

Analyse de données sur les différentes thématiques des journaux télévisés diffusées sur 06 chaînes françaises (TF1, France2, France3, Canal+, Arte, M6)

Projet personnel

Manuella Kevine MEBU
2023/06/21

Table des matières

1	Classement thématique des sujets de journaux télévisés (Janvier 2005 - Septembre 2020)	3
1.1	Démarche de l'analyse des données	3
1.2	ASK.....	3
1.2.1	L'objectif.....	3
1.3	PREPARE.....	3
1.3.1	Source de données	3
1.3.2	Interlocuteur.....	4
1.3.3	L'outil	4
1.4	PROCESS	4
1.4.1	Emplacement, compréhension schématique et relationnelle des fichiers	4
1.4.2	Importation, Nettoyage, Transformation, Validation et Chargement des données	5
1.5	ANALYSE.....	8
1.5.1	Contour des données	8
1.5.2	Le total des chaines, des thématiques, des sujets de journaux télévisés et la durée	9
1.5.3	Le nombre de thématiques par chaine	9
1.5.4	Le total des sujets de journaux télévisés par mois.....	9
1.5.5	Le total, la durée moyenne/ mois, le nombre moyen/ mois des sujets de journaux télévisés par thématique	10
1.5.6	Le pourcentage des sujets de journaux télévisés par chaine	11
1.5.7	La durée d'un sujet de journaux télévisés par chaine	11
1.5.8	Fichier d'Analyse	12
1.6	SHARE et ACT.....	12
1.7	Découvertes.....	12
1.7.1	Rappelons l'objectif	12
1.7.2	Découvertes.....	12
2	Annexe	13

1 Classement thématique des sujets de journaux télévisés (Janvier 2005 - Septembre 2020)

ELITEHIVE (entreprise fictive) est une chaîne française nouvellement créée, voulant connaître les différentes thématiques proposées par leurs concurrents, afin d'établir leur propre programme des sujets de journaux télévisés.

Je suis une analyste de données junior de l'équipe marketing digital. Le Manager Christian DDARK, m'a sollicité dans l'optique d'analyser l'ensemble des sujets de journaux télévisés (JT) des chaînes TF1, France2, France3, Canal+, Arte et M6, de Janvier 2005 – Septembre 2020 (données fournies par INA (Institut national de l'audiovisuel)).

Cette analyse doit pouvoir répondre à l'objectif demandé. De plus, il précise que je dois fournir un tableau de bord (TDB), à présenter à l'équipe exécutive pour prise de décisions.

Institut national de l'audiovisuel (INA) est une entreprise publique créée en 1975. INA collecte et conserve les données de programmes télé, pour le grand public, en France comme à l'étranger.

1.1 Démarche de l'analyse des données

Nous allons utiliser la démarche d'analyse de « Google Data Analytics Professional Certificate ». Qui est : « ASK », « PREPARE », « PROCESS », « ANALYSE », « SHARE » et « ACT », pour effectuer la tâche.

1.2 ASK

1.2.1 L'objectif

ELITEHIVE veut être informée sur les différentes thématiques proposées par leurs concurrents. De fait, elle veut s'en inspirer afin de mettre sur pied la sienne. Pour ce faire elle aura besoin de savoir :

- Le nombre de thématiques proposées, par mois, années
- Le temps alloué pour chaque thématique
- Les variations mensuelles et annuelles
- Etc.

1.3 PREPARE

1.3.1 Source de données

- Les données proviennent de l'Institut national de l'audiovisuel (INA)
- La source de données obtenue sur le site [data.gouv.fr](https://www.data.gouv.fr/fr/datasets/classement-thematique-des-sujets-de-journaux-televises-janvier-2005-septembre-2020/#/discussions)
<https://www.data.gouv.fr/fr/datasets/classement-thematique-des-sujets-de-journaux-televises-janvier-2005-septembre-2020/#/discussions>
- INA est un « Second-party data ». Elle obtient ses données des chaînes de télévision nationale (TF1, France2, France3, Canal+, Arte, M6)
- Nous avons (02) deux fichiers CSV, l'un porte sur le nombre de thématiques et l'autre la durée de Janvier 2005 - Septembre 2020
- Les métadonnées qui décrivent le jeu de données se trouvent sur ce lien
<http://www.inatheque.fr/publications-evenements/ina-stat/ina-stat-methodologie.html>.
- Il s'agit de données historiques de Janvier 2005 - Septembre 2020

