



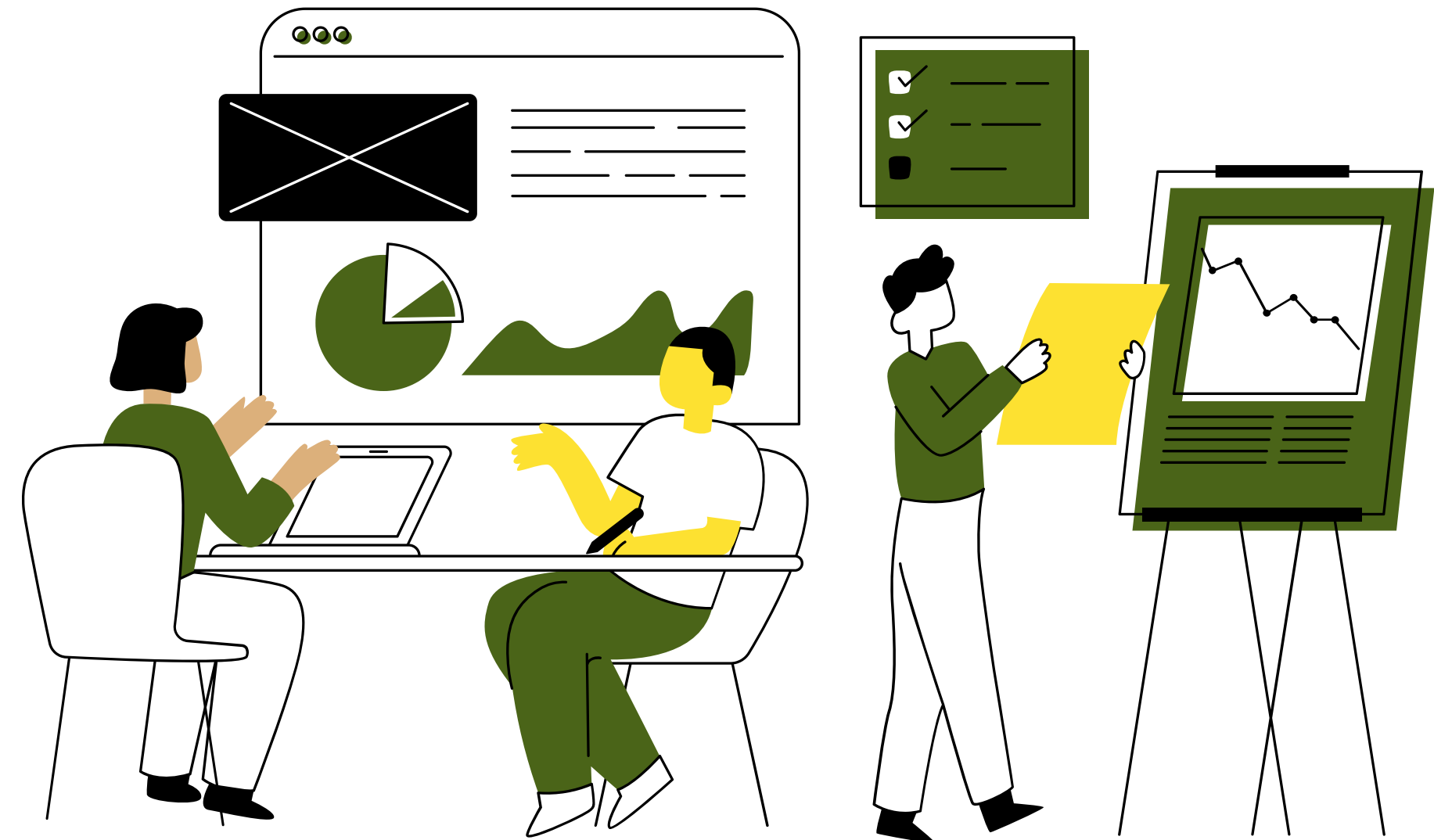
# Predictive Analytics for Disease Diagnosis

# Healthcare Industry

by Azmah, Darsh, Navya and Aarav

# Introduction

The CRISP-DM model, which stands for Cross-Industry Standard Process for Data Mining, is a structured approach used by data scientists and analysts to tackle data mining and analytics projects. The goal is to develop a predictive analytics model for disease diagnosis. This model should be able to analyze various types of data (e.g., patient history, clinical variables, biomarkers, treatment information) and predict disease progression to guide treatment decisions.



# Business Problem

In healthcare, diagnosing diseases accurately is vital for effective treatment. Using the CRISP-DM model, we developed a predictive analytics system. We gathered patient data, cleaned it, and selected key features. After training various models, we integrated the best-performing one into the diagnostic process. This system helps healthcare providers make more accurate diagnoses and tailor treatments. By continuously updating the model, we ensure its effectiveness, leading to better patient care and reduced costs. So let's dive deeper into CRISP Data modeling steps.



# Case Study

Chronic diseases are the leading causes of death and disability in the US, as well as the main drivers of the country's \$3.5 trillion in annual health costs. Five chronic diseases account for 75% of healthcare spending: cancer, cardiovascular disease, diabetes, obesity, and kidney disease. Chronic disease management depends on the ability of healthcare professionals to prevent the development of these diseases and to control them. Managing and preventing chronic conditions, however, is a challenging task. Predictive analytics can empower healthcare providers to make timely and fact-based informed decisions to provide more effective treatments while reducing the costs of this care to patients.

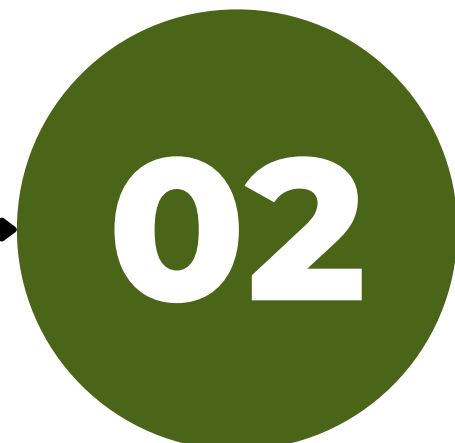




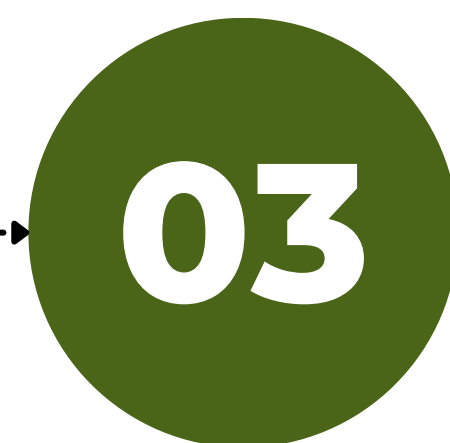
# Business Understanding



Clearly defined the objective of increasing client retention rates as a primary business goal.



Identified crucial business objectives related to client retention, such as reducing churn and increasing profitability.



Utilized the CRISP-DM model to structure the data mining project, ensuring alignment with business needs.



Conducted a thorough analysis to identify key elements causing the loss of clients, understanding the business context and requirements.



2

## Data Understanding

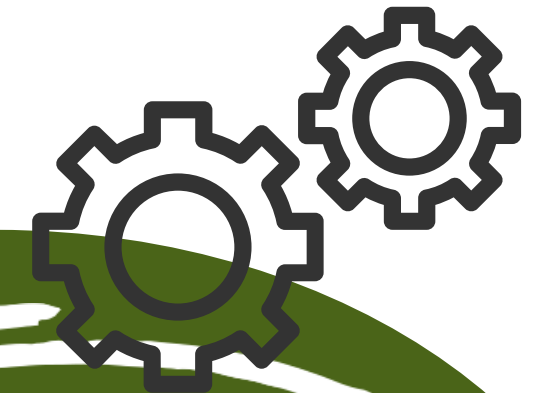
Examined the quality and completeness of the data, identifying any inconsistencies or missing values.



Conducted comprehensive data collection efforts, gathering data on consumer characteristics, product preferences, and buying habits.

Recognized the importance of understanding the data's structure and content to inform subsequent steps in the project.

Analyzed the dataset to uncover patterns and correlations related to client retention rates, gaining insights into customer behavior.



3

## Data Preparation

Transformed the data through normalization, scaling, or encoding to prepare it for modeling, ensuring consistency and compatibility.

Cleansed the dataset by removing inconsistencies, errors, and outliers to ensure data accuracy.

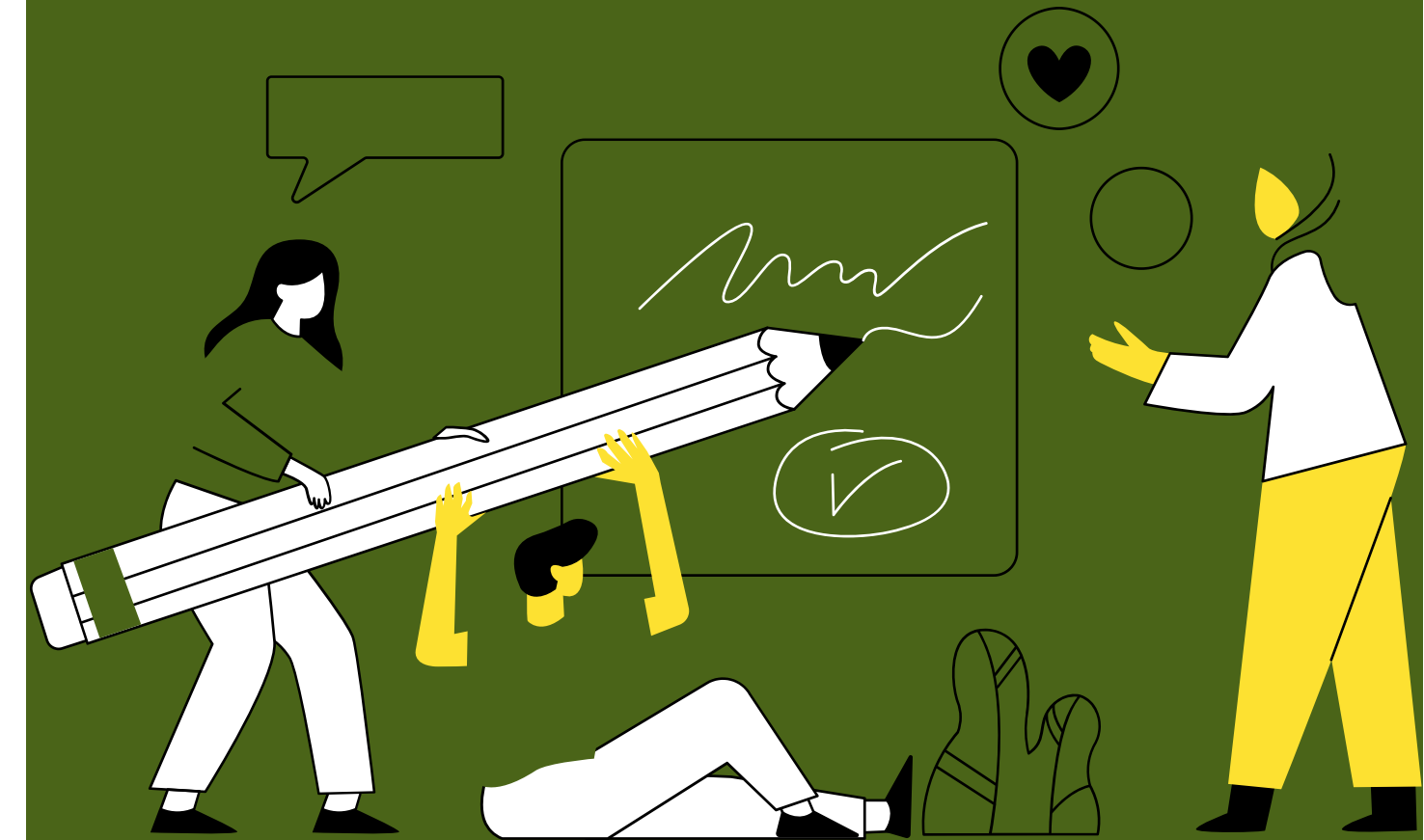
Integrated data from multiple sources, such as customer databases and transaction records, to create a unified dataset.

Reduced the dimensionality of the dataset by selecting relevant features and removing redundant or irrelevant variables.

# 4

## Modeling

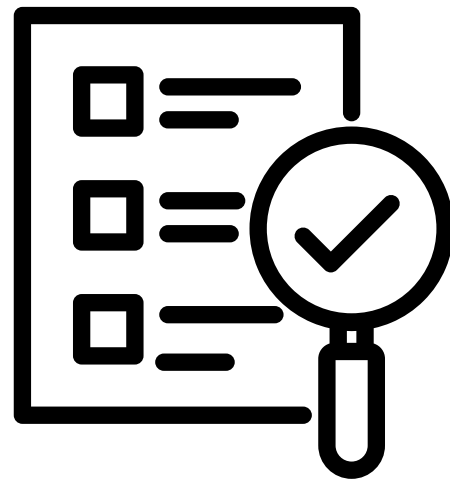
- Developed predictive models to estimate the probability of customer churn, leveraging statistical and machine learning techniques.
- Iteratively trained and evaluated the models to optimize performance and accuracy, considering factors such as precision, recall, and F1 score.
- Incorporated feedback and insights from evaluation results to refine and improve the models, ensuring alignment with business objectives.
- Used decision-tree algorithms to create focused retention strategies based on customer characteristics and behavior.





Assessed the effectiveness of the models in fulfilling business needs by comparing their performance against real-world data.

Considered the practical implications of the models' recommendations and their impact on client retention rates.



## Evaluation



Iteratively refined the models based on evaluation results to enhance their effectiveness in achieving business goals.

Analyzed metrics such as accuracy, precision, recall, and F1 score to evaluate the models' performance and generalization ability.

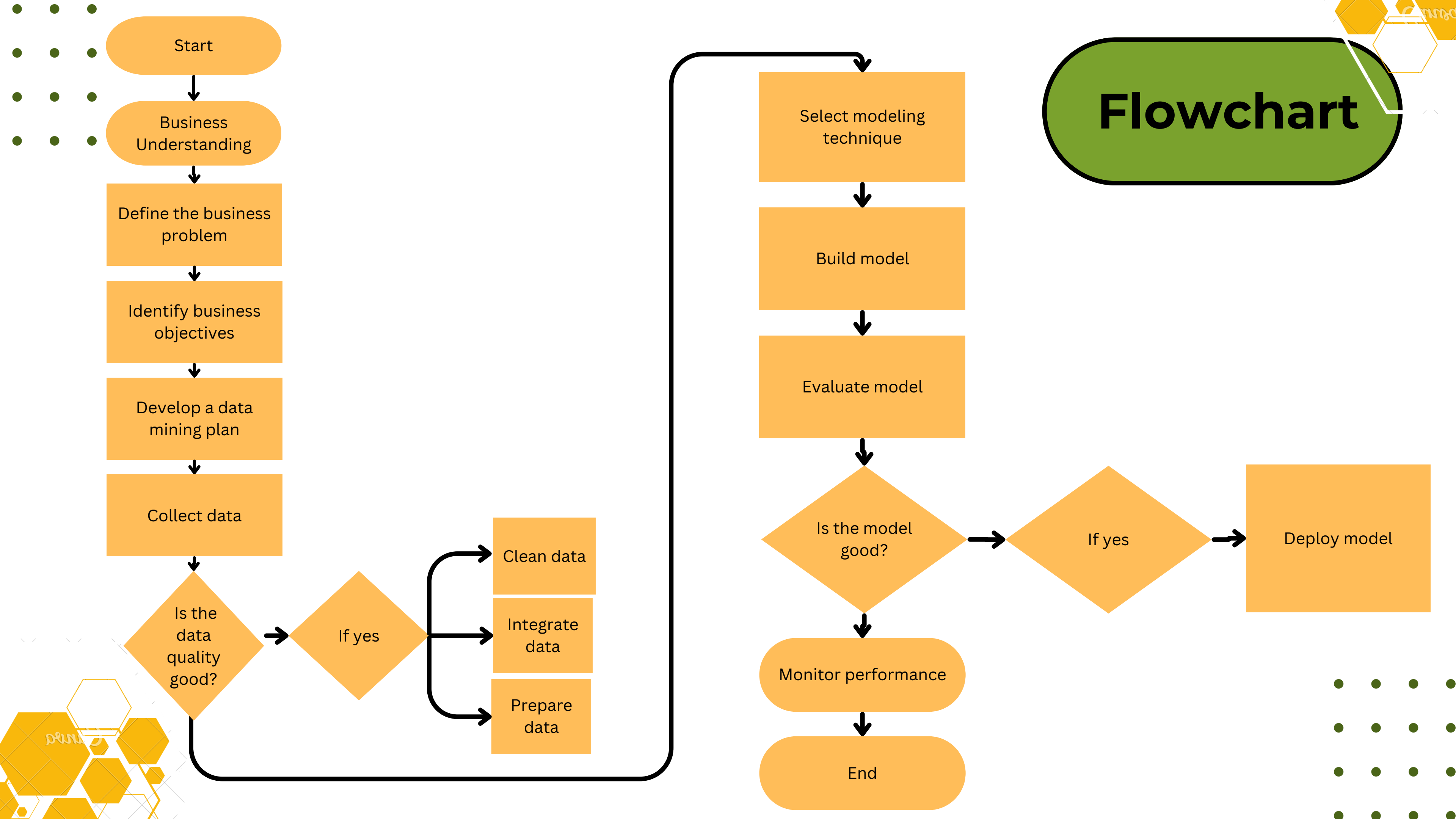


# 6

## Deployment

- Developed a deployment plan to roll out the retention strategies across the retail chain, ensuring consistency and scalability.
- Ensured the integration of data-driven decision-making processes into the company's operational workflows, fostering a culture of data-driven innovation and continuous improvement.
- Monitored the effectiveness of the retention tactics in reducing client loss and increasing profitability, adjusting strategies as needed.
- Implemented customized retention tactics based on the insights derived from the data mining project, targeting at-risk customers.

# Flowchart



# Conclusion

The comprehensive client retention approach utilizes the CRISP-DM model for structured data mining. It involves thorough data collection, analysis, modeling with advanced techniques, and iterative model refinement. Customized retention tactics derived from the models are deployed consistently across the retail chain, with continuous monitoring and adjustment. This data-driven, holistic approach aligns with business objectives, fosters innovation, and optimizes client retention, representing a significant advancement over previous models.



