USTH

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

Bachelor Thesis

Analysis of new sales performance of the "Ultra Fast" product and revenue forecasting

ID - Student - Major

22BI13220 -   Nguyen Minh Khoi - Data Science

Supervisor:

Bui Thi Thu Hoa

Academic year: 2022 - 2025

Hanoi, July 2025

Acknowledgements

I would like to extend my deep gratitude to Ms Bui Thi Thu Hoa, my supervisor. My project would never have been possible without her wisdom and experience guiding me along the way. Her wealth of experience and advice was critical to overcoming obstacles and keeping me on track with my mission. Despite her busy schedule, she always made time for me with patience and precious advice. I deeply appreciate her unwavering support and are honored to have had the privilege of working under her supervision.

Abstract

Contents

List of Figures

List of Tables

1 Introduction

This section begins by exploring the key challenges related to the sales performance of the "Ultra Fast" product and highlights the importance of accurate revenue forecasting to guide future decisions.

1.1 Ultra Fast

Ultra Fast (UF) [1] is an advanced gaming network optimization solution designed to deliver the ultimate online gaming experience. It leverages cutting-edge Artificial Intelligence (AI) and Big Data technologies to intelligently separate gaming traffic from other online activities, prioritizing it to ensure smooth and uninterrupted gameplay.

With its Game Connection Optimization system, UF analyzes and determines the best routing paths to connect gamers with global game servers, minimizing latency and maximizing stability.

Best of all, no installation is required. Simply use a compatible modem or Access Point (AP) along with a single UF license, and up to five gaming devices in your household will be automatically optimized simultaneously.

1.2 Problem Statement

The launch of the UF product — built to reduce ping and lag for gamers — marked a strategic move to meet the growing demand for smoother online gaming. While the product addresses a clear market need, understanding its early sales performance is crucial to evaluating its success and guiding future decisions.

At this point, there's limited clarity on how well the product is doing in terms of sales trends, customer adoption, and regional performance. Without this insight, it becomes challenging for the company to assess the product's reception or plan effectively for its growth.

In addition, the absence of a reliable revenue forecasting model makes it difficult to project future earnings, allocate resources wisely, or refine marketing strategies. To move forward with confidence, a deeper analysis of current sales data and well-informed revenue predictions are essential. This project aims to uncover those insights and provide a clearer picture of where UF stands — and where it could be heading.

1.3 Goal

2 Methodology

In this section, we will list all the tools and techniques used in this project, the reasons why they are chosen with overall system architecture and the detailed use cases implemen tations.

2.1 Research Approach

This project employs both descriptive and predictive approaches. Initially, descriptive analysis is conducted to provide a comprehensive overview of customer demographics, sales performance, and package popularity. Through summary statistics, visualizations, and trend analysis, we gain insights into historical patterns and business performance.

Building upon this foundation, predictive modeling techniques are applied to anticipate future outcomes. Specifically, machine learning models are used to predict customer churn and forecast sales revenue. These models leverage historical data patterns identified in the descriptive phase to generate actionable predictions, supporting strategic decision-making

This study follows a quantitative approach, focusing on numerical data analysis. Statistical techniques and machine learning models are applied to structured datasets to measure customer behavior, sales performance, and predict future trends based on historical data.

This combined approach ensures both a thorough understanding of past business performance and the ability to make data-driven forecasts, which is essential for proactive decision-making in a competitive market.

2.2 Data Analysis Pipeline

The project follows a structured data analysis pipeline comprising four main stages: data collection, preprocessing, analysis, and modeling. Data is initially collected from business databases and cleaned to address inconsistencies and missing values. Preprocessing involves transforming the data into a suitable format for analysis, including normalization and feature selection. Descriptive analysis is then conducted to explore key trends, followed by the development of predictive models to forecast customer churn and revenue. Each step builds upon the previous, ensuring accurate and actionable insights.
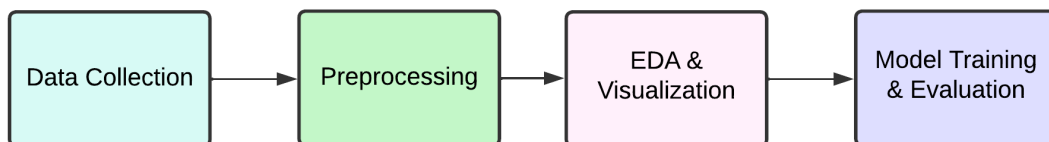
Figure 1: Data Analysis Pipeline

