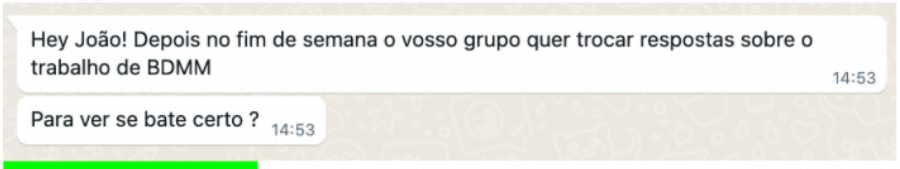
1. **Before meeting**
   a. Gaspar suggestion
      i.



      ii. Todos concordam --> Eu digo-lhe respostas 3, 4, 5

2. **During meeting**
   a. <u>Data Exploration</u>
      i. To do
         1) Confirmar valores das tabelas - (Umi)
            a)

Node	Property	Count	Distinct Count	Mean	Min	Max	Missing Count	Missing Count (%)
BEERS	abv	372718	939	6.53	0.01	100	45028	12.08
	availability	417746	20	-	-	-	0	0
	brewery_id	417746	16569	24 592.84	1	54 144	0	0
	id	417746	358873	189 196.88	3	374 406	0	0
	name	417746	298567	-	-	-	0	0
	notes	417746	48313	-	-	-	55	0.01
	retired	417746	2	-	-	-	0	0
	state	417746	68	-	-	-	70831	16.96
BREWERIES	id	100694	50347	27 870.51	1	54 156	0	0
	name	100694	45245	-	-	-	0	0
	notes	100694	3271	-	-	-	170	0.17
	state	100694	68	-	-	-	22542	22.39
	types	100694	30	-	-	-	0	0
CITIES	name	23330	11665	-	-	-	2	0.01
COUNTRIES	name	400	200	-	-	-	2	0.50
REVIEWS	beer_id	2549252	189645	77 459.01	3	373 128	0	0
	date	2549252	6379	-	-	-	19	0.00
	feel	1484819	17	3.89	1	5	1064433	71.69
	id	2549252	2546141	4 517 442.86	1	9 073 127	0	0
	look	1484819	17	3.95	1	5	1064433	71.69
	overall	1484819	17	3.92	1	5	1064433	71.69
	score	2549252	401	3.89	1	5	0	0
	smell	1484819	17	3.89	1	5	1064433	71.69
	taste	1484819	17	3.92	1	5	1064433	71.69
	text	2549252	814333	-	-	-	19	0.00
STYLE	name	113	113	-	-	-	1	0.88
USER	name	123935	123935	-	-	-	1	0.00

            b) Confirmar valores com tabela em 2.1

         2) Cofirmar valores (Umi)
            a)

BEERS

- `abv`: According to Hops & Hopes (2025) the world's strongest beer is **Snake Venom** with **67.5% ABV**. We can see that the **maximum** value in the database is **100**, so we will need to verify the beers with more than **67.5% ABV**.
- `availability` have 20 different values. We will check what are the possible values.
- `brewery_id`: In total we have 417 746 brewery_ids, but 16 569 distinct brewery_ids. It is possible that some breweries have more than one beer.
- `id`: In total we have 417 746 ids, but 358 873 distinct ids. We need to check the duplicates.
- `name`: In total we have 417 746 names, but 298 567 distinct names. It is possible that some beers have the same name.
- `notes`: In total we have 417 746 notes, but 48 313 distinct notes. Same as the `name` property, it is possible that some beers have the same notes.
- `retired` have 2 different values - `True` and `False`.
- `state` have 68 different values. We will check what are the possible values.

BREWERIES

- `id`: In total we have 100 694 ids, but 50 347 distinct ids. As in `BEERS`, it is possible that some breweries have more than one beer.
- `name`: In total we have 100 694 names, but 45 245 distinct names. We need to check the duplicates.
- `notes`: In total we have 100 694 notes, but 3 271 distinct notes. Same as the `name` property, it is possible that some breweries have the same notes.
- `state` have 68 different values (the same as in `BEERS`). We will check what are the possible values.
- `types` have 30 different values. We will check what are the possible values.

CITIES

- `name`: In total we have 23 330 names, but 11 665 distinct names. We need to check the duplicates.

