Database Designs

Section 3: Advanced Retrieving Data Queries

3.1 Aggregate Operators

Aggregate operators allows us to return more interesting information from our data. Below is an example Sales_Data table containing sales for two ASDA stores.

Sales_Data			
storeLocation	productName	salesDate	revenue
ASDA Coventry	Milk	November 02, 2020	1,233.32
ASDA Birmingham	Bread	November 02, 2020	5,434.74
ASDA Coventry	Coffee	November 02, 2020	3,855.96
ASDA Coventry	Coffee	November 02, 2020	2,280.90
ASDA Birmingham	Coffee	November 02, 2020	2,110.95
ASDA Coventry	Milk	November 01, 2020	4,558.24
ASDA Birmingham	Milk	November 01, 2020	6,849.99
ASDA Birmingham	Bread	November 01, 2020	2,543.57

The insights an analyst or business owner may be interested in retrieving from the above dataset are: Total Revenue, Best Performing Stores, Best Performing Products, Best Product-Store Combination that Sold Best.

We would like to find patterns in the data so that we can drive business decisions that would help the business to perform better. Below are some example Aggregate Operations which can help us retrieve more interesting data insights.

```
SELECT SUM(revenue) totalRevenue
FROM Sales_Data;
```

The **SUM** aggregate operator wraps the column in brackets for the column data we wish to apply the sum function. Whatever result is returned from the aggregate operator is then returned in a column we have decided to name as totalRevenue. By omitting the **WHERE** clause from the select statement means that it will sum over all data within the Sales_Data table.

The **SUM** is a function that operates over an entire column and not just a single cell and the result returned is a single value which can be named. The **SUM** can operate over an entire column or any subset of a column and does not necessarily have to be an entire column.

The reason SUM is called an aggregate function is because it acts on an aggregation of cells and not just a single cell.

```
SELECT AVG(revenue) averageRevenue
FROM Sales_Data;
```

The **AVG** aggregate operator is a function that operates over an entire column or a subset of a column and not just a single cell and returns a single value which can be named. This function averages all the values in the column. The syntax is the same as the **SUM** function above but it returns the average instead.