**\section{Analyse des données}**

**\subsection{Distribution des données}**

Pour l’immense majorité des lieux recensés, seuls quelques concerts y ont été organisés entre 2010 et 2019, comme le montre les données du tableau \ref{tab\_concerts\_venues}. Près de la moitié des lieux (48\%) n’ont qu’un seul événement répertorié, tandis que pour 95\% des salles, il y a eu au maximum 38 événements reconnus par Songkick. Toutefois, si la majorité des lieux accueillent peu de concerts, les artistes recensés dans notre base de données sont largement représentés par ces 5\% de salles plus actives, comme on peut le voir sur le tableau \ref{tab\_salles\_actives}. Chacune de ces 251 salles a accueilli en dix ans en moyenne plus de 170 concerts. Plus de huit artistes sur 10 présents dans notre base de données a fréquenté au moins une de ces salles. Les genres issus de Spotify sont presque tous représentés dans ces salles les plus actives, tandis que les 23 top genres définis se retrouvent dans ces lieux. Ce déséquilibre flagrant dans la distribution des données interroge sur la nature des salles qui semblent dominer, mais qui ne font pas preuve d’une influence considérable sur la scène suisse. \par

\begin{table}[h]

\centering

\begin{tabular}{|l|c|c|c|c|c|c|c|c|c|}

\hline

Nbr concerts par salle & 1 & 2 & 3 & 4 & 5 & 6 & 10 & 20 & 38 \\ \hline

Part des salles de concert & 0.48 & 0.14 & 0.08 & 0.05 & 0.03 & 0.02 & 0.01 & 0.03 & 0.001 \\ \hline

Part cumulée des salles & 0.48 & 0.62 & 0.70 & 0.74 & 0.77 & 0.80 & 0.86 & 0.92 & 0.95 \\ \hline

\end{tabular}

\captionsetup{width=.8\textwidth}

\caption{Distribution du nombre de concerts par salle.}

\label{tab\_concerts\_venues}

\end{table}

\begin{table}[h]

\centering

\begin{tabular}{|l|c|c|c|c|c|}

\hline

& Salles & Concerts & Artistes & Genres Spotify & Top genres \\ \hline

Toutes les données & 5'063 & 62'827 & 24'719 & 3'649 & 23 \\ \hline

Salles (n concerts \textgreater 39) & 251 & 43'478 & 20'417 & 3'422 & 23 \\ \hline

\% (n concerts \textgreater 39) & 0.05 & 0.70 & 0.83 & 0.94 & 1.00 \\ \hline

\end{tabular}

\captionsetup{width=.8\textwidth}

\caption{Distribution des données dans les salles les plus actives.}

\label{tab\_salles\_actives}

\end{table}

L’examen d’un échantillon aléatoire de 30 salles qui n’ont accueilli qu’un seul concert en dix ans révèle qu’aucun de ces lieux n’est en fait une salle de concert à proprement parler. Il s’agit de restaurants, de bars, d’hôtels, de lieux publics, de festivals, de lieux culturels non-dédiées à la musique, de lieux éphémères et d’autres lieux divers (cf. tableau \ref{tab\_sample\_lieux}. Sur cet échantillon, certes très réduit par rapport aux milliers de salles peu actives, on ne retrouve donc aucune salle susceptible d’organiser des concerts de manière régulière, et ces lieux sont caractérisés par l’aspect éphémère et rare des concerts. Cette représentation massive des lieux avec peu de concerts suggère que Songkick a une compréhension très large des concerts, qui va au-delà du réseau habituel de salles et lieux de concerts. \par

\begin{table}[h]

\centering

\begin{tabular}{|l|c|c|c|c|c|c|c|}

\hline

& Rest. & Ephémères & Lieux-dits & Festival & Culturels & Autres & Total \\ \hline

Nombre & 7 & 2 & 4 & 6 & 4 & 7 & 30 \\ \hline

Proportion & 0.23 & 0.07 & 0.13 & 0.20 & 0.13 & 0.23 & 1.00 \\ \hline

\end{tabular}

\captionsetup{width=.8\textwidth}

\caption{Types de salles d’un échantillon de 30 lieux qui ont accueilli un seul concert entre 2010 et 2019.}

