



# **Exploring Breast Cancer Trends in Females:**

## **A comprehensive look at Factors and Hotspots.**

Presented by: Warda Rashid

## Abstract:

This study is about the breast cancer patterns in females, using data from a vizhub health data website. This study focus on three region Asia, Europe and America, exploring four age groups (20-24, 25-29, 30-34, 35-39) from 2017 to 2019.

we check which age groups are more prone to breast cancer and why.

In this study we try to understand which risk factors connect to breast cancer.

We also check where breast cancer causes the more deaths.

```
[ ] breast_cancer.head(2)
```

	measure_name	location_name	sex_name	age_name	cause_name	rei_name	year	val	upper	lower
0	Deaths	America	Female	20-24 years	Breast cancer	Alcohol use	2018	9.226472	11.335756	7.232823
1	Deaths	America	Female	25-29 years	Breast cancer	Tobacco	2018	7.757920	13.598633	1.811088

```
[ ] breast_cancer.shape
```

```
(135, 10)
```

```
[ ] breast_cancer.columns
```

```
Index(['measure_name', 'location_name', 'sex_name', 'age_name', 'cause_name',  
      'rei_name', 'year', 'val', 'upper', 'lower'],  
      dtype='object')
```

In this research we have total 135 sample data of three region. Each region include same amount of sample data.

```
[ ] location_name = breast_cancer["location_name"].value_counts()  
location_name  
  
America    45  
Asia       45  
Europe     45  
Name: location_name, dtype: int64
```

we considered these risk factors to see how much of a role they play in causing breast cancer

- Alcohol use
- Tobacco
- Iron deficiency
- Zinc deficiency
- Drug use
- Dietary risk
- Low physical activity
- Smoking
- Air pollution
- Vitamin A deficiency
- Environmental/occupational risks

# Findings

From the data we found that these risk factors causing the most breast cancer .

```
[ ] rei_name = breast_cancer["rei_name"].value_counts()  
rei_name
```

Alcohol use	36
Tobacco	27
Dietary risks	27
Low physical activity	27
Smoking	18

Name: rei\_name, dtype: int64

```
[ ] rename_location= breast_cancer.groupby("location_name")["rei_name"].value_counts()  
rename_location
```

location_name	rei_name	
America	Alcohol use	12
	Dietary risks	9
	Low physical activity	9
	Tobacco	9
	Smoking	6
Asia	Alcohol use	12
	Dietary risks	9
	Low physical activity	9
	Tobacco	9
	Smoking	6
Europe	Alcohol use	12
	Dietary risks	9
	Low physical activity	9
	Tobacco	9
	Smoking	6

Name: rei\_name, dtype: int64

# Death Rate

```
[ ] Avg_val= breast_cancer.groupby("location_name" )["val"].mean().sort_values(ascending=False)
```

```
[ ] Avg_val
```

