

IP-adressen

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	07-10-2013	<div>0.2.e</div>	Initieel document

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Bron data.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	4
3. Beslissingen.....	4

2. BESCHRIJVING

Deze bron data wordt gebruikt in IP_lijst.

De bron data wordt maandelijks geleverd door 10.2.e

De bron data wordt opgeleverd als xlsx-bestand en wordt opgeslagen op het netwerk:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\IP adressen\Source

Vervolgens word de bron data ingelezen in SAS.

2.1. BRON DATA

De bron data van IP adressen:

Bronbestand	Omschrijving	Aantal records
IP_lijst_input.xlsx	Lijst met BSN en IP-adressen	

2.1. VELDEN

Onderstaande velden zijn opgenomen:

IP_adres,
BSN

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

DEELJAAR BESCHIKKING KOT

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	21-05-2013	10.2.e	Initieel document
0.02	24-02-2014		Het programma voor het genereren van deeljaar beschikking en doelgroep heeft een andere naam
1.0	30-11-2015		Informatie aanvullen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving.....	4
2.1.	Bron data.....	4
2.2.	Velden.....	4
2.3.	Indicatoren.....	5
3.	Beslissingen.....	5

2. BESCHRIJVING¹

Deze brondata bevat drie lijsten met deeljaar beschikkingen voor KOT. Deze lijsten worden gebruikt in het halfproduct DTM_KOT.

Maandelijks vraagt het risicoselectieteam de brondata **per mail op bij** 10.2.e Na formeel beschikken wordt de data **door BCA/BICC** opgeleverd als SAS-bestanden en **geplaatst op de netwerkschijf**:

Q:\VEPROW52\TSL_UT_Datawisseling\Noodoplossingen voor EDW 2014.1\Profiling Kot

Deze bestanden heten over het algemeen: 'doelgroep_<datum>', 'kot_deeljaar_bedragen_<datum>' en 'kot_deeljaar_lrk_<datum>'.

Vervolgens wordt de data door het risicoselectieteam **gekopieerd naar een andere locatie**, van waaruit met de data wordt gewerkt:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\ DEELJAAR BESCHIKKING KOT \Source

Op deze locatie zijn de namen van de bronbestanden aangepast tot **'doelgroep'**, **'kot_deeljaar_bedragen'** en **'kot_deeljaar_lrk'**. Deze bestanden worden gebruikt bij het creëren van het halfproduct.

2.1. BRONDATA

De brondata van DEELJAAR BESCHIKKING KOT:

Bronbestand	Omschrijving
doelgroep.sas7bdat	Lijst met deeljaar beschikkingen KOT en indicatie doelgroep en dagtaaktype voor toeslagaanvrager en toeslagpartner
kot_deeljaar_bedragen.sas7bdat	Lijst met deeljaar beschikkingen KOT en bijbehorende bedragen
kot_deeljaar_lrk.sas7bdat	Lijst met deeljaar beschikkingen KOT en bijbehorende gegevens kinderopvanginstellingen

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Name	Type	Length	Format
------	------	--------	--------

¹ Na overgang tot AWS zal het groen gemarkeerde (aanvraag / oplevering / locatie en naam brondata) moeten worden aangepast. Daarnaast zal te zijner tijd de query, gebruikt om de data op te halen uit de database, toegevoegd worden.

Bsn	Numeric	6	11.
Toeslagjaar	Numeric	3	20.
ToetsingsinkomenAanvrager	Numeric	5	20.
Draagkracht	Numeric	5	20.
Maandtoeslag	Numeric	8	14.2
Inkomen_partner	Numeric	5	20.
Toeslagmaandbedrag	Numeric	8	14.2
Soort_instelling_sk	Numeric	3	11.
Type_opvang_sk	Numeric	3	11.
TotaalAantalUrenOpvang	Numeric	3	20.
GemiddeldeUurprijs	Numeric	3	14.2
TotaalAantalUrenNietAfgetopt	Numeric	4	20.
GemiddeldeUurprijsNietAfgetopt	Numeric	3	14.2
Datum_Dagtekening	Character	10	\$10.
Beschikkingnummer	Character	17	\$17.
Bsn_partner	Character	9	\$9.
BsnKind	Character	9	\$9.
Ingangsdatum_Deeljaar_datum	Character	10	\$10.
Einddatum_Deeljaar_datum	Character	10	\$10.
Uren_Afgetopt	Character	1	
Uurtarief_Afgetopt	Character	1	

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

Indeling resultaat

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	30-11-2015		Initieel document
1.1	04-08-2016	10.2.e	AWS-aanpassingen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Bron data.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	5
3. Beslissingen.....	5

2. BESCHRIJVING

Deze brondata bevat een lijst van verschillende karakteristieken van een aanvraag en alle combinaties van mogelijke waarden. Op basis van deze lijst wordt bepaald op welke actielijst een aanvrager komt te staan. Dit bronbestand is op 22 september 2014 gecreëerd door het risicoselectieteam en staat opgeslagen op de AWS:

/prj/bi/data/AD010/data/RisicoSelectie/Brondata/INDELING RESULTAAT

Dit bronbestand wordt tijdens elke risicoselectierun ingelezen in SAS bij het runnen van KOT_Model en HT_Model.

2.1. BRONDATA

De brondata van Indeling resultaat:

Bronbestand	Omschrijving
Schema_wie_op_welke_lijt_v6.1.csv	Lijst met schema dat bepaalt op welke actielijst de aanvrager komt te staan.

2.1. VELDEN

Onderstaande velden zijn opgenomen:

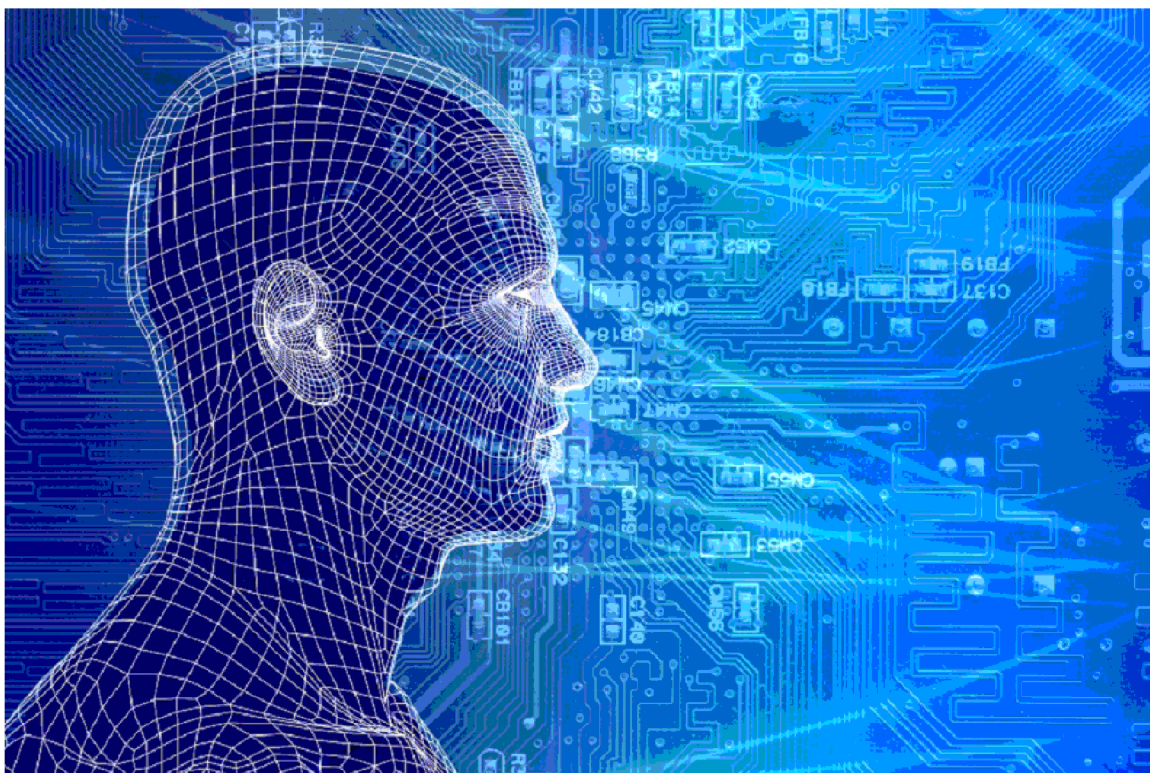
Naam	Getalnotatie (csv)
Type aanvraag	Standaard
Laatste beschikking is concept?	Standaard
Deeljaar nu actueel?	Standaard
Bijvangst?	Standaard
Hothor	Standaard
Onbekende aanvrager	Standaard
Niet uitsluiten	Standaard
Geen bijvangst	Standaard
Geen onbekende aanvrager	Standaard
Actie	Standaard
Opmerking	Standaard
Verklaring	Standaard

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
------	-------------------------	--------------	---

N.v.t.

