

```
In [1]: import numpy as np
```

```
In [2]: import pandas as pd
```



CMSC 320 / HomeWork4

Published at Mar 27, 2023 Unlisted

```
df = pd.read_csv('Dataset3.csv')
# Count the number of people who got superpowers with the serum
num_super_serum = df[df['SuperSerum'] == 1]['GotPowers'].sum()

# Count the number of people who got superpowers naturally
num_natural = df[df['SuperSerum'] == 0]['GotPowers'].sum()

# Print the results
print(f"Number of people who got superpowers with the serum: {num_super_serum}")
print(f"Number of people who got superpowers naturally: {num_natural}")
```

Number of people who got superpowers with the serum: 44902

Number of people who got superpowers naturally: 44145

```
from scipy.stats import chi2_contingency

# Load the dataset into a pandas dataframe
df = pd.read_csv('Dataset3.csv')
# Create a contingency table
contingency_table = pd.crosstab(df['GotPowers'], df['SuperSerum'])

from scipy.stats import chi2_contingency

stat, p, dof, expected = chi2_contingency(contingency_table)
print('chi-square statistic:', stat)
print('p-value:', p)
print('degrees of freedom:', dof)
print('expected frequencies:', expected)
```

chi-square statistic: 102.89321293670491

p-value: 3.53709332538169e-24

degrees of freedom: 1

