



Bike Sales Analysis Report

21.02.2025

Emmanuella Oluwakemi Odiete
emmanuellaodiete04@gmail.com

Project Overview

Objective

This project aimed to analyze bike sales data to gain insights into sales trends, customer behavior, product performance, and store performance. The goal was to identify key factors influencing sales, understand customer preferences, and provide actionable recommendations to improve business strategies, optimize inventory management, and enhance overall sales performance. Specifically, the analysis sought to answer questions such as:

- What are the top-selling products and categories?
- Which brands are performing best?
- How do sales vary across different stores and staff members?
- What are the key characteristics of our customer base?
- Are there any seasonal trends in sales?

Context

The dataset used for this analysis was sourced from Kaggle and represents bike sales data, encompassing nine interconnected tables: brands, categories, customers, orders, order items, products, staffs, stores, and stock. This rich dataset provided a comprehensive view of the bike sales operations, including product details, customer information, order history, inventory levels, and staff and store details. Effective analysis of this data is crucial for making informed business decisions in a competitive market.

Duration

The project was completed over a period of one month. This time frame included data cleaning and preparation, exploratory data analysis (EDA), SQL query development, Power BI report creation, and the compilation of findings and recommendations.

Role

I was responsible for all stages of the analysis, from understanding the business context and defining the objectives to extracting, transforming, and loading (ETL) the data, performing the analysis using SQL, visualizing the results in Power BI, and finally, creating a comprehensive report summarizing the findings and providing actionable recommendations.

Tools and Methodologies

- **SQL (MySQL):** Used for data extraction, transformation, and analysis. SQL queries were developed to aggregate data, filter records, join tables, and calculate key metrics such as sales revenue, customer demographics, and product performance.
- **Power BI:** Employed for data visualization and report creation. Power BI dashboards were designed to present the findings of the SQL analysis in a clear, interactive, and user-friendly format. This included creating charts, graphs, and tables to effectively communicate key insights.

Approach and Processes

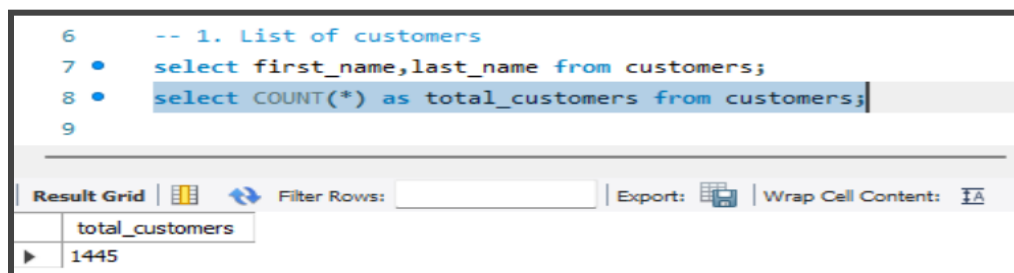
Data Cleaning and Preprocessing

The project commenced with a detailed examination of the dataset's structure, including the nine interconnected tables (brands, categories, customers, orders, order items, products, staffs, stores, and stocks). The relationships between these tables were carefully mapped out to understand how data could be effectively joined and aggregated for analysis. Data types for each column were reviewed, and potential data quality issues, such as missing values, inconsistencies, or unusual data distributions, were identified for subsequent handling.

Exploratory Data Analysis (EDA)

SQL queries formed the basis of the exploratory data analysis process. This phase was central to the project, using SQL as the primary tool for data exploration and analysis. The process involved a tight integration of query development, statistical analysis, and visualization to uncover meaningful insights from the bike sales data.

1. **Descriptive Statistics:** Basic descriptive statistics (counts) were calculated for key categorical variables using the **COUNT** aggregate function. For instance, the total number of customers, brands, and product categories were determined to provide an initial overview of the dataset's scope.



```
6      -- 1. List of customers
7 •    select first_name,last_name from customers;
8 •    select COUNT(*) as total_customers from customers;
9
```

total_customers
1445

- 2. Basic Retrieval and Filtering:** **SELECT**, **WHERE**, **ORDER BY**, and **LIMIT** clauses were used extensively for retrieving specific data subsets and sorting results. For example, queries were used to list all customers, brands, and product categories, and to find products within a specific category and orders placed on a specific date.

```
24 -- 5. Orders placed on a specific date.
25 • select * from orders where order_date="2016-01-01";
26 • select products.product_name,orders.order_date,orders.store_id,orders.staff_id from orders join products on orders.product_id=products.product_id where order_date="2016-01-01";
```

	order_id	customer_id	order_status	order_date	required_date	shipped_date	store_id	staff_id
▶	1	259	4	2016-01-01	2016-01-03	2016-01-03	1	2
	2	1212	4	2016-01-01	2016-01-04	2016-01-03	2	6

