

Charlotte Dugrain
n° étudiant : 21502249
(Paris Diderot)

pour le 28/02/2020

Datavisualisation

Analyse du concours Eurovision de la chanson

Synopsis

À partir des données déjà disponibles ainsi que des données obtenues en analysant des tweets concernant l'évènement, on présentera le concours Eurovision de la chanson, plus communément appelé Eurovision, sous différents aspects : pays ayant le plus remporté le concours, analyse de sentiments des tweets d'une édition spécifique... Ce concours, créé en 1956, a eu lieu chaque année sans interruption, ce sera donc la 65^{ème} édition qui se tiendra en 2020. S'étendant d'année en année passant de 7 pays à sa création à 41 pays en 2020, il sera intéressant d'étudier cet évènement dans sa globalité, mais aussi son évolution à travers les années et enfin une de ses éditions en particulier. Il sera également intéressant de voir si les avis exprimés sur les prestations sur les réseaux sociaux (ici Twitter) sont cohérents avec les résultats du concours. Ce concours se déroule en 3 étapes : 2 demi finales et une finale. Le pays hôte ainsi que l'Allemagne, l'Espagne, la France, l'Italie et le Royaume-Uni sont d'office qualifiés pour la finale chaque année, tandis que les autres pays participent aux demis finales dans l'espoir de figurer parmi les 26 pays accédant à la finale. Le classement est déterminé pour moitié par le vote des téléspectateurs, et par moitié par un jury composé de membres de chaque pays. Les téléspectateurs tout comme le jury ne peuvent pas voter ou attribuer de points à leur propre pays.

Synthèse

Sources

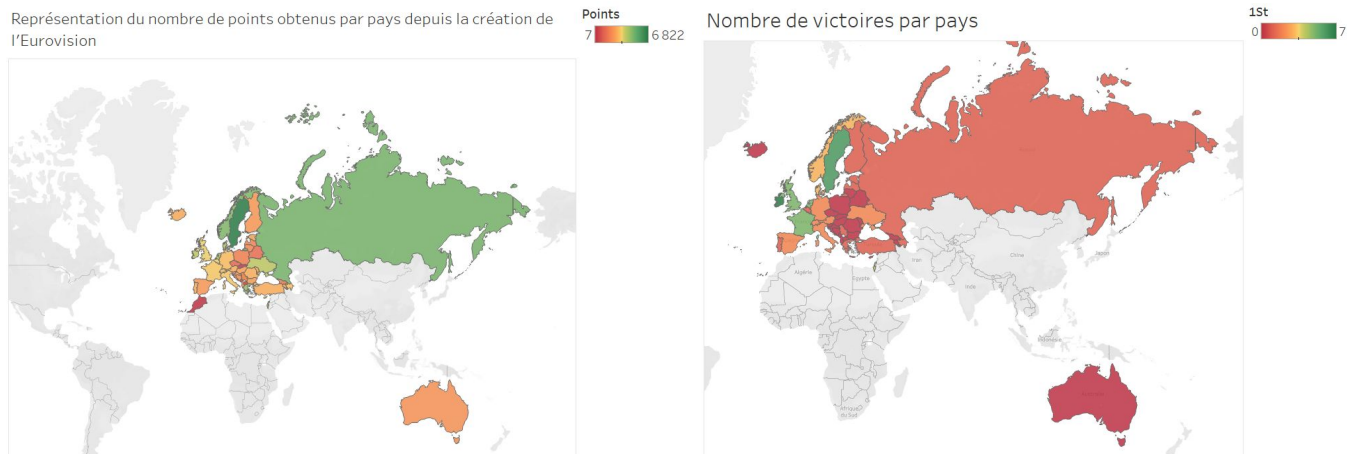
- Dataworld
- Eurovision song contest database : <https://eschome.net/>
- Twitter : tweets postés le jour de la finale 2019 avec un des hashtag des différents pays participants à la finale (26 pays)
- <https://lyricstranslate.com/en/collection/eurovision-song-contest-winners>

Datavis générées

https://eu-west-1a.online.tableau.com/#/site/datavizdugraincharlotte/projects/141380?origin=card_share_link

Comme nous faisons face à un concours on peut s'intéresser dans un premier temps aux pays ayant obtenu le plus de points ou le plus de victoires. On pourrait s'attendre à ce que ces deux notions soient équivalentes étant donné que gagner revient à remporter le plus de points. Hors, ce serait faire abstraction de l'histoire de l'Eurovision, en effet certains pays ne prennent part au concours que depuis quelques années et ont donc bien moins de points bien qu'ils aient plus ou autant de victoires que d'autres, ce qui explique la différence entre les 2 cartes de nos visualisations.

