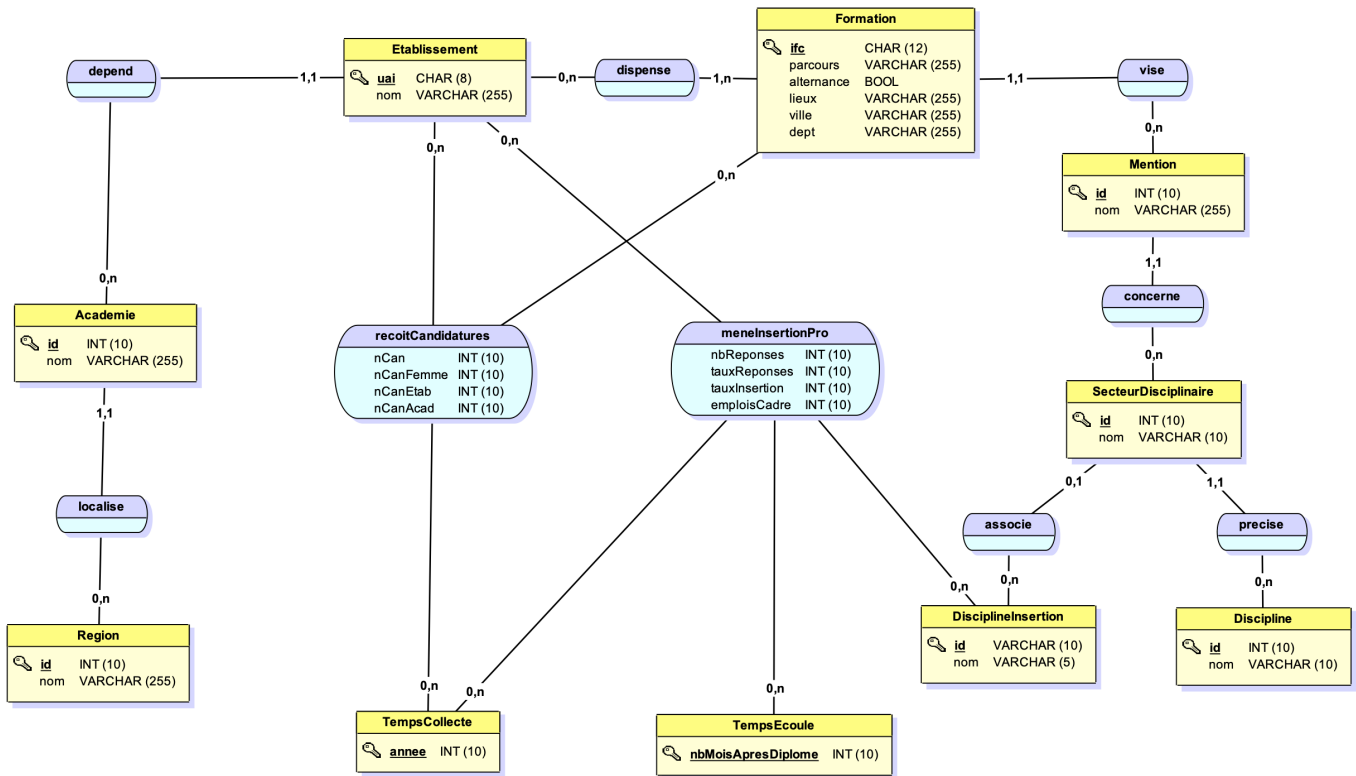


Modèle de données



Note: les attributs des associations "recoitCandidatures" et "meneInsertionPro" ne sont exhaustifs sur le modèle, par soucis de lisibilité

Elements de compréhension

À la lecture des deux jeux de données et suite à leur analyse, plusieurs informations étaient importantes pour construire ce modèle de données :

- Candidatures :
 - les liens de hiérarchie stricte existants (qui mènent à des associations 1,1 - 1,N) :
 - établissement -> académie -> région
 - formation -> mention -> secteur disciplinaire -> discipline
 - remarque : le parcours, car définit par l'université, n'ajoute pas de niveau supplémentaire dans la hiérarchie précédente
 - la formation peut appartenir à plusieurs établissements
 - d'après les texte une formation pourrait être portée par 2 établissements, même si toutefois, les données montrent un lien de hiérarchie unique entre formation et établissement (i.e. : une formation n'a qu'un établissement). Nous considérons cela pour s'assurer que notre modèle permette de prendre en compte cette possibilité
 - même si la colonne annee a toujours la même valeur, celle-ci montre que d'autres collectes (et d'autres années) auront lieu, cette information est donc à considérer
 - des identifiants existent déjà pour les régions (acad_reg), les académies (acad), les établissements (uai), les disciplines (discipline), les secteurs disciplinaires (secteur_disci) et les formations (ifc)

