Perincian RAW Data

Kolum	Tujuan / Fungsi
SessionID	ID unik untuk setiap sesi lawatan, elak duplikasi dan analisis lawatan individu.
Date	Tarikh sesi berlaku, untuk analisis trend harian/mingguan/bulanan & kesan kempen.
UserID	ID unik pengguna, kesan pelanggan baru vs pelanggan tetap, analisis cohort.
Gender	Demografi jantina pengguna, untuk segmentasi pasaran & strategi pemasaran khusus.
Age	Umur pengguna, kenal pasti kelompok umur sasaran (Gen Z, Millennial, dsb).
Region	Lokasi pelanggan (North, South, West, East), analisis pasaran & strategi logistik.
Device	Jenis peranti (Mobile/Desktop), banding tingkah laku & optimasi reka bentuk laman.
Channel	Saluran utama (Social, Email, Direct, Paid Ads), analisis keberkesanan pemasaran.
SourceMedium	Sumber trafik terperinci (contoh: google/organic, instagram/cpc), kira ROI & banding organic vs paid.
Campaign	Nama kempen iklan (contoh: NewArrivals, FlashDeal), ukur keberkesanan kempen.
Pageviews	Bilangan halaman dilihat, indikator tahap minat & engagement pelanggan.
SessionDurationMin	Tempoh lawatan (minit), lebih lama → lebih peluang conversion.
CartAdds	Item ditambah ke troli, ukur niat beli & kesan isu *cart abandonment*.
Purchases	Item yang benar-benar dibeli, asas kiraan jualan & revenue.
Revenue	Nilai jualan (RM/USD), indikator kewangan utama.

AdSpend	Kos iklan untuk sesi, analisis pulangan (ROAS: Return on Ad Spend).
Satisfaction	Skor kepuasan pelanggan (1–5), nilai pengalaman pengguna & servis.
Converted *(clean only)*	Penanda 1 = ada pembelian, 0 = tiada; analisis conversion rate.
AOV (Average Order Value) *(clean only)*	Nilai purata setiap transaksi (Revenue / Purchases), ukur kesan upselling & cross-selling.

Isi Nilai Kosong (Missing Values)

Kolum	Jika Blank / Kosong	Cadangan Isi (Imputasi)	Rasional
SessionID	Tak boleh kosong	**Buang rekod**	ID unik – kalau kosong tak sah.
Date	Tak boleh kosong	**Buang rekod** atau isi dengan tarikh anggaran dari log sistem	Tarikh penting untuk trend analysis.
UserID	Boleh jadi kosong (anonymous user)	Isi **“Guest”** atau kod U0000	Bezakan antara user tetap & guest.
Gender	Kosong/tak pasti	Isi **“Unknown”**	Elak bias → analisis masih boleh jalan.
Age	Kosong	Isi dengan **median umur** (lebih stabil berbanding mean)	Median lebih tahan outlier.
Region	Kosong	Isi dengan **“Unknown”** atau guna IP lokasi jika ada	Elak buang data terlalu banyak.
Device	Kosong	Isi dengan **“Unknown”**	Supaya tetap ada kategori.
Channel	Kosong	Isi dengan **“Direct”** (jika masuk terus tanpa channel) atau **“Unknown”**	Praktikal untuk funnel analisis.
SourceMedium	Kosong	Isi dengan **“Unknown”**	Elak hilang info → boleh asingkan dalam analisis.
Campaign	Kosong	Isi dengan **“Organic”** (jika bukan dari iklan) atau **“Unknown”**	Bezakan kempen vs trafik biasa.
Pageviews	Kosong	Isi **0**	Logiknya kalau

			kosong = tiada page dilihat.
SessionDurationMin	Kosong	Isi **0**	Kosong = user terus keluar (bounce).
CartAdds	Kosong	Isi **0**	Kalau kosong, anggap user tak tambah troli.
Purchases	Kosong	Isi **0**	Kosong = tiada pembelian.
Revenue	Kosong	Isi **0** (jika Purchases=0), atau imputasi ikut Purchases * AOV	Pastikan konsisten dengan pembelian.
AdSpend	Kosong	Isi **0** (jika trafik organic), atau imputasi ikut kos purata kempen	Bezakan organic vs paid.
Satisfaction	Kosong	Isi dengan **median (3)**, atau **“Unknown”**	Elak bias terlalu positif/negatif.
Converted *(clean only)*	Kosong	Auto-kira: IF(Purchases>0,1,0)	Derived column – boleh dikira.
AOV *(clean only)*	Kosong	Auto-kira: IF(Purchases>0, Revenue/Purchases, 0)	Derived column – boleh dikira.