2.3 Tools and Technologies

- Mention the programming languages (e.g., Python, R), libraries (Pandas, Scikit-learn, TensorFlow), and platforms (e.g., Jupyter, Power BI, SQL) you used.

- Explain why you selected these tools.

2.4 Techniques Applied

Summarize techniques used for:

- EDA: visualizations, descriptive statistics.

- Modeling: machine learning algorithms (e.g., classification for churn, regression for forecasting).

- Validation: how you evaluated model performance (cross-validation, metrics like accuracy, RMSE, etc.).

2.5 Assumptions and Limitations

3 Dataset

This section outlines the dataset used, including its collection, cleaning, structure, and storage. The data covers customer profiles, contracts, transactions, and sales for FPT Telecom's ultra-fast internet services.

3.1 Data Collecting

The data have been collected from FPT Telecom's system; however, due to access restrictions, not all information could be retrieved. The focus was placed on gathering the essential datasets that could provide valuable insights into customer behavior, service usage, and sales performance.

Specifically, four main datasets were extracted:

- Contract Data: This table acts as the central entity, containing unique Contract IDs and essential timestamps such as Creation Date, Active Date, Service Start and End Dates, as well as the Cancellation Date. It also includes details about the Internet Package associated with each contract, making it critical for tracking the service lifecycle.

- Transaction Data: Connected via the Contract ID, this table details each transaction linked to a contract. Key fields include the Order ID/Reg Code, Transaction Code, and Type of transaction. Additionally, it records the Payment Month, Revenue generated, and status flags such as FG Package Command and Status FGame. This table enables revenue tracking and usage pattern analysis over time.

- Sales Data: Also linked by Contract ID, this dataset captures the sales perspective, including details such as Managing Branch, Sales Unit, Selling Branch, and Sales Region.

- Customer Data: Containing fields for Gender, Age, and Region of the customer, all linked through Contract ID.

3.2 Data Preprocessing

The following data cleaning steps were implemented:

- **Date Standardization**: The `[Creation Date]` column originally contained two date formats (`year-month-day hour:minute:second` and `day/month/year`). Both were standardized into a single `day/month/year` format.

- **Branch (`[Managing Branch]`) Cleaning**: For branches other than Ho Chi Minh (HCM) and Hanoi (HN), any extraneous text following the main branch names was removed.

- **New Column Creation (`[Sales Method]`)**: A new column, `[Sales Method]`, was added and categorized as follows:

- - [Type] = "code" → labeled as "code".
  - [FG Package Command] = "UF Bundle" → labeled as "Bundle".
  - All other cases → labeled as "Extra".

- **Row Filtering by [Sales Channel]**: Rows were removed where the [Sales Channel] was "inside" but the [FG Package Command] was not "Ultra Fast Bundle Internet".

- **Row Filtering by [Sales Method] and [Status F-game]**: Entries labeled as "Bundle" in the [Sales Method] column were eliminated if their corresponding [Status F-game] was not "Activated".

- **Branch Columns ([Sales Branch] and [Managing Branch]) Standardization**:

  - Each province was represented as one distinct branch.
  - Due to their large scale, HCM and HN branches consisted of multiple sub-branches, each equivalent to a provincial branch in terms of hierarchical structure.

After completing the data cleaning process, the cleaned and validated dataset was stored systematically for further analysis. Instead of using a database system, the processed data was saved in structured Excel sheets, ensuring accessibility and ease of use for analysis and reporting purposes. The Excel files were organized clearly by data categories to maintain data integrity and facilitate efficient retrieval. This storage approach was chosen to support the workflow of subsequent data analysis and visualization tasks.

