Mais quelles sont donc les thématiques favorites de nos JT ?

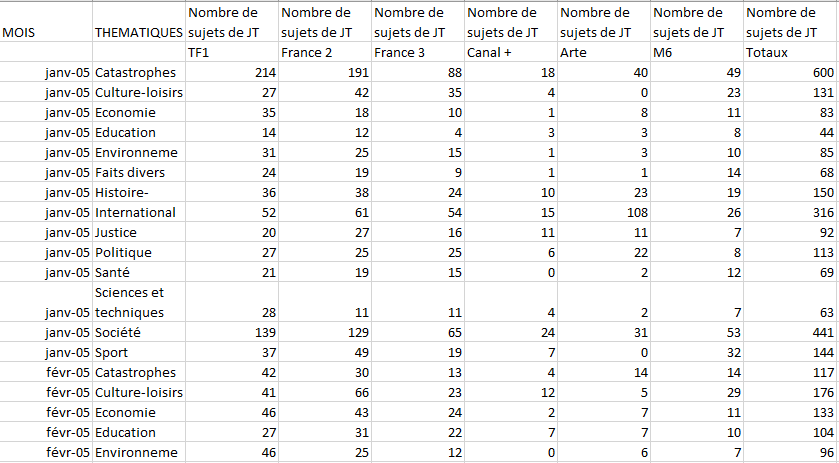
De nos jours, un bon nombre de chaînes de télévision possèdent un journal télévisé qu’ils diffusent aux heures de grande écoute. Ces JT traitent de différents sujets et chaque journal possède une ligne éditoriale selon l’ADN de la chaîne. Mais quelles sont les lignes éditoriales de ces JT ? Ont-elles évolué avec le temps ? C’est ce que nous allons voir dans cet article...juste après... le rappel des titres ! (On me dit dans l’oreillette que c’est inutile...alors c’est parti !)

Pour observer ces possibles évolutions, nous avons décidé dans un premier temps de travailler sur des données open source issues du site [data.gouv](https://www.data.gouv.fr/fr/datasets/classement-thematique-des-sujets-de-journaux-televises-janvier-2005-septembre-2020/). Ces données représentent le nombre de sujets diffusés par mois suivant la chaîne entre janvier 2005 et septembre 2020. Ces sujets sont aussi triés suivant leur thématique : culture, politique, catastrophes…L’analyse sera effectuée en Python.

# Présentation des données

C’est un jeu de données statique, importé sous format csv dont la dernière date présente est antérieure à aujourd’hui.

Dans notre base importée, nous comptons 2646 lignes et 9 variables. Nous obtenons donc la table suivante :



Les deux premières variables sont de type qualitatif. Il s’agit du mois analysé ainsi que de la thématique traitée sur ce mois. Quant aux autres variables (de type quantitatif), elles représentent le nombre de sujets par thématique traité dans les JT par chaîne. Les chaines TV analysées sont les plus populaires : TF1, France2, France3, Canal+, Arte, M6. Dans le jeu de données, nous avons une variable qui a été ajoutée pour comptabiliser le nombre de sujets pour l’ensemble de toutes ces chaînes.

# Nos objectifs

Ainsi, à travers ces données, nous avons souhaité analyser les thématiques traitées par chaque chaîne suivant la période.

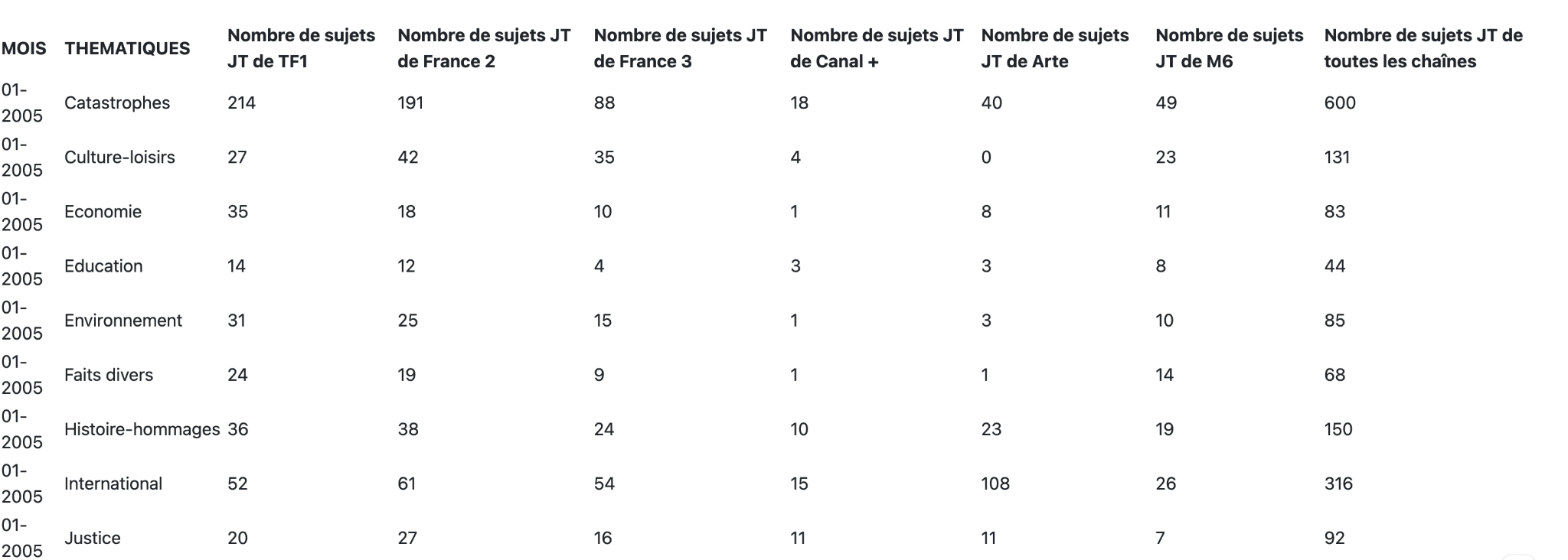
Avant de commencer à coder le Dashboard, nous avons défini les différents graphiques et informations qui nous paraissaient intéressantes et pertinentes à restituer. Nous avons donc retenu les points suivants :

* **Le nombre de sujets de JT traités par chaînes** : Cela avait pour but de voir si une chaîne monopolise les sujets diffusés ou non par rapport aux autres chaînes.
* **Le pourcentage de chaque thématique** : Cela a été fait dans l’intégralité des chaînes et suivant chacune. L’objectif était de voir principalement suivant la chaine quelle était la thématique de JT majoritaire sans distinction de période.
* **L’évolution de la part du nombre de sujets par thématique au cours du temps pour toutes les chaînes réunies** : Ici, nous souhaitions voir si au cours du temps, il y avait des périodes qui traitaient davantage un sujet plutôt qu’un autre. Nous avons, après réflexion, décidé de ne pas afficher le nombre car il n’était pas aussi représentatif que le pourcentage.
* **La répartition par thématique de la part de sujets suivant la chaîne** : L’idée était d’avoir une vision globale du pourcentage que représente chaque thème dans chacune des chaînes diffusant des JT.

