

Chapitre #2 – Analyse et visualisation des données

Le présent chapitre propose une démarche progressive d'exploitation de données en quatre parties.

1. Création d'un dictionnaire de données et exploration d'un jeu de données à l'aide de statistiques descriptives et de visualisations simples ;
2. Utilisation de la fonction Tableau croisé dynamique d'Excel pour l'exploration et l'analyse des données ;
3. Calcul de nouveaux indicateurs, à partir du jeu de données, et croisement avec d'autres variables pour répondre à des besoins d'affaires ;
4. Construction d'un tableau de bord dynamique combinant les différents indicateurs.

Liste de contrôle pour les gestionnaires de données

Cette liste comprend une série de questions permettant de vous préparer lors de l'analyse des données de votre projet. Il s'agit en quelque sorte d'un aide-mémoire en lien avec le contenu de cette partie.

Avant l'analyse de données

- **D'où viennent vos données ? Est-ce que leurs sources sont fiables ?**
Ne faire des analyses que si vous êtes certains de la qualité des données
- **Vos données ont-elles été nettoyées conformément aux bonnes pratiques (corrections des erreurs de saisies, identification des doublons, choix des bons formats, etc.) ?**
Si la réponse est non ou « je ne sais pas », consultez le Chapitre 1 intitulé « La collecte de données et l'évaluation de leur qualité » pour apprendre comment les nettoyer.
- **Avez-vous un dictionnaire de données ?**
Celui-ci simplifie l'analyse en rendant intelligible votre jeu de données et rend plus aisés le travail collaboratif. Vous pouvez consulter la section « Pourquoi concevoir un dictionnaire de données ? » de cette partie pour découvrir ce qu'est un dictionnaire de données et comment en faire un bon usage.

Pendant l'analyse

- **Avez-vous précisé la question à laquelle vous tentez de répondre et identifié les informations qui permettront d'y répondre ?**
Connaître en amont les questions qui vous préoccupent permet de diriger le travail d'analyse.
- **Avez-vous défini vos indicateurs clés de performance (KPI) ? Sont-ils :**
 - Spécifiques : définis en termes clairs ;
 - Mesurables : permettent d'évaluer le résultat ;
 - Utiles : permettent de prendre le pouls du projet, de répondre à un besoin d'information et de vous soutenir dans vos prises de décisions ;
 - Simples : se mesurent facilement et sont compris par l'ensemble des acteurs concernés ?

Pour concevoir un tableau de bord

Connaissez-vous les réponses aux questions suivantes :

- **À quoi servira le tableau de bord ? À quelle(s) question(s) répondra-t-il ?**
- **À quelle fréquence sera-t-il consulté (quotidiennement, mois, trimestre, ponctuellement, etc.) ?**
- **Quelles personnes (cibles) consulteront le tableau de bord ?**
- **Quels sont les indicateurs les plus importants pour le travail de ces personnes ?**

Le jeu de données

Tout au long de cette partie, nous nous référerons à un jeu de données fictives, « **jeudedonnees.xlsx** ». En téléchargeant ce fichier, vous pourrez l'explorer et reproduire les différents techniques et analyses qui sont présentées dans les prochaines sections. Ce jeu de données fictives regroupe l'ensemble des transactions effectuées par les clients d'un festival multidisciplinaire annuel ayant lieu au début du mois de juillet dans la région de Montréal. Ce jeu rassemble plus de 10.000 transactions de billetterie effectuées par un peu plus de 6.000 clients différents entre 2016 et 2020.

Les fichiers Excel

Dans le fichier « **jeudedonnees.xlsx** », vous trouverez plusieurs feuilles pertinentes pour la réalisation des différentes étapes de cette partie. D'abord, la feuille « données initiales » présente le jeu de données de départ.

Le dictionnaire de données (voir ci-dessous) est exposé dans la feuille du même nom tandis que les feuilles qui suivent présentent, une à une, les différentes étapes de transformation et d'analyse des données. Un second fichier Excel, « **tableaudebord.xlsx** », propose différentes feuilles menant à la création d'un tableau de bord (Partie 4 du présent chapitre). Dans ces fichiers, le jeu de données a été transformé en format « Tableau ». Cette mise en forme a l'avantage qu'Excel reconnaît rapidement l'ensemble de votre jeu de données simplement en plaçant votre souris dans une seule cellule, ce qui sera fort utile lors de la réalisation de tableaux croisés dynamiques. Il est maintenant temps d'ouvrir le fichier Excel « **jeudedonnees.xlsx** ».

	A	B	C	D	E	F
1	ID_Client	ID_Transaction	Date d'achat	Date de l'événement	DélaiAchatEnJour	Quantité
2	Cli_1	Trans_425221347	2016-07-09	2016-07-09	0	2
3	Cli_10	Trans_434671200	2019-04-06	2019-07-02	87	1
4	Cli_10	Trans_427441115	2017-07-02	2017-07-02	0	2
5	Cli_10	Trans_425231544	2016-07-11	2016-07-11	0	3
6	Cli_100	Trans_434841358	2019-04-02	2019-07-03	92	1

Partie 1. Ne pas négliger les premières étapes

Pourquoi concevoir un dictionnaire de données ?

Lorsque l'on développe un jeu de données, peu importe le nombre d'utilisateurs, il est important de créer une documentation précise du contenu : ce qu'on appelle un dictionnaire de données.

Plus précisément, un dictionnaire de données est « un répertoire de métadonnées » qui sert à présenter et à décrire le contenu d'une base ou d'un jeu de données¹.

C'est donc la liste de toutes les variables qui se trouvent dans le jeu de données avec un descriptif, permettant ainsi de s'y retrouver.

Un dictionnaire de données contient généralement :

- Le nom des champs (ou variables) contenus dans le jeu de donnée (ex. : Type de client) ;
- La description des champs (ex. : type de client peut référer à si celui-ci est abonné ou non) ;
- Le format des données des différents champs. Les principaux formats observés sont :
 - Nombre entier ;
 - Nombre décimal ;
 - Format texte composé de caractères (lettres, ponctuation, combinaison de lettres et de nombres (ex. code postal) ;
 - Date.

Dans notre exemple, le champ *type de client* est en format texte.

Le dictionnaire de données peut également contenir des métadonnées, c'est-à-dire des données « qui renseignent sur la nature de certaines autres données et qui permettent ainsi leur utilisation pertinente »². Ces métadonnées peuvent concerner :

- La source des données de chaque champ (billetterie interne, site Web, Facebook, etc.) ;
- La relation entre les champs, lorsqu'il s'agit d'une base de données relationnelle³ ;
- Les valeurs par défaut, s'il y a lieu ;
- L'étendue des valeurs possibles (ex. : le type de client n'a que deux valeurs possibles, soit « abonné » ou « occasionnel ») ;
- Toute autre notion pertinente pour les analyses.

¹ Traduction libre de : Bourne, K. C. (2013). *Application administrators handbook: installing, updating and troubleshooting software*. Newnes. Repéré ici : [Data Dictionary - an overview | ScienceDirect Topics](#)

² Définition de l'OQLF, [Métadonnée](#) (page consultée en février 2022)

³ D'après [Oracle](#), une base de données relationnelle « est un type de base de données où les données sont liées à d'autres informations au sein des bases de données. Les bases de données relationnelles sont composées d'un ensemble de tables qui peuvent être accessibles et reconstruites de différentes manières, sans qu'il soit nécessaire de réarranger ces tables de quelque façon que ce soit. » (page consultée en février 2022)

Le **Tableau 1**, ci-dessous, représente le dictionnaire de données pour le jeu de données fictives. Vous trouverez ce même dictionnaire dans le fichier Excel sur la feuille « dictionnaire de données ».

Grâce à ce dictionnaire, nous savons maintenant que la variable « type de spectacle » réfère à la discipline artistique du contenu du spectacle (humour, danse, opéra, théâtre ou musique) et non pas, par exemple, au fait que ce soit un événement intérieur ou extérieur ou encore tenu en mode virtuel.

Nous constatons également, dans le même tableau, que notre jeu de données compte deux colonnes associées aux identifiants. Toutefois, une seule de ses colonnes ne contient que des identifiants uniques⁴, c'est-à-dire qu'elle ne contient aucune valeur en double.



À vous de jouer !

