



OpenRefine

Minor Datavisualisatie & Infographics

Handleiding OpenRefine



Ellis Buth

Mendy Hogenboom

Svenja van Leuven

Amélie Ong

Fabiënne van Wijaarda

Inleiding

In deze handleiding wordt stap voor stap uitgelegd wat de basis van OpenRefine is. Met behulp van screenshots leiden we je door de belangrijkste functies, zodat je daarna genoeg kennis hebt om zelf met de tool aan de slag te gaan. We hopen dat na het doorlopen van deze tutorial duidelijk is wat de kracht van OpenRefine is en hoe je de tool effectief kunt inzetten in je eigen projecten.

Binnen de minor Datavisualisatie en Infographics wordt geleerd hoe data vertaald kan worden naar een visualisatie. De eerste stap hiervoor is data opschonen die nog niet geschikt is voor een visualisatie, want in de praktijk zijn datasets vaak rommelig, inconsistent en vol met fouten.

Deze handleiding focust zich bewust niet op alle aspecten van de tool, maar op de functies die het meest relevant zijn voor de basis van de tool.

Leerdoelen van deze handleiding

Het doel van deze handleiding is om de basisvaardigheden in OpenRefine te leren. In de handleiding komen de meest voorkomende stappen in OpenRefine voor. Hiermee hebben beginners een goede basis. De belangrijkste leerdoelen van deze handleiding zijn:

- 1 Een dataset succesvol inladen en voorbereiden.
- 2 Inconsistenties in data snel identificeren.
- 3 Data systematisch opschonen en standaardiseren.
- 4 Data in bulk transformeren voor consistentie en leesbaarheid.
- 5 Opgeschoonde dataset exporteren voor verder gebruik.

Inhoud

1

Introductie tool

1.1 Wat is OpenRefine	4
1.2 Geschiedenis	4
1.3 Kernfuncties	4
1.4 Kansen	5
1.5 Beperkingen	6
1.6 Waarom OpenRefine en niet Excel	6

2

Uitleg

2.1 Dowloaden OpenRefine	7
2.2 Dataset Inladen	9
2.3 Text Facet	10
2.4 Numeric Facet	11
2.5 Cluster and Edit	14
2.6 Transform	16
2.7 Common transforms	17
2.8 Add column based on this column	18
2.9 Split into several columns	21
2.10 Timeline Facet	22
2.11 Facet by blank	24
2.12 History & Undo/Red	25
2.13 Exporteren	26

3

Verder leren

3.1 GREL-formules	28
3.2 Verder leren met externe bronnen	30

4

Bijlagen

4.1 Cheatsheet	31
4.2 Antwoordenlijst vragen	34
4.2 Literatuurlijst	35

1 - Introductie tool

1.1 Wat is OpenRefine

OpenRefine is een gratis open-source programma dat wordt gebruikt om datasets op te schonen, transformeren en verrijken. Het wordt vooral gebruikt door data-analisten, onderzoekers en journalisten. Zij gebruiken dit, omdat het erg handig is om rommelige gegevens te verwerken voordat deze worden geanalyseerd of gepubliceerd.

OpenRefine draait lokaal op de computer, maar wordt geopend in de webbrowser. Dit betekent dat het niet in de Cloud draait en data dus veilig op de computer blijft. Gevoelige data zal dus niet zomaar gedeeld worden met externe servers. Het kan gekoppeld worden aan online functies, deze zijn online dus blijft de data niet alleen maar lokaal.

1.2 Geschiedenis

De geschiedenis van OpenRefine begint in 2010 toen Metaweb Technologies de open-source tool Freebase Gridworks ontwikkelde. Deze tool is bedacht door David Huynh. De tool was in eerste instantie bedoeld om data op te schonen en te transformeren voor import in Freebase, een grote kennisbank.

In juli 2010 werd Metaweb overgenomen door Google en de tool kreeg een nieuwe naam en dat werd Google Refine. Het project bleef open-source en werd tussen 2010 en 2012 door Google-ingenieurs en de community onderhouden. Google was eigenlijk voornamelijk geïnteresseerd in de data van Freebase en niet in de functionaliteit van de tool zelf.

In oktober 2012 kondigde Google dan ook aan dat de actieve ondersteuning ging stoppen. Het project werd volledig overgedragen aan de open-sourcegemeenschap. Die de tool hernoemde naar OpenRefine en de code naar GitHub verhuisde. Sindsdien is het een door de community geleid, en later door subsidies ondersteund.

1.3 Kernfuncties

OpenRefine is een krachtige, gratis en open-source tool voor het werken met "rommelige data". Het programma draait lokaal op de computer, waardoor de gegevens privé blijven. Dit is handig voor het werken met gevoelige data.

Data opschonen en transformeren is de kern van OpenRefine. Met de functie Faceting is er snel een overzicht van alle unieke waarden in een kolom te krijgen, wat helpt om inconsistenties te vinden. Clustering groepeert automatisch gelijkende waarden (bijvoorbeeld "McDonalds" en "Mc Donalds") en stelt deze in staat deze samen te voegen.

Hele kolommen kunnen ook tegelijk bewerkt worden met "common transforms" (zoals spaties verwijderen) of via de GREL-taal, bijvoorbeeld om tekst naar hoofdletters om te zetten. Ook kan data van het ene formaat naar het andere omgezet worden.

In OpenRefine kan een dataset uitgebreid worden met externe data. Een functie genaamd Reconciliation is hierbij een sleutelfunctie: het koppelt waarden in de dataset aan entiteiten in externe databases zoals Wikidata. Dit standaardiseert niet alleen de data, maar stelt de dataset ook in staat om nieuwe informatie uit die externe bronnen toe te voegen. Ook kan data opgehaald worden via API-calls door URL's te gebruiken.

Tot slot is er een goede gebruiksvriendelijkheid en controle. De tool heeft een gebruiksvriendelijke point-and-click interface, wat het toegankelijk maakt voor gebruikers zonder programmeerkennis. Een belangrijke functie is de oneindige undo/redo, waarmee elke bewerking teruggedraaid kan worden en de volledige bewerkingsgeschiedenis kunt opslaan en opnieuw kunt toepassen. Dit zorgt voor volledige controle en reproduceerbaarheid van eerder werk.

1.4 Kansen

Er zijn veel kansen om OpenRefine technisch beter en krachtiger te maken. Ontwikkelaars werken aan het moderniseren van de software door de voorkant (de knoppen die zichtbaar zijn) los te koppelen van de achterkant (de motor die het werk doet). Dit is een belangrijke stap, want hierdoor kan OpenRefine in de toekomst samenwerken met extra krachtige "data-engines". Hierdoor zouden gebruikers nog grotere datasets kunnen bewerken zonder dat hun computer traag wordt. Ook wordt de gebruikersinterface verbeterd zodat deze sneller werkt en meer aanvoelt als een moderne spreadsheet.

Ook ligt er een grote kans in het bereiken van een breder publiek en het versterken van de gemeenschap. OpenRefine is erg gebruiksvriendelijk hierdoor kunnen ook mensen zonder technische kennis, zoals journalisten, bibliothecarissen en studenten, de tool gebruiken om "verhalen met data te vertellen". Door tutorials te maken met relevante Nederlandse voorbeelden, wordt het voor nog meer mensen duidelijk hoe ze de tool in hun eigen werk kunnen inzetten. De sterke samenwerking met Wikidata (de database achter Wikipedia) maakt OpenRefine voor iedereen die wil bijdragen aan 's werelds grootste open kennisbank.

Tot slot is er de kans om de toekomst van het project veilig te stellen door te zorgen voor stabiliteit en groei. Zoals hierboven genoemd wordt OpenRefine niet meer door één groot bedrijf geleid, maar door een actieve gemeenschap. Het project ontvangt nu subsidies van grote organisaties, zoals de Wikimedia Foundation en de Chan Zuckerberg Initiative. Hierdoor kunnen ze een professioneel team betalen om constant aan de tool te werken. Door deel te nemen aan programma's zoals Google Summer of Code, worden jonge en getalenteerde ontwikkelaars aangetrokken.

1.5 Beperkingen

Hoewel OpenRefine een krachtige tool is, heeft het ook beperkingen. OpenRefine heeft wat beperkingen als het grote datasets zijn. OpenRefine draait volledig lokaal op een computer. Dit is erg handig voor privacygevoelige data maar het betekent ook dat de snelheid en de grootte van de data afhangen van hoe krachtig een computer is. Het is niet gebouwd als een cloudoplossing die de rekenkracht van Google kan gebruiken.

Een andere uitdaging is dat de technologie achter de schermen een beetje verouderd is. Dit maakt het voor nieuwe ontwikkelaars soms lastig om in te springen en mee te helpen bouwen aan verbetering van de tool. Functies die meestal in een moderne spreadsheets tools zitten bijvoorbeeld het aanpassen van kolombreedtes zijn nog steeds niet beschikbaar.

Daarnaast is het project sterk afhankelijk van vrijwilligers en subsidies. Hoewel er een geweldige gemeenschap is, betekent dit dat de ontwikkeling van nieuwe functies soms lang kan duren. Er is een lange lijst met wensen, zoals betere ondersteuning voor nog grotere datasets en een volledig nieuwe interface. Door dit is er geen garantie wanneer deze klaar zullen zijn. Het is een project van de gemeenschap, wat betekent dat de vooruitgang afhangt van de tijd en middelen die beschikbaar zijn.

1.6 Waarom OpenRefine en niet Excel

Excel is een handig programma om berekeningen te maken, grafieken te tekenen en nette data bij te houden. OpenRefine is meer een specialistische tool die gebruikt wordt voor een grote, rommelige dataset eerst opgeruimd moet worden.

Het eerste verschil is hoe OpenRefine fouten kan vinden. De functie Faceting toont direct een overzicht van alle unieke waarden in een kolom, hierdoor kunnen spelfouten en inconsistenties meteen opmerkt worden. Vervolgens kan de Clustering-functie deze gelijkende waarden automatisch groeperen en voorstellen om ze samen te voegen. Dit is veel krachtiger dan handmatig zoeken en vervangen in Excel.

Een tweede, unieke functie is Reconciliation. Hiermee kan OpenRefine data in een bestand, zoals straatnamen of personen, koppelen aan een externe, officiële database zoals Wikidata. Nadat de koppeling is gemaakt, kan de eigen dataset verrijkt worden met extra informatie uit die bron bijvoorbeeld coördinaten of geboortedata. Dit is iets wat Excel niet zomaar kan.

Tot slot houdt OpenRefine een complete geschiedenis bij van al de bewerkingen. Dit maakt het mogelijk om altijd terug te gaan naar een eerdere stap, en een hele reeks opschoonacties opslaan en opnieuw toepassen op een nieuwe dataset. Dit maakt eerder werk volledig transparant en herhaalbaar, wat in Excel veel lastiger is.

Excel is het meest geschikt voor het analyseren en visualiseren van schone data. OpenRefine is beter om de chaos in rommelige data eerst op te ruimen en om data te verrijken.

2 - Uitleg

In dit hoofdstuk laten we stap voor stap zien hoe je met OpenRefine aan de slag gaat. We behandelen de belangrijkste functies, van het inladen van een dataset tot het opschonen, transformeren en exporteren van gegevens. Bij bijna alle functies vind je een korte oefenvraag, zodat je de uitleg direct kunt toepassen en zelf kunt oefenen.

