Dans cette partie, nous verrons en profondeur comment les données relatives aux concerts peuvent renseigner sur la scène musicale suisse et comment mettre les mettre en valeur grâce à des outils de visualisation de données.

**\section{Analyse des données}**

Nous nous intéresserons à la distribution des données, à l’existence de corrélations entre certaines variables et la pertinence d’appliquer un partitionnement de ces 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 montrent 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}[ht]

\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}[ht]

\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}[ht]

\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}**

Si certaines variables obtenues sont logiquement liées entre elles, comme par exemple la mobilité avec la mobilité pondérée ou le nombre de followers avec le nombre d’auditeurs, les relations entre les variables acoustiques nécessitent une analyse dédiée. Afin de mettre en évidence l’existence ou non de liens entre certains de ces indicateurs, une matrice de corrélation est calculée sur l’ensemble des variables acoustiques (acousticness, danceability, duration, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence). 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

**\subsection{Partitionnement des données}\label{section\_kmeans}**

La méthode des k-moyennes, un algorithme de partitionnement (\emph{k-means clustering} en anglais) qui sépare un jeu de données en un nombre $k$ de partitions \footnote{Le k-means est un algorithme courant pour des tâches de classification non-supervisée. \url{https://fr.wikipedia.org/wiki/K-moyennes}}, a été appliquée sur les données pour observer comment les différentes caractéristiques acoustiques, spatiales, temporelles et de popularité permettent de grouper les salles de concert. Pour cette section, seules les salles où au moins 100 artistes sont référencés ont été retenues, ce qui représente 119 salles. Le nombre de clusters $k$ pour chaque type de variable a été défini selon la méthode empirique du coude (\emph{elbow method} en anglais). \par

Le clustering des salles en fonction des données acoustiques (acousticness, danceability, duration, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence), avec $k=7$, semble plutôt concluant puisque dans chaque cluster on retrouve des salles relativement homogènes que l’on peut catégoriser ainsi :

\begin{itemize}

\item Salles historiques des années 1980 et 1990 : Ebullition à Bulle, Fri-Son à Fribourg, Rote Fabrik à Zurich.

\item Salles des arts classiques :Victoria Hall à Genève, KKL à Lucerne.

\item Discothèques organisant des concerts : D! Club à Lausanne, Case à Chocs à Neuchâtel, Zukunft à Zurich.

\item Discothèques plutôt avec des DJ : Borderline à Bâle, Hive à Zurich.

\item Salles rock tendance hard ou métal : Caves du Manoir à Martigny, Ebrietas à Zurich, Horst Klub à Kreuzlingen.

\item Salles grand public : Aréna à Genève, Messe à Bâle.

\item Salles éclectiques : Alhambra à Genève, Bad Bonn à Guin.

\end{itemize}

Si l’on prend en compte les variables de popularité, définies par le nombre d’auditeurs mensuels et le nombre de followers répertoriés sur Spotify, le clustering, avec $k = 4$, fait apparaître quatre catégories de salles qui peuvent être différenciées par leur capacité d’accueil :

\begin{itemize}

\item Salle unique : le Hallenstadion à Zurich. Il s’agit de la plus grande salle de Suisse, pouvant accueillir 15'000 personnes.

\item Grands complexes : Aréna à Genève, Halle St-Jacques à Bâle, Auditorium Stravinski à Montreux. Ces salles ont une capacité d’accueil de plusieurs milliers de personnes.

\item Salles moyennes : les Docks à Lausanne, le Théâtre du Léman à Genève, le X-TRA à Zurich. Ces salles sont dotées d’environ 1'000 places. Il manque toutefois quelques salles dans ce cluster qui pourrait faire partie de cette catégorie, comme Fri-Son à Fribourg ou le KKL à Lucerne.

\item Salles plus petites. L’immense majorité des salles se trouvent dans cette catégorie.

\end{itemize}

L’application de l’algorithme du k-means en fonction des variables spatiales et temporelles, avec $k=3$ pour les deux, n’a pas donné des clusters qui permettent de catégoriser les salles. Ce partitionnement des données met donc en évidence la capacité des variables acoustiques et de popularité de mieux saisir la diversité des salles de concert en Suisse.

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

Pour mieux comprendre ce que les données récoltées représentent au sein de la scène musicale, des outils de visualisation montrent comment les concerts et les salles se répartissent sur le territoire helvétique, comment ces lieux de concert se distinguent les uns des autres et enfin comment chaque salle est définie par les artistes qu’elle accueille.