Sans regarder le fichier Excel, identifiez, dans le dictionnaire de données, la colonne qui contient que les identifiants uniques ? Pourquoi avoir un autre champ aussi appelé identifiant dans le jeu de données ?

Tableau 1. Dictionnaire de données pour le jeu de données fictives

Noms de champs	Définition	Format
ID_Client	Numéro d'identification du client (ou Identifiant client)	Texte
ID_Transaction	Numéro d'identification unique de la transaction (ou Identifiant transaction)	Texte
Date d'achat	Date à laquelle la transaction a été effectuée	Date
Date de l'événement	Date du spectacle pour lequel a été une transaction	Date
Quantité	Quantité de billets achetés lors de la transaction	Numérique
Prix	Prix d'un seul billet	Numérique
Type de client	Transaction effectuée par un occasionnel ou un abonné	Texte
Code postal	Code postal du client	Texte
RTA	3 premiers caractères du code postal	Texte
Canal d'achat	Par guichet, Internet ou téléphone	Texte
Nom événement	Identifiant donné au spectacle (ici, il s'agit d'un nombre entre 1 et 10 000)	Texte
Type de spectacle	Discipline artistique (musique, danse, théâtre, humour ou opéra)	Texte
Nom de la scène	Nom donné à la scène sur laquelle se déroule le spectacle (ici, il s'agit d'un nombre de 1 à 5)	Texte
Capacité de la scène	Nombre de spectateurs maximal	Nombre entier

L'exploration des données : une démarche importante

Avant de se lancer dans des analyses complexes, il est utile d'explorer son jeu de données à l'aide de statistiques descriptives ou de visualisations simples. Cette étape d'exploration permet de se familiariser à notre jeu de données et à ses caractéristiques, ainsi qu'à réfléchir

⁴ Un identifiant unique est une chaîne de caractères numérique ou alphanumérique associé à une seul entité dans une base de données. Le numéro d'assurance social ou le numéro de permis de conduire sont des exemples d'identifiants uniques.

aux analyses possibles. Pour l'exploration, nous pouvons, par exemple, utiliser des fonctions arithmétiques simples qui se trouvent sous l'onglet « formules » comme la moyenne et l'étendue de nos variables numériques (voir **Tableau 2**). Vous trouverez ce même tableau dans le fichier Excel sous l'onglet « Statistiques descriptives ».

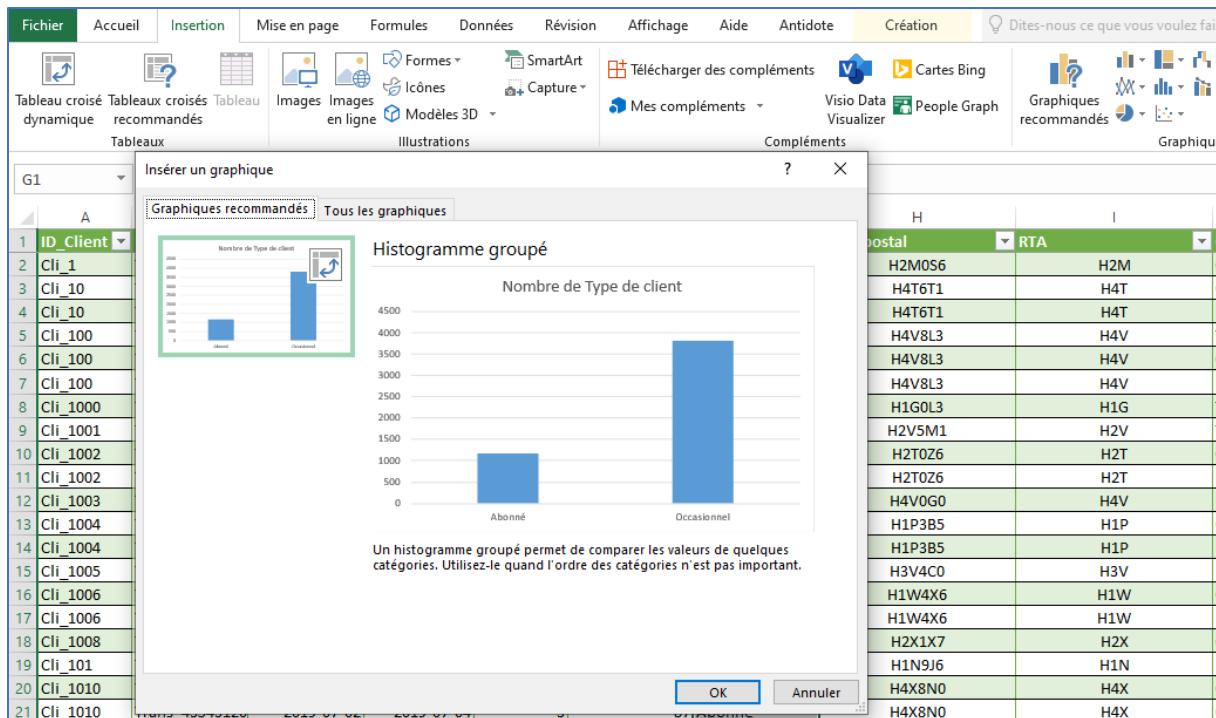
Tableau 2. Statistiques descriptives pour les variables numériques.

Quantité	2,71	1	68
Prix	59,99	20	100
Formule	=Moyenne	=Min	=Max

Il est aussi pertinent de créer quelques visualisations simples, comme des histogrammes. En effet, la visualisation est une façon de synthétiser l'information et d'illustrer rapidement ce que renferment les données. Il suffit d'utiliser la fonction Graphiques recommandés dans Excel après avoir sélectionné la colonne du jeu de données à partir de laquelle nous voulons créer un graphique et de choisir la visualisation qui nous plaît.

Dans notre exemple, cette étape permet rapidement de voir que les transactions réalisées par des clients occasionnels sont environ trois fois plus nombreuses que celles effectuées par des abonnés.

	A	B	C	D	E	F	G	H	I
1	ID_Client	ID_Transaction	Date d'achat	Date de l'év.	Quantité	Prix	Type de client	Code postal	RTA
2	Cli_1	Trans_42522134	2016-07-09	2016-07-09	2	26	Occasionnel	H2MOS6	H2M
3	Cli_10	Trans_42523154	2016-07-11	2016-07-11	3	52	Occasionnel	H4T6T1	H4T
4	Cli_10	Trans_42744111	2017-07-02	2017-07-02	2	50	Occasionnel	H4T6T1	H4T
5	Cli_100	Trans_43484135	2019-04-02	2019-07-03	1	89	Occasionnel	H4V8L3	H4V



À vous de jouer !

Testez quelques graphiques à partir d'autres variables du jeu de données comme le canal d'achat ou le type de spectacle. Quelles conclusions pouvez-vous en tirer ?

