An introduction to data and data analytics

The data practitioner's guide

Attributes of a data professional

Data professionals have many attributes that make them **effective** in **working with data**, **extracting insights**, and **deriving value** from data-driven approaches.

Analytical thinking

Tend to possess strong analytical skills and are capable of **breaking down complex problems into manageable components**. Can **identify patterns**, **trends**, and **insights** within data and use **critical thinking** to derive meaningful conclusions.

Curiosity and continuous learning

Tend to have a **natural curiosity** and a **passion** for **exploring** and **understanding data.** Eager to learn new techniques, tools, and methodologies to enhance their data analysis skills. Stay updated with the latest trends and advancements.

Technical proficiency

Tend to have a solid **understanding of data-related technologies**, **programming languages**, and **tools**. Proficient in data manipulation, analysis, and visualization.

Attributes of a data professional

Attention to detail

Pays attention to detail when working with data. They ensure **data accuracy**, **completeness**, and **quality**. They are **meticulous** in data cleaning, pre-processing, and validation to minimize errors and biases that could impact analysis outcomes.

Collaboration and teamwork

Collaborate effectively with colleagues from different backgrounds, including stakeholders. They can **work together** to identify project goals, share insights, and contribute to data-driven decision-making.

Communication and storytelling

Skilled communicators who can effectively **convey complex data concepts**. Can translate data insights into clear, actionable recommendations and present findings through **effective use of visualizations** and **storytelling**.

While the specific combination and emphasis of attributes may vary depending on the role and organization, these attributes are **commonly sought after** in data professionals, such as data analysts, data scientists, and data engineers.

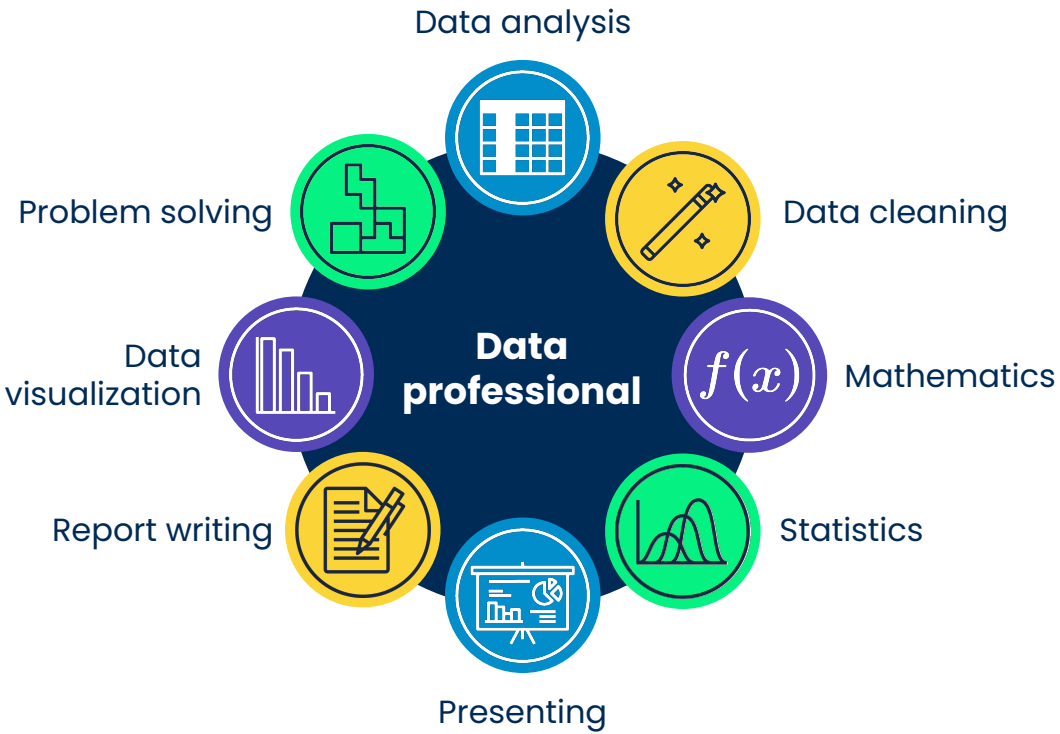
Since these **attributes are in high demand**, it is valuable for data professionals to **develop these attributes**.

