



DUT STID 2<sup>ème</sup> année  
Année universitaire 2020 / 2021  
IUT Paris Descartes

---

## Description du profil et des comportements d'audience des foyers panélistes qui participent à la mesure depuis au moins dix ans

---

*Davia Moujabber*



**Entreprise d'accueil :** Médiamétrie, Levallois-Perret (92)

**Maître de stage :** Quiterie Le Bourgeois

**Tuteur universitaire :** Marc Métivier



## Remerciements

Ce stage représente l'aboutissement de deux années d'études épanouissantes et confirme mon souhait de poursuivre mes études dans ce domaine. Je tiens à remercier plusieurs personnes pour ce qu'elles ont pu m'apporter.

Tout d'abord, je tiens à remercier Nicolas Goncalves, un data scientist de l'entreprise et mon ancien professeur de sondage. Il m'a présentée l'entreprise Médiamétrie et m'a partagée son expérience qui m'a permis de découvrir le monde de l'entreprise dans les meilleures conditions.

Quiterie Le Bourgeois, mon maître de stage au sein de Médiamétrie, qui m'a suivi tout au long du stage ainsi que toute la liberté qu'elle m'a accordée dans la réalisation de mes missions. Au-delà de la confiance qu'elle m'a accordée, je souhaite lui faire part de ma gratitude pour son accueil et ses précieux conseils.

Je souhaite remercier Caroline Coinon, un membre au sein de l'équipe qui a su me mettre tout de suite dans les meilleures dispositions pour ce stage et pour ses précieux conseils. Je remercie également Quentin Hell et Audrey Renot pour leurs avis et suggestions et pour leur aide sur SAS.

Je remercie par ailleurs, Aurélie Vanheuverzwyn et Fabienne Le Sager pour m'avoir accueillie au sein de la Direction Data Science.

Merci à Quentin, Caroline et Théotime pour les bons moments passés en leur compagnie.

Concernant l'IUT Paris Descartes, je tiens à remercier Marc Métivier pour m'avoir accompagnée durant toute la durée du stage. Plus globalement, je remercie toute l'équipe pédagogique et enseignante du DUT, pour la qualité de leurs enseignements et leur bienveillance.

Enfin merci à mes proches, et plus particulièrement, mes parents, pour leur soutien quotidien.

# Sommaire

1	Introduction.....	5
2	Contexte .....	6
2.1	Le monde de la mesure des médias .....	6
2.1.1	Histoire de l'entreprise.....	6
2.1.2	Mesure d'audience aujourd'hui .....	6
2.2	Entreprise .....	6
2.2.1	Médiamétrie, leader des études médias .....	6
2.2.2	Organisation de Médiamétrie .....	7
2.3	Une étude pour la base de données .....	8
2.3.1	Médiamat .....	8
3	Cartographie de l'architecture des données.....	10
3.1	Circuit des données .....	10
3.1.1	Création des données.....	10
3.1.2	Préparation et pré-traitement .....	11
3.1.2.1	Mesure semi- automatique .....	11
3.2	Environnement technique .....	12
3.2.1	Organisation du travail.....	12
3.2.2	Outils informatiques et statistiques .....	13
3.3	Base de données.....	13
3.4	Gestion des données à caractère personnel .....	13
4	Présentation de la mission .....	14
4.1	Contexte de la mission .....	14
4.1.1	Enjeux et objectifs .....	14
4.1.2	Calendrier et définitions mathématiques .....	14
4.2	Compétences sollicitées .....	15
4.3	Déroulement de la mission.....	16
4.3.1	Elaboration des profils des panélistes ayant 10 ans d'ancienneté ou plus .....	16
4.3.2	Evaluation de leur audience .....	19
5	Conclusion .....	21
6	Ressources.....	22
6.1	Tables des figures .....	22
6.2	Glossaire .....	22
7	Annexe.....	23

# 1 Introduction

Concernant la recherche de stage, j'ai décidé de trouver un stage qui puisse me laisser l'opportunité de mettre en pratique les enseignements théoriques et relationnelles acquises durant mon cursus et de découvrir le monde du travail. La recherche de stage a représenté pour moi, une sorte de première recherche d'emploi dans le domaine de la statistique et de l'informatique. Ce stage a été un bon moyen de découvrir de nouvelles technologies et d'approfondir les méthodes statistiques. L'offre de stage proposée par MÉDIAMÉTRIE répondait aux attentes que je voulais puisqu'elle m'offrait une grande liberté dans mon travail et la possibilité de mettre en pratique mes compétences sur les logiciels SAS et Excel dans un cas réel. De plus, ce stage à dominante statistique s'inscrivait dans mon projet professionnel qui est de construire une carrière dans la data.

Le secteur des sondages et des études de marché, dont Médiamétrie est acteur en grande partie, représente en France plus de 2 milliards d'euros (4<sup>ème</sup> marché mondial du secteur) et comporte environ 2 000 entreprises. Néanmoins, Médiamétrie n'a pas de concurrent direct car son principal marché d'étude, l'audience des médias, est un marché dont les acteurs se sont regroupés pour créer Médiamétrie. En effet, Médiamétrie mesure et analyse les audiences médias (TV, Radio, Internet, Cinéma, ...) en France. Elle publie des résultats récurrents comme les audiences TV des principales chaînes en France chaque jour à 9h et répond aussi à des demandes spécifiques de ses clients qui sont des chaînes de télévisions, des stations de radios et des publicitaires. L'univers des médias étant un secteur en constante évolution, Médiamétrie se doit d'être informé des dernières tendances de consommation.

Lors de ce stage, j'ai eu deux missions principales à réaliser. La première était de faire une description des foyers qui participent au panel Médiamat, la mesure d'audience TV de référence en France, depuis 10 ans ou plus. C'est-à-dire définir le profil de ces individus. La deuxième réalisation était une analyse de leur audience. Dans le but de mener à bien ces missions, il a fallu notamment répondre à ces questions :

- Qui sont ces foyers ?
- Participent-ils bien à la mesure ?
- Ont-ils des comportements d'audience différents ?

L'objectif principal de ces deux missions est de déterminer si la durée de panélisation entraîne un biais de participation. Pour cela, différentes données présentes sur le réseau de l'entreprise ont été mis à ma disposition et des codes SAS existants nécessaires à la mise en place des analyses. A l'issue de ces analyses, la rédaction d'une note méthodologique décrivant la problématique, les analyses effectuées et les conclusions a été demandée.

## 2 Contexte

### 2.1 Le monde de la mesure des médias

#### 2.1.1 Histoire de l'entreprise

Médiamétrie<sup>1</sup> est une entreprise créée en 1985. C'est une société anonyme spécialisée dans la mesure d'audience et de l'étude des usages des médias audiovisuels et numériques en France. Elle répond en ce sens aux nouveaux besoins du paysage audiovisuel français : l'apparition des radios privées de la bande FM, la privatisation d'Europe 1 et de la première chaîne, la création de Canal+, le lancement de la Cinq et de la Six. Face à cette situation nouvelle, les pouvoirs publics ont encouragé la naissance d'une société indépendante, Médiamétrie, pour assurer la mesure scientifique d'audience des principaux médias audiovisuels.

L'indépendance de Médiamétrie est garantie par la présence, dans toutes ses instances de décisions et dans son capital, de l'ensemble des professionnels (médias, annonceurs, agences) sans qu'aucun ne détienne une majorité pour décider seul. La règle aussi est le dialogue et le débat en vue de parvenir à un consensus dans l'intérêt général de la profession. La rencontre, au sein d'une même instance, de représentants des radios, télévisions, annonceurs, agences de publicité et centrales, contribue au succès de l'entreprise et à la mise au point d'instruments d'étude indispensables et performants : ils permettent de mesurer l'émergence de nouvelles stations et - ou antennes comme l'évolution des leaders du marché.

#### 2.1.2 Mesure d'audience aujourd'hui

Médiamétrie est aujourd'hui la référence en matière de mesure d'audience. L'entreprise développe sa gamme de services et élargit son champ d'action en travaillant sur les nouveaux médias, la téléphonie, les nouvelles pratiques multimédias, le Cross Média... Elle propose des produits originaux destinés à des utilisateurs spécifiques et lance sur le marché international des offres rendues indispensables par l'évolution des comportements du public en matière d'audiovisuel. Elle est notamment connue pour sa marque Audimat qui se nomme aujourd'hui Médiamat.

