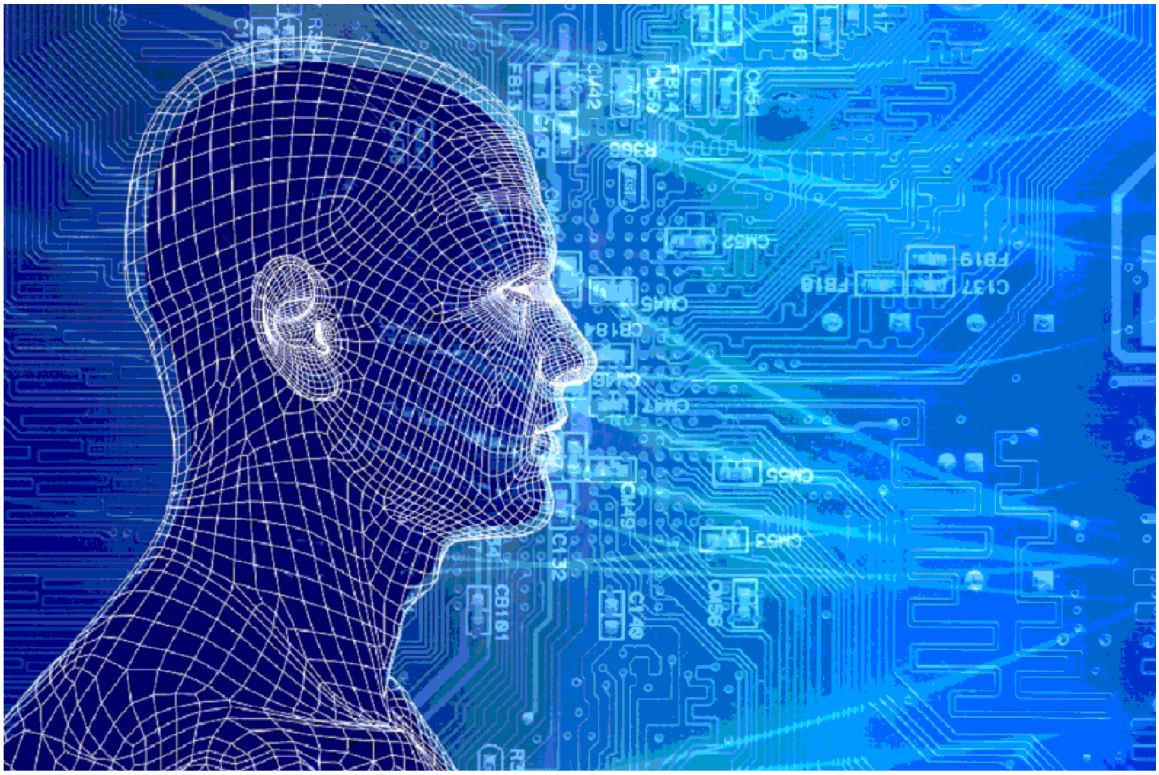


DTM_KOT

Versies van het model

Versie nummer	Datum	Initialen	Belangrijkste wijziging
0.01	23-05-2013		Initieel document
0.02	04-06-2013	10.2.e	Flows etc
1.0	24-12-2015		Aanvullen informatie

1. INHOUDSOPGAVE

1. Inhoudsopgave 3
2. Beschrijving Beschrijving proces 4
2.1. Input 4
2.1.1. Brondata 4
2.1.2. Halfproducten 5
2.2. Flows 6
2.2.1. Create DTM KOT 6
2.3. Beslissingen 10

1075483 00067

2. BESCHRIJVING BESCHRIJVING PROCES[1]

Het SAS-script DTM_KOT.egp wordt op de dag van de risicoselectie gedraaid door het risicoselectieteam. Dit resulteert in het bestand dtm_kot.sas7bdat, dat wordt gebruikt tijdens de risicoselectierun.

