BLUETHUMB CANVAS SUCCESS PROJECT PROPOSAL:

Data analysis and development of an artwork growth score model to steer new artists towards market trends and improve their artwork sales potential

bluethumb

By: Thinithi Bulathsinghala

Student ID: 27523306

Report words:1,991

CONTENT

1.1 Introduction

Creating a brand name in the art industry is indeed a challenge on its own. Bluethumb was a company established in 2012 with a mission to empower Australian Artists. They did not agree with the habit of having to wait for an artist to reach the level of having their own gallery or exhibition, instead started Australia's online art gallery which today represents over 20,000 emerging and established artists (Bluethumb, 2023). I, myself, having the passion for art, stumbled upon their website which was very well made indeed. But as a new artist and user myself, there were some points I realized it lacked. They have an enormous customer base and an even larger artwork portfolio. Data science could revolutionize the way Bluethumb operates, if the website data could be utilized to provide insights into what are the current market trends and where do one's artwork stands before the artist can start selling. This will enable emerging artists to increase the artwork sales potential by aligning with market trends whilst improving their own profiles to the level of those who are well established. For Bluethumb, this will not only increase their sales due to more artist engagement but also it will be a significant value add to their corporate social responsibility (CSR).

"Bluethumb Canvas Success Project" aims to provide insights on the market trends in the art industry within Australia to guide new artists to areas with high demand. Further summary insights into making better data driven decisions when it comes to sizing, texture, topic…etc. The artwork data and artists profile information coupled with these insights will be incorporated into developing an art growth score model to indicate an artwork's potential to sell. The model initially will not take into account the image of the artwork itself but rather other variables that influence a buyer's decision. For example, art style, topic, size, price, frame, artist popularity, follower, count..etc. It could be later extended to include image analysis in phase 2. The objective of a growth score model is to provide a comparison with the artists who are selling and to highlight weak areas to improve on or to make better artwork wise decisions. As per Martin et al (2020), the moment the human brain encounters a mismatch between the goal and capacity it initiates a learning process. This development will provide new artists with a guide they can refer to in order to improve and focus their efforts into achieving the end goal of selling and becoming a more established artist. Whilst for Blue thumber it will drive more artist engagement which converts to better sales and increased CSR for the brand name. The combinations of advanced analytics for market analysis, artist comparison, propensity to buy mapped into an artist growth journey in a user friendly platform will be a novel data science initiative that stands out in the art industry.

1.2 Goal and Objectives

The goal of this project is to enable artists to drive growth which eventually converts to increase in sales. The project is developed particularly for Bluethumb. The objectives are briefly mentioned below:

- **Efficient use of resources** – As it will enable emerging artists to focus their efforts on making creatives in areas of market interest.
- **Drive growth** – As the artwork grade would change along with the artist profile and other variables which will act as a growth indicator.
- **Drive Artist Interaction** - The more an artist can see a quantified indication of results from their efforts and not just sales the more motivated they are to continue interacting on Bluethumb
- **Drive sales** - Support emerging artists into making their first sale and much more as they grow.

As of today, there are several challenges artists utilizing this online platform face, the project outcome on addressing them are shown below:

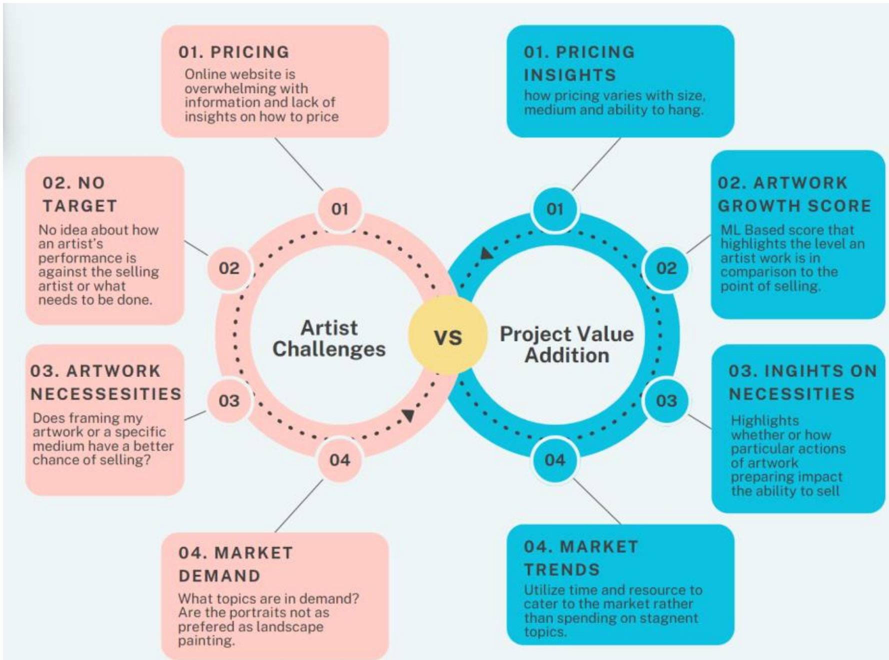


Figure 1.2.1: Project value addition addressing artist challenges.

Source: Author

1.3 Data Science Roles

- **Business Analyst** - Pitching ideas on how to maximize the business value of the project, managing team collaboration and tracking project progress.
- **Data Analyst** - Conducting data collection, wrangling, pre-processing data followed by data analysis and visualization.
- **Data Scientist** - Identifying data collection sources and automating processes for the project. Collaborating with IT to ensure smooth data integration of multiple systems. Overlooking the data science project direction and developing the machine learning model prior conducting project operationalization.
- **Data Engineer** - Design and build scalable data ecosystems for database management.
- **Data Architect** - Make the blueprints for system integration for smooth project roll out to the end user.
- **Database Administrator** - Maintain database access, project documentation and ensure project alignment with data governance.

