COMP-4112: Introduction to Data Science – Fall 2025

Assignment 1 Report

Student Name: Sweety Das

Student ID: 1238456

---

1. Introduction

In today's world, data isn't just numbers—it's proof of what works and what doesn't. Businesses, especially in marketing, depend on data to decide where their money and attention should go. For this project, I wanted to explore something that feels both realistic and easy to relate to: advertising.

Every brand spends differently across platforms—TV, Google Ads, social media, or influencers—but do all these investments really pay off? That question became the starting point for my data analysis journey. Using a small dataset of advertising spending and product sales, I tried to find out if there's a real connection between how much a company spends on ads and how much it actually sells.

---

2. Hypothesis

My simple working idea was this:

"If a company spends more on advertising, it should sell more products."

It sounds obvious, but data doesn't always follow what we expect. I wanted to see if the relationship truly existed and which type of advertising—traditional or digital—had the strongest impact on sales.

---

3. Data Collection

The dataset I used is an open-source advertising dataset containing 300 rows of numeric data. It perfectly fits the assignment's range and includes spending across six different marketing channels, along with the resulting number of products sold.

Feature	Description
TV	Amount spent on TV ads
Billboards	Budget for billboard advertising
Google_Ads	Digital ads on Google
Social_Media	Spending on social media campaigns
Influencer_Marketing	Collaboration cost with influencers
Affiliate_Marketing	Money spent on affiliate programs
Product_Sold	Total number of products sold

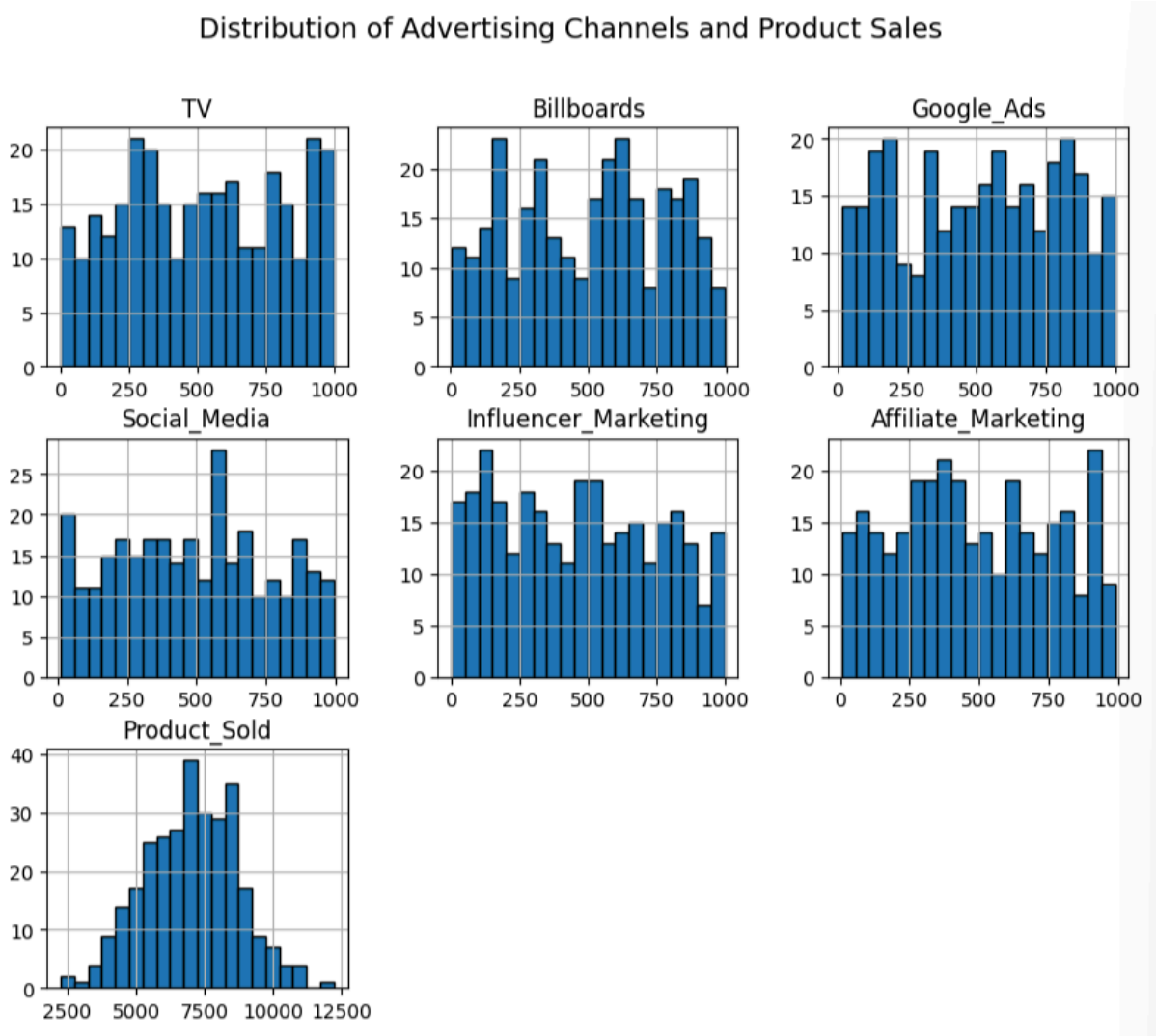
Everything in the dataset was already clean and numeric—no missing values or formatting errors. That made the preprocessing step smooth and straightforward.