- les champs inm et inmp ne peuvent pas être considérés comme identifiants de mention et parcours respectivement car ils sont propre à chaque université
- une collecte de statistique de candidatures (i.e.: une ligne du fichier) dépend donc de :
 - l'année de collecte
 - la formation
 - l'établissement (car non n'avons pas de hierarchie directe avec la formation)
- Insertion professionnelle :
 - pas de formation présente dans ce jeu de données : on travaille ici à un niveau plus haut d'aggregation
 - hiérarchies présentes :
 - discipline -> domaine
 - etablissement -> académie
 - on a également une année de collecte mais aussi 2 collectes différentes par année selon le nombre de mois écoulé depuis le diplôme
 - les disciplines et domaines ne sont alignés aux mentions, secteurs disciplinaires et disciplines du fichier de candidatures.
 - une collecte de statistique d'insertion pro (i.e.: une ligne du fichier) dépend donc de :
 - l'année de collecte
 - le nombre de mois depuis l'obtention du diplôme
 - une discipline "d'insertion profesionnelle"
 - un établissement
- fusion des deux sources
 - les 2 sources n'ont pas la même granularité (les collectes ne dépendent pas des mêmes informations). Notamment, les candidatures sont à l'échelle d'une formation alors que les insertions pro sont à l'échelle d'un établissement
 - les établissements utilisent bien le même code (uai)
 - il n'y a pas de bijectivité entre discipline de candidatures et discipline d'insertion pro (pas de possibilité de trouver un lien unique entre l'un et l'autre), et entre domaine d'insertion pro et tout autre éléments de la hiérarchie mention -> sect. disciplinaire -> discipline des candidatures
 - il faut donc analyser les données pour tenter de trouver un mapping très certainement imparfait
 - il est possible d'avoir un mapping hierarchique "correcte" : secteur disciplinaire -> discipline d'insertion

Entités

Région : une région est une zone administrative du territoire français. Elle est désignée par un identifiant unique (nombre) et un nom, unique également.

Académie : circonscription administrative du Ministère de l'Education, dont dépend les universités en France. Définie par son identifiant unique (nombre) et son nom ; appartient à une région.

Etablissement : Etablissement public d'enseignement supérieur (pour la plupart du temps : une université). Identifié par son *uai*, identifiant unique (chaîne de caractère) et qualifié par son nom ; rattaché à une

académie.

Discipline : Discipline d'enseignement de la nomenclature officielle BCN (Base centrale des nomenclatures). Plus haut niveau de regroupement disciplinaire. Défini par un identifiant unique (nombre) et un nom.

DisciplineInsertion : Discipline utilisée pour les statistiques d'insertion professionnelle, non officiel. Défini par un identifiant unique (chaîne de caractère) et un nom.

SecteurDisciplinaire : Secteur disciplinaire de la nomenclature officielle BCN, appartenant et raffinant une discipline. Défini par un identifiant unique (nombre) et un nom. Lié également à une discipline d'insertion professionnel.

Mention : Mention de diplôme de master, définit nationalement. Défini par son identifiant unique (nombre auto-généré), son nom, et dépendante d'un secteur disciplinaire.

Formation : Formation dispensée dans un ou plusieurs établissements (si co-habilitation), délivrant un master rattaché à une mention et précisé possiblement par un parcours. Outre son identifiant unique national de formation, l'*ifc*, son parcours, la formation est qualifiée par un indicateur d'alternance possible (ou obligatoire), un lieu d'enseignement (adresse complète ou non), la ville et le code de département d'enseignement (également présent dans le lieu).

Tempscollecte : entité calendaire qualifiant une année de collecte statique, que ce soit pour les candidatures ou l'insertion professionnelle.

TempsEcoule : entité temporelle, qualifiant un nombre de mois écoulés entre une collecte statistique sur l'insertion professionnelle et l'année de diplomation du répondant.

Associations importantes

recoitCandidatures : informations de candidatures pour une année, une formation et un établissement. Contient tous les indicateurs statistiques des candidatures. **menInsertioPro** : informations d'insertion professionnelle pour une année, un nombre de mois après diplôme, un établissement et une discipline d'insertion pro.

Notes

Ce modèle est bien un modèle conceptuel. Il ne relate pas la manière dont les données seront implantées en BD (le modèle logique). Par exemple, il n'y aurait pas d'intérêt à créer une table pour les années et une table pour le nombre de mois écoulé après diplôme. En effet, ces deux tables n'auraient respectivement qu'un attribut, leur clé primaire, recopiée dans les autres tables s'en servant comme clé étrangère.

API REST

Un ensemble de services REST permet d'accéder facilement aux données unifiées sur les masters, et d'y rechercher des statistiques de candidatures et d'insertion professionnelles. Cette API REST est publique, sans authentification ni limitation particulière et ne fournit que des données au format JSON. Cette API se divise en deux parties :

- l'accès aux informations contextuelles (les entités du modèle : établissements, disciplines...) qui permet d'obtenir les informations de chacune des entités :
- l'accès au service de recherche qui permet une requête de collecte avancée d'après des filtres et des spécifications des données à retourner.

Accès aux informations contextuelles

Académie et leurs régions

[GET] /api/rest/academies

Retourne la liste des academies (*id, nom*) et de leur région associée (*regionId, regionNom*).

Paramètres de requête possible : aucun

Code de retour possibles :

- 200 : OK
- 500 : Erreur serveur

Exemple de retour :

```
[
  {
    "id": 1,
    "nom": "Paris",
    "regionId": 11,
    "regionNom": "Île-de-France"
  },
  {
    "id": 2,
    "nom": "Aix-Marseille",
    "regionId": 93,
    "regionNom": "Provence-Alpes-Côte d'Azur"
  }
]
```

Etablissements

[GET] /api/rest/etablissements

Retourne la liste des établissements (*uid, nom*) et de l'id de l'académie d'appartenance (*academiId*).

Paramètres de requête possible : aucun

Code de retour possibles :

- 200 : OK
- 500 : Erreur serveur

Exemple de retour :

```
[
  {
    "uai": "0062205P",
    "nom": "Université Côte d'Azur",
    "academieId": 23
  },
  {
    "uai": "0101060Y",
    "nom": "Université de technologie de Troyes",
    "academieId": 19
  }
]
```

Secteurs disciplinaires et discipline associées

[GET] /api/rest/secteurs-disciplinaires

_Retourne la liste des secteur disciplinaires (id, nom), de leurs disciplines auxquelles ils appartiennent (disciplineId, disciplineNom) et de la discipline d'insertion professionnelle (utilisée pour identifier les statistiques idoine) auxquelles ils appartiennent (insDiscId) _

Paramètres de requête possible : aucun

Code de retour possibles :

- 200 : OK
- 500 : Erreur serveur

Exemple de retour :

```
[
  {
    "id": 1,
    "nom": "Mathématiques",
    "disciplineId": 15,
    "disciplineNom": "Sciences fondamentales et applications",
    "insDiscId": "disc14"
  },
  {
    "id": 2,
    "nom": "Physique",
    "disciplineId": 15,
    "disciplineNom": "Sciences fondamentales et applications",
    "insDiscId": "disc14"
  }
]
```

Mentions de diplômes de Master

[GET] /api/rest/mentions

Retourne la liste des mentions de diplômes de Master (id, nom), de l'id de secteur disciplinaire d'appartenance (secDiscId).

Paramètres de requête possible : aucun

Code de retour possibles :

- 200 : OK
- 500 : Erreur serveur

Exemple de retour :

```
[
  {
    "id": 0,
    "nom": "ACOUSTIQUE",
    "secDiscId": 11
  },
  {
    "id": 1,
    "nom": "ACOUSTIQUE ET MUSICOLOGIE",
    "secDiscId": 11
  }
]
```

Formations

[GET] /api/rest/formations

Retourne la liste des formations (ifc) avec les informations de diplôme délivré (parcours, id de mention mentionId, id de secteur disciplinaire secDiscId), les informations de modalités pédagogiques (indicateur booléen d'alternance proposé) et de lieux de formation (lieu complet lieux, ville, code de département dept)

Paramètres de requête possible :

- **uai** : filtre sur identifiant d'établissement. Optionnel, repetable.
- **sdid** : filtre sur identifiant de secteur disciplinaire. Optionnel, repetable.
- **dept** : filtre sur departement de localisation de la formation. Optionnel, repetable.

Code de retour possibles :

- 200 : OK
- 400 : paramètre de requête invalide
- 500 : Erreur serveur

Exemple de requête

GET <http://.../api/rest/formations?uai=0755976N&uai=0931238R&sdid=2&sdid=6>

Cette requête récupère toutes les formations dispensées dans les établissements d'identifiants 0900816N1CXR ou 0900816NDLTV, et dont l'identifiant du secteur disciplinaire est 2 ou 6.

Exemple de retour :

```
[
  {
    "ifc": "0900816N1CXR",
    "parcours": "Ingenierie Physique des Energies - (master, sciences, physique)",
    "alternance": false,
    "lieux": "UFR Physique - Campus Grands Moulins - PARIS (75), Université Paris Cité - PARIS (75), UFR Physique - PARIS (75)",
    "etabUai": "0755976N",
    "mentionId": 242,
    "secDiscId": 2,
    "ville": "PARIS",
    "dept": "75"
  },
  {
    "ifc": "0900816NDLTV",
    "parcours": "Ingenierie Physique des Energies - (master, sciences, physique)",
    "alternance": true,
    "lieux": "UFR Physique - Campus Grands Moulins - PARIS (75), Université Paris Cité - PARIS (75), UFR Physique - PARIS (75)",
    "etabUai": "0755976N",
    "mentionId": 242,
    "secDiscId": 2,
    "ville": "PARIS",
    "dept": "75"
  }
]
```

Recherche statistique

Le service de recherche statistique permet de récupérer à la fois des statistiques de candidatures et d'insertions professionnelles à partir d'un ensemble de filtres donné. Ce service est centré autour d'un unique point d'entrée, une requête POST dans laquelle on transmet au format JSON les filtres et paramètres de statistiques à retourner.

Template de recherche

[GET] /api/rest/stats/search

Retourne le modèle de requête à transmettre pour une recherche, incluant tous les filtres possibles et les paramètres de données à retourner. Ce modèle explique chaque filtre et chaque paramètre.

Paramètres de requête possible : aucun

Code de retour possibles :

- 200 : OK
- 500 : Erreur serveur

Exemple de retour :

```
{
  "filters": {
    "##Description##": "filtres de statistiques. Multiples filtres possibles, combinés en \"ET\"",
    "regionIds": "Identifiant de région (int) ou tableau d'identifiants de région. Optionnel.",
    "academieIds": "Identifiant d'académie (int) ou tableau d'identifiants d'académie. Optionnel.",
    "etablissementIds": "Identifiant d'université (str) ou tableau d'identifiants d'université. Optionnel.",
    "mentionIds": "Identifiant de mention (int) ou tableau d'identifiants de mention. Optionnel.",
    "secteurDisciplinaireIds": "Identifiant de secteur disciplinaire (int) ou tableau d'identifiants de secteurs disciplinaires. Optionnel.",
    "disciplineIds": "Identifiant de discipline (int) ou tableau d'identifiants de disciplines. Optionnel.",
    "annees": "Année de collecte (int) ou tableau d'années de collecte. Optionnel.",
    "anneeMin": "Année minimale de collecte, incluse (int). Optionnel.",
    "anneeMax": "Année maximale de collecte, exclue (int). Optionnel.",
    "moisApresDiplome": "nombre de mois après obtention du diplôme (int). Ne s'applique que pour les statistiques d'insertion professionnelle. Optionnel."
  },
  "harvest": {
    "##Description##": "Choix des données retournées. Multiples options possibles.",
    "typeStats": "Type de statistiques retournées (str). Valeur possible: {'all', 'candidatures', 'insertionsPro'}. Optionnel. Valeur par défaut : 'all'",
    "candidatureDetails": "Element de statistiques de candidature à retourner (str) ou tableau d'éléments. Valeurs possibles: {'all', 'general', 'experience', 'origine'}. Optionnel. Valeur par défaut : 'general'",
    "insertionProDetails": "Element de statistiques d'insertion professionnelle à retourner (str) ou tableau d'éléments. Valeurs possibles: {'all', 'general', 'emplois', 'salaire', 'refRegion'}. Optionnel. Valeur par défaut : 'general'"
  }
}
```

Recherche de statistiques

[POST] /api/rest/stats/search

_Effectue une recherche statistiques sur les candidatures et/ou les insertions professionnelles à partir des filtres et paramètre de données à retourner fournis en corps de requête

Paramètres de requête possible : aucun

Corps de requête : doit respecter le modèle de template, fournit par la requête précédente.