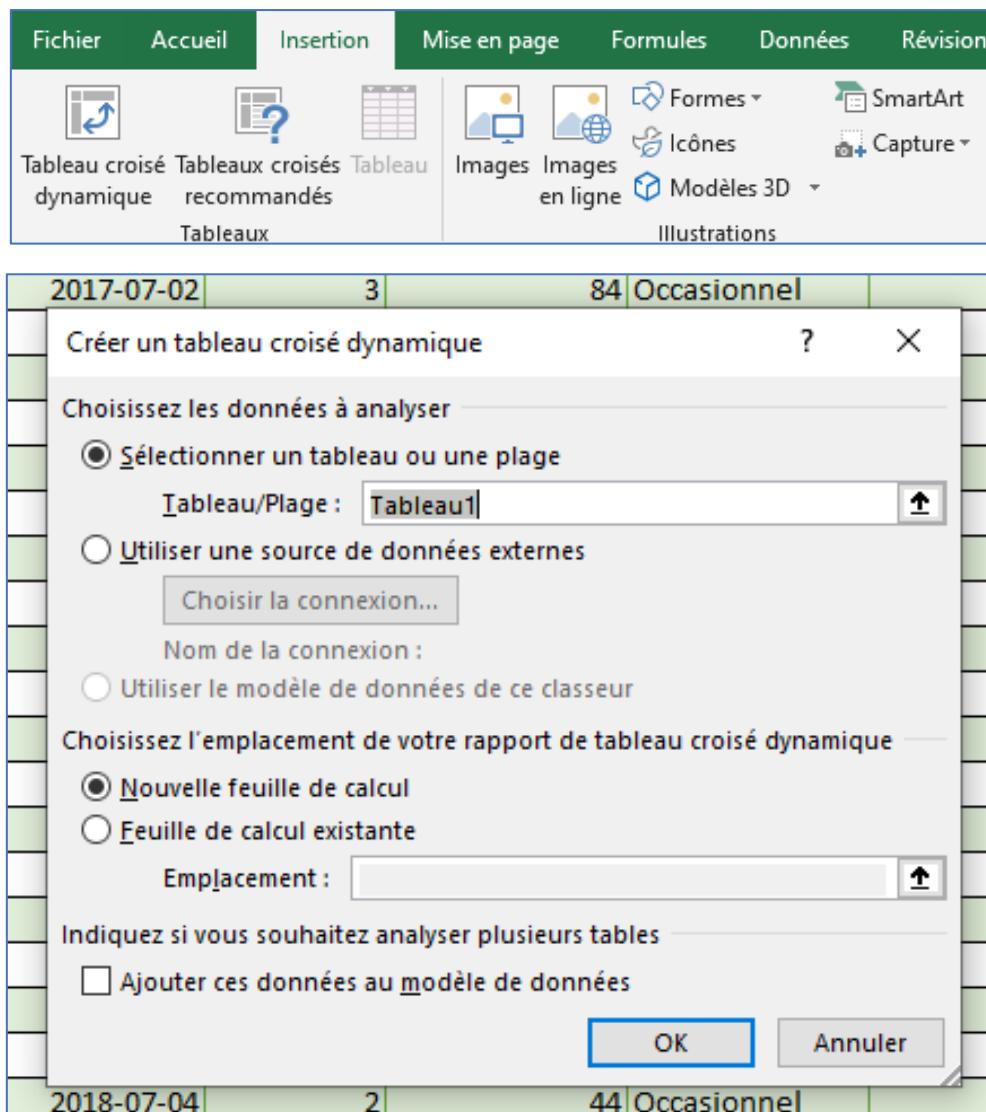
Partie 2. Explorer et analyser les données à l'aide des tableaux croisés dynamiques d'Excel

Pour tirer plus d'information du jeu de données, par exemple sur l'évolution du nombre d'abonnés par année, nous devons utiliser une autre fonctionnalité d'Excel, soit le tableau croisé dynamique. En effet, cette fonctionnalité permet de croiser des variables, d'analyser des données et de déceler des tendances en quelques étapes. En un seul clic, vous obtenez automatiquement des tableaux synthèses de votre jeu de données.

Savoir utiliser les tableaux croisés dynamiques en 5 étapes

Avant de créer un tableau croisé, il faut s'assurer que chaque colonne du jeu de données ait un titre.

Pour créer un tableau croisé, il faut d'abord sélectionner une cellule de votre jeu de données pour qu'Excel le reconnaisse automatiquement lors de la création du tableau croisé dynamique. Excel demandera ensuite de choisir l'emplacement du tableau croisé. Il est préférable de créer le tableau croisé dynamique sur une nouvelle feuille.



Étape 1 : Préparation du tableau croisé dynamique

Pour utiliser le tableau croisé, nous devons nous diriger vers la nouvelle feuille où il a été créé et sélectionner une cellule dans la zone spécifique du tableau croisé. Les différents champs du jeu de données initial vont ainsi apparaître à droite dans une nouvelle fenêtre de formatage. Cela permettra de glisser des champs dans les cases *Colonnes*, *Lignes*, *Valeurs* et *Filtres*, pour commencer à expérimenter avec le jeu de données.

Commençons notre initiation avec une des variables de type « Date ». Avec un jeu de données transactionnelles comptant un nombre important de dates différentes, comme c'est le cas pour celui avec lequel nous travaillons, il est préférable de commencer par regrouper les transactions en un nombre prédéterminé de périodes. Pour ce faire, il suffit d'insérer le champ « Date de l'événement » en tant que « *Lignes* » dans votre tableau et de cliquer sur une des dates de la feuille pour avoir accès à la fonction « *Grouper la sélection* ». Nous pouvons ensuite sélectionner le grain désiré (ici, nous choisissons les années, mais il est possible de travailler par semaine, par mois, etc.).

The screenshot illustrates the preparation of a PivotTable. The Excel ribbon is visible at the top, with the 'Analyse' tab selected. A 'Grouper' dialog box is displayed, allowing the user to group data by year ('Années') from July 2016 to July 2020. To the right, the 'Champs de tableau crois...' (PivotTable Fields) window is open, showing a list of available fields including 'ID_Client', 'ID_Transaction', 'Date d'achat', and 'Date de l'événement'. The 'Date de l'événement' field is checked, indicating it is included in the report. The PivotTable itself is partially visible on the left.

Il est d'ailleurs possible dans le tableau croisé dynamique d'utiliser différents grains de données dans le même tableau. C'est le cas de cet exemple qui regroupe des données par année, mais offre aussi la possibilité de déplier la vente de billet mensuelle (Vous trouverez ce tableau croisé sur la feuille « Exemple tableau croisé » du fichier Excel).

Somme de Quantité				
	Guichet	Internet	Téléphone	Total général
2016	1476	298	3	1777
2017	4462	1520	602	6584
janv		156	104	260
févr		93	65	158
mars		87	104	191
avr		129	121	250
mai		109	103	212
juin		101	105	206
juil	4462	845		5307
2018	400	5684	3467	9551
2019	234	6207	2782	9223
2020		243	108	351
Total général	6572	13952	6962	27486

Champs de tableau crois...

Choisissez les champs à inclure dans le rapport :

Rechercher

ID_Client
 ID_Transaction
 Date d'achat
 Date de l'événement
 Quantité
 Prix
 Type de client
 Code postal
 RTA
 Canal d'achat
 Nom événement
 Type de spectacle
 Nom de la scène
 Capacité de la scène
 Années

Faites glisser les champs dans les zones voulues ci-dessous:

Filtres	Colonnes
	Canal d'achat
Lignes	Valeurs
Années	Somme de Quantité
Date d'achat	

Étape 2 : Croisements de bases et choix des variables

Pour commencer les analyses, glissons différentes variables dans la case « Valeurs » pour faire des observations et des explorations.

Nombre de ID_Transaction		Somme de Quantité	
2016	510	2016	1777
2017	2513	2017	6584
2018	3578	2018	9551
2019	3433	2019	9223
2020	113	2020	351
Total général	10147	Total général	27486

vs

Dans cet exemple, nous remarquons que la variable « ID_transaction », qui ne décrit que l'existence d'une transaction, offre un portrait incomplet du nombre de festivaliers. En contrepartie, la variable « Quantité », qui décrit le nombre de billets par transaction, offre ce chiffre précis. Dans ce cas-ci, il est préférable d'analyser la « Somme de Quantité » pour avoir une meilleure analyse de l'évolution des ventes.

Pour poursuivre l'exploration des données, glissons maintenant une variable qualitative⁵ dans la case « Colonnes » du tableau croisé. C'est la force du tableau croisé : il permet rapidement de croiser presque toutes les variables ensemble.

Somme de Quantité	Danse	Humour	Musique	Opéra	Théâtre	Total général
2016	373	731	358	207	108	1777
2017	1939	1436	2622	167	420	6584
2018	1896	2766	3867	1022		9551
2019	2342	3068	2717	612	484	9223
2020	157	74	120			351
Total général	6707	8075	9684	2008	1012	27486

À vous de jouer !

À l'aide du tableau croisé de la feuille « Tableau croisé exercice », essayez de croiser la variable « date d'achat » avec le « type de client ». Que constatez-vous ?

