



ANALYSE DE DONNÉES AVEC EXCEL

Objectifs du cours



Analyser un ensemble de données à l'aide d'outils Excel natifs



Appliquer les outils et formules Excel pour transformer et structurer les données



Créer des tableaux croisés dynamiques pour explorer et filtrer les données

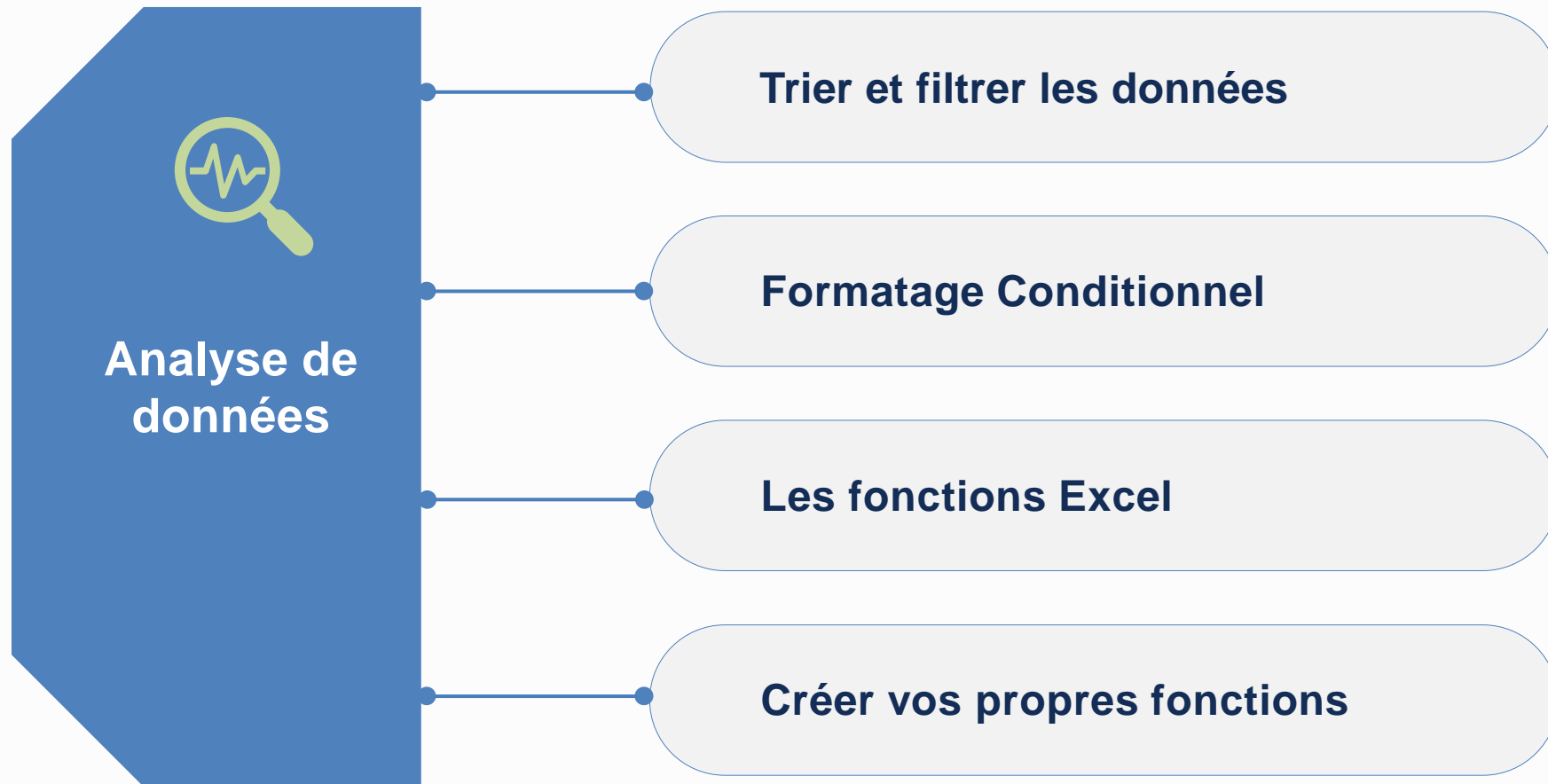


Visualisez les données avec des tableaux croisés dynamiques et des graphiques Excel



Explorez le paysage de l'analyse et de la visualisation des données

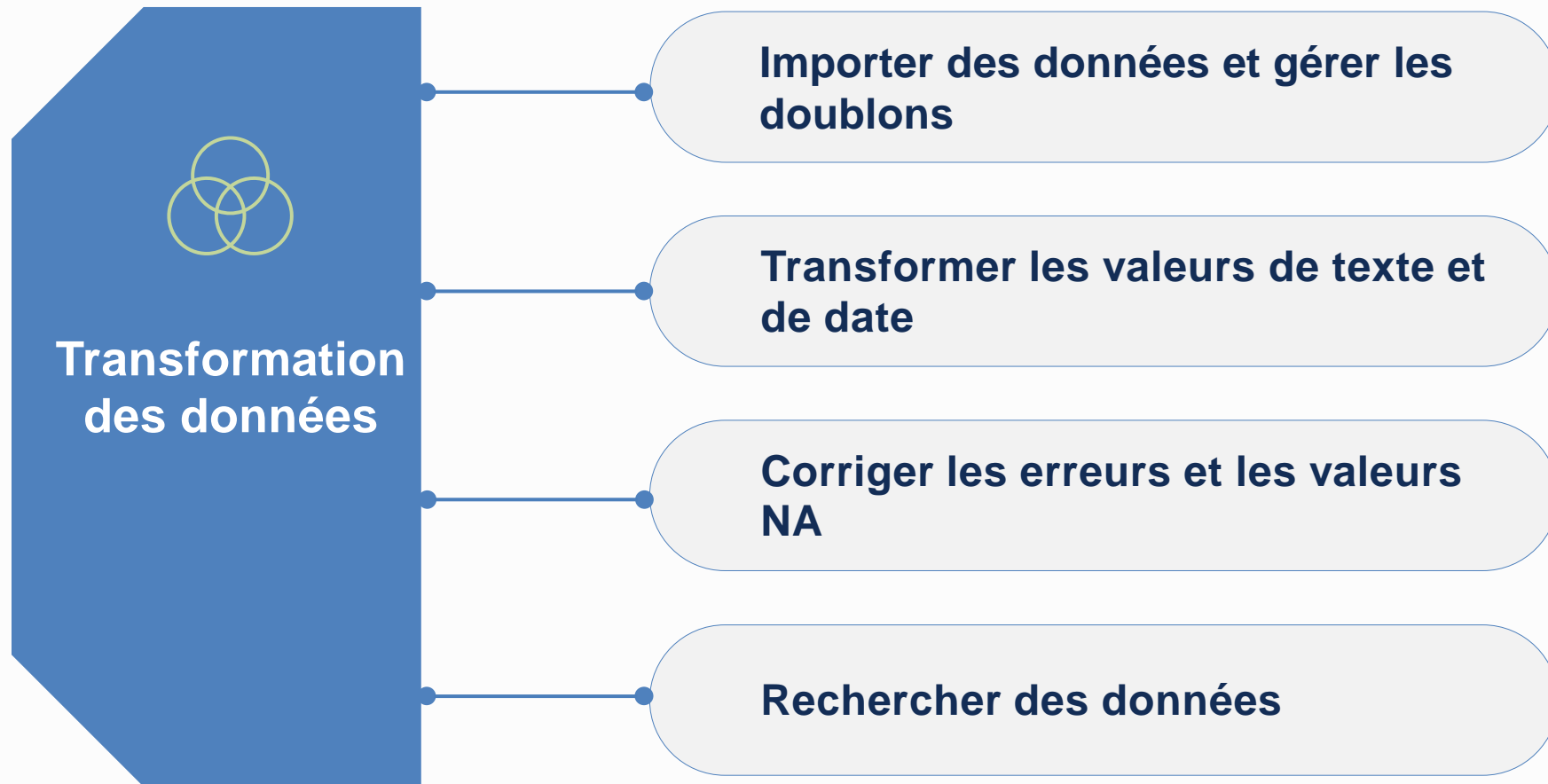
1– Analyse de données



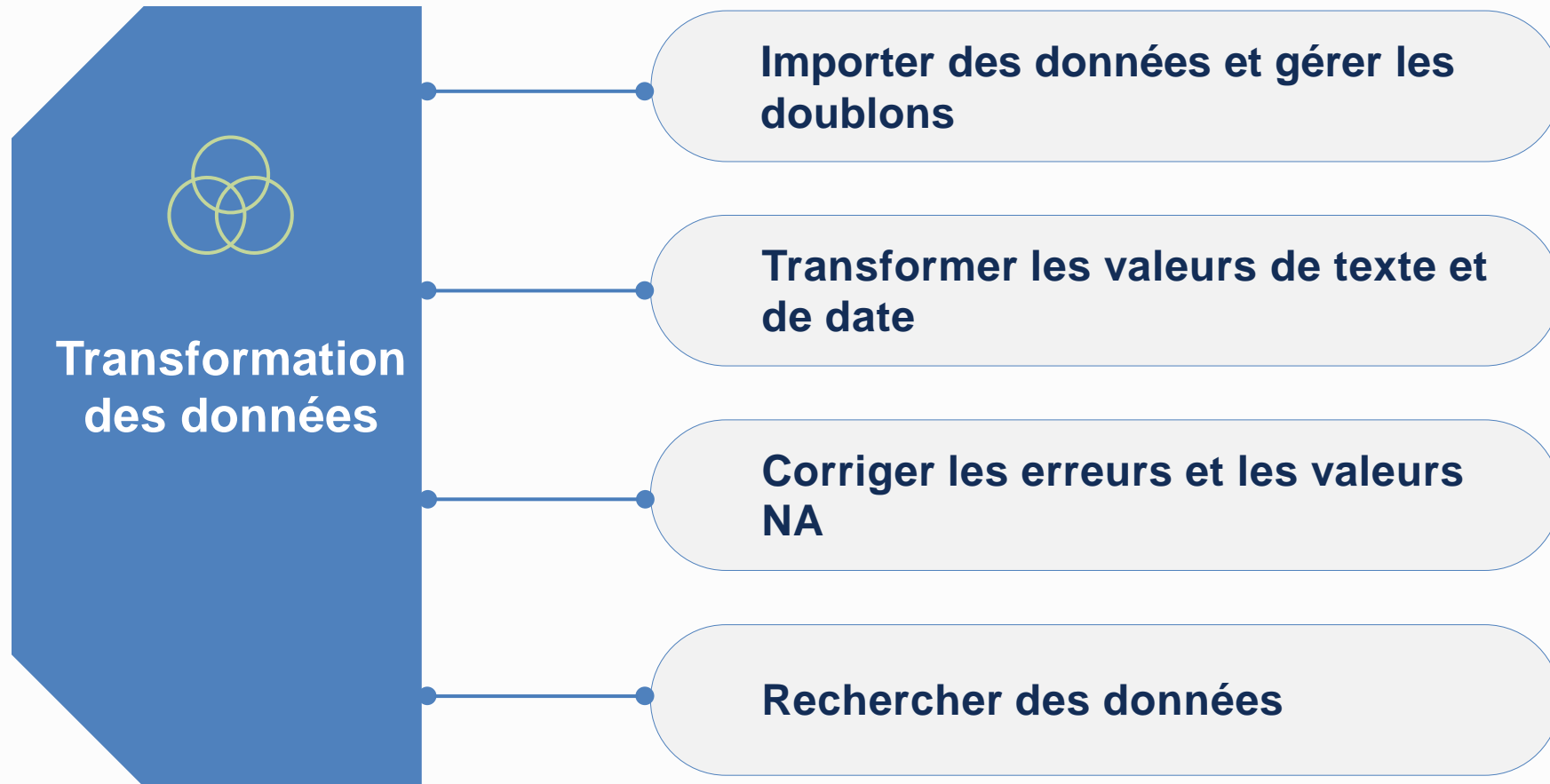
1– Analyse de données



2- Transformation des données



2- Transformation des données



3- Les tableaux Excel



Tableaux Excel

Gérer et analyser les données associées



Power Query

Outil de préparation et de transformation des données dans Excel



Types de données liées

Fournir la possibilité de se connecter à des données sur une variété de sujets

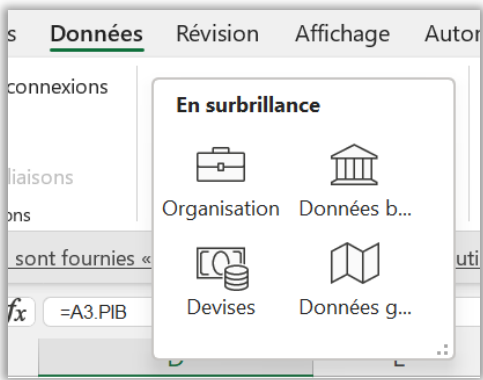
3- Les tableaux Excel

Les types de données liées

Les types de données liées sont placées dans l'onglet « Données »

