

Aggregate and filter to summarize data

So far, you've learned that aggregation is the process of gathering and summarizing data sets to produce meaningful, actionable insights. Building on your previous exploration of data processing methods, this reading will dive deeper into the specifics of aggregating with dimensions. You'll also gain more insight into the usefulness of filters in data processing through the Looker Studio platform.

Aggregation review

As a reminder, data aggregation is the process of collecting and analyzing data, and expressing it in a summary form. Data aggregation is an essential component of data analysis because it streamlines datasets and presents meaningful information at a high-level. Previously, you explored different aggregation methods in SQL, like **SUM**, **AVG**, and **COUNT**. These methods help you extract important information from a dataset without having to sift through every single data point. Not only can you aggregate data points, but you can apply these methods on dimensions, or columns of data.

Aggregate with dimensions

Dimensions categorize, slice, and segment your data. Aggregating with dimensions means summarizing data based on specific categories or segments. Consider a retail chain that operates in multiple cities. For each sale, the retailer records the date, city, product sold, the quantity sold, and the total sale amount. This table provides an overview of total sales, total units sold, and top selling products from the example.

City	Total Sales (\$)	Total Units Sold	Top Selling Product
New York	\$500,000	10,000	Product B
Los Angeles	\$450,000	9,500	Product A
Chicago	\$480,000	9,800	Product C

Choosing “city” as a dimension will group your data by location. Grouping sales data by city answers questions like:

- Which city had the highest sales month?
- How does the sale of a specific product in City 1 compare to City 2?
- Which city has seen the most significant growth in sales over the past year?

Categorizing and summarizing with a dimension allows for easier comparison and analysis

between groups. Dimensions help data professionals zoom in on very specific parts of their data. Depending on the needs of the project, or the preferences of the data professional, dimensions also allow for significant flexibility in data analysis. You can start with a specific segment or category you want to investigate, but you can also reverse engineer your analysis by starting with questions.

Aggregate data in Looker Studio

Looker Studio is a software for effective data analysis that offers various tools to aggregate data seamlessly. Aggregations in Looker Studio are similar to that of SQL, with these methods:

- Average
- Count and Count Distinct
- Max and Min
- Sum

Returning to the previous example, you can use Looker Studio to aggregate the same dimensions of city, total sales, total units, and top selling product. Use the **Average** function in Looker Studio to aggregate the dimension “Total Sales” to calculate the average total sales for each city, or the average total sales for each product.

Metric	Value
New York	\$500,000
Los Angeles	\$450,000
Chicago	\$480,000
Average	\$476,667

Utilize filters on dimensions

Not only are dimensions useful for grouping data, but you can group and summarize data further using filters. Filtering is the process of showing only the data that meets a specified criteria, while hiding the rest. As an analyst, you can use filtering tools to focus only on the data you need for specific projects and desired outcomes.

If you’re working on a project like the retailer’s, you may want to focus on the sales for a specific product, or the sales performance for a specific city. You can use filters to display only this information.

Note: Filtering and aggregating data are similar methods for grouping data. Filtering narrows down data using a specific criteria, while aggregating summarizes data at a higher level.

Filter data in Looker Studio

Not only can you aggregate on dimensions in Looker Studio, but you can also filter data. Looker Studio allows you to add chart filters, report/page-level filters, and filters for an entire report. You can filter on any applicable dimension or measure. As a reminder, a measure is an aggregation of one or more dimensions, like count or average.

Filters applied to charts will filter data on a specific visualization. At the report level, a filter will apply to all data on the page. Filters applied to the entire report will filter data on every single page.

You can create a filter by clicking CREATE NEW FILTER under the Setup tab in the properties panel. Then select the dimension you want to filter by. You'll need to select the comparison operator (**Equals**, **Contains**, **Greater than**, **Less than**), and provide a comparison value for the expression. Comparison operators compare numbers or strings, and perform evaluations. In this case, the operators compare dimensions with a provided target or value. Like SQL, you can add an **AND** or **OR** clause. Then, click SAVE.

To contextualize filtering data in Looker Studio, consider the major retailer example from earlier in this reading. You can filter on the dimension "Total Sales" to be **Less than** 500,000 to determine which locations are underperforming based on this target.

Key takeaways

Data aggregation is a crucial step in the data analysis process. Aggregation methods provide summarized insights from data sets. Software like Looker Studio can streamline the aggregation process, and provide filtering capabilities. Aggregating and filtering data is essential to refining data sets to ensure you can focus on the most important information.