

# RAPPORT PROJET 8

(option C)

*Le stand-up américain au Cirque  
du Soleil*

---



CIRQUE DU SOLEIL®

9 DECEMBRE 2020

---

Parcours Data Analyst OpenClassrooms  
Levallois Matthieu



OPENCCLASSROOMS



---

## Table des matières

Introduction.....	4
1. Diversité des spectacles du Cirque du Soleil.....	6
2. L'humour omniprésent aux USA.....	9
3. Des thèmes grand public.....	14
4. Un public très large .....	16
5. Des interprètes aux multiples facettes .....	17
Conclusion.....	22

---

## Introduction

Le parcours Data Analyst de la formation OpenClassrooms s'articule en plusieurs projets abordant l'analyse de données dans divers domaines professionnels. Or, s'il en est un que j'affectionne particulièrement, c'est celui du spectacle vivant.

En effet, mon projet professionnel est d'intégrer le service de l'analyse de données au Cirque du Soleil afin d'en étudier toutes ses perspectives. Cette firme québécoise opte depuis ses débuts en 1984 pour une stratégie Océan Bleu et réinvente le cirque en mêlant ses traditionnels numéros d'acrobatie et de clowns aux chorégraphies, chants, costumes et musiques de l'opéra.

Repoussant toujours ses limites, le Cirque du Soleil ne cesse alors de proposer de nouvelles thématiques et possède aujourd'hui un catalogue de spectacles très diversifiés, allant du classique *Alegria* au très novateur *Messi10*, inspiré de la vie du célèbre footballeur.

Aussi, propre à cette stratégie Océan Bleu, il me paraît judicieux d'analyser un domaine qui pourrait servir de base de réflexion sur la dimension humoristique des prochaines productions : celui du stand-up américain.

Effectivement, d'une part, les comédiens américains connaissent beaucoup de succès et d'autre part, les spectacles fixes du Cirque du Soleil se situent sur le sol des États-Unis, à Las Vegas, représentant 60 % des revenus de l'entreprise.

L'objectif de cette étude sera donc de répondre à la problématique suivante : est-il possible d'intégrer le stand-up américain dans une production du Cirque du Soleil ?

Pour y répondre, nous nous appuierons sur les données du site <https://scomedy.com/> répertoriant plus de 18 000 citations d'humoristes américains, complétées par leur nombre de vues et leur thème.

L'étude se développe en cinq parties. La première partie présente le Cirque du Soleil. La seconde aborde l'attrait de l'humour pour les américains. La troisième traite de

---

la diversité des thèmes en humour. La quatrième détaille le cœur de cible. Et la dernière partie aborde les représentants du stand-up.

---

## 1. Diversité des spectacles du Cirque du Soleil

Fleuron du Québec, le Cirque du Soleil est créé par Gilles Sainte-Croix et Guy Laliberté. Dès leurs débuts, ils cassent les codes du cirque traditionnel en abandonnant la présence d'animaux dans leurs spectacles et en intégrant davantage de chants et de chorégraphies.

Cette formule gagnante pose les jalons de leur stratégie d'entreprise : celle de l'Océan Bleu<sup>1</sup>. Cette stratégie consiste à se diversifier en empruntant tous les champs des possibles.

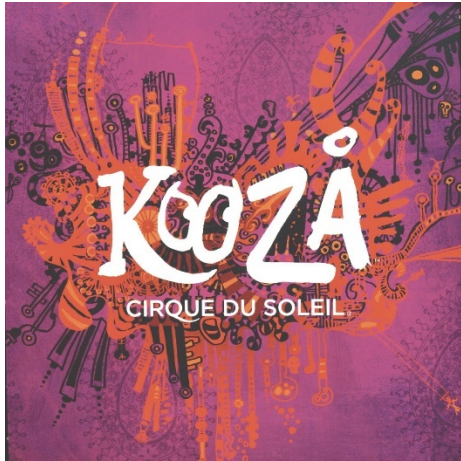
En effet, le Cirque du Soleil a brillé à travers ses spectacles tels que *Corteo*, œuvre onirique inspirée de la commedia dell'arte, *Kooza*, monde burlesque dont l'esthétique rappelle les peintures de Gustav Klimt, ou encore *Alegria*, plus grand succès et désormais classique de l'entreprise.



---

<sup>1</sup> [https://plus.lapresse.ca/screens/cc9b6434-a5c3-47b7-8f56-ca17fd7b984e\\_7C\\_0.html](https://plus.lapresse.ca/screens/cc9b6434-a5c3-47b7-8f56-ca17fd7b984e_7C_0.html)





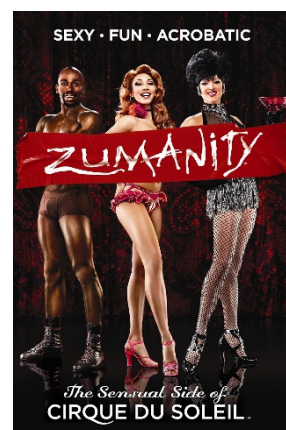
Et si ces spectacles définissent la signature esthétique de l'entreprise, d'autres s'en éloignent considérablement. *Toruk* s'inspire du film *Avatar* de James Cameron.



O puise ses références dans la natation synchronisée.



*Zumanity* exploite les codes du cabaret.



---

*R.U.N.*, écrit par le réalisateur Robert Rodriguez, renvoie aux films de série B.



Enfin, inutile de préciser les références des spectacles *Michael Jackson : The Immortal World Tour* et *The Beatles LOVE*.



Ainsi, le Cirque du Soleil ne se fixe aucune limite dans la création. Et si son œuvre, propre au cirque traditionnel, remplit toujours la fonction d'impressionner par ses acrobaties, elle a également pour objectif de faire rire avec des scènes de clowns.

Dans ces conditions, pourquoi ne pas intégrer l'univers du stand-up américain ? Les spectacles cités prouvent que le Cirque du Soleil ne se fixe aucune limite d'inspiration dans la mise en scène, la danse, l'acrobatie et la musique. La comédie a donc toute sa place en tant que référence pour une prochaine production à Las Vegas.

D'autant plus que les États-Unis regorgent d'humoristes.



## 2. L'humour omniprésent aux USA

À l'image du succès de la chaîne de télévision américaine *Comedy Central*, l'humour tient une place importante aux États-Unis. Le site Internet *SComedy*<sup>2</sup> atteste cette aversion. Ce site répertorie 18873 citations écrites par 291 humoristes exerçant aux États-Unis, en précisant leurs thèmes et leur nombre de vues.

Le site Internet *SComedy* est donc notre point de départ de l'analyse en récupérant ses données par Web scraping.

	citation	auteur	nbre_vues	theme	page_site
0	Well, birthdays are merely symbolic of how ano...	Jerry Seinfeld	1632	[birthday]	1
1	My perfect date night: I pick you up. In my Ki...	Aziz Ansari	1416	[danger, dating, restaurants]	1
2	I mean, they don't grade fathers. But if your ...	Chris Rock	1212	None	1
3	Jack and Jill went up the hill Both with a buc...	Andrew Dice Clay	1158	None	1
4	"Money doesn't buy happiness." Uh, do you live...	Daniel Tosh	1036	[happiness]	1
...	...	...	...	...	...
18868	A couple weeks ago I was on the street and I s...	Demetri Martin	4	None	1258
18869	Whenever I see an autobiography for sale in th...	Demetri Martin	4	None	1258
18870	I saw a door that said exit only. So I entered...	Demetri Martin	4	None	1259
18871	I bought a dictionary. First thing I did was, ...	Demetri Martin	4	None	1259
18872	I was seeing this girl and she wanted to get m...	Demetri Martin	4	None	1259

18873 rows × 5 columns

### *Jeu de données issues du site Web SComedy*

Le jeu de données se compose de 5 colonnes :

- citation : répertoriant toutes les citations du site Internet *SComedy*
- auteur : associant un auteur à chaque citation
- nbre\_vues : indiquant le nombre de vues de chaque citation
- theme : renseignant le ou les thèmes abordés par chaque citation
- page\_site : le numéro de la page du site sur laquelle apparaît la citation

---

<sup>2</sup> <https://scomedy.com/>

Une première vérification détecte 228 doublons de citations, soit seulement 1,2% des données extraites. Le jeu de données est fiable. Parmi ces doublons, 16 citations sont attribuées à deux auteurs différents. Le numéro de page du site Internet correspondant permet de vérifier que les citations sont créditées de la sorte sur le site *SComedy*. D'autres sources Internet permettent de déduire que la paternité de l'œuvre revient toujours à l'auteur totalisant le plus grand nombre de vues. Sont donc supprimés tous les doublons de citations à deux auteurs différents dont le nombre de vues est le plus faible.

### Qu'est-ce que le Web scraping ?

*Le Web scraping est une technique d'extraction du contenu d'un site Web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte.*

Définition [Wikipédia](#)

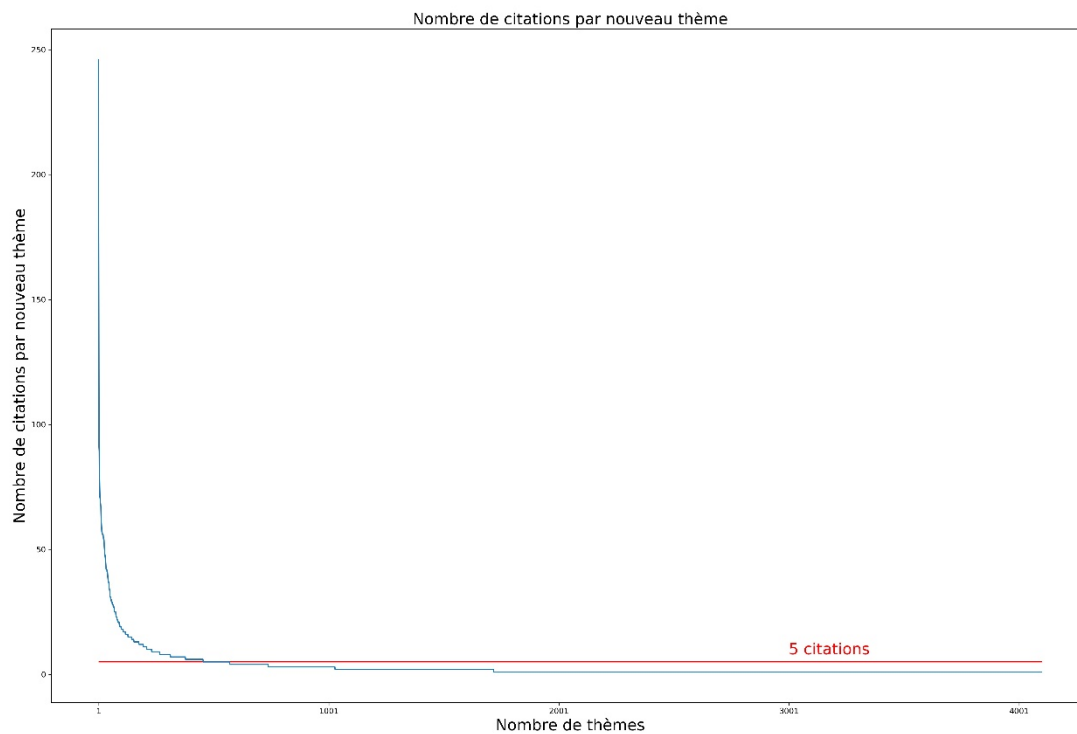
Les autres doublons de citations ont le même auteur mais un nombre de vues différents. De même, une vérification sur le site *SComedy* avec le numéro de page correspondant à chaque citation permet de vérifier que ces citations apparaissent bien deux fois sur le site. Chaque citation sur le site *SComedy* dispose d'un permalink. Un clic sur ce permalink fait augmenter le nombre de vues de la citation. Il faut donc additionner les nombres de vues des doublons de citations ayant le même auteur puisqu'elles sont visibles sur deux pages différentes.

70% des citations du jeu de données n'ont pas de thèmes. Il faut donc leur en attribuer un en effectuant une analyse sémantique de chaque citation. Le jeu de données dispose initialement de 201 thèmes, mais avec l'analyse sémantique, il y en a désormais plus de 4000.

### Qu'est-ce que l'analyse sémantique ?

*L'analyse sémantique consiste à établir une liste de mots utilisés, en précisant le nombre de fois, pour chaque citation. Cette technique est possible dans Python grâce à la librairie NLTK. Cette dernière permet de faire des associations de mots (par exemple en anglais, les termes en 'ing' et les verbes : 'do' avec 'doing') et d'occulter les mots n'ayant pas d'intérêts (les articles, les adverbes) via une liste couramment appelée 'Stopwords'. Une fois l'analyse sémantique effectuée, il ne reste qu'à retenir le mot le plus fréquent pour chaque citation et qui va être son thème.*

En groupant le nombre de citations par nouveau thème, 3645 de ces thèmes rassemblent au mieux 5 citations. D'autre part, le nombre de vues des citations concernées se situe entre 4 et 1172, quand la moyenne de nombre de vues cumulées pour chaque thème est de 2385. Ces thèmes n'ont donc pas de succès et sont des *outliers*.



L'étude se focalise sur les thèmes ayant le plus de succès. Par conséquent, les 3645 thèmes rassemblant au mieux 5 citations sont remplacés par le thème 'Autres'. Le jeu de données dispose finalement de 632 thèmes, et plus de 70 % des citations ont maintenant un thème attribué. Les 30% restant sont classés dans le thème 'Autres'.

Afin d'approfondir l'étude, il convient d'identifier le sexe de chaque auteur. Pour cela, il faut interroger l'API publique *Genderize*<sup>3</sup> qui associe un genre à chaque prénom. En appliquant une fonction aux valeurs de la colonne 'auteur', une colonne 'prenom' est créée dans le jeu de données, recensant le prénom de chaque auteur. Après avoir établi la liste des 203 différents prénoms, le genre est attribué à chaque prénom grâce à l'API. Cependant, bien que l'API soit fiable, des artistes ont un prénom prêtant à confusion. Par exemple, Jackie Mason est un humoriste masculin, bien que son prénom soit féminin. Cela peut être un nom de scène destiné à attirer l'attention, comme c'est le cas pour le chanteur Marilyn Manson par exemple. Ainsi, en établissant une base de données manuelle des auteurs, l'API *Genderize* atteint un taux de mismatch de seulement 5.91%.

### Qu'est-ce qu'une API?

L'interface de programmation d'application ou API (Application Programming Interface) est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels.

Définition [Wikipédia](#)

Il y a 12 erreurs, soit un mismatch de 5.91 % :

	genre_manuel	prenom	gender_API
26	Femme	Brett	male
33	Homme	Carrot	female
44	Homme	Dana	female
47	Homme	Dara	female
87	Homme	Henny	female
93	Homme	Jackie	female
103	Femme	Joan	male
114	Homme	Katt	female
158	Femme	Rene	male
178	Homme	Shelley	female
189	Femme	Tig	male
195	Homme	Tracy	female

### Comparaison des résultats entre API et genre manuel

<sup>3</sup> <https://pypi.org/project/Genderize/>

La base de données manuelle des 291 auteurs étant établie et vérifiée, les résultats de l'API ne seront pas retenus. Cette base de données manuelle dispose également des informations suivantes :

- l'origine ethnique
- la religion
- la date de naissance
- le statut mort ou vivant
- l'orientation sexuelle<sup>4</sup>

Le jeu de données définitif comporte finalement 18756 lignes pour 10 colonnes.

