LEON MARCO A. DEVELA

COM232
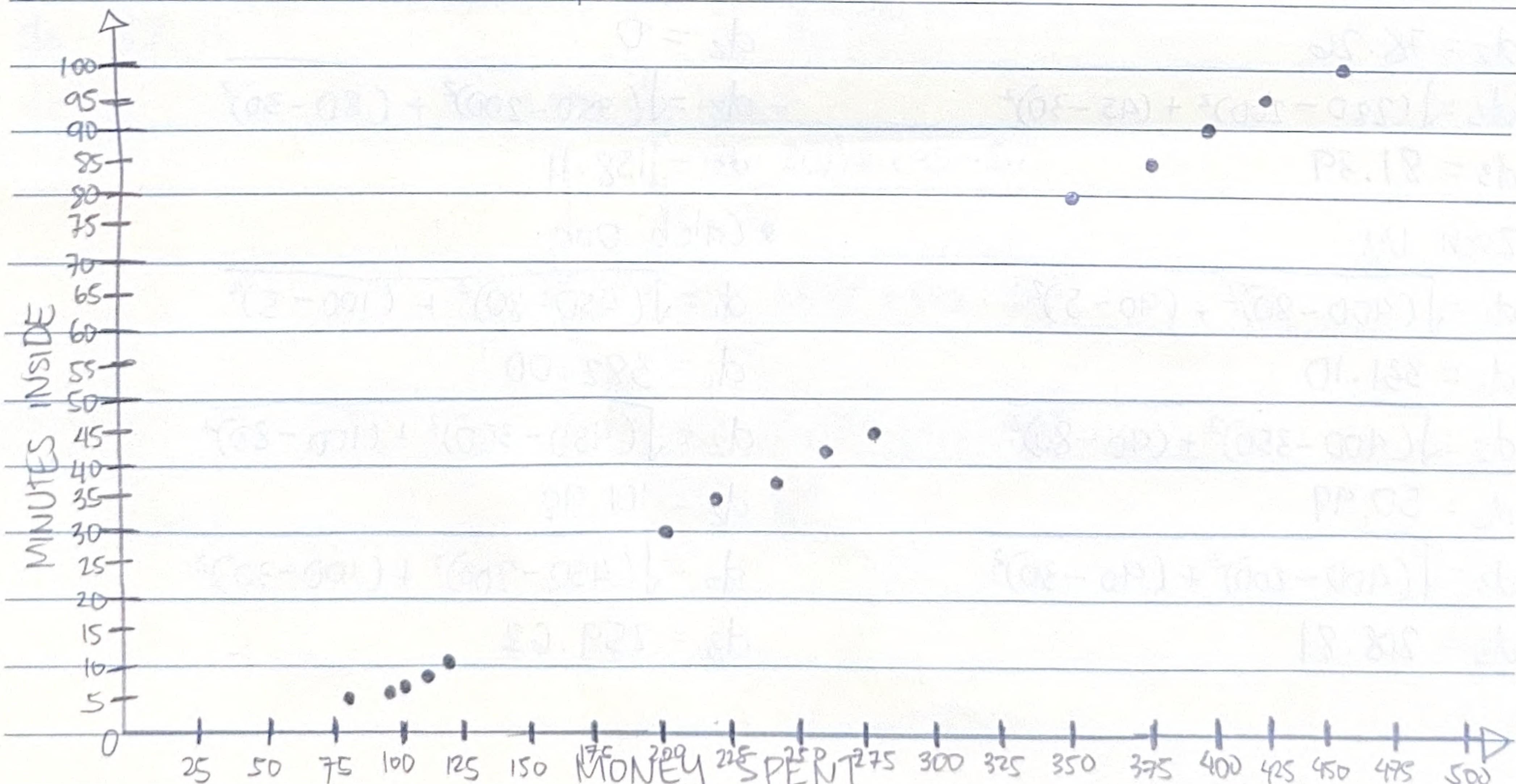
ADVANCED MACHINE LEARNING

ACTIVITY 1

GIVEN:

CUSTOMER NAME	MONEY SPENT	MINUTES INSIDE
Chloe Mendoza	380	85
Anna Reyes	80	5
Mika Tan	280	45
Zach Uy	400	90
Kevin Ramos	220	35
Sofia Dela Peña	110	9
Brian Lim	350	80
Caleb Ong	450	100
Liam Cruz	100	8
Ella Navaro	240	38
John Mercado	120	10
Jared Flores	260	42
Mark Santos	95	7
Paula Gomez	200	30
Hannah Roque	420	95

① Plot the data in a scatter plot

② Using the randomly selected initial clusters, calculate the distances of all data points using Euclidean Distance