3. BESLISSINGEN



HT_Tijdreizen

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	20-11-2013	10.2.e	Initieel document
0.02	24-01-2014		Nieuwe maand toegevoegd: 2014-01
0.03	27-01-2014		Opmerkingen MS verwerkt
0.04	10-02-2014		Aanpassing toewijzen cases aan run
1.0	20-12-2015		Wijziging nav overzetten naar AWS en bevroren toeslagjaar 2013.
1.1	07-03-2015		Toevoegen Stappenplan (H6)
1.2	30-11-2018		Nieuwe opzet, 2013 apart

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Doel.....	4
3.	Beschrijving project.....	4
3.1.	Bron data.....	4
3.2.	Flows.....	5
4.	Beschrijving Halfproduct.....	6
4.1.	Autoexec.....	6
4.1.1.	A. Initialize.....	6
4.1.2.	B. Steekproef cases 2013.....	6
4.1.3.	C. Create Testcases mm-YY.....	6
4.1.4.	Z. Append cases.....	6
5.	Update HT_Tijdreizen.....	7
6.	Stappenplan.....	8
6.1.	Programma tijdreizen.....	8
6.2.	Input data toevoegen.....	8
7.	Nieuwe opzet.....	8
8.	Resultaten per tijdreizen.....	10
8.1.	Maart 2019.....	10

2. DOEL

Doel van het tijdreizen is trainingscases voorzien van de informatie zoals deze op het moment van beoordeling (als goed of fout) actueel was.

Dit heeft tevens tot gevolg dat als er nieuwe indicatoren worden toegevoegd aan het model, deze ook met terugwerkende kracht voor alle trainingscases toegevoegd moeten kunnen worden.

3. BESCHRIJVING PROJECT

3.1. BRON DATA

Er is generieke brondata en brondata per risicoselectierun:

- Generiek is het SAS-bestand 'Trainingscases_HT' dat volgt uit het SAS project 'Trainingscases' (Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Trainingscases).
- Daarnaast is er per risicoselectierun die in het verleden is uitgevoerd een tabel met alle indicatoren vereist. Dit is de tabel HT_Score. Let op: de tabel Export_Temp3 kan ook, die bevat dezelfde indicatoren, en daar boven op een aantal andere variabelen die uit de risicoselectierun volgen, maar niet gebruikt worden tijdens het tijdreizen.

3.2. FLOWS

Het SAS project wordt met ingang van december 2015 uitgevoerd op de AWS omgeving:

AD010\data\RisicoSelectie\Tijdreizen\HT\SAS\HT_Tijdreizen

Het project bestaat uit meerdere process flows:

Autoexec: toewijzen van libnames

A. Initialize: Initialisatie van scripts en input Trainingscases_HT

B. Steekproef cases 2013: trekken van steekproef uit bestand 2013

C. Create Testcases dd-yy: het toevoegen van details vanuit een historische risicoselectierun die uitgevoerd is vlak voor de datum waarop de case als trainingscase beoordeeld is.

Z. Append: voeg alle testcases samen in 1 tabel

In december 2015 is besloten om de trainingscases voor toeslagjaar 2013 te bevriezen. De process flows die dit regelden waren omvangrijk en er werden nauwelijks meer nieuwe trainingscases voor dit toeslagjaar toegevoegd.

De bevriezing houdt in dat het tijdreizen voor deze toeslagjaren eenmalig is gerund (zie aparte documentatie HT_Tijdreizen_2013), waarna het bestand met cases inclusief alle indicatoren opgeslagen is. Dit bestand wordt aan de gehele set met trainingscases toegevoegd in de process flow Z. Append. De implicatie is dat geen nieuwe trainingscases voor 2013 meer toegevoegd kunnen worden. Omdat het een omvangrijk en niet representatief bestand is, wordt een steekproef van 2013 aan het totaal toegevoegd (zie par. 3.3.3).

4. BESCHRIJVING HALFPRODUCT

4.1. AUTOEXEC

Deze process flow wordt gebruikt voor het aanmaken om:

- Verschillende libraries aan te maken, 1 generieke, en 1 per maand;
- De generieke brondata (BSN's van alle testcases in te lezen).

4.1.1. A. INITIALIZE

In het script A02_SetMonth de bepaling van de meest historische data per testcase plaats. Hiervoor wordt eerst per testcase bepaald wat de datum van beoordeling was, op basis van de diverse beschikbare datumkolommen.

Vervolgens worden aan de testcases alle beschikkingen uit de risicoselectieruns gelijk of voor de beoordeeldatum gekoppeld. Principe hierachter is dat de beoordeling van de BSN waarschijnlijk plaatsvindt naar aanleiding van een conceptbeschikking, en dat die conceptbeschikking terug te vinden zou moeten zijn in de run die op of voor die datum plaats heeft gevonden. Als de BSN in meerdere risicoselectieruns terugkomt, wordt alleen de laatste mutatie behouden.

Dus bijvoorbeeld:

- BSN is beoordeeld op 2 augustus 2013.
- Er zijn gescoorde mutaties beschikbaar voor 31 mei, 2 augustus en 30 september 2013.
- De testcase wordt toegewezen aan maand augustus 2013.

En een tweede voorbeeld:

- BSN is beoordeeld op 20 december 2013.
- Er zijn gescoorde mutaties beschikbaar voor 2 augustus, 30 september en 13 december 2013.
- De testcase wordt toegewezen aan maand december 2013.

4.1.2. B. STEEKPROEF CASES 2013

In deze process flow wordt het steekproefbestand voor 2013 aangemaakt. Het programma is in deze process flow opgenomen zodat het mogelijk is om een andere steekproef te trekken indien dat wenselijk is. Het resultaat van de steekproef staat in de map Results/HT_training_2013_selectie.

4.1.3. C. CREATE TESTCASES 2014

Per jaar is er momenteel 1 process flow.

Brondata specifiek voor deze flow:

HT_score_yyyymm	Output van het model zoals gedraaid tijdens een risicoselectierun (bijvoorbeeld januari 2014 over toeslagjaar 2014).
-----------------	--

Deze flows kennen maar 1 stap: selecteer in tabel met alle trainingscases de cases die beoordeeld zijn op het moment van de betreffende risicoselectierun run, en gebruik hiervoor alle indicatoren beschikbaar in de HT_score tabel voor die risicoselectierun.

4.1.4. Z. APPEND CASES

Z01: In deze flow worden alle testcases zoals samengesteld in de voorgaande flows samengevoegd tot één tabel: HT_TRAININGSCASES_FINAL. Alleen velden die nodig zijn om te modelleren in Enterprise Miner worden meegenomen, de rest wordt gedropt.

Bovendien wordt een laatste filter gezet op de cases die uiteindelijk als trainingscase naar Enterprise Miner gaan. In de volgende situatie wordt een case uitgesloten:

1. Er is niet bekend of de case frauduleus was of niet (Y ne.); OF
2. De kale huur is onbekend (Kale_huur ne.);

In het verleden werden posten ook uitgesloten als de volgende variabelen missing waren:

3. Het oppervlakte per persoon is onbekend (oppervlakte_per_persoon ne.); OF
4. De oppervlakteprijs ten opzichte van de gemiddelde huurprijs in de Postcode 5-omgeving is onbekend (perc_oppervlakteprijs_omg ne.); OF
5. De huurprijs ten opzichte van de gemiddelde huurprijs in de Postcode 5-omgeving is onbekend (perc_huurprijs_huurprijs_omg ne.).

In gevallen 2-5 gaat het om indicatoren die in het verleden voor een kleine subset niet te reconstrueren bleek. Omdat deze cases de training van het model sterk beïnvloedden (want bij waarde missing was eventrate 100%) is besloten deze cases uit te sluiten. Echter op dit moment is de verdeling goed/fout bij de cases met missings op deze variabelen 50/50. Besloten is dan ok om deze wel in de trainingsset te laten.

Het bestand met trainingscases uit 2013 (zie hierboven) wordt vervolgens afzonderlijk toegevoegd.

Vervolgens wordt een steekproef getrokken op het totale bestand omdat het bestand onevenwichtig is opgebouwd voor wat betreft herkomst. We hebben geëxperimenteerd met de SAS Enterprise Miner en op basis daarvan gekozen voor een steekproef van:

- 4.000 posten op de DTCheck van 2017 (dit betreft alleen goede posten)
- 4.000 goede posten uit het bestand van eerder gecontroleerde posten
- 4.000 foute posten uit het bestand van eerder gecontroleerde posten

Het programma is zodanig opgezet dat de steekproef eenvoudig aangepast kan worden.

Z02: check op alle numerieke velden of missende waarden voorkomen. Deze output wordt verder niet gebruikt, is bedoeld voor controledoeleinden. Wel altijd draaien!

5. UPDATE HT_TIJDREIZEN

Het project is per risicoselectierun opgezet. Dat betekent dat:

1. Er diverse scripts zijn waarin hard-coded de betreffende maanden aangeroepen worden;
2. Er 1 process flow per risicoselectierun is, waarin de relevante testcases geselecteerd worden en de indicatoren vanuit die risicoselectierun toegevoegd worden aan de cases die vlak na die risicoselectierun zijn beoordeeld.

Bij het toevoegen van een nieuwe risicoselectierun moet dus:

1. Nieuwe (bron)data die relevant is voor deze risicoselectierun ontsloten worden. De HT_Score tabel uit de betreffende risicoselectierun wordt tijdens het maken van de back-up van de desbetreffende risicoselectierun in de HT_Tijdreizen folder op de AWS omgeving gezet.
2. Een tab ingevoegd worden waarin testcases voor de betreffende risicoselectierun geselecteerd en indien nodig aangevuld worden.
3. Diverse algemene scripts aangepast worden zodat ook de nieuwe risicoselectierun meeloopt. Het gaat om:
Autoexec: A01: nieuwe library
A02: nieuwe risicoselectierun hardcoded toevoegen aan stap 3 (2x).
Z01: finaal testcases bestand voor de nieuwe risicoselectierun toevoegen aan set statement in stap 1.
4. Bepaald worden welke indicatoren nog niet beschikbaar zijn, en (indien van toepassing) hoe deze op basis van de tijdens de run voor die risicoselectierun gebruikte gegevens toegevoegd kunnen worden. Dit komt momenteel niet voor.

