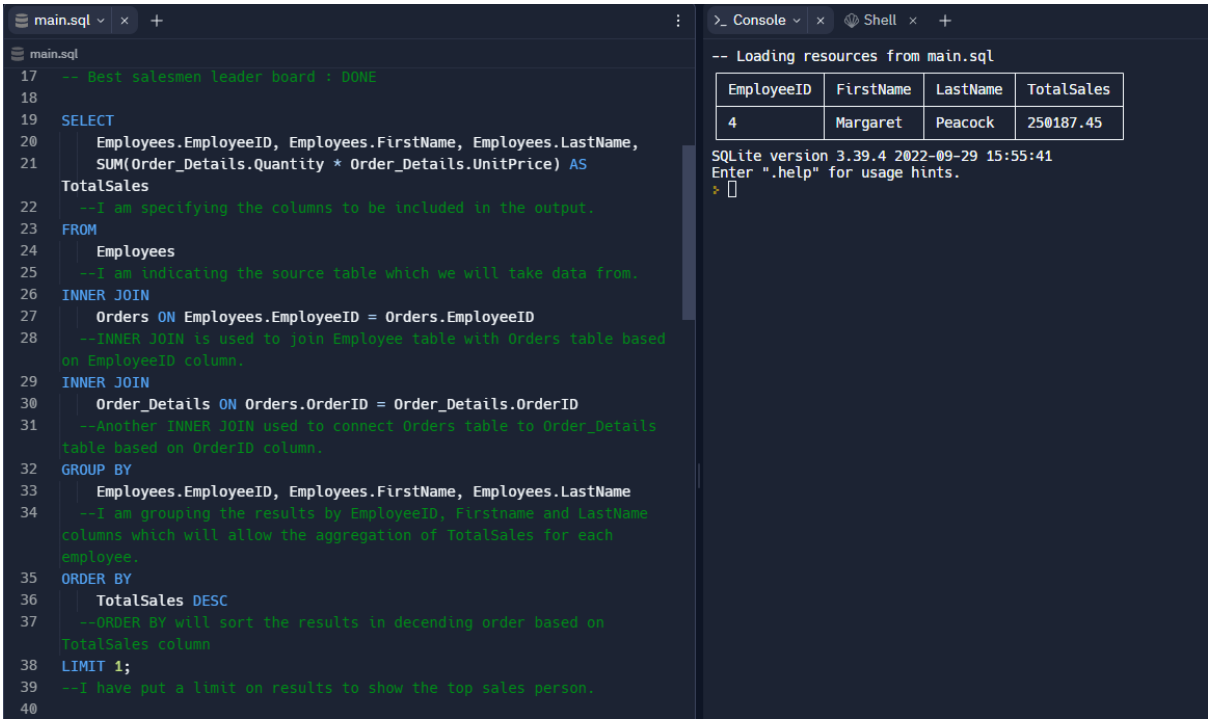
Data Analysis Concepts Assignment

Part 1 (50 marks):

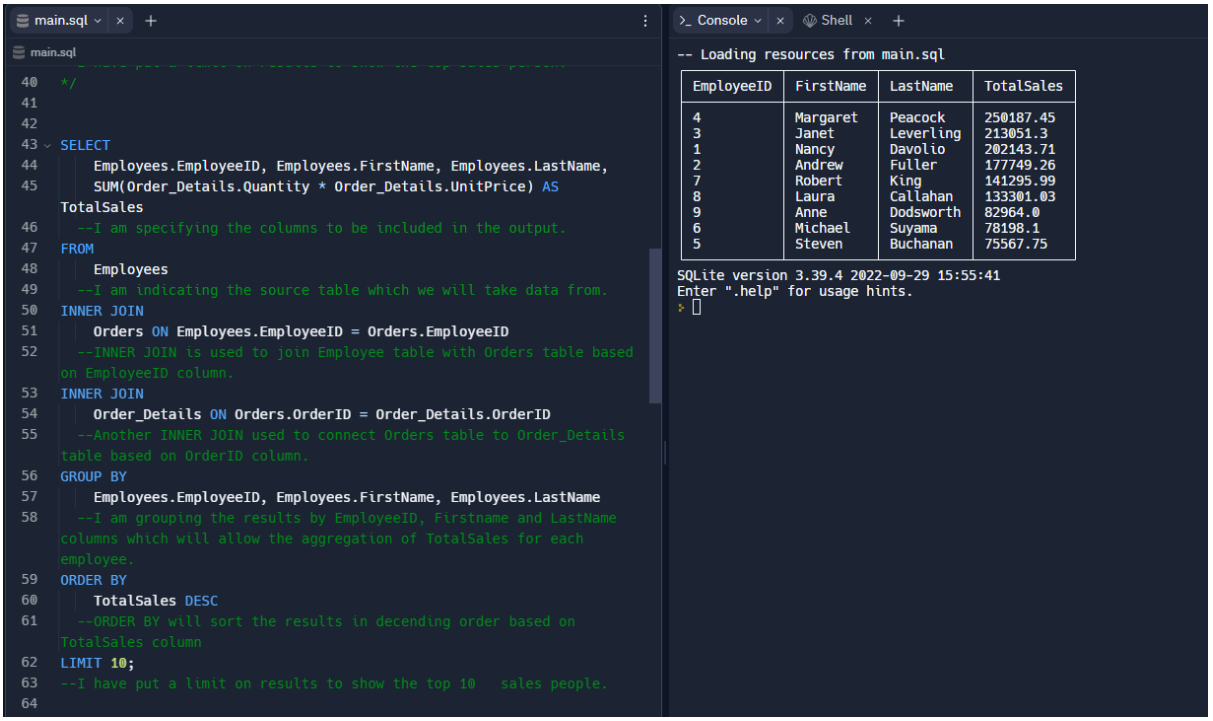
The Northwind database is a sample database that was originally created by Microsoft. It contains the sales data for a fictitious company called “Northwind Traders,” who import and export specialty foods from around the world.

Explanations of the SQL code are within the code themselves as the green comments and are not in the writing of this document

Best salesmen leaderboard (you can compare them by sales, territories and both)



(SQL code generating the TOP sales person)



```
main.sql
*/

SELECT
    Employees.EmployeeID, Employees.FirstName, Employees.LastName,
    SUM(Order_Details.Quantity * Order_Details.UnitPrice) AS
TotalSales
  --I am specifying the columns to be included in the output.
FROM
    Employees
  --I am indicating the source table which we will take data from.
INNER JOIN
    Orders ON Employees.EmployeeID = Orders.EmployeeID
  --INNER JOIN is used to join Employee table with Orders table based
on EmployeeID column.
INNER JOIN
    Order_Details ON Orders.OrderID = Order_Details.OrderID
  --Another INNER JOIN used to connect Orders table to Order_Details
table based on OrderID column.
GROUP BY
    Employees.EmployeeID, Employees.FirstName, Employees.LastName
  --I am grouping the results by EmployeeID, Firstname and LastName
columns which will allow the aggregation of TotalSales for each
employee.
ORDER BY
    TotalSales DESC
  --ORDER BY will sort the results in decending order based on
TotalSales column
LIMIT 10;
--I have put a limit on results to show the top 10   sales people.
```

```
-- Loading resources from main.sql
```

EmployeeID	FirstName	LastName	TotalSales
4	Margaret	Peacock	250187.45
3	Janet	Leverling	213051.3
1	Nancy	Davolio	202143.71
2	Andrew	Fuller	177749.26
7	Robert	King	141295.99
8	Laura	Callahan	133301.03
9	Anne	Dodsworth	82964.0
6	Michael	Suyama	78198.1
5	Steven	Buchanan	75567.75

```
SQLite version 3.39.4 2022-09-29 15:55:41
Enter ".help" for usage hints.
```

(SQL code generating the TOP 10 salespeople)

The EmployeeID, FirstName and LastName columns in the Employees table serve as the nominal data and categorical data since they are identifiers and labels for the employees and allow the employees to be categorised by these labels. I have deduced that none of the columns exhibits any ordinal data because there is no data that represents the data with a meaningful order or ranking. Similar to Ordinal, I believe there are no columns in this code that exhibit interval data since there is no data that measures the difference between values. We can consider the TotalSales column to include ratio data because it stores a quantitative measurement of sales data which has the potential to be used for meaningful mathematical operations.

The TotalSales column represents discrete data because it is a countable quantity (We can do this by using the sum of the quantities multiplied by the unit price). I believe that there is no representation of continuous data because there is no data present which can be infinitely divided, suggesting a lack of continuous data

Output and its meaningfulness:

The code outputs the EmployeeID, FirstName, LastName and TotalSales for the employees who have the highest sales or the top 10 employees with the highest sales. This makes the output meaningful because as it allows for recognition and motivation, Benchmarking and goal setting, performance evaluation and development.

An invoice

main.sql

```sql
-- An invoice : DONE

SELECT
    Orders.OrderID, Orders.OrderDate,
    Customers.CompanyName, Customers.ContactName, Customers.Address, Customers.City,
    Customers.PostalCode, Products.ProductName, Order_Details.UnitPrice,
    Order_Details.Quantity,
    (Order_Details.UnitPrice * Order_Details.Quantity) AS TotalPrice
    --The SELECT clause specifies that we will be including data orders, orderdetails, products and customers tables.
FROM
    Orders
  --The FROM statement specifies that we will be taking data from the orders table.
INNER JOIN
    Customers ON Orders.CustomerID = Customers.CustomerID
  --This first INNER JOIN is used to join the oreders table with the customers table using the CustomerID column.
INNER JOIN
    Order_Details ON Orders.OrderID = Order_Details.OrderID
  --This second INNER JOIN will be used to join the orderdetails table based on the OrderID column.
INNER JOIN
    Products ON Order_Details.ProductID = Products.ProductID
  --This last INNER join will be used to join the products table using the ProductID column.
WHERE
    Orders.OrderID = '10248';
--The WHERE statement allows us to add a filter to show result where the order id IS 10248.
```

Console | Shell

```
-- Loading resources from main.sql
```

OrderID	OrderDate	CompanyName	ContactName	Address	City	PostalCode	ProductName	UnitPrice	Quantity	TotalPrice
10248	1996-07-04 00:00:00.000	Vins et alcools Chevalier	Paul Henriot	59 rue de l'Abbaye	Reims	51100	Queso Cabrales	14	12	168
10248	1996-07-04 00:00:00.000	Vins et alcools Chevalier	Paul Henriot	59 rue de l'Abbaye	Reims	51100	Singaporean Hokkien Fried Mee	9.8	10	98.0
10248	1996-07-04 00:00:00.000	Vins et alcools Chevalier	Paul Henriot	59 rue de l'Abbaye	Reims	51100	Mozzarella di Giovanni	34.8	5	174.0

```
SQLite version 3.39.4 2022-09-29 15:55:41
Enter ".help" for usage hints.
```

(SQL code generating an invoice for the order with the order ID of 10248)

For the invoice the CompanyName, City, Address and PostalCode columns within the Customers table is out nominal data types as they are identifiers, I believe that there are no Interval or ordinal data types

present within this code as there is no data that has a significant ranking or any data that measures the difference between values. We can see that the Quantity, TotalPrice and UnitPrice columns all represent a ratio data type as they represent quantitative measurements that give a potential for meaningful mathematical operations.

The OrderID, OrderDate and ProductName all act as our categorical data since they categorise the orders and products. The quantity column represents our discrete data because it is a countable quantity of products that are ordered. Finally, the TotalPrice and UnitPrice columns represent our continuous data since they are values that can be infinitely divided.

Output and its meaningfulness:

This code outputs an invoice which displays the order ID, Order date, Company name of the customer, Contact name of the customer, Customer address, City and postal code of the customer, the product they ordered, how many they ordered and the total price.

For someone managing a sales team, this output can be useful in several ways:

Order-specific Details: The output provides comprehensive information about a specific order, including customer details, product details, and order-specific attributes. This can be helpful for tracking specific sales transactions and understanding the context of the order.

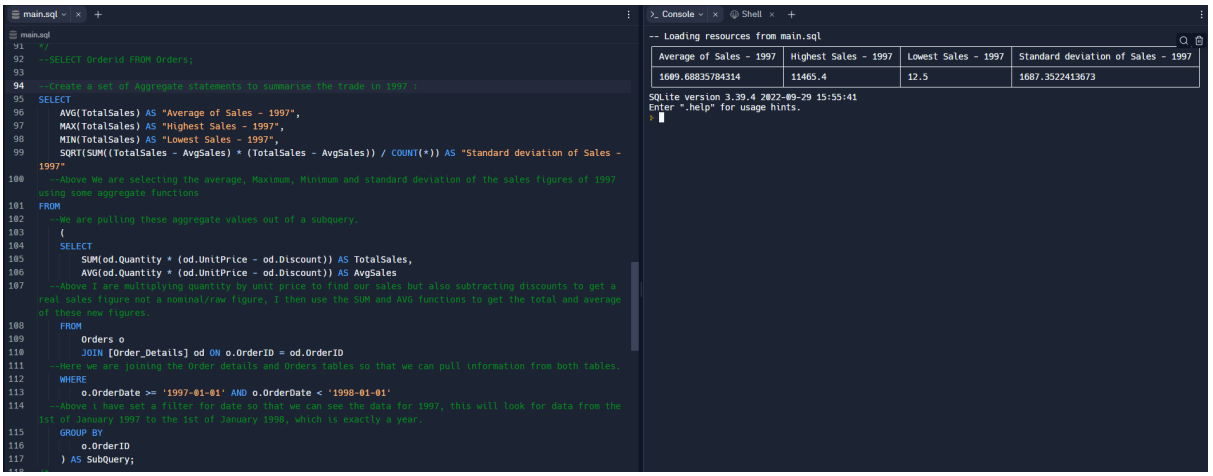
Customer Insights: The output includes customer-related information, such as company name, contact name, address, city, and postal code. This data can provide insights into customer preferences, locations, and potential customer segments.

Product Performance: The output includes the product name, unit price, and quantity ordered. This information allows sales managers to analyze

the performance of specific products, identify popular or high-demand items, and assess pricing strategies.

Sales Analysis: The output also includes the total price of the ordered product. By summing up the total prices of multiple orders, sales managers can analyze revenue, sales volume, and profitability. This information can help identify top-selling products, analyze revenue contributions from different customers, and evaluate overall sales performance.

A set of aggregate statements to summarise the trade in 1997



(SQL code generating a set of aggregate values for the sales of the year 1997)

The SQL code above does not involve any Nominal data and the data included does not refer to categories with any inherent hierarchy or order however, it does process some categorical data indirectly since it aggregates sales data based on the "Order_Details" table, such as Unit price, quantity and discount. The SQL code also doesn't make use of any Ordinal data as the data used does not represent any categories with a meaningful hierarchy or order. The SQL code doesn't include any sort of interval data since the numerics don't have a consistent interval between them i.e. the sales of each month. The data is not explicitly manipulated as ratio data which means that ratio data is not present. Still, we can understand that there would be no meaningful ratio between sales figures due to their volatile nature. Finally, I believe that there is no

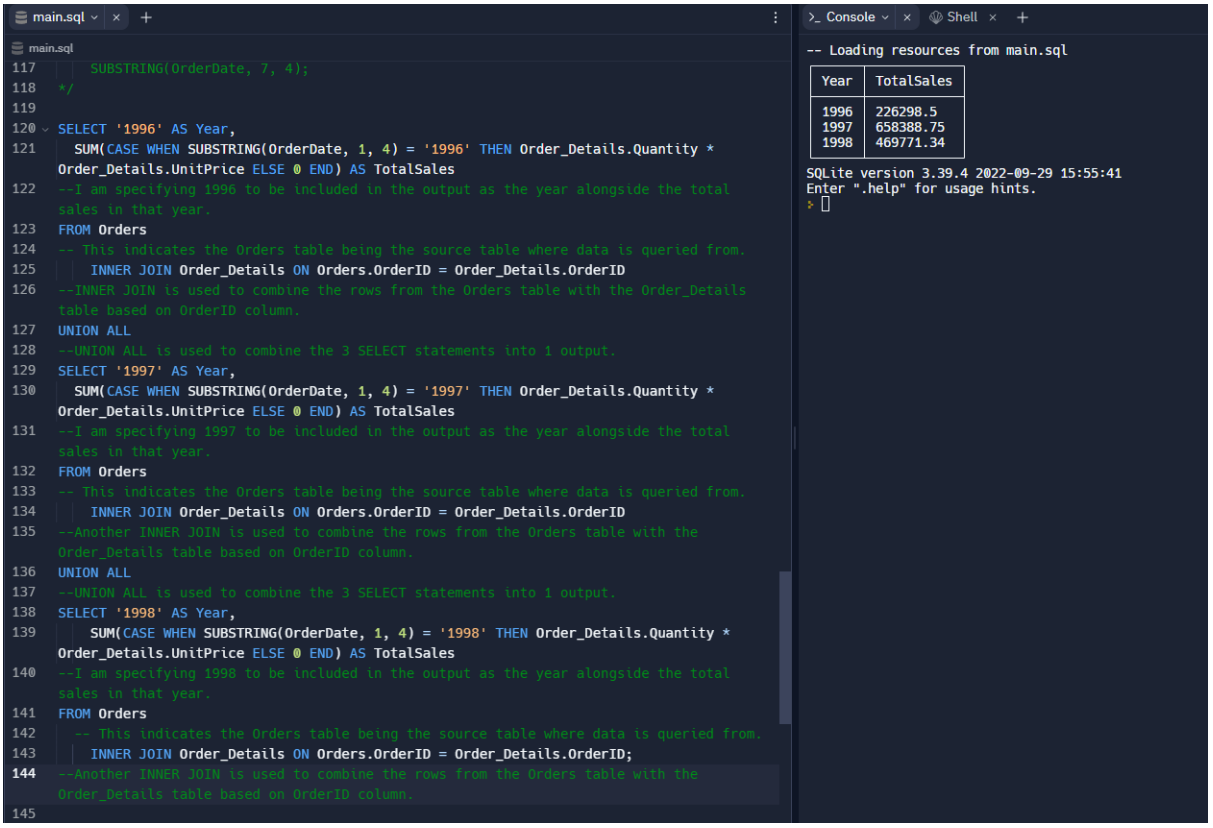
Discrete data present since there are no counts/distinct counts present however since sales figures are involved, we can say that there is continuous data involved as sales figures are an example of continuous data.

Output and its meaningfulness:

The output of this code will display the average sales during a specified time period, the highest and lowest amount of sales within this period, also the standard deviation of sales which would be the dispersion of sales amounts from the average.

For a person managing a sales team, the output of this code would be meaningful and useful. It provides statistical insights into sales performance during a specific time period. The average sales can give an overall understanding of the team's performance. The maximum and minimum sales help identify outliers or exceptional sales figures. The standard deviation indicates the variability or consistency of sales performance. These insights can assist sales managers in evaluating team performance, setting targets, identifying trends, and making informed decisions to improve sales strategies.

A summary of yearly sales



```
    SUBSTRING(OrderDate, 7, 4);
*/

SELECT '1996' AS Year,
  SUM(CASE WHEN SUBSTRING(OrderDate, 1, 4) = '1996' THEN Order_Details.Quantity *
Order_Details.UnitPrice ELSE 0 END) AS TotalSales
--I am specifying 1996 to be included in the output as the year alongside the total
sales in that year.
FROM Orders
-- This indicates the Orders table being the source table where data is queried from.
    INNER JOIN Order_Details ON Orders.OrderID = Order_Details.OrderID
--INNER JOIN is used to combine the rows from the Orders table with the Order_Details
table based on OrderID column.
UNION ALL
--UNION ALL is used to combine the 3 SELECT statements into 1 output.
SELECT '1997' AS Year,
  SUM(CASE WHEN SUBSTRING(OrderDate, 1, 4) = '1997' THEN Order_Details.Quantity *
Order_Details.UnitPrice ELSE 0 END) AS TotalSales
--I am specifying 1997 to be included in the output as the year alongside the total
sales in that year.
FROM Orders
-- This indicates the Orders table being the source table where data is queried from.
    INNER JOIN Order_Details ON Orders.OrderID = Order_Details.OrderID
--Another INNER JOIN is used to combine the rows from the Orders table with the
Order_Details table based on OrderID column.
UNION ALL
--UNION ALL is used to combine the 3 SELECT statements into 1 output.
SELECT '1998' AS Year,
    SUM(CASE WHEN SUBSTRING(OrderDate, 1, 4) = '1998' THEN Order_Details.Quantity *
Order_Details.UnitPrice ELSE 0 END) AS TotalSales
--I am specifying 1998 to be included in the output as the year alongside the total
sales in that year.
FROM Orders
  -- This indicates the Orders table being the source table where data is queried from.
    INNER JOIN Order_Details ON Orders.OrderID = Order_Details.OrderID;
--Another INNER JOIN is used to combine the rows from the Orders table with the
Order_Details table based on OrderID column.
```

```
-- Loading resources from main.sql
```

