

SQL Data Cleaning Guide PDF Page

SQL Data Cleaning Guide PDF Page: Content + `<head>` (Uncodemy)

FINAL URL: <https://uncodemy.com/sql-data-cleaning-guide-pdf>

Publish this page at exactly this URL (non-www, no trailing slash). Canonical, og:url and sitemap entry must match it.

Prepared: 26 October 2026 · **Primary keyword:** `sql data cleaning guide pdf` · **Search volume:** not checked (keyword tool out of credits); verify in Search Console.

Read first (5 flags):

- Your course PDF title, page count and size were not shared.** Download cards are fill-in slots (`[CONFIRM]`). Card 1 is your course PDF; cards 2 and 3 only if you have more files.
- One unique file per button.** The same course PDF repeated on many pages is fine as a resource, but each page's text must stay unique and useful without the file.
- Free or gated?** Not stated. The page says "Download", not "Free". If a form is used, add a consent checkbox and Privacy Policy link (DPPD Act 2023).
- No made-up stats or testimonials;** no placement or salary claims without a named, dated source.
- Author and reviewer are placeholders;** use real named trainers and one real "Last reviewed" date.

PART 1: Page plan

Item	Decision
Final URL	<code>https://uncodemy.com/sql-data-cleaning-guide-pdf</code>
Intent (anti-cannibalisation)	CLEANING DATA INSIDE SQL. The tool-neutral cleaning process is on the data cleaning guide; Python cleaning is on its own page.
Why not thin	Explainer-style sections, FAQs and cross-links make the page useful even without a download
Schema	WebSite, Organization, Breadcrumb, CollectionPage, FAQPage (only FAQs visible on the page). PDF <code>ItemList</code> provided commented, to enable after titles are final
Share image	<code>https://uncodemy.com/image/logo-og.jpg</code> for now; custom 1200x630 card recommended

PART 2: Page content (copy-paste ready)

SEO Title (38 chars):

SQL Data Cleaning Guide PDF | Uncodemy

Meta Description (136 chars):

Download the SQL data cleaning guide PDF from Uncodemy: NULLs, duplicates, trimming, casting and standardising text with tested queries.

H1: `SQL Data Cleaning Guide PDF Downloads` **Breadcrumb:** `Home > SQL Data Cleaning Guide PDF` **Final URL:** `https://uncodemy.com/sql-data-cleaning-guide-pdf`

Byline: Prepared by the Uncodemy trainer team `[CONFIRM: author name and role]` · Reviewed by `[CONFIRM: reviewer name and role]` · Last reviewed: `[DATE]`

Quick Answer

This page offers Uncodemy's SQL data cleaning guide as a PDF download. When data already lives in a database, cleaning it with SQL is fast and repeatable. The guide shows queries to profile data, find and remove duplicates, handle NULLs, standardise text and dates, and validate the result, using examples that run in SQLite.

Download SQL Data Cleaning Guide PDF

`[CONFIRM: free or gated, and whether details are needed before download]`

PDF 1 (main): `[CONFIRM TITLE: your Data Analytics course PDF]`
`[CONFIRM: one line on what it covers and who it is for]` · `[N]` pages · `[X]` MB · PDF · Updated `[DATE]`
[Download PDF] → `https://uncodemy.com/pdf/sql-data-cleaning-guide.pdf` (*your course PDF, uploaded under this file name*)

PDF 2 (optional): `[CONFIRM TITLE]`
Only if you have a second PDF that genuinely fits this topic. **[Download PDF]** → `https://uncodemy.com/pdf/sql-data-cleaning-guide-2.pdf`

PDF 3 (optional): `[CONFIRM TITLE]`

Same rule. **[Download PDF]** → <https://uncodemy.com/pdf/sql-data-cleaning-guide-3.pdf>

Each button must open its own file. Do not point several buttons at one file. Delete unused optional cards.

Why clean data in SQL

If data is already stored in a database, SQL is often the quickest way to inspect and fix it, because you avoid moving large files around. A saved SQL cleaning query is also repeatable and auditable: anyone can read exactly what was changed. The best practice is not to modify the raw table but to create a cleaned view or a new table derived from it, so the original stays intact. This guide uses a messy example table called `raw_customers` with columns `customer_id`, `name`, `email`, `city`, `signup_date` and `monthly_spend`, which contains duplicates, inconsistent text, NULLs and mixed date styles.

A cleaning workflow in SQL

1. Profile the table: row counts, NULLs per column and distinct values.
2. Find duplicates on the columns that define a unique record.
3. Standardise text with `TRIM`, `UPPER` or `LOWER` and `REPLACE`.
4. Handle NULLs with `COALESCE` or an explicit label.
5. Convert types and dates explicitly.
6. Write the cleaned result to a view or table and validate counts and totals.

Profiling the table

Counts rows, missing values and distinct cities in one query.

```
SELECT COUNT(*)                AS total_rows,
       COUNT(DISTINCT customer_id) AS distinct_ids,
       SUM(CASE WHEN email IS NULL THEN 1 ELSE 0 END) AS missing_email,
       SUM(CASE WHEN monthly_spend IS NULL THEN 1 ELSE 0 END) AS missing_spend,
       COUNT(DISTINCT city)      AS distinct_city_values
FROM raw_customers;
```

Finding duplicates

Shows customer ids that appear more than once, and then ranks rows within each id so that you can keep one.

```
-- which ids are duplicated?
SELECT customer_id, COUNT(*) AS copies
FROM raw_customers
GROUP BY customer_id
HAVING COUNT(*) > 1;

-- keep the row with the latest signup_date for each id
SELECT *
FROM (
  SELECT r.*,
         ROW_NUMBER() OVER (PARTITION BY customer_id ORDER BY signup_date DESC) AS rn
  FROM raw_customers r
)
WHERE rn = 1;
```

Standardising text and handling NULLs

Trim spaces, fix case, standardise a misspelling and give missing values a default or a clear label.

```
SELECT customer_id,
       TRIM(name) AS name_clean,
       LOWER(TRIM(email)) AS email_clean,
       CASE UPPER(TRIM(city))
         WHEN 'NEW DELHI' THEN 'Delhi'
         WHEN 'DELHI'     THEN 'Delhi'
         ELSE SUBSTR(UPPER(TRIM(city)), 1, 1) || LOWER(SUBSTR(TRIM(city), 2))
       END AS city_clean,
       COALESCE(monthly_spend, 0) AS spend_filled,
       COALESCE(NULLIF(TRIM(email), ''), 'unknown') AS email_or_unknown
FROM raw_customers;
```

Useful cleaning functions (names vary slightly by database)

Task	Function idea
Remove spaces	TRIM, LTRIM, RTRIM
Change case	UPPER, LOWER
Replace text	REPLACE
Part of a string	SUBSTR or SUBSTRING, LEFT, RIGHT
Default for NULL	COALESCE
Turn empty text into NULL	NULLIF(column, '')
Conditional fixes	CASE WHEN ... THEN ... END
Change type	CAST(column AS type)
Detect pattern	LIKE; regular expressions where supported

Dates and numbers

Dates are a frequent source of trouble: text values in different styles, impossible dates and time zones. Where possible store dates in a proper date type and write them in ISO form, year-month-day, which sorts correctly as text too. When converting, test the conversion on a sample and check for rows that fail. For numbers, watch for values stored as text, currency symbols and commas. CAST a cleaned string to a number, but first remove characters that do not belong, and report rows that could not be converted instead of hiding them.

Creating a cleaned view and validating it

A view keeps the raw data intact. The check afterwards compares counts with the raw table.

```
CREATE VIEW clean_customers AS
SELECT customer_id,
       TRIM(name) AS name,
       LOWER(TRIM(email)) AS email,
       TRIM(city) AS city,
       COALESCE(monthly_spend, 0) AS monthly_spend
FROM (
  SELECT r.*, ROW_NUMBER() OVER (PARTITION BY customer_id ORDER BY signup_date DESC) AS rn
  FROM raw_customers r
)
WHERE rn = 1;

SELECT (SELECT COUNT(*) FROM raw_customers) AS raw_rows,
       (SELECT COUNT(*) FROM clean_customers) AS clean_rows,
       (SELECT COUNT(DISTINCT customer_id) FROM raw_customers) AS raw_distinct_ids;
```

Common SQL cleaning mistakes

- **Updating the raw table in place** and losing the original.
- **Deleting duplicates without deciding which row to keep.**
- ****Comparing with NULL using = ****instead of IS NULL.
- **Filling NULLs with zero** when a missing value should stay unknown.
- **Not validating** row counts and totals afterwards.

