

## **BUSINESS INTELLIGENCE INOM AVFALLSHANTERING**

---

Effektivisera sophämtning genom att analysera avvikelserapporter

## **BUSINESS INTELLIGENCE IN WASTE MANGEMENT**

---

Garbage collection efficiency by analyzing error reports

Examensarbete inom informationssystemutveckling  
Grundnivå 22,5 högskolepoäng  
Vårtermin 2014

Mikael Staf

Handledare: Kristens Gudfinnsson  
Examinator: Eva Söderström

## Sammanfattning

I stort sett alla hushåll och företag använder sig av sopor och vi är beroende att det sköts på ett korrekt sätt. Avfallshantering i Östra Skaraborg rapporterar idag in avvikelser på ungefär var tionde soptömning vilket förhindrar effektiviteten. Denna studie handlar om att analysera dessa rapporter för att kunna effektivisera sophantering och på så vis minska antalet avvikelser. Med Business Intelligence (BI) har det skapats en relationsdatabas för att kunna lagra intern och extern data på ett ställe och utföra analyser på dessa. En stor del i studien är extrahering, transformering och inladdning (ETL) där fokuset legat på transformeringen för att få så bra datakvalité som möjligt. Resultatet visar att ingen generell typ av kund står för alla avvikelser utan olika typer av kunder står för olika typer av avvikelser. Utifrån resultatet har det tagits fram beslutsunderlag vilket ska hjälpa till att kunna effektivisera sophantering inom Östra Skaraborg.

**Nyckelord:** Business Intelligence, ETL, databas, informationssystem, avfallshantering

## **Förord**

Jag vill tacka alla inblandade parter i mitt examensarbete, Högskolan i Skövde och Avfallshantering Östra Skaraborg vilket jag har utfört studien med, min handledare Kristens Gudfinnsson som väglett mig genom arbetet, och examinatorn Eva Söderström vilket också har givit förslag på förbättringar. Vidare vill jag tacka intervjudeltagare samt vänner och familj som givit stöd till att utföra denna studie.

# Innehållsförteckning

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduktion</b>                     | <b>1</b>  |
| <b>2</b> | <b>Bakgrund</b>                         | <b>3</b>  |
| 2.1      | Business Intelligence                   | 3         |
| 2.1.1    | Geographical Information System         | 4         |
| 2.1.2    | ETL                                     | 5         |
| 2.1.3    | Datakvalité                             | 6         |
| 2.1.4    | Sammanfattning av Business Intelligence | 7         |
| 2.2      | Kommunal avfallshantering               | 7         |
| <b>3</b> | <b>Problembeskrivning</b>               | <b>9</b>  |
| 3.1      | Avgränsning                             | 10        |
| 3.2      | Förväntat resultat                      | 10        |
| <b>4</b> | <b>Metod</b>                            | <b>11</b> |
| 4.1      | Insamling av data                       | 11        |
| 4.1.1    | Intervjuer                              | 11        |
| 4.1.2    | Enkätundersökning                       | 12        |
| 4.1.3    | Skapande av frågor                      | 13        |
| 4.1.4    | Etiska aspekter                         | 13        |
| 4.2      | Metod för datatvättning                 | 13        |
| 4.3      | Metod för analysering                   | 15        |
| 4.4      | Val av verktyg                          | 15        |
| 4.5      | Genomförande                            | 16        |
| 4.5.1    | Intervju och enkäter                    | 16        |
| 4.5.2    | Datatvättning                           | 17        |
| <b>5</b> | <b>Empiri</b>                           | <b>18</b> |
| 5.1      | Enkäter och intervjuer                  | 18        |
| 5.1.1    | Projektansvarig på Högskolan i Skövde   | 18        |
| 5.1.2    | Sopbilschaufför                         | 18        |
| 5.2      | Databas                                 | 19        |
| 5.2.1    | Kärlbeskrivning                         | 23        |
| 5.2.2    | Väder                                   | 23        |
| <b>6</b> | <b>Analys</b>                           | <b>25</b> |
| 6.1      | Utförande av datatvättning              | 25        |
| 6.1.1    | Dataanalys                              | 25        |
| 6.1.2    | Verifiera                               | 28        |
| 6.1.3    | Transformerering                        | 28        |
| 6.1.4    | Ändra äldre data                        | 28        |
| 6.2      | Analys av avvikelser                    | 29        |
| 6.2.1    | Trasigt kärl                            | 29        |
| 6.2.2    | Felsorterad                             | 30        |
| 6.2.3    | Annan orsak                             | 31        |
| 6.2.4    | Fel kärlstorlek                         | 32        |
| 6.2.5    | Felaktigt stopp                         | 32        |
| 6.2.6    | Blockerat                               | 33        |
| 6.2.7    | Extra säckar                            | 34        |
| 6.2.8    | Ej utställt                             | 34        |
| 6.2.9    | Överfullt kärl                          | 35        |

|   |           |
|---|-----------|
| 6.2.10 Sammanfattning av avvikelser.....                | 35        |
| <b>7 Slutsats.....</b>                                  | <b>37</b> |
| <b>8 Diskussion.....</b>                                | <b>39</b> |
| 8.1 Forskningsmetod.....                                | 40        |
| 8.2 Etiska, samhällliga och vetenskapliga aspekter..... | 41        |
| 8.3 Framtida arbete.....                                | 41        |
| <b>Referenser.....</b>                                  | <b>42</b> |

# 1 Introduktion

Tekniken blir allt mer avancerad dag för dag vilket leder till att mer och mer data skapas. Idag finns det över 2,5 kvintijoner bytes av data, varav 90% av den datan skapades under de senaste två åren (IBM, 2013). Denna data ses som tillgång till information vilket det inte går att ha för mycket av så länge den hanteras på ett korrekt sätt (Trujullo och Mora, 2003).

Business Intelligence (BI) är ett samlingsbegrepp vilket går ut på att samla in, integrera, analysera och presentera stora volymer av data för att kunna göra bättre beslut (Dayal et al., 2009). Extraction-Transformation-Loading (ETL) är en fundamental aktivitet i BI vilket går ut på att hämta data från en eller flera källsystem, ändra den till ett lämpligt format, och till sist för att sedan ladda in det till en databas (Simitsis, et al., 2005). För att få bra kvalitet på data behövs datatvättning (en. Data Cleaning). Detta är en process för att se till att data blir relevant inom arbetet, vilket sker i ETL-processen. Rahm och Do (2000) skriver att datatvättningen inte har fått den uppmärksamhet inom forskningen som den behöver, och på så vis räknas det som det största problemet inom relationsdatabaser.

Avfallshanteringsbranschen är ett område vilket vi stöter på i vår vardag över stora delar av planeten. Hushållen har soptunnor, sorterar avfall och slänger det på återvinningsstationer. Med hjälp av IT kan en effektivare sophantering ske via diverse optimeringar vilket även kan påverka resultatet globalt.

Inom avfallshanteringsbranschen används Business Intelligence mer och mer under namnet Geographical Information System (GIS), vilket översätts till "Geografiskt informationssystem". GIS är ett system som tillåter en att hämta, lagra, integrera och analysera data utifrån geografiska positioner (Rada, et al., 2013). Ett sådant system omfattar flera olika tekniker som använder sig av geografiska system, till exempel GPS och radardata, men även icke-geografisk data kan lagras. Det har skett flera olika fallstudier där resultatet visar att ett GIS hjälper till att minska utsläppen, bland annat inom sophanteringsbranschen (Zsigraiova, et al., 2013). Zsigraiova, et al. (2013) gjorde en fallstudie för att mäta hur GIS kan förbättra ruttoptimeringen inom sophantering där resultatet ledde till 40 % minskat utsläpp och 57 % i kostnadsbesparingar. I deras resultat tas inte problemet upp angående om sopbilarna inte kommer fram eller inte har möjligheten att tömma soporna som exempel, vilket denna studie gör.

Avfallshantering Östra Skaraborg använder ett GIS som är i bruk men fortfarande under utveckling. Deras GIS lagrar information om vilka tunnor som inte har kunnat tömmas genom att sopbilschaufförerna rapporterar in avvikelser hos hushållen. I dessa rapporter lagras information om avvikelserna för att sedan laddas till en databas, men i dagsläget utnyttjas inte den informationen för vidare optimering. Varje avvikelse innehåller information om problemet och vilken adress det är, vilket kan vara i form av bild, ljud eller text. Idag saknas det forskning inom de problem som hindrar sopbilarna att komma fram till hushållens sopor eller varför vissa soptunnor inte går att tömma. Den här studien kommer därav att inrikta sig på hur ett BI-system kan hjälpa till att omvandla rådata i avvikelserapporterna till information vilket kan hjälpa till att motverka dessa, detta kommer att tas upp mer i det tredje

kapitlet ”problembeskrivning”. I dagsläget blir ungefär var tionde soptunna felrapporterad inom Östra Skaraborg och sopbilschaufförerna anser att det tar tid att identifiera och rapportera problemet. Genom att använda sig av en BI-lösning kommer avvikelserapporterna av tömningar från Avfallshantering Östra Skaraborg att analyseras för att identifiera problemen. Denna process kommer att ske genom att skapa en relationsdatabas och med ETL-processen identifiera, hämta, ändra och ladda både intern och extern data till systemet för att sedan genom analyser ta fram information för beslutsfattning. Ett av problemen gäller att identifiera vilken data som är relevant inom både interna och externa källor, detta då det kan finnas många faktorer som påverkar varför hushållens sopor inte går att tömma, bland annat vägar som är avstängda eller väder. När datan kommer från olika källor läggs stort fokus på ETL-processen för att datan ska få rätt kvalité och följa företagets dataregler. Denna information ska sedan användas för att lämna fram ett förbättringsförslag hur det ska gå att effektivisera sophämtningen.

Syftet med studien är att ta fram beslutsunderlag med hjälp av ett BI-system för att kunna effektivisera sophämtningen. Med hjälp av ett BI-system ska jag ta reda på varför det inte går att tömma alla soptunnor hos hushållen för att sedan komma fram med ett beslutsunderlag vilket kan effektivisera sophantering. Genom att ta reda på vad som orsakar dessa problem kan det bli en effektivare sophämtning då varje avvikelse kräver tid i form av identifiering av problem och rapportering vilket påverkar miljön.

## 2 Bakgrund

Detta kapitel innehåller information vilket behövs för att förstå bakgrunden till denna studie. Bakgrunden börjar med en övergripande bild över Business Intelligence för att sedan gå djupare in på dess komponenter. Den sista delen går igenom sopbranschen.

### 2.1 Business Intelligence

Business Intelligence (BI) har sitt ursprung från 1960-talet då det först började forskas om beslutsstödssystem (en. decision support systems, DSS) (Urban, 1966). Det var först på 80-talet klivet gick till ett mer utvecklat DSS och började användas för att få tillgång till stora mängder data som sedan kunna analyseras (Power, 2007). I början på 1990-talet togs ännu ett kliv då Data Warehouse och On-Line Analytical Processing (OLAP) började bli mer utvecklat. Innan OLAP var utvecklat var det svårt att få ut data från relationsdatabaserna då varje fråga tog mycket datorkraft samt att systemen inte var flexibla nog att hantera frågorna effektivt (Intricity, 2011). Utveckling av OLAP gjorde att frågorna inte ställdes direkt till databasen utan istället räknade OLAP ut kombinationer av utfall innan användaren ställde en fråga vilket ledde till i stort sett svar på direkten. När DSS blev allt mer och mer utvecklat så delades de upp i fem olika kategorier enligt Illustration 1 för att lättare kunna definiera de olika delarna (Kopáčková och Škrobáčková, 2006).

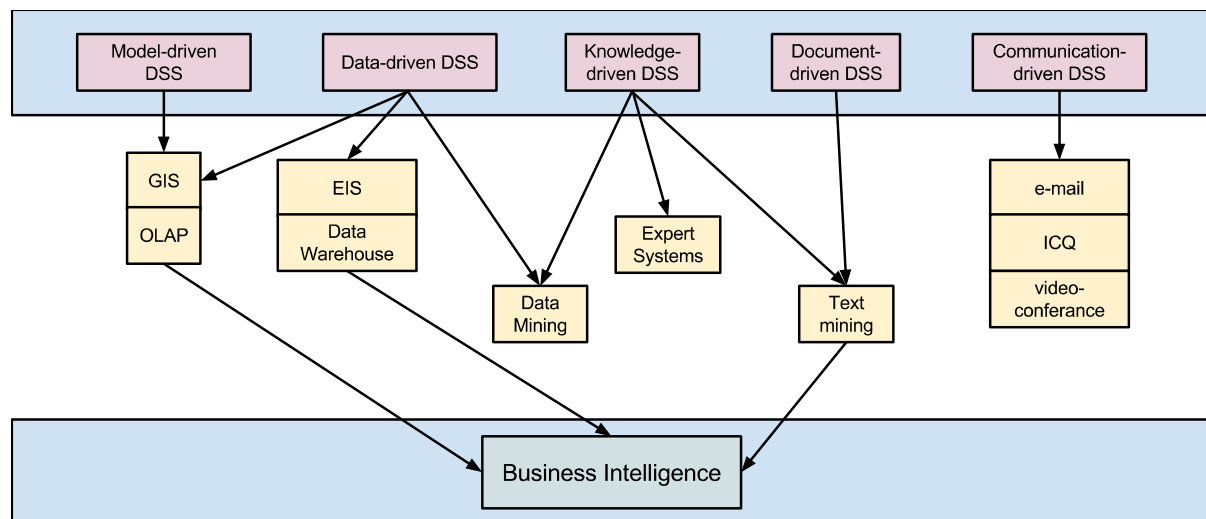


Illustration 1: Utvecklingen av DSS till BI. Egen version av Kopáčková och Škrobáčková, 2006, s.102.

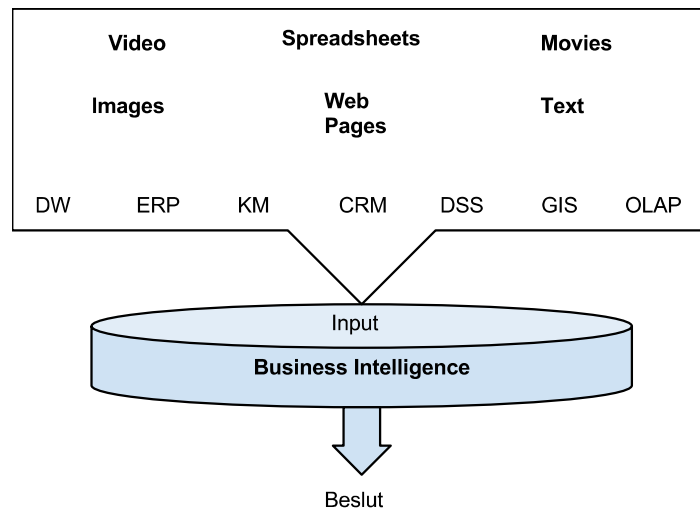
Kopáčková och Škrobáčková (2006) skriver att Business Intelligence innefattar de fyra första delarna av DSS, men även andra eller färre komponenter kan användas utifrån kundens önsningar.

Det råder delade meningar vad definitionen för Business Intelligence är. Golfarelli, et al (2004) beskriver det kortfattat som en process som omvandlar data till information, som sedan omvandlas till kunskap. Golfarellis definition tar upp själva målet med BI men



beskriver inte själva processen för att komma dit. En annan definition som mer går in på själva processen för att omvandla data är Negash (2004). Han beskriver att ett Business Intelligence är ett system som kombinerar datainsamling, datalagring och Knowledge Management med analysering för att utvärdera information som ska presenteras till beslutsfattarna, detta med målet att kunna öka effektiviteten och tidsaspekterna för att skapa bättre beslut. Dayal, et al., går i samma riktlinje med de tekniska bitarna och skriver ”Business Intelligence (BI) refers to technologies, tools, and practices for collecting, integrating, analyzing, and presenting large volumes of information to enable better decision making” (Dayal, et al., 2009, s. 1). De menar alltså att BI handlar om teknologier, verktyg och kännedom för insamling, integrering, analysering och presentation av stora volymer av information för att kunna få bättre beslutsunderlag.

Både Dayal, et al., (2009) och Negash & Gray (2003) skriver att termen Business Intelligence innefattar flera olika teknologier. Illustration 2 visar att det kommer in data i flera olika format (bilder, text, filmer etc.) och från olika system (Data Warehouse, Knowledge Management Systems etc.).