. Il en existe quatre (4) :

- Organisation
- Données boursières
- Devises
- Données géographiques



Pays
Espagne
Canada
Allemagne
Italie
Belgique
Chine
Maroc
Sénégal
Burkina Faso
Nigeria
Mali
Gabon



Pays	Code devise	Espérance de vie	PIB	Population	Population urbaine
🇪🇸 Espagne	EUR	83,3	\$ 1 394 116 310 769	47 076 781	37 927 409
🇨🇦 Canada	CAD	81,9	\$ 1 736 425 629 520	36 991 981	30 628 482
🇩🇪 Allemagne	EUR	80,9	\$ 3 845 630 030 824	83 132 799	64 324 835
🇮🇹 Italie	EUR	82,9	\$ 2 001 244 392 042	60 297 396	42 651 966
🇧🇪 Belgique	EUR	81,6	\$ 529 606 710 418	11 484 055	11 259 082
🇨🇳 Chine	CNY	77,0	\$ 18 460 000 000 000	1 397 715 000	842 933 962
🇲🇦 Maroc	MAD	76,5	\$ 118 725 279 596	36 910 560	22 975 026
🇸🇳 Sénégal	XOF	67,7	\$ 23 578 084 052	16 296 364	7 765 706
🇧🇫 Burkina Faso	XOF	61,2	\$ 15 745 810 235	20 321 378	6 092 349
🇳🇬 Nigeria	NGN	54,3	\$ 448 120 428 859	200 963 599	102 806 948
🇲🇱 Mali	XOF	58,9	\$ 17 510 141 171	19 658 031	8 479 688
🇬🇦 Gabon	XAF	66,2	\$ 16 657 960 228	2 172 579	1 949 694