## 2.2 Entreprise

### 2.2.1 Médiamétrie, leader des études médias

Médiamétrie est un Leader des études médias, elle observe, mesure et analyse les comportements du public et les tendances du marché. Leur fiabilité, leur capacité d'innovation et l'engagement de leurs équipes contribuent à fédérer leurs clients et à faire d'eux **la référence**. Elle s'est développée en France sur plusieurs supports (TV, Radio, Internet, Cinéma, téléphonie mobile, tablette et le cross médias) et à l'international avec 3 types d'activités :

---

<sup>1</sup> Médiamétrie : <https://en.wikipedia.org/wiki/Médiamétrie>

- L'exportation de technologies (audimètres) en Inde et Norvège et la participation à GAMMA<sup>2</sup> (Global Alliance for the Measurement of Media Audiences) ;
- La mesure d'audience en Afrique, au Maroc avec sa filiale Marocmétrie et en Côte d'Ivoire, au Sénégal, au Cameroun et à Madagascar avec sa filiale Omedia ;
- La veille de programmes à l'international et les données d'audience de différents pays avec son département Eurodata TV Worldwide.

Médiamétrie a employé, en novembre 2018, 679 collaborateurs et enregistre un chiffre d'affaires de 102,7 millions d'euros. Son siège social est situé à Levallois-Perret (92) et elle possède deux centres d'appels à Amiens (80) et Petit-Quevilly (76) contribuant à la réalisation de 1,5 millions d'interviews par an.

### 2.2.2 Organisation de Médiamétrie

Médiamétrie est organisé autour de 3 *business units* (Annexe I) :

- **La Direction des Mesures d'Audience**, qui produit et commercialise les études de référence de Médiamétrie avec pour mission de garantir la qualité et l'évolution de la mesure de chaque média
- **La Direction Performance et Cross Media**, conçoit, produit et commercialise les études d'environnement
- **Eurodata TV Worldwide®**, le service mondial de référence pour le suivi des audiences des programmes de télévision dans plus de 80 pays.

Des Directions Support et Directions Corporate viennent compléter l'organisation de Médiamétrie. Mon stage s'est déroulé au sein de la Direction Data Science (DDS), une Direction Support dirigée par Aurélie Vanheuverzwyn, qui est composée d'environ 30 personnes (Annexe II). La DDS est divisée en *pôles* qui se concentrent sur un média ou une thématique (pôle TV et Internet, pôle Radio, Outre-mer et Cinéma ...).

---

<sup>2</sup> GAMMA : une alliance internationale entre des sociétés de mesure d'audience des médias dont Médiamétrie (France), BARC (Inde), Numeris (Canada) et Video Research (Japon).

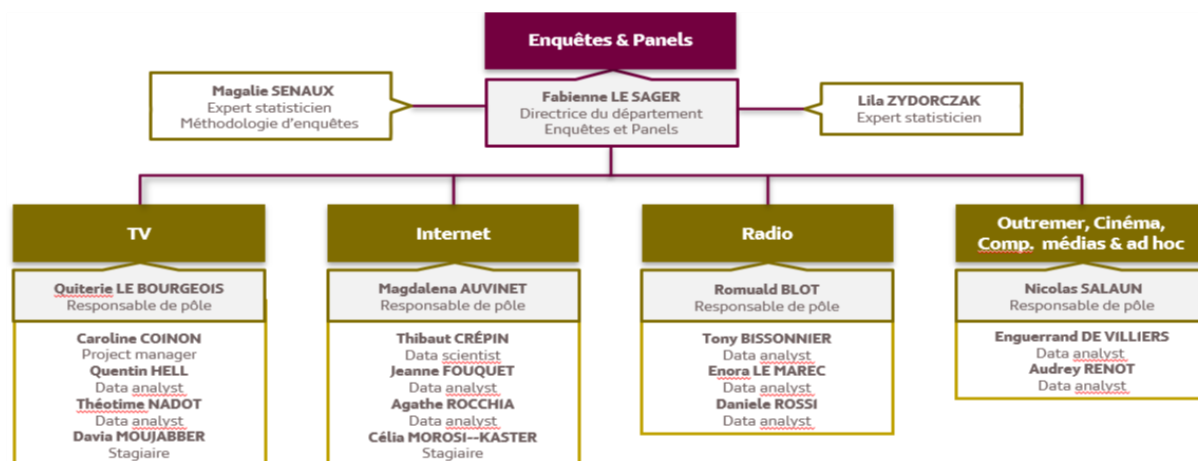


Figure 1-Organisation de la DDS (Direction Data Science)

Ainsi, j'ai intégré le pôle TV dirigé par Quiterie le Bourgeois. L'équipe TV est composée de trois membres :

- Caroline Coinon, Project manager
- Théotime Nadot, Data analyst
- Quentin Hell, Data analyst

## 2.3 Une étude pour la base de données

Avant de présenter le circuit des données, de leur création à leur analyse, il faut présenter le contexte dans lequel elles sont créées. Ainsi, à Médiamétrie, les données sont issues d'études. Pour remplir mes missions, j'ai utilisé des données issues du Médiamat.

### 2.3.1 Médiamat

Le Médiamat est un dispositif s'appuyant sur un panel, des conventions et des outils informatiques qui permet de restituer des résultats d'audience vendus aux acteurs (chaînes, agences médias) du marché. Les résultats sont restitués quotidiennement via le logiciel Restit'TV (outil développé en interne). Cette étude est la référence de la mesure d'audience des chaînes de la télévision en France. Elle se base sur un panel de 5 000 foyers de France métropolitaine équipés d'au moins une télévision active<sup>3</sup>, soit environ 11 700 individus âgés de 4 ans et plus. Ce panel est représentatif des foyers équipés d'une TV en France métropolitaine sur :

- Des critères sociodémographiques de l'Insee ;
- Des critères d'équipements et de modes de réception de l'étude Home Devices de Médiamétrie.

Le recueil d'audience est automatique, il fonctionne grâce à la technologie du watermarking<sup>4</sup> qui permet à un boîtier connecté à la télévision (audimètre) de reconnaître la chaîne (sous forme d'identifiant), l'heure et la date de diffusion du programme. À chaque individu du foyer est attribué une touche sur une télécommande reliée à l'audimètre pour

<sup>3</sup> Télévision active : utilisée au moins une fois par mois.

<sup>4</sup> Watermarking : insertion d'une marque audio inaudible à l'oreille humaine dans les programmes des chaînes de télévision.



déclarer leur présence dans la pièce où la télévision est allumée. De plus un bouton invité permet de recueillir l'audience des amis et de la famille. On obtient ainsi les audiences des chaînes de télévision watermarquées sur la télévision du domicile et en Live ou replay (Catch-up) à la seconde près pour chaque individu du foyer.



*Figure 2-Processus de Watermarking*

L'objectif de cette technologie est de détecter la marque que l'on a insérée dans le son. Chaque régie dispose d'un inserteur lui permettant d'insérer ce tatouage dans le son. La chaîne diffuse ses programmes marqués. L'audimètre collecte des informations, notamment ce que les panélistes ont vu à la télévision.

## 3 Cartographie de l'architecture des données

### 3.1 Circuit des données

Les données suivent un parcours différent selon l'étude dont elles font parties. Ici nous allons nous focaliser sur les circuits liés au **Médiamat**. Il y a trois vagues de recrutements :

- Février – Mars
- Avril – Juin
- Septembre – Décembre

Une fois le recrutement effectué, une photographie du panel est faite chaque lundi. Ces données sont complétées par d'autres informations notamment les données collectées par l'audimètre, entre autres, les données. Médiamétrie réalise tous les ans une enquête de recensement auprès des foyers du panel afin de mettre à jour les informations sur la composition et l'équipement des foyers panélisés. L'enquête a lieu de février à mai. Elle comporte :

- Un volet sur le foyer, à remplir par la personne de référence du ménage ou par son conjoint,
- Et un volet individuel, à remplir par chacun des membres du foyer (ainsi que par les éventuels nouveaux membres). Il y est également intégré des questions servant à définir des cibles de consommation ou comportementales.

