

```
# data about the blood contents is crucial to compare to healthy levels
```

```
blood_pressure_data = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
GROUP BY sex, age
ORDER BY sex, age ASC
""")

blood_pressure_data.show(30)
```

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Female	20	12661	111.31	69.3
Female	25	29618	111.04	69.42
Female	30	28701	111.38	70.11
Female	35	26468	112.67	71.05
Female	40	60071	115.21	72.63
Female	45	54655	117.64	73.89
Female	50	64610	120.02	75.1
Female	55	55286	121.79	75.57
Female	60	55036	124.31	75.89
Female	65	26388	127.46	76.35
Female	70	26717	129.89	76.38
Female	75	13932	131.59	76.56
Female	80	8547	132.77	76.84
Female	85	2241	131.93	76.3
Male	20	9310	119.32	73.41
Male	25	34752	121.14	74.86
Male	30	48899	122.56	76.6
Male	35	58258	123.57	77.94
Male	40	70314	123.85	78.73
Male	45	63700	124.2	79.1
Male	50	64824	124.93	79.24
Male	55	55937	125.88	78.94
Male	60	51027	127.02	78.37
Male	65	26573	128.21	77.4
Male	70	23949	129.28	76.47
Male	75	11401	130.26	75.66
Male	80	6421	130.79	75.25
Male	85	1050	130.59	74.93

```
blood_pressure_data_female = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Female'
GROUP BY sex, age
ORDER BY age ASC
""")
```

```
blood_pressure_data_female.show(15)
```

```
blood_pressure_data_male = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Male'
GROUP BY sex, age
```

```

ORDER BY age ASC
""")

blood_pressure_data_male.show(15)
# for males
x = blood_pressure_data_male.select('age').toPandas()['age'].to_numpy()
y =
blood_pressure_data_male.select('Systolic_blood_pressure_mmHg').toPandas(
['Systolic_blood_pressure_mmHg']).to_numpy()
z =
blood_pressure_data_male.select('Diastolic_blood_pressure_mmHg').toPandas(
['Diastolic_blood_pressure_mmHg']).to_numpy()

# Extracting data for females
a = blood_pressure_data_female.select('age').toPandas()['age'].to_numpy()
b =
blood_pressure_data_female.select('Systolic_blood_pressure_mmHg').toPandas(
)['Systolic_blood_pressure_mmHg'].to_numpy()
c =
blood_pressure_data_female.select('Diastolic_blood_pressure_mmHg').toPandas(
)['Diastolic_blood_pressure_mmHg'].to_numpy()

# Plotting
plt.figure(figsize=(10, 6))

# Plot for Systolic Blood Pressure
plt.plot(x, y, label='Male - SBP', marker='o')
plt.plot(a, b, label='Female - SBP', marker='o')

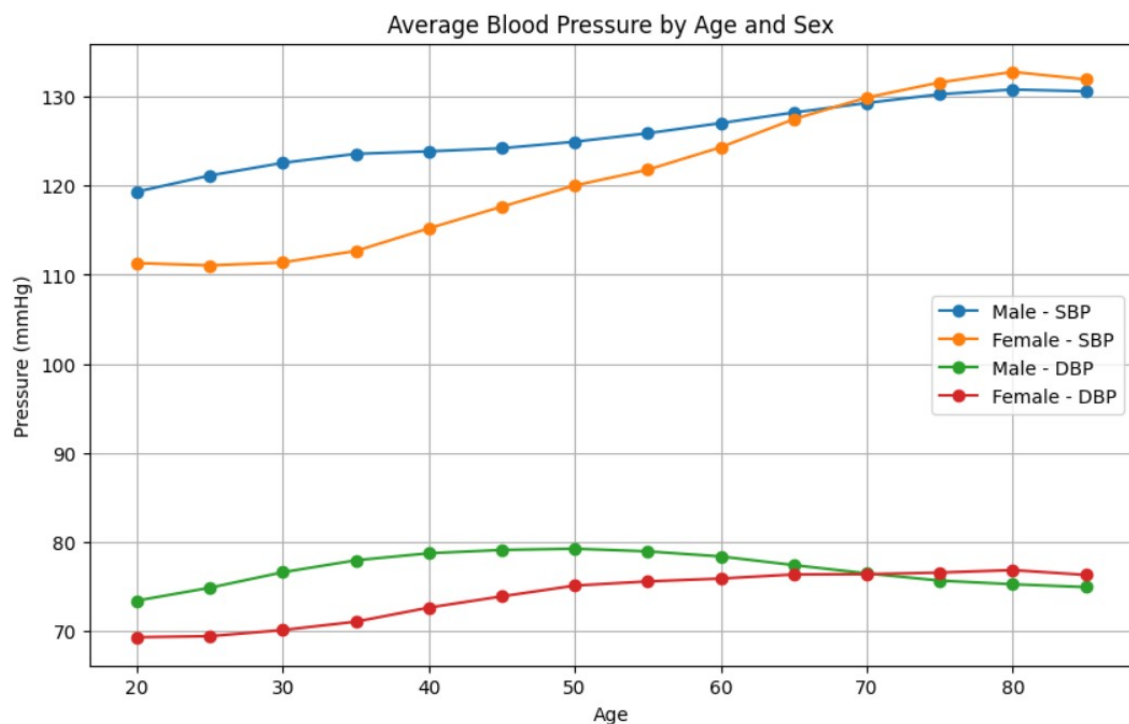
# Plot for Diastolic Blood Pressure
plt.plot(x, z, label='Male - DBP', marker='o')
plt.plot(a, c, label='Female - DBP', marker='o')

# Customize the plot
plt.title('Average Blood Pressure by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Pressure (mmHg)')
plt.legend()
plt.grid(True)
plt.show()

```

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Female	20	12661	111.31	69.3
Female	25	29618	111.04	69.42
Female	30	28701	111.38	70.11
Female	35	26468	112.67	71.05
Female	40	60071	115.21	72.63
Female	45	54655	117.64	73.89
Female	50	64610	120.02	75.1
Female	55	55286	121.79	75.57
Female	60	55036	124.31	75.89
Female	65	26388	127.46	76.35
Female	70	26717	129.89	76.38
Female	75	13932	131.59	76.56
Female	80	8547	132.77	76.84
Female	85	2241	131.93	76.3

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Male	20	9310	119.32	73.41
Male	25	34752	121.14	74.86
Male	30	48899	122.56	76.6
Male	35	58258	123.57	77.94
Male	40	70314	123.85	78.73
Male	45	63700	124.2	79.1
Male	50	64824	124.93	79.24
Male	55	55937	125.88	78.94
Male	60	51027	127.02	78.37
Male	65	26573	128.21	77.4
Male	70	23949	129.28	76.47
Male	75	11401	130.26	75.66
Male	80	6421	130.79	75.25
Male	85	1050	130.59	74.93



sex	age	number_of_people	average_weight
Female	20	12661	54.1
Female	25	29618	53.92
Female	30	28701	54.96
Female	35	26468	55.78
Female	40	60071	56.13
Female	45	54655	56.35
Female	50	64610	56.25
Female	55	55286	56.03
Female	60	55036	56.09
Female	65	26388	55.73
Female	70	26717	55.01
Female	75	13932	53.11
Female	80	8547	50.59
Female	85	2241	46.11

sex	age	number_of_people	average_weight
Male	20	9310	69.06
Male	25	34752	72.54
Male	30	48899	74.43
Male	35	58258	74.54
Male	40	70314	72.95
Male	45	63700	71.38
Male	50	64824	69.44
Male	55	55937	67.86
Male	60	51027	66.5
Male	65	26573	64.8
Male	70	23949	63.8
Male	75	11401	62.12
Male	80	6421	59.86
Male	85	1050	56.95

```
weight_data_female = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(weight), 2) AS average_weight
FROM S_D_Data
WHERE sex = 'Female'
GROUP BY sex, age
ORDER BY age ASC

""")
weight_data_female.show(30)

weight_data_male = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(weight), 2) AS average_weight
FROM S_D_Data
WHERE sex = 'Male'
GROUP BY sex, age
```

```

ORDER BY age ASC

""")
weight_data_male.show(30)

#MATPLOTLIB

# blood_pressure_data_male.show(15)
# for males
x = weight_data_male.select('age').toPandas()['age'].to_numpy()
y = weight_data_male.select('average_weight').toPandas()
['average_weight'].to_numpy()

# Extracting data for females
a = weight_data_female.select('age').toPandas()['age'].to_numpy()
b = weight_data_female .select('average_weight').toPandas()
['average_weight'].to_numpy()

# Plotting
plt.figure(figsize=(10, 6))

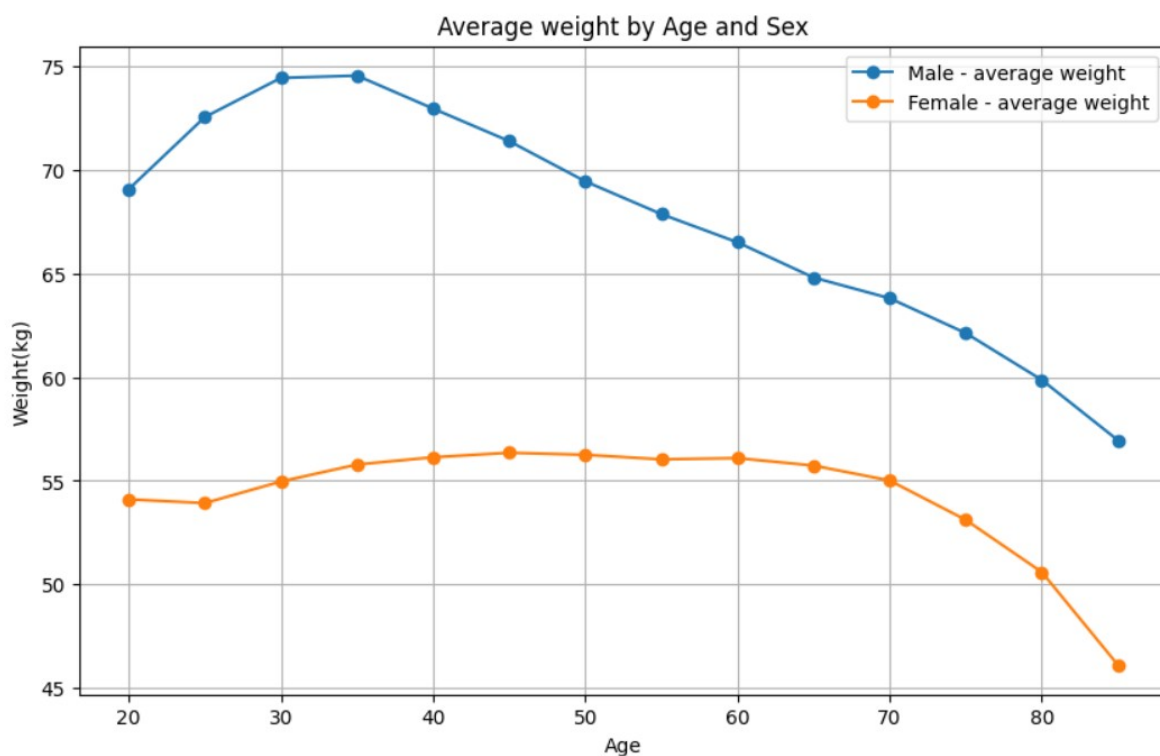
# Plot for weight
plt.plot(x, y, label='Male - average weight', marker='o')
plt.plot(a, b, label='Female - average weight', marker='o')

# Customize the plot
plt.title('Average weight by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Weight(kg)')
plt.legend()
plt.grid(True)
plt.show()

