Business Intelligence Project: ETL, Dimensional Modeling, and Report Generation in Power BI

Gili, Paula

Master's in Business Intelligence and Big Data Analytics
Universitat Oberta de Catalunya
paualgili.bi@gmail.com

This project is part of the Master's in Business Intelligence and Big Data Analytics at the Universitat Oberta de Catalunya. The goal of the project is to apply ETL (Extract, Transform, Load) principles, dimensional modeling, and report generation using Power BI for data analysis in the hotel industry.

1. Project Overview

A Business Intelligence (BI) project begins with the capture and preprocessing of incoming data, a process known as ETL. The ETL process consists of three key stages:

a) Extraction: Collecting data from various relevant sources.
b) Transformation: Processing the data according to predefined business rules.
c) Loading: Storing the transformed data in a target data warehouse.

For this project, we'll be working with three primary data sources, alongside an additional source selected by the team. In the first phase, we focus on loading the main data repository, which holds reservation data for a hotel group. In the subsequent phase, we will enhance this data using web scraping, integrating external sources to complement the original dataset.

2. Dimensional Modeling

Within the scope of this project, we will design a star schema data model by identifying the different dimensions and the fact table required. This model is straightforward due to the absence of data inconsistencies and cyclic dependencies, making it easier to structure. We will also define the relationships between the fact table and the various dimensions.

3. Using DAX for Metric Creation

A critical element of this project is the use of DAX (Data Analysis Expressions), the scripting language used within Power BI. While DAX expressions are not mandatory for basic report creation, they are essential for key tasks, including:

a) Aggregation calculations (e.g., totals, percentages).
b) Date tables for filtering data by time dimension.
c) Iterators and complex calculations.

This project emphasizes the use of measures over calculated columns (which are generally discouraged). We will explore the advantages and disadvantages of both approaches, guiding you to choose the best solution based on the project requirements.

4. Report Generation

Once the ETL and modeling steps are complete, the next phase involves generating interactive reports. Reports will visually represent key business data (KPIs), offering actionable insights. In this phase, the project focuses on utilizing the most effective Power BI visuals to build interactive and user-friendly reports. Customization and formatting options for each visual will also be covered to ensure a polished presentation.

5. Project Objectives

- Apply an ETL process to collect and transform data from various sources.
- Model the data using a star schema dimensional model.
- Leverage DAX to create key metrics (KPIs) and advanced calculations.
- Generate interactive, visual reports in Power BI to support data-driven decision-making.

6. Tools and Resources

- Power BI: Used for data modeling and generating visual reports.
- Internal data sources: Hotel group reservation data.
- External data sources: Collected via web scraping.
- DAX: For creating measures and performing advanced calculations.