Year	TotalSales
1996	226298.5
1997	658388.75
1998	469771.34

```
SQLite version 3.39.4 2022-09-29 15:55:41
Enter ".help" for usage hints.
```

(SQL code generating sales summary for each year)

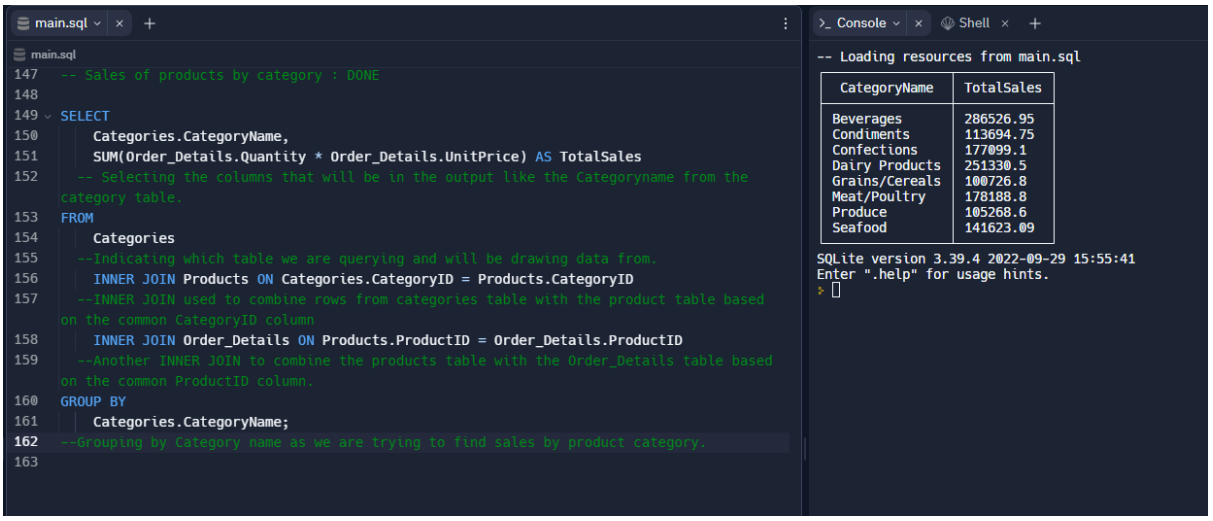
The values “1996”, “1997” and “1998” in the Year column represent our nominal data because they serve as the labels for the corresponding years, I believe that there are no columns containing ordinal or interval data in this code as there is no data that has a meaningful order, ranking or data that measures the difference between values. The TotalSales column is a representation of ratio data because it is a quantitative measurement of the sales which means meaningful mathematical operations can be carried out.

The year column displays categorical data since it categorises the sales into different years like 1996, 1997 and 1998. The TotalSales column is discrete data as it is a countable quantity. I believe that there is no continuous data in this code because there is no data that can be infinitely divided.

Output and its meaningfulness:

The output shows the total sales that have been generated in the years; 1996, 1997 and 1998, this is meaningful as it provides insights into the sales trends over the three years and can be used for analysing sales growth, helping to evaluate the success and profitability of specific periods. This can allow the business to find a way to take advantage of these specific periods.

Sales of products by categories



(SQL code generating the sales of all the products grouped by category)

The CategoryName column represents our nominal and categorical data as it is qualitative data used to name variables (e.g. the category of product) without providing any numeric value. I believe this code does not include any ordinal or interval data, this is because there is no data that can be categorised and ranked nor is there any data which uses values with fixed measurement units, where the distance between two points is known to me. The TotalSales column is an example of ratio and discrete data as it is quantitative data where 0 can be treated as the point of origin.

This code will output the name of the product category as well as the total sales for the products in these categories. This can be meaningful in a number of ways, for example :

Category-wise Sales Analysis: The output provides the total sales for each product category. This information allows sales managers to analyze sales performance across different categories. They can identify

the best-selling categories, compare sales figures between categories, and determine the relative contribution of each category to overall revenue.

Product Category Insights: By examining the total sales for each product category, sales managers can gain insights into customer preferences and market demand. They can identify product categories that generate the highest sales and focus on developing strategies to further promote and enhance those categories.

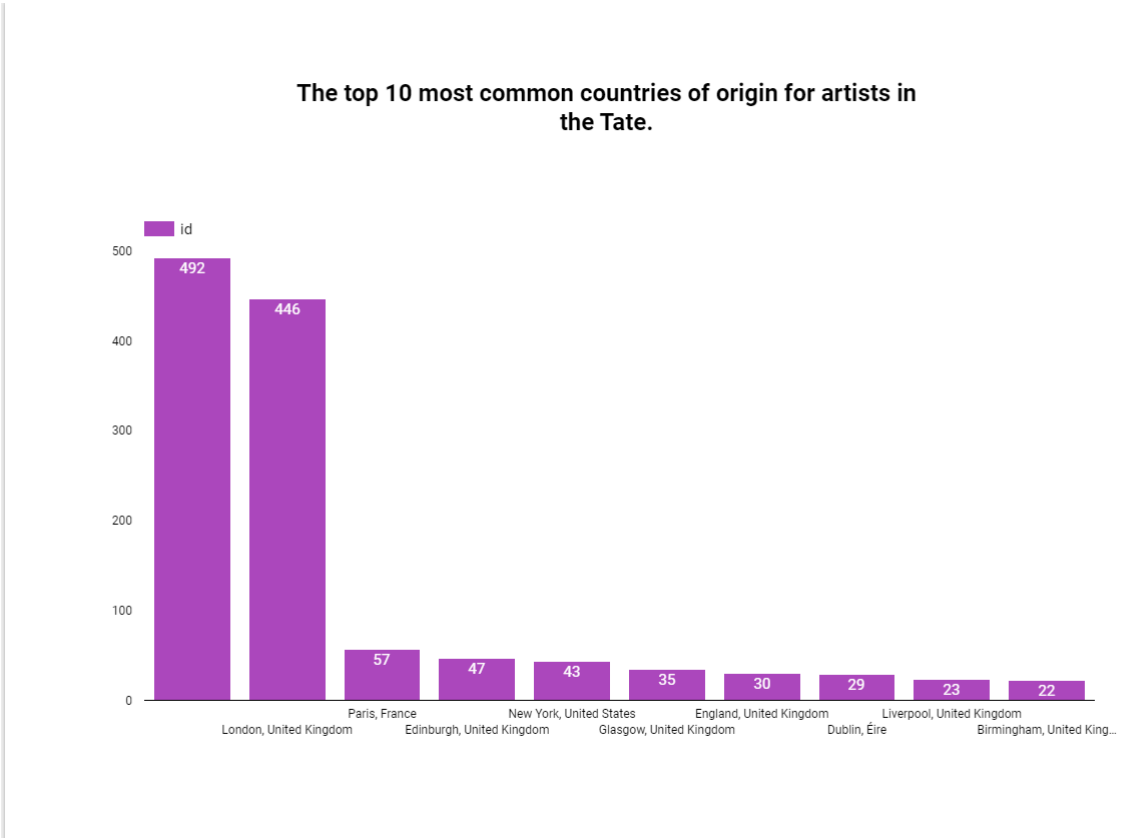
Resource Allocation: The output can aid in resource allocation decisions. Sales managers can allocate resources, such as sales representatives or marketing efforts, based on the performance of different product categories. Categories with higher sales may require additional resources or targeted marketing campaigns to maximize their potential.