1.3.2 Interlocuteur

- Le Manager Christian DDARK est mon interlocuteur direct. Nous aurons plusieurs séances de travail à chaque étape de l'analyse pour s'assurer de l'intégrité des données, et que cette dernière répond toujours à l'objectif escompté
- L'équipe exécutive est le client final, à qui je présenterai le résumé de mon travail

1.3.3 L'outil

- J'utilise l'ETL Power Query de MS EXCEL pour le nettoyage et la transformation de données
- Tableau croisé dynamique, fonctions, graphiques, formatage pour le tableau de bord sur EXCEL
- Un résumé sur MS Word, sur la démarche de l'analyse

1.4 PROCESS

1.4.1 Emplacement, compréhension schématique et relationnelle des fichiers

Nous avons deux fichiers CSV. L'un pour le nombre de thématiques diffusées par chaîne et par mois, et l'autre constitué de la durée de diffusion des thématiques par chaîne et par mois. Les deux fichiers contiennent 2 colonnes communes : **mois** et **thématique**.

	A	B	C	D	E	F	G	H	I
1	MOIS	THEMATIQUES	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT
2			TF1	France 2	France 3	Canal +	Arte	M6	Totaux
3	janv-05	Catastrophes	05:39:50	05:13:56	02:35:01	00:22:41	00:51:51	01:07:28	15:50:47
4	janv-05	Culture-loisirs	00:50:22	01:24:04	01:12:27	00:04:12	00:00:00	00:27:44	03:58:49
5	janv-05	Economie	00:52:25	00:31:12	00:16:39	00:01:59	00:07:59	00:11:04	02:01:18

Figure 1 : Extrait sur la durée de diffusion des thématiques

	A	B	C	D	E	F	G	H	I
	MOIS	THEMATIQUES	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT
			TF1	France 2	France 3	Canal +	Arte	M6	Totaux
	janv-05	Catastrophes	214	191	88	18	40	49	600
	janv-05	Culture-loisirs	27	42	35	4	0	23	131
	janv-05	Economie	35	18	10	1	8	11	83

Figure 2 : Extrait sur le nombre des thématiques diffusées

- Téléchargez les données, créez un dossier et placez les deux fichiers

Documents > Projet > Projet EXCEL > THEMATIQUE_SUJET_TELEWISE_DUREE > DATA SET				
	Nom	Modifié le	Type	Taille
	ina-barometre-jt-tv-donnees-mensuelles...	18/06/2023 04:36	Fichier CSV Micro...	222 Ko
	ina-barometre-jt-tv-donnees-mensuelles...	18/06/2023 04:36	Fichier CSV Micro...	116 Ko

Figure 3 : Emplacement jeux de données INA

- Faire une copie des feuilles de chaque classeur téléchargé vers un autre classeur (un seul classeur). Nous obtenons un nouveau classeur (nom : DONNEES_SUJETS_TELEWISES) avec toutes les données

févr-05	Environnem	46	25	12	0	6	7	96
ina-barometre-jt-tv-donnees-men		ina-barometre-jt-tv-donnees						
Accessibilité : vérification terminée								

Figure 4 : Regrouper les données dans un seul classeur

- 2 646 enregistrements (lignes) et 9 colonnes par feuille du classeur

- Nous avons une relation 1 – 1 pour les données dans les deux feuilles du classeur : au minimum un enregistrement de la première feuille (durée) est lié au minimum à un enregistrement de la deuxième feuille (nombre de sujets télévisés) et au maximum à un enregistrement, et inversement pareil.

1.4.2 Importation, Nettoyage, Transformation, Validation et Chargement des données

1.4.2.1 Importation des données

Ouvrir un nouveau fichier EXCEL> DONNEES > OBTENIR DES DONNEES > A PARTIR D'UN CLASSEUR EXCEL. La fenêtre ci-dessous s'affiche, choisir toutes les tables et faire un clic sur « Transformer les données »