L'Enquête peut être réalisée par téléphone ou, à défaut, par courrier lorsque les foyers ou certains individus ne peuvent pas être joints autrement. L'enquête est à réaliser sur la totalité des foyers du panel, allégée pour les foyers récemment intégrés.

Le questionnaire comporte une partie foyer et une partie individuelle, à renseigner par chacun des membres du foyer :

- **Partie foyer :**
  - Descriptif détaillé des installations audiovisuelles (TV, câble, lecteurs DVD, consoles de jeu, modes de réception et abonnements TV)
  - Descriptif sociodémographique du foyer et de son équipement général, dont internet, et téléphone.
- **Partie individu :**
  - Situation personnelle et profession,
  - Utilisation des équipements audiovisuels,
  - Fréquentation du cinéma,
  - Téléphonie mobile,
  - Questions diverses (identiques à celles de l'enquête « recueil » lors du recrutement)

#### 3.1.1 Création des données

- Médiamat

Les données sont issues d'une mesure automatique grâce aux audimètres reliés aux postes de télévision des panélistes. Chaque événement d'usage (ON, OFF, Zapping et chaîne ...) et

chaque déclaration de présence individuelle (via une télécommande, les individus doivent appuyer sur la touche qui leur est attribuée) est enregistrée par l'audimètre. Ces données sont traitées et transformées en tickets d'audience. Un ticket est affecté à une journée et chaque ticket d'audience est composé des identifiants Foyer/Individu/Poste, d'une plage horaire (heures de début/fin), de l'identifiant physique de la chaîne et de l'offset<sup>5</sup>.

Par exemple : Quiterie a regardé « Polars du Terroir » sur Arte le 27/10/2017 sur le téléviseur de la cuisine. Seul l'identifiant de la chaîne est différent entre les deux tickets générés en sortie du prétraitement, l'offset est à zéro car on est en live.

### ➤ Live

Tableau 1-Ticket d'audience en mode Live

	ID foy	ID ind	ID poste	H deb	H fin	Chaîne	Offset	Date
	Le Bourgeois	Quiterie	Chambre	12 : 20 : 00	12 : 50 : 00	Arte	00 : 00 : 00	27/10/2017
JDV	1	1	1	12 : 20 : 00	12 : 50 : 00	18	00 : 00 : 00	27/10/2017
JDD	1	1	1	12 : 20 : 00	12 : 50 : 00	307	00 : 00 : 00	27/10/2017

Pour un individu, il y a autant de tickets d'audience qu'il y a de séquences de visionnage et de chaînes visionnées au cours de la journée.

Par exemple : Lorie a regardé « Danse avec les Stars » (TF1) pendant 30 minutes, 2h après sa diffusion. L'identifiant est toujours différent entre JDV et JDD et l'heure de début est réaffectée à l'heure de diffusion en JDD. L'offset est mis à zéro en JDV mais est conservé en JDD. La date reste la même car c'est du différé du jour même

### ➤ En différé du jour même (programme enregistré)

Tableau 2-Ticket d'audience en mode différé du jour même

	ID foy	ID ind	ID poste	H deb	H fin	Chaîne	Offset	Date
	Dudoignon	Lorie	Salon	23 : 00 : 00	23 : 30 : 00	TF1	02 : 00 : 00	21/10/2017
JDV	2	2	3	23 : 00 : 00	23 : 30 : 00	1	00 : 00 : 00	21/10/2017
JDD	2	2	3	21 : 00 : 00	21 : 30 : 00	301	02 : 00 : 00	21/10/2017

## 3.1.2 Préparation et pré-traitement

### 3.1.2.1 Mesure semi- automatique

Toutes les données ne sont pas utilisées pour les analyses. En effet, une première phase de validation intervient chaque jour pour garantir la validité des résultats et constitue une première étape de filtrage des foyers. Par exemple, des résultats sont considérés invalides lorsque la télévision est restée allumée toute la journée sur la même chaîne. A noter qu'une « journée TV » est de 3h à 27h (3h le lendemain matin). Un foyer est considéré « répondant » si l'audience d'une chaîne ne dépasse pas 12h en continu et si les audiences sont cohérentes.

<sup>5</sup> Offset : heure de vision – heure de diffusion

Chaque jour, entre 5h et 7h, deux chaînes de production fonctionnent en parallèle, chaque lundi une troisième vient compléter le processus :

- Référentiel TV : donnée quotidienne fournie par Kantar (PIGE) ;
- Tickets d'audience : tous les tickets d'audiences de la veille de tous les audimètres sont traités ;
- Panel de référence : photographie hebdomadaire du panel avec l'ensemble des caractéristiques socio-démographiques.

Après l'intégration de la base de données vient l'étape de redressement (les données brutes sont conservées). Elle s'effectue à partir de 53 critères au niveau du foyer et 18 critères au niveau de l'individu, par exemple : région, âge, sexe. Chaque jour à 9h, les résultats redressés sont envoyés aux clients.

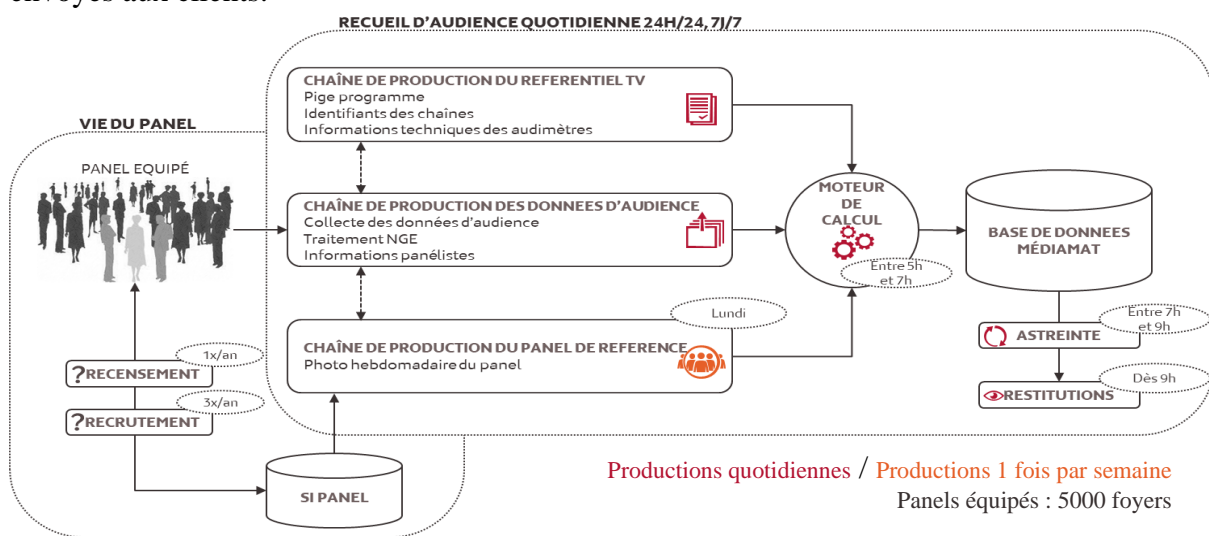


Figure 3-Processus de restitution d'audience quotidienne

## 3.2 Environnement technique

### 3.2.1 Organisation du travail

Tout d'abord, j'ai effectué mon stage en télétravail et en présentiel à cause de la crise sanitaire liée à l'épidémie de la Covid-19. L'accès au réseau en télétravail se faisait via une application **Token**, qui me permettait de connecter l'ordinateur portable que j'avais à ma disposition au réseau de l'entreprise et d'avoir accès aux données et aux informations nécessaires afin d'effectuer mes missions. La première tâche a été d'identifier les profils des panélistes selon leurs critères socio-démographiques puis d'équipement, abonnement et type de réception. Chaque semaine un point été mis en place pour pouvoir étudier l'avancement de mes analyses.

### 3.2.2 Outils informatiques et statistiques

J'ai tout d'abord travaillé sur Excel, une application du pack office afin de découvrir les bases de données. Pour vérifier la cohérence de mes analyses je disposais de **Restit'TV**. C'est un outil, interne à Médiamétrie, qui me permettait de récupérer sous forme de fichier **Excel** les données selon des cibles socio-démographiques et sous plusieurs formes (indicateurs d'audience, audience sur une période, ...). Une fois toutes les données à ma disposition récupérées, j'effectuais un requêtage des données sous SAS.

