- **Introduction: Introduce the problem, provide some background**
  - ➢ Our boss wants a descriptive analysis to understand what drives the performance of competing stores of the Subway franchise. We acquired data from last year's sales and the demographic factors of each store. We have 27 Subway store observations.

- **Objectives:**
  - ➢ To determine and understand the relevant sales performance of Subway franchise competing stores using descriptive analytics, in order for management to utilize these findings to implement future sales efforts. This descriptive analysis is based on sales performance data from 2019 (prior year) sales and the relevant demographic factors of 27 observations Store with data on five independent variables. The demographic factors include: square feet (SQFT), inventory, advertising, families and competing stores location (categories).

- **Problem Statement: describe the problem you are proposing to solve**
  - ❖ What drives the performance of competing stores of the Subway Franchise?
  - ❖ How do the individual factors affect our sales in the franchises?
  - ❖ How can the store achieve a certain level of sales?
  - ❖ Which factors are critical to the design of the store meaning: area square feet, inventory, advertising spend, size of the sale district, number of competing stores
  - ➢ As our team looks at last year's data from the Subway franchise, we look for what drives our store to what affects it can create. We look at the individual factors

- **State the assumptions (if any)**
  - ➢ Negative assumptions sales performance decreases: Competing stores will have a significant impact on sales performance at Subway stores; as it is demonstrated by a strong negative correlation with a P-value (3.5159E-11)

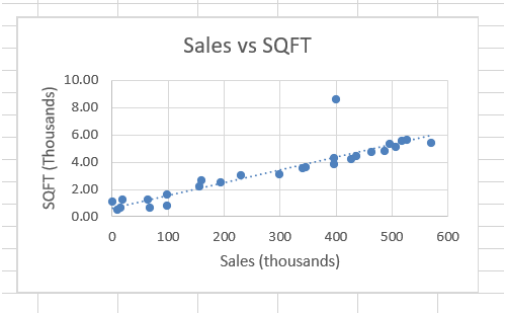
- **Data Sources: list the sources of data you expect to extract data for developing the analysis.**
  - ➢ The sources of our data will all be from the excel worksheet. We will be extracting correlation between two variables sales vs. one of the demographic factors. We will also conduct a simple linear regression model and a multivariate regression model.

- **Data Description: describe the data (explain any terms that are specific to your dataset**
  - ➢ The demographic factors describe the following: Sales is the annual net sales/$1000, square feet is the number of sq.ft/1000, inventory is the inventory/$1000, advertising is the amount spent on advertising/$1000, families is the size of sales district/1000 families and competing stores is the number of competing stores in the district.

- **Methods:**
  - ➢ **Discuss your modelling approach, what are the independent variables (co-variates), what is the dependent variable.**
    - ○ Our dependent variable is sales (Y), and our independent variables (X) are the demographic factors. Our co-variates is square feet, inventory, advertising, families and competing stores.

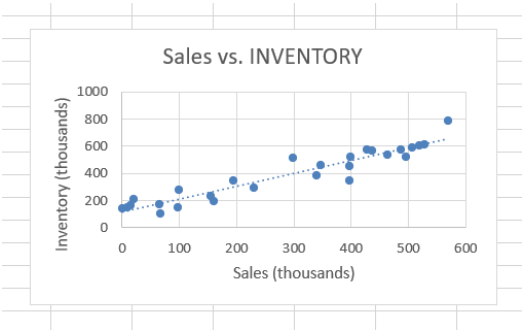
➢ Simple Regression Models: One dependent vs. one independent

➢ **Sales vs. Square Feet (SQFT)**



- **Residual Equation:** $\hat{y} = 85.389x + 2.577$
- **The correlation (Multiple R) is 89.41% this explains the degree of linear relationship between Sales and Square Feet.**
- **The adjusted r-square shows that the proportion that 79.14 is the variation between the dependent variable and the independent variable.**

