Department of Information and Communication Technology

Defence 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 Problem Statement

The launch of the "Ultra Fast" 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 "Ultra Fast" stands — and where it could be heading.

1.2 Goal

2 Dataset

2.1 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.

2.2 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 "Đã kích hoạt" (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.

2.3 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 23,623 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.

2.4 Database

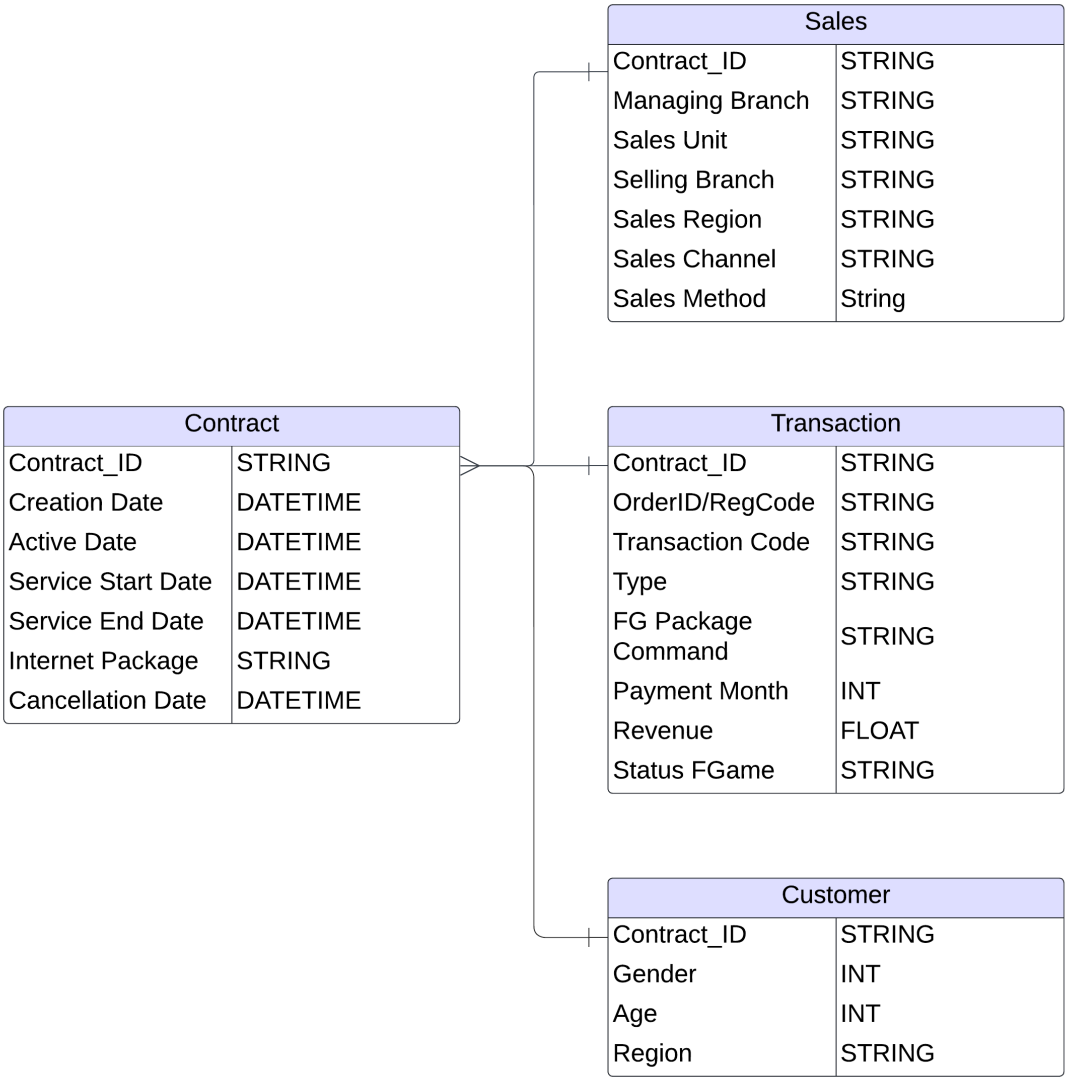

Figure 1: Database Schema

3 Model

4 Conclusion and Future Work

5 References

Appendix A