1. **Project Context and Aims**

Cultural capital of a city is largely moulded by the artistic productions of its inhabitants. Spatiotemporal of artwork provides a powerful medium to feasibly understand the social and cultural developments of a certain time and place.

An overarching objective was to present an exploratory tool that enables the user to navigate various typologies of museum data on a map, providing visualisations that tell a story about the objects of interest.

1. **Project Rationale**
2. **Data Collection**

The Metropolitan Museum (MET) published a csv of **n** artworks from their collection on their GitHub (insert hyperlink to MET github). Table 1 provides an overview of columns in the raw dataset, column data, data type and number of missing values.

1. **Data Handling**
2. **Data Cleaning**
   1. Country Data

The spatial element of each artwork was the country, which contained a significant amount of inconsistencies such as undesired strings (e.g. probably China) and characters (e.g. Ghana?). Employing the FlashText python library (insert link), keyword extraction was implemented on the dataset to obtain a clean set of country names. An independent world countries dataset (insert hyperlink) was served as the set of keywords. For various unique situations, manual cleaning of the data was needed.

* 1. Classification

General artwork classifications were needed to easily navigate the dataset.

* 1. Medium Data

To prepare medium data for analysis, materials and techniques needed to be automatically identified to efficiently extract the words required. To illustrate, a medium description can be “Blown and etched glass”. Materials would involve nouns (“glass”), while techniques would involve verbs (“blown, “etched”). Point of speech (POS) tagging was implemented using TextBlob, a Natural Language Processing (NLP) library, to extract relevant materials and techniques for the clustering analysis. A python function was built to vectorize the dataset according to materials or techniques used in the artwork. This was simply a dataframe with medium items a columns, a value of 1 indicated that the medium item was used in the artwork and 0 indicated the contrary.

**5.3** Donation Data

Donator and year of donation were given in the credit line column. POS tagging aided in the extraction of donation year from the string values.

|  |  |  |
| --- | --- | --- |
| Column Name | Column Data | Type |
| Object Number | Unique ID | Int |
| Is Highlight |  | Boolean |
| Is Public Domain |  | Boolean |
| Department |  | String |
| Object Name | Name describing the object | String |
| Title | Title of the object | String |
| Culture |  | String |
| Period |  | String |
| Dynasty |  | String |
| Reign |  | String |
| Portfolio |  | String |
| Artist Role |  | String |
| Artist Prefix |  | String |
| Artist Display Name |  | String |
| Artist Display Bio |  | String |
| Artist Suffix |  | String |
| Artist Alpha Sort |  | String |
| Artist Nationality |  | String |
| Artist Begin Date |  | Int |
| Artist End Date |  | Int |
| Object Date |  | Int |
| Object Begin Date |  | Int |
| Object End Date |  | Int |
| Medium | Materials and techniques of the object |  |
| Dimensions |  |  |
| Credit Line |  |  |
| Geography Type |  |  |
| City |  |  |
| State |  |  |
| County |  |  |
| Country |  |  |
| Region |  |  |
| Subregion |  |  |
| Locale |  |  |
| Locus |  |  |
| Excavation |  |  |
| River |  |  |
| Classification |  |  |
| Rights and Reproduction |  |  |
| Link Resource |  |  |
| Metadata Date |  |  |
| Repository |  |  |
| Tags |  |  |

1. **Data analysis methods and major findings**
2. **Data Visualization**
3. **Technical integration between elements**