6. STAPPENPLAN

In dit deel staan kort de verschillende stappen die ondernomen moeten worden bij het aanvullen van het tijdreizen. In het stappenplan moet de input data worden aangevuld, en het programma moet worden aangepast.

6.1. PROGRAMMA TIJDREIZEN

[AWS/files/AD10/data/RisicoSelectie/tijdreizen/ht_training/SAS/HT_tijdreizen.epg]

- 1.) In process flow 'A0.Initialize' in programma 'A01_SetParameters' onderaan een libname toevoegen met de nieuwe run. Voorbeeld: LIBNAME LIB0316 "&PATH./Input/2016";
- 2.) Zelfde process flow in programma 'A02_Set month' de libnames toevoegen en onderaan de nieuwe risicoselectie run met betreffende datum en het jaar waarop de run betrekking heeft toevoegen. Voorbeeld: LIB0316.HT_score_201603 (IN = IN032016 KEEP = BSN)

```
En: IF IN032016 THEN DO;
      runnr = 201603;
      Toeslagjaar = 2016;
      Rundatum = INPUT('29/01/2016', ddmmyy10.);
END;
```

- 3.) Vervolgens wordt er in de process flow van het huidige jaar een nieuw stukje aan het programma geplakt waarin nieuwe runs worden toegevoegd. Voorbeeld van een programma:

```
PROC SQL;
CREATE TABLE LIB0118.HT_TRAINING_201801_final AS
SELECT A.*,
      B.Type,
      B.Datum_beeoordeeld,
      B.Rundatum,
      B.Herkomst,
      B.Omschrijving AS werkopdracht
FROM LIB0118.HT_score_201801 AS A
INNER JOIN HT_Train.Trainingscases_HT_final AS B
ON A.BSN = B.BSN AND a.toeslagJaar = B.toeslagJaar
WHERE B.runnr = 201801 ;
QUIT;
```

- 4.) In de process flow 'Z. Append' wordt tot slot in het programma 'Z01_append' de libname ook toegevoegd, voorbeeld: LIB0316.HT_TRAINING_201603_FINAL

6.2. INPUT DATA TOEVOEGEN

De data van elke risicoselectie run worden automatisch tijdens het maken van de back-up na elke run toegevoegd in:

AWS/ files/AD10/data/RisicoSelectie/tijdreizen/ht_training/Input/"betreffende
jaar"/ht_score_'jaar'runnr'.

6.3. BACK-UP MAKEN

Als het proces is afgerond het programma en de datasets opslaan op de q-schijf:

Q:\VEPROW63\TSL_DM_Handhavingregie\Profiling 2013\Trainingscases\Backup

Het project in de map sas_epg en de bestanden in de map: bestanden.

Telkens opslaan met bestandsnaam_jjjjmmdd.

7. NIEUWE OPZET

Met ingang van april 2019 zijn ook de jaren 2014 en 2015 bevroren. Deze bestanden zijn opgeslagen onder de namen: ht_trainingset_2014 en ht_trainingset_2015 in de map Results en deze worden vanaf april 2019 rechtstreeks in de append cases toegevoegd.

8. RESULTATEN PER TIJDREIZEN

Vanaf maart 2019 worden per keer ook de resultaten van het tijdreizen opgenomen

8.1. MAART 2019

Herkomst	Toeslagjaar	totaal			steekproef		
		GOED	FOUT	TOTAAL	GOED	FOUT	TOTAAL
DT Check	2017	12.252	nvt	12.252	4.000	nvt	4.000
Excel	2013	1.500	2.088	3.588	1.500	2.088	3.588
	2014	638	2.684	3.322	638	2.684	3.322
	2015	88	209	297	88	209	297
	totaal	2.226	4.981	7.207	2.226	4.981	7.207
Fraudeteams	2014	0	60	60	0	60	60
	2015	0	52	52	0	52	52
	2016	0	48	48	0	48	48
	2017	0	92	92	0	92	92
	2018	0	37	37	0	37	37
	2019	0					
	totaal	0	289	289	0	289	289
GreenLane	2016	7.555	0	7.555	2.000	0	2.000
	2017	659	127	786	659	127	786
	totaal	8.214	127	8.341	2.659	127	2.786
Zaak	2014	787	2.225	3.012	148	737	885
	2015	4.206	3.151	7.357	784	1.045	1.829
	2016	6.126	2.469	8.595	1.126	844	1.970
	2017	7.420	2.947	10.367	1.432	1.029	2.461
	2018	2.357	997	3.354	491	327	818
	2019	70	49	119	19	18	37
	totaal	20.966	11.838	32.804	4.000	4.000	8.000
Totaal		43.658	17.235	60.893	12.885	9.397	22.282

HT CONCEPTBESCHIKKINGEN

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	11-07-2013		Initieel document
1.0	30-11-2015	10.2.0	Informatie aanvullen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Brondata.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	5
3. beslissingen.....	5

2. BESCHRIJVING¹

Deze brondata bevat een lijst met conceptbeschikkingen HT en wordt gebruikt in het halfproduct DTM_HT.

Maandelijks vraagt het risicoselectieteam de brondata **per mail op bij Toeslagen IM Portfoliomanagement_Postbus**. Vóór de risicoselectierun levert **BCA\BICC** de brondata op als SAS-bestand en wordt de data opgeslagen **op het netwerk**:

Q:\VEPROW52\TSL_UT_Datawisseling\Profiling van CIV\jjjmmdd

Vervolgens wordt het bestand door het risicoselectieteam gekopieerd naar de volgende locatie **op de netwerkschijf**:

*Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\HT
CONCEPTBESCHIKKINGEN\Source*

Het bronbestand op deze locatie wordt ingelezen in SAS.

2.1. BRONDATA

De brondata van HT CONCEPTBESCHIKKINGEN (inclusief aantal records):

Bronbestand	Omschrijving	Aantal records
Conceptbeschikkingen ht.sas7bdat	Lijst met conceptbeschikkingen HT	122.489 (11-07-2013)

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Name	Type	Length	Format
Aanvrager_bsn	Numeric	8	11.
Toeslagjaar	Numeric	8	20.
ContentVolgnummer	Numeric	8	20.
jaarbedragToeslag	Numeric	8	14.2
TransactieTijdstip	Date	8	DATETIME22.3
ToetsingsinkomenAanvrager	Numeric	8	20.
DatumIngang	Date	8	DATETIME22.3
DatumEinde	Date	8	DATETIME22.3
Draagkracht	Numeric	8	20.

¹ Na overgang tot AWS zal het groen gemarkeerde (aanvraag / oplevering / locatie en naam brondata) moeten worden aangepast. Daarnaast zal te zijner tijd de query, gebruikt om de data op te halen uit de database, toegevoegd worden.

Maandtoeslag	Numeric	8	14.2
Bsn_partner	Numeric	8	11.
Inkomen_kind	Numeric	8	20.
Bsn_kind	Numeric	8	11.
Inkomen_kind	Numeric	8	20.
Bsn_medebewoner	Numeric	8	11.
Inkomen_medebewoner	Numeric	8	20.
Woonobjectidentificatie	Numeric	8	20.
KaleHuur	Numeric	8	14.2
RekenHuur	Numeric	8	14.2
DienstEnRecruitmenkosten	Numeric	8	14.2
Energiekosten	Numeric	8	14.2
Huismeesterkosten	Numeric	8	14.2
Schoonmaakkosten	Numeric	8	14.2
Standplaatshuur	Numeric	8	14.2
Postcode	Character	6	\$6.
Huisnummer	Character	6	\$6.
HuisnummerToevoeging	Character	12	\$12.

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

2. BESLISSINGEN

10.2.g

10.2.g

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	30-11-2015	10.2.e	Initieel document
1.1	25-04-2016		AWS-aanpassingen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Brondata.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	4
3. Beslissingen.....	4

2. BESCHRIJVING

Deze brondata bevat een lijst met aanvragen HT van buitenlandse studenten. De woningbouwvereniging 10.2.g levert volgens afspraak met het risicoselectieteam de data maandelijks per mail op als xlsx-bestand. Het bestand 'Lijst aanvragen 10.2.g' wordt ontvangen in de risicoselectie postbus en wordt door het risicoselectieteam gekopieerd op de AWS:

/prj/bi/data/AD010/data/RisicoSelectie/Brondata/10.2.g/Source

Vervolgens wordt het ontvangen bestand met het SAS-script 10.2.g ingelezen en omgezet in een SAS-bestand.

2.1. BRONDATA

De brondata van 10.2.g:

Bronbestand	Omschrijving
10.2.g sas7bdat	Lijst met aanvragen HT buitenlandse studenten

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Name	Type	Length	Format
BSN	Numeric	8	BEST12.
Name	Character	28	\$CHAR28.
Date of Birth	Character	11	\$CHAR11.
Address	Character	26	\$CHAR26.
Datum doorgegeven aan Belastingdienst	Date	8	DATE9.