3.3   Dataset

The dataset is organized into four logical tables: **Customer**, **Contract**, **Transaction**, and **Sales**, interconnected by the common field Contract_ID. It contains a total of 28,623 records, capturing customer demographics, service details, transaction history, and sales operations for ultra-fast internet services.

The **Customer** table includes essential demographic information such as gender, age, and region of the customer. The **Contract** table stores details related to the service lifecycle, including creation, activation, and cancellation dates, as well as the specific internet package associated with each contract.

The **Transaction** table records transactional data, including order identifiers, package commands, sales type, revenue, and game service status. Complementing this, the **Sales** table contains operational details such as managing and selling branches, sales region, sales channel, and sales method.

The dataset integrates various data types, including categorical fields (e.g., region, package, branches), numerical fields (e.g., revenue, age), and multiple date fields that track the customer journey across different stages. While some missing values are present, particularly in the Sales Unit and Cancellation Date fields, the dataset remains robust and suitable for comprehensive analysis.

3.4 Database

The structured database was designed to efficiently store and manage comprehensive information related to customer profiles, service contracts, transactions, and sales activities for ultra-fast internet services. Following a thorough data cleaning and validation process, the dataset was organized into four interrelated tables—`Customer`, `Contract`, `Transaction`, and `Sales`—linked via the common field `Contract_ID`. This relational structure ensures data integrity and enables efficient querying and analysis. By storing the data in this format, the project facilitates streamlined exploration of customer behaviors, contract lifecycles, revenue tracking, and sales performance across various regions and branches.
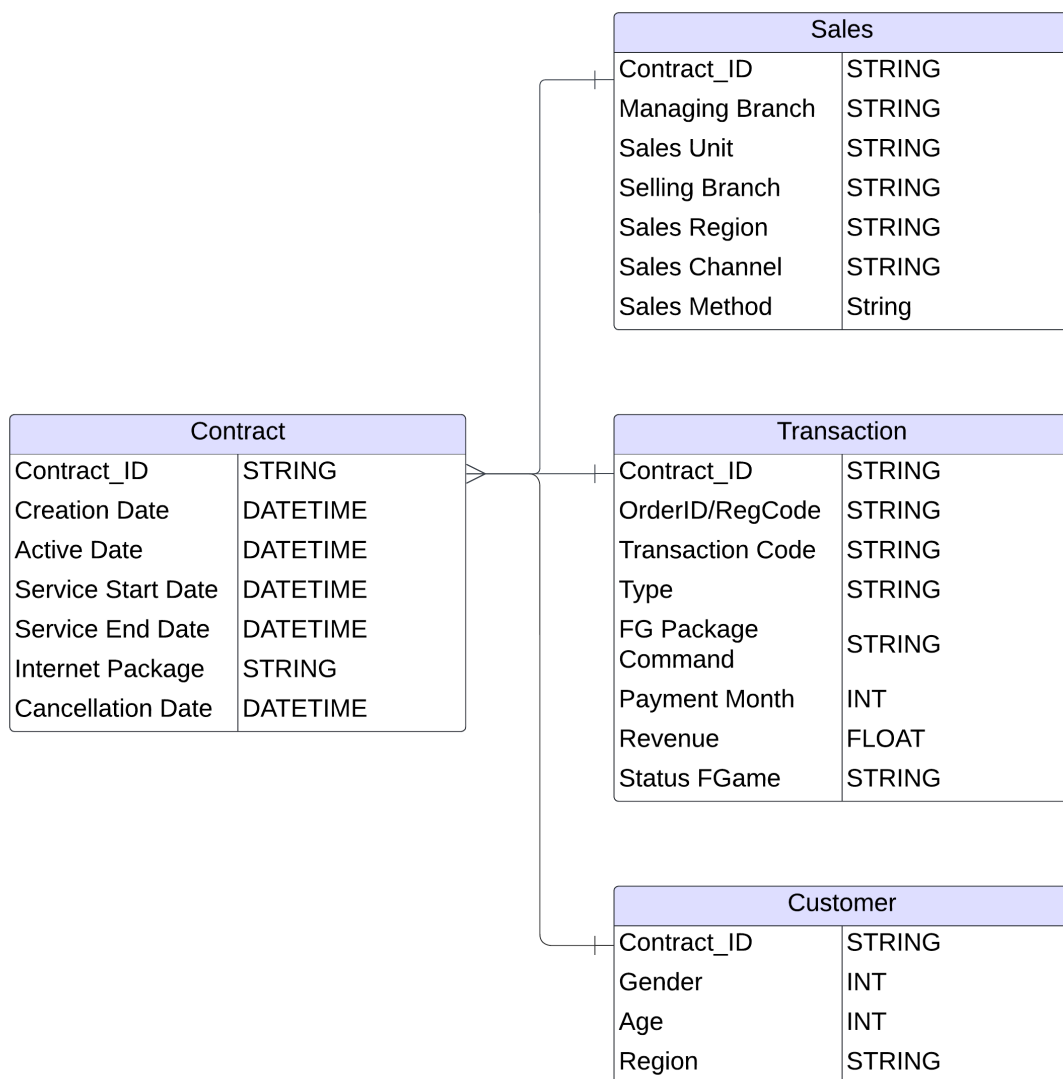
Sales	
Contract_ID	STRING
Managing Branch	STRING
Sales Unit	STRING
Selling Branch	STRING
Sales Region	STRING
Sales Channel	STRING
Sales Method	String