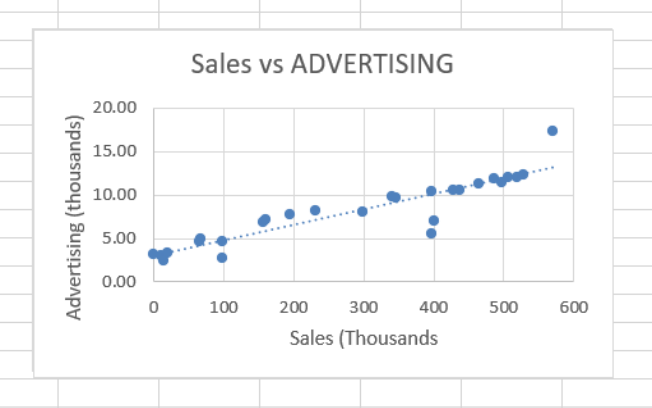
➢ **Sales vs. Inventory**



- **Residual Equation:** $\hat{y} = (-81.504) + 0.9499x$
- **The correlation (Multiple R) is 94.55% this explains the degree of linear relationship between Sales and Inventory.**
- **The Adjusted R-square shows that the proportion that 88.97 is the variation between the dependent variable and the independent variable.**

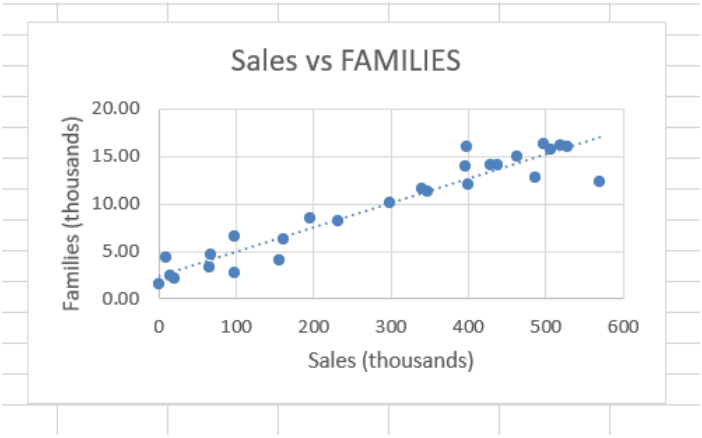
➢

➢ **Sales vs. Advertising**



- **Residual Equation:** $\hat{y} = (-90.1496) + 46.509x$
- **The correlation (Multiple R) is 91.40% this explains the degree of linear relationship between Sales and Advertising.**
- **The Adjusted R-square shows that the proportion that 82.89 is the variation between the dependent variable and the independent variable.**

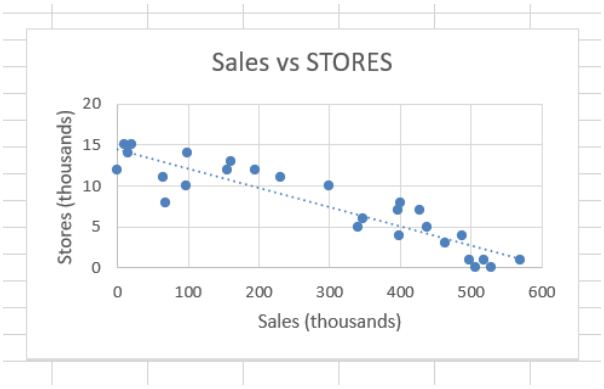
➢ **Sales vs. Families**



- **Residual Equation:** $\hat{y} = 35.635x - 58.823$
- **The correlation (Multiple R) is 95.37% this explains the degree of linear relationship between Sales and Families.**
- **The Adjusted R-square shows that the proportion that 90.95 is the variation between the dependent variable and the independent variable.**

➢

➢ **Sales vs. Stores**



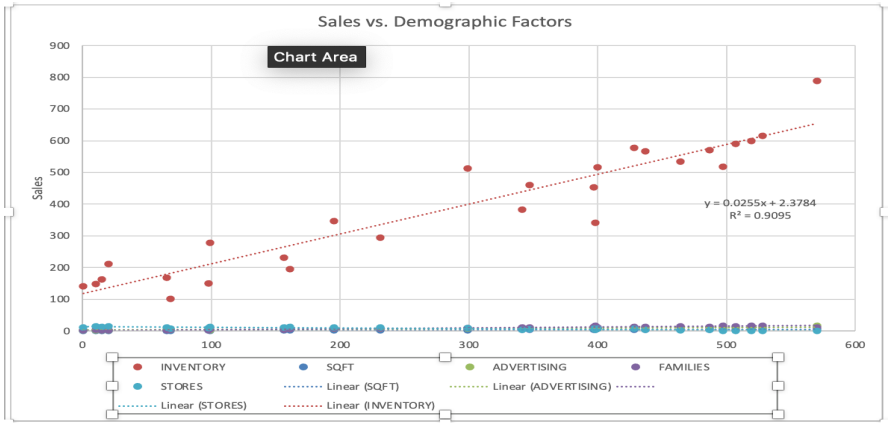
	Coefficients	*P-value*
Intercept	563.5926245	1.66993E-16
STORES	-35.78708546	3.51586E-11