# Réalisation de nos objectifs

Pour donner suite aux objectifs que nous nous étions fixés, nous avons tenté de tous les réaliser. Nous avons aussi souhaité diversifier les types de représentation de données mis à notre disposition (diagramme en barres, diagramme circulaire, KPI, …)

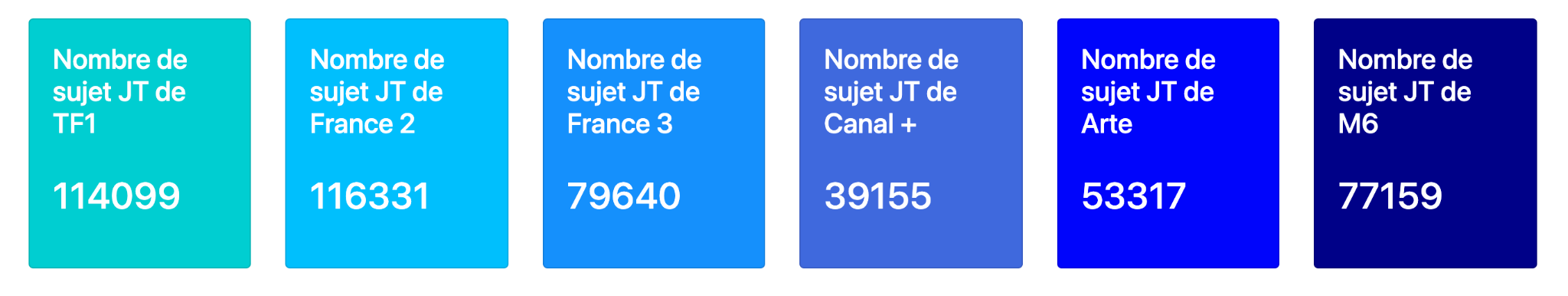
Ainsi, pour débuter, nous avons affiché un tableau affichant les 100 premières lignes du jeu de données de manière brute, sans mise en forme. C’est que nous pouvons voir sur la capture d’écran ci-dessous :



Puis, à partir de là, nous avons commencé à implémenter des graphiques petit à petit. Nous avons donc commencé par afficher les valeurs du nombre de sujets de JT diffusés suivant la chaîne sous forme de KPI. Cela nous semblait être la représentation graphique la mieux adaptée. En effet, nous n'avons affiché qu’une seule valeur pour chaque chaîne. Pour récupérer le nombre de sujets traités par chaque chaîne nous avons donc écrit le code suivant :

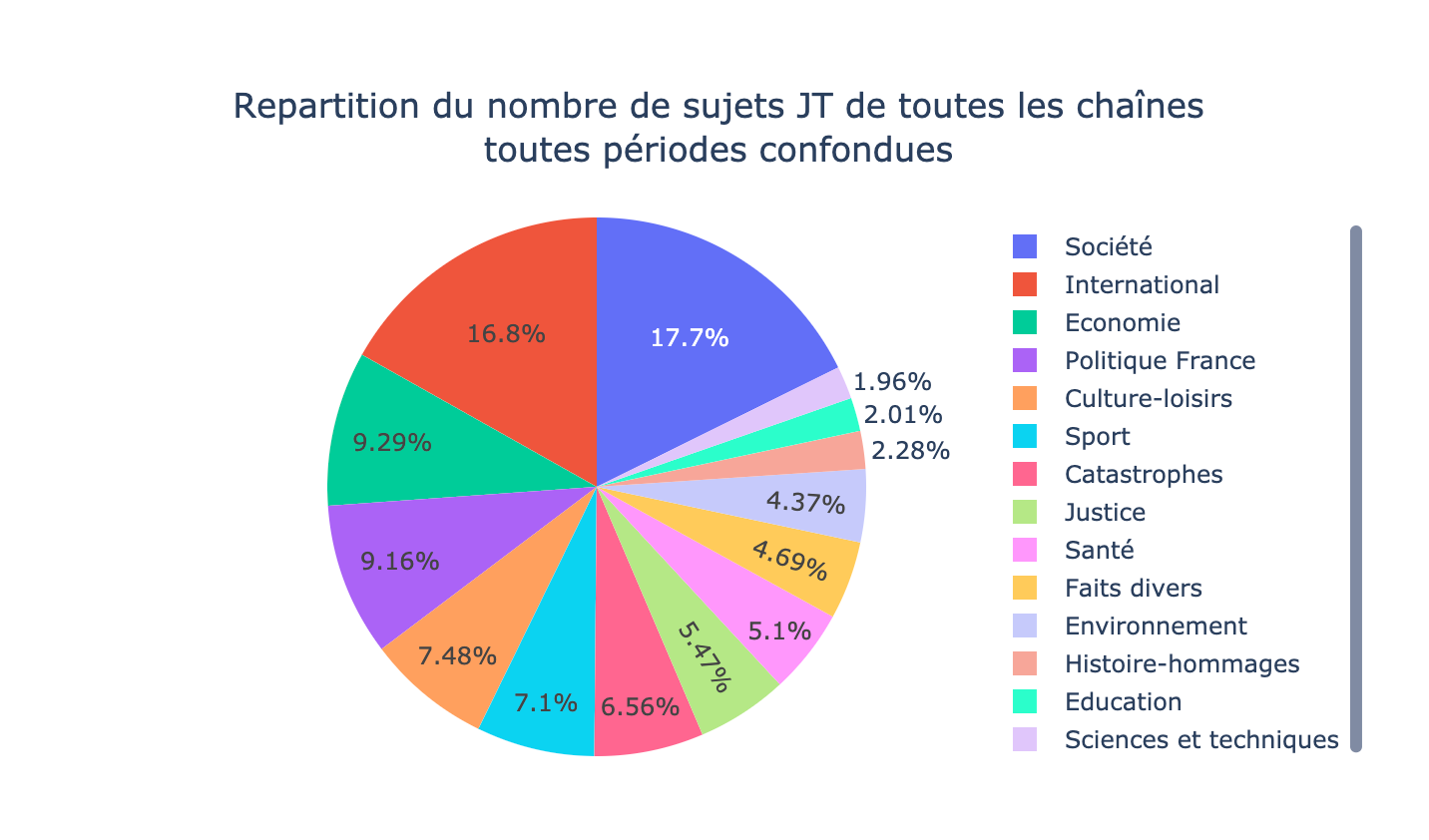


Cela est répété pour chaque chaîne de manière distincte. Ainsi, grâce à cela, nous avons pu obtenir l’affichage suivant :