2.1 Downloaden OpenRefine

2.1.1 Windows

Stappen:

1. Ga naar de OpenRefine website: <https://openrefine.org>
2. Download de windows-versie, kies de Windows **zip-file** en download deze.

Other platforms and versions

Version	Windows (installer)	Windows (zip file)	Mac OS	Linux
3.9.5 (2025-09-02)	↓ EXE	↓ ZIP	↓ DMG	↓ TAR.GZ
3.8.7 (2024-11-21)	↓ EXE	↓ ZIP	↓ DMG	↓ TAR.GZ
3.7.9 (2024-02-10)		↓ ZIP	↓ DMG	↓ TAR.GZ

See the [full list of releases on GitHub](#) for other versions.

3. Zodra het .zip-bestand is gedownload kies je voor **alles uitpakken**.

The screenshot shows a Windows File Explorer window with the following files and folders:

Naam	Type	Gecomprimeerde gr...	Met wacht...	Grootte	Compressi...	Gewijzigd op
licenses	Bestandsmap					2-9-2025 22:49
server	Bestandsmap					2-9-2025 23:04
webapp	Bestandsmap					2-9-2025 23:04
LICENSE	Text Source File	1 kB	Nee	2 kB	50%	2-9-2025 22:49
licenses	Microsoft Edge HTML Do...	2 kB	Nee	13 kB	92%	2-9-2025 22:56
openrefine	Toepassing	32 kB	Nee	125 kB	75%	2-9-2025 23:03
openrefine.i4j	INI Source File	1 kB	Nee	1 kB	40%	2-9-2025 22:49
README	Markdown Source File	2 kB	Nee	4 kB	59%	2-9-2025 22:49
refine	Windows-batchbestand	4 kB	Nee	10 kB	68%	2-9-2025 22:49
refine	INI Source File	1 kB	Nee	2 kB	44%	2-9-2025 22:49

4. Open vervolgens de OpenRefine Toepassing met het diamant icoontje. (Soms krijg je een beveiligingswaarschuwing van windows defender, wanneer dit gebeurt klik je op meer informatie en dan op toch uitvoeren.

Vandaag				
LICENSE	2-10-2025 14:43	Text Source File	2 kB	
licenses	2-10-2025 14:43	Microsoft Edge HT...	13 kB	
openrefine	2-10-2025 14:43	Toepassing	125 kB	
openrefine.l4j	2-10-2025 14:43	INI Source File	1 kB	
README	2-10-2025 14:43	Markdown Source ...	4 kB	
refine	2-10-2025 14:43	Windows-batchbe...	10 kB	
refine	2-10-2025 14:43	INI Source File	2 kB	
licenses	2-10-2025 14:43	Bestandsmap		
server	2-10-2025 14:43	Bestandsmap		
webapp	2-10-2025 14:43	Bestandsmap		

5. Vervolgens opent OpenRefine in je webbrowser.

2.1.2 Mac OS

Stappen:

1. Ga naar de OpenRefine website: <https://openrefine.org>
2. Download de MacOS-versie, kies de Mac OS **.dmg-file** en download deze.

Other platforms and versions				
Version	Windows (installer)	Windows (zip file)	Mac OS	Linux
3.9.5 (2025-09-02)	⬇ EXE	⬇ ZIP	⬇ DMG	⬇ TAR.GZ
3.8.7 (2024-11-21)	⬇ EXE	⬇ ZIP	⬇ DMG	⬇ TAR.GZ
3.7.9 (2024-02-10)		⬇ ZIP	⬇ DMG	⬇ TAR.GZ

See the [full list of releases on GitHub](#) for other versions.

3. Open **Finder** en ga naar de map **Applicaties**, zoek hier het OpenRefine icoon en klik op de rechtermuisknop om deze te **openen**. (soms krijg je een beveiligingswaarschuwing dat de applicatie van het internet komt, klik dan nogmaals op openen)

4. Vervolgens opent OpenRefine in je webbrowser.

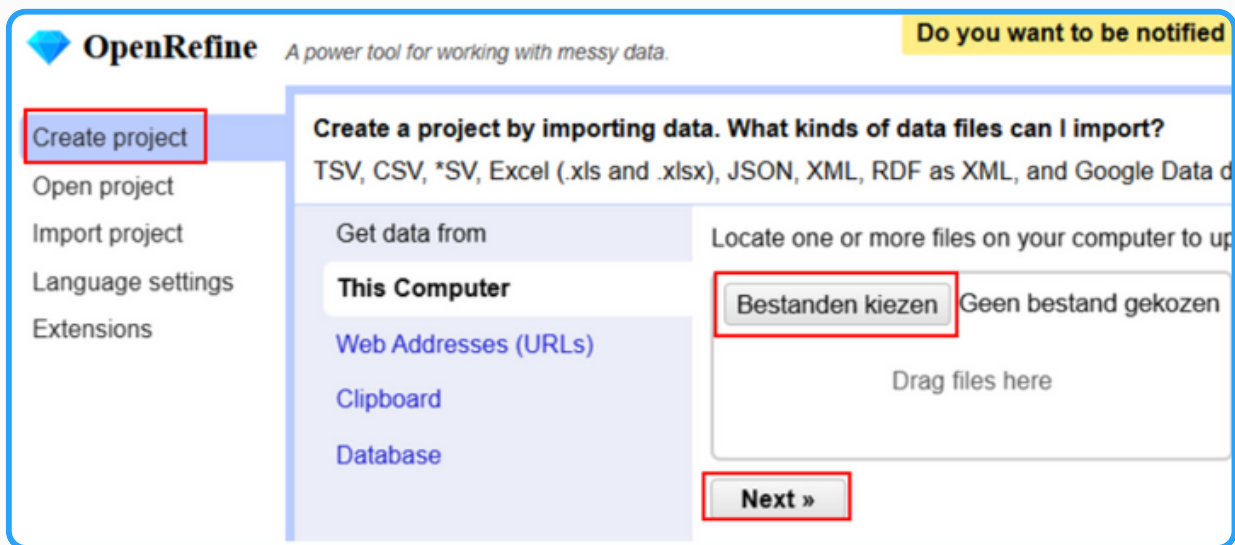
2.2 Dataset inladen

Voordat je met OpenRefine kunt werken, moet je eerst een dataset importeren.

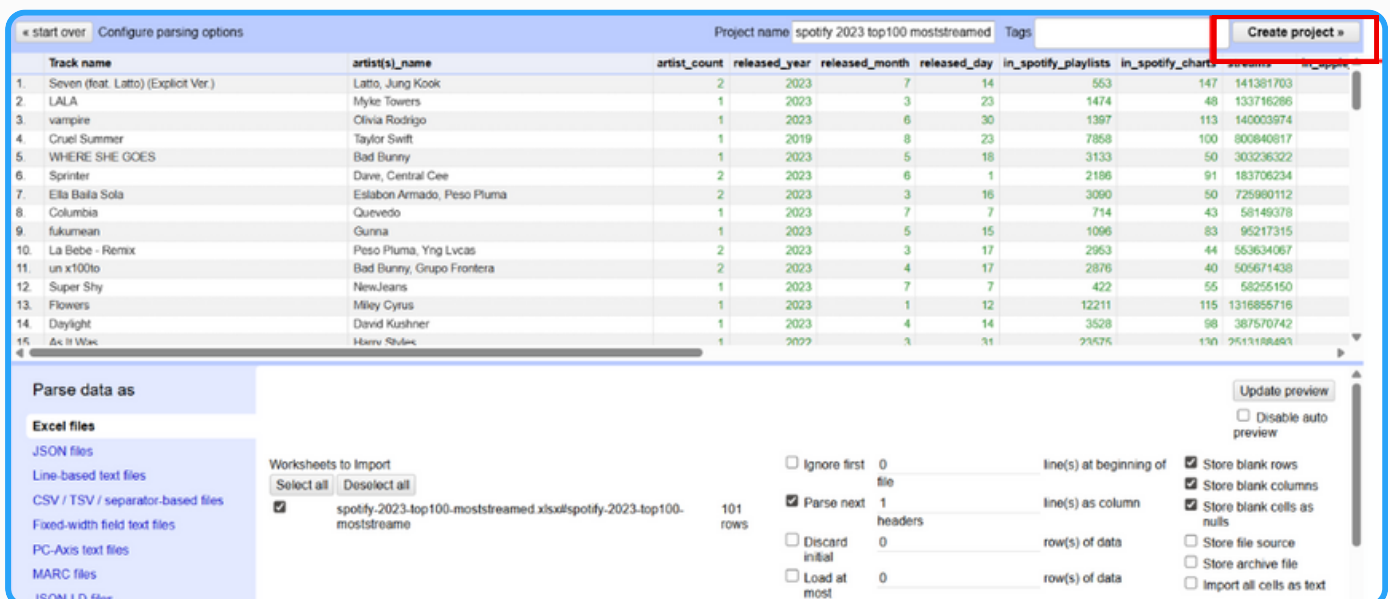
OpenRefine kan omgaan met bestanden zoals CSV, TSV en Excel. Tijdens het inladen zie je een preview van de data, zodat je kunt controleren of de kolommen goed worden ingelezen.

Stappen:

1. Start OpenRefine.
2. Klik op Create Project.
3. Kies het bestand *spotify2023_clean.csv* en druk op **Next**.



4. Controleer in de preview of de kolommen goed ingelezen zijn (track_name, artists_name, streams).
5. Klik bovenaan op **Create Project** om de dataset te openen.



2.3 Text Facet

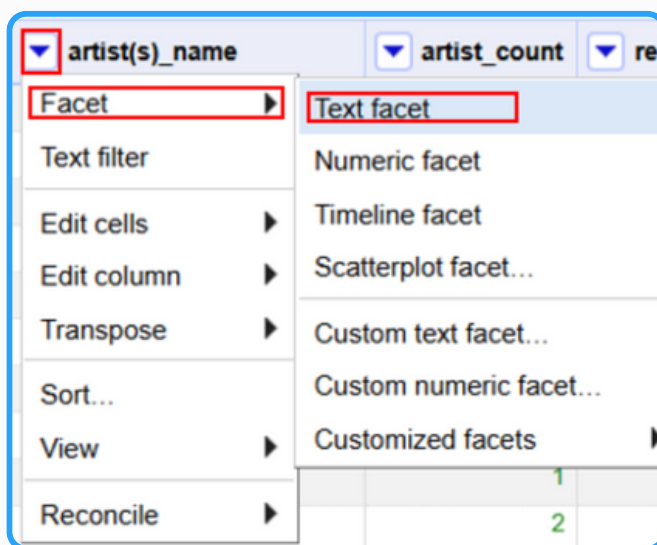
Een **Text Facet** laat je in één oogopslag zien welke unieke waarden er in een kolom staan en hoe vaak ze voorkomen. Dit is vooral nuttig om snel overzicht te krijgen van categorieën of namen.