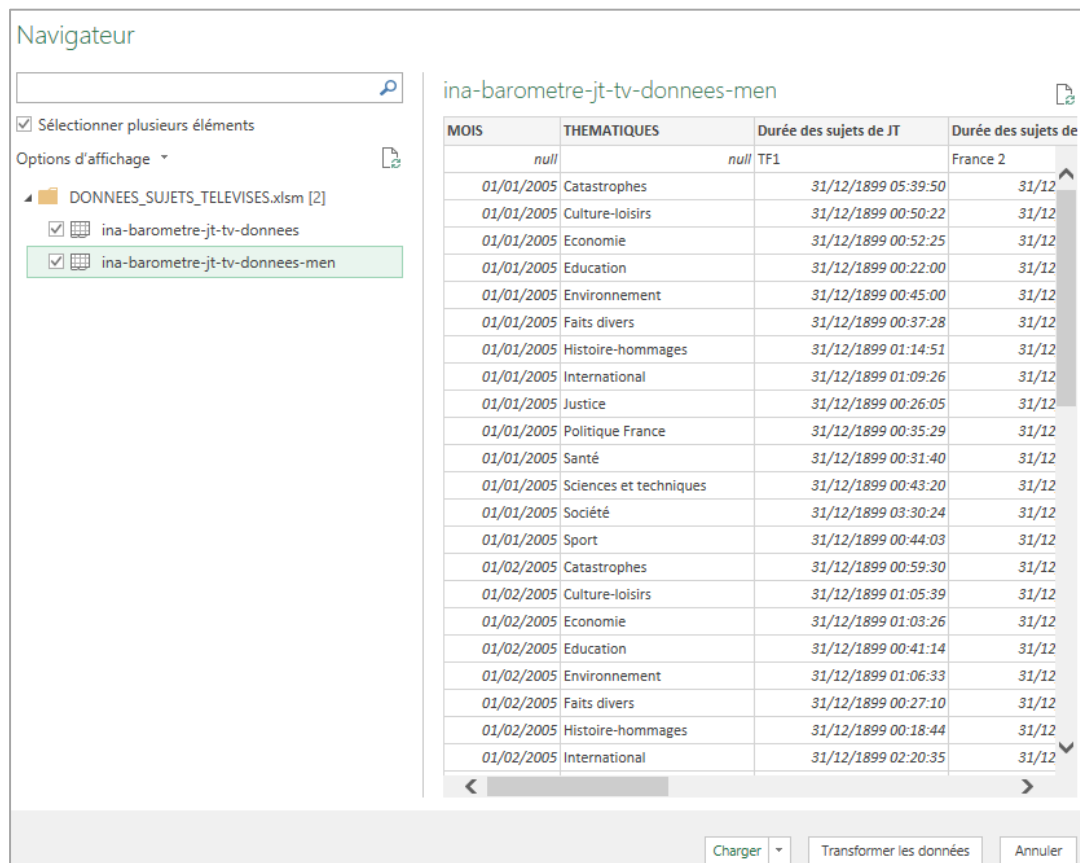


Figure 5 : Navigateur pour sélectionner les tables du classeur à importer dans l'éditeur Power Query

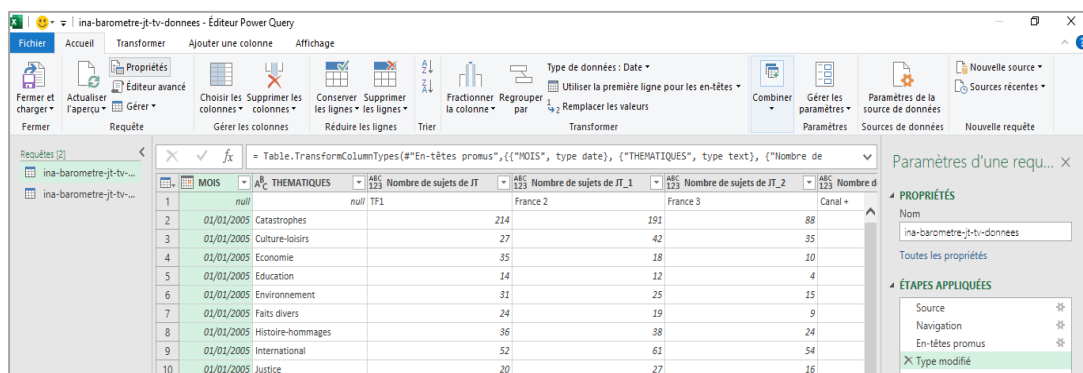


Figure 6 : L'ouverture de l'éditeur Power Query après importation des données pour transformation

1.4.2.2 Transformation des données avec l'éditeur Power query

Cette partie consiste à faire le nettoyage, la transformation, les validations des données et la fusion des deux tables importées.