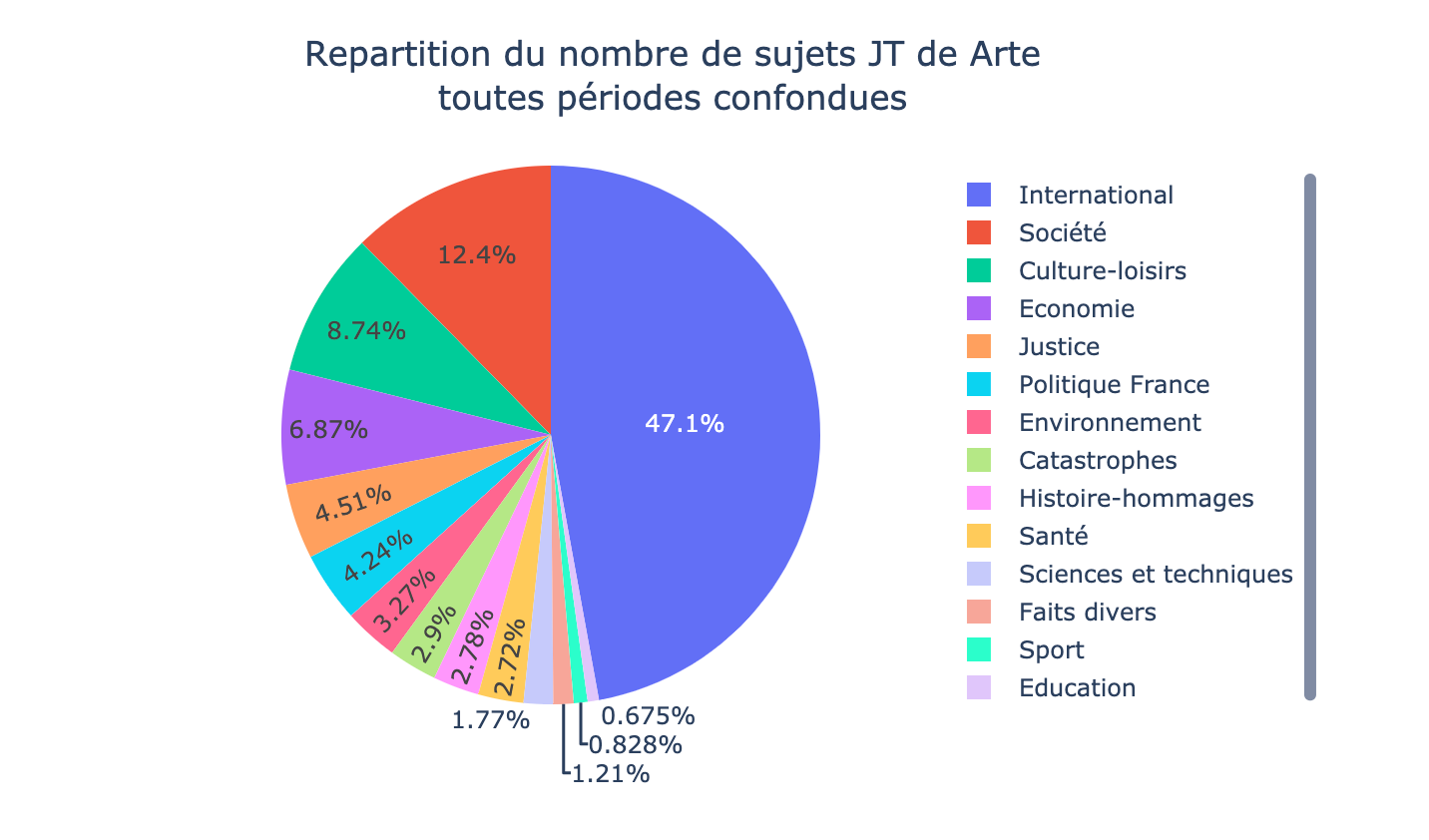
On peut ainsi voir que TF1 et France 2 ont traité davantage de sujets de JT que les autres chaînes. La chaîne qui a diffusé le moins de JT entre 2005 et septembre 2020 est Canal +. En faisant de plus amples recherches, cette observation s’explique par le fait que TF1 et France 2 ont deux JT journaliers qui durent approximativement 40 mn chacun. Les autres chaînes comme Canal + par exemple, n’avait qu’un JT de 15 mn, deux fois par jour jusqu’en 2016 (année où le JT de Canal+ s’est arrêté). L’explication est similaire pour les autres chaînes (les durées de ces JT diffèrent à celles de TF1 et France 2).

Concernant l’affichage du pourcentage de chaque thématique au sein du jeu de données par chaîne, nous avons choisi de représenter cette information sous forme de diagramme circulaire. Cela résulte du fait que nous disposions d’un nombre raisonnable de thématiques (13) et que nous voulions la part de chaque thématique dans l’ensemble de ces dernières. Notons aussi que c’est le graphique qui nous a demandé le plus de temps, car il s’agissait du tout premier élément de visualisation que nous codions seul. Malgré le temps passé, nous avons obtenu le résultat souhaité, qui est le suivant :



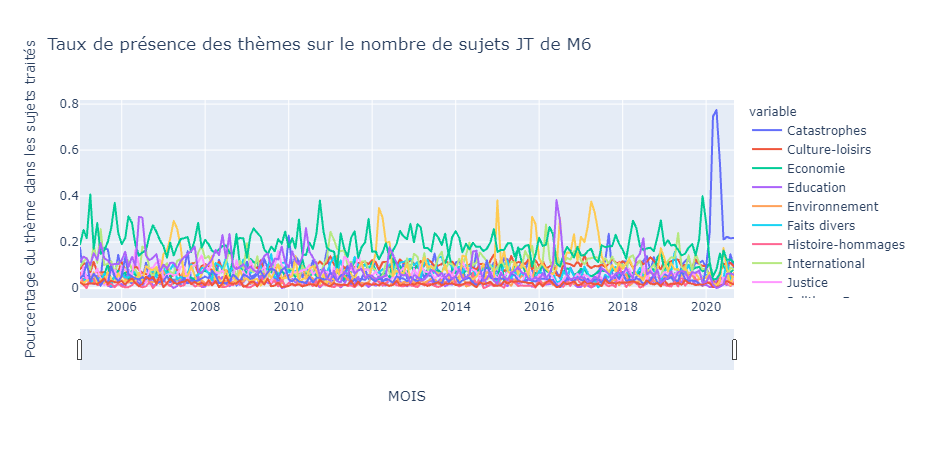
Ainsi, nous pouvons voir que sur l’ensemble des chaînes, les deux sujets les plus traités concernent la société et l'international et ils représentent respectivement 17,7% et 16,8% des diffusions. A l’inverse, on peut voir que les sujets les moins présentés sont les sciences et techniques, l'éducation et l’histoire-hommage.

De plus, ce graphique est filtré suivant la chaîne TV grâce à une liste déroulante qui appelle les différentes colonnes numériques de notre jeu de données (soit les différentes chaînes TV). Ainsi on peut comparer plusieurs chaînes et s’apercevoir que pour la majorité d’entre elles, on retrouve les mêmes ordres de grandeur pour la part de chaque thématique. Une chaîne se distingue des autres par rapport à la répartition des diffusions des thèmes au JT : Arte. La part de chaque thématique est représenté sur le graphique suivant :

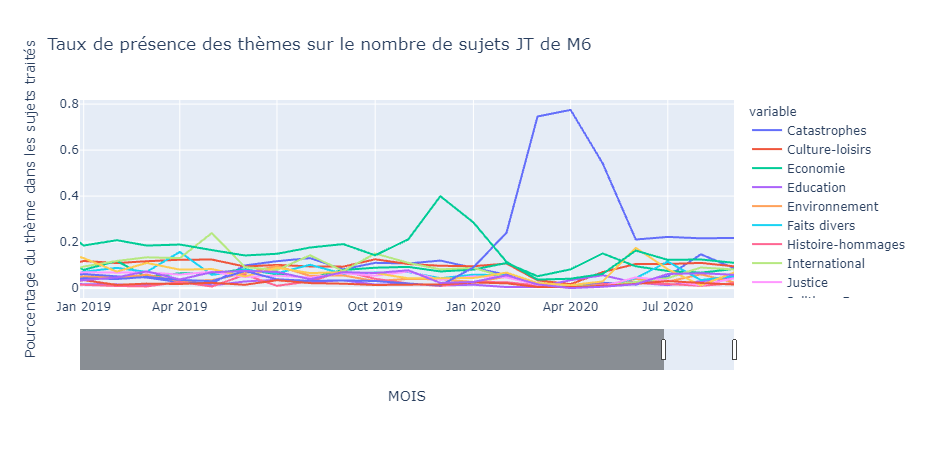


Ainsi, comme on peut le voir sur cette chaîne, près de la moitié des sujets diffusés au JT concerne l’international. Cela n’est pas étonnant car Arte a toujours été réputée pour être une chaîne publiant des reportages sur le monde, des documentaires en langues étrangères…