- Datavis 1 et 2 : Cartes des nombres de points reçus par pays depuis la création de l'Eurovision vs nombre de victoires par pays également depuis la création du concours



https://eu-west-1a.online.tableau.com/#!/site/datavizdugraincharlotte/workbooks/291917?:origin=card_share_link

https://eu-west-1a.online.tableau.com/#!/site/datavizdugraincharlotte/workbooks/291919?:origin=card_share_link

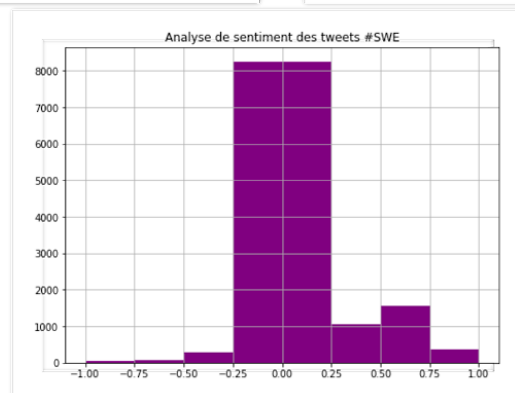
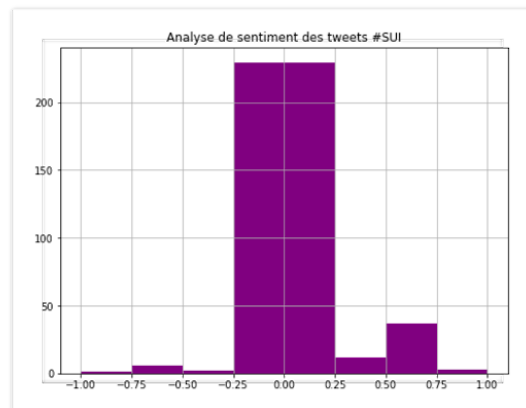
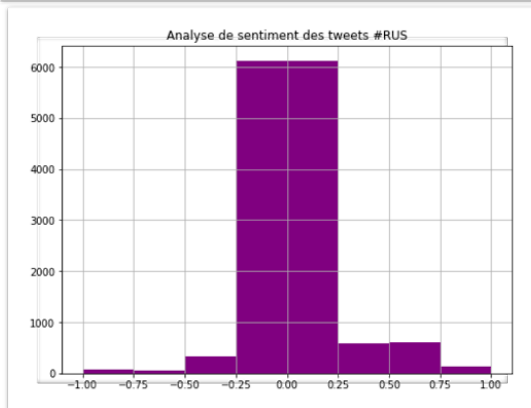
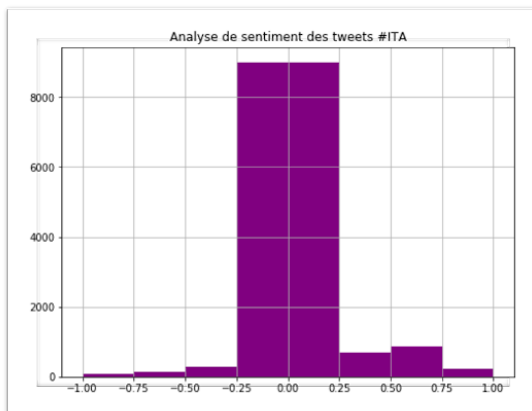
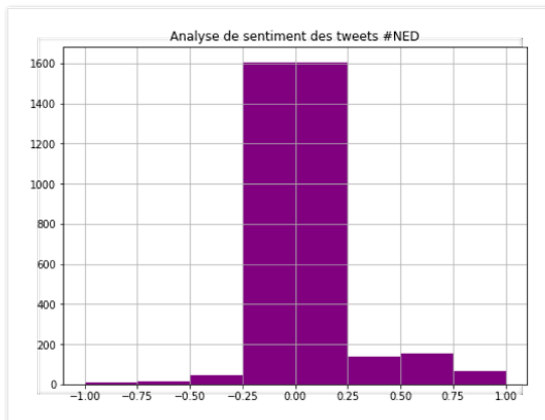
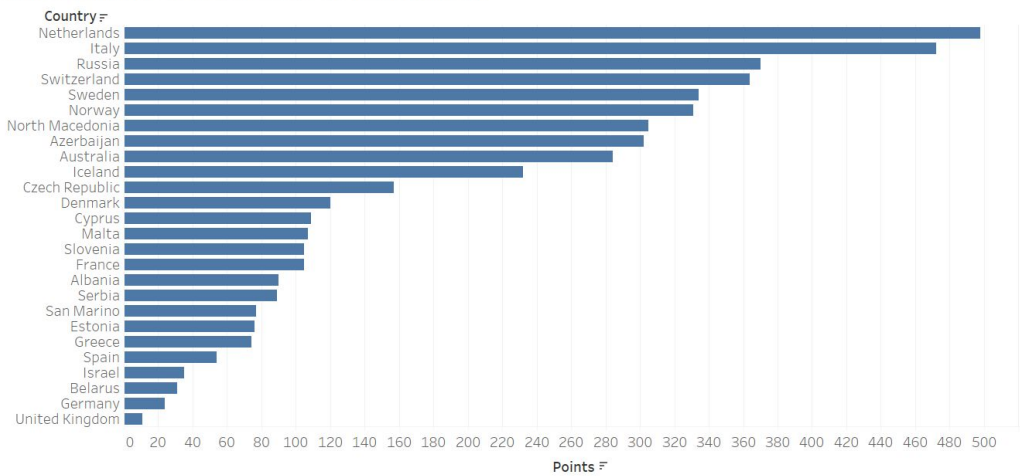
Comme expliqué dans notre synopsis, l'Eurovision est très suivi et commenté sur les réseaux sociaux et plus particulièrement sur Twitter, il pourrait donc être intéressant de voir si les internautes sont d'accord avec le classement de ce concours.