location\_name

Asia        141.061732

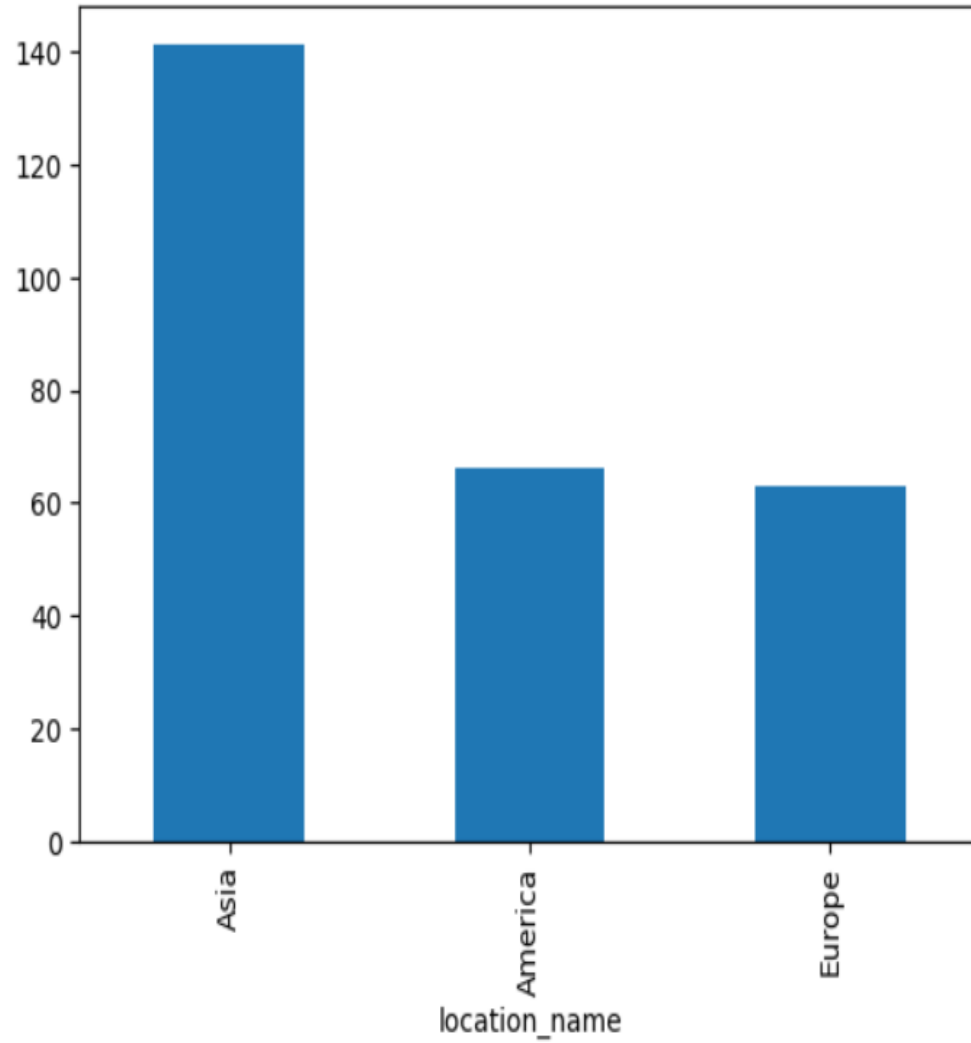
America     66.232981

Europe      63.055788

Name: val, dtype: float64

▶ Avg\_val.plot.bar()

↳ <Axes: xlabel='location\_name'>





# Age group that affected the most

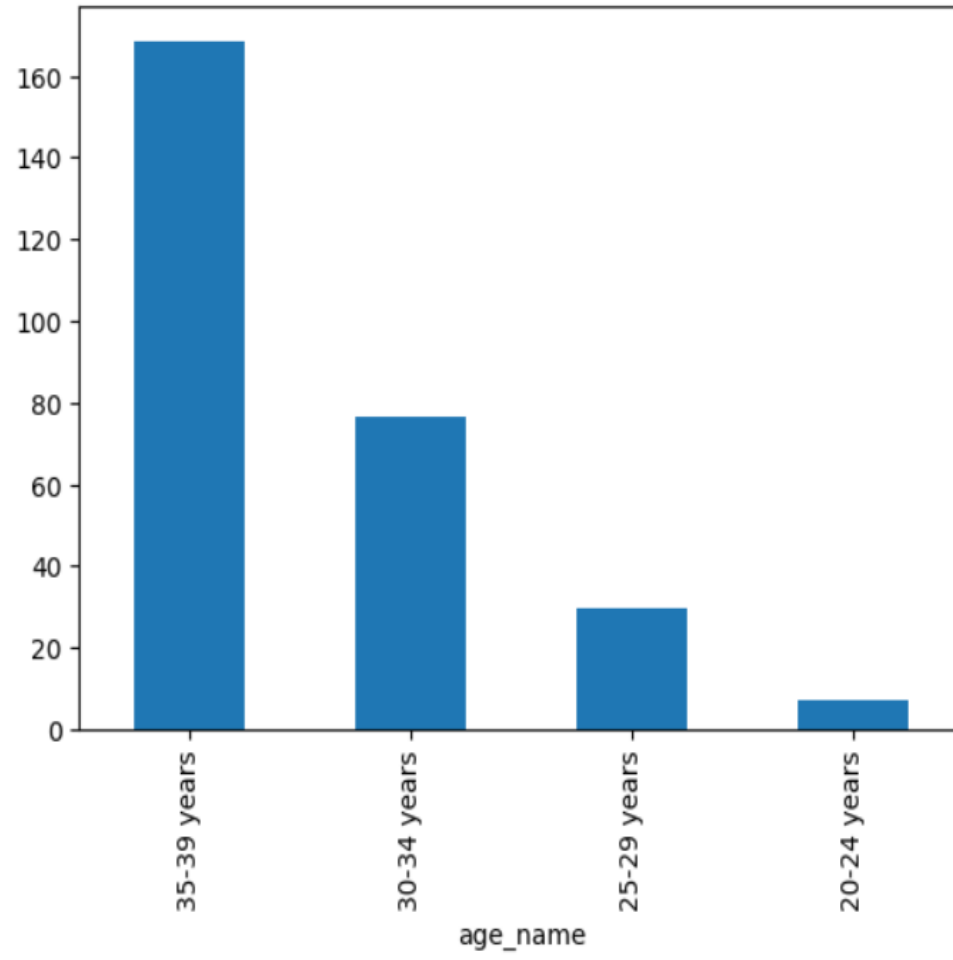
```
[ ] age_val= breast_cancer.groupby("age_name" )["val"].mean().sort_values(ascending=False)
```

```
[ ] age_val
```

```
age_name
35-39 years    168.385339
30-34 years     76.559582
25-29 years     29.931927
20-24 years      7.300193
Name: val, dtype: float64
```

▶ `age_val.plot.bar()`

↳ `<Axes: xlabel='age_name'>`



# How many times each age groups affected

```
▶ rename_year= breast_cancer.groupby("age_name" )["rei_name"].value_counts()  
rename_year
```

age_name	rei_name	
20-24 years	Alcohol use	9
25-29 years	Alcohol use	9
	Dietary risks	9
	Low physical activity	9
	Tobacco	9
30-34 years	Alcohol use	9
	Dietary risks	9
	Low physical activity	9
	Smoking	9
	Tobacco	9
35-39 years	Alcohol use	9
	Dietary risks	9
	Low physical activity	9
	Smoking	9
	Tobacco	9

Name: rei\_name, dtype: int64

## Which year has most breast cancer cases

```
[ ] breast_cancer_avg = breast_cancer.groupby("year" )["val"].mean().sort_values(ascending=False)
breast_cancer_avg
```

```
year
2019    90.664120
2018    90.137467
2017    89.548915
Name: val, dtype: float64
```

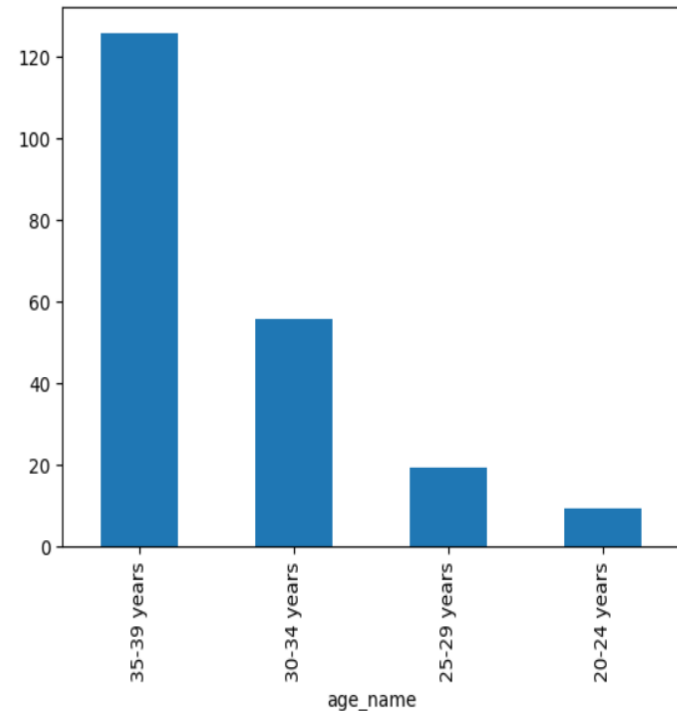
# In America which year has most cases and which age groups affected

```
▶ america_avg_val_year = america_df.groupby("year")["val"].mean().sort_values(ascending=False)  
america_avg_val_year
```

```
↳ year  
2017    67.118018  
2018    66.272401  
2019    65.308523  
Name: val, dtype: float64
```

```
[ ] america_avg.plot.bar()
```

