitatud varasemalt.
- Cluster 2 hooned on keskmiselt umbes 7 korrust, kompaktsema ehitisealuse pinnaga, kuid kõrgemate keskmise kõrguse ja suhteliselt suure laiusega. Need hooned on ehitatud suhteliselt hiljem.



Klastrid on samasugused nagu meetodis K-means.

### Eksperiment 5, k=3:

<p>Cluster 1</p> <p>Buildings: [120857076]</p> <p>Average Values:</p> <p>MaksKorrust: 12.00</p> <p>EhitisealunePind: 521.30</p> <p>Kõrgus: 38.80</p> <p>Pikkus: 28.30</p> <p>Laius: 21.20</p> <p>LiftideArv: 2.00</p> <p>Aasta: 2020.00</p>	<p>Cluster 2</p> <p>Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344]</p> <p>Average Values:</p> <p>MaksKorrust: 5.50</p> <p>EhitisealunePind: 637.03</p> <p>Kõrgus: 8.52</p> <p>Pikkus: 49.52</p> <p>Laius: 12.22</p> <p>LiftideArv: 1.00</p> <p>Aasta: 1996.00</p>	<p>Cluster 3</p> <p>Buildings: [120847092, 120847091, 120847090, 120791952, 120857095]</p> <p>Average Values:</p> <p>MaksKorrust: 6.40</p> <p>EhitisealunePind: 468.38</p> <p>Kõrgus: 21.04</p> <p>Pikkus: 34.18</p> <p>Laius: 15.32</p> <p>LiftideArv: 0.80</p> <p>Aasta: 2020.00</p>
---	--	--



Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned selles klastris on keskmiselt 12 korrust kõrged, nende ehitisealune pind on umbes 521.30 m<sup>2</sup> ning keskmine kõrgus on ligikaudu 38.80 meetrit. Need hooned on ehitatud aastal 2020 ja neil on keskmiselt 2 lifti.
- Cluster 2 hooned, millel on keskmiselt 5.50 korrust, ehitisealune pind ligikaudu 637.03 m<sup>2</sup> ning keskmine kõrgus umbes 8.52 meetrit. Hooned selles klastris on ehitatud aastatel 1996. Need

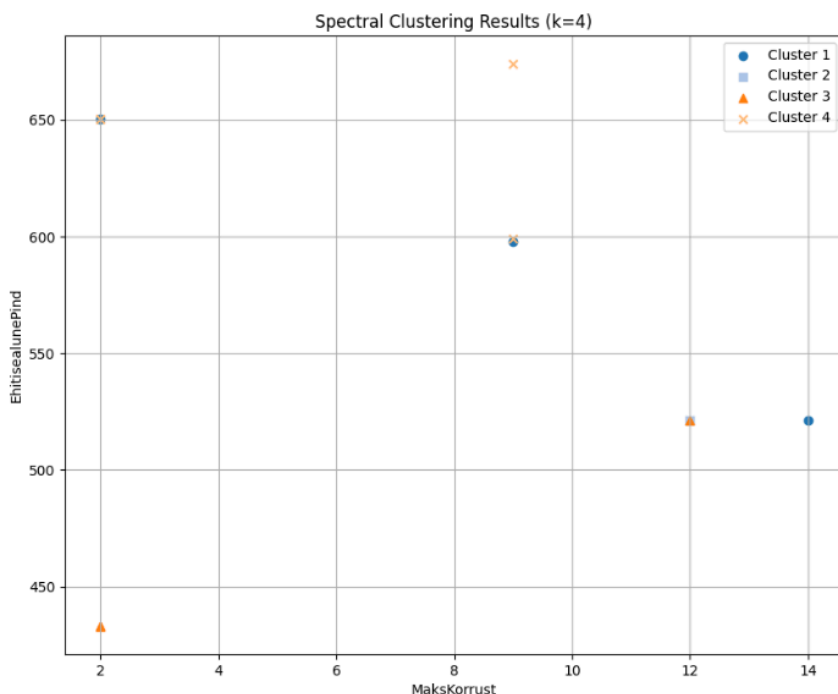
hooned on suhteliselt pikad (keskmine pikkus 49.52 meetrit) ja kitsad (keskmine laius 12.22 meetrit). Liftide arv keskmiselt 1.

