**Projeto Data Mining**

**To do List**

* ~~Mudar a variável flight para o plural~~

### **Introduction**

• Client: Amazing International Airlines Inc. (AIAI)

• Objective: Customer segmentation for personalized marketing

**Amazing International Airlines Inc. (AIAI)**

Challenge: Design personalized services and marketing strategies

Context: Highly competitive airline industry

Opportunity: Leverage data‐driven customer segmentation

Dataset: 3‐year loyalty program and flight activity data

### **Project Overview**

We act as consultants for AIAI and our task is to analyze customer loyalty data and corresponding flight activity collected over a three-year period [3, 2].

Based on this information, we have to develop a data-driven segmentation strategy.

The segmentation should be approached from multiple perspectives, such as:

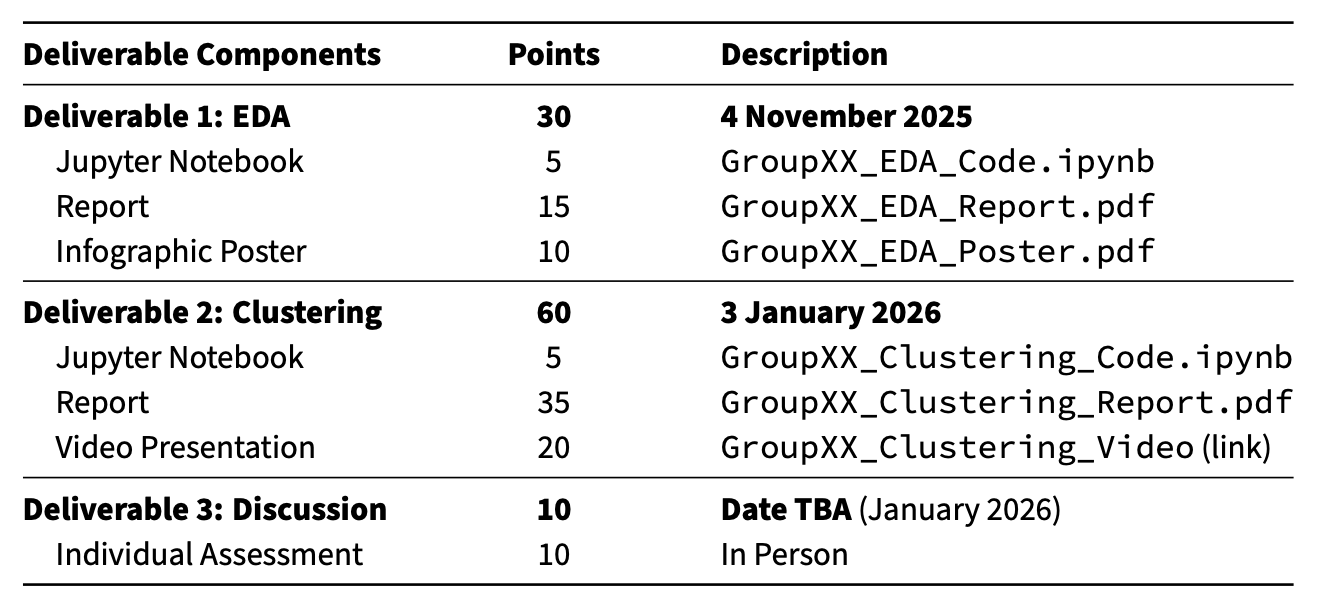
* Value-based segmentation, grouping customers according to their economic contribution.
* Behavioral segmentation, analyzing purchasing habits and travel behaviors.
* Demographic segmentation, categorizing customers by age, occupation, or other attributes to reveal different interaction patterns.

**Ultimate objective:** integrate these perspectives into a final segmentation framework that supports AIAI in crafting a comprehensive marketing strategy.

Teacher recommendation: To structure the project, we recommend applying the **CRISP-DM** methodology [5, 6], progressing systematically from business understanding and data preparation through modeling and evaluation.