Performance Evaluation: The output allows for the evaluation of sales team performance in relation to different product categories. It provides a basis for setting sales targets and evaluating the performance of individual sales representatives or teams based on their contribution to sales in specific categories.

Part 2 (30 marks):
Explore the following Datasets - artists and artworks from the Tate.
You may wish to use this Google Sheets or access them via BigQuery (artculture.museums.tate_artists and artculture.museums.tate_artworks)

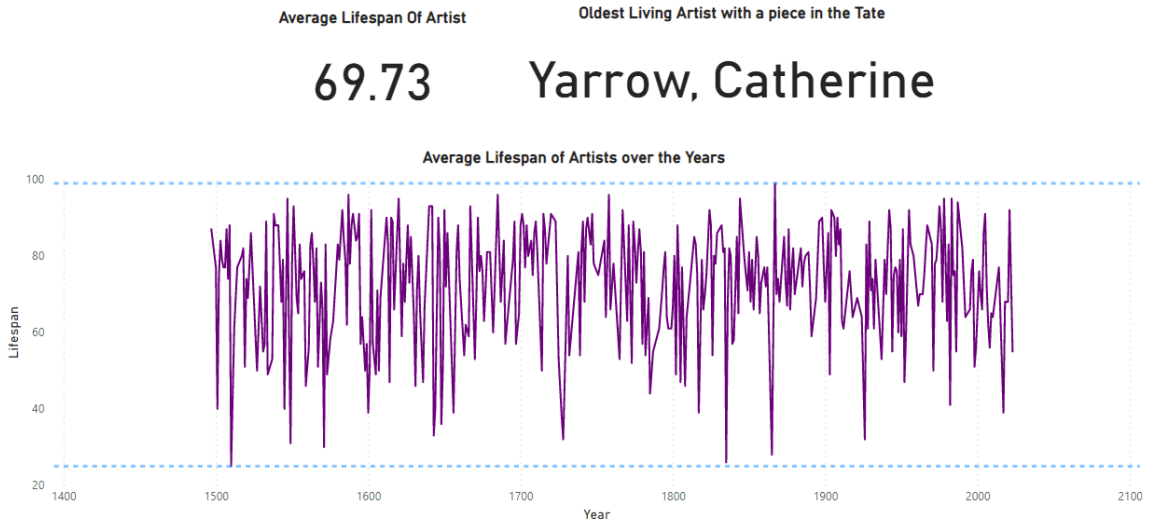
1. Answer the following questions, while showing your work through screenshots or commentary.

a) What are the most common countries of origin for artists in the Tate? Visualise your results in an appropriate way. (4 marks)



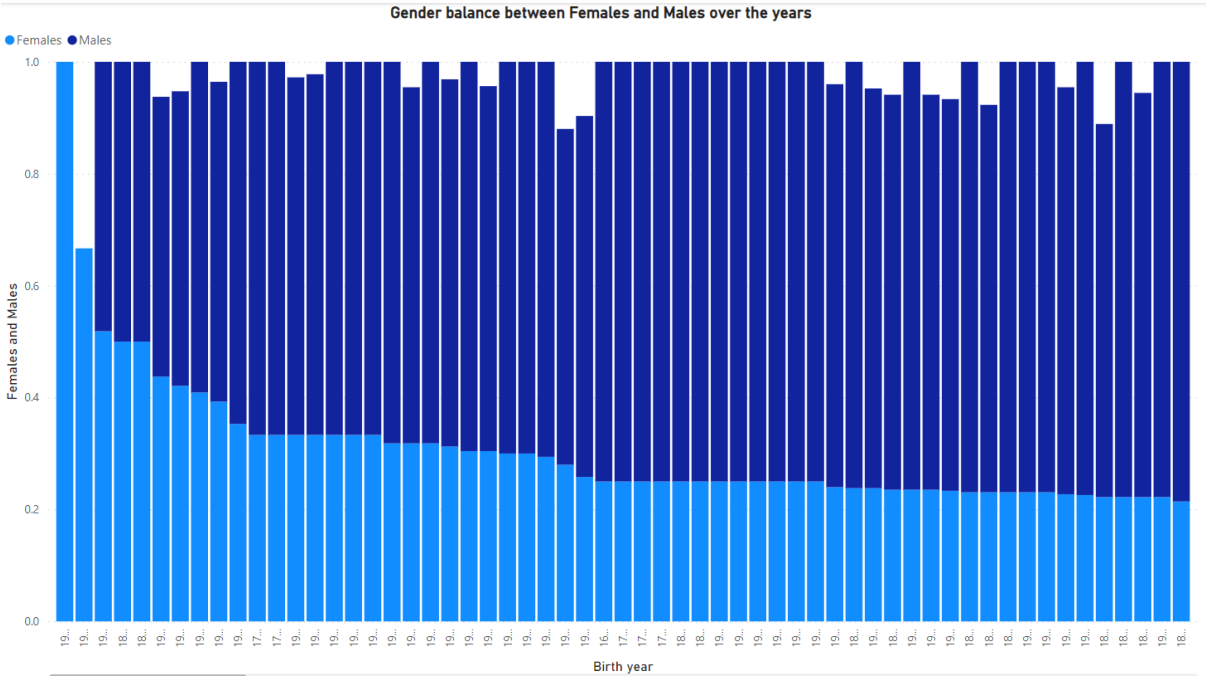
In my answer i have used the bar chart to visualise the top 10 most common countries of origin for artists in the Tate, I used a bar chart as it allows for a clear comparison of categorical data where the length or height of each bar represents its value, this makes understanding the visualisation almost instantaneous which can be useful when looking for insights. Another reason is that the bar chart is a very familiar visualisation which again makes it easy to understand but also it will help the viewer make more data-driven decisions as the bar chart makes it easy to identify trends and anomalies.

