# Amazing International Airlines Inc.

## Exploratory Data Analysis for Customer Segmentation

Grupo 27 - Gonçalo Rilhas, Guilherme Santos, Luana Pinto, Marta Pedro
Instructors: Fernando Bação, Farina Pontejos, Gaspar Pereira, Ana Caleiro

NOVA IMS
INFORMATION MANAGEMENT SCHOOL

Course: Data Mining

## Introduction

**Amazing International Airlines Inc.** (AIAI) operates in a **highly competitive** aviation market where **customer loyalty** and **personalized experiences** are critical to success. While facing diverse traveler behaviors and evolving expectations, AIAI seeks to better understand its customer base through **data-driven insights**.
This case study explores AIAI's loyalty and flight activity data to uncover key patterns in demographics, value and travel behavior. By identifying factors that **differentiate customers**, the analysis enables **targeted strategies** that strengthen satisfaction, retention and loyalty.

## Project Objective

The **primary objective** of this study is to analyze customer loyalty and flight activity data from AIAI to uncover **meaningful behavioral** and **demographic patterns**.
By conducting **exploratory data analysis**, this project aims to provide a data-driven foundation for **customer segmentation**, enabling the airline to design more **targeted marketing strategies**, improve customer retention and enhance loyalty program effectiveness.

## Methodology

The methodology applied in this project follows the main stages of a Data Science and Data Mining process, guided by the **CRISP-DM framework.** This first Deliverable focuses on the following **key stages**:

- Business Understanding
- Data Understanding and Visualization

## Collected Data

**Customer's Database**
- 16,921 records
- 20 features

**Flights' Database**
- 608,436 records
- 10 features

## Data Quality

To ensure the **reliability** of the analysis, a series of data quality checks were performed on the available data:

- Basic validation ensured data accuracy and consistency.
- Detected duplicates and outlined approaches to address them, as well as missing and anomalous values.
- Analyzed outliers to assess their relevance and consistency within the business context.
- Combined the Flights and Customers databases to enable one final temporary integrated analysis.

## EngineeredFeatures

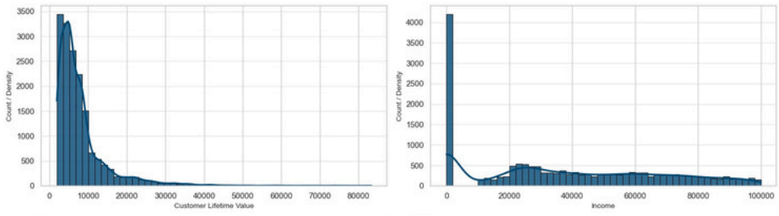
To enrich customer insights and prepare data for segmentation, new features were created and flight data was aggregated at the customer level:

- ***'Is_Active'***: distinguishes active vs. inactive customers based on '*CancellationDate*'.
- ***'Customer_Tenure'***: measures program duration since the enrollment date in the program.
- ***'Total_Flights'***, ***'Total_DistanceKM'*** and ***'Total_PointsAccumulated'***: summarize overall travel activity.
- ***'Redemption_Behavior'***: classifies customers by how much of their earned points they redeem each month.
- ***'Companion_Rate'***: proportion of flights with companions.
- ***'Fidelity_Age_Years':*** how long each customer has been enrolled in the loyalty program, in years.

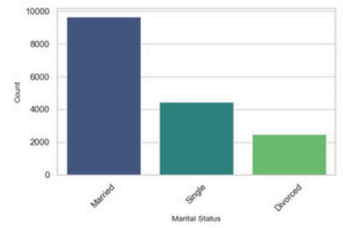
## Exploratory Data Analysis

The exploratory analysis uncovered key behavioral and structural patterns in AIAI's customer and flight data. These findings reveal distinct loyalty dynamics, spending behaviors, and travel trends that will guide the next modeling phase.

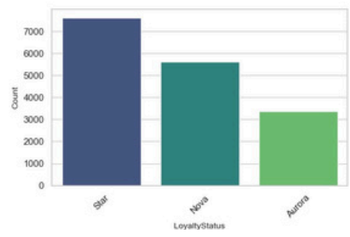
Most of the number-based data (like **income** and **customer value**) is heavily **uneven** - meaning a small number of customers have very high values, while most have much lower ones. In other words, a few big spenders make up most of the flight activity.




The **loyalty program** has a pyramidal structure - there are many customers in the lower levels and fewer in the higher ones. **Married customers** tend to travel more often with a companion. These patterns show that the customer base is quite diverse and give a good starting point for grouping and analyzing customers in the next steps.

