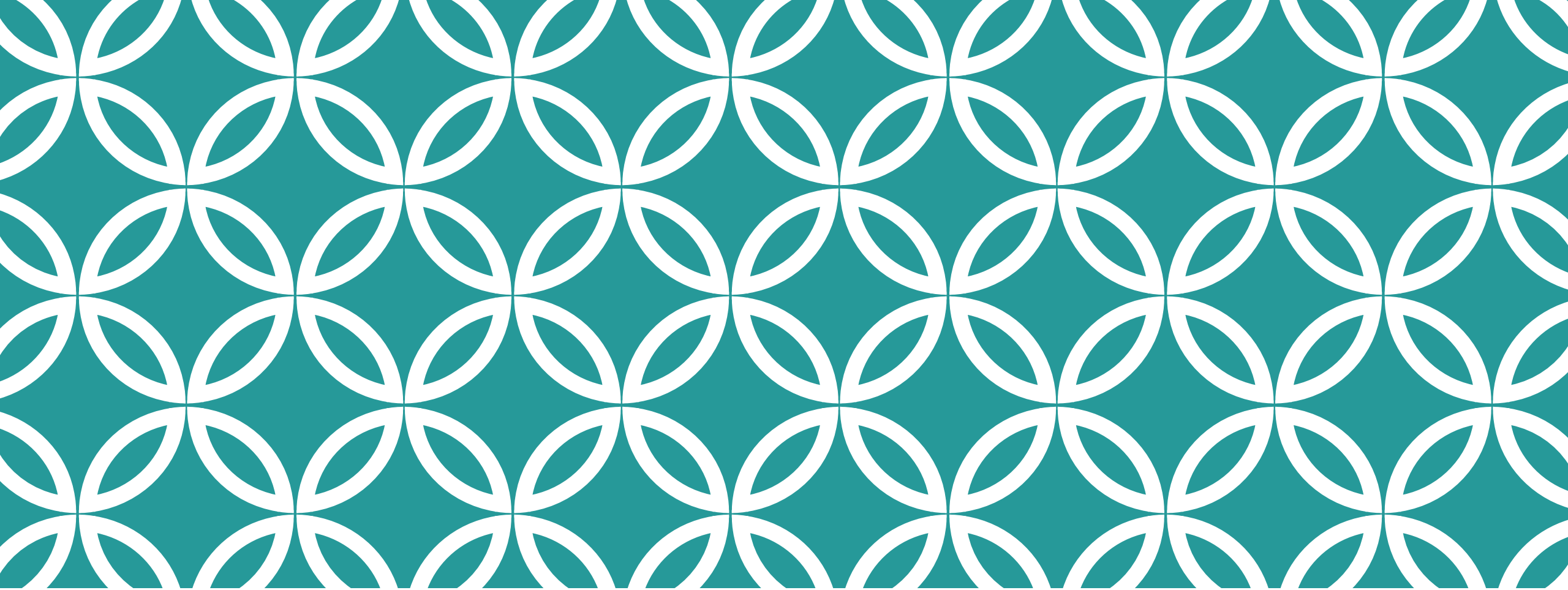

CHAPTER 5: GROUPING DATA

2023

CHAPTER 5: GROUPING DATA

Content:

1. Group By Clause
2. Having Clause
3. Order of Clauses
4. Distinct Clause

GROUP BY

```
SELECT column_name(s)
FROM table_name
WHERE condition
GROUP BY column_name(s)
ORDER BY column_name(s);
```

```
SELECT Country, COUNT(CustomerID) as NrOfCustomers
FROM Customers
GROUP BY Country
ORDER BY COUNT(CustomerID) DESC;
```

Select Department, AVG(salary) from tbl

Group by Department

Order by Avg(salary) DESC

Practice GROUP BY

1. What is the most expensive price by room_type?
2. Show the average price for each host_id. Show their host_id host_name and average price as price_on_avg
3. Calculate how many listings are found in each neighbourhood and only show the 5 highest ones (You can show the neighbourhood and rename the calculation column as nr_of_listings)
4. Display the 10 most expensive places for each neighbourhood

AIRBNB

HAVING

The HAVING Clause enables you to specify conditions that filter which group results appear in the results.

The WHERE clause places conditions on the selected columns whereas the HAVING clause places conditions on groups created by the GROUP BY clause.

HAVING

```
SELECT column_name(s)
FROM table_name
WHERE condition
GROUP BY column_name(s)
HAVING condition
ORDER BY column_name(s);
```

```
SELECT COUNT(CustomerID)  Country
FROM Customers
GROUP BY Country
HAVING COUNT(CustomerID) > 5
ORDER BY COUNT(CustomerID) DESC;
```

Practice HAVING

1. Find the average price of Airbnb listings in each neighborhood, and only show neighborhoods where the average price is greater than R1000
2. Find the number of reviews for each host, and only show hosts who have more than 50 reviews
3. Find the average number of days per year that each room type is available to book, and only show room types with more than 240 days of availability:

AIRBNB

ORDER OF CLAUSES

1. SELECT
2. FROM
3. WHERE
4. GROUP BY
5. HAVING
6. ORDER BY

SELECT DISTINCT

The SELECT DISTINCT statement is used to return only distinct (different) values.

Inside a table a column often contains many duplicate values; and sometimes you only want to list the different (distinct) values.

eg.

```
SELECT DISTINCT column1,  column2  ...
FROM table_name;
```

SELECT DISTINCT

SELECT DISTINCT *column1 column2 ...*
FROM *table_name;*

SELECT DISTINCT Country FROM Customers;

SELECT COUNT(DISTINCT Country) FROM Customers;

SELECT COUNT(DISTINCT F_Name, L_Name) FROM Customers;

SELECT COUNT(distinct(Titles)) FROM BOOKS

Where title like '%the%';

```sql
SELECT DISTINCT sale_date, store_state,
       sum(sale_amount) as total_sales
FROM customers
GROUP BY store_state, sale_date
ORDER BY sale_date ASC;
```

Data Output | Messages | Notifications | Explain

	sale_date date	store_state character varying (255)	total_sales numeric
1	2020-05-21	NY	1200
2	2020-05-22	MH	1500
3	2020-05-22	NULL	2410
4	2020-05-23	KA	1200
5	2020-05-23	MH	2520

Practice DISTINCT

1. Show a unique list of all the hostnames and neighbourhoods mentioned
2. How many different room types are there?
3. How many listings have catering mentioned in their name?

AIRBNB