Contract	
Contract_ID	STRING
Creation Date	DATETIME
Active Date	DATETIME
Service Start Date	DATETIME
Service End Date	DATETIME
Internet Package	STRING
Cancellation Date	DATETIME

Transaction	
Contract_ID	STRING
OrderID/RegCode	STRING
Transaction Code	STRING
Type	STRING
FG Package Command	STRING
Payment Month	INT
Revenue	FLOAT
Status FGame	STRING

Customer	
Contract_ID	STRING
Gender	INT
Age	INT
Region	STRING

Figure 2: Database Schema

4 Exploratory Data Analysis (EDA)

4.1 Customer Profile Analysis
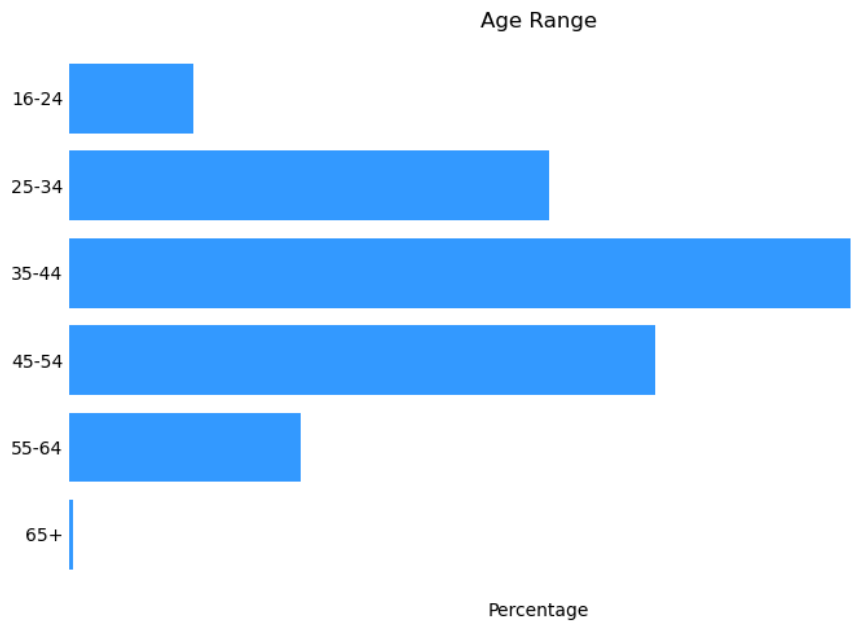
4.1.1 Age Distribution

