WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

Masters Programmes: Assignment Cover Sheet

Student Number:	5609013, 5640147, 5607485, 5663439, 5647189, 5667437
Module Code:	IB9HP0
Module Title:	Data Management
Submission Deadline:	20/03/2025
Date Submitted:	19/03/2025
Word Count:	1979 (Introduction, Thematic main body and Conclusion)
Number of Pages:	30
Question Attempted: *(question number/title, or description of assignment)*	*Group Assignment Questions*
Have you used Artificial Intelligence (AI) in any part of this assignment?	**Yes, used for data generation, for grammar check**

Academic Integrity Declaration

We're part of an academic community at Warwick. Whether studying, teaching, or researching, we're all taking part in an expert conversation which must meet standards of academic integrity. When we all meet these standards, we can take pride in our own academic achievements, as individuals and as an academic community.

Academic integrity means committing to honesty in academic work, giving credit where we've used others' ideas and being proud of our own achievements.

In submitting our work, we confirm that:

- We have read the guidance on academic integrity provided in the Student Handbook and understand the University regulations in relation to Academic Integrity. We are aware of the potential consequences of Academic Misconduct.
- We declare that this work is being submitted on behalf of us and is all our own, except where we have stated otherwise.
- No substantial part(s) of the work submitted here has also been submitted by us in other credit bearing assessments courses of study (other than in certain cases of a resubmission of a piece of work), and we acknowledge that if this has been done this may lead to an appropriate sanction.
- Where a generative Artificial Intelligence such as ChatGPT has been used we confirm we have abided by both the University guidance and specific requirements as set out in the Student Handbook and the Assessment brief. We have clearly acknowledged the use of any generative Artificial Intelligence in our submission, our reasoning for using it and which generative AI (or AIs) we have used. We have taken help from ChatGPT for professional formatting of statements and searching formal synonyms of key words.
- We understand that should this piece of work raise concerns requiring investigation in relation to any of points above, it is possible that other work we have submitted for assessment will be checked, even if marks (provisional or confirmed) have been published.
- Where a proof-reader, paid or unpaid was used, we confirm that the proof-reader was made aware of and has complied with the University's proofreading policy.

Upon electronic submission of your assessment, you will be required to agree to the statements above

Table of Contents

1. Introduction

This report presents the development of a data product for an online audiobook company, ‘Litstream’. The objective is to design and implement a database that supports key business functions, and the data product will enable the company to generate meaningful insights through structured queries and reports. The project involves defining the business context, designing the database schema, implementing the database in SQLite, and generating synthetic data to simulate real-world operations. The final deliverable includes SQL-based business reports and visualisations to support data-driven decision-making.

2. Business Context

2.1 Key Business Functions

Litstream is a fictional online audiobook provider offering a subscription-based model for accessing audiobooks. Litstream's key business functions include user management, audiobook cataloguing, tracking purchases and subscriptions, and capturing user engagement through ratings, reviews, and listening history. Marketing and user analytics are also vital for enhancing recommendations and improving retention strategies.

2.2 Database Mini world

The mini world of this database consists of several key entities. Users represent the customers who sign up for the platform, while Audiobooks encompass the collection of available audiobooks. Transactions record subscription payments, whereas Subscriptions store details of users enrolled in different plans. Additionally, Reviews capture user feedback for audiobooks, and Analytics tracks user activity, providing insights into listening behaviour and preferences.

2.3 Purpose of Data Product

The primary purpose of this data product is to manage customer transactions and provide actionable insights. The reports generated will analyse user preferences, monitor sales and revenue, and support user engagement through personalised recommendations. The system will also identify top-selling audiobooks and popular genres to refine marketing strategies and enhance overall business performance.

2.4 Expected Reports from the Database

The database will generate several key reports for business insights. The User Engagement Report will analyse customer interactions, including average session duration, monthly active users, device preferences, return rates and content satisfaction based on reviews. These metrics will help to assess the company's growth and audience needs.

The Audiobook Performance Report will assess content popularity and user preferences, including the most listened-to audiobooks, highest-rated titles based on reviews and most popular genres by listeners. These insights will help enhance user engagement and retention and drive revenue growth by aligning content with audience interests.

The Revenue Analysis Report will track the platform's financial performance, user spending behaviour, monthly revenue trends, subscription breakdowns and payment method preferences. These insights will inform operational decisions, marketing campaigns, and strategic pricing to maximise revenue and customer satisfaction.

3. Database Design

3.1 Schema design

Entities and attributes of LitStream database are detailed in Appendix A, Table 1.

Junction table, detailed in Appendix A, Table 2, manages many-to-many cardinalities by containing the primary keys of the connected tables.

Relationship and Cardinalities:

- One user is linked to one active subscription at a time (1-to-1).
- Users make multiple payments; each payment is associated with one user (1-to-M).
- Each payment is linked to one subscription, but a subscription can have multiple payments (M-to-1).
- An audiobook can have multiple authors, narrators, and genres; hence it can be linked to multiple audiobooks (M-to-N).
- Narrators speak one language, but a language can have multiple narrators (M-to-1).
- Users write multiple reviews, with each review linked to one user and one audiobook (1-to-M from users to reviews, 1-to-M from audiobooks to reviews).
- Users listen to many audiobooks, and audiobooks have many listeners (M-to-N).

- Users generate multiple analytics entries for different audiobook sessions, with each entry linked to one user and one audiobook (1-to-M from users to analytics, 1-to-M from audiobooks to analytics).
- Each user has one wishlist (1-to-1), but a wishlist can contain multiple audiobooks, and an audiobook can be on multiple wishlists (M-to-N).

3.2 Entity-Relationship Diagram

The Chen's ER diagram is shown in Appendix B. It depicts entities, attributes and relationships described above.

3.3 Database Assumptions

The database was designed with a few assumptions that are highlighted below.

Each user can only have one active subscription at a time. Updating a subscription will override the existing plan. The payments table tracks subscription periods, resulting in constant updates in billing and allowing users to have different billing addresses per payment. The payment date is a subscription start date. Additionally, users can review an audiobook without fully listening. Audiobooks can contain multiple tracks in different languages, with two narrators per language. Users can listen only on registered devices but may switch, overwriting the previous device.

Basic and Premium are the two subscription tiers. Basic users have limited access, while premium users get offline listening, priority service, and full language support with multiple narrators. Monthly or annual billing is an option for both categories, with annual payments offering a discount. To promote premium plans, a short free trial may be offered.

Users can listen to their audiobooks while their membership is "Completed," meaning it hasn't expired or failed in the current cycle. If payment fails or subscription expires, access is paused until payment goes through. User data will be saved.