Nous avons ensuite réalisé un graphique de série temporelle. Comme dans notre base de données, nous disposions de données concernant le mois et l'année d’étude, nous nous en sommes servi afin de représenter l’évolution des thématiques dans le temps. Ce graphique nous a demandé de bien formater nos données, de faire un pivot\_table() dans le but d’avoir les bonnes données pour l’affichage. Nous avons aussi dû regrouper nos données suivant les thématiques et calculer leur pourcentage pour chaque mois. Nous avons obtenu le graphique suivant pour M6 :



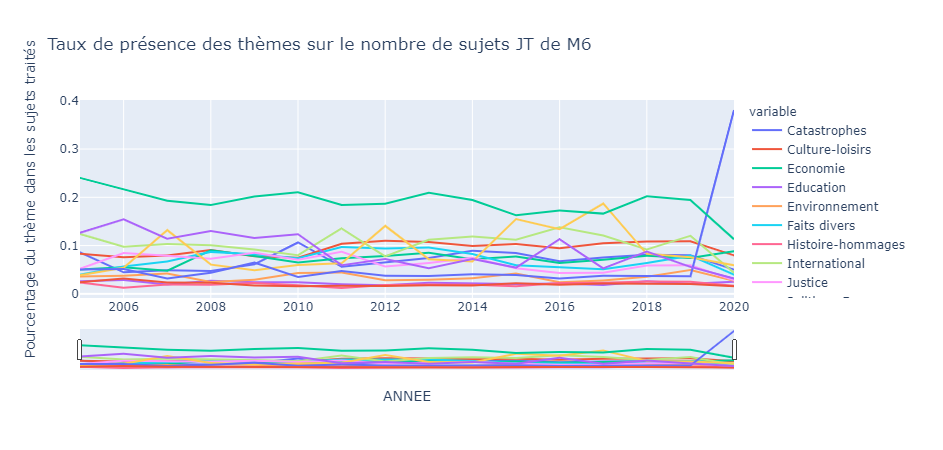
La barre située en dessous nous permet de naviguer librement sur le graphique et réaliser un focus sur une période précise. Par exemple, en zoomant sur les années 2019 et 2020, nous obtenons le résultat suivant :



Ainsi, on peut mieux voir les données. On peut ainsi se rendre compte qu' entre mars 2020 et mai 2020, plus de 80% des sujets diffusés au JT parlaient de la santé. Cela est en réalité dû à l’apparition du COVID-19, du taux de mortalité élevé à cause de ce virus, du confinement… Ce pic inhabituel, que l’on ne retrouve nul par ailleurs sur la courbe, est dû à cette période exceptionnelle.

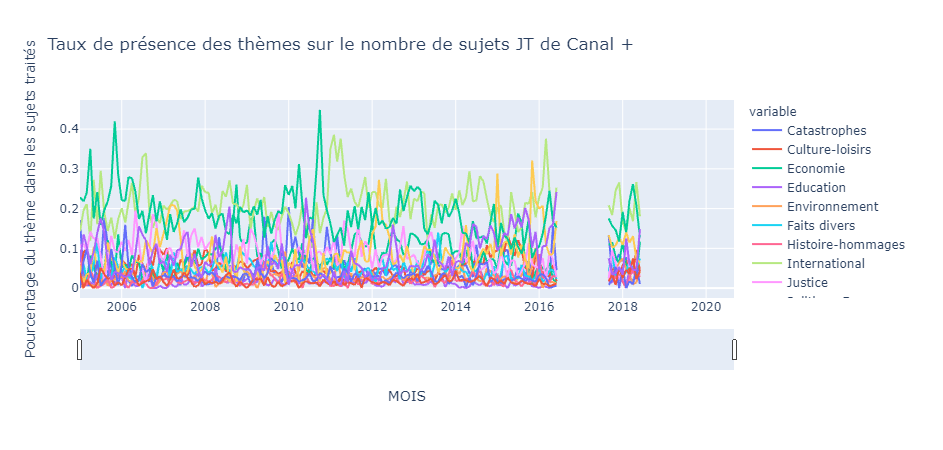
De plus, ce graphique peut être filtré par chaine TV mais aussi suivant la sélection par mois ou par année. Cela permet de lisser la courbe de manière annuelle. Le graphique est plus lisible mais ne donne pas autant d'informations que l’affichage mensuel. Cela se fait grâce au bouton suivant :





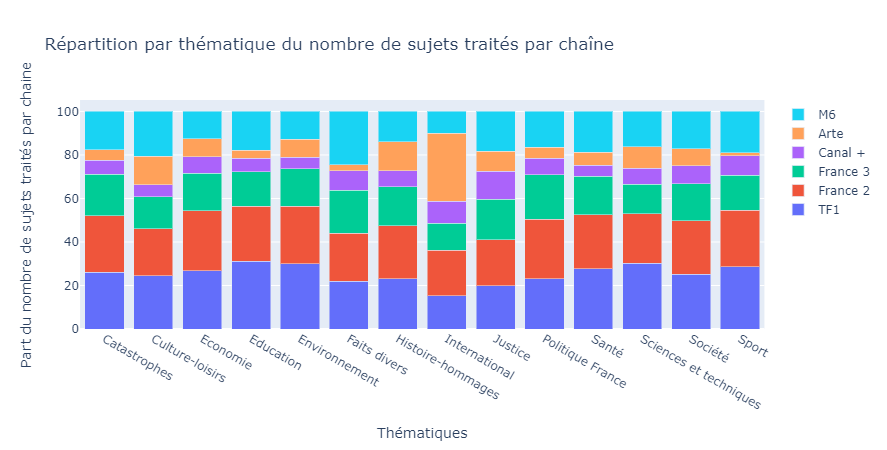
Avant de passer à la description d’un autre graphique, nous aimerions décrire davantage les informations pertinentes que nous avons pu observer sur ce graphique :

* De manière générale, le pourcentage de chaque thématique diffusé au JT est semblable. Le thème dont on parle le plus de manière mensuelle est la société. Il s’agit d’un thème relativement large qui englobe pas mal de choses, qui sont communes à la vie que l’on même tous les jours. Étant un sujet plutôt généraliste et au centre des discussions de manière quotidienne c’est bon pour cela que nous notons un pourcentage plus élevé de diffusion de cette catégorie de manière globale.
* Nous avons aussi pu noter que tous les 5 ans (2007, 2012…) on avait un pique de diffusions de sujet portant sur la politique au JT. Après avoir regardé les dates, tout semble correspondre aux périodes d'élections. Ces dernières étant très médiatisées, il est donc normal d’obtenir ces irrégularités sur la courbe.
* Nous avons aussi noté la présence d’une augmentation du nombre de sujets au JT concernant le sport tous les 2 ans environ. Cela semble correspondre aux périodes de coupe du monde ou de l’Euro. Étant là aussi des sujets très médiatisés, il est donc logique de voir apparaître ces hausse dans la courbe.
* En explorant notre graphique, nous sommes arrivés sur le cas suivant :



Comme on peut le voir, on note une absence de données pour l’année 2017 et au-delà de 2018. Nous nous sommes donc posés la question pour savoir si ce résultat était normal ou non. Après plusieurs recherches, nous nous sommes rendus compte que le Journal de Canal+ a été retiré en 2016 puis rajouté pour seulement une année. Cela explique donc ces creux dans le graphique.

