

Here's a step-by-step data wrangling process using Python and the pandas library.

Import Libraries

```
import pandas as pd
import numpy as np
```

Load Data

```
# Load the data from a CSV file
data = pd.read_csv('data.csv')

# Print the first few rows of the data
print(data.head())
```

Check for Missing Values

```
# Check for missing values
print(data.isnull().sum())
```

Data Cleaning

```
# Remove any rows with missing values
data = data.dropna()

# Remove the '$' symbol from the 'price' column and convert it to numeric
data['price'] = data['price'].str.replace('$', '').str.replace(',', '').astype(float)

# Remove the ',' symbol from the 'currency' column
data['currency'] = data['currency'].str.replace(',', '')

# Convert 'prev_sold_year', 'prev_sold_month', 'prev_sold_day' to datetime
data['prev_sold_date'] = pd.to_datetime(data[['prev_sold_year', 'prev_sold_month', 'prev_sold_day']].assign(DA

# Drop 'prev_sold_year', 'prev_sold_month', 'prev_sold_day' columns
data = data.drop(['prev_sold_year', 'prev_sold_month', 'prev_sold_day'], axis=1)
```

Data Transformation

```
# Extract the state from the 'address' column
data['state'] = data['address'].apply(lambda x: x.split(', ')[-2])

# Extract the city from the 'address' column
data['city'] = data['address'].apply(lambda x: x.split(', ')[-3])

# Create a new column 'acre_lot_category' based on 'acre_lot'
data['acre_lot_category'] = pd.cut(data['acre_lot'], bins=[0, 0.1, 0.5, 1, np.inf], labels=['Small', 'Medium',

# Create a new column 'price_category' based on 'price'
data['price_category'] = pd.cut(data['price'], bins=[0, 200000, 500000, 1000000, np.inf], labels=['Low', 'Medi
```

Data Exploration

```
# Print the summary statistics of the data
print(data.describe())

# Print the unique values in each column
for column in data.columns:
    print(f"Unique values in {column}: {data[column].nunique()}")

# Print the top 5 most frequent states
print(data['state'].value_counts().head(5))
```

```
# Print the top 5 most frequent cities
print(data['city'].value_counts().head(5))
```

This is a basic example of data wrangling. The specific steps and transformations will depend on the characteristics of the data and the goals of the analysis.

Here is the full code:

```
import pandas as pd
import numpy as np

def load_data(file_path):
    data = pd.read_csv(file_path)
    return data

def clean_data(data):
    data = data.dropna()
    data['price'] = data['price'].str.replace('$', '').str.replace(',', '').astype(float)
    data['currency'] = data['currency'].str.replace(',', '')
    data['prev_sold_date'] = pd.to_datetime(data[['prev_sold_year', 'prev_sold_month', 'prev_sold_day']].assign(
    data = data.drop(['prev_sold_year', 'prev_sold_month', 'prev_sold_day'], axis=1)
    return data

def transform_data(data):
    data['state'] = data['address'].apply(lambda x: x.split(',')[2])
    data['city'] = data['address'].apply(lambda x: x.split(',')[3])
    data['acre_lot_category'] = pd.cut(data['acre_lot'], bins=[0, 0.1, 0.5, 1, np.inf], labels=['Small', 'Medi
    data['price_category'] = pd.cut(data['price'], bins=[0, 200000, 500000, 1000000, np.inf], labels=['Low', '
    return data

def explore_data(data):
    print(data.describe())
    for column in data.columns:
        print(f"Unique values in {column}: {data[column].nunique()}")
    print(data['state'].value_counts().head(5))
    print(data['city'].value_counts().head(5))

def main():
    data = load_data('data.csv')
    data = clean_data(data)
    data = transform_data(data)
    explore_data(data)

if __name__ == "__main__":
    main()
```