

Créer un diagramme de Pareto dans Excel

par SilkyRoad (silkyroad.developpez.com)

Date de publication : 08 Novembre 2008

Dernière mise à jour :

Cet article montre comment mettre en oeuvre un diagramme de Pareto dans Excel, sans VBA.

Les descriptions correspondent à une utilisation dans Excel 2007. Le principe de fonctionnement est identique pour les versions antérieures.

| | |
|--|----|
| I - Introduction..... | 3 |
| II - La collecte des données..... | 3 |
| III - Les calculs..... | 4 |
| III-A - Lister les catégories sans doublons..... | 4 |
| III-B - Compter le nombre d'occurrences par catégorie..... | 6 |
| III-C - Déterminer le rang de chaque catégorie..... | 6 |
| III-D - Trier les catégories par totaux décroissants..... | 6 |
| III-E - Calculer la part de chaque catégorie..... | 6 |
| III-F - Cumuler les occurrences..... | 7 |
| III-G - Visualiser les limites A, B et C dans le graphique..... | 7 |
| IV - Le résultat..... | 7 |
| IV-A - Le diagramme de Pareto..... | 7 |
| IV-B - L'ajout des séries dans le graphique..... | 8 |
| IV-C - La mise en forme du graphique..... | 9 |
| IV-C-1 - Redéfinir les limites de l'axe des ordonnées..... | 9 |
| IV-C-2 - Modifier les types de séries d'histogramme en courbe..... | 9 |
| IV-C-3 - Afficher toutes les catégories sur l'axe des abscisses..... | 9 |
| IV-C-4 - Supprimer les espaces entre chaque donnée en abscisses..... | 10 |
| V - Conclusion..... | 10 |
| VI - Liens..... | 11 |
| VII - Remerciements..... | 11 |
| VIII - Téléchargement..... | 11 |

I - Introduction

Le diagramme de Pareto est un outil statistique qui permet d'identifier l'importance relative de chaque catégorie dans une liste d'enregistrements, en comparant leur fréquence d'apparition.

Un diagramme de Pareto est mis en évidence lorsque 20 % des catégories produisent 80 % du nombre total d'effets. Cette méthode permet donc de déterminer rapidement quelles sont les priorités d'actions. Si on considère que 20 % des causes représentent 80% des occurrences, agir sur ces 20 % aide à solutionner un problème avec un maximum d'efficacité.

Citation

En préparant leurs ambitions, les managers expérimentés savent que seuls quelques éléments majeurs sont décisifs. Le reste sera traité par la même occasion en tant que parties de ces éléments.

Joseph Moser Juran - 1964

Un diagramme de Pareto est généralement présenté sous forme d'histogramme en colonne. Les catégories sont affichées en abscisses, par ordre décroissant de fréquence. Le nombre d'apparition est indiqué sur l'axe des ordonnées. Une courbe en ligne contenant les valeurs cumulées complète le diagramme.

Le diagramme de Pareto est également appelé :

- * Méthode "ABC".
- * Règle des 80/20.

L'exemple présenté dans cet article a été mis en oeuvre dans Excel 2007. Le principe reste identique sous les versions antérieures.

Il existe bien entendu d'autres solutions Excel pour visualiser un diagramme de Pareto.

Cette démo reposait sur plusieurs attendues :

- * Ne pas utiliser de VBA.
- * Automatiser au maximum le classeur pour que l'utilisateur ait simplement à renseigner les données collectées.
- * Détailler les formules pour expliquer le principe de calcul.
- * Afficher les valeurs dans un graphique pour obtenir un résultat visuel.
- * Pouvoir réutiliser directement le classeur démo pour d'autres analyses.

Le classeur démo montre également comment :

- * Extraire des données sans doublons puis les trier, par formule.
- * Créer des plages nommées dynamiques dans Excel 2007.
- * Utiliser ces noms dans des graphiques Excel 2007.

II - La collecte des données

La collecte des données est une action essentielle car toute l'analyse va être basée sur la validité des informations recueillies.

Il convient donc de vérifier :

- * La justesse des périodes de mesures.
- * La véracité des enregistrements.
- * Les unités de mesure identiques pour l'ensemble de la source de données.
- * La cohérence des catégories.
- * Le regroupement des causes mineures en une seule catégorie "Divers".