1.4 Business Model

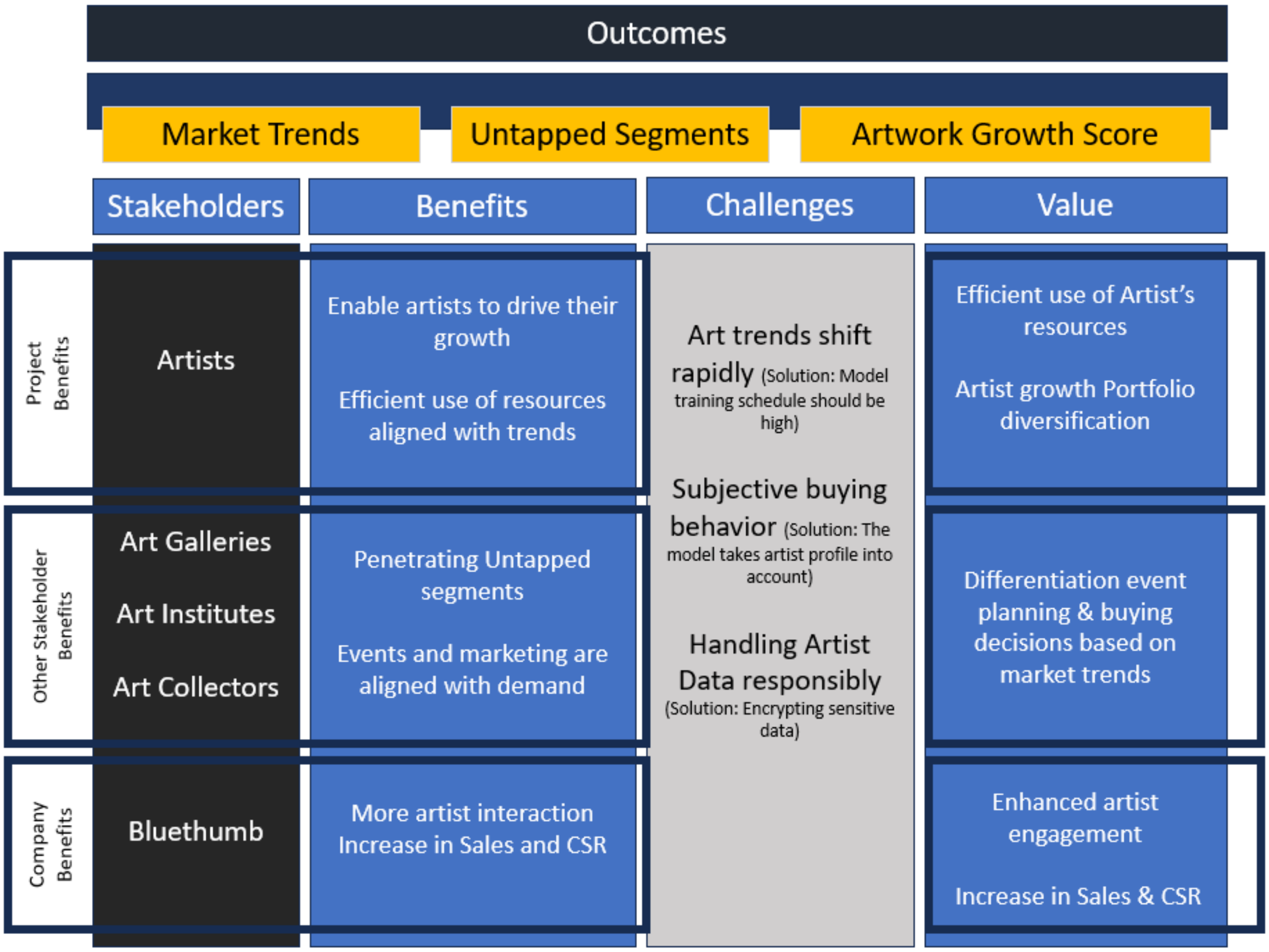


Figure 1.4.1: Business Model Summary

Source: Author

1.5 Data Sources and Collection

The required for this project is already available on the Bluethumb website. When developing this project within Bluethumb it is a matter of tapping into their own website data utilizing APIs which will be enabled by the data engineer, data architect and the data scientist through collaboration. In order to develop a prototype the data is obtained from the website using a web scraper known as Octoparse. The format of the feed page is shown below where all published art is posted in the order it was published. The points that have been extracted are indicated by the red marker

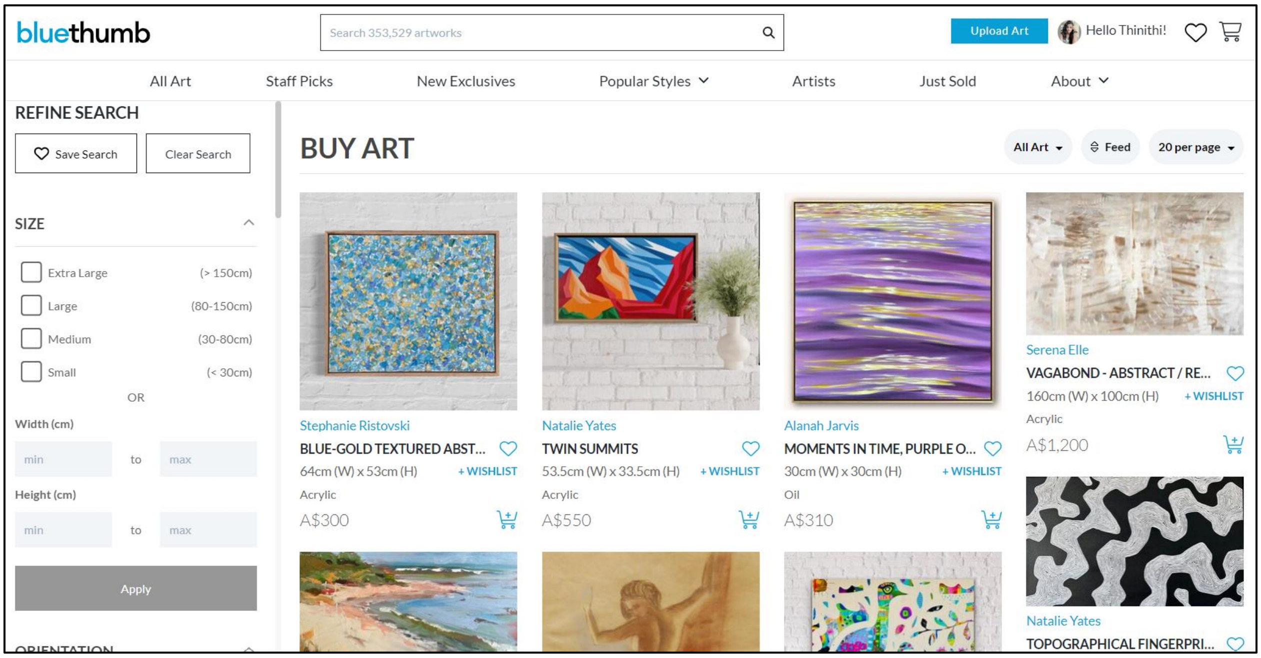
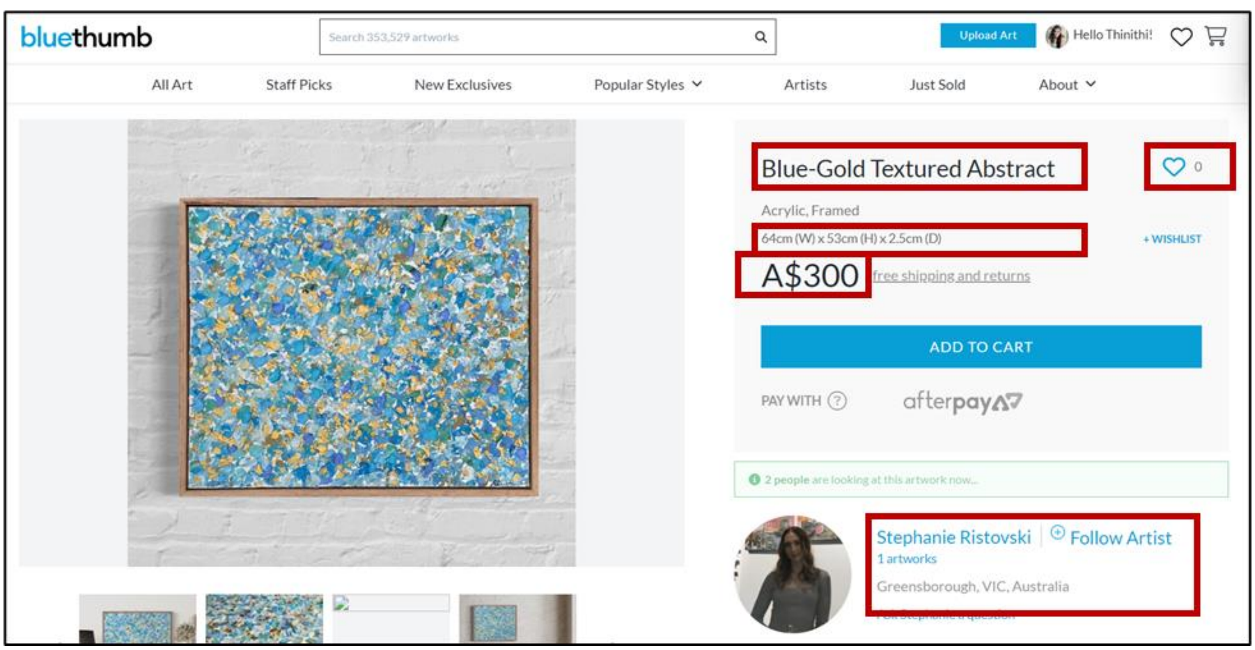