Figure 3: Customer's age Range

4.1.2 Gender Distribution

The gender distribution of the user base is relatively balanced, with males accounting for 58.1% and females 41.9% (see Figure 4). This indicates a broad interest across genders in gaming performance enhancement tools. Marketing strategies should be inclusive, targeting both male and female gamers, while the product design should remain accessible and appealing to a diverse audience.
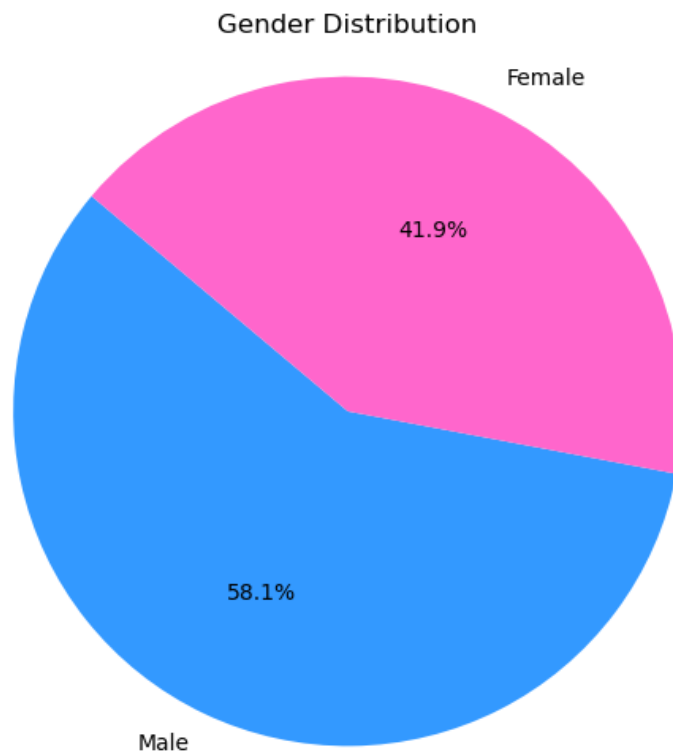

Figure 4: Gender Distribution
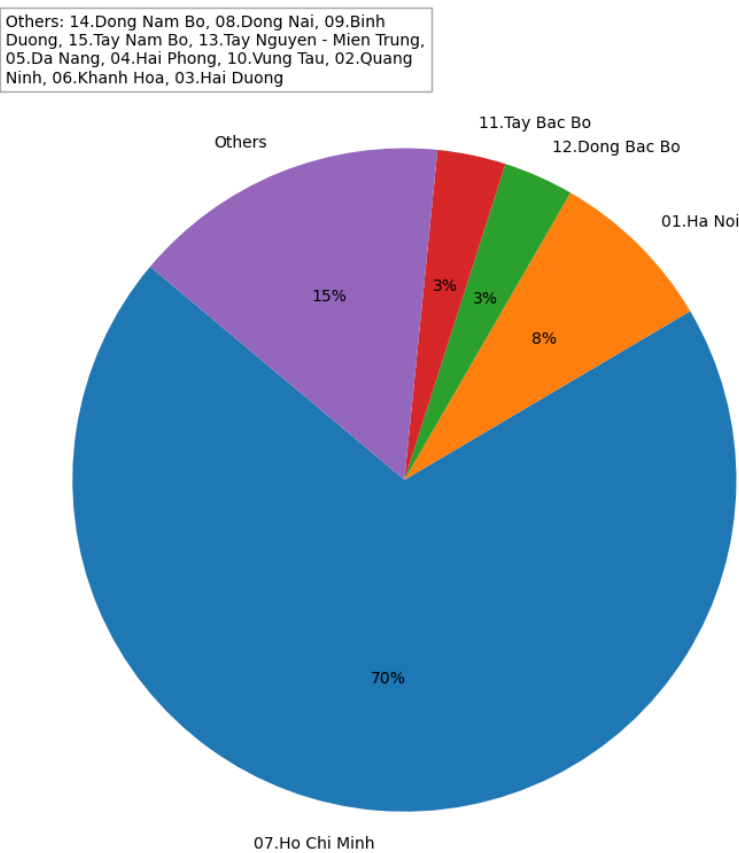
4.1.3 Region Distribution