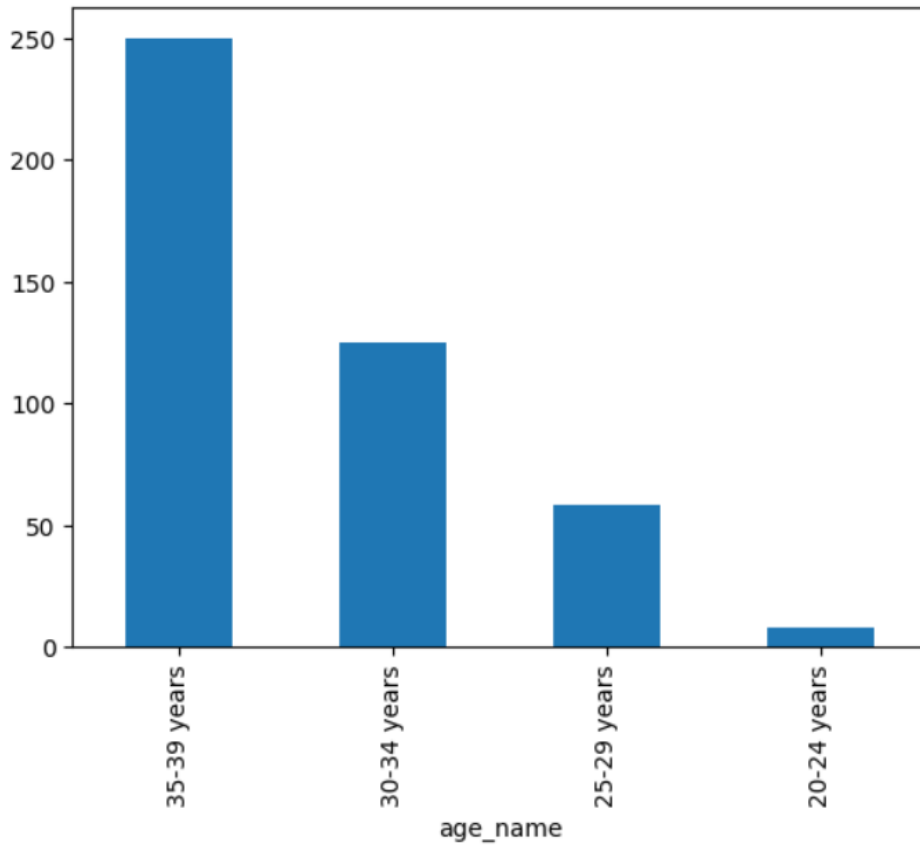
<Axes: xlabel='age\_name'>



# In Asia which year has most cases and which age groups affected

```
asia_avg.plot.bar()
```

```
<Axes: xlabel='age_name'>
```



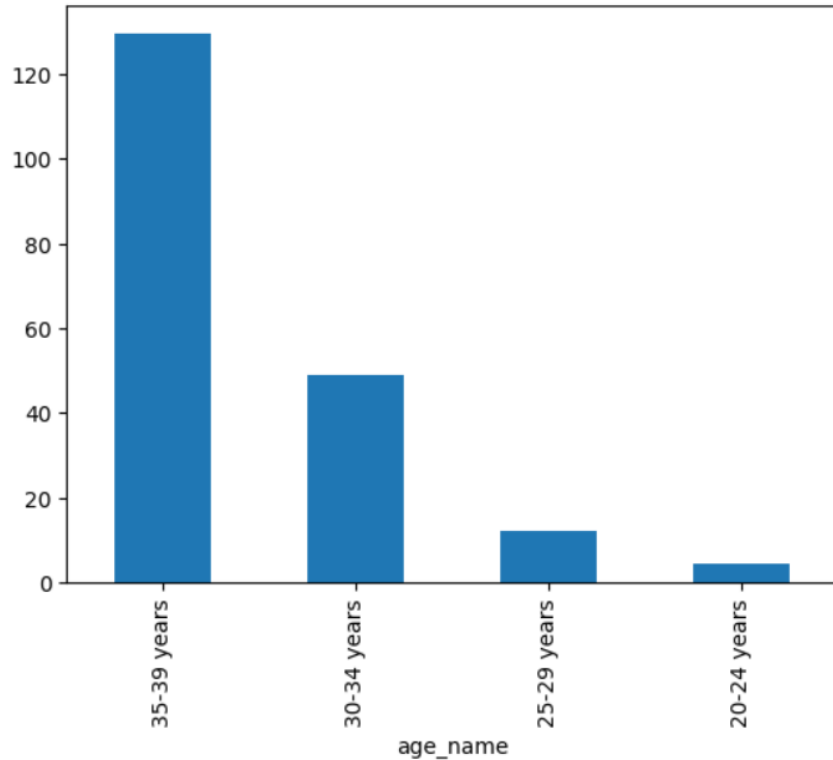
```
asia_avg_val_year = asia_df.groupby("year")["val"].mean().sort_values(ascending=False)  
asia_avg_val_year
```

```
year  
2019    144.055899  
2018    141.151533  
2017    137.977765  
Name: val, dtype: float64
```

# In Europe which year has most cases and which age groups affected

▶ europe\_avg.plot.bar()

↳ <Axes: xlabel='age\_name'>



```
[ ] europe_avg_val_year = europe_df.groupby("year")["val"].mean().sort_values(ascending=False)
europe_avg_val_year
```

```
year
2017    63.550961
2018    62.988466
2019    62.627937
Name: val, dtype: float64
```