Figure 1.5.1: Bluethumb Feed Page

Source: https://bluethumb.com.au/?gclid=Cj0KCQjw1aOpBhCOARIsACXYv-cuMsAMRdOqb9kvzg7I34NS4fvIKS3wMTdO-UitwVTrS3HuWELHtLkaArLJEALw_wcB



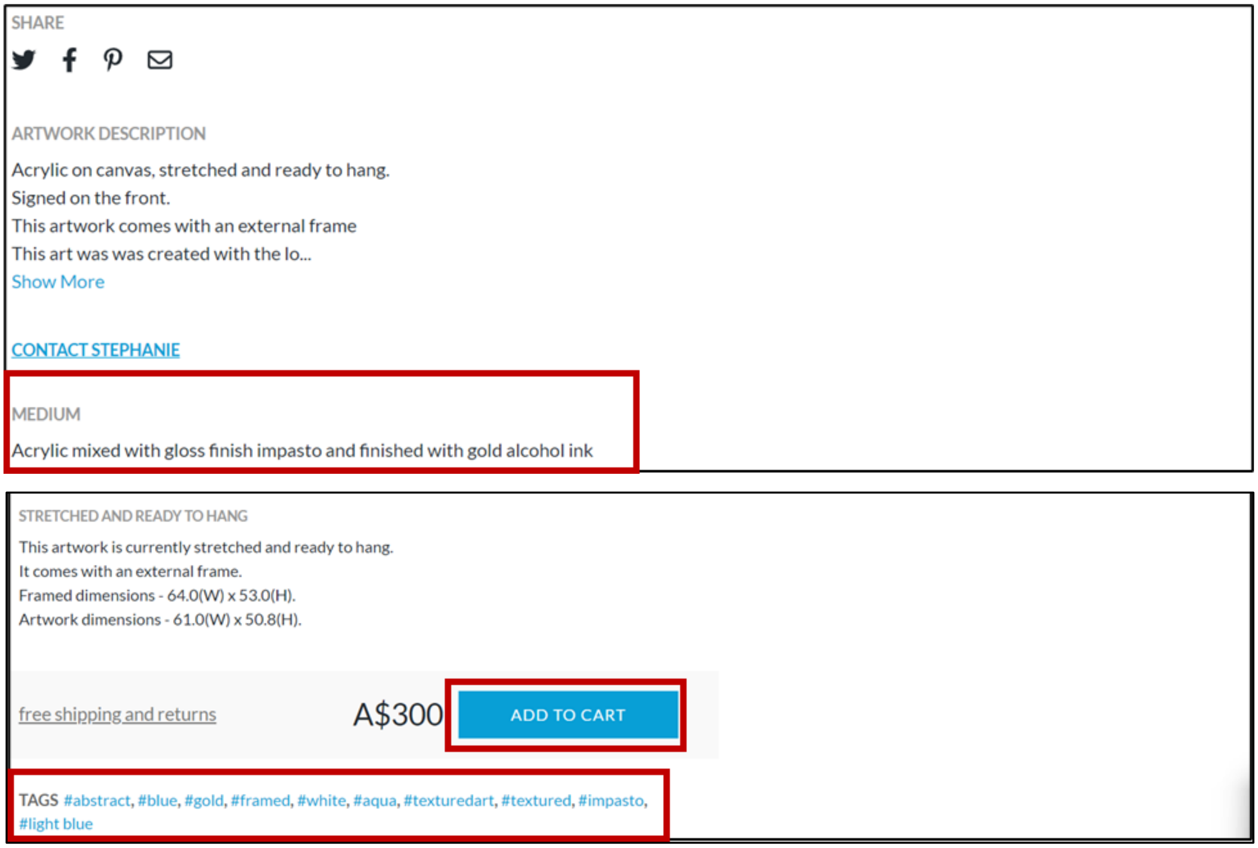


Figure 1.5.2: Webpage once a specific artwork has been clicked on (Red signifies data to be extracted)

Source: https://bluethumb.com.au/?gclid=Cj0KCQjw1aOpBhCOARIsACXYv-cuMsAMRdOqb9kvzg7I34NS4fvIKS3wMTdO-UitwVTrS3HuWELHtLkaArLJEALw_wcB

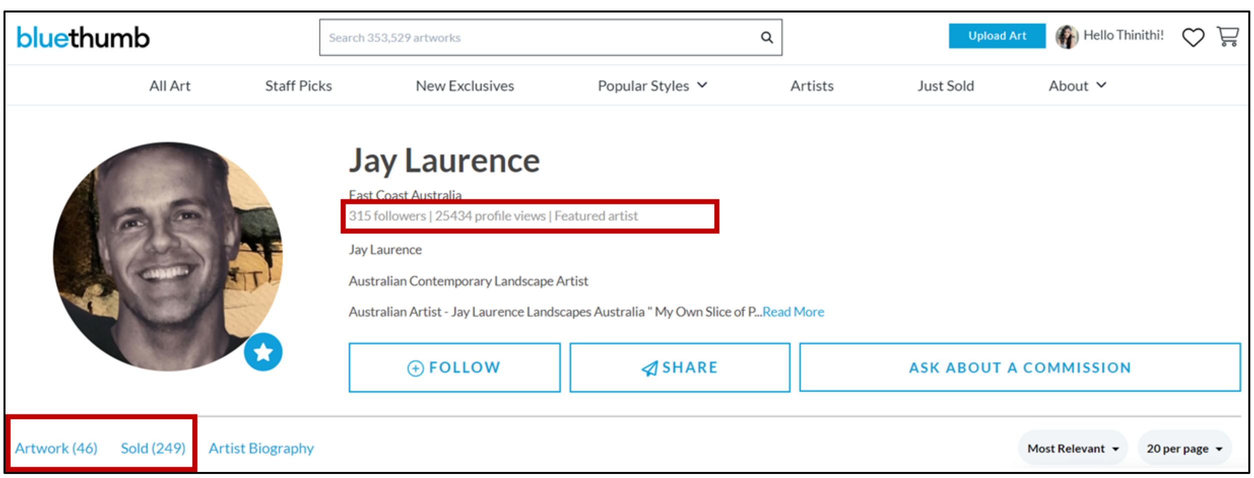


Figure 1.5.3: Artist profile once the relevant artist name has been clicked on(Red signifies data to be extracted)

Source: https://bluethumb.com.au/?gclid=Cj0KCQjw1aOpBhCOARIsACXYv-cuMsAMRdOqb9kvzg7I34NS4fvIKS3wMTdO-UitwVTrS3HuWELHtLkaArLJEALw_wcB

The data is extracted in three stages to speed up the process. Initially, the artwork url of all the artworks in the feed page is obtained after which, the workflow in the scraper is set to visit each artwork webpage shown in figure 1.52 and extract the relevant data. Finally, the url of each artwork is trimmed to obtain the artist page link from the extracted data and is fed into the scraper to visit each artist profile and obtain the data shown in figure 1.5.3, after which it is stored in *Artist_profile_data.csv*. The three datasets are compiled into a single dataset named *final_data.csv via excel* and loaded into R Studio for data pre-processing and analysis. The octoparse workflows for each stage are shown below.