```

sex	age	number_of_people	average_weight
Female	20	12661	54.1
Female	25	29618	53.92
Female	30	28701	54.96
Female	35	26468	55.78
Female	40	60071	56.13
Female	45	54655	56.35
Female	50	64610	56.25
Female	55	55286	56.03
Female	60	55036	56.09
Female	65	26388	55.73
Female	70	26717	55.01
Female	75	13932	53.11
Female	80	8547	50.59
Female	85	2241	46.11

sex	age	number_of_people	average_weight
Male	20	9310	69.06
Male	25	34752	72.54
Male	30	48899	74.43
Male	35	58258	74.54
Male	40	70314	72.95
Male	45	63700	71.38
Male	50	64824	69.44
Male	55	55937	67.86
Male	60	51027	66.5
Male	65	26573	64.8
Male	70	23949	63.8
Male	75	11401	62.12
Male	80	6421	59.86
Male	85	1050	56.95



```

cholesterol_data_female = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`BLDS`), 2) AS BLDS_fasting_blood_glucose_mgdL,
       ROUND(AVG(`tot_chole`), 2) AS total_cholesterol_mgdL,
       ROUND(AVG(`HDL_chole`), 2) AS HDL_cholesterol_mgdL,
       ROUND(AVG(`LDL_chole`), 2) AS LDL_cholesterol_mgdL,
       ROUND(AVG(`triglyceride`), 2) AS triglyceride_mgdL,
       ROUND(AVG(`serum_creatinine`), 2) AS serum_creatinine_mgdL
FROM S_D_Data
WHERE sex = 'Female'
GROUP BY sex, age
ORDER BY age ASC
""")

```

```

cholesterol_data_female.show(15)

cholesterol_data_male = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`BLDS`), 2) AS BLDS_fasting_blood_glucose_mgdL,
       ROUND(AVG(`tot_chole`), 2) AS total_cholesterol_mgdL,
       ROUND(AVG(`HDL_chole`), 2) AS HDL_cholesterol_mgdL,
       ROUND(AVG(`LDL_chole`), 2) AS LDL_cholesterol_mgdL,
       ROUND(AVG(`triglyceride`), 2) AS triglyceride_mgdL,
       ROUND(AVG(`serum_creatinine`), 2) AS serum_creatinine_mgdL
FROM S_D_Data
WHERE sex = 'Male'
GROUP BY sex, age
ORDER BY age ASC
""")
cholesterol_data_male.show(15)

#MATPLOTLIB

# for males
x = cholesterol_data_male.select('age').toPandas()['age'].to_numpy()
y =
cholesterol_data_male.select('BLDS_fasting_blood_glucose_mgdL').toPandas()
['BLDS_fasting_blood_glucose_mgdL'].to_numpy()
z = cholesterol_data_male.select('total_cholesterol_mgdL').toPandas()
['total_cholesterol_mgdL'].to_numpy()
m = cholesterol_data_male.select('HDL_cholesterol_mgdL').toPandas()
['HDL_cholesterol_mgdL'].to_numpy()
n = cholesterol_data_male.select('LDL_cholesterol_mgdL').toPandas()
['LDL_cholesterol_mgdL'].to_numpy()
o = cholesterol_data_male.select('triglyceride_mgdL').toPandas()
['triglyceride_mgdL'].to_numpy()
p = cholesterol_data_male.select('serum_creatinine_mgdL').toPandas()
['serum_creatinine_mgdL'].to_numpy()
# Extracting data for females
a = cholesterol_data_female.select('age').toPandas()['age'].to_numpy()
b =
cholesterol_data_female.select('BLDS_fasting_blood_glucose_mgdL').toPandas()
['BLDS_fasting_blood_glucose_mgdL'].to_numpy()
c = cholesterol_data_female.select('total_cholesterol_mgdL').toPandas()
['total_cholesterol_mgdL'].to_numpy()
d = cholesterol_data_female.select('HDL_cholesterol_mgdL').toPandas()
['HDL_cholesterol_mgdL'].to_numpy()
e = cholesterol_data_female.select('LDL_cholesterol_mgdL').toPandas()
['LDL_cholesterol_mgdL'].to_numpy()
f = cholesterol_data_female.select('triglyceride_mgdL').toPandas()
['triglyceride_mgdL'].to_numpy()
g = cholesterol_data_female.select('serum_creatinine_mgdL').toPandas()
['serum_creatinine_mgdL'].to_numpy()
# Plotting
plt.figure(figsize=(10, 6))

```

```

# Plot for BLDS
plt.plot(x, y, label='Male - BLDS', marker='o')
plt.plot(a, b, label='Female - BLDS', marker='o')

plt.title('Average BLDS by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL) ')
plt.legend()
plt.grid(True)
plt.show()

# Plot for total_cholesterol
plt.plot(x, z, label='Male - total_cholesterol', marker='o')
plt.plot(a, c, label='Female - total_cholesterol', marker='o')

plt.title('Average total_cholesterol by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL) ')
plt.legend()
plt.grid(True)
plt.show()

# Plot for HDL
plt.plot(x, m, label='Male - HDL', marker='o')
plt.plot(a, d, label='Female - HDL', marker='o')

plt.title('Average HDL by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL) ')
plt.legend()
plt.grid(True)
plt.show()

# Plot for LDL
plt.plot(x, n, label='Male - LDL', marker='o')
plt.plot(a, e, label='Female - LDL', marker='o')
plt.title('Average LDL by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL) ')
plt.legend()
plt.grid(True)
plt.show()

# Plot for triglyceride
plt.plot(x, o, label='Male - triglyceride', marker='o')
plt.plot(a, f, label='Female - triglyceride', marker='o')

plt.title('Average triglyceride by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL) ')

