

CHINOOK DATA ANALYSIS

By : Mahdieh Alavi



Activate Windows
Go to Settings to activate

About CHINOOK ...

CHINOOK is a chain of **music stores** which manages a wide collection of music albums, songs and user interactions.

This store has **9 employees** that all of them are from **CANADA**, 3 of them are sellers and the rest are in charge of management and coordination.



Available DATA

There are basic information including the **invoices** table as the main information and descriptive tables including :

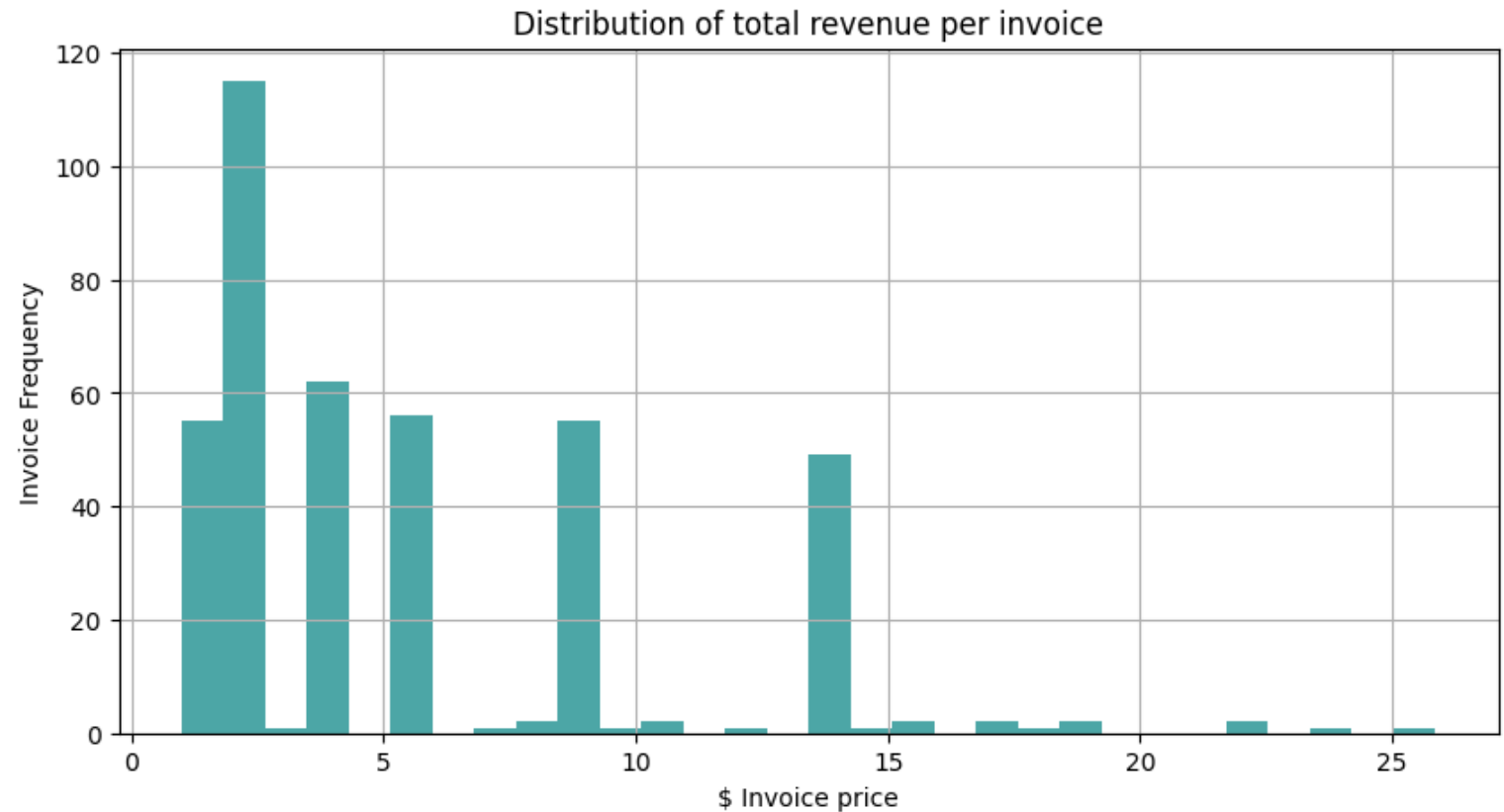
- Customers
- Invoice line
- Employees
- Tracks
- Genre
- Playlist
- Album
- Artist

