

Bike Sales Analytics Dashboard

An interactive Power BI dashboard providing comprehensive insights into bike sales performance, customer demographics, and product trends through robust data preparation and advanced modeling techniques.





Executive Summary

Project Objective

Develop an interactive Power BI dashboard providing key insights into bike sales performance, customer demographics, and product trends, supported by robust data preparation and modeling.

Key Findings

- Sales heavily dominated by 'Bikes' (96% of total revenue)
- 'Road Bikes' and 'Mountain Bikes' are top-performing subcategories
- Clear seasonal sales trend peaking in late spring/early summer
- 2014 showing strong year-over-year growth
- Majority of customers concentrated in United States and Australia
- Primary customer age group: 35-49 years

Strategic Recommendation

Focus marketing efforts on the dominant 'Bikes' category during peak seasonality (April-July) and explore strategies to boost sales in underperforming categories like 'Clothing'.



Project Context & Problem Statement

Business Challenge

This project analyzes sales data for a hypothetical bike retailer to understand historical performance, identify key growth drivers, and inform strategic decisions.

The business lacked a centralized, interactive view of its sales data, making it difficult to track KPIs, identify product performance, and understand its customer base effectively.

Solution Overview

A robust data pipeline involving:

- **Python** for initial data cleansing and transformation
- **MySQL** for structured storage and relational modeling
- **Power BI** for dynamic visualization and interactive reporting

This integrated approach transforms raw data into actionable business intelligence.



Data Source & Initial Assessment

01

Data Source

Simulated bike sales datasets provided as CSV files, representing real-world business scenarios.

02

Initial State

Multiple CSV files for sales facts, product details, and customer information, each requiring integration and standardization.

03

Data Volume

Approximately 50,000 sales transactions across 3 main tables, providing substantial analytical depth.

04

Tools Used

Python (Pandas, NumPy) for loading CSV files into DataFrames and conducting initial assessment using `df.info()`, `df.describe()`, and `df.head()` methods.

The initial assessment identified missing values, duplicates, and inconsistent formats that required systematic cleaning and transformation before analysis could begin.

Data Cleaning & Transformation

The initial raw datasets (bikesales_db_fact_sale.csv, bikesales_db_dim_product.csv, bikesales_db_dim_customer.csv, dim_date.csv) required several cleaning and transformation steps to ensure data quality, consistency, and suitability for relational database loading and Power BI analysis.

1

Date Parsing

The sls_order_dt column was converted from object (string) to datetime using pd.to_datetime() to enable chronological analysis and proper date hierarchy creation in Power BI.

2

Missing Values Handling

Missing values in geographical columns (cst_country) were filled with 'Unknown' placeholder. This approach allows Power BI to display unclassified records without distorting numerical aggregates.

3

Data Type Conversion

Crucial numerical columns (sales_amount, sls_quantity) were converted to appropriate numeric types using pd.to_numeric() to enable aggregation and calculations. Key columns (cst_id, prd_key, sls_order_num) were ensured to be appropriate integer types.

4

Standardization

Text standardization performed on categorical columns: cst_country values unified to standard format, cst_gndr standardized to 'Male'/'Female', and MAINTENANCE converted to consistent 'Yes'/'No' strings for user-friendly display.

Database Design: Star Schema Architecture

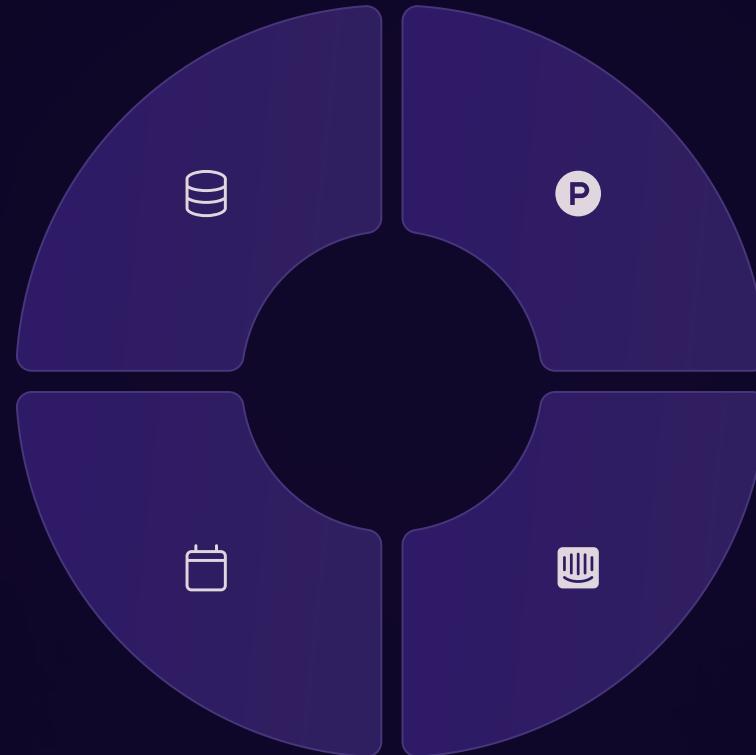
The database was designed following a **star schema methodology**, which is ideal for analytical reporting. This involves a central fact table (fact_sales) containing measurable events, linked to several dimension tables that describe the "who, what, where, when" of those events.

fact_sales (Central)

Stores individual sales transactions with measurable metrics: sls_sales (revenue) and sls_quantity. Contains foreign keys linking to dimension tables.

dim_date

Comprehensive date-related attributes (Year, Month, Day, Month Name) to support time-intelligence analysis. Uses date_key as primary key.



