Carved Into the City: Melbourne’s Story Through Public Art

Introduction

Melbourne has more than 202 public art installations in 37 square kilometres. You don’t need to visit galleries or museums to witness art in Melbourne, you’re likely to trip over the whimsical *Public Purse* outside the GPO. They’re more than ornamentation, these are permanent witnesses to shifting values, civic ambitions, and the quiet transformations that shape the city’s inherently artistic soul.

This project aims to explore the evolution of thematic content in Melbourne’s public art by applying topic modelling techniques to the “Outdoor Artworks” dataset published by the City of Melbourne (City of Melbourne 2024). The dataset includes records of permanent public artworks with fields such as title, artist, installation year, location and artwork description. Our primary focus is on the "description" field, which offers unstructured textual content that can be computationally analysed for recurring themes.

By examining how the themes of public art change over time, this project seeks to uncover cultural patterns, shifts in public sentiment, and possibly trace the influence of socio-political events or urban development initiatives on artistic expression (Roose, Roose, and Daenekindt 2018). This project focuses on three major questions:

- What are the dominant themes expressed in Melbourne's public art?

- How have these themes changed across decades?
- And are certain themes concentrated in specific neighbourhoods or periods?

To answer these questions, this project will apply natural language processing (NLP) tools and machine learning-based topic modelling. We selected the **Natural Language Toolkit (NLTK)** and **Scikit-Learn** for our analysis due to their capabilities in handling text data and performing topic modelling, respectively (Bird, Klein, and Loper 2009; Pedregosa et al. 2011).

Natural Language Toolkit (NLTK)

NLTK is a powerful library for pre-processing textual data before further analysis, by enabling methods like tokenisation, stop word removal and lemmatisation. These are foundational steps before moving to further modelling or analysis. In this project, NLTK would be used to preprocess the “description” field of each public artwork, which contains unstructured human-focused text, varying in tone, length and vocabulary. Basic NLTK preprocessing is essential for reducing noise such that modelling algorithms can detect semantic meaning, rather than nuances of the language. NLTK alone would not be sufficient for identifying themes, but it can open doors to basic exploratory steps like word clouds or common terms across periods.

Scikit-Learn

Scikit-Learn is a widely used library for Machine Learning and provides an implementation of Latent Dirichlet Allocation (LDA) (Scikit-learn 2020). LDA is an unsupervised probabilistic model that represents documents as mixtures of topics - a topic being a distribution of words (Blei, Ng, and Jordan 2003). After processing with NLTK, we can use Scikit-Learn’s TF-IDF vectoriser to convert our cleaned descriptions to numerical representations. We can then apply LDA to extract 5-10 topics. These topics can be interpreted by examining their most frequent words. By correlating these artwork topics with time, location and socio-political scenarios, we can visualise the changing concerns of Melbourne’s public art scene. Sci-kit learn is computationally efficient and is designed to integrate well with exports from NLTK. This project will make arbitrary assumptions about a number of topics, but can be informed by coherence scores and manual interpretations, looking at the distributions with a subjective lens. Despite LDA being unsupervised, strong domain knowledge about Melbourne’s cultural history is required to interpret its results.

Supplementary Tools

pandas

Pandas is the backbone of tabular data wrangling and cleaning in Python, and is essential for loading and cleaning the dataset (McKinney 2010).

seaborn

Great plots are essential to good data-based storytelling, and seaborn offers high-level APIs for creating informative and attractive statistical graphics (Waskom 2021).