Zowel het SAS-script als het halfproduct staat opgeslagen op de C-schijf van de stand alone laptop:

Script:

C:\Risico Selectie\Model \DTM_KOT\SAS Enterpriseguide

Halfproduct:

C:\Risico Selectie\Model \DTM_KOT\Results

2.1. INPUT

De input in het SAS-script betreft zowel brondata als halfproducten:

2.1.1. BRONDATA

De bronbestanden gebruikt als input voor het model zijn:

1. **Beschikkingen_kot.sas7bdat**
   Dit bestand bevat een lijst met beschikkingen kinderopvangtoeslag. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\BESCHIKKING KOT

2. **Kot_deeljaar_bedragen.sas7bdat**
   Dit bestand bevat een lijst met deeljaar beschikkingen KOT en bijbehorende bedragen. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\DEELJAAR BESCHIKKING KOT

3. **Kot_deeljaar_lrk.sas7bdat**
   Dit bestand bevat een lijst met deeljaar beschikkingen KOT en bijbehorende gegevens kinderopvanginstellingen. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\DEELJAAR BESCHIKKING KOT

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

1075483 00067

4. **Doelgroep.sas7bdat**
   Dit bestand bevat een lijst met deeljaar beschikkingen KOT en indicatie doelgroep en dagtaaktype voor toeslagaanvrager en toeslagpartner. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\DEELJAAR BESCHIKKING KOT

5. **Conceptbeschikkingen_kot.sas7bdat**
   Dit bestand bevat een lijst met conceptbeschikkingen. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Brondata\KOT CONCEPTBESCHIKKINGEN

2.1.2. HALFPRODUCTEN

De halfproducten gebruikt als input voor het model zijn:

1. **BVR.sas7bdat**
   Dit bestand betreft een adressenbestand van de Belastingdienst met persoonsdetails. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\BVR

2. **Ouderkind_svb.sas7bdat**
   Dit bestand bevat een lijst met ouder-kind relaties. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\OUDER_KIND

3. **LRK.sas7bdat**
   Dit bestand bevat een lijst met kinderopvanginstellingen in Nederland. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\LRK

4. **Geblokkeerde_lrks.sas7bdat**
   Dit bestand bevat een lijst met LRKnummers in CAFzaken. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Geblokkeerde LRKs

5. **Postcodes.sas7bdat**
   Dit bestand bevat een lijst met de postcodes van Nederland. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\6PP

1075483 00067

6. **Groenevlagkot.sas7bdat**

   Dit bestand bevat een lijst met BSNs die niet mogen worden uitgesloten van beschikken. Voor het bestand en de documentatie zie de map:
   Q:\VEPROW63\TSL_DM_Handhavingsregie\Profiling 2013\Halfproducten\Eerdere Acties

2.2. FLOWS

Het SAS-script bestaat uit drie procesflows:

- Create DTM KOT
  - Deze flow is bestemd om het deelproduct DTM_KOT te maken.
- Data checks
  - In deze flow kunnen alle eventuele data checks in gebouwd worden. Je kunt hier alle tussengelegen stapjes in het model queryen en inspecteren.
- Analytics
  - In deze flow kunnen alle analytics binnen de data set gedaan worden

Onderstaand is de eerste flow 'Create DTM KOT' beschreven.

2.2.1. CREATE DTM KOT

In deze flow worden de beschikkingen plat geslagen tot een tabel op deeljaar niveau. Hieraan worden de conceptbeschikkingen toegevoegd, en vervolgens gekoppeld aan een aantal halfproducten.

Stap 1: Het toeslagjaar wordt (handmatig) bepaald.

Stap 2: De libraries worden gedefinieerd.

Stap 3: De brondata KOT_DEELJAAR_LRK wordt met een Left Join gekoppeld aan KOT_DEELJAAR_BEDRAGEN op de volgende velden / criteria:

- BSN
- Toeslagjaar
- Datum_dagtekening
- BsnKind
- Soort_instelling_sk
- Type_opvang_sk
- Ingangsdatum_Deeljaar_datum[2] < Einddatum_Deeljaar_datum[3]

- Einddatum_Deeljaar_datum[4] > Einddatum_Deeljaar_datum[5]

Hierbij wordt ook een selectie gemaakt op de velden die worden meegenomen in het outputbestand **Deeljaar_beschikking_kot**.