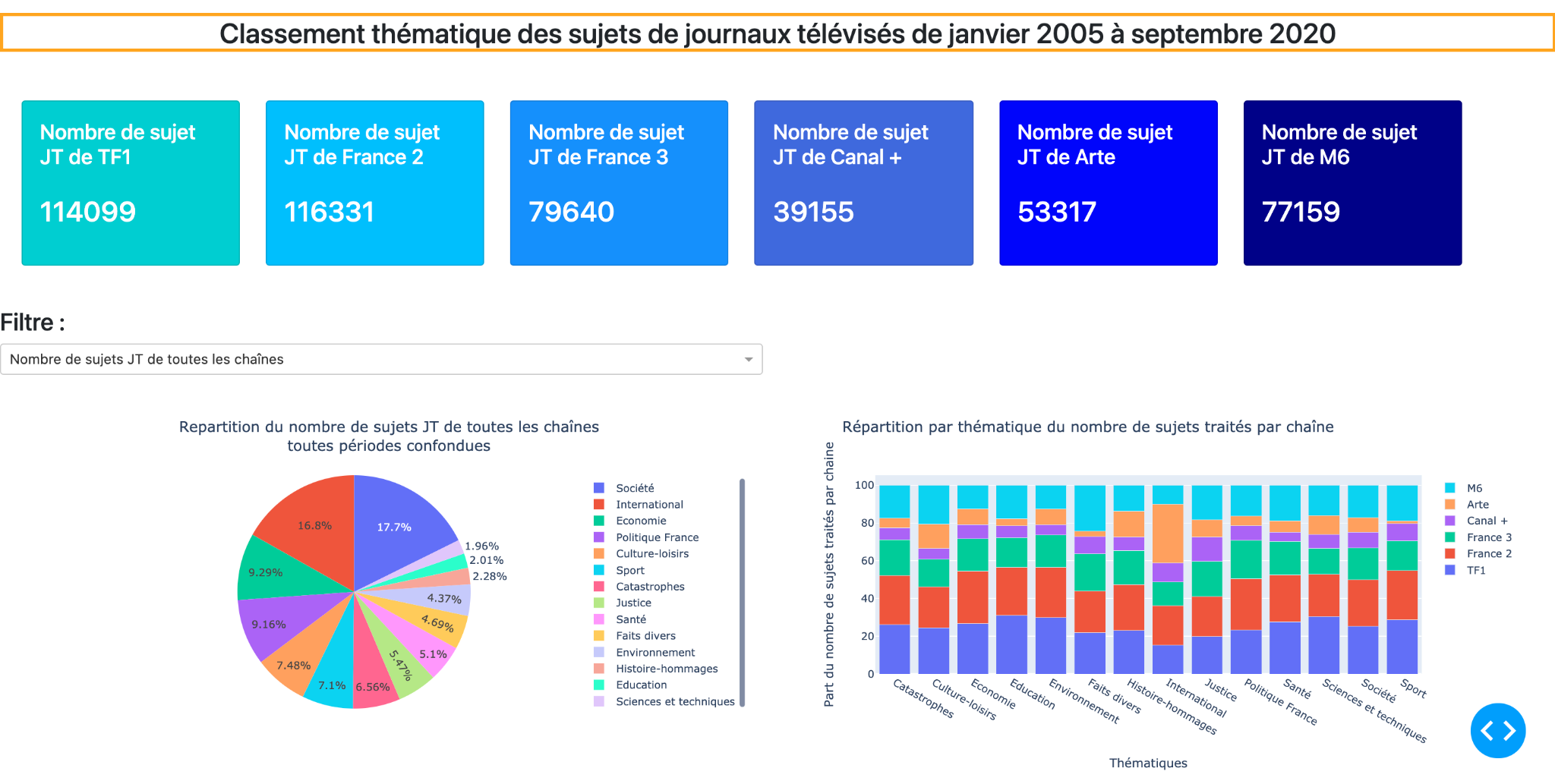
Nous avons ensuite réalisé un dernier graphique nous permettant de représenter la part de chaque chaîne suivant la thématique. Pour ce graphique, nous souhaitions afficher pour une thématique à la fois, la part que représentaient les sujets diffusées au JT pour TF1, France2, France3, Canal+, Arte, M6. Nous avons donc décidé de confectionner un graphique en barre empilée car cela représentait bien la part de chaque chaîne et grâce aux couleurs, le résultat était visible très facilement. Le résultat de ce graphique est le suivant :

