Project Proposal: The potential regeneration for Tate Modern

**1. General description:**

Tate is a museum that houses the United Kingdom's national collection of British art, and international modern and contemporary art. Recently, Tate announced  across the galleries that new space is urgently needed that  acquisitions from Latin America, Africa, Asia, the Middle East and Eastern Europe, including Meschac Gaba and Cildo Meireles will be brought together on a global scale.Therefore,for this project we are looking to play the role of consultants who are helping the museum inventory its collection, in preparation for a possible expansion. Apart of the inventory process is to explore what “kind” of art Tate has - that is, the origin, the style, the artists, the time period, and the gender (of the artist). Understanding this will help us understanding what they have, and how they should expand their collection, how to diversify and what strengths they can focus on and promote. Knowing these details will also help us understand how the country perceives culture and value.

On a logistical level, we also want to understand the size and material of the pieces so that we can understand what sort of space we need, and what sort of resources are required to move these pieces. This will help us estimate the cost of this expansion and for future storage.

**2. Identifying datasets:**

We found two datasets that contain the metadata for around 70,000 artworks that Tate owns or jointly owns with the National Galleries of Scotland, as well as data for around 3,500 associated artists is also included. We can join/merge these two datasets to explore the artwork available and the artists associated with those pieces.

<https://www.kaggle.com/rtatman/the-tate-collection>

<https://www.kaggle.com/vivianapinto/the-tate-artists>

**3. Literature review:**

Currently, the studies of Tate Modern can be divided into **two categories**: The first group is the academic studies based on the development of Tate Modern as a profound contemporary art organization. For example, Corinna Dean, Caroline Donnellan and Andy C Pratt wrote an article called “Tate Modern: Pushing the limits of regeneration” on Elsevier which questioned the conventional limits to regeneration, Identifies the manifest tensions between the instrumentalization of museums and galleries. However, this kind of study uses too many museology terms for people to understand.

Later on , with the development of data science, we can play more games with data to tell the story in a more illustrated way. But people mostly are interested in the artists, not the natural qualities bring by the masterpieces themselves. In 2015, Heatber Bircball and David Fraser Jenkins published an article on The British Art Journal called: “Is it better to buy artists young? An investigation of the database of Modern British art at Tate” [1] The results shows that today Tate is more likely to pursue the work of young, emerging artists. Also, the research is narrowly confined within artists. Very little research was done into the artworks and the artists together.

Therefore, our project decided to join two datasets together. Current studies based on either of the two datasets are still in basic level. Vivianna Pinto published a well-structured analysis to study the basic information about the artworks including acquisition time and title test analysis, which covered nothing on its dimension and artists gender compositions.  As consultants from Heinz College, we want to search for the intrinsic social problems behind the artworks and try to offer Tate some suggestions for regeneration with the support of numbers.

[1]<https://www.jstor.org/stable/41614651?Search=yes&resultItemClick=true&searchText=tate&searchText=modern&searchText=database&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dtate%2Bmodern%2Bdatabase&refreqid=search%3Aa5ac525062317adf7b8b888c749039b4&seq=1#metadata_info_tab_contents>

**4. Set of questions to be explored form the data**

Our statistical analysis will be an investigation between different style characteristics (region, art-type, year, materials), and the size of the object. This will tell us in the future what sort of collections the museum can bid for, and whether the museum has the space to hold those collections (and how much space that is).

* How many artists have appeared in Tate Modern
* Are the artists mostly male or female?
* What is the general space needed for storing/displaying the art on hand?
* What is the medium of art in possession (which will dictate what is required for transportation of the artwork).

**5. Set of hypotheses to be tested from the data**

* Throughout the acquisition years, Tate Modern tended to acquire works from male artists.
* Tate tended to get artworks with that was bigger as they gained a larger endowment and had more space.
* There is an over-abundance of certain art styles (paintings on paper) and not enough of a balance with other types of artwork.
* The Tate Collection is dominated by artwork of the early 20th century.

**6. Our strategy** **for testing these hypotheses and visualizing:**

To facilitate some of this analysis we plan on building functions that take characteristic parameters (size/volume, materials, art-type) and return pieces that fit those parameters. That will allow us to identify which pieces can be moved or grouped together (to save time transportation logistics).

Other functions we plan on creating would be to extract pieces based on the material type. This will require more sophisticated string subsetting that can extract single words from a compound phrase (i.e. “Screenprint and mezzotint on paper”)

We plan on displaying some of the location-based information with a GIS map as well and seeing if we can correlate the types of art to different locations. We will be using the leaflet package to integrate that with R.

Conclusion: more female artist needed for regeneration

*(A more advanced data mining task may be to see how names of the pieces can be clustered together. Many art pieces have similar sounding names, it may be that those names and artists have a strong correlation, or that names of the piece have a strong correlation to a time period. This is a more advanced unstructured data analysis task that will be optional)*

**7. R-code:**

1. *Data cleaning*
2. *Merge the two datasets*
3. *Find unique values for certain columns*
4. *Do a bar chart of collection years*
5. *Show some general exploratory*

*Etc.*