

Projet TER par l'équipe ARS

TER ARTIST-RUN-SPACES

KARAMI Aya

JEBALI Anas

GUYON NOAH

EL OUALYDY Mohamed-Amine

VITOFFODJI Adjimon

November 30, 2024

## Analyse des données du projet TER

Dans le cadre de notre projet, l’exploration et la préparation des données représentent une étape essentielle pour assurer la qualité des analyses prévues. Nous avons commencé par identifier les sources de données importantes, puis nous les avons nettoyées et organisées pour pouvoir les exploiter efficacement. Ce travail préliminaire a également permis de repérer les éventuelles lacunes et de structurer les informations dans un format adapté aux besoins du projet.

Les données ont été fournies par Sabrina Issa, qui recense des artist-run spaces à l’international avec le soutien de diverses institutions. Ces données incluent des informations géographiques (adresse, latitude, longitude), temporelles (dates d’ouverture/fermeture), et descriptives (activités, publications, expositions). La majorité des données nécessaires est déjà disponible. Cependant, M. Collin, notre tuteur, a indiqué qu’il nous fournirait progressivement des données supplémentaires, que nous pourrions intégrer à notre travail.

Les fichiers initiaux, majoritairement en format CSV et JSON, n’étaient pas organisés en tables relationnelles. Pour transformer ces données, nous avons commencé par charger les fichiers sources dans des DataFrames avec pandas pour une exploration initiale. Les colonnes contenant des valeurs manquantes ont été nettoyées : les données numériques ont été remplacées par 0, et les valeurs textuelles manquantes par "Information Manquante". Les doublons ont été supprimés, et les balises HTML résiduelles ont été retirées à l’aide de BeautifulSoup.

Les données ont ensuite été réorganisées dans une base relationnelle SQLite, avec des tables dédiées aux espaces (**Espace**), aux questions/réponses (**Question.Reponse**), et aux responsables (**Responsable**), interconnectées via une table de liaison (**tenir**).

## **0.1 Exploration descriptive des données**

### *0.1.1 Analyse univariée*

Les premières explorations incluent des analyses univariées simples, accompagnées de visualisations, notamment un nuage de mots basé sur les descriptions des artist-run spaces.

### *0.1.2 Analyse bivariable*

Les premières explorations incluent la corrélation entre la localisation géographique et le type d'activités proposées, ainsi que l'influence des dates d'ouverture et de fermeture sur la durée de vie des espaces.

## **0.2 Analyses avancées et visualisations**

### *0.2.1 Carte interactive*

Une carte interactive a été créée avec Folium, affichant la localisation géographique des artist-run spaces validés. Cette visualisation met en lumière leur répartition mondiale.

### *0.2.2 Tendances temporelles*

Un graphique interactif a révélé une augmentation des ouvertures d'espaces au cours des dernières décennies.

## **0.3 Conclusion**

Cette étape a permis de poser les bases nécessaires pour une analyse approfondie. Les données identifiées sont pertinentes et ont été nettoyées pour garantir leur exploitabilité. Bien que des informations supplémentaires, notamment sur les horaires et les responsables des espaces, puissent enrichir les analyses, l'organisation en tables relationnelles adaptées à SQLite est désormais finalisée.