On procédera pour cela de la façon suivante : lorsque les internautes commentent l'eurovision, ils sont très nombreux à utiliser les hashtags donnés pendant l'émission, un hashtag par pays. On récupérera donc les tweets avec les hashtags des pays sur la journée du 18 mai 2019 jour de la finale. Puis on estimera leur polarité à l'aide de textblob, plus le tweet est positif est plus il est proche de 1, plus il est négatif plus il est proche de -1, 0 correspond à un tweet considéré comme neutre. On décide d'observer les tweets concernant le top 5 de l'édition 2019, dans l'ordre : Pays-Bas (#NED), Italie #ITA, Russie (#RUS), Suisse (#SUI), Suède (#SWE).

On s'aperçoit qu'il y a bien des différences par exemple les pays ayant proportionnellement le plus de tweet très positifs ($>0,50$) sont les 4 et 5ème de ce top 5 et non les 2 premiers. On voit également que le nombre de tweets est bien moindre pour la Suisse et les Pays-Bas, une vérification plus poussée de l'extraction serait peut être nécessaire.

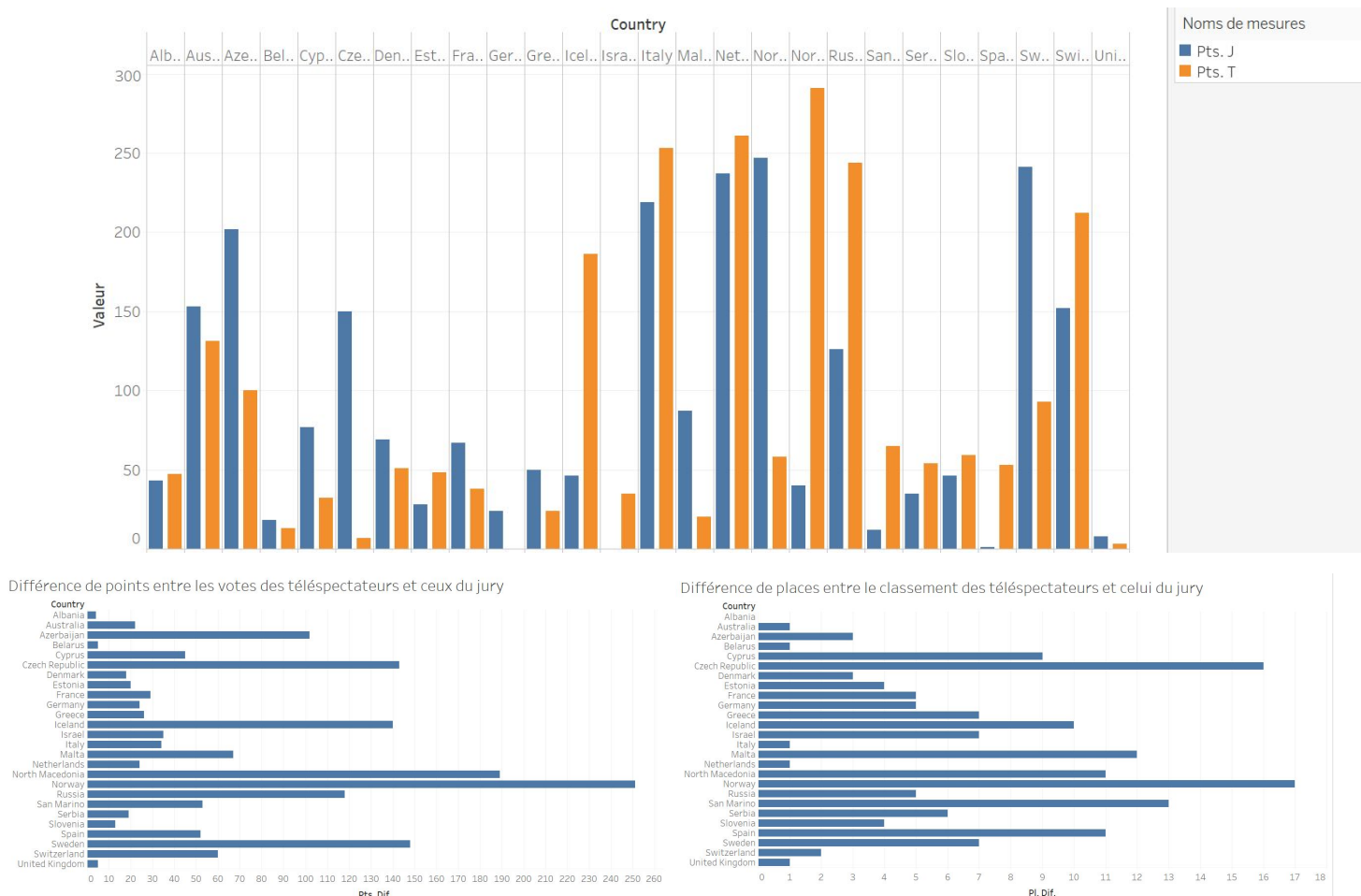
- Datavis 3 et 4 : Classement 2019 vs polarité des tweets 2019

Classement de la finale 2019 de l'Eurovision



