The Curious Case of the Copycat Cat Foods

Duplicates in Cat Food Data

Prompt

Check for duplicates in the dataset and delete if you find any

Strategy

1. Find copies
2. Delete copies

Code & Results

1. Find copies

This code **checks** for duplicates.

```
# Check for duplicates
duplicates = data.duplicated()
print(f"Number of duplicate rows: {duplicates.sum()}")
```

```
Number of duplicate rows: 10
```

We found ten copies of the same thing.

2. Delete copies

This code **drops** duplicates.

```
# Remove duplicates
data = data.drop_duplicates()

# Check the dataset shape after removing duplicates
print(f"Dataset shape after removing duplicates: {data.shape}")
```

```
Dataset shape after removing duplicates: (218, 23)
```

After dropping copies, only 218 of 228 entries remain.

💡 Tips

- Know what a duplicate is
  Not all repeated rows are mistakes — make sure it's really a copy and not just a similar record.
- Check if the whole row is the same
  Some rows may look similar but have tiny differences — like a missing phone number or a different time.
- Clean up before removing
  Fix things like extra spaces, different letter cases (e.g., "Cat" vs "cat"), or weird date formats — these can hide duplicates.
- Choose which one to keep
  If you find copies, decide: keep the first one, the last one, or maybe the one with more information.
- Keep a backup
  Always save a copy of your original data before deleting anything — just in case!
- Look for almost-duplicates
  Sometimes duplicates have small typos or formatting issues. You might need to fix those first to catch them.
- Check after deleting
  When you're done, take a quick look — does the data still make sense? Did anything important disappear?
- Write down what you removed
  It's good practice to note how many duplicates you found and what you did with them.

Made by: @katerynakononova / Machine learning: for humans on cats
Watch the short: https://youtube.com/shorts/V3gSk4YQI2A?feature=share

Website: https://katerynakononova.github.io/meowlearning/

Dive deeper into the world of AI with *Machine Learning: for Humans on Cats* — now on Amazon!
📘👉 https://www.amazon.com/dp/B0CW9SFYXF