- 3. Advanced Filtering and Comparisons:** Queries were used to find the most expensive and cheapest products, to list products ordered by price, and to find orders placed by a specific customer. These queries demonstrate the use of **ORDER BY** and **LIMIT**.

```
41
42 -- 8. Most expensive product.
43 • select products.product_name,order_items.list_price from products join order_items
44 on products.product_id=order_items.product_id order by list_price desc limit 1;
45
```

	product_name	list_price
▶	Trek Domane SLR 9 Disc - 2018	11999.99

- 4. Joining Tables:** **INNER JOINs** were used to combine data from multiple related tables. This was crucial for linking sales data with product details, customer information, and store locations. For example, queries were used to list customers and their corresponding orders, products and their brand names, and order items for a specific order.

```
54 -- 11. List of customers and their corresponding orders.
55 • select customers.first_name,customers.last_name,products.product_name,order_items.quantity
56 from customers join orders on customers.customer_id=orders.customer_id join order_items on
57 orders.order_id=order_items.order_id join products on order_items.product_id=products.product_id;
58
```

	first_name	last_name	product_name	quantity
	Emmitt	Sanchez	Electra Girl's Hawaii 1 (16-inch) - 2015/2016	1
	Caren	Stephens	Surly Ice Cream Truck Frameset - 2016	1
	Caren	Stephens	Electra Girl's Hawaii 1 (16-inch) - 2015/2016	1
	Caren	Stephens	Pure Cycles Vine 8-Speed - 2016	2
	Georgetta	Hardin	Surly Ice Cream Truck Frameset - 2016	2

- 5. Data Aggregation and Trend Identification:** Data was aggregated using **GROUP BY** clauses in conjunction with aggregate functions to summarize information across different dimensions. For example, the number of products in each category was calculated to understand product distribution. This aggregation helped identify trends and patterns.

```
86 -- 17. Number of products in each category.
87 • select c.category_name, count(*) as no_of_products from categories as c join
88 products as p where c.category_id=p.category_id group by category_name;
89
```

category_name	no_of_products
Mountain Bikes	60
Electric Bikes	24
Cyclocross Bicycles	10
Cruisers Bicycles	78
Children Bicycles	59

- 6. Aggregation and Grouping with Joins:** Queries were used to find the total number of orders placed, the total number of orders placed by each customer, the top 5 customers who have placed the most orders, the average order value, the top 3 brands with the highest revenue, the customer who has spent the most money, the product that has been sold the most, the store that has generated the most revenue, the staff member who has processed the most orders, and the average order value for each product category. These queries combined joins, aggregation, and grouping to answer more complex business questions.

```
93 -- 18. Top 3 brands with the highest revenue.
94 • SELECT b.brand_name, SUM(oi.quantity * oi.list_price) as total_revenue FROM brands as b
95 JOIN products as p ON b.brand_id = p.brand_id JOIN order_items as oi ON
96 p.product_id = oi.product_id GROUP BY b.brand_name ORDER BY total_revenue DESC LIMIT 3;
97
```

brand_name	total_revenue
Trek	5129381.610000127
Electra	1344143.7899999814
Surly	1063135.8199999952

- 7. Other queries:** Several queries focused on operational aspects of the business. These included retrieving staff and store assignments, specific order details (like items within an order or orders on a given date), and analyzing sales performance by identifying top-selling products, highest-revenue stores, and staff members with the most processed orders.

```

98  -- 19. List of staff members and the store they work at.
99  • select s.staff_id,s.first_name,s.last_name,st.store_name from staffs
100  as s join stores as st on st.store_id=s.store_id;
101

```

	staff_id	first_name	last_name	store_name
▶	1	Fabiola	Jackson	Santa Cruz Bikes
	2	Mireya	Copeland	Santa Cruz Bikes
	3	Genna	Serrano	Santa Cruz Bikes
	4	Virgie	Wiggins	Santa Cruz Bikes
	5	Jannette	David	Baldwin Bikes

```

37  -- 7. List all order items for a specific order.
38  • select * from order_items where order_id=29;
39  • select products.product_name,order_items.* from products join order_items
40  on products.product_id=order_items.product_id where order_id=29;
41