1	MOIS	THEMATIQUES	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée des sujets de JT	Durée
2		TF1		France 2	France 3	Canal +	Arte
3	01/01/2005	Catastrophes	31/12/1899 05:39:50	31/12/1899 05:13:56	31/12/1899 02:35:01	31/12/1899 00:22:41	
4	01/01/2005	Culture-loisirs	31/12/1899 00:50:22	31/12/1899 01:24:04	31/12/1899 01:12:27	31/12/1899 00:04:12	
5	01/01/2005	Economie	31/12/1899 00:52:25	31/12/1899 00:31:12	31/12/1899 00:16:39	31/12/1899 00:01:59	
6	01/01/2005	Education	31/12/1899 00:22:00	31/12/1899 00:20:42	31/12/1899 00:09:46	31/12/1899 00:02:18	
7	01/01/2005	Environnement	31/12/1899 00:45:00	31/12/1899 00:32:55	31/12/1899 00:23:09	31/12/1899 00:00:41	
8	01/01/2005	Faits divers	31/12/1899 00:37:28	31/12/1899 00:27:43	31/12/1899 00:15:50	31/12/1899 00:01:22	

Figure 7 : Table des durées importée à l'état brute

1	MOIS	THEMATIQUES	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nombre de sujets de JT	Nomb
2		TF1		France 2	France 3	Canal +	Arte
3	01/01/2005	Catastrophes	214	191	88	18	
4	01/01/2005	Culture-loisirs	27	42	35	4	
5	01/01/2005	Economie	35	18	10	1	
6	01/01/2005	Education	14	12	4	3	
7	01/01/2005	Environnement	31	25	15	1	
8	01/01/2005	Faits divers	24	19	9	1	

Figure 8 : Table de nombres de sujets télévisés

1.4.2.2.1 Les différentes transformations

- Déplacer vers le haut les premières lignes de chaque table en en-têtes
- Renommer les deux premières colonnes des deux tables : **Mois** et **Thématique**
- Supprimer les deux dernières colonnes (totaux). Nous pouvons les avoir à partir d'un TCD
- Nous n'avons pas de doublons et pas des valeurs nulles
- Changer les colonnes de types **date heure** à **date**
- Pivoter les colonnes quantitatives des deux tables, renommer les colonnes engendrées

1	Mois	Thématique	Chaîne	N° de sujets télévisés
2	01/01/2005	Catastrophes	TF1	214
3	01/01/2005	Catastrophes	France 2	191
4	01/01/2005	Catastrophes	France 3	88
5	01/01/2005	Catastrophes	Canal +	18
6	01/01/2005	Catastrophes	Arte	40
7	01/01/2005	Culture-loisirs	TF1	27
8	01/01/2005	Culture-loisirs	France 2	42
9	01/01/2005	Culture-loisirs	France 3	35
10	01/01/2005	Culture-loisirs	Canal +	4
11	01/01/2005	Culture-loisirs	Arte	0
12	01/01/2005	Culture-loisirs	M6	23
13	01/01/2005	Economie	TF1	35
14	01/01/2005	Economie	France 2	18
15	01/01/2005	Economie	France 3	10
16	01/01/2005	Economie	Canal +	1

Figure 9 : Les différentes étapes appliquées à la table des sujets télévisés

1	Mois	Thématique	Chaîne	Durée (hh/mm/ss)
2	01/01/2005	Catastrophes	TF1	05:39:50
3	01/01/2005	Catastrophes	France 2	05:13:56
4	01/01/2005	Catastrophes	France 3	02:35:01
5	01/01/2005	Catastrophes	Canal +	00:22:41
6	01/01/2005	Catastrophes	Arte	00:51:51
7	01/01/2005	Culture-loisirs	M6	01:07:28
8	01/01/2005	Culture-loisirs	TF1	00:50:22
9	01/01/2005	Culture-loisirs	France 2	01:24:04
10	01/01/2005	Culture-loisirs	France 3	01:12:27
11	01/01/2005	Culture-loisirs	Canal +	00:04:12
12	01/01/2005	Culture-loisirs	Arte	00:00:00
13	01/01/2005	Culture-loisirs	M6	00:27:44
14	01/01/2005	Economie	TF1	00:52:25
15	01/01/2005	Economie	France 2	00:31:12
16	01/01/2005	Economie	France 3	00:16:39
17	01/01/2005	Economie	Canal +	00:01:59

Figure 10 : Les différentes étapes appliquées à la table des durées de sujets télévisés

- Convertir la colonne Durée (hh/mm/ss) en minute

Figure 11 shows the Power Query interface with the formula bar set to: `= Table.RenameColumns(#"Colonnes supprimées1",{{"Minutes", "Minute"}})`. The table displays data for various months (01/01/2005), themes (Catastrophes, Culture-loisirs, Economie, Education), and channels (TF1, France 2, France 3, Canal +, Arte, M6). The 'Durée (hh/mm/ss)' column is being converted to 'Minute'.