COUNTRIES

- `name`: In total we have 400 names, but 200 distinct names. We need to check the duplicates.

REVIEWS

- `beer_id`: In total we have 2 549 252 beer_ids, but 189 645 distinct beer_ids. It is possible that some beers have more than one review.
- `date`: We need to convert this property to `date` format.
- `feel`, `look`, `overall`, `smell`, `taste` and `score`: We need to convert these properties to `float` format, but we can see that the values are already clean, because the `Min` and `Max` values are between 1 an 5.
- `id`: In total we have 2 549 252 ids, but 2 546 141 distinct ids. We need to check the duplicates.
- `text`: In total we have 2 549 252 texts, but 814 333 distinct texts. It is possible that some reviews have the same text.

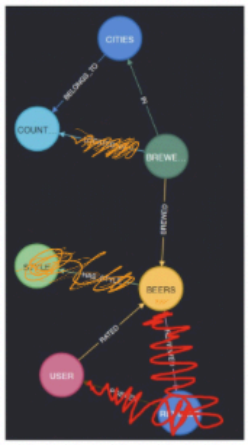
STYLE

- `name`: In total we have 113 names and distinct names, so we don't need to check the duplicates.

b. Database schema

i. Decisão final

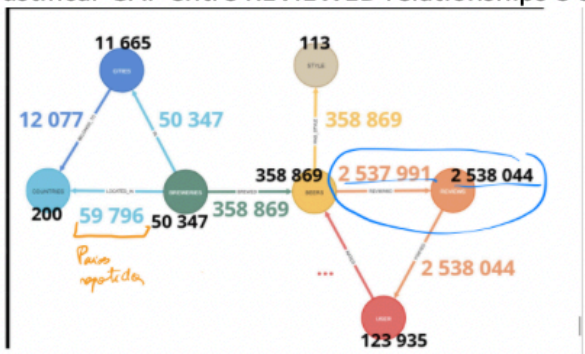
1)



ii. To do list

1) Alterar nome nodes: Não mexer com código. Apenas sugestão. Colocar tudo no Singular
2) Justificar
   a) Remover Style, Has_Style - Done
   b) Remoção de REVIEWS, POSTED + REVIEWED - Done
   c) Não implementação de ligação BREWERIES to COuntries - Done
   d) Justificar GAP entre REVIEWED relationships e o REVIEWES - Done

e)



   f) Propery State na BREWERIES e BEERS

i)

```
MATCH (brewery:BREWERIES)-[:BREWED]→(beer:BEERS)
WHERE brewery.state <> beer.state
RETURN brewery.name AS Brewery, brewery.state AS BreweryState,
       beer.name AS Beer, beer.state AS BeerState
LIMIT 10
```

(no changes, no records)

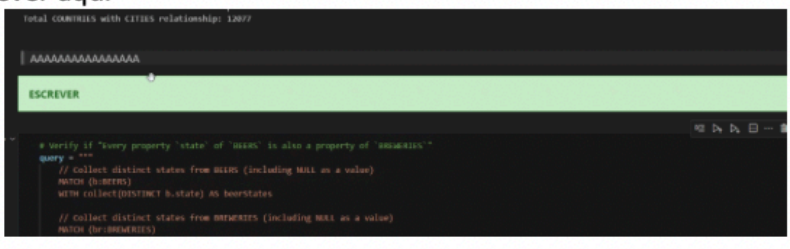
ii) Solução:

Um. Eliminar state de BEERS e manter apenas na de BREWERIES - Código (André)

Dois. Justificação para o porquê de não podermos colocar a property state na relação belongs_to (Filipa)

1. Queriamos colocar a property state na releção Belongs_to the city country mas não podemos porque.....
2. porque..... há cidades associadas a mais do que um país
3. Escrever aqui

1.



   g) Justificar parte da remoção do Style (João)
      i) Fazer na minha versao e nao na minha original

c. Query optimization
   i. Comentários com base nas imagens (Alex)

d. Graph algorithms
   i. 5.1 - comentário (João)
      1) Assumptions
      2) Results
      3)

   ii. 5.2 - comentários (André)

3. Task geral para todos
a. Reler notebook inteiro
b. Criar container novo + Correr tudo
c. Até meeting 16.03 - 18:30