Introduction to Data Science

Name:_____

QUIZ 3

SCORE = _____/25

1. A Greek food truck with a limited menu would like to assess the popularity of its couscous salad. In particular, it would like to evaluate the association rule: {*Chicken Gyro*} → {*Couscous Salad*}. Using this small data set of 12 visitors, do a market basket analysis of this rule.

Chicken Gyro	Lamb Gyro	Grilled Cheese	Feta Tomato Wrap	Stuffed Grape Leaves	Hummus & Pita	Greek Salad	Couscous Salad	Spinach Pie	chicken + Couscous
1	1	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	1	0	1
0	1	0	0	0	1	0	0	0	0
0	0	1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	1	1
0	0	0	0	0	1	0	1	0	0
0	0	1	0	0	0	1	0	0	0
0	0	1	0	0	0	0	1	0	0
0	1	0	1	1	0	0	0	0	0
1	0	0	0	0	0	0	1	1	1
1	0	0	0	1	0	0	1	0	1
0	1	0	0	0	0	0	1	0	0
5							7		4

a. Create a column in the table above for 'Chicken Gyro and Couscous Salad', as we've done in class. (1 pt)

b. Calculate the needed quantities, as we did in the Association Rules activity, and state values below. Show work for each calculation. (4 pts)

Rule #1	{*Chicken Gyro*} → {*Couscous Salad*}

Chicken Gyro: Number of times purchased: 5 Support .4167 ≈ 5/12

Couscous Salad: Number of times purchased: 7 Support .5833 7/12

Chicken Gyro and Couscous Salad: Number of times purchased: 4 Support .3333 4/12

Confidence of the Rule: $\frac{4/12}{5/12} = 4/5 = .8$

Lift of the Rule: $.8 / .583 \approx 1.371$

c. Complete the sentence below, in terms of the Chicken Gyro and Couscous Salad association rule. *(2 pts)*

The Lift value means that we are 1.371 *times more likely to believe*

they'll buy couscous given they bought gyro

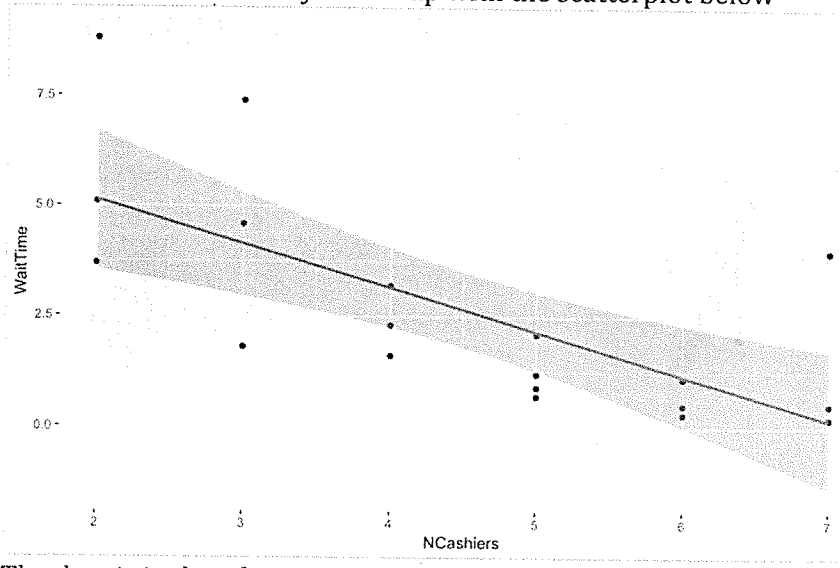
Compared to if we didn't know about their gyro purchase

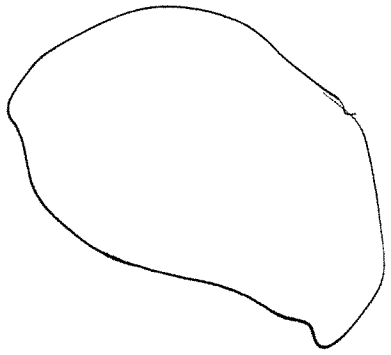
(= if we had no information on gyro purchase)

Introduction to Data Science

Name:_____

2. A bookstore wondered how much customer wait times would be reduced by having more cashiers. Each hour, they randomly assigned whether to have 2, 3, 4, 5, or 6 cashiers, then they recorded the average wait times for the hour. They ended up with the scatterplot below





a. The data is in data frame ~~**Wait.**~~ Write the code to make the plot above, using dplyr/ggplot: (3 pts)

```
ggplot( data = Wait,
        mapping = aes(x = NCashiers, y = WaitTime))
  geom_point() +
  geom_smooth(method = 'lm')
```

b. Looking at the plot, the correlation coefficient is closest to (circle one): -.9 **(-.7)** -.3 0 .3 .7 .9 (1 pt)

c. The output below is most of the summary of the linear model object **regwait**. Write the line of R code that would create **regwait**. (2 pts)

regwait ← lm(Wait$WaitTime ~ Wait$NCashiers)

> summary(regwait)

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.3515 -1.1099 -0.0641  0.1880  3.8317

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.1389     1.2001   5.949 1.59e-05
wt$NCashiers  -0.9958     0.2488  -4.002 0.000923
```

y ~ x
↑ explained by

d. Using values in the output above, write the estimated linear regression equation, in terms of this problem, as we've done in class: (1 pts)

Predicted Wait Time = 7.1389 − .9958(NCashiers)

Name:_____

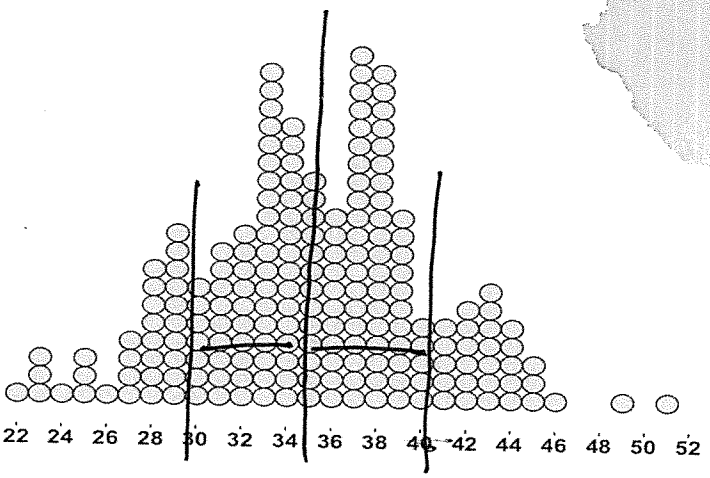
e. Using your equation in d, predict Wait Time for a day with 3 cashiers. Show your work below: (1 pt)

7.1389 − .9958 * 3 = 4.1515

f. Finally, using the value of the slope that you wrote down in part d, interpret it in terms of this problem, as we've done in class: "For each ...." (2 pts)

For each additional cashier, there's an associated decrease in wait time of .9958 minutes

3. Visually estimate the mean and standard deviation of the data set below, showing your work on the plot: (2 pts)



mean ≈ 35

sd ≈ 5

(4-6)

4. Hunchlab and Predpol are used for predictive policing (1 pt)
Northpointe is an algorithm for predictive sentencing (1 pt)

5. Describe a primary **advantage** of the algorithms noted in question 4, and a primary **disadvantage**. (4 pts)

Advantage: Main advantage is to remove human bias
-- Algorithm gives same answers to everyone.
(Also, cheap!)

Disadvantage: Targets past crime locations, thus profiling.
Not tested! Not sure of accuracy