

Advanced SQL untuk Analisa Data

➤ Progate

- ❑ Materi-materi SQL di Progate membahas dasar-dasar SQL dan terdapat latihan (dojo) yang memberikan Anda latihan dengan query yang lebih menantang.
- ❑ Di latihan ini, kita akan mendalami lebih jauh apa yang sudah Anda pelajari di Progate dengan mencoba menjawab beberapa pertanyaan bisnis untuk keputusan perusahaan yang lebih baik berdasarkan data.
- ❑ Disini, Anda berperan sebagai seorang Marketing executive dan Anda mencoba untuk memberikan masukan kepada tim dan atasan Anda, keputusan apa yang harus diambil oleh perusahaan berdasarkan data yang Anda dapatkan!

Apa yang akan kita kerjakan di latihan ini

- ❑ Kita akan menganalisis data situs e-commerce di kursus latihan di Progate lebih dalam lagi.
- ❑ Kita akan belajar bagaimana melakukan analisis data selangkah demi selangkah dan mencoba mencari solusi untuk meningkatkan keuntungan situs e-commerce kita.
- ❑ Mari akses salah satu halaman dojo dan hapus kueri yang ada. Kita akan menggunakan editor dan database di Progate.

<https://progate.com/sql/dojo/1/10#/19>

Meninjau Ulang Struktur Tabel

- ❏ Ada 3 tabel di database kita: `items`, `sales_records` dan `users`.
- ❏ Anda dapat melihat hal berikut:
 - 30 item terdaftar di situs.
 - Ada 910 rekor penjualan pada Juli 2018.
 - 325 pengguna terdaftar di tabel pengguna.
- ❏ Ayo mulai!

Permintaan Data no. 1:

Tim keuangan bertanya kepada Anda untuk memberikan angka total untuk pendapatan (revenue), laba, dan persentase laba website pembelian

Apa yang ingin kita dapatkan

- ❑ Ayo cari total pendapatan, laba serta persentase laba di website pembelian untuk mencari tau gambaran besar dari jumlah penjualan dan laba yang didapatkan website tersebut.

Latihan 1

Dapatkan total pendapatan, laba dan persentase laba untuk seluruh website pembelian.

- ❑ Total pendapatan adalah jumlah keseluruhan harga barang yang terjual
- ❑ Laba adalah total dari perbedaan antara harga jual dan harga beli setiap barang yang terjual.
- ❑ Persentase laba dapat dikalkulasikan dengan formula sebagai berikut:
$$\text{Total persentase laba} = \frac{\text{total laba}}{\text{total pendapatan}} * 100$$

One more point

- ❑ Dengan formula ini, Anda akan mendapatkan hasil, 0.

$$\text{Total persentase laba} = \text{total laba} / \text{total pendapatan} * 100$$

- ❑ Ini disebabkan karena tipe data pendapatan dan laba yang dimasukkan adalah integer. Anda harus mengubah integer ke nomor decimal untuk mendapatkan hasil yang diinginkan.

- ❑ Jika Anda mengkalikan nomor desimal, kalkulasi akan diubah ke desimal secara otomatis. Formula ini akan memberikan hasil yang diinginkan!!

$$\text{Persentase laba} = 1.0 * \text{total laba} / \text{total pendapatan} * 100$$

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

Jawaban

[Code]

```
exercise1.sql exercise2.sql exercise3.sql
1 SELECT SUM(items.price) AS total_pendapatan,
2       SUM(items.price - items.cost) AS total_laba,
3       1.0 * SUM(items.price - items.cost) / SUM(items.price) * 100 AS persentase_laba
4 FROM sales_records
5 JOIN items ON sales_records.item_id = items.id
6
```

[Hasil]

total_revenue	total_profit	profit_rate
46779	31428	67.1839928172898

- ❑ Kita telah mendapatkan gambaran besar untuk website belanjaan di mana total pendapatan: \$46,779, total laba: \$31,428 and persentase laba: 67.2%.
- ❑ Berdasarkan data yang kita miliki sebagai langkah lanjutan, kita dapat membandingkan data setiap item dengan seluruh item (*Setting the standard*). Kita dapat melihat item mana yang di atas standar, item mana yang berkontribusi ke pendapatan dan mana yang tidak, dan item mana yang memiliki persentase laba terbesar dan sebaliknya.
- ❑ Perlu diingat bahwa ketika Anda melakukan analisis data, mulailah dari gambaran besar, bukan dari data individu.

Permintaan Data no.2 :

Apakah Anda tahu 5 item yang berkontribusi paling banyak terhadap pendapatan?

