

FINAL URL: <https://uncodemy.com/data-analytics-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:** data analytics 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 (DPPA 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	https://uncodemy.com/data-analytics-data-cleaning-guide-pdf
Intent (anti-cannibalisation)	DATA CLEANING PROCESS across tools. SQL-only cleaning is on the SQL data cleaning page; Python-only cleaning is on the Python data cleaning 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	https://uncodemy.com/image/logo-og.jpg for now; custom 1200x630 card recommended

PART 2: Page content (copy-paste ready)

SEO Title (49 chars):

Data Analytics Data Cleaning Guide PDF | Uncodemy

Meta Description (152 chars):

Download the data cleaning guide PDF from Uncodemy: missing values, duplicates, outliers, formats, a cleaning log and examples in Excel, SQL and Python.

H1: Data Analytics Data Cleaning Guide PDF Downloads **Breadcrumb:** Home > Data Analytics Data Cleaning Guide PDF **Final URL:** <https://uncodemy.com/data-analytics-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 data cleaning guide as a PDF download. Data cleaning is the work of finding and fixing problems such as missing values, duplicates, wrong data types and inconsistent labels so that analysis is trustworthy. The guide explains a general process and shows examples in Excel, SQL and Python.

Download Data Analytics 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/data-analytics-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/data-analytics-data-cleaning-guide-2.pdf>

PDF 3 (optional): [CONFIRM TITLE]

Same rule. **[Download PDF]** → <https://uncodemy.com/pdf/data-analytics-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 data cleaning matters

Analysts often spend more of their time preparing data than analysing it, and for good reason. A chart built on dirty data looks as convincing as one built on clean data, but its message may be wrong. Duplicate orders inflate revenue, text dates sort in the wrong order, and a stray space in a city name splits one city into two. Cleaning is not a boring chore before the real work, it is part of the real work, because the quality of every conclusion depends on it. A well-cleaned dataset also saves time later, since the same fixes do not need to be repeated.

A practical data cleaning workflow

- 1. Keep an untouched copy of the raw data and never overwrite it.
- 2. Profile the data: size, types, missing values, duplicates and value ranges.
- 3. Decide the rule for each problem and write it down.
- 4. Apply fixes in a repeatable way, using a script or query instead of manual edits.
- 5. Validate: recount rows, recheck totals and compare with the source.
- 6. Document what changed, how many rows were affected and why.

Common data problems and typical fixes

Problem	Example	Typical fix
Missing values	Blank phone or null amount	Fill with a justified value, mark as unknown or exclude, depending on the cause
Duplicates	Same order loaded twice	Remove exact copies; investigate near-duplicates using a key
Wrong data type	Date stored as text '12/03/2026'	Convert with an explicit date format
Inconsistent text	'Noida', 'noida ', 'NOIDA'	Trim spaces and standardise case
Invalid values	Age of -5 or 250	Check the source; correct or set to missing
Mixed units	Weights in kg and grams	Convert to one unit
Structural issues	Several values packed in one column	Split into separate columns

Handling missing values with care

There is no single correct way to treat missing data, because the right choice depends on why the values are missing. If a small share of rows is missing in a column you do not need, dropping them may be fine. If the column matters, consider filling with the median for skewed numbers, the most common category for labels, or a clear label such as Unknown so that the gap stays visible. Never fill silently without a note, and avoid filling values that should stay missing, such as a refund date for an order that was never refunded. Always report how many values you changed so readers can judge the impact.

Cleaning in pandas

A short pipeline: trim and standardise text, convert dates, remove duplicates and handle missing numbers. Column names are examples.

```
import pandas as pd

df = pd.read_csv("customers.csv")

df["city"] = df["city"].str.strip().str.title()
df["signup_date"] = pd.to_datetime(df["signup_date"], format="%d/%m/%Y", errors="coerce")
df = df.drop_duplicates(subset=["customer_id"])
df["monthly_spend"] = df["monthly_spend"].fillna(df["monthly_spend"].median())

print(df.isna().sum())
```

Cleaning in SQL

The same ideas with SQL functions, on a customers table that has city, monthly_spend and email columns for illustration. Function names vary slightly between databases; the syntax is similar in SQLite, MySQL and PostgreSQL.

```
SELECT customer_id,
       TRIM(UPPER(city))           AS city_clean,
       COALESCE(monthly_spend, 0) AS monthly_spend_filled,
       LOWER(TRIM(email))         AS email_clean
FROM customers
WHERE customer_id IS NOT NULL;
```

Cleaning in Excel

- Use TRIM and CLEAN to remove extra spaces and hidden characters.
- Use Text to Columns or Flash Fill to split combined values.
- Use Remove Duplicates on the right key columns, after backing up the data.
- Use Power Query for repeatable steps: it records each transformation so that it can be refreshed.
- Use Data Validation to stop bad values entering new data.

Validation after cleaning

Cleaning is only finished when you have checked the result. Compare row counts before and after and explain every difference. Recalculate key

totals and confirm they still match the source system or an approved report. Re-run your data quality checks to make sure no missing values or duplicates remain unexpectedly. Spot-check a few records by hand against the original. If anything does not reconcile, find out why before building any report on top of the data.

Common cleaning mistakes

- **Overwriting raw data**, leaving no way back.
- **Deleting rows too eagerly** and biasing the results.
- **Manual edits that cannot be repeated** when new data arrives.
- **Changing values without a note**, so nobody knows what is original.
- **Skipping validation** and discovering errors in the final report.

Cleaning log: what to record

Item	Example entry
Issue found	312 rows with missing city
Rule applied	Filled from the pin-code lookup table; remaining 18 rows labelled Unknown
Rows affected	312 of 48,200 (0.6 percent)
Reason	City is needed for the regional sales report
Date and person	Date of the change and the analyst's name

Automating repeated cleaning

If the same data arrives every week, turn your cleaning steps into a script, a saved query or a Power Query flow instead of repeating manual edits. Automation removes human error, makes the work reproducible and frees time for analysis. Keep the script under version control and add a few automatic checks, for example a warning when the share of missing values suddenly rises, because a sudden change often signals a problem in the source system rather than in your code.

Related Uncodemy resources

- [Data Analytics EDA Guide PDF](#)
- [SQL Data Cleaning Guide PDF](#)
- [Python Data Cleaning PDF](#)
- [Real-World Datasets for Data Analytics 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

What is data cleaning?

Data cleaning is the process of detecting and correcting problems in a dataset, such as missing values, duplicates, wrong types and inconsistent labels, so that analysis is reliable.

How do you handle missing values?

It depends on why they are missing. Options include dropping rows, filling with a justified value such as the median, or labelling the gap as unknown. Always document the choice.

What is the best tool for data cleaning?

Excel with Power Query suits small tasks, SQL suits data in databases and Python with pandas suits larger or repeatable work. The best choice depends on the data size and your workflow.

How do I remove duplicates safely?

Decide which columns define a unique record, back up the data, remove exact duplicates first and investigate near-duplicates before deleting them.

Why should I keep the raw data untouched?

A pristine copy lets you repeat or correct the cleaning steps and prove where each change came from.

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/data-analytics-data-cleaning-guide-pdf>

```

<!-- =====
UNCODEMY <head> | FINAL URL: https://uncodemy.com/data-analytics-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>Data Analytics Data Cleaning Guide PDF | Uncodemy</title>
<meta name="description" content="Download the data cleaning guide PDF from Uncodemy: missing values, duplicates, outliers, formats, a cleaning log and examples in Excel, SQL and Python.">
<link rel="canonical" href="https://uncodemy.com/data-analytics-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/data-analytics-data-cleaning-guide-pdf">
<link rel="alternate" hreflang="x-default" href="https://uncodemy.com/data-analytics-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/data-analytics-data-cleaning-guide-pdf">
<meta property="og:title" content="Data Analytics Data Cleaning Guide PDF | Uncodemy">
<meta property="og:description" content="Download the data cleaning guide PDF from Uncodemy: missing values, duplicates, outliers, formats, a cleaning log and examples in Excel, SQL and Python.">
<meta property="og:image" content="https://uncodemy.com/image/logo-og.jpg">
<meta property="og:image:alt" content="Uncodemy Data Analytics 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="Data Analytics Data Cleaning Guide PDF | Uncodemy">
<meta name="twitter:description" content="Download the data cleaning guide PDF from Uncodemy: missing values, duplicates, outliers, formats, a cleaning log and examples in Excel, SQL and Python.">
<meta name="twitter:image" content="https://uncodemy.com/image/logo-og.jpg">
<meta name="twitter:image:alt" content="Uncodemy Data Analytics 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/data-analytics-data-cleaning-guide-pdf#breadcrumb",
      "itemListElement": [
        {
          "@type": "ListItem",
          "position": 1,
          "name": "Home",
          "item": "https://uncodemy.com/"
        },
        {
          "@type": "ListItem",
          "position": 2,
          "name": "Data Analytics Data Cleaning Guide PDF",
          "item": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf"
        }
      ]
    },
    {
      "@type": "CollectionPage",