**\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, en cartographiant soit les genres musicaux, soit 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 top 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 d’apporter une dimension spatiale aux données musicales, à travers les genres prédominants dans les salles, et aux données événementielles, en montrant tous 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

\begin{figure}[ht]

\centering

\includegraphics[width=0.8\textwidth]{images/resultats/viz\_genre\_distrib.png}

\captionsetup{width=.8\textwidth}

\caption{Visualisation de la distribution de la fréquence des genres musicaux dans les salles de L’Usine à Genève et du Marian’s Jazz Club à Berne. On remarque que le rock, la musique électro et les musiques du monde prédominent à L’Usine, où 60\% des artistes sont apparentés au rock. Le jazz, les musiques classiques et la pop sortent du lot au Marian’s Jazz Club, où plus de la moitié des artistes sont apparentés au jazz.}

\label{fig\_viz\_genre\_distrib}

\end{figure}

Dans la continuité des données mises en valeur dans l’outil de cartographie, un diagramme en barres (cf. figure \ref{fig\_viz\_genre\_distrib}) permet de comparer la fréquence de chacun des 23 top genres entre plusieurs salles sélectionnées au préalable. Etant donné que des artistes peuvent être associés à plusieurs genres musicaux, la somme des fréquences des genres d’une salle peut être supérieures à 1. Grâce à cet outil, il est ainsi possible de voir avec plus de précision comment des salles repérées sur la carte se distinguent au niveau des genres. \par

\begin{figure}[ht]

\centering

\includegraphics[width=0.8\textwidth]{images/resultats/viz\_venues\_scatter.png}

\captionsetup{width=.8\textwidth}

\caption{Visualisation des salles de concert en fonction de deux variables acoustiques, danceability et acousticness.}

\label{fig\_viz\_venues\_scatter}

\end{figure}

La comparaison entre les salles n’est pas seulement réalisable avec les données relatives aux genres, mais également avec les données acoustiques, spatiales, temporelles et de popularité. Un diagramme en nuage de points (\emph{scatter plot} en anglais) met en évidence comment les salles se distinguent les unes des autres en fonction des artistes qui y jouent (cf. figure \ref{fig\_viz\_venues\_scatter}). Chaque point représente une salle unique dont la position sur le graphique est définie par deux variables sélectionnées au préalable parmi les variables à disposition. Les valeurs des variables ont été établies pour chaque salle en faisant la moyenne de ces variables de tous les artistes ayant fréquenté cette salle. Comme pour la cartographie, il est possible de diminuer l’influence de la distribution inégale de l’activité des salles en filtrant le nombre de concerts minimum organisés par salle. L’algorithme de partitionnement \emph{k-means} est appliqué sur les données du diagramme pour voir, en fonction des variables comparées, quelles salles partagent des similarités ou à l’inverse, qu’est-ce qui distingue une salle d’une autre. Comme nous l’avons vu au point \ref{section\_kmeans}, les salles que l’on trouve dans les clusters ne sont pas immuables et les particularités des salles apparaissent en altérant les variables de comparaison. Par exemple, des salles qui ont a priori peu en commun, comme le Victoria Hall de Genève, une salle majestueuse habituée aux arts classiques, et les Caves du Manoir à Martigny, une salle sous-terraine consacrée au rock, partagent des similitudes sur certains aspects acoustiques, comme la danceability, mais divergent sur d’autres variables, mettant en avant la diversité des salles de concert. \par

\begin{figure}[ht]

\centering

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

\centering

\includegraphics[width=\textwidth]{images/resultats/viz\_venue\_concerts\_scatter.png}

\caption{Visualisation des concerts. La teinte des points varie du bleu au jaune plus un concert compte d’artistes.}

\end{subfigure}

\hfill

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

\centering

\includegraphics[width=\textwidth]{images/resultats/viz\_venue\_artists\_scatter.png}

\caption{Visualisation des artistes. La teinte des points varie du bleu au jaune plus un artiste a joué de concerts dans ce lieu.}

\end{subfigure}

\captionsetup{width=.8\textwidth}

\caption{Visualisation des concerts et des artistes qui ont joué dans la salle du Romandie à Lausanne, en fonction des variables acoustiques de danceability et d’acousticness.}

\label{fig\_individual\_venue\_scatter}

\end{figure}

Les différentes variables peuvent donc aider à comprendre les variations d’une salle à l’autre, mais elles peuvent aussi rendre compte de la diversité au sein même des salles, grâce à un nuage de points qui montre comment les artistes définissent un lieu et comment ils se distinguent entre eux (cf. figure \ref{fig\_individual\_venue\_scatter}). Comme pour les comparaisons entre les salles, il est nécessaire de sélectionner deux variables pour observer leur répartition parmi les individus d’une salle. Ces individus justement sont représentés de deux manières différentes : les concerts et les artistes. Bien que généralement les artistes d’un même concert partagent de nombreuses similarités (dans l’idée de proposer un événement homogène au public), les valeurs des variables associées à ces artistes peuvent diverger. Deux nuages de données proposent donc de montrer pour le premier les données associées aux concerts et le second montre les artistes individuellement. Les points représentant les concerts sont obtenus en faisant la moyenne des valeurs des artistes d’un concert. Il est ainsi possible d’obtenir une vue globale des concerts et une vue précise des artistes dans chaque salle. \par

Ainsi, les données relatives aux concerts en Suisse collectées sur Songkick, Spotify et Wikidata mettent en évidence une distribution inégale du nombre de concerts organisés par salle, suggérant un poids important d’une minorité de salles dans la scène musicale, des corrélations entre certaines des variables acoustiques, pouvant servir de point de départ pour comparer des salles de concert, et un partitionnement pertinent des salles de concert selon certaines variables. La visualisation de ces données de son côté montre la diversité et la concentration spatiale des artistes et des genres grâce à l’outil de cartographie, les divergences dans la programmation des salles en comparant la fréquence des genres, les caractéristiques qui définissent les salles entre elles grâce au nuage de points et enfin la diversité trouée au sein de chaque salle.