

music Album Sales data analysis

By Megha S M



Contents

[Introduction 2](#_Toc70522935)

[Objective 2](#_Toc70522936)

[Initial database and tool setup 2](#_Toc70522937)

[Data Analysis using SQL 2](#_Toc70522938)

[Data Exploration 2](#_Toc70522939)

[Advanced SQL analysis 3](#_Toc70522940)

[Data Visualization using Power BI 4](#_Toc70522941)

[Advanced visualization features 4](#_Toc70522942)

[Conclusion and Learning 5](#_Toc70522943)

[Reference 6](#_Toc70522944)

# Introduction

## Objective

Analyzing music album sales data to identify highest selling artist/tracks. Exploring various data analysis tools and technologies like SQL, PowerBI in this journey.

## Initial database and tool setup

* **PowerBI** – Microsoft Business Intelligence tool.
* **DBeaver** – DBeaver is a SQL client software application and a database administration tool. For relational databases it uses the JDBC application programming interface to interact with databases via a JDBC driver.
* **SQL Lite** – Used SQLlite sample database available in DBeaver.
* **Github** – Used Git as version control tool storing all project artifacts.

[Git repository link](https://github.com/meghasiddappa/1st-data-analysis-project)

# Data Analysis using SQL

## Data Exploration

**Downloaded the Sample dataset for Project (Album and Sales Tracker)**

This database contains various tables, below are the detail.

* Album: it contains details AlbumId, Name, artist number
* Artist: it has the details like artist name, artist id
* Customer: the details of the customer who has purchased the tracks
* Employee: Details of the employee who has sold the tracks and some of the employees are artist too.
* Genre: it has the details of the different genres
* Invoice: It contains all the invoice line details.
* Invoice Line: For every track sale there is invoice which contains the details of customers who purchased and invoice id.
* Playlist: it has the details of all the tracks and which type of playlist it belongs.
* Track: Contains details of artist, genre, invoiceid, media typeid.
* Media Type: Detailed Mediatype names with their id.

**Analyze the relationship between the tables and values**

There are different relationships between the tables like, one to one, one to many, many to many, many to one. In the dataset the relationship will be built between the columns. Identified Primary and Foreign keys in the different tables.

**Exploratory data analysis using simple SQL queries.**

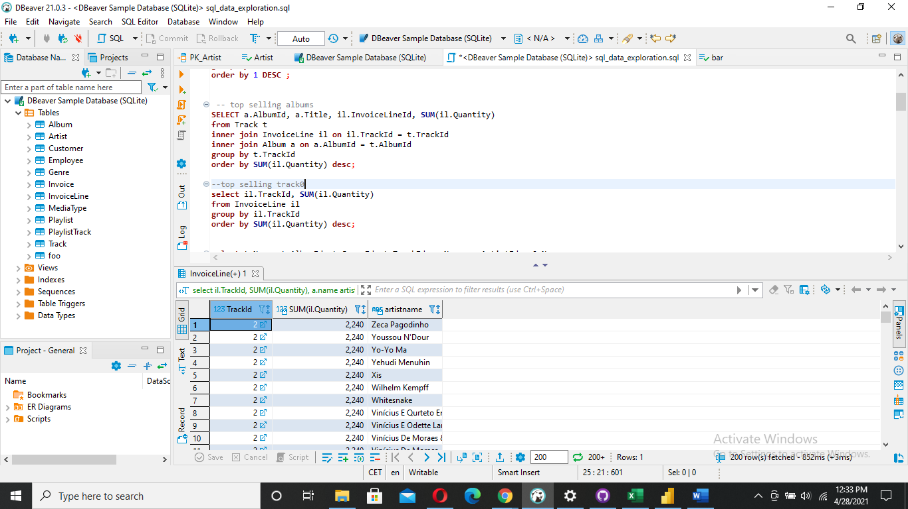


Fig1: SQL queries Execution.

Explored various SQL functionalities:

* Select
* From
* Group By
* Having
* Order by

## Advanced SQL analysis

Explored Advanced SQL analytics functions for in-depth analysis.

