EPITA - International Programs
AI Project Methodology
2023-2024

Graded Project : Part 1 (AI project Functional Methodologies)

⇒**Case Study Goal**
The goal of this case study is to be more familiar with functional aspects of Data & AI projects through a case study

⇒**Tools**
It is recommended to use tables and diagrams to explain and organize your answers, it's also proposed to use a project management free tool (trello), in addition, you can use the tools of your choice if relevant for this case study.

⇒**References**
Add all sources you may rely on in your answers (web articles, sites, generative AI…)

⇒**Restitution**
Report you group case study with all requested items in a google document or a pdf in you **group dedicated repository** you can choose here : AI PM_Graded Project_Groups_092024

Case Study: Predicting Customer Churn at RetailGenius

In this case study, we will explore an AI project for a fictional company "**RetailGenius"**, a cutting-edge technology company specializing in E-commerce and online marketplaces.

The company's primary objective is to retain users and enhance their experience on RetailGenius platform by leveraging data and AI technologies.

Presentation of RetailGenius :

RetailGenius is a rapidly growing e-commerce company with a diverse range of **products**, **sellers**, and **customers**. They operate globally and handle a substantial volume of data.

RetailGenius launched a strategic program to derive value from its huge amount of Data. To achieve this goal, they started **a first AI project to predict customers churn**.

The first step was to collect some insights from interviews with **key stakeholders**:

Interviews with key RetailGenius stakeholders :

Interviews was conducted with the teams below, to understand the company's data needs and goals :

Interview with Business Stakeholders
How does data support your day-to-day operations and long-term strategies?

Data helps us make **informed decisions** in product development, marketing strategies, and inventory management. It's crucial for long-term customer retention and growth.

What data elements are critical for your decision-making processes?

We rely on **customer** demographics, **purchase** history, **product** performance data, and marketing **campaign** effectiveness.

What are the key performance indicators (KPIs) you use to measure success?

KPIs include conversion rates, customer retention, revenue growth, and customer satisfaction scores.

Interview with RetailGenius's Data Team

The team consists of data engineers, data scientists, and data analysts. below the questions asked with the answers :

What business objectives are being addressed with data and AI?

Data and AI support our efforts to enhance the user experience and optimize operations. We aim to improve personalization, recommendation systems, fraud detection, and inventory management.

What are the primary data sources used by RetailGenius?

RetailGenius primarily sources data from **user interactions** on our platform, including **product** views, **searches**, **purchases**, and **reviews**. We also collect data from **seller** activities, **order** processing, and **customer** support.

What challenges do you face in managing and utilizing data effectively?

Challenges include **data silos, data quality** issues, and ensuring **data privacy and security**. **Integrating and cleaning** data from various sources can be time-consuming.

What data governance and security measures are in place?

We have a data governance framework that defines roles, responsibilities, and data ownership. Security measures include encryption, access controls, and regular audits.

Interview with Technology Team

What databases and data storage systems are currently in use?

We use a combination of **SQL and NoSQL databases, cloud** storage solutions, and **data lakes** to accommodate various data types.

What are the existing data integration and data transformation tools?

We employ **ETL** (Extract, Transform, Load) processes using tools like **Apache Nifi** and **Talend** to integrate and transform data.

Are there any scalability or performance considerations for data storage and processing?

Yes, as our data volume increases, we're actively exploring **distributed data storage** solutions and optimizing query performance.

RetailGenius Stakeholder Workshops

Stakeholder workshops to gather collective requirements and expectations and to identify common data elements and relationships that need to be considered.

Workshop	Key Insights
Product Managers	Product managers emphasized the importance of data-driven decision-making in product development. They highlighted the need for **real-time access t**o user behavior data to refine product features and optimize user experiences.
Marketing Executives	Marketing executives expressed the need for data to support targeted marketing campaigns. They expressed interest in **analyzing user demographics** and **campaign engagement metrics** to tailor marketing strategies.
Data Team	The data team underscored the significance of **data quality** and **data governance.** They highlighted the importance of **data modeling** and **processing** to ensure quality and compliance with data privacy regulations.
Technology and Infrastructure Team	The technology team provided insights into the current tech stack and infrastructure. They expressed a desire for data models that could be easily integrated into their existing systems, emphasizing the importance of data security and scalability.
Data Consumers	In addition to these stakeholders, data consumers from various departments shared their specific data requirements. These include customer **support teams, finance, and operations**, each with unique needs for data access and analysis.

Common Data Elements	Stakeholder workshops identified common data elements, such as **customer** information, **product** data, and **order** history, that are crucial for decision-making across various departments. This understanding will guide the data modeling process.