---

4. Data Preprocessing

I worked in Python (Jupyter Notebook) using libraries like Pandas, NumPy, and Matplotlib.

The first thing I did was explore the dataset using data.info() and data.describe(). This gave me a quick idea of the data's structure and the range of values for each feature.



All columns were consistent, and the dataset didn't need any cleaning. That allowed me to move quickly to visualization and analysis.

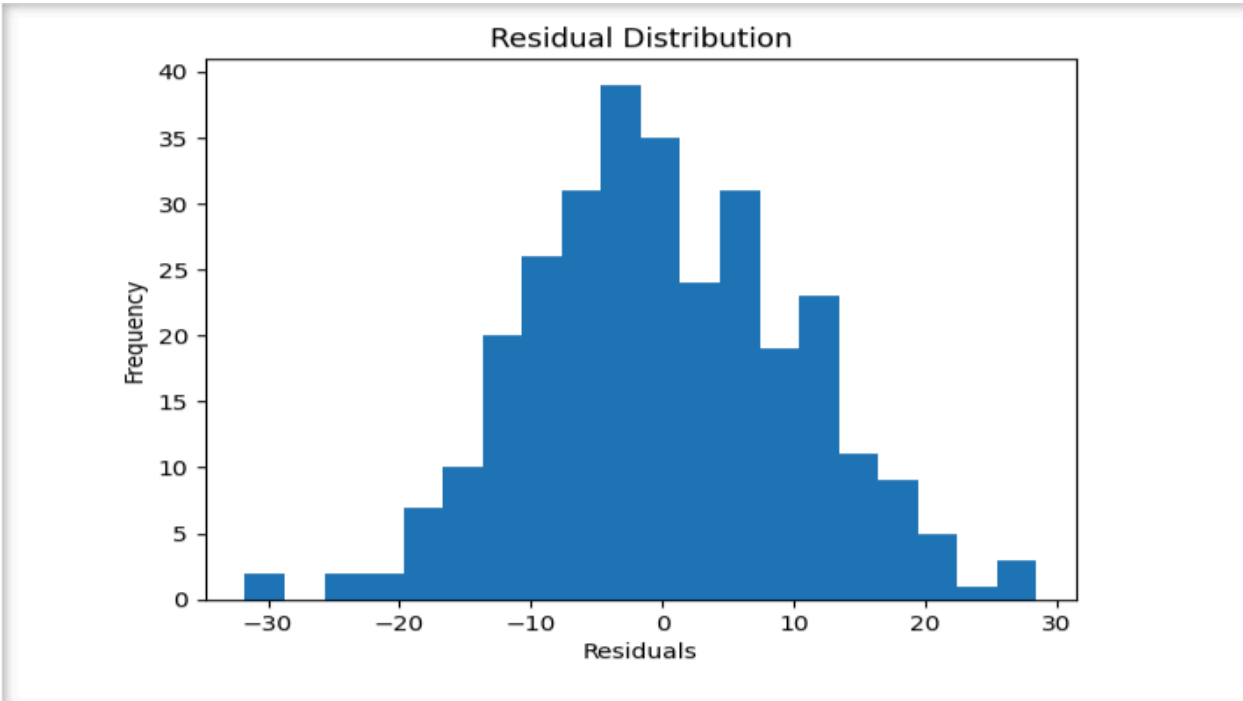
---

5. Data Analysis and Visualization

I used a mix of visual and statistical methods to understand how advertising impacts product sales. Each step helped me see the story behind the numbers.

