

## 1. Load the Excel File:

```
import pandas as pd
import numpy as np
from sklearn.impute import KNNImputer
from openpyxl import load_workbook
from openpyxl.styles import PatternFill

file_path = "Data-Melbourne_F_fixed.xlsx" # Replace with the path to your Excel
file
sheet_name = "Data-Melbourne_F_fixed" # Replace with the sheet name

# Read the data into a DataFrame
df = pd.read_excel(file_path, sheet_name=sheet_name)
data = df.copy() # Make a copy for processing
```

## 2. Define Outlier Detection Function:

```
# Function to detect outliers using the IQR method
def detect_outliers_iqr(column):
    Q1 = column.quantile(0.25)
    Q3 = column.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return (column < lower_bound) | (column > upper_bound)
```

## 3. Apply Outlier Detection:

```
# Identify outliers in each column and set them to NaN
for col in data.columns:
    if np.issubdtype(data[col].dtype, np.number): # Check if the column is
numeric
        outliers = detect_outliers_iqr(data[col])
        data.loc[outliers, col] = np.nan # Set outliers to NaN
```

#### 4. K-Nearest Neighbors (KNN) Imputation:

```
# Replace NaN values (outliers) using K-Nearest Neighbors imputation
imputer = KNNImputer(n_neighbors=3)
data_imputed = pd.DataFrame(imputer.fit_transform(data),
                             columns=data.columns)
```

#### 5. Highlight and Replace Outliers

```
# Write the data back to Excel with highlighted outliers
wb = load_workbook(file_path)
ws = wb[sheet_name]

# Define the yellow fill for highlighting
yellow_fill = PatternFill(start_color="FFFF00", end_color="FFFF00",
                           fill_type="solid")

# Highlight outliers in yellow and replace values in the Excel sheet
for col in df.columns:
    if np.issubdtype(df[col].dtype, np.number): # Check if the column is
numeric
        outliers = detect_outliers_iqr(df[col])
        for idx, is_outlier in outliers.items():
            if is_outlier: # If the cell is an outlier
                cell =ws.cell(row=idx+2, column=df.columns.get_loc(col)+1)
# +2 for header row offset
                cell.value = data_imputed.loc[idx, col] # Replace with
imputed value
                cell.fill = yellow_fill # Highlight with yellow
```

#### 6. Save as new file:

```
# Save the modified Excel file
wb.save("Melbourne_file_with_outliers_highlighted.xlsx")
```

code :

```
import pandas as pd
import numpy as np
from sklearn.impute import KNNImputer
from openpyxl import load_workbook
from openpyxl.styles import PatternFill

# Load the Excel file

file_path = r"C:\Users\ASUS\Desktop\Data-Melbourne_F_fixed.xlsx" # Replace with
the path to your Excel file
sheet_name = "Data-Melbourne_F_fixed" # Replace with the sheet name

# Read the data into a DataFrame
df = pd.read_excel(file_path, sheet_name=sheet_name)
data = df.copy() # Make a copy for processing

# Function to detect outliers using the IQR method
def detect_outliers_iqr(column):
    Q1 = column.quantile(0.25)
    Q3 = column.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return (column < lower_bound) | (column > upper_bound)

# Identify outliers in each column and set them to NaN
for col in data.columns:
    if np.issubdtype(data[col].dtype, np.number): # Check if the column is
numeric
        outliers = detect_outliers_iqr(data[col])
        data.loc[outliers, col] = np.nan # Set outliers to NaN

# Replace NaN values (outliers) using K-Nearest Neighbors imputation
imputer = KNNImputer(n_neighbors=3)
data_imputed = pd.DataFrame(imputer.fit_transform(data), columns=data.columns)

# Write the data back to Excel with highlighted outliers
wb = load_workbook(file_path)
ws = wb[sheet_name]

# Define the yellow fill for highlighting
yellow_fill = PatternFill(start_color="FFFF00", end_color="FFFF00",
fill_type="solid")
```

```

# Highlight outliers in yellow and replace values in the Excel sheet
for col in df.columns:
    if np.issubdtype(df[col].dtype, np.number): # Check if the column is numeric
        outliers = detect_outliers_iqr(df[col])
        for idx, is_outlier in outliers.items():
            if is_outlier: # If the cell is an outlier
                cell = ws.cell(row=idx + 2, column=df.columns.get_loc(col) +
1) # +2 for header row offset
                cell.value = data_imputed.loc[idx, col] # Replace with imputed
value
                cell.fill = yellow_fill # Highlight with yellow

# Save the modified Excel file
wb.save("Melbourne_file_with_outliers_highlighted.xlsx")

```

## Result:



Data-Melbourne\_F\_fixed.xlsx