Voorbeeld in Spotify-data:

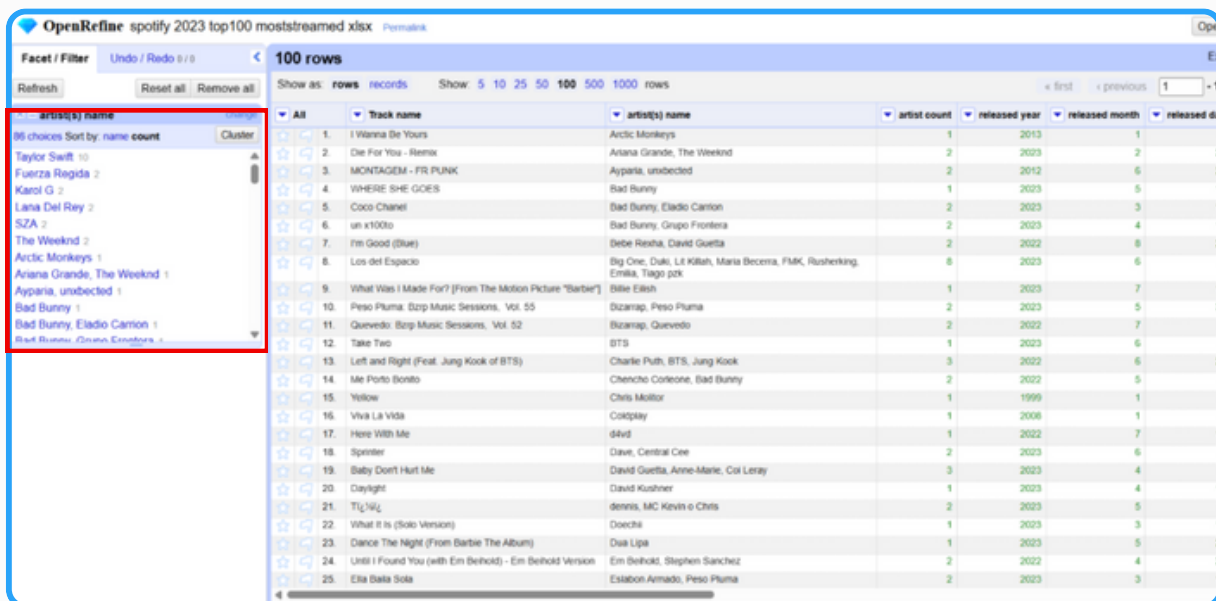
Je kunt een Text Facet gebruiken op de kolom **artist(s)_name** om te zien welke artiesten meerdere keren in de top 100 staan.

Stappen:

1. Klik op het pijltje bij de kolom **artists_name**.
2. Kies Facet → Text Facet.



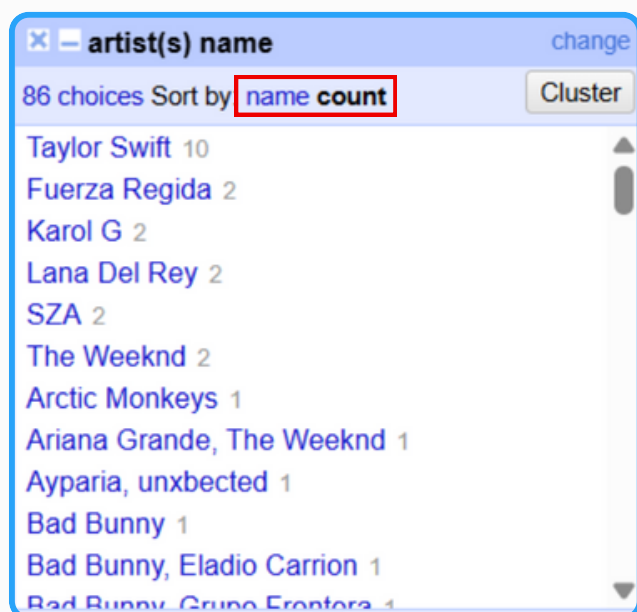
3. Helemaal links verschijnt een lijst met alle unieke artiesten en hoe vaak ze voorkomen.

A screenshot of the OpenRefine interface showing the '100 rows' view of the Spotify data. The 'artist(s)_name' column is faceted, and a list of unique artists is shown on the left. The list includes Taylor Swift (10), Fuerza Regida (2), Karol G (2), Lana Del Rey (2), SZA (2), The Weeknd (2), Arctic Monkeys (1), Ariana Grande, The Weeknd (1), Ayrina, unobscured (1), Bad Bunny (1), and Bad Bunny, Eladio Carmon (1). The main table shows the top 100 tracks with columns for Track name, artist(s) name, artist count, released year, released month, and released day. The first row is 'I Wanna Be Yours' by Arctic Monkeys, released in 2013.

Track name	artist(s) name	artist count	released year	released month	released day
1. I Wanna Be Yours	Arctic Monkeys	1	2013	1	1
2. Die For You - Remix	Ariana Grande, The Weeknd	2	2023	2	2
3. MONTAGEM - FR PUNK	Ayrina, unobscured	2	2012	6	6
4. WHERE SHE GOES	Bad Bunny	1	2023	5	5
5. Coco Chanel	Bad Bunny, Eladio Carmon	2	2023	3	3
6. un x1000	Bad Bunny, Grupo Frontera	2	2023	4	4
7. I'm Good (Blue)	Bebe Rexha, David Guetta	2	2022	8	8
8. Los del Espacio	Big One, Duke, Li Kilah, Maria Becerra, FMK, Rusherking, Emika, Tiago pzk	8	2023	6	6
9. What Was I Made For? (From The Motion Picture "Barbie")	Billie Eilish	1	2023	7	7
10. Peso Pluma: Bzrp Music Sessions, Vol. 55	Bizarrap, Peso Pluma	2	2023	5	5
11. Quevedo: Bzrp Music Sessions, Vol. 52	Bizarrap, Quevedo	2	2022	7	7
12. Take Two	BTS	1	2023	6	6
13. Left and Right (Feat. Jung Kook of BTS)	Charlie Puth, BTS, Jung Kook	3	2022	6	6
14. Me Porto Bonito	Chencho Corleone, Bad Bunny	2	2022	5	5
15. Yellow	Chris Mottola	1	1999	1	1
16. Viva La Vida	Coldplay	1	2006	1	1
17. Here With Me	d4vd	1	2022	7	7
18. Sprinter	Dave, Central Cee	2	2023	6	6
19. Baby Don't Hurt Me	David Guetta, Anne-Marie, Colley	3	2023	4	4
20. Daylight	David Kushner	1	2023	4	4
21. Tíu/14	deniro, MC Kevin o Chris	2	2023	5	5
22. What It Is (Solo Version)	Doechii	1	2023	3	3
23. Dance The Night (From Barbie The Album)	Dua Lipa	1	2023	5	5
24. Until I Found You (with Em Behold) - Em Behold Version	Em Behold, Stephen Sanchez	2	2022	4	4
25. Ella Baila Sola	Estaban Armado, Peso Pluma	2	2023	3	3

4. Bovenaan deze lijst kun je wisselen tussen **name** en **count**:

- Bij **name** worden de artiesten alfabetisch gesorteerd.
- Bij **count** worden ze gesorteerd op frequentie: de artiest die het vaakst voorkomt staat bovenaan.



Oefenvraag Text Facet

Hoeveel nummers uit de Spotify Most Streamed 2023 Top 100 zijn uitgebracht in 2022?

antwoord op blz.34

2.4 Numeric Facet

De **numeric facet** sorteert alle getallen in de kolom van laagste naar hoogste. Hiervan wordt er een histogram gemaakt waar snel in te zien is hoe vaak getallen voorkomen. Het gaat hierbij niet om exacte getallen zoals bij de text facet, maar om de verdeling van de getallen binnen het bereik te zien.

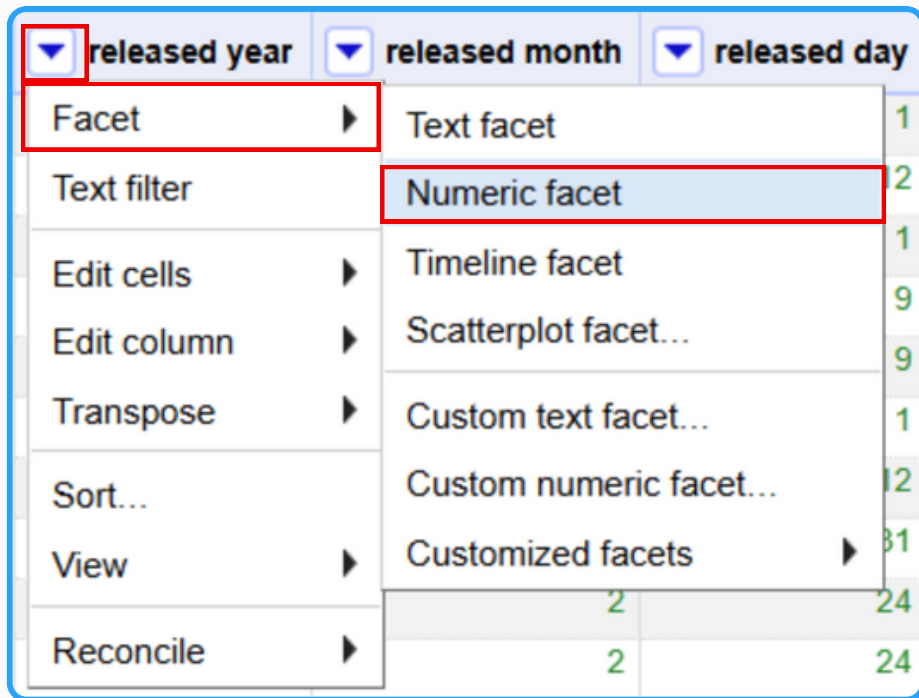
Aan de zijkanten van het histogram zitten schuifjes waarmee je kunt filteren. Zodat alleen de waardes binnen de schuifjes zichtbaar zijn.

Voorbeeld in Spotify-data:

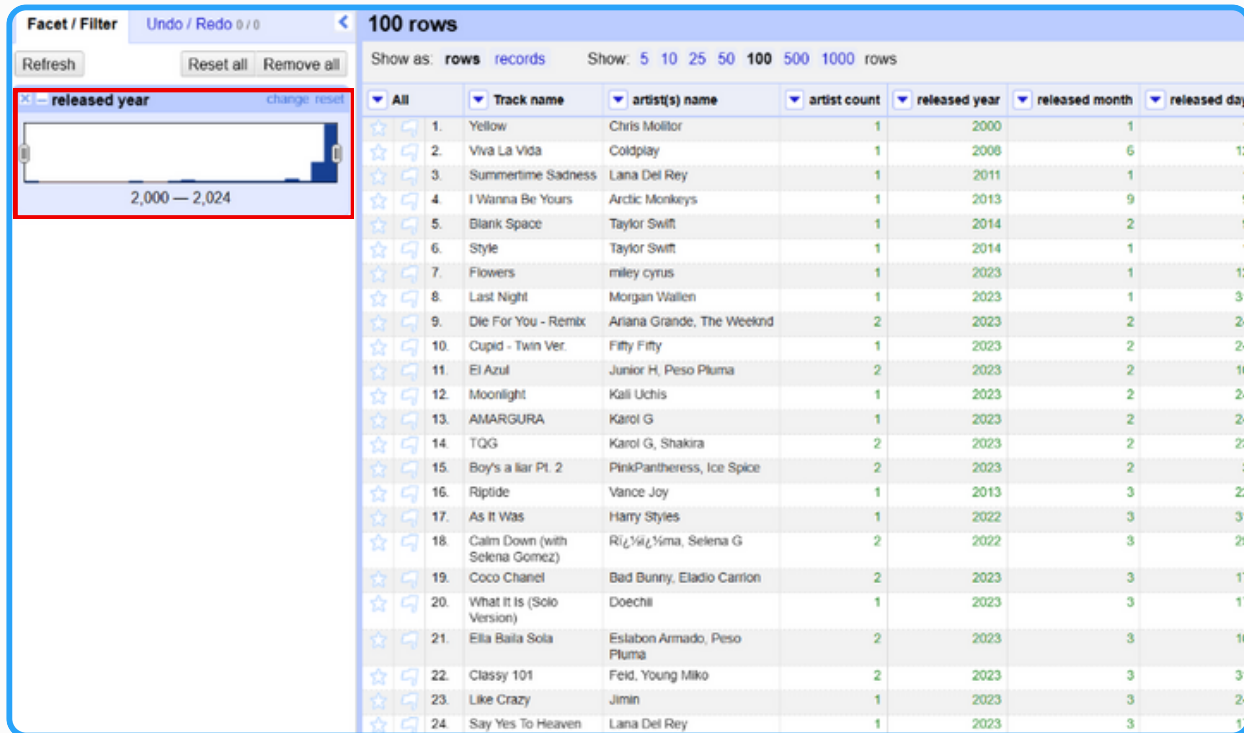
Je kunt de numeric facet gebruiken op de kolom **released_year** om te zien in welk jaar/jaren de meeste nummers gereleased zijn die in de top 100 van spotify staan. Met de schuifjes kan je vervolgens verder filteren op bepaalde periodes/jaren.