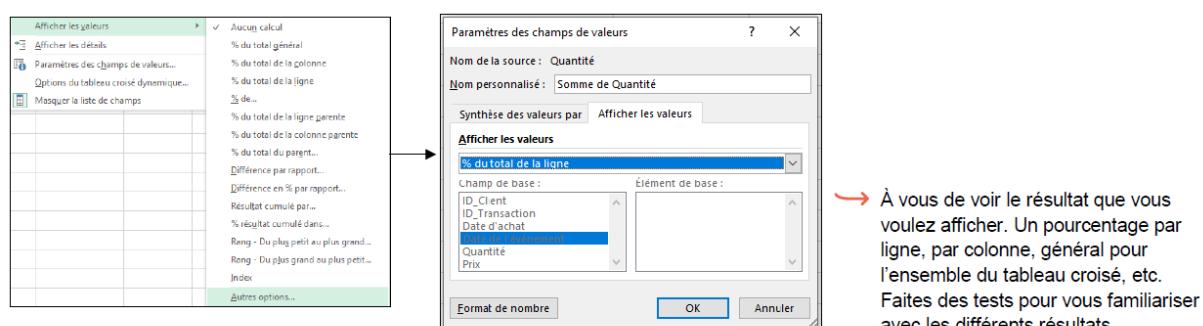
⁵ Une variable qualitative est une variable qui prend pour valeur des catégories, des niveaux ou des modalités (dans ce cas-ci, la discipline artistique du spectacle). Elle ne mesure donc pas de quantité, contrairement aux variables quantitatives.

Étape 3 : Le tableau croisé comme calculatrice

Parfois, de simples nombres ne permettent pas de dégager facilement des tendances ou de faire une lecture rapide de la situation. Il est préférable, pour mieux présenter l'information, d'afficher les résultats sous un autre format, dont le pourcentage.

	Guichet	Internet	Téléphone	Total général
2016	83,06%	16,77%	0,17%	100,00%
2017	67,77%	23,09%	9,14%	100,00%
2018	4,19%	59,51%	36,30%	100,00%
2019	2,54%	67,30%	30,16%	100,00%
2020	0,00%	69,23%	30,77%	100,00%
Total général	23,91%	50,76%	25,33%	100,00%

Pour obtenir ce genre de tableau, il suffit d'utiliser le tableau croisé comme une calculatrice : avec un clic droit sur une valeur du tableau et le choix de l'onglet afficher les valeurs, on obtient alors la boîte suivante (*Paramètres des champs de valeurs*) qui donne accès à une panoplie de solution arithmétique.



Étape 4 : Visualiser par année à l'aide des chronologies

Avec des données de type date, comme c'est le cas du jeu de données fictives, il devient intéressant d'ajouter un filtre temporel pour analyser les données sur certaines périodes précises. Pour ce faire, il faut simplement utiliser la fonction Insérer une chronologie. On obtient ainsi un filtre rapide et intuitif qui permet d'isoler les données d'une période précise en un seul clic (pour un exemple, voir la feuille « Tableau croisé chronologie » du fichier Excel).

Étape 5 : La mise en forme conditionnelle, un bon outil pour dégager des constats

Lorsqu'il est difficile de visualiser à l'œil nu les écarts entre les différentes cases d'un tableau, la mise en forme conditionnelle s'avère très utile.

La mise en forme conditionnelle offre en effet plusieurs façons de mettre de l'avant certaines valeurs selon des critères précis. On peut par exemple colorer les cases avec les 10 valeurs les plus élevées. Avec l'option *Nuances de couleurs*, Excel peut également créer un gradient de couleurs qui s'adapte en fonction des valeurs du tableau. Dans cet exemple, il devient donc plus facile de constater que plus de billets pour des spectacles d'humour que pour des concerts musicaux ont été vendus en 2016, mais que cette tendance s'est inversée l'année suivante (2017).

The screenshot shows an Excel spreadsheet titled "Données_formation_AvecIndicateurs_EAQv3.xlsx - Excel". The ribbon is visible at the top with tabs like Accueil, Insertion, Mise en page, Formules, Données, Révision, Affichage, Aide, Antidote, Analyse, and Création. The "Outils de tableau croisé dynamique" tab is selected. The main area displays a table with columns labeled "Danse", "Humour", "Musique", "Opéra", "Théâtre", and "Total général". Rows represent years from 2016 to 2020. The "Nuances de couleurs" option under the conditional formatting dropdown is highlighted. The table data is as follows:

	Danse	Humour	Musique	Opéra	Théâtre	Total général
2016	20,99%	41,14%	20,15%	11,65%	6,08%	100,00%
2017	29,45%	21,81%	39,82%	2,54%	6,38%	100,00%
2018	19,85%	28,96%	40,49%	10,70%	0,00%	100,00%
2019	25,39%	33,26%	29,46%	6,64%	5,25%	100,00%
2020	44,73%	21,08%	34,19%	0,00%	0,00%	100,00%
Total général	24,40%	29,38%	35,23%	7,31%	3,68%	100,00%

Somme de Quantité						
	Danse	Humour	Musique	Opéra	Théâtre	Total général
2016	20,99%	41,14%	20,15%	11,65%	6,08%	100,00%
2017	29,45%	21,81%	39,82%	2,54%	6,38%	100,00%
2018	19,85%	28,96%	40,49%	10,70%	0,00%	100,00%
2019	25,39%	33,26%	29,46%	6,64%	5,25%	100,00%
2020	44,73%	21,08%	34,19%	0,00%	0,00%	100,00%
Total général	24,40%	29,38%	35,23%	7,31%	3,68%	100,00%

À vous de jouer !

Sur la feuille « Tableau croisé exercice » du fichier Excel, essayez différentes nuances de couleurs pour mieux explorer le tableau. Par la suite, utilisez les règles de valeurs de plage pour identifier les cinq valeurs les plus élevées : que constatez-vous ?

Partie 3. Prolonger l'analyse à l'aide d'indicateurs intelligents

Dans les sections précédentes, nous avons vu comment explorer les données, sans transformer nos variables, de façon à obtenir un premier aperçu des informations qui s'y cachent. Or, le jeu de données ne permet pas encore de répondre à des questions plus complexes comme :

- « **Quelle(s) discipline(s) artistique(s) attirent le plus de visiteurs de l'extérieur de Montréal ?** » ;
- « **Quelle est l'évolution du taux de remplissage moyen des différents spectacles par année ?** ».

Pour répondre à ces questions, lesquelles peuvent correspondre à des besoins d'information stratégique, il est nécessaire de créer de nouvelles variables à partir des données. Ces nouvelles variables permettront ensuite de répondre aux indicateurs intelligents choisis.

Qu'est-ce qu'un indicateur et comment le choisir ?

Un indicateur⁶ est une mesure quantitative utilisée pour l'aide à la décision. Quatre principaux critères de qualité décrivent un bon indicateur :

- Il est **mesurable**, c'est-à-dire qu'on peut évaluer le résultat ;
- Il est **utile**, car il permet de prendre le pouls de l'organisation, répond à un besoin d'information et soutient la prise de décisions ;
- Il est **spécifique**, c'est-à-dire qu'il est défini en termes clairs ;
- Il est **simple**, car il se mesure facilement et est compris par l'ensemble des acteurs concernés.

Ajoutons qu'un bon indicateur peut être suivi dans le temps, permettant ainsi de savoir si nos objectifs ont été atteints et d'identifier ceux sur lesquels il importe d'agir.

Dans le cas de notre festival fictif, plusieurs indicateurs pertinents peuvent être définis pour avoir un meilleur portrait de la situation ou encore pour créer un tableau de bord (voir Partie 4) :

1. Total des revenus du festival par année ou selon le type de spectacle ;
2. Taux (%) de visiteurs provenant de l'extérieur de Montréal dans les spectacles ;
3. Proportion des clients qui sont des abonnés du festival ;
4. Proportion des transactions réalisées en ligne ;
5. Fréquence des achats des différents clients (le nombre total de transactions par client) ;
6. Délai d'achat moyen (le nombre de jours avant l'événement lorsque le billet a été acheté à l'avance) ;
7. Taux de remplissage des différents spectacles.

⁶ Les termes « indicateurs clés de performance » (ICP) ou *Key Performance Indicators* (KPI) sont également des termes souvent employés.