*Illustration 2: Input till Business Intelligence. Fet stil är format på filer och vanlig är olika system. Egen version av Negash and Gray, 2003.*

Även om BI-definitionerna skiljer sig åt har de ändå en sak gemensamt; en process som går ut på att omvandla rätt data till användbar information som hjälpa till att förbättra besluten inom organisationen. Därav blir min egen definition av Business Intelligence följande: ”Business Intelligence (BI) är en term vilket innebär omvandling av rådata till värdefull information för beslutsfattning genom datainsamling, integrering och analysering”.

### 2.1.1 Geographical Information System

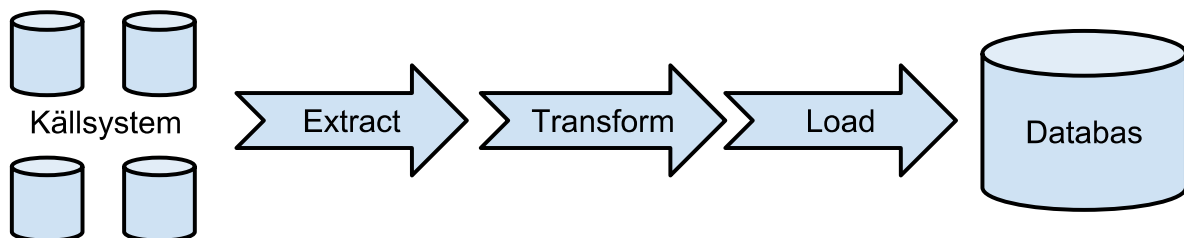
Geographical Information System (GIS) är en teknologi inom BI vilket är ett system som tillåter användaren att hämta, lagra, integrera och analysera data utifrån geografiska positioner (Rada, et al., 2013). Chrisman (1997) delar upp GIS i tre komponenter: områden, tid och attribut. Område innefattar den yta eller plats själva systemet analyserar, tiden är

tidpunkten området gäller och attribut är något som beskriver det, till exempel gatunamn. GIS har växt på det senaste åren då teknologin blir allt mer avancerad och idag finns GPS i stort sett i varje smartphone och även smartklockor (Fonseca et al, 2002). En av fördelarna med GIS är att det kan göra ”tänk om” analyseringar vilket kan räkna ut framtidsscenario (Moeinaddini, et al., 2010, och Tavares, et al., 2011). Till exempel kan en GPS-applikation göra tänk om du skulle svänga vänster om 100 meter så sparar du 5 minuter. Idag används GIS främst för att studera den vanliga fordonstrafiken eller för att optimera rutter (Zsigraiova, Semiao and Beijoco, 2013).

I dagens forskning finns det inget som visar hur avvikelserapportering inom avfallshanteringen kan minskas då fokuset har legat på att göra ruttoptimering inom GIS.

### 2.1.2 ETL

ETL är en förkortning för Extraction-Transformation-Loading (sv. extrahering, omvandling, inladdning), och används inom BI för att hämta och omvandla eventuellt felaktiga data vilket sedan laddas in i en databas (Simitsis, et al., 2005). Genom att använda ETL hjälper det till med BI-analysen då du får rätt data på valt ställe istället för utspridd data vilket kan innehålla fel. I början ansågs ETL-processen bara att vara ett hjälpverktyg och på så vis ignorerades det in inom forskningen (Dayal, et al., 2009). Idag används det mer flitigt och hjälper till att få ut den endast nödvändigaste informationen. Data cleaning (sv. datatvättning) är en av nyckelkomponenterna inom ETL-processen, men även den största utmaningen (Raman och Hellerstein, 2001; Rahm och Do, 2000). Datatvättning handlar om att data blir rensad/tvättad från onödig data och gör den mer relevant och genom detta så blir det en bättre datakvalité. Illustration 3 visar övergripande faserna i ETL-processen.



*Illustration 3: Processen inom ETL-processen. Hämtar data från ett källsystem för att sedan omvandla den innan inladdningen sker till en databas (egen illustration).*

Simitsis, et al (2005) delar upp ETL-fasen i 5 olika uppdrag.

1. Identifiera relevant information från källsystemet
2. Uthämtningen (extraction) av denna information.
3. Anpassa och integrera data från olika system till ett och samma format.
4. Rensa (clean/transform) data efter organisationens regler och mål.
5. Skicka (load) in data till databasen.

Trujullo och Mora (2003) beskriver ETL-fasen som en process då du hämtar data från heterogena operationella system, ändrar data genom bland annat omvandling, normalisering etc. och sedan inladdningen till databas. De beskriver även att ETL-fasen är nyckelkomponenten i ett databassystem, detta då inkorrekt eller missledande data kommer ge felaktig information vilket leder till att felaktiga beslut tas. Dayal, et al. (2009) beskriver ETL-processen på ett liknande sätt som Trujullo och Mora.

### 2.1.3 Datakvalité

Datakvalité (en. data quality) är när data är anpassad för användning av dataanvändarna (Wang and Strong, 1996). Det gäller att ha bra datakvalité inom Business Intelligence då det minskar chansen att felaktig data påverkar analysen till fel resultat (Ponniah, 2010). Enligt en studie från Ponniah (2010) svarar cirka 45 procent att datakvalité står för den största utmaningen inom datalagring. Datakvalitén kan förbättras genom datatvättning för att ändra de värdena som är fel eller motsägelsefulla. Wang and Strong kom fram till följande definition av datakvalité i Tabell 1.

**Tabell 1** Definition av datakvalité

|                      |   |
|----------------------|---|
| <b>Noggrann</b>      | Att värdet för ett element är korrekt. Om det står att lagersaldot är 10 i systemet så ska även saldot vara 10. |
| <b>Aktuell</b>       | Den data som finns ska fortfarande gälla. Om en person har flyttat så ska den nya adressen finnas.              |
| <b>Tillförlitlig</b> | Datan ska vara tillförlitlig och trovärdig för användaren.  |
| <b>Fullständig</b>   | Att det inte finns några missade värden. Det måste stå kortnummer i onlineordern för att köpet ska gå igenom.   |
| <b>Relevant</b>      | Datan ska vara relevant för arbetet. Skostorlek behöver inte finnas när du köper elektronikprylar.              |
| <b>Tolkningsbar</b>  | Användaren av datan ska kunna tolka och förstå innebörden av den.   |

Ponniah (2010) ägnar ett helt kapitel i sin bok om Data Warehouse angående just datakvalité då det är en såpass viktig del inom datalagring. Ponniah listar i utöver Wang och Strong (1996) punkter även följande på datakvalité i Tabell 2.

**Tabell 2** Utökad definition av datakvalité.

|                   |   |
|-------------------|---|
| <b>Redundans</b>  | Det ska inte finnas några dubletter av data i systemet.   |
| <b>Konsistent</b> | Datan och dess värde ska betyda likadant i varje system. Datumet 140213 ska innebära året 2014 i alla system. |

|                               |  |
|-------------------------------|--|
| <b>Följa företagsreglerna</b> | Datan får inte bryta de företagsregler som finns. I en bank kan du inte låna ett negativt värde eller att köp-nupriset är lägre än högsta budet. |
|-------------------------------|--|

#### **2.1.4 Sammanfattning av Business Intelligence**

Ett BI-system tar hand om rådatan i en fil eller en databas (extract) och omvandlar den till bättre datakvalité genom datatvättning (transform). Efter datatvättningen laddas den in till en relationsdatabas (load). Utifrån detta ska data från databasen analyseras för att ta fram information.

## **2.2 Kommunal avfallshantering**

Användningsområdet för BI i denna rapport är kommunal avfallshantering. Den kommunala avfallshanteringen ingår i riksdagens miljöbalk och går ut på följande:

8 § Varje kommun skall, om inte annat föreskrivs med stöd av 6 §, svara för att

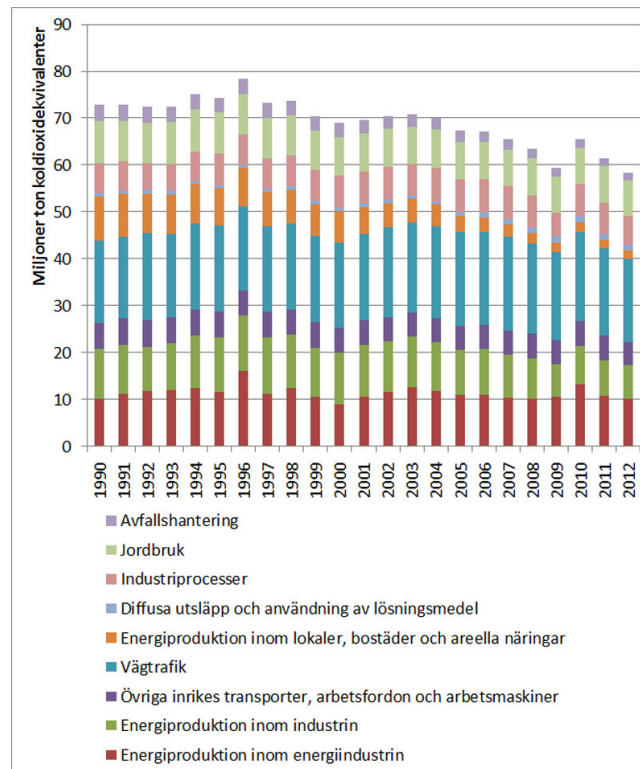
1. hushållsavfall inom kommunen transporteras till en behandlingsanläggning, om det behövs för att tillgodose såväl skyddet för människors hälsa och miljön som enskilda intressen, och

2. hushållsavfall från kommunen återvinns eller bortskaffas.

(Riksdagen, 2013)

Kommunal avfallshantering går alltså ut på att kommunen har ett ansvar att transportera bort, återvinna eller bortskaffa allt avfall från hushåll, och i vis mån industrier. Den kommunala avfallshanteringen består av flera aktiviteter som innefattar generering, källsortering, lagring, insamling, överföringar, transport och tömning (Rada, et al., 2013).

Idag finns det 2442 sopbilar registrerade i Sverige vilket kan jämföras med 4 487 658 personbilar som är i trafik (Transportstyrelsen, 2014). Utsläppen från avfallshantering i Sverige under 2012 ligger på cirka 1,3 miljoner ton koldioxidekvivalenter (mått för att mäta hur olika gaser påverkar växthuseffekten) (Illustration 4). Detta motsvarar 2,2 % av allt utsläpp i Sverige.



*Illustration 4: Utsläpp av olika aktörer inom Sverige. Bild från (Westermarck, 2013).*

SKL, Sveriges Kommuner och Landsting, skriver att kommunerna idag arbetar för att förbättra avfallshanteringen ännu mer och att fokuset också har flyttats mot att minska mängden avfall (Fredriksson, 2012).

Rada, et al. skriver hur den kommunala sophanteringen kan förbättras:

*”enhanced information is needed, in order to optimize and reduce the management costs and in order to improve the quality of the service toward the citizen”* (Rada, et al., 2013, s. 785).

De säger alltså att det behövs stödjande information för att kunna optimera och reducera kostnaderna för att kunna öka kvaliteten av service gentemot invånarna.

### 3 Problembeskrivning

*Problemformuleringskapitlet beskriver de problem och hinder det finns inom projektet vilket sedan kommer att undersökas. Andra delen av kapitlet kommer gå igenom de avgränsningar projektet har samt det förväntade resultatet av studien.*

Inom den kommunala avfallshanterings databas har det rapporterats in 18 000 avvikelser av cirka 170 000 tömningar. Detta är cirka 11 % vilket anses vara högt över den godkända gränsen enligt dem själva. Varje avvikelse kräver sin tid i form av identifiering och rapportering om varför det inte går att tömma. Detta kan påverka körningen då sopbilarna i onödan står i tomgång när de skriver sina rapporter vilket leder till att miljön påverkas negativt. Att främja miljön är något varje kommun har ansvar över vilket leder till att problemet inte bara hämmar effektiviteten hos sopbilarna utan för alla personer i kommunen. Enligt sopbilschaufförerna tar det tid genomföra avvikelserprocessen istället för att kunna tömma som vanligt.

Utifrån detta ska jag sätta upp ett BI-system innefattande en databas och med hjälp av ETL-processen ta fram, tvätta och ladda in data. Utifrån databasen ska datan analyseras för att ta fram beslutsunderlag vilket ska leda till en effektivisering av sophanteringen.

I och med att avvikelserrapporteringar innefattar flera olika format (text, bild, ljud) behövs det datatvättning av viss data för att kunna öka datakvalitén inför dataanalysen. Problemet ligger också att ta reda på vilken intern data som är nödvändig för att kunna få information angående avvikelserna.

Den interna datan inom företaget (avvikelser) innefattar rik data om vilka hushåll och varför den inte går att tömma. Detta kommer dock inte enbart att räcka då den endast tar upp vad problemet är. Det gäller också att identifiera nya källor utav data då det kan finnas yttre faktorer som påverkar varför inte tömningar sker. Det kan finnas flera olika orsaker, bland annat vägar har stängts av olyckor eller väder ställer till det. Genom att använda flera olika datakällor kommer det största problemet ligga inom ETL-processen där tvättningen (en. cleaning) utgör en stor del då datan måste tvättas till företagets egna regler. Tvättningen av data är en del av Transform (sv. omvandling) i ETL-processen vilket går ut på att ändra datan till en bättre kvalitet. Den forskning som har gjorts inom datatvättning handlar främst om strukturerad tabelldata och inte om ostrukturerad textdata (Tang, et al, 2005). Efter att datan är tvättad kan en analys ske för att ta fram beslutsunderlag.

Genom dessa beskrivna problem kommer grunden för den frågeställning vilket studien ska forskas på:

*Hur kan avfallshanteringsbranschen effektivisera sophämtningen med hjälp av en Business Intelligence-lösning?*

För att besvara på frågan ska följande tre steg utföras:

1. Göra en kvalitetsgranskning på befintlig data.

- a) Gå igenom en ETL-process.
2. Analysera befintlig data för att ta fram beslutsunderlag.
3. Ta fram rekommendationer för Avfallshantering Östra Skaraborg.

Med hjälp av denna frågeställning ska införandet av ett Business Intelligence-system användas för att omvandla *rådata till värdefull information för beslutsfattning genom datainsamling, integrering och analysering*. Frågeställning kopplas ihop med syftet av studien som är att med hjälp av en BI-lösning kunna effektivisera sopheringen vilket ska leda till att antalet avvikelser minskas.

### **3.1 Avgränsning**

Den data som samlas in kunde som tidigare nämnt vara ett flertalet olika format, de ostrukturerade är bilder, ljud, fritext, och de strukturerade är färdiga kommandon. Vid skrivande stund är 1378 av cirka 22189 avvikelser bilder eller ljud, vilket motsvarar 6,2 %.

När en sådan liten del av datan består av ljud och bild blir kostnaden i både tid och pengar större än nyttan om datavättning ska ske på dessa. Både bilder och ljud är extra svårt att analysera då det inte går att läsa ut viktiga nyckelord utan någon form av omvandling. Utifrån projektets tidsram och budget har det därför valts att fokusera endast på strukturella informationen, rådata samt fritexten.

### **3.2 Förväntat resultat**

Det förväntade resultatet är att få en effektivare sophering med hjälp av de beslutsmaterial som har tagits fram utifrån BI-analysen. Med en effektivare sophering kommer antalet avvikelser att minska vilket leder till att både sopsbilschaufförerna får en bättre arbetsprocess genom att slippa rapportera lika många avvikelser, samt att miljön blir bättre genom att slippa onödiga stopptider för avvikelserrapportering.

## 4 Metod

*Det här kapitlet tar upp de olika metoder som har använts inom studien och motiveringar till varför. Första delen inriktar sig på insamling av data medan sista delen går in på val av metod inom BI-processen.*

### 4.1 Insamling av data

Det finns flera metoder att använda för att samla in information om en verksamhet. Vid insamling av information finns det två typer av metoder, kvalitativ och kvantitativ. Myers (1997) beskriver de bägge forskningsmetoderna följande:

1. Kvalitativ forskningsmetod innebär att du använder kvalitativ data, till exempel intervjuer, dokument och observationer för att förstå och förklara sociala fenomen.
2. Kvantitativ forskningsmetod är när du studerar naturliga fenomen. Exempel på detta är undersökningar, experiment och numeriska metoder.

I början av studien gällde det att få reda på så mycket som möjligt om företaget för att veta hur de arbetade och hur data samlades in. Genom att ta reda på information om hur arbetsprocesserna går till inom verksamheten blir det lättare att förstå hur den fungerar. För att få en djupare förståelse om något så behövs det en kvalitativ forskningsmetod som intervjuer, dokument och observationer för att förstå hur allting går till och varför. Detta kallas för triangulering när det används flera metoder för att studera en och samma sak (Jick, 1979). Den kvantitativa metoden funkar alltså inte för detta då den mest fokuserar på nummer vilket inte förklarar hur verksamheten fungerar. Detta stärks upp av Berndtsson, et al (2008) som förklarar kvantitativa metoder som hur något är konstruerat, hur det är uppbyggt, eller hur det funkar.