Stappen:

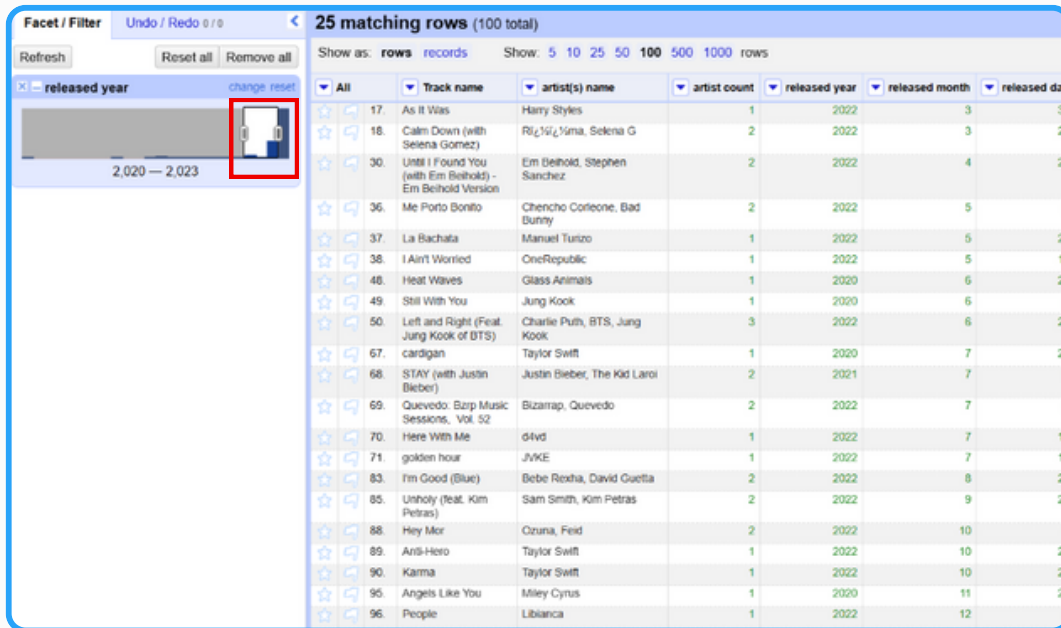
1. Klik op het pijltje bij de kolom **released_year**.
2. Kies **Facet** → **Numeric Facet**.



3. Links verschijnt een histogram met de verdeling van de getallen.



4. Door de schuifjes te verplaatsen kun je filteren, in de tabel is vervolgens te zien waarop je hebt gefilterd.



The screenshot shows a data table with a facet filter on the left. The facet is for 'released year' and has a range of 2,020 to 2,023. A red box highlights the 'change reset' button. The table has 25 matching rows (100 total) and shows columns for Track name, artist(s) name, artist count, released year, released month, and released day. The table is sorted by released year, with rows 17 to 96 visible.

	Track name	artist(s) name	artist count	released year	released month	released day
17.	As It Was	Harry Styles	1	2022	3	3
18.	Calm Down (with Selena Gomez)	Rihanna, Selena G	2	2022	3	25
30.	Unstoppable (feat. Em Beihold)	Em Beihold, Stephen Sanchez	2	2022	4	22
36.	Me Porto Bonito	Chencho Corleone, Bad Bunny	2	2022	5	6
37.	La Bachata	Manuel Turizo	1	2022	5	26
38.	I Ain't Worried	OneRepublic	1	2022	5	15
48.	Heat Waves	Glass Animals	1	2020	6	26
49.	Still With You	Jung Kook	1	2020	6	5
50.	Left and Right (Feat. Jung Kook of BTS)	Charlie Puth, BTS, Jung Kook	3	2022	6	24
67.	cardigan	Taylor Swift	1	2020	7	24
68.	STAY (with Justin Bieber)	Justin Bieber, The Kid Laroi	2	2021	7	1
69.	Quevedo: Bzrp Music Sessions, Vol. 52	Bizarrap, Quevedo	2	2022	7	6
70.	Here With Me	d4vd	1	2022	7	11
71.	golden hour	JVKE	1	2022	7	15
83.	I'm Good (Blue)	Bebe Rexha, David Guetta	2	2022	8	26
85.	Unholy (feat. Kim Petras)	Sam Smith, Kim Petras	2	2022	9	22
88.	Hey Mor	Ciara, Feid	2	2022	10	6
89.	Anti-Hero	Taylor Swift	1	2022	10	21
90.	Karma	Taylor Swift	1	2022	10	21
95.	Angels Like You	Miley Cyrus	1	2020	11	2
96.	People	Libianca	1	2022	12	2

Oefenvraag Numeric Facet

Hoe ziet de verdeling van released year eruit?

antwoord op blz.34

2.5 Cluster & Edit

In datasets komen vaak varianten van dezelfde waarde voor, bijvoorbeeld door spelfouten of verschillende schrijfwijzen. **Cluster & Edit** helpt om deze varianten te herkennen en samen te voegen, zodat je data consistent wordt.

Voorbeeld in Spotify-data:

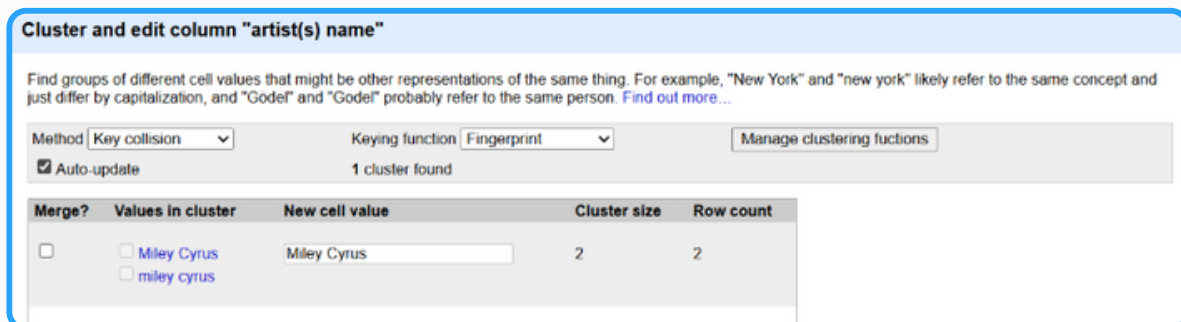
Als een artiest per ongeluk op twee manieren is geschreven, worden de streams verdeeld en lijkt het alsof die artiest minder populair is dan hij eigenlijk is.

Stappen:

1. Maak eerst een Text Facet (bijvoorbeeld op **artists_name**).
2. Klik in het facet-venster op **Edit → Cluster**.



3. OpenRefine toont varianten die op elkaar lijken (bijvoorbeeld “miley cyrus” en “Miley Cyrus”).



4. Kies een **Method** en een **Keying function** om te bepalen hoe de vergelijkingen worden gemaakt:

Method:

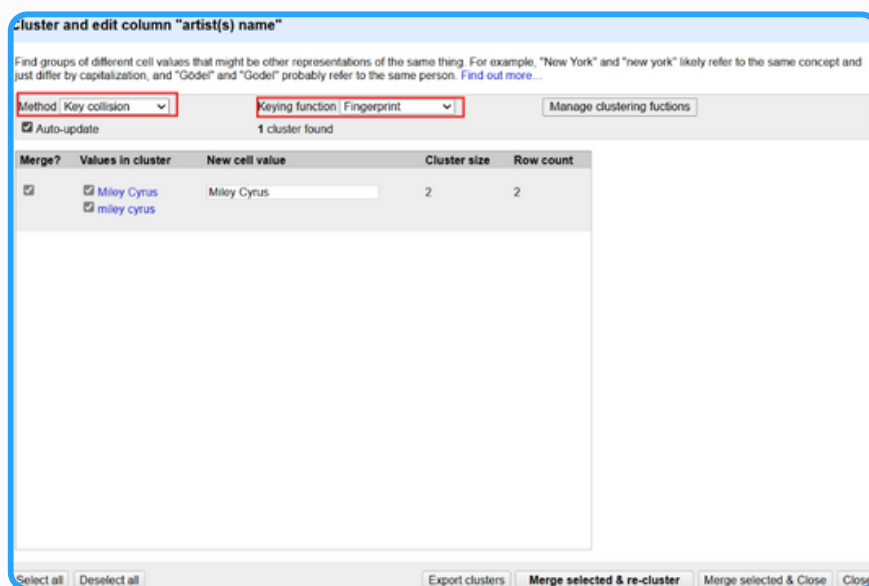
- **Key collision:** zet waarden om naar een eenvoudige vorm (bijvoorbeeld alles naar kleine letters, leestekens weg). Als de vereenvoudigde waarden gelijk zijn, ziet OpenRefine ze als hetzelfde.
- Voorbeeld: “Beyoncé” en “beyonce” worden samengevoegd.

- **Nearest neighbor:** vergelijkt woorden op basis van gelijkenis (zoals spellingscontrole).
- Voorbeeld: “Weekend” en “Weeknd” kunnen worden herkend als varianten.

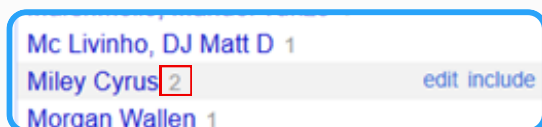
Keying functions (meest gebruikt):

- **Fingerprint:** zet alles om naar kleine letters, verwijdert leestekens en sorteert woorden alfabetisch
- Voorbeeld: “The Weeknd” en “weeknd the” → worden hetzelfde.
- **N-gram fingerprint:** splitst woorden in korte stukjes en vergelijkt die, zodat spelfouten ook herkend worden.
- Voorbeeld: “Rihanna” en “Rihana”.
- **Metaphone3:** kijkt naar de klank van woorden. Handig voor namen.
- Voorbeeld: “Katherine” en “Catherine”.

5. Klik op Merge Selected & Re-Cluster.



6. Beide varianten zijn nu samengevoegd tot één waarde.



Oefenvraag Cluster & Edit

Welke **methode** en **keying function** zou je in deze gevallen gebruiken?

