This SQL project centers on the Chinook record store. They are seeking to make business descisons based on the data in their database.
Please note that since I'm using jupyter all of my SQL queries must start with " %%sql ".

Objective 1: Assess which albums American Chinook customers would most likely purchase
Objective 2 : Determine the total amount of sales assigned to each sales support agent
Objective 3: Analyze Sales by Country

In [1]:
```
%%capture
%load_ext sql
%sql sqlite:///chinook.db
```

Out[1]:
```
'Connected: None@chinook.db'
```

The Chinook Database

In [2]:
```
%%sql

SELECT
    name,
    type
FROM sqlite_master
WHERE type IN ( "table", "view");
```

```
Done.
```

Out[2]:

name	type
album	table
artist	table
customer	table
employee	table
genre	table
invoice	table
invoice_line	table
media_type	table
playlist	table
playlist_track	table
track	table

Chinook DATABASE SCHEMA

In [5]:
```
%%sql


SELECT * FROM album
limit 3;
```

Done.

Out[5]:

album_id	title	artist_id
1	For Those About To Rock We Salute You	1
2	Balls to the Wall	2
3	Restless and Wild	2

In [8]:
```
%%sql

SELECT * FROM artist
LIMIT 3;
```

Done.

Out[8]:

artist_id	name
1	AC/DC
2	Accept
3	Aerosmith

In [9]:
```
%%sql

SELECT * FROM customer
LIMIT 3;
```

Done.

Out[9]:

customer_id	first_name	last_name	company	address	city	state	country	postal_cod
1	Luís	Gonçalves	Embraer - Empresa Brasileira de Aeronáutica S.A.	Av. Brigadeiro Faria Lima, 2170	São José dos Campos	SP	Brazil	12227-00
2	Leonie	Köhler	None	Theodor-Heuss-Straße 34	Stuttgart	None	Germany	7017
3	François	Tremblay	None	1498 rue Bélanger	Montréal	QC	Canada	H2G 1A

In [10]:
```
%%sql

SELECT * FROM employee
LIMIT 3;
```

Done.

Out[10]:

employee_id	last_name	first_name	title	reports_to	birthdate	hire_date	address	cit
1	Adams	Andrew	General Manager	None	1962-02-18 00:00:00	2016-08-14 00:00:00	11120 Jasper Ave NW	Edmonto
2	Edwards	Nancy	Sales Manager	1	1958-12-08 00:00:00	2016-05-01 00:00:00	825 8 Ave SW	Calgar
3	Peacock	Jane	Sales Support Agent	2	1973-08-29 00:00:00	2017-04-01 00:00:00	1111 6 Ave SW	Calgar

In [12]:
```
%%sql

SELECT * FROM genre
LIMIT 3;
```

Done.

Out[12]:

genre_id	name
1	Rock
2	Jazz
3	Metal

In [13]:
```
%%sql

SELECT * FROM invoice
LIMIT 3;
```

Done.

Out[13]:

invoice_id	customer_id	invoice_date	billing_address	billing_city	billing_state	billing_country
1	18	2017-01-03 00:00:00	627 Broadway	New York	NY	USA
2	30	2017-01-03 00:00:00	230 Elgin Street	Ottawa	ON	Canada
3	40	2017-01-05 00:00:00	8, Rue Hanovre	Paris	None	France

In [14]:
```
%%sql

SELECT * FROM invoice_line
LIMIT 3;
```

Done.

Out[14]:

invoice_line_id	invoice_id	track_id	unit_price	quantity
1	1	1158	0.99	1
2	1	1159	0.99	1
3	1	1160	0.99	1

In [15]:
```
%%sql

SELECT * FROM media_type
LIMIT 3;
```

Done.

Out[15]:

media_type_id	name
1	MPEG audio file
2	Protected AAC audio file
3	Protected MPEG-4 video file

In [16]:
```
%%sql

SELECT * FROM playlist
LIMIT 3;
```

Done.