	citation	auteur	nbre_vues	theme	genre	ethnie	religion	date_naissance	vivant	orientation_sexuelle
0	Well, birthdays are merely symbolic of how ano...	Jerry Seinfeld	1666	[birthday]	Homme	blanc	athee	1954-04-29	Oui	heterosexuel
1	"Wait up!" That's what kids say. They don't sa...	Jerry Seinfeld	422	[kids, life, parents, future]	Homme	blanc	athee	1954-04-29	Oui	heterosexuel
2	Women have two orgasms, the real ones and the ...	Jerry Seinfeld	340	[women, accidents]	Homme	blanc	athee	1954-04-29	Oui	heterosexuel
3	Dogs have no money. Isn't that amazing? They'r...	Jerry Seinfeld	235	[dogs]	Homme	blanc	athee	1954-04-29	Oui	heterosexuel
4	My parents didn't want to move to Florida, but...	Jerry Seinfeld	219	[parents, laws, florida]	Homme	blanc	athee	1954-04-29	Oui	heterosexuel
...	...	...	...	...	...	...	...	...	...	...
18751	I miss having a pet. We're not allowing to hav...	John Caparulo	18	allow	Homme	blanc	athee	1975-09-22	Oui	heterosexuel
18752	I never understood how people could come to a ...	John Caparulo	16	Autres	Homme	blanc	athee	1975-09-22	Oui	heterosexuel
18753	They got this program called Intervention. Whi...	Tracy Morgan	21	Autres	Homme	afro-american	chretien	1968-11-10	Oui	heterosexuel
18754	I had one DWI, which was a bogus charge, becau...	Ron White	20	Autres	Homme	blanc	athee	1956-12-18	Oui	heterosexuel
18755	The N-Word. And everyone says, "Don't say it a...	Whoopi Goldberg	19	word	Femme	afro-american	athee	1955-11-13	Oui	heterosexuel

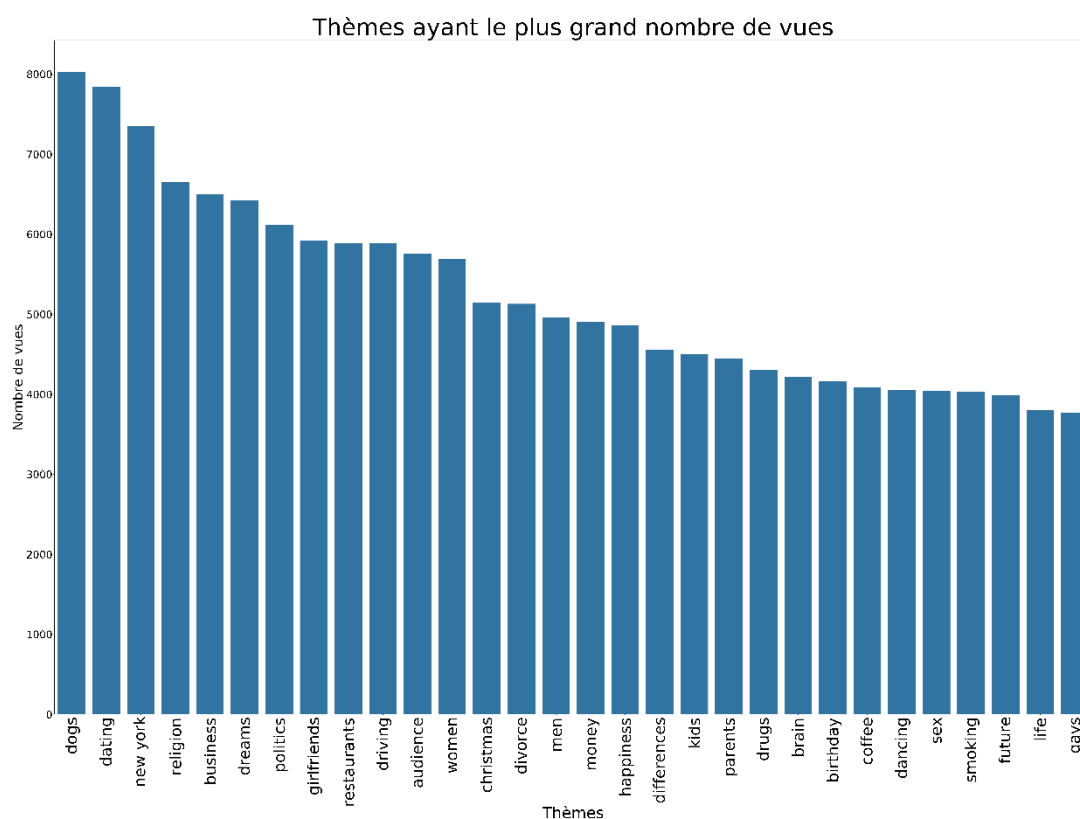
### *Jeu de donnée définitif*

<sup>4</sup> Toutes les informations sur les artistes, leur date de naissance, de décès (pour certains), leur origine, leur religion et leur orientation sexuelle, proviennent de Wikipédia. Concernant la religion et l'orientation sexuelle, lorsque celles-ci ne sont pas précisées, l'artiste est considéré comme athée et hétérosexuel.