Document your rules

Write a short note or comments in the SQL for every rule: what the problem was, how you fixed it and how many rows it touched. The count is easy to get with a query before and after. These notes turn a one-off fix into reusable knowledge, and they make it easy for a colleague to review your work.

Data cleaning checklist for SQL

Check	Query idea
Count rows and distinct keys	SELECT COUNT(*), COUNT(DISTINCT id) FROM table
Find NULLs per column	SUM(CASE WHEN col IS NULL THEN 1 ELSE 0 END)
Find duplicates	GROUP BY key HAVING COUNT(*) > 1
Find leading and trailing spaces	WHERE col <> TRIM(col)
Check value ranges	MIN, MAX and ordered samples
Check category spellings	SELECT DISTINCT col ORDER BY col

Keep a cleaning log

- Record each rule, such as treat blank strings as NULL.
- Note how many rows each rule changed.
- Keep the raw table unchanged and write cleaned data into a new table or view.

Related Uncodemy resources

- [Data Analytics Data Cleaning Guide PDF](#)
- [Python Data Cleaning PDF](#)
- [SQL Subqueries Guide PDF](#)
- [SQL for Data Analytics Guide PDF](#)

Learn with a trainer [use only the line that is true]

- **Data Analytics course:** <https://uncodemy.com/course/data-analytics-training-course-in-noida>. URL verified on the live site.

(One soft call to action: course link or free demo. No pop-up.)

About these PDFs

These PDFs are study material for learning and revision. They are not a certificate, a degree or a promise of a job, and technology changes, so check official documentation for the latest details. Spot an error? Write to [CONFIRM: [feedback_email](#)] and the trainer team will review it.

Frequently Asked Questions

How do I remove duplicates in SQL?

Decide which columns define a unique record, then use ROW_NUMBER over a partition by those columns and keep rn equal to 1, or group and aggregate. Avoid deleting from the raw table without a backup.

How do I handle NULL values in SQL?

Use IS NULL to find them, COALESCE to supply a default, and NULLIF to turn empty strings into NULL. Choose defaults carefully.

How do I standardise text in SQL?

Use TRIM to remove spaces, UPPER or LOWER to normalise case, REPLACE for fixed substitutions and CASE for mapping variants to one label.

Should I clean data in SQL or Python?

If data is in a database and rules are simple, SQL is efficient and repeatable. Python suits complex logic or files outside a database.

Why create a view for cleaned data?

A view leaves the raw table untouched and shows exactly what rules you applied.

Are these PDFs free to download?

[CONFIRM: state clearly whether the PDFs are free, or whether a short form is needed first.]

The last FAQ contains [CONFIRM], so it is NOT in the FAQ schema. Fill it and keep it on the page, or delete it.

PART 3: <head> section (copy-paste ready)