```



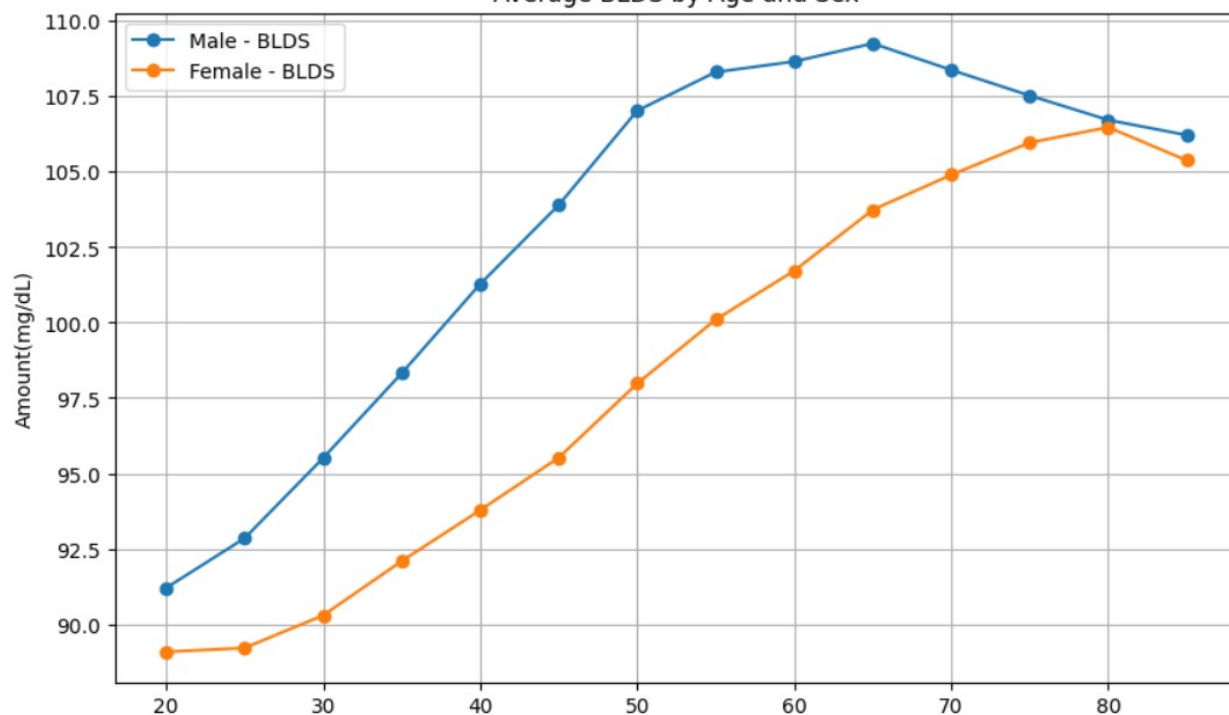
```
plt.legend()
plt.grid(True)
plt.show()

# Plot for serum_creatinine
plt.plot(x, p, label='Male - serum_creatinine', marker='o')
plt.plot(a, g, label='Female - serum_creatinine', marker='o')

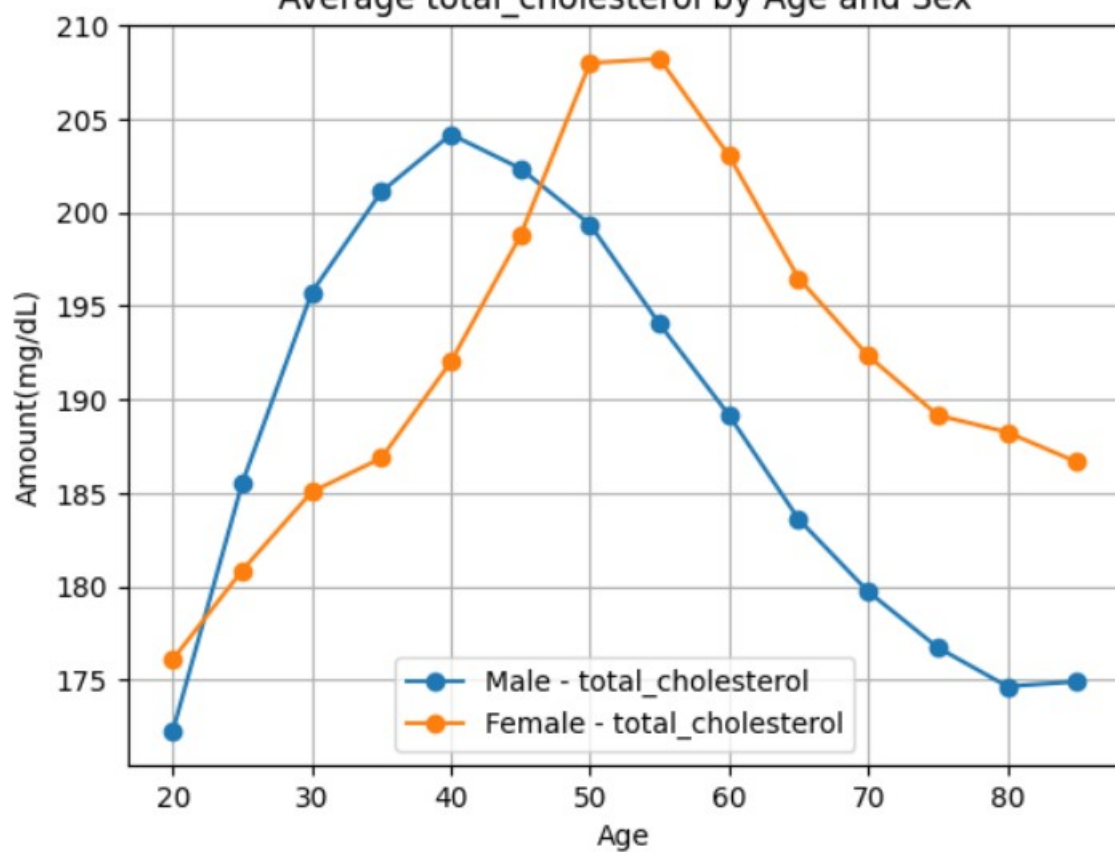
plt.title('Average serum_creatinine by Age and Sex')
plt.xlabel('Age')
plt.ylabel('Amount (mg/dL)')
plt.legend()
plt.grid(True)
plt.show()
```

sex	age	number_of_people	BLDS_fasting_blood_glucose_mgdl	total_cholesterol_mgdl	HDL_cholesterol_mgdl	LDL_cholesterol_mgdl	triglyceride_mgdl	serum_creatinine_mgdl
Female	20	12661	89.08	185.05	66.56	101.54	74.91	0.7
Female	25	29618	89.21	186.86	64.62	104.25	77.17	0.7
Female	30	28701	90.29	192.01	63.48	109.39	85.94	0.7
Female	35	26468	92.09	198.76	62.75	115.28	90.61	0.71
Female	40	60071	93.78	198.76	62.75	115.28	96.33	0.72
Female	45	54655	95.52	207.99	61.96	123.23	104.02	0.73
Female	50	64610	97.98	208.21	60.22	124.07	114.81	0.73
Female	55	55286	100.1	203.03	58.55	119.72	119.86	0.74
Female	60	55036	101.72	196.48	57.21	114.22	124.21	0.74
Female	65	26388	103.74	192.33	56.21	110.29	126.03	0.75
Female	70	26717	104.9	189.15	54.96	107.83	129.12	0.78
Female	75	13932	105.97	188.24	54.1	107.47	131.86	0.8
Female	80	8547	106.48	186.65	52.3	107.93	133.72	0.85
Female	85	2241	105.38				131.66	0.94
Male	20	9310	91.2	172.26	56.67	96.08	99.83	0.95
Male	25	34752	92.85	185.54	55.9	106.19	122.01	0.96
Male	30	48899	95.51	195.73	53.69	114.18	147.91	0.96
Male	35	58258	98.32	201.11	52.53	117.75	164.86	0.95
Male	40	70314	101.28	204.17	52.44	119.55	171.56	0.96
Male	45	63700	103.9	202.35	52.51	117.77	171.06	0.96
Male	50	64824	107.03	199.35	52.44	115.98	163.68	0.97
Male	55	55937	108.31	194.03	52.43	112.2	153.78	0.97
Male	60	51027	108.66	189.15	52.36	108.62	144.63	0.97
Male	65	26573	109.26	183.59	52.24	104.75	134.97	0.99
Male	70	23949	108.38	179.72	51.78	102.59	127.76	1.01
Male	75	11401	107.53	176.73	51.34	100.81	123.73	1.06
Male	80	6421	106.72	174.65	50.69	100.2	118.57	1.09
Male	85	1050	106.22	174.9	51.05	100.98	113.49	1.15

