Case study: Cannondale Bicycle Corporation

2025-09-10

Table of contents

# 1. Introduction

This case study is based on a business example from the course [Data Science for Business Part 1](https://university.business-science.io/p/ds4b-101-r-business-analysis-r/) offered by [Business Science](https://www.business-science.io/). It has been adapted to align with the topics discussed in Chapman and Feit ([2019](#ref-chapman_r_2019)) and implemented using the [R programming language](https://www.r-project.org/). The analysis uses the bike\_sales database included in the [sweep](https://business-science.github.io/sweep/)[[1]](#footnote-24) package.

# 2. Case study description

The purpose of this case study is to use data from the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale) to apply marketing research and analytics tools. The analysis incorporates the following topics from Chapman and Feit ([2019](#ref-chapman_r_2019)):

* **Basics of R**
  + The R programming language (Chapter 1)
  + Exploring the R language (Chapter 2)
* **Fundamentals of Data Analysis**
  + Describing data (Chapter 3)
  + Relationships between continuous variables (Chapter 4)
  + Comparing groups: tables and visualizations (Chapter 5)
  + Comparing groups: statistical tests (Chapter 6)
* **Modeling**
  + Identifying drivers of outcomes: linear models (Chapter 7)
  + Reducing data complexity (Chapter 8)
  + Segmentation: clustering (Chapter 11)

## 2.1 Instructions

Students enrolled in the course will be divided into teams and assigned to work for the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale), a company located in Wilton, Connecticut, United States, that supplies bicycle frames and related components to bike shops across the country.

Each team will be responsible for completing the same two tasks:

* Support the Research and Development (R & D) division by identifying potential new products and determining appropriate pricing based on data collected from bicycle shops.
* Support the marketing team by developing a segmentation and clustering model using data collected from bicycle shops.

### 2.1.1 Business unit, product and distribution channel

The business unit in this case study is the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale), which acts as a manufacturer and supplies bicycle shops across the United States.

The products analyzed are bicycle frames, made of either carbon or aluminum, designed for mountain or road use. These frames are further divided into specialized secondary categories (see [Section 2.1.3](#sec-products) for more details).

The distribution channel considered in this case study is the Business-to-Business (B2B) subchannel (manufacturer to retailer), rather than the Business-to-Customer (B2C) subchannel (retailer to customer) (see [Figure 1](#fig-distribution-channel)).

|  |
| --- |
| Figure 1: Distribution channel |

### 2.1.2 Bikeshops (Retailers)

The clients of the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale) are 30 bicycle shops located across 21 states and 30 cities in the United States (see [Figure 2](#fig-bike-shops-location) and [Table 1](#tbl-bicycle-shops)).

|  |
| --- |
| Figure 2: Bikeshops locations throughout the United States |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Bicycle shops   | **Bike shop** | **City** | | --- | --- | | Ithaca Mountain Climbers | Ithaca | | Kansas City 29ers | Kansas City | | Louisville Race Equipment | Louisville | | Ann Arbor Speed | Ann Arbor | | Denver Bike Shop | Denver | | Phoenix Bi-peds | Phoenix | | Minneapolis Bike Shop | Minneapolis | | New Orleans Velocipedes | New Orleans | | Portland Bi-peds | Portland | | San Francisco Cruisers | San Francisco | | Nashville Cruisers | Nashville | | Albuquerque Cycles | Albuquerque | | Oklahoma City Race Equipment | Oklahoma City | | Cincinnati Speed | Cincinnati | | Las Vegas Cycles | Las Vegas | | Seattle Race Equipment | Seattle | | Columbus Race Equipment | Columbus | | Tampa 29ers | Tampa | | Providence Bi-peds | Providence | | New York Cycles | New York | | Austin Cruisers | Austin | | Wichita Speed | Wichita | | Pittsburgh Mountain Machines | Pittsburgh | | San Antonio Bike Shop | San Antonio | | Indianapolis Velocipedes | Indianapolis | | Los Angeles Cycles | Los Angeles | | Miami Race Equipment | Miami | | Detroit Cycles | Detroit | | Philadelphia Bike Shop | Philadelphia | | Dallas Cycles | Dallas | |

The segmentation model, related to the first task described in [Section 2.1](#sec-instructions), focuses on bike shops and uses their orders as the basis for analysis. Each shop offers a range of bicycle frame models, and the objective is to identify distinct segments and shop characteristics by analyzing the patterns and behaviors in their orders placed with the manufacturer, the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale).

### 2.1.3 Products (Bicycle frames)

The [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale) manufactures 97 different models, which are sold to bicycle shops and subsequently offered to final customers.

Each model belongs to a primary category with possible values: Mountain or Road. In addition, each model is assigned to a secondary category that provides a more detailed description of the product (see [Table 2](#tbl-secondary-category)).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Secondary category labels   | **Category secondary** | | --- | | Over Mountain | | Trail | | Elite Road | | Endurance Road | | Sport | | Cross Country Race | | Cyclocross | | Triathalon | | Fat Bike | |

Considering the values of the frame material and the values of the secondary category, there are possible combinations. However, not all these combinations are currently produced by the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale). To address the first task described in [Section 2.1](#sec-instructions), the [product gap](https://en.wikipedia.org/wiki/Gap_analysis) must be identified in order to detect potential new products and determine their pricing using a statistical model.

### 2.1.4 Data

To carry out the tasks described in [Section 2.1](#sec-instructions), a transaction database is used. This database is built from invoice information and organized within a [Customer Management Relational System](https://en.wikipedia.org/wiki/Consumer_relationship_system) owned by the [Cannondale Bicycle Corporation](https://en.wikipedia.org/wiki/Cannondale).

