Hemali Chotalia
23BHI10073

**GDG_ML_TASK1**

## 1)Are foodborne disease outbreaks increasing or decreasing?

Answer

**Steps:**

1. Group by Year: Count the number of outbreaks per year.
2. Visualise Trend: Plot the yearly trend.

**Code:**

```
outbreaks_per_year = df.groupby('Year').size()
```

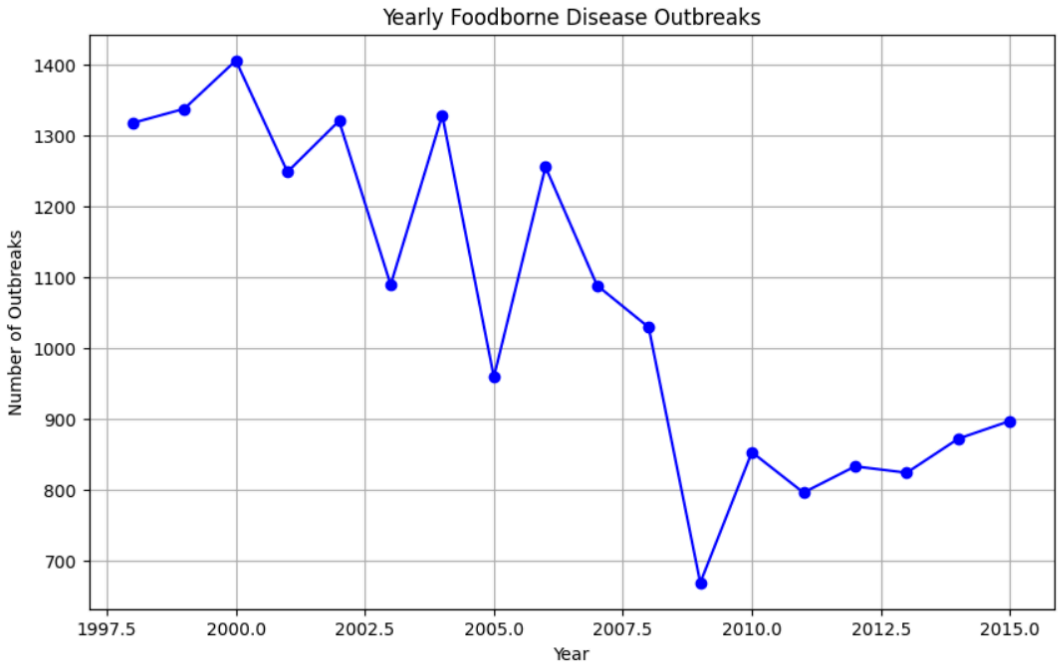
```
plt.figure(figsize=(8, 4))

plt.plot(outbreaks_per_year.index, outbreaks_per_year.values,
marker='o', colour='b')

plt.title('Yearly Foodborne Disease Outbreaks')

plt.xlabel('Year')

plt.ylabel('Number of Outbreaks')

plt.grid(True)

plt.show()

print("\nYearly Illnesses Trend:\n", outbreaks_per_year)
```

**Output:**

```
Yearly Illnesses Trend:
 Year
1998    1317
1999    1337
2000    1405
2001    1248
2002    1320
2003    1089
2004    1328
2005     959
2006    1255
2007    1088
2008    1029
2009     669
```

```
2010      853
2011      796
2012      833
2013      824
2014      872
2015      897
dtype: int64
```

**GRAPH:**




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.

## 2)Which contaminant has caused the most illnesses, hospitalizations, and deaths?

Answer

**Steps:**

1. Group by Contaminant: Sum illnesses, hospitalizations, and fatalities.
2. Sort and Identify Top Contaminants: Find the contaminant causing the most harm.