Out[16]:

playlist_id	name
1	Music
2	Movies
3	TV Shows

In [17]:
```
%%sql

SELECT * FROM playlist_track
LIMIT 3;
```

Done.

Out[17]:

playlist_id	track_id
1	3402
1	3389
1	3390

In [18]:
```
%%sql

SELECT * FROM track
LIMIT 3;
```

Done.

Out[18]:

track_id	name	album_id	media_type_id	genre_id	composer	milliseconds	bytes	unit_pr
1	For Those About To Rock (We Salute You)	1	1	1	Angus Young, Malcolm Young, Brian Johnson	343719	11170334	0
2	Balls to the Wall	2	2	1	None	342562	5510424	0
3	Fast As a Shark	3	2	1	F. Baltes, S. Kaufman, U. Dirkscneider & W. Hoffman	230619	3990994	0

Objective 1: Assess which albums American Chinook customers would most likely purchase

Genres Popular with USA Chinook Customers

To help determine which albums from the new record label should be added to the store first, I will use Chinook's data to find the best-selling genres in the USA.

Code Explanation:

USA_invoices has the invoice_id of every invoice that was billed in the USA. total_tracks_purchased is the total amount of tracks purchased by Americans. USA_invoice_line is a filtered table showing invoice_id, track_id, and quanitity of tracks purchased on invoices that were billed in America. Genre_tracks contains the number of tracks purchased by Genre in the USA.

In [21]:
```
%%sql

WITH USA_invoices AS
    (
    SELECT
        invoice_id
    FROM invoice
    WHERE billing_country = "USA"
    ),

    total_tracks_purchased AS

        (
            SELECT
            SUM(il.quantity) total_tracks
        FROM invoice_line il
        WHERE invoice_id IN USA_invoices
        ),

    USA_invoice_line AS
            (
            SELECT
                 invoice_id,
                 track_id,
                 quantity
             FROM invoice_line
             WHERE invoice_id IN USA_invoices
            ),


    genre_tracks AS
            (
             SELECT
                g.name Genre_Name,
                SUM(usail.quantity) Tracks_Purchased
            FROM USA_invoice_line usail
            INNER JOIN track t
            ON USAIL.TRACK_ID = t.track_id
            INNER JOIN genre g
            ON t.genre_id =g.genre_id
            GROUP BY Genre_Name)

SELECT
    Genre_Name,
    Tracks_Purchased,
    ROUND(Cast(Tracks_Purchased as float) / total_tracks * 100, 1) AS Sales_Pe
rcentage
FROM genre_tracks, total_tracks_purchased
ORDER BY Tracks_Purchased DESC;
```

```
Done.
```

Out[21]:

Genre_Name	Tracks_Purchased	Sales_Percentage
Rock	561	53.4
Alternative & Punk	130	12.4
Metal	124	11.8
R&B/Soul	53	5.0
Blues	36	3.4
Alternative	35	3.3
Latin	22	2.1
Pop	22	2.1
Hip Hop/Rap	20	1.9
Jazz	14	1.3
Easy Listening	13	1.2
Reggae	6	0.6
Electronica/Dance	5	0.5
Classical	4	0.4
Heavy Metal	3	0.3
Soundtrack	2	0.2
TV Shows	1	0.1

The most popular genres among American Chinook consumers are Rock, Alternative & Punk, and Metal. Songs in the rock genre consist of over 50% of purchased songs for American Chinook users. When looking for new record labels to add, Chinook should prioritize Rock, Alternative & Punk, and Metal before other genres.

In [27]:
```
%%sql

With USA_invoice
        AS (
            SELECT
                customer_id,
                SUM(total) customer_total
            FROM invoice
            GROUP BY customer_id
            HAVING billing_country ="USA"
        )

    SELECT
            (e.first_name|| " " || e.last_name) employee_name,
            SUM(usai.customer_total) total_sales,
            e.reports_to,
            e.hire_date
    FROM employee e
    INNER JOIN customer c
    ON e.employee_id = c.support_rep_id
    INNER JOIN USA_invoice usai
    ON c.customer_id = usai.customer_id
    GROUP BY employee_id
    ORDER BY total_sales DESC;
```

Done.