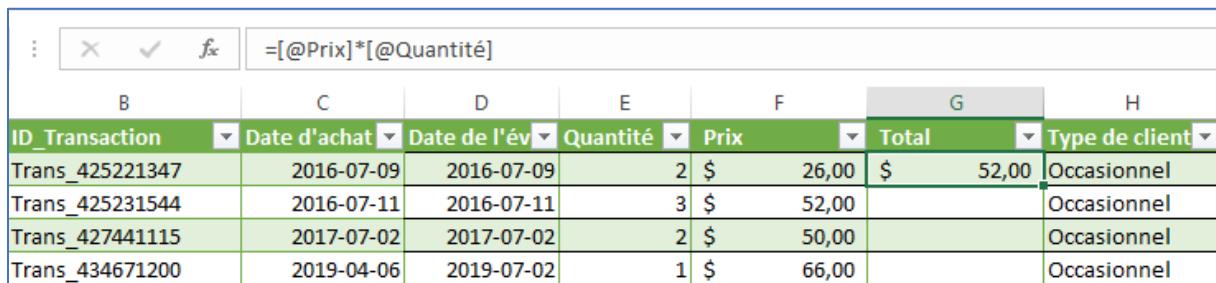
Si certains des indicateurs ci-dessus peuvent déjà être calculés rapidement à l'aide des méthodes vues dans la Partie 2 (tableau croisé), comme c'est le cas pour la proportion des transactions réalisées en ligne, d'autres nécessitent la création de nouvelles variables. Heureusement, toutes ces nouvelles variables peuvent facilement être calculées dans Excel, soit à l'aide de formules arithmétiques simples, soit à l'aide de fonctions conditionnelles, ou soit à l'aide d'un tableau croisé dynamique.

Création de nouvelles variables dans Excel

La création de variables à l'aide d'opérations arithmétiques simples

Pour créer de nouvelles variables dans Excel, lorsque le jeu de données est en format « Tableau », il faut tout simplement choisir la première cellule d'une colonne vierge pour créer la nouvelle variable et d'insérer⁷ le calcul dans la barre de formule 

Par exemple, pour obtenir le total des revenus de chaque transaction, il suffira de multiplier les valeurs contenues dans les colonnes « Prix » et « Quantité » (voir feuille « Création 1 (total) » du fichier Excel). Ensuite, pour recopier la formule dans les cellules sous-jacentes, un double-clic dans le coin du bas à droite de la première cellule est suffisant. La colonne se complètera par automatiquement. Cette nouvelle colonne ainsi obtenue pourra être utilisée dans un tableau croisé pour construire ou calculer le résultat de l'indicateur de notre choix (par exemple, les revenus totaux par année).



B	C	D	E	F	G	H
ID_Transaction	Date d'achat	Date de l'év	Quantité	Prix	Total	Type de client
Trans_425221347	2016-07-09	2016-07-09	2	\$ 26,00	\$ 52,00	Occasionnel
Trans_425231544	2016-07-11	2016-07-11	3	\$ 52,00		Occasionnel
Trans_427441115	2017-07-02	2017-07-02	2	\$ 50,00		Occasionnel
Trans_434671200	2019-04-06	2019-07-02	1	\$ 66,00		Occasionnel

L'utilisation d'opérateurs logiques dans une formule

Dans certains cas, la création d'une variable précise requiert l'utilisation d'opérateurs logiques tels que « si », « ou », « et », etc. Dans l'exemple, les opérateurs ont été utilisés pour classer en deux groupes distincts les spectateurs selon leur provenance. Cette nouvelle variable sera utile pour calculer le nombre de transactions par clients et pour définir le nombre total de billets vendus par spectacle.

⁷ Lorsque l'on n'utilise pas le format « tableau », la procédure est similaire, mais nous devons indiquer à Excel les coordonnées des cellules qui nous intéressent (ex. : E4, G6, etc.)

Tableau 3. Les nouvelles variables créées directement dans la feuille « Données initiales » et la méthode employée pour le faire

Nouvelles variables créées	Définition	Format	Dans le même tableau	Méthode (formule utilisée dans Excel)
Revenu transaction	Sous-total de la transaction	Numérique	Oui	=[@Prix]*[@Quantité]
Provenance	Client de Montréal (local) ou non (extérieur)	Texte	Oui	=SI(OU(GAUCHE([@][Code postal]];2)="H1"; GAUCHE([@][Code postal]];2)="H2"; GAUCHE([@][Code postal]];2)="H3"; GAUCHE([@][Code postal]];2)="H4"; GAUCHE([@][Code postal]];2)="H8"; GAUCHE([@][Code postal]];2)="H9");"Local";"Extérieur")
Délai d'achat	Délai en jour entre l'achat et le spectacle	Nombre entier	Oui	=[@[Date de l'événement]]-[@Date d'achat]]

→ Pour en savoir plus sur les fonctions Excel, consultez le [site de Microsoft](#).

Dans le premier cas, grâce aux opérateurs « si » et « ou » et à la fonction « Gauche », nous demandons à Excel de donner la valeur « Local » à une observation lorsque son code postal commence par les caractères H1, H2, H3, H4, H8 ou H9, soit les deux premiers caractères des codes postaux associés à l'île de Montréal, et la valeur « Extérieur » aux autres observations (voir feuille « Création2 (provenance) » du fichier Excel).

À vous de jouer !

Testez vos habiletés Excel en essayant par vous-même de créer les trois indicateurs ci-dessous dans le jeu de données. Les réponses à ces exercices se trouvent dans la feuille « Exercice indicateur » du fichier Excel.

- Une variable qui répartie géographiquement les clients en trois régions distinctes : Région de Montréal, Est-du-Québec et Ouest-du-Québec, sachant que les codes postaux pour la région de Montréal commencent tous par « H », ceux de l'Est-du-Québec commencent par « G » et ceux de l'ouest commence par « J ».
- Une variable qui mesure la part des transactions effectuées durant l'événement sachant que celui-ci se déroule chaque année du 1er au 12 juillet. Il faudra alors créer une variable qui distingue les transactions réalisées « durant l'événement » de celles réalisées « hors de la période de l'événement ».
- Une variable qui permet de caractériser les consommateurs qui viennent seuls à un spectacle de ceux qui viennent accompagnés.

La création de variables dans une nouvelle feuille de données

Dans d'autres cas, comme lorsque l'on veut calculer le nombre total de billets vendus ou encore le taux de remplissage pour chacun des spectacles, il est préférable de créer ces variables sur une nouvelle feuille de données.

Pour ce faire, il suffit de créer un tableau croisé (voir feuille « Tableau croisé (BilletsParSpec) » dans le fichier Excel) avec les variables qui nous intéressent. Dans ce cas-ci, il s'agit du nom de l'événement⁸ et de la scène, le type de spectacle, la somme des quantités de billets vendus par transaction ainsi que la capacité de la scène.

Dans sa forme compactée, le tableau ainsi obtenu est difficile à utiliser pour la création de nouvelles variables. Ce problème se règle facilement à l'aide de la fonction « disposition du rapport » qui se trouve dans l'onglet « création » des outils de tableau croisé dynamique. En choisissant l'option « afficher sous forme tabulaire », nous obtenons un format plus approprié pour la création de nouvelles variables ; il ne reste donc plus qu'à coller les valeurs dans une nouvelle feuille pour travailler directement dessus (voir feuille « Données spectacle (remplissage) » dans le fichier Excel). Pour créer la nouvelle variable « taux de remplissage » de la scène, par exemple, il suffit ensuite de diviser les valeurs de la colonne « billets vendus » par celles de la colonne « capacité de la scène ».

⁸ Pour que cette étape se déroule bien, il est important d'avoir des données propres. Ici, les noms de spectacle suivent un même format ce qui nous assure d'avoir un nouveau tableau avec des données valides. Si les spectacles étaient parfois désignés par une chaîne de caractères (« spectacle 23 »), parfois par un seul nombre (« 23 »), il serait beaucoup plus difficile d'arriver rapidement aux résultats attendus.

Pour obtenir des informations concernant les clients, par exemple le nombre de transactions par clients, nous pouvons répéter ces étapes en glissant de nouvelles variables dans les champs de tableau croisé. Un exemple est présenté dans la feuille « Tableau croisé (clients) » du fichier Excel « **jeudedonnées.xlsx** ». Notez que l'utilisation de la fonction de « Champ calculé » dans un tableau croisé dynamique permet d'accélérer la démarche en fournissant directement le résultat dans le tableau croisé. Cette méthode réalise les calculs sur la somme de vos données.