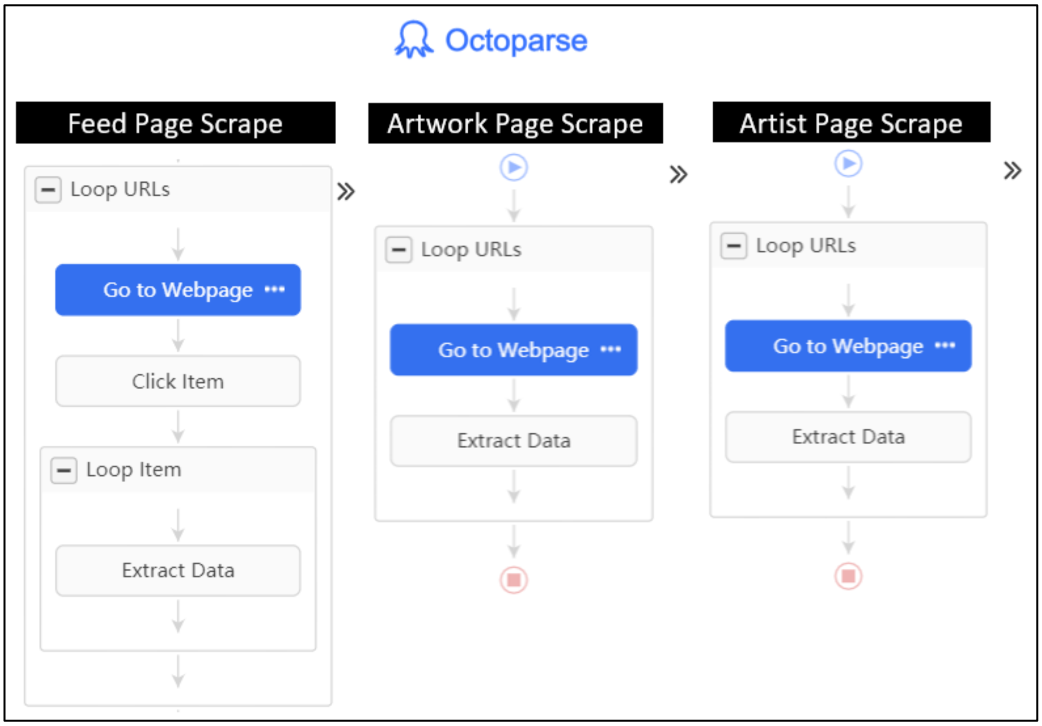
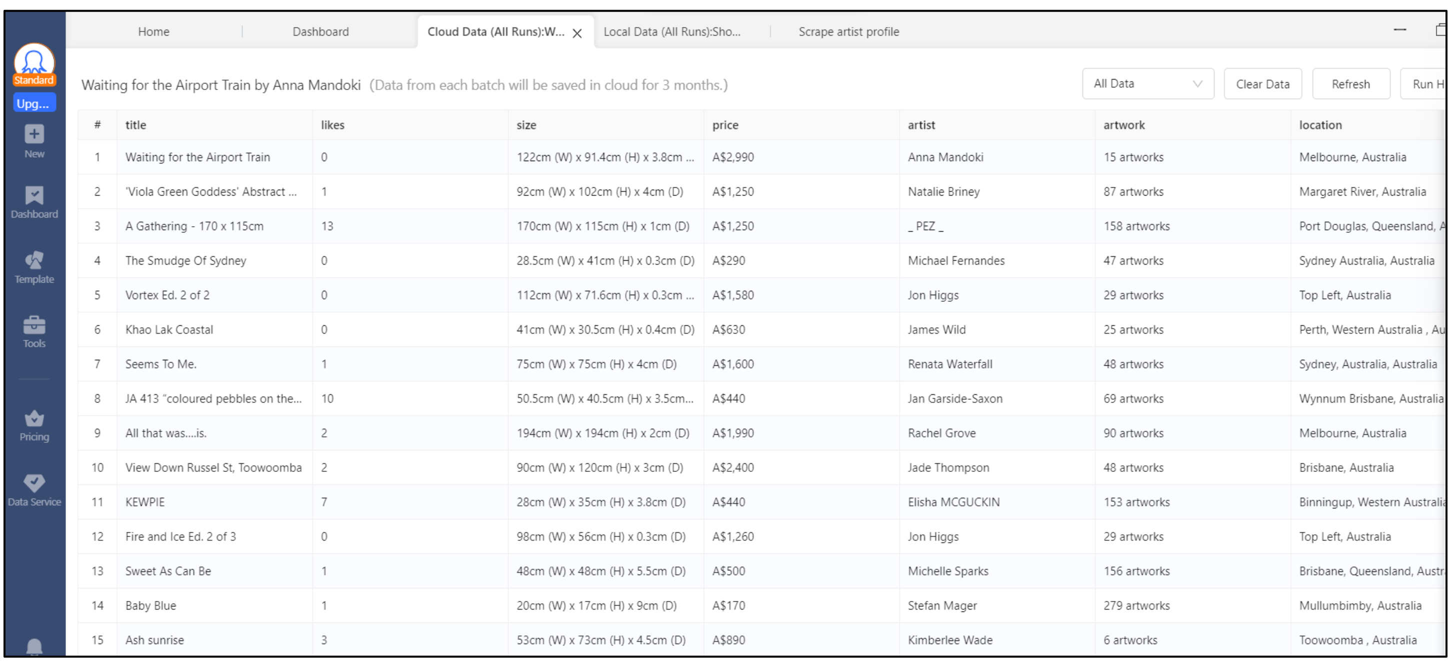


Figure 1.5.4: Octoparse work flows for each stage



#	title	likes	size	price	artist	artwork	location
1	Waiting for the Airport Train	0	122cm (W) x 91.4cm (H) x 3.8cm ...	A$2,990	Anna Mandoki	15 artworks	Melbourne, Australia
2	'Viola Green Goddess' Abstract ...	1	92cm (W) x 102cm (H) x 4cm (D)	A$1,250	Natalie Briney	87 artworks	Margaret River, Australia
3	A Gathering - 170 x 115cm	13	170cm (W) x 115cm (H) x 1cm (D)	A$1,250	_ PEZ _	158 artworks	Port Douglas, Queensland, A
4	The Smudge Of Sydney	0	28.5cm (W) x 41cm (H) x 0.3cm (D)	A$290	Michael Fernandes	47 artworks	Sydney Australia, Australia
5	Vortex Ed. 2 of 2	0	112cm (W) x 71.6cm (H) x 0.3cm ...	A$1,580	Jon Higgs	29 artworks	Top Left, Australia
6	Khao Lak Coastal	0	41cm (W) x 30.5cm (H) x 0.4cm (D)	A$630	James Wild	25 artworks	Perth, Western Australia , Au
7	Seems To Me.	1	75cm (W) x 75cm (H) x 4cm (D)	A$1,600	Renata Waterfall	48 artworks	Sydney, Australia, Australia
8	JA 413 "coloured pebbles on the...	10	50.5cm (W) x 40.5cm (H) x 3.5cm...	A$440	Jan Garside-Saxon	69 artworks	Wynnum Brisbane, Australia
9	All that was....is.	2	194cm (W) x 194cm (H) x 2cm (D)	A$1,990	Rachel Grove	90 artworks	Melbourne, Australia
10	View Down Russel St, Toowoomba	2	90cm (W) x 120cm (H) x 3cm (D)	A$2,400	Jade Thompson	48 artworks	Brisbane, Australia
11	KEWPIE	7	28cm (W) x 35cm (H) x 3.8cm (D)	A$440	Elisha MCGUCKIN	153 artworks	Binningup, Western Australia
12	Fire and Ice Ed. 2 of 3	0	98cm (W) x 56cm (H) x 0.3cm (D)	A$1,260	Jon Higgs	29 artworks	Top Left, Australia
13	Sweet As Can Be	1	48cm (W) x 48cm (H) x 5.5cm (D)	A$500	Michelle Sparks	156 artworks	Brisbane, Queensland, Austr
14	Baby Blue	1	20cm (W) x 17cm (H) x 9cm (D)	A$170	Stefan Mager	279 artworks	Mullumbimby, Australia
15	Ash sunrise	3	53cm (W) x 73cm (H) x 4.5cm (D)	A$890	Kimberlee Wade	6 artworks	Toowoomba , Australia

