

11 novembre 2025

CMN 4500/JOU4500 Journalisme numérique II

Bonavena Gatima, Debora Adebiyi, Djeneba Diakité

Présenté à Jean-Sébastien Marier

Aanalyse exploratoire de données (AED) et proposition

1. Introduction

Dans ce travail, nous analyserons un jeu de données de la Ville d'Ottawa portant sur le revenu total médian en 2020 parmi les bénéficiaires. Cette variable nous permet d'observer combien les gens gagnent dans différents quartiers d'Ottawa et d'identifier d'éventuelles différences importantes entre eux. L'objectif est d'utiliser ces données pour dégager un angle de reportage portant sur les inégalités de revenu dans la ville.

Les données proviennent du recensement de 2021, colligé par Statistique Canada et diffusé ensuite par la Ville d'Ottawa. Elles ont été recueillies au moyen du questionnaire abrégé, envoyé à 25% des ménages canadiens. Le fichier regroupe ensuite ces informations pour chacun des quartiers de la ville.

Le jeu de données est accessible sur [Ottawa ouverte](#) et la version CSV utilisée dans ce travail se trouve sur [GitHub](#).

Ce travail suivra les sections suivantes : obtention des données, compréhension des données (analyse VIMA, nettoyage et analyse exploratoire), formulation d'un récit potentiel, conclusion et références.

2. Obtenir les données

Pour débiter l'analyse, nous avons d'abord téléchargé le fichier CSV contenant les données du *Questionnaire détaillé du recensement de 2021 – Données par quartier* à partir du portail GitHub de JSMarier. Une fois le fichier enregistré sur notre ordinateur, nous avons ouvert Google Feuilles de calcul, créé une nouvelle feuille, puis cliqué sur Fichier > Importer. Dans le menu d'importation, nous avons sélectionné le fichier CSV du recensement et choisi l'option permettant de remplacer la feuille actuelle. Le jeu de données original est ainsi apparu immédiatement dans Google Sheets.

Image 1

Le fichier de Google Feuilles de calcul importé.

Image 2

Le fichier Google Feuilles de calcul des données extraites.

<https://docs.google.com/spreadsheets/d/1VksFI3gRm7LuZUExl3b86OL0BqQ6D7cVPPiMW-EPKZc/edit?>

usp=sharing

3. Comprendre les données

3.1. Analyse VIMA

L'une des méthodes pour explorer la validité des données est d'effectuer une analyse VIMA. Cet acronyme signifie Valide, Invalide, Manquante et aberrante (Statistique Canada, 2022). Cette démarche nous permet d'apprécier la qualité, l'exactitude et la fiabilité de notre jeu de données.

1. Valeur valide : Toutes les valeurs du revenu total médian sont présentes, numériques et cohérentes avec ce que l'on peut attendre pour les quartiers d'Ottawa. Les montants correspondent à une fourchette réaliste pour des revenus médians urbains. Aucune valeur manquante, nulle ou illogique n'a été trouvée, ce qui confirme la validité du jeu de données.
2. Valeur invalide : Aucune valeur invalide n'a été détectée. Une valeur invalide aurait été, par exemple, une cellule contenant du texte (« N/A », « inconnu »), un symbole (\$\$\$) ou un chiffre négatif. La présence exclusive de nombres entiers positifs et cohérents indique que les données ont été correctement saisies.
3. Valeur manquantes : Les valeurs manquantes (cellules vides) sont inexistantes dans notre tableau de données. Chaque variable nominale indique un revenu en dollar ce qui renforce la fiabilité du jeu de données.
4. Valeur Aberrantes : Ce sont celles extrêmement élevées ou extrêmement basses par rapport à ce que l'on s'attendait à observer. Les plus basses dans notre jeu de données sont à 39 200 \$, ce qui représente un écart important avec les quartiers les plus élevés. Mais cela ne veut pas dire qu'elles sont anormales. Aucun revenu ne dépasse de façon démesurée la tendance générale.

En résumé, les données sont valides, exactes et fiables.

