Artwork Master Database

The following document describes the contents of the Artwork Database and the process through which the Database was compiled.

The data was collected for the research of Renée B. Adams at the Saïd Business School, University of Oxford. Professor Adams’ research focuses on gender and its’ effects on the pricing of art on the primary markets.

<https://www.renee-adams.com>

<https://www.sbs.ox.ac.uk/about-us/people/renee-b-adams>

Table of contents

1. General description of the database
2. Process of data collection
3. Variable descriptions
4. Codes

1) General description of the database

The database contains artwork price data collected from the primary art market.

Due to Covid-19 most fairs during 2020 and 2021 were held online, which made the digital data collection possible. The majority of artworks in the database originate from [artsy.net](https://www.artsy.net), who have been hosting online art fairs since 2013. We have been actively collecting data since July 2020, however we also obtained data from fairs that took place earlier, thanks to artsy maintaining an archive.

We compiled a large database of artists and their gender from various sources, the primary being [artfacts.net](https://artfacts.net/). They provide a large range of data and detailed descriptions of artists and their works. Our primary interest was identifying the gender of the individual, however Artfacts provided us with a much wider range of variables, that can be utilised in the models.

Most of the data collection was automated via Python and falls under the U.K. Government’s “Exemptions to copyright” guideline as non-commercial research. [(for further reference)](https://www.gov.uk/guidance/exceptions-to-copyright%23content)

The current database contains 484,296 observations of which 189,550 (39.14%) are unique artworks and 179,052 (36.97%) are unique artworks from fairs. The duplicates are either due to us recording some fairs on multiple occasions or some artworks appearing at different fairs. We use the URL of an artwork to identify it as unique. We also assigned each observation a unique ID with which we can link it to the image of the artwork.

2) Process of data collection

3) Variable descriptions

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Variable Type** | **Description** |
| artwork\_url\_id | Float | *group(url)*  Consistent throughout fairs and saves. |
| image\_download\_save\_id | Float | ID assigned when downloading images. No consistency throughout fairs or saves |
| image\_id | String | ID to match artwork 1:1 to the downloaded image based on its filename. |
| is\_fair | Dummy/Byte | Marks whether the observation was collected from an art fair.  0: non-fair data 1: fair data |
| artsy\_data | Dummy/Byte | Marks whether the observation was collected from artsy.  0: non-artsy data 1: artsy data |
| artfacts\_data\_available | Dummy/Byte | Marks whether the observation’s artist could be matched to the artfacts artist database.  0: no match 1: match |
| title | String | The title of the artwork. |
| year\_completed | Integer | The year when the artwork was completed. |
| years | Integer | The number of years taken to create the artwork. |
| bc\_dummy | Dummy/Byte | Marks whether the years are BC.  0: AD 1: BC |
| price | Float | *(low\_price\_USD + high\_price\_USD) / 2*  The mean of low & high end range converted to USD. |
| ln\_price | Float | *log(price)*  The natural log of price. Useful for analysis since prices are log-normally distributed. |
| low\_price\_USD | Float | *low\_price \* exchange\_rate*  The lower end of the price range converted to USD. |
| high\_price\_USD | Float | *high\_price \* exchange\_rate*  The higher end of the price range converted to USD. |
| low\_price | Float | The lower end of the price range specified in the original currency. |
| high\_price | Float | The higher end of the price range specified in the original currency. |