L'exemple proposé dans la feuille 'Collecte' (voir le classeur en téléchargement à la fin de cette page) liste les causes d'immobilisation d'une machine de production. Un nombre d'heures d'arrêt, mesuré sur une année, est associé à chaque catégorie. L'objectif final est d'identifier les causes principales d'arrêt et de savoir sur quelles catégories agir en priorité pour augmenter le taux de rendement de l'installation.

| | A | B | C | D |
|----|-------------------------|-------------------------|------------|---|
| 1 | Catégorie | Immobilisation (heures) | Date | |
| 2 | manque matière première | 1 | 02/01/2007 | |
| 3 | panne mécanique | 2 | 03/02/2007 | |
| 4 | Manque de personnel | 3 | 03/02/2007 | |
| 5 | manque matière première | 2 | 10/02/2007 | |
| 6 | panne mécanique | 1 | 06/03/2007 | |
| 7 | panne mécanique | 3 | 07/03/2007 | |
| 8 | panne mécanique | 1 | 01/04/2007 | |
| 9 | panne électrique | 2 | 05/04/2007 | |
| 10 | intercampagne | 2 | 19/04/2007 | |
| 11 | panne mécanique | 1 | 07/05/2007 | |
| 12 | intercampagne | 2 | 16/05/2007 | |

Si vous souhaitez réutiliser le classeur pour votre projet, supprimez les données existantes dans la feuille 'Collecte', à partir de la deuxième ligne, puis insérez vos valeurs.

La colonne A contient les catégories (les causes).

Insérez les éléments à la suite dans la feuille, sans ligne vide. Ne vous préoccupez pas des calculs ou des tris préalables. L'application va s'en charger automatiquement comme vous le découvrirez dans la suite du tutoriel.

Pensez juste à regrouper les catégories mineures en une seule catégorie "Divers" afin de ne pas polluer l'analyse par des données non représentatives.

La colonne B contient le nombre d'occurrences pour chaque catégorie.

Indiquez les valeurs en prenant soins de conserver une seule unité (Par exemple, ne mélangez pas des kilos avec des tonnes).

Attention à ne pas utiliser plusieurs libellés pour une même cause.

Excel dispose d'une fonctionnalité dans le menu contextuel des cellules afin de choisir rapidement parmi les éléments existants dans les cellules contiguës d'une colonne :

Clic droit dans la première cellule vide de la colonne / Liste déroulante de choix.

| | | | |
|----|-------------------------|---|------------|
| 38 | intercampagne | 3 | 07/12/2007 |
| 39 | panne mécanique | 3 | 18/12/2007 |
| 40 | manque matière première | 2 | 20/12/2007 |
| 41 | | | |
| 42 | essai | | |
| 43 | intercampagne | | |
| 44 | Manque de personnel | | |
| 45 | manque matière première | | |
| 46 | panne instrumentation | | |
| 47 | panne mécanique | | |
| 48 | panne électrique | | |
| 49 | travaux modification | | |

Nota :

Si vous disposez d'une liste brute de catégories sans comptage d'occurrences, placez simplement les groupes dans la colonne A et indiquez 1 dans chaque cellule de la colonne B.

III - Les calculs

Toutes les opérations sont réalisées dans la feuille 'Calculs'.

Ce chapitre décrit en détail le processus qui va permettre d'obtenir le résultat définitif.

III-A - Lister les catégories sans doublons

La première phase consiste à extraire la liste des catégories sans doublons, de la feuille 'Collecte'.

La formule ci-dessous est saisie dans la cellule A2 (colonne 'IndexUniques') puis étirée vers le bas :

Formule

```
=SI(Collecte!A2="" ; "" ; SI (NB.SI(Collecte!A$2:A2;Collecte!A2)>1 ; "" ; MAX($A$1:$A1) +1))
```

Ainsi, vous repérez et indexez la position de chaque catégorie de manière unique.

 **Important :**

Cette formule doit être recopiée dans autant de lignes qu'il y a de données dans la feuille 'Collecte' afin de pouvoir prendre en compte toutes les catégories.

Dans la cellule B2 (colonne 'Liste catégories'), est saisie la formule :

Formule

```
=SI(LIGNES($A$2:$A2)>MAX(IndexUniques) ; "" ; DECALER(Collecte!$A$2;EQUIV(LIGNES($A$2:$A2);IndexUniques;0)-1;0))
```

Vous obtenez la liste des catégories sans doublons.