3.2. Nettoyage des données

Le nettoyage d'un jeu de données est essentiel pour garantir que l'analyse repose sur des informations justes et fiables. Nous avons donc utilisé trois méthodes différentes afin de corriger les erreurs et nous assurer que nos données sont propres.

La première étape a été d'utiliser l'outil de nettoyage intégré proposé par Google Feuilles de calcul sur la feuille 2: **Données > Nettoyage de données > Suggestions de nettoyage**

Deux modifications nous ont alors été suggérées: supprimer les espaces blancs dans les colonnes A15 et B1. Ceux-ci ont été acceptés.

Ensuite, Nous avons utilisé l'outil Statistiques des colonnes pour examiner les valeurs présentes dans chaque colonne; **Données > Statistiques des colonnes**

Cet outil a permis de remarquer que la cellule A15 contenait seulement l'inscription « Rideau-Rockcliffe », sans le numéro du quartier. Comme les quartiers étaient classés en ordre et que le numéro 13 était manquant, nous avons ajouté « Quartier 13 » dans le tableau. Cette information a également été confirmée en consultant la liste officielle des quartiers sur le site de la Ville d'Ottawa.

Pour terminer, nous avons importé les données dans OpenRefine afin d’effectuer une dernière vérification. Nous avons utilisé le filtre **Facet > Text facet** qui permet d’observer les tendances d’une colonne en regroupant toutes les valeurs uniques qu’elle contient. Un facet sert à repérer rapidement les différents éléments présents dans une variable, à voir combien de fois ils apparaissent et à identifier d’éventuelles anomalies (Exploring Facets | OpenRefine, 2024). En appliquant ce filtre à la colonne « Quartier », nous avons pu vérifier si certaines valeurs se répétaient ou s’il existait des erreurs de saisie. Aucune répétition n’a été détectée. Cependant, nous avons constaté que la valeur « Ville d’Ottawa » avait été incluse par erreur dans la colonne des quartiers. Nous l’avons donc supprimée manuellement dans Google Sheets. Après cette correction, notre jeu de données est propre et prêt pour l’analyse.

Image 3

OpenRefine

Questionnaire détaillé du recensement de 2021 Données par quartier.xlsx

Permalink

Open...

Export

Help

Facet / Filter

Undo / Redo 1 / 1

25 rows

Extensions Wikibase

Refresh

Reset all

Remove all

Quartier

change

25 choices Sort by: name count

Cluster

Alta Vista - Quartier 18 1

Baie - Quartier 7 1

Barrhaven-Est - Quartier 24 1

Barrhaven-Ouest - Quartier 3 1

Beacon Hill-Cyrville - Quartier 11 1

Capitale - Quartier 17 1

Collège - Quartier 8 1

Gloucester-Southgate - Quartier 10 1

Kanata-Nord - Quartier 4 1

Kanata-Sud - Quartier 23 1

Kitchissippi - Quartier 15 1

Knoxdale-Merivale - Quartier 9 1

Orléans-Est-Cumberland - Quartier 1 1

Orléans-Ouest-Innes - Quartier 2 1

Orléans-Sud-Navan - Quartier 19 1

Osgoode - Quartier 20 1

Rideau-Jock - Quartier 21 1

Rideau-Rockcliffe- Quartier 13 1

Rideau-Vanier - Quartier 12 1

Riverside-Sud-Findlay Creek - Quartier 22 1

Rivière - Quartier 16 1

Somerset - Quartier 14 1

Stittsville - Quartier 6 1

Ville d'Ottawa 1

West Carleton-March - Quartier 5 1

Facet by choice counts

▼ All

▼ Quartier

▼ Revenu total médian en 2020 parmi les bénéficiaires (\$)

