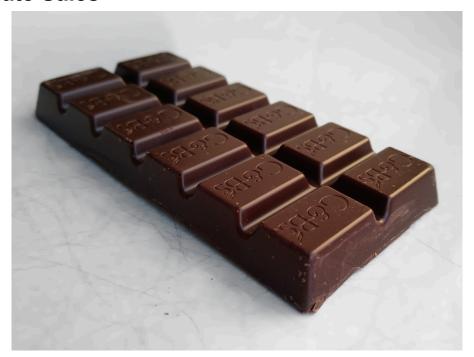


Data Science Assignment Final Data Science Project Overview

Chocolate Sales



Problem: Unwrapping Insights from Chocolate Sales Data

What influences chocolate sales the most—price, brand, season, or something else entirely?

In this project, you'll explore the **Chocolate Sales Dataset**, which includes detailed information on chocolate products, pricing, sales quantities, and customer behavior. Your goal is to analyze sales trends, identify top-performing products, and generate data-driven strategies to boost future sales.

This is a business-focused project that blends market analysis with predictive modeling—perfect for practicing real-world retail analytics.

You'll work on tasks such as:



- Data cleaning: formatting dates, handling missing values, standardizing product categories
- Exploratory data analysis (EDA): uncovering patterns in sales by product type, brand, region, or time
- Optional modeling: forecasting sales or clustering customer/product profiles
- Modeling: Training a Regression Model
- Evaluation: Use R2 and RMSE

Skills you'll practice:

- Business analytics and product performance evaluation
- Predictive modeling or segmentation strategies
- Turning raw sales data into actionable business recommendations

If you've ever wondered how data drives decisions in retail or product marketing, this project gives you a sweet introduction.

Data:

The dataset chocolate_sales.csv has features related to chocolate sale translations

Client Information (Demographics & Socioeconomic)

Sales Person: Name of the salesperson Responsible
 Product: Name of the chocolate product sold
 Date: The transaction date of the chocolate
 Amount: Total revenue generated from the sale
 Boxes Shipped: Number of chocolate boxes shipped

Suggestions:



- Try using only a few sales people in your model, rather than attempt to create a dummy variable for all of them.
- Maybe look at the transactions on a monthly basis, rather than daily
- Create new features to improve predictability (seasonal indicators, price-per-unit, brand category, etc.