Key Performance Indicators

- ➔ Total revenue per invoice
- ➔ Quantity of tracks per invoice

Total revenue per invoice distribution

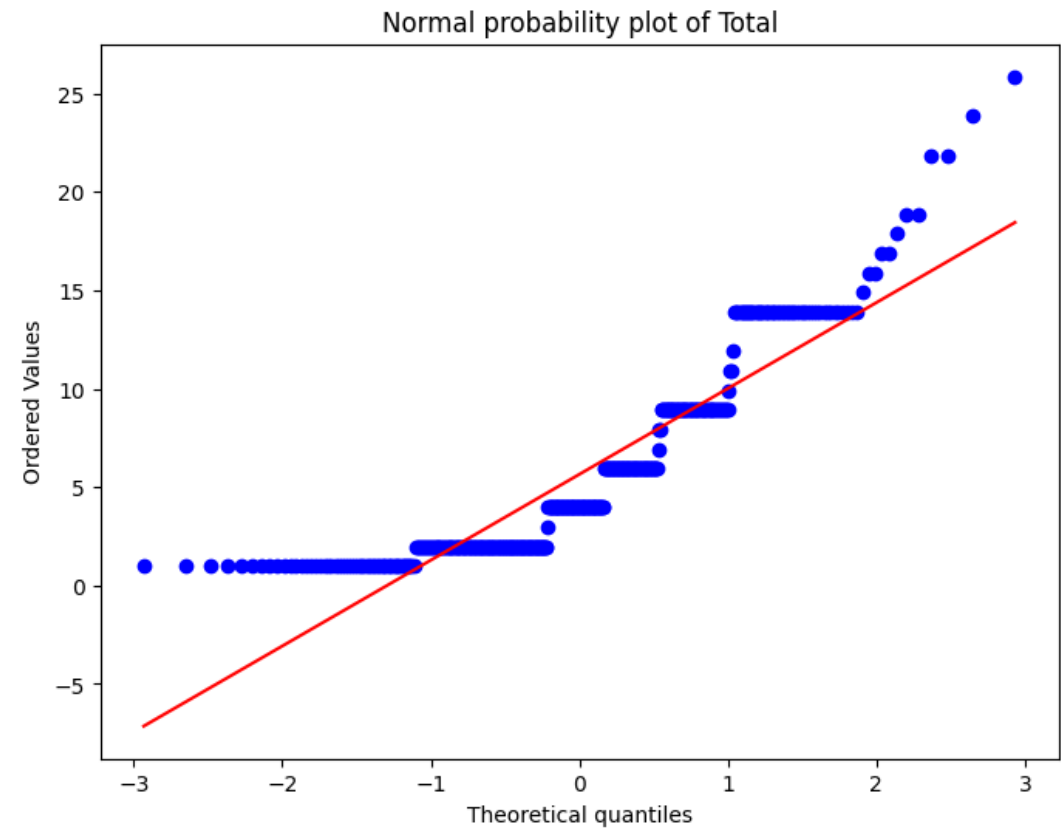
- The price range of invoices is between 1\$ and 25 \$
- Most invoices have an amount between 1\$ and 4\$



Total revenue per invoice distribution

Normality tests results:

- X** Shapiro-Wilk test $<< 0.05$
- X** Kolmogorov-Smirnov test $<< 0.05$
- X** Graphical Assessment



Total revenue per invoice distribution

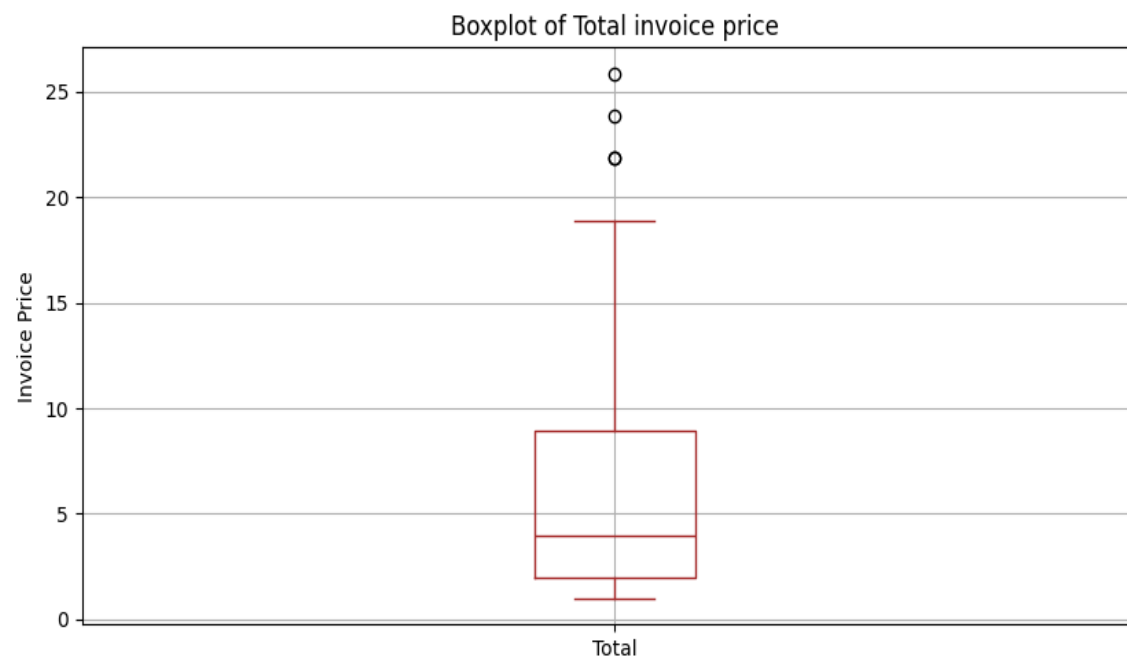
50% of invoice prices are between **1.98 \$** and **8.91 \$**