Nous pouvons, par exemple, utiliser cette technique pour calculer le prix moyen des billets achetés par chacun des clients. Pour ce faire, dirigez-vous vers la feuille « Tableau croisé (clients) » du fichier Excel et cliquez sur l'option « Champs, Éléments et jeux » de l'onglet « Calculs ». Nous pouvons ainsi ajouter directement dans notre tableau croisé un « champ calculé », soit une variable calculée à partir d'une formule arithmétique simple. Dans ce cas-ci, en divisant la quantité de billets achetés par un client par les revenus totaux qui y sont associés, nous obtenons le prix moyen de chaque billet directement dans notre tableau croisé.

Cette option permet aussi de faire des calculs à l'aide des autres variables calculées (nous pourrions utiliser le prix moyen des billets achetés dans une nouvelle formule). Elle est toutefois restreinte aux opérations mathématiques simples et ne prend en compte que la somme des variables (il est donc impossible de faire des calculs à partir des moyennes d'une variable par exemple).

The screenshot shows the Microsoft Excel ribbon with the 'Analyse' tab selected. A context menu is open over a table, and the 'Insérer un champ calculé...' option is highlighted. An 'Insertion d'un champ calculé' dialog box is displayed, showing the formula '=Revenu transaction'/Quantité'.

Date de l'événement	DélaiAchat	Quantité	Prix
26,00 \$			
53,67 \$			
60,55 \$			
21,20 \$			
57,00 \$			
46,80 \$			
32,00 \$			
60,80 \$			
36,00 \$			

À vous de jouer !

À l'aide du tableau de la feuille « Tableau croisé (clients) », calculez le délai d'achat moyen pour chaque consommateur.

Partie 4. Exploiter de nouvelles habiletés par la conception d'un tableau de bord

Maintenant que nous savons créer des graphiques, réaliser des tableaux croisés dynamiques, définir des indicateurs intelligents et créer des variables qui permettront de mesurer certains indicateurs, nous sommes prêts à concevoir un premier tableau de bord.

Brièvement, un tableau de bord est un outil qui permet de visualiser et d'analyser de façon simple et rapide les différentes mesures clés ou indicateurs et ainsi de suivre l'évolution de la performance de votre organisation. Le tableau de bord est utile parce qu'il permet de transformer les données brutes en information pertinente pour l'aide à la décision.

Dans le cas de notre festival fictif, le tableau de bord permettra de suivre en un seul coup d'œil plusieurs indicateurs reliés à la billetterie, tels que les revenus liés aux spectacles, le nombre de billets vendus, les canaux utilisés par les clients pour effectuer leurs transactions, le taux de remplissage moyen par scène, et ce, par année et selon le type de discipline artistique. Le tout aidera à savoir quelles sont les disciplines artistiques qui sont les plus populaires auprès des festivaliers locaux ou encore à suivre l'évolution du nombre d'abonnés au fil des ans. Le fichier « **tableaudebord.xlsx** » vous sera nécessaire pour reproduire les différentes manipulations expliquées dans les prochaines sections.

Quelques règles d'or pour un bon tableau de bord

Définir un but précis pour le tableau de bord et s'en tenir à l'essentiel

Avant de concevoir un tableau de bord, il faut d'abord définir son but et ses objectifs en se posant les questions suivantes :

- **À quoi servira le tableau de bord ? Quelles sont les questions auxquelles il répondra ?**
- **Sera-t-il consulté quotidiennement ou de façon ponctuelle (par mois, par trimestre, etc.) ?**
- **Quelles sont les personnes (cibles) qui le consulteront ?**
- **Quels sont les indicateurs les plus importants pour le travail de ces personnes ?**

Cette étape est essentielle puisqu'elle permet de construire un tableau de bord aligné aux besoins de l'organisation ou des personnes concernées, incitant ainsi à le consulter plus régulièrement. En identifiant un but précis, nous évitons de faire un tableau de bord confus avec un trop grand nombre de dimensions. En effet, pour concevoir un bon tableau de bord, il faut s'en tenir aux seules informations qui sont pertinentes pour le public visé par le tableau de bord.

À titre d'exemple, un tableau de bord mélangeant des variables de comptabilité (ex. : le total des factures de nos fournisseurs), de ressources humaines (ex. : le nombre de nos bénévoles) et du site Web (ex. : le nombre de pages vues) aura comme défaut d'être peu utile aux différents départements et apportera plus de confusion que de réponses aux questions.

Dans le cadre de notre festival fictif, nous avons choisi de construire un tableau de bord destiné aux responsables de la billetterie et de la programmation. C'est pourquoi le tableau comprend des informations sur la performance des spectacles, selon la discipline artistique ou encore la scène utilisée, ainsi que sur le profil des clients (abonnement, canal d'achat, provenance).

Concevoir un visuel simple et attrayant

Comme c'est le cas pour la plupart des visualisations de données, la force d'un tableau de bord réside dans sa conception graphique ou, en d'autres mots, dans sa capacité à transformer des données complexes en visuels simples et attrayants. Il suffit de suivre certaines consignes générales :

- **Éviter d'inclure une trop grande quantité d'information :**
 - Sachant que la mémoire de travail de l'être humain peut traiter de 5 à 9 éléments d'information à la fois, il est préférable d'éviter les tableaux de bord massif avec des dizaines de graphiques, chiffres ou jauge et de s'en tenir au minimum nécessaire.
 - Arrondir les nombres contenus dans les visualisations contribue à réduire la lourdeur du tableau de bord, voire dépouiller les visualisations pour accentuer le message en n'y faisant apparaître que les valeurs que l'on veut porter à l'attention du lecteur.
- **Jouer avec la taille, la forme et la position des éléments pour établir une structure hiérarchique claire :**
 - L'œil est habituellement guidé vers les visualisations en haut de tableau et celles qui sont de grandes tailles. Nous pouvons nous servir de cette dynamique pour mettre en évidence les indicateurs les plus importants.
 - Les couleurs et les formes peuvent être très utiles pour regrouper les indicateurs qui sont interreliés et bien communiquer ce lien aux lecteurs.
- **Définir clairement les indicateurs :**
 - Un bon tableau de bord doit se passer d'explication. Il est donc important de choisir des titres qui soient clairs et permettent d'éviter au lecteur de mal interpréter les données. Qui plus est, annoncer la conclusion dans le titre d'un graphique est une façon de faciliter la compréhension du message.
 - Lorsque les abréviations sont nécessaires (par exemple, pour ne pas alourdir trop le visuel), un simple lexique des indicateurs peut être ajouté en bas du tableau de bord.

Comment construire un tableau de bord dans Excel ?

Étape 1 : Créer des visualisations avec l'outil de graphique croisé dynamique.

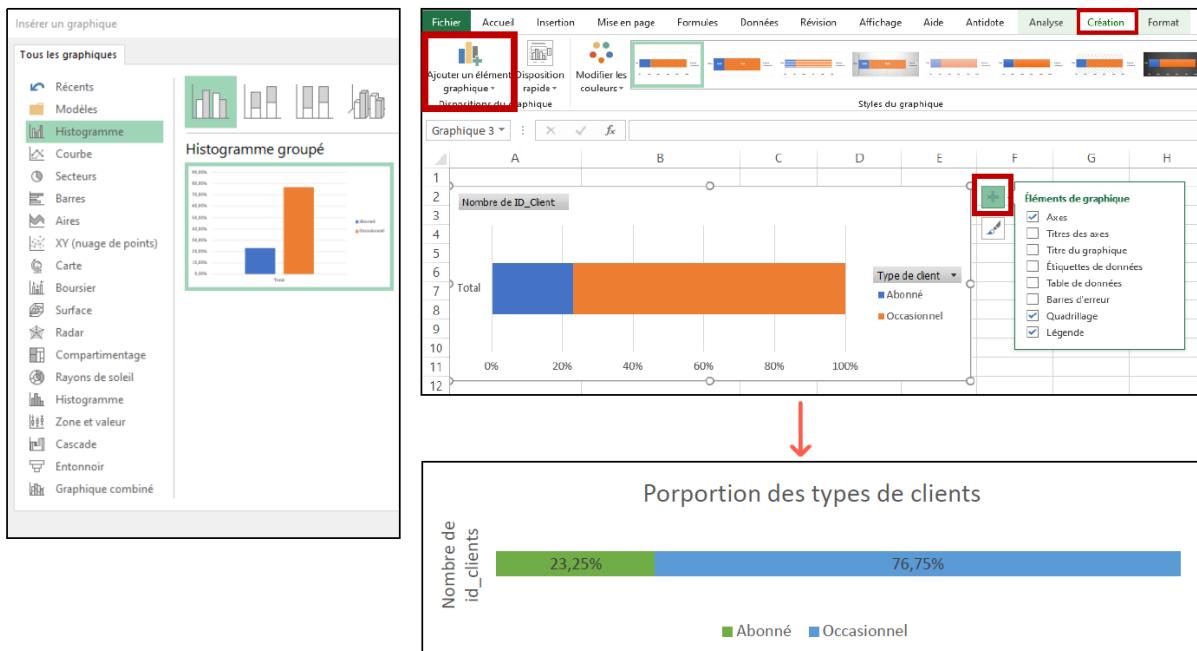
Pour concevoir un tableau de bord dans Excel, il faut simplement maîtriser l'outil de tableau croisé dynamique vu dans la Partie 2 de ce guide ainsi que son proche compagnon, le graphique croisé dynamique. En effet, pour créer un tableau de bord dans Excel, nous devons d'abord créer les tableaux croisés nécessaires et les graphiques qui y sont associés, et ensuite les coller sur une feuille destinée à héberger le tableau de bord.