A many-to-many relationship is proposed between authors and books, allowing multiple authors per book. Each user has a one-to-one relationship with a wishlist, allowing only a single wishlist per user. Users can add many audiobooks to their wishlist, which is captured by many-to-many relationship between audiobooks and wishlist.

4. SQL Schema

4.1 Implementation

The schema creation process involved selecting appropriate data types. TEXT was chosen for most of the variables where text was present as well as for fields including dates and numbers with special characters (i.e. phone numbers starting with +44). Other data types include INTEGER for such variables as rating and publication year as those are represented as numbers, and REAL for progress, amount and price variables as those are represented by decimal numbers. The SQL relational schema, highlighting cardinalities and keys, is visually represented in Appendix C.

4.2 Data Generation

Synthetic data was generated for sixteen distinct entities using Python with the Faker package and Open package API to fill in realistic data, such as usernames, audiobook titles, author names, addresses, and timestamps.

A key aspect is time-based trends in subscriptions and payments. Subscriptions are split into basic (400) and premium (100) and distributed month by month over the past year. Each subscription record thus has a created_date that ensures fewer early-month entries and gradually rising towards the current month. For payments, premium subscriptions are more frequent on recent dates (70% probability of being within the last six months) while basic subscriptions are more common in older windows (70% probability of being 6–24 months ago). Additionally, 20 payment records failed (start/end date = “N/A,” amount=0) and 30 remain pending.

Audiobook_narrators guarantees that each audiobook can have multiple narrators in various languages and that each narrator can narrate multiple audiobooks. There are a total of 20 narrators with 10 different languages where each language is assigned to 2 different narrators (one male and one female). A one-to-many or many-to-many pattern is also used to assign authors to audiobooks, allowing for one to three authors per audiobook. For the first 500 records, genres link to audiobooks one-to-one, but the system allows for a more flexible connection.

Phone numbers follow a ten-digit pattern with country prefixes like +44 or +91, and each user's email is generated from their first and last name with randomly selected parts. Each entity generates a CSV file, which makes it simple to import the data into any relational database and captures trends, many-to-many relationships, and realistic limitations.

Generated tables can be found in Appendix D.

SQL DDL statements can be found in Appendix E.

5. Key Business Findings

5.1 User Engagement Report

User engagement impacts retention, satisfaction, and revenue. Key metrics include:

Returning rate (Figure 1) shows how many users return to listen to more audiobooks, offering insights into customer loyalty and churn. The data indicates that 239 users, representing 47.8% of the total user base, returned to listen. The retention rate remains below 50%, suggesting a relatively low level of user engagement and repeat interactions.

	Total Users	Returning Users	Retention Rate (%)
0	500	239	47.8

Figure 1. Retention Rate

Most audiobook reviews are average or negative, accounting for majority of all reviews, suggesting a need for content improvement (Figure 2) Sentiment analysis could help identify dissatisfaction causes and align content with user needs.

	Rating Category	User Count
0	Average	399
1	Negative	394
2	Positive	207

Figure 2. Review Rating Split

Monthly new subscribers track platform growth and marketing effectiveness. While overall subscription numbers are modest, recent months show slight gradual growth (Figure 3).

	Month	New Subscriptions
0	2025-03	13
1	2025-02	29
2	2025-01	34
3	2024-12	35
4	2024-11	28
5	2024-10	30
6	2024-09	16
7	2024-08	8
8	2024-07	24
9	2024-06	17

Figure 3. New subscribers by month (limited to last 10)

Monthly active users measure platform health by tracking returning users who engage with audiobooks (Figure 4). The number of monthly active users remained stable but experienced a sudden decline from January to March 2025, dropping from 32 to 9, indicating possible retention or engagement challenges that may require investigation and adjustments.

	Month	Active Users
0	2025-03	9
1	2025-02	16
2	2025-01	32
3	2024-12	14
4	2024-11	26
5	2024-10	16
6	2024-09	15
7	2024-08	24
8	2024-07	27
9	2024-06	19

Figure 4. Active Users by month (limited to last 10)

Device preferences reveal insights into how users access the platform, which can guide optimization efforts. The data shows that most users access the platform via mobile devices (phones and tablets), while slightly bigger audience uses desktops (Figure 5).

	Device Type	Total Sessions
0	Desktop	172
1	Mobile	165
2	Tablet	163

Figure 5. Device Split

Session duration measures the average time users spend listening. A decline may suggest dissatisfaction or usability issues. Results indicate a slight decrease in session duration (Figure 6).

	month	Avg Session Duration per User
0	2025-03	01:50:53
1	2025-02	02:54:02
2	2025-01	03:09:55
3	2024-12	03:21:53
4	2024-11	02:53:46
5	2024-10	02:47:27
6	2024-09	02:31:00
7	2024-08	02:53:14
8	2024-07	03:28:34
9	2024-06	02:56:00

Figure 6. Average Session per month (limited to last 10)

5.2 Audiobook Performance Report

Audiobook performance provides insights into audiobook engagement, user preferences, and platform performance. Key metrics include:

Justine holds the highest rating at 5, followed by several books with only two reviews. *Genji Monogatari* leads the most-reviewed books in the top 10 ratings, with an average rating of 4.8, indicating its popularity (Figure 7).

	title	average_rating	total_reviews
0	Justine	5.00	4
1	Confessions	5.00	2
2	Hard Times	5.00	2
3	Harry Potter and the Deathly Hallows	5.00	2
4	Principles of Anatomy and Physiology	5.00	2
5	The art of money getting, or, Golden rules for...	5.00	2
6	Мастер и Маргарита	5.00	2
7	Genji monogatari	4.80	5
8	King Henry V	4.75	4
9	Kritik der reinen Vernunft	4.67	3

Figure 7. Top 10 Highest Average Rated Books

The most popular audiobook languages are Spanish with 188 titles, Russian with 178 titles, and Hindi with 174 titles, indicating a strong demand for multilingual content. English, with 147 titles, ranks lower, suggesting a diverse user demographic with varying language preferences (Figure 8).

	language_name	total_listens
0	Spanish	376
1	Russian	356
2	Hindi	348
3	Indonesian	332
4	Bengali	324
5	French	320
6	Mandarin Chinese	302
7	English	294
8	Portuguese	278
9	Arabic	258

Figure 8. Top 10 Popular Languages

Fantasy, with 103 listens, and Historical, with 96 listens, are the most popular genres. In contrast, Comedy, with 51 listens, ranks as the lowest-performing genre, indicating it remains a niche and underperforming category within the audiobook market (Figure 9).

