ENHANCING CHINOOK MUSIC STORE PERFORMANCE

Exploratory Data Analysis (EDA)

Proposed by: Alanoud Alosaimi

INTRODUCTION

The Chinook Music Store contains information about the artists, songs, and albums, as well as information on the shop's employees, customers, and the customers purchases. In this project, I will use Exploratory Data Analysis (EDA) with the use of summary statistics and graphical representations, refers to the crucial process of completing early investigations on data in order to uncover patterns, spot anomalies, test hypotheses, and check assumptions. I'll assist the Chinook team in understanding the media in their shop, their customers and employees, and their invoice information.

First, I'll begin the analysis by identifying the best-selling genres. Second, analyzing employee sales performance, when a customer makes their first purchase at the Chinook store, they are assigned a sales support agent. I'll examine this connection to determine which sales support agents are performing the best. Then, analyze sales by country to determine the total number of customers and total sales value for each country.

DATA DESCRIPTION

There are 11 tables in the chinook sample database:

- employees table stores employees data such as employee id, last name, first name, etc. It also has a field named ReportsTo to specify who reports to whom.
- customers table stores customers data.
- invoices & invoice_items tables: these two tables store invoice data. The invoices table stores invoice header data and the invoice_items table stores the invoice line items data.
- artists table stores artists data. It is a simple table that contains only the artist id and name.
- albums table stores data about a list of tracks. Each album belongs to one artist. However, one artist may have multiple albums.
- media_types table stores media types such as MPEG audio and AAC audio files.
- genres table stores music types such as rock, jazz, metal, etc.
- tracks table stores the data of songs. Each track belongs to one album.
- playlists & playlist_track tables: playlists table store data about playlists. Each playlist contains a list of tracks. Each track may belong to multiple playlists. The relationship between the playlists table and tracks table is many-to-many. The playlist_track table is used to reflect this relationship.

TOOLS

I will be using the Jupyter notebook to explore and analyze the data, and I'll be utilizing Python libraries like:

- Sqlalchemy to write SQL code.
- Numpy and pandas for reading the data and writing operations.
- Matplotlib for data visualization.