**Nota:** CRISP-DM - (Cross-Industry Standard Process for Data Mining), que é um padrão de processo amplamente utilizado para projetos de mineração de dados e ciência de dados. Esta metodologia é dividida em 6 fases (chat)

O CRISP-DM é cíclico, ou seja, pode-se retornar a fases anteriores conforme surgem novos insights ou problemas.

**Deliverable Summary**

**Deliverable 1: Exploratory Data Analysis (30 points)**

**Focus:** uncovering meaningful patterns, identifying limitations, and generating initial hypotheses about customer groups.

**key Tasks:**

* Conduct descriptive statistics and visualizations to highlight distributions, trends, and anomalies, while noting which variables appear most relevant for segmentation.
* Assess data quality issues and evaluate how these may affect clustering reliability.
* Identify preliminary behavioral signals that suggest distinct types of customers.
* Develop and justify engineered features. Show how these derived variables capture richer aspects of customer behavior and explain their potential contribution to clustering models.

**In presenting results, consider addressing the following:**

1. Which findings were most unexpected or insightful, and what do they reveal about likely customer clusters?
2. What data limitations pose the greatest risks for clustering, and how might they be mitigated?
3. Which patterns in customer activity, including those revealed by engineered features, suggest natural groupings, and what cluster characteristics do you anticipate?
4. How would you communicate these insights to non-technical stakeholders? Include a clear explanation of the expected number of clusters, the most important differentiating features, and any anticipated challenges.

**Component 1: Jupyter Notebook (5 points)**

**Expectations:**

- Jupyter notebook following file naming format GroupXX\_EDA\_Code.ipynb

- All code cells executed with visible outputs

- Systematic data exploration workflow

- Organized code, using Markdown headings to separate sections

- Clean, well-documented code with consistent variable naming

- Error-free execution in fresh environment

**Component 2: Report (15 points)**

This component evaluates analytical thinking and business insight as you transform technical exploration into strategic insights, following CRISP-DM methodology to prepare for segmentation analysis.

**Expectations:**

- Report following file naming format GroupXX\_EDA\_Report.pdf

- Maximum of 5 pages of content using provided framework and template

- Insights and interpretations (not just data descriptions)

**Component 3: Infographic Poster (10 points)**

This component assesses your ability to communicate key insights to business stakeholders, translating technical findings into understandable visual summaries. This deliverable should be targeted at AIAI management team seeking data-driven insights for customer strategy.

**Expectations:**

- Single page poster following file naming format GroupXX\_EDA\_Poster.pdf

- A3 size poster in PDF

- 3-4 key findings with supporting visualizations

- Non-technical language accessible to executives

- Professional design with clear information hierarchy

- Focus on insights that inform segmentation strategy

### **General Polices**

**Use of Generative AI tools**

- The use of generative AI tools is permitted but must be fully disclosed. It is essential that the students’ own contributions to the report exceed that of any AI tools used.

- Students must include a section in the annex of the report documenting the use of AI tools (Annex AI Usage Statement).

- Students must document specific AI tool usage (ChatGPT, Claude, Gemini, etc.)

- Students must specify how AI tool was used (ideation, code assistance, refining, proofreading).

- If the group did not use AI tools, this statement must explicitly state this.

- Students are fully responsible for the contents of the report they submit, including any material generated or assisted by AI tools.

- Do not just copy and paste the results if AI tools are used.

- Students may be asked to explain the meaning of any AI-generated content submitted to ensure their comprehension.

Components Overview

* Component 1: Jupyter Notebook (5 points)

File: GroupXX\_EDA\_Code.ipynb

Clean, documented code with systematic workflow

* Component 2: Report (15 points)

File: GroupXX\_EDA\_Report.pdf

Maximum 5 pages using CRISP‐DM framework

* Component 3: Infographic Poster (10 points)

File: GroupXX\_EDA\_Poster.pdf

A3 size, executive‐focused communication