In [Figure 3](#fig-invoice-example-transaction), an example invoice is shown representing the first 2 rows in the bike\_sales data set (See [Section 4.2](#sec-consolidated-data)). The order ID (*identity document*) is , the transaction date is --, and the bike shop is *Ithaca Mountain Climbers* located in Ithaca, New York (NY). The transaction includes order lines corresponding to products:

* Jekyll Carbon 2 – Over Mountain (Carbon)
* Trigger Carbon 2 – Over Mountain (Carbon)

In the first order line, one bicycle frame is sold at a unit price of US dollars. In the second order line, one bicycle frame is sold at a unit price of US dollars. The total value of the invoice is therefore US dollars.

|  |
| --- |
| Figure 3: Invoice example representing a transaction |

The bike\_sales database, available in the [sweep](https://business-science.github.io/sweep/) package, contains transactions and invoices. In [Figure 4](#fig-database-erd), an Entity–Relationship Diagram[[2]](#footnote-46) provides a representation of the database structure.

|  |
| --- |
| Figure 4: Database Entity Relationship Diagram (ERD) |

The database can be explored in two ways:

* As separate tables, as represented in [Figure 4](#fig-database-erd).
* As a single consolidated table, accessible through the [sweep](https://business-science.github.io/sweep/) package as bike\_sales.

In addition, [Section 4](#sec-annexes) provides a set of files that allow the data to be accessed in [CSV (comma-separated values)](https://en.wikipedia.org/wiki/Comma-separated_values) format.

#### 2.1.4.1 Variable dictionary

To understand the database as provided in the [sweep](https://business-science.github.io/sweep/) package (bike\_sales) and represented in [Figure 4](#fig-database-erd), the entities can be described as follows:

* **Product**
  + Product Id (product.id): unique product identification number
  + Model: model name of the bicycle
  + Category primary (category.primary): main bicycle category (Mountain, Road)
  + Category secondary (category.secondary): more specific bicycle category (9 categories)
  + Frame (frame): bicycle frame material (Carbon, Aluminum)
* **Retailer**
  + Bike shop Id (customer.id): unique bike shop identification number
  + Bike shop name (bikeshop.name): the client name
  + Bike shop state (bikeshop.state): state that the bike shop is located
  + Bike shop city (bikeshop.city): city that the bike shop is located
  + Latitude (latitude): geographic latitude of the bike shop location
  + Longitude (longitude): geographic longitude of the bike shop location
* **Closed order**
  + Order Id (order.id): unique order identification number
  + Order date (order.date): date the order was placed
  + Order line (order.line): sequential identification number for products on an order
  + Quantity (quantity): number of units purchased by the retailer
  + Price (price): unit price of the bicycle
  + Bike shop Id (customer.id): unique bike shop identification number
  + Product Id (product.id): unique product identification number

In the consolidated dataset, an additional variable price.ext is included, which corresponds to (price \* quantity) and serves as a helper for calculating revenue per order line.

# 3. Classroom project

The project developed throughout the course is carried out in stages over weeks. Teams generate marketing reports in [HTML format](https://quarto.org/docs/output-formats/html-basics.html) using the open-source scientific and technical publishing system [Quarto](https://quarto.org/) and primarily the [Tidyverse](https://www.tidyverse.org/) ecosystem[[3]](#footnote-59):

* **Stage 1:** Each team learns to use the [R programming language](https://www.r-project.org/) by exploring the database described in [Section 2.1.4](#sec-data). This involves calculating summary statistics, visualizing variables, and answering questions with data.
* **Stage 2:** Each team examines relationships between continuous variables, computes descriptive statistics by group, and applies statistical tests.
* **Stage 3:** Each team identifies the [product gap](https://en.wikipedia.org/wiki/Gap_analysis), estimates the price of a new product based on this gap using a linear model, and segments the bike shops with a k-means clustering algorithm.

# 4. Annexes

This section provides the files related to the data used in the case study.

## 4.1 Identity tables

The following files contain the data described in [Section 2.1.4](#sec-data) and illustrated in [Figure 4](#fig-database-erd):

* **Product**: products.csv
* **Closed Order**: closed\_orders.csv
* **Retailer**: customers.csv

## 4.2 Consolidated data

The following file contains the consolidated dataset, which can also be accessed through [sweep](https://business-science.github.io/sweep/) package as bike\_sales:

* **Bike sales**: bike\_sales.csv

# 5. References

Abba, Ihechikara Vincent. 2022. “Crow’s Foot Notation – Relationship Symbols And How to Read Diagrams.” <https://www.freecodecamp.org/news/crows-foot-notation-relationship-symbols-and-how-to-read-diagrams/>.

Chapman, Chris, and Elea McDonnell Feit. 2019. *R For Marketing Research and Analytics*. 2nd ed. 2019. Use R! Cham: Springer International Publishing : Imprint: Springer. <https://doi-org.ezproxy.umng.edu.co/10.1007/978-3-030-14316-9>.

Dancho, Matt, and Davis Vaughan. 2017. “Sweep: Tidy Tools for Forecasting.” <https://doi.org/10.32614/CRAN.package.sweep>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the Tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

1. See Dancho and Vaughan ([2017](#ref-dancho_sweep_2017)) for more details about this package [↑](#footnote-ref-24)
2. See Abba ([2022](#ref-abba_crows_2022)) [↑](#footnote-ref-46)
3. See Wickham et al. ([2019](#ref-wickham_welcome_2019-1)) for more details about this set of packages [↑](#footnote-ref-59)