1. Je vindt “DJ Khaled” en “D. J. Khaled” in dezelfde kolom.
2. Twee waarden verschillen alleen door een spelfout: “Post Malone” en “Post Malon”

antwoord op blz.34

2.6 Transform

Met Transform kun je alle waarden in een kolom automatisch aanpassen door middel van formules (een aantal GREL formules die veel worden gebruikt worden toegelicht in het hoofdstuk Handige GREL formules in OpenRefine). Dit is handig om inconsistenties in hoofdletters, spaties of notaties in één keer op te lossen.

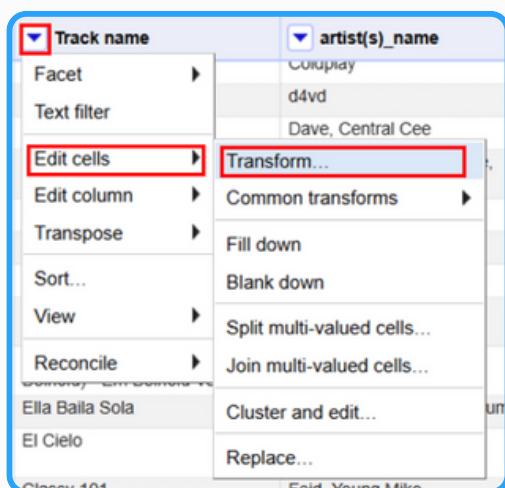
Waar **Common transforms** alleen een paar standaardopties aanbieden (zoals to lowercase, to uppercase, to titlecase en trim whitespace), biedt **Transform** veel meer flexibiliteit. Je kunt er zelf formules intypen en combineren, waardoor je ook complexere bewerkingen kunt doen, zoals tekst vervangen of kolommen splitsen.

Voorbeeld in Spotify-data:

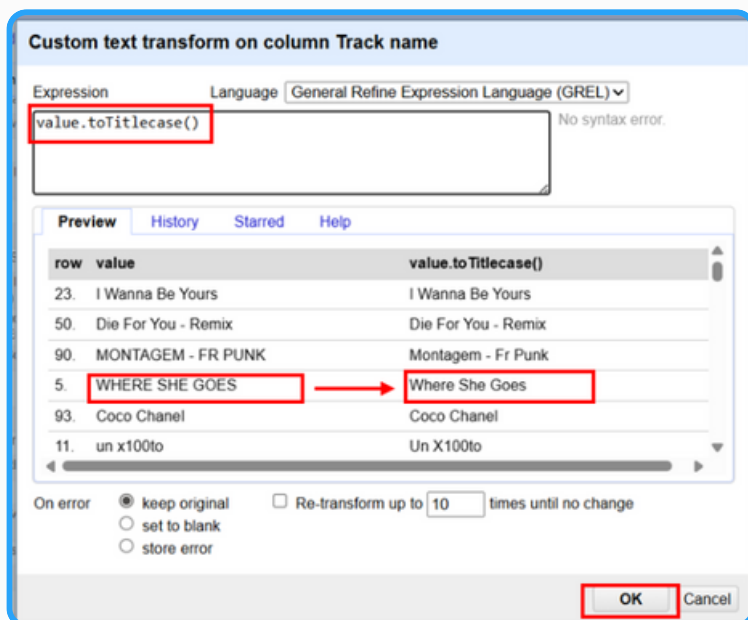
Sommige songtitels beginnen in kleine letters, andere in hoofdletters. Met een Transform kun je dit standaardiseren, zodat je een nettere dataset hebt.

Stappen:

1. Klik op het pijltje bij de kolom **track_name**.
2. Kies **Edit cells** → **Transform**.



3. Typ de formule: `value.toTitlecase()`
4. Klik op **OK**. Nu beginnen alle songtitels met een hoofdletter.



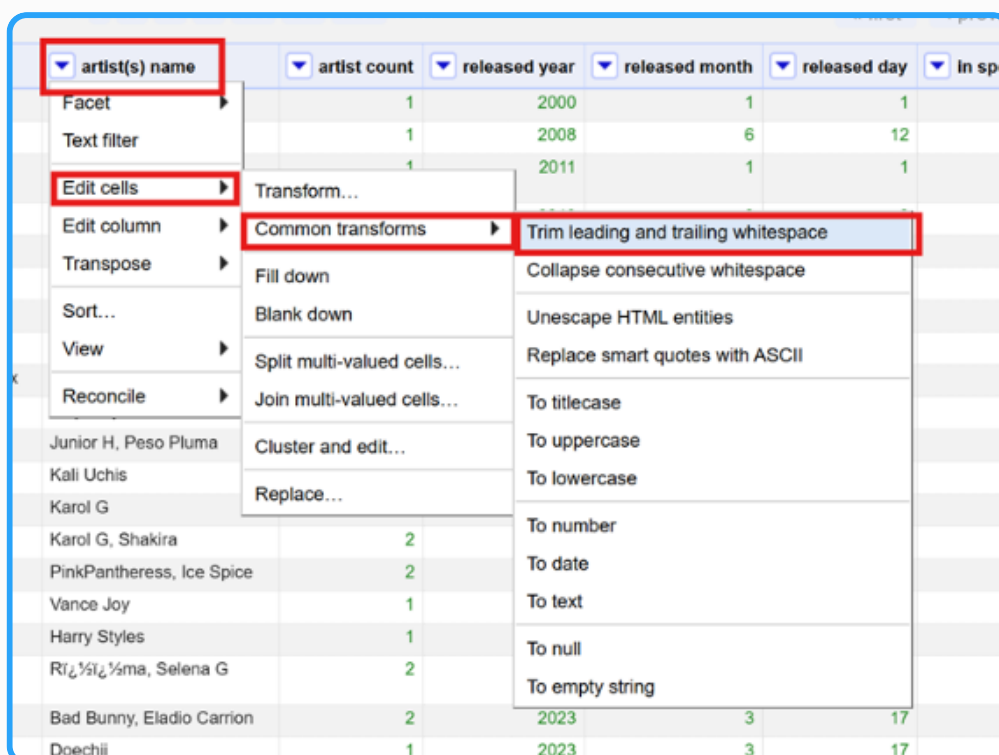
2.7 Common Transforms

Common transforms zijn vooraf ingestelde, veelgebruikte bewerkingen die met een klik op een hele kolom toegepast worden. Het zijn eigenlijk handige snelkoppelingen voor standaard transformaties en komen altijd met een preview en de mogelijkheid om te undo'en.

Doel	Formule	Uitleg
Alles naar kleine letters	<code>value.toLowerCase()</code>	"Taylor Swift" → "taylor swift"
Alles naar hoofdletters	<code>value.toUpperCase()</code>	"taylor swift" → "TAYLOR SWIFT"
Eerste letter elk woord hoofdletter	<code>value.titlecase()</code>	"the weeknd" → "The Weeknd"
Begin- en eindspaties weghalen	<code>value.trim()</code>	" Drake " → "Drake"
Tekst vervangen	<code>value.replace("Pop","POP")</code>	vervangt "Pop" door "POP"
Splitsen	<code>value.split(",")</code>	"Pop, R&B" → ["Pop", "R&B"]

Stappen (Trim whitespace):

1. Klik op de kolomkop **Artist name** → **Edit cells** → **Common transforms** → **Trim leading and trailing whitespace**. Nu zijn alle onnodige spaties verwijderd uit deze kolom.



Stappen (Title case):

1. Kolom **Artist Name** → **Edit cells** → **Common transforms** → **To titlecase**.
Nu beginnen alle namen met een hoofdletter.

artist(s) name	artist count	released year	released month	released day
Facet	1	2000	1	1
Text filter	1	2008	6	12
	1	2011	1	1
Junior H, Peso Pluma				
Kali Uchis				
Karol G				
Karol G, Shakira	2			
PinkPantheress, Ice Spice	2			
Vance Joy	1			
Harry Styles	1			
Ritzy, Selena G	2			
Bad Bunny, Eladio Carrion	2	2023	3	17

Edit cells

Common transforms

To titlecase

Oefenvraag Common Transforms

Je hebt in je dataset de kolom Artist met waarden zoals "taylor SWIFT " en "drake", met welke twee common transforms kun je dat het beste oplossen?

antwoord op blz.34

2.8 Add column based on this column

Met **Add column based on this column** kun je een nieuwe kolom maken op basis van de waarden in één of meerdere bestaande kolommen. Dit is handig als je data wilt combineren of een nieuwe berekende waarde wilt toevoegen.

Waarom deze functie gebruiken?

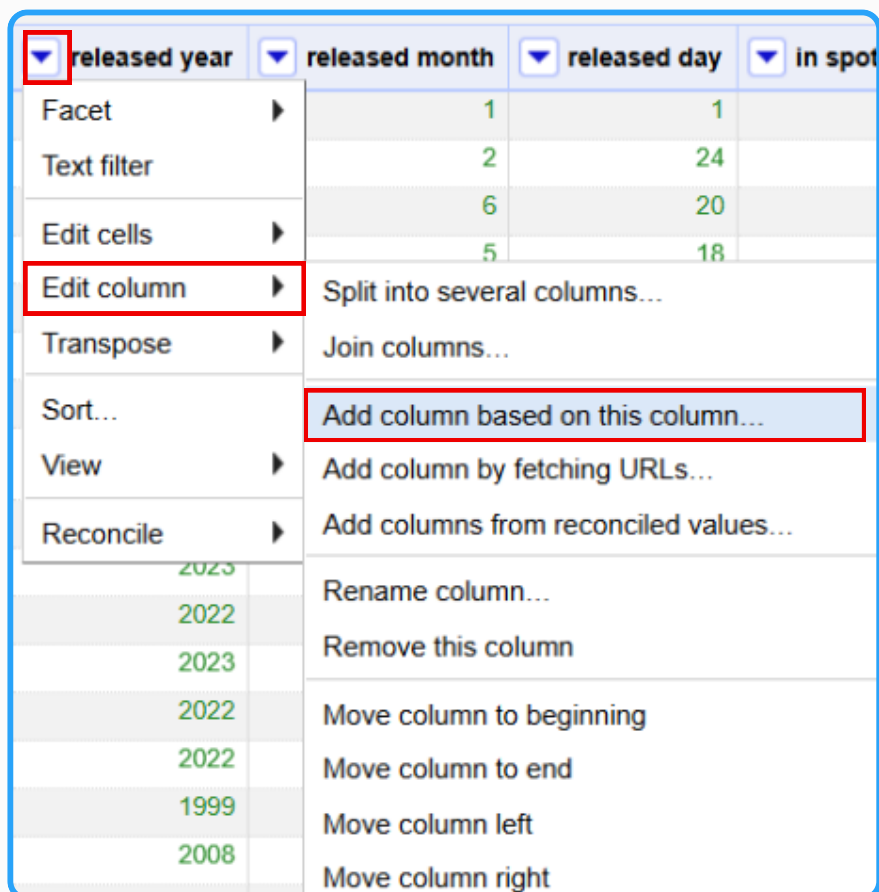
- Je kunt meerdere kolommen samenvoegen tot één nieuwe kolom (bijvoorbeeld jaar, maand en dag combineren tot een datum).
- Je kunt extra categorieën toevoegen (bijvoorbeeld een label “boven 1 miljard streams: ja/nee”).
- Je kunt extra kolommen toevoegen of bestaande kolommen aanpassen op basis van bestaande data, zonder dat de originele data verandert.

