Morris Patin

Professor Ayyagari

Intro to Machine Learning

21 October 2025

Lab 4 Report

Throughout this exploratory data analysis project, I learned how to interpret data, uncover meaningful insights, and visualize patterns that reflect real-world customer behavior. This activity gave me a hands-on opportunity to apply everything I've learned about cleaning, summarizing, and exploring data using Python. By working through each challenge, I gained a much clearer picture of how education, gender, and city tier can influence a customer's purchasing habits.

In **Challenge 1**, I created a bar chart to show the relationship between education level and average order value. The visualization revealed a clear trend: customers with higher education levels tended to have higher average order values. This finding suggests that education may be linked to income or lifestyle factors that influence spending behavior. Visualizing this relationship helped me see how important demographic variables can be when trying to understand purchasing power and customer value.

Moving on to **Challenge 2**, I analyzed different combinations of gender, education, and city tier to find which segment represented the most valuable customers. By grouping the data and calculating average total spending, I discovered that highly educated customers living in Tier 1

cities, particularly males, spent the most overall. This kind of insight could be very useful for a business that wants to focus its marketing or promotional efforts on its top-spending customer groups. It also showed me how segmentation can transform raw data into clear and actionable insights.

In **Challenge 3**, I made several recommendations to improve data quality. My first suggestion was to handle missing values and ensure that key columns like education and total spending are complete. Next, I recommended standardizing categorical data to avoid inconsistencies (for example, making sure "Bachelor's" and "bachelors" are labeled the same way). Lastly, I noted that duplicate records should be checked and removed, since they can affect the accuracy of spending calculations. Paying attention to these details helps make future analyses more reliable and accurate.

Overall, this lab helped me strengthen my data analysis skills and understand how to extract insights that matter. I practiced loading and exploring datasets, cleaning and validating data, and visualizing trends that tell a story. I also learned the importance of presenting results in a clear and organized way so that others can easily interpret them. This project didn't just teach me how to code in pandas or create charts—it showed me how data can explain real business questions and support smart decision-making. Completing this analysis made me feel more confident about working with data and applying these techniques to future projects.