- Cluster 3 hooned on keskmiselt 6.40 korrust kõrged, ehitisealune pind on ligikaudu 468.38 m<sup>2</sup> ja keskmine kõrgus umbes 21.04 meetrit. Need hooned on ehitatud aastal 2020. Keskmine pikkus on 34.18 meetrit ja keskmine laius 15.32 meetrit. Liftide arv keskmiselt 0.80.

Väga huvitav, et see meetod eraldas klastrisse 1 ainult ühe hoone 120857076, mis on tegelikult peaaegu identne 120857095 hoonega, kuid ainult aasta erinev. Näen, et Cluster 2 võttis hooned suurema pinnaga ja need, mis on olid ehitatud varasemalt, kui klustris 3.

#### Eksperiment 6, k=4:

<p>Cluster 1 Buildings: [120801759, 101019312, 120791952]</p> <p>Average Values: MaksKorrust: 8.33 EhitisealunePind: 589.90 Kõrgus: 17.30 Pikkus: 42.90 Laius: 15.20 LiftideArv: 1.33 Aasta: 2004.00</p>	<p>Cluster 2 Buildings: [120857076]</p> <p>Average Values: MaksKorrust: 12.00 EhitisealunePind: 521.30 Kõrgus: 38.80 Pikkus: 28.30 Laius: 21.20 LiftideArv: 2.00 Aasta: 2020.00</p>	<p>Cluster 3 Buildings: [120847092, 120847091, 120847090, 120857095]</p> <p>Average Values: MaksKorrust: 4.50 EhitisealunePind: 455.15 Kõrgus: 15.10 Pikkus: 35.35 Laius: 13.85 LiftideArv: 0.50 Aasta: 2020.00</p>	<p>Cluster 4 Buildings: [120801756, 120801732, 101014868, 101019344]</p> <p>Average Values: MaksKorrust: 5.50 EhitisealunePind: 643.45 Kõrgus: 11.00 Pikkus: 49.47 Laius: 12.22 LiftideArv: 1.00 Aasta: 1996.00</p>
--	---	---	---



- Cluster 1 hooned on sarnased keskmise kõrguse (umbes 17.30 meetrit) ja suhteliselt suure ehitisealuse pinna (umbes 589.90 m<sup>2</sup>) poolest. Need hooned on ehitatud aastal 2004 ning neil on keskmine liftide arv (umbes 1.33) võrreldes teiste klastritega.

- Cluster 2 hooned erinevad teistest klastritest kõrge keskmise kõrguse (umbes 38.80 meetrit) ja suure laiuse (umbes 21.20 meetrit) poolest. Need hooned on ehitatud aastal 2020 ning neil on kõrge liftide arv (umbes 2).

- Cluster 3 hooned on omavahel sarnased madala keskmise kõrguse (umbes 15.10 meetrit) ja väikese ehitisealuse pinna (umbes 455.15 m<sup>2</sup>) poolest. Need hooned on samuti

ehitatud aastal 2020, kuid neil puudub täielikult lift.

- Cluster 4 hooned eristuvad teistest kõrge keskmise kõrguse (umbes 11.00 meetrit) ja suure ehitisealuse pinna (umbes 643.45 m<sup>2</sup>) poolest. Need hooned on ehitatud varakult, aastal 1996, ning neil on keskmine liftide arv (umbes 1).

Kui vaadata neid klastreid ja hooneomadusi, võib märgata järgmisi seaduspärasusi. Klastreid eristab oluliselt keskmine kõrgus ja ehitisealune pind. Näiteks Cluster 2-s on hooned märkimisväärselt kõrgemad ja suurema ehitisealuse pinnaga võrreldes teiste klastritega. Cluster 3-s on hoone keskmine kõrgus märgatavalt madalam, samuti on ka ehitisealune pind väiksem. Hoone ehitamise aastal on samuti märgatav erinevus. Näiteks Cluster 4 hooned on ehitatud varakult, aastal 1996, samas kui Cluster 2 hooned on hiljutisemad, ehitatud aastal 2020. Liftide arv erineb ka oluliselt erinevate klastrihoonete vahel. Näiteks Cluster 2-s on keskmiselt rohkem lifte (2), samas kui Cluster 3-s on lifte üldse puudunud.