Final URL: <https://uncodemy.com/sql-data-cleaning-guide-pdf>

```
<!-- =====
UNCODEMY <head> | FINAL URL: https://uncodemy.com/sql-data-cleaning-guide-pdf
Prepared 26 October 2026. Paste into the page <head>; keep your site-wide favicon, fonts, analytics.
BEFORE GO-LIVE: (1) set dateModified in the JSON-LD to the real publish date;
(2) replace og:image with a 1200x630 share card when available;
(3) once PDF titles are final, uncomment the ItemList block at the bottom.
===== -->

<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1">

<title>SQL Data Cleaning Guide PDF | Uncodemy</title>
<meta name="description" content="Download the SQL data cleaning guide PDF from Uncodemy: NULLs, duplicates, trimming, casting and standardising text with tes
queries.">
<link rel="canonical" href="https://uncodemy.com/sql-data-cleaning-guide-pdf">
<meta name="robots" content="index, follow, max-snippet:-1, max-image-preview:large, max-video-preview:-1">
<meta name="author" content="Uncodemy">
<meta name="publisher" content="Uncodemy Edutech Pvt. Ltd.">
<meta name="language" content="en-IN">
<link rel="alternate" hreflang="en-in" href="https://uncodemy.com/sql-data-cleaning-guide-pdf">
<link rel="alternate" hreflang="x-default" href="https://uncodemy.com/sql-data-cleaning-guide-pdf">

<!-- Open Graph -->
<meta property="og:type" content="website">
<meta property="og:site_name" content="Uncodemy">
<meta property="og:locale" content="en-IN">
<meta property="og:url" content="https://uncodemy.com/sql-data-cleaning-guide-pdf">
<meta property="og:title" content="SQL Data Cleaning Guide PDF | Uncodemy">
<meta property="og:description" content="Download the SQL data cleaning guide PDF from Uncodemy: NULLs, duplicates, trimming, casting and standardising text w
tested queries.">
<meta property="og:image" content="https://uncodemy.com/image/logo-og.jpg">
<meta property="og:image:alt" content="Uncodemy SQL Data Cleaning Guide PDF study material">

<!-- X (Twitter) -->
<meta name="twitter:card" content="summary_large_image">
<meta name="twitter:site" content="@uncodemy">
<meta name="twitter:title" content="SQL Data Cleaning Guide PDF | Uncodemy">
<meta name="twitter:description" content="Download the SQL data cleaning guide PDF from Uncodemy: NULLs, duplicates, trimming, casting and standardising text
tested queries.">
<meta name="twitter:image" content="https://uncodemy.com/image/logo-og.jpg">
<meta name="twitter:image:alt" content="Uncodemy SQL Data Cleaning Guide PDF study material">

<!-- Structured data: one @graph (WebSite, Organization, Breadcrumb, CollectionPage, FAQPage) -->
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "WebSite",
      "@id": "https://uncodemy.com/#website",
      "url": "https://uncodemy.com/",
      "name": "Uncodemy",
      "publisher": {
        "@id": "https://uncodemy.com/#organization"
      },
      "inLanguage": "en-IN"
    },
    {
      "@type": "EducationalOrganization",
      "@id": "https://uncodemy.com/#organization",
      "name": "Uncodemy",
      "legalName": "Uncodemy Edutech Pvt. Ltd.",
      "url": "https://uncodemy.com/",
      "logo": {
        "@type": "ImageObject",
        "url": "https://uncodemy.com/uc-layout/img/Logo.webp"
      },
      "email": "info@uncodemy.com",
      "identifier": {
        "@type": "PropertyValue",
        "name": "CIN",
        "value": "U85499UP2024PTC195902"
      },
      "contactPoint": [
        {
          "@type": "ContactPoint",
          "telephone": "+91-9818366550",
          "contactType": "admissions",
          "areaServed": "IN",
          "availableLanguage": [
            "en",
            "hi"
          ]
        }
      ]
    }
  ]
}
```

```

    }
  },
  "sameAs": [
    "https://x.com/uncodemy",
    "https://www.facebook.com/uncodemyofficial/",
    "https://www.linkedin.com/company/uncodemy/",
    "https://www.instagram.com/uncodemyofficial/"
  ]
},
{
  "@type": "BreadcrumbList",
  "@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#breadcrumb",
  "itemListElement": [
    {
      "@type": "ListItem",
      "position": 1,
      "name": "Home",
      "item": "https://uncodemy.com/"
    },
    {
      "@type": "ListItem",
      "position": 2,
      "name": "SQL Data Cleaning Guide PDF",
      "item": "https://uncodemy.com/sql-data-cleaning-guide-pdf"
    }
  ]
},
{
  "@type": "CollectionPage",
  "@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#webpage",
  "url": "https://uncodemy.com/sql-data-cleaning-guide-pdf",
  "name": "SQL Data Cleaning Guide PDF Downloads",
  "description": "Download the SQL data cleaning guide PDF from Uncodemy: NULLs, duplicates, trimming, casting and standardising text with tested queries.",
  "inLanguage": "en-IN",
  "isPartOf": {
    "@id": "https://uncodemy.com/#website"
  },
  "publisher": {
    "@id": "https://uncodemy.com/#organization"
  },
  "breadcrumb": {
    "@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#breadcrumb"
  },
  "about": {
    "@type": "Thing",
    "name": "SQL Data Cleaning Guide"
  },
  "primaryImageOfPage": {
    "@type": "ImageObject",
    "url": "https://uncodemy.com/image/logo-og.jpg"
  },
  "dateModified": "2026-10-26"
},
{
  "@type": "FAQPage",
  "@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#faq",
  "inLanguage": "en-IN",
  "isPartOf": {
    "@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#webpage"
  },
  "mainEntity": [
    {
      "@type": "Question",
      "name": "How do I remove duplicates in SQL?",
      "acceptedAnswer": {
        "@type": "Answer",
        "text": "Decide which columns define a unique record, then use ROW_NUMBER over a partition by those columns and keep rn equal to 1, or group and aggregate. Avoid deleting from the raw table without a backup."
      }
    },
    {
      "@type": "Question",
      "name": "How do I handle NULL values in SQL?",
      "acceptedAnswer": {
        "@type": "Answer",
        "text": "Use IS NULL to find them, COALESCE to supply a default, and NULLIF to turn empty strings into NULL. Choose defaults carefully."
      }
    },
    {
      "@type": "Question",
      "name": "How do I standardise text in SQL?",
      "acceptedAnswer": {
        "@type": "Answer",
        "text": "Use TRIM to remove spaces, UPPER or LOWER to normalise case, REPLACE for fixed substitutions and CASE for mapping variants to one label."
      }
    },
    {
      "@type": "Question",
      "name": "Should I clean data in SQL or Python?",
      "acceptedAnswer": {
        "@type": "Answer",
        "text": "If data is in a database and rules are simple, SQL is efficient and repeatable. Python suits complex logic or files outside a database."
      }
    },
    {
      "@type": "Question",
      "name": "Why create a view for cleaned data?",
      "acceptedAnswer": {
        "@type": "Answer",
        "text": "A view leaves the raw table untouched and shows exactly what rules you applied."
      }
    }
  ]
}
}
}
</script>

<!-- ADD AFTER THE PDF TITLES ARE FINAL (remove this wrapper, fill every [...] value):
<script type="application/ld+json">
{