geopandas & folium

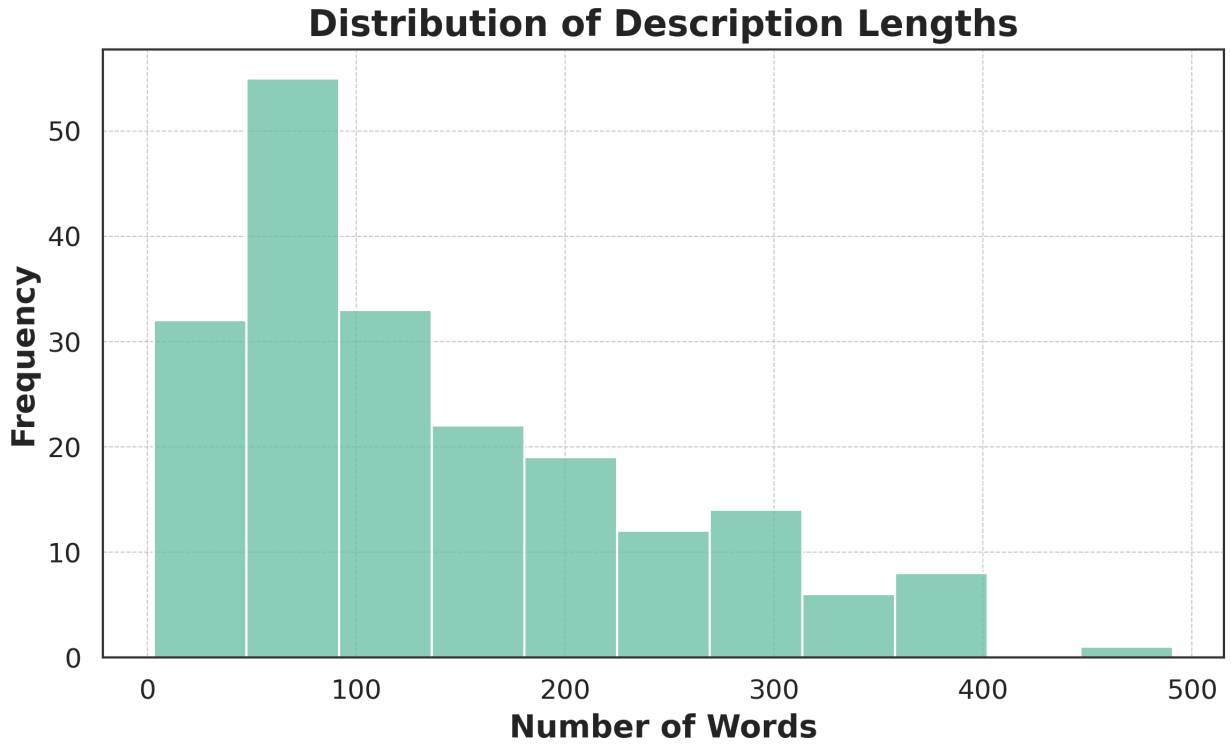
The dataset includes precise coordinates of the artworks, which can be plotted using these libraries. Overlays with topic classifications and timeline can give us insights into how urban development influenced art, and compare how different areas engage with particular narratives.

quarto

Quarto documents can integrate code, visualisations and interpretations into simple reproducible academic reports, with great support for referencing and viewer-friendly exports. It encourages literate programming, keeping analysis and explanation together (Posit 2023).

