THE UNIVERSITY OF
MELBOURNE

MAST90106
Project Report

GROUP 7

Haonan Zhong *867492*

Samy Allouache *1210426*

Supanuth Amorntiyanggoon *1211674*

Xuan Hung Ho *1276655*

Haocong Chen *987916*

1 Introduction

1.1 Project Domain

Canvas paint can be thought of as a composite object, and it generally consists of five components. The first one is *auxiliary support*, it is the framework on which a painting canvas is stretched, which can be categorised into stretcher and strainers. The former are rigid frames with flexible and expandable corners, which allows adjustment of tension of the canvas, while the latter are frames with fixed and non-expandable corners. Then, the *flexible support*, the canvas fabric, e.g. cotton, linen, and bast fibre. After that, the *sizing layer*, which is a material applied to the flexible support to seal the surface. Following that is the *ground layer*, a priming layer applied to the flexible support on top of the size layer, and its primary purpose is to seal and prepare the surface to accept oil paint. Finally, it is the *paint layer*.

The practice of oil paint was spread from Western Europe to Southeast Asia in the nineteenth and early twentieth centuries, reflecting the colonial development and religious conversation in the region at the time. Conservation is an important procedure to preserve cultural heritage. However, the conservation of Southeast Asia paintings poses two main challenges, the effect of tropical climate and the used materials. These challenges present a significant source of problems for the storage of the paintings due to the high relative humidity (RH) and the elevated temperature in the tropical region, since components react differently to various temperatures and humidity. For example, the sizing layer with rabbit skin glue are considered as highly responsive materials to RH, which might swell and cause the paint layer on top to fall off. And due to war and expense limitations in the early 20^{th} century, the supply of imported artists' materials is limited. Local artists without access to imported materials started to source painting materials locally instead. Therefore, it is important to study what materials are used in a painting and learn their expected behaviour in the tropical climate.

1.2 Project Overview

Our client for this project is *Dr Nicole Tse* from Grimwade Centre for Cultural Materials Conservation. Most of the conservation protocols were initially developed from research taken from the northern hemisphere. Nicole's study aims to develop regional relevant conservation solutions for Southeast Asia paintings. The dataset provided to us consists of 208 Southeast Asia painting condition reports. Condition attributes related to the five components of paintings are manually investigated and collected by Nicole during her PhD studies. As stated by Nicole, our expected outcome of this project can be categorized as:

- **Clean and transform the raw dataset**, which comprises the conservation record of 208 early 20^{th} century canvas paintings from Southeast Asia, **into an interactive dashboard for visualisation**.
- **Identify the correlation** between tropical climate, e.g. relative humidity and temperature, and painting conditions.
- **Map the relationship** between artists and their used painting materials. In addition identify the commonly used materials at the time.
- **Deliver a presentation** of our findings and the dashboard to Nicole and her research student at the end of next semester.

Our project involves data cleaning, data visualisation, and correlation analysis from the data science perspective. The challenge that we faced this semester was that none of our team members had any profound knowledge of canvas painting and painting conservation before this semester. But as the semester progresses, we've gained a better understanding of the domain through background readings and participating in Nicole's lab demonstration. Another challenge is related to the given dataset. The dataset was put together by our client almost twenty years ago using a legacy version of the software FileMaker Pro. Throughout the year, the software has gone through multiple updates. Therefore, we notice some inconsistencies and errors when exporting the dataset, which makes it more difficult for us when clean the data. This challenge will be discussed further in the data analysis section.

2 Related Work

In this section, we will be reviewing works and literature that are closely related or helpful to our project.