**Code:**

```
contaminant_data = df.groupby('Species')[['Illnesses',
'Hospitalizations', 'Fatalities']].sum()
```

```
top_contaminants = contaminant_data.sort_values(by='Illnesses',
ascending=False)

print("\nTop 5 Contaminants by Illnesses:\n")

print(top_contaminants.head())
```

**Output:**

```
Top 5 Contaminants by Illnesses:

                         Illnesses  Hospitalizations  Fatalities

Species

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
```

### 3)What location for food preparation poses the greatest risk of foodborne illness?

Answer

**Steps:**

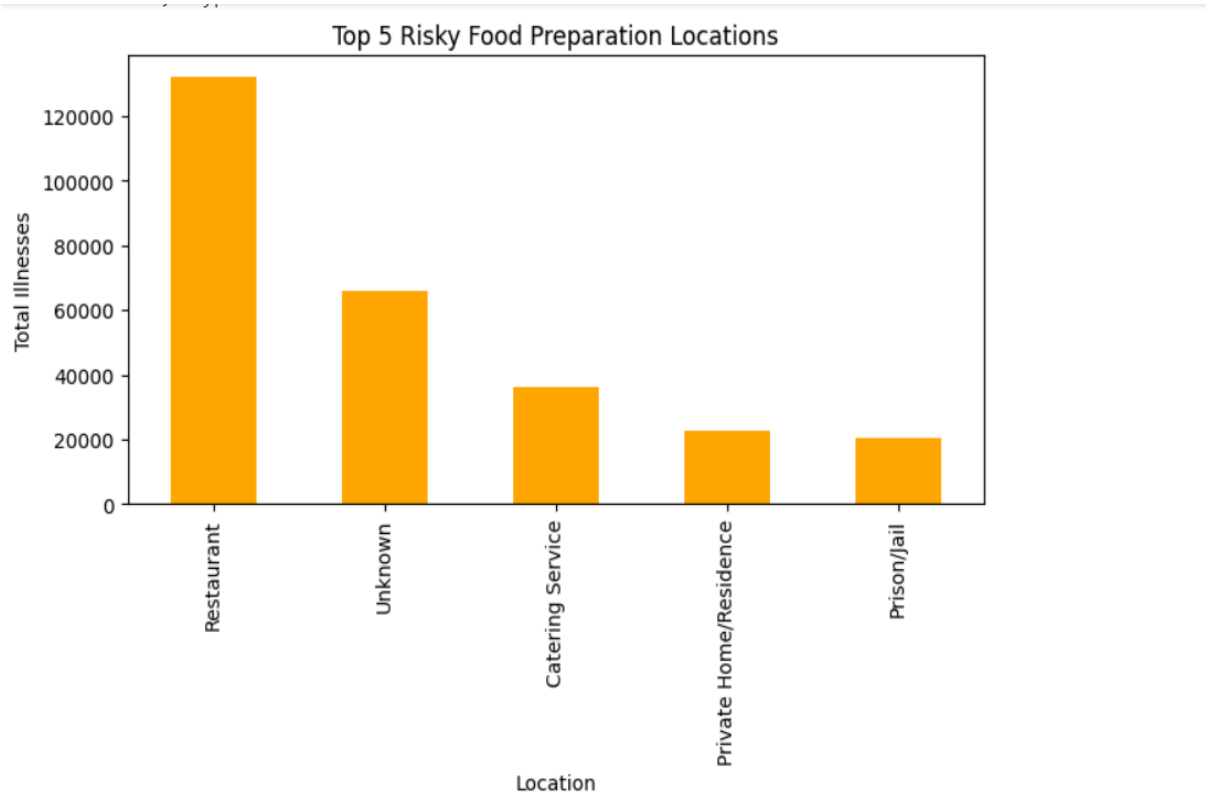
1. Group by Location: Sum illnesses.
2. Sort and Identify Risky Locations: Find locations with the most illnesses.

**Code:**

```
location_risks =
df.groupby('Location')['Illnesses'].sum().sort_values(ascending=False)

print("\nTop 5 Risky Locations:\n", location_risks.head())



location_risks.head(5).plot(kind='bar', figsize=(8, 4), color='orange')
```

```
plt.title('Top 5 Risky Food Preparation Locations')

plt.xlabel('Location')

plt.ylabel('Total Illnesses')

plt.show()
```

**GRAPH:**




**Output:**

```
Top 5 Risky Locations:

 Location

Restaurant                131970

Unknown                    66015

Catering Service           36044

Private Home/Residence     22564
```

```
Prison/Jail                20608

Name: Illnesses, dtype: int64
```

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.