Utöver den kvalitativa metoderna intervju och kvalitativ enkät har också en kvantitativ metod använts för att undersöka avvikelserapporterna då dessa innehåller statistik om soptömningarna. Genom att både använda kvalitativa och kvantitativa metoder för denna forskning ges information som kompletterar varandra. Den kvantitativa metoden undersöker data i avvikelserapporterna och får fram nummer och statistik om felen, och den kvalitativa metoden får fram hur arbetsprocessen går till, hur arbetarna själva upplever problemet och om de tycker informationen i dokumenten stämmer. Tillsammans ger de en helhet över problemet vilket behövs för att analyseringen inte endast ska lösa ett problem som stämmer i pappret men inte enligt intervjuerna, och vice versa.

#### 4.1.1 Intervjuer

Den person som ska intervjuas är den ansvarige för projektet hos Högskolan i Skövde för att få en djupare förståelse för verksamheten, vilka mål som finns och en övergripande bild av projektet. Genom att ta reda på de arbetsprocesser och mål som finns inom verksamheten blir det lättare att förstå hur den fungerar.

Det finns fyra stycken olika intervjutekniker (Opdenakker, 2006).



1. Ansikte mot ansikte (en. Face-to-face)
2. Telefonintervjuer
3. Chatintervjuer
4. E-postintervjuer

Opdenakker (2006) delar upp dessa i två delar, synkrona (realtid) och asynkrona (tids och platsberoende). Ansikte mot ansikte är synkron i både tid och plats medan telefon- och chatintervjuer är synkrona i tid men asynkron i plats. E-postintervjuer hamnar under asynkron inom både tid och plats.

Den intervjuteknik som användes blev de synkrona teknikerna med primärt ansikte mot ansikte och chatintervjuer som sekundär teknik. Valet på dessa baserades utifrån fördelarna med respektive teknik. Genom att ha intervjun ansikte mot ansikte får intervjuaren extra fördelar genom den sociala påverkan, vilket bland annat är tonläge och kroppsspråk (Opdenakker, 2006). Dessa sociala aspekter förstärker svaren och kan ge extra information på ett svar som en asynkron intervju inte kan ge. Genom ansikte mot ansikte får intervjuaren även information på direkten utan att behöva vänta på svar vilket gör att arbetet inte skjuts upp. Denna teknik användes också för möjligheten att spela in.

Chatintervjuerna användes via verktyget Skype. Denna teknik var sekundär då ansikte mot ansikte hade mest fördelar. Chatintervjuer användes endast för kortare frågor eller följdfrågor som kom på i efterhand. Denna metod användes då bägge parter skulle förlora tid och eventuellt resekostnader på att arrangera ett möte där det endast kommer att ställas några korta frågor.

Det finns tre sorters intervjumetoder: strukturerade, semi-strukturerade och ostrukturerade (Dicicco-Bloom och Crabtree, 2006). Den strukturerade är förutbestämd och ger oftast korta svar, medan den ostrukturerade är när frågor kommer fram efter hand.

Vid val av intervjumetod valdes den semi-strukturerade intervjumetoden för ansikte mot ansikte, detta då den ställer öppna men förberedda frågor vilket ger mer utvecklande svar. Denna metod ger också möjlighet till följdfrågor som gör det möjligt att få djupare förståelse för något som är oklart eller har missat. Utifrån detta förbereddes öppna frågor där vissa även hade del- eller följdfrågor. Frågorna baserades utifrån att ta reda på så mycket som möjligt om verksamheten. Genom att ställa öppna frågor kunde gå det att få svar på många av delfrågorna på direkten utifrån den första frågan men även saker som inte hade tänkts på innan. Vid tillfällen då alla mina delfrågor inte blev besvarade kunde de ställas som följdfrågor för att få reda på alla svar jag var ute efter. Under intervjuerna blev det då lätt att upprepa, få ytterligare förklaring eller bekräfta fakta (Dicicco-Bloom and Crabtree, 2006).

#### **4.1.2 Enkätundersökning**

En enkätundersökning används för att ta reda på ”vad är det som händer” och är väl definierade, har exakta rutiner och ger lätt tolkningsbar data (Pinsonneault and Kraemer (1993). Enkäterna kommer att användas för att reda på vad alla sju sopsbilschaufförerna hos AÖS anser om rapportering av avvikelser och vilka faktorer som påverkar deras tömningar.

Genom att skicka ut till alla sopbilschaufförer inom AÖS ökar reliabiliteten av undersökningen. Det finns tre typer av enkätundersökningar enligt Pinsonneault and Kraemer (1993).

1. Utforskande
2. Beskrivande
3. Förklarande

Utforskande metod används för att ta reda på mer om ett ämne, beskrivande är att någon ska beskriva ett event eller process, och förklarande är när någon ska beskriva en relation mellan två variabler.

Den metod som har använts är den beskrivande enkäten då sopbilschaufförerna ska beskriva vad de anser om avvikelserapportering. Då de ska beskriva en process används öppna frågor istället för stängda så att de kan beskriva med egna ord hur processen går till. Enkäter räknas i detta fall som kvalitativ forskning då öppna frågor används och för att få en förståelse hur de arbetar.

#### **4.1.3 Skapande av frågor**

Vid skapandet av frågorna för intervjun och enkäter gick följande steg igenom. Denna metod är en egen modifierad version av Wellington och Szczerbiński (2007):

1. Skriv ner alla frågor som dyker upp i huvudet.
2. Säll ut de frågor som är relevanta och ta bort de frågor som är lika varandra.
3. Ändra följderna på frågorna så det blir en röd tråd mellan alla frågor.

#### **4.1.4 Etiska aspekter**

Alla intervjuade personer kommer att vara anonyma förutom arbetstitel. Anonyma personer inte är lika försiktiga under intervjuer och kan på så vis berätta mer material utan att vara rädd för att någon ska spåra dem (Gergen et al, 1973). Detta valdes också för att om till exempel sex personer hade sitt namn och en ville vara anonym så går det att räkna ut vilken person som är anonym genom utslutning. Vid intervjuer som skedde ansikte mot ansikte behövdes ett godkännande från den intervjuade om en inspelning av intervjun fick ske.

## **4.2 Metod för datavätning**

Den metod som ska användas för datavätningen är Rahm och Do (2000) "Data cleaning approaches". Valet föll på denna metod för att den tar upp hur datan identifieras till att skicka tillbaka den äldre datan till systemet istället för bara att tvätta datan. Rahm och Do har delat upp datavätningen i fem olika steg enligt tabell 3.

**Tabell 3** Metod för datavätning

|               |   |
|---------------|---|
| Data analysis | Göra en dataanalys för att ta reda på vilken data som har |
|---------------|---|

|  |   |
|--|---|
| (Dataanalys)   | fel och motsägelser. Denna dataanalys görs både manuellt och via färdiga verktyg.   |
| Definition of transformation workflow and mapping rules<br>(Definera regler) | Definiera de olika reglerna datatvättningprocessen ska gå igenom. Genom att göra detta steg kommer framtida data att bli lättare att tvätta då reglerna redan finns.              |
| Verification<br>(Verifiera)  | Verifiera så att tidigare steg uppfyller kraven genom att göra tester på datan. Förbättra sedan eventuella fel och gör om processen.  |
| Transformation<br>(Ändra)  | Efter verifikation sker själva tvättningen/ändringen av data.   |
| Backflow of Cleaned Data<br>(Ändra äldre data)                               | Om möjligt så gå tillbaka till den ursprungliga källan och ändra felen där. Detta så att hela tvättprocessen inte behöver göras om varje gång en hämtning sker från det systemet. |

Vid definition av regler för datatvättningen kommer reglerna i tabell 4 att användas. Valet på dessa regler är baserade utifrån Poniah (2010) där han listar olika typer av omvandlingar och de vanligaste typerna av ändringarna enligt OpenRefine (2014).

**Tabell 4** Metod för definiera regler

| <b>Regler</b>                           | <b>Beskrivning</b>   |
|---|--|
| Ändra versaler                          | Beroende på vilket sätt mestadels av texten är skriven (endast små bokstäver, bara stora eller börjar med stor bokstav) ändras alla till samma typ.          |
| Ta bort extra mellanrum efter en mening | Tar bort onödiga mellanrum efter texter, t.ex "vägen" och "vägen ".  |
| Ta bort dubbla mellanrum i rad.         | Tar bort om två mellanslag har skrivits i rad. Till exempel "Fel tunna" och "Fel tunna".   |
| Dela multivärden i celler.              | Tar bort dubbla värden eller attribut som står i en och samma cell. Till exempel kan "Högskolevägen 6" delas upp i gatan "Högskolevägen" och gatunummer "6". |
| Ändra datum                             | Skriva tid på rätt sätt. T ex skriver vissa 12/02/2004 som 12 feb, men i USA så är det 2 dec. Denna är viktig när det hämtas data från                       |

|                  |  |
|------------------|--|
|                  | flera källor.  |
| Deduplicering    | Ta bort dubletter.   |
| Omvandling       | Omvandla till standardisering och gör det förståeligt för användaren.  |
| Klustring        | Det finns två olika klustringsmetoder som ska användas. "Key collision" identifierar nyckelvärden i varje kolumn och "Nearest Neighbor" letar efter små avvikelser som har skett n steg tillbaka.  |
| Key collision    | Tar fram liknelser utifrån: <ul style="list-style-type: none"> <li>• Ta bort onödiga tecken som punkter och frågetecken</li> <li>• Ta bort dubletter</li> <li>• Ändrar västerländska tecken till ASCII (ö blir o)</li> <li>• Delar upp texten i strängar efter varje mellanslag</li> </ul> |
| Nearest Neighbor | Denna algoritm letar efter liknelser n antalet steg tillbaka. Genom att kolla närliggande ord kan den identifiera meningar där endast ett fåtal tecken eller ett ord skilljer.   |

### 4.3 Metod för analysering

Analysering kommer att ske via ett externt verktyg som visar en visualisering av data. Genom att visualisera datan (visa grafer) blir det lättare att få en överblick gentemot de textsvaren SQL-fråga ger. Analysering ska användas för att ta reda på mer information om avvikelserapporterna vilket ska användas för beslutsunderlag. Alla analyseringar börjar först på en övergripande nivå för att sedan gå nedåt i hierarkin för specifika händelser. Genom att arbeta enligt den metoden går det lättare att hitta källan till felet. Till exempel är det lättare att börja söka fel per kontinent och sedan gå neråt på länder istället för att bläddra igenom alla länder för sig för att hitta det land som ger mest fel.

### 4.4 Val av verktyg

Val av verktyg för datatvättning och visualisering baseras utifrån två kriterier:

- Kunna utföra de tekniska processerna från "Data cleaning approaches".

- Vara Open Source och/eller gratis då studien inte har någon budget.

Databashantering sker i Microsoft SQL Server 2012 och Microsoft SQL Management Studios 2014 för att använda samma databas som används av Högskolan i Skövde. Denna mjukvara tillhandahålls av Högskolan i Skövde. För datatvättningsverktyg föll valet på OpenRefine, tidigare Google Refine, vilket uppfyller dessa krav. OpenRefine beskriver sig själv som "A free, open source, powerful tool for working with messy data" (Openrefine, 2014). I verktyget finns möjligheter att tvätta datan via förinställda metoder men också genom programmeringsspråk. Det finns en inbyggd förhandsvisning/verifiering vilket gör att du inte behöver ändra datan och sen backa om det skulle visa sig vara fel. Det går även att spara reglerna för framtida tvättningar. På så vis underlättas definition av regler och verifieringsfasen i "Data cleaning approaches". För analyseringen kommer mjukvaran "Qlikview" att användas. Qlikview är en gratis Business Intelligence-programvara vilket visualiserar data för att lättare kunna upptäcka ny information (Qlikview, 2014).

## **4.5 Genomförande**

### **4.5.1 Intervju och enkäter**

Både ansikte mot ansikte intervjun och enkäterna utfördes på respondentens arbetsplats för att personerna skulle känna sig bekväma med sin omgivning och på så vis få högre reliabilitet. Gällande chattintervjuerna är det svårt att avgöra var personen befinner sig, dock genomfördes de flesta sådana intervjuer under dagtid på vardagar för att undvika arbete på fritiden.

#### **Respondent – Ansvarig projektledare**

Intervjun av den ansvarige projektledaren skedde i början av studien för att ta reda mer på om systemet och hur arbetsprocesserna ser ut. Intervjun tog ungefär 40 minuter och var vid hennes kontor. Intervjun spelades in efter medhåll och transkriberades efter att intervjun var klar. Transkriberingen skickades tillbaka för kontrollering och gav möjligheten att göra korrigeringar vid behov för att förstärka reliabiliteten, dock skedde inga ändringar. Ett flertalet chattintervjuer användes för att ställa nya frågor som kommit upp under studiens gång vilket varierade varade mellan 5 och 30 minuter.

Utifrån transkriberingen delades svaren upp i olika stycken med relevant och ickerelevant text då alla svar inte behövdes i studien. Därefter markerades nyckelord i texten för att lättare kunna hitta de viktiga delarna. Genom att strukturera dokumentet på detta vis blir det lättare att analysera texten då all relevant text redan finns uppdelad.

#### **Respondent – Sopbilschaufförer**

Enkäten lämnades till en ansvarig av projektet på Högskolan i Skövde som sedan skickade vidare den till Avfallshantering Östra Skaraborg. AÖS lämnade sedan ut enkäterna till sopbilschaufförerna där de fick fylla i den på arbetsplatsen innan de skickade tillbaka den. Alla sju sopbilschaufförer på AÖS var med och fyllde i enkäten tillsammans.

#### **4.5.2 Datatvättning**

Med hjälp av datatvättningsverktyget Google Refine sattes regler upp utifrån tabell 4. Vissa av reglerna finns redan förprogrammerade i programmet vilket minskade antalet egenprogrammerad kod. De reglerna som inte fanns förprogrammerade användes Open Refines egna programmeringsspråk GREL (Google Refine Expression Language). Både innan och efter varje ändring kommer datan att kontrolleras för att se att den har gått rätt till. Datatvättning skedde innan inladdningen av filerna till databasen.

## 5 Empiri

*Det är kapitlet presenterar det material som fåtts in via intervjuer och databasen samt hur de är kopplade till problempreciseringen.*

### 5.1 Enkäter och intervjuer

#### 5.1.1 Projektansvarig på Högskolan i Skövde

Arbetet är ett samarbete mellan Högskolan i Skövde (HiS) på avdelningen Produktion - Automation i Institutionen för ingenjörsvetenskap och Avfallshantering Östra Skaraborg (AÖS). Deras projekt handlar om att utveckla ett Geographical Information System (GIS) inom den kommunala sophantering. HiS står för utvecklingen av själva systemet vilket AÖS kommer att använda sig utav. Systemet lagrar sopbilarnas rutter och samlar in data, bland annat genom att hämta information om hur många soptunnor som behöver tömmas, vilka tunnor som de inte kunde tömma, och när de ska tömmas för att få en effektivare sophantering. Sopbilarnas roll har den största rollen i systemet. *”Allt det som har med bilen att göra är väldigt centralt i verksamheten. Det handlar om att hämta soptunnor. Så allt det som kommer från bilen är verklighetskritiskt, det vill man även jobba mycket med, det är oerhört mycket och viktig data.”*

Systemet används redan idag och är i bruk, dock är det fortfarande under utveckling. Målet med projektet är att få effektivare kommunal sophantering, vilket innefattar mindre utsläpp och lättare att ta hand om soporna.

Systemet får inte bara in data utan kan också skicka ut information till chaufförerna. *”Det är inte bara bilen som skickar information till systemet, utan även systemet kan skicka information till bilen. Så det är en dubbelriktad kommunikation. Systemet skickar vad bilen ska göra och vilken rutt den ska göra. Det går även att skicka extrauppdrag i realtid om till exempel en kund behöver en akut tömning.”*

Det har runt 46 000 kunder och cirka 31 000 hämtplatser inom östra Skaraborg. På de 170 000 tömningarna de hade i i början på februari efter 10 månader i bruk hade 18 000 avvikelser anmälts, vilket är cirka 11 %.

#### 5.1.2 Sopbilschaufför

Sopbilschaufförerna beskriver sin process för avvikelserapportering att de stannar vid hämtplatsen och upptäcker avvikelser, till exempel att ett lock är sönder. När problemet är identifierat trycker de på knappen ”Avvikelse” på surfplattan där de får välja vilken typ av avvikelse det gäller (beskrivs i tabell 6 på sidan 21). De får också en möjlighet att ge kompletterande information genom text, bild eller ljud, detta är dock något som sällan behövs. Därefter väljer dom att spara och avvikelserna blir rapporterad till databasen.