Figure 11 : Colonne Durée (hh/mm/ss) en minute

- Fusionner (combiner) les deux tables : la relation est de 1-1, vue précédemment

Figure 12 shows the Power Query interface with the formula bar set to: `= Table.ExpandTableColumn(#"Requêtes fusionnées", "ina-barometre-jt-tv-donnees-men", {"Durée (hh/mm/ss)", "Minute"})`. The table now includes an additional column 'N° de sujets télévisés'.

Figure 12 : Tables combinées

Charger la table combinée, nettoyer, transformer et valider dans le classeur EXCEL

A	B	C	D	E	F
Mois	Thématique	Chaîne	N° de sujets télévisés	Durée (hh/mm/ss)	Minute
01/01/2005	Catastrophes	TF1	214	05:39:50	339,8333333
01/01/2005	Catastrophes	France 2	191	05:13:56	313,9333333
01/01/2005	Catastrophes	France 3	88	02:35:01	155,0166667
01/01/2005	Catastrophes	Canal +	18	00:22:41	22,6833333
01/01/2005	Catastrophes	Arte	40	00:51:51	51,85
01/01/2005	Catastrophes	M6	49	01:07:28	67,4666667
01/01/2005	Culture-loisirs	TF1	27	00:50:22	50,3666667
01/01/2005	Culture-loisirs	France 2	42	01:24:04	84,0666667
01/01/2005	Culture-loisirs	France 3	35	01:12:27	72,45
01/01/2005	Culture-loisirs	Canal +	4	00:04:12	4,2
01/01/2005	Culture-loisirs	Arte	0	00:00:00	0
01/01/2005	Culture-loisirs	M6	23	00:27:44	27,7333333
01/01/2005	Economie	TF1	35	00:52:25	52,4166667
01/01/2005	Economie	France 2	18	00:31:12	31,2
01/01/2005	Economie	France 3	10	00:16:39	16,65
01/01/2005	Economie	Canal +	1	00:01:59	1,9833333
01/01/2005	Economie	Arte	8	00:07:59	7,9833333
01/01/2005	Economie	M6	11	00:11:04	11,0666667
01/01/2005	Education	TF1	14	00:22:00	22
01/01/2005	Education	France 2	12	00:20:42	20,7

Figure 13 shows the Excel spreadsheet with the combined data. The table is named 'nombres_durées_sujetsTV'.

Figure 13 : Jeu de données transformé pour assurer l'intégrité des données, et chargé dans EXCEL

1.5 ANALYSE

L'analyse se fait sur MS EXCEL. J'utilise ces différentes fonctionnalités :

- Le tableau croisé dynamique (TCD)
- Les graphiques et leurs personnalisations
- Les fonctions comme : **sous.total**, **nb.si** et **division et rechercheV**
- Les formes pour le design du tableau de bord

L'analyse se porte sur les axes suivants :

- Faire une analyse sur le contour des données
- Le total des chaînes, des thématiques, des sujets de journaux télévisés et la durée
- Le nombre de thématiques par chaîne
- Le total des sujets de journaux télévisés par mois
- Le total, la durée moyenne par mois, le nombre moyen par mois des sujets de journaux télévisés par thématique
- Le pourcentage des sujets de journaux télévisés par chaîne
- La durée d'un sujet de journaux télévisés par chaîne

1.5.1 Contour des données

Cette partie consiste à faire une analyse rapide pour répertorier les valeurs max, min et l'étendu des données.

Nous devons garder à l'esprit qu'un enregistrement est constitué du mois, la thématique, la chaîne, le nombre des sujets de journaux télévisés et la durée.

1.5.1.1 Maximum

Mois	Thématique	Chaîne	N° de sujets télévisés	Durée (hh/mm/ss)	Durée (min)
01/03/2020	Santé	TF1	569	21:13:33	1273,55

Figure 14 : au mois (toutes les années confondues), TF1 fait un total de 569 sujets de JT avec une durée de 1 273,55 minutes

1.5.1.2 Minimum

La valeur minimale de la durée et le nombre des sujets télévisés est de zéro pour un mois d'une année. Ceci pour toutes les thématiques, chaîne et années confondues.