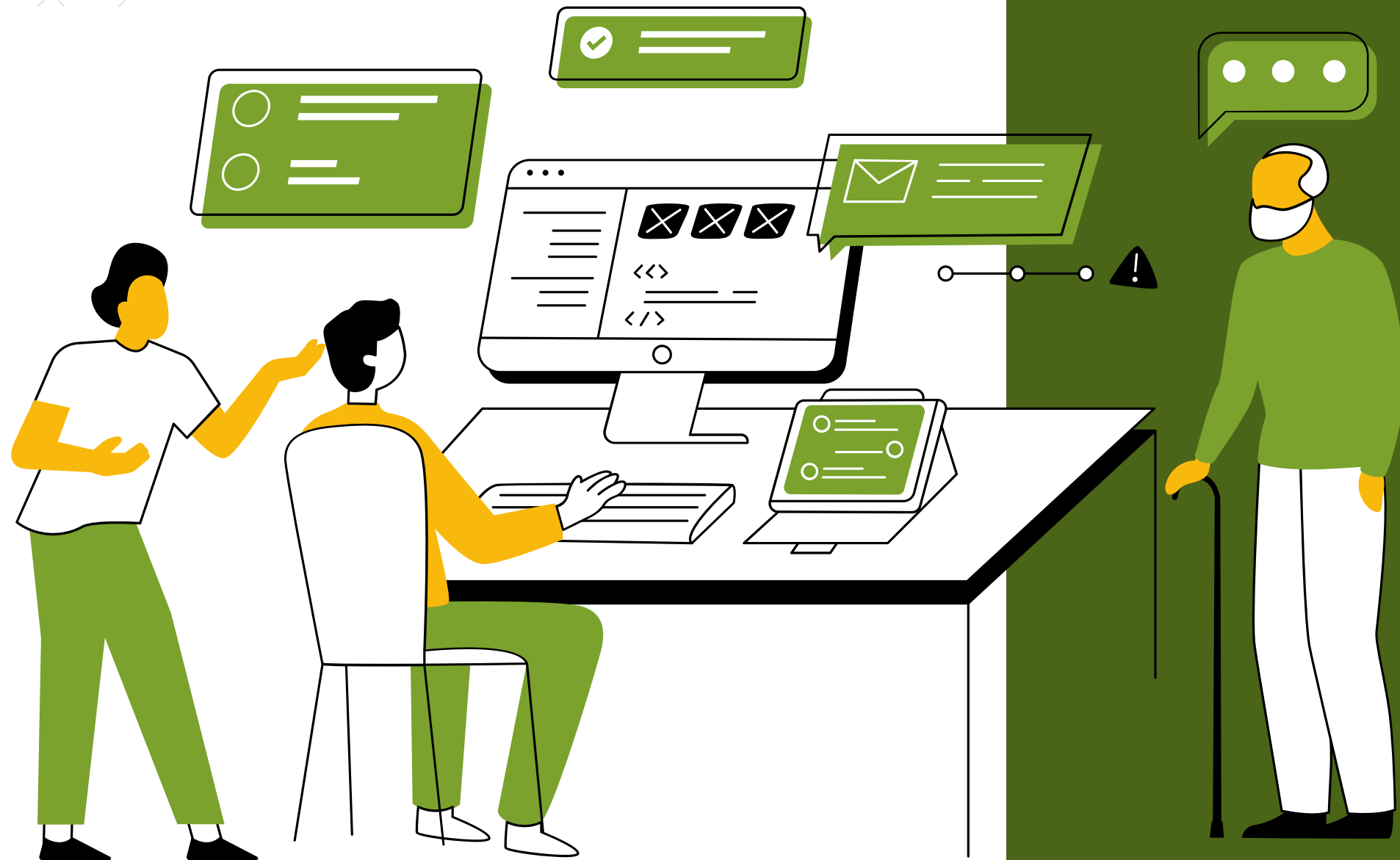


# References



- <https://insightssuccess.com/data-mining-in-business-applications/>
  - <https://www.matillion.com/blog/5-real-life-applications-of-data-mining-and-business-intelligence>
  - <https://visme.co/blog/presentation-tips/>
  - <https://www.wikihow.com/Write-a-Use-Case>
  - <https://segment.com/data-hub/predictive-analytics/healthcare/>
  - <https://www.analyticsvidhya.com/blog/2023/02/use-of-ml-in-healthcare-predictive-analytics-and-diagnosis/>
  - <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>
  - <https://www.datascience-pm.com/crisp-dm-2/>
  - [https://medium.com/@avikumart\\_/crisp-dm-framework-a-foundational-data-mining-process-model-86fe642da18c](https://medium.com/@avikumart_/crisp-dm-framework-a-foundational-data-mining-process-model-86fe642da18c)
  - <https://www.youtube.com/watch?v=NinRBxDVdnM>
  - <https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview>
  - <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>
  - <https://www.cdc.gov/chronicdisease/about/costs/index.htm>
- 





# THANK YOU

Navya - PPT  
Aarav - Information  
Darsh - Flowchart information  
Azmah - Case Study and Flowchart