Ce concours partageant les points attribués entre un jury des différents pays et d'autres parts des téléspectateurs (sachant que les habitants d'un pays ne peuvent pas voter pour leur propre candidat) on peut se demander s'il y a de grandes disparités entre l'avis des téléspectateurs et celui du public. On observera ces disparités en terme de différence de points et de différence de places. On voit que les différences peuvent être très importantes, allant jusqu'à 250 points et 17 places.

- Datavis 5 : différence points jury points téléspectateurs 2019



https://eu-west-1a.online.tableau.com/#/site/datavizdugraincharlotte/workbooks/291998?:origin=card_share_link

On peut enfin se demander ce qui fait d'une chanson une bonne prétendante à la victoire dans ce concours, on va donc analyser les paroles des dix dernières chansons en langue anglaise ayant gagnées l'Eurovision. On choisira les dix dernières chansons gagnantes en anglais et non pas les 10 dernières chansons afin de simplifier l'analyse et de ne pas faire face à des problèmes ou imprécisions de traduction.

On en ressortira tout d'abord un wordcloud avec tous les mots, puis le résultat n'étant pas très lisible on choisira de conserver les mots apparaissant au minimum 10 fois. On s'aperçoit qu'on a des mots très communs à la langue anglaise ce qui explique leur présence mais par exemple on voit 'I' qui est seulement le 20ème mot le plus courant de la langue anglaise est dans les mots les plus présents de ces paroles, alors, égocentriques ces gagnants ?