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

2. BESLISSINGEN

10.2.g

Geblokkeerde LRK's

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	30-04-2014	10.2.e	Initieel document

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Bron data.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	4
3. Beslissingen.....	4

2. BESCHRIJVING

Deze bron data wordt gebruikt in het halfproduct Geblokkeerde LRKs.

De brondata zijn afkomstig van de Intelligence Analisten en worden op onregelmatige basis aangevuld door

10.2.e

De brondata wordt opgeleverd als xlsx-bestand en wordt als CSV opgeslagen op het netwerk:
Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\Geblokkeerde LRKs\Source

Vervolgens word de bron data ingelezen in SAS.

Het bestand bevat een aantal LRK-nummers die betrokken zijn bij een CAF-zaak. Het is hierdoor niet meer wenselijk dat er kinderopvangtoeslag wordt aangevraagd op deze LRK's. Het risicoselectiemodel voegt de aanvragers die een van deze LRK's gebruiken toe aan de lijst harde stoppers.

2.1. BRON DATA

De bron data van IP adressen:

Bronbestand	Omschrijving	Aantal records
Geblokkeerde LRKs.csv	Lijst met LRK-nummers in CAF-zaken	

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Geblokkeerde_LRK

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

FRS Meldingen

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	07-10-2013		Initieel document
1.0	30-11-2015	10.2.e	Informatie aanvullen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Brondata.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	4
3. Beslissingen.....	5

2. BESCHRIJVING¹

Deze brondata bevat lijsten met FRS (Feiten Registratie Systeem) meldingen gegenereerd door TVS die worden gebruikt in halfproducten Melding8 en MeldingFRS.

Maandelijks vraagt het risicoselectieteam de brondata **per mail** op **bij BCA / UPG**. Deze aanvraag verloopt **via** **10.2.e** Vóór het draaien van de risicoselectierun wordt de brondata opgeleverd in de vorm van csv-bestanden en op **de netwerkschijf opgeslagen**:

Q:\VEPROW52\TSL_UT_Datawisseling\Queries Formeel Beschikken

Vervolgens worden de bestanden door het risicoselectieteam gekopieerd naar **de volgende locatie**:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\MELDINGENFRS\Source

Vanaf deze locatie worden de bestanden ingelezen in SAS.

2.1. BRONDATA

De brondata van MeldingenFRS:

Bronbestand	Omschrijving
MLD006-2.csv	Meldingen van opgegeven verhuizingen
MLD008-2.csv	Meldingen van nieuwe toeslagaanvragen
MLD015-2.csv	Meldingen van veranderde betaalwijzen
MLD018-2.csv	Meldingen van gewijzigd aantal uren opvang
MLD019-2.csv	Meldingen van huurprijswijzigingen

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Name	Getalnotatie (csv)
BSN	Standaard
Aantal meldingen	Standaard
Datum ingang (<i>alleen MLD008.csv</i>)	Datum
Datum aanvraag (<i>alleen MLD008.csv</i>)	Datum

¹ Na overgang tot AWS zal het groen gemarkeerde (aanvraag / oplevering / locatie en naam brondata) moeten worden aangepast. Daarnaast zal te zijner tijd de query, gebruikt om de data op te halen uit de database, toegevoegd worden.

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

2. BESLISSINGEN

Excepties

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	30-11-2015		Initieel document
1.1	04-08-2016	10.2.6	Bronbestand op de AWS

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Bron data.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	5
3. Beslissingen.....	5

2. BESCHRIJVING

Deze brondata bevat een lijst met BSNs die niet uitgesloten mogen of juist moeten worden van beschikken. De brondata wordt naar afspraak met het risicoselectieteam in principe maandelijks opgeleverd door **team handhavingsregie**. Een xlsx-bestand wordt **per mail** aangeleverd indien er nieuwe gevallen aan de lijst zijn toegevoegd. Deze oplevering gebeurt de dag vóór de risicoselectierun of op zijn laatst om 9.00 uur 's ochtends op de dag van de run zelf.

Het ontvangen bronbestand wordt door het risicoselectieteam opgeslagen op de AWS:
 /prj/bi/data/AD010/data/RisicoSelectie/Brondata/EXCEPTIES

Vervolgens wordt de data ingelezen in SAS in HT- en KOT-model.

2.1. BRONDATA

De brondata van Excepties:

Bronbestand	Omschrijving
BSNS niet uitsluiten wel uitsluiten.xlsx	Lijst met BSNs die niet uitgesloten mogen of juist moeten worden van beschikken

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Naam	Getalnotatie (csv)
Bsn	Standaard
Toeslagjaar	Getal
Regeling	Standaard
Reden	Standaard

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

IGS

In Gebreke Stelling

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	18-12-2015		Initieel document
1.1	25-04-2016	10.2.e	AWS-aanpassingen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Brondata.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	5
3. Beslissingen.....	5

2. BESCHRIJVING

Deze bron data bevat een lijst met werkitens die uitval of uitworp veroorzaken of in een bundel aangeboden zijn en handmatig opgepakt moeten worden. Dagelijks wordt een bestand zaak met nieuwe werkitens gecreëerd door 10.2.e werkzaam bij IM Toeslagen.

Dit bestand wordt opgeslagen op de netwerkschijf:

Q:\VEPROW52\TSL_UT_Datawisseling

10.2.e

Eens per week worden alle nieuw gecreëerde bestanden door 10.2.e van team handhavingsregie toegevoegd aan één groot SAS-bestand '**zaak**'. Dit bestand bevat een historisch overzicht van de werkitens uit eerdere perioden. Dit bestand wordt door 10.2.e op de AWS geplaatst:

/prj/bi/data/AD010/data/RisicoSelectie/Brondata/IGS/Source

Het bronbestand op deze locatie wordt gebruikt voor het creëren van de halfproducten IGS en Eerdere Acties.

2.1. BRONDATA

De brondata van IGS:

Bronbestand	Omschrijving
zaak.sas7bdat	Lijst met werkitens (uitval, uitworp of aangeboden in werkbundel)

2.1. VELDEN

Onderstaande velden zijn opgenomen:

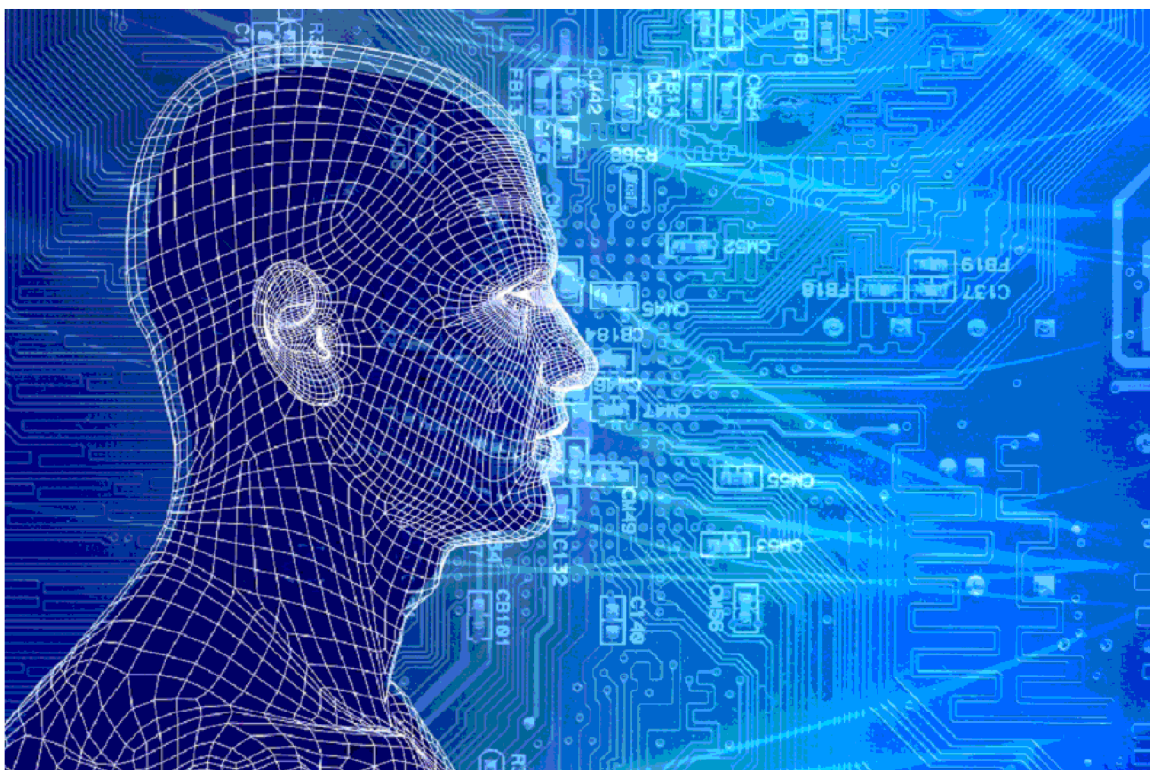
Name	Type	Length	Format
ProcesCode	Character	50	\$50.
ProcessInstanceID	Numeric	8	20.
DatumEinde	Numeric	8	E8601DT23.3
DatumEindeTermijn	Numeric	8	E8601DT23.3
DatumStart	Numeric	8	E8601DT23.3
BSN	Numeric	8	Z9.
EventID	Character	50	\$50.
Status	Character	20	\$20.
BurgerIsVIP	Character	11	\$1.
BurgerIsAmbtenaar	Character	17	\$1.
BurgerIsGeprivilegeerd	Character	23	\$1.
Type	Character	128	\$128.