Code de retour possibles :

- 200 : OK
- 400 : données transmises invalide (format incorrect ou non respect du template de requête)
- 500 : Erreur serveur

Strucure de réponse :

La structure de réponse générale est la suivante :

- *request* : Objet de la requête transmise
- *candidatures* : Tableau de statistiques de candidatures (décrites ci-dessous), si demandé
- *insertionsPro* : Tableau de statistiques d'insertion pro. (décrites ci-dessous), si demandé

La structure de statistique de candidatures est la suivante :

- *identifiants* : informations permettant d'identifier de manière unique la statistique
 - *anneeCollecte* : année de collecte de la statistique
 - *etabUai* : identifiant (uai) de l'établissement
 - *formationIfc* : identifiant (ifc) de la formation
- *relations* : informations de lien de la statistique avec d'autres entités
 - *academiId* : identifiant de l'académie
 - *regionId* : identifiant de la région
 - *secDisclId* : identifiant du secteur disciplinaire
 - *disclId* : identifiant de la discipline
 - *mentionId* : identifiant de la mention de la formation
- *general* : mesures générales de la statistique
 - *capacite* : capacité (nombre de places maximales) annoncée pour la formation
 - *nb* : nombre de candidat.es
 - *nbFemmes* : nombre de candidates
 - *clas* : nombre de candidat.es classé.es
 - *clasFemmes* : nombre de candidates classées
 - *prop* : nombre de candidat.es ayant reçu une proposition d'acceptation
 - *propFemmes* : nombre de candidates ayant reçu une proposition d'acceptation
 - *accept* : nombre de candidat.es ayant accepté la proposition
 - *acceptFemmes* : nombre de candidates ayant accepté la proposition
 - *nbComp* : nombre de candidats en effectif complémentaire
 - *rangDernier* : rang (dans le classement) du dernier candidat ayant reçu une proposition
- *experience* : mesures de candidatures par type de formation précédente des candidats
 - *lg3*: licence générale
 - *nb* : nombre de candidat.es provenant de ce type de formation
 - *nbFemmes* : nombre de candidates provenant de ce type de formation
 - *clas* : nombre de classé.es

- *clasFemmes* : nombre de femmes classées
 - *prop* : nombre de candidat.es ayant reçu une proposition d'acceptation
 - *propFemmes* : nombre de femmes ayant reçu une proposition d'acceptation
 - *accept* : nombre de candidat.es ayant accepté la proposition
 - *acceptFemmes* : nombre de femmes ayant accepté la proposition
- *lp3* : licence professionnelle, même structure que pour lg3
- *master* : master, même structure que pour lg3
- *autre* : autre formation, même structure que pour lg3
- *nonInscrit* : non inscrit l'année précédente, même structure que pour lg3
- *origine* : mesures de candidatures par origine géographique
 - *etablissement* : candidats provenant du même établissement (université)
 - *nb* : nombre de candidats
 - *clas* : nombre de candidat.es classé.es
 - *prop* : nombre de candidat.es ayant reçu une proposition d'acceptation
 - *accept* : nombre de candidat.es ayant accepté la proposition
 - *academie* : candidats provenant de la même académie (établissement compris), même structure que pour etablissement
 - *region* : candidats provenant de la même région (académie comprise), même structure que pour etablissement

La structure de statistique de l'insertion professionnelle est la suivante :