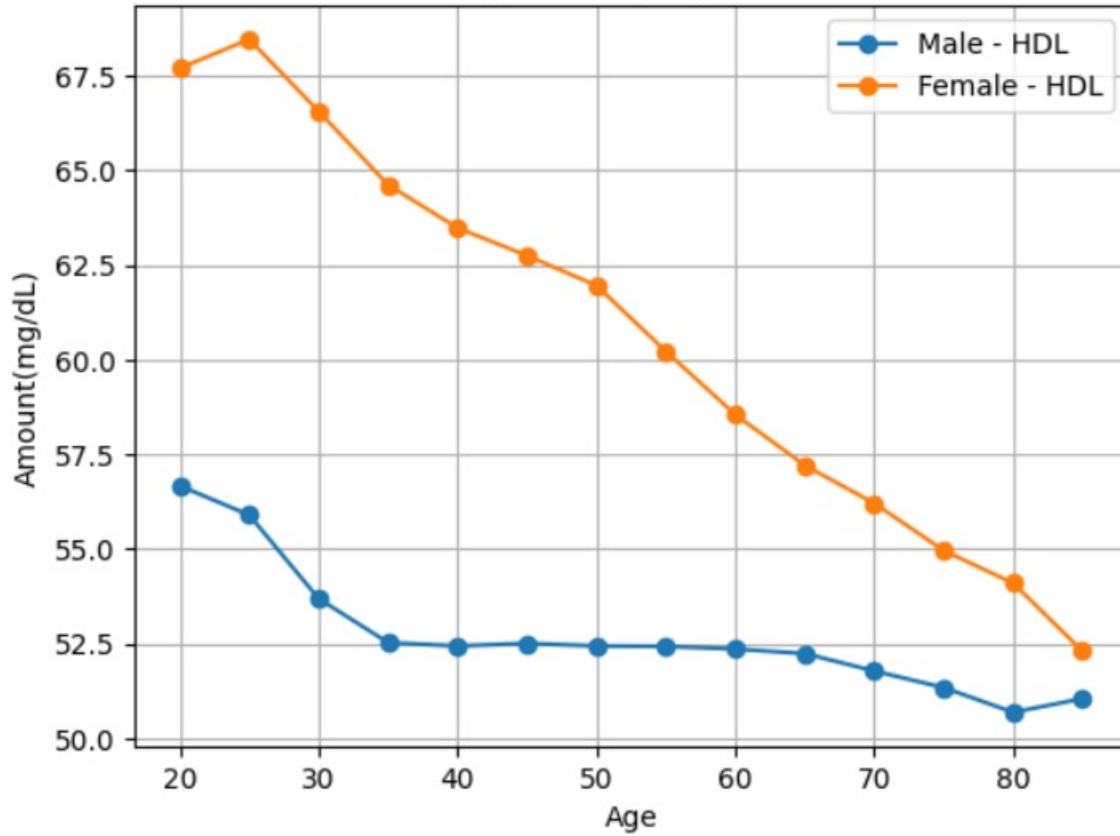
Average BLDS by Age and Sex



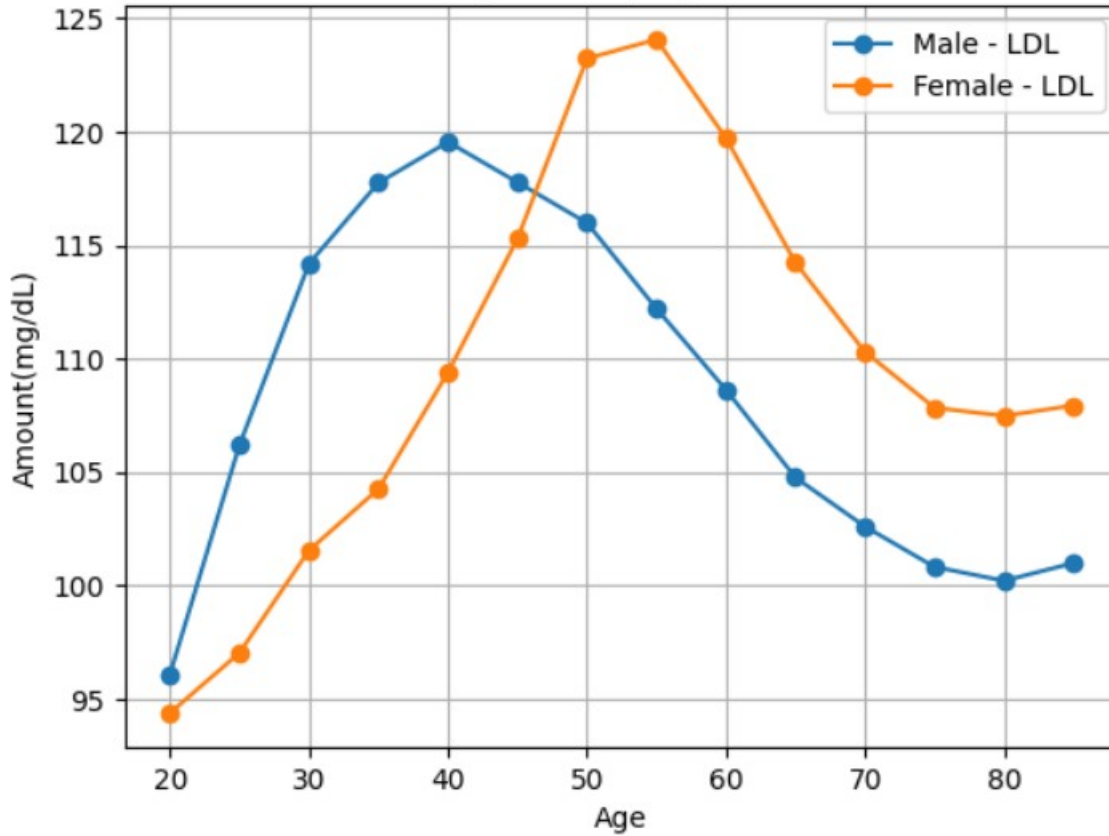
Average total_cholesterol by Age and Sex