3- Les tableaux Excel

Les tableaux Excel sont importants pour une analyse de données moderne



Power Query



Type de données liées



Tableaux dynamiques



Tableaux croisés dynamiques

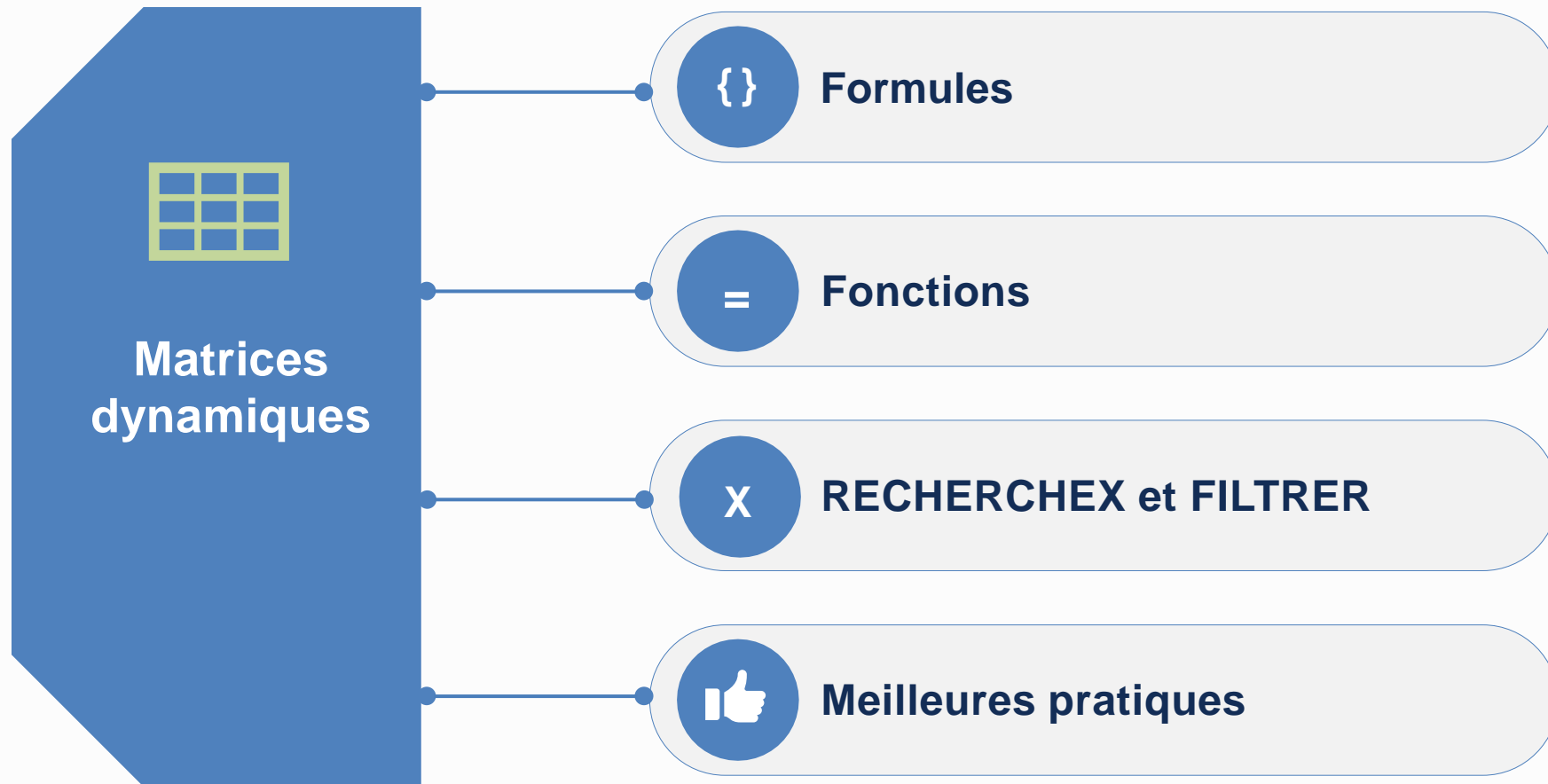


Visualisation des données

ID Etudiant	Prénom	Nom	Note moyenne	Faculté	Frais de scolarité	Heures de participation	Tutoriels Suivis	Cours ratés
20123456	John	Park	B	Arts	44191	0	10	5
20123457	Alex	Great	B	Science	32245	4	10	10
20123458	Sebastian	Taylor	B	Affaire	42679	6	3	7
20123459	Michael	Bay	A	Mathématiques	46478	15	3	2
20123460	Scott	Foster	A	Ingénierie	36784	5	9	8
20123461	Amy	Winehouse	B	Arts	36537	10	6	3
20123462	Ralph	Wiggins	B	Affaire	40762	2	6	8
20123463	Homer	Simpson	C	Ingénierie	47669	4	8	7
20123464	Marge	Simpson	B	Mathématiques	39429	10	5	3
20123465	Peter	Gryffin	D	Arts	31956	7	9	7
20123466	Louise	King	D	Affaire	33227	6	2	7
20123467	Megan	Botts	A	Science	34751	25	5	1
20123468	Cyrus	Wong	A	Science	49298	20	0	0
20123469	Michelle	Chang	B	Affaire	35046	5	10	4
20123470	Zachary	Chua	A	Affaire	31210	10	8	0
20123471	Angus	Helmsworth	B	Affaire	47515	10	6	1
20123472	Aaron	McDowell	B	Affaire	43421	13	7	2
20123473	Carol	Kuo	B	Ingénierie	43063	7	0	4
20123474	Tim	James	B	Science	46775	9	3	4
20123475	Johnson	Curry	B	Ingénierie	40539	9	4	3
20123476	Paul	Reed	A	Affaire	41397	14	2	2
20123477	Josh	Hart	A	Ingénierie	46468	4	8	8
20123478	Justin	Kang	A	Ingénierie	35859	2	4	9
20123479	Kevin	Yoo	A	Arts	41048	8	9	1
20123480	Rosaline	Jun	A	Arts	44915	25	6	3
20123481	Jimin	Park	B	Ingénierie	34570	20	1	3
20123482	Joseph	Kim	A	Mathématiques	33376	12	5	6
20123483	Chris	Dang	F	Affaire	44737	3	8	8
20123484	Robbie	Tee	B	Ingénierie	49682	10	5	6
20123485	Shelly	Yoon	A	Mathématiques	33585	5	7	10
20123486	Namjoon	Yoongi	A	Arts	53585	7	8	2

4- Les matrices dynamiques

Les formules matricielles dynamiques nous permettent de référencer plusieurs valeurs de données à la fois.



4- Les matrices dynamiques

Les tableaux Excel sont importants pour une analyse de données moderne



Plusieurs valeurs dans une formule



Fonctions matricielles dynamiques



Répansion des matrices

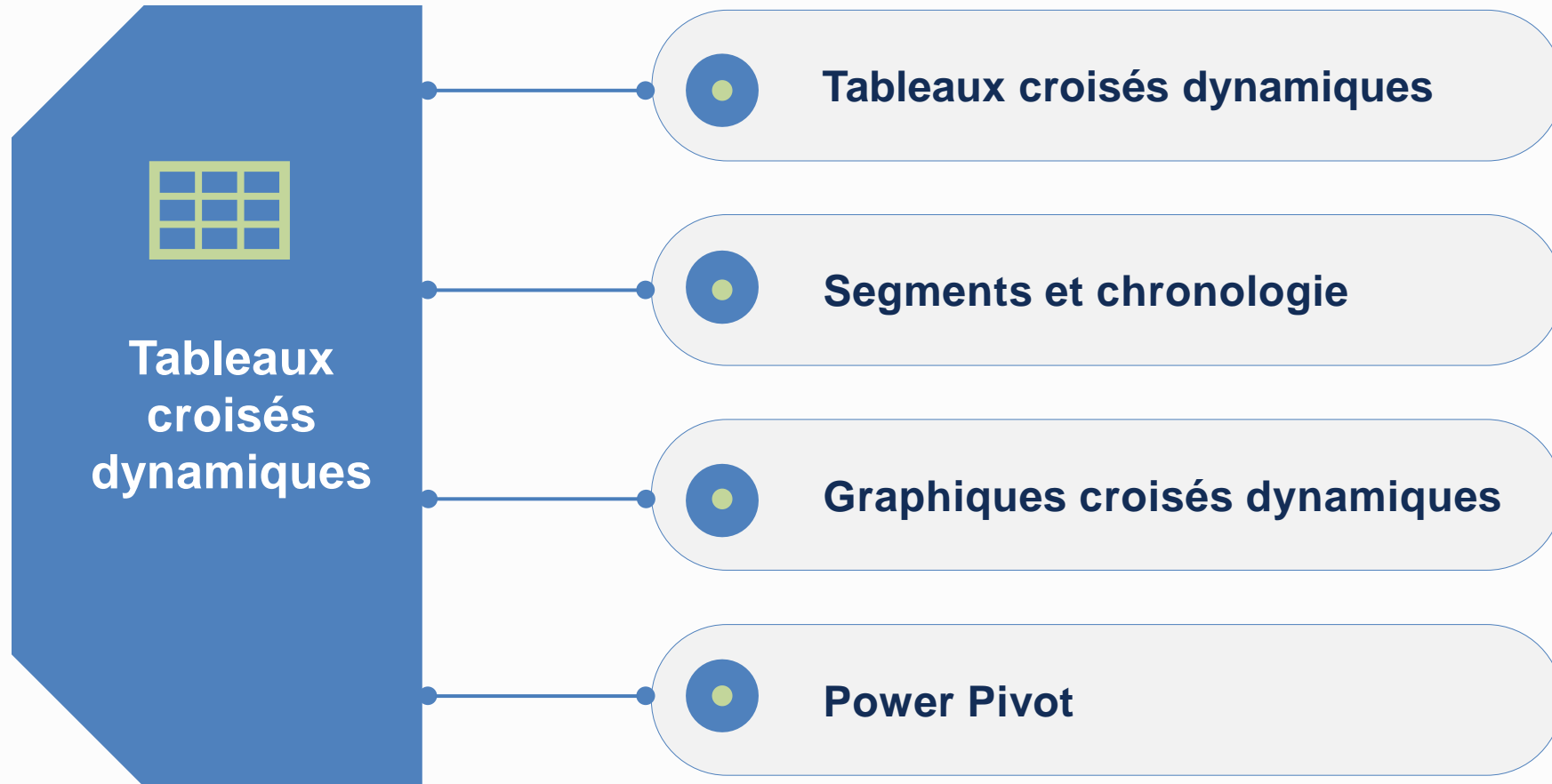


Solution pérenne

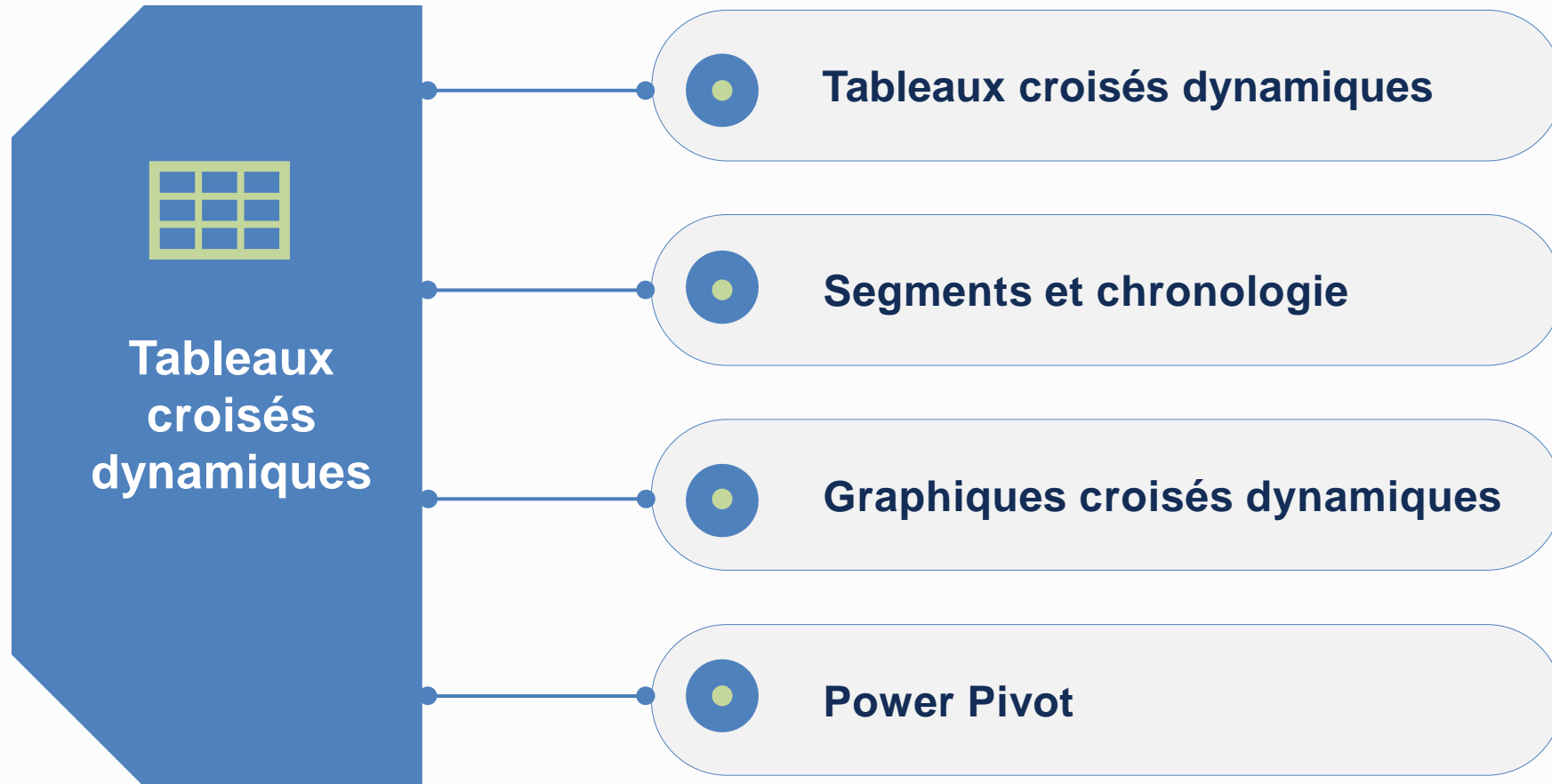
ID Etudiant	Prénom	Nom	Note moyenne	Faculté	Frais de scolarité	Heures de participation	Tutoriels Suivis	Cours ratés	
20123456	John	Park	B	Arts	44191		0	10	5
20123457	Alex	Great	B	Science	32245		4	10	10
20123458	Sebastian	Taylor	B	Affaire	42679		6	3	7
20123459	Michael	Bay	A	Mathématiques	46478		15	3	2
20123460	Scott	Foster	A	Ingénierie	36784		5	9	8
20123461	Amy	Winehouse	B	Arts	36537		10	6	3
20123462	Ralph	Wiggins	B	Affaire	40762		2	6	8
20123463	Homer	Simpson	C	Ingénierie	47669		4	8	7
20123464	Marge	Simpson	B	Mathématiques	39429		10	5	3
20123465	Peter	Gryffin	D	Arts	31956		7	9	7
20123466	Louise	King	D	Affaire	33227		6	2	7
20123467	Megan	Botts	A	Science	34751		25	5	1
20123468	Cyrus	Wong	A	Science	49298		20	0	0
20123469	Michelle	Chang	B	Affaire	35046		5	10	4
20123470	Zachary	Chua	A	Affaire	31210		10	8	0
20123471	Angus	Helmsworth	B	Affaire	47515		10	6	1
20123472	Aaron	McDowell	B	Affaire	43421		13	7	2
20123473	Carol	Kuo	B	Ingénierie	43063		7	0	4
20123474	Tim	James	B	Science	46775		9	3	4
20123475	Johnson	Curry	B	Ingénierie	40539		9	4	3
20123476	Paul	Reed	A	Affaire	41397		14	2	2
20123477	Josh	Hart	A	Ingénierie	46468		4	8	8
20123478	Justin	Kang	A	Ingénierie	35859		2	4	9
20123479	Kevin	Yoo	A	Arts	41048		8	9	1
20123480	Rosaline	Jun	A	Arts	44915		25	6	3
20123481	Jimin	Park	B	Ingénierie	34570		20	1	3
20123482	Joseph	Kim	A	Mathématiques	33376		12	5	6
20123483	Chris	Dang	F	Affaire	44737		3	8	8
20123484	Robbie	Tee	B	Ingénierie	49682		10	5	6
20123485	Shelly	Yoon	A	Mathématiques	33585		5	7	10
20123486	Namjoon	Yoongi	A	Arts	53585		7	8	2

5- Les tableaux croisés dynamiques

Les tableaux croisés dynamiques peuvent être utilisés pour résumer de grands ensembles de données et générer rapidement de nombreuses informations pertinentes.