Data and Documentation analysis

A detailed analysis of RetailGenius's existing data landscape was conducted to examine sample datasets, data dictionaries, and any existing data-related documentation available within the organization. Here are some key findings and documentation:

Data Samples: RetailGenius collected sample datasets from various sources, including user interactions, product information, and marketing campaign data. These samples helped us understand the structure and quality of the data.

Data Dictionaries: RetailGenius maintains comprehensive data dictionaries that define data elements, their meanings, and their relationships. This documentation was essential for understanding the semantics of data.

Data Quality Assessment: RetailGenius conducted a preliminary assessment of data quality, identifying issues such as missing values, inconsistencies, and outliers. This information will guide data cleaning and transformation efforts.

Data Flow Diagrams: RetailGenius created high-level data flow diagrams to visualize how data moves through the organization, from data sources to data consumers. These diagrams illustrate data pipelines and integration points.

Metadata Catalog: RetailGenius has an established metadata catalog, which documents the source, lineage, and transformations of data. It's a valuable resource for understanding the data ecosystem.

External Data Sources:

RetailGenius recognizes the value of incorporating external data sources into its data ecosystem to enrich its understanding of customer behavior and market dynamics. Here are some external data sources that were identified:

Market Trends Data: External market data sources provide insights into industry trends, consumer preferences, and competitor performance. Subscribing to market research reports and aggregating data from industry publications can help RetailGenius stay informed about the competitive landscape.

Customer Reviews and Sentiment Analysis: Leveraging customer reviews from third-party platforms and conducting sentiment analysis can offer valuable insights into user satisfaction, preferences, and pain points. RetailGenius can utilize APIs to gather this data.

Competitor Data: Monitoring the pricing, product offerings, and strategies of competitors is crucial for maintaining a competitive edge. Data on competitor products, pricing, and customer reviews can be obtained through web scraping and API integrations.

Economic and Demographic Data: External data sources such as government census data, economic indicators, and demographic statistics can inform marketing and sales strategies. Accessing this data through open data platforms and government agencies can be beneficial.

Social Media Data: Data from social media platforms provides real-time feedback and user-generated content. APIs provided by platforms like Twitter, Facebook, and Instagram can be used to collect user-generated data relevant to RetailGenius's market and brand sentiment.

Weather Data: For businesses that are impacted by weather conditions, such as retail sales, access to real-time weather data can help in demand forecasting and inventory management. Weather data providers offer APIs and data feeds.

Third-Party Data Providers: Various third-party data providers offer specialized datasets that can enhance RetailGenius's understanding of customer behavior and preferences. These may include consumer behavior data, geographical data, and more.

RetailGenius can establish data pipelines and integration processes to continuously collect, cleanse, and integrate external data sources into their data ecosystem. By leveraging these external data

sources, RetailGenius can gain a more comprehensive view of its market and customers, enabling data-driven decision-making and competitive advantage.

Your mission : Framing of Churn prediction project for RetailGenius

RetailGenius aims to proactively address customer churn by leveraging AI. The project's goal is to predict which customers are at risk of churning and take preventive measures to retain them.

Through the sections below, you will be guided to prepare **a functional framing of this AI project.**

Project Strategy

- *Define the strategic objectives of the AI project in the context of customer churn.*
- *What key performance indicators (KPIs) would you use to measure the success of the churn prediction model?*
- *How can AI contribute to improving customer retention at RetailGenius?*

Project Design

Data

- *What data sources are relevant for predicting customer churn?*
- *Discuss potential challenges in using the necessary data for the project.*

Models

- *What AI models would be suitable for predicting customer churn?*
- *How would you handle model training, validation, and testing?*
- *How would you manage the model versioning and serving ?*

Deployment

- *Discuss the different deployment strategies for the churn prediction model.*
- *What considerations should be taken into account for deploying this AI project in a production environment?*

Monitoring

- *How will you monitor the performance of the churn prediction model over time?*
- *Outline a plan for handling model drift and maintaining model accuracy.*

Project Team

- *Identify the roles and expertise required for the AI project team.*
- *How can you ensure cross-functional collaboration within the team?*
- *Discuss the skills and expertise needed for each role in the team ?*
- *How do you ensure team alignment with project strategy ?*
- *How will the AI team collaborate with other departments, such as marketing and customer support?*

Project Governance & Communication

- *Identify key stakeholders for the AI project and define a communication plan.*
- *What are the governance instances you will put in place to properly inform and align stakeholders ?*
- *How will you communicate model outputs and predictions to both technical and non technical teams ?*

AI Project Management Methodology

- *Choose a project management methodology suitable for this project.*
- *Discuss why this methodology is suitable for the project ?*
- *Identify potential risks associated with this AI project and propose some risk mitigation strategies.*
- *How would you handle costs and planning derivation due to AI models iterations ?*
- *Set up a mock AI project on **trello** (free and easy to use tool): create a Kanban board to simulate the management of churn prediction project from start to deployment and monitoring.*