

# Too Many Features, Not Enough Meat

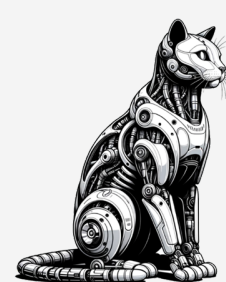
## Feature Selection in Cat Food Data

### Prompt



I have a dataset of 23 features that include numbers and other types of data.  
How do I select the important ones?  
Please give me the Python code to clear my dataset.

### Strategy



1. Remove features with low variance
2. Drop highly correlated numerical features

### Code & Results

#### 1. Low variance filter

This code **removes** features that barely change — they don't help your model learn.

```
from sklearn.feature_selection import VarianceThreshold

# Separate numerical and object features
numerical_data = data.select_dtypes(exclude=['object'])
object_data = data.select_dtypes(include=['object'])

# Apply variance threshold
selector = VarianceThreshold(threshold=0.15)
numerical_reduced = selector.fit_transform(numerical_data)

# Recover selected column names
selected_columns = numerical_data.columns[selector.get_support()]
numerical_reduced_df = pd.DataFrame(numerical_reduced, columns=selected_columns, index=data.index)

# Combine with original object columns
data = pd.concat([object_data, numerical_reduced_df], axis=1)
```

Let's check the columns **before**

```
print(data.columns.tolist())
```

```
['company', 'price', 'weight', 'star_rating', 'total_comments', 'adult', 'chicken', 'wet', 'from_ocean', 'fish', 'tuna', 'salmon', 'mackerel', 'shrimps', 'milk', 'can', 'gravy', 'pouch', 'liver', 'grilled', 'turkey', 'ham', 'vegan']
```

And **after** the code is executed.

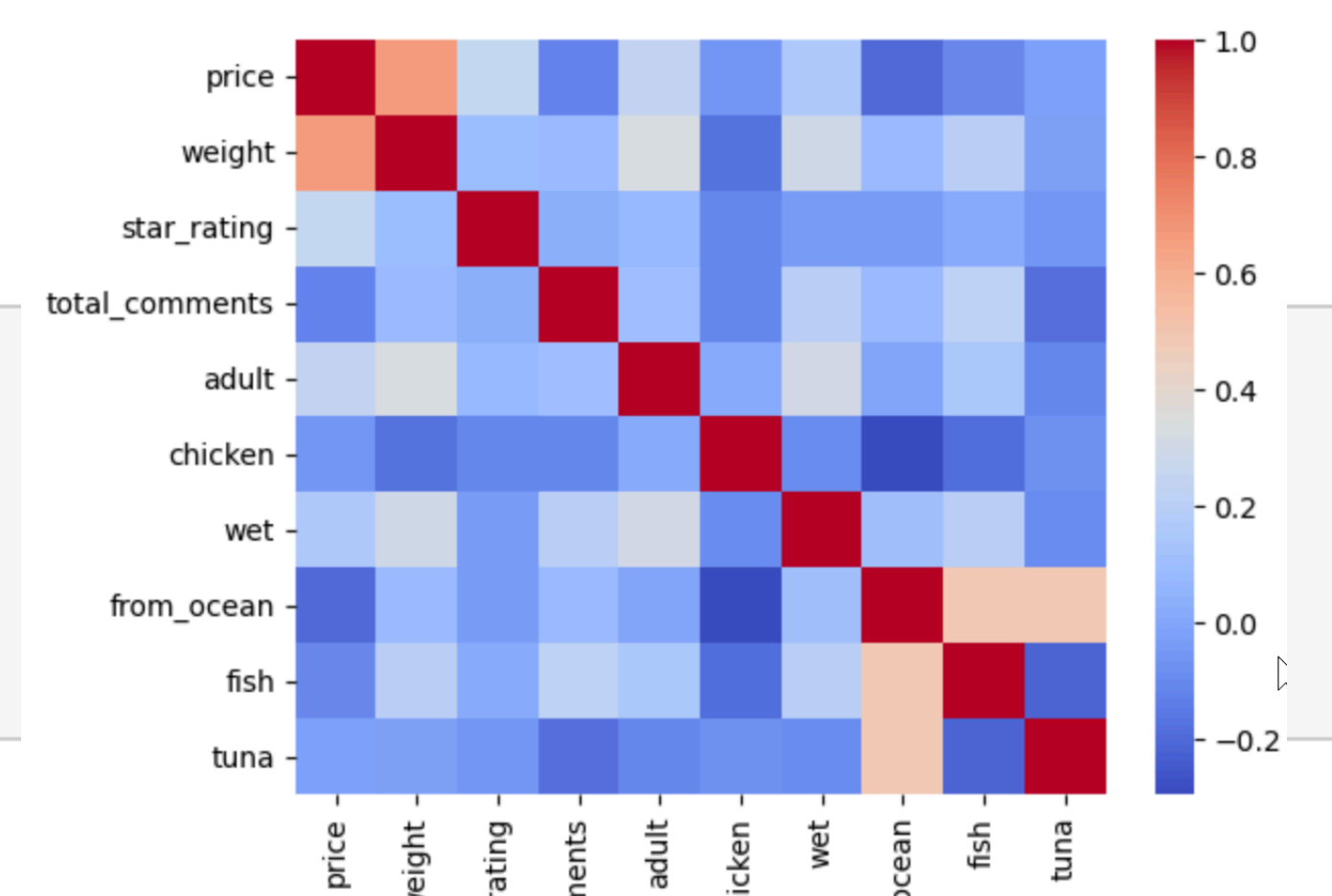
```
['company', 'price', 'weight', 'star_rating', 'total_comments', 'adult', 'chicken', 'wet', 'from_ocean', 'fish', 'tuna']
```

#### 2. Correlation analysis

This code creates a **heatmap** that shows feature pairs that are too similar.

```
import seaborn as sns
import matplotlib.pyplot as plt

corr = data.corr()
sns.heatmap(corr, annot=False, cmap="coolwarm")
```



Drop one per pair!

```
data.drop(['fish', 'tuna'], axis='columns', inplace=True)
print(data.columns.tolist())
```

```
['company', 'price', 'weight', 'star_rating', 'total_comments', 'adult', 'chicken', 'wet', 'from_ocean']
```

### Tips

- 💡 Keep features that teach the model something new.
- 🚀 One clean dataset = one giant leap for smart ML!

Made by: [@katerynakononova](#) / Machine learning: for humans on cats

Watch the short: <https://youtube.com/shorts/V3gSk4YQI2A?feature=share>

Dive deeper into the world of AI with *Machine Learning: for Humans on Cats* — now on Amazon!

📖 <https://www.amazon.com/dp/BOCW9SFYXF>