Création du tableau croisé dynamique

Pour la première visualisation, l'objectif est de représenter la part de clients abonnés à l'événement par rapport aux occasionnels (variable « type de client »). En insérant la variable « ID_Client » dans le champ valeur et la variable « type de client » dans le champ colonnes, nous obtenons le tableau requis. En cliquant sur une cellule du tableau, nous avons maintenant accès à l'outil graphique croisé dynamique.

Création du graphique croisé dynamique

Grâce à l'outil graphique croisé dynamique, nous avons accès à plusieurs types de graphiques pour illustrer les données du tableau croisé. Une fois le graphique créé, il peut être redimensionné et formaté selon les besoins. Après avoir modifié les couleurs, ajouté un titre et des étiquettes de données et retiré les autres éléments superflus, nous obtenons une visualisation prête à être copiée et collée dans le tableau de bord. Il ne reste qu'à répéter ces étapes pour créer les autres visualisations qui composeront le tableau de bord. Cet exemple se trouve sur la feuille « Tableau (type de client) » du fichier Excel « [tableaubord.xlsx](#) ».

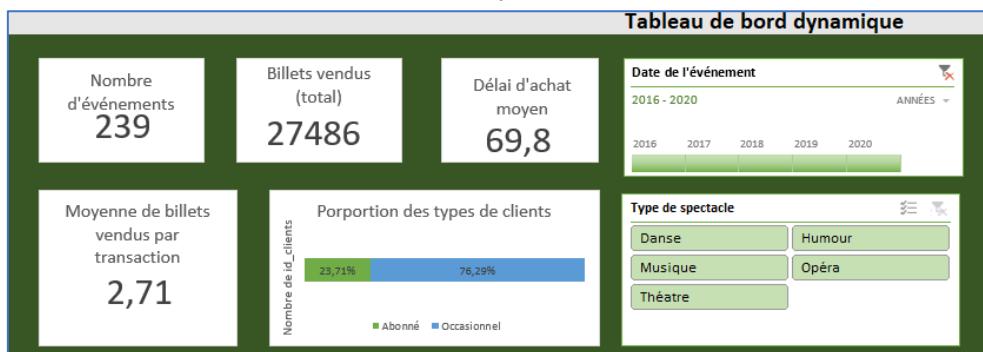
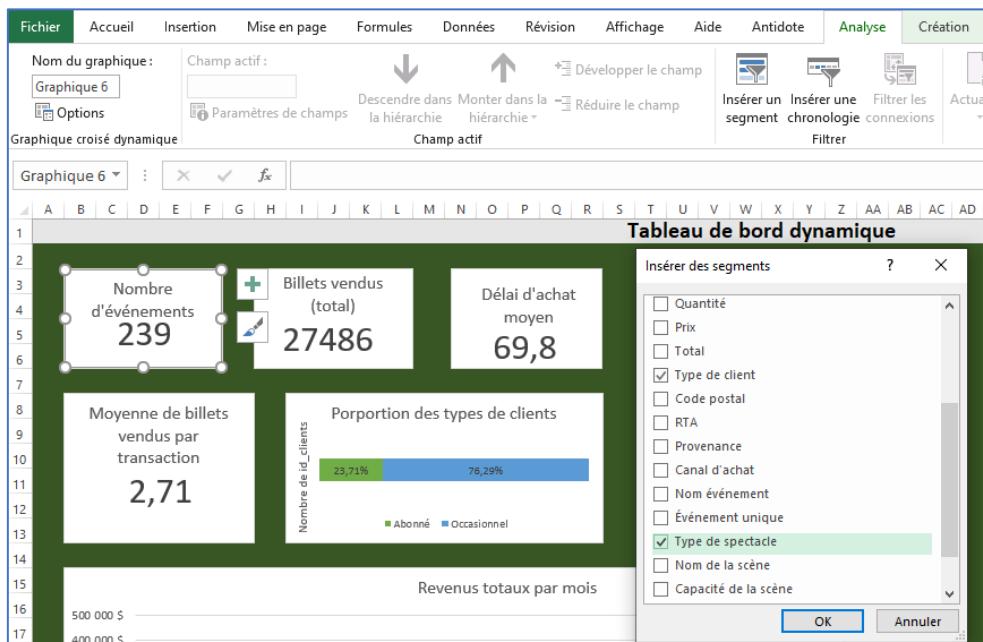
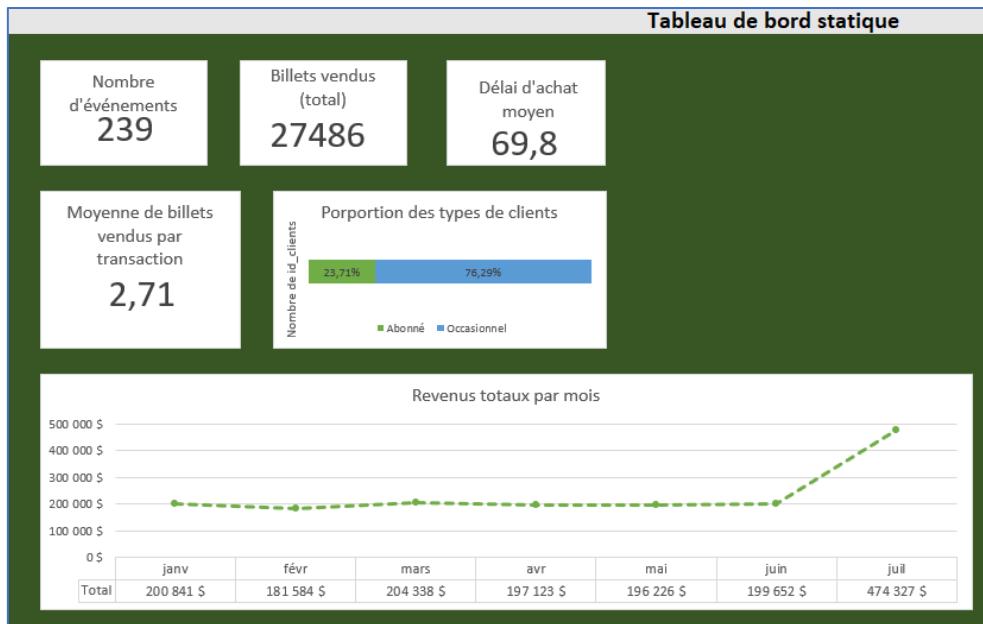


Étape 2 : Assembler le tableau de bord

L'ajout de filtres : les segments et les chronologies

En combinant quelques visualisations, le tableau de bord commence à prendre forme. Il possède toutefois un défaut important qu'il faut corriger : il s'agit d'un tableau de bord statique qui brosse le portrait du jeu de données complet, mais qui ne permet pas d'isoler les données pour une année précise ou encore pour certains types de spectacles (par exemple, la musique et l'humour seulement). Autrement dit, il ne permet pas de « creuser » les données.