\label{tab\_sample\_lieux}

\end{table}

**\subsection{Corrélations des données acoustiques}**

Une matrice de corrélation est calculée sur l’ensemble des variables acoustiques (acousticness, danceability, duration, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence) afin de mettre en évidence l’existence ou non de liens entre certains de ces indicateurs. Des variables dépendantes l’une de l’autre pourraient servir de point de comparaison initial entre plusieurs salles de concert en observant si les salles suivent cette corrélation ou non. Les variables qui ont une forte corrélation (coefficient \textgreater 0.50) représentent des aspects qui, logiquement, sont liés : energy et loudness ; energy et acousticness (négativement) et loudness et acousticness (négativement). On peut en effet facilement concevoir que plus une musique est énergique, plus son volume sonore est élevé et que plus une musique est énergique ou forte, moins elle est acoustique. Viennent ensuite deux paires de variables avec un lien modéré (coefficient \textgreater 0.30 et \textless 0.50) : danceability et valence et loudness et instrumentalness (négativement). Là aussi les rapprochements entre ces éléments sont plutôt évidents puisque plus une musique est dansante, plus elle suppose un sentiment positif, et plus une musique est forte, moins l’aspect instrumental se fait ressentir. Enfin, on observe plusieurs de variables avec un coefficient de corrélation faible (coefficient \textless 0.30) : energy-liveness, energy-tempo, energy-valence, loudness-valence, loudness-tempo, loudness-duration, acousticness-tempo, acousticness-valence, instrumentalness-duration, valence-duration. Comme pour les liens forts et moyens, les relations entre ces variables représentent une certaine évidence musicale. La présence de liens établis entre plusieurs variables permet donc à la fois de confirmer que ces données représentent une certaine réalité acoustique et un point de départ pour les analyses des salles de concert. \par

**\section{Mise en valeur des données collectées}**

**\subsection{Cartographie des salles de concert}**

Tout d’abord, un outil capable de représenter la distribution géographique des salles de concert en Suisse a été développé. La cartographie s’articule autour de deux paramètres, les genres musicaux et les artistes. \par

\begin{figure}[htp]

\centering

\includegraphics[width=1\textwidth]{images/resultats/viz\_geo\_genre.png}

\captionsetup{width=.8\textwidth}

\caption{Visualisation de la distribution géographique des genres jazz, hip hop et dance. Les points de couleur représentent une salle de concert. Les points noirs entourés d’une couleur représentent le centroïde des salles d’un genre. Les lignes grises représentent le lien entre des salles et leur centroïde.}

\label{fig\_viz\_geo\_genre}

\end{figure}

Dans le premier cas (voir la figure \ref{fig\_viz\_geo\_genre}), l’outil permet de voir quels genres musicaux prédominent dans quelles salles, en montrant les lieux où un genre est programmé au-delà d’une certaine fréquence. Pour cela, il est nécessaire de sélectionner un ou plusieurs genres et un seuil de fréquence minimale. Par exemple, un seuil de 40\% signifie que les salles retenues sont seulement celles où au moins 40\% des artistes sont apparentés aux genres sélectionnés. Pour tenir compte de la distribution inégale de l’activité des salles, les salles peuvent être filtrées en fonction du nombre de concerts minimum qui y ont été organisés. La carte est ainsi épurée des lieux peu pertinents pour représenter la scène musicale. La position moyenne des salles apparaît sous la forme d’un centroïde qui met en évidence la région dans lequel un genre est le plus fréquent. Les liens qui convergent des salles vers leur centroïde, affichés en transparence, indiquent si une zone a un poids plus important pour un genre. Par exemple si plusieurs salles apparentées à un genre se trouvent dans une même ville, leurs liens vers le centroïde se superposent et apparaissent d’une manière plus visible. \par

\begin{figure}[htp]