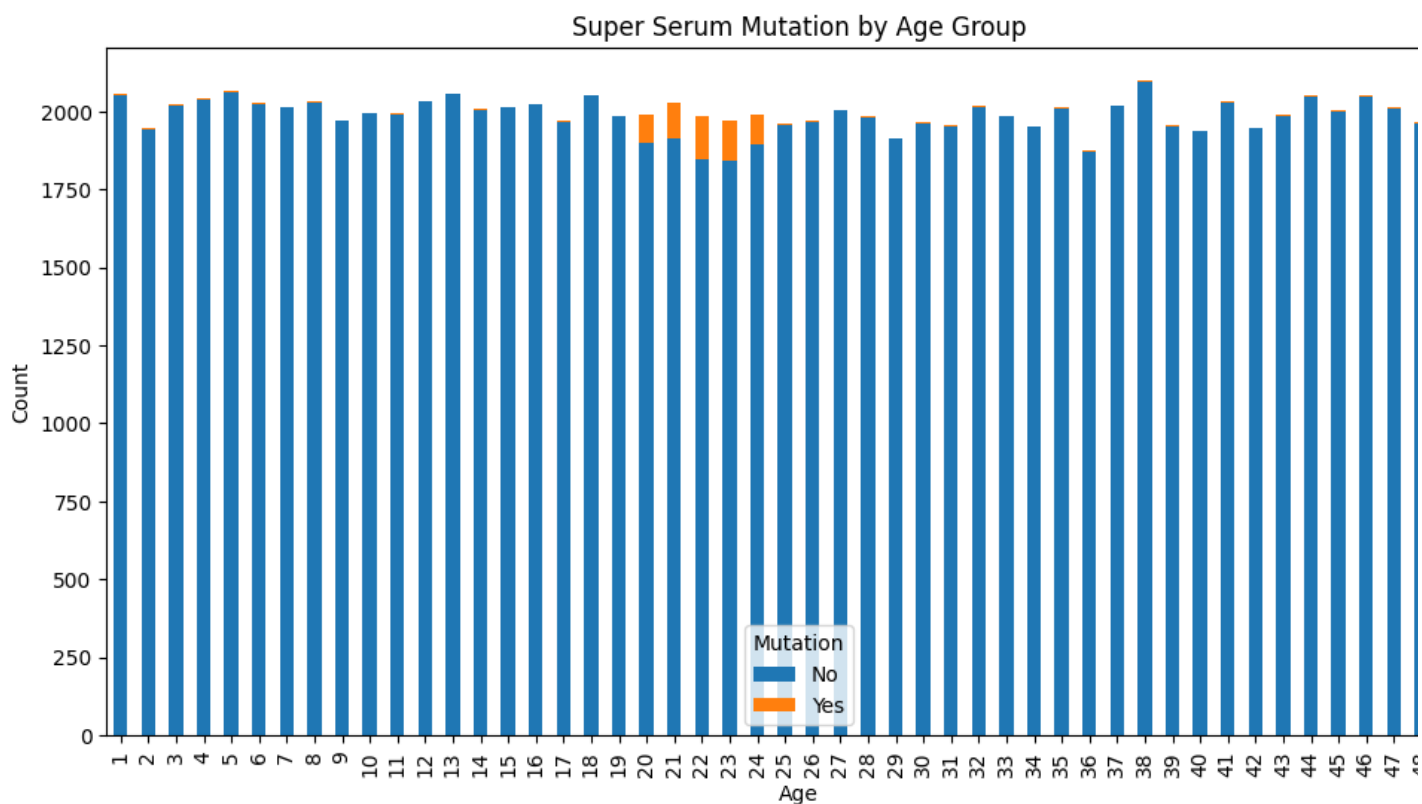
expected frequencies: [[5491.61514 5461.38486]
[44646.38486 44400.61514]]

```
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('Dataset3.csv')

# create a new column for mutation status
df["Mutation"] = df["GotMutation"].apply(lambda x: "No" if x==0 else "Yes")

# group the data by age and mutation status
age_groups = df.groupby(["Age", "Mutation"]).size().unstack()

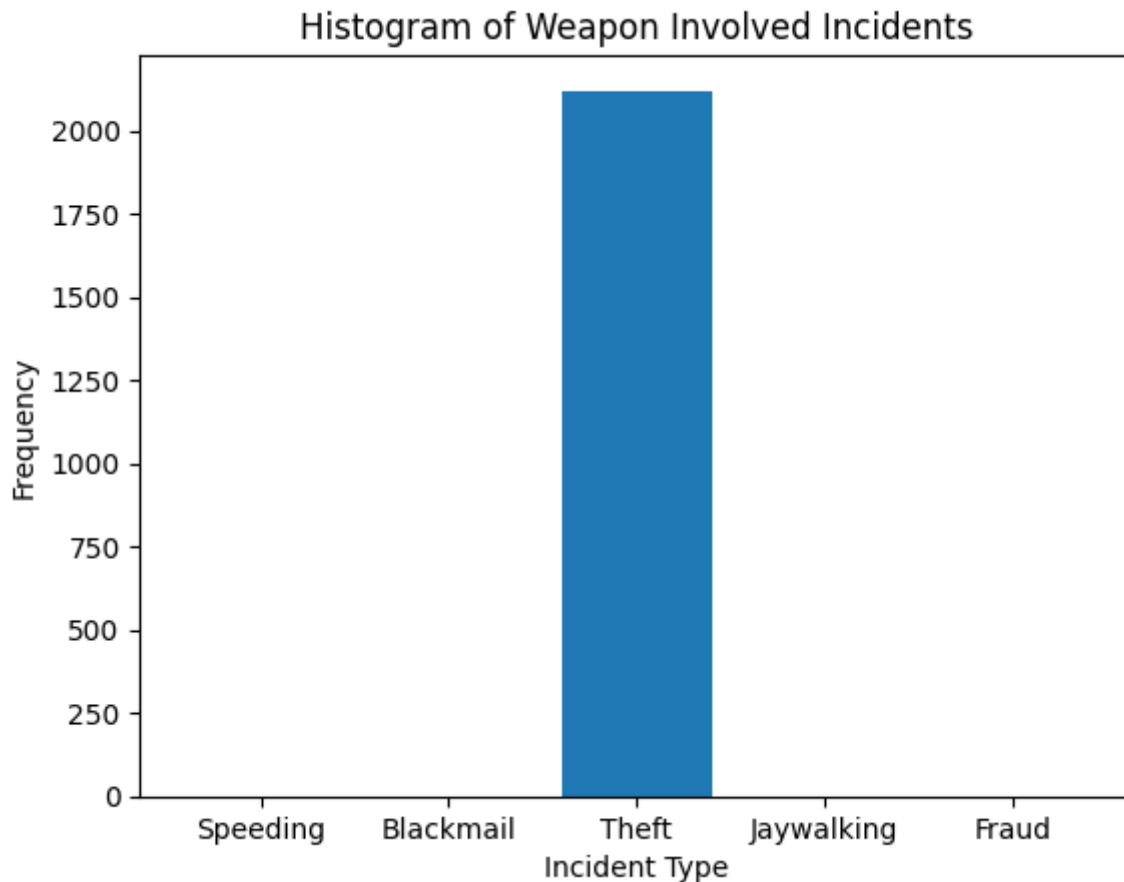
# plot the bar chart
ax = age_groups.plot(kind="bar", stacked=True, figsize=(12,6))
ax.set_xlabel("Age")
ax.set_ylabel("Count")
ax.set_title("Super Serum Mutation by Age Group")
plt.show()
```



```
import matplotlib.pyplot as plt
# Load the dataset into a pandas dataframe
df = pd.read_csv('Dataset1.csv')
# Identify incidents of Speeding, Blackmail, Theft, Jaywalking, and Fraud in the Weapon_Invol
ic = ['Speeding', 'Blackmail', 'Theft', 'Jaywalking', 'Fraud']
weapon_involved = df.loc[df['Weapon_Involved'] == True, 'Type']
```

```
types = weapon_involved.value_counts().reindex(ic, fill_value=0)

# Plot a histogram of the identified incidents
plt.bar(types.index, types.values)
plt.title("Histogram of Weapon Involved Incidents")
plt.xlabel("Incident Type")
plt.ylabel("Frequency")
plt.show()
```