- **Residual Equation:** $\hat{y} = -35.787x + 563.59$
- **The correlation (Multiple R) is 91.22% this explains the degree of linear relationship between Sales and competing stores.**
- **The Adjusted R-square shows that the proportion that 82.55 is the variation between the dependent variable and the independent variable.**

Outcome of all SLR Models:

- **Each regression line shows a strong positive linear relationship between sales and the demographic factors in exception with stores. There is a strong linear relationship between sales and stores, as the area of each store occupies an increase, sales will also increase. There is a strong positive linear relationship between sales and inventory, as the amount of items in each store increases, sales will also increase. There is a strong positive linear relationship between sales and advertising, as the amount of money each store spends on promotion increases, sales will also increase. There is a strong positive linear relationship between sales and families, as the size of the customer base around each competing store increases, sales will also increase. The regression line shows a strong negative linear relationship between sales and competing stores, as the number of competing stores near by each Subway store in our franchise increases, sales will decrease.**
  - **Null Ho=B1=0 (slope is equal to 0)**
  - **Alternative H1=B1≠0**
  - **At 95% significance level the P-Values for all the demographic factors are less than the alpha value of (0.05) which therefore allows us to reject the null hypothesis. There is sufficient evidence to show that there is a significant relationship between sales and all the demographic factors.**

Sales vs. All Demographic Factors



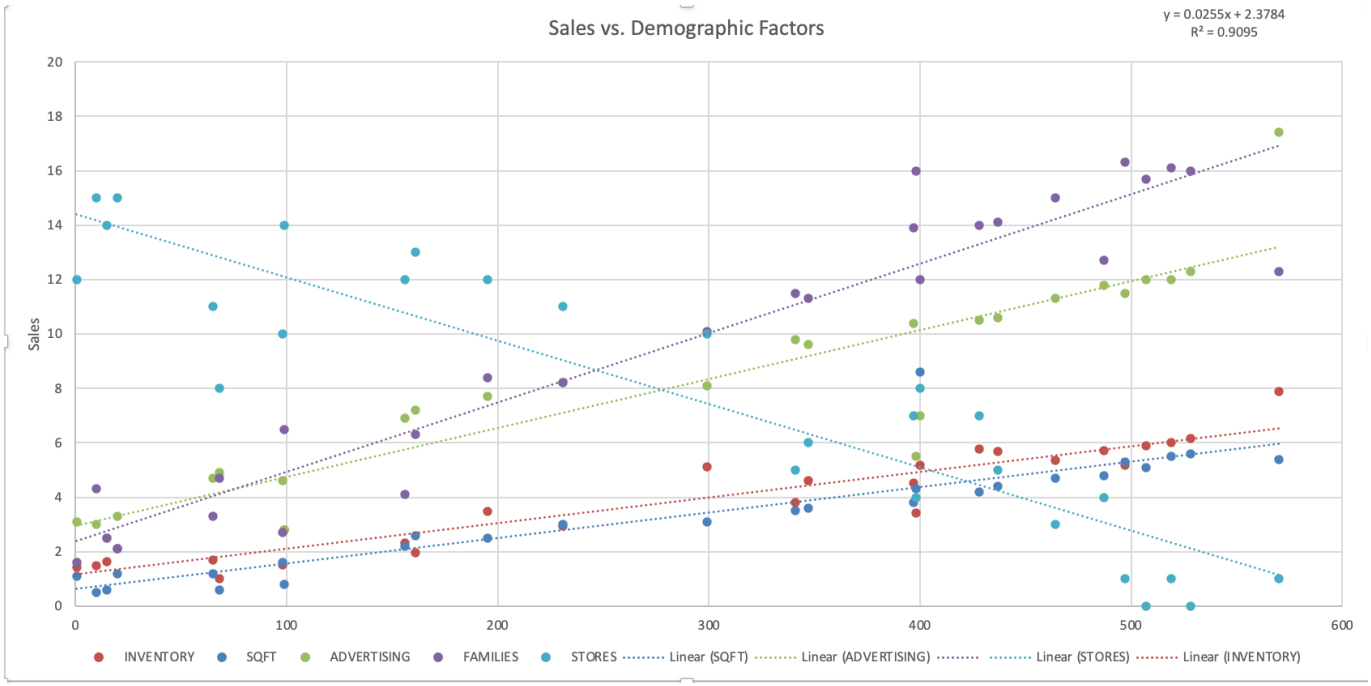
Regression Statistics	
Multiple R	0.996583914
R Square	0.993179497
Adjusted R Square	0.991555568
Standard Error	17.64924165
Observations	27

	Coefficients	Standard Error	t Stat	P-value
Intercept	-18.85941416	30.15022791	-0.62551481	0.538372333
SQFT	16.20157356	3.544437306	4.570986073	0.000165985
INVENTORY	0.174635154	0.057606068	3.031540961	0.006346793
ADVERTISING	11.52626903	2.5321033	4.55205324	0.000173652
FAMILIES	13.5803129	1.770456609	7.670514392	1.60543E-07
STORES	-5.31097141	1.70542654	-3.11416017	0.005248873

- **Residual Equation:**
  $\hat{y} = -18.859 + 16.202(SQFT) + .1746(Inventory) + 11.526(Advertising) + 13.580(Families$
- **The group finds that the intercept shows no economic sense for projected sales, it is much more on a mathematical coefficient for the residual equation.**
- **The correlation (Multiple R) is 99.66% this explains the degree of linear relationship between Sales and Inventory.**
- **The Adjusted R-square shows that the proportion that 99.16 is the variation between the dependent variables and the independent variable.**
- **The regression line shows a strong linear relationship between square feet, inventory, advertising, families and sales; it also shows a negative linear relationship between competing stores and sales.**
- **Null Ho=B1=0 (slope is equal to 0) Alternative H1=B1≠0**
- **At 95% significance level our P-Values for the demographic factors (dependent variables) are less than our alpha value of (0.05) which therefore allows us to reject the null hypothesis. There is sufficient evidence to show that there is a significant relationship between sales and square feet, inventory, advertising, families and competing stores.**

- **Did you need to transform any variables? No, we found that it was just necessary to adjust our variables rather than to transform.**

The inventory variable had been previously divided by $10,000. After running the numbers, it was found that the residual equation showed there was no major difference in the projected sales as compared to the below graph. As a group we found it was best to adjust inventory by dividing it by $100,000. This shows the ultimate trendlines, and it goes coherently to the regression models and the correlations. In the diagram below, it will show the adjustment, since inventory had higher values than the other factors, this graph shows what we got from the regression models. In the diagram, families had a higher stronger linear relationship than any other factors.



- **We just opened a store in a neighborhood with 5,000 families, the store is 5,000 sq ft, we are planning to spend $5,000 a month in advertising, carry $250,000 in inventory and there are 5 competing stores in the neighborhood.**
  - **What are the projected sales?**

^y = projected sales

Inventory/$1,000

$\hat{y} = -18.859 + 16.202(SQFT) + .1746(Inventory) + 11.526(Advertising) + 13.580(Families) -$

Projected Sales $206,538.24 the variable of inventory 250,000 for inventory

Inventory/$100,000

$\hat{y} = -18.859 + 16.202(SQFT) + 17.464(Inventory) + 11.526(Advertising) + 13.580(Families)$

Projected Sales $ 206,538.25 changing the variable of inventory to 2.5 due to the adjustment (250,000/100,000).

- **How well do the variables fit the model? How well do the variables describe the model?**
  - ➢ After adjusting the variable (inventory), the variables now fit the model more accordingly. Inventory will have a higher value because it's the way that stores do business. Having high inventory is better than having a shortage or out of stock which leads to having more sales.

- **Results: What is the outcome of your analysis across the different models?**
  - ➢ The outcome of our analysis across the different models is that families have the highest contributing (strong) positive correlation to sales. Throughout the trendlines and scatter plots and the regression models, families have a significant say in the projected sales. In addition the p-value for families, had the lowest in comparison with the remaining demographic factors. When the p-value is low this shows that the t-statistic was a high value, which indicates that there was a high significance level between the dependent and independent variables. Competing stores have the highest contributing (strong) negative correlation to sales. As per assumption, when organizing the data, the data concluded that competing stores did have a strong negative correlation with the projected sales. In addition, Advertising and Inventory had a significant correlation with sales but not as high a correlation as families did.

- **Conclusion, Recommendations are as follows:**
  - ➢ Per our findings, management should focus on Competing Stores and Families (the customer base's size) around Subway stores. In this manner, management can add more focus to advertising as it is shown to increase Sales. Management should also keep the number of competing stores nearby each store in our franchise at a low. Since there is a negative relationship between competing stores and sales, look for a location with the least competing stores around.