

OPDRACHT 2 VOOR PORTFOLIO

STAP 1:

Na wat rondzoeken heb ik een dataset over iemand's verzameling fifastickers van Panini omtrent het aankomende WK in Qatar gekozen omdat er geen copyright aan gekoppeld was

en omdat de data nog voldoende overzichtelijk was hoewel ik er enkele aanpassingen in kan maken om de dataset op te kuisen.

Link: https://p.cygnus.cc.kuleuven.be/webapps/blackboard/content/listContent.jsp?course_id=_1126307_1&content_id=_35483416_1

STAP 2: Opkuisen

1) De eerste kolom heeft een onnodige nummering aangezien het id van een kaartje steeds uniek is

voor dit specifieke prentje. We verwijderen deze kolom dus allereerst.

2) De kolommenbenamingen zijn wat onduidelijk omdat sommige in het Engels zijn en andere dan weer in het Spaans.

Om het universeeler te maken, zetten we alle kolomtitels in het Engels, laten we ze beginnen met een kleine letter en vermijden we afkortingen.

3) De laatste kolom geeft aan of hij de kaart al eerder had gepacked heeft of niet. In principe kunnen we in deze

kolom simpelweg aangeven hoeveel dubbele stickers hij van dit exemplaar heeft waardoor we de rijen met dubbele exemplaren kunnen verwijderen (bv. bij het prentje van Xhaka). Alhoewel dit een erg groot en precies werkje is zou dit ervoor moeten zorgen dat het prentjes-id nu inderdaad als KEY gebruikt kan worden/

4) In kolom 9 is "attack" fout geschreven en dit passen we dus aan.

5) In kolom 5 start de string soms met een hoofdletter en soms niet dus we maken het universeel door de

gehele kolom met een kleine letter te laten beginnen behalve voor FIFA want dit mag eigenlijk niet in kleine letters geschreven worden).

6) FIFA is daarnaast een organisatie en geen land dus ook dit mag verwijderd worden.

7) Kolom 7 heet first_name wat een voornaam van een speler aanduidt, maar in deze kolom kunnen ook

de namen van stadia, etc. staan dus we veranderen de naam van deze kolom in mainname waardoor benaming van kolom

8 nog wel gewoon hetzelfde kan blijven en gericht is op de achternamen bij spelerskaarten.

8) Bij squadkaarten heb je zowel in kolom 7 als 8 een invoer namelijk "Team" en "Shot" wat erg omslachtig is,

en daarom combineren we ze simpelweg in kolom 7. Hetzelfde geldt voor de logo- en museumprentjes.

9) Als laatste willen we graag dat de lengte van de spelers overal hetzelfde aantal cijfers na de komma heeft zodat we

makkelijk een variabele ervoor kunnen definiëren.

STAP 3: Tabel aanmaken

Commando's in query-tool van database TX36:

```
CREATE TABLE r0932517.importcsv(  
card_id int NOT NULL,  
figure_id char(6) NOT NULL,  
buy_id smallint NOT NULL,  
package_id smallint NOT NULL,  
figure_type varchar(25) NOT NULL,  
country varchar(20),  
mainname varchar(25) NOT NULL,  
last_name varchar(25),  
player_position varchar(10),  
birth date,  
weight numeric(3,1),  
height numeric(3,2),  
amount_doubles smallint NOT NULL,  
CONSTRAINT pk_importcsv_card_id PRIMARY KEY ( card_id )  
);
```

STAP 4: Importeren

Nu importeren we ons opgekuist CSV-bestand bij onze nieuwe tabel en zijn uiteindelijk erg opgelucht dat alles van de eerste keer correct geïmporteerd wordt. We werken nu in de Query-tool van de tabel.

STAP 5: Vragen

1) Van welke landen heeft hij de meeste spelers in zijn verzameling?

```
SELECT country, count(*)  
FROM r0932517.importcsv  
GROUP BY country  
HAVING country IS NOT NULL  
ORDER BY 2 DESC
```

2) Wat is de gemiddelde lengte van de spelers die hij al heeft uit een bepaald land (van klein naar groot)?

```
SELECT country, avg(height) AS gemiddelde_lengte  
FROM r0932517.importcsv  
GROUP BY country  
HAVING country IS NOT NULL  
ORDER BY 2
```

3) Welke spelers die lichter zijn dan 70kg heeft hij momenteel in zijn bezit?

```
SELECT mainname, last_name, weight  
FROM r0932517.importcsv  
WHERE weight IS NOT NULL AND weight<70  
ORDER BY 3
```

4) Uit welk jaar komt de oudste speler uit elk land in zijn bezit? En de jongste? (Op alfabetische volgorde van landen)

```
SELECT country, min(birth) AS geboortedatum_oudste_speler, max(birth) AS  
geboortedatum_jongste_speler  
FROM r0932517.importcsv  
GROUP BY country
```

```
HAVING country IS NOT NULL  
ORDER BY COUNTRY
```

5) Welke spelers van Duitsland heeft hij in zijn verzameling?

```
SELECT country, mainname, last_name  
FROM r0932517.importcsv  
WHERE country = 'Germany' AND last_name IS NOT NULL  
ORDER BY 2
```