809 enregistrements sont concernés (nous avons en total 15 875 enregistrements). Le jeu de données sera fourni en copie de ce rapport.

1.5.1.3 Conclusion

Nous avons des données manquantes et ceci affectera les calculs statistiques, tels que la moyenne d'une colonne quantitative par mois ou par année. Il faudra informer l'équipe exécutive de ces données manquantes.

Les données manquantes peuvent expliquer les chutes dans les graphiques qui représentent la tendance dans le temps, par exemple le nombre des sujets de journaux télévisés allant d'un mois à un autre.

- 1.5.2 Le total des chaines, des thématiques, des sujets de journaux télévisés et la durée
Le TDB offre la possibilité de filtrer les sujets de journaux télévisés et la durée par chaines, thématiques et mois.



Figure 15 : Les grandes valeurs

- 1.5.3 Le nombre de thématiques par chaine
Nous pouvons filtrer ces valeurs par mois et thématique sur le TDB.

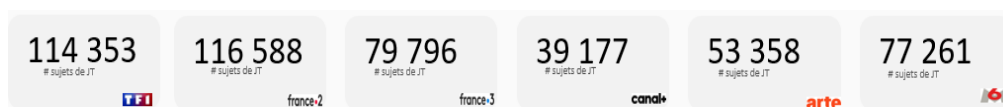


Figure 16 : Thématique par chaine

- 1.5.4 Le total des sujets de journaux télévisés par mois
En Mars, Mai et Juillet les sujets de JT sont les plus élevés, d'une valeur approximative de 42 000. A partir d'Aout jusqu'en Décembre, nous observons une baisse des sujets de JT, qui se stabilise à 37 000.

Nous pouvons filtrer ce graphique par année, chaine et thématique sur le TDB.

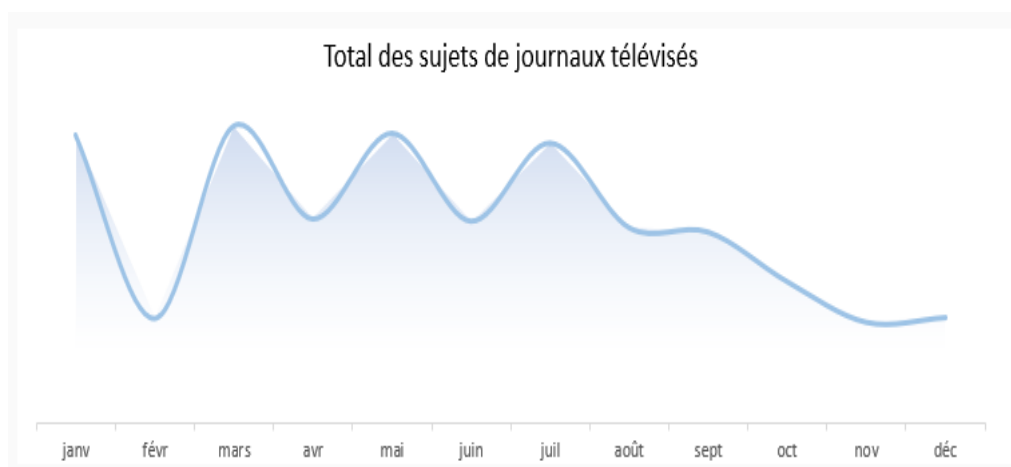


Figure 17 : Total des sujets de JT/ mois

- 1.5.5 Le total, la durée moyenne par mois, le nombre moyen par mois des sujets de journaux télévisés par thématique
- Par thématique, nous avons la moyenne par mois des sujets de JT ainsi que la durée. Nous pouvons filtrer par chaîne ou par année pour une analyse fine.

Les top 05 thématiques sont : société, International, Economie, Politique France, Culture-loisirs.

