

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 Growth Manager 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 buka halaman SQL Latihan I di Progate dan hapus kueri apapun yang ada di text editor agar tampilan kosong. Kita akan menggunakan text editor dan database yang sudah tersedia di bagian tersebut:

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

Persiapan

Kali ini, abaikan instruksi di sebelah kiri, dan tidak perlu menekan tombol **irim**. Setiap selesai mengerjakan soal, cukup tekan tombol ▶ dan jika sudah benar, jangan lupa untuk screenshot sebagai tanda bukti sudah melewati setiap soal. Total screenshot yang harus Anda submit ada 7.

The screenshot shows a web-based SQL practice environment. The breadcrumb navigation at the top reads "SQL Latihan I > Menganalisis Database Situs Fashion > 2. Pengguna". The user's name "Norman Ganto (Lv.55)" is visible in the top right. The interface is divided into two main sections: "Instruksi" on the left and "Latihan" on the right. The "Instruksi" section contains two tasks: "Cari usia rata-rata semua pengguna" and "Cari nilai dari setiap kolom untuk pengguna pria yang berusia kurang dari 20 tahun." This section is enclosed in a red box with a large red 'X' over it, indicating it should be ignored. The "Latihan" section features a code editor with tabs for "exercise1.sql", "exercise2.sql", and "exercise3.sql". Below the editor are tabs for "Jawaban Anda (exercise1.sql)" and "Jawaban Benar (exercise1.sql)". A large play button (▶) is circled in green on the right side of the interface. At the bottom right, a "Kirim" button is highlighted with a red box and a red diagonal line through it, indicating it should not be clicked. The status bar at the bottom shows "1 / 10 Sudah diselesaikan" and a "Hubungi kami!" button.

Cara Pengumpulan Hasil Tugas Akhir

1. Setiap Anda berhasil menyelesaikan latihan dan hasil tabel sudah sama dengan yang ada disini, screenshot penuh layar komputer Anda
2. Lakukan untuk semua latihannya (ada 7 total)
3. Setelah selesai, masukkan screenshot-screenshot Anda ke sebuah dokumen google slides / ppt
4. Namakan dokumen Anda dengan nama Anda dan kelas Anda, sebagai contoh berikut:

Space Up Bootcamp - Norman Ganto - SQL

5. Kirim dokumen Anda ke info-indonesia@progate.com

Meninjau Ulang Struktur Tabel

- Ada 3 tabel di database: `items`, `sales_records` dan `users`

tabel `items`

kolom	tipe data
id	angka
name	teks
gender	angka
price	angka
cost	angka

tabel `sales_records`

kolom	tipe data
id	angka
item_id	angka
user_id	angka
purchased_at	tanggal

tabel `users`

kolom	tipe data
id	angka
name	teks
gender	angka
age	angka

- Anda dapat melihat hal berikut:
 - 30 items terdaftar di situs.
 - Ada 910 rekor penjualan (`sales_records`) pada Juli 2018.
 - 325 pengguna terdaftar di tabel `users`.

Chapter 1

Ayo kita mulai!

Permintaan Data no. 1:

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

Anda

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

Satu point penting

- ❑ 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!!

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

Hasil yang benar:

total_revenue	total_profit	profit_rate
46779	31428	67.1839928172898

Analisis Hasil

- ❑ 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.

Hasil yang benar:

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 teratas

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.

Hasil yang benar:

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

Catatan

```
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 = 23 OR items.id = 30 OR items.id = 3 OR items.id = 17 OR items.id = 1  
8 GROUP BY items.name, profit, profit_rate  
9 ORDER BY total_profit 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
```

Analisis Hasil

- ❑ 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.

Permintaan Data no. 4:

Cari tahu 5 item yang memberikan penghasilan laba teratas

Apa yang ingin kita dapatkan

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

Hasil yang benar:

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

Analisis Hasil

- ❑ 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 tahu item-item mana saja yang paling berpengaruh terhadap pendapatan perusahaan. 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.

Hasil yang benar:

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.

Hasil yang benar:

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

Analisis hasil

- ❑ 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!

Permintaan Data no. 7:

Cari tahu keuntungan total dan tingkat keuntungan item berdasarkan jenis kelamin pembeli

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.

Hasil yang benar:

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!

- ❑ Selamat! Anda sudah selesai Chapter 1 dengan baik!!
- ❑ 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!