5- Les tableaux croisés dynamiques



6- Visualisation des données

Visualisation de données avec des graphiques Excel :



Sélectionner des données



Choisir un graphique



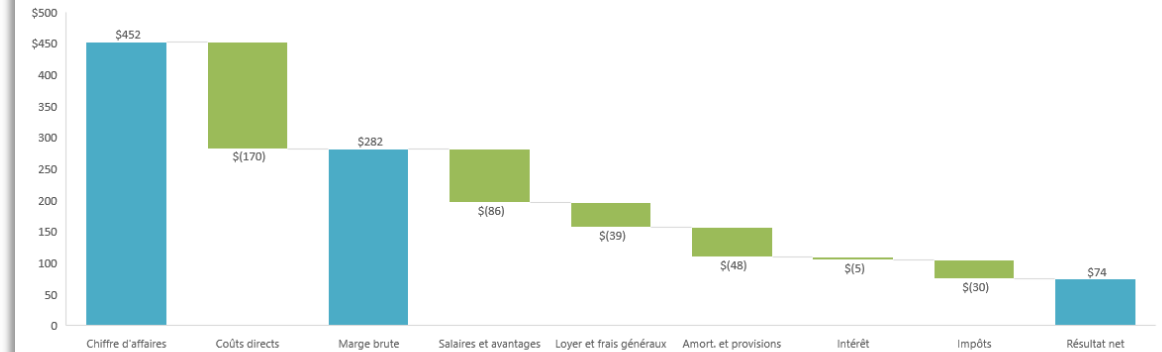
Formater le graphique

© Xelmore

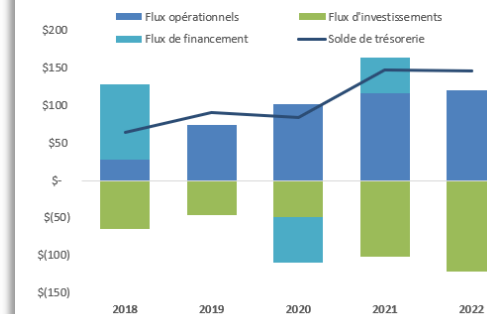
Visualisation des données dans Excel



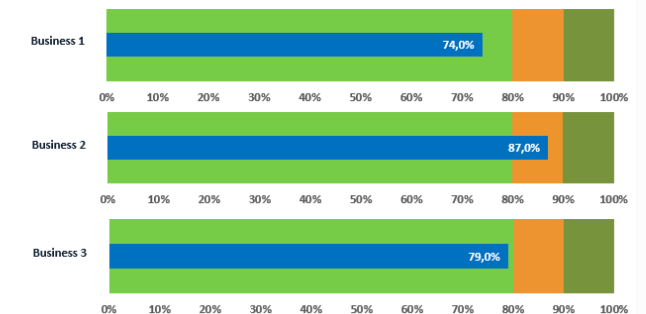
Graphique en cascade - Bénéfices nets 2022 ('000 \$)



Flux de trésorerie et solde de trésorerie (000 \$)

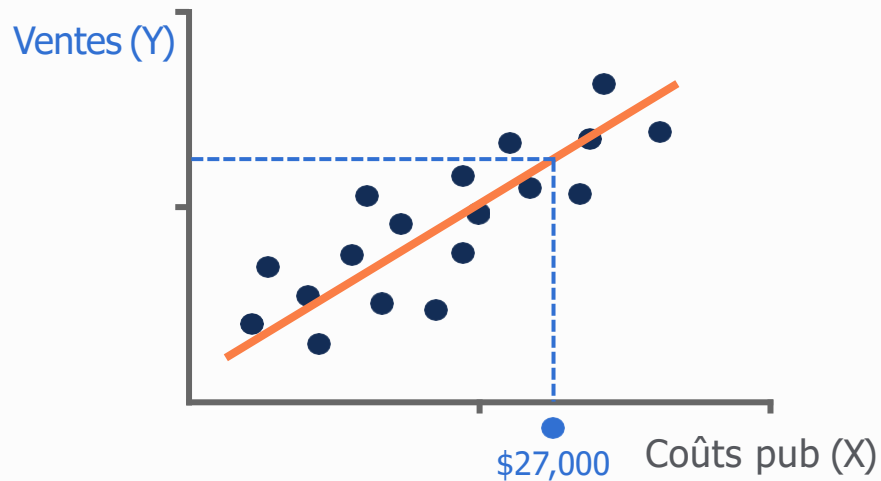


Taux de productivité 2022

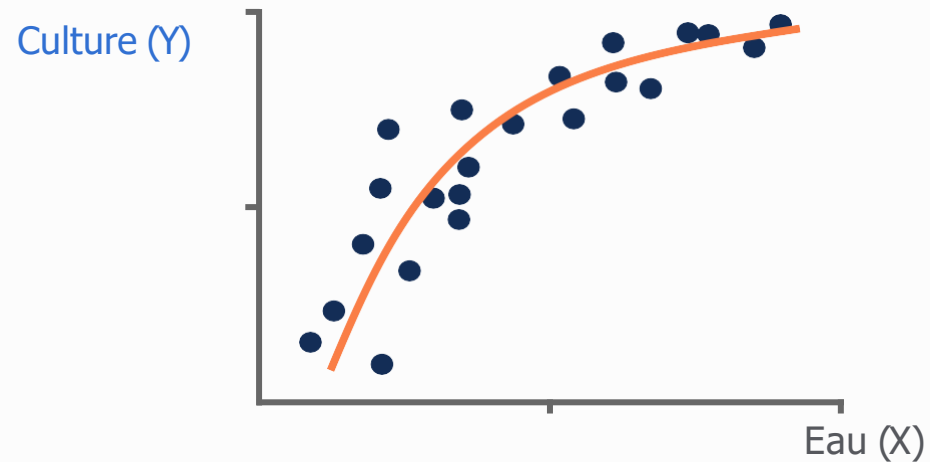


7- Régression linéaire avec Excel

Le but de la régression est d'évaluer la relation entre une ou plusieurs variables d'entrée (X) et une variable de sortie continue (Y).



Comment les ventes sont affectées par les dépenses publicitaires.



Rendement des cultures par rapport à l'eau

Notre ligne de meilleur ajustement nous permet de faire des prédictions sur la valeur de la variable cible dans un scénario donné.

7- Régression linéaire avec Excel

Terminologie

Variable(s) prédictive(s)
Variable(s) d'entrée
Variable(s) indépendante(s)
Variable(s) explicative(s)
X (X1, X2, X3 ...)



Variable cible
Variable de sortie
Variable dépendante
Variable de réponse
Y

7- Régression linéaire avec Excel

La régression linéaire est la forme la plus simple d'analyse de régression.

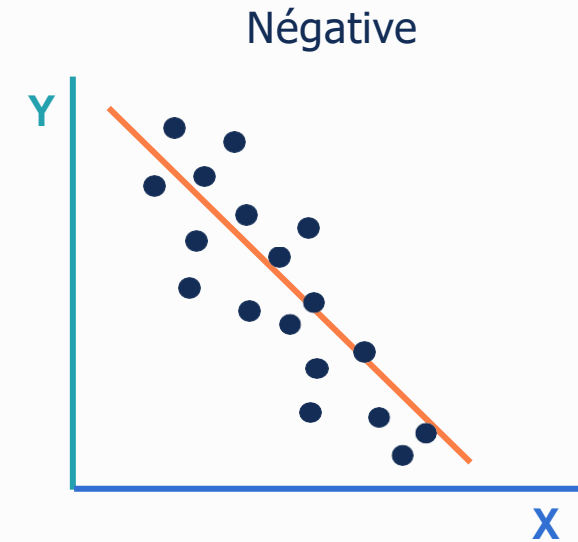
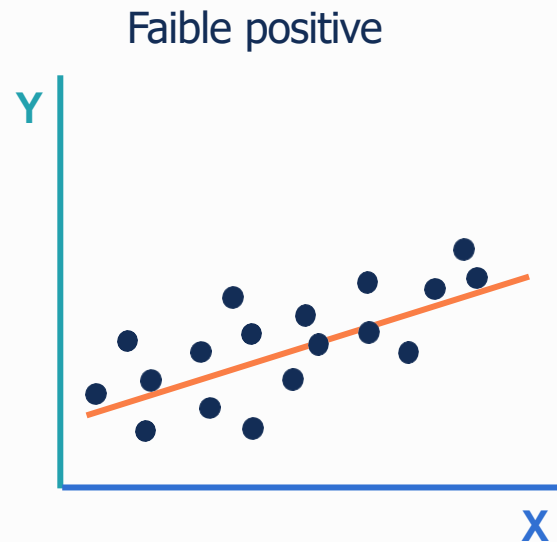
$$y = mx + b$$

Y est la variable de sortie (cible)

X est la variable d'entrée

M est la valeur du coefficient (pente)

B est l'intercept (de l'axe Y)

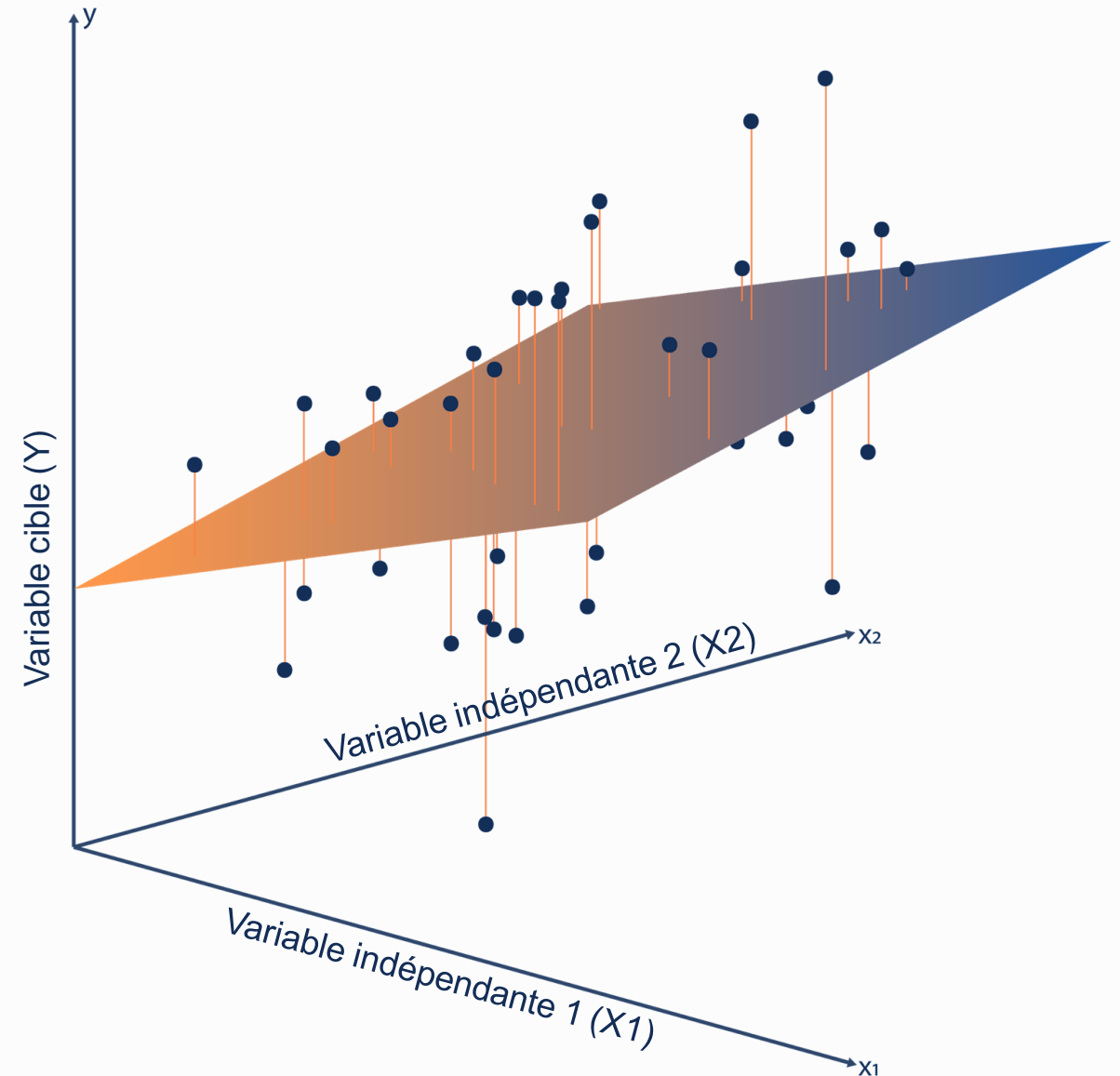


7- Régression linéaire avec Excel

Régression linéaire multiple

- Une variable d'entrée est rarement suffisante pour faire des prédictions sur une variable cible.
- **La régression linéaire multiple** nous permet de prédire une variable cible à l'aide de plusieurs variables indépendantes.
- Lorsque nous avons deux variables indépendantes, nous ajustons un plan aux données au lieu d'une ligne droite