```

```

    "id": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf#webpage",
    "url": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf",
    "name": "Data Analytics Data Cleaning Guide PDF Downloads",
    "description": "Download the data cleaning guide PDF from Uncodemy: missing values, duplicates, outliers, formats, a cleaning log and examples in Excel,
and Python.",
    "inLanguage": "en-IN",
    "isPartOf": {
      "@id": "https://uncodemy.com/#website"
    },
    "publisher": {
      "@id": "https://uncodemy.com/#organization"
    },
    "breadcrumb": {
      "@id": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf#breadcrumb"
    },
    "about": {
      "@type": "Thing",
      "name": "Data Analytics Data Cleaning Guide"
    },
    "primaryImageOfPage": {
      "@type": "ImageObject",
      "url": "https://uncodemy.com/image/logo-og.jpg"
    },
    "dateModified": "2026-10-26"
  },
  {
    "@type": "FAQPage",
    "id": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf#faq",
    "inLanguage": "en-IN",
    "isPartOf": {
      "@id": "https://uncodemy.com/data-analytics-data-cleaning-guide-pdf#webpage"
    },
    "mainEntity": [
      {
        "@type": "Question",
        "name": "What is data cleaning?",
        "acceptedAnswer": {
          "@type": "Answer",
          "text": "Data cleaning is the process of detecting and correcting problems in a dataset, such as missing values, duplicates, wrong types and
inconsistent labels, so that analysis is reliable."
        }
      },
      {
        "@type": "Question",
        "name": "How do you handle missing values?",
        "acceptedAnswer": {
          "@type": "Answer",
          "text": "It depends on why they are missing. Options include dropping rows, filling with a justified value such as the median, or labelling the ga
unknown. Always document the choice."
        }
      },
      {
        "@type": "Question",
        "name": "What is the best tool for data cleaning?",
        "acceptedAnswer": {
          "@type": "Answer",
          "text": "Excel with Power Query suits small tasks, SQL suits data in databases and Python with pandas suits larger or repeatable work. The best ch
depends on the data size and your workflow."
        }
      },
      {
        "@type": "Question",
        "name": "How do I remove duplicates safely?",
        "acceptedAnswer": {
          "@type": "Answer",
          "text": "Decide which columns define a unique record, back up the data, remove exact duplicates first and investigate near-duplicates before delet
them."
        }
      },
      {
        "@type": "Question",
        "name": "Why should I keep the raw data untouched?",
        "acceptedAnswer": {
          "@type": "Answer",
          "text": "A pristine copy lets you repeat or correct the cleaning steps and prove where each change came from."
        }
      }
    ]
  }
]
}
</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/data-analytics-data-cleaning-guide-pdf#pdfs",
  "name": "Data Analytics 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/data-analytics-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/data-analytics-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/data-analytics-data-cleaning-guide-3.pdf",
        "inLanguage": "en",
        "publisher": {
          "@id": "https://uncodemy.com/#organization"
        }
      }
    }
  ]
}
</script>
-->

```

DEVELOPER CHECKLIST

- ☐ Final URL live at <https://uncodemy.com/data-analytics-data-cleaning-guide-pdf>; canonical, og:url and sitemap identical.
- ☐ Each button: `<a href="/pdf/data-analytics-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