Apa yang ingin kita dapatkan

- ❑ Karena kita mendapatkan data pendapatan, laba, dan persentase laba dari seluruh situs belanja, mari kita lanjutkan dengan menganalisis "pendapatan" terlebih dahulu.
- ❑ Dalam latihan ini, kita akan mendapatkan 5 item teratas yang berkontribusi pada pendapatan.

Latihan 2

- ❑ Dapatkan "nama item", "jumlah penjualan", "harga" dan "total pemasukkan" dari 5 item yang memberikan pendapatan terbesar.
- ❑ Susun item berdasarkan pendapatan dari yang terbesar ke terkecil.

- ❑ Kuantitas penjualan berarti jumlah data dalam tabel sales_record.
- ❑ Pendapatan total dapat dihitung dengan harga * kuantitas penjualan.
- ❑ Anda perlu menggunakan GROUP BY untuk menggabungkan pendapatan setiap item.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
exercise1.sql exercise2.sql exercise3.sql
1 SELECT items.name,
2     COUNT(*) AS jumlah_penjualan,
3     items.price,
4     COUNT(*) * items.price AS total_pendapatan
5 FROM sales_records
6 JOIN items ON sales_records.item_id = items.id
7 GROUP BY items.name, items.price
8 ORDER BY COUNT(*) * items.price DESC
9 limit 5
```

[Result]

name	jumlah_penjualan	price	total_pendapatan
jaket kulit	27	320	8640
mantel merah	25	270	6750
celana denim	61	60	3660
sepatu lari	35	75	2625
rok	52	40	2080

Analisis latihan

- ❑ Sekarang kita mendapat 5 item yang menghasilkan pendapatan tertinggi dan mengapa itu menghasilkan pendapatan tinggi.
- ❑ Dalam data ini, item teratas adalah "jaket kulit". Kita bisa lihat kalau jumlah penjualannya tidak terlalu besar tapi harga satuannya sangat tinggi, yang membuatnya menjadi item teratas.
- ❑ Sebaliknya, "celana denim" dan "rok" memiliki harga yang lebih rendah tetapi menghasilkan keuntungan besar karena jumlah penjualan yang besar.
- ❑ Dari Permintaan data no. 1 dan 2, Anda juga dapat melihat berapa jumlah 5 item teratas untuk total pendapatan situs ini, yaitu $(\text{Jumlah dari total_revenue dari 5 teratas}) / 46779 = 50,8\%$.

Permintaan data no. 3:

Cari tahu persentase laba dan laba dari 5 item penghasil pendapatan terbatas

Apa yang ingin kita dapatkan

- ❑ Sekarang kita mengetahui 5 item penghasil pendapatan teratas, tetapi itu tidak berarti mereka juga menghasilkan keuntungan karena mungkin merupakan item yang sangat mahal.
- ❑ Dalam latihan ini, kita akan mendapatkan keuntungan dan tingkat keuntungan 5 item teratas.

Latihan 3

- ❑ Dapatkan "nama item", "keuntungan setiap item", "tingkat keuntungan setiap item" dan "keuntungan total setiap item" dari 5 item penghasil pendapatan teratas.
- ❑ Urutkan hasil dari item yang memiliki "keuntungan total" tertinggi ke yang terendah.

- ❑ Gunakan **WHERE** untuk hanya mendapatkan data dari 5 item penghasil pendapatan teratas. Dan sebaiknya gunakan **id** setiap item untuk kondisi **WHERE** karena mungkin ada nama item yang sama tetapi id unik.
(Anda perlu mendapatkan id dari item menggunakan kueri exercise2.)
- ❑ Anda perlu menggunakan **GROUP BY** untuk menggabungkan **total_laba** dari setiap item.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
1 SELECT items.name,  
2     (items.price - items.cost) AS laba,  
3     1.0 * (items.price - items.cost) / (items.price) * 100 AS persentase_laba,  
4     COUNT(*) * (items.price - items.cost) AS total_profit  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 WHERE items.id = 23 OR items.id = 30 OR items.id = 3 OR items.id = 17 OR items.id = 1  
8 GROUP BY items.name, (items.price - items.cost), 1.0 * (items.price - items.cost) / (items.price) * 100  
9 ORDER BY COUNT(*) * items.price DESC
```

[Hasil]

name	laba	persentase_laba	total_profit
jaket kulit	222	69.375	5994
mantel merah	195	72.22222222222222	4875
sepatu lari	47	62.66666666666667	1645
rok	32	80.0	1664
sepatu merah	26	61.9047619047619	1248

```
1 SELECT items.name,  
2      (items.price - items.cost) AS laba,  
3      1.0 * (items.price - items.cost) / (items.price) * 100 AS persentase_laba,  
4      COUNT(*) * (items.price-items.cost) AS total_profit  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 WHERE items.id = 23 OR items.id = 30 OR items.id = 3 OR items.id = 17 OR items.id = 1  
8 GROUP BY items.name, (items.price - items.cost), 1.0 * (items.price - items.cost) / (items.price) * 100  
9 ORDER BY COUNT(*) * items.price DESC
```