	0	1	2	3	4	5	6	7	8	9
genre_name	Fantasy	Historical	Horror	Drama	Romance	Thriller	Poetry	Crime	Adventure	Comedy
total_listens	103	96	91	91	89	72	71	68	68	51

Figure 9. Top 10 Popular Genres

5.3 Revenue Analysis Report

Revenue Analysis evaluates the financial performance of the LitStream platform. Key metrics include:

The monthly revenue trend exhibits significant fluctuations over time (Figure 10). A sharp increase in revenue was observed from August to December 2024, culminating in a peak of $15,627 in December. However, the trend indicates a subsequent decline in early 2025.

	month	total_revenue
0	2023-03	1595.0
1	2023-04	5031.0
2	2023-05	6325.0
3	2023-06	6832.0
4	2023-07	4385.0
5	2023-08	4139.0
6	2023-09	8430.0
7	2023-10	3239.0
8	2023-11	3738.0
9	2023-12	6433.0

Figure 10. Monthly Revenue Trends (limited to last 10)

Subscription revenue presents that long-term subscriptions are a key factor in revenue stability, particularly within the basic tier (Figure 11). Basic yearly plans accounted for 61.8% of total revenue, generating $94,311, while basic monthly plans contributed 30.11%, amounting to $45,954. Monthly subscriptions have a comparatively lower impact on overall revenue.

	subscription_level	subscription_type	total_subscriptions	total_revenue	revenue_percentage
0	basic	yearly	189	94311.0	61.8%
1	premium	yearly	46	45954.0	30.11%
2	basic	monthly	179	8771.0	5.75%
3	premium	monthly	36	3564.0	2.34%

Figure 11. Subscription-Based Revenue Breakdown

Total revenue indicates that bank transfers generated the highest earnings at $44,276, followed closely by PayPal transactions at $41,182. The transaction method distribution further supports this trend. Bank transfers accounted for 29.01% of all transactions, making them the leading payment method in terms of total payments. Digital wallets followed closely, comprising 26.99% of the total transaction count (Figure 12).

	payment_method	total_transactions	total_revenue	percent_share
0	Bank Transfer	124	44276.0	29.01%
1	PayPal	118	41182.0	26.99%
2	Debit Card	111	35639.0	23.35%
3	Credit Card	97	31503.0	20.64%

Figure 12. Payment Method Preference

All SQL queries for business insights can be found in Appendix F.

6. Conclusion

The Litstream database was developed using Chen's notation, with synthetic data generated to support key business functions. Assumptions were made to simulate real-world scenarios. The database was tested to validate its integrity and usability. The resulting database facilitates insightful reporting on user engagement, audiobook performance, and revenue analysis, supporting strategic decisions such as optimising audiobook recommendations, refining marketing strategies, and improving customer retention. The structured approach to database development, coupled with reflective problem-solving, ensures that the Litstream database is both functionally robust and scalable for future enhancements.

Appendices

Appendix A: Figures

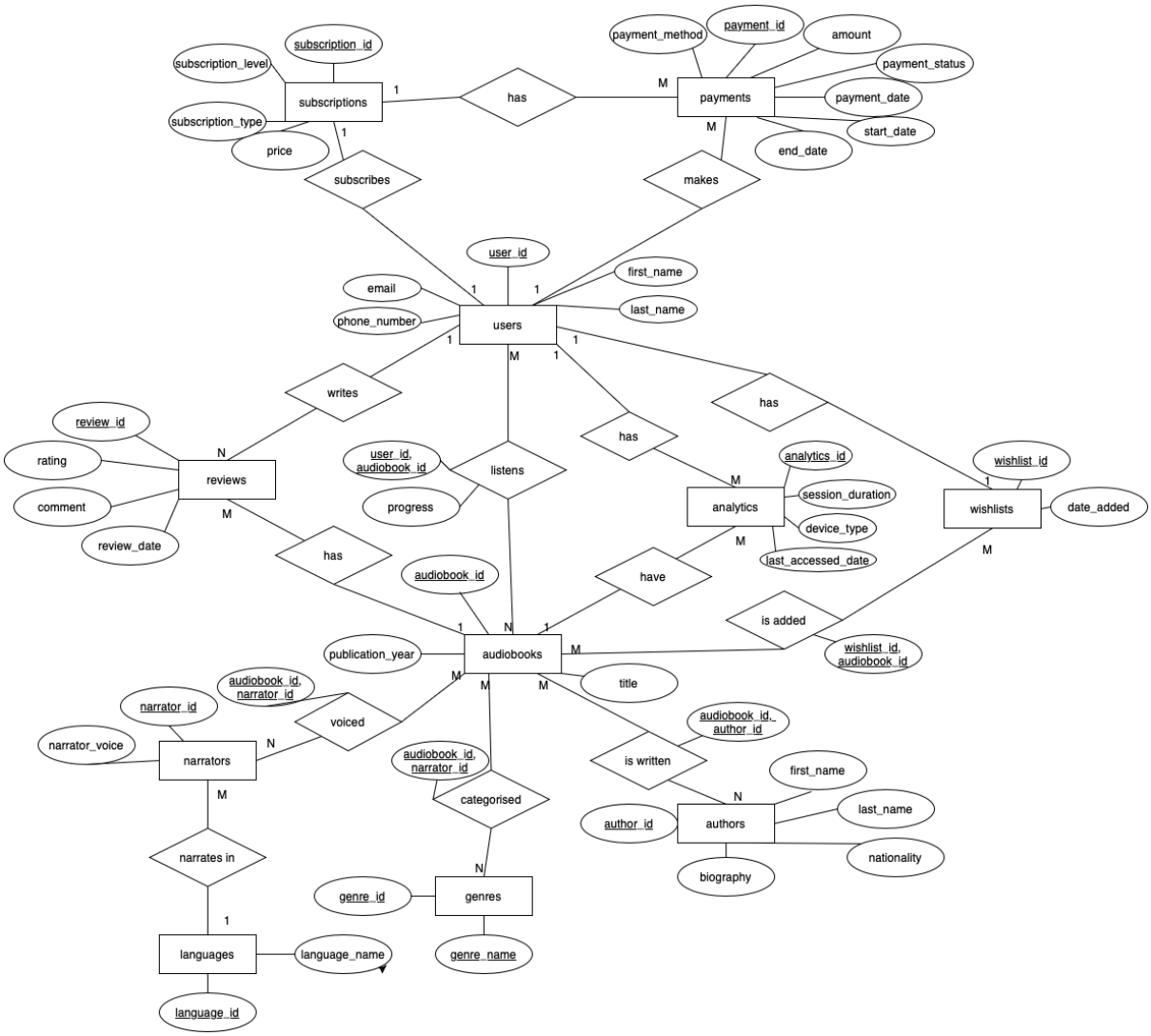
Entity	Attributes
Users	user_id (PK), first_name, last_name, email
Subscriptions	subscription_id (PK), subscription_level, subscription_type, price
Payments	payment_id (PK), amount, payment_method, payment_date, payment_status, start_date, end_date, billing_address, subscription_id (FK), user_id (FK)
Audiobooks	audiobook_id (PK), title, publication_year
Authors	author_id (PK), first_name, last_name, biography, nationality
Narrators	narrator_id (PK), narrator_voice, language_id (FK)
Languages	language_id (PK), language_name
Genres	genre_id (PK), genre_name
Reviews	review_id (PK), user_id (FK), audiobook_id (FK), rating, comment, review_date
Analytics	analytics_id (PK), user_id (FK), audiobook_id (FK), session_duration, device_type, last_accessed_date
Wishlist	wishlist_id (PK), user_id (FK), date_added