2.1 Dashboard

Dashboard is an information management tool with a graphical interface that allows users to visualise and understand the data more efficiently. Before we can proceed to build our dashboard, we must first identify the suitable tool among the many dashboard packages. Our supervisor has suggested that we use the R Shiny package to build our dashboard, given its simplicity and the product user does not need to understand R to use it, which is perfect for the client. Shiny dashboard has been implemented for a variety of areas, such as health care and environment. In the early stage of the COVID-19 pandemic, the dashboard web map developed by the London School of Hygiene and Tropical Medicine provides the latest information on the spread of the virus. It allows people to make comparison with other recent disease outbreak. Another example is polluter alert dashboard, which effectively let people in Warsaw to monitor and report air pollution sources. These examples effectively help and gave us inspiration on how to help to implement and improve the user interface of our proposed dashboard.

2.2 Correlation Analysis

As the first challenge mentioned, The client requirement would like to see the pattern or model to predict the paint condition response from predictor parameters such as temperature and humidity, strainer/ stretcher, elastic/plastic material behaviour, size layer visible etc. Based on this requirement, the relevant approach to find the relation among them could be binary logistics regression. Our group found the relevant paper, which relates to client requirement [2], The researchers assert the binary logistics regression for assessing the environmental impact on historical artifacts by using air quality temperature and humidity. The researcher used 794 records to develop and validate the binary logistic regression model and also, want to find parameters that have a significant impact on those artifacts. Furthermore, the model achieves an accuracy prediction around 95 percent. Hence, our group believes that the model could be a useful tool in decision-making regarding preventive preservation in the future.

3 Data Analysis and Preliminary Visualisation

3.1 Dataset

As we mentioned in the project overview, the data provided to us consists of 208 painting record gathered in a single FileMaker Pro file, where each record is an condition report. An example of the condition report is shown in figure 1.