TSLHeventID	Character	200	\$200.
TijdstipTransactie	Character	30	\$30.
AfhandelingsresultaatCode	Character	4	\$CHAR4.
Omschrijving afhrescode	Character	35	
BedragVoor	Numeric	8	
BedragNa	Numeric	8	
DatumZichtbaar	Numeric	8	E8601DT26.3
PreviousProcesStapCode	Character	50	\$50.
PreviousProcesStapEindDatum	Numeric	8	E8601DT23.3
Status_bericht	Character	50	\$50.
Omschrijving	Character	255	\$255.
TSLcodes_tekstueel	Character	6	
Eerste_bhl	Character	255	\$50.
Tweede_bhl	Character	255	\$50.
Startdatum_wachtstap	Numeric	8	E8601DT26.3
DestinationName	Character	128	\$128.
DestinationType	Character	15	\$5.
ProcesStapCode	Character	50	\$50.
ProcessStatus	Character	11	\$11.
ActivityStatus	Character	11	\$11.
Postcode	Character	8	\$4.
ActivityStartDate	Numeric	8	E8601DT23.3
ActivityEndDate	Numeric	8	E8601DT23.3
Behandelaar	Character	50	\$50.
ToeslagjaarVan	Character	6	\$6.
ToeslagjaarTm	Character	6	\$6.
UDD	Date	8	YYMMDD10.
DTVerdaagdTot	Date	8	YYMMDD10.
RegelingID	Numeric	8	
StartDatumBehandeling	Numeric	8	
UitgeslotenBeschikkingstype	Character	6	\$6.
Omschrijving_beschikkingstype	Character	15	E8601DT23.3
UDA	Date	8	YYMMDD10.

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN



SVB

Sociale Verzekeringsbank

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.00	08-10-2013	10.2.e	Initieel document

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving model.....	4
2.1.	Bron Data.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	4

2. BESCHRIJVING MODEL

In het script word de csv file met ouder kind relaties ingelezen vervolgens wordt het zo bewerkt dat de sas file de juiste output bevat.

2.1. BRON DATA

Als Bron data wordt het bestand Meld057.csv.

Zie de map Brondata\SVB voor meer informatie.

2.2. VELDEN IN HET MODEL

Het eind resultaat bevat de volgende velden:

svb_ouderkind.sas7bdat							
	Variable	Type	Len	DLen	Format	InFormat	Label
1	KIND	Num	8	9	Z9.	.	KIND
2	OUDER1	Num	8	12	12.	.	OUDER1
3	OUDER2	Num	8	12	12.	.	OUDER2

2.3. FLOWS

In het script word de csv file met ouder kind relaties ingelezen vervolgens wordt het zo bewerkt dat de sas file de juiste output bevat.

In principe is in de SVB data elk kind gekoppeld aan 1 ouder. Het komt echter voor dat een kind aan 2 of meerdere ouders is gekoppeld. Zie onderstaande tabel

COUNT_OF_KIND	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	3909590	98.36	3909591	98.36
2	64760	1.63	3974351	99.99
3	474	0.01	3974825	100.00
4	6	0.00	3974831	100.00

Bij navraag blijkt dat het kan voorkomen dat een kind begint met 1 ouder en in de loop van het jaar aan een tweede, derde of vierde ouder wordt gekoppeld. We zien in de data niet welke de actuele ouder is bij meerdere ouders.

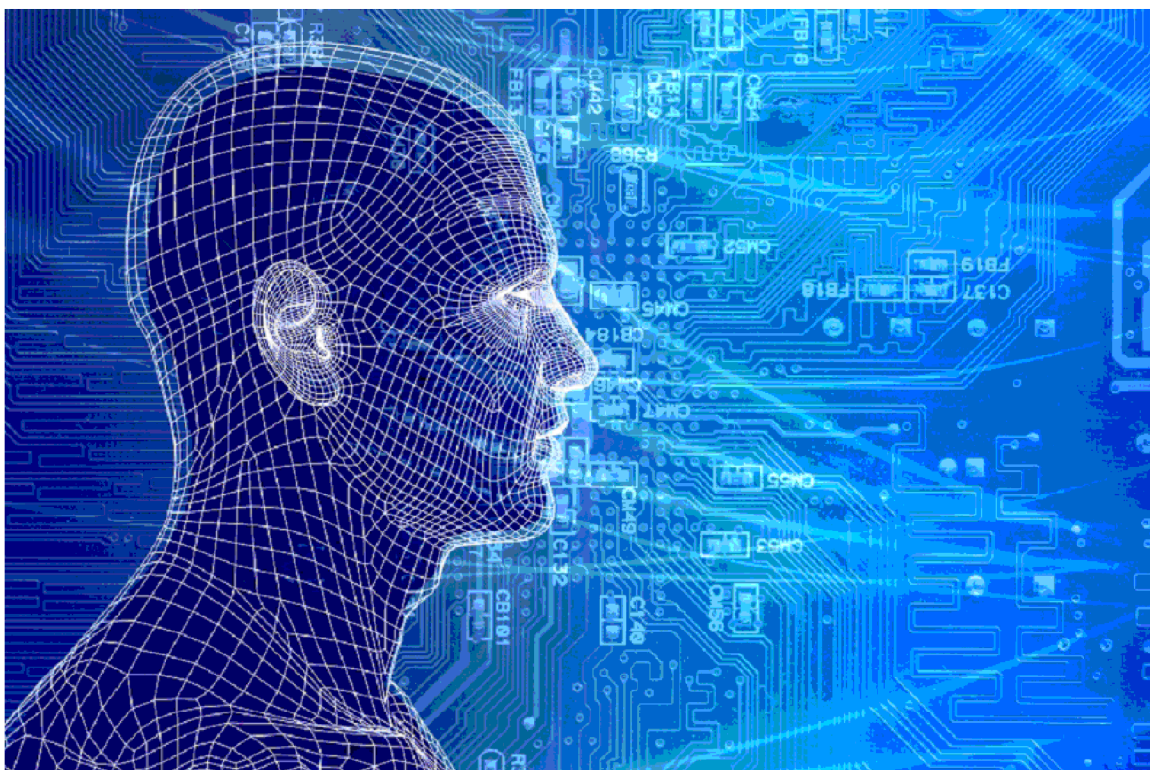
Een ander raar punt is dat er 99 gevallen zijn waar de ouder BSN gelijk is aan de kinder BSN. Het vermoeden van 10.2.4 is dat dit wezen zijn, waar van het kind de beide ouders overleden zijn. Dan is wettelijk het kind ook meerderjarig. Aangezien dit een hele kleine fractie is besloten met 10.2.4 om dit te negeren.

We aggregeren de data set tot op uniek kind niveau, met maximaal twee ouders. Het half fabricaat **svb_ouderkind.sas7bat** heeft dan de volgende kolommen

- KIND (BSN van kind)
- OUDER1 (BSN van ouder 1)
- OUDER2 (BSN van ouder 2)

KIND	OUDER1	OUDER2
1	78	78
2	585	47

Voor het overgrote deel (98.36%) zal ouder 1 en ouder 2 dus gelijk zijn. Dit halffabricaat svb_ouderkind wordt verder gebruikt in de ouderkind relatie. Om te zien of een GBA ouder relatie ook bij SVB bekend is



PIT

Blacklist/fraude

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	15-05-2013		Initieel document
0.02	08-10-2013	10.2.e	Kleine toevoegingen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving model.....	4
2.1.	Bron data.....	4
2.2.	Velden in het model.....	4
2.3.	Indicatoren.....	4

2. BESCHRIJVING MODEL

Een blacklist met BSN vanuit de FIOD en andere diensten. Voorkomen op de PIT lijst betekent ernstig vermoeden van fraude. (Hoog risico, maar niet bewezen fraude)

2.1. BRON DATA

DagboekPIT_Export.mdb van 10.2.e / 10.2.e ontvangen.

Zie de map Brondata\PIT voor meer informatie.

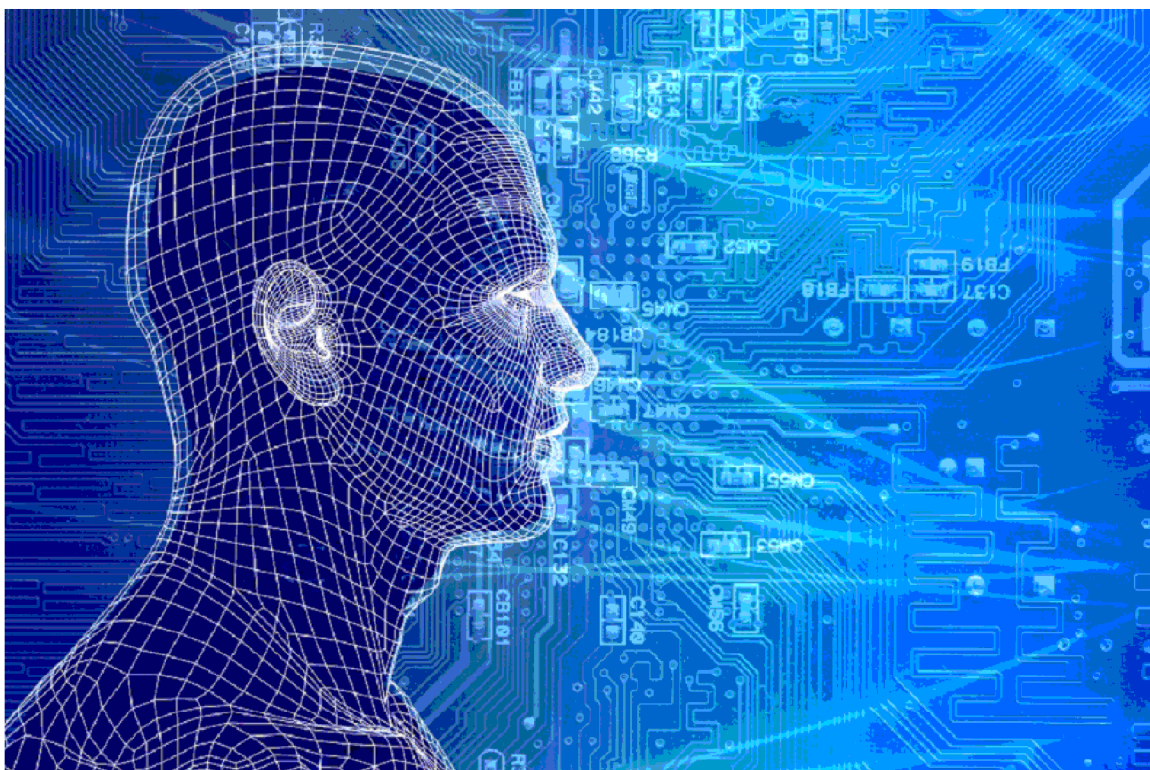
2.2. VELDEN IN HET MODEL

Het eind resultaat bevat alleen het veld

BSN

2.3. INDICATOREN

Veld	Indicator	Ontbrekende waarde
BSN	Koppelveld	Geen indicatie in deze lijst is mogelijk een "false negative"



Ouder Kind

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.00	08-10-2013	10.2.e	Initieel document
1.0	21-12-2015		Informatie aanvullen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving proces.....	4
2.1.	Brondata.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	4

2. BESCHRIJVING PROCES¹

Het SAS-script **Kind_Ouder.egg** wordt één dag voor de risicoselectierun gedraaid door het risicoselectieteam. Dit resulteert in het bestand **ouderkind_svb.sas7bdat** dat wordt gebruikt tijdens de risicoselectierun.

Zowel het SAS-script als het halfproduct staan opgeslagen **op de netwerkschijf**:

Script:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\OUDER_KIND\SAS Enterpriseguide

Halfproducten:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\OUDER_KIND\Results

2.1. BRONDATA

De input gebruikt in het SAS-script zijn de bronbestanden **Klantbeeld ouderkind.csv** en **FRS Mld057.csv**. Deze bestanden bevatten een lijst met ouder-kind relaties (Klantbeeld ouder kind) en een lijst waarin voor elk kind de ouder wordt aangegeven die de kinderbijslag ontvangt (FRS Mld057).

Voor de bronbestanden en de documentatie zie de **mappen**:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\OUDER_KIND

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\SVB

2.2. VELDEN IN HET MODEL

Het eind resultaat bevat de volgende velden:

Name	Type	Length
Kind	Numeric	8
OUDER1	Numeric	8
OUDER2	Numeric	8
SVBOUDER1	Numeric	8
SVBOUDER2	Numeric	8

¹ Na overgang tot AWS zal het groen gemarkeerde (locatie en naam brondata/halfproduct/SAS-script) moeten worden aangepast.

2.3. FLOWS

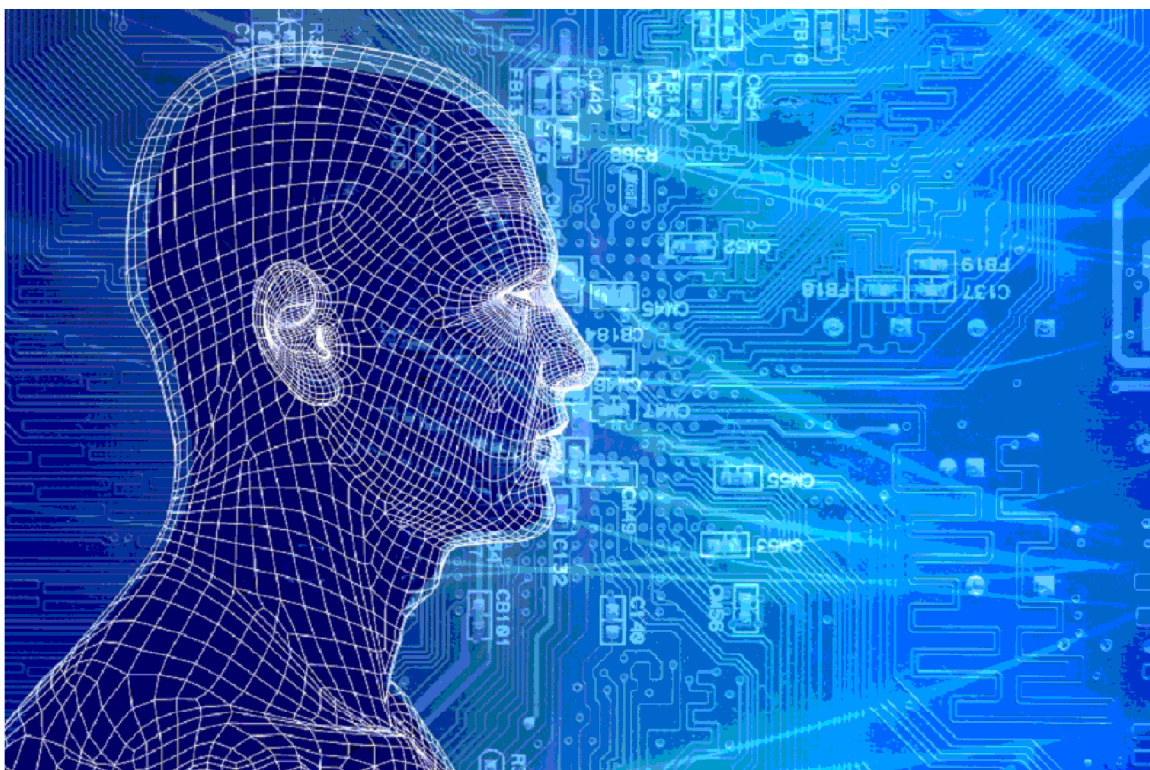
In het script worden de bronbestanden ingelezen en gekoppeld op kind BSN:

Stap 1: De bronbestanden worden ingelezen en voor elk veld wordt het format bepaald.

Stap 2: De ingelezen tabel met de gegevens van de SVB (FRS Mld057) wordt aangepast: In principe is elk kind gekoppeld aan één ouder. Echter, het komt voor dat een kind in de loop der tijd aan meerdere ouders wordt gekoppeld. Het is hier niet bekend wie van de ouders de actuele ouder is. De dataset wordt geaggregeerd op uniek kind niveau, met daarbij maximaal twee ouders. De eerste ouder is de ouder met het hoogste BSN, en de tweede ouder is de ouder met de laagste BSN.

Stap 3: De ingelezen tabel met ouder-kind relaties wordt gekoppeld aan de geaggregeerde tabel afkomstig van de SVB. Deze koppeling wordt gedaan door middel van een full outer join op de BSN van het kind.

In het resulterende halfproduct is te zien of een ouder kind relatie ook bekend is bij de SVB.



BAG

Basisregistratie adressen en gebouwen.

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.00	04-10-2013	10.2.e	Initieel document

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving model.....	4
2.1.	Bron Data.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	4

2. BESCHRIJVING MODEL

Het aangeleverde BAG bestand wordt met de gewenste velden omgezet naar sas.

2.1. BRON DATA

Uitgangspunt is de platte bag bron: bag_convert.csv

Voor meer informatie zie Brondata\BAG

2.2. VELDEN IN HET MODEL

Het eindresultaat bevat de volgende velden

bag.sas7bdat							
	Variable	Type	Len	DLen	Format	InFormat	Label
1	PHT	Char	30	30	\$30.	.	PHT
2	M2	Num	8	12	12.	.	M2
3	GEBRUIKSDOE	Char	30	30	\$30.	.	GEBRUIKSDOEL

2.3. FLOWS

Stap 1

Inlezen van data, In totaal bevat deze data set 19,085,286 records. 79.39% van deze records heeft een woonfunctie, er zijn ruim 872,612 records zonder postcode en dus ook niet te matchen.

Stap 2

We maken een KeyAdres aan door:

```
compress( uncase(postcode hoofdadres || "&" || huisnummer hoofdadres ||  
"&" || huisletter_hoofdadres || huisnummertoevoeging_hoofdadres )))
```

En we pakken allen records met een doel_vbo = woonfunctie EN waar een postcode is EN waar status_pand = 'Pand in Gebruik'

We houden na deze stap 11,629,668 records over.

Stap 3

We nemen per KeyAdres het meest recente record gebaseerd op de datum in de kolom begin_vbo omdat we indicatoren baseren op doel vbo en oppervlakte vbo.

Het bag halffabricaat heet **bag.sas7bat** en heeft totaal 7,495,281 records en bevat de volgende velden:

- KeyAdres
- M2
- Bag_Gebruiksdoel

Aangeleverde Stellaposten

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	24-12-2015	10.2.e	Informatie aanvullen
1.1	29-04-2016		AWS-aanpassingen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving proces.....	4
2.1.	Brondata.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	4

2. BESCHRIJVING PROCES¹

Het SAS-script **Aangeleverd.egp** wordt vlak voor de risicoselectierun gedraaid door het risicoselectieteam. Dit resulteert in het halfproduct **aangeleverd.sas7bdat**, welke wordt gebruikt tijdens de risicoselectie-run.

Zowel het SAS-script als het halfproduct staan opgeslagen op de AWS:

Script:

/prj/bi/data/AD010/data/RisicoSelectie/Halfproducten/BEZWAREN/Results/SAS Enterprise Guide

Halfproduct:

/prj/bi/data/AD010/data/RisicoSelectie/Halfproducten/BEZWAREN/Results

2.1. BRONDATA

De input gebruikt in het SAS-script is het bronbestand **aangeleverd.xlsx**. Dit bestand bevat een lijst met stellaposten die niet mogen worden meegenomen in de risicoselectie. Voor het bestand zie de map:

/prj/bi/data/AD010/data/RisicoSelectie/Brondata/BEZWAREN/Source

2.2. VELDEN IN HET MODEL

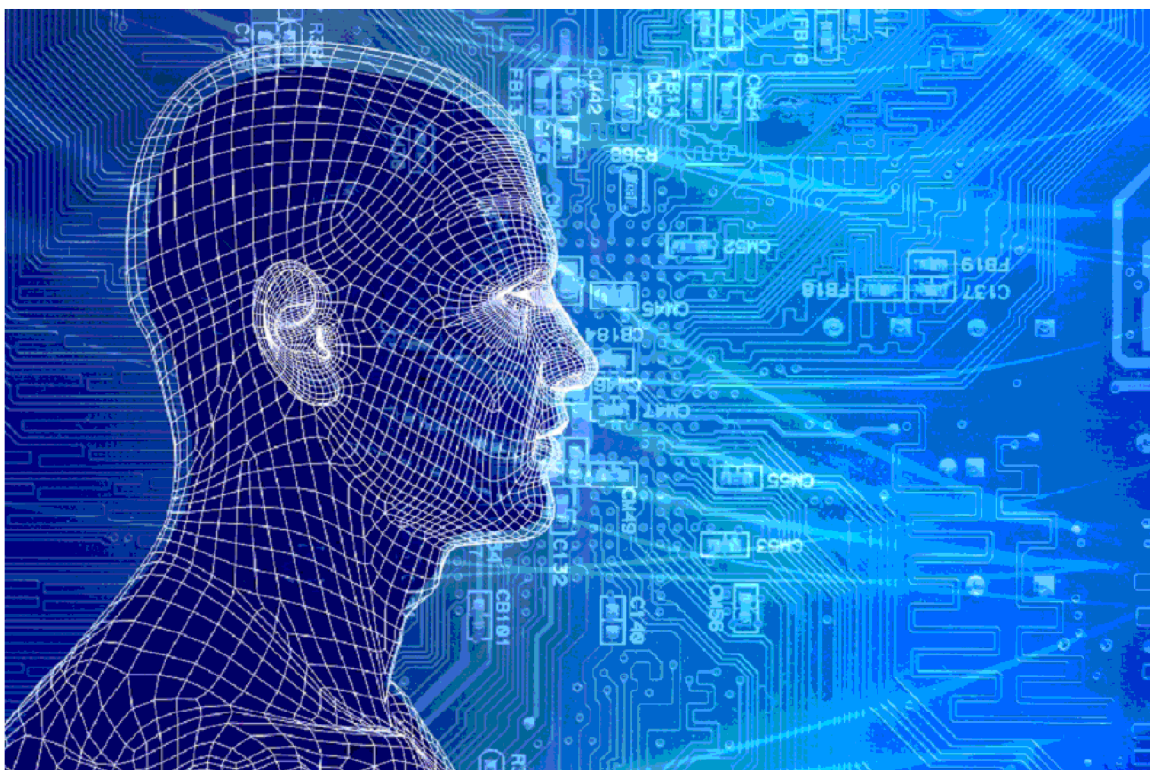
Het eind resultaat bevat de volgende velden:

Name	Type	Length	Format	Informat
BSN	Numeric	8	BEST9.	BEST9.
Toeslagjaar	Numeric	8	BEST4.	BEST4.
KOT	Character	1	\$CHAR1.	\$CHAR1.
Huur	Character	1	\$CHAR1.	\$CHAR1.
Zorg	Character	1	\$CHAR1.	\$CHAR1.
KGB	Character	1	\$CHAR1.	\$CHAR1.

2.3. FLOWS

In het script wordt het bronbestand ingeladen in SAS en opgeslagen als SAS-bestand.

¹ Na overgang tot AWS zal het groen gemarkeerde (locatie en naam brondata/halfproduct/SAS-script) moeten worden aangepast.



6PP

postcodes

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.00	04-10-2013		Initieel document
1.00	13-10-2014	10.2.e	Aanpassing n.a.v. nieuwe wijze van levering
1.01	23-12-2015		Uitbreiden informatie

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving proces.....	4
2.1.	Brondata.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	4

2. BESCHRIJVING PROCES¹

Het SAS-script **6PP.egp** wordt één dag voor elke risicoselectierun gedraaid door het risicoselectieteam. Dit resulteert in het halfproduct **postcodes.sas7bdat**, welke wordt gebruikt tijdens de risicoselectierun.

Zowel het SAS-script als het halfproduct staan opgeslagen **op de netwerkschijf**:

Script: *Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\6PP\Result*

Halfproduct: *Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\6PP\SAS Enterpriseguide*

2.1. BRONDATA

De input gebruikt in het SAS-script is het bronbestand **6PP.csv**. Dit bestand bevat een lijst met postcodes van Nederland met lengte- en breedtegraad. Voor het bestand en de documentatie zie de **map**:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\6PP

2.2. VELDEN IN HET MODEL

Het eindresultaat bevat de volgende velden.

Name	Type	Length
Postcode	Character	6
Lat	Numeric	8
Lng	Numeric	8
CijfersPostcode	Character	6

2.3. FLOWS

Stap 1: Het bronbestand wordt ingeladen.

Stap 2: Een nieuw bestand wordt gecreëerd met enkel de velden Postcode, Lat en Lng.

Stap 3: De regels met dezelfde postcode worden gegroepeerd. Hierbij worden de lat- en lng-waarden vervangen door het gemiddelde.

¹ Na overgang tot AWS zal het groen gemarkeerde (locatie en naam brondata/halfproduct/SAS-script) moeten worden aangepast.

Stap 4: De regels worden ontdubbeld op postode.

Stap 5: Een nieuw veld Cijferspostcode wordt aangemaakt met de cijfers van elke postcode.

WOZ
waardering onroerende zaken

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	12-07-2013	10.2.e	Initieel document
0.02	08-10-2013		Info bron data toegevoegd

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving.....	4
2.1.	Bron data.....	4
2.2.	Velden.....	4
2.3.	Indicatoren.....	4
3.	Beslissingen.....	5

2. BESCHRIJVING

Deze bron data wordt gebruikt voor het halfproduct WOZ.

De bron data wordt geleverd/opgevraagd bij of van B/CA – CIV.

De data wordt niet maandelijks opgevraagd.

De bron data wordt opgeleverd als txt-bestand en wordt opgeslagen op het netwerk:

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\WOZ\Source

Vervolgens word de bron data ingelezen in SAS (halfproduct WOZ).

2.1. BRON DATA

De bron data van WOZ (inclusief aantal records):

Bronbestand	Omschrijving	Aantal records
U130156A1_TOESL_WOZ.TXT	WOZ-lijst	2.287.133

2.1. VELDEN

Onderstaande velden zijn opgenomen:

BSN_BETROKKENE

BSN_AANVRAGER

POSTCODE

HUISNUMMER

HUISLETTER

HUISNUMMER_TOEVOEGING

COMBI_PC_HN_HL

INGDAT_HTSL

EINDDAT_HTSL

WZ230FINR

WZ230JAAR

WOZ_POSTCD

WOZ_HUISNR

WOZ_HUISLETTER

WOZ_TOEV

WOZ_AAND

COMBI_PC_HNR_HL

WOZ_STRATNM
WOZ_WOONPLTS
WOZ_WAARDE
INGANGSDATUM
EINDDATUM

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

Bezwaren

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	18-12-2015	10.2.e	Initieel document
1.1	29-04-2016		AWS-aanpassingen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving proces.....	4
2.1.	Brondata.....	4
2.2.	Velden in het model.....	4
2.3.	Flows.....	5
2.4.	Indicatoren.....	5

2. BESCHRIJVING PROCES¹

Het SAS-script filter op **bezwaren.sas** wordt één dag voor de risicoselectie-run gedraaid door het risicoselectieteam. Dit resulteert in het halfproduct **bezwaren_filter_1.sas7bdat**, welke wordt gebruikt tijdens de risicoselectie-run.