```

```

"@context": "https://schema.org",
"@type": "ItemList",
"@id": "https://uncodemy.com/sql-data-cleaning-guide-pdf#pdfs",
"name": "SQL Data Cleaning Guide PDF downloads",
"numberOfItems": 3,
"itemListElement": [
  {
    "@type": "ListItem",
    "position": 1,
    "item": {
      "@type": "DigitalDocument",
      "name": "[PDF 1 TITLE]",
      "description": "[PDF 1 ONE-LINE DESCRIPTION]",
      "encodingFormat": "application/pdf",
      "url": "https://uncodemy.com/pdf/sql-data-cleaning-guide.pdf",
      "inLanguage": "en",
      "publisher": {
        "@id": "https://uncodemy.com/#organization"
      }
    }
  },
  {
    "@type": "ListItem",
    "position": 2,
    "item": {
      "@type": "DigitalDocument",
      "name": "[PDF 2 TITLE]",
      "description": "[PDF 2 ONE-LINE DESCRIPTION]",
      "encodingFormat": "application/pdf",
      "url": "https://uncodemy.com/pdf/sql-data-cleaning-guide-2.pdf",
      "inLanguage": "en",
      "publisher": {
        "@id": "https://uncodemy.com/#organization"
      }
    }
  },
  {
    "@type": "ListItem",
    "position": 3,
    "item": {
      "@type": "DigitalDocument",
      "name": "[PDF 3 TITLE]",
      "description": "[PDF 3 ONE-LINE DESCRIPTION]",
      "encodingFormat": "application/pdf",
      "url": "https://uncodemy.com/pdf/sql-data-cleaning-guide-3.pdf",
      "inLanguage": "en",
      "publisher": {
        "@id": "https://uncodemy.com/#organization"
      }
    }
  }
]
}
</script>
-->

```

DEVELOPER CHECKLIST

- ☐ Final URL live at `https://uncodemy.com/sql-data-cleaning-guide-pdf`; canonical, og:url and sitemap identical.
- ☐ Each button: `<a href="/pdf/sql-data-cleaning-guide.pdf" download type="application/pdf">`, its own file, size shown.
- ☐ If a form precedes download: consent checkbox, Privacy Policy link, nothing pre-ticked.
- ☐ One `<h1>`; real HTML tables; code samples in `<pre><code>`.
- ☐ Remove any `<meta name="keywords">` from the template.
- ☐ Test in Google Rich Results Test and validator.schema.org after publishing.
- ☐ Add to sitemap and any PDF/resources hub.

STILL NEEDED FROM YOU

#	Item
1	Course PDF title, description, page count, size, and the file URL for each page
2	Free or gated (and form fields if gated)
3	Author, reviewer, publish and last-reviewed dates
4	Course URL for the CTA (lines marked <code>[CONFIRM]</code>)
5	Feedback email
6	Confirmation that Uncodemy owns the rights to every PDF