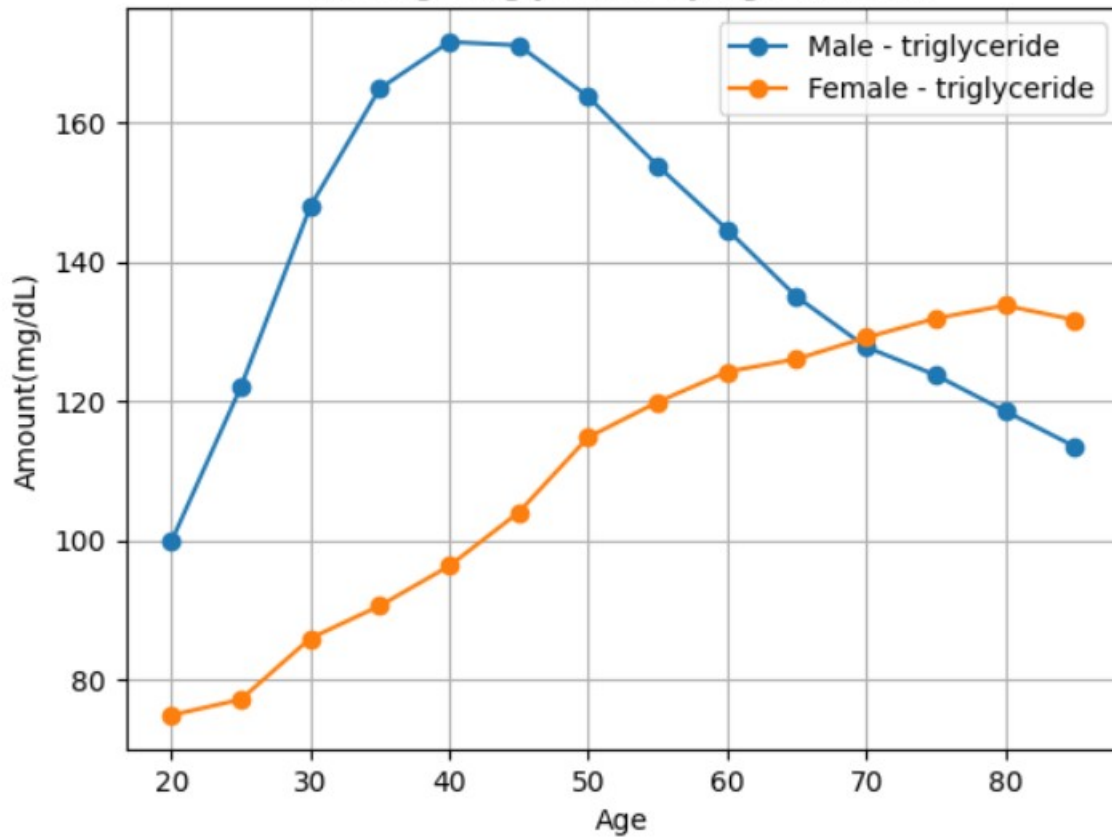
Average HDL by Age and Sex



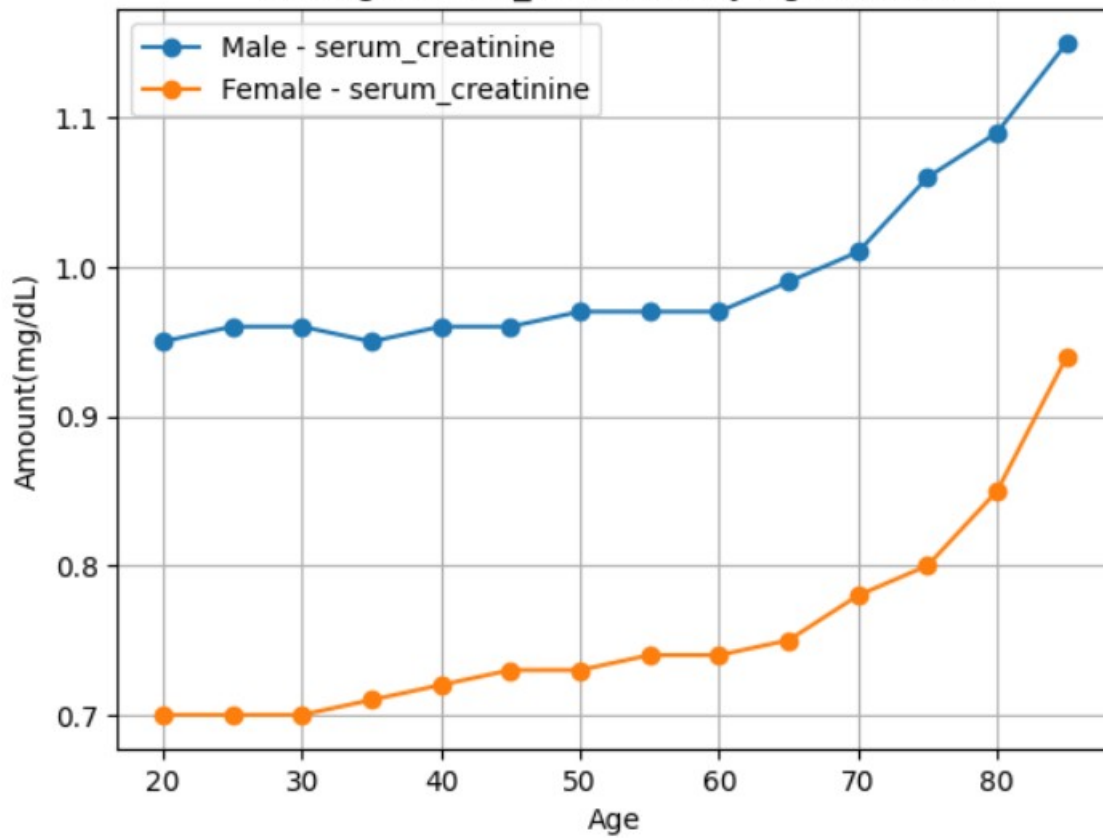
Average LDL by Age and Sex



Average triglyceride by Age and Sex



Average serum_creatinine by Age and Sex

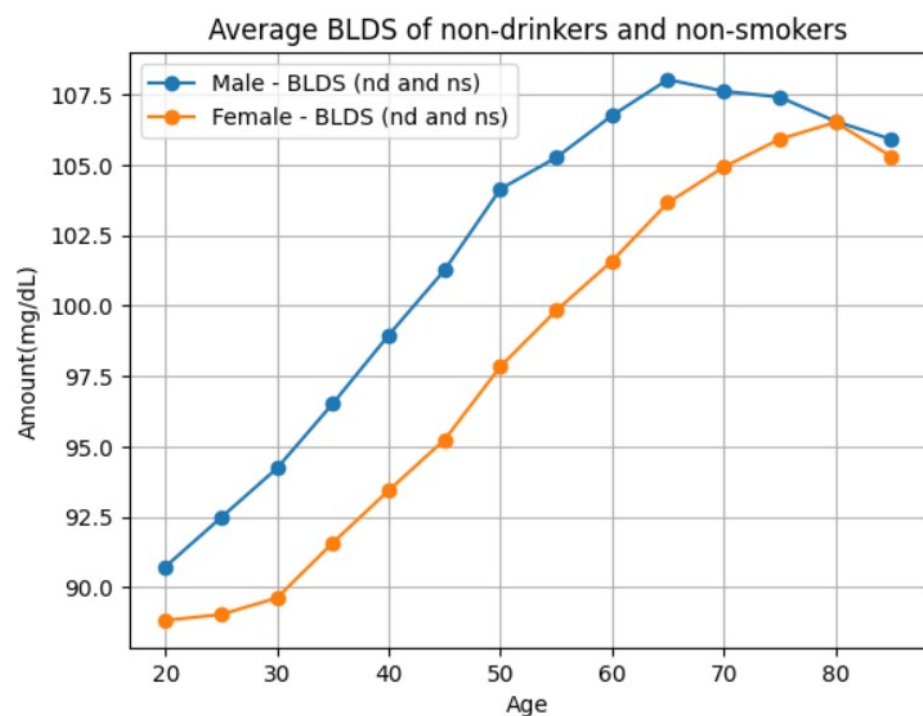
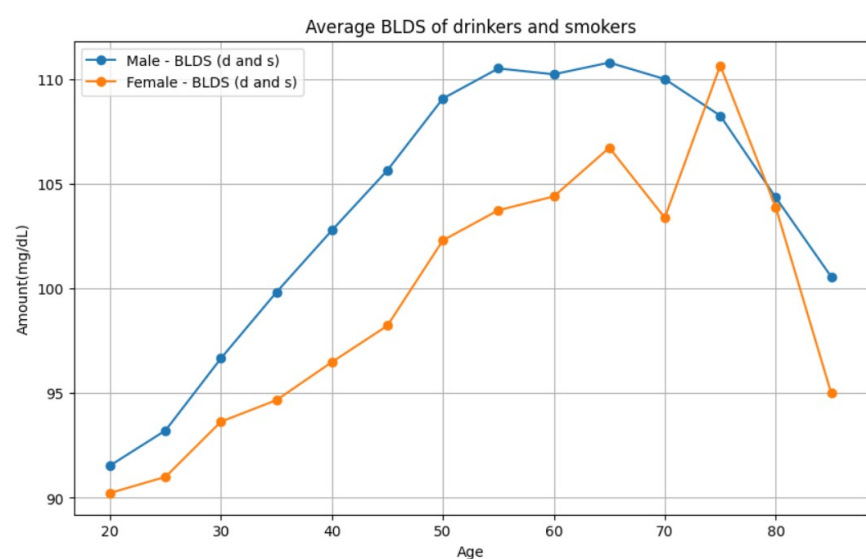


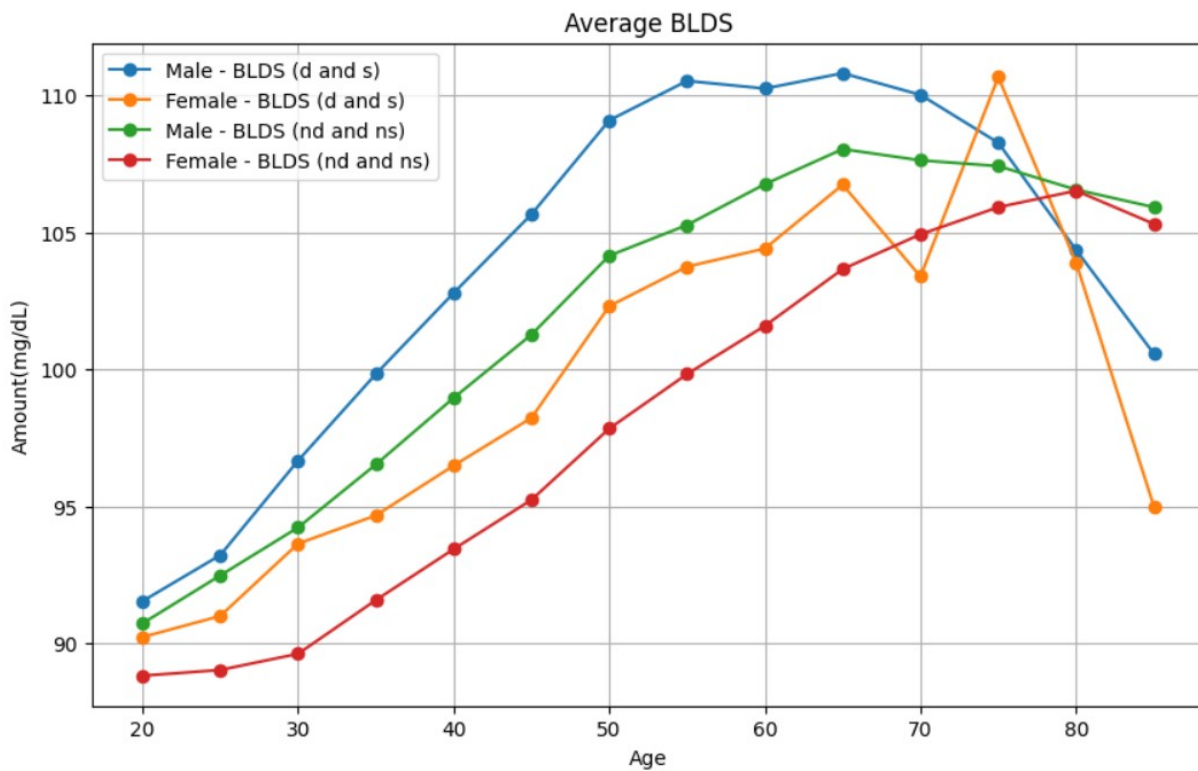
sex	age	number_of_people	BLDS_s_and_d_mgdL
Female	20	931	90.23
Female	25	1261	91.0
Female	30	1044	93.64
Female	35	849	94.67
Female	40	1997	96.49
Female	45	1293	98.23
Female	50	1254	102.31
Female	55	723	103.75
Female	60	532	104.41
Female	65	107	106.74
Female	70	66	103.39
Female	75	21	110.67
Female	80	18	103.89
Female	85	2	95.0

sex	age	number_of_people	BLDS_ns_and_nd_mgdL
Female	20	4583	88.82
Female	25	11203	89.03
Female	30	13085	89.62
Female	35	13226	91.58
Female	40	32446	93.44
Female	45	33720	95.24
Female	50	44738	97.84
Female	55	42887	99.83
Female	60	45678	101.59
Female	65	23521	103.66
Female	70	24666	104.93
Female	75	13152	105.92
Female	80	8144	106.52
Female	85	2131	105.31

sex	age	number_of_people	BLDS_s_and_d_mgdL
Male	20	3371	91.54
Male	25	11868	93.22
Male	30	16476	96.67
Male	35	21793	99.83
Male	40	27042	102.79
Male	45	22041	105.66
Male	50	19422	109.09
Male	55	13465	110.53
Male	60	9519	110.25
Male	65	3472	110.81
Male	70	2078	110.02
Male	75	654	108.27
Male	80	255	104.36
Male	85	23	100.57

sex	age	number_of_people	BLDS_ns_and_nd_mgdL
Male	20	1953	90.73
Male	25	5556	92.48
Male	30	6819	94.23
Male	35	6633	96.53
Male	40	7004	98.96
Male	45	6451	101.27
Male	50	7197	104.15
Male	55	7392	105.27
Male	60	8065	106.76
Male	65	5519	108.04
Male	70	6301	107.63
Male	75	3799	107.42
Male	80	2630	106.56
Male	85	511	105.92





```
# blood pressure for drinkers and smokers female
```

```
blood_pressure_data_s_and_d_female = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Female' AND SMK_stat_type_cd = 3 AND DRK_YN = 'Y'
GROUP BY sex, age
ORDER BY age ASC
""")
blood_pressure_data_s_and_d_female.show(15)
```

```
# blood pressure for non_drinkers and non_smokers female
```

```
blood_pressure_data_ns_and_nd_female = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Female' AND SMK_stat_type_cd = 1 AND DRK_YN = 'N'
GROUP BY sex, age
ORDER BY age ASC
""")
blood_pressure_data_ns_and_nd_female.show(15)
```

```
# blood pressure for drinkers and smokers male
```

```

blood_pressure_data_s_and_d_male = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Male' AND SMK_stat_type_cd = 3 AND DRK_YN = 'Y'
GROUP BY sex, age
ORDER BY age ASC
""")
blood_pressure_data_s_and_d_male.show(15)

# blood pressure for non-drinkers and non-smokers male

blood_pressure_data_ns_and_nd_male = spark.sql("""
SELECT sex, age, COUNT(*) as number_of_people,
       ROUND(AVG(`SBP`), 2) AS Systolic_blood_pressure_mmHg,
       ROUND(AVG(`DBP`), 2) AS Diastolic_blood_pressure_mmHg
FROM S_D_Data
WHERE sex = 'Male' AND SMK_stat_type_cd = 1 AND DRK_YN = 'N'
GROUP BY sex, age
ORDER BY age ASC
""")
blood_pressure_data_ns_and_nd_male.show(15)

```

	sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
	Female	20	931	112.32	70.15
	Female	25	1261	113.21	71.6
	Female	30	1044	115.32	73.66
	Female	35	849	114.9	73.3
	Female	40	1997	117.49	74.94
	Female	45	1293	119.32	76.22
	Female	50	1254	120.15	76.14
	Female	55	723	122.3	76.59
	Female	60	532	124.23	76.18
	Female	65	107	125.57	75.22
	Female	70	66	130.39	76.98
	Female	75	21	129.38	75.86
	Female	80	18	131.39	75.33
	Female	85	2	124.5	61.5