b) What is the average lifespan for a Tate artist? Does this change over time? Who is the oldest living artist with a piece in the Tate? (4 marks)



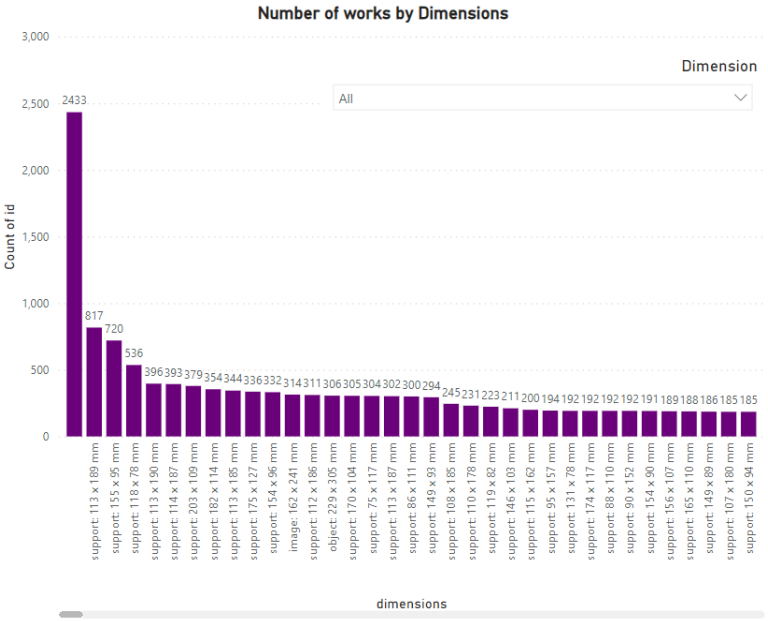
The Average lifespan for a tate artist is 69 - 70 years but this has changed quite drastically over time from the earliest year of 1497 to the current year, We can see that the highest average lifespan was in the year 1867 with a value of 99 years and the lowest average lifespan was in the year 1510 with a value of 25 years. The oldest artist that is alive and has a piece in the Tate is Yarrow Catherine, For these insights, I used the card visualisations as I needed to display single values like an average and a name, this would effectively display these single values without adding any unnecessary information, for the average lifespan over the years I used a line graph as it is easy to understand, it deals with time series analysis very well and will allow me to use further analysis like adding a min and max line.

c) The gender balance of artists in the Tate's collection is regrettable, but is the balance any better for artists born in the latter half of the twentieth century? Visualise your results in an appropriate way. (4 marks)

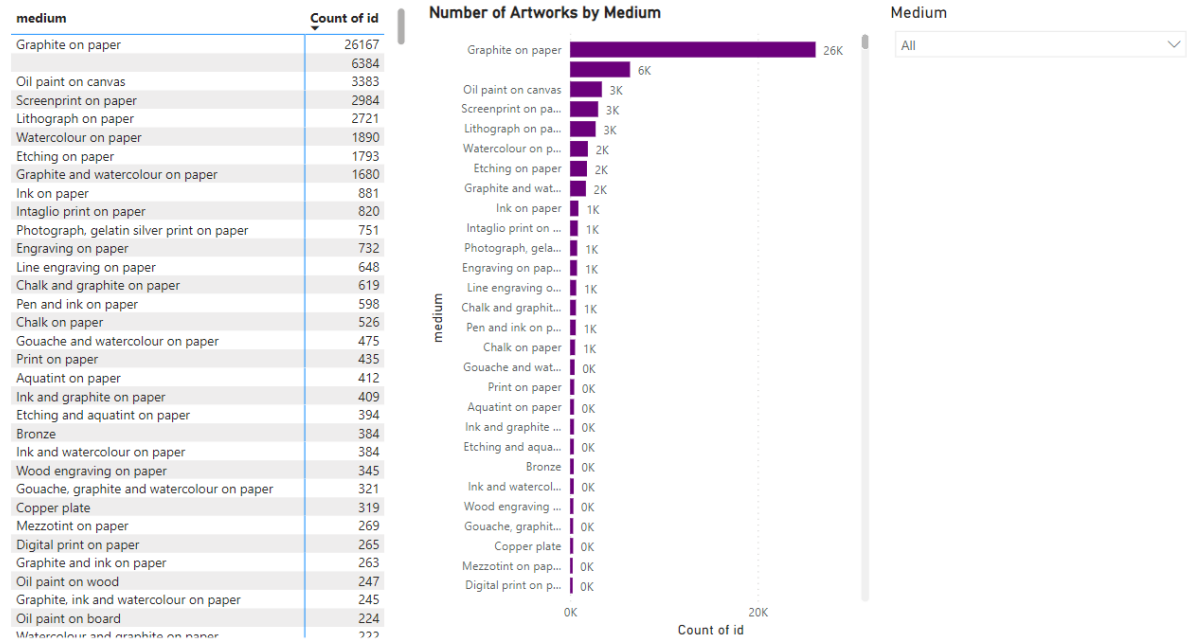


2. Produce three charts that provide an insight into the size (dimensions), medium (what they are made with) and provenance of works in the collection (how they ended up in the Tate). For each, you may need to do some data cleaning or shaping. (6 marks each)

dimensions	Count of id
	2433
10.5ft x 20 x 10 -overall display dimensions -variable	1
102 x 127 mm	1
102 x 178 mm	2
102 x 83 mm	1
104 x 164 mm	1
105 x 221 mm	1
105 x 460 mm	1
1053 x 758 mm	1
1072 x 811 mm	1
1075x1075x120mm weight: 28kg	1
10min	1
110 x 440 x 1000 mm	1
111 x 309 mm	1
1147.5 x 191 mm	1
115 x 490 x 300 mm	1
122 x 108 mm	1
1220 x 1830 mm	1
127 x 102 mm	3
127 x 172 mm	1
127 x 178 mm	1
127 x 184 mm	1
129 x 240 mm	1
1290 x 1220 x 700 mm	1
12min	1
130 x 127 mm	1
133 x 152 mm	1
133 x 178 mm	1
1346 x 2083 x 914 mm	1
136 x 210 mm	1
1440 x 560 mm	1
146 x 187 mm	1