Figure 1.5.5: Sample Octoparse extracted data for stage 2

1.6 Data characteristics and processing

After the dataset was imported into R, the structure of the dataset was viewed via the glimpse() function . There are total of 17 columns in the dataset and 8,714 rows of unique artwork from which 253 are sold by now (Approx 3% conversion rate)

Column Name	Description
title	Artwork title
likes	Artwork likes count
size	Artwork dmiensions (Height x Width x Diameter)
price	Artwork price
artist	Artist name
artwork	Total count of artworks done by artist
location	Artist city location
medium	Artwork medium used(Ex: Acryllic, canvas, board..etc)
hand	Ready to hang, not framed..etc
sold_tag	"add_to_cart" or "sold"
sold	Binary 1-sold,0-add_to_card (Not sold)
description	Artwork category and description hashtags
page_url	Artwork page url
artist_url	Artist profile url
follow	Artist follower count
featured_artist	Artist Bluethumb recongition status
artwork_sold	Total artworks sold

Figure 1.6.1: Data set column description

```
> glimpse(data)
Rows: 8,714
Columns: 17
$ title           <chr> "Waiting for the Airport Train", "'Viola Green Goddess' Abstract Green Pansy Flower", "A Gathering - 170 x 11…
$ likes           <dbl> 0, 1, 13, 0, 0, 0, 1, 10, 2, 2, 7, 0, 1, 3, 0, 0, 0, 4, 0, 1, 13, 7, 0, 4, 0, 2, 6, 1, 14, 8, 16, 0, 1, 0, 0,…
$ size            <chr> "122cm (W) x 91.4cm (H) x 3.8cm (D)", "92cm (W) x 102cm (H) x 4cm (D)", "170cm (W) x 115cm (H) x 1cm (D)", "2…
$ price           <chr> "A$2,990", "A$1,250", "A$1,250", "A$290", "A$1,580", "A$630", "A$1,600", "A$440", "A$1,990", "A$2,400", "A$44…
$ artist          <chr> "Anna Mandoki", "Natalie Briney", "_ PEZ _", "Michael Fernandes", "Jon Higgs", "James Wild", "Renata Waterfal…
$ artwork         <chr> "15 artworks", "87 artworks", "158 artworks", "47 artworks", "29 artworks", "25 artworks", "48 artworks", "69…
$ location        <chr> "Melbourne, Australia", "Margaret River, Australia", "Port Douglas, Queensland, Australia", "Sydney Australia…
$ medium          <chr> "Oil, acrylic, soil, bitumen and image transfer on canvas", "Acrylic paint, moulding and crackle paste", "Lar…
$ hang            <chr> "This artwork is currently stretched and ready to hang.", "This artwork is currently stretched and ready to h…
$ sold_tag        <chr> "free shipping and returns  A$2,990  Add to Cart", "free shipping and returns  A$1,250  Add to Cart", "free s…
$ sold            <chr> "add_to_cart", "add_to_cart", "add_to_cart", "add_to_cart", "add_to_cart", "add_to_cart", "add_to_cart", "add…
$ description     <chr> "POLITICAL ART, PEOPLE & PORTRAIT ART, BIRD ART\n#birds, #people, #group, #suitcase, #travel, #texture, #ligh…
$ Page_URL        <chr> "https://bluethumb.com.au/anna-mandoki/Artwork/waiting-for-the-airport-train", "https://bluethumb.com.au/nata…
$ artist_url      <chr> "https://bluethumb.com.au/anna-mandoki", "https://bluethumb.com.au/natalie-briney", "https://bluethumb.com.au…
$ follow          <chr> "9 followers | 1390 profile views", "10 followers | 1390 profile views", "11 followers | 1390 profile views",…
$ featured_artist <chr> "featured-ico.6af4c6f5.svg", "featured-ico.6af4c6f5.svg", "established-ico.205c386c.svg", "0", "photograph-ic…
$ artwork_sold    <chr> "Sold (1)", "Sold (37)", "Sold (98)", "Sold (11)", "Artist Biography", "Sold (9)", "Sold (15)", "Sold (13)", …
```

Figure 1.6.2: Data structure overview

The table below summarizes the characteristics of the data and the technologies that will be utilized to process and store it.

Data Characteristics	**Data Source**	Bluethumb artist profiles, artwork information and artwork sale information on the webpage.
	Data Type	Mainly categorical variables and some numeric data in string format and stored in csv files.
	Volume	The company has sold over 90,000 artworks for first time buyers and having over 20,000 emerging and established artists. If 90000x5imagesx5mb data = 900,000,000 (880GB). So terabytes would be a reasonable estimate.
	Velocity	Artwork statuses update daily when sold. Artist followers, profile views and likes change daily, which means the data for the project needs to be refreshed regularly to be on top of the market trends.
	Variety	Data is combination of artwork data and artist profile information with data formats ranging from double, to categorical to string including hashtags describing each artwork.
	Variability	Artist tends to change their style of art, topics and level of engagement with the online platform.
	Veracity	There may be a handful of artists who have sold to people they know in order improve their profile status. But majority would include data truthfully.
Data Processing	**Visualization**	Visualization will be extensively used in the exploratory data analysis component not so much on the artwork score model output. R visualization corporate to the webpage
	System	Amazon EC2, a scalable and inexpensive computing capacity in cloud.
	Storage	HDFS
	Networking	Good networking to download data and process. The computed score could be incorporated into the artwork itself which is only visible to the artist itself.
	Software	In the prototype the development is done entirely on R studio and by using Octoparse. Within the company python, R or Apache Spark could be used.

1.7 Data pre-processing

R is utilized to clean the data starting with splitting the data columns with multiple information ex: dimension column into width, height and diameter and calculating area. Followed by removing the unnecessary text within the data columns extracted and converting numeric string values into numeric values.

Cleansing Activity	Description
Column separation into sub columns	1) split “size” column into separate dimensions. 2) split “follow” into followers and profile_views 3) split “featured” column into status and image code. 4) split “location” into city location and other information. 5) split “description” into category and hashtags
Data cleaning using regex pattern identification	Use R function str_remove() with the pattern functionality where regex is used to remove unnecessary text within each column data. Ex: “A$” in price, sold(41) in artworks sold count..etc
Drop unnecessary columns	Such as columns named feature status image code, page_url, other information of city location and artist_url
Identifying and treating missing values	1) artwork_sold “NA” signifies no sales, replace with zero 2) followers “NA” signifies no followers, replace with zero 3) profile view “NA” signifies no views, replace with zero 4) status view “NA” signifies artist without special status, replace with zero
Convert “sold” column to binary	“sold” is indicated by 1 and “add_to_cart” is indicated by 0
Outlier Removal	To further analyze price and its relationship with other attribute, its extreme values are removed using “IQR Outlier Removal” approach.

Figure 1.7.1: Data Description

Missing value Treatment Before and After

The “NA” within the “artwork_sold” were successfully replaced by zero as it signified no artwork sale at the date of the analysis.