### 3. Des thèmes grand public

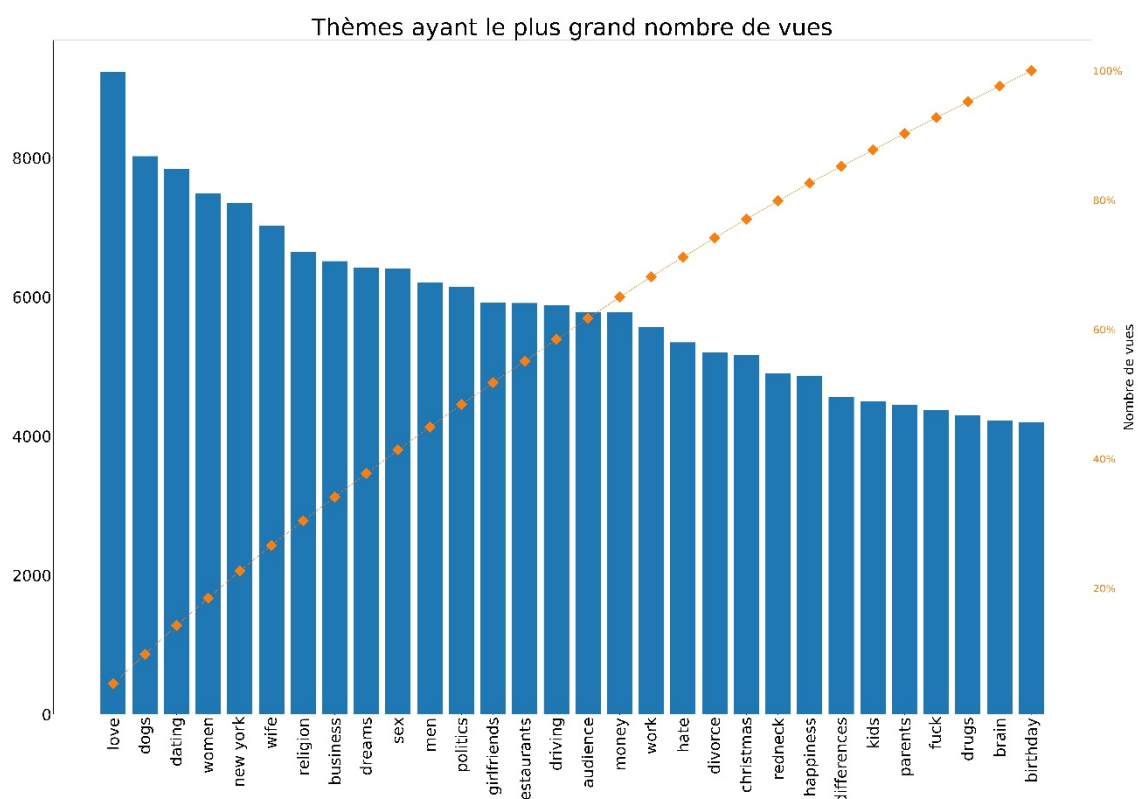
Premier élément-clé : cette étude révèle que les thèmes ayant le plus de succès sont totalement adaptables aux productions du Cirque du Soleil puisqu'ils s'adressent à tous publics.



*Top 30 des thèmes créés par SComedy*

Il s'agit du classement des thèmes déjà existants sur SComedy. Il y a bien sûr des sujets clivants, tels que la religion et la politique. Il y a également des thèmes qui ne conviendraient pas aux enfants : le sexe et la drogue. Mais arrivent finalement en tête des sujets grand publics : les chiens, les rendez-vous amoureux et New York, la Mecque de l'humour.

Une fois les thèmes rajoutés par l'analyse sémantique, voici le classement final :



### Top 30 définitif des thèmes

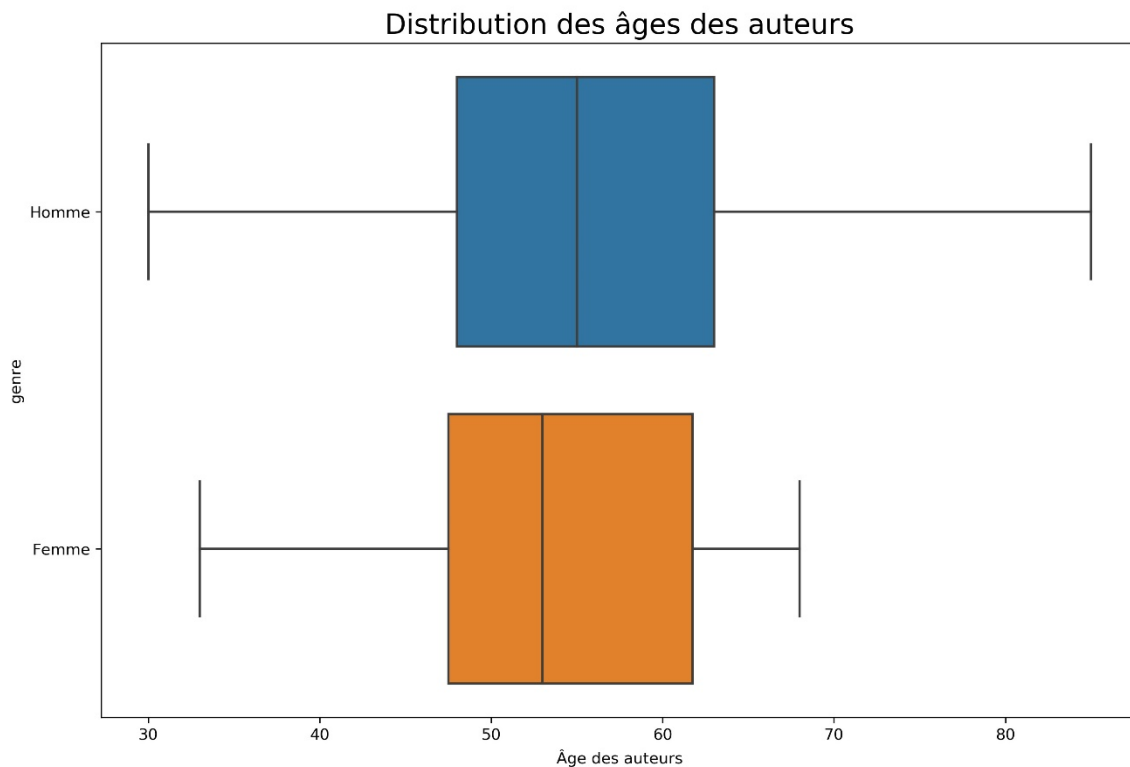
Le classement des thèmes a changé. Arrive désormais en tête l'amour avec plus de 9000 nombres de vues. Le thème des femmes passe en quatrième position avec plus de 7400 nombres de vues. Il paraît difficile à ce stade de l'analyse de savoir si ce thème traite plus du sexisme ou du féminisme. Mais le choix de l'angle permet d'adapter ces sujets à un plus large public.