```

	product_name	order_id	item_id	product_id	quantity	list_price	discount
▶	Electra Townie Original 21D - 2016	29	1	24	2	549.99	0.05
	Surly Ice Cream Truck Frameset - 2016	29	2	6	1	469.99	0.2
	Electra Cruiser 1 (24-Inch) - 2016	29	3	21	2	269.99	0.1
	Trek Remedy 29 Carbon Frameset - 2016	29	4	8	1	1799.99	0.2
	Electra Girl's Hawaii 1 (16-inch) - 2015/2016	29	5	14	2	269.99	0.05

```

122  -- 23.Staff member who has processed the most orders.
123  • select s.first_name,s.last_name,count(o.order_id) as no_of_orders from staffs as s join
124  orders as o on s.staff_id=o.staff_id group by s.first_name,s.last_name order by no_of_orders desc;
125

```

	first_name	last_name	no_of_orders
▶	Marcelene	Boyer	553
	Venita	Daniel	540
	Genna	Serrano	184
	Mireya	Copeland	164
	Kali	Vargas	88

Data Visualization and Reporting

Power BI was used to create interactive dashboards and reports, visualizing the insights derived from the SQL analysis. The Bike Sales Dashboard effectively utilizes a variety of visualizations to present key performance indicators (KPIs) and insights derived from the data. The dashboard is structured to provide a high-level overview of sales performance while also allowing for deeper exploration of specific aspects. Key features and visualizations include:

- 1. KPI Summary:** The dashboard prominently displays key metrics at the top, providing an immediate snapshot of overall performance including the total orders, total number of customers, revenue generated and quantity sold.
- 2. Revenue by Store:** A donut chart effectively illustrates the proportion of revenue generated by different stores (Baldwin Bikes, Santa Cruz Bikes,

Rowlett Bikes). The size of each slice corresponds to the revenue contribution, allowing for quick comparison of store performance.

- 3. Top Selling Brands by Quantity:** A horizontal bar chart ranks the top-selling brands based on the quantity sold. This visualization clearly shows which brands are most popular among customers, aiding in inventory and marketing decisions.
- 4. Revenue by City:** A map-based visualization uses color intensity to represent revenue generated by different cities. This provides a geographic perspective on sales performance and can highlight key urban markets.
- 5. Revenue by Year and Month:** A line chart displays the trend of revenue over time. This allows for the identification of seasonal patterns, growth trends, and the impact of any specific events or campaigns on sales.
- 6. Quantity Sold by Category:** A bar chart shows the total quantity sold for each product category. This helps understand customer preferences for different bike types (Cruisers, Mountain Bikes, Road Bikes, etc.) and informs inventory management strategies.



Challenges and Solutions

One of the main challenges I encountered with this Bike Sales dataset was its size and complexity. Comprising nine interconnected tables, the dataset required me to construct queries involving joins across three to four tables (and sometimes even more). This necessitated a deep understanding of the database schema and the relationships between tables to accurately navigate these

connections. For example, calculating total revenue per brand required joining **brands**, **products**, and **order_items** tables—a complex undertaking. To address this, I meticulously mapped out the table relationships and carefully planned my queries before execution. I also focused on writing efficient queries from the start and optimizing filtering and aggregation logic to minimize processing time. This experience significantly enhanced my ability to work with large, relational datasets and reinforced the importance of careful planning, efficient query design, and a solid grasp of database optimization techniques.

Key Findings

- 1. Sales Performance:** The business generated \$7.689 million in revenue from January 2016 to December 2018, with a total of 1615 orders and 7078 units sold. The average order value is \$1212.71. Revenue trend shows that there was a spike in sales in April 2018 followed by a steep decline.
- 2. Product Performance:** The top-selling product is the 'Electra Cruiser 1 (24-Inch) - 2016' while the 5 top-selling brands are Electra, Trek, Surly, Sun Bicycles, and Pure Cycles. The most popular product categories are Cruisers Bicycles, Mountain Bikes, and Children Bicycles.
- 3. Store Performance:** The store with the highest revenue is Baldwin Bikes.
- 4. Customer Insights:** The customer base comprises 1445 customers. The customer who has spent the most money is Pamela Newman and the top 5 customers who have placed the most orders are Tameka Fisher, Emmitt Sanchez, Pamela Newman, Mozelle Carter, and Elinore Aguilar.
- 5. Staff Performance:** The staff member who has processed the most orders is Marcelene Boyer.

Recommendations

To capitalize on identified opportunities and enhance business performance, I recommend a four-pronged strategic approach: First, **focus on our top-performing brands and categories** (Electra, Trek, Cruisers, Mountain Bikes) by optimizing inventory, developing targeted marketing campaigns, and exploring product line expansions. Second, **replicate the success of the highest-performing store (Baldwin Bikes)** across all locations by identifying and implementing its best practices, while also investing in enhancing the overall in-store experience. Third, **strengthen customer relationships** through a loyalty program and personalized marketing initiatives. Finally, **embrace data-driven decision making** by continuously monitoring sales data and KPIs to identify emerging trends and measure the effectiveness of new strategies.