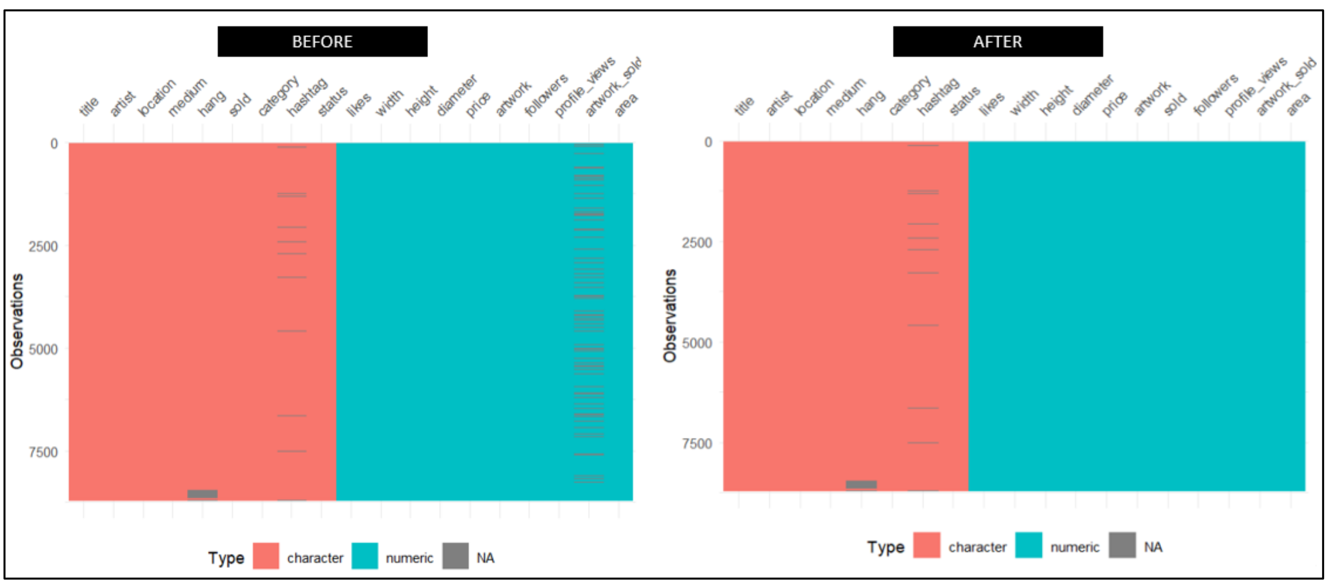


Figure 1.7.1: Data structure overview

Outlier Treatment Before and After

Outliers need to be treated to avoid the statistical analysis and machine learning algorithm being affected negatively and resulting in inaccurate results. The box plot below reveals the outliers in the numeric columns.

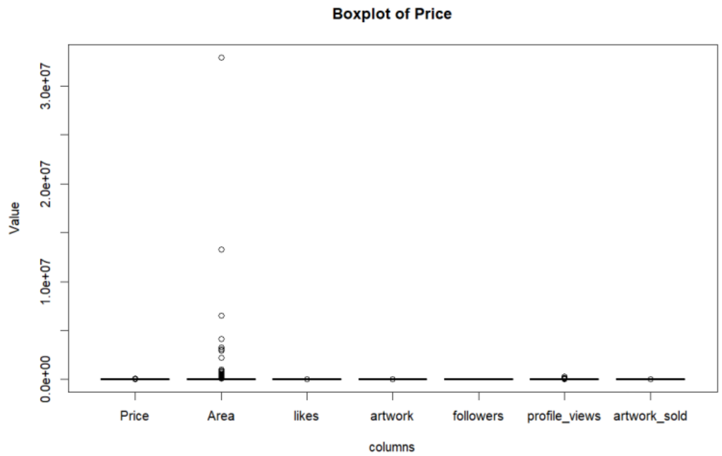


Figure 1.7.2: Data structure overview

The relationship between the size of the artwork and price needs to be analyzed via a scatterplot. Outlier removal via IQR approach was carried out resulting in the improved scatterplot in the right hand side.

Figure 1.7.3: price vs artwork area for sold and not sold artworks before and after outlier treatment.

1.8 Data Analysis

Pricing Insights: How to price artwork?

The correlation matrix was derived to better understand the relationship between the price and other attributes.

```
> print(correlation_matrix)
                     price       area      height        width artwork_sold    followers profile_views       likes
price          1.000000000 0.64799132 0.61046742 0.602554969   0.03385989  0.020284306  -0.001025278  0.07760851
area           0.647991316 1.00000000 0.89376734 0.908309380   0.16601886  0.005195910   0.055190910  0.05818019
height         0.610467420 0.89376734 1.00000000 0.704326149   0.11068518  0.018278126   0.033381663  0.05433367
width          0.602554969 0.90830938 0.70432615 1.000000000   0.16311725  0.008518096   0.048208127  0.05752228
artwork_sold   0.033859893 0.16601886 0.11068518 0.163117255   1.00000000 -0.022480007   0.165527466  0.11761012
followers      0.020284306 0.00519591 0.01827813 0.008518096  -0.02248001  1.000000000  -0.090547033 -0.01017243
profile_views -0.001025278 0.05519091 0.03338166 0.048208127   0.16552747 -0.090547033   1.000000000  0.01479253
likes          0.077608513 0.05818019 0.05433367 0.057522276   0.11761012 -0.010172434   0.014792530  1.00000000
```

Figure 1.8.1: Correlation matrix

The highest correlation appears to be between price and area which is 0.65, which signifies a strong relationship between the two variables that's highlighted by the scatterplot. The best pricing reference for an artist can be the price vs artwork area shown below.

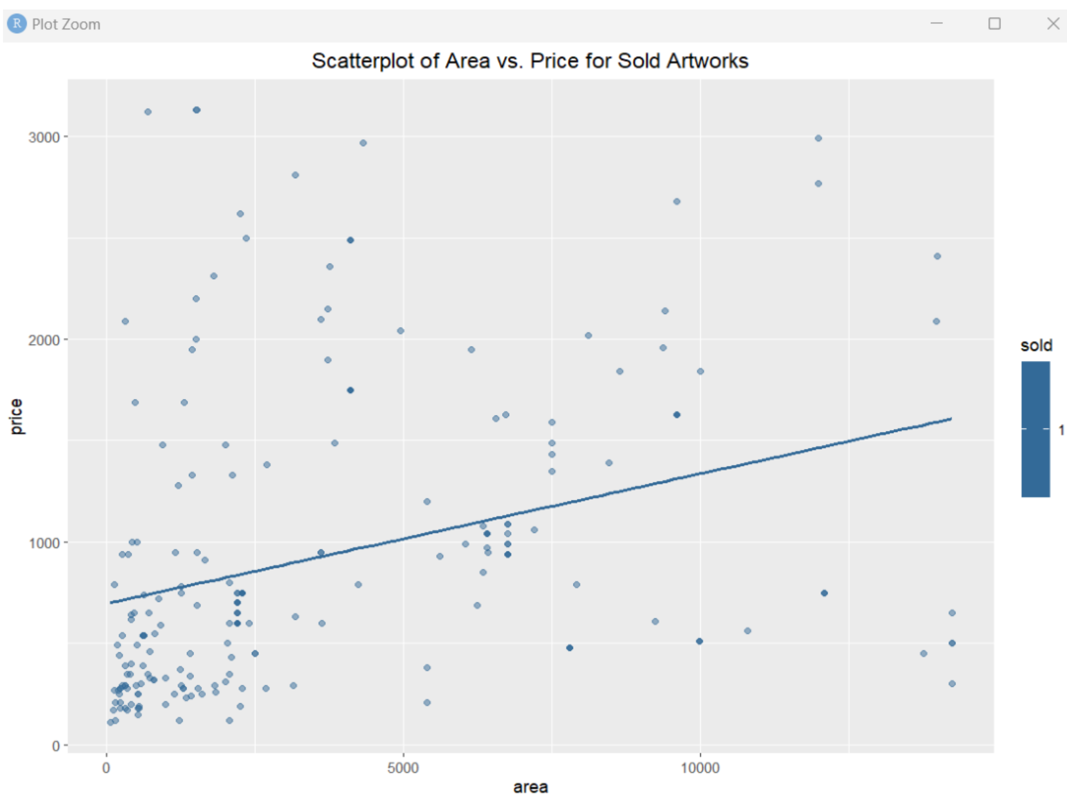


Figure 1.8.2: Price Vs Artwork area scatterplot for artworks sold

Market Trends: What artworks mostly sell?