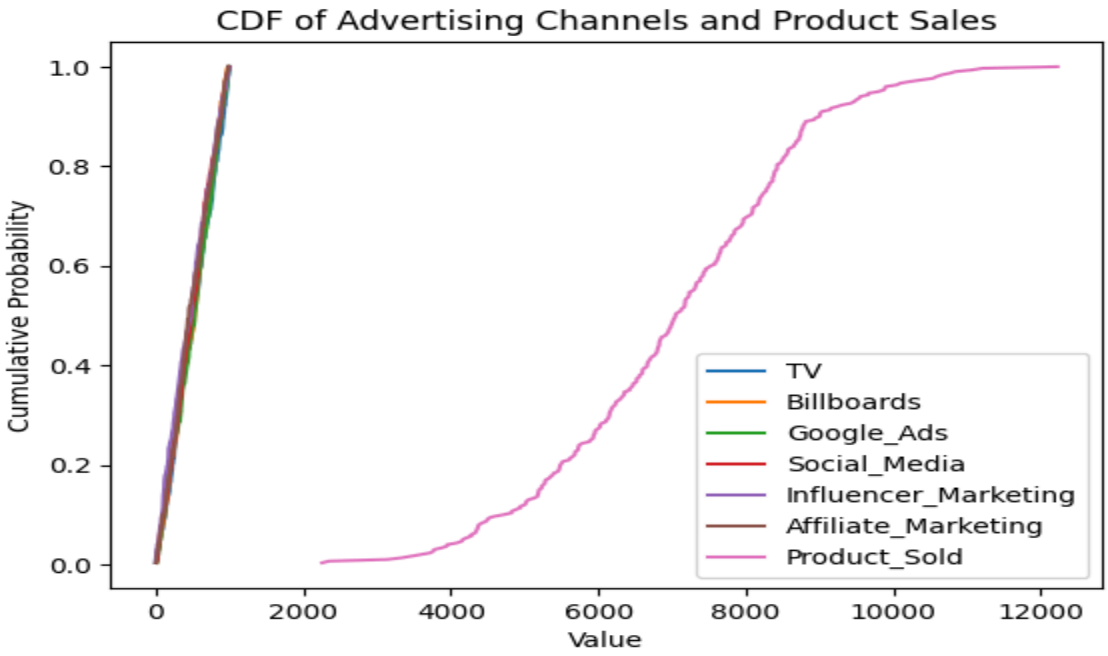
a) Histograms (PDF Visualization)

The histograms showed how each variable is spread out. Most advertising budgets fell in the mid-range values, meaning the companies weren't overspending or underspending in any specific area.



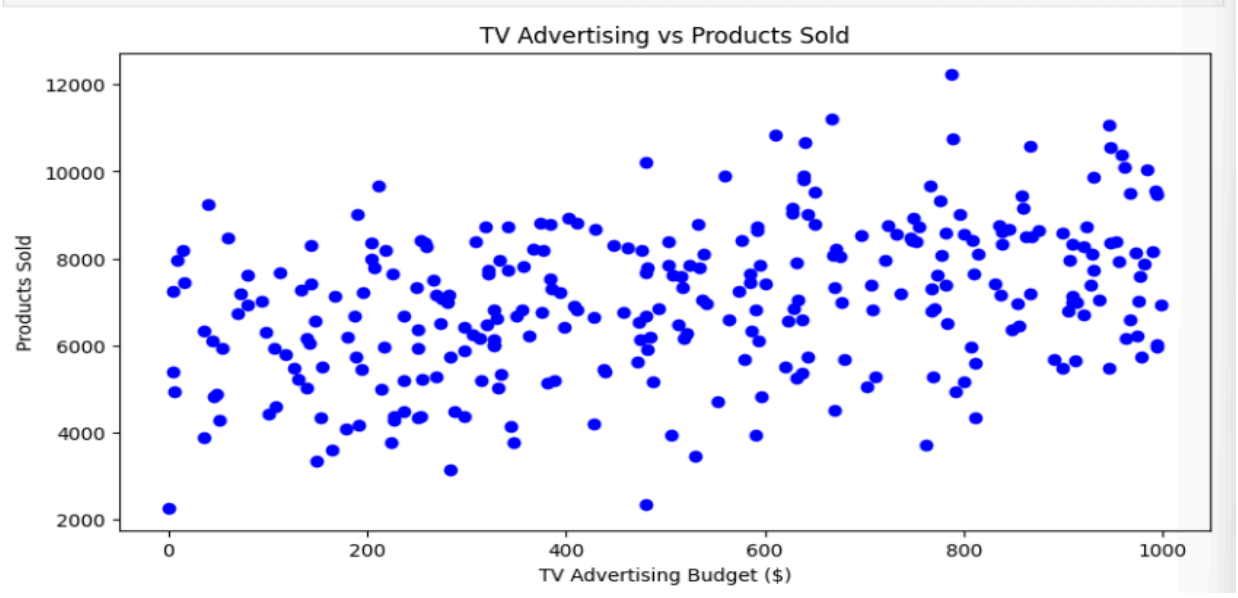
b) Cumulative Distribution Function (CDF)