Utifrån frågan vad de anser om rapporteringen så anser sopbilschaufförerna att rapporteringen är en nödvändig process i deras arbete vilket det har alltid gjort. Innan GIS utvecklades skrevs det mesta istället ner på postit-lappar istället för surfplattan. De anser att all typ av

rapportering tar tid att rapportera. ”Dock tar all typ av rapportering tid, men det är oundvikligt då vi ju måste hantera fel som uppstår.”. Utifrån frågan hur externa faktorer angående externa faktorer svaras ”Saker som kan påverka är väder (t.ex. snö eller halka), blockerade/oåtkomliga hämtplatser (pga t.ex. vägarbete eller att någon parkerat i vägen) eller att en kontaktperson är oanträffbar (personen ska t.ex. låsa upp till en container).”

## 5.2 Databas

För att få tillgång till all data hos AÖS har en backup-fil av deras databassystem använts. Avvikelseberapporterna lagras som rådata, det vill säga ej bearbetad data som inte har gått igenom någon omvandling. När datan rapporteras in sker alltså ingen tvättning av rapporterna utan det kan finnas fel och motsägelser som till exempel att en text som beskriver problemet innehåller felaktig data om adressen. Genom att ha en backup kommer därav ingen ny data att tillkomma från AÖS system. Eftersom GIS endast har varit uppe i 10 månader kommer inte alla månader att finnas med i undersökningen. Då systemet även är under utveckling kan resultaten vara olika under de första månaderna då allting var nytt, men även att alla funktioner inte var färdigutvecklade. Dessa punkter påverkar resultatet då det endast skett en insamling per kalendermånad och kan därav inte se trender eller liknande problem som skiljer mellan de olika åren. Då systemet inte är färdigutvecklat finns det även flera ”dummy-value” (fel ifyllda fält) och attribut som inte gäller längre.

Tabell 5 beskriver övergripande de tabeller ifrån AÖS som använts i analysen och Illustration 5 visar en modell hur de sitter ihop.

**Tabell 5** Tabeller vilket ska undersökas.

|                            |  |
|----------------------------|--|
| <b>IssueHeaders</b>        | I denna tabell sker själva problembeskrivningen. Ett problem kan få en rubrik tilldelad sig, t.ex trasigt kärl, men även en beskrivning på själva problemet varför den inte går att tömma. |
| <b>IssueFacility</b>       | I denna tabell finns kopplingen mellan problemet och hämtplatsen.  |
| <b>IssueMessages</b>       | Innehåller en detaljerad beskrivning om problemet.   |
| <b>IssueRouteDeviation</b> | Koppling mellan problem och avvikelse.   |
| <b>IssueReceptacles</b>    | Koppling mellan problemet och vilket kärl det gäller.  |
| <b>Location</b>            | Beskriver platsen där avvikelsen rapporterades. Delas upp i flera attribut, bland annat gata, postnummer och gatunummer.   |
| <b>Facility</b>            | Varje hämtplats som finns registrerad i databasen, det vill säga varje kund.   |



|                         |   |
|-------------------------|---|
| <b>Receptacles</b>      | Alla registrerade kärl.   |
| <b>Receptaclemodels</b> | Vilken typ av modell kärlet är.   |
| <b>RouteDeviations</b>  | <p>Alla avvikelser som rapporteras in. Själva grunden i analyseringen och består av 11 olika områden:</p> <ol style="list-style-type: none"> <li>1. Blockerat</li> <li>2. Trasigt kärl</li> <li>3. Kontaktperson oanträffbar</li> <li>4. Extra säckar</li> <li>5. Ej utställt</li> <li>6. Annan orsak</li> <li>7. Felaktigt stopp</li> <li>8. Felsorterad</li> <li>9. Felaktig kärldiameter</li> <li>10. Överfullt kärl</li> <li>11. Ospecificerat</li> </ol> |
| <b>Väder</b>            | Medeltemperatur av dygnstemperaturen hämtad från SMHI.  |

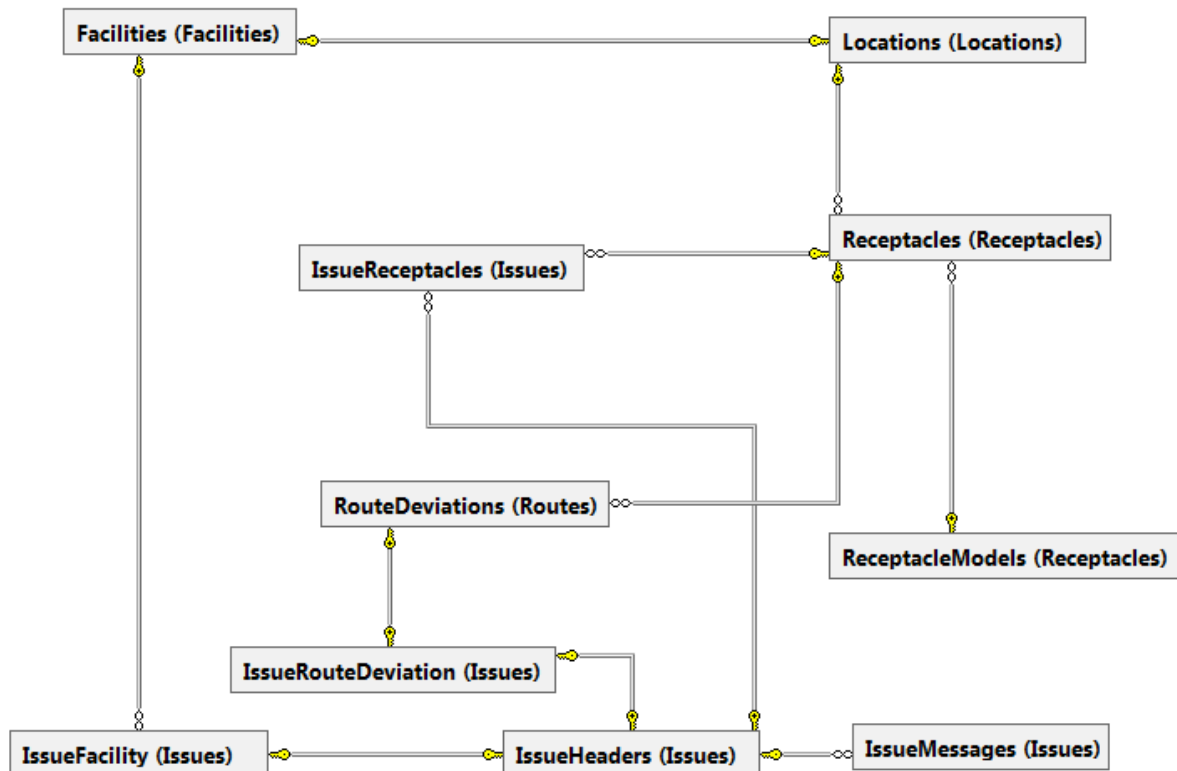


Illustration 5: Modell över vilka tabeller som ska användas i analysen. Endast tabellnamnen visas då det inte får plats på skärmen med alla attribut. Väder är exkluderat. Bild genererad från Microsoft Sql Management 2014.

I illustration 6 visas alla attribut i de tabellerna locations, receptacles och deviations och respektives nycklar.

| Locations (Locations) | Receptacles (Receptacles) | RouteDeviations (Routes) |
|-----------------------|---------------------------|--------------------------|
| <b>Id</b>             | <b>Id</b>                 | <b>RouteGroupId</b>      |
| ImportId              | SerialId                  | RouteDay                 |
| RegionId              | ImportId                  | ReceptacleId             |
| Position              | DivisionId                | Timestamp                |
| Name                  | LocationId                | UserId                   |
| Street                | ReceptacleModelId         | Type                     |
| StreetNumber          | PeriodId                  | AdditionalBags           |
| StreetExtension       | IntervalId                | Note                     |
| PostalCode            | Tag                       | LastModifiedBy           |
| PostalTown            | Position                  | Created                  |
| CountryCode           | Note                      | Modified                 |
| PropertyCounty        | LastModifiedBy            | Deleted                  |
| PropertyArea          | Created                   |                          |
| PropertyBlock         | Modified                  |                          |
| PropertyUnit          | Deleted                   |                          |
| Note1                 |                           |                          |
| Note2                 |                           |                          |
| Note3                 |                           |                          |
| LastModifiedBy        |                           |                          |
| Created               |                           |                          |
| Modified              |                           |                          |
| Deleted               |                           |                          |

Illustration 6: Attribut i tabellerna Locations, Receptacles och Deviations. Skärmdump från Microsoft SQL Mangement Studios 2014.

I illustration 7 är data hämtat utifrån tabellen ”Location”. Den vänstra visar städer som finns och den högra visar de gator som finns.

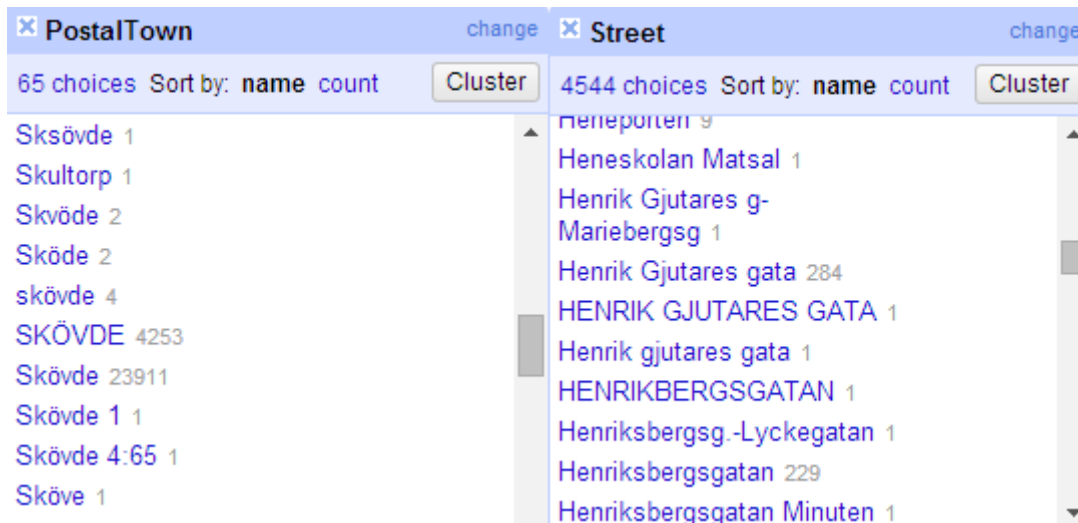


Illustration 7: Attribut i tabellen Location sorterad efter bokstavsordning. Skärmdump från OpenRefine.

Alla problemämnena härstammar från tabellen RouteDeviations (sv. ruttavvikelser) där alla avvikelser har rapporterats. Utifrån dessa elva ämnen är det nio stycken som har använts för rapportering. I tabell 6 visas en beskrivning av dessa nio problemämnena vilket kommer att användas i analysen, och i illustration 8 hur en avvikelse ser ut i en databas.

| RouteGroupId | R...        | ReceptacleId | Timestamp   | Userid                      | Type       | Additi... | Note | LastModifiedBy | Created                     | Modified                    | Deleted |
|--------------|-------------|--------------|-------------|-----------------------------|------------|-----------|------|----------------|-----------------------------|-----------------------------|---------|
| 1            | DFC69DC4... | 0            | 4A234FDF... | 2014-01-13 10:23:00.4500000 | 802EF68... | 3         | 4    | 802EF689-D...  | 2014-01-13 10:20:02.6870000 | 2014-01-13 10:20:02.6870000 | 0       |
| 2            | DFC69DC4... | 0            | 4A234FDF... | 2014-01-20 10:00:57.3590000 | 802EF68... | 3         | 15   | 802EF689-D...  | 2014-01-20 10:00:08.9300000 | 2014-01-20 10:00:08.9300000 | 0       |
| 3            | DFC69DC4... | 0            | 4A234FDF... | 2014-01-27 10:27:24.0120000 | 992506E... | 3         | 15   | 992506E7-6...  | 2014-01-27 09:48:48.1870000 | 2014-01-27 09:48:48.1870000 | 0       |
| 4            | DFC69DC4... | 0            | 1FDE16A9... | 2014-01-20 08:54:26.7900000 | 802EF68... | 4         | NULL | 802EF689-D...  | 2014-01-20 08:53:28.6300000 | 2014-01-20 08:53:28.6300000 | 0       |
| 5            | DFC69DC4... | 0            | 087D6809... | 2014-01-27 10:36:24.9240000 | 992506E... | 0         | NULL | 992506E7-6...  | 2014-01-27 09:57:31.4500000 | 2014-01-27 09:57:31.4500000 | 0       |
| 6            | DFC69DC4... | 0            | 7D08929C... | 2014-01-27 10:50:23.4530000 | 992506E... | 4         | NULL | 992506E7-6...  | 2014-01-27 10:13:27.1030000 | 2014-01-27 10:13:27.1030000 | 0       |
| 7            | DFC69DC4... | 0            | 819CF184... | 2013-10-28 12:58:01.6000000 | 992506E... | 6         | NULL | 992506E7-6...  | 2013-10-28 12:57:50.3500000 | 2013-11-19 15:50:47.4200000 | 0       |
| 8            | DFC69DC4... | 0            | 819CF184... | 2013-11-04 10:03:08.4770000 | 992506E... | 4         | NULL | 992506E7-6...  | 2013-11-04 10:02:38.0770000 | 2013-11-19 15:50:47.4200000 | 0       |
| 9            | DFC69DC4... | 0            | DCE7F168... | 2013-12-09 09:07:16.6210000 | 992506E... | 0         | NULL | 992506E7-6...  | 2013-12-09 09:05:20.6670000 | 2013-12-09 09:05:20.6670000 | 0       |
| 10           | DFC69DC4... | 0            | DCE7F168... | 2014-01-20 08:39:07.1460000 | 802EF68... | 1         | NULL | 802EF689-D...  | 2014-01-20 08:38:09.1670000 | 2014-01-20 08:38:09.1670000 | 0       |
| 11           | DFC69DC4... | 0            | DCE7F168... | 2014-01-27 09:07:34.2000000 | 992506E... | 1         | NULL | 992506E7-6...  | 2014-01-27 08:31:34.1030000 | 2014-01-27 08:31:34.1030000 | 0       |
| 12           | DFC69DC4... | 0            | 6355F40C... | 2013-12-09 07:26:54.1210000 | 992506E... | 0         | NULL | 992506E7-6...  | 2013-12-09 07:25:00.9100000 | 2013-12-09 07:25:00.9100000 | 0       |
| 13           | DFC69DC4... | 0            | 27ABDCF4... | 2013-11-25 13:08:49.3610000 | 992506E... | 6         | NULL | 992506E7-6...  | 2013-11-25 13:08:48.8330000 | 2013-11-25 13:08:48.8330000 | 0       |
| 14           | DFC69DC4... | 0            | 557D7A16... | 2013-10-28 09:57:15.7620000 | 992506E... | 0         | NULL | 992506E7-6...  | 2013-10-28 09:56:59.7500000 | 2013-10-28 09:56:59.7500000 | 0       |
| 15           | DFC69DC4... | 0            | 33D7EA79... | 2013-12-09 10:42:38.6880000 | 992506E... | 5         | NULL | 992506E7-6...  | 2013-12-09 10:40:40.9570000 | 2013-12-09 10:40:40.9570000 | 0       |
| 16           | DFC69DC4... | 0            | D666F399... | 2013-12-16 06:18:06.1700000 | 992506E... | 1         | NULL | 992506E7-6...  | 2013-12-16 06:15:37.4600000 | 2013-12-16 06:15:37.4600000 | 0       |
| 17           | DFC69DC4... | 0            | DA8784C1... | 2013-11-25 10:14:13.0110000 | 992506E... | 6         | NULL | 992506E7-6...  | 2013-11-25 10:14:19.6000000 | 2013-11-25 10:14:19.6000000 | 0       |
| 18           | DFC69DC4... | 1            | 4A234FDF... | 2013-10-29 07:53:34.6710000 | AEAF4F0... | 0         | NULL | AEAF4F05-8...  | 2013-10-29 07:53:12.7300000 | 2013-10-29 07:58:22.1900000 | 0       |
| 19           | DFC69DC4... | 1            | 6ED4DF44... | 2014-01-28 08:57:55.0340000 | CEB70EF... | 5         | NULL | CEB70EF7-B...  | 2014-01-28 18:05:48.0930000 | 2014-01-28 18:05:48.0930000 | 0       |
| 20           | DFC69DC4... | 1            | B1E73E8F... | 2014-01-28 08:57:38.5410000 | CEB70EF... | 5         | NULL | CEB70EF7-B...  | 2014-01-28 18:05:38.3570000 | 2014-01-28 18:05:38.3570000 | 0       |

Illustration 8: Exempel på hur en avvikelse ser ut i databasen. Skärmdump från Microsoft SQL Management Studio 2014.