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Female	20	4583	111.12	69.17
Female	25	11203	110.75	69.15
Female	30	13085	110.86	69.56
Female	35	13226	112.21	70.59
Female	40	32446	114.67	72.09
Female	45	33720	117.19	73.4
Female	50	44738	119.82	74.82
Female	55	42887	121.67	75.39
Female	60	45678	124.32	75.78
Female	65	23521	127.52	76.32
Female	70	24666	129.93	76.36
Female	75	13152	131.62	76.55
Female	80	8144	132.83	76.9
Female	85	2131	131.9	76.25

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Male	20	3371	119.81	74.03
Male	25	11868	121.57	75.52
Male	30	16476	123.54	77.52
Male	35	21793	124.52	78.78
Male	40	27042	124.52	79.32
Male	45	22041	124.81	79.62
Male	50	19422	125.52	79.71
Male	55	13465	126.3	79.18
Male	60	9519	127.39	78.34
Male	65	3472	128.45	77.12
Male	70	2078	129.06	75.96
Male	75	654	129.6	75.2
Male	80	255	129.77	74.94
Male	85	23	128.91	78.0

sex	age	number_of_people	Systolic_blood_pressure_mmHg	Diastolic_blood_pressure_mmHg
Male	20	1953	118.58	72.91
Male	25	5556	119.64	73.89
Male	30	6819	120.88	75.34
Male	35	6633	121.74	76.47
Male	40	7004	122.28	77.31
Male	45	6451	122.37	77.75
Male	50	7197	123.48	78.07
Male	55	7392	124.75	78.14
Male	60	8065	126.18	77.89
Male	65	5519	127.85	77.32
Male	70	6301	129.02	76.57
Male	75	3799	130.35	75.82
Male	80	2630	130.97	75.75
Male	85	511	129.87	74.6

```

#MATPLOTLIB

# for males

x = blood_pressure_data_s_and_d_male.select('age').toPandas()
['age'].to_numpy()
y =
blood_pressure_data_s_and_d_male.select('Systolic_blood_pressure_mmHg').to
Pandas()['Systolic_blood_pressure_mmHg'].to_numpy()
z =
blood_pressure_data_s_and_d_male.select('Diastolic_blood_pressure_mmHg').t
oPandas()['Diastolic_blood_pressure_mmHg'].to_numpy()
t =
blood_pressure_data_ns_and_nd_male.select('Systolic_blood_pressure_mmHg').
toPandas()['Systolic_blood_pressure_mmHg'].to_numpy()
u =
blood_pressure_data_ns_and_nd_male.select('Diastolic_blood_pressure_mmHg')
.toPandas()['Diastolic_blood_pressure_mmHg'].to_numpy()
# Extracting data for females
a = blood_pressure_data_s_and_d_female.select('age').toPandas()
['age'].to_numpy()
b =
blood_pressure_data_s_and_d_female.select('Systolic_blood_pressure_mmHg').
toPandas()['Systolic_blood_pressure_mmHg'].to_numpy()
c =
blood_pressure_data_s_and_d_female.select('Diastolic_blood_pressure_mmHg')
.toPandas()['Diastolic_blood_pressure_mmHg'].to_numpy()
d =
blood_pressure_data_ns_and_nd_female.select('Systolic_blood_pressure_mmHg'
).toPandas()['Systolic_blood_pressure_mmHg'].to_numpy()
e =
blood_pressure_data_ns_and_nd_female.select('Diastolic_blood_pressure_mmHg
').toPandas()['Diastolic_blood_pressure_mmHg'].to_numpy()

# Plotting
plt.figure(figsize=(10, 6))

# Plot for Systolic Blood Pressure
plt.plot(x, y, label='Male - SBP(s_and_d)', marker='o')
plt.plot(a, b, label='Female - SBP(s_and_d)', marker='o')
plt.plot(x, t, label='Male - SBP(ns_and_nd)', marker='o')
plt.plot(a, d, label='Female - SBP(ns_and_nd)', marker='o')

# Customize the plot
plt.title('SBP for s_and_d / ns_and_nd')
plt.xlabel('Age')
plt.ylabel('Pressure (mmHg)')
plt.legend()
plt.grid(True)
plt.show()

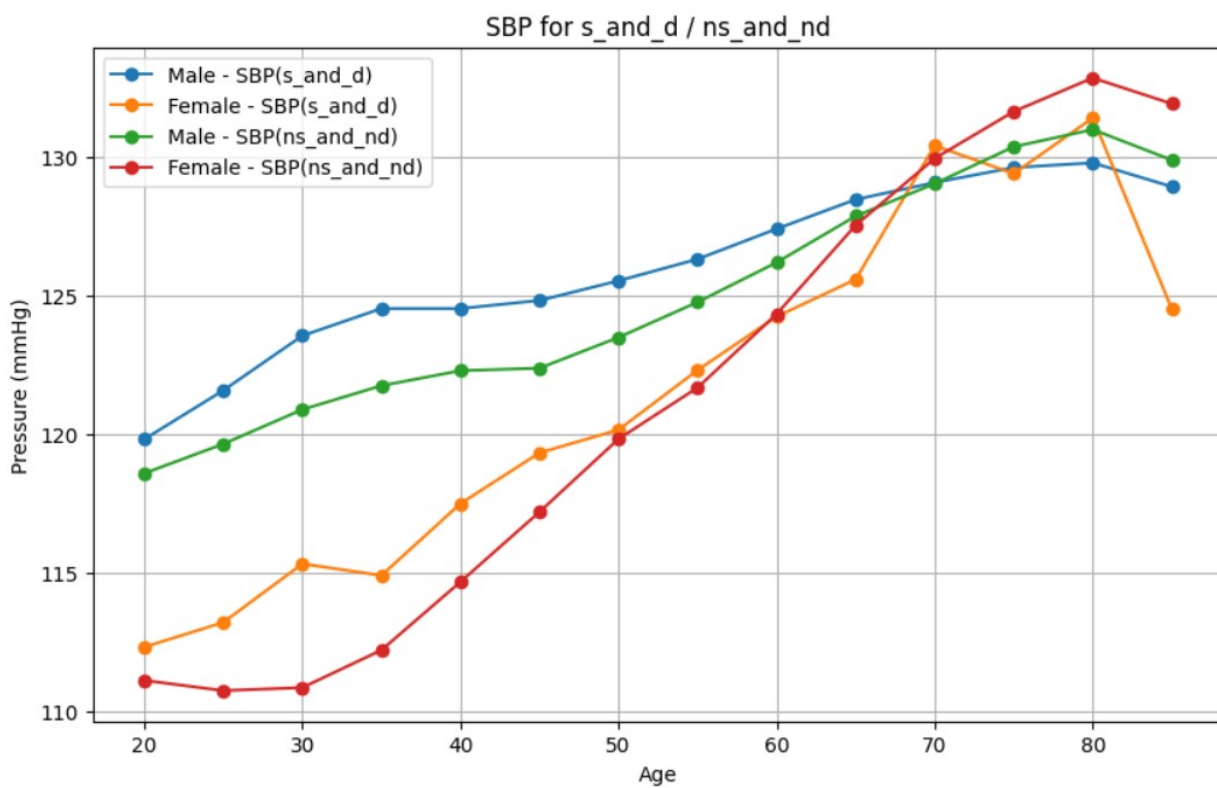
```

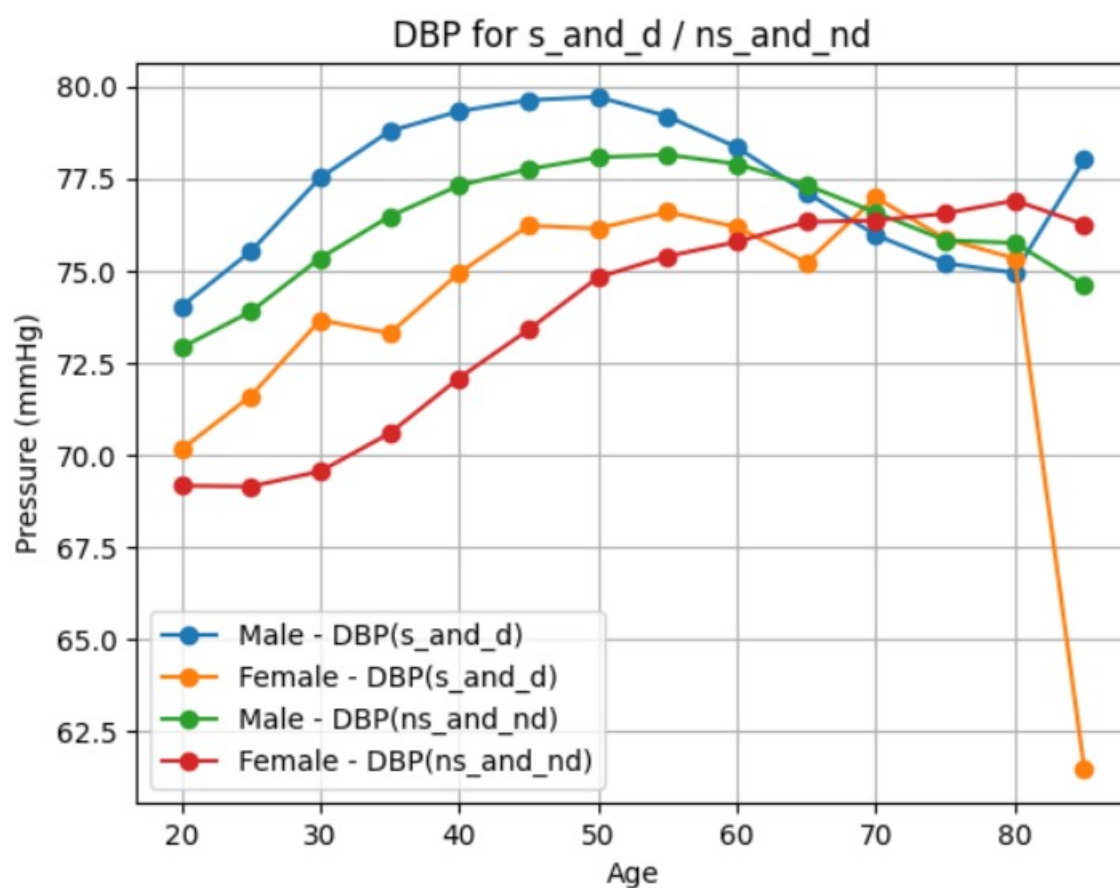
```

# Plot for Diastolic Blood Pressure
plt.plot(x, z, label='Male - DBP(s_and_d)', marker='o')
plt.plot(a, c, label='Female - DBP(s_and_d)', marker='o')
plt.plot(x, u, label='Male - DBP(ns_and_nd)', marker='o')
plt.plot(a, e, label='Female - DBP(ns_and_nd)', marker='o')

# Customize the plot
plt.title('DBP for s_and_d / ns_and_nd')
plt.xlabel('Age')
plt.ylabel('Pressure (mmHg)')
plt.legend()
plt.grid(True)
plt.show()