* Rank, Dense Rank
* Partition By clause
* Various aggregation functions like Min, Max, Count, Sum etc

[SQL Queries Github link](https://github.com/meghasiddappa/1st-data-analysis-project/blob/main/sql_data_exploration.sql)

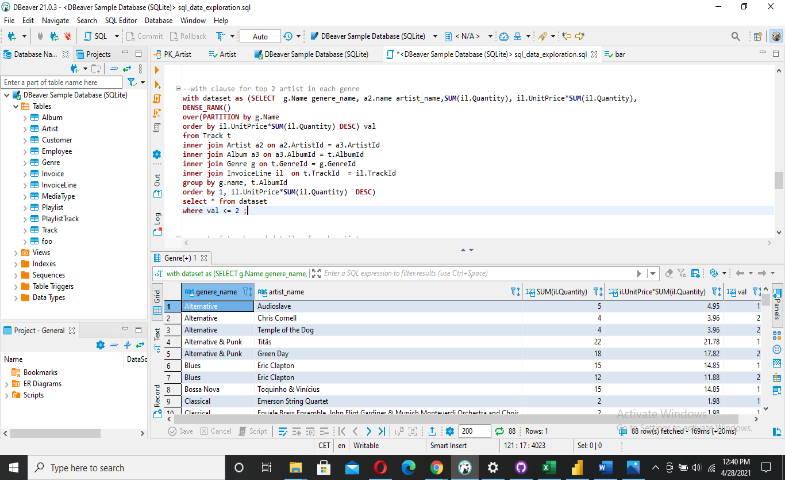


Fig 2: Execution of Advanced Query

# Data Visualization using Power BI

Power BI is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities. In this step using PowerBI platform 4 different analysis are shown.

1. Artist table with highest sales
2. Album with highest sales
3. Genre with highest sale
4. Customer with highest sale

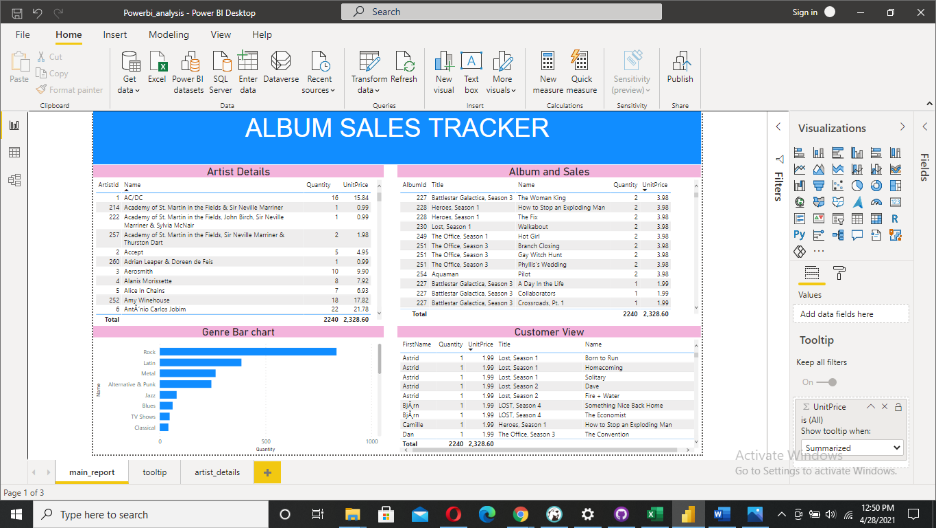


Fig 3: 4 different analysis

## Advanced visualization features

Implemented Visual tool tip and Drill Through functionality.