- Chloe Mendoza

$$d_1 = \sqrt{(380-80)^2 + (85-5)^2}$$
$$d_1 = 310.48$$
$$d_2 = \sqrt{(380-350)^2 + (85-80)^2}$$
$$d_2 = 30.41$$
$$d_3 = \sqrt{(380-200)^2 + (85-30)^2}$$
$$d_3 = 188.22$$

- Anna Reyes

$$d_1 = \sqrt{(80-80)^2 + (5-5)^2}$$
$$d_1 = 0$$
$$d_2 = \sqrt{(80-350)^2 + (5-80)^2}$$
$$d_2 = 280.22$$
$$d_3 = \sqrt{(80-200)^2 + (5-30)^2}$$
$$d_3 = 122.58$$

- Mika Tan

$$d_1 = \sqrt{(280-80)^2 + (45-5)^2}$$
$$d_1 = 203.96$$
$$d_2 = \sqrt{(280-350)^2 + (45-80)^2}$$
$$d_2 = 78.26$$
$$d_3 = \sqrt{(280-200)^2 + (45-30)^2}$$
$$d_3 = 81.39$$

- Zach Uy

$$d_1 = \sqrt{(400-80)^2 + (90-5)^2}$$
$$d_1 = 331.10$$
$$d_2 = \sqrt{(400-350)^2 + (90-80)^2}$$
$$d_2 = 50.99$$
$$d_3 = \sqrt{(400-200)^2 + (90-30)^2}$$
$$d_3 = 208.81$$

- Kevin Ramos

$$d_1 = \sqrt{(220-80)^2 + (35-5)^2}$$
$$d_1 = 143.18$$
$$d_2 = \sqrt{(220-350)^2 + (35-80)^2}$$
$$d_2 = 137.57$$
$$d_3 = \sqrt{(220-200)^2 + (35-30)^2}$$
$$d_3 = 20.62$$

- Sofia Dela Peña

$$d_1 = \sqrt{(110-80)^2 + (9-5)^2}$$
$$d_1 = 30.27$$
$$d_2 = \sqrt{(110-350)^2 + (9-80)^2}$$
$$d_2 = 250.28$$
$$d_3 = \sqrt{(110-200)^2 + (9-30)^2}$$
$$d_3 = 92.42$$

- Brian Lim

$$d_1 = \sqrt{(350-80)^2 + (80-5)^2}$$
$$d_1 = 280.22$$
$$d_2 = \sqrt{(350-350)^2 + (80-80)^2}$$
$$d_2 = 0$$
$$d_3 = \sqrt{(350-200)^2 + (80-30)^2}$$
$$d_3 = 158.11$$

- Caleb Ong

$$d_1 = \sqrt{(450-80)^2 + (100-5)^2}$$
$$d_1 = 382.00$$
$$d_2 = \sqrt{(450-350)^2 + (100-80)^2}$$
$$d_2 = 101.98$$
$$d_3 = \sqrt{(450-200)^2 + (100-30)^2}$$
$$d_3 = 259.62$$

- Liam Cruz

$$d_1 = \sqrt{(100-80)^2 + (8-5)^2}$$
$$d_1 = 20.22$$
$$d_2 = \sqrt{(100-350)^2 + (8-80)^2}$$
$$d_2 = 260.16$$
$$d_3 = \sqrt{(100-200)^2 + (8-30)^2}$$
$$d_3 = 102.39$$

- Ella Navarro

$$d_1 = \sqrt{(240-80)^2 + (38-5)^2}$$
$$d_1 = 163.37$$
$$d_2 = \sqrt{(240-350)^2 + (38-80)^2}$$
$$d_2 = 117.75$$
$$d_3 = \sqrt{(240-200)^2 + (38-30)^2}$$
$$d_3 = 40.79$$

- John Mercado

$$d_1 = \sqrt{(120-80)^2 + (10-5)^2}$$
$$d_1 = 40.31$$
$$d_2 = \sqrt{(120-350)^2 + (10-80)^2}$$
$$d_2 = 240.42$$
$$d_3 = \sqrt{(120-200)^2 + (10-30)^2}$$
$$d_3 = 82.46$$

- Jared Flores

$$d_1 = \sqrt{(260-80)^2 - (42-5)^2}$$
$$d_1 = 183.76$$
$$d_2 = \sqrt{(260-350)^2 - (42-80)^2}$$
$$d_2 = 97.69$$
$$d_3 = \sqrt{(260-200)^2 - (42-30)^2}$$
$$d_3 = 61.89$$

- Mark Santos

$$d_1 = \sqrt{(95-80)^2 + (7-5)^2}$$
$$d_1 = 15.13$$
$$d_2 = \sqrt{(95-350)^2 + (7-80)^2}$$
$$d_2 = 265.24$$
$$d_3 = \sqrt{(95-200)^2 + (7-30)^2}$$
$$d_3 = 107.49$$

