COMP4447 Final Project

Peter Strimbu

May 16, 2022

# Metropolitan Museum of Art Data Analysis

# Columns

|  |
| --- |
| 1. Is Highlight |
| 1. Object ID |
| 1. AccessionYear |
| 1. Culture |
| 1. Reign |
| 1. Artist Role |
| 1. Artist Display Bio |
| 1. Artist Nationality |
| 1. Artist Gender |
| 1. Object Date |
| 1. Medium |
| 1. Geography Type |
| 1. County |
| 1. Subregion |
| 1. Excavation |
| 1. Rights and Reproduction |
| 1. Metadata Date |
| 1. Tags AAT URL |

# Column Types / Range / Cleaning Tasks

## Object ID

* int64
* 1 through 860873
* No cleanup, ok as-is

## AccessionYear

* 4 digit year
* Ex: nan, 1889.0, 1928, 19171917.0, 1956-08-24
* Clean: keep only first 4 digits will fix all 4 issues

## Culture

* category
* Ex: nan, 'reign of Amenhotep III', 'reign of Xerxes I'
* remove duplicate spaces
* possibly remove 'reign of', 'reigns of', 'or later', 'possibly', 'or slightly later', ',early', ',probably', (anything after comma), (anything in parens), (question marks), (space at end), (split on slash,dash/'and'/'or' and process individually), 'to xyz' and group by remaining names

## Artist Role

* object
* Ex: 'nan', 'Maker', 'Designer|Manufacturer'
* Clean: split on vert bar?

## Artist Display Bio

* object
* Ex: '1785–1844', 'nan', 'British, London 1873–1952 Hailsham, Sussex|British, Wiltshire 1877–1952 Oxford', '||||Female|Female'
* Issues: not sure what the vert bars separate
* None - don't use or use as-is

## Object Date

* Numeric
* Ex: '1853', '1901', '1909–27', '1782-1784', 'December 1, 1925', 'after 1773'
* Medium
* object
* Ex: 'Gold' 'Silver' 'Bronze or copper' ... 'Overlay for 23.112.2889, graphite and ink on glazed linen tracing paper'
* Don't use or use as-is

# AccessionYear

Text

Description automatically generated

# Culture

Graphical user interface, text

Description automatically generated

Text

Description automatically generated

### MASK:

Logo

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated