**Statistical Hypothetical question Summary**

* **Outcome of your EDA**

○ **Context:** A retail firm has many products in their inventory, and very few of them tend to sell (only about 10% sell each year) and many of the products only have a single sale in the course of a year

○ **Objective:** The sales and growth team of the retail firm wants to determine which products from their inventory should they retain to sell and the ones to discard. The outcome of the project is to build a binary classifier which gives us a list of product ID which need to retained in the inventory or list of products that need to be removed.

* **What do you feel was missed during the analysis?**

○ I feel like I have missed using some more data modeling and prediction techniques like SVM and deep learning methods.

* **Were there any variables you felt could have helped in the analysis?**

The bivariate distribution plots help us to study the relationship between two variables by analyzing the scatter plot, I have used this distribution to look out for scatter plots that follow a clear linear pattern with an either increasing or decreasing slope so that we can draw conclusions, but don’t notice these patterns in this particular dataset.

* **Were there any assumptions made you felt were incorrect?**

No

* **What challenges did you face, what did you not fully understand?**

○ I could not fully understand the use other descriptive characteristics like Mean, Mode, Spread, and Tails in my dataset so I did not consider it in my analysis.

○ Initial challenge was to select the right variables for my analysis.

○ From technical perspective I had a challenge understanding random forest classifier for my dataset.