

The 6 Dimensions of Data Quality

ACCURACY

How closely the data reflects the **true values** or **facts**

COMPLETENESS

All necessary data is **present** and **accounted for** in a dataset

TIMELINESS

Reflects how **up-to-date** the data is

CONSISTENCY

Data is **uniform** across different datasets and systems

UNIQUENESS

Each data entry is **distinct** and **not repeated** in the dataset

VALIDITY

The data is **correct** and **appropriate** for its intended use



Office of Data
Governance
and Analytics

ACCURACY

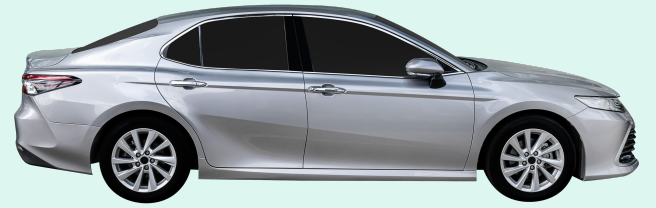
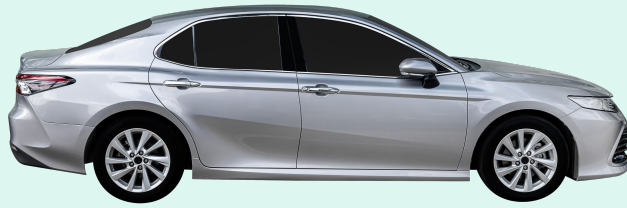
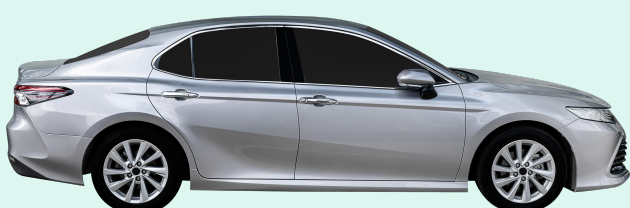
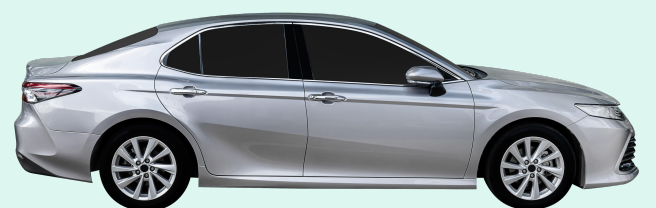
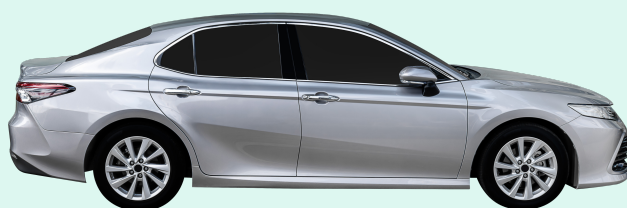
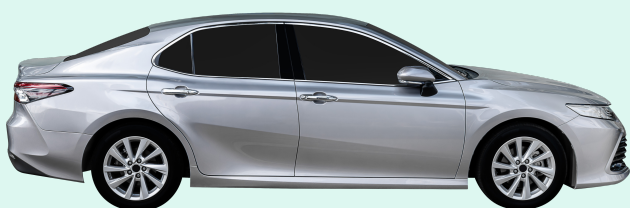
Accuracy is how closely the data reflects the **true values** or **facts**



Ways to Improve Data Accuracy:

- **Double-check** all entries before saving, especially numbers and dates
- Use **copy-paste** for complex values (like serial numbers) to avoid typing errors
- **Report** any discrepancies or errors you notice to the appropriate team

Example: It is inaccurate to say there are five cars in the group below



UNIQUENESS

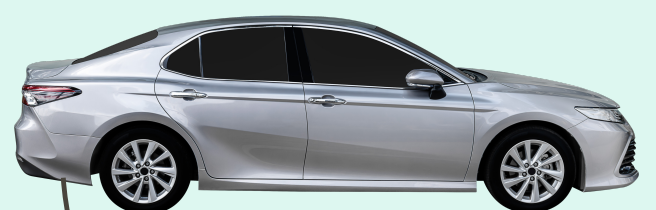
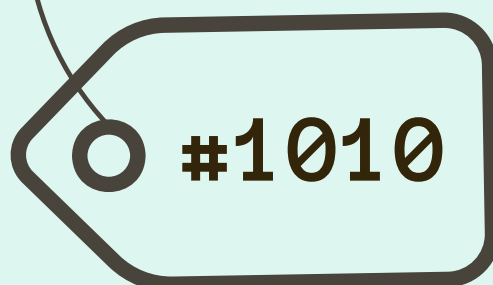
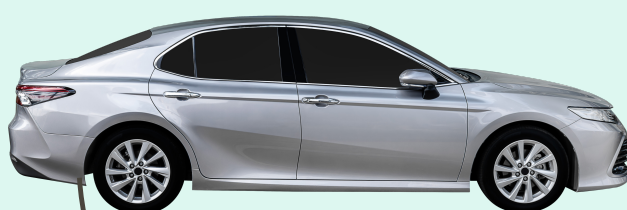
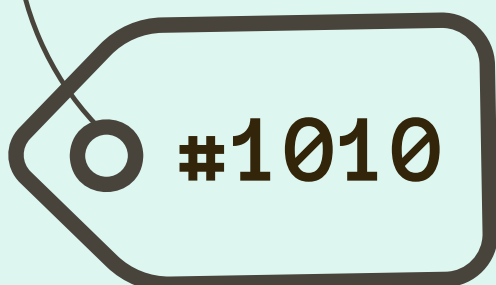
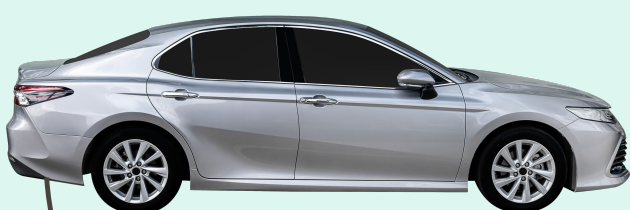
Uniqueness means each data entry is **distinct** and **not repeated** in the dataset



Ways to Improve Data Uniqueness:

- If you find duplicates, **report** them to the appropriate team
- Use the system's **search function** with different keywords to ensure the record doesn't already exist

Example: This group of cars does not meet uniqueness standards, as car #1010 is duplicated in the dataset



COMPLETENESS

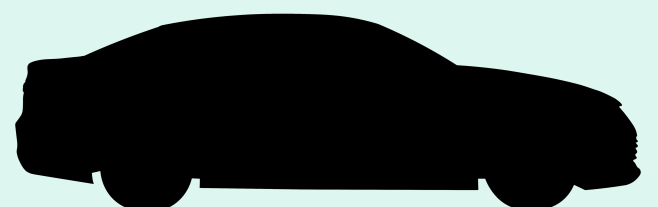
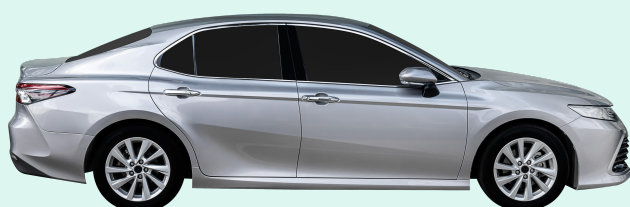
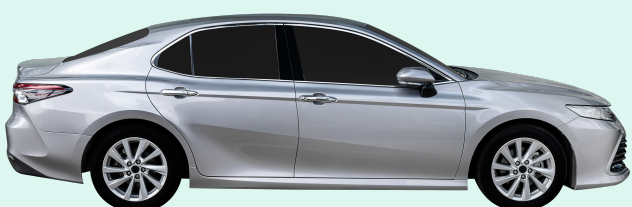
Completeness refers to the extent to which all required data is **present** and **accounted for**.



Ways to Improve Data Completeness:

- Always fill in all **required fields**, don't leave them blank unless instructed
- Add **placeholder values** (e.g. "TBD") if you're unsure but need to flag missing information
- **Follow up** on any pending information and update records once available

Example: The group of cars below is incomplete as an existing third car is not represented in the dataset



TIMELINESS

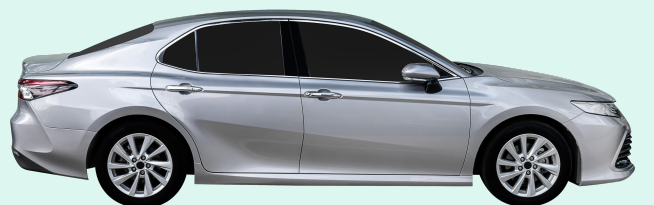
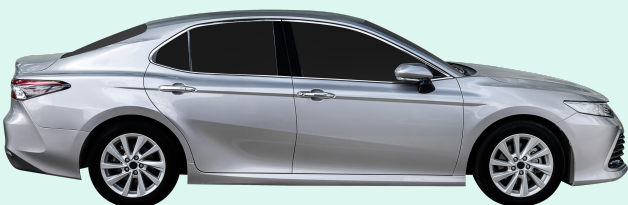
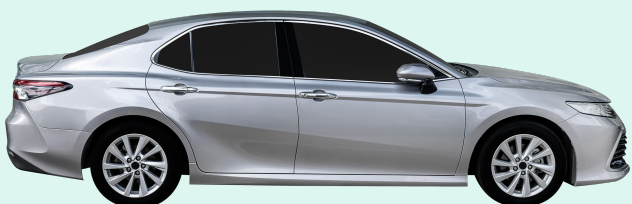
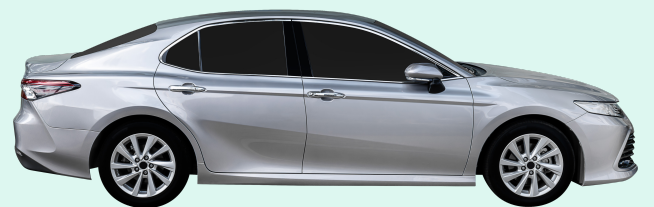
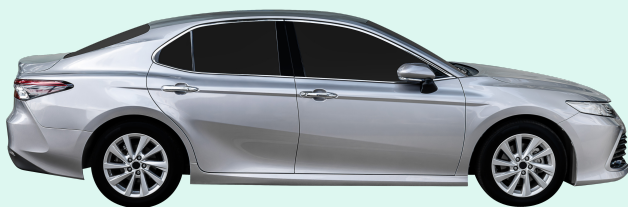
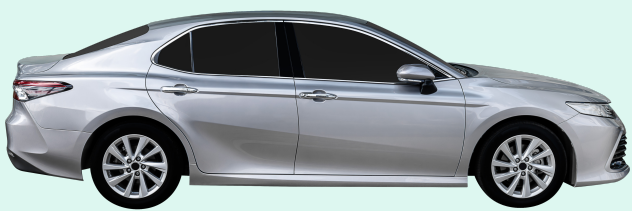
Timeliness means the data is current and up-to-date



Ways to Improve Data Timeliness:

- Enter data as **soon** as you receive it
- **Flag** any outdated information you notice during your daily work

Example: The cars below were counted last year and new cars have since been added. The dataset has not been updated to reflect the new cars. Therefore, the data is no longer timely.