Di baris ke 8 query:

GROUP BY items.name, (items.price - items.cost), 1.0 * (items.price - items.cost) / (items.price) * 100

Dapat di persingkat menggunakan alias yang di berikan ke kalkulasi terkait. Untuk mempersingkat code, Anda dapat menulisnya seperti berikut:

GROUP BY items.name, laba, persentase_laba

Anda bisa mendapatkan hasil yang sama dengan code ini menggunakan subquery dan klausa IN. Anda belum mempelajari IN di Progate tetapi terkadang berguna.

[Code]

```
1 SELECT items.name,  
2      (items.price-items.cost) AS profit,  
3      1.0 * (items.price - items.cost) / (items.price) * 100 AS profit_rate,  
4      COUNT(*) * (items.price-items.cost) AS total_profit  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 WHERE items.id IN (  
8     SELECT items.id  
9     FROM sales_records  
10    JOIN items ON sales_records.item_id = items.id  
11   GROUP BY items.id  
12   ORDER BY COUNT(*) * items.price desc  
13   limit 5  
14 )  
15 GROUP BY items.name, (items.price - items.cost), 1.0 * (items.price - items.cost) / (items.price) * 100  
16 ORDER BY COUNT(*) * (items.price-items.cost) desc
```

- ❑ Apakah Anda ingat bahwa tingkat keuntungan seluruh situs adalah 67% (dari permintaan data no. 1)? Anda dapat membandingkan tingkat keuntungan setiap item dengannya.
- ❑ Di antara 5 item yang menghasilkan pendapatan teratas, "jaket kulit", "mantel merah", dan "rok" memiliki tingkat keuntungan yang lebih tinggi dan 2 lainnya memiliki tingkat keuntungan yang lebih rendah daripada yang ada di seluruh situs.
- ❑ Selain itu, Anda juga dapat melihat berapa jumlah 5 item akun total keuntungan situs ini
(Jumlah total laba dari 5 item) / 31.428 = 51,7%.
- ❑ Sekarang Anda dapat menyimpulkan bahwa 5 item penghasil pendapatan teratas juga menghasilkan untung besar.

Latihan 4:

Cari tahu 5 item yang memberikan penghasilan laba terbatas

Apa yang ingin kita dapatkan

- ❑ Kami hanya mengambil data laba dari 5 item penghasil pendapatan terbatas. Tapi mungkin ada barang yang lebih menguntungkan.
- ❑ Mari kita dapatkan 5 item penghasil keuntungan terbatas dalam latihan ini.

Latihan 4

- ❑ Dapatkan "nama item", "keuntungan setiap item", "tingkat keuntungan setiap item" dan "total keuntungan setiap item" dari 5 item penghasil keuntungan terbatas.
- ❑ Urutkan hasil dari item yang memiliki "keuntungan total" tertinggi ke yang terendah.

- ❑ Tidak ada petunjuk untuk latihan ini. Anda hanya perlu mengubah kueri latihan 3.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
1 SELECT items.name,  
2      (items.price - items.cost) AS laba,  
3      1.0 * (items.price - items.cost) / (items.price) * 100 AS persentase_laba,  
4      COUNT(*) * (items.price-items.cost) AS total_laba  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 GROUP BY items.name, laba, persentase_laba  
8 ORDER BY COUNT(*) * laba DESC  
9 limit 5
```

[Hasil]

name	laba	persentase_laba	total_profit
jaket kulit	222	69.375	5994
mantel merah	195	72.22222222222222	4875
celana denim	34	56.66666666666667	2074
sweater	32	94.1176470588235	1856
rok	32	80.0	1664

- ❑ Dari 5 item penghasil pendapatan teratas, 4 item juga menempati peringkat 5 item penghasil laba teratas.
- ❑ Sekarang Anda dapat memutuskan untuk tetap menjual 5 item penghasil pendapatan teratas karena juga menguntungkan. Tapi terkadang ada barang yang laku tapi untung sedikit. Dalam hal ini, Anda dapat menyimpulkan mungkin Anda harus mengubah taktik penjualan Anda.
- ❑ Kita mendapat "sweater" dalam latihan ini sebagai item penghasil keuntungan No. 4. Anehnya, tingkat keuntungannya luar biasa tinggi, 94%. Artinya, Anda bisa mencoba mendiskon barang (masih menguntungkan) dan meningkatkan kuantitas penjualan untuk memaksimalkan total keuntungan.

Permintaan data no. 5:

Sekarang tim Anda sudah tau item-item mana saja yang paling berpengaruh. Agar dapat mengalokasikan biaya marketing secara efektif, Cari tahu 5 item penghasil pendapatan terburuk

Apa yang ingin kita dapatkan

- ❑ Mari lakukan hal yang sama untuk 5 item yang menghasilkan pendapatan terburuk!
- ❑ Penting juga untuk memeriksa item terburuk karena mungkin lebih baik berhenti menjualnya dan menghemat biaya penyimpanan.

Exercise 5

- ❑ Dapatkan "nama item", "kuantitas penjualan", "harga", dan "total pendapatan" dari 5 item terendah.
- ❑ Urutkan hasil dari item yang memiliki pendapatan terendah hingga tertinggi.

- ❑ Tidak ada petunjuk untuk latihan ini. Anda hanya perlu mengubah kueri permintaan data no. 2.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
1 SELECT items.name,  
2     COUNT(*) AS jumlah_penjualan,  
3     items.price,  
4     COUNT(*) * items.price AS total_pendapatan  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 GROUP BY items.name, items.price  
8 ORDER BY total_pendapatan ASC  
9 limit 5
```

