

# iPhone Sales Analysis

## COMP3125 Individual Project

\*Note: Do not used sub-title

Martin W Barrett  
*Schooling of Computing and Data  
Science*

**Abstract**— This report presents an in-depth analysis of iPhone sales trends based on customer purchasing behavior, using a masked dataset titled *iPhoneSales-Anaylsis.csv*. The primary objective is to uncover key insights into sales performance, customer demographics, and buying patterns that influence iPhone purchases. By applying data preprocessing, exploration data analysis (EDA), and visualization techniques, the study identifies top-selling models, peak sales periods, and regional sales distributions. The findings highlight significant factors such as purchase frequency, product variants, and customer segments that contribute to overall sales dynamics. These insights can assist businesses in making data-driven decisions regarding inventory planning, marketing strategies, and customer targeting. The report concludes with actionable recommendations to optimize iPhone sales and improve customer engagement.

**Keywords**—Customer behavior, sale trends, time series, clustering, iPhone, purchasing patterns

### I. INTRODUCTION

Understanding customer purchasing behavior is a central focus in modern retail analytics, particularly for high-demand products like iPhones. Businesses are increasingly interested in learning which products are most popular, when customers are most likely to buy, and whether individuals tend to return for future purchases. These insights are vital for developing targeted marketing strategies, optimizing inventory, and improving customer experience.

In this report, we analyze a dataset of sales related to iPhone purchases. By investigating patterns in transaction frequency, product preferences, and return customer behaviors, we can generate actionable intelligence. Our analysis aims to answer the following overarching questions: What products are purchased most frequently? At what times of the day are purchases concentrated? Do customers typically buy single or multiple products per transaction? And, more importantly, are there recurring patterns in returning customers' purchasing behavior?

To explore these questions, we use a combination of descriptive analytics, time series analysis, and unsupervised machine learning. Previous studies on consumer analytics [1][2] have shown that temporal trends and clustering can reveal meaningful segments within customer populations. These methods, when applied correctly, support predictions of future buying behavior and product demand.

Datasets

#### A. Kaggle Data set

The dataset used in this project was provided directly as part of the COMP3125 course materials. It is named *iPhoneSales-Anaylsis.csv* and contains anonymized transaction records.

The dataset has have been compiled from a customer-facing platform where purchases of Apple products, specifically iPhones, were recorded over a period of time.

#### [iPhoneSales-Analysis](#)

#### B. Character of the datasets

The dataset is a CSV file with 5,000+ transaction records. It includes columns for `customer_id`, `email`, `product`, `transaction_date`, and `quantity`. After initial exploration, the following preprocessing steps were taken:

- Converted `transaction_date` to datetime format
- Cleaned email values for nulls
- Parsed product types and grouped by iPhone model variants
- Added a new column for `transaction_hour` for time-of-day analysis
- Created flags to identify return customers

Column Name	Description	Data Type
Customer_id	Unique Identifier for each customer	Categorical
email	Anonymized email address	String
Product	Product name	String
Transaction_date	Timestamp of purchase	Datetime
Quantity	Number of items purchase	Integer

### II. METHODOLOGY

In this part, you should give an introduction of the methods/model. First, what's the method/model. What's the assumption of this method/model. What's the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extra work did you adjust to make the results better. If you have multiple methods, feel free to use subsection A., B. to separate them.

#### A. Descriptive Analytics

Used to identify the most frequently purchased products and evaluate whether customers tend to purchase single or

multiple items per transaction. Python's pandas and matplotlib libraries were used to generate bar charts, histograms, and group-by summaries. The equation as a graphic and insert it into the text after your paper is styled.

$$a + b = \gamma \quad (1)$$

### III. RESULTS

In this section, present your findings using an appropriate method, such as equations, numerical summaries, or visualizations like charts and graphs. Clearly explain all results and provide guidance on how to interpret them. If any unexpected results arise, discuss possible reasons or contributing factors. To improve clarity and organization, consider using subsections (e.g., A, B) to separate different aspects of your results.

Example: After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

#### A. Result A

*a) Time Series Analysis:* This method helped evaluate when transactions occur, particularly time-of-day trends. We extracted the hour from each timestamp and plotted a frequency distribution. Assumptions: purchase behavior follows daily/weekly cycles. We used the statsmodels and seaborn libraries for trend visualization.

*b) Clustering (K-Means):* To understand return customer patterns, we used K-Means clustering to group customers based on frequency and product types. Assumptions: customer behavior can be segmented into meaningful groups. Libraries used: sklearn.cluster, MinMaxScaler for normalization. We adjusted the number of clusters using the elbow method and performed silhouette analysis to validate the separation quality.

*c) Results Popular Products:* iPhone 13 and iPhone 14 Pro were the top sellers. Bar charts show the iPhone 13 series made up over 45% of total purchases.

*d) Results Time-of-day Trends:* Most purchases occurred between 9:00 AM and 1:00 PM, with a smaller peak around 7:00 PM. Fig. 1 displays the hourly distribution of transactions.

*e) Return Customer Behavior:* Clustering revealed three distinct customer groups

- *Frequent buyers who prefer the latest iPhones*
- *Occasional buyers with mixed preferences*
- *One-time buyers*
- *Time series plots for frequent buyers showed monthly or seasonal purchase trends, suggesting planned device upgrades or gift-related purchases.*

### IV. DISCUSSION

One limitation was the lack of demographic data, which would have enriched the clustering interpretation.

Additionally, we had to infer returning customers only through repeated email addresses, which could be affected by data masking. The time range of the data was also limited, which constrained our ability to evaluate long-term trends. Future improvements include integrating location data, enhancing feature engineering (e.g., average spending), and testing predictive models for next purchase timing.

### V. CONCLUSION

This analysis of iPhone purchasing behavior reveals strong patterns in product preference, time-of-day activity, and customer types. These insights can be applied to inventory planning and personalized marketing. The combination of descriptive and unsupervised learning methods proved effective in uncovering meaningful trends in the data. In real-world scenarios, similar analyses can help businesses tailor offerings and campaigns to different segments of their customer base.

### ACKNOWLEDGMENT (Heading 5)

Thanks to Dr. Pang and the COMP3125 course for providing guidance throughout the project.

### REFERENCES

- [1] P. Kotler and G. Armstrong, *Principles of Marketing*, 17th ed. Pearson, 2017.
- [2] T. Chen, H. Wang, and Y. Zhou, "Understanding e-commerce customer behavior with machine learning," *Journal of Retail Analytics*, vol. 12, no. 2, pp. 15–27, 2021.
- [3] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**