Zowel het SAS-script als het halfproduct staan opgeslagen op de AWS:

Script:

/prj/bi/data/AD010/data/RisicoSelectie/Halfproducten/BEZWAREN / SAS Enterprise Guide

Halfproduct:

/prj/bi/data/AD010/data/RisicoSelectie/Halfproducten/BEZWAREN /Results

2.1. BRONDATA

De input gebruikt in het SAS-script is het bronbestand **overzicht_afgedane_bezwaren.xlsx**. Dit bestand bevat alle behandelde en lopende bezwaren. Voor het bestand zie de map:

/prj/bi/data/AD010/data/RisicoSelectie/Brondata/BEZWAREN /Source

2.2. VELDEN IN HET MODEL

Het eind resultaat bevat de volgende velden:

Name	Type	Length	Format	Informat
ProcesCode	Character	7	\$7.	\$7.
ProcessInstanceID	Numeric	8	BEST12.	BEST32.
DatumEinde	Date	8	DDMMYY10.	DDMMYY10.
DatumEindeTermijn	Date	8	DDMMYY10.	DDMMYY10.
DatumStart	Date	8	DDMMYY10.	DDMMYY10.
BSN	Numeric	8	BEST12.	BEST32.
Status	Character	9	\$9.	\$9.
Type	Character	26	\$26.	\$26.
TSLHeventID	Character	1	\$1.	\$1.
TijdstipTransactie	Date	8	DATETIME.	ANYDTCM40.
datumZichtbaar	Date	8	DDMMYY10.	DDMMYY10.
Eerste_bhl	Character	7	\$7.	\$7.
Tweede_bhl	Character	7	\$7.	\$7.

¹ Na overgang tot AWS zal het groen gemarkeerde (locatie en naam brondata/halfproduct/SAS-script) moeten worden aangepast.

Startdatum_wachtstap	Date	8	DDMMYY10.	DDMMYY10.
previousProcesStapCode	Character	29	\$29.	\$29.
Status_bericht	Character	20	\$20.	\$20.
DestinationName	Character	1	\$1.	\$1.
DestinationType	Character	1	\$1.	\$1.
ProcesStapCode	Character	1	\$1.	\$1.
ProcesStatus	Character	1	\$1.	\$1.
ActivityStatus	Character	1	\$1.	\$1.
Postcode	Character	1	\$1.	\$1.
Behandelaar	Character	7	\$7.	\$7.
BurgerIsVIP	Character	1	\$1.	\$1.
BurgerIsAmbtenaar	Character	1	\$1.	\$1.
Omschrijving	Character	33	\$33.	\$33.
BurgerIsGeprivilegeerd	Character	1	\$1.	\$1.
RegelingID	Character	1	\$1.	\$1.
DatumPrimaireBeschikking	Date	8	DDMMYY10.	DDMMYY10.
PrimaireBeschikkingKenmerk	Character	17	\$17.	\$17.
Team	Character	7	\$7.	\$7.
Regeling	Numeric	8	BEST12.	BEST32.
Naam	Character	19	\$19.	\$19.
Toeslagjaar	Numeric	8	BEST12.	BEST31.

2.2. FLOWS

In het script word er een selectie gemaakt op de bezwaren die aan de volgende eisen voldoen:

- Start datum na 30 april 2013
- Einddatum na de laatste run van formeel beschikking.
- Het toeslagjaar van de primaire beschikking is niet 2012 of 2013.
- De regeling waar de primaire beschikking betrekking op heeft is niet Kinderopvangtoeslag of Kindergebondenbudget.
- De datum van de primaire beschikking is leeg of eerder dan 30 april 2013.
- De omschrijving van het bezwaar is niet gelijk aan '9001 – reactie vraagbrief bezwaar'.

De einddatum moet per run worden geüpdatet met de variabele DatumEinde. DatumStart hoeft niet aangepast te worden.

Wanbetalers

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	18-12-2015		Initieel document
1.1	04-08-2016	10.2.e	AWS-aanpassingen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Brondata.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	4
3. Beslissingen.....	4

2. BESCHRIJVING

Deze brondata bevat een lijst met BSN's waarvan vanuit CVZ bekend is dat geen zorgpremie wordt afgedragen.

Het risicoselectieteam vraagt de data maandelijks per mail op bij Toeslagen IM Portfoliomanagement_Postbus. Deze aanvraag stuurt IM middels een queryverzoek door naar BICC.

Vóór de risicoselectierun levert BCA\BICC de brondata op als SAS-bestand en wordt de data opgeslagen op het netwerk:

Q:\VEPROW52\TSL_UT_Datawisseling\Profiling van CIV\jjjmmdd

en op de AWS:

/prj/bi/data/AD010/data/RisicoSelectie/CIV

Vanaf deze locatie wordt de data ingelezen in SAS om te gebruiken tijdens de risicoselectierun.

2.1. BRONDATA

De brondata van Wanbetalers:

Bronbestand	Omschrijving
civ_prof_wanbetalers.sas7bdat	Lijst met BSNs waarvan bekend dat geen zorgpremie wordt afgedragen.

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Name	Type	Length	Format
Bsn	Numeric	8	11.

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

VIP

Versie nummer	Datum	Initialen	Belangrijkste wijziging
1.0	18-12-2015		Initieel document
1.1	04-08-2016	10.2.e	AWS-aanpassingen

1. INHOUDSOPGAVE

1.	Inhoudsopgave.....	3
2.	Beschrijving.....	4
2.1.	Brondata.....	4
2.2.	Velden.....	4
2.3.	Indicatoren.....	4
3.	Beslissingen.....	4

2. BESCHRIJVING

Deze brondata bevat een lijst van binnen de Belastingdienst beschermde BSNs die niet voor elke behandelaar inzichtelijk zijn. De brondata wordt gebruikt in halfproducten HT_Model en KOT_Model.

Maandelijks stuurt het risicoselectieteam een mail naar Toeslagen IM Portfoliomanagement_Postbus. Deze aanvraag stuurt IM middels een queryverzoek door naar BICC.

Vóór de risicoselectierun wordt brondata als SAS-bestand opgeleverd en wordt de data opgeslagen op het netwerk:

Q:\VEPROW52\TSL_UT_Datawisseling\Profiling van CIV\jjjmmdd

en op de AWS:

/prj/bi/data/AD010/data/RisicoSelectie/CIV

Vanaf deze locatie wordt de brondata ingelezen in SAS.

2.1. BRONDATA

De brondata van VIP:

Bronbestand	Omschrijving
civ_prof_privacy_doelgroep. sas7bdat	Lijst met binnen Belastingdienst beschermde BSNs

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Naam	Type	Length	Format
Bsn	Numeric	8	11.

2.1. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

3. BESLISSINGEN

VHA

Verhuurders Administratie

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	12-07-2013		Initieel document
1.0	17-12-2015	10.2.e	Informatie aanvullen

1. INHOUDSOPGAVE

1. Inhoudsopgave.....	3
2. Beschrijving.....	4
2.1. Bron data.....	4
2.2. Velden.....	4
2.3. Indicatoren.....	5
3. Beslissingen.....	5

2. BESCHRIJVING¹

Deze brondata bevat een lijst samengesteld door de Belastingdienst met huurbedragen per fysiek adres volgens de verhuurders. De brondata wordt gebruikt in het halfproduct VHA.

Maandelijks vraagt het risicoselectieteam de data **per mail** aan **bij BCA / UPG**. Deze aanvraag verloopt **via** **10.2.e** Vóór het draaien van de risicoselectierun wordt de brondata opgeleverd als csv-bestand en op **de netwerkschijf opgeslagen**:

Q:\VEPROW52\TSL_UT_Datawisseling\Queries Formeel Beschikken

Over het algemeen heeft dit bestand de naam 'VHA', 'VHA huur' of 'VHA huurprijzen'.

Vervolgens wordt het bronbestand door het risicoselectieteam gekopieerd **naar de volgende locatie**, waarbij de bestandsnaam is aangepast tot 'VHA':

Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\VHA\Source

Vanaf deze locatie wordt de data ingelezen in SAS.

2.1. BRONDATA

De brondata van VHA:

Bronbestand	Omschrijving
VHA.csv	Lijst van huurbedragen per fysiek adres volgens verhuurders.

2.1. VELDEN

Onderstaande velden zijn opgenomen:

Naam	Getalnotatie (csv)
Postcode	Standaard
Adrescode	Standaard
Datum1	Standaard
Huisnummer	Standaard
Huisnummer_toevoeging	Standaard
Huisletter	Standaard
Datum2	Standaard

¹ Na overgang tot AWS zal het groen gemarkeerde (aanvraag / oplevering / locatie en naam brondata) moeten worden aangepast. Daarnaast zal te zijner tijd de query, gebruikt om de data op te halen uit de database, toegevoegd worden.

Kale_huur	Standaard
Energiekosten	Standaard
Huismeester_kosten	Standaard
Schoonmaakkosten	Standaard
Standplaats_huur	Standaard
Diensten_recreatieruimten_kosten	Standaard
Bedrijfsruimte_huur	Standaard
Garage_huur	Standaard

2.2. INDICATOREN

Veld	Benodigd voor Indicator	Omschrijving	Ontbrekende waarde indicator mits gekoppeld aan beschikking
N.v.t.			

2. BESLISSINGEN