1. Ville d'Ottawa 50000

2. Orléans-Est-Cumberland - Quartier 1 56400

3. Orléans-Ouest-Innes - Quartier 2 54400

4. Barrhaven-Ouest - Quartier 3 54800

5. Kanata-Nord - Quartier 4 54400

6. West Carleton-March - Quartier 5 56400

7. Stittsville - Quartier 6 59200

8. Baie - Quartier 7 43200

9. Collège - Quartier 8 45600

10. Knoxdale-Merivale - Quartier 9 45200

11. Gloucester-Southgate - Quartier 10 39200

12. Beacon Hill-Cyrville - Quartier 11 45200

13. Rideau-Vanier - Quartier 12 39200

14. Rideau-Rockcliffe- Quartier 13 44800

15. Somerset - Quartier 14 49600

16. Kitchissippi - Quartier 15 64000

17. Rivière - Quartier 16 42800

18. Capitale - Quartier 17 52800

19. Alta Vista - Quartier 18 44000

20. Orléans-Sud-Navan - Quartier 19 57600

21. Osgoode - Quartier 20 56000

22. Rideau-Jock - Quartier 21 58400

23. Riverside-Sud-Findlay Creek - Quartier 22 58000

24. Kanata-Sud - Quartier 23 53600

25. Barrhaven-Est - Quartier 24 48000

Analyse facet dans OpenRefine

Image 4

Questionnaire détaillé du recensement de 2021 - Données par quartier

FichierÉditionAffichageInsertionFormatDonnéesOutilsExtensionsAideDemander à Gemini

100%\$%0.00123Arial11B

1:1Quartier

1	A	B	C	D	E	F	G	H	I	J	K
	Quartier	Revenu total médian en 2020 parmi les bénéficiaires (\$)									
2	Orléans-Est-Cumberland - Quartier 1	56400									
3	Orléans-Ouest-Innes - Quartier 2	54400									
4	Barrhaven-Ouest - Quartier 3	54800									
5	Kanata-Nord - Quartier 4	54400									
6	West Carleton-March - Quartier 5	56400									
7	Stittsville - Quartier 6	59200									
8	Baie - Quartier 7	43200									
9	Collège - Quartier 8	45600									
10	Knoxdale-Merivale - Quartier 9	45200									
11	Gloucester-Southgate - Quartier 10	39200									
12	Beacon Hill-Cyrville - Quartier 11	45200									
13	Rideau-Vanier - Quartier 12	39200									
14	Rideau-Rockcliffe - Quartier 13	44800									
15	Somerset - Quartier 14	49600									
16	Kitchissippi - Quartier 15	64000									
17	Rivière - Quartier 16	42800									
18	Capitale - Quartier 17	52800									
19	Alta Vista - Quartier 18	44000									
20	Orléans-Sud-Navan - Quartier 19	57600									
21	Osgoode - Quartier 20	56000									
22	Rideau-Jock - Quartier 21	58400									
23	Riverside-Sud-Findlay Creek - Quartier 22	58000									
24	Kanata-Sud - Quartier 23	53600									
25	Barrhaven-Est - Quartier 24	48000									
26											

Jeu de données après le processus de nettoyage

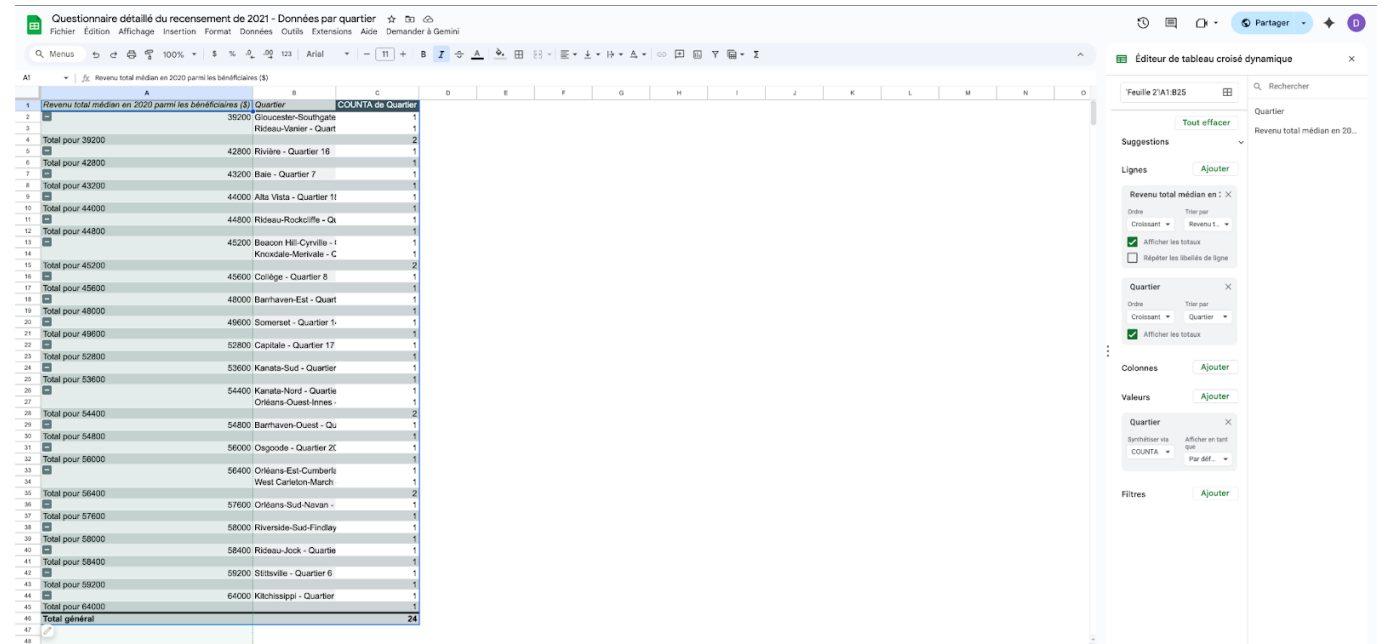
3.3. Analyse exploratoire des données (AED)

Pour commencer l'analyse exploratoire, nous avons créé un tableau croisé dynamique dans Google Feuilles de calcul. L'ensemble du tableau a été sélectionné, puis l'option Données > Tableau croisé dynamique a été appliquée. Dans la nouvelle feuille générée, nous avons choisis :

✳️Ligne : le revenu médian total par bénéficiaire ainsi que les Quartiers ✳️Valeur : les quartiers (option COUNTA)

Ce tableau croisé dynamique a permis de réorganiser automatiquement les revenus médians des 24 quartiers afin de faciliter leur analyse. En triant les valeurs par ordre croissant, nous avons pu repérer rapidement les quartiers aux revenus les plus faibles et ceux aux revenus les plus élevés. Il a aussi permis de compter le nombre de quartiers associés à chaque montant de revenu grâce à la fonction COUNTA. Cette étape nous a aidés à vérifier si certains quartiers partageaient le même revenu médian.

