

Kelly Christensen
Soline Doat
Margaux Faure

Projet Données - Les lieux de tournages à Paris

Lien vers le dataset initial : <https://www.data.gouv.fr/fr/datasets/lieux-de-tournage-a-paris-2/>

Ce dataset répertorie plus de 10 000 tournages s'étant déroulés en extérieur dans la capitale entre 2016 et 2021. Pour chaque identifiant de lieu de tournage, sont renseignés : le titre de l'œuvre, les dates de tournage, les informations de production et de l'adresse ayant accueilli les scènes filmées.

Ce dataset met donc en relation des œuvres audiovisuelles avec des lieux parisiens qui n'ont pas été choisis par hasard par la production comme emplacement de tournage pour une ou plusieurs scènes. Aussi nous a-t-il semblé intéressant de problématiser ces données autour du choix géographique effectué au sein de la capitale : les monuments parisiens attirent-ils particulièrement les productions cinématographiques ? Cette question suppose donc de lier les adresses de tournage à celles des bâtiments parisiens étant reconnus comme patrimoniaux mais également de s'intéresser à ce que peut représenter le fait de tourner à Paris. Quelles types de production viennent filmer à Paris ? Parmi elles, quelle est la part de productions françaises et de productions étrangères ? Une nationalité étrangère est-elle particulièrement présente parmi les réalisateurs ayant tourné au moins une scène à Paris ? Quelle est la durée des tournages de scène dans la capitale ? C'est l'ensemble de ces interrogations qui ont accompagné notre travail sur les données utilisées et leurs visualisations.

1. Le travail sur les données du set initial

La première étape a été de travailler sur les éléments présents dans notre dataset de départ. Il s'agissait là de nettoyer et uniformiser les données en vue de leur enrichissement avec des données extérieures. L'objectif étant de traiter uniquement l'ensemble en masse, des choix de nettoyage ont été faits dans ce sens.

- ❖ Les titres d'œuvres ont été uniformisés, tous en majuscule pour faciliter de potentielles jointures. La principale difficulté s'est portée sur le traitement des titres de série qui sont enregistrés de manière assez irrégulière, mentionnant notamment les numéros de saison ou d'épisode. Ceux-ci ont été supprimés pour uniformiser au mieux les enregistrements.
- ❖ Pour les adresses de tournage, la mention de « Paris » et des arrondissements a été supprimée dans la colonne « Localisation de la scène ». Puisque tout notre travail se concentre sur cette ville, la précision de la ville et de "75" dans l'adresse n'est pas nécessaire. La logique est la même pour les arrondissements qui disposent d'une colonne qui leur est propre. Les tournages ayant eu lieu en région parisienne n'ont pas non plus été retenus.
- ❖ Les dates de début et de fin de tournage ont été parsées afin de pouvoir calculer la durée de tournage de chaque scène avec la formule « *Compute difference between dates* » disponible sur Dataiku. Cependant, avec cette méthode, la première journée de tournage n'est pas comptabilisée puisque nous ne disposons pas des précisions des heures de tournage mais seulement des jours. Les tournages d'une seule journée donnaient donc un résultat de 0. Nous avons donc décidé d'incrémenter chaque résultat afin de comptabiliser les premières 24 heures de tournage comme un jour entier.

- ❖ Afin de rendre possible une jointure à partir des coordonnées géographiques, une colonne contenant un *GeoPoint* a été créée à partir des latitudes et des longitudes.

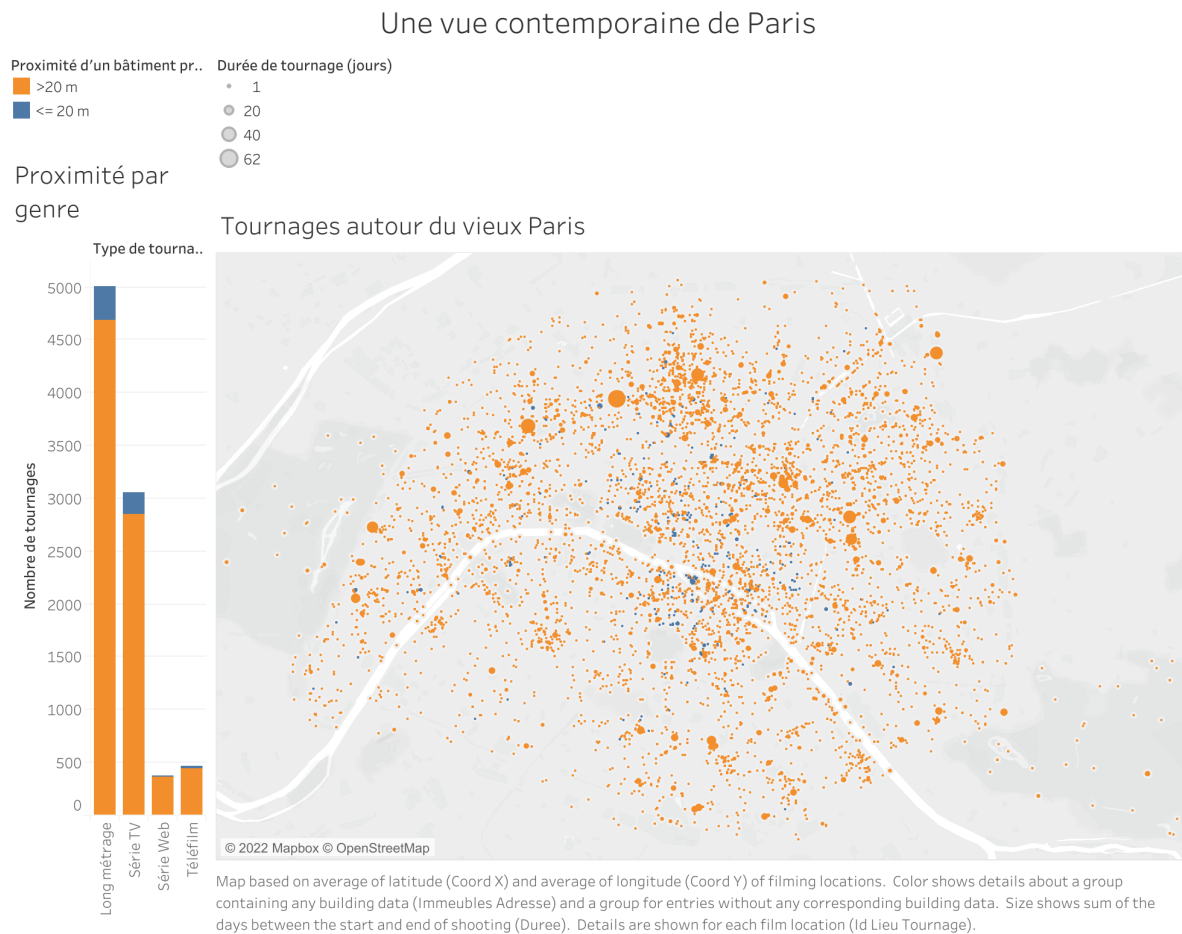
2. L'enrichissement du dataset avec d'autres données

Afin d'étudier l'axe que nous avons défini en introduction, des données extérieures viennent enrichir le dataset initial.

- ❖ Tout d'abord, nous avons récupéré un [dataset des bâtiments classés à Paris](#) afin d'étudier la proximité des tournages avec ceux-ci. Toutes les entrées dont le code départemental ne correspondait pas à 75 ont été supprimées. Des coordonnées *GeoPoint* ont été ajoutées pour faire la jointure avec le dataset des lieux de tournage. Nous avons fait une jointure gauche pour conserver les tournages qui n'ont pas été tournés près de bâtiments classés. En revanche, un nettoyage de multiples colonnes a été effectué dans le contenu du dataset sur les monuments parisiens. On garde surtout ce qui est nécessaire pour étudier la logique de localisation des tournages : le nom du bâtiment, son adresse, les coordonnées géographiques pour faire les visualisations, le siècle de construction et la clef de référence comme identifiant qui pourrait servir ultérieurement. Pour la datation du bâtiment, nous avons gardé seulement le siècle le plus récent, partant du principe que celui-ci représentait au mieux l'état actuel de la façade. La jointure se fait donc à partir des *GeoPoints* présents sur les deux *datasets*, avec une zone géographique autorisée de 20 mètres afin de mettre en relation les tournages ayant une distance raisonnable avec les monuments.
- ❖ Les nationalités des réalisateurs ont été récupérées à partir d'une requête SPARQL dans Wikidata. Nous avons récupéré l'information pour les réalisateurs nés après 1920 afin d'étudier la proportion de réalisateurs français et étrangers. Un problème s'est posé avec certains réalisateurs dont les résultats affichaient deux dates de naissance via l'utilisation d'un *group by*. Nous avons fait le choix de n'en conserver qu'une seule. Les doubles nationalités apparaissent aussi chacune sur une ligne : nous avons choisi de n'en garder qu'une en faisant un *group by* sur le réalisateur et un *count(group keys)* pour additionner. La jointure avec le dataset original a été faite sur les noms de réalisateurs.

3. La visualisation des données

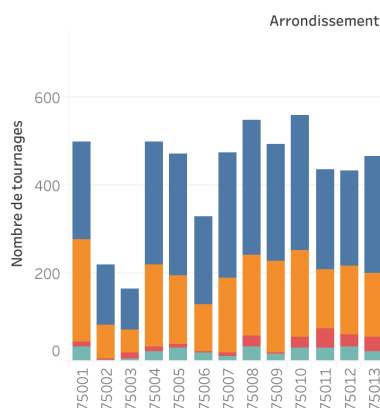
Nous avons travaillé avec Tableau public pour faire la datavisualisation. Nous avons voulu constituer des dashboards thématiques.



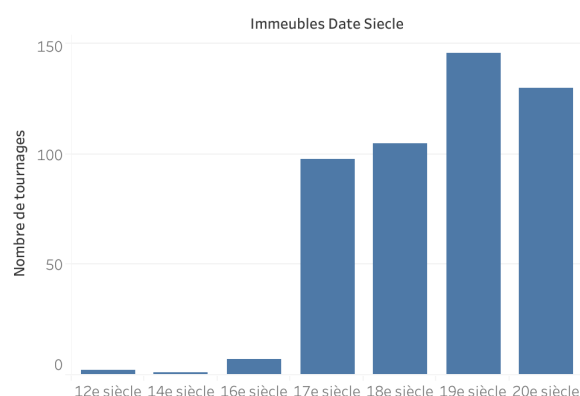
- ❖ Le premier nommé “Une vue contemporaine de Paris” insiste sur la géographie des lieux de tournage parisiens. Une carte de Paris permet de les mettre en évidence, la couleur bleue indique les tournages situés près de bâtiments historiques parisiens et la couleur orange ceux dont ce n’est pas le cas en fonction du nombre de jours de tournage. Enfin, une visualisation permet d’apprécier cette proximité en fonction du genre.

Caractéristiques des tournages à Paris

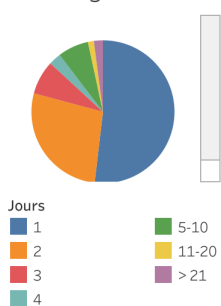
Tournages par arrondissement



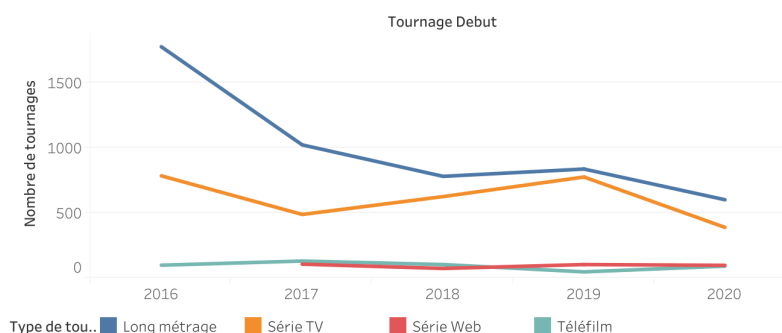
Ancienneté des bâtiments servant des lieux de tournage



