

各 位

2024年7月31日
株式会社インプレス

小規模データの分析で使える SQL の書き方がわかる！
『データ分析に強くなる SQL レシピ 小規模データの前処理・分析の書き方&テクニック』
を8月5日に発売！

インプレスグループでIT関連メディア事業を展開する株式会社インプレス（本社：東京都千代田区、代表取締役社長：高橋隆志）は、SQLで数万件程度までの小規模データを扱うエンジニアやビジネスパーソン向けの書籍『データ分析に強くなるSQLレシピ 小規模データの前処理・分析の書き方&テクニック』を2024年8月5日（月）に発売いたします。



■データ分析で役立つSQLの書き方がわかる！基礎から応用テクニックまでを標準SQLサンプルコードとともに徹底解説！

本書では、中小企業でよくある、数百件から数万件程度の小規模なデータを想定し、それらデータの前処理や分析を行なうSQLの書き方やテクニックをレシピとしてまとめています。データの前処理を中心に、合計や平均、最小値・最大値の計算、小数処理、グループ集計、重複除外、日付・文字列処理、条件分岐などの基本的なSQL操作から、複数テーブルの結合やサブクエリ、外れ値・欠損値処理、データの分布調査、RFM分析やアソシエーション分析などの応用的な技術まで、実用的なサンプルコードとともに徹底解説しています。

「AIやビッグデータが注目されますが、手元にあるちょっとしたデータをSQLで分析してみませんか？」
(著者Xより引用：<https://twitter.com/masuiepe/status/1809117862772232385>)

■本書は以下のような方におすすめです

- ・ 数万件程度までの小規模データを使ってSQLでデータ分析をしたい方
- ・ SQLによる小規模データの前処理や分析の基本やテクニックを学びたい人方
- ・ エンジニア、データサイエンティスト、マーケター、ビジネスパーソン（日常業務でちょっとしたデータ分析をする方）

■紙面イメージ

1-1 データベースとは

企業などの組織において複数人でデータを管理するとき、データベースという仕組みがよく使われます。Excelのような表計算ソフトなどを使っても表形式のデータを扱えますが、なぜデータベースを使う必要があるのか、その特徴と求められる機能について考えてみましょう。

1-1-1 ファイルの特徴を知る

私たちがコンピュータを使うとき、一般的にはデータをファイルに保存します。Wordのような文書作成ソフトで文書を作成すれば、そのデータは文書ファイルとして保存されますし、インターネットを使っていて何か気になるデータがあれば、それをダウンロードしてファイルに保存しておくこともあるでしょう。スマートフォンで写真を撮影すれば、それもファイルとして保存されていますし、スマートフォンのアプリなどもすべてファイルです。

このように、私たちはコンピュータやスマートフォンを使うとき、ファイル単位でさまざまなデータを管理しています。このとき、1つのファイルを1つの単位として扱います。音楽であれば1曲の音楽が1つのファイル、写真であれば1枚の写真が1つのファイル、アプリであれば1つのアプリが1つのファイルです(図1-1)。

●図1-1: ファイルへの保存



これはわかりやすい一方で、複数のデータをまとめて扱うほうが便利な場面があります。たとえば、住所録(アドレス帳)を考えてみましょう。

連絡先の情報を1件ずつそれぞれのファイルとして保存することもできますが、登録した連絡先の一覧を表示したい場合には、それらすべてのファイルを1つずつ開かなければいけません。同様に、名前や住所の一部で検索したいときも、すべてのファイルを開いて、検索条件に当てはまるかを1件ずつ調べなければなりません。

これでは不便なので、住所録のようなデータを作成するときは、表計算ソフトがよく使われます。表計算ソフトではそれぞれのデータを行単位に格納し、列単位で名前や住所、電話番号などを管理します(図1-2)。

●図1-2: 表計算ソフトでのデータの管理

	A	B	C	D
1	名前	住所	電話番号	
2	山田太郎	東京都新宿区	090-1111-1111	
3	鈴木花子	東京都港区	090-2222-2222	
4				

そして、これを1つのファイルとして保存します。これにより、ファイルを1つ開くだけで、保存している連絡先を一覧としてすべて表示することもでき、何らかの条件を指定したデータを容易に検索できます。表計算ソフトを使うと項目を自由に追加、編集することもでき、便利に使えます。

一方で、誤った項目を編集しても問題なく更新できてしまいます。郵便番号の項目に住所を入れたり、電話番号の項目に名前を入れたりしても登録できてしまうのです。名前を入力することなく連絡先を保存してしまうと、住所や電話番号が記録されていてもそれが誰の連絡先なのか分からなくなります。また、項目を自由にいくつも追加できるため、目的のデータがどこに保存されているのかわからなくなることもあるでしょう。

表計算ソフトを使って複数のデータを1つのファイルに保管できることは便利ですが、特定のデータを更新しただけなのに、すべてのデータを開かなければいけないという問題もあります。複数人でこのファイルを使おうとすると、ある人が更新している最中は他の人が更新できない、もしくは誰かが更新した内容が知らないうちに他の人の更新によって失われる、といった事態が発生します。

前半章では、SQLやデータベースの基本、オーソドックスな関数について学びます

リスト 5-1 度数分布表の作成

```
SELECT
CASE
WHEN english >= 0 AND english < 20 THEN '0-20'
WHEN english >= 20 AND english < 40 THEN '20-40'
WHEN english >= 40 AND english < 60 THEN '40-60'
WHEN english >= 60 AND english < 80 THEN '60-80'
WHEN english >= 80 AND english <= 100 THEN '80-100'
ELSE 'Other'
END AS スコア,
COUNT(*) AS 度数
FROM scores
GROUP BY スコア
ORDER BY スコア;
```

実行結果

スコア	度数
40-60	1
60-80	2
80-100	1

これは、CASEによって英語の点数によって分岐するとともに、その値に「スコア」という名前をつけ、GROUP BYやORDER BYに使用しています。

5-1-2 累積度数と相対度数を計算する

度数分布表を作成したときに、各階級の度数を見るだけでなく、その階級までの度数の合計を知りたいこともあります。このような、ある階級までの度数の合計を累積度数といいます。

たとえば、表5-1の度数分布表に累積度数を追加すると、表5-4のようになります。

●表5-4: 累積度数を追加した度数分布表

身長	度数	累積度数
130cm以上140cm未満	6人	6人
140cm以上150cm未満	8人	14人
150cm以上160cm未満	12人	26人
160cm以上170cm未満	5人	31人
170cm以上180cm未満	1人	32人
合計	32人	-

この表を作成しておくことで、160cm未満の生徒の人数を知りたい人は、度数を順に足し算しなくても、累積度数を見るだけで26人であることがわかります。このような累積度数を計算するとき、WITHを使ったサブクエリと分析関数を使うと便利です。

たとえば、リスト5-2のようなSQLが考えられます。

リスト 5-2 度数分布表と累積度数の作成

```
WITH temp AS (
SELECT
CASE
WHEN english >= 0 AND english < 20 THEN '0-20'
WHEN english >= 20 AND english < 40 THEN '20-40'
WHEN english >= 40 AND english < 60 THEN '40-60'
WHEN english >= 60 AND english < 80 THEN '60-80'
WHEN english >= 80 AND english <= 100 THEN '80-100'
ELSE 'Other'
END AS score_range,
COUNT(*) AS frequency
FROM scores
GROUP BY score_range
ORDER BY score_range
)
SELECT
score_range AS スコア,
frequency AS 度数,
SUM(frequency) OVER (ORDER BY score_range) AS 累積度数
FROM temp;
```

後半にかけてテーブルの結合やサブクエリの使い方を学習し、統計データの作成にも触れていきます

■本書の構成

- 第1章 SQLの基本
- 第2章 基本的な関数を知る
- 第3章 複数のテーブルを結合する
- 第4章 サブクエリの使用
- 第5章 統計データを作成する
- 第6章 複雑な問題を解く
- 付録（データベースの作成など）
- 練習問題の解答

■書誌情報



書名：データ分析に強くなるSQLレシピ 小規模データの前処理・分析の書き方&テクニック

著者：増井 敏克 著

発売日：2024年8月5日（金）

ページ数：312ページ

サイズ：A5判

定価：2,970円（本体2,700円＋税10%）

電子版価格：2,970円（本体2,700円＋税10%）※インプレス直販価格

ISBN：9784295019848

◇Amazonの書籍情報ページ：<https://www.amazon.co.jp/dp/4295019844>

◇インプレスの書籍情報ページ：

<https://book.impress.co.jp/books/1123101037>

■著者プロフィール

増井 敏克（ますいとしかつ）

増井技術士事務所代表

技術士（情報工学部門）

1979年奈良県生まれ。大阪府立大学大学院修了。

テクニカルエンジニア（ネットワーク、情報セキュリティ）、その他情報処理技術者試験にも多数合格。また、ビジネス数学検定1級に合格し、公益財団法人日本数学検定協会認定トレーナーとしても活動。「ビジネス」×「数学」×「IT」を組み合わせ、コンピュータを「正しく」「効率よく」使うためのスキルアップ支援や、各種ソフトウェアの開発を行なっている。著作実績として、『実務で使える メール技術の教科書』

以上

【株式会社インプレス】 <https://www.impress.co.jp/>

シリーズ累計 8,000 万部突破のパソコン解説書「できる」シリーズ、「デジタルカメラマガジン」等の定期雑誌、IT 関連の専門メディアとして国内最大級のアクセスを誇るデジタル総合ニュースサービス「Impress Watch シリーズ」等のコンシューマ向けメディア、「IT Leaders」をはじめとする企業向け IT 関連メディアなどを総合的に展開・運営する事業会社です。IT 関連出版メディア事業、およびデジタルメディア&サービス事業を幅広く展開しています。

【インプレスグループ】 <https://www.impressholdings.com/>

株式会社インプレスホールディングス（本社：東京都千代田区、代表取締役：松本大輔、証券コード：東証スタンダード市場 9479）を持株会社とするメディアグループ。「IT」「音楽」「デザイン」「山岳・自然」「航空・鉄道」「モバイルサービス」「学術・理工学」を主要テーマに専門性の高いメディア&サービスおよびソ

リユース事業を展開しています。さらに、コンテンツビジネスのプラットフォーム開発・運営も手がけています。

【本件に関するお問合せ先】

株式会社インプレス 広報担当：丸山

E-mail: pr-info@impress.co.jp URL : <https://www.impress.co.jp/>

※弊社はテレワーク推奨中のため電話でのお問い合わせを停止しております。メールまたは Web サイトからお問い合わせください。