Abnormal value : **21.86 \$**

21.86 \$

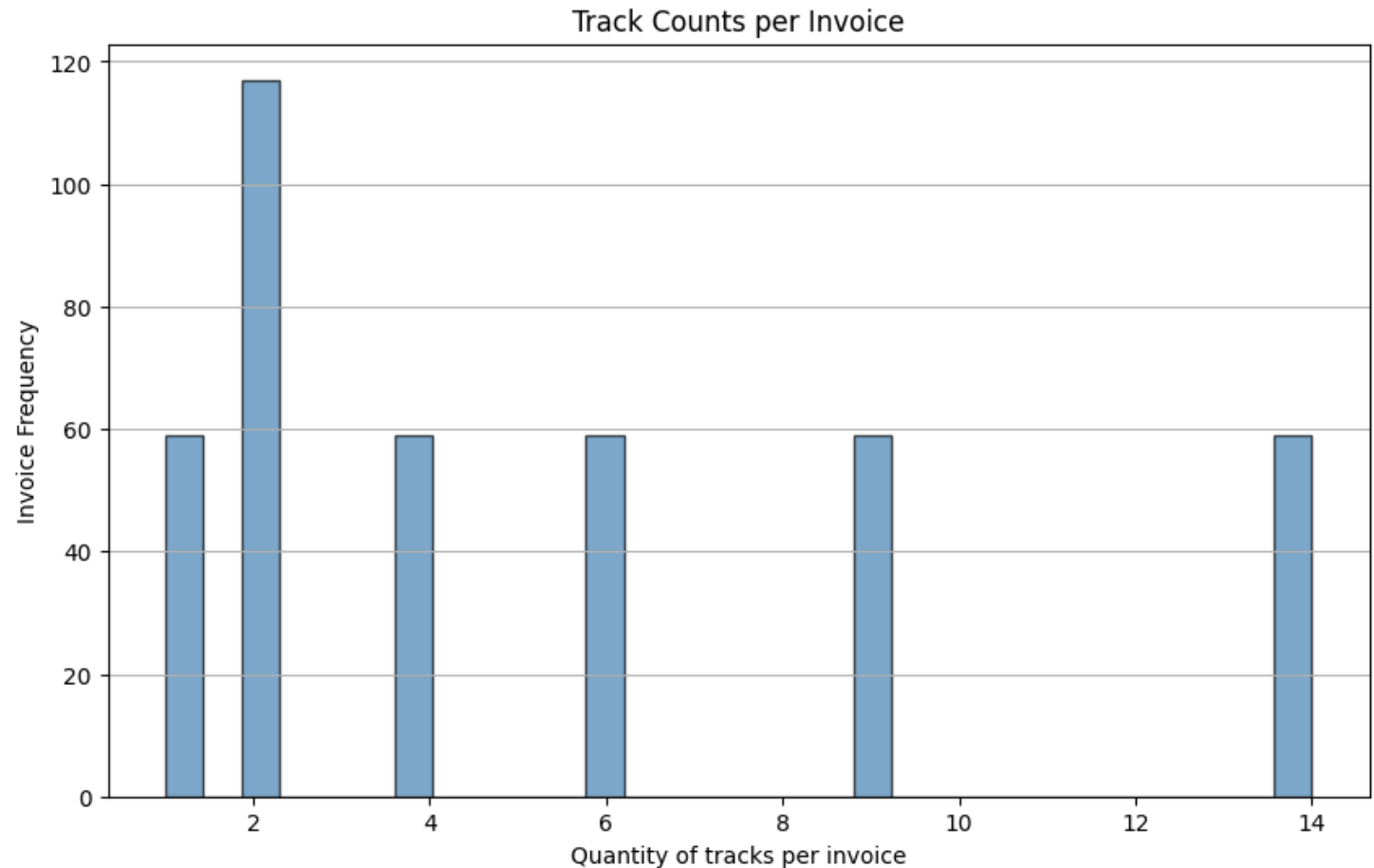
23.86 \$

25.86 \$



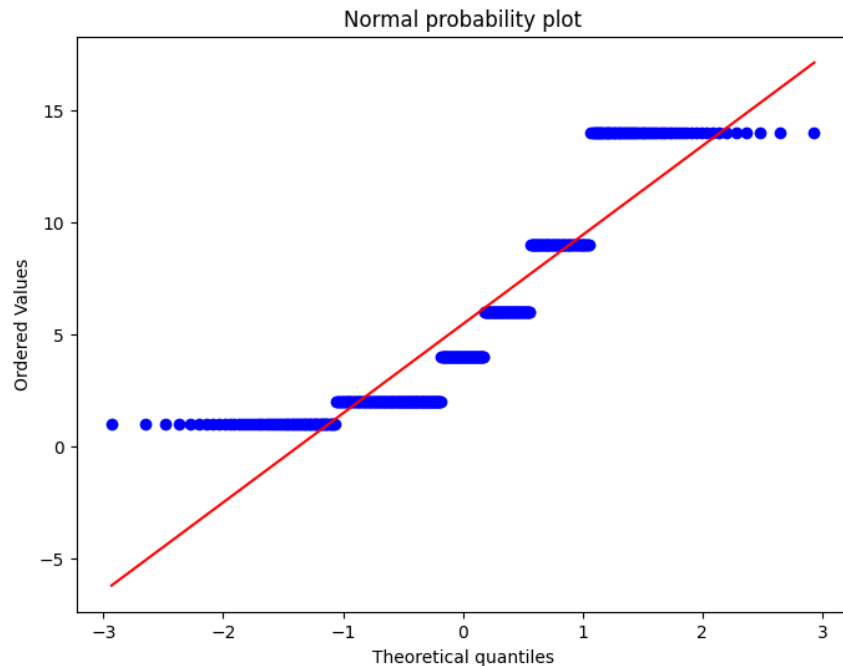
Quantity of tracks per invoice distribution

- Quantity of tracks in invoices is distributed from **1** to **14**
- Most of invoices contain **1** to **4** tracks