|  |  |  |
| --- | --- | --- |
| currency | String | The 3 letter abbreviation of the currency in which the artwork was originally advertised. |
| category | Integer | Categorical variable. |
| category\_inferred | Integer | The same variable as category, with additional observations’ classification inferred based on common materials in each group. |
| materials | String | The materials used in the creation of the artwork. |
| length\_in | Float | The length of the artwork in inches. |
| width\_in | Float | The width of the artwork in inches. |
| depth\_in | Float | The depth of the artwork in inches. |
| sold | Dummy/Byte | Marks whether the artwork was sold or not.  0: not sold 1:sold |
| sold\_inferred | Dummy/Byte | The same variable as sold, with additional observations’ classification inferred based on the sold status of other artworks within the fair. |
| collection\_date | Date (Integer) | The date the artwork’s data was collected in Stata date format. |
| collection\_times | Integer | The number of times an artwork appears in the database. |
| collection\_date1 | Date (Integer) |  |
| collection\_date2 | Date (Integer) |  |
| collection\_date1\_fair | Integer |  |
| collection\_date2\_fair | Integer |  |
| sold\_date | Date (Integer) | The date the artwork was sold. |
| exit\_date | Integer |  |
| gallery\_id | Integer | *group(gallery)*  The unique identifier of a gallery based on its name. |
| gallery | String | The name of the gallery representing the artist. |
| gallery\_slug | String | The artsy slug of the gallery. |
| fair\_id | String | The ID of fair. |
| fair\_name | String | The name of the fair. |
| fair\_image\_id | String | A combination of the fair\_id and fair\_year variables such that fairs in different years may be distinguished for the purpose of downloading images. |
| fair\_year | Integer | The year of the fair. |
| number\_of\_artworks | Long | The number of artworks that appeared on the fair. |
| fair\_start\_date | Date (Integer) | The date the fair started in Stata format. |
| fair\_end\_date | Date (Integer) | The date the fair ended in Stata format (if applicable). |
| artist\_id | Integer | *group(artist)*  Unique identifier of an artist based on their name. Consistent throughout fairs. |
| artist | String | The name of the artist |
| birth\_year | Date (Integer) | The birth year of the artist based on various sources. |
| death\_year | Date (Integer) | The death date of the artist based on various sources (if applicable). |
| artist\_slug |  |  |
| acclaim |  |  |
| gender |  |  |
| gender\_source |  |  |
| artist\_g7 |  |  |
| artist\_advanced\_econ |  |  |
| artist\_eu |  |  |
| artist\_first\_nationality |  |  |
| artist\_secondary\_nationality |  |  |
| artist\_dual\_nationality |  |  |
| inconsistency |  |  |
| manual\_correction |  |  |
| artist\_data\_available |  |  |
| aF\_gender |  |  |
| au\_gender\_MANUAL |  |  |
| au\_gender |  |  |
| bl\_gender |  |  |
| ar\_gender |  |  |
| nn\_gender |  |  |
| exact\_match\_gender |  |  |
| aF\_birth\_date |  |  |
| aF\_birth\_year |  |  |
| aF\_death\_date |  |  |
| aF\_death\_year |  |  |
| aF\_birth\_location |  |  |
| aF\_nationality |  |  |
| aF\_countries |  |  |
| aF\_sector |  |  |
| aF\_ranking |  |  |
| aF\_exhibitions\_in\_country\_1 |  |  |
| aF\_exhibitions\_in\_count\_1 |  |  |
| aF\_exhibitions\_in\_country\_2 |  |  |
| aF\_exhibitions\_in\_count\_2 |  |  |
| aF\_exhibitions\_in\_country\_3 |  |  |
| aF\_exhibitions\_in\_count\_3 |  |  |
| aF\_exhibitions\_at\_institution\_1 |  |  |
| aF\_institution\_id\_1 |  |  |
| aF\_exhibitions\_at\_institution\_2 |  |  |
| aF\_institution\_id\_2 |  |  |
| aF\_exhibitions\_at\_institution\_3 |  |  |
| aF\_institution\_id\_3 |  |  |
| aF\_national\_ranking |  |  |
| aF\_exhibitions\_total |  |  |
| au\_derived\_nationality |  |  |
| au\_artist\_birth\_place |  |  |
| au\_artist\_year\_birth |  |  |
| au\_deceased |  |  |
| bl\_nationality |  |  |