Durées de tournage



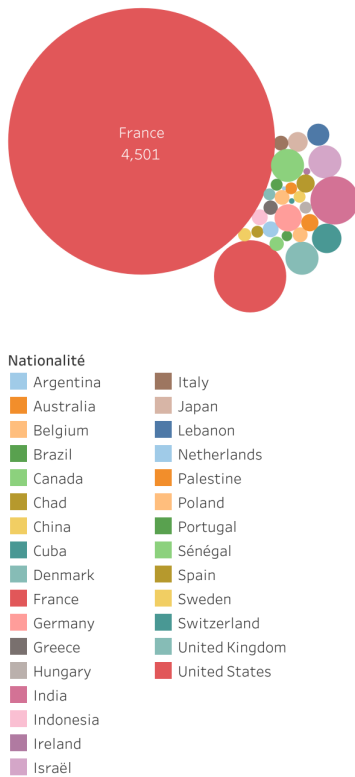
Nombre de tournages à Paris par année



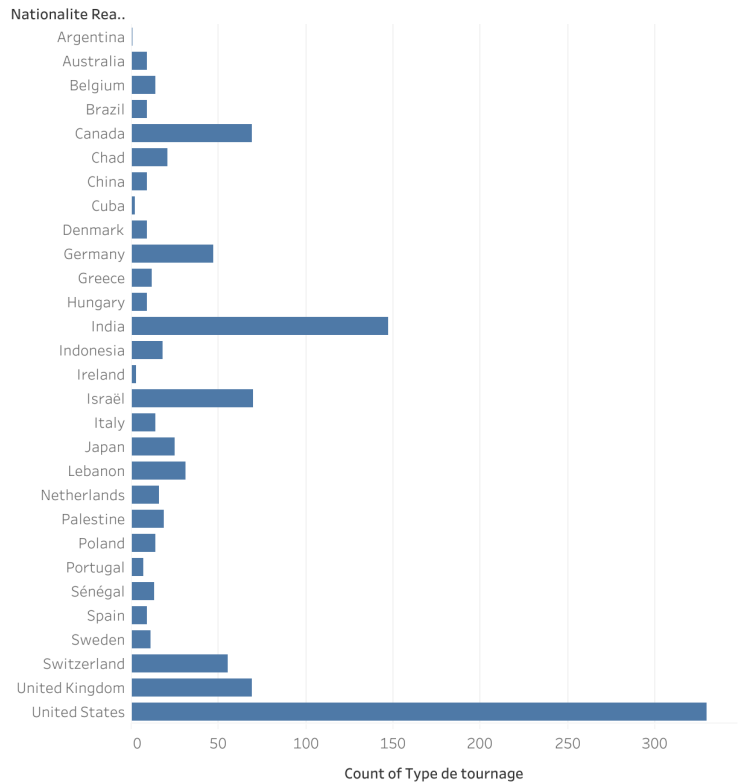
- ❖ Un deuxième dashboard intitulé “Caractéristiques des tournages à Paris” se concentre sur l’environnement de tournage de manière plus détaillée en prenant en compte le nombre de tournages réalisés par arrondissement afin d’examiner si l’un d’eux est mis en avant dans les choix de lieu de tournage. Il est également possible de constater l’ancienneté des bâtiments à proximité et servant donc potentiellement de décor. Un graphique montre la durée des tournages qui peut être comparée au nombre de tournages par année.

Nationalités des réalisateurs

Nationalités des réalisateurs tournant à Paris



Nationalités des réalisateurs étrangers



- ❖ Le dernier dashboard intitulé “Nationalités des réalisateurs” est orienté vers l’identité des productions tournant à Paris : les nationalités des producteurs sont représentées afin de prendre connaissance des pays d’origine de ceux venant tourner à Paris. Une première visualisation montre que les réalisateurs français sont largement majoritaires. Une seconde les exclut donc afin de comparer la présence des réalisateurs étrangers et met en évidence une grande majorité de réalisateurs originaires des Etats-Unis.