Word clouds are effective visual representation of text data that allows the user to understand which words are mostly used. This can be applied in this project to understand the artwork categories done by artists and the categories which have sold.

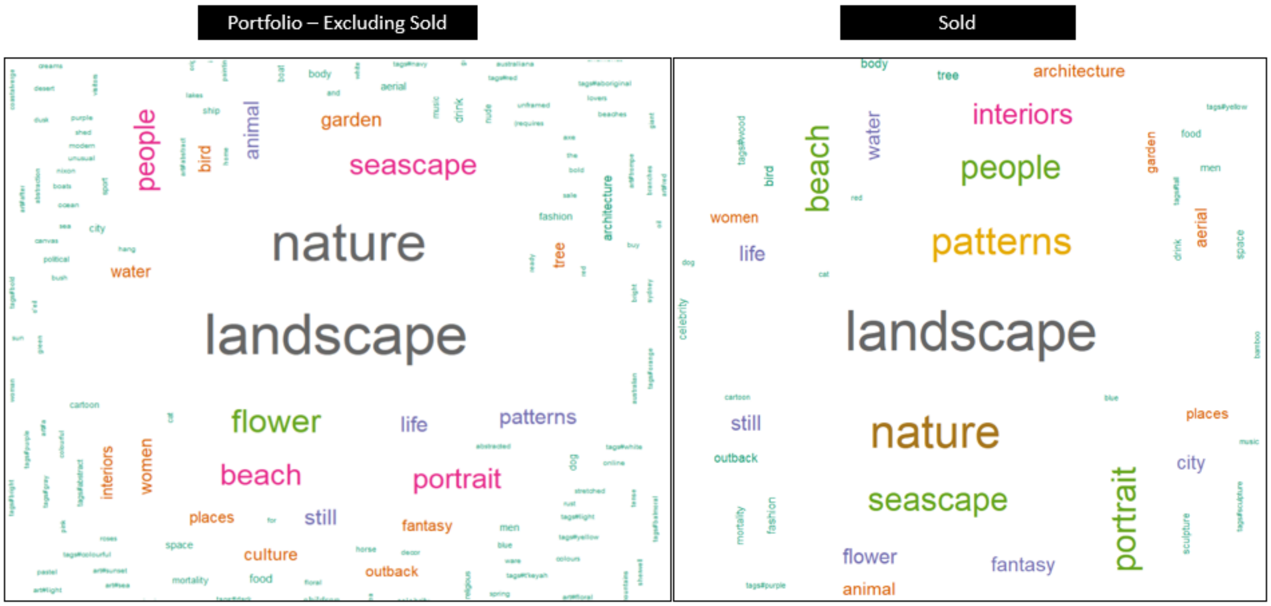


Figure 1.8.3: Word cloud of artwork description for those sold and haven't

```
> print(subset_df_sorted)
           word freq.x freq.y    percent
11         city     15    148 0.101351351
21    interiors     26    377 0.068965517
1        aerial     11    179 0.061452514
30     patterns     39    649 0.060092450
3  architecture     13    252 0.051587302
16      fashion      6    132 0.045454545
37        space      5    116 0.043103448
15      fantasy     16    402 0.039800995
7          body      7    179 0.039106145
44        water     16    441 0.036281179
31       people     33    979 0.033707865
34     portrait     33    982 0.033604888
36    seascape     31   1020 0.030392157
4         beach     31   1023 0.030303030
24          men      4    146 0.027397260
22    landscape     59   2294 0.025719268
12      culture     12    471 0.025477707
```

Figure 1.8.4: Conversion rate for art categories (Freq.x = sold, Freq.y=total)

Though some topics maybe be preferred by artists due to their assumption that it is the market demand, in reality that might not be the case. For instance, landscape is one of the most popular topics but when looking at the conversion rate, it's merely 2.6% whilst city artworks have a 10% conversion rate.

Insights on necessities: Do more prepared artwork sell better?

Acrylic, oil and watercolors are some of the most popular mediums but when compared with artworks that are selling, watercolor artwork are not as preferred. More diverse forms of artwork are selling rather than the common mediums.

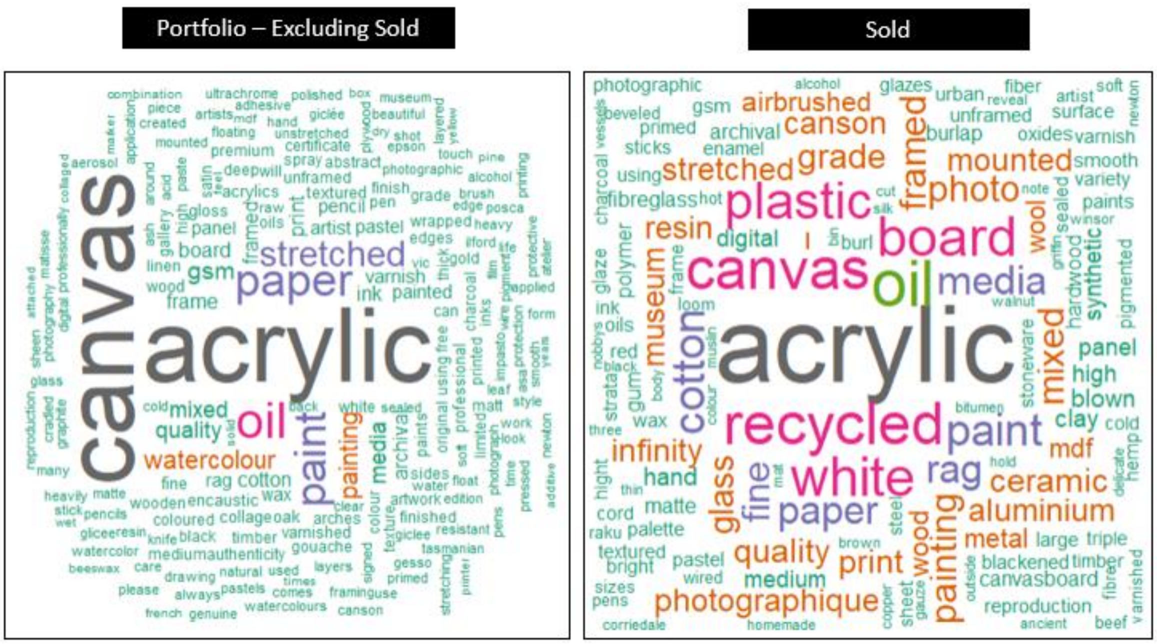


Figure 1.8.5: Word cloud of artwork medium for those sold and haven't

```
> print(subset_df_sorted)
          word freq.x freq.y     percent
20       board    299     25 0.083612040
151        rag    213     16 0.075117371
37      cotton    320     20 0.062500000
55      framed    338     14 0.041420118
102      media    424     17 0.040094340
144      print    359     12 0.033426184
105      mixed    439     14 0.031890661
149    quality    407     12 0.029484029
121   painting    517     14 0.027079304
114        oil   1518     31 0.020421607
128     pastel    212      4 0.018867925
195        wax    215      4 0.018604651
119      paint   1289     20 0.015515904
2      acrylic   3674     56 0.015242243
7     archival    337      5 0.014836795
54       frame    322      4 0.012422360
173  stretched    936     11 0.011752137
127      paper   1363     16 0.011738811
190    varnish    271      3 0.011070111
88         ink    338      3 0.008875740
70         gsm    459      4 0.008714597
8       artist    248      2 0.008064516
25      canvas   3528     27 0.007653061
120    painted    266      2 0.007518797
194 watercolour    545      3 0.005504587
```

Figure 1.8.6: Conversion rate for art mediums (Freq.y = sold, Freq.x=total)

There seems to be more preference for diverse forms of mediums and as highlighted framed artwork have a higher conversion that those which are done on the canvas only. These insights could help improve the end products an artist delivers.