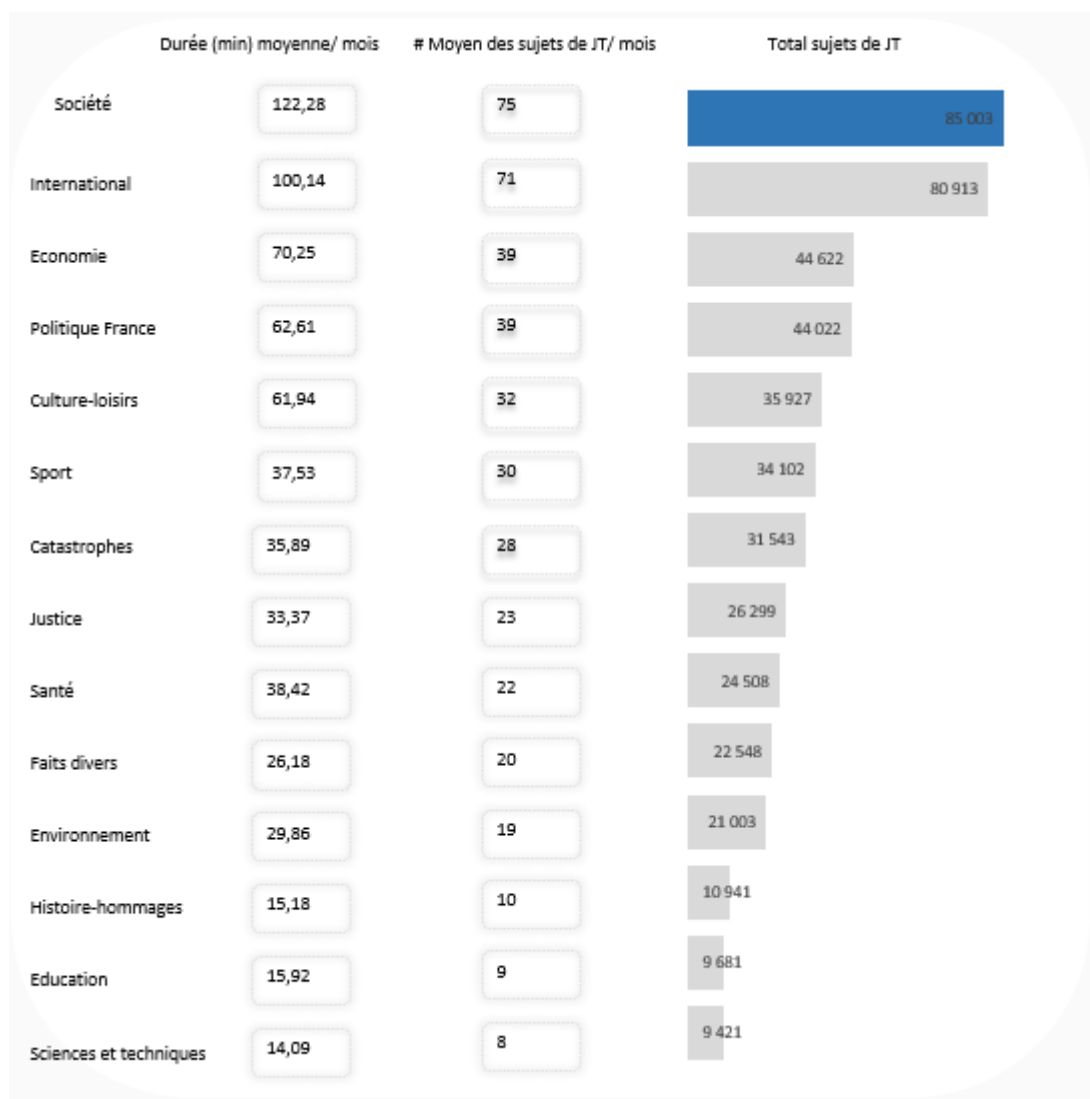


Figure 18 : Le total, la moyenne des sujets de JT et durée par mois pour chaque thématique

1.5.6 Le pourcentage des sujets de journaux télévisés par chaîne

Nous observons que TF1 et France2 offrent plus des sujets télévisés que les autres chaînes.

Le TDB offre la possibilité de filtrer par thématique et mois.