Key data skills

Data has become an integral part of daily life and many of these skills are **critical** but **transferable** in many careers – we are all **data users**.

As data professionals, there are various skills we should have.

Depending on the specific role and responsibilities, **additional skills** such as machine learning, domain knowledge, database systems, and programming may be required.



Key data tools

Data professionals use several **different tools** to leverage their data skills and knowledge, depending on their role. Although we don't need to be proficient in all of them, it's important to know what the toolkit of a data professional could look like.

Spreadsheets:
Google Sheets, Microsoft Excel, Numbers, LibreOffice Calc

Database management systems:
MySQL, PostgreSQL, Oracle, Microsoft SQL Server, MongoDB, SQLite, Apache Cassandra

Data visualization and dashboarding tools:
Microsoft Power BI, Tableau, QlikView

Programming languages:
SQL, Python, R

Integrated development environments (IDEs):
Visual Studio Code, Jupyter Notebook, PyCharm, RStudio

Version control systems:
Git, Mercurial, Subversion

Cloud computing platforms:
Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure

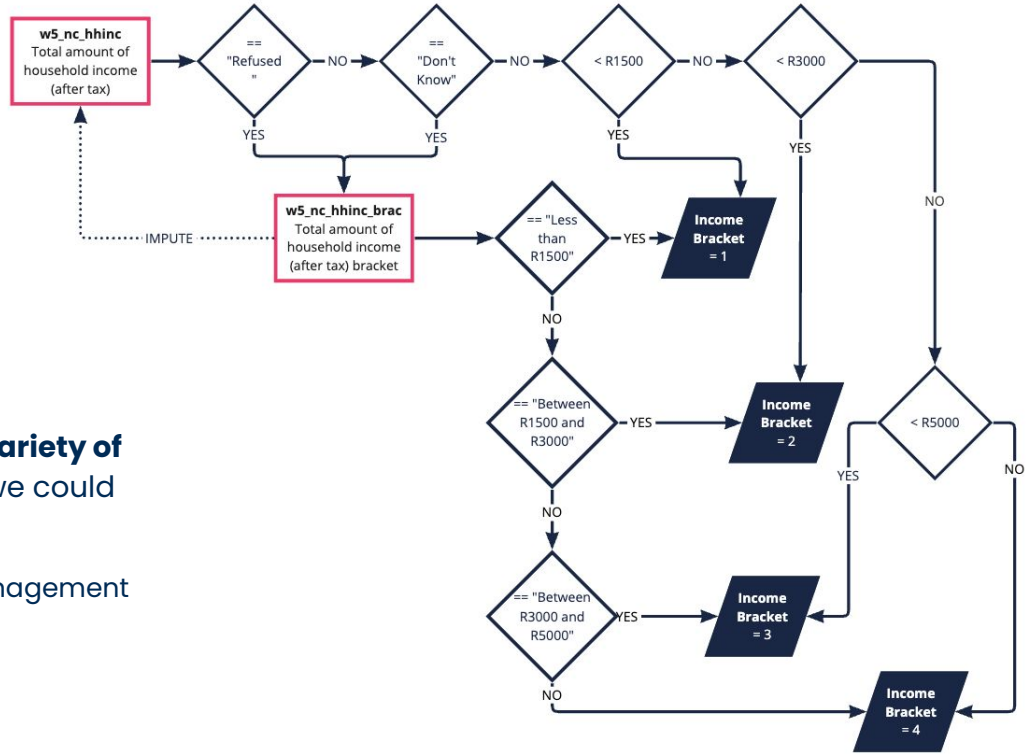
Machine learning frameworks:
scikit-learn (Python), TensorFlow, Keras

And many more.

Problem solving

Problem solving is a critical skill for all data users because it provides **structured**, **transferable**, and **logical** ways of approaching data problems.

Depending on the problem we need to solve, we use critical thinking "tools" such as **logic trees** and **flowcharts** to understand the problem and plan a solution.



When we understand the problem, we apply a **variety of skills and tools to solve** the problem. The tools we could use include:

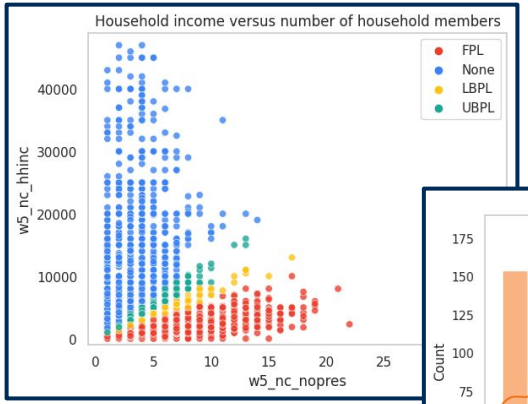
- Spreadsheets
- Database management systems
- Data visualization and dashboarding tools
- Programming languages

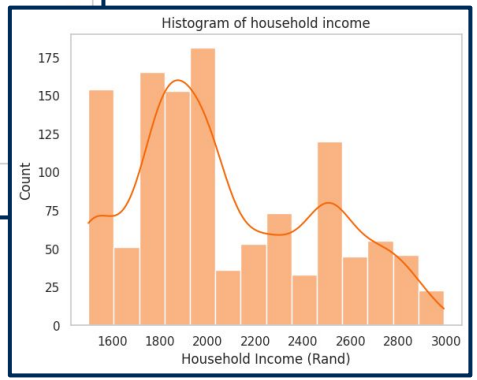
Data preparation and analysis

Data preparation (cleaning) and **analysis** are critical skills for data professionals as they **enable effective decision-making**, **uncovering insights**, **identifying patterns**, and **detecting trends**.

The information extracted with these skills can **drive business strategy**, **optimization**, **problem solving**, and **innovation**.

Column	w5_nc_hhinc_brac	w5_nc_hhinc
Count non NaN	1561	5862
Count after drop "Refused" and "Don't Know" in **w5_nc_hhinc_brac**	1167	5468





The tools we can use:

- Spreadsheets
- Database management systems
- Data visualization and dashboarding tools
- Programming languages

Mathematics and statistics

Although **mathematics** and **statistics** may feel overwhelming to many, understanding the **fundamentals** of these fields makes **using data easier**.

Mathematics and statistics are **critical for data analysis**, **problem solving**, and **machine learning**.

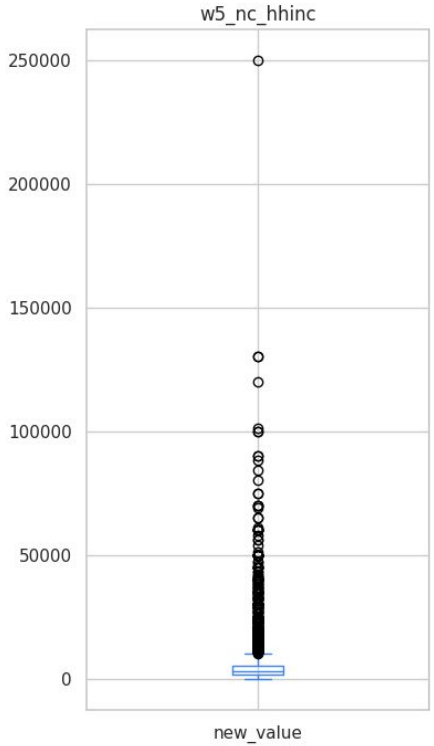
We can apply this knowledge and skills across various tools, including:

- Spreadsheets
- Database management systems
- Machine learning frameworks
- Programming languages

Column	w5_nc_hhinc
Count	4301.00
Mean	5653.48
Standard deviation	10057.09
Minimum	0.00
25%	1600.00
50%	3000.00
75%	5000.00
Maximum	250000.00

IQR = 3400
Higher outliers > R 10100

The number of outliers: 508 households (11.81%)



Communication and storytelling

Strong communication skills are essential for data practitioners to effectively share insights and findings to non-technical stakeholders.

The ability to **translate complex data concepts** into clear meaningful narratives helps organizations to make data-driven decisions.

We use **data visualization**, **dashboards**, **reports**, and **presentations** to communicate data insights.

The tools we can use:

- Spreadsheets
- Data visualization and dashboarding tools
- Programming languages