```



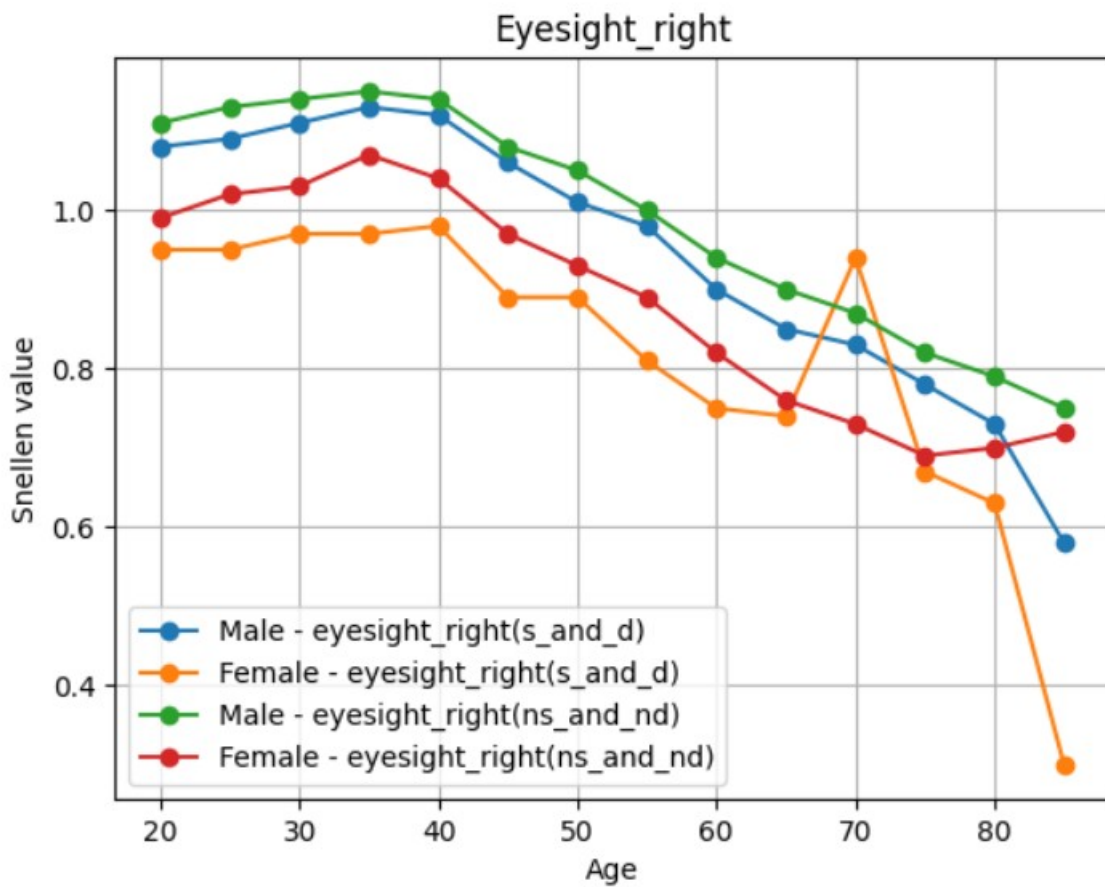
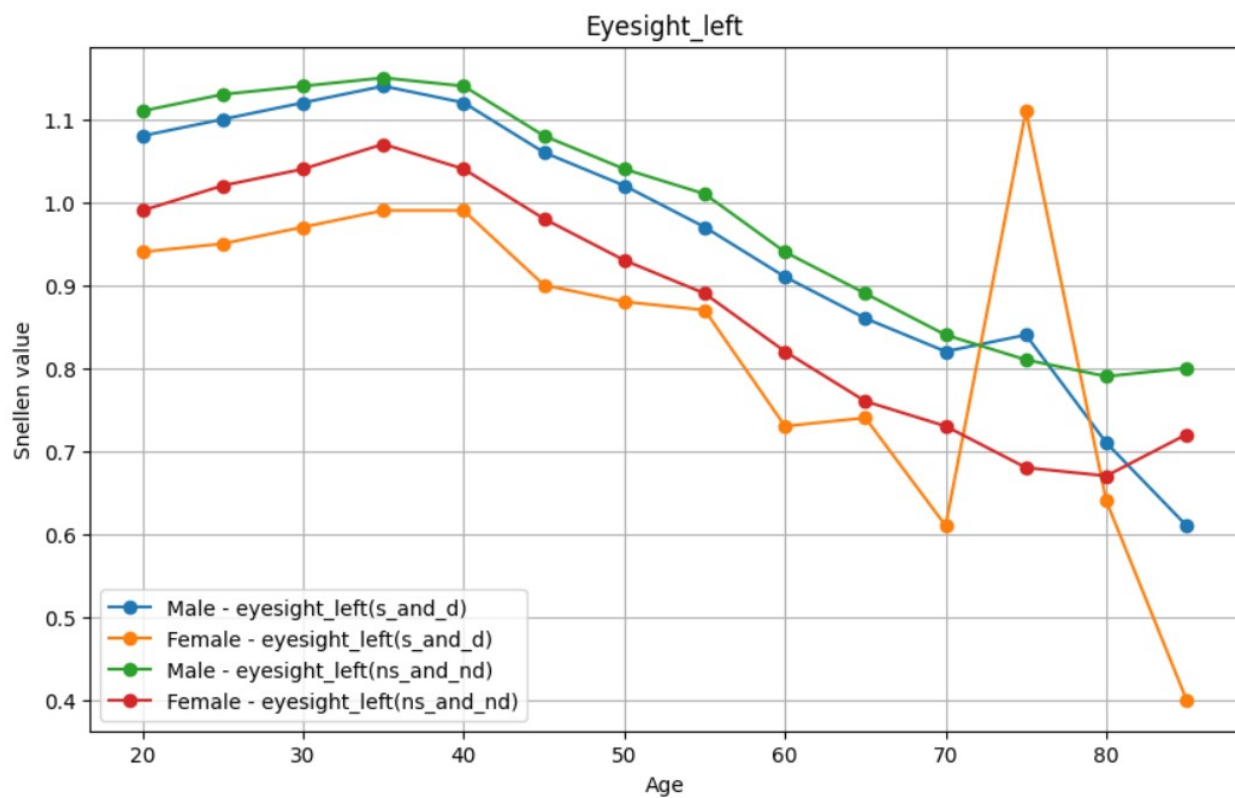


sex	age	number_of_people	Snellen_value_left	Snellen_value_right
Female	20	931	0.94	0.95
Female	25	1261	0.95	0.95
Female	30	1044	0.97	0.97
Female	35	849	0.99	0.97
Female	40	1997	0.99	0.98
Female	45	1293	0.9	0.89
Female	50	1254	0.88	0.89
Female	55	723	0.87	0.81
Female	60	532	0.73	0.75
Female	65	107	0.74	0.74
Female	70	66	0.61	0.94
Female	75	21	1.11	0.67
Female	80	18	0.64	0.63
Female	85	2	0.4	0.3