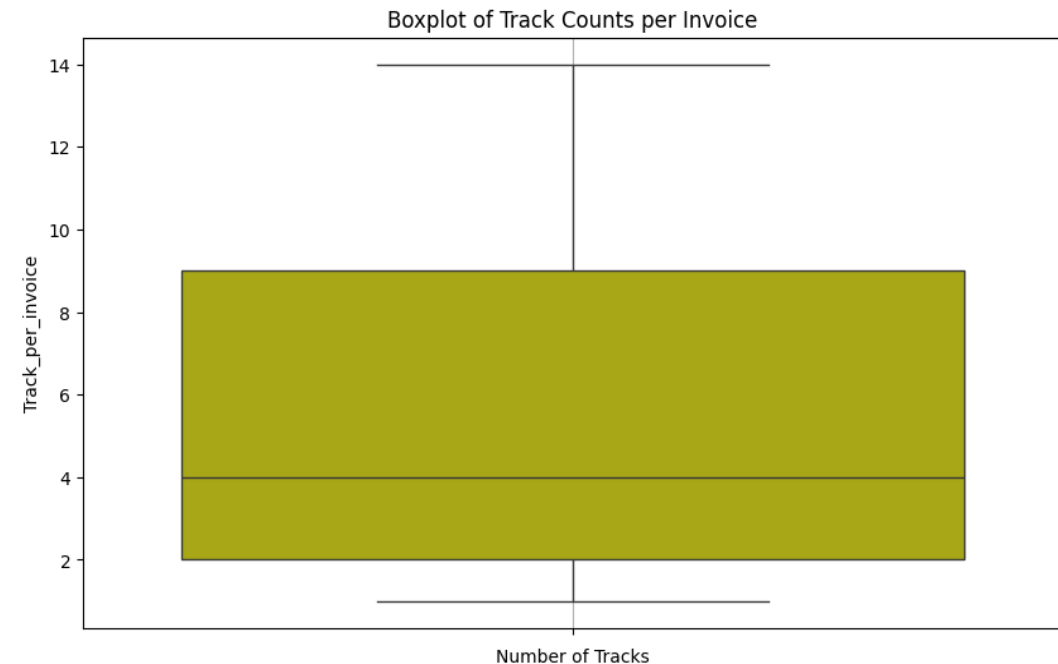


Quantity of tracks per invoice distribution

- ✗ Shapiro-Wilk test $<< 0.05$
- ✗ Kolmogorov-Smirnov test $<< 0.05$
- ✗ Graphical Assessment



50% of invoice tracks count are between 2 and 9



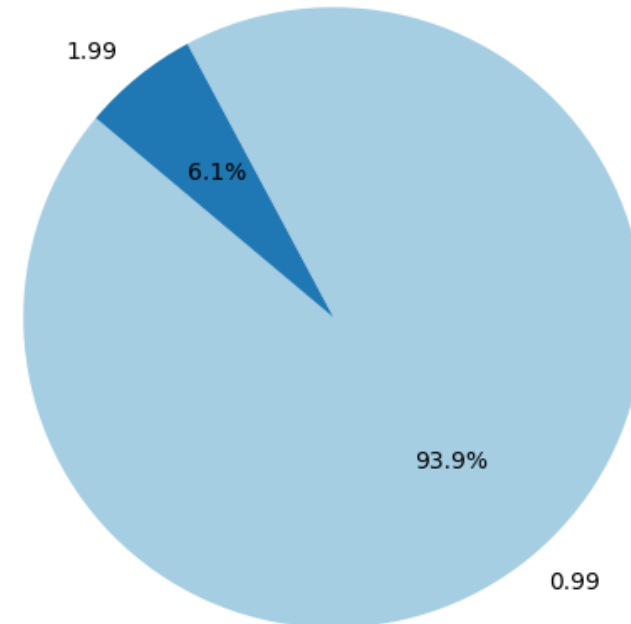
Checking the price range of Tracks

➤ Total number of tracks: 3503

➤ Number of Tracks for price categories:

Unit Price	Frequency
0.99	3290
1.99	213

Distribution of Track Prices



Genre information

➤ **Most Expensive Tracks are in:**

Science Fiction
TV Shows
Sci Fi & Fantasy
Drama
Comedy

➤ **Top 3 popular Genres:**

Number of sell:

Rock 835
Latin 386
Metal 264

Mean price of Genre tracks:

Rock 0.99
Latin 0.99
Metal 0.99

Genre information

Max Average length of song:

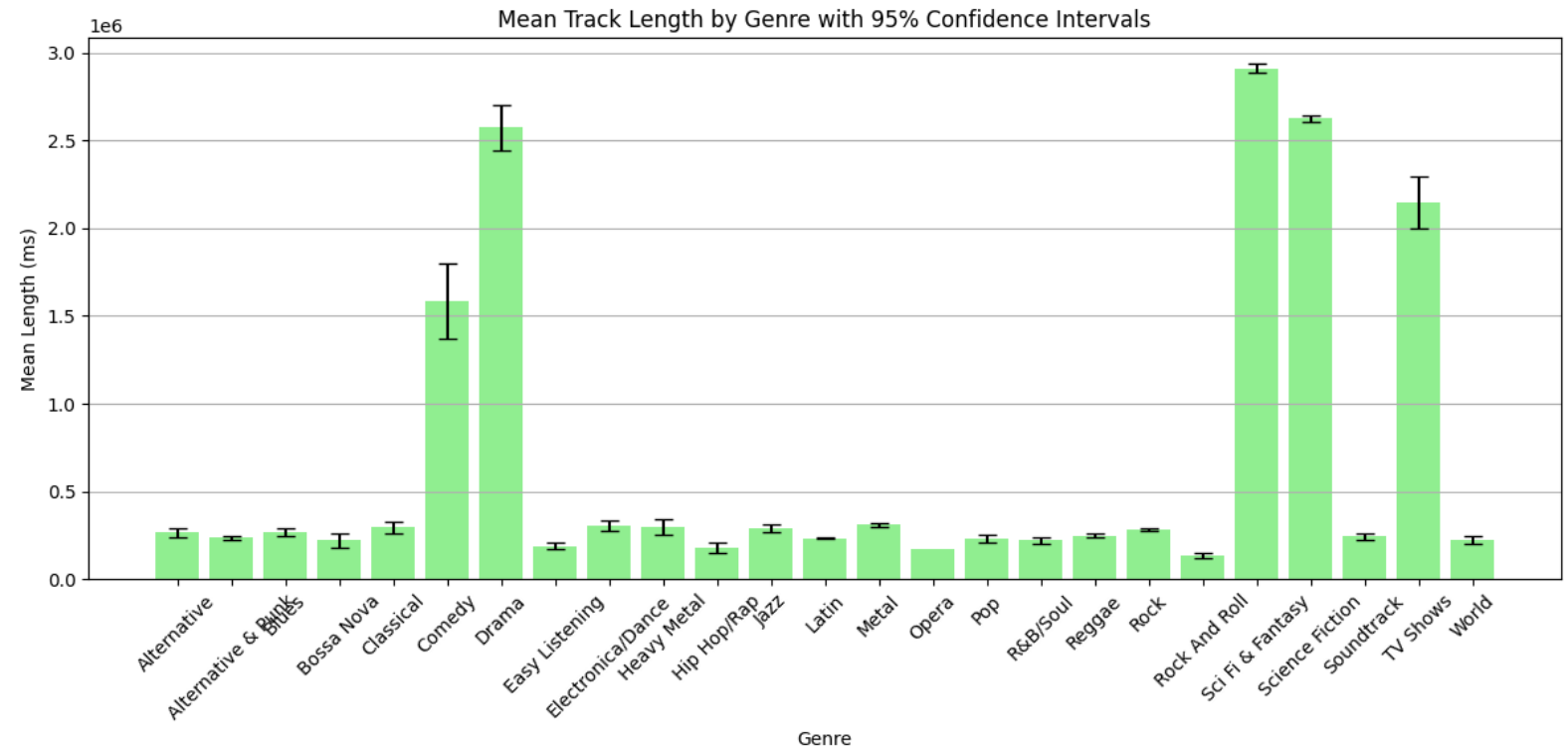
48.52 (Minutes)

Sci Fi & Fantasy

Min Average length of song:

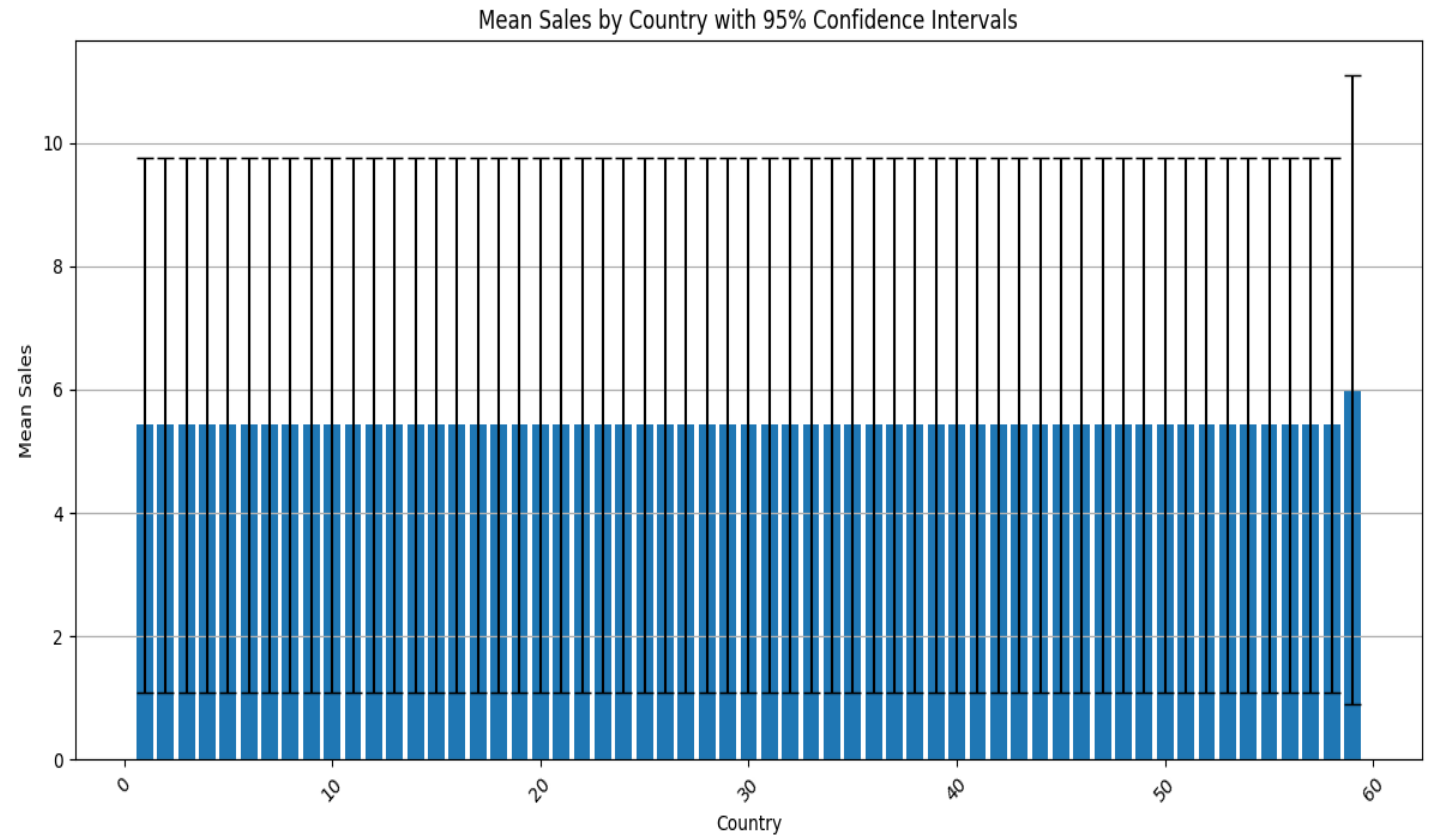
2.24 (Minutes)

Rock And Roll



Analysis of information based on “customers”

- Number of customers: 59
 - Number of Country: 24
 - Average number of purchased songs: 5.43
 - Average number of Purchased by
- 95% Confidence Interval: (5.418, 5.457)