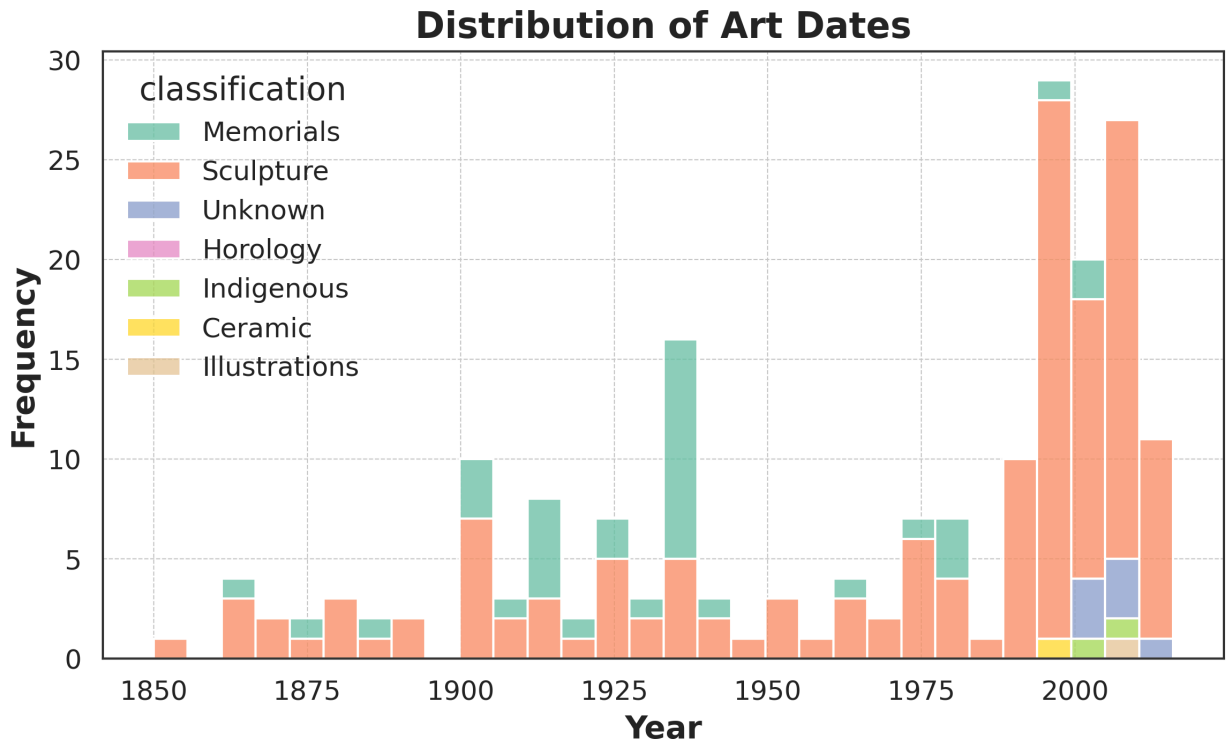
Preliminary Analysis

To gauge the feasibility of the project, some exploratory analysis was conducted on the dataset. Since the “description” field is a core of the research, the field must contain sufficient information for topic modelling.



We see that the descriptions are generally detailed, with most installations having around 50-100 words, with some descriptions extending up to 450 words. From empirical information, we know this is generally sufficient for modelling, and since they’re written by subject matter experts, they’d be information-dense.

Melbourne has a rich development history, and we notice that it’s also reflected in its art installation history.



We notice a sudden peak in memorial installations around the 1930s, a significant slowdown in new installations from the 1950s to 1970s and a sudden rise from the 1990s. These shifts aren't just data points - they're openings into stories of urban stasis and a city expressing its voice through public form.

Previous Related Works

Although temporal topic modelling on public art is relatively rare, some studies and blog posts explore similar intersections of text, art, and urban data that this project can take inspiration from:

- Trends in Contemporary Art Discourse: Using Topic Models to Analyze 25 years of Professional Art Criticism (Roose, Roose, and Daenekindt 2018)
- The Arts Engagement Project (University of Michigan 2021)

Bibliography

- Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, 2009.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (2003): 993–1022.
- City of Melbourne. "Outdoor Artworks." *Data.Melbourne.vic.gov.au*, 2024. https://data.melbourne.vic.gov.au/explore/dataset/outdoor-artworks/information/

- McKinney, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, 2010.
- Pedregosa, Fabian, et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (2011): 2825–2830.
- Posit. "Quarto – Scientific and Technical Publishing." 2023. https://quarto.org/
- Roose, Henk, Elke Roose, and Stijn Daenekindt. "Trends in Contemporary Art Discourse: Using Topic Models to Analyze 25 Years of Professional Art Criticism." *Cultural Sociology* 12, no. 3 (2018): 364–388. https://repub.eur.nl/pub/105624/
- Scikit-learn. "sklearn.decomposition.LatentDirichletAllocation." Accessed May 8, 2025. https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.LatentDirichletAllocation.html
- University of Michigan. "The Arts Engagement Project." *ArtsEngine*, 2021. https://artsengine.engin.umich.edu/the-arts-engagement-project/
- Waskom, Michael. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6, no. 60 (2021): 3021. https://doi.org/10.21105/joss.03021