Durant ces deux mois, j'ai pu me perfectionner dans la pratique de logiciels que je connaissais déjà mais aussi découvrir d'autres façons d'utiliser des logiciels que je connaissais. SAS (Statistical Analysis System) est un langage de programmation édité par SAS Institute depuis 1976. Excel (logiciel issu du pack office) est un tableur permettant d'effectuer toutes sortes de modifications de données (calculs, comparaisons, analyses...). C'est un logiciel très connu, facile d'utilisation et qui permet par exemple de mettre en forme des résultats obtenus sous SAS.

### 3.3 Base de données

Les données auxquelles j'ai accédé durant mon stage proviennent du Médiamat. J'ai utilisé plusieurs bases de données sur SAS. Pour les comprendre on a mis à ma disposition des tables Excel où je pouvais retrouver les variables qui me servais de critère afin d'évaluer le profil des panélistes et leur qualité de participation. J'ai réalisé mon propre dictionnaire contenant uniquement les variables de critères et donc celles que j'allais être amenée à utiliser par la suite (*Annexe II*).

### 3.4 Gestion des données à caractère personnel

Depuis la mise en place du RGPD<sup>6</sup>, les données à caractère personnel font l'objet d'un contrôle strict. La Direction CATI est à la charge de leur gestion, elle gère les autorisations et accès aux collaborateurs de Médiamétrie.

---

<sup>6</sup> RGPD : Règlement général sur la protection des données

## 4 Présentation de la mission

### 4.1 Contexte de la mission

#### 4.1.1 Enjeux et objectifs

Le panel Médiamat est renouvelé tout au long de l'année mais certains foyers sont présents dans le panel depuis 10 ans ou plus. Depuis plusieurs années, Médiamétrie observe un nombre important de panélistes de 10 ans ou plus d'ancienneté.

La première partie de ma mission a consisté à faire une description des foyers qui participent à la mesure depuis des années c'est-à-dire d'établir un profil type à l'aide de critères d'abord socio-démographiques puis des critères d'équipement, d'abonnement ou encore de type de réception.

La seconde partie de ma mission était de faire une analyse de leur audience et de voir s'il participe correctement à la mesure.

Le but est de déterminer si la durée de panélisation entraîne un biais de participation.

J'ai dû livrer à la fin de mon stage une note méthodologique récapitulant les différentes analyses menées et une présentation des résultats obtenus. Je n'avais pas de cahier des charges à suivre, en dehors de l'utilisation de SAS et de Excel.

#### 4.1.2 Calendrier et définitions mathématiques

Le stage a débuté le 06/07 et se termine le 31/08. Vous retrouverez en (*Annexe III*) le calendrier prévisionnel détaillé.

Avant de détailler le déroulement du stage il est nécessaire de donner les définitions mathématiques des indicateurs d'audience utilisés durant les analyses :

DEI = Durée d'Écoute par Individu

La durée d'écoute par individu est la durée d'écoute moyenne par individu.

$$DEI(c, T) = \frac{\sum_{i \text{ membres}} \pi(i) D(i, c, T) + \sum_{i \text{ invités}} \pi(i) D(i, c, T)}{\sum_{i \text{ membres}} \pi(i)}$$

PDA = Part D'Audience

La part d'audience ou la part de marché permet à une chaîne de se positionner par rapport aux autres.

$$PDA\%(c, T) = \frac{DEI(c, T)}{DEI(TTV, T)}$$

La PDA s'interprète comme :

- Un individu consacre PDA% de son temps TV à la **c**

- PDA% des individus regardant la TV, le font sur la chaîne **c**

TC = Taux Cumulé d'audience

Le taux cumulé est la proportion de téléspectateurs.

$$TC\%(c, T) = \frac{\sum_{i \text{ membres}} \pi(i) 1(i, c, T) + \sum_{i \text{ invités}} \pi(i) 1(i, c, T)}{\sum_{i \text{ membres}} \pi(i)}$$

Avec :

- $i$  l'individu,
- $c$  la chaîne regardée,
- $T$  la tranche horaire (15 minutes, 1 jour, ...),
- TTV la liste des chaînes du référentiel (chaînes de la TNT, chaînes watermarquées, ...),
- $\pi(i)$  Correspond au poids attribué à l'individu  $i$ ,
- $D(i, c, T)$  Correspond à la durée passée par l'individu  $i$  sur la chaîne  $c$  durant la tranche horaire  $T$ ,
- $1(i, c, T)$  Vaut 1 si l'individu  $i$  a eu au moins 1 contact (10 secondes consécutives) avec la chaîne  $c$  durant la tranche horaire  $T$ .

TM%= Taux moyen pourcentage

$$TM\%(c, T) = \frac{DEI(c, T)}{D(c, T)}$$

Avec :

- $D(c, T)$  La durée de diffusion de l'événement

Le taux moyen d'audience s'interprète de deux façons :

- En moyenne, un individu a regardé la chaîne **c** sur TM% de la durée de la tranche horaire **T**.
- En moyenne sur la tranche horaire, TM% des individus ont regardé la chaîne **c**.

A noter que la formule de calcul exacte varie selon l'étude, ici ce sont les formules de calcul de l'étude **Médiamat**.

## 4.2 Compétences sollicitées

Les compétences sollicitées étaient techniques et relationnelles, j'ai acquis des nouvelles compétences durant ce stage. Tout d'abord, j'ai pu approfondir mes connaissances en SAS pour le traitement des données. Je suis devenue ainsi plus rigoureuse et efficace sur certains éléments :

- Organisation des répertoires de travail et des programmes ;
- Importation et Exportation des données ;
- Langage Macro et macro-variables ;
- Etape DATA et manipulation de variables ;
- Calculs d'indicateurs et variables indicatrices ;
- Utilisation des procédures descriptives ;
- Validation des résultats et de leur cohérence ;

- Synthétiser et sélectionner les informations à mettre en valeur.

Ensuite, j'ai mis en application les méthodes de reporting apprises sur Excel dans la création de graphique. J'ai appris à utiliser des logiciels internes de Médiamétrie Restit'TV (c'est une interface de restitution des résultats produits).

Sur le plan relationnel, la première compétence sollicitée était la capacité à communiquer par le dialogue et les diverses formes d'échanges en s'habituant rapidement au vocabulaire technique du monde professionnel. Deuxièmement, la manière de rendre compte des résultats sous forme d'une note méthodologique ou lors d'une présentation en respectant les codes propres à l'entreprise. Ces compétences d'échanges et de mise en avant de nos résultats sont le fruit de nos travaux d'équipe et des diverses présentations effectuées au cours du DUT.

### 4.3 Déroulement de la mission

Le stage s'est déroulé en deux parties : une partie où je devais établir les profils des panélistes de 10 ans ou plus d'ancienneté et une deuxième partie sur l'évaluation de leur audience.

#### 4.3.1 Elaboration des profils des panélistes ayant 10 ans d'ancienneté ou plus

##### Etape préliminaire

Notre base de données est constituée de 5001 foyers. Ces foyers vont être analysés à l'aide de différents critères sur la date du 18 novembre 2019. L'objectif de cette 1<sup>ère</sup> phase d'analyse est de déterminer et de définir les profils types en fonction de l'ancienneté dans le panel. Pour cela, nous avons effectué quatre classes :

- Jusqu'à 1 an d'ancienneté
- 2 à 5 ans d'ancienneté
- 6 à 9 ans d'ancienneté
- 10 ans ou plus d'ancienneté

En effet, le panel est composé en majorité de panélistes ayant entre deux à cinq ans d'ancienneté. Les panélistes de dix ans ou plus d'ancienneté représentent **14% du panel**, ce qui est majeur puisque d'après le CESP<sup>9</sup>, Médiamétrie devrait limiter la durée de panélisation à 10 ans maximum afin de réduire le nombre de panélistes de 10 ans d'ancienneté ou plus.

---

<sup>9</sup> Cesp : le Centre d'Etude des Supports de Publicité qui contrôle la qualité du Médiamat,



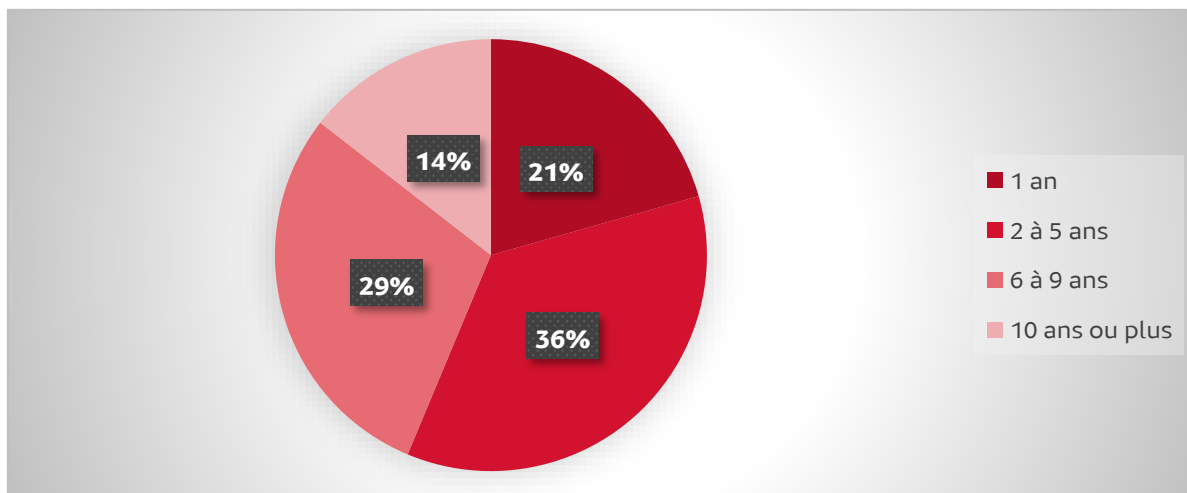


Figure 4-Répartition des panélistes dans le foyer

## Résultat

Tout d'abord, l'analyse se divise en deux étapes : Les critères socio-démographiques puis les critères d'équipement, abonnement et type de réception.

### Phase 1.1- Critère socio-démographique

Le premier constat est que les panélistes de 10 ans d'ancienneté ou plus sont plus des inactifs de plus de 65 ans que des agriculteurs de moins de 35 ans. De plus, ils vivent plus en Ile-de-France dans l'agglomération rurale que dans le Nord dans l'agglomération parisienne. Ils sont le plus souvent composé de deux personnes dans leur foyer que 5 personnes ou plus et pour la plupart ne dispose pas d'enfant de moins de 15 ans (*Annexe IV*).

Néanmoins, on peut émettre l'hypothèse que ce sont le plus souvent des retraités.

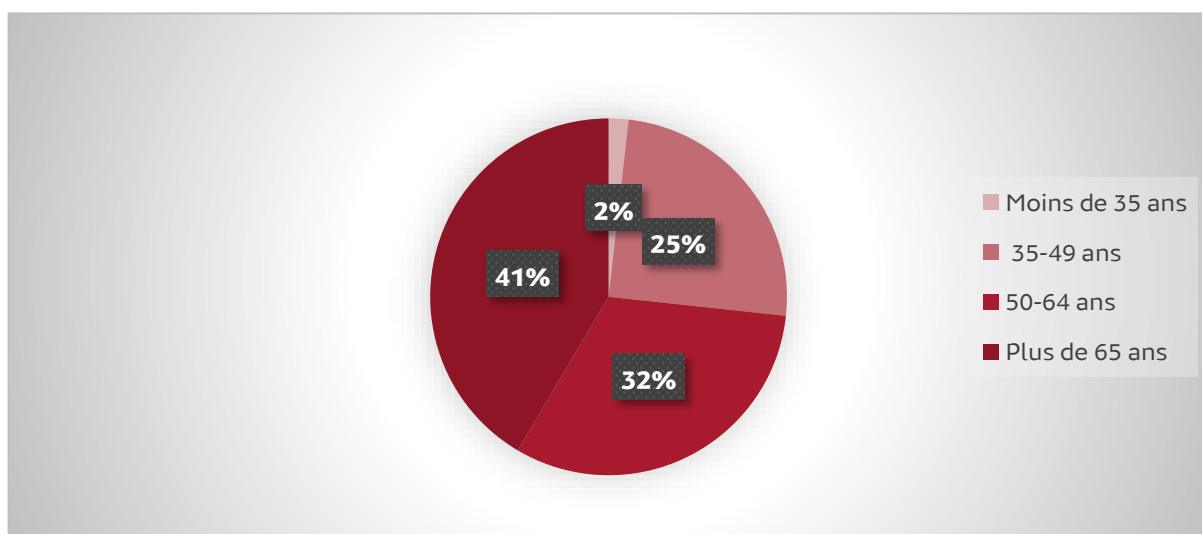


Figure 5-Age du chef du foyer qui ont dix ans ou plus d'ancienneté

## Phase 1.2- Critères d'équipement, abonnement et type de réception.

Cette deuxième étape est faite pour savoir de quels équipements disposent-ils et s'ils possèdent un abonnement ou non et quel est leur type de réception.

Tous les gens ont au moins une TV active dans le panel. Ceux qui n'en ont qu'une sont les monos équipés. Les autres sont multi-équipés TV. Leur type de réception est la réception **Hertzienne**, et sont surreprésentés sur ce critère par rapport aux autres niveaux d'ancienneté. Ils sont également ceux qui ont moins de réception ADSL/FIBRE ou abonnement aux câbles par rapport au panel. Seulement 13% des panélistes ont déclaré avoir un abonnement Canalsat et **14% des panélistes ayant 10 ans ou plus d'ancienneté ont un abonnement Canalsat**. Ils sont également les moins équipés d'ordinateur que le reste du panel ainsi que ceux qui **souscrivent le moins à la VOD.** (*Annexe V*)

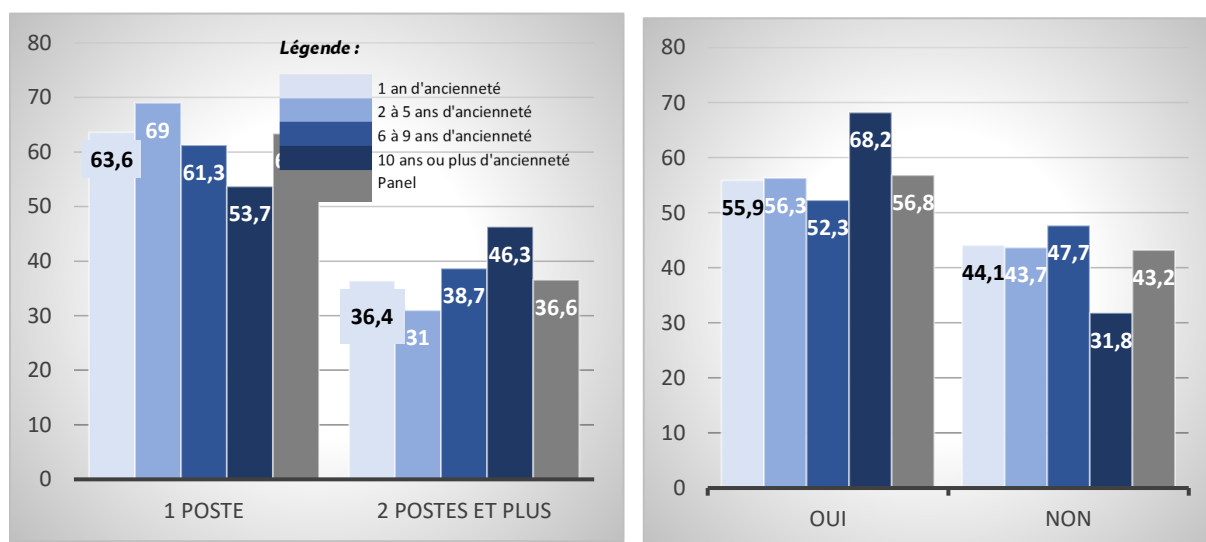
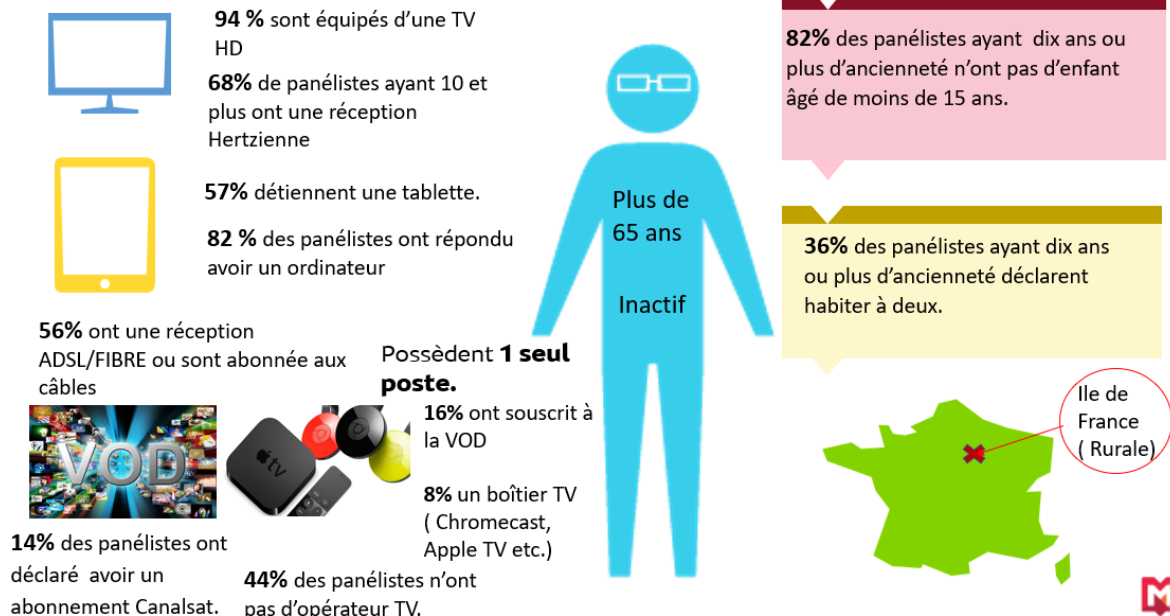


Figure 6-Nombre de TV actives et réception Hertzienne par niveau d'ancienneté (en %)

Chez les panélistes de 10 ans d'ancienneté ou plus **8%** ont un boîtier TV type Chromecast, Google TV, Apple TV, Fibre TV Stick ou Cube S. Ce qui représente peu.

## Phase 1.3 -Profil type

### Profil type d'un panéliste ayant 10 ans ou plus d'ancienneté



## 4.3.2 Evaluation de leur audience

### Etape préliminaire

Nous allons à présent étudier les audiences des panélistes. Pour débiter les analyses, nous allons nous focaliser sur la tranche d'horaire 3H-27H<sup>10</sup>. La période utilisée est du 18 au 24 novembre 2019. De plus, nous avons établi des tranches horaires à analyser qui sont :

- Matinale : 6h-9h
- Midi: 12h-14h30
- Access: 18h-21h
- Prime: 21h-23h30

### Résultat

Afin d'analyser les comportements d'audience des panélistes et voir s'il y a une différence entre les plus anciens et ceux recrutés récemment, on va se servir des indicateurs tels que la DEI (durée d'écoute par individu), la PDA (la part d'audience), le TC (Taux Cumulé d'audience) et le TM (Taux moyen pourcentage) .

<sup>10</sup> 3H-27H : on compte les heures entre 3H du jour j et 3H du jour suivant

### Phase 1.1- Tranche 3H-27H

Après avoir étudié la tranche horaire 3H-27H, nous avons vu que **les plus anciens panélistes ne sont pas ceux qui regardent le plus la télévision**. Le pic de consommation des panélistes de 10 ans d'ancienneté ou plus est le dimanche. Tandis que le mardi est le jour où les panélistes de 10 ans d'ancienneté ou plus regardent le moins la télévision.

En moyenne, les panélistes de 10 ans ou plus d'ancienneté écoutent plus la télévision le dimanche que le lundi.

Tableau 3-Durée d'écoute par individu de 10 ans ou plus d'ancienneté en mode Live

Période	18/11/19	19/11/19	20/11/19	21/11/19	22/11/19	23/11/19	24/11/19
<b>DEI (durée d'écoute par individu)</b>	03:29:57	03:21:26	03:24:50	03:16:42	03:24:39	03:38:09	04:01:54

### Phase 1.2- Groupe de tranches horaires

Entre 6H et 9H, nous pouvons voir que le **dimanche**, il y a beaucoup moins de personnes qui regardent la télévision (*Annexe VI*). Nous pouvons émettre l'hypothèse que c'est un jour de repos. La proportion de personnes qui regardent la télévision pour tous les niveaux d'ancienneté confondus a considérablement **diminué entre 6H-9H**. Puis le vendredi, la proportion augmente de nouveau. En audience consolidé (diffusion live consolidée du replay et du différé), nous pouvons voir que la proportion de téléspectateur **ayant 10 ans d'ancienneté ou plus est nettement inférieur qu'en live**. Nous pouvons supposer que les personnes âgées de plus de 65 ans consomment moins les chaînes en différé. Nous pouvons voir que la durée d'écoute chez les téléspectateurs de 10 ans d'ancienneté ou plus est plus élevée le vendredi que le lundi.

Tableau 4-Durée d'écoute par individu de 10 ans ou plus d'ancienneté en mode Live

Période	18/11/19	19/11/19	20/11/19	21/11/19	22/11/19	23/11/19	24/11/19
<b>DEI (durée d'écoute par individu)</b>	00:08:58	00:08:55	00:09:06	00:09:27	00:09:15	00:07:32	00:05:41

Entre 12H et 14H30 : **Le week-end**, la proportion de téléspectateurs augmente bien qu'il y **ait plus de téléspectateurs de 10 ans ou plus d'ancienneté. Le week-end, le taux moyen est plus important que la semaine** que ce soit en mode Live ou consolidé. **Les plus anciens panélistes restent ceux qui regardent le plus la télévision** que ce soit le week-end ou la semaine. Chez les panélistes de 10 ans ou plus d'ancienneté, la durée d'écoute moyenne la plus élevée est le dimanche.

Entre 21H et 23H30, tout comme les autres tranches horaires précédentes, les téléspectateurs consomment essentiellement la télévision le dimanche. Néanmoins, en semaine, le second pic n'est pas le mercredi mais le vendredi. La durée d'écoute chez les panélistes ayant 10 ans ou plus d'ancienneté sur cette tranche horaire est plus importante le lundi et le dimanche.

Les panélistes ayant 10 ans d'ancienneté ou plus sont ceux qui regardent le plus la télévision que ce soit le dimanche entre 12H et 14H30, entre 18H et 21H ou encore entre 21H et 23H30. Néanmoins, en moyenne ils écoutent beaucoup la télévision le lundi et le dimanche entre 21H-23H30 et le vendredi entre 6H-9H.

### Phase 1.3- Croiser les profils avec leur audience

Cette analyse sera réalisée après la rédaction de ce rapport.

## 5 Conclusion

Mon stage chez Médiamétrie a été très didactique, c'était pour moi une première expérience aussi bien en grande entreprise que dans le domaine de la Statistique et de l'Informatique. J'ai pu découvrir le métier de Data Analyst en travaillant sur l'étude des profils et des audiences TV des panélistes.

La principale difficulté est arrivée au début, il s'agissait de comprendre rapidement et de manière éclairée le panel Médiamat c'est-à-dire d'où provenaient les données, de comprendre son fonctionnement. Malgré la complexité due au contexte c'est-à-dire moitié du temps en télétravail et le reste en présentiel. Je pense que ça n'a pas été simple pour prendre en main un sujet de stage dans ces conditions. C'était une expérience enrichissante et stimulante.

Ce stage a validé mon souhait de continuer mes études dans ce domaine par le biais d'une école d'ingénieur comme l'INGEDIS, je souhaite en effet me spécialiser dans le Big Data et l'Intelligence Artificielle.

## 6 Ressources

### 6.1 Tables des figures

Figure 1-Organisation de la DDS (Direction Data Science) .....	8
Figure 2-Processus de Watermarking .....	9
Figure 3-Processus de restitution d'audience quotidienne .....	12
Figure 4-Répartition des panélistes dans le foyer .....	17
Figure 5-Age du chef du foyer qui ont dix ans ou plus d'ancienneté .....	17
Figure 6-Nombre de TV actives et réception Hertzienne par niveau d'ancienneté (en %). ....	18
Figure 7-Répartition des panélistes ayant dix ans ou plus d'ancienneté (en %). ....	24
Figure 8-Présence d'enfant de moins de 15 ans chez les foyers de 10 ans ou plus d'ancienneté .....	24
Figure 9-Réponse à la question "êtes-vous abonnés à Canalsat ?" des panélistes ayant 10 ans ou plus d'ancienneté .....	24
Figure 10-Réponse à la question "Souscrivez-vous à la VOD ?" (En %) .....	24
Figure 11-Proportion de téléspectateurs dans le panel pour le mode Live sur la tranche horaire 6-9h. ....	24
Tableau 1-Ticket d'audience en mode Live.....	11
Tableau 2-Ticket d'audience en mode différé du jour même .....	11
Tableau 3-Durée d'écoute par individu de 10 ans ou plus d'ancienneté en mode Live .....	20
Tableau 4-Durée d'écoute par individu de 10 ans ou plus d'ancienneté en mode Live.....	20

### 6.2 Glossaire

Quelques définitions techniques :

*AC/TC* : audience cumulée ou taux cumulé

*Catch-up* : télévision de rattrapage, une émission, film ou série TV est proposé temporairement en streaming vidéo

*DEI/DET* : durée d'écoute par individu ou par téléspectateur

*Live* : diffusion en direct

*Panel* : échantillon de personnes auprès desquelles est faite une enquête d'opinion

*PDA/PDM* : part d'audience ou part de marché

*Session de visionnage* : durée pendant laquelle un individu regarde un programme télévisé

*Svod* : « subscription video on demand », abonnement à un service de vod

*Vod* : « video on demand », vidéo à la demande

## Annexe I : Organigramme de Médiamétrie



Annexe II : dictionnaire des données

Critère socio-démographique	Nom SAS	Modalité
CSP de la personne de référence	CSP	0- AGRICULTEUR
		1- ARTISAN- COMMERCEANT- CCHEF D'ENTREPRISE
		3-PROFESSION INTERMEDIAIRE
		4- EMPLOYE
		5- OUVRIER QUALIFIE
		6- OUVRIER NON QUALIFIE
		7- RETRAITE
		8- ETUDIANT-ELEVE-MILITAIRE DU CONTINGENT
		9- AUTRES INACTIFS
Nombre de personne dans le foyer	NPF	0 -NON REPONSE
		1 - 1 PERSONNE
		2-2 PERSONNES
		3-3 PERSONNES
		4-4 PERSONNES
Région		0 -NON REPONSE
		1 - ILE DE FRANCE
		2- NORD
		3- EST
		4 - BASSIN PARISIEN EST
		5- BASSIN PARISIEN OUEST
		6- OUEST
		7- SUD OUEST
		8- SUD EST
		9- MEDITERRANEE



Taille de l'agglomération	TAGG	0- COMMUNES RURALES 1-AGGLO 2-5 000 HAB 2- AGGLO 5-10 000 HAB 3- AGGLO 10-20 000 HAB 4- AGGLO 20-50 000 HAB 5- AGGLO 50-100 000 HAB 6- AGGLO 100- 200 000 HAB 7- AGGLO 200 000 HAB 8-AGGLO PARIS HORS PARIS VILLE 9-PARIS VILLE
Enfant de moins de 15 ans	NEF15	0 -Aucun 1 -AU MOINS 1
Age du chef de famille	AGECM	0 - NON REPONSE 1 -35 ans 2 -35-49 ANS 3 -50-64 ANS 4 -65 ANS ET +
<b>Critères d'équipement, abonnement et type de réception</b>		
Equipé d'un T.V. HD		1- OUI 2- NON
Nombre de postes TV	NPOTT	0- NON REPONSE 1- 1 POSTE 2- 2 POSTES ET +
Abonnement Canalsat	ABO-	1- OUI 2- NON
Réception ADSL/ FIBRE ou sont		1- OUI 2- NON
Réception Hertzienne	RECEPTION_HERTZIENNE	0 - NON 1- OUI
Souscription à la VOD	SOUSCRIPTION SVOD	0- NON REPONSE 1- OUI 2- NON
Possession d'un boîtier TV type Chromecast, Google TV, Apple		0- NON REPONSE 1- OUI 2- NON

### Annexe III : calendrier prévisionnel détaillé

Semaines	Activités
6 juil. 2020 au 10 juil. 2020	<ul style="list-style-type: none"> <li>6 juil. 2020 <ul style="list-style-type: none"> <li>Mise en place du poste</li> <li>Lectures : MMat en Bref / Book MMat / Présentation DDS et organigramme</li> </ul> </li> <li>7 juil. 2020 <ul style="list-style-type: none"> <li>Lectures : suite Book MMat / Général MM</li> </ul> </li> <li>8 juil. 2020 <ul style="list-style-type: none"> <li>Formation Restit TV et lecture bases SAS</li> <li>lecture de l'audit médiamat 2019</li> <li>Réunion présentation de la base de donnée</li> </ul> </li> <li>9 juil. 2020 <ul style="list-style-type: none"> <li>Manipulation de Resti TV + documentation dessus</li> <li>Prise de connaissance de la base de données</li> </ul> </li> <li>10 juil. 2020 <ul style="list-style-type: none"> <li>Création de la variable niveau d'ancienneté sur SAS</li> <li>Début de création des regroupements des classes de l'année d'entrée dans le Panel</li> <li>Rédaction du début du rapport ( la partie présentation de Médiamétrie)</li> </ul> </li> </ul>
13 juil. 2020 au 17 juil. 2020	<ul style="list-style-type: none"> <li>13 juil. 2020 <ul style="list-style-type: none"> <li>Continuation de la rédaction du rapport de stage</li> </ul> </li> <li>15 juil. 2020 <ul style="list-style-type: none"> <li>Regroupement de classes fini</li> <li>Début des analyses univariés et bivariés</li> </ul> </li> <li>16 juil. 2020 <ul style="list-style-type: none"> <li>Création de macro pour la mise en forme des différentes variables sur un même tableau afin de faciliter l'interprétation et la macro permettant l'incrémentation des valeurs à différentes variables</li> </ul> </li> <li>17 juil. 2020 <ul style="list-style-type: none"> <li>Réalisation de différents onglets Excel avec les 4 groupes : 1 an / 2 à 5 ans / 6 à 10 ans / + de 10 ans à l'aide d'une macro</li> </ul> </li> </ul>

20 juil. 2020 au 24 juil. 2020	<ul style="list-style-type: none"><li>• 20 juil. 2020<ul style="list-style-type: none"><li>◦ Début des analyses avec les graphiques</li><li>◦ Sélection de certains critères</li><li>◦ Début de rédaction de la note méthodologique</li></ul></li><li>• 21 juil. 2020<ul style="list-style-type: none"><li>◦ Rédaction de la note méthodologique ( suite)</li><li>◦ Graphiques</li><li>◦ Établir des profils types pour chaque niveau d'ancienneté</li></ul></li><li>• 22 juil. 2020<ul style="list-style-type: none"><li>◦ Rédaction de la suite du rapport de stage</li><li>◦ Attente de validation des premières analyses effectuées</li><li>◦ Problème de classe recommencer toutes les analyses</li></ul></li><li>• 23 juil. 2020<ul style="list-style-type: none"><li>◦ Achèvement des premières analyses</li><li>◦ Rendu d'un premier rapport</li><li>◦ Début du PowerPoint pour la soutenance, suite du rapport de stage</li></ul></li><li>• 24 juil. 2020<ul style="list-style-type: none"><li>◦ Correction des premières analyses</li><li>◦ Rajout des trois autres variables</li></ul></li></ul>
-----------------------------------	---

27 juil. 2020 au 31 juil. 2020	<ul style="list-style-type: none"> <li>27 juil. 2020 <ul style="list-style-type: none"> <li>Poursuite des analyses et des corrections</li> </ul> </li> <li>28 juil. 2020 <ul style="list-style-type: none"> <li>Relecture du Book Médiamat ( partie audience)</li> <li>Réunion pour faire le pont sur la deuxième mission</li> <li>Commencement du plan d'analyse pour la deuxième partie de la mission</li> </ul> </li> <li>29 juil. 2020 <ul style="list-style-type: none"> <li>Réunion sur le plan d'analyse audience</li> <li>Manipulation de Restit'TV</li> <li>Première analyse sur toutes les chaines et le panel sur Restit'TV</li> </ul> </li> <li>30 juil. 2020 <ul style="list-style-type: none"> <li>Préparation de la nouvelle base de donnée pour les audiences</li> <li>Début des analyses sur la nouvelle base</li> </ul> </li> <li>31 juil. 2020 <ul style="list-style-type: none"> <li>Début de rédaction de la suite du rapport sur les audiences</li> <li>Suite des analyses sur les audiences</li> </ul> </li> </ul>
3 août 2020 au 7 août 2020	<ul style="list-style-type: none"> <li>3 août 2020 <ul style="list-style-type: none"> <li>Analyses de l'audience sur TTV et les tranches horaires (3-27h et 6-9h)</li> </ul> </li> <li>4 août 2020 <ul style="list-style-type: none"> <li>Analyses de l'audience sur TTV sur les tranches horaires( suite 6h-9h + début 12h-14h30)</li> </ul> </li> <li>5 août 2020 <ul style="list-style-type: none"> <li>Analyses de l'audience sur TTV sur les tranches horaires( suite 12h-14h30 + début 18h-21h)</li> </ul> </li> <li>6 août 2020 <ul style="list-style-type: none"> <li>Analyse de l'audience sur TTV sur les tranches horaires( 18h-21h+ début 21h-23h30)</li> </ul> </li> <li>7 août 2020 <ul style="list-style-type: none"> <li>Rédaction du rapport de stage + vérification de la cohérence des analyses</li> </ul> </li> </ul>
10 août 2020 au 14 août 2020	<ul style="list-style-type: none"> <li>10 août 2020 <ul style="list-style-type: none"> <li>Réunion de présentation des analyses</li> <li>Suite du rapport de stage</li> </ul> </li> <li>11 août 2020 <ul style="list-style-type: none"> <li>Correction sur les analyses</li> </ul> </li> <li>12 août 2020 <ul style="list-style-type: none"> <li>Suite des analyses</li> </ul> </li> <li>13 août 2020 <ul style="list-style-type: none"> <li>Poursuite des analyses</li> <li>Correction du rapport de stage</li> </ul> </li> </ul>

## Annexe IV : Analyses socio-démographiques

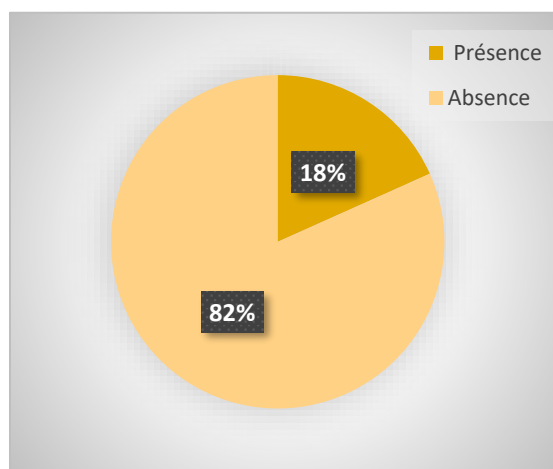


Figure 8-Présence d'enfant de moins de 15 ans chez les foyers de 10 ans ou plus d'ancienneté

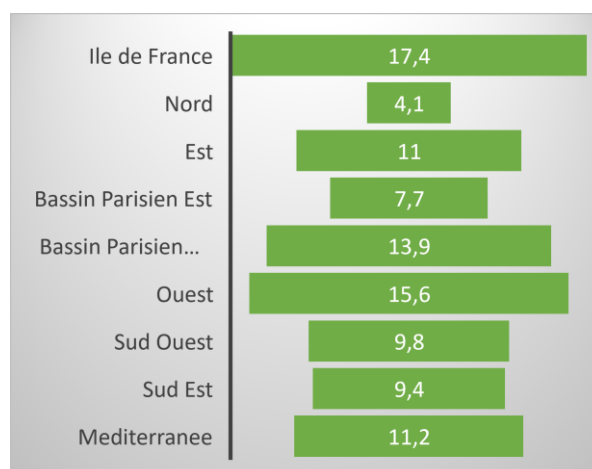


Figure 7-Répartition des panélistes ayant dix ans ou plus d'ancienneté (en %)

## Annexe V : Analyses d'abonnement

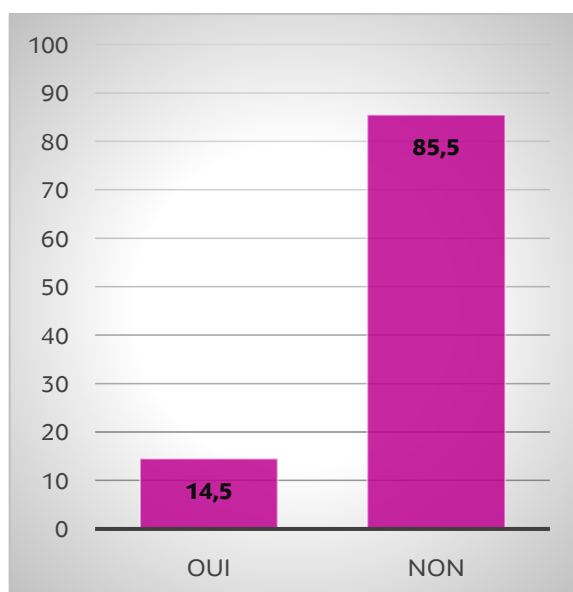


Figure 9-Réponse à la question "êtes-vous abonnés à Canalsat ?" des panélistes ayant 10 ans ou plus d'ancienneté

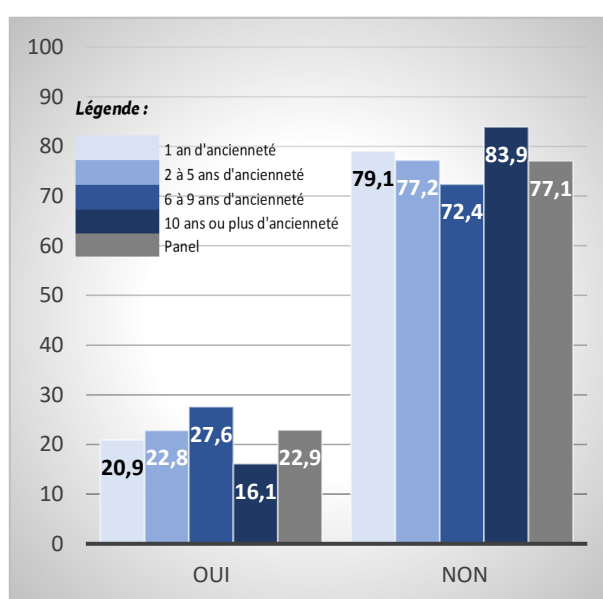
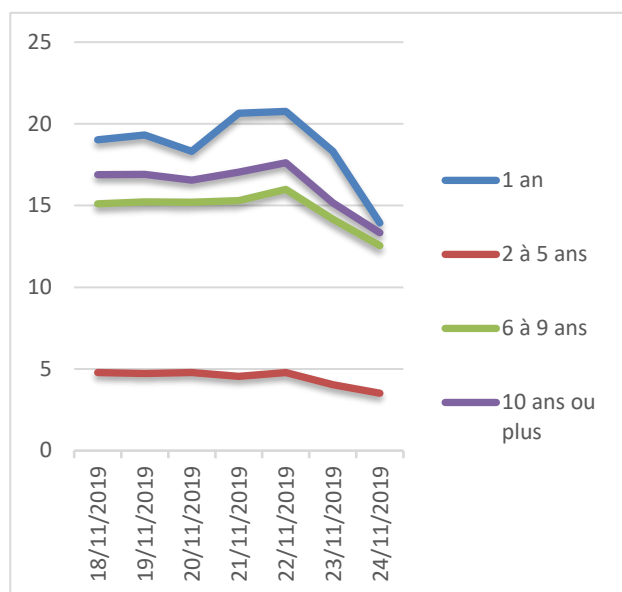


Figure 10-Réponse à la question "Souscrivez-vous à la VOD ?" (En %)

Annexe VI : Analyses audience



*Figure 11-Proportion de téléspectateurs dans le panel pour le mode Live sur la tranche horaire 6-9h*