$$y = B + M1X1 + M2X2...MnXn$$

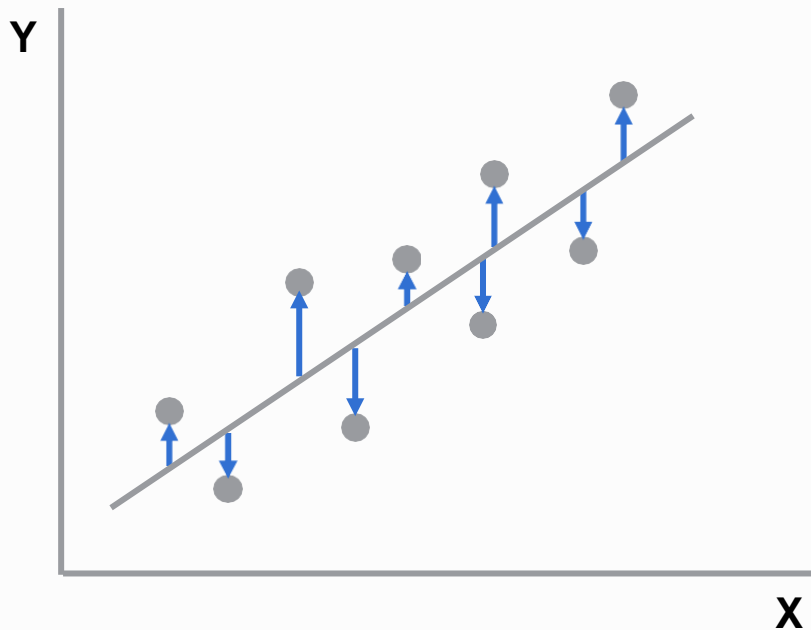


7- Régression linéaire avec Excel

The Optimal Line of Best Fit

How do we decide where exactly the line of best fit sits? Generally, we try to **minimize the size of errors** in our predictions.

Errors represent the amount by which the target variable is different from the value predicted.



Other Metrics	Sensitive to Outliers
Sum of Squared Error (SSE)	Yes
Mean Squared Error (MSE)	Yes
Root Mean Squared Error (RMSE)	Yes

Other Metrics	Sensitive to Outliers
Sum Absolute Error (SAE)	No
Mean Absolute Error (MAE)	No

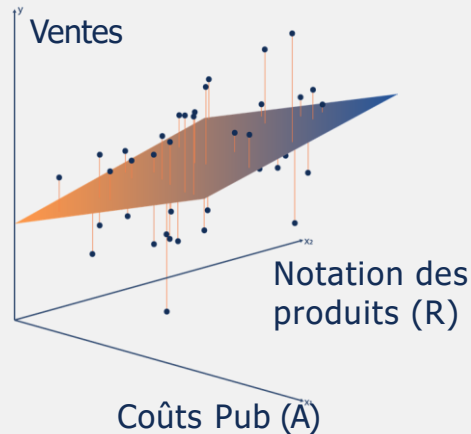
The most common approach **minimizes the squared errors**, and is known as **Ordinary Least Squares**

7- Régression linéaire avec Excel

Interprétation des coefficients

Les coefficients de régression nous aident à comprendre l'interaction entre les variables

Scénario marketing



$$\text{Ventes} = 2\,000 + 1,3A + 0,11R$$

Où **A** est « Coûts Pub »

et **R** est Notation des produits

Le **coefficient des coûts** est 1,3

Nous pouvons dire que pour chaque unité de coûts Pub supplémentaire, **les ventes augmentent de 1,3**

Le **coefficient notation produit** est 0,11

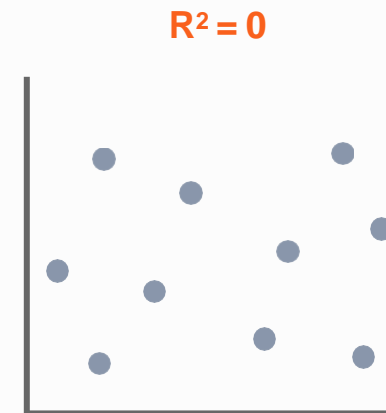
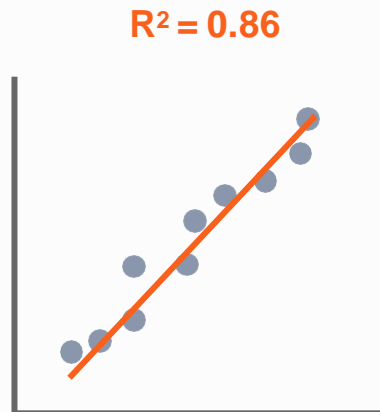
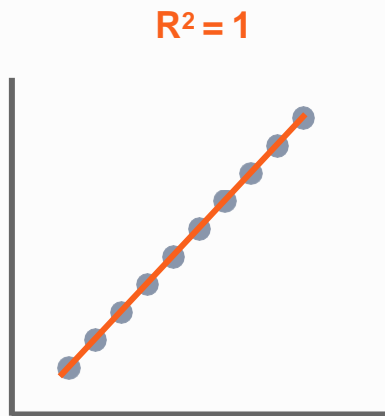
Nous pouvons dire que pour chaque unité d'augmentation de la notation des produits, **les ventes croîtront de 0,11**

7- Régression linéaire avec Excel

Regression Metrics

Coefficient of Determination (R^2) is one of the most used metrics to evaluate regression models.

R^2 measures how close the data are to the fitted regression line. In other words, how much of the variability in Y is explained by changes in X.



Higher R^2 indicates better fit of the model, and therefore smaller errors.

R^2 can sometimes be biased, so a related measure called **Adjusted R^2** can also be used.

7- Régression linéaire avec Excel



Conclusion