- Paula Gomez

$$d_1 = \sqrt{(200-80)^2 + (30-5)^2}$$
$$d_1 = 122.58$$
$$d_2 = \sqrt{(200-350)^2 + (30-80)^2}$$
$$d_2 = 158.11$$
$$d_3 = \sqrt{(200-200)^2 + (30-30)^2}$$
$$d_3 = 0$$

- Hannah Rogue

$$d_1 = \sqrt{(420-80)^2 + (95-5)^2}$$
$$d_1 = 351.71$$
$$d_2 = \sqrt{(420-350)^2 + (95-80)^2}$$
$$d_2 = 71.59$$
$$d_3 = \sqrt{(420-200)^2 + (95-30)^2}$$
$$d_3 = 229.40$$

③ Assign a cluster for all data points

CUSTOMER NAME	MONEY SPENT	MINUTES INSIDE	d_1	d_2	d_3	CLUSTER
Chloe Mendoza	380	85	310.48	30.41	188.22	RICH
Anna Reyes	80	5	0	280.22	122.58	POOR
Mika Tan	280	45	203.96	78.26	81.39	RICH
Zach Uy	400	90	331.10	50.99	208.81	RICH
Kevin Ramos	220	35	143.18	137.57	20.62	MIDDLE
Sofia Dela Peña	110	9	30.27	250.28	92.42	POOR
Brian Lim	350	80	280.22	0	158.11	RICH
Caleb Ong	450	100	382.0	101.98	259.62	RICH
Liam Cruz	100	8	20.22	260.16	102.39	POOR
Ella Navarro	240	38	163.37	117.75	40.79	MIDDLE
John Mercado	120	10	40.31	240.42	82.46	POOR
Jared Flores	260	42	183.76	97.69	61.89	MIDDLE
Mark Santos	95	7	15.3	265.24	107.49	POOR
Paula Gomez	200	30	122.58	158.11	0	MIDDLE
Hannah Roque	420	95	351.71	71.59	229.40	RICH

④ Calculate the mean of each cluster

$$a_2 = \frac{(380 + 280 + 400 + 350 + 450 + 420)}{6} = 380$$

New centroid for RICH: (380, 82.5)

$$b_2 = \frac{(85 + 45 + 90 + 80 + 100 + 95)}{6} = 82.5$$

$$a_1 = \frac{(80 + 110 + 100 + 120 + 95)}{5} = 101$$

New centroid for POOR: (101, 7.8)

$$b_1 = \frac{(5 + 9 + 8 + 10 + 7)}{5} = 7.8$$

$$a_3 = \frac{(220 + 240 + 260 + 200)}{4} = 230$$

New centroid for MIDDLE: (230, 36.25)

$$b_3 = \frac{(35 + 38 + 42 + 30)}{40} = 36.25$$

5.) Recalculate the distance of all data points from the mean using Euclidean Distance