Cleaned and transformed data was loaded into MySQL tables using Load Data Infile method, establishing a robust, query-optimized source for Power BI that ensures data integrity and analytical flexibility.

dim_product

Provides descriptive attributes: prd_nm (product name), CAT (category), SUBCAT (subcategory), MAINTENANCE status. Uses surrogate key prd_key for data integrity.

dim_customer

Stores customer attributes: cst_gndr (gender), cst_marital_status, cst_country, BDATE (birth date). Uses surrogate key cst_key for stable relationships.

Power BI Data Modeling & DAX Measures

Data Model Design

Power BI Desktop connected directly to the MySQL database using Import mode, optimizing performance by caching data. The star schema was faithfully recreated with one-to-many relationships:

- fact_sales to dim_product (via prd_key)
- fact_sales to dim_customer (via cst_key)
- fact_sales to dim_date (via order date)

The dim_date table was enhanced using CALENDARAUTO() with custom columns for Month Name and Year. Customer Age was calculated using DATEDIFF, then binned into Age Groups for demographic analysis.

Key DAX Measures

- **Total Sales** = SUM(fact_sales[sls_sales])
- **Total Quantity** = SUM(fact_sales[sls_quantity])
- **Total Customers** = DISTINCTCOUNT(dim_customer[cst_id])
- **Total Orders** = DISTINCTCOUNT(fact_sales[sls_ord_num])
- **Avg Order Value** = DIVIDE([Total Sales], [Total Orders])
- **Avg Customer Value** = DIVIDE([Total Sales], [Total Customers])
- **Sales YTD** = TOTALYTD([Total Sales], dim_date[Date])

Dashboard Design & Visualization Strategy

The dashboard employs a **multi-page, structured design** to guide users through the data narrative logically, moving from high-level summaries to detailed performance and demographic breakdowns.



Global Sales Overview

Executive summary with KPIs, high-level trends, and market dominance. Designed for 30-second understanding.

Product Performance

Deep-dive into category hierarchies, profitability, maintenance impact, and subcategory comparisons.

Customer Demographics

Customer segmentation by age, gender, marital status, and geographical distribution.

Visual Design Philosophy

Theme: Dark Mode selected for modern, high-contrast appearance with deep red/purple background providing unified base.

Accent Colors: Vibrant Blue/Cyan for KPIs and primary metrics, Electric Green for dominant "Bikes" category, White text for maximum readability.

Navigation: Custom, persistent navigation pane on left side using Button visuals with Page Navigation Action for seamless user experience.

KPI Cards: Prominently placed at top with large fonts and contrasting blue background for immediate business health visibility.

Key Insights & Business Recommendations

96%

30-49

Apr-Aug

Bikes Category Dominance

Overwhelming revenue concentration in Bikes category, with Road Bikes and Mountain Bikes as primary drivers.

Core Customer Age

Largest customer segment falls within this age range, providing clear target for marketing efforts.

Peak Season

Consistent seasonal pattern with sales peaking during spring and summer months across all years.

Strategic Recommendations

1 Optimize for Peak Seasonality

Capitalize on strong seasonal peaks by strategically launching marketing campaigns and ensuring optimal inventory levels for Bikes between April and August. Investigate sales drops in early 2011 and 2013 to prevent recurrence.

2 Strategic Investment in Core Categories

Allocate additional R&D and marketing spend towards development and promotion of new bike models, particularly in high-performing Road Bikes and Mountain Bikes subcategories.

3 Targeted Demographic Marketing

Implement targeted campaigns specifically designed for the 30-49 age group, tailoring product messaging and channels to resonate with this largest customer segment.

4 Data Quality Improvement

Address 'N/A' classification issues by reviewing data collection processes. Properly classifying transactions will provide more accurate and complete picture of sales origins.

5 Expand Market Penetration

Focus on deepening market presence in countries showing high customer numbers but relatively lower total sales, particularly beyond US, Australia, and Germany.

6 Diversify Product Strategy

Develop long-term strategy to assess and potentially revitalize underperforming categories like Clothing and Accessories through market research, product innovation, or revised pricing strategies.

Conclusion & Future Enhancements

This project successfully delivered a **robust, end-to-end data analytics solution** for bike sales performance, culminating in a highly interactive and insightful Power BI dashboard. By meticulously processing raw data through Python (Pandas) for cleaning and transformation, structuring it efficiently within a MySQL relational database, and finally visualizing key metrics and trends in Power BI, a powerful analytical tool has been established.

Business Value Delivered

- **Centralized Information:** Consolidating disparate sales, product, and customer data into single, accessible platform
- **Data-Driven Decisions:** Offering immediate insights into sales trends, product performance, and customer demographics
- **Strategic Planning:** Equipping stakeholders to identify dominant product lines, understand customer segments, and detect seasonal patterns
- **Self-Service Analytics:** Enabling users to dynamically explore data through intuitive navigation and interactive filters

Future Enhancement Opportunities

Integrate Cost Data

Enable comprehensive profitability analysis across products, categories, and customer segments.

Predictive Forecasting

Leverage historical sales data to forecast future demand and optimize inventory management.

Expand Data Scope

Incorporate inventory, supply chain, or marketing campaign data for holistic business view.

Advanced Analytics

Explore customer lifetime value (CLTV) or market basket analysis to uncover further sales opportunities.

This dashboard stands as a testament to the power of integrating diverse data tools to transform raw data into actionable business intelligence, providing a clear roadmap for strategic growth and operational excellence.