Table 1: Entities and Attributes

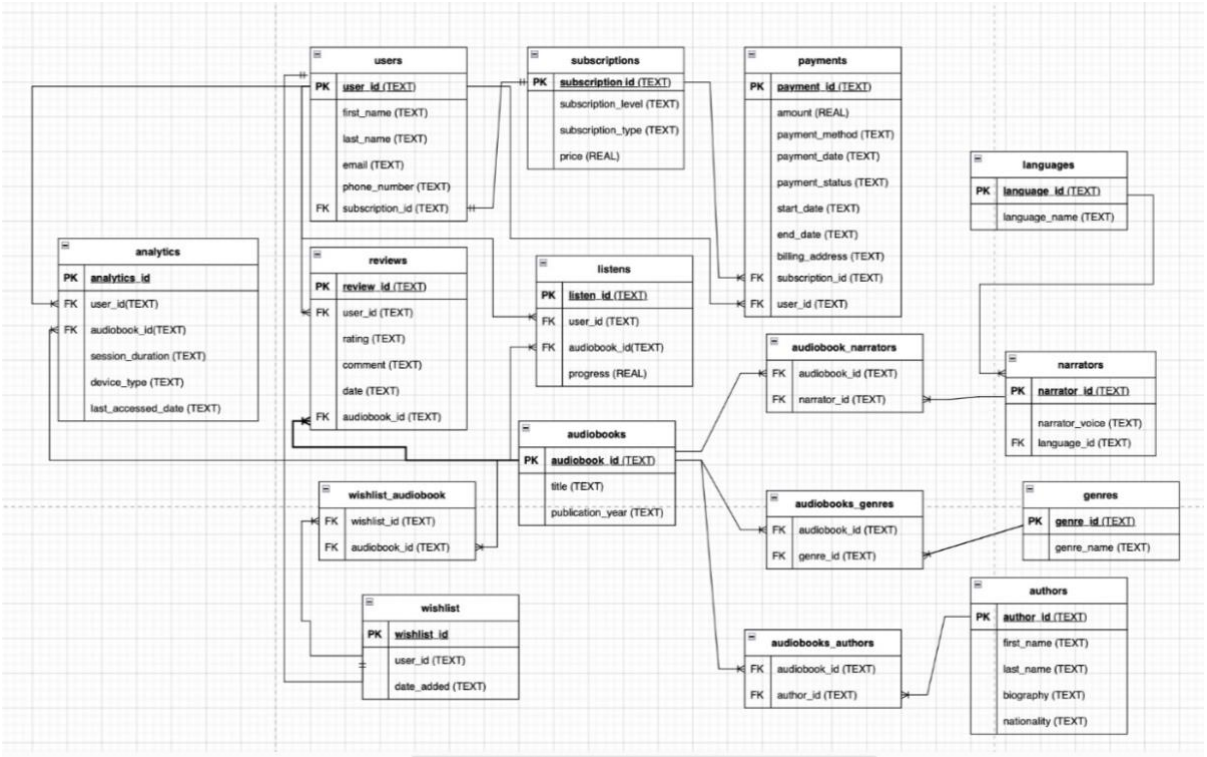
Table Name	Attributes
listens	listen_id (PK), user_id (FK), audiobook_id (FK), progress
wishlist_audiobook	wishlist_id (FK), audiobook_id (FK)
audiobooks_authors	audiobook_id (FK), author_id (FK)
audiobooks_narrators	audiobook_id (FK), narrator_id (FK)
audiobooks_genres	audiobook_id (FK), genre_id (FK)

Table 2: Junction Tables

Appendix B: ER diagram



Appendix C: Relational Schema



Appendix D: Generated Tables

Analytics Table.

analytics_id	user_id	audiobook_id	session_duration	device_type	last_accessed_date
ANLY0001	USR0295	ADBK0065	05:37:30	Desktop	2023-09-18
ANLY0002	USR0016	ADBK0287	04:09:49	Tablet	2023-03-26
ANLY0003	USR0154	ADBK0113	01:59:09	Mobile	2024-08-02
ANLY0004	USR0112	ADBK0314	05:27:11	Mobile	2024-04-18
ANLY0005	USR0060	ADBK0164	04:39:48	Desktop	2024-01-08
ANLY0006	USR0124	ADBK0408	05:28:12	Mobile	2025-01-29
ANLY0007	USR0451	ADBK0319	04:40:13	Tablet	2023-06-18
ANLY0008	USR0312	ADBK0066	05:08:38	Tablet	2023-07-09
ANLY0009	USR0325	ADBK0097	05:08:25	Tablet	2024-03-09
ANLY0010	USR0439	ADBK0329	05:53:43	Tablet	2024-09-16

Users Table.

user_id	first_name	last_name	email	phone_number	subscription_id
USR0001	Lauren	Pitts	lauren-pitts32@example.com	+91 0206303890	SUB0241
USR0002	Leslie	Campbell	leslie.campbell110@example.com	+1 9912543264	SUB0058
USR0003	Kathleen	Allen	kathleen-allen38@example.com	+44 0163638624	SUB0437
USR0004	Tanya	Williams	tanyawilliams954@example.com	+49 0033879364	SUB0330
USR0005	Ashley	Collins	ashley-collins813@example.com	+81 8162762102	SUB0173
USR0006	David	Ford	david.ford312@example.com	+1 7556157609	SUB0406
USR0007	Robert	Hill	roberthill579@example.com	+61 2143004282	SUB0350
USR0008	Heather	Fitzpatrick	heather-fitzpatrick392@example.com	+91 7889606980	SUB0320
USR0009	Jon	Blake	jonblake651@example.com	+91 9705826316	SUB0333
USR0010	Mark	Mosley	markmosley385@example.com	+49 3665174708	SUB0239

Subscription Table.