### **Meetod 3 agglomerative clustering:**

Miks võtsin: on kirjas, et see on hierarhiline klasterdamine, seega tahav vaadata, kuidas see erineb tavalisest hierarhilise klasterdamisest. Samuti vaatasin internetis, et selle kood on tavaliselt suure keerukusega. On lihtsalt huvi vaadata kuidas see toimib.

#### Eksperiment 7, k=2:

Cluster 1 Buildings: [120847092, 120847091, 120847090, 120857076, 120791952, 120857095] Average Values: MaksKorrust: 7.33 EhitisealunePind: 477.20 Kõrgus: 24.00 Pikkus: 33.20 Laius: 16.30 LiftideArv: 1.00 Aasta: 2020.00	Cluster 2 Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344] Average Values: MaksKorrust: 5.50 EhitisealunePind: 637.03 Kõrgus: 8.52 Pikkus: 49.52 Laius: 12.22 LiftideArv: 1.00 Aasta: 1996.00
--	---

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on keskmiselt umbes 7 korrust, kompaktsema ehitisealuse pinnaga, kuid kõrgemate keskmise kõrguse ja suhteliselt suure laiusega. Need hooned on ehitatud hiljem.
- Cluster 2 hooned on keskmiselt umbes 5.5 korrust, suurema ehitisealuse pinnaga, kuid madalama keskmise kõrguse ja suhteliselt pikema pikkusega. Need hooned on ehitatud varasemalt.

Klastrid on samasugusud nagu meetodis K-means.

### Eksperiment 8, k=3:

Cluster 1 Buildings: [120801759, 120801756, 120801732, 101019312, 101014868, 101019344] Average Values: MaksKorrust: 5.50 EhitisealunePind: 637.03 Kõrgus: 8.52 Pikkus: 49.52 Laius: 12.22 LiftideArv: 1.00 Aasta: 1996.00	Cluster 2 Buildings: [120857076, 120791952, 120857095]  Average Values: MaksKorrust: 12.67 EhitisealunePind: 521.30 Kõrgus: 40.80 Pikkus: 28.70 Laius: 21.20 LiftideArv: 2.00 Aasta: 2021.00	Cluster 3 Buildings: [120847092, 120847091, 120847090]  Average Values: MaksKorrust: 2.00 EhitisealunePind: 433.10 Kõrgus: 7.20 Pikkus: 37.70 Laius: 11.40 LiftideArv: 0.00 Aasta: 2019.00
--	---	---

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on omavahel sarnased, kuna nendel on keskmine kõrgus umbes 8.52 meetrit, suurema ehitisealuse pinnaga, umbes 637.03 ruutmeetrit ning need on ehitatud varasemalt, aastal 1996.
- Cluster 2 hooned eristuvad teistest kõrgema keskmise korruse (umbes 12.67 korrust) ja suure ehitisealuse pinna (umbes 521.30 ruutmeetrit) poolest. Need hooned on suhteliselt uued, ehitatud aastal 2021, ning neil on keskmiselt 2 lifti.
- Cluster 3 hooned erinevad teistest väiksema keskmise kõrguse (umbes 7.20 meetrit) ja ehitisealuse pinna (umbes 433.10 ruutmeetrit) poolest. Need hooned on ehitatud aastal 2019 ja neil puudub täielikult lift.

Klastrid on samasugused nagu meetodis K-means.