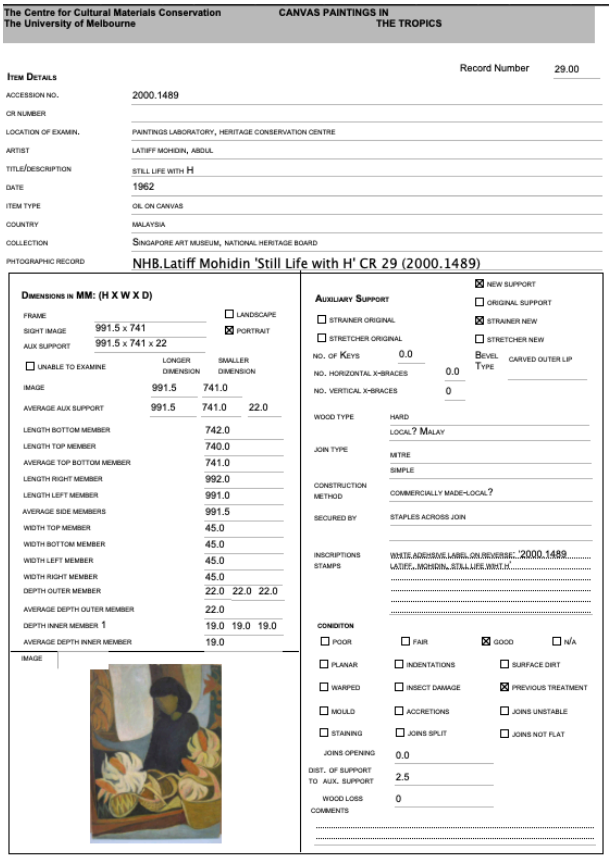

Figure 1: Example of a FileMaker Pro Condition Report

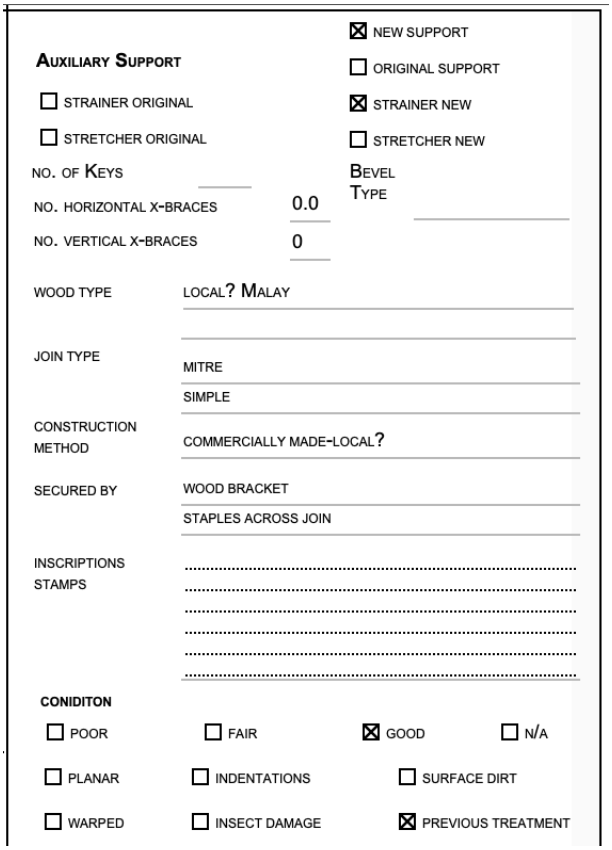

Figure 2: Zoom on the auxiliary support part

A condition report starts with some general information such as the painting's accession number (which is an identifier), the name of the painting, the name of the artist, the country and the museum collection it belongs to, the dimensions and a thumbnail of the painting if available. It then focuses on the condition of every main component of the piece of art which are the auxiliary support, the painting support, the ground layer, the painting layer, the surface coating and some miscellaneous information.

For each of these main components, the report offers a large amount of attributes that need to be filled by the author of the report. There are two different types of attributes: they can either be text fields that the reporter can fill or boxes that can be ticked. An example for the text fields would be the "comment" attribute from the auxiliary support and one for the boxes would be the condition (poor, fair, good, excellent or N/A) of the ground layer (figure 2).

3.2 Data Preprocessing and Cleaning

As said before, the FileMaker Pro file could not directly be used in the data processing pipeline. It was exported into an excel file using the "export to excel" option from FileMaker Pro. Unfortunately, the excel file could not directly be used either. Indeed, it presented two types of problems:

- Some of the exported values were just not correct like shown in the red rectangle in figure 3.
- Categorical and ordinal variable that have several possible values were exported as separated variables in the excel file (cf. blue rectangle in figure 3)

	A	C	J	N	X	Y	Z
1	accession no	artist	stretcher original	strainer original	poor (aux support)	fair (aux support)	good (aux support)
2	UPVMA-III.00240	Gallardo, Alladin		strainerstrainer original			good
3	UPVMA-III.00085	Ancheta, isidro		strainerstrainer original	poor		
4	UPVMA-III.00292	Lagniton, Jose		strainerstrainer original			good
5	UPVMA-III.00126	Buenaventura y Espana, Oscar		strainerstrainer original		fair	
6	UPVMA-III.00161	Cristobal, bonifacio		strainerstrainer original			good
7	UPVMA-III.00289	Jervosa, Fortunato		strainerstrainer original			good
8	UPVMA-III.00180	Dumlao, Antonio	stretcher original	strainer			good
9	UPVMA-III.00177	dizon, vicente alvarez		strainerstrainer original			good
10	UPVMA-III.00205	Enriquez, Romeo		strainerstrainer original	poor		

Figure 3: Exported excel file

We couldn't find any explanation nor any pattern for the first problem even when after talking about it with Nicole. This resulted in the impossibility to implement an algorithm solve it. Therefore, we had to clean it manually by correcting all the errors. We then wrote a program that processes the clean excel file to have a preprocessed, ready for computation dataset. The output of our program is a CSV file with fused columns for categorical and ordinal variables. A screenshot of the cleaned CSV file is shown in the appendix.

3.3 Preliminary Visualisation and Analysis

In our preliminary analysis, we have chosen to visualise several attributes from the dataset to establish a qualitative understanding of the data. First, we aim to investigate what time frame the given set of paintings was painted. As shown in figure 4, we can observe a heavily left-skewed distribution, as most of the paintings were clustered together between the 1930s and the 1960s. Furthermore, the earliest period can date back to the 1850s.

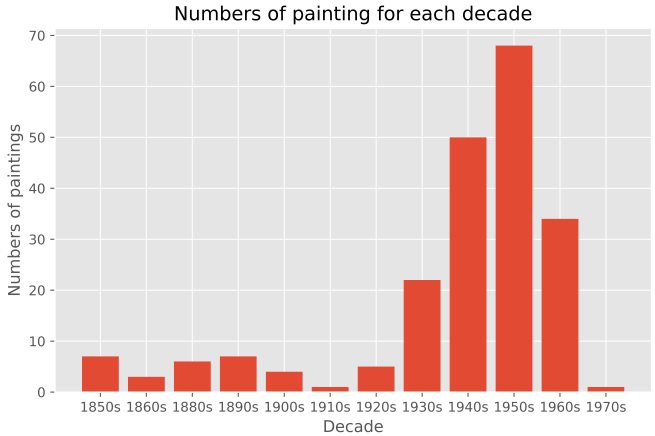

Figure 4: Numbers of Painting for each Decade

Secondly, we aim to investigate the origin of the examined paintings. As figure 5 depicted, of the 208 examined works, 29.8%, 27.4%, and 26.4% of the paintings have an origin from the Philippines, Singapore, and Malaysia, respectively. Whilst Thailand has the least amount of paintings, only accounting for 16.3% of the examined works.

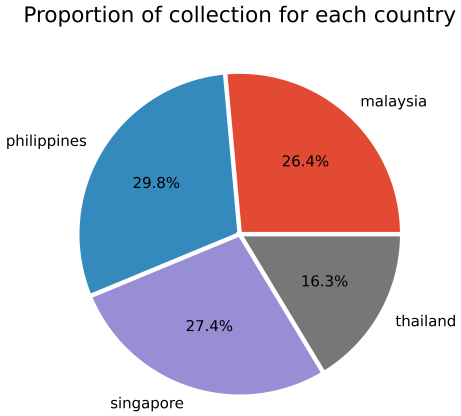

Figure 5: Painting Distribution by Country

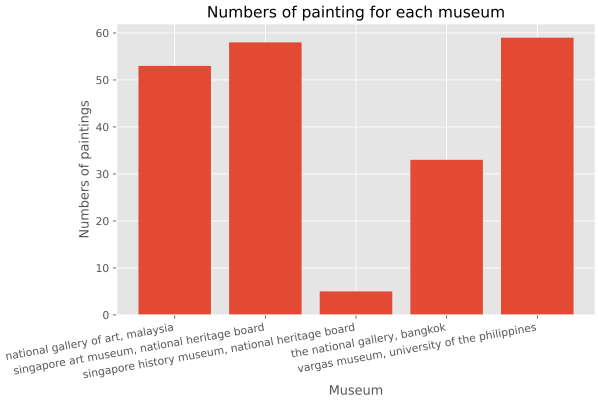

Figure 6: Painting Distribution by Museum

Next, we have shown the number of paintings examined in each museum in figure 6. Of the five museums, the Singapore History Museum has the least number of paintings investigated, with only 5 paintings, while Philippines' Vargas Museum, Singapore Art Museum and the Malaysia National Gallery of Art top the chart, with 59, 58, and 53 paintings investigated, respectively. Furthermore, the National Gallery from Bangkok has contributed 33 paintings for examination.

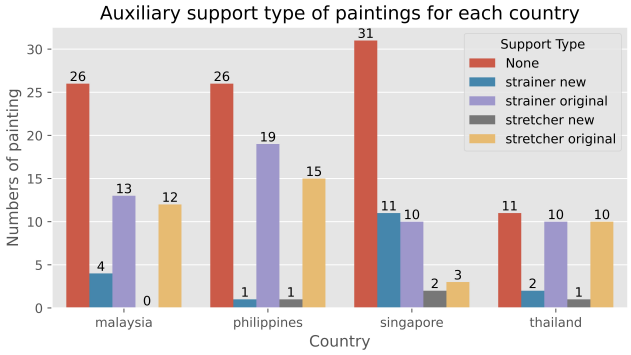

Figure 7: Distribution of Auxiliary Support

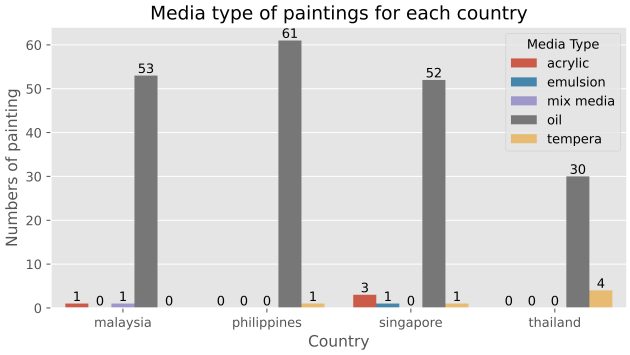

Figure 8: Media Type Popularity

As we mentioned in the introduction, components from a painting may react differently to temperatures and humidity. Here, we have shown some visualisation of different components, which may allow us to grasp the basic concept of the components from the examined works. The bar plot in figure 7 illustrates the distribution of auxiliary support type of paintings for each country, we can see the numbers of examined paintings without auxiliary support account for a large proportion in each country. While original strainer and stretcher share roughly the same amount from paintings originate from Malaysia, the Philippines and Thailand, respectively. Figure 8 on the right shows the media type used in each country's paintings, we can observe that oil is the most popular media type in each country.

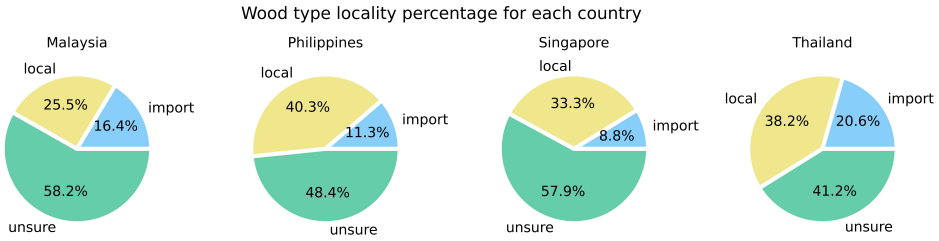

Figure 9: Wood Type Locality percentage for each Country

Now, we aim to investigate the used wood's locality in each painting. As we can see from figure 9, a large proportion is either unsure or does not have wood support installed. And it is worth mentioning that, in each of the four countries, there are more paintings with local wood support, accounting for at least 25% of the examined paintings from each country, than those with imported wood support. This well reflects our client's conclusion that imported materials were interrupted by World War II and the high cost at the time. Therefore local artists are sourcing local alternatives instead.

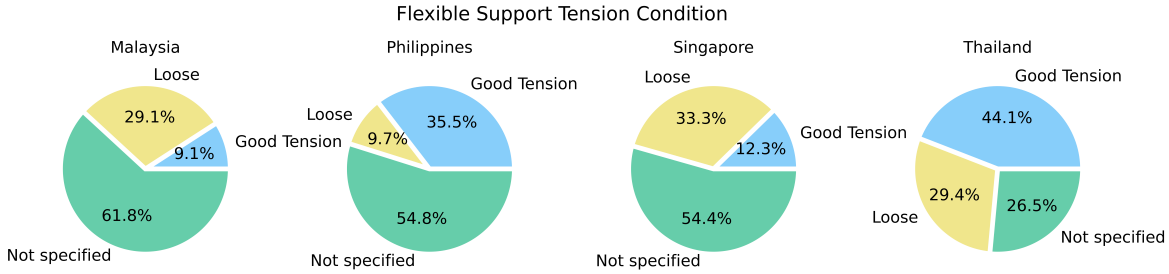

Figure 10: Flexible Support Tension Condition

Finally, we aim to investigate the flexible support, in terms of their tension, mould, stain, and insect damage condition. As figure 10 depicted, a large proportion of the tension condition were not specified in the examined works from Malaysia, the Philippines, and Singapore. On the other hand, paintings with loose flexible support accounts for approximately 30% of the examined works from Malaysia, the Philippines, and Thailand. While paintings with good tension accounts for 9.1%, 35.5%, 12.3%, and 44.1% of the examined works from Malaysia, the Philippines, Singapore, and Thailand, respectively.

- **Mould:** As we can see from figure 11, 12.9% of the examined paintings from Philippines has mould presented in the flexible layer, while the mould condition are less severe in examined paintings from Malaysia, Singapore, and Thailand, with 0%, 1.8%, and 2.9%, respectively.

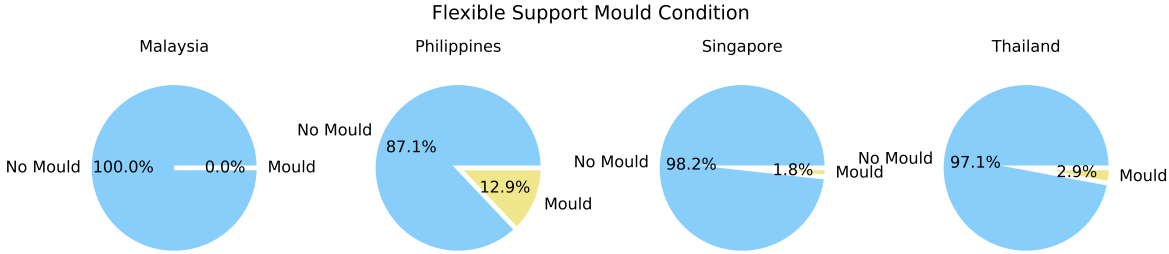

Figure 11: Flexible Support Mould Condition

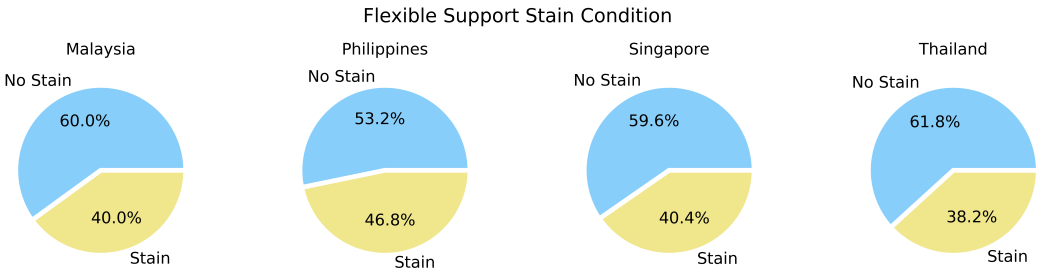

Figure 12: Flexible Support Stain Condition

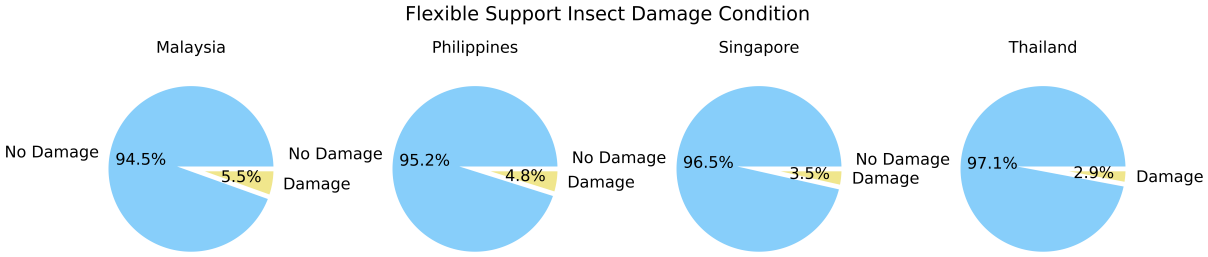

Figure 13: Flexible Support Insect Damage Condition

- **Stain:** Stain are presented in 40%, 46.8%, 40.4%, and 38.2% of the examined works from Malaysia, the Philippines, Singapore, and Thailand, respectively, as shown in figure 12.
- **Insect Damage:** As for insect damage, we can clearly see that only a small portion of the paintings are damaged by insect in all four countries. In which accounted for 5.5%, 4.8%, 3.5%, and 2.9% of the examined works Malaysia, the Philippines, Singapore, and Thailand, respectively.

Lastly, throughout the stage of visualisation, our group has combined the given dataset with the background readings provided by Nicole, which enables us to form a better understanding in terms of historical context of the dataset, and sets up a good foundation for correlation analysis in the next semester.

4 Proposal for Next Semester

Based on the work we have accomplished during the first semester, we would like to propose a plan for the second part of this data science project. We will focus on the two main goals:

1. **Interactive Dashboard for Visualisation**: We will focus on developing an informative and interactive dashboard using Shiny package from RStudio. We plan to deliver the product by sprints, therefore the client could give feedback periodically and we can improve the product as well as being adaptive to any request to changes from the client.
2. **Further data exploration** Given the fact that the dataset we were given might not be sufficient for the dashboard, our team has taken into consideration steps to process and transform the data in order to get more valuable insights. We will focus on two stages of data science pipeline:
   (a) *Processing the data to analyse*: In this stage, we will try to apply data augmentation techniques to enrich the small dataset that we are given. Plus, we will keep working with the client to collect more data if available. In addition, data cleaning script will be implemented simultaneously.

(b) *Exploring the data*: We will try to apply several statistical or machine learning models in a bid to find whether there is any relationship between materials and artists (as suggested by the client) as well as to learn which attributes have major effects on painting conditions such as temperature or RH. Classification will be taken into account to clustering the group of artist with material and regression models would also be taken into account to predict the result of the condition based on given dataset.

The most important goal of our team is to satisfy client's requirements, which is the Dashboard for Visualisation. Data exploration will facilitate the development of the dashboard, but we will prioritise team's resources for the dashboard implementation.

5 Our Project Timeline

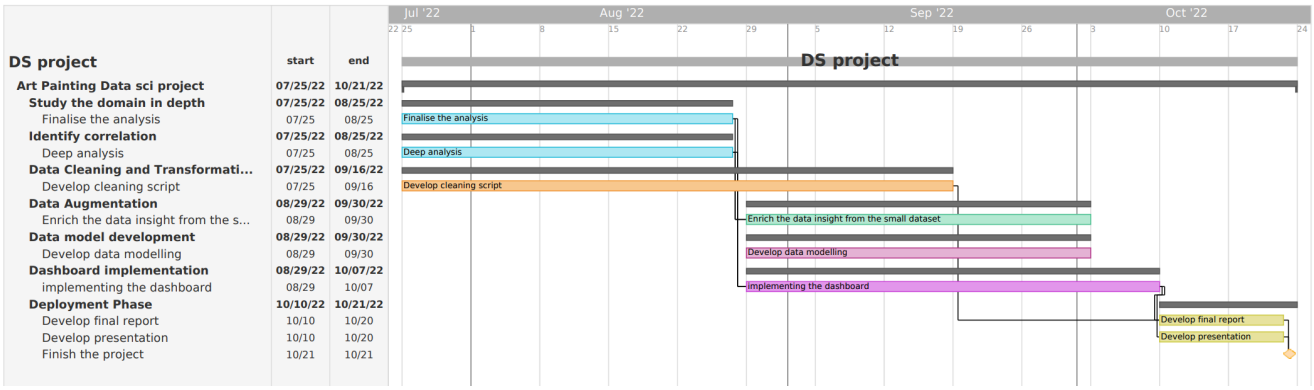

Figure 14: Project timeline Gantt Chart

The Gantt Chart illustrates the estimated time and person in charge for tasks we plan to do during the next semester.