**Tabell 6** Problemämnena

| <b>Avvikelser</b> | <b>Beskrivning</b>   |
|-------------------|--|
| Trasigt kärl      | Rapportering av vilka kärl som är trasiga.   |
| Felsorterad       | När kärl innefattar material som inte ska finnas i det givna kärlet.               |
| Fel kärlstorlek   | Det är inte samma kärlstorlek utplacerade som står i GIS.                          |
| Felaktigt stopp   | Sopbilen har i onödan stannat vid ett kärl som inte ska tömmas vid denna tidpunkt. |
| Blockerat         | De kärl som är blockerade och inte kan tömmas.                                     |
| Extra säckar      | De kärl som behöver extra säckar att lägga soporna i.                              |
| Ej utställt       | Kärl som ej är utställda för tömning vid rätt tidpunkt.                            |
| Överfullt kärl    | Kärl som är överfulla.   |
| Annan orsak       | Avvikelser som inte finns fördefinierade i systemet.                               |

### 5.2.1 Kärlbeskrivning

Det finns ett flertalet olika kärl då olika kunder har olika behov för avfall. Varje typ av kärl har en egen beteckning vilket förklaras nedan.

- K = Kärl
- C = Container
- Siffran = Antalet liter
- mat = matavfall
- rest = brännbart avfall
- umat = utan matabonnemang
- HT = månadstömning

De större kärnen är oftast representerade av industrier medan mindre är för hushåll. Vid skrivande stund kan endast hushåll ha kärl angående mat, det vill säga mat och umat. Exempel på kärl är ”190 rest” vilket är ett kärl med 190 liter och avser brännbart avfall.

### 5.2.2 Väder

Utifrån avvikelserapporterna förekommer även väderrelaterade problem, bland annat fastfrusna tunnor och halka, vilket även ansågs vara ett problem av sobilschaufförerna. En

av de externa datorkällorna som kommer att användas är SMHI öppna väderdata för att kunna se vilka kopplingar vädret har med antalet avvikelserapporteringar. Formatet på datafilen tabellbaserad (csv-fil). Dokumentet innehåller kolumnerna datum (från 1960 fram till första januari 2014), lufttemperatur enligt Ekholm-Modéns formel, tidpunkt och kvalitet (godkänd data mot misstänkta värden). Utöver detta fanns det även beskrivningar i text om tabellerna och om väderstationen. Illustration 9 visar hur dokumentet ser ut.

|    | A   | B                   | C                            | D                       | E                        |
|----|---|---------------------|------------------------------|-------------------------|--------------------------|
| 1  | Stationsnamn  | Klimatnummer        | Mätthöjd (meter över marken) |                         |                          |
| 2  | Skövde  | 83230               | 2.0                          |                         |                          |
| 3  |   |                     |                              |                         |                          |
| 4  | Parameternamn   | Beskrivning         | Enhet                        |                         |                          |
| 5  | Lufttemperatur  | Temperatur          | dygnsmedel Ekholm-Moden      | 1 per dygn 00           | degree celsius           |
| 6  |   |                     |                              |                         |                          |
| 7  | Tidsperiod (fr.o.m)   | Tidsperiod (t.o.m)  | Höjd (meter över havet)      | Latitud (decimalgrader) | Longitud (decimalgrader) |
| 8  | 1961-01-01 00:00:00   | 2014-04-30 00:00:00 | 150.0                        | 58.3949                 | 13.8436                  |
| 9  |   |                     |                              |                         |                          |
| 10 | Tidsutsnitt:  |                     |                              |                         |                          |
| 11 | Flödet innehåller alla data som genomgått kvalitetskontroll. För data från de tre senaste hela kalendermånaderna pågår rättningsprocessen. Dessa ingår därför inte. |                     |                              |                         |                          |
| 12 | Tidsperiod (fr.o.m.) = 1961-01-01 00:00:00 (UTC)  |                     |                              |                         |                          |
| 13 | Tidsperiod (t.o.m.) = 2014-02-01 06:00:00 (UTC)   |                     |                              |                         |                          |
| 14 | Samplingstid = 24 timmar  |                     |                              |                         |                          |
| 15 |   |                     |                              |                         |                          |
| 16 | <u>Kvalitetskoderna:</u>  |                     |                              |                         |                          |
| 17 | Grön (G) = Kontrollerade och godkända värden.   |                     |                              |                         |                          |
| 18 | Gul (Y) = Misstänkta eller aggregerade värden. Grovt kontrollerade arkivdata och okontrollerade realtidsdata (senaste 2 tim).                                       |                     |                              |                         |                          |
| 19 |   |                     |                              |                         |                          |
| 20 | Datum   | Tid (UTC)           | Värde                        | Kvalitet                |                          |
| 21 | 1961-01-02  | 00:00:00            | -0.6                         | Y                       |                          |
| 22 | 1961-01-03  | 00:00:00            | 0.4                          | Y                       |                          |
| 23 | 1961-01-04  | 00:00:00            | 0.8                          | Y                       |                          |
| 24 | 1961-01-05  | 00:00:00            | 0.3                          | Y                       |                          |
| 25 | 1961-01-06  | 00:00:00            | -0.1                         | Y                       |                          |
| 26 | 1961-01-07  | 00:00:00            | -1.5                         | Y                       |                          |
| 27 | 1961-01-08  | 00:00:00            | -1.0                         | Y                       |                          |
| 28 | 1961-01-09  | 00:00:00            | -3.6                         | Y                       |                          |
| 29 | 1961-01-10  | 00:00:00            | -1.8                         | Y                       |                          |
| 30 | 1961-01-11  | 00:00:00            | -0.7                         | Y                       |                          |
| 31 | 1961-01-12  | 00:00:00            | -4.6                         | Y                       |                          |
| 32 | 1961-01-13  | 00:00:00            | -5.8                         | Y                       |                          |
| 33 | 1961-01-14  | 00:00:00            | 1.4                          | Y                       |                          |
| 34 | 1961-01-15  | 00:00:00            | 0.5                          | Y                       |                          |
| 35 | 1961-01-16  | 00:00:00            | 2.7                          | Y                       |                          |
| 36 | 1961-01-17  | 00:00:00            | 0.4                          | Y                       |                          |
| 37 | 1961-01-18  | 00:00:00            | -2.4                         | Y                       |                          |
| 38 | 1961-01-19  | 00:00:00            | -4.3                         | Y                       |                          |
| 39 | 1961-01-20  | 00:00:00            | -7.2                         | Y                       |                          |
| 40 | 1961-01-21  | 00:00:00            | -8.7                         | Y                       |                          |

Illustration 9: Väderdata över Skövde från SMHI. Skärmdump från LibreOffice Calc

## 6 Analys

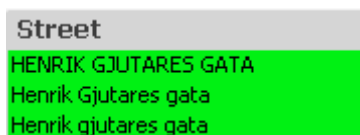
*Detta kapitel analyserar den insamlade datan och går igenom steg för steg hur den gick till och varför på detta vis.*

### 6.1 Utförande av datatvättning

Rahm och Do (2000) "Data cleaning approaches" kommer att användas för datatvättningen vilket har beskrivits under metodkapitlet. Det skedde flera hundra olika typer av tvättningar på all data och tusentals attribut ändrades. Alla ändringar kan inte antecknas utan två klustringar, en manuell splitt och en med olika versaler har valts för att visa några olika typer av fel. Tvättningen är nödvändig för att få en bättre datakvalité till analysen av avvikelserna.

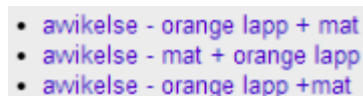
#### 6.1.1 Dataanalys

Genom att kolla på databasen och dess tabeller upptäcktes det förekomster av både fel och motsägelser, framförallt i "Issueheaders" och "Locations". Tabellerna "Issuefacility" och "Facilities" är formella och statiska då datan är autogenererad eller skapats en gång utifrån andra parametrar eller databaser, detta gör att ingen tvättning behövs i dessa tabeller då inga manuella inlägg har lagts till. I "Issueheaders" och "Locations" finns också både formell och statistisk data, men även manuellt inskriven. Den manuella texten är skriven av sopsbilschaufförerna som utfört själva avvikelserna. En databas hanterar text olika om den inte är skriven på ett exakt likadant vis, detta gör att både versaler och blanksteg räknas som ett annat tecken och bildar på så vis ett nytt attribut. När datan skrivs in manuellt uppstår problem då varje person skriver vad de anser vara rätt, till exempel börjar vissa gatunamnen med stor bokstav och andra inte. Illustration 10 visar på hur databasen har delat upp en och samma gata till tre olika på grund av felaktig inmatning av adress. Den första gatan är bara skriven i versaler, den andra har versaler på varje namn, och den sista har endast versal på det första namnet. Illustration 11 visar hur en och samma text kan skrivas på flera olika vis beroende på vilken person som rapporterar felet, eller om samma person skriver olika för olika gånger han eller hon upptäcker felet. Bägge dessa exempel bildar var sina attribut i en dataanalys och är på så vis dåligt datakvalité då det finns redundans.



Street  
HENRIK GJUTARES GATA  
Henrik Gjutares gata  
Henrik gjutarens gata

*Illustration 10: Exempel på samma gata som har olika versaler vilket gör den till tre olika. Skärmdump från Qlikview.*



- awikelse - orange lapp + mat
- awikelse - mat + orange lapp
- awikelse - orange lapp +mat

*Illustration 11: Exempel på data som betyder samma men står olika skrivit. Skärmdump från OpenRefine.*

Ett annat datatvättningsproblem som identifierades var som Ponnier (2010) kallar ”Splitting of Single Fields” vilket är när det står flera attribut i samma fält när de egentligen kan vara en egen. Illustration 12 visar hur fälten innehåller flera olika attribut, både datum, klockslag och delsekunder.

| Created                        | Modified                       |
|--------------------------------|--------------------------------|
| 2013-03-21<br>12:38:16.5070000 | 2013-03-21<br>12:39:46.5000000 |
| 2013-03-21<br>12:42:36.3930000 | 2013-03-21<br>12:44:46.3370000 |
| 2013-03-22<br>07:55:57.8400000 | 2013-03-22<br>07:55:57.8400000 |
| 2013-03-22<br>09:59:42.3000000 | 2013-03-22<br>09:59:42.3000000 |
| 2013-03-22<br>09:59:47.3170000 | 2013-03-22<br>09:59:49.5930000 |
| 2013-03-22<br>09:59:47.5670000 | 2013-03-22<br>09:59:49.4700000 |
| 2013-04-08<br>11:24:13.6230000 | 2013-04-10<br>13:30:48.2900000 |
| 2013-04-17<br>08:06:27.8900000 | 2013-04-17<br>08:06:27.8900000 |
| 2013-04-18<br>12:47:59.4670000 | 2013-04-18<br>12:48:03.4000000 |
| 2013-04-18<br>12:48:06.1970000 | 2013-04-18<br>12:48:12.5770000 |

*Illustration 12: Ett fält som innefattar flera attribut. Skärmdump från OpenRefine.*

Med hjälp av datatvättningsverktyget OpenRefine användes de regler som tagits fram för att ta bort bland annat ta bort felen i illustration 10 och 11. För illustration 10 ändrades så att alla namn står med små bokstäver vilket gör att de blir ett attribut då det står likadant, och i illustration 11 användes klustringen ”key collision” vilket tar upp den data som står exakt likadant fast i olika ordningar och omvandlar den till likadan. Den andre metoden i klustringen ”nearest neighbor” går ut på att leta upp liknande ord eller bokstäver som skiljer sig lite från varandra och slå ihop dessa till en, till exempel ”felsorterad” och ”felsorterat” enligt illustration 13. Alla dessa tvättningsmetoder letar efter redundans vilket är ett av kraven för att få så bra datakvalité som möjligt.

För att utgöra en split utifrån Illustration 12 användes regeln att dela upp i kolumner utifrån ett visst tecken, vilket är ett kommando i OpenRefine. I detta fall skilde det ett mellanrum mellan datum och tid, och en punkt mellan sekund och delsekund. Genom att dela upp efter dessa identifikationer skapades det tre nya kolumner (datum, klockslag och delsekunder) vilket syns i illustration 14. Delsekunder anses vara orelevanta när det gäller avvikelser och togs på så vis bort för att få en bättre datakvalité.

## Cluster & Edit column "Subject"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method  Distance Function  Radius  Block Chars  15 clusters found

| Cluster Size | Row Count | Values in Cluster   | Merge?                              | New Cell Value                  |
|--------------|-----------|---|-------------------------------------|---------------------------------|
| 2            | 14        | <ul style="list-style-type: none"> <li>• awikelse - orange lapp + mat (13 rows)</li> <li>• awikelse - orangelapp + mat (1 rows)</li> </ul>                  | <input checked="" type="checkbox"/> | awikelse - orange lapp + mat    |
| 2            | 2         | <ul style="list-style-type: none"> <li>• awikelse - annan orsak, trång soprum (1 rows)</li> <li>• awikelse - annan orsak, trångt soprum (1 rows)</li> </ul> | <input checked="" type="checkbox"/> | awikelse - annan orsak, trång s |
| 2            | 208       | <ul style="list-style-type: none"> <li>• awikelse - felsorterad (180 rows)</li> <li>• awikelse - felsorterat (28 rows)</li> </ul>                           | <input checked="" type="checkbox"/> | awikelse - felsorterad          |
| 2            | 2         | <ul style="list-style-type: none"> <li>• trasigt hjul (1 rows)</li> <li>• trasiga hjul (1 rows)</li> </ul>  | <input checked="" type="checkbox"/> | trasigt hjul                    |
| 2            | 3         | <ul style="list-style-type: none"> <li>• andra kärstorlek (2 rows)</li> <li>• andra kärstorlek (1 rows)</li> </ul>  | <input checked="" type="checkbox"/> | andra kärstorlek                |
| 2            | 4         | <ul style="list-style-type: none"> <li>• byt kärstorlek (2 rows)</li> <li>• byt kärstorlek (2 rows)</li> </ul>  | <input checked="" type="checkbox"/> | byt kärstorlek                  |

# Rows in Cluster

Average Length of Choices

Length Variance of Choices

Select All

Illustration 13: Skärmdump från OpenRefine vilket visar vilka rader som är liknande genom klustret "Nearest Neighbor".

| Created                        | CreatedHour | CreatedDay |
|--------------------------------|-------------|------------|
| 2013-03-21<br>12:38:16.5070000 | 12:38:16    | 2013-03-21 |
| 2013-03-21<br>12:42:36.3930000 | 12:42:36    | 2013-03-21 |
| 2013-03-22<br>07:55:57.8400000 | 07:55:57    | 2013-03-22 |
| 2013-03-22<br>09:59:42.3000000 | 09:59:42    | 2013-03-22 |
| 2013-03-22<br>09:59:47.3170000 | 09:59:47    | 2013-03-22 |
| 2013-03-22<br>09:59:47.5670000 | 09:59:47    | 2013-03-22 |
| 2013-04-08<br>11:24:13.6230000 | 11:24:13    | 2013-04-08 |
| 2013-04-17<br>08:06:27.8900000 | 08:06:27    | 2013-04-17 |
| 2013-04-18<br>12:47:59.4670000 | 12:47:59    | 2013-04-18 |
| 2013-04-18<br>12:48:06.1970000 | 12:48:06    | 2013-04-18 |

Illustration 14: Före och efter en split har gjorts. Skärmdump från OpenRefine.



Väderdatan var redan i en tabellbaserad fil, dock behövdes det ändå datatvättning för att få den att passa med övriga data. Enligt dokumentbeskrivningen hade alla dagar misstänkta eller aggregerade värden vilket går emot kriteriet ”tillförlitlighet” under datakvalité, dock kommer denna data ändå att användas för att i alla fall få en ungefärlig bild hur temperaturen var. Det som gjordes var att ta bort kommentarer och orelevant data från dokumentet så att endast temperatur och datum fanns med. Tidpunkten togs bort då temperaturen redan är uträknad utifrån Ekholm-Modéns formel och används inte ens. Dokumentet innehöll även data från 1960 vilket rensades bort då den utifrån datakvalité inte är relevant på grund av sopherteringsystemet först börjades att använda våren 2013. När den relevanta datan var utvald gjordes datatvättningen på denna. Den andra delen som tvättades var temperaturen. Temperaturen stod i decimalform och bildade på så vis många attribut som i stort sätt är samma temperatur. Istället för att avrunda temperaturen i källdata så användes funktionen ”.round()” vilket automatiskt avrundar till närmsta heltal. På så vis finns även ursprungsdata kvar och kan vid behov även komma åt originalvärdet. Datumet var skrivit på ett likadant sätt som i AÖS databas vilket uppfyller kravet konsistent under datakvalité. Den enda nackdelen med denna väderdata är att de senaste tre månaderna inte finns med. När denna data hämtades var det sista datumet den första januari 2014. Detta gör att vintermånaden januari saknas vilket var den kallaste månaden denna vinter.