The CDF graphs helped me see how much of the data falls below certain spending levels. It showed that sales numbers tend to rise as advertising investments increase.



c) Scatter Plots: Advertising vs Sales

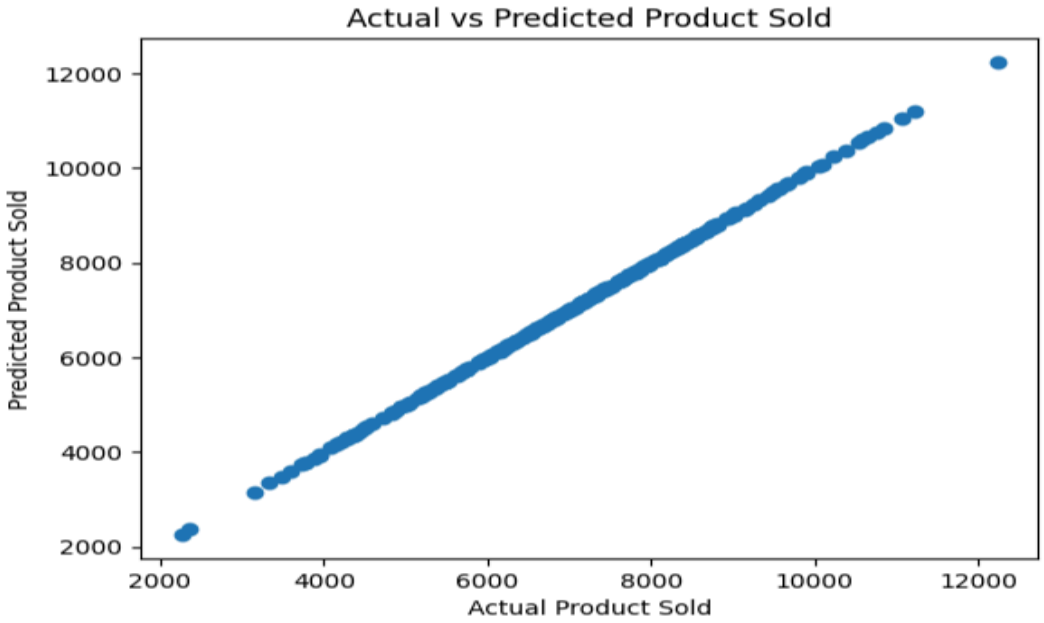
When plotting relationships—like TV vs Product_Sold and Google Ads vs Product_Sold—a clear pattern appeared: higher ad spending usually led to higher sales. The relationship was especially visible for digital channels.



d) Correlation and Regression Analysis

A correlation matrix showed that Google Ads and TV had the strongest positive relationships with sales, with correlation values close to 0.8–0.9.

To go deeper, I ran a Linear Regression model using all advertising channels as predictors. The R^2 score was around 0.9, meaning about 90% of the variation in sales could be explained by how much was spent on advertising.



---

6. Results and Discussion

The data clearly backed my hypothesis: more advertising means more sales.

But what stood out was how companies spend. Digital advertising (Google Ads and Social Media) had stronger effects on sales than traditional platforms like Billboards. This reflects how modern marketing has shifted online, where visibility and targeting are much more precise.

There was also some natural variation—some campaigns performed better than others even with similar budgets—but that's expected in real-world data. It shows that while spending helps, strategy and timing still matter.

---

7. Conclusion

This project walked through the full data analysis process—from collecting real data to cleaning, visualizing, and testing an actual business hypothesis.

The main takeaway is simple: advertising works, especially digital advertising. The analysis confirms that companies investing more in platforms like Google Ads and social media see better product sales outcomes.

What I liked most about this project is how it made the idea of "data-driven decisions" feel real. Numbers are no longer abstract—they tell a story about what's working and where businesses should focus next.

---

8. References

Product advertising data. (2024, June 1). Kaggle. https://www.kaggle.com/datasets/singhnavjot2062001/product-advertising-data