Sarah Theriot

DSC550

Final Write Up

Understanding the Relationship Between Apple Product Sales and Services Revenue Across U.S. Regions

Introduction

Apple is one of the largest technology companies in the world, best known for its flagship products like the iPhone, iPad, and Mac, as well as its growing Services segment. These product lines don't perform the same across different regions, which makes it valuable to understand where and how Apple generates the most sales. This project looks at how product sales are distributed across U.S. states and global regions, with the goal of identifying which areas contribute most to Apple's revenue and how product-level sales relate to Services income.

Recognizing these patterns is useful from a business strategy perspective. If certain states see especially strong iPad sales, for instance, Apple could invest more in marketing or offer product bundles in those areas. On the other hand, if one region shows a clear connection between product purchases and increased Services revenue, it might be a good opportunity to introduce or expand subscription offerings. The more Apple understands how product sales and Services revenue interact geographically, the better it can target investments, promotions, and innovation.

If I were presenting this analysis to a group of Apple stakeholders, I'd describe it as a data-backed approach to making smarter regional decisions. Instead of treating every market the same, this analysis uses sales trends to guide where resources should go. That can lead to better performance and more efficient spending.

The dataset for this project came from a CSV file provided in a course. It included Apple's product unit sales and Services revenue, organized by state and region. The product categories in the data were iPhone, iPad, Mac, and Wearables, with revenue figures for Services. Each row reflected a unique combination of region and sales data, allowing for both geographic and product-based analysis.

Milestone 1: Exploratory Data Analysis

The first phase of the project focused on getting familiar with the data and identifying any early patterns. The dataset showed product sales in millions of units and Services revenue in billions of dollars. Each entry was labeled with a U.S. state and broader region, which made it possible to compare how different parts of the country perform across product categories.

To get a quick overview, I used the .describe() function. This confirmed that iPhones had the highest total sales by far, which is consistent with Apple's overall business performance. Services revenue also appeared to be a significant part of the picture. I created visualizations to explore how product sales varied by region, and certain states—

like California, New York, Florida, and Texas—consistently showed higher sales across multiple categories. These visuals helped identify key markets where Apple is already strong and others where there may be potential for growth.

Initially, my plan was to build a model that could predict total product sales. However, after receiving feedback from my instructor, I realized that wasn't the best approach. Since total product sales are just the sum of individual product sales, trying to predict that wouldn't provide any new insights. Based on this, I shifted the focus of the modeling phase to predict Services revenue instead. That change allowed for more meaningful analysis and a better understanding of how product sales relate to revenue generated through Apple's service ecosystem.

Milestone 2: Data Preparation and Feature Engineering

With the goal of modeling Services revenue, I moved on to preparing the data. I began by standardizing the numeric columns for iPhone, iPad, Mac, and Wearables using StandardScaler. This was necessary to make sure that the model didn't give extra weight to features with larger numeric values just because of their scale. By normalizing the data, all features could be compared more evenly.

After that, I explored some feature engineering. I created a column for Total Product Sales, even though it was simply the sum of the individual products. While it wasn't useful for the model, it did help with quick referencing during analysis. I also calculated an iPhone Ratio, which measures what portion of a region's total product sales came from iPhones.

This was interesting for understanding product dominance, though neither of these engineered features ended up being directly useful in the predictive model.

The dataset didn't require much cleaning, which made this step straightforward. I used .isnull().sum() to confirm that there were no missing values. Initially, I thought I would need to create dummy variables for the State and Region columns, but those had already been generated in the dataset. Because of that, I avoided re-encoding them, which helped prevent unnecessary errors or redundancy in the model.

Milestone 3: Modeling and Evaluation

Once the data was ready, I moved on to the modeling phase. The goal was to predict Services revenue using the four product sales features: iPhone, iPad, Mac, and Wearables. I selected linear regression for this part of the project because it is easy to interpret and provides direct insights into how each independent variable relates to the target.

I trained the model on 80 percent of the data and tested it on the remaining 20 percent. The model's performance wasn't particularly strong. The R-squared value came out close to zero, and the Mean Squared Error was about 18.76. While this wasn't ideal from a predictive standpoint, the model still revealed some useful relationships. Wearables had the highest positive coefficient, followed by Mac and iPad. Interestingly, the iPhone variable had a slightly negative coefficient. That result was unexpected but possibly reflects regions where iPhone sales are high but don't lead to increased use of Apple's paid services. It might mean that those users are less likely to subscribe to things like iCloud, Apple Music, or other paid features.

This outcome highlighted an important insight: not all products contribute equally to Services revenue. That's exactly the kind of information that could help Apple fine-tune its regional strategy. If Wearables are more likely to lead to service usage, it might make sense to bundle service offerings with those products in regions where sales are already strong.

Conclusion

This project provided a clearer picture of how product-level sales influence Services revenue across different states and regions. From the beginning, it was obvious that iPhones dominate in terms of units sold. However, the modeling process made it clear that sales volume doesn't always translate into increased revenue from Services. The strongest relationships came from Wearables and Macs, which could have implications for how Apple promotes and packages those products moving forward.

Even though the linear regression model didn't perform well as a predictor, it still added value. Its simplicity allowed me to see which products are more closely tied to Services revenue, even if the overall accuracy was low. In future versions of this project, I would try more advanced models such as decision trees or random forests. These models are better at capturing complex patterns and might reveal interactions between features that linear regression missed.

The dataset itself also had limitations. It only included unit sales and didn't account for factors like pricing, device usage, or customer engagement. These elements probably play a big role in determining whether someone subscribes to Apple services. If I had access to more data, I would look at things like App Store usage, subscription history, and how long customers keep their devices. Adding that type of information could make future models much more effective.

Time is another factor worth exploring. This dataset only captured a snapshot, which means it can’t account for changes in trends or customer behavior over time. For example, maybe iPhone users didn’t engage with services during the timeframe of this dataset, but that could change with new features or software updates. A time series model using quarterly or monthly data might offer more accurate and dynamic insights.

So, is this model ready for deployment? Not yet. But it is a strong starting point. It highlights areas where Apple’s Services revenue is stronger or weaker and shows where further analysis is needed. With better data and more advanced modeling techniques, there’s real potential to turn this into a tool Apple could actually use.

My final recommendation is that Apple should continue to explore the link between product sales and Services usage. Knowing which products lead to higher engagement can shape everything from marketing and pricing to product development. For instance, if Wearables drive Services revenue more than other products, Apple could offer bundled subscriptions or tailor its promotions around that connection. Regional differences are also important to consider. What works in California might not work in another country, and this project helped uncover those kinds of patterns. Even though this was just a class project, it shows how valuable data-driven decisions can be when it comes to growing a business.