### **6.1.2 Verifiera**

När reglerna var uppsatta gjordes tester för att se om reglerna var rätt och om det önskade resultatet uppfylldes. Vid varje regel kollades det om testresultatet inte skapade några mer motsägelser eller fel, och om detta gjordes så skapades det nya regler för att få det önskade resultatet. Det gavs extra stor vikt på att verifiera klustringen ”nearest neighbor” då den vill slå ihop andra ord och tecken som inte alltid stämmer. Till exempel ansåg regeln att ”brunt kär!” och ”grönt kär!” borde slås ihop till endast ”brunt kär!” vilket inte betyder samma sak då olika färger har olika betydelse.

### **6.1.3 Transformer**

När alla regler var verifierade utfördes själva tvättningen eller transformering av den felaktiga datan. Den största förändringen var i kolumnen ”Postaltown” där staden Skövde stod i flera varianter. Genom att bara byta till samma versaler på texten kunde ändringar på 311 attribut ändras så att de fick samma versaler som de övriga 628 som hade skrivits mest. I vissa tabeller ökades datakvaliteten med upp mot 50 % då flera attribut som betydde likadant kunde slås ihop till en.

### **6.1.4 Ändra äldre data**

Då denna data är hämtad av en backup-fil så kommer inte denna datatvättning inte att påverka det riktiga systemet. De tvättade filerna kommer ändå att skickas in i databasen för principens skull. Varje regel för tvättningen har sparats ner vilket gör att det kan ske på det riktiga systemet vid ett eventuellt behov.

## 6.2 Analys av avvikelser

När datan är tvättad sker analysen på avvikelserna i tabell 6 för att ta fram beslutsunderlag. För varje analys börjas det med att kolla på en generell nivå (postnummer, hela året) för att sedan gå in på en mer detaljerad nivå (gata, månad). Genom att börja mer övergripande går det att få en bra överbild för att sedan grota ner sig i själva detaljerna vid behov. Månaderna mars, april och maj 2013 är inte med i analysen då det inte anses finns tillräckligt med data från dessa. Juni anses ha tillräckligt med data, dock har den ungefär 30 % mindre data än från de resterande månaderna vilket gör att den inte blir lika relevant. Månaderna januari, februari är från 2014 och övriga är från 2013. Alla procenttal är avrundade till närmsta heltal. Fullständiga adresser och postnummer kommer inte att skrivas ut på grund av etiska skäl utan kommer istället presentera vilken typ av kunder som bor i dessa (hushåll, gårdar och industrier/affärer).

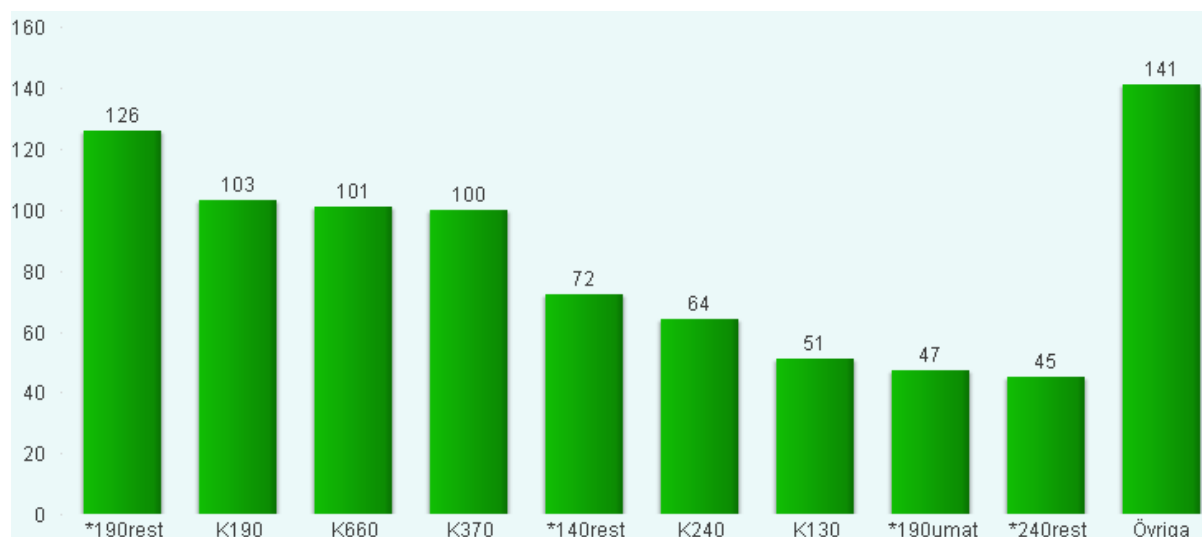
Vilken typ av kunder har tagits fram genom att analysera vilka fastigheter som ligger på respektive postnummer och gata.

### 6.2.1 Trasigt kärl

Utifrån postnummer är det två stycken som sticker ut mer än de övriga. Dessa två postnummer är mindre samhällen på stor yta och innefattar både vanliga hushåll och gårdar. De postnummer som följer dessa är industri- och affärsområden, vilket oftast har andra och fler sorters sopor än hushållen. Det går se en tydlig trend att det är industrier och landet som har de största problemen med trasiga kärl.

Genom att gå in på gatunivå är det en gata som sticker ut med nästan dubbelt mycket som tvåan. Bägge dessa gator har gemensamt att de ligger i ett område med affärer, vilket stärker sambandet med postnummer.

De kärl som ligger högst är 190 rest, K190, K660 och K370. Att kärlet K660 är bland de högsta stärker tesen att industrier/affärer står för flertalet av trasiga kärl då detta kärl endast står för 5 % av de totalt antal kärl som finns. Kärl 190 rest och K190 är bland mest vanliga kärnen och står i förhållanden inte lika högt upp, dessa kärl kan dock användas hos både industrier, hushåll och landet. Endast ett matrelaterat kärl (190 umat) finns med bland topp 10 vilket tyder på att hushållen inte står för stor andel av de trasiga kärnen (Illustration 15).



*Illustration 15: Vilka kärl som är trasiga. Skärmdump från Qlikview.*

En teori innan var att vädret, framförallt kyla, skulle ha en stor påverkan på trasiga kärl. Utifrån SMHI väderdata har det varit en mild vinter, och under kvartal 4 år 2013 har endast två dagar haft medeltemperatur under noll grader. Utifrån dessa två dagar har 8 tunnor rapporterats vara trasiga, det går dock inte att dra någon slutsats om kylan ligger bakom dessa när endast två dagar är inräknade.

Trasigt kärl står för 4,1 % av alla avvikelser.

### 6.2.2 Felsorterad

Det är tre stycken postnummer som står för cirka en tredjedel av alla felsorteringar. Alla dessa postnummer är områden som består utav villor, lägenheter och radhus, det vill säga inga industrier. Utifrån gator är det ingen som sticker ut markant jämfört med de andra. När det gäller kärl så står 190 liter mat för 60 % av hushållen, vilket är den vanligaste soptunnan i bostadsområden. De fem största kärnen som följer är också av typen mat. Denna typ av kärl finns vid skrivande stund endast finns hos hushåll och står för 94 % av de felsorterade (illustration 16). Alla punkter tyder på att hushållen står för den största delen av felsortering.

Felsortering står för 1,4 % av alla avvikelser.

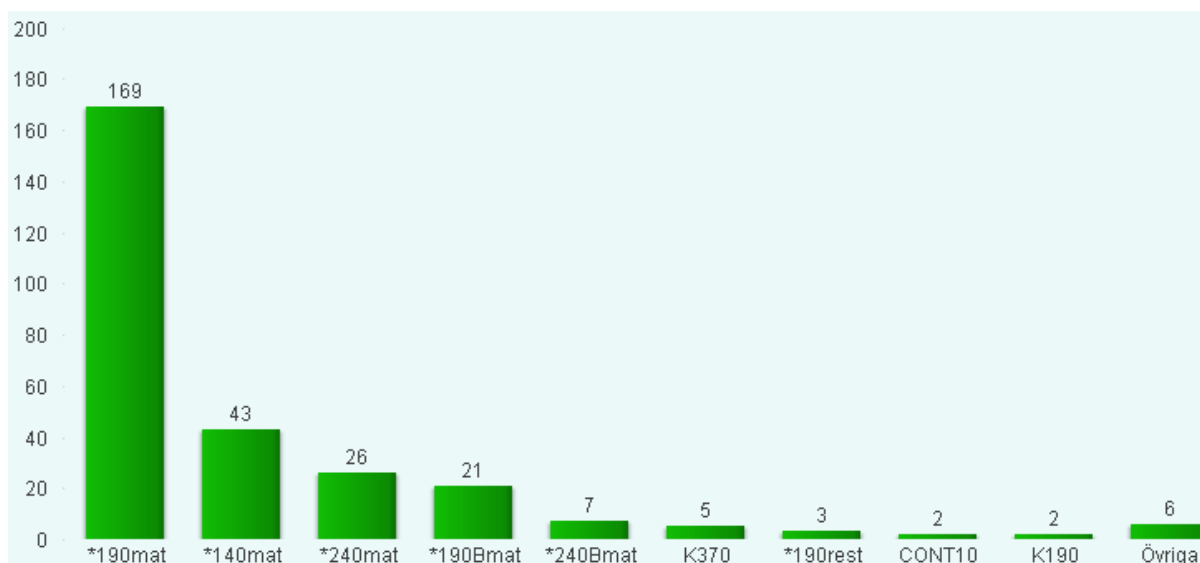


Illustration 16: Vilka kärl som felsorteras mest. Skärmdump från Qlikview.

### 6.2.3 Annan orsak

Utifrån postnummer rapporteras denna typ problem över alla områden och inga gator som har rapporterats markant mer än andra. Det går på så vis inte att ta reda på utifrån adresser vilka som orsakar majoriteten av annan orsak.

Kärlet K660 står för 23 % av annan orsak och 15 % kärlen K190 (Illustration 17). Att kärlet K660 är representerat så högt tyder på att annan orsak oftast förekommer vid industrier/affärer. Detta är logiskt då industrier är mer dynamiska gällande sopor än vanliga hushåll som oftast täcks in i de övriga avvikelserna.

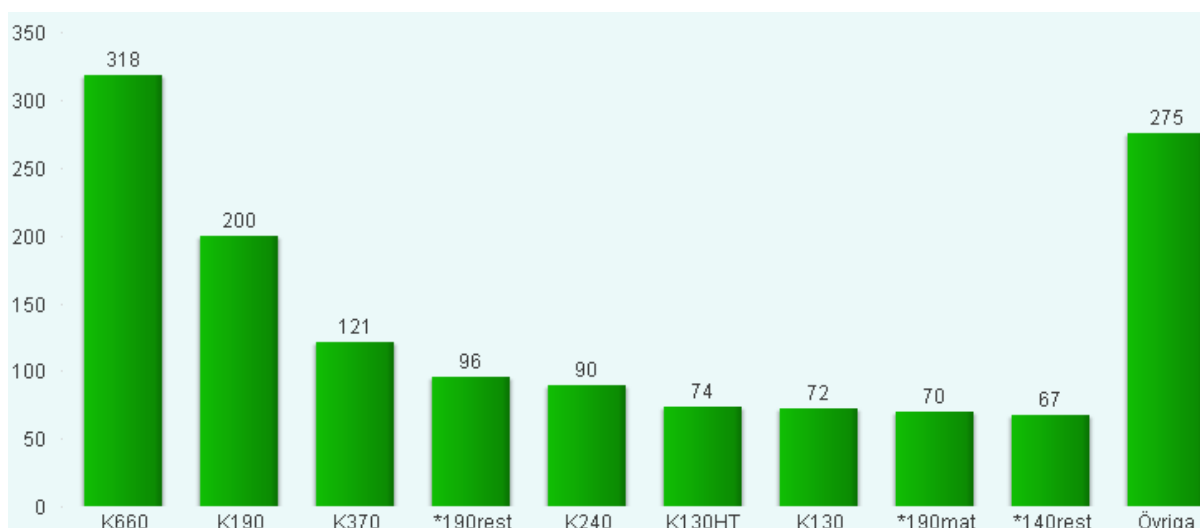


Illustration 17: Vilka tunnor som har rapporterats för annan orsak. Skärmdump från Qlikview.

Annan orsak står för 6,6 % av alla avvikelser.

#### 6.2.4 Fel kärstorlek

Vid analysering av postnummer står fem stycken för 38 % av den totala avvikelserapporteringen varav de två största för 24 %. Gemensamt för de två första är att de är småorter med många gårdar. Vid gatunivå skiljer sig det inte markant mellan gatorna.

De kärll som är mest representerade är vanliga kärll med olika sorters volym (K xxx) där K370 toppar med 17 % (Illustration 18). Dessa kärll är sådana som slänger mer än ett normalt hushåll som oftast har 190 eller 140 liter, vilket kan tyda på gårdar. I stort sett inga matkärll finns med i fel kärllstorlek vilket exkluderar hushåll.

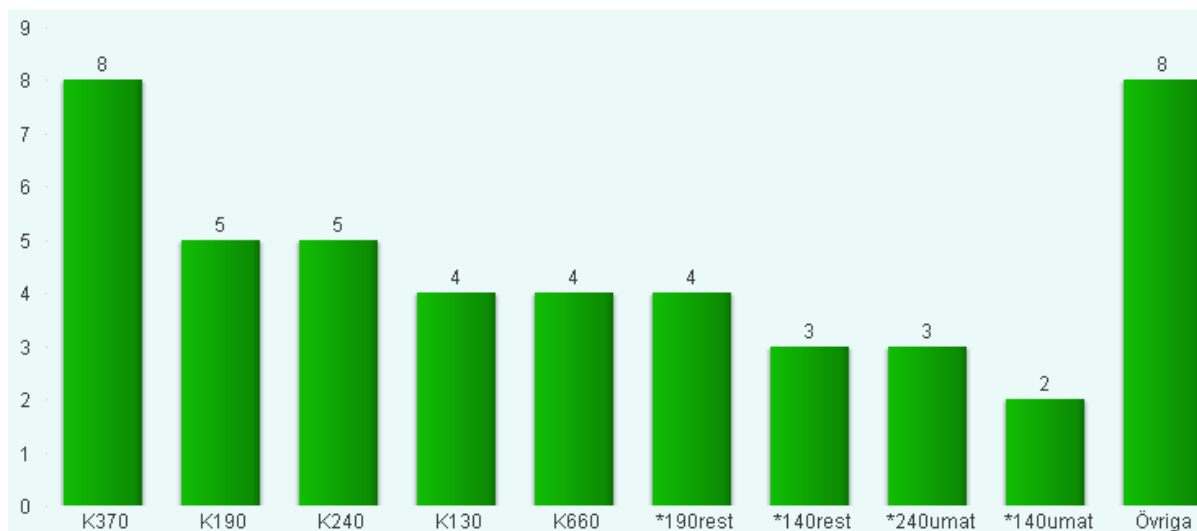


Illustration 18: Vilka tunnor som är fel kärllstorlek. Skärmdump från Qlikview.

Fel kärllstorlek står för 0,2 % av alla avvikelser.

#### 6.2.5 Felaktigt stopp

Vid postnummer är det främst hushållsområden med lägenheter och gårdar. Endast en gata har rapporterats mer markant än de andra, och då rapporterades hela gatan på en och samma dag med kommentaren ”kunde ej tömma idag”. Detta tyder på att det endast var en engångsföreteelse då det inte har hänt igen.

På felaktigt stopp är det flest av kärlet K130HT följt av två matkärll vilket tyder på att det innefattar hushåll (illustration 19).

Felaktig stopp står för 0,9 % av alla avvikelser.