Vaadates neid kolme klastrit ja nende hoonete omadusi, võib välja tuua mõningaid seaduspärasusi. Cluster 2 (keskmine kõrgus 40.80 meetrit ja ehitisealune pind 521.30 ruutmeetrit) erineb oluliselt teistest klastritest, kus keskmine kõrgus on väiksem (Cluster 1) või palju väiksem (Cluster 3). Cluster 1 (keskmine kõrgus 8.52 meetrit ja ehitisealune pind 637.03 ruutmeetrit) erineb samuti teistest klastritest, kus keskmine kõrgus on kas kõrgem (Cluster 2) või madalam (Cluster 3), kuid ehitisealune pind on suurem. Cluster 2 (ehitatud aastal 2021) on uusimate hoonete klastrina kõige hilisem. Cluster 3 (ehitatud aastal 2019) ja Cluster 1 (ehitatud aastal 1996) sisaldavad vanemaid ehitisi. Cluster 2 (keskmine liftide arv 2.00) eristub teistest klastritest, kus liftide arv on kas madalam (Cluster 1) või puudub täielikult (Cluster 3).



#### Eksperiment 9, k=4:

Cluster 1 Buildings: [120801759, 120801756, 120801732, 101019344] Average Values: MaksKorrust: 3.75 EhitisealunePind: 656.30 Kõrgus: 5.32 Pikkus: 50.85 Laius: 11.67 LiftideArv: 0.50 Aasta: 2007.00	Cluster 2 Buildings: [101019312, 101014868] Average Values: MaksKorrust: 9.00 EhitisealunePind: 598.50 Kõrgus: 14.90 Pikkus: 46.85 Laius: 13.30 LiftideArv: 2.00 Aasta: 1974.00	Cluster 3 Buildings: [120847092, 120847091, 120847090] Average Values: MaksKorrust: 2.00 EhitisealunePind: 433.10 Kõrgus: 7.20 Pikkus: 37.70 Laius: 11.40 LiftideArv: 0.00 Aasta: 2019.00	Cluster 4 Buildings: [120857076, 120791952, 120857095] Average Values: MaksKorrust: 12.67 EhitisealunePind: 521.30 Kõrgus: 40.80 Pikkus: 28.70 Laius: 21.20 LiftideArv: 2.00 Aasta: 2021.00
---	--	---	---

Klastrid on samasugusud nagu meetodis K-means.

Klastrite põhjal võib järeldada, et:

- Cluster 1 hooned on madalama kõrgusega ja suurema ehitisealuse pinnaga. Neil on vähem liftide arv võrreldes teiste klastritega. Hooned on ehitatud suhteliselt hiljuti, aastal 2007.
- Cluster 2 sisaldab kõrgemaid hooneid, millel on keskmiselt rohkem korruseid ja suurem ehitisealune pind. Hooneid iseloomustab ka suurem liftide arv. Need hooned on ehitatud varasemalt, aastal 1974.
- Cluster 3 hooned on madalama kõrguse ja väiksema ehitisealuse pinnaga ning neil puudub täielikult lift. Hooned on suhteliselt hiljuti ehitatud, aastal 2019.
- Cluster 4 sisaldab kõrgeid hooneid, millel on kõige kõrgem keskmine kõrgus ja suhteliselt suur ehitisealune pind. Hooneid iseloomustab ka rohkem liftide arv. Need hooned on ehitatud hiljuti, aastal 2021. On sarnane klastriga 2.

#### **Mis sarnasusi ja erinevusi sa märkad erinevate klastrite arvu puhul?**

Tavaliselt klastrid, kus on rohkem elemente, jagunesid teiste klastritesse, mis võib muuta klastrite jaotust detailsemaks ja rohkemateks gruppideks ja eristada erinevaid omadusi ja mustreid. Erinevus aga seisnes selles, et kui klastrite arvu suurendada, võib see viia üleklasterdamiseni (overfitting), kus iga element kuulub omaenda klastrisse, ilma selge struktuurita või mustrita (spectral clustering). Selle probleemiga toimusid palju parem k-means ja agglomerative clustering.