Pour rendre le tableau de bord dynamique ou interactif, nous pouvons utiliser des filtres qu'Excel nomme *segment* et *chronologie*. Ces outils permettent de filtrer en un clic les données selon une variable catégorielle (comme le « type de spectacle ») ou encore selon des étendues de temps (années, mois, trimestres, etc.).



Une fois le filtre ajouté au tableau de bord, nous pouvons modifier son apparence (couleurs, nombre de colonnes, etc.) dans le menu « Options » qui s'affiche en sélectionnant celui-ci.

Il ne reste maintenant qu'à connecter ce filtre à tous les autres tableaux croisés dynamiques afin qu'ils s'actualisent tous en même temps. Pour ce faire, il faut cliquer sur l'outil « Connexions de rapport » et choisir les différents tableaux que l'on souhaite voir être modifiés par le filtre.

The screenshot shows the Microsoft Power BI ribbon with the 'Options' tab selected. Below the ribbon, a dynamic dashboard is displayed. The dashboard features several cards: 'Nombre d'événements' (261), 'Billets vendus (total)' (27486), 'Délai d'achat moyen' (69,8), and a date range selector for 'Date de l'événement' from 2016 to 2020. A callout box points to the 'Connexions de rapport' (Report Connections) icon in the ribbon. A separate callout box highlights the 'Connexions de rapport (Type de spectacle)' dialog box, which lists various report items and their connection status.

Connexions de rapport (Type de spectacle)

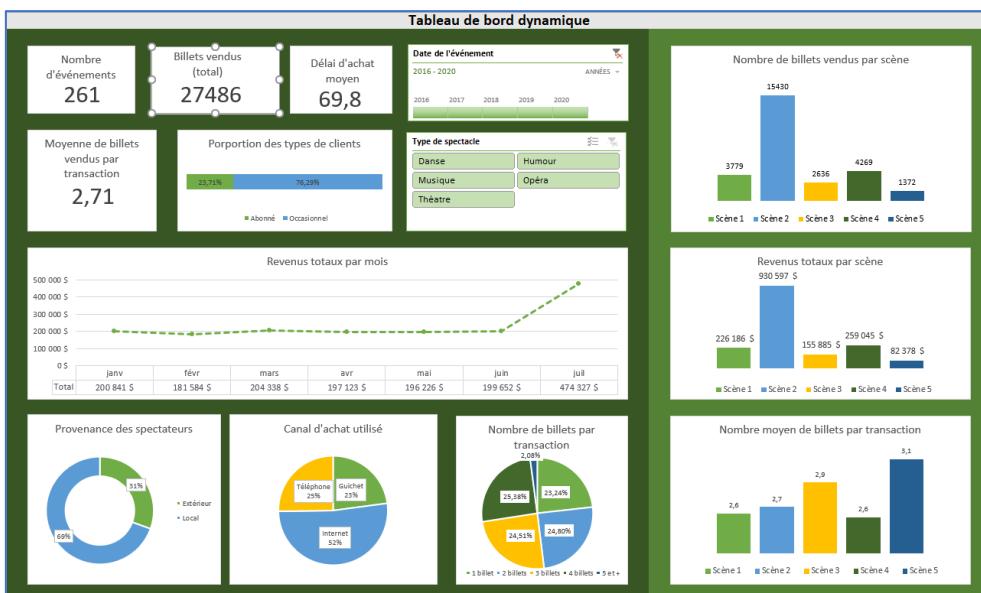
Selectionner les rapports de tableaux croisés dynamiques et de graphiques croisés dynamiques à connecter à ce filtre

Nom	Feuille
<input checked="" type="checkbox"/> Tableau croisé dynamique8	Remplissage
<input checked="" type="checkbox"/> Tableau croisé dynamique7	Tableau (canal d'achat)
<input checked="" type="checkbox"/> Tableau croisé dynamique5	Tableau (nb de billets vendus)
<input checked="" type="checkbox"/> Tableau croisé dynamique5	Tableau (nb d'événements)
<input type="checkbox"/> Tableau croisé dynamique5	Tableau (Prévalence)

OK Annuler

Nous pouvons maintenant contempler le résultat final et, surtout, commencer à explorer les données à l'aide de ces filtres !

Le résultat final

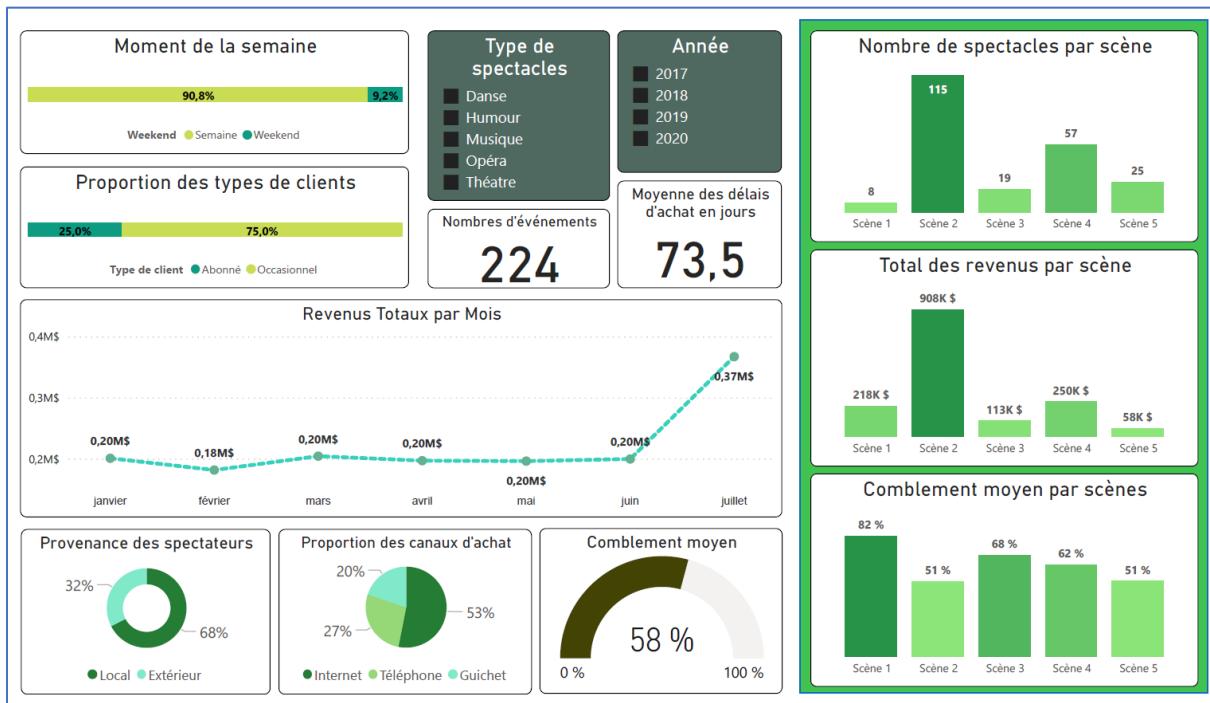


Pour aller plus loin : le tableau de bord dans Power BI

Bien que le logiciel Excel permet de faire des tableaux de bord plus que convenables, nous vous invitons à vous renseigner sur le logiciel Power BI, aussi offert avec les pack Office 365, et sur ses possibilités analytiques. En effet, Power BI a plusieurs avantages :

- Il peut se connecter à un grand nombre de sources de données ;
- Il travaille mieux avec de gros jeux de données qu'Excel ;
- Il permet de créer plus rapidement des tableaux de bord visuellement supérieurs à ceux d'Excel.

Puisque le fait que la présentation de Power BI déborde de la portée de cette ressource, l'exemple ci-dessous présente un avant-goût d'un autre tableau de bord conçu avec Power BI à partir du même jeu de données.



En conclusion

Grâce à ce chapitre consacré à l'analyse et à la visualisation des données, vous savez maintenant comment créer un dictionnaire de données et explorer un jeu de données à l'aide de statistiques descriptives et de visualisations simples. De plus, vous en savez maintenant plus sur le Tableau croisé dynamique d'Excel : vous pouvez maintenant l'utiliser pour explorer et analyser des données ou encore pour construire un tableau de bord avec vos principaux indicateurs de performance.

À vous de jouer !

C'est maintenant l'heure pour vous de reproduire les différentes techniques vues dans ce chapitre avec un autre jeu de données réelles cette fois.