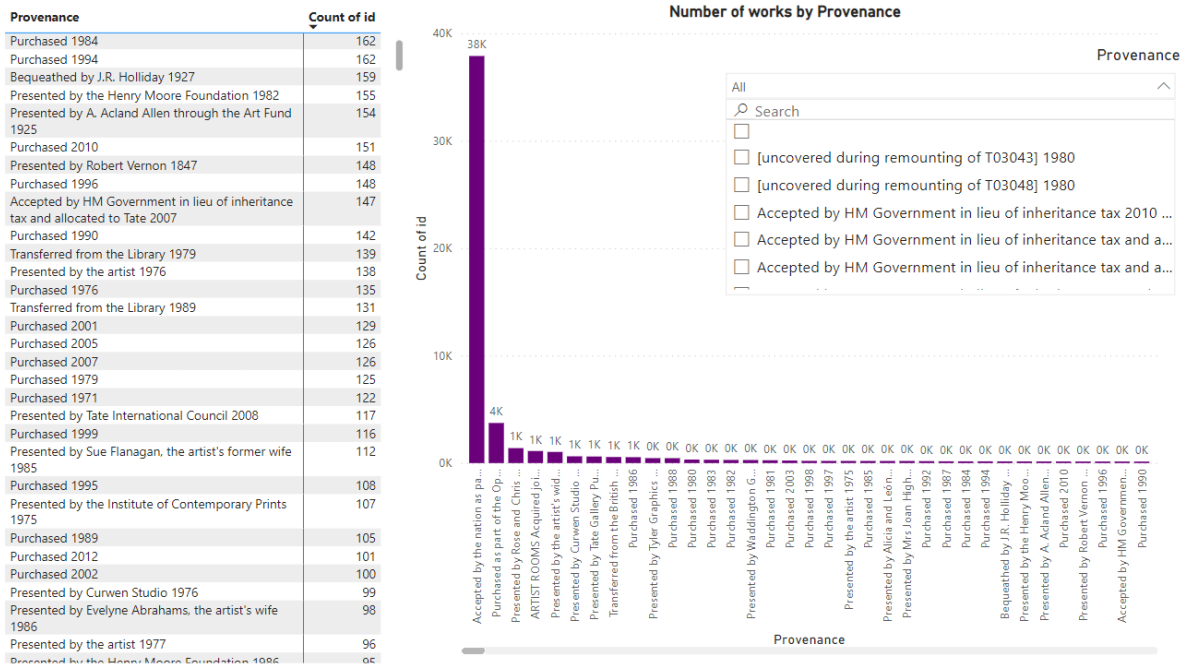


(Visualisation page showing the dimensions of works in the collection)



medium	Count of id
Graphite on paper	26167
	6384
Oil paint on canvas	3383
Screenprint on paper	2984
Lithograph on paper	2721
Watercolour on paper	1890
Etching on paper	1793
Graphite and watercolour on paper	1680
Ink on paper	881
Intaglio print on paper	820
Photograph, gelatin silver print on paper	751
Engraving on paper	732
Line engraving on paper	648
Chalk and graphite on paper	619
Pen and ink on paper	598
Chalk on paper	526
Gouache and watercolour on paper	475
Print on paper	435
Aquatint on paper	412
Ink and graphite on paper	409
Etching and aquatint on paper	394
Bronze	384
Ink and watercolour on paper	384
Wood engraving on paper	345
Gouache, graphite and watercolour on paper	321
Copper plate	319
Mezzotint on paper	269
Digital print on paper	265
Graphite and ink on paper	263
Oil paint on wood	247
Graphite, ink and watercolour on paper	245
Oil paint on board	224
Watercolour and graphite on paper	222

(Visualisation page showing the medium of works in the collection)



Provenance	Count of id
Purchased 1984	162
Purchased 1994	162
Bequeathed by J.R. Holliday 1927	159
Presented by the Henry Moore Foundation 1982	155
Presented by A. Acland Allen through the Art Fund 1925	154
Purchased 2010	151
Presented by Robert Vernon 1847	148
Purchased 1996	148
Accepted by HM Government in lieu of inheritance tax and allocated to Tate 2007	147
Purchased 1990	142
Transferred from the Library 1979	139
Presented by the artist 1976	138
Purchased 1976	135
Transferred from the Library 1989	131
Purchased 2001	129
Purchased 2005	126
Purchased 2007	126
Purchased 1979	125
Purchased 1971	122
Presented by Tate International Council 2008	117
Purchased 1999	116
Presented by Sue Flanagan, the artist's former wife 1985	112
Purchased 1995	108
Presented by the Institute of Contemporary Prints 1975	107
Purchased 1989	105
Purchased 2012	101
Purchased 2002	100
Presented by Curwen Studio 1976	99
Presented by Evelyne Abrahams, the artist's wife 1986	98
Presented by the artist 1977	96
Presented by the Henry Moore Foundation 1986	95

(Visualisation page showing the provenance of works in the collection)

In the three pages above I have utilised a Matrix to show the dimensions, medium and provenance of works alongside a distinct count of ID, this will show how many artworks are in each of these categories however there are so many categories it can be difficult to read, I also included either a clustered column chart or car chart to make the contrast between values much easier to visualise however this also

sufferers from the same issue I mentioned earlier. To counter this I included a slicer which opens a dropdown menu with a search bar, containing the categorical data, this allows the user to filter by one or many different categories which enables you to easily see the values and comparison of the data.

Part 3 (20 marks):

1. Briefly discuss the value of Open Access Datasets for Cultural Institutions. Why might they benefit from making data available in this way? What might some of the risks be? (3 marks)

Open access datasets provide us with a significant amount of value to cultural institutions, I will list some of the benefits which offer value:

- Open access datasets encourage knowledge sharing and collaboration among cultural institutions as people from different institutions can access the data, analyse the data and gather insight which could then be shared.
- Open access datasets push cultural institutions to reach a wider audience since it removes any restrictive licences or paywalls, this means it will be easier reaching students, developers, researchers and even the general public. This provides broadened access to use the data for educational purposes.
- Open datasets can lay out a foundation for innovation and the development of new applications as it provides a test environment which allows people to test new ideas which can then lead to new innovations that can be shared.
-

However, there are also some potential risks associated with open access datasets like the following:

- The quality and integrity of open datasets may not be good so it is crucial to ensure it is up to standards, cultural institutions need to establish data collection processes in order to maintain high data accuracy and reliability, they will also need to be cautious about any privacy implications of sharing datasets as it may contain personal or sensitive information.

- The misuse or misrepresentation of open access datasets, Cultural institutions should be proactive in providing context to the dataset, guidelines and revoke or put in place access restrictions to prevent misuse.
- Cultural institutions need to be extremely cautious about property rights and copyright issues as carelessly including this in a dataset that is open access may have severe consequences.

