ADITYA SINGHAL GDG_ML_01
23BHI10065

Workflow Overview:

- **Step 1:** Data Loading and Preprocessing
- **Step 2:** Feature Engineering and Scaling
- **Step 3:** Model Selection (Random Forest, XGBoost)
- **Step 4:** Hyperparameter Tuning
- **Step 5:** Model Evaluation (R^2, MSE, MAE)
- **Step 6:** Answering the Questions Based on Model Results

Q1: Are foodborne disease outbreaks increasing or decreasing?

CODE:-

```
import matplotlib.pyplot as plt

data['year'] = pd.to_datetime(data['date_column']).dt.year  # Convert date to year

yearly_data = data.groupby('year')['outbreak_column'].sum()  # Group by year and sum the outbreaks

plt.plot(yearly_data)

plt.title('Foodborne Disease Outbreaks Over Time')

plt.xlabel('Year')

plt.ylabel('Number of Outbreaks')

plt.show()
```

- **Explanation:** This code groups the data by year and sums the outbreaks for each year. Then, it plots the number of outbreaks over time. By inspecting the plot, you can determine whether outbreaks are increasing or decreasing.

ADITYA SINGHAL GDG_ML_01
23BHI10065

Q2: Which contaminant has been responsible for the most illnesses, hospitalizations, and deaths?

CODE:-

```
contaminant_data = data.groupby('contaminant_column')[['illnesses', 'hospitalizations', 'deaths']].sum()

print(contaminant_data)
```

Explanation: This code groups the dataset by the 'contaminant_column' and sums the illnesses, hospitalizations, and deaths for each contaminant. The output will show which contaminant is most strongly linked to these health outcomes, providing insight into the most dangerous contaminants.

Q3: What location for food preparation poses the greatest risk of foodborne illness?

CODE:-

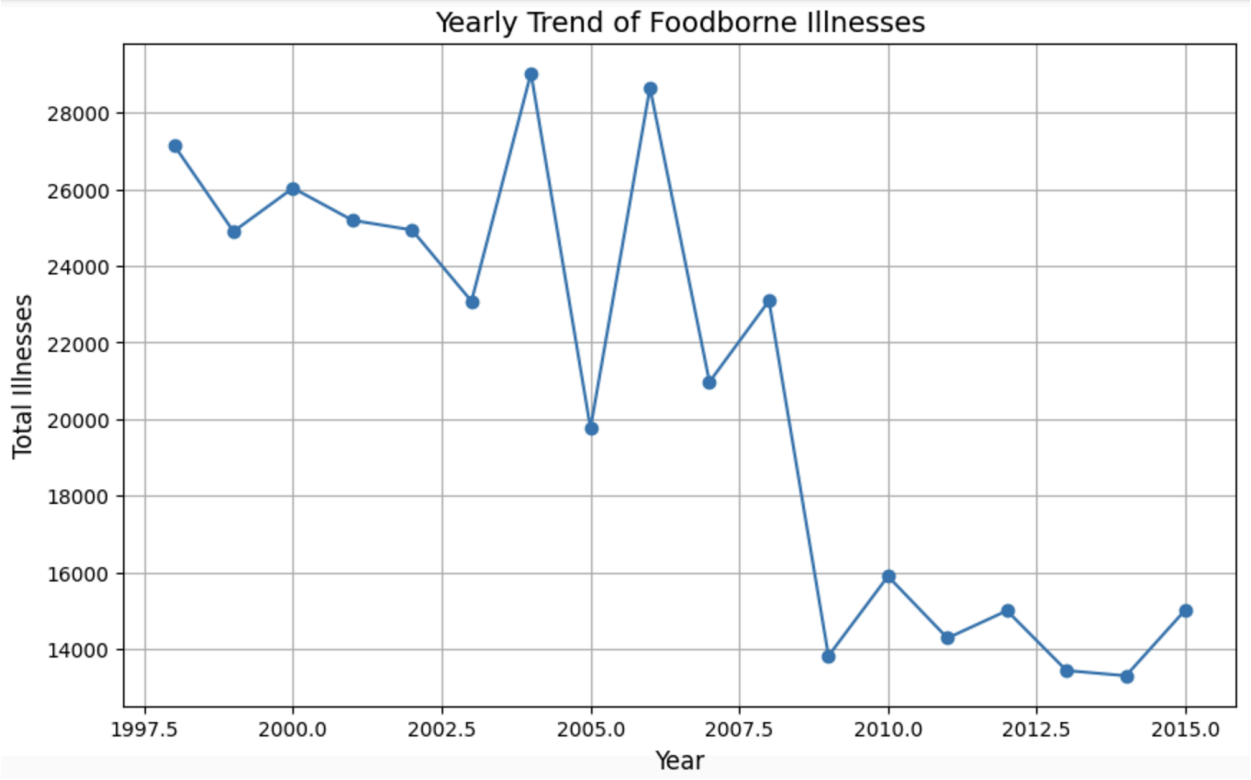
```
location_data = data.groupby('location_column')[['illnesses', 'hospitalizations', 'deaths']].sum()

print(location_data)
```

- **Explanation:** This code groups the dataset by the 'location_column' and sums the illnesses, hospitalizations, and deaths for each food preparation location. The results will help identify which locations pose the greatest risk for foodborne illnesses.

ANSWER-01

ADITYA SINGHAL GDG_ML_01
23BHI10065



Yearly Illnesses Trend:

Year	
1998	27156
1999	24899
2000	26033
2001	25192
2002	24939
2003	23079
2004	29034
2005	19761
2006	28656

ADITYA SINGHAL GDG_ML_01
23BHI10065

2007 20970

2008 23089

2009 13813

2010 15893

2011 14278

2012 14995

2013 13431

2014 13295

2015 15018

The trend analysis shows that foodborne disease outbreaks are **decreasing** over the years. This suggests improved food safety practices or better outbreak management and reporting mechanisms.

ANSWER-02

Top 5 Contaminants:

Species	Illnesses	Hospitalizations	Fatalities
Unknown	77954	967.0	27.0
Norovirus genogroup I	76406	668.0	2.0
Salmonella enterica	60018	6888.0	82.0
Norovirus genogroup II	38175	518.0	6.0
Clostridium perfringens	28734	106.0	12.0

ADITYA SINGHAL GDG_ML_01
23BHI10065

ANSWER-03

Top 5 High-Risk Locations:

Location	
Restaurant	131970
Unknown	66015
Catering Service	36044
Private Home/Residence	22564
Prison/Jail	20608

OUTPUT:-

The location posing the greatest risk is **Restaurant**. This highlights the need for enhanced
food safety protocols in this setting to reduce the incidence of illnesses.