Stap 4: In deze stap worden de beschikkingen platgeslagen tot een tabel op deeljaar niveau, met alle relevante kolommen: **Deeljaar_beschikking_kot** wordt met een Left Join op BSN en Beschikkingsnummer gekoppeld aan de brondata beschikkingen_kot. Hierbij worden uit beschikkingen_kot meegenomen alleen de beschikkingen meegenomen die het juiste toeslagjaar betreffen. Daarnaast wordt er een selectie gemaakt op de velden.

Stap 4: De brondata doelgroep wordt met een Left Join op BSN, Beschikkingsnummer en Ingangsdatum gekoppeld aan de resulterende tabel. De outputtabel

Deeljaar_met_besch_en_doelgroep bevat nu de deeljaren in het huidige toeslagjaar, met daarbij contra-informatie over de kinderopvanginstellingen en doelgroep en dagtaaktype van de toeslagaanvrager en partner.

Stap 5: De brondata conceptbeschikkingen_KOT wordt ingelezen en de beschikkingen behorend bij het juiste toeslagjaar worden geselecteerd. Deze beschikkingen worden opgeslagen in de tabel **CONCEPTDEELJAREN2**. Hier is één nieuw veld aan toegevoegd, namelijk het Beschikkingsnummer. Deze wordt per aanvraag als volgt bepaald:

(PUT(CONCEPTDEELJAREN.AANVRAGER_BSN, Z9.) || "T" || PUT(toeslagjaar - 2000 ,2.) || "09991")

Alle beschikkingen met beschikkingsnummer eindigend op 09991 betreffen dus een conceptbeschikking.

Stap 6: De tabel **CONCEPTDEELJAREN2** wordt gesorteerd op BSN, Beschikkingsnummer, Ingangsdatum en BSN kind. Hierbij wordt een selectie gemaakt van alle velden, die grotendeels worden hernoemd. Het resultaat wordt opgeslagen in de nieuwe tabel **CONCEPTDEELJAREN3**. Deze bevat nu exact dezelfde velden als **Deeljaar_met_besch_en_doelgroep.**

Stap 7: De tabel **Deeljaar_met_besch_en_doelgroep** wordt samengevoegd met **CONCEPTDEELJAREN2.** Dit resulteert in de outputtabel **TOTAL_DEELJAAR_BESCH**. Dit is een tabel met alle formele en conceptbeschikkingen in het geselecteerde toeslagjaar.

Stap 8: Van het outputbestand uit de vorige stap worden de volgende velden geselecteerd:

[2] Uit bestand KOT_DEELJAAR_BEDRAGEN
[3] Uit bestand KOT_DEELJAAR_BEDRAGEN
[4] Uit bestand KOT_DEELJAAR_LRK
[5] Uit bestand KOT_DEELJAAR_LRK

- BSN
- Toeslagjaar
- Beschikkingsnummer
- BSN_Kind
- Maandtoeslag_per_kind

Deze informatie wordt opgeslagen in de tabel **Query_for_deeljaar_met_besch.**

Stap 9: De totale maandelijkse toeslag die een de toeslagaanvrager ontvangt, wordt hier berekend: **Query_for_deeljaar_met_besch** wordt gegroepeerd op BSN, Toeslagjaar, Beschikkingsnummer en Ingangsdatum. Hierbij wordt het nieuwe veld 'Toeslag_per_maand' toegevoegd aan de dataset. Dit veld wordt als volgt berekend: SUM(maandtoeslag_per_kind).

Het resultaat wordt opgeslagen in **Maandbedragen.**

Stap 10: De maandbedragen worden gekoppeld aan de tabel **TOTAL_DEELJAAR_BESCH**. Deze koppeling wordt gedaan op een aantal velden:

- BSN
- Toeslagjaar
- Beschikkingsnummer
- Ingangsdatum.

Het outputbestand is **TOTAL_DEELJAAR_BESCH2**.

Stap 11: Het outputbestand van de vorige stap wordt gegroepeerd op BSN en Beschikkingsnummer. Hierbij worden een aantal velden geselecteerd en de volgende nieuwe velden worden aangemaakt en berekend:

- INGANGSDATUM_bes: MIN(INGANGSDATUM)
- eindDATUM_bes: Max(eindDATUM)
- max_maandbedrag: max(toeslag_per_maand)

Deze informatie wordt opgeslagen in **QUERY_FOR_TOTAL_DEELJAAR_BE_0001**.

