Reflective Journal: Video Game Sales Prediction Lab

Working on this lab gave me a deeper understanding of how data science can uncover meaningful insights in real-world contexts, like the video game industry. One of the most interesting patterns I noticed was how global sales relate to regional performance. For example, some games sold extremely well worldwide but had uneven success in specific regions, which surprised me. I realized that a game's success isn't just about overall sales, it's also influenced by platform, genre, and publisher, showing that multiple factors interact to determine a hit game.

Evaluating the model's performance was an insightful experience. The model was fairly accurate at predicting which games would be hits, but there were some limitations. While accuracy was decent, precision and recall revealed that it sometimes misclassified games that had unusual sales patterns. I noticed that hyperparameter tuning could improve performance by adjusting the model to better recognize patterns in the data, highlighting how small changes in settings can have a meaningful impact on predictions.

From a business perspective, these insights could be valuable to game developers and publishers. For instance, understanding which genres or platforms are more likely to produce hit games could guide development priorities. Regional sales ratios could help companies decide where to focus marketing efforts. I think actionable recommendations would include targeting high-performing regions for specific game types and planning release strategies around platforms with historically high sales.

This lab also introduced me to several new concepts and techniques. Troubleshooting file handling issues taught me the importance of file paths and correct file types, especially when resolving FileNotFoundError and BadZipFile errors. Understanding error messages and variable scope was another key lesson; the NameError we encountered reminded me that a model must be properly defined and trained before making predictions. The most challenging part was ensuring the proper sequence of operations—from loading and preprocessing the data to training the model and making predictions—but careful reading of error messages and step-by-step verification helped me overcome these hurdles.

Critically reflecting on the limitations, I see that the model could improve with more detailed data, such as user reviews, game ratings, or marketing spend. Additionally, exploring more advanced modeling techniques like ensemble methods or neural networks could capture complex patterns better than the basic model we used.

Finally, this practical experience tied closely to theoretical concepts from class. We learned about feature engineering, target encoding, and evaluation metrics in theory, and seeing them applied to real data helped solidify my understanding. For example, calculating precision, recall, and ROC AUC made me appreciate how evaluation metrics give different insights into model performance. I'd say in total, this lab strengthened my ability to think critically about data, troubleshoot issues, and connect theory to practice in data science.