Figure 5: Region Distribution

4.2 Sales Performance Overview
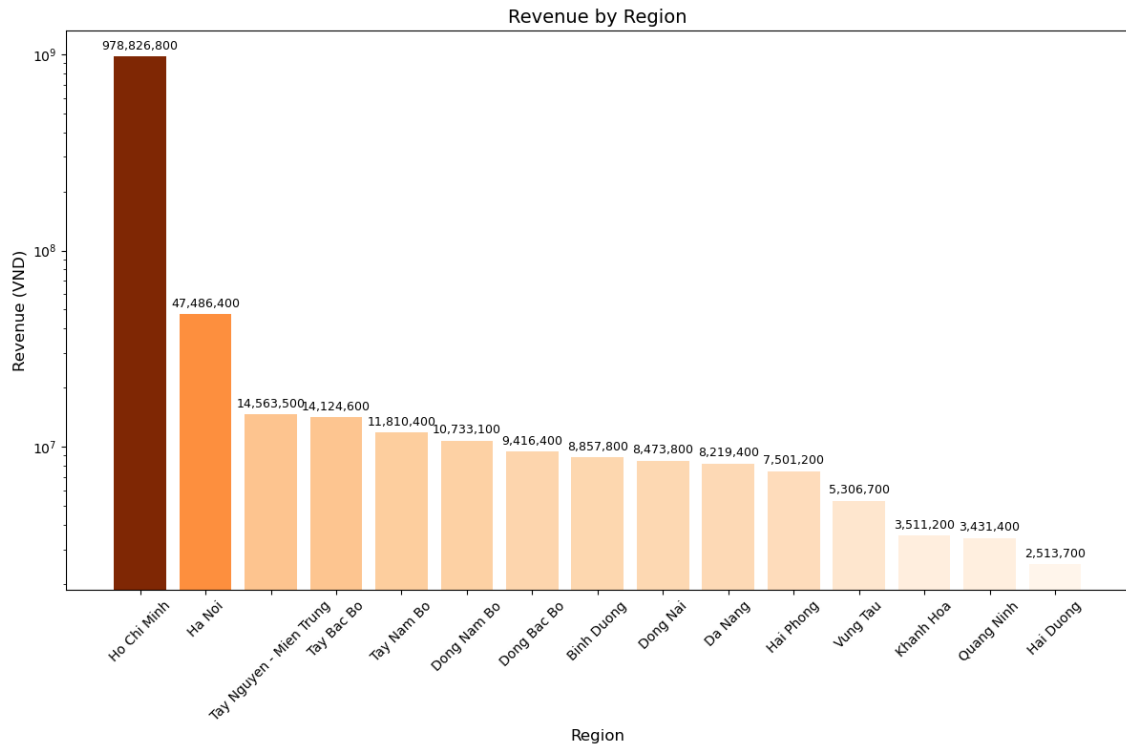
4.2.1 Total revenue

4.2.2 Revenue by region

Figure 6: Revenue each Region

4.2.3 Monthly trends

4.3 Package Popularity

4.3.1 Most sold packages

4.3.2 Packages by region

5 Diagnostic Analysis

5.1 Contract Lifecycle Analysis

5.2 Cancellation Analysis

5.3 Sales Channel Effectiveness

6   Advanced Analysis

6.1   Customer Churn Prediction

6.2   Sales Forecasting

6.3   Customer Segmentation

7 Conclusion and Recommendations

Appendix A

Acronym

AI Artificial Intelligence

UF Ultra Fast

AP Access Point

HCM Ho Chi Minh

HN Hanoi

References

[1] FPT Telecom. *Ultra-fast Internet Service.* Available at: https://fpt.vn/shop/ultra-fast