Stap 12: De beschikkingen uit **QUERY_FOR_TOTAL_DEELJAAR_BE_0001** worden gesorteerd op BSN, Toeslagjaar en Beschikkingsnummer. Vervolgens worden de meest actuele beschikkingen geselecteerd. Hier worden twee velden aan toegevoegd: verschil_vorige_bes en verschil_maandbedrag. Dit is het verschil met de voorgaande beschikking in jaarbedrag en maandbedrag respectievelijk.

Deze stap resulteert in het bestand **select_beschikking.**

Stap 13: De deeljaren uit **TOTAL_DEELJAAR_BESCH2** worden gekoppeld aan de meest actuele beschikkingen in **select_beschikking.** Deze koppeling wordt gedaan op de volgende velden / vereisten:

- BSN
- Beschikkingsnummer
- Jaarbedrag (uit tabel select_beschikking) > 0
- Maandbedrag (uit tabel TOTAL_DEELJAAR_BESCH2) > 0

Hierbij wordt een selectie gemaakt op de velden uit beide bestanden en worden een aantal velden toegevoegd, waaronder het percentage van de inkomen voor eigen bijdrage (per_inkomen_eigen_bijdrage), en de eigen bijdrage per maand (eigen_bijdrage_per_maand).

De outputtabel **Query_for_SELECT_BESCHIKKING** bevat nu de deeljaren horende bij de meest actuele beschikkingen met contra-informatie betreffende het inkomen en toeslagbedragen.

Stap 14: Hier wordt van alle deeljaren per BSN het meest actuele deeljaar geselecteerd: De outputtabel uit de vorige stap wordt gesorteerd op BSN, Toeslagjaar, Beschikkingsnummer en Ingangsdatum. Voor elke BSN wordt de beschikking behorende bij het laatste beschikkingsnummer geselecteerd. Deze beschikking betreft het meest actuele deeljaar. Hierbij worden een aantal velden berekend, waaronder de ingangsdatum en einddatum horende bij het deeljaar, het veld 'einddatum_bekend' (geeft aan of het einddatum van de beschikking bekend is) en hele_jaar (geeft aan of de beschikkking een geheel jaar betreft). Deze meest actuele beschikkingen met actuele deeljaren worden opgeslagen in de tabel **select_deeljaar_0.**

Stap 15: Voor alle beschikkingen in **select_deeljaar_0** wordt bepaald of er sprake is van Hot Hor. Er is sprake van Hot Hor indien een beschikking aan ten minste een van de volgende criteria voldoet:

- De aanvraag is nieuw en boven de Hot Hor grens (actueel maandbedrag of jaarbedrag)
- De aanvraag is niet nieuw, maar komt door een mutatie voor de eerste keer boven de Hot Hor grens
- De aanvraag is niet nieuw en heeft een mutatie omhoog groter dan 4000 euro (ook al is er eerder sprake geweest van Hot Hor.)

Indien er aan een Hot Hor criterium wordt voldaan wordt dit type opgeslagen in het veld 'Type_HotHor'

De resulterende tabel **select_deeljaar** bevat alle gegevens uit **select_deeljaar_0 en** extra contra-informatie over Hot Hor.

1075483 00067

Stap 16: Er wordt een tabel **SELECTIE_DEELJAAR** gecreëerd, welke alle contra-informatie horende bij de deeljaren uit **select_deeljaar_0** bevat. Aan deze tabel met deeljaren wordt **TOTAL_DEELJAAR_BESCH_2** gekoppeld op BSN, Beschikkingsnummer en Ingangsdatum.

Stap 17: Het halproduct BVR wordt met meerdere Left Joins op BSN (van aanvrager, kind en partner) gekoppeld aan **SELECTIE_DEELJAAR.** Hiermee worden gegevens (waaronder leeftijd, adres, Leeftijd van BSN) over de toeslagaanvrager, het kind en de toeslagpartner toegevoegd. Het outputbestand van deze stap heet **DEELJAAR_MET_BESCH_1b.**

Stap 18: de ouder-kind relatie uit het haldproduct Ouderkind_SVB wordt gekoppeld aan het outputbestand uit de vorige stap. Deze koppeling gebeurt met een Left Join op BSN van het kind. Het outputbestand **DEELJAAR_MET_BESCH_1C** wordt gesorteerd op BSN, Beschikkingsnummer, Ingangsdatum en BSN_Kind.

Stap 19: De IDs van de kinderopvanginstellingen worden vervangen door de postcodes of eventueel adressen van de kinderopvanginstellingen.

Stap 20: In deze stap wordt de tabel **DEELJAAR_MET_BESCH_4** aangemaakt. Hierbij wordt de tabel uit de vorige stap gesorteerd op BSN, Toeslagjaar, Beschikkingsnummer, Ingangsdatum en BSN_Kind en worden vervolgens nieuwe variabelen aangemaakt en de waarden voor elke BSN berekend. Deze variabelen betreffen onder andere het aantal kinderen in de opvang, de leeftijd van de ouder(s) bij de geboorte van het kind, een indicator die aangeeft of het kind te oud is voor BSO en/of DO, en een indicator die aangeeft of de nationaliteit van het kind gelijk is aan het nationaliteit van een van de ouders.

Stap 21: Het halfproduct 'Groenevlagkot' wordt gekoppeld aan tabel **DEELJAAR_MET_BESCH_4** op BSN. Hierbij wordt een selectie gemaakt van de velden. Dit wordt opgeslagen in: **DTM_KOT_01.**

Stap 22: De waarde van de variabele GroeneVlag wordt eventueel, als voldaan wordt aan bepaalde restricties, aangepast. Daarnaast wordt een nieuw veld 'Oudbedrag' toegevoegd. Dit geeft het jaarbedrag van de vorige beschikking en wordt als volgt berekend: B_JAARBEDRAG – verschil_vorige_bes. Dit wordt opgeslagen in **DTM_KOT_2.**

Stap 23: In deze stap wordt het eindbestand **DTM_KOT** gecreëerd. Dit bevat alle beschikkingen uit **DTM_KOT_2**, waarbij een selectie is gemaakt op de velden.

2.3. BESLISSINGEN

Voor `av_aant_kot_besch_2013` worden *alle* beschikkingen geteld, ook degenen die op 0 zijn gezet.

De laatste beschikking wordt gebruikt voor de indicatoren TENZIJ het jaarbedrag van de laatste beschikking op nul is gezet, in dit geval wordt ook de indicator `op_nul_gezet` geactiveerd (dit is puur voor controle achteraf, geen model indicator). We gebruiken dan de laatste beschikking die NIET op nul is beschikt.

Het actuele deeljaar word gebruikt, tenzij dit er niet is, dan gebruiken we het laatste deeljaar op de beschikking.

Als er van de BSN een conceptbeschikking ontvangen is, is B_LAATSTE_BESCH_IS_CONCEPT 1.

De postcode-huisnr-toevoeging opvang locatie wordt geparsed met de volgende reguliere expressie: `/([0-9]{4}[A-Z]{2})\s([0-9]+)[^0-9A-Z]*([0-9A-Z]*)/`

De postcode wordt omgezet naar longetude lattitude vanuit de 6PP database. (http://www.d-centralize.nl/projects/6pp/). Indien de postcode niet gevonden wordt in deze database. Wordt de gemiddelde longetude/lattitude combinatie gebruikt van de eerste 4 cijfers.

De afstand tussen opvanglocatie en adres aanvrager wordt als volgt berekend:
`(ARCOS(SIN(Lat)*SIN(Lat_kdv)+COS(Lat)*COS(Lat_kdv)*COS(Lon-Lon_kdv))*6371)`.

Indien het BSN_KIND niet aanwezig is, dan wordt de indicator bsn_kind_mist op 1 gezet. Verdere risico indicatoren (vlaggetjes) worden niet gezet voor dit kind.

Max uren BSO en DO worden per 30-8-2013 per kind opgeteld, en niet meer de hoogst voorkomende waarde.