

1.0 Informatie behoefte

Doelen

- “CO2-reductie/NOx/SOx: verlagen van het brandstofverbruik met 10% (brandstofverbruik/per ton vracht per zeemijl)”
- “Utilisatiegraad verbeteren: Reduceren van de wachttijden in de havens, de zogenaamde idle time, met 5%.”
- “Beladingsgraad; Verder optimaliseren van de beladingsgraad naar 90%”

Informatievragen

1. Beladingsgraad per schip per haven.
2. Wachttijden ('idle time') per haven en per schip inclusief trends (afgelopen jaren).
3. Brandstofverbruik (in relatie tot optimale snelheid) in tonnen brandstof per ton vracht per zeemijl; een en ander per type boot, vaarroute en seizoen.
4. Kosten per ton vracht per schip per route.

2.0 Opvallende punten in de raw data sources

2.1 Consignor Excel bestand

Kolommen die wel nodig zijn

- De discount in combinatie met de container (EuroPricePerKm) is zeker van belang.
- Mogelijk is discount ook indirect van toepassing op de beladingsgraad.
- De discount moet mogelijk ook verrekend worden in de kosten (informatievraag 4)

Kolommen die mogelijk niet nodig zijn

- De country/city van de consignor zijn mogelijk niet belangrijk.
- De naam van de consignor is niet belangrijk.

Opvallende punten

- R. Claus komt 2x voor, is dit dezelfde persoon of iemand met een dezelfde naam. (mogelijk vereist dit een groupby).
- Er is missende data (soms 0, void, of -).
- Er is een outlier van 89.

Idee

- Aan de hand van de discount, aantal containers, reisafstand, en type container kan de gegeven korting berekend worden. Ook de opbrengsten worden berekend.
- Beladingsgraad = Het vervoerde gewicht delen door het maximaal toelaatbare vervoerde gewicht. Dit voor iedere haven berekenen.

2.2 Container TxT bestand

Kolommen die wel nodig zijn

- Length & inhoud zijn mogelijk wel nodig, maar deze moeten misschien eerst omgezet worden naar TEU.
 - o $1 \text{ TEU} = \text{Lengte container in feet} / 20$ (1 foot = 0.3048m)
- EuroPricePerKm is sowieso nodig.
- RefrigerationFlag is van toe passing (sommige producten vereisen een gekoelde container)

Kolommen die mogelijk niet nodig zijn

- PowerFlag is het type outlet van een stroomaansluiting
 - o De powerflag heeft geen invloed op de informatiebehoeften.
 - o PowerFlag wordt soms met land iso code aangegeven.

Opvallende punten

- Items zitten niet altijd in de juiste container.
- Items zitten mogelijk in een te grote container.

Idee

- Eenheden van length en cubes zijn niet gegeven. Length van 6,10 vertaalt zich naar (20foot, TEU)

2.3 Item Access bestand

Kolommen die wel nodig zijn

- Item_storage_type is van toepassing om het type container te bepalen.
- Item_hazardflag is van toepassing om het type container te bepalen.
- Item_Category is van toepassing om het type container te bepalen
- Item_mfgr is mogelijk van toepassing omdat zij mogelijk de verkeerde type containers gebruiken (bv. Te grote container).

Kolommen die mogelijk niet nodig zijn

- Item_Description is van een te hoge granulariteit, kan me niet inbeeld waarvoor deze kolom handig kan zijn.
-

Opvallende punten

- Producten worden niet altijd in de juiste container geladen.
- Producten worden mogelijk onnodig in te grote container gestopt.
- Mogelijk zijn bovenstaande te koppelen aan item_mfgr.

Idee

n.v.t

- Item_mfgr bevat xxx waarde. Is dit om anoniem te blijven?

Kolommen die mogelijk niet nodig zijn

- Doet item_description ertoe?
- Doet item_category ertoe?
- Doet item_mfgr ertoe?
- Als bovenstaande waar is, dan kan er een groupby gedaan worden.

2.4 Port Access bestand

Bevindingen

- Wat moeten we met distance to oslo en distance to piraeus.
 - o Is het relevant om deze informatie überhaupt te bewaren?
- P_PortOrder is niet van het type int.
- Het PortId is niet opeenvolgend ingedeeld. Er missen dus records

Kolommen die mogelijk niet nodig zijn

- P_DistanceFromOslo: Relevantie?
- P_DistanceFromPiraeus: Relevantie?

Ideeën

- Hoe meer aanmeerplekken er zijn hoe langer je kan wachten. De port order is dus waarschijnlijk van belang, omdat dit de grote van de havens betreft.

2.5 Ship Access bestand

Bevindingen

- Naam geving van kolommen vreemd
- Enkel cosignor id 1t/m4 wordt gebruikt, toeval?