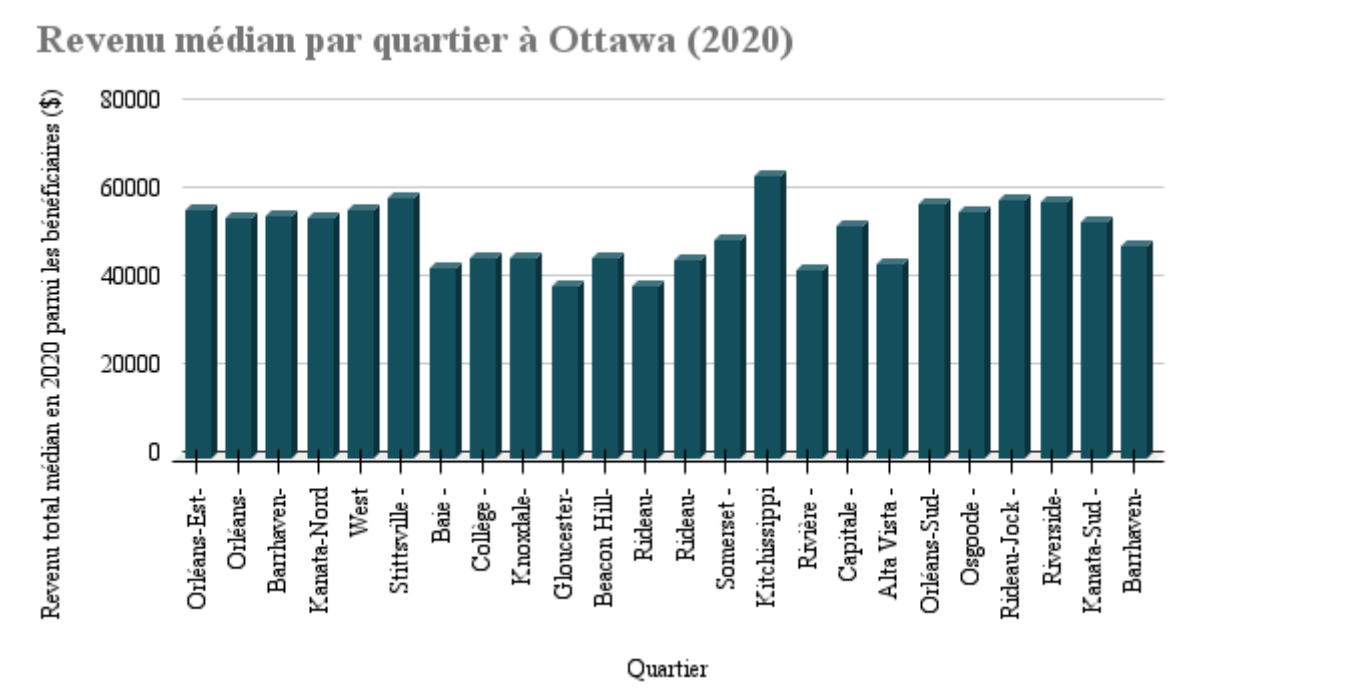
Image 5



Ce tableau croisé dynamique montre le revenu median des quartiers par ordre croissant ainsi que le nombre de quartiers associés à chaque revenu

À partir de ce tableau, un graphique exploratoire a été créé à l'aide de l'outil graphique de Google Feuilles de calcul. L'option Insertion > Graphique a été utilisée, puis l'onglet Personnaliser dans l'éditeur de graphique a permis d'ajuster la présentation. Nous avons opté pour un graphique à colonnes en 3D, avec des couleurs vives et modifié le titre. De même, des quadrillages principaux et secondaires ont été sélectionnés pour améliorer la lisibilité. Sur l'axe vertical, nous avons placé les revenus totaux, et sur l'axe horizontal, les quartiers. Ce graphique rend les variations de revenus beaucoup plus visibles et facilite l'interprétation des données.

Image 6



Ce graphique exploratoire montre le revenu total médian en 2020 selon les quartiers d'Ottawa révélant ainsi les inégalités économiques entre eux

4. Récit potentiel

Étant donné que l'analyse des revenus médians par quartier à Ottawa révèle un écart important entre les quartiers à faible revenu, comme Rideau-Vanier ou Gloucester-Southgate, et ceux à revenu plus élevé, tels que Kitchissippi ou Orléans Sud-Navan, notre histoire cherchera à comprendre les raisons qui expliquent ces différences. Nous explorerons les facteurs socio-économiques, historiques ou structurels qui contribuent à cette inégale répartition des revenus à travers la ville.

Pour raconter cette histoire de manière complète, il sera nécessaire de rassembler d'autres données complémentaires. Par exemple, le niveau d'éducation, le taux d'emploi, le coût du logement, la proportion de nouveaux arrivants, le type de ménages ou encore l'accès aux services publics pourraient aider à expliquer pourquoi certains quartiers demeurent moins favorisés que d'autres. Ces variables permettront de mieux comprendre les contextes locaux et les dynamiques sociales qui influencent le revenu médian.