2. Some works in the dataset are in the public domain, but others are not – identify sensitive information in the dataset and suggest mechanisms to avoid risks. (3 marks)

Some of the sensitive information that could be present in the open-access collection for the Metropolitan Museum of art is the following:

- Some of the artworks in the collection may be copyrighted which means they are protected and any sort of reproduction or even use without the permission of the right authorities will infringe on the rights of the copyright holder.
- A few of the artworks may have some religious or cultural significance so any inappropriate use or mishandling of the artworks can become an offence or at least be seen as disrespectful.

I will provide some measures to avoid these risks below :

- Clearly specify any sort of restrictions for the use of an artwork that is not in the public domain, this can just be restrictions on modifications or commercial use.
- Display clear information about whether an artwork is copyright protected, including whether it is in the public domain or still protected, you could also include what you can't do if it is copyrighted.

- Create a reporting mechanism that allows people to report any sort of misuse or mishandling of any artwork, this will aid in addressing any issues with compliance with the museum guidelines.

3. Construct a data dictionary with the data types of columns represented in the dataset. Briefly discuss the data quality, and range of possible results, of each. (7 marks)

Field name	**Data Type**	**Data Format**	**Description**	**Example**	**Data Quality**
Object Number	String/ Text	xxxx.xxx.x	Identifier for object	1979.486.1	Good
Is Highlight	String/ Text	XXXXX	This is a boolean to show if a condition is met	False	Good
Is Timeline Work	String/ Text	XXXXX	This is a boolean to show if a condition is met	False	Good
Is Public Domain	String/ Text	XXXXX	This is a boolean to show if a condition is met	False	Good

Object ID	Integer	x	This is a unique identifier	2	Good
Gallery Number	Integer	xxx	Number to identify different galleries	774	Missing some data
Department	String/ Text	“Xxx xxxxxxx x xxxx”	Identify different departments	The American Wing	Missing some data
AccessionYear	Integer	xxxx	Year date	1979	Good
Object Name	String/ Text	“XXXX”	Name of the object	Coin	Good
Title	String/ Text	“Xxx xxxxx xxx”	The name of art piece	Ten-do dollar Liberty Head Coin	Good
Culture	String/ Text	“xxxxxxx”	Identify the culture from with an artwork comes from	Mexican	Poor - Missing data

Period	Missing Data.	Missing Data.	Ambiguous	Missing Data.	Poor - Missing data
Dynasty	Missing Data.	Missing Data.	Ambiguous	Missing Data.	Poor - missing data
Reign	Missing Data.	Missing Data.	Ambiguous	Missing Data.	Poor Missing data
Portfolio	Missing Data.	Missing Data.	Ambiguous	Missing Data.	Poor - missing data
Constituent ID	Integer	xxx	This is a unique identifier	107	Poor
Artist Role	String/ Text	“xxxxxx”	This shows what roles the artists have	Maker	Poor
Artist Display Name	String/ Text	“xxxxxx”	This shows the name of artists that re displayed	James Barton Longa cre	Poor

Artist Display Bio	String/ Text	“xxxxxxx x”	This can show an address	American, Delaware County, Pennsylvania 1794–1869 Philadelphia, Pennsylvania	Poor
Artist Alpha Sort	String/ Text	“xxxxxx”	Ambiguous	Longacre, James Barton	Poor
Artist Nationality	String/ Text	“xxxx”	This tells us the nationality of the artist in each row	American	Poor
Artist Begin Date	Integer	xxxx	This tells us the date when someone became an artist	1794	Poor
Artist End Date	Integer	xxxx	This tells us the date when someone stopped being an artist	1869	Poor

Artist ULAN URL	String/ Text	“xxxx”	Ambiguous	http://vocab.getty.edu/page/ulan/500011409	Poor
Artist Wikidata URL	String/ Text	“xxxx”	Ambiguous	https://www.wikidata.org/wiki/Q3806459	Poor
Object Date	Date	YYYY	This tells us the date of an object	1853	Poor - Different formats used
Object Begin Date	Date	YYYY	This tells us the end date of an object	1853	Poor
Object End Date	Date	YYYY	This tells us the date of an object	1853	Poor
Medium	String/ Text	“xxxxx”	Ambiguous	Gold	Poor

Dimensions	String/ Text	“XXXXX X”	Ambiguous	Diam. 11/16 in. (1.7 cm)	Poor - Different formats used
Credit Line	String/ Text	“xxxxxxx”	Ambiguous	Gift of Heinz L. Stoppelmann, 1979	Good
Geography Type	String/ Text	“XXXX”	Tells us if the piece was made domestically	Made In	poor
City	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
State	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
County	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Country	String/ Text	“xxxxx”	Tells us which country an artwork comes from	Mexico	poor

Region	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Subregion	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Locale	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Locus	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Excavation	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
River	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor
Classification	Missing Data.	Missing Data.	Ambiguous	Missing Data.	poor

Rights and Reproduction	String/ Text	“xxxxxxx”	Ambiguous	© 2022 Artists Rights Society (ARS), New York	Poor
Link Resource	String/ Text	“xxxxx”	Ambiguous	http://www.metmuseum.org/art/collection/search/1	poor
Object Wikidata URL	String/ Text	“xxxxx”	Ambiguous	https://www.wikidata.org/wiki/Q83545838	Poor
Metadata Date	Missing Data.	Missing Data.	Ambiguous	Missing Data.	Poor
Repository	String/ Text	“xxxxx”	Ambiguous	Metropolitan Museum of Art, New York, NY	Good

4. What surprising things have you discovered about the museum's collections? You may wish to compare it with the collection of the Tate. (7 marks)

The Artworks in the Museum collections (The Metropolitan Museum of Art and the Tate Gallery) consist of a diverse plethora of pieces that represent various cultures and time periods, It is also quite varied in the mediums that it encompasses as it includes paintings, photographs, sculptures and other art forms.

Both the Tate Gallery and the Metropolitan Museum of Art have very distinct areas of focus in regard to regional and cultural focus. The Tate Gallery emphasises British Art and incorporates international modern and contemporary artworks, In contrast, The Metropolitan Museum of Art takes a more concentrated global approach by showcasing artworks from different regions worldwide.

Both museums feature everything by famous artisans and highlight various beautiful shifts. The Metropolitan Museum of Art displays everything by inventors in the way that Leonardo da Vinci, Vincent van Gogh, and Pablo Picasso, between possible choices. The Tate Gallery, in another way, showcases artworks by British artisans like J.M.W. Turner, Damien Hirst, and Tracey Emin, in addition to worldwide modern and existing experts.

Additionally, the Metropolitan Museum of Art and the Tate Gallery are unconnected organisations accompanying their own unique curatorial approaches and plans for expanding their groups. Although two organisations endeavour to preserve and exhibit skill, their group preference and addition processes can influence differences in the composition and theme focus of their groups.