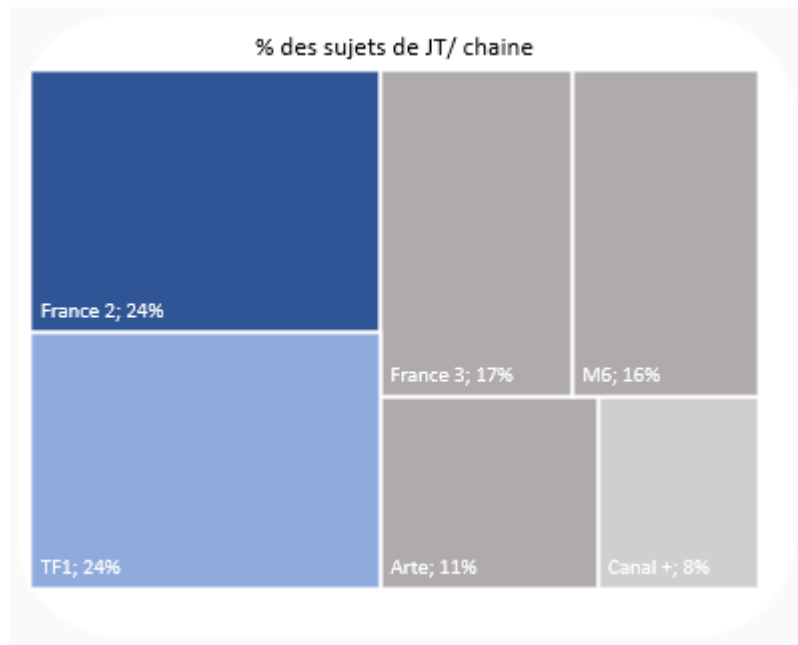


Figure 19 : % des sujets de JT/ chaîne

1.5.7 La durée d'un sujet de journaux télévisés par chaîne

Cette image illustre la durée (minute) d'un sujet de JT par chaîne. À partir du TDB, nous pouvons filtrer par thématique pour une analyse fine.



Figure 20 : Durée (minute) d'un sujet de JT

1.5.8 Fichier d'Analyse

Nous avons un fichier EXCEL (nom : « nombres_durées_sujetsTV »), contenant dans chaque feuille (03) les informations ci-dessous :

- La base de données : la table finale nettoyée, transformée et validée en utilisant Power Query
- Le traitement : ici se trouve les différents TCD et les calculs créés pour les visuels du TDB
- Tableau de bord : le résultat final de l'analyse

14	01/01/2005	Economie	TF1	35	00:52:25	52,41666667
15	01/01/2005	Economie	France 2	18	00:31:12	31,2
16	01/01/2005	Economie	France 3	10	00:16:39	16,65
17	01/01/2005	Economie	Canal +	1	00:01:59	1,983333333
18	01/01/2005	Economie	Arte	8	00:07:59	7,983333333
19	01/01/2005	Economie	M6	11	00:11:04	11,06666667
20	01/01/2005	Education	TF1	14	00:22:00	22
21	01/01/2005	Education	France 2	12	00:20:42	20,7
22	01/01/2005	Education	France 3	4	00:09:46	9,766666667
23	01/01/2005	Education	Canal +	3	00:02:18	2,3
24	01/01/2005	Education	Arte	3	00:04:35	4,583333333
25	01/01/2005	Education	M6	8	00:13:10	13,16666667
<div> <div>◀ ▶</div> <div>nombres_durées_sujetsTV</div> <div>TRAITEMENT</div> <div>TABEAU DE BORD</div> <div>⊕</div> </div>						

Figure 3 : Extrait de la présentation du fichier d'analyse MS EXCEL

1.6 SHARE et ACT

Cette partie consiste à présenter son travail final à l'équipe exécutive pour prise de décision. Etant donné que c'est un projet personnel, nous ne pouvons pas prendre en compte ces étapes. Je parlerai de mes découvertes au paragraphe suivant.

1.7 Découvertes

1.7.1 Rappel sur l'objectif

ELITEHIVE veut être informée sur les différentes thématiques proposées par leurs concurrents. De fait, elle veut s'en inspirer afin de mettre sur pied la sienne.

1.7.2 Découvertes

- Nous avons 14 thématiques proposées par leurs concurrents (TF1, France2, France3, Canal+, Arte et M6)
- Le Top 05 des thématiques sont : Société, International, Economie, Politique France et Culture-loisirs
 - Ici, le nombre moyen des sujets de JT varie entre 75 – 32 par mois
 - Toujours pour ce top, la durée moyenne des sujets de JT varie entre 61 – 122 minutes par mois
- Les mois où nous avons le plus des sujets de JT sont : Mars, Mai et Juillet
- France2 (24%) et TF1 (24%) proposent un plus grand nombre des sujets de JT par rapport aux autres chaînes
- De façon générale, la durée d'un sujet de JT varie entre 1 – 2 minutes par chaîne

2 Annexe

TDB : Tableau de bord

INA : Institut national de l'audiovisuel

JT : Journaux télévisés

J'espère que vous avez aimé le document. Je suis ouverte aux commentaires.

Contact : mebu.kevin@yahoo.com