Voorbeeld in Spotify-data:

De dataset heeft drie losse kolommen: released year, released month en released day. Met deze functie kun je die combineren tot een datum (release_date).

Stappen:

1. Klik op het pijltje bij de kolom **released year**.
2. Kies **Edit column** → **Add column based on this column...**



3. Geef de nieuwe kolom een naam, bijvoorbeeld `release_date`.

4. Vul in het formuleveld (GREL):

```
cells["released year"].value + "-" + cells["released month"].value + "-" +  
cells["released day"].value
```

Je ziet meteen in het voorbeeldscherm onder het formuleveld hoe de nieuwe waarden in de kolom eruit zullen zien.

Add column based on column released year

New column name:

On error: ☒ set to blank ☐ store error ☐ copy value from original column

Expression: Language: No syntax error.

Preview History Starred Help

row	value	cells["released year"].value + ...
1.	2013	2013-1-1
2.	2023	2023-2-24
3.	2012	2012-6-20
4.	2023	2023-5-18
5.	2023	2023-3-17
6.	2023	2023-4-17

OK Cancel

5. Klik op **OK**. Er wordt een nieuwe kolom toegevoegd met de samengestelde datum.

6. Zet de nieuwe kolom **release_date** om naar een datumtype:

- Klik op het pijltje bij `release_date`.
- Kies **Edit cells** → **Common transforms** → **To date**

Oefenvraag Add column based on this column

Maak met *Add column based on this column* een nieuwe kolom waarin je aangeeft of een nummer meer dan 1 miljard streams heeft of niet. Hoeveel nummers in de Top 100 hebben meer dan 1 miljard streams?

De GREL-formule die je hiervoor nodig hebt is:

```
if(cells["streams"].value > 1000000000, "Ja", "Nee")
```

antwoord op blz.34

2.9 Split into several columns

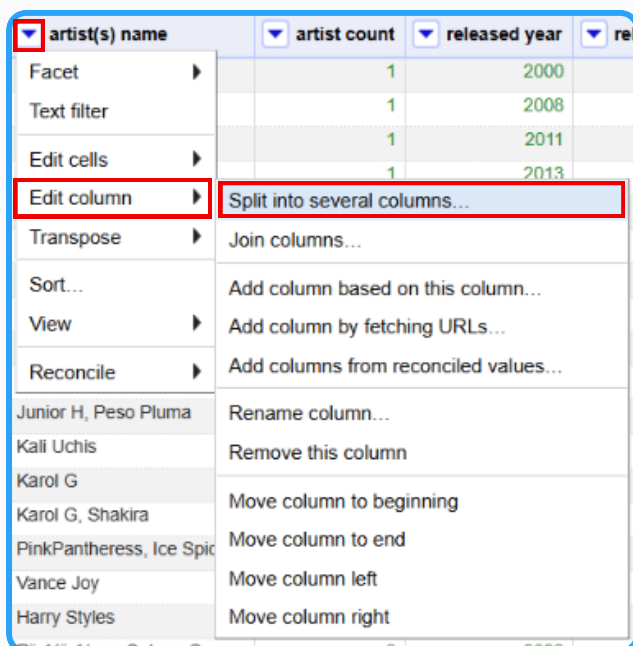
Met de **Split into several columns** kun je de inhoud van cellen opsplitsen naar meerdere kolommen.

Voorbeeld in Spotify-data:

De dataset heeft één kolom met de namen van de artiesten. In een aantal kolommen staan twee of meer artiesten gescheiden met een komma. Met de Split into several columns wordt de inhoud van cellen opgesplitst naar meerdere kolommen.

Stappen:

1. Klik op het pijltje bij de kolom **artist(s)_name**.
2. Kies **Edit column** → **Split into several**



3. Vul in het menu de juiste scheidingsteken in bij **separator** en het aantal kolommen waarin het wordt opgesplitst bij **split into ... columns at most**.

A dialog box titled 'Split column "artist(s)_name" into several columns'. It has two main sections: 'How to split column' and 'After splitting'. In the 'How to split column' section, the 'by separator' radio button is selected. Below it, the 'Separator' field contains a comma (,) and is highlighted with a red box. There is an unchecked checkbox for 'regular expression'. The 'Split into' field contains the number '2' and is also highlighted with a red box. The text 'columns at most (leave blank for no limit)' follows. The 'by field lengths' radio button is unselected. Below it is a text area for a list of integers, with the example text 'List of integers separated by commas, e.g. 5, 7, 15'. In the 'After splitting' section, there are two checked checkboxes: 'Guess cell type' and 'Remove this column'. At the bottom right are 'OK' and 'Cancel' buttons.

4. Klik op **OK**, nu is de inhoud van de cellen met meerdere artiesten in meerdere kolommen gesplitst.

All	Track name	artist(s) name 1	artist(s) name 2	artist(s) name 3	artist count
1.	Yellow	Chris Molitor			1
2.	Viva La Vida	Coldplay			1
3.	Summertime Sadness	Lana Del Rey			1
4.	I Wanna Be Yours	Arctic Monkeys			1
5.	Blank Space	Taylor Swift			1
6.	Style	Taylor Swift			1
7.	Flowers	miley cyrus			1
8.	Last Night	Morgan Wallen			1
9.	Die For You - Remix	Ariana Grande	The Weeknd		2
10.	Cupid - Twin Ver.	Fifty Fifty			1
11.	El Azul	Junior H	Peso Pluma		2
12.	Moonlight	Kali Uchis			1
13.	AMARGURA	Karol G			1
14.	TQG	Karol G	Shakira		2
15.	Boy's a liar Pt. 2	PinkPantheress	Ice Spice		2
16.	Riptide	Vance Joy			1
17.	As It Was	Harry Styles			1
18.	Calm Down (with Selena Gomez)	RiĲ/šĲ/šma	Selena G		2
19.	Coco Chanel	Bad Bunny	Eladio Carrion		2
20.	What It Is (Solo Version)	Doechii			1
21.	Ella Baila Sola	Eslabon Armado	Peso Pluma		2
22.	Classy 101	Feid	Young Miko		2
23.	Like Crazy	Jimin			1
24.	Say Yes To Heaven	Lana Del Rey			1

Oefenvraag Split into several columns

Wat gebeurt er als je split into... columns open laat?

antwoord op blz.34

2.10 Timeline Facet

Met **Timeline Facet** kun je data die een datum bevat op een tijdlijn bekijken en filteren. Zo zie je in één oogopslag hoe je data verdeeld is over de tijd.

Belangrijk vooraf:

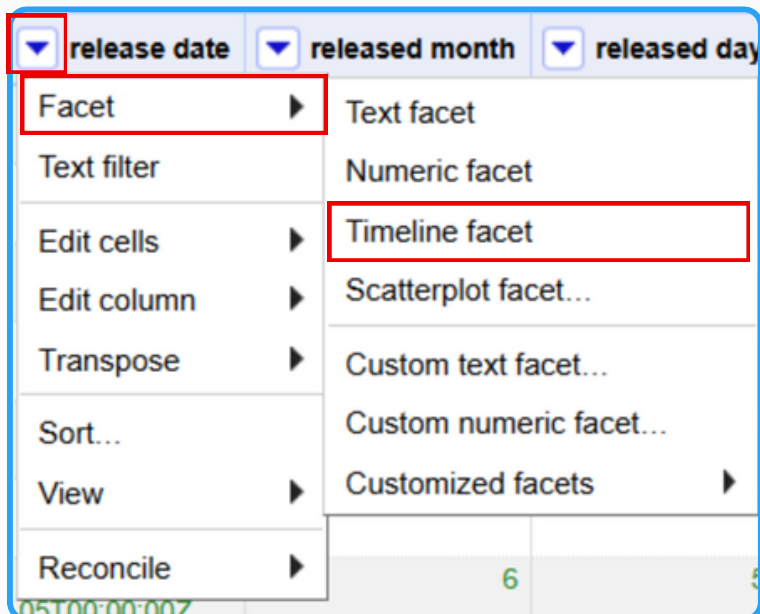
De Timeline Facet werkt alleen als de kolom echt als datum is gegeven. In de Spotify-dataset zijn jaar, maand en dag eerst samengevoegd in een nieuwe kolom *release_date* (via *Add column based on this column*). En daarna zijn de cellen omgezet naar een datumformat (via *Common transforms*).

Voorbeeld in Spotify-data:

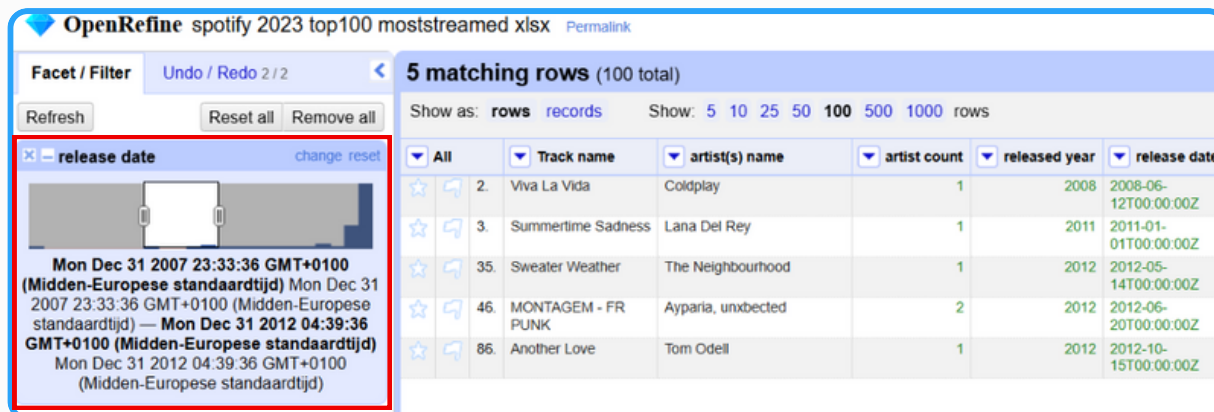
Door een nieuwe kolom *release_date* te maken van jaar, maand en dag, kun je met een Timeline Facet analyseren hoeveel nummers in de Top 100 in 2023 zelf zijn uitgebracht.

Stappen:

1. Klik op het pijltje bij de kolom **release_date**.
2. Kies **Facet** → **Timeline facet**.



3. Er verschijnt een tijdlijn aan de linkerkant met een schuifbalk. Met deze schuifbalk kun je in- en uitzoomen op specifieke jaren of maanden. Zodra je de schuif verschuift, filtert OpenRefine de dataset automatisch en zie je alleen de rijen die binnen de geselecteerde periode vallen



Oefenvraag Timeline Facet

Gebruik de Timeline Facet en zoom in op de periode 2015–2020. Hoeveel nummers uit de Spotify Most Streamed 2023 Top 100 zijn in deze periode uitgebracht?

antwoord op blz.34

2.11 Facet by blank

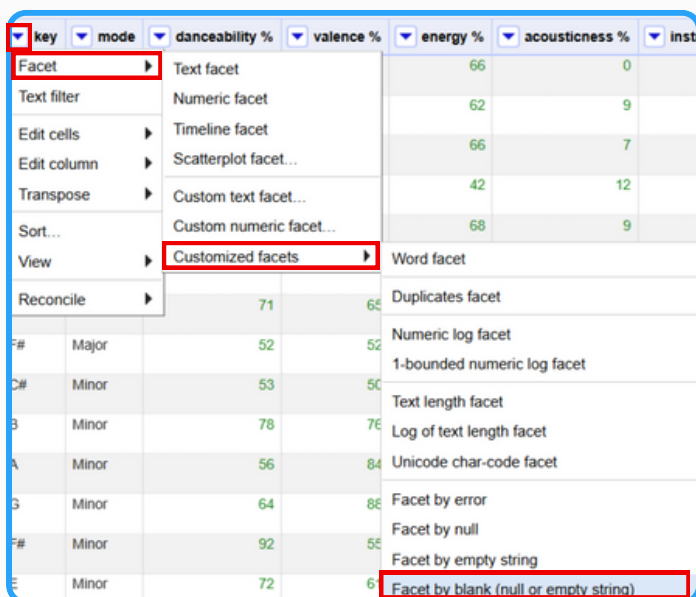
Met **Facet by blank** kun je in één keer zien welke cellen in een kolom leeg zijn. Dit is handig om ontbrekende data op te sporen en te beslissen of je die wilt aanvullen, corrigeren of negeren.

Voorbeeld in Spotify-data:

In deze dataset zijn er een paar kolommen waar niet alle waarden zijn ingevuld, zoals bij *key*, *in_shazam_charts* en *artist_count*. Door Facet by blank te gebruiken, zie je meteen hoeveel rijen in die kolommen leeg zijn en kun je die selecteren.

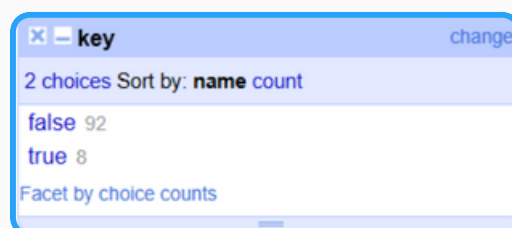
Stappen:

1. Klik op het pijltje bij de kolom die je wilt controleren (bijvoorbeeld *key*).
2. Kies **Facet** → **Customized facets** → **Facet by blank**.

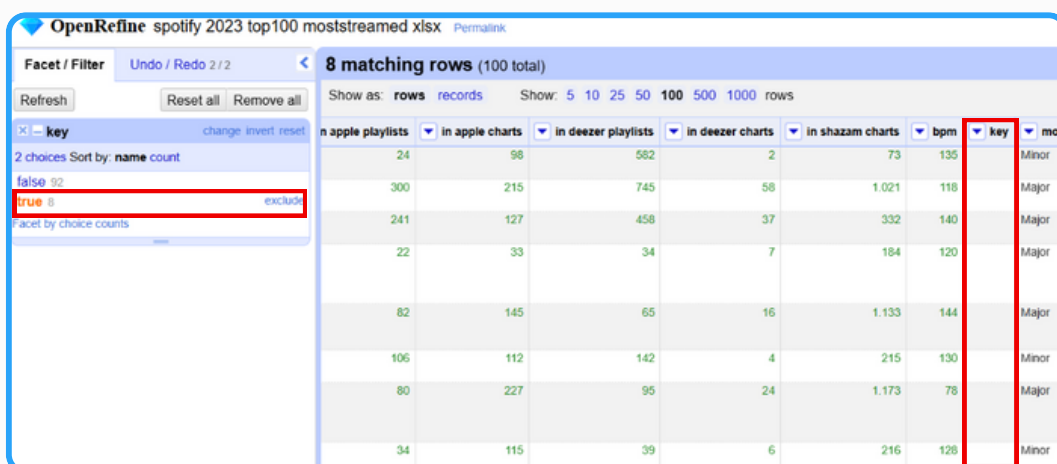


3. Links verschijnt een facet met twee waarden:

- **true** = de cel is leeg.
- **false** = de cel heeft een waarde.



4. Klik op **true** om alleen de rijen met lege cellen te bekijken.



Oefenvraag Facet by blank

Gebruik Facet by blank op de kolom *in_shazam_charts*. Hoeveel rijen in de Spotify *Most Streamed 2023 Top 100* hebben hier geen waarde ingevuld?

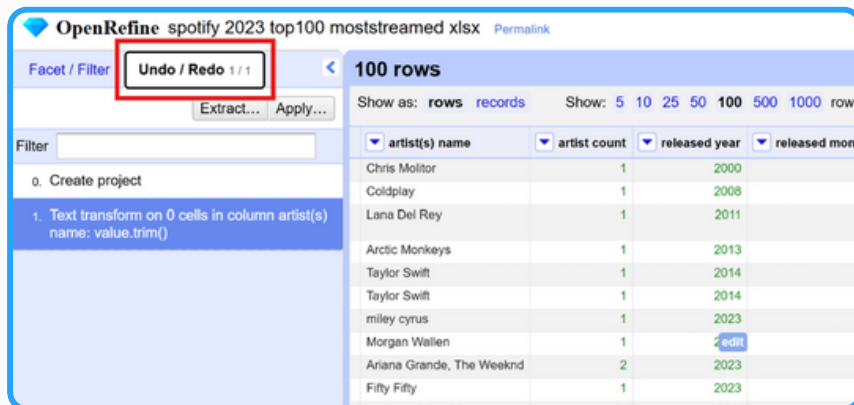
antwoord op blz.34

2.12 History & Undo/Redo

Met de History functie van OpenRefine kan elke stap die is uitgevoerd terugzien worden, terugdraaien (undo) of opnieuw uitvoeren (redo).

Waar vind je het?

- In het linkerpaneel van je project staat een tabblad Undo / Redo.
- Hier zie je een lijst van alle bewerkingen die je gedaan hebt, vanaf het moment dat je je dataset hebt geopend.
- Elke regel is één stap, zoals “Text transform on Artist” of “Common transform: trim whitespace”.



Hoe werkt het?

1. Automatische opslaan

- Elke bewerking wordt automatisch opgeslagen. Je hoeft dit dus niet zelf bij te houden.

2. Undo (teruggaan)

- Klik op een eerdere stap in de lijst.
- Alles wat daarna gebeurde wordt tijdelijk ongedaan gemaakt.
- Voorbeeld: als je 5 stappen hebt gedaan en je klikt terug op stap 2, dan zie je het resultaat alsof je alleen stap 1 en 2 hebt uitgevoerd.

3. Redo (opnieuw doen)

- Als je terug bent gegaan naar een eerdere stap, kun je ook weer vooruit klikken naar de nieuwere stappen.
- Zo kun je gemakkelijk vergelijken of een bewerking nuttig was of niet.

4. Permanent schrappen

- Als je teruggaat naar een eerdere stap en vervolgens een nieuwe bewerking uitvoert, worden de stappen daarna overschreven.
- Dat is geen probleem, maar belangrijk om te weten: je kunt dus niet “twee paden” tegelijk bewaren in één project.

Oefenvraag History & Undo/Redo

Je hebt per ongeluk een transform uitgevoerd waardoor een hele kolom leeg is geworden. Hoe gebruik je de Undo/Redo functie om terug te gaan naar de staat vóór deze fout?

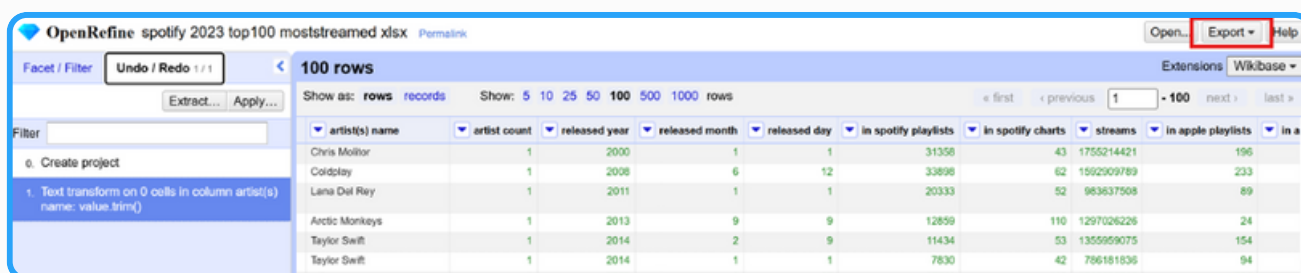
antwoord op blz.34

2.13 Exporteren

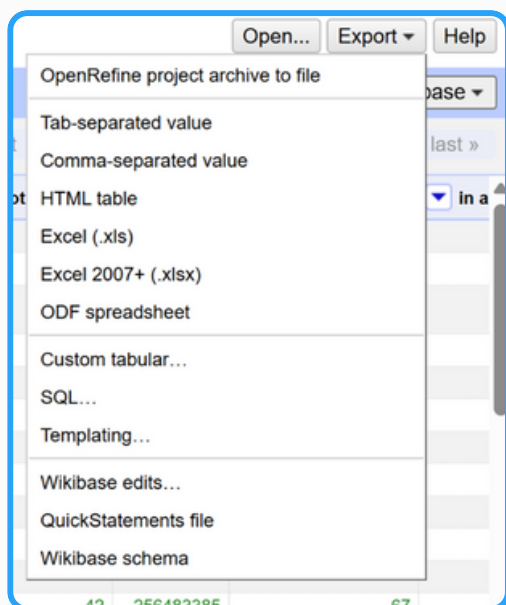
Met de Exportfunctie van OpenRefine kan de opgeschoonde dataset opgeslagen worden in verschillende formaten.

Stappen:

1. Rechtsboven in je project staat de knop Export.



2. Klik hierop en je ziet een menu met verschillende exportopties.



Welke exportmogelijkheden zijn er?

1. Bestandsformaten

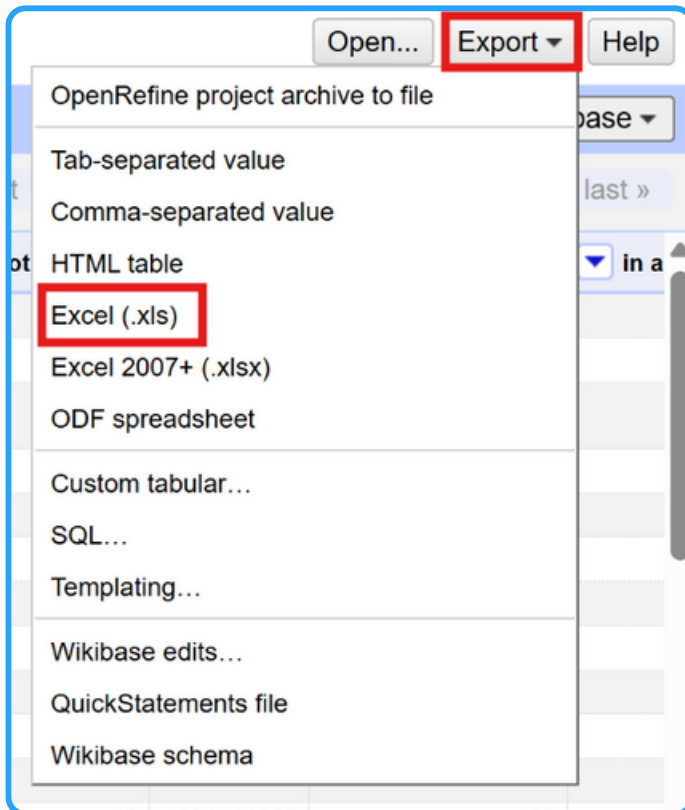
- **Excel (.xls of .xlsx):** handig voor gebruik in Excel of Google Sheets.
- **CSV (comma separated values):** een standaardformaat voor data-analyse en uitwisseling.
- **TSV (tab-separated values):** vergelijkbaar met CSV, maar met tabs als scheidingstekens.
- **HTML table:** je dataset als een eenvoudige tabel in HTML.