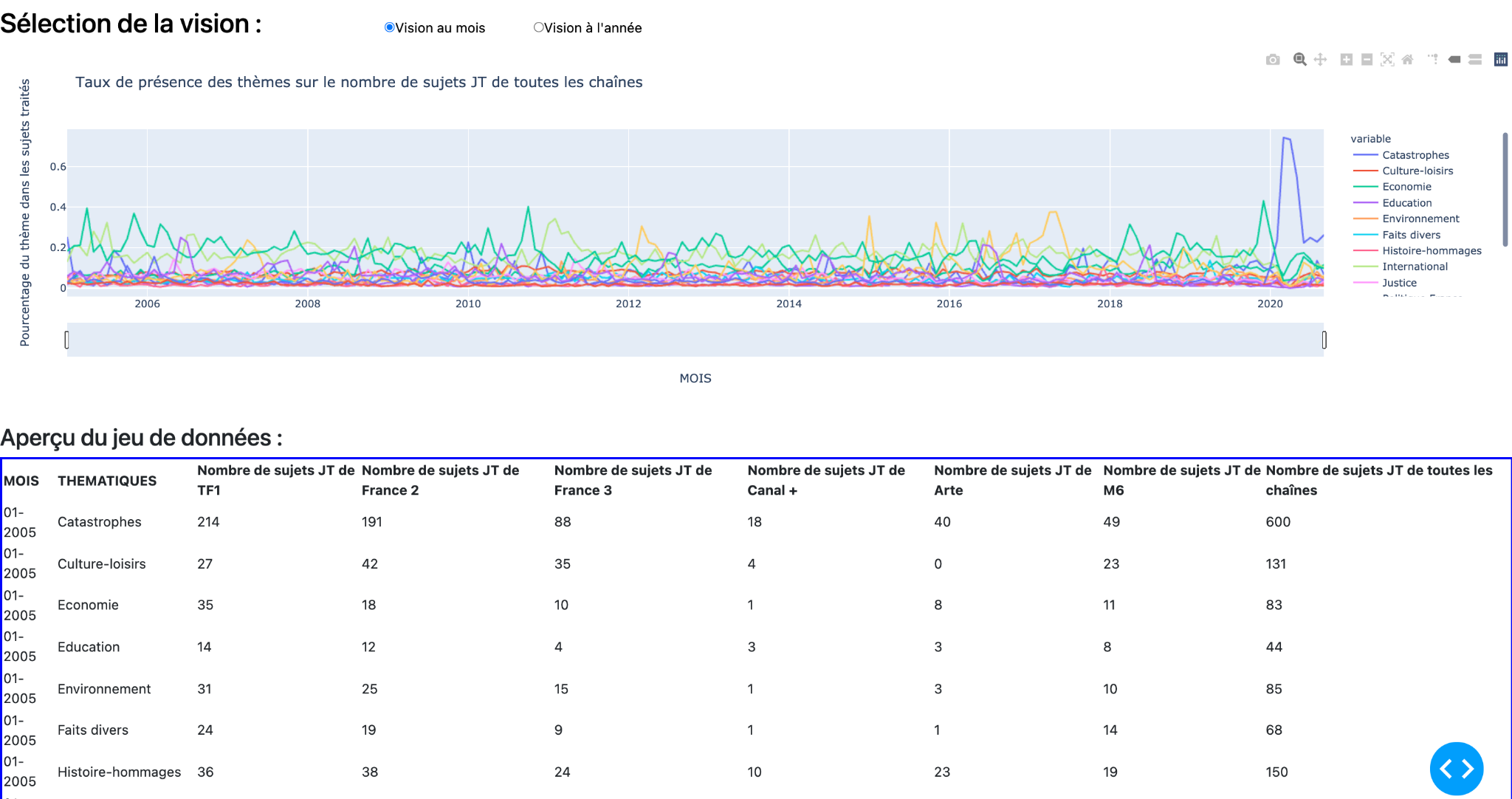


Comme on peut le voir, la part de chaque chaîne diffère suivant la thématique. Comme on a pu l’évoquer précédemment, on observe que la thématique internationale est principalement diffusée par Arte. A l’inverse, Arte diffuse très peu de sport. La part de sujet portant sur le sport pour cette chaîne représente 1,3%.

# Résultat final

En regroupant l’ensemble de ces graphiques, nous avons obtenu le Dashboard suivant :





Pour pouvoir voir l’intégralité de la page, il faut scroller car nous avons décidé d’afficher les graphiques de manière lisible au lieu de tout centraliser sur un même affichage.

On notera que le diagramme circulaire et le graphique de série temporelle peuvent être filtrés par la chaîne TV. Cela se fait grâce à une liste déroulante sur laquelle l’utilisateur peut choisir la chaîne TV qu’il souhaite analyser. Par défaut, il nous a semblé pertinent d’afficher les valeurs sur l’ensemble des chaînes.

Le graphique correspondant à la série temporelle peut aussi être filtré suivant les années ou les mois. Ceci se fait suivant des boutons radios comme montré précédemment. Par défaut, cet élément de visualisation de données est filtré par mois.

Et enfin, chaque graphique est dynamique. En passant la souris dessus, les valeurs sont affichées avec les libellés. On peut aussi se déplacer suivant l’espace temps de nos données sur le graphique de série temporelle. Quant au diagramme circulaire et diagramme en barre, on peut retirer des modalités de l'affichage en cliquant simplement sur la valeur souhaitée.

# Accès au projet

Comme dit précédemment, ce projet a été réalisé sous Python et plus précisément avec l’utilisation de Dash. Ainsi, pour pouvoir visualiser le Dashboard, il faut dans un premier temps lancer le projet python à partir du fichier app.py . Ce dernier appellera successivement les fichiers nécessaires (l’architecture de la page sera vue dans la partie suivante). Pour retourner l’affichage de la page, nous pouvons nous connecter à l’URL fourni en lançant le code. Cette dernière est la suivante : http://127.0.0.1:8050/

Nous noterons aussi que nous avons développé ce Dashboard sous mac et sous windows simultanément. Nous avons ainsi pu lancer ce projet sous différents navigateurs (Firefox, Chrome, Safari, Edge). Nous nous sommes aperçus que sous Firefox, l’affichage ne se faisait pas correctement. Ainsi pour ouvrir le projet et obtenir un affichage correct, il vaut mieux utiliser Chrome, Safari ou Edge comme moteur de recherche.

# Méthode architecture du site

Afin de réaliser ce tableau de bord, nous avons dû organiser nos fichiers. Pour cela, nous avons séparé nos fichiers de la manière suivante :

* Dans un dossier datagraph, nous avons créé deux fichiers python functions\_graph.py et functions\_graph.py. Ces deux fichiers sont constitués des fonctions que l’on réutilise dans le fichier python app.py.
  + Dans le fichier functions\_data, nous avons importé notre jeu de données puis créé diverses fonctions afin de traiter les données pour ensuite les réutiliser dans les autres fichiers. Par exemple, on a créé la fonction suivante :

def data\_pie(Chaine\_TV):

# Préparation des données pour réaliser le graphique circulaire

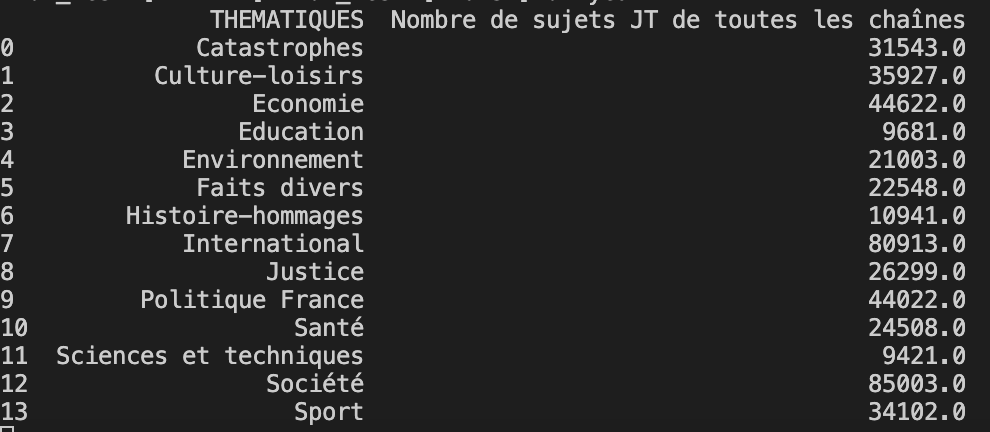
df\_test0= pd.to\_numeric(data[Chaine\_TV])

col = ["THEMATIQUES", Chaine\_TV]

df\_test0= data[col].groupby("THEMATIQUES").sum().reset\_index()

return df\_test0

Cette fonction nous permet de récupérer le nombre de sujets par thématique filtré suivant une chaîne TV dont le nom est entré en paramètre. Cela nous permet donc d’obtenir le dataFrame suivant (ici filtré sur toutes les chaînes TV) :



Cette structure nous a été nécessaire pour afficher le diagramme circulaire.

* + Dans le fichier functions\_graph, nous avons créé diverses fonctions afin de concevoir les différents graphiques que nous souhaitions afficher. Par exemple, on a créé la fonction suivante pour afficher le graphique circulaire à partir des données traitées précédemment :

def affichage\_pie(Chaine\_TV):

df = data\_pie(Chaine\_TV)

# Réalisation du graphique circulaire

fig = px.pie(df, values=Chaine\_TV, names=df["THEMATIQUES"],title="Repartition du "+(Chaine\_TV[0].lower())+Chaine\_TV[1:]+"<br>toutes périodes confondues")

fig.update\_layout(title\_x=0.5)

return fig

Comme on peut le voir, nous récupérons les données traitées en appelant la fonction data\_pie() se situant dans le fichier functions\_data.py. Cette nouvelle fonction est ensuite appelée dans le fichier app.py.

* Enfin, nous avons conçu un fichier à la racine du projet nommé app.py. Ce dernier permet de regrouper tous les callback des fonctions en ayant besoin. C’est le cas par exemple pour le graphique circulaire présenté. Nous avons donc obtenu le code suivant :

@app.callback(

Output("pie-chart", "figure"),

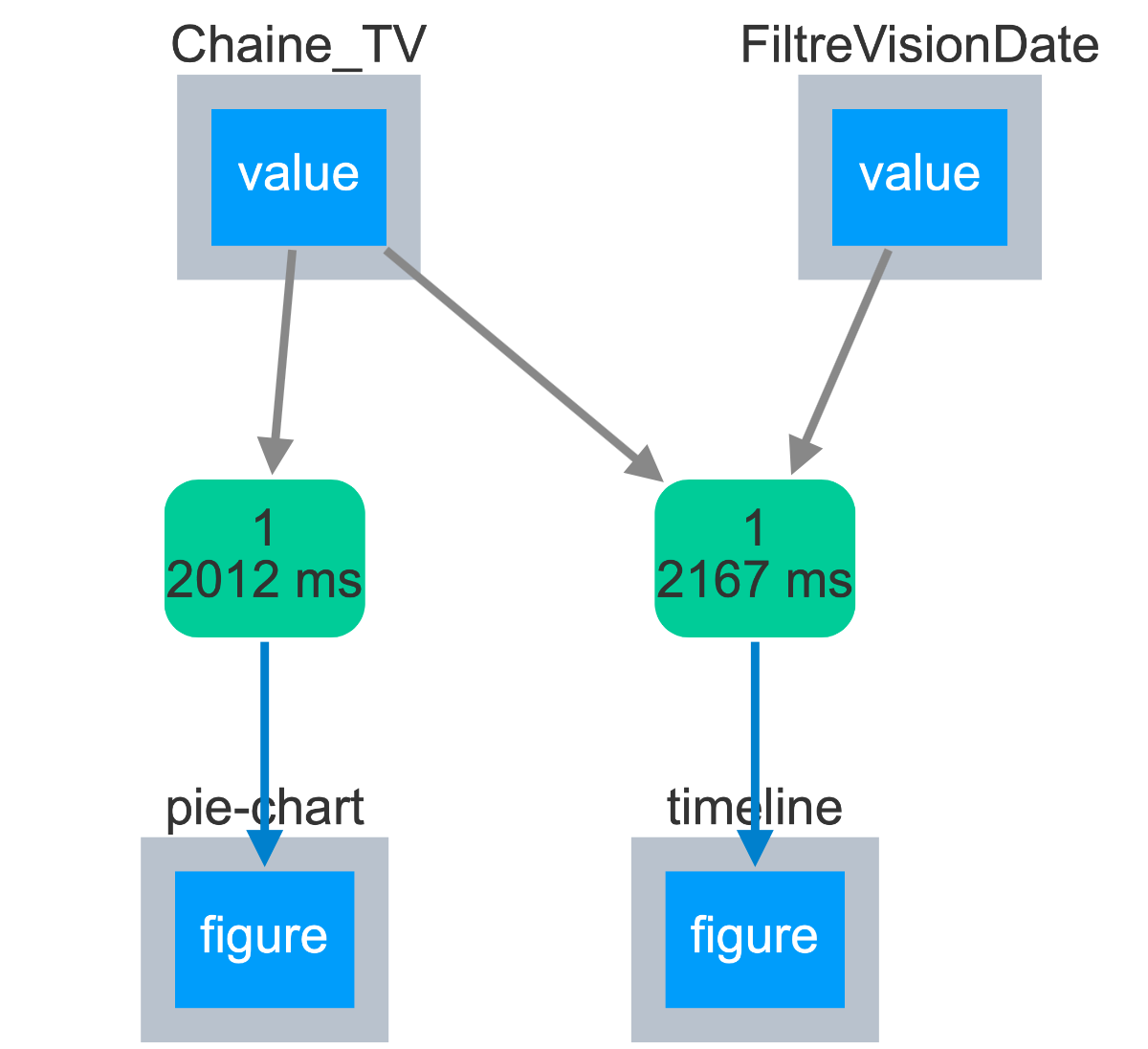
Input("Chaine\_TV", "value"))

def generate\_chart(Chaine\_TV):

return affichage\_pie(Chaine\_TV)

Ceci nous permet donc suivant la chaîne que l’on choisit de filtrer et d’actualiser les données. Néanmoins, l’affichage de ce graphique se fait grâce à son appel dans le Layout situé dans le fichier app.py. C’est cette partie de code qui nous permet d’afficher le graphique tel quel.

Ainsi, en revenant au fichier app.py qui permet d’appeler les fonctions et réaliser les callback (actualisation des données suivant les filtres), nous pouvons voir l’appel de ces derniers se fait de la manière suivante :



On peut donc observer que l’id Chaine TV, rapporté à la liste déroulante permettant de filtrer nos données suivant la chaîne s’applique sur le graphique circulaire et le graphique de série temporelle. Sur ce dernier graphique, le filtre nommé filtreVisionDate s’applique aussi. Ce dernier nous permet de lisser nos données sur les mois ou les années.

