Domain Knowledge: Travel Package Purchase Dataset

This document provides domain-level understanding of the Travel Package Purchase dataset. The dataset is commonly used in customer analytics, sales prediction, and marketing intelligence to identify customers who are most likely to purchase travel packages after a sales pitch.

Business Context

Travel and tourism companies rely heavily on direct sales and customer engagement. Sales representatives pitch travel packages to customers and follow up to convert leads. This dataset captures demographic, behavioral, and interaction-based attributes that influence a customer's decision to purchase a travel package.

Business Objective

The primary objective is to predict whether a customer will purchase a travel package (ProdTaken = 1) or not (ProdTaken = 0). Accurate predictions help businesses improve lead targeting, optimize sales efforts, and increase overall conversion rates.

Feature Categories

- **Customer Demographics:** Age, Gender, MaritalStatus, CityTier, Occupation, Designation, MonthlyIncome
- **Sales Interaction Details:** TypeofContact, DurationOfPitch, NumberOfFollowups, PitchSatisfactionScore
- **Travel Preferences:** ProductPitched, PreferredPropertyStar, NumberOfTrips
- **Household & Mobility:** NumberOfPersonVisiting, NumberOfChildrenVisiting, OwnCar, Passport

Key Columns Explained

- **ProdTaken:** Target variable indicating whether the customer purchased the package.
- **Age:** Helps understand purchase behavior across different age groups.
- **MonthlyIncome:** Strong indicator of affordability and purchasing power.
- **DurationOfPitch:** Longer pitches may indicate higher customer engagement.
- **NumberOfFollowups:** Reflects sales effort intensity.
- **ProductPitched:** Type of package offered influences purchase likelihood.
- **PitchSatisfactionScore:** Direct measure of customer response to sales pitch.
- **Passport:** Indicates readiness for international travel.

Business Use Cases

- Lead scoring and prioritization for sales teams
- Targeted marketing and personalized offers
- Improving sales pitch strategies
- Customer segmentation and profiling

Relevance to Machine Learning

This dataset is well-suited for binary classification problems. It contains a mix of numerical and categorical features, making it ideal for models such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and other ensemble techniques. Preprocessing using encoding and scaling is essential for optimal model performance.