2. Project export

- Je kunt ook het **hele project** exporteren (inclusief je geschiedenis en bewerkingen).
- Dit levert een .tar.gz bestand op dat je later weer in OpenRefine kunt importeren.

Stappen:

1. Klik rechtsboven op **Export**.
2. Kies **Excel (.xlsx)**.
3. OpenRefine download automatisch een bestand naar je computer.
4. Open het bestand in Excel - je ziet de opgeschoonde dataset netjes terug.



Oefenvraag Exporteren

Stel, je wilt je opgeschoonde dataset delen met iemand die alleen met Excel werkt. Welke exportoptie kies je?

antwoord op blz.34

3 - Verder leren

In dit hoofdstuk wordt gekeken naar mogelijkheden om nog verder te leren na de basis uit de uitleg, zo worden de GREL formules kort behandeld en wordt er verwezen naar externe bronnen waar je zelf verder kunt experimenteren.

3.1 GREL-formules

GREL staat voor **General Refine Expression Language**. Het is de “taal” die je in OpenRefine gebruikt bij functies als *Transform* of *Add column based on this column*. Met GREL kun je data **bewerken, corrigeren of combineren**. Denk aan tekst netjes maken (hoofdletters/spaties), kolommen samenvoegen of een nieuwe waarde berekenen. Een groot voordeel van GREL is dat je **zonder te programmeren** kleine formules kunt toepassen, terwijl je in het voorbeeldscherm meteen ziet wat het resultaat wordt.

Wanneer gebruik je GREL?

je gebruikt GREL zodra je:

- Waarden in een kolom wilt aanpassen (Transform).
- Een nieuwe kolom wil maken op basis van bestaande kolommen (Add column based on this column).
- Patronen of foutjes in tekst wilt opschonen (bijv. dubbele spaties, verschillende schrijfwijzen).
- Tekst wil omzetten naar getallen of datums, zodat je er verder mee kunt rekenen of filteren.

Met andere woorden: **GREL gebruik je als je handmatig aanpassen te veel werk is en je data in één keer wilt standaardiseren of structureren.**

3.1.1 Veel gebruikte GREL functies

Doel	Formule	Uitleg	Voorbeeld
Tekst naar kleine letters	<code>value.toLowerCase()</code>	Maakt tekst uniform voor vergelijkingen, handig bij namen of categorieën.	"Taylor Swift" → "taylor swift"
Tekst naar hoofdletters	<code>value.toUpperCase()</code>	Voor export of labels die overal in hoofdletters moeten staan.	"Taylor Swift" → "TAYLOR SWIFT"
Eerste letter per woord hoofdletter (Titelcase)	<code>value.toTitlecase()</code>	Maakt namen of titels netjes leesbaar.	"the weeknd" → "The Weeknd"

Doel	Formule	Uitleg	Voorbeeld
Spaties weghalen	<code>value.trim()</code>	Haalt begin- en eindspaties weg die tot dubbele waarden leiden.	" Drake " → "Drake"
Tekst vervangen	<code>value.replace("Pop", "POP")</code>	Corrigeert waarden; met regex ook patronen vervangen.	"Pop" → "POP"
Tekst splitsen	<code>value.split(",")</code>	Zet meerdere waarden uit één cel in een array (later opsplitsen naar kolommen).	"Pop, R&B" → ["Pop", "R&B"]
Waarden samenvoegen	<code>cells["Artiest"].value + " - " + cells["Titel"].value</code>	Maak nieuwe velden zoals "Artiest – Titel".	"Adele" + "Hello" → "Adele - Hello"
Voorwaarden toevoegen	<code>if(value > 1000000, "Hit", "Geen Hit")</code>	Label data op basis van een regel.	2000000 → "Hit", 50000 → "Geen Hit"
Tekst naar getal	<code>toNumber(value)</code>	Nodig als getallen als tekst zijn geïmporteerd.	"1000" → 1000
Lengte berekenen	<code>value.length()</code>	Spot lege of afwijkend lange waarden.	"Hello" → 5
Tekst naar datum	<code>toDate(value)</code>	Zet tekst om naar een datumtype (vereist voor Timeline Facet).	"2023-10-02" → [date object]
Datumonderdeel selecteren	<code>value.datePart("years")</code>	Haal jaartal, maand of dag uit een datum.	"2023-10-02" → 2023
Regex zoeken	<code>value.contains(/\s+/)</code>	Zoek patronen, bijv. meerdere spaties.	"Bad Bunny" → true
Regex extract	<code>value.match(/.*(\d{4}).*/)</code>	Haal specifieke stukken uit tekst (bijv. een jaartal).	"Album (2020)" → ["2020"]
Eerste niet-lege waarde	<code>coalesce(value, "Onbekend")</code>	Zet lege cellen om naar een fallback.	null → "Onbekend"

3.2 Verder leren met externe bronnen

Naast de basisfuncties zijn er nog veel meer geavanceerde toepassingen waar je je zelf in kunt verdiepen. Hiervoor zijn veel verschillende websites waar de mogelijkheid is om deze toepassingen te verkennen.

Neem eens een kijkje op onderstaande websites;

- **OpenRefine Wiki (Github):** <https://github.com/OpenRefine/OpenRefine/wiki>
- **Officiële website OpenRefine:** hier vind je uitgebreide handleidingen.
<https://openrefine.org/documentation.html>
- **Librabry Carpentry tutorials:** [Library Carpentry: OpenRefine: Summary and Setup](#)

4 - Bijlagen

4.1 Cheatsheet

Deze cheat sheet biedt een handig overzicht van de belangrijkste functies, GREL-expressies en tips om efficiënt met OpenRefine te werken.

4.1.1 De Basis

Project	Beschrijving
Facet	Groeperen en filteren van data op basis van bepaalde waarden om snel patronen en afwijkingen te ontdekken
Cluster	automatisch vergelijkbare of foutief gespelde waarden detecteren en samenvoegen tot één consistente waarde.
Transform	Bewerken van celwaarden met behulp van GREL-expressies.

4.1.2 Facets

Facet	Beschrijving
Text Facet	Groep en filter op unieke waarden
Numeric Facet	Filter via schuifbalk op numerieke waarden
Timeline Facet	Filteren op datum/tijd
Custom Facet	Maak eigen filters met GREL-expressies
Scatterplot Facet	Ontdek relatie tussen twee numerieke kolommen

4.1.3 Clusters

Cluster	Beschrijving
Key Collision	Groepeert waarden die bijna identiek zijn.
Fingerprint	Herkent varianten door spaties, leestekens en volgorde van woorden te negeren.
N-Gram Fingerprint	Vergelijkt waarden op basis van overlappende lettergroepjes
Nearest Neighbour	Fuzzy matching (vergelijkbare waarden)

4.1.4 GREL (General Refine Expression Language)

GREL-Expressie	Voorbeeld	Beschrijving
Tekstfuncties		
toUppercase	toUppercase(value)	Zet tekst om naar hoofdletters
toLowercase	toLowercase(value)	Zet tekst om naar kleine letters
trim	trim(value)	Verwijder spaties
replace	replace(value,"oud","nieuw")	Vervang tekst
split	split(value, ",")	Splits string op delimiter
join	join(["a","b"], ";")	Voeg lijst samen
length	length(value)	lengte van tekst
Rekenfuncties		
toNumber	toNumber(value)	Zet om naar getal
toDate	toDate(value)	Zet om naar datum
round	round(value)	Afronden
floor	floor(value)	Naar beneden afronden
ceil	ceil(value)	Naar boven afronden

4.1.5 Bewerken van gegevens

Methode	Beschrijving
Edit cells → Transform	Bewerken van celwaarden met behulp van GREL-expressies.
Edit column → Add column	Maak nieuwe kolom gebaseerd op bestaande
Edit column → Split	Splits waarden in meerdere kolommen
Edit rows → Remove rows	Verwijder ongewenste rijen

4.1.6 Tips

Methode	Beschrijving
Undo/Redo	Draai bewerkingen terug
History Exporteren	Herhaal bewerkingen op nieuwe dataset
Star/Flag	Markeren van records



4.2 Antwoorden oefenvragen

Hoeveel nummers uit de Spotify Most Streamed 2023 Top 100 zijn uitgebracht in 2022?

- Er zijn **20** nummers in de lijst die in 2022 zijn uitgebracht.

Hoe ziet de verdeling van released year eruit?

- De meeste nummers zijn tussen 2022-2023 gereleased die in de top 100 staan. Voor die tijd niet zoveel.

Welke *methode* en *keying function* zou je in deze gevallen gebruiken?

Je vindt “DJ Khaled” en “D. J. Khaled” in dezelfde kolom.

- *Key collision* met *Fingerprint* → omdat leestekens (puntjes) worden verwijderd en beide vormen dezelfde sleutel krijgen.

Twee waarden verschillen alleen door een spelfout: “Post Malone” en “Post Malon”

- *Nearest neighbor* met *N-gram fingerprint* → herkent dat de woorden bijna hetzelfde zijn en clustert ze.

Je hebt in je dataset de kolom Artist met waarden zoals "taylor SWIFT " en "drake", met welke twee common transforms kun je dat het beste oplossen?

- Trim leading and trailing whitespace en To titlecase

Maak met Add column based on this column een nieuwe kolom waarin je aangeeft of een nummer meer dan 1 miljard streams heeft of niet. Hoeveel nummers in de Top 100 hebben meer dan 1 miljard streams? De GREL-formule die je hiervoor nodig hebt is: `if(cells["streams"].value > 1000000000, "Ja", "Nee")`

- Er zijn **9** nummers in de Spotify Most Streamed 2023 Top 100 die meer dan 1 miljard streams hebben.

Wat gebeurt er als je split into... columns open laat?

- De kolom wordt in acht kolommen gesplitst

Gebruik de Timeline Facet en zoom in op de periode 2015–2020. Hoeveel nummers uit de Spotify Most Streamed 2023 Top 100 zijn in deze periode uitgebracht?

- In de periode 2015–2020 zijn er **14** nummers uitgebracht. Het grootste deel van de lijst bestaat juist uit recentere releases (2021–2023), wat laat zien dat de meest gestreamde muziek vooral nieuw is.

Gebruik Facet by blank op de kolom in_shazam_charts. Hoeveel rijen in de Spotify Most Streamed 2023 Top 100 hebben hier geen waarde ingevuld?

- Het resultaat laat zien dat er **6** rijen leeg zijn (true), en 94 rijen wel een waarde bevatten (false).

Je hebt per ongeluk een transform uitgevoerd waardoor een hele kolom leeg is geworden. Hoe gebruik je de Undo/Redo functie om terug te gaan naar de staat vóór deze fout?

- In het linkerpaneel klik je in de Undo/Redo geschiedenis op de stap vóór de fout, je dataset springt terug naar de vorige versie.

Stel, je wilt je opgeschoonde dataset delen met iemand die alleen met Excel werkt. Welke exportoptie kies je?

- Je kiest voor Export → Excel (.xlsx).

4.3 Literatuurlijst

- OpenRefine. (z.d.). Geraadpleegd op 16 september 2025, van <https://openrefine.org/>