# Les difficultés rencontrées

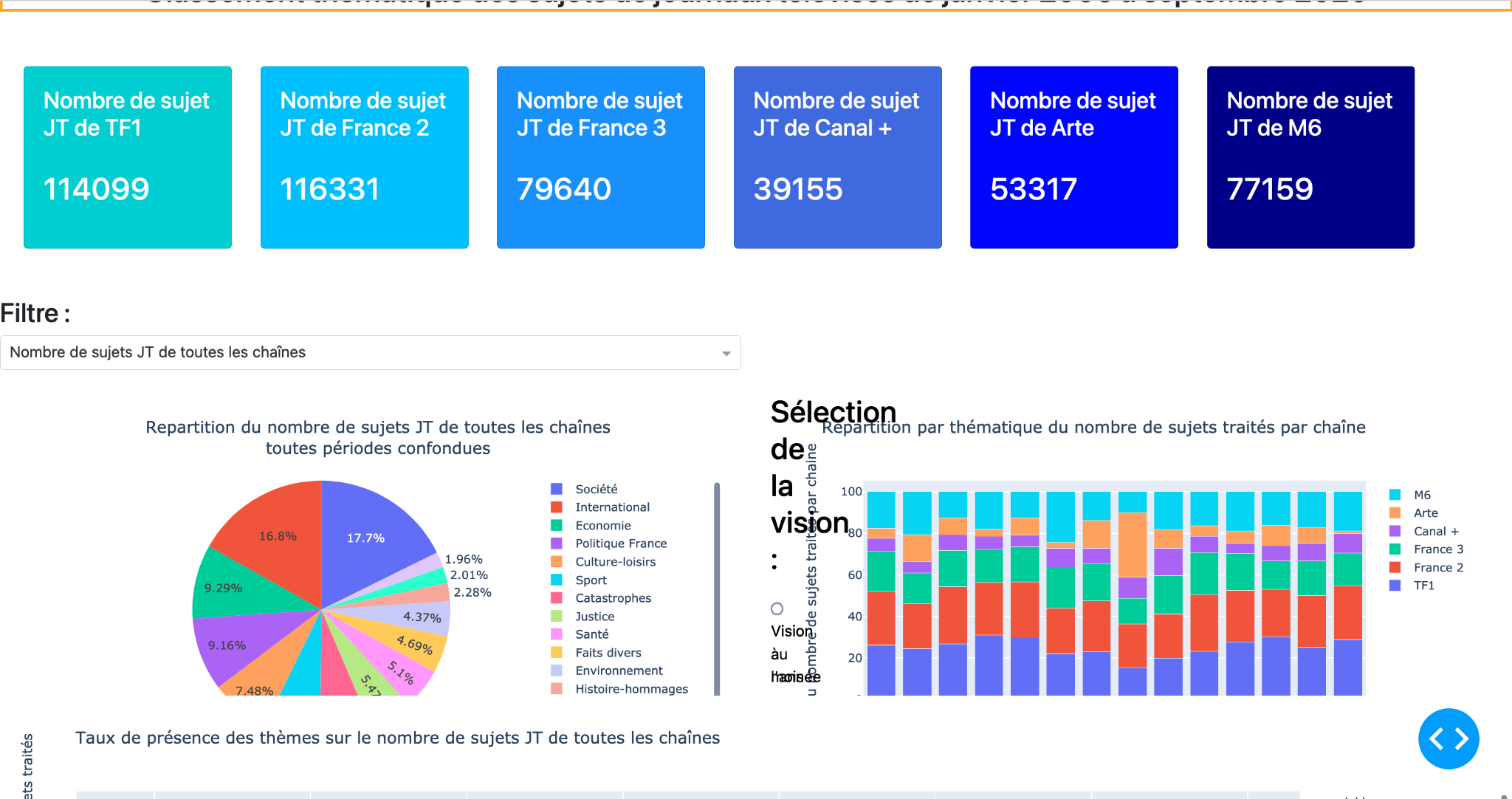
A travers ce projet, nous avons rencontré des difficultés de tous types.

Dans un premier temps, nous avons eu du mal à créer notre premier graphique. Ce dernier nous a demandé plus de temps que prévu. Nous avions un problème au niveau de la récupération des données car ces dernières n’étaient pas au bon format (numérique). Nous avons donc été obligés de convertir chaque colonne en valeur numérique.

Un deuxième problème est apparu au cours de la réalisation du projet, un message d’erreur s’affichait la première fois que l’on lançait le code. En rechargeant la page, ce message disparaissait et ne réapparaissait plus. Nous ne comprenions pas pourquoi ce message s’affichait uniquement de manière ponctuelle. En cherchant sur internet, nous n'avons pas trouvé d’explications nous permettant de résoudre ce souci.

De plus, le plus gros inconvénient que nous avons rencontré a été le temps. Nous avions 6h de projet banalisé. En parallèle de cela, nous étions en semaine de partiel avec d’autres projets à réaliser. Ainsi, dégager du temps pour ce projet a été nécessaire, mais nous ne pouvions pas faire plus. Nous avons fait au maximum pour afficher un Dashboard le plus propre, intéressant, stable et exploitable possible. Avoir une semaine de plus nous aurait permis de réaliser davantage de graphiques, et de plus « vivre » ce projet que de le « subir » dans le stress. Du fait de ce court délai, nous avons opté pour effectuer notre projet sous DASH car nous craignions de ne pas avoir le temps d’obtenir un rendu correct (et heureusement).

Actuellement, un problème persiste, nous n’avons pas réussi à ouvrir notre projet proprement sous le navigateur Firefox. Nous avons travaillé sous mac et windows simultanément, et plus précisément sous les navigateurs de recherche Safari, Google Chrome. Dans ces derniers, nous n’avons pas remarqué de problème d’affichage. Mais une fois le tableau de bord réalisé, nous l’avons ouvert sous Edge et Firefox. Et sous Firefox, nous avons observé que les graphiques s’affichaient mal, ainsi que les boutons pour filtrer par mois ou par année. Nous avions donc l’affichage suivant que nous n’avons pas réussi à corriger :



Peut-être qu’avec un peu plus de temps, nous aurions pu prendre un peu plus de recul sur notre travail, réussi à résoudre ces problèmes et nous aurions pu rendre un travail plus complet et stable.

# Ressenti de ce projet

Malgré les difficultés rencontrées et citées précédemment, nous avons quand même apprécié réaliser ce projet. C’est une application concrète des connaissances apprises lors de ce cours d'architecture logicielle.

Nous avons aussi apprécié le fait d’être libre dans le résultat attendu, l’outil à utiliser (Js ou Dash), le choix des éléments de visualisation de données (graphique circulaire, à barres…), le choix de la base de données, du sujet à traiter. Cela nous a permis de traiter un sujet autre que le jeu de données iris et d’avoir des données qui nous plaisaient.

De plus, le fait de travailler en groupe a été très constructif, car nous n’avions pas les mêmes bases en python et cela nous a tous permis de progresser à notre rythme. Dès que l’un de nous était bloqué, nous pouvions l’aider. Nous travaillons principalement ensemble sur un même graphique. Mais si nous n'arrivions pas à l’afficher au bout d’un certain temps, nous le laissions de côté et on le reprenait à un autre moment chacun de notre côté suivant nos disponibilités.

# Où retrouver notre projet ?

Notre projet se trouve à l’adresse :

<https://github.com/shirleyclay/ArchitectureLogicielle/>

Celui-ci est en public afin que vous puissiez, si vous le souhaitez, le récupérer à tout moment.

# Conclusion

Ainsi de manière générale, nous avons pu nous familiariser avec Dash et le principe de l’architecture logicielle. Il nous a fallu comprendre le fonctionnement de création de graphique sous Dash, de mise en forme, de filtre…

Nous avons pu tirer quelques analyses pertinentes par rapport à notre étude. Nous avons donc pu observer que récemment, entre mars et mai 2020, la santé a été le principal sujet diffusé au JT dans l’ensemble des chaînes. De plus, nous avons pu constater que Arte diffusait principalement des sujets internationaux par rapport aux autres chaînes. Quant à Canal+, nous avons pu apprendre que cette chaîne avait suspendue son JT durant un an pour ensuite le remettre l’année suivante avant de le retirer définitivement.

Malgré tout, ce Dashboard peut encore être amélioré. Nous pourrions tout d’abord corriger le bug remonté et permettre d’ouvrir ce rendu sous n’importe quel navigateur. Nous pourrions aussi rajouter d’autres graphiques et informations.

Néanmoins, nous sommes assez contents du résultat que nous présentons. Il a répondu à nos objectifs (tous les graphiques souhaités ont pu être réalisés). De plus, nous trouvons le rendu plutôt agréable à regarder (au niveau des couleurs, des choix de visualisation graphiques…). Mais surtout nous avons pu faire des analyses pertinentes sur les résultats retournés par ce Dashboard.

Nous espérons que le résultat sera à la hauteur de vos attentes. Merci beaucoup pour votre aide et votre accompagnement.

*Bonne continuation !!!!*