[Hasil]

name	jumlah_penjualan	price	total_pendapatan
kaos kaki	18	10	180
kaos abu-abu	9	23	207
topi	20	15	300
syal	17	18	306
sandal	21	18	378

- ❑ Jumlah pendapatan barang adalah \$ 1371, yang hanya 2,9% dari seluruh pendapatan.
- ❑ Selama kuantitas penjualan kecil (terutama "kaos abu-abu"), Anda dapat mengurangi stok untuk menghemat biaya penyimpanan atau menjalankan kampanye untuk meningkatkan kuantitas penjualan.
- ❑ Tetapi Anda juga perlu memeriksa tingkat keuntungan setiap item karena sulit untuk menjalankan kampanye jika tidak menguntungkan. Upaya untuk menjual barang tidak akan membuahkan hasil jika tidak menghasilkan keuntungan yang cukup.

Permintaan data no. 6:

Cari tahu 5 item dengan penghasilan laba
terburuk

Apa yang ingin kita dapatkan

- ❑ Berikutnya, mari cari tahu 5 item dengan penghasilan laba terburuk

Exercise 6

- ❑ Dapatkan "nama item", "laba setiap item", "persentase laba setiap item" dan "total laba setiap item" dari 5 item dengan hasil laba terburuk.
- ❑ Urutkan hasil dari item yang memiliki "total laba" terendah ke tertinggi.

- ❑ Tidak ada petunjuk untuk latihan ini. Anda hanya perlu mengubah kueri permintaan data no. 4.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
1 SELECT items.name,  
2      (items.price - items.cost) AS laba,  
3      1.0 * (items.price - items.cost) / (items.price) * 100 AS persentase_laba,  
4      COUNT(*) * (items.price-items.cost) AS total_laba  
5 FROM sales_records  
6 JOIN items ON sales_records.item_id = items.id  
7 GROUP BY items.name, laba, persentase_laba  
8 ORDER BY total_laba ASC  
9 limit 5
```

[Result]

name	laba	persentase_laba	total_laba
kaos abu-abu	13	56.5217391304348	117
kaos kaki	8	80.0	144
kaos biru tua	11	55.0	220
syal	13	72.2222222222222	221
topi	12	80.0	240

- ❑ Jumlah keuntungan barang adalah \$ 942, yang hanya 3,0% dari seluruh pendapatan.
- ❑ Tingkat keuntungan dari “kaos abu-abu” dan “kaos biru tua” sangat kecil dan “kaos abu-abu” juga merupakan salah satu item yang menghasilkan pendapatan terburuk. Anda harus melakukan sesuatu untuk barang-barang ini.
- ❑ Tingkat keuntungan dari “kaus kaki” tidak buruk meskipun keuntungannya sendiri sangat kecil. Dalam hal ini, menjual 3 pasang kaus kaki dengan sedikit diskon mungkin berhasil!

Latihan 7:

Cari tahu keuntungan total dan tingkat keuntungan item tergantung pada gender

Apa yang ingin kita dapatkan

- ❏ Karena tabel `items` berisi data tentang `gender`, mari kita kategorikan item berdasarkan gender.

Latihan 7

- ❏ Dapatkan "gender", "total laba dari setiap gender" dan "persentase laba dari setiap gender".

- ❏ Anda perlu menggunakan `sum` untuk mendapatkan laba dan persentase laba karena ini penting untuk mendapatkan total item per gender.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

[Code]

```
1 SELECT gender,  
2      SUM(items.price - items.cost) AS total_laba,  
3      1.0 * SUM(items.price - items.cost) / SUM(items.price) * 100 AS persentase_laba  
4 FROM sales_records  
5 JOIN items ON sales_records.item_id = items.id  
6 GROUP BY Gender
```

[Hasil]

gender	total_laba	persentase_laba
0	3644	60.5818786367415
1	18450	70.8362128541811
2	9334	63.4189427911401

- ❑ Item untuk "wanita" menghasilkan banyak laba dan memiliki tingkat laba tertinggi.
- ❑ Tetapi Anda perlu melihat lebih dalam setiap kategori untuk mencapai kesimpulan. Katakanlah hanya ada 2 item yang laris manis di kategori wanita, hanya beberapa item tersebut dapat berdampak besar pada hasil.
- ❑ Penting untuk selalu mencari tahu alasan dari hasil yang Anda peroleh. Kami tidak melakukan analisis lebih rinci tentang gender dalam latihan ini, tetapi Anda dapat mencobanya sendiri!

- ❑ Sudah selesai dilakukan dengan baik!! Anda telah menyelesaikan bab pertama dari latihan ini!
- ❑ Seperti yang Anda pelajari, mampu menulis kueri adalah keterampilan yang sangat penting, tetapi penting juga untuk mendapatkan data dari berbagai sudut pandang dan membuat sekumpulan angka menjadi data yang bermakna untuk membuat keputusan bisnis dan keuangan yang lebih baik!

Chapter2

- ❑ Ada 8 latihan lagi di bab ini.
- ❑ Karena kita telah selesai menganalisis pendapatan dan keuntungan item dengan mendapatkan data dari tabel `items` dan `sales_record`, mari kita lanjutkan untuk mendapatkan data yang terkait dengan `user` di bab ini.
- ❑ Bab ini lebih advance ketimbang bab sebelumnya, tetapi jangan khawatir, majulah dengan kecepatan Anda sendiri.

Latihan 8:

Cari tahu jumlah pengguna aktif dan persentasenya.

- ❑ Selamat datang di latihan pertama dalam bab ini!
- ❑ Pertama-tama, mari kita mulai dari gambaran besar lagi. Dalam latihan ini, kita akan mengetahui berapa banyak pengguna aktif yang ada di situs belanja tersebut.

Latihan 8

- ❑ Dapatkan jumlah pengguna aktif (mereka yang membeli item setidaknya sekali) dan persentase pengguna aktif.

Hint

- ❑ Anda dapat menggunakan `user_id` di tabel `items` untuk mendapatkan jumlah pengguna aktif.
- ❑ Berhati-hatilah karena pengguna yang sama dapat membeli item beberapa kali dan ini berarti `user_id` yang sama dicatat di tabel item beberapa kali. `DISTINCT` berguna dalam kasus ini.
- ❑ Persentase pengguna aktif dapat dihitung dengan rumus: $(\text{jumlah pengguna aktif}) / (\text{jumlah total pengguna}) * 100$
- ❑ Anda perlu menggunakan subquery untuk menghitung Persentase pengguna aktif.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

Answer

[Code]

```
exercise1.sql exercise2.sql exercise3.sql
1 SELECT COUNT(distinct sales_records.user_id) AS pengguna_aktif,
2     1.0 * 100 * COUNT(distinct sales_records.user_id) / (
3     SELECT COUNT(*)
4     FROM users
5     ) AS persentase_aktif
6 FROM sales_records
7
```

[Result]

pengguna_aktif	persentase_aktif
283	87.0769230769231

- ❑ Dari 325 pengguna yang mendaftar di situs belanja, 283 orang (87%) benar-benar membeli barang.
- ❑ Untuk mendapatkan lebih banyak keuntungan, Anda perlu mengaktifkan pengguna yang tidak aktif. Anda dapat mencoba mengirim email atau semacamnya.
- ❑ Cara lain untuk mendapatkan lebih banyak keuntungan adalah membuat pengguna aktif berbelanja lebih sering atau membeli barang yang lebih mahal.
- ❑ Selanjutnya, mari kita pelajari tentang frekuensi rata-rata dan pengeluaran rata-rata pengguna aktif!

Latihan 9:

Cari tahu frekuensi rata-rata pembelian, pengeluaran rata-rata, dan pengeluaran rata-rata per pembelian pengguna aktif

- ❑ Mari dapatkan informasi tentang pembelian pengguna aktif. Ini terlihat rumit tetapi Anda bisa mendapatkan datanya dengan menggabungkan pengetahuan yang sudah Anda pelajari!

Latihan 9

- ❑ Dapatkan 3 data pengguna aktif berikut:
 - ❑ Frekuensi rata-rata pembelian
 - ❑ Pengeluaran rata-rata
 - ❑ Pengeluaran rata-rata per pembelian

Petunjuk

- ❑ Anda bisa mendapatkan data dengan perhitungan berikut.
- ❑ Frekuensi rata-rata pembelian:
$$1.0 * (\text{Jumlah total pembelian di situs}) / (\text{Jumlah pengguna aktif})$$
- ❑ Pengeluaran rata-rata:
$$1.0 * (\text{Jumlah harga item yang terjual}) / (\text{Jumlah pengguna aktif})$$
- ❑ Pengeluaran rata-rata per pembelian
$$1.0 * (\text{Belanja rata-rata}) / (\text{Frekuensi rata-rata pembelian})$$
- ❑ Kami akan menampilkan semua data dalam satu kueri tetapi tidak buruk untuk menulis beberapa kueri untuk mendapatkan setiap data.

Jawaban:

Cobalah untuk mengerjakannya
terlebih dahulu sebelum melihat
jawabannya!!

Answer

[Code]

```
1 SELECT 1.0 * COUNT(*) / COUNT(distinct sales_records.user_id) AS rata_jumlah_penjualan,
2       1.0 * SUM (items.price) / COUNT(distinct sales_records.user_id) AS rata_pengeluaran,
3       (
4         1.0 * SUM (items.price) / COUNT(distinct sales_records.user_id)
5       ) / (
6         1.0 * COUNT(*) / COUNT(distinct sales_records.user_id)
7       ) AS rata_pengeluaran_per_belanjaan
8 FROM sales_records
9 JOIN items ON sales_records.item_id = items.id
```

[Hasil]

rata_jumlah_penjualan	rata_pengeluaran	rata_pengeluaran_per_belanjaan
3.21554770318021	165.296819787986	51.4054945054945

[Different Answer]

- ❑ In the previous answer, the code on line 3-7 became a bit too long and the same codes are used twice.
- ❑ With subquery, you can use the name after **AS** for calculation.

```
1 SELECT rata_jumlah_penjualan,  
2        rata_pengeluaran,  
3        1.0 * rata_pengeluaran / rata_jumlah_penjualan AS rata_pengeluaran_per_belanjaan  
4 FROM (  
5     SELECT 1.0 * COUNT(*) / COUNT(distinct sales_records.user_id) AS rata_jumlah_penjualan,  
6     1.0 * SUM(items.price) / COUNT(distinct sales_records.user_id) AS rata_pengeluaran  
7 FROM sales_records  
8 JOIN items ON sales_records.item_id = items.id  
9 )
```


[Different Answer]

- ❑ Of course it's not bad to write several queries like the answer below.

```
1 SELECT 1.0 * COUNT(*) / COUNT(distinct sales_records.user_id) AS rata_jumlah_penjualan
2 FROM sales_records
```

```
1 SELECT 1.0 * SUM (items.price) / COUNT(distinct sales_records.user_id) AS rata_pengeluaran,
2 FROM sales_records
3 JOIN items ON sales_records.item_id = items.id
```

Analysis from this result

- ❑ We got the overall data of active users. Active users bought items 3.2 times, spent \$165.3 and spent \$51.4 per purchase in a month.
- ❑ This is the standard for comparisons. For example, you can compare it to the same data of a different month to find out which month went well.
- ❑ You can also use the data to analyze users to find out who bought items more often and who spent more than the average. Let's do this from the next exercises!

- ❑ There are advantages and disadvantages of getting multiple data in one query. You should choose the appropriate way depending on the situation!

Advantage:

- ❑ You can see all the data at a glance, which makes it easier to compare and analyze.

Disadvantage:

- ❑ The result is easy to see but the query itself becomes long and complicated, which makes it hard for other people to understand and edit the query.

Exercise 10:

Get the list of users who purchased more often than the average

- ❑ We got the average frequency of purchases in the previous exercise. This is an important data because we can find excellent customers who purchased more often than the average using the data.

Exercise 10

- ❑ Get the following data of users who purchased more often than the average:
 - ❑ The name of users
 - ❑ The number of purchases of each user
- ❑ Order the result from the highest frequency of purchases to the lowest.

Hint

- ❑ Step1: Find out what tables are needed to get the data.
- ❑ Step2: Get the number of purchases of all users using `GROUP BY`.
- ❑ Step3: Using `HAVING`, add a condition of "bigger than average frequency of purchases".
- ❑ Step4: Use `ORDER BY` to line up the result based on the number of purchases.

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1 SELECT users.name, count(*) AS number_of_purchases
2 FROM sales_records
3 JOIN users ON sales_records.user_id = users.id
4 GROUP BY users.name
5 HAVING number_of_purchases >= (
6     SELECT 1.0 * COUNT(*) / COUNT(distinct sales_records.user_id) AS average_number_of_purchases
7     FROM sales_records
8 )
9 ORDER BY number_of_purchases desc
```

[Result]

name	number_of_purchases
Hobie	14
Burch	12
Sam	12
Cale	11
Rori	10

Analysis from this result

- ❑ We got the list of customers who purchased very often. We can send them a coupon to keep their loyalty.
- ❑ The frequency of purchases is an important factor to increase the sales of the shopping site. You can think of incentives like “giving gold status” or “giving a special coupon” to the users who purchases more than 5 times in a month. Since the average frequency was 3.2(exercise 9), this can work well.

Exercise 11:

Find out how many and what percentage of users spent more than the average

- ❑ We got the average spending of active users in the exercise 9. But we are not sure about how many people actually spent that much. Sometimes the average doesn't work because there is the case that only a small part of people buy a lot and most people don't spend. Let's find it out in this exercise!

Exercise 11

- ❑ Get the following data:
 - ❑ The number of users whose spending is more than the average
 - ❑ The percentage of the high-spending users out of all active users

Hint

- ❑ Step1: Get the total spending of each user using `GROUP BY`.
- ❑ Step2: Using `HAVING`, add a condition of "bigger than average spending".
- ❑ Step3: Using `COUNT`, get the number of the users from the result of the Step2.
- ❑ Step4: Using a subquery, divide the Step3 result by the total number of active users.

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1  SELECT COUNT(user_id) AS number_of_users,  
2         1.0 * 100 * COUNT(user_id) / (  
3         SELECT COUNT(distinct sales_records.user_id)  
4         FROM sales_records  
5         ) AS rate  
6  FROM (  
7      SELECT user_id,  
8             SUM(price) AS total_spending  
9      FROM sales_records  
10     JOIN items ON sales_records.item_id = items.id  
11     GROUP BY user_id  
12     HAVING total_spending >= (  
13         SELECT 1.0 * SUM(price) / COUNT(distinct sales_records.user_id) AS average_spending  
14         FROM items  
15         JOIN sales_records ON items.id = item_id  
16     )  
17 )
```

[Result]

number_of_users	rate
94	33.2155477031802

Analysis from this result

- ❑ We now know that 94 people spent more than the average and it accounts for 33% of active users.
- ❑ If your website is like “only 5% of people spend a lot”, you should think of selling high brand expensive items but you don’t have to this time.
- ❑ We got all the data in a query but of course you can get the same data writing several queries.

Exercise 12:

Find out the items which have higher prices than the average spending per purchase

What We Want to Get

- ❑ We got the average spending per purchase in the exercise 9. If we sell more expensive items than the average, it means we can increase the average spending! Let's find out the items!

Exercise 12

- ❑ Get the following data of items which have higher prices than the average spending per purchase:
 - ❑ The name
 - ❑ The price
- ❑ Order the items based on the price(highest->lowest)

Hint

- ❏ No hint for this exercise. You can utilize the exercise 9 query to get the average spending per purchase.

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1  SELECT name, price
2  FROM items
3  WHERE price >= (
4      SELECT 1.0 * average_spending / average_number_of_purchases AS average_spending_per_purchase
5      FROM (
6          SELECT 1.0 * COUNT(*) / COUNT(distinct sales_records.user_id) AS average_number_of_purchases,
7          1.0 * SUM(price) / COUNT(distinct sales_records.user_id) AS average_spending
8          FROM items
9          JOIN sales_records ON items.id = sales_records.item_id
10     )
11 )
12 ORDER BY price desc
```

[Result]

name	price
leather jacket	320
red coat	270
blouse	80
running shoes	75
denim pants	65
denim pants	60

Analysis from this result

- ❑ It turns out that 6 items are more expensive than the average spending per purchase. If customers buy those items, you can increase the average. Displaying the items on the top page of the website can work.
- ❑ You can also get the items which have lower prices than average. For those items, you can sell them by packages like 2 T-shirts and 3 socks with some discount.

Exercise 13:

Retrieve purchases data based the gender of users

- ❑ Looking into data from different perspectives is always important. Let's get data based on the gender of users.

Exercise 13

- ❑ Get the following data of each gender
 - ❑ Gender
 - ❑ The number of users
 - ❑ Total spending
 - ❑ Total number of purchases
 - ❑ Average spending per user
 - ❑ Average spending per purchase

Hint

- ❏ You need to use `GROUP BY` to get the data based on `gender`.
- ❏ We don't give you other hints in this exercise, you can try by yourself using the knowledge you learned so far!

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1  SELECT users.gender,  
2         COUNT(distinct sales_records.user_id) AS number_of_users,  
3         SUM(items.price) AS total_spending,  
4         COUNT(*) AS number_of_purchases,  
5         1.0 * SUM(items.price) / COUNT(distinct sales_records.user_id) AS average_spending_per_user,  
6         1.0 * SUM(items.price) / COUNT(*) AS average_spending_per_purchase  
7  FROM items  
8  JOIN sales_records ON sales_records.item_id = items.id  
9  JOIN users ON users.id = sales_records.user_id  
10 GROUP BY users.gender
```

[Result]

gender	number_of_users	total_spending	number_of_purchases	average_spending_per_user	average_spending_per_purchase
0	112	20112	397	179.571428571429	50.6599496221662
1	171	26667	513	155.947368421053	51.9824561403509

Analysis from this result

- ❑ It turned out that the shopping site has more female users and gets more revenue from them. But looking at the average spending per user, male users spend more than female users.
- ❑ As you can see, you can get different insight when you get data based on a specific category. Bear in mind that it's important to see data from various perspectives.

Exercise 14:

Retrieve the purchase data of users in their 20s.

- ❑ Next, let's get data based on the age of users.

Exercise 14

- ❑ Get the following data of users in their 20s
 - ❑ The number of users
 - ❑ Total spending
 - ❑ Total number of purchases
 - ❑ Average spending per user
 - ❑ Average spending per purchase

Hint

- ❏ You can express the condition to get only the data of users in their 20's using `where`.

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1 SELECT COUNT(distinct sales_records.user_id) AS number_of_users,  
2        SUM(items.price) AS total_spending,  
3        COUNT(*) AS the_number_of_purchases,  
4        1.0 * SUM(items.price) / COUNT(distinct sales_records.user_id) AS average_spending_per_user,  
5        1.0 * SUM(items.price) / COUNT(*) AS average_spending_per_purchase  
6 FROM items  
7 JOIN sales_records ON items.id = item_id  
8 JOIN users ON user_id = users.id  
9 WHERE age >= 20 AND age < 30;
```

[Result]

number_of_users	total_spending	the_number_of_purchases	average_spending_per_user	average_spending_per_purchase
107	21716	375	202.953271028037	57.9093333333333

Analysis from this result

- ❑ The query itself is simple and it's easy to understand. Since you successfully got the data of users in their 20s and you already know the data of all users, you can compare the data and you will know how much the users in their 20's account for the total numbers.
- ❑ You can do the same for 30s, 40s and 50s one by one. But let's try to get all the data at once in the next exercise.

Exercise 15 (Last exercise!):

Find out the total spending of each age (10s, 20s, 30s...) and what percentage they account for.

Exercise 15

- ❑ This is the last exercise! Let's get the total spending and percentages of each age like the following image.

age	total_spending	rate
10	xxx	xxx
20	xxx	xxx
30	xxx	xxx
40	xxx	xxx
50	xxx	xxx
60	xxx	xxx

Hint

- ❑ As the age stored in the database is not like the form of 10, 20 and 30, you need to use a technique to change the actual ages to the form we want.
- ❑ To do that, you can utilize the characteristics of integer type data. Try the calculation:
$$(\text{age} / 10) * 10$$
- ❑ Let's say the age is 25, "25 / 10 = 2" if it's integer. Then you can multiply 10 to get 20 from the actual age.

Answer:

Please think well before going to the
answer!!

Answer

[Code]

```
1  SELECT (users.age / 10) * 10 as age,  
2         SUM(items.price) as total_spending,  
3         1.0 * 100 * SUM(items.price) / (  
4         SELECT SUM(items.price)  
5         FROM items  
6         JOIN sales_records ON items.id = sales_records.item_id  
7         ) AS rate  
8  from users  
9  JOIN sales_records on users.id = sales_records.user_id  
10 JOIN items on items.id = sales_records.item_id  
11 GROUP BY (age / 10) * 10;
```

[Result]

age	total_spending	rate
10	3515	7.51405545223284
20	21716	46.4225400286453
30	13181	28.1771735180316
40	4574	9.77789178905064
50	2034	4.34810491887385
60	1759	3.76023429316574

Analysis from this result

- ❑ You now understand that the 70% of the revenue came from users in their 20s and 30s.
- ❑ Apparently the age of customers is important to think of what items you should sell on the shopping site.
- ❑ Great job to reach the end of this exercises!!!

The end of the exercises

- ❏ We hope you have had fun and got a deeper understanding of data analysis with SQL!
- ❏ Python and R have been said to be the languages for data science recently, but SQL can do lots of things and it's easier to use in your work.
- ❏ Repetition is the key to acquire a new skill. Please continue to try getting more data using what you learned with us!! Enjoy learning!!!

Thank you and
keep learning!!

