

Project Report

Researching & Quantifying the relationship between Modern art & historical art by the prices and their relationships with different factors such as movements, periods, artists, conditions and ages of the art.

Introduction

Project Overview

This project explores a dataset of artworks, focusing on analysing trends, pricing, and movements within the art market. The dataset was sourced from Kaggle-<https://www.kaggle.com/datasets/flkuhm/art-price-dataset/data>, and it contains details such as artist names, the year that the artwork was created, the condition of the artwork whether it was signed or not or initialised, and the prices it was sold for. The goal of my project was to clean, analyse, and visualise the data to extract insights based on mainly the prices of the art work and the time periods of the pieces. This Dataset could also help me to understand what artist or styles have lost or gained value over time. That would be beneficial in helping private galleries or even collectors understand what may be their best market.

Objectives

- Clean and preprocess the dataset for consistency and accuracy.
- Explore and visualise trends in the dataset, such as price distribution, popularity of the artists, and art movement influences.
- Although this is to find a representation of the artwork sold globally, its also to gain insights in how Sothebys has conducted their sales and their influences.
- Provide and demonstrate an analysis using Python coding.

Methodology

Data Sources

The dataset that I will be using is the 'Art price Dataset' found on Kaggle.com made by FI Kuhm. My source, the link to this Dataset is: <https://www.kaggle.com/datasets/flkuhm/art-price-dataset/data>

It was made in 03/19/2022 @misc{fl_kuhm_2022, title={Art Price Dataset}, url={<https://www.kaggle.com/ds/2558650>}, DOI={10.34740/KAGGLE/DS/2558650}, publisher={Kaggle}, author={Fl.Kuhm}, year={2022}}

How the data was collected and it's limitations/constraints

There is limited information on this dataset, it was made using data from the Sotheby's, art auction website: <https://www.sothebys.com/en/> by webscraping. The data being taken in 2022 could mean that the data is not an accurate representation currently. ### The number of records: 754 entries ##### The number of features: 9, they include; unnamed(0) price, artist, title, yearCreation, signed, condition, period and movement The dataset used in this project is: original (oldartDataset.csv) and recent one artDataset2.csv, which contains information about various artworks sold by the Sothebys art gallery. The data has columns such as: * Title: The name of the artwork. * Artist: The artist of the artwork. * Year created: The year in which the artwork was made. * Time period: Post war, modern, contemporary art, etc. * Price: The selling price of the artwork. * Art movement: The artistic movement associated with the work. * Condition: The state of preservation of the artwork. * Signed: Has the artwork been signed or not?

- Columns that I have added, include: Age of artwork(years) and I replaced Price with Price(USD)

Data Cleaning & Processing

s ### Handling missing values: Replaced empty fields and NaN/nan values with pd.NA, then dropped rows with critical missing data or replacing ti with 'unknown' to keep the data more consistent by keeping rows that I can afford to keep. ### Reformatted columns such as year created and Price(USD) to ensure they contain only numeric values, This way they could be sorted easily, such as ascending the numbers ### Conversion of data types: Transformed price (USD) into a integer for analysis, such as above. ### I've also created new columns such as age of artwork (years) to help in trend analysis. I've done reaserch here where i can use the age of the pieces to determine any relationships between the ages and prices.

Exploring my data analysis

My explorations and questions

- Price analysis: What's the average price of artworks by different artists, periods, or movements?
- Trends over time: How has the value of artworks changed over different periods or years?
- Movement popularity: Which artistic movements are most represented in the dataset?
- Artist comparisons: Who are the most expensive or prolific artists in the collection?
- Time periods: How many artworks fall into each artistic period, and how does that correlate with price?

Signed vs. unsigned: Do signed works command a premium?

Key Findings

- The distribution of artwork prices varies significantly across different periods and movements.
- Certain artists have a higher frequency of sales and higher average prices, these artists include, cell 39:
 - Russell Young - \$226.600
 - Alex Katz-\$187.000
 - Ed Ruscha-\$154.500
 - Cindy Sherman-\$141.000
 - Vija Celmins - \$135.000
 - Frank Schroeder - \$120.000
 - Gustave Blache III - \$93.000
 - Biff Elrod - \$87.000
 - Jasper Johns - \$75.000
 - Larry Bell - \$71.500
- Then I evaluated the movements that the top 10 artists are part of based on the pieces sold, cell 40: The top 3 artists are in Pop art, from this I drew the conclusion that pop art would sell better than other movements if sold by an gallery.
- In the next cell, I tried to find the percentages of the artists so that I can then put them into a pie chart, cell 41 & cell 43. I decided to find the top five movements and put that in a pie chart instead to see how this data would translate in the entire dataset. Turns out the most pieces sold by thr Top 10 artists were in Abstract pieces at 29%, followed by Realism and Expressionism at 18.5% at 24.5% found in cell 42.
- In cell 44, I was l looking at the top 10 artist by sales value in comparison to the top 10 most prolific artists by the number of artworks they have had represented by the Sothebys and sold.
- From cell 44 I have gathered that Russel Young is the top artists in both followed by Alex Katz and Ed Rucha, reinforcing that the top art movement sold or demanded is Pop art. Some artists have a higher total revenue despite selling fewer pieces, indicating a strong demand for their work.
- The condition of the artwork influences pricing trends.

- Older artworks generally fetch higher prices, though exceptions exist based on artist reputation and rarity.

- In cell 45, seen in a bar chart, I got the data that the number of artwork and their average price in each art period going from high to low is

Time period

- 19th Century \$9800.00
- Contemporary \$9062.82
- Modern \$3576.92
- Post-War \$3470.37
- unknown \$1500.00
- Based on average price by art movement Photorealism is of the highest value, followed by Baroque then Abstract expressionism.
- Based on Average artwork price by time period - by calculating the mean value of the entire dataset it happens to be that the 19th century sold for the most dollars, then contemporary, followed by modern artwork then Post-War and the unknown time pieces last.

Lastly, building up towards the end of my artwork, the final analysis, the average artwork price by time period over time stacked chart provides great insights into market trends and the financial evolution of different art periods. My key takeaways are:

Price Trends Over Time: The Post-War period consistently sees higher average prices, likely due to the demand for well-known artists from that period.

<https://www.widewalls.ch/magazine/famous-paintings-post-war-art>, such as Andy Warhol, David Hockney and Jackson Pollock

Contemporary art has a fluctuating market, sometimes surpassing older periods.

Modern art prices have shown steady growth, reinforcing its status as a foundational period for art collectors.

Visualisations

- Price trends over time: A line chart illustrating the fluctuation in average prices of artworks per decade.
- Artist popularity: A bar chart showing the top 10 artists based on total sales revenue.
- Art movements: A pie chart and bar chart depicting the distribution of artworks across different movements.
- Artwork age vs. Price: A scatter plot examining the correlation between an artwork's age and its price
- Artwork Prices Over Time by Period bar chart
- Interactive Price Analysis by Art Movement using the `ipywidgets.interact()` so that I could use a dropdown menu
- Average Artwork Price by Time Period in a bar chart comparison of Expensive vs. Prolific Artists
- Top 10 Artists by Total Sales Revenue using a horizontal bar chart using two bar charts side by side: One showing the top 10 artists by total sales revenue. Another showing the top 10 artists by the number of artworks sold.
- Most Popular Art Periods bar chart

Challenges & Learnings

Challenges Faced * Dealing with inconsistent data formats, especially in the year created and signed column, condition column. * Handling missing and unknown values in crucial columns. * Ensuring the dataset was clean enough for meaningful visualisation and analysis.

Lessons Learned

- Data preprocessing is crucial to ensure reliable results in analysis.
- Using pandas and seaborn effectively helped me in extracting and presenting insights.
- Structuring a project with reproducibility in mind makes future modifications easier.
- Not to waste time of creating structures that I won't use

Ethical Considerations

Biases & Limitations * Selection Bias: The dataset may not be fully representative of the global art market or the most up to date due to the last time it was edited. * Missing Information: Some artworks lacked complete data, affecting overall analysis. * Potential Misinterpretations: Price variations may be influenced by external factors like auction house prestige and collector demand.

Conclusion

Conclusion This project cleaned, analysed, and visualised key trends in the art market. By identifying patterns in pricing and artist popularity, it offers valuable insights into how different factors impact the sale value of artworks across a range of artwork. **Artwork prices go up and down over time**

- Prices don't just increase or decrease steadily; some years see big jumps.
- This could be because of famous artists, economic changes, or rare expensive pieces being sold.

Older art is often more expensive

- Some historical time periods have higher average prices than others.
- However, some modern artworks still sell for high prices.

Art movements affect prices

- Some art movements have higher prices, often because prolific artists are part of them.
- Modern art can sometimes be as valuable as older art, depending on the movement and artist.

Older art doesn't always mean higher prices

- While old art is usually worth more, some new pieces still sell for a lot.
- The artist, the style, and the condition of the artwork all affect the price.

- **Old artworks are often more valuable but new art can also be expensive.**
- **Certain historical periods and art movements are worth more.**
- **Artwork prices are unpredictable and change based on many factors, like time periods, art movements, artists and age.**

What I would do different

- Expanding the dataset with additional sources for a broader analysis.
- Implementing machine learning models to predict artwork prices based on historical data & exploring external factors such as auction houses and geographic trends to refine insights. I wish that I would've used my time more wisely by not creating random data frames that I may not need for my research and instead use an API such as Artprice API, Artnet API or Christie's API, which is similar to Sotheby's to determine market influences like:
 - Economic factors such as recessions and booms and looking at how they influence fluctuations in prices, especially in the Contemporary and Post-War categories.
 - Auctions and gallery sales trends indicate a shift toward newer art forms to see if they command premium prices.
 - It would have also been useful to find Predictive Insights to research whether those trends continue.
 - Analysing contemporary art on if it could overtake Post-War in average prices due to increased global interest in digital art, NFTs, and emerging artists.
 - Art periods with fewer available works, some of which I have labeled as unknown as well might see an appreciation in value over time due to scarcity.

Sources

- Class Exercises & Homework: [https://github.com/farencia/Farencia-DS-final-project/blob/main/edited_story.txt] [<https://github.com/farencia/Farencia-DS-final-project/blob/main/newfile.txt>] [https://github.com/farencia/Farencia-DS-final-project/blob/main/session3_data.ipynb] [<https://github.com/farencia/Farencia-DS-final-project/blob/main/mostpopularwords.txt>] [<https://github.com/farencia/Farencia-DS-final-project/blob/main/farencia>] [<https://github.com/farencia/Farencia-DS-final-project/blob/main/modulo.py>] [<https://github.com/farencia/Farencia-DS-final-project/blob/main/popularwords.py>]

Mention any sources, collaborators, or tools that were helpful.

[https://github.com/peckham-daz/24-intro-to-data-science/tree/main/2025_01]
[https://matplotlib.org/stable/gallery/statistics/boxplot_color.html#sphx-glr-gallery-statistics-boxplot-color-py] [https://seaborn.pydata.org/tutorial/color_palettes.html]
[https://proclusacademy.com/blog/customize_matplotlib_piechart/]
[<https://www.kaggle.com/datasets/flkuhm/art-price-dataset/data>]
[<https://www.statology.org/pandas-remove-characters-from-string/>]

[<https://stackoverflow.com/questions/17468878/pandas-python-how-to-count-the-number-of-records-or-rows-in-a-dataframe>] [<https://www.geeksforgeeks.org/python-pandas-dataframe-replace/>] [<https://saturncloud.io/blog/how-to-update-a-pandas-dataframe-row-with-new-values/#:~:text=This%20can%20be%20done%20using,the%20column%20names%20as%20keys>] [https://www.w3schools.com/python/pandas/trypandas.asp?filename=demo_pandas_cleaning_dropna] [<https://matplotlib.org/ipyml/>] [<https://chatgpt.com>] [<https://www.geeksforgeeks.org/interactive-graphs-in-jupyter-notebook/>] [https://www.w3schools.com/python/pandas/ref_df_nunique.asp]https://proclusacademy.com/blog/customize_matplotlib] [https://seaborn.pydata.org/tutorial/color_palettes.html] [https://matplotlib.org/stable/gallery/color/named_colors.html] [https://matplotlib.org/stable/gallery/statistics/boxplot_color.html#sphx-glr-gallery-statistics-boxplot-color-py] [https://matplotlib.org/stable/gallery/lines_bars_and_markers/stackplot_demo.html#sphx-glr-gallery-lines-bars-and-markers-stackplot-demo-py] [<https://www.widewalls.ch/magazine/famous-paintings-post-war-art>] I also got help from my colleague, a python coding tutor, Ben

Submission & Repository

File Format: This report is available in .pdf, .docx, .odt, and .md formats.

GitHub Repository: [<https://github.com/farencia/Farencia-DS-final-project>]