Isi Nilai ?

Kolum	Jika nilai = ?	Cadangan Isi
SessionID / Date	Kalau ? → data tidak sah	**Buang rekod**
UserID	Kalau ? → user tanpa ID	Ganti dengan **“Guest”**
Gender	?	Tukar ke **“Unknown”**
Age	?	Isi dengan **median umur**
Region	?	Tukar ke **“Unknown”**
Device	?	Tukar ke **“Unknown”**
Channel	?	Tukar ke **“Unknown”** atau **“Direct”**
SourceMedium	?	Tukar ke **“Unknown”**
Campaign	?	Tukar ke **“Organic”** (jika bukan iklan) atau **“Unknown”**
Pageviews / Duration / CartAdds	?	Isi **0**
Purchases	?	Isi **0**
Revenue	?	Jika Purchases=0 → isi **0**. Jika Purchases>0 → anggar guna AOV purata
AdSpend	?	Isi **0** (trafik organic)
Satisfaction	?	Isi **3 (median)** atau **“Unknown”**
Converted / AOV	?	Auto-kira (tak perlu isi manual)

Excel Formula Ganti ?

- **Untuk kategori** (contoh Gender): =IF(A2="?", "Unknown", A2)

🔹 **Untuk numerik** (contoh Age, ganti dengan median di sel H1): =IF(A2="?", H1, A2)

🔹 **Untuk transaksi (Purchases, Revenue):** =IF(A2="?", 0, A2)

🔹 **Cari & ganti semua ? sekali gus**:

- **Ctrl + H** → Find: ? → Replace with: Unknown (untuk text) atau 0 (untuk nombor).
  ⚠️ *Tip:* Tandakan kolum tertentu sahaja bila guna Find & Replace, supaya tak kacau semua kolum.

1) Buang duplikasi

Terlibat: SessionID (utama), opsyen tambahan: UserID, Date.

Excel (cara mudah):
Data → Remove Duplicates → tick SessionID (dan kolum lain jika perlu) → OK.

Semak cepat (flag): di kolum bantu IsDuplicate:

=IF(COUNTIF(A:A, A2)>1, "Duplicate","Unique")

Tukar A:A kepada kolum SessionID.

2) Tangani Missing Values

Terlibat: Age, Channel, SourceMedium, Campaign, dll.

Kira median (numerik) untuk imputasi: letak di satu sel bantuan, contohnya H1:

=MEDIAN(IF(ISNUMBER(E2:E2001), E2:E2001))

Tekan **Ctrl+Shift+Enter** *jika Excel lama (array)*. Tukar E:E ikut kolum Age.

Impute Age (jika kosong):

=IF(ISBLANK(E2), H1, E2)

Impute kategori (contoh Channel → "Unknown" jika kosong):

=IF(OR(ISBLANK(H2), H2=""), "Unknown", H2)

Kira nilai paling kerap (mode) untuk kategori (jika mahu guna mode, 365):

=MODE.SNGL(IF(H2:H2001<>"", MATCH(H2:H2001, H2:H2001, 0)))

(lebih praktikal: guna "Unknown" atau mapping kamus seperti langkah #3)

3) Standardisasi Label Kategori (kamus)

Terlibat: Gender, Region, Device, Channel.

Sediakan sheet Dictionary (contoh):

A (Asal)	B (Standard)
F	Female
M	Male
S	South
N	North
Phone	Mobile
IG	Social
newsletter	Email

Guna XLOOKUP (365) / VLOOKUP:

=IFERROR(XLOOKUP(C2, Dictionary!A:A, Dictionary!B:B, C2), C2)

atau

=IFERROR(VLOOKUP(C2, Dictionary!A:B, 2, FALSE), C2)

Ulang untuk setiap kolum kategori yang perlu diseragamkan.

Kemas ejaan & spasi (jika perlu):

=PROPER(TRIM(SUBSTITUTE(C2,CHAR(160)," ")))

4) Betulkan Outlier & Nilai Pelik

Terlibat: Purchases, Revenue, AdSpend, Pageviews, SessionDurationMin.

a) Clamp nilai negatif ke 0 (contoh Purchases):

=MAX(0, N2)

b) Peraturan konsisten revenue-purchase:

- Jika Purchases=0 → Revenue=0

=IF(Purchases2=0, 0, Revenue2)

c) Kenal pasti outlier (IQR):
Kira di sel bantuan:

Q1: =QUARTILE.INC(N2:N2001, 1)

Q3: =QUARTILE.INC(N2:N2001, 3)

IQR:= Q3 - Q1

LowerBound:= Q1 - 1.5*IQR

UpperBound:= Q3 + 1.5*IQR

Flag outlier:

=IF(OR(N2<$LowerBound$, N2>$UpperBound$), "Outlier","OK")

Guna Conditional Formatting untuk highlight.