CONSISTENCY

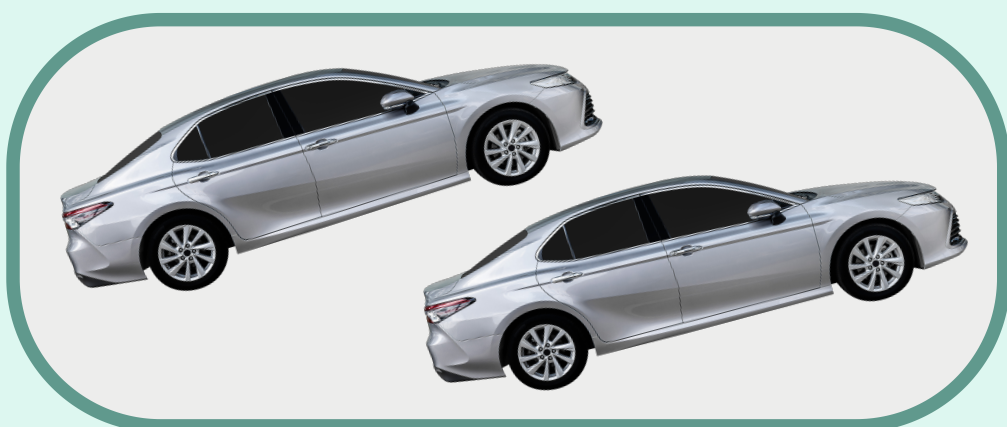
Consistency means data is **uniform** across different datasets and systems



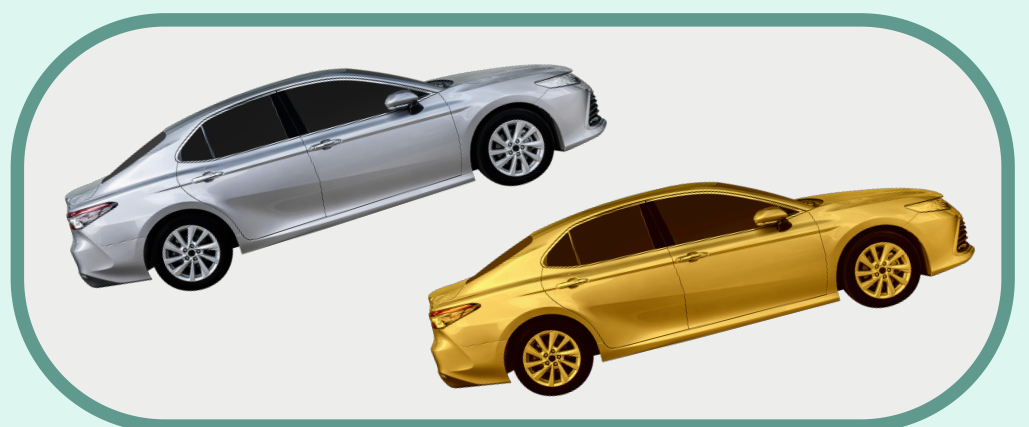
Ways to Improve Data Consistency:

- Check the **dropdown menus** first instead of creating new options
- Use **consistent** formats, language, codes, or terms as outline in your team's guidelines

Example: The color of the cars should be the same across different systems, therefore, these data sets are inconsistent



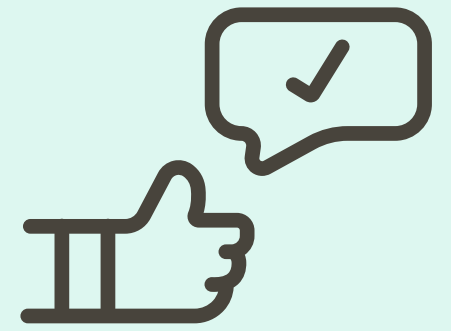
System A



System B

VALIDITY

Validity means the data is **correct** and **appropriate** for its intended use



Ways to Improve Data Validity:

- **Learn** and **follow** your department's data entry standards
- Keep an eye out for values that **don't make sense**, like a driver's license holder who is 3 years old

Example: The "car" below is invalid, as it does not meet the set rules for a car (i.e. four wheels, headlights, bumpers, etc.)