L'histogramme informe également que 80% des nombres de vues sont réalisées par des citations regroupées par seulement 22 thèmes, soit moins de 3,5% des thèmes :

- |                         |             |                 |              |               |
|-------------------------|-------------|-----------------|--------------|---------------|
| 1. Amour                | 6. Épouse   | 11. Hommes      | 16. Audience | 21. Christmas |
| 2. Chiens               | 7. Religion | 12. Politique   | 17. Argent   | 22. Redneck   |
| 3. Rendez-vous amoureux | 8. Business | 13. Petite amie | 18. Travail  |               |
| 4. Femmes               | 9. Rêves    | 14. Restaurants | 19. Haine    |               |
| 5. New York             | 10. Sexe    | 15. Conduite    | 20. Divorce  |               |

## 4. Un public très large

L'humour est générationnel. Chaque humoriste séduit particulièrement un public de la même tranche d'âge. Partant de ce constat, l'âge des humoristes présentés sur *SComedy* permet de cibler n'importe quel adulte.



Les humoristes ont en effet entre 30 et 93 ans. Le plus jeune est Bo Burnham et le plus âgé est Mort Sahl. La moyenne d'âge est de 55 ans. À noter que 50 humoristes répertoriées sur *SComedy* sont morts.

Pour les femmes, la plus jeune humoriste, Sabrina Matthews, a 33 ans tandis que la plus âgée, Roseanne Barr, a 68 ans. Leur moyenne d'âge est de 53 ans, ce qui est aussi l'âge médian. La majorité d'entre elles ont 50 ans.

Chez les hommes, la moyenne d'âge est de 56 ans. L'âge médian est de 55 ans et la majorité d'entre eux ont 52 ans.

## 5. Des interprètes aux multiples facettes

Les États-Unis sont une société multiethnique. Cette diversité se retrouve au sein des artistes répertoriés sur *SComedy*. Cependant, les artistes blancs sont surreprésentés avec 224 humoristes, soit plus de 75% d'entre eux. Les artistes afro-américains sont la deuxième population la mieux représentées avec plus de 16% des humoristes.

Ethnies	Nombre d'humoristes	Pourcentage d'humoristes
Blanc	224	76.98 %
Afro-américain	48	16.49 %
Hispanique	7	2.41 %
Asiatique	4	1.37 %
Indo-américain	3	1.03 %
Italo-américain	3	1.03 %
Américano-iranien	1	0.34 %
Russo-américain	1	0.34 %

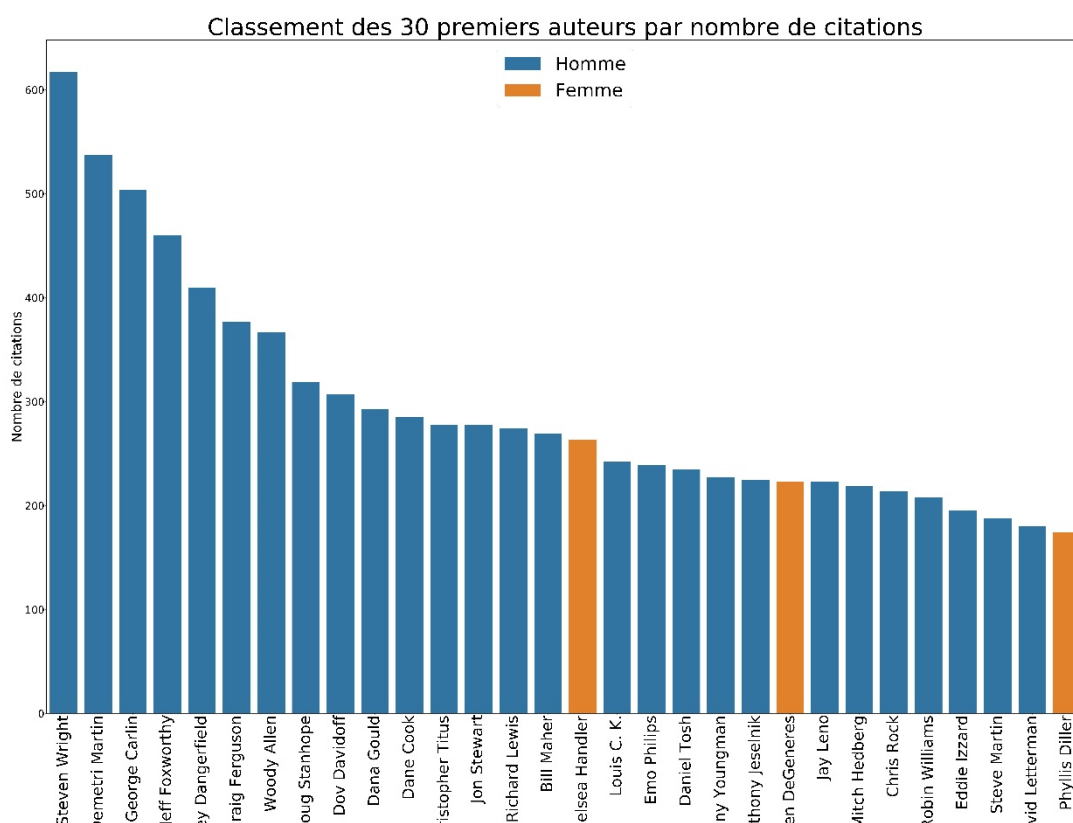
### *Répartition des humoristes selon leur ethnie*

D'autre part, seulement 14% des 291 artistes sont des femmes, soit 41 au total. Le stand-up américain semble être un monde d'homme. Cette tendance se confirme au niveau du nombre de citations. Sur les 18756 citations, seulement 1787 sont écrites par les artistes féminines, soit moins de 10%. De même, en terme de nombre de vues cumulées, seulement 7.7% du nombre de vues total est attribué aux citations dont l'auteur est une femme.

Au niveau du classement des comédiens par nombre de citations, le classement est une nouvelle fois dominé par les hommes. L'auteur le plus actif est Steven Wright avec plus de 600 citations. Cet auteur aborde des thèmes grand public tels que l'anniversaire, les chiens ou encore le travail.

*« I named my dog Stay, so I can say, 'Come here, Stay ! Come here, Stay !' »*

*Steven Wright*



La première femme de ce classement est Chelsea Handler et se place à la 16<sup>ème</sup> position avec 263 citations.

Si nous classons les comédiens par nombre de vues cumulées, c'est encore une fois Steven Wright qui arrive en tête avec plus de 33000 vues, soit 3,75% du total de vues. Le classement obtenu des 10 meilleurs comédiens diffère de celui publié sur le *SComedy*<sup>5</sup>.

C'est effectivement Brett Butler qui arrive en tête du classement publié sur *SComedy*.

*« There are no accidents. God's just trying to remain anonymous. »*

*Brett Butler*

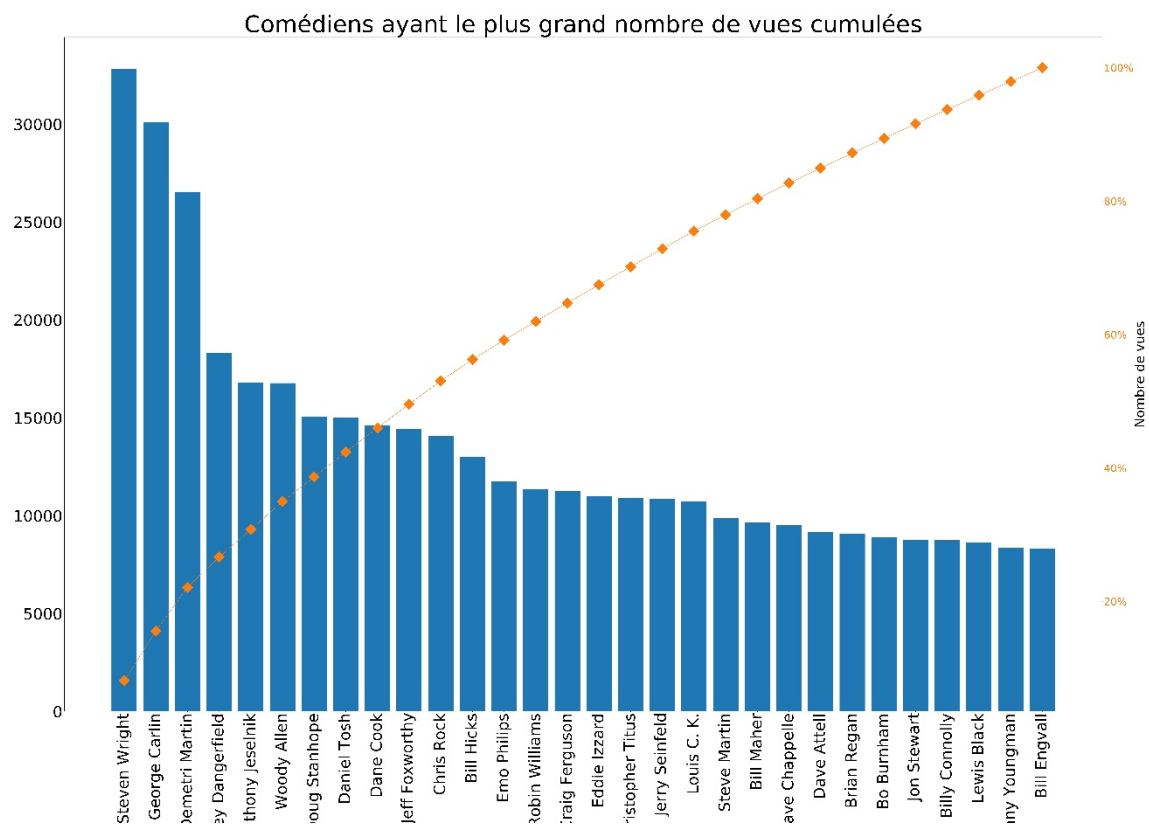
<sup>5</sup> <https://scomedy.com/>



Classement SComedy	Classement obtenu	Pourcentage de nombre de vues pour classement obtenu
Brett Butler	Steven Wright	3,76%
Anthony Jeselnik	George Carlin	3,46%
Adam Sandler	Demetri Martin	3,04%
Bill Hicks	Rodney Dangerfield	2,11%
Bonnie McFarlane	Anthony Jeselnik	1,94%
Bobcat Goldthwait	Woody Allen	1,93%
Daniel Tosh	Doug Stanhope	1,74%
Bill Burr	Daniel Tosh	1,73%
Ray Romano	Dane Cook	1,68%
Steven Wright	Jeff Foxworthy	1,65%