Chapter 2

- ❑ Ada 8 jenis permintaan data lagi di chapter 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.
- ❑ Chapter ini lebih advanced ketimbang chapter sebelumnya, tetapi jangan khawatir, cobalah dengan kesediaan waktu yang Anda miliki, tidak perlu terburu-buru

Latihan 8:

Cari tahu jumlah pengguna aktif dan persentasenya.

Apa yang ingin kita dapatkan

- ❑ Selamat datang di latihan pertama dalam Chapter 2 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.

- ❑ 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.

Hasil yang benar:

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!

Permintaan Data no. 9:

Cari tahu frekuensi rata-rata

pembelian, pengeluaran rata-rata, dan

pengeluaran rata-rata per pembelian

pengguna aktif

Apa yang ingin didapatkan

- ❑ 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.

[Cara Lain]

- ❑ Pada jawaban sebelumnya, code pada baris 3-7 agak terlalu panjang dan code yang sama digunakan dua kali.
- ❑ Dengan subkueri, Anda dapat menggunakan nama setelah AS untuk penghitungan.

```
1 SELECT rata_jumlah_penjualan,  
2         rate_pembelian,  
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_pembelian,  
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 )
```

[Cara lain]

- ❑ Kamu juga dapat membuat beberapa kueri seperti gambar dibawah.

```
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
```

Hasil yang benar:

rata_jumlah_penjualan	rata_pengeluaran	rata_pengeluaran_per_belanjaan
3.21554770318021	165.296819787986	51.4054945054945

- ❑ Kita mendapat data keseluruhan dari pengguna aktif. Pengguna aktif membeli item 3,2 kali, membelanjakan \$ 165,3, dan \$ 51,4 per pembelian dalam sebulan.
- ❑ Ini adalah standar perbandingan. Misalnya, Anda dapat membandingkannya dengan data yang sama pada bulan yang berbeda untuk mengetahui bulan mana yang berjalan dengan baik.
- ❑ Anda juga dapat menggunakan data untuk menganalisis pengguna untuk mengetahui siapa yang lebih sering membeli barang dan siapa yang belanja di atas rata-rata. Ayo lakukan ini dari latihan berikutnya!

- ❑ Ada keuntungan dan kerugian mendapatkan banyak data dalam satu kueri. Anda harus memilih cara yang tepat tergantung pada situasinya!

Keuntungan:

- ❑ Anda dapat melihat semua data bersamaan, sehingga lebih mudah untuk membandingkan dan menganalisisnya.

Kerugian:

- ❑ Hasilnya mudah dilihat tetapi kueri itu sendiri menjadi panjang dan rumit, yang membuat orang lain sulit memahami dan mengedit kueri tersebut.

Permintaan data no. 10:

Dapatkan daftar pengguna yang membeli lebih sering daripada rata-rata

Apa yang ingin kita dapatkan

- ❑ Kami mendapatkan frekuensi rata-rata pembelian pada latihan sebelumnya. Ini adalah data penting karena kami dapat menemukan pelanggan yang sangat baik yang membeli lebih sering daripada rata-rata menggunakan data.

Exercise 10

- ❑ Dapatkan data berikut dari pengguna yang membeli di atas rata-rata:
 - ❑ Nama pengguna
 - ❑ Jumlah pembelian setiap pengguna
- ❑ Urutkan hasil dari frekuensi pembelian tertinggi ke yang terendah.

- ❑ Step1: Cari tahu tabel apa yang dibutuhkan untuk mendapatkan data.
- ❑ Step 2: Dapatkan jumlah pembelian dari semua pengguna menggunakan **GROUP BY**.
- ❑ Step3: Gunakan **HAVING**, menambahkan kondisi "frekuensi pembelian lebih besar dari rata-rata".
- ❑ Step4: Gunakan **ORDER BY** untuk mengantre hasil berdasarkan jumlah pembelian.

Hasil yang benar:

name	jumlah_pembelian
Hobie	14
Burch	12
Sam	12
Cale	11

- ❑ Kami mendapat daftar pelanggan yang sangat sering membeli. Kami dapat mengirimkan kupon untuk menjaga loyalitas mereka.
- ❑ Frekuensi pembelian merupakan faktor penting untuk meningkatkan penjualan situs belanja tersebut. Anda dapat memikirkan insentif seperti "memberikan status emas" atau "memberikan kupon khusus" kepada pengguna yang membeli lebih dari 5 kali dalam sebulan. Karena frekuensi rata-rata adalah 3,2 (latihan 9), ini dapat bekerja dengan baik.

Permintaan Data no. 11:

Cari tahu berapa banyak dan berapa persentase pengguna yang berbelanja di atas rata-rata

Apa yang ingin didapatkan

- ❑ Kami mendapatkan rata-rata pengeluaran pengguna aktif dalam latihan 9. Tetapi kami tidak yakin tentang berapa banyak orang yang sebenarnya membelanjakan sebanyak itu. Terkadang rata-rata tidak tepat karena ada kasus dimana hanya sebagian kecil orang yang membeli dalam jumlah besar dan sebagian besar orang tidak. Mari kita cari tahu di latihan ini!

Latihan 11

- ❑ Dapatkan data berikut:
 - ❑ Jumlah pengguna yang pembelanjannya di atas rata-rata
 - ❑ Persentase pengguna yang berbelanja banyak dari semua pengguna aktif

- ❑ Step1: Dapatkan total pembelanjaan yang digunakan setiap pengguna **GROUP BY**.
- ❑ Step2: Gunakan **HAVING**, tambahkan ketentuan "lebih besar dari pengeluaran rata-rata".
- ❑ Step3: Gunakan **COUNT**, dapatkan jumlah pengguna dari hasil Step2.
- ❑ Step4: Menggunakan subkueri, bagi hasil Step3 dengan jumlah total pengguna aktif.

Hasil yang benar:

jumlah_pengguna	persentase
94	33.2155477031802

Analysis dari this result

- ❑ Kami sekarang tahu bahwa 94 orang membelanjakan lebih dari rata-rata dan itu menyumbang 33% dari pengguna aktif.
- ❑ Jika situs web Anda seperti "hanya 5% orang yang membelanjakan banyak", Anda harus memikirkan untuk menjual barang-barang bermerek mahal tetapi Anda tidak perlu melakukannya saat ini.
- ❑ Kami mendapatkan semua data dalam satu kueri tetapi tentu saja Anda bisa mendapatkan data yang sama dengan menulis beberapa kueri.

Permintaan Data no. 12:

Cari tahu barang-barang yang memiliki harga lebih tinggi dari pengeluaran rata-rata per pembelian

Apa yang ingin didapatkan

- ❑ Kita mendapatkan **pengeluaran rata-rata per pembelian** dalam latihan 9. Jika kita menjual barang lebih mahal daripada rata-rata, itu berarti kita bisa meningkatkan pengeluaran rata-rata! Ayo cari tahu itemnya!

Exercise 12

- ❑ Dapatkan data item berikut yang memiliki harga lebih tinggi dari **rata-rata pengeluaran per pembelian**:
 - ❑ Nama
 - ❑ Harga
- ❑ Pesan barang berdasarkan harga (tertinggi->terendah)

- ❑ Tidak ada petunjuk untuk latihan ini. Anda dapat menggunakan kueri latihan 9 untuk mendapatkan pengeluaran rata-rata per pembelian.

Hasil yang benar:

name	price
jaket kulit	320
mantel merah	270
blus	80
sepatu lari	75

- ❑ Ternyata 6 item lebih mahal dari pengeluaran rata-rata per pembelian. Jika pelanggan membeli barang-barang tersebut, Anda dapat meningkatkan rata-rata. Menampilkan item di halaman atas situs web akan dapat membantu.
- ❑ Anda juga bisa mendapatkan item yang memiliki harga lebih rendah dari rata-rata. Untuk barang-barang tersebut bisa dijual dengan paket seperti 2 kaos dan 3 kaos kaki dengan potongan harga.

Permintaan Data no. 13:

Ambil data pembelian berdasarkan jenis kelamin pengguna

Apa yang ingin kita dapatkan

- ❑ Lihat data dari perspektif yang berbeda selalu penting. Mari kita dapatkan data berdasarkan jenis kelamin pengguna.

Latihan 13

- ❑ Dapatkan data berikut dari setiap jenis kelamin
 - ❑ Jenis kelamin
 - ❑ Jumlah pengguna
 - ❑ Total pengeluaran
 - ❑ Jumlah total pembelian
 - ❑ Pembelanjaan rata-rata per pengguna
 - ❑ Pengeluaran rata-rata per pembelian

- ❑ Gunakan **GROUP BY** untuk mendapatkan data berdasarkan **jenis kelamin (gender)**.
- ❑ Kami tidak memberi Anda petunjuk lain dalam latihan ini, Anda dapat mencobanya sendiri menggunakan pengetahuan yang telah Anda dapatkan.

Hasil yang benar:

gender	jumlah_pengguna	total_pengeluaran	jumlah_pembelian	rata_pengeluaran_p...	rata_pengeluarn_per...
0	283	5874042	114263	20756.332155477	51.4080848568653
1	283	9282354	180577	32799.8374558304	51.4038554190179

- ❑ Ternyata situs belanja tersebut memiliki lebih banyak pengguna wanita dan mendapat lebih banyak pendapatan dari mereka. Namun melihat pembelanjaan rata-rata per pengguna, pengguna pria membelanjakan lebih banyak daripada pengguna wanita.
- ❑ Seperti yang Anda lihat, Anda bisa mendapatkan wawasan yang berbeda saat mendapatkan data berdasarkan kategori tertentu. Ingatlah bahwa penting untuk melihat data dari berbagai perspektif.

Permintaan Data no. 14:

Ambil data pembelian pengguna berusia 20-an.

Apa yang ingin kita dapatkan

- ❑ Selanjutnya, dapatkan data berdasarkan usia pengguna.

Latihan 14

- ❑ Dapatkan data berikut dari pengguna berusia 20-an
 - ❑ Jumlah pengguna
 - ❑ Total pengeluaran
 - ❑ Jumlah total pembelian
 - ❑ Pembelanjaan rata-rata per pengguna
 - ❑ Pengeluaran rata-rata per pembelian

- ❑ Anda dapat mengekspresikan kondisi untuk hanya mendapatkan data pengguna berusia 20-an dengan menggunakan `where`.

Hasil yang benar:

jumlah_pengguna	total_pengeluaran	jumlah_pembelian	rata_pengeluaran_per_pen...	rata_pengeluaran_per_pe...
107	21716	375	202.953271028037	57.9093333333333

- ❑ Kueri itu sendiri sederhana dan mudah dipahami. Karena Anda berhasil mendapatkan data pengguna berusia 20-an dan sudah mengetahui data dari semua pengguna, Anda dapat membandingkan data tersebut dan Anda akan mengetahui berapa banyak jumlah total pengguna akun yang berumur 20-an.
- ❑ Anda dapat melakukan hal yang sama untuk pengguna berusia 30-an, 40-an, dan 50-an satu per satu. Tapi mari kita coba mendapatkan semua data sekaligus di latihan berikutnya.

Permintaan Data no. 15 (latihan terakhir!):

Cari tahu total pengeluaran setiap usia (10-an, 20-an, 30-an ...) dan berapa persentase mereka.

Latihan 15

- Ini latihan terakhir! Mari kita dapatkan total pembelanjaan dan persentase tiap usia seperti gambar berikut.

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

- ❑ Karena umur yang disimpan di database tidak seperti bentuk 10, 20 dan 30, Anda perlu menggunakan teknik untuk mengubah umur sebenarnya ke bentuk yang kita inginkan.
- ❑ Untuk melakukan itu, Anda dapat memanfaatkan karakteristik data tipe integer. Coba kalkulasi ini:
 $(age / 10) * 10$
- ❑ Misalkan usianya adalah 25, "25/10 = 2" jika bilangan bulat. Kemudian kamu bisa mengalikannya dengan 10 agar menjadi 20 dari umur sebenarnya.

Hasil yang benar:

umur	total_spending	persentase
10	3515	7.51405545223284
20	21716	46.4225400286453
30	13181	28.1771735180316
40	4574	9.77789178905064

Analisis Hasil

- ❑ Sekarang Anda memahami bahwa 70% pendapatan berasal dari pengguna berusia 20-an dan 30-an.
- ❑ Ternyata usia pelanggan penting untuk memikirkan barang apa yang harus Anda jual di situs belanja tersebut.
- ❑ Selamat telah menyelesaikan latihan ini!

- ❑ Kami harap Anda bersenang-senang dan mendapatkan pemahaman yang lebih dalam tentang analisis data dengan SQL!
- ❑ Python dan R telah dikatakan sebagai bahasa untuk ilmu data baru-baru ini, tetapi SQL tidak kalah, dan dapat melakukan banyak hal dan (terkadang) lebih mudah digunakan.
- ❑ Pengulangan adalah kunci untuk memperoleh keterampilan baru. Silakan untuk mencoba mendapatkan lebih banyak data menggunakan apa yang telah Anda pelajari bersama kami !! Selamat belajar !!!

Selamat! Walau sudah selesai, jangan berhenti berlatih!!