Nous enrichirons notre récit en interviewant des résidents de quartiers aux profils différents, des conseillers municipaux ainsi que des chercheurs spécialisés en urbanisme ou en géographie sociale. Leurs perspectives aideraient à illustrer les réalités vécues et à contextualiser les écarts observés dans les données.

Pour soutenir cette analyse et offrir un contexte plus large, voici quelques articles et études pertinents à notre histoire :

– Inégalités entre les quartiers des villes canadiennes – ARCHIVED

<https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2000160-fra.pdf>

– L'Étude des quartiers d'Ottawa <https://www.neighbourhoodstudy.ca/fr/>

– La hausse des prix et ses répercussions sur les plus vulnérables financièrement : un profil des personnes faisant partie du quintile inférieur de revenu familial <https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00002-fra.htm>

5. Conclusion

Ce projet nous a permis de mieux comprendre comment transformer un jeu de données brutes en une analyse claire et visuelle. Le processus de nettoyage, de création du tableau croisé dynamique et du graphique à colonnes 3D nous a aidé à voir concrètement les écarts économiques entre les quartiers d'Ottawa.

Le plus difficile fut d'apprendre à manipuler les fonctions de Google Feuilles de calcul (comme le tableau croisé dynamique etc...) tout en gardant les données exactes et bien formatées. Mais c'est aussi ce qui fut le plus gratifiant : réussir à produire une visualisation lisible et à interpréter les résultats de manière significative. Sur le plan personnel, ce travail nous a montré nos difficultés en analyse statistique et en exploration de données. Nous avons compris qu'au-delà de la technique, il faut toujours replacer les chiffres dans leur contexte social.

Si nous devions recommencer, on ajouterait des données complémentaires (ex. éducation, logement, emploi) pour élargir l'analyse et expliquer plus finement les écarts observés.

En résumé, cette expérience nous a appris à allier rigueur méthodologique et sens critique, tout en ouvrant la voie à une réflexion plus large sur les inégalités de revenu et la justice sociale à Ottawa.

6. Références

Exploring facets | OpenRefine. (2024, April 29). <https://openrefine.org/docs/manual/facets>

Marier, J.S. (2024). Fichier CSV du recensement 2021 – Données par quartier (Ottawa) [GitHub]. https://raw.githubusercontent.com/jsmarier/files-for-course-assignments/refs/heads/main/Questionnaire_d%C3%A9tail%C3%A9_du_recensement_de_2021_Donn%C3%A9es_par_quartier.csv

Statistique Canada. (2022, February 25). Exactitude et validation des données : méthodes pour assurer la qualité des données [Vidéo]. <https://www.statcan.gc.ca/fr/afc/litteratie-donnees/catalogue/892000062020008>

Ville d'Ottawa. (2023.). Questionnaire détaillé du recensement de 2021 Données par quartier. Ottawa ouverte. <https://ouverte.ottawa.ca/datasets/ottawa::questionnaire-d%C3%A9tail%C3%A9-du-recensement-de-2021-donn%C3%A9es-par-quartier/about>

Ville d'Ottawa. (n.d.). Cartes des quartiers et zones des conseils scolaires. Site Web officiel de la ville d'Ottawa. <https://ottawa.ca/fr/hotel-de-ville/elections/cartes-des-quartiers-et-zones-des-conseils-scolaires#section-9ac4699c-ed4f-4179-b386-78953e3ab0d0>

Myles, J., Picot, G., & Pyper, W. (Eds.). (2000). *Inégalités entre les quartiers des villes canadiennes - ARCHIVED*. Statistique Canada. <https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2000160-fra.pdf>

L'Étude des quartiers d'Ottawa <https://www.neighbourhoodstudy.ca/fr/>

<https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00002-eng.htm>

Voici un exemple :

Bounegru, L., & Gray, J. (Eds.). (2021). *The Data Journalism Handbook 2: Towards A Critical Data Practice*. Amsterdam University Press. https://ocul-crl.primo.exlibrisgroup.com/permalink/01OCUL_CRL/hgdufh/alma991022890087305153