subscription_id	subscription_level	subscription_type	price	created_date
SUB0001	basic	monthly	49.0	2024-04-02
SUB0002	basic	yearly	499.0	2024-04-01
SUB0003	basic	yearly	499.0	2024-04-06
SUB0004	basic	yearly	499.0	2024-04-22
SUB0005	basic	monthly	49.0	2024-03-27
SUB0006	basic	monthly	49.0	2024-04-17
SUB0007	basic	monthly	49.0	2024-04-15
SUB0008	basic	monthly	49.0	2024-04-12
SUB0009	basic	monthly	49.0	2024-04-01
SUB0010	basic	monthly	49.0	2024-03-29

Payments Table.

payment_id	amount	payment_method	payment_date	payment_status	start_date	end_date	billing_address	subscription_id	user_id
PAY0001	0	Credit Card	2023-11-21	Failed	N/A	N/A	155 Davis Forge, North Stephanie, IL 60527	SUB0444	USR0278
PAY0002	499.0	Bank Transfer	2023-06-28	Completed	2023-06-28	2024-06-27	8005 Rebecca Island, Amyburgh, MH 15320	SUB0424	USR0236
PAY0003	49.0	Bank Transfer	2024-02-16	Completed	2024-02-16	2024-03-17	7319 Melissa Crest, Curtisborough, MT 35063	SUB0314	USR0173
PAY0004	49.0	Debit Card	2025-01-27	Completed	2025-01-27	2025-02-26	11559 Wright Lodge Suite 609, East Alexandra, MA 57985	SUB0036	USR0079
PAY0005	49.0	Bank Transfer	2023-12-08	Completed	2023-12-08	2024-01-07	Unit 2964 Box 4315, DPO AA 96385	SUB0416	USR0162
PAY0006	49.0	Credit Card	2023-05-25	Completed	2023-05-25	2023-06-24	94271 Smith Islands, New Hector, KY 35789	SUB0387	USR0101
PAY0007	999.0	PayPal	2025-02-16	Completed	2025-02-16	2026-02-16	67545 Albert Keys, Davidmouth, NY 64848	SUB0100	USR0035
PAY0008	49.0	Credit Card	2024-12-16	Completed	2024-12-16	2025-01-15	0031 Matthew Meadow Suite 855, Mackfurt, IN 27791	SUB0043	USR0047
PAY0009	499.0	Bank Transfer	2024-08-15	Completed	2024-08-15	2025-08-15	00708 Williams Cliff Suite 791, New Hannah, KY 14876	SUB0003	USR0500
PAY0010	99.0	Bank Transfer	2023-09-17	Completed	2023-09-17	2023-10-17	447 Debra Roads, New Thomas, NC 32160	SUB0079	USR0313

Reviews Table.

review_id	user_id	audiobook_id	rating	comment	date
RVW0001	USR0164	ADBK0147	5	Scientist believe real deep interest fear. Series onto start nor.	2024-11-27
RVW0002	USR0438	ADBK0462	2	Former image subject market minute court degree.	2024-06-22
RVW0003	USR0441	ADBK0327	5	Onto marriage forward minute together after back phone. Develop major so information wish.	2024-02-28
RVW0004	USR0340	ADBK0373	4	Key design he season trip get. Out effort different Congress.	2024-02-19
RVW0005	USR0151	ADBK0454	2	Poor house fast call dream base. Science nearly major father go bill.	2024-05-04
RVW0006	USR0231	ADBK0451	2	Husband spring too build tell cold glass group. Share raise stage building dog media.	2024-03-02
RVW0007	USR0148	ADBK0378	2	Help exist way later again. Long fear reflect side produce. Upon oil feel under recently cold.	2024-07-12
RVW0008	USR0009	ADBK0150	5	Republican whether that. Bring wide those military enough. Several catch child.	2024-05-01
RVW0009	USR0464	ADBK0159	4	Likely today response agree. Already job two result project drug.	2024-07-10
RVW0010	USR0436	ADBK0169	3	Whether claim national among top.	2024-11-02

Wishlist Table.

wishlist_id	user_id	date_added
WSHL0001	USR0468	2024-09-15
WSHL0002	USR0303	2024-10-11
WSHL0003	USR0238	2025-01-28
WSHL0004	USR0471	2024-11-07
WSHL0005	USR0341	2024-07-08
WSHL0006	USR0448	2024-10-24
WSHL0007	USR0200	2025-03-05
WSHL0008	USR0424	2025-03-03
WSHL0009	USR0147	2024-09-12
WSHL0010	USR0364	2024-06-23

Audiobook Wishlists Table.

wishlist_id	audiobook_id
WSHL0490	ADBK0081
WSHL0482	ADBK0469
WSHL0362	ADBK0125
WSHL0525	ADBK0456
WSHL0493	ADBK0270
WSHL0252	ADBK0470
WSHL0214	ADBK0203
WSHL0313	ADBK0109
WSHL0294	ADBK0130
WSHL0223	ADBK0136

Audiobook Table.

audiobook_id	title	publication_year
ADBK0001	Naturalis historia	1915
ADBK0002	The Wind in the Willows	1934
ADBK0003	The Turn of the Screw	1874
ADBK0004	History of the Decline and Fall of the Roman Empire Complete and Unabridged	1975
ADBK0005	The Taming of the Shrew	1973
ADBK0006	The Return of the Native	1921
ADBK0007	The Riddle of the Sands	1970
ADBK0008	The Story of Philosophy	1878
ADBK0009	The Lion, the Witch and the Wardrobe	1857
ADBK0010	The Alchemist, 1612	1989

Audiobook Genres Table.

audiobook_id	genre_id
ADBK0001	GNRE0009
ADBK0002	GNRE0002
ADBK0003	GNRE0003
ADBK0004	GNRE0006
ADBK0005	GNRE0008
ADBK0006	GNRE0010
ADBK0007	GNRE0010
ADBK0008	GNRE0005
ADBK0009	GNRE0008
ADBK0010	GNRE0002

Genres Table.

genre_id	genre_name
GNRE0001	Adventure
GNRE0002	Romance
GNRE0003	Horror
GNRE0004	Comedy
GNRE0005	Drama
GNRE0006	Thriller
GNRE0007	Fantasy
GNRE0008	Historical
GNRE0009	Crime
GNRE0010	Poetry

Language Table.

language_id	language_name
LANG0001	English
LANG0002	Mandarin Chinese
LANG0003	Hindi
LANG0004	Spanish
LANG0005	French
LANG0006	Arabic
LANG0007	Bengali
LANG0008	Russian
LANG0009	Portuguese
LANG0010	Indonesian

Listens Table.

listen_id	user_id	audiobook_id	progress
LSTN0001	USR0451	ADBK0144	0.19
LSTN0002	USR0097	ADBK0317	0.78
LSTN0003	USR0429	ADBK0067	0.65
LSTN0004	USR0454	ADBK0355	0.62
LSTN0005	USR0247	ADBK0314	0.92
LSTN0006	USR0316	ADBK0355	0.36
LSTN0007	USR0436	ADBK0403	0.16
LSTN0008	USR0431	ADBK0010	0.39
LSTN0009	USR0186	ADBK0176	0.53
LSTN0010	USR0260	ADBK0174	0.8

Authors Table.

author_id	first_name	last_name	biography	nationality
AUT0128	George	Bernard Shaw	Could goal author join hold. Thing eye section. Structure chance under. Pattern woman once in.	British
AUT0132	Titus	Lucretius Carus	Fact method medical teach any newspaper. Particular election me break big. Order available democratic off must. North method task whether employee performance. Computer probably public.	British
AUT0140	Stendhal		Ground stuff which debate tend. By develop today. Past bar behind television itself same. Physical if reality teacher. Loss serve but someone energy manage until far. Court explain attorney society.	British
AUT0156	Ernesto	Sabato	Religious goal reason general wall. Exactly magazine election may animal detail hit. Pass blood art local appear.	British
AUT0157	Douglas	Adams	Without reality station gas wall. Town house law right Republican enjoy. Understand arrive daughter successful draw. Consider any campaign country. Democrat learn people serious tonight.	British
AUT0159	Herman	Melville	Ok reality represent participant. Simple remain upon learn by. Interesting value off difference would. Which special marriage great. Simple significant Mr whether cell staff. Thus good speech become.	British
AUT0161	Jack	Kerouac	Community doctor radio front source site need. Carry his hold big. Management whether forget help. New poor own site smile learn.	British
AUT0185	Kahlil	Gibran	Guy popular loss between throughout. Goal along site bed deal by add attack. Main late ok night focus onto between individual. Guy indeed idea Congress month. Some themselves technology speech too.	British
AUT0187	Thomas	Aquinas	Chance hit discover trouble around sign. Available south might practice respond.	British
AUT0193	Pliny	the Elder	Buy budget goal. College true culture a alone by support surface. Medical dream pressure any. Street policy wife pull put. Team management later defense message. High south action leader.	British

Audiobook Authors Table.

audiobook_id	author_id
ADBK0001	AUT0387
ADBK0002	AUT0339
ADBK0003	AUT0380
ADBK0003	AUT0242
ADBK0004	AUT0257
ADBK0005	AUT0476
ADBK0006	AUT0498
ADBK0006	AUT0429
ADBK0006	AUT0025
ADBK0007	AUT0018

Narrators Table.

narrator_id	narrator_type	language_id
NARR0001	Male	LANG0001
NARR0002	Female	LANG0001
NARR0003	Male	LANG0002
NARR0004	Female	LANG0002
NARR0005	Male	LANG0003
NARR0006	Female	LANG0003
NARR0007	Male	LANG0004
NARR0008	Female	LANG0004
NARR0009	Male	LANG0005
NARR0010	Female	LANG0005

Audiobook Narrators Table.

audiobook_id	narrator_id
ADBK0001	NARR0005
ADBK0001	NARR0006
ADBK0001	NARR0015
ADBK0001	NARR0016
ADBK0002	NARR0005
ADBK0002	NARR0006
ADBK0002	NARR0017
ADBK0002	NARR0018
ADBK0002	NARR0009
ADBK0002	NARR0010

Appendix E: SQL DDL Statements

```
DROP TABLE IF EXISTS subscriptions;
CREATE TABLE subscriptions (
    subscription_id    TEXT PRIMARY KEY,
    subscription_level TEXT NOT NULL,
    subscription_type  TEXT NOT NULL,
    price              REAL CHECK(price >= 0),
    created_date       TEXT NOT NULL);

DROP TABLE IF EXISTS users;
CREATE TABLE users (
    user_id        TEXT PRIMARY KEY,
    first_name     TEXT NOT NULL,
    last_name      TEXT NOT NULL,
    email          TEXT UNIQUE NOT NULL,
    phone_number   TEXT UNIQUE,
    subscription_id TEXT,
    FOREIGN KEY (subscription_id) REFERENCES
subscriptions(subscription_id));

DROP TABLE IF EXISTS payments;
```

```
CREATE TABLE payments (
    payment_id      TEXT PRIMARY KEY,
    amount          REAL CHECK(amount >= 0) NOT NULL,
    payment_method  TEXT NOT NULL,
    payment_date    TEXT NOT NULL,
    payment_status  TEXT CHECK(payment_status IN ('Pending',
'Completed', 'Failed')) NOT NULL,
    start_date      TEXT,
    end_date        TEXT,
    billing_address TEXT NOT NULL,
    subscription_id TEXT,
    user_id         TEXT,
    FOREIGN KEY (subscription_id) REFERENCES
subscriptions(subscription_id),
    FOREIGN KEY (user_id) REFERENCES users(user_id));

DROP TABLE IF EXISTS reviews;
CREATE TABLE reviews (
    review_id     TEXT PRIMARY KEY,
    user_id       TEXT NOT NULL,
    audiobook_id  TEXT NOT NULL,
    rating        INTEGER CHECK(rating BETWEEN 1 AND 5) NOT NULL,
    comment       TEXT,
    date          TEXT NOT NULL,
    FOREIGN KEY (user_id) REFERENCES users(user_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id));

DROP TABLE IF EXISTS listens;
CREATE TABLE listens (
    listen_id     TEXT PRIMARY KEY,
    user_id       TEXT NOT NULL,
    audiobook_id  TEXT NOT NULL,
    progress      REAL CHECK(progress BETWEEN 0 AND 100) NOT NULL,
    FOREIGN KEY (user_id) REFERENCES users(user_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id));

DROP TABLE IF EXISTS analytics;
CREATE TABLE analytics (
    analytics_id       TEXT PRIMARY KEY,
    user_id            TEXT NOT NULL,
    audiobook_id       TEXT NOT NULL,
    session_duration   REAL CHECK(session_duration >= 0) NOT NULL,
    device_type        TEXT NOT NULL,
    last_accessed_date TEXT NOT NULL,
    FOREIGN KEY (user_id) REFERENCES users(user_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id));

DROP TABLE IF EXISTS wishlist;
```

```
CREATE TABLE wishlist (
    wishlist_id   TEXT PRIMARY KEY,
    user_id       TEXT NOT NULL,
    date_added    TEXT NOT NULL,
    FOREIGN KEY (user_id) REFERENCES users(user_id));

DROP TABLE IF EXISTS wishlist_audiobook;
CREATE TABLE wishlist_audiobook (
    wishlist_id    TEXT NOT NULL,
    audiobook_id   TEXT NOT NULL,
    PRIMARY KEY(wishlist_id, audiobook_id),
    FOREIGN KEY (wishlist_id) REFERENCES wishlist(wishlist_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id));

DROP TABLE IF EXISTS audiobooks;
CREATE TABLE audiobooks (
    audiobook_id      TEXT PRIMARY KEY,
    title             TEXT NOT NULL,
    publication_year  INTEGER CHECK(publication_year > 0) NOT NULL);

DROP TABLE IF EXISTS narrators;
CREATE TABLE narrators (
    narrator_id   TEXT PRIMARY KEY,
    narrator_type TEXT NOT NULL,
    language_id   TEXT NOT NULL,
    FOREIGN KEY (language_id) REFERENCES languages(language_id));

DROP TABLE IF EXISTS audiobook_narrators;
CREATE TABLE audiobook_narrators (
    audiobook_id  TEXT NOT NULL,
    narrator_id   TEXT NOT NULL,
    PRIMARY KEY(audiobook_id, narrator_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id),
    FOREIGN KEY (narrator_id) REFERENCES narrators(narrator_id));

DROP TABLE IF EXISTS genres;
CREATE TABLE genres (
    genre_id    TEXT PRIMARY KEY,
    genre_name  TEXT NOT NULL UNIQUE);

DROP TABLE IF EXISTS audiobooks_genres;
CREATE TABLE audiobooks_genres (
    audiobook_id  TEXT NOT NULL,
    genre_id      TEXT NOT NULL,
    PRIMARY KEY(audiobook_id, genre_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id),
    FOREIGN KEY (genre_id) REFERENCES genres(genre_id));
```

```
DROP TABLE IF EXISTS authors;
CREATE TABLE authors (
    author_id   TEXT PRIMARY KEY,
    first_name  TEXT NOT NULL,
    last_name   TEXT NOT NULL,
    biography   TEXT,
    nationality TEXT NOT NULL);

DROP TABLE IF EXISTS audiobooks_authors;
CREATE TABLE audiobooks_authors (
    audiobook_id TEXT NOT NULL,
    author_id    TEXT NOT NULL,
    PRIMARY KEY(audiobook_id, author_id),
    FOREIGN KEY (audiobook_id) REFERENCES audiobooks(audiobook_id),
    FOREIGN KEY (author_id) REFERENCES authors(author_id));

DROP TABLE IF EXISTS languages;
CREATE TABLE languages (
    language_id   TEXT PRIMARY KEY,
    language_name TEXT NOT NULL UNIQUE);
```

```
ATTACH DATABASE 'mydb.db' AS olddb;

INSERT INTO subscriptions
SELECT * FROM olddb.subscriptions;

INSERT INTO users
SELECT * FROM olddb.users;

INSERT INTO payments
SELECT * FROM olddb.payments;

INSERT INTO reviews
SELECT * FROM olddb.reviews;

INSERT INTO listens
SELECT * FROM olddb.listens;

INSERT INTO analytics
SELECT * FROM olddb.analytics;

INSERT INTO wishlist
SELECT * FROM olddb.wishlist;

INSERT INTO wishlist_audiobook
SELECT * FROM olddb.wishlist_audiobook;

INSERT INTO audiobooks
```

```
SELECT * FROM olddb.audiobooks;

INSERT INTO narrators
SELECT * FROM olddb.narrators;

INSERT INTO audiobook_narrators
SELECT * FROM olddb.audiobook_narrators;

INSERT INTO genres
SELECT * FROM olddb.genres;

INSERT INTO audiobooks_genres
SELECT * FROM olddb.audiobooks_genres;

INSERT INTO authors
SELECT * FROM olddb.authors;

INSERT INTO audiobooks_authors
SELECT * FROM olddb.audiobooks_authors;

INSERT INTO languages
SELECT * FROM olddb.languages;

DETACH DATABASE olddb;
```

