Bike Insights

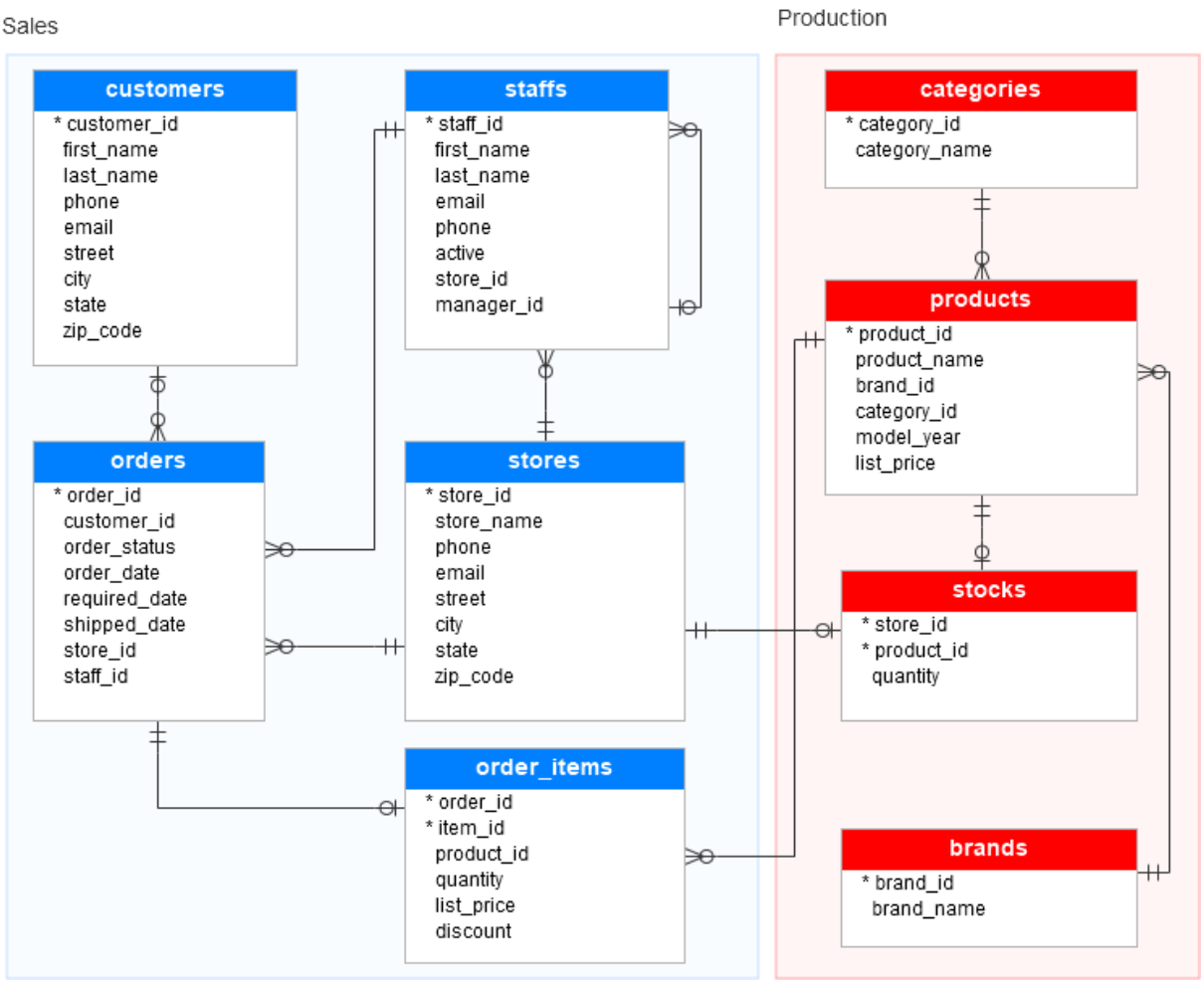
Unveiling Trends, Optimising Sales, and Empowering Growth

Introduction

The problem statement revolves around a fictional bike company that has a huge dataset of their sales and production. It includes information about customers, employees, orders, products, and suppliers at the bike company. Each table contains relevant columns representing different attributes and relationships between entities. Through this analysis, we explore the database's structure, delve into the data, and apply SQL queries to perform tasks such as retrieving sales information, analysing product trends, and understanding customer behaviour. We further analyse the specifics of data using Pivot Tables and create a final dashboard to present our analysis.

Contents of the data:

The company has it's data in over 9 different tables. The two major teams are sales and production. It is important to understand the relationships between each entity to perform the analysis.



Understanding the problem:
There are 9 individual but related tables to help understand the scope of the problem better:

Sales Team:

1. Customers: This table contains information about the customers, such as customer ID, name, email, and address.
2. Staffs: The employees table includes details about the store's staff, such as employee ID, name, position, and contact information.
3. Orders: This table captures information about customer orders, including order ID, customer ID, employee ID, order date, and order status.
4. OrderItems: The order items table provides details about the individual items within each order, such as the order item ID, order ID, product ID, quantity, and unit price
5. Stores: This table has information about all the stores selling the bikes. This includes storeid, name, location and contact information

Production Team:

1) Products: This table holds information about the products available for sale, including product ID, product name, category, and price.
2) Categories: The categories table categorises the products into different categories, such as bikes, accessories, and clothing.
3) Stocks: Includes the quantity of stocks every product holds
4) Brands: includes brand, id and name

Creating a database using SQL:
Now I know what is in the data, it is time to put together a dataset. Dealing with 9 different tables is difficult, our next step would be to create one main dataset to help us analyse and visualise the dataset better.

The first step would definitely be creating these nine tables in the database and importing the data. We would be creating a comprehensive join operation on these tables to retrieve aggregated information:

Breaking down the queries:

1. The SELECT statement is used to specify the columns to be included in the query result.
2. The columns selected include:
    - ord.order_id: Order ID from the "orders" table.
    - CONCAT(cus.first_name," ", cus.last_name) AS "customers": Concatenated full name of the customers from the "customers" table.
    - cus.city: City of the customers.

- cus.state: State of the customers.
- ord.order_date: Date of the orders.
- SUM(ite.quantity) AS "total_units": Total units (quantity) of products in the order, aggregated using the SUM function.
- SUM(ite.quantity*ite.list_price) AS "revenue": Revenue generated from the order, calculated as the sum of quantity multiplied by the list price of each item.
- pro.product_name: Name of the products from the "products" table.
- cat.category_name: Name of the categories from the "categories" table.
- sto.store_name: Name of the stores from the "stores" table.
- CONCAT(sta.first_name, " ", sta.last_name) AS "sales_rep": Concatenated full name of the sales representative from the "staffs" table.

3. The FROM clause specifies the tables to be used in the query: "orders", "customers", "order_items", "products", "categories", "stores", and "staffs".
4. The JOIN statements define the relationships between the tables using common columns:
    - ord.customer_id=cus.customer_id: Joining the "orders" table with the "customers" table based on the customer ID.
    - ite.product_id= pro.product_id: Joining the "order_items" table with the "products" table based on the product ID.
    - ord.order_id= ite.order_id: Joining the "orders" table with the "order_items" table based on the order ID.
    - pro.category_id= cat.category_id: Joining the "products" table with the "categories" table based on the category ID.
    - ord.store_id= sto.store_id: Joining the "orders" table with the "stores" table based on the store ID.
    - ord.staff_id= sta.staff_id: Joining the "orders" table with the "staffs" table based on the staff ID.
5. The GROUP BY clause groups the result set based on specific columns to calculate aggregated values:
    - ord.order_id, CONCAT(cus.first_name," ", cus.last_name), cus.city, cus.state, pro.product_name, cat.category_name, sto.store_name, CONCAT(sta.first_name, " ", sta.last_name)

Summarising the dataset- Pivot Tables (Excel)

To further understand and analyse the dataset, it is important to get a jist of it. We are going to use excel to get a summary of the values in this huge dataset. We would be creating pivot tables to understand the effect on revenue due to different factors

Product Brand	Sum of revenue
Baldwin Bikes	5826242.21
Rowlett Bikes	962600.76
Santa Cruz Bikes	1790145.91
Grand Total	**8578988.88**

Yearly	Sum of revenue
2016	2709484.47
2017	3845515.02
2018	2023989.39
Grand Total	**8578988.88**

Sales Rep	Sum of revenue
Marcelene Boyer	2938888.73
Venita Daniel	2887353.48
Genna Serrano	952722.26
Mireya Copeland	837423.65
Kali Vargas	516695.17
Layla Terrell	445905.59
Grand Total	**8578988.88**

Brand Categorie	Sum of revenue
Children Bicycles	327888.21
Comfort Bicycles	438506.87
Cruisers Bicycles	1109151.04
Cyclocross Bicycles	799874.6
Electric Bikes	1020236.85
Mountain Bikes	3030775.71
Road Bikes	1852555.6
Grand Total	**8578988.88**

States	Sum of revenue
CA	1790145.91
NY	5826242.21
TX	962600.76
Grand Total	**8578988.88**

Monthly	Sum of revenue
2016	**2709484.47**
Jan	241184.15
Feb	175768.1
Mar	202157.14
Apr	187223.55
May	228701.13
Jun	231120.29
Jul	222854.21
Aug	253130.83
Sep	303282.61
Oct	235051.79
Nov	205315.47
Dec	223695.2
2017	**3845515.02**
Jan	316954.77
Feb	348740.47
Mar	348177.13
Apr	254105.57
May	297754.66

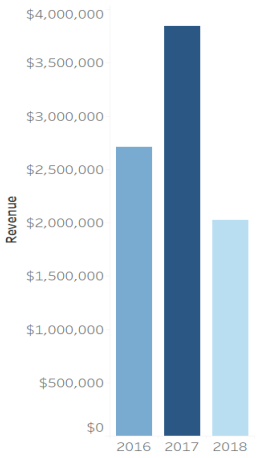
Customer Names	Sum of revenue
Pamelia Newman	37801.84
Abby Gamble	37500.89
Sharyn Hopkins	37138.86
Lyndsey Bean	35857.86
Emmitt Sanchez	34503.82
Melanie Hayes	34390.88
Debra Burks	30645.87
Elinore Aguilar	29661.83
Corrina Sawyer	29214.89
Shena Carter	27618.95
Robby Sykes	27157.88
Abram Copeland	26913.92
Genoveva Baldwin	26679.78
Tameka Fisher	26249.81
Lorrie Becker	25940.88
Teofila Fischer	25772.86
Mozelle Carter	25382.84
Guillermina Noble	24779.86
Damien Dorsey	24686.84

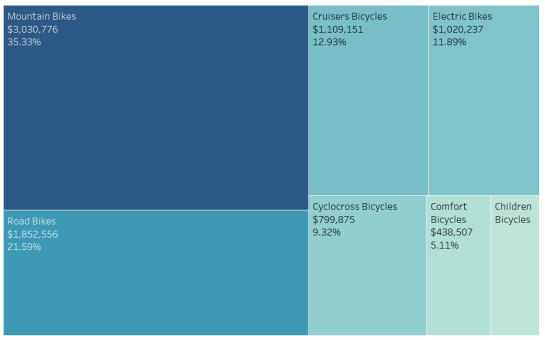
These pivot tables allow us to summarise large amounts of data quickly and efficiently. You can easily aggregate and consolidate data based on different criteria, such as totals, averages, counts, or percentages. This summarization helps gain insights and identify patterns or trends in your data. Pivot tables help in making data-driven decisions by presenting information in a concise and organised manner. Now that we “know” our data it gets easier to create visualisations out of it.

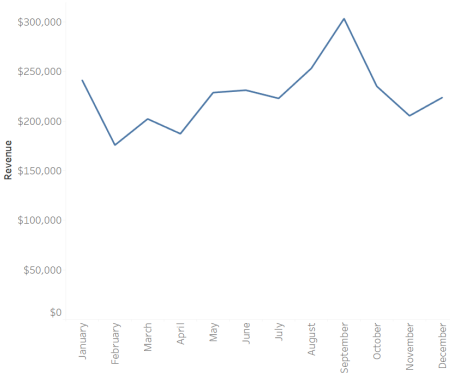
Visualising in Tableau:

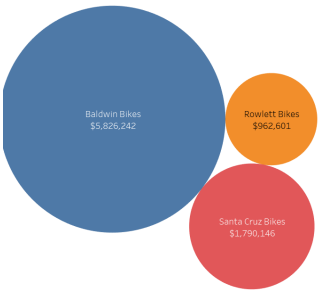
Visualisations are important because they provide a clear and intuitive representation of data, making it easier to interpret and understand complex information. They can reveal patterns, trends, and relationships that may not be immediately apparent in raw data or tabular formats

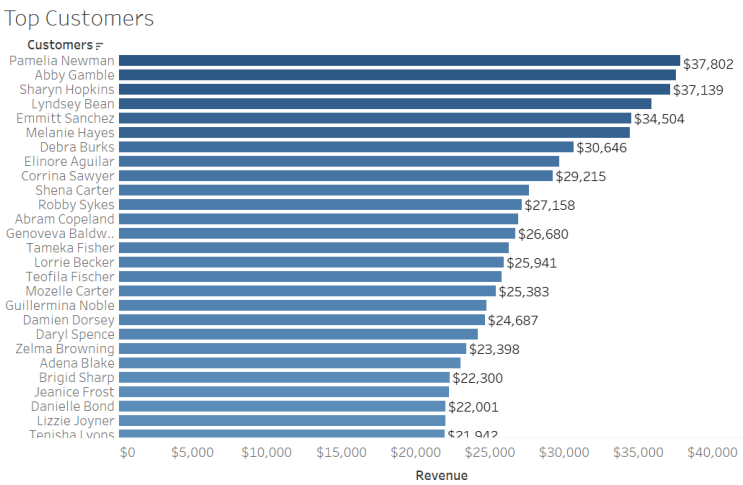
Here are some visualisations that better represent the pivot tables.

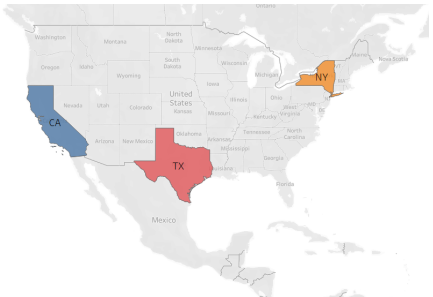












Conclusion:

This project involved the analysis of a fictional bike database. Through the utilisation of SQL queries and pivot tables, we delved into the dataset, examined relationships between tables, and derived valuable information.

The analysis focused on various aspects, including customer behavior, sales performance, and product trends. By aggregating data and performing calculations, we gained a comprehensive understanding of the dataset.

The findings revealed important insights, such as top-selling products, revenue generated from different categories, customer distribution across cities and states, and the performance of sales representatives. These insights can guide decision-making processes and provide actionable recommendations for the bike shop.