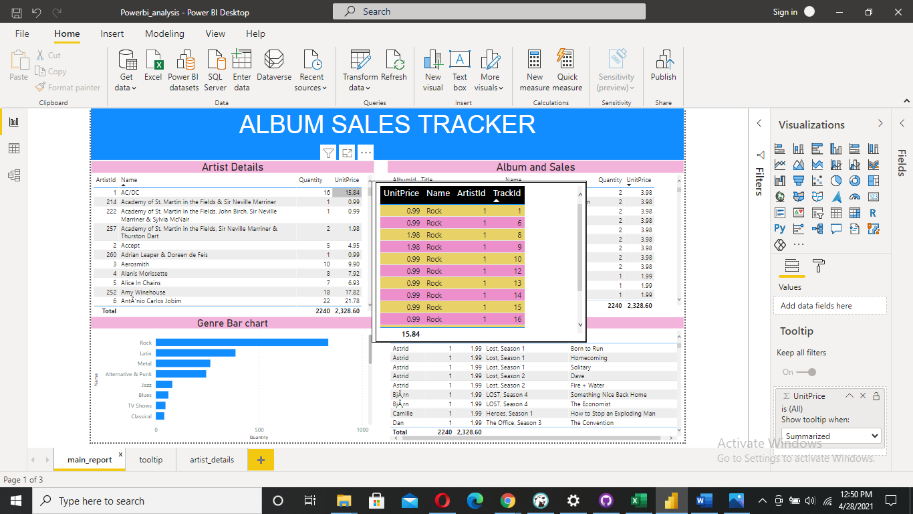


Fig4: Customize tool tip usage

Drill through is one more feature of the PowerBI. With drill through in Power BI reports, you can create a page in your report that focuses on a specific entity such as a customer, artist or quantity.

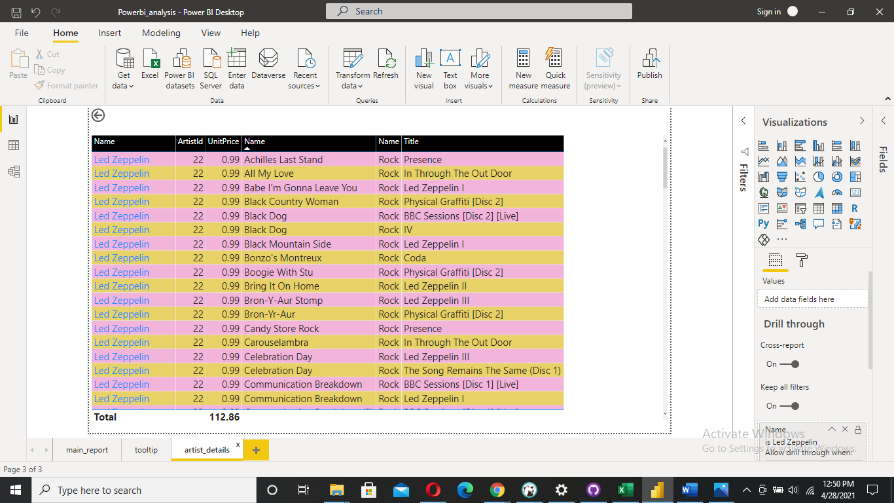


Fig5: Drill through feature

# Conclusion and Learning

Analyzed Music album dales data and identified best and poor performing Genre, Artists, Tracks.

Top 3 best performing Genre: ROCK, Latin, Metal

Top 3 best performing Artist: Iron maden, U2, Metallica

Top 3 best performing Album: the women king, the exploring man, the fix

In this journey of data analysis self-project:

* Envisioned own project, laid out the project plan and completed it successfully.
* Understood the usage of database client DBeaver.
* Analyzed the data using SQL queries, joining multiple tables.
* Explored various SQL analytical functions.
* Imported data to Power BI and built relationship between different datasets using Power BI data modeler.
* Saved all queries in version control tool: Git.
* Created multiple interactive visualizations in Power BI desktop.
* Used advanced visualization features like Visual tool tip and Drill through.
* Prepared project document and documented all steps and processes followed.

# Reference

1. <https://www.google.com>
2. <https://www.udemy.com/course/data-analytics-essentials-with-power-bi/learn/lecture/17506134#overview>
3. <https://www.youtube.com/watch?v=4Qy-bksqExI&ab_channel=GuyinaCube>