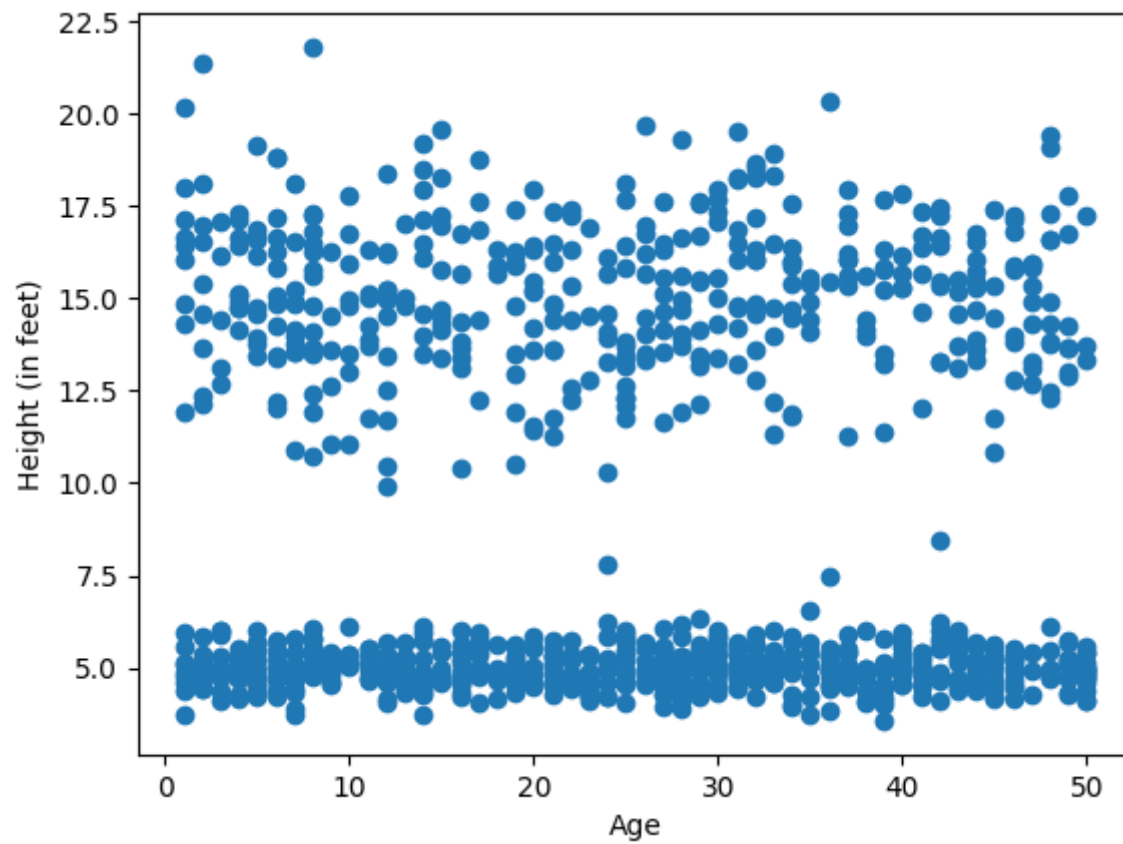
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv('Dataset2.csv')

# Create a scatter plot of Age vs. Height_in_feet
plt.scatter(df['Age'], df['Height_in_feet'])
plt.xlabel('Age')
plt.ylabel('Height (in feet)')
plt.show()

# Group the data by Age and calculate the mean height for each group
age_groups = df.groupby('Age')['Height_in_feet'].mean().reset_index()
```

```
# Display the results  
print(age_groups)
```



```

37  38      8.346501
38  39      8.140776
39  40      8.259033
40  41      9.867789
41  42      8.076909
42  43     10.044014
43  44      7.893874
44  45      9.348085
45  46     12.165970
46  47     11.459068
47  48      8.548599
48  49      6.344203

```

```

import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
# create the dataframe
df = pd.read_csv('Dataset4.csv')

# filter for unemployed individuals
unemployed_df = df[df['Profession'] == 'Unemployed']

# check for non-zero values in the income column
non_zero_income = unemployed_df[unemployed_df['Income'] != 0]

print(non_zero_income)

# create histogram of income for unemployed individuals
plt.hist(unemployed_df['Income'])
plt.xlabel('Income')
plt.ylabel('Frequency')
plt.title('Income Distribution for Unemployed Individuals')
plt.show()

```

	Unnamed: 0	Age	Gender	Profession	Income	Job_Sat
3	3	38	Male	Unemployed	3700.764831	0
4	4	55	Male	Unemployed	5555.823704	2
6	6	44	Male	Unemployed	7222.011737	2
17	17	54	Male	Unemployed	6242.092048	5
21	21	31	Female	Unemployed	7096.535467	1
...
9976	9976	58	Female	Unemployed	2198.367847	0
9979	9979	28	Female	Unemployed	2189.989185	3
9980	9980	51	Female	Unemployed	4372.559115	2
9981	9981	23	Female	Unemployed	5788.317711	4
9985	9985	39	Male	Unemployed	2627.314924	4

[1512 rows x 6 columns]

