Bike sales in Europe, USA, Australia, and United Kingdom data set.

Data source

The data set is the external data for me.

Data access (source from):
https://www.kaggle.com/datasets/sadiqshah/bike-sales-in-europe

This data belongs to Kaggle.com and was created and managed by Sadiq Shah. The data set is more or less trustworthy.

The data is administrative and collected by the years 2011-2016, cities and states. I think the data is collected automatically by one of the sales online platforms. Then through the API, the data analyst (Sadiq Shah) extracts it from the main source.

There is no time lag in collecting the data because this is operation data and is collected every year, month, and day.

The data set has 18 columns which contain different variables (integers and objects).

The data set has 113 036 rows.

```
#   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   Date              113036 non-null  object
 1   Day               113036 non-null  int64
 2   Month             113036 non-null  object
 3   Year              113036 non-null  int64
 4   Customer_Age      113036 non-null  int64
 5   Age_Group         113036 non-null  object
 6   Customer_Gender   113036 non-null  object
 7   Country           113036 non-null  object
 8   State             113036 non-null  object
 9   Product_Category  113036 non-null  object
 10  Sub_Category      113036 non-null  object
 11  Product           113036 non-null  object
 12  Order_Quantity    113036 non-null  int64
 13  Unit_Cost         113036 non-null  int64
 14  Unit_Price        113036 non-null  int64
 15  Profit            113036 non-null  int64
 16  Cost              113036 non-null  int64
 17  Revenue           113036 non-null  int64
```

The limitation could be the number of sales in unofficial shops, which could not provide official statistics.

Of course, there could be a bias in the data set, if some retailer provides untrue information. But for now, I do not know.

Data had been collected frequently during the period of 5 Years.

The data set could have some manual errors when it was transferred, saved, or transformed. However, the percentage of errors and mistakes is low.

Why did I choose this data set?

This data set provides a comprehensive collection of sales data of the wide range of product categories split by region. My primary focus is to uncover valuable business insights and answer key questions that can drive strategic decisions for bike sales in these markets.

Business Questions and Insights

Here are some of the business questions I aim to answer through this analysis:

Sales Trends: What are the overall sales trends for bikes in Europe and North America? Are there noticeable peaks or troughs during specific times of the year?

Regional Performance: How do bike sales compare between Europe and North America? Which region shows higher sales volume and profitability?

Market Segmentation: What are the key customer segments driving bike sales in each region? How do demographic factors and purchasing behaviors differ between these segments?

Product Performance: Which specific bike models or sub-categories are performing well in each region? Are there any products that consistently outperform others?

Data Profile

Clean data and consistency check.

1. Duplicates: 1000 rows have been deleted.
2. Not a null: data set has no not a null information (cells)
3. Mix-type: data set has no mix-type variables
4. No missing data.
5. Outliers. As we have outliers in Column Profit, Cost, and Revenue we need to address them correctly. We need to identify them as outliers or fair data.

	Data Types			
Variables	Time-variant / -invariant	Structured / Unstructured	Qualitative / Quantitative	Qualitative: Nominal / Ordinal Quantitative: Discrete / Continuous
Date	time-variant	structured	quantitative	continuous
Day	time-variant	structured	quantitative	continuous
Month	time-variant	structured	Qualitative	continuous
Year	time-variant	structured	quantitative	continuous
Customer age	time-variant	structured	quantitative	continuous
Age Group	time-variant	structured	quantitative	continuous
Gender	time-invariant	structured	Qualitative	nominal
Country	time-invariant	structured	Qualitative	nominal
State	time-invariant	structured	Qualitative	nominal
Product Category	time-invariant	structured	Qualitative	nominal
Sub Category	time-invariant	structured	Qualitative	nominal
Product	time-invariant	structured	Qualitative	nominal
Order Quantity	time-variant	structured	quantitative	continuous
Unit Cost	time-variant	structured	quantitative	continuous
Unit Price	time-variant	structured	quantitative	continuous
Profit	time-variant	structured	quantitative	continuous
Cost	time-variant	structured	quantitative	continuous
Revenue	time-variant	structured	quantitative	continuous

Motivation

The motivation behind this project stems from the need to gain a deeper understanding of the bike sales market in Europe and North America. In recent years, the bicycle industry has seen

significant growth due to increasing awareness of environmental sustainability, health benefits of cycling, and the surge in demand for alternative modes of transportation.

Objective

The primary objective of this project is to perform a comprehensive analysis of bike sales data to extract actionable insights that can drive business growth and operational efficiency. I am going to do the following type of analysis:

Geographic, demographic, product-specific, temporal, and financial performance.

Scope

In my data set, I have resources and limitations.

a. the years 2011-2016.
b. the regions are France and Germany, the United Kingdom, Australia, and the USA.
c. age 17-87
d. age group: Youth less than < 25, Young adults (24-35), Adults (35-64), Seniors (64+)
e. gender: Male, Female.

It is good to break down the project into the tasks (type of analysis) and define the deliverables that'll be executed to meet goals and stakeholder requirements and deliver the project successfully.

By the end of this project, we aim to deliver a comprehensive set of insights and strategic recommendations that can inform business decisions and drive growth in the bike sales market in Europe and North America.