\centering

\includegraphics[width=1\textwidth]{images/resultats/viz\_geo\_artist.png}

\captionsetup{width=.8\textwidth}

\caption{Visualisation de la distribution géographique des artistes Duck Duck Grey Duck et Greis. Les points de couleur représentent une salle de concert. Les points orangés plus grands représentent le centroïde des salles fréquentées par un artiste. La couleur du centroïde représente le coefficient de mobilité pondérée d’un artiste et varie du jaune (peu mobile) au rouge (très mobile). Les lignes grises représentent le lien entre des salles et leur centroïde.}

\label{fig\_viz\_geo\_artist}

\end{figure}

Dans le second cas, la carte indique dans quelles salles les artistes se produisent en marquant d’un point tous les lieux fréquentés par un artiste. Pour cela, il suffit de sélectionner un ou plusieurs artistes pour afficher son historique de concerts en Suisse. La position moyenne des salles fréquentées par un artiste apparaît sous la forme d’un centroïde qui met en évidence la région dans lequel cet artiste y est le plus souvent. Le centroïde rend aussi compte du coefficient de mobilité pondérée d’un artiste en variant sa couleur. Plus un artiste est mobile, c’est-à-dire plus il se déplace et moins il ne retourne dans des mêmes salles, plus la couleur du centroïde tend vers le rouge. En revanche, moins un artiste est mobile plus la couleur du centroïde tend vers le jaune. Comme pour la cartographie des genres, les liens entre les salles et leur centroïde sont légèrement transparents pour montrer quels sont les « couloirs » les endroits où un artiste passe le plus souvent. Ces lignes pointent vers les centroïdes et non entre les salles pour des raisons de lisibilité et de compréhension. Si les lignes reliaient une salle à une autre, il en résulterait un amas de lignes sans nœud ni direction qui n’aideraient en rien la compréhension de la distribution géographique des salles de concert. \par

\begin{figure}[htp]

\centering

\begin{subfigure}[t]{0.45\textwidth}

\centering

\includegraphics[width=\textwidth]{images/resultats/viz\_geo\_jitter\_1.jpg}

\caption{Plusieurs individus pointent vers le même endroit. Il apparaît évident que ces points représentent le même lieu.}

\end{subfigure}

\hfill

\begin{subfigure}[t]{0.45\textwidth}

\centering

\includegraphics[width=\textwidth]{images/resultats/viz\_geo\_jitter\_2.jpg}

\caption{En zoomant sur le lieu, on peut facilement distinguer tous les individus. Le point bleu central est le point de départ. Les points forment ensuite une spirale rectangulaire dans le sens des aiguilles d'une montre.}

\end{subfigure}

\captionsetup{width=.8\textwidth}

\caption{Effet de l'algorithme de fluctuation sur les données spatiales.}

\label{fig\_jitter}

\end{figure}

Si plusieurs points, représentant des genres ou des artistes, se partagent le même lieu (par exemple, une salle propose régulièrement du jazz et de la musique classique), ils ne se superposent pas puisqu’un algorithme de fluctuation (\emph{jittering} en anglais) est appliqué sur tous les points qui partagent les mêmes coordonnées (cf. figure \ref{fig\_jitter}). Pour chacun de ces points, l’algorithme décale d’un cran sa position par rapport à celle du point précédent, pour former une spirale rectangulaire autour de l’emplacement du point de base (cf. Algorithme \ref{anx\_alg\_spirale} pour le fonctionnement en détail). Ce cran représente environ un mètre sur la carte, soit 0.01 millième d’un degré de coordonnées géographique \footnote{Un degré de latitude correspond à environ 110 km, un degré de longitude à environ 80 km à la hauteur de la Suisse.}. Ce décalage minime par rapport aux coordonnées de base altère de manière négligeable la précision des points sur la carte, mais améliore considérablement la lisibilité des données \autocite{ross\_unhide\_2014, boyles\_how\_2019}. \par

**\subsection{Caractéristiques des salles de concert}**

L’outil de cartographie permet de mettre en relation des données spatiales (la position des salles) avec des données musicales (les genres prédominants dans les salles) et avec des données événementielles (les concerts des artistes). D’autres outils de visualisation ont été conçus pour mettre en évidence ce qui définit chaque salle par rapport aux autres et en son sein. \par