Final Project – Professional Master in Data Analysis

Raw Dataset: **Yoga Dataset**

Post-analysis Dataset: **Tableau Dataset**

--> **Python Analysis** <--

--> **Tableau Dashboard** <--

Guidelines

The project is composed of three main files, to be reviewed in the following logical order:

1. **Python Notebook (.ipynb)**

The analytical core of the project. It contains all data processing, modeling, and visualization phases, divided into five steps:
Setup, **Data Quality & Preprocessing**, **Introductive EDA**, **Data Manipulation & Visualization**, **ML Implementation & Strategy**.
Each phase closely reflects the structure of the master course, starting from the raw dataset transformation up to the implementation of machine learning algorithms covered in class.
Markdown cells describe the workflow step by step, supported by comments within the Python code. This enables clear understanding of how the data is transformed, analyzed, and interpreted, leading up to the strategic questions outlined at the end.
Data cleaning was minimal, given the initial good quality of the dataset. Only a few adjustments were made: converting some columns to datetime, reordering fields for clarity, and engineering key features for further analysis.
Charts use both static (Seaborn/Matplotlib) and interactive (Plotly) libraries, following the techniques learned throughout the course. All main chart types were employed, including bar charts, line charts, pie charts, scatterplots, boxplots, and heatmaps.
Implemented machine learning algorithms include: **Linear, Polynomial, and Logistic Regression**, **Random Forest**, and **K-Means Clustering**.

2. **Tableau Dashboard(.twbx)**

A visual summary of the analysis, built from the most significant KPIs and variables identified in the notebook.
The dashboard shows time trends, coach performance, most active weekdays, and class distribution by geography, category, level, and platform.
All charts are interactive and interconnected: selecting a feature in one chart updates the others accordingly.
The layout is designed to be clear, intuitive, and accessible to non-technical stakeholders—ideal for use during a professional presentation.

3. **PDF Document (Current)**

Summarizes the project's main insights, with a final section dedicated to strategic reflections, operational recommendations, and key questions for potential decision-makers.
It can be used as a standalone document or as supporting material for a portfolio or professional presentation.

Strategic Conclusions & Recommendations

Key Business Questions

These strategic questions emerged directly from the insights found in:

- **Step 3** (*Data Manipulation & Visualization*)
- **Step 4** (*ML Implementation & Strategy*)

From user behavior (Step 3):

1. Why do many users wait over a year before attending their first class?
2. Would it make sense to launch seasonal campaigns to stimulate activity during colder months (Oct–Feb)?
3. Can we reduce early churn from "one-timers" by encouraging them to join Intermediate (higher-rated) instead of Basic classes?
4. Should we rethink coach assignment, given that the most active coaches receive lower average ratings?

From modeling and strategy (Step 4*)*

5. With a recall of 54%, are we willing to lose 46 out of 100 churners?
6. Is it better to optimize the model or invest in generalized retention strategies?
7. What's the ROI of saving a user with a €50 LTV via a €5 retention campaign?
8. Which variables are the most predictive of churn (e.g., days until first class)?
9. Does it make more sense to predict low ratings rather than exact values from 1 to 5?
10. Do different user clusters react differently to churn or satisfaction drivers?
11. Should we reward loyal users (Cluster 0) or focus efforts on the least active ones (Cluster 2)?
12. What factors promote transition from Cluster 1 (moderate) to Cluster 0 (loyal)?

Operational Conclusions & Improvement Actions

Churn Prediction

• Use **Logistic Regression** as a first tool to identify at-risk users.
• Integrate churn score into the app, triggering notifications within the first 7 days after registration.
• Prioritize **Cluster 2 (new users)** in retention strategies.

Rating & Class Quality

• Switch from predicting exact ratings to a classification approach (low / medium / high).
• Reevaluate or reposition Basic-level classes, which are often underrated.
• Promote Intermediate-level classes to new users.

User Segmentatiom

• Use **Cluster 0 (loyal users)** as a test group for new features or loyalty rewards
• Launch targeted re-engagement campaigns for **Cluster 2** (7–14 days of inactivity).
• Track transition from **Cluster 1 to 0** as a key loyalty KPI.

Temporal & Behavioral Strategy

• Offer themed classes or bonus events during low-frequency times (e.g., Sundays).
• Introduce a guided welcome flow to reduce time to first class.
• Improve experience on underperforming but widely used platforms (e.g., TV).