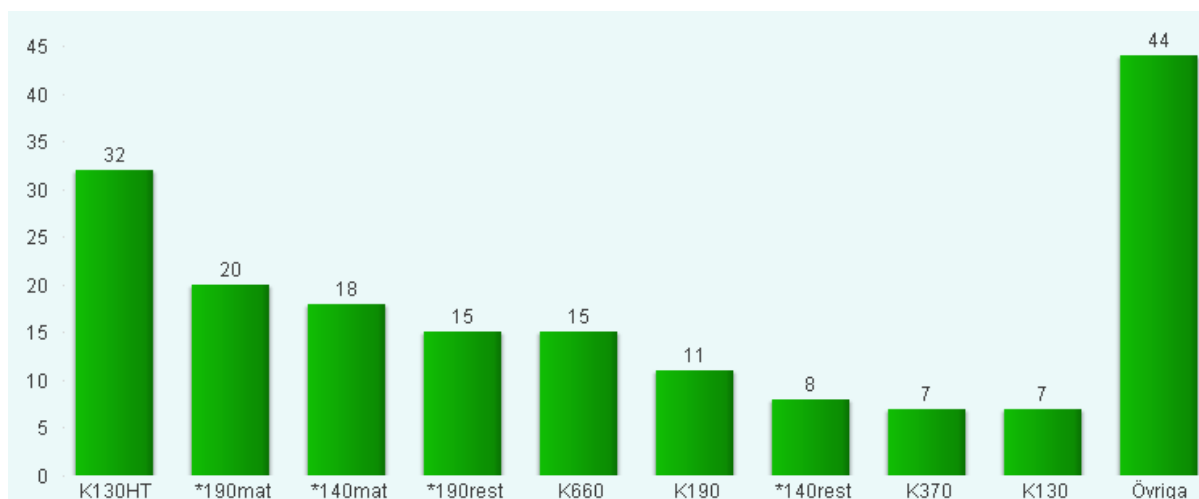


Illustration 19: Vilka tunnor som står för felaktigt stopp. Skärmdump från Qlikview.

### 6.2.6 Blockerat

Utifrån postnummer står både gårdar, industrier och hushållsområden för ungefär lika stor andel. Vid gatunivå står hushållsgator för de som ligger högst i antalet rapporterade blockeringar.

När det gäller kärl står K660 står för 34 % av de blockerade kärlen och K190 för 19 % (illustration 20). Dessa två kärl är främst representerade hos industrier/affärer och har väldigt hög kvot jämfört med totala tunnor, vilket är en motsägelse från gatunivå där väldigt få hushåll använder K660 kärl. De kärl som följer är matkärl och vanliga kärl, dock i små kvantiteter. Gällande blockering är det svårt att hitta någon generell aktör som står för felen utan verkar finnas överallt, dock står K660 för hög andel vilket oftast används i industrier/affärer.

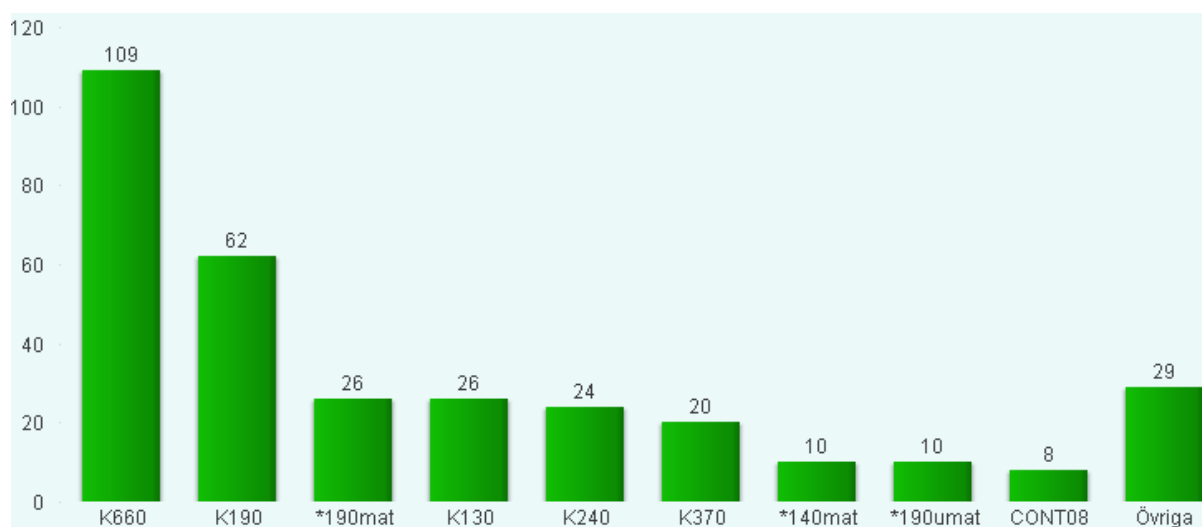


Illustration 20: Vilka tunnor som är mest blockerade. Skärmdump från Qlikview.

Blockerad står för 1,6 % av alla avvikelser.

### 6.2.7 Extra säckar

Vid postnummer står området kring torget för 29 % av de totala extra säckarna. De gator som behöver flest extra säckar är även de inom samma område, vilket är ett affärsområde. K660 står för 59 % av de kärl som behöver extra säckar vilket tyder att industrier/affärer står för majoriteten av dessa rapporter (Illustration 21). Dessa följs av vanliga kärl i varierande volym. Endast 2 % av kärnen av matkärl behöver extra säckar vilket gör att hushåll knappt finns representerad under denna kategori.

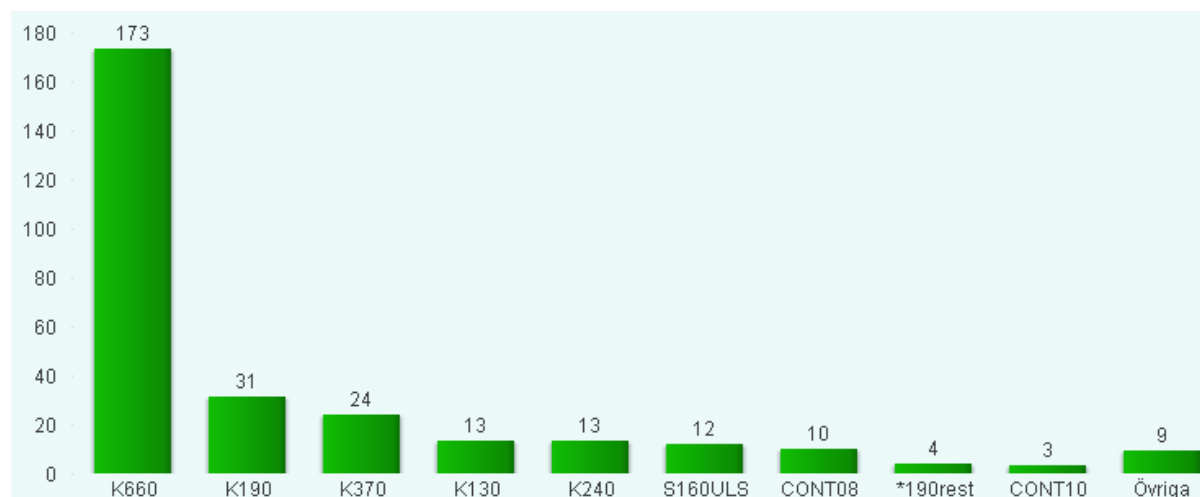


Illustration 21: Vilka kärl som behöver extra säckar. Skärmdump från Qlikview.

Extra säckar står för 1,4 % av alla avvikelser.

### 6.2.8 Ej utställt

Utifrån postnummer är det områden med hushåll, vilket också stärks med gatorna vilket alla i toppen är från hushåll. Kärlet 190mat står för 35 % av de kärl som ej är utställda och 140 mat för 24 % (illustration 22). Vid skrivande stund kan inte industrier ha kärl med anteckningen mat, detta tyder att hushållen står för majoriteten av de kärl som inte ställs fram vilket även kopplas ihop med adresserna. En teori är att flera hushåll inte har sina kärl vid gatan utan måste ställa fram dem innan sopbilen kommer.

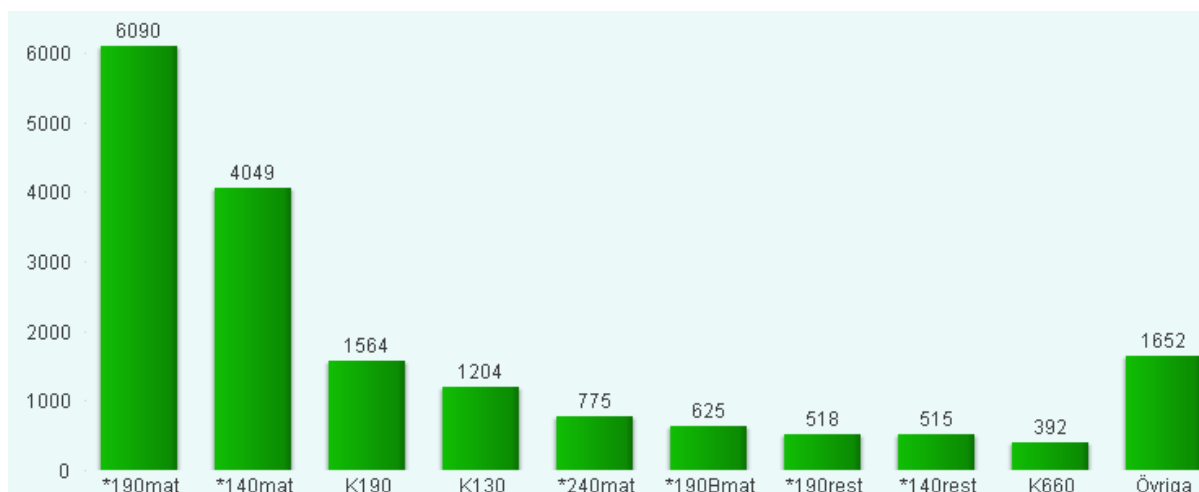


Illustration 22: Vilka kärl som inte blir utställda. Skärmdump från Qlikview.

Ej utställd står för 84% av alla avvikelser och står för den absolut största andelen.

### 6.2.9 Överfullt kärl

De överfulla kärnen finns utifrån postnummer vid gårdar och mindre samhällen. Vid gatunivå är det ingen gata som sticker ut jämfört med andra.

Kärnen K660 står för 23 % av de överfulla kärnen följt av K190 som står för 20 % (illustration 23). Dessa kärl tyder att industrier/gårdar står för majoriteten av de överfulla kärnen, speciellt kärl med matavfall som endast finns hos hushåll knappt är representerade.

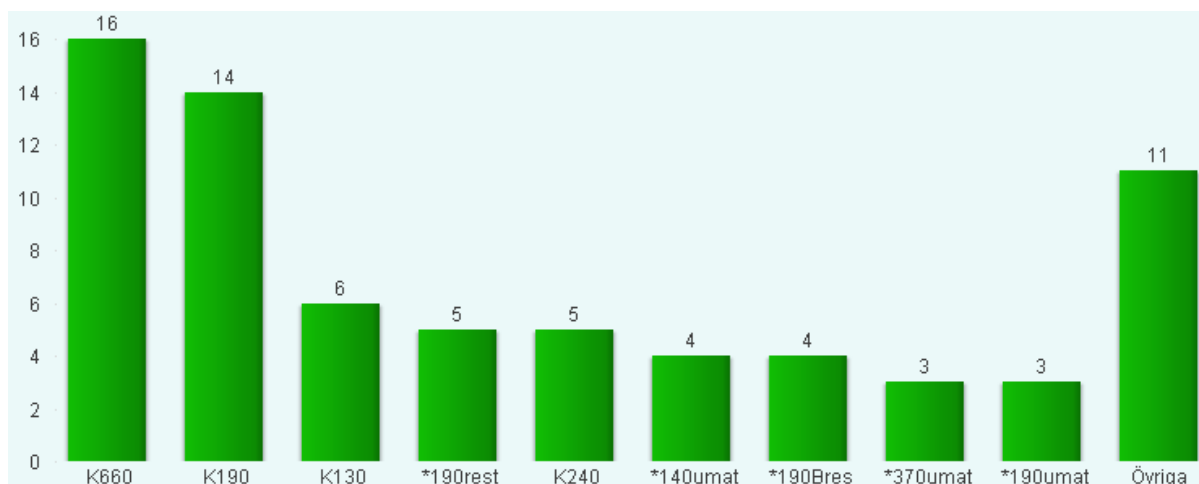


Illustration 23: Vilka kärl som mest är överfulla. Skärmdump från Qlikview.

Överfullt kärl står för 0,3 % av alla avvikelser.

### 6.2.10 Sammanfattning av avvikelser

Tabell 7 visar vilka typer av fastigheter som står för respektive avvikelse. Tre cirklar betyder stark koppling (röd), två cirklar betyder medel koppling (gul), och en cirkel betyder svag koppling (grön). Tabell 8 visar sammanfattning av de olika problemområdena och vilka som orsakar dessa.



**Tabell 7** Sammanfattning av avvikelser - kunder

|                 | Hushåll | Gårdar | Industrier/affärer |
|-----------------|---------|--------|--------------------|
| Trasigt kärl    | ●       | ● ● ●  | ● ● ●              |
| Felsorterad     | ● ● ●   | ●      | ●                  |
| Annan orsak     | ● ●     | ● ●    | ● ●                |
| Fel kärlstorlek | ●       | ● ● ●  | ● ●                |
| Felaktigt stopp | ● ●     | ● ● ●  | ● ●                |
| Blockerat       | ● ●     | ● ●    | ● ●                |
| Extra säckar    | ●       | ●      | ● ● ●              |
| Ej utställt     | ● ● ●   | ●      | ●                  |
| Överfullt kärl  | ●       | ● ● ●  | ●                  |

**Tabell 8** Sammanfattning av avvikelser – kunder och kärl

|                        |  |
|------------------------|--|
| <b>Trasigt kärl</b>    | Kärnen 190 rest, K190 och K660, främst inom industri och gårdar.                     |
| <b>Felsorterad</b>     | Hushåll med kärl för matavfall   |
| <b>Annan orsak</b>     | Större kärl än normalt.  |
| <b>Fel kärlstorlek</b> | K370 och K240 och områden där matkärl inte används, främst gårdar.                   |
| <b>Felaktigt stopp</b> | Kärnen K130HT och matavfallskärl, främst vid gårdar men även hushåll och industrier. |
| <b>Blockerat</b>       | Främst på kärnen K660 men utspritt över gårdar, industrier och hushåll.              |
| <b>Extra säckar</b>    | Butiker/industrier med kärlet K660.  |
| <b>Ej utställt</b>     | Hushållsområden som har kärl med matavfall.  |
| <b>Överfullt kärl</b>  | Gårdar med kärnen K660 och K190  |

## 7 Slutsats

*Detta kapitel svarar på den problemformulering vilket ställdes tidigare i studien.*

Den här studien handlar om att analysera avvikelserapporter inom avfallshanteringsbranschen för att kunna ta fram beslutsunderlag som ska kunna effektivisera sophämtningen. Problemformuleringen som ställdes var:

*Hur kan avfallshanteringsbranschen effektivisera sophämtningen med hjälp av en Business Intelligence-lösning?*

Denna problemformulering skulle lösas med hjälp av tre punkter:

1. Göra en kvalitetsgranskning på befintlig data.
  - a) Gå igenom en ETL-process.
2. Analysera befintlig data för att ta fram beslutsunderlag.
3. Ta fram rekommendationer för Avfallshantering Östra Skaraborg.

Den första punkten beskrivs i kapitel 6.1 där en kvalitetsgranskning av datan sker för att hitta felaktiga och motsägelsefull data. Genom att tvätta den data som är fel kommer leda till en effektivare analys då datan har fått bättre datakvalité. Steg två påbörjas i kapitel 6.2 där en analys sker på den kvalitetsgranskade datan för att ta fram beslutsunderlag angående vilka grupperingar som står för respektive fel. Utifrån resultatet från analysen visar att det inte bara en speciell grupp som står för alla problem utan att det är utspridda beroende på vilken typ av problem det är. Det går alltså att inte dra en generell slutsats för vilka som står för alla avvikelser. Utifrån beslutsunderlaget i fråga punkt 2 har det tagits fram rekommendationer för varje avvikelse till Avfallshantering Östra Skaraborg. Varje rekommendation beskrivs i tabell 9. Tabellen är sorterad utifrån de avvikelser som rapporterats in mest och bör prioriteras. De avvikelser som inte har någon rekommendation hamnar längst ned.