Remarquez dans la formule, l'utilisation d'une plage nommée 'IndexUniques' qui représente la colonne A dans la feuille 'Calculs'.

Procédez de la manière suivante pour créer le nom :

Onglet "Formules" dans le ruban.

Bouton "Gestionnaire de noms" dans le groupe "Noms définis".

Bouton "Nouveau".

"IndexUniques" est saisi dans le champ "Nom". C'est la référence qui sera utilisée dans la cellule B2.

La formule `=DECALER(Calculs!A2;;;NBVAL(Collecte!$A:$A)-1;1)` est indiquée dans le champ "Fait référence à".

Le bouton OK permet de valider le nom.

La fonction DECALER peut être résumée ainsi dans notre exemple :

`DECALER(Cellule de base;;;Nombre de lignes;Nombre de colonnes)`

Le nombre de lignes de la plage nommée est dépendant de la partie `NBVAL(Collecte!$A:$A)-1`.

Elle compte le nombre de données dans la colonne A de la feuille 'Collecte'.

La formule DECALER permet donc la mise en oeuvre d'une plage nommée dynamique, c'est à dire qui sera redéfinie automatiquement en fonction du nombre de données dans la feuille 'Collecte'.

Les autres fonctions du classeur utilisent également des plages nommées dynamiques. Les mises à jour sont donc automatisées et l'utilisateur n'a pas besoin de modifier toutes les formules quand le nombre et le contenu des données changent.

Nota

C'est volontairement que la plage nommée 'IndexUniques' est basée sur le nombre de lignes de la feuille 'Collecte'.

Ce choix limite le risque d'oublier des catégories dans la source de données.

Le tableau suivant récapitule les noms et les références utilisés dans le classeur démo :

| Nom | Référence | Formule |
|------------------|------------------------------|--|
| Categories | Feuille Collecte / colonne A | <code>=DECALER(Collecte!\$A\$2;;;NBVAL(Collecte!\$A:\$A)-1;1)</code> |
| Occurrences | Feuille Collecte / colonne B | <code>=DECALER(Collecte!\$B\$2;;;NBVAL(Collecte!\$B:\$B)-1;1)</code> |
| IndexUniques | Feuille Calculs / colonne A | <code>=DECALER(Calculs!\$A\$2;;;NBVAL(Collecte!\$A:\$A)-1;1)</code> |
| NbOccurrences | Feuille Calculs / colonne C | <code>=DECALER(Calculs!\$C\$2;;;SOMMEPROD((Calculs!\$C\$1:\$C\$100<>"")*1)-1;1)</code> |
| RangUnique | Feuille Calculs / colonne D | <code>=DECALER(Calculs!\$D\$2;;;SOMMEPROD((Calculs!\$D\$1:\$D\$100<>"")*1)-1;1)</code> |
| OrdreOccurrences | Feuille Calculs / colonne F | <code>=DECALER(Calculs!\$F\$2;;;SOMMEPROD((Calculs!\$C\$1:\$C\$100<>"")*1)-1;1)</code> |
| CategoriesTriees | Feuille Calculs / colonne G | <code>=DECALER(Calculs!\$G\$2;;;SOMMEPROD((Calculs!\$G\$1:\$G\$100<>"")*1)-1;1)</code> |
| Part | Feuille Calculs / colonne H | <code>=DECALER(Calculs!\$H\$2;;;SOMMEPROD((Calculs!\$C\$1:\$C\$100<>"")*1)-1;1)</code> |
| CumulOccurrences | Feuille Calculs / colonne I | <code>=DECALER(Calculs!\$I\$2;;;SOMMEPROD((Calculs!\$C\$1:\$C\$100<>"")*1)-1;1)</code> |
| Pourcentage | Feuille Calculs / colonne J | <code>=DECALER(Calculs!\$J\$2;;;SOMMEPROD((Calculs!\$C\$1:\$C\$100<>"")*1)-1;1)</code> |
| Limite_A | Feuille Calculs / colonne K | <code>=DECALER(Calculs!\$K\$2;;;SOMMEPROD((Calculs!\$D\$1:\$D\$100<>"")*1)-1;1)</code> |
| Limite_B | Feuille Calculs / colonne L | <code>=DECALER(Calculs!\$L\$2;;;SOMMEPROD((Calculs!\$D\$1:\$D\$100<>"")*1)-1;1)</code> |

III-B - Compter le nombre d'occurrences par catégorie

La colonne C (champ 'Nb occurrences') de la feuille 'Calculs' va comptabiliser le total de valeurs pour chaque groupe.