| bl\_yobirth |  |  |
| bl\_yodeath |  |  |
| birth\_error |  |  |
| death\_error |  |  |
| price\_unformatted |  |  |
| fair |  |  |
| filename |  |  |
| year |  |  |
| start\_year |  |  |
| end\_year |  |  |
| century\_dummy |  |  |
| cent\_early |  |  |
| cent\_late |  |  |
| ln\_low\_price\_USD |  |  |
| ln\_high\_price\_USD |  |  |
| datecur |  |  |
| exchange\_rate |  |  |
| dimensions |  |  |
| length |  |  |
| width |  |  |
| depth |  |  |
| measurement |  |  |
| length\_cm |  |  |
| width\_cm |  |  |
| depth\_cm |  |  |
| url |  |  |
| image\_url |  |  |
| collection\_url |  |  |
| sold\_unformatted |  |  |
| collectiondate\_unformatted |  |  |
| details |  |  |
| description |  |  |
| \_dupe |  |  |
| fair\_date |  |  |
| fair\_artsy\_url |  |  |
| aF\_id |  |  |
| aF\_first\_name |  |  |
| aF\_last\_name |  |  |
| aF\_aliases |  |  |
| aF\_birth\_date\_unformatted |  |  |
| aF\_death\_date\_unformatted |  |  |
| aF\_death\_location |  |  |
| aF\_movements |  |  |
| aF\_media |  |  |
| aF\_link\_artfacts |  |  |
| aF\_link\_personal |  |  |
| aF\_links\_wiki |  |  |
| aF\_ranking\_trend |  |  |
| aF\_cities |  |  |
| aF\_artists |  |  |
| aF\_groups |  |  |
| aF\_institution\_1\_link\_artfacts |  |  |
| aF\_institution\_1\_link |  |  |
| aF\_institution\_1\_link\_facebook |  |  |
| aF\_institution\_1\_link\_twitter |  |  |
| aF\_institution\_1\_link\_wiki |  |  |
| aF\_institution\_2\_link\_artfacts |  |  |
| aF\_institution\_2\_link |  |  |
| aF\_institution\_2\_link\_facebook |  |  |
| aF\_institution\_2\_link\_twitter |  |  |
| aF\_institution\_2\_link\_wiki |  |  |
| aF\_institution\_3\_link\_artfacts |  |  |
| aF\_institution\_3\_link |  |  |
| aF\_institution\_3\_link\_facebook |  |  |
| aF\_institution\_3\_link\_twitter |  |  |
| aF\_institution\_3\_link\_wiki |  |  |
| aF\_description |  |  |
| aF\_exhibitions\_solo |  |  |
| aF\_exhibitions\_group |  |  |
| aF\_exhibitions\_artfair |  |  |
| aF\_exhibitions\_collective |  |  |
| aF\_exhibitions\_current |  |  |
| aF\_exhibitions\_biennial |  |  |
| exhibitions\_region\_eastafrica |  |  |
| exhibitions\_region\_middleafrica |  |  |
| exhibitions\_region\_northafrica |  |  |
| exhibitions\_region\_southafrica |  |  |
| exhibitions\_region\_westafrica |  |  |
| exhibitions\_region\_caribbean |  |  |
| exhibitions\_region\_centralameric |  |  |
| exhibitions\_region\_southamerica |  |  |
| exhibitions\_region\_northamerica |  |  |
| exhibitions\_region\_centralasia |  |  |
| exhibitions\_region\_eastasia |  |  |
| exhibitions\_region\_southasia |  |  |
| exhibitions\_region\_southeastasia |  |  |
| exhibitions\_region\_westasia |  |  |
| exhibitions\_region\_easteurope |  |  |
| exhibitions\_region\_northeurope |  |  |
| exhibitions\_region\_southeurope |  |  |
| exhibitions\_region\_westeurope |  |  |
| exhibitions\_region\_australianewz |  |  |
| exhibitions\_region\_melanesia |  |  |
| exhibitions\_region\_polynesia |  |  |
| exhibitions\_region\_micronesia |  |  |
| aF\_exhebitions\_museum |  |  |
| aF\_exhebitions\_gallery |  |  |
| aF\_exhebitions\_artfair |  |  |
| aF\_exhebitions\_biennial |  |  |
| aF\_exhebitions\_other |  |  |
| aF\_catalog |  |  |
| aF\_dealer |  |  |
| aF\_collection |  |  |
| ar\_url |  |  |
| ar\_artsy\_prediction\_dummy |  |  |
| ar\_description |  |  |
| ar\_artsy\_female\_counter |  |  |
| ar\_artsy\_male\_counter |  |  |
| ar\_artsy\_prediction |  |  |
| au\_artist\_surname |  |  |
| au\_artist\_forename |  |  |
| au\_example\_artwork\_title |  |  |
| au\_medium |  |  |
| au\_nationality\_code |  |  |
| au\_artist\_id |  |  |
| au\_artwork\_id |  |  |
| au\_logsurface |  |  |
| au\_marked |  |  |
| au\_style |  |  |
| bl\_malepronouns |  |  |
| bl\_femalepronouns |  |  |
| bl\_predictionaccuracy |  |  |
| bl\_url |  |  |
| exact\_match |  |  |
| nn\_valid\_input |  |  |
| nn\_value |  |  |
| nn\_accuracy |  |  |
| sector |  |  |
| category\_missing |  |  |
| fair\_group |  |  |