**Tabell 9** Rekommendationer att göra för att få en effektivare sophantering.

|                     |  |
|---------------------|--|
| <b>Ej utställt</b>  | Hushållsområden som har kärl med matavfall ska få information om vilka regler och tider som gäller angående utställning av kärl.               |
| <b>Trasigt kärl</b> | Tillverka ett nytt extra tåligt kärl vilket kan användas främst av gårdar och industrier som slänger material som kan skada kärlet.            |
| <b>Felsorterad</b>  | Skicka ut information till hushåll med kärl för matavfall angående hur de ska sortera sina sopor och påföljder som kan hända vid felsortering. |
| <b>Extra säckar</b> | Ge butiker med kärlet K660 större kvantitet av säckar så att de inte behöver beställa oftare.  |

|                             |  |
|-----------------------------|--|
| <b>Felaktigt stopp</b>      | Ge information till kunder med månadstömning att den inte kommer att tömmas även fast den står framme. |
| <b>Överfullt kärl</b>       | Gårdar med kärnen K660 och K190 behöver större eller extra kärl för att minska risken för överfyllnad. |
| <b>Felaktig kärlstorlek</b> | Informera gårdar att de måste rapportera byte av kärl.   |
| <b>Blockerat</b>            | Svårt med generell lösning då det finns överallt och på flera kärl.                                    |
| <b>Annan orsak</b>          | Går inte dra någon generell slutsats då det innefattar massor utav olika ickedefinierade fel.          |

Utifrån de här rekommendationerna som är framtagna med hjälp av en BI-lösning kan de hjälpa avfallshanteringsbranschen att få en effektivare sophantering.

## 8 Diskussion

Att analysera avvikelser inom avfallshanteringen blev en väldigt intressant studie med många lärdomar. Alla hushåll och företag använder sig av sopor och vi är beroende att det sköts på ett korrekt sätt. Först i och med denna studie har jag insett att sophantering är mer avancerat när jag väl har blivit insatt i området, samt att jag fått en insikt att det finns väldigt mycket problem som hindrar att sophantering sker felfritt. Det har varit ett givande projekt där jag hoppas att mitt resultat även ger ett bättre resultat ute i fältet. Det finns förhoppningar med resultatet att förbättra det både för arbetarna inom branschen samt andra intressenter som kunderna, och även minskad effekt på miljön.

I denna studie var det flera steg som skulle genomföras för att få fram resultatet. Första steget var att leta efter liknande projekt om någon hade gjort något liknande. Efter att gått igenom tänkbara litteraturlösningsdatabaser, bland annat flera som inriktar sig just på sophantering, har det inte hittats någon som gjort ett liknande projekt med analys av avvikelser inom sophanteringsbranschen. När inget liknande hade hittats kändes det bra att denna studie kunde bidra med något nytt i forskning som inte bara hjälper sophantering utan även miljön. Även om problemet inte finns rapporterat av andra inom branschen känns det som att alla kommuner drabbas av att soptunnor inte går att tömma och kan på så vis använda sig av resultatet. De kan även använda sig av samma forskning om de också samlar in data om avvikelserrapporter vilket leder till att studien kan användas nationellt.

Det praktiska arbetet gick generellt sätt bra och uppsättningen av en databas gick utan större problem. Med hjälp av de inbyggda verktygen i databashanteraren kunde det snabbt sorteras ut vilken data som var orelevant och vilka som var betydelsefulla. Resultatet av datavättningen blev också den lyckad men med vissa svårigheter. När det fanns såpass mycket rådata tog tvättningen majoriteten av tiden med att identifiera alla problem samt definiera de algoritmer för att omvandla datan till rätt format. Även om detta tog väldigt stor del av tiden blev resultatet bra, och uppemot 50 % av datan i vissa tabeller kunde tvättas utan att förlora någon betydelsefull data. En del av datan var svårtolkad vid tvättningen vilket i vissa fall gjorde att den lämnades för att undvika ändring av korrekt data. Den data som var tveksam fanns det inga större kvantiteter utav och gjorde på så vis ingen större skillnad i analysen.

När allt detta var klart var det dags för analysen vilket ibland gav vissa motsägelser, till exempel att avvikelsen ”ej utställt” resulterade i felärendet ”extra säckar”. Till stor del av detta är på grund av att den data som användes i arbetet är fortfarande under utvecklingen av verksamheten och av mänskliga fel hos chaufförerna. Deras system har inte ens varit ute i ett år för testning, vilket leder till att resultatet är svårt att analysera då det skiljer sig så mycket från månad till månad. Mellan månaderna har nya saker både lagts till och tagits bort då det fortfarande är under utveckling. När antalet avvikelser från maj till juni ökar med 500 % för att sedan från augusti till september öka med ytterligare 390 % blir resultatet väldigt olika under dessa månader. Vissa ämnen har inte heller funnits med från början vilket gör att det går från inget till massor på kort tid.

I början på analysfasen analyserades fel data då jag gick ett steg för djupt i hierarkin över avvikelser vilket resulterade att en stor del av tiden gick till spillo på fel data. När detta identifierades i analysverktyget att datan såg felaktig ut ändrades parametrarna till de rätta och för säkerhets skull kontrollerades svaren med SQL-frågor direkt i databasen för att kontrollera att inget misstag gjordes igen.

Den information som fanns vid analysen bekräftade i stort sett alltid varandra vilket gjorde analysen lättare, till exempel att felsorterade visade hushåll utifrån postnummer och de kärn som mest var felsorterade var sådana som används i hushåll. Det blev svårt att se trender då systemet är nytt och att varje månad bara finns representerad en gång, speciellt när det är såpass stora ändringar per månad som beskrevs i föregående stycke.

I efterhand skulle mer tid ha lagts på analysen då den nu enligt mig endast ger övergripande statistik. Genom att gå ännu mer på nivån går det att få fram ännu mer statistik om mer exakta saker, till exempel vilken veckodag under kvartal tre sker mest felsorteringar. Denna typ detaljrikedom fanns inte tillräckligt med tid i denna studie.

Vädret som extern källa angavs av chaufförerna som en av problemen av tömningen. Vilket tidigare sagt går det inte få ”real-time” data på vädret utifrån SMHIs öppna väderdatabas, detta leder till att i början av året där det var som kallast inte finns någon statistik över. Därav blev denna externa källa inte till stor nytta under kvartal fyra av 2013 då vädret endast hade minusgrader två gånger under denna tidsperiod. Även här går det inte se om det är en engångsförekomst eller en trend då data inte finns för flera år. De övriga externa faktorerna som sopbilschaufförerna påpekade går det inte att hitta någon bra datakälla på.

## 8.1 Forskningsmetod

Det är första gången jag gjorde intervjuade personer för ett formellt arbete. Generellt sätt gick det bra men vissa saker kunde såhär i efterhand ha gjorts annorlunda, bland annat vara tydligare med definitioner. I början av studien blev det en del missförstånd angående begreppen ”avvikelse” och ”felrapport” då jag tolkade de likadant. Först en bit in i studien via en chatintervju kom det fram till att det var en stor skillnad på begreppen vilket gjorde att en del av studien behövdes göra om då jag hade analyserat fel data.

I början var det tänkt att även intervju sopbilschaufförerna, men på grund bristfällig kommunikation mellan parterna blev det uppskjutet och sedan var sopbilschaufförerna upptagna. Istället för att göra en intervju gjordes då en enkät med liknande frågor som skickades ut vilket de kunde svara på när de hade tid. Då frågorna handlade om att beskriva processer anses resultatet inte vara någon större skillnad om det skulle sagts muntligt eller skriftligt då hela processförloppet ändå kom med i svaret.

En annan sak som skulle gjorts annorlunda är att vara med och observera sopbilschaufförerna under en vanlig dag. På så vis skulle jag kunnat få mer information om själva processen och kunde räknat ut tiden det tar att rapportera ett fel. Det skulle även kunnat användas vid senare tillfälle för att se hur mycket det har förbättrats gentemot innan.

## 8.2 Etiska, samhällliga och vetenskapliga aspekter

De flesta etiska aspekter har beskrivits i respektive kapitel de används i. Överlag har jag försökt att vara tydlig att endast den nödvändiga informationen ska publiceras för att undvika problem med att information sprids som inte behöver spridas. Varje representant är anonym förutom eventuell titel och ingen data om vilka kunder som står för de flesta felen angående sophantering har angetts. Detsamma gäller geografiska positioner där ingen adress eller postnummer visas för utomstående personer än inom verksamheterna som är inkluderade i studien. Den information som innehåller detta kommer att separat att lämnas in till Avfallshantering Östra Skaraborg och Högskolan i Skövde som sedan själva får bestämma vad som ska bli publikt eller ej.

En effektivare sophantering påverkar samhället till det bättre. Genom att använda sig av rekommendationerna för att förhindra antalet avvikelser behöver sopbilarna inte stanna lika ofta för att identifiera och rapportera avvikelserna vilket tar tid. Att undvika att skriva en avvikelserapport leder till att sopbilen inte behöver stå lika mycket i tomgång vilket stör samhället med både buller från sopbilen och extra utsläpp. Det underlättar också för sopbilschaufförerna då de slipper identifiera och rapportera lika många avvikelser vilket gör deras arbetsprocess mer komplext och tidskrävande.

Utifrån tidigare forskning har det till stor del forskats på ruttoptimering inom GIS och inget om varför soptunnor inte går att tömma. I studiens resultat går det att se att avvikelserapportering krävs på ungefär var tionde soptunna samt att det tar tid att utföra. Denna forskning tar upp en aspekt som tidigare inte forskats på och lyfter fram att det finns en stor potential att effektivare sophantering genom att ta reda på orsakerna till avvikelser. Genom att forska om avvikelser går det också att ta reda på mer hur samhället hanterar sina sopor vilket kan leda till ytterligare effektivisering.

## 8.3 Framtida arbete

Denna analys var som tidigare sagt att kunna motverka framtida problem när deras produkt är släppt i en fullständig version. Efter att den fullständiga versionen har släppts går det att fortsätta göra analyser vilket kommer ge ett bättre resultat då inga större uppdateringar tillkommer vilket ändrar datan. Det kommer också gå att se större trender mellan åren då endast 8 månader finns med i denna analys. Som tidigare nämndes går det även att dyka ner ännu längre i hierarkin för att ta reda på mer specifika problem som vilken dag under sommarperioden står för mest rapporteringar. Externa källor såsom vädret går också att få en tydligare bild utöver när det finns mer data och jämförelse med kalla mot milda vintrar.

En uppskattning var att miljön skulle förbättras genom att minska antalet avvikelser. Då denna studie inte innefattar processen efter ett förslag går det på så vis inte mäta hur miljön påverkas. Ett framtida arbete kan på så vis vara att göra före- och efteranalyser av utsläppen för att räkna ut hur mycket som sparas.

Denna analys utgår bara från Östra Skaraborg vilket är en liten del av Sverige. Genom att analysera övriga områden går det att se likheter och skillnader vilket kan användas för att förbättra resultatet i en annan kommun/område.

## Referenser

Berndtsson, M., Hansson, J., Olsson, B. and Lundell, B. 2008. Computer Science and Information Systems Research Projects. Thesis Projects: *A Guide for Students in Computer Science and Information Systems*, pp. 9—14.

Chrisman, N. (1997). *Exploring geographic information systems*. 1st ed. New York: J. Wiley & Sons.

Dayal, U., Castellanos, M., Simitsis, A. and Wilkinson, K. 2009. Data integration flows for business intelligence. pp. 1--11.

Dicicco-Bloom, B. and Crabtree, B. F. 2006. The qualitative research interview. *Medical education*, 40 (4), pp. 314--321.

Fredriksson, G. 2012. *Insamling och behandling av avfall*. [online] Available at: [http://www.skl.se/vi\\_arbetar\\_med/tillvaxt\\_och\\_samhallsbyggnad/avfall/insamling-och-behandling-av-avfall](http://www.skl.se/vi_arbetar_med/tillvaxt_och_samhallsbyggnad/avfall/insamling-och-behandling-av-avfall) [Accessed: 23 Feb 2014].

Fonseca, F., Egenhofer, M., Agouris, P. and Câmara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), pp.231--257.

Gergen, Kenneth J.; Gergen, Mary M., & Barton, William H. (1973). Deviance in the dark. In *Psychology Today*, 7, 129-130.

Golfarelli, M., Rizzi, S. and Cella, I. 2004. Beyond data warehousing: what's next in business intelligence?. pp. 1--6

IBM. (2013). *What is big data?* [Online], Available at: <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>. Accessed 26 Feb 2014.

Inmon, W. H. 1996. The data warehouse and data mining. *Communications of the ACM*, 39 (11), pp. 49--50

Jain, A., Doan, A. and Gravano, L. 2007. SQL queries over unstructured text databases. pp. 1255--1257.

Jick, T. D. 1979. *Mixing qualitative and quantitative methods: Triangulation in action*. *Administrative science quarterly*, pp. 602—611.

Kopáčková och Škrobáčková. 2006. Decision support systems or business intelligence: what can help in decision making?. *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration. 10 (2006)*.

Moeinaddini, M., Khorasani, N., Danehkar, A., Darvishsefat, A. A. and Others. 2010. Siting MSW landfill using weighted linear combination and analytical hierarchy process (AHP) methodology in GIS environment (case study: Karaj). *Waste management*, 30 (5), pp. 912--920.



- Mooney, R. J. and Bunescu, R. 2005. Mining knowledge from text using information extraction. *ACM SIGKDD explorations newsletter*, 7 (1), pp. 3--10.
- Myers, M. D. 1997. Qualitative research in information systems. *Management Information Systems Quarterly*, 21 pp. 241—242.
- Negash, S. 2004. Business intelligence. *Communications of the Association for Information Systems*, 13 (1), pp. 177—195.
- Negash, Solomon and Gray, Paul, "Business Intelligence" (2003). *AMCIS 2003 Proceedings*. Paper 423.
- Opdenakker, R. 2006. Advantages and disadvantages of four interview techniques in qualitative research. 7 (4).
- Openrefine (2014). *OpenRefine*. [online] Available at: <http://openrefine.org/> [Accessed 25 Apr. 2014].
- Pinsonneault, A. and Kraemer, K. (1993). Survey research methodology in management information systems: an assessment.
- Ponniah, P. 2010. *Data warehousing fundamentals for IT professionals*. Hoboken, N.J.: John Wiley & Sons
- Power, D. J. 2007. A brief history of decision support systems. DSSResources. COM, World Wide Web, <http://DSSResources.com/history/dsshistory.html>, version, 4. [Accessed; 14 Feb 2014]
- Qlikview, (2014). *QlikView*. [online] Available at: <http://www.qlik.com/se> [Accessed 5 May. 2014].
- Rada, E. C., Ragazzi, M. and Fedrizzi, P. 2013. Web-GIS oriented systems viability for municipal solid waste selective collection optimization in developed and transient economies. *Waste management*, 33 (4), pp. 785--792.
- Rahm, E. and Do, H. H. 2000. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23 (4), pp. 3--13.
- Raman, V. and Hellerstein, J. M. 2001. Potter's wheel: An interactive data cleaning system. 1 pp. 381--390.
- Riksdagen. 2013. *Svensk författningssamling 1998:808 Miljöbalk (1998:808)*. [online] Available at: [http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Miljobalk-1998808\\_sfs-1998-808/?bet=1998:808](http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Miljobalk-1998808_sfs-1998-808/?bet=1998:808) [Accessed: 23 Feb 2014].
- Simitsis, A., Vassiliadis, P. and Sellis, T. 2005. Optimizing ETL processes in data warehouses. pp. 564--575.

- Tavares, G., Zsigraiová, Z. and Semiao, V. 2011. Multi-criteria GIS-based siting of an incineration plant for municipal solid waste. *Waste management*, 31 (9), pp. 1960--1972.
- Trujillo, J. and Luján-Mora, S. (2003). A UML based approach for modeling ETL processes in data warehouses. *Springer*, pp.307--320.
- Transportstyrelsen. 2014. *Fordonsstatistik januari 2014*. [online] Available at: <https://www.transportstyrelsen.se/sv/Press/Statistik/Vag/Fordonsstatistik/Fordonsstatistik-januari-2014/> [Accessed: 10 Mar 2014].
- Urban, G. L. 1966. SPRINTER: a tool for new product decision making. [*Cambridge, Mass., MIT*].
- Wang, R. Y. and Strong, D. M. 1996. Beyond accuracy: What data quality means to data consumers. *J. of Management Information Systems*, 12 (4), pp. 5—33.
- Wellington, J. & Szczerbiński, M., 2007. *Research methods for the social sciences*. London: Continuum International Publishing.
- Westermark, L. 2013. *Rekordlåga utsläpp år 2012*. [online] Available at: <http://www.naturvardsverket.se/klimat2012> [Accessed: 2 Mar 2014].
- Zsigraiova, Z., Semiao, V. and Beijoco, F. 2013. Operation costs and pollutant emissions reduction by definition of new collection scheduling and optimization of MSW collection routes using GIS. The case study of Barreiro, Portugal. *Waste management*, 33 (4), pp. 793--806.