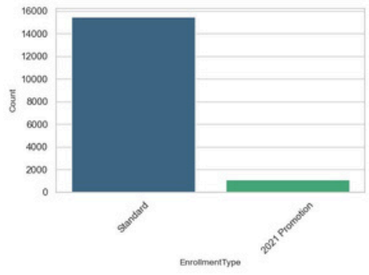



**Loyalty Status Distribution by Marital Status**

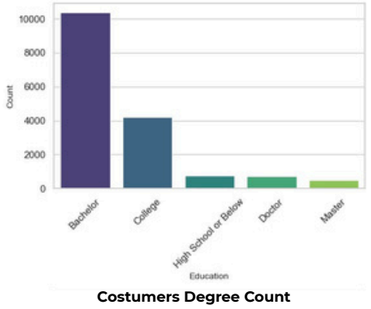



**Loyalty Status Absolute Count**

**Educationally**, most members held at least a **bachelor's degree**, and "Standard" remained the main enrollment channel.

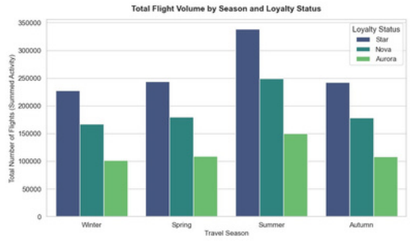



**Enrollment Channel Count**

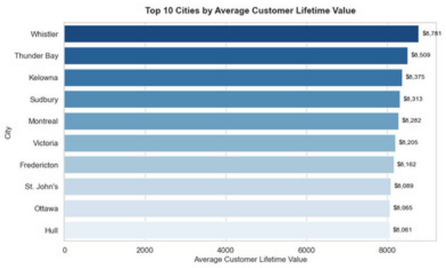



**Costumers Degree Count**

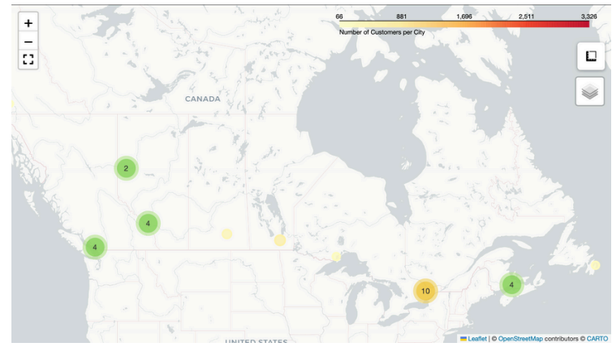
Flight activity peaks during **summer** across all loyalty tiers, with Star members consistently showing the highest engagement. This suggests that marketing efforts should focus on **rewarding Star** loyalty while boosting Nova and Aurora participation in lower-demand seasons.



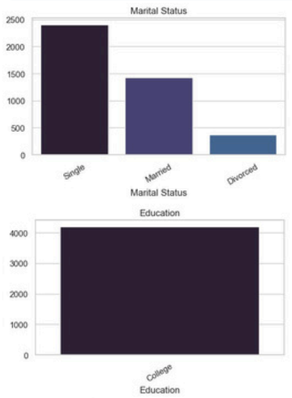

**Flight Volume by Season and Loyalty Status**




**Whistler**, **Thunder Bay**, and **Kelowna** show the highest average customer lifetime value, suggesting that customers in these cities may have stronger loyalty, higher travel frequency, or greater spending capacity. These patterns could be influenced by **demographic or economic factors**, such as higher income levels or stronger brand engagement.



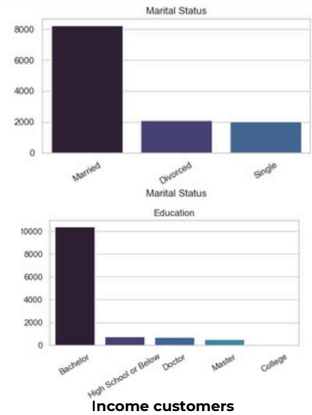

This map was generated to better understand and visualize **customer distribution**. The highest concentrations are found in **Ontario** and the **western provinces**, particularly around major cities such as Toronto and Vancouver. In contrast, central and northern regions show lower customer density, indicating potential markets where brand visibility and customer engagement could be strengthened.




**No-income customers**

It was noticed that around a quarter of the customers reported no income, most of whom are single and have a college education, suggesting they might be students or early-career individuals.

In contrast, income customers are mainly married and highly educated, indicating a more stable and affluent segment with stronger purchasing potential.




**Income customers**

## Next steps

These analyses revealed AIAI's customer structure, data quality, and behavioral patterns, providing a strong foundation for the next phase of Data Preparation and Clustering:

### Feature Refinement

Keep:
- *'Is_Active'*
- *'Customer_Tenure'*
- *'Fidelity_Age_Years'*
- *'Companion_Rate'*

as core behavioral indicators.

### Group Preservation

Maintain the "Income = 0" group as a meaningful behavioral segment.

### Data Transformation

Apply scaling or logarithmic transformation to reduce the impact of skewed variables.

### Feature Selection

Remove redundant travel metrics (e.g., Total_Flights vs Total_Distance) to improve clustering efficiency.