- **Study domain in depth**: Supanuth Amorntiyanggoon will be in charge of this task. In this semester he had dived deep and gained profound knowledge about the domain. In order to explore deeper and search for relationship between features in the data, more domain knowledge is required. Thus, our team will keep studying the art field in depth.
- **Identify correlation**: This task will be managed by Samy Allouache, because of his experience with this dataset. He had been working on cleaning the dataset since the client gave dataset to the team.
- **Dashboard Implementation**: Haonan Zhong will be responsible for this task, since he had been working closely with the client for the semester. He will manage the team to deliver the dashboard every sprint.
- **Data Augmentation**: Xuan Hung Ho will be accountable for data augmentation. This needs to be done to facilitate the data modelling task.
- **Implement statistical / machine learning models**: Haocong Chen will be held accountable for this task.

Appendix

Complementary figures

	Accession number	Country	Support Type	Auxiliary support Condition	planar (auxiliary support)	warped (auxiliary support)
0	UPVMA-III.00240	philippines	strainer original	2.0	1	0
1	UPVMA-III.00085	philippines	strainer original	0.0	0	1
2	UPVMA-III.00292	philippines	strainer original	2.0	0	0
3	UPVMA-III.00126	philippines	strainer original	1.0	0	1
4	UPVMA-III.00161	philippines	strainer original	2.0	1	0
5	UPVMA-III.00289	philippines	strainer original	2.0	1	0
6	UPVMA-III.00180	philippines	stretcher original	2.0	1	0
7	UPVMA-III.00177	philippines	strainer original	2.0	1	0
8	UPVMA-III.00205	philippines	strainer original	0.0	0	1
9	UPVMA-III.00430	philippines	strainer original	2.0	1	0
10	39/2521	thailand		0.0	0	0
11	UPVMA-III.000452	philippines	strainer original	2.0	1	1
12	UPVMA-III.00293	philippines	strainer original	2.0	0	0
13	UPVMA-III.00269	philippines	strainer original	1.0	1	0
14	UPVMA-III.00338	philippines	strainer original	1.0	0	1
15	UPVMA - III.00296	philippines	strainer original	2.0	1	0
16	UPVMA - III.00003	philippines	strainer original	2.0	1	0
17		malaysia	stretcher original	0.0	0	0
18	UPVMA - III.00058	philippines	stretcher original	2.0	1	0
19	UPVMA - III.00055	philippines	stretcher original	2.0	0	0
20	UPVMA - III.00208	philippines	strainer original	1.0	0	1
21	UPVMA - III.00239	philippines	strainer original	2.0	1	0
22	38/2530	thailand		0.0	0	0
23	UPVMA - III.00432	philippines		0.0	0	0
24	UPVMA - III.00453	philippines	strainer original	2.0	1	0
25	p0820	singapore		0.0	0	0

Figure 15: Cleaned CSV file

References

[1] Tse, N., & Sloggett, R. (2008). Southeast Asian oil paintings: supports and preparatory layers. Archetype Publications.

[2] Bucur, E., Danet, A. F., Lehr, C. B., Lehr, E., & Nita-Lazar, M. (2017). Binary logistic regression—Instrument for assessing museum indoor air impact on exhibits. Journal of the Air & Waste Management Association, 67(4), 391–401.

[3] Tse, N., & Sloggett, R. (2008a). A preliminary understanding of the behaviour of oil paintings in tropical Southeast Asia. ICOM COMMITTEE FOR CONSERVATION.