Artwork Growth Score: Where do my artwork stand compared to those that sell?

In order to generate a growth score for each artwork a logistics regression is used for modelling. The target variable being whether the artwork will sell or not and the numeric columns are used as exploratory. Each column is scaled to 1-0 due to the data set having volatile and diver sizes of data. The data set is spilt into a 80% train and 20% test set and the model is utilized to obtain prediction for the test set along with the probability for each one to sell.

```
R 4.3.1 · C:/Users/Thinithi/Monash/Sem2_2023/FIT5145_INTRO_TO_DS/Assignment/Bluethumb Data/
> print(conf_matrix)
          Truth
Prediction    0    1
         0 1439   18
         1    3   25
>
> #calculate accuracy
> accuracy(results, truth = sold, estimate = .pred_class)
# A tibble: 1 × 3
  .metric  .estimator .estimate
  <chr>    <chr>          <dbl>
1 accuracy binary         0.986
>
> # Calculate precision
> precision <- 25/ (25+3)
> print(precision)
[1] 0.8928571
>
> # Calculate precision
> recall <- 25/(25+18)
> print(recall)
[1] 0.5813953
```

Figure 1.8.7: Logistic regression model performance

	hashtag	followers	profile_views	status	artwork_sold	area	artwork_growth
T ART	Portrait, realism, cupcake, candle, crown, jewel, blush, black	6	1670	featured	4	2349.00	Level_03
E ART	seascape, morning, whitewash, beach, pink, sky, waves, brea...	9	2171	0	7	2679.00	Level_02
LL LIFE ART, PATTERNS ART	stilllife, oranges, chineseporcelain, blueandorange, walnuts, ...	53	984	0	1	1520.00	Level_02
T ART	josstacey, paintandpony, joy, freestyle, doobie, peace, oils, a...	119	6318	featured	14	2071.63	Level_03
MAL ART	dynamic, icon, auspicious, animal, Sculpture, dynamic anima...	549	17768	featured	206	340.00	Level_05
& SEASCAPE ART, LANDSCAPE ART	Earth, Autumn, Shapes, Rocks, Leaves, Butterflies, orange, pi...	1213	36282	rising	235	6552.00	Level_10
RNS ART, FLOWER ART	Violet, hard edge abstract, burlap, hessian, Australian art, Bl...	638	23197	rising	91	13984.00	Level_07

Figure 1.8.8: The artwork growth level compared is derived for each artwork

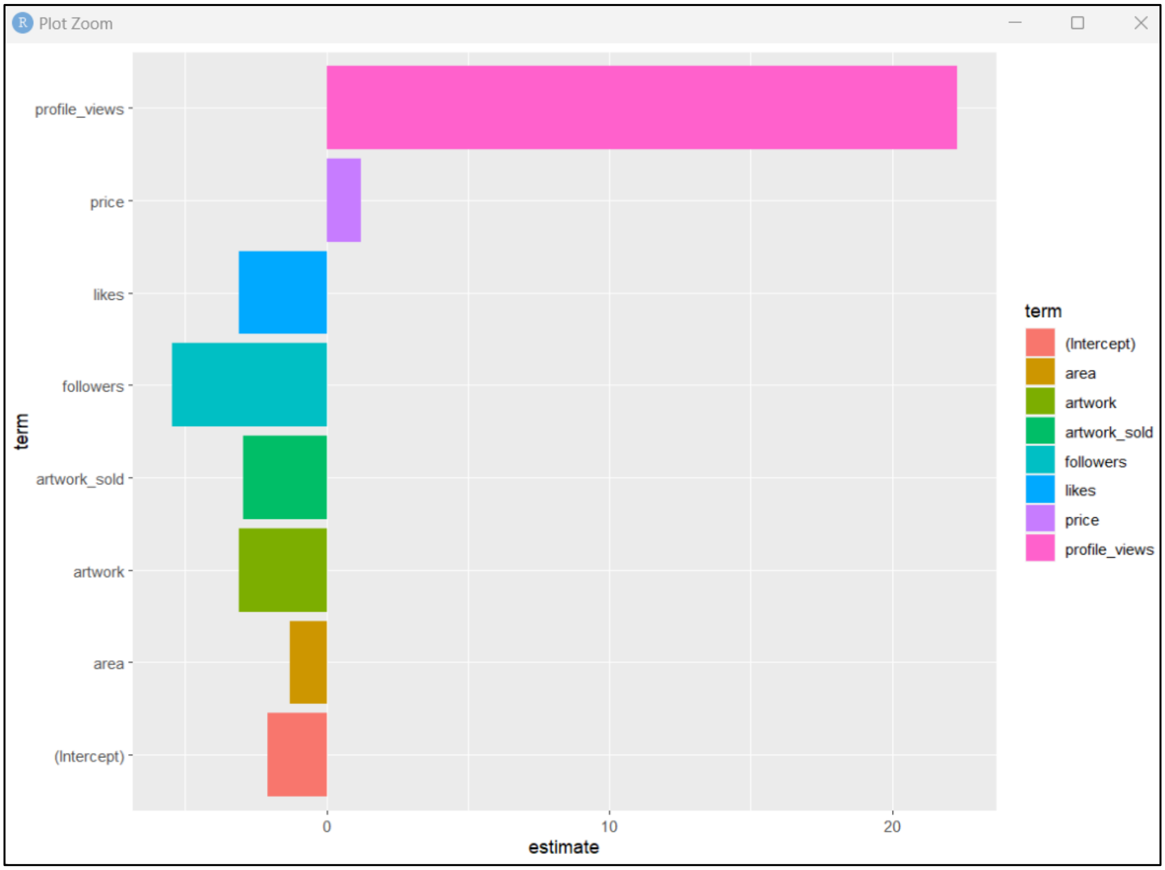


Figure 1.8.9: Attributes sorted by the coefficient value (Order of importance top-bottom)

1.9 Data governance and management

The data governance team of Bluethumb needs to be brought in to ensure the project initiatives and plan align governance rules and regulations. First aspect being regarding data security, to ensure data is safely utilized and stored. Ethics aspect should also be taken into account when utilizing the customer and artist base to ensure the data will only be sued to provide a service to these stakeholders and this should be clearly communicated. All data should be backed up and audited regularly to maintain all practices adhere to the best practices and regulations on data governance.

2.0 Conclusion

As highlighted in the analysis above, Bluethumb project “Canvas success” needs to be implemented to ensure on this day and age data science mechanisms are incorporated into supporting the art industry. Bluethumb being an industry leader could be the first to open the pathway to a data driven platform that supports the art community and create an Bluethumb journey for each artist to grow along. Increasing the company’s CSR and user engagement by atleast 10% (Assuming 1% increase in Conversion rate) would mean an annual profit increase by AUD $645,000 where this project can be build with the existing resources, hence there is barely an investment.

2.1 References

About Us. Bluethumb Online Art Gallery Blog RSS2. (n.d.). https://bluethumb.com.au/about

LÃ¶vdÃ©n, Martin; GarzÃ³n, BenjamÃ-n; Lindenberger, Ulman (2020). Human Skill Learning: Expansion, Exploration, Selection, and Refinement. Current Opinion in Behavioral Sciences, 36(), 163–168. doi:10.1016/j.cobeha.2020.11.002

Chugh, V. (2023, March 17). Logistic regression in R tutorial. DataCamp. https://www.datacamp.com/tutorial/logistic-regression-R

Growjo. (n.d.). Bluethumb revenue and competitors. Bluethumb: Revenue, Competitors, Alternatives. https://growjo.com/company/Bluethumb

Rul, C. V. den. (2019, October 20). How to generate word clouds in R. Medium. https://towardsdatascience.com/create-a-word-cloud-with-r-bde3e7422e8a