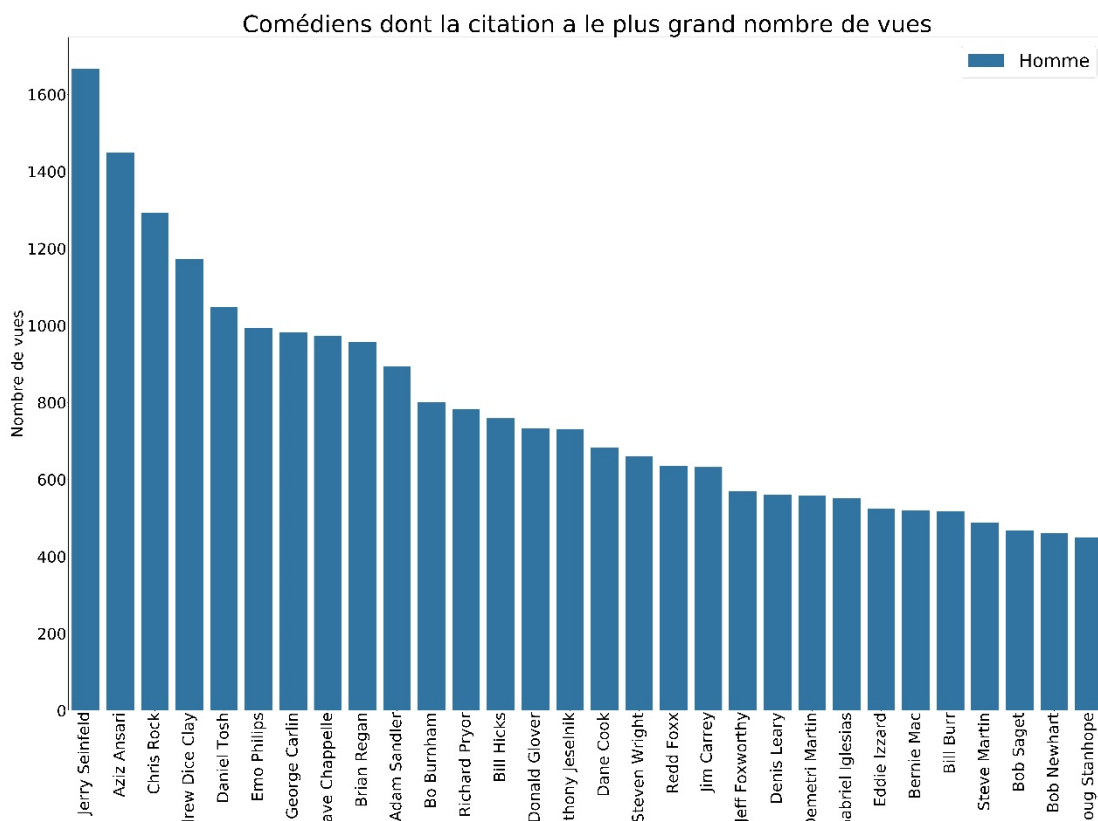
### Différences de classement entre celui de SComedy et celui obtenu dans l'analyse

À noter que 80% des vues sont réalisées par seulement 22 artistes.



Concernant le classement des citations par nombre de vues, c'est cette fois-ci Jerry Seinfeld qui arrive en tête avec plus de 1600 vues pour sa citation sur les anniversaires :

*« Well, birthdays are merely symbolic of how another year has gone by and how little we've grown. No matter how desperate we are that someday a better self will emerge, with each flicker of the candles on the cake, we know it's not to be, that for the rest of our sad, wretched pathetic lives, this is who we are to the bitter end. Inevitably, irrevocably; happy birthday? No such thing. »*



Enfin, quels thèmes abordent ces auteurs ? Et surtout, les artistes s'autorisent-ils à évoquer des sujets qui ne les concernent pas ?

Concernant l'homosexualité, sur les 29 citations abordant ce thème, seulement deux sont écrites par Sandra Bernhard qui est bisexuelle. Paula Poundstone, qui se revendique asexuelle, est également l'autrice d'une de ces 29 citations. Les autres citations sont écrites par des auteurs hétérosexuels.

---

Au sujet du racisme, 49 citations sont regroupées par le thème ‘black’ ou ‘racism’ et 26 sont attribuées à des auteurs afro-américains.

Sur l’antisémitisme, parmi les 9 citations classées dans le thème ‘jews’, seulement un tiers sont écrites par des artistes juifs.

Ces sujets précités sont ciblés pour leur caractère discriminatoire. Cette analyse montre qu’ils peuvent être abordés par des artistes qui n’en sont pas les sujets.

---

## Conclusion

L'analyse du site Internet *SComedy* révèle que les thèmes les plus récurrents dans le stand-up américain sont parfaitement compatibles avec une production du Cirque du Soleil : l'amour, les chiens et les rendez-vous amoureux.

La diversité des spectacles de la firme québécoise prouve que cet art a toute sa place parmi ses références. D'autant plus que si chaque détail du cirque traditionnel a été revisité, jamais les scènes de clowns ne s'en sont éloignées.

D'autre part, le public des principaux humoristes exerçant aux États-Unis assure au Cirque du Soleil un cœur de cible n'épargnant aucune tranche d'âge.

Enfin, au sujet des choix des interprètes, l'étude montre que malgré la diversité au sein du monde des humoristes et la présence de femmes, cet univers reste majoritairement représenté par les hommes blancs. Après les mouvements *#MeToo* et *Black lives matter*, le Cirque du Soleil se doit de faire preuve d'exemplarité. L'analyse montre que le choix de l'interprète ne fait nullement obstacle aux thèmes abordés.

Car ce n'est pas le choix du thème, clivant ou non, qui pose problème, mais le choix de son angle. Un thème clivant peut donc être abordé par n'importe quel artiste du moment que son angle ne prête pas à confusion.