5) Tukar Format Text → Number/Date

Terlibat: Revenue, AdSpend, Date.

a) Buang simbol & tukar ke nombor (mata wang teks):

=NUMBERVALUE(SUBSTITUTE(SUBSTITUTE(G2,"RM",""),"$",""))

b) Tukar tarikh teks → tarikh Excel:

=DATEVALUE(TEXT(D2,"yyyy-mm-dd"))

atau jika datang dd/mm/yyyy:

=DATEVALUE(TEXT(D2,"dd/mm/yyyy"))

c) Buang koma ribu/kurungan negatif:

=--SUBSTITUTE(SUBSTITUTE(SUBSTITUTE(G2,",",""),"(","-"),")","")

6) Seragamkan Jenis Numerik & Presisi

Terlibat: Pageviews, CartAdds, SessionDurationMin.

Pastikan nombor (tiada teks tersembunyi):

=VALUE(F2)

Bulatkan presisi (2 perpuluhan):

=ROUND(F2, 2)

7) Feature Engineering (kolum baharu)

Terlibat: Converted, AOV.

a) Converted (1 jika beli, 0 jika tidak):

=IF(Purchases2>0, 1, 0)

b) AOV (Average Order Value):

=IFERROR(Revenue2 / Purchases2, 0)

Tambahan idea (jika perlu):

- CTR = CartAdds / Pageviews

=IFERROR(CartAdds2 / Pageviews2, 0)

- RPM (Revenue per Minute) = Revenue / SessionDurationMin

=IFERROR(Revenue2 / SessionDurationMin2, 0)

8) Tapis Rekod Tak Relevan / ‘Zero-Info’

Terlibat: semua metrik engagement.

Flag sesi tanpa makna (tiada aktiviti):

=IF(AND(Pageviews2=0, CartAdds2=0, Purchases2=0, Revenue2=0), "Drop","Keep")

Kemudian tapis Drop.

Aturan minimum aktiviti (opsyen):

=IF(OR(Pageviews2>=1, SessionDurationMin2>=0.5), "Keep","Drop")

9) Standardkan Unit (masa & mata wang)

Terlibat: SessionDuration, Revenue, AdSpend.

a) Saat → minit:

=IFERROR(Seconds2/60, "")

b) USD → RM (contoh kadar di Setup!B2):

=IF(Currency2="USD", Amount2 * Setup!B2, Amount2)

c) Pastikan satu mata wang sahaja pada kolum akhir.
Buat kolum Revenue_RM, AdSpend_RM berasingan jika perlu.

10) Validasi Akhir (quality gates)

Terlibat: semua kolum.

a) Kiraan rekod:
— Pastikan perubahan 2000 → ~1204 (bergantung tapis).

=COUNTA(A2:A2001)

b) Tiada duplikasi SessionID:

=SUMPRODUCT((A2:A2001<>"")/COUNTIF(A2:A2001, A2:A2001))=COUNTA(A2:A2001)

(*TRUE* kalau semua unik)

c) Tiada revenue tanpa purchases:

=COUNTIFS(Purchases2:Purchases2001,0, Revenue2:Revenue2001,">0")

(Hasil sepatutnya 0)

d) Nilai negatif tak wajar:

=COUNTIF(Purchases2:Purchases2001,"<0")

(Hasil sepatutnya 0)

e) Missing penting antara 0:

=SUMPRODUCT(--(ISBLANK(A2:A2001))) 'contoh untuk SessionID

f) Ujian pantas Pivot:
Insert Pivot → Rows: Region, Values: SUM of Revenue → periksa masuk akal.

Bonus: Data Validation (elak salah eja masa input)

- Pilih kolum (contoh Gender) → Data → Data Validation → List → sumber: Female,Male,Unknown.
- Untuk Region: rujuk senarai di Dictionary!B:B.

Bonus: Conditional Formatting

- Highlight sel kosong penting (contoh SessionID, Date):
  Home → Conditional Formatting → New Rule → “Format only cells that contain” → Blanks.