

Ashwin Gopal  
Class: DAT565  
Instructor: Kenneth Le Cour

## Week 2 Predictive and Quantitative Analysis Report

### Scope and descriptive statistics:

Pastas R Us Inc. has been the go-to restaurant for many people. The company's business development team has been trying to open new restaurants in areas with target demographics such as median age from 25 to 45 years old, Household median income being above the national wage and at least fifteen percent of adults being college-educated within a three-mile radius. Our company also wants to analyze key available information to determine whether the current business strategy for the restaurant's expansion criteria and our loyalty card strategy is working. Furthermore, last year the marketing team released a loyalty card strategy to increase sales. To do this I will analyze the available data, then prepare a report on the analysis and my findings. The objective of this report is to analyze and draw insights about the effectiveness of the current expansion criteria, loyalty card program, and marketing position using the company's available data. So far our company has collected data from 74 of its restaurants to track important variables such as average sales per customer, year-to-year sales growth, sales per sq. ft., loyalty card usage as a percent of sales, and others. With this, I have analyzed and used descriptive statistics for two key sales variables. These include annual sales and annual sales per sq. ft. The variables might help us with making an effective financial performance strategy for the restaurant. To start first I calculated all of the annual sales values. Then I calculated the 5 summary values. Furthermore, I created a box plot chart for the annual sales variable and a histogram for the annual sales per sq. ft. variable. With this, I have made some observations on these variables. For annual sales, the box plot chart is not symmetric. This means that there are values that might be high but don't align with the minimum value. So for example the max value on the chart is \$1746 and the min value is \$500. It would be symmetric if the max value was \$1500. Another observation on annual sales is describing the dispersion. When describing this, standard deviation is preferred. This is because the chart is not symmetric and has no outliers. Now with annual sales per sq. ft distribution, the histogram shows that it's skewed right with outliers. The value of the skew is 1.2358. There is one outlier, in the range of 948.56 sales per sqft or greater. The value is 987.12 sales per sqft., and the SqFt of the outlier is 1251. Finally, another observation on annual sales per sq. ft is describing the appropriate measure of central tendency. In my opinion, it's median. The reason is the median is usually unaffected when there are outliers.

### Analysis:

Apart from these observations, there are some observations on scatter plots created over variables such as sales per sq ft, college graduates, median age, median income, and sales growth. The purpose of scatter plots is to find patterns, trends, and key associations between variables. Now let's look at the scatter plot for sales per sq. ft versus college graduates. Some trends I noticed are, as the percentage of bachelor degrees college graduates hold increases the number of sales per sq. ft increases. This means, there is an increasing and positive relationship. This is determined by looking at the trend line and the linear equation. The linear equation helps determine whether the trend line is increasing or decreasing. For example, if the linear equation's slope value is 6.699 it's positive, this means the trend line is increasing and positive. The R squared value is a statistical measurement that indicates how much of a variation of a dependent variable is explained by an independent variable in a regression model(FERNANDO, 2023). For example, based on the scatter plot for sales per sq. ft versus college graduates if the R squared value is 0.1169 then this means there's only a 10% variation of the data. With the scatter plot for sales per sq. ft versus median age, the trend is that there might be no relationship. This may be due to the outliers or observation of ages between 30 to 40 having the highest sales. So in this case even though the equation indicates the trend line is decreasing and negative, some patterns were determined. So this plot lets us determine the target age for the highest amount of sales. With the scatter plot for sales per sq. ft versus median income, the trend here is that there is not a proper relationship between the variables. The reason is the slope value is -0.0002, which is closer to 0. The conclusion here is that the majority of the sales occurred between an income of \$50,000 and \$100,000. With the scatter plot for loyalty cards versus sales growth, there is a negative and decreasing relationship. This means as the rate of the loyalty cards increases the rate of sales growth decreases. From this, I concluded that the majority of the sales growth is concentrated at a loyalty rate of 2%

### **Recommendations and implementation:**

Now based on the findings in the scatter plots, the expansion criteria that seem to be more effective is at least 15% of the adult population is college-educated. The criteria that could be changed is the median age. The reason is, that the highest concentration of sales is from age 30 to 40. If we change it to that we might see some effect. Based on the findings, the loyalty card appears to be negatively correlated with sales growth. From this, I would recommend changing the loyalty card rate between 1.0% to 2.0. Based on my findings, I would recommend giving discounts for ages 20 to 30. In terms of local culture and communities, the younger ages might choose based on price, quality, and safety. With this, some information that should be collected is the number of people purchased by price and age, food quality rating, and safety inspection rating. This is collected through surveys.

References:

FERNANDO, J. (2023, April 8). *R-Squared: Definition, Calculation Formula, Uses, and Limitations*. Investopedia. <https://www.investopedia.com/terms/r/r-squared.asp>