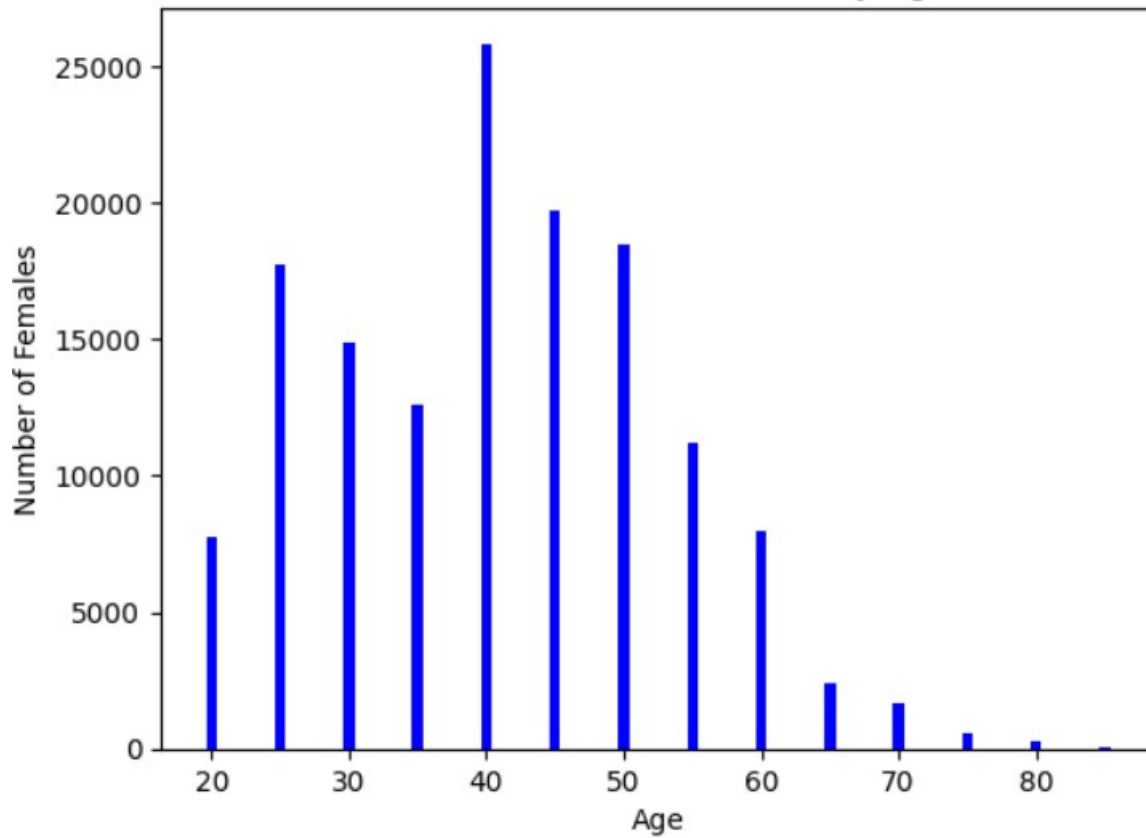
sex	age	number_of_people	Snellen_value_left	Snellen_value_right
Female	20	4583	0.99	0.99
Female	25	11203	1.02	1.02
Female	30	13085	1.04	1.03
Female	35	13226	1.07	1.07
Female	40	32446	1.04	1.04
Female	45	33720	0.98	0.97
Female	50	44738	0.93	0.93
Female	55	42887	0.89	0.89
Female	60	45678	0.82	0.82
Female	65	23521	0.76	0.76
Female	70	24666	0.73	0.73
Female	75	13152	0.68	0.69
Female	80	8144	0.67	0.7
Female	85	2131	0.72	0.72

sex	age	number_of_people	Snellen_value_left	Snellen_value_right
Male	20	3371	1.08	1.08
Male	25	11868	1.1	1.09
Male	30	16476	1.12	1.11
Male	35	21793	1.14	1.13
Male	40	27042	1.12	1.12
Male	45	22041	1.06	1.06
Male	50	19422	1.02	1.01
Male	55	13465	0.97	0.98
Male	60	9519	0.91	0.9
Male	65	3472	0.86	0.85
Male	70	2078	0.82	0.83
Male	75	654	0.84	0.78
Male	80	255	0.71	0.73
Male	85	23	0.61	0.58

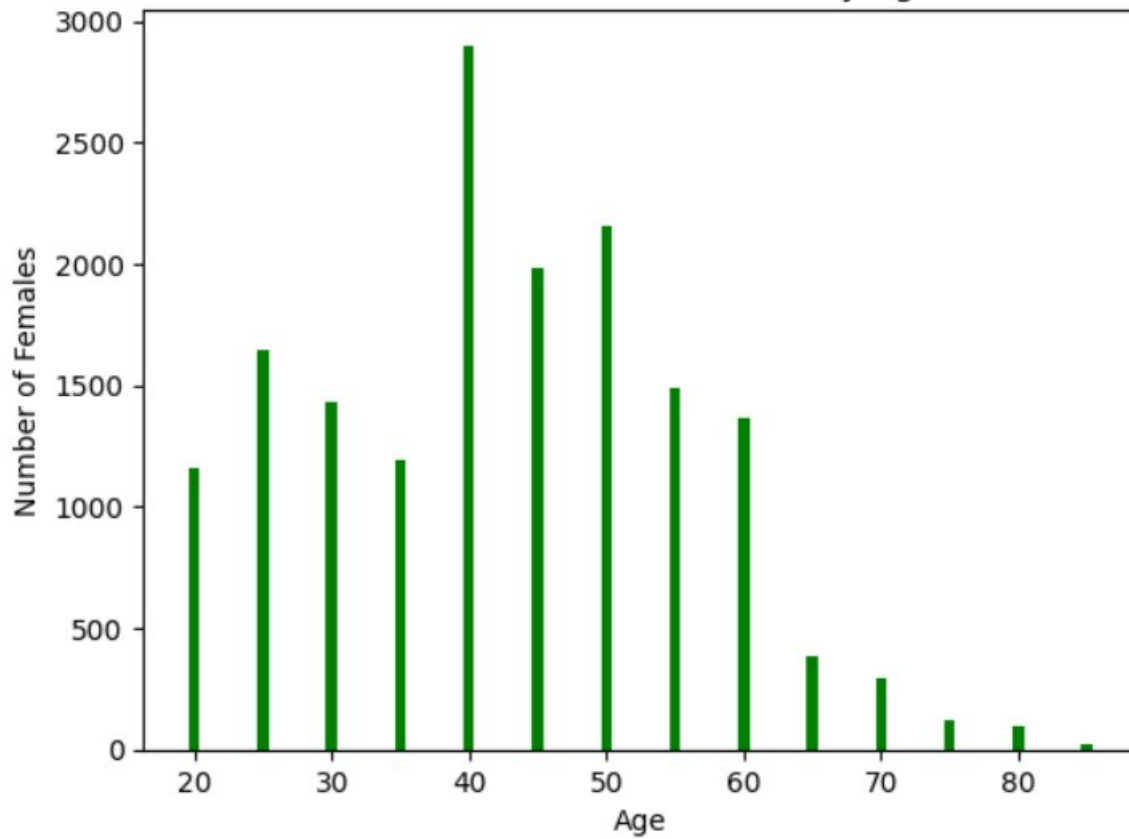
sex	age	number_of_people	Snellen_value_left	Snellen_value_right
Male	20	1953	1.11	1.11
Male	25	5556	1.13	1.13
Male	30	6819	1.14	1.14
Male	35	6633	1.15	1.15
Male	40	7004	1.14	1.14
Male	45	6451	1.08	1.08
Male	50	7197	1.04	1.05
Male	55	7392	1.01	1.0
Male	60	8065	0.94	0.94
Male	65	5519	0.89	0.9
Male	70	6301	0.84	0.87
Male	75	3799	0.81	0.82
Male	80	2630	0.79	0.79
Male	85	511	0.8	0.75



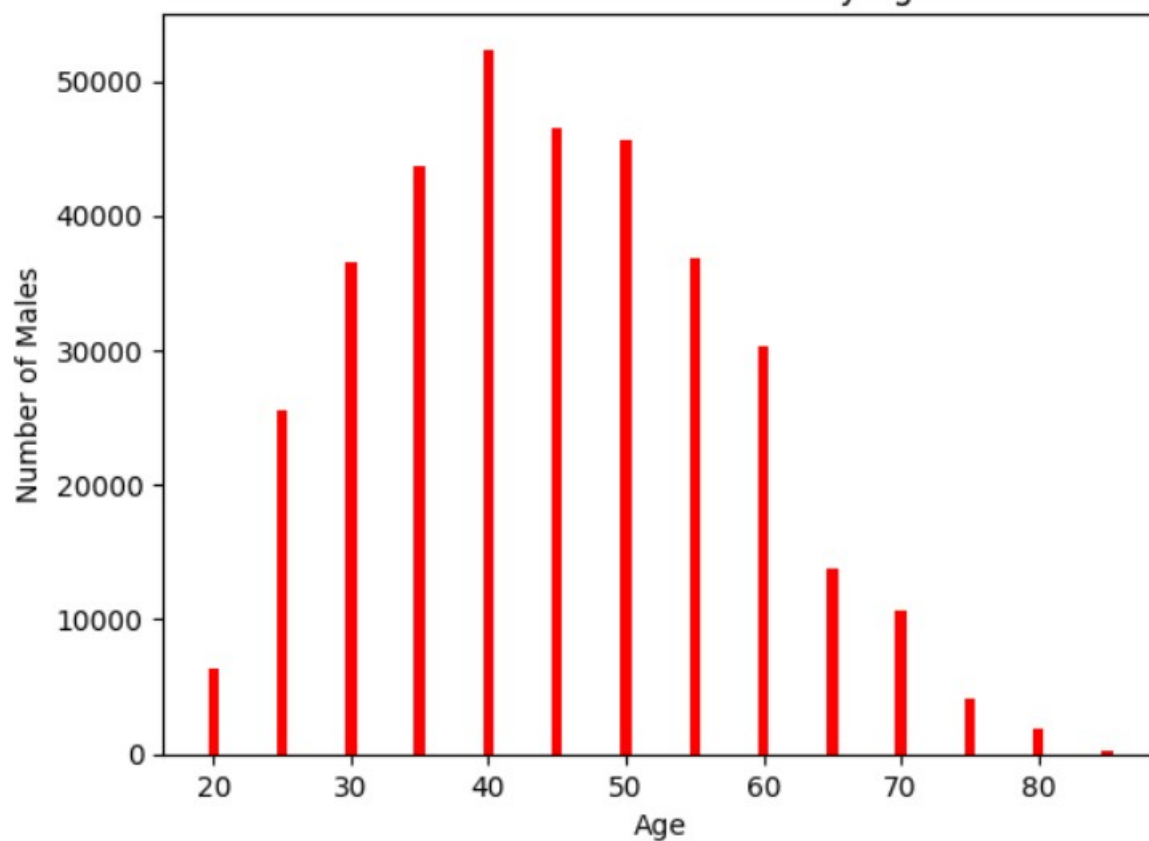
Number of Female Drinkers by Age



Number of Female Smokers by Age



Number of male Drinkers by Age



Number of male Smokers by Age

