

## Overview

This automation uses **Zapier + Python** to clean insurance data every time a new row is added to a Google Sheet. The Python script removes duplicates, fills missing values, and outputs clean data to a separate sheet.

---

### Step 1: Trigger – New Spreadsheet Row in Google Sheets

- **App:** Google Sheets
  - **Event:** New Spreadsheet Row
  - **Spreadsheet:** Select the one where you uploaded `insurance.csv`
  - **Worksheet:** Choose the specific sheet (e.g., `Raw_Data`)
- 

### Step 2: Action – Code by Zapier (Python)

- **App:** Code by Zapier
  - **Event:** Run Python
  - **Input Data Fields:**  
Add input fields based on column headers like:
    - `age, sex, bmi, children, smoker, region, charges`
- 

### Python Code (Copy into Code Box)

python

CopyEdit

```
import pandas as pd
```

```
# Step 1: Convert Zapier inputs to a DataFrame
```

```
data = {  
    'age': [input_data['age']],
```

```

    'sex': [input_data['sex']],
    'bmi': [input_data['bmi']],
    'children': [input_data['children']],
    'smoker': [input_data['smoker']],
    'region': [input_data['region']],
    'charges': [input_data['charges']]
}
df = pd.DataFrame(data)

# Step 2: Clean missing values
df['age'] = pd.to_numeric(df['age'], errors='coerce').fillna(30)
df['bmi'] = pd.to_numeric(df['bmi'], errors='coerce').fillna(30.0)
df['children'] = pd.to_numeric(df['children'],
errors='coerce').fillna(0)
df['charges'] = pd.to_numeric(df['charges'],
errors='coerce').fillna(12000)

df['sex'] = df['sex'].fillna('female')
df['smoker'] = df['smoker'].fillna('no')
df['region'] = df['region'].fillna('southwest')

# Step 3: Return cleaned row
output = df.to_dict(orient='records')[0]
return output

```

---

### Step 3: Output – Create Spreadsheet Row in Google Sheets

- **App:** Google Sheets
  - **Event:** Create Spreadsheet Row
  - **Spreadsheet:** Same sheet as before
  - **Worksheet:** A new tab called `Cleaned_Data`
  - **Mapped Fields:** Use output from the Python code step
-

## Example Workflow Summary

pgsql

CopyEdit

Trigger: New Row in "Raw\_Data" Sheet

→ Code by Zapier (Python script cleans the row)

→ Action: Add Cleaned Row to "Cleaned\_Data" Sheet