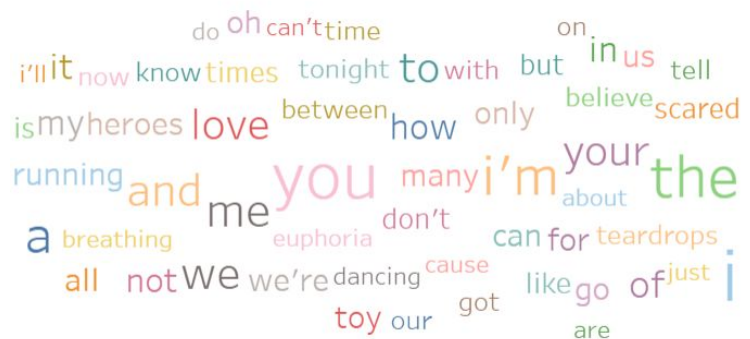
On voit que 'love' est le premier nom très représenté, ce qui nous surprend peu l'amour étant un thème très universel et très utilisé dans la musique, on voit que les mots des refrains des paroles apparaissent également beaucoup, il aurait pu être plus précis de représenter le classement de la proportion des mots dans chaque chanson puis de faire un classement plus global.

- Datavis 6 et 7 : Répartition des mots les plus utilisés dans les paroles des 10 dernières chansons en anglais ayant gagnées l'Eurovision

Répartition des mots dans les 10 dernières chansons en anglais ayant gagnées l'Eurovision



Répartition des mots dans les 10 dernières chansons en anglais ayant gagnées l'Eurovision
(mots apparaissant au moins 10 fois)



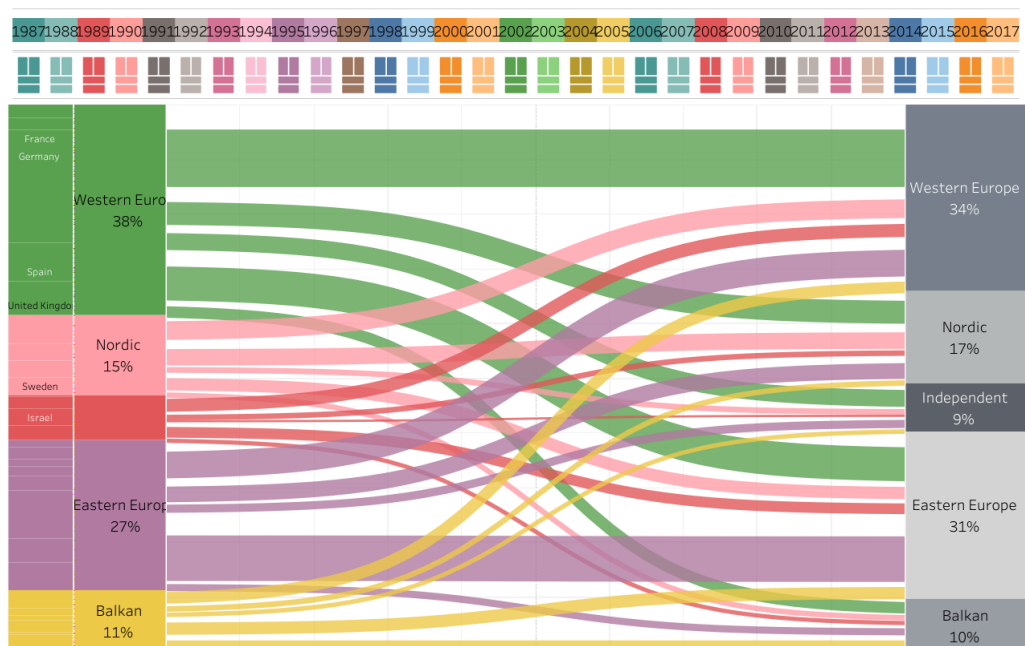
https://eu-west-1a.online.tableau.com/#/site/datavizdugraincharlotte/workbooks/291095?:origin=card_share_link

https://eu-west-1a.online.tableau.com/#/site/datavizdugraincharlotte/workbooks/291098?:origin=card_share_link

On aurait pu également analyser d'autres aspects du concours tel que le genre des participants et des gagnant : est ce qu'il y a parité au sein de ce concours ?

Ou bien encore la distribution des points entre les pays, voir si comme certains peuvent le penser la plupart des pays votent pour leurs pays voisins/alliés.

On peut d'ailleurs voir sur dataworld que certaines personnes se sont penchés sur la question :



Lífið er aðeins betra með Sankey línunum:

Myndin sýnir hvernig landa-blokkir greiða atvæði til mismunandi blokka. Ef eitt ár er valið eru eingöngu gögn tekin fyrir það ár. Fjöldi atkvæða ræður stærð línanna, en prósentur sýna hlut atkvæða sem eru gefin (vinstra megin) og þegin (hægra megin)