Out[27]:

employee_name	total_sales	reports_to	hire_date
Margaret Park	497.97	2	2017-05-03 00:00:00
Steve Johnson	337.59000000000003	2	2017-10-17 00:00:00
Jane Peacock	204.93	2	2017-04-01 00:00:00

Objective 2 : Determine the total amount of sales assigned to each sales support agent

In [29]:
```
%%sql



    SELECT
            (e.first_name|| " " || e.last_name) employee_name,
            SUM(i.total) total_sales,
            e.hire_date
    FROM employee e
    INNER JOIN customer c
    ON e.employee_id = c.support_rep_id
    INNER JOIN invoice i
    ON c.customer_id = i.customer_id
    GROUP BY employee_id
    ORDER BY total_sales DESC;
```

Done.

Out[29]:

employee_name	total_sales	hire_date
Jane Peacock	1731.5100000000039	2017-04-01 00:00:00
Margaret Park	1584.0000000000034	2017-05-03 00:00:00
Steve Johnson	1393.920000000002	2017-10-17 00:00:00

Jane Peacock is has the most sales associted with her. This may be due to the extra experience she gained by being employeed in this position the longest.

In [ ]:
```
Sales data by Country
total number of customers
total value of sales
average value of sales per customer
average order value

customer
```

Objective 3 : Analyze Sales by Country

In [12]:
```
%%sql

WITH total_sales
      AS
            (
              SELECT SUM(i.total) total_sales
              FROM invoice i
                  )


SELECT
    c.country Country,
    COUNT(DISTINCT(c.customer_id)) Number_Customers ,
    ROUND (SUM(i.total)/ total_sales * 100, 1) Sales_Percentage,
    ROUND(CAST(SUM(i.total) as float), 2) Country_Total_Sales,
    ROUND(CAST(SUM(i.total) as float)/ COUNT(DISTINCT(c.customer_id)) , 2) Avg
_Sales,
    ROUND(CAST(SUM(i.total) as float)/ COUNT(i.invoice_id), 2) Avg_Order_Value
FROM customer c, total_sales
INNER JOIN invoice i
ON c.customer_id = i.customer_id
GROUP BY c.country
ORDER BY 2 DESC;
```

```
Done.
```

Out[12]:

Country	Number_Customers	Sales_Percentage	Country_Total_Sales	Avg_Sales	Avg_Order_V
USA	13	22.1	1040.49	80.04	
Canada	8	11.4	535.59	66.95	
Brazil	5	9.1	427.68	85.54	
France	5	8.3	389.07	77.81	
Germany	4	7.1	334.62	83.66	
United Kingdom	3	5.2	245.52	81.84	
Czech Republic	2	5.8	273.24	136.62	
India	2	3.9	183.15	91.57	
Portugal	2	3.9	185.13	92.57	
Argentina	1	0.8	39.6	39.6	
Australia	1	1.7	81.18	81.18	
Austria	1	1.5	69.3	69.3	
Belgium	1	1.3	60.39	60.39	
Chile	1	2.1	97.02	97.02	
Denmark	1	0.8	37.62	37.62	
Finland	1	1.7	79.2	79.2	
Hungary	1	1.7	78.21	78.21	
Ireland	1	2.4	114.84	114.84	
Italy	1	1.1	50.49	50.49	
Netherlands	1	1.4	65.34	65.34	
Norway	1	1.5	72.27	72.27	
Poland	1	1.6	76.23	76.23	
Spain	1	2.1	98.01	98.01	
Sweden	1	1.6	75.24	75.24	

Chinook gets 22% of its sales from USA customers and 11.4% from its Canadian customers. Those two countries also have the most customers. The average amount a Chinook customer spends is around $50-100$. Notably the Czech Republic has the highest average sales per customer with $136.62 with only 2 customers. It would be worthwhile to do additional research and marketing for countries in Europe like the Czech Republic.

In [ ]: