**ONLINE MUSIC STORE PROJECT**

In this project, I will be working with the dataset of an online music store to explore, analyze and visualize the insights gotten from the dataset

The data was gotten from Rishabh Mishra and includes 11 tables.

The analysis includes 4 stages which were done using two tools

1) Defining the questions

2) Processing: Microsoft SQL Server

3) Analyzing: Microsoft SQL Server

4) Visualization: Microsoft Excel

Defining the questions: the first step was to ask questions that could give better insight into the dataset and help the analysis stage.

The following were the questions asked:

Q1: From the employee table, who is the highest-ranking officer?

Q2: Which countries have the most invoice?

Q3: What are the top 3 values of total invoices?

Q4: What city has the highest sum of invoice total?

Q5: Who is the best customer? (ie the customer that spent the most)

Q6: What are the email, first name, last name, and genre of all Rock music listeners (Order alphabetically by email)

Q7: What are the top 10 rock bands, the artist's name and their total count of tracks?

Q8: what songs are longer than the average milliseconds?

Q9: Who are the top 20 selling artist (artists whose music were purchased the most)

Q10: Who are the top 10 fans of the top artist? From the previous query, the top artist is Queen.

Q11: who are the favourite artists of each customer?

Q12: What is the most popular genre in each country?

Q13: How much did each customer spend on music in each country?

Q14: what are the most purchased tracks?

Q15: What are the top 10 best-selling albums

Q16: How many albums does each artist have?

Q17: How many albums does each artist have?

Q18; What is the most popular media type?

Processing: This stage includes gathering the dataset, inspecting it, and cleaning them.

The first step was to create a database in the SQL server named Music data store and then import all 11 tables namely:

Album2

Artist

Employees

Invoice

Invoice line

Customer

Genre

Media type

Playlist

Playlist tracks

Tracks

**The other steps taken include:**

Rename the album2 table to just ‘album’

View each table to check if they are formatted properly

Check if all columns have the right data type

Check for null/empty values

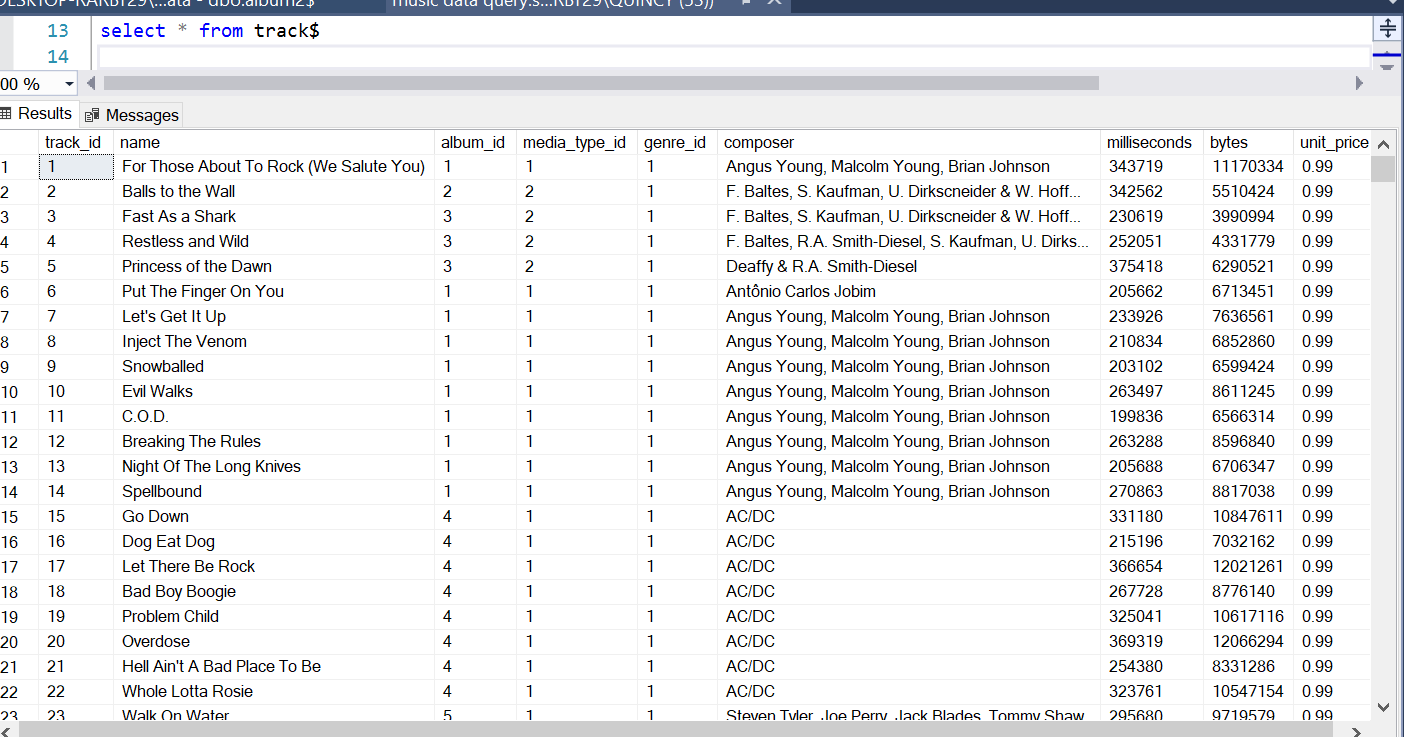
Check for duplicate values

This stage reveals that the album table has 3 columns and 347 rows, the artist table has 2 columns and 275 rows, the employee has 16 columns and 9 rows, the invoice has 9 columns and 614 rows, the invoice line has5 columns 4,757 rows, the customer table has 13 columns and 59 rows, genre table has 2 columns and 25 rows, the media type table had 2 columns and 5 rows, the playlist has 2 columns and 18 rows, playlist track has 2 columns and 8715 rows and finally the track table has 9 columns and 3503 rows.

The customer table has 58 null values in the state column but after inspection, I realized the respective countries do not have states, just cities and provinces.

The track table has 977 null values in the composer’s column. Some of them I was able to clean as the respective artist were also the composers of the tracks.

There are no duplicate values in the columns.



A screenshot of a computer

Description automatically generated

A screenshot of a computer program

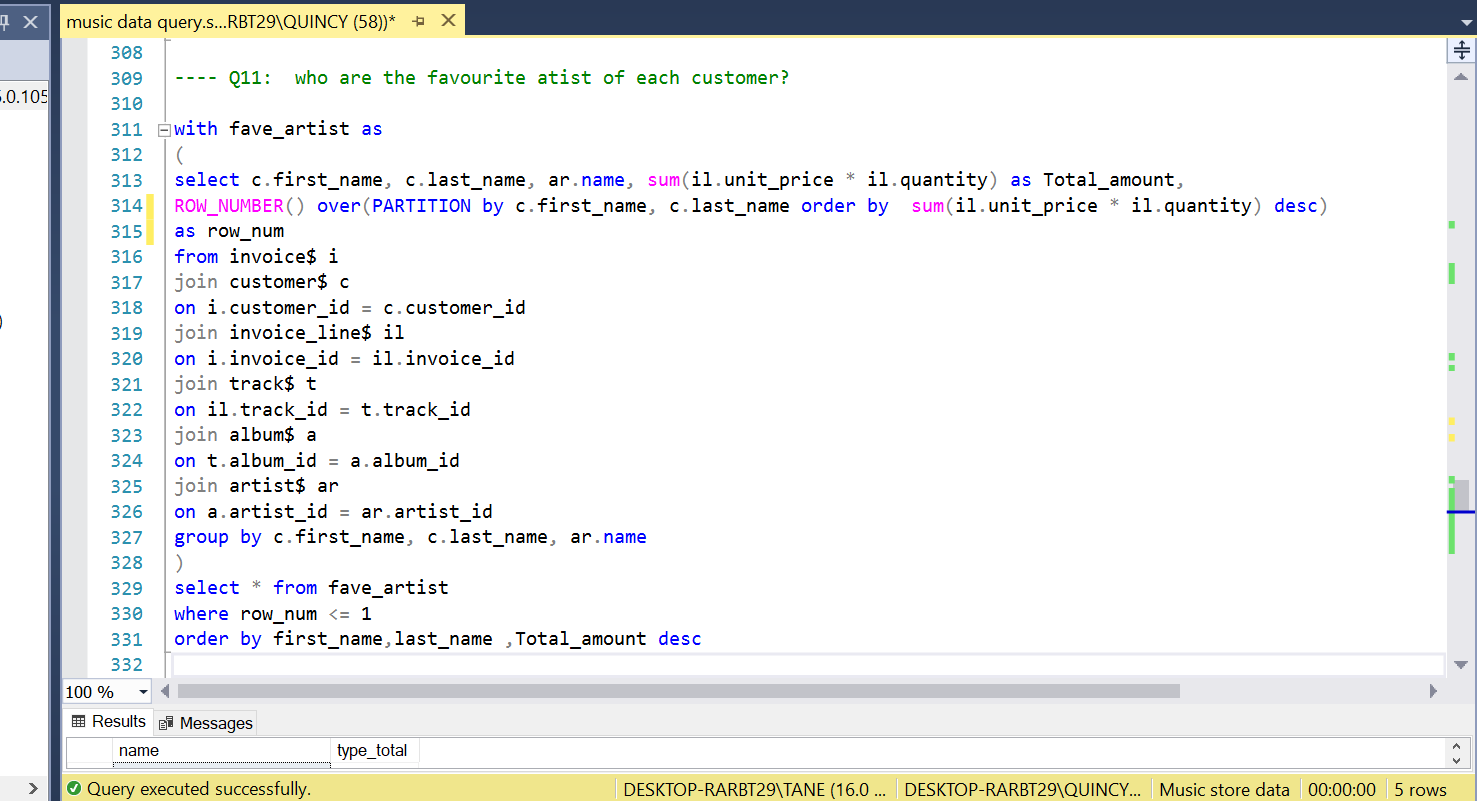
Description automatically generated with medium confidence

Analyzing: This stage of the analysis was also done in the Sql server. Before I started the analysis, I designed the schema of the database to view the data columns easily.

A screenshot of a computer

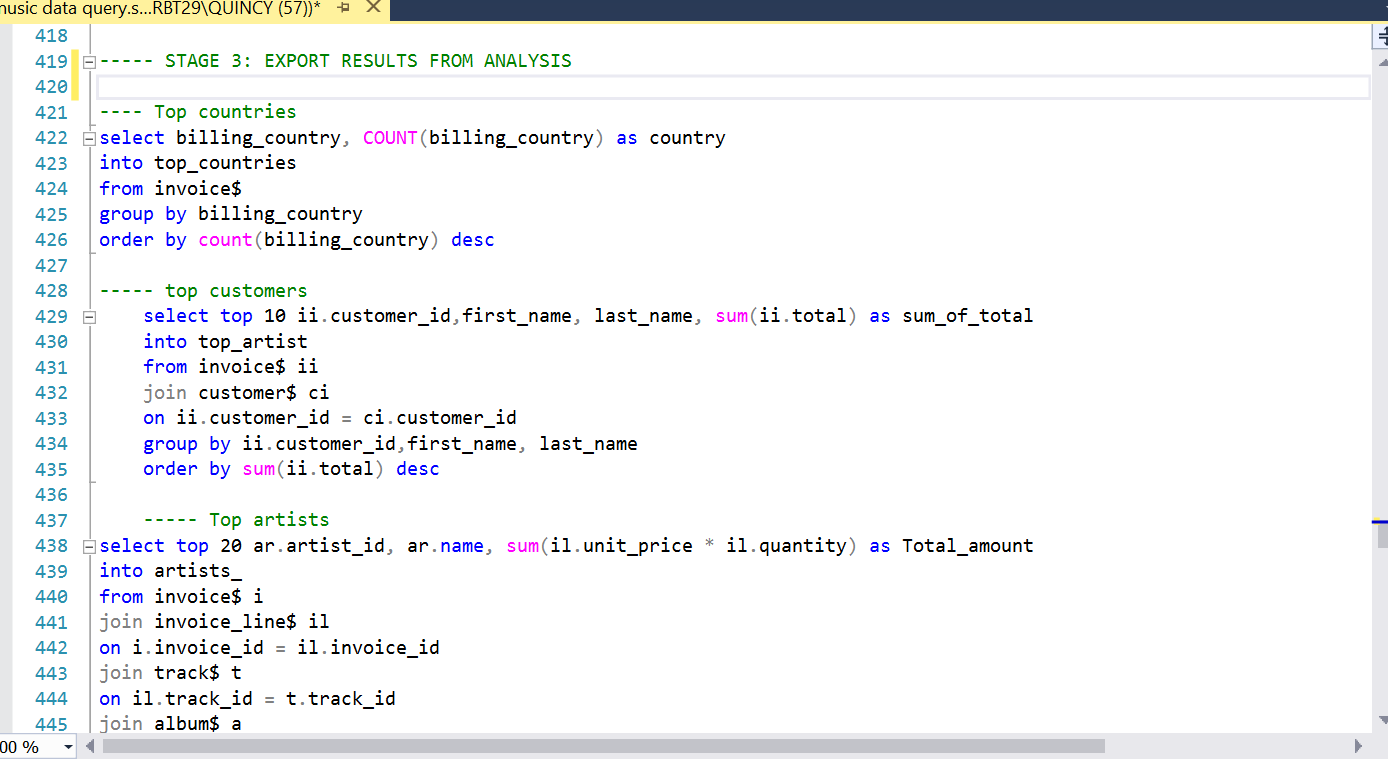
Description automatically generated with medium confidence

After this, I identified the KPIs (Key Performance Indication) such as the top artists, the top-selling song, the most popular genre etc.



A screenshot of a computer program

Description automatically generated with medium confidence



Visualization: After importing the result from sql into excel, it was time to visualize the insights from the analysis. This includes: The top 10 best-selling artists, top 10 best selling albums, top 5 customers, and most listened genre. The visuals were input in an interactive dashboard.

I added all imported data into a data model using power pivot in order to extablish the relationship between the tables, then I created the dashboard using pivot charts.