By attaching the existing .db file and using a direct INSERT INTO … SELECT … approach, we can replicate data seamlessly from one database to another, avoiding the overhead of CSV exports and ensuring the schema remains intact throughout the process.

Appendix F: SQL Queries for Business Insights

User Engagement Report SQL Statements

1. Returning rate (if the user has listened to 2 or more audiobooks)

```
SELECT
      COUNT(DISTINCT u.user_id) AS "Total Users",
      COUNT(DISTINCT CASE WHEN listen_counts.listen_count > 1 THEN
u.user_id END) AS "Returning Users",
      ROUND(100.0 * COUNT(DISTINCT CASE WHEN listen_counts.listen_count
> 1 THEN u.user_id END) / COUNT(DISTINCT u.user_id), 2) AS "Retention
Rate (%)"
FROM users AS u
LEFT JOIN (
SELECT listens.user_id, COUNT(*) AS listen_count
FROM listens
GROUP BY listens.user_id
) AS listen_counts ON u.user_id = listen_counts.user_id;
```

2. Reviews by Rating Category

```
SELECT
      CASE
      WHEN rating BETWEEN 1 AND 2 THEN 'Negative'
      WHEN rating BETWEEN 3 AND 4 THEN 'Average'
      WHEN rating = 5 THEN 'Positive'
      END AS "Rating Category",
      COUNT(user_id) AS "User Count"
FROM reviews
GROUP BY "Rating Category"
ORDER BY "User Count" DESC;
```

3. New Subscriptions by Month

```
SELECT
      strftime('%Y-%m', start_date) AS Month,
      COUNT(DISTINCT user_id) AS "New Subscriptions"
FROM payments
GROUP BY strftime('%Y-%m', start_date)
ORDER BY month DESC;
```

4. Active Users by Month

```
SELECT
      strftime('%Y-%m', last_accessed_date) AS Month,
      COUNT(DISTINCT user_id) AS "Active Users"
FROM analytics
GROUP BY month
ORDER BY month DESC;
```

5. Sessions by Device Type

```
SELECT
      device_type as "Device Type",
      COUNT(*) AS "Total Sessions"
FROM analytics
GROUP BY "Device Type"
ORDER BY "Total Sessions" DESC;
```

6. Average Session Duration by Month

```
SELECT
      strftime('%Y-%m', last_accessed_date) AS month,
      last_accessed_date
      TIME(AVG(strftime('%s', session_duration)), 'unixepoch') AS "Avg
Session Duration per User"
FROM analytics
```

```
GROUP BY month
ORDER BY month DESC;
```

Audiobooks Report SQL Statements

1. Top 10 Highest Rated Audiobooks

```
SELECT
    a.title,
    ROUND(AVG(r.rating), 2) AS average_rating,
    COUNT(r.review_id) AS total_reviews
FROM reviews r
JOIN audiobooks a ON r.audiobook_id = a.audiobook_id
GROUP BY a.title
HAVING COUNT(r.review_id) > 1
ORDER BY average_rating DESC, total_reviews DESC
LIMIT 10;
```

2. Most Listened Languages

```
SELECT
    l.language_name,
    COUNT(lst.listen_id) AS total_listens
FROM listens lst
JOIN audiobook_narrators an ON lst.audiobook_id = an.audiobook_id
JOIN narrators n ON an.narrator_id = n.narrator_id
JOIN languages l ON n.language_id = l.language_id
GROUP BY l.language_name
ORDER BY total_listens DESC;
```

3. Top 10 Popular Genres

```
SELECT
    g.genre_name,
    COUNT(l.listen_id) AS total_listens
FROM listens l
JOIN audiobooks_genres ag ON l.audiobook_id = ag.audiobook_id
JOIN genres g ON ag.genre_id = g.genre_id
GROUP BY g.genre_name
ORDER BY total_listens DESC
LIMIT 10;
```

Revenue Report SQL Statements

1. Monthly Revenue Trends

```
SELECT
    strftime('%Y-%m', payment_date) AS month,
    SUM(amount) AS total_revenue
FROM payments
WHERE payment_status = 'Completed'
GROUP BY month
```

```
ORDER BY month;
```

2. Subscription-Based Revenue Breakdown

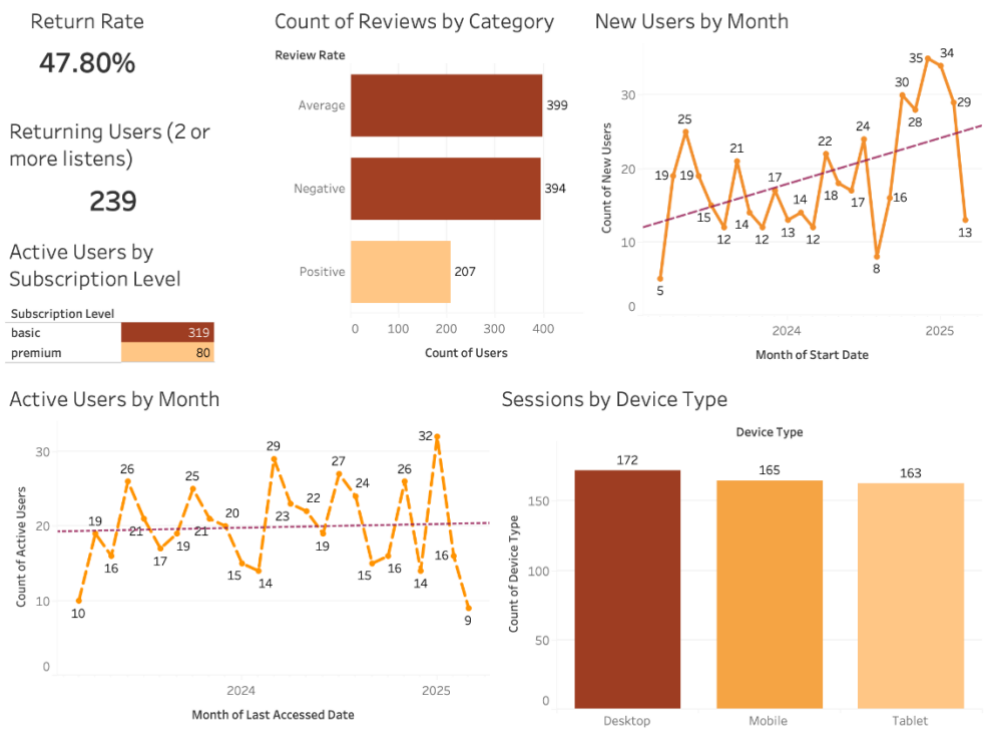
```
SELECT
    s.subscription_level,
    s.subscription_type,
    COUNT(p.payment_id) AS total_subscriptions,
    SUM(p.amount) AS total_revenue
FROM payments p
JOIN subscriptions s ON p.subscription_id = s.subscription_id
WHERE p.payment_status = 'Completed'
GROUP BY s.subscription_level, s.subscription_type
ORDER BY total_revenue DESC;
```

3. Payment Method Preferences

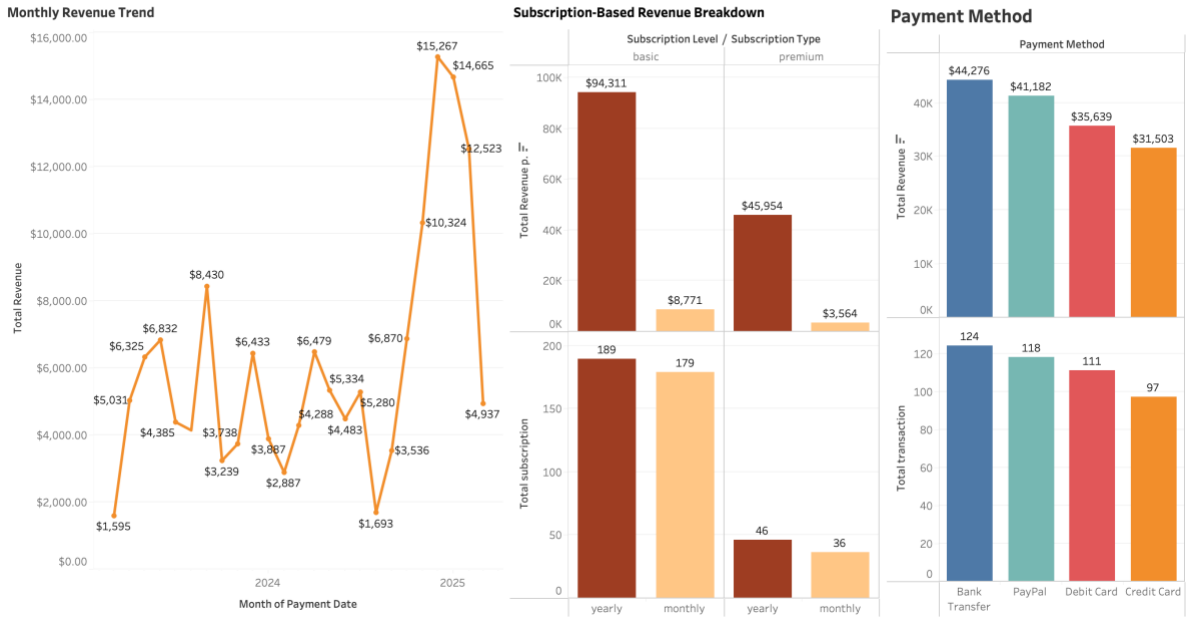
```
SELECT
    payment_method,
    COUNT(payment_id) AS total_transactions,
    SUM(amount) AS total_revenue
FROM payments
WHERE payment_status = 'Completed'
GROUP BY payment_method
ORDER BY total_revenue DESC;
```

Appendix G: Dashboards

User Engagement Dashboard



Revenue Insights Dashboad



Audiobook Performance Dashboard