- **identifiants** : informations permettant d'identifier de manière unique la statistique
 - **anneeCollecte** : année de collecte de la statistique
 - *etabUai* : identifiant (uai) de l'établissement
 - *moisApresDip* : nombre de mois passée entre l'obtention du diplôme et la collecte des répondants
 - *insDisclD* : identifiant de la discipline d'insertion professionnelle
- *relations* : informations de lien de la statistique avec d'autres entités
 - *academiD* : identifiant de l'académie
 - *regionD* : identifiant de la region
 - *secDisclDs* : identifiants des secteurs disciplinaire qui correspondent à la discipline d'insertion pro. de la statistiques
 - *disclDs* : identifiants des disciplines qui correspondent à la discipline d'insertion pro. de la statistiques
 - *mentionDs* : identifiants des mentions qui correspondent à la discipline d'insertion pro. de la statistiques
- *general* : mesures générales de la statistique
 - *nbResponses* : nombre de réponses
 - *tauxReponse* : taux de réponse
 - *pbEchantillon* : indicateur d'un problème avec la statistique (ex.: échantillon trop faible)
 - *pbEchantillonRaison* : raison du problème avec la statistique
- *emplois* : mesures des types d'emploi et de la sociographie des emplois
 - *cadreProIntermediaire* : nombre de répondants cadres ou de professions intermédiaire
 - *cadre* : nombre de répondant.es cadres
 - *stable* : nombre de répondant.es ayant un emploi stable
 - *tempsPlein* : nombre de répondant.es ayant un emploi à temps plein

- *exterieurRegion* : nombre de répondant.es ayant un emploi en dehors de la région
 - *femmes* : nombre de répondantes
 - *boursier* : nombre de répondant.es étant boursier
- *salaire* : mesures des salaires
 - *netMedianTempsPlein* : médiane des salaires net des répondant.es
 - *brutAnnuelEstime* : salaire brut annuel estimé (estimation sur la base des réponses)
- *refRegion* : mesures des emplois et salaires de la région, utilisable comme base de référence de comparaison
 - *tauxChomageRegional* : taux de chômage régional
 - *netQ1Regional* : premier quartile des salaires net temps plein de la région
 - *netMedianRegional* : médiane des salaires net temps plein de la région
 - *netQ3Regional* : troisième quartile des salaires net temps plein de la région

Exemple de requête

POST <http://.../api/rest/stats/search>

Corps de requête :

```
{
  "filters": {
    "academieIds": 1,
    "secteurDisciplinaireIds": [1, 2],
    "anneeMin": 2016,
    "anneeMax": 2024,
    "moisApresDiplome": 30
  },
  "harvest": {
    "typeStats": "all",
    "candidatureDetails": ["general", "origine"],
    "insertionProDetails": ["general", "salaire"]
  }
}
```

Cette requête récupère les statistiques de candidatures et d'insertion pro. pour toutes les années de collecte comprise entre 2016 incluse et 2024 exclue, pour les secteurs disciplinaire d'id 1 et 2 et pour l'académie d'id 1. Concernant les statistiques d'insertion pro, ne sont demandées que celles à 30 mois après le diplôme. Il est demandé de retourner, pour les statistiques sur les candidatures, les composantes "général" et "origine" en plus de celles de base (identifiants et relations). Il est demandé de retourner, pour les statistiques sur les insertion pro., les composantes "général" et "salaire" en plus de celles de base (identifiants et relations).

Exemple de réponse (tronquée par souci de lisibilité)

```
{
  "request": {
    "academies_filter": [
      1
    ]
  }
}
```

```

    ],
    "sec_disc_filter": [
      1,
      2
    ],
    "annee_mini_filter": 2016,
    "annee_maxi_filter": 2024,
    ...
  },
  "candidatures": [
    {
      "identifiants": {
        "anneeCollecte": 2023,
        "etabUai": "0755976N",
        "formationIfc": "0900816N1CXR"
      },
      "relations": {
        "academieId": 1,
        "regionId": 11,
        "secDiscId": 2,
        "discId": 15,
        "mentionId": 242
      },
      "general": {
        "capacite": 5,
        "nb": 97,
        ...
      }
    },
    ...
  ],
  "insertionsPro": [
    {
      "identifiants": {
        "anneeCollecte": 2020,
        "etabUai": "0755890V",
        "moisApresDip": 30,
        "insDisc": "disc14"
      },
      "relations": {
        "academieId": 1,
        "regionId": 11,
        ...
      },
      "general": {
        "nbResponses": 65,
        "tauxReponse": 52,
        "pbEchantillon": true,
        "pbEchantillonRaison": "Université dont les résultats sont plus fragiles (taux de réponse inférieur à 50%)"
      },
      "salaire": {
        "netMedianTempsPlein": 2810,
        "brutAnnuelEstime": 43800
      }
    }
  ]

```

```
    }  
    },  
    ...  
  ]  
}
```