Formule

```
=SI (B2=""; "", SOMME.SI (Categories;B2;Occurrences) )
```

Le nom 'Categories' représente la colonne A dans la feuille 'Collecte' :

=DECALER(Collecte!\$A\$2;;;NBVAL(Collecte!\$A:\$A)-1;1)

Le nom 'Occurrences' représente la colonne B dans la feuille 'Collecte' :

=DECALER(Collecte!\$B\$2;;;NBVAL(Collecte!\$B:\$B)-1;1)

La formule additionne toutes les valeurs par catégorie.

III-C - Déterminer le rang de chaque catégorie

Un diagramme de Pareto représente généralement un histogramme de données décroissantes et une courbe des valeurs cumulées. Pour obtenir le même résultat visuel dans un graphique Excel, nous allons devoir classer chaque total par ordre décroissant.

Tout d'abord il faut identifier le rang de chaque total : 1 pour le plus grand total, 2 pour le deuxième total, ...etc ...

La formule suivante est insérée dans la colonne D (champ 'RangUnique').

Formule

```
=SI (B2=""; "", RANG (C2;NbOccurrences) +NB.SI ($C$2:C2;C2) -1)
```

Le nom 'NbOccurrences' représente la colonne C dans la feuille 'Calculs'.

+NB.SI(\$C\$2:C2;C2)-1 permet de différencier la numérotation des ex-aequo. Cela sera utile pour lister les catégories triées, comme vous pourrez le voir dans le chapitre suivant.

III-D - Trier les catégories par totaux décroissants

La phase suivante consiste à trier les valeurs et les catégories associées.

La colonne F (champ 'OrdreOccurrences') trie les totaux par ordre décroissant.

Formule

```
=SI (C2=""; "", GRANDE.VALEUR (NbOccurrences;LIGNE () -1) )
```

La colonne G (champ 'CatégoriesTriées') trie les catégories par ordre décroissant.

Cette plage de cellules servira à créer l'axe des abscisses du graphique.

Formule

```
=SI (C2=""; "", INDEX (DECALER (RangUnique; -2; SOMMEPROD ((Calculs!$C$1:$C$100<>"") *1) -1; 3); EQUIV (LIGNE () -1; RangUnique; 0); 1) )
```

III-E - Calculer la part de chaque catégorie

Les données de la colonne H (champ 'Part') vont servir de valeurs d'ordonnées pour l'histogramme du graphique.

Il s'agit donc ici de redéfinir les valeurs de la colonne F (champ 'OrdreOccurrences') sous forme de pourcentage par rapport au total de cette colonne.

Formule

```
=SI (C2=""; ""; F2/SOMME (OrdreOccurences))
```

III-F - Cumuler les occurrences

Un diagramme de Pareto est également basé sur la représentation d'une courbe cumulée de 0 à 100%. La colonne I (champ 'OrdreOccurences') va additionner les valeurs triées dans la colonne F.

Formule

```
=SI (F2=""; ""; SOMME ($F$2:F2))
```

La colonne J (champ 'Pourcentage') va convertir les valeurs de la colonne I en pourcentage par rapport au total de cette colonne.

Cette plage de cellules servira d'ordonnées pour la courbe de cumul dans le graphique.

Formule

```
=SI (I2=""; ""; I2/SOMME (NbOccurences))
```

III-G - Visualiser les limites A, B et C dans le graphique

L'efficacité d'un graphique Excel reposant en partie sur son attrait visuel, il reste à définir des droites à 80 et 95 %.

Vous pourrez ainsi visualiser rapidement les limites qui définissent l'analyse de Pareto :

La plage A : Catégories contenant de 0 à 80% des effets.

La plage B : Catégories contenant de 80 à 95 % des effets.

La plage C : les 5 % d'effets restants.

Le calcul effectué dans les colonnes K et L permettra d'afficher ces droites limites dans le graphique.

Formule

```
=SI (C2=""; ""; 0,8)
```

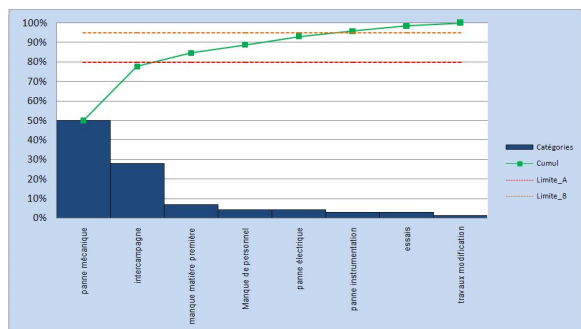
Formule

```
=SI (C2=""; ""; 0,95)
```

IV - Le résultat

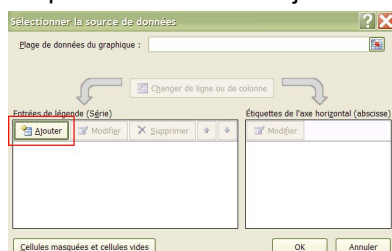
IV-A - Le diagramme de Pareto

Désormais, tous les paramètres sont en place pour afficher le diagramme de Pareto dans un graphique Excel. L'objectif consiste à obtenir la présentation finale suivante :

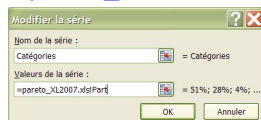


IV-B - L'ajout des séries dans le graphique

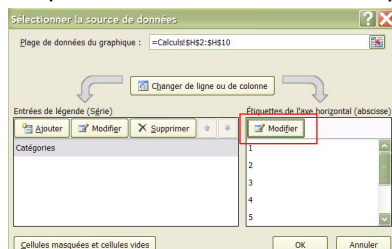
Ajoutez une nouvelle feuille de calcul.
Sélectionnez l'onglet "Insertion" dans le ruban.
Cliquez sur le bouton "Colonne" dans le groupe "Graphiques".
Sélectionnez "Histogramme 2D" / groupe.
Un nouvel objet graphique vierge est inséré dans la feuille.
Sélectionnez cet objet graphique.
Faites un clic droit.
Cliquez sur l'option "Sélectionner des données" dans le menu contextuel.
Cliquez sur le bouton "Ajouter".



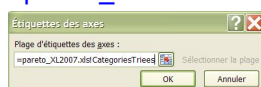
Nous allons créer l'histogramme décroissant (données de la colonne H) :
Indiquez un nom pour la série.
Dans le champ "Valeurs de la série", saisissez la plage nommée correspondant à la colonne H :
`=pareto_XL2007.xls!Part`



Cliquez sur le bouton OK pour valider.
Cliquez sur le bouton "Modifier" pour spécifier l'abscisse (la liste des catégories).



Saisissez la plage nommée correspondant à la colonne G :
`=pareto_XL2007.xls!CategoriesTries`



Cliquez sur le bouton OK dans toutes les boîtes de dialogue pour valider.
Reproduisez la même opération pour ajouter les autres séries.
La courbe de cumul :

=pareto_XL2007.xls!Pourcentage

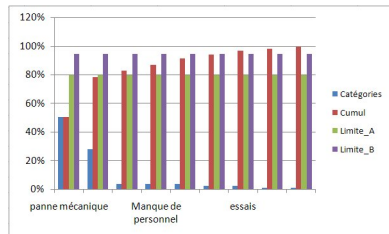
La droite de limite des 80 % :

=pareto_XL2007.xls!Limite_A

La droite de limite des 95 % :

=pareto_XL2007.xls!Limite_B

Toutes les séries sont insérées dans le graphique mais on est encore loin du compte pour ce qui est du résultat visuel :



Le chapitre suivant montre comment mettre en forme le graphique et finaliser le projet.

IV-C - La mise en forme du graphique

IV-C-1 - Redéfinir les limites de l'axe des ordonnées

Par défaut, les valeurs d'ordonnées sont indiquées de 0 à 120 %.

Pour afficher les valeurs de 0 à 100 %, sélectionnez l'axe des ordonnées.

Clic droit.

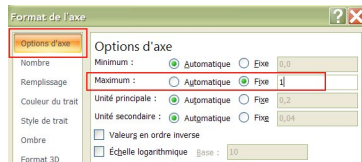
Sélectionnez "Mise en forme de l'axe" dans le menu contextuel.

Cliquez sur le menu "Options de l'axe".

Sélectionnez l'option "Fixe" pour la valeur "Maximum".

Indiquez 1 (ce qui représente 100 %) dans le champ.

Cliquez sur le bouton "Fermer".



IV-C-2 - Modifier les types de séries d'histogramme en courbe

Les séries Cumul, Limite_A et Limite_B doivent être transformées en courbes dans le diagramme de Pareto.

Procédez de la même manière pour les 3 séries :

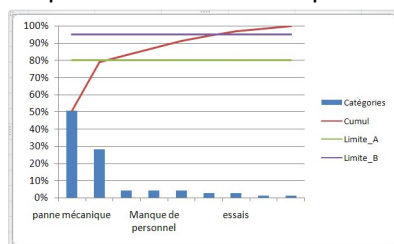
Sélectionnez la série.

Clic droit.

Sélectionnez "Modifier le type de graphique Série de données" dans le menu contextuel.

Choisissez le type de graphique "Courbe".

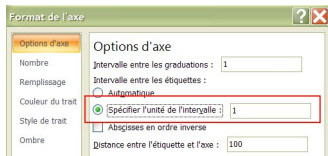
Cliquez sur le bouton OK pour valider.



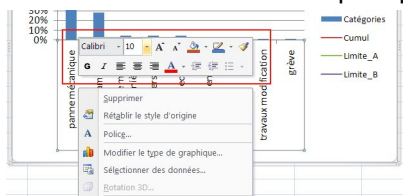
IV-C-3 - Afficher toutes les catégories sur l'axe des abscisses

Pour visualiser la liste complète des catégories sur l'axe des abscisses :

Sélectionnez l'axe des abscisses.
Clic droit.
Sélectionnez "Mise en forme de l'axe" dans le menu contextuel.
Choisissez le menu "Options de l'axe".
Cliquez sur l'option "Spécifier l'unité de l'intervalle".
Indiquez la valeur 1.

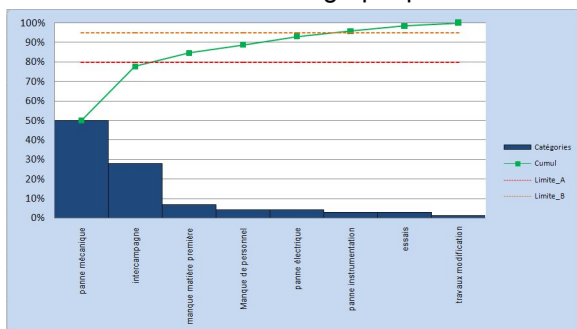


Ensuite, Cliquez sur le menu "Alignement".
Choisissez "Faire pivoter tout le texte de 270°" dans le menu déroulant "Orientation du texte".
Cliquez sur le bouton "Fermer".
Nota :
Lorsque vous faites un clic droit sur l'axe, apparait également la mini barre d'outils Excel 2007.
Utilisez ce menu contextuel pour personnaliser le format du texte.



IV-C-4 - Supprimer les espaces entre chaque donnée en abscisses

Sélectionnez la série en histogramme ('Catégories').
Clic droit.
Sélectionnez l'option "Mettre en forme une série de données".
Menu "Options des séries".
Ramenez le curseur "Largeur de l'intervalle" à 0%.
Pendant que la fenêtre de personnalisation de la série est ouverte, vous pouvez mettre en forme :
* La couleur de fond
* Les bordures
* Les types de traits et les marqueurs
Cliquez sur le bouton "Fermer" pour terminer.
La mise en forme de votre graphique est désormais terminée et vous pouvez visualiser le diagramme de Pareto.



On voit clairement qu'il faut agir en priorité sur les temps de panne mécanique et sur la durée des intercampagnes pour diminuer significativement l'immobilisation de la machine.

V - Conclusion

J'espère que les informations fournies dans cet article vous aideront à finaliser vos projets.
Adaptez et modifiez librement le classeur démo en fonction de vos besoins.

Les formules sont ici volontairement détaillées sur plusieurs colonnes mais vous pouvez regrouper les fonctions, masquer ou réarranger les colonnes de calcul intermédiaire comme bon vous semble.

VI - Liens

Le diagramme de Pareto, sur WikiPedia.

Utiliser les fonctions Excel

Optimisez vos recherches avec les fonctions Index et Equiv

La gestion des doublons dans Excel

VII - Remerciements

Je remercie toute l'équipe Office de DVP et particulièrement **Philippe Jochmans**, **Jean Ballat** , pour la relecture et la correction du tutoriel.

VIII - Téléchargement

Téléchargez le classeur démo (Testé sous Excel2002 et Excel2007).