Belangrijke gegevens

- Sh_shipname: Relevant voor een lookup in het MRV Excel bestand
- Sh_maxTEU: Relevant voor het berekenen van de belading per reis
- sh_speedinknots: Relevant voor het berekenen van brandstofverbruik en CO2 uitstoot per zijmijl.

Kolommen die mogelijk niet nodig zijn

- sh_country: De vlag waaronder het schip vaart (Niet relevant?)
- sh_Yearcost: De onderhoudskosten van het schip per jaar (Wellicht relevant voor kostenplaatje. Is verder geen infobehoeft)
- sh_length: Meting van het schip (Relevantie?)
- sh_width: Meting van het schip (Relevantie?)
- sh_speedinkm_h: Je hebt al een gegeven (speedInKnots) (Relevantie?)

2.6 Shipment Access bestand

Bevindingen

- Foreign key tussen tabellen Shipment en Port (Kolom PortIdTo) bestaat niet

Belangrijke gegevens

- Distance: Afstand is belangrijk voor het berekenen van brandstof kosten en CO2 uitstoot.
- NumberContainers: Het aantal containers die geladen worden bij de huidige haven
- Gegevens Consignor: Eerder is de relevantie van de consignor details in twijfel gebracht. Wel is de korting wellicht belangrijk om kosten te berekenen.

Idee

Mocht de korting uit de consignor data relevant zijn, dan zou deze als extra kolom bijgevoegd kunnen worden om deze data te kunnen plotten en berekeningen te maken

2.7 shipmentdetail Access bestand

Kolommen die wel nodig zijn

- containNr is van belang, echter kan deze beter omgezet worden naar het aantal container dan slechts het bereik. Het is dan makkelijk om er mee te rekenen. Hoe de container genummerd zijn, is niet van belang. Het aantal container is belangrijk om dat het gebruikte container type variabele prijzen hebben.
- Item is van belang, omdat daarmee gekeken kan worden of het juiste container type/grootte gebruikt is.
- ShipmentId is uiteraard ook belangrijk.

Kolommen die mogelijk niet nodig zijn

N.v.t

Opvallende punten

- Het aangeven van de range van containers is een overbodig gegeven. Aantal is beter.

Idee

- Vooruit denkend moet waarschijnlijk container & item, gedenormaliseerd worden als shipmentdetail dimensie of zelfs in de shipment fact. Idem voor de discount van een cosignor. Deze moet betrokken worden in de fact/dimensie.

2.8 Voyage Access bestand

Kolommen die wel nodig zijn

- voyageld, shipId,

Kolommen die mogelijk niet nodig zijn

- Vp_legDateDepart is al aanwezig in de VoyagePort
- PortIdStart en PortIdEnt zijn ook uit VoyagePort te halen. Redundant.

Opvallende punten

Idee

Merk op dat de tabellen Voyage en VoyagePort niet genormaliseerd zijn en dus bijwerken anomalieën kunnen opleveren. Of beide tabellen geen tegenstrijdigheden bevatten is een aandachtspunt

2.9 VoyagePort Access bestand

Bevindingen

- VoyagePort en Shipment beschrijven bijna hetzelfde feit. Shipment kent meer gegevens over de lading en VoyagePort beschrijft hoe lang een schip over de reis doet.
- Vp_Traject_Distance is een gegeven in kilometer. Dit wijkt af van de andere gegevens (Zeemijl per uur)
- Vp_PortOrder: Voor dit gegeven is een kolom aangemaakt, maar is nergens ingevuld

Belangrijke gegevens

- Vp_LegDateDepart: Het gegeven van vertrek is belangrijk om de wachttijd in een haven te bepalen (komt terug op informatiebehoefte)
- Vp_LegDateArrival: Het gegeven van aankomst in de haven is belangrijk om de duur van de reis te bepalen.

Redundante gegevens

- Vp_PortIdCurrent: Dit gegeven is ook terug te vinden in de shipment tabel
- Vp_PortIdNext: Dit gegeven is ook terug te vinden in de shipment tabel
- Vp_Traject_Distance: Dit gegeven is in ander formaat (km_u) en is ook terug te vinden in de shipment tabel.
- Vp_PortOrder: Deze kolom is in geen enkele record ingevuld

Idee

Het samenvoegen van de VoyagePort en Shipment tabellen zal een hoop dubbele data weghalen. De data uit shipment en voyageport zijn namelijk bijna identiek. De unieke gegevens kunnen samengebundeld worden in de shipment tabel. Deze zal dan een meer op meer relatie kennen tussen de voyage en port tabellen.

2.10 MRV Bestand

- Schip Xin Beijing heet eigenlijk Xin Being
- Schip Mærsk Stepnica bestaat niet. Is dit MSC STEPNICA? Die bestaat namelijk wel.

Les week 6 Erik Tips

Voyage bevat dubbele records in combinatie met schip. Eerste instantie krijg je ongeveer 700 records daarna verwacht je er rond de 100.