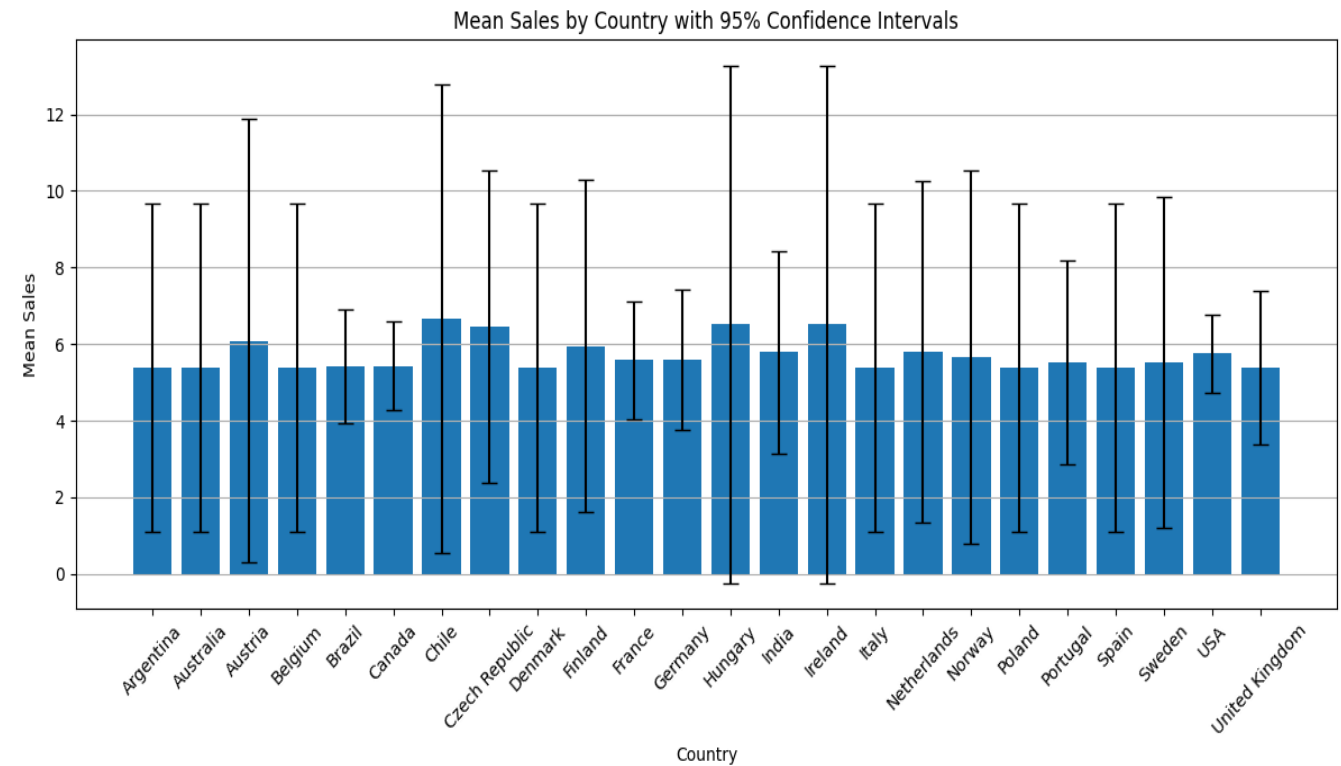


Analysis of information based on “customers”

Average sales in different countries:

The confidence interval in some columns has significant difference from the average, which may be due to:

- The small number of sales
- The large difference between the orders



Examining the relations between parameters

1. Correlation between “track length” and “price” :

To check the independence between the length of the song (Milliseconds) and its price (Unit Price) in the Track table , First we check whether the two distributions are normal or not. Here the distributions were not normal, so we use the “**Spearman hypothesis test**”.

Tests results :

Shapiro-Wilk Test for Track Length: Statistic=0.4069350980582944, p-value=2.7287436202313686e-75

Shapiro-Wilk Test for Price: Statistic=0.25366636273367993, p-value=2.7287436202313686e-75

Spearman Correlation: 0.4099206080846446, p-value: 4.576036687486155e-142

There is a correlation between the length of the song and its price.

Examining the relation between parameters

2. Correlation between “ track size “ and “ price “ by using “Pearson test“ :

To check the independence between the length of the song (BYTS) and its price (Unit Price) in the Track table , First we check whether the two distributions are normal or not. Here the distributions were not normal, so we use the “**Spearman hypothesis test**”.

Tests results :

Shapiro-Wilk Test for Track Size: Statistic=0.2756405817618164, p-value=1.7275565060666367e-79

Shapiro-Wilk Test for Price: Statistic=0.25366636273367993, p-value=1.7275565060666367e-79

Spearman Correlation: 0.41386143347943216, p-value: 4.831360302670962e-145

Reject the null hypothesis: There is a significant correlation between track size and price.

Examining the relation between parameters

3. correlation between 'MediaType' and 'Genre':

For investigating the independence of "**Genre**" and "**MediaType**" using **Chi-Squared**. This test is suitable for checking the independence between two **qualitative** variables.

Test result:

Chi-Squared Statistic: 13519.037479520488

p-value: 0.0012092442869131447

There is a significant association between genre and media type.