#### **Mis sarnasusi ja erinevusi sa märkad erinevate meetodite vahel?**

Meetod k-means ja agglomerative clustering toimusid väga sarnaselt, e. tekkisid samasugused klastrid, kuid võib olla teises järjekorras. Kolme meetodite sarnasused olid suurema arvu klastrite kasutamisel võib anda detailsema jaotuse, võimaldades eristada erinevaid alarühmi või kategooriaid. Spectral clustering, nagu mainitud, võib suurema klastrite arvu korral kalduvus olla üleklasterdamisele, kus iga element kuulub omaenda klastrisse ilma selge struktuuri või mustrita. K-means ja agglomerative clustering näitavad paremat vastupidavust üleklasterdamisele, kuna need meetodid põhinevad erinevatel algoritmilistel lähenemisviisidel ja klastrikeskuste või hierarhilise struktuuri määramisel.

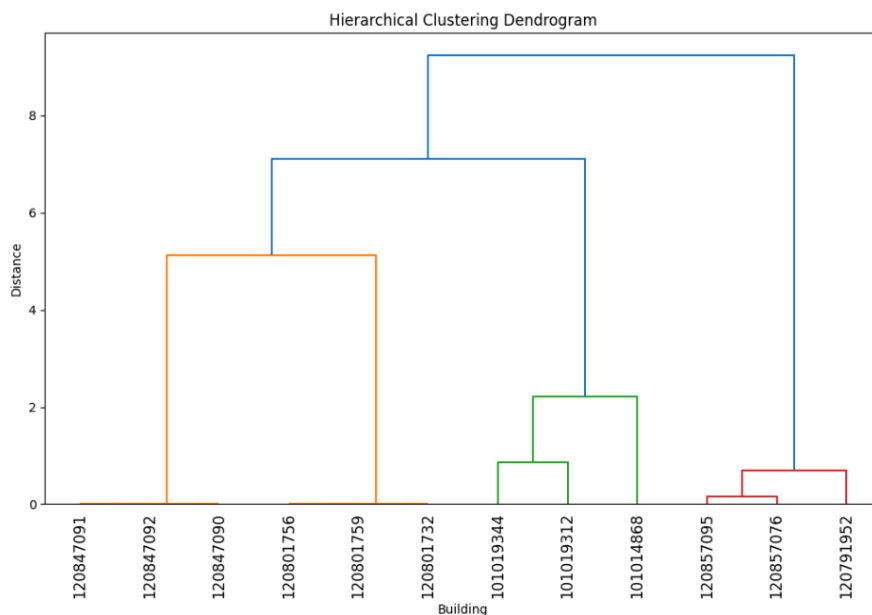
## Meetod hierarhiline klasterdamine:

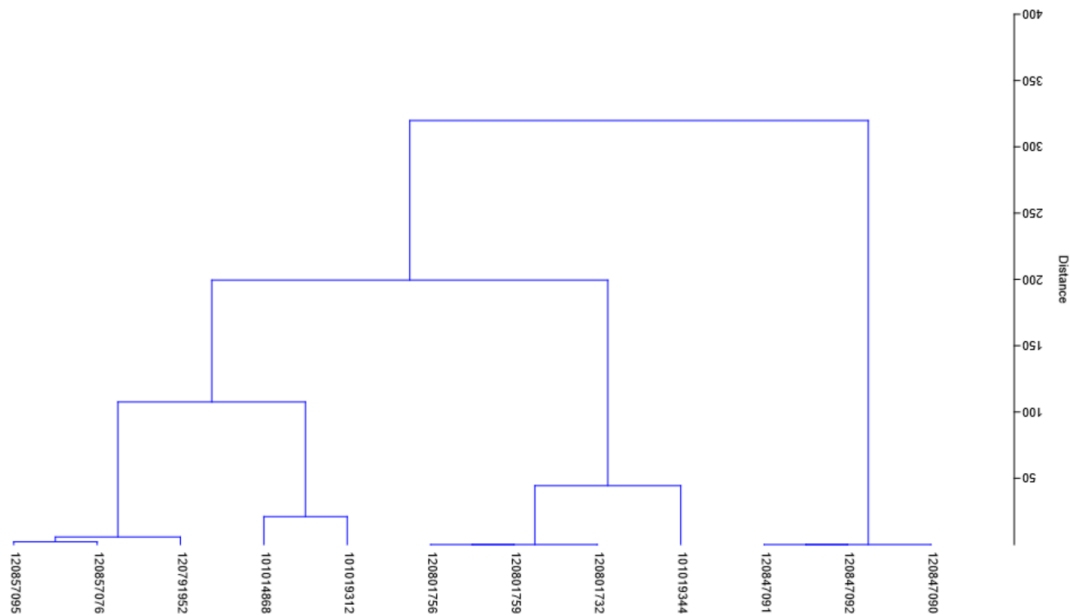
### Eksperiment 10:

Cluster 1 Buildings: [120801759, 120801756, 120801732, 120847092, 120847091, 120847090] Average Values: MaksKorrust: 2.00 EhitisealunePind: 541.75 Kõrgus: 7.15 Pikkus: 45.05 Laius: 11.25 LiftideArv: 0.00 Aasta: 2018.50	Cluster 2 Buildings: [120857076, 120791952, 120857095]  Average Values: MaksKorrust: 12.67 EhitisealunePind: 521.30 Kõrgus: 40.80 Pikkus: 28.70 Laius: 21.20 LiftideArv: 2.00 Aasta: 2021.00	Cluster 3 Buildings: [101019312, 101014868, 101019344]  Average Values: MaksKorrust: 9.00 EhitisealunePind: 623.67 Kõrgus: 9.93 Pikkus: 46.63 Laius: 13.33 LiftideArv: 2.00 Aasta: 1974.00
--	---	---

### **Kuidas hierarhilise klasterdamise tulemus on sarnane ja erineb eelnevast üheksast eksperimentidist?**

On sarnane k-means ja agglomerative clustering meetodiga. Kõik need meetodid jagavad andmeid klastritesse vastavalt nende omadustele. Hierarchical clustering, nagu ka k-means ja agglomerative clustering, kasutab tunnuste sarnasust andmete jagamiseks klastritesse. On hästi näha, et hierarchical clustering kasutab korruste arvu, kõrgust, laiust, liftide arvu ja ehitamise aastat. Kui andmetel on palju sarnaseid tunnuseid, paneb kõik need meetodid need tavaliselt samasse klastrisse. Kuid aga hierarchical clustering ei nõua etteantud klastrite arvu. See moodustab puu struktuuri, millest saab valida sobiva klastrite arvu, lõigates puud soovitud kõrguselt. K-means ja agglomerative clustering nõuavad eelnevat määratletud klastrite arvu. Vaadates kõik 10 eksperimente, on milu meelet hierarhiline klasterdamine palju arusaadav, miks just need hooned pandi niisuguse klatri sise.





### **Kuidas hierarhilise klasterdamise tulemus on sarnane ja erineb PAST ülesandes tehtud hierarhilisest klasterdamisest?**

On mitu hooneid, mis on sarnased tulemustes: 120857076, 120791952, 120857095. Omavahel nad paiknevad kõrval. Kuid aga koodi tulemus pani neid teistest kaugemale, PAST tulemus aga minu meelest on parem neid klasterdanud, kuna saab öelda, et hoonete 101019312, 101019344 ja 101014868 tunnused on sarnased nende tunnustustega. Hoonete 120847092, 120847091, 120847090 tulemused ka erinevad. Võiks öelda, et PAST vaatas pigem korruste arvu ja pinna järgi, ning kood kõrguse järgi, mis võiks ka seletada, miks hooned 120857076, 120791952, 120857095 paiknevad nii kaugel. Minu andmetes 101019312, 101019344 hoonedel on puudu andmed kõrguse kohta, kuid korruste arv on ka suur, võrreldes teiste hoonetega. Seega võiks järeldada, et kui mul oleks olnud nende kõrgus, siis koodi tulemus võiks olnud olla sarnasem PAST'iga.

Hoonete 120847092, 120847091, 120847090 pind on palju väiksem võrreldes teiste hoonetega, seega on näha miks PAST tulemuses nad on kõige kaugemad. Kokkuvõtteks ütlen, et PASTi tulemus tundub usaldusväärsem, mida ma kasutaks rohkem.