- **Chloe Mendoza**

$d_1 = \sqrt{(380-101)^2 + (85-7.8)^2}$

$d_1 = 289.48$

$d_2 = \sqrt{(380-380)^2 + (85-82.5)^2}$

$d_2 = 2.5$

$d_3 = \sqrt{(380-230)^2 + (85-36.25)^2}$

$d_3 = 157.72$

- **Anna Reyes**

$d_1 = \sqrt{(80-101)^2 + (5-7.8)^2}$

$d_1 = 21.19$

$d_2 = \sqrt{(80-380)^2 + (5-82.5)^2}$

$d_2 = 309.85$

$d_3 = \sqrt{(80-230)^2 + (5-36.25)^2}$

$d_3 = 153.22$

- **Mika tan**

$d_1 = \sqrt{(280-101)^2 + (45-7.8)^2}$

$d_1 = 182.82$

$d_2 = \sqrt{(280-380)^2 + (45-82.5)^2}$

$d_2 = 106.80$

$d_3 = \sqrt{(280-230)^2 + (45-36.25)^2}$

$d_3 = 50.76$

- **Zach Uy**

$d_1 = \sqrt{(400-101)^2 + (90-7.8)^2}$

$d_1 = 310.09$

$d_2 = \sqrt{(400-380)^2 + (90-82.5)^2}$

$d_2 = 21.36$

$d_3 = \sqrt{(400-230)^2 + (90-36.25)^2}$

$d_3 = 178.29$

- **Kevin Ramos**

$d_1 = \sqrt{(220-101)^2 + (35-7.8)^2}$

$d_1 = 122.07$

$d_2 = \sqrt{(220-380)^2 + (35-82.5)^2}$

$d_2 = 166.90$

$d_3 = \sqrt{(220-230)^2 + (35-36.25)^2}$

$d_3 = 10.08$

- **Sofia Dela Peña**

$d_1 = \sqrt{(110-101)^2 + (9-7.8)^2}$

$d_1 = 9.08$

$d_2 = \sqrt{(110-380)^2 + (9-82.5)^2}$

$d_2 = 279.83$

$d_3 = \sqrt{(110-230)^2 + (9-36.25)^2}$

$d_3 = 123.06$

- **Brian Lim**

$d_1 = \sqrt{(350-101)^2 + (80-7.8)^2}$

$d_1 = 259.26$

$d_2 = \sqrt{(350-380)^2 + (80-82.5)^2}$

$d_2 = 30.10$

$d_3 = \sqrt{(350-230)^2 + (80-36.25)^2}$

$d_3 = 127.73$

- **Caleb Ong**

$d_1 = \sqrt{(450-101)^2 + (100-7.8)^2}$

$d_1 = 360.97$

$d_2 = \sqrt{(450-380)^2 + (100-82.5)^2}$

$d_2 = 72.15$

$d_3 = \sqrt{(450-230)^2 + (100-36.25)^2}$

$d_3 = 229.05$

- Liam Cruz

$d_1 = \sqrt{(100-101)^2 + (8-7.8)^2}$

$d_1 = 1.02$

$d_2 = \sqrt{(100-380)^2 + (8-82.5)^2}$

$d_2 = 289.74$

$d_3 = \sqrt{(100-230)^2 + (8-36.25)^2}$

$d_3 = 133.03$

- Ella Navarro

$d_1 = \sqrt{(240-101)^2 + (38-7.8)^2}$

$d_1 = 142.24$

$d_2 = \sqrt{(240-380)^2 + (38-82.5)^2}$

$d_2 = 146.90$

$d_3 = \sqrt{(240-230)^2 + (38-36.25)^2}$

$d_3 = 10.15$

- John Mercado

$d_1 = \sqrt{(120-101)^2 + (10-7.8)^2}$

$d_1 = 19.13$

$d_2 = \sqrt{(120-380)^2 + (10-82.5)^2}$

$d_2 = 269.92$

$d_3 = \sqrt{(120-230)^2 + (10-36.25)^2}$

$d_3 = 113.09$

- Jared Flores

$d_1 = \sqrt{(260-101)^2 + (42-7.8)^2}$

$d_1 = 162.64$

$d_2 = \sqrt{(260-380)^2 + (42-82.5)^2}$

$d_2 = 126.650$

$d_3 = \sqrt{(260-230)^2 + (42-36.25)^2}$

$d_3 = 30.55$

- Mark Santos

$d_1 = \sqrt{(95-101)^2 + (7-7.8)^2}$

$d_1 = 6.05$

$d_2 = \sqrt{(95-380)^2 + (7-82.5)^2}$

$d_2 = 294.83$

$d_3 = \sqrt{(95-230)^2 + (7.-36.25)^2}$

$d_3 = 138.13$

- Paula Gomez

$d_1 = \sqrt{(200-101)^2 + (30-7.8)^2}$

$d_1 = 101.46$

$d_2 = \sqrt{(200-380)^2 + (30-82.5)^2}$

$d_2 = 187.5$

$d_3 = \sqrt{(200-230)^2 + (30-36.25)^2}$

$d_3 = 30.64$

- Hannah Roque

$d_1 = \sqrt{(420-101)^2 + (95-7.8)^2}$

$d_1 = 330.70$

$d_2 = \sqrt{(420-380)^2 + (95-82.5)^2}$

$d_2 = 41.91$

$d_3 = \sqrt{(420-230)^2 + (95-36.25)^2}$

$d_3 = 198.88$

⑥ Assign a cluster for all data points

CUSTOMER NAME	MONEY SPENT	MINUTES INSIDE	d₁	d₂	d₃	CLUSTER
Chloe Mendoza	380	85	289.48	2.5	157.72	RICH
Anna Reyes	80	5	21.19	309.85	153.22	POOR
Mika tan	280	45	182.82	106.80	50.76	MIDDLE
Zach Uy	400	90	310.09	21.36	178.29	RICH
Kevin Ramos	220	35	122.07	166.90	10.08	MIDDLE
Sofia Dela Peña	110	9	9.08	279.83	123.06	POOR
Brian Lim	350	80	259.26	30.10	127.73	RICH
Caleb Ong	450	100	360.97	72.15	229.05	RICH
Liam Cruz	100	8	1.02	289.74	133.03	POOR
Ella Navarro	240	38	142.24	146.90	10.15	MIDDLE
John Mercado	120	10	19.13	269.92	113.09	POOR
Jared Flores	260	42	162.64	126.65	30.55	MIDDLE
Mark Santos	95	7	6.05	294.83	138.13	POOR
Paula Gomez	200	30	101.46	187.5	